22nd Edition

HARRISON'S®

Principles of Internal Medicine

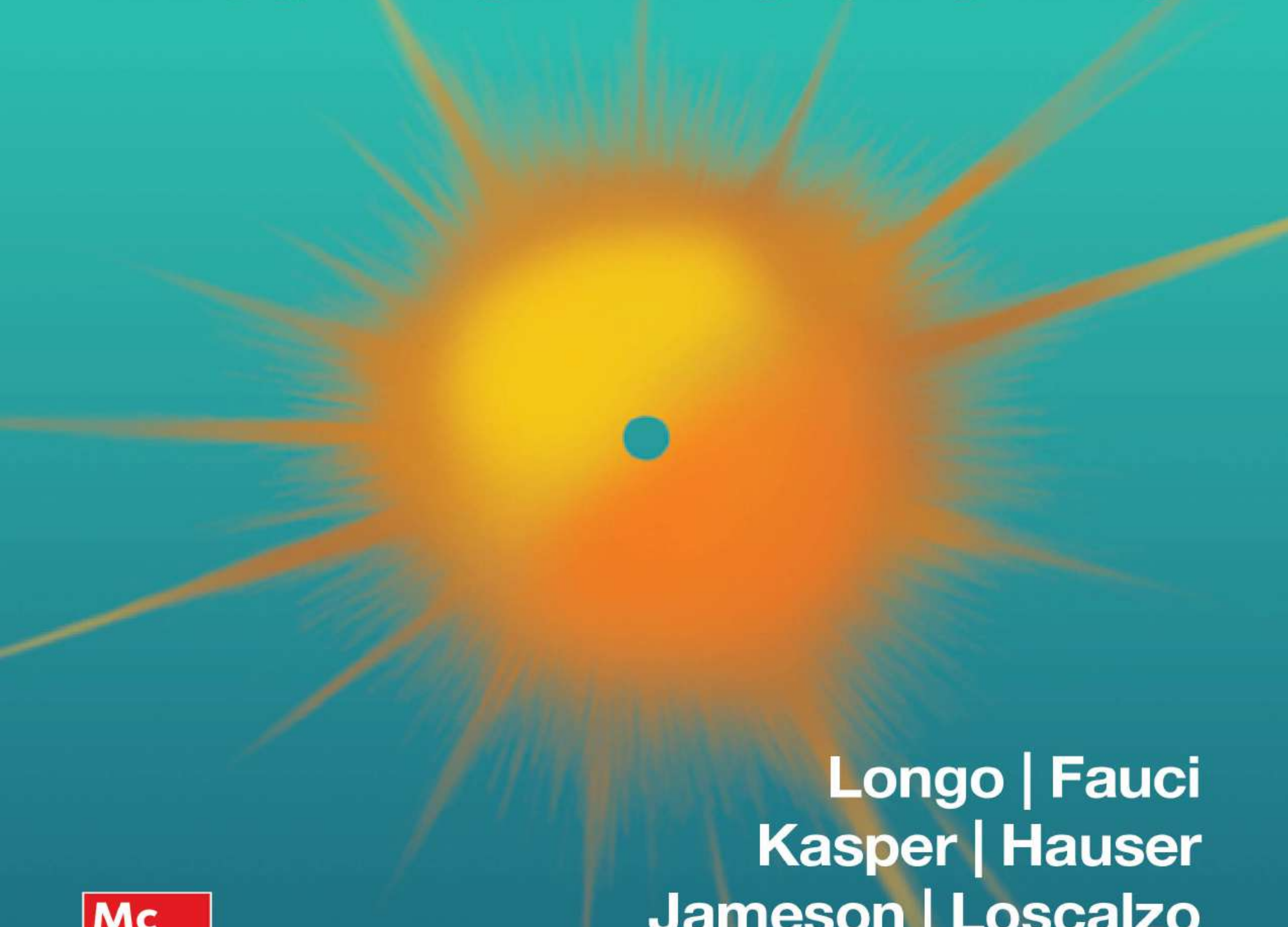

Longo | Fauci
Kasper | Hauser
Jameson | Loscalzo
Holland | Langford

McGraw Hill

22nd Edition

HARRISON'S®
PRINCIPLES OF
INTERNAL MEDICINE

Editors of Current and Previous Editions

T. R. Harrison
Editor-in-Chief, Editions 1, 2, 3, 4, 5

W. R. Resnick
Editor, Editions 1, 2, 3, 4, 5

M. M. Wintrobe
Editor, Editions 1, 2, 3, 4, 5
Editor-in-Chief, Editions 6, 7

G. W. Thorn
Editor, Editions 1, 2, 3, 4, 5, 6, 7
Editor-in-Chief, Edition 8

R. D. Adams
Editor, Editions 2, 3, 4, 5, 6, 7, 8, 9, 10

P. B. Beeson
Editor, Editions 1, 2

I. L. Bennett, Jr.
Editor, Editions 3, 4, 5, 6

E. Braunwald
Editor, Editions 6, 7, 8, 9, 10, 12, 13, 14, 16, 17
Editor-in-Chief, Editions 11, 15

K. J. Isselbacher
Editor, Editions 6, 7, 8, 10, 11, 12, 14
Editor-in-Chief, Editions 9, 13

R. G. Petersdorf
Editor, Editions 6, 7, 8, 9, 11, 12
Editor-in-Chief, Edition 10

J. D. Wilson
Editor, Editions 9, 10, 11, 13, 14
Editor-in-Chief, Edition 12

J. B. Martin
Editor, Editions 10, 11, 12, 13, 14

A. S. Fauci
Editor, Editions 11, 12, 13, 15, 16, 18, 19, 20, 21, 22
Editor-in-Chief, Editions 14, 17

R. Root
Editor, Edition 12

D. L. Kasper
Editor, Editions 13, 14, 15, 17, 18, 20, 21, 22
Editor-in-Chief, Editions 16, 19

S. L. Hauser
Editor, Editions 14, 15, 16, 17, 18, 19, 20, 21, 22

D. L. Longo
Editor, Editions 14, 15, 16, 17, 19, 20, 21
Editor-in-Chief, Editions 18, 22

J. L. Jameson
Editor, Editions 15, 16, 17, 18, 19, 21, 22
Editor-in-Chief, Edition 20

J. Loscalzo
Editor, Editions 17, 18, 19, 20, 22
Editor-in-Chief, Edition 21

S. M. Holland
Editor, Edition 22

C. A. Langford
Editor, Edition 22

22nd Edition

HARRISON'S®

PRINCIPLES OF INTERNAL MEDICINE

Editors

Dan L. Longo, MD
Professor of Medicine, Harvard Medical School; Senior Physician, Brigham and Women's Hospital; Deputy Editor, *New England Journal of Medicine,* Boston, Massachusetts

Anthony S. Fauci, MD
Distinguished University Professor, Georgetown University School of Medicine and the McCourt School of Public Policy, Washington, DC

Dennis L. Kasper, MD
William Ellery Channing Professor of Medicine and Professor of Immunology, Department of Immunology, Harvard Medical School; Division of Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts

Stephen L. Hauser, MD
Robert A. Fishman Distinguished Professor, Department of Neurology; Director, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, California

J. Larry Jameson, MD, PhD
Robert G. Dunlop Professor of Medicine; President, University of Pennsylvania, Philadelphia, Pennsylvania

Joseph Loscalzo, MD, PhD
Hersey Distinguished Professor of the Theory and Practice of Medicine, Samuel A. Levine Professor of Medicine, Harvard Medical School; Physician-in-Chief Emeritus, Soma Weiss, MD, Honorable Endowed Chair in Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts

Steven M. Holland, MD
Bethesda, Maryland

Carol A. Langford, MD, MHS
Harold C. Schott Endowed Chair in Rheumatic and Immunologic Diseases; Professor of Medicine, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University School of Medicine; Director, Center for Vasculitis Care and Research, Department of Rheumatic and Immunologic Diseases, Cleveland Clinic, Cleveland, Ohio

VOLUME I

Mc
Graw
Hill

Note: Dr. Holland's work as editor and author was performed outside the scope of his employment as a U.S. government employee. This work represents his personal and professional views and not necessarily those of the U.S. government.

Harrison's®
Principles of Internal Medicine
Twenty-Second Edition

Copyright © 2025, 2022 by McGraw Hill LLC. Previous editions copyright © 2018, 2015 by McGraw-Hill Education; 2012, 2008, 2005, 2001, 1998, 1994, 1991, 1987, 1983, 1980, 1977, 1974, 1970, 1966, 1962, 1958 by The McGraw-Hill Companies, Inc. All rights reserved. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a data base or retrieval system, without the prior written permission of the publisher.

Harrison's® is a trademark of McGraw Hill.

$PrintCode

Two Volume Set ISBN 9781265973872; MHID 1265973873
Volume 1 ISBN 9781265969905; MHID 1265969906
Volume 2 ISBN 9781265971724; MHID 1265971722
eBook Two Volume Set ISBN 9781265979522; MHID 1265979529
Volume 1 eBook ISBN 9781265978037; MHID 1265978034
Volume 2 eBook ISBN 9781265979102; MHID 1265979103

FOREIGN LANGUAGE EDITIONS
Albanian: (17e): Tabernakul Publishing, Skopje, Macedonia; (20e, 21e): Life Shpk, Tirane
Arabic: (13e): McGraw-Hill Libri Italia srl (1996)
Azerbaijani: (21e): Gadim Gala Publishing
Chinese Long Form: (15e): McGraw-Hill International, Enterprises, Inc., Taiwan
Chinese Short Form: (15e, 19e): McGraw-Hill Education (Asia), Singapore; (21e): Peking University Medical Press
Croatian: (16e): Placebo, Split, Croatia
French: (16e, 18e): Medecine-Sciences Flammarion, Paris, France
Georgian: (19e): Tbilisi State Medical University, Tbilisi, Georgia
German: (17e, 18e, 19e, 20e): ABW Wissenschaftsverlagsgesellschaft GmbH, Berlin, Germany
Greek: (17e): Parissianos, S.A., Athens, Greece; (19e, 21e): Parisianou, S.A., Athens, Greece
Italian: (17e, 18e): The McGraw-Hill Companies, Srl, Milan, Italy; (19e, 20e): Casa Editrice Ambrosiana, Milan, Italy
Japanese: (17e, 18e, 19e): MEDSI-Medical Sciences International Ltd, Tokyo, Japan
Korean: (17e, 18e): McGraw-Hill Korea, Inc., Seoul, Korea
Macedonian: (17e): Tabernakul Publishing, Skopje, Macedonia
Polish: (17e): Czelej Publishing Company, Lubin, Poland
Portuguese: (17e, 18e): McGraw-Hill Interamericana Editores, SA de C.V., Mexico City, Mexico; (19e, 20e): AMGH Editora Ltda., Porto Alegre, Brazil; (21e): GA Educação S.A.
Romanian: (17e): Editura All, Bucharest, Romania; (19e): ALL Publishing House, Bucharest, Romania
Serbian: (15e): Publishing House Romanov, Bosnia & Herzegovina, Republic of Serbska; (19e): Data Status, Novi Beograd, Serbia
Spanish: (17e, 18e): McGraw-Hill Interamericana Editores, SA de C.V., Mexico City, Mexico; (19e, 20e): McGraw-Hill Mexico, Mexico City, Mexico; (21e): McGraw-Hill Interamericana Editores, S.A. de C.V.
Turkish: (17e, 19e, 20e): Nobel Tip Kitabevleri, Ltd., Istanbul, Turkey
Vietnamese: (15e): McGraw-Hill Education (Asia), Singapore

This book was set in Minion Pro by KnowledgeWorks Global Ltd. The editors were James F. Shanahan, Samantha C. Shapiro, MD, Kim J. Davis, and Katerina Heidhausen. The production manager was Jeffrey Herzich. The digital production managers were Priscilla Beer and Steve Brown. Project management was provided by Revathi Viswanathan, KnowledgeWorks Global Ltd. The index was prepared by Susan Hunter. The text designer was Janice Bielawa; the cover design was by the Velo Marketing Agency and Anthony Landi.

Library of Congress Cataloging-in-Publication Data

Names: Longo, Dan L. (Dan Louis), 1949- editor.
Title: Harrison's principles of internal medicine / editors, Dan L. Longo, Anthony S. Fauci, Dennis L. Kasper, Stephen L. Hauser, J. Larry Jameson, Joseph Loscalzo, Steven M. Holland, Carol A. Langford.
Other titles: Principles of internal medicine
Description: 22nd edition. | New York : McGraw Hill, [2025] | Includes bibliographical references and index. | Summary: "This text provides a sharp focus on the clinical presentation of disease, expert in-depth summaries of pathophysiology and treatment, and highlights of emerging frontiers of science and medicine"—Provided by publisher.
Identifiers: LCCN 2024049465 (print) | LCCN 2024049466 (ebook) | ISBN 9781265979317 (hardback ; set) | ISBN 9781265977061 (hardback ; v. 1) | ISBN 9781265978631 (hardback ; v. 2) | ISBN 9781265979522 (ebook ; set) | ISBN 9781265978037 (ebook ; v. 1) | ISBN 9781265979102 (ebook ; v. 2)
Subjects: MESH: Internal Medicine
Classification: LCC RC46 (print) | LCC RC46 (ebook) | NLM WB 115 | DDC 616—dc23/eng/20250408
LC record available at https://lccn.loc.gov/2024049465
LC ebook record available at https://lccn.loc.gov/2024049466

NOTICE

Medicine is an ever-changing science. As new research and clinical experience broaden our knowledge, changes in treatment and drug therapy are required. The authors and the publisher of this work have checked with sources believed to be reliable in their efforts to provide information that is complete and generally in accord with the standards accepted at the time of publication. However, in view of the possibility of human error or changes in medical sciences, neither the authors nor the publisher nor any other party who has been involved in the preparation or publication of this work warrants that the information contained herein is in every respect accurate or complete, and they disclaim all responsibility for any errors or omissions or for the results obtained from use of the information contained in this work. Readers are encouraged to confirm the information contained herein with other sources. For example and in particular, readers are advised to check the product information sheet included in the package of each drug they plan to administer to be certain that the information contained in this work is accurate and that changes have not been made in the recommended dose or in the contraindications for administration. This recommendation is of particular importance in connection with new or infrequently used drugs.

Cover Illustration

Beginning with the 6th edition, the cover of *Harrison's* has included an image of a bright light—a patient's perception of being examined with an ophthalmoscope. This allegorical symbol of *Harrison's* is a reminder of how the light of knowledge empowers physicians to better diagnose and treat diseases that ultimately afflict all of humankind.

Author Disclosure Policy: McGraw Hill and the *Harrison's* Editorial Board require all contributors to disclose to the Editors and the Publisher any potential financial or professional conflicts that would raise the possibility of distorting the preparation of a *Harrison's* chapter.

🧬 The genetic icons identify a clinical issue with an explicit genetic relationship.

Contents

Contributors ... xix

Preface ... xli

Related Harrison's Resources ... xlii

PART 1 The Profession of Medicine

1 The Practice of Medicine ... 1
The Editors

2 Promoting Good Health ... 8
Donald M. Lloyd-Jones, Kathleen M. McKibbin

3 Vaccine Opposition and Hesitancy ... 14
Julie A. Bettinger, Hana Mitchell

4 Decision-Making in Clinical Medicine ... 20
Daniel B. Mark, John B. Wong

5 Precision Medicine and Clinical Care ... 29
The Editors

6 Screening and Prevention of Disease ... 37
Gary J. Martin

7 The Safety and Quality of Health Care ... 42
David W. Bates

8 The Value of the Physical Examination in Modern Medicine ... 46
Abraham Verghese, Steven McGee

9 Physician Well-Being ... 52
Lisa M. Bellini, Vivek H. Murthy

10 Diagnosis: Reducing Errors and Improving Quality ... 56
Gordon Schiff

11 Racial and Ethnic Disparities in Health Care ... 61
Lenny López, Joseph R. Betancourt

12 Ethical Issues in Clinical Medicine ... 70
Christine Grady, Bernard Lo

13 Palliative and End-of-Life Care ... 74
Ezekiel J. Emanuel, John W. Urwin

PART 2 Cardinal Manifestations and Presentation of Diseases

SECTION 1 Pain

14 Pain: Pathophysiology and Management ... 93
James P. Rathmell, Howard L. Fields

15 Chest Discomfort ... 102
David A. Morrow

16 Abdominal Pain ... 109
Danny O. Jacobs

17 Headache ... 113
Peter J. Goadsby

18 Low Back Pain ... 118
Steven P. Cohen, Eric J. Wang

19 Neck Pain ... 129
Fan Jiang, Soichiro Takamiya, Michael G. Fehlings

SECTION 2 Alterations in Body Temperature

20 Fever ... 133
Charles A. Dinarello, Neeraj K. Surana

21 Fever and Rash ... 137
Elaine T. Kaye, Kenneth M. Kaye

22 Fever of Unknown Origin ... 149
Catharina M. Mulders-Manders, Chantal P. Rovers

SECTION 3 Nervous System Dysfunction

23 Syncope ... 155
Roy Freeman, Satish R. Raj

24 Dizziness and Vertigo ... 162
Mark F. Walker, Robert B. Daroff

25 Fatigue ... 165
Jeffrey M. Gelfand, Vanja C. Douglas

26 Neurologic Causes of Weakness and Paralysis ... 168
Stephen L. Hauser

27 Numbness, Tingling, and Sensory Loss ... 172
Stephen L. Hauser

28 Gait Disorders, Imbalance, and Falls ... 177
Jessica M. Baker, Gerald Pankratz

29 Confusion and Delirium ... 181
S. Andrew Josephson, Bruce L. Miller

30 Coma ... 186
S. Andrew Josephson, Allan H. Ropper, Stephen L. Hauser

31 Dementia ... 193
Gil D. Rabinovici, William W. Seeley, Bruce L. Miller

32 Aphasia, Memory Loss, and Other Cognitive Disorders ... 200
M.-Marsel Mesulam

33 Sleep Disorders ... 209
Thomas E. Scammell, Clifford B. Saper, Charles A. Czeisler

SECTION 4 Disorders of Eyes, Ears, Nose, and Throat

34 Disorders of the Eye ... 220
Jonathan C. Horton

35 Disorders of Smell and Taste ... 237
Richard L. Doty, Steven M. Bromley

36 Disorders of Hearing ... 244
Anil K. Lalwani

37 Upper Respiratory Symptoms, Including Earache, Sinus Symptoms, and Sore Throat ... 252
Rachel L. Amdur, Jeffrey A. Linder

38 Oral Manifestations of Disease ... 260
Samuel C. Durso

SECTION 5 Alterations in Circulatory and Respiratory Functions

39 Dyspnea ... 267
Rebecca M. Baron

40 Cough ... 272
Bruce D. Levy

41 Hemoptysis ... 274
Erin M. DeBiasi, Carolyn D'Ambrosio

42 Hypoxia and Cyanosis ... 276
Joseph Loscalzo

43 Edema ... 279
Joseph Loscalzo

44 Approach to the Patient with a Heart Murmur ... 283
Patrick T. O'Gara, Joseph Loscalzo

45 Palpitations....................291
Joseph Loscalzo

46 Exercise Intolerance....................291
Joseph Loscalzo, William M. Oldham

SECTION 6 Alterations in Gastrointestinal Function

47 Dysphagia....................293
Ikuo Hirano, Peter J. Kahrilas

48 Nausea, Vomiting, and Indigestion....................296
William L. Hasler

49 Diarrhea and Constipation....................302
Michael Camilleri, Joseph A. Murray

50 Unintentional Weight Loss....................314
J. Larry Jameson

51 Gastrointestinal Bleeding....................316
Loren Laine

52 Jaundice....................320
Savio John, Daniel S. Pratt

53 Abdominal Swelling and Ascites....................326
Lawrence S. Friedman

SECTION 7 Alterations in Renal and Urinary Tract Function

54 Interstitial Cystitis/Bladder Pain Syndrome....................330
R. Christopher Doiron, J. Curtis Nickel

55 Azotemia and Urinary Abnormalities....................337
David B. Mount

56 Fluid and Electrolyte Disturbances....................344
David B. Mount

57 Hypercalcemia and Hypocalcemia....................362
Dwight A. Towler

58 Acidosis and Alkalosis....................366
Thomas D. DuBose, Jr.

SECTION 8 Alterations in the Skin

59 Approach to the Patient with a Skin Disorder....................377
Kim B. Yancey, Thomas J. Lawley

60 Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders....................382
Justin T. Cheeley, Leslie P. Lawley

61 Skin Manifestations of Internal Disease....................392
Jean L. Bolognia, Jonathan S. Leventhal, Irwin M. Braverman

62 Immunologically Mediated Skin Diseases....................409
Kim B. Yancey, Benjamin F. Chong, Thomas J. Lawley

63 Cutaneous Drug Reactions....................416
Robert G. Micheletti, Misha Rosenbach, Elizabeth J. Phillips, Bruce U. Wintroub, Kanade Shinkai

64 Photosensitivity and Other Reactions to Sunlight....................429
Alexander G. Marneros, David R. Bickers

SECTION 9 Hematologic Alterations

65 Interpreting Peripheral Blood Smears....................437
Dan L. Longo

66 Anemia and Polycythemia....................443
Thomas G. DeLoughery, Dan L. Longo

67 Disorders of Granulocytes and Monocytes....................453
Steven M. Holland, John I. Gallin

68 Primary and Secondary Hemophagocytic Lymphohistiocytosis....................464
Jan-Inge Henter

69 Bleeding and Thrombosis....................469
Barbara A. Konkle

70 Enlargement of Lymph Nodes and Spleen....................476
Dan L. Longo

PART 3 Pharmacology

71 Principles of Clinical Pharmacology....................483
Dan Roden

72 Pharmacogenomics....................491
Dan Roden

PART 4 Oncology and Hematology

SECTION 1 Neoplastic Disorders

73 Approach to the Patient with Cancer....................497
Dan L. Longo

74 Symptom Control in Patients with Cancer....................506
Charles L. Loprinzi, Thomas J. Smith

75 Prevention and Early Detection of Cancer....................511
Jennifer M. Croswell, Otis W. Brawley, Barnett S. Kramer

76 Cancer Genetics....................519
Fred Bunz, Bert Vogelstein

77 Cancer Cell Biology....................528
Jeffrey W. Clark, Dan L. Longo

78 Principles of Cancer Treatment....................551
Edward A. Sausville, Dan L. Longo

79 Infections in Patients with Cancer....................576
Brahm H. Segal, Juan C. Gea-Banacloche

80 Oncologic Emergencies....................585
Rasim Gucalp, Janice P. Dutcher

81 Cancer of the Skin....................598
Brendan D. Curti, John T. Vetto, Sancy A. Leachman

82 Head and Neck Cancer....................610
Everett E. Vokes

83 Neoplasms of the Lung....................614
Eric K. Singhi, Christine M. Lovly

84 Breast Cancer....................632
Nancy E. Davidson

85 Upper Gastrointestinal Tract Cancers....................641
David Kelsen

86 Colorectal Cancer....................653
David P. Ryan

87 Tumors of the Liver and Biliary Tree....................659
Josep M. Llovet

88 Pancreatic Cancer....................671
Eileen M. O'Reilly, Andrew M. Lowy, Daniel D. Von Hoff

89 Gastrointestinal Neuroendocrine Tumors....................676
Matthew H. Kulke

90 Renal Cell Carcinoma....................687
Robert J. Motzer, Martin H. Voss

91 Cancer of the Bladder and Urinary Tract....................690
Noah M. Hahn

92 Benign and Malignant Diseases of the Prostate....................695
Oliver Sartor, James A. Eastham

93 Testicular Cancer....................704
David J. Vaughn

94 Gynecologic Malignancies....................709
David Spriggs

95 Primary and Metastatic Tumors of the Nervous System 716
Mary Jane Lim-Fat, Patrick Y. Wen

96 Soft Tissue and Bone Sarcomas and Bone Metastases 727
Shreyaskumar R. Patel

97 Carcinoma of Unknown Primary 731
Kanwal Raghav, James L. Abbruzzese

98 Paraneoplastic Syndromes: Endocrinologic/Hematologic 736
J. Larry Jameson, Dan L. Longo

99 Paraneoplastic Neurologic Syndromes and Autoimmune Encephalitis 743
Josep Dalmau, Francesc Graus

100 Cancer Survivorship and the Long-Term Impact of Cancer and Its Treatment 752
Mark Roschewski, Dan L. Longo

SECTION 2 Hematopoietic Disorders

101 Hematopoietic Stem Cells 758
David T. Scadden, Dan L. Longo

102 Iron Deficiency and Other Acquired Anemias Due to Decreased Erythrocyte Production 762
Clara Camaschella, Domenico Girelli

103 Disorders of Hemoglobin 769
Vijay G. Sankaran, Martin H. Steinberg

104 Megaloblastic Anemias 780
A. Victor Hoffbrand

105 Hemolytic Anemias 791
Lucio Luzzatto, Lucia De Franceschi

106 Anemia Due to Acute Blood Loss 807
Dan L. Longo

107 Bone Marrow Failure Syndromes Including Aplastic Anemia and Myelodysplasia 808
Neal S. Young, Emma M. Groarke

108 Polycythemia Vera and Other Myeloproliferative Neoplasms 818
Jerry L. Spivak

109 Acute Myeloid Leukemia 825
William Blum

110 Chronic Myeloid Leukemia 834
Hagop Kantarjian, Elias Jabbour

111 Acute Lymphoid Leukemia 844
Dieter Hoelzer

112 Chronic Lymphocytic Leukemia 849
Jennifer A. Woyach, John C. Byrd

113 Non-Hodgkin's Lymphoma 857
Caron A. Jacobson, Dan L. Longo

114 Hodgkin's Lymphoma 868
Caron A. Jacobson, Dan L. Longo

115 Less Common Lymphoid and Myeloid Malignancies 871
Ayalew Tefferi, Dan L. Longo

116 Plasma Cell Disorders 883
Nikhil C. Munshi, Dan L. Longo, Kenneth C. Anderson

117 Amyloidosis 895
John L. Berk, Vaishali Sanchorawala

118 Transfusion Therapy and Biology 901
Pierre Tiberghien, Olivier Garraud, Jacques Chiaroni

119 Hematopoietic Cell Transplantation 914
Frederick R. Appelbaum

SECTION 3 Disorders of Hemostasis

120 Disorders of Platelets and Vessel Wall 921
Barbara A. Konkle

121 Coagulation Disorders 928
Jean M. Connors

122 Arterial and Venous Thrombosis 937
Jane E. Freedman, Joseph Loscalzo

123 Antiplatelet, Anticoagulant, and Fibrinolytic Drugs 941
Jeffrey I. Weitz

PART 5 Infectious Diseases

SECTION 1 Basic Considerations in Infectious Diseases

124 Approach to the Patient with an Infectious Disease 957
Neeraj K. Surana, Dennis L. Kasper

125 Molecular Mechanisms of Microbial Pathogenesis 964
Jordan B. Jastrab, Marcia B. Goldberg

126 Microbial Genomics and Infectious Disease 975
Roby P. Bhattacharyya, Yonatan H. Grad, Deborah T. Hung

127 Approach to the Acutely Ill Infected Febrile Patient 989
Tamar F. Barlam

128 Principles of Immunization 996
Kathleen M. Neuzil, Kathryn M. Edwards

129 Immunization Principles and Vaccine Use 1001
Sarah Meyer, Amanda Cohn, Georgina Peacock

130 Health Recommendations for International Travel 1010
Jesse Waggoner, Henry M. Wu

SECTION 2 Clinical Syndromes: Community-Acquired Infections

131 Pneumonia 1023
Lionel A. Mandell, Michael S. Niederman

132 Lung Abscess 1034
Rebecca M. Baron, Beverly W. Baron, Miriam Baron Barshak

133 Infective Endocarditis 1036
Sara E. Cosgrove, Michael T. Melia

134 Infections of the Skin, Muscles, and Soft Tissues 1049
Dennis L. Stevens, Amy E. Bryant

135 Infectious Arthritis 1055
Nongnooch Poowanawittayakom, Lawrence C. Madoff

136 Osteomyelitis 1061
Werner Zimmerli

137 Intraabdominal Infections and Abscesses 1069
Miriam Baron Barshak

138 Acute Infectious Diarrheal Diseases and Bacterial Food Poisoning 1076
Ana A. Weil, Regina C. LaRocque

139 *Clostridioides difficile* Infection, Including Pseudomembranous Colitis 1082
Dale N. Gerding, Stuart Johnson

140 Urinary Tract Infections: Cystitis, Prostatitis, and Pyelonephritis 1086
Barbara W. Trautner, Prathit A. Kulkarni, Nicolás W. Cortés-Penfield, Kalpana Gupta

141 Sexually Transmitted Infections: Overview and Clinical Approach 1096
Jeanne M. Marrazzo

142 Encephalitis ... 1111
Karen L. Roos, Michael R. Wilson, Kenneth L. Tyler

143 Acute Meningitis ... 1118
Karen L. Roos, Kenneth L. Tyler

144 Chronic and Recurrent Meningitis ... 1128
Avindra Nath, Walter J. Koroshetz, Michael R. Wilson

145 Brain Abscess and Empyema ... 1135
Karen L. Roos, Kenneth L. Tyler

146 Infectious Complications of Bites ... 1142
Nongnooch Poowanawittayakom, Lawrence C. Madoff

SECTION 3 Clinical Syndromes: Health Care–Associated Infections

147 Infections Acquired in Health Care Facilities ... 1146
Mini Kamboj, Tara N. Palmore

148 Infections in Transplant Recipients ... 1154
Jennifer M. Cuellar-Rodriguez, Juan C. Gea-Banacloche

SECTION 4 Therapy for Bacterial Diseases

149 Treatment and Prophylaxis of Bacterial Infections ... 1165
David C. Hooper, Erica S. Shenoy, Alyssa R. Letourneau, Ramy H. Elshaboury

150 Bacterial Resistance to Antimicrobial Agents ... 1181
David C. Hooper

SECTION 5 Diseases Caused by Gram-Positive Bacteria

151 Pneumococcal Infections ... 1187
David Goldblatt, Katherine L. O'Brien

152 Staphylococcal Infections ... 1195
Franklin D. Lowy, Anne-Catrin Uhlemann

153 Streptococcal Infections ... 1206
Michael R. Wessels

154 Enterococcal Infections ... 1215
William R. Miller, Cesar A. Arias, Barbara E. Murray

155 Diphtheria and Other Corynebacterial Infections ... 1221
William R. Bishai, John R. Murphy

156 *Listeria monocytogenes* Infections ... 1227
Karen P. Neil, Jennifer P. Collins, Patricia M. Griffin

157 Tetanus ... 1230
C. Louise Thwaites, Hai Duong Ha Thi

158 Botulism ... 1233
Carolina Lúquez, Jeremy Sobel

159 Gas Gangrene and Other Clostridial Infections ... 1238
Amy E. Bryant, Dennis L. Stevens

SECTION 6 Diseases Caused by Gram-Negative Bacteria

160 Meningococcal Infections ... 1244
Manish Sadarangani, Andrew J. Pollard

161 Gonococcal Infections ... 1253
Sanjay Ram, Peter A. Rice

162 *Haemophilus* and *Moraxella* Infections ... 1260
Kristian Riesbeck, Timothy F. Murphy

163 Infections Due to the HACEK Group and Miscellaneous Gram-Negative Bacteria ... 1265
Tamar F. Barlam

164 *Legionella* Infections ... 1269
Steven A. Pergam, Thomas R. Hawn

165 Pertussis and Other *Bordetella* Infections ... 1276
May S. ElSherif, Scott A. Halperin

166 Diseases Caused by Gram-Negative Enteric Bacilli ... 1281
Thomas A. Russo, Yohei Doi

167 *Acinetobacter* Infections ... 1294
Rossana Rosa, Christine A. Vu

168 *Helicobacter pylori* Infections ... 1298
Emad M. El-Omar, Martin J. Blaser

169 Infections Due to *Campylobacter* and Related Organisms ... 1303
Martin J. Blaser

170 Infections Due to *Pseudomonas*, *Burkholderia*, and *Stenotrophomonas* Species ... 1307
Kartikeya Cherabuddi, Reuben Ramphal

171 Salmonellosis ... 1314
David A. Pegues, Samuel I. Miller

172 Shigellosis ... 1322
Philippe J. Sansonetti, Jean Bergounioux

173 Cholera and Other Vibrioses ... 1325
Matthew K. Waldor, Edward T. Ryan

174 Brucellosis ... 1331
Nicholas J. Beeching

175 Tularemia ... 1335
Anders F. Johansson

176 Plague and Other *Yersinia* Infections ... 1341
Michael B. Prentice

177 *Bartonella* Infections, Including Cat-Scratch Disease ... 1350
Michael Giladi, Moshe Ephros

178 Donovanosis ... 1356
Nigel O'Farrell

SECTION 7 Miscellaneous Bacterial Infections

179 Nocardiosis ... 1358
Michael S. Abers, Gregory A. Filice

180 Actinomycosis ... 1363
Thomas A. Russo, John C. Hu

181 Whipple Disease ... 1367
Thomas A. Russo, Seth R. Glassman

182 Infections Due to Mixed Anaerobic Organisms ... 1371
Neeraj K. Surana, Dennis L. Kasper

SECTION 8 Mycobacterial Diseases

183 Tuberculosis ... 1380
Mario C. B. Raviglione, Andrea Gori

184 Leprosy ... 1404
Jan H. Richardus, Hemanta K. Kar, Zoica Bakirtzief, Wim H. van Brakel

185 Nontuberculous Mycobacterial Infections ... 1414
Steven M. Holland

186 Antimycobacterial Agents ... 1418
Divya Reddy, Sebastian G. Kurz, Max R. O'Donnell

SECTION 9 Spirochetal Diseases

187 Syphilis ... 1428
Sheila A. Lukehart

188 Endemic Treponematoses ... 1436
Sheila A. Lukehart, Lorenzo Giacani

189 Leptospirosis 1439
Marco Goeijenbier, Jiři F. P. Wagenaar

190 Relapsing Fever..... 1444
Alan G. Barbour

191 Lyme Borreliosis 1448
Allen C. Steere, Jacob E. Lemieux

SECTION 10 Diseases Caused by Rickettsiae, Mycoplasmas, and Chlamydiae

192 Rickettsial Diseases..... 1454
David H. Walker, J. Stephen Dumler, Lucas S. Blanton, Chantal P. Rovers

193 Infections Due to Mycoplasmas 1464
R. Doug Hardy

194 Chlamydial Infections 1467
Matthew M. Hamill, Thomas C. Quinn

SECTION 11 Viral Diseases: General Considerations

195 Principles of Medical Virology 1477
David M. Knipe, Max L. Nibert

196 Antiviral Chemotherapy, Excluding Antiretroviral Drugs..... 1484
Jeffrey I. Cohen, Marc G. Ghany

SECTION 12 Infections Due to DNA Viruses

197 Herpes Simplex Virus Infections 1495
Lawrence Corey

198 Varicella-Zoster Virus Infections..... 1504
Richard J. Whitley

199 Epstein-Barr Virus Infections, Including Infectious Mononucleosis 1508
Jeffrey I. Cohen

200 Cytomegalovirus and Human Herpesvirus Types 6, 7, and 8..... 1512
Camille Nelson Kotton, Martin S. Hirsch

201 Monkeypox, Molluscum Contagiosum, and Other Poxvirus Infections..... 1517
Faisal Syed Minhaj, Christina L. Hutson

202 Parvovirus Infections 1522
Maria Söderlund-Venermo

203 Human Papillomavirus Infections 1525
Darron R. Brown, Aaron C. Ermel

SECTION 13 Infections Due to DNA and RNA Respiratory Viruses

204 Common Viral Respiratory Infections, Other Than COVID-19..... 1531
James E. Crowe, Jr.

205 SARS-CoV-2 and COVID-19..... 1539
James E. Crowe, Jr.

206 Influenza 1544
Kathleen M. Neuzil

SECTION 14 Infections Due to Human Immunodeficiency Virus and Other Human Retroviruses

207 The Human Retroviruses..... 1550
Dan L. Longo, Anthony S. Fauci

208 Human Immunodeficiency Virus Disease: AIDS and Related Disorders..... 1556
Anthony S. Fauci, Gregory K. Folkers, H. Clifford Lane

SECTION 15 Infections Due to RNA Viruses

209 Viral Gastroenteritis 1627
Umesh D. Parashar, Roger I. Glass

210 Enterovirus, Parechovirus, and Reovirus Infections..... 1632
Jeffrey I. Cohen

211 Measles (Rubeola)..... 1638
Alex C. Kong, William J. Moss

212 Rubella (German Measles) 1643
Alan C. Ou, Ludmila M. Perelygina, Laura A. Zimmerman, Susan E. Reef

213 Mumps 1646
Jessica Leung, Nina B. Masters

214 Rabies and Other Rhabdovirus Infections 1650
Alan C. Jackson

215 Arthropod-Borne and Rodent-Borne Virus Infections 1655
Jens H. Kuhn, Ian Crozier

216 The Filovirids: Orthoebolavirus and Orthomarburgvirus Infections 1679
Jens H. Kuhn, Ian Crozier

SECTION 16 Fungal Infections

217 Pathogenesis, Diagnosis, and Treatment of Fungal Infections..... 1687
Michail S. Lionakis, John E. Edwards, Jr.

218 Histoplasmosis..... 1692
Chadi A. Hage, L. Joseph Wheat

219 Coccidioidomycosis 1696
Neil M. Ampel

220 Blastomycosis 1699
Gregory M. Gauthier, Bruce S. Klein

221 Cryptococcosis..... 1703
Arturo Casadevall, Shmuel Shoham

222 Candidiasis 1706
Michail S. Lionakis, Shakti Singh, Ashraf S. Ibrahim, John E. Edwards, Jr.

223 Aspergillosis 1713
David W. Denning

224 Mucormycosis..... 1717
Brad Spellberg, Ashraf S. Ibrahim

225 Superficial Fungal Infections..... 1722
Joseph Pechacek, Carol A. Kauffman, Michail S. Lionakis

226 Uncommon Disseminated Fungal Infections 1725
Joseph Pechacek, Carol A. Kauffman, Michail S. Lionakis

227 *Pneumocystis* Infections 1730
Alison Morris, Henry Masur

SECTION 17 Protozoal and Helminthic Infections: General Considerations

228 Introduction to Parasitic Infections 1735
Sharon L. Reed, Charles E. Davis

229 Agents Used to Treat Parasitic Infections..... 1740
Thomas A. Moore

SECTION 18 Protozoal Infections

230 Amebiasis and Infection with Free-Living Amebae 1753
Rosa M. Andrade, Sharon L. Reed

231 Malaria 1760
Elizabeth A. Ashley, Nicholas J. White

232 Babesiosis 1775
Edouard Vannier, Jeffrey A. Gelfand

233 Leishmaniasis 1781
Shyam Sundar

234 Chagas Disease and African Trypanosomiasis 1788
François Chappuis, Yves Jackson

235 *Toxoplasma* Infections 1797
Kami Kim

236 Protozoal Intestinal Infections and Trichomoniasis 1804
Peter F. Weller, Edward T. Ryan

SECTION 19 Helminthic Infections

237 Introduction to Helminthic Infections 1809
Peter F. Weller, Edward T. Ryan

238 Trichinellosis and Other Tissue Nematode Infections 1810
Peter F. Weller, Edward T. Ryan

239 Intestinal Nematode Infections 1814
David J. Diemert, Thomas B. Nutman

240 Filarial and Related Infections 1819
Philip J. Budge, Thomas B. Nutman

241 Schistosomiasis and Other Trematode Infections 1825
Birgitte Jyding Vennervald, Anna-Sofie Stensgaard

242 Cestode Infections 1832
A. Clinton White, Jr., Miguel M. Cabada

PART 6 Disorders of the Cardiovascular System

SECTION 1 Introduction to Cardiovascular Disorders

243 Approach to the Patient with Possible Cardiovascular Disease 1841
Joseph Loscalzo

244 Basic Biology of the Cardiovascular System 1843
Joseph Loscalzo, John F. Keaney, Jr., Calum A. MacRae

245 Epidemiology of Cardiovascular Disease 1854
J. Michael Gaziano, Thomas A. Gaziano

SECTION 2 Diagnosis of Cardiovascular Disorders

246 Physical Examination of the Cardiovascular System 1859
Patrick T. O'Gara, Joseph Loscalzo

247 Electrocardiography 1868
Ary L. Goldberger

248 Noninvasive Cardiac Imaging: Echocardiography, Nuclear Cardiology, and Magnetic Resonance/Computed Tomography Imaging 1876
Marcelo F. DiCarli, Raymond Y. Kwong, Scott D. Solomon

249 Diagnostic Cardiac Catheterization and Coronary Angiography 1903
Jane A. Leopold, Ajar Kochar

SECTION 3 Disorders of Rhythm

250 Principles of Clinical Cardiac Electrophysiology 1910
William H. Sauer, Bruce A. Koplan, Paul C. Zei

251 The Bradyarrhythmias: Disorders of the Sinoatrial Node 1917
William H. Sauer, Bruce A. Koplan

252 The Bradyarrhythmias: Disorders of the Atrioventricular Node 1924
William H. Sauer, Bruce A. Koplan

253 Approach to Supraventricular Arrhythmias 1932
William H. Sauer, Paul C. Zei

254 Physiologic and Nonphysiologic Sinus Rhythm 1935
William H. Sauer, Paul C. Zei

255 Focal Atrial Tachycardia 1937
William H. Sauer, Paul C. Zei

256 Paroxysmal Supraventricular Tachycardias 1938
William H. Sauer, Paul C. Zei

257 Common Atrial Flutter and Macroreentrant and Multifocal Atrial Tachycardias 1943
William H. Sauer, Jorge E. Romero, Paul C. Zei

258 Atrial Fibrillation 1947
William H. Sauer, Jorge E. Romero, Paul C. Zei

259 Approach to Ventricular Arrhythmias 1954
William H. Sauer, Usha B. Tedrow

260 Premature Ventricular Contractions, Nonsustained Ventricular Tachycardia, and Accelerated Idioventricular Rhythm 1960
William H. Sauer, Usha B. Tedrow

261 Sustained Ventricular Tachycardia 1963
William H. Sauer, Usha B. Tedrow

262 Polymorphic Ventricular Tachycardia and Ventricular Fibrillation 1968
William H. Sauer, Usha B. Tedrow

263 Electrical Storm and Incessant Ventricular Tachycardia 1972
William H. Sauer, Usha B. Tedrow

SECTION 4 Disorders of the Heart, Muscles, Valves, and Pericardium

264 Heart Failure: Pathophysiology and Diagnosis 1975
Michael M. Givertz, Mandeep R. Mehra

265 Heart Failure: Management 1987
Akshay Desai, Mandeep R. Mehra

266 Classification of Cardiomyopathy 2001
Lynne Warner Stevenson, Neal K. Lakdawala, Joseph Loscalzo

267 Genetic Cardiomyopathies 2003
Neal K. Lakdawala, Lynne Warner Stevenson, Joseph Loscalzo

268 Acute and Chronic Myocarditis 2011
Neal K. Lakdawala, Lynne Warner Stevenson, Joseph Loscalzo

269 Dilated Cardiomyopathies 2015
Neal K. Lakdawala, Lynne Warner Stevenson, Joseph Loscalzo

270 Amyloidosis and Other Restrictive Cardiomyopathies 2020
Rodney H. Falk, Neal K. Lakdawala, Lynne Warner Stevenson, Joseph Loscalzo

271 Cardiac Transplantation and Prolonged Assisted Circulation 2028
Mandeep R. Mehra

272 Aortic Stenosis 2033
Patrick T. O'Gara, Joseph Loscalzo

273 Aortic Regurgitation 2043
Patrick T. O'Gara, Joseph Loscalzo

274 Mitral Stenosis 2048
Patrick T. O'Gara, Joseph Loscalzo

275 Mitral Regurgitation 2052
Patrick T. O'Gara, Joseph Loscalzo
276 Mitral Valve Prolapse 2057
Patrick T. O'Gara, Joseph Loscalzo
277 Tricuspid Valve Disease 2059
Patrick T. O'Gara, Joseph Loscalzo
278 Pulmonic Valve Disease 2063
Patrick T. O'Gara, Joseph Loscalzo
279 Multiple and Mixed Valvular Heart Disease 2065
Patrick T. O'Gara, Joseph Loscalzo
280 Congenital Heart Disease in the Adult 2067
Anne Marie Valente, Michael J. Landzberg
281 Pericardial Disease 2078
Joseph Loscalzo
282 Atrial Myxoma and Other Cardiac Tumors 2084
Stephen Tsaur, Eric H. Awtry
283 Cardiac Trauma 2087
Alexandra R. Pipilas, Eric H. Awtry

SECTION 5 Coronary and Peripheral Vascular Disease

284 Ischemic Heart Disease 2090
Robert P. Giugliano, Elliott M. Antman, Joseph Loscalzo
285 Non-ST-Segment Elevation Acute Coronary Syndrome (Non-ST-Segment Elevation Myocardial Infarction and Unstable Angina) 2106
Robert P. Giugliano, Christopher P. Cannon, Eugene Braunwald
286 ST-Segment Elevation Myocardial Infarction 2113
David A. Morrow, Elliott M. Antman, Joseph Loscalzo
287 Percutaneous Coronary Interventions and Other Interventional Procedures 2126
Deepak L. Bhatt, Pinak B. Shah
288 Hypertension 2133
Paul K. Whelton
289 Renovascular Disease 2150
Stephen C. Textor
290 Deep-Venous Thrombosis and Pulmonary Thromboembolism 2154
Samuel Z. Goldhaber
291 Diseases of the Aorta 2165
Mark A. Creager, Joseph Loscalzo
292 Arterial Diseases of the Extremities 2172
Mark A. Creager, Joseph Loscalzo
293 Chronic Venous Disease and Lymphedema 2179
Mark A. Creager, Robert T. Eberhardt, Joseph Loscalzo
294 Pulmonary Hypertension 2185
Bradley A. Maron, Joseph Loscalzo

PART 7 Disorders of the Respiratory System

SECTION 1 Diagnosis of Respiratory Disorders

295 Approach to the Patient with Disease of the Respiratory System 2199
Bruce D. Levy
296 Disturbances of Respiratory Function 2201
George R. Washko, William M. Oldham
297 Diagnostic Procedures in Respiratory Disease 2207
George R. Washko, Hilary J. Goldberg, Mājid Shafiq

SECTION 2 Diseases of the Respiratory System

298 Asthma 2215
Elliot Israel
299 Hypersensitivity Pneumonitis and Pulmonary Infiltrates with Eosinophilia 2228
Praveen Akuthota, Michael E. Wechsler
300 Occupational and Environmental Lung Disease 2234
John R. Balmes, Mehrdad Arjomandi
301 Bronchiectasis 2242
Rebecca M. Baron, Beverly W. Baron, Miriam Baron Barshak
302 Cystic Fibrosis 2245
Eric J. Sorscher
303 Chronic Obstructive Pulmonary Disease 2249
Craig P. Hersh, Edwin K. Silverman, Dawn DeMeo
304 Interstitial Lung Disease 2259
Gary M. Hunninghake, Ivan O. Rosas
305 Disorders of the Pleura 2267
Khalid Ismail, Hilary J. Goldberg, Rebecca M. Baron
306 Disorders of the Mediastinum 2271
Stefi F. Lee, Rebecca M. Baron, Hilary J. Goldberg
307 Disorders of Ventilation 2273
Khalid Ismail, George R. Washko
308 Sleep Apnea 2277
Andrew Wellman, Daniel J. Gottlieb, Susan Redline
309 Lung Transplantation 2283
Hilary J. Goldberg, Hari R. Mallidi
310 Interventional Pulmonary Medicine 2287
Lonny Yarmus, David Feller-Kopman

PART 8 Critical Care Medicine

SECTION 1 Respiratory Critical Care

311 Approach to the Patient with Critical Illness 2291
Rebecca M. Baron, Anthony F. Massaro
312 Acute Respiratory Distress Syndrome 2299
Rebecca M. Baron, Bruce D. Levy
313 Mechanical Ventilatory Support 2304
Scott Schissel

SECTION 2 Shock and Cardiac Arrest

314 Approach to the Patient with Shock 2309
Rebecca M. Baron, Anthony F. Massaro
315 Sepsis and Septic Shock 2316
Jeffrey R. Strich, Daniel S. Chertow
316 Cardiogenic Shock and Pulmonary Edema 2324
David H. Ingbar, Holger Thiele
317 Cardiovascular Collapse, Cardiac Arrest, and Sudden Cardiac Death 2332
Christine M. Albert, William H. Sauer

SECTION 3 Neurologic Critical Care

318 Nervous System Disorders in Critical Care 2342
J. Claude Hemphill, III, Wade S. Smith, S. Andrew Josephson, Daryl R. Gress

PART 9 Disorders of the Kidney and Urinary Tract

319 Approach to the Patient with Renal Disease or Urinary Tract Disease 2353
Julian L. Seifter
320 Cell Biology and Physiology of the Kidney 2363
Alfred L. George, Jr., Eric G. Neilson
321 Acute Kidney Injury 2372
Sushrut S. Waikar, Joseph V. Bonventre
322 Chronic Kidney Disease 2385
Joanne M. Bargman, Karl L. Skorecki
323 Dialysis in the Treatment of Kidney Failure 2398
Kathleen D. Liu, Glenn M. Chertow
324 Interventional Nephrology 2402
Dirk M. Hentschel
325 Kidney Transplantation 2405
Jamil R. Azzi, Naoka Murakami, Anil K. Chandraker
326 Glomerular Diseases 2411
Julia B. Lewis, Eric G. Neilson
327 Polycystic Kidney Disease and Other Inherited Disorders of Tubule Growth and Development 2430
Jing Zhou, Martin R. Pollak
328 Tubulointerstitial Diseases of the Kidney 2437
Laurence H. Beck, Jr., David J. Salant
329 Thrombotic Renovascular Disorders 2444
Elisabeth M. Battinelli, Rebecca L. Zon
330 Nephrolithiasis 2448
Gary C. Curhan
331 Urinary Tract Obstruction 2453
Julian L. Seifter

PART 10 Disorders of the Gastrointestinal System

SECTION 1 Disorders of the Alimentary Tract

332 Approach to the Patient with Gastrointestinal Disease 2459
William L. Hasler, Chung Owyang
333 Gastrointestinal Endoscopy 2464
Louis Michel Wong Kee Song, Vinay Chandrasekhara, Mark Topazian
334 Diseases of the Esophagus 2503
Peter J. Kahrilas, Ikuo Hirano
335 Peptic Ulcer Disease and Related Disorders 2514
John Del Valle
336 Disorders of Absorption 2540
Deborah C. Rubin
337 Inflammatory Bowel Disease 2551
Sonia Friedman, Richard S. Blumberg
338 Irritable Bowel Syndrome 2571
Chung Owyang, Prashant Singh
339 Diverticular Disease and Common Anorectal Disorders 2578
Susan L. Gearhart
340 Mesenteric Vascular Insufficiency 2586
Daniel Willie-Permor, Mahmoud Malas
341 Acute Intestinal Obstruction 2589
Danny O. Jacobs
342 Acute Appendicitis and Peritonitis 2594
Danny O. Jacobs

SECTION 2 Nutrition

343 Nutrient Requirements and Dietary Assessment 2598
Johanna T. Dwyer
344 Vitamin and Trace Mineral Deficiency and Excess 2605
Paolo M. Suter
345 Malnutrition and Nutritional Assessment 2615
Charles Chin Han Lew, Charlene W. Compher
346 Enteral and Parenteral Nutrition 2619
L. John Hoffer, Bruce R. Bistrian

SECTION 3 Liver and Biliary Tract Disease

347 Approach to the Patient with Liver Disease 2627
Marc G. Ghany, Jay H. Hoofnagle
348 Evaluation of Liver Function 2633
Emily D. Bethea, Daniel S. Pratt
349 The Hyperbilirubinemias 2637
Allan W. Wolkoff
350 Acute Viral Hepatitis 2643
Esperance A. K. Schaefer, Raymond T. Chung, Jules L. Dienstag
351 Toxic and Drug-Induced Hepatitis 2665
William M. Lee, Jules L. Dienstag
352 Chronic Hepatitis 2673
Esperance A. K. Schaefer, Raymond T. Chung, Jules L. Dienstag
353 Alcohol-Associated Liver Disease 2697
Bernd Schnabl
354 Metabolic Dysfunction–Associated Steatotic Liver Disease and Steatohepatitis 2700
Manal F. Abdelmalek, Anna Mae Diehl
355 Cirrhosis and Its Complications 2708
Alex S. Befeler, Bruce R. Bacon
356 Liver Transplantation 2718
Emily D. Bethea, Raymond T. Chung, Jules L. Dienstag
357 Diseases of the Gallbladder and Bile Ducts 2727
Norton J. Greenberger, Gustav Paumgartner, Daniel S. Pratt

SECTION 4 Disorders of the Pancreas

358 Approach to the Patient with Pancreatic Disease 2738
Somashekar G. Krishna, Darwin L. Conwell, Phil A. Hart
359 Acute and Chronic Pancreatitis 2742
Phil A. Hart, Darwin L. Conwell, Somashekar G. Krishna

PART 11 Immune-Mediated, Inflammatory, and Rheumatologic Disorders

SECTION 1 The Immune System in Health and Disease

360 Introduction to the Immune System 2755
Barton F. Haynes, Kelly A. Cuttle, Anthony S. Fauci
361 Mechanisms of Regulation and Dysregulation of the Immune System 2784
Barton F. Haynes, Kelly A. Cuttle, Anthony S. Fauci
362 Primary Immune Deficiency Diseases 2793
Alain Fischer

SECTION 2 Disorders of Immune-Mediated Injury

363 Urticaria, Angioedema, and Allergic Rhinitis................ 2804
Dinah Foer, Joshua A. Boyce

364 Anaphylaxis 2812
David Hong, Joshua A. Boyce

365 Desensitization.. 2814
Mariana Castells

366 Mastocytosis.. 2819
Matthew P. Giannetti, Joshua A. Boyce

367 Autoimmunity and Autoimmune Diseases 2821
Mariana J. Kaplan

368 Systemic Lupus Erythematosus.................................. 2827
Chelsey J.F. Smith, Paul J. Hoover, Kenneth Kalunian

369 Antiphospholipid Syndrome 2837
Haralampos M. Moutsopoulos, Maria G. Tektonidou

370 Rheumatoid Arthritis.. 2839
Ankoor Shah, E. William St. Clair

371 Acute Rheumatic Fever 2854
Joseph Kado, Jonathan Carapetis

372 Systemic Sclerosis (Scleroderma) and Related Disorders 2859
John Varga

373 Sjögren's Disease 2876
Haralampos M. Moutsopoulos, Clio P. Mavragani

374 Spondyloarthritis 2879
Atul Deodhar, Dirk Elewaut

375 The Vasculitis Syndromes 2893
Carol A. Langford, Anthony S. Fauci

376 Behçet Syndrome.. 2908
Yusuf Yazici

377 Inflammatory Myopathies.. 2909
Steven A. Greenberg, Anthony A. Amato

378 Relapsing Polychondritis... 2917
Marcela A. Ferrada, Peter C. Grayson

379 Sarcoidosis.. 2921
Alicia K. Gerke

380 IgG4-Related Disease... 2930
John H. Stone

381 Familial Mediterranean Fever and Other Hereditary Autoinflammatory Diseases 2934
Daniel L. Kastner

SECTION 3 Disorders of the Joints and Adjacent Tissues

382 Approach to Articular and Musculoskeletal Disorders 2938
John J. Cush

383 Osteoarthritis 2949
David T. Felson, Tuhina Neogi

384 Gout and Other Crystal-Associated Arthropathies....... 2956
Hyon K. Choi

385 Fibromyalgia.. 2963
Leslie J. Crofford

386 Arthritis Associated with Systemic Disease, and Other Arthritides.. 2966
Carol A. Langford, Brian F. Mandell

387 Periarticular Disorders of the Extremities..................... 2973
Carol A. Langford

PART 12 Endocrinology and Metabolism

SECTION 1 Endocrinology

388 Approach to the Patient with Endocrine Disorders 2977
J. Larry Jameson

389 Mechanisms of Hormone Action 2980
J. Larry Jameson

390 Physiology of Anterior Pituitary Hormones 2987
Shlomo Melmed, J. Larry Jameson

391 Hypopituitarism 2992
Shlomo Melmed, J. Larry Jameson

392 Pituitary Tumor Syndromes 2998
Shlomo Melmed, J. Larry Jameson

393 Disorders of the Neurohypophysis............................... 3014
Mirjam Christ-Crain, Mark Sherlock

394 Thyroid Gland Physiology and Testing......................... 3020
J. Larry Jameson, Anthony P. Weetman, Susan J. Mandel

395 Hypothyroidism... 3027
Anthony P. Weetman, Susan J. Mandel, J. Larry Jameson

396 Hyperthyroidism and Other Causes of Thyrotoxicosis.................................... 3032
Anthony P. Weetman, Susan J. Mandel, J. Larry Jameson

397 Thyroid Nodular Disease and Thyroid Cancer............. 3040
Susan J. Mandel, Anthony P. Weetman, J. Larry Jameson

398 Disorders of the Adrenal Cortex.................................. 3050
Wiebke Arlt, Alessandro Prete

399 Pheochromocytoma 3073
Frederic Castinetti, Hartmut P. H. Neumann

400 Multiple Endocrine Neoplasia Syndromes................... 3079
Rajesh V. Thakker

401 Autoimmune Polyendocrine Syndromes....................... 3089
Peter A. Gottlieb, Aaron W. Michels

SECTION 2 Sex- and Gender-Based Medicine

402 Sex Development... 3094
Courtney Finlayson, J. Larry Jameson, John C. Achermann

403 Disorders of the Testes and Male Reproductive System.. 3102
Shalender Bhasin, J. Larry Jameson

404 Disorders of the Female Reproductive System 3124
Janet E. Hall, Anuja Dokras

405 Menstrual Disorders and Pelvic Pain 3132
Janet E. Hall, Anuja Dokras

406 Hirsutism 3137
David A. Ehrmann

407 Menopause and Postmenopausal Hormone Therapy............................. 3142
JoAnn E. Manson, Shari S. Bassuk

408 Infertility and Contraception 3149
Anuja Dokras, Janet E. Hall

409 Sexual Dysfunction 3155
Kevin T. McVary

410 Women's Health.. 3162
Emily Nosova, Andrea Dunaif

411 Men's Health 3168
Shalender Bhasin

412 Lesbian, Gay, Bisexual, Transgender, Queer or Questioning, Intersex, Asexual, and More (LGBTQIA+) Health 3178
Baligh R. Yehia, Todd P. Levin

SECTION 3 Obesity, Diabetes Mellitus, and Metabolic Syndrome

413 Pathobiology of Obesity 3181
Stephen O'Rahilly, I. Sadaf Farooqi
414 Evaluation and Management of Obesity 3187
Robert F. Kushner
415 Diabetes Mellitus: Diagnosis, Classification, and Pathophysiology 3195
Alvin C. Powers, Kevin D. Niswender, Carmella Evans-Molina
416 Diabetes Mellitus: Management and Therapies 3205
Alvin C. Powers, Kevin D. Niswender, Michael R. Rickels
417 Diabetes Mellitus: Complications 3221
Alvin C. Powers, John M. Stafford, Michael R. Rickels
418 Hypoglycemia 3230
Stephen N. Davis
419 Disorders of Lipoprotein Metabolism 3237
Daniel J. Rader
420 The Metabolic Syndrome 3252
Robert H. Eckel

SECTION 4 Disorders of Bone and Mineral Metabolism

421 Bone and Mineral Metabolism in Health and Disease 3259
F. Richard Bringhurst, Henry M. Kronenberg, Eva S. Liu, Marc N. Wein
422 Disorders of the Parathyroid Gland and Calcium Homeostasis 3271
Michael Mannstadt, Harald Jüppner
423 Osteoporosis 3291
Peter R. Ebeling
424 Paget's Disease and Other Dysplasias of Bone 3311
Rajesh K. Jain, Tamara J. Vokes

SECTION 5 Disorders of Intermediary Metabolism

425 Heritable Disorders of Connective Tissue 3319
Joan C. Marini, Fransiska Malfait
426 Hemochromatosis 3333
Darrell H. G. Crawford, David M. Frazer
427 Wilson's Disease 3338
Stephen G. Kaler
428 The Porphyrias 3340
Robert J. Desnick, Manisha Balwani
429 Lysosomal Storage Diseases 3351
Robert J. Hopkin, Gregory A. Grabowski
430 Glycogen Storage Diseases and Other Inherited Disorders of Carbohydrate Metabolism 3358
Rebecca L. Koch, Priya S. Kishnani
431 Inherited Disorders of Amino Acid Metabolism in Adults 3367
Nicola Longo
432 Inherited Defects of Membrane Transport 3373
Nicola Longo

PART 13 Neurologic Disorders

SECTION 1 Diagnosis of Neurologic Disorders

433 Approach to the Patient with Neurologic Disease 3377
Daniel H. Lowenstein, S. Andrew Josephson, Stephen L. Hauser
434 Neuroimaging in Neurologic Disorders 3383
William P. Dillon
435 Pathobiology of Neurologic Diseases 3394
Stephen L. Hauser, Arnold R. Kriegstein, Stanley B. Prusiner

SECTION 2 Diseases of the Central Nervous System

436 Seizures and Epilepsy 3408
Patricia Dugan, Vikram R. Rao
437 Introduction to Cerebrovascular Diseases 3428
Wade S. Smith, J. Claude Hemphill, III
438 Ischemic Stroke 3439
Wade S. Smith, Anthony S. Kim, J. Claude Hemphill, III
439 Intracerebral Hemorrhage 3452
J. Claude Hemphill, III, Edilberto Amorim, Wade S. Smith
440 Subarachnoid Hemorrhage 3457
Wade S. Smith, Nerissa U. Ko, J. Claude Hemphill, III
441 Migraine and Other Primary Headache Disorders 3462
Peter J. Goadsby
442 Alzheimer's Disease 3475
Gil D. Rabinovici, Peter A. Ljubenkov, William W. Seeley, Bruce L. Miller
443 Frontotemporal Dementia 3485
William W. Seeley, Bruce L. Miller
444 Vascular Dementia 3489
Steven M. Greenberg, William W. Seeley
445 Dementia with Lewy Bodies 3491
Irene Litvan, William W. Seeley, Bruce L. Miller
446 Parkinson's Disease 3493
C. Warren Olanow, Anthony H. V. Schapira, Christine Klein
447 Tremor, Chorea, and Other Movement Disorders 3508
C. Warren Olanow, Christine Klein
448 Amyotrophic Lateral Sclerosis and Other Motor Neuron Diseases 3519
Robert H. Brown, Jr.
449 Prion Diseases 3525
Stanley B. Prusiner, Michael D. Geschwind
450 Ataxic Disorders 3533
Roger N. Rosenberg, Vikram G. Shakkottai
451 Disorders of the Autonomic Nervous System 3538
Safwan S. Jaradeh
452 Trigeminal Neuralgia, Bell's Palsy, and Other Cranial Nerve Disorders 3551
Vanja C. Douglas, Stephen L. Hauser
453 Diseases of the Spinal Cord 3559
Stephen L. Hauser
454 Concussion and Other Traumatic Brain Injuries 3570
Geoffrey T. Manley, Benjamin L. Brett, Michael McCrea
455 Multiple Sclerosis 3577
Bruce A. C. Cree, Stephen L. Hauser
456 Neuromyelitis Optica 3590
Bruce A. C. Cree, Stephen L. Hauser

SECTION 3 Nerve and Muscle Disorders

457 Peripheral Neuropathy ... 3594
Anthony A. Amato, Richard J. Barohn

458 Guillain-Barré Syndrome and Other Immune-Mediated Neuropathies ... 3617
Stephen L. Hauser, Anthony A. Amato

459 Myasthenia Gravis and Other Diseases of the Neuromuscular Junction ... 3624
Anthony A. Amato, Amanda C. Guidon

460 Muscular Dystrophies and Other Muscle Diseases ... 3634
Anthony A. Amato, Robert H. Brown, Jr.

SECTION 4 Syndromes Associated with Chronic Fatigue

461 Myalgic Encephalomyelitis/ Chronic Fatigue Syndrome ... 3649
Elizabeth R. Unger, Jin-Mann S. Lin, Jeanne Bertolli

SECTION 5 Psychiatric and Addiction Disorders

462 Biology of Psychiatric Disorders ... 3651
Robert O. Messing, Eric J. Nestler, Matthew W. State

463 Psychiatric Disorders ... 3657
Victor I. Reus

464 Alcohol and Alcohol Use Disorders ... 3675
Marc A. Schuckit

465 Nicotine Addiction ... 3681
David M. Burns

466 Cannabis and Cannabis Use Disorder ... 3686
Nora D. Volkow, Aidan Hampson, Ruben Baler

467 Opioid-Related Disorders ... 3688
Thomas R. Kosten, Colin N. Haile

468 Cocaine, Other Psychostimulants, and Hallucinogens ... 3692
Karran A. Phillips, Wilson M. Compton

PART 14 Poisoning, Drug Overdose, and Envenomation

469 Heavy Metal Poisoning ... 3699
Howard Hu

470 Poisoning and Drug Overdose ... 3703
Mark B. Mycyk

471 Disorders Caused by Venomous Snakebites and Marine Animal Exposures ... 3716
Nathaniel Mann, Michael D. Simpson, Erik Fisher

472 Ectoparasite Infestations and Arthropod Injuries ... 3728
Silas A. Davidson, Scott A. Norton

PART 15 Disorders Associated with Environmental Exposures

473 Health Effects of Climate Change ... 3739
Eugene Richardson, Maxine A. Burkett

474 Climate Change and Infectious Disease ... 3744
Aaron S. Bernstein, Jonathan A. Patz

475 Altitude Illness ... 3754
Buddha Basnyat, Geoffrey Tabin, Steven Roy

476 Hyperbaric and Diving Medicine ... 3760
Simon J. Mitchell, Richard E. Moon

477 Hypothermia and Peripheral Cold Injuries ... 3768
Daniel F. Danzl

478 Heat-Related Illnesses ... 3772
Daniel F. Danzl

PART 16 Genes, the Environment, and Disease

479 Principles of Human Genetics ... 3777
J. Larry Jameson, Peter Kopp

480 The Integration of Inherited Genetics into Clinical Practice ... 3799
Susan M. Domchek, J. Larry Jameson

481 Mitochondrial DNA and Heritable Traits and Diseases ... 3805
Karl L. Skorecki, Bruce H. Cohen

482 Telomere Disease ... 3818
Rodrigo T. Calado, Neal S. Young

483 Gene and Cell-Based Therapy in Clinical Medicine ... 3822
Katherine A. High, Marcela V. Maus

484 The Human Microbiome in Health and Disease ... 3830
Neeraj K. Surana, Dennis L. Kasper

PART 17 Global Medicine

485 Global Issues in Medicine ... 3843
Joseph J. Rhatigan, Paul E. Farmer

486 Emerging and Re-Emerging Infectious Diseases ... 3853
David M. Morens, Anthony S. Fauci

487 Primary Care and Global Health ... 3859
Tim Evans, Kumanan Rasanathan

PART 18 Aging

488 Biology of Aging ... 3869
Rafael de Cabo, David G. Le Couteur

489 Caring for the Geriatric Patient ... 3876
Joseph G. Ouslander, Bernardo Reyes

PART 19 Consultative Medicine

490 Approach to Medical Consultation ... 3897
Jeffrey S. Berns, Jack Ende

491 Medical Disorders During Pregnancy ... 3898
Sarah Rae Easter, Robert L. Barbieri

492 Medical Evaluation of the Patient Undergoing Noncardiac Surgery ... 3905
Prashant Vaishnava, David I. Sahar, Kim A. Eagle

PART 20 Emerging Topics in Clinical Medicine

493 Point-of-Care Ultrasound ... 3911
Wilma Chan, Nilam J. Soni, Paul H. Mayo

494 Improving Health and Health Care Using Behavioral Economics ... 3915
Kevin G. Volpp, George Loewenstein, David A. Asch

495 Complementary and Integrative Therapies and Practices 3924
Helene M. Langevin

496 Placebo and Nocebo Effects 3930
Kathryn T. Hall, Alia J. Crum

497 The Role of Epigenetics in Disease and Treatment 3935
Brian C. Capell, Shelley L. Berger

498 The Role of Circadian Biology in Health and Disease 3943
Jonathan Cedernaes, Kathryn Moynihan Ramsey, Joseph Bass

499 Network Medicine and Multiomic Approaches to Health and Disease 3955
Joseph Loscalzo

500 Emerging Neurotherapeutic Technologies 3963
Jyoti Mishra, Karunesh Ganguly

501 Machine Learning and Augmented Intelligence in Clinical Medicine 3970
Arjun K. Manrai, Isaac S. Kohane

502 Metabolomics 3974
Jared R. Mayers, Mathew G. Vander Heiden

503 Circulating Nucleic Acids as Liquid Biopsies and Noninvasive Disease Biomarkers 3980
Ash A. Alizadeh, Kiran K. Khush, Yair J. Blumenfeld

504 Protein Folding Disorders 3990
Richard I. Morimoto, G. Scott Budinger

505 Novel Approaches to Diseases of Unknown Etiology 3995
David R. Adams, Camilo Toro, Joseph Loscalzo

Index I-1

Related Harrison's 22nd Edition Content

The following chapters are available online. They can be viewed by opening the table of contents of *Harrison's* 22nd edition at *accessmedicine.com/harrisons.*

Video Collection

V1 Video Library of Gait Disorders
Gail A. Kang, Nicholas B. Galifianakis, Michael D. Geschwind

V2 Primary Progressive Aphasia, Memory Loss, and Other Focal Cerebral Disorders
Maria Luisa Gorno-Tempini, Joel Kramer, Bruce L. Miller, Gil D. Rabinovici, Maria Carmela Tartaglia

V3 Video Library of Neuro-Ophthalmology
Jonathan C. Horton

V4 Examination of the Comatose Patient
S. Andrew Josephson

V5 Video Atlas of Gastrointestinal Endoscopic Lesions
Louis Michel Wong Kee Song, Vinay Chandrasekhara, Mark Topazian

V6 The Neurologic Screening Exam
Daniel H. Lowenstein

V7 Video Atlas of the Detailed Neurologic Examination
Martin A. Samuels

V8 Bedside Examination Tests in the Patient with Low Back Pain
Steven P. Cohen, Eric J. Wang

Supplementary Topics

S1 Fluid and Electrolyte Imbalances and Acid-Base Disturbances: Case Examples
David B. Mount, Thomas D. DuBose, Jr.

S2 Cerebrospinal Fluid Disturbances: Case Examples
Prashanth Ramachandran, Michael R. Wilson

S3 Technique of Lumbar Puncture
Elizabeth Robbins, Stephen L. Hauser

S4 Microbial Bioterrorism
H. Clifford Lane, Anthony S. Fauci

S5 Chemical Terrorism
Jonathan Newmark

S6 Radiation Terrorism
Cecilia Jiang, Zelig A. Tochner, Christine E. Hill-Kayser

S7 Infections in War Veterans
Andrew W. Artenstein

S8 Health Care for Military Veterans
Stephen C. Hunt, Lucile Burgo-Black, Charles W. Hoge

S9 Primary Immunodeficiencies Associated with (or Secondary to) Other Diseases
Alain Fischer

S10 The Clinical Laboratory in Modern Health Care
Anthony A. Killeen

S11 Laboratory Diagnosis of Infectious Diseases
Manfred Brigl, Alexander J. McAdam

S12 Laboratory Diagnosis of Parasitic Infections
Sanjay R. Mehta, Sharon L. Reed

S13 Antithrombotic Therapy in Adult Patients
Arvind K. Pandey, Joseph Loscalzo

Atlases

A1 Atlas of Rashes Associated with Fever
Kenneth M. Kaye, Elaine T. Kaye

A2 Atlas of Blood Smears of Malaria and Babesiosis
Nicholas J. White, Elizabeth A. Ashley

A3 Atlas of Oral Manifestations of Disease
Samuel C. Durso, Janet A. Yellowitz

A4 Atlas of Urinary Sediments and Renal Biopsies
Agnes B. Fogo

A5 Atlas of Skin Manifestations of Internal Disease
Justin T. Cheeley, Thomas J. Lawley, Benjamin K. Stoff

A6 Atlas of Hematology
Dan L. Longo

A7 Atlas of Electrocardiography
Ary L. Goldberger

A8 Atlas of Cardiac Arrhythmias
Ary L. Goldberger

A9 Atlas of Noninvasive Imaging
Marcelo F. DiCarli, Raymond Y. Kwong, Scott D. Solomon

A10 Atlas of Atherosclerosis
Peter Libby

A11 Atlas of Percutaneous Revascularization and Adult Structural Heart Interventions
Jane A. Leopold, Deepak L. Bhatt

A12 Atlas of Chest Imaging
Samuel Y. Ash, George R. Washko

A13 Atlas of Liver Biopsies
Jules L. Dienstag, Atul K. Bhan

A14 Atlas of the Vasculitic Syndromes
Carol A. Langford, Anthony S. Fauci

A15 Atlas of Clinical Manifestations of Endocrine and Metabolic Diseases
J. Larry Jameson

A16 Atlas of Neuroimaging
Michael F. Regner, Andre D. Furtado, Luciano Villarinho, William P. Dillon

Contributors

James L. Abbruzzese, MD, FACP, FASCO, DSc (Hon)
Professor, Division of Medical Oncology, Duke Cancer Institute, Durham, North Carolina [97]

Manal F. Abdelmalek, MD, MPH
Professor of Medicine, Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota [354]

Michael S. Abers, MD
Assistant Clinical Investigator, National Institutes of Health (NIH); Chief, Opportunistic Bacterial Pathogenesis Unit, Laboratory of Clinical Immunology and Microbiology, Bethesda, Maryland [179]

John C. Achermann, MD, PhD
Professor of Pediatric Endocrinology, Genetics and Genomic Medicine, UCL GOS Institute of Child Health, University College London, London, United Kingdom [402]

David R. Adams, MD, PhD
Co-Director NIH Undiagnosed Diseases Program, Intramural Research Program, National Institutes of Health, Bethesda, Maryland [505]

Praveen Akuthota, MD
Professor of Medicine, Division of Pulmonary, Critical Care and Sleep Medicine, University of California, San Diego, San Diego, California [299]

Christine M. Albert, MD, MPH
Chair, Department of Cardiology; Lee and Harold Kapelovitz Endowed Chair in Research Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, California [317]

Ash A. Alizadeh, MD, PhD
Professor of Medicine (Oncology), Stanford University School of Medicine, Stanford, California [503]

Anthony A. Amato, MD
Professor of Neurology, Harvard Medical School; Distinguished Chair of Neurology and Chief, Neuromuscular Division, Brigham and Women's Hospital, Boston, Massachusetts [377, 457-460]

Rachel L. Amdur, MD
Assistant Professor of Medicine, Division of General Internal Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois [37]

Edilberto Amorim, MD
Assistant Professor, Department of Neurology, University of California, San Francisco; Director, Epilepsy and Clinical Neurophysiology Service, Zuckerberg San Francisco General Hospital, San Francisco, California [439]

Neil M. Ampel, MD
Professor Emeritus of Medicine and Immunobiology, University of Arizona, Tucson, Arizona [219]

Kenneth C. Anderson, MD
Kraft Family Professor of Medicine, Harvard Medical School; Chief, Jerome Lipper Multiple Myeloma Center, Dana-Farber Cancer Institute, Boston, Massachusetts [116]

Rosa M. Andrade, MD
Assistant Professor of Medicine, University of California, Irvine School of Medicine, Irvine, California [230]

Elliott M. Antman, MD
Professor of Medicine, Senior Physician, Harvard Medical School, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts [284, 286]

Frederick R. Appelbaum, MD
Executive Vice President, Fred Hutchinson Cancer Research Center, Seattle, Washington [119]

Cesar A. Arias, MD, PhD, MSc, FIDSA
Chief, Division of Infectious Diseases, Houston Methodist Hospital; Professor and John F. III and Ann H. Bookout Distinguished Chair; Co-Director Center for Infectious Diseases Research, Houston Methodist Research Institute and Weill Cornell Medical College, Houston, Texas [154]

Mehrdad Arjomandi, MD
Professor of Medicine, Division of Pulmonary, Critical Care, Allergy, and Sleep Medicine, Division of Occupational, Environmental, and Climate Medicine, University of California, San Francisco; Director, University of California San Francisco Human Exposure Laboratory; Director, San Francisco Veterans Affairs Health Care System Environmental Medicine Clinic; Director, San Francisco Veterans Affairs Health Care System Airborne Hazards and Burn Pit Center of Excellence, San Francsico, California [300]

Wiebke Arlt, MD, DSc, FRCP, FMedSci
Director of the Medical Research Council Laboratory of Medical Sciences; Professor of Transdisciplinary Medicine, Imperial College London; Consultant Endocrinologist, Imperial College Healthcare NHS Trust, Hammersmith Hospital, London, United Kingdom [398]

Andrew W. Artenstein, MD, MBA
Chief Physician Executive and Chief Academic Officer, Baystate Health; President, Baystate Medical Practices; Regional Executive Dean and Professor of Medicine, University of Massachusetts Chan Medical School-Baystate, Springfield, Massachusetts [S7]

David A. Asch, MD, MBA
John Morgan Professor, Perelman School of Medicine and the Wharton School; Senior Vice President for Strategic Initiatives, University of Pennsylvania, University of Pennsylvania, Philadelphia, Pennsylvania [494]

Samuel Y. Ash, MD, MPH
Assistant Professor of Medicine, Tufts University School of Medicine; Staff Physician, Critical Care Medicine, South Short Health, South Weymouth, Massachusetts [A12]

Elizabeth A. Ashley, MB, BS, FRCP, FRCPath
Professor of Tropical Medicine, Oxford University; Director, Lao-Oxford-Mahosot Hospital-Wellcome Trust Research Unit, Vientiane, Lao PDR [231, A2]

Eric H. Awtry, MD
Associate Professor of Medicine, Boston University School of Medicine; Associate Chair for Clinical Affairs, Section of Cardiology, Boston Medical Center, Boston, Massachusetts [282, 283]

Jamil R. Azzi, MD, PhD
Associate Professor of Medicine, Harvard Medical School; Medical Director, Vascularized Composite Allotransplantation; Medical Director Kidney and Pancreas Transplantation, Director, Renal Transplant Fellowship; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts [325]

Bruce R. Bacon, MD
Emeritus Professor of Internal Medicine, Saint Louis University School of Medicine, St. Louis, Missouri [355]

Jessica M. Baker, MD
Assistant Professor, Department of Neurology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin [28]

Zoica Bakirtzief, PhD
Substance Abuse Disorder and Mental Health Treatment Technician, NorthStar Regional, Chaska, Minnesota [184]

Ruben Baler, PhD
Health Scientist, Office of Science Policy and Communications, National Institute on Drug Abuse, National Institutes of Health, Bethesda, Maryland [466]

John R. Balmes, MD
Professor Emeritus of Medicine, University of California San Francisco School of Medicine, San Francisco, California [300]

Manisha Balwani, MD, MS
Professor, Department of Genetics and Genomic Sciences and Medicine, Icahn School of Medicine at Mount Sinai, New York, New York [428]

Robert L. Barbieri, MD
Kate Macy Ladd Distinguished Professor of Obstetrics, Gynecology and Reproductive Biology, Harvard Medical School; Chief of Obstetrics, Brigham and Women's Hospital, Boston, Massachusetts [491]

Alan G. Barbour, MD
Distinguished Professor of Medicine and Microbiology and Molecular Genetics, University of California, Irvine, Irvine, California [190]

Joanne M. Bargman, MD, FRCPC
Professor of Medicine, University of Toronto; Staff Nephrologist, University Health Network; Clinician Investigator, Toronto General Hospital Research Institute; Director, Peritoneal Dialysis Program, Co-Director, Renal-Rheumatology Lupus Clinic, University Health Network, Toronto, Ontario, Canada [322]

Tamar F. Barlam, MD, MSc
Professor of Medicine, Boston University Chobanian and Avedisian School of Medicine; Chief, Section of Infectious Diseases, Boston Medical Center, Boston, Massachusetts [127, 163]

Richard J. Barohn, MD
Executive Vice Chancellor for Health Affairs; Executive Director, NextGen Precision Health, University of Missouri, Columbia, Missouri [457]

Beverly W. Baron, MD
Professor of Pathology (Retired), University of Chicago, Chicago, Illinois [132, 301]

Rebecca M. Baron, MD
Associate Professor of Medicine, Harvard Medical School; Associate Physician, Brigham and Women's Hospital, Division of Pulmonary and Critical Care Medicine, Boston, Massachusetts [39, 132, 301, 305, 306, 311, 312, 314]

Miriam Baron Barshak, MD
Associate Professor of Medicine, Harvard Medical School; Physician, Massachusetts General Hospital, Boston, Massachusetts [132, 137, 301]

Buddha Basnyat, MSc, MD, FACP, FRCP (Edinburgh)
Director, Oxford University Clinical Research Unit—Nepal, Patan Hospital, Kathmandu, Nepal [475]

Joseph Bass, MD, PhD
Division of Endocrinology, Metabolism and Molecular Medicine, Department of Medicine, Feinberg School of Medicine, Department of Neurobiology, Northwestern University, Chicago, Illinois [498]

Shari S. Bassuk, ScD
Epidemiologist, Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts [407]

David W. Bates, MD, MSc
Professor of Medicine, Harvard Medical School; Chief, Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Phyllis Jen Center for Primary Care, Boston, Massachusetts [7]

Elisabeth M. Battinelli, MD, PhD
Associate Professor of Medicine, Harvard Medical School; Associate Chief, Hematology Division, Brigham and Women's Hospital, Boston, Massachusetts [329]

Laurence H. Beck, Jr., MD, PhD
Associate Professor of Medicine, Boston University Chobanian and Avedisian School of Medicine, Boston, Massachusetts [328]

Nicholas J. Beeching, FRCP, FRACP, FFTM RCPS (Glasg), FESCMID, FISTM, DCH, DTM&H
Emeritus Professor of Tropical and Infectious Diseases, Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, United Kingdom [174]

Alex S. Befeler, MD
Medical Director Liver Transplant, Professor of Internal Medicine, Division of Gastroenterology and Hepatology, Saint Louis University, St. Louis, Missouri [355]

Lisa M. Bellini, MD, MACP
Professor of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [9]

Shelley L. Berger, PhD
Daniel S. Och University Professor, Departments of Cell and Developmental Biology; Biology; Genetics; Director, Penn Epigenetics Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania [497]

Jean Bergounioux, MD, PhD
Professor of Medicine, Versailles Saint Quentin University – Paris Saclay, UFR Simone Veil – Motigney le Bretonneux, France; Director, Department of Pediatric Neurology and Intensive Care Medicine, Assistance Publique des Hôpitaux de Paris, Garches, France [172]

John L. Berk, MD
Professor of Medicine, Boston University School of Medicine, Assistant Director, Amyloidosis Center, Boston Medical Center, Boston, Massachusetts [117]

Jeffrey S. Berns, MD
Vice President for Graduate Medical Education and DIO-UPHS Associate Dean for Graduate Medical Education-Perelman School of Medicine Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania [490]

Aaron S. Bernstein, MD, MPH
Assistant Professor of Pediatrics, Harvard Medical School; Hospitalist, Division of General Pediatrics, Boston Children's Hospital, Boston, Massachusetts [474]

Jeanne Bertolli, PhD
Division of High-Consequence Pathogens and Pathology, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [461]

Joseph R. Betancourt, MD, MPH
President of the Commonwealth Fund, Associate Professor of Medicine, Massachusetts General Hospital; Harvard Medical School, Boston, Massachusetts [11]

Emily D. Bethea, MD
Assistant Professor of Medicine, Harvard Medical School; Medical Director of Liver Transplantation; Physician, Gastroenterology and Hepatology Division, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts [348, 356]

Julie A. Bettinger, PhD, MPH
Professor, Department of Pediatrics, Vaccine Evaluation Center, BC Children's Hospital Research Institute, University of British Columbia, Vancouver, British Columbia, Canada [3]

Atul K. Bhan, MBBS, MD
Professor of Pathology, Harvard Medical School, Associate Director, Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital, Boston, Massachusetts [A13]

Shalender Bhasin, MB, BS
Professor of Medicine, Harvard Medical School; Director, Research Program in Men's Health: Aging and Metabolism; Director, Boston Claude D. Pepper Older Americans Independence Center; Brigham and Women's Hospital, Boston, Massachusetts [403, 411]

Deepak L. Bhatt, MD, MPH, MBA, FACC, FAHA, FESC, MSCAI
Director of Mount Sinai Fuster Heart Hospital; Dr. Valentin Fuster Professor of Cardiovascular Medicine, Icahn School of Medicine at Mount Sinai, New York, New York [287, A11]

Roby P. Bhattacharyya, MD, PhD
Assistant Professor of Medicine, Harvard Medical School and Massachusetts General Hospital; Associate Member, Broad Institute of MIT and Harvard, Boston, Massachusetts [126]

David R. Bickers, MD
Carl Truman Nelson Professor, Department of Dermatology, Columbia University Irving Medical Center, New York, New York [64]

William R. Bishai, MD, PhD
Professor of Medicine, Division of Infectious Diseases, Johns Hopkins School of Medicine, Baltimore, Maryland [155]

Bruce R. Bistrian, MD, PhD, MPH
Professor of Medicine, Harvard Medical School; Chief, Clinical Nutrition, Beth Israel Deaconess Medical Center, Boston, Massachusetts [346]

Lucas S. Blanton, MD
Associate Professor, Division of Infectious Diseases, Department of Internal Medicine University of Texas Medical Branch, Galveston, Texas [192]

Martin J. Blaser, MD
Henry Rutgers Chair of the Human Microbiome; Director, Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, New Jersey [168, 169]

William Blum, MD
Director, Acute Leukemia Program; Professor, Department of Hematology and Oncology, Winship Cancer Institute and Emory University, Atlanta, Georgia [109]

Richard S. Blumberg, MD
Vice-Chair for Research in Department of Medicine, Brigham and Women’s Hospital, Professor of Medicine, Harvard Medical School, Boston, Massachusetts [337]

Yair J. Blumenfeld, MD
Professor of Obstetrics & Gynecology (Maternal-Fetal Medicine and Obstetrics), Stanford University School of Medicine, Stanford, California [503]

Jean L. Bolognia, MD
Professor, Department of Dermatology, Yale University School of Medicine, New Haven, Connecticut [61]

Joseph V. Bonventre, MD, PhD
Samuel A. Levine Distinguished Professor, Harvard Medical School; Chief, Renal Division and Engineering in Department of Medicine, Brigham and Women’s Hospital, Boston, Massachusetts [321]

Joshua A. Boyce, MD
Professor of Medicine and Pediatrics; Albert L. Sheffer Professor of Medicine, Harvard Medical School; Director, Inflammation and Allergic Disease Research Section; Chief, Division of Allergy and Clinical Immunology, Brigham and Women’s Hospital, Boston, Massachusetts [363, 364, 366]

Eugene Braunwald, MD
Distinguished Hersey Professor of the Theory and Practice of Medicine, Harvard Medical School; BWH/Founding Chair, TIMI Group, Brigham and Women’s Hospital, Boston, Massachusetts [285]

Irwin M. Braverman, MD
Professor Emeritus; Senior Research Scientist, Department of Dermatology, Yale University School of Medicine, New Haven, Connecticut [61]

Otis W. Brawley, MD, MACP, FRCP(L), FASCO, FACE
Bloomberg Distinguished Professor, Johns Hopkins School of Medicine and Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland [75]

Benjamin L. Brett, PhD
Medical College of Wisconsin, Assistant Professor, Departments of Neurosurgery and Neurology (Division Neuropsychology), Milwaukee, Wisconsin [454]

Manfred Brigl, MD
Assistant Professor of Pathology, Harvard Medical School; Medical Director, Clinical Microbiology Laboratory, Brigham and Women’s Hospital, Boston, Massachusetts [S11]

F. Richard Bringhurst, MD
Associate Professor of Medicine, Harvard Medical School, Endocrine Unit, Massachusetts General Hospital, Boston, Massachusetts [421]

Steven M. Bromley, MD, FAAN
Director, South Jersey MS Center and Bromley Neurology, Audubon and Linwood, New Jersey [35]

Darron R. Brown, MD
Professor of Medicine, Microbiology, and Immunology, Division of Infectious Diseases, Indiana University School of Medicine, Indianapolis, Indiana [203]

Robert H. Brown, Jr., MD, PhD
Director of Neurotherapeutics, Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts [448, 460]

Amy E. Bryant, PhD
Affiliate Research Professor, Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, Idaho State University, Meridian, Idaho [134, 159]

Philip J. Budge, MD, PhD
Associate Professor of Medicine, Division of Infectious Diseases, Washington University School of Medicine in St. Louis, St. Louis, Missouri [240]

G. Scott Budinger, MD
Ernest S. Bazley Professor of Airway Diseases; Chief, Pulmonary and Critical Care Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois [504]

Fred Bunz, MD, PhD
Associate Professor, Johns Hopkins University School of Medicine, Baltimore, Maryland [76]

Lucile Burgo-Black, MD, FACP
National Co-Director, VA Post-Deployment Integrated Care Initiative, Assistant Clinical Professor of Medicine, Department of General Internal Medicine, Yale University School of Medicine, New Haven, Connecticut [S8]

Maxine A. Burkett, JD
Professor of Law, William S. Richardson School of Law, University of Hawai’i at Mānoa, Honolulu, Hawai’i [473]

David M. Burns, MD
Professor Emeritus, University of California, San Diego School of Medicine, Del Mar, California [465]

John C. Byrd, MD
The Gordon and Helen Hughes Taylor Professor and Chair, Department of Internal Medicine, University of Cincinnati College of Medicine, Cincinnati, Ohio [112]

Miguel M. Cabada, MD, MSc, FASTMH
Associate Professor, Division of Infectious Diseases, Department of Medicine, University of Texas Medical Branch, Galveston, Texas; Director, Cusco Branch – Tropical Medicine Institute, Universidad Peruana Cayetano Heredia, Lima, Peru [242]

Rodrigo T. Calado, MD, PhD
Professor of Medicine, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil [482]

Clara Camaschella, MD
Professor of Medicine, Consultant San Raffaele Scientific Institute, Milan, Italy [102]

Michael Camilleri, MD, DSc
Atherton and Winifred W. Bean Professor; Professor of Medicine, Pharmacology, and Physiology, Mayo Clinic College of Medicine and Science, Rochester, Minnesota [49]

Christopher P. Cannon, MD
Professor of Medicine, Harvard Medical School; Education Director, Cardiovascular Innovation, Preventive Cardiology Section, Brigham and Women’s Hospital, Boston, Massachusetts [285]

Brian C. Capell, MD, PhD
Assistant Professor of Dermatology and Genetics, Departments of Dermatology and Genetics, Penn Epigenetics Institute, Abramson Cancer Center, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania [497]

Jonathan Carapetis, MBBS, PhD
Executive Director, The Kids Research Institute Australia, Nedlands; Professor, University of Western Australia, Perth; Consultant, Perth Children’s Hospital, Nedlands, Australia [371]

Arturo Casadevall, MD, PhD
Professor and Chair, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland [221]

Mariana Castells, MD, PhD
Professor of Medicine, Harvard Medical School; Director, Drug Hypersensitivity and Desensitization Center; Director, Mastocytosis Center, Brigham and Women’s Hospital; Dana Farber Cancer Institute, Boston, Massachusetts [365]

Frederic Castinetti, MD, PhD
Professor of Endocrinology, Marseille Medidal Genetics, Aix Marseille University, INSERM; Chief, Department of Endocrinology, La Conception Hospital, Assistance Publique Hopitaux de Marseille, Marseille, France [399]

Jonathan Cedernaes, MD, PhD
Senior Researcher, Division of Endocrinology, Metabolism and Molecular Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois [498]

Wilma Chan, MD, EdM
Associate Professor of Emergency Medicine, University of Pennsylvania Perelman School of Medicine; Director, Ultrasound Education at Perelman School of Medicine, University of Pennsylvania Health System, Philadelphia, Pennsylvania [493]

Anil K. Chandraker, MB, ChB
Professor, UMass Chan Medical School; Lecturer on Medicine, Harvard Medical School; Chief, Division of Renal Medicine, Brigham and Women's Hospital, Boston, Massachusetts [325]

Vinay Chandrasekhara, MD
Professor of Medicine; Director, Advanced Endoscopy and Pancreas Groups; Program Director, Advanced Endoscopy Fellowship, Mayo Clinic, Rochester, Minnesota [333, V5]

François Chappuis, MD, MCTM, PhD
Head of Division, Division of Tropical and Humanitarian Medicine, Geneva University Hospitals, Geneva, Switzerland [234]

Justin T. Cheeley, MD, FAAD
Assistant Professor, Divisions of Dermatology and Internal Medicine and Geriatrics, Emory University School of Medicine, Atlanta, Georgia [60, A5]

Kartikeya Cherabuddi, MD
Chief Epidemiology Officer, Tampa General Hospital; Professor of Medicine, Division of Infectious Diseases, University of South Florida, Tampa, Florida [170]

Daniel S. Chertow, MD, MPH
United States Public Health Service; Senior Investigator, National Institutes of Health; Head, Emerging Pathogens Section, Critical Care Medicine Department, NIH Clinical Center; Laboratory of Virology, National Institute of Allergy and Infectious Diseases; and Critical Care Medicine Brach, National Heart Lung and Blood Institute, Bethesda, Maryland [315]

Glenn M. Chertow, MD, MPH
Professor of Medicine, Division of Nephrology, Stanford University School of Medicine, Palo Alto, California [323]

Jacques Chiaroni, MD, PhD
Professor of Medicine, University of Aix Marseille University, ULR7268, ADES, Marseille, Etablissement Francais du Sang Provence Alpes Côté d'Azur et Corse, Marseille, France [118]

Hyon K. Choi, MD, DrPH
Professor of Medicine, Harvard Medical School; Director, Gout and Crystal Arthropathy Center; Director, Clinical Epidemiology and Health Outcomes, Division of Rheumatology, Allergy, and Immunology, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts [384]

Benjamin F. Chong, MD, MSCS
Associate Professor, Department of Dermatology, University of Texas Southwestern Medical Center, Dallas, Texas [62]

Mirjam Christ-Crain, MD, PhD
Professor of Endocrinology, Medical faculty, University of Basel, Switzerland, Deputy Chief, Clinic of Endocrinology, Diabetes and Metabolism, University Hospital of Basel, Switzerland [393]

Raymond T. Chung, MD
Professor of Medicine, Harvard Medical School; Zhou Family Endowed Chair in Gastroenterology; Director of Hepatology and Liver Center; Vice Chief, Division of Gastroenterology, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts [350, 352, 356]

Jeffrey W. Clark, MD
Associate Professor of Medicine, Harvard Medical School; Member, NCCN Investigator Steering Committee; Gastrointestinal Cancer Center, Massachusetts General Hospital, Boston, Massachusetts [77]

Bruce H. Cohen, MD
Professor of Pediatrics; Professor of Integrative Medical Science, Northeast Ohio Medical University; Director, NeuroDevelopmental Science Center and Director, Scientific Affairs and Clinical Trials, Rebecca D. Considine Research Institute, Akron Children's Hospital, Akron, Ohio [481]

Jeffrey I. Cohen, MD
Chief, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [196, 199, 210]

Steven P. Cohen, MD
Vice Chair of Pain Medicine; Professor of Anesthesiology, Physical Medicine and Rehabilitation, Neurology, Psychiatry and Neurosurgery, Northwestern Feinberg School of Medicine, Chicago, Illinois; Director of Pain Research; Professor of Physical Medicine and Rehabilitation and Anesthesiology, Walter Reed National Military Medical Center, Uniformed Services University of the Health Sciences, Bethesda, Maryland [18, V8]

Amanda Cohn, MD
Director, Division of Birth Defects and Infant Disorders, National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia [129]

Jennifer P. Collins, MD, MSc
Medical Epidemiologist, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [156]

Charlene W. Compher, PhD
Professor of Nutrition Science, Biobehavioral Health Science Department, University of Pennsylvania, School of Nursing; Advanced Practice Clinical Dietitian Specialist, Philadelphia, Pennsylvania [345]

Wilson M. Compton, MD, MPE
Deputy Director, National Institute on Drug Abuse, National Institutes of Health, U.S. Department of Health and Human Services, Bethesda, Maryland [468]

Jean M. Connors, MD
Hematology Division, Brigham and Women's Hospital; Professor of Medicine, Harvard Medical School, Boston, Massachusetts [121]

Darwin L. Conwell, MD, MS
Professor of Medicine, Chair, Department of Internal Medicine, University of Kentucky, Lexington, Kentucky [358, 359]

Lawrence Corey, MD
Professor of Medicine and Laboratory Medicine and Pathology, University of Washington; Past President and Director, Fred Hutchinson Cancer Research Center; Professor, Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington [197]

Nicolás W. Cortés-Penfield, MD, FACP, FIDSA
Assistant Professor in Infectious Diseases, University of Nebraska Medical Center; Medical Director, Outpatient Parenteral Antimicrobial Therapy Program and Co-Director, Orthopedic Infectious Diseases Program, Omaha, Nebraska [140]

Sara E. Cosgrove, MD, MS
Professor of Medicine, Division of Infectious Diseases, Johns Hopkins University School of Medicine, Baltimore, Maryland [133]

Darrell H. G. Crawford, MD, FRACP, FAASLD, FGESA
Emeritus Professor, The University of Queensland; Director of Research, Gallipoli Medical Research, Greenslopes Private Hospital, Brisbane, Australia [426]

Mark A. Creager, MD
Professor of Medicine and Surgery, Geisel School of Medicine at Dartmouth, Heart and Vascular Center, Dartmouth-Hitchcock Medical Center, One Medical Center Drive, Lebanon, New Hampshire [291, 292, 293]

Bruce A. C. Cree, MD, PhD, MAS
George A. Zimmermann Endowed Professor in Multiple Sclerosis; Professor of Clinical Neurology; Clinical Research Director, UCSF Weill Institute for Neurosciences, Department of Neurology, University of California San Francisco, San Francisco, California [455, 456]

Leslie J. Crofford, MD
Professor, Departments of Medicine and Pathology, Microbiology and Immunology, Vanderbilt University; Chief, Division of Rheumatology and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee [385]

Jennifer M. Croswell, MD, MPH
Medical Officer, Healthcare Delivery Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland [75]

James E. Crowe, Jr., MD
Director, Vanderbilt Vaccine Center; Ann Scott Carell Chair and Professor, Departments of Pediatrics, Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee [204, 205]

Ian Crozier, MD
Clinical Monitoring Research Program Directorate, Frederick National Laboratory for Cancer Research, Frederick, Maryland [215, 216]

Alia J. Crum, PhD
Associate Professor of Psychology and, by courtesy, of Medicine (Primary Care & Population Health), Stanford University, Stanford, California [496]

Jennifer M. Cuellar-Rodriguez, MD
Director, Transplant Infectious Diseases Consult Service, NIH Clinical Research Center; and Staff Clinician, Laboratory of Clinical Immunology and Microbiology, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [148]

Gary C. Curhan, MD, ScD
Senior Lecturer, Harvard Medical School, Senior Investigator, Channing Division of Network Medicine; Physician, Renal Division, Brigham and Women's Hospital, Boston, Massachusetts [330]

Brendan D. Curti, MD
Director, Melanoma Program; Robert W. Franz Endowed Chair for Clinical Research, Earle A. Chiles Research Institute, a Division of the Providence Cancer Institute, Portland, Oregon [81]

John J. Cush, MD
Executive Editor, RheumNow.com; Professor of Internal Medicine, TCU Burnett School of Medicine, Fort Worth, Texas [382]

Kelly A. Cuttle, PhD
Chief of Staff, Duke Human Vaccine Institute, Department of Medicine, Duke University School of Medicine; Duke University Medical Center, Durham, North Carolina [360, 361]

Charles A. Czeisler, MD, PhD
Frank Baldino, Jr., PhD Professor of Sleep Medicine, Professor of Medicine and Director, Division of Sleep Medicine, Harvard Medical School; Chief, Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, Massachusetts [33]

Josep Dalmau, MD, PhD
Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS)-CaixaResearch Institute, Hospital Clínic de Barcelona, Barcelona, Spain; Adjunct Professor, University of Pennsylvania, Philadelphia, Pennsylvania [99]

Carolyn D'Ambrosio, MS, MD
Associate Professor of Internal Medicine, Yale University; Division of Pulmonary, Critical Care and Sleep Medicine, Yale University School of Medicine, New Haven, Connecticut [41]

Daniel F. Danzl, MD
Professor and Emeritus Chair, Department of Emergency Medicine, University of Louisville, Louisville, Kentucky [477, 478]

Robert B. Daroff, MD*
Professor and Chair Emeritus, Department of Neurology, Case Western Reserve University School of Medicine; University Hospitals – Cleveland Medical Center, Cleveland, Ohio [24]

Nancy E. Davidson
Executive Vice President Clinical Affairs and Professor, Fred Hutchinson Cancer Center; Professor of Medicine, University of Washington Medical Center, Seattle, Washington [84]

Silas A. Davidson, PhD
Entomologist and Force Health Protection Officer, 18th Medical Command, US Army, Fort Shafter, Hawaii [472]

Charles E. Davis, MD
Professor of Pathology and Medicine, Emeritus, University of California, San Diego School of Medicine; Director Emeritus, Microbiology, University of California, San Diego Medical Center, La Jolla, California [228]

Stephen N. Davis, MBBS, FRCP, FACE, MACP
Theodore E. Woodward Professor of Medicine; Professor of Physiology; Chairman, Department of Medicine, University of Maryland School of Medicine; Director, Institute for Clinical and Translational Research; Vice President of Clinical Translational Science University of Maryland, Baltimore; Physician-in-Chief, University of Maryland Medical Center, Baltimore, Maryland [418]

Erin M. DeBiasi, MD
Assistant Professor of Internal Medicine, Yale University; Division of Pulmonary, Critical Care and Sleep Medicine, Yale University School of Medicine, New Haven, Connecticut [41]

Rafael de Cabo, PhD
Chief, Translational Gerontology Branch, National Institute on Aging, National Institutes of Health, Baltimore, Maryland [488]

Lucia De Franceschi, MD
Department of Engineering for Innovative Medicine, University of Verona and Azienda Ospedaliera Universitaria Integrata Verona, Verona, Italy [105]

Thomas G. DeLoughery, MD, MACP, FAWM
Professor of Medicine, Pathology and Pediatrics, Oregon Health and Science University, Portland, Oregon [66]

John Del Valle, MD
Professor and Vice Chair of Medicine, Department of Internal Medicine, University of Michigan School of Medicine, Ann Arbor, Michigan [335]

Dawn DeMeo, MD, MPH
Associate Professor, Harvard Medical School; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts [303]

David W. Denning, MBBS, FRCP, FRCPath, FMedSci
Professor of Infectious Diseases in Global Health, The University of Manchester, Manchester, United Kingdom [223]

Atul Deodhar, MD
Professor of Medicine; Medical Director, Rheumatology Clinics, Oregon Health and Science University, Portland, Oregon [374]

Akshay Desai, MD, MPH
Associate Professor of Medicine, Harvard Medical School; Center for Advanced Heart Disease, Cardiovascular Division, Brigham and Women's Hospital, Boston, Massachusetts [265]

Robert J. Desnick, PhD, MD
Dean for Genetic and Genomic Medicine Emeritus; Professor and Chairman Emeritus, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, Mount Sinai Health System, New York, New York [428]

Marcelo F. DiCarli, MD
Professor, Department of Radiology and Medicine, Harvard Medical School; Chief, Division of Nuclear Medicine and Molecular Imaging; Executive Director, Noninvasive Cardiovascular Imaging Program, Brigham and Women's Hospital, Boston, Massachusetts [248, A9]

Anna Mae Diehl, MD
Florence McAlister Professor of Medicine; Director, Duke Liver Center, Duke University, Durham, North Carolina [354]

David J. Diemert, MD, FRCP(C)
Professor, Department of Medicine and Department of Microbiology, Immunology and Tropical Medicine, George Washington University School of Medicine and Health Sciences, Washington, DC [239]

Jules L. Dienstag, MD
Carl W. Walter Professor of Medicine, Harvard Medical School; Physician, Gastroenterology and Hepatology Division, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts [350, 351, 352, 356, A13]

---

*Deceased

William P. Dillon, MD
Emeritus Professor and Executive Vice-Chair, Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, California [434, A16]

Charles A. Dinarello, MD
Distinguished Professor of Medicine, University of Colorado School of Medicine, Aurora, Colorado [20]

Yohei Doi, MD, PhD
Professor of Medicine, Division of Infectious Diseases, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; Professor of Medicine, Departments of Microbiology and Infectious Diseases, Fujita Health University School of Medicine, Toyoake, Japan [166]

R. Christopher Doiron, MD, MPH, FRCS(C)
Assistant Professor, Department of Urology, Queen's University, Kingston; Attending Urologist, Kingston Health Sciences Centre, Kingston, Ontario, Canada [54]

Anuja Dokras, MD, MHCI, PhD
Founders Professor of Women's Health, Department of Obstetrics and Gynecology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [404, 405, 408]

Susan M. Domchek, MD
Basser Professor of Oncology, Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [480]

Richard L. Doty, PhD, FAAN
Professor, Department of Otorhinolaryngology: Head and Neck Surgery; Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania Director, Smell and Taste Evaluation Center, Haddon Heights, New Jersey [35]

Vanja C. Douglas, MD
Professor of Neurology and Sara and Evan Williams Foundation Endowed Neurohospitalist Chair, University of California, San Francisco, San Francisco, California [25, 452]

Thomas D. DuBose, Jr., MD, MACP
Emeritus Professor of Medicine, Wake Forest School of Medicine, Winston Salem, North Carolina [58, S1]

Patricia Dugan, MD
Professor of Neurology (Clinical), NYU Grossman School of Medicine; Director, Adult Epilepsy; Director, Epilepsy Fellowship Program, NYU Langone Health, New York, New York [436]

J. Stephen Dumler, MD
Professor and Chairman, Department of Pathology, Uniformed Services University of the Health Sciences, Walter Reed National Military Medical Center, Joint Pathology Center, Bethesda, Maryland [192]

Andrea Dunaif, MD
Lillian and Henry M. Stratton Professor of Molecular Medicine, System Chief, Hilda and J. Lester Gabrilove Division of Endocrinology, Diabetes and Bone Disease, Icahn School of Medicine and Mount Sinai Health System, New York, New York [410]

Samuel C. Durso, MD, MBA
Jim and Patty Dresher and Dresher Family Foundation Professor of Medicine; Chair, Department of Medicine, Johns Hopkins Bayview Medical Center, Baltimore, Maryland [38, A3]

Janice P. Dutcher, MD
Associate Director, Cancer Research Foundation of New York, Chappaqua, New York; Former Professor of Medicine, Albert Einstein College of Medicine; Former Professor of Medicine, New York Medical College, Valhalla, New York [80]

Johanna T. Dwyer, DSc, RD
Professor of Medicine and Community Health, Schools of Medicine and Friedman School of Nutrition Science and Policy at Tufts University and Senior Nutrition Scientist (Contractor) ICF serving the Office of Dietary Supplements, National Institutes of Health), Washington DC [343]

Kim A. Eagle, MD
Albion Walter Hewett Professor of Internal Medicine; Director, Samuel and Jean Frankel Cardiovascular Center, University of Michigan, Ann Arbor, Michigan [492]

Sarah Rae Easter, MD
Assistant Professor of Obstetrics, Gynecology and Reproductive Biology, Harvard Medical School; Director, Obstetrical Critical Care, Brigham and Women's Hospital, Boston, Massachusetts [491]

James A. Eastham, MD
Chief, Urology Service; Peter T. Scardino Chair in Oncology, Department of Surgery, Sidney Kimmel Center for Prostate and Urologic Cancers, Memorial Sloan Kettering Cancer Center, New York, New York [92]

Peter R. Ebeling, MBBS, MD, FRACP, FRCP
Head, School of Clinical Sciences, Monash University; Chair, Division of Medicine, Monash Health, Clayton, Victoria, Australia [423]

Robert T. Eberhardt, MD
Boston Medical Center, Clinical Professor of Medicine and Surgery, Sections of Cardiovascular Medicine and Vascular and Endovascular Surgery, BU Chobanian and Avedisian School of Medicine, Boston, Massachusetts [293]

Robert H. Eckel, MD
Professor of Medicine, Emeritus, University of Colorado Anschutz Medical Campus, Division of Endocrinology, Metabolism and Diabetes; Division of Cardiology, Aurora, Colorado [420]

John E. Edwards, Jr., MD
Distinguished Professor of Medicine Emeritus, David Geffen School of Medicine, University of California, Los Angeles; Senior Investigator, The Lundquist Institute and Emeritus Chief, Division of Infectious Disease at Harbor-UCLA Medical Center, Torrance, California [217, 222]

Kathryn M. Edwards, MD
Professor of Pediatrics Emerita, Vanderbilt University Medical Center, Nashville, Tennessee [128]

David A. Ehrmann, MD
Professor of Medicine, Section of Endocrinology; Director, University of Chicago Center for PCOS, University of Chicago, Chicago, Illinois [406]

Dirk Elewaut, MD, PhD
Professor of Medicine, Department of Internal Medicine and Paediatrics; Group Leader, VIB Center for Inflammation Research, Ghent University; Chair, Department of Rheumatology, Ghent University Hospital, Ghent, Belgium [374]

Emad M. El-Omar, MB ChB, MD
Professor of Medicine, University of New South Wales Sydney; Director, Microbiome Research Centre, St George and Sutherland Clinical Campuses, Kogarah, New South Wales, Australia [168]

Ramy H. Elshaboury, PharmD
Director, Clinical Pharmacy Services; Director, Graduate Pharmacy Education, Massachusetts General Hospital, Boston, Massachusetts [149]

May S. ElSherif, MD, MPH, BSc
Adjunct Assistant Professor, Department of Microbiology and Immunology, Dalhousie University, Halifax, Nova Scotia, Canada [165]

Ezekiel J. Emanuel, MD, PhD
Vice Provost for Global Initiatives; Co-Director, Healthcare Transformation Institute; Levy University Professor, Perelman School of Medicine and The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania [13]

Jack Ende, MD, MACP
The Schaeffer Professor of Medicine; Executive Medical Director, Penn Signature Services, Perelman School of Medicine of the University of Pennsylvania, Philadelphia, Pennsylvania [490]

Moshe Ephros, MD
Clinical Associate Professor of Pediatrics, Faculty of Medicine, Technion – Israel Institute of Technology, Haifa, Israel [177]

Aaron C. Ermel, MD
Associate Professor of Clinical Medicine, Department of Internal Medicine, Division of Infectious Diseases, Indiana University School of Medicine, Indianapolis, Indiana [203]

Tim Evans, DPhil, MD
Vice President Research, Innovation and Impact, Concordia University, Montreal, Quebec, Canada [487]

Carmella Evans-Molina, MD, PhD
Director, Indiana Diabetes Research Center; Eli Lilly and Company Professor of Pediatric Diabetes; Professor of Pediatrics, Medicine, Indiana University School of Medicine, Indianapolis, Indiana [415]

Rodney H. Falk, MD
Associate Professor of Medicine, Harvard Medical School; Director, Amyloidosis Program, Brigham and Women's Hospital, Boston, Massachusetts [270]

Paul E. Farmer, MD, PhD*
Kolokotrones University Professor, Harvard University; Chair, Department of Global Health and Social Medicine, Harvard Medical School; Chief, Division of Global Health Equity, Brigham and Women's Hospital; Chief Strategist, Co-Founder, Partners In Health, Boston, Massachusetts [485]

Sadaf Farooqi, PhD, FRCP, FMedSci, FRS
Professor of Metabolism and Medicine, Wellcome-MRC Institute of Metabolic Science, University of Cambridge; Hon Consultant Physician, Addenbrooke's Hospital, Cambridge, United Kingdom [413]

Anthony S. Fauci, MD
Distinguished University Professor, Georgetown University, School of Medicine and McCourt School of Public Policy, Washington, DC [207, 208, 360, 361, 375, 486, S4, A14]

Michael G. Fehlings, MD, PhD
Division of Genetics and Development, Krembil Brain Institute, University Health Network, Toronto, Ontario, Canada [19]

David Feller-Kopman, MD
Professor of Medicine, Dartmouth Geisel School of Medicine; Chief, Pulmonary and Critical Care Medicine, Dartmouth Hitchcock Medical Center, Lebanon, New Hampshire [310]

David T. Felson, MD, MPH
Professor of Medicine and Epidemiology, Boston University Schools of Medicine and Public Health, Boston, Massachusetts [383]

Marcela A. Ferrada, MD
Clinical Associate Professor, Division of Rheumatology, Allergy and Clinical Immunology, Department of Medicine, University of Maryland School of Medicine, Bethesda, Maryland [378]

Howard L. Fields, MD, PhD
Professor, Department of Neurology, University of California, San Francisco, San Francisco, California [14]

Gregory A. Filice, MD
Staff Physician, Veterans Affairs Medical Center, Professor of Medicine and Adjunct Professor of Public Health, University of Minnesota, Minneapolis, Minnesota [179]

Courtney Finlayson, MD
Associate Professor, Division of Endocrinology, Department of Pediatrics, Ann and Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine, Chicago, Illinois [402]

Alain Fischer, MD, PhD
Imagine Institute; Emeritus Professor at College de France, Paris, France [362, S9]

Erik Fisher, MD
Clinical Assistant Professor of Emergency Medicine, University of South Carolina School of Medicine Greenville; Director, Medical Toxicology, Department of Emergency Medicine, Prisma Health-Update, Greenville, South Carolina [471]

Dinah Foer, MD
Assistant Professor of Medicine, Harvard Medical School; Associate Physician, Division of Allergy and Clinical Immunology, Brigham and Women's Hospital, Boston, Massachusetts [363]

Agnes B. Fogo, MD
John L. Shapiro Endowed Chair in Pathology; Professor of Medicine and Pediatrics; Director, Renal Pathology/Electron Microscopy Laboratory, Vanderbilt University Medical Center, Nashville, Tennessee [A4]

Gregory K. Folkers, MS, MPH
Washington, DC [208]

David M. Frazer, PhD
Associate Professor, Molecular Nutrition Laboratory, QIMR Berghofer, Brisbane, Australia [426]

Jane E. Freedman, MD
Director, Division of Cardiology, Physician-in-Chief, Vanderbilt Medical Center, Nashville, Tennessee [122]

Roy Freeman, MD
Professor of Neurology, Harvard Medical School; Director, Center for Autonomic and Peripheral Nerve Disorders, Beth Israel Deaconess Medical Center, Boston, Massachusetts [23]

Lawrence S. Friedman, MD
Professor of Medicine, Harvard Medical School; Professor of Medicine, Tufts University School of Medicine; The Anton R. Fried, MD Chair, Department of Medicine, Newton-Wellesley Hospital, Newton, Massachusetts; Assistant Chief of Medicine, Massachusetts General Hospital, Boston, Massachusetts [53]

Sonia Friedman, MD
Professor of Medicine, Tufts University School of Medicine; Chief of Gastroenterology, Tufts Medical Center, Boston, Massachusetts [337]

Andre D. Furtado, MD
Assistant Professor, Department of Radiology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania [A16]

Nicholas B. Galifianakis, MD, MPH
Associate Professor of Neurology, Movement Disorder and Neuromodulation Center Weill Institute for Neurosciences, Department of Neurology, University of California, San Francisco, San Francisco, California [V1]

John I. Gallin, MD*
Associate Director for Clinical Research; Chief Scientific Officer, Clinical Center, National Institutes of Health, Bethesda, Maryland [67]

Karunesh Ganguly, MD, PhD
Department of Neurology, University of California, San Francisco; Neurology and Rehabilitation Service, San Francisco VA Medical Center, San Francisco, California [500]

Olivier Garraud, MD, PhD
Professor of Medicine, Lyon University , Saint-Etienne Faculty of Medicine, Saint-Etienne, France [118]

Gregory M. Gauthier, MD
Associate Professor, Department of Medicine, Division of Infectious Disease, School of Medicine and Public Health, University of Wisconsin, Madison, Madison, Wisconsin [220]

J. Michael Gaziano, MD
Professor of Medicine, Harvard Medical School; Departments of Medicine, Brigham and Women's Hospital, VA Boston Healthcare System, Boston, Massachusetts [245]

Thomas A. Gaziano, MD, MSC
Associate Professor, Harvard Medical School; Associate Professor, Department of Health Policy and Management, Harvard T.H. Chan School of Public Health, Cardiovascular Medicine, Brigham and Women's Hospital, Boston, Massachusetts [245]

Juan C. Gea-Banacloche, MD
Senior Consultant in Infectious Diseases, Division of Clinical Research, National Institute of Allergy and Infectious Diseases; Staff Clinician, Transplant Infectious Diseases Consult Service, NIH Clinical Research Center, Bethesda, Maryland [79, 148]

Susan L. Gearhart, MD, MEd
Professor of Surgery, Division of Colorectal Surgery, Johns Hopkins Medical Institutions, Baltimore, Maryland [339]

Jeffrey A. Gelfand, MD
Professor of Medicine, Harvard Medical School; Attending Physician, Infectious Diseases Division, Massachusetts General Hospital, Boston, Massachusetts [232]

---

*Deceased

Jeffrey M. Gelfand, MD, MAS
Associate Professor of Neurology, University of California, San Francisco (UCSF), San Francisco, California [25]

Alfred L. George, Jr., MD
Professor and Chair, Department of Pharmacology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois [320]

Dale N. Gerding, MD
Research Physician, Edward Hines Jr. VA Hospital, Hines, Illinois; Professor of Medicine (Retired), Loyola University Chicago Stritch School of Medicine, Maywood, Illinois [139]

Alicia K. Gerke, MD
Associate Professor of Medicine, Division of Pulmonary, Critical Care, and Occupational Medicine, University of Iowa, Iowa City, Iowa [379]

Michael D. Geschwind, MD, PhD
Professor of Neurology; Michael J. Homer Chair in Neurology, Memory and Aging Center, University of California, San Francisco, San Francisco, California [449, V1]

Marc G. Ghany, MD, MHSc
Chief, Clinical Hepatology Research Section; Tenure-Track Investigator, Liver Diseases Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland [196, 347]

Lorenzo Giacani, PhD
Professor of Medicine, Department of Medicine, Division of Allergy and Infectious Diseases, University of Washington, Seattle, Washington [188]

Matthew P. Giannetti, MD
Division of Allergy and Clinical Immunology, Brigham and Women's Hospital; Harvard Medical School, Boston, Massachusetts [366]

Michael Giladi, MD, MSc
Associate Professor of Medicine, Sackler Faculty of Medicine, Tel Aviv University; Senior Physician, The Infectious Disease Unit; Director, The Bernard Pridan Laboratory for Molecular Biology of Infectious Diseases, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel [177]

Domenico Girelli, MD, PhD
Professor of Medicine, University of Verona; Chief, Internal and Emergency Medicine, University Hospital, Verona, Italy [102]

Robert P. Giugliano, MD, SM, FACC, FAHA
Professor of Medicine, Harvard Medical School; Senior Investigator, TIMI Study Group; Senior Physician, Cardiovascular Medicine, Brigham and Women's Hospital, Boston, Massachusetts [284, 285]

Michael M. Givertz, MD
Professor of Medicine, Harvard Medical School; Medical Director, Heart Transplant and Mechanical Circulatory Support, Brigham and Women's Hospital, Boston, Massachusetts [264]

Roger I. Glass, MD, PhD
President, The CMB Foundation, Atlanta, Georgia [209]

Seth R. Glassman, MD
Assistant Clinical Professor of Medicine, Division of Infectious Diseases, University at Buffalo Jacobs School of Medicine and the Biomedical Sciences, Buffalo, New York [181]

Peter J. Goadsby, MBBS, MD, PhD, DSc, FRACP, FRCP, FMedSci, FRS
Professor of Neurology and Director, King's Clinical Research Facility, King's College, London, United Kingdom; Professor (Emeritus), Department of Neurology, University of California, Los Angeles, California [17, 441]

Marco Goeijenbier, MD, PhD, EDIC
Internist, Acute Internal Medicine and Intensive Care, Department of Intensive Care, Spaarne Gasthuis, Haarlem/Hoofddorp, The Netherlands; Senior Researcher, Department of Intensive Care, Erasmus Medical Center, Rotterdam, The Netherlands [189]

Hilary J. Goldberg, MD, MPH
Assistant Professor of Medicine, Harvard Medical School; Medical Director, Lung Transplant Program; Clinical Director, Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts [297, 305, 306, 309]

Marcia B. Goldberg, MD
Professor of Medicine and of Microbiology, Harvard Medical School; Professor of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health; Distinguished Physician, Massachusetts General Hospital, Boston, Massachusetts [125]

Ary L. Goldberger, MD
Professor of Medicine, Harvard Medical School; Director, Margret and H.A. Rey Institute for Nonlinear Dynamics in Medicine; Chief, Division of Interdisciplinary Medicine and Biotechnology (IMBIO); Beth Israel Deaconess Medical Center, Boston, Massachusetts [247, A7, A8]

David Goldblatt, MB, ChB, PhD
Professor of Vaccinology and Immunology, University College London Great Ormond Street Institute of Child Health, London, United Kingdom [151]

Samuel Z. Goldhaber, MD
Professor of Medicine, Harvard Medical School; Director, Thrombosis Research Group, Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston, Massachusetts [290]

Andrea Gori, MD
Full Professor of Infectious Diseases, Department of Infectious Diseases, "Ospedale Luigi Sacco," Co-Director Centre for Multidisciplinary Research in Health Science (MACH), University of Milan, Milan, Italy [183]

Maria Luisa Gorno-Tempini, MD, PhD
Professor, Department of Neurology; Language Neurobiology Lab, Memory and Aging Center; Dyslexia Center, University of California, San Francisco, San Francisco, California [V2]

Daniel J. Gottlieb, MD, MPH
Associate Professor of Medicine, Harvard Medical School; Director, Sleep Disorders Center, VA Boston Healthcare System, Sleep Medicine Division, Brigham and Women's Hospital, Boston, Massachusetts [308]

Peter A. Gottlieb, MD
Professor of Pediatrics and Medicine, Barbara Davis Center for Childhood Diabetes, University of Colorado School of Medicine, Aurora, Colorado [401]

Gregory A. Grabowski, MD
Professor Emeritus, University of Cincinnati College of Medicine; Departments of Pediatrics, and Molecular Genetics, Biochemistry and Microbiology, Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio [429]

Yonatan H. Grad, MD, PhD
Professor of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts [126]

Christine Grady, RN, PhD
Chief, Department of Bioethics, National Institutes of Health Clinical Center, Bethesda, Maryland [12]

Francesc Graus, MD, PhD
Neuroimmunology Program, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Hospital Clínic de Barcelona, Barcelona, Spain [99]

Peter C. Grayson, MD, MSc
Tenure Track Investigator, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland [378]

Steven A. Greenberg, MD, PhD
Professor of Neurology, Harvard Medical School; Associate Neurologist, Brigham and Women's Hospital, Boston, Massachusetts [377]

Steven M. Greenberg, MD, PhD
Professor of Neurology, Harvard Medical School; Vice Chair of Neurology, Massachusetts General Hospital, MGH Stroke Research Center, Boston, Massachusetts [444]

Norton J. Greenberger, MD*
Clinical Professor of Medicine, Harvard Medical School; Senior Physician, Division of Gastroenterology, Brigham and Women's Hospital, Boston, Massachusetts [357]

Daryl R. Gress, MD
Professor of Neurological Sciences; Director of Neurocritical Care, University of Nebraska Medical Center, Omaha, Nebraska [318]

Patricia M. Griffin, MD
Adjunct Professor, Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, Georgia [156]

Emma M. Groarke, MB, BCh, BAO
Associate Research Physician, Hematology Branch, National Heart, Lung and Blood Institute, Bethesda, Maryland [107]

Rasim Gucalp, MD, FACP
Professor of Medicine/Oncology, Albert Einstein College of Medicine; Associate Chairman for Educational Programs, Department of Oncology; Director, Hematology/Oncology Fellowship, Montefiore Medical Center, Bronx, New York [80]

Amanda C. Guidon, MD, MPH
Assistant Professor, Harvard Medical School; Chief, Division of Neuromuscular Medicine; Director, EMG Laboratory, Massachusetts General Hospital (MGH); Program Director, MGB/Harvard Neuromuscular Medicine Fellowship, Director, MGH Myasthenia Gravis Program, Boston, Massachusetts [459]

Kalpana Gupta, MD, MPH
Associate Chief of Staff and Chief, Infectious Diseases, Veterans Affairs Boston Healthcare System, West Roxbury, Massachusetts; Professor of Medicine, Boston University School of Medicine, Boston, Massachusetts [140]

Chadi A. Hage, MD, ATSF
Professor of Medicine, University of Pittsburgh; Division of Pulmonary, Allergy, Critical Care, and Sleep Medicine; Medical Director, Lung Transplant, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania [218]

Noah M. Hahn, MD
Professor of Oncology and Urology, Johns Hopkins University School of Medicine; Johns Hopkins University Greenberg Bladder Cancer Institute, Baltimore, Maryland [91]

Colin N. Haile, MD, PhD
Associate Research Professor of Psychology, University of Houston, Houston, Texas [467]

Janet E. Hall, MD
Senior Investigator, Division of Intramural Research, NIH/NIEHS, Research Triangle Park, North Carolina [404, 405, 408]

Kathryn T. Hall, PhD, MPH
Assistant Professor of Medicine, Part Time, Harvard Medical School; Associate Molecular Biologist, Brigham and Women's Hospital, Boston, Massachusetts; Senior Vice President of Research, New York Academy of Medicine, New York, New York [496]

Scott A. Halperin, MD
Professor of Pediatrics and Microbiology and Immunology, Dalhousie University, Halifax, Nova Scotia, Canada [165]

Matthew M. Hamill, MBChB, PhD
Assistant Professor of Medicine, Johns Hopkins School of Medicine, Division of Infectious Diseases, Bayview Medical Center, Baltimore, Maryland [194]

Aidan Hampson, PhD
Senior Advisor and Special Content Expert on Cannabis, Division of Therapeutics & Medical Consequences, National Institute on Drug Abuse, National Institutes of Health, Rockville, Maryland [466]

R. Doug Hardy, MD
Prism Health North Texas, Dallas, Texas [193]

Phil A. Hart, MD
Professor of Medicine; Director, Section of Pancreatic Disorders, Division of Gastroenterology, Hepatology, and Nutrition, The Ohio State University Wexner Medical Center, Columbus, Ohio [358, 359]

William L. Hasler, MD
Professor of Medicine, Division of Gastroenterology and Hepatology, Mayo Clinic Arizona, Scottsdale, Arizona [48, 332]

Stephen L. Hauser, MD
Robert A. Fishman Distinguished Professor, Department of Neurology; Director, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, California [26, 27, 30, 433, 435, 452, 453, 455, 456, 458, S3]

Thomas R. Hawn, MD, PhD
Professor, Department of Medicine, Division of Allergy and Infectious Diseases, University of Washington, Seattle, Washington [164]

Barton F. Haynes, MD
Director, Duke Human Vaccine Institute; Frederic M. Hanes Professor of Medicine; Professor of Integrative Immunobiology, Departments of Medicine and Immunology, Duke University Medical Center, Durham, North Carolina [360, 361]

J. Claude Hemphill, III, MD, MAS
Professor of Neurology and Neurological Surgery, University of California, San Francisco; Chief, Neurology Service, Zuckerberg San Francisco General Hospital, San Francisco, California [318, 437–440]

Jan-Inge Henter, MD, PhD
Professor of Pediatrics, Department of Women's and Children's Health, Karolinska Institute; Senior Consultant, Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden [68]

Dirk M. Hentschel, MD
Assistant Professor of Medicine, Harvard Medical School; Director of Interventional Nephrology, Brigham Health; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts [324]

Craig P. Hersh, MD, MPH
Associate Professor of Medicine, Harvard Medical School; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts [303]

Katherine A. High, MD
Professor Emerita, Perelman School of Medicine of the University of Pennsylvania, Philadelphia, Pennsylvania; Visiting Professor, Rockefeller University, New York, New York [483]

Christine E. Hill-Kayser, MD
Assistant Professor of Radiation Oncology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [S6]

Ikuo Hirano, MD*
Professor of Medicine, Division of Gastroenterology, Northwestern University Feinberg School of Medicine, Chicago, Illinois [47, 334]

Martin S. Hirsch, MD
Professor of Medicine, Harvard Medical School; Senior Physician, Massachusetts General Hospital, Boston, Massachusetts [200]

Dieter Hoelzer, PhD, MD
Emeritus Director of Internal Medicine, University of Frankfurt, Frankfurt, Germany [111]

A. Victor Hoffbrand, DM
Emeritus Professor of Haematology, University College, London, United Kingdom [104]

L. John Hoffer, MD, PhD
Professor, Faculty of Medicine, McGill University; Senior Physician, Divisions of Internal Medicine and Endocrinology, Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada [346]

Charles W. Hoge, MD
Senior Scientist, Center for Psychiatry and Neuroscience, Walter Reed Army Institute of Research, Silver Spring, Maryland [S8]

Steven M. Holland, MD
Bethesda, Maryland [67, 185]

---

*Deceased

David Hong, MD
Assistant Professor of Medicine, Harvard Medical School; Division of Allergy and Immunology, Brigham and Women's Hospital, Boston, Massachusetts [364]

Jay H. Hoofnagle, MD
Director, Liver Diseases Research Branch, Division of Digestive Diseases and Nutrition, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland [347]

David C. Hooper, MD
Professor of Medicine, Harvard Medical School; Division of Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts [149, 150]

Paul J. Hoover, MD, PhD
Director of Translational Research in Lupus Nephritis; Instructor in Medicine, Harvard Medical School; Associate Physician, Division of Rheumatology, Inflammation and Immunity, Brigham and Women's Hospital, Boston, Massachusetts [368]

Robert J. Hopkin, MD
Professor of Clinical Pediatrics, Cincinnati Children's Hospital Medical Center Division of Human Genetics, Cincinnati, Ohio [429]

Jonathan C. Horton, MD, PhD
William F. Hoyt Professor of Neuro-ophthalmology, Professor of Ophthalmology, Neurology and Physiology, University of California, San Francisco School of Medicine, San Francisco, California [34, V3]

Howard Hu, MD, MPH, ScD
Professor and Flora L. Thornton Chair, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, California [469]

John C. Hu, MD, PhD
Assistant Professor of Medicine, Division of Infectious Diseases, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York; Staff Physician, Department of Medicine, Division of Infectious Diseases, Veterans Administration Western New York Healthcare Center, Buffalo, New York [180]

Deborah T. Hung, MD, PhD
Professor of Genetics, Harvard Medical School, Brigham and Women's Hospital and Massachusetts General Hospital, Boston, Massachusetts; Co-Director, Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts [126]

Gary M. Hunninghake, MD, MPH
Associate Professor of Medicine, Harvard Medical School; Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts [304]

Stephen C. Hunt, MD, MPH
National Director, VA Post-Deployment Integrated Care Initiative; Clinical Professor of Medicine, Department of Medicine, Division of General Internal Medicine, Occupational and Environmental Medicine Program, University of Washington, Seattle, Washington [S8]

Christina L. Hutson, PhD, MS
Chief, Poxvirus and Rabies Branch, Division of High Consequence Pathogens and Pathology, National Center for Emerging Zoonotic Infectious Diseases, US Centers for Disease Control and Prevention, Atlanta, Georgia [201]

Ashraf S. Ibrahim, PhD
Professor of Medicine, David Geffen School of Medicine at the University of California, Los Angeles, The Lindquist Institute for Biomedical Innovation, Torrance, California [222, 224]

David H. Ingbar, MD
Professor, Medicine, Pediatrics and Integrative Biology and Physiology, Pulmonary, Allergy, Critical Care and Sleep Division; Assistant Dean for Faculty Development and Mentoring, CTSI Associate Director; Director, Education, Career Development and Training (CTSI-Ed), Executive Director, Center for Lung Science and Health; University of Minnesota; Minnesota [316]

Khalid Ismail, MD
Instructor of Medicine, Harvard Medical School; Medical Director, Non-invasive Ventilation Clinic, Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [305, 307]

Elliot Israel, MD, MSc (Hon)
Professor of Medicine, Harvard Medical School; Gloria M. and Anthony C. Simboli Distinguished Chair in Asthma Research; Director, MGB Asthma Center; Director of Clinical Research, Pulmonary and Critical Care Division, Allergy and Immunology Division, Brigham and Women's Hospital, Boston, Massachusetts [298]

Elias Jabbour, MD
Professor, Section Chief, Acute Lymphocytic Leukemia, Department of Leukemia, Division of Cancer Medicine, MD Anderson Cancer Center, Houston, Texas [110]

Alan C. Jackson, MD
Adjunct Clinical Professor of Clinical Neurosciences, Section of Neurology, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; Neurology Locums, Thompson General Hospital, Thompson, Manitoba, Canada; Flin Flon General Hospital, Flin Flon, Manitoba, Canada; Lake of the Woods District Hospital, Kenora, Ontario, Canada [214]

Yves Jackson, MD, PhD
Assistant Professor, Division of Primary Care Medicine, Geneva University Hospitals, Geneva, Switzerland [234]

Danny O. Jacobs, MD, MPH, FACS
President, Oregon Health and Science University, Portland, Oregon [16, 341, 342]

Caron A. Jacobson, MD
Associate Professor of Medicine, Harvard Medical School; Dana-Farber Cancer Institute, Boston, Massachusetts [113, 114]

Rajesh K. Jain, MD
Assistant Professor, Department of Medicine, Section of Endocrinology, University of Chicago, Chicago, Illinois [424]

J. Larry Jameson, MD, PhD
Robert G. Dunlop Professor of Medicine; President, University of Pennsylvania, Philadelphia, Pennsylvania [50, 98, 388, 389, 390, 391, 392, 394, 395, 396, 397, 402, 479, 480, A15]

Safwan S. Jaradeh, MD
Professor of Neurology and Neurological Sciences; Director, Autonomic Disorders Program, Stanford University School of Medicine, Palo Alto, California [451]

Jordan B. Jastrab, MD, PhD
Instructor in Medicine, Harvard Medical School; Associate Physician in Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts [125]

Cecilia Jiang, MD
Assistant Professor of Radiation Oncology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [S6]

Fan Jiang, MD
Division of Neurosurgery, Krembil Neuroscience Centre, Toronto Western Hospital, University Health Network, Toronto, Ontario, Canada [19]

Anders F. Johansson, MD, PhD
Professor in Infectious Diseases, Umeå University, Umeå Sweden; Head of Infection Prevention and Control, Region Västerbotten, Sweden [175]

Savio John, MD
Chief of Gastroenterology, State University of New York Upstate Medical University, Syracuse, New York [52]

Stuart Johnson, MD
Professor of Medicine Emeritus, Loyola University Chicago Stritch School of Medicine, Maywood, Illinois; Staff Physician, Edward Hines Jr. VA Hospital, Hines, Illinois [139]

S. Andrew Josephson, MD
Professor and Chair, Department of Neurology, University of California, San Francisco, San Francisco, California [29, 30, 318, 433, V4]

Harald Jüppner, MD
Professor of Pediatrics, Endocrine Unit and Pediatric Nephrology Unit, Harvard Medical School; Massachusetts General Hospital, Boston, Massachusetts [422]

Joseph Kado, MBBS, DCH, MMed, PhD
Senior Research Fellow, Rheumatic Heart Disease (Pacific Islands), The Kids Research Institute Australia, Perth Children's Hospital, Nedlands, Western Australia [371]

Peter J. Kahrilas, MD
Gilbert H. Marquardt Professor of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois [47, 334]

Stephen G. Kaler, MD, MPH
CAPT, US Public Health Service (Ret), Professor of Pediatrics, Columbia University Vagelos College of Physicians and Surgeons NewYork-Presbyterian/Columbia University Irving Medical Center Physicians and Surgeons Building, New York, New York [427]

Kenneth Kalunian, MD
Wolfe Family Director, Lupus Center of Excellence; Professor of Medicine, University of California, San Diego, La Jolla, California [368]

Mini Kamboj, MD, FIDSA, FSHEA
Professor of Medicine, Weill Cornell Medical College; Chief Medical Epidemiologist, Infection Control; Attending Physician, Infectious Disease Service, Memorial Sloan Kettering Cancer Center, New York, New York [147]

Gail A. Kang, MD
Private Practice, Berkeley, California [V1]

Hagop Kantarjian, MD
Chairman, Leukemia Department; Professor of Leukemia, The University of Texas MD Anderson Cancer Center, Houston, Texas [110]

Mariana J. Kaplan, MD
NIH Distinguished Investigator; Chief, Systemic Autoimmunity Branch; Deputy Scientific Director, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland [367]

Hemanta K. Kar, MD, MAMS
Professor Emeritus Dermatology, Department of Dermatology, STD and Leprosy, Kalinga Institute of Medical Sciences, KIIT University, Bhubaneswar, Odisha, India [184]

Dennis L. Kasper, MD
William Ellery Channing Professor of Medicine and Professor of Immunology, Department of Immunology, Harvard Medical School; Division of Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts [124, 182, 484]

Daniel L. Kastner, MD, PhD
NIH Distinguished Investigator and Scientific Director Emeritus, Division of Intramural Research, National Human Genome Research Institute, Bethesda, Maryland [381]

Carol A. Kauffman, MD
Professor Emerita Internal Medicine, Infectious Diseases, University of Michigan Medical School, VA Medical Center, Ann Arbor, Michigan [225, 226]

Elaine T. Kaye, MD
Assistant Professor of Dermatology, Harvard Medical School, Boston Children's Hospital, Boston, Massachusetts [21, A1]

Kenneth M. Kaye, MD
Professor of Medicine, Harvard Medical School; Senior Physician, Division of Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts [21, A1]

John F. Keaney, Jr., MD
Professor of Medicine, Harvard Medical School; Chief, Division of Cardiovascular Medicine; Executive Director, Heart and Vascular Center, Brigham and Women's Hospital, Boston, Massachusetts [244]

David Kelsen, MD
Professor of Medicine, Weill Cornell Medical College; Edward S. Gordon Chair in Medical Oncology, Memorial Sloan Kettering Cancer Center, New York, New York [85]

Kiran K. Khush, MD, MAS
Professor of Medicine (Cardiovascular Medicine), Stanford University School of Medicine, Stanford, California [503]

Anthony A. Killeen, MD, MSc, PhD
Ellis S. Benson Professor and Vice-Chair for Clinical Affairs, Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota [S10]

Anthony S. Kim, MD, MAS
Professor of Neurology, Department of Neurology, University of California, San Francisco, San Francisco, California [438]

Kami Kim, MD
Andor Szentivanyi Professor of Medicine and Director, Division of Infectious Disease and International Medicine, University of South Florida; Medical Director, Infectious Disease, Tampa General Hospital, Tampa, Florida [235]

Priya S. Kishnani, MD
C.L. and Su Chen Professor of Pediatrics; Medical Director, YT and Alice Chen Pediatrics Genetics and Genomics Center; Professor of Molecular Genetics and Microbiology; Chief, Division of Medical Genetics, Duke University Medical Center, Durham, North Carolina [430]

Bruce S. Klein, MD
Gerard B. Odell Professor and Shirley S. Matchette Professor; Chief, Division of Pediatric Infectious Disease, Departments of Pediatrics, Medicine and Medical Microbiology and Immunology, University of Wisconsin, Madison, Madison, Wisconsin [220]

Christine Klein, MD
Professor of Neurology and Neurogenetics, Institute of Neurogenetics and Department of Neurology, University of Lübeck and University Hospital Schleswig-Holstein, Lübeck, Germany [446, 447]

David M. Knipe, PhD
Higgins Professor of Microbiology and Molecular Genetics; Head, Program in Virology, Department of Microbiology, Blavatnik Institute, Harvard Medical School, Boston, Massachusetts [195]

Nerissa U. Ko, MD, MAS
Professor and Vice Chair, Department of Neurology, University of California, San Francisco, San Francisco, California [440]

Rebecca L. Koch, PhD, RDN
Assistant Professor, Division of Medical Genetics, Duke University Medical Center, Durham, North Carolina [430]

Ajar Kochar, MD, MHS
Instructor in Medicine, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [249]

Isaac S. Kohane, MD, PhD
Marion V. Nelson Professor and Chair, Biomedical Informatics; Harvard Medical School; Faculty Member, Informatics Program, Boston Children's Hospital; Boston, Massachusetts [501]

Alex C. Kong, PhD, PharmD
Department of International Health, Health Systems Program, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland [211]

Barbara A. Konkle, MD
Professor of Medicine, Division of Hematology/Oncology Research Director, Washington Center for Bleeding Disorders, University of Washington, Seattle, Washington [69, 120]

Bruce A. Koplan, MD, MPH
Assistant Professor of Medicine, Harvard Medical School; Director, Electrophysiology Laboratory, Brigham and Women's Hospital, Boston, Massachusetts [250, 251, 252]

Peter Kopp, MD
Professor of Medicine/Médecin Chef, Division of Endocrinology, Diabetology and Metabolism, University of Lausanne, Lausanne, Switzerland; Adjunct Professor, Division of Endocrinology, Metabolism and Molecular Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois [479]

Walter J. Koroshetz, MD
Director, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland [144]

Thomas R. Kosten, MD
J. H. Waggoner Professor of Psychiatry, Pharmacology, Immunology, Neuroscience, Baylor College of Medicine, Houston, Texas [467]

Camille Nelson Kotton, MD, FIDSA, FAST
Clinical Director, Transplant and Immunocompromised Host Infectious Diseases, Infectious Diseases Division, Massachusetts General Hospital, Boston, Massachusetts [200]

Barnett S. Kramer, MD, MPH, FACP
Member, Lisa Schwartz Foundation for Truth in Medicine, Hanover, Vermont [75]

Joel Kramer, PsyD
John Douglas French Alzheimer's Foundation Endowed Professor of Neuropsychology in Neurology; Director of Neuropsychology, Memory and Aging Center, University of California, San Francisco, San Francisco, California [V2]

Arnold R. Kriegstein, MD, PhD
Professor of Neurology, University of California, San Francisco, San Francisco, California [435]

Somashekar G. Krishna, MD, MPH
Professor of Medicine, Director of Advanced Endoscopy, Division of Gastroenterology, Hepatology, and Nutrition, The Ohio State University Wexner Medical Center, Columbus, Ohio [358, 359]

Henry M. Kronenberg, MD
Professor of Medicine, Harvard Medical School, Endocrine Unit, Massachusetts General Hospital, Boston, Massachusetts [421]

Jens H. Kuhn, MD, PhD, PhD, MS
Principal Scientist and Director of Virology (contractor), NIH/NIAID/DCR/Integrated Research Facility at Fort Detrick, Frederick, Maryland [215, 216]

Prathit A. Kulkarni, MD
Assistant Chief of Medicine, Michael E. DeBakey Veterans Affairs Medical Center; Assistant Professor and Fellowship Program Director, Infectious Diseases Section, Department of Medicine, Baylor College of Medicine, Houston, Texas [140]

Matthew H. Kulke, MD
Zoltan Kohn Professor of Medicine, Chobanian and Avedisian School of Medicine; Chief, Section of Hematology and Medical Oncology, Boston University and Boston Medical Center, Boston, Massachusetts [89]

Sebastian G. Kurz MD, PhD, FCCP
Associate Professor of Medicine, Division of Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, Connecticut [186]

Robert F. Kushner, MD
Professor of Medicine and Medical Education, Northwestern University Feinberg School of Medicine, Chicago, Illinois [414]

Raymond Y. Kwong, MD, MPH
Professor of Medicine, Harvard Medical School; Director of Cardiac Magnetic Resonance Imaging, Cardiovascular Division, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [248, A9]

Loren Laine, MD
Professor of Medicine; Chief, Section of Digestive Diseases, Yale School of Medicine, New Haven, Connecticut; VA Connecticut Healthcare System, West Haven, Connecticut [51]

Neal K. Lakdawala, MD, MSc
Assistant Professor of Medicine, Harvard Medical School; Associate Physician; Program Director, Advanced Heart Failure Fellowship Program, Brigham and Women's Hospital, Boston, Massachusetts [266–270]

Anil K. Lalwani, MD
Associate Dean for Student Research, Columbia University Vagelos College of Physicians and Surgeons; Professor and Vice Chair for Research; Co-Director, Columbia Cochlear Implant Center, Columbia University Vagelos College of Physicians and Surgeons; Medical Director of Perioperative Services, New York Presbyterian–Columbia University Irving Medical Center, New York, New York [36]

Michael J. Landzberg, MD
Associate Professor of Medicine, Harvard Medical School; Boston Adult Congenital Heart Disease and Pulmonary Hypertension Program, Boston Children's Hospital, Brigham and Women's Hospital, Boston, Massachusetts [280]

H. Clifford Lane, MD
Clinical Director, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [208, S4]

Helene M. Langevin, MD
Director, National Center for Complementary and Integrative Health, National Institutes of Health, Bethesda, Maryland [495]

Carol A. Langford, MD, MHS
Harold C. Schott Endowed Chair in Rheumatic and Immunologic Diseases; Professor of Medicine, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University School of Medicine; Director, Center for Vasculitis Care and Research, Department of Rheumatic and Immunologic Diseases, Cleveland Clinic, Cleveland, Ohio [375, 386, 387, A14]

Regina C. LaRocque, MD, MPH
Associate Professor of Medicine, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts [138]

Leslie P. Lawley, MD
Associate Professor, Department of Dermatology, School of Medicine, Emory University, Atlanta, Georgia [60]

Thomas J. Lawley, MD
William Patterson Timmie Professor of Dermatology, Former Dean, Emory University School of Medicine, Atlanta, Georgia [59, 62, A5]

David G. Le Couteur, MD, PhD
Professor of Geriatric Medicine, University of Sydney; Senior Staff Geriatrician, Concord Hospital, Sydney, Australia [488]

Sancy A. Leachman, MD, PhD
Professor and Vice-Chair of Faculty Engagement, Department of Dermatology, University of Utah, Salt Lake City, Utah [81]

Stefi F. Lee, MD
Instructor of Medicine, Harvard Medical School; Associate Physician, Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts [306]

William M. Lee, MD, FACP, FAASLD, AGAF
Professor of Internal Medicine; Meredith Mosle Chair in Liver Diseases, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas [351]

Jacob E. Lemieux, MD, DPhil
Assistant Professor of Medicine and Biology, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts [191]

Jane A. Leopold, MD, FACC, FAHA, FSCAI
Associate Professor of Medicine, Harvard Medical School; Director, Women's Interventional Cardiology Health Initiative, Director, Cardiovascular Research Cores, Brigham and Women's Hospital; Boston, Massachusetts [249, A11]

Alyssa R. Letourneau, MD, MPH
Assistant Professor of Medicine, Harvard Medical School; Medical Director, Massachusetts General Hospital Antimicrobial Stewardship Program, Boston, Massachusetts [149]

Jessica Leung, MPH
Epidemiologist, Viral Vaccine Preventable Diseases Branch, Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [213]

Jonathan S. Leventhal, MD
Associate Professor of Dermatology, Yale University School of Medicine, New Haven, Connecticut [61]

Todd P. Levin, DO, MS-HQS
Jefferson Health, Philadelphia, Pennysylvania [412]

Bruce D. Levy, MD, MSc (Hon)
Hersey Professor of the Theory and Practice of Physic, Harvard Medical School; Co-Chair, Department of Medicine, Mass General Brigham [40, 295, 312]

Charles Chin Han Lew, PhD
Adjunct Assistant Professor, Faculty of Health and Social Sciences, Singapore Institute of Technology Senior Principal Dietitian, Dietetics and Nutrition Department, Ng Teng Fong General Hospital, Singapore, Singapore [345]

Julia B. Lewis, MD
Professor of Medicine, Division of Nephrology and Hypertension, Vanderbilt University Medical Center, Nashville, Tennessee [326]

Peter Libby, MD
Mallinckrodt Professor of Medicine, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [A10]

Mary Jane Lim-Fat, MD, MSc, FRCPC
Assistant Professor, Division of Neurology, Department of Medicine, University of Toronto; Neuro-Oncologist, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada [95]

Jin-Mann S. Lin, PhD
Division of High-Consequence Pathogens and Pathology, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [461]

Jeffrey A. Linder, MD, MPH, FACP
Michael A. Gertz Professor of Medicine, Chief, Division of General Internal Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois [37]

Michail S. Lionakis, MD, ScD
Chief, Laboratory of Clinical Immunology and Microbiology; Chief, Fungal Pathogenesis Section, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [217, 222, 225, 226]

Irene Litvan, MD, MSc, FAAN, FANA
Tasch Endowed Professor in Parkinson Disease Research; Director of the Parkinson and Other Movement Disorders Center, University of California, San Diego, La Jolla, California [445]

Eva S. Liu, MD
Assistant Professor of Medicine, Harvard Medical School, Division of Endocrinology, Diabetes, and Hypertension, Brigham and Women's Hospital, Boston, Massachusetts [421]

Kathleen D. Liu, MD, PhD, MAS
Professor, Division of Nephrology, Department of Medicine, Division of Critical Care Medicine, Department of Anesthesiology, University of California, San Francisco, San Francisco, California [323]

Peter A. Ljubenkov, MD
Associate Professor of Neurology, Memory and Aging Center, Department of Neurology, UCSF Weill Institute for Neurosciences, University of California, San Francisco, California [442]

Josep M. Llovet, MD, PhD
Liver Cancer Program, Division of Liver Diseases, Tisch Cancer Institute, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York; Liver Cancer Translational Research Laboratory, Program of Translational Research in Cancer, Liver Unit, IDIBAPS-Hospital Clínic, University of Barcelona; Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain [87]

Donald M. Lloyd-Jones, MD, ScM, FACC, FAHA
Eileen M. Foell Professor of Heart Research; Professor of Preventive Medicine, Medicine, and Pediatrics; Chair, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine; President, American Heart Association 2021–22, Chicago, Illinois [2]

Bernard Lo, MD
Professor of Medicine Emeritus and Director Emeritus of the Program in Medical Ethics, University of California, San Francisco, San Francisco, California; President Emeritus, The Greenwall Foundation, New York, New York [12]

George Loewenstein, PhD
Herbert A. Simon University Professor of Economics and Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania [494]

Dan L. Longo, MD
Professor of Medicine, Harvard Medical School; Senior Physician, Brigham and Women's Hospital; Deputy Editor, *New England Journal of Medicine*, Boston, Massachusetts [65, 66, 70, 73, 77, 78, 98, 100, 101, 106, 113–116, 207, A6]

Nicola Longo, MD, PhD
Professor and Vice Chair of Human Genetics, Allen and Charlotte Ginsburg Chair in Precision Genomic Medicine; Chief, Division of Clinical Genetics, Department of Human Genetics, University of California, Los Angeles, California [431, 432]

Lenny López, MD, MPH, MDiv
Professor of Medicine, University of California San Francisco; San Francisco VA Medical Center, San Francisco, California [11]

Charles L. Loprinzi, MD
Mayo Clinic, Rochester, Minnesota [74]

Joseph Loscalzo, MD, PhD
Hersey Distinguished Professor of the Theory and Practice of Medicine, Samuel A. Levine Professor of Medicine, Harvard Medical School; Physician-in-Chief Emeritus, Soma Weiss, MD, Honorable Endowed Chair in Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [42–46, 122, 243, 244, 246, 266–270, 272–279, 281, 284, 286, 291–294, 499, 505, S13]

Christine M. Lovly, MD, PhD
Associate Professor of Medicine, Division of Hematology and Oncology; Ingram Associate Professor of Cancer Research, Vanderbilt University Medical Center and Vanderbilt Ingram Cancer Center, Nashville, Tennessee [83]

Daniel H. Lowenstein, MD
Distinguished Professor of Neurology Education, University of California, San Francisco, San Francisco, California [433, V6]

Andrew M. Lowy, MD
Professor of Surgery; Chief, Division of Surgical Oncology; Adjunct Professor of Pharmacology; Levine Family Chancellor's Endowed Chair in Surgical Oncology, University of California, San Diego; Associate Clinical Director, Surgery, Moore's Cancer Center, La Jolla, California [88]

Franklin D. Lowy, MD
Clyde '56 and Helen Wu Professor Emeritus of Medicine and Professor Emeritus of Pathology and Cell Biology (in Epidemiology), Columbia University College of Physicians and Surgeons, New York, New York [152]

Sheila A. Lukehart, PhD
Emeritus Professor of Medicine, Division of Allergy and Infectious Diseases and Global Health, University of Washington School of Medicine, Seattle, Washington [187, 188]

Carolina Lúquez, PhD
Team Lead, National Botulism Laboratory, Enteric Diseases Laboratory Branch, Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [158]

Lucio Luzzatto, MD, FRCP, FRCPath
Honorary Professor of Hematology, University of Florence, Firenze, Italy [105]

Calum A. MacRae, MD, PhD
Professor of Medicine, Harvard Medical School; Vice Chair for Scientific Innovation, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [244]

Lawrence C. Madoff, MD
Professor of Medicine, University of Massachusetts Chan Medical School, Worcester, Massachusetts; Medical Director, Bureau of Infectious Disease and Laboratory Sciences, Massachusetts Department of Public Health, Hinton State Laboratory Institute, Jamaica Plain, Massachusetts [135, 146]

Mahmoud Malas, MD, MHS, RPVI, FACS
Professor in Residence; Vice Chair of Surgery for Clinical Research; Chief Division Vascular and Endovascular Surgery, University of California, San Diego, Health System, La Jolla, California [340]

Fransiska Malfait, MD, PhD
Associate Professor, Center for Medical Genetics, Ghent University Hospital and Department for Biomolecular Medicine, Ghent University, Ghent, Belgium [425]

Hari R. Mallidi, MD
Professor of Clinical Surgery, University of Miami; Chief of Cardiac Surgery, Jackson Health System, Miami, Florida [309]

Susan J. Mandel, MD, MPH
Sylvan H. Eisman Professor of Medicine; Chief, Division of Endocrinology, Diabetes and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [394–397]

Brian F. Mandell, MD, PhD
Professor and Chairman of Medicine, Cleveland Clinic Lerner College of Medicine, Department of Rheumatic and Immunologic Disease, Cleveland Clinic, Cleveland, Ohio [386]

Lionel A. Mandell, MD FRCPC, FRCP (LOND), FAMMI
Professor Emeritus of Medicine, McMaster University School of Medicine Hamilton, Ontario, Canada [131]

Geoffrey T. Manley, MD, PhD
Professor and Vice Chairman of Neurological Surgery, University of California, San Francisco; Chief of Neurosurgery, Zuckerberg San Francisco General Hospital and Trauma Center; Co-Director, Brain and Spinal Injury Center, University of California, San Francisco, San Francisco, California [454]

Nathaniel Mann, MD
Assistant Clinical Professor, University of South Carolina School of Medicine, Greenville; Director of Wilderness Medicine, Prisma Health-Upstate, Greenville, South Carolina [471]

Michael Mannstadt, MD
Associate Professor in Medicine, Harvard Medical School; Chief, Endocrine Unit, Massachusetts General Hospital, Boston, Massachusetts [422]

Arjun K. Manrai, PhD
Assistant Professor, Biomedical Informatics, Harvard Medical School; Senior Deputy Editor, NEJM AI, Boston, Massachusetts [501]

JoAnn E. Manson, MD, DrPH
Professor of Medicine and the Michael and Lee Bell Professor of Women's Health, Harvard Medical School; Chief, Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts [407]

Joan C. Marini, MD, PhD
Scientist Emeritus, Section on Heritable Disorders of Bone and Extracellular Matrix, National Institute of Child Health and Human Development (NICHD), National Institutes of Health, Bethesda, Maryland [425]

Daniel B. Mark, MD, MPH
Professor of Medicine, Duke University Medical Center; Founding Director, Outcomes Research, Duke Clinical Research Institute, Durham, North Carolina [4]

Alexander G. Marneros, MD, PhD
Associate Professor, Department of Dermatology, Harvard Medical School; Cutaneous Biology Research Center, Massachusetts General Hospital, Boston, Massachusetts [64]

Bradley A. Maron, MD
Senior Associate Dean for Precision Medicine and Professor of Medicine, University of Maryland School of Medicine, Baltimore, Maryland [294]

Jeanne M. Marrazzo, MD, MPH
Director, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [141]

Gary J. Martin, MD
Raymond J. Langenbach, MD Professor of Medicine; Senior Vice Chairman, Department of Medicine, Northwestern University Medical School, Chicago, Illinois [6]

Anthony F. Massaro, MD
Instructor, Harvard Medical School; Director, Medical Intensive Care Unit, Division of Pulmonary and Critical Care, Brigham and Women's Hospital, Boston, Massachusetts [311, 314]

Nina B. Masters, PhD, MPH
Epidemiologist, Mumps Program Lead, Division of Viral Diseases, National Center of Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [213]

Henry Masur, MD
Chief, Critical Care Medicine Department, National Institutes of Health Clinical Center, Bethesda, Maryland [227]

Marcela V. Maus, MD, PhD
Director, Cell Therapy Program; Paula O'Keefe Endowed Chair, Massachusetts General Hospital Cancer Center; Professor of Medicine, Harvard Medical School; Attending Physician, Hematopoietic Cell Transplant and Cell Therapy Program, Massachusetts General Hospital; Associate Member, Broad Institute of MIT and Harvard, Charlestown, Massachusetts [483]

Clio P. Mavragani, MD
Professor of Physiology/Clinical Physiology, Attending Rheumatologist, Departments of Physiology and Pathophysiology, National and Kapodistrian University of Athens, Athens, Greece [373]

Jared R. Mayers, MD, PhD
Assistant Professor, Fred Hutchinson Cancer Center and University of Washington, Seattle, Washington [502]

Paul H. Mayo, MD, FCCP, FASE
Professor of Clinical Medicine, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell; Academic Director of Critical Care, Long Island Jewish Medical Center and North Shore University Hospital, Hempstead, New York [493]

Alexander J. McAdam, MD, PhD
Associate Professor of Pathology, Harvard Medical School; Chief, Department of Laboratory Medicine,Boston Children's Hospital, Boston, Massachusetts [S11]

Michael McCrea, PhD, ABPP
Professor and Eminent Scholar; Vice Chair of Research; Co-Director, Center for Neurotrauma Research (CNTR), Department of Neurosurgery, Medical College of Wisconsin, Milwaukee, Wisconsin [454]

Steven McGee, MD
Professor Emeritus, Medicine, University of Washington School of Medicine, Seattle, Washington [8]

Kathleen M. McKibbin, MD
Staff Physician, Northwestern University Health Services, Evanston, Illinois [2]

Kevin T. McVary, MD, FACS
Director of the Center for Male Health; Professor of Urology, Department of Urology, Stritch School of Medicine, Loyola University Medical Center, Maywood, Illinois [409]

Mandeep R. Mehra, MBBS, MSc, FRCP (London)
Professor of Medicine, Harvard Medical School; The William Harvey Distinguished Chair in Advanced Cardiovascular Medicine; Executive Director, Center for Advanced Heart Disease, Brigham and Women's Hospital, Boston, Massachusetts [264, 265, 271]

Sanjay R. Mehta, MD, DTM&H, D(ABMM)
Professor of Medicine and Pathology, University of California, San Diego School of Medicine, San Diego, California [S12]

Michael T. Melia, MD
Associate Professor of Medicine, Division of Infectious Diseases, Johns Hopkins University School of Medicine, Baltimore, Maryland [133]

Shlomo Melmed, MBChB, MACP, FRCP
Executive Vice President and Dean of the Medical Faculty; Distinguished Professor of Medicine, Cedars-Sinai Medical Center, Los Angeles, California [390, 391, 392]

Robert O. Messing, MD
Professor and Chair of Neuroscience; Professor of Neurology; Director, Waggoner Center for Alcohol and Addiction Research, University of Texas at Austin, Austin, Texas [462]

M.-Marsel Mesulam, MD
Ruth Dunbar Davee Professor of Neuroscience and Neurology, Mesulam Center for Cognitive Neurology and Alzheimer's Disease, Northwestern University Feinberg School of Medicine, Chicago, Illinois [32]

Sarah Meyer, MD, MPH
Chief Medical Officer, Immunization Services Division, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [129]

Robert G. Micheletti, MD
Associate Professor of Dermatology and Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [63]

Aaron W. Michels, MD
Professor of Pediatrics, Medicine, and Immunology, Barbara Davis Center for Childhood Diabetes, University of Colorado School of Medicine, Aurora, Colorado [401]

Bruce L. Miller, MD
A. W. and Mary Margaret Clausen Distinguished Professor of Neurology, Memory and Aging Center, Global Brain Health Institute, University of California, San Francisco School of Medicine, San Francisco, California [29, 31, 442, 443, 445, V2]

Samuel I. Miller, MD
Professor of Medicine, Microbiology and Genome Sciences, University of Washington, Seattle, Washington [171]

William R. Miller, MD
Assistant Professor of Medicine, Division of Infectious Diseases, Houston Methodist Hospital, Center for Infectious Diseases Research, Houston Methodist Research Institute and Weill Cornell Medical College, Houston, Texas [154]

Faisal Syed Minhaj, PharmD, MPH
Epidemiologist, Poxvirus and Rabies Branch, Division of High Consequence Pathogens and Pathology, National Center for Emerging Zoonotic Infectious Diseases, US Centers for Disease Control and Prevention, Atlanta, Georgia [201]

Jyoti Mishra, PhD
Associate Professor, Department of Psychiatry, University of California, San Diego, La Jolla, California [500]

Hana Mitchell, MD, MSc
Clinical Assistant Professor, Department of Pediatrics, Vaccine Evaluation Center, BC Children's Hospital Research Institute, University of British Columbia, Vancouver, British Columbia, Canada [3]

Simon J. Mitchell, MBChB, PhD, FUHM, FANZCA
Professor of Anaesthesiology, University of Auckland; Anesthesiologist, Department of Anesthesia, Auckland City Hospital, Auckland, New Zealand [476]

Richard E. Moon, MD, FUHM, FRCPC, FACP
Professor of Anesthesiology and Medicine, Medical Director, Center for Hyperbaric Medicine and Environmental Physiology, Duke University Medical Center, Durham, North Carolina [476]

Thomas A. Moore, MD, FACP, FIDSA
Clinical Professor of Medicine, University of Kansas School of Medicine, Wichita, Kansas [229]

David M. Morens, MD
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [486]

Richard I. Morimoto, PhD
Bill and Gayle Cook Professor of Biology; Director, Rice Institute for Biomedical Research Department of Molecular Biosciences, Northwestern University, Evanston, Illinois [504]

Alison Morris, MD, MS
Chief, Pulmonary, Allergy, Critical Care, and Sleep Medicine; Professor of Medicine; UPMC Chair of Translational Pulmonary and Critical Care Medicine; Director, University of Pittsburgh Center for Medicine and the Microbiome, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania [227]

David A. Morrow, MD, MPH
Professor of Medicine, Harvard Medical School; Director, Samuel A. Levine Cardiac Intensive Care Unit, Cardiovascular Division, Brigham and Women's Hospital, Boston, Massachusetts [15, 286]

William J. Moss, MD, MPH
Professor, Departments of Epidemiology, International Health, and Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland [211]

Robert J. Motzer, MD
Jack and Dorothy Byrne Chair in Clinical Oncology, Associate Attending Physician, Genitourinary Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York [90]

David B. Mount, MD, FRCPC
Assistant Professor of Medicine, Harvard Medical School; Clinical Chief, Renal Division, Brigham and Women's Hospital; VA Boston Healthcare System, Boston, Massachusetts [55, 56, S1]

Haralampos M. Moutsopoulos, MD, FACP, FRCP(hc), Master ACR
Professor, Chair Medical Sciences-Immunology, Academy of Athens, Athens, Greece [369, 373]

Catharina M. Mulders-Manders, MD, PhD
Department of Internal Medicine, Radboud University Medical Center, Nijmegen, The Netherlands [22]

Nikhil C. Munshi, MD
Professor of Medicine, Harvard Medical School; Boston VA Healthcare System; Kraft Family Chair and Director of Basic and Correlative Sciences; Associate Director, Jerome Lipper Myeloma Center, Dana-Farber Cancer Institute, Boston, Massachusetts [116]

Naoka Murakami, MD, PhD
Assistant Professor of Medicine, Harvard Medical School; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts [325]

John R. Murphy, PhD
Professor of Medicine, Division of Infectious Diseases, Johns Hopkins School of Medicine, Baltimore, Maryland [155]

Timothy F. Murphy, MD
SUNY Distinguished Professor; Director, UB Clinical and Translational Science Institute; Senior Associate Dean for Clinical and Translational Research, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, The State University of New York, Buffalo, New York [162]

Barbara E. Murray, MD*
J. Ralph Meadows Professor of Medicine, Division of Infectious Diseases; Professor, Microbiology and Molecular Genetics, University of Texas Medical School, Houston, Texas [154]

Joseph A. Murray, MD
Professor of Medicine, John and Shirley Berry Professor of Gastrointestinal Sciences, Mayo Clinic School of Medicine, Rochester, Minnesota [49]

Vivek H. Murthy, MD, MBA
19th and 21st Surgeon General of the United States, Washington, DC [9]

Mark B. Mycyk, MD
Professor, Department of Emergency Medicine, Rush Medical College; Professor, Department of Emergency Medicine, Feinberg School of Medicine, Northwestern University; Chair of Research, Department of Emergency Medicine, Cook County Health, Chicago, Illinois [470]

Avindra Nath, MD
Chief, Section of Infections of the Nervous System; Clinical Director, National Institute of Neurological Disorders and Stroke (NINDS), National Institutes of Health, Bethesda, Maryland [144]

Karen P. Neil, MD, MSPH
Medical Epidemiologist, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [156]

Eric G. Neilson, MD
Vice President for Medical Affairs, Lewis Landsberg Dean; Professor of Medicine and Cell and Molecular Biology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois [320, 326]

Tuhina Neogi, MD, PhD
Professor of Medicine and Epidemiology, Chair, Section of Rheumatology, Boston University Schools of Medicine and Public Health, Boston, Massachusetts [383]

Eric J. Nestler, MD, PhD
Nash Family Professor, Department of Neuroscience; Director, Friedman Brain Institute; Dean for Academic and Scientific Affairs, Ichan School of Medicine at Mount Sinai, New York, New York [462]

Hartmut P. H. Neumann, MD
Professor of Medicine, Unit for Preventive Medicine, Department of Nephrology and General Medicine, Albert-Ludwigs University of Freiburg, Freiburg, Germany [399]

Kathleen M. Neuzil, MD, MPH
Director, Fogarty International Center, Associate Director for International Health, National Institutes of Health, Bethesda, Maryland [128, 206]

Jonathan Newmark, MD, MM
Colonel (Retired), Medical Corps, US Army; Adjunct Professor, Neurology, F. Edward Hebert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland; Clinical Assistant Professor, Neurology, School of Medicine and Health Sciences, George Washington University, Washington, DC; Department of Neurology, Washington DC Veterans' Affairs Medical Center, Washington, DC; Senior Medical Advisor, Office of Biodefense Research and Surety, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland [S5]

Max L. Nibert, MD, PhD
Professor of Microbiology; Program in Virology, Department of Microbiology, Blavatnik Institute, Harvard Medical School, Boston, Massachusetts [195]

J. Curtis Nickel, MD, FRCS(C)
Professor Emeritus, Department of Urology, Queen's University, Kingston, Ontario, Canada [54]

Michael S. Niederman, MD
Lauder Family Professor of Medicine, Division of Pulmonary and Critical Care Medicine, New York Presbyterian/Weill Cornell Medical Center, Weill Cornell Medical College, New York, New York [131]

Kevin D. Niswender, MD, PhD
Associate Professor of Medicine, Molecular Physiology and Biophysics; Director, Clinical Research Center, Vanderbilt University Medical Center, Nashville, Tennessee [415, 416]

Scott A. Norton, MD, MPH, MSc
Professor of Dermatology and Pediatrics, George Washington University School of Medicine and Health Sciences; Chief of Dermatology, Division of Dermatology, Children's National Health System, Washington, DC [472]

Emily Nosova, MD
Assistant Professor of Medicine, Division of Endocrinology, Diabetes and Bone Disease, Icahn School of Medicine and Mount Sinai Health System, New York, New York [410]

Thomas B. Nutman, MD
Head, Helminth Immunology Section; Head, Clinical Parasitology Section; Chief, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [239, 240]

Katherine L. O'Brien, MD, MPH
Director, IVB, World Health Organization, Geneva, Switzerland [151]

Max R. O'Donnell, MD, MPH
Associate Professor of Medicine and Epidemiology, Division of Pulmonary, Allergy, and Critical Care Medicine and Department of Epidemiology, Columbia University Irving Medical Center, New York, New York [186]

Nigel O'Farrell, MD, FRCP
Pasteur Suite Ealing Hospital, London, United Kingdom [178]

Patrick T. O'Gara, MD
Professor of Medicine, Harvard Medical School; Watkins Family Distinguished Chair in Cardiology, Brigham and Women's Hospital, Boston, Massachusetts [44, 246, 272–279]

C. Warren Olanow, MD, FRCPC, FRCP (Hon)[*]
Professor and Chairman Emeritus, Department of Neurology; Professor Emeritus, Department of Neuroscience, Mount Sinai School of Medicine, New York, New York; CEO, Clintrex, LLC [446, 447]

William M. Oldham, MD, PhD
Assistant Professor, Harvard Medical School; Associate Physician, Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [46, 296]

Stephen O'Rahilly, MD, FRCP, FMedSci, FRS
Professor of Clinical Biochemistry and Medicine; Director, Medical Research Council Metabolic Diseases Unit; Co-Director, Wellcome-MRC Institute of Metabolic Science; Head of Department of Clinical Biochemistry, University of Cambridge; Scientific Director, Cambridge NIHR Biomedical Research Centre; Hon Consultant Physician, Addenbrooke's Hospital, Cambridge, United Kingdom [413]

Eileen M. O'Reilly, MD, FASCO
Professor of Medicine, Weill Cornell Medicine; Winthrop Rockefeller Endowed Chair in Medical Oncology; Section Head, Hepatopancreaticobiliary and Neuroendocrine Cancers; Chair, Human Research Protection Program and IRB; Co-Director Medical, David M. Rubenstein Center for Pancreatic Cancer; Attending Physician, Member, Memorial Sloan Kettering Cancer Center, New York, New York [88]

Alan C. Ou, MD, MPH
Medical Epidemiologist, Centers for Disease Control and Prevention, Atlanta, Georgia [212]

Joseph G. Ouslander, MD
Charles E. Schmidt College of Medicine, Florida Atlantic University, Boca Raton, Florida [489]

Chung Owyang, MD
Professor of Medicine and Physiology/Internal Medicine (Retired), University of Michigan Health System, Ann Arbor, Michigan [332, 338]

Tara N. Palmore, MD
Senior Medical Advisor, Office of the Director, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [147]

Arvind K. Pandey, MD
Instructor, Harvard Medical School; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts [S13]

Gerald Pankratz, MD
Clinical Associate Professor, Department of Medicine, Division of Geriatrics and Gerontology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin [28]

Umesh D. Parashar, MBBS, MPH
Chief, Viral Gastroenteritis Branch, Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [209]

Shreyaskumar R. Patel, MD
Robert R. Herring Distinguished Professor of Medicine; Center Medical Director, Sarcoma Center, The University of Texas MD Anderson Cancer Center, Houston, Texas [96]

Jonathan A. Patz, MD, MPH
Vilas Distinguished Achievement Professor; John P. Holton Chair of Health and the Environment, Nelson Institute and the Department of Population Health Sciences, University of Wisconsin-Madison, Madison, Wisconsin [474]

Gustav Paumgartner, MD[*]
Professor Emeritus of Medicine, University of Munich, Munich, Germany [357]

Georgina Peacock, MD, MPH
Director, Immunization Services Division, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [129]

---

[*] Deceased

Joseph Pechacek, MD, MS
Staff Clinician, Fungal Pathogenesis Section, Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [225, 226]

David A. Pegues, MD
Professor of Medicine, Division of Infectious Diseases, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania [171]

Ludmila M. Perelygina, PhD
Research Microbiologist, Centers for Disease Control and Prevention, Atlanta, Georgia [212]

Steven A. Pergam, MD, MPH
Professor, Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center; Professor, Department of Medicine, Division of Allery & Infectious Diseases, University of Washington; Medical Director, Infection Prevention, Fred Hutchinson Cancer Center, Seattle, Washington [164]

Elizabeth J. Phillips, MD
Professor of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee [63]

Karran A. Phillips, MD, MSc
Deputy Director, Center for Substance Abuse Treatment, Substance Abuse and Mental Health Service Administration, U.S. Department of Health and Human Services, Rockville, Maryland [468]

Alexandra R. Pipilas, MD
Assistant Professor of Medicine, Boston University School of Medicine; Cardiovascular Medicine, Boston Medical Center, Boston, Massachusetts [283]

Martin R. Pollak, MD
George C. Reisman Professor of Medicine, Harvard Medical School; Beth Israel Deaconess Medical Center, Boston, Massachusetts [327]

Andrew J Pollard, BSc, MBBS, DIC, MA, PhD, FRCPCH, FMedSci
Ashall Professor of Infection and Immunity, Director of the Oxford Vaccine Group, University of Oxford, and Consultant in Paediatric Infectious Disease, Oxford Children's Hospital, Oxford, United Kingdom [160]

Nongnooch Poowanawittayakom, MD, MPH
Assistant Professor, Division of Infectious Diseases, Allergy, and Immunology, St. Louis University, St. Louis, Missouri [135, 146]

Alvin C. Powers, MD
Joe C. Davis Chair in Biomedical Science; Professor of Medicine, Molecular Physiology and Biophysics; Director, Vanderbilt Diabetes Center; Chief, Division of Diabetes, Endocrinology, and Metabolism, Vanderbilt University Medical Center, Nashville, Tennessee [415–417]

Daniel S. Pratt, MD
Assistant Professor of Medicine, Harvard Medical School; Director, Autoimmune and Cholestatic Liver Center, Massachusetts General Hospital, Boston, Massachusetts [52, 348, 357]

Michael B. Prentice, MBChB, PhD, FRCP (UK), FRCPath, FFPRCPI, FRCPI
Emeritus Professor of Medical Microbiology, School of Medicine, University College Cork, Cork, Ireland [176]

Alessandro Prete, MD, PhD
Clinical Associate Professor of Endocrinology, Department of Metabolism and Systems Science School of Medical Sciences, College of Medicine and Health, University of Birmingham, Birmingham, United Kingdom [398]

Stanley B. Prusiner, MD
Director, Institute for Neurodegenerative Diseases; Professor, Department of Neurology, UCSF Weill Institute for Neurosciences, University of California, San Francisco; Professor, Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California [435, 449]

Thomas C. Quinn, MD, MSc
Professor of Medicine and Pathology, Johns Hopkins University School of Medicine; Director, Johns Hopkins Center for Global Health, Baltimore, Maryland [194]

Gil D. Rabinovici, MD
Edward and Pearl Fein Distinguished Professor, Memory and Aging Center, Department of Neurology, Department of Radiology and Biomedical Imaging, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, California [31, 442, V2]

Daniel J. Rader, MD
Seymour Gray Professor of Molecular Medicine; Chair, Department of Genetics; Chief, Division of Translational Medicine and Human Genetics, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania [419]

Kanwal Raghav, MBBS, MD
Associate Professor, GI Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas [97]

Satish R. Raj, MD, MSCI
Professor of Cardiac Sciences, Cumming School of Medicine, Medical Director, Calgary Autonomic Investigation and Management Clinic University of Calgary, Calgary, Alberta, Canada [23]

Sanjay Ram, MBBS
Professor of Medicine, Division of Infectious Diseases and Immunology, University of Massachusetts Chan Medical School, Worcester, Massachusetts [161]

Prashanth Ramachandran, MBBS, BMedSci, PhD, FRACP
Laboratory Head, Department of Infectious Diseases, The Peter Doherty Institute for Infection and Immunity, Melbourne Medical School; Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Victoria, Australia [S2]

Reuben Ramphal, MD
Courtesy Professor, Department of Medicine, University of Florida College of Medicine, Gainesville, Florida [170]

Kathryn Moynihan Ramsey, PhD
Research Assistant Professor, Division of Endocrinology, Metabolism and Molecular Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois [498]

Vikram R. Rao, MD, PhD
Distinguished Professor in Neurology; Associate Professor of Clinical Neurology; Chief, Epilepsy Division, Department of Neurology, University of California, San Francisco, San Francisco, California [436]

Kumanan Rasanathan, MBChB, MPH, FAFPHM
Executive Director, Alliance for Health Policy and Systems Research, World Health Organization, Geneva, Switzerland [487]

James P. Rathmell, MD
Leroy D. Vandam Professor of Anaesthesia, Harvard Medical School; Chair, Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Boston, Massachusetts [14]

Mario C. B. Raviglione, MD, FERS, Hon RSP (RF)
Full Professor of Global Health; Co-Director, Centre for Multidisciplinary Research in Health Science (MACH), University of Milan, Milan, Italy [183]

Divya Reddy, MBBS, MPH
Associate Professor of Medicine; Program Director, Pulmonary and Critical Care Fellowship; Medical Director, Bronchiectasis and Nontuberculous Mycobacterial (NTM) Disease Program, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, New York [186]

Susan Redline, MD, MPH
Professor of Epidemiology, Peter C. Farrell Professor of Sleep Medicine, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [308]

Sharon L. Reed, MD, MScCTM, D(ABMM)
Professor of Pathology and Medicine, San Diego School of Medicine; Director, Microbiology Laboratory; Microbiology Laboratory, University of California, La Jolla, California [228, 230, S12]

Susan E. Reef, MD
Medical Epidemiologist, Centers for Disease Control and Prevention, Atlanta, Georgia [212]

Michael F. Regner, MD, MS
Assistant Professor, Department of Radiology, Neuroradiology Section, Oregon Health and Science University, Portland, Oregon [A16]

Victor I. Reus, MD
Distinguished Professor Emeritus, Department of Psychiatry and Behavioral Sciences, University of California, San Francisco School of Medicine; UCSF Weill Institute for Neurosciences, San Francisco, California [463]

Bernardo Reyes, MD
Dr. Kiran C. Patel College of Allopathic Medicine, Nova Southern University, Fort Lauderdale, Florida [489]

Joseph J. Rhatigan, MD
Chief, Division of Global Health Equity, Brigham and Women's Hospital; Associate Professor, Harvard Medical School and Harvard T.H. Chan School of Public Health, Boston, Massachusetts [485]

Peter A. Rice, MD
Professor Emeritus (Medicine), Division of Infectious Diseases and Immunology, University of Massachusetts Chan Medical School, Worcester, Massachusetts [161]

Eugene Richardson, MD, PhD
Assistant Professor, Global Health and Social Medicine, Harvard Medical School, Boston, Massachusetts [473]

Jan H. Richardus, MD, PhD
Professor Emeritus Infectious Diseases and Public Health, Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands [184]

Michael R. Rickels, MD, MS
Willard and Rhoda Ware Professor in Diabetes and Metabolic Diseases; Medical Director, Pancreatic Islet Cell Transplant Program, Hospital of the University of Pennsylvania; Director, Translational Research Program, Institute for Diabetes, Obesity and Metabolism, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania [416, 417]

Kristian Riesbeck, MD, PhD
Senior consultant and professor in Clinical Bacteriology, Lund University, Malmö, Sweden [162]

Elizabeth Robbins, MD
Clinical Professor, Pediatrics, Emeritus, University of California, San Francisco, San Francisco, California [S3]

Dan Roden, MD
Professor of Medicine, Pharmacology, and Biomedical Informatics; Senior Vice-President for Personalized Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee [71, 72]

Jorge E. Romero, MD
Associate Professor of Medicine, Harvard Medical School; Director, Arrhythmia and Electrophysiology Research; Associate Director, Ventricular Arrhythmia Program, Brigham and Women's Hospital, Boston, Massachusetts [257, 258]

Karen L. Roos, MD
John and Nancy Nelson Professor Emerita; Professor Emerita Neurological Surgery, Department of Neurology, Indiana University School of Medicine, Indianapolis, Indiana [142, 143, 145]

Allan H. Ropper, MD, FRCP, FACP
Professor of Neurology, Harvard Medical School; Associate Editor, *New England Journal of Medicine*, Boston, Massachusetts [30]

Rossana Rosa, MD, MSc
Infectious Diseases Specialist, Jackson Memorial Hospital, Miami, Florida [167]

Ivan O. Rosas, MD
Professor and Section Chief, Pulmonary, Critical Care, and Sleep Medicine, Baylor College of Medicine, Houston, Texas [304]

Mark Roschewski, MD
Clinical Director, Lymphoid Malignancies Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland [100]

Misha Rosenbach, MD
Professor of Dermatology and Medicine, Perelman School of MedicineUniversity of Pennsylvania, Philadelphia, Pennsylvania [63]

Roger N. Rosenberg, MD
Emeritus Zale Distinguished Chair and Professor of Neurology, Department of Neurology, University of Texas Southwestern Medical Center, Dallas, Texas [450]

Chantal P. Rovers, MD, PhD
Professor Outbreaks of Infectious Diseases, Department of Internal Medicine, Radboud University Medical Center, Nijmegen, The Netherlands [22, 192]

Steven Roy, MDCM, DiMM, DiMEM, DiWEM, DipROM, FAWM, FRGS, FRCPC
Adjunct Assistant Professor, Department of Critical Care Medicine, University of Calgary, Calgary, AB; Co-Program Director, Resident Physician Elective in Wilderness Medicine, McGill University, Montreal, Quebec, Canada [475]

Deborah C. Rubin, MD
William B. Kountz Professor of Medicine and of Developmental Biology; Associate Chair for Faculty Affairs, GI Division, and Director, Women's GI Committee, Washington University School of Medicine, St. Louis, Missouri [336]

Thomas A. Russo, MD, CM
SUNY Distinguished Professor of Medicine and Microbiology and Immunology; Chief, Division of Infectious Diseases, Jacobs School of Medicine and the Biomedical Sciences, University at Buffalo, State University of New York; Chief, Division of Infectious Diseases, Veterans Administration Western New York Healthcare Center, Buffalo, New York [166, 180, 181]

David P. Ryan, MD
Shelby Memorial Professor of Medicine, Harvard Medical School; Physician in Chief, Mass General Brigham Cancer Institute, Boston, Massachusetts [86]

Edward T. Ryan, MD
Professor of Medicine, Harvard Medical School; Professor of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health; Director, Global Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts [173, 236–238]

Manish Sadarangani, BM BCh, FRCPC, DPhil
Associate Professor, Division of Infectious Diseases, Department of Pediatrics, University of British Columbia; Director, Vaccine Evaluation Center, BC Children's Hospital Research Institute; Sauder Family Chair in Pediatric Infectious Diseases, Vancouver, British Columbia, Canada [160]

David I. Sahar, MD
Professor of Medicine; Kenneth A. Forde MD Clinical Educator, Milstein Division of Cardiology, Vagelos College of Physicians & Surgeons; Columbia University Irving Medical Center, New York, New York [492]

David J. Salant, MD
Professor of Medicine, Boston University Chobanian and Avedisian School of Medicine, Boston, Massachusetts [328]

Martin A. Samuels, MD, DSc (Hon), FACP, FAAN, FRCP, FANA[*]
Miriam Sydney Joseph Distinguished Professor of Neurology, Harvard Medical School; Founding Chair Emeritus, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts [V7]

Vaishali Sanchorawala, MD
Professor of Medicine; Director, Amyloidosis Center; Director, Autologous Stem Cell Transplantation Program, Boston University School of Medicine and Boston Medical Center, Boston, Massachusetts [117]

Vijay G. Sankaran, MD, PhD
Jan Ellen Paradise, MD Professor of Pediatrics, Harvard Medical School; Division of Hematology/Oncology, Boston Children's Hospital, Boston, Massachusetts [103]

Philippe J. Sansonetti, MD
Professor, Collège de France; Emeritus Professor, Institut Pasteur, Paris, France [172]

Clifford B. Saper, MD, PhD
James Jackson Putnam Professor of Neurology and Neuroscience, Harvard Medical School; Department of Neurology, Beth Israel Deaconess Medical Center, Boston, Massachusetts [33]

---

[*] Deceased

Oliver Sartor, MD
Director of Radiopharmaceutical Trials; Chief, Genitourinary Cancer Disease Group, Mayo Clinic, Rochester, Minnesota [92]

William H. Sauer, MD
Associate Professor of Medicine, Harvard Medical School; Section Chief, Cardiac Arrhythmia Service, Brigham and Women's Hospital, Boston, Massachusetts [250–263, 317]

Edward A. Sausville, MD, PhD
National Cancer Institute, Bethesda, Maryland; Greenebaum Comprehensive Cancer Center, University of Maryland, Baltimore, Maryland (Retired) [78]

David T. Scadden, MD
Gerald and Darlene Jordan Professor of Medicine; Chair Emeritus and Professor, Department of Stem Cell and Regenerative Biology, Harvard University; Director, Center for Regenerative Medicine; Massachusetts General Hospital, Co-director, Harvard Stem Cell Institute, Cambridge, Massachusetts [101]

Thomas E. Scammell, MD
Professor, Harvard Medical School; Beth Israel Deaconess Medical Center; Boston Children's Hospital, Boston, Massachusetts [33]

Esperance A. K. Schaefer, MD, MPH
Instructor, Harvard Medical School; Assistant in Medicine, Division of Gastroenterology, Massachusetts General Hospital, Boston, Massachusetts [350, 352]

Anthony H. V. Schapira, MD, DSc, FRCP, FMedSci
Head and Professor, Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology; Director of UCL Royal Free Campus; Vice-Dean UCL, London, United Kingdom [446]

Gordon Schiff, MD
Quality and Safety Director, Harvard Medical School Center for Primary Care; Associate Professor of Medicine, Harvard Medical School; Associate Director, Center for Patient Safety Research and Practice, Brigham and Women's Hospital, Boston, Massachusetts [10]

Scott Schissel, MD, PhD
Chief Medical Officer, Newark Beth Israel Medical Center, RWJBarnabas Health; Rutgers, New Jersey Medical School, Newark, New Jersey [313]

Bernd Schnabl, MD
Professor of Medicine, Department of Medicine, Division of Gastroenterology, University of California San Diego, La Jolla, California [353]

Marc A. Schuckit, MD
Distinguished Professor of Psychiatry, University of California, San Diego Medical School, La Jolla, California [464]

William W. Seeley, MD
Zander Family Distinguished Professor, Memory and Aging Center, Department of Neurology, Department of Pathology, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, California [31, 442–445]

Brahm H. Segal, MD
Professor of Medicine, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo; Chair, Department of Internal Medicine; Chief, Division of Infectious Diseases; Professor of Oncology; Member, Department of Immunology, Roswell Park Comprehensive Cancer Center, Buffalo, New York [79]

Julian L. Seifter, MD
Associate Professor of Medicine, Harvard Medical School; James G. Haidas Family Endowed Chair, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [319, 331]

Mājid Shafiq, MD, MPH
Assistant Professor of Medicine, Harvard Medical School; Medical Director, Interventional Pulmonology, Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts [297]

Ankoor Shah, MD
Associate Professor, Department of Medicine, Division of Rheumatology and Immunology, Duke University Medical Center, Durham, North Carolina [370]

Pinak B. Shah, MD
Associate Professor of Medicine, Harvard Medical School; Director, Cardiac Catheterization Laboratory, Brigham and Women's Hospital, Boston, Massachusetts [287]

Vikram G. Shakkottai, MD, PhD
Professor of Neurology, UT Southwestern Medical Center, Dallas, Texas [450]

Erica S. Shenoy, MD, PhD
Associate Professor of Medicine, Harvard Medical School; Medical Director, Infection Control, Mass General Brigham; Associate Chief, Infection Control Unit, Massachusetts General Hospital, Boston, Massachusetts [149]

Mark Sherlock, MB (Hons), BCh, BAO, LRCPI/SI, MRCPI, MD, PhD
Professor of Medicine, Royal College of Surgeons in Ireland, University of Medicine and Health Sciences; Consultant Endocrinologist, Beaumont Hospital, Dublin, Republic of Ireland [393]

Kanade Shinkai, MD, PhD
Professor of Dermatology, University of California, San Francisco, San Francisco, California [63]

Shmuel Shoham, MD
Professor of Clinical Medicine, Division of Infectious Diseases, Johns Hopkins University School of Medicine, Baltimore, Maryland [221]

Edwin K. Silverman, MD, PhD
Professor of Medicine, Harvard Medical School; Chief, Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts [303]

Michael D. Simpson, MD
Instructor, Harvard Medical School; Emergency Medicine Attending Physician, Beth Israel Deaconess Medical Center and Beth Israel Deaconess Needham Hospital; Medical Toxicology Attending Physician, Beth Israel Deaconess Medical Center, Boston, Massachusetts [471]

Prashant Singh, MBBS
Assistant Professor of Internal Medicine, University of Michigan Health System, Ann Arbor, Michigan [338]

Shakti Singh, MSc, PhD
Research Scientist, The Lundquist Institute at Harbor-UCLA Medical Center, Torrance, California [222]

Eric K. Singhi, MD
Assistant Professor, Department of Thoracic/Head and Neck Medical Oncology and Department of General Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas [83]

Karl L. Skorecki, MD, FRCP(C), FASN
Professor Emeritus and Former Dean, Azrieli Faculty of Medicine, Bar-Ilan University and Rappaport Faculty of Medicine, Technion – Israel Institute of Technology; Consultant in Nephrology and Population Genetics Research, Rambam Health Care Campus, Haifa, Israel [322, 481]

Chelsey J. F. Smith, MD
Wolfe Family Associate Director, Lupus Center of Excellence; Associate Professor of Medicine, University of California, San Diego, La Jolla, California [368]

Thomas J. Smith, MD, FACP, FASCO, FAAHPM
Director of Palliative Medicine, Johns Hopkins Medical Institutions, Division of General Internal Medicine, Section of Palliative Medicine, JHUSOM; Professor of Oncology, Sidney Kimmel Comprehensive Cancer Center; Harry J. Duffey Family Professor of Palliative Medicine, Baltimore, Maryland [74]

Wade S. Smith, MD, PhD
Professor of Neurology, Department of Neurology, University of California, San Francisco, San Francisco, California [318, 437, 438, 439, 440]

Jeremy Sobel, MD, MPH
Associate Director for Epidemiologic Science, Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [158]

Maria Söderlund-Venermo, PhD
Adjunct Professor of Virology, Department of Virology, University of Helsinki, Helsinki, Finland [202]

Scott D. Solomon, MD
The Edward D. Frohlich Distinguished Chair; Professor of Medicine, Harvard Medical School; Senior Physician, Brigham and Women's Hospital, Boston, Massachusetts [248, A9]

Nilam J. Soni, MD, MS, FACP, SFHM
Professor of Medicine, Joe R. and Teresa Lozano Long School of Medicine, University of Texas Health San Antonio; Director, Critical Care Ultrasound Education; Director, VA Point-of-Care Ultrasound Patient Safety Center, South Texas Veterans Health Care System, San Antonio, Texas [493]

Eric J. Sorscher, MD
Hertz Endowed Professorship, Emory University School of Medicine, Children's Healthcare of Atlanta, Atlanta, Georgia [302]

Brad Spellberg, MD
Chief Medical Officer, Los Angeles General Medical Center, Los Angeles, California [224]

Jerry L. Spivak, MD
Emeritus Professor of Medicine, Hematology Division, Johns Hopkins University School of Medicine, Baltimore, Maryland [108]

David Spriggs, MD, FASCO
Faculty Member, Harvard Medical School; Gynecologic Oncology, Program, Massachusetts General Hospital Cancer Center, Boston, Massachusetts [94]

E. William St. Clair, MD
W. Lester Brooks, Jr. Professor of Medicine; Professor of Immunology, Department of Medicine, Duke University Medical Center, Durham, North Carolina [370]

John M. Stafford, MD, PhD
Associate Professor of Medicine; Associate Director, Director, Vanderbilt Diabetes Research and Training Center; Director, Vanderbilt and NIDDK Medical Student Research Programs, Vanderbilt University Medical Center, Nashville, Tennessee [417]

Matthew W. State, MD, PhD
Oberndorf Family Distinguished Professor; Chair, Department of Psychiatry and Behavioral Sciences; Director, Langley Porter Psychiatric Hospital and Clinics, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, California [462]

Allen C. Steere, MD
Professor of Medicine, Harvard Medical School and Massachusetts General Hospital, Boston, Massachusetts [191]

Martin H. Steinberg, MD
Professor of Medicine, Pediatrics, Pathology and Laboratory Medicine, Boston University School of Medicine, Boston, Massachusetts [103]

Anna-Sofie Stensgaard, MSc, PhD
Associate Professor, Section for Parasitology and Aquatic Pathobiology, Department of Veterinary and Animal Sciences, University of Copenhagen, Frederiksberg, Denmark [241]

Dennis L. Stevens, MD, PhD
Professor of Medicine, University of Washington School of Medicine, Seattle, Washington [134, 159]

Lynne Warner Stevenson, MD
Lisa Jacobson Professor of Cardiovascular Medicine, Director of Cardiomyopathy Program, Vanderbilt University Medical Center, Nashville, Tennessee [266–270]

Benjamin K. Stoff, MD, MAB
Associate Professor of Dermatology, Emory University School of Medicine; Senior Faculty Fellow, Emory Center for Ethics, Atlanta, Georgia [A5]

John H. Stone, MD, MPH
Professor of Medicine, Harvard Medical School; The Edward Fox Chair in Medicine, Massachusetts General Hospital, Boston, Massachusetts [380]

Jeffrey R. Strich, MD, MHS
United States Public Health Service; Staff Clinician, National Institutes of Health; Head, Pathogenesis and Therapeutics Section, Critical Care Medicine Department, NIH Clinical Center and Critical Care Medicine Brach, National Heart Lung and Blood Institute, Bethesda, Maryland [315]

Shyam Sundar, MD
Distinguished Professor, Department of General Medicine, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India [233]

Neeraj K. Surana, MD, PhD
Assistant Professor of Pediatrics, Molecular Genetics and Microbiology, Integrative Immunobiology, and Cell Biology, Duke University, Durham, North Carolina [20, 124, 182, 484]

Paolo M. Suter, MD, MS
Department of Endocrinology, Diabetology, and Clinical Nutrition, University Hospital, Zurich, Switzerland [344]

Geoffrey Tabin, MD
Director, Department of Ophthalmology, Stanford University, Stanford, California [475]

Soichiro Takamiya, MD, PhD
Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada [19]

Maria Carmela Tartaglia, MD
Professor, Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Ontario, Canada [V2]

Usha B. Tedrow, MD, MSc
Associate Professor of Medicine, Harvard Medical School; Director, Clinical Cardiac Electrophysiology Fellowship; Clinical Director, Ventricular Arrhythmia Program, Brigham and Women's Hospital, Boston, Massachusetts [259–263]

Ayalew Tefferi, MD
Barbara Woodward Lips Professor of Medicine and Hematology, Mayo Clinic, Rochester, Minnesota [115]

Maria G. Tektonidou, MD, PhD
Professor of Rheumatology, Head of Rheumatology Unit, First Department of Propaedeutic and Internal Medicine, School of Medicine, National and Kapodistrian University of Athens, Athens, Greece [369]

Stephen C. Textor, MD*
Professor of Medicine, Division of Nephrology and Hypertension, Mayo Clinic College of Medicine, Rochester, Minnesota [289]

Rajesh V. Thakker, MD, FRCP, FRS
Professor of Medicine, University of Oxford, Oxford Centre for Diabetes, Endocrinology and Metabolism, Churchill Hospital, Oxford [400]

Hai Duong Ha Thi, MD, MSc OR
Vice Head of Adult Intensive Care Unit, Hospital for Tropical Diseases; Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam [157]

Holger Thiele, MD
Professor of Medicine, Director, Department of Internal Medicine/Cardiology; Heart Center Leipzig at University of Leipzig, Leipzig, Germany [316]

C. Louise Thwaites, MBBS, BSc, MD
Clinical Research Fellow, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam; Associate Professor, Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, United Kingdom [157]

Pierre Tiberghien, MD, PhD
Professor of Medicine, Franche-Comté University, Besançon; Etablissement Français du Sang, Paris, France [118]

Zelig A. Tochner, MD
Professor Emeritus of Radiation Oncology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania [S6]

Mark Topazian, MD
Professor of Medicine, Mayo Clinic, Rochester, Minnesota [333, V5]

Camilo Toro, MD
Director, Adult Undiagnosed Diseases Program, Office of the Director and Clinical Director/NHGRI, National Institutes of Health, Bethesda, Maryland [505]

Dwight A. Towler, MD, PhD
Professor of Medicine, UT Southwestern Medical Center, Dallas, Texas [57]

---

*Deceased

Barbara W. Trautner, MD, PhD
Professor, Section of Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, Texas [140]

Stephen Tsaur, MD
Assistant Professor of Medicine, Boston University School of Medicine; Associate Director Ambulatory Cardiology Clinic, Boston Medical Center, Boston, Massachusetts [282]

Kenneth L. Tyler, MD
Louise Baum Endowed Chair and Chairman of Neurology; Professor of Medicine and Immunology-Microbiology, University of Colorado School of Medicine, Aurora, Colorado; Neurologist, Rocky Mountain VA Medical Center, Aurora, Colorado [142, 143, 145]

Anne-Catrin Uhlemann, MD, PhD
Associate Professor of Medicine, Columbia University Irving Medical Center, New York, New York [152]

Elizabeth R. Unger, PhD, MD
Division of High-Consequence Pathogens and Pathology, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [461]

John W. Urwin, MD, MBA
Research Fellow, Harvard Medical School; Chief Fellow in Cardiovascular Disease, Massachusetts General Hospital, Boston, Massachusetts [13]

Prashant Vaishnava, MD, FACC, FACP
Director, Inpatient Cardiology Services Columbia Interventional Cardiovascular Care Program Assistant Professor of Medicine Columbia University College of Physicians&Surgeons, New York, New York [492]

Anne Marie Valente, MD
Professor of Pediatrics, Harvard Medical School; Director, Boston Adult Congenital Heart Disease and Pulmonary Hypertension Program, Boston Children's Hospital, Brigham and Women's Hospital, Boston, Massachusetts [280]

Wim H. van Brakel, MD, PhD
Medical Director, Until No Leprosy Remains, Amsterdam, The Netherlands [184]

Mathew G. Vander Heiden, MD, PhD
Professor and Director, Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts [502]

Edouard Vannier, PharmD, PhD
Assistant Professor of Medicine, Division of Geographic Medicine and Infectious Diseases, Tufts Medical Center; Tufts University School of Medicine, Boston, Massachusetts [232]

John Varga, MD
Frederick Huetwell Professor; Chief, Division of Rheumatology, University of Michigan, Ann Arbor, Michigan [372]

David J. Vaughn, MD
Genitourinary Medical Oncology Professor, Perelman School of Medicine at the University of Pennsylvania, Perelman Center for Advanced Medicine, Philadelphia, Pennsylvania [93]

Birgitte Jyding Vennervald, MD, MSA
Professor, Section for Parasitology and Aquatic Pathobiology, Department of Veterinary and Animal Sciences, University of Copenhagen, Frederiksberg, Denmark [241]

Abraham Verghese, MD, FRCP (Edin)
Linda R. Meier and Joan F. Lane Provostial Professor and Vice Chair, Department of Medicine, Stanford University School of Medicine, Stanford, California [8]

John T. Vetto, MD, FACS
Professor of Surgery, Division of Surgical Oncology; Director, Cutaneous Oncology Program, Department of Surgery, Oregon Health & Science University; Program Leader, Melanoma Disease Site Team, OHSU Knight Cancer Institute, Portland, Oregon [81]

Luciano Villarinho, MD
Neuroradiologist, South County Hospital, Wakefield, Rhode Island [A16]

Bert Vogelstein, MD
Professor, Ludwig Center for Cancer Genetics and Therapeutics, Johns Hopkins University School of Medicine; Investigator, Howard Hughes Medical Institute, Baltimore, Maryland [76]

Everett E. Vokes, MD
John E. Ultmann Distinguished Service Professor; Chairman, Department of Medicine; Physician-in-Chief, University of Chicago Medicine and Biological Sciences, Chicago, Illinois [82]

Tamara J. Vokes, MD
Professor, Department of Medicine, Section of Endocrinology, University of Chicago, Chicago, Illinois [424]

Nora D. Volkow, MD
Director, National Institute on Drug Abuse (NIDA), National Institutes of Health, Rockville, Maryland [466]

Kevin G. Volpp, MD, PhD
Mark V. Pauly President's Distinguished Professor, Perelman School of Medicine and the Wharton School; Director Penn Center for Health Incentives and Behavioral Economics, University of Pennsylvania, Philadelphia, Pennsylvania [494]

Daniel D. Von Hoff, MD, FACP, FASCO, FAACR
Distinguished Professor, Translational Genomics Research Institute (TGEN), Phoenix, Arizona; Virginia G. Piper Distinguished Chair for Innovative Cancer Research and Chief Scientific Officer, Honor Health Research Institute; Senior Consultant-Clinical Investigations, City of Hope; Professor of Medicine, Mayo Clinic, Scottsdale, Arizona [88]

Martin H. Voss, MD
Associate Attending Physician, Genitourinary Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York [90]

Christine A. Vu, PharmD
Clinical Infectious Diseases Pharmacist, Jackson Memorial Hospital, Miami, Florida [167]

Jiři F. P. Wagenaar, MD, PhD
Internist and Infectious Disease Specialist, Northwest Clinics, Alkmaar, The Netherlands [189]

Jesse Waggoner, MD
Assistant Professor, Department of Medicine, Division of Infectious Diseases, Emory University, Atlanta, Georgia [130]

Sushrut S. Waikar, MD, MPH
Chief, Section of Nephrology; Norman G. Lewinsky Professor of Medicine, Boston University Chobanian and Avedisian School of Medicine,, Boston, Massachusetts [321]

Matthew K. Waldor, MD, PhD
Edward H. Kass Professor of Medicine, Harvard Medical School, Division of Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts [173]

David H. Walker, MD
The Carmage and Martha Walls Distinguished University Chair in Tropical Diseases; Professor, Department of Pathology; Executive Director, Center for Biodefense and Emerging Infectious Diseases, University of Texas Medical Branch, Galveston, Texas [192]

Mark F. Walker, MD
Associate Professor, Neurology, Case Western Reserve University; Director, Daroff-Dell'Osso Ocular Motility Laboratory, VA Northeast Ohio Healthcare System, Cleveland, Ohio [24]

Eric J. Wang, MD
Associate Clinical Professor; Division of Pain Medicine, Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland [18, V8]

George R. Washko, MD, MMSc
Professor of Medicine, Harvard Medical School; Associate Physician, Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [296, 297, 307, A12]

Michael E. Wechsler, MD, MMSc
Professor of Medicine; Director, Asthma Program, Department of Medicine, National Jewish Health, Denver, Colorado [299]

Anthony P. Weetman, MD, DSc
Emeritus Professor of Medicine, School of Medicine and Population Health, University of Sheffield, Sheffield, United Kingdom [394–397]

Ana A. Weil, MD, MPH
Assistant Professor, Department of Medicine, Division of Allergy and Infectious Diseases; Adjunct Assistant Professor, Department of Global Health, University of Washington, Seattle, Washington [138]

Marc N. Wein, MD, PhD
Associate Professor of Medicine, Harvard Medical School, Endocrine Unit, Massachusetts General Hospital, Boston, Massachusetts [421]

Jeffrey I. Weitz, OC, MD, FRCP(C), FRSC, FACP
Professor of Medicine and Biochemistry and Biomedical Sciences, McMaster University; Executive Director, Thrombosis and Atherosclerosis Research Institute, Hamilton, Ontario, Canada [123]

Peter F. Weller, MD
William Bosworth Castle Professor of Medicine, Harvard Medical School; Professor of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health; Chief Emeritus, Infectious Diseases Division and Vice Chair of Research, Chief, Allergy Division Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts [236–238]

Andrew Wellman, MD, PhD
Professor of Medicine, Harvard Medical School; Director, Sleep Disordered Breathing Lab, Brigham and Women's Hospital, Boston, Massachusetts [308]

Patrick Y. Wen, MD
Professor of Neurology, Harvard Medical School; Director, Center for Neuro-Oncology, Dana-Farber Cancer Institute; Director, Division of Neuro-Oncology, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts [95]

Michael R. Wessels, MD
John F. Enders Professor of Pediatrics and Professor of Medicine, Harvard Medical School; Senior Physician, Division of Infectious Diseases, Boston Children's Hospital, Boston, Massachusetts [153]

L. Joseph Wheat, MD
Medical Director, MiraVista Diagnostics, Indianapolis, Indiana [218]

Paul K. Whelton, MB, MD, MSc
Professor of Epidemiology and Medicine; Show Chwan Chair in Global Public Health, Tulane University Health Sciences Center, New Orleans, Louisiana [288]

A. Clinton White, Jr., MD, FACP, FIDSA, FASTMH
Professor, Infectious Disease Division, Department of Internal Medicine, University of Texas Medical Branch, Galveston, Texas [242]

Nicholas J. White, DSc, MD, FRCP, F Med Sci, FRS
Professor of Tropical Medicine, Mahidol Oxford Research Unit, Bangkok, Thailand [231, A2]

Richard J. Whitley, MD
Distinguished Professor; Loeb Eminent Scholar in Pediatrics; Professor of Pediatrics, Microbiology, Medicine and Neurosurgery, University of Alabama at Birmingham, Birmingham, Alabama [198]

Daniel Willie-Permor, MD, MPH
Postdoctoral Scholar, University of California, San Diego, San Diego, California [340]

Michael R. Wilson, MD, MAS
Rachleff Family Distinguished Professor in Neurology, University of California San Francisco Weill Institute for Neurosciences; Staff Physician, University of California San Francisco Medical Center and Zuckerberg San Francisco General Hospital, San Francisco, California [142, 144, S2]

Bruce U. Wintroub, MD
Professor of Dermatology, Department of Dermatology, University of California, San Francisco, San Francisco, California [63]

Allan W. Wolkoff, MD
The Herman Lopata Chair in Liver Disease Research; Professor of Medicine and Developmental and Molecular Biology; Associate Chair of Medicine for Research; Chief, Division of Hepatology; Director, Marion Bessin Liver Research Center, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, New York [349]

John B. Wong, MD
Professor of Medicine, Tufts University School of Medicine; Vice Chair, Academic Affairs, Chief, Division of Clinical Decision Making, Department of Medicine, Tufts Medical Center, Boston, Massachusetts [4]

Louis Michel Wong Kee Song, MD
Professor of Medicine, Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, Minnesota [333, V5]

Jennifer A. Woyach, MD
Professor of Medicine, Division of Hematology, The Ohio State University, Columbus, Ohio [112]

Henry M. Wu, MD, DTM&H, FIDSA
Associate Professor of Medicine, Division of Infectious Diseases, Emory University; Director, Emory TravelWell Center, Atlanta, Georgia [130]

Kim B. Yancey, MD
Professor and Chair, Department of Dermatology, University of Texas Southwestern Medical Center in Dallas, Dallas, Texas [59, 62]

Lonny Yarmus, DO, MBA
Professor of Medicine and Oncology; Vice Chair, Department of Medicine, Clinical Business Development and Strategic Planning; Director, Interventional Pulmonary, Johns Hopkins University School of Medicine, Baltimore, Maryland [310]

Yusuf Yazici, MD
Clinical Associate Professor of Medicine, New York University Grossman School of Medicine, New York, New York [376]

Baligh R. Yehia, MD, MPP, MSc
Thomas Jefferson University, Philadelphia, Pennsylvania [412]

Janet A. Yellowitz, DMD, MPH
Associate Professor; Director, Special Care and Geriatric Dentistry, University of Maryland School of Dentistry, Baltimore, Maryland [A3]

Neal S. Young, MD
Chief, Hematology Branch, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland [107, 482]

Paul C. Zei, MD, PhD
Associate Professor of Medicine, Harvard Medical School; Director, Clinical Atrial Fibrillation Program, Brigham and Women's Hospital, Boston, Massachusetts [250, 253–258]

Jing Zhou, MD, PhD, FASN
Professor of Medicine, Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts [327]

Werner Zimmerli, MD
Professor of Medicine, Basel University, Interdisciplinary Unit of Orthopaedic Infections, Kantonsspital Baselland, Liestal, Switzerland [136]

Laura A. Zimmerman, MPH
Epidemiologist, Centers for Disease Control and Prevention, Atlanta, Georgia [212]

Rebecca L. Zon, MD
Clinical Instructor, Harvard Medical School; Dana-Farber Cancer Institute, Brigham and Women's Hospital, Boston, Massachusetts [329]

Preface

The Editors are pleased to present the 22nd edition of *Harrison's Principles of Internal Medicine. Harrison's* is a landmark in medicine, spanning 75 years and multiple generations of trainees and practicing clinicians. While medicine and medical education have evolved, readers will appreciate how this classic textbook has retained enduring features that have distinguished it among medical texts—a sharp focus on the clinical presentation of disease, expert in-depth summaries of pathophysiology and treatment, and highlights of emerging frontiers of science and medicine. Indeed, *Harrison's* retains its conviction that, in the profession of medicine, we are all perpetual students with lifelong learning as our common goal.

Harrison's is intended for learners throughout their careers. For *students*, Part 1, Chapter 1 begins with an overview of "The Practice of Medicine." In this introductory chapter, the editors continue the tradition of orienting clinicians to the *science* and the *art* of medicine, emphasizing the values of our profession while incorporating new advances in technology, science, and clinical care. Part 2, "Cardinal Manifestations and Presentation of Diseases," is a signature feature of *Harrison's*. These chapters eloquently describe how patients present with common clinical conditions, such as headache, fever, cough, palpitations, or anemia, and provide an overview of typical symptoms, physical findings, and differential diagnoses. Mastery of these topics prepares students for subsequent chapters on specific diseases they will encounter in courses on pathophysiology and in clinical clerkships. For *residents* and *fellows* caring for patients and preparing for board examinations, *Harrison's* remains a definitive source of trusted content written by internationally renowned experts. Trainees will be fully informed by the depth of content, comprehensive tables, and illuminating figures and clinical algorithms. *Practicing clinicians and health professionals who deliver direct patient care, like physician associates and nurse practitioners,* must keep up with an ever-expanding knowledge base and clinical guidelines as part of lifelong learning. Clinicians can trust that chapters are updated extensively with each edition of *Harrison's*. The text is an excellent point-of-care reference for clinical questions, differential diagnosis, and patient management. In addition to the expanded and detailed "Treatment" sections, *Harrison's* continues its tradition of including "Approach to the Patient" sections, which provide an expert's overview of the practical management of common but often complex clinical conditions.

This new edition has been modified extensively and offers a standardized format for each disease chapter. The authors and editors have curated rigorously and synthesized the vast amount of information that comprises general internal medicine—and each of the major specialties—into a highly readable and informative two-volume book. Readers will appreciate the concise writing style and substantive quality that have always characterized *Harrison's*. This book has a sharp focus on essential information with a goal of providing clear and definitive answers to clinical questions.

In the 22nd edition, examples of new chapters include "The Value of the Physical Examination in Modern Medicine," "Physician Well-Being," "Placebo and Nocebo Effects," "Symptom Control in Patients with Cancer," "Principles of Immunization," "Desensitization," "Point-of-Care Ultrasound," and five new chapters on cardiomyopathies. Many chapters have new authors. Updated information is included on artificial intelligence, the use of gliptins (DPP-4 inhibitors) and glucagon-like peptide agonists, drug-resistant tuberculosis, and HIV prevention. Current state-of-the-art summaries on gene therapies for a wide range of disorders are included. The latest results on the explosion of biologic therapies for cancer and autoimmune diseases are detailed. New therapies are transforming the outlook for patients suffering from neurologic conditions, and throughout the 22nd edition, readers will find the updated management guidelines for these complex conditions. Updated coverage of new therapies spanning all of adult medicine is presented in this 22nd edition, carefully assessed and presented by leading authorities in the respective fields.

Harrison's content is available in a variety of print and digital formats, including eBooks, apps, and a popular, widely used online platform available at *accessmedicine.com.* A very useful companion, *Harrison's Internal Medicine Review Questions and Explanations,* includes over 1000 Q&A and offers comprehensive explanations of the correct answer. All Q&A provide links to the relevant chapters in the textbook. This popular content is available on *AccessMedicine.com* and is regularly updated with new and revised Q&A. The podcast based on this self-assessment text, *Harrison's Podclass,* has enjoyed great global popularity with more than 3 million streams and downloads.

This edition has begun a transition for the book. The editors have enjoyed a remarkable stability as colleagues working together to create an informative, updated, and practical book. With this edition, we have added two new editors who will work alongside two current editors for whom this 22nd edition will be their last. Anthony S. Fauci has been a *Harrison's* editor since the 11th edition and was Editor-in-Chief for editions 14 and 17. It isn't possible to accurately describe his role in shaping the book. He will ultimately be replaced by Carol Langford, MD, Harold C. Schott Endowed Professor and Director, Center for Vasculitis Care and Research, Cleveland Clinic Foundation, our first woman to join the Editorial Board. She has worked closely with Dr. Fauci on this edition and will take over his sections on her own in the next edition.

Dennis L. Kasper, MD, the Editor responsible for the Infectious Disease section of the book since the 13th edition, has carried the heaviest load in chapter numbers and has written about a field of medicine that has been buffeted the most by unforeseen emerging epidemics that need to be analyzed and understood in a timely fashion by readers of our pages. He was Editor-in-Chief for editions 16 and 19. For our 22nd edition, Dr. Kasper is joined by Steven M. Holland, MD, who will ultimately take over this important part of *Harrison's*. Dr. Holland is a distinguished physician-scientist with expertise in both microbial and host mechanisms in infectious diseases and in acquired and congenital immunodeficiency.

This is also the final edition for our superb Neurology/Psychiatry editor, Stephen L. Hauser, MD. Steve joined the editors with the 14th edition and handled his section again in this edition without assistance. The new Neurology/Psychiatry editor will be S. Andrew Josephson, MD, Professor and Chairman, Department of Neurology, University of California, San Francisco.

We have many people to thank for their efforts in producing this book. First, the authors have done a superb job of producing authoritative chapters that synthesize vast amounts of scientific and clinical data to create informative and practical approaches to managing patients. In today's information-rich, rapidly evolving environment, they have ensured that this information is current. We are most grateful to our colleagues who work closely with each editor to facilitate communication with the authors and help us keep *Harrison's* content current. In particular, we wish to acknowledge the expert support of Lauren Bauer, Rosemary Cannistraro, Gregory K. Folkers, Jenn Kuzmuk, Laura Nickrosz, Elizabeth Robbins, Marie Scurti, Katherine Seropian, Geeta Strange, and Stephanie C. Tribuna. Scott Grillo and James Shanahan, our long-standing partners at McGraw Hill's Professional Educational Publishing group, have inspired the creative and dynamic evolution of *Harrison's*, guiding the development of the book and its related products in new formats. Samantha Shapiro, MD, provided guidance on new formats for the content. Kim Davis, as Managing Editor, has adeptly ensured that the complex production of this multi-authored textbook proceeded smoothly and efficiently. Priscilla Beer and Steve Brown oversaw the production of our videos and digital content; Jeffrey Herzich, Melinda Avelar, and Rachel Norton, along with other members of the McGraw Hill staff, and Revathi Viswanathan of KnowledgeWorks Global Ltd., shepherded the production of this new edition.

We are privileged to have compiled this 22nd edition and are enthusiastic about all that it offers our readers. We learned much in the process of editing *Harrison's* and hope that you will find this edition uniquely valuable as a clinical and educational resource.

We dedicate this edition to the students, trainees, and physicians utilizing *Harrison's* who remain steadfast in their quest for knowledge to equip themselves to offer the best care to their patients. We also express our gratitude to the many thousands of patients who courageously participate each year in clinical trials, the engine that fuels advances in clinical medicine.

The Editors

Related Harrison's Resources

A complete collection to meet your educational, clinical, and board prep needs.

Harrison's Online

Harrison's Principles of Internal Medicine is available online at *accessmedicine.com* and via the AccessMedicine app with an institutional or individual subscription. *Harrison's Online* features all the chapters from the print edition, plus more than two dozen supplementary chapters in print, atlas, and video formats. The content is regularly updated to reflect important new developments in medical research and practice. Easily search across the entire *Harrison's* content set, access high-yield key points of *Harrison's* chapters, download images and tables for presentations and lectures, view step-by-step videos on common clinical procedures, access the text of the *Harrison's Manual of Medicine,* set up a personalized test exam for board prep from over 1000 review questions, get access to chapters from new editions of *Harrison's* months before book publication, and more.

Harrison's Internal Medicine Review Questions and Explanations

Formerly the *Harrison's Self-Assessment and Board Review, Harrison's Review Questions* provides more than 1000 self-assessment questions, with new questions added quarterly. Questions are presented in board-style clinical vignette format with multiple choice answers, making this a convenient resource for primary and recertification exam prep, rotational shelf exams, and general assessment of understanding of the principles of clinical medicine. The explanations for the questions are comprehensive and provide detailed guidance on correct and incorrect answers. Question and answer sets include references to related chapters in *Harrison's Principles of Internal Medicine* for more comprehensive understanding. This resource is available on *accessmedicine.com*, where users can create personalized testing experiences and receive instant scores on practice tests.

Harrison's Podclass

Harrison's Podclass delivers illuminating and engaging discussions led by Catherine Handy Marshall and Charles M. Wiener of the John Hopkins School of Medicine on key topics in medicine and features board-style case vignettes from *Harrison's Review Questions* and chapters from *Harrison's Principles of Internal Medicine*. The hosts work through correct and incorrect answer choices and summarize cases with practical pearls that all students and clinicians will find helpful and interesting. *Harrison's Podclass* is updated weekly and available on most podcast outlets and on *accessmedicine.com*.

1

The Practice of Medicine

The Editors

ENDURING VALUES OF THE MEDICAL PROFESSION

> *No greater opportunity, responsibility, or obligation can fall to the lot of a human being than to become a physician. In the care of the suffering, [the physician] needs technical skill, scientific knowledge, and human understanding. Tact, sympathy, and understanding are expected of the physician, for the patient is no mere collection of symptoms, signs, disordered functions, damaged organs, and disturbed emotions. [The patient] is human, fearful, and hopeful, seeking relief, help, and reassurance.*
>
> —*Harrison's Principles of Internal Medicine,* 1950

The practice of medicine has changed in significant ways since the first edition of this book was published in 1950. The advent of molecular genetics, sophisticated new imaging techniques, robotics, and advances in bioinformatics and information technology have contributed to an explosion of scientific information that has changed fundamentally the way physicians define, diagnose, treat, and attempt to prevent disease. This growth of scientific knowledge continues to evolve at an accelerated pace.

The widespread use of electronic medical records and the Internet have altered the way physicians and other health care providers access and exchange information as a routine part of medical education and practice **(Fig. 1-1)**. As today's physicians strive to integrate an ever-expanding body of scientific knowledge into everyday practice, it is critically important to remember two key principles: first, the ultimate goal of medicine is to prevent disease and, when it occurs, to diagnose it early and provide effective treatment; and second, despite 70 years of scientific advances since the first edition of this text, a trusting relationship between physician and patient still lies at the heart of effective patient care.

THE SCIENCE AND ART OF MEDICINE

Deductive reasoning and applied technology form the foundation for the approach and solution to many clinical problems. Extraordinary advances in biochemistry, cell biology, immunology, and genomics, coupled with newly developed imaging techniques, provide a window into the most remote recesses of the body and allow access to the innermost parts of the cell. Revelations about the nature of genes and single cells have opened a portal for understanding physiology and unraveling the causes of disease. Researchers are deciphering the complex mechanisms by which genes are regulated, and increasingly, physicians are learning how subtle changes in many different genes, acting in an integrative contextual way, can affect the function of cells and organisms. Clinicians have developed a new appreciation of the role of stem cells in the renewal and repair of normal tissues, in the development of cancer and other disorders, and in the treatment of certain diseases. Entirely new areas of research, including studies of the human microbiome, epigenetics, and noncoding RNAs as regulatory features of the genome, have become important for understanding both health and disease. Extraordinary advances in vaccine platform technology and the use of cryo-electron microscopy for the structure-based design of vaccine immunogens have transformed the field of vaccinology, resulting in the unprecedented speed and success with which COVID-19 vaccines were developed. Information technology enables the interrogation of medical records from millions of individuals, yielding new insights into the etiology, characteristics, prognosis, and stratification of many diseases. With the increasing availability of very large data sets ("big data") from omic analyses and the electronic medical record, there is now a growing interest in machine learning and artificial intelligence for unbiased analyses that enhance clinical predictive accuracy. The field of artificial intelligence in particular is expanding rapidly, perhaps more rapidly than our understanding of how to use it optimally. In the coming years, it will be essential to learn the value and limitations of artificial intelligence methods and not abdicate decision-making entirely to a computer algorithm. The knowledge gleaned from the ***science of medicine*** continues to enhance the understanding by physicians of complex pathologic processes and to provide new approaches to disease prevention, diagnosis, and treatment. With continued refinement of unique omic signatures coupled with nuanced clinical pathophenotypes, the profession moves ever closer to practical precision medicine. Yet, skill in the most sophisticated applications of laboratory technology and in the use of the latest therapeutic modality alone does not make a good physician.

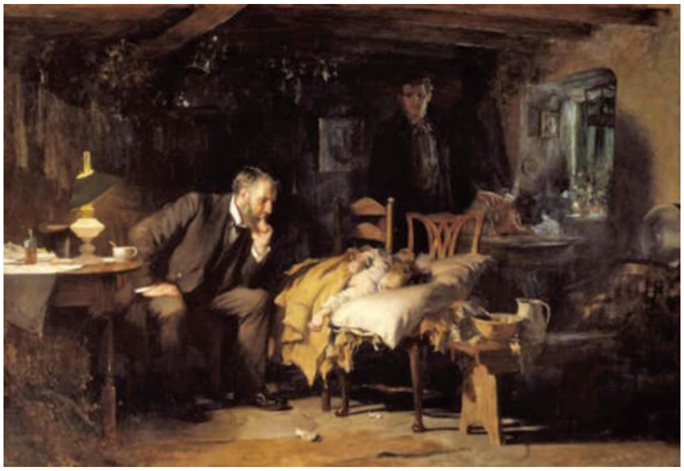

FIGURE 1-1 ***The Doctor*** **by Luke Fildes depicts the caring relationship** between this Victorian physician and a very ill child. Painted in 1891, the painting reflects the death of the painter's young son from typhoid fever and was intended to reflect the compassionate care provided by the physician even when his tools were not able to influence the course of disease. *(Source: History and Art Collection/Alamy Stock Photo.)*

When a patient poses challenging clinical problems, an effective physician must be able to identify the crucial elements in a complex history and physical examination; order the appropriate laboratory, imaging, and diagnostic tests; and extract the key results from densely populated computer screens to determine whether to treat or to "watch." As the number of tests increases, so does the likelihood that some incidental finding, completely unrelated to the clinical problem at hand, will be uncovered. Deciding whether a clinical clue is worth pursuing or should be dismissed as a "red herring" and weighing whether a proposed test, preventive measure, or treatment entails a greater risk than the disease itself are essential judgments that a skilled clinician must make many times each day. This combination of medical knowledge, intuition, experience, and judgment defines the ***art of medicine***, which is as necessary to the practice of medicine and the precision medicine of the future as is a sound scientific base, and as important for contemporary medical practice as it has been in earlier eras.

CLINICAL SKILLS

History-Taking The recorded history of an illness should include all the facts of medical significance in the life of the patient. Recent events should be given the most attention. Patients should, at some early point, be given the opportunity to tell their own story of the illness without frequent interruption and, when appropriate, should receive expressions of interest, encouragement, and empathy from the physician. Any event related by a patient, however trivial or seemingly irrelevant, may provide the key to solving the medical problem. A methodical review of systems is important to elicit features of an underlying disease that might not be mentioned in the patient's narrative. In general, patients who feel comfortable with the physician will offer more complete information; thus, putting the patient at ease contributes substantially to obtaining an adequate history.

An informative history involves more than eliciting an orderly listing of symptoms. By listening to patients and noting the ways in which they describe their symptoms, physicians can gain valuable insight. Inflections of voice, facial expression, gestures, and attitude (i.e., "body language") may offer important clues to patients' perception of and reaction to their symptoms. Because patients vary considerably in their medical sophistication and ability to recall facts, the reported medical history should be corroborated whenever possible. The social history also can provide important insights into the types of diseases that should be considered and can identify practical considerations for subsequent management. The family history not only identifies rare genetic disorders or common exposures, but often reveals risk factors for common disorders, such as coronary heart disease, hypertension, autoimmunity, certain cancers, and asthma. A thorough family history may require input from multiple relatives to ensure completeness and accuracy. An experienced clinician can usually formulate a relevant differential diagnosis from the history alone, using the physical examination and diagnostic tests to narrow the list or reveal unexpected findings that lead to more focused inquiry.

The very act of eliciting the history provides the physician with an opportunity to establish or enhance a unique bond that can form the basis for a good patient–physician relationship. This process helps the physician develop an appreciation of the patient's view of the illness, the patient's expectations of the physician and the health care system, and the financial and social implications of the illness for the patient. Although current health care settings may impose time constraints on patient visits, it is important not to rush the encounter. A hurried approach may lead patients to believe that what they are relating is not of importance to the physician, and, as a result, they may withhold relevant information. The confidentiality of the patient–physician relationship cannot be overemphasized.

Physical Examination The purpose of the physical examination is to identify physical signs of disease. The significance of these objective indications of disease is enhanced when they confirm a functional or structural change already suggested by the patient's history. At times, however, physical signs may be the only evidence of disease and may not have been suggested by the history.

The physical examination should be methodical and thorough, with consideration given to the patient's comfort and modesty. Although attention is often directed by the history to the diseased organ or part of the body, the examination of a new patient must extend from head to toe in an objective search for abnormalities or other diagnostic clues. The results of the examination, like the details of the history, should be recorded at the time they are elicited—not hours later, when they are subject to the distortions of memory. Physical examination skills should be learned under direct observation of experienced clinicians. Even highly experienced clinicians can benefit from ongoing coaching and feedback. Simulation laboratories and standardized patients play an increasingly important role in the development of clinical skills. Although the skills of physical diagnosis are acquired with experience, it is not merely technique that determines success in identifying signs of disease. The detection of a few scattered petechiae, a faint diastolic murmur, or a small mass in the abdomen is not only a question of keen eyes and ears or more sensitive fingers but also of a mind alert to those findings. Because physical findings can change with time, the physical examination should be repeated as frequently as the clinical situation warrants.

Given the many highly sensitive diagnostic tests now available (particularly imaging techniques), it may be tempting to place less emphasis on the physical examination. Some are critical of physical diagnosis based on perceived low levels of specificity and sensitivity. Indeed, many patients are seen by consultants only after a series of diagnostic tests have been performed and the results are known. This fact should not deter the physician from performing a thorough physical examination since important clinical findings may have escaped detection by diagnostic tests. Especially important, a thorough and thoughtful physical examination may render a laboratory finding unimportant (i.e., certain echocardiographic regurgitant lesions). The act of a hands-on examination of the patient also offers an opportunity for communication and may have reassuring effects that foster the patient–physician relationship.

Diagnostic Studies Physicians rely increasingly on a wide array of laboratory and imaging tests to make diagnoses and ultimately to solve clinical problems; however, such information does not relieve the physician from the responsibility of carefully observing and examining the patient. It is also essential to appreciate the limitations of diagnostic tests. By virtue of their apparent precision, these tests often gain an aura of certainty regardless of the fallibility of the tests themselves, the instruments used in the tests, and the individuals performing or interpreting the tests. Physicians must weigh the expense involved in laboratory procedures against the value of the information these procedures are likely to provide.

Single laboratory tests are rarely ordered. Instead, physicians generally request "batteries" of multiple tests, which often prove useful and can be performed with a single specimen at relatively low cost. For example, abnormalities of hepatic function may provide the clue to nonspecific symptoms such as generalized weakness and increased fatigability, suggesting a diagnosis of chronic liver disease. Sometimes a single abnormality, such as an elevated serum calcium level, points to a particular disease, such as hyperparathyroidism.

The thoughtful use of screening tests (e.g., measurement of low-density lipoprotein cholesterol) may allow early intervention to prevent disease **(Chap. 6)**. Screening tests are most informative when they are directed toward common diseases and when their results indicate whether other potentially useful—but often costly—tests or interventions are needed. On the one hand, biochemical measurements, together with simple laboratory determinations such as routine serum chemistries, blood counts, and urinalysis, often provide a major clue to the presence of a pathologic process. On the other hand, the physician must learn to evaluate occasional screening-test abnormalities that do not necessarily connote significant disease. An in-depth workup after the report of an isolated laboratory abnormality in a person who is otherwise well is often wasteful and unproductive. Because so many tests are performed routinely for screening purposes, it is not unusual for one or two values to be slightly outside the normal range or to reflect physiologic variation. Nevertheless, even if there is no reason to suspect an underlying illness, tests yielding abnormal results ordinarily are repeated to rule out laboratory error. If an abnormality is confirmed, it is important to consider its potential significance in the context of the patient's condition and other test results.

There is almost continual development of technically improved imaging studies with greater sensitivity and specificity. These tests provide remarkably detailed anatomic information that can be pivotal in informing medical decision-making. MRI, CT, ultrasonography, a variety of isotopic scans, and positron emission tomography (PET) have supplanted older, more invasive approaches and opened new diagnostic vistas. In light of their capabilities and the rapidity with which they can lead to a diagnosis, it is tempting to order a battery of imaging studies. All physicians have had experiences in which imaging studies revealed findings that led to an unexpected diagnosis. Nonetheless, patients must endure each of these tests, and the added cost of unnecessary testing is substantial. Furthermore, investigation of an unexpected abnormal finding may lead to an iatrogenic complication or to the diagnosis of an irrelevant or incidental problem. A skilled physician must learn to use these powerful diagnostic tools judiciously, always considering whether the results will alter management and benefit the patient.

MANAGEMENT OF PATIENT CARE

Team-Based Care Medical practice has long involved teams, particularly physicians working with nurses and, more recently, with physician assistants and nurse practitioners. Advances in medicine have increased our ability to manage very complex clinical situations (e.g., intensive care units [ICUs], bone marrow transplantation) and have shifted the burden of disease toward chronic illnesses. Because an individual patient may have multiple chronic diseases, they may be cared for by several specialists as well as a primary care physician. In the inpatient setting, care may involve multiple consultants along with the primary admitting physician. Communication through the medical record is necessary but not sufficient, particularly when patients have complex medical problems or when difficult decisions need to be made about the optimal management plan. Physicians should optimally meet face-to-face or by phone to ensure clear communication and thoughtful planning. It is important to note that patients often receive or perceive different messages from various care providers; thus, attempts should be made to provide consistency among these messages to the patient. Management plans and treatment options should be outlined succinctly and clearly for the patient.

Another dimension of team-based care involves allied health professions. It is not unusual for a hospitalized patient to encounter physical therapists, pharmacists, respiratory therapists, radiology technicians, social workers, dieticians, and transport personnel (among others) in addition to physicians and nurses. Each of these individuals contributes to clinical care as well as to the patient's experience with the health care system. In the outpatient setting, disease screening and chronic disease management are often carried out by nurses, physician assistants, or other allied health professionals. The roles and responsibilities of various team members can be confusing for patients. Within reason, different providers should introduce themselves, explain their role, and wear nametags or other identifying information.

The growth of team-based care has important implications for medical culture, student and resident training, and the organization of health care systems. Despite diversity in training, skills, and responsibilities among health care professionals, common values need to be espoused and reinforced. Many medical schools have incorporated interprofessional teamwork into their curricula. Effective communication is inevitably the most challenging aspect of implementing team-based care. While communication can be aided by electronic devices, including medical records, apps, or text messages, it is vitally important to balance efficiency with taking the necessary time to speak directly with colleagues.

The Dichotomy of Inpatient and Outpatient Internal Medicine The hospital environment has undergone sweeping changes over the past few decades. Emergency departments and critical care units have evolved to manage critically ill patients, allowing them to survive formerly fatal conditions. In parallel, there is increasing pressure to reduce the length of stay in the hospital and to manage complex disorders in the outpatient setting. This transition has been driven not only by efforts to reduce costs but also by the availability of new outpatient technologies, such as imaging and percutaneous infusion catheters for long-term antibiotics or nutrition, minimally invasive surgical procedures, and evidence that outcomes often are improved by reducing inpatient hospitalization. Telehealth and remote monitoring tools can also enhance connectivity with patients and reduce visits to the clinic or hospital.

In addition to traditional medical beds, hospitals now encompass multiple distinct levels of care, such as the emergency department, procedure rooms, overnight observation units, critical care units, and palliative care units. A consequence of this differentiation has been the emergence of new specialties (e.g., emergency medicine and end-of-life care) and the provision of in-hospital care by hospitalists and intensivists. Most *hospitalists* are board-certified internists who bear primary responsibility for the care of hospitalized patients and whose work is limited entirely to the hospital setting. The shortened length of hospital stay means that most patients receive only acute care while hospitalized; the increased complexities of inpatient medicine make the presence of an internist with specific training, skills, and experience in the hospital environment extremely beneficial. *Intensivists* are board-certified physicians who are further certified in critical care medicine and who direct and provide care for very ill patients in critical care units. Clearly, an important challenge in internal medicine today is to ensure the continuity of communication and information flow between a patient's primary care physician and those who are in charge of the patient's hospital care. Maintaining these channels of communication is frequently complicated by patient "handoffs"—i.e., transitions from the outpatient to the inpatient environment, from the critical care unit to a general medicine floor, from a medical to a surgical service and vice versa, from the hospital environment to the recently developed "home hospital" setting (for select patients with adequate home support), and from the hospital or home hospital to the outpatient environment.

The involvement of many care providers in conjunction with these transitions can threaten the traditional one-to-one relationship between patient and primary care physician. Of course, patients can benefit greatly from effective collaboration among a number of health care professionals; however, *it is the duty of the patient's principal or primary physician to coordinate these collaborations and provide cohesive guidance through an illness*. To meet this challenge, primary care physicians must be familiar with the techniques, skills, and objectives of specialist physicians and allied health professionals who care for their patients in the hospital. In addition, primary care physicians must ensure that their patients benefit from scientific advances and the expertise of specialists, both in and out of the hospital. Primary care physicians should explain the role of these specialists to reassure patients that they are in the hands of physicians best trained to manage their current illness. However, the primary care physician should assure patients and their families that decisions are being made in consultation with these specialists. The evolving concept of the "medical home" incorporates team-based primary care with subspecialty care in a cohesive environment that ensures smooth transitions of care.

Mitigating the Stress of Acute Illness Few people are prepared for a new diagnosis of cancer or anticipate the occurrence of a myocardial infarction, stroke, or major accident. The care of a frightened or distraught patient is confounded by these understandable responses to life-threatening events. The physician and other health providers can reduce the shock of life-changing events by providing information in a clear, calm, consistent, and reassuring manner. Often, information and reassurance need to be repeated. Caregivers should also recognize that, for the typical patient, hospital emergency rooms, operating rooms, ICUs, and general medical floors represent an intimidating and often frightening environment. Hospitalized patients find themselves surrounded by air jets, buttons, and glaring lights; invaded by tubes and wires; and beset by the numerous members of the health care team—hospitalists, specialists, nurses, nurses' aides, physician assistants, social workers, technologists, physical therapists, medical students, house officers, attending and consulting physicians, and many others. They may be transported to special laboratories and imaging facilities replete with blinking lights, strange sounds, and unfamiliar personnel; they may be left unattended at times; and they may be obligated to share a room with other patients who have their own health problems. It is little wonder that patients may find this environment bewildering and stressful. The additive effects of an acute illness, unfamiliar environment, multiple medications, and sleep deprivation can lead to confusion or delirium, especially in older hospitalized patients. Physicians who appreciate the hospital experience from the patient's perspective and who make an effort to guide and comfort the patient through this experience may make a stressful situation more tolerable and enhance the patient's chances for an optimal recovery. The advent of home-hospital care represents an attempt to improve the patient experience during uncomplicated acute illness by providing care in the familiar surroundings of the home environment, mitigating the stress of conventional hospitalization and shortening recovery times.

Medical Decision-Making Medical decision-making is a fundamental responsibility of the physician and occurs at each stage of the diagnostic and therapeutic process. The decision-making process involves the ordering of additional tests, requests for consultations, decisions about treatment, and predictions concerning prognosis. This process requires an in-depth understanding of the pathophysiology and natural history of disease. Formulating a differential diagnosis requires not only a broad knowledge base but also the ability to assess the relative probabilities of various diseases for a given patient. Application of the scientific method, including hypothesis formulation and data collection, is essential to the process of accepting or ruling out a particular diagnosis. Analysis of the differential diagnosis is an iterative process. As new information or test results are acquired, the group of disease processes being considered can be contracted or expanded appropriately. Whenever possible, decisions should be evidence-based, taking advantage of data from rigorously designed clinical trials or objective comparisons of different diagnostic tests. *Evidence-based medicine* stands in sharp contrast to anecdotal experience, which is often biased. Unless attuned to the importance of using larger, objective studies for making decisions, even the most experienced physicians can be influenced to an undue extent by recent encounters with selected patients. Evidence-based medicine has become an increasingly important part of routine medical practice and has led to the publication of many useful practice guidelines. It is important to remember, however, that only a fraction of the many decisions made in clinical practice are based on rigorous clinical trial evidence; other guideline recommendations are, therefore, predicated on expert consensus and weaker evidentiary support.

Thus, the importance of evidence-based medicine notwithstanding, much medical decision-making still relies on good clinical judgment, an attribute that is difficult to quantify or even to assess qualitatively. Physicians must use their knowledge and experience as a basis for weighing known factors, along with the inevitable uncertainties, and then making a sound judgment; this synthesis of information is particularly important when a relevant evidence base is not available. Several quantitative tools may be invaluable in synthesizing the available information, including diagnostic tests, Bayes' theorem (the probability of an event predicated on prior knowledge of conditions possibly related to the event), and multivariate statistical models **(Chap. 4)**. Diagnostic tests serve to reduce uncertainty about an individual's diagnosis or prognosis and help the physician decide how best to manage that individual's condition. The battery of diagnostic tests complements the history and physical examination. The accuracy of a particular test is ascertained by determining its sensitivity (true-positive rate) and specificity (true-negative rate), as well as the predictive value of a positive and a negative result. **See Chap. 4 for a more thorough discussion of decision-making in clinical medicine.**

Practice Guidelines Many professional organizations and government agencies have developed formal clinical-practice guidelines to aid physicians and other caregivers in making diagnostic and therapeutic decisions that are evidence-based, cost-effective, and most appropriate to a particular patient and clinical situation. As the evidence base of medicine increases, guidelines can provide a useful framework for managing patients with particular diagnoses or symptoms. Clinical guidelines can protect patients—particularly those with inadequate health care benefits—from receiving substandard care. These guidelines also can protect conscientious caregivers from inappropriate charges of malpractice and society from the excessive costs associated with the overuse of medical resources. There are, however, caveats associated with clinical-practice guidelines since they tend to oversimplify the complexities of medicine. Furthermore, groups with different perspectives may develop divergent recommendations regarding issues as basic as the need for screening of women by mammography or of men with serum prostate-specific antigen (PSA) measurements. Finally, guidelines, as the term implies, do not—and cannot be expected to—account for the uniqueness of each individual and their illness. The physician's challenge is to integrate into clinical practice the useful recommendations offered by experts without accepting them blindly or being inappropriately constrained by them.

Precision Medicine The concept of *precision* or *personalized medicine* reflects the growing recognition that diseases once lumped together can be further stratified based on genetic, biomarker, phenotypic, and/or psychosocial characteristics that distinguish a given patient from other patients with similar clinical presentations. Inherent in this concept is the goal of targeting therapies in a more specific way to improve clinical outcomes for the individual patient and minimize unnecessary side effects for those less likely to respond to a particular treatment. In some respects, precision medicine represents the evolution of clinical practice guidelines, which are usually developed for populations of patients or a particular diagnosis (e.g., hypertension, thyroid nodule). As the pathobiology, prognosis, and treatment responses of subgroups within these diagnoses become better understood (i.e., through refined genomic analysis or enhanced deep phenotyping), the relevant clinical guidelines incorporate progressively more refined recommendations for individuals within these subgroups. The role of precision medicine is best illustrated for cancers in which genetic testing is able to predict responses (or the lack thereof) to targeted therapies **(Chap. 78)**. Increasingly, biomarkers, including assessment of specific genetic alterations, are being used to define more precisely subsets of disease and identify therapeutic approaches tailored to the cellular pathobiology. Effective management often relies on the detection of biomarker expression. The validity of such tests is certified by the U.S. Food and Drug administration based on the collection of data from clinical trials. One can anticipate similar applications of precision medicine in pharmacogenomics, immunologic disorders, and diseases in which biomarkers can predict treatment responses. **See Chap. 5 for a more thorough discussion of precision medicine.**

Evaluation of Outcomes Clinicians generally use *objective* and readily measurable parameters to judge the outcome of a therapeutic intervention. These measures may oversimplify the complexity of a clinical condition as patients often present with a major clinical problem in the context of multiple complicating background illnesses. For example, a patient may present with chest pain and cardiac ischemia, but with a background of chronic obstructive pulmonary disease and renal insufficiency. For this reason, outcome measures, such as mortality, length of hospital stay, or readmission rates, are typically risk-adjusted. An important point to remember is that patients usually seek medical attention for *subjective* reasons; they wish to obtain relief from pain, to preserve or regain function, and to enjoy life. The components of a patient's health status or quality of life can include bodily comfort, capacity for physical activity, personal and professional function, sexual function, cognitive function, and overall perception of health. Each of these important domains can be assessed through structured interviews or specially designed questionnaires. Such assessments provide useful parameters by which a physician can judge patients' subjective views of their disabilities and responses to treatment, particularly in chronic illness. The practice of medicine requires consideration and integration of both objective and subjective outcomes.

Many health systems use survey and patient feedback data to assess qualitative features such as patient satisfaction, access to care, and communication with nurses and physicians. In the United States, HCAHPS (Hospital Consumer Assessment of Healthcare Providers and Systems) surveys are used by many systems and are publicly reported. Social media is also being used to assess feedback in real time as well as to share patient experiences with health care systems, potentially enriching the information available for use in medical decisions.

Errors in the Delivery of Health Care A series of reports from the Institute of Medicine (now the National Academy of Medicine [NAM]) called for an ambitious agenda to reduce medical error rates and improve patient safety by designing and implementing fundamental changes in health care systems **(Chap. 7)**. It is the responsibility of hospitals and health care organizations to develop systems to reduce risk and ensure patient safety. Medication errors can be reduced through the use of ordering systems that rely on electronic processes or, when electronic options are not available, that eliminate misreading of handwriting, or that highlight important drug interactions. Whatever the clinical situation, it is the physician's responsibility to use powerful therapeutic measures wisely, with due regard for their beneficial actions, potential dangers, and cost. Implementation of infection control systems, enforcement of hand-washing protocols, and careful oversight of antibiotic use can minimize the complications of nosocomial infections. Central-line infection rates and catheter-associated urinary tract infections have been dramatically reduced at many centers by careful adherence of trained personnel to standardized protocols for introducing, maintaining, and removing central lines and urinary catheters, respectively. Rates of surgical infection and wrong-site surgery can likewise be reduced by the use of standardized protocols and checklists. Falls by patients can be minimized by judicious use of sedatives and appropriate assistance with bed-to-chair and bed-to-bathroom transitions. Taken together, these and other measures are saving thousands of lives each year.

Electronic Medical Records Both the growing reliance on computers and the strength of information technology now play central roles in medicine, including efforts to reduce medical errors. Laboratory data are accessed almost universally through computers. Many medical centers now have electronic medical records (EMRs), computerized order entry, and bar-coded tracking of medications. Some of these systems are interactive, sending reminders or warning of potential medical errors.

EMRs offer rapid access to information that is invaluable in enhancing health care quality and patient safety, including relevant data, historical and clinical information, imaging studies, laboratory results, and medication records. These data can be used to monitor and reduce unnecessary variations in care and to provide real-time information about processes of care and clinical outcomes. Ideally, patient records are easily transferred across the health care system; however, technological limitations and concerns about privacy and cost continue to limit broad-based use of EMRs in many clinical settings.

For all of the advantages of EMRs, they can create distance between the physician and patient if care is not taken to preserve face-to-face contact. EMRs also require training and time for data entry. Many providers spend significant time entering information to generate structured data and to meet billing requirements. They may feel pressured to take short cuts, such as "cutting and pasting" parts of earlier notes into the daily record, thereby increasing the risk of errors. EMRs also structure information in a manner that disrupts the traditional narrative flow across time and among providers. These features, which may be frustrating for some providers, must be weighed against the advantages of ready access to past medical history, imaging, laboratory data, and consultant notes. Furthermore, the effort, time, and attention needed to maintain and utilize the EMR have led to a growing sense of dissatisfaction among physicians, lessening professional and personal well-being as a result. Clearly, this is an area of daily practice that requires improvement both for the delivery of safe and optimal care and physician wellness.

It is important to emphasize that information technology is merely a tool and can never replace the clinical decisions that are best made by the physician. Clinical knowledge and an understanding of a patient's needs, supplemented by quantitative tools, still represent the best approach to decision-making in the practice of medicine.

THE PATIENT–PHYSICIAN RELATIONSHIP

> *The significance of the intimate personal relationship between physician and patient cannot be too strongly emphasized, for in an extraordinarily large number of cases both the diagnosis and treatment are directly dependent on it. One of the essential qualities of the clinician is interest in humanity,* ***for the secret of the care of the patient is in caring for the patient****.*
>
> —Francis W. Peabody, October 21, 1925,
> Lecture at Harvard Medical School

Physicians must never forget that patients are individuals with problems that all too often transcend their physical complaints. They are not "cases" or "admissions" or "diseases." Patients do not fail treatments; treatments fail to benefit patients. This point is particularly important in this era of high technology in clinical medicine. Most patients are anxious and fearful. Physicians should instill confidence and offer reassurance, but they must never come across as arrogant, patronizing, impatient, or hurried. A professional attitude, coupled with warmth and openness, can do much to alleviate anxiety and to encourage patients to share all aspects of their medical history. Empathy and compassion are the essential features of a caring physician. The physician needs to consider the setting in which an illness occurs—in terms not only of patients themselves but also of their familial, social, and cultural backgrounds. The ideal patient–physician relationship is based on thorough knowledge of the patient, mutual trust, and the ability to communicate.

Informed Consent The fundamental principles of medical ethics require physicians to act in the patient's best interest and to respect the patient's autonomy. Both principles are reflected in the process of informed consent. Patients are required to sign consent forms for most diagnostic or therapeutic procedures. Many patients possess limited medical knowledge and must rely on their physicians for advice. Communicating in a clear and understandable manner, physicians must fully discuss the alternatives for care and explain the risks, benefits, and likely consequences of each alternative. The physician is responsible for ensuring that the patient thoroughly understands these risks and benefits; encouraging questions is an important part of this process. It may be necessary to go over certain issues with the patient more than once. This is the very definition of *informed consent*. Complete, clear explanation and discussion of the proposed procedures and treatment can greatly mitigate the fear of the unknown that commonly accompanies hospitalization. Often the patient's understanding is enhanced by repeatedly discussing the issues in an unthreatening and supportive way and answering new questions that occur to the patient as they arise. Continuing efforts to educate the patient are essential. Patients are frequently inhibited from understanding by the fear of an uncertain future and potential impact of the illness on themselves and their families. Clear communication can also help alleviate misunderstandings in situations where complications of intervention occur. Special care should also be taken to ensure that a physician seeking a patient's informed consent has no real or apparent conflict of interest.

Approach to Grave Prognoses and Death No circumstance is more distressing than the diagnosis of an incurable disease, particularly when premature death is inevitable. What should the patient and family be told? What measures should be taken to maintain life? What can be done to optimize quality of life?

Transparency of information, delivered in an appropriate manner, is essential in the face of a terminal illness. Even patients who seem unaware of their medical circumstances, or whose family members have protected them from diagnoses or prognoses, often have keen insights into their condition. They may also have misunderstandings that can lead to additional anxiety. The patient must be given an opportunity to speak with the physician and ask questions. A wise and insightful physician uses such open communication as the basis for assessing what the patient wants to know and when they want to know it. On the basis of the patient's responses, the physician can assess the most appropriate time and pace for sharing information. Ultimately, the patient must understand the expected course of the disease so that appropriate plans and preparations can be made. The patient should participate in decision-making with an understanding of the goal of treatment (palliation) and its likely effects. The patient's religious beliefs should be taken into consideration. Some patients may find it easier to share their feelings about death with their physician, nurses, or members of the clergy than with family members or friends.

The physician should provide or arrange for emotional, physical, and spiritual support, and must be compassionate, unhurried, and open. In many instances, there is much to be gained by the laying on of hands. Pain should be controlled adequately, human dignity maintained, and isolation from family and close friends avoided. These aspects of care tend to be overlooked in hospitals, where the intrusion of life-sustaining equipment can detract from attention to the individual person and encourage concentration instead on the life-threatening disease, against which the battle ultimately will be lost in any case. In the face of terminal illness, the goal of medicine must shift from *cure* to *care* in the broadest sense of the term. *Primum succurrere*, first to help, is a guiding principle. In offering care to a dying patient, a physician should be prepared to provide information to family members and deal with their grief and sometimes their feelings of guilt or even anger. It is important for the physician to assure the family that everything reasonable is being done. A substantial challenge in these discussions is that the physician often does not know exactly how to gauge the prognosis. In addition, various members of the health care team may offer different opinions. Good communication among providers is essential so that consistent information is provided to patients. This is especially important when the best path forward is uncertain. Advice from experts in palliative and terminal care should be sought whenever appropriate to ensure that clinicians are not providing patients with unrealistic expectations. **For a more complete discussion of end-of-life care, see Chap. 13.**

Maintaining Humanism and Professionalism Many trends in the delivery of health care tend to make medical care feel impersonal. These trends, some of which have been mentioned already, include (1) vigorous efforts to reduce the escalating costs of health care; (2) the growing number of managed-care programs, which are intended to reduce costs but where the patient may have little choice in selecting a physician; (3) increasing reliance on technological advances and computerization; (4) the need for numerous physicians and other health professionals to be involved in the care of most patients who are seriously ill; and (5) the growth of telemedicine and virtual encounters that place physical limitations on patient interactions.

In light of these changes in the medical care system, it is a major challenge for physicians to maintain the *humane* aspects of medical care. The American Board of Internal Medicine, working together with the American College of Physicians–American Society of Internal Medicine and the European Federation of Internal Medicine, has published a *Charter on Medical Professionalism* that underscores three main principles in physicians' contract with society: (1) the primacy of patient welfare, (2) patient autonomy, and (3) social justice. While medical schools appropriately place substantial emphasis on professionalism, a physician's personal attributes, including integrity, respect, and compassion, also are extremely important.

Availability to the patient, expression of sincere concern, willingness to take the time to explain all aspects of the illness, and a nonjudgmental attitude when dealing with patients whose cultures, lifestyles, attitudes, and values differ from those of the physician are just a few of the characteristics of a humane physician. Every physician will, at times, be challenged by patients who evoke strongly negative or positive emotional responses. Physicians should be alert to their own reactions to such situations and should consciously monitor and control their behavior so that the patient's best interest remains the principal motivation for their actions at all times. Many organizations use implicit bias training to assist with these skills.

Another important aspect of patient care involves an appreciation of the patient's "quality of life," a subjective assessment of what each patient values most. This assessment requires detailed, sometimes intimate knowledge of the patient, which usually can be obtained only through candid, unhurried, and often repeated conversations. Time pressures will always threaten these interactions, but they should not diminish the importance of understanding and seeking to fulfill the priorities of the patient. The patient encounter should be focused on the patient and not on distractions such as data available in the EMR displayed on the office or bedside computer.

■ EXPANDING FRONTIERS IN MEDICAL PRACTICE

The Era of "Omics" In the spring of 2003, announcement of the complete sequencing of the human genome officially ushered in the genomic era. However, even before that landmark accomplishment, the practice of medicine had been evolving as a result of insights into both the human genome and the genomes of a wide variety of microbes. The clinical implications of these insights are illustrated by the complete genome sequencing of H1N1 influenza virus in 2009 and even faster sequencing of COVID-19 in early 2020, leading to the swift development and dissemination of safe and effective vaccines. Today, gene expression profiles are being used to guide therapy and inform prognosis for a number of diseases, and genotyping is providing a new means to assess the risk of certain diseases as well as variations in response to a number of drugs. Despite these advances, the use of complex genomics in the diagnosis, prevention, and treatment of disease is still in its early stages. The task of physicians is complicated by the fact that phenotypes generally are determined not by genes alone but by the complex interactions among genes and gene products and by the interplay of genetic and environmental factors.

Rapid progress is also being made in other areas of molecular medicine. *Epigenetics* is the study of alterations in chromatin and histone proteins and methylation of DNA sequences (often caused by environmental factors) that influence gene expression **(Chap. 497)**. Every cell of the body has identical DNA sequences; the diverse phenotypes a person's cells manifest are, in part, the result of epigenetic regulation of gene expression. Epigenetic alterations are associated with a number of cancers and other diseases. *Proteomics* is the study of the entire library of proteins made in a cell or organ and the complex relationship of these proteins to disease. This huge repertoire of proteins is generated through alternate splicing of genes, posttranslational processing, and posttranslational modifications that often have unique functional consequences. The presence or absence of particular proteins in the circulation or in cells is being explored for many diagnostic and disease-screening applications. *Microbiomics* is the study of the resident microbes in humans and other mammals, which together compose the microbiome. The human haploid genome has ~23,000 genes, whereas the microbes residing on and in the human body encompass more than 3–4 million genes; these resident microbes are likely to be of great significance with regard to health status of the host. Ongoing research is demonstrating that the microbes inhabiting human mucosal and skin surfaces play a critical role in maturation of the immune system, in metabolic balance, in brain function, and in disease susceptibility. A variety of environmental factors, including the use and overuse of antibiotics, have been tied experimentally to substantial increases in disorders such as obesity, metabolic syndrome, atherosclerosis, and immune-mediated diseases in both adults and children. *Metagenomics*, of which microbiomics is a part, is the genomic study of environmental species that have the potential to influence human biology directly or indirectly. An example is the study of exposures to microorganisms in farm environments that may be responsible for the lower incidence of asthma among children raised on farms. *Metabolomics* is the study of the range of metabolites in cells or organs and the ways they are altered in disease states. The aging process itself may leave telltale metabolic footprints that allow the prediction (and possibly the prevention) of organ dysfunction and disease. It seems likely that disease-associated patterns will be found in lipids, carbohydrates, membranes, mitochondria and mitochondrial function, and other vital components of cells and tissues. *Exposomics* is the study of the exposome—i.e., the environmental exposures such as smoking, sunlight, diet, exercise, education, and violence that together have an enormous impact on health. All of this new information represents a challenge to the traditional reductionist approach to medical thinking. The variability of results in different patients, together with the large number of variables that can be assessed, creates challenges in identifying preclinical disease and defining disease states unequivocally. Accordingly, the tools of *systems biology* and *network medicine* are being applied to the enormous body of information ("big data") now obtainable for every patient and may eventually provide new approaches to classifying, diagnosing, treating, and preventing disease. **For a more complete discussion of a complex systems and network science approach to human disease, see Chap. 499.**

The rapidity of these advances may seem overwhelming to practicing physicians; however, physicians have an important role to play in ensuring that these powerful technologies and sources of new information are applied judiciously to patient care. Since omics are evolving so rapidly, physicians and other health care professionals must engage in continuous learning so that they can apply this new knowledge to the benefit of their patients' health and well-being. Genetic testing requires wise counsel based on an understanding of the value and limitations of the tests as well as the implications of their results for specific individuals. **For a more complete discussion of genetic testing, see Chap. 480.**

The Globalization of Medicine Physicians should be cognizant of diseases and health care services beyond local boundaries. Global travel has critical implications for disease spread, and it is not uncommon for diseases endemic to certain regions to be seen in other regions after a patient has traveled to and returned from those regions. The outbreak of Zika virus infections in the Americas is a cogent example of this phenomenon. In addition, factors such as wars, the migration of refugees, travel patterns, and the impact of increasing climate extremes are contributing to changing disease profiles worldwide. Patients have broader access to unique expertise or clinical trials at distant medical centers, even those in other countries, and the cost of travel may be offset by the quality of care at those distant locations. As much as any other factor influencing global aspects of medicine, the Internet has transformed the transfer of medical information throughout the world. This change has been accompanied by the transfer of technological skills through telemedicine and international consultation—for example, interpretation of radiologic images and pathologic specimens. **For a complete discussion of global issues, see Chap. 485.**

Medicine on the Internet On the whole, the Internet has had a positive effect on the practice of medicine; through personal computers, a wide range of information is available to physicians and patients almost instantaneously at any time and from anywhere in the world. This medium holds enormous potential for the delivery of current information, practice guidelines, state-of-the-art conferences, journal content, textbooks (including this text), and direct communications with other physicians and specialists, expanding the depth and breadth of information available to the physician regarding the diagnosis and care of patients. Medical journals are now accessible online, providing rapid sources of new information. By bringing them into direct and timely contact with the latest developments in medical care, this medium also serves to lessen the information gap that has hampered physicians and health care providers in remote areas.

Patients, too, are turning to the Internet in increasing numbers to acquire information about their illnesses and therapies and to join Internet-based support groups. Patients often arrive at a clinic visit with sophisticated information about their illnesses. In this regard, physicians are challenged in a positive way to keep abreast of the latest relevant information while serving as an information "editor" as patients navigate this seemingly endless source of information, the accuracy and validity of which are not uniform.

A critically important caveat is that virtually anything can be published on the Internet, with easy circumvention of the peer-review process that is an essential feature of academic publications. Both physicians and patients who search the Internet for medical information must be aware of this danger. We have recently seen the adverse consequences of misinformation and disinformation (often in service of some political agenda) in social media. It is difficult to calculate the harm done by the widespread deceptions that prevented perhaps millions of people from accepting life-saving preventions like the COVID-19 vaccine. Notwithstanding this limitation, appropriate use of the Internet is revolutionizing information access for physicians and patients, and in this regard, the Internet represents a remarkable resource that was not available to practitioners a generation ago.

Public Expectations and Accountability The general public's level of knowledge and sophistication regarding health issues has grown rapidly over the past few decades. As a result, expectations of the health care system in general and of physicians in particular have risen. Physicians are expected to master rapidly advancing fields (the *science* of medicine) while considering their patients' unique needs (the *art* of medicine). Thus, physicians are held accountable not only for the technical aspects of the care they provide but also for their patients' satisfaction with the delivery and costs of care.

In many parts of the world, physicians increasingly are expected to account for the way in which they practice medicine by meeting certain standards prescribed by federal and local governments. The hospitalization of patients whose health care costs are reimbursed by the government and other third parties is subjected to utilization review. Thus, a physician must defend the cause for and duration of a patient's hospitalization if it falls outside certain "average" standards. Authorization for reimbursement increasingly is based on documentation of the nature and complexity of an illness, as reflected by recorded elements of the history and physical examination. A growing "pay-for-performance" movement seeks to link reimbursement to quality of care. The goal of this movement is to improve standards of health care and contain spiraling health care costs. In many parts of the United States, managed (capitated) care contracts with insurers have replaced traditional fee-for-service care, placing the onus of managing the cost of all care directly on the providers and increasing the emphasis on preventive strategies. In addition, physicians are expected to give evidence of their current competence through mandatory continuing education, patient record audits, maintenance of certification, and relicensing.

Medical Ethics and New Technologies The rapid pace of technological advances has profound implications for medical applications that go far beyond the traditional goals of disease prevention, treatment, and cure. Cloning, gene therapy, gene editing, human–computer interfaces, nanotechnology, and use of targeted therapies have the potential to modify inherited predispositions to disease, select desired characteristics in embryos, augment "normal" human performance, replace failing tissues, and substantially prolong life span. Given their unique training, physicians have a responsibility to help shape the debate on the appropriate uses of and limits placed on these new technologies and to consider carefully the ethical issues associated with the implementation of such interventions. As medicine becomes more complex, shared decision-making is increasingly important, not only in areas such as genetic counseling and end-of-life care but also in diagnostic and treatment options.

Learning Medicine More than a century has passed since the publication of the Flexner Report, a seminal study that transformed medical education and emphasized the scientific foundations of medicine as well as the acquisition of clinical skills. In an era of burgeoning information and access to medical simulation and informatics, many schools are implementing new curricula that emphasize lifelong learning and the acquisition of competencies in teamwork, communication skills, system-based practice, and professionalism. The tools of medicine also change continuously, necessitating formal training in the use of EMRs, large datasets, ultrasound, robotics, and new imaging techniques. These and other features of the medical school curriculum provide the foundation for many of the themes highlighted in this chapter and are expected to allow physicians to progress, with experience and learning over time, from competency to proficiency to mastery.

At a time when the amount of information that must be mastered to practice medicine continues to expand, increasing pressures both within and outside of medicine have led to the implementation of restrictions on the amount of time a physician-in-training can spend in the hospital and in clinics. Because the benefits associated with continuity of medical care and observation of a patient's progress over time were thought to be outstripped by the stresses imposed on trainees by long hours and by fatigue-related errors, strict training parameters were introduced by the Accreditation Council for Graduate Medical Education (ACGME) to reduce fatigue and enhance learning. The impact of these changes is still being assessed, and there are continual efforts to optimize training and move toward competency-based endpoints. An unavoidable by-product of fewer hours at the bedside is an increase in the number of "handoffs" of patient responsibility from one physician to another. It is imperative that these transitions of responsibility be handled with care and thoroughness, with all relevant information exchanged and acknowledged. These issues highlight the challenge our profession has in establishing a reliable measure of physician effectiveness.

The Physician as Perpetual Student From the time physicians graduate from medical school, it becomes all too apparent that this milestone is symbolic and that they must embrace the role of a "perpetual student." This realization is at the same time exhilarating and anxiety-provoking. It is exhilarating because physicians can apply constantly expanding knowledge to the treatment of their patients; it is anxiety-provoking because physicians realize that they will never know as much as they want or need to know. Ideally, physicians will translate the latter feeling into energy through which they can continue to improve and reach their potential. It is the physician's responsibility to pursue new knowledge continually by reading, attending conferences and courses, and consulting colleagues and the Internet. This is often a difficult task for a busy practitioner; however, a commitment to continued learning is an integral part of being a physician and must be given the highest priority.

The Physician as Citizen Being a physician is a privilege. The capacity to apply one's skills for the benefit of fellow human beings is a noble calling. The physician–patient relationship is inherently unbalanced in the distribution of power. In light of their influence, physicians must always be aware of the potential impact of what they do and say and must always strive to strip away individual biases and preferences to find what is best for their patients. To the extent possible, physicians should also act within their communities to promote health and alleviate suffering. Meeting these goals begins by setting a healthy example and continues in taking action to deliver needed care even when personal financial compensation may not be available.

Research, Teaching, and the Practice of Medicine The word *doctor* is derived from the Latin *docere*, "to teach." As teachers, physicians should share information and medical knowledge with colleagues, students of medicine and related professions, and their patients. The practice of medicine is dependent on the sum total of medical knowledge, which in turn is based on an unending chain of scientific discovery, clinical observation, analysis, and interpretation. Advances in medicine depend on the acquisition of new information through research, and improved medical care requires the transmission of that information. As part of their broader societal responsibilities, physicians should encourage patients to participate in ethical and properly approved clinical investigations if these studies do not impose undue hazard, discomfort, or inconvenience. Physicians engaged in clinical research must be alert to potential conflicts of interest between their research goals and their obligations to individual patients. The best interests of the patient must always take priority.

> *To wrest from nature the secrets which have perplexed philosophers in all ages, to track to their sources the causes of disease, to correlate the vast stores of knowledge, that they may be quickly available for the prevention and cure of disease—these are our ambitions.*
>
> —William Osler, 1849–1919

FURTHER READING

Cheston CC et al: Social media use in medical education: A systematic review. Acad Med 88:893, 2013.

Cooke M et al: American medical education 100 years after the Flexner report. N Engl J Med 355:1339, 2006.

Excel JL et al: Vaccine development for emerging infectious diseases. Nat Med 27:591, 2021.

Institute of Medicine: *Dying in America: Improving Quality and Honoring Individual Preferences Near the End of Life*. Washington, DC, National Academies Press, 2015.

Institute of Medicine: *Improving Diagnosis in Health Care*. Washington, DC, National Academies of Sciences, Engineering, and Medicine, 2015.

Kohane IS: Injecting artificial intelligence into medicine. NEJM AI 1, 2023.

Levine DM et al: Hospital-level care at home for acutely ill adults: A qualitative evaluation of a randomized controlled trial. J Gen Intern Med 36:1965, 2021.

Romanella M et al: The 2022 report of the Lancet Countdown on health and climate change: Health at the mercy of fossil fuels. Lancet 400:1619, 2022.

Stern DT, Papadakis M: The developing physician—becoming a professional. N Engl J Med 355:1794, 2006.

West P et al: Intervention to promote physician well-being, job satisfaction, and professionalism. A randomized clinical trial. JAMA Intern Med 174:527, 2014.

2

Promoting Good Health

Donald M. Lloyd-Jones,
Kathleen M. McKibbin

GOALS AND APPROACHES TO PREVENTION

Prevention of acute and chronic diseases before their onset has been recognized as one of the hallmarks of excellent medical practice for centuries and is now used as a metric for highly functioning health care systems. The ultimate goal of preventive strategies is to avoid premature death. However, as longevity has increased dramatically worldwide over the last century (largely as a result of public health practices), increasing emphasis is placed on prevention for the purpose of preserving quality of life and extending the health span, not just the life span. Given that all patients will eventually die, the goal of prevention ultimately becomes compression of morbidity toward the end of the life span; that is, reduction of the amount of burden and time spent with disease prior to dying. As shown in **Fig. 2-1**, normative aging tends to involve a steady decline in the stock of health, with accelerating decline over time. Successful prevention offers the opportunity both to extend life and to extend healthy life, thus "squaring the curve" of health loss during aging.

Prevention strategies have been characterized as tertiary, secondary, primary, and primordial. *Tertiary prevention* requires rapid action to prevent imminent death and preserve organ function in the setting of acute illness, such as through thrombolysis or thrombectomy in acute ischemic stroke. *Secondary prevention* strategies focus on avoiding the recurrence of disease and death in an individual who is already affected. For example, tamoxifen is recommended for women with surgically treated early-stage, estrogen receptor–positive breast cancer, because it reduces the risk of recurrent breast cancer (including in the contralateral breast) and death. *Primary prevention* attempts to reduce the risk of incident disease among individuals with one or more risk factors. Treatment of elevated blood pressure in individuals who have not yet experienced cardiovascular disease represents one example of primary prevention that has proven effective in reducing the incidence of stroke, heart failure, and coronary heart disease.

Primordial prevention is a more recent concept (first introduced in 1979) that focuses on prevention of the development of *risk factors* for disease, not just prevention of disease. Primordial prevention strategies emphasize upstream determinants of risk for chronic diseases, such as eating patterns, physical activity, and environmental and social determinants of health. It therefore encompasses medical treatment strategies for some individuals as well as a strong reliance on public health and social policy. It is increasingly clear that primordial prevention represents the ultimate means for reducing the burden of chronic

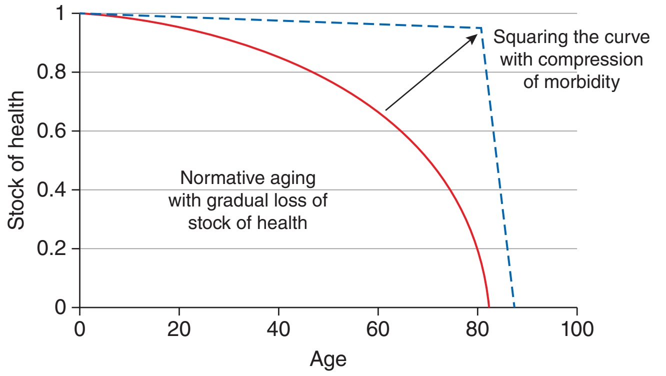

FIGURE 2-1 Loss of health with aging. Representation of normative aging with loss of the full stock of health with which individuals are born (indicating gain of morbidity), contrasted with a squared curve with greater longevity and fuller stock of health (less morbidity) until shortly before death. The "squared curve" represents the likely ideal situation for most patients.

diseases of aging. Once risk factors develop, it is difficult to restore risk to the low level of someone who never developed the risk factor. The time spent with adverse levels of the risk factor often causes irreversible damage that precludes complete restoration of low risk. For example, individuals with hypertension who are treated back to optimal levels (<120/<80 mmHg) do have a lower risk compared with untreated patients with hypertension, but they still have twice the risk of cardiovascular events as those who maintained optimal blood pressure without medications. Patients with elevated blood pressure that is subsequently treated have greater left ventricular mass index, worse renal function, and more evidence of atherosclerosis and other target organ damage as a result of the time spent with elevated blood pressure; such damage cannot be fully reversed despite efficacious therapy with antihypertensive medications. Conversely, as described below in greater detail, individuals who maintain optimal levels of all major cardiovascular risk factors into middle age through primordial prevention essentially abolish their lifetime risk of developing cardiovascular disease while also living substantially longer and having a lower burden and later onset of other chronic comorbid illnesses (compression of morbidity).

Prevention strategies should be distinguished from disease-screening strategies. Screening attempts to detect evidence of disease at its earliest stages, when treatment is likely to be more efficacious than for advanced disease **(Chap. 6)**. Screening can be performed in service of prevention, especially if it aids in identifying preclinical markers, such as dyslipidemia or hyperglycemia, associated with elevated disease risk.

HEALTH PROMOTION

In recent decades, medical practice has increasingly focused on clinical and public health approaches to promote health and not just prevent disease. Prevention of disease is a worthy individual and societal goal in and of itself, but it does not necessarily guarantee health. Health is a broader construct encompassing more than just absence of disease. It includes biologic, physiologic, and psychological domains (among others) in a continuum, rather than occurring as a dichotomous trait. Health is therefore somewhat subjective, but attempts have been made to use more objective criteria to define health in order to raise awareness, prevent disease, and promote healthy longevity.

For example, in 2010, the American Heart Association (AHA) formally defined a new construct of "cardiovascular health" based on evidence of associations with longevity, disease avoidance, healthy longevity, and quality of life. The concept was updated and expanded in 2022 to encompass eight metrics that help define an individual's or population's current health status. The eight metrics (termed Life's Essential 8™) include diet, physical activity, sleep, nicotine exposure, body mass index, blood lipids, glycemia, and blood pressure. Each metric is scored on a scale of 0 to 100 points (higher is better), and overall health is measured as the average of the scores (also from 0 to 100). Higher cardiovascular health scores at all ages have been associated with greater longevity, lower incidence of cardiovascular disease, lower incidence of other chronic diseases of aging (including dementia, cancer, diabetes, and more), compression of morbidity, greater quality of life, and lower health care costs, achieving both individual and societal goals for healthy aging and further establishing the critical importance of primordial prevention and health promotion.

Focusing on health promotion, rather than just disease prevention, may also provide greater motivation for patients to pursue lifestyle changes or adhere to clinician recommendations. Extensive literature suggests that providing patients solely with information regarding disease risk, or risk reduction with treatment, is unlikely to motivate desired behavior change. Empowering patients with strategies to achieve positive health goals after discussing risks can provide more effective adherence and better long-term outcomes. In the case of smoking cessation, enumerating only the risks of smoking can lead to patient inertia and therapeutic nihilism and has proven to be an ineffective approach, whereas strategies that incorporate positive health messaging, support, and feedback, with appropriate use of evidence-based therapies, have proven far more effective.

PRIORITIZING PREVENTION STRATEGIES

In secondary prevention, the patient already has manifest clinical disease and is therefore at high risk for progression. The approach should be to work with the patient to implement all evidence-based strategies that will help to prevent recurrence or progression. This will typically include drug therapy as well as therapeutic lifestyle changes to control ongoing risk factors that may have caused disease in the first place. Juggling priorities can be difficult, and barriers to implementation are many, including costs, time, patient health literacy, and patient and caregiver capacity to organize the regimen. Addressing these potential barriers with the patient can help to forge a therapeutic bond and may improve adherence; ignoring them will likely lead to therapeutic failure. Numerous studies demonstrate that, even in high-functioning health systems, only ~50% of patients are taking recommended, evidence-based secondary prevention medications, such as statins, by 1 year after a myocardial infarction.

In patients who are eligible for primary prevention strategies, it is important to frame the discussion around the overall evidence base as well as an individual patient's likelihood of benefit from a given preventive intervention. A first step is to understand the patient's estimated absolute risk for disease in the foreseeable future or during their remaining life span. However, absolute risk estimation and presentation of those risks are generally insufficient to motivate behavior change. It is critical to assess the patient's understanding and tolerance of the risk, their readiness to implement lifestyle changes or adhere to drug therapy, and their overall preferences regarding use of drug therapy to prevent an event (e.g., cancer, myocardial infarction, stroke). The clinician can help the patient by informing them of the risks for disease and potential for absolute benefits (and harms) from the available evidence-based choices. This may take more than one conversation, but given that diseases, such as cancer and cardiovascular disease, are the leading causes of premature death and disability, the time is well spent.

Partnering with the patient through motivational interviewing may assist in the process of selecting initial approaches to prevention. Selecting an area that the patient feels they are ready to change can lead to better adherence and greater achievement of success in the short and longer term. If the patient is uncertain what course to choose, prudence would dictate focusing on control of risk factors that may lead to the most rapid reduction in risk for acute events. For example, blood pressure is both a chronic risk factor and an acute trigger for cardiovascular events. Thus, if a patient has both significant elevations in blood pressure and dyslipidemia, it would be appropriate to focus initial efforts on blood pressure control. Likewise, a focus on smoking cessation can lead to more rapid reductions in risk for acute events than some other lifestyle interventions.

PREVENTION AND HEALTH PROMOTION ACROSS THE LIFE COURSE

Periodic Health Evaluations The "routine annual physical" has in many ways become an expected part of the patient-physician relationship in primary care practice. However, evidence for the efficacy of the periodic health evaluation in asymptomatic adults unselected for risk factors or disease is mixed and depends on the study design and outcome. Systematic reviews and meta-analyses of published randomized trials have consistently observed lack of benefit (and also lack of harm) in terms of total mortality in association with periodic health evaluations. Data are more heterogeneous but overall suggest no benefit for cancer- or cardiovascular-specific mortality, with the potential for either benefit or harm depending on number of evaluations and patient-level factors. Well-designed studies on nonfatal clinical events and morbidity have been sparsely reported, but there appear to be no large effects.

Periodic health evaluations do appear to lead to greater diagnosis of certain conditions such as hypertension and dyslipidemia, as expected. Likewise, periodic health examinations also improve the delivery of recommended preventive services, such as gynecologic examinations and Papanicolaou smears, fecal occult blood testing, and cholesterol screening. The benefits and risks associated with screening tests are discussed in detail in Chap. 6. Risks of routine evaluations include inappropriate testing or overtesting, or false-positive findings that require follow-up and induce patients to worry. Periodic health examinations appear to be associated with less patient worry. On balance, given the lack of convincing evidence of harm and the potential for better delivery of appropriate screening, counseling, and preventive services, periodic health evaluations appear reasonable for general populations at average risk for chronic conditions.

It is important to note that routine annual comprehensive *physical* examinations of asymptomatic adult patients have very low yield and may take an inordinate amount of time in a wellness visit. Such time may be better spent on assessing and counseling the patient on other aspects of their health, as discussed below. Evidence-based components that should be included in periodic evaluations focused on health and prevention include a number of age-appropriate screening tests for chronic disease and risk factors, preventive interventions including immunizations and chemoprevention for at-risk individuals, and preventive counseling. The U.S. Preventive Services Task Force publishes its *Guide to Clinical Preventive Services*, which contains evidence-based recommendations from the Task Force on preventive services for which there is a high degree of certainty that the service provides at least moderate net clinical benefit (i.e., benefits outweigh harms significantly and to a reasonable magnitude).

Healthy Behaviors and Lifestyles Owing to the heterogeneity of study designs and the diverse nature of lifestyle interventions studied, many clinicians are uncertain as to how to deliver advice regarding healthy behaviors and lifestyles. Nevertheless, adverse behaviors and lifestyles contribute to >75% of premature, preventable deaths and disability. Estimates from the U.S. National Health and Nutrition Examination Survey indicate that fewer than 1% of Americans achieve an optimal heart-healthy eating pattern. Thus, whereas there are many demands on time during a typical patient-clinician encounter, few things may have more impact on longevity, health, and quality of life for asymptomatic patients than an efficient approach to assessing, documenting, and improving patients' health behaviors. Indeed, the mere act of assessing health behaviors has been shown to affect patients' health behaviors. Facility with tools for assessment of lifestyle and with strategies for counseling are therefore of paramount importance.

Healthy Eating Patterns (see Chap. 343) Despite the existence of numerous "fad" diets and seemingly inconsistent recommendations on dietary composition, there is remarkable agreement about what should constitute a healthy eating pattern for the broad population to avoid nutritional deficits (i.e., vitamin deficiency) and excesses (i.e., excessive caloric intake) and to maximize potential health (Table 2-1). Optimal eating patterns consist of whole fruits and vegetables, whole grains, lean proteins, and healthy oils, and allow for nonfat or low-fat dairy intake. They tend to exclude frequent ingestion of foods high in refined sugars and starches, saturated fat, and sodium. Since sodium and refined sugars and starches are the hallmark of much of the processed/packaged food supply, a simple rule of thumb is to provide or cook the majority of one's own meals starting from whole foods and emphasizing fruits and vegetables. Likewise, foods prepared outside of the home tend to have higher fat and sodium content, so special attention to menu choices focused on fruits, vegetables, lean proteins, and whole grains, while minimizing sauces and dressings, can help most individuals follow healthier eating patterns when eating food prepared outside the home. In all cases, sugar-sweetened beverages and nonnutritious snack foods should be minimized. If snacks are included, small amounts of healthy nuts and seeds or more fruits and vegetables should be encouraged.

Specific conditions and diseases, such as diabetes or hypertension, other metabolic disorders, allergies, and gastrointestinal disorders, may require tailored approaches to diet. In counseling most patients,

TABLE 2-1 Guidelines and Key Recommendations from the *Dietary Guidelines for Americans*, 2020–2025

GUIDELINES	KEY RECOMMENDATIONS
1. **Follow a healthy dietary pattern at every life stage.** For the first 6 months of life, infants should exclusively be fed human milk, or iron-fortified formula if human milk is unavailable. From 6 to 12 months, infants should be introduced to a variety of complementary nutrient-dense foods. From 12 months to older adulthood, the dietary pattern should meet nutrient needs, help achieve a healthy body weight, and reduce the risk of chronic disease. 2. **Customize and enjoy nutrient-dense food and beverage choices to reflect personal preferences, cultural traditions, and budgetary considerations.** The Dietary Guidelines provide a framework of several dietary patterns intended to be customized to individual needs and preferences, as well as the foodways of the diverse cultures in the United States. 3. **Focus on meeting food group needs with nutrient-dense foods and beverages and stay within calorie limits.** Nutrient-dense foods provide vitamins, minerals, and other health-promoting components and have no or little added sugars, saturated fat, and sodium. A healthy dietary pattern consists of nutrient-dense forms of foods and beverages across all food groups, in recommended amounts, and within calorie limits. 4. **Limit foods and beverages higher in added sugars, saturated fat, and sodium, and limit alcoholic beverages.** At every life stage, meeting food group recommendations, even with nutrient-dense choices, fulfills most of a person's daily calorie needs and sodium limits, with little room for extra added sugars, saturated fat, or sodium, or for alcoholic beverages.	**The Dietary Guidelines' Key Recommendations for healthy eating patterns should be applied in their entirety, given the interconnected relationship that each dietary component can have with others. They are also intended as a framework to accommodate personal preferences, cultural traditions, and budgetary considerations.** **Focus on meeting food group needs with nutrient-dense foods and beverages, and stay within calorie limits to achieve a healthy weight and reduce the risk of chronic disease.** **The core elements that make up a healthy dietary pattern include:** • Vegetables of all types—dark green; red and orange; beans, peas, and lentils; starchy; and other vegetables • Fruits, especially whole fruit • Grains, at least half of which are whole grain • Dairy, including fat-free or low-fat milk, yogurt, and cheese, and/or lactose-free versions and fortified soy beverages and yogurt as alternatives • Protein foods, including lean meats, poultry, and eggs; seafood; beans, peas, and lentils; and nuts, seeds, and soy products • Oils, including vegetable oils and oils in food, such as seafood and nuts **A healthy eating pattern limits:** • Added sugars—Less than 10% of calories per day starting at age 2. Avoid foods and beverages with added sugars for those younger than age 2. • Saturated fat—Less than 10% of calories per day starting at age 2. • Sodium—Less than 2300 mg/d—and even less for children younger than age 14. • Alcoholic beverages—Adults of legal drinking age can choose not to drink or to drink in moderation by limiting intake to 2 drinks or less in a day for men and 1 drink or less in a day for women, when alcohol is consumed. Drinking less is better for health than drinking more. There are some adults who should not drink alcohol, such as women who are pregnant. **Meet the U.S. Department of Health and Human Services' *Physical Activity Guidelines for Americans*** In tandem with the recommendations above, Americans of all ages—children, adolescents, adults, and older adults—should meet the *Physical Activity Guidelines for Americans* to help promote health and reduce the risk of chronic disease. Americans should aim to achieve and maintain a healthy body weight. The relationship between diet and physical activity contributes to calorie balance and managing body weight.

Source: Adapted from the *Dietary Guidelines for Americans, 2020-2025.* Washington, DC: U.S. Department of Agriculture and U.S. Department of Health and Human Services; 2020. Available at *https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf.*

the general approach should focus on whole foods, eating patterns, and appropriate calorie balance, rather than on specific micronutrients such as electrolytes or selected vitamins. It should be remembered that most patients have difficulty understanding nutritional labels on packaged foods, with the attendant demands on numeracy and health literacy.

Dietary guidelines are published by the U.S. Department of Agriculture (USDA) and U.S. Department of Health and Human Services every 5 years, and these guidelines have undergone substantial evolution over time. The current U.S. Dietary Guidelines and Key Recommendations for 2020–2025 are summarized in Table 2-1 and emphasize the importance of healthy eating patterns for every stage of life, to avoid chronic diseases including obesity, diabetes, cancer, and cardiovascular disease. The core elements include eating patterns with nutrient-dense (rather than calorie-dense) whole foods and appropriate caloric intake to achieve and maintain healthy weight. The USDA guidelines focus on the concept of a healthy plate (rather than the prior food pyramid) for ease of counseling and adoption. Fifty percent of the plate should consist of vegetables and whole fruits, with remaining portions for whole grains and lean protein foods. When using fat for cooking, it should be done by sauteing in healthier oils (e.g., canola oil), and addition of judicious amounts of healthy raw oils (e.g., olive oil, nuts) to dishes is appropriate. Recommendations also focus on limitation of foods and beverages higher in added sugars, saturated fat, and sodium, and moderation or avoidance of alcohol intake.

The USDA guidelines focus on specific healthy eating patterns that adhere to these broad recommendations and are appropriate for ~97% of the general population. They identify a “Healthy U.S.-Style Dietary Pattern” that adheres closely to the evidence-based Dietary Approaches to Stop Hypertension (DASH) eating pattern but is customizable for different cultural or personal preferences. Alternative patterns, which vary more in emphasis than in content, include a “Healthy Mediterranean-Style Dietary Pattern” and a “Healthy Vegetarian Dietary Pattern.”

AGE- AND SEX-SPECIFIC RECOMMENDATIONS Current dietary framework recommendations are generally similar for all life stages from ages ≥12 months, but recommended levels of caloric intake (and hence amounts of foods) differ by age, sex, and physical activity level. For example, recommended caloric intake ranges from 1000 calories/d for sedentary 2-year-old children to as high as 3200 calories/d for active 16- to 18-year-old young men. Recommended caloric intakes peak in late adolescence or early adulthood for men and women and gradually decrease over ensuing decades.

As with all lifestyle counseling aimed at behavior change, dietary approaches that partner with the patient and utilize motivational interviewing strategies and shared goals and commitments tend to work best, as described below (see “Approach to the Patient”).

Physical Activity Similar to the approach to counseling regarding healthy eating patterns, recommendations on participation in physical activity emphasize the point that any physical activity is better than none. A simple rule of thumb for patients is: “If you are doing nothing, do something; and if you are doing something, do more, every day.” The evidence base for physical activity indicates that the marginal benefits from physical activity are greatest in advancing from no activity to low levels of moderate activity. With increasing duration and intensity of activity, there is a continued curvilinear increase in health benefits, but the marginal gains for each additional minute of moderate-to-vigorous activity slowly diminish. Thus, for adults, the recommended amount of physical activity is 150 min of moderate-intensity or 75 min of vigorous-intensity aerobic activity per week, performed in episodes of at least 5 min, and preferably spread throughout the week, plus participation in muscle-strengthening activity at least 2 days per week. Additional health benefits can be realized by engaging in physical activity beyond this amount.

In counseling patients regarding physical activity, it is important to note that sedentary time (e.g., seated at work or at home in front of electronic screens) has adverse health consequences independent of the lack of physical activity during these episodes. Therefore, even modest efforts like standing at the desk and doing gentle stretching for periods during the day may be beneficial. It is also important to emphasize that participating in a variety of aerobic activities (biking, swimming, walking, jogging, rowing, elliptical training, stair-climbing, etc.) can be beneficial and may help to avoid overuse injuries and boredom with the exercise regimen. Addition of resistance (muscle-strengthening) activities is also beneficial for health improvement. Emphasis should be placed on body-weight resistance or weights that allow more repetitions (e.g., 3 sets of 15–20 repetitions that can be performed comfortably, with a rest period in between) and on avoiding breath-holding and straining against a closed glottis.

SUDDEN CARDIAC DEATH RISK Patients may express concerns regarding the risk of sudden cardiac death during exercise. Whereas the risk of sudden death during exercise does increase directly with the amount of time spent exercising, this association is substantially mitigated by training effects. Thus, patients embarking on an exercise program should be encouraged to increase the duration of aerobic exercise gradually as tolerated, aiming for episodes of at least 30 min 5 times a week as an ideal. Once a comfortable duration is reached, incorporating interval training periods of more intensive activity interspersed during the exercise can provide greater fitness gains.

EXTREME ENDURANCE ACTIVITIES As with other forms of exercise, extreme endurance activities such as triathlons and marathons should be undertaken only with appropriate and graded training. Such activities tend to take a greater toll on the musculoskeletal system over time than less extreme activities, and they are also associated with measurable damage to the myocardium and greater risks for other organ damage. Athletes participating in endurance activities routinely have elevations in cardiac troponin (a specific circulating marker of myocardial cell damage and death) at the end of the race, although elevations are lower in those who are well trained. Patients and clinicians should consider the patient’s overall health, specific limitations, potential for injury, and ability to train in decision-making regarding participation in endurance events.

AGE-SPECIFIC RECOMMENDATIONS The U.S. Department of Health and Human Services’ *Physical Activity Guidelines for Americans*, second edition (2018) **(Table 2-2)**, recommend that preschool-aged children (aged 3–5 years) should be physically active throughout the day in a variety of activity types to enhance growth and development. Children and adolescents aged 6–17 years should participate in ≥60 min of physical activity daily, most of which should be moderate- or vigorous-intensity aerobic activity, including vigorous, muscle-strengthening, and bone-strengthening activities at least 3 days a week each. As noted above, adults aged 18–64 years are recommended to pursue at least 150 min of moderate-intensity or 75 min of vigorous-intensity aerobic activity per week (or equivalent combinations), with at least 2 days of muscle-strengthening activities. Adults aged ≥65 years should follow the adult guidelines or be as active as possible as abilities and conditions allow. For older adults, special emphasis is also placed on multicomponent physical activity that includes balance training as well as aerobic and muscle-strengthening activities.

Sleep Hygiene Sleeping between 7 and 9 h per night appears to be optimal for health in adults aged ≥18 years. Sleeping <7 h is associated with adverse outcomes, including obesity, diabetes, elevated blood pressure, cardiovascular disease, depression, and all-cause mortality, as well as physiologic disturbances such as impaired immune function, increased pain sensitivity, and impaired cognitive performance. Conversely, achieving appropriate levels of sleep is associated with more success in weight loss, better blood pressure control among patients with hypertension, and improved mental health and performance. Regular sleep more than 9 h per night is appropriate for children and adolescents or individuals recovering from sleep deprivation or illness, but for most individuals, the effects on health are uncertain.

Patients often express concerns about the quantity and quality of their sleep. With aging, both aspects of sleep tend to decline, even without overt sleep disorders. Documentation of sleep using a sleep log may assist in understanding different types of insomnia and sleep disorders.

TABLE 2-2 Recommendations from *Physical Activity Guidelines for Americans*, 2nd Edition (2018)

AGE	RECOMMENDATIONS
3–5 years	• Preschool-aged children (ages 3 through 5 years) should be physically active throughout the day to enhance growth and development. • Adult caregivers of preschool-aged children should encourage active play that includes a variety of activity types.

6–17 years	• It is important to provide young people opportunities and encouragement to participate in physical activities that are appropriate for their age, that are enjoyable, and that offer variety.
	• Children and adolescents ages 6 through 17 years should do 60 min (1 h) or more of moderate-to-vigorous physical activity daily:
	  • Aerobic: Most of the 60 min or more per day should be either moderate- or vigorous-intensity aerobic physical activity and should include vigorous-intensity physical activity on at least 3 days a week.
	  • Muscle-strengthening: As part of their 60 min or more of daily physical activity, children and adolescents should include muscle-strengthening physical activity on at least 3 days a week.
	  • Bone-strengthening: As part of their 60 min or more of daily physical activity, children and adolescents should include bone-strengthening physical activity on at least 3 days a week.

18–64 years	• Adults should move more and sit less throughout the day. Some physical activity is better than none. Adults who sit less and do any amount of moderate-to-vigorous physical activity gain some health benefits. • For substantial health benefits, adults should do at least 150 min (2 h and 30 min) to 300 min (5 h) a week of moderate-intensity or 75 min (1 h and 15 min) to 150 min (2 h and 30 min) a week of vigorous-intensity aerobic physical activity, or an equivalent combination of moderate- and vigorous-intensity aerobic activity. Preferably, aerobic activity should be spread throughout the week. • Additional health benefits are gained by engaging in physical activity beyond the equivalent of 300 min (5 h) of moderate-intensity physical activity a week. • Adults should also do muscle-strengthening activities of moderate or greater intensity and that involve all major muscle groups on 2 or more days a week, as these activities provide additional health benefits.

≥65 years	• The key guidelines for adults also apply to older adults. In addition, the following key guidelines are just for older adults:
	  • As part of their weekly physical activity, older adults should do multicomponent physical activity that includes balance training as well as aerobic and muscle-strengthening activities.
	  • Older adults should determine their level of effort for physical activity relative to their level of fitness.
	  • Older adults with chronic conditions should understand whether and how their conditions affect their ability to do regular physical activity safely.
	  • When older adults cannot do 150 min of moderate-intensity aerobic activity a week because of chronic conditions, they should be as physically active as their abilities and conditions allow.

Moderate-intensity physical activity: Aerobic activity that increases a person's heart rate and breathing to some extent. On a scale relative to a person's capacity, moderate-intensity activity is usually a 5 or 6 on a 0 to 10 scale. Brisk walking, dancing, swimming, or bicycling on a level terrain are examples. Vigorous-intensity physical activity: Aerobic activity that greatly increases a person's heart rate and breathing. On a scale relative to a person's capacity, vigorous-intensity activity is usually a 7 or 8 on a 0 to 10 scale. Jogging, singles tennis, swimming continuous laps, or bicycling uphill are examples. Muscle-strengthening activity: Physical activity, including exercise that increases skeletal muscle strength, power, endurance, and mass. It includes strength training, resistance training, and muscular strength and endurance exercises. Bone-strengthening activity: Physical activity that produces an impact or tension force on bones, which promotes bone growth and strength. Running, jumping rope, and lifting weights are examples.

Source: Adapted from U.S. Department of Health and Human Services. *Physical Activity Guidelines for Americans, 2nd edition.* Washington, DC: U.S. Department of Health and Human Services; 2018. Available at *https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf.*

Encouraging daily activity to promote fatigue, avoidance of eating and drinking alcohol too close to bedtime, and regular daily sleep habits may help patients achieve better sleep. Regular use of sedative medications should generally be discouraged given the high potential for dependence, addiction, and altered sleep quality.

DISORDERS OF SLEEP The prevalence of sleep-related breathing disorders, including obstructive sleep apnea (OSA), is high and increasing in our aging and increasingly overweight population. A recent systematic review suggested that that the prevalence of clinically important OSA in the general adult population may be between 9 and 38%, with higher rates in men versus women, older versus younger adults, and those with higher versus lower body mass index (BMI). In the United States, ~40 million adults, and worldwide, nearly a billion adults, are affected by OSA. Patients with persistent complaints of poor sleep quality or excessive daytime somnolence or with witnessed apneic spells may benefit from screening for sleep disorders prior to consideration of a formal sleep study. A number of clinical tools have been developed to screen for sleep apnea, including the Epworth Sleepiness Scale, the STOP (*s*noring, *t*iredness, *o*bserved apnea, high blood *p*ressure) Questionnaire, and the STOP-Bang Questionnaire (STOP plus assessment of BMI, age, neck circumference, and gender), among others. The U.S. Preventive Services Task Force found that current evidence is insufficient to assess the balance of benefits and harms of screening for OSA in asymptomatic adults owing to a lack of validation data in primary care settings. Nonetheless, the high prevalence and significant health consequences of sleep apnea suggest that clinicians should be alert for its potential presence, particularly in patients who are obese with symptoms of excessive daytime somnolence or witnessed apnea episodes. Other sleep disorders, such as restless leg syndrome, may be identified with simple history.

Weight Management Overweight and obesity are prevalent in epidemic proportions in the United States and other industrialized nations **(Chaps. 413 and 414)**. Since 1985, the prevalence of obesity in the United States has increased from ~10 to ~35%, and the prevalence of overweight is now ~40%. Overweight and obesity disproportionately affect individuals in lower socioeconomic strata and in many underrepresented populations, including individuals who identify as Black, Latino, and American Indian. In all race/ethnic groups, both overweight and obesity are associated with adverse health consequences, including diabetes, certain cancers, cardiovascular diseases, and degenerative joint disease. Eating disorders such as anorexia and bulimia are much less common but pose major health consequences for affected patients and should be suspected particularly in younger women with history of rapid weight shifts, electrolyte disturbances, or underweight status.

Weight loss is one of the most difficult preventive interventions to achieve and sustain over time. However, several key factors can assist the patient and clinician, and early referral to a dietician can be very helpful. The first therapeutic goal is to aim for weight stabilization. Many of the risks of overweight and obesity are driven more strongly by continued weight gain, rather than overweight/obese status per se. Working with the patient to find initial strategies for weight maintenance can be a successful initial step with success for many patients. For those who can progress to considering weight loss, it is critical to help the patient understand that there is no standard solution. Experimentation and documentation are key. Tools to assist patients can include food and weight logs, activity logs, and smart phone apps. Some patients respond best to structured approaches such as intermittent fasting regimens or commercial dietary programs where meals are provided. Any of these approaches can be tried with or without social group supports.

The key construct for weight loss is, of course, negative calorie balance. This is achieved through a combination of reduced caloric intake and increased physical activity. Patients may already understand, from prior weight loss attempts, what combination works best for them to achieve this. Some patients find that they cannot lose weight without increasing their exercise. For many, reduction of caloric intake is most efficient. Encouraging the patient to find what works for them is most important. The same principle holds for dietary content. Well-done feeding studies indicate that weight loss is dependent far more on the reduction of caloric intake than on the relative composition of fat, protein, and carbohydrate in the diet. There may be other medical reasons to choose one approach over another, but if not, encouraging the patient to pick one approach and document the results is an important start. Newer pharmacologic agents (e.g., GLP1 receptor agonists) appear to be remarkably successful in helping patients achieve substantial weight loss (up to 20% of baseline body weight), with proven short-term benefits in cardiovascular risk reduction, but issues of access, cost, and long-term safety remain. Once weight loss is achieved, increase in activity is often required for its successful maintenance.

Tobacco Cessation (see Chap. 465) Escaping nicotine dependence is another major, but critical, challenge to prevention and wellness efforts. The addictive effects of nicotine have been well documented, with effects that can last for years after successful cessation. Assessing a patient's past history of cessation attempts and current readiness for change are key first steps in forging a successful approach. Frequent follow-up and reinforcement, as well as use of nicotine replacement therapy and other cessation-promoting medications, are additional critical elements. Recidivism is the rule, and patients should expect to resume smoking and attempt again as they journey to tobacco cessation. Electronic cigarettes have some evidence for benefit in adult cessation of combustible tobacco, but this approach may transfer and worsen nicotine addiction, and the potential for use by adolescents and young adults who are not smokers represents a major public health threat for a new generation of nicotine addiction, with unknown health consequences as a result of the high doses of nicotine delivered to developing organs, including the brain. Vaping of other substances, often in association with flavoring compounds, has also been associated with pulmonary and cardiovascular damage and should be actively discouraged.

VACCINATION (CHAP. 129)

One of the major advances in public health that has contributed to increases in health and longevity worldwide is the development of safe and effective vaccinations against endemic and epidemic infectious diseases. Patients should be counseled regarding age-appropriate vaccinations for their children and for themselves. Some individuals may be reluctant to receive a vaccination; in these cases, listening to the patient's concerns is important, followed by explanation of the benefits to the individual, their family, and their community and review of the low risk for potential harms. It is true to say that no current vaccines are ever worse than the disease they prevent, although side effects may occur rarely. Thorough knowledge of the data on side effect rates and of efficacy will aid the clinician in helping the patient make a fully informed decision.

MENTAL HEALTH AND ADDICTION

Assessment for depression and cognitive impairment is important to address when patients exhibit symptoms or they or their family members express concerns. Both of these common conditions play a major role in reducing quality of life and are high on patients' lists of concerns, even if not clearly expressed. Screening tools for depression are reviewed in Chap. 463. Cognitive function decline with aging or comorbid illness, including depression, should be anticipated. Assessment tools such as the General Practitioner Assessment of Cognition or the Mini-Cog test are widely available and effective rapid assessment tools.

Alcohol and Opioids (see Chaps. 464 and 467) Alcohol dependence and abuse are common and underdiagnosed. Rapid screening tools have proven efficacy for identifying patients with alcohol problems. In a systematic review, the CAGE (*c*ut down, *a*nnoyed, *g*uilty, *e*ye opener) questionnaire was most effective at identifying alcohol abuse and dependence, with reasonable sensitivity and high specificity. The present opioid epidemic in the United States presents a new and substantial public health challenge given the high potential for dependency and abuse of these drugs. Rapid screening tools are available to assist clinicians in screening for opioid dependence.

ACCIDENTS AND SUICIDE

Regular assessment of patient safety through simple questions about seat belt use, domestic violence, and gun safety in the home continues to be an important part of health promotion and wellness. Longstanding recommendations for assessment of suicidal ideation among patients with depression or a history of suicide attempts also continue to be relevant.

APPROACH TO THE PATIENT

In the context of a clinical visit focused on health assessment, health promotion, and prevention, the basic skills of history-taking are of paramount importance. Much of the evaluation, counseling, and management that focus on health promotion and prevention also require engagement and buy-in from the patient in order to assist with recognition of contributing behaviors and to promote adherence to therapeutic plans. Therefore, in addition to standard history-taking, additional skills such as motivational interviewing and eliciting patient commitments and contracting may prove of significant value. The availability of additional tools to assist with screening, monitoring, and chronic management, both online and through wearable devices and mobile health technologies, is rapidly expanding, with uncertain implications for the future. Major research gaps exist in our understanding of how best to employ these newer technologies to improve health outcomes. Concepts of behavioral economics are being explored to better understand the psychology of decision-making and incentives as a means to improve lifestyle choices and adherence to treatment plans (Chap. 494).

The limited time available to clinicians and patients during a wellness visit or periodic health examination (not driven by specific patient issues) makes it important to prioritize assessment and counseling for factors that affect longevity, health span, and quality of life over approaches that may have low yield, such as the annual comprehensive physical examination in an asymptomatic patient. Setting clear expectations for the content of a wellness visit may be a first step, and scheduling follow-up visits for findings or to continue indicated counseling are important steps to achieving better health outcomes.

FURTHER READING

Boulware LE et al: Systematic review: The value of the periodic health evaluation. Ann Intern Med 146:289, 2007.

Dietary Guidelines for Americans, 2020–2025. Washington, DC: U.S. Department of Agriculture and U.S. Department of Health and Human Services; 2020. Available at *https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf*.

Krogsboll LT et al: General health checks in adults for reducing morbidity and mortality from disease. Cochrane Database Syst Rev 1:CD009009, 2019.

Lloyd-Jones DM et al: Life's Essential 8: Updating and enhancing the American Heart Association's Construct of Cardiovascular Health: A presidential advisory from the American Heart Association. Circulation 146:e18, 2022.

U.S. Department of Health and Human Services: *Physical Activity Guidelines for Americans,* 2nd ed. Washington, DC: U.S. Department of Health and Human Services; 2018. Available at *https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf*.

U.S. Preventive Services Task Force webpage. Available at *https://www.uspreventiveservicestaskforce.org/uspstf/*.

3

Vaccine Opposition and Hesitancy

Julie A. Bettinger, Hana Mitchell

Vaccines have been recognized as one of the top public health achievements of the twentieth century. Dramatic declines in the morbidity and mortality of vaccine-preventable diseases have been observed, and the contribution of vaccines to the elimination, control, and prevention of infectious disease cannot be overstated. However, opposition and hesitancy to vaccines occurred, even prior to the COVID-19 pandemic, and are not new. Vaccine hesitancy has existed since Edward Jenner introduced the first vaccine against smallpox in the eighteenth century and the World Health Organization (WHO) ranked these attitudes as one of the ten greatest threats to public health in 2019. Are current opposition and hesitancy any different from what has been seen before? Many sociologists, public health experts, and health care providers (HCPs) argue yes. Recent social and cultural trends, combined with new communication formats and further amplified by the COVID-19 pandemic, have converged to create a particularly potent form of hesitancy and what some have labeled a crisis of confidence. This crisis manifests as a lack of trust in specific vaccines, vaccine programs, researchers, HCPs, the health care system, pharmaceutical companies, academics, policymakers, governments, and authority in general.

The roots of modern vaccine hesitancy and opposition—defined as delay or rejection of vaccines despite availability—vary depending on the place and the population. For some individuals and communities, pseudoscience and false claims about the safety of existing vaccines (e.g., an unsupported link between measles vaccine and autism) have driven fears, increased hesitancy, and decreased acceptance. For others, real safety events, such as the association of narcolepsy with a specific pandemic influenza vaccine (Pandemrix), have justified concerns. In a few locations, vaccine hesitancy is the result of failed health systems or even state failures. Finally, for some groups, including some fundamentalist religious groups and alternative-culture communities, vaccine hesitancy and opposition reflect exclusion from and rejection of mainstream society and allopathic health care and manifest as a deep distrust of these institutions and their HCPs. Although the genesis of modern vaccine hesitancy is multifactorial, its outcomes are uniform: a decrease in vaccine demand and uptake, a decrease in coverage by childhood and adult vaccines, and an increase in vaccine-preventable diseases, outbreaks, and epidemics of disease. Addressing this crisis and moving people from vaccine hesitancy and refusal to acceptance and active demand require intervention at multiple levels: the individual, the health system (including public health), and the state.

This chapter will define vaccine hesitancy and briefly describe its determinants and effects in North America (the United States and Canada). Physicians and other HCPs are well positioned to address the crisis of confidence many patients feel toward HCPs and the health care system. Studies demonstrate that an unambiguous, strong recommendation by trusted HCPs is most often the reason that patients, including those who are vaccine hesitant, choose to vaccinate. Strategies for counseling vaccine-hesitant and vaccine-resistant patients will be presented and examples of strong vaccine recommendations provided. Presenting strategies to increase vaccine demand at a system and policy level is beyond the scope of this chapter. While some physicians may have roles that allow them to act at this level, all physicians can act and influence their individual patients. Strategies to create active vaccine demand at the individual level alone will not solve vaccine hesitancy, but vaccine hesitancy cannot be addressed without these efforts. **For further discussion of immunization principles and vaccine use, see Chap. 129.**

VACCINE COVERAGE AND OUTBREAKS

The epidemiologic data from measles outbreaks over the past 15 years provide an interesting illustration of the effects of vaccine opposition and hesitancy. **For further discussion of measles, see Chap. 211.**

North America *Herd immunity* occurs when enough individuals in a population become immune to an infectious disease, usually through vaccination, that transmission of the infection stops. The level of immunity (or level of vaccine coverage) required to confer herd immunity varies with the specific infectious disease. Because measles is a highly contagious virus, a coverage rate of 93–95% must be achieved for vaccination to confer herd immunity and interrupt measles transmission. National coverage estimates place one-dose measles vaccine coverage rates in 2-year-old children at 92% in the United States and 92% in Canada. Despite these relatively high levels of coverage in young children, numerous measles outbreaks have occurred in both countries since 2010 **(Table 3-1)**.

The vast majority (>80%) of measles cases described in Table 3-1 occurred in under- or completely unvaccinated individuals. Of note, many of these outbreaks highlight pockets of significantly under- or unvaccinated individuals that are not apparent in national vaccine coverage statistics. Moreover, many of the outbreaks listed in Table 3-1 were ignited by unvaccinated returned travelers from areas with existing outbreaks or epidemics, who spread disease into an unvaccinated or undervaccinated community. Many of the outbreaks were contained within the nonvaccinating community, but several spread to other undervaccinated communities geographically contiguous with the outbreak community. More concerning still are the cases and outbreaks originating in communities that had not previously been identified as nonvaccinating. These cases likely highlight pockets of unvaccinated individuals who object for cultural rather than religious reasons. In the past, these nonvaccinating individuals did not exist in large enough clusters to sustain the spread of measles. Of further concern is the number of individuals included in outbreak statistics who have had one or sometimes even two doses of vaccine and who were thought to be protected but who still end up with the disease. The assumption is that one or two doses provide full disease immunity, but this is not always true. Often, individual-level characteristics (e.g., age, immunocompromise) affect the individual's response to the vaccine and their level of protection. In other instances, vaccine protection can wane over time, thus leaving fully immunized individuals susceptible to infection. In fact, when herd immunity breaks (i.e., the level of immunity in a community becomes too low to prevent transmission of disease), the occurrence of cases even in fully immunized persons is seen, as reflected in outbreak statistics. As a result of decreased vaccination rates and the resulting disruption of herd immunity, these individuals may become more identifiable as nonimmune.

Outside North America Although overall coverage rates may still be high in North America, they are lower in other parts of the world and further decreased during the COVID-19 pandemic. In 2022, for example, only 34% of countries met the WHO goal of providing one dose of measles vaccine to at least 95% of their 1-year-old children, a decrease of 10% from prepandemic levels. Twenty years ago, vaccine coverage was sufficiently high in some parts of the world, including Europe, that an unvaccinated traveler from a nonvaccinating community to most regions would have been protected by herd immunity at their destinations. Today that is not the case: such travelers are likely to become infected in a country with active measles transmission and return home to spread the infection into their communities and possibly beyond. Thus, active measles transmission, whether at home or abroad, places individuals who rely on herd immunity (e.g., immunocompromised persons and young infants) at increased risk.

FACTORS IN VACCINE HESITANCY

Vaccination coverage rates provide an estimate of the proportion of children or adults in the population who have been vaccinated, but they do not indicate the proportion of individuals who are vaccine hesitant. An individual may be fully vaccinated but still be hesitant about the safety and effectiveness of vaccines, or an individual may be unvaccinated as a result of access issues but may not be hesitant. Therefore, in attempts to understand a patient's lack of vaccination, it is important to distinguish persons who are hesitant and refuse vaccines

TABLE 3-1 Measles Outbreaks in North America

YEAR/PLACE	NO. OF CASES	REASON
2010/Canada	70	An infected traveler to the 2010 Winter Olympics transmitted infection to an under- and unvaccinated local population in British Columbia.
2011/Canada	776	Disease was imported from France by an unvaccinated returned traveler to Quebec. The outbreak spread in a nonvaccinating religious community and outside that community. A majority of cases occurred in under- and unvaccinated persons.
2011/United States	118	Of 118 cases, 46 were in returned travelers from Europe and Asia/Pacific regions; 105 cases (89%) occurred in unvaccinated persons.
2013/United States	58	Disease was imported by a returned unvaccinated traveler from Europe. The outbreak spread in a nonvaccinating religious community in New York.
2014/Canada	433	Disease was imported from the Netherlands. The outbreak spread in a nonvaccinating religious community in British Columbia.
2014/United States	383	The outbreak occurred in nonvaccinating religious communities in Ohio.
2015/United States	147	A multistate/multicountry outbreak was linked to Disneyland amusement park. More than 80% of cases occurred in unvaccinated persons.
2015/Canada	159	Disease was imported from the United States (part of the Disneyland outbreak) by an unvaccinated traveler. The outbreak spread in a nonvaccinating religious community in Quebec.
2017/United States	75	The outbreak occurred in an undervaccinated community in Minnesota; 95% of patients were unvaccinated.
2018/United States	375	Disease was imported by returned unvaccinated travelers from Israel. The outbreak spread in nonvaccinating religious communities in New York and New Jersey.
2019/Canada	31	Disease was imported from Vietnam by a returned traveler to British Columbia. The outbreak spread throughout local area schools in under- and unvaccinated persons and resulted in a province-wide measles mass immunization campaign for schoolchildren.
2019/United States	1282	Outbreaks occurred in 10 states; 73% of cases (~935) were linked to outbreaks in nonvaccinating religious communities in New York.

Source: Centers for Disease Control and Prevention and Public Health Agency of Canada.

from those who need assistance to access the health care system and successfully complete vaccination. To this end, an understanding of vaccine hesitancy and its determinants is needed.

Vaccine hesitancy and opposition are defined by the WHO's SAGE Working Group on Vaccine Hesitancy as a "delay in acceptance or refusal of vaccines despite availability of vaccination services." The SAGE group describes vaccine hesitancy as "complex and context specific, varying across time, place, and vaccines."

It is useful to frame vaccine acceptance as a continuum pyramid, with active demand for all vaccines representing the largest group at the bottom of the pyramid and outright refusal of all vaccines depicted in the smallest group at the top. In the middle lies vaccine hesitancy, in which the degree of vaccine demand and acceptance varies. Fortunately, for disease control efforts, most individuals fall within the active-demand category or, if they are hesitant, still accept all vaccines. Hesitancy can be influenced by complacency, convenience, and confidence (Fig. 3-1).

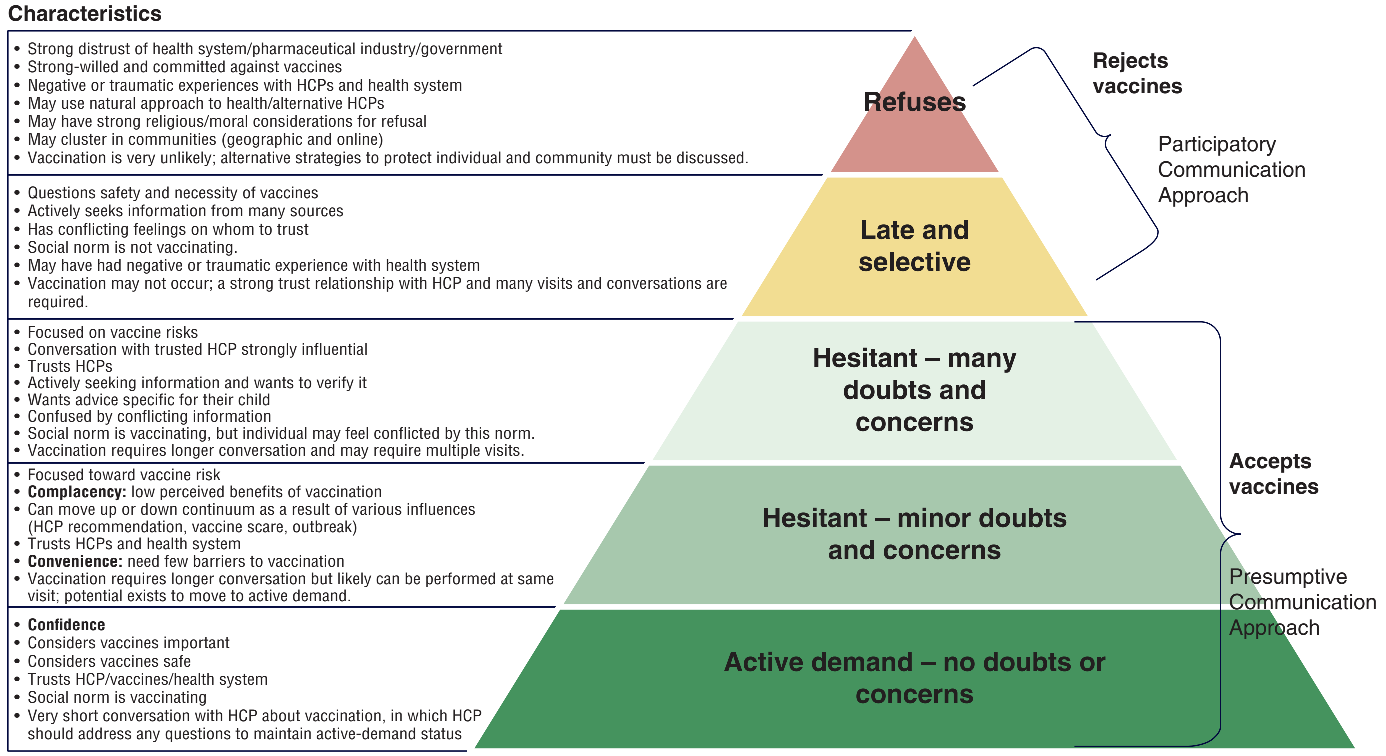

Characteristics

- Strong distrust of health system/pharmaceutical industry/government
- Strong-willed and committed against vaccines
- Negative or traumatic experiences with HCPs and health system
- May use natural approach to health/alternative HCPs
- May have strong religious/moral considerations for refusal
- May cluster in communities (geographic and online)
- Vaccination is very unlikely; alternative strategies to protect individual and community must be discussed.

- Questions safety and necessity of vaccines
- Actively seeks information from many sources
- Has conflicting feelings on whom to trust
- Social norm is not vaccinating.
- May have had negative or traumatic experience with health system
- Vaccination may not occur; a strong trust relationship with HCP and many visits and conversations are required.

- Focused on vaccine risks
- Conversation with trusted HCP strongly influential
- Trusts HCPs
- Actively seeking information and wants to verify it
- Wants advice specific for their child
- Confused by conflicting information
- Social norm is vaccinating, but individual may feel conflicted by this norm.
- Vaccination requires longer conversation and may require multiple visits.

- Focused toward vaccine risk
- **Complacency:** low perceived benefits of vaccination
- Can move up or down continuum as a result of various influences (HCP recommendation, vaccine scare, outbreak)
- Trusts HCPs and health system
- **Convenience:** need few barriers to vaccination
- Vaccination requires longer conversation but likely can be performed at same visit; potential exists to move to active demand.

- **Confidence**
- Considers vaccines important
- Considers vaccines safe
- Trusts HCP/vaccines/health system
- Social norm is vaccinating
- Very short conversation with HCP about vaccination, in which HCP should address any questions to maintain active-demand status

Refuses
Late and selective
Hesitant – many doubts and concerns
Hesitant – minor doubts and concerns
Active demand – no doubts or concerns
Rejects vaccines
Participatory Communication Approach
Accepts vaccines
Presumptive Communication Approach

FIGURE 3-1 Vaccine acceptance continuum. HCPs, health care providers. *(Adapted from J Leask et al: BMC Pediatrics 12:154, 2012; AL Benin et al: Pediatrics 117:1532, 2006; and E Dubé, NE MacDonald: The Vaccine Book, 2016, pp. 507-528.)*

Complacency is self-satisfaction when accompanied by a lack of awareness for real dangers or deficiencies. Complacency exists in communities and individuals when the perceived risks of vaccine-preventable diseases are low and vaccination is not deemed a necessary preventive action. This attitude can apply to vaccination in general or to specific vaccines, such as influenza vaccines. Actual or perceived vaccine efficacy and effectiveness contribute to complacency. Patients who are complacent about vaccine-preventable diseases prioritize other lifestyle or health factors over vaccination. These individuals can be influenced toward vaccination by a strong recommendation from a trusted HCP or a local influenza outbreak. They can be influenced away from vaccination by a vaccine scare or misinformation on social media. Finally, the real or perceived ability of patients to take the action required for vaccination (i.e., self-efficacy) influences the role complacency plays in hesitancy and willingness to seek vaccination.

Convenience is determined by the degree to which conversations about vaccination and other services can be provided in culturally safe contexts that are convenient and comfortable for the individual. Clearly, convenience varies by community, health clinic, and even patient. Persons who are criticized or scolded for not vaccinating themselves or their children may not feel comfortable or safe accessing health services. Factors such as affordability, geographic accessibility, language, and health literacy are important considerations when evaluating the convenience of existing clinical care. Any of these factors can affect vaccine acceptance and can push a patient who has some hesitancy toward vaccinating or not vaccinating.

Confidence is based on trust in the safety and efficacy of vaccines, in the health care system that delivers vaccines (including HCPs), and in the policymakers or governments who decide which vaccines are needed and used. A continual erosion of confidence around vaccination, health systems, and governments drives today's hesitancy and has been amplified by larger social and cultural trends in medicine, parenting, and information availability.

SOCIAL AND CULTURAL TRENDS

Individualized Health Care Over the past 30 years, the focus of medicine and health care has shifted to patient-oriented, individualized care, with an increasing emphasis on treatment and prevention options tailored to the individual patient. In vaccination programs, this shift has manifested as requests for individualized vaccine recommendations and customized immunization schedules. The increasing personalization of medicine, while positive overall, has forced public health away from a focus on the community and its common good and has created tension between individual rights and community health, which was further exacerbated during the COVID-19 pandemic.

Parenting Trends The desire for an individualized approach to medicine and vaccination reflects broader cultural trends concerning individual risk management: accordingly, the individual is to blame for bad outcomes, and public institutions cannot be trusted to manage technological (i.e., vaccine-related) risks. This viewpoint is directly linked with cultural shifts in parenting and social norms defining what it means to be a "good parent." The image of a good parent has been reframed to refer to someone whom several investigators have described as "a critical consumer of health services and products, accounting for their own individual situation as they see it with little regard for the implications of their decision on other children." The archetypical good parent no longer unquestioningly trusts HCPs and other authorities and experts. According to this social norm, "good parents" should seek individual medical advice that is tailored for their child and specific to that child's needs. While in essence not a bad thing, this norm can conflict directly with public health vaccine recommendations and schedules that are organized to maximize community health and to facilitate efficient provision of care at a community level.

Traditional Media Newspapers, radio, and television have been criticized for their coverage of vaccines and in particular their coverage of the alleged link between the measles-mumps-rubella (MMR) vaccine and autism. By offering equal coverage throughout the early to mid-2000s for both the scientific evidence and unproven claims of MMR vaccine harms, traditional media outlets provided a forum and a megaphone for the spread of pseudoscience. Equal coverage leads to false equivalencies. Celebrity advocates further amplified the message via this channel. The boost that traditional media provided to active vaccine resistance and, less directly, to vaccine hesitancy has not been adequately measured but must be considered in any discussion of vaccine hesitancy. After headlines about multiple outbreaks of measles and other vaccine-preventable diseases and continued direct criticism of the equal-coverage approach, some traditional media began rejecting this approach and attempted to discredit pseudoscience. During the COVID-19 pandemic, the approach of traditional media toward scientific evidence further diversified based on the political orientation of the news source or organization. At the patient level, the political orientation of their news source can affect their risk perceptions toward disease as well as their acceptance of vaccines.

The Internet and Social Media Approximately 92% of Americans and 95% of Canadians use the Internet, and 90% of Americans and 86% of Canadians have an active social network profile. Widespread access to social media can be empowering, but it is also problematic. The Internet and social media require users to select their information sources, creating an environment described as an "echo chamber" in which individuals choose information sources harboring beliefs or opinions similar to their own and thereby reinforcing their existing views. This situation has created a new platform for further spread of vaccine *misinformation* (inaccuracies due to error) and *disinformation* (deliberate lies) and has provided a forum for vaccine-resistant individuals, including celebrities, to organize and raise funds to support their efforts. The harmful effects of Internet and social media use on vaccine hesitancy have been well documented. Vaccine hesitancy increases for parents who seek their information from the Internet. In this medium, personal stories and anecdotes are now viewed as data and disproportionately influence vaccine decision-making, while traditional, more authoritative, fact-based information sources are deemphasized.

As with traditional media, the social media landscape appears to be shifting. In 2019, the proliferation of antivaccination information combined with measles outbreaks in North America and increasing pressure from health leaders led large social media companies (Facebook, Instagram, Pinterest) to deemphasize antivaccination information by removing relevant advertisements and recommendations and decreasing their prominence in search results. While this resulted in an initial decrease in both pro- and antivaccine content, misleading content was unfortunately still widely available. Moreover, antivaccination users switched to alternate platforms without restrictions, and the level of engagement with antivaccine content remained unchanged.

The COVID-19 pandemic further accelerated the spread of misinformation and disinformation circulating on social media to the point that it was termed an "infodemic" and forced public health and health care institutions to respond. Centralized monitoring by jurisdiction of vaccine misinformation and disinformation, with summaries of the relevant discourses and rebuttals provided to HCPs, is a potential way to counter the influence of social media on vaccine hesitancy. Some early work is occurring with this through the WHO Early AI-Supported Response with Social Listening Platform (WHO EARS), which was used in 30 countries during the pandemic to provide centralized monitoring of the COVID-19 discourses on social media and the WHO/Centers for Disease Control and Prevention (CDC) infodemic management training (see "Further Reading" below). While such strategies have been applied in single jurisdictions and appear to have had some success, their applicability beyond a pandemic context is unknown. Moreover, while the resources for a coordinated response were available in some jurisdictions during the pandemic, it is unclear if they will continue to be provided. Most individual HCPs have been left to counter popular, shifting, viral communications on their own, patient by patient, or to adapt the general materials provided by the WHO/CDC to their local context using their own resources.

Given these social and cultural trends, no one should be surprised when individuals now question vaccination, express confusion about conflicting information and information sources, and feel unsure about whom to trust. Their broader social context is telling them they should question everything and trust no one. This message is reinforced via misinformation and disinformation on social media and a politicized traditional media and public. Recent vaccine-preventable disease outbreaks illustrate that effective engagement with individuals cannot be accomplished through one-way, top-down information provision (which still is often the de facto choice for health system communication), but rather requires a dialogue that takes into account the social processes surrounding individual vaccination decisions. It is at the interface between the individual and the health system in which conversations between HCPs and their patients can have the greatest impact. It is critical for all HCPs to discuss vaccines and provide strong vaccine recommendations—including HCPs who do not administer vaccines but who have established trust with their patients.

APPROACH TO THE PATIENT

An ideal vaccine-hesitancy intervention would result in full compliance with vaccination, the patient's satisfaction with the health care encounter, and sustained trust in the HCP's recommendations. On a programmatic level, vaccine-hesitancy interventions should be multicomponent, dialogue based, and tailored to specific undervaccinated populations.

Communicating with vaccine-hesitant individuals can be challenging and time-consuming. HCPs may feel that vaccine-hesitant patients cast doubt on their personal and professional integrity, their authority as medical experts, and their competence as communicators. Some HCPs may be reluctant to initiate conversations about vaccination because of concerns that discussing a sensitive topic may compromise their clinical rapport with their patients. Other HCPs may believe that they have not received sufficient training to confidently recommend vaccines and answer questions. Discussing vaccines with hesitant patients, while not always easy, provides an opportunity to honor the principles of patient-centered care by demonstrating an interest in patients' opinions, engaging in dialogue, and ideally increasing patients' confidence in vaccine recommendations.

FACTORS IN EFFECTIVE VACCINE RECOMMENDATIONS

Vaccine recommendations ideally should be made within an established, trusting patient–provider relationship in which patients are comfortable asking questions and voicing concerns, even if their views on vaccines contradict the HCP's recommendations. Recommending vaccines requires both provision of information and effective communication. There is no single "best practice" for how providers should approach recommending vaccines to vaccine-hesitant individuals. In general, all vaccine recommendations should be (1) strong, making it clear that the provider supports and recommends vaccination; (2) tailored, acknowledging the vaccine attitudes and potential concerns of individual patients; (3) transparent and accurate, highlighting the benefits of vaccines while also communicating the risks; (4) supported by trustworthy information resources that patients can access and review after the clinical encounter; and (5) revisited, with repetition and reinforcement during follow-up health care encounters.

Strength of the Recommendation HCPs should make it explicit (in the absence of medical contraindications) that vaccination based on the recommended schedule is the best option. While HCPs should take time to elicit patients' questions and address concerns, the recommendation for vaccination should be made in clear and unambiguous terms.

Tailored Communication Vaccine hesitancy occurs on a continuum (Fig. 3-1). Therefore, it is helpful for HCPs to have some understanding of their patients' attitudes toward vaccination at the start of the health care appointment. Unfortunately, vaccine-hesitancy surveys for use as part of vaccine consultation visits have not been validated on a large scale. However, the following are some examples of questions that can be asked, depending on the setting. (1) Did you have a chance to review the vaccine leaflet/online resource we provided? Did you have any questions about it? (2) Have you ever been reluctant or hesitant about getting a vaccination for yourself or your child? If so, what were the reasons? (3) Are there other pressures in your life that prevent you from getting yourself or your child immunized on time? (4) Whom/what resources do you trust the most for information about vaccines? Whom/what resources do you trust the least?

Communication style and content for patients in the active-demand category for vaccination will be different from those for individuals who are hesitant, late and selective, or strongly inclined to refuse vaccines. Two communication styles have been proposed for vaccine recommendations. Evidence shows that a *presumptive/directive* approach ("Your child is due for MMR vaccination.") results in higher rates of vaccine uptake than a *participatory/guiding* approach ("What are your thoughts about the MMR vaccine?"). However, adopting a strictly presumptive/directive approach may alienate some patients, especially those who are higher up on the hesitancy pyramid and who may feel that they are being pressured into vaccination before their concerns have been heard and addressed. Adopting a participatory/guiding approach and clarifying receptivity to vaccines may be more suitable for hesitant individuals with many doubts and concerns, persons with a late or selective attitude, and those who are strongly inclined to refuse vaccines. In addition, a participatory/guiding approach provides an opportunity for ongoing clinical rapport and dialogue between unvaccinated or undervaccinated patients and their HCPs, even when it does not result in immediate vaccine uptake. Regardless of which approach is used, a strong vaccine recommendation should be made at each encounter.

Transparency and Accuracy Vaccine recommendations should be transparent, should include accurate information about both the benefits and the risks of the vaccine, and should emphasize why the benefits outweigh the risks. For example, when evidence supports an association between a vaccine and an adverse event, the occurrence of the adverse event is often very rare and the event quickly resolves **(Chap. 129).** U.S. Federal law (under the National Childhood Vaccine Injury Act) requires HCPs to provide a copy of the current Vaccine Information Statement from the CDC, which describes both benefits and risks of vaccines to an adult patient or to a child's parent/legal representative before vaccination.

CDC Vaccine Information Statements should not replace a discussion with the HCP. Depending on the provider and the patient, a description of benefits and risks may include words and numbers, graphics, and personal anecdotes (e.g., why the provider vaccinates their own children). Personal anecdotes are powerful, and many hesitant patients seek and are influenced by them.

A discussion of benefits and risks provides an opportunity to address specific misconceptions about a particular vaccine or about vaccines overall. For example, patients may be concerned about adverse events following vaccination that are not supported by evidence, such as autism following MMR vaccination or myocardial infarction following influenza vaccination in the elderly.

Most adults—even those whose children are fully immunized—still have questions, misconceptions, or concerns about vaccines that should be addressed. A risk/benefit discussion allows HCPs to describe the vaccine safety monitoring systems in place. Providers should emphasize that vaccines are developed and approved through a highly regulated process that includes prelicensure clinical trials, review and approval by designated regulatory authorities (e.g., the U.S. Food and Drug Administration, Health Canada), strict manufacturing regulations, and ongoing postmarketing safety surveillance.

Support from Accessible Information Sources All vaccine recommendations should be supported by additional information sources patients can assess after the health care encounter. HCPs play an important role as information intermediaries for their patients. They can navigate information (and misinformation) about vaccines and direct patients toward reliable, appropriate resources. HCPs should consider what resources will be suitable for a patient or patient population. Vaccine information resources are available in different media formats and use a combination of images and text to communicate the information to various audiences. See "Further Reading," below, for suggestions or refer to resources provided by local health authorities.

Revisiting and Reinforcement of Vaccine Recommendations All health care encounters offer an opportunity to revisit and reinforce vaccine recommendations. Vaccine-hesitant individuals who do not accept vaccines but are willing to review information should be offered a follow-up appointment to reinforce previously made recommendations and address further questions. Vaccine-hesitant patients who accept vaccines should be seen at a follow-up appointment to confirm and document vaccine receipt (if vaccine is not given at the point of care), ascertain whether the vaccine was well tolerated, and reinforce the message about vaccine safety and effectiveness. Patients who actively demand vaccines usually do not require much follow-up other than to confirm and document the receipt of vaccine (if it is not given at the point of care) and to address additional questions or concerns arising subsequent to vaccination. Often this follow-up can be covered without an office visit.

WHAT TO SAY TO VACCINE-HESITANT PATIENTS

Engaging vaccine-hesitant individuals requires confidence, knowledge, skills, time, and creativity to tailor the approach to each individual patient. Examples for each part of the vaccine recommendation are listed in **Table 3-2.**

OTHER CONSIDERATIONS DURING CLINICAL ENCOUNTERS

Missed Opportunities The WHO defines a missed opportunity for vaccination as "any contact with health services by an individual (child or person of any age) who is eligible for vaccination

TABLE 3-2 Sample Vaccine Conversations

STRONG VACCINE RECOMMENDATION
"We are headed into the respiratory virus season. Getting flu, RSV, and COVID vaccines not only protects you, but it helps protect other people around you who can get very sick from flu, RSV, or COVID. I strongly recommend you get shots. Do you know where to get them?"
"You will be turning 50 next year. This means you will be eligible for a vaccine that prevents shingles, and I strongly recommend you receive it. Have you heard about this vaccine before? Can I answer your questions about it?"
"I know you are not comfortable getting vaccinated today. I do want to make it clear that I recommend vaccines because I am convinced they are the best way to protect you from some serious diseases. Is there something that would lead you to think about getting vaccinated in the future?"
TAILORED COMMUNICATION
"I recommend that children and adults stay up to date on recommended vaccines. I see from your vaccine record that you've had your childhood vaccines, but you haven't gotten any adult vaccines. I wanted to clarify whether this is because you decided not to get vaccines or something else prevented you from getting vaccinated."
"I understand that you are here for your pneumococcal vaccine. This is the best way to protect yourself and those around you from pneumonia. Do you have any questions before I give you the vaccine?"
"There is strong evidence that COVID-19 vaccines work well for all people, regardless of their ethnic or genetic background. What particular concerns did you have about the vaccine?"
"Thank you for telling me about your fear of needles. This is quite common in children and in adults. Would you like to talk about some potential strategies to help you with getting vaccinated?"
TRANSPARENCY AND ACCURACY
"Serious side effects can develop after MMR vaccination but are very rare. On average, 3 out of 10,000 children who get MMR vaccine will have a febrile seizure/convulsion in the days after vaccination. Febrile seizures can be frightening, but nearly all children who have a febrile seizure recover very quickly and without any long-term consequences. On the other hand, 1 out of 1000 children who get measles will develop encephalitis (brain inflammation) that not only causes seizures but can also lead to permanent damage."
"About 10 out of every 10,000 Americans who do not get vaccinated against flu die because of influenza every year, and many more are hospitalized. While flu vaccine does not prevent all cases of influenza, it is the most effective vaccine we have. By getting the vaccine, you also help protect people around you from getting sick."
"You are correct, aluminum is used in some vaccines to help the body's immune system respond. However, aluminum is also present in food and drinking water. In fact, the amount of aluminum present in vaccines is similar to or less than what is present in breast milk or infant formulas."
SUPPORT FROM ACCESSIBLE INFORMATION SOURCES
"Your child and other boys and girls his age will be eligible for the human papillomavirus vaccine this coming school year. Have you heard about this vaccine before? What questions do you have about it? Here's a list of websites for parents and teenagers that explain what it is about."
"There's a lot of information about vaccines on the Internet, and a lot of that information is not based on facts. Here is a list of websites that have been reviewed by health care professionals and accurately describe benefits and risks of each vaccine, including information resources written by the LGBTQ community that many of my patients have found useful."
REVISITING AND REINFORCEMENT OF THE RECOMMENDATION
"During our last visit, we talked about why COVID vaccine is recommended for your son and some of the concerns you had about potential side effects, especially myocarditis. It is important to weigh the risks of side effects against the risks of infection. Have you had a chance to look at the take-home information I gave you? Was there anything else you or your partner would like to ask about?"
"When you were here last month, we talked about receiving a pertussis booster during pregnancy and where you can get vaccinated. Have you had a chance to get your pertussis vaccine?"
"I see that you got your vaccines at the public health clinic last week. How did it go? Did you have any questions?"
"It's possible that the symptoms you experienced after receiving the vaccine were an adverse reaction to the vaccine. I will report this to the health authority. Let's discuss what we can do next time to prevent symptoms from occurring again."

Note: Specific vaccine recommendations, vaccine eligibility guidelines, and statistics used to communicate benefits and risks will vary with the health jurisdiction and the country. Several sample statements here are adapted from the Australian National Centre for Immunisation Research and Surveillance website (*www.skai.org.au/healthcare-professionals*). For patient vaccine information resources, see also the Immunization Action Coalition website for the public developed in partnership with the Centers for Disease Control and Prevention (*vaccineinformation.org*).

(e.g., unvaccinated or partially vaccinated and free of contraindications to vaccination), which does not result in the person receiving one or more of the vaccine doses for which he or she is eligible." HCPs who do not offer point-of-care vaccination frequently miss the opportunity to recommend vaccines to their patients. Missed opportunities for recommending and providing vaccines during routine health care encounters contribute to undervaccination. Studies show that up to 45% of undervaccinated children could be up to date with all age-appropriate vaccines and up to 90% of female adolescents could be up to date with human papillomavirus (HPV) vaccination if all opportunities to vaccinate were taken.

Vaccine counseling and vaccination should be incorporated into clinical care for individuals of all ages, not just young children. Clinical encounters should be used as an opportunity to remind patients about seasonal vaccines (influenza, COVID-19) as well as new vaccines as they become available, such as respiratory syncytial virus for older individuals. Because many adolescents and adults do not have regular health care follow-up, providers should take advantage of every health care encounter to recommend and provide vaccines. For example, a visit to an emergency department, a routine follow-up visit at a diabetes clinic, or a visit planning for elective orthopedic surgery offer opportunities to inquire about the patient's vaccination status and to recommend vaccines. Depending on the jurisdiction, adolescents may or may not have the legal ability to consent to or decline vaccines. Adolescents' views, questions, and concerns related to vaccines may differ from those of their caregivers and should ideally be explored as part of adolescent health care.

HCPs should make preemptive vaccine recommendations (e.g., initiating discussions about infant vaccines during pregnancy, informing parents about HPV vaccine before their child becomes eligible). Such advance discussions may be especially helpful in identifying vaccine-hesitant patients and ensuring that they have enough time to ask questions and make decisions before vaccines are due.

HCPs should ensure that a vaccine recommendation is followed by vaccination. Providers who recommend vaccines but do not vaccinate at the point of care should inform patients where they can be vaccinated. This discussion may include information about public health clinics, travel clinics, and pharmacies or a referral to another provider. HCPs should follow up with their patients at subsequent appointments to confirm that they were vaccinated.

HCPs should be prepared to discuss newer vaccines. While safety concerns about new vaccines can be anticipated on the basis of past experience with other new vaccines, the COVID-19 pandemic highlighted the need for HCPs to understand and be able to explain the newer vaccine platforms (mRNA, DNA, and viral vector vaccines) and to provide examples of other, older vaccines that have been developed by similar techniques. HCPs also need to be able to explain and provide information resources around how vaccines are evaluated before being approved for use and how vaccine safety is monitored after vaccines are used in the population. It is important to be honest, to describe known, rare side effects (e.g., myocarditis in young males following COVID-19 vaccine) and the positive outcomes in these cases. Placing potential vaccine risks in the context of known disease risks is helpful for some patients. Depending on the context, HCPs should explain why specific high-risk groups may have been prioritized to receive the vaccine.

Adverse Events Following Vaccination Although rare, adverse events **(Chap. 129)** may influence vaccine acceptance and willingness to be vaccinated in the future. Frequent, acute adverse effects can be captured in clinical trial data, whereas worries about rare and long-term side effects can be addressed only by direct evidence after the initiation of a new vaccination program. Providing patients with information on the incidence of common or expected health events in an unvaccinated population (i.e., background rates) over a 4-week period is helpful in distinguishing what is normal and expected from a point of concern. It is important to ensure that more specific background-rate information is available to HCPs with regard to the individual groups being vaccinated (e.g., pregnant individuals, children, immunocompromised people) whenever possible. HCPs, public health programs, and vaccine manufacturers can anticipate these questions and should develop answers and information to respond to them.

It is also essential for providers to identify and follow up with all patients who experience an adverse event, regardless of the patients' vaccine attitudes prior to the event. Adverse events following vaccination should be reported to the relevant vaccine monitoring system: the U.S. Vaccine Adverse Event Reporting System or the Canadian Adverse Event Following Immunization Surveillance System.

Addressing Inequities in Vaccine Access Discrepancies in access to health care services create inequitable access to vaccines for children and adults and contribute to undervaccination, disproportionally affecting black people, indigenous populations, and people of color. HCPs must recognize that socially disadvantaged individuals and populations are often at greater risk of vaccine-preventable diseases (e.g., as a result of crowded living conditions, limited access to sanitation, poor nutrition, or substance abuse). They are also at greater risk of being undervaccinated because they have limited access to health care services and continue to face pervasive discrimination within the health care system.

Depending on the setting and the patient, some recommended vaccines may not be covered through public funding or private insurance coverage. HCPs should be aware of alternative funding models, such as the Vaccines for Children Program, which provides free vaccines for U.S. children (<19 years of age) with financial barriers to vaccine access. When vaccines are not publicly funded or covered by private insurance and patients perceive that they cannot afford a vaccine, HCPs should not withhold a vaccine recommendation. The risks and benefits of vaccination still need to be communicated, with a strong recommendation, and the patient should be provided the opportunity to decide whether they can afford the vaccine.

Providing Culturally Safe Care Cultural safety in health care is defined as an outcome based on respectful engagement between the patient and the HCP that recognizes and strives to address power imbalances inherent in the health care system. It results in an environment free of racism and discrimination, where people feel safe when receiving health care. HCPs need to be aware of the legacy of discrimination, racism, and medical experimentation and the distrust in vaccines this has fostered for many individuals and communities[1] and strive to approach clinical practice with cultural humility and self-reflection. While SARS-CoV-2 has critically highlighted fractures in our health care system for minority and marginalized communities, addressing these underlying issues goes beyond addressing vaccine hesitancy and is clearly needed for all types of medical care in these communities.

Further Communication With Patients Who Refuse Vaccines Fortunately, the proportion of people who completely refuse all vaccines and are not willing to talk to their HCP is small. Nevertheless, in some cases, attempts to initiate discussion and address vaccine refusal may be futile. When possible, HCPs should focus on the common goals of care and preserve the therapeutic relationship. Vaccine refusal should be well documented in the patient's chart. The HCP should continue with tailored communication and be open to future discussions. Vaccine demand and vaccine refusal are rarely static over time.

CONCLUSION

In summary, vaccine hesitancy is complex and context specific. It varies with time, place, patient, and vaccine. HCPs are well positioned to address vaccine hesitancy and should develop the skills, knowledge, and confidence to make strong vaccine recommendations to their patients.

---

[1] The Tuskegee Syphilis Study is the most infamous example of medical experimentation in black communities in the United States. (See Brandt [1978] for details.)

FURTHER READING

American Academy of Pediatrics: Talking with vaccine hesitant parents. Available at *www.aap.org/en/patient-care/immunizations/communicating-with-families-and-promoting-vaccine-confidence/talking-with-vaccine-hesitant-parents/*. Accessed December 15, 2023.

Brandt AM: Racism and research: The case of the Tuskegee Syphilis Study. Hastings Cent Rep 8:21, 1978.

Centers for Disease Control and Prevention: How to address COVID-19 vaccine misinformation. Available at *www.cdc.gov/vaccines/covid-19/health-departments/addressing-vaccine-misinformation.html*. Accessed January 15, 2024.

Centers for Disease Control and Prevention: Vaccinate with confidence: Strategy to reinforce confidence in Covid-19 vaccines. Available at *www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence.html*. Accessed January 15, 2024.

DeStefano F et al: Principal controversies in vaccine safety in the United States. Clin Infect Dis 69:726, 2019.

Dudley MZ et al: The state of vaccine safety science: Systematic reviews of the evidence. Lancet Infect Dis 20:e80, 2020.

Immunization Action Coalition: For healthcare professionals. Available at *www.immunize.org*. Accessed December 15, 2023.

Immunization Action Coalition: For the public: Vaccine information you need. Available at vaccineinformation.org. Accessed December 15, 2023.

Leask J et al: Communicating with parents about vaccination: A framework for health professionals. BMC Pediatr 12:154, 2012.

Lurie N et al: Developing Covid-19 vaccines at pandemic speed. N Engl J Med 382:21, 2020.

MacDonald N et al: Vaccine hesitancy: Definition, scope and determinants. Vaccine 33:4161, 2015.

Quinn S et al: Addressing vaccine hesitancy in BIPOC communities: Toward trustworthiness, partnership, and reciprocity. N Engl J Med 385:8, 2021.

World Health Organization: Infodemic. Available at *www.who.int/health-topics/infodemic#tab=tab_1*. Accessed January 15, 2024.

World Health Organization: Reducing missed opportunities for vaccination (MOV). Available at *www.who.int/teams/immunization-vaccines-and-biologicals/essential-programme-on-immunization/implementation/reducing-missed-opportunities-for-vaccination*. Accessed December 15, 2023.

4

Decision-Making in Clinical Medicine

Daniel B. Mark, John B. Wong

Practicing medicine at its core requires making decisions. What makes medical practice so difficult is not only the specialized technical knowledge required but also the intrinsic uncertainty that surrounds each decision. Mastering the technical aspects of medicine alone, unfortunately, does not ensure a mastery of the practice of medicine. Sir William Osler's familiar quote "Medicine is a science of uncertainty and an art of probability" captures well this complex duality. Although the science of medicine is often taught as if the mechanisms of the human body operate with Newtonian predictability, every aspect of medical practice is infused with an element of irreducible uncertainty that the clinician ignores at their peril. Although deeply rooted in science, more than 100 years after the practice of medicine took its modern form, it remains at its core a craft, to which individual doctors bring varying levels of skill, knowledge, and understanding. With the exponential growth in medical literature and other technical information and an ever-increasing number of testing and treatment options, twenty-first century physicians who seek excellence in their craft must master a more diverse and complex set of skills than any of the generations that preceded them. This chapter introduces three of the pillars upon which the craft of modern medicine rests: (1) expertise in clinical reasoning (what it is and how it can be developed); (2) rational diagnostic test use and interpretation; and (3) integration of the best available research evidence with clinical judgment in the care of individual patients (evidence-based medicine [EBM]).

BRIEF INTRODUCTION TO CLINICAL REASONING

Clinical Expertise Defining "clinical expertise" remains surprisingly difficult. Chess has an objective ranking system based on skill and performance criteria. Athletics, similarly, have ranking systems to distinguish novices from Olympians. But in medicine, after physicians complete training and pass the boards (or get recertified), no tests or benchmarks are used to identify those who have attained the highest levels of clinical performance. At each institution, there are often a few "elite" clinicians who are known for their "special problem-solving prowess" when particularly difficult or obscure cases have baffled everyone else. Yet despite their skill, even such master clinicians typically cannot explain their exact processes and methods, thereby limiting the acquisition and dissemination of the expertise used to achieve their impressive results. Furthermore, clinical virtuosity appears not to be generalizable, e.g., an expert on hypertrophic cardiomyopathy may be no better (and possibly worse) than a first-year medical resident at diagnosing and managing a patient with neutropenia, fever, and hypotension.

Broadly construed, clinical expertise encompasses not only cognitive dimensions involving the integration of disease knowledge with verbal and visual cues and test interpretation but also potentially the complex fine-motor skills necessary for invasive procedures and tests. In addition, "the complete package" of expertise in medicine requires effective communication and care coordination with patients and members of the medical team. Research on medical expertise remains sparse overall and mostly centered on diagnostic reasoning, so this chapter focuses primarily on the cognitive elements of clinical reasoning.

Objective study of the clinical reasoning process is difficult as it occurs in the heads of clinicians. One research approach asks clinicians to "think out loud" as they receive increments of clinical information in a manner meant to simulate a clinical encounter. Another research approach focuses on how doctors should reason diagnostically, to identify remediable "errors," rather than on how they actually do reason. Much of what is known about clinical reasoning comes from empirical studies of nonmedical problem-solving behavior. Because of the diverse perspectives contributing to this area, with important contributions from cognitive psychology, medical education, behavioral economics, sociology, informatics, and decision sciences, no single integrated model of clinical reasoning exists, and not infrequently, different terms and reasoning models describe similar phenomena.

Intuitive Versus Analytic Reasoning A useful contemporary model of reasoning, the dual-process theory distinguishes two general conceptual modes of thinking as fast or slow. *Intuition* (System 1) provides rapid effortless judgments from memorized associations using pattern recognition and other simplifying "rules of thumb" (i.e., heuristics). For example, a very simple pattern that could be useful in certain situations is "black woman plus hilar adenopathy equals sarcoid." Because no effort is involved in recalling the pattern, the clinician is often unable to say how those judgments were formulated. In contrast, *analysis* (System 2), the other form of reasoning in the dual-process model, is slow, methodical, deliberative, and effortful. A student might read about causes of hilar adenopathy and from that list (e.g., **Chap. 70**), identify diseases more common in black women or examine the patient for skin or eye findings that occur with sarcoid. These dual processes, of course, represent two exemplars taken from the cognitive continuum. They provide helpful descriptive insights but very little guidance in how to develop expertise in clinical reasoning. How these idealized systems interact in different decision problems, how experts use them differently from novices, and when their usage can lead to errors in judgment remain the subject of study and considerable debate.

Pattern recognition, an important part of System 1 reasoning, is a complex cognitive process that appears largely effortless. One can recognize people's faces, the breed of a dog, an automobile model, or a piece of music from just a few notes within milliseconds without necessarily being able to articulate the specific features that prompted the recognition. Analogously, experienced clinicians often recognize familiar diagnostic patterns very quickly. The key here is having a large library of stored patterns that can be rapidly accessed. In the absence of an extensive stored repertoire of diagnostic patterns, students (as well as experienced clinicians operating outside their area of expertise and familiarity) often must use the more laborious System 2 analytic approach along with more intensive and comprehensive data collection to reach the diagnosis.

The following brief patient scenarios illustrate three distinct patterns associated with hemoptysis that experienced clinicians recognize without effort:

- A 46-year-old man presents to his internist with a chief complaint of hemoptysis. An otherwise healthy, nonsmoker, he is recovering from an apparent viral bronchitis. This presentation pattern suggests that the small amount of blood-streaked sputum is due to acute bronchitis, so that a chest x-ray provides sufficient reassurance that a more serious disorder is absent.
- In the second scenario, a 46-year-old patient who has the same chief complaint but with a 100-pack-year smoking history, a productive morning cough with blood-streaked sputum, and weight loss fits the pattern of carcinoma of the lung. Consequently, along with the chest x-ray, the clinician obtains a sputum cytology examination and refers this patient for a chest computed tomography (CT) scan.
- In the third scenario, the clinician hears a soft diastolic rumbling murmur at the apex on cardiac auscultation in a 46-year-old patient with hemoptysis who immigrated from a developing country and orders an echocardiogram as well, because of possible pulmonary hypertension from suspected rheumatic mitral stenosis.

Pattern recognition by itself is not, however, sufficient for secure diagnosis. Without deliberative systematic reflection, undisciplined pattern recognition can result in premature closure: mistakenly jumping to the conclusion that one has the correct diagnosis before all the relevant data are in. A critical second step, therefore, even when the diagnosis seems obvious, is *diagnostic verification*: considering whether the diagnosis adequately accounts for all of the presenting symptoms and signs and can explain all the ancillary findings. The following case based on a real clinical encounter provides an example of premature closure. A 45-year-old man presents with a 3-week history of a "flulike" upper respiratory infection (URI) including dyspnea and a productive cough. The emergency department (ED) clinician pulled out a "URI assessment form," which defines and standardizes the information gathered. After quickly acquiring the requisite structured examination components and noting in particular the absence of fever and a clear chest examination, the physician prescribed a cough suppressant for acute bronchitis and reassured the patient that his illness was not serious. Following a sleepless night at home with significant dyspnea, the patient developed nausea and vomiting and collapsed. He was brought back to the ED in cardiac arrest and was unable to be resuscitated. His autopsy showed a posterior wall myocardial infarction (MI) and a fresh thrombus in an atherosclerotic right coronary artery. What went wrong? Presumably, the ED clinician felt that the patient was basically healthy (one can be misled by the way the patient appears on examination—a patient that does not appear "sick" may be incorrectly assumed to have an innocuous illness). So, in this case, the physician, upon hearing the overview of the patient from the triage nurse, elected to use the URI assessment protocol even before starting the history, closing consideration of the broader range of possibilities and associated tests required to confirm or refute these possibilities. Specifically, by concentrating on the abbreviated and focused URI protocol, the clinician failed to elicit the full dyspnea history, which was precipitated by exertion and accompanied by chest heaviness and relieved by rest, suggesting a far more serious disorder.

Heuristics or rules of thumb are a part of the intuitive system. These cognitive shortcuts provide a quick and easy path to reaching conclusions and making choices, but when used improperly, they can lead to errors. Two major research programs have studied heuristics in a mostly nonmedical context and have reached different conclusions about the value of these cognitive tools. The "heuristics and biases" program focuses on how these mental shortcuts can lead to incorrect judgments. So far, however, little evidence exists that educating physicians and other decision makers to watch for the >100 cognitive biases identified to date has had any effect on the rate of diagnostic errors. In contrast, the "fast and frugal heuristics" research program explores how and when relying on simple heuristics can produce good decisions. Although many heuristics have relevance to clinical reasoning, only four will be mentioned here.

When diagnosing patients, clinicians usually develop diagnostic hypotheses based on the similarity of that patient's symptoms, signs, and other data to their mental representations (memorized patterns) of the disease possibilities. In other words, clinicians pattern match to identify the diagnoses that share the most similar findings to the patient at hand. This cognitive shortcut is called the representativeness heuristic. Consider a patient with hypertension who has headache, palpitations, and diaphoresis. Given this classic presenting symptom triad suggesting pheochromocytoma, clinicians might judge pheochromocytoma to be quite likely based on the representativeness heuristic. Doing so, however, would be incorrect given that other causes of hypertension are much more common than pheochromocytoma and this triad of symptoms can occur in patients who do not have it. Thus, clinicians using the representativeness heuristic may overestimate the likelihood of a particular disease based on the presence of representative symptoms and signs, failing to account for its low underlying prevalence (i.e., the prior, or pretest, probabilities). Conversely, atypical presentations of common diseases may lead to underestimating the likelihood of a particular disease. Thus, inexperience with a specific disease and with the breadth of its presentations may also lead to diagnostic delays or errors, e.g., diseases that affect multiple organ systems, such as sarcoid or tuberculosis, may be particularly challenging to diagnose because of the many different patterns they may manifest.

A second commonly used cognitive shortcut, the availability heuristic, involves judgments based on how easily prior similar cases or outcomes can be brought to mind. For example, a clinician may recall a case from a morbidity and mortality conference in which an elderly patient presented with painless dyspnea of acute onset and was evaluated for a pulmonary cause but was eventually found to have acute MI, with the diagnostic delay likely contributing to the development of ischemic cardiomyopathy. If the case was associated with a malpractice accusation, such examples may be even more memorable. Errors with the availability heuristic arise from several sources of recall bias. Rare catastrophic outcomes become memorable cases with a clarity and force disproportionate to their likelihood for future diagnosis—for example, a patient with a sore throat eventually found to have leukemia or a young athlete with leg pain subsequently found to have an osteosarcoma—and those publicized in the media or recently experienced are, of course, easier to recall and therefore more influential on clinical judgments.

The third commonly used cognitive shortcut, the anchoring heuristic (also called conservatism or stickiness), involves insufficiently adjusting the initial probability of disease up (or down) following a positive (or negative test) when compared with Bayes' theorem, i.e., sticking to the initial diagnosis. For example, a clinician may still judge the probability of coronary artery disease (CAD) to be high despite a negative exercise perfusion test and go on to cardiac catheterization (see "Measures of Disease Probability and Bayes' Rule," below).

The fourth heuristic states that clinicians should use the simplest explanation possible that will adequately account for the patient's symptoms and findings (Occam's razor or, alternatively, the simplicity heuristic). Although this is an attractive and often used principle, it is important to remember that no biologic basis for it exists. Errors from the simplicity heuristic include premature closure leading to the neglect of unexplained significant symptoms or findings.

For complex or unfamiliar diagnostic problems, clinicians typically resort to analytic reasoning processes (System 2) and proceed methodically using the *hypothetico-deductive model of reasoning*. Based on the patient's stated reasons for seeking medical attention, clinicians develop an initial list of diagnostic possibilities in *hypothesis generation*. During the history of the present illness, the initial hypotheses evolve in *diagnostic refinement* as emerging information is tested against the mental models of the diseases being considered with diagnoses increasing and decreasing in likelihood or even being dropped from or added to consideration as the working hypotheses of the moment. These mental models often generate additional questions that distinguish the diagnostic possibilities from one another. The focused physical examination contributes to further distinguishing the working hypotheses. Is the spleen enlarged? How big is the liver? Is it tender? Are there any palpable masses or nodules? *Diagnostic verification* involves testing the adequacy (whether the diagnosis accounts for all symptoms and signs) and coherency (whether the signs and symptoms are consistent with the underlying pathophysiologic causal mechanism) of the working diagnosis. For example, if the enlarged and quite tender liver felt on physical examination is due to acute hepatitis (the hypothesis), then certain specific liver function tests will be markedly elevated (the prediction). Should the tests come back normal, the hypothesis may have to be discarded and others reconsidered.

Although often neglected, negative findings are as important as positive ones because they reduce the likelihood of the diagnostic hypotheses under consideration. Chest discomfort that is not provoked or worsened by exertion and not relieved by rest in an active patient lowers the likelihood that chronic ischemic heart disease is the underlying cause. The absence of a resting tachycardia and thyroid gland enlargement reduces the likelihood of hyperthyroidism in a patient with paroxysmal atrial fibrillation.

The acuity of a patient's illness may override considerations of prevalence and the other issues described above. "Diagnostic imperatives" recognize the significance of relatively rare but potentially catastrophic conditions if undiagnosed and untreated. For example, clinicians should consider aortic dissection routinely as a possible cause of acute severe chest discomfort. Although the typical presenting symptoms of dissection differ from those of MI, dissection may mimic MI, and because it is far less prevalent and potentially fatal if mistreated, diagnosing dissection remains a challenging diagnostic imperative **(Chap. 291)**. Clinicians taking care of acute, severe chest pain patients should explicitly and routinely inquire about symptoms suggestive of dissection, measure blood pressures in both arms for discrepancies, and examine for pulse deficits. When these are all negative, clinicians may feel sufficiently reassured to discard the aortic dissection hypothesis. If, however, the chest x-ray shows a possible widened mediastinum, the hypothesis should be reinstated and an appropriate imaging test ordered (e.g., thoracic CT angiography or transesophageal echocardiogram). In nonacute situations, the prevalence of potential alternative diagnoses should play a much more prominent role in diagnostic hypothesis generation.

Cognitive scientists studying the thought processes of expert clinicians have observed that clinicians group data into packets, or "chunks," that are stored in short-term or "working memory" and manipulated to generate diagnostic hypotheses. Because short-term memory is limited (classically humans can accurately repeat a list of 7 ± 2 numbers read to them), the number of diagnoses that can be actively considered in hypothesis-generating activities is similarly limited. For this reason, the cognitive shortcuts discussed above play a key role in the generation of diagnostic hypotheses, many of which are discarded as rapidly as they are formed, thereby demonstrating that the distinction between analytic and intuitive reasoning is an arbitrary and simplistic, but nonetheless useful, representation of cognition.

Research into the hypothetico-deductive model of reasoning has had difficulty identifying the elements of the reasoning process that distinguish experts from novices. This has led to a shift from examining the problem-solving process of experts to analyzing the organization of their knowledge for pattern matching as exemplars, prototypes, and illness scripts. For example, diagnosis may be based on the resemblance of a new case to patients seen previously (exemplars). As abstract mental models of disease, prototypes incorporate the likelihood of various disease features. Illness scripts include risk factors, pathophysiology, and symptoms and signs. Experts have a much larger store of exemplar and prototype cases, an example of which is the visual long-term memory of experienced radiologists. However, clinicians do not simply rely on literal recall of specific cases but have constructed elaborate conceptual networks of memorized information or models of disease to aid in arriving at their conclusions (illness scripts). That is, expertise involves an enhanced ability to connect symptoms, signs, and risk factors to one another in meaningful ways; relate those findings to possible diagnoses; and identify the additional information necessary to confirm the diagnosis.

No single theory accounts for all the key features of expertise in medical diagnosis. Experts have more knowledge about presenting symptoms of diseases and a larger repertoire of cognitive tools to employ in problem solving than nonexperts. One definition of expertise highlights the ability to make powerful distinctions. In this sense, expertise involves a working knowledge of the diagnostic possibilities and those features that distinguish one disease from another. Memorization alone is insufficient, e.g., photographic memory of a medical textbook would not make one an expert. But having access to detailed case-specific relevant information is critically important. In the past, clinicians primarily acquired clinical knowledge through their patient experiences, but now clinicians have access to a plethora of information sources. Clinicians of the future will be able to leverage the experiences of large numbers of other clinicians using electronic tools, but, as with the memorized textbook, the data alone will be insufficient to create expertise.

Despite all the research seeking to understand expertise in medicine and other disciplines, it remains uncertain whether any didactic program can actually accelerate the progression from novice to expert or from experienced clinician to master clinician. Deliberate effortful practice (over an extended period of time, sometimes said to be 10 years or 10,000 practice hours) and personal coaching are two strategies often used outside medicine (e.g., music, athletics, chess) to develop expertise. Their use in the context of medical practice has not yet been adequately explored. Some studies in medicine suggest that the most beneficial approach to education exposes students to both the signs and symptoms of specific diseases (disease pattern recognition) and, in addition, the lists of diseases that can present with specific symptoms and signs (differential diagnosis). Active learning opportunities useful for those in training include developing a personal learning system, e.g., systematically reflecting on diagnostic processes used (metacognition) and following up to identify diagnoses and treatments for patients in their care.

PERSONALIZED DECISION-MAKING

The modern ideal of medical therapeutic decision-making is to "personalize" treatment recommendations. In the abstract, personalizing treatment involves combining the best available evidence about what works with an individual patient's unique features (e.g., risk factors, genomics, and comorbidities) and their preferences and health goals to craft an optimal treatment recommendation with the patient. Operationally, two different and complementary levels of personalization are possible: individualizing the risk of harm and benefit for the options being considered based on the specific patient characteristics (precision medicine) and personalizing the therapeutic decision process by incorporating the patient's preferences and values for the possible health outcomes. This latter process is sometimes referred to as shared decision-making and typically involves clinicians sharing their knowledge about the options and the associated consequences and trade-offs and patients sharing their health goals (e.g., avoiding a short-term risk of dying from coronary artery bypass grafting to see their grandchild get married in a few months).

Individualizing the evidence about therapy **does not** mean relying on physician impressions of benefit and harm from their personal experience. Because of nonrandom selection, small sample sizes, and rare events, the chance of drawing erroneous causal inferences from one's own clinical experience is very high. For most chronic diseases, the treatment response is a counterfactual concept, only demonstrable statistically in large patient populations. Because of this, it would be incorrect to infer with any certainty, for example, that treating a hypertensive patient with angiotensin-converting enzyme (ACE) inhibitors necessarily prevented a stroke from occurring during treatment, or that an untreated patient would definitely have avoided their stroke had they been treated. For many chronic diseases, a majority of patients will remain event free over long periods of time regardless of treatment choices; some will have events regardless of which treatment is selected; and those who avoided having an event through treatment cannot be individually identified. Blood pressure lowering, a readily observable surrogate endpoint, does not have a tightly coupled relationship with strokes prevented. Consequently, in most situations, demonstrating therapeutic effectiveness cannot rely simply on observing the outcome of an individual patient but should instead be based on large groups of patients carefully studied and properly analyzed.

Therapeutic decision-making, therefore, should be based on the best available evidence from clinical trials and well-done outcome studies. Trustworthy clinical practice guidelines that synthesize such evidence offer normative guidance for many testing and treatment decisions. However, all guidelines recognize that "one size fits all" recommendations may not apply to individual patients. Increased research into the heterogeneity of treatment effects seeks to understand how best to adjust group-level clinical evidence of treatment harms and benefits to account for the absolute level of risks faced by subgroups and even by individual patients, using, for example, validated clinical risk scores.

NONCLINICAL INFLUENCES ON CLINICAL DECISION-MAKING

More than three decades of research on variations in clinician practice patterns have identified important nonclinical forces that shape clinical decisions. These factors can be grouped conceptually into three overlapping categories: (1) factors related to an individual physician's practice, (2) factors related to practice setting, and (3) factors related to payment systems.

Practice Style To ensure that necessary care is provided at a high level of quality, physicians fulfill a key role in medical care by serving as the patient's advocate. Factors that influence performance in this role include the physician's knowledge, training, and experience. Clearly, physicians cannot practice EBM if they are unfamiliar with the evidence. As would be expected, specialists generally know the evidence in their field better than do generalists. Beyond published evidence and practice guidelines, a major set of influences on physician practice can be subsumed under the general concept of "practice style." The practice style serves to define norms of clinical behavior. Differing practice styles may be based on training, personal experience, and medical evidence. Beliefs about effectiveness of different therapies and preferred patterns of diagnostic test use are examples of different facets of a practice style. For example, cardiologists evaluating patients with lower risk chest pain symptoms often conceptualize their primary diagnostic objective as maximizing the detection of ischemia. For this reason, they may strongly favor stress imaging. Internists caring for the same patients may be more comfortable with initial use of exercise electrocardiogram (ECG) testing without imaging. This latter practice style focuses less on ischemia detection and more on following guideline recommendations that indicate no outcome advantage for stress imaging in this context. Cardiologists, relative to general internists, may also favor a more liberal use of coronary angiography and revascularization in patients with stable ischemic symptoms, i.e., the "oculostenotic reflex."

Beyond the patient's welfare, physician perceptions about the risk of a malpractice suit resulting from either an erroneous decision or a bad outcome may drive clinical decisions and create a practice referred to as defensive medicine. This practice involves ordering tests and therapies with very small marginal benefits, ostensibly to preclude future criticism should an adverse outcome occur. Over time, such patterns of care may become accepted as part of the practice norm, thereby perpetuating their overuse, e.g., annual cardiac exercise testing in asymptomatic patients.

Practice Setting Factors in this category relate to work systems including tasks and workflow (e.g., interruptions, inefficiencies, workload), technology (e.g., electronic health record design or implementation issue,), organizational characteristics (e.g., culture, leadership, staffing, scheduling), and the physical environment (e.g., noise, lighting, layout). *Physician-induced demand* is a term that refers to the repeated observation that once medical facilities and technologies become available to physicians, they will find ways to use them. Other environmental factors that can influence decision-making include the local availability of specialists for consultations and procedures; "high-tech" advanced imaging or procedure facilities such as magnetic resonance imaging (MRI) machines and proton beam therapy centers; and fragmentation of care.

Payment Systems Economic incentives are closely related to the other two categories of practice-modifying factors. Financial issues can exert both stimulatory and inhibitory influences on clinical practice. Historically, physicians have been paid on a fee-for-service, capitation, or salary basis. In fee-for-service, physicians who do more generally get paid more, thereby encouraging overuse, consciously or unconsciously. When fees are reduced (discounted reimbursement), clinicians tend to increase the number of services provided to maintain revenue. Capitation, in contrast, provides a fixed payment per patient per year to encourage physicians to consider a global population budget in managing individual patients and ideally reducing the use of interventions with small marginal benefit. In recognition of the unsustainability of continued growth in medical expenditures and the opportunity costs associated with that (funds that might be more beneficially applied to education, energy, social welfare, or defense), current efforts seek to transition to a value-based payment system to reduce overuse and to reflect benefit. Work to define how to tie payment to value has mostly focused so far on "pay for performance" models. High-quality clinical trial evidence for the effectiveness of these models is still mostly lacking.

DIAGNOSTIC TEST PERFORMANCE: UNDERSTANDING TEST ACCURACY

The purpose of performing a test on a patient is to reduce uncertainty about the patient's diagnosis or prognosis to facilitate appropriate management. Although diagnostic tests commonly refer to laboratory (e.g., blood count) or imaging tests or procedures (e.g., colonoscopy or bronchoscopy), any information that changes a clinician's understanding of the patient's problem qualifies as a diagnostic test. Thus, even the history and physical examination can be considered as diagnostic tests. In clinical medicine, it is common to reduce the results of a test to a dichotomous outcome, such as positive or negative, normal or abnormal. Although this simplification often suppresses useful information (such as the degree of abnormality), it facilitates illustrating some important principles of test interpretation that are described below.

The accuracy of any diagnostic test is best assessed relative to a "gold standard," where a positive gold standard test defines the patients who have disease and a negative test securely rules out disease **(Table 4-1)**. Characterizing the diagnostic performance of a new test requires identifying an appropriate population (ideally, patients representative of those in whom the new test would be used) and applying both the new and the gold standard tests to all subjects. Biased estimates of test performance occur when diagnostic accuracy is defined using an inappropriate population or one in which gold standard determination of disease status is incomplete. The accuracy of the new test in distinguishing disease from health is determined relative to the gold standard results and summarized in four estimates. The sensitivity or true-positive rate reflects how well the new test identifies patients with disease. It is the proportion of patients with disease (defined by the gold standard) who have a positive test. The proportion of patients with disease who have a negative test is the false-negative rate, calculated as

TABLE 4-1 Measures of Diagnostic Test Accuracy

	DISEASE STATUS	
TEST RESULT	PRESENT	ABSENT
Positive	True positives (TP)	False positives (FP)
Negative	False negatives (FN)	True negatives (TN)
Test Characteristics in Patients with Disease		
True-positive rate (sensitivity) = TP/(TP + FN)		
False-negative rate = FN/(TP + FN) = 1 – true-positive rate		
Test Characteristics in Patients without Disease		
True-negative rate (specificity) = TN/(TN + FP)		
False-positive rate = FP/(TN + FP) = 1 – true-negative rate		

1 – sensitivity. The specificity, or true-negative rate, reflects how well the new test correctly identifies patients without disease. It is the proportion of patients without disease (defined by the gold standard) who have a negative test. The proportion of patients without disease who have positive test is the false-positive rate, calculated as 1 – specificity. A theoretically perfect test then would have a sensitivity of 100% and a specificity of 100% and would completely distinguish patients with disease from those without it. A useful mnemonic to help remember the somewhat paradoxical relationship between what the test is best at technically versus what it is most useful for clinically is: a test with a very high sensitivity (Sn) when *negative* (N) helps *rule out* (out) disease (SnNout), and a test with a very high specificity (Sp) when *positive* (P) helps *rule in* (in) disease (SpPin).

Calculating sensitivity and specificity requires selection of a threshold value or cut point above which the test is considered "positive." Making the cut point "stricter" (e.g., raising it) lowers sensitivity but improves specificity, while making it "laxer" (e.g., lowering it) raises sensitivity but lowers specificity. This dynamic trade-off between more accurate identification of patients with disease versus those without disease is often displayed graphically as a receiver operating characteristic (ROC) curve **(Fig. 4-1)** by plotting sensitivity (*y* axis) versus 1 – specificity (*x* axis). Each point on the curve represents a potential cut point with an associated sensitivity and specificity value. The area under the ROC curve often is used as a quantitative measure of the information content of a test. Values range from 0.5 (no diagnostic information from testing at all; the test is equivalent to flipping a coin) to 1.0 (perfect test). The choice of cut point should ideally reflect the relative harms and benefits of treatment for those without versus those with disease. For example, if treatment was safe with substantial benefit, then choosing a high-sensitivity cut point (upper right of the ROC curve) for a low-risk test may be appropriate (e.g., phenylketonuria in newborns), but if treatment had substantial risk for harm, then choosing a high-specificity cut point (lower left of the ROC curve) may be appropriate (e.g., chemotherapy for cancer). The choice of cut point may also depend on the prevalence of disease, with low prevalence placing a greater emphasis on the harms of false-positive tests (e.g., HIV testing in marriage applicants) or the harms of false-negative tests (e.g., HIV testing in blood donors).

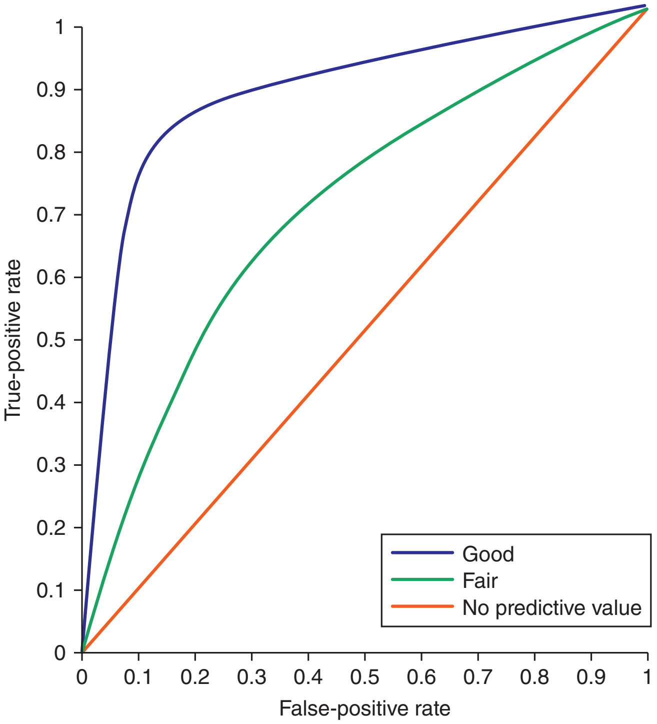

FIGURE 4-1 Each receiver operating characteristic (ROC) curve illustrates a trade-off that occurs between improved test sensitivity (accurate detection of patients with disease) and improved test specificity (accurate detection of patients without disease), as the test value defining when the test turns from "negative" to "positive" is varied. A 45° line would indicate a test with no predictive value (sensitivity = specificity at every test value). The area under each ROC curve is a measure of the information content of the test. Thus, a larger ROC area signifies increased diagnostic accuracy.

MEASURES OF DISEASE PROBABILITY AND BAYES' RULE

In the absence of perfect tests, the true disease state of the patient remains uncertain after every test. Bayes' rule provides a way to quantify the revised uncertainty using simple probability mathematics (and thereby avoid anchoring bias). It calculates the *posttest probability*, or likelihood of disease after a test result, from three parameters: the pretest probability of disease, the test sensitivity, and the test specificity. The *pretest probability* is a quantitative estimate of the likelihood of the diagnosis before the test is performed and is usually estimated from the prevalence of the disease in the underlying population (if known) or clinical context (e.g., age, sex, and type of chest pain). For some common conditions, such as CAD, existing nomograms and statistical models generate estimates of pretest probability that account for history, physical examination, and test findings. The posttest probability (also called the predictive value of the test, see below) is a recalibrated statement of the probability of the diagnosis, accounting for both pretest probability and test results. For the probability of disease following a positive test (i.e., positive predictive value), Bayes' rule is calculated as:

$$\text{Posttest probability} = \frac{\text{Pretest probability} \times \text{test sensitivity}}{\text{Pretest probability} \times \text{test sensitivity} + (1 - \text{Pretest probability}) \times (\text{false-positive test rate})}$$

For example, consider a 64-year-old woman with atypical chest pain who has a pretest probability of 0.50 and a "positive" diagnostic test result (assuming test sensitivity = 0.90 and specificity = 0.90).

$$\text{Posttest probability} = \frac{(0.50)(0.90)}{(0.50)(0.90) + (0.50)(0.10)} = 0.90$$

The term *predictive value* has often been used as a synonym for the posttest probability. Unfortunately, clinicians commonly misinterpret reported predictive values as intrinsic measures of test accuracy rather than calculated probabilities. Studies of diagnostic test performance compound the confusion by calculating predictive values from the same sample used to measure sensitivity and specificity. Such calculations are misleading unless the test is applied subsequently to populations with exactly the same disease prevalence. For these reasons, the more descriptive term, *posttest probability following a positive or a negative test,* is preferred over *predictive value.*

The nomogram version of Bayes' rule **(Fig. 4-2)** helps us to understand at a conceptual level how it estimates the posttest probability of disease. In this nomogram, the impact of the diagnostic test result is summarized by the likelihood ratio, which is defined as the ratio of the probability of a given test result (e.g., "positive" or "negative") in a

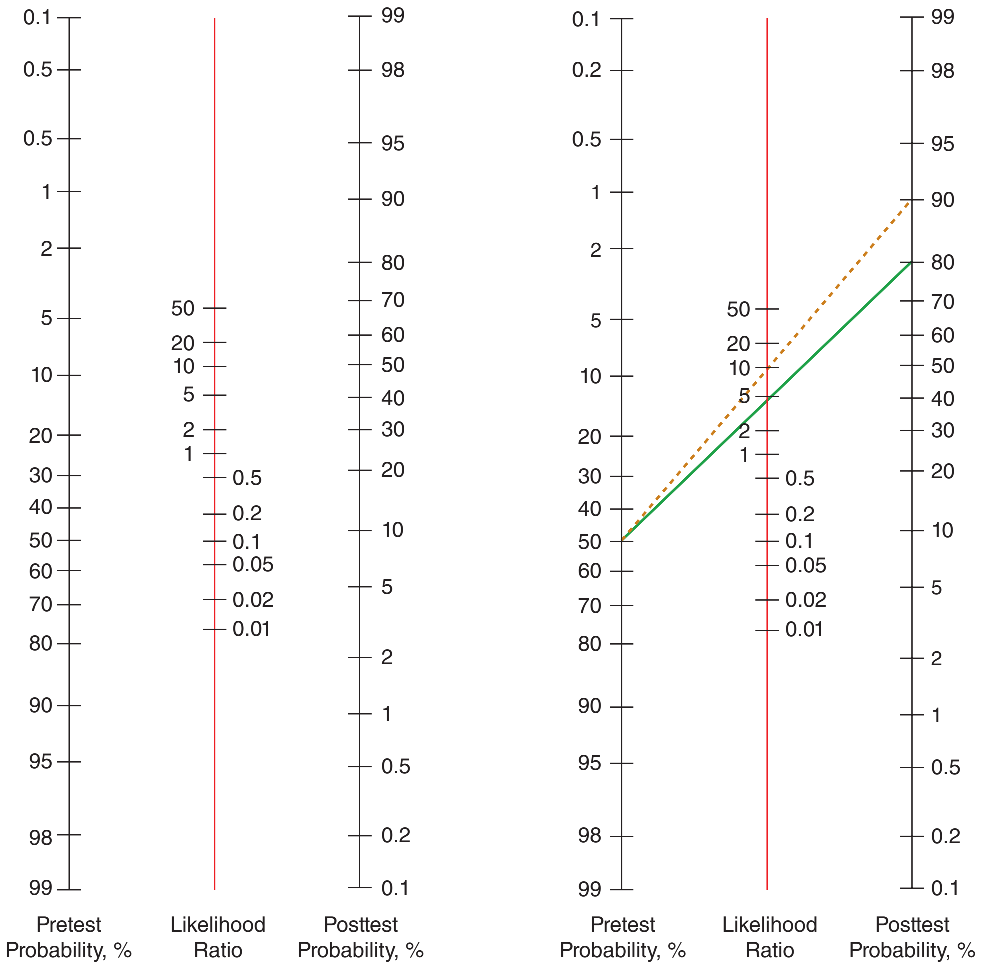

FIGURE 4-2 Nomogram version of Bayes' theorem used to predict the posttest probability of disease (right-hand scale) using the pretest probability of disease (left-hand scale) and the likelihood ratio for a positive or a negative test (middle scale). See text for information on calculation of likelihood ratios. To use, place a straightedge connecting the pretest probability and the likelihood ratio and read off the posttest probability. The right-hand part of the figure illustrates the value of a positive exercise treadmill test (likelihood ratio 4, *green line*) and a positive exercise thallium single-photon emission CT perfusion study (likelihood ratio 9, broken *brown line*) in a patient with a pretest probability of coronary artery disease of 50%. *(Adapted from Centre for Evidence-Based Medicine: Likelihood ratios. Available at http://www.cebm.net/likelihood-ratios/.)*

patient with disease to the probability of that result in a patient without disease, thereby providing a measure of how well the test distinguishes those with from those without disease.

The *likelihood ratio for a positive test* is calculated as the ratio of the true-positive rate to the false-positive rate (or sensitivity/[1 – specificity]). For example, a test with a sensitivity of 0.90 and a specificity of 0.90 has a likelihood ratio of 0.90/(1 – 0.90), or 9. Thus, for this hypothetical test, a "positive" result is 9 times more likely in a patient with the disease than in a patient without it. Most tests in medicine have likelihood ratios for a positive result between 1.5 and 20. Higher values are associated with tests that more substantially increase the posttest likelihood of disease. A very high likelihood ratio positive (>10) usually implies high specificity, so a positive high-specificity test helps "rule in" disease (the "SpPin" mnemonic introduced earlier). If sensitivity is excellent but specificity is less so, the likelihood ratio positive will be reduced substantially (e.g., with a 90% sensitivity but a 55% specificity, the likelihood ratio positive is 2.0).

The corresponding *likelihood ratio for a negative test* is the ratio of the false-negative rate to the true-negative rate (or [1 – sensitivity]/specificity). Lower likelihood ratio negative values more substantially lower the posttest likelihood of disease. A very low likelihood ratio negative (falling below 0.10) usually implies high sensitivity, so a negative high-sensitivity test helps "rule out" disease (the SnNout mnemonic). The hypothetical test considered above with a sensitivity of 0.9 and a specificity of 0.9 would have a likelihood ratio for a negative test result of (1 – 0.9)/0.9, or 0.11, meaning that a negative result is about one-tenth as likely in patients with disease than in those without disease (or about 10 times more likely in those without disease than in those with disease).

APPLICATIONS TO DIAGNOSTIC TESTING IN CAD

Consider two tests commonly used in the diagnosis of CAD: an exercise treadmill and an exercise single-photon emission CT (SPECT) myocardial perfusion imaging test **(Chap. 248)**. A positive treadmill ST-segment response has an average sensitivity of ~60% and an average specificity of ~75%, yielding a likelihood ratio positive of 2.4 (0.60/[1 – 0.75]) (consistent with modest discriminatory ability because it falls between 2 and 5). A 41-year-old man with atypical chest pain and no other risk factors has about a 10% pretest probability of CAD. After a positive result, the posttest probability of disease rises to only ~30%. For a 60-year-old man with typical angina and multiple risk factors, the pretest probability of CAD is about 80%. After a positive test result, the posttest probability of disease rises to ~95%.

In contrast, exercise SPECT myocardial perfusion test is more accurate for diagnosis of CAD. For simplicity, assume that the finding of a reversible exercise-induced perfusion defect has both a sensitivity and a specificity of 90% (a bit higher than reported), yielding a likelihood ratio for a positive test of 9.0 (0.90/[1 – 0.90]) (consistent with intermediate discriminatory ability because it falls between 5 and 10). For the same 10% pretest probability patient, a positive test raises the probability of CAD to 50% (Fig. 4-2). However, despite the differences in posttest probabilities between these two tests (30 vs 50%), the more accurate test may not improve diagnostic likelihood enough to change patient management (e.g., decision to refer to cardiac catheterization) because the more accurate test has only moved the physician from being fairly certain that the patient did not have CAD to a 50:50 chance of disease. In a patient with a pretest probability of 80%, exercise SPECT test raises the posttest probability to 97% (compared with 95% for the exercise treadmill). Again, the more accurate test does not provide enough improvement in posttest confidence to alter management, and neither test has improved much on what was known from clinical data alone.

In general, positive results with an accurate test (e.g., likelihood ratio for a positive test of 10) when the pretest probability is low (e.g., 20%) do not move the posttest probability to a range high enough to rule in disease (e.g., 80%). In screening situations, pretest probabilities are often particularly low because patients are asymptomatic. In such cases, specificity becomes especially important. For example, in screening first-time female blood donors without risk factors for HIV, a positive test raised the likelihood of HIV to only 67% despite a specificity of 99.995% because the prevalence was 0.01%. Conversely, with a high pretest probability, a negative test may not rule out disease adequately if it is not sufficiently sensitive. Thus, the largest change in diagnostic likelihood following a test result occurs when the clinician is most uncertain (i.e., pretest probability between 30 and 70%). For example, a 70-year-old woman with typical angina and multiple risk factors has a pretest probability for CAD of ~50%. A positive exercise treadmill test moves the posttest probability to 80%, and a positive exercise SPECT perfusion test moves it to 90% (Fig. 4-2).

As presented above, Bayes' rule employs a number of important simplifications that should be considered. First, few tests provide only "positive" or "negative" results. Many tests have multidimensional outcomes (e.g., extent of ST-segment depression, exercise duration, and exercise-induced symptoms with exercise testing). Although Bayes' theorem can be adapted to this more detailed test result format, it is computationally more complex to do so. Similarly, when multiple sequential tests are performed, the posttest probability may be used as the pretest probability to interpret the second test. However, this simplification assumes conditional independence—that is, that the results of the first test do not affect the likelihood of the second test result—and this is often not true.

Finally, many texts assert that sensitivity and specificity are prevalence-independent parameters of test accuracy. This statistically useful assumption, however, is often incorrect. A treadmill exercise test, for example, has a sensitivity of ~30% in a population of patients with one-vessel CAD, whereas its sensitivity in patients with severe three-vessel CAD approaches 80%. Thus, the best estimate of sensitivity to use in a particular decision may vary, depending on the severity of disease in the local population. A hospitalized, symptomatic, or referral population typically has a higher prevalence of disease and, importantly, a higher prevalence of more advanced disease than does an outpatient population. Consequently, test sensitivity will likely be higher in hospitalized patients and test specificity higher in outpatients.

RISK PREDICTION MODELS

Bayes' rule, when used as presented above, is useful in studying diagnostic testing concepts, but predictions based on multivariable statistical models can more accurately address these more complex problems by simultaneously accounting for additional relevant patient characteristics. These models explicitly account for multiple, even possibly overlapping, pieces of patient-specific information and assign a relative weight to each on the basis of its unique independent contribution to the prediction in question. For example, a logistic regression model to predict the probability of CAD ideally considers all the relevant independent factors from the clinical examination and diagnostic testing and their relative importance instead of the limited data that clinicians can manage in their heads or with Bayes' rule. However, despite this strength, prediction models are usually too complex computationally to use without a calculator or computer. Guideline-driven treatment recommendations based on statistical prediction models available online, e.g., the American College of Cardiology/American Heart Association risk calculator for primary prevention with statins and the CHA_2DS_2-VASC calculator for anticoagulation for atrial fibrillation, have generated more widespread usage. Some predictive models are now embedded into electronic health record (EHR) systems, most commonly addressing issues related to thrombosis/anticoagulation and to sepsis. Evidence about the impact of these EHR-based models on patient outcomes is mostly observational and suggests more work is needed to deliver the risk information to the right clinician at the right time in a way that supports clinical workflow.

One reason for limited clinical use is that, to date, only a handful of prediction models have been validated sufficiently (e.g., Wells criteria for pulmonary embolism; Table 4-2). The importance of independent validation in a population separate from the one used to develop the model cannot be overstated. An unvalidated risk prediction model should be viewed with the skepticism appropriate for any new drug or medical device that has not had rigorous clinical trial testing.

When statistical survival models in cancer and heart disease have been compared directly with clinicians' predictions, the survival models have been found to be more consistent, as would be expected, but not always more accurate. On the other hand, comparison of clinicians with websites and apps that generate lists of possible diagnoses to help patients with self-diagnosis found that physicians outperformed the currently available programs. For students and less-experienced clinicians, the biggest value of diagnostic decision support may be in extending diagnostic possibilities and triggering "rational override," but their impact on knowledge, information-seeking, and problem-solving needs additional research.

TABLE 4-2 Wells Clinical Prediction Rule for Pulmonary Embolism (PE)

CLINICAL FEATURE	POINTS
Clinical signs of deep-vein thrombosis	3
Alternative diagnosis is less likely than PE	3
Heart rate >100 beats/min	1.5
Immobilization ≥3 days or surgery in previous 4 weeks	1.5
History of deep-vein thrombosis or PE	1.5
Hemoptysis	1
Malignancy (with treatment within 6 months) or palliative	1
INTERPRETATION	
Score >6.0	High
Score 2.0–6.0	Intermediate
Score <2.0	Low

FORMAL DECISION SUPPORT TOOLS

DECISION SUPPORT SYSTEMS AND ARTIFICIAL INTELLIGENCE

Over the past 50 years, many attempts have been made to develop computer systems to aid clinical decision-making and patient management. Conceptually, computers offer several levels of potentially useful support for clinicians. At the most basic level, they provide ready access to vast reservoirs of information, which may, however, be quite difficult to sort through to find what is needed. At higher levels, computers can support care management decisions by making accurate predictions of outcome, or can simulate the whole decision process, and provide algorithmic guidance. Computer-based predictions using Bayesian or statistical regression models inform a clinical decision but do not actually reach a "conclusion" or "recommendation." Recent advances in artificial intelligence (AI) suggest that medicine is on the threshold of developing much more powerful digital tools, but current enthusiasm for such tools still exceeds demonstrated utility in clinical care. Work on AI dates back to the 1950s and can be separated into three major subtypes: neural networks, machine learning (and its subtype deep learning), and generative AI. Machine learning methods are being applied to pattern recognition tasks such as the examination of skin lesions and the interpretation of x-rays. Generative AI (AI models that generate new content) is a term that covers several different types of systems, including large language models (which generate language-based content; an example is GPT-4). Large language models offer some promise in helping to create clinical notes. Their use in support of clinical decision-making, however, is still at a very preliminary stage with need for independent validation in a population separate from the one used to develop the model. Early evidence suggests that clinicians are willing to rely on AI-based tools even when the information provided is clearly inaccurate or contradictory. Concerns about model confabulation and the potential for patient harms mandate careful and comprehensive testing before AI tools are implemented in clinical care.

Reminder or protocol-directed systems do not make predictions but use existing algorithms, such as guidelines or appropriate utilization criteria, to direct clinical practice. In general, however, decision support systems have so far had little impact on practice. Reminder systems built into EHRs have shown the most promise, particularly in correcting drug dosing and promoting adherence to guidelines. Checklists may also help avoid or reduce errors.

DECISION ANALYSIS

Compared with the decision support methods discussed earlier, decision analysis represents a normative prescriptive approach to decision-making in the face of uncertainty. Its principal application is in complex decisions. For example, public health policy decisions often involve *trade-offs* in length versus quality of life, benefits versus resource use, population versus individual health, and *uncertainty* regarding efficacy, effectiveness, and adverse events as well as *values* or preferences regarding mortality and morbidity outcomes.

One recent analysis using this approach involved the optimal screening strategy for breast cancer, which has remained controversial, in part because a randomized controlled trial to determine when to begin screening and how often to repeat screening mammography is impractical. In 2016, the National Cancer Institute–sponsored Cancer Intervention and Surveillance Network (CISNET) examined eight strategies differing by whether to initiate mammography screening at age 40, 45, or 50 years and whether to screen annually, biennially, or annually for women in their forties and biennially thereafter (hybrid). The six simulation models found biennial strategies to be the most efficient for average-risk women. Biennial screening for 1000 women from age 50–74 years versus no screening avoided seven breast cancer deaths. Screening annually from age 40–74 years avoided three additional deaths but required 20,000 additional mammograms and yielded 1988 more false-positive results. Factors that influenced the results included patients with a two- to fourfold higher risk for developing breast cancer in whom annual screening from age 40–74 years yielded similar benefits as biennial screening from age 50–74. For average-risk patients with moderate or severe comorbidities, screening could be stopped earlier, at age 66–68 years.

This analysis involved six models that reproduced epidemiologic trends and a screening trial result, accounted for digital technology and treatments advances, and considered quality of life, risk factors, breast density, and comorbidity. It provided novel insights into a public health problem in the absence of randomized clinical trials examining alternative start age, stop age, and screening frequencies, and helped weigh the pros and cons of such a health policy recommendation. Although such models have been developed for selected clinical problems, their benefit and application to individual real-time clinical management has yet to be demonstrated.

DIAGNOSIS AS AN ELEMENT OF QUALITY OF CARE

High-quality medical care begins with accurate diagnosis. The incidence of diagnostic errors has been estimated by a variety of methods including postmortem examinations, medical record reviews, and medical malpractice claims, with each yielding complementary but different estimates of this quality of care patient-safety problem. In the past, diagnostic errors tended to be viewed as a failure of individual clinicians. The modern view is that they are mostly a system of care deficiencies. Current estimates suggest that nearly everyone will experience at least one diagnostic error in their lifetime, leading to mortality, morbidity, unnecessary tests and procedures, costs, and anxiety.

Solutions to the "diagnostic errors as a system of care" problem have focused on system-level approaches, such as decision support and other tools integrated into EHRs. The use of checklists has been proposed as a means of reducing some of the cognitive errors discussed earlier in the chapter, such as premature closure. While checklists have been shown to be useful in certain medical contexts, such as operating rooms and intensive care units, their value in preventing diagnostic errors that lead to patient adverse events remains to be shown.

EVIDENCE-BASED MEDICINE

Clinical medicine is defined traditionally as a practice combining medical knowledge (including scientific evidence), intuition, and judgment in the care of patients **(Chap. 1)**. EBM updates this construct by placing much greater emphasis on the processes by which clinicians gain knowledge of the most up-to-date and relevant clinical research to determine for themselves whether medical interventions alter the disease course and improve the length or quality of life. The phrase "evidence-based medicine" is now used so often and in so many different contexts that many practitioners are unaware of its original meaning. The intention of the EBM program, as described in the early 1990s by its founding proponents at McMaster University, becomes clearer through an examination of its four key steps:

1. Formulating the management question to be answered
2. Searching the literature and online databases for applicable research data
3. Appraising the evidence gathered with regard to its validity and relevance
4. Integrating this appraisal with knowledge about the unique aspects of the patient (including the patient's preferences about the possible outcomes)

The process of searching the world's research literature and appraising the quality and relevance of studies can be time-consuming and requires skills and training that most clinicians do not possess. In a busy clinical practice, the work required is also logistically not feasible. This has led to a focus on finding recent systematic overviews of the problem in question as a useful shortcut in the EBM process. Systematic reviews are regarded by some as the highest level of evidence in the EBM hierarchy because they are intended to comprehensively summarize the available evidence on a particular topic. To avoid the potential biases found in narrative review articles, predefined reproducible explicit search strategies and inclusion and exclusion criteria seek to find all of the relevant scientific research and grade its quality. The prototype for this kind of resource is the Cochrane Database of Systematic Reviews. When appropriate, a meta-analysis is used to quantitatively summarize the systematic review findings (discussed further below).

Unfortunately, systematic reviews are not uniformly the acme of the EBM process they were initially envisioned to be. In select circumstances, they can provide a much clearer picture of the state of the evidence than is available from any individual clinical report, but their value is less clear when only a few trials are available, when trials and observational studies are mixed, or when the evidence base is only observational. They cannot compensate for deficiencies in the underlying research available, and many are created without the requisite clinical insights. The medical literature is now flooded with systematic reviews of varying quality and clinical utility. The peer review system has, unfortunately, not proved to be an effective arbiter of quality of these papers. Therefore, systematic reviews should be used with circumspection in conjunction with selective reading of some of the best empirical studies.

■ SOURCES OF EVIDENCE: CLINICAL TRIALS AND REGISTRIES

The notion of learning from observation of patients is as old as medicine itself. Over the past 50 years, physicians' understanding of how best to turn raw observation into useful evidence has evolved considerably. Medicine has received a hard refresher lesson in this process from the COVID-19 pandemic. Starting in the spring of 2020, case reports, personal and institutional anecdotal experience, and small single-center case series started appearing in the peer-reviewed literature and within months turned into a flood of confusing and often contradictory evidence. Observational reports of treatments for COVID-19 fueled the confusion. Despite >40,000 publications appearing in the first 7 months of the pandemic, an enormous amount of uncertainty around prevention, diagnosis, treatment, and prognosis of the disease remained. Many of the early 2020 publications were either small observational series or reviews of published series, neither of which can resolve the key uncertainties clinicians need to address in caring for these patients. These small observational studies often have substantial limitations in validity and generalizability, and although they may generate important hypotheses or be the first reports of adverse events or therapeutic benefit, they have no role in formulating modern standards of practice. The major tools used to develop reliable evidence consist of randomized clinical trials supplemented strategically by large (high-quality) observational registries. A registry or database typically is focused on a disease or syndrome (e.g., different types of cancer, acute or chronic CAD, pacemaker capture, or chronic heart failure), a clinical procedure (e.g., bone marrow transplantation, coronary revascularization), or an administrative process (e.g., claims data used for billing and reimbursement).

By definition, in observational data, the investigator does not control patient care. Carefully collected prospective observational data, however, can at times achieve a level of evidence quality approaching that of major clinical trial data through trial emulation (specifying eligibility criteria, interventions, outcome, follow-up, causal contrast, and statistical analysis) using causal inference methods. At the other end of the spectrum, data collected retrospectively (e.g., chart review) are limited in form and content to what previous observers recorded and may not include the specific research data being sought (e.g., claims data). Advantages of observational data include the inclusion of a broader population as encountered in practice than is typically represented in clinical trials because of their restrictive inclusion and exclusion criteria. In addition, observational data provide primary evidence for research questions when a randomized trial cannot be performed. For example, it would be difficult to randomize patients to test diagnostic or therapeutic strategies that are unproven but widely accepted in practice, and it would be unethical to randomize based on sex, racial/ethnic group, socioeconomic status, or country of residence or to randomize patients to a potentially harmful intervention, such as smoking or deliberately overeating to develop obesity.

A well-done prospective observational study of a particular management strategy differs from a well-done randomized clinical trial most importantly by its lack of protection from treatment selection bias. The use of observational data to compare diagnostic or therapeutic strategies assumes that sufficient uncertainty and heterogeneity exists in clinical practice to ensure that similar patients will be managed differently by diverse physicians. In short, the analysis assumes that a sufficient element of randomness (in the sense of disorder rather than in the formal statistical sense) exists in clinical management. In such cases, statistical models attempt to adjust for important imbalances to "level the playing field" so that a fair comparison among treatment options can be made. When management is clearly not random (e.g., all eligible left main CAD patients are referred for coronary bypass surgery), the problem may be too confounded (biased) for statistical correction, and observational data may not provide reliable evidence.

In general, the use of concurrent controls is vastly preferable to that of historical controls. For example, comparison of current surgical management of left main CAD with medically treated patients with left main CAD during the 1970s (the last time these patients were routinely treated with medicine alone) would be extremely misleading because "medical therapy" has substantially improved in the interim.

Randomized controlled clinical trials include the careful prospective design features of the best observational data studies but also include the use of random allocation of treatment. This design provides the best protection against measured and unmeasured confounding due to treatment selection bias (a major aspect of internal validity). However, the randomized trial may not have good external validity (generalizability) if the process of recruitment into the trial resulted in the exclusion of many potentially eligible subjects or if the nominal eligibility for the trial describes a very heterogeneous population.

Consumers of medical evidence need to be aware that randomized trials vary widely in their quality and applicability to practice. The process of designing such a trial often involves many compromises. For example, trials designed to gain U.S. Food and Drug Administration (FDA) approval for an investigational drug or device must fulfill regulatory requirements (such as the use of a placebo control) that may result in a trial population and design that differ substantially from what practicing clinicians would find most useful.

META-ANALYSIS

The Greek prefix *meta* signifies something at a later or higher stage of development. Meta-analysis is research that combines and summarizes the available evidence quantitatively. Although it is used to examine nonrandomized studies, meta-analysis is most useful for summarizing all available randomized trials examining a particular therapy used in a specific clinical context. Ideally, unpublished trials should be identified and included to avoid publication bias (i.e., missing "negative" trials that may not be published). Furthermore, the best meta-analyses obtain and analyze individual patient-level data from all trials rather than using only the summary data from published reports. Nonetheless, not all published meta-analyses yield reliable evidence for a particular problem, so their methodology should be scrutinized carefully to ensure proper study design and analysis. The results of a well-done meta-analysis are likely to be most persuasive if they include at least several large-scale, properly performed randomized trials. Meta-analysis can especially help detect benefits when individual trials are inadequately powered (e.g., the benefits of streptokinase thrombolytic therapy in acute MI demonstrated by ISIS-2 in 1988 were evident by the early 1970s through meta-analysis). However, in cases in which the available trials are small or poorly done, meta-analysis should not be viewed as a remedy for deficiencies in primary trial data or trial design.

Meta-analyses typically focus on summary measures of relative treatment benefit, such as odds ratios or relative risks. Clinicians should also examine what absolute risk reduction (ARR) can be expected from the therapy. A metric of absolute treatment benefit that is frequently reported is the number needed to treat (NNT) to prevent one adverse outcome event (e.g., death, stroke). **NNT should not be interpreted literally as a causal statement.** NNT is simply 1/ARR. For example, if a hypothetical therapy reduced mortality rates over a 5-year follow-up by 33% (the relative treatment benefit) from 12% (control arm) to 8% (treatment arm), the ARR would be 12% – 8% = 4% and the NNT would be 1/0.04, or 25. This does not mean literally that 1 patient benefits and 24 do not. However, it can be conceptualized as an informal measure of treatment efficiency. If the hypothetical treatment was applied to a lower-risk population, say, with a 6% 5-year mortality, the 33% relative treatment benefit would reduce absolute mortality by 2% (from 6 to 4%), and the NNT for the same therapy in this lower-risk group of patients would be 50. Although often not made explicit, comparisons of NNT estimates from different studies should account for the duration of follow-up used to create each estimate. In addition, the NNT concept assumes a homogeneity in response to treatment that may not be accurate. The NNT is simply another way of summarizing the absolute treatment difference and does not provide any unique information.

CLINICAL PRACTICE GUIDELINES

Per the 1990 Institute of Medicine definition, clinical practice guidelines are "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances." This definition emphasizes several crucial features of modern guideline development. First, guidelines are created by using the tools of EBM. In particular, the core of the development process is a systematic literature search followed by a review of the relevant peer-reviewed literature. Second, guidelines usually are focused on a clinical disorder (e.g., diabetes mellitus, stable angina pectoris) or a health care intervention (e.g., cancer screening). Third, the primary objective of guidelines is to improve the quality of medical care by identifying care practices that should be routinely implemented, based on high-quality evidence and high benefit-to-harm ratios for the interventions. Guidelines are intended to "assist" decision-making, not to define explicitly what decisions should be made in a particular situation, in part because guideline-level evidence alone is never sufficient for clinical decision-making (e.g., deciding whether to intubate and administer antibiotics for pneumonia in a terminally ill individual, in an individual with dementia, or in an otherwise healthy 30-year-old mother).

Guidelines are narrative documents constructed by expert panels whose composition often is determined by interested professional organizations. These panels vary in expertise and in the degree to which they represent all relevant stakeholders. The guideline documents consist of a series of specific management recommendations, a summary indication of the quantity and quality of evidence supporting each recommendation, an assessment of the benefit-to-harm ratio for the recommendation, and a narrative discussion of the recommendations. Many recommendations simply reflect the expert consensus of the guideline panel because literature-based evidence is insufficient or absent. An examination of this issue in cardiovascular guidelines showed that <15% of guideline recommendations were based on the highest level of clinical trial evidence, and this proportion had not improved in 10 years despite a substantial number of trials being conducted and published. The final step in guideline construction is peer review, followed by a final revision in response to the critiques provided.

Guidelines are closely tied to the process of quality improvement in medicine through their identification of evidence-based best practices. Such practices can be used as quality indicators. Examples include the proportion of acute MI patients who receive aspirin upon admission to a hospital and the proportion of heart failure patients with a depressed ejection fraction treated with an ACE inhibitor.

CONCLUSIONS

Thirty years after the introduction of the EBM movement, it is tempting to think that all the difficult decisions practitioners face have been or soon will be solved and digested into practice guidelines and computerized reminders. However, EBM provides practitioners with an ideal rather than a finished set of tools with which to manage patients. Moreover, even with such evidence, it is always worth remembering that the response to therapy of the "average" patient represented by the summary clinical trial outcomes may not be what can be expected for the specific patient sitting in front of a clinician in the clinic or hospital. In addition, meta-analyses cannot generate evidence when no adequate randomized trials exist, and most of what clinicians confront in practice will never be thoroughly tested in a randomized trial. For the foreseeable future, excellent clinical reasoning skills and experience supplemented by well-designed quantitative tools and a keen appreciation for the role of individual patient preferences in their health care will continue to be of paramount importance in the practice of clinical medicine.

FURTHER READING

Chen JH et al: Decoding artificial intelligence to achieve diagnostic excellence. JAMA 328:709, 2022.

Fanaroff AC et al: Levels of evidence supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008-2018. JAMA 321:1069, 2019.

Goldberg CB et al: To do no harm - and the most good - with AI in health care. Nat Med 30:623, 2024.

Hernán MA, Robins JM: *Causal Inference: What If*. Boca Raton, FL, Chapman & Hall/CRC, 2020.

Khera R et al: Automation bias and assistive AI. Risk of harm from AI-driven clinical decision support. JAMA 330:2255, 2023.

Mandelblatt JS et al: Collaborative modeling of the benefits and harms of associated with different U.S. breast cancer screening strategies. Ann Intern Med 164:215, 2016.

Monteior S et al: The 3 faces of clinical reasoning: Epistemological explorations of disparate error reduction strategies. J Eval Clin Pract 24:666, 2018.

Murthy VK et al: An inquiry into the early careers of master clinicians. J Grad Med Educ 10:500, 2018.

Park YJ et al: Assessing the research landscape and clinical utility of large language models: A scoping review. BMC Med Inform Decis Mak 12:72, 2024.

Richards JB et al: Teaching clinical reasoning and critical thinking: From cognitive theory to practical application. Chest 158:1617, 2020.

Royce CS et al: Teaching critical thinking: A case for instruction in cognitive biases to reduce diagnostic errors and improve patient safety. Acad Med 94:187, 2019.

Saposnik G et al: Cognitive biases associated with medical decisions: A systematic review. BMC Med Inform Decis Mak 16:138, 2016.

Schuwirth LWT et al: Assessment of clinical reasoning: three evolutions of thought. Diagnosis (Berl) 7:191, 2020.

5

Precision Medicine and Clinical Care

The Editors

DISEASE NOSOLOGY AND PRECISION MEDICINE

Modern disease nosology arose in the late nineteenth century and represented a clear departure from the holistic, limited descriptions of disease dating to Galen. In this rubric, the definition of any disease is largely based on clinicopathologic observation. As the correlation between clinical signs and symptoms with pathoanatomy required autopsy material, diseases tended to be characterized by the end organ in which the primary syndrome was manifest and by late-stage presentations. Morgagni institutionalized this framework with the publication of *De Sedibus et Causis Morborum per Anatomen Indagatis* in 1761, in which he correlated the clinical features of patients with more than 600 autopsies at the University of Padua, demonstrating an anatomic basis for disease pathophysiology. Clinicopathologic observation served as the basis for inductive generalization coupled with the application of Occam's razor in which disease complexity was reduced to its simplest possible form. While this approach to defining human disease has held sway for over a century and facilitated the conquest of many diseases previously considered incurable, overly inclusive and simplified Oslerian diagnostics suffer from significant shortcomings. These include, but are not limited to, failure to distinguish the underlying etiology of different diseases with common pathophenotypes. For example, many different diseases can cause end-stage kidney disease or heart failure. Over time, the classification of neurodegenerative disorders or lymphomas, as well as many other diseases, is becoming more refined and precise as the underlying molecular etiologies are identified. These distinctions are important for providing predictable prognostic information for individual patients with even highly prevalent diseases. Additionally, therapies may be ineffective owing to a lack of understanding of the often subtle molecular complexities of specific disease drivers.

Beginning in the mid-twentieth century, the era of molecular medicine offered the idealized possibility of identifying the underlying molecular basis of every disease. Using a conventional reductionist paradigm, physician-scientists explored disease mechanism at ever-increasing molecular depth, seeking the single (or limited number of) molecular cause(s) of many human diseases. Yet, as effective as this now conventional scientific approach was at uncovering many disease mechanisms, the clinical manifestations of very few diseases could be explained on the basis of a single molecular mechanism. Even knowledge of the globin β chain mutation that causes sickle cell disease does not predict the many different manifestations of the disease (stroke syndrome, painful crises, and hemolytic crisis, among others). Clearly, the profession had expected too much from oversimplified reductionism and failed to take into consideration the extraordinary biologic variety and its accompanying molecular and genetic complexity that underpin both normal and pathologic diversity. The promise of the Human Genome Project provided new tools and approaches and unleashed efforts to identify a monogenic, oligogenic, or polygenic cause for every disease (allowing for environmental influences). Yet, once again, disappointment reigned as the pool of genomes expanded without the expected revelations (aside from rare variants). These shortcomings are explained in part by the important roles of epigenetics, which is significantly modulated by environmental exposures and individual experiences. The arc of progressive reductionism (as illustrated for tuberculosis in **Fig. 5-1**) in refining and explaining disease reached a humbling plateau, revealing the need for new approaches to understand better the etiology, manifestations, and progression of most diseases. The stage was set for a return to holism. However, in contrast to the holism of ancient physicians, a more useful approach is one that is integrative, taking genomic context into account in all dimensions. In the course of elaborating this complex pathobiological landscape,

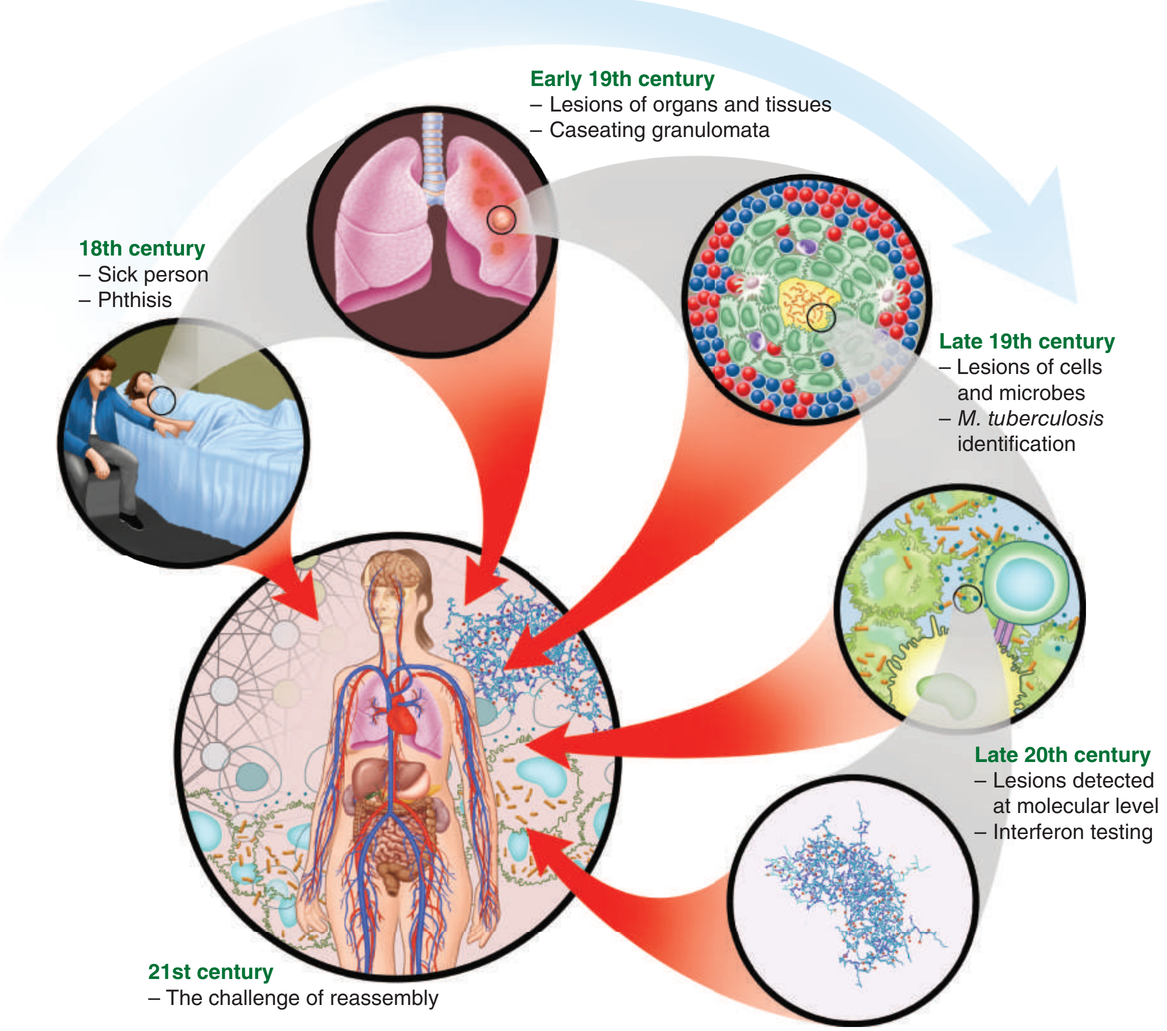
Early 19th century
– Lesions of organs and tissues
– Caseating granulomata
18th century
– Sick person
– Phthisis
Late 19th century
– Lesions of cells and microbes
– *M. tuberculosis* identification
Late 20th century
– Lesions detected at molecular level
– Interferon testing
21st century
– The challenge of reassembly

FIGURE 5-1 Arc of reductionism in medicine. *(From JA Greene, J Loscalzo: Putting the patient back together–social medicine, network medicine, and the limits of reductionism. N Engl J Med 377:2493, 2017. Copyright © 2017 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

disease definition must become more precise and progressively more individualized, setting the stage for *precision medicine*.

Oversimplification of phenotype is a natural outgrowth of the observational scientific method. Categorizing individuals as falling into groups or clusters that are reasonably similar simplifies the task of the diagnostician and also facilitates the application of "specific" therapies more broadly. Biomedicine has been viewed as less quantitative and precise than other scientific disciplines, with biological and pathobiological diversity (biologic "noise") viewed as the norm. Thus, distilling such observational complexity to a fundamental group of symptoms or signs that are reasonably invariant across a group of sick individuals has served as the basis for the approach to disease and its treatment since the earliest days of medicine. This approach to diagnosis and therapy has remained in place into the twenty-first century, serving as the basis for the development of standard diagnostic tests and of broadly applied drug therapies. Targeting larger groups of patients is efficient when applied to large populations. As successful as this approach has been in advancing medical care, it is important to point out its limitations, which include significant predictive inaccuracies and sizeable segments of the disease population that do not respond to the most "effective" drugs (upward of 60% by some estimates). Clearly, a more nuanced approach to diagnosis and therapy is required to achieve better prognostic and therapeutic outcomes.

Turning first to phenotype, astute clinicians know full well the subtle and vivid differences in presentation that are often manifest among individuals with the same disease. In some cases, these differences in pathophenotype lead to new subclassifications of the disease, such as heart failure with preserved ejection fraction versus heart failure with reduced ejection fraction. Often, these relatively crude efforts at making diagnoses more precise are driven by new technologies or new ways of applying established technologies. In other cases, differences in pathophenotype are more subtle, not necessarily clinically apparent, and often driven by measures of endophenotype, such as distinctions among vasculitides facilitated by refinements in serologies or immunophenotyping. The impetus to create these subclasses of disease is largely determined by the need to improve prognosis and apply more precise and effective therapies. Based on these guiding principles, many experienced clinicians will argue—and rightly so—that they have been practicing personalized, precision medicine throughout their careers: they characterize each patient's illness in great detail and choose therapies that respect and are guided by those individualized clinical and laboratory features, limited though they may be.

For many diseases, genomic variation, whether inherited or acquired, provides opportunities to refine diagnostic precision with even greater fidelity and predictive accuracy. For this reason, the field of precision medicine has now entered a new era that couples the molecular reductionism of the last century with an integrative, systems-level understanding of the basis for pathophenotype. Equally important, modern genomics has established that genomic context, sometimes referred to as modifier genes, is distinctive for each individual person; hence, understanding that context provides the insight necessary to predict how a primary disease driver or drivers may manifest a clinical pathophenotype—e.g., why some individuals with sickle cell anemia will develop stroke, while others will develop acute chest syndrome. *This concept that primary genetic and/or environmental drivers of a disease differentially affect disease expression based on an individual's* *unique genomic context serves as the ultimate basis for much of what is now called precision medicine.*

To develop a precision medicine strategy for any disease, the clinician needs to be aware of two important, confounding principles. First, patients with different diseases can manifest similar pathophenotypes, i.e., *convergent phenotypes*. Examples of this principle include the hypertrophied myocardium found in hypertrophic cardiomyopathy, infiltrative cardiomyopathies, critical aortic stenosis, and untreated, long-standing hypertension; and the thrombotic microangiopathy found in malignant hypertension, scleroderma renal crisis, thrombotic thrombocytopenic purpura, eclampsia, and antiphospholipid syndrome. As genetic analysis has been applied to neoplasms originating from the same organ and sharing a common name, substantial heterogeneity has been detected often with pathophysiologic and therapy-affecting consequences. Second, patients with the same basic disease can manifest very different pathophenotypes, i.e., *divergent phenotypes* **(Chap. 479)**. Examples of this principle include the different clinical manifestations of cystic fibrosis or sickle cell disease and the incomplete penetrance of many common genetic diseases. These common presentations of different diseases and different presentations of the same disease are both a consequence of genomic context coupled with unique exposures over an individual's lifetime **(Fig. 5-2)**. Understanding the interplay among these many complex molecular determinants of disease expression is essential for the success of precision medicine.

Given the complexity of the genomic and environmental context of an individual, one must ask the question: How precise do we need to be in order to practice effective precision medicine? Complete knowledge of a person's comprehensive genome (DNA, gene expression, mitochondrial function, proteome, metabolome, posttranslational modification of the proteome, and metagenome, among others) and quantitative assessments of environmental and social history are not possible to acquire; yet, this shortcoming does not render the general problem intractable. Ultimately, more precision is needed when it is actionable; otherwise, it leads to excessive testing, anxiety, and accompanying risks. Owing to the fact that the molecular networks that govern phenotype are overdetermined (i.e., redundant) and that there are primary drivers of disease expression that are modified in a weighted way by other genomic features of an individual, the practice of precision medicine can be realized without complete knowledge of all dimensions of the genome. Examples of how best to realize this strategy are discussed later in this chapter.

REQUIREMENTS FOR PRECISION MEDICINE

The essential elements of any precision medicine effort include phenotyping, endophenotyping (defining the characteristics of a disorder that are not readily observable), genomic profiling, and understanding social determinants of health **(Fig. 5-3)**. While subtle distinctions among individuals with the same disease are well known to clinicians, formalizing these nuanced differences is critical for achieving more precise phenotypes. Deep phenotyping requires a detailed history, including family history and environmental exposures, as well as relevant (physiologic) functional studies and imaging, including molecular imaging where appropriate. Biochemical, immunologic, and molecular tests of body fluids provide additional detail to the overall phenotype. Importantly, these objective laboratory tests together with functional studies compose an assessment of the endophenotype (or endotype) of an individual, refining the overall discriminant power of the evaluation. One additional concept that has gained traction in recent years is the notion of orthogonal phenotyping, i.e., assessing clinical, molecular, imaging, or functional (endo)phenotypes seemingly unrelated to the clinical presentation. These features further enhance the ability to distinguish (sub)phenotypes and derive from the fact that diseases can be subtly (subclinically) manifest in organ systems different from that in which the primary symptoms or signs are expressed. While some diseases are well known to affect multiple organ systems (e.g., systemic lupus erythematosus) and in many cases involvement of those many systems is assessed at initial diagnosis, such is not the case for most other diseases. As we begin to understand the differences in the organ-specific expression of genomic variants that drive or modify disease, it is becoming increasingly apparent that orthogonal—or more appropriately, unbiased comprehensive—phenotyping should become the norm.

Genomic profiling must next be coupled to detailed phenotyping. The complex levels of genomic assessment continue to mature and include DNA sequencing (exome, whole genome), gene expression (mRNA and protein expression), and metabolomics. In addition, the epigenome, the posttranslationally modified proteome, and the metagenome (the personal microbiome of an individual) are gaining traction as additional elements of comprehensive genomics **(Chap. 497)**. Most of these genomic features are not yet available for clinical laboratory testing, and those that are available are largely confined to blood testing. An emerging area is immunophenotyping, using the immune system as an indicator of disease or prior exposures, as well as a sensing system for the emergence of new diseases. While DNA sequencing using whole blood would generally apply to any organ-based disease, gene expression, metabolomics, and epigenomics are often tissue-specific. As tissue specimens cannot always or easily be obtained from the organ of interest, attempts at correlating whole-blood mRNA, protein, or metabolite profiles with those of the involved organ are critical for precise prognostics and therapeutic choices. In many cases, systemic

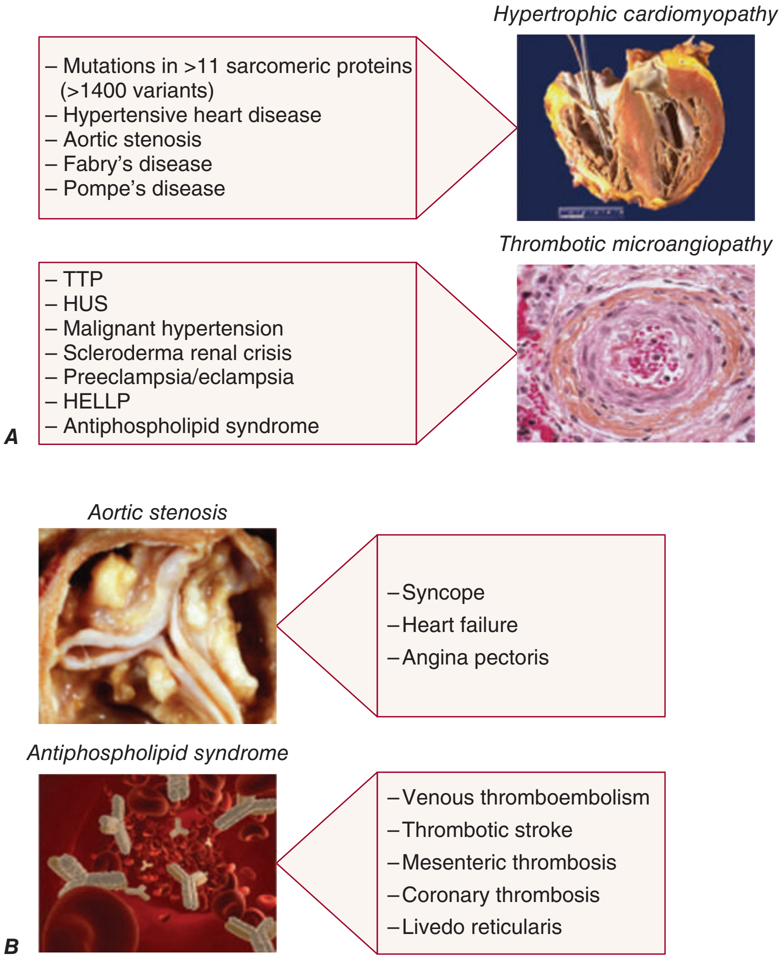

FIGURE 5-2 Convergent and divergent phenotypes. Examples of the former (***A***) include hypertrophic cardiomyopathy and thrombotic microangiopathy, and examples of the latter (***B***) include aortic stenosis and antiphospholipid syndrome, each of which can have several distinct clinical presentations. HELLP, hemolysis, elevated liver enzymes, and a low platelet count; HUS, hemolytic-uremic syndrome; TTP, thrombotic thrombocytopenic purpura.

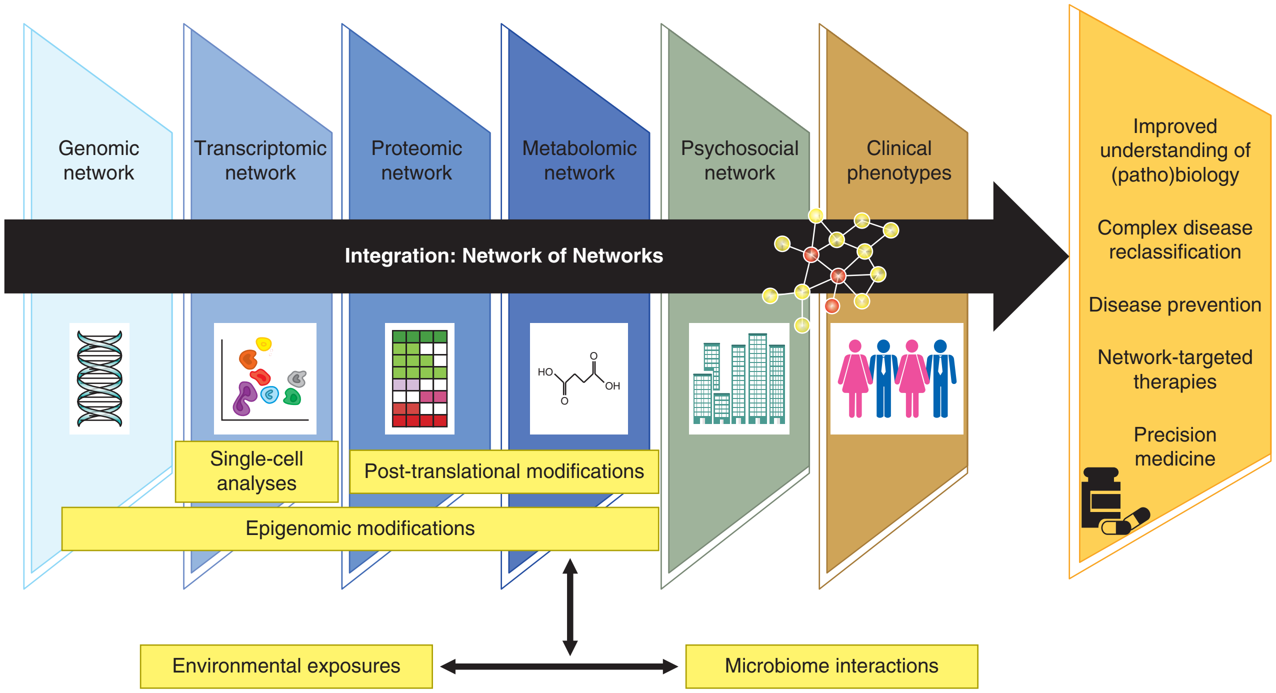

FIGURE 5-3 Universe of precision medicine. The totality of precision medicine incorporates multidimensional biologic networks, the integration of which leads to a network of networks whose components interact with each other and with environmental exposures to yield a distinctive phenotype or pathophenotype. *(Reproduced with permission from LYH Lee, J Loscalzo: Network medicine in pathobiology. Am J Pathol 189:1311, 2019.)*

consequences to an organ-specific disease (e.g., systemic inflammatory responses in individuals with atherosclerosis) can be ascertained and may provide useful prognostic information or therapeutic strategies. These biomarker signatures are the subject of ongoing discovery and have provided useful guidance toward improved diagnostic precision in many diseases. However, in many diseases, the correlations between these plasma or blood markers and organ-based diseases are weak, indicating a need to analyze each condition and each resulting signature before applying it to clinical decision-making. It is important to note that one of the key determinants of the functional consequences of a genetic variant believed to drive a disease phenotype is not simply its expression in a tissue of interest but, more importantly, the coexpression of protein binding partners in that same tissue comprising specific (dys)functional pathways that govern phenotype **(Fig. 5-4)**. An alternative strategy currently under investigation is the conversion of induced pluripotent stem cells from a patient into a cell type of interest for gene expression or metabolomics study. As rational as this approach seems from first principles, it is important to note that gene expression patterns in these induced, differentiated cell types are not completely consonant with their native counterparts, offering often limited additional information at potentially great additional expense.

Single-cell gene expression data are yet another area of modern genomics that will add even more complexity to understanding the genesis of disease phenotype. These data are becoming increasingly available for different cell and tissue types, including their spatial distribution. What role this differential expression may have on ultimate integrative pathophenotype and how intercellular communication (homologous and heterologous) within an organ or tissue may influence gene expression or be influenced by differential gene expression remain interesting questions of ongoing study.

While phenotype features of many chronic diseases are assessed longitudinally, at the current time, genomic features tend to be limited to single time point sampling. Time trajectories are extremely informative in precision genotyping and phenotyping, with gene expression patterns and phenotypes changing over time in different ways among different patients with the same overarching phenotype. Cost, feasible sampling frequency, predictive power, and therapeutic choices will all drive the optimal strategy for the acquisition of timed samples in any given patient; however, with continued cost reduction in genomics technologies, this limitation may be progressively mitigated and clinical application may become a reality.

One important class of diseases that does not have most of these limitations in genomic profiling is cancer. Cancers can be (and are) sampled (biopsied) frequently to monitor temporal changes in the somatically mutating oncogenome and its consequences for the limited number of well-defined oncogenic driver pathways **(Chap. 76)**. A unique limitation of cancer in this regard, however, is that the frequency of somatic mutations over time (and, especially, with treatment) is great and the functional consequences of many of these mutations unknown. Equally important, assessment of single-cell mRNA sequencing patterns demonstrates great variability between apparently similar cells, challenging functional interpretation. Lastly, in solid tumors, stromal cells interact in a variety of ways (e.g., metabolically) with the associated malignant cells, and their gene expression signatures are also modified by the changing somatic mutational landscape of the primary malignancy. Thus, while much more information can be obtained over time in most cancer patients, the interpretation of these rich data sets continues to remain largely semi-empirical.

The possibility of identifying specific therapeutic targets remains a major goal of precision medicine. Doing so usually requires more than simple DNA sequencing and may include analysis of some level of gene expression, ideally in the involved organ(s). In addition to demonstrating the expression of a variant protein in the organ, one must ideally also demonstrate its functional consequences, which requires ascertaining the expression of binding partner proteins and the functional pathways they comprise. To achieve this goal, a variety of approaches have been tried, one of the most successful of which is the construction of the protein-protein interaction network (the interactome), which is a comprehensive network map of the protein-protein interactions in a cell or organ of interest **(Chap. 499)**. This template provides information on the subnetworks that govern a disease phenotype (disease modules), which can be further individualized by incorporating individual variants and differentially expressed proteins that are patient specific. This type of analysis leads to the creation of an individual "reticulome" or reticulotype (after the Latin for network), which links the genotype to the phenotype of an individual **(Fig. 5-5)**. Using this approach, one can identify potential drug targets in a rational way or can even repurpose existing drugs by demonstrating the proximity of a known

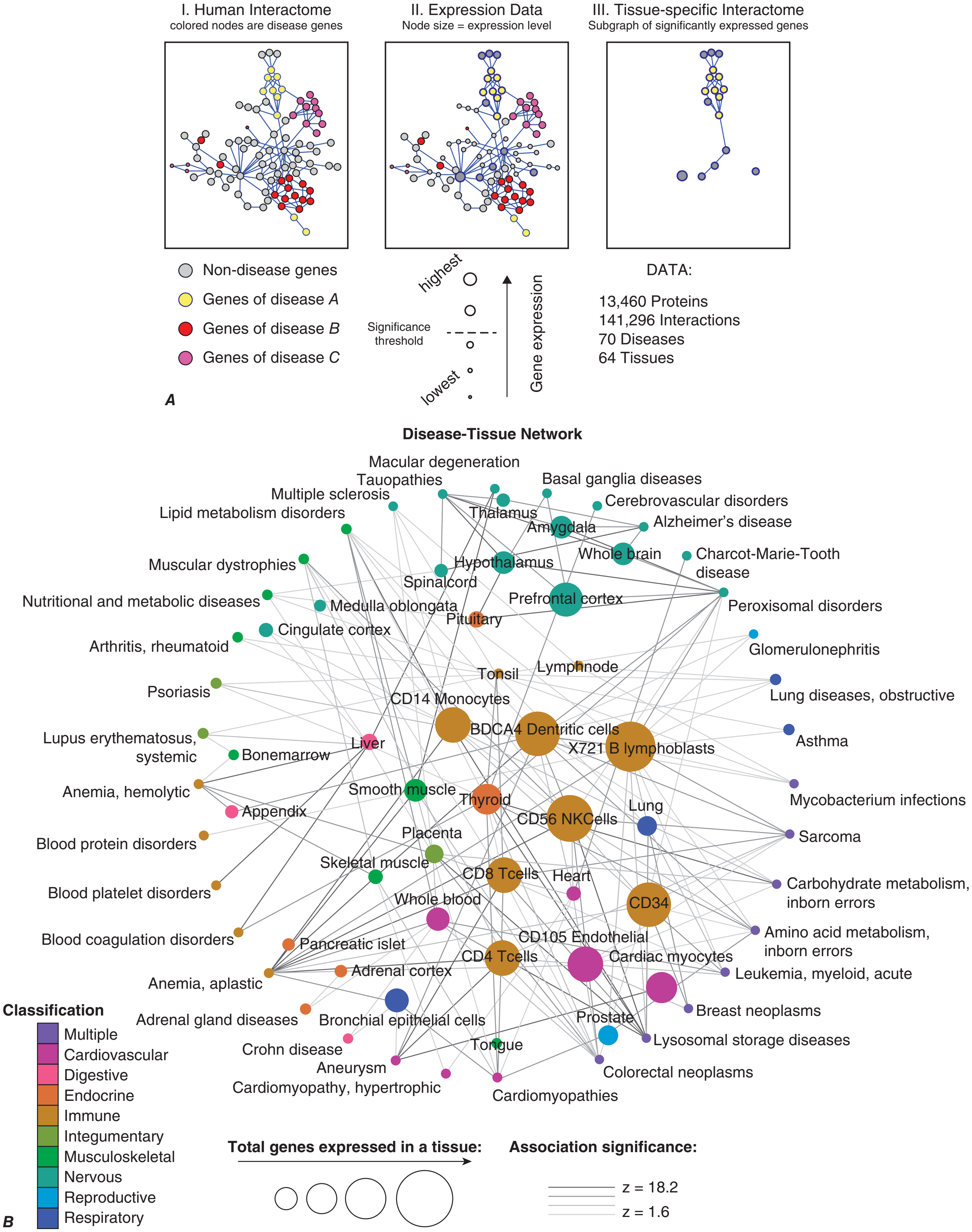
I. Human Interactome
colored nodes are disease genes
II. Expression Data
Node size = expression level
III. Tissue-specific Interactome
Subgraph of significantly expressed genes
Non-disease genes
Genes of disease *A*
Genes of disease *B*
Genes of disease *C*
highest
Significance threshold
lowest
Gene expression
DATA:
13,460 Proteins
141,296 Interactions
70 Diseases
64 Tissues
A
Disease-Tissue Network
Macular degeneration
Tauopathies
Basal ganglia diseases
Multiple sclerosis
Cerebrovascular disorders
Lipid metabolism disorders
Thalamus
Amygdala
Alzheimer's disease
Muscular dystrophies
Hypothalamus
Whole brain
Charcot-Marie-Tooth disease
Spinalcord
Nutritional and metabolic diseases
Prefrontal cortex
Medulla oblongata
Peroxisomal disorders
Pituitary
Arthritis, rheumatoid
Cingulate cortex
Glomerulonephritis
Tonsil
Lymphnode
Psoriasis
Lung diseases, obstructive
CD14 Monocytes
BDCA4 Dentritic cells
Lupus erythematosus, systemic
Liver
X721 B lymphoblasts
Asthma
Bonemarrow
Anemia, hemolytic
Smooth muscle
Mycobacterium infections
Thyroid
Appendix
CD56 NKCells
Lung
Placenta
Sarcoma
Blood protein disorders
Skeletal muscle
CD8 Tcells
Heart
Carbohydrate metabolism, inborn errors
Blood platelet disorders
Whole blood
CD34
Amino acid metabolism, inborn errors
Blood coagulation disorders
Pancreatic islet
CD105 Endothelial
CD4 Tcells
Cardiac myocytes
Adrenal cortex
Leukemia, myeloid, acute
Anemia, aplastic
Classification
Adrenal gland diseases
Bronchial epithelial cells
Breast neoplasms
Prostate
Multiple
Lysosomal storage diseases
Cardiovascular
Crohn disease
Tongue
Digestive
Aneurysm
Colorectal neoplasms
Endocrine
Cardiomyopathy, hypertrophic
Cardiomyopathies
Immune
Integumentary
Musculoskeletal
Nervous
Reproductive
Respiratory
Total genes expressed in a tissue:
Association significance:
z = 18.2
z = 1.6
B

FIGURE 5-4 Tissue-specific gene expression and phenotype. ***A.*** The human protein-protein interactome is constructed, and a specific disease module is identified (I); gene expression within this module is ascertained (II); and the tissue specificity of gene expression is determined (III). This analysis leads to a reduction of the total number of disease module genes that govern phenotype in a specific organ, which is a reflection of the specific pathway (or pathways) that is (or are) expressed in their functional entirety in that tissue. ***B.*** A disease-tissue bipartite network is constructed wherein specific tissues are placed within the circle and linked to diseases shown on the circumference. Nodes are colored according to tissue classification, the sizes of nodes are proportional to the total number of genes expressed in them, and the widths (shades) of the lines or edges correspond to the significance of the associations with specific diseases. *(From M Kitsak et al: Tissue specificity of human disease module. Sci Rep 6:35241, 2016, Figure 4.)*

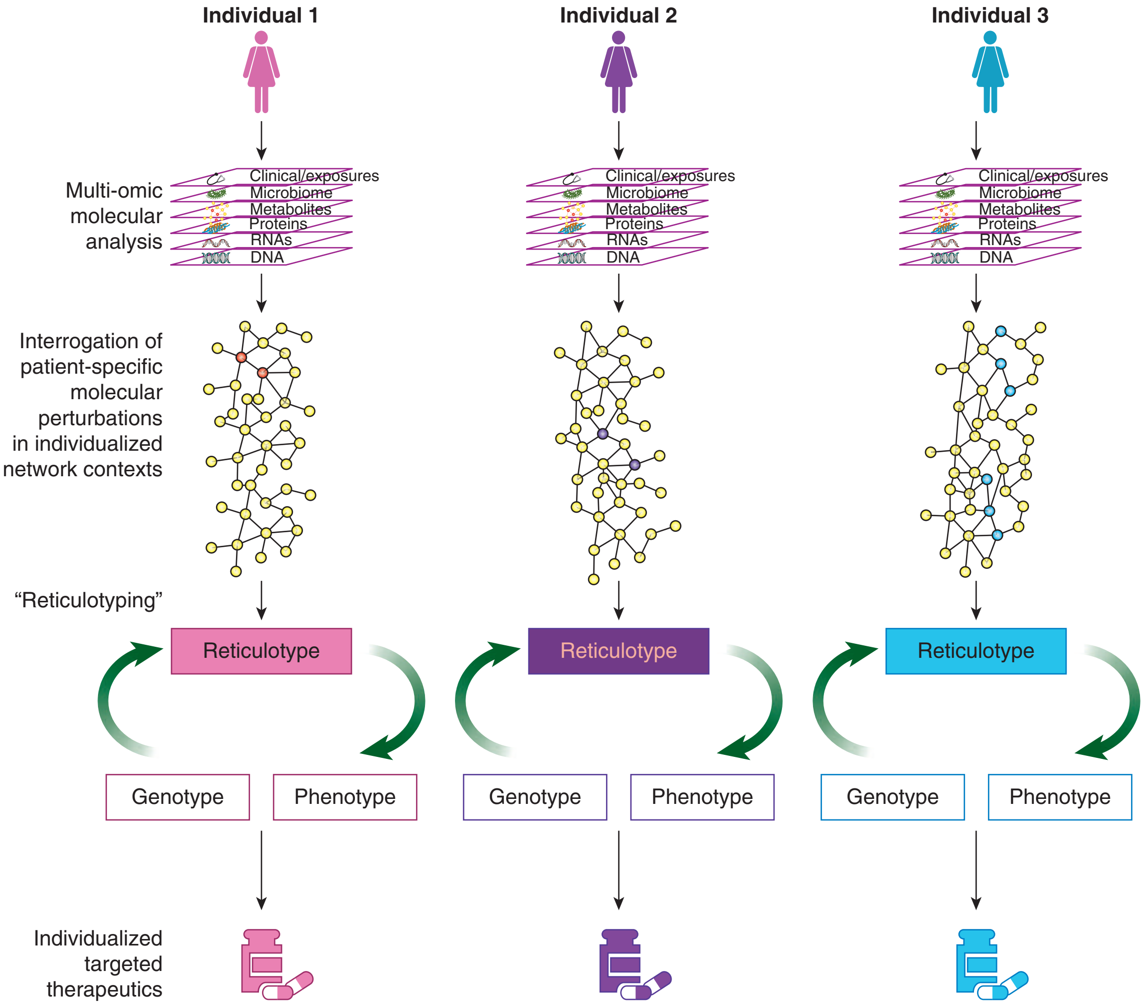
Individual 1
Individual 2
Individual 3
Multi-omic molecular analysis
Clinical/exposures
Microbiome
Metabolites
Proteins
RNAs
DNA
Interrogation of patient-specific molecular perturbations in individualized network contexts
"Reticulotyping"
Reticulotype
Genotype
Phenotype
Individualized targeted therapeutics

FIGURE 5-5 Reticulotype. Patient-specific genotype-phenotype relationships by multiomic network structures are depicted for three individuals. Each individual's unique molecular perturbations (genetic variants, differentially expressed genes) are examined within the context of the subject's unique integrative biologic network or reticulome derived from these multiomic analyses. These unique reticulotypes then serve as the basis for patient-specific, precision therapies. *(Reproduced with permission from LY-H Lee, J Loscalzo: Network medicine in pathobiology. Am J Pathol 189:1311, 2019.)*

drug target to a disease module of interest **(Fig. 5-6)**. For example, in multicentric Castleman's disease, a disorder of unclear etiology, recognition that the PI3K/Akt/mTOR pathway is highly activated led to trials with an existing drug approved for other purposes, sirolimus. Precision medicine offers additional opportunities for optimizing the utilization of a drug by assessing the individualized pharmacogenomics of its disposition and metabolism, as demonstrated for the adverse consequences of variants in *TPMT* on azathioprine metabolism and variants in *CYP2C19* on clopidogrel metabolism **(Chap. 72)**.

EXAMPLES OF PRECISION MEDICINE APPLICATIONS

The field of precision medicine did not appear abruptly in medical history but, rather, evolved gradually as clinicians became more aware of differences among patients with the same disease. With the advent of modern genomics, in the ideal situation, these phenotype differences can now be mapped to genotype differences. Thus, we can consider precision medicine from the perspective of the pregenomic era and the postgenomic era. Pregenomic precision medicine was applied to many diseases as therapeutic classes expanded for those disorders. For example, many endocrine disorders (i.e., type 1 diabetes mellitus) are treated with a goal of restoring metabolic normality, precisely titrating hormone treatments (i.e., insulin) to a metabolic endpoint (i.e., glucose). Another prime example of this approach is in the field of heart failure, where diuretics, digoxin, beta blockers, afterload-reducing agents, venodilators, renin-angiotensin-aldosterone inhibitors, and brain natriuretic

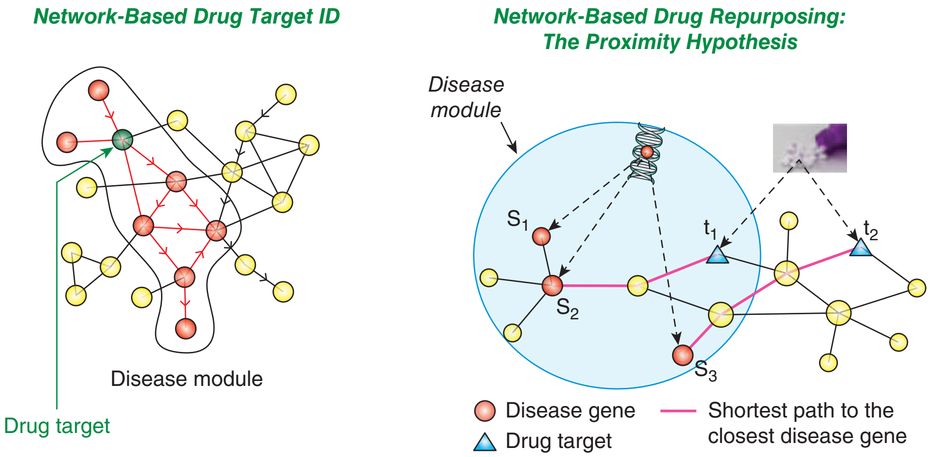
Network-Based Drug Target ID
Disease module
Drug target
Network-Based Drug Repurposing: The Proximity Hypothesis
Disease module
S1
S2
S3
t1
t2
Disease gene
Drug target
Shortest path to the closest disease gene

FIGURE 5-6 Network-based precision drug repurposing. *(Adapted from F Cheng et al: A genome-wide positioning systems network algorithm for in silico drug repurposing. Nat Commun 10:3476, 2019.)*

peptide (nesiritide) are commonly used in some combination for most patients. The choice of agents is governed by the evidence basis for their use, but tailored to the primary pathophysiologic phenotypes manifest in a patient, such as congestion, hypertension, and impaired contractility. These treatments were developed in the latter half of the last century based on empiric observation, reductionist experiments of specific pathways believed to be involved in the pathophysiology, and clinical response in prospective trials. As phenotyping became more refined (e.g., echocardiographic assessments of ventricular function and tissue Doppler characterization of ventricular relaxation), the syndrome was subclassified into heart failure with reduced ejection fraction and heart failure with preserved ejection fraction, the latter of which does not respond well to most of the classes of therapeutic agents currently available. In the postgenomic era, ever more refined and detailed methods are under investigation to characterize pathophenotypes as well as genotypes, which may then be matched to the idealized combination of therapeutic classes of agents.

Pulmonary arterial hypertension is another disease for which definitive therapies straddle the pre- and postgenomic eras of precision medicine. Prior to the 1990s, there were no effective therapies for this highly morbid and lethal condition. With the advent of molecular and biochemical characterization of vascular abnormalities in individuals with established disease, however, therapies with agents that restored normal vascular function improved morbidity and mortality. These included calcium channel blockers, prostacyclin congeners, and endothelin receptor antagonists. As genomic characterization of the disease has progressed over the past two decades, there is increasing recognition of distinct genotypes that yield unique phenotypes **(Chap. 294)**, such as the demonstration of a primarily fibrotic endophenotype governed by the (oxidized) scaffold protein NEDD-9 and its aldosterone-dependent, TGF-β–independent enhancement of collagen III expression. This approach will continue to evolve as therapies become more effective (e.g., for perivascular fibrosis) and therapeutic choices better targeted to individual patients.

Precision diagnostics has also led to a new classification of the dementias, conditions previously thought to have a single cause with varied clinical expression. These disorders can now be categorized based on the genes and pathways involved and the site where aggregated proteins first form and then spread in the nervous system. For example, the varied clinical presentations of frontotemporal dementia, including progressive aphasia, behavioral disturbances, and dementia with amyotrophic lateral sclerosis, can now be linked to specific genotypes and susceptible cell types **(Chap. 443)**. In prion diseases, the clinical phenotype is determined by specific germline mutations present in the prion protein **(Chap. 449)**. Discovery of autoantibodies against aquaporin-4 (AQP-4) and myelin oligodendrocyte glycoprotein (MOG) has allowed neuromyelitis optica, previously considered a multiple sclerosis–like disorder, to be classified as a separate entity requiring different treatment **(Chap. 456)**. Similarly, in myasthenia gravis, the identification of novel autoantibodies now permits stratification and a more finely tuned precision approach to therapy **(Chap. 459)**.

Precision medicine approaches to cancers have, of course, become the prime example of the opportunity that this strategy offers. In the pregenomic era, chemotherapy was widely used based upon the tissue affected and histologic characteristics of the tumors. Treatment success was variable despite continued efforts to characterize the molecular features of the specific tumors and their semi-empiric responses to specific chemotherapeutic agents. As cancer genome sequencing evolved, however, it became apparent that there are a limited number of oncogenic pathways (<20) that are represented in the great majority of malignancies, without regard for the organ in which the disease is primarily manifest. These genomic signatures served as a template for precisely targeted therapies that have led to dramatic changes in response to treatment, including, for example, imatinib (and congeners) for Bcr-Abl tyrosine kinase activity in chronic myeloid leukemia, erlotinib for *EGFR*-mutant non-small cell lung cancers, and ibrutinib for Bruton tyrosine kinase in chronic lymphocytic leukemia, among many others.

As exciting as these approaches have been, at least four primary challenges associated with precision therapeutics are unique to cancer: (1) the mutational landscape continues to evolve as the disease progresses, and therapy often (if not invariably) leads to selection for resistant clones; (2) the likelihood that any cancer can be definitively cured by any single agent, no matter its exquisite precision, is quite limited, necessitating the development of rational polypharmaceutical approaches that take into account alternative pathways that achieve the same oncogenic goals as the primary targeted pathway, complicating drug development; (3) marked genomic heterogeneity characterizes many malignancies, arguing that targeting a specific pathway—even with multiple drugs—may not ultimately succeed over the long term owing to the continued and heterogeneous evolution of the genomic landscape within a tumor within a patient; and (4) there is variability in the characteristics of patients and their ability to withstand treatment and mount a complementary immune response to neoplastic cells. Despite these serious shortcomings, the application of progressively more refined and precisely targeted therapies used alone and in combination, such as with immune modulators, continues to offer great promise for the treatment of these diseases.

In some ways, these approaches in cancer mirror earlier strategies in the treatment of infectious diseases in which the identification of the causative organism and its sensitivity to potential antimicrobials allows precision approaches to treatment. Combinatorial antimicrobial treatments represent an effective strategy to address acquired resistance. These diagnostic and therapeutic strategies can be applied without detailed knowledge of personalized responses to the infection or treatment (aside from serious adverse effects) with good outcomes in most cases. Yet, individuals do respond differently to specific infections and their treatments, possibly driven by different endophenotypes (e.g., different inflammatory responses), suggesting that more precise knowledge of these precise mechanistic differences may yield improved prognosis and therapeutic approaches. As with cancer, immune modulation, particularly for immune exhaustion in chronic infections, represents a new frontier, again amenable to the personalized, precise analyses described above.

THE FUTURE OF PRECISION MEDICINE

Precision medicine clearly holds great promise for the future of the practice of medicine. For precision medicine to continue to evolve successfully, however, several requirements will need to be met. First, both deeply refined personal phenotypic data and genomic data are essential as the information with which precision analysis is performed. These data sets are quite large and require sufficient storage for analysis, especially for individuals in whom time trajectories are acquired (as should be the case for every person whenever feasible). Equally important, the analytical methods required to extract useful information from these data sets are evolving and themselves quite complex. While great progress has been made in genomics and biochemical testing, our ability to capture meaningful immunologic endophenotypes and environmental exposures is limited by comparison. Machine learning and artificial (auxiliary) intelligence methods will be essential for extracting optimal information from these data sets, which include not only pathways that can be uniquely targeted therapeutically but also individualized genomic or phenotypic signatures that are highly predictive of outcome, with or without therapy. Gathering sufficient information on the “normal” segments of the population is also required to ensure appropriate comparison data sets for optimal prediction.

Second, phenotyping must continue to expand and become dimensionally richer. The phenotypic features included in this data gathering must incorporate not only data relevant to the clinical presentation but also orthogonal phenotypic data that may yield useful information on disease trajectory or preclinical disease markers. Personal device data, environmental exposure history, social network interactions, and health system data will all be incorporated increasingly in defining phenotype and will require great efforts on the

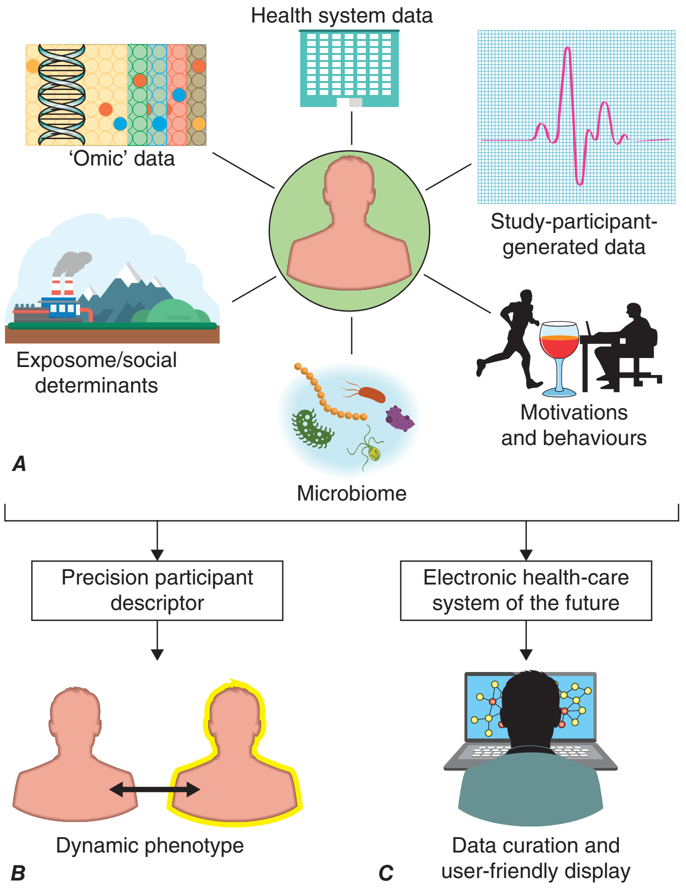
Health system data
'Omic' data
Study-participant-generated data
Exposome/social determinants
Motivations and behaviours
Microbiome
A
Precision participant descriptor
Electronic health-care system of the future
Dynamic phenotype
B
C
Data curation and user-friendly display

FIGURE 5-7 Big data in precision medicine. ***A.*** Six dimensions by which individuals may be characterized in the precision medicine era are described. ***B.*** The precision participant descriptor integrates the data from these six dimensions and varies over time. ***C.*** The electronic medical record increasingly must evolve to provide curated precision data in a user-friendly way. *(Reproduced with permission from E Antman, J Loscalzo: Precision medicine in cardiology. Nat Rev Cardiol 13:591, 2016.)*

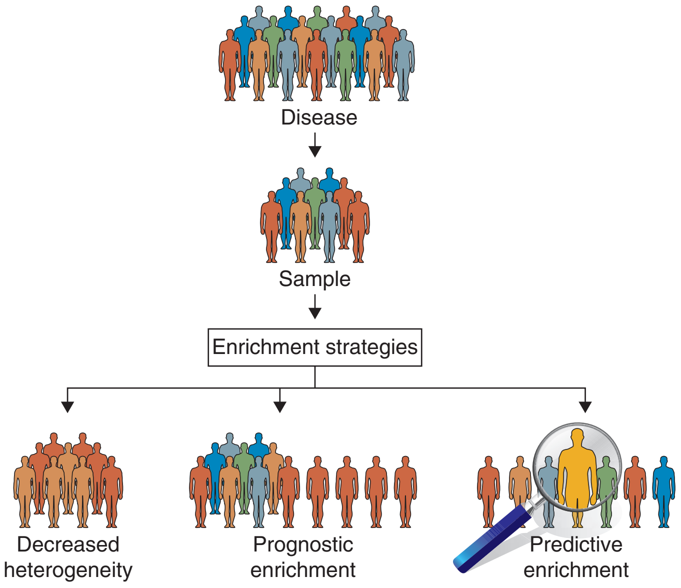

FIGURE 5-8 The basis for precision medicine. The notion of precision medicine evolved, in part, from clinical trial design. From the entire population of patients with the disease of interest, a sample cohort of individuals is enrolled in the trial that ideally is representative of the entire distribution. Enrichment strategies developed to decrease heterogeneity or increase the representation of individuals with a high risk of observed outcomes (prognostic enrichment) facilitate trial conduct but do not necessarily improve precision in defining treatment response. The predictive enrichment strategy utilizes both trial participant characteristics and data from experiments conducted before or during (adaptive design) the trial to improve the prediction of who is likely to have a more pronounced response to the treatment under study. *(Reproduced with permission from E Antman, J Loscalzo: Precision medicine in cardiology. Nat Rev Cardiol 13:591, 2016.)*

part of the medical informatics community to harmonize data sets, standardize data collection, and optimize/standardize data analysis **(Fig. 5-7)**.

Third, perhaps the greatest challenge to making precision medicine the standard approach to illness will be to determine the minimal data set required to predict outcome and response to therapy. Gathering data is comparatively simple; however, analyzing it to eliminate redundant information in these overdetermined biologic systems, weighting the determinants of an outcome, and using the data as phenomic/genomic signatures that are easier to collect than comprehensive, unbiased data sets are the ideal goals—a major challenge, but not insurmountable. Rapidly evolving machine learning and artificial intelligence strategies will also be essential for maximal success.

To return to the question of how precise precision medicine needs to be in order to be useful, please refer to **Fig. 5-8** where the approaches to clinical trial design meant to improve therapeutic signal are illustrated. Decreasing heterogeneity and enriching the study population will enhance the effect size, but these strategies are based on analyses of prior data sets that define those individuals who are more likely than not to respond to a therapy. By contrast, the notion of predictive enrichment follows from the information provided by a detailed, big data–driven analysis of individuals that explores phenotypic and genomic features used to predict response. These features need not be precisely met by each patient; however, they can be collated or clustered to define a reasonably sized cohort predicted to respond in a particular way within certain confidence bounds. In this way, the boundaries to the practice of precision medicine are imprecise strictly speaking, but sufficiently predictive to be practical from the perspectives of clinical care and cost-effectiveness.

FURTHER READING

Antman EM, Loscalzo J: Precision medicine in cardiology. Nat Rev Cardiol 13:591, 2016.

Cheng F et al: A genome-wide positioning systems network algorithm for in silico drug repurposing. Nat Commun 10:3476, 2019.

Cheng F et al: Comprehensive characterization of protein-protein interactions perturbed by disease mutations. Nat Genet 53:342, 2021.

Greene JA, Loscalzo J: Putting the patient back together—Social medicine, network medicine, and the limits of reductionism. N Engl J Med 377:2493, 2017.

Gupta RM et al: Multiomic analysis and CRISPR perturbation screens identify endothelial cell programs and novel therapeutic targets for coronary artery disease. Arterioscler Thromb Vasc Biol 43:608, 2023.

Kitsak M et al: Tissue specificity of human disease module. Sci Rep 6:35241, 2016.

Lee LY, Loscalzo J: Network medicine in pathobiology. Am J Pathol 189:1311, 2019.

Leopold JA et al: The application of big data to cardiovascular disease: Paths to precision medicine. J Clin Invest 130:29, 2020.

Loscalzo J et al: Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. Mol Syst Biol 3:124, 2007.

Loscalzo J et al: Molecular interaction networks and drug development: Novel approach to drug target identification and drug repurposing. FASEB J 37:e22660, 2023.

Maiorino E, Loscalzo J: Phenomics and robust multiomics data for cardiovascular disease subtyping. Arterioscler Thromb Vasc Biol 43:1111, 2023.

Maron BA et al: Individualized interactomes for network-based precision medicine in hypertrophic cardiomyopathy with implications for other clinical pathophenotypes. Nat Commun 12:873, 2021.

Menche J et al: Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science 347:1257601, 2015.

Samokhin AO et al: NEDD9 targets *COL3A1* to promote endothelial fibrosis and pulmonary arterial hypertension. Sci Transl Med 10:eaap7294, 2018.

6

Screening and Prevention of Disease

Gary J. Martin

A primary goal of health care is to prevent disease or detect it early enough that intervention will be more effective. Tremendous progress has been made toward this goal over the past 50 years. Screening tests are available for many common diseases and encompass biochemical (e.g., cholesterol, glucose), physiologic (e.g., blood pressure, growth curves), radiologic (e.g., mammogram, bone densitometry), and cytologic (e.g., Pap smear) approaches. Effective preventive interventions have resulted in dramatic declines in mortality from many diseases, particularly infections. Preventive interventions include counseling about risk behaviors, vaccinations, medications, and, in some relatively uncommon settings, surgery. Preventive services (including screening tests, preventive interventions, and counseling) are different than other medical interventions because they are proactively administered to healthy individuals instead of in response to a symptom, sign, or diagnosis. Thus, the decision to recommend a screening test or preventive intervention requires a particularly high bar of evidence that testing and intervention are both practical and effective.

Because population-based screening and prevention strategies must be extremely low risk to have an acceptable benefit-to-harm ratio, the ability to target individuals who are more likely to develop disease could enable the application of a wider set of potential approaches and increase efficiency. Currently, there are many types of data that can predict disease incidence in an asymptomatic individual. Germline genomic data have received the most attention to date, at least in part because mutations in high-penetrance genes have clear implications for preventive care **(Chap. 480)**. Women with mutations in either *BRCA1* or *BRCA2*, the two major breast cancer susceptibility genes identified to date, have a markedly increased risk (five- to twentyfold) of breast and ovarian cancer. Genetic counseling should be provided. Screening and prevention recommendations often include prophylactic oophorectomy and breast magnetic resonance imaging (MRI), both of which are considered to incur too much harm for women at average cancer risk. Some women with *BRCA* mutations opt for prophylactic mastectomy to greatly reduce their breast cancer risk. Although the proportion of common disease explained by high-penetrance genes appears to be relatively small (5–10% of most diseases), mutations in rare, moderate-penetrance genes and variants in low-penetrance genes also contribute to the prediction of disease risk. Most recently, polygenic risk scores combining information about variants across hundreds of genes are being evaluated for identifying individuals at high risk of coronary heart disease and other conditions. The advent of affordable whole exome/whole genome sequencing is likely to speed the dissemination of these tests into clinical practice and may transform the delivery of preventive care.

Other forms of "omic" data also have the potential to provide important predictive information. Proteomics and metabolomics can provide insight into gene function, but it has proven challenging to develop reliable, predictive measures using these platforms. More recently, it has become possible to measure the presence of mutations in DNA circulating in the bloodstream and in stool, with early promising evidence that these assays can be used to detect cancer before existing screening tests.

In addition to "omic" data, imaging data are increasingly being integrated into risk-stratified prevention approaches as evidence grows about the predictive ability of these data. For example, coronary computed tomography (CT) scans are used in many preventive cardiology programs to inform decisions about beginning statin therapy when there is conflicting or uncertain information from other risk assessment approaches. Of course, these data may also be helpful in predicting the risk of harms from screening or prevention, such as the risk of a false-positive mammogram.

TABLE 6-1 Principles of Screening

The condition should be an important health problem.
There should be a treatment for the condition.
Facilities for diagnosis and treatment should be available.
There should be a latent stage of the disease.
There should be a test or examination for the condition.
The test should be acceptable to the population.
The natural history of the disease should be adequately understood.
There should be an agreed policy on whom to treat.
The cost of finding a case should be balanced in relation to overall medical expenditure.

Source: Reproduced with permission from The basic principles of screening populations for disease. World Health Organization; 1968.

In addition to advances in risk prediction, there are several other reasons that screening and prevention are likely to gain importance in medical care in the near term. New imaging modalities are being developed that promise to detect changes at the cellular and subcellular levels, greatly increasing the probability that early detection improves outcomes. The rapidly growing understanding of the biologic pathways underlying initiation and progression of many common diseases has the potential to transform the development of preventive interventions, including chemoprevention. Furthermore, screening and prevention offer the promise of both improving health and sparing the costs of disease treatment, an issue that will continue to gain importance as long as health care costs in the United States remain a concern to patients, government agencies, and insurers.

This chapter will review the basic principles of screening and prevention in the primary care setting. Recommendations for specific disorders such as cardiovascular disease, diabetes, and cancer are provided in the chapters dedicated to those topics.

BASIC PRINCIPLES OF SCREENING

The basic principles of screening populations for disease were published by the World Health Organization in 1968 **(Table 6-1)**.

In general, screening is most effective when applied to relatively common disorders that carry a large disease burden **(Table 6-2)**. The five leading causes of mortality in the United States are heart diseases, malignant neoplasms, chronic obstructive pulmonary disease, accidents, and cerebrovascular diseases. Thus, many screening strategies are targeted at these conditions. From a global health perspective, these conditions are priorities, but malaria, malnutrition, AIDS, tuberculosis, and violence also carry a heavy disease burden **(Chap. 485)**.

Having an effective treatment for early disease has proven challenging for some common diseases. For example, although Alzheimer's disease is the sixth leading cause of death in the United States, there are no curative treatments and a paucity of data that early treatment improves outcomes. Lack of facilities for diagnosis and treatment is a particular challenge for developing countries and may change screening strategies, including the development of "see and treat" approaches such as those currently used for cervical cancer screening in some countries. A long latent or preclinical phase where early treatment increases the chance of cure is a hallmark of many cancers; for example, polypectomy prevents progression to colon cancer. Similarly, early identification of hypertension or hyperlipidemia allows therapeutic

TABLE 6-2 Lifetime Cumulative Risk

Breast cancer for women	10%
Colon cancer	6%
Cervical cancer for women[a]	2%
Domestic violence for women	Up to 15%
Hip fracture for white women	16%

[a] Assuming an unscreened population without HPV vaccination.

interventions that reduce the long-term risk of cardiovascular or cerebrovascular events. In contrast, lung cancer screening has historically proven more challenging because most tumors are not curable by the time they can be detected on a chest x-ray. However, the length of the preclinical phase also depends on the level of resolution of the screening test, and this situation changed with the development of chest CT. Low-dose chest CT scanning can detect tumors earlier and has been demonstrated to reduce lung cancer mortality by 20% in individuals who had at least a 30-pack-year history of smoking. The short interval between the ability to detect disease on a screening test and the development of incurable disease also contributes to the limited effectiveness of mammography screening in reducing deaths from some forms of breast cancer. At the other end of the spectrum, the early detection of prostate cancer may not lead to a difference in the mortality rate because the disease is often indolent and competing morbidities, such as coronary artery disease, may ultimately cause mortality **(Chap. 75)**. This uncertainty about the natural history is also reflected in the controversy about treatment of prostate cancer, further contributing to the challenge of screening in this disease. Finally, screening programs can incur significant economic costs that must be considered in the context of the available resources and alternative strategies for improving health outcomes.

METHODS OF MEASURING HEALTH BENEFITS

Because screening and preventive interventions are recommended to asymptomatic individuals, they are held to a high standard for demonstrating a favorable risk-benefit ratio before implementation. In general, the principles of evidence-based medicine apply to demonstrating the efficacy of screening tests and preventive interventions, where randomized controlled trials (RCTs) with mortality outcomes are the gold standard. However, because RCTs are often not feasible, observational studies, such as case-control designs, have been used to assess the effectiveness of some interventions such as colonoscopy for colorectal cancer screening. For some strategies, such as Pap smear screening for cervical cancer, the only data available are ecologic data demonstrating dramatic declines in mortality.

Irrespective of the study design used to assess the effectiveness of screening, it is critical that disease incidence or mortality is the primary endpoint rather than length of disease survival. This is important because lead time bias and length time bias can create the appearance of an improvement in disease survival from a screening test when there is no actual effect. Lead time bias occurs because screening identifies a case before it would have presented clinically, thereby creating the perception that a patient lived longer after diagnosis simply by moving the date of diagnosis earlier rather than the date of death later. Length time bias occurs because screening is more likely to identify slowly progressive disease than rapidly progressive disease. Thus, within a fixed period of time, a screened population will have a greater proportion of these slowly progressive cases and will appear to have better disease survival than an unscreened population. Disease-free survival is an alternative to consider since many metastatic cancers are controllable for many years with newer therapies. Differences in cancer mortality may not be as easily detected, but years of chemotherapy would still be worth avoiding.

A variety of endpoints are used to assess the potential gain from screening and preventive interventions.

1. *The absolute and relative impact of screening on disease incidence or mortality*. The absolute difference in disease incidence or mortality between a screened and nonscreened group allows the comparison of size of the benefit across preventive services. A meta-analysis of Swedish mammography trials (ages 40–70) found that ~1.2 fewer women per 1000 would die from breast cancer if they were screened over a 12-year period. By comparison, at least ~3 lives per 1000 would be saved from colon cancer in a population (aged 50–75) screened with annual fecal occult blood testing (FOBT) over a 13-year period, and an estimated 20–24 lives per 1000 would be saved over the entire 25-year period. Based on this analysis, colon cancer screening may actually save more women's lives than does mammography. However, the relative impact of FOBT (30% reduction in colon cancer death) is similar to the relative impact of mammography (14–32% reduction in breast cancer death), emphasizing the importance of both relative and absolute comparisons.
2. *The number of subjects screened to prevent disease or death in one individual*. The inverse of the absolute difference in mortality is the number of subjects who would need to be screened or receive a preventive intervention to prevent one death. For example, 731 women aged 65–69 would need to be screened by dual-energy x-ray absorptiometry (DEXA) (and treated appropriately) to prevent one hip fracture from osteoporosis.
3. *Increase in average life expectancy for a population*. Predicted increases in life expectancy for various screening and preventive interventions are listed in **Table 6-3**. It should be noted, however, that the increase in life expectancy is an average that applies to a population, not to an individual. In reality, the vast majority of the population does not derive any benefit from a screening test. A small subset of patients, however, will benefit greatly. For example, Pap smears do not benefit the 98% of women who never develop cancer of the cervix. However, for the 2% who would have developed cervical cancer, Pap smears may add as much as 25 years to their lives. Some studies suggest that a 1-month gain of life expectancy is a reasonable goal for a population-based screening or prevention strategy.

TABLE 6-3 Estimated Average Increase in Life Expectancy for a Population

SCREENING OR PREVENTIVE INTERVENTION	AVERAGE INCREASE
Mammography:	
Women, 40–50 years	0–5 days
Women, 50–70 years	1 month
Pap smears, age 18–65	2–3 months
Getting a 35-year-old smoker to quit	3–5 years
Beginning regular exercise for a 40-year-old man (30 min, 3 times a week)	9 months–2 years

ASSESSING THE HARMS OF SCREENING AND PREVENTION

Just as with most aspects of medical care, screening and preventive interventions also incur the possibility of adverse outcomes. These adverse outcomes include side effects from preventive medications and vaccinations, false-positive screening tests, overdiagnosis of disease from screening tests, anxiety, radiation exposure from some screening tests, and discomfort from some interventions and screening tests. The risk of side effects from preventive medications is analogous to the use of medications in therapeutic settings and is considered in the U.S. Food and Drug Administration (FDA) approval process. Side effects from currently recommended vaccinations are primarily limited to discomfort and minor immune reactions. However, the concern about associations between vaccinations and serious adverse outcomes continues to limit the acceptance of many vaccinations despite the lack of data supporting the causal nature of these associations.

The possibility of a false-positive test occurs with nearly all screening tests, although the definition of what constitutes a false-positive result often varies across settings. For some tests such as screening mammography and screening chest CT, a false-positive result occurs when an abnormality is identified that is not malignant, requiring either a biopsy diagnosis or short-term follow-up. For other tests such as Pap smears, a false-positive result occurs because the test identifies a wide range of potentially premalignant states, only a small percentage of which would ever progress to an invasive cancer. This risk is closely tied to the risk of overdiagnosis in which the screening test identifies disease that would not have presented clinically in the patient's lifetime.

Assessing the degree of overdiagnosis from a screening test is very difficult given the need for long-term follow-up of an unscreened population to determine the true incidence of disease over time. Recent estimates suggest that as much as 15–40% of breast cancers identified by mammography screening and 15–37% of prostate cancers identified by prostate-specific antigen testing may never have presented clinically. Screening tests also have the potential to create unwarranted anxiety, particularly in conjunction with false-positive findings. Although multiple studies have documented increased anxiety through the screening process, there are few data suggesting this anxiety has long-term adverse consequences, including subsequent screening behavior. Screening tests that involve radiation (e.g., mammography, chest CT) add to the cumulative radiation exposure for the screened individual. The absolute amount of radiation is very small from any of these tests, but the overall impact of repeated exposure from multiple sources is still being determined. Some preventive interventions (e.g., vaccinations) and screening tests (e.g., mammography) may lead to discomfort at the time of administration, but again, there is little evidence of long-term adverse consequences.

WEIGHING THE BENEFITS AND HARMS

The decision to implement a population-based screening and prevention strategy requires weighing the benefits and harms, including the economic impact of the strategy. The costs include not only the expense of the intervention but also time away from work, downstream costs from false-positive results, "incidentalomas" or adverse events, and other potential harms. Cost-effectiveness is typically assessed by calculating the cost per year of life saved, with adjustment for the quality of life impact of different interventions and disease states (i.e., quality-adjusted life-year). Typically, strategies that cost $50,000–100,000 per quality-adjusted year of life saved are considered "cost-effective" **(Chap. 4)**.

The U.S. Preventive Services Task Force (USPSTF) is an independent panel of experts in preventive care that provides evidence-based recommendations for screening and preventive strategies based on an assessment of the benefit-to-harm ratio **(Tables 6-4 and 6-5)**. Because there are multiple advisory organizations providing recommendations for preventive services, the agreement among the organizations varies across the different services. For example, all advisory groups support screening for hyperlipidemia and colorectal cancer, whereas consensus is lower for breast cancer screening among women in their forties and for prostate cancer screening. Because the guidelines are only updated periodically, differences across advisory organizations may also reflect the data that were available when the guideline was issued.

For many screening tests and preventive interventions, the balance of benefits and harms may be uncertain for the average-risk population but more favorable for individuals at higher risk for disease. Although age is the most commonly used risk factor for determining screening and prevention recommendations, the USPSTF also recommends some screening tests in populations based upon the presence of other risk factors for the disease. In addition, being at increased risk for the disease often supports initiating screening at an earlier age than that recommended for the average-risk population. For example, when there is a significant family history of colon cancer, it is prudent to initiate screening 10 years before the age at which the youngest family member was diagnosed with cancer.

TABLE 6-4 Screening Tests Recommended by the U.S. Preventive Services Task Force for Average-Risk Adults

DISEASE	TEST	POPULATION	FREQUENCY	CHAPTER
Abdominal aortic aneurysm	Ultrasound	Men 65–75 who have ever smoked	Once	
Alcohol misuse	Alcohol Use Disorders Identification Test	All adults	Unknown	464
Breast cancer	Mammography with or without clinical breast examination	Women (40?) 50–75	Every 2 years	84
Cervical cancer	Pap smear	Women 21–65	Every 3 years	75
	Pap smear and/or HPV testing	Women 30–65	Every 5 years if HPV negative	
Chlamydia/gonorrhea	Nucleic acid amplification test on urine or cervical swab	Sexually active women <25	Unknown	194
Colorectal cancer	Fecal occult blood testing	45–75	Every year	75, 86
	Fecal immunochemical-DNA	45–75	Every 1–3 years	
	Sigmoidoscopy	45–75	Every 5 years	
	Colonoscopy (or occult blood testing combined with sigmoidoscopy)	45–75	Every 10 years	
Depression+anxiety	Screening questions	All adults	Periodically	
Diabetes	Fasting blood glucose or HgbA1c	Adults overweight, obese, or with hypertension	Every 3 years	415
Hepatitis C	Anti-HCV antibody followed by confirmatory PCR	18–79	Once	350
HIV	Reactive immunoassay or rapid HIV followed by confirmatory test	15–65	At least once	208
Hyperlipidemia	Cholesterol	40–75	Unknown	419
Hypertension	Blood pressure	All adults	Periodically	288
Intimate partner violence	Screening questions	Women of childbearing age	Unknown	
Lung cancer	Low-dose computed tomography	Adults 50–80 years who have a 20-pack-year smoking history and currently smoke or have quit within the past 15 years	Yearly	83
Obesity	Body mass index	All adults	Unknown	414
Osteoporosis	DEXA	Women >65 or >60 with risk factors	Unknown	423

Abbreviations: DEXA, dual-energy x-ray absorptiometry; HCV, hepatitis C virus; HPV, human papillomavirus; PCR, polymerase chain reaction.

Source: Adapted from the U.S. Preventive Services Task Force, *https://www.uspreventiveservicestaskforce.org/uspstf/recommendation-topics/uspstf-a-and-b-recommendations*.

TABLE 6-5 Preventive Interventions Recommended for Average-Risk Adults

INTERVENTION	DISEASE	POPULATION	FREQUENCY	CHAPTER
Adult immunization				127, 129
COVID-19		>18		
Tetanus-diphtheria		>18	Every 10 years	
Varicella		Susceptibles only, >18	Two doses	
Measles-mumps-rubella		Women, childbearing age	One dose	
Pneumococcal		>64	20 valent option or 15/23	
Influenza		>18	Yearly	
Human papillomavirus		Up to age 27	If not done prior	
Zoster		>60	Once	
Chemoprevention				
Aspirin	Cardiovascular disease	Aged 40–59 years with a ≥10% 10-year cardiovascular disease risk (bleeding risk may = benefit for some groups)		
Folic acid	Neural tube defects in baby	Women planning or capable of pregnancy		
Tamoxifen/raloxifene	Breast cancer	Women at high risk for breast cancer		
Vitamin D	Fracture/falls	>64 at increased risk for falls		

Although informed consent is important for all aspects of medical care, shared decision-making may be a particularly important approach to decisions about preventive services when the benefit-to-harm ratio is uncertain for a specific population. For example, many expert groups, including the American Cancer Society, recommend an individualized discussion about prostate cancer screening because the decision-making process is complex and relies heavily on personal issues. Some men may decline screening, whereas others may be more willing to accept the risks of an early detection strategy. Another example of shared decision-making involves the choice of techniques for colon cancer screening **(Chap. 75)**. In controlled studies, the use of annual FOBT reduces colon cancer deaths by 15–30%. Flexible sigmoidoscopy reduces colon cancer deaths by ~40–60%. Colonoscopy appears to offer a greater benefit than flexible sigmoidoscopy with a reduction in risk of ~50–70% under optimal conditions of adherence and adenoma detection rates, but its use incurs additional costs and risks. These screening procedures have not been compared directly in the same population, but models suggest that appropriate frequencies of each technique may be associated with similar numbers of lives saved and cost to society per life saved ($10,000–25,000). Thus, although one patient may prefer the ease of preparation and less time disruption of a DNA-based stool test, others may prefer the sedation, thoroughness, and time interval of colonoscopy.

COUNSELING ON HEALTHY BEHAVIORS

In considering the impact of preventive services, it is important to recognize that tobacco and alcohol use, diet, and exercise constitute the vast majority of factors that influence preventable deaths in developed countries. Perhaps the single greatest preventive health care measure is to help patients quit smoking **(Chap. 465)**. However, efforts in these areas frequently require behavior changes (e.g., weight loss, exercise) or the management of addictive conditions (e.g., tobacco and alcohol use) that are often recalcitrant to intervention. Although these are challenging problems, evidence strongly supports the role of counseling by health care providers **(Table 6-6)** in effecting health behavior change. Educational campaigns, public policy changes, and community-based interventions have also proven to be important parts of a strategy for addressing these factors in some settings. Although the USPSTF found that the evidence was conclusive to recommend a relatively small set of counseling activities, counseling in areas such as physical activity and injury prevention (including seat belts and bicycle and motorcycle helmets) has become a routine part of primary care practice. Preventing and treating excess weight is also a top priority given the multiple complications of obesity **(Chap. 414)**.

IMPLEMENTING DISEASE PREVENTION AND SCREENING

The implementation of disease prevention and screening strategies in practice is challenging. A number of techniques can assist physicians with the delivery of these services. An appropriately configured electronic health record can provide reminder systems that make it easier for physicians to track and meet guidelines. Some systems give patients secure access to their medical records, providing an additional means to enhance adherence to routine screening. Systems that provide nurses and other staff with standing orders are effective for immunizations. The USPSTF has developed flow sheets and electronic tools to assist clinicians (*https://www.uspreventiveservicestaskforce.org/uspstf/about-uspstf/task-force-resources*). Many of these tools use age categories to help guide implementation. Age-specific recommendations for screening and counseling are summarized in **Table 6-7**.

Many patients see a physician for ongoing care of chronic illnesses, and this visit provides an opportunity to include a "measure of prevention" for other health problems. For example, a patient seen for management of hypertension or diabetes can have breast cancer screening incorporated into one visit and a discussion about colon cancer screening at the next visit. Other patients may respond more favorably to a clearly defined visit that addresses all relevant screening and prevention interventions. Because of age or comorbidities, it may be appropriate with some patients to abandon certain screening and prevention activities, although there are fewer data about when to "sunset" these services. For many screening tests, the benefit of screening does not accrue until 5–10 years of follow-up, and there are generally few data to support continuing screening for most diseases past age 75. In addition, for patients with advanced diseases and limited life expectancy, there is considerable benefit from shifting the focus from screening procedures to the conditions and interventions more likely to affect quality and length of life.

TABLE 6-6 Preventive Counseling Recommended by the U.S. Preventive Services Task Force (USPSTF)

TOPIC	CHAPTER REFERENCE
Alcohol and drug use	464, 467, 468
Genetic counseling for *BRCA1/2* testing among women at increased risk for deleterious mutations	84, 480
Nutrition and diet	343, 344
Sexually transmitted infections	141, 208
Sun exposure	64
Tobacco use	465

TABLE 6-7 Age-Specific Causes of Mortality and Corresponding Preventive Options

AGE GROUP	LEADING CAUSES OF AGE-SPECIFIC MORTALITY	SCREENING PREVENTION INTERVENTIONS TO CONSIDER FOR EACH SPECIFIC POPULATION
15–24	1. Accident 2. Homicide 3. Suicide 4. Malignancy 5. Heart disease	• Counseling on routine seat belt use, bicycle/motorcycle/ATV helmets (1) • Counseling on diet and exercise (5) • Discuss dangers of alcohol use while driving, swimming, boating (1) • Assess and update vaccination status (tetanus, diphtheria, hepatitis B, MMR, rubella, varicella, meningitis, HPV, COVID-19) • Ask about gun use and/or gun possession (2,3) • Assess for substance abuse history including alcohol (2,3) • Screen for domestic violence (2,3) • Screen for depression and/or suicidal/homicidal ideation (2,3) • Pap smear for cervical cancer screening after age 21 (4) • Discuss skin, breast awareness, and testicular self-examinations (4) • Recommend UV light avoidance and regular sunscreen use (4) • Measurement of blood pressure, height, weight, and body mass index (5) • Discuss health risks of tobacco use, consider emphasis on cosmetic and economic issues to improve quit rates for younger smokers (4,5) • Chlamydia and gonorrhea screening and contraceptive counseling for sexually active females, discuss STD prevention • Hepatitis B, and syphilis testing if there is high-risk sexual behavior(s) or any prior history of sexually transmitted disease • Hepatitis C screening starting at age 18 to 79 • HIV testing • Continue annual influenza vaccination
25–44	1. Accident 2. Malignancy 3. Heart disease 4. Suicide 5. Homicide 6. HIV	*As above plus consider the following:* • Readdress smoking status, encourage cessation at every visit (2,3) • Obtain detailed family history of malignancies and begin early screening/prevention program if patient is at significant increased risk (2) • Assess all cardiac risk factors (including screening for diabetes and hyperlipidemia); consider statin therapy for higher-risk patients • Assess for chronic alcohol abuse, risk factors for viral hepatitis, or other risks for development of chronic liver disease • Consider individualized breast cancer screening with mammography at age 40 (2)
45–64	1. Malignancy 2. Heart disease 3. Accident 4. Diabetes mellitus 5. Cerebrovascular disease 6. Chronic lower respiratory disease 7. Chronic liver disease and cirrhosis 8. Suicide	• Consider prostate cancer screen with annual PSA and digital rectal examination at age 50 (or possibly earlier in African Americans or patients with family history) (1) • Begin colorectal cancer screening at age 45 or 50 with fecal occult blood testing, stool DNA testing, flexible sigmoidoscopy, or colonoscopy (1) • Reassess and update vaccination status at age 50 and vaccinate all smokers against *Streptococcus pneumoniae* at age 50 (6) • Consider screening for coronary disease in higher-risk patients (2,5) • Zoster vaccination at age 60 • Begin mammography screening by age 50 • Lung cancer screening at age 50–80 years if a 20-pack-year smoking history and currently smoke or have quit within the past 15 years, yearly
≥65	1. Heart disease 2. Malignancy 3. Cerebrovascular disease 4. Chronic lower respiratory disease 5. Alzheimer's disease 6. Influenza and pneumonia 7. Diabetes mellitus 8. Kidney disease 9. Accidents 10. Septicemia	*As above plus consider the following:* • Readdress smoking status, encourage cessation at every visit (1,2,3,4) • One-time ultrasound for AAA in men 65–75 who have ever smoked • Consider pulmonary function testing for all long-term smokers to assess for development of chronic obstructive pulmonary disease (4,6) • Screen all postmenopausal women (and all men with risk factors) for osteoporosis • Continue annual influenza vaccination and vaccinate against *S. pneumoniae* at age 65 (4,6) • Screen for visual and hearing problems, home safety issues, and elder abuse (9) • Consider fall prevention exercise intervention if at higher risk (9)

Note: The numbers in parentheses refer to areas of risk in the mortality column affected by the specified intervention.

Abbreviations: AAA, abdominal aortic aneurysm; ATV, all-terrain vehicle; HPV, human papillomavirus; MMR, measles-mumps-rubella; PSA, prostate-specific antigen; STD, sexually transmitted disease; UV, ultraviolet.

Acknowledgment

The author is grateful to Dr. Katrina A. Armstrong for contributions to this chapter in prior editions of Harrison's.

FURTHER READING

Bretthauer M et al: Effect of colonoscopy screening on risks of colorectal cancer and related death. N Engl J Med 387:1547, 2022.

Davidson KW et al: Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. JAMA 325:1965, 2021.

Hugosson J et al: Mortality results from the Goteborg randomized population-based prostate-cancer screening trial. Lancet Oncol 11:725, 2010.

Woloshin S et al: The new USPSTF mammography recommendations—a dissenting view. N Engl J Med 389:1061, 2023.

7

The Safety and Quality of Health Care

David W. Bates

Safety and quality are two of the central dimensions of health care. In recent years, it has become easier to measure safety and quality, and it is increasingly clear that performance in both dimensions could be much better. The public is—with good justification—demanding measurement and accountability, and payment for services will increasingly be based on performance in these areas. Thus, physicians must learn about these two domains, how they can be improved, and the relative strengths and limitations of the current ability to measure them.

Safety and quality are closely related but do not completely overlap. The Institute of Medicine has suggested in a seminal series of reports that safety is the first part of quality and that the health care system must first and foremost guarantee that it will deliver safe care, although quality is also pivotal. In the end, it is likely that more net clinical benefit will be derived from improving quality than from improving safety, though both are important and safety is in many ways more tangible to the public. The first section of this chapter will address issues relating to the safety of care and the second will cover quality of care.

SAFETY IN HEALTH CARE

Safety Theory and Systems Theory *Safety theory* clearly points out that individuals make errors all the time. Think of driving home from the hospital: you intend to stop and pick up a quart of milk on the way home but find yourself entering your driveway without realizing how you got there. Everybody uses low-level, semiautomatic behavior for many activities in daily life; this kind of error is called a *slip*. Slips occur often during care delivery—e.g., when people intend to write an order but forget because they must complete another action first. *Mistakes*, by contrast, are errors of a higher level; they occur in new or nonstereotypic situations in which conscious decisions are being made. An example would be dosing of a medication with which a physician is not familiar. The strategies used to prevent slips and mistakes are often different.

Systems theory suggests that most accidents occur as the result of a series of small failures that happen to line up in an individual instance so that an accident can occur **(Fig. 7-1)**. It also suggests that most individuals in an industry such as health care are trying to do the right thing (e.g., deliver safe care) and that most accidents thus result from defects in systems. Systems should be designed both to make errors less likely and to identify those that do inevitably occur.

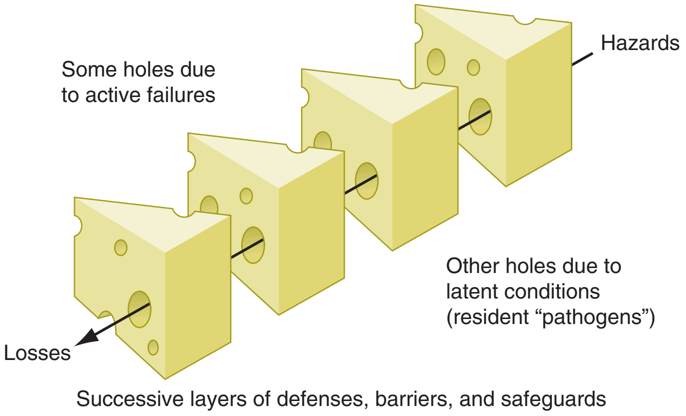
Hazards
Some holes due to active failures
Other holes due to latent conditions (resident "pathogens")
Losses
Successive layers of defenses, barriers, and safeguards

FIGURE 7-1 "Swiss cheese" diagram. Reason argues that most accidents occur when a series of "latent failures" are present in a system and happen to line up in a given instance, resulting in an accident. Examples of latent failures in the case of a fall might be that the unit is unusually busy and the floor happens to be wet. *(Adapted from J Reason: BMJ 320:768, 2000.)*

Factors That Increase the Likelihood of Errors Many factors ubiquitous in health care systems can increase the likelihood of errors, including fatigue, stress, interruptions, complexity, and transitions. The effects of fatigue in other industries are clear, but its effects in health care have been more controversial until recently. For example, the accident rate among truck drivers increases dramatically if they work over a certain number of hours in a week, especially with prolonged shifts. A study of house officers in the intensive care unit demonstrated that they were about one-third more likely to make errors when they were on a 24-h shift than when they were on a schedule that allowed them to sleep 8 h the previous night. The American College of Graduate Medical Education moved to address this issue by putting in place the 80-h workweek. Although this stipulation is a step forward, it does not address the most important cause of fatigue-related errors: extended-duty shifts. High levels of stress and heavy workloads also can increase error rates. Thus, in extremely high-pressure situations, such as cardiac arrests, errors are more likely to occur. Strategies such as using protocols in these settings can be helpful, as can simple recognition that the situation is stressful.

Interruptions also increase the likelihood of error and occur frequently in health care delivery. It is common to forget to complete an action when one is interrupted partway through it by a page, for example. Approaches that may be helpful in this area include minimizing interruptions and setting up tools that help define the urgency of an interruption.

Complexity represents a key issue that contributes to errors. Providers are confronted by streams of data (e.g., laboratory tests and vital signs), many of which provide little useful information but some of which are important and require action or suggest a specific diagnosis. Tools that emphasize specific abnormalities or combinations of abnormalities may be helpful in this area.

Transitions between providers and settings are also common in health care, especially with the advent of the 80-h workweek, and generally represent points of vulnerability. Tools that provide structure in exchanging information—for example, when transferring care between providers—may be helpful.

The Frequency of Adverse Events in Health Care Most large studies focusing on the frequency and consequences of adverse events have been performed in the inpatient setting; some data are available for nursing homes, but much less information is available about the outpatient setting. The Harvard Medical Practice Study, one of the largest studies to address this issue, was performed with hospitalized patients in New York. The primary outcome was the adverse event: an injury caused by medical management rather than by the patient's underlying disease. In this study, an event either resulted in death or disability at discharge or prolonged the length of hospital stay by at least 2 days. Key findings were that the adverse event rate was 3.7% and that 58% of the adverse events were considered preventable. Although New York is not representative of the United States as a whole, the study was replicated later in Colorado and Utah, where the rates were essentially similar. Several recent studies suggest that the frequency of harm related to medical care now approaches one in four admissions. The rates appear to be higher for several reasons—the techniques for finding events have improved with the use of "triggers" such as an unexpected transfer to the intensive care unit; records are now electronic and are easier to search; and the complexity of care continues to grow. Overall, it is quite concerning that rates remain so high, even though the frequency of some types of harm such as hospital-acquired infections appears to have decreased.

In the Harvard Medical Practice Study, adverse drug events (ADEs) were most common, accounting for 19% of all adverse events, and were followed in frequency by wound infections (14%) and technical complications (13%). Almost half of adverse events were associated with a surgical procedure. Among nonoperative events, 37% were ADEs, 15% were diagnostic mishaps, 14% were therapeutic mishaps, 13% were procedure-related mishaps, and 5% were falls.

ADEs have been studied more than any other error category. Studies focusing specifically on ADEs have found that they appear to be much more common than was suggested by the Harvard Medical Practice Study, although most other studies use more inclusive criteria. Detection approaches in the research setting include chart review and the use of a computerized ADE monitor, a tool that explores the database and identifies signals that suggest an ADE may have occurred. Studies that use multiple approaches find more ADEs than does any individual approach, and this discrepancy suggests that the true underlying rate in the population is higher than would be identified by a single approach. About 6–10% of patients admitted to U.S. hospitals experience an ADE.

Injuries caused by drugs are also common in the outpatient setting. One study found a rate of 21 ADEs per every 100 patients per year when patients were called to assess whether they had had a problem with one of their medications. The severity level was lower than in the inpatient setting, but approximately one-third of these ADEs were preventable.

The period immediately after a patient is discharged from the hospital appears to be very risky. A recent study of patients hospitalized on a medical service found an adverse event rate of 19%; about one-third of those events were preventable, and another one-third were ameliorable (i.e., they could have been made less severe). ADEs were the single leading error category.

Prevention Strategies Most work on strategies to prevent adverse events has targeted specific types of events in the inpatient setting, with nosocomial infections and ADEs having received the most attention. Nosocomial infection rates have been reduced greatly in intensive care settings, especially by using checklists. For ADEs, several strategies have been found to reduce the medication error rate, although it has been harder to demonstrate that they reduce the ADE rate overall, and no studies with adequate power to show a clinically meaningful reduction have been published.

Implementation of checklists to ensure that specific actions are carried out has had a major impact on rates of catheter-associated bloodstream infection and ventilator-associated pneumonia, two of the most serious complications occurring in intensive care units. The checklist concept is based on the premise that several specific actions can reduce the frequency of these issues; when these actions are all taken for every patient, the result has been an extreme reduction in the frequency of the associated complication. These practices have been disseminated across wide areas in the state of Michigan.

Computerized physician order entry (CPOE) linked with clinical decision support reduces the rate of serious medication errors, defined as those that harm someone or have the potential to do so. In one study, CPOE, even with limited decision support, decreased the serious medication error rate by 55%. CPOE can prevent medication errors by suggesting a default dose, ensuring that all orders are complete (e.g., that they include dose, route, and frequency), and checking orders for allergies, drug–drug interactions, and drug–laboratory issues. In addition, clinical decision support can suggest the right dose for a patient, tailoring it to the level of renal function and age. In one study, patients with renal insufficiency received the appropriate dose only one-third of the time without decision support, whereas that fraction increased to approximately two-thirds with decision support; moreover, with such support, patients with renal insufficiency were discharged from the hospital half a day earlier. As of 2019, over 95% of U.S. hospitals had implemented CPOE, although the decision support often is still limited.

Another technology that can improve medication safety is bar coding linked with an electronic medication administration record. Bar coding can help ensure that the right patient gets the right medication at the right time. Electronic medication administration records can make it much easier to determine what medications a patient has received. Studies to assess the impact of bar coding on medication safety are under way, and the early results are promising. Another technology to improve medication safety is "smart pumps." These pumps can be set according to which medication is being given and at what dose; the health care professional will receive a warning if too high a dose is about to be administered.

The National Safety Picture Several organizations, including the National Quality Forum and The Joint Commission, have made recommendations for improving safety. The National Quality Forum has released recommendations to U.S. hospitals about what practices will most improve the safety of care, and all hospitals are expected to implement these recommendations. Many of these practices arise frequently in routine care. One example is "readback," the practice of recording all verbal orders and immediately reading them back to the physician to verify the accuracy of what was heard. Another is the consistent use of standard abbreviations and dose designations; some abbreviations and dose designations are particularly prone to error (e.g., 7U may be read as 70).

Measurement of Safety Measuring the safety of care is difficult and expensive, since adverse events are, fortunately, rare. Most hospitals rely on spontaneous reporting to identify errors and adverse events, but the sensitivity of this approach is very low, with only ~1 in 20 ADEs reported. Promising research techniques involve searching the electronic record for signals suggesting that an adverse event has occurred. These methods are not yet in wide use but will probably be used routinely in the future. Claims data have been used to identify the frequency of adverse events; this approach works much better for surgical care than for medical care and requires additional validation. The net result is that, except for a few specific types of events (e.g., falls and nosocomial infections), hospitals have little idea about the true frequency of safety issues.

Nonetheless, all providers have the responsibility to report problems with safety as they are identified. All hospitals have spontaneous reporting systems, and if providers report events as they occur, those events can serve as lessons for subsequent improvement.

Conclusions about Safety It is abundantly clear that the safety of health care can be improved substantially—nearly one inpatient in four suffers harm today. As more areas are studied closely, more problems are identified. Much more is known about the epidemiology of safety in the inpatient setting than in outpatient settings. Many effective strategies for improving inpatient safety have been identified, and they are increasingly being applied. Some effective strategies are also available for the outpatient setting. Transitions appear to be especially risky. The solutions to improving care often entail the consistent use of systematic techniques such as checklists and often involve leveraging of information technology. Nevertheless, solutions will also include many other domains, such human factors techniques, team training, and a culture of safety.

QUALITY IN HEALTH CARE

Assessment of quality of care has remained somewhat elusive, although the tools for this purpose have increasingly improved. Selection of health care and measurement of its quality are components of a complex process.

Quality Theory Donabedian has suggested that quality of care can be categorized by type of measurement into structure, process, and outcome. *Structure* refers to whether a particular characteristic is applicable in a particular setting—e.g., whether a hospital has a catheterization laboratory or whether a clinic uses an electronic health record (EHR). *Process* refers to the way care is delivered; examples of process measures are whether a Pap smear was performed at the recommended interval or whether an aspirin was given to a patient with a suspected myocardial infarction. *Outcome* refers to what happens—e.g., the mortality rate in myocardial infarction. It is important to note that good structure and process do not always result in a good outcome. For instance, a patient may present with a suspected myocardial infarction to an institution with a catheterization laboratory and receive recommended care, including aspirin, but still die because of the infarction.

Quality theory also suggests that overall quality will be improved more in the aggregate if the performance level of all providers is raised rather than if a few poor performers are identified and punished. This view suggests that systems changes are especially likely to be helpful in improving quality, since large numbers of providers may be affected simultaneously.

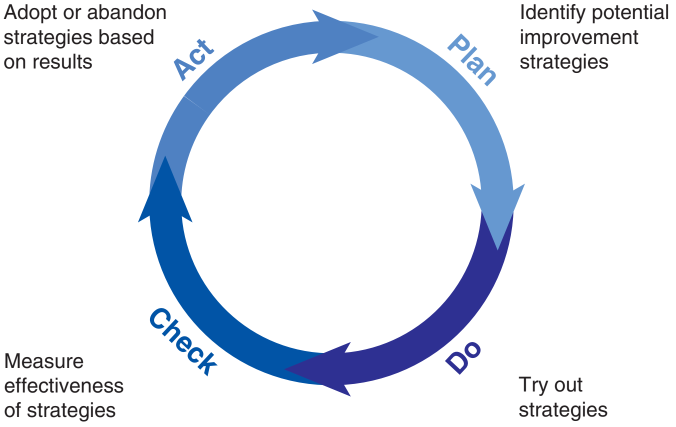
Adopt or abandon strategies based on results
Act
Plan
Identify potential improvement strategies
Check
Do
Measure effectiveness of strategies
Try out strategies

FIGURE 7-2 Plan-Do-Check-Act cycle. This approach can be used to improve a specific process rapidly. First, planning is undertaken, and several potential improvement strategies are identified. Next, these strategies are evaluated in small "tests of change." "Checking" entails measuring whether the strategies have appeared to make a difference, and "acting" refers to acting on the results.

The theory of *continuous quality improvement* suggests that organizations should be evaluating the care they deliver on an ongoing basis and continually making small changes to improve their individual processes. This approach can be very powerful if embraced over time.

Several specific tools have been developed to help improve process performance. One of the most important is the Plan-Do-Check-Act cycle **(Fig. 7-2)**. This approach can be used for "rapid cycle" improvement of a process—e.g., the time that elapses between a diagnosis of pneumonia and administration of antibiotics to the patient. Some statistical tools, such as control charts, are often used in conjunction to determine whether progress is being made. Because most medical care includes one or many processes, this tool is especially important for improvement.

Factors Relating to Quality Many factors can decrease the level of quality, including stress to providers, high or low levels of production pressure, and poor systems. Stress can have an adverse effect on quality because it can lead providers to omit important steps, as can a high level of production pressure. Low levels of production pressure sometimes can result in worse quality, as providers may be bored or have little experience with a specific problem. Poor systems can have a tremendous impact on quality, and even extremely dedicated providers typically cannot achieve high levels of performance if they are operating within a poor system.

Data About the Current State of Quality A study published by the RAND Corporation in 2006 provided the most complete picture of quality of care delivered in the United States to date. The results were sobering. The authors found that, across a wide range of quality parameters, patients in the United States received only 55% of recommended care overall; there was little variation by subtype, with scores of 54% for preventive care, 54% for acute care, and 56% for care of chronic conditions. The authors concluded that, in broad terms, the chances of getting high-quality care in the United States were little better than those of winning a coin flip.

Work from the Dartmouth Atlas of Health Care evaluating geographic variation in use and quality of care demonstrates that, despite large variations in utilization, there is no positive correlation between the two variables at the regional level. An array of data demonstrate, however, that providers with larger volumes for specific conditions, especially for surgical conditions, do have better outcomes.

Strategies for Improving Quality and Performance Many specific strategies can be used to improve quality at the individual level, including rationing, education, feedback, incentives, and penalties. *Rationing* has been effective in some specific areas, such as persuading physicians to prescribe within a formulary, but it generally has been resisted. *Education* is effective in the short run and is necessary for changing opinions, but its effect decays fairly rapidly with time. *Feedback* on performance can be given at either the group or the individual level. Feedback is most effective if it is individualized and is given in close temporal proximity to the original events. *Incentives* can be effective, and many believe that they will prove to be a key to improving quality, especially if pay-for-performance with sufficient incentives is broadly implemented (see below). *Penalties* produce provider resentment and are rarely used in health care.

Another set of strategies for improving quality involves changing the systems of care. An example would be introducing reminders about which specific actions need to be taken at a visit for a specific patient—a strategy that has been demonstrated to improve performance in certain situations, such as the delivery of preventive services. Another approach that has been effective is the development of "bundles" or groups of quality measures that can be implemented together with a high degree of fidelity. Many hospitals have implemented a bundle for ventilator-associated pneumonia in the intensive care unit that includes five measures (e.g., ensuring that the head of the bed is elevated). These hospitals have been able to improve performance substantially. Another technique is SCAMPs, or Standardized Clinical Assessment and Management Plans. These are care guidelines developed by clinicians who identify key steps in workflow and decisions to help improve the process outcomes.

Perhaps the most pressing need is to improve the quality of care for chronic diseases. The Chronic Care Model has been developed by Wagner and colleagues **(Fig. 7-3)**; it suggests that a combination of strategies is necessary (including self-management support, changes in delivery system design, decision support, and information systems) and that these strategies must be delivered by a practice team composed of several providers, not just a physician.

Available evidence about the relative efficacy of strategies in reducing hemoglobin A$_{1c}$ (HbA$_{1c}$) in outpatient diabetes care supports this general premise. It is especially notable that the outcome was the HbA$_{1c}$ level, as it has generally been much more difficult to improve outcome measures than process measures (such as whether HbA$_{1c}$ was measured). In this meta-analysis, a variety of strategies were effective, but the most effective ones were the use of team changes and the use of a case manager. When cost-effectiveness is considered in addition, it appears likely that an amalgam of strategies will be needed. However, the more expensive strategies, such as the use of case managers, probably will be implemented widely only if pay-for-performance takes hold.

The evidence linking better performance on quality metrics assessing process and outcomes varies greatly by condition. For example, there is strong evidence that performing Pap smears results in better

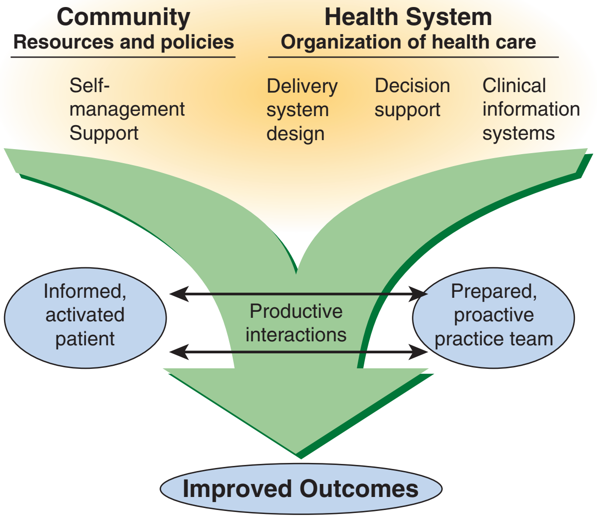

FIGURE 7-3 The Chronic Care Model, which focuses on improving care for chronic diseases, suggests that (1) delivery of high-quality care requires a range of strategies that must closely involve and engage the patient and (2) team care is essential. *(From EH Wagner et al: Eff Clin Pract 1:2, 1998.)*

outcomes in patients who develop cervical cancer, but the evidence for many other conditions is far more tenuous.

National State of Quality Measurement In the inpatient setting, quality measurement is now being performed by a very large proportion of hospitals for several conditions, including myocardial infarction, congestive heart failure, pneumonia, and surgical infection prevention; 20 measures are included in all. This is the result of the Hospital Quality Initiative, which represents a collaboration among many entities, including the Hospital Quality Alliance, The Joint Commission, the National Quality Forum, and the Agency for Healthcare Research and Quality. The data are housed at the Centers for Medicare and Medicaid Services, which publicly releases performance data on the measures on a website called *Hospital Compare* (*www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalCompare.html*). These data are reported voluntarily and are available for a very high proportion of the nation's hospitals. Analyses demonstrate substantial regional variation in quality and important differences among hospitals. Analyses by The Joint Commission for similar indicators reveal that performance on measures by hospitals has improved over time and that, as might be hoped, lower performers have improved more than higher performers.

The biggest change recently in this domain is that Medicare is now moving to electronic clinical quality metrics (ECQMs). Historically, care has been measured mainly through claims data, but now it is being measured through data being extracted from EHRs, though these new metrics need to be validated. *https://www.cms.gov/medicare/regulations-guidance/promoting-interoperability-programs/electronic-clinical-quality-measures-basics#:~:text=CMS%20has%20finalized%20the%20Electronic%20clinical%20quality%20measure,to%20measure%20the%20quality%20of%20health%2.*

Public Reporting Overall, public reporting of quality data is becoming increasingly common. There are now commercial websites that have quality-related data for most regions of the United States, and these data can be accessed for a fee. Similarly, national data for hospitals are available. The evidence to date indicates that patients have not made much use of such data, but that the data have had an important effect on provider and organization behavior. Instead, patients have relied on provider reputation to make choices, partly because little information was available until very recently and the information that was available was not necessarily presented in ways that were easy for patients to access. Problems still exist with quality metrics; many can be "gamed," and even though providers are now nearly universally using EHRs, most metrics come from claims that include many inaccuracies. More metrics that leverage EHRs are sorely needed. However, many authorities think that, as more information about quality becomes available, it will become increasingly central to patients' choices about where to access care.

Pay-for-Performance Currently, providers in the United States get paid the same amount for a specific service, regardless of the quality of care delivered. The pay-for-performance theory suggests that, if providers are paid more for higher-quality care, they will invest in strategies that enable them to deliver that care. The current key issues in the pay-for-performance debate relate to (1) how effective it is, (2) what levels of incentives are needed, and (3) what perverse consequences are produced. The evidence on effectiveness is limited, although several studies are ongoing. With respect to incentive levels, most quality-based performance incentives have accounted for merely 1–2% of total payment in the United States to date. In the United Kingdom, however, 40% of general practitioners' salaries have been placed at risk according to performance across a wide array of parameters; this approach has been associated with substantial improvements in reported quality performance, although it is still unclear to what extent this change represents better performance versus better reporting. The potential for perverse consequences exists with any incentive scheme. One problem is that, if incentives are tied to outcomes, there may be a tendency to transfer the sickest patients to other providers and systems. Another concern is that providers will pay too much attention to quality measures with incentives and ignore the rest of the quality picture. The validity of these concerns remains to be determined. Nonetheless, it appears likely that, under health care reform, the use of various pay-for-performance schemes is likely to increase.

■ CONCLUSIONS

The safety and quality of care in the United States could be improved substantially. Many available interventions have been shown to improve the safety of care and should be used more widely; others are undergoing evaluation or soon will be. Quality also could be dramatically better, and the science of quality improvement continues to mature. Medicare is rapidly moving toward electronic clinical quality measures. Implementation of value-based approaches such as accountable care that include pay-for-performance related to safety and quality should make it much easier for organizations to justify investments in improving safety and quality parameters, including health information technology. However, many improvements will also require changing the structure of care—e.g., moving to a more team-oriented approach and ensuring that patients are more involved in their own care. Payment reform focusing on value seems very likely to progress and will likely include both positive incentives and penalties related to safety and quality performance. Measures of safety are still relatively immature and could be made much more robust; it would be particularly useful if organizations had measures they could use in routine operations to assess safety at a reasonable cost, and substantial research is addressing this. Although the quality measures available are more robust than those for safety, they still cover a relatively small proportion of the entire domain of quality, and more measures need to be developed. The public and payers are demanding better information about safety and quality as well as better performance in these areas. The clear implication is that these domains will have to be addressed directly by providers.

■ FURTHER READING

Bates DW et al: Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. JAMA 280:1311, 1998.

Bates DW et al: The safety of inpatient health care. N Engl J Med 388:142, 2023.

Bates DW et al: Two decades since to err is human: An assessment of progress and emerging priorities in patient safety. Health Aff (Millwood) 37:1736, 2018.

Berwick DM: Era 3 for medicine and health care. JAMA 315:1329, 2016.

Brennan TA et al: Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. N Engl J Med 324:370, 1991.

Centers for Medicare and Medicaid Services: Inspector General's Report. *https://www.cms.gov/medicare/regulations-guidance/promoting-interoperability-programs/electronic-clinical-quality-measures-basics#:~:text=CMS%20has%20finalized%20the%20Electronic%20clinical%20quality%20measure,to%20measure%20the%20quality%20of%20health%20care%20provided*.

Chertow GM et al: Guided medication dosing for inpatients with renal insufficiency. JAMA 286:2839, 2001.

Institute of Medicine: Report: To err is human: Building a safer health system. 1999. *https://www.nap.edu/resource/9728/To-Err-is-Human-1999–report-brief.pdf*.

Institute of Medicine: Crossing the quality chasm: A new health system for the 21st century. 2001. *https://www.nap.edu/catalog/10027/crossing-the-quality-chasm-a-new-health-system-for-the*.

Landrigan C et al: Effect of reducing interns' work hours on serious medical errors in intensive care units. N Engl J Med 351:1838, 2004.

McGlynn EA et al: The quality of health care delivered to adults in the United States. N Engl J Med 348:2635, 2003.

Pronovost P et al: An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl J Med 355:2725, 2006. Erratum in: N Engl J Med 356:2660, 2007.

Starmer AJ et al: Rates of medical errors and preventable adverse events among hospitalized children following implementation of a resident handoff bundle. JAMA 310:2262, 2013.

8

The Value of the Physical Examination in Modern Medicine

Abraham Verghese, Steven McGee

Medicine is experiencing unprecedented progress in understanding the biology of disease and developing new therapies; concurrently, artificial intelligence (AI), with its ability to train on and interpret large medical data sets, is transforming many aspects of clinical medicine. Meanwhile, the rising costs of health care in the United States and many other nations have led to the merging of health care systems and the decline of individual physician practices. For the patient, the profoundly personal and often isolating nature of the illness experience remains largely unchanged since antiquity: individuals who are ill want to feel they are in the hands of a caring and attentive physician, and they will judge the quality of their care based on that interaction. It is not uncommon to hear patient dissatisfaction expressed as "my doctor never touched me" or "never laid a hand on me."

HISTORICAL EVOLUTION OF THE PHYSICAL EXAMINATION

A student of ancient medical texts finds few good descriptions of physical findings of disease; longstanding proscriptions against dissecting the dead meant that for centuries physicians had a poor understanding of normal anatomy or physiology. One exception is in the writings of Hippocrates (460–370 BCE), who described clubbing in the setting of empyema and the "succussion splash" (heard without a stethoscope) in a patient with hydropneumothorax. Only in 1543, with the publication of Vesalius's *De Humani Corporis Fabrica Libri Septem* (*On the fabric of the human body in seven books*) did things begin to change.

The Dawn of Physical Diagnosis In 1761, the Viennese physician Leopold Auenbrugger published his treatise, *Inventum Novum ex Percussione Thoracis Humani Interni Pectoris Morbos Detegendi* (*A new discovery that enables the physician from the percussion of the human thorax to detect the diseases hidden within the chest*). He recalled his father, an innkeeper, tapping the sides of casks of wine and listening for where the sound changed in character, thereby revealing the position of the fluid meniscus and how much wine remained. Auenbrugger could similarly, by tapping on his patients' chests, detect cardiomegaly as well as the presence of fluid in the pleural space. He confirmed the validity of his findings in a few autopsies. His discovery was revolutionary because it was the first means of "looking" into the living body; it was the ultrasound of its day. Auenbrugger's text was so complete that little new has been added to the art of percussion since, with the exception of "Skodaic resonance" (an arc of hyperresonance above a pleural effusion) and "Kronig's isthmus" (a narrow band of resonance in the supraclavicular area connecting the resonance of the front and back of the chest, an isthmus that is narrowed on one side in the presence of apical lobe diseases such as tuberculosis).

New Bedside Instruments Soon after percussion came the invention of the stethoscope by Laënnec in 1816, the ophthalmoscope by Helmholtz in 1850, the clinical thermometer by Wunderlich in 1868, the reflex hammer to elicit the muscle stretch reflexes described by Erb and Westphal in 1875, and the blood pressure cuff by Riva-Rocci in 1896. In the 1800s, Paris was the place to learn clinical skills under individuals such as Corvisart, Laënnec, Dupuytren, Bichat, and Pierre Louis. By the 1850s, the "German School" epitomized by Johannes Müller emerged, with its emphasis on experimental medicine and laboratory skills. For many North American physicians, including William Osler (1849–1919), who shaped medical education in America, a sojourn in Europe was obligatory. The use of tuning forks, oximeters, electronic stethoscopes, and Dopplers (for pulse detection and for ankle brachial index) and other ultrasound devices brings us to the present. Just as modern physicians are adopting point-of-care ultrasound and even pocket ultrasounds, the addition of new instruments has always been part of the evolution of physical diagnosis.

An Illustration of Clinical Examination Excellence Almost all diseases in the latter part of the nineteenth century and the early twentieth century were defined by their bedside findings, as clinical imaging had not yet developed. Physicians therefore prioritized careful bedside examination by means of inspection, palpation, percussion, and auscultation. Professor Joseph Bell (1837–1911) at the University of Edinburgh was legendary for his inferential skills. Among his medical students was the future physician and writer, Arthur Conan Doyle, who based his famous sleuth, Sherlock Holmes, on Professor Bell. The following anecdote in *Lancet* (by another one of Bell's students) illustrates Bell's skill in inspection before he begins palpation, percussion, and auscultation.

> *A woman with a small child was shown in. Joe Bell said good morning to her and she said good morning in reply. "What sort of a crossing di' ye have fra' Burntisland?"*
> *"It was guid."*
>
> *"And had ye a guid walk up Inverleith Row?"*
> *"Yes."*
>
> *"And what did ye do with th' other wain (child)?"*
> *"I left him with my sister in Leith."*
>
> *"And would ye still be working at the linoleum factory?"*
> *"Yes I am."*
>
> *Dr. Bell then explains to the students: "You see gentlemen, when she said good morning I noted her Fife accent, and, as you know, the nearest town in Fife is Burntisland. You noticed the red clay on the edges of the soles of her shoes, and the only such clay within 20 miles of Edinburgh is in the Botanical Gardens. Inverleith Row borders the gardens and is her nearest way here from Leith. You observed that the coat she carried over her arm is too big for the child who is with her, and therefore she set out from home with two children. Finally she has a dermatitis on the fingers of the right hand which is peculiar to workers in the linoleum factory at Burntisland."*

With the ascendance of diagnostic technology and laboratory testing, bedside examination skills declined, a trend that began in the 1970s and has accelerated. "Assessment drives learning" is an axiom of education; in medicine, the use of multiple-choice exams without a clinical assessment by direct observation of trainees as they examine patients with known physical findings diminishes the trainee's incentive to develop these skills.

FIVE REASONS THE PHYSICAL EXAMINATION REMAINS VERY RELEVANT

The bedside examination remains important and necessary for at least the following five reasons:

1. For a host of disorders, there are no laboratory or imaging studies that make the diagnosis, and thus the bedside findings *are* the gold standard; Parkinson's disease is an example.
2. Focused evidence-based physical examination maneuvers allow the physician to assign a greater or smaller value to the probability of a particular disease than prior to the examination and can thus direct and complement imaging and laboratory testing.
3. The physical examination frequently uncovers clues to other disease states that are asymptomatic or unrelated to the patient's presenting complaint but that are potentially treatable.
4. Missing critical findings in the examination can lead to errors that delay diagnosis and treatment; subject the patient to unnecessary contrast, radiation, or even surgery; and at times lead to significant morbidity and death. Such errors are embarrassing to the physician and can lead to malpractice claims.
5. The physical examination, when viewed through an anthropologic lens, has all the classic markers of a ritual. When performed well, the ritual elicits the patient's confidence and trust, while at the same time symbolically validating and localizing their disease or

symptom on the canvas of their body and in an organ, as opposed to on an image or biopsy report.

It is also worth noting that the physical examination is safe, immediately accessible, has no added cost, and uses no technology beyond the instruments carried by the clinician or available in the room. Its pedagogic value with trainees is in the opportunity to teach proper examination techniques as well as clinical reasoning. It allows the clinician to model empathy, consideration for the patient's comfort, and the establishment of a connection that builds trust and reduces anxiety.

The five themes are elaborated below, with particular attention to the second theme above.

When the Physical Examination Is the Sole Means of Making a Diagnosis For numerous medical disorders, the physical examination findings *are* the gold standard, the *only* means of making a diagnosis. This is the case for many diseases of the skin and of the eye (a reason why internists should be well versed in examining both organs and be familiar with common disorders associated with these). Amyotrophic lateral sclerosis is diagnosed principally by the clinical examination—there are no diagnostic blood tests or imaging findings. Table 8-1 lists other examples.

Evidence-Based Physical Examination Improves Diagnosis and Complements Imaging and Laboratory Testing Trainees often assume that evidence-based medicine (EBM) is properly focused on "external" data such as from laboratory tests or diagnostic images; these are perceived as more "objective." However, as the late clinical epidemiologist Alvan Feinstein (often thought of as one of the fathers of EBM and whose book *Clinical Judgment* remains a seminal work) took pains to emphasize, the *clinical* evidence base is of great importance. He wrote, "clinicians can bring science to clinical judgment by better exercise of the very human capacities that appear to impair it, and by giving increased attention not to laboratory substances and inanimate technology, but to sick people and the human methods of evaluating sick people." Rather than devaluing the objective data to be found on the clinical examination, physicians should take pride in its careful collection. These data include not just the physical findings but also the physician's unique understanding of the patient's history, background, social situation, work environment, aspirations, beliefs, family structure, and local trends in disease.

BEYOND THE STUDENT'S FIRST PHYSICAL DIAGNOSIS TEXTBOOK Medical students typically learn their patient examination skills in preclinical courses that utilize one of several excellent physical diagnosis textbooks. In later years, however, few students revisit these texts to remind themselves of the rationale for each test or the correct technique. This might be because when the student enters the hospital in their clinical years, they often find that the workflow revolves more around the virtual patient in the computer, the ordering and retrieving of tests, and data entry, with less emphasis on examining the embodied patient. The introductory physical examination texts remain an important and fundamental base from which to build skills. The trainee in internal medicine should progress to more advanced texts that embody Feinstein's philosophy of "bringing science to clinical judgment," such as *Evidence-Based Physical Diagnosis* by Steven McGee and the *JAMA Rational Clinical Exam* series begun in 1998 by David Simel and Drummond Rennie.

Trainees may falsely believe that physical examination findings are unreliable because clinicians can disagree about the presence or absence of a finding. Also, trainees may incorrectly assume that chest radiograph, computed tomography (CT) scan, or tissue biopsy reports are "gold standards," while physical examination findings are inaccurate.

RELIABILITY Whether it is the interpretation of a bedside finding or of a CT scan, human observations are accompanied by a certain degree of interobserver disagreement. In clinical studies, *interobserver agreement* is captured by the test statistic *kappa* or κ, which is a normalized measure of the increase in observed agreement over what would be expected by chance. Its value ranges from 0 (for agreement by chance) to 1 (for perfect agreement); 0.2–0.4 is considered fair agreement,

TABLE 8-1 Selected Examples of Conditions in Which the Physical Exam Is the Diagnostic Standard or the Primary Method of Diagnosis

DISEASE CONDITIONS	NOTES
Dermatology (selected list) Cellulitis, psoriasis, rosacea, acne vulgaris, eczema, urticaria, pityriasis rosea, cutaneous lichen planus, erythema multiforme, hereditary hemorrhagic telangiectasia, rubella, measles, herpes zoster, erythema nodosum	• Many dermatologic conditions are diagnosed by observation, even if confirmatory tests may exist • Skin biopsy not needed for the selected examples unless there is diagnostic uncertainty
Neurology Amyotrophic lateral sclerosis, Parkinson's disease (tremor, bradykinesia, rigidity; absence of atypical features), Bell's palsy	• Only the neurologic exam can determine the functional deficit, if any, of an abnormality on computed tomography scan or magnetic resonance imaging • Peripheral neuropathy is common; diagnosis is by symptoms and signs (decreased sensation, decreased or absent reflexes, and motor weakness); electromyography and nerve conduction studies are painful and must be ordered for specific reasons • Sensory testing for pain and temperature (carried by small unmyelinated fibers) may be more sensitive than conventional nerve conduction tests, which test larger myelinated nerves fibers • Patterns of sensory loss help localization: glove and stocking (peripheral neuropathy); radicular (root); Brown-Sequard type (spinal cord); "crossed" signs (cranial nerve deficits on one side and motor deficit on the other with brainstem lesions)
Rheumatology Rheumatoid arthritis, lupus erythematosus, scleroderma, relapsing polychondritis, patterns suggesting vasculitis (acute mononeuritis, palpable purpura)	• The clinical exam is essential for diagnosis and follow-up • In rheumatoid arthritis or lupus, serology alone does not make the diagnosis. A positive rheumatoid factor *without* multiple joint involvement, constitutional symptoms, rheumatoid nodules, morning stiffness, palpable synovitis, etc., is likely falsely positive. Seronegative rheumatoid arthritis is a clinical diagnosis
Cardiology Pericarditis	• Diagnosis depends on combinations of characteristic chest pain, pericardial friction rub, and characteristic electrocardiogram; 34–50% have normal echocardiograms
Ophthalmology Diabetic retinopathy, conjunctivitis, suspicion of keratitis, iritis, scleritis, acute angle glaucoma, retinal infarction	• "Red eye": the internist must recognize symptoms and signs that require immediate ophthalmology referral, including peri-limbic erythema (ciliary flush), photophobia, acute progressive pain, vision loss, foreign body sensation, anisocoria, fluid in the anterior chamber (hypopyon) • With conjunctivitis, the redness is most intense on the palpebral and peripheral bulbar conjunctiva • Bacterial versus viral versus allergic causes of conjunctivitis: bacterial discharge is purulent, sticky; eyes are stuck shut and crusted. Viral and allergic conjunctivitis can have mucoid, watery discharge. History of hay fever, itching, and previous episodes suggests allergic

0.4–0.6 is moderate agreement, and 0.6–0.8 is substantial agreement. It might surprise the trainee to learn, for example, that there is only fair to moderate agreement (κ of 0.38–0.58) in recognizing the presence of an infiltrate on chest radiograph, while there is better agreement (κ of 0.83) for recognizing interstitial edema. The interobserver agreement for the classification of coronary artery lesions on angiogram is only fair (κ of 0.33); it is slightly better for determining the severity of valvular regurgitation on echocardiogram (κ of 0.32–0.55). Pathologists interpreting liver biopsies show fair agreement in noting cholestasis (κ of 0.4) and moderate agreement on the existence of cirrhosis (κ of 0.59).

In the same manner, physical signs have varying degrees of interobserver agreement. Some signs, such as percussing for liver span or the use of auscultatory percussion, have notoriously low interobserver agreement and should be abandoned, while other signs have high agreement. A wide range in κ for a physical examination maneuver often reflects studies of varying rigor, different levels of training and experience, the transient nature of a sign, and other factors. Determining whether neck veins are normal or elevated shows a κ ranging from 0.08 to 0.71. Nevertheless, when the method of examination is agreed upon and the technique performed correctly, the interobserver agreement for many physical signs is good. The clock-drawing test (Wolf-Klein method) for dementia has a κ of 0.73; eliciting a positive abdominojugular test has a κ of 0.92, or almost perfect agreement.

DIAGNOSTIC ACCURACY: PRETEST PROBABILITY, LIKELIHOOD RATIO, AND POSTTEST PROBABILITY From the patient's history, the experienced clinician formulates hypotheses that are then accepted or rejected, or new ones are added as data are obtained by the physical examination (iterative hypothesis testing). For example, in a patient with cough and fever, the clinician might suspect pneumonia, and with those symptoms, the *pretest probability* of pneumonia is between 15 and 35%. It is useful for the clinician to have handy or memorize the pretest probabilities for common diagnoses, given particular clinical settings **(Table 8-2)**. In a patient with cirrhosis, for example, the pretest probability of their having hepatopulmonary syndrome (HPS), which considerably alters their prognosis, is in the range from 14 to 37%. The finding of clubbing in this patient has a high specificity of 64–96% for HPS, but a variable sensitivity of 12–91%.

Of more utility than sensitivity and specificity is the likelihood ratio (LR), a measure that allows the clinician to rapidly estimate *posttest probability*. The LR is calculated as the ratio of the probability of a particular finding in patients *with* disease (i.e., the sensitivity) divided by the probability of the identical finding in patients with mimicking conditions but *without* disease (i.e., the false-positive probability, which is 1 minus the specificity). LRs serve as diagnostic weights: values >1 increase probability from pretest to posttest (and the greater the LR, the more the probability increases); LRs with values <1 decrease probability (and the closer the value is to zero, the more probability decreases). McGee has popularized a useful rule of thumb in interpreting LRs: LRs of 2, 5, and 10 translate to increased probability of disease of 15, 30, and 45%, respectively, whereas LRs of 0.5, 0.2, and 0.1 reduce the probability of disease by 15, 30, and 45%, respectively. These estimates are accurate to within 5–10% of the actual calculated posttest probability and serve well for bedside decisions.

TABLE 8-3 Physical Signs and Their Likelihood Ratios

	LIKELIHOOD RATIO IF FINDING IS:	
DIAGNOSIS AND FINDING	**PRESENT**	**ABSENT**
Diagnosing ascites in patients with abdominal distension		
Edema	3.8	0.2
Flank dullness	NS	0.3
Fluid wave	5.0	0.5
Shifting dullness	2.3	0.4
Diagnosing hepatopulmonary syndrome in patients with chronic liver disease		
Clubbing	4.3	0.6
Cyanosis	4.4	0.7
Ascites	NS	NS
Jaundice	NS	NS
Diagnosing pneumonia in patients with cough and fever		
Percussion dullness	3.6	NS
Egophony	4.1	NS
Crackles	2.8	0.8
Bronchial breath sounds	3.3	0.9
Diagnosing elevated left heart filling pressures in patients with chest pain and dyspnea		
Heart rate >100 beats/min	5.5	NS
Crackles	NS	NS
Displaced apical impulse	5.8	NS
Positive abdominojugular test	8.0	0.3
S3 gallop	3.9	0.8

Abbreviation: NS, not significant (i.e., the likelihood ratio of the finding is statistically no different from the value of 1.0 and therefore useless to the clinician when considering this diagnosis).

Source: Adapted from SR McGee: *Evidence-Based Physical Diagnosis*, 5th ed. Philadelphia, Elsevier, 2022.

In the previous example of the patient with cirrhosis, the finding of finger clubbing in detecting HPS has an LR of 4.3 **(Table 8-3)**. Suppose the clinician's estimate of pretest probability of HPS is about 30%: the

TABLE 8-2 Pretest Probability of Selected Conditions[a]

		PRETEST PROBABILITY (%)	
CLINICAL SETTING	**DIAGNOSIS**	**MEDIAN**	**RANGE**
Hospitalized with fever	Bacteremia	18	7–37
Cough and fever	Pneumonia	22	15–35
Pleuritic chest pain, dyspnea, or hemoptysis	Pulmonary embolism	25	9–43
Murmur of aortic regurgitation	Moderate-to-severe aortic regurgitation	42	24–56
Chronic liver disease	Hepatopulmonary syndrome	25	14–37
Abdominal distension	Ascites	29	24–33
Ankle injury	Ankle fracture	12	10–14
Lymphadenopathy persisting several weeks	Serious disease (mostly cancer)	26	14–41
Diabetic foot ulcer	Osteomyelitis	65	52–68
Acute calf pain or swelling	Deep venous thrombosis	25	6–43

[a] These pretest probabilities derive from published studies of physical findings. Clinicians can further refine them by considering their own clinical experience. For example, the pretest probability of hepatopulmonary syndrome (range, 14–37%) is probably closer to 14% in primary care patients with liver disease and closer to 37% in hepatology patients.

Source: Adapted from SR McGee: *Evidence-Based Physical Diagnosis*, 5th ed. Philadelphia, Elsevier, 2022.

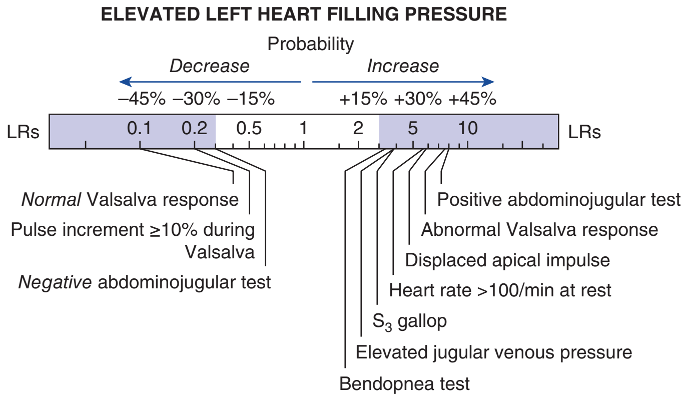

FIGURE 8-1 With just a glance, the clinician can immediately identify those few findings that significantly increase probability of heart failure (*right* side of ruler, likelihood ratio [LR] = 3 or more) and those that significantly decrease it (*left* side of ruler, LR = 0.3 or less). For example, the figure shows that the *presence* of S_3 gallop (LR = 3.9) and *presence* of displaced apical impulse (LR = 5.8) significantly increase probability of heart failure. On the other hand, the *absence* of either finding does not appear on the ruler because these LRs lie between 0.3 and 3.0 (Table 8-3) and therefore are diagnostically unhelpful. *(Reproduced with permission from SR McGee: Evidence-Based Physical Diagnosis, 4th ed. Philadelphia, Elsevier; 2022.)*

LR of 4.3 raises the probability of HPS by about 25–30%; adding this to the pretest probability results in a posttest probability of HPS of about 55–60%. One can also combine findings, if they have different pathogenesis (i.e., are independent). If the same patient with cirrhosis and clubbing also has cyanosis, which has an LR of 4.4 (increasing probability another 25–30%), the posttest probability of HPS (combining the two findings and their probabilities with the pretest probability) is now about 80–90%.

LRs allow the clinician to compare all traditional physical findings for a given diagnosis and quickly identify those few findings that accurately increase or decrease probability, a practice that improves the clinician's accuracy, efficiency, and confidence. Table 8-3 provides selected findings for four different conditions. If contemplating the diagnosis of ascites, for example, the clinician consults the table and focuses only on those findings with large LR values (3.0 or more, which increase probability) and those findings with LRs close to zero (0.3 or less, which decrease probability), disregarding those findings with values close to 1 (i.e., 0.3–3.0) because these latter values change probability minimally or not at all. In the example of abdominal distension, only the presence of the fluid wave (LR = 5) and edema (LR = 3.8) increase probability sufficiently, and only the absence of edema (LR = 0.2) and absence of flank dullness (which is the same as presence of flank tympany, LR = 0.3) decrease probability. These are the only findings the clinician applies at the bedside. LRs for a variety of conditions can be easily looked up in the two sources mentioned above. In *Evidence-Based Physical Diagnosis,* McGee has created an original and easily remembered graphic representation of physical signs related to a host of conditions **(Fig. 8-1)**.

PHYSICAL SIGNS AS DECISION POINTS IN PATIENT MANAGEMENT The clinician should also master some common physical diagnosis maneuvers and "rules" that are helpful in distinguishing serious conditions from more benign ones **(Table 8-4)**. These are sometimes called "stop rules" because the workup can safely end if all the conditions of the rule are met. One example is the HINTS battery (Head Impulse, Nystagmus, Skew Deviation Test) in evaluating the patient with dizziness: in emergency room patients with acute sustained vertigo, nausea, and vomiting, the combination of positive head impulse test, absence of direction-changing nystagmus, and absence of skew deviation markedly decreases the probability of posterior circulation stroke (LR = 0.02). This and a few other useful maneuvers and clinical rules are shown in Table 8-4.

Finding Clues to Unsuspected and Asymptomatic Disorders The attentive physician often picks up clues to diseases that the patient may be unaware of but that could be consequential. Noting diffuse enlargement of the thyroid (in goiter or in Grave's disease), or xanthelasma (hypercholesterolemia), or acanthosis nigricans (in insulin resistance) presents the opportunity to intervene. Such clues abound, as an observant physician in busy public places such as airports cannot help but notice. A common observation is the coxalgic gait: the patient's trunk leans dramatically to the side (the "lateral lurch") when bearing weight on a painful hip. On the other hand, if the patient's lateral lean over the hip is less dramatic and accompanied by a drop in the contralateral pelvis, the opposing sways of the shoulder and pelvis give the impression of a hinge between sacrum and lumbar spine. This is the Trendelenburg gait, a sign of weakness of the hip abductor, the gluteus medius. At one time, the Trendelenburg gait was commonly seen after polio or other neuromuscular diseases, or with congenital hip dislocation; more recently, it can be seen after damage to the gluteus medius or the superior gluteal nerve after hip arthroplasty by the lateral approach. Another commonly observed gait abnormality is "circumduction" of the foot (which swings in a small semicircle with each step instead of moving forward directly), along with a reduced arm swing on the same side, suggesting past hemiplegia. **Table 8-5** lists some other such observations that are meaningful to the clinical eye. Developing this type of clinical gaze requires conscious practice and alertness. Too narrow an examination based on the presenting symptom can miss important clues and be a disservice to the patient. Conversely, detecting such findings is satisfying because it might present the opportunity to intervene earlier in the course of a disease.

Medical Error from Oversights in the Physical Examination

Failure to do the examination, or an incomplete examination, can lead to diagnostic delay, inappropriate or delayed treatment, unnecessary exposure of the patient to radiation, surgical misadventure, or even death. Examples abound: the febrile patient whose petechiae or ecchymosis is missed because clothing is not completely removed, delaying the consideration of and empiric treatment for suspected Rocky Mountain spotted fever or meningococcemia; the patient with wrist and ankle pain seen by the primary care physician and referred

TABLE 8-4 Bedside Rules That Decrease Probability of Serious Conditions ("Stop Rules")

CONDITION	DIAGNOSIS	BEDSIDE RULE[a]	LIKELIHOOD RATIO IF BEDSIDE RULE SATISFIED
Ankle injury	Ankle fracture	Negative Ottawa ankle rule	0.1
Acute calf pain and swelling	Deep venous thrombosis	Original Wells score 0 or less[b]	0.2
Acute sustained vertigo, nausea, and vomiting	Posterior circulation stroke	HINTS peripheral	0.02
Acute abdominal pain	Acute appendicitis	Alvarado score 4 or less	0.1
Acute cough and fever	Pneumonia	Heckerling score 0 or 1	0.3
Diabetic foot ulcer	Osteomyelitis	Negative probe-to-bone test	0.2

[a]Definition of rules: for "Ottawa ankle rule," "Original Wells rule," and "Alvarado score," see reference (McGee); for "HINTS peripheral" see text; for "Heckerling score," the clinician scores 1 point for each of the following findings if present: temperature >37.8°C, heart rate >100 beats/min, crackles, diminished breath sounds, and *absence* of asthma; for "probe-to-bone" test, the clinician gently probes the foot ulcer with a blunt metal probe and identifies a rock-hard, gritty base without intervening soft tissue (positive test) or fails to observe this (negative test). [b]Clinicians combine a Wells score ≤0 with negative quantitative D-dimer before stopping workup.

Source: Adapted from SR McGee: *Evidence-Based Physical Diagnosis*, 5th ed. Philadelphia, Elsevier, 2022.

TABLE 8-5 Important Clues to Disorders That May Not Be Related to the Patient's Presenting Symptoms

TYPE OF OBSERVATION, WITH SELECTED EXAMPLES	NOTES
Faces and expression Acromegaly, Parkinson's disease, Cushing's syndrome, myxedema, hyperthyroidism, myasthenia gravis, Hippocratic facies, amiodarone facies, myotonic dystrophy, multiple endocrine neoplasia 2b (MEN2b), Down's syndrome, congenital syphilis, facial lipodystrophy with antiretroviral therapy in HIV, scleroderma	• Blue pigmentation around malar region with amiodarone • Myotonic dystrophy: "hatchet face" from temporal and facial muscle wasting, baldness in males, cataracts. • MEN2b: mucosal neuromas on lips, tongue, marfanoid habitus
Gait Coxalgic gait, Trendelenburg gait, high-stepping gait, Parkinson's gait, hemiplegic gait, diplegic gait or spastic gait, ataxic gait, sensory ataxic gait	• Parkinson's features are short, shuffling steps, forward flexion, appearance of hurrying up ("festination"), hesitancy in turning, absent arm swing, tremor • Peripheral neuropathy or posterior column disease cause sensory ataxia with broad-based "stamping" gait—striking foot down hard; worse at night when visual cues are diminished • Both hemiplegic and diplegic or "spastic" gait have hypertonia in upper limb flexors and lower limb extensors and adductors, with ankles extended and toes pointed in, circumduction with each step, "scissor gait" with diplegia (cerebral palsy)
Hands (Selected examples in the right column.) There are many eponymous *nail changes* (Beau's line, Terry's nails, Mee's line, Muehrcke's lines, half-and-half nails, etc.); they are frequently seen but not diagnostically helpful as they are associated with multiple conditions, including renal failure, liver failure, past or current severe illness, and hypoproteinemia. Koilonychia is associated with iron deficiency anemia but not exclusively. Pitting is seen in psoriasis and many other conditions.	• *Cardiac conditions:* splinter hemorrhages, Osler's nodes, Janeway lesions in endocarditis; "fingerization of the thumb" in Holt-Oram syndrome (with atrial septal defect); Marfan syndrome with arachnodactyly, thumb and wrist sign; Ehlers-Danlos syndrome with joint hypermobility, lax thin skin • *Liver dysfunction:* palmar erythema, spider angioma, white nails, asterixis • *Endocrine:* moist, warm, tremulous extremities, onycholysis in hyperthyroidism • *Neurologic:* myotonic grip with myotonic dystrophy; wrist drop of radial palsy; claw hand of ulnar nerve palsy • *Rheumatology:* nail fold and capillary changes in vasculitis; Heberden's and Bouchard's nodes in osteoarthritis; swan neck deformity, subluxation and ulnar deviation of fingers in rheumatoid arthritis and other chronic inflammatory arthritis; telescoping hand in destructive psoriatic or rheumatoid arthritis • *Pulmonary:* nicotine staining; clubbing; cyanosis • *Congenital/developmental changes:* Single transverse palmar crease in Down's syndrome
Odors	• Odor of tobacco on clothing • Grapelike odor of *Pseudomonas* wound infection • Fetid breath of anerobic lung abscess • Uriniferous odor in renal failure • Ammoniacal mousy odor in hepatic failure • Acetone-like fruity odor in diabetic ketoacidosis • Fish odor in trimethylaminuria • Bitter almond scent with cyanide poisoning • Alcohol metabolites with intoxication or alcohol-induced delirium
Clothing	• Inappropriate clothing for the ambient weather in hypothyroidism and hyperthyroidism • Neglect of clothing or color mismatch in dementia or delirium • Untied shoelaces in edema, toe of shoe cut out in chronic gout
Stature	• Short stature (growth hormone deficiency, Turner's syndrome) • Tall stature in Marfan syndrome and those with a Marfanoid habitus. In one study, the phenotypic features that favor Marfan syndrome are pectus carinatum, reduced elbow extension, high-arched or "gothic" palate, arm span–to–height ratio (ASHR) >1.05, hindfoot deformity, downslanting palpebral fissures, the thumb sign, lens subluxation, myopia, dental crowding, joint laxity, and micrognation

Source: Adapted from multiple sources including SL Berk, A Verghese: General Appearance, in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed. Walker HK et al (eds). Boston, Butterworths, 1990.

to consultants in rheumatology or orthopedics, and serologic tests for lupus, vasculitis, and other conditions ordered, until at some point the presence of clubbing (and even Horner's syndrome) is noted, suggesting pulmonary hypertrophic osteoarthropathy caused by a malignancy in the lung; or the patient with chest pain taken to the cardiac catheterization lab where contrast is injected before a rash looking like "dew drops on rose petals" involving a dermatome on the left chest is noted (herpes zoster). Studies show that when patients with "cellulitis" are first routed through a dermatology clinic, about a third will have an alternative diagnosis such as eczema or lymphedema, avoiding hospitalization and antibiotics.

RADIOLOGISTS REPORTING WHAT SHOULD HAVE BEEN OBVIOUS ON THE EXAMINATION It is unfortunately commonplace in hospital practice for the physician to be notified by the radiologist on the day after admission that the admission plain films of the abdomen show the patient labeled "gastroenteritis" has strangulated bowel in a hernial orifice or has gas in the scrotal tissue (Fournier's gangrene). These represent surgical emergencies that should be diagnosed by physical examination. Radiologists often report breast masses on CT scan of the chest, gynecomastia in a male, or thyroid masses, all of which should have been palpable. Such errors are consequential to the patient and an embarrassment to the clinician. Conversely (and less frequently), a careful physical examination might raise doubt on a radiologic interpretation. For example, even though all patients with acute atraumatic third cranial nerve palsy should undergo urgent neuroimaging, the "rule of the pupil" still applies: a dilated pupil suggests a compressive etiology such as an aneurysm, whereas a normal pupil suggests an ischemic cause. One particular study described two patients whose pupillary findings challenged the radiologic report. One patient's image showed a cavernous meningioma, which was suggested as causal, yet the pupil was spared; the clinicians elected to follow the patient, who made a full recovery, suggesting ischemia was the cause. A second patient had an abnormal pupil but a normal CT angiogram; noting this discrepancy, the clinicians discussed the dilated pupil with the radiologist, and on restudying the image, a posterior communicating aneurysm was noted.

THE CAUSE OF OVERSIGHTS IN THE PHYSICAL EXAMINATION This variety of medical error usually goes unnoted and is difficult to study. In one paper based on physician-reported anecdotes of such oversights (either their own or those of others), it was because the physical examination was cursory or had not been done, although the patient's chart suggested a complete exam. The electronic medical record with its templates, dropdown boxes, and cut-and-paste functions makes it easy to suggest (and bill for) a complete examination. These errors of omission and commission are magnified by rapid patient turnover; frequent "handoffs" from the emergency room to the night team and then to the morning admitting team, with overreliance placed on the initial diagnosis; and insufficient continuity of care for any particular physician to be fully invested in the patient. Pride in their profession and a desire to avoid egregious mistakes that harm the patient should give clinicians a healthy skepticism for diagnostic labels given to new patients handed over for continuing care.

The Physical Examination as a Ritual Meaningful to Patient and Physician Busy physicians may not appreciate that what may be routine for them—seeing a patient in the clinic or the hospital—is far from routine for the patient. An ethnographer observing a new patient–physician encounter sees one stranger presenting themselves to another stranger and revealing personal and sensitive information that they may not admit to their spouse or their spiritual advisor; the setting is a room that has unique furnishings not to be found in either individual's home. One of the two participants might be wearing a white shamanistic garment with specialized tools in the pockets, while the other is asked to don a cloth or paper gown for the occasion. Then, at some point, one participant disrobes and allows the other to touch and manipulate their limbs and body; peer into their eyes, ears, and throat; probe their hernial orifices; and at times, examine the genitalia and rectum. These actions are far from the norm in society and could even be construed as assault. The great privilege of being a physician is that the physical examination is part of the contractual agreement to provide care for the patient, and it therefore comes with great fiduciary responsibility. To the ethnographer, the patient–physician encounter has all the trappings of a ritual.

CHARACTERISTICS OF RITUALS All rituals (such as weddings, baptisms, funerals, inaugurations, or graduations) are characterized by the crossing of a threshold, by a commitment, and by some type of transformation. The physical examination ritual can signal the patient's transition from self-sufficiency to seeking help, or from illness to wellness; it also signals the placement of trust in the physician. The willingness to disrobe and allow touch—markers of vulnerability—indicates the patient's acceptance that this ritual is important for the transfer of knowledge.

PATIENTS ARE EXPERT JUDGES OF RITUAL Patients of different ethnicities and cultural backgrounds, and with differing beliefs about illness, nevertheless recognize and appreciate ritual, even when the physician fails to see it. Patients are also good judges of the skills of other professionals such as automobile mechanics, hairdressers, or cooks, and they can tell if the work is being done well and with pride. In studies where lay subjects assess videos of surgeons operating (crowdsourced assessment), the subjects' assessments when compared to that of experienced surgeons rating the same videos showed moderate to strong correlation. Patients can feel let down when a physical examination feels perfunctory, such as when it consists only of a prodding of the belly and the placing of the stethoscope on the clothing instead of on the skin. To quote William Osler, "Remember, however, that every patient upon whom you wait will examine you critically and form an estimate of you by the way in which you conduct yourself at the bedside. Skill and nicety in manipulation, whether in the simple act of feeling the pulse or in the performance of any minor operation, will do more towards establishing confidence in you than a string of Diplomas, or the reputation of extensive hospital experience." When well executed, the physical examination preserves the patient's sense of identity and affirms their humanity and personhood. It validates the presenting symptoms by localizing them on their soma, on and in their body, rather than on a distant radiologic report or a lab test. Imaging and laboratory tests strip away the markers of individuality and personhood. Patients who chance to see their own imaging studies see little that is recognizable to them as self.

PASSING ON THE SKILLS Skilled rituals are typically learned by a lengthy apprenticeship. The gradual erosion of bedside rounds in teaching hospitals, replaced by rounds in a conference room where the patient's data on the computer are discussed, is detrimental to trainees. Bedside skills can only be passed on by role models who, by repeated demonstration, by observing the trainees' technique, and by giving appropriate feedback, keep this useful and essential skill alive.

PLACEBO, NOCEBO, AND THE PHYSICAL EXAMINATION The physical examination can be dehumanizing when done poorly, but when done well, it can have a salutary effect. Research on the placebo effect shows that a placebo can be something other than an inert tablet. One can have a "placebo without a placebo," meaning that the context, the ritual and its manner of execution, the setting, and the tone of voice of the examiner induce measurable change in levels of neurotransmitters and can produce a psychobiological effect. Trainees must appreciate their role in bringing about this effect. When patient expectations of the physical examination rituals are fulfilled, there is a positive effect; conversely, a clumsy and indifferent examination could have a nocebo (unpleasant or harmful) effect that impairs subsequent interactions.

■ CONCLUSION

The physical examination of the patient remains a critical element in diagnosis as well as in the ongoing assessment of the patient. The history generates hypotheses that are subsequently confirmed by useful evidence-based physical examination maneuvers, which then allow judicious ordering of further diagnostic tests. The ritual epitomizes the art and science of medicine and is itself an important means of satisfying the patient's need to feel cared for in a uniquely human and personal way. Physicians who over a lifetime cultivate skill in "reading the body" will find their practice more rewarding, will excite those apprenticing with them, and will pass on this important aspect of medicine to another generation.

ACKNOWLEDGMENT

The authors thank George Verghese, PhD, for his close reading and invaluable input.

■ FURTHER READING

Elder A et al: How valuable is physical examination of the cardiovascular system? BMJ 354:i3309, 2016.

Elder A et al: The road back to the bedside. JAMA 323:1672, 2020.

Maitra A, Verghese A: Diagnosis and the illness experience: Ways of knowing. JAMA 326:1907, 2021.

McGee S: *Evidence-Based Physical Diagnosis*, 5th ed. Philadelphia, Elsevier, 2022.

Osler W: Valedictory address to the graduates in Medicine and Surgery, McGill University. Can Med Sur J 3:433-42, 1874-75. *https://www.canadiana.ca/view/oocihm.8_05177_34/4*

Simel DL, Rennie D (eds): *The Rational Clinical Examination: Evidence-Based Clinical Diagnosis*. New York, McGraw-Hill Education, 2009. Ongoing updates at *https://jamaevidence.mhmedical.com/content.aspx?bookid=845&sectionid=61356244*.

Verghese A et al: The bedside evaluation: Ritual and reason. Ann Intern Med 155:550, 2011.

Verghese A et al: Inadequacies of physical examination as a cause of medical errors and adverse events: A collection of vignettes. Am J Med 128:1322, 2015.

Zaman J et al: The value of physical examination: A new conceptual framework. South Med J 109:754, 2016.

9

Physician Well-Being

Lisa M. Bellini, Vivek H. Murthy

PHYSICIAN WELL-BEING: AN HISTORICAL PERSPECTIVE

The practice of medicine is often viewed as a calling and as a way of life. The term *resident* originated from Dr. William Osler in 1890, when he established the first full-time, live-in medical residency training program at Johns Hopkins Hospital in Baltimore, Maryland. There, residents lived in the administration building. With the duration of their training undefined, many stayed for years. Personal sacrifices came at the cost of professional expectations and norms of the time.

The twentieth century ushered an era of standards and policies for medical education, and for the profession more broadly. In 1904, the American Medical Association (AMA) formed the Council for Medical Education (CME) to strengthen medical education training programs and policies. In 1910, the AMA supported an evaluation of the quality of these programs in the United States and Canada, which resulted in the Flexner Report, which called for the creation of a new model for medical education by recommending that medical schools enact higher admission and graduation standards and adhere strictly to the protocols of mainstream science in their teaching and research. There was no reference to physician well-being. The AMA's CME supported the report, leading to the birth of the Federation of State Medical Boards in 1912, the American Board of Medical Specialties in 1933, the Licensing Commission of Medical Education (LCME) in 1972, and the Accreditation Council of Graduate Medical Education (ACGME) in 1981. Again, physician well-being was not prioritized in causes for reform. Through this time, residents worked long hours, with little sleep or days off for illness and time with their families. Most were expected to work 36 hours every other night, often exceeding 100 hours per week.

Concerns mounted about the potential impact of fatigue on resident well-being and patient care; both were inevitably and inextricably linked. Yet, changes in policy were slow to follow. In 1975, New York City medical residents went on strike calling for fewer hours on duty. The tragic death of Libby Zion in 1984, following a misdiagnosis, was linked to medical resident fatigue and inadequate supervision. By 1989, New York became the first state to regulate resident duty hours. These regulations limited duty hours to an average of no more than 80 hours per week, with no more than 24 hours of continuous duty, and at least 24 hours free from clinical duties weekly. Subsequently, medical residents across the country petitioned the Occupational Safety and Health Administration in 2001 to nationally limit duty hours. The ACGME responded to concerns by 2003 with national standards limiting resident duty hours for all of its programs.

In 2000, the Institute of Medicine (IOM) landmark report "To Err Is Human: Building a Safer Health System" shifted the discourse away from individual-level factors toward system-level changes needed to address alarming patient safety concerns. Efforts soon after accelerated to ensure patient safety and quality of care alongside physician well-being.

Over the past 20 years, there has been an increasing focus on physician well-being with the introduction of wellness programs to promote self-care. Such programs have diffused into undergraduate medical education and biomedical training, but there are fewer standards governing physicians in independent practice. While individual-level stressors are important determinants of physician well-being, the most important drivers are the workplace systems in which students and residents learn and where physicians work.

CHALLENGES IN THE CURRENT HEALTH CARE LANDSCAPE

As health care systems have grown more complex and the demand for care has outpaced resources, chronic workplace stress and burnout have defined the contemporary experience of many physicians. Burnout is an occupational syndrome resulting from chronic workplace stress due to an imbalance between job demands and resources and other organizational, societal, and cultural factors in health care. The concept and contributing factors are included in the 11th Revision of the International Classification of Diseases (ICD-11) and the *Surgeon General's Advisory on Addressing Health Worker Well-Being* **(Fig. 9-1)**.

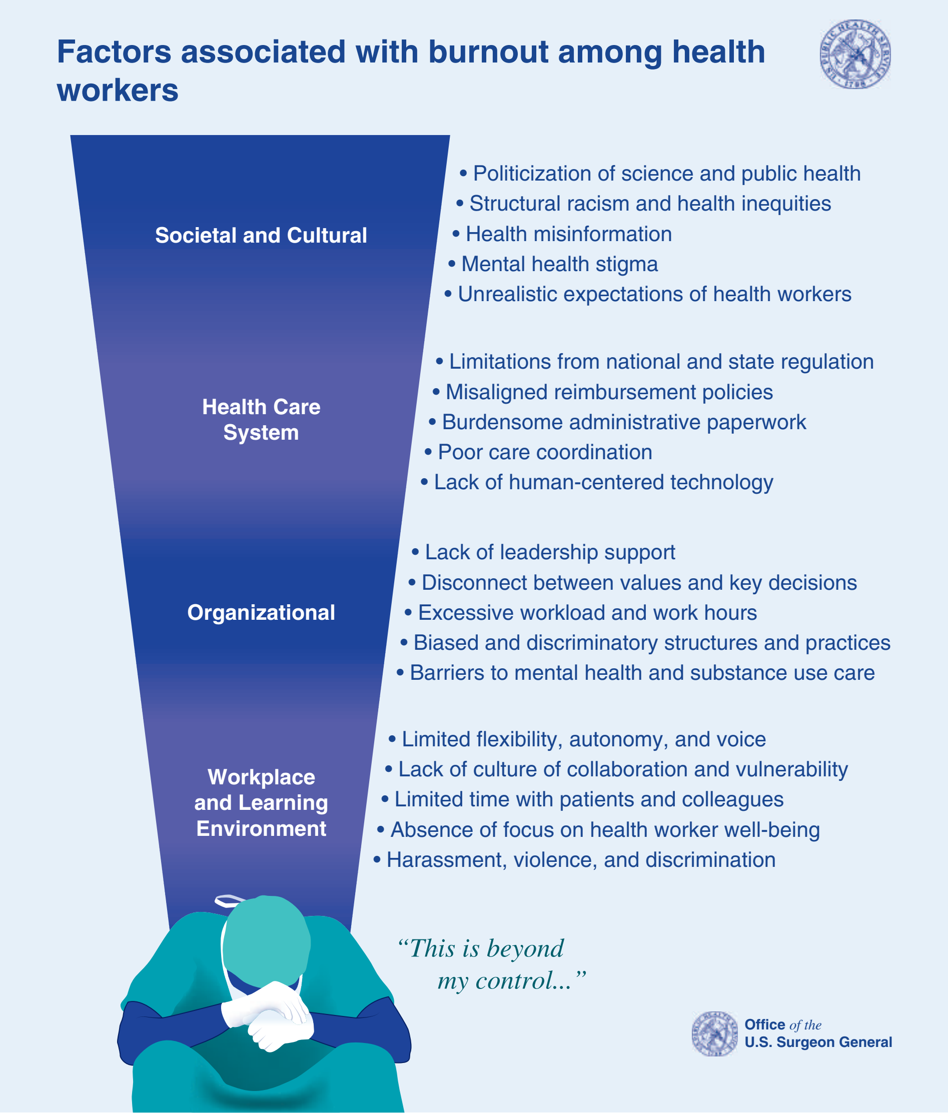
Factors associated with burnout among health workers
Societal and Cultural
• Politicization of science and public health
• Structural racism and health inequities
• Health misinformation
• Mental health stigma
• Unrealistic expectations of health workers
Health Care System
• Limitations from national and state regulation
• Misaligned reimbursement policies
• Burdensome administrative paperwork
• Poor care coordination
• Lack of human-centered technology
Organizational
• Lack of leadership support
• Disconnect between values and key decisions
• Excessive workload and work hours
• Biased and discriminatory structures and practices
• Barriers to mental health and substance use care
Workplace and Learning Environment
• Limited flexibility, autonomy, and voice
• Lack of culture of collaboration and vulnerability
• Limited time with patients and colleagues
• Absence of focus on health worker well-being
• Harassment, violence, and discrimination
"This is beyond my control..."
Office of the U.S. Surgeon General

FIGURE 9-1 Factors associated with burnout among health workers. *(Adapted from* Addressing Health Worker Burnout: The U.S. Surgeon General's Advisory on Building a Thriving Health Workforce, p. 12. *Available at https://www.hhs.gov/sites/default/files/health-worker-wellbeing-advisory.pdf.)*

It is characterized by a high degree of emotional exhaustion, depersonalization (e.g., job cynicism), and a low sense of personal achievement at work.

In 2014, the Institute for Healthcare Improvement advanced the *Quadruple Aim*, adding a fourth goal to improve health workforce well-being, directly acknowledging that the existing goals in health care—to enhance the patience experience, improve population health, and reduce costs—cannot be met without the health and well-being of our nation's healers. Subsequently, in 2017, the National Academy of Medicine (NAM) formed a *Clinician Well-Being Action Collaborative*, gathering leaders to highlight and act on the alarming burnout crisis across all health care specialties and settings. NAM declared "crisis levels" of burnout in 2019, even before the pandemic, with 35–54% of nurses and physicians and 45–60% of medical students and residents experiencing varying degrees of burnout.

The COVID-19 pandemic acutely brought the nation's attention to the unprecedented demands placed on physicians and other health workers. It also highlighted the moral distress, moral injury, and compassion fatigue experienced by many physicians and other health care providers who were forced to choose between caring for patients and keeping their families safe; who witnessed countless patients suffer and die, without family or friends at their bedside; and who had to helplessly ration inadequate supplies, beds, or treatment to their patients. One survey published in *Mayo Clinic Proceedings* found that an alarming 63% of physicians reported burnout during the pandemic in 2021, compared with 38% in 2020 and 44% in 2017.

Decades of research and evidence have found, as referenced by the *Surgeon General's Advisory on Addressing Health Worker Burnout* in 2022, that chronic work stress and burnout have harmful effects on physician health and well-being. Chronic work stress is associated with poor health outcomes such as impaired cognitive function, increased risk of cardiovascular disease, type 2 diabetes, fertility issues, sleep disruptions, isolation, relationship conflict, and risk for substance use and misuse. Burnout is also associated with mental health challenges including anxiety, depression, and suicidal ideation. Surveys by health care professional associations have also found widening gender and racial gaps, especially among female physicians and groups underrepresented in medicine. Physicians and other health workers who report burnout are more likely to reduce working hours, report intent to leave jobs or medical school, or leave medicine altogether. One in every five physicians has reported intent to leave medical practice due to burnout, contributing to a projected shortfall of up to 86,000 physicians in the United States by 2036. These shortages will only compound the vicious cycle of increasing physician work demands that often lead to burnout.

CORE ELEMENTS FOR BUILDING A THRIVING PHYSICIAN WORKFORCE

A thriving physician workforce requires a dynamic, multipronged, and collective approach to solve the complex array of institutional, structural, cultural, and societal factors that impact physician health and well-being (Figs. 9-1 **and 9-2**). Fundamental change in the organizational environment, including the systems and cultures where all physicians learn, train, and work, is the necessary first step.

PROTECTION FROM HARM

Strengthen physician protections from physical and psychological harms in the clinical learning and work environment. This includes ensuring adequate personal protective equipment during and outside of public health emergencies, sufficient staffing, shift coverage and rest, and clearly communicated policies that protect physicians from all threats and acts of harassment, intimidation, and violence at work and in their communities. The Joint Commission released workplace violence prevention standards in 2021 to guide leaders with implementation. Health care organizations must also explicitly support inclusion and equitable access to policies and programs (e.g., paid leave, career advancement) that can comprehensively address diversity and accessibility. Doing so can address the microaggressions, implicit bias, discrimination, and racism that many physicians face.

Increase access to quality mental health care for all physicians, including residents and students. These services can be offered on-site or via telehealth options including peer support groups, confidential physician health programs, or employee assistance program (EAP) services (e.g., counseling, referrals, caregiver support). It is most important to ensure that access to any service is convenient, meeting unique needs and work schedules.

In the 2023 Physicians Foundation Survey of America's Current and Future Physicians, more than half of physicians (up since 2021) reported knowing of a physician who has ever considered, attempted, or died by suicide. A proactive, evidence-based approach to suicide prevention must be incorporated to embed voluntary, anonymous screening and tailored referrals and follow-up care.

Furthermore, we must end barriers to mental health care, specifically stigma and policies that deter physicians and other health workers from mental health support services. Many physicians may be reluctant to seek formal short-term and recovery care for mental health challenges and conditions, given concerns regarding potential repercussions on their license, hospital credentials, careers, and credibility. Health care organizations, academic institutions, and policymakers at all levels can review and remove intrusive, stigmatizing questions on all applications and forms, ensuring alignment with national recommendations set by The Joint Commission in 2020.

REDUCE ADMINISTRATIVE BURDENS

Administrative burdens must be reduced to give physicians more time for what matters—their patients. A rapidly changing health care ecosystem, with new payment options, market consolidation, insurance and regulatory requirements, and advances in health information technology, has contributed to the loss of physician autonomy and a significant reduction in time for patient care. For example, with the growing consolidation of the health care market and the corporate privatization of practices, many physicians feel their voices, decision-making, and sense of value have diminished over time. Systems-level changes are needed to ensure that the physician–patient relationship is at the very center of the health care system. Further, there are additional organization-level practice changes that can also address inefficiencies and give physicians time back.

Improve and streamline workflow processes (e.g., documentation requirements, inbox notifications, prior authorizations). As an example, changes to documentation guidelines for outpatient evaluation and management (E/M) visit codes were updated in 2021 to better align with current medical practice and patient care. Health professional associations offer practical tools, including the AMA *Saving Time Playbook* and *Electronic Health Records (EHR) Playbook*. Researchers have found notable progress across health systems when leaders visibly apply the "Getting Rid of Stupid Stuff," or GROSS, model, seeing measurable reductions in the volume of unnecessary daily documentation tasks for physicians and the care team.

Eliminate unnecessary and inefficient prior authorizations. In 2022, a majority of physicians surveyed (94%) reported that prior authorizations from insurers delayed access to necessary care. Many physicians also reported treatment abandonment by their patients because of prior authorization delays. Many policymakers and health professional association programs exempt clinicians from prior authorization requirements if they meet specific performance measures. Although evidence is mixed on the benefits of these efforts, it seems likely that any measurable reduction in administrative burden will improve costs and, if thoughtfully implemented, improve quality of care.

With workforce shortages, organizations can harness, adopt, and implement technology in a physician- and patient-centered way. While the role of the physician will always be indispensable, particularly for direct patient interactions and complex clinical decision-making, there are growing opportunities to leverage trustworthy artificial intelligence (AI), machine learning, and other digital automation tools to support various stages of patient care, such as previsit planning, encounter documentation, prior authorizations, and follow-up communication. Importantly, AI companies have the responsibility to also ensure that such platforms safeguard patient safety and privacy, mitigate risk of

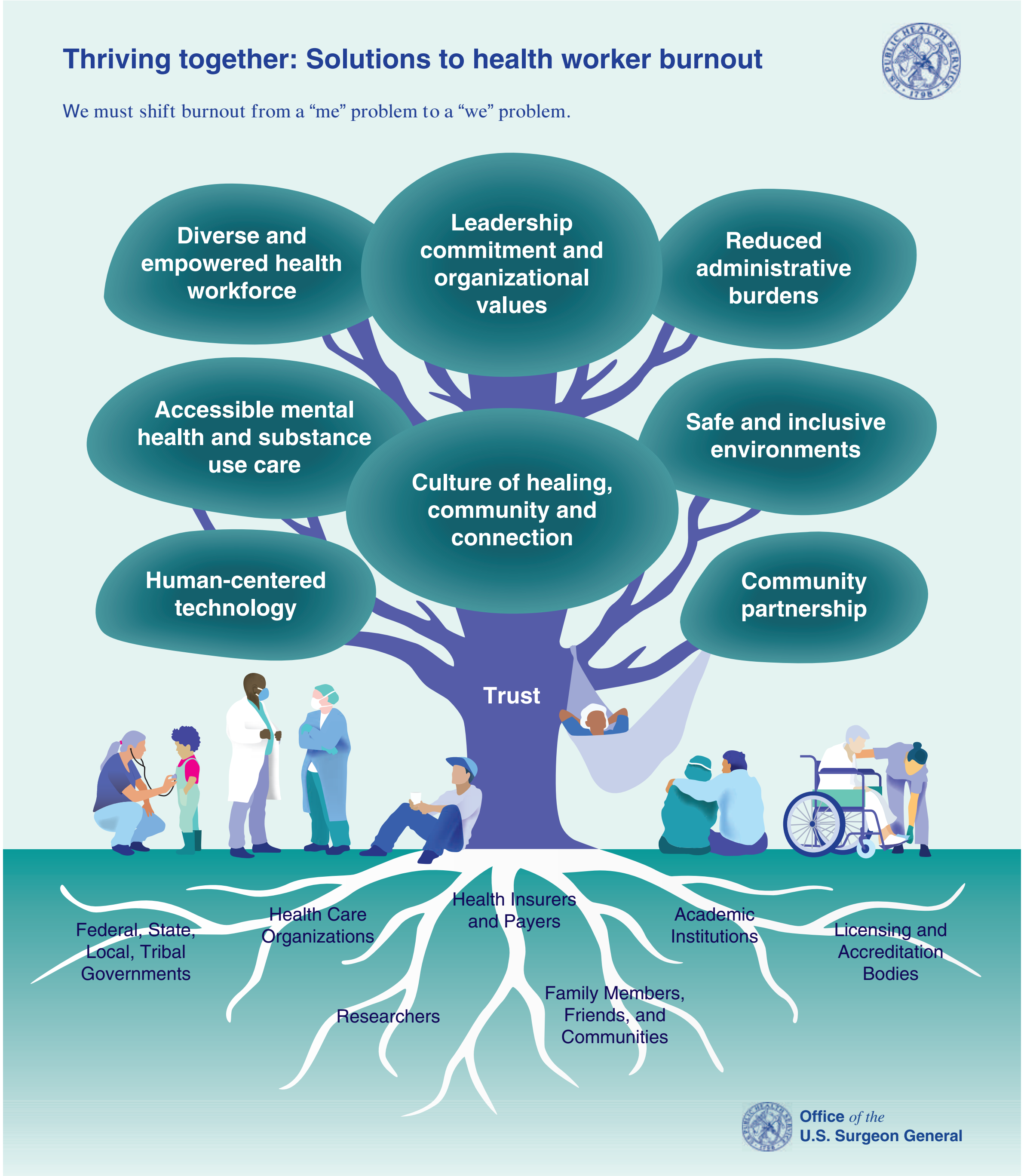
Thriving together: Solutions to health worker burnout
We must shift burnout from a "me" problem to a "we" problem.
Diverse and empowered health workforce
Leadership commitment and organizational values
Reduced administrative burdens
Accessible mental health and substance use care
Safe and inclusive environments
Culture of healing, community and connection
Human-centered technology
Community partnership
Trust
Federal, State, Local, Tribal Governments
Health Care Organizations
Health Insurers and Payers
Academic Institutions
Licensing and Accreditation Bodies
Researchers
Family Members, Friends, and Communities
Office of the U.S. Surgeon General

FIGURE 9-2 Thriving together: solutions to health worker burnout. *(Adapted from* Addressing Health Worker Burnout: The U.S. Surgeon General's Advisory on Building a Thriving Health Workforce, p. 20. *Available at https://www.hhs.gov/sites/default/files/health-worker-wellbeing-advisory.pdf.)*

clinical harm, and measurably improve the quality of care. Further, EHR companies must strengthen platforms to focus on key features needed for patient care and design the systems to meet the needs of clinicians in ways that optimize usability as a clinical decision support tool. Finally, health care technology companies must work with physicians and other health workers to ensure accessible, meaningful, and unbiased processes and data. This includes assurance that interventions do not add more burden onto the health care team.

BUILD CONNECTION AND COMMUNITY

Opportunities for strengthening social connection and community in medical education, clinical training programs, and all health care work environments should be integrated into our systems. Physicians, including students and residents, should be included in these social networks, as appropriate. They may have fewer routine opportunities to connect meaningfully with colleagues and identify new colleagues and mentors, particularly when new to the community. The value of workplace connection and social support for physicians cannot be overstated, especially in an era of widespread societal loneliness and social isolation. Acknowledging that the common experience of loneliness and isolation can be felt at all stages of medical education and throughout a career in medicine is a necessary and important first step.

Peers, faculty, and health care organization and academic institution leaders are each well-positioned to model and foster opportunities for social connection and community. This can be through applying teamwork and team-based care models, establishing peer support groups and informal learning networks, reviving the modern-day *doctors' lounge*, and investing in mentoring and coaching programs.

These efforts to cultivate social connection must be operationalized with protected time and infrastructure built into core working hours and the physical environment. This can help all physicians feel seen and heard, with time to pause and reflect on challenging circumstances and morally distressing dilemmas together with peers and mentors. These efforts also help physicians build trusting relationships with one another and promote a culture that values connection within the workplace. Finally, such efforts also offer necessary time and space to reinforce connection through shared purpose, professional fulfilment, and a celebration of collective achievements in health care, research, and medicine.

BOLSTER THE HEALTH WORKFORCE

Investments that bolster the health workforce need to be expanded. Health workforce shortages across all specialties and settings can negatively impact timely access, quality, and patient safety. As noted earlier, the American Academy of Medical Colleges has estimated a major shortage of physicians in the United States, and the challenge is much greater in most parts of the world. Health care needs will only increase as we address the ongoing mental health crises and a growing and aging population and strengthen our responses to climate change and its impact on health.

Despite these gaps, there are hopeful signs and solutions. According to data compiled by the Federation for State Medical Boards, the U.S. physician workforce is 20% larger than it was over a decade ago. New medical schools and postgraduate training programs are being developed at an unprecedented rate in most industrialized countries and in many less developed parts of the world. There are renewed efforts to increase and diversify the workforce, especially to ensure more culturally appropriate care. Many of these efforts encourage practice in specialties with shortages and in rural and underserved communities. We can also ensure support for equitable pathways via scholarships and tuition support. Additional career advancement, faculty salary support, and apprenticeship training programs can further retain and sustain instructors, preceptors, and mentors in the nursing, direct patient care, and behavioral health workforce.

We must also sustain and increase investments for addressing unmet underlying social needs, such as housing, food, and transportation. Doing so can not only reduce demands upstream for health care but can also address the moral distress physicians may experience when they encounter root causes of their patients' suffering and the obstacles to quality care that they cannot address. An annual survey by the Physicians Foundation in 2022 found that many (61%) felt they had little to no time to effectively address their patients' social determinants of health. Social determinants of health are estimated to account for upward of 80–90% of modifiable factors to health outcomes **(Chap. 11)**. Health systems, together with community leaders in public health, are well-poised to focus on prevention and whole-person care, investing in evidence-informed models that keep individuals and communities healthy for the long term.

FOSTER A CULTURE OF PHYSICIAN WELL-BEING

The culture of medicine must be transformed to center on physician health and well-being. For medical education and training programs, institutions can begin by addressing the *hidden curriculum*—the unwritten and unofficial values, and unintended lessons, that students learn by observing a teacher's actions, which may be at odds with the formalized curriculum. The American College of Physicians (ACP) reported that more than half of medical students experienced disconnect between what they were explicitly taught and what they perceived from faculty behaviors in practice. The ACP 2018 position paper is a helpful resource for physician leaders and faculty with recommendations and strategies for fostering values of respect, honesty, empathy, inquiry, and ethics, while promoting clinician wellness.

Across any organization, leadership is critical in fostering a culture that values physician health and well-being. Concrete steps include operationalizing well-being as an organizational value; integrating it into strategic plans, performance indicators, and training; and establishing a chief well-being officer role. This position must have dedicated resources and decision-making authority. The role should collaborate with leadership in health care administration, human resources and talent management, finances, health information technology (IT), and equity, while proactively engaging physician and other health worker representatives. They must regularly assess the work environment for factors contributing to chronic work stress and burnout while also evaluating changes in policies and programs that impact the role, function, and well-being of physicians. NAM's Action Collaborative offers a compendium of resources, with validated tools for engaging staff, measuring physician well-being, and benchmarking success.

All health care leaders must model and create environments where conversations about physical and mental health are normalized. Physicians should be able to talk openly about work stressors and their mental health without stigma or fear of repercussions on their licensing, credentialing, or careers. This includes encouraging open and honest conversations about the mental health challenges they face in their day-to-day work and offering support and validation along their journey. This must begin early in medical training. Periodic physician well-being support and awareness campaigns can help physicians feel heard, supported, and valued and serve as useful tools for communicating information and updates about well-being services. The National Institute for Occupational Safety and Health, the NAM Action Collaborative, and numerous professional associations offer practical tools to support leaders with this type of communication and programming to support physician well-being.

THE FUTURE OF PHYSICIAN WELL-BEING

Medicine remains a calling, yet one that should be built on, not at the expense of, the well-being of our nation's healers. Failing to value and center their health and well-being at the core of our health care system puts us all at risk. When physicians look ahead, they should see a future where their dedication is not taken for granted, but one where their health, safety, and well-being are as much a priority as the well-being of the people and communities in their care.

FURTHER READING

The Complexities of Physician Supply and Demand: Projections From 2021 to 2036. *https://www.aamc.org/media/75236/download?attachment*. Accessed April 29, 2024.

Kohn L et al (eds): *To Err Is Human: Building a Safer Health System*. Institute of Medicine (U.S.) Committee on Quality of Health Care in America, Washington, DC, National Academies Press, 2000. *https://pubmed.ncbi.nlm.nih.gov/25077248/*.

Lehmann LS et al: Hidden curricula, ethics, and professionalism: Optimizing clinical learning environments in becoming and being a physician. A position paper of the American College of Physicians, 2018. Ann Intern Med 168:506, 2018.

Linzer M et al: Trends in clinician burnout with associated mitigating and aggravating factors during the COVID-19 pandemic. JAMA Health Forum 3:11, 2022.

Lyubarova R et al: Gender differences in physician burnout: Driving factors and potential solutions. Perm J 27:2, 2023.

National Academy of Medicine: *National Plan for Health Workforce Well-Being*. Washington, DC: The National Academies Press. *https://doi.org/10.17226/26744*. Accessed July 28, 2023.

Philibert I, Taradejna C: Duty hour standards. Chapter 2: A brief history of duty hours and resident education, in *Enhancing Quality of Care, Supervision, and Resident Professional Development*. ACGME, 2011. *https://www.acgme.org/globalassets/pdfs/jgme-11-00-5-111.pdf*.

Shanafelt T et al: Physician well-being 2.0: Where are we and where are we going? Mayo Clin Proc 96:10, 2021.

US Department of Health and Human Services: Addressing Health Worker Burnout: The U.S. Surgeon General's Advisory on Building a Thriving Health Workforce. *https://www.hhs.gov/sites/default/files/health-worker-wellbeing-advisory.pdf.* Accessed July 28, 2023.

WEBSITES

American Medical Association: *www.ama-assn.org/*
Physicians Foundation: *www.physiciansfoundation.org/*

10

Diagnosis: Reducing Errors and Improving Quality

Gordon Schiff

Diagnosing patients' illnesses is the essence of medicine. Patients present to doctors seeking an answer to the question, "What is wrong with me?" Ideally, no clinician would want to treat a patient without knowing the diagnosis or, worse yet, erroneously treat a misdiagnosed illness. From the earliest moments of medical school, the defining quest toward becoming a knowledgeable and proficient physician is learning how to put a diagnostic label on patients' symptoms and physical findings, and clinicians pride themselves on being "good diagnosticians." Yet the centuries-old paradigm of mastering a long list of diseases, understanding their pathophysiology, and knowing the cardinal ways they manifest themselves in signs and symptoms, while still of fundamental importance, is being challenged by new insights illuminated by the glaring spotlight of diagnostic errors. Basic internal medicine diseases, such as asthma, pulmonary embolism, congestive heart failure, seizures, strokes, ruptured aneurysms, depression, and cancer, are misdiagnosed at shockingly high rates, often with 20–50% of patients either being mislabeled as having these conditions (false-positive diagnoses) or having their diagnosis missed or delayed (false negatives). How and why do physicians so often get it wrong, and what can we do to both diagnose and treat the problem of delayed diagnosis or misdiagnosis?

Diagnosis is both an ancient art and a modern science. The current science of diagnosis, however, goes far beyond what typically comes to clinicians' and patients' minds when they conjure up images of state-of-the-art molecular, genetic, or imaging technologies. Improvements in diagnosis are just as likely to come from other areas, many with origins outside of medicine, as they are from advanced diagnostic testing modalities. These diverse sciences that the field of diagnostic safety has, and must, draw from include systems and human factors engineering, reliability science, cognitive psychology, decision sciences, forensic science, clinical epidemiology, health services research, decision analysis, network medicine, learning health systems theory, medical sociology, team dynamics and communication, risk assessment and communication, information and knowledge management, and health information technology, especially artificial intelligence and clinical decision support. A clinician reading this chapter is likely to find this list of overlapping and intersecting domains quite daunting. However, rather than feeling overwhelmed, we urge readers to view them as the basic science supports that will ultimately make their lives easier and diagnosis more accurate and timely. Rather than feeling intimidated, clinicians should feel a sense of relief and assurance in understanding that good diagnosis does not rest entirely on their shoulders. Instead, it is a systems property, where an infrastructure and a team, one that especially includes the patient, can in a coordinated way work together to achieve more reliable and optimal diagnosis.

EMERGENCE OF DIAGNOSIS ERROR AS AN IMPORTANT PATIENT SAFETY ISSUE

Over the past two decades, a series of studies culminating in a landmark report from the U.S. National Academy of Medicine (NAM), *Improving Diagnosis in Health Care,* have shone a spotlight on diagnostic errors. Reports from patient surveys, malpractice claims, and safety organizations have found that diagnostic errors are the leading type of medical error. Although errors in diagnosis can be defined in various ways, the NAM Committee defined diagnostic error as "the failure to (a) establish an accurate and timely explanation of the patient's health problem(s) or (b) communicate that explanation to the patient."*

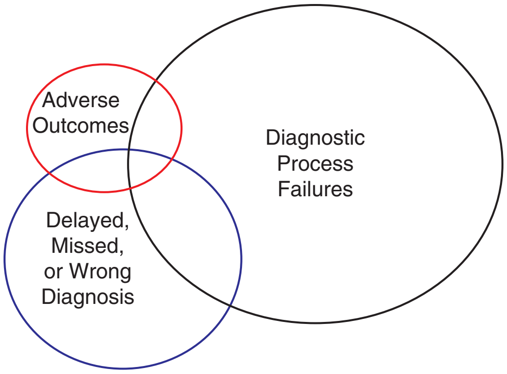
Adverse Outcomes
Diagnostic Process Failures
Delayed, Missed, or Wrong Diagnosis

FIGURE 10-1 What is a diagnosis error? *(Adapted from GD Schiff et al: Diagnosing diagnosis errors: Lessons from a multi-institutional collaborative project, in Advances in Patient Safety: from Research to Implementation. Vol. 2 Concepts and Methodology, Rockville, MD, 2005, pp. 255-278, and GD Schiff, L Leape: Acad Med 87:135, 2012.)*

One way to visualize diagnostic errors is through a Venn diagram **(Fig. 10-1)**, which illustrates the fact that many things can go wrong in the diagnostic process (e.g., failure to ask an important history question, physical examination sign overlooked, laboratory specimen erroneously switched between two patients, x-ray not followed up), but this usually does not result in a wrong diagnosis or patient harm. Similarly, a patient can be misdiagnosed but unharmed, without any identifiable error in the care received. Our greatest concern is where these three circles intersect, with conservative estimates suggesting that 40,000–80,000 patients die each year in U.S. hospitals alone from diagnostic errors. The NAM report outlined eight recommendations that are the foundation for this chapter **(Table 10-1)**.

NEW WAYS TO THINK ABOUT DIAGNOSIS AND DIAGNOSTIC ERRORS

Medical textbooks have historically given attention to "clinician reasoning" and associated cognitive heuristics and biases. Errors in clinical reasoning can be summarized in three broad groups: (1) hasty judgments, (2) biased judgments, and (3) inaccurate probability estimates. Research from cognitive psychology has identified scores of common mental shortcuts or "heuristics" humans are prone to use in everyday life, many of which are useful for efficient diagnosis but can also lead to biases and errors. **Table 10-2** lists some of the common cognitive biases that can lead diagnosis astray (this topic is discussed further in **Chap. 4**).

TABLE 10-1 National Academy of Medicine Recommendations for Improving Diagnosis in Health Care

1. Facilitate more effective teamwork in the diagnostic process among health care professionals, patients, and their families.
2. Enhance professional education and training in the diagnostic process in areas such as clinical reasoning; teamwork; communication with patients, families, and other health care professionals; and appropriate use of diagnostic tests.
3. Ensure that health information technologies support patients and health care professionals in the diagnostic process.
4. Develop and deploy approaches to identify, learn from, and reduce diagnostic errors and near misses in clinical practice including providing systematic feedback on diagnostic performance.
5. Establish a work system and culture that supports the diagnostic process and improvements in diagnostic performance.
6. Develop a reporting environment and medical liability system that facilitates improved diagnosis by learning from diagnostic errors and near misses.
7. Design a payment and care delivery environment that supports the diagnostic process.
8. Provide dedicated funding for research on the diagnostic process and diagnostic errors.

Source: National Academies of Sciences, Engineering, and Medicine. 2015. Improving Diagnosis in Health Care. *https://doi.org/10.17226/21794.* Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.

**Source:* National Academies of Sciences, Engineering, and Medicine. 2015. Improving Diagnosis in Health Care. *https://doi.org/10.17226/21794*. Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.

TABLE 10-2 Selected Cognitive Biases Contributing to Diagnostic Errors

1. Premature closure: accepting a diagnosis before it has been fully verified
2. Anchoring: tendency to fixate on a specific symptom or piece of information early in the diagnostic process with subsequent failure to appropriately adjust
3. Confirmation bias: tendency to look for confirming evidence to support one's diagnostic hypothesis, rather than disconfirming evidence to refute it
4. Search satisficing: tendency to call off a search, satisfied once a piece of data or presumed explanation is found, and not considering/searching for additional findings or diagnoses
5. Availability bias: tendency to give too much weight to diagnoses that come more readily to mind (e.g., recent dramatic case)
6. Base-rate neglect: failing to adequately take into account prevalence of a particular disease (e.g., erroneously interpreting a positive test as indicating disease in a low-prevalence population using a test with 5% false-positive rate)
7. Knowledge deficit (on part of provider, with accompanying lack of awareness)
8. Framing bias: judgement overly influenced by the way the problem was presented (how it was framed in words, settings, or situations)
9. Social/demographic/stereotype bias: biases from personal or cultural beliefs about women, historically marginalized populations such as African Americans, people with differing sexual identities, or other patient groups for whom prejudices may distort diagnostic assessment

However, clinicians will also benefit from having a better understanding of diagnosis as a "system" rather than just what takes place in clinicians' minds. Classic teaching exhorting trainees and practicing physicians to have a broad differential and "high index of suspicion" for various diseases is challenged not only by these unconscious biases but also by limitations of human memory, information shortfalls, constrained encounter time, system process failures, and the myriad nonspecific symptoms that patients bring to clinicians. Many symptoms are self-limited, defy a precise diagnosis or etiology, and do not portend harmful outcomes. Insights from safety and cognitive sciences call for rethinking traditional approaches to diagnosis and suggest new approaches to overcome current limitations **(Table 10-3)**.

UNCERTAINTY IN DIAGNOSIS

Given variations in ways patients present, illnesses evolve, and tests perform, it is often not possible or practical to "make" a definitive diagnosis, particularly in the primary care setting early in the course of a patient's illness. Clinicians need to harness these uncertainties to both have enhanced situational awareness of where things can go wrong and create safety nets to protect patients against harms from delayed diagnosis and misdiagnosis. Terms such as *preliminary diagnosis*, *working diagnosis*, *differential diagnosis*, *deferred diagnosis*, *undiagnosed illness*, *diagnoses with uncertain or multifactorial etiologies*, *intermittent diagnoses*, *multiple/dual diagnoses*, *self-diagnosis*, or at times *contested diagnosis* need to be part of our vocabulary, thinking, and communications with patients to convey that diagnosis is often imprecise. Anxious patients worried about a condition, for example, cancer, COVID-19 infection, or a diagnosis to which a relative or a friend has recently succumbed, come seeking reassurance and may not welcome an uncertain answer. Thus, we have to work with patients, listen to and respect their concerns, and take their symptoms seriously yet modestly acknowledge our limitations. We need to tailor this approach to patients' differing levels of health literacy, trust in our clinical advice, and experiences with the health system.

DON'T MISS DIAGNOSES AND RED FLAGS

Uncertainty should not be a license for complacency. Particularly for diseases that (1) progress rapidly, (2) require specific treatments that depend on making the correct diagnosis, or (3) have public health or contagion implications, clinicians need to be poised, and systems designed, to consider and, where appropriate, pursue critical "don't miss" diagnoses. While clinicians are generally aware of more common "don't miss" diagnoses (e.g., acute myocardial infarction, sepsis), **Table 10-4** illustrates examples of less common diagnoses that warrant similar consideration. Throughout this textbook, readers should orient themselves to recognize such critical diagnoses and think about presentations and syndromes where they may be lurking.

An important related concept is so-called "red flags" or "alarm symptoms." This construct has its origins in guidelines for back pain but has increasingly been applied to many other problems, such as headache, red eye, swollen joint, or even abdominal pain and chest

TABLE 10-3 New Models for Conceptualizing Diagnosis and Diagnosis Improvement

TRADITIONAL WAYS OF THINKING ABOUT DIAGNOSIS AND DIAGNOSTIC ERROR	NEW PARADIGMS/BETTER WAYS TO THINK ABOUT DIAGNOSIS AND IMPROVING DIAGNOSIS
General	
A good diagnostician gets it right the first time, almost all of the time	Diagnosis is an inexact science with inherent uncertainties Goal is to minimize errors and delays via more reliable systems and follow-up
Lore of masterful/skillful academic expert diagnostician who knows/recalls everything; need to look to them if seeking diagnostic excellence	Less reliance on (fallible) human memory Quality diagnosis is based on well-coordinated distributed network/team of people and reliable processes All patients entitled to receive quality diagnosis, regardless of where and from whom they receive care
Diagnosis is the doctor's job	Co-production of diagnosis among clinicians (including lab, radiology, specialists, nurses, social workers) and, especially, the patient and family
Patients often viewed as overly anxious, exaggerating, time-consuming, questioning, with sometimes unreasonable demands and expectations	Patients are key allies in diagnosis; hold key information Need to address understandable/legitimate fears, desires for explanations Leveraging patient questions and questioning of diagnosis to stimulate rethinking the diagnosis where needed
Diagnosis and treatment as separate stages in patient care (i.e., make a diagnosis, then treat)	Prioritizing diagnostic efforts to target treatable conditions More integrated strategies and timing for testing and treatment depending on urgency for treatment
Clinical practices	
Order lots of tests to avoid missing diagnoses	Judicious ordering: targeted, well-organized data and testing Appreciation of test limitations (false positives or negatives, incidental findings, overdiagnosis, test risks) and resulting harms
More referrals to avoid missing rarer/specialized diagnoses; concomitant utilization barriers (copays, prior authorization) to minimize overuse	"Pull systems" to lower barriers and make it easier to pose questions, obtain real-time virtual consults Co-management approaches to enable collaborative watch-and-wait conservative strategies where appropriate
Frequent empirical drug trials when uncertain of diagnosis	Conservative use of drugs to avoid confusing clinical picture or labeling patients with diseases they may not have
Physician attention/efforts to ensure disease screening	Automating, delegating clerical functions; teamwork to free up physician cognitive time

(*Continued*)

TABLE 10-3 New Models for Conceptualizing Diagnosis and Diagnosis Improvement (Continued)

TRADITIONAL WAYS OF THINKING ABOUT DIAGNOSIS AND DIAGNOSTIC ERROR	NEW PARADIGMS/BETTER WAYS TO THINK ABOUT DIAGNOSIS AND IMPROVING DIAGNOSIS
Diagnosis errors and challenges	
Diagnostic error viewed as a personal failing Errors classified as either "system" or "cognitive"	Many errors/delays rooted in processes and system design/failures Errors multifactorial with interwoven, interacting, and inseparable cognitive and system factors
Errors are infrequent; hit-and-miss ways to learn about errors	Errors are common; systematic proactive follow-up is needed to recognize potential for errors Surveilling of high-risk situations and one's own diagnostic performance and outcomes
Clinicians' reactions: denial, defensive, others to blame, pointing to others also making similar errors	Culture of actively and nondefensively seeking to uncover, dig deep to learn from, and share errors and lessons
Dreading complex, frustrating diagnostic dilemmas	Welcoming/enjoying intellectual/professional challenges Adequate support (time, help, consultations) for more complex patients
Diagnoses as distinct labels, events	Diagnoses can be indistinct, interacting comorbidities, socially constructed, multifactorial, evolving over time, or have overlapping genotype-phenotype expressions
Documentation/communication	
Viewed as time-consuming, mindless, primarily to document for billing code and/or bulwark against malpractice claims	Documentation as useful tool for reflecting, crafting, sharing assessments, differential diagnosis, reflecting about unanswered questions Opportunities for decision support interacting with computer Notes open for patients to read to help understand and critique diagnosis
Say and write as little as possible about uncertainties, lest it be used against you in malpractice allegation	Share uncertainties to maximize communication and engagement with other caregivers, patients
Don't let patient know about errors so they don't become angry, mistrustful, or sue	Patients have right to honest disclosure; often find out about errors anyway (e.g., cancer evolves); anticipate, engage their concerns
Patients advised to call if not better; no news is good news (test results: "We'll call if anything is abnormal.")	Systematic proactive follow-up to close loop on all tests and any worrisome symptoms, to check how patient is doing, monitor outcomes
Global remedies	
Knowing/memorizing more medical knowledge	Knowing more about the patient (including psychosocial, past history, environmental contexts)
Attention to the "objective" data (physical exam, tests) to reliably make diagnoses	Renewed emphasis on history, history-taking, listening Acknowledgment of ubiquitous subjective cognitive biases; efforts to anticipate, recognize, counteract
Exhortations to have "high index of suspicion" of various diagnoses	Less reliance on memory recall of lectures/reading; more just-in-time info look-up Affordances, alerts to red flags engineered into workflow Delineation of "don't miss" diagnoses with design of context-relevant decision support reminders
Ensuring physician is copied on everything, thorough/voluminous notes, widespread reminders/alerts	Biggest problem no longer lack of access to information, but rather information overload; strategies to organize, minimize
Continuing medical education (CME) courses to expand medical knowledge	Real-time, context-aware reminders of pitfalls, critical differential diagnoses, and key differentiating features. Ready access to medical references, second opinions
Redundancies, double-checks	Recognition that single, highly reliable systems are often better than multiple halfway solutions Clear delineation of responsibilities for follow-up tasks
Fear of malpractice suits to motivate physicians to be more careful and practice defensive medicine	Drive out fear, make it safe to learn from and share errors Shared situational awareness of where pitfalls lurk
More accountability, financial incentives, and penalties tied to performance metrics	Clinician engagement in improvement based on trust, collaboration, professionalism, financial neutrality Metric modesty, recognizing many best practices yet to be defined/proven
More rules, requirements; target outlier physicians for better compliance	Standardization with flexibility; learning from deviations
More time with patients	Better time spent with patients: offloading distractions, more efficient history collection/organization, longitudinal continuity, and, where needed, additional time to talk/think/explain during, before, or after visits Easier access for patients to reach or be seen by clinicians when experiencing concerning symptoms
Reflex changes in response to errors	Avoiding "tampering," which entails understanding/diagnosing difference between "special cause" versus "common cause" (random) variation

Source: Modified from GD Schiff: Quality and Safety in Health Care 2013.

pain. Examples of widely cited red flags for back pain that should trigger consideration of more serious etiologies include fever, weight loss, history of malignancy or intravenous drug use, or neurologic signs and symptoms. In theory, many presenting syndromes could benefit from identification of such clues to more serious diagnoses. Evidence-based medicine calls for better data on the sensitivity, specificity, yield, and discriminatory ability of various clinical "red flag" clues; yet few have been rigorously evaluated. Nonetheless, clinicians find them useful as simple ways to reassure themselves and their patients that a common symptom such as back pain or headache is, or is not, likely an indicator of more urgent or serious pathology.

Interwoven with the challenges of not missing critical diagnoses is the problem of overtesting and overdiagnosis—performing unnecessary and even potentially harmful tests whose benefit does not justify the risks or costs or that may lead to diagnoses that would have never caused any symptoms or problems. Diagnosticians need to weigh carefully this "other side of the coin" of missed diagnosis to avoid such harms and expenses. Thus, being more conservative in diagnostic testing should not be primarily about conserving resources, but more an approach for ensuring laboratory or imaging studies truly benefit patients, while minimizing short- and longer-term harms.

DIAGNOSTIC PITFALLS

One of the important ways of learning in medicine is learning from the missteps of those who have walked the path ahead of us. By learning about commonly missed diagnoses and the ways accurate, timely

TABLE 10-4 Examples of "Don't Miss" Diagnoses

INFECTIONS/ INFLAMMATION	CARDIAC/ISCHEMIC/ BLEEDING	METABOLIC/ HEMATOLOGIC/ ENVIRONMENTAL
Spinal epidural abscess	Aortic dissection Leaking/ruptured abdominal aortic aneurysm	Diabetes ketoacidosis Hyperosmolar hyperglycemia
Necrotizing fasciitis	Pericardial tamponade	Myxedema/ thyrotoxicosis
Meningitis	Wolff-Parkinson-White Prolonged QT	Addison's disease
Endocarditis	Pulmonary embolism	B$_{12}$ deficiency anemia
Peritonsillar abscess	Tension pneumothorax	von Willebrand's disease
Tuberculosis-active pulmonary, other	Acute mesenteric ischemia Sigmoid volvulus	Hemochromatosis
COVID-19 infection	Esophageal, bowel perforation	Celiac sprue
Guillain-Barré syndrome	Cerebellar hemorrhage	Carbon monoxide, lead, pesticide poisoning
Ebola infection	Spinal cord compression	Food poisoning
Temporal arteritis	Testicular, ovarian torsion	Malignant hyperthermia
Rhabdomyolysis	Ectopic pregnancy	Alcohol, benzodiazepine, barbiturate withdrawal
Angioedema	Retroperitoneal hemorrhage	Tumor lysis syndrome Hypo-/hypercalcemia

diagnosis went astray, we can avoid making similar mistakes. Anticipating the potential for similar types of errors can both create situational awareness of traps to avoid and contribute to learning from our own personal and collective patterns of mistakes. Several studies have examined common or recurring pitfalls in diagnosis. An example of a common disease-specific diagnostic pitfall in breast cancer diagnosis is ordering a mammogram for a woman with a palpable breast lump and, when the mammogram returns as normal, reassuring her that cancer has been "ruled out" by the negative test. Any unexplained mass or lesion palpable on physical examination needs assessment including further testing and biopsy, where warranted. Diagnostic pitfalls can be classified into a number of generic scenarios (Table 10-5). We now have large databases that have the potential to track "diagnoses outcomes"—i.e., whether a new diagnosis emerges that suggests an initial diagnosis was incorrect or a diagnosis of a patient's symptoms was suboptimally delayed. This should, in the future, allow us to more rigorously focus on these cases, to identify contributing factors and recurring patterns, and to help point the way for systemwide improvement strategies.

TABLE 10-5 Generic Types of Diagnostic Pitfalls

PITFALL	EXAMPLES
Disease A mistaken for disease B Diseases often mistaken/misdiagnosed with each other	• Aortic dissection misdiagnosed as acute myocardial infarction • Bipolar disorder misdiagnosed as depression
Misinterpretation of test result(s) False-positive or false-negative results with failure to recognize test limitations	• Breast lump dismissed after negative mammogram • Negative COVID-19 test early or late in course
Failure to recognize atypical presentation, signs, and symptoms	• Apathetic hyperthyroidism • Sepsis in elderly patient who is afebrile or hypothermic
Failure to assess appropriately the urgency of diagnosis Urgency of the clinical situation was not appreciated and/or delays critical diagnoses	• Compartment syndrome • Pericardial tamponade • Tension pneumothorax
Perils of intermittent symptoms or misleading evolution Intermittent symptoms dismissed due to normal findings (exam, lab, electrocardiogram) when initially seen	• "Lucid interval" in traumatic epidural hematoma • Paroxysmal arrhythmias • Intermittent hydrocephalus (Bruns' syndrome)
Confusion arising from response/ masking by empiric treatment	• Empiric treatment with steroids, proton pump inhibitors, antibiotics, pain medication erroneously masking serious diagnosis
Chronic disease or comorbidity presumed to account for new symptoms Especially in medically complex patients	• Septic joint signs misattributed to chronic rheumatoid arthritis • Mental status change due to infection or medication misattributed to underlying dementia
Rare diagnosis: failure to consider or know	• Many; fortunately, by definition, rare, but still warrant consideration especially if urgent or treatable
Drug or environmental factor not considered/overlooked Underlying etiology causing/ contributing to symptoms, or disease progression not sought, uncovered	• Ventricular arrhythmia related to QT-prolonging drug • Achilles tendon rupture related to quinolone
Failure to appreciate risk factors for particular disease	• Family history of breast, colorectal cancer not solicited and/or weighed in diagnostic evaluation or screening
Failure to appreciate limitations of physical exam Now with ↑ telemedicine, missing physical exam entirely	• Overweighing absence of tenderness, swelling in deep vein thrombosis • Missing pill-rolling tremor during telemedicine visit

DIAGNOSIS SAFETY CULTURE

Just as diagnosing bacterial infections relies on a proper culture medium to grow and identify etiologic organisms, good diagnosis also requires a healthy safety culture that will allow it to grow and flourish. While clinicians may be inclined to view "safety culture" as something too subjective to be important in their quest to make a definitive diagnosis, this view is misguided. Multiple studies have demonstrated adverse consequences resulting from organizational cultures that inhibit openness, learning, and sharing and create a climate where staff and patients are afraid to speak up when they observe problems or have questions. Most importantly, patients need to be encouraged to question diagnoses and be heard, particularly when they are not responding to treatment as expected or developing symptoms that are either not consistent with the diagnosis or represent possible red flags for other diagnoses or complications.

Studies examining "high-reliability organizations" outside of medicine and "learning health care organizations" have distilled a series of fundamental properties that are correlated with more reliable and safer outcomes. Just as a thermometer or recording of a pulse can suggest how ill a patient is, we now have instruments that can measure safety culture. These safety measurement tools typically are validated staff surveys that assess (1) communication about errors with staff willingness to report mistakes because they do not feel these mistakes are held against them; (2) openness and encouragement to talk about hospital/office problems; (3) existence of a learning culture that seeks to learn from errors and improve based on lessons learned; (4) leadership commitment to safety, prioritizing safety over production speed and the "bottom line" by providing adequate staffing and resources to operate safely; and (5) accountability and transparency for following up safety events and concerns. Each of these generic culture attributes translates into specific implications for diagnostic safety. These include the following:

- Making it "safe" for clinicians to admit and share diagnostic errors
- Proactive identification, ownership, and accountability regarding error-prone diagnostic workflow processes (particularly around test results, referrals, and patient follow-up)
- Leadership making diagnosis improvement a top priority based on recognition that patients and malpractice insurers report that diagnostic errors are the leading patient safety problem

- Mutual trust and respect for challenges that clinicians often face in making diagnoses and caution in applying the lens of hindsight bias in judging what in retrospect might seem like an "obvious" diagnosis that a clinician initially missed.

HEALTH INFORMATION TECHNOLOGY AND THE FUTURE OF DIAGNOSIS

Clinicians now spend more time interacting with computers than they do interacting with patients. This is especially true for diagnosis and will likely be even more so in the future. Interactions with patients, consultants, and other staff are increasingly mediated through the computer. Key activities, such as collecting patients' history (past and current), interpreting data to make a diagnosis, conveying diagnostic assessments (to others on the team and, increasingly, to the patient via open notes), and tracking diagnostic trajectories as they evolve over time, are now computer based. With the rise of telemedicine, even elements of the physical examination have been rerouted to electronic encounters, with important implications for diagnostic safety.

While many complain the computer has "gotten in the way" of good diagnosis, distracting clinicians from quality time listening to patients and miring doctors in reading and writing notes filled with copied/pasted/templated information of questionable currency and accuracy, medicine needs to harness the computer's capabilities to improve diagnosis **(Table 10-6)**. Although supporting these basic diagnosis capabilities should be the foundation of health information technology and everyday workflows, electronic medical records have historically been largely designed around other needs, such as ordering medications, billing, and documentation to guard against malpractice claims. They instead need to be redesigned radically to better support diagnostic processes, as well as save clinicians time. With the rise of generative artificial intelligence large language learning models, many are looking to the computer to take over the job of making diagnoses, answering patients' diagnostic questions, or resolving diagnostic dilemmas. However, despite its significant cabilities for image and data analysis, pattern recognition, creating clinical notes, and decision support including generating differential diagnoses, there are fundamental limitations, challenges, and unanswered questions related to data accuracy and how to incorporate human relational elements into AI-driven diagnostic processes.

TABLE 10-6 Areas Where Health Information Technology Has Potential to Help Improve Diagnosis and Reduce Errors

FUNCTION	EXAMPLES
Facilitate collection/ gathering of information	• Quickly access past history from prior care at same and outside institutions • Electronic collection of history of present illness, review of systems, and social determinant risks in advance of visits
Enhanced information entry, organization, and display	• Visually enhanced flowsheets showing trends, relationships to treatment • Reorganized notes to facilitate summarization and simplification and prevent items from getting lost
Generating differential diagnosis	• Automated creation of lists of diagnoses to consider based on patient's symptoms, demographics, risks • ChatGPT augmenting physician's diagnostic considerations
Weighing diagnoses likelihoods	• Tools to assist in calculation of posttest (Bayesian) probabilities
Aids for formulating diagnostic plan, intelligent test ordering	• Entering a diagnostic consideration (e.g., celiac disease, pheochromocytoma) and computer suggests most appropriate diagnostic test(s) and how to order
Access to diagnostic reference information	• Info-buttons instantly linking symptom- or diagnosis-relevant questions to Harrison's, Up-to-Date chapters, references
Ensuring more reliable follow-up	• Hardwiring "closed loops" to ensure abnormal labs, missed referrals, worrisome symptoms are tracked and followed up
Support screening for early detection	• Collaborative tools that patients, clinicians, and offices can use to know when due, order and track screening based on individualized demographics, risk factors, prior tests
Collaborative diagnosis; access to specialist	• Real-time posing/answering of questions • Electronic consults; virtual co-management
Facilitating feedback on diagnoses	• Feeding back new diagnoses (from downstream providers, patients) that emerge suggesting potential misdiagnosis/errors to clinicians, emergency rooms that saw patient previously

Source: Modified from G Schiff, DW Bates: N Engl J Med 362:1066, 2010, and R El-Karah et al: BMJ Qual Saf Suppl 2:ii40, 2013.

DIAGNOSIS OF DIAGNOSIS ERRORS AND SAFETY: PRACTICAL CONCLUSIONS

In practice, there are frequent and meaningful opportunities for improving diagnosis in each of the three NAM-defined areas to make it (1) more reliable and (2) timely, and (3) to improve diagnosis-related communication with patients. Clinicians in training, practicing physicians, nurses, and others should develop the habit of regularly asking themselves three questions about individual patients in their care, and another three questions regarding the systems in which they work. For each patient being assessed, clinicians should ask:

1. What else might this be? (forcing a differential diagnosis to be made)
2. What doesn't fit? (making sure unexplained abnormal findings are not dismissed)
3. What critical diagnoses are important not to miss? (injecting consideration of "don't miss" diagnoses, red flags, and known pitfalls)

And to diagnose safely, each practitioner must recognize that they are working within a larger system. Questions to be asking continually, ensuring we are maximizing reliability and timeliness and minimizing potential for errors, include:

1. Do we have reliable "closed loop" systems to provide reliable, ideally automated tracking and following up of patients' symptoms, abnormal laboratory or imaging findings, and critical referrals that we order?
2. What is the culture-of-safety climate in our organization, office, or clinic?
3. How does the electronic (or even paper) medical record as currently implemented help versus impair efficient, timely, accurate, and fail-safe diagnosis, and how can it be improved?

The challenge will be to take these questions to the next developmental stage in order to ensure diagnostic errors are recognized both in individual patients as well as prioritized for systemic changes in medicine. In the decade following the ground-breaking National Academy of Medicine report there have been a host of quality improvement and research efforts to better understand the epidemiology, causes, and ways to prevent diagnostic errors. While these efforts have had mixed success, they have enhanced understanding of where, how, and why things go wrong. Transforming diagnostic safety culture, ensuring reliable follow-up and feedback, learning from diagnostic errors, leveraging health IT, and more deeply partnering with patients will be essential elements for highest quality diagnosis in the future.

FURTHER READING

Moore Foundation: Viewpoint series aims to broaden understanding of diagnostic excellence. *https://www.moore.org/article-detail?newsUrlName=viewpoint-series-aims-to-broaden-understanding-of-diagnostic-excellence*.

National Academies of Sciences, Engineering, and Medicine. 2015. *Improving Diagnosis in Health Care. https://doi.org/10.17226/21794*. Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, DC.

Newman-Toker D et al: Rate of diagnostic errors and serious misdiagnosis-related harms for major vascular events, infections, and cancers: Toward a national incidence estimate using the "Big Three." Diagnosis 8:1, 67, 2021.

Schiff GD et al: Ten principles for more conservative, care-full diagnosis. Ann Intern Med 169:643, 2018.

Singh H et al: Developing the Safer Dx Checklist of ten safety recommendations for health care organizations to address diagnostic errors. Jt Comm J Qual Pat Saf 48:581, 2022.

Society to Improve Diagnosis in Medicine (SIDM): Resource page. *https://www.improvediagnosis.org/resources-for/*.

11

Racial and Ethnic Disparities in Health Care

Lenny López, Joseph R. Betancourt

Over the course of its history, the United States has experienced dramatic improvements in overall health and life expectancy, largely as a result of initiatives in public health, health promotion, disease prevention, and chronic care management. Our ability to prevent, detect, and treat diseases in their early stages has allowed us to target and reduce rates of morbidity and mortality. Despite interventions that have improved the overall health of the majority of Americans, racial and ethnic minorities (blacks, Hispanics/Latinos, Native Americans/Alaskan Natives, Asian/Pacific Islanders) have benefited less from these advances than whites and have suffered poorer health outcomes from many major diseases, including cardiovascular disease, cancer, and diabetes. In 2018, the estimated economic burden of racial and ethnic health inequities was between $421 billion and $451 billion and between $940 billion and $978 billion for adults without a 4-year college degree. These disparities highlight the importance of recognizing and addressing the multiple factors that impact health outcomes, including structural racism, *social determinants of health* (SDOH), access to care, and health care quality. On this last point, research has revealed that minorities may receive less care and lower-quality care than whites, even when confounders such as stage of presentation, comorbidities, and health insurance are controlled. These differences in quality are called *racial and ethnic disparities in health care*. These health care disparities have taken on greater importance with the significant transformation of the U.S. health care system and value-based purchasing. The shift toward creating financial incentives and disincentives to achieve quality goals makes focusing on those who receive lower-quality care more important than ever before. This chapter will provide an overview of racial and ethnic disparities in health and health care, identify root causes, and provide key recommendations to address these disparities at both the clinical and health system levels.

NATURE AND EXTENT OF DISPARITIES

Life expectancy at birth is an important measure of the health of a nation's population. Although the overall life expectancy in the United States has been increasing since 1900, differences due to race/ethnicity, education, and socioeconomic status have persisted. For example, at every level of education and income, African Americans have lower life expectancy at age 25 than whites and Hispanics/Latinos. Blacks with a college degree or more education have lower life expectancy than whites and Hispanics who graduated from high school. Blacks have had lower life expectancy compared to whites for as long as data have been collected. From 1975 to 2003, the largest difference in life expectancy between blacks and whites was substantial (6.3 years for males and 4.5 years for females). The gap in life expectancy between the black and white populations decreased by 2.3 years between 1999 and 2013 from 5.9 to 3.6 years (4.4 years for males and 3.0 years for women). During 1999–2017, mortality decreases occurred in all racial/ethnic groups except non-Hispanic American Indian and Alaskan Native adults, who experienced steady increases in midlife mortality through 2017. The pattern of high midlife mortality rates among non-Hispanic American Indian and Alaskan Native and non-Hispanic black adults exceeded rates among other racial/ethnic groups. This pattern remained through 2020, with male non-Hispanic American Indian or Alaskan Native and black males having the lowest life expectancies at birth **(Fig. 11-1)**.

The life expectancy gap is augmented by worse health and higher disease burden. Cardiovascular-related diseases remain the leading cause of black-white differences in life expectancy. If all cardiovascular causes and diabetes are considered together, they account for 35% and 52% of the gap for males and females, respectively. During the COVID-19 pandemic, the U.S. population experienced the most significant 2-year decline in life expectancy in roughly a century, disproportionately impacting people of color and exacerbating longstanding racial disparities in life expectancy. While overall U.S. life expectancy declined by 2.7 years between 2019 and 2021, American Indian and Alaskan Native people experienced a decline of 6.6 years and Hispanic people and black people experienced a decline of 4.2 and 4 years, respectively, compared to a decline of 2.4 years for white people and 2.1 years for Asian people. Finally, place matters for health. Analysis of data from 2010 to 2015 demonstrate large geographic life expectancy

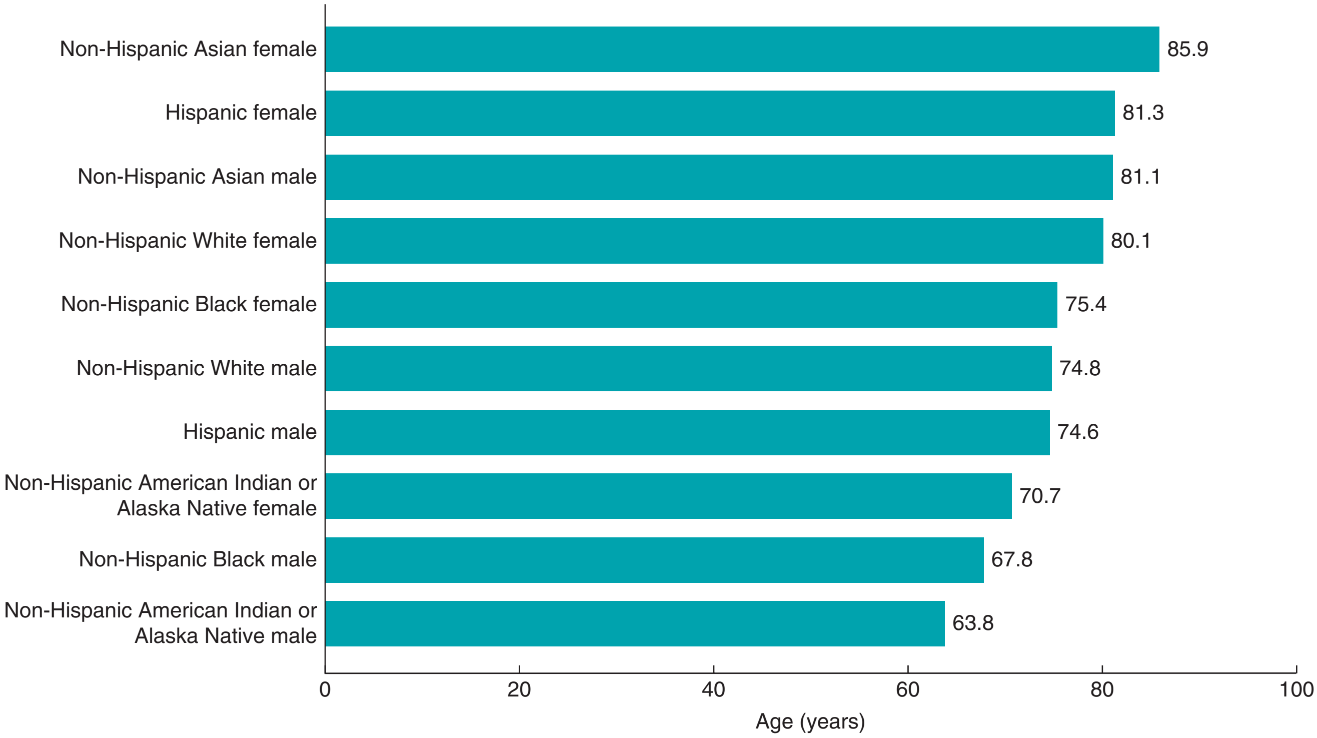

FIGURE 11-1 Life expectancy at birth, by Hispanic origin and race and sex: United States, 2020. *(From E Arias, JQ Xu: United States life tables, 2020. National Vital Statistics Reports; vol 71 no 1. Hyattsville, MD: National Center for Health Statistics. 2022. Retrieved from https://www.cdc.gov/nchs/data/nvsr/nvsr71/nvsr71-01.pdf.)*

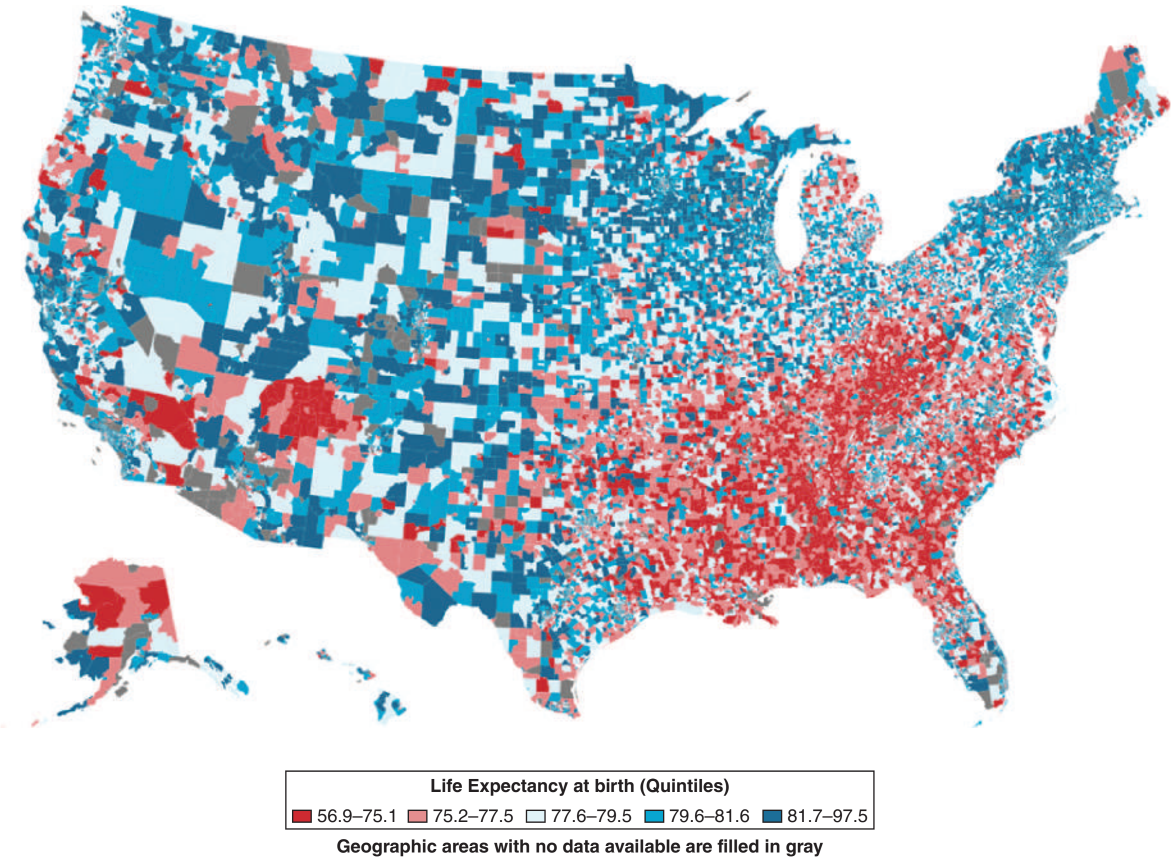
Life Expectancy at birth (Quintiles)
🟥 56.9–75.1
🟥 75.2–77.5
☐ 77.6–79.5
🟦 79.6–81.6
🟦 81.7–97.5
Geographic areas with no data available are filled in gray

FIGURE 11-2 Life expectancy at birth for U.S. census tracts, 2010–2015. *(From A New View of Life Expectancy, Surveillance and Data - Blogs and Stories, Centers for Disease Control and Prevention. Retrieved from https://www.cdc.gov/surveillance/blogs-stories/life-expectancy.html.)*

gap variation at the census tract level **(Fig. 11-2)**. Socioeconomic and race/ethnicity factors, behavioral and metabolic risk factors (prevalence of obesity, leisure-time physical inactivity, cigarette smoking, hypertension, diabetes), and health care factors (percentage of the population younger than 65 years who are insured, primary care access and quality, number of physicians per capita) explained 60%, 74%, and 27% of county-level variation in life expectancy, respectively. Combined, these factors explained 74% of this variation. Most of the association between socioeconomic and race/ethnicity factors and life expectancy was mediated through behavioral and metabolic risk factors.

In addition to racial and ethnic disparities in *health*, there are racial and ethnic disparities in the *quality of care* for persons with access to the health care system. Seminal studies over several decades have consistently documented disparities in health care. For instance, studies have documented disparities in the treatment of pneumonia and congestive heart failure, with blacks receiving less optimal care than whites when hospitalized for these conditions. Moreover, blacks with end-stage renal disease are referred less often to the transplant list than are their white counterparts **(Fig. 11-3)**. Disparities have been found, for example, in the use of cardiac diagnostic and therapeutic procedures (with blacks being referred less often than whites for cardiac catheterization and bypass grafting), prescription of analgesia for pain control (with blacks and Hispanics/Latinos receiving less pain medication than whites for long-bone fractures and cancer), and surgical treatment of lung cancer (with blacks receiving less curative surgery than whites for non-small-cell lung cancer). Again, many of these disparities have occurred even when variations in factors such as insurance status, income, age, comorbid conditions, and symptom expression are taken into account. Finally, disparities in the quality of care provided at the sites where minorities tend to receive care have been shown to be an important additional contributor to overall disparities.

The 2022 National Healthcare Quality and Disparities Report, released by the Agency for Healthcare Research and Quality, tracks more than 400 health care process, outcome, and access measures, across many diseases and settings. This annual report is particularly important because most studies of disparities have not been longitudinally repeated with the same methodology to document trends and changes in disparities over time. This report found that some disparities were getting smaller from 2000 through 2016–2018, but disparities persisted and some even worsened, especially for poor and uninsured populations. For about 45% of quality measures, blacks (85 of 190 measures) and American Indians and Alaska Natives (47 of 110 measures) received worse care than whites. For more than one-third of quality measures, Hispanics (73 of 190 measures), Asians, and Native Hawaiians/Pacific Islanders received worse care than whites, but Asians also received better care for about 28% of quality measures **(Fig. 11-4)**. Of note, for those quality measures that demonstrated disparities at baseline, >90% of these measures showed no improvement since 2000 **(Fig. 11-5)**.

ROOT CAUSES OF DISPARITIES

Race, Racism, and Health Race and racism are core elements of any explanatory model on racial and ethnic disparities in health and health care. Our nation's history of slavery, segregation, separate but "equal" health care, and medical experimentation, among a myriad of other ways in which racism has manifested in the United States, has played a key role in the existence and persistence of these disparities. It is now well accepted that race is a social category without biologic foundation and a product of historical racism. Nevertheless, it is clear that racism has a biologic impact as a form of psychosocial stress. It is now well established that psychosocial stress negatively impacts health through psychophysiologic reactivity causing hyperstimulation of the sympathetic-adrenal-medullary system and the hypothalamic-pituitary-adrenal axis, leading to vascular inflammation, endothelial dysfunction, and neurohormonal dysregulation, causing an acceleration of cardiovascular disease. Behavioral changes occurring as adaptations or coping responses to stressors such as

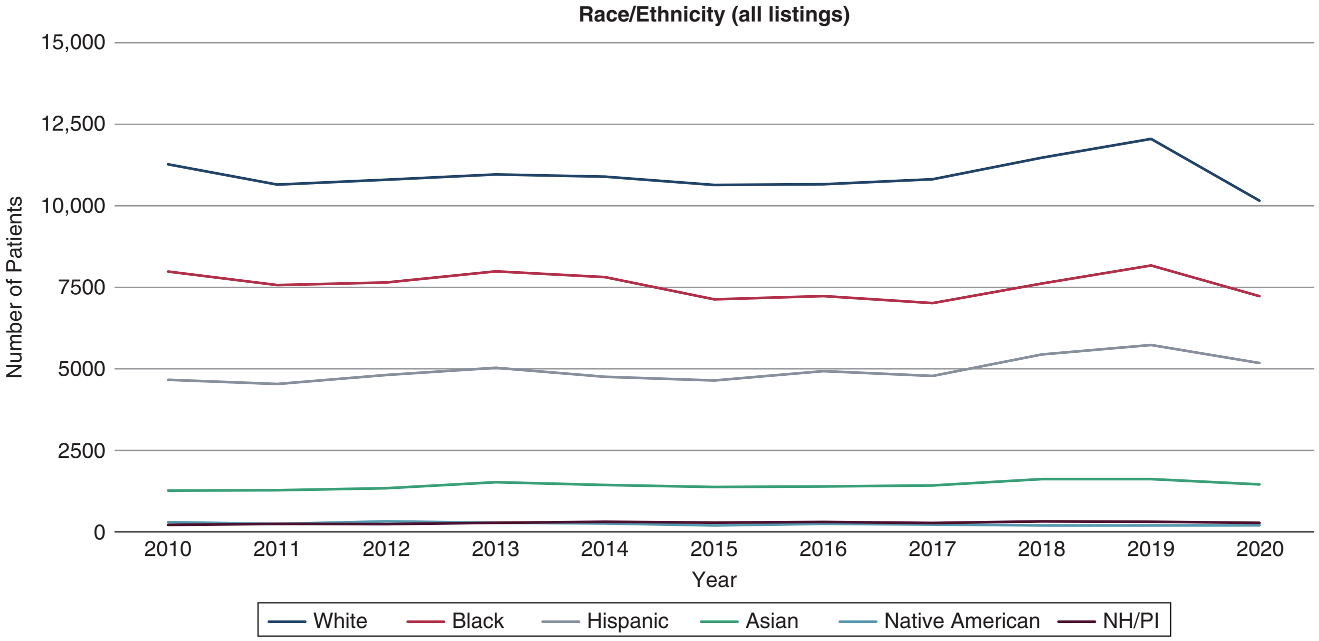

FIGURE 11-3 Number of end-stage renal disease patients added to the waitlist for a kidney transplant, 2010–2020. NH/PI, Native Hawaiian/Pacific Islander. *(From US Renal Data System. 2022 USRDS Annual Data Report: Epidemiology of Kidney Disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2022. https://usrds-adr.niddk.nih.gov/2022/end-stage-renal-disease/7-transplantation. Accessed October 25, 2023.)*

increased smoking, decreased exercise and sleep, and poorer adherence to medical regimens provide an additional important pathway through which stressors influence disease risk. This accelerated disease risk, aging, and premature death has been termed the *weathering effect*.

While most empiric research focuses on interpersonal racial/ethnic discrimination, structural racism (sometimes called institutional racism) provides a more holistic framework. Structural racism refers to the totality of ways that a society fosters, sustains, and reinforces discrimination through sociopolitical, legal, economic, and health structures that determine differential access to risks, opportunities, and resources that drive health and health care disparities. Structural racism explains how racism's structure and ideology can persist in governmental and institutional policies in the absence of individual actors who are explicitly racially prejudiced. For example, the history of residential segregation has had lasting negative effects generationally on equal access for racial/ethnic minorities to employment, banking, earnings, high-quality education, and health care. Policies that do not address root structural causes will not address health and health care inequities.

With the promise of individualizing clinical decisions, the use of race in clinical and risk assessment algorithms has long been a part of modern medicine. The evidence is now clear that race is not a reliable proxy for genetic difference and that race adjustment has the potential to create inadvertent disparities in health care. One clinical example is from nephrology. Blacks have higher rates of end-stage kidney disease and death due to kidney failure than the overall population. The most widely used cohort-derived equation to estimate glomerular filtration rate (GFR), the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation, has the limitation that it produces 80–90% estimated GFR (eGFR) values that are within ±30% of a patient's measured GFR. In addition, this equation uses a black race-related factor, which increases eGFR for any given serum creatinine by 15.9% compared to a nonblack patient with the same age, sex, and serum creatinine. The increase in eGFR is likely to disadvantage blacks for early referral to a nephrologist, early treatment of advanced chronic kidney disease, and kidney transplantation. It is also not clear how to apply the race factor when the patient's race is unknown and/or ambiguous, as in those who are multiracial. This disparity-inducing scenario could be avoided through the use of cystatin C–based eGFR estimation, which has been demonstrated to be more accurate than the CKD-EPI equation and for which race is not required in estimation.

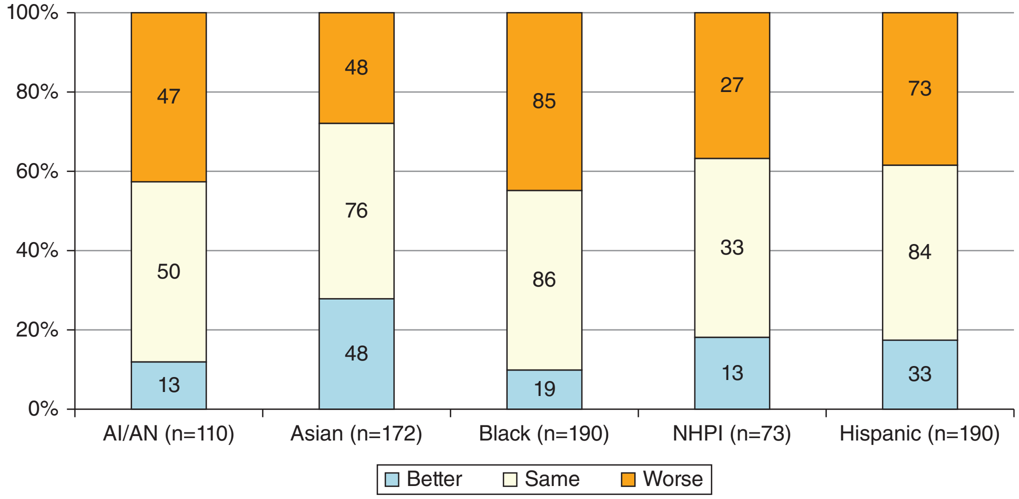

FIGURE 11-4 Number and percentage of quality measures for which members of selected groups experienced better, same, or worse quality of care compared with white people for the most recent data year, 2017, 2018, 2019, or 2020. AI/AN, American Indian or Alaska Native; NHPI, Native Hawaiian/Pacific Islander *(From 2022 National Healthcare Quality and Disparities Report. Rockville, MD: Agency for Healthcare Research and Quality; October 2022. AHRQ Publication No. 22(23)-0030. https://www.ahrq.gov/sites/default/files/wysiwyg/research/findings/nhqrdr/2022qdr-appendix-combined.pdf. Accessed October 25, 2023.)*

The application of artificial intelligence (AI) analytics to large amounts of clinical electronic data—big data—holds the promise to better understand health care costs, utilization, resource allocation, and population health monitoring. Machine learning models can identify the statistical patterns in large amounts of

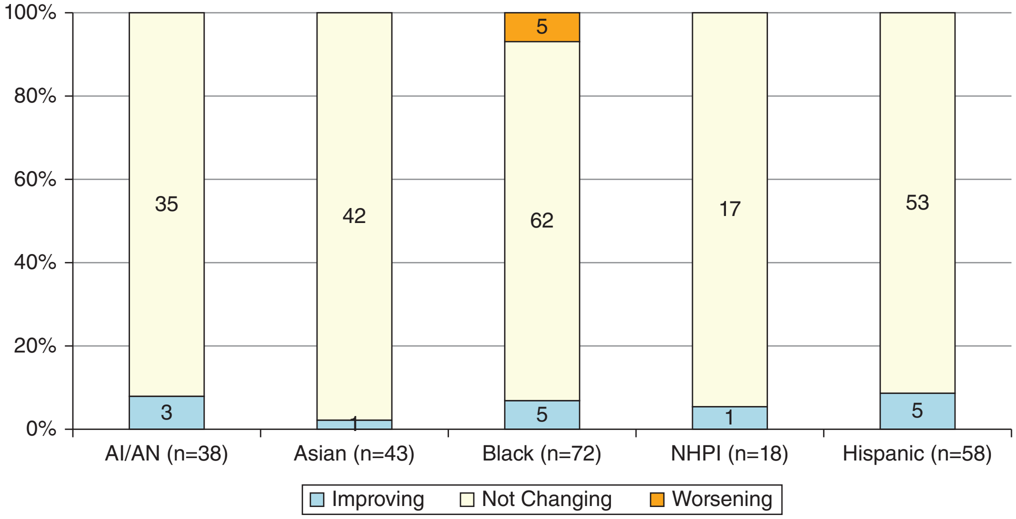

FIGURE 11-5 Number and percentage of quality measures with disparity at baseline for which disparities related to race and ethnicity were improving, not changing, or worsening over time, 2000–2020. AI/AN, American Indian or Alaska Native; NHPI, Native Hawaiian/Pacific Islander. *(From 2022 National Healthcare Quality and Disparities Report. Rockville, MD: Agency for Healthcare Research and Quality; October 2022. AHRQ Publication No. 22(23)-0030. https://www.ahrq.gov/sites/default/files/wysiwyg/research/findings/nhqrdr/2022qdr-appendix-combined.pdf. Accessed October 25, 2023.)*

historically collected data. These data naturally contain the patterning of preexisting health care disparities created by socially and historically structured inequities. This biased patterning can lead to incorrect predictions, withholding of resources, and worse outcomes for vulnerable populations. Recently, analysis of a commercial, national, proprietary prediction algorithm, affecting millions of patients, exhibited racial bias. Historical cost data were used to predict clinical risk and allocate additional clinical services for high-cost patients. Algorithmic bias arose because black patients historically have less access to health care and thus less money is spent on their care compared to white patients. Thus, blacks, who tended to be sicker than white patients, received lower clinical risk scores and thus were less likely to receive additional clinical services. The observed allocation bias was remedied using direct measures of illness and illness severity. Thus, machine learning algorithms are not inherently free of bias and should be assessed for accuracy and fairness.

In summary, there are many ways in which racism has contributed and does and will continue to contribute to racial and ethnic disparities in health and health care.

SOCIAL DETERMINANTS OF HEALTH

Minority Americans have poorer health outcomes than whites from preventable and treatable conditions such as cardiovascular disease, diabetes, asthma, cancer, and HIV/AIDS. Multiple factors contribute to these racial and ethnic disparities in health. The landmark National Academy of Medicine (formerly, the Institute of Medicine [IOM]) report, *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*, published in 2002, summarized the scientific evidence on health disparities and provided an important framework for conceptualizing and defining racial/ethnic disparities. Since the *Unequal Treatment* report, there has been a growing empiric evidence base on how racism and the SDOH, often working in synergy, create and sustain disparities. Mechanistically, the biopsychosocial model brings together the social and physical characteristics of the environment with individual physical and psychological attributes. These environmental and individual characteristics, in turn, influence health behaviors and stress-related physiologic pathways that directly impact health. The National Institute on Minority Health and Health Disparities SDOH model builds on prior models and adds the time element across the life course of the individual in recognition of the long-lasting health effects of socioeconomic exposures **(Fig. 11-6)**. The resulting matrix has the domains of influence of health (biological, behavioral, physical and built environment, sociocultural environment, health care system) along the y-axis and the levels of influence on health (individual, interpersonal, community, societal) along the x-axis. Cells are not mutually exclusive, and examples of factors within each cell are illustrative and not comprehensive. This framework emphasizes the complex multidomain etiologies of disparities across the factors in the conceptual matrix, thus highlighting the limitation of individual-level focused research and policy.

The Centers for Disease Control and Prevention's (CDC) Social Vulnerability Index (SVI) is one of the most comprehensive indicators of SDOH for every U.S. county and census tract **(Fig. 11-7)**. The 15 social factors are grouped into four broad themes: (1) socioeconomic status (below poverty, unemployed, income level, and no high school diploma); (2) household composition and disability (aged 65 years or older, aged 17 years or younger, individuals >5 years with a disability, and single-parent households); (3) minority status and language (minority and individual speaks English "less than well"); and (4) housing type and transportation (multiunit structure, mobile home, crowding, group quarters, and no vehicle). Recent studies using national databases have demonstrated a strong geographic concentration of negative factors by counties and regions in the United States. States in the highest tertile of social vulnerability had predominantly black and Hispanic adults, lower levels of education, lower income, and higher rates of unemployment and substance use. These individuals also had higher rates of hypertension, diabetes, hyperlipidemia, chronic kidney disease, smoking, and atherosclerotic cardiovascular disease compared with those living in counties in the first tertile of SVI. Importantly, there is a graded increase in mortality related to cancer, cardiovascular disease, and comorbid cancer and cardiovascular disease in U.S. counties with the worst SVI.

In addition to race and racism, *Unequal Treatment* identified a set of root causes that included health system, provider-level, and patient-level factors.

Health System Factors • **HEALTH SYSTEM COMPLEXITY** Even among persons who are insured and educated and who have a high degree of health literacy, navigating the U.S. health care system can be complicated and confusing. Some individuals may be at higher risk for receiving substandard care because of their difficulty navigating the system's complexities. These individuals may include those from cultures unfamiliar with the Western model of health care delivery, those with limited English proficiency, those with low health literacy, and those who are mistrustful of the health care system. These individuals may have difficulty knowing how and where to go for a referral to a specialist; how to prepare for a procedure such as a colonoscopy; or how to follow up on an abnormal test result such as a mammogram. Since people of color in the United States tend to be overrepresented among the groups listed above, the inherent complexity of navigating the health care system has been seen as a root cause for racial/ethnic disparities in health care.

OTHER HEALTH SYSTEM FACTORS Racial/ethnic disparities are due not only to differences in care provided within hospitals but also to where and from whom minorities receive their care; i.e., certain specific providers, geographic regions, or hospitals are lower-performing on certain aspects of quality. For example, one study showed that 25% of hospitals cared for 90% of black Medicare patients in the United States and that these hospitals tended to have lower performance scores on certain quality measures than other hospitals. That said, health systems generally are not well prepared to measure, report, and intervene to reduce disparities in care. Few hospitals or health plans stratify their quality data by race/ethnicity or language to measure disparities, and even fewer use data of this type to develop disparity-targeted

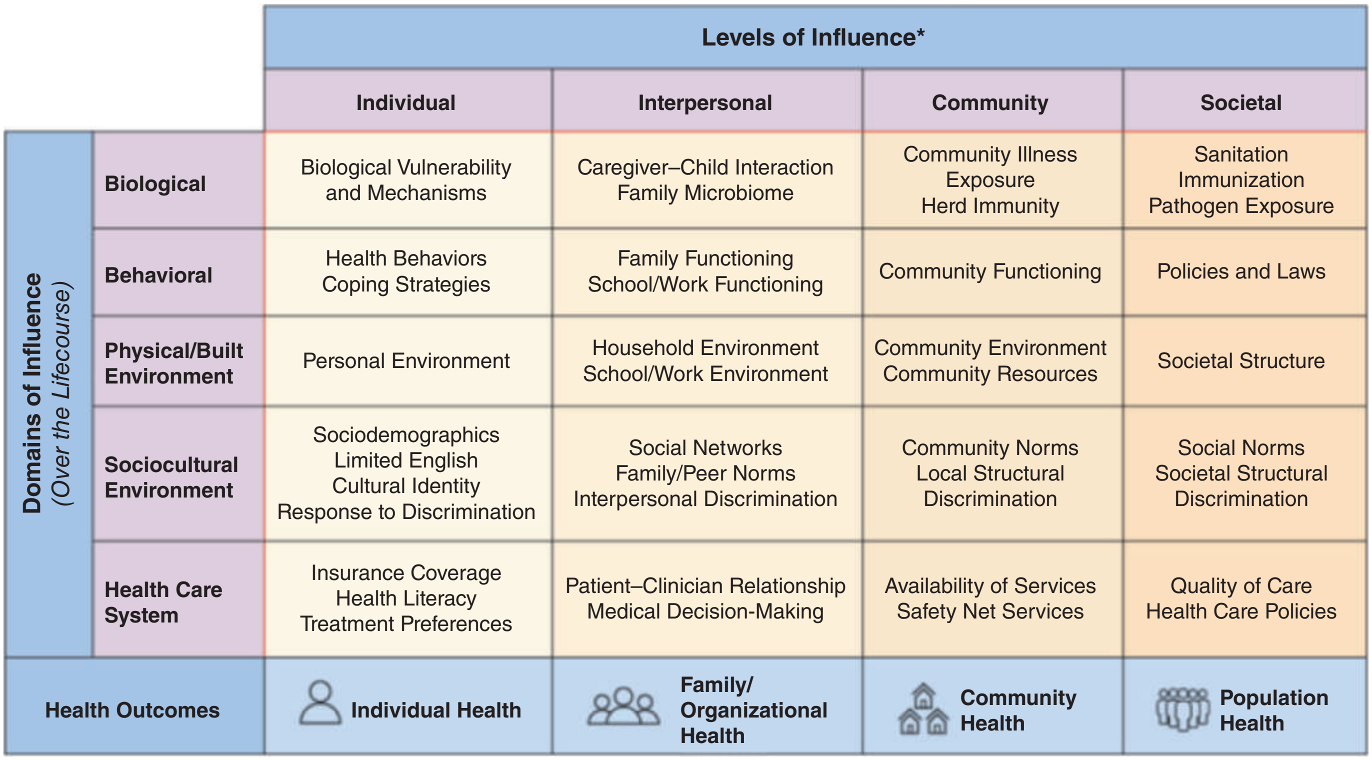

Domains of Influence (Over the Lifecourse)	Levels of Influence*: Individual	Interpersonal	Community	Societal
Biological	Biological Vulnerability and Mechanisms	Caregiver–Child Interaction Family Microbiome	Community Illness Exposure Herd Immunity	Sanitation Immunization Pathogen Exposure
Behavioral	Health Behaviors Coping Strategies	Family Functioning School/Work Functioning	Community Functioning	Policies and Laws
Physical/Built Environment	Personal Environment	Household Environment School/Work Environment	Community Environment Community Resources	Societal Structure
Sociocultural Environment	Sociodemographics Limited English Cultural Identity Response to Discrimination	Social Networks Family/Peer Norms Interpersonal Discrimination	Community Norms Local Structural Discrimination	Social Norms Societal Structural Discrimination
Health Care System	Insurance Coverage Health Literacy Treatment Preferences	Patient–Clinician Relationship Medical Decision-Making	Availability of Services Safety Net Services	Quality of Care Health Care Policies
Health Outcomes	**Individual Health**	**Family/ Organizational Health**	**Community Health**	**Population Health**

FIGURE 11-6 National Institute on Minority Health and Health Disparities social determinants research framework. *Health disparity populations: race/ethnicity, low socioeconomic status, rural, sexual and gender minority. Other fundamental characteristics: sex and gender, disability, geographic region. *(From National Institute on Minority Health and Health Disparities. NIMHD Research Framework. 2017. Retrieved from https://www.nimhd.nih.gov/about/overview/research-framework.html.)*

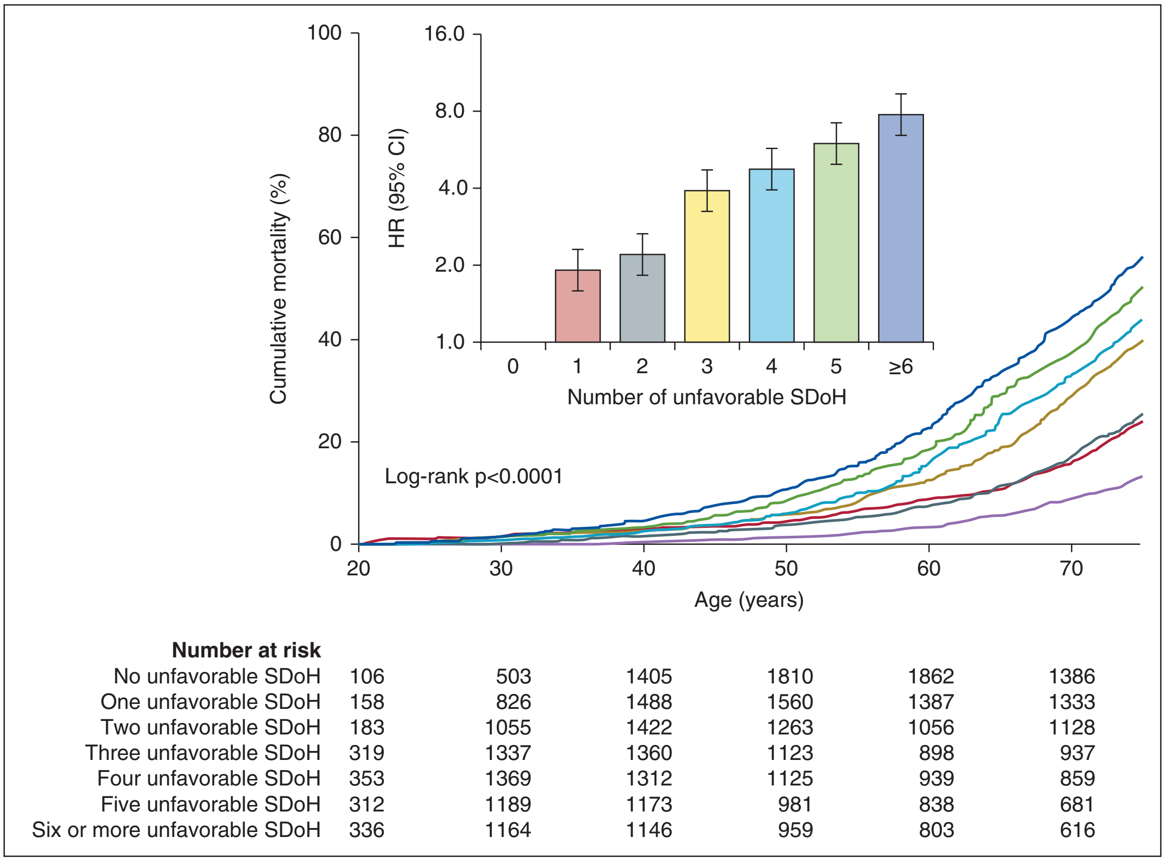

Number at risk	20	30	40	50	60	70
No unfavorable SDoH	106	503	1405	1810	1862	1386
One unfavorable SDoH	158	826	1488	1560	1387	1333
Two unfavorable SDoH	183	1055	1422	1263	1056	1128
Three unfavorable SDoH	319	1337	1360	1123	898	937
Four unfavorable SDoH	353	1369	1312	1125	939	859
Five unfavorable SDoH	312	1189	1173	981	838	681
Six or more unfavorable SDoH	336	1164	1146	959	803	616

FIGURE 11-7 Premature all-cause mortality and hazard ratios (HRs) in U.S. adults aged 20–74 years according to the number of unfavorable social determinants of health (SDoHs). Kaplan-Meier curves show cumulative mortality by age and number of unfavorable SDoHs. Bar chart shows HRs of premature all-cause mortality associated with number of unfavorable SDoHs, adjusted for age, gender, and race and ethnicity; error bars are 95% confidence intervals. *(Reproduced with permission from JD Bundy et al: Social determinants of health and premature death among adults in the USA from 1999 to 2018: A national cohort study. Lancet Public Health 8:e422, 2023.)*

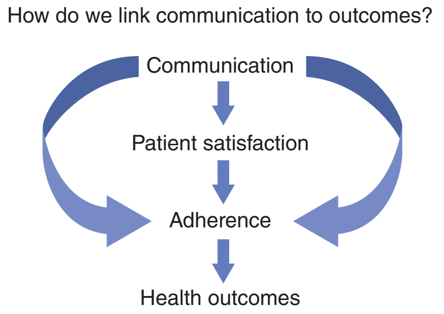

FIGURE 11-8 The link between effective communication and patient satisfaction, adherence, and health outcomes. *(From the Institute of Medicine. 2003. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. https://doi.org/10.17226/12875. Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)*

interventions. Similarly, despite regulations concerning the need for professional interpreters, research demonstrates that many health care organizations and providers fail to routinely provide this service for patients with limited English proficiency. Despite the link between limited English proficiency and health care quality and safety, few providers or institutions monitor performance for patients in these areas.

Provider-Level Factors • **PROVIDER–PATIENT COMMUNICATION** Significant evidence highlights the impact of sociocultural factors, race, ethnicity, and limited English proficiency on health and clinical care. Health care professionals frequently care for diverse populations with varied perspectives, values, beliefs, and behaviors regarding health and well-being. The differences include variations in the recognition of symptoms, thresholds for seeking care, comprehension of management strategies, expectations of care (including preferences for or against diagnostic and therapeutic procedures), and adherence to preventive measures and medications. In addition, sociocultural differences between patient and provider influence communication and clinical decision-making and are especially pertinent: evidence clearly links provider–patient communication to improved patient satisfaction, regimen adherence, and better health outcomes **(Fig. 11-8)**. Thus, when sociocultural differences between patient and provider are not appreciated, explored, understood, or communicated effectively during the medical encounter, patient dissatisfaction, poor adherence, poorer health outcomes, and racial/ethnic disparities in care may result.

A survey of 6722 Americans ≥18 years of age is particularly relevant to this important link between provider–patient communication and health outcomes. Whites, African Americans, Hispanics/Latinos, and Asian Americans who had made a medical visit in the past 2 years were asked whether they had trouble understanding their doctors; whether they felt the doctors did not listen; and whether they had medical questions they were afraid to ask. The survey found that 19% of all patients experienced one or more of these problems, yet whites experienced them 16% of the time as opposed to 23% of the time for African Americans, 33% for Hispanics/Latinos, and 27% for Asian Americans **(Fig. 11-9)**.

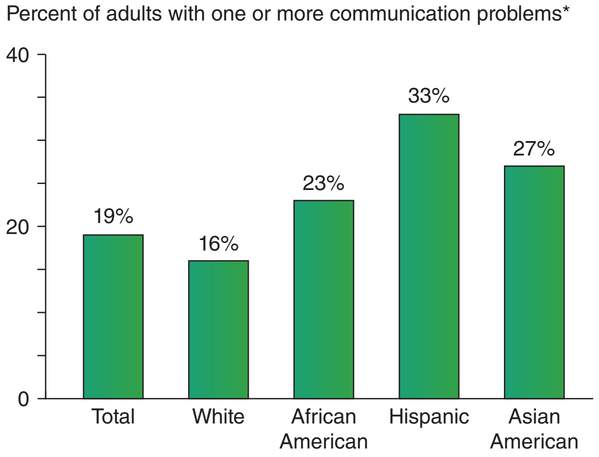

Base: Adults with health care visit in past two years
*Problems include understanding doctor, feeling doctor listened, had questions but did not ask.

FIGURE 11-9 Communication difficulties with physicians, by race/ethnicity. The reference population consisted of 6722 Americans ≥18 years of age who had made a medical visit in the previous 2 years and were asked whether they had had trouble understanding their doctors, whether they felt that the doctors had not listened, and whether they had had medical questions they were afraid to ask. *(Reproduced with permission from the Commonwealth Fund Health Care Quality Survey, 2001.)*

In addition, in the setting of even a minimal language barrier, provider–patient communication without an interpreter is recognized as a major challenge to effective health care delivery. These communication barriers for patients with limited English proficiency lead to frequent misunderstanding of diagnosis, treatment, and follow-up plans; inappropriate use of medications; lack of informed consent for surgical procedures; high rates of adverse events with more serious clinical consequences; and a lower-quality health care experience than is provided to patients who speak fluent English. Physicians who have access to trained interpreters report a significantly higher quality of patient–physician communication than physicians who use other methods. Communication issues related to discordant language disproportionately affect minorities and likely contribute to racial/ethnic disparities in health care.

CLINICAL DECISION-MAKING Theory and research suggest that variations in clinical decision-making may contribute to racial and ethnic disparities in health care. Two factors are central to this process: clinical uncertainty and stereotyping.

First, a doctor's decision-making process is nested in *clinical uncertainty*. Doctors depend on inferences about severity based on what they understand about illness and the information obtained from the patient. A doctor caring for a patient whose symptoms he or she has difficulty understanding and whose "signals"—the set of clues and indications that physicians rely on to make clinical decisions—are hard to read may make a decision different from the one that would be made for another patient who presents with exactly the same clinical condition. Given that the expression of symptoms may differ among cultural and racial groups, doctors—the overwhelming majority of whom are white—may understand symptoms best when expressed by patients of their own racial/ethnic groups. The consequence is that white patients may be treated differently from minority patients. Differences in clinical decisions can arise from this mechanism even when the doctor has the same regard for each patient (i.e., is not prejudiced).

Second, the literature on social cognitive theory highlights how natural tendencies to stereotype may influence clinical decision-making. *Stereotyping* can be defined as the way in which people use social categories (e.g., race, gender, age) in acquiring, processing, and recalling information about others. Faced with enormous information loads and the need to make many decisions, people often subconsciously simplify the decision-making process and lessen cognitive effort by using "categories" or "stereotypes" that bundle information into groups or types that can be processed more quickly. Although functional, stereotyping can be systematically biased, as people are automatically classified into social categories based on dimensions such as *race*, *gender*, and *age*. Many people may not be aware of their attitudes, may not consciously endorse specific stereotypes, and paradoxically may consider themselves egalitarian and not prejudiced.

Stereotypes may be strongly influenced by the messages presented consciously and unconsciously in society. For instance, if the media and our social/professional contacts tend to present images of minorities as being less educated, more violent, and nonadherent to health care recommendations, these impressions may generate stereotypes that unnaturally and unjustly impact clinical decision-making. As signs of racism, classism, gender bias, and ageism are experienced (consciously or unconsciously) in our society, stereotypes may be created that impact the way doctors manage patients from these groups. On the basis of training or practice location, doctors may develop certain perceptions about race/ethnicity, culture, and class that may evolve into stereotypes. For example, many medical students and residents are trained—and minorities cared for—in academic health centers or public hospitals located in socioeconomically disadvantaged areas. As a result, doctors may begin to equate certain races and ethnicities with specific health beliefs and behaviors (e.g., "these patients" engage in risky behaviors, "those patients" tend to be noncompliant) that are more associated with the social environment (e.g., poverty) than with a patient's racial/ethnic background or cultural traditions. This "conditioning" phenomenon may also be operative if doctors are faced with certain racial/ethnic patient groups who frequently do not choose aggressive forms of diagnostic or therapeutic intervention. The result over time may be that doctors begin to believe that "these patients" do not like invasive procedures; thus, they may not offer these procedures as options. A wide range of studies have documented the potential for provider biases to contribute to racial/ethnic disparities in health care. For example, one study measured physicians' unconscious (or implicit) biases and showed that these were related to differences in decisions to provide thrombolysis for a hypothetical black or white patient with a myocardial infarction.

It is important to differentiate stereotyping from prejudice and discrimination. *Prejudice* is a conscious prejudgment of individuals that may lead to disparate treatment, and *discrimination* is conscious and intentional disparate treatment. All individuals *stereotype* subconsciously, yet, if left unquestioned, these subconscious assumptions may lead to lower-quality care for certain groups because of differences in clinical decision-making or differences in communication and patient-centeredness. For example, one study tested physicians' unconscious racial/ethnic biases and showed that patients perceived more biased physicians as being less patient-centered in their communication. What is particularly salient is that stereotypes tend to be activated most in environments where the individual is stressed, multitasking, and under time pressure—the hallmarks of the clinical encounter. In fact, in a survey of close to 16,000 physicians, 42% admitted that bias—including by race and ethnicity—impacted their clinical decision-making. Interestingly, emergency medicine physicians, who work in environments of stress, time pressure, risk, and where they are multitasking, topped the list by discipline at 62%.

Patient-Level Factors Lack of trust has become a major concern for many health care institutions today. For example, an IOM report, *To Err Is Human: Building a Safer Health System*, documented alarming rates of medical errors that made patients feel vulnerable and less trustful of the U.S. health care system. The increased media and academic attention to problems related to quality of care (and of disparities themselves) has clearly diminished trust in doctors and nurses.

Trust is a crucial element in the therapeutic alliance between patient and health care provider. It facilitates open communication and is directly correlated with adherence to the physician's recommendations and the patient's satisfaction. In other words, patients who mistrust their health care providers are less satisfied with the care they receive, and mistrust of the health care system greatly affects patients' use of services. Mistrust can also result in inconsistent care, "doctor-shopping," self-medication, and an increased demand by patients for referrals and diagnostic tests.

On the basis of historic factors such as discrimination, segregation, and medical experimentation, blacks may be especially mistrustful of providers. The exploitation of blacks by the U.S. Public Health Service during the Tuskegee syphilis study from 1932 to 1972 left a legacy of mistrust that persists even today among this population. Other populations, including Native Americans/Alaskan Natives, Hispanics/Latinos, and Asian Americans, also harbor significant mistrust of the health care system. A national survey conducted by the Kaiser Family Foundation found that there is significant mistrust for the health care system among minority populations. Of the 3884 individuals surveyed, 36% of Hispanics and 35% of blacks (compared to 15% of whites) felt they were treated unfairly in the health care system in the past based on their race and ethnicity. Perhaps even more alarming—65% of blacks and 58% of Hispanics (compared to 22% of whites) were afraid of being treated unfairly in the future based on their race/ethnicity **(Fig. 11-10)**.

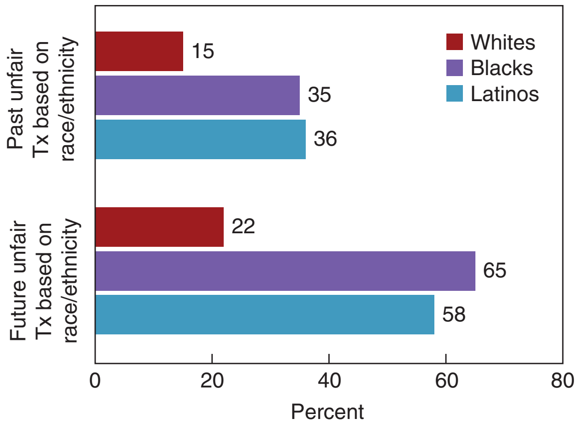

FIGURE 11-10 Patient perspectives regarding unfair treatment (Tx) based on race/ethnicity. The reference population consisted of 3884 individuals surveyed about how fairly they had been treated in the health care system in the past and how fairly they felt they would be treated in the future on the basis of their race/ethnicity. *(From Race, Ethnicity & Medical Care: A Survey of Public Perceptions and Experiences. Kaiser Family Foundation, 2005.)*

This mistrust may contribute to wariness in accepting or following recommendations, undergoing invasive procedures, or participating in clinical research, and these choices, in turn, may lead to misunderstanding and the perpetuation of stereotypes among health professionals.

KEY RECOMMENDATIONS TO ADDRESS RACIAL/ETHNIC DISPARITIES IN HEALTH CARE

Unequal Treatment provides recommendations to address the root causes of racial/ethnic disparities organized as *health system interventions*, *provider interventions*, *patient interventions*, and *general recommendations*.

Health System Interventions • **COLLECTING, REPORTING, AND TRACKING OF DATA ON HEALTH CARE ACCESS AND USE, BY PATIENTS' RACE/ETHNICITY** *Unequal Treatment* found that the appropriate systems to track and monitor racial and ethnic disparities in health care are lacking and that less is known about the disparities affecting minority groups other than African Americans (Hispanics, Asian Americans, Pacific Islanders, Native Americans, and Alaskan Natives). For instance, only in the mid-1980s did the Medicare database begin to collect data on patient groups outside the standard categories of "white," "black," and "other." Federal, private, and state-supported data-collection efforts are scattered and unsystematic, and many health care systems and hospitals still do not collect data on the race, ethnicity, or primary language of enrollees or patients. A survey by the Institute for Diversity in Health Management and the Health Research and Educational Trust in 2015 found that 98% of 1083 U.S. hospitals collected information on race, 95% collected data on ethnicity, and 94% collected data on primary language. However, only 45% collected data on race, 40% collected data on ethnicity, and 38% collected data on primary language to benchmark gaps in care. A survey by America's Health Insurance Plans Foundation in 2008 and 2010 showed that the proportion of enrollees in plans that collected race/ethnicity data of some type increased from 75 to 79%; however, the total percentage of plan enrollees whose race/ethnicity and language are recorded is still much lower than these figures.

COLLECTING, REPORTING, AND TRACKING OF SDOH DATA In 2014, the IOM Committee on Recommended Social and Behavioral Domains and Measures for Electronic Health Records recommended the routine collection, in the electronic health record, of a parsimonious panel of clinically significant SDOH measures that may be obtained by self-report in advance of or during the health care encounter and, when used together, provide a psychosocial vital sign. The IOM-recommended questionnaire includes 25 items addressing the following domains: race and ethnicity, education, financial resource strain, stress, depression, physical activity, tobacco use, alcohol use, social connection or isolation, intimate partner violence, residential address, and geocoded census tract median income. Implementation studies have demonstrated that collection of these data takes about 5 minutes, and both patients and providers saw this data collection as appropriate and important. Given that data access and monitoring is an essential component to disparities elimination, we highlight several important sources of up-to-date racial/ethnic disparities monitoring initiatives that are available to the general public and are updated regularly. We highlight only three examples of national data sources.

- Since 2003, the Agency for Healthcare Research and Quality has led the yearly compilation of *The National Healthcare Quality and Disparities Report*, which reports trends for measures related to access to health care, affordable care, care coordination, healthy living, patient safety, and the quality of care across acute and chronic disease management by race/ethnicity, income, and other SDOH *(https://www.ahrq.gov/research/findings/nhqrdr/index.html)*.
- Since 2011, the Geospatial Research, Analysis, and Services Program (GRASP) created and maintains the CDC SVI. This database maps, for all U.S. Census tracts, 15 social factors (grouped in four SDOH categories: socioeconomic status, housing composition and disability, minority status and language, and housing and transportation) and is updated every 2 years *(https://www.atsdr.cdc.gov/placeandhealth/svi/index.html)*.
- *Launched in 2018, the* Health Opportunity and Equity (HOPE) Initiative benchmarks and tracks 27 indicators by race, ethnicity, and socioeconomic status. The indicators measure social and economic factors, community and safety, physical environment, access to health care, and health outcomes for the United States *(https://www.nationalcollaborative.org/our-programs/hope-initiative-project/)*.

INCREASE INSURANCE COVERAGE AND ACCESS Lack of access to high-quality health care is an important driver of racial/ethnic disparities. Signed into law in 2010, the Affordable Care Act (ACA) fundamentally transformed health insurance by decreasing the uninsured population from 16.3% in 2010 (~49.9 million) to 8.8.% in 2016 (~28.1 million) to 7.7% in early 2023. This represents the largest expansion of health insurance since the creation of Medicare and Medicaid in 1965. Prior to the ACA, non-Hispanic blacks were 70% and Hispanics nearly three times more likely to be uninsured than non-Hispanic whites. Of note, Medicaid expansion accounted for an estimated 60% of the ACA's effect through a combination of expanded eligibility and increased enrollment of previously eligible but unenrolled people. This is important given the higher number of racial/ethnic minorities who obtain insurance through Medicaid. Many studies have demonstrated that increased insurance coverage has also translated to greater improvement for blacks and Hispanics in access to care, more access to a usual source of care, and improved health outcomes. In 2023, as Medicaid returns to normal eligibility rules following the end of the COVID-19 public health emergency, many Americans are likely to transition out of Medicaid coverage with the concomitant risk of returning to high uninsurance rates in the United States among poor people and racial/ethnic minorities.

ENCOURAGEMENT OF THE USE OF EVIDENCE-BASED GUIDELINES AND QUALITY IMPROVEMENT *Unequal Treatment* highlights the subjectivity of clinical decision-making as a potential cause of racial and ethnic disparities in health care by describing how clinicians—despite the existence of well-delineated practice guidelines—may offer (consciously or unconsciously) different diagnostic and therapeutic options to different patients on the basis of their race or ethnicity. Therefore, the widespread adoption and implementation of evidence-based guidelines is a key recommendation in eliminating disparities. For instance, evidence-based guidelines are now available for the management of diabetes, HIV/AIDS, cardiovascular diseases, cancer screening and management, and asthma—all areas where significant disparities exist. As part of ongoing quality-improvement efforts, particular attention should be paid to the implementation of evidence-based guidelines for all patients, regardless of their race and ethnicity.

SUPPORT FOR THE USE OF LANGUAGE INTERPRETATION SERVICES IN THE CLINICAL SETTING As described previously, a lack of efficient and effective interpreter services in a health care system can lead to patient dissatisfaction, to poor comprehension and adherence, and thus to ineffective/lower-quality care for patients with limited English proficiency. *Unequal Treatment*'s recommendation to support the use of interpretation services has clear implications for delivery of quality health care by improving doctors' ability to communicate effectively with these patients.

INCREASES IN THE PROPORTION OF UNDERREPRESENTED MINORITIES IN THE HEALTH CARE WORKFORCE Data for 2021 from the Association of American Medical Colleges indicate that of active physicians, 63.9% identified as white, 6.9% identified as Hispanic, 5.7% identified as black or African American, and 0.3% identified as Native American or Alaskan Natives. Furthermore, U.S. national data show that only 3.6% of full-time faculty are black or African American, and 5.5% are Hispanic, Latino, or of Spanish origin (alone or in combination with another race/ethnicity), compared to 63.9% who identified as white. Longitudinal data demonstrate that minority faculty are more likely to be at or below the rank of assistant professor, whereas whites composed the highest proportion of full professors. Similarly, several studies have found that both Hispanic and black faculty were promoted at lower rates than their white counterparts. Despite representing ~30% of the U.S. population (a number projected to almost double by 2050), minority students are still underrepresented in medical schools. In 2018, matriculates to U.S. medical schools were 6.2% Latino, 7.1% African American, 0.1% Native Hawaiian or Other Pacific Islander, and 0.2% Native American or Alaskan Native. These percentages have decreased or remained nearly the same since 2007. It will be difficult to develop a diverse physician workforce that can meet the needs of an increasingly diverse population without dramatic changes in the racial and ethnic composition of medical student bodies. Long-term investment in pathway programs and the nearly universal adoption of holistic admissions (a process by which schools consider each applicant individually to determine how they might contribute to the learning environment and the workforce instead of relying just on test scores and grades) have produced modest results. Institutional change in medical schools, focused on creating nurturing, inclusive, and equity-focused environments that dismantle the structural racism that has created the opportunity gap faced by many minority students, is needed to address this important workforce challenge.

Provider Interventions • **INTEGRATION OF CROSS-CULTURAL EDUCATION INTO THE TRAINING OF ALL HEALTH CARE PROFESSIONALS** The goal of cross-cultural education is to improve providers' ability to understand, communicate with, and care for patients from diverse backgrounds. Such education focuses on enhancing awareness of sociocultural influences on health beliefs and behaviors and on building skills to facilitate understanding and management of these factors in the medical encounter. Cross-cultural education includes curricula on health care disparities, use of interpreters, and effective communication and negotiation across cultures. These curricula can be incorporated into health professions training in medical schools, residency programs, nursing schools, and other health professions programs, and can be offered as a component of continuing education. Despite the importance of this area of education and the attention it has attracted from medical education accreditation bodies, a national survey of senior resident physicians by Weissman and colleagues found that up to 28% felt unprepared to deal with cross-cultural issues, including caring for patients who have religious beliefs that may affect treatment, patients who use complementary medicine, patients who have health beliefs at odds with Western medicine, patients who mistrust the health care system, and new immigrants. In a study at one medical school, 70% of fourth-year students felt inadequately prepared to care for patients with limited English proficiency. Efforts to incorporate cross-cultural education into medical education will contribute to improving communication and to providing a better quality of care for all patients.

INCORPORATION OF TEACHING ON THE IMPACT OF RACE, ETHNICITY, AND CULTURE ON CLINICAL DECISION-MAKING *Unequal Treatment* and more recent studies found that stereotyping by health care providers can lead to disparate treatment based on a patient's race or ethnicity. The Liaison Committee on Medical Education, which accredits medical schools, issued a directive that medical education should include instruction on how a patient's race, ethnicity, and culture might unconsciously impact communication and clinical decision-making.

Patient Interventions Difficulty navigating the health care system and obtaining access to care can be a hindrance to all populations, particularly to minorities. Similarly, lack of empowerment or involvement in the medical encounter by minorities can be a barrier to care. Patients need to be educated on how to navigate the health care system and how best to access care. Interventions should be used to increase patients' participation in treatment decisions.

General Recommendations • **INCREASE AWARENESS OF RACIAL/ETHNIC DISPARITIES IN HEALTH CARE** Efforts to raise awareness of racial/ethnic health care disparities have done little for the general public but have been fairly successful among physicians, according to a Kaiser Family Foundation report. In 2006, nearly 6 in 10 people surveyed believed that blacks received the same quality of care as whites, and 5 in 10 believed that Latinos received the same quality of care as whites. These estimates are similar to findings in a 1999 survey. Despite this lack of awareness, most people believed that all Americans deserve quality care, regardless of their background. In contrast, the level of awareness among physicians has risen sharply. In 2002, the majority (69%) of physicians said that the health care system "rarely or never" treated people unfairly on the basis of their racial/ethnic background. In 2005, less than one-quarter (24%) of physicians disagreed with the statement that "minority patients generally receive lower-quality care than white patients." More recently, a survey by WebMD showed that 42% of 16,000 physicians admitted that their own personal biases impact their clinical decision-making, including on characteristics such as race and ethnicity. Increasing awareness of racial and ethnic health disparities, and their root causes, among health care professionals and the public is an important first step in addressing these disparities. The ultimate goals are to generate discourse and to mobilize action to address disparities at multiple levels, including health policymakers, health systems, and the community.

CONDUCT FURTHER RESEARCH TO IDENTIFY SOURCES OF DISPARITIES AND PROMISING INTERVENTIONS While the literature that formed the basis for the findings reported and recommendations made in *Unequal Treatment* provided significant evidence for racial and ethnic disparities, additional research is needed in several areas. First, most of the literature on disparities focuses on black-versus-white differences; much less is known about the experiences of other minority groups. Improving the ability to collect racial and ethnic patient data should facilitate this process. However, in instances where the necessary systems are not yet in place, racial and ethnic patient data may be collected prospectively in the setting of clinical or health services research to more fully elucidate disparities for other populations. Second, much of the literature on disparities to date has focused on defining areas in which these disparities exist, but less has been done to identify the multiple factors that contribute to the disparities or to test interventions to address these factors. There is clearly a need for research that identifies promising practices and solutions to disparities.

■ IMPLICATIONS FOR CLINICAL PRACTICE

Individual health care providers can do several things in the clinical encounter to address racial and ethnic disparities in health care.

Be Aware That Disparities Exist Increasing awareness of racial and ethnic disparities among health care professionals is an important first step in addressing disparities in health care. Only with greater awareness can care providers be attuned to their behavior in clinical practice and thus monitor that behavior and ensure that all patients receive the highest quality of care, regardless of race, ethnicity, or culture.

Practice Culturally Competent Care Previous efforts have been made to teach clinicians about the attitudes, values, beliefs, and behaviors of certain cultural groups—the key practice "dos and don'ts" in caring for "the Hispanic patient" or the "Asian patient," for example. In certain situations, learning about a particular local community or cultural group, with a goal of following the principles of community-oriented primary care, can be helpful; when broadly and uncritically applied, however, this approach can actually lead to stereotyping and oversimplification of culture, without respect for its complexity.

Cultural competence has thus evolved from merely learning information and making assumptions about patients on the basis of their backgrounds to focusing on the development of skills that follow the principles of patient-centered care. *Patient-centeredness* encompasses the qualities of compassion, empathy, and responsiveness to the needs, values, and expressed preferences of the individual patient. *Cultural competence* aims to take things a step further by expanding the repertoire of knowledge and skills classically defined as "patient-centered" to include those that are especially useful in cross-cultural interactions (and that, in fact, are vital in all clinical encounters). This repertoire includes effectively using interpreter services, eliciting the patient's understanding of their condition, assessing decision-making preferences and the role of family, determining the patient's views about biomedicine versus complementary and alternative medicine, recognizing sexual and gender issues, and building trust. For example, while it is important to understand all patients' beliefs about health, it may be particularly crucial to understand the health beliefs of patients who come from a different culture or have a different health care experience. With the individual patient as teacher, the physician can adjust their practice style to meet the patient's specific needs.

Avoid Stereotyping Several strategies can allow health care providers to counteract, both systemically and individually, the normal tendency to stereotype. For example, when racially/ethnically/culturally/socially diverse teams in which each member is given equal power are assembled and are tasked to achieve a common goal, a sense of camaraderie develops and prevents the development of stereotypes based on race/ethnicity, gender, culture, or class. Thus, health care providers should aim to gain experiences working with and learning from a diverse set of colleagues. In addition, simply being aware of the operation of social cognitive factors allows providers to actively check up on or monitor their behavior. Physicians can constantly reevaluate to ensure that they are offering the same things, in the same ways, to all patients. Understanding one's own susceptibility to stereotyping—and how disparities may result—is essential in providing equitable, high-quality care to all patients.

Work to Build Trust Patients' mistrust of the health care system and of health care providers impacts multiple facets of the medical encounter, with effects ranging from decreased patient satisfaction to delayed care. Although the historic legacy of discrimination can never be erased, several steps can be taken to build trust with patients and to address disparities. First, providers must be aware that mistrust exists and is more prevalent among minority populations, given the history of discrimination in the United States and other countries. Second, providers must reassure patients that they come first, that everything possible will be done to ensure that they always get the best care available, and that their caregivers will serve as their advocates. Third, interpersonal skills and communication techniques that demonstrate honesty, openness, compassion, and respect on the part of the health care provider are essential tools in dismantling mistrust. Finally, patients indicate that trust is built when there is shared, participatory decision-making and the provider makes a concerted effort to understand the patient's background. When the doctor–patient relationship is reframed as one of solidarity, the patient's sense of vulnerability can be transformed into one of trust. The successful elimination of disparities requires trust-building interventions and strengthening of this relationship.

■ CONCLUSION

The issue of racial and ethnic disparities in health care has gained national prominence, both with the release of the IOM report *Unequal*

Treatment and with more recent articles that have confirmed their persistence and explored their root causes. Furthermore, another influential IOM report, *Crossing the Quality Chasm*, has highlighted the importance of equity—i.e., no variations in quality of care due to personal characteristics, including race and ethnicity—as a central principle of quality. Current efforts in health care reform and transformation, including a greater focus on value (high-quality care and cost control), will sharpen the nation's focus on the care of populations who experience low-quality, costly care. Addressing disparities will become a major focus, and there will be many obvious opportunities for interventions to eliminate them. Greater attention to addressing the root causes of disparities will improve the care provided to all patients, not just those who belong to racial and ethnic minorities.

FURTHER READING

Buchmueller TC et al: The ACA's impact on racial and ethnic disparities in health insurance coverage and access to care. Health Aff (Millwood) 39:395, 2020.

Dwyer-Lindgren L et al: Inequalities in life expectancy among us counties, 1980 to 2014: Temporal trends and key drivers. JAMA Intern Med 177:1003, 2017.

Kreuter MW et al: Addressing social needs in health care settings: Evidence, challenges and opportunities for public health. Annu Rev Public Health 42:11, 2021.

Krieger N: Measures of racism, sexism, heterosexism, and gender binarism for health equity research: From structural injustice to embodied harm: An ecosocial analysis. Annu Rev Public Health 41:37, 2020.

LaVeist TA et al: The economic burden of racial, ethnic, and educational health inequities in the US. JAMA 329:1682, 2023.

Medscape: Medscape Lifestyle Report 2016: Bias and burnout. *http://www.medscape.com/features/slideshow/lifestyle/2016/public/overview*.

Vyas DA et al: Hidden in plain sight: Reconsidering the use of race correction in clinical algorithms. N Engl J Med 383:874, 2020.

Williams DR et al: Racism and health: Evidence and needed research. Annu Rev Public Health 40:105, 2019.

12

Ethical Issues in Clinical Medicine

Christine Grady, Bernard Lo

Physicians face both enduring and novel ethical dilemmas that can be perplexing and emotionally draining and can sometimes lead to moral distress. For example, physicians may worry about the intensity of care that is appropriate at the end of life, how to distribute time between critically ill patients, or their responsibilities for patients with limited access to needed care. Newer technologies such as artificial intelligence promise more coordinated and efficient care, but also raise concerns about confidentiality, the doctor–patient relationship, and accountability. New social and economic developments, such as greater attention to health disparities and the growth of for-profit enterprises, present additional issues. This chapter suggests how physicians can address important vexing ethical issues they encounter in their work. Physicians make ethical judgments about clinical situations every day and should prepare for lifelong learning about ethical issues so they can respond appropriately. Traditional professional codes and ethical principles provide instructive guidance but need to be interpreted and applied to each situation. When facing a challenging ethical issue, physicians may need to reevaluate their basic convictions, tolerate uncertainty, and act in a manner that maintains their integrity. Physicians should articulate their concerns and reasoning, listen to and respect the views of others involved in the patient's care, and utilize available resources, including members of the interdisciplinary health care team as well as ethics consultants in some cases. Ethics consultation services or a hospital ethics committee can help to clarify issues and identify strategies for resolution, including improving communication and dealing with strong or conflicting emotions. Through these efforts, physicians can gain deeper insight into the ethical issues they face and reach mutually acceptable resolutions to complex problems.

APPROACHES TO ETHICAL PROBLEMS

Several approaches are useful for addressing ethical issues, including approaches based on ethical principles, virtue ethics, professional oaths, and personal values. Sometimes these various sources of guidance may seem to conflict. In a diverse society, different individuals may turn to different sources of moral guidance. In addition, general moral precepts often need to be interpreted and applied to particular clinical situations.

ETHICAL PRINCIPLES

Ethical principles serve as general guidelines to help physicians determine the right thing to do.

Respecting Patients Physicians should always treat patients with respect, which entails understanding patients' goals, providing information, communicating effectively, obtaining informed and voluntary consent, respecting informed refusals, and protecting confidentiality. Different clinical goals and approaches are often feasible, and interventions can result in both benefit and harm. Individuals differ in how they value health and medical care and weigh the benefits and risks of medical interventions. Generally, physicians should respect patients' values and informed choices. Respect is especially important when patients are responding to experiences of, or concerns about, disrespect and discrimination. Physicians should consider patient confidentiality, professional boundaries, and therapeutic relationships in all spoken and written communications.

GOALS AND TREATMENT DECISIONS Physicians should provide relevant and accurate information for patients about current clinical circumstances, diagnoses, expected future course, prognosis, treatment options, and uncertainties, and discuss how different options might realize patients' goals of care. Physicians may be tempted to withhold a serious diagnosis, misrepresent it by using ambiguous terms, or limit discussions of prognosis or risks for fear that patients will become anxious or depressed. However, providing honest information about clinical situations promotes patients' autonomy and trust. When sharing bad news with patients, physicians should adjust the pace of disclosure, offer empathy and hope, provide emotional support, and call on other resources such as spiritual care or social work. Some patients may choose not to receive such information or may ask surrogates to make decisions on their behalf, as is common with serious diagnoses in some cultures.

SHARED DECISION-MAKING AND OBTAINING INFORMED CONSENT Physicians should engage in shared decision-making with patients, whenever appropriate. Physicians promote shared decision-making by informing and educating patients, eliciting and answering their questions, checking that they understand key issues, making recommendations, and helping them to deliberate. Medical jargon, needlessly complicated explanations, or the provision of too much information at once may overwhelm patients. Increasingly, decision aids can help patients play a more active role in decision-making, improve the accuracy of their perception of risk and benefit, and help them clarify their values and goals. Informed consent is more than obtaining signatures on consent forms and involves disclosure of honest and understandable information to promote understanding and choice. Competent, informed patients may refuse recommended interventions and choose among reasonable alternatives. In an emergency, treatment can be given without informed consent if patients cannot give their own consent and delaying treatment while surrogates are contacted would jeopardize patients' lives or health. People are presumed to want such emergency care unless they have previously indicated otherwise, such as through a portable order for life-sustaining treatment (POLST).

Respect for patients does not entitle patients to insist on any care or treatment that they want. Physicians are not obligated to provide interventions that have no physiologic rationale, have already failed, or are contrary to evidence-based practice recommendations or good clinical judgment. Public policies and laws also dictate certain decisions—e.g., allocation of scarce medical resources during public health crises such as the COVID-19 pandemic, use of cadaveric organs for transplantation, and responding to requests for physician aid in dying.

CARING FOR PATIENTS WHO LACK DECISION-MAKING CAPACITY Patients with decision-making capacity can express choices and appreciate their medical situation; the nature, risks, and benefits of proposed care; and the consequences of each alternative. Patient choices should be consistent with their values and not the result of delusions, hallucinations, or misinformation. Physicians should use available and validated assessment tools, resources such as psychiatry or ethics consultation, and clinical judgment to ascertain whether individuals have the capacity to make decisions for themselves. Some patients are unable to make informed decisions because of unconsciousness, advanced dementia, delirium, or other medical conditions that impair cognition. Courts have the legal authority to determine that a patient is legally incompetent, but physicians usually determine when patients lack the capacity to make particular health care decisions and arrange for authorized surrogates to make decisions, without involving the courts. Patients who disagree with recommendations or refuse treatment should not be assumed to lack capacity. Such decisions should be probed, however, to ensure the patient is not deciding based on misunderstandings and has the capacity to make an informed decision. When impairments are fluctuating or reversible, decisions should be postponed if possible and revisited when the patient recovers decision-making capacity.

Physicians seek an appropriate surrogate for patients who lack decision-making capacity. Patients may designate a health care proxy through a POLST or an advance directive; such choices should be respected **(see Chap. 13)**. For patients who lack decision-making capacity and have not previously designated a health care proxy, family members usually serve as surrogate decision-makers. Statutes in most U.S. states delineate a priority order of relatives to make medical decisions. For unrepresented patients who do not have relatives or friends who can be found and no previously appointed proxy, making decisions presents difficult dilemmas and should follow a fair and careful institutional or legal process. Patients' values, goals, and previously expressed preferences, when known, should guide surrogate decisions. However, the patient's current best interests may sometimes justify overriding earlier preferences if an intervention is likely to provide significant benefit, previous statements do not fit the situation well, or the patient gave the surrogate leeway in decisions.

MAINTAINING CONFIDENTIALITY Maintaining confidentiality is essential to respecting patients' autonomy and privacy; it encourages patients to seek treatment and discuss problems candidly. However, confidentiality may be overridden to prevent serious harm to third parties or to the patient. Exceptions to confidentiality are justified when the risk is serious and probable, no less restrictive measures can avert risk, and the adverse effects of overriding confidentiality are minimized and deemed acceptable by society. For example, laws require physicians to report cases of tuberculosis, sexually transmitted infections, elder or child abuse, and domestic violence.

Beneficence or Acting in Patients' Best Interests The principle of *beneficence* requires physicians to act for the patient's benefit. Patients typically lack medical expertise, and illness often makes them vulnerable. Patients rely on and trust physicians to treat them with compassion and provide sound recommendations and treatments aimed to promote their well-being. Physicians encourage such trust and have a fiduciary duty to act in the best interests of patients, which should prevail over physicians' self-interest or the interests of third parties such as hospitals or insurers. A related principle, "first do no harm," obliges physicians to prevent unnecessary harm by recommending interventions that maximize benefit and minimize harm and forbids physicians from providing known ineffective interventions or acting without due care. Although often cited, this precept alone provides limited guidance because many beneficial interventions also pose serious risks.

Physicians often provide care within interdisciplinary teams and rely on consultation with or referral to specialists. Team members and consultants contribute different expertise to the provision of comprehensive, high-quality care for patients. Physicians should collaborate with and respect the contributions of interdisciplinary team members and should initiate and participate in regular communication and planning to avoid diffusion of responsibility and ensure accountability for quality patient care.

INFLUENCES ON PATIENTS' BEST INTERESTS Conflicts arise when patients' refusals or requests for interventions thwart their own goals for care, cause serious harm, or jeopardize their best medical interests. For example, simply accepting a young asthmatic adult's refusal of mechanical ventilation for readily reversible respiratory failure, in the name of respecting autonomy, is morally constricted. Physicians should elicit patients' expectations and concerns, correct their misunderstandings, and try to persuade them to accept beneficial therapies. If disagreements persist after such efforts, physicians should call on institutional resources for assistance, but patients' informed choices and views of their own best interests should prevail.

Drug prices and out-of-pocket expenses for patients have escalated in many parts of the world and may compromise care in the patients' best interests. Physicians should recognize that patients, especially those with high copayments or inadequate insurance, may not be able to afford prescribed tests and interventions. Physicians should strive to prescribe indicated medications that are affordable and acceptable to the patient. Knowing what kind of insurance, if any, the patient has and whether certain medications are likely to be covered may help in determining appropriate prescriptions. Physicians or their team should follow up with patients who don't fill prescriptions, don't take their medications, or skip doses to explore whether cost and affordability are obstacles. Clinical pharmacists, technicians, or social workers can help the prescribing physician appeal for coverage or find funding for needed medications. It may be reasonable for physicians to advocate for coverage of nonformulary products for sound reasons, such as when the formulary drugs are ineffective or not tolerated.

Organizational policies and workplace conditions may sometimes conflict with patients' best interests. Physicians' focus and dedication to the well-being and interests of patients may be negatively influenced by perceived or actual staffing inadequacies, unfair wages, deficiencies in equipment or infrastructure, work-hour limitations, corporate culture, violence, and other risks in the workplace. Physicians should work with institutional leaders to ensure that policies and practices support their ability to provide quality care focused on patients' best interests.

Patients' interests are served by the increasing use of evidence-based practice guidelines and performance benchmarking. However, practice guideline recommendations may not serve the interests of each individual patient when another plan of care would provide substantially greater benefits. In prioritizing their duty to act in the patient's best interests, physicians should be familiar with relevant practice guidelines, be able to recognize situations that might justify exceptions, and advocate for reasonable exceptions.

Acting Justly The principle of *justice* provides guidance to physicians about how to ethically treat patients and make decisions about allocating important resources, including their own time. *Justice* in a general sense means fairness: people should receive what they deserve or are owed. Justice also entails similar treatment of people who are similar in ethically relevant ways, in order to avoid arbitrary, biased, and unfair decisions. Justice forbids discrimination in health care based on race, religion, gender, sexual orientation, disability, age, or other personal characteristics **(Chap. 11)**. Distributive justice requires fair and equitable allocation of limited health care resources, which should be distributed equally or according to relevant moral considerations such as need, probability of benefit, or other factors.

ALLOCATION OF RESOURCES Universal access to medically needed health care remains an unrealized moral aspiration in the United States and around the world. Patients with no or inadequate health insurance, especially those with chronic diseases, often cannot afford needed care and lack access to safety-net services. Even among insured patients, insurers sometimes deny coverage for interventions recommended by their physician. Physicians should advocate for patients' affordable access to indicated care, try to help patients obtain needed care, and work with institutions and policies to promote wider access.

Allocation of health care resources is unavoidable when resources are limited, such as during medication shortages due to manufacturing and quality problems or during public health crises. Allocation policies should be fair, transparent, accountable, responsive to the concerns of those affected, and proportionate to the situation, including the supply relative to the need. Difficult decisions about allocation of available scarce resources are inevitable. A fair process should aim to prioritize patients who are likely to have a large and irreversible decrement in health outcomes without the interventions; avoid discrimination; mitigate health disparities; save the most lives; and promote the good of the community.

First-come, first-served allocation is often not fair, because it disadvantages patients who experience barriers to accessing care. To avoid discrimination, allocation decisions should not consider personal social characteristics such as race, gender, or disability, nor consider insurance status or wealth. Allocation policies should aspire to reduce health care disparities often due to social determinants of health including poverty, unhealthy living conditions, and poor access to health care.

Fair and well-considered guidelines help mitigate emotional and moral distress for clinicians and patients when difficult allocation decisions are made. Physicians should act as patient advocates within constraints set by society, reasonable insurance policies, and evidence-based practice. Ad hoc resource allocation by individual physicians may be inconsistent, unfair, and ineffective. Many allocation decisions are made at the institutional or public policy level with physician and public input, for example, decisions and criteria for allocating cadaveric organs for transplantation.

■ VIRTUE ETHICS

Virtue ethics focuses on physicians' character and qualities, with the expectation that doctors will cultivate virtues such as compassion, trustworthiness, intellectual honesty, humility, and integrity. Proponents argue that, if such characteristics become ingrained, they help guide physicians in unforeseen situations. Moreover, simply following ethical precepts or principles without any of these virtues could lead to uncaring doctor–patient relationships.

■ PROFESSIONAL OATHS AND CODES

Professional oaths and codes are useful guides for physicians. Most physicians take oaths during their medical training, and many are members of professional societies with codes of ethics. Physicians pledge to the public and to their patients that they will be guided by the principles and values in these oaths or codes and commit to the spirit of the ethical ideals and precepts represented in oaths and professional codes of ethics.

■ PERSONAL VALUES

Personal values, cultural traditions, and religious beliefs are important sources of personal morality that help physicians address ethical issues and cope with any moral distress they may experience in practice. While essential, personal morality alone is a limited ethical guide in clinical practice. Physicians have role-specific ethical obligations that go beyond their obligations as good people, including the duties to obtain informed consent and maintain confidentiality discussed earlier. Furthermore, in a culturally and religiously diverse world, physicians should expect that many patients and colleagues will have personal moral beliefs that differ from their own.

ETHICALLY COMPLEX PROFESSIONAL ISSUES FOR PHYSICIANS

■ CLAIMS OF CONSCIENCE

Some physicians, based on personal values, have legally protected conscientious objections to providing, or referring patients for, certain treatments such as contraception, abortion, or physician aid in dying. Although physicians should not be asked to violate deeply held moral beliefs or religious convictions, patients need medically appropriate and timely care, rely on physicians for medical advice, and should always be treated with respect. Institutions such as clinics and hospitals have a collective ethical duty to provide medical care that will avert serious risks and suffering to patients while accommodating health care workers' conscientious objections—for example, when possible, by arranging for another professional to provide the service in question or by referral to another institution. Ethically, patients seeking a relationship with a doctor or health care institution should be notified in advance of any conscientious objections to the provision of specific interventions. Since insurance often constrains patients' selection of physicians or health care facilities, switching providers can be problematic after an illness is recognized. There are also important limits on claims of conscience. Health care workers may not insist that patients receive unwanted medical interventions or refuse to treat or illegally discriminate against patients because of their race, ethnicity, disability, genetic information, or diagnosis. Refusal to treat patients because of personal characteristics, such as sexual orientation, gender identity, or immigration status is ethically inappropriate because it falls short of helping patients in need and respecting them as persons.

Importantly, some physicians may object to laws, institutional policies, or insurance policies that restrict them from providing or even describing medically appropriate care to their patients. In some of these cases, workarounds are justified on the ethical basis of sound clinical practice.

■ PHYSICIAN AS GATEKEEPER

In some cases, patients may ask their physicians to facilitate access to services that the physician has ethical and clinical qualms about providing. For example, a patient might request a prescription for a cognitive enhancing medication to temporarily augment cognitive abilities in order to take an exam or apply for employment, or request that the physician sign a disability form when the patient does not meet disability criteria. Patients may request more pain medication than the physician believes is warranted for the given situation, or the physician may feel uncomfortable prescribing attention-deficit/hyperactivity disorder medications to a young child because of concerns that the possible benefit does not justify the risks despite the parent's request. In these and similar circumstances, the physician should work with the patient or parent to understand reasons for their requests, some of which might be legitimate. In addition to considering possible risks and benefits to the patient, the physician should consider how meeting the request might affect other patients, societal values, and public trust in the medical profession. If the physician determines that fulfilling the request requires deception, is unfair, jeopardizes professional responsibilities, could undermine the physician's credibility or trustworthiness, or is inconsistent with the patient's best medical interests, the physician should decline and explain the reasons to the patient.

■ OCCUPATIONAL RISKS AND BURDENS

Physicians accept some risks in fulfilling their professional responsibilities, including exposure to infectious agents or toxic substances, abuse from patients or families, distress, and violence or threats in the workplace. Nonetheless, most physicians, nurses, and other hospital staff willingly care for patients, despite personal risk, grueling hours, and sometimes inadequate resources or information. Health care institutions are responsible for preventing and addressing occupational risks and burdens and for providing proper information, training and supervision, resources, infrastructure and workflow modifications, and emotional and psychological support. Clinical leaders should acknowledge fears about personal safety and take steps to mitigate the impact of work on family responsibilities, moral distress, and burnout.

MORAL DISTRESS AND WELL-BEING

Clinicians, including residents, medical students, and experienced physicians, experience moral distress when they feel that ethically appropriate action is hindered by institutional policies or culture, decision-making hierarchies, limited resources, or other reasons. Moral distress can lead to anger, anxiety, depression, frustration, fatigue, work dissatisfaction, and burnout. A physician's health and well-being can affect how they care for patients and interact with colleagues. Discussing complex or unfamiliar clinical situations with colleagues and seeking assistance with difficult decisions can help mitigate moral distress, as can a healthy work environment characterized by open communication, mutual respect, adequate support and resources, and an emphasis on the common goal of good patient care. In addition, physicians should take good care of their own health and well-being and be aware of the personal and system factors associated with stress, burnout, and depression. Health care organizations should provide a supportive work environment, counseling, and other support services as needed.

CONFLICTS OF INTEREST

Acting in patients' best interests may sometimes conflict with a physician's self-interest or the interests of third parties such as insurers or hospitals. From an ethical viewpoint, patients' interests are paramount. Transparency, appropriate disclosure, and management of conflicts of interest are essential to maintain the trust of patients, colleagues, and the public. Disclosure requirements vary for different purposes, and available software can assist physicians in complying with specific requirements. Importantly, not all conflicts are financial. Physicians sometimes face conflicts of commitment between their patient's interests and their own personal interests, professional goals, responsibilities, and aspirations or the interests of the health care institutions or corporations that employ them. Physicians should prioritize patients' interests while recognizing possible conflicts and use disclosure, discussion with institutional officials, and management of the conflict or recusal when appropriate.

Medical institutions themselves may have conflicts of interest arising from patent rights, industry-funded research programs, donations from individuals and companies, and priorities of the health care system, pharmacy benefits managers, insurers, and private equity or corporate owners. Institutions should be transparent about the presence and amount of such relationships and about the steps taken to prevent such relationships from influencing clinical or financial decisions. Institutions should take steps not to benefit from donations or honor donors when there is good evidence that a donor breached ethical or legal standards.

FINANCIAL INCENTIVES

Physicians have financial incentives to improve the quality or efficiency of care that might lead some to avoid patients who are older, chronically ill, or have more complicated problems, or to focus on benchmarked outcomes even when not in individual patients' best interests. In contrast, fee-for-service payments might encourage physicians to order more interventions than necessary or to refer patients to laboratory, imaging, or surgical facilities in which they have a financial stake. Regardless of financial incentives, physicians should recommend available care that is in the patient's best interests—no more and no less.

RELATIONSHIPS WITH PHARMACEUTICAL COMPANIES

Financial relationships between physicians and industry are increasingly scrutinized. Many medical centers have banned drug-company gifts, including branded pens or notepads or meals to physicians, to reduce risk of undue influence or subconscious feelings of reciprocity and to decrease possible influences on public trust or the costs of health care.

The federal Open Payments website provides public information on the payments and amounts that drug and device companies give to individual physicians. The challenge is to distinguish payments for scientific consulting and research contracts—which should be encouraged as promoting professional and academic missions—from those for promotional speaking and other activities that aim to increase sales of company products.

LEARNING CLINICAL SKILLS

Medical students', residents', and physicians' interests in learning, which fosters the long-term goal of benefiting future patients, may sometimes conflict with the short-term goal of providing optimal care to current patients. When trainees are learning procedures on patients, they lack the proficiency of experienced physicians, and patients may experience inconvenience, discomfort, longer procedures, or increased risk. Increasingly, institutions are developing simulation-based clinical skills laboratories and requiring students to demonstrate proficiency before carrying out procedures such as venipuncture and intravenous lines in patients. Furthermore, teaching hospitals are establishing proceduralist services in which procedure-specialist faculty members directly supervise interns for procedures such as lumbar puncture and thoracentesis and certify their proficiency. Medical students may need to defer learning such invasive procedures until internship. It is always important to seek patients' consent for trainee participation in their care, particularly for intimate examinations, such as pelvic, rectal, breast, and testicular examinations, and for invasive procedures. Patients should be told who is providing care and how trainees are supervised. Failing to introduce students or trainees who will be performing procedures undermines trust, may lead to more elaborate deception, and hinders patients' informed choices about their care. Most patients, when informed, allow trainees to play an active role in their care.

RESPONSE TO MEDICAL ERRORS

Errors are inevitable in clinical medicine, and some errors cause harm to patients. Most errors are caused by lapses of attention or flaws in the system of delivering health care; only a small number result from blameworthy individual behavior. Many health care institutions have adopted a communication and resolution system that encourages open and honest reporting of errors as essential to quality of care and shifts the focus from individual blame to system redesign for improvement in quality and safety **(Chap. 7)**. This approach is more likely than a punitive approach to improve patient safety. However, professional discipline is appropriate for cases of gross incompetence, reckless behavior, physician impairment, and boundary violations. Physicians and students may fear that disclosing errors will damage their careers. Physicians and health care institutions show respect for patients by disclosing and explaining errors, offering apologies, offering appropriate compensation for harm done, and using errors as opportunities to improve the quality of care.

PHYSICIAN IMPAIRMENT

Physicians may hesitate to intervene when colleagues impaired by alcohol, drugs, or psychiatric or medical illness place patients at risk. However, society relies on physicians to regulate themselves. Colleagues of an impaired physician should take steps to protect patients and help their impaired colleague, starting with reporting their concerns to their clinical supervisor or director, and institutions should have programs in place for assessing physician competence.

ETHICAL ISSUES IN CLINICAL RESEARCH

Clinical research is essential to translate scientific discoveries into beneficial interventions for patients. However, clinical research raises ethical concerns because participants face inconvenience and risks in research designed to advance scientific knowledge and not specifically to benefit them. Ethical guidelines require researchers to rigorously design and conduct research, minimize risk to participants, and obtain informed and voluntary participant consent and approval from an institutional review board (IRB). IRBs determine whether risks to participants are acceptable and adequately minimized, and recommend appropriate additional safeguards when research includes vulnerable participants.

Physicians may be clinical research investigators themselves or in a position to refer or recommend research participation to their patients. Physician-investigators are likely to feel some inherent tension between conducting research and providing health care. Awareness of this tension, familiarity with research ethics, collaboration with research and clinical team members, and utilizing research ethics consultation can help to mitigate tensions. Before starting clinical research, investigators should complete training in the ethics of clinical research, which is widely available. Physicians also should be critical consumers of clinical research results and keep up with research advances that change standards of practice.

Forms of research that generate knowledge useful for improving health might pose concerns about informed consent, privacy and confidentiality, and psychosocial risk without physical risks. Research with routinely collected clinical data available in electronic health records, leftover clinical specimens, data from mobile devices, administrative data, and combinations could encompass traditional discovery research as well as quality improvement, comparative effectiveness research, and learning health care systems.

EMERGING TECHNOLOGIES

Scientific advances in genome sequencing, gene editing (e.g., with CRISPR-Cas9), machine learning, artificial intelligence, brain–computer interfaces, cell-based therapies, and other technologies offer great promise for research and clinical care with the ultimate goal of improving the prediction, prevention, and treatment of disease. Groundbreaking innovations that have strong scientific plausibility still need to be evaluated in rigorous clinical studies for efficacy and safety.

Physicians should keep up to date on the status of novel and often complex technologies as research evolves, data emerge, and technologies are incorporated into clinical practice. They can help their patients understand research findings and the evidence for clinical use, correct any misunderstandings, facilitate shared decision-making, and advocate for fair access to such therapies. Furthermore, physicians should engage in professional and public discussion related to allocation of resources and fair access to expensive new therapies and emerging technologies and their impact on overall health care affordability.

Certain cell-based therapies, such as peripheral blood stem cell transplantation **(Chap. 119)** and chimeric antigen receptor (CAR)-T cell therapy **(Chap. 78)**, are approved for use in several serious hematologic disorders. Somatic cell gene therapies have been approved as safe and effective for clinical use in β thalassemia, hemophilia A, sickle cell disease, and some rare pediatric-onset diseases. Numerous other gene therapy applications are under review by regulatory agencies. The approved therapies have a very high cost, which often limits access. Germline gene editing in blastocysts or embryos raises many ethical questions and is not permitted in the United States and many other countries.

Some patients may request cell-based, gene, or other complex, highly technical, and expensive therapies for unproven indications. Yet, claims of cures through unproven stem cell or gene-based “therapies” pose significant health and financial risks to patients without evidence of benefit. Physicians should help patients distinguish approved therapies from unproven claims and refer interested patients to well-designed clinical trials.

Artificial intelligence (AI) using computers to carry out tasks typically done by humans, and machine learning (ML), a type of AI that automatically learns and improves its performance without explicit programming, are increasingly used in clinical practice. AI clinical algorithms are used to make diagnoses using images from radiology studies, retinal scans, pathology slides, or skin photographs; to search for potential diagnoses associated with a genomic sequence; and to predict outcomes for patients, such as surgical complications, hospital readmissions, or not coming to scheduled appointments. Physicians should use these predictive algorithms when they are confident that they are clinically meaningful, are unbiased, and do not lead to worse patient outcomes. Generative AI, large language models (LLMs), or chatbots **(Chap. 501)** promise to help reduce burdensome administrative tasks that contribute to health care worker burnout, such as drafting clinical notes and summaries for patients and writing letters to obtain authorization or appeal denials of insurance coverage. Although rapidly evolving in sophistication, LLMs can sometimes “hallucinate” and provide false information. Physicians should be confident that the LLM or bots they use are providing medically accurate information and advice. Collaboration between AI developers and clinicians can utilize the strengths of each. The impact of AI on clinician workflow and satisfaction, as well as patient outcomes, needs to be assessed.

Physicians should stay informed of emerging evidence about such technologies and the ethical challenges that accompany their use and always keep their patients’ best interests and informed preferences at the forefront.

GLOBAL CONSIDERATIONS

INTERNATIONAL RESEARCH

Clinical research is often conducted across multiple sites and across national borders. Societal, legal, and cultural norms and perspectives about research may vary, and there are many ethical challenges. Physician-investigators involved in international research should be familiar with federal regulations, applicable international regulatory and good clinical practice guidelines, as well as national and local laws where research is taking place. Partnering with local researchers and communities is essential not only to demonstrate respect but also to facilitate successful clinical research.

INTERNATIONAL CLINICAL EXPERIENCES

Providing patient care in international settings can be a humanitarian service and/or broaden clinical experience. Such arrangements, however, raise ethical challenges—for example, as a result of differences in beliefs about health and illness, expectations regarding health care and physicians’ roles, standards of clinical practice, resource limitations, and cultural norms regarding disclosure and other issues. Visiting physicians and trainees should receive training and mentoring about cultural and clinical practices in the host community, respect local customs and values, collaborate closely with local professionals and staff, and recognize their own limits.

CONCLUSION

Ethical issues are common in clinical medicine and occur in circumstances that may be foreseeable, novel, or unexpected. Physicians address these ethical issues by being prepared, informed, and reflective, and using appropriate available resources.

FURTHER READING

Beauchamp T, Childress J: *Principles of Biomedical Ethics*, 8th ed. New York, Oxford University Press, 2019.

Berwick D: *Salve Lucrum*: The existential threat of greed in US health care. JAMA 329:629, 2023.

Harris L: Divisions, new and old: Conscience and religious freedom at HHS. N Engl J Med 378:1369, 2018.

Hightower M et al: Is medicine ready for AI: Intention to treat episode 6. N Engl J Med 388:e49, 2023.

National Academy of Medicine. Action Collaborative on Clinician Wellbeing and Resilience. *https://nam.edu/initiatives/clinician-resilience-and-well-being/*.

13

Palliative and End-of-Life Care

Ezekiel J. Emanuel, John W. Urwin

EPIDEMIOLOGY

CAUSES OF DEATH

In 2019, before the COVID-19 pandemic, 2,854,838 individuals died in the United States **(Table 13-1)**. Approximately 74% of these deaths occurred in those aged ≥65 years. The epidemiology of death has changed significantly since 1900 and even since 1980. In 1900, heart

TABLE 13-1 Ten Leading Causes of Death in the United States and Britain

CAUSE OF DEATH	UNITED STATES (2019) NUMBER OF DEATHS, ALL AGES (%)	UNITED STATES (2019) NUMBER OF DEATHS, PEOPLE ≥65 YEARS OF AGE	ENGLAND AND WALES (2019) NUMBER OF DEATHS, ALL AGES (%)	ENGLAND AND WALES (2019) NUMBER OF DEATHS, PEOPLE ≥65 YEARS OF AGE
All deaths	2,854,838	2,117,332	530,841	449,047
Heart disease[a]	659,041 (23.1)	531,583 (25.1)	87,095 (16.4)	74,967 (16.7)
Malignant neoplasms	599,601 (21.0)	435,462 (20.6)	147,419 (27.8)	118,982 (26.5)
Chronic lower respiratory diseases	156,979 (5.5)	133,246 (6.3)	31,221 (5.9)	28,235 (6.3)
Accidents	173,040 (6.1)	60,527 (2.9)	15,141 (2.9)	8999 (2.0)
Cerebrovascular diseases	150,005 (5.3)	129,193 (6.1)	29,816 (5.6)	27,210 (6.0)
Alzheimer's disease	121,499 (4.3)	120,090 (5.7)	20,400 (3.8)	20,279 (4.5)
Diabetes mellitus	87,647 (3.1)	62,397 (2.9)	6528 (1.2)	5552 (1.2)
Influenza and pneumonia	49,783 (1.7)	40,399 (1.9)	26,398 (5.0)	24,269 (5.4)
Nephritis, nephritic syndrome, nephrosis	51,565 (1.8)	42,230 (2.0)	3575 (0.7)	3323 (0.7)
Intentional self-harm	47,511 (1.7)	—	4832 (0.9)	751 (0.2)

[a] Calculated using International Classification of Diseases codes I00–I09, I11, I13, I20–I51.

Source: National Center for Health Statistics (United States, 2019), *http://www.cdc.gov/nchs*; National Statistics (Great Britain, 2019), *http://www.statistics.gov.uk*.

disease caused ~8% of all deaths, and cancer accounted for <4% of all deaths. In 1980, heart disease accounted for 38.2% of all deaths, cancer 20.9%, and cerebrovascular disease 8.6% of all deaths. By 2019, there had been a dramatic drop in deaths from cardiovascular and cerebrovascular diseases. In 2021, 23.1% of all deaths were from cardiovascular disease and just 5.3% from cerebrovascular disease. Deaths attributable to cancer, however, had increased slightly to 21.0%. The proportions of deaths due to chronic lower respiratory disease, diabetes, Alzheimer's, and suicide have increased. Interestingly, in 2019, HIV/AIDS accounted for <0.18% of all U.S. deaths. While unlikely to continue being a leading cause of death in the future, COVID-19 was the cause for >600,000 deaths in 2020–2021, and the official figure is almost certainly an undercount of the actual death toll.

This change in the epidemiology of death is also reflected in the costs of illness. In the United States, about 90% of all health care spending goes to patients with chronic illnesses, and around 10% of total personal health care spending—slightly <$450 billion in 2022—goes to the 0.98% of the population in the last year of their lives.

In upper-middle- and upper-income countries, an estimated 70% of all deaths are preceded by a disease or condition, making it reasonable to plan for dying in the foreseeable future. Cancer has served as the paradigm for terminal care, but it is not the only type of illness with a recognizable and predictable terminal phase. Since heart failure, chronic obstructive pulmonary disease (COPD), chronic liver failure, dementia, and many other conditions have recognizable terminal phases, a systematic approach to end-of-life care should be part of all medical specialties. Many patients with chronic illness–related symptoms and suffering benefit from palliative care regardless of prognosis. Palliative care is an essential part of comprehensive care for all chronically ill patients. Strong evidence demonstrates that palliative care can be improved by coordination between caregivers, doctors, and patients for advance care planning, as well as dedicated teams of physicians, nurses, and other providers.

SITE OF DEATH

Where patients die varies by country. In Belgium and Canada, for instance, over half of all cancer patients still die in the hospital. The past few decades have seen a steady shift, both in the United States and other countries like the Netherlands, out of the hospital, as patients and their families list their own homes as the preferred site of death. In the early 1980s, approximately 70% of American cancer patients died in the hospital. Today, that percentage is closer to 25% **(Fig. 13-1)**. Since 2000, there has been a shift in the United States from inpatient to home deaths, especially for patients with cancer, COPD, congestive heart

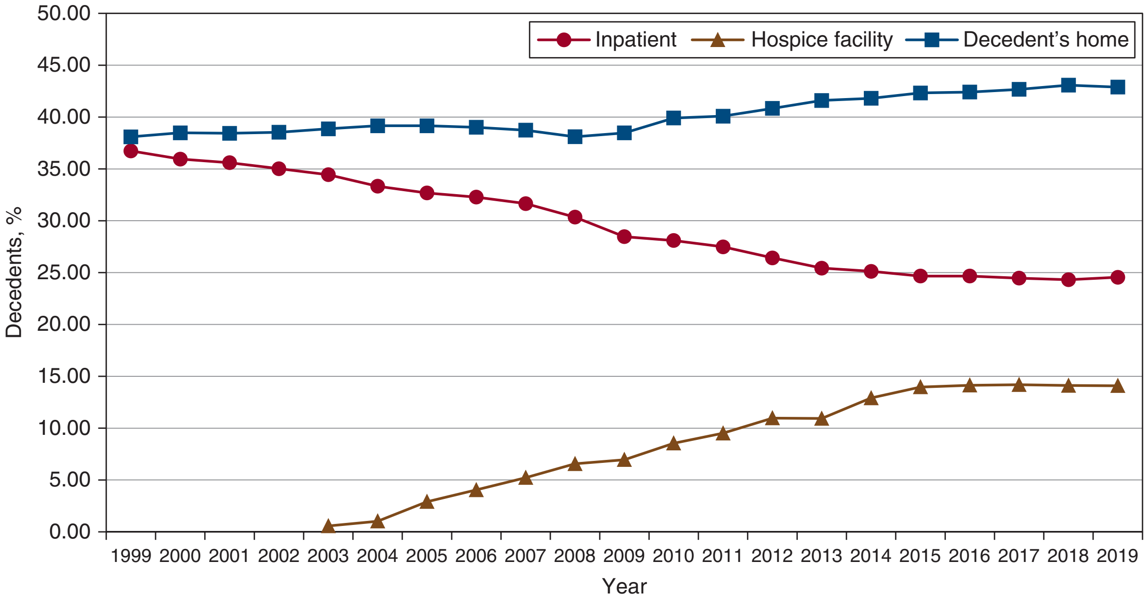

FIGURE 13-1 Graph showing trends in cancer decedents' site of death 1999–2019. *(Source: Centers for Disease Control and Prevention, National Center for Health Statistics. Underlying Cause of Death 1999-2019 on CDC WONDER Online Database. http://wonder.cdc.gov.)*

failure (CHF), and dementia. For instance, among Medicare beneficiaries, 30.1% of deaths due to cancer in 2000 occurred in acute care hospitals; by 2009, this figure had dropped to 22.1%; by 2015, it was 19.8%.

Paradoxically, while deaths in acute care hospitals have declined in the United States since 2000, both hospitalizations in the last 90 days of life and—even more troublingly—admission to the intensive care unit (ICU) in the last 30 days have actually increased. Over 40% of cancer patients in the United States are admitted to the ICU in their last 6 months of life, and >25% of cancer patients are admitted to the hospital in the last 30 days.

The shift in deaths out of the hospital has been accompanied by an increase in the use of hospice in the United States. In 2000, 21.6% of Medicare decedents used hospice at the time of death; by 2009, 42.2% were using hospice; and by 2019, 51.6% of Medicare decedents were enrolled in hospice at the time of death. Among those aged 85 or older, 60.8% of Medicare decedents were enrolled in hospice in 2021. Since the COVID-19 pandemic, this number has been a 29% over the 2021-2023 time period. Among cancer patients, ~60% were using hospice at the time of death. Hospice is also increasingly being used by noncancer patients. Today, cancer patients constitute only 25% of hospice users. Indeed, since 2014, the proportion of patients with other diagnoses using hospice has grown substantially. Those with cardiovascular disease has grown significantly, from 14% in 2014 to 29% in 2021, while neurocognitive disorders now account for 24% of cases. Of 2019 Medicare hospice decedents, 51.5% died at home, 17.4% in a nursing facility, 12.8% in a hospice inpatient facility, and 12.3% in assisted living.

Unfortunately, significant racial disparities exist in end-of-life care and the use of hospice, especially for noncancer deaths. Racial and ethnic minorities are less likely to receive hospice services than white decedents and are more likely to receive invasive or aggressive care in end-of-life treatment. Of people who died of head and neck cancers between 1999 and 2017, African Americans and Asians/Pacific Islanders were less likely to die at home or in hospice. Among Medicare beneficiaries who had a pancreatectomy for pancreatic cancer and lived at least 30 days, racial and ethnic minority patients remained 22% less likely than white patients to have hospice care before death.

In 2008, for the first time, the American Board of Medical Specialties (ABMS) offered certification in hospice and palliative medicine. With the shortening of hospital stays, many serious conditions are now being treated at home or on an outpatient basis. Consequently, providing optimal palliative and end-of-life care requires ensuring that appropriate services are available in a variety of settings, especially noninstitutional settings.

HOSPICE AND THE PALLIATIVE CARE FRAMEWORK

Central to this type of care is an interdisciplinary team approach that typically encompasses pain and symptom management, spiritual and psychological care for the patient, and support for family caregivers during the patient's illness and the bereavement period.

One of the more important changes in this field is beginning palliative care many months before death in order to focus on symptom relief and then switching to hospice in the patient's last few months. This approach avoids leaving hospice until the very end by introducing palliative care earlier, thereby allowing patients and families time to accommodate and transition. Phasing palliative care into end-of-life care means that patients will often receive palliative interventions long before they are formally diagnosed as terminally ill, or likely to die within 6 months.

Fundamental to ensuring quality palliative and end-of-life care is a focus on four broad domains: (1) physical symptoms; (2) psychological symptoms; (3) social needs that include interpersonal relationships, caregiving, and economic concerns; and (4) existential or spiritual needs.

■ ASSESSMENT AND CARE PLANNING

Comprehensive Assessment Standardized methods for conducting a comprehensive assessment focus on evaluating the patient's condition in all four domains affected by the illness: physical, psychological, social, and spiritual.

A comprehensive assessment should follow a modified version of the traditional medical history and physical examination and should emphasize both physical and psychological or mental symptoms. Questions should aim to elucidate symptoms, discern sources of suffering, and gauge how much those symptoms interfere with the patient's quality of life. Standardized and repeated assessments to evaluate the effectiveness of interventions are critical. Thus, clinicians should use shorter, validated instruments, such as (1) the revised Edmonton Symptom Assessment Scale; (2) Condensed Memorial Symptom Assessment Scale (MSAS); (3) MD Anderson Brief Symptom Inventory; (4) Rotterdam Symptom Checklist; (5) Symptom Distress Scale; (6) Patient-Reported Outcomes Measurement Information System; and (7) Interactive Symptom Assessment and Collection (ISAAC) tool.

MENTAL HEALTH With respect to mental health, many practices use the Patient Health Questionnaire-9 (PHQ-9) to screen for depression and the Generalized Anxiety Disorder-7 (GAD-7) to screen for anxiety. Using such tools ensures that the assessment is comprehensive and does not focus excessively on only pain.

INVASIVE TESTS Invasive tests are best avoided in end-of-life care, and even minimally invasive tests should be evaluated carefully for their benefit-to-burden ratio for the patient. Aspects of the physical examination that are uncomfortable and unlikely to yield useful information that changes patient management should be omitted.

SOCIAL NEEDS Health care providers should also assess the status of important relationships, financial burdens, caregiving needs, and access to medical care. Relevant questions include the following: *How often is there someone to feel close to? How has this illness been for your family? How has it affected your relationships? How much help do you need with things like getting meals and getting around? How much trouble do you have getting the medical care you need?*

EXISTENTIAL NEEDS To determine a patient's existential needs, providers should assess distress, the patient's sense of emotional and existential well-being, and whether the patient believes they have found purpose or meaning. Helpful assessment questions can include the following: *How much are you able to find meaning since your illness began? What things are most important to you at this stage?*

PERCEPTION OF CARE In addition, it can be helpful to ask how the patient perceives their care: *How much do you feel your doctors and nurses respect you? How clear is the information from us about what to expect regarding your illness? How much do you feel that the medical care you are getting fits with your goals?* If concern is detected in any of these areas, deeper evaluative questions are warranted.

Communication Particularly when an illness is life-threatening, there exists the potential for many emotionally charged and potentially conflict-creating moments—collectively called "bad news" situations—in which empathic and effective communication skills are essential. Those moments include the sharing of a terminal diagnosis with the patient and/or family, the discussion of the patient's prognosis and any treatment failures, the consideration of deemphasizing efforts to cure and prolong life while focusing more on symptom management and palliation, advance care planning, and the patient's actual death. Although these conversations can be difficult, research indicates that end-of-life discussions can lead to earlier hospice referrals, rather than overly aggressive treatment, ultimately benefiting quality of life for patients and improving the bereavement process for families.

Just as surgeons prepare for major operations and investigators rehearse a presentation of research results, physicians and health care providers caring for patients with significant or advanced illnesses should develop a standardized approach for sharing important information and planning interventions. In addition, physicians must be aware that families often care not only about how prepared the physician was to deliver bad news, but also the setting in which it was delivered. For instance, one study found that 27% of families making critical decisions for patients in an ICU desired better and more private physical space to communicate with physicians.

Multiple communications frameworks and guides have been developed to assist clinicians prepare for and conduct these serious illness conversations, including SPIKES, GUIDE, BREAKS, SHARE, ABCDE, and the Serious Illness Conversation Guide. SPIKES is the oldest and perhaps most commonly used in practice guide and consists of a six-step process for communicating bad news: (1) ***s***et up a suitable environment, (2) begin the discussion by finding out what the patient and/or family's ***p***erception and understand of their condition is, (3) obtain the patient's ***i***nvitation to receive new information and determine how much they want to know, (4) provide needed new ***k***nowledge accordingly, (5) allow for ***e***motional responses empathetically, and (6) ***s***ummarize and share next steps in care (Table 13-2).

Continuous Goal Assessment Major barriers to providing high-quality palliative and end-of-life care include the difficulty in determining an accurate prognosis and the emotional resistance of patients and their families to accepting the implications of a poor prognosis. A practical solution to these barriers is to integrate palliative care interventions or home visits from a palliative care visiting nurse months before the estimated final 6 months of life. Under this approach, palliative care no longer conveys the message of failure, having no more treatments, or "giving up hope." The transition from palliative to end-of-life care or hospice also feels less hasty and unexpected to the family. Fundamental to integrating palliative care with curative therapy is the inclusion of a continuous goal assessment as part of the routine patient reassessments that occur at most patient-physician encounters.

Goals for care are numerous, ranging from curing a specific disease, to prolonging life, to relieving a particular symptom, to adapting to a progressive disability without disrupting the family, to finding peace of mind or personal meaning, to dying in a manner that leaves loved ones with positive memories. Discerning a patient's goals for care can be approached through a seven-step protocol: (1) ensure that medical and other information is as complete as reasonably possible and is understood by all relevant parties (see above); (2) explore what the patient and/or family is hoping for, while also identifying relevant and realistic goals; (3) share all the options with the patient and family; (4) respond with empathy as they adjust to changing expectations; (5) make a plan that emphasizes what can be done to achieve the realistic goals; (6) follow through with the plan; and (7) periodically review the plan and consider at every encounter whether the goals of care should be revised with the patient and/or family. Each of these steps need not be followed in rote order, but together they provide a helpful framework for interactions with patients and their families regarding their goals

TABLE 13-2 Elements of Communicating Bad News—The P-SPIKES Approach

ACRONYM	STEPS	AIM OF THE INTERACTION	PREPARATIONS, QUESTIONS, OR PHRASES
P	Preparation	Mentally prepare for the interaction with the patient and/or family.	Review what information needs to be communicated. Plan how you will provide emotional support. Rehearse key steps and phrases in the interaction.
S	Setting of the interaction	Ensure the appropriate setting for a serious and potentially emotionally charged discussion.	Ensure that patient, family, and appropriate social supports are present. Devote sufficient time. Ensure privacy and prevent interruptions by people or beeper. Bring a box of tissues.
P	Patient's perception and preparation	Begin the discussion by establishing the baseline and whether the patient and family can grasp the information. Ease tension by having the patient and family contribute.	Start with open-ended questions to encourage participation. Possible questions to use: *What do you understand about your illness?* *When you first had symptom X, what did you think it might be?* *What did Dr. X tell you when Dr. X sent you here?* *What do you think is going to happen?*
I	Invitation and information needs	Discover what information needs the patient and/or family have and what limits they want regarding the bad information.	Possible questions to use: *If this condition turns out to be something serious, do you want to know?* *Would you like me to tell you all the details of your condition? If not, who would you like me to talk to?*
K	Knowledge of the condition	Provide the bad news or other information to the patient and/or family sensitively.	Do not just dump the information on the patient and family. Check for patient and family understanding. Possible phrases to use: *I feel badly to have to tell you this, but…* *Unfortunately, the tests showed…* *I'm afraid the news is not good…*
E	Empathy and exploration	Identify the cause of the emotions—e.g., poor prognosis. Empathize with the patient's and/or family's feelings. Explore by asking open-ended questions.	Strong feelings in reaction to bad news are normal. Acknowledge what the patient and family are feeling. Remind them such feelings are normal, even if frightening. Give them time to respond. Remind the patient and family you won't abandon them. Possible phrases to use: *I imagine this is very hard for you to hear.* *You look very upset. Tell me how you are feeling.* *I wish the news were different.* *We'll do whatever we can to help you.*
S	Summary and planning	Delineate for the patient and the family the next steps, including additional tests or interventions.	It is the unknown and uncertain that can increase anxiety. Recommend a schedule with goals and landmarks. Provide your rationale for the patient and/or family to accept (or reject). If the patient and/or family are not ready to discuss the next steps, schedule a follow-up visit.

Source: Adapted from R Buckman: *How to Break Bad News: A Guide for Health Care Professionals.* Baltimore, Johns Hopkins University Press, 1992.

for care. Such interactions can be especially challenging if a patient or family member has difficulty letting go of an unrealistic goal. In such cases, the provider should help them refocus on more realistic goals and should also suggest that while it is fine to hope for the best, it is still prudent to plan for other outcomes as well.

Advance Care Planning • PRACTICES Advance care planning is the process of planning for future medical care in case the patient becomes incapable of making medical decisions. A 2010 study of adults aged ≥60 who died between 2000 and 2006 found that while 42% of adults were required to make treatment decisions in their final days of life, 70% lacked decision-making capacity. Among those lacking decision-making capacity, approximately one-third did not have advance planning directives. Ideally, such planning would occur before a health care crisis or the terminal phase of an illness. Unfortunately, diverse barriers prevent this. Approximately 80% of Americans endorse advance care planning and living wills. However, few Americans have completed advance care documents. One systematic review of the literature from 2017 found that 38.2% of chronically ill patients completed advance care documents, higher than 32.7% of healthy adults. Importantly, 45.6% of older adults completed documents, but fewer than a third of people under 65 did so. According to a 2020 Gallup survey, 45% of American adults claim to have a living will. Other studies report that even fewer Americans—with some estimates as low as 26% of adults—have filled out advance care directives. Overall, the results suggest that between 55 and 67% of adults have not completed advance care documents. However, larger numbers of adults, between 50 and 70%, claim to have talked with someone about their treatment wishes. Americans aged 65 and older are more likely to complete an advance directive compared to younger adults (46 vs 32%), but this is still a minority of patients who are likely to need to plan.

Effective advance care planning should follow six key steps: (1) introducing the topic, (2) structuring a discussion, (3) reviewing plans that have been discussed by the patient and family, (4) documenting the plans, (5) updating them periodically, and (6) implementing the advance care directives (Table 13-3). Two of the main barriers to advance care planning are problems in raising the topic and difficulty in structuring a succinct discussion. Raising the topic can be done efficiently as a routine matter, noting that it is recommended for all patients, analogous to purchasing insurance or estate planning. Many of the most difficult cases have involved unexpected, acute episodes of brain damage in young individuals.

Structuring a focused discussion is an important communication skill. To do so, a provider must first identify the health care proxy and recommend their involvement in the advance care planning process. Next, a worksheet must be selected that has been demonstrated to produce reliable and valid expressions of patient preferences, and the patient and proxy must be oriented to it. Such worksheets exist for both general and disease-specific situations. The provider should then discuss with the patient and proxy one example scenario to demonstrate how to think about the issues. It is often helpful to begin with a scenario in which the patient is likely to have settled preferences for care, such

TABLE 13-3 Steps in Advance Care Planning

STEP	GOALS TO BE ACHIEVED AND MEASURES TO COVER	USEFUL PHRASES OR POINTS TO MAKE
Introduce advance care planning	Ask the patient what they know about advance care planning and if they have already completed an advance care directive.	*I'd like to talk with you about something I try to discuss with all my patients. It's called advance care planning. In fact, I feel that this is such an important topic that I have done this myself. Are you familiar with advance care planning or living wills?*
	Indicate that you as a physician have completed advance care planning.	*Have you thought about the type of care you would want if you ever became too sick to speak for yourself? That is the purpose of advance care planning.*
	Indicate that you try to perform advance care planning with all patients regardless of prognosis.	*There is no change in health that we have not discussed. I am bringing this up now because it is sensible for everyone, no matter how well or ill, old or young.*
	Explain the goals of the process as empowering the patient and ensuring that you and the proxy understand the patient's preferences.	*Have many copies of advance care directives available, including in the waiting room, for patients and families.*
	Provide the patient relevant literature, including the advance care directive that you prefer to use.	*Know resources for state-specific forms (available at www.nhpco.org).*
	Recommend the patient identify a proxy decision-maker who should attend the next meeting.	
Have a structured discussion of scenarios with the patient	Affirm that the goal of the process is to follow the patient's wishes if the patient loses decision-making capacity.	*Use a structured worksheet with typical scenarios.*
	Elicit the patient's overall goals related to health care. Elicit the patient's preferences for specific interventions in a few salient and common scenarios. Help the patient define the threshold for withdrawing and withholding interventions. Define the patient's preference for the role of the proxy.	*Begin the discussion with persistent vegetative state and consider other scenarios, such as recovery from an acute event with serious disability; then ask the patient about their preferences regarding specific interventions, such as ventilators, artificial nutrition, and CPR; finally, proceeding to less invasive interventions, such as blood transfusions and antibiotics.*
Review the patient's preferences	After the patient has made choices of interventions, review them to ensure they are consistent and the proxy is aware of them.	
Document the patient's preferences	Formally complete the advance care directive and have a witness sign it.	
	Provide a copy for the patient and the proxy.	
	Insert a copy into the patient's medical record and summarize it in a progress note.	
Update the directive	Periodically, and with major changes in health status, review the directive with the patient and make any modifications.	
Apply the directive	The directive goes into effect only when the patient becomes unable to make medical decisions for themself.	
	Reread the directive to be sure about its content.	
	Discuss your proposed actions based on the directive with the proxy.	

Abbreviation: CPR, cardiopulmonary resuscitation.

as being in a persistent vegetative state. Once the patient's preferences for interventions in this scenario are determined, the provider should suggest that the patient and proxy discuss and complete the worksheet for each other. If appropriate, the patient and proxy should consider involving other family members in the discussion. During a subsequent return visit, the provider should go over the patient's preferences, checking and resolving any inconsistencies. After having the patient and proxy sign the document, the provider should place the document in the patient's medical chart and make sure that copies are provided to relevant family members and care sites. Since patients' preferences can change, these documents must be reviewed periodically.

TYPES OF DOCUMENTS Advance care planning documents are of two broad types. The first includes living wills, also known as instructional directives; these are advisory documents that describe the types of decisions that should direct a patient's care. Some are more specific, delineating different scenarios and interventions for the patient to choose from. Among these, some are for general use and others are designed for use by patients with a specific type of disease, such as cancer, renal failure, or HIV. Less specific directives can be general statements, such as not wanting life-sustaining interventions, or forms that describe the values that should guide specific discussions about terminal care. The second type of advance directive allows the designation of a health care proxy (sometimes also referred to as a durable attorney for health care), an individual selected by the patient to make decisions. The choice is not either/or; a combined directive that includes a living will and designates a proxy is often used, and the directive should indicate clearly whether the specified patient preferences or the proxy's choice takes precedence if they conflict. Some states have begun to put into practice a "Physician Orders for Life-Sustaining Treatment (POLST)" directive, which builds on communication between providers and patients by including guidance for end-of-life care in a color-coordinated form that follows the patient across treatment settings. The procedures for completing advance care planning documents vary according to state law.

A potentially misleading distinction relates to statutory, as opposed to advisory, documents. Statutory documents are drafted to fulfill relevant state laws. Advisory documents are drafted to reflect the patient's wishes. Both are legal, the former under state law and the latter under common or constitutional law.

LEGAL ASPECTS As of 2024, 48 states and the District of Columbia had enacted living will legislation. Massachusetts and Michigan are the two states without living will legislation. Indiana has a life-prolonging procedures declaration. States differ in the requirements for advanced directives, including whether they need to be witnessed and, if so, by how many witnesses and whether they need to be notarized. Importantly, in 26 states, the laws state that the living will is not valid if a woman is pregnant. All states except Alaska have enacted durable power of attorney for health care laws that permit patients to designate a proxy decision-maker with authority to terminate life-sustaining treatments. Only in Alaska does the law prohibit proxies from terminating life-sustaining treatments for pregnant women.

The U.S. Supreme Court has ruled that patients have a constitutional right to decide any issues related to refusing or terminating medical interventions, including life-sustaining interventions, and that mentally incompetent patients can exercise this right by providing "clear and convincing evidence" of their preferences. Since advance care directives permit patients to provide such evidence, commentators agree that they are constitutionally protected. Most commentators believe that a state is required to honor any clear advance care directive, regardless of whether it is written on an "official" form. Many states have enacted laws for the explicit purpose of honoring out-of-state directives. If a patient is not using a statutory form, it may be advisable to attach a statutory form to the advance care directive being used. State-specific forms are readily available free of charge for health care providers, patients, and families through the website of the National Hospice and Palliative Care Organization (*http://www.nhpco.org*). The National Hospice and Palliative Care Organization website also has created CaringInfo (*https://www.caringinfo.org/*), a free online resource that contains free patient-centered educational materials on end-of-life care and advanced care planning.

REIMBURSEMENT As of January 1, 2016, the Centers for Medicare and Medicaid Services amended the physician fee schedule to reimburse discussions of advance care planning under Current Procedural Terminology codes 99497 and 99498. The session must be voluntary and include an explanation of advance care planning but need not include a completed advance care document. There can be multiple bills for the discussion if it extends over several encounters. A study found that patients who engaged in a billed advance care planning encounter were more likely to be enrolled in hospice and less likely to receive intensive therapies, despite being more likely to be hospitalized in the ICU. However, a billing incentive in and of itself may not increase advance care planning discussions by clinicians. In 2016, just 1.6% of Medicare Advantage patients had a discussion of advance care planning that was billed. Moreover, one retrospective study examining Medicare claims data from 2016 to 2018 found that only 15% of practices billed for any advanced care planning visits. Factors beyond reimbursement, such as clinicians' lack of comfort and skill in carrying out advance care planning discussions and lack of time, appear to impede discussions of advance care planning.

INTERVENTIONS

PHYSICAL SYMPTOMS AND THEIR MANAGEMENT

Great emphasis has been placed on addressing dying patients' pain. In order to emphasize its importance, pain assessment has frequently been included as the fifth vital sign. Heightened consideration of pain has been advocated by large health care systems such as the Veterans' Administration and accrediting bodies such as The Joint Commission. Although this embrace of pain has been symbolically important, available data suggest that making pain the fifth vital sign does not lead to improved pain management practices. In light of the opioid crisis in the United States, the emphasis on pain management has begun to be reexamined. For instance, in 2017 standards, The Joint Commission recommends nonpharmacologic pain treatment as well as identification of psychosocial risk factors for addiction. Importantly, good palliative care requires much more than good pain management. The frequency of symptoms varies by disease and other factors. The most common physical and psychological symptoms among all terminally ill patients include pain, fatigue, insomnia, anorexia, dyspnea, depression, anxiety, nausea, and vomiting. In the last days of life, terminal delirium and excessive secretions are also common. Assessments of patients with advanced cancer have shown that patients experienced an average of 11.5 different physical and psychological symptoms (**Table 13-4**).

TABLE 13-4 Common Physical and Psychological Symptoms of Terminally Ill Patients

PHYSICAL SYMPTOMS	PSYCHOLOGICAL SYMPTOMS
Pain	Anxiety
Fatigue and weakness	Depression
Dyspnea	Hopelessness
Insomnia	Meaninglessness
Dry mouth	Irritability
Anorexia	Impaired concentration
Nausea and vomiting	Confusion
Constipation	Delirium
Cough	Loss of libido
Swelling of arms or legs	
Itching	
Diarrhea	
Dysphagia	
Dizziness	
Fecal and urinary incontinence	
Numbness/tingling in hands/feet	

In the vast majority of cases, evaluations to determine the etiology of these symptoms should be limited to the history and physical examination. In some cases, radiologic or other diagnostic examinations will provide sufficient benefit in directing optimal palliative care to warrant the risks, potential discomfort, and inconvenience, especially to a seriously ill patient. Only a few of the common symptoms that present difficult management issues will be addressed in this chapter. **Additional information on the management of other symptoms, such as nausea and vomiting, insomnia, and diarrhea, can be found in Chaps. 48, 33, and 49, respectively. Information on the management of patients with cancer is provided in Chaps. 73 and 74.**

Pain • **FREQUENCY** The frequency of pain among terminally ill patients varies significantly. Cancer (~85%), CHF (~75%), and HIV/AIDS have been associated with a higher prevalence of pain compared to other advanced illnesses, such as COPD (~45%), chronic kidney disease (~40%), and dementia (~40%). One meta-analysis of adults with advanced or terminal illness found pain prevalence of 30–94% in patients with cancer, compared to 21–77% for COPD, 14–78% for CHF, 11–83% for end-stage renal disease, 14–63% for dementia, and 30–98% for AIDS.

ETIOLOGY There are two types of pain: nociceptive and neuropathic. Nociceptive pain is further divided into somatic or visceral pain. *Somatic pain* is the result of direct mechanical or chemical stimulation of nociceptors and normal neural signaling to the brain. It tends to be localized, aching, throbbing, and cramping. The classic example is bone metastases. *Visceral pain* is caused by nociceptors in gastrointestinal (GI), respiratory, and other organ systems. It is a deep or colicky type of pain classically associated with pancreatitis, myocardial infarction, or tumor invasion of viscera. *Neuropathic pain* arises from disordered nerve signals. It is described by patients as burning, electrical, or shock-like pain. Classic examples are poststroke pain, tumor invasion of the brachial plexus, and herpetic neuralgia.

ASSESSMENT Pain is a subjective experience. Depending on the patient's circumstances, perspective, and physiologic condition, the same physical lesion or disease state can produce different levels of reported pain and need for pain relief. Systematic assessment includes eliciting the following: (1) type: throbbing, cramping, burning, etc.; (2) periodicity: continuous, with or without exacerbations, or incident; (3) location; (4) intensity; (5) modifying factors; (6) effects of treatments; (7) functional impact; and (8) impact on patient. Several validated pain assessment measures may be used, including the Visual Analogue Scale (VAS), the Defense and Veterans Pain Rating Scale (DVPRS), the Brief Pain Inventory (BPI), or the Numerical Pain Rating Scale (NRS-11). Other scales have been developed for neuropathic pain, such as the Neuropathic Pain Scale and the DN4 Questionnaire. Frequent reassessments on a consistent scale are essential to assess the impact of and need to readjust interventions.

INTERVENTIONS Interventions for pain must be tailored to each individual, with the goal of preempting chronic pain and relieving breakthrough pain. At the end of life, there is rarely reason to doubt a patient's report of pain. With the opioid crisis in the United States, there is more emphasis on making opioids one component of multimodal analgesia. Nevertheless, at the end of life, pain medications, especially opioids, remain the cornerstone of management **(Fig. 13-2)**. If they are failing and nonpharmacologic interventions—including radiotherapy and anesthetic or neurosurgical procedures such as peripheral nerve blocks or epidural medications—are required, a pain consultation is appropriate.

Pharmacologic interventions still largely follow the World Health Organization three-step, "analgesic ladder" approach, which involves nonopioid analgesics, "mild" opioids, and "strong" opioids, with or without adjuvants **(Chap. 14)**. Nonopioid analgesics, especially nonsteroidal anti-inflammatory drugs (NSAIDs), are the initial treatments for mild pain. They work primarily by inhibiting peripheral prostaglandins and reducing inflammation but may also have central nervous system (CNS) effects. Additionally, NSAIDs have a ceiling effect. Ibuprofen, up to 2400 mg/d qid, has a minimal risk of causing bleeding and renal impairment and is a good initial choice. In patients with a history of severe GI or other bleeding, however, ibuprofen should be

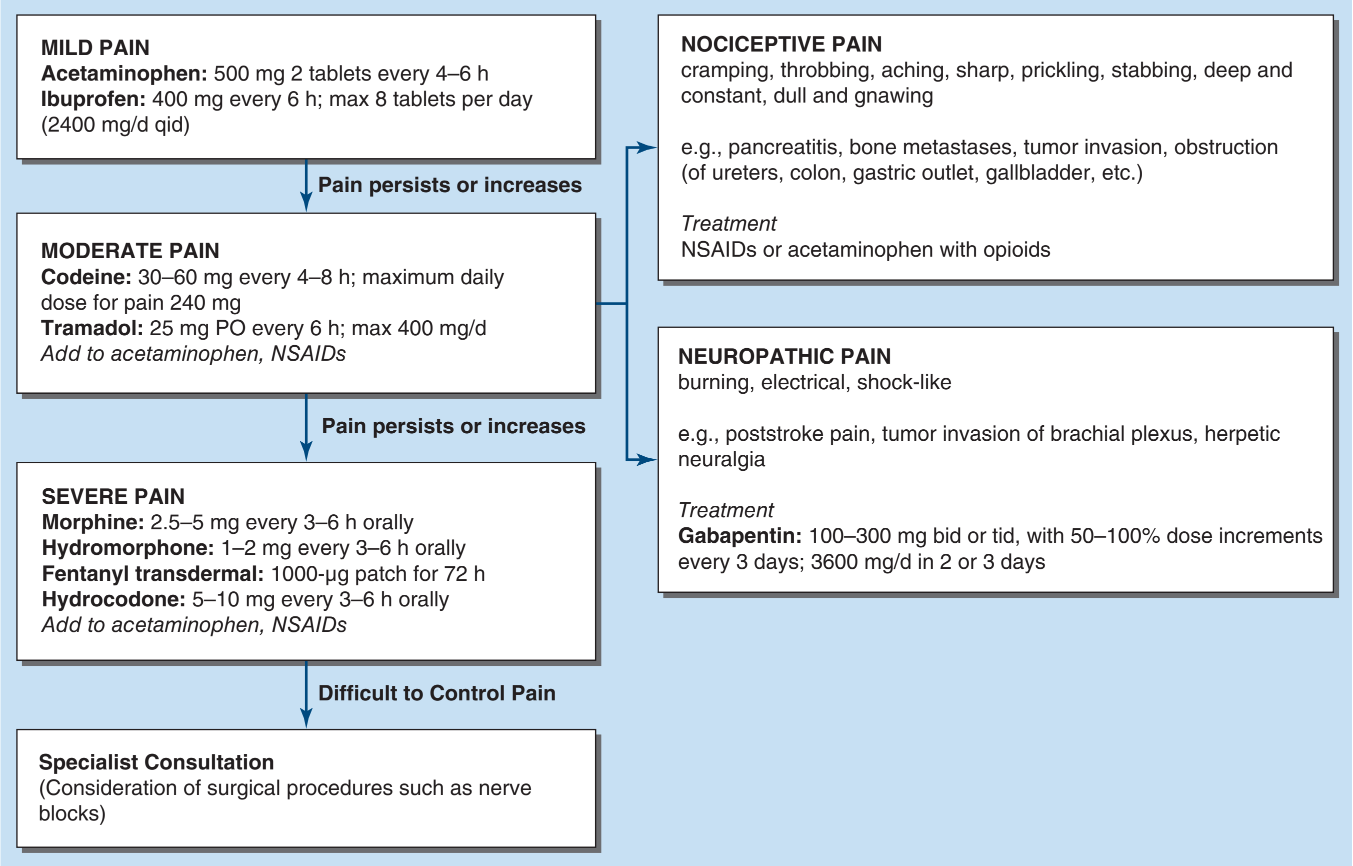
MILD PAIN
Acetaminophen: 500 mg 2 tablets every 4–6 h
Ibuprofen: 400 mg every 6 h; max 8 tablets per day (2400 mg/d qid)
Pain persists or increases
MODERATE PAIN
Codeine: 30–60 mg every 4–8 h; maximum daily dose for pain 240 mg
Tramadol: 25 mg PO every 6 h; max 400 mg/d
Add to acetaminophen, NSAIDs
Pain persists or increases
SEVERE PAIN
Morphine: 2.5–5 mg every 3–6 h orally
Hydromorphone: 1–2 mg every 3–6 h orally
Fentanyl transdermal: 1000-μg patch for 72 h
Hydrocodone: 5–10 mg every 3–6 h orally
Add to acetaminophen, NSAIDs
Difficult to Control Pain
Specialist Consultation
(Consideration of surgical procedures such as nerve blocks)
NOCICEPTIVE PAIN
cramping, throbbing, aching, sharp, prickling, stabbing, deep and constant, dull and gnawing
e.g., pancreatitis, bone metastases, tumor invasion, obstruction (of ureters, colon, gastric outlet, gallbladder, etc.)
Treatment
NSAIDs or acetaminophen with opioids
NEUROPATHIC PAIN
burning, electrical, shock-like
e.g., poststroke pain, tumor invasion of brachial plexus, herpetic neuralgia
Treatment
Gabapentin: 100–300 mg bid or tid, with 50–100% dose increments every 3 days; 3600 mg/d in 2 or 3 days

FIGURE 13-2 Terminal pain management flow chart. NSAIDs, nonsteroidal anti-inflammatory drugs.

avoided. In patients with a history of mild gastritis or gastroesophageal reflux disease (GERD), acid-lowering therapy, such as a proton pump inhibitor, should be used. Acetaminophen is an alternative in patients with a history of GI bleeding and can be used safely at up to 4 g/d qid. In patients with liver dysfunction due to metastases or other causes and in patients with heavy alcohol use, doses should be reduced. Patients with cirrhosis can receive doses up to 2 g/d.

If nonopioid analgesics are insufficient, opioids should be introduced. Opioids primarily work by interacting with μ opioid receptors to activate pain-inhibitory neurons in the CNS, although they also interact variably with δ and κ receptors. Receptor agonists, such as morphine, codeine, and fentanyl, produce analgesia by activating pain-inhibitory neurons in the CNS. Partial agonists, such as buprenorphine, have a ceiling effect for analgesia and a lower potential for abuse. They are useful for postacute pain but should not be used for chronic pain in end-of-life care. Pure antagonists, such as naloxone and methylnaltrexone, are used for reversal of opioid effects.

Traditionally, "weak" opioids such as codeine were used first. If they failed to relieve pain after dose escalation, "strong" opioids like morphine were used in doses of 5–10 mg every 4 h. However, this breakdown between "weak" and "strong" opioids is no longer commonly accepted, with smaller doses of "stronger" opioids frequently being preferred over similar or larger doses of "weaker" opioids, and different pain syndromes having different preferred therapies. Regardless, nonopioid analgesics should be combined with opioids, as they potentiate the effect of opioids.

Importantly, the goal is to prevent patients from experiencing pain. Consequently, for continuous pain, opioids should be administered on a regular, around-the-clock basis consistent with their duration of analgesia, and the next dose should occur before the effect of the previous dose wears off. They should not be provided only when the patient experiences pain. Patients should also be provided rescue medication, such as liquid morphine, for breakthrough pain, generally at 20% of the baseline dose. Patients should be informed that using the rescue medication does not obviate the need to take the next standard dose of pain medication. If the patient's pain remains uncontrolled after 24 h and recurs before the next dose, requiring the patient to utilize the rescue medication, the daily opioid dose can be increased by the total dose of rescue medications used by the patient or by 50% of the standing opioid daily dose for moderate pain and 100% for severe pain.

It is inappropriate to start with extended-release preparations. Instead, an initial focus on using short-acting preparations to determine how much is required in the first 24–48 h will allow clinicians to determine opioid needs. Once pain relief is obtained using short-acting preparations, the switch should be made to extended-release preparations. Even with a stable extended-release preparation regimen, the patient may experience incident pain, such as during movement or dressing changes. Short-acting preparations should be taken before such predictable episodes. Although less common, patients may have "end-of-dose failure" with long-acting opioids, meaning that they develop pain after 8 h in the case of an every-12-h medication. In these cases, a trial of giving an every-12-h medication every 8 h is appropriate.

Due to differences in opioid receptors, cross-tolerance among opioids is incomplete, and patients may experience different side effects with different opioids. Therefore, if a patient is not experiencing pain relief or is experiencing too many side effects, a change to another opioid preparation is appropriate. When switching, one should begin with 50–75% of the published equianalgesic dose of the new opioid.

Unlike NSAIDs, opioids have no ceiling effect; therefore, there is no maximum dose, no matter how many milligrams the patient is receiving. The appropriate dose is the dose needed to achieve pain relief. This is an important point for clinicians to explain to patients and families. Addiction or excessive respiratory depression is extremely unlikely in the terminally ill; fear of these side effects should neither prevent escalating opioid medications when the patient is experiencing insufficient pain relief nor justify using opioid antagonists.

Opioid side effects should be anticipated and treated preemptively. Nearly all patients experience constipation that can be debilitating (see below). Failure to prevent constipation often results in noncompliance with opioid therapy. The preferred treatment is prevention. Fiber supplementation with bulk-forming agents (psyllium 1 tbsp qd) as well as osmotic laxatives (polyethylene glycol 1 tbsp qd) and stimulant laxatives (senna 2 tbsp qhs) are considered first-line treatment. For refractory cases, opioid antagonists or other therapies, such as lubiprostone, should be considered.

Methylnaltrexone is the best-studied opioid antagonist for use in refractory opioid-induced constipation. It reverses opioid-induced constipation by blocking peripheral opioid receptors, but not central receptors, for analgesia. In placebo-controlled trials, it has been shown to cause laxation within 24 h of administration. As with the use of opioids, about a third of patients using methylnaltrexone experience nausea and vomiting, but unlike with opioid usage, tolerance usually develops within a week. Therefore, when one is beginning opioids, an antiemetic such as metoclopramide or a serotonin antagonist is often prescribed prophylactically and stopped after 1 week. Olanzapine has also been shown to have antinausea properties and can be effective in countering delirium or anxiety, with the advantage of some weight gain.

Drowsiness, a common side effect of opioids, also usually abates within a week. For refractory or severe cases, pharmacologic therapy should be considered. The best-studied agents are the psychostimulants dextroamphetamine, methylphenidate, and modafinil, although evidence regarding their efficacy is weak. Modafinil has the advantage of once-a-day dosing compared to methylphenidate's twice daily dosing.

Seriously ill patients who require chronic pain relief rarely become addicted. Suspicion of addiction should not be a reason to withhold pain medications from terminally ill patients. Nonetheless, patients and families may withhold prescribed opioids for fear of addiction or dependence. Physicians and health care providers should reassure patients and families that the patient will not become addicted to opioids if they are used as prescribed for pain relief; this fear should not prevent the patient from taking the medications around the clock. However, diversion of drugs for use by other family members or illicit sale may occur. It is necessary to advise the patient and caregiver about secure storage of opioids. Contract writing with the patient and family can help. If that fails, transfer to a safe facility may be necessary.

Tolerance describes the need to increase medication dosage for the same pain relief without a concurrent change in disease. In the case of patients with advanced disease, the need for increasing opioid dosage for pain relief usually is caused by disease progression rather than tolerance. Physical dependence is indicated by symptoms resulting from the abrupt withdrawal of opioids and should not be confused with addiction.

In recent years, the potential dangers of opioid drugs have become increasingly apparent. To help mitigate the risk of these powerful drugs, several strategies should be used to reduce the risk of aberrant drug use. To start, all patients should be assessed for their individual levels of risk. While there are multiple surveys available, including the Opioid Risk Tool, none have gained widespread use or validation. It is also important to screen for prior substance use disorder, major psychiatric disorders, and in-state and out-of-state controlled substance fill histories.

For patients deemed to be high risk, a multidisciplinary effort should be pursued to reduce the risk of adverse consequences, such as addiction and diversion. Prescribing strategies include selecting opioids with longer durations of action and lower street values, such as methadone, and prescribing smaller quantities with more frequent follow-up. Monitoring options include periodic urine screening and referral to pain specialists. In some cases, it may also be reasonable to consider not offering short-acting opioids for breakthrough pain. In no situation, however, should adequate pain relief be withheld due to risk.

Adjuvant analgesic medications are nonopioids that potentiate the analgesic effects of opioids. They are especially important in the management of neuropathic pain. Gabapentin, an anticonvulsant initially studied in the setting of herpetic neuralgia, is now the first-line treatment for neuropathic pain resulting from a variety of causes. It is begun at 100–300 mg bid or tid, with 50–100% dose increments every

3 days. Usually 900–3600 mg/d in two or three doses is effective. The combination of gabapentin and nortriptyline may be more effective than gabapentin alone. Two potential side effects of gabapentin to be aware of are confusion and drowsiness, especially in the elderly. Other effective adjuvant medications include pregabalin, which has the same mechanism of action as gabapentin but is absorbed more efficiently from the GI tract. Several antidepressants have also been shown to be beneficial, including duloxetine, venlafaxine, nortriptyline, and amitriptyline. Second-line agents include topical medications such as capsaicin cream and lidocaine patches. Lamotrigine is a novel agent whose mechanism of action is unknown but has been shown to be effective. It is recommended to begin at 25–50 mg/d, increasing to 100 mg/d. Carbamazepine, a first-generation agent, has been proven effective in randomized trials for neuropathic pain. Other potentially effective anticonvulsant adjuvants include topiramate (25–50 mg qd or bid, rising to 100–300 mg/d) and oxcarbazepine (75–300 mg bid, rising to 1200 mg bid).

Glucocorticoids, preferably dexamethasone given once a day, can be useful in reducing inflammation that causes pain, while also elevating mood, energy, and appetite. Its main side effects include confusion, sleep difficulties, and fluid retention. Glucocorticoids are especially effective for bone pain and abdominal pain from distention of the GI tract or liver. Other drugs, including clonidine and baclofen, can be effective in providing pain relief. These drugs are adjuvants and generally should be used in conjunction with—not instead of—opioids. Methadone, carefully dosed because of its unpredictable half-life in many patients, has activity at the *N*-methyl-D-aspartame (NMDA) receptor and is useful for complex pain syndromes and neuropathic pain. It is generally reserved for cases in which first-line opioids (morphine, oxycodone, hydromorphone) are either ineffective or unavailable.

Radiation therapy can treat bone pain from single metastatic lesions. Bone pain from multiple metastases can be amenable to radiopharmaceuticals such as strontium-89 and samarium-153. Bisphosphonates, such as pamidronate (90 mg every 4 weeks) and calcitonin (200 IU intranasally once or twice a day), also provide relief from bone pain but have multiday onsets of action.

Constipation • **FREQUENCY** Constipation is reported in up to 70–100% of patients requiring palliative care.

ETIOLOGY Although hypercalcemia and other factors can cause constipation, it is most frequently a predictable consequence of the use of opioids for pain and dyspnea relief and of the anticholinergic effects of tricyclic antidepressants, as well as due to the inactivity and poor diets common among seriously ill patients. If left untreated, constipation can cause substantial pain and vomiting and also is associated with confusion and delirium. Whenever opioids and other medications known to cause constipation are used, preemptive treatment for constipation should be instituted.

ASSESSMENT Assessing constipation can be difficult because people describe it differently. Four commonly used assessment scales are the Bristol Stool Form Scale, the Constipation Assessment Scale, the Constipation Visual Analogue Scale, and the Eton Scale Risk Assessment for Constipation. The Bowel Function Index can be used to quantify opioid-induced constipation. The physician should establish the patient's previous bowel habits, as well as any changes in subjective and objective qualities such as bloating or decreased frequency. Abdominal and rectal examinations should be performed to exclude impaction or an acute abdomen. Radiographic assessments beyond a simple flat plate of the abdomen in cases in which obstruction is suspected are rarely necessary.

INTERVENTION Any measure to address constipation during end-of-life care should include interventions to reestablish comfortable bowel habits and to relieve pain or discomfort. Although physical activity, adequate hydration, and dietary treatments with fiber can be helpful, each is limited in its effectiveness for most seriously ill patients, and fiber may exacerbate problems in the setting of dehydration or if impaired motility is the etiology. Fiber is contraindicated in the presence of opioid use. Stimulant and osmotic laxatives, stool softeners, fluids, and enemas are the mainstays of therapy (Table 13-5). To prevent constipation from opioids and other medications, a combination of a laxative and a stool softener (such as senna and docusate) should be used. If after several days of treatment a bowel movement has not occurred, a rectal examination to remove impacted stool and place a suppository is necessary. For patients with impending bowel obstruction or gastric stasis, octreotide to reduce secretions can be helpful. For patients in whom the suspected mechanism is dysmotility, metoclopramide can be helpful.

TABLE 13-5 Medications for the Management of Constipation

INTERVENTION	DOSE	COMMENT
Stimulant laxatives		These agents directly stimulate peristalsis and may reduce colonic absorption of water.
Prune juice	120–240 mL/d	Work in 6–12 h.
Senna (Senokot)	2–8 tablets PO bid	
Bisacodyl	5–15 mg/d PO, PR	
Osmotic laxatives		These agents are not absorbed. They attract and retain water in the gastrointestinal tract.
Lactulose	15–30 mL PO q4–8h	Lactulose may cause flatulence and bloating.
Magnesium hydroxide (Milk of Magnesia)	15–30 mL/d PO	Lactulose works in 1 day, magnesium products in 6 h.
Magnesium citrate	125–250 mL/d PO	
Stool softeners		These agents work by increasing water secretion and as detergents, increasing water penetration into the stool.
Sodium docusate (Colace)	300–600 mg/d PO	Work in 1–3 days.
Calcium docusate	300–600 mg/d PO	
Suppositories and enemas		
Bisacodyl	10–15 PR qd	
Sodium phosphate enema	PR qd	Fixed dose, 4.5 oz, Fleet's.

Nausea • **FREQUENCY** Up to 70% of patients with advanced cancer have nausea, defined as the subjective sensation of wanting to vomit.

ETIOLOGY Nausea and vomiting are both caused by stimulation at one of four sites: the GI tract, the vestibular system, the chemoreceptor trigger zone (CTZ), and the cerebral cortex. Medical treatments for nausea are aimed at receptors at each of these sites: the GI tract contains mechanoreceptors, chemoreceptors, and 5-hydroxytryptamine type 3 (5-HT$_3$) receptors; the vestibular system probably contains histamine and acetylcholine receptors; and the CTZ contains chemoreceptors, dopamine type 2 receptors, and 5-HT$_3$ receptors. An example of nausea that most likely is mediated by the cortex is anticipatory nausea before a dose of chemotherapy or other noxious stimuli.

Specific causes of nausea include metabolic changes (liver failure, uremia from renal failure, hypercalcemia), bowel obstruction, constipation, infection, GERD, vestibular disease, brain metastases, medications (including antibiotics, NSAIDs, proton pump inhibitors, opioids, and chemotherapy), and radiation therapy. Anxiety can also contribute to nausea.

INTERVENTION Medical treatment of nausea is directed at the anatomic and receptor-mediated cause revealed by a careful history and physical examination. When no specific cause of nausea is identified, many advocate beginning treatment with metoclopramide; a serotonin type 3 (5-HT$_3$) receptor antagonist such as ondansetron, granisetron, palonosetron, dolasetron, tropisetron, or ramosetron; or a dopamine antagonist such as chlorpromazine, haloperidol, or prochlorperazine. When decreased motility is suspected, metoclopramide can be an effective treatment. When inflammation of the GI tract is suspected, glucocorticoids, such as dexamethasone, are an appropriate treatment. For nausea that follows chemotherapy and radiation therapy, one of the 5-HT$_3$ receptor antagonists or neurokinin-1 antagonists, such as aprepitant or fosaprepitant, is recommended. Clinicians should attempt prevention of postchemotherapy nausea, rather than simply providing treatment after the fact. Current clinical guidelines recommend tailoring the strength of treatments to the specific emetic risk posed by a specific chemotherapy drug. When a vestibular cause (such as "motion sickness" or labyrinthitis) is suspected, antihistamines, such as meclizine (whose primary side effect is drowsiness), or anticholinergics, such as scopolamine, can be effective. In anticipatory nausea, patients can benefit from nonpharmacologic interventions, such as biofeedback and hypnosis. The most common pharmacologic intervention for anticipatory nausea is a benzodiazepine, such as lorazepam. As with antihistamines, drowsiness and confusion are the main side effects.

The use of medical marijuana or oral cannabinoids for palliative treatment of nausea is controversial, as there are no controlled trials showing its effectiveness for patients at the end of life. A 2018 meta-analysis concluded that "there is reasonable evidence that cannabinoids improve nausea and vomiting after chemotherapy," and such treatments are not as good as 5-HT$_3$ receptor antagonists and can sometimes even cause cannabis hyperemesis syndrome. Older patients, who compose the vast majority of dying patients, seem to tolerate cannabinoids poorly.

Dyspnea • **FREQUENCY** Dyspnea is the subjective experience of being short of breath. Over 50%, and as many as 75%, of dying patients, especially those with lung cancer, metastases to the lung, CHF, and COPD, experience dyspnea at some point near the end of life. Dyspnea is among the most distressing of physical symptoms and can be even more distressing than pain.

ASSESSMENT As with pain, dyspnea is a subjective experience that may not correlate with objective measures of PO$_2$, PCO$_2$, or respiratory rate. Consequently, measurements of oxygen saturation through pulse oximetry or blood gases are rarely helpful in guiding therapy. Despite the limitations of existing assessment methods, physicians should regularly assess and document patients' experience of dyspnea and its intensity. Guidelines recommend visual analogue dyspnea scales to assess the severity of symptoms and the effects of treatment. Potentially reversible or treatable causes of dyspnea include infection, pleural effusions, pulmonary emboli, pulmonary edema, asthma, and tumor encroachment on the airway. However, the risk-versus-benefit ratio of the diagnostic and therapeutic interventions for patients with little time left to live must be considered carefully before undertaking diagnostic steps. Frequently, the specific etiology cannot be identified, and dyspnea is the consequence of progression of the underlying disease that cannot be treated. The anxiety caused by dyspnea and the choking sensation can significantly exacerbate the underlying dyspnea in a negatively reinforcing cycle.

INTERVENTIONS When reversible or treatable etiologies are diagnosed, they should be treated as long as the side effects of treatment, such as repeated drainage of effusions or anticoagulants, are less burdensome than the dyspnea itself. More aggressive treatments such as stenting a bronchial lesion may be warranted if it is clear that the dyspnea is due to tumor invasion at that site and if the patient and family understand the risks of such a procedure.

Usually, treatment will be symptomatic (Table 13-6). Supplemental oxygen does not appear to be effective. "A systematic review of the literature failed to demonstrate a consistent beneficial effect of oxygen inhalation over air inhalation for study participants with dyspnea due to end-stage cancer or cardiac failure." Therefore, oxygen may be no more than an expensive placebo. Low-dose opioids reduce the sensitivity of the central respiratory center and relieve the sensation of dyspnea. If patients are not receiving opioids, weak opioids can be initiated; if patients are already receiving opioids, morphine or other stronger opioids should be used. Controlled trials do not support the use of nebulized opioids for dyspnea at the end of life. Phenothiazines and chlorpromazine may be helpful when combined with opioids. Benzodiazepines can be helpful in treating dyspnea, but only if anxiety is present. Benzodiazepines should not be used as first-line therapy or if there is no anxiety. If the patient has a history of COPD or asthma, inhaled bronchodilators and glucocorticoids may be helpful. If the patient has pulmonary edema due to heart failure, diuresis with a medication such as furosemide is indicated. Excess secretions can be transdermally or intravenously dried with scopolamine. More general interventions that medical staff can perform include sitting the patient upright, removing smoke or other irritants like perfume, ensuring a supply of fresh air with sufficient humidity, and minimizing other factors that can increase anxiety.

TABLE 13-6 Medications for the Management of Dyspnea

INTERVENTION	DOSE	COMMENTS
Weak opioids		For patients with mild dyspnea
Codeine (or codeine with 325 mg acetaminophen)	30 mg PO q4h	For opioid-naïve patients
Hydrocodone	5 mg PO q4h	
Strong opioids		For opioid-naïve patients with moderate to severe dyspnea
Morphine	5–10 mg PO q4h	For patients already taking opioids for pain or other symptoms
	30–50% of baseline opioid dose q4h	
Oxycodone	5–10 mg PO q4h	
Hydromorphone	1–2 mg PO q4h	
Anxiolytics		Give a dose every hour until the patient is relaxed; then provide a dose for maintenance
Lorazepam	0.5–2.0 mg PO/SL/IV qh then q4–6h	
Clonazepam	0.25–2.0 mg PO q12h	
Midazolam	0.5 mg IV q15min	

Fatigue • **FREQUENCY** Fatigue is one of the most commonly reported symptoms not only of cancer treatment but also of the palliative care of multiple sclerosis, COPD, heart failure, and HIV. More than 90% of terminally ill patients experience fatigue and/or weakness. Fatigue is frequently cited as one of the most distressing symptoms in these patients.

ETIOLOGY The multiple causes of fatigue in the terminally ill can be categorized as resulting from the underlying disease; from disease-induced factors such as tumor necrosis factor and other cytokines; and from secondary factors such as dehydration, anemia, infection, hypothyroidism, and drug side effects. In addition to low caloric intake, loss of muscle mass and changes in muscle enzymes may play an important role in fatigue during terminal illness. The importance of changes in the CNS, especially the reticular activating system, have been hypothesized based on reports of fatigue in patients receiving cranial radiation, experiencing depression, or having chronic pain in the absence of cachexia or other physiologic changes. Finally, depression and other causes of psychological distress can contribute to fatigue.

ASSESSMENT Like pain and dyspnea, fatigue is subjective, as it represents a patient's sense of tiredness and decreased capacity for physical work. Objective changes, even in body mass, may be absent. Consequently, assessment must rely on patient self-reporting. Scales used to measure fatigue, such as the Edmonton Functional Assessment Tool, the Fatigue Self-Report Scales, and the Rhoten Fatigue Scale, are usually appropriate for research but not clinical purposes. In clinical practice, a simple performance assessment such as the Karnofsky performance status or the Eastern Cooperative Oncology Group (ECOG)'s question "How much of the day does the patient spend in bed?" may be the best measure. In the ECOG 0–4 performance status assessment, 0 = normal activity; 1 = symptomatic without being bedridden; 2 = requiring some, but <50%, bed time; 3 = bedbound more than half the day; and 4 = bedbound all the time. Such a scale allows for assessment over time and correlates with overall disease severity and prognosis. A 2008 review by the European Association of Palliative Care also described several longer assessment tools that contained 9–20 items, including the Piper Fatigue Inventory, the Multidimensional Fatigue Inventory, and the Brief Fatigue Inventory (BFI).

INTERVENTIONS Reversible causes of fatigue, such as anemia and infection, should be treated. However, at the end of life, it must be realistically acknowledged that fatigue will not be "cured." The goal is to ameliorate fatigue and help patients and families adjust expectations. Behavioral interventions should be utilized to avoid blaming the patient for inactivity and to educate both the family and the patient that the underlying disease causes physiologic changes that produce low energy levels. Understanding that the problem is physiologic and not psychological can help alter expectations regarding the patient's level of physical activity. Practically, this may mean reducing routine activities such as housework, cooking, and social events outside the house and making it acceptable to receive guests while lying on a couch. At the same time, the implementation of exercise regimens and physical therapy can raise endorphins, reduce muscle wasting, and decrease the risk of depression. In addition, ensuring good hydration without worsening edema may help reduce fatigue. Sleep hygiene should also be emphasized, and effort should be taken to eliminate poor sleep hygiene behaviors. Discontinuing medications that worsen fatigue may help, including cardiac medications, benzodiazepines, certain antidepressants, or opioids if the patient's pain is well-controlled. As end-of-life care proceeds into its final stages, fatigue may protect patients from further suffering, and continued treatment could be detrimental.

Only a few pharmacologic interventions target fatigue and weakness. Randomized controlled trials suggest glucocorticoids can increase energy and enhance mood. Dexamethasone (8 mg/d) is preferred for its once-a-day dosing and minimal mineralocorticoid activity. Benefit, if any, is usually seen within the first month. For fatigue related to anorexia, megestrol (480–800 mg) can be helpful. Psychostimulants such as dextroamphetamine (5–10 mg PO) and methylphenidate (2.5–5 mg PO) may enhance energy levels, although controlled trials have not shown these drugs to be effective for fatigue induced by mild to moderate cancer. Doses should be given in the morning and at noon to minimize the risk of counterproductive insomnia. Modafinil and armodafinil, developed for narcolepsy, have had mixed results in the treatment of fatigue but have the advantage of once-daily dosing. Their precise role in fatigue at the end of life has not been documented but may be worth trying if other interventions are not beneficial.

■ PALLIATIVE SEDATION

Palliative sedation is used in distressing situations that cannot be addressed in other ways. When patients experience severe symptoms, such as pain or dyspnea, that cannot be relieved by conventional interventions or experience acute catastrophic symptoms, such as uncontrolled seizures, then palliative sedation should be considered as an intervention of last resort. It can be abused if done to hasten death (which it usually does not), when done at the request of the family rather than according to the patient's wishes, or when there are other interventions that could still be tried. The use of palliative sedation in cases of extreme existential or spiritual distress remains controversial. Typically, palliative sedation should be introduced only after the patient and family have been assured that all other interventions have been tried and after the patient and their loved ones have been able to "say goodbye."

Palliative sedation can be achieved by significantly increasing opioid doses until patients become unconscious and then putting them on a continuous infusion. Another commonly used medication for palliative sedation is midazolam at 1–5 mg IV every 5–15 min to calm the patient, followed by a continuous IV or subcutaneous infusion of 1 mg/h. In hospital settings, a continuous propofol infusion of 5 μg/kg per min can be used. There are also other, less commonly used medications for palliative sedation that include levomepromazine, chlorpromazine, and phenobarbital.

PSYCHOLOGICAL SYMPTOMS AND THEIR MANAGEMENT

Depression • **FREQUENCY AND IMPACT** Depression at the end of life presents an apparently paradoxical situation. Many people believe that depression is normal among seriously ill patients because they are dying. People frequently say, "Wouldn't you be depressed?" Although sadness, anxiety, anger, and irritability are normal responses to a serious condition, they are typically of modest intensity and transient. Persistent sadness and anxiety and the physically disabling symptoms that they can lead to are abnormal and suggestive of major depression. The precise number of terminally ill patients who are depressed is uncertain, primarily due to a lack of consistent diagnostic criteria and screening. Careful follow-up of patients suggests that while as many as 75% of terminally ill patients experience depressive symptoms, ~25% of terminally ill patients have major depression. Depression at the end of life is concerning because it can decrease the quality of life, interfere with closure in relationships and other separation work, obstruct adherence to medical interventions, and amplify the suffering associated with pain and other symptoms.

ETIOLOGY Previous history of depression, family history of depression or bipolar disorder, and prior suicide attempts are associated with increased risk for depression among terminally ill patients. Other symptoms, such as pain and fatigue, are associated with higher rates of depression; uncontrolled pain can exacerbate depression, and depression can cause patients to be more distressed by pain. Many medications used in the terminal stages, including glucocorticoids, and some anticancer agents, such as tamoxifen, interleukin 2, interferon α, and vincristine, also are associated with depression. Some terminal conditions, such as pancreatic cancer, certain strokes, and heart failure, have been reported to be associated with higher rates of depression, although this is controversial. Finally, depression may be attributable to grief over the loss of a role or function, social isolation, or loneliness.

ASSESSMENT Unfortunately, many studies suggest that most depressed patients at the end of life are not diagnosed, or if they are diagnosed, they are not properly treated. Diagnosing depression among seriously ill patients is complicated, as many of the vegetative symptoms in the *DSM-V* (*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*) criteria for clinical depression—insomnia, anorexia and weight loss, fatigue, decreased libido, and difficulty concentrating—are associated with the process of dying itself. The assessment of depression in seriously ill patients therefore should focus on the dysphoric mood, helplessness, hopelessness, and lack of interest, enjoyment, and concentration in normal activities. It is now recommended that patients near the end of life should be screened with either the PHQ-9 or the PHQ-2, which asks "Over the past 2 weeks, how often have you been bothered by any of the following problems? (1) Little interest or pleasure in doing things and (2) feeling down, depressed or hopeless." The answer categories are as follows: not at all, several days, more than half the days, and nearly every day. Other possible diagnostic tools include the short form of the Beck Depression Index or a visual analogue scale.

Certain conditions may be confused with depression. Endocrinopathies, such as hypothyroidism and Cushing's syndrome, electrolyte abnormalities, such as hypercalcemia, and akathisia, especially from dopamine-blocking antiemetics such as metoclopramide and prochlorperazine, can mimic depression and should be excluded.

INTERVENTIONS Undertreatment of depressed, terminally ill patients is common. Physicians must treat any physical symptom, such as pain, that may be causing or exacerbating depression. Fostering adaptation to the many losses that the patient is experiencing can also be helpful.

Unfortunately, there are few randomized trials to guide such interventions. Thus, treatment typically follows the treatment used for non–terminally ill depressed patients.

In the absence of randomized controlled trials, nonpharmacologic interventions, including group or individual psychological counseling, and behavioral therapies such as relaxation and imagery can be helpful, especially in combination with drug therapy.

Pharmacologic interventions remain at the core of therapy. The same medications are used to treat depression in terminally ill as in non–terminally ill patients. Psychostimulants may be preferred for patients with a poor prognosis or for those with fatigue or opioid-induced somnolence. Psychostimulants are comparatively fast-acting, working within a few days instead of the weeks required for selective serotonin reuptake inhibitors (SSRIs). Dextroamphetamine or methylphenidate should be started at 2.5–5.0 mg in the morning and at noon, the same starting doses used for treating fatigue. The doses can eventually be escalated up to 15 mg bid. Modafinil is started at 100 mg qd and can be increased to 200 mg if there is no effect at the lower dose. Pemoline is a nonamphetamine psychostimulant with minimal abuse potential. It is also effective as an antidepressant beginning at 18.75 mg in the morning and at noon. Because it can be absorbed through the buccal mucosa, it is preferred for patients with intestinal obstruction or dysphagia. If it is used for prolonged periods, liver function must be monitored. The psychostimulants can also be combined with more traditional antidepressants while waiting for the antidepressants to become effective, then tapered down after a few weeks if necessary. Psychostimulants have side effects, particularly initial anxiety, insomnia, and very rarely paranoia, which may necessitate lowering the dose or discontinuing treatment.

Mirtazapine, an antagonist at the postsynaptic serotonin receptors, is a promising psychostimulant. It should be started at 7.5 mg before bed and titrated up no more than once every 1–2 weeks to a maximal dose of 45 mg/d. It has sedating, antiemetic, and anxiolytic properties, with few drug interactions. Its side effect of weight gain may be beneficial for seriously ill patients; it is available in orally disintegrating tablets.

For patients with a prognosis of several months or longer, SSRIs, including fluoxetine, sertraline, paroxetine, escitalopram, and citalopram, and serotonin-noradrenaline reuptake inhibitors, such as venlafaxine and duloxetine, are the preferred treatments, due to their efficacy and comparatively few side effects. Because low doses of these medications may be effective for seriously ill patients, one should use half the usual starting dose as for healthy adults. The starting dose for fluoxetine is 10 mg once a day. In most cases, once-a-day dosing is possible. The choice of which SSRI to use should be driven by (1) the patient's past success or failure with the specific medication and (2) the most favorable side effect profile for that specific agent. For instance, for a patient in whom fatigue is a major symptom, a more activating SSRI (fluoxetine) would be appropriate. For a patient in whom anxiety and sleeplessness are major symptoms, a more sedating SSRI (paroxetine) would be appropriate. Importantly, it can take up to 4 weeks for these drugs to have an effect.

Atypical antidepressants are recommended only in select circumstances, usually with the assistance of a specialty consultation. Trazodone can be an effective antidepressant but is sedating and can cause orthostatic hypotension and, occasionally, priapism. Therefore, it should be used before bed and only when a sedating effect is desired and is often used for patients with insomnia at a dose starting at 25 mg. Bupropion can also be used. In addition to its antidepressant effects, bupropion is energizing, making it useful for depressed patients who experience fatigue. However, it can cause seizures, preventing its use for patients with a risk of CNS neoplasms or terminal delirium. Finally, alprazolam, a benzodiazepine, starting at 0.25–1.0 mg tid, can be effective in seriously ill patients who have a combination of anxiety and depression. Although it is potent and works quickly, it has many drug interactions and may cause delirium, especially among very ill patients, because of its strong binding to the benzodiazepine–γ-aminobutyric acid (GABA) receptor complex.

Unless used as adjuvants for the treatment of pain, tricyclic antidepressants are not recommended. While they can be effective, their therapeutic window and serious side effects typically limit their utility. Similarly, monoamine oxidase (MAO) inhibitors are not recommended because of their side effects and dangerous drug interactions.

Psychedelic drugs, such as psylocibin, have been tried for patients with cancer and major depressive disorder and shown promise. In smaller trials, psylocibin (a single dose of 25 mg) has been shown to be effective in up to 70% of patients. Importantly, psylocibin has a relatively rapid onset of relief of depressive symptoms—within 1 week—that can last for months. While still not U.S. Food and Drug Administration approved for general clinical use, there are multiple trials that may be available for patients.

Delirium (See Chap. 29) • **FREQUENCY** In the weeks or months before death, delirium is uncommon, although it may be significantly underdiagnosed. However, delirium becomes relatively common in the days and hours immediately before death. Up to 85% of patients dying from cancer may experience terminal delirium.

ETIOLOGY Delirium is a global cerebral dysfunction characterized by alterations in cognition and consciousness. It is frequently preceded by anxiety, changes in sleep patterns (especially reversal of day and night), and decreased attention. In contrast to dementia, delirium has an acute onset, is characterized by fluctuating consciousness and inattention, and is reversible, although reversibility may be more theoretical than real for patients near death. Delirium may occur in a patient with dementia; indeed, patients with dementia are more vulnerable to delirium.

Causes of delirium include metabolic encephalopathy arising from liver or renal failure, hypoxemia, or infection; electrolyte imbalances such as hypercalcemia; paraneoplastic syndromes; dehydration; and primary brain tumors, brain metastases, or leptomeningeal spread of tumor. Among dying patients, delirium is commonly caused by side effects of treatments, including radiation for brain metastases and medications, such as opioids, glucocorticoids, anticholinergic drugs, antihistamines, antiemetics, benzodiazepines, and chemotherapeutic agents. The etiology may be multifactorial; e.g., dehydration may exacerbate opioid-induced delirium.

ASSESSMENT Delirium should be recognized in any terminally ill patient exhibiting new onset of disorientation, impaired cognition, somnolence, fluctuating levels of consciousness, or delusions with or without agitation. Delirium must be distinguished from acute anxiety, depression, and dementia. The central distinguishing feature is altered consciousness, which usually is not noted in anxiety, depression, or dementia. Although "hyperactive" delirium, characterized by overt confusion and agitation, is probably more common, patients should also be assessed for "hypoactive" delirium, which is characterized by sleep-wake reversal and decreased alertness.

In some cases, use of formal assessment tools such as the Mini-Mental Status Examination (which does not distinguish delirium from dementia) and the Delirium Rating Scale (which does distinguish delirium from dementia) may be helpful in distinguishing delirium from other processes. The patient's list of medications must be evaluated carefully. Nonetheless, a reversible etiologic factor for delirium is found in fewer than half of all terminally ill patients. Given that most terminally ill patients experiencing delirium are very close to death and often at home, extensive diagnostic evaluations such as lumbar punctures and neuroradiologic examinations are inappropriate.

INTERVENTIONS One of the most important objectives of terminal care is to provide terminally ill patients the lucidity to say goodbye to the people they love. Delirium, especially when in combination with agitation during the final days, is distressing to family and caregivers. A strong determinant of bereavement difficulties is witnessing a difficult death. Thus, terminal delirium should be treated aggressively.

At the first sign of delirium, such as day-night reversal with slight changes in mentation, the physician should let the family members know that it is time to be sure that everything they want to say has been said. The family should be informed that delirium is common just before death.

If medications are suspected of being a cause of the delirium, unnecessary agents should be discontinued. Other potentially reversible

TABLE 13-7 Medications for the Management of Delirium

INTERVENTIONS	DOSE
Neuroleptics	
Haloperidol	0.5–5 mg q2–12h, PO/IV/SC/IM
Thioridazine	10–75 mg q4–8h, PO
Chlorpromazine	12.5–50 mg q4–12h, PO/IV/IM
Atypical neuroleptics	
Olanzapine	2.5–5 mg qd or bid, PO
Risperidone	1–3 mg q12h, PO
Anxiolytics	
Lorazepam	0.5–2 mg q1–4h, PO/IV/IM
Midazolam	1–5 mg/h continuous infusion, IV/SC
Anesthetics	
Propofol	0.3–2.0 mg/h continuous infusion, IV

causes, such as constipation, urinary retention, and metabolic abnormalities, should be treated. Supportive measures aimed at providing a familiar environment should be instituted, including restricting visits only to individuals with whom the patient is familiar and eliminating new experiences; orienting the patient, if possible, by providing a clock and calendar; and gently correcting the patient's hallucinations or cognitive mistakes.

Pharmacologic management focuses on the use of neuroleptics and, in extreme cases, anesthetics **(Table 13-7)**. Haloperidol remains the first-line therapy. Usually, patients can be controlled with a low dose (1–3 mg/d), given every 6 h, although some may require as much as 20 mg/d. Haloperidol can be administered PO, SC, or IV. IM injections should not be used, except when this is the only way to address a patient's delirium. Olanzapine, an atypical neuroleptic, has shown significant effectiveness in completely resolving delirium in cancer patients. It also has other beneficial effects for terminally ill patients, including antinausea, antianxiety, and weight gain. Olanzapine is useful for patients with longer anticipated life expectancies because it is less likely to cause dysphoria and has a lower risk of dystonic reactions. Additionally, because olanzapine is metabolized through multiple pathways, it can be used in patients with hepatic and renal dysfunction. Olanzapine has the disadvantage that it is only available orally and takes a week to reach steady state. The usual dose is 2.5–5 mg PO bid. Chlorpromazine (10–25 mg every 4–6 h) can be useful if sedation is desired and can be administered IV or PR in addition to PO. Dystonic reactions resulting from dopamine blockade are a side effect of neuroleptics, although they are reported to be rare when these drugs are used to treat terminal delirium. If patients develop dystonic reactions, benztropine should be administered. Neuroleptics may be combined with lorazepam to reduce agitation when the delirium is the result of alcohol or sedative withdrawal.

If no response to first-line therapy is observed, a specialty consultation should be obtained with a goal to change to a different medication. If the patient fails to improve after a second neuroleptic, sedation with either an anesthetic such as propofol or continuous-infusion midazolam may be necessary. By some estimates, as many as 25% of patients at the very end of life who experience delirium, especially restless delirium with myoclonus or convulsions, may require sedation.

Physical restraints should be used with great reluctance and only when patients' violence is threatening to themselves or others. If restraints are used, their appropriateness should be frequently reevaluated.

Insomnia • **FREQUENCY** Sleep disorders, defined as difficulty initiating sleep or maintaining sleep, sleep difficulty at least 3 nights a week, or sleep difficulty that causes impairment of daytime functioning, occurs in 19–63% of patients with advanced cancer. Some 30–74% of patients with other end-stage conditions, including HIV/AIDS, heart disease, COPD, and renal disease, experience insomnia.

ETIOLOGY Patients with cancer may experience changes in sleep efficiency, such as an increase in stage I sleep. Insomnia may also coexist with both physical illnesses, like thyroid disease, and psychological illnesses, like depression and anxiety. Medications, including antidepressants, psychostimulants, glucocorticoids, and β agonists, are significant contributors to sleep disorders, as are caffeine and alcohol. Multiple over-the-counter medications contain caffeine and antihistamines, which can contribute to sleep disorders.

ASSESSMENT Assessments should include specific questions concerning sleep onset, sleep maintenance, and early-morning wakening, as these will provide clues to both the causative agents and management of insomnia. Patients should be asked about previous sleep problems, screened for depression and anxiety, and asked about symptoms of thyroid disease. Caffeine and alcohol are prominent causes of sleep problems, and a careful history of the use of these substances should be obtained. Both excessive use and withdrawal from alcohol can be causes of sleep problems.

INTERVENTIONS The mainstays of any intervention include improvement of sleep hygiene (encouragement of regular time for sleep, decreased nighttime distractions, and elimination of caffeine and other stimulants and alcohol), interventions to treat anxiety and depression, and treatment for the insomnia itself. For patients with depression who have insomnia and anxiety, a sedating antidepressant such as mirtazapine can be helpful. In the elderly, trazodone, beginning at 25 mg at nighttime, is an effective sleep aid at doses lower than those that cause its antidepressant effect. Zolpidem may have a decreased incidence of delirium in patients compared with traditional benzodiazepines, but this has not been clearly established. When benzodiazepines are prescribed, short-acting ones (such as lorazepam) are favored over longer-acting ones (such as diazepam). Patients who receive these medications should be observed for signs of increased confusion and delirium.

■ SOCIAL NEEDS AND THEIR MANAGEMENT

Financial Burdens • **FREQUENCY** Dying can impose substantial economic strains on patients and families, potentially causing distress. This is known as financial toxicity. In the United States, which has the least comprehensive health insurance systems among wealthy countries, a fifth to a half of families coping with end-stage cancer report that care was a major financial burden. Among caregivers of Medicare beneficiaries with cancer, average out-of-pocket costs within the last year of life ranged from $4,645 to $10,547 (14–28% household income). Between 10 and 30% of families are forced to sell assets, use savings, or take out a mortgage to pay for the patient's health care costs, while around 40% of primary caregivers felt forced to quit their jobs due to caregiving requirements.

The patient is likely to reduce hours worked and eventually stop working altogether. In 20% of cases, a family member of the terminally ill patient also must stop working to provide care. The major underlying causes of economic burden are related to poor physical functioning and care needs, such as the need for housekeeping, nursing, and personal care. More debilitated patients and poor patients experience greater economic burdens.

INTERVENTION The economic burden of end-of-life care should not be ignored as a private matter. It has been associated with a number of adverse health outcomes, including preferring comfort care over life-prolonging care, as well as consideration of euthanasia or physician-assisted suicide (PAS). Economic burdens increase the psychological distress of the families and caregivers of terminally ill patients, and poverty is associated with many adverse health outcomes. Importantly, studies have found that "patients with advanced cancer who reported having end-of-life conversations with physicians had significantly lower health care costs in their final week of life. Higher costs were associated with worse quality of death." Assistance from a social worker, early on if possible, to ensure access to all available benefits may be helpful. Many patients, families, and health care providers are unaware of options for long-term care insurance, respite care, the Family Medical Leave Act (FMLA), and other sources of assistance. Some of these options (such as respite care) may be part of a formal hospice program, but others (such as the FMLA) do not require enrollment in a hospice program.

Relationships • **FREQUENCY** Settling personal issues and closing the narrative of lived relationships are universal needs. When asked if sudden death or death after an illness is preferable, respondents often initially select the former, but soon change to the latter as they reflect on the importance of saying goodbye. Bereaved family members who have not had the chance to say goodbye often have a more difficult grief process.

INTERVENTIONS Care of seriously ill patients requires efforts to facilitate the types of encounters and time spent with family and friends that are necessary to meet those needs. Family and close friends may need to be accommodated in hospitals and other facilities with unrestricted visiting hours, which may include sleeping near the patient, even in otherwise regimented institutional settings. Physicians and other health care providers may be able to facilitate and resolve strained interactions between the patient and other family members. Assistance for patients and family members who are unsure about how to create or help preserve memories, whether by providing materials such as a scrapbook or memory box or by offering them suggestions and informational resources, can be deeply appreciated. Taking photographs and creating videos can be especially helpful to terminally ill patients who have younger children or grandchildren.

Family Caregivers • **FREQUENCY** Caring for seriously ill patients places a heavy burden on families. Families are frequently required to provide transportation and homemaking, as well as other services. Typically, paid professionals, such as home health nurses and hospice workers, supplement family care; only about a quarter of all caregiving consists of exclusively paid professional assistance. Over the past 40 years, there has been a significant decline in the United States of deaths occurring in hospitals, with a simultaneous increase in deaths in other facilities and at home. Over a third of deaths occur in patients' homes despite 70% of Americans expressing a preference to die at home. This increase in out-of-hospital deaths increases reliance on families for end-of-life care. Increasingly, family members are being called upon to provide physical care (such as moving and bathing patients) and medical care (such as assessing symptoms and giving medications) in addition to emotional care and support.

Three-quarters of family caregivers of terminally ill patients are women—wives, daughters, sisters, and even daughters-in-law. Since many are widowed, women tend to be able to rely less on family for caregiving assistance and may need more paid assistance. About 20% of terminally ill patients report substantial unmet needs for nursing and personal care. The impact of caregiving on family caregivers is substantial: both bereaved and current caregivers have a higher mortality rate than that of non-caregiving controls.

INTERVENTIONS It is imperative to inquire about unmet needs and to try to ensure that those needs are met either through the family or by paid professional services when possible. Community assistance through houses of worship or other community groups often can be mobilized by telephone calls from the medical team to someone the patient or family identifies. Sources of support specifically for family caregivers should be identified through local sources or nationally through groups such as the National Family Caregivers Association (*www.nfcacares.org*), the American Cancer Society (*www.cancer.org*), and the Alzheimer's Association (*www.alz.org*).

EXISTENTIAL NEEDS AND THEIR MANAGEMENT

Frequency Religion and spirituality are often important to dying patients. Nearly 70% of patients report becoming more religious or spiritual when they became terminally ill, and many find comfort in religious or spiritual practices such as prayer. However, ~20% of terminally ill patients become less religious, frequently feeling cheated or betrayed by becoming terminally ill. For other patients, the need is for existential meaning and purpose that is distinct from, and may even be antithetical to, religion or spirituality. When asked, patients and family caregivers frequently report wanting their professional caregivers to be more attentive to religion and spirituality.

Assessment Health care providers are often hesitant about involving themselves in the religious, spiritual, and existential experiences of their patients because it may seem private or not relevant to the current illness. But physicians and other members of the care team should be able at least to detect spiritual and existential needs. Screening questions have been developed for a physician's spiritual history taking. Spiritual distress can amplify other types of suffering and even masquerade as intractable physical pain, anxiety, or depression. The screening questions in the comprehensive assessment are usually sufficient. Deeper evaluation and intervention are rarely appropriate for the physician unless no other member of a care team is available or suitable. Pastoral care providers may be helpful, whether from the medical institution or from the patient's own community.

Interventions Precisely how religious practices, spirituality, and existential explorations can be facilitated and improve end-of-life care is not well established. What is clear is that for physicians, one main intervention is to inquire about the role and importance of spirituality and religion in a patient's life. This will help a patient feel heard and help physicians identify specific needs. In one study, only 36% of respondents indicated that a clergy member would be comforting. Nevertheless, the increase in religious and spiritual interest among a substantial fraction of dying patients suggests inquiring of individual patients how this need can be addressed. Some evidence supports specific methods of addressing existential needs in patients, ranging from establishing a supportive group environment for terminal patients to individual treatments emphasizing a patient's dignity and sources of meaning.

MANAGING THE LAST STAGES

PALLIATIVE CARE SERVICES: HOW AND WHERE

Determining the best approach to providing palliative care to patients will depend on patient preferences, the availability of caregivers and specialized services in close proximity, institutional resources, and reimbursement. Hospice is a leading, but not the only, model of palliative care services. In the United States, slightly more than a third—35.7%—of hospice care is provided in private residential homes, with 14.5% of hospice care in nursing homes. In the United States, Medicare pays for hospice services under Part A, the hospital insurance part of reimbursement. Two physicians must certify that the patient has a prognosis of ≤6 months if the disease runs its usual course. Prognoses are probabilistic by their nature; patients are not required to die within 6 months but rather to have a condition from which half the individuals with it would not be alive within 6 months. While many have advocated for allowing hospice patients to continue with curative treatments, Medicare does not cover hospice services if patients are receiving curative therapy related to their terminal illness. However, hospice patients can still receive medical services for other comorbid conditions. Patients also can withdraw enrollment and reenroll later; the hospice Medicare benefit can be revoked later to secure traditional Medicare benefits. Payments to the hospice are per diem (or capitated), not fee-for-service. Payments are intended to cover physician services for the medical direction of the care team; regular home care visits by registered nurses and licensed practical nurses; home health aide and homemaker services; chaplain services; social work services; bereavement counseling; and medical equipment, supplies, and medications. No specific therapy is excluded, and the goal is for each therapy to be considered for its symptomatic (as opposed to disease-modifying) effect. Additional clinical care, including services of the primary physician, is covered by Medicare Part B even while the hospice Medicare benefit is in place.

The Affordable Care Act directed the secretary of Health and Human Services to gather data on Medicare hospice reimbursement with the goal of reforming payment rates to account for resource use over an entire episode of care. The legislation also required additional evaluations and reviews of eligibility for hospice care by hospice physicians or nurses. The Center for Medicare and Medicaid Innovation (CMMI) sponsors and carries out demonstration projects to test models and evaluate the potential of new methods. Between 2012 and 2021, CMMI carried out four palliative care demonstration projects that demonstrated improvement in patient and caregiver experience of care and quality of life. For instance, the Medicare Care Choices Model (MCCM) that ran from 2016 to 2021 offered participants supportive services without forgoing payment for curative services and found that offering supportive services reduced expenditures per beneficiary by 14% ($7,254 per capita), driven by a reduction in utilization of resource-intensive services at the end of life, notably a 38% reduction in ICU days, a 26% reduction in inpatient admissions, and a 14% reduction in emergency department visits. The Medicare Advantage (MA) Value-Based Insurance Design (VBID) Model, which was started in 2021 and extended through 2030, is currently testing the inclusion of hospice in MA, which covers 8% of the market and includes important health plans.

By 2021, the average length of enrollment in a hospice for Medicare beneficiaries was 92 days. However, the median length of stay was just 17 days, suggesting most patients are in hospice for a short time. In fact, over a third of patients referred to hospice die within 7 days of arriving and 1 in 10 die within 72 h. Around 5% of patients live longer than 6 months, with patients with neurologic diagnoses (155 days), COPD (140 days), and heart/circulatory disease (104 days) having the longest average lengths of hospice. Such short stays create barriers to establishing high-quality palliative services in patients' homes and also place financial strains on hospice providers since the initial assessments are resource intensive. Physicians should initiate early referrals to the hospice to allow more time for patients to receive palliative care.

Several cognitive biases predispose physicians to referring patients to hospice services late. Studies have shown that when asked to prognosticate, physicians are both inaccurate and systematically optimistic. When asked to predict the number of days patients referred to hospice would survive, the referring providers overestimated prognosis by a factor of 5.3. Only one in five predictions was within 33% of actual survival, while two-thirds were overestimated by over a third. Notably, the longer physicians had been with their patient and the more recently they had seen them, the less accurate their predictions were, suggesting that physicians overestimate how long patients will live the better they knew them.

In response to these innate biases, the "surprise question" was developed as a clinical heuristic. Using this tool, instead of asking "Do I expect this patient to do X in the next Y period of time," the clinician instead asks the question "Would I *be surprised* if this patient did X in the next Y period of time?" While the "surprise question" was originally studied to predict 12-month mortality to aid in referral to palliative care in the primary care setting, it has since been validated in trials in a variety of settings, including dialysis patients, cancer patients, overall 1-year mortality, and overall 1-month mortality.

In the United States, hospice care has been the main method for securing palliative services for terminally ill patients. However, leading physicians have increasingly emphasized the need to introduce palliative care much earlier in a patient's illness, and efforts are being made to develop palliative care services that can be provided before the last 6 months of life and across a variety of settings. Studies of terminally ill patients indicate that those who received in-home palliative care delivered by an interdisciplinary team compared to usual care were more satisfied, more likely to die at home, and had fewer visits to the emergency room and lower per-day costs. More companies and home health agencies are now offering nonhospice palliative care services in patients' homes in an effort to increase quality of life and forestall emergency room visits and hospitalizations. Similarly, palliative care services are increasingly available via consultation, rather than being available only in hospital, day care, outpatient, and nursing home settings. Palliative care consultations for nonhospice patients can be billed as for other consultations under Medicare Part B. It is argued that using palliative care earlier in a patient's illness allows patients and family members to become more acculturated to avoiding life-sustaining treatments, facilitating a smoother transition to hospice care closer to death.

WITHDRAWING AND WITHHOLDING LIFE-SUSTAINING TREATMENT

Legal Aspects For centuries, it has been deemed ethical to withhold or withdraw life-sustaining interventions. The current legal consensus in the United States and most wealthy countries is that patients have a moral as well as legal right to refuse medical interventions. American courts also have held that incompetent patients have a right to refuse medical interventions. For patients who are incompetent and terminally ill and who have not completed an advance care directive, next of kin can exercise that right, although this may be restricted in some states, depending on how clear and convincing the evidence is of the patient's preferences. Courts have limited families' ability to terminate life-sustaining treatments in patients who are conscious and incompetent but not terminally ill. In theory, a patient's right to refuse medical therapy can be limited by four countervailing interests: (1) preservation of life, (2) prevention of suicide, (3) protection of third parties such as children, and (4) preservation of the integrity of the medical profession. In practice, these interests almost never override the right of competent patients and incompetent patients who have left explicit wishes or advance care directives.

For incompetent patients who either appointed a proxy without specific indications of their wishes or never completed an advance care directive, three criteria have been suggested to guide the decision to terminate medical interventions. First, some commentators suggest that ordinary care should be administered but extraordinary care could be terminated. Because the ordinary/extraordinary distinction is too vague, courts and commentators widely agree that it should not be used to justify decisions about stopping treatment. Second, many courts have advocated the use of the substituted-judgment criterion, which holds that the proxy decision-makers should try to imagine what the incompetent patient would do if they were competent. However, multiple studies indicate that many proxies, even close family members, cannot accurately predict what the patient would have wanted. Therefore, substituted judgment becomes more of a guessing game than a way of fulfilling the patient's wishes. Finally, the best-interests criterion holds that proxies should evaluate treatments by balancing their benefits and risks and select those treatments where the benefits maximally outweigh the burdens of treatment. Clinicians have a clear and crucial role in this by accurately and dispassionately explaining the known benefits and burdens of specific treatments. Yet even when that information is as clear as possible, different individuals can have very different views of what is in the patient's best interests, and families may have disagreements or even overt conflicts. This criterion has been criticized because there is no single way to determine the balance between benefits and burdens; it depends on a patient's personal values. For instance, for some people, being alive even if mentally incapacitated is a benefit, whereas for others, it may be the worst possible existence. As a matter of practice, physicians rely on family members to make decisions that they feel are best and object only if those decisions seem to demand treatments that the physicians consider not beneficial.

Practices Withholding and withdrawing acutely life-sustaining medical interventions from terminally ill patients are now standard practice. More than 90% of American patients die without cardiopulmonary resuscitation (CPR), and just as many forgo other potentially life-sustaining interventions. For instance, in ICUs in the period of 1987–1988, CPR was performed 49% of the time, but it was performed only 10% of the time in 1992–1993 and on just 1.8% of admissions from 2001 to 2008. On average, 3.8 interventions, such as vasopressors and transfusions, were stopped for each dying ICU patient. However, up to 19% of decedents in hospitals received interventions such as extubation, ventilation, and surgery in the 48 h preceding death. There is wide variation in practices among hospitals and ICUs, suggesting an important element of physician preferences rather than consistent adherence to professional society recommendations.

Mechanical ventilation may be the most challenging intervention to withdraw. The two approaches are *terminal extubation*, which is the removal of the endotracheal tube, and *terminal weaning*, which is the gradual reduction of the fraction of inspired oxygen (FIO_2) or ventilator rate. One-third of ICU physicians prefer to use the terminal weaning technique, and 13% extubate; the majority of physicians utilize both techniques. The American Thoracic Society's 2008 clinical policy guidelines note that there is no single correct process of ventilator withdrawal and that physicians use and should be proficient in both methods but that the chosen approach should carefully balance benefits and burdens as well as patient and caregiver preferences. Some recommend terminal weaning because patients do not develop upper airway obstruction and the distress caused by secretions or stridor; however, terminal weaning can prolong the dying process and not allow a patient's family to be with the patient unencumbered by an endotracheal tube. To ensure comfort for conscious or semiconscious patients before withdrawal of the ventilator, neuromuscular blocking agents should be terminated and sedatives and analgesics administered. Removing the neuromuscular blocking agents permits patients to show discomfort, facilitating the titration of sedatives and analgesics; it also permits interactions between patients and their families. A common practice is to inject a bolus of midazolam (2–4 mg) or lorazepam (2–4 mg) before withdrawal, followed by a bolus of 5–10 mg of morphine and continuous infusion of morphine (50% of the bolus dose per hour) during weaning. In patients who have significant upper airway secretions, IV scopolamine at a rate of 100 μg/h can be administered. Additional boluses of morphine or increases in the infusion rate should be administered for respiratory distress or signs of pain. Higher doses will be needed for patients already receiving sedatives and opioids.

The median time to death after stopping of the ventilator is 1 h. However, up to 10% of patients unexpectedly survive for 1 day or more after mechanical ventilation is stopped. Women and older patients tend to survive longer after extubation. Families need to be reassured about both the continuations of treatments for common symptoms, such as dyspnea and agitation, after withdrawal of ventilatory support and the uncertainty of length of survival after withdrawal of ventilatory support.

FUTILE CARE

Beginning in the late 1980s, some commentators argued that physicians could terminate futile treatments demanded by the families of terminally ill patients. Although no objective definition or standard of futility exists, several categories have been proposed. Physiologic futility means that an intervention will have no physiologic effect. Some have defined qualitative futility as applying to procedures that "fail to end a patient's total dependence on intensive medical care." Quantitative futility occurs "when physicians conclude (through personal experience, experiences shared with colleagues, or consideration of reported empiric data) that in the last 100 cases, a medical treatment has been useless." The term conceals subjective value judgments about when a treatment is "not beneficial." Deciding whether a treatment that obtains an additional 6 weeks of life or a 1% survival advantage confers benefit depends on patients' preferences and goals. Furthermore, physicians' predictions of when treatments are futile deviate markedly from the quantitative definition. When residents thought CPR was quantitatively futile, more than one in five patients had a >10% chance of survival to hospital discharge. Most studies that purport to guide determinations of futility are based on insufficient data and therefore cannot provide statistical confidence for clinical decision-making. Quantitative futility rarely applies in ICU settings.

Many commentators reject using futility as a criterion for withdrawing care, preferring instead to consider futility situations as ones that represent conflict that calls for careful negotiation between families and health care providers. Importantly, the American Medical Association clearly states that "respecting patient autonomy does not mean that patients should receive specific interventions simply because they (or their surrogates) request them." That is, ethically, physicians can refuse patient or caregiver requests for treatments. The American Medical Association and other professional societies have developed process-based approaches to resolving cases clinicians feel are futile. These process-based measures mainly suggest involving consultants and/or ethics committees when there are seemingly irresolvable differences. Some hospitals have enacted "unilateral do-not-resuscitate" policies to allow clinicians to provide a do-not-resuscitate order in cases in which consensus cannot be reached with families and medical opinion is that resuscitation would be futile if attempted. This type of a policy is not a replacement for careful and patient communication and negotiation but recognizes that agreement cannot always be reached.

In 1999, Texas enacted the so-called Futile Care Act. Other states, such as Virginia, Maryland, and California, have also enacted such laws that provide physicians a "safe harbor" from liability if they refuse a patient's or family's request for life-sustaining interventions. It appears that 19 state laws protect a physician's futility judgment. For instance, in Texas, when a disagreement about terminating interventions between the medical team and the family has not been resolved by an ethics consultation, the physician is tasked with trying to facilitate transfer of the patient to an institution willing to provide treatment. If this fails after 10 days, the hospital and physician may unilaterally withdraw treatments determined to be futile. The family may appeal to a state court. Data suggest that the law increases futility consultations for the ethics committee and that, although most families concur with withdrawal, ~10–15% of families refuse to withdraw treatment. As of 2007, there had been 974 ethics committee consultations on medical futility cases and 65 in which committees ruled against families and gave notice that treatment would be terminated. In 2007, a survey of Texas hospitals showed that 30% of hospitals had used the futility law in 213 adult cases and 42 pediatric cases. Treatment was withdrawn for 27 of those patients, and the remainder were transferred to other facilities or died while awaiting transfer.

EUTHANASIA AND PHYSICIAN-ASSISTED SUICIDE

Euthanasia and PAS are defined in Table 13-8. Terminating life-sustaining care and providing opioid medications to manage symptoms such as pain or dyspnea have long been considered ethical by the medical profession and legal by courts and should not be conflated with euthanasia or PAS.

Some advocates, especially in the United States, urge the use of the term physician-assisted death (PAD), rather than physician-assisted suicide (PAS), on the idea that it is not suicide. This is not accepted in many countries, and the terms themselves remain under dispute and are used inconsistently. For instance, in the Netherlands, PAD applies to terminating life-sustaining treatments, such as a ventilator, not providing pills to a patient who takes them to end their life. In the United Kingdom, PAD applies to prescribing life-ending interventions to terminally ill patients, while PAS applies to prescribing life-ending interventions to patients who are not necessarily terminally ill. Thus,

TABLE 13-8 Definitions of Physician-Assisted Suicide and Euthanasia

TERM	DEFINITION	LEGAL STATUS
Voluntary active euthanasia	Intentionally administering medications or other interventions to cause the patient's death with the patient's informed consent	Netherlands, Belgium, Luxembourg, Canada, Colombia, Spain, Western Australia, New Zealand
Involuntary active euthanasia	Intentionally administering medications or other interventions to cause the patient's death when the patient was competent to consent but did not—e.g., the patient may not have been asked	Nowhere
Passive euthanasia	Withholding or withdrawing life-sustaining medical treatments from a patient to let the patient die (terminating life-sustaining treatments)	Everywhere
Physician-assisted suicide	A physician provides medications or other interventions to a patient with the understanding that the patient can use them to commit suicide	Netherlands, Belgium, Luxembourg, Canada, Colombia, Germany, Switzerland, Oregon, Washington, Montana, Vermont, California, Colorado, District of Columbia, Hawaii, Maine, New Jersey, New Mexico

in many countries, PAD is a misnomer if it is meant to replace or cover PAS. That is why we use the term PAS.

Legal Aspects Euthanasia and PAS are legal in the Netherlands, Belgium, Luxembourg, Colombia, Canada, Spain, Portugal, Western Australia, and New Zealand. Euthanasia was legalized in the Northern Territory of Australia in 1996, but that legislation was repealed 9 months later in 1997. Under certain conditions, a layperson in Switzerland or Germany can legally elect assisted suicide. In the United States, PAS is legal in Washington, D.C., and 10 states: Oregon, Washington State, Montana, Vermont, California, Colorado, Hawaii, Maine, New Jersey, and New Mexico. No state in the United States has legalized euthanasia. In 2009, the state supreme court of Montana ruled that state law permits PAS for terminally ill patients. In 2021 and 2023, respectively, Spain and Portugal became the most recent countries to legalize both euthanasia and PAS. Many other countries are actively debating the legalization of euthanasia and/or PAS.

In the United States, multiple criteria must be met for PAS: the patient must have a terminal condition of <6 months and must be determined eligible through a process that includes a 15-day waiting period. Many countries, such as the Netherlands, do not require patients to be terminally ill to obtain euthanasia or PAS. Instead, the key criterion is that patients must have “unbearable suffering” regardless of whether they have a terminal illness. In Canada, patients requesting euthanasia or PAS need not be terminally ill. Canadian law allows patients with chronic conditions to receive euthanasia or PAS.

Practices Fewer than 10–20% of terminally ill patients actually consider euthanasia and/or PAS for themselves. Use of euthanasia and PAS is increasing but remains relatively rare. In all countries, even the Netherlands and Belgium where these practices have been tolerated and legal for many years, at most 5% of deaths occur by euthanasia or PAS. As of the most recent data in the Netherlands, 5.1% of all deaths were by euthanasia or PAS (2022) and 2.5% of all deaths were by euthanasia in Belgium (2013). Canada enacted its Medical Assistance in Dying Laws <8 years ago and expanded which patients could be eligible in 2021. In 1 year, the use of euthanasia or PAD increased 30% with >4% of all deaths in Canada being euthanasia or PAS. Conversely, after 25 years of enactment, just 0.82% of all deaths in Oregon in 2023 (367 of 44,681 deaths) and 0.53% of all deaths in Washington State in 2022 (363 of 69,116 deaths) were reported to be by PAS, although these may be underestimates since the cause of some deaths of patients who received medications could not be verified.

In Belgium, the Netherlands, Oregon, and Washington, >70% of patients utilizing these interventions are dying of cancer; <10% of deaths by euthanasia or PAS involve patients with AIDS or amyotrophic lateral sclerosis. While the numbers are small, in the Netherlands, the numbers of euthanasia or PAS cases in patients with psychiatric disorders, dementia, and the accumulation of health issues are increasing.

Pain is not the primary motivator for patients’ requests for or interest in euthanasia and/or PAS. Among the first patients to receive PAS in Oregon, only 1 of the 15 patients had inadequate pain control, compared with 15 of the 43 patients in a control group who experienced inadequate pain relief. In 2022, 31% of patients in Oregon seeking PAS currently cite pain or fear of pain as their main reason for doing so compared to 46% in Washington. Concerningly, the percentage of patients who cite inadequate pain control has increased over the past 15 years, with only 10% of patients in Oregon and 25% of patients in Washington citing inadequate pain control as a motivating factor in 2009. Conversely, depression and hopelessness are strongly associated with patient interest in euthanasia and PAS. Losing autonomy (86% Oregon [OR], 83% Washington [WA]), not being able to enjoy activities (89% OR, 83% WA), and fear of losing dignity (62% OR, 69% WA) are the most-cited end-of-life concerns in both states. Importantly, a recent study of 25 years of experience in Oregon reported that “there was an increase in patients feeling a burden and describing financial concerns as reasons for choosing an assisted death.” A study from the Netherlands showed that depressed terminally ill cancer patients were four times more likely to request euthanasia and confirmed that uncontrolled pain was not associated with greater interest in euthanasia.

Importantly, a growing number of cases in Oregon appear to be transactional rather than arising from a long-standing physician–patient relationship. Over 25 years, there has been a substantial decline in physician–patient relationship before the doctor provides medications, from an average of 18 weeks to just 5 weeks. Moreover, <1% of patients are referred for formal psychiatric assessment before prescription.

Similarly, in Canada, where >4.1% of all deaths are accounted for by medically assisted suicide, the two most common motivating factors for assisted suicide are loss of ability to engage in meaningful life activities (86%) and loss of ability to perform activities of daily living (82%). Again, concerningly, 59% cited inadequate pain control, 47% cited inadequate symptom control, and 35% cited feeling like they were a burden on their family, friends, or caregivers as motivating factors. In addition, cases of health care providers not just responding to patients’ inquiries for euthanasia or PAS but actively encouraging them to use these interventions have been reported.

Euthanasia and PAS are no guarantee of a painless, quick death. Data from the Netherlands indicate that in as many as 20% of euthanasia and PAS cases, technical and other problems arose, including patients waking from coma, not becoming comatose, regurgitating medications, and experiencing a prolonged time to death. Data from Oregon between 1998 and 2017 and Washington between 2009 and 2017 indicate that of patients who received PAS prescriptions, 81% died at home and prescribers were present in 9.7% of cases. The time between drug intake and coma ranged from 1 min to 11 h, and the time from drug intake to death ranged from 1 min to 104 h. The median time from ingestion to coma was 5 min and from ingestion to death was 25 min. In Oregon between 1998 and 2015, 53% of patients had no complications, 44% of patients had no data on complications, and 2.4% of patients had regurgitation after taking the prescribed medicine as the only complication. In addition, six patients awakened. In Washington State between 2014 and 2015, 1.4% of patients had regurgitation, one patient had a seizure, and the reported range of time to death extended to 30 h. In the Netherlands, problems were significantly more common in PAS, sometimes requiring the physician to intervene and provide euthanasia.

Regardless of whether they practice in a setting where euthanasia is legal or not, many physicians over the course of their careers will receive a patient request for euthanasia or PAS. In the United States, 18% of physicians have received a request for PAS and 11% have received a request for euthanasia. Three percent complied with a request for PAS, while 5% complied with a request for euthanasia. In the Netherlands, where the practices are legal, 77% of physicians have received a request for PAS or euthanasia and 60% have performed these interventions. In Oregon in 2022, 167 physicians participated in PAS out of a total of ~6300 (2.6%).

Competency in dealing with such a request is crucial. Although challenging, the request can also provide a chance to address intense suffering. After receiving a request for euthanasia and/or PAS, health care providers should carefully clarify the request with empathic, open-ended questions to help elucidate the underlying cause for the request, such as, “What makes you want to consider this option?” Endorsing either moral opposition or moral support for the act tends to be counterproductive, giving an impression of being judgmental or of endorsing the idea that the patient’s life is worthless. Health care providers must reassure the patient of continued care and commitment. The patient should be educated about alternative; less laden options, such as symptom management and withdrawing any unwanted treatments; and the reality of euthanasia and/or PAS, since the patient may have misconceptions about their effectiveness as well as the legal implications of the choice. Depression, hopelessness, and other symptoms of psychological distress, as well as physical suffering and economic burdens, are likely factors motivating the request, and such factors should be assessed and treated aggressively. After these interventions and clarification of options, most patients proceed with another approach, declining life-sustaining interventions, possibly including refusal of nutrition and hydration.

CARE DURING THE LAST HOURS

Most laypersons have limited experiences with the actual dying process and death. They frequently do not know what to expect of the final hours and afterward. The family and other caregivers must be prepared, especially if the plan is for the patient to die at home.

Patients in the last days of life typically experience extreme weakness and fatigue and become bedbound; this can lead to pressure sores. The issue of turning patients who are near the end of life, however, must be balanced against the potential discomfort that movement may cause. Patients stop eating and drinking with drying of mucosal membranes and dysphagia. Careful attention to oral swabbing, lubricants for lips, and use of artificial tears can provide a form of care to substitute for attempts at feeding the patient. With loss of the gag reflex and dysphagia, patients may also experience accumulation of oral secretions, producing noises during respiration sometimes called "the death rattle." Scopolamine can reduce the secretions. Patients also experience changes in respiration with periods of apnea or Cheyne-Stokes breathing. Decreased intravascular volume and cardiac output cause tachycardia, hypotension, peripheral coolness, and livedo reticularis (skin mottling). Patients can have urinary and, less frequently, fecal incontinence. Changes in consciousness and neurologic function generally lead to two different paths to death.

Each of these terminal changes can cause patients and families distress, requiring reassurance and targeted interventions (Table 13-9). Informing families that these changes might occur and providing them with an information sheet can help preempt problems and minimize distress. Understanding that patients stop eating because they are dying, not dying because they have stopped eating, can reduce family and caregiver anxiety. Similarly, informing the family and caregivers that the "death rattle" may occur and that it is not indicative of suffocation, choking, or pain can reduce their worry from the breathing sounds.

Families and caregivers may also feel guilty about stopping treatments, fearing that they are "killing" the patient. This may lead to demands for interventions, such as feeding tubes, that may be ineffective. In such cases, the physician should remind the family and caregivers about the inevitability of events and the palliative goals. Interventions may prolong the dying process and cause discomfort. Physicians also should emphasize that withholding treatments is both legal and ethical and that the family members are not the cause of the

TABLE 13-9 Managing Changes in the Patient's Condition during the Final Days and Hours

CHANGES IN THE PATIENT'S CONDITION	POTENTIAL COMPLICATION	FAMILY'S POSSIBLE REACTION AND CONCERN	ADVICE AND INTERVENTION
Profound fatigue	Bedbound with development of pressure ulcers that are prone to infection, malodor, and pain, and joint pain	Patient is lazy and giving up.	Reassure family and caregivers that terminal fatigue will not respond to interventions and should not be resisted. Use an air mattress if necessary.
Anorexia	None	Patient is giving up; patient will suffer from hunger and will starve to death.	Reassure family and caregivers that the patient is not eating because they are dying; not eating at the end of life does not cause suffering or death. Forced feeding, whether oral, parenteral, or enteral, does not reduce symptoms or prolong life.
Dehydration	Dry mucosal membranes (see below)	Patient will suffer from thirst and die of dehydration.	Reassure family and caregivers that dehydration at the end of life does not cause suffering because patients lose consciousness before any symptom distress. Intravenous hydration can worsen symptoms of dyspnea by pulmonary edema and peripheral edema as well as prolong the dying process.
Dysphagia	Inability to swallow oral medications needed for palliative care		Do not force oral intake. Discontinue unnecessary medications that may have been continued, including antibiotics, diuretics, antidepressants, and laxatives. If swallowing pills is difficult, convert essential medications (analgesics, antiemetics, anxiolytics, and psychotropics) to oral solutions, buccal, sublingual, or rectal administration.
"Death rattle"—noisy breathing		Patient is choking and suffocating.	Reassure the family and caregivers that this is caused by secretions in the oropharynx and the patient is not choking. Reduce secretions with scopolamine (0.2–0.4 mg SC q4h or 1–3 patches q3d). Reposition patient to permit drainage of secretions. Do not suction. Suction can cause patient and family discomfort and is usually ineffective.
Apnea, Cheyne-Stokes respirations, dyspnea		Patient is suffocating.	Reassure family and caregivers that unconscious patients do not experience suffocation or air hunger. Apneic episodes are frequently a premorbid change. Opioids or anxiolytics may be used for dyspnea. Oxygen is unlikely to relieve dyspneic symptoms and may prolong the dying process.
Urinary or fecal incontinence	Skin breakdown if days until death Potential transmission of infectious agents to caregivers	Patient is dirty, malodorous, and physically repellent.	Remind family and caregivers to use universal precautions. Frequent changes of bedclothes and bedding. Use diapers, urinary catheter, or rectal tube if diarrhea or high urine output.
Agitation or delirium	Day/night reversal Hurt self or caregivers	Patient is in horrible pain and going to have a horrible death.	Reassure family and caregivers that agitation and delirium do not necessarily connote physical pain. Depending on the prognosis and goals of treatment, consider evaluating for causes of delirium and modifying medications. Manage symptoms with haloperidol, chlorpromazine, diazepam, or midazolam.
Dry mucosal membranes	Cracked lips, mouth sores, and candidiasis can also cause pain Odor	Patient may be malodorous, physically repellent.	Use baking soda mouthwash or saliva preparation q15–30 min. Use topical nystatin for candidiasis. Coat lips and nasal mucosa with petroleum jelly q60–90 min. Use ophthalmic lubricants q4h or artificial tears q30 min.

patient's death. This reassurance may have to be provided multiple times.

Hearing and touch are said to be the last senses to stop functioning. Whether this is the case or not, families and caregivers can be encouraged to communicate with the dying patient. Encouraging them to talk directly to the patient, even if the patient is unconscious, and hold the patient's hand or demonstrate affection in other ways can be an effective way to channel their urge "to do something" for the patient.

When the plan is for the patient to die at home, the physician must inform the family and caregivers how to determine that the patient has died. The cardinal signs are cessation of cardiac function and respiration; the pupils become fixed; the body becomes cool; muscles relax; and incontinence may occur. Remind the family and caregivers that the eyes may remain open even after the patient has died.

The physician should establish a plan for who the family or caregivers will contact when the patient is dying or has died. Without a plan, family members may panic and call 911, unleashing a cascade of unwanted events, from arrival of emergency personnel and resuscitation to hospital admission. The family and caregivers should be instructed to contact the hospice (if one is involved), the covering physician, or the on-call member of the palliative care team. They should also be told that the medical examiner need not be called unless the state requires it for all deaths. Unless foul play is suspected, the health care team need not contact the medical examiner either.

Just after the patient dies, even the best-prepared family may experience shock and loss and be emotionally distraught. They need time to assimilate the event and be comforted. Health care providers are likely to find it meaningful to themselves and the patient's caregivers to write a bereavement card or letter to the family. The purpose is to communicate about the patient, perhaps emphasizing the patient's virtues and the honor it was to care for the patient, and to express concern for the family's hardship. Some physicians attend the funerals of their patients. Although this is beyond any medical obligation, the presence of the physician can be a source of support to the grieving family and provides an opportunity for closure for the physician.

Death of a spouse is a strong predictor of poor health, and even mortality, for the surviving spouse. It may be important to alert the spouse's physician about the death so that they are aware of symptoms that might require professional attention.

■ FURTHER READING

Emanuel E et al: Attitudes and practices of euthanasia and physician-assisted suicide in the United States, Canada, and Europe. JAMA 316:79, 2016.

Kelley AS, Meier DE: Palliative care—A shifting paradigm. N Engl J Med 363:781, 2010.

Kelley AS et al: Hospice enrollment saves money for Medicare and improves care quality across a number of different lengths-of-stay. Health Aff 32:552, 2012.

Kelley AS et al: Palliative care for the seriously ill. N Engl J Med 373:747, 2015.

Mack JW et al: Associations between end-of-life discussion characteristics and care received near death: A prospective cohort study. J Clin Oncol 30:4387, 2012.

Murray SA et al: Illness trajectories and palliative care. BMJ 330:1007, 2005.

Neuman P et al: Medicare per capita spending by age and service: New data highlights oldest beneficiaries. Health Aff (Millwood) 34:335, 2015.

Nicholas LH et al: Regional variation in the association between advance directives and end-of-life Medicare expenditures. JAMA 306:1447, 2011.

Ornstein KA et al: Evaluation of racial disparities in hospice use and end-of-life treatment intensity in the REGARDS cohort. JAMA Netw Open 3:e2014639, 2020.

Quinn KL et al: Association of receipt of palliative care interventions with health care use, quality of life, and symptom burden among adults with chronic noncancer illness: A systematic review and meta-analysis. JAMA 324:1439, 2020.

Teno JM et al: Change in end-of-life care for Medicare beneficiaries: Site of death, place of care, and health transitions in 2000, 2005, and 2009. JAMA 309:470, 2013.

Teno JM et al: Site of death, place of care, and health care transitions among US Medicare beneficiaries, 2000-2015. JAMA 320:264, 2018.

Van Den Beuken-VanEverdingen MH et al: Update on prevalence of pain in patients with cancer: Systematic review and meta-analysis. J Pain Symptom Manage 51:1070, 2016.

■ WEBSITES

American Academy of Hospice and Palliative Medicine: *www.aahpm.org*

Center to Advance Palliative Care: *http://www.capc.org*

Education in Palliative and End of life Care (EPEC): *http://www.epec.net*

Family Caregiver Alliance: *http://www.caregiver.org*

National Hospice and Palliative Care Organization (including state-specific advance directives): *http://www.nhpco.org*

NCCN: The National Comprehensive Cancer Network palliative care guidelines: *http://www.nccn.org*

Our Care Wishes Advance Care Planning Tool: *https://www.ourcarewishes.org*

Section 1 Pain

14 Pain: Pathophysiology and Management

James P. Rathmell, Howard L. Fields

The province of medicine is to preserve and restore health and to relieve suffering. Understanding pain is essential to both goals. Because pain is universally understood as a signal of disease, it is the most common symptom that brings a patient to a physician's attention. The function of the pain sensory system is to protect the body and maintain homeostasis. It does this by detecting, localizing, and identifying potential or actual tissue-damaging processes. Because different diseases produce characteristic patterns of tissue damage, the quality, time course, and location of a patient's pain lend important diagnostic clues. It is the physician's responsibility to assess each patient promptly for any remediable cause underlying the pain and to provide rapid and effective pain relief whenever possible.

THE PAIN SENSORY SYSTEM

Pain is an unpleasant sensation localized to a part of the body. It is often described in terms of a penetrating or tissue-destructive process (e.g., stabbing, burning, twisting, tearing, squeezing) and/or of a bodily or emotional reaction (e.g., terrifying, nauseating, sickening). Furthermore, any pain of moderate or higher intensity is accompanied by anxiety and the urge to escape or terminate the feeling. These properties illustrate the duality of pain: it is both sensation and emotion. When it is acute, pain is characteristically associated with behavioral arousal and a stress response consisting of increased blood pressure, heart rate, pupil diameter, and plasma cortisol levels. In addition, local muscle contraction (e.g., limb flexion, abdominal wall rigidity) is often present.

PERIPHERAL MECHANISMS

The Primary Afferent Nociceptor A peripheral nerve consists of the axons of three different types of neurons: primary sensory afferents, motor neurons, and sympathetic postganglionic neurons (Fig. 14-1). The cell bodies of primary sensory afferents are in the dorsal root ganglia within the vertebral foramina. The primary afferent axon has two branches: one projects centrally into the spinal cord and the other projects peripherally to innervate tissues. Primary afferents are classified by their diameter, degree of myelination, and conduction velocity. The largest diameter afferent fibers, A-beta (Aβ), respond maximally to light touch and/or moving stimuli; they are present primarily in nerves that innervate the skin. In normal individuals, the activity of these fibers does not produce pain. There are two other classes of primary afferent nerve fibers: the small diameter myelinated A-delta (Aδ) and the unmyelinated (C) axons (Fig. 14-1). These fibers are present in nerves to the skin and to deep somatic and visceral structures. Some tissues, such as the cornea, are innervated only by Aδ and C fiber afferents. Most Aδ and C fiber afferents respond maximally to intense (painful) stimuli and produce the subjective experience of pain when they are activated; this defines them as *primary afferent nociceptors* (*pain receptors*). The ability to detect painful stimuli is completely abolished when conduction in Aδ and C fiber axons is blocked.

Individual primary afferent nociceptors can respond to several different types of noxious stimuli. For example, most nociceptors respond to heat; intense cold; intense mechanical distortion, such as a pinch; changes in pH, particularly an acidic environment; and application of chemical irritants including adenosine triphosphate (ATP), serotonin, bradykinin (BK), and histamine. The transient receptor potential cation channel subfamily V member 1 (TRPV1), also known as the vanilloid receptor, mediates perception of some noxious stimuli, especially heat sensations, by nociceptive neurons; it is activated by heat, acidic pH, endogenous mediators, and capsaicin, a component of hot chili peppers.

Sensitization When intense, repeated, or prolonged stimuli are applied to damaged or inflamed tissues, the threshold for activating primary afferent nociceptors is lowered, and the frequency of firing is higher for all stimulus intensities. Inflammatory mediators such as BK, nerve-growth factor, some prostaglandins (PGs), and leukotrienes contribute to this process, which is called *sensitization*. Sensitization occurs at the level of the peripheral nerve terminal (*peripheral sensitization*) as well as at the level of the dorsal horn of the spinal cord (*central sensitization*). Peripheral sensitization occurs in damaged or inflamed tissues, when inflammatory mediators activate intracellular signal transduction in nociceptors, prompting an increase in the production, transport, and membrane insertion of chemically gated and voltage-gated ion channels. These changes increase the excitability of nociceptor terminals and lower their threshold for activation by mechanical, thermal, and chemical stimuli. Central sensitization occurs when activity, generated by nociceptors during inflammation, enhances the excitability of nerve cells in the dorsal horn of the spinal cord. Following injury and resultant sensitization, normally innocuous stimuli can produce pain (termed *allodynia*). Sensitization is a clinically important process that contributes to tenderness, soreness, and *hyperalgesia* (increased pain intensity in response to the same noxious stimulus; e.g., pinprick causes severe pain). A striking example of sensitization is sunburned skin, in which severe pain can be produced by a gentle slap or a warm shower.

Sensitization is of particular importance for pain and tenderness in deep tissues. Viscera are normally relatively insensitive to noxious mechanical and thermal stimuli, although hollow viscera do generate

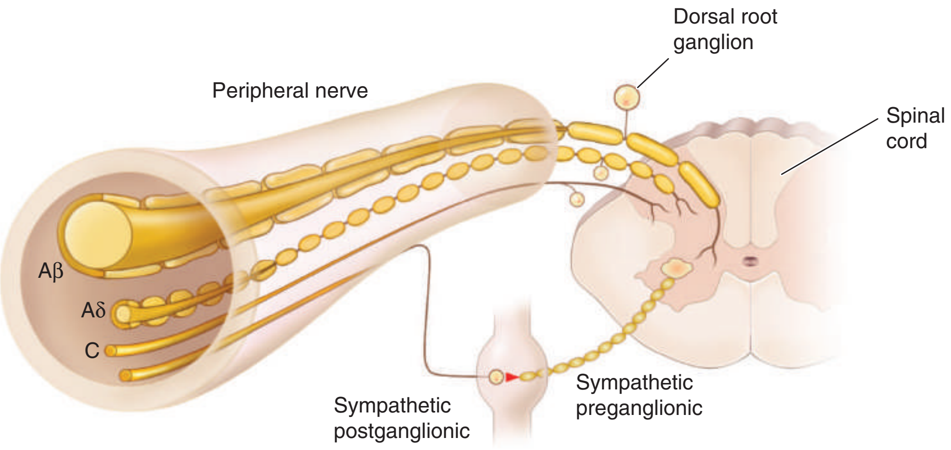

FIGURE 14-1 Components of a typical cutaneous nerve. There are two distinct functional categories of axons: primary afferents with cell bodies in the dorsal root ganglion and sympathetic postganglionic fibers with cell bodies in the sympathetic ganglion. Primary afferents include those with large-diameter myelinated (Aβ), small-diameter myelinated (Aδ), and unmyelinated (C) axons. All sympathetic postganglionic fibers are unmyelinated.

significant discomfort when distended. In contrast, when affected by a disease process with an inflammatory component, deep structures such as joints or hollow viscera characteristically become exquisitely sensitive to mechanical stimulation.

A large proportion of Aδ and C fiber afferents innervating viscera are completely insensitive in normal noninjured, noninflamed tissue. That is, they cannot be activated by known mechanical or thermal stimuli and are not spontaneously active. However, in the presence of inflammatory mediators, these afferents become sensitive to mechanical stimuli. Such afferents have been termed *silent nociceptors*, and their characteristic properties may explain how, under pathologic conditions, the relatively insensitive deep structures can become the source of severe and debilitating pain and tenderness. Low pH, PGs, leukotrienes, and other inflammatory mediators such as BK play a significant role in sensitization.

Nociceptor-Induced Inflammation Primary afferent nociceptors are not simply passive messengers of threats to tissue injury but also play an active role in tissue protection through a neuroeffector function. Most nociceptors contain polypeptide mediators, including substance P, calcitonin gene related peptide (CGRP), and cholecystokinin, that are released from their peripheral terminals when they are activated **(Fig. 14-2)**. Substance P is an 11-amino-acid peptide that is released in peripheral tissues from primary afferent nociceptors and has multiple biologic activities. It is a potent vasodilator, causes mast cell degranulation, is a chemoattractant for leukocytes, and increases the production and release of inflammatory mediators. Interestingly, depletion of substance P from joints reduces the severity of experimental arthritis.

CENTRAL MECHANISMS

The Spinal Cord and Referred Pain The axons of primary afferent nociceptors enter the spinal cord via the dorsal root. They terminate in the dorsal horn of the spinal gray matter **(Fig. 14-3)**. The terminals of primary afferent axons contact spinal neurons that transmit the pain signal to brain sites involved in pain perception. When primary afferents are activated by noxious stimuli, they release neurotransmitters from their terminals that excite the spinal cord neurons. The major neurotransmitter released is glutamate, which rapidly excites the second-order dorsal horn neurons. Primary afferent nociceptor terminals also release substance P and CGRP, which produce a slower and longer-lasting excitation of the dorsal horn neurons. The axon of each primary afferent contacts many spinal neurons, and each spinal neuron receives convergent inputs from many primary afferents.

The convergence of sensory inputs to a single spinal pain-transmission neuron is of great importance because it underlies the phenomenon of referred pain. All spinal neurons that receive input from the viscera and deep musculoskeletal structures also receive input from the skin. The convergence patterns are determined by the spinal segment of the dorsal root ganglion that supplies the afferent innervation of a structure. For example, the afferents that supply the central diaphragm are derived from the third and fourth cervical dorsal root ganglia. Primary afferents with cell bodies in these same ganglia supply the skin of the shoulder and lower neck. Thus, sensory inputs from both the shoulder skin and the central diaphragm converge on pain-transmission neurons in the third and fourth cervical spinal segments. *Because of this convergence and the fact that the spinal neurons are most often activated by inputs from the skin, activity evoked in spinal neurons by input from deep structures is often mislocalized by the patient to a bodily location that roughly corresponds with the region of skin innervated by the same spinal segment*. Thus, inflammation near the central diaphragm is often reported as shoulder discomfort. This spatial displacement of pain sensation from the site of the injury that produces it is known as *referred pain*.

Ascending Pathways for Pain Most spinal dorsal horn neurons activated by primary afferent nociceptors send their axons to the contralateral thalamus. These axons form the contralateral spinothalamic tract, which lies in the anterolateral white matter of the spinal cord, the lateral edge of the medulla, and the lateral pons and midbrain. The spinothalamic pathway is crucial for pain sensation in humans. Interruption of this pathway produces permanent deficits in pain and temperature discrimination.

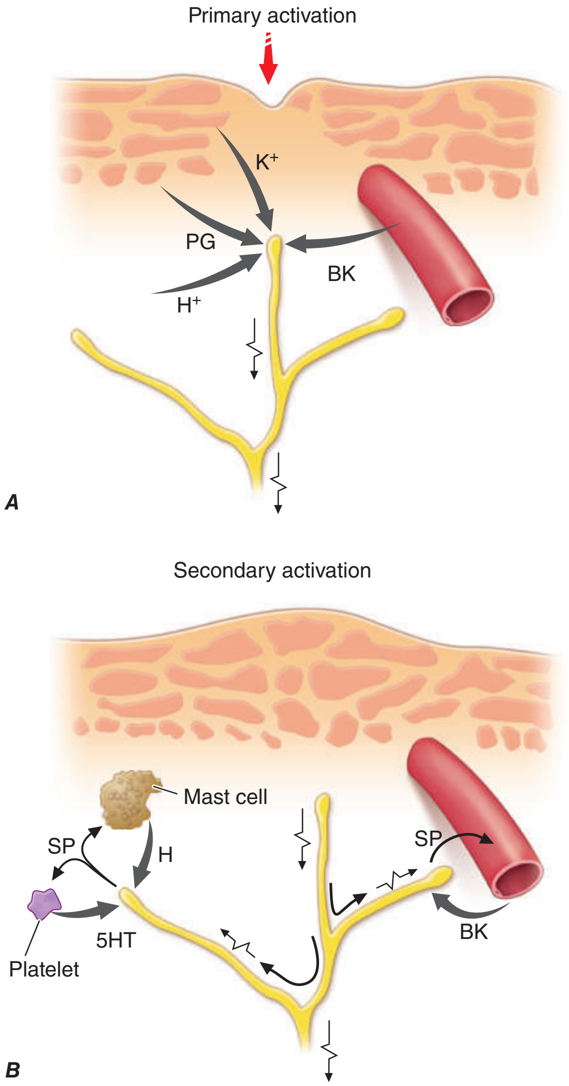

FIGURE 14-2 Events leading to activation, sensitization, and spread of sensitization of primary afferent nociceptor terminals. ***A***. Direct activation by intense pressure and consequent cell damage. Cell damage induces lower pH (H^+) and leads to release of potassium (K^+) and to synthesis of prostaglandins (PGs) and bradykinin (BK). PGs increase the sensitivity of the terminal to BK and other pain-producing substances. ***B***. Secondary activation. Impulses generated in the stimulated terminal propagate not only to the spinal cord but also into other terminal branches where they induce the release of peptides, including substance P (SP). Substance P causes vasodilation and neurogenic edema with further accumulation of BK. Substance P also causes the release of histamine (H) from mast cells and serotonin (5HT) from platelets.

Spinothalamic tract axons ascend to several regions of the thalamus. There is significant divergence of the pain signal from these thalamic sites to several distinct areas of the cerebral cortex that subserve different aspects of the pain experience **(Fig. 14-4)**. One of the thalamic projections is to the somatosensory cortex. This projection mediates the sensory discriminative aspects of pain, i.e., its location, intensity, and quality. Other thalamic neurons project to cortical regions that are linked to emotional responses, such as the anterior cingulate and insular cortex. These pathways to the cortex subserve the affective or unpleasant emotional dimension of pain. This affective dimension of pain produces suffering and exerts potent control of behavior. Because of this dimension, fear is a constant companion of pain. Consequently, injury or surgical lesions to areas of the frontal cortex activated by painful stimuli can diminish the emotional

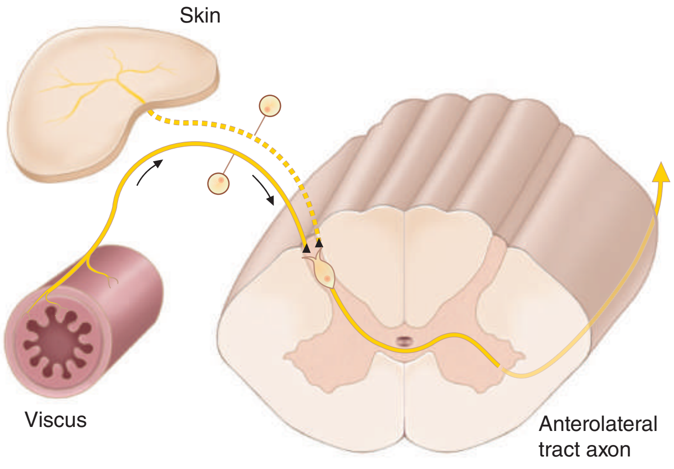

FIGURE 14-3 The convergence-projection hypothesis of referred pain. According to this hypothesis, visceral afferent nociceptors converge on the same pain-projection neurons as the afferents from the somatic structures in which the pain is perceived. The brain has no way of knowing the actual source of input and mistakenly "projects" the sensation to the somatic structure.

impact of pain while largely preserving the individual's ability to localize and recognize stimuli as painful.

PAIN MODULATION

The pain produced by injuries of similar magnitude is remarkably variable in different situations and in different individuals. For example, athletes have been known to sustain serious fractures with only minor pain, and Beecher's classic World War II survey revealed that many soldiers in battle were unbothered by injuries that would have produced agonizing pain in civilian patients. Furthermore, even the suggestion that a treatment will relieve pain can have a significant analgesic effect (the *placebo effect*). On the other hand, many patients find even minor injuries such as venipuncture frightening and unbearable, and the expectation of pain can induce pain even without a noxious stimulus. The suggestion that pain will worsen following administration of an inert substance can increase its perceived intensity (the *nocebo effect*).

The powerful effect of expectation and other psychological variables on the perceived intensity of pain is explained by brain circuits that modulate the activity of the pain-transmission pathways. One of these circuits has links to the hypothalamus, midbrain, and medulla, and it selectively controls spinal dorsal horn pain-transmission neurons through a descending pathway (Fig. 14-4).

Human brain-imaging studies have implicated this pain-modulating circuit in the pain-relieving effect of attention, suggestion, and opioid analgesic medications **(Fig. 14-5)**. Furthermore, each of the component structures of the pathway contains opioid receptors and is sensitive to the direct application of opioid drugs. In animals, lesions of this

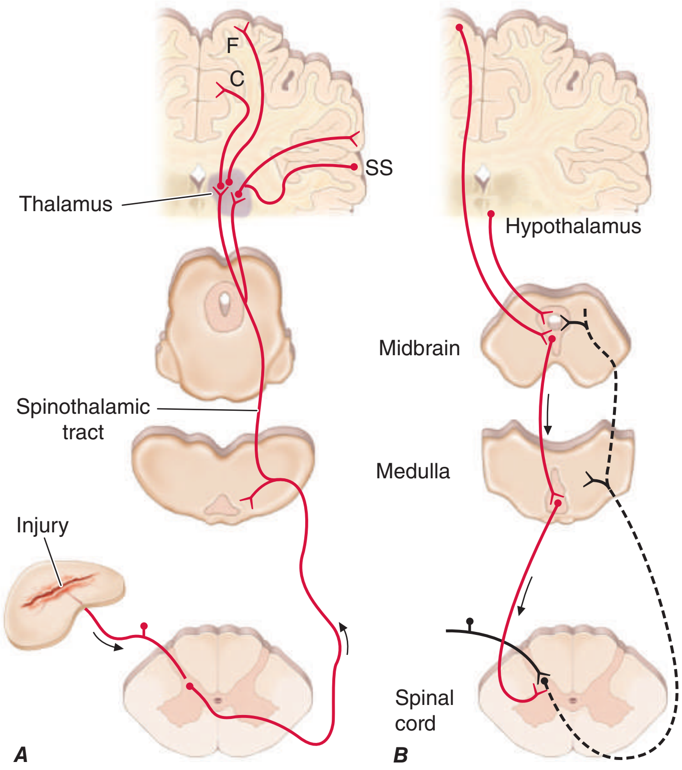

FIGURE 14-4 Pain-transmission and modulatory pathways. ***A***. Transmission system for nociceptive messages. Noxious stimuli activate the sensitive peripheral ending of the primary afferent nociceptor by the process of transduction. The message is then transmitted over the peripheral nerve to the spinal cord, where it synapses with cells of origin of the major ascending pain pathway, the spinothalamic tract. The message is relayed in the thalamus to the anterior cingulate (C), frontal insular (F), and somatosensory cortex (SS). ***B***. Pain-modulation network. Inputs from frontal cortex and hypothalamus activate cells in the midbrain that control spinal pain-transmission cells via cells in the medulla.

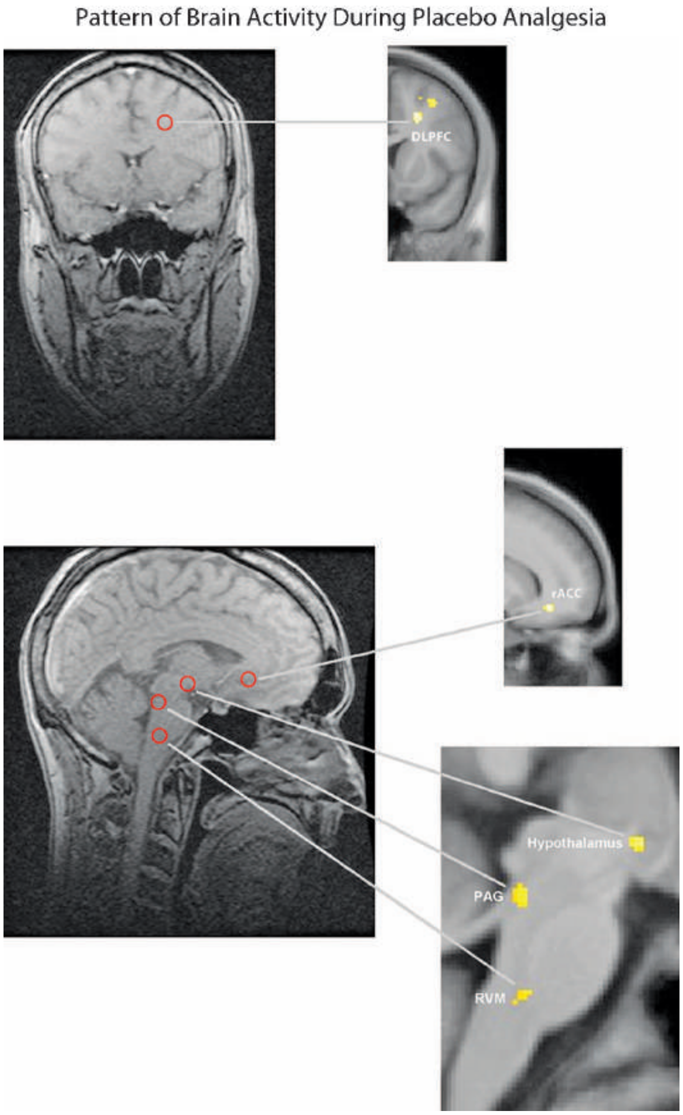

FIGURE 14-5 Functional magnetic resonance imaging (fMRI) demonstrates placebo-enhanced brain activity in anatomic regions correlating with the opioidergic descending pain control system. *Top panel:* Frontal fMRI image shows placebo-enhanced brain activity in the dorsal lateral prefrontal cortex (DLPFC). *Bottom panel:* Sagittal fMRI images show placebo-enhanced responses in the rostral anterior cingulate cortex (rACC), the rostral ventral medullae (RVM), the periaqueductal gray (PAG) area, and the hypothalamus. The placebo-enhanced activity in all areas was reduced by naloxone, demonstrating the link between the descending opioidergic system and the placebo analgesic response. *(Reproduced with permission from F Eippert et al: Activation of the opioidergic descending pain control system underlies placebo analgesia. Neuron 63:533, 2009.)*

descending modulatory system reduce the analgesic effect of systemically administered opioids such as morphine. Along with the opioid receptor, the component nuclei of this pain-modulating circuit contain endogenous opioid peptides such as the enkephalins and β-endorphin.

The most reliable way to activate this endogenous opioid-mediated modulating system is by suggestion of pain relief or by intense emotion directed away from the pain-causing injury (e.g., during severe threat or an athletic competition). In fact, pain-relieving endogenous opioids are released following surgical procedures and in patients given a placebo for pain relief.

Pain-modulating circuits can enhance as well as suppress pain. Both pain-inhibiting and pain-facilitating neurons in the medulla project to and control spinal pain-transmission neurons. Because pain-transmission neurons can be activated by modulatory neurons, it is theoretically possible to generate a pain signal with no peripheral noxious stimulus. In fact, human functional imaging studies have demonstrated increased activity in this circuit during migraine headaches. A central circuit that facilitates pain could account for the finding that pain can be induced by suggestion or enhanced by expectation and provides a framework for understanding how psychological factors can contribute to chronic pain.

■ NEUROPATHIC PAIN

Lesions of the peripheral or central nociceptive pathways typically result in a loss or impairment of pain sensation. Paradoxically, damage to or dysfunction of these pathways can also produce pain. For example, damage to peripheral nerves, as occurs in diabetic neuropathy, or to primary afferents, as in herpes zoster infection, can result in pain that is referred to the body region innervated by the damaged nerves. Pain may also be produced by damage to the central nervous system (CNS), for example, in some patients following trauma or vascular injury to the spinal cord, brainstem, or thalamic areas that contain central nociceptive pathways. Such pains are termed *neuropathic* and are often severe and resistant to standard treatments for pain.

Neuropathic pain typically has an unusual burning, tingling, or electric shock-like quality and may occur spontaneously, without any stimulus, or be triggered by very light touch. These features are rare in other types of pain. On examination, a sensory deficit is characteristically co-extensive with the area of the patient's pain. *Hyperpathia*, a greatly exaggerated pain response to innocuous or mild nociceptive stimuli, especially when applied repeatedly, is also characteristic of neuropathic pain; patients often complain that the very lightest moving stimulus evokes exquisite pain (allodynia). In this regard, it is of clinical interest that a topical preparation of 5% lidocaine in patch form is effective for patients with postherpetic neuralgia who have prominent allodynia.

A variety of mechanisms contribute to neuropathic pain. As with sensitized primary afferent nociceptors, damaged primary afferents, including nociceptors, become highly sensitive to mechanical stimulation and may generate impulses in the absence of stimulation. Increased sensitivity and spontaneous activity are due, in part, to an increased density of sodium channels in the damaged nerve fiber. Damaged primary afferents may also develop sensitivity to norepinephrine. Interestingly, spinal cord pain-transmission neurons cut off from their normal input may also become spontaneously active. Thus, both central and peripheral nervous system hyperactivity contribute to neuropathic pain.

Sympathetically Maintained Pain Patients with peripheral nerve injury occasionally develop spontaneous pain in or beyond the region innervated by the nerve. This pain is often described as having a burning quality. The pain typically begins after a delay of hours to days or even weeks and is accompanied by swelling of the extremity, periarticular bone loss, and arthritic changes in the distal joints. Early in the course of the condition, the pain may be relieved by a local anesthetic block of the sympathetic innervation to the affected extremity. Damaged primary afferent nociceptors acquire adrenergic sensitivity and can be activated by stimulation of the sympathetic outflow. This constellation of spontaneous pain and signs of sympathetic dysfunction following injury has been termed *complex regional pain syndrome* (CRPS) (**Chap. 19**). When this occurs after an identifiable nerve injury, it is termed CRPS type II (also known as posttraumatic neuralgia or, if severe, *causalgia*). When a similar clinical picture appears without obvious nerve injury, it is termed CRPS type I (also known as *reflex sympathetic dystrophy*). CRPS can be produced by a variety of injuries, including fractures of bone, soft tissue trauma, myocardial infarction, and stroke. CRPS type I typically resolves with symptomatic treatment; however, when it persists, detailed examination often reveals evidence of peripheral nerve injury. Although the pathophysiology of CRPS is poorly understood, the pain and the signs of inflammation, when acute, can be rapidly relieved by blocking the sympathetic nervous system. This implies that sympathetic activity can activate undamaged nociceptors when inflammation is present. Signs of sympathetic hyperactivity should be sought in patients with posttraumatic pain and inflammation and no other obvious explanation.

TREATMENT

Acute Pain

The ideal treatment for any pain is to remove the cause; thus, while treatment can be initiated immediately, efforts to establish the underlying etiology should always proceed as treatment begins. Sometimes, treating the underlying condition does not immediately relieve pain. Furthermore, some conditions are so painful that rapid and effective analgesia is essential (e.g., the postoperative state, burns, trauma, cancer, or sickle cell crisis). Analgesic medications are a first line of treatment in these cases, and all practitioners should be familiar with their use.

ASPIRIN, ACETAMINOPHEN, AND NONSTEROIDAL ANTI-INFLAMMATORY AGENTS (NSAIDS)

These drugs are considered together because they are used for similar problems and may have a similar mechanism of action (**Table 14-1**). All these compounds inhibit cyclooxygenase (COX), and except for acetaminophen, all have anti-inflammatory actions, especially at higher dosages. They are particularly effective for mild to moderate headache and for pain of musculoskeletal origin.

Because they are effective for these common types of pain and are available without prescription, COX inhibitors are by far the most used analgesics. They are absorbed well from the gastrointestinal tract and, with occasional use, have only minimal side effects. With chronic use, gastric irritation is a common side effect of aspirin and NSAIDs and is the problem that most frequently limits the dose that can be given. Gastric irritation is most severe with aspirin, which may cause erosion and ulceration of the gastric mucosa leading to bleeding or perforation. Because aspirin irreversibly acetylates platelet COX and thereby interferes with coagulation of the blood, gastrointestinal bleeding is a particular risk. Older age and history of gastrointestinal disease increase the risks of aspirin and NSAIDs. In addition to the well-known gastrointestinal toxicity of NSAIDs, nephrotoxicity is a significant problem for patients using these drugs on a chronic basis. Patients at risk for renal insufficiency, particularly those with significant contraction of their intravascular volume as occurs with chronic diuretic use or acute hypovolemia, should avoid NSAIDs. NSAIDs can also increase blood pressure in some individuals. Long-term treatment with NSAIDs requires regular blood pressure monitoring and treatment if necessary. Although toxic to the liver when taken in high doses, acetaminophen rarely produces gastric irritation and does not interfere with platelet function.

The introduction of parenteral forms of NSAIDs, ketorolac and diclofenac, extends the usefulness of this class of compounds in the management of acute severe pain. Both agents are sufficiently potent and rapid in onset to supplant opioids as first-line treatment for many patients with acute severe headache and musculoskeletal pain.

There are two major classes of COX: COX-1 is constitutively expressed, and COX-2 is induced in the inflammatory state.

TABLE 14-1 Drugs for Relief of Pain

GENERIC NAME	DOSE, mg	INTERVAL	COMMENTS
Nonnarcotic Analgesics: Usual Doses and Intervals			
Acetylsalicylic acid	650 PO	q4h	Enteric-coated preparations available
Acetaminophen	650 PO	q4h	Side effects uncommon
Ibuprofen	400 PO	q4–6h	Available without prescription
Naproxen	250–500 PO	q12h	Naproxen is the common NSAID that poses the least cardiovascular risk, but it has a somewhat higher incidence of gastrointestinal bleeding
Fenoprofen	200 PO	q4–6h	Contraindicated in renal disease
Indomethacin	25–50 PO	q8h	Gastrointestinal side effects common
Ketorolac	15–60 IM/IV	q4–6h	Available for parenteral use
Celecoxib	100–200 PO	q12–24h	Useful for arthritis
Valdecoxib	10–20 PO	q12–24h	Removed from U.S. market in 2005

GENERIC NAME	PARENTERAL DOSE, mg	PO DOSE, mg	COMMENTS
Narcotic Analgesics: Usual Doses and Intervals			
Codeine	30–60 q4h	30–60 q4h	Nausea common
Oxycodone	—	5–10 q4–6h	Usually available with acetaminophen or aspirin
Oxycodone extended-release	—	10-40 q12h	Oral extended-release tablet; high potential for misuse
Morphine	5 q4h	30 q4h	
Morphine sustained release	—	15–60 bid to tid	Oral slow-release preparation
Hydromorphone	1–2 q4h	2–4 q4h	Shorter acting than morphine sulfate
Levorphanol	2 q6–8h	4 q6–8h	Longer acting than morphine sulfate; absorbed well PO
Methadone	5–10 q6–8h	5–20 q6–8h	Due to long half-life, respiratory depression and sedation may persist after analgesic effect subsides; therapy should not be initiated with >40 mg/d, and dose escalation should be made no more frequently than every 3 days
Meperidine	50–100 q3–4h	300 q4h	Poorly absorbed PO; normeperidine is a toxic metabolite; routine use of this agent is not recommended
Butorphanol	—	1–2 q4h	Intranasal spray
Fentanyl	25–100 μg/h	—	72-h transdermal patch
Buprenorphine	5–20 μg/h		7-day transdermal patch
Buprenorphine	0.3 q6–8h		Parenteral administration
Tramadol	—	50–100 q4–6h	Mixed opioid/adrenergic action

GENERIC NAME	UPTAKE BLOCKADE		SEDATIVE POTENCY	ANTICHOLINERGIC POTENCY	ORTHOSTATIC HYPOTENSION	CARDIAC ARRHYTHMIA	AVERAGE DOSE, mg/d	RANGE, mg/d
	5-HT	NE						
Antidepressants[a]								
Doxepin	++	+	High	Moderate	Moderate	Less	200	75–400
Amitriptyline	++++	++	High	Highest	Moderate	Yes	150	25–300
Imipramine	++++	++	Moderate	Moderate	High	Yes	200	75–400
Nortriptyline	+++	++	Moderate	Moderate	Low	Yes	100	40–150
Desipramine	+++	++++	Low	Low	Low	Yes	150	50–300
Venlafaxine	+++	++	Low	None	None	No	150	75–400
Duloxetine	+++	+++	Low	None	None	No	40	30–60

GENERIC NAME	PO DOSE, mg	INTERVAL	COMMENTS
Anticonvulsants and Antiarrythmics[a]			
Carbamazepine	200–300	q6h	Rare aplastic anemia, GI irritation, hepatoitoxicity
Oxcarbamazepine	300	bid	Similar to carbamazepine
Gabapentin[b]	600–1200	q8h	Dizziness, GI irritation; useful in trigeminal neuralgia
Pregabalin	150–600	bid	Similar to gabapentin; dry mouth, edema

[a]Antidepressants, anticonvulsants, and antiarrhythmics have not been approved by the U.S. Food and Drug Administration (FDA) for the treatment of pain. [b]Gabapentin in doses up to 1800 mg/d is FDA approved for postherpetic neuralgia.

Abbreviations: GI, gastrointestinal; 5-HT, serotonin; NE, norepinephrine; NSAID, nonsteroidal anti-inflammatory agent.

COX-2-selective drugs have similar analgesic potency and produce less gastric irritation than the nonselective COX inhibitors. The use of COX-2-selective drugs does not appear to lower the risk of nephrotoxicity compared to nonselective NSAIDs. On the other hand, COX-2-selective drugs offer a significant benefit in the management of acute postoperative pain because they do not affect blood coagulation. Nonselective COX inhibitors (especially aspirin) are usually contraindicated postoperatively because they impair platelet-mediated blood clotting and are thus associated with increased bleeding at the operative site. COX-2 inhibitors, including celecoxib (Celebrex), are associated with increased cardiovascular risk, including cardiovascular death, myocardial infarction, stroke, heart failure, or a thromboembolic event. It appears that this is a class effect of NSAIDs, excluding aspirin. These drugs are contraindicated in patients in the immediate period after coronary artery bypass surgery and should be used with caution in elderly patients and those with a history of or significant risk factors for cardiovascular disease.

OPIOID ANALGESICS

Opioids are the most potent pain-relieving drugs currently available. Of all analgesics, they have the broadest range of efficacy and provide the most reliable and effective treatment for rapid pain relief. Although side effects are common, most are reversible: nausea, vomiting, pruritus, sedation, and constipation are the most frequent and bothersome side effects. Respiratory depression is uncommon at standard analgesic doses but can be life-threatening. Opioid-related side effects can be reversed rapidly with the narcotic antagonist naloxone. Many physicians, nurses, and patients have a certain trepidation about using opioids that is based on a fear of initiating addiction in their patients. There is a small chance of patients becoming addicted to narcotics as a result of their appropriate medical use. For chronic pain, particularly chronic noncancer pain, the risk of addiction in patients taking opioids on a chronic basis remains small, and the risk appears to increase with dose escalation and duration of treatment. The physician should not hesitate to use opioid analgesics in patients with acute severe pain. Table 14-1 lists the most commonly used opioid analgesics.

Opioids produce analgesia by actions in the CNS. They activate pain-inhibitory neurons and directly inhibit pain-transmission neurons. Most of the commercially available opioid analgesics act at the same opioid receptor (μ-receptor), differing mainly in potency, speed of onset, duration of action, and optimal route of administration. Some side effects are due to accumulation of nonopioid metabolites that are unique to individual drugs. One striking example of this is normeperidine, a metabolite of meperidine. At higher doses of meperidine, typically >1 g/d, accumulation of normeperidine can produce hyperexcitability and seizures that are not reversible with naloxone. Normeperidine accumulation is increased in patients with renal failure.

The most rapid pain relief is obtained by intravenous administration of opioids; relief with oral administration is significantly slower. Because of the potential for respiratory depression, patients with any form of respiratory compromise must be kept under close observation following opioid administration; an oxygen-saturation monitor may be useful, but only in a setting where the monitor is under constant surveillance. Opioid-induced respiratory depression is primarily manifest as a reduction in respiratory rate and is typically accompanied by sedation. A fall in oxygen saturation represents a critical level of respiratory depression and the need for immediate intervention to prevent life-threatening hypoxemia. Newer monitoring devices that incorporate capnography or pharyngeal air flow can detect apnea at the point of onset and are now commonly used in hospitalized patients. Ventilatory assistance should be maintained until the opioid-induced respiratory depression has resolved. The opioid antagonist naloxone should be readily available whenever opioids are used at high doses or in patients with compromised pulmonary function. Opioid effects are dose-related, and there is great variability among patients in the doses that relieve pain and produce side effects. Synergistic respiratory depression is common when opioids are administered with other CNS depressants. Co-administration of benzodiazepines is particularly likely to produce respiratory depression and should be avoided, especially in outpatient pain management. Because of this variability in patient response, initiation of therapy requires titration to optimal dose and interval. The most important principle is to provide adequate pain relief. This requires determining whether the drug has adequately relieved the pain and timely reassessment to determine the optimal interval for dosing. *The most common error made by physicians in managing severe pain with opioids is to prescribe an inadequate dose. Because many patients are reluctant to complain, this practice leads to needless suffering.* In the absence of sedation at the expected time of peak effect, a physician should not hesitate to repeat the initial dose to achieve satisfactory pain relief.

A now standard approach to the problem of achieving adequate pain relief is the use of patient-controlled analgesia (PCA). PCA uses a microprocessor-controlled infusion device that can deliver a baseline continuous dose of an opioid drug as well as preprogrammed additional doses whenever the patient pushes a button. The patient can then titrate the dose to the optimal level. This approach is used most extensively for the management of postoperative pain, but there is no reason it should not be used for any hospitalized patient with persistent severe pain. PCA is also used for short-term home care of patients with intractable pain, such as that caused by metastatic cancer.

It is important to understand that the PCA device delivers small, repeated doses to maintain pain relief; in patients with severe pain, the pain must first be brought under control with a loading dose before transitioning to the PCA device. The bolus dose of the drug (typically 1 mg of morphine, 0.2 mg of hydromorphone, or 10 μg of fentanyl) can then be delivered repeatedly as needed. To prevent overdosing, PCA devices are programmed with a lockout period after each demand dose is delivered (typically starting at 10 min) and a limit on the total dose delivered per hour. Although some have advocated the use of a simultaneous continuous or basal infusion of the PCA drug, this may increase the risk of respiratory depression and has not been shown to increase the overall efficacy of the technique.

The availability of new routes of administration has extended the usefulness of opioid analgesics. Most important is the availability of spinal administration. Opioids can be infused through a spinal catheter placed either intrathecally or epidurally. By applying opioids directly to the spinal or epidural space adjacent to the spinal cord, regional analgesia can be obtained using relatively low total doses. Indeed, the dose required to produce effective analgesia when using morphine intrathecally (0.1–0.3 mg) is a fraction of that required to produce similar analgesia when administered intravenously (5–10 mg). In this way, side effects such as sedation, nausea, and respiratory depression can be minimized. This approach has been used extensively during labor and delivery and for postoperative pain relief following surgical procedures. Continuous intrathecal delivery via implanted spinal drug-delivery systems is now commonly used, particularly for the treatment of cancer-related pain that would require sedating doses for adequate pain control if given systemically. Opioids can also be given intranasally (butorphanol), rectally, and transdermally (fentanyl and buprenorphine), or through the oral mucosa (fentanyl), thus avoiding the discomfort of frequent injections in patients who cannot be given oral medication. The fentanyl and buprenorphine transdermal patches have the advantage of providing steady plasma levels, which may improve patient comfort.

Recent additions to the armamentarium for treating opioid-induced side effects are the peripherally acting opioid antagonists alvimopan (Entereg) and methylnaltrexone (Rellistor). Alvimopan is available as an orally administered agent that is restricted to the intestinal lumen by limited absorption; methylnaltrexone is available in a subcutaneously administered form that has virtually no penetration into the CNS. Both agents act by binding to peripheral μ-receptors, thereby inhibiting or reversing the effects of opioids at these peripheral sites. The action of both agents is restricted to receptor sites outside of the CNS; thus, these drugs can reverse the adverse effects of opioid analgesics that are mediated through their peripheral receptors without reversing their CNS-mediated analgesic effects. Alvimopan has proven effective in lowering the duration of persistent ileus following abdominal surgery in patients receiving opioid analgesics for postoperative pain control. Methylnaltrexone has proven effective for relief of opioid-induced constipation in patients taking opioid analgesics on a chronic basis.

Opioid and COX Inhibitor Combinations When used in combination, opioids and COX inhibitors have additive or synergistic effects. Because a lower dose of each can be used to achieve the same degree of pain relief and their side effects are typically dose related, such combinations are used to lower the severity of dose-related side effects. However, fixed-ratio combinations of an opioid with acetaminophen carry an important risk. Dose escalation because of increased severity of pain or decreased opioid effect as a result of tolerance may lead to ingestion of levels of acetaminophen that are toxic to the liver. Although acetaminophen-related hepatotoxicity is uncommon, it remains a significant cause for liver failure. Thus, many practitioners have moved away from the use of opioid-acetaminophen combination analgesics to avoid the risk of excessive acetaminophen exposure as the dose of the analgesic is escalated.

CHRONIC PAIN

Managing patients with chronic pain is intellectually and emotionally challenging. Sensitization of the nervous system can occur without an obvious precipitating cause, e.g., fibromyalgia, or chronic headache. In many patients, chronic pain becomes a distinct disease unto itself. The pain-generating mechanism is often difficult or impossible to determine with certainty; such patients are demanding of the physician's time and often appear emotionally distraught. The traditional medical approach of seeking an obscure organic pathology is often unhelpful. On the other hand, psychological evaluation and behaviorally based treatment paradigms are frequently helpful, particularly in the setting of a multidisciplinary pain-management center. Unfortunately, this approach, while effective, remains largely underused in current medical practice.

There are several factors that can cause, perpetuate, or exacerbate chronic pain. First, of course, the patient may simply have a disease that is characteristically painful for which there is presently no cure. Arthritis, cancer, chronic daily headaches, fibromyalgia, and diabetic neuropathy are examples of this. Second, there may be secondary perpetuating factors that are initiated by disease and persist after that disease has resolved. Examples include damaged sensory nerves, sympathetic efferent activity, and painful reflex muscle contraction (spasm). Finally, a variety of psychological conditions can exacerbate or even cause pain.

There are certain areas to which special attention should be paid in a patient's medical history. Because depression **(Chap. 463)** is the most common emotional disturbance in patients with chronic pain, patients should be questioned about their mood, appetite, sleep patterns, and daily activity. A simple standardized questionnaire, such as the Beck Depression Inventory, can be a useful screening device. It is important to remember that major depression is a common, treatable, and potentially fatal illness.

Other clues that a significant emotional disturbance is contributing to a patient's chronic pain complaint include pain that occurs in multiple, unrelated sites; a pattern of recurrent, but separate, pain problems beginning in childhood or adolescence; pain beginning at a time of emotional trauma, such as the loss of a parent or spouse; a history of physical or sexual abuse; and past or present substance abuse.

On examination, special attention should be paid to whether the patient guards the painful area and whether certain movements or postures are avoided because of pain. Discovering a mechanical component to the pain can be useful both diagnostically and therapeutically. Painful areas should be examined for deep tenderness, noting whether this is localized to muscle, ligamentous structures, or joints. Chronic myofascial pain is very common, and in these patients, deep palpation may reveal highly localized trigger points that are firm bands or knots in muscle. Relief of the pain following injection of local anesthetic into these trigger points supports the diagnosis. A neuropathic component to the pain is indicated by evidence of nerve damage, such as sensory impairment, exquisitely sensitive skin (allodynia), weakness, and muscle atrophy, or loss of deep tendon reflexes. Evidence suggesting sympathetic nervous system involvement includes the presence of diffuse swelling, changes in skin color and temperature, and hypersensitive skin and joint tenderness compared with the normal side. Relief of the pain with a sympathetic block supports the diagnosis, but once the condition becomes chronic, the response to sympathetic blockade is of variable magnitude and duration; the role for repeated sympathetic blocks in the overall management of CRPS is unclear.

A guiding principle in evaluating patients with chronic pain is to assess both emotional and somatic causal and perpetuating factors before initiating therapy. Addressing these issues together, rather than waiting to address emotional issues after somatic causes of pain have been ruled out, improves compliance in part because it assures patients that a psychological evaluation does not mean that the physician is questioning the validity of their complaint. Even when a somatic cause for a patient's pain can be found, it is still wise to look for other factors. For example, a cancer patient with painful bony metastases may have additional pain due to nerve damage and may also be depressed. Optimal therapy requires that each of these factors be assessed and treated.

TREATMENT

Chronic Pain

Once the evaluation process has been completed and the likely causative and exacerbating factors identified, an explicit treatment plan should be developed. An important part of this process is to identify specific and realistic functional goals for therapy, such as getting a good night's sleep, being able to go shopping, or returning to work. A multidisciplinary approach that uses medications, counseling, physical therapy, nerve blocks, and even surgery may be required to improve the patient's quality of life. There are also some newer, minimally invasive procedures that can be helpful for some patients with intractable pain. These include image-guided interventions such as epidural injection of glucocorticoids for acute radicular pain and radiofrequency treatment of the facet joints for chronic facet-related back and neck pain. For patients with severe and persistent pain that is unresponsive to more conservative treatment, placement of electrodes on peripheral nerves or within the spinal canal on nerve roots or in the space overlying the dorsal columns of the spinal cord (spinal cord stimulation) has shown significant benefit. The criteria for predicting which patients will respond to these procedures continue to evolve. They are generally reserved for patients who have not responded to conventional pharmacologic approaches. Referral to a multidisciplinary pain clinic for a full evaluation should precede any invasive procedure. Such referrals are clearly not necessary for all chronic pain patients. For some, pharmacologic management alone can provide adequate relief.

ANTIDEPRESSANT MEDICATIONS

The tricyclic antidepressants (TCAs), particularly nortriptyline and desipramine (Table 14-1), are useful for the management of chronic pain. Although developed for the treatment of depression, the TCAs have a spectrum of dose-related biologic activities that include analgesia in a variety of chronic clinical conditions. Although the mechanism is unknown, the analgesic effect of TCAs has a more rapid onset and occurs at a lower dose than is typically required for the treatment of depression. Furthermore, patients with chronic pain who are not depressed obtain pain relief with antidepressants. There is evidence that TCAs potentiate opioid analgesia, so they may be useful adjuncts for the treatment of severe persistent pain such as occurs with malignant tumors. **Table 14-2** lists some of the painful conditions that respond to TCAs. TCAs are of particular value in the management of neuropathic pain such as occurs in diabetic neuropathy and postherpetic neuralgia, for which there are few other therapeutic options.

The TCAs that have been shown to relieve pain have significant side effects (Table 14-1; **Chap. 463**). Some of these side effects, such as orthostatic hypotension, drowsiness, cardiac conduction delay, memory impairment, constipation, and urinary retention,

TABLE 14-2 Painful Conditions That Respond to Tricyclic Antidepressants

Postherpetic neuralgia[a]
Diabetic neuropathy[a]
Fibromyalgia[a]
Tension headache[a]
Migraine headache[a]
Rheumatoid arthritis[a,b]
Chronic low back pain[b]
Cancer
Central poststroke pain

[a]Controlled trials demonstrate analgesia. [b]Controlled studies indicate benefit but not analgesia.

are particularly problematic in elderly patients, and several are additive to the side effects of opioid analgesics. The selective serotonin reuptake inhibitors such as fluoxetine (Prozac) have fewer and less serious side effects than TCAs, but they are much less effective for relieving pain. In contrast, venlafaxine (Effexor) and duloxetine (Cymbalta), which are nontricyclic antidepressants that block both serotonin and norepinephrine reuptake, appear to retain most of the pain-relieving effects of TCAs with a side effect profile more like that of the selective serotonin reuptake inhibitors. Because of their favorable adverse effect profiles, these drugs have largely supplanted the use of TCAs in treatment of chronic pain.

ANTICONVULSANTS AND ANTIARRHYTHMICS

These drugs are useful primarily for patients with neuropathic pain. Phenytoin (Dilantin) and carbamazepine (Tegretol) were first shown to relieve the pain of trigeminal neuralgia **(Chap. 452)**. This pain has a characteristic brief, shooting, electric shock-like quality. In fact, anticonvulsants seem to be particularly helpful for pains that have such a lancinating quality. Newer anticonvulsants, the calcium channel alpha-2-delta subunit ligands gabapentin (Neurontin) and pregabalin (Lyrica), are effective for a broad range of neuropathic pains. Furthermore, because of their favorable side effect profiles, these newer anticonvulsants are often used as first-line agents.

CANNABINOIDS

These agents are widely used for their analgesic properties, although published evidence suggests that any analgesic effects are modest, with small increases in pain threshold reported and variable reductions in clinical pain intensity. Cannabis more consistently reduces the unpleasantness of the pain experience and, in cancer-related pain, can lessen the nausea and vomiting associated with chemotherapy use. **Marijuana and related compounds are discussed in Chap. 466.**

CHRONIC OPIOID MEDICATION

The long-term use of opioids is accepted for patients with pain due to malignant disease. Although opioid use for chronic pain of nonmalignant origin is controversial, for many patients, opioids are the only option that produces meaningful pain relief. This is understandable because opioids are the most potent and have the broadest range of efficacy of any analgesic medications. Although addiction is rare in patients who first use opioids for pain relief, some degree of tolerance and physical dependence is likely with long-term use. Furthermore, studies suggest that long-term opioid therapy may worsen pain in some individuals, termed *opioid-induced hyperalgesia*. Therefore, before embarking on opioid therapy, other options should be explored, and the limitations and risks of opioids should be explained to the patient. It is also important to recognize that some opioid analgesic medications have mixed agonist-antagonist properties (e.g., butorphanol and buprenorphine). From a practical standpoint, this means that they may worsen pain by inducing an abstinence syndrome in patients who are actively being treated with other opioids and are physically dependent.

With long-term outpatient use of orally administered opioids, it may be desirable to use long-acting compounds such as levorphanol, methadone, extended-release morphine or oxycodone, or transdermal fentanyl (Table 14-1). The pharmacokinetic profiles of these drug preparations enable the maintenance of sustained analgesic blood levels, potentially minimizing side effects such as sedation that are associated with high peak plasma levels and reducing the likelihood of rebound pain associated with a rapid fall in plasma opioid concentration. Extended-release opioid formulations are approved primarily for patients who are already taking other opioids and should not be used as first-line opioids for pain. Although long-acting opioid preparations may provide superior pain relief in patients with a continuous pattern of ongoing pain, others suffer from intermittent severe episodic pain and experience superior pain control and fewer side effects with the periodic use of short-acting opioid analgesics. Constipation is a virtually universal side effect of opioid use and should be treated expectantly. As noted earlier in the discussion of acute pain treatment, an advance for patients is the development of peripherally acting opioid antagonists that can reverse the constipation associated with opioid use without interfering with analgesia.

Soon after the introduction of an extended-release oxycodone formulation (OxyContin) in the late 1990s, a dramatic rise in emergency department visits and deaths associated with oxycodone ingestion appeared. This appears to be due primarily to individuals using a prescription opioid nonmedically. Drug-induced deaths have rapidly risen and are now the second leading cause of death in Americans, just behind motor vehicle fatalities. In 2011, the Office of National Drug Control Policy established a multifaceted approach to address prescription drug abuse, including prescription drug monitoring programs (PDMPs) that allow practitioners to determine if patients are receiving prescriptions from multiple providers and use of law enforcement to eliminate improper prescribing practices. In 2016, the Centers for Disease Control and Prevention (CDC) released the *CDC Guideline for Prescribing Opioids for Chronic Pain*, with recommendations for primary care clinicians who are prescribing opioids for chronic noncancer pain. A modified approach to opioid prescribing was published in 2019 by the Health and Human Services Task Force on chronic pain best medical practices and updated in 2022. These guidelines address (1) when to initiate or continue opioids for chronic pain; (2) opioid selection, dosage, duration, follow-up, and discontinuation; and (3) assessing risk and addressing harms of opioid use. The recent increase in scrutiny leaves many practitioners hesitant to prescribe opioid analgesics, other than for brief periods to control pain associated with illness or injury. For now, the choice to begin chronic opioid therapy for a given patient is left to the individual practitioner. The new CDC guidelines aim to help practitioners reduce the risk of opioid misuse and overdose deaths while providing effective pain management for their patients. Pragmatic guidelines for properly selecting and monitoring patients receiving chronic opioid therapy are shown in **Table 14-3**; a checklist for primary care clinicians prescribing opioids for noncancer pain is shown in **Table 14-4**. Practitioners should refer to the 2022 Clinical Practice Guideline for prescribing opioids for pain (referenced in Further Reading, below). **Opioid use disorders are comprehensively discussed in Chap. 467.**

TREATMENT OF NEUROPATHIC PAIN

It is important to individualize treatment for patients with neuropathic pain. Several general principles should guide therapy: the first is to move quickly to provide relief, and the second is to minimize drug side effects. For example, in patients with postherpetic neuralgia and significant cutaneous hypersensitivity, topical lidocaine (e.g. Lidoderm patches) can provide immediate relief without side effects. The anticonvulsants gabapentin or pregabalin

TABLE 14-3 Guidelines for Selecting and Monitoring Patients Receiving Chronic Opioid Therapy (COT) for the Treatment of Chronic, Noncancer Pain

Patient Selection

- Conduct a history, physical examination, and appropriate testing, including an assessment of risk of substance abuse, misuse, or addiction.
- Consider a trial of COT if pain is moderate or severe, pain is having an adverse impact on function or quality of life, and potential therapeutic benefits outweigh potential harms.
- A benefit-to-harm evaluation, including a history, physical examination, and appropriate diagnostic testing, should be performed and documented before and on an ongoing basis during COT.

Informed Consent and Use of Management Plans

- Informed consent should be obtained. A continuing discussion with the patient regarding COT should include goals, expectations, potential risks, and alternatives to COT.
- Consider using a written COT management plan to document patient and clinician responsibilities and expectations and assist in patient education.

Initiation and Titration

- Initial treatment with opioids should be considered as a therapeutic trial to determine whether COT is appropriate.
- Opioid selection, initial dosing, and titration should be individualized according to the patient's health status, previous exposure to opioids, attainment of therapeutic goals, and predicted or observed harms.

Monitoring

- Reassess patients on COT periodically and as warranted by changing circumstances. Monitoring should include documentation of pain intensity and level of functioning, assessments of progress toward achieving therapeutic goals, presence of adverse events, and adherence to prescribed therapies.
- In patients on COT who are at high risk or who have engaged in aberrant drug-related behaviors, clinicians should periodically obtain urine drug screens or other information to confirm adherence to the COT plan of care.
- In patients on COT not at high risk and not known to have engaged in aberrant drug-related behaviors, clinicians should consider periodically obtaining urine drug screens or other information to confirm adherence to the COT plan of care.

Source: Reproduced with permission from R Chou et al: Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. J Pain 10:113-130, 2009.

TABLE 14-4 Centers for Disease Control and Prevention Checklist for Prescribing Opioids for Chronic Pain

For Primary Care Providers Treating Adults (18+) with Chronic Pain ≥3 Months, Excluding Cancer, Palliative, and End-of-Life Care

CHECKLIST

WHEN CONSIDERING LONG-TERM OPIOID THERAPY

- Set realistic goals for pain and function based on diagnosis (e.g., walk around the block).
- Check that nonopioid therapies tried and optimized.
- Discuss benefits and risks (e.g., addiction, overdose) with patient.
- Evaluate risk of harm or misuse.
  - Discuss risk factors with patient.
  - Check prescription drug monitoring program (PDMP) data.
  - Check urine drug screen.
- Set criteria for stopping or continuing opioids.
- Assess baseline pain and function (e.g., Pain, Enjoyment, General Activity [PEG] scale).
- Schedule initial reassessment within 1–4 weeks.
- Prescribe short-acting opioids using lowest dosage on product labeling; match duration to scheduled reassessment.

IF RENEWING WITHOUT A PATIENT VISIT

- Check that return visit is scheduled ≤3 months from last visit.

WHEN REASSESSING AT A PATIENT VISIT

- Continue opioids only after confirming clinically meaningful improvements in pain and function without significant risks or harm.
- Assess pain and function (e.g., PEG); compare results to baseline.
- Evaluate risk of harm or misuse:
  - Observe patient for signs of oversedation or overdose risk. If yes: Taper dose.
  - Check PDMP.
  - Check for opioid use disorder if indicated (e.g., difficulty controlling use). If yes: Refer for treatment.
- Check that nonopioid therapies optimized. Determine whether to continue, adjust, taper, or stop opioids.
- Calculate opioid dosage morphine milligram equivalent (MME).
  - If ≥50 MME/day total (≥50 mg hydrocodone; ≥33 mg oxycodone), increase frequency of follow-up; consider offering naloxone.
  - Avoid ≥90 MME/day total (≥90 mg hydrocodone; ≥60 mg oxycodone), or carefully justify; consider specialist referral.
- Schedule reassessment at regular intervals (≤3 months).

Source: Centers for Disease Control and Prevention, available at *https://stacks.cdc.gov/view/cdc/38025*. Accessed February 18, 2024 (Public Domain).

(see above) or antidepressants (nortriptyline, desipramine, duloxetine, or venlafaxine) can be used as first-line drugs for patients with neuropathic pain. Systemically administered antiarrhythmic drugs such as lidocaine and mexiletine are less likely to be effective. Although intravenous infusion of lidocaine can provide analgesia for patients with different types of neuropathic pain, the relief is usually transient, typically lasting just hours after the cessation of the infusion. The oral lidocaine congener mexiletine is poorly tolerated, producing frequent gastrointestinal adverse effects. There is no consensus on which class of drug should be used as a first-line treatment for any chronically painful condition. However, because relatively high doses of anticonvulsants are required for pain relief, sedation is not uncommon. Sedation is also a problem with TCAs but is much less of a problem with serotonin/norepinephrine reuptake inhibitors (SNRIs; e.g., venlafaxine and duloxetine). Thus, in the elderly or in patients whose daily activities require high-level mental activity, these drugs should be considered the first line. In contrast, opioid medications should be used as a second- or third-line drug class. Although highly effective for many painful conditions, opioids are sedating, and their effect tends to lessen over time, leading to dose escalation and, occasionally, a worsening of pain. A couple of interesting alternatives to pure opioids are two drugs with mixed opioid and norepinephrine reuptake action: tramadol and tapentadol. Tramadol is a relatively weak opioid but is sometimes effective for pain unresponsive to nonopioid analgesics. Tapentadol is a stronger opioid, but its analgesic action is apparently enhanced by the norepinephrine reuptake blockade. Similarly, drugs of different classes can be used in combination to optimize pain control.

It is worth emphasizing that many patients, especially those with chronic pain, seek medical attention primarily because they are suffering and because only physicians can provide the medications required for pain relief. A primary responsibility of all physicians is to minimize the physical and emotional discomfort of their patients. Familiarity with pain mechanisms and analgesic medications is an important step toward accomplishing this aim.

FURTHER READING

De Vita MJ et al: Association of cannabinoid administration with experimental pain in healthy adults a systematic review and meta-analysis. JAMA Psychiatry 75:1118, 2018.

Dowell D et al: CDC Clinical Practice Guideline for prescribing opioids for pain—United States, 2022. MMWR Recomm Rep 71:1, 2022.

Finnerup NB et al: Neuropathic pain: From mechanisms to treatment. Physiol Rev 101:259, 2021.

Sun EC et al: Incidence of and risk factors for chronic opioid use among opioid-naive patients in the postoperative period. JAMA Intern Med 176:1286, 2016.

15

Chest Discomfort

David A. Morrow

Chest discomfort is among the most common reasons for which patients present for medical attention at either an emergency department (ED) or an outpatient clinic. The evaluation of nontraumatic chest discomfort is inherently challenging owing to the broad variety of possible causes, a minority of which are life-threatening conditions that should not be missed. It is helpful to frame the initial diagnostic assessment and triage of patients with acute chest discomfort around three categories: (1) myocardial ischemia; (2) other cardiopulmonary causes (myopericardial disease, aortic emergencies, and pulmonary conditions); and (3) noncardiopulmonary causes. Although rapid identification of high-risk conditions is a priority of the initial assessment, strategies that incorporate routine liberal use of testing carry the potential for adverse effects of unnecessary investigations.

EPIDEMIOLOGY AND NATURAL HISTORY

Chest discomfort is one of the three most common reason for visits to the ED in the United States, resulting in 6–7 million emergency visits each year. More than 60% of patients with this presentation are hospitalized for further testing, and most of the remainder undergo additional investigation in the ED. Fewer than 15% of evaluated patients are eventually diagnosed with acute coronary syndrome (ACS), with rates of 10–20% in most series of unselected populations, and a rate as low as 5% in some studies. The most common diagnoses are gastrointestinal causes **(Fig. 15-1)**, and as few as 5% are other life-threatening cardiopulmonary conditions. In a large proportion of patients with transient acute chest discomfort, ACS or another acute cardiopulmonary cause is excluded but the cause is not determined. Therefore, the resources and time devoted to the evaluation of chest discomfort *in the absence of a severe cause* are substantial. Nevertheless, historically, a disconcerting 2–6% of patients with chest discomfort of presumed nonischemic etiology who are discharged from the ED were later deemed to have had a missed myocardial infarction (MI).

The natural histories of ACS, myocarditis, acute pericardial diseases, pulmonary embolism, and aortic emergencies are discussed in **Chaps. 281, 284, 285, 286, 290, and 291**, respectively. The estimated rate of major cardiovascular events through 30 days in patients with acute chest pain who had been stratified as low risk was 2.5% in a large population-based study that excluded patients with ST-segment elevation or definite noncardiac chest pain.

CAUSES OF CHEST DISCOMFORT

The major etiologies of chest discomfort are discussed in this section and summarized in **Table 15-1**. Additional elements of the history, physical examination, and diagnostic testing that aid in distinguishing these causes are discussed in a later section (see "Approach to the Patient").

MYOCARDIAL ISCHEMIA/INJURY

Myocardial ischemia causing chest discomfort, termed *angina pectoris*, is a primary clinical concern in patients presenting with chest symptoms. Myocardial ischemia is precipitated by an imbalance between myocardial oxygen requirements and myocardial oxygen supply, resulting in insufficient delivery of oxygen to meet the heart's metabolic demands. Myocardial oxygen consumption may be elevated by increases in heart rate, ventricular wall stress, and myocardial contractility, whereas myocardial oxygen supply is determined by coronary blood flow and coronary arterial oxygen content. When myocardial ischemia is sufficiently severe and prolonged in duration (as little as 20 min), irreversible cellular injury occurs, resulting in MI.

The pathobiology of ischemic heart disease is described in **Chap. 284**. *Stable angina* is characterized by ischemic episodes that are typically precipitated by a superimposed increase in oxygen demand during physical exertion and relieved upon resting. Unstable ischemic heart disease **(Chaps. 285 and 286)** is classified clinically by the presence or absence of detectable acute myocardial injury and the presence or absence of ST-segment elevation on the patient's electrocardiogram (ECG). Unstable ischemic heart disease is classified as *unstable angina* when there is no detectable acute myocardial injury and either as *non–ST elevation MI* (NSTEMI; see **Chap. 285**) or *ST-elevation MI* (STEMI; see **Chap. 286**) when there is evidence of acute myocardial injury.

Clinicians should be aware that unstable ischemic symptoms may also occur predominantly because of increased myocardial oxygen demand (e.g., during intense psychological stress) or because of decreased oxygen delivery due to anemia, hypoxia, or hypotension. However, the term *acute coronary syndrome*, which encompasses unstable angina, NSTEMI, and STEMI, is in general reserved for ischemia precipitated by acute coronary atherothrombosis. To guide therapeutic strategies, a standardized system for classification of MI has been expanded to discriminate MI resulting from acute coronary thrombosis (type 1 MI) from MI occurring secondary to other imbalances of myocardial oxygen supply and demand (type 2 MI; see **Chap. 285**). Conditions associated with extreme myocardial oxygen demand and impaired endocardial blood flow, such as aortic valve disease **(Chap. 291)**, hypertrophic cardiomyopathy, or idiopathic dilated cardiomyopathy **(Chaps. 266–270)**, can precipitate myocardial ischemia in patients with or without underlying obstructive atherosclerosis. Ischemic causes of chest pain and myocardial injury are additionally distinguished from nonischemic causes of acute myocardial injury, such as myocarditis.

Characteristics of Ischemic Chest Discomfort The clinical characteristics of angina pectoris, often referred to simply as "angina," are highly similar in quality and location whether the ischemic discomfort is a manifestation of stable ischemic heart disease, unstable angina, or MI (Table 15-1). Heberden initially described angina as a sense of

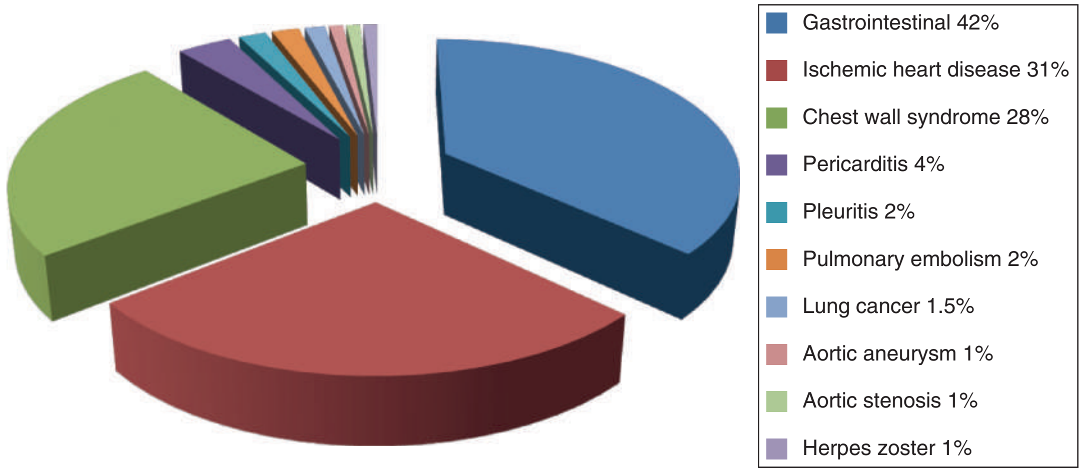

FIGURE 15-1 Distribution of final discharge diagnoses in patients with nontraumatic acute chest pain. *(Figure prepared from data in P Fruergaard et al: Eur Heart J 17:1028, 1996.)*

TABLE 15-1 Typical Clinical Features of Major Causes of Acute Chest Discomfort

SYSTEM	CONDITION	ONSET/DURATION	QUALITY	LOCATION	ASSOCIATED FEATURES
Cardiopulmonary					
Cardiac	Myocardial ischemia	*Stable angina:* Precipitated by exertion, cold, or stress; 2–10 min *Unstable angina:* Increasing pattern or at rest *Myocardial infarction:* Usually >30 min	Pressure, tightness, squeezing, heaviness, burning	Retrosternal; often radiation to neck, jaw, shoulders, or arms; sometimes epigastric	S_4 gallop or mitral regurgitation murmur (rare) during pain; S_3 or rales if severe ischemia or complication of myocardial infarction
	Pericarditis	Variable; hours to days; may be episodic	Pleuritic, sharp	Retrosternal or toward cardiac apex; may radiate to left shoulder	May be relieved by sitting up and leaning forward; pericardial friction rub
Vascular	Acute aortic syndrome	Sudden onset of unrelenting pain	Tearing or ripping; knifelike	Anterior chest, often radiating to back, between shoulder blades	Associated with hypertension and/or underlying connective tissue disorder; murmur of aortic insufficiency; loss of peripheral pulses
	Pulmonary embolism	Sudden onset	Pleuritic; may manifest as heaviness with massive pulmonary embolism	Often lateral, on the side of the embolism	Dyspnea, tachypnea, tachycardia, and hypotension
	Pulmonary hypertension	Variable; often exertional	Pressure	Substernal	Dyspnea, signs of increased venous pressure
Pulmonary	Pneumonia or pleuritis	Variable	Pleuritic	Unilateral, often localized	Dyspnea, cough, fever, rales, occasional rub
	Spontaneous pneumothorax	Sudden onset	Pleuritic	Lateral to side of pneumothorax	Dyspnea, decreased breath sounds on side of pneumothorax
Noncardiopulmonary					
Gastrointestinal	Esophageal reflux	10–60 min	Burning	Substernal, epigastric	Worsened by postprandial recumbency; relieved by antacids
	Esophageal spasm	2–30 min	Pressure, tightness, burning, intense squeezing	Retrosternal	Can closely mimic angina. May be relieved by nitroglycerin or dihydropyridine calcium channel antagonists
	Esophageal injury	Prolonged	Intense squeezing	Retrosternal	Includes Mallory-Weiss tear or esophageal rupture (Boerhaave's syndrome) from vomiting
	Peptic ulcer	Prolonged; 60–90 min after meals	Burning	Epigastric (most common), substernal	Relieved with food or antacids
	Gallbladder disease, including cholecystitis and biliary colic	Prolonged; steady, usually hours	Aching or colicky	Epigastric, right upper quadrant; sometimes to the back, scapula, and lower chest	May follow meal and may subside spontaneously
	Pancreatitis	Prolonged	Typically aching	Epigastric, radiates to the back	
Neuromuscular	Costochondritis	Variable	Aching	Sternal	Sometimes swollen, tender, warm over joint; may be reproduced by localized pressure on examination
	Cervical disk disease	Variable; may be sudden	Aching; may include numbness	Arms and shoulders	May be exacerbated by movement of neck
	Trauma or strain	Usually constant	Aching	Localized to area of strain	Reproduced by movement or palpation
	Herpes zoster	Usually prolonged	Sharp or burning	Dermatomal distribution	Vesicular rash in area of discomfort
Psychological	Emotional and psychiatric conditions	Variable; may be fleeting or prolonged	Variable; often manifests as tightness and dyspnea with feeling of panic or doom	Variable; may be retrosternal	Situational factors may precipitate symptoms; history of panic attacks, depression

"strangling and anxiety." Chest discomfort characteristic of myocardial ischemia is typically described as aching, heavy, squeezing, crushing, or constricting. However, in a substantial minority of patients, the quality of discomfort is extremely vague and may be described as a mild tightness, or merely an uncomfortable feeling, that sometimes is experienced as numbness or a burning sensation. The site of the discomfort is usually retrosternal, but radiation is common and generally occurs down the ulnar surface of the left arm; the right arm, both arms, neck, jaw, or shoulders may also be involved. These and other characteristics of ischemic chest discomfort pertinent to discrimination from other causes of chest pain are discussed later in this chapter (see "Approach to the Patient").

Stable angina usually begins gradually and reaches its maximal intensity over a period of minutes before dissipating within several minutes with rest or with nitroglycerin. The discomfort typically occurs predictably at a characteristic level of exertion or psychological stress. By definition, unstable angina is manifest by anginal chest discomfort that occurs with progressively lower intensity of physical activity or even at rest. Chest discomfort associated with MI is commonly more severe, is prolonged (usually lasting ≥30 min), and is not relieved by rest.

OTHER CARDIOPULMONARY CAUSES

Pericardial and Other Myocardial Diseases (See also Chap. 281) Inflammation of the pericardium due to infectious or noninfectious causes can be responsible for acute or chronic chest discomfort. The visceral surface and most of the parietal surface of the pericardium are insensitive to pain. Therefore, the pain of pericarditis is thought to arise principally from associated pleural inflammation. Because of this pleural association, the discomfort of pericarditis is usually pleuritic pain that is exacerbated by breathing, coughing, or changes in position. Moreover, owing to the overlapping sensory supply of the central diaphragm via the phrenic nerve with somatic sensory fibers originating in the third to fifth cervical segments, the pain of pleural and pericardial inflammation is often referred to the shoulder and neck. Involvement of the pleural surface of the lateral diaphragm can lead to pain in the upper abdomen.

Acute inflammatory and other nonischemic myocardial diseases can also produce chest discomfort. The symptoms of acute myocarditis are highly varied. Chest discomfort may either originate with inflammatory injury of the myocardium or be due to severe increases in wall stress related to poor ventricular performance. The symptoms of *Takotsubo (stress-related) cardiomyopathy* often start abruptly with chest pain and shortness of breath. This form of cardiomyopathy, in its most recognizable form, is triggered by an emotionally or physically stressful event and may mimic acute MI because of its commonly associated ECG abnormalities, including ST-segment elevation, and elevated biomarkers of myocardial injury.

Diseases of the Aorta (See also Chap. 291) *Acute aortic dissection* (Fig. 15-1) is a less common cause of chest discomfort but is important because of the catastrophic natural history of certain subsets of cases when recognized late or left untreated. Acute aortic syndromes encompass a spectrum of acute aortic diseases related to disruption of the media of the aortic wall, including *penetrating ulcer* and *intramural hematoma* (see Chap. 291).

Each of these subtypes of acute aortic syndrome typically presents with chest discomfort that is often severe, sudden in onset, and sometimes described as "tearing" in quality. Acute aortic syndromes involving the *ascending* aorta tend to cause pain in the midline of the anterior chest, whereas *descending* aortic syndromes most often present with pain in the back. Therefore, dissections that begin in the ascending aorta and extend to the descending aorta tend to cause pain in the front of the chest that extends toward the back, between the shoulder blades.

Nontraumatic aortic dissections are very rare in the absence of hypertension or conditions associated with deterioration of the elastic or muscular components of the aortic media, including pregnancy, bicuspid aortic disease, or inherited connective tissue diseases, such as Marfan and Ehlers-Danlos syndromes. Aortitis, whether of noninfectious or infectious etiology, in the absence of aortic dissection is a rare cause of chest or back discomfort.

Pulmonary Conditions Pulmonary and pulmonary-vascular conditions that cause chest discomfort usually do so in conjunction with dyspnea and often produce symptoms that have a pleuritic nature.

PULMONARY EMBOLISM (SEE ALSO CHAP. 290) Pulmonary emboli can produce dyspnea and chest discomfort that is sudden in onset. Typically pleuritic in pattern, the chest discomfort associated with pulmonary embolism may result from (1) involvement of the pleural surface of the lung adjacent to a resultant pulmonary infarction; (2) distention of the pulmonary artery; or (3) possibly, right ventricular wall stress and/or subendocardial ischemia related to acute pulmonary hypertension. The pain associated with small pulmonary emboli is often lateral and pleuritic and is believed to be related to the first of these three possible mechanisms. In contrast, massive pulmonary emboli may cause severe substernal pain that may mimic an MI and that is plausibly attributed to the second and third of these potential mechanisms. Massive or submassive pulmonary embolism may also be associated with syncope, hypotension, and signs of right heart failure. Other typical characteristics that aid in the recognition of pulmonary embolism are discussed later in this chapter (see "Approach to the Patient").

PNEUMOTHORAX (SEE ALSO CHAP. 305) *Primary spontaneous pneumothorax* is a rare cause of chest discomfort. Risk factors include male sex, smoking, family history, and Marfan syndrome. The symptoms are usually sudden in onset, and dyspnea may be mild; thus, presentation to medical attention is sometimes delayed. *Secondary spontaneous pneumothorax* may occur in patients with underlying lung disorders, such as chronic obstructive pulmonary disease, asthma, or cystic fibrosis, and usually produces symptoms that are more severe. Tension pneumothorax is a medical emergency caused by trapped intrathoracic air that precipitates hemodynamic collapse.

Other Pulmonary Parenchymal, Pleural, or Vascular Disease (See also Chaps. 294, 295, and 305) Most pulmonary diseases that produce chest pain, including pneumonia and malignancy, do so because of involvement of the pleura or surrounding structures. Pleurisy is typically described as a knifelike pain that is worsened by inspiration or coughing. In contrast, chronic pulmonary hypertension can manifest as chest pain that may be very similar to angina in its characteristics, suggesting right ventricular myocardial ischemia in some cases. Reactive airways diseases similarly can cause chest tightness associated with breathlessness rather than pleurisy.

NONCARDIOPULMONARY CAUSES

Gastrointestinal Conditions (See also Chap. 332) Gastrointestinal disorders are the most common cause of nontraumatic chest discomfort and often produce symptoms that are difficult to discern from more serious causes of chest pain and are described in Table 15-1. Esophageal disorders, in particular, may simulate angina in the character and location of the pain. Gastroesophageal reflux and disorders of esophageal motility are common (Fig. 15-1 and Table 15-1).

Musculoskeletal and Other Causes (See also Chap. 372) Chest discomfort can be produced by any musculoskeletal disorder involving the chest wall or the nerves of the chest wall, neck, or upper limbs. Costochondritis causing tenderness of the costochondral junctions (*Tietze's syndrome*) is relatively common. Cervical radiculitis may manifest as a prolonged or constant aching discomfort in the upper chest and limbs. The pain may be exacerbated by motion of the neck. Occasionally, chest pain can be caused by compression of the brachial plexus by the cervical ribs, and tendinitis or bursitis involving the left shoulder may mimic the radiation of angina. Pain in a dermatomal distribution can also be caused by cramping of intercostal muscles or by herpes zoster (Chap. 198).

Emotional and Psychiatric Conditions As many as 10% of patients who present to EDs with acute chest discomfort have a panic disorder or related condition (Table 15-1).

APPROACH TO THE PATIENT

Chest Discomfort

Given the broad set of potential causes and the heterogeneous risk of serious complications in patients who present with acute nontraumatic chest discomfort, the priorities of the initial clinical encounter include assessment of (1) the patient's clinical stability and (2) the probability that the patient has an underlying cause of the discomfort that may be life-threatening. The high-risk conditions of principal concern are acute cardiopulmonary processes, including ACS, acute aortic syndrome, pulmonary embolism, tension pneumothorax, and pericarditis with tamponade. Fulminant myocarditis also carries a poor prognosis but is usually also manifest by heart failure symptoms. Among noncardiopulmonary causes

TABLE 15-2 Considerations in the Assessment of the Patient with Chest Discomfort

1. Could the chest discomfort be due to an acute, potentially life-threatening condition that warrants urgent evaluation and management?			
Unstable ischemic heart disease	Aortic dissection	Pneumothorax	Pulmonary embolism
2. If not, could the discomfort be due to a chronic condition likely to lead to serious complications?			
Stable angina	Aortic stenosis	Pulmonary hypertension	
3. If not, could the discomfort be due to an acute condition that warrants specific treatment?			
Pericarditis	Pneumonia/pleuritis	*Herpes zoster*	
4. If not, could the discomfort be due to another treatable chronic condition?			
Esophageal reflux		Cervical disk disease	
Esophageal spasm		Arthritis of the shoulder or spine	
Peptic ulcer disease		Costochondritis	
Gallbladder disease		Other musculoskeletal disorders	
Other gastrointestinal conditions		Anxiety state	

Source: Developed by Dr. Thomas H. Lee for the 18th edition of *Harrison's Principles of Internal Medicine.*

of chest pain, esophageal rupture likely holds the greatest urgency for diagnosis. The remaining population with noncardiopulmonary conditions has a more favorable prognosis during completion of the diagnostic workup. A rapid targeted assessment for a serious cardiopulmonary cause is most relevant for patients with acute ongoing pain who have presented for emergency evaluation. Among patients presenting in the outpatient setting with chronic pain or pain that has resolved, a general diagnostic assessment is reasonably undertaken (see "Outpatient Evaluation of Chest Discomfort," below). A series of questions that can be used to structure the clinical evaluation of patients with chest discomfort is shown in **Table 15-2**.

HISTORY

The evaluation of nontraumatic chest discomfort relies heavily on the clinical history and physical examination to direct subsequent diagnostic testing. The evaluating clinician should assess the quality, location (including radiation), and pattern (including onset and duration) of the pain as well as any provoking or alleviating factors. The presence of associated symptoms may also be useful in establishing a diagnosis.

Quality of Pain The quality of chest discomfort alone is never sufficient to establish a diagnosis. However, the characteristics of the pain are pivotal in formulating an initial clinical impression and assessing the likelihood of a serious cardiopulmonary process (Table 15-1; **Fig. 15-2**). Pressure or tightness is consistent with a typical presentation of myocardial ischemic pain. Nevertheless, the clinician must remember that some patients with ischemic chest symptoms deny any "pain" but rather complain of dyspnea or a vague sense of anxiety. The severity of the discomfort has poor diagnostic accuracy. It is often helpful to ask about the similarity of the discomfort to previous definite ischemic symptoms. It is unusual for angina to be sharp, as in knifelike, stabbing, or pleuritic; however, patients sometimes use the word "sharp" to convey the intensity of discomfort rather than the quality. Pleuritic discomfort is suggestive of a process involving the pleura, including pericarditis, pulmonary embolism, or pulmonary parenchymal processes. "Tearing" or "ripping" pain is often described by patients with acute aortic dissection. However, acute aortic emergencies also present commonly with knifelike pain. A burning quality can suggest acid reflux or peptic ulcer disease but may also occur with myocardial ischemia. Esophageal pain, particularly with spasm, can be a severe squeezing discomfort identical to angina.

Location of Discomfort A substernal location with radiation to the neck, jaw, shoulder, or arms is typical of myocardial ischemic discomfort. Radiation to both arms has a particularly high association with MI as the etiology. Some patients present with aching in sites of radiated pain as their only symptoms of ischemia. However, pain that is highly localized—e.g., that which can be demarcated by the tip of one finger—is highly unusual for angina. A retrosternal

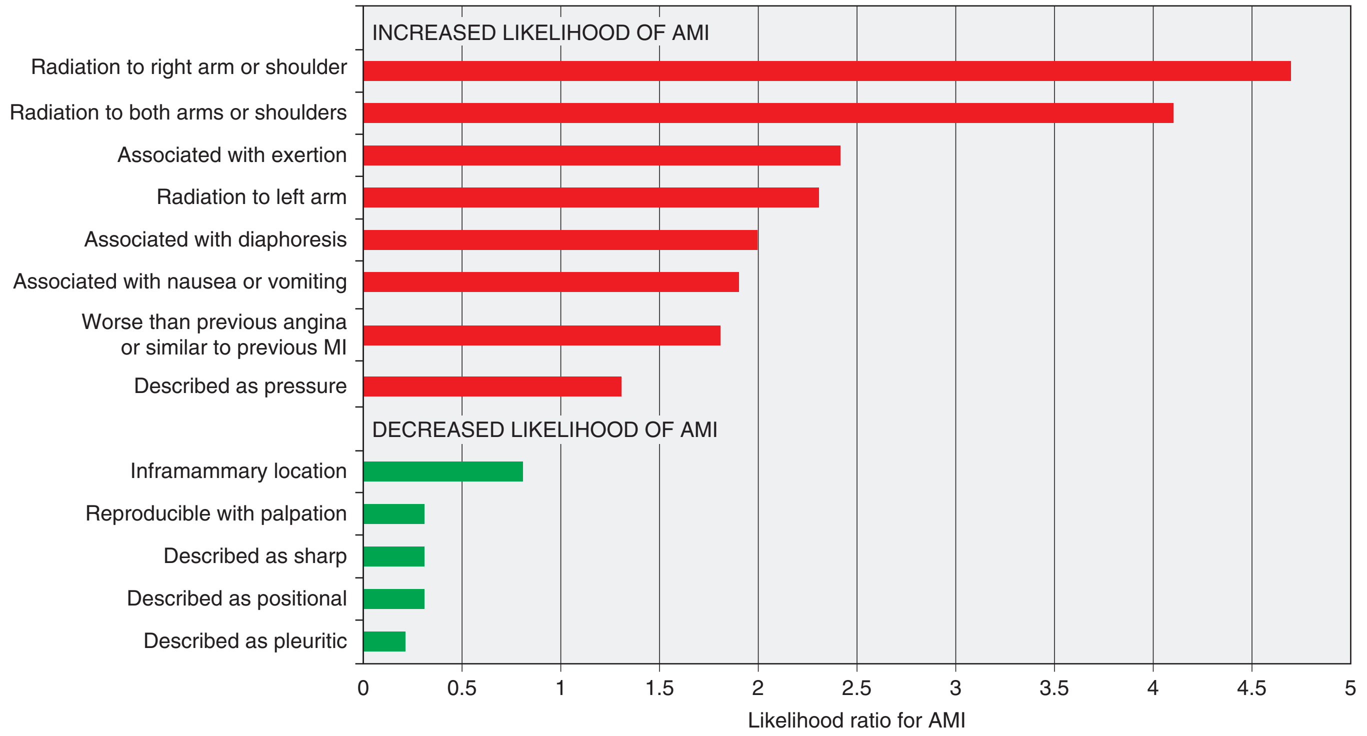

FIGURE 15-2 Association of chest pain characteristics with the probability of acute myocardial infarction (AMI). Note that a subsequent larger study showed a nonsignificant association with radiation to the right arm. *(Figure prepared from data in CJ Swap, JT Nagurney: JAMA 294:2623, 2005.)*

location should prompt consideration of esophageal pain; however, other gastrointestinal conditions usually present with pain that is most intense in the abdomen or epigastrium, with possible radiation into the chest. Angina may also occur in an epigastric location. Pain that occurs solely above the mandible or below the epigastrium is rarely angina. Severe pain radiating to the back, particularly between the shoulder blades, should prompt consideration of acute aortic syndrome. Radiation to the trapezius ridge is characteristic of pericardial pain and does not usually occur with angina.

Pattern Myocardial ischemic discomfort usually builds over minutes and is exacerbated by activity and mitigated by rest. In contrast, pain that reaches its peak intensity immediately is more suggestive of aortic dissection, pulmonary embolism, or spontaneous pneumothorax. Pain that is fleeting (lasting only a few seconds) is rarely ischemic in origin. Similarly, pain that is constant in intensity for a prolonged period (many hours to days) is unlikely to represent myocardial ischemia if it occurs in the absence of other clinical consequences, such as abnormalities of the ECG, elevation of cardiac biomarkers, or clinical sequelae (e.g., heart failure or hypotension). Both myocardial ischemia and acid reflux may have their onset in the morning.

Provoking and Alleviating Factors Patients with myocardial ischemic pain usually prefer to rest, sit, or stop walking. However, clinicians should be aware of the phenomenon of "warm-up angina" in which some patients experience relief of angina as they continue at the same or even a greater level of exertion **(Chap. 284)**. Alterations in the intensity of pain with changes in position or movement of the upper extremities and neck are less likely with myocardial ischemia and suggest a musculoskeletal etiology. The pain of pericarditis, however, often is worse in the supine position and relieved by sitting upright and leaning forward. Gastroesophageal reflux may be exacerbated by alcohol, some foods, or a reclined position. Relief can occur with sitting.

Exacerbation by eating suggests a gastrointestinal etiology such as peptic ulcer disease, cholecystitis, or pancreatitis. Peptic ulcer disease tends to become symptomatic 60–90 min after meals. However, in the setting of severe coronary atherosclerosis, redistribution of blood flow to the splanchnic vasculature after eating can trigger postprandial angina. The discomfort of acid reflux and peptic ulcer disease is usually diminished promptly by acid-reducing therapies. In contrast with its impact in some patients with angina, physical exertion is very unlikely to alter symptoms from gastrointestinal causes of chest pain. Relief of chest discomfort within minutes after administration of nitroglycerin is suggestive of but not sufficiently sensitive or specific for a definitive diagnosis of myocardial ischemia. Esophageal spasm may also be relieved promptly with nitroglycerin. A delay of >10 min before relief is obtained after nitroglycerin suggests that the symptoms either are not caused by ischemia or are caused by severe ischemia, such as during acute MI.

Associated Symptoms Symptoms that accompany myocardial ischemia may include diaphoresis, dyspnea, nausea, fatigue, faintness, and eructations. In addition, these symptoms may exist in isolation as anginal equivalents, particularly in women and the elderly. Dyspnea may occur with multiple conditions considered in the differential diagnosis of chest pain and thus is not discriminative, but the presence of dyspnea is important because it suggests a cardiopulmonary etiology. Sudden onset of significant respiratory distress should lead to consideration of pulmonary embolism and spontaneous pneumothorax. Hemoptysis may occur with pulmonary embolism or as blood-tinged frothy sputum in severe heart failure but usually points toward a pulmonary parenchymal etiology of chest symptoms. Presentation with syncope or presyncope should prompt consideration of hemodynamically significant pulmonary embolism or aortic dissection as well as ischemic arrhythmias. Although nausea and vomiting suggest a gastrointestinal disorder, these symptoms may occur in the setting of MI (more commonly inferior MI), presumably because of activation of the vagal reflex or stimulation of left ventricular receptors as part of the Bezold-Jarisch reflex.

Past Medical History The past medical history is useful in assessing the patient for risk factors for coronary atherosclerosis and venous thromboembolism **(Chap. 290)** as well as for conditions that may predispose the patient to specific disorders. For example, a history of connective tissue diseases such as Marfan syndrome should heighten the clinician's suspicion of acute aortic syndrome or spontaneous pneumothorax.

PHYSICAL EXAMINATION

In addition to providing an initial assessment of the patient's clinical stability, the physical examination of patients with chest discomfort can provide direct evidence of specific etiologies of chest pain (e.g., unilateral absence of lung sounds) and can identify potential precipitants of acute cardiopulmonary causes of chest pain (e.g., uncontrolled hypertension), relevant comorbid conditions (e.g., obstructive pulmonary disease), and complications of the presenting syndrome (e.g., heart failure). However, because the findings on physical examination may be normal in patients with unstable ischemic heart disease, an unremarkable physical exam is not definitively reassuring.

General The patient's general appearance is helpful in establishing an initial impression of the severity of illness. Patients with acute MI or other acute cardiopulmonary disorders often appear anxious, uncomfortable, pale, cyanotic, or diaphoretic. Patients who are massaging or clutching their chests may describe their pain with a clenched fist held against the sternum (*Levine's sign*). Occasionally, body habitus is helpful—e.g., in patients with Marfan syndrome or the prototypical young, tall, thin man with spontaneous pneumothorax.

Vital Signs Significant tachycardia and hypotension are indicative of important hemodynamic consequences of the underlying cause of chest discomfort and should prompt a rapid survey for the most severe conditions, such as acute MI with cardiogenic shock, massive pulmonary embolism, pericarditis with tamponade, or tension pneumothorax. Acute aortic emergencies usually present with severe hypertension but may be associated with profound hypotension when there is coronary arterial compromise or dissection into the pericardium. Sinus tachycardia is an important manifestation of submassive pulmonary embolism. Tachypnea and hypoxemia point toward a pulmonary cause. The presence of low-grade fever is nonspecific because it may occur with MI and with thromboembolism in addition to infection.

Pulmonary Examination of the lungs may localize a primary pulmonary cause of chest discomfort, as in cases of pneumonia, asthma, or pneumothorax. Left ventricular dysfunction from severe ischemia/infarction as well as acute valvular complications of MI or aortic dissection can lead to pulmonary edema, which is an indicator of high risk.

Cardiac The jugular venous pulse is often normal in patients with acute myocardial ischemia but may reveal characteristic patterns with pericardial tamponade or acute right ventricular dysfunction **(Chaps. 246 and 281)**. Cardiac auscultation may reveal a third or, more commonly, a fourth heart sound, reflecting myocardial systolic or diastolic dysfunction. Murmurs of mitral regurgitation or a ventricular-septal defect may indicate mechanical complications of STEMI. A murmur of aortic insufficiency may be a complication of ascending aortic dissection. Other murmurs may reveal underlying cardiac disorders contributory to ischemia (e.g., aortic stenosis or hypertrophic cardiomyopathy). Pericardial friction rubs reflect pericardial inflammation.

Abdominal Localizing tenderness on the abdominal exam is useful in identifying a gastrointestinal cause of the presenting syndrome. Abdominal findings are infrequent with purely acute cardiopulmonary problems, except in the case of right-sided heart failure leading to hepatic congestion.

Extremities Vascular pulse deficits may reflect underlying chronic atherosclerosis, which increases the likelihood of coronary artery disease. However, evidence of acute limb ischemia with loss of the pulse and pallor, particularly in the upper extremities, can indicate catastrophic consequences of aortic dissection. Unilateral lower-extremity swelling should raise suspicion about venous thromboembolism.

Musculoskeletal Pain arising from the costochondral and chondrosternal articulations may be associated with localized swelling, redness, or marked localized tenderness. Pain on palpation of these joints is usually well localized and is a useful clinical sign, although deep palpation may elicit pain in the absence of costochondritis. Although palpation of the chest wall often elicits pain in patients with various musculoskeletal conditions, it should be appreciated that chest wall tenderness does not exclude myocardial ischemia. Sensory deficits in the upper extremities may be indicative of cervical disk disease.

ELECTROCARDIOGRAPHY

Electrocardiography is crucial in the evaluation of nontraumatic chest discomfort. The ECG is pivotal for identifying patients with ongoing ischemia as the principal reason for their presentation as well as secondary cardiac complications of other disorders. Professional society guidelines recommend that an ECG be obtained within 10 min of presentation, with the primary goal of identifying patients with ST-segment elevation diagnostic of MI who are candidates for immediate interventions to restore flow in the occluded coronary artery. ST-segment depression and symmetric T-wave inversions at least 0.2 mV in depth are useful for detecting myocardial ischemia in the absence of STEMI and are also indicative of higher risk of death or recurrent ischemia. Serial performance of ECGs (every 30–60 min) is recommended early in the ED evaluation of suspected ACS. In addition, an ECG with right-sided lead placement should be considered in patients with clinically suspected ischemia and a nondiagnostic standard 12-lead ECG. Despite the value of the resting ECG, its sensitivity for ischemia is poor—as low as 20% in some studies.

Abnormalities of the ST segment and T wave may occur in a variety of conditions, including pulmonary embolism, ventricular hypertrophy, acute and chronic pericarditis, myocarditis, electrolyte imbalance, and metabolic disorders. Notably, hyperventilation associated with panic disorder can also lead to nonspecific ST and T-wave abnormalities. Pulmonary embolism is most often associated with sinus tachycardia but can also lead to rightward shift of the ECG axis, manifesting as an S-wave in lead I, with a Q-wave and T-wave in lead III **(Chaps. 247 and 290)**. In patients with ST-segment elevation, the presence of diffuse lead involvement not corresponding to a specific coronary anatomic distribution and PR-segment depression can aid in distinguishing pericarditis from acute MI.

CHEST RADIOGRAPHY

(See Chap. A12) Plain radiography of the chest is performed routinely when patients present with acute chest discomfort and selectively when individuals who are being evaluated as outpatients have subacute or chronic pain. The chest radiograph is most useful for identifying pulmonary processes, such as pneumonia or pneumothorax. Findings are often unremarkable in patients with ACS, but pulmonary edema may be evident. Other specific findings include widening of the mediastinum in some patients with aortic dissection, Hampton's hump or Westermark's sign in patients with pulmonary embolism **(Chaps. 290 and A12)**, or pericardial calcification in chronic pericarditis.

CARDIAC BIOMARKERS

Laboratory testing in patients with acute chest pain is focused on the detection of myocardial injury. Such injury can be detected by the presence of circulating proteins released from damaged cardiomyocytes. Owing to the time necessary for this release, initial biomarkers of injury may be in the normal range, even in patients with STEMI. Cardiac troponin is the preferred biomarker for the diagnosis of MI and should be measured in all patients with suspected ACS. It is not necessary or advisable to measure troponin in patients without suspicion of ACS unless this test is being used specifically for risk stratification (e.g., in pulmonary embolism or heart failure).

The development of cardiac troponin assays with progressively greater analytical sensitivity has facilitated detection of myocardial injury, enhanced the overall accuracy of a diagnosis of MI, and improved risk stratification in suspected ACS. For these reasons, high-sensitivity assays are preferred over prior generation troponin assays. The greater negative predictive value of a negative troponin result with high-sensitivity assays is an advantage in the evaluation of chest pain in the ED. Rapid rule-out protocols that use serial testing and changes in troponin concentration over as short a period as 1–2 h perform well for diagnosis of ACS when using a high-sensitivity troponin assay. Troponin should be measured at presentation and repeated at 1–3 h using high-sensitivity troponin and 3–6 h using conventional troponin assays. Additional troponin measurements may be warranted beyond 3–6 h when the clinical condition still suggests possible ACS or if there is diagnostic uncertainty. In patients presenting more than 2–3 h after symptom onset, a very low concentration of cardiac troponin, at the time of hospital presentation, using a high-sensitivity assay may be sufficient to exclude MI with a negative predictive value >99%.

With the use of high-sensitivity assays for troponin, myocardial injury is detected in a larger proportion of patients who have non-ACS cardiopulmonary conditions than with previous, less sensitive assays. Therefore, other aspects of the clinical evaluation are critical to the practitioner's determination of the probability that the symptoms represent ACS. In addition, observation of a change in cardiac troponin concentration between serial samples is necessary for discriminating acute causes of myocardial injury from chronic elevation due to underlying structural heart disease, end-stage renal disease, or the rare presence of interfering antibodies. The diagnosis of MI is reserved for acute myocardial injury that is marked by a rising and/or falling pattern—with at least one value exceeding the 99th percentile reference limit—*and that is caused by ischemia*. Other nonischemic insults, such as myocarditis, may result in acute myocardial injury but should not be labeled MI **(Fig. 15-3)**.

Other laboratory assessments may include the D-dimer test to aid in exclusion of pulmonary embolism **(Chap. 290)**. Measurement of a B-type natriuretic peptide is useful when considered in conjunction with the clinical history and exam for the diagnosis of heart failure.

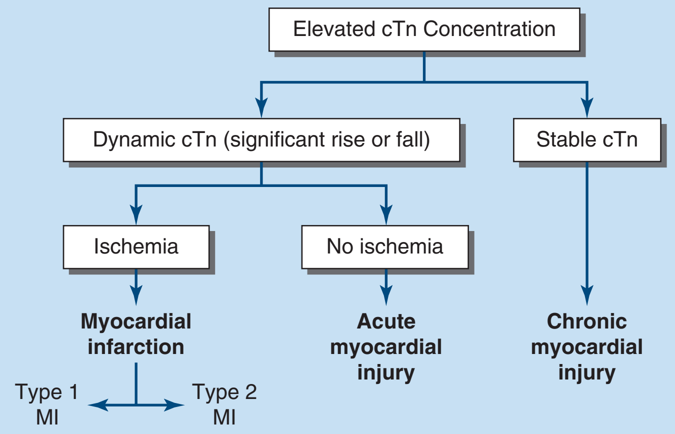

FIGURE 15-3 Clinical classification of patients with elevated cardiac troponin (cTn). MI, myocardial infarction.

INTEGRATIVE DECISION-AIDS

Multiple clinical algorithms have been developed to aid in decision-making during the evaluation and disposition of patients with acute nontraumatic chest pain. Such decision-aids estimate either of two closely related but not identical probabilities: (1) the probability of a final diagnosis of ACS and (2) the probability of major cardiac events during short-term follow-up. Use of clinical decision pathways (CDPs) is recommended to categorize patients as low, intermediate, and high risk. Evidence-based CDPs may be used to identify patients with a low clinical probability of ACS who are candidates for discharge from the ED without additional noninvasive testing. Two CDPs used commonly in current practice are shown in **Fig. 15-4**. Elements common across multiple risk stratification tools are (1) symptoms typical for ACS; (2) older age; (3) risk factors for or known atherosclerosis; (4) ischemic ECG abnormalities; and (5) elevated cardiac troponin level. Clinical application of such integrated CDPs incorporating ECGs and serial high-sensitivity cardiac troponin has been reported to achieve overall "miss rates" for ACS of <0.5% and are useful for identifying patients who are reasonable to discharge home without admission or urgent cardiac testing. In some studies, provision of such protocol-driven care in dedicated chest pain units has decreased costs and overall duration of hospital evaluation with no detectable excess of adverse clinical outcomes.

Clinicians should differentiate between the algorithms discussed above and risk scores derived for stratification of prognosis (e.g., the TIMI and GRACE risk scores, **Chap. 286**) in patients *who already have an established diagnosis of ACS*. The latter risk scores were not designed to be used for *diagnostic* assessment.

CORONARY AND MYOCARDIAL STRESS IMAGING

Among patients for whom other life-threatening causes of chest pain have been reasonably excluded and serial biomarker and clinical assessment have determined the patient to remain eligible for further testing because of intermediate or undetermined risk, diagnostic coronary imaging with coronary computed tomographic (CT) angiography or functional testing, preferably with nuclear or echocardiographic imaging, is recommended. Patient characteristics (e.g., body habitus and renal function), prior cardiac testing, history of known coronary artery disease, existing contraindications for a given test modality, and patient preferences are considerations when choosing among these diagnostic tests **(Chaps. 248 and A9)**.

CT Angiography (See Chap. 248) CT angiography has emerged as a preferred modality for the evaluation of such patients. Coronary CT angiography is a sensitive technique for detection of obstructive coronary disease. CT appears to enhance the speed to disposition of patients with a low-intermediate probability for ACS, with its major strength being the negative predictive value of a finding of no significant stenosis or coronary plaque. At the same time, CT angiography can exclude aortic dissection, pericardial effusion, and pulmonary embolism.

Stress Nuclear Perfusion Imaging or Stress Echocardiography (See Chaps. 248 and A9) Functional testing with stress nuclear perfusion imaging and stress echocardiography are alternatives for the evaluation of patients with acute chest pain who are candidates for further testing and are preferred over coronary CT angiography in patients with known obstructive epicardial disease. The selection of stress test modality may depend on institutional availability and expertise. Stress testing with myocardial imaging, either with nuclear perfusion imaging or echocardiography, offers superior diagnostic performance over exercise ECG. In patients selected for stress myocardial imaging who are able to exercise, exercise stress testing is preferred over pharmacologic testing. When available,

HEART Score (without cTn)		
History	Highly suspicious	2
	Moderately suspicious	1
	Slightly suspicious	0
ECG	Significant ST depression	2
	Nonspecific abnormality	1
	Normal	0
Age	≥65 y	2
	45–<65 y	1
	<45 y	0
Risk factors	≥3 risk factors	2
	1–2 risk factors	1
	None	0
	TOTAL	
	Low risk: 0–3 **Not low risk:** ≥4	

EDACS Score		
Age	86+ y	20
	81–85 y	18
	76–80 y	16
	Step down by 5-y increments	(–2)
	46–50 y	4
	18–45 y	2
Known CAD or risk factors	Known CAD (prior MI, PCI, or CABG) or ≥3 cardiac risk factors in patient aged ≤50 y	4
Sex	Male	6
	Female	0
Symptoms	Radiation to arm, shoulder, neck, or jaw	5
	Diaphoresis	3
	Pain with inspiration	–4
	Reproduced by palpation	–6
	TOTAL	
	Low risk: 0–15 **Not low risk:** ≥16	

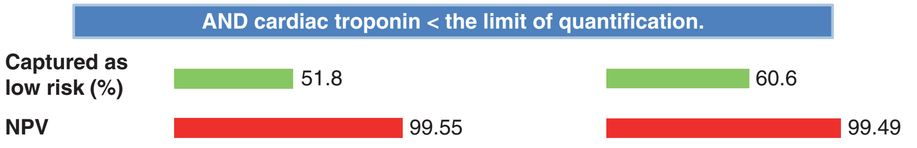

FIGURE 15-4 Examples of decision-aids used in conjunction with serial measurement of cardiac troponin (cTn) for evaluation of acute chest pain. The HEART score was modified by the authors in the presented study and omitting the assignment of 0, 1, or 2 points based on troponin. The negative predictive value (NPV) reported is for the composite endpoint of myocardial infarction (MI), cardiogenic shock, cardiac arrest, and all-cause mortality by 60 days. CABG, coronary artery bypass graft; CAD, coronary artery disease; ECG, electrocardiogram; PCI, percutaneous coronary intervention. *(Figure prepared from data in DG Mark et al: J Am Coll Cardiol 13:606, 2018.)*

positron emission tomography offers advantages of improved diagnostic performance and fewer nondiagnostic studies than single-photon emission CT.

EXERCISE ELECTROCARDIOGRAPHY

Exercise electrocardiography has historically been commonly employed for completion of risk stratification of patients who have undergone an initial evaluation that has not revealed a specific cause of chest discomfort and has identified a low risk of ACS. Early exercise testing is safe in patients without ongoing chest pain or high-risk findings and may assist in refining their prognostic assessment. However, for patients with chest pain for whom both cardiac troponin and clinical risk stratification have determined the patient to have *low* probability of ACS, there is insufficient evidence that stress testing or cardiac imaging improves their outcomes. This evolution in evidence supports a change from past practice in which stress testing within 72 hours was broadly used for patients with acute chest pain. Exercise ECG is an alternative for patients without known coronary artery disease who remain eligible for additional testing based on intermediate risk of ACS but should be performed with cardiac imaging in patients with known coronary atherosclerosis.

OTHER NONINVASIVE STUDIES

Other noninvasive imaging studies of the chest can be used selectively to provide additional diagnostic and prognostic information in patients with chest discomfort.

Echocardiography Echocardiography (nonstress) is not routinely necessarily in patients with chest discomfort. However, in patients with an uncertain diagnosis, particularly those with nondiagnostic ST elevation, ongoing symptoms, or hemodynamic instability, detection of abnormal regional wall motion provides evidence of possible ischemic dysfunction. Echocardiography is diagnostic in patients with mechanical complications of MI or in patients with pericardial tamponade. Transthoracic echocardiography is poorly sensitive for aortic dissection, although an intimal flap may sometimes be detected in the ascending aorta.

MRI (See Chap. 248) Cardiac magnetic resonance (CMR) imaging is a versatile technique for structural and functional evaluation of the heart and the vasculature of the chest. CMR can be performed as a modality for pharmacologic stress perfusion imaging and is an alternative for completing risk assessment in patients with intermediate risk of ACS. Gadolinium-enhanced CMR can provide early detection of MI, defining areas of myocardial necrosis accurately, and can delineate patterns of myocardial disease that are often useful in discriminating ischemic from nonischemic myocardial injury. Although not practical in some institutions for the urgent evaluation of acute chest discomfort, CMR can be a useful modality for cardiac structural evaluation of patients with elevated cardiac troponin levels in the absence of definite coronary artery disease. CMR coronary angiography is in its early stages. MRI also permits highly accurate assessment for aortic dissection but is infrequently used as the first test because CT and transesophageal echocardiography are usually more practical.

OUTPATIENT EVALUATION OF CHEST DISCOMFORT

Chest pain is common in outpatient practice, with a lifetime prevalence of 20–40% in the general population. More than 25% of patients with MI have had a related visit with a primary care physician in the previous month. The diagnostic principles are the same as in the ED. However, the pretest probability of acute cardiopulmonary cause is significantly lower. Therefore, testing paradigms are less intense, with an emphasis on the history, physical examination, and ECG. Moreover, decision-aids developed for settings with a high prevalence of significant cardiopulmonary disease have lower positive predictive value when applied in the practitioner's office. However, in general, if the level of clinical suspicion of ACS is sufficiently high to consider troponin testing, the patient should be referred to the ED for evaluation.

FURTHER READING

Anand A et al: High-sensitivity cardiac troponin on presentation to rule out myocardial infarction: A stepped-wedge cluster randomized controlled trial. Circulation 153:2215, 2021.

Fanaroff AC et al: Does this patient with chest pain have acute coronary syndrome? JAMA 315:1955, 2015.

Gulati M et al: 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR guideline for the evaluation and diagnosis of chest pain: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 154:e368, 2021.

Hsia RY et al: A national study of the prevalence of life-threatening diagnoses in patients with chest pain. JAMA Intern Med 176:1029, 2016.

Kontos MC et al: 2022 ACC expert consensus decision pathway on the evaluation and disposition of acute chest pain in the emergency department. J Am Coll Cardiol 80:1925, 2022.

Mahler SA et al: Safely identifying emergency department patients with acute chest pain for early discharge: HEART pathway accelerated diagnostic protocol. Circulation 138:2456, 2018.

16

Abdominal Pain

Danny O. Jacobs

Correctly diagnosing acute abdominal pain can be quite challenging. Few clinical situations require greater judgment, because the most catastrophic of events may be heralded by the subtlest of symptoms and signs. In every instance, the clinician must distinguish those conditions that require urgent intervention from those that do not and are best managed nonoperatively. A meticulously executed, detailed history and physical examination are critically important for focusing the differential diagnosis and allowing the diagnostic evaluation to proceed expeditiously **(Table 16-1)**.

The etiologic classification in **Table 16-2**, although not complete, provides a useful framework for evaluating patients with abdominal pain.

Any patient with abdominal pain of recent onset requires an early and thorough evaluation. The most common causes of abdominal pain on admission are nonspecific abdominal pain, acute appendicitis, pain of urologic origin, and intestinal obstruction. A diagnosis of "acute or surgical abdomen" is not acceptable because of its often misleading and erroneous connotations. Although most patients who present with acute abdominal pain will have self-limited disease processes, it is important to remember that pain severity does not necessarily

TABLE 16-1 Some Key Components of the Patient's History

Age
Time and mode of onset of the pain
Pain characteristics
Duration of symptoms
Location of pain and sites of radiation
Associated symptoms and their relationship to the pain
Nausea, emesis, and anorexia
Diarrhea, constipation, or other changes in bowel habits
Menstrual history

TABLE 16-2 Some Important Causes of Abdominal Pain

Pain Originating in the Abdomen	
Parietal peritoneal inflammation Bacterial contamination Perforated appendix or other perforated viscus Pelvic inflammatory disease Chemical irritation Perforated ulcer Pancreatitis Mittelschmerz Mechanical obstruction of hollow viscera Obstruction of the small or large intestine Obstruction of the biliary tree Obstruction of the ureter	Vascular disturbances Embolism or thrombosis Vascular rupture Pressure or torsional occlusion Sickle cell anemia Abdominal wall Distortion or traction of mesentery Trauma or infection of muscles Distension of visceral surfaces, e.g., by hemorrhage Hepatic or renal capsules Inflammation Appendicitis Typhoid fever Neutropenic enterocolitis or "typhlitis"
Pain Referred from Extraabdominal Source	
Cardiothoracic Acute myocardial infarction Myocarditis, endocarditis, pericarditis Congestive heart failure Pneumonia (especially lower lobes) Pulmonary embolus	Pleurodynia Pneumothorax Empyema Esophageal disease, including spasm, rupture, or inflammation Genitalia Torsion of the testis
Metabolic Causes	
Diabetes Uremia Hyperlipidemia Hyperparathyroidism	Acute adrenal insufficiency Familial Mediterranean fever Porphyria C1 esterase inhibitor deficiency (angioneurotic edema)
Neurologic/Psychiatric Causes	
Herpes zoster Tabes dorsalis Causalgia Radiculitis from infection or arthritis	Spinal cord or nerve root compression Functional disorders Psychiatric disorders
Toxic Causes	
Lead poisoning Insect or animal envenomation Black widow spider bites Snake bites	
Uncertain Mechanisms	
Narcotic withdrawal Heat stroke	

correlate with the severity of the underlying condition. In addition, the presence or absence of various degrees of "hunger" is unreliable as a sole indicator of the severity of intraabdominal disease. The most obvious of "acute abdomens" may not require operative intervention, but the mildest of abdominal pain could.

SOME MECHANISMS OF PAIN ORIGINATING IN THE ABDOMEN

Inflammation of the Parietal Peritoneum The pain of parietal peritoneal inflammation is steady and aching in character and is located directly over the inflamed area and is transmitted by somatic nerves. The intensity of the pain is dependent on the type and amount of material to which the peritoneal surfaces are exposed in a given time period. For example, the sudden release of a small quantity of *sterile* acidic gastric juice into the peritoneal cavity causes much more pain than the same amount of grossly contaminated pH neutral feces. Enzymatically active pancreatic juice incites more pain and inflammation than does the same amount of sterile bile containing no potent enzymes. Because blood is normally only a mild irritant, and the response to urine is also typically bland, exposure of blood and urine to the peritoneal cavity may go unnoticed unless it is sudden and massive. Bacterial contamination, such as may occur with pelvic inflammatory disease or perforated distal intestine, causes low-intensity pain until multiplication causes significant amounts of inflammatory mediators to be released. Patients with perforated upper gastrointestinal ulcers may present entirely differently depending on how quickly gastric juices enter the peritoneal cavity and their pH. Thus, the rate at which any inflammatory material irritates the peritoneum is important.

The pain of peritoneal inflammation is invariably accentuated by pressure or changes in tension of the peritoneum, whether produced by palpation or by movement such as with coughing or sneezing. The patient with peritonitis characteristically lies quietly in bed, preferring to avoid motion, in contrast to the patient with colic, who may be thrashing in discomfort.

Another characteristic feature of peritoneal irritation is tonic reflex spasm of the abdominal musculature, localized to the involved body segment. Its intensity depends on the integrity of the nervous system, the location of the inflammatory process, and the rate at which it developed. Spasm over a perforated retrocecal appendix or perforation into the lesser peritoneal sac may be minimal or absent because of the protective effect of overlying viscera. Catastrophic abdominal emergencies may be associated with minimal or no detectable pain or muscle spasm in obtunded, seriously ill, debilitated, immunosuppressed, or psychotic patients. A slowly developing process also often greatly attenuates the degree of muscle spasm.

Obstruction of Hollow Viscera Intraluminal obstruction classically elicits intermittent or colicky abdominal pain that is not as well localized as the pain of parietal peritoneal irritation. However, the absence of cramping discomfort can be misleading because distention of a hollow viscus may also produce steady pain with only rare paroxysms.

Small-bowel obstruction often presents as poorly localized, intermittent periumbilical or supraumbilical pain. As the intestine progressively dilates and loses muscular tone, the colicky nature of the pain may diminish. With superimposed strangulating obstruction, pain may spread to the lower lumbar region if there is traction on the root of the mesentery. The colicky pain of colonic obstruction is of lesser intensity, is commonly located in the infraumbilical area, and may often radiate to the lumbar region.

Sudden distention of the biliary tree produces a steady rather than colicky type of pain; hence, the term *biliary colic* is misleading. Acute distention of the gallbladder typically causes pain in the right upper quadrant with radiation to the right posterior region of the thorax or to the tip of the right scapula, but discomfort is also not uncommonly found near the midline. Distention of the common bile duct often causes epigastric pain that may radiate to the upper lumbar region. Considerable variation is common, however, so that differentiation between gallbladder or disease affecting the common bile duct may be impossible.

Gradual dilatation of the biliary tree, as can occur with carcinoma of the head of the pancreas, may cause no pain or only a mild aching sensation in the epigastrium or right upper quadrant. The pain of distention of the pancreatic ducts is similar to that described for distention of the common bile duct but, in addition, is very frequently accentuated by recumbency and relieved by the upright position.

Obstruction of the urinary bladder usually causes dull, low-intensity pain in the suprapubic region. Restlessness, without a specific complaint of pain, may be the only sign of a distended bladder in an obtunded patient. In contrast, acute obstruction of the intravesicular portion of the ureter is characterized by severe suprapubic and flank pain that radiates to the penis, scrotum, or inner aspect of the upper thigh. Obstruction of the ureteropelvic junction manifests as pain near the costovertebral angle, whereas obstruction of the remainder of the ureter is associated with flank pain that often extends into the same side of the abdomen.

Vascular Disturbances A frequent misconception is that pain due to intraabdominal vascular disturbances is sudden and catastrophic in nature. Certain disease processes, such as embolism or thrombosis of the superior mesenteric artery or impending rupture of an abdominal aortic aneurysm, can certainly be associated with diffuse, severe pain. Yet, just as frequently, the patient with occlusion of the superior mesenteric artery only has mild continuous or cramping diffuse pain for 2 or 3 days before vascular collapse or findings of peritoneal inflammation appear. The early, seemingly insignificant discomfort is caused by hyperperistalsis rather than peritoneal inflammation. Indeed, absence of tenderness and rigidity in the presence of continuous, diffuse pain (e.g., "pain out of proportion to physical findings") in a patient likely to have vascular disease is quite characteristic of occlusion of the superior mesenteric artery. Abdominal pain with radiation to the sacral region, flank, or genitalia should always signal the possible presence of a rupturing abdominal aortic aneurysm. This pain may persist over a period of several days before rupture and collapse occur.

Abdominal Wall Pain arising from the abdominal wall is usually constant and aching. Movement, prolonged standing, and pressure accentuate the discomfort and associated muscle spasm. In the relatively rare case of hematoma of the rectus sheath, now most frequently encountered in association with anticoagulant therapy, a mass may be present in the lower quadrants of the abdomen. Simultaneous involvement of muscles in other parts of the body usually serves to differentiate myositis of the abdominal wall from other processes that might cause pain in the same region.

REFERRED PAIN IN ABDOMINAL DISEASE

Pain referred to the abdomen from the thorax, spine, or genitalia may present a diagnostic challenge because diseases of the upper part of the abdominal cavity such as acute cholecystitis or perforated ulcer may be associated with intrathoracic complications. A most important, yet often forgotten, dictum is that the possibility of intrathoracic disease must be considered in every patient with abdominal pain, especially if the pain is in the upper abdomen.

Systematic questioning and examination directed toward detecting myocardial or pulmonary infarction, pneumonia, pericarditis, or esophageal disease (the intrathoracic diseases that most often masquerade as abdominal emergencies) will often provide sufficient clues to establish the proper diagnosis. Diaphragmatic pleuritis resulting from pneumonia or pulmonary infarction may cause pain in the right upper quadrant and pain in the supraclavicular area, the latter radiation to be distinguished from the referred subscapular pain caused by acute distention of the extrahepatic biliary tree. The ultimate decision as to the origin of abdominal pain may require deliberate and planned observation over a period of several hours, during which repeated questioning and examination will provide the diagnosis or suggest the appropriate studies.

Referred pain of thoracic origin is often accompanied by splinting of the involved hemithorax with respiratory lag and a decrease in excursion more marked than that seen in the presence of intraabdominal disease. In addition, apparent abdominal muscle spasm caused by referred pain will diminish during inspiration, whereas it persists throughout both respiratory phases if it is of abdominal origin. Palpation over the area of referred pain in the abdomen also does not usually accentuate the pain and, in many instances, actually seems to relieve it.

Thoracic disease and abdominal disease frequently coexist and may be difficult or impossible to differentiate. For example, the patient with known biliary tract disease often has epigastric pain during myocardial infarction, or biliary colic may be referred to the precordium or left shoulder in a patient who has suffered previously from angina pectoris. **For an explanation of the radiation of pain to a previously diseased area, see Chap. 14.**

Referred pain from the spine, which usually involves compression or irritation of nerve roots, is characteristically intensified by certain motions such as cough, sneeze, or strain and is associated with hyperesthesia over the involved dermatomes. Pain referred to the abdomen from the testes or seminal vesicles is generally accentuated by the slightest pressure on either of these organs. The abdominal discomfort experienced is of dull, aching character and is poorly localized.

METABOLIC ABDOMINAL CRISES

Pain of metabolic origin may simulate almost any other type of intraabdominal disease. Several mechanisms may be at work. In certain instances, such as hyperlipidemia, the metabolic disease itself may be accompanied by an intraabdominal process such as pancreatitis, which can lead to unnecessary laparotomy unless recognized. C1 esterase deficiency associated with angioneurotic edema is often associated with episodes of severe abdominal pain. Whenever the cause of abdominal pain is obscure, a metabolic origin always must be considered. Abdominal pain is also the hallmark of familial Mediterranean fever **(Chap. 381)**. Rarely, some patients with COVID-19 may present with severe abdominal pain in the absence of pulmonary symptoms.

The pain of porphyria and of lead colic is usually difficult to distinguish from that of intestinal obstruction because severe hyperperistalsis is a prominent feature of both. The pain of uremia or diabetes is nonspecific, and the pain and tenderness frequently shift in location and intensity. Diabetic acidosis may be precipitated by acute appendicitis or intestinal obstruction, so if prompt resolution of the abdominal pain does not result from correction of the metabolic abnormalities, an underlying organic problem should be suspected. Black widow spider bites produce intense pain and rigidity of the abdominal muscles and back, an area infrequently involved in intraabdominal disease.

IMMUNOCOMPROMISE

Evaluating and diagnosing causes of abdominal pain in immunosuppressed or otherwise immunocompromised patients is very difficult. This includes those who have undergone organ transplantation; who are receiving immunosuppressive treatments for autoimmune diseases, chemotherapy, or glucocorticoids; who have AIDS; and who are very old. In these circumstances, normal physiologic responses may be absent or masked. In addition, unusual infections may cause abdominal pain where the etiologic agents include cytomegalovirus, mycobacteria, protozoa, and fungi. These pathogens may affect all gastrointestinal organs, including the gallbladder, liver, and pancreas, as well as the gastrointestinal tract, causing occult or overtly symptomatic perforations of the latter. Splenic abscesses due to *Candida* or *Salmonella* infection should also be considered, especially when evaluating patients with left upper quadrant or left flank pain. Acalculous cholecystitis may be observed in immunocompromised patients or those with AIDS, where it is often associated with cryptosporidiosis or cytomegalovirus infection.

Neutropenic enterocolitis (typhlitis) is often identified as a cause of abdominal pain and fever in some patients with bone marrow suppression due to chemotherapy. Acute graft-versus-host disease should be considered in this circumstance. Optimal management of these patients requires meticulous follow-up including serial examinations to assess the need for more surgical intervention, for example, to address perforation.

NEUROGENIC CAUSES

Diseases that injure sensory nerves may cause causalgic pain. This pain has a burning character and is usually limited to the distribution of a given peripheral nerve. Stimuli that are normally not painful such as touch or a change in temperature may be causalgic and are often present even at rest. The demonstration of irregularly spaced cutaneous "pain spots" may be the only indication that an old nerve injury exists. Even though the pain may be precipitated by gentle palpation, rigidity of the abdominal muscles is absent, and the respirations are not usually disturbed. Distention of the abdomen is uncommon, and the pain has no relationship to food intake.

Pain arising from spinal nerves or roots comes and goes suddenly and is of a lancinating type **(Chap. 18)**. It may be caused by herpes zoster, impingement by arthritis, tumors, a herniated nucleus pulposus, diabetes, or syphilis. It is not associated with food intake, abdominal distention, or changes in respiration. Severe muscle spasms, when present, may be relieved by, but are usually not accentuated by, abdominal palpation. The pain is made worse by movement of the spine and is usually confined to a few dermatomes. Hyperesthesia is very common.

Pain due to functional causes conforms to none of the aforementioned patterns. Here mechanisms of disease are not as clearly established. For example, irritable bowel syndrome (IBS) is a functional gastrointestinal disorder characterized by abdominal pain and altered bowel habits. The diagnosis is made on the basis of clinical criteria **(Chap. 338)** and after exclusion of demonstrable structural abnormalities. The episodes of abdominal pain may be brought on by stress, and the pain varies considerably in type and location. Nausea and vomiting are rare. Localized tenderness and muscle spasm are inconsistent or absent. The causes of IBS or related functional disorders are not yet fully understood, although proinflammatory cells and lipotoxic lipids likely play a role.

TABLE 16-3 Differential Diagnoses of Abdominal Pain by Usual Location

Right Upper Quadrant	Epigastric	Left Upper Quadrant
Cholecystitis	Peptic ulcer disease	Splenic infarct
Cholangitis	Gastritis	Splenic rupture
Pancreatitis	GERD	Splenic abscess
Pneumonia/empyema	Pancreatitis	Gastritis
Pleurisy/pleurodynia	Myocardial infarction	Gastric ulcer
Subdiaphragmatic abscess	Pericarditis	Pancreatitis
Hepatitis	Ruptured aortic aneurysm	Subdiaphragmatic abscess
Budd-Chiari syndrome	Esophagitis	
Right Lower Quadrant	**Periumbilical**	**Left Lower Quadrant**
Appendicitis	Early appendicitis	Diverticulitis
Salpingitis	Gastroenteritis	Salpingitis
Inguinal hernia	Bowel obstruction	Inguinal hernia
Ectopic pregnancy	Ruptured aortic aneurysm	Ectopic pregnancy
Nephrolithiasis		Nephrolithiasis
Inflammatory bowel disease		Irritable bowel syndrome
Mesenteric lymphadenitis		Inflammatory bowel disease
Typhlitis		
Diffuse Nonlocalized Pain		
Gastroenteritis	Malaria	
Mesenteric ischemia	Familial Mediterranean fever	
Bowel obstruction	Metabolic diseases	
Irritable bowel syndrome	Psychiatric disease	
Peritonitis		
Diabetes		

Abbreviation: GERD, gastroesophageal reflux disease.

APPROACH TO THE PATIENT

Abdominal Pain

Few abdominal conditions require such urgent operative intervention that an orderly approach needs to be abandoned, no matter how ill the patient is. Only patients with exsanguinating intraabdominal hemorrhage (e.g., ruptured aneurysm) must be rushed to the operating room immediately, but in such instances, only a few minutes are required to assess the critical nature of the problem. Under these circumstances, all obstacles must be swept aside, adequate venous access for fluid replacement obtained, and the operation begun. Unfortunately, many of these patients may die in the radiology department or the emergency room while awaiting unnecessary examinations. *There are no absolute contraindications to operation when massive intraabdominal hemorrhage is present.* Fortunately, this situation is relatively rare. This statement does not necessarily apply to patients with intraluminal gastrointestinal hemorrhage, who can often be managed by other means **(Chap. 51)**. In these patients, obtaining a *detailed history when possible* can be extremely helpful even though it can be laborious and time-consuming. Decision-making regarding next steps is facilitated and a reasonably accurate diagnosis can be made before any further diagnostic testing is undertaken.

In cases of *acute* abdominal pain, a diagnosis can be readily established in most instances, whereas success is not so frequent in patients with *chronic* pain. IBS is one of the most common causes of abdominal pain and must always be kept in mind **(Chap. 338)**. The location of the pain can assist in narrowing the differential diagnosis **(Table 16-3)**; however, the *chronological sequence of events* in the patient's history is often more important than the pain's location. Careful attention should be paid to the extraabdominal regions. Narcotics or analgesics should *not* be withheld until a definitive diagnosis or a definitive plan has been formulated; obfuscation of the diagnosis by adequate analgesia is unlikely.

An accurate menstrual history in a female patient is essential. It is important to remember that normal anatomic relationships can be significantly altered by the gravid uterus. Abdominal and pelvic pain may occur during pregnancy due to conditions that do not require operation. Lastly, some otherwise noteworthy laboratory values (e.g., leukocytosis) may represent the normal physiologic changes of pregnancy.

In the examination, simple critical inspection of the patient, for example, of facies, position in bed, and respiratory activity, provides valuable clues. The amount of information to be gleaned is directly proportional to the *gentleness* and thoroughness of the examiner. Once a patient with peritoneal inflammation has been examined brusquely, accurate assessment by the next examiner becomes almost impossible. Eliciting rebound tenderness by sudden release of a deeply palpating hand in a patient with suspected peritonitis is cruel and unnecessary. The same information can be obtained by gentle percussion of the abdomen (rebound tenderness on a miniature scale), a maneuver that can be far more precise and localizing. Asking the patient to cough will elicit true rebound tenderness without the need for placing a hand on the abdomen. Furthermore, the forceful demonstration of rebound tenderness will startle and induce protective spasm in a nervous or worried patient in whom true rebound tenderness is not present. A palpable gallbladder will be missed if palpation is so aggressive that voluntary muscle spasm becomes superimposed on involuntary muscular rigidity.

As with history taking, sufficient time should be spent in the examination. Abdominal signs may be minimal but, nevertheless, if accompanied by consistent symptoms, may be exceptionally meaningful. Abdominal signs may be virtually or totally absent in cases of pelvic peritonitis, so careful *pelvic and rectal examinations are mandatory in every patient with abdominal pain.* Tenderness on pelvic or rectal examination in the absence of other abdominal signs can be caused by operative indications such as perforated appendicitis, diverticulitis, twisted ovarian cyst, and many others. Much attention has been paid to the presence or absence of peristaltic sounds, their quality, and their frequency. Auscultation of the abdomen is one of the least revealing aspects of the physical examination of a patient with abdominal pain. Catastrophes such as a strangulating small-intestinal obstruction or perforated appendicitis may occur in the presence of normal peristaltic sounds. Conversely, when the proximal part of the intestine above obstruction becomes markedly distended and edematous, peristaltic sounds may lose the characteristics of borborygmi and become weak or absent, even when peritonitis is not present. It is usually the severe chemical peritonitis of sudden onset that is associated with the truly silent abdomen.

Laboratory examinations may be valuable in assessing the patient with abdominal pain, yet, with few exceptions, they rarely establish a diagnosis. Leukocytosis should never be the single deciding factor as to whether or not operation is indicated. A white blood cell count >20,000/μL may be observed with perforation of a viscus, but pancreatitis, acute cholecystitis, pelvic inflammatory disease, and intestinal infarction may also be associated with marked leukocytosis. A normal white blood cell count is not rare in cases of perforation of abdominal viscera. A diagnosis of anemia may be more helpful than the white blood cell count, especially when combined with the history.

The urinalysis may reveal the state of hydration or rule out severe renal disease, diabetes, or urinary infection. Blood urea nitrogen, glucose, and serum bilirubin levels and liver function tests may be helpful. Serum amylase levels may be increased by many diseases other than pancreatitis, for example, perforated ulcer, strangulating intestinal obstruction, and acute cholecystitis; thus, elevations of serum amylase do not rule in or rule out the need for an operation.

Plain and upright or lateral decubitus radiographs of the abdomen have limited utility and may be unnecessary in some patients who have substantial evidence of some diseases such as acute appendicitis or strangulated external hernia. Where the indications for surgical or medical intervention are not clear, low-dose computed tomography is preferred to abdominal radiography when evaluating nontraumatic acute abdominal pain.

Very rarely, barium or water-soluble contrast study of the upper part of the gastrointestinal tract is an appropriate radiographic investigation and may demonstrate partial intestinal obstruction that may elude diagnosis by other means. If there is any question of obstruction of the colon, oral administration of barium sulfate should be avoided. On the other hand, in cases of suspected colonic obstruction (without perforation), a contrast enema may be diagnostic.

In the absence of trauma, peritoneal lavage has been replaced as a diagnostic tool by CT scanning and laparoscopy. Ultrasonography has proved to be useful in detecting an enlarged gallbladder or pancreas, the presence of gallstones, an enlarged ovary, or a tubal pregnancy. Laparoscopy is especially helpful in diagnosing pelvic conditions, such as ovarian cysts, tubal pregnancies, salpingitis, acute appendicitis, and other disease processes. Laparoscopy has a particular advantage over imaging in that the underlying etiologic condition can often be definitively addressed.

Radioisotopic hepatobiliary iminodiacetic acid scans (HIDAs) may help differentiate acute cholecystitis or biliary colic from acute pancreatitis. A CT scan may demonstrate an enlarged pancreas, ruptured spleen, or thickened colonic or appendiceal wall and streaking of the mesocolon or mesoappendix characteristic of diverticulitis or appendicitis.

Sometimes, even under the best circumstances with all available aids and with the greatest of clinical skill, a definitive diagnosis cannot be established at the time of the initial examination. And, in some cases, operation may be indicated based on clinical grounds alone. Should that decision be questionable, watchful waiting with repeated questioning and examination will often elucidate the true nature of the illness and indicate the proper course of action.

Acknowledgment

The author gratefully acknowledges the enormous contribution to this chapter and the approach it espouses to William Silen, who authored this chapter for many previous editions.

FURTHER READING

Bhangu A et al: Acute appendicitis: Modern understanding of pathogenesis, diagnosis and management. Lancet 386:1278, 2015.

Cartwright SL, Knudson MP: Diagnostic imaging of acute abdominal pain in adults. Am Fam Phys 91:452, 2015.

Huckins DS et al: Diagnostic performance of a biomarker panel as a negative predictor for acute appendicitis in acute emergency department patients with abdominal pain. Am J Emerg Med 35:418, 2017.

Nayor J et al: Tracing the cause of abdominal pain. N Engl J Med 375:e8, 2016.

Phillips MT: Clinical yield of computed tomography scans in the emergency department for abdominal pain. J Invest Med 64:542, 2016.

Silen W, Cope Z: *Cope's Early Diagnosis of the Acute Abdomen*, 22nd ed. New York, Oxford University Press, 2010.

17

Headache

Peter J. Goadsby

Headache is among the most common reasons patients seek medical attention and is responsible, on a global basis, for more disability than any other neurologic problem. Diagnosis and management are based on a careful clinical approach augmented by an understanding of the anatomy, physiology, and pharmacology of the nervous system pathways mediating the various headache syndromes. This chapter will focus on the general approach to a patient with headache; migraine and other primary headache disorders are discussed in **Chap. 441**.

GENERAL PRINCIPLES

A classification system developed by the International Headache Society (*www.ihs-headache.org/en/resources/guidelines/*) characterizes headache disorders as primary or secondary **(Table 17-1)**. *Primary headaches* are those in which headache and its associated features form the disorder itself, whereas *secondary headaches* are those caused by exogenous disorders (Headache Classification Committee of the International Headache Society, 2018). Primary headache often results in considerable disability and a decrease in the patient's quality of life. Mild secondary headache, such as that seen in association with upper respiratory tract infections, is common but rarely worrisome. Life-threatening headache is relatively uncommon, with vigilance required to recognize and appropriately treat such patients.

ANATOMY AND PHYSIOLOGY OF HEADACHE

Pain usually occurs when peripheral nociceptors are stimulated in response to tissue injury, visceral distension, or other factors **(Chap. 14)**. In such situations, pain perception is a normal physiologic response mediated by a healthy nervous system. Pain can also result when pain-producing pathways of the peripheral or central nervous system (CNS) are damaged or activated inappropriately. Headache may originate from either or both mechanisms. Relatively few cranial structures are pain-producing; these include the scalp, meningeal arteries, dural sinuses, falx cerebri, and proximal segments of the large pial arteries. The ventricular ependyma, choroid plexus, pial veins, and much of the brain parenchyma are not pain-producing.

The key structures involved in primary headache appear to be:

- The large intracranial vessels and dura mater and the peripheral terminals of the trigeminal nerve that innervate these structures
- The caudal portion of the trigeminal nucleus, which extends into the dorsal horns of the upper cervical spinal cord and receives input from the first and second cervical nerve roots (the trigeminocervical complex)
- Rostral pain-processing regions, such as the ventroposteromedial thalamus and the cortex
- The pain-modulatory systems in the brain that modulate input from trigeminal nociceptors at all levels of the pain-processing pathways and influence vegetative functions, such as hypothalamus and brainstem structures

TABLE 17-1 Common Causes of Headache

PRIMARY HEADACHE		SECONDARY HEADACHE	
TYPE	**%**	**TYPE**	**%**
Tension-type	69	Systemic infection	63
Migraine	16	Head injury	4
Idiopathic stabbing	2	Vascular disorders	1
Exertional	1	Subarachnoid hemorrhage	<1
Cluster	0.1	Brain tumor	0.1

Source: After J Olesen et al: *The Headaches*. Philadelphia, Lippincott Williams & Wilkins, 2005.

The innervation of the large intracranial vessels and dura mater by the trigeminal nerve is known as the *trigeminovascular system*. Cranial autonomic symptoms, such as *lacrimation, conjunctival injection, nasal congestion, rhinorrhea, periorbital swelling, aural fullness*, and *ptosis*, are prominent in the trigeminal autonomic cephalalgias (TACs), including cluster headache and paroxysmal hemicrania, and are often seen in migraine, even in children. These cranial autonomic symptoms reflect activation of cranial parasympathetic pathways, and functional imaging studies indicate that vascular changes in migraine and cluster headache, when present, are similarly driven by these cranial autonomic systems. Thus, they are secondary, and not causative, events in the headache cascade. Moreover, they can often be mistaken for symptoms or signs of cranial sinus inflammation, which is then overdiagnosed and inappropriately managed. Migraine and other primary headache disorders are not "vascular headaches"; these disorders do not reliably manifest vascular changes, and treatment outcomes cannot be predicted by vascular effects. Migraine is a brain disorder and is best understood and managed as such.

CLINICAL EVALUATION OF ACUTE, NEW-ONSET HEADACHE

The patient who presents with a new, severe headache has a differential diagnosis that is quite different from the patient with recurrent headaches over many years. In new-onset and severe headache, the probability of finding a potentially serious cause is considerably greater than in recurrent headache. Patients with recent onset of pain require prompt evaluation and appropriate treatment. Serious causes to be considered include meningitis, subarachnoid hemorrhage, epidural or subdural hematoma, glaucoma, tumor, and purulent sinusitis. When worrisome symptoms and signs are present **(Table 17-2)**, rapid diagnosis and management are critical.

A careful neurologic examination is an essential part of the first step evaluation. In most cases, patients with an abnormal examination, or a history of recent-onset headache, should be evaluated by a computed tomography (CT) or magnetic resonance imaging (MRI) study of the brain. As an initial screening procedure for intracranial pathology in this setting, CT and MRI methods appear to be equally sensitive. In some circumstances, a lumbar puncture (LP) is also required, unless a benign etiology can be otherwise established. A general evaluation of acute headache might include cranial arteries by palpation; cervical spine by the effect of passive movement of the head and by imaging; the investigation of cardiovascular and renal status by blood pressure monitoring and urine examination; and eyes by funduscopy, intraocular pressure measurement, and refraction.

The psychological state of the patient should also be evaluated because a relationship exists between head pain, depression, and anxiety. This is intended to identify comorbidity rather than provide an explanation for the headache, because troublesome headache is seldom simply caused by mood change. Although it is notable that some medicines with antidepressant actions are also effective in the preventive treatment of both tension-type headache and migraine, each symptom must be treated optimally.

TABLE 17-2 Headache Symptoms That Suggest a Serious Underlying Disorder
Sudden-onset headache
First severe headache
Vomiting that precedes headache
Subacute worsening over days or weeks
Pain induced by bending, lifting, cough
Pain that disturbs sleep or presents immediately upon awakening
Known systemic illness
"Worst" headache ever
Onset after age 55
Fever or unexplained systemic signs
Abnormal neurologic examination
Pain associated with local tenderness, e.g., region of temporal artery

Underlying recurrent headache disorders may be activated by pain that follows otologic or endodontic surgical procedures. Thus, pain about the head as the result of diseased tissue or trauma may reawaken an otherwise quiescent migraine syndrome. Treatment of the headache is largely ineffective until the cause of the primary problem is addressed.

Serious underlying conditions that are associated with headache are described below. Brain tumor is a rare cause of isolated headache and even less commonly a cause of severe pain. The vast majority of patients presenting with severe headache have a benign cause, usually migraine.

SECONDARY HEADACHE

The management of secondary headache focuses on diagnosis and treatment of the underlying condition.

MENINGITIS

Acute, severe headache with stiff neck and fever suggests meningitis. LP is mandatory. Often there is striking accentuation of pain with eye movement. Meningitis can be easily mistaken for migraine in that the cardinal symptoms of pounding headache, photophobia, nausea, and vomiting are frequently present, perhaps reflecting the underlying biology of some of the patients.

Meningitis is discussed in Chaps. 143 and 144.

INTRACRANIAL HEMORRHAGE

Acute, maximal in <5 min, severe headache lasting >5 min with stiff neck but without fever suggests subarachnoid hemorrhage. A ruptured aneurysm, arteriovenous malformation, or intraparenchymal hemorrhage may also present with headache alone. Rarely, if the hemorrhage is small or below the foramen magnum, the head CT scan can be normal. Therefore, LP may be required to diagnose definitively subarachnoid hemorrhage.

Subarachnoid hemorrhage is discussed in Chap. 440, and intracranial hemorrhage in Chap. 439.

BRAIN TUMOR

Approximately 30% of patients with brain tumors consider headache to be their chief complaint. The head pain is usually nondescript—an intermittent deep, dull aching of moderate intensity, which may worsen with exertion or change in position and may be associated with nausea and vomiting. This pattern of symptoms results from migraine far more often than brain tumor. The headache of brain tumor disturbs sleep in about 10% of patients. Vomiting that precedes the appearance of headache by weeks is highly characteristic of posterior fossa brain tumors. A history of amenorrhea or galactorrhea with headache suggests a possible prolactin-secreting pituitary adenoma or the polycystic ovary syndrome. Headache arising de novo in a patient with known malignancy suggests either cerebral metastases or carcinomatous meningitis. Head pain appearing abruptly after bending, lifting, or coughing can be due to a posterior fossa mass, a Chiari malformation, or low cerebrospinal fluid (CSF) volume.

Brain tumors are discussed in Chap. 95.

TEMPORAL ARTERITIS

(See also Chaps. 34 and 375) Temporal (giant cell) arteritis is an inflammatory disorder of arteries that frequently involves the extracranial carotid circulation. It is a common disorder of the elderly; its annual incidence is 77 per 100,000 individuals aged ≥50. The average age of onset is 70 years, and women account for 65% of cases. About half of patients with untreated temporal arteritis develop blindness due to involvement of the ophthalmic artery and its branches; indeed, ischemic optic neuropathy induced by giant cell arteritis is the major cause of rapidly developing bilateral blindness in patients >60 years. Because treatment with glucocorticoids is effective in preventing this complication, prompt recognition of the disorder is important.

Typical presenting symptoms include headache, polymyalgia rheumatica **(Chap. 375)**, jaw claudication, fever, and weight loss. Headache is the dominant symptom and often appears in association with malaise and muscle aches. Head pain may be unilateral or bilateral and is located temporally in 50% of patients but may involve any and all aspects of the cranium. Pain usually appears gradually over a few hours before peak intensity is reached; occasionally, it is explosive in onset. The quality of pain is infrequently throbbing; it is almost invariably described as dull and boring, with superimposed episodic stabbing pains similar to the sharp pains that appear in migraine. Most patients can recognize that the origin of their head pain is superficial, external to the skull, rather than originating deep within the cranium (the pain site usually identified by migraineurs). Scalp tenderness is present, often to a marked degree; brushing the hair or resting the head on a pillow may be impossible because of pain. Headache is usually worse at night and often aggravated by exposure to cold. Additional findings may include reddened, tender nodules or red streaking of the skin overlying the temporal arteries, and tenderness of the temporal or, less commonly, the occipital arteries.

The erythrocyte sedimentation rate (ESR) is often, although not always, elevated; a normal ESR does not exclude giant cell arteritis. A temporal artery biopsy followed by immediate treatment with prednisone 80 mg daily for the first 4–6 weeks should be initiated when clinical suspicion is high. The prevalence of migraine among the elderly is substantial, considerably higher than that of giant cell arteritis. Migraineurs often report amelioration of their headaches with prednisone; thus, caution must be used when interpreting the therapeutic response.

■ GLAUCOMA

Glaucoma may present with a prostrating headache associated with nausea and vomiting. The headache often starts with severe eye pain. On physical examination, the eye is often red with a fixed, moderately dilated pupil.

Glaucoma is discussed in Chap. 34.

PRIMARY HEADACHE DISORDERS

Primary headaches disorders generally manifest as headache and associated features occurring in the absence of any exogenous cause. The most common are migraine, tension-type headache, and the trigeminal autonomic cephalalgias (TACs), notably cluster headache. These entities are discussed in detail in **Chap. 441**.

■ CHRONIC DAILY OR NEAR-DAILY HEADACHE

The broad description of chronic daily headache (CDH) can be applied when a patient experiences headache on 15 days or more per month. CDH is not a single entity, nor a diagnosis; it encompasses a number of different headache syndromes, both primary and secondary **(Table 17-3)**. In aggregate, this group presents considerable disability and is thus specially mentioned here. Population-based estimates suggest that about 4% of adults have daily or near-daily headache.

TABLE 17-3 Classification of Daily or Near-Daily Headache

Primary		
>4 H DAILY	**<4 H DAILY**	**SECONDARY**
Chronic migraine[a]	Chronic cluster headache[b]	Posttraumatic Head injury Iatrogenic Postinfectious
Chronic tension-type headache[a]	Chronic paroxysmal hemicrania	Inflammatory, such as Giant cell arteritis Sarcoidosis Behçet's syndrome
Hemicrania continua[a]	SUNCT/SUNA	Chronic CNS infection
New daily persistent headache[a]	Hypnic headache	Medication-overuse headache[a]

[a]May be complicated by medication overuse. [b]Some patients may have headache >4 h/d.

Abbreviations: CNS, central nervous system; SUNA, *s*hort-lasting *u*nilateral *n*euralgiform headache *a*ttacks with cranial autonomic symptoms; SUNCT, *s*hort-lasting *u*nilateral *n*euralgiform headache attacks with *c*onjunctival injection and *t*earing.

APPROACH TO THE PATIENT

Chronic Daily Headache

The first step in the management of patients with CDH is to diagnose any secondary headache and treat that problem (Table 17-3). This can sometimes be a challenge when the underlying cause triggers worsening of a primary headache. For patients with primary headache disorders, diagnosis of the headache type will guide therapy. Preventive treatments such as tricyclics, either amitriptyline or nortriptyline at doses up to 1 mg/kg, are very useful in patients with CDH arising from migraine or tension-type headache or where the secondary cause has activated the underlying primary headache. Tricyclics are started in low doses (10–25 mg) daily and may be given 12 h before the expected time of awakening in order to avoid excessive morning sleepiness. Medicines including topiramate, valproate, propranolol, flunarizine (not available in the United States), candesartan, and the newer calcitonin gene-related peptide (CGRP) pathway monoclonal antibodies and CGRP receptor antagonists (gepants; **see Chap. 441**) are also useful when the underlying issue is migraine.

MANAGEMENT OF MEDICALLY INTRACTABLE DISABLING PRIMARY HEADACHE

The management of medically intractable headache is difficult, although recent developments in therapy have proved effective. Monoclonal antibodies to CGRP or its receptor, or CGRP receptor antagonists (gepants), are effective and well-tolerated in chronic migraine and licensed for use in clinical practice. Noninvasive neuromodulatory approaches, such as single-pulse transcranial magnetic stimulation and remote electrical neuromodulation, which appear to modulate thalamic processing or brainstem mechanisms, respectively, in migraine have been used in clinical practice with success **(Chap. 441)**. Noninvasive vagal nerve stimulation has also shown promise, particularly in cluster headache, chronic paroxysmal hemicrania, and hemicrania continua, and possibly in short-lasting unilateral neuralgiform headache attacks with cranial autonomic symptoms (SUNA) and short-lasting unilateral neuralgiform headache attacks with conjunctival injection and tearing (SUNCT) **(Chap. 441)**. Other modalities are discussed in **Chap. 441**.

MEDICATION-RELATED AND MEDICATION-OVERUSE HEADACHE

Overuse of analgesic medication for headache can aggravate headache frequency, markedly impair the effect of preventive medicines, and induce a state of refractory daily or near-daily headache called *medication-overuse headache*. A proportion of patients who stop taking analgesics will experience substantial improvement in the severity and frequency of their headache. However, even after cessation of analgesic use, many patients continue to have headache, although they may feel clinically improved in some way, especially if they have been using opioids or barbiturates regularly. The residual symptoms probably represent the underlying primary headache disorder, and most commonly, this issue occurs in patients prone to migraine.

Management of Medication Overuse: Outpatients For patients who overuse analgesic medications, it is often helpful to reduce and eliminate them, although this strategy is far from universally effective. One approach is to reduce the medication dose by 10% every 1–2 weeks. Immediate cessation of analgesic use is possible for some patients, provided there is no contraindication. Both approaches are facilitated by use of a medication diary maintained during the month or two before cessation; this helps to identify the scope of the problem. A small dose of a nonsteroidal anti-inflammatory drug (NSAID) such as naproxen, 500 mg bid, if tolerated, will help relieve residual pain as analgesic use is reduced.

NSAID overuse is not usually a problem for patients with daily headache when a NSAID with a longer half-life is taken once or twice daily. Once the patient has substantially reduced analgesic use, a preventive medication should be introduced. Another widely used approach is to commence the preventive at the same time the analgesic reduction is started; both approaches are supported by controlled trial data. A common cause of unresponsiveness to treatment is the use of a preventive when analgesics continue to be used regularly. For some patients, discontinuing analgesics is very difficult; often the best approach is to inform the patient that some degree of pain is inevitable during this initial period.

Management of Medication Overuse or Treatment-Refractory Headache: Inpatients Some patients will require hospitalization for detoxification and management. Such patients have typically failed efforts at outpatient withdrawal or have a significant medical condition, such as diabetes mellitus or epilepsy, that would complicate withdrawal as an outpatient. Following admission to the hospital, medications are withdrawn completely on the first day, in the absence of a contraindication. Antiemetics and fluids are administered as required; clonidine is used for opioid withdrawal symptoms. For acute intolerable pain during the waking hours, aspirin, 1 g IV (not approved in United States), is useful. IM chlorpromazine can be helpful at night; patients must be adequately hydrated. Three to five days into the admission, as the effect of the withdrawn substance wears off, a course of IV dihydroergotamine (DHE) can be used. DHE, administered every 8 h for 5 consecutive days, a treatment that is not stopped short if headache settles, can induce a significant remission that allows a preventive treatment to be established. Serotonin 5-HT$_3$ receptor antagonists, such as ondansetron or granisetron, or the neurokinin receptor antagonist aprepitant may be required with DHE to prevent significant nausea, and domperidone (not approved in the United States) orally or by suppository can be very helpful. Avoiding sedating or otherwise side effect–prone antiemetics is helpful. Alternatives include a 7- to 10-day course of IV lidocaine or IV divalproex sodium or use of the CGRP antagonist eptinezumab intravenously **(Table 441-6)**.

NEW DAILY PERSISTENT HEADACHE

New daily persistent headache (NDPH) is a clinically distinct syndrome with important secondary causes; these are listed in **Table 17-4**.

Clinical Presentation NDPH presents with headache on most if not all days, and the patient can clearly, and often vividly, recall the moment of onset. The headache usually begins abruptly, but onset may be more gradual; evolution over 3 days has been proposed as the upper limit for this syndrome. Patients typically recall the exact day and circumstances of the onset of headache; the new, persistent head pain does not remit. The first priority is to distinguish between a primary and a secondary cause of this syndrome. Subarachnoid hemorrhage is the most serious of the secondary causes and must be excluded either by history or appropriate investigation **(Chap. 440)**.

Secondary NDPH • **Low CSF Volume Headache** In these syndromes, head pain is positional: it begins when the patient sits or stands upright and resolves upon reclining. The pain, which is occipitofrontal, is usually a dull ache but may be throbbing. Patients with chronic low CSF volume headache typically present with a history of headache from one day to the next that is generally not present on waking but worsens during the day. Recumbency usually improves the headache within minutes, and it can take only minutes to an hour for the pain to return when the patient resumes an upright position.

The most common cause of headache due to persistent low CSF volume is CSF leak following LP **(Chap. S3)**. Post-LP headache usually begins within 48 h but may be delayed for up to 12 days. Its incidence is between 10 and 30%. Beverages with caffeine may provide temporary relief. Besides LP, index events may include epidural injection or a vigorous Valsalva maneuver, such as from lifting, straining, coughing, clearing the eustachian tubes in an airplane, or multiple orgasms. Spontaneous CSF leaks are well recognized, and the diagnosis should be considered whenever the headache history is typical, even when there is no obvious index event. As time passes from the index event, the postural nature may become less apparent; cases in which the index event occurred several years before the eventual diagnosis have been recognized. Symptoms appear to result from low volume rather than low pressure: although low CSF pressures, typically 0–50 mm CSF, are usually identified, a pressure as high as 140 mm CSF has been noted with a documented leak.

Postural orthostatic tachycardia syndrome (POTS; **Chap. 451**) can present with orthostatic headache similar to low CSF volume headache and is a diagnosis that needs consideration in this setting.

When imaging is indicated to identify the source of a presumed leak, an MRI with gadolinium is the initial study of choice **(Fig. 17-1)**. A striking pattern of diffuse meningeal enhancement is so typical that in the appropriate clinical context the diagnosis is established. Chiari malformations may sometimes be noted on MRI; in such cases, surgery to decompress the posterior fossa is *not* indicated and usually worsens the headache. Spinal MRI with T2 weighting may reveal a leak, and spinal MRI may demonstrate spinal meningeal cysts whose role in these syndromes is yet to be elucidated. The source of CSF leakage may be identified by spinal MRI with appropriate sequences, or by CT, preferably lateral decubitus digital subtraction, myelography. In the absence of a directly identified site of leakage, ^{111}In-DTPA CSF studies may demonstrate early emptying of the tracer into the bladder or slow progress of tracer across

TABLE 17-4 Differential Diagnosis of New Daily Persistent Headache

PRIMARY	SECONDARY
Migrainous-type	Subarachnoid hemorrhage
Featureless (tension-type)	Low cerebrospinal fluid (CSF) volume headache
	Raised CSF pressure headache
	Posttraumatic headache[a]
	Chronic meningitis

[a]Includes postinfectious forms.

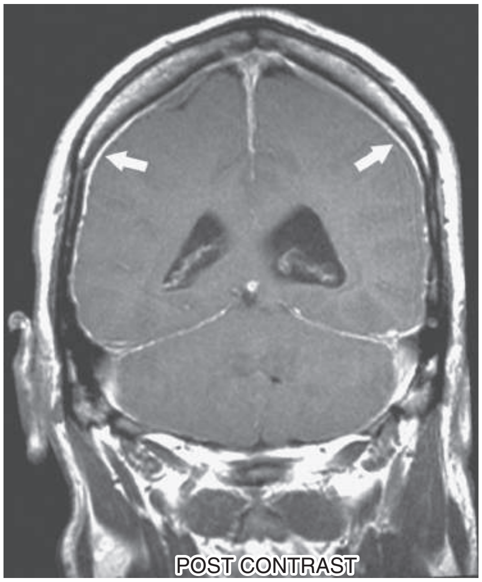

FIGURE 17-1 Magnetic resonance image showing diffuse meningeal enhancement after gadolinium administration in a patient with low cerebrospinal fluid (CSF) volume headache.

the brain suggesting a CSF leak; this procedure is now only rarely employed.

Initial treatment for low CSF volume headache is bed rest. For patients with persistent pain, IV caffeine (500 mg in 500 mL of saline administered over 2 h) can be very effective. An electrocardiogram (ECG) to screen for arrhythmia should be performed before administration. It is reasonable to administer at least two infusions of caffeine before embarking on additional tests to identify the source of the CSF leak. Because IV caffeine is safe and can be curative, it spares many patients the need for further investigations. If unsuccessful, an abdominal binder may be helpful. If a leak can be identified, an autologous blood patch is usually curative. A blood patch is also effective for post-LP headache; in this setting, the location is empirically determined to be the site of the LP. In patients with intractable headache, oral theophylline is a useful alternative that can take some months to be effective. CSF-venous fistulas require closure.

Raised CSF Pressure Headache Raised CSF pressure is well recognized as a cause of headache. Brain imaging can often reveal the cause, such as a space-occupying lesion.

Idiopathic Intracranial Hypertension (Pseudotumor Cerebri) NDPH due to raised CSF pressure can be the presenting symptom for patients with idiopathic intracranial hypertension, a disorder associated with obesity, female gender, and, on occasion, pregnancy. The syndrome can also occur without visual problems, particularly when the fundi are normal. These patients typically present with a history of generalized headache that is present on awakening, improves as the day goes on, and worsens with recumbency. Transient visual obscurations are frequent and may occur when the headaches are most severe. The diagnosis is relatively straightforward when papilledema is present, but the possibility must be considered even in patients without funduscopic changes. Formal visual field testing should be performed even in the absence of overt ophthalmic involvement. Partial obstructions of the cerebral venous sinuses are found in a small number of cases. In addition, persistently raised intracranial pressure can trigger a syndrome of chronic migraine. Other conditions that characteristically produce headache on rising in the morning or nocturnal headache are obstructive sleep apnea or poorly controlled hypertension.

Evaluation of patients suspected to have raised CSF pressure requires brain imaging. It is most efficient to obtain an MRI, including an MR venogram, as the initial study. If there are no contraindications, the CSF pressure should be measured by LP; this should be done when the patient is symptomatic so that both the pressure and the response to removal of 20–30 mL of CSF can be determined. An elevated opening pressure and improvement in headache following removal of CSF are diagnostic in the absence of fundal changes.

Initial treatment is with acetazolamide (250–500 mg bid); the headache may improve within weeks. A plan for weight loss should also be instituted when required. If ineffective, topiramate is the next treatment of choice; it has many actions that may be useful in this setting, including carbonic anhydrase inhibition, weight loss, and neuronal membrane stabilization, likely mediated via effects on phosphorylation pathways. Severely disabled patients who do not respond to medical treatment require intracranial pressure monitoring and may require shunting.

Posttraumatic Headache A traumatic event can trigger a headache process that lasts for many months or years after the event. The term *trauma* is used here in a very broad sense: headache can develop following an injury to the head, but it can also develop after an infectious episode, typically viral meningitis, a flulike illness (see below), or a parasitic infection. Complaints of dizziness, vertigo, and impaired memory can accompany the headache. Symptoms may remit after several weeks or persist for months and even years after the injury. Typically, the neurologic examination is normal and CT or MRI studies are unrevealing. Chronic subdural hematoma may on occasion mimic this disorder. Posttraumatic headache may also be seen after carotid dissection and subarachnoid hemorrhage and after intracranial surgery. The underlying theme appears to be that a traumatic event involving the pain-producing meninges can trigger a headache process that lasts for many years.

Other Causes In one series, one-third of patients with NDPH reported headache beginning after a transient flulike illness characterized by fever, neck stiffness, photophobia, and marked malaise. Evaluation typically reveals no apparent cause for the headache. There is no convincing evidence that persistent Epstein-Barr virus infection plays a role in NDPH. A complicating factor is that many patients undergo LP during the acute illness; iatrogenic low CSF volume headache must be considered in these cases. Post-COVID-19 onset of NDPH is now well documented.

Treatment Treatment is largely empirical and directed at the headache phenotype. Tricyclic antidepressants, notably amitriptyline, and anticonvulsants, such as topiramate, valproate, candesartan, and gabapentin, have been used with reported benefit, as have CGRP pathway blockers (monoclonal antibodies and gepants). The monoamine oxidase inhibitor phenelzine may also be useful in carefully selected patients. The headache usually resolves within 3–5 years, but it can be quite disabling.

PRIMARY CARE AND HEADACHE MANAGEMENT

Most patients with headache will be seen first in a primary care setting. The challenging task of the primary care physician is to identify the very few worrisome secondary headaches from the very great majority of primary and less dangerous secondary headaches (Table 17-2).

Absent any warning signs, a reasonable approach is to treat when a diagnosis is established. As a general rule, the investigation should focus on identifying worrisome causes of headache or on helping the patient to gain confidence if no primary headache diagnosis can be made.

After treatment has been initiated, follow-up care is essential to identify whether progress has been made against the headache complaint. Not all headaches will respond to treatment, but, in general, worrisome headaches will progress and will be easier to identify.

When a primary care physician feels the diagnosis is a primary headache disorder, it is worth noting that >90% of patients who present to primary care with a complaint of headache will have migraine **(Chap. 441)**.

In general, patients who do not have a clear diagnosis, have a primary headache disorder other than migraine or tension-type headache, or are unresponsive to two or more standard therapies for the considered headache type should be referred to a specialist. In a practical sense, the threshold for referral is also determined by the experience of the primary care physician in headache medicine and the availability of secondary care options.

■ FURTHER READING

Headache Classification Committee of the International Headache Society: The International Classification of Headache Disorders, 3rd ed. Cephalalgia 38:1, 2018.

Kernick D, Goadsby PJ: *Headache: A Practical Manual*, 2nd ed. Oxford, Oxford University Press, 2024.

Lance JW, Goadsby PJ: *Mechanism and Management of Headache*, 7th ed. New York, Elsevier, 2005.

Olesen J et al: *The Headaches*. Philadelphia, Lippincott, Williams & Wilkins, 2005.

Silberstein SD et al: *Wolff's Headache and Other Head Pain*, 9th ed. New York, Oxford University Press, 2021.

18

Low Back Pain

Steven P. Cohen, Eric J. Wang

EPIDEMIOLOGY

Low back pain (LBP) is among the leading causes of years lived with disability worldwide and the principal cause of work-related disability in nearly all industrialized countries. Between 28 and 34% of Americans experienced LBP in the past 3 months, with LBP accounting for >57 million unique patient visits. The all-cause medical costs in the United States are estimated to exceed $300 billion per year. Risk factors for chronic LBP include female sex, African-American race, older age, being unemployed, obesity, and sedentary lifestyle.

PAIN CATEGORIZATION

The categorization of pain is important because it predicates treatment decisions at all levels of care.

Nociceptive pain is the most common form of chronic pain in general and LBP in particular and results from activity in neural pathways secondary to actual or potentially tissue-damaging stimuli. Nociceptive LBP typically worsens with activities that stress the structures responsible for pain, is usually secondary to degenerative changes that occur over time, and with the exception of myofascial pain, tends to be progressive in nature. Mechanical pain can radiate to the upper and sometimes lower leg depending on the structure and level(s) involved and the magnitude of the stimulus (greater stimulation results in more distal radiation). However, referral patterns of mechanical pain tend to be more variable and more proximal than for radicular pain, and do not follow a dermatomal distribution.

Neuropathic pain is defined as pain caused by injury or disease affecting the somatosensory nervous system. In contrast to nociceptive pain, neuropathic pain is often accompanied by sensory abnormalities such as paresthesias, numbness, and sometimes allodynia; is more unpredictable and associated with wide fluctuations and paroxysms; and often presents with focal neurologic findings (e.g., loss of sensorimotor functions or reflexes). It is important to recognize that radiculopathy can occur without pain, and radicular pain frequently occurs in the absence of neurologic deficits.

The newest recognized category of pain is *nociplastic*, which is pain that develops due to abnormal processing of pain signals without evidence of tissue damage or pathology involving the somatosensory system (e.g., central sensitization). Nociplastic back pain, often termed "nonspecific LBP," is characterized by diffuse pain, superficial tenderness, and pain patterns that deviate from normal neuroanatomy. Patients may experience pain-induced weakness, multiple concomitant pain conditions, and sensory deficits outside of classic dermatomal maps. **Table 18-1** provides a summary of the distinguishing characteristics among nociceptive, neuropathic, and nociplastic pain conditions.

Clinical studies estimate that more than one-third (range <10–55%) of patients with chronic LBP report neuropathic qualities, with 10–20% of the overall back pain population having nociplastic pain. However, different pain categories may occur simultaneously. For example, the predisposing pathology that results in herniated disks (disk degeneration with annular tears) and spinal stenosis (bulging, degenerative disks, facet joint and ligamentum flavum hypertrophy, spondylolisthesis [e.g., anterior (anterolisthesis) or posterior (retrolisthesis) displacement of a vertebral body, decreasing the diameter of the spinal canal]) frequently results in concomitant nociceptive pain, and individuals with central sensitization often experience neuropathic and nociceptive pain at lower thresholds than other people. Studies performed in

TABLE 18-1 Distinguishing Characteristics of Nociceptive, Neuropathic, and Nociplastic Low Back Pain

CLINICAL CHARACTERISTIC	NOCICEPTIVE PAIN	NEUROPATHIC PAIN	NOCIPLASTIC PAIN
Etiology	Cumulative stress	Usually preceded by spine degeneration; herniated disk may sometimes occur after inciting event	Usually insidious, but sometimes occurs after physically or psychologically traumatic event
Onset	Insidious	Usually insidious	Usually insidious
Examples/causes	Degenerative spondylosis, myofascial pain	Herniated disk, spinal stenosis	Nonspecific back pain; may present as mechanical or radicular pain
Descriptors	Aching, deep, throbbing	Sharp, shooting, lancinating	Usually similar to neuropathic descriptors, but may include nociceptive ones as well
Sensory deficits	Uncommon	Common	Occur sometimes, but often outside of any dermatomal distribution
Motor deficits	May be pain-induced	Frequent	Pain-induced weakness, fatigue common
Hypersensitivity	Occasionally, with myofascial pain	Common	Extremely common
Pain pattern	May be referred into leg (usually proximally) in nondermatomal distribution	Reflects dermatomal pattern	Diffuse, often outside of any anatomic pain patterns
Precipitating/relieving factors	Worse with activities that stress structure	More unpredictable; spinal stenosis may be alleviated by forward flexion	Unpredictable, typically worse with stress
Autonomic signs	Uncommon	Present in up to 25% of patients	Sympathetic nervous system hyperactivity and postural orthostatic tachycardia syndrome (POTS) very common
Accompanying symptoms	Co-existing psychopathology common, and increased rate of neck pain	Higher levels of psychological stress and quality of life decrements than in nociceptive pain	Very high levels of psychological distress and sleep abnormalities; generally co-prevalent with other nociplastic conditions
Diagnosis	Imaging correlated with history and physical exam and diagnostic blocks	History and neurologic exam, instruments such as s-LANSS and painDETECT[a]	History (e.g., multiple concomitant nociplastic and nonnociplastic conditions), physical exam (e.g., diffuse tenderness), instruments (central sensitization inventory), and psychophysical tests (conditioned pain modulation)

[a] S-LANSS (self-report version of the Leeds Assessment of Neuropathic Symptoms and Signs) and painDETECT are patient-reported questionnaires used to distinguish neuropathic from nonneuropathic pain. Developed before the term *nociplastic pain* was proposed, individuals scoring in the predominantly or likely "neuropathic pain" range in the absence of identifiable nerve damage are often presumed to have nociplastic pain.

orthopedic populations suggest that over half of individuals may have mixed pain phenotypes.

NATURAL COURSE

■ MECHANICAL LBP

The distinction between acute (<3 months' duration) and chronic LBP is important as it is the major factor in determining prognosis. In one systematic review involving 11 studies with acute nonradicular pain, 80% (95% confidence interval [CI], 61–100%), 67% (95% CI, 50–83%), 57% (95% CI, 46–68%), and 65% (95% CI, 54–75%) of patients experienced pain at 1, 3, 6, and 12 months, respectively. In systematic reviews evaluating patients with predominantly chronic LBP, stagnant rates of improvement in pain were reported, with few patients improving dramatically after 6 months and a small percentage worsening.

■ RADICULAR LBP

In patients with radicular pain, between 15 and 40% of individuals experience persistent symptoms at 6 months to 1 year, with most studies also finding that herniated disks typically resorb within 2 years but often reherniate. For spinal stenosis, most patients also remain stable, with a small percentage progressing, although unlike disk herniations, the underlying pathology does not recede. Risk factors for pain persistence and poor outcomes for both axial and radicular symptoms include greater disease burden, older age, psychopathology, poor job satisfaction, and secondary gain.

ETIOLOGIES

■ NOCICEPTIVE PAIN

Myofascial Pain Muscles, ligaments, and fascia may be sources of mechanical pain, as they are imbued with nociceptors, collectively comprise a large surface area within spinal structures, are heavily involved in loadbearing and movements, and provide structural support to other potential pain generators. Studies examining muscle histology have found higher levels of neuropeptides (substance P, bradykinin), neurotransmitters (norepinephrine, 5-hydroxytryptamine), and inflammatory cytokines (e.g., tumor necrosis factor α, interleukins); lower pH levels; and more numerous vascular abnormalities in active trigger points than in latent trigger points and normal muscle. Studies using electromyography have also found higher myoelectric tone in patients with back pain compared to controls.

Back muscles can be divided into deep intrinsic muscles that connect to the vertebral column (semispinalis, rotatores, multifidus), intermediate muscles (erector spinae), and superficial muscles (e.g., latissimus dorsi). Although trigger points are frequently associated with muscle pain, these are more challenging to palpate in the low back compared to the mid-back and neck. Often co-prevalent with other etiologies or misdiagnosed as nonspecific (or nociplastic) pain, individuals with myofascial pain may present with focal or diffuse tenderness (and occasionally discrete trigger points), limited range of motion, increased muscle tension (and functional scoliosis in severe cases), and normal neurologic exams.

Discogenic Pain Disk degeneration is reported to account for 26–42% of patients with axial LBP, although selection bias (i.e., only those with suspected discogenic pain are included in discography prevalence studies), concomitant pain generators (e.g., disk degeneration predisposes to facet degeneration), the lack of a reference standard for identifying painful disks (high false-positive rate of discography), and flaws in studies utilizing diagnostic tests to identify painful disks (lack of multiple diagnostic tests with adequate controls) limit the precision of prevalence estimates. In healthy disks, nerve fibers are limited to the outer annulus, but in those with disk degeneration, they populate the inner annulus and even the nucleus pulposus. Disk degeneration is associated with upregulation of inflammatory cytokines and other molecules, which may sensitize intradiscal nerve endings and cause hypermobility, which increases the mechanical stress on disks. Macroscopically, the tearing and degeneration of annular fibers increases the stress on intact annular rings to the point of exceeding the mechanical pain threshold and facilitates contact between intradiscal cytokines and sensitized nerve endings.

Clinically, discogenic pain manifests as pain worsened with sitting or bending forward. It is more likely to be bilateral than facet or sacroiliac (SI) joint pain and frequently radiates into the upper and sometimes lower leg in a nondermatomal distribution. Since most individuals have evidence of disk degeneration by their fourth decade of life and a majority of individuals will experience LBP at least once, it can be challenging to establish a cause–effect relationship between pathology and symptoms. Provocative and analgesic discography are sometimes used to correlate degenerated disks with pain but are characterized by high false-positive rates in some populations (e.g., those with psychiatric morbidities, somatization, multiple other pain conditions) (**Fig. 18-1**).

Facet Joint Pain Facet joint pain affects approximately 10–15% of individuals with axial LBP, increasing with age. It may arise from the synovial lining, fibrous capsule, and bone, all of which are innervated with nociceptors. Disk degeneration generally precedes facet degeneration and increases loadbearing on the joints. Individuals with facet joint pain are more likely to experience unilateral, paraspinal pain and tenderness than those with predominantly discogenic pain, although the referral patterns overlap and advanced disease is usually bilateral. Individuals with facet joint arthritis may experience morning stiffness, and unlike those with discogenic pain, sitting may alleviate their symptoms. The diagnosis of facet joint pain is made via anesthetic blocks of the medial branches innervating the joints or the joints themselves, but anesthetic blocks are subject to high false-positive rates.

Sacroiliac Joint Pain Pain arising from the SI complex may be secondary to pathology involving the ligaments connecting the ilia and sacrum posteriorly and anteriorly (extraarticular) or the joint itself (intraarticular, e.g., internal bony structures, capsule, synovial lining). The SI joint is a true synovial joint, with the upper third being a syndesmosis, the lower two-thirds lined by synovium, and the lower third containing an anteriorly situated joint capsule, all of which contain nociceptors. Studies have found equal prevalence rates between extraarticular and intraarticular pathology, with the former being more common in younger individuals, after trauma, and in those with prominent tenderness and less degeneration on imaging. Pain from the SI joint is more likely to be unilateral than discogenic or facetogenic pain and is generally most marked inferior to the L5 vertebral level, with about half of patients experiencing nondermatomal pain radiating into the leg(s), including below the knee in about a quarter of cases. Depending on the pathology, pain from the SI joint may also be referred into the groin and be mistaken for hip pathology. The reference standard

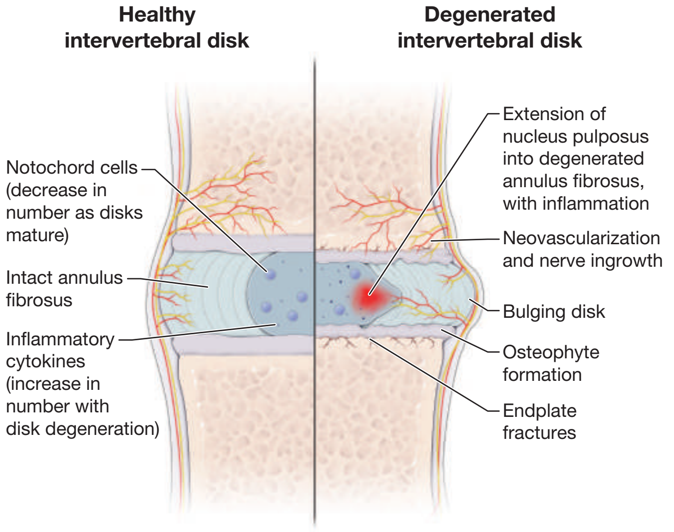
Healthy intervertebral disk
Degenerated intervertebral disk
Notochord cells (decrease in number as disks mature)
Intact annulus fibrosis
Inflammatory cytokines (increase in number with disk degeneration)
Extension of nucleus pulposus into degenerated annulus fibrosus, with inflammation
Neovascularization and nerve ingrowth
Bulging disk
Osteophyte formation
Endplate fractures

FIGURE 18-1 Schematic drawing of a coronal view demonstrating a healthy intervertebral disc (left) and a degenerated disc (right). *(Redrawn with permission from Seffrah Jin.)*

for diagnosis is low-volume anesthetic blocks, although some studies have found that a battery of three or more provocative tests on physical examination (e.g., Patrick's test [external rotation of the hip with the patient supine and the knee flexed], Gaenslen's test [leg hyperextension off the edge of the exam table in the supine position, with the other leg flexed at the knee toward the chest], SI joint distraction [dorsolateral pressure on the anterior superior iliac spines of the iliac crests with the patient supine], or compression [downward pressure on the front side of the iliac crest in the lateral position with the affected side up and the hips and knees flexed]) has high sensitivity and specificity for detecting intra-articular SI joint pain (Table 18-2).

■ RADICULAR PAIN

Herniated Disc The annual incidence of symptomatic lumbar disk herniation is about 1%, with a point prevalence between 1.5 and 4%. However, the prevalence of asymptomatic disk herniation is much higher, ranging from 29 to 43%, increasing with age. Between 38 and 56% of symptomatic individuals report an inciting event, with falls, lifting, and motor vehicle collisions being the three most common causes.

Persons with a herniated disk typically present with LBP radiating into the lower leg following a dermatomal distribution, although there is significant overlap and variability in dermatomes, and up to 40% of individuals have multilevel involvement. Patients frequently report sensory deficits and neurologic motor deficits (25–30%) and occasionally are found to have asymmetrical or diminished reflexes, most pronounced when L4 or S1 is involved (<20%). The sensitivity of the straight leg raising test (Table 18-2) is ~80% for L5 and more caudad nerve roots, with the sensitivity of the femoral stretch test (Table 18-2) exceeding 50% for mid-lumbar nerve root involvement.

Spinal Stenosis Spinal stenosis affects approximately 11% of the U.S. population, with the prevalence dramatically increasing with age. Stenosis may be central (<10 mm anteroposterior diameter) or involve the lateral recesses or foramina (<3 mm). Anatomic etiologies include bulging or herniated disks, facet joint hypertrophy, spondylolisthesis, and ligamentum flavum buckling and hypertrophy, all of which can also independently cause axial pain. Neurogenic claudication, which has a sensitivity of 88% (95% CI, 78–98%), is a hallmark of spinal stenosis but has low specificity. Symptoms of neurogenic claudication include back pain radiating into the legs that is exacerbated by activity and improved by rest, especially sitting. The most common levels affected by spinal stenosis are L4–5 (92%) and L3–4 (66%), with most people having multiple nerve root involvement. Typically, leaning forward (e.g., shopping cart sign) alleviates symptoms. Other signs and symptoms of spinal stenosis include a wide-based gait, poor balance, pain worsened by lumbar extension, and diminished vibratory perception. As with radicular symptoms secondary to a herniated disk, the diagnosis of lumbosacral stenosis (with or without neurogenic claudication) is made by a combination of history, physical examination, and imaging (e.g., magnetic resonance imaging [MRI]); see Figs. 18-2 through 18-4.

HISTORY AND PHYSICAL EXAM (SEE ALSO CHAP. V8)

History and physical examination may be used to identify patients who require further diagnostic workup and have indications for advanced therapies, including surgery, but are rarely pathognomonic. Inspection may provide clues of congenital or unusual pathology (e.g., birthmarks and doughy lipomas can indicate spina bifida, and an unusual patch of hair over the spine may indicate underlying bony pathology), while observation of gait can suggest nonspinal pathology (e.g., Parkinson's disease or antipsychotic drug use causing propulsive gait; a central lesion causing spastic gait; muscular dystrophy, spinal or gluteal muscle weakness, or hip pathology causing waddling gait; peroneal neuropathy, a large herniated disk, Guillain-Barré syndrome, multiple sclerosis, or another neurologic condition causing steppage gait or foot drop). Paraspinal tenderness overlying an area of "fullness" or increased muscle tension can indicate muscle spasm or a muscle tear, which can sometimes be distinguished through ultrasound, while midline tenderness may indicate ligamentous injury.

Spine alignment should be viewed from multiple dimensions. Scoliosis can predispose patients to disk and facet joint degeneration. Scoliosis, defined as a sideways curvature of the spine, or curvature in a coronal plane, can predispose patients to disk and facet joint degeneration. Functional scoliosis or decreased lordosis (natural inward or anterior curvature of the spine) can indicate muscle spasm or postural dysfunction (which may disappear with flexion), and exaggerated lordosis can be secondary to a tethered spinal cord or abdominal muscle weakness. Range of motion can indicate specific pathology but is most frequently associated with nonspecific pain-induced limitations. For example, decreased extension can indicate spinal stenosis or spondylolisthesis, diminished forward flexion can suggest discogenic pain, and pain when rising from sitting or with transitional movements might warrant workup for SI joint pain. True or apparent leg length discrepancies (20% have a clinically relevant leg length discrepancy exceeding 9 mm), which can be distinguished by measurements from the umbilicus, anterior superior iliac spine, or greater trochanter to the medial malleolus, may predispose patients to a host of biomechanical problems including SI joint pain, accelerated disk and facet joint degeneration, and myofascial strain. Nonorganic signs (e.g., overreaction, pain with sham stimulation) may signify underlying psychopathology and are associated with treatment failure.

Specific tests are generally more specific for radicular than axial pain. Clinical studies have found that older age, positive treadmill test (a decrease in ambulatory capacity and an increase in pain with progressively greater grades of inclination), positive Romberg's test (Chap. 433), pain that disappears with sitting, and perineal numbness have strong predictive value for lumbosacral stenosis. For detecting a herniated disk, the straight leg raising test (with the patient prone, the examiner gently straightens and raises the leg of the affected side by flexing the hip, reproducing radicular pain at an elevation between 30° and 70°) has high sensitivity but widely variable specificity for L5–S2 (L5–S1 are most commonly affected) nerve root involvement, with the femoral stretch test (the knee is passively flexed to the thigh while the hip is gently extended, reproducing radicular pain in the anterior thigh) being less studied but reasonably sensitive for mid-lumbar nerve root involvement. In contrast, the crossed straight leg raising test (eliciting radicular pain on the affected side when raising the leg on the unaffected, contralateral side) demonstrates consistently high (>85%) specificity but low sensitivity. For mechanical back pain, centralization (referred pain that is perceived as receding toward the midline with repeated movements), pain worse with sitting, and midline tenderness suggest discogenic pain, whereas paraspinal tenderness is weakly associated with injection-confirmed (e.g., lumbar medial branch nerve block) facet joint pain. An array of at least three positive SI provocation tests (e.g., Patrick's, Gaenslen's, compression, distraction; Table 18-2) is associated with accurately identifying the intraarticular SI joint as the principal pain generator. Compared to other sources of LBP, pain below L5 and radiation into the groin are also more likely to indicate SI joint pain.

A neurologic exam can indicate nerve root involvement (i.e., radicular pain), with reflexes (patellar reflex indicating L4 and sometimes L2 or L3 involvement, Achilles reflex indicating S1 pathology) being the most objective measure. However, these too must be considered in context as about 5% of younger individuals but over one-third of older individuals have absent reflexes and about one-quarter of people have asymmetrical reflexes. Although tenderness over the sciatic notch with internal rotation of the extended hip (Freiberg's test) may suggest piriformis syndrome, tenderness elicited on rectal or pelvic examination may improve selection for diagnostic injections. When cauda equina syndrome is suspected, assessing sensation in the perianal area and a rectal exam to evaluate sphincter tone is necessary, and urgent confirmation through MRI might be needed.

Table 18-3 summarizes the main etiologies of LBP and their usual clinical features, diagnostic tests, and treatments.

RED FLAGS

The term "red flags" has been used to denote signs or symptoms that suggest the potential presence of serious spinal (e.g., cauda equina syndrome) or nonspinal pathology (e.g., infectious, visceral [pelvic and

TABLE 18-2 Summary of Common Physical Exam Maneuvers for the Low Back

TEST	DESCRIPTION	COMMENTS
Lumbar Radiculopathy		
Straight leg raising (SLR)	The patient is in a supine position. The examiner passively flexes the leg of the affected side at the hip, reproducing radicular pain.	SLR has greatest sensitivity (80%) for impingement of the L5 or S1 nerve roots; sensitivity markedly decreases for nerve roots cephalad to L4. Unreliable for eliciting radicular symptoms from spinal stenosis.
Crossed SLR	The patient is in a supine position. The examiner passively flexes the leg of the nonaffected (contralateral) side at the hip, reproducing radicular pain in the affected leg.	High specificity (>85%) but low sensitivity.
Femoral stretch	The patient is in a prone position. The examiner passively extends the leg of the affected side at the hip, reproducing radicular pain in the thigh.	Modest sensitivity (50%) for L2–L4 nerve root impingement.
Sacroiliac (SI) Joint Provocation		
Compression	The patient is in a lateral decubitus position with the affected side up, with hips and knees flexed; the examiner exerts downward pressure on the superior border of the iliac crest.	There are no physical examination maneuvers that reliably distinguish between intraarticular and extraarticular pathologies, but most tests have been studied based on blocks diagnosing intraarticular pathology. Estimated sensitivities and specificities for individual tests vary greatly. The thigh thrust, FABER, and pelvic torsion tests have relatively greater sensitivities (up to 50–80%), whereas the compression and distraction tests have relatively greater specificities (up to 70–80%). A battery of ≥3 tests is generally accepted as having the greatest overall sensitivity (potentially ≥90%) and specificity (potentially ≥80%).
Thigh thrust (posterior shear test [POSH]; femoral shear test)	The patient extends their unaffected leg while in a supine position. On the affected side, the examiner flexes the patient's hip to 90° and simultaneously flexes the ipsilateral knee while applying downward pressure along the longitudinal axis of the femur.	
Distraction (gapping test)	The patient is in a supine position. On the affected side, the examiner applies downward (dorsolateral) pressure on the ipsilateral anterior superior iliac spine (ASIS).	
Flexion, abduction, and external rotation (FABER; Patrick's test)	The patient is in a supine position. On the affected side, the examiner flexes the patient's hip and knee and positions the foot under the contralateral knee (abduction). While stabilizing the contralateral ASIS with one hand, the examiner uses their other hand to apply downward pressure on the knee of the affected side (external rotation).	
Pelvic torsion (Gaenslen's test)	The patient is in a supine position, usually on the edge of an examining table. The examiner hyperextends the leg of the affected side while maximally flexing the hip and knee of the unaffected side against the patient's abdomen.	
SI Joint Mobility/Alignment		
Standing hip flexion test (SHFT; Gillet's test; Stork test)	The patient stands upright with both feet on level ground. The patient is instructed to lift one leg by flexing their hip and knee toward the chest. The examiner stands behind the patient and observes the spine and pelvis. The test is repeated in the other leg for comparison.	During hip flexion, the ipsilateral ASIS should rise slightly while the posterior superior iliac spine (PSIS) drops slightly. If these motions are paradoxical (i.e., PSIS rises with hip flexion) or asymmetrical, this suggests SI joint mobility dysfunction.
Deep Gluteal Syndrome/Piriformis Syndrome		
Freiberg's sign	The patient is in a supine position. The examiner passively extends, adducts, and internally rotates the thigh and calf ("log roll") on the affected side.	Assesses potential impingement of the sciatic nerve by nonspine structures (e.g., piriformis muscle, gluteal muscles) by stretching the piriformis and associated (e.g., gemelli) muscles, which can produce symptoms similar to those of lumbar radiculopathy.
Flexion, adduction, and internal rotation (FADIR; FAIR test)	The patient is in a supine position. On the affected side, the examiner flexes the patient's hip and knee, and while maximally adducting the thigh, internally rotates the hip.	
Pace test	In a sitting position, the patient is asked to abduct and externally rotate their hip, eliciting pain.	There are variants in which the patient lies supine with the hips and knees flexed.
Beatty test	The patient is positioned in a lateral decubitus position with the affected side up. Elevating the affected leg elicits pain in the buttocks.	Unlike the FAIR test and Freiberg's sign, this test causes contraction of the piriformis muscle and thus may not reliably elicit sciatic nerve entrapment symptoms.
Spondyloarthropathy/Ankylosing Spondylitis		
Schober test	The patient stands upright, and horizontal lines are drawn across L5 and 10 cm superior to L5. The patient is asked to bend forward and touch their toes. If the distance between the drawn lines increases <5 cm, this indicates decreased range of motion and is a positive result.	Numerous modifications exist regarding the location of the drawn lines. Nonspecific for inflammatory spinal arthritis (i.e., other conditions such as discogenic pain associated with decreased forward flexion can lead to a positive result).
Nonorganic Signs/Functional Disorders		
(Neurologic and/or psychiatric consultations potentially indicated prior to interventional procedures)		
Hoover's sign	The patient is in a supine position. The examiner asks the patient to flex the leg of the affected side at the hip, against resistance. If an organic source of neuropathy or paresis is present, with normal effort, the unaffected leg will involuntarily push downward on the examination table.	A discordant response might suggest malingering or the presence of a functional neurologic disorder (e.g., conversion disorder).
Tripod sign	With the patient in a seated position, elevating the affected leg may result in pain in the leg and back.	Leaning back and resting both hands on the table should reduce the pain. Failure to do this may suggest nonorganic pathology or malingering.
Waddell signs	Five categories of signs: (1) nonanatomic distribution of tenderness; (2) pain from sham stimulation (i.e., lumbar pain from gentle downward force on the shoulders); (3) distraction (i.e., positive SLR test in supine position but not while sitting, or while preoccupied); (4) regional disturbances (i.e., motor or sensory findings that do not correlate with areas of pathology); and (5) overreaction (i.e., disproportionate physical or emotional responses on exam).	A greater number of positive signs is associated with a greater risk of treatment failure.

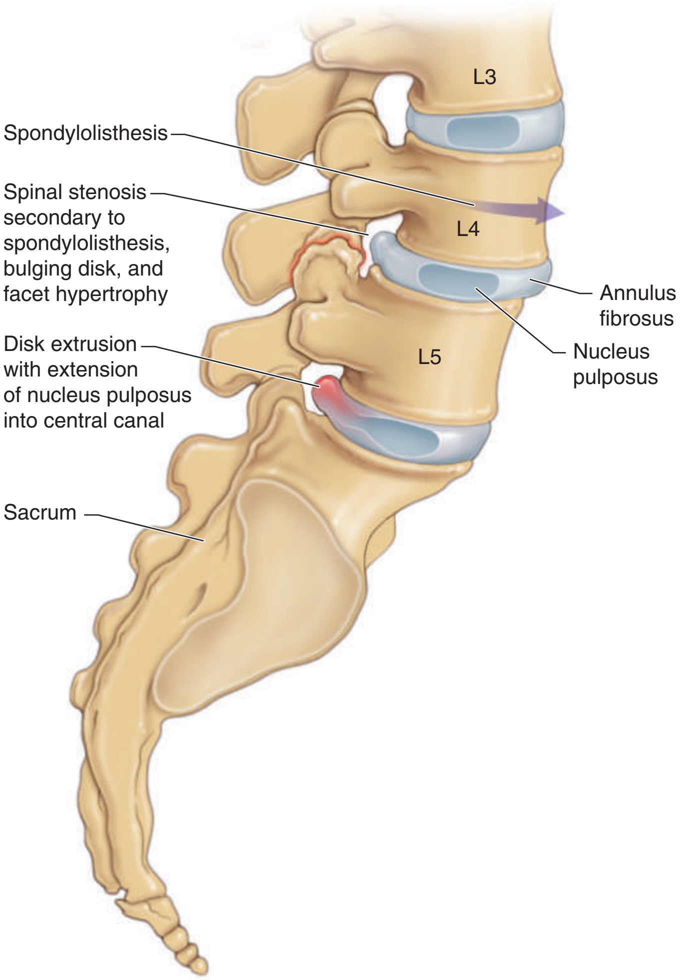
L3
Spondylolisthesis
Spinal stenosis secondary to spondylolisthesis, bulging disk, and facet hypertrophy
L4
Annulus fibrosus
Nucleus pulposus
Disk extrusion with extension of nucleus pulposus into central canal
L5
Sacrum

FIGURE 18-2 Sagittal view of the lumbar spine depicting L4-5 spinal stenosis secondary to spondylolisthesis, a bulging disc and facet joint hypertrophy, and an L5-S1 herniated nucleus pulposus. *(Redrawn with permission from Seffrah Jin.)*

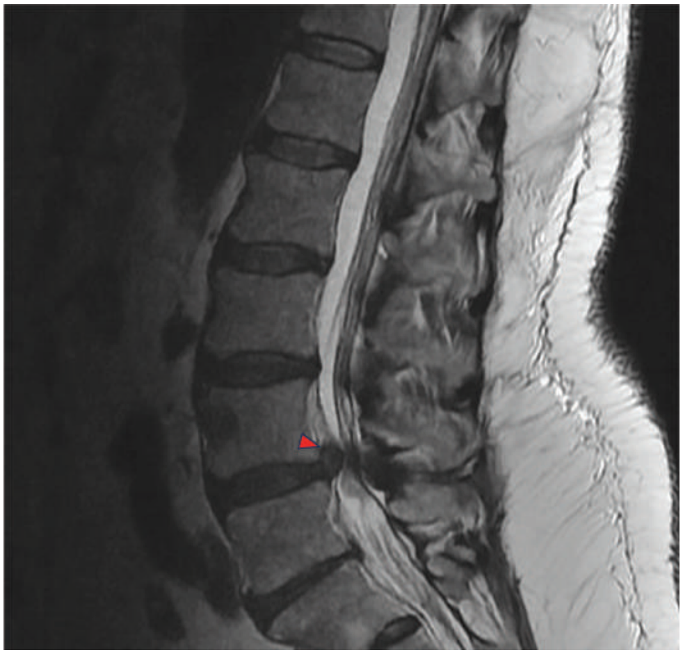

FIGURE 18-3 In this T2-weighted sagittal lumbar MRI, severe central canal stenosis secondary to a disk herniation is visualized at L4–5 (*arrowhead*). This patient appears to also have partial lumbarization of the sacrum, a risk factor for back pain.

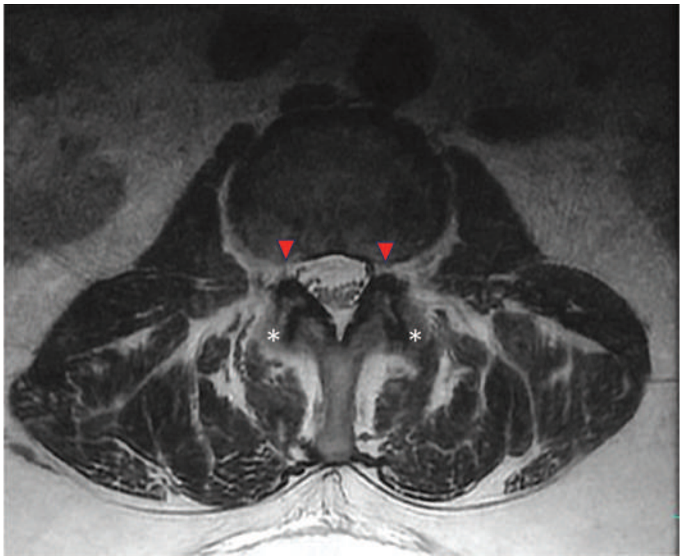

FIGURE 18-4 In this T2-weighted axial lumbar MRI, bilateral neuroforaminal narrowing secondary to protrusion of the L4–5 disk (*arrowheads*) and severe bilateral facet joint hypertrophy (*asterisks*) are visualized.

retroperitoneal organs], traumatic, vascular, neoplastic, inflammatory, or endocrine) that may lead to permanent neurologic deficits if not urgently treated. In one review of 9940 patients with a chief complaint of LBP, 92.6% of patients endorsed at least one red flag, with the most common being night pain (58.1%); although the presence of one or more red flags could predict that a neurologic emergency was present, their absence did not meaningfully decrease the likelihood. Table 18-4 summarizes red flag findings and their potential causes.

ANCILLARY TESTS

IMAGING

Advanced imaging is often used to associate symptoms with a potential etiology, but lacks specificity, with most studies demonstrating no significant correlation between imaging and symptoms, and a high percentage of abnormalities in populations without back pain. For acute LBP, red flags and serious or progressive neurologic deficits warrant imaging. For chronic LBP, MRI can be considered on a case-to-case basis, especially when considering invasive interventions such as surgery. For lumbar epidural steroid injections (ESIs), a randomized trial failed to demonstrate that MRI improved outcomes or meaningfully affected decision-making.

MRI is considered the gold standard for detecting soft tissue abnormalities including herniation, and can reveal active inflammation on certain sequences (Chap. 434). Although MRI is superior for contrast resolution between structures, computed tomography (CT) can provide better spatial contrast, is more sensitive for some bone abnormalities, and has sensitivity >90% for identifying most lumbar pathology including disk herniations. Plain films can be used to evaluate scoliosis and spondylolisthesis and detect fractures (including pars interarticularis), although MRI is needed to determine acuity and chronicity (e.g., presence or absence of edema).

As a general principle, when pain symptoms persist or worsen despite optimal nonpharmacologic and pharmacologic treatments or if red flag symptoms arise, dedicated imaging (e.g., x-ray, CT, MRI) can clarify potential diagnoses. Focal lesions on radiologic exams that are physiologically or anatomically plausible sources of pain (e.g., disk herniation with nerve root impingement) might be targets for subsequent interventional pain procedures or surgical management, but because symptom severity does not always correlate with imaging findings, a thorough history and physical examination remain the foundation for accurate contextualization. Treatment decisions should rarely be predicated on imaging findings alone.

TABLE 18-3 Clinical Evaluation, Diagnosis, and Treatment of Low Back Pain Etiologies

ETIOLOGY	RISK FACTORS	TYPICAL ONSET	CLINICAL PRESENTATION	PHYSICAL EXAM	DIAGNOSTIC TESTS	TREATMENTS[a]
Mechanical Pain						
Myofascial pain	Strenuous activities, sedentary lifestyle, abrupt movements (sneezing)	Usually insidious (chronic) but sometimes acute (muscle tear or spasm)	Axial pain, occasionally referred into thigh, mid-back, or groin. Often superimposed with other etiologies.	Tenderness, occasionally trigger points. Midline tenderness in between spinous processes may indicate ligamentous injury.	Occasionally ultrasound or electromyography	*Pharmacologic*: NSAIDs; nonbenzodiazepine muscle relaxants; antidepressants (e.g., duloxetine, tricyclic compounds); topical analgesics (e.g., diclofenac, lidocaine in those with irritable nociceptors or allodynia) *Interventional*: trigger point injections
Intervertebral disks	Advanced age (although patients younger than those with facet joint pain), acute (torsional event) or repetitive trauma, genetic predisposition	Insidious, occasionally abrupt (e.g., annular tear)	Axial pain, may radiate into extremities, more likely to be bilateral than SI or facet joint pain, worse with sitting	Midline tenderness, reduced range of motion (especially forward flexion)	Discography (high false-positive rate), bony vibration test; MRI generally nonspecific but high-intensity zone correlates with discography results	*Pharmacologic*: NSAIDs; antidepressants (e.g., duloxetine) *Interventional*: biacuplasty or IDET can be considered, but evidence for long-term benefit is mixed and disk puncture may increase risk of future disk degeneration; the benefit for intradiscal corticosteroid injections or ESI unclear
Vertebrogenic	Advanced age/osteoporosis, trauma	Abrupt (trauma) or insidious	Axial pain that may radiate into upper leg, worsened with activities	Midline tenderness	MRI to identify endplate abnormalities and acuity, plain films for acute fractures. Bony vibration test sometimes used.	*Pharmacologic*: NSAIDs; antidepressants (e.g., duloxetine); topical analgesics (e.g., diclofenac, lidocaine) only if superimposed soft tissue pain. Judicious use of opioids can be considered for severe pain secondary to acute fracture(s). *Interventional*: In acute vertebral compression fracture(s) (<6 weeks) with severe pain or disability, percutaneous vertebral augmentation (vertebroplasty or kyphoplasty) can be considered. The benefit for vertebral augmentation in chronic fractures is mixed. In patients with chronic axial pain with Modic type 1 or type 2 vertebral endplate/bone marrow changes on MRI and in whom other lumbar structures are less likely to be the predominant source of pain, basivertebral nerve RFA can be considered.
Facet joints	Increases with age, repetitive strain, genetic predisposition	Insidious	Axial pain may radiate into upper and occasionally lower legs. About half are bilateral.	Paraspinal tenderness, improved with sitting	CT or MRI generally reveals degeneration. Diagnostic medial branch or intraarticular injections.	*Pharmacologic*: NSAIDs; antidepressants (e.g., duloxetine) *Interventional*: RFA of the relevant medial branch nerves when preceded by a positive diagnostic block (e.g., ≥50% reduction in pain)
SI joint	Bimodal prevalence (younger individuals with unilateral extraarticular pathology often secondary to trauma, older people with bilateral intraarticular pathology). True and apparent leg length discrepancies, hip pathology, inflammatory arthritis (intraarticular), and lumbar spine surgery are predisposing factors.	Abrupt for extraarticular, insidious for intraarticular.	Unilateral (younger individuals, post-trauma) or bilateral (older individuals, intraarticular). About half of cases radiate into leg, sometimes below knee.	Tenderness below L5; positive provocative tests. Most painful area located near posterior superior iliac spine.	Advanced imaging (especially CT) may indicate intraarticular pathology. Battery of provocative tests correlate well with diagnostic injections.	*Pharmacologic*: NSAIDs; antidepressants (e.g., duloxetine); topical analgesics (e.g., diclofenac, lidocaine) only if superimposed soft tissue pain *Interventional*: Intra- and/or extra-articular SI joint injection; RFA of the sacral lateral branch nerves for extraarticular pathology; minimally invasive fusion of the SI joint in refractory cases involving joint degeneration or instability

(*Continued*)

TABLE 18-3 Clinical Evaluation, Diagnosis, and Treatment of Low Back Pain Etiologies (Continued)

ETIOLOGY	RISK FACTORS	TYPICAL ONSET	CLINICAL PRESENTATION	PHYSICAL EXAM	DIAGNOSTIC TESTS	TREATMENTS[a]
Predominantly Radicular Pain						
Herniated disk	Peak prevalence 30–50 years, preexisting disk degeneration, trauma, genetic predisposition, lifestyle (heavy lifting, smoking), obesity	Often abrupt, may be insidious	May be unilateral or bilateral (large, central herniation) pain radiating into leg in dermatomal distribution	Most patients also have back pain. Sensory loss and neurologic weakness common; may be associated with diminished reflexes. Straight leg raising test usually positive for lower lumbar levels.	MRI or CT scan, myelography when contraindicated. Electrodiagnostic tests or selective nerve root blocks may be confirmatory.	*Pharmacologic*: NSAIDs; nonbenzodiazepine muscle relaxants; antidepressants (e.g., duloxetine). Gabapentinoids are commonly used (gabapentin, pregabalin), but systematic reviews and meta-analyses suggest minimal benefit when balanced against potential harms (e.g., dizziness, somnolence, gait disturbance). *Interventional*: ESIs, which are likely most effective in the acute phase of radicular pain secondary to disk herniation. ESIs might confer a weak but potentially meaningful surgery-sparing effect.
Spinal stenosis	Advanced age, concomitant spinal pathology (disk degeneration, spondylolisthesis, facet hypertrophy)	Usually insidious	Usually unilateral for lateral recess or foraminal stenosis; central stenosis may be unilateral or bilateral. Wide-based gait, often affects multiple dermatomes. Pain often improves with sitting. Although the severity of pain and disability may wax and wane, the causes of spinal stenosis (e.g., spondylolisthesis, degenerative changes/osteophyte formation) often progress with age.	Sensory and motor deficits; SLR test usually negative. Improved walking ability when bending forward.	MRI or CT scan	*Pharmacologic*: NSAIDs; nonbenzodiazepine muscle relaxants; antidepressants (e.g., duloxetine). Gabapentinoids are commonly used (gabapentin, pregabalin), but systematic reviews and meta-analyses suggest minimal benefit when balanced against potential harms, especially in the elderly (e.g., dizziness, somnolence, gait disturbance). *Interventional*: ESIs commonly used and may confer short-term and sometimes intermediate-term benefit, but evidence for long-term benefit is mixed given the progressive pathology.
Neuropathy	Underlying neuropathic pain condition (e.g., herpes zoster, diabetic amyotrophy)	Generally insidious, acute for herpes zoster	Lumbosacral dermatomes less common with herpes zoster than thoracic dermatomes. Unilateral (herpes zoster), occasionally bilateral (advanced neuropathies). Advanced neuropathies usually multidermatomal.	Sensory and motor loss, diminished reflexes	Workup for underlying cause; advanced imaging may rule out disk pathology	*Pharmacologic*: gabapentinoids (gabapentin, pregabalin), antidepressants (e.g., SNRIs, TCAs), topical local anesthetics (e.g., lidocaine patches or creams), topical capsaicin. Other anticonvulsants (e.g., oxcarbazepine, lamotrigine) are occasionally used, but minimal evidence is available. *Interventional*: peripheral nerve blocks might confer analgesia, but evidence for long-term benefit is minimal or negative for certain conditions (e.g., PHN).
Nociplastic Pain						
Nonspecific back pain	Central sensitization, more common in females, peak age 20s to 50s, genetic predisposition	Usually insidious, but >20% report a physical or psychological inciting event	Bilateral, often involves other spinal regions, frequently radiates into legs. Most patients have other nociplastic and nonnociplastic pain conditions. High co-prevalence rate of mood disorders and sleep dysfunction.	Marked diffuse tenderness, provocative tests often positive but nonspecific, neurologic exam nonfocal	Based on clinical presentation. Patients often score in "neuropathic" range on validated instruments (painDETECT); central sensitization inventory and conditioned pain modulation may confirm central sensitization.	*Pharmacologic*: NSAIDs; nonbenzodiazepine muscle relaxants; antidepressants (e.g., duloxetine); gabapentinoids (gabapentin, pregabalin); over-the-counter topical analgesics (e.g., diclofenac, lidocaine, menthol). Nonpharmacologic therapies (e.g., physical therapy, targeted exercise) should be prioritized.

Mixed Pain Phenotypes						
Most neuropathic conditions have concomitant mechanical pain (e.g., spinal stenosis may be from facet hypertrophy or degenerated/herniated disks, which in themselves can cause mechanical pain); herpes zoster can cause cutaneous inflammatory pain.						
Postlaminectomy syndrome	Younger age, female sex, opioid use, greater presurgical disease burden, more concurrent pain conditions (e.g., baseline central sensitization)	Poor patient selection (e.g., irrelevant surgical levels), pseudoarthroses, and retained fragments may result in no benefit. Epidural scar tissue, arachnoiditis, and muscle wasting may present with worsening pain weeks or months after surgery. Adjacent segment disease can take >1 year to manifest.	Often involves axial and radicular pain in a multidermatomal fashion	Sensory loss and weakness common. Neurologic signs may be in multiple dermatomes. Muscle wasting and/or loss or lordosis may be appreciated. Superficial tenderness is common in individuals with nociplastic pain.	MRI with gadolinium, or even myelography. Diagnostic blocks (e.g., discography, facet blocks) are associated with a high false-positive rate.	*Pharmacologic*: NSAIDs; nonbenzodiazepine muscle relaxants; antidepressants; gabapentinoids (depending on whether neuropathic or nociceptive symptoms predominate) *Interventional*: Epidural lysis of adhesions can be considered in individuals with epidural scar tissue. SCS can be considered in refractory cases, especially those in whom neuropathic symptoms predominate.

[a] Nonpharmacologic treatments such as physical therapy, along with the management of concomitant mood disorders and maladaptive coping mechanisms (e.g., via pain psychology or CBT), should be considered for all patients with chronic pain.

Abbreviations: CBT, cognitive-based therapy; CT, computed tomography; ESI, epidural steroid injection; IDET, intradiscal electrothermal therapy; MRI, magnetic resonance imaging; NSAID, nonsteroidal anti-inflammatory drug; PHN, postherpetic neuralgia; RFA, radiofrequency ablation; SCS, spinal cord stimulation; SI, sacroiliac; SLR, straight leg raise; SNRI, serotonin-norepinephrine reuptake inhibitor; TCA, tricyclic antidepressant.

TABLE 18-4 "Red Flag" Symptoms and Corresponding Pathology

Demographics	
Age ≤18 years	Congenital defect, tumor, spondylolysis, or spondylolisthesis
Age >50 years	Tumor, fracture, vascular abnormality (aortic aneurysm)
Social and Treatment Related	
Intravenous drug use	Infection
Anticoagulant use	Hematoma
Recent procedure	Hematoma (complication after spine procedure) or infection
Immunocompromised state	Infection
Trauma	Fracture, hematoma
History of cancer	Tumor
Symptoms	
Fever, night sweats, chills	Infection, tumor
Weight loss	Tumor, infection
Saddle anesthesia	Cauda equina syndrome
Urinary or rectal incontinence, sexual dysfunction	Cauda equina syndrome
Rapidly progressive or severe neurologic symptoms	Cauda equina syndrome
Pain not relieved by rest or at night	Tumor, infection
Physical Exam Signs	
Saddle anesthesia	Cauda equina syndrome
Decreased rectal tone	Cauda equina syndrome

■ ELECTRODIAGNOSTIC TESTING

Electromyography and nerve conduction studies are often used to identify peripheral sources of nerve and muscle injury. This determination is particularly pertinent in cases of multiple dermatomal involvement or atypical extremity pain, when symptoms and imaging findings are conflicting, and in cases of transitional anatomy or aberrant innervation. Transitional anatomy, which includes variations of "lumbarization" of the S1 spinal segment (i.e., a sixth vertebral body is present), or more commonly "sacralization" of L5 (i.e., a partial or complete anatomic fusion of L5 with S1), is present in 15–35% of the population and is associated with an increased prevalence of back pain. Studies have generally found sensitivity rates ranging from 36 to 64% for radicular pain in the absence of focal neurologic findings and from 51 to 86% in patients with an abnormal neurologic examination. The specificity of electrodiagnostic testing is also variable, ranging between 50 and 60%.

■ SELECTIVE NERVE ROOT BLOCKS

Selective nerve root blocks (SNRBs) can also be used to identify a symptomatic nerve root in ambiguous cases. SNRBs are performed in a manner similar to that of transforaminal ESIs but, by definition, involve the blockade of only a single nerve root (i.e., avoidance of epidural spread), with studies finding that high-volume injections (>0.5 mL) undermine specificity. There have been no randomized studies evaluating the effect of SNRB on postsurgical decompression outcomes, although retrospective studies have generally found a modest correlation between pain relief after SNRB and surgery. One systematic review assessing the accuracy of SNRB in detecting radiculopathy in patients with LBP and lower limb pain found 93% sensitivity and 26% specificity based on very low-quality studies and concluded that the addition of SNRB to routine presurgical workup is not cost-effective.

TREATMENT

Low Back Pain

MULTIMODAL AND INTERDISCIPLINARY STRATEGIES

Contemporary strategies for managing acute and chronic LBP prioritize the optimization of nonpharmacologic modalities (e.g., physical therapy, exercise, cognitive-behavioral therapy, heat, massage) with a graded, patient-centered incorporation of pharmacologic, interventional (e.g., fluoroscopic injections), and surgical treatments for increasingly refractory symptoms. Because most cases of acute LBP resolve within 6 weeks, over-the-counter analgesics in addition to nonpharmacologic treatments such as targeted physical therapy, core-strengthening exercises, and education frequently suffice without the need for imaging or prescription analgesics. The biopsychosocial model recognizes that psychological (e.g., underlying mood disorders) and social factors (e.g., systemic barriers to care) contribute to the overall perception and experience of pain, and these issues should be identified and addressed when possible. An interdisciplinary team that involves specialists from pain medicine, orthopedic, and/or neurosurgical spine surgery, physical therapy, and psychology/psychiatry can facilitate a more nuanced, individualized, and comprehensive treatment plan. Studies have found that interdisciplinary treatment programs involving these specialties provide better improvements in pain, function, and quality of life, but restrictions in coverage from health care payors have hindered widespread implementation.

Pharmacotherapy Around a quarter of patients with acute LBP develop chronic symptoms. Chronic LBP is defined by symptoms that persist for >3 months. Pharmacologic options include nonsteroidal anti-inflammatory drugs (NSAIDs), nonbenzodiazepine muscle relaxants, and antidepressants (e.g., duloxetine, a serotonin and norepinephrine reuptake inhibitor, or tricyclic antidepressants) **(Chap. 14)**. Although frequently used, acetaminophen is unlikely to provide significant analgesia for back pain and is no longer recommended as a first-line agent. There is insufficient evidence for gabapentinoids (gabapentin or pregabalin) for either axial or radicular back pain. Opioids have not demonstrated significant long-term benefits for analgesia or function, but a temporary course may be considered on a case-by-case basis for debilitating acute pain or severe exacerbations of chronic LBP. Opioids are associated with risks of serious potential harms (e.g., respiratory depression, addiction, endocrinologic disturbances). If used, opioids should be prescribed at the lowest effective dose for the shortest duration of time feasible and with clearly defined treatment goals collaboratively made with the patient. Concomitant use of opioids with benzodiazepines should be avoided due to the increased risk of respiratory depression.

Psychological Therapies There is a high co-prevalence between psychopathology (e.g., depression, anxiety disorders, catastrophization, poor coping skills, somatic symptom disorder, fear avoidance, posttraumatic stress disorder, substance use disorders) and chronic back pain. Studies have found co-prevalence rates for depression ranging between 33 and 67%, for anxiety between 10 and 30%, for substance misuse disorders between 13 and 40%, and for axis II disorders (e.g., personality disorders), >50% in some studies. The lifetime co-prevalence rates of axis I and axis II conditions in individuals with chronic back pain are even higher. It is important to recognize that psychiatric conditions are not binary (present or absent), but rather exist along a continuum **(Chap. 463)**. Many chronic pain sufferers may still benefit from precision psychotherapies despite not formally meeting contemporaneous diagnostic criteria.

It is therefore important to identify and address underlying psychiatric and mood conditions. Studies have shown that targeted education, mindfulness-based stress reduction, operant therapy, biofeedback, progressive relaxation, and cognitive-behavioral therapies may benefit patients with back pain from various etiologies, with evidence generally being greatest for patients with chronic pain. No individual psychological therapy has demonstrated consistent superiority, and it is likely that the effectiveness of these therapies is greatly dependent upon the provider-specific and patient-specific characteristics that undergird the therapeutic relationships of psychiatric care. There is stronger and more consistent evidence for short-term than long-term benefit on pain and function, with the benefits waning without ongoing follow-up.

Physical Therapies Physical therapies have been a cornerstone of back pain treatment for decades. Physical therapists evaluate and educate patients regarding kinesiologic or functional abnormalities that contribute to pain and provide minimally invasive procedural interventions to help reduce symptoms and dysfunction. Physical therapists develop exercise regimens to address underlying causes of pain (e.g., correcting gait abnormalities) and provide treatments (e.g., hot and cold packs, manual therapies including manipulation, massage, neuromuscular reeducation). Exercise has been shown to reduce pain and increase function for radicular and nonradicular back pain, although the effects diminish over time if exercises cease. For acute back pain, although early resumption of activities including exercise (within 2 weeks of symptom onset) is widely recommended, studies are mixed regarding its long-term effect on pain and function. Most studies have failed to demonstrate one type of exercise as more beneficial than another, with yoga and Tai Chi being two of the more commonly studied therapies.

Integrative Medicine The use of integrative treatments for back pain has grown substantially but continues to be characterized by low-quality studies. Integrative medicine therapies can be provided via specialists (e.g., acupuncturists, chiropractors), physical therapies, and physicians. Although there are no trials comparing therapies stratified by specialty, specialists (acupuncturists, chiropractors) may have a greater knowledge base and more experience than generalists, which could theoretically be helpful for refractory cases. Massage may be beneficial in individuals with acute and chronic pain with prominent soft tissue symptoms (e.g., spasmodic or tension-based pain), but the analgesic benefits tend to be short-lived. Spinal manipulation may provide small benefits for acute and chronic back pain and physical function compared to the absence of therapy, but the effects diminish over time. Spinal manipulation is noninferior to other recommended physical therapies; there is mixed evidence for its benefit compared to sham or as an add-on treatment to other physical therapies. Acupuncture has been shown to be effective for pain and, to a lesser degree, several secondary outcomes in patients with acute and chronic LBP, although the effects tend to be modest and short-lived without continued therapy. Reviews have found similar effects for a wide variety of different types of acupuncture (e.g., electroacupuncture, moxibustion, auricular, cupping), with true acupuncture being slightly more effective than sham acupuncture (e.g., needles placed outside standard acupoints or applying pressure that fails to penetrate the skin). In turn, sham acupuncture is more effective than the absence of treatment, although this is likely due to placebo effects. There is no evidence to support one form of integrative treatment compared to others.

INTERVENTIONAL PAIN PROCEDURES

Most cases of LBP cannot be attributed to one anatomic source (nonspecific LBP). However, a thorough history and physical examination and the appropriate use of imaging can help identify potential targets for interventional procedures, which may provide analgesia and improve physical function when surgical indications have not been met or if contraindications for surgery exist. Most interventional pain procedures are performed fluoroscopically or with ultrasound in some cases.

LBP Without Radicular Symptoms (Axial LBP) Axial LBP most commonly involves the facet joints (e.g., zygapophyseal joint, or Z-joint, referring to the paired posterolateral articulations between the inferior articular process of a vertebra with the superior articular process of the subjacent vertebra), SI joints, intervertebral disks, vertebrae, or the paraspinal muscles and ligaments. It is important to recognize that patients can have pain from more than one of these structures simultaneously. However, estimating the prevalence of concomitant sources of spine pain remains challenging; among several structural abnormalities that may be present, only a few might be contributing to a patient's overall pain symptoms. Patients with similar radiologic findings frequently experience different severities and locations of pain from one another, likely as a consequence of each individual's confluence of biopsychosocial factors.

LUMBAR MEDIAL BRANCH NERVE BLOCKS AND ABLATION

Facet joint pain comprises 10–15% of cases of axial LBP, with prevalence increasing with age. The diagnosis of facetogenic pain can only be established through diagnostic medial branch blocks, which entail the administration of local anesthetic onto the medial branch nerves that innervate the facet joints thought to be contributing to the patient's axial pain. If the patient experiences significant improvement in pain and physical function, the diagnosis of facetogenic pain is confirmed and the same nerves can be ablated via radiofrequency ablation (RFA).

SI JOINT INTERVENTIONS

The SI joints are confirmed as the primary etiology in 20–35% of suspected cases of lower axial LBP. SI pain manifests predominantly inferior to the L5 vertebral level and is more likely to be unilateral than facetogenic pain or discogenic pain. The source of SI pain may be intraarticular, which is more likely to occur bilaterally and in older individuals, or extraarticular, which is frequently unilateral, more common in younger individuals (especially after trauma), and can be associated with unremarkable imaging. As noted earlier, the likelihood of SI pain is greatly increased when there are three or more positive provocative physical exam maneuvers (Table 18-2). In addition to being therapeutic, a low-volume intraarticular SI joint injection is the reference standard for diagnosis. Although the prevalence of intraarticular and extraarticular pathology is similar, the treatments are different. RFA of the sacral lateral branch nerves that innervate the extraarticular SI joint ligaments can be considered if intraarticular SI joint injections provide only transient relief, whereas minimally invasive fusion techniques may be indicated in refractory cases of intraarticular pathology or joint malalignment.

INTRADISCAL INJECTIONS AND THERMAL-BASED THERAPIES

Discogenic pain is the main pain generator in 26–42% of individuals with chronic, axial LBP. The diagnosis is suggested by certain physical exam findings, such as increased pain with forward lumbar flexion, sitting, or Valsalva maneuver. Provocative discography, which entails the administration of contrast into the nucleus pulposus to increase intradiscal pressure and reproduce the patient's symptoms, is purported to identify the specific disks contributing to the patient's pain. However, discography is characterized by high false-positive rates in certain populations (e.g., those with somatization and other psychiatric conditions; patients who have undergone prior spine surgery or have multifocal pain symptoms), and based on both animal and clinical studies, the procedure is associated with concerns regarding subsequent accelerated disk degeneration or injury. If the relevant disk(s) have not fragmented and there is no extrusion of intradiscal contents, intradiscal electrothermal therapy (IDET) or biacuplasty, which entails ablating the nervous tissue in the disk, might be considered, but the evidence for intermediate-term benefit is mixed. Intradiscal administration of corticosteroids or ESIs has also demonstrated mixed efficacy for short-term benefit. The intradiscal administration of bone marrow concentrate is a topic of emerging study, but concerns regarding hastened disk degeneration secondary to disk penetration with a large-bore needle and theoretical risks of tumor formation led to a 2019 U.S. Food and Drug Administration warning about stem cell therapies.

VERTEBRAL AUGMENTATION AND BASIVERTEBRAL NERVE ABLATION

Vertebrogenic pain can result from vertebral compression fractures (most commonly due to osteoporosis) or vertebral endplate inflammation. The vertebral endplates are anatomically discrete structures composed of an epiphyseal bone ring surrounding a cartilaginous interior that form the interface between vertebrae and adjacent disks. Due to the transition from the rigid, rib-bearing thoracic spine to the more flexible lumbar spine, most fractures occur at the thoracolumbar junction (T11–L2), with the lower lumbar region being the second most common location. Most vertebral compression fractures are associated with mild to moderate pain that improves within 6–8 weeks of conservative therapy (e.g., physical therapy, oral analgesics), but up to 40% of cases might result in chronic pain. Infrequently (<10%), posterior lumbar compression fractures may be associated with nerve root impingement or spinal cord injury, in which case surgical consultation is warranted.

Although evidence is conflicting, vertebral augmentation via the percutaneous administration of cement into the fracture (vertebroplasty or kyphoplasty) can be considered for patients with severe pain or disability due to an acute (<6 weeks) compression fracture. This may work not only by stabilizing the fracture but also possibly by denervating nociceptive fibers. Posterior compression fractures may also cause facetogenic pain as the superior and inferior articular processes collapse on themselves.

Vertebral endplate inflammation due to trauma or degenerative changes may be present (though not necessarily be the primary etiology of pain) in up to 40% of patients with chronic axial lumbar pain. Vertebrogenic pain from endplate fractures or degeneration presents similarly to discogenic pain (e.g., worsened pain with bending forward, sitting, or activity), also occurs most frequently in the lower lumbar area, and can co-occur with discogenic pain. For patients in whom other structures of the spine (e.g., facet joints) are less likely to be the predominant source of their symptoms, and in whom Modic type 1 (hypointense signal on T1-weighted and hyperintense signal on T2-weighted MRI, indicating marrow edema) or type 2 changes (hyperintense signal on T1-weighted and isointense on T2-weighted MRI, signifying conversion of red hemopoietic bone marrow into yellow fatty marrow due to ischemia) are demonstrated on MRI, RFA of the basivertebral nerves that innervate the vertebral endplates can be considered.

TRIGGER POINT INJECTIONS OF MYOFASCIAL TAUT BANDS

Pain from paraspinal ligaments and muscles is frequently due to spasmodic activity that results in myofascial taut bands. Trigger point injections can relieve spasmodic activity by improving local blood flow via vasodilation and facilitate the removal of inflammatory mediators and cytokines. Evidence has not demonstrated the superiority of any specific injectate (e.g., saline, higher concentrations of local anesthetic with or without steroids, botulinum toxin), suggesting that the mechanism of action of trigger point injections is independent of the medication(s) administered.

LBP WITH RADICULAR SYMPTOMS (RADICULAR LBP)

Epidural Steroid Injections Radicular LBP is usually from spinal stenosis (e.g., central canal, lateral recess, or neuroforaminal) or disk pathology (e.g., bulge, herniation, extrusion) causing compression or irritation of one or more spinal nerve roots. Although most patients with acute radicular pain due to a herniated disk will clinically improve within 3 months with conservative management (e.g., over-the-counter analgesics and physical therapy), the natural history of pain from lumbar stenosis is more guarded. Whereas over two-thirds of herniated disks will retract within 2 years, the causes of spinal stenosis (e.g., spondylolisthesis, facet hypertrophy, ligamentum flavum hypertrophy) often progress with age. Although some studies have reported that spinal stenosis is less responsive to ESI than herniated disks, most studies have failed to show any differences in response rates, with several showing a poorer response rate for noncompressive (e.g., degenerative disks causing chemical irritation of nerve roots) pathologies. There is also some evidence for a weak but potentially meaningful surgery-sparing effect from ESIs for radicular pain.

Surgery Lumbar surgeries can be broadly categorized as entailing fusion of adjacent vertebral bodies, decompression of spinal nerve roots (i.e., diskectomy or laminectomy), or a combination. Lumbar interbody fusion is commonly performed for spinal instability (i.e., symptomatic or severe spondylolisthesis) or severe axial pain refractory to nonsurgical management and associated with severe functional disability. The precise indications for lumbar fusion remain controversial. Data are mixed regarding the effectiveness for axial back pain and disability based on randomized trials and observational cohorts, with the strongest evidence being for spondylolisthesis, which is often associated with instability. In individuals with single-level and sometimes two-level discogenic pain without posterior element involvement, the less invasive disk arthroplasty procedure has been shown to be at least as effective as circumferential or anterior fusion for pain and function with better preservation of motion; however, the quality of these studies has generally been poor, with most being industry funded. Adjacent segment disease (e.g., subsequent accelerated degeneration of adjacent disks and facet joints) can occur following both procedures but is more likely after a fusion, while the risk of postsurgical instability is greater following disk arthroplasty.

Lumbar nerve root decompression is indicated for severe or progressive neurologic deficits (i.e., loss of motor or sensory function or reflexes, indicating progressive radiculopathy) or for severe radicular pain refractory to interventional treatments (e.g., ESIs), medications, and physical therapy. The role of lumbar interbody fusion in addition to decompression is controversial, but studies have generally shown higher complication rates without greater benefit for nerve root compression without instability. Studies are mixed regarding whether surgical decompression is superior to conservative management for spinal stenosis, with meta-analyses generally finding a small benefit that diminishes after 2 years. For herniated disk, randomized trials have generally shown greater short-term reduction in pain and disability, but small and questionably meaningful benefits after 1 year. Minimally invasive surgical techniques are available for both lumbar stenosis (e.g., interspinal spacers, minimally invasive lumbar decompression) and herniated disks (e.g., chemonucleolysis, endoscopic discectomy), with low-quality data supporting at least short-term benefit for these procedures in well-selected patients.

FURTHER READING

Brinjikji W et al: Systematic literature review of imaging features of spinal degeneration in asymptomatic populations. AJNR Am J Neuroradiol 36:811, 2015.

Cohen SP et al: Effect of MRI on treatment results or decision making in patients with lumbosacral radiculopathy referred for epidural steroid injections: A multicenter, randomized controlled trial. Arch Intern Med 172:134, 2012.

Cohen SP et al: Chronic pain: An update on burden, best practices, and new advances. Lancet 397:2082, 2021.

Cohen SP et al: Multicenter study evaluating factors associated with treatment outcome for low back pain injections. Reg Anesth Pain Med 47:89, 2022.

Cook CJ et al: Systematic review of diagnostic accuracy of patient history, clinical findings, and physical tests in the diagnosis of lumbar spinal stenosis. Eur Spine J 29:93, 2020.

Itz CJ et al: Clinical course of non-specific low back pain: A systematic review of prospective cohort studies set in primary care. Eur J Pain 17:5, 2013.

Kasch R et al: Association of lumbar MRI findings with current and future back pain in a population-based cohort study. Spine (Phila Pa 1976) 47:201, 2022.

Katz JN et al: Diagnosis and management of lumbar spinal stenosis: A review. JAMA 327:1688, 2022.

Knezevic NN et al: Low back pain. Lancet 398:78, 2021.

Lo J et al: A systematic review of the incidence, prevalence, costs, and activity and work limitations of amputation, osteoarthritis, rheumatoid arthritis, back pain, multiple sclerosis, spinal cord injury, stroke, and traumatic brain injury in the United States: A 2019 Update. Arch Phys Med Rehabil 102:115, 2021.

19 Neck Pain

Fan Jiang, Soichiro Takamiya, Michael G. Fehlings

Neck pain is a highly prevalent global problem. In the United States, it is the fourth leading cause of disability and can affect people of all ages, genders, and professions. The lifetime prevalence is nearly 50%, women appear to be at a higher risk than men, and the incidence increases with age, peaking in late middle life. It is associated with sport- and work-related injuries, with low job satisfaction and poor work support having been found to play a role in work-related neck pain. Other risk factors include genetics, headaches, sleep disorders, smoking, obesity, sedentary lifestyle, secondary gain, history of neck pain, trauma, back pain, and poor overall physical or mental health.

Neck pain usually arises from diseases of the cervical spine and soft tissues of the neck, is typically precipitated by movement, and may be accompanied by focal tenderness and limitation of motion. Pain arising from the brachial plexus, shoulder, or peripheral nerves can sometimes be confused with cervical spine disease, but the history and examination usually identify a more distal origin for the pain (discussed later). The underlying causes of neck pain are diverse, and patients often present with nonspecific and vague symptoms; occipital headache is a common complaint. It is especially important to develop a systematic approach to evaluation and management, as well as to recognize dangerous etiologies including infection, malignancy, or spinal cord involvement from any cause.

When a patient presents with neck pain, the clinician should identify if the pain has a neuropathic component, indicating that the pain originates from identifiable nerves producing cervical radiculopathy. Causes of neuropathic neck pain include compression or irritation of cervical spinal nerves secondary to disk herniation, ligamentous hypertrophy, or facet overgrowth, among other etiologies, discussed below.

APPROACH TO THE PATIENT

Neck Pain

A complete clinical history is essential to gain an appreciation of the patient's chief complaints and to look for red flags. A patient's occupation, general health, and past medical history are important to identify likely pain generators and risk factors. A history of cancer, fever, and weight loss should prompt investigations to rule out malignancy and infectious etiologies. Characteristics of the pain (dull, sharp, electric, stabbing, spasms) and aggravating and alleviating factors should be determined. Any radiating pain into the occiput, arms, or hands can raise concern for radiculopathy. Patients should be specifically asked if they have experienced numbness or tingling in extremities; clumsiness in hands; change in handwriting; difficulty with buttons; unsteady gait; saddle paresthesia; and bladder or bowel incontinence.

Physical examination of the patient begins with observation. The head, neck, and shoulder should be examined for any deformity, mass, skin changes, or signs of trauma. Determining the location of the pain is imperative, as well as the quantity of pain and whether it is interfering with daily function and activities. The range of motion of the neck and shoulder should be assessed, and any restriction or guarding against movement noted. Fifty percent of cervical motion (flexion, extension, and rotation) originates from the atlantoaxial joint (C1-C2) and 50% from the subaxial (C3-C7); in general, any motion restriction needs to be further investigated to rule out underlying pathologies.

Radicular pain in the arms typically follows dermatomal distributions (**Table 19-1**). Obtaining a precise description of the pain pattern can assist in determining if the pain is neuropathic and where the problem might originate. If a patient describes radiculopathy-like or other neuropathic pain, a complete neurologic examination is required, including assessment of the cranial nerves, motor strength, sensation, coordination, reflexes, and gait (**Chaps. 433, V6, and V7**).

A few bedside tests are of particular value when assessing for radiculopathy or possible spinal cord involvement in patients presenting with neck pain. Spurling's maneuver is a test for radiculopathy originating from the cervical spine. It is performed by passively rotating and flexing a patient's neck laterally and performing axial compression to the top of the head. A test is considered positive if it triggers or worsens symptoms of upper extremity radiculopathy. Lhermitte's sign is elicited by gently flexing a patient's neck. It is considered positive and signifies underlying cervical spinal pathology if the maneuver reproduces electric pain down the spine or extremities. Also useful is Hoffmann's test, which should be performed whenever there is any suspicion of cervical spinal cord pathology. It is a sign of hyperreflexia

TABLE 19-1 Cervical Radiculopathy: Neurologic Features

CERVICAL NERVE ROOT	EXAMINATION FINDINGS			PAIN DISTRIBUTION
	REFLEX	SENSORY	MOTOR	
C5	Biceps	Lateral deltoid	Rhomboids[a] (elbow extends backward with hand on hip) Infraspinatus[a] (arm rotates externally with elbow flexed at the side) Deltoid[a] (arm raised laterally 30°–45° from the side)	Lateral arm, medial scapula
C6	Biceps	Palmar thumb/index finger Dorsal hand/lateral forearm	Biceps[a] (arm flexed at the elbow in supination) Pronator teres (forearm pronated)	Lateral forearm, thumb/index fingers
C7	Triceps	Middle finger Dorsal forearm	Triceps[a] (forearm extension, flexed at elbow) Wrist/finger extensors[a]	Posterior arm, dorsal forearm, dorsal hand
C8	Finger flexors	Palmar surface of little finger Medial hand and forearm	Abductor pollicis brevis (abduction of thumb) First dorsal interosseous (abduction of index finger) Abductor digiti minimi (abduction of little finger)	Fourth and fifth fingers, medial hand and forearm
T1	Finger flexors	Axilla, medial arm, anteromedial forearm	Abductor pollicis brevis (abduction of thumb) First dorsal interosseous (abduction of index finger) Abductor digiti minimi (abduction of little finger)	Medial arm, axilla

[a] These muscles receive the majority of innervation from this root.

Source: From JW Engstrom: Back and neck pain, in *Harrison's Principles of Internal Medicine*, 21st ed. New York, NY: McGraw Hill; 2022, Table 17-4, p. 127.

whereby a patient involuntarily flexes and adducts the thumb and index finger while the examiner snaps the distal phalanx of the middle finger. Similarly, the presence of more than three beats of ankle clonus when the ankle is briskly dorsiflexed and held under pressure by the examiner is an upper motor neuron sign signifying underlying spinal cord pathology. Patients should also be observed for tandem or heel-to-toe gait, the integrity of which is dependent on proprioception, as well as coordination and strength. Patients are asked to walk in a straight line with one foot in front of the other. If there are any signs of imbalance (i.e., the patient is not able to complete this task), the possibility of spinal cord compression should be considered.

Imaging and additional tests are often necessary to finalize the diagnosis and should be guided by the patient's history and physical examination findings. Cervical spine x-rays are simple diagnostic tools that are readily accessible and can provide an excellent initial assessment for spinal column pathologies. Performed in the anterior-posterior, lateral, and flexion-extension views, these radiographs are useful tools to screen for fractures, instability, and osteoarthritis, and can often prompt further investigations. Computed tomography (CT) scan is usually indicated if there are any concerns for fractures in the spinal column or if there have been previous surgical instrumentations to assess for any hardware complications. However, due to the poor visualization of soft tissues, CT is not very helpful when evaluating the spinal cord or ruling out intervertebral disk pathologies. Magnetic resonance imaging (MRI) is the gold standard for assessing these soft tissue structures. MRI is indicated for patients presenting with radiculopathy symptoms or signs of myelopathy or other neurologic deficits.

MRI is also the imaging modality of choice when ruling out malignancy or infectious causes of neck pain. However, given the high rate of abnormal findings on MRI in asymptomatic individuals, and degenerative changes in particular, it should be performed with caution and only in subjects with strong indications based on history, physical exam, and other screening radiographs. Finally, for individuals in whom MRI is contraindicated **(Chap. 434)**, CT myelography, whereby contrast is injected into the spinal canal prior to the CT scan, can be used to assess the spinal cord and surrounding structures.

Electromyography (EMG) and nerve conduction studies (NCS) are diagnostic studies often utilized in evaluating a patient with neck pain and associated neuropathic symptoms. While these electrodiagnostic studies are typically unnecessary when the diagnosis is clear, they can be helpful when a patient's symptoms do not correlate with the MRI findings. By measuring the electrical response to nerve stimulation in the muscles and the speed by which an electrical impulse travels, the combination of EMG and NCS can distinguish radiculopathy arising from the spine from peripheral neuropathy and brachial plexopathy **(Chap. 457)**. Referral to a neurologist is recommended in the setting of high clinical suspicion and negative electrodiagnostic findings.

Routine laboratory tests have limited value in assessment of most patients with neck pain. However, if there is clinical suspicion of an underlying infection, malignancy, inflammatory arthritis, or neuromuscular disorder, appropriate blood panels should be obtained.

While the differential diagnosis of neck pain is very broad, the most important goal of the clinical evaluation is recognizing nonmuscular causes of pain and identifying clinically dangerous underlying pathology. **Table 19-2** summarizes clinical findings that should prompt a clinician to conduct further investigations to rule out dangerous underlying etiologies. The following section will provide an overview and discussion of various causes of clinically important neck pain.

TABLE 19-2 Clinical Features Suggesting a Need for Further Investigation in Patients With Neck Pain

HISTORY	PHYSICAL EXAM
History of congenital disorders	Fever
History of cancer	Neck stiffness
History of substance abuse	Severe neck pain and tenderness
History of inflammatory arthritis	Torticollis
Recent trauma to the head and neck	Motor weakness in upper or lower extremities
Signs or symptoms of infection	Sensory changes in upper or lower extremities
Family history of spinal conditions	Upper motor neuron signs: Hoffman, clonus, Babinski
Unexplained weight loss	Hyperreflexia
Progressive neurologic deficits	Unsteady gait
Saddle anesthesia	Difficulty with tandem or heel-to-toe walk
Bladder or bowel incontinence	

CAUSES OF NECK PAIN

DEGENERATIVE

Cervical radiculopathy occurs when a cervical spinal nerve root is compressed. Herniated cervical disks cause about 25% of cervical radiculopathy, whereas protruding cervical disks and the narrowing intervertebral foramen due to osteophytic spurs, hypertrophic facet joints, and uncovertebral joints can be solely or jointly responsible for ~70% of cervical radiculopathy. Although the mechanisms underlying radiculopathy are not fully understood, ischemia and hypoxia of the affected nerve root appear to contribute to injury and subsequent symptoms. The nerve root most frequently affected is C7, followed by C6. Besides neck pain, cervical radiculopathy is commonly accompanied by symptoms such as pain, sensory disturbance, and motor weakness in the shoulder, interscapular, or upper limb, depending on the affected nerve root; however, neck pain can be the only symptom, especially when arising from nerve roots at C4 or higher.

Degenerative cervical disk disease is a relatively common pathology. Degenerative disks contain inflammatory mediators, including proinflammatory cytokines, that can lead to stimulation of inflammation-responsive sensory nerve fibers, which innervate intervertebral disks, resulting in nociceptive or "discogenic" pain. When cervical disks are herniated, they can cause radiculopathy and myelopathy **(Fig. 19-1 *A, B*)**. Neck pain in patients with cervical disk herniation is often caused by a combination of the mechanisms mentioned above—ischemia of the compressed root and/or spinal cord and inflammation arising from degenerative disks.

Degenerative cervical myelopathy, which is the comprehensive term used to describe the various degenerative diseases causing symptomatic cervical spine narrowing such as cervical spondylosis, ossification of the posterior longitudinal ligament, degenerative disk disease, and ossification of the ligamentum flavum, is also a common cause of neck pain. In degenerative cervical myelopathy, the conditions listed above can compress the cervical spinal cord and lead to spinal cord injury **(Fig. 19-1 *C, D*)**. Clinicians need to understand this condition well, given that the initial presenting symptoms can often be vague and misleading. It has been reported that the delay in obtaining a diagnosis of degenerative cervical myelopathy can be up to 2 years after the onset of symptoms. Given the shifting demographics of an aging society, this will become more and more relevant at the primary care level, where these patients are typically first encountered.

The early signs of cervical myelopathy can be extremely mild, such as paresthesia, minor loss of hand dexterity, or feelings of "clumsiness." These can easily be overlooked during the initial clinical encounter, especially when the chief presenting complaint is neck pain. Severe dysfunction, including gait disturbance, quadriparesis, and bowel or bladder incontinence, often presents at a later stage. However, prompt diagnosis and timely surgical intervention at the initial phase of the disease often result in better clinical outcomes in these individuals, hence the importance of early recognition and prompt referral for definitive treatment. Fortunately, the deterioration in degenerative cervical myelopathy is typically slow and occurs in a stepwise progressive fashion. **Spinal cord disorders are discussed in Chap. 454.**

NEOPLASTIC

Spinal tumors can cause neck pain when they occur in the cervical spine. Metastases are the most common extradural spinal tumors. Although the cervical spine is involved in <20% of patients with spinal

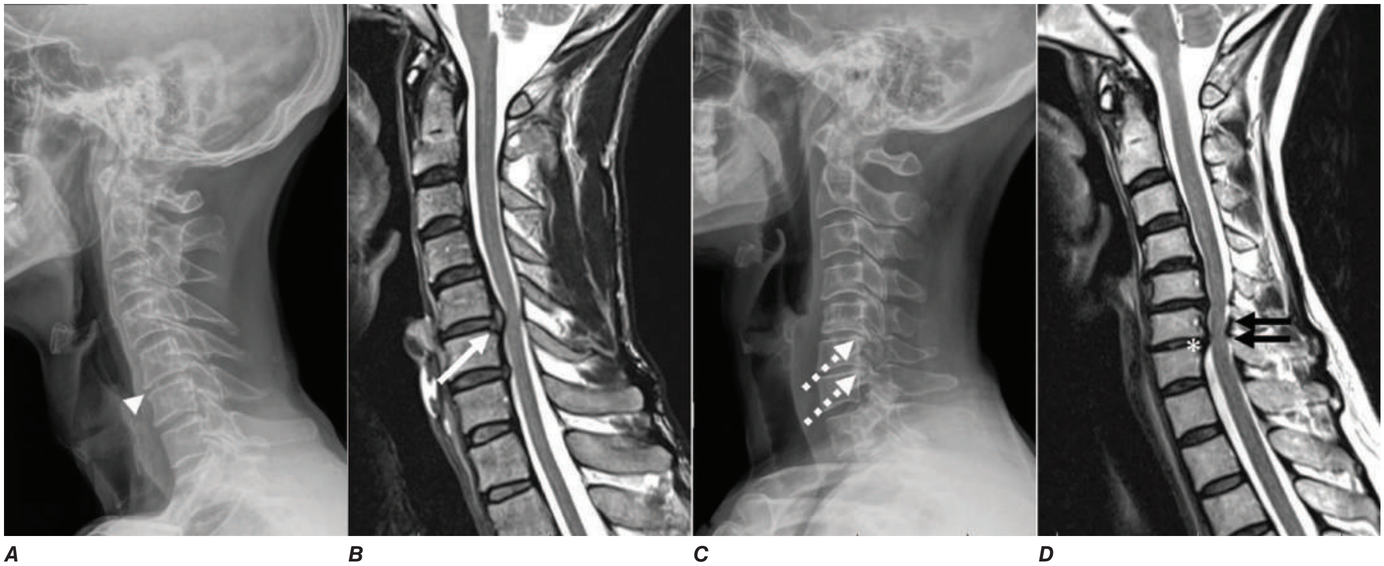

FIGURE 19-1 Cases of degenerative cervical disk disease and degenerative cervical myelopathy. Cervical x-ray shows a disk height reduction at the C5/C6 level (arrowhead) (***A***), and cervical magnetic resonance imaging (MRI) shows a herniated disk compressing the spinal cord (arrow) (***B***) in a case of degenerative cervical disk disease. Cervical x-ray shows osteophytic spurs at C5/C6 and C6/C7 levels (dotted arrow) (***C***), and cervical MRI shows a protruded disk (asterisk) and thickened ligamentum flavum (double arrows) (***D***) in a case of degenerative cervical myelopathy.

metastatic tumors, neck pain is the most common symptom in such patients. This pain tends to be unrelieved by rest and worse at night. Metastases sometimes cause pathological fractures, which can also be responsible for neck pain **(Fig. 19-2 *A, B*)**.

Primary spinal cord tumors including schwannomas, meningiomas, and intramedullary tumors can induce neuropathic neck pain, whereas primary spinal bone tumors, such as hemangiomas or chordomas, typically cause nociceptive neck pain. Systemic tumors such as multiple myeloma or lymphoma may also be responsible for neck pain when the cervical spine is involved.

INFECTIOUS

Cervical vertebral osteomyelitis and diskitis, which is less common compared with thoracic or lumbar regions, can cause constant neck pain **(Chap. 136)**. Although fever is present in some patients, it is absent in many others, which can lead to a delay in diagnosis. Whenever patients present with neck pain and unidentified fever, physicians should always consider these etiologies, especially in patients with a history of intravenous drug abuse, immunocompromised status, diabetes mellitus, or other disorders that predispose to infection.

Cervical epidural abscess **(Chap. 145)** can cause severe neck pain and progressive neurologic deterioration due to cord compression **(Fig. 19-2 *C, D*)**, often requiring urgent surgical intervention.

Meningitis **(Chap. 143)** is another important cause of neck pain, usually accompanied by fever and headache.

VASCULAR

Neck pain associated with vascular pathologies often occurs suddenly. Vertebral and carotid artery dissections can cause sudden neck pain as well as headache. Acute coronary syndromes may present as a referred neck pain. Cervical arteriovenous malformations and other vascular pathologies can cause acute neck pain when they bleed. Cervical epidural hematomas typically present with sudden neck pain and neurologic deficits. They can be misdiagnosed as stroke; however, the neck pain aggravated by palpation of the spinous process can be a helpful distinguishing sign.

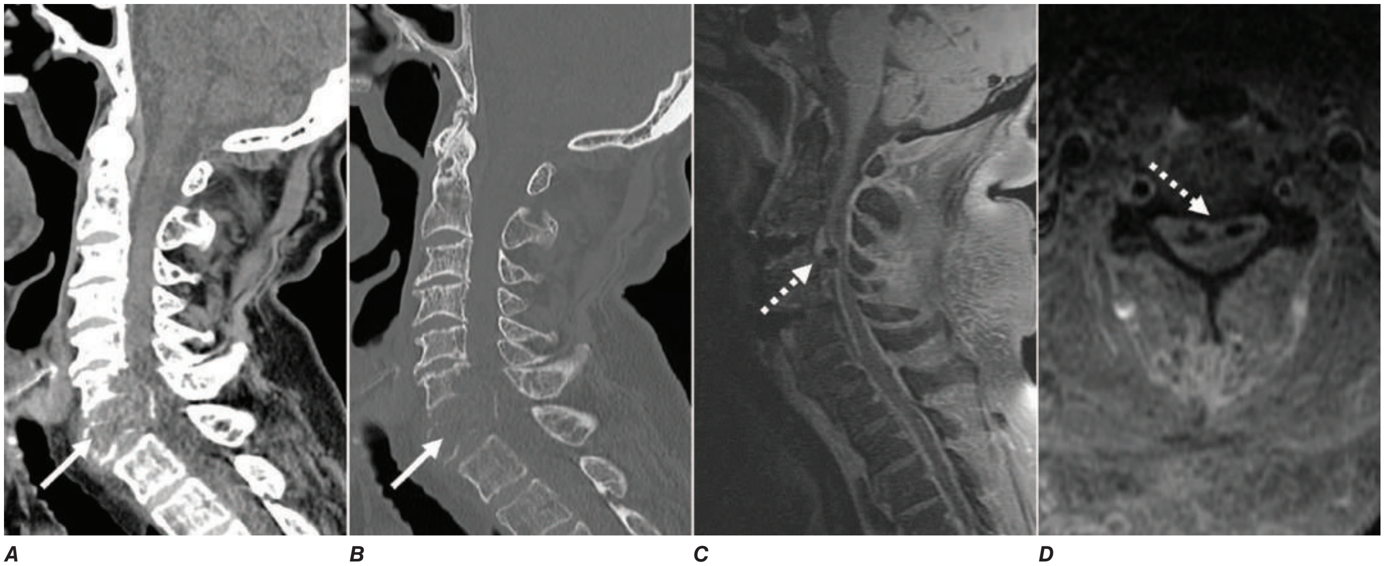

FIGURE 19-2 Cases of cervical metastasis and cervical epidural abscess. Cervical computed tomography shows osteolytic lesions at C6 and C7 that protrude into the spinal canal (arrow) (***A, B***) in a case of cervical metastasis. Sagittal (***C***) and axial (***D***) views of cervical fat-suppressed contrast-enhanced magnetic resonance imaging show enhanced epidural lesions at C2/3–C4/5, which compress the spinal cord drastically (dotted arrows) in this case of cervical epidural abscess.

CONGENITAL

Chiari type 1 malformations **(Chap. 453)** can cause headache and neck pain aggravated by cough. The pathophysiology remains unclear, but dissociation between intracranial and intraspinal pressures or traction on pain-sensitive nerves of the dura mater may be responsible.

AUTOIMMUNE/INFLAMMATORY

Rheumatoid arthritis **(Chap. 370)** commonly affects the cervical spine and produces neck pain most often due to cervical instability. Approximately 50% of patients with rheumatoid arthritis have atlantoaxial subluxation, which may lead to neurologic deterioration. *Polymyalgia rheumatica* **(Chap. 375)** presents with pain and stiffness in the neck, shoulder, and pelvic girdle and is accompanied by other systemic symptoms. *Ankylosing spondylitis* **(Chap. 374)** also causes neck pain, though chronic back pain is more common, and these pains typically worsen with rest or inactivity. Crowned dens syndrome, also known as periodontoid calcium pyrophosphate dihydrate crystal deposition disease, is another cause of severe neck pain and neck stiffness.

ENDOCRINE/METABOLIC

Paget's disease **(Chap. 424)** is a bone disorder characterized by an imbalance in bone modeling and remodeling. Paget's disease in the cervical spine is less common compared with the thoracolumbar regions but can cause neck pain that tends to be worse at rest. *Osteoporosis* **(Chap. 423)**, which is sometimes induced by hyperparathyroidism and glucocorticoid therapy, and *osteomalacia* due to abnormal vitamin D metabolism **(Chap. 421)**, can produce pathological bone fractures. Although these disorders usually affect lumbar or lower thoracic vertebrae, neck pain can occur when the cervical spine is affected.

TRAUMATIC

Traumatic fracture and dislocation of the cervical spine are life-threatening conditions when accompanied by injury to the cervical spinal cord. Nociceptive and neuropathic neck pains can be combined, and immobilization of the neck is essential to reduce neck pain and prevent further cervical cord injury due to spinal instability. *Whiplash injury* is also responsible for neck pain. It is usually caused by motor vehicle collisions, and injury to the facet joints is believed to underlie this pain syndrome. Approximately 50% of patients with whiplash injury will have had persistent neck pain for 1 year.

OTHER CAUSES

Cervical myofascial pain is a very common clinical condition that can involve diffuse areas around the neck and shoulder. Restriction of motions of the cervical spine can cause this nonspecific neck pain. Myofascial trigger points, which are palpable nodules housed inside taut muscle bands, are regarded as a hallmark of myofascial pain.

TREATMENT

Neck Pain

Given the broad differential diagnosis of neck pain, the management and treatment should be tailored toward the underlying condition. Most important is to recognize red flags that indicate the possible presence of myelopathy, malignancy, infection, or a severe spinal column injury; in such cases, urgent assessment is required. The discussion below will focus on the management of the common causes of neck pain seen in general medical practice. In general, the evidence regarding treatment for neck pain is less comprehensive than that for low back pain **(Chap. 18)**, but the approach is similar in many respects.

NECK PAIN WITHOUT RADICULOPATHY OR OTHER NEUROLOGIC FINDINGS

For many patients with acute neck pain, spontaneous improvement is the norm, and the usual goals of therapy are to promote a rapid return to normal function and provide pain relief while healing proceeds. Acute neck pain is often treated with nonsteroidal anti-inflammatory drugs (NSAIDs), acetaminophen, cold packs, or heat, alone or in combination while awaiting recovery. Patients should be specifically educated regarding the favorable natural history of acute neck pain to avoid unrealistic fear and inappropriate requests for imaging and other tests. For patients kept awake by symptoms, cyclobenzaprine (5–10 mg) at night can help relieve muscle spasm and promote drowsiness. A trial of physiotherapy such as supervised exercise with or without mobilization appears to be effective. Exercises often include shoulder rolls and neck stretches. The McKenzie method of physical therapy is one option in widespread use. Some patients obtain modest pain relief using a soft neck collar; there is little risk or cost. Massage can produce temporary pain relief.

For patients with chronic neck pain, supervised exercise programs can provide symptom relief and improve function. Acupuncture provided short-term benefit for some patients when compared to a sham procedure and is an option. Spinal manipulation alone has not been shown to be effective and carries a risk for injury.

Surgery plays a very limited role in patients with acute and chronic neck pain without neurologic findings, spine instability, or fractures. For individuals with pain refractory to conservative management, referral to a pain specialist can be of benefit. The literature shows that neck pain originating from cervical facet joints can be effectively treated with fluoroscopically guided radiofrequency ablation. However, careful patient selection is essential to ensure optimal outcomes, and the procedure is not available in many centers. Referral to a specialist for diagnostic and treatment guidance is highly recommended for any patient in whom such procedures are considered.

NECK PAIN WITH RADICULOPATHY OR OTHER NEUROLOGIC FINDINGS

The natural history of acute neck pain with radiculopathy due to disk disease is also favorable, and many patients will improve without specific therapy. Although there are no randomized trials of NSAIDs for neck pain, a course of NSAIDs, acetaminophen, or both, with or without muscle relaxants, and avoidance of activities that trigger symptoms are reasonable as initial therapy. Gentle supervised exercise and avoidance of inactivity are reasonable as well. A short course of high-dose oral glucocorticoids with a rapid taper or epidural steroids administered under imaging guidance can be effective for acute or subacute disk-related cervical radicular pain but have not been subjected to rigorous trials. The risk of injection-related complications is higher in the neck than the low back; vertebral artery dissection, dural puncture, spinal cord injury, and embolism in the vertebral arteries have all been reported. Opioid analgesics can be used in the emergency department and for short courses as an outpatient. Soft cervical collars can be modestly helpful by limiting spontaneous and reflex neck movements that exacerbate pain; hard collars are in general poorly tolerated.

If cervical radiculopathy is due to bony compression from cervical spondylosis with foraminal narrowing, periodic follow-up to assess for progression is indicated and consideration of surgical decompression is reasonable. Surgical treatment can produce rapid pain relief, although it is unclear if long-term functional outcomes are improved over nonsurgical therapy. Indications for cervical disk surgery include a progressive motor deficit due to nerve root compression, functionally limiting pain that fails to respond to conservative management, or spinal cord compression. In other circumstances, clinical improvement over time regardless of therapeutic intervention is common.

Surgical treatments include anterior cervical diskectomy/corpectomy and fusion, posterior cervical laminectomy, laminoplasty or laminectomy with fusion. The risk of subsequent radiculopathy or myelopathy at cervical segments adjacent to a fusion is ~3% per year and 26% per decade. Although this risk is sometimes portrayed as a late complication of surgery, it may also reflect the natural history of degenerative cervical disk disease.

PAIN ARISING IN THE SHOULDER REGION

THORACIC OUTLET SYNDROMES

The thoracic outlet contains the first rib, the subclavian artery and vein, the brachial plexus, the clavicle, and the lung apex. Injury to these structures may result in postural or movement-induced pain around the shoulder and supraclavicular region, classified as follows.

True neurogenic thoracic outlet syndrome (TOS) is an uncommon disorder resulting from compression of the lower trunk of the brachial plexus or ventral rami of the C8 or T1 nerve roots, caused most often by an anomalous band of cartilaginous tissue connecting an elongate transverse process at C7 with the first rib. Pain is mild or may be absent. Signs include weakness and wasting of intrinsic muscles of the hand and diminished sensation on the palmar aspect of the fifth digit. An anteroposterior cervical spine x-ray will show an elongate C7 transverse process (an anatomic marker for the anomalous cartilaginous band), and EMG and NCSs confirm the diagnosis. Treatment consists of surgical resection of the anomalous band. The weakness and wasting of intrinsic hand muscles typically do not improve, but surgery halts the insidious progression of weakness.

Arterial TOS results from compression of the subclavian artery by a cervical rib, resulting in poststenotic dilatation of the artery and in some cases secondary thrombus formation. Blood pressure is reduced in the affected limb, and signs of emboli may be present in the hand. Neurologic signs are absent. Ultrasound can confirm the diagnosis noninvasively. Treatment is with thrombolysis or anticoagulation (with or without embolectomy) and surgical excision of the cervical rib compressing the subclavian artery.

Venous TOS is due to subclavian vein thrombosis resulting in swelling of the arm and pain. The vein may be compressed by a cervical rib or anomalous scalene muscle. Venography is the diagnostic test of choice.

Disputed TOS accounts for 95% of patients diagnosed with TOS; chronic arm and shoulder pain are prominent and of unclear cause. The lack of sensitive and specific findings on physical examination or specific markers for this condition results in diagnostic uncertainty. The role of surgery in disputed TOS is controversial. Multidisciplinary pain management is a conservative approach, although treatment is often unsuccessful.

BRACHIAL PLEXUS AND NERVES

Pain from injury to the brachial plexus or peripheral nerves of the arm can occasionally mimic referred pain of cervical spine origin, including cervical radiculopathy, but the pain typically begins distal to the posterior neck region in the shoulder girdle or upper arm. Neoplastic infiltration of the lower trunk of the brachial plexus may produce shoulder or supraclavicular pain radiating down the arm, numbness of the fourth and fifth fingers or medial forearm, and weakness of intrinsic hand muscles innervated by the lower trunk and medial cord of the brachial plexus. Delayed radiation injury may produce weakness in the upper arm or numbness of the lateral forearm or arm due to involvement of the upper trunk and lateral cord of the plexus. Pain is less common and less severe than with neoplastic infiltration. A Pancoast tumor of the lung **(Chap. 83)** is another cause of injury to the brachial plexus and should be considered especially when a concurrent Horner's syndrome is present. *Acute brachial neuritis* is often confused with radiculopathy; the acute onset of severe shoulder or scapular pain is followed typically over days by weakness of the proximal arm and shoulder girdle muscles innervated by the upper brachial plexus. The onset may be preceded by an infection, vaccination, or minor surgical procedure. The long thoracic nerve may be affected, resulting in a winged scapula. Brachial neuritis may also present as an isolated paralysis of the diaphragm with or without involvement of other nerves of the upper limb. Recovery may take up to several years, and full functional recovery can be expected in the majority of patients.

Occasional cases of carpal tunnel syndrome produce pain and paresthesias extending into the forearm, arm, and shoulder resembling a C5 or C6 root lesion. Lesions of the radial or ulnar nerve can also mimic radiculopathy at C7 or C8, respectively. EMG and NCSs can accurately localize lesions to the nerve roots, brachial plexus, or peripheral nerves.

For further discussion of peripheral nerve disorders, see Chap. 457.

SHOULDER

Pain arising from the shoulder can also on occasion mimic pain from the spine. If symptoms and signs of radiculopathy are absent, then the differential diagnosis includes mechanical shoulder pain (bicipital tendonitis, frozen shoulder, bursitis, rotator cuff tear, dislocation, adhesive capsulitis, or rotator cuff impingement under the acromion) and referred pain (subdiaphragmatic irritation, angina, or Pancoast tumor). Mechanical pain is often worse at night, associated with local shoulder tenderness, and aggravated by passive abduction, internal rotation, or extension of the arm. Demonstrating normal passive full range of motion of the arm at the shoulder without worsening of the pain can help exclude mechanical shoulder pathology as a cause of neck region pain. Pain from shoulder disease may radiate into the arm or hand, but focal neurologic signs (sensory, motor, or reflex changes) are absent.

ACKNOWLEDGMENT

The authors acknowledge the contributions of John W. Engstrom to earlier editions of this chapter.

FURTHER READING

Cohen SP, Hooten WM: Advances in the diagnosis and management of neck pain. BMJ 358:j3221, 2017.

Popescu A, Lee H: Neck pain and lower back pain. Med Clin North Am 104:279, 2020.

Rupp R et al: International standards for neurological classification of spinal cord injury: Revised 2019. Top Spinal Cord Inj Rehabil 27:1, 2021.

Shin DW et al: Global, regional, and national neck pain burden in the general population, 1990-2019: An analysis of the global burden of disease study 2019. Front Neurol 13:955367, 2022.

Vazirizadeh-Mahabadi M, Yarahmadi M: Canadian C-spine Rule versus NEXUS in screening of clinically important traumatic cervical spine injuries: A systematic review and meta-analysis. Arch Acad Emerg Med 11:e5, 2023.

Section 2 Alterations in Body Temperature

20

Fever

Charles A. Dinarello, Neeraj K. Surana

Body temperature is controlled by the hypothalamus. Neurons in both the preoptic anterior hypothalamus and the posterior hypothalamus receive two kinds of signals: one from peripheral nerves that transmit information from warmth/cold receptors in the skin and the other from the temperature of the blood bathing the region. These two types of signals are integrated by the thermoregulatory center of the hypothalamus to maintain normal temperature. In a neutral temperature environment, the human metabolic rate produces more heat than is necessary to maintain the core body temperature in the range of 36.5–37.5°C (97.7–99.5°F).

NORMAL BODY TEMPERATURES

A normal body temperature is ordinarily maintained despite environmental variations because the hypothalamic thermoregulatory center balances the excess heat production derived from metabolic activity in muscle and the liver with heat dissipation from the skin and lungs. According to a study of >35,000 individuals ≥18 years of age seen in routine medical visits, the mean oral temperature is 36.6°C (95% confidence interval, 35.7–37.3°C). In light of this study, *a temperature of >37.7°C (>99.9°F), which represents the 99th percentile for healthy individuals, defines a fever.* Importantly, higher ambient temperatures are linked to higher baseline body temperatures. Additionally, body temperatures have diurnal and seasonal variation, with low levels at 8 A.M. and during summer and higher levels at 4 P.M. and during winter. Baseline temperatures are also affected by age (lower by 0.02°C for every 10-year increase in age), demographics (African-American women have temperatures 0.052°C higher than white men), and comorbid conditions (cancer is associated with 0.02°C higher temperatures; hypothyroidism is linked to temperatures lower by 0.01°C). After controlling for age, sex, race, vital signs, and comorbidities, an increase in baseline temperature of 0.15°C (or 1 standard deviation) intriguingly translates into a 0.52% absolute increase in 1-year mortality.

Rectal temperatures are generally 0.4°C (0.7°F) higher than oral readings. The lower oral readings are probably attributable to mouth breathing, which is a factor in patients with respiratory infections and rapid breathing. Lower-esophageal temperatures closely reflect core temperature. Tympanic membrane thermometers measure radiant heat from the tympanic membrane and nearby ear canal and display that absolute value (*unadjusted mode*) or a value automatically calculated from the absolute reading on the basis of nomograms relating the radiant temperature measured to actual core temperatures obtained in clinical studies (*adjusted mode*). These measurements, although convenient, may be more variable than directly determined oral or rectal values. Studies in adults show that readings are lower with unadjusted-mode than with adjusted-mode tympanic membrane thermometers and that unadjusted-mode tympanic membrane values are 0.8°C (1.6°F) lower than rectal temperatures.

In women who menstruate, the A.M. temperature is generally lower during the 2 weeks before ovulation; it then rises by ~0.6°C (1°F) with ovulation and stays at that level until menses occur. During the luteal phase, the amplitude of the circadian rhythm remains the same.

FEVER VERSUS HYPERTHERMIA

Fever is an elevation of body temperature that exceeds the normal daily variation and occurs *in conjunction with an increase in the hypothalamic set point* (e.g., from 37°C to 39°C). This shift of the set point from "normothermic" to febrile levels very much resembles the resetting of the home thermostat to a higher level in order to raise the ambient temperature in a room. Once the hypothalamic set point is raised, neurons in the vasomotor center are activated and vasoconstriction commences. The individual first notices vasoconstriction in the hands and feet. Shunting of blood away from the periphery to the internal organs essentially decreases heat loss from the skin, and the person feels cold. For most fevers, body temperature increases by 1–2°C. Shivering, which increases heat production from the muscles, may begin at this time; however, shivering is not required if mechanisms of heat conservation raise blood temperature sufficiently. Nonshivering heat production from the liver also contributes to increasing core temperature. Behavioral adjustments (e.g., putting on more clothing or bedding) help raise body temperature by decreasing heat loss.

The processes of heat conservation (vasoconstriction) and heat production (shivering and increased nonshivering thermogenesis) continue until the temperature of the blood bathing the hypothalamic neurons matches the new "thermostat setting." Once that point is reached, the hypothalamus maintains the temperature at the febrile level by the same mechanisms of heat balance that function in the afebrile state. When the hypothalamic set point is again reset downward (in response to either a reduction in the concentration of pyrogens or the use of antipyretics), the processes of heat loss through vasodilation and sweating are initiated. Loss of heat by sweating and vasodilation continues until the blood temperature at the hypothalamic level matches the lower setting. Behavioral changes (e.g., removal of clothing) facilitate heat loss.

A fever of >41.5°C (>106.7°F) is called *hyperpyrexia*. This extraordinarily high fever can develop in patients with severe infections but most commonly occurs in patients with central nervous system (CNS) hemorrhages. In the preantibiotic era, fever due to a variety of infectious diseases rarely exceeded 106°F, and there has been speculation that this natural "thermal ceiling" is mediated by neuropeptides functioning as central antipyretics.

In rare cases, the hypothalamic set point is elevated as a result of local trauma, hemorrhage, tumor, or intrinsic hypothalamic malfunction. The term *hypothalamic fever* is sometimes used to describe elevated temperature caused by abnormal hypothalamic function. However, most patients with hypothalamic damage have *sub*normal, not *supra*normal, body temperatures.

Although most patients with elevated body temperature have fever, there are circumstances in which elevated temperature represents not fever but *hyperthermia* (*heat stroke*). Hyperthermia is characterized by an uncontrolled increase in body temperature that exceeds the body's ability to lose heat. The setting of the hypothalamic thermoregulatory center is unchanged. In contrast to fever in infections, hyperthermia does not involve pyrogenic molecules. Exogenous heat exposure and endogenous heat production are two mechanisms by which hyperthermia can result in dangerously high internal temperatures. Excessive heat production can easily cause hyperthermia despite physiologic and behavioral control of body temperature. For example, work or exercise in hot environments can produce heat faster than peripheral mechanisms can lose it. **For a detailed discussion of hyperthermia, see Chap. 478.**

It is important to distinguish between fever and hyperthermia since hyperthermia can be rapidly fatal and characteristically does not respond to antipyretics. In an emergency situation, however, making this distinction can be difficult. For example, in systemic sepsis, fever (hyperpyrexia) can be rapid in onset, and temperatures can exceed 40.5°C (104.9°F). Hyperthermia is often diagnosed on the basis of the events immediately preceding the elevation of core temperature—e.g., heat exposure or treatment with drugs that interfere with thermoregulation. In patients with heat stroke syndromes and in those taking drugs that block sweating, the skin is hot but dry, whereas in fever, the skin can be cold as a consequence of vasoconstriction. Antipyretics do not reduce the elevated temperature in hyperthermia, whereas in fever—and even in hyperpyrexia—adequate doses of either aspirin or acetaminophen usually result in some decrease in body temperature.

PATHOGENESIS OF FEVER

■ PYROGENS

The term *pyrogen* (Greek *pyro*, "fire") is used to describe any substance that causes fever. *Exogenous* pyrogens are derived from outside the patient; most are microbial products, microbial toxins, or whole microorganisms (including viruses). The classic example of an exogenous pyrogen is the lipopolysaccharide (endotoxin) produced by all gram-negative bacteria. Endotoxin is a highly pyrogenic molecule in humans: when injected intravenously into volunteers, a dose of 2–3 ng/kg produces fever, leukocytosis, acute-phase proteins, and generalized symptoms of malaise. While cell wall components of gram-positive organisms are less pyrogenic than endotoxin, these organisms often produce toxins that act as *superantigens* and induce fever (e.g., *Staphylococcus aureus* toxic shock syndrome toxin-1, staphylococcal enterotoxins, streptococcal pyrogenic exotoxins). These products of staphylococci and streptococci cause fever in experimental animals when injected intravenously at concentrations of 1–10 μg/kg.

■ PYROGENIC CYTOKINES

Cytokines are small proteins (molecular mass, 10,000–20,000 Da) that regulate immune, inflammatory, and hematopoietic processes. For example, the elevated leukocytosis seen in several infections with an absolute neutrophilia is attributable to the cytokines interleukin (IL) 1 and IL-6. Some cytokines also cause fever; formerly referred to as *endogenous pyrogens*, they are now called *pyrogenic cytokines*. The pyrogenic cytokines include IL-1, IL-6, tumor necrosis factor (TNF), and ciliary neurotropic factor, a member of the IL-6 family. Fever is a prominent side effect of interferon α therapy. Each pyrogenic cytokine is encoded by a separate gene, and each has been shown to cause fever in laboratory animals and in humans. When injected into humans at low doses (10–100 ng/kg), IL-1 and TNF produce fever; in contrast, for IL-6, a dose of 1–10 μg/kg is required for fever production.

A wide spectrum of bacterial and fungal products induce the synthesis and release of pyrogenic cytokines. However, fever can be a manifestation of disease in the absence of microbial infection. For example, inflammatory processes such as pericarditis, trauma, stroke, and routine immunizations induce the production of IL-1, TNF, and/or IL-6; individually or in combination, these cytokines trigger the hypothalamus to raise the set point to febrile levels.

ELEVATION OF THE HYPOTHALAMIC SET POINT BY CYTOKINES

During fever, levels of prostaglandin E_2 (PGE_2) are elevated in hypothalamic tissue and the third cerebral ventricle. The concentrations of PGE_2 are highest near the circumventricular vascular organs (organum vasculosum of lamina terminalis)—networks of enlarged capillaries surrounding the hypothalamic regulatory centers. Destruction of these organs reduces the ability of pyrogens to produce fever. Most studies in animals have failed to show, however, that pyrogenic cytokines pass from the circulation into the brain itself. Thus, it appears that both exogenous pyrogens and pyrogenic cytokines interact with the endothelium of these capillaries and that this interaction is the first step in initiating fever—i.e., in raising the set point to febrile levels.

The key events in the production of fever are illustrated in **Fig. 20-1**. Myeloid and endothelial cells are the primary cell types that produce pyrogenic cytokines. Pyrogenic cytokines such as IL-1, IL-6, and TNF are released from these cells and enter the systemic circulation. Although these circulating cytokines lead to fever by inducing the synthesis of PGE_2, they also induce PGE_2 in peripheral tissues. The increase in PGE_2 in the periphery accounts for the nonspecific myalgias and arthralgias that often accompany fever. It is thought that some systemic PGE_2 escapes destruction by the lung and gains access to the hypothalamus via the internal carotid. However, it is the elevation of PGE_2 in the brain that starts the process of raising the hypothalamic set point for core temperature.

There are four receptors for PGE_2, and each signals the cell in different ways. Of the four receptors, the third (EP-3) is essential for fever: when the gene for this receptor is deleted in mice, no fever follows the injection of IL-1 or endotoxin. Deletion of the other PGE_2 receptor genes leaves the fever mechanism intact. Although PGE_2 is essential for fever, it is not a neurotransmitter. Rather, the release of PGE_2 from the brain side of the hypothalamic endothelium triggers the PGE_2 receptor on glial cells, and this stimulation results in the rapid release of cyclic adenosine 5′-monophosphate (cAMP), which is a neurotransmitter. As shown in Fig. 20-1, the release of cAMP from glial cells activates neuronal endings from the thermoregulatory center that extend into the area. The elevation of cAMP is thought to account for changes in the hypothalamic set point either directly or indirectly (by inducing the release of neurotransmitters). Distinct receptors for microbial products are located on the hypothalamic endothelium. These receptors are called *Toll-like receptors* and are similar in many ways to IL-1 receptors. IL-1 receptors and Toll-like receptors share the same signal-transducing mechanism. Thus, the direct activation of Toll-like receptors or IL-1 receptors results in PGE_2 production and fever.

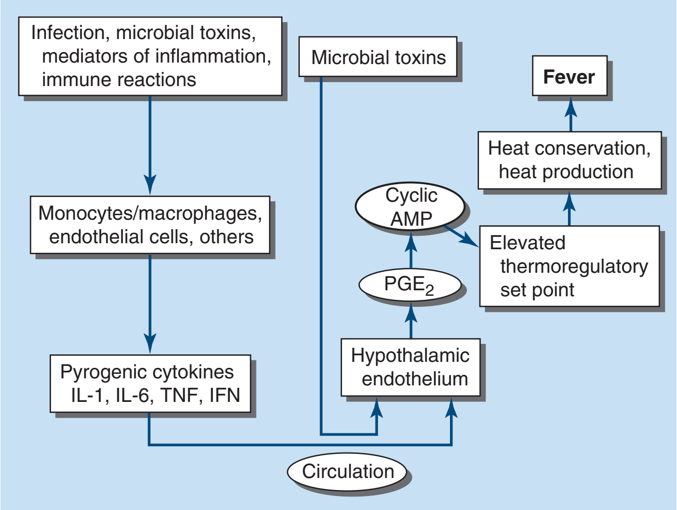
Infection, microbial toxins, mediators of inflammation, immune reactions
Microbial toxins
Fever
Heat conservation, heat production
Monocytes/macrophages, endothelial cells, others
Cyclic AMP
Elevated thermoregulatory set point
PGE2
Hypothalamic endothelium
Pyrogenic cytokines IL-1, IL-6, TNF, IFN
Circulation

FIGURE 20-1 Chronology of events required for the induction of fever. AMP, adenosine 5′-monophosphate; IFN, interferon; IL, interleukin; PGE_2, prostaglandin E_2; TNF, tumor necrosis factor.

PRODUCTION OF CYTOKINES IN THE CNS

Cytokines produced in the brain may account for the hyperpyrexia of CNS hemorrhage, trauma, or infection. Viral infections of the CNS induce microglial and possibly neuronal production of IL-1, TNF, and IL-6. In experimental animals, the concentration of a cytokine required to cause fever is several orders of magnitude lower with direct injection into the brain substance or brain ventricles than with systemic injection. Therefore, cytokines produced in the CNS can raise the hypothalamic set point, bypassing the circumventricular organs. CNS cytokines likely account for the hyperpyrexia of CNS hemorrhage, trauma, or infection.

APPROACH TO THE PATIENT

Fever

HISTORY AND PHYSICAL EXAMINATION

There are a range of disease processes that present with fever as a cardinal manifestation, and a thorough history can help distinguish between these broad categories **(Table 20-1)**. The chronology of events preceding fever, including exposure to other symptomatic individuals or to vectors of disease, should be ascertained. Electronic devices for measuring oral, tympanic membrane, or rectal temperatures are reliable, but the same site should be used consistently to monitor a febrile disease. Moreover, physicians should be aware that newborns, elderly patients, patients with chronic hepatic or renal failure, and patients taking glucocorticoids or being treated with an anticytokine may have active disease in the absence of fever because of a blunted febrile response.

LABORATORY TESTS

The workup should include a complete blood count; a differential count should be performed manually or with an instrument sensitive to the identification of neutrophil juvenile or band forms, toxic granulations, and Döhle bodies, which are suggestive of bacterial infection. Neutropenia may be present with some viral infections.

Measurement of circulating cytokines in patients with fever is not helpful since levels of cytokines such as IL-1 and TNF in the circulation often are below the detection limit of the assay or do not coincide with fever. However, in patients with low-grade fevers or with suspected occult disease, the most valuable measurements are the C-reactive protein (CRP) level and the erythrocyte sedimentation rate. These markers of inflammatory processes are particularly helpful in detecting occult disease. Measurement of circulating IL-6, which induces CRP, can be useful. However, whereas IL-6

TABLE 20-1 Disease Categories That Present with Fever as a Cardinal Sign

Infectious diseases
Autoimmune and noninfectious inflammatory disorders
Cancer
Medication related (e.g., vaccines, drug fever)
Endocrine disorders (e.g., hyperthyroidism)
Intrinsic hypothalamic malfunction

levels may vary during a febrile disease, CRP levels remain elevated. Acute-phase reactants are discussed in **Chap. 315**.

FEVER IN PATIENTS RECEIVING ANTICYTOKINE THERAPY

Patients receiving long-term treatment with anticytokine-based regimens are at increased risk of infection because of lowered host defenses. For example, latent *Mycobacterium tuberculosis* infection can disseminate in patients receiving anti-TNF therapy. With the increasing use of anticytokines to reduce the activity of IL-1, IL-6, IL-12/23, IL-17, or TNF in patients with Crohn disease, rheumatoid arthritis, or psoriasis, the possibility that these therapies blunt the febrile response should be kept in mind.

The blocking of cytokine activity has the distinct clinical drawback of lowering the level of host defenses against both routine bacterial and opportunistic infections such as *M. tuberculosis* and fungal infections. The use of monoclonal antibodies to reduce IL-17 in psoriasis increases the risk of systemic candidiasis.

In nearly all reported cases of infection associated with anticytokine therapy, fever is among the presenting signs. However, the extent to which the febrile response is blunted in these patients remains unknown. Therefore, low-grade fever in patients receiving anticytokine therapies is of considerable concern. The physician should conduct an early and rigorous diagnostic evaluation in these cases. The febrile response is also blunted in patients receiving chronic glucocorticoid therapy or anti-inflammatory agents such as nonsteroidal anti-inflammatory drugs (NSAIDs).

TREATMENT

Fever

THE DECISION TO TREAT FEVER

In deciding whether to treat fever, it is important to remember that fever itself is not an illness: it is an ordinary response to a perturbation of normal host physiology. Most fevers are associated with self-limited infections, such as common viral diseases. The use of antipyretics is not contraindicated in these infections: no significant clinical evidence indicates either that antipyretics delay the resolution of viral or bacterial infections or that fever facilitates recovery from infection or acts as an adjuvant to the immune system. In short, treatment of fever and its symptoms with routine antipyretics does no harm and does not slow the resolution of common viral and bacterial infections.

However, in bacterial infections, the withholding of antipyretic therapy can be helpful in evaluating the effectiveness of a particular antibiotic, especially in the absence of positive cultures of the infecting organism, and the routine use of antipyretics can mask an inadequately treated bacterial infection. Withholding antipyretics in some cases may facilitate the diagnosis of an unusual febrile disease. Temperature–pulse dissociation (*relative bradycardia*) occurs in typhoid fever, brucellosis, leptospirosis, some drug-induced fevers, and factitious fever. As stated earlier, in newborns, elderly patients, patients with chronic liver or kidney failure, and patients taking glucocorticoids, fever may not be present despite infection. Hypothermia can develop in patients with septic shock.

Some infections have characteristic patterns in which febrile episodes are separated by intervals of normal temperature. For example, *Plasmodium vivax* causes fever every third day, whereas fever occurs every fourth day with *Plasmodium malariae*. Another relapsing fever is related to *Borrelia* infection, with days of fever followed by a several-day afebrile period and then a relapse into additional days of fever. In the Pel-Ebstein pattern, fever lasting 3–10 days is followed by afebrile periods of 3–10 days; this pattern can be classic for Hodgkin disease and other lymphomas. In cyclic neutropenia, fevers occur every 21 days and accompany the neutropenia. There are also a number of periodic fever syndromes (e.g., familial Mediterranean fever, TNF receptor–associated periodic syndrome [TRAPS]) that differ in their periodicity, duration of attack, constellation of clinical features, genetic causes, and therapies **(Chap. 381)**. Understanding these clinical differences can help tailor diagnostic testing to confirm the diagnosis and guide therapy.

ANTICYTOKINE THERAPY TO REDUCE FEVER IN AUTOIMMUNE AND AUTOINFLAMMATORY DISEASES

Recurrent fever is documented at some point in most autoimmune diseases and many autoinflammatory diseases, which include the periodic fever syndromes as well as disorders of inflammasomes (e.g., NLRP3, pyrin) and other components of the innate immune system **(Chap. 360)**. Although fever can be a manifestation of autoimmune diseases, recurrent fevers are characteristic of autoinflammatory diseases, including uncommon diseases such as adult and juvenile Still disease, familial Mediterranean fever, and hyper-IgD syndrome but also common diseases such as idiopathic pericarditis and gout. In addition to recurrent fevers, neutrophilia and serosal inflammation characterize autoinflammatory diseases. The fevers associated with many of these illnesses are dramatically reduced by blocking of IL-1 activity with anakinra or canakinumab. Anticytokines therefore reduce fever in autoimmune and autoinflammatory diseases. Although fevers in autoinflammatory diseases are mediated by IL-1β, patients also respond to antipyretics.

MECHANISMS OF ANTIPYRETIC AGENTS

The reduction of fever by lowering of the elevated hypothalamic set point is a direct function of reduction of the PGE_2 level in the thermoregulatory center. The synthesis of PGE_2 depends on the constitutively expressed enzyme cyclooxygenase. The substrate for cyclooxygenase is arachidonic acid released from the cell membrane, and this release is the rate-limiting step in the synthesis of PGE_2. Therefore, inhibitors of cyclooxygenase are potent antipyretics. The antipyretic potency of various drugs is directly correlated with the inhibition of brain cyclooxygenase. Acetaminophen is a poor cyclooxygenase inhibitor in peripheral tissue and lacks noteworthy anti-inflammatory activity; in the brain, however, acetaminophen is oxidized by the P450 cytochrome system, and the oxidized form inhibits cyclooxygenase activity. Moreover, in the brain, the inhibition of another enzyme, COX-3, by acetaminophen may account for the antipyretic effect of this agent. However, COX-3 is not found outside the CNS.

Oral aspirin and acetaminophen are equally effective in reducing fever in humans. NSAIDs such as ibuprofen and specific inhibitors of COX-2 also are excellent antipyretics. Chronic, high-dose therapy with antipyretics such as aspirin or any NSAID does not reduce normal core body temperature. Thus, PGE_2 appears to play no role in normal thermoregulation.

As effective antipyretics, glucocorticoids act at two levels. First, similar to the cyclooxygenase inhibitors, glucocorticoids reduce PGE_2 synthesis by inhibiting the activity of phospholipase A_2, which is needed to release arachidonic acid from the cell membrane. Second, glucocorticoids block the transcription of the mRNA for the pyrogenic cytokines. Limited experimental evidence indicates that ibuprofen and COX-2 inhibitors reduce IL-1-induced IL-6 production and may contribute to the antipyretic activity of NSAIDs.

REGIMENS FOR THE TREATMENT OF FEVER

The objectives in treating fever are first to reduce the elevated hypothalamic set point and second to facilitate heat loss. Reducing fever with antipyretics also reduces systemic symptoms of headache, myalgias, and arthralgias.

Oral aspirin and NSAIDs effectively reduce fever but can adversely affect platelets and the gastrointestinal tract. Therefore, acetaminophen is preferred as an antipyretic. In children, acetaminophen or oral ibuprofen must be used because aspirin increases the risk of Reye syndrome with certain viral infections. If the patient cannot take oral antipyretics, parenteral preparations of NSAIDs and rectal suppositories of various antipyretics can be used.

Treatment of fever in some patients is highly recommended. Fever increases the demand for oxygen (i.e., for every increase of 1°C over 37°C, there is a 13% increase in oxygen consumption) and can aggravate the condition of patients with preexisting impairment of cardiac, pulmonary, or CNS function. Children with a history of febrile or nonfebrile seizure should be aggressively treated to reduce fever. However, it is unclear what triggers the febrile seizure, and there is no correlation between absolute temperature elevation and onset of a febrile seizure in susceptible children.

In hyperpyrexia, the use of cooling blankets facilitates the reduction of temperature; however, cooling blankets should not be used without oral antipyretics. In hyperpyretic patients with CNS disease or trauma (CNS bleeding), reducing core temperature mitigates the detrimental effects of high temperature on the brain.

For a discussion of treatment for hyperthermia, see Chap. 478.

■ FURTHER READING

Dinarello CA et al: Treating inflammation by blocking interleukin-1 in a broad spectrum of diseases. Nature Rev 11:633, 2012.

Gattorno M et al: Classification criteria for autoinflammatory recurrent fevers. Ann Rheum Dis 78:1025, 2019.

Kullenberg T et al: Long-term safety profile of anakinra in patients with severe cryopyrin-associated periodic syndromes. Rheumatology 55:1499, 2016.

Sakkat A et al: Temperature control in critically ill patients with fever: A meta-analysis of randomized controlled trials. J Crit Care 61:89, 2021.

21

Fever and Rash

Elaine T. Kaye, Kenneth M. Kaye

The acutely ill patient with fever and rash often presents a diagnostic challenge for physicians, yet the distinctive appearance of an eruption in concert with a clinical syndrome can facilitate a prompt diagnosis and the institution of life-saving therapy or critical infection-control interventions. **Representative images of many of the rashes discussed in this chapter are included in Chap. A1.**

APPROACH TO THE PATIENT

Fever and Rash

A thorough history of patients with fever and rash includes the following relevant information: immune status, medications taken within the previous month, specific travel history, immunization status, exposure to domestic pets and other animals, history of animal (including arthropod) bites, recent dietary exposures, existence of cardiac abnormalities, presence of prosthetic material, recent exposure to ill individuals, and sexual exposures. The history should also include the site of onset of the rash and its direction and rate of spread.

PHYSICAL EXAMINATION

A thorough physical examination entails close attention to the rash, with an assessment and precise identification of its salient features. First, it is critical to determine what *type* of lesions make up the eruption. *Macules* are flat lesions defined by an area of changed color (i.e., a blanchable erythema). *Papules* are raised, solid lesions <5 mm in diameter; *plaques* are lesions >5 mm in diameter with a flat, plateau-like surface; and *nodules* are lesions >5 mm in diameter with a more rounded configuration. *Wheals* (urticaria, hives) are papules or plaques that are pale pink and may appear annular (ring-like) as they enlarge; classic (nonvasculitic) wheals are transient, lasting only 24 h in any defined area. *Vesicles* (<5 mm) and *bullae* (>5 mm) are circumscribed, elevated lesions containing fluid. *Pustules* are raised lesions containing purulent exudate; vesicles such as due to varicella or herpes simplex may evolve to pustules. *Nonpalpable purpura* is a flat lesion that is due to bleeding into the skin. If <3 mm in diameter, the purpuric lesions are termed *petechiae*; if >3 mm, they are termed *ecchymoses*. *Palpable purpura* is a raised lesion that is due to inflammation of the vessel wall (vasculitis) with subsequent hemorrhage. An *ulcer* is a defect in the skin extending at least into the upper layer of the dermis, and an *eschar* (tâche noire) is a necrotic lesion covered with a black crust.

Other pertinent features of rashes include their *configuration* (i.e., annular or target), the *arrangement* of their lesions, and their *distribution* (i.e., central or peripheral).

For further discussion, see Chaps. 59, 61, 127, and 134.

■ CLASSIFICATION OF RASH

This chapter reviews rashes that reflect systemic disease, but it does not include localized skin eruptions (i.e., cellulitis, impetigo) that may also be associated with fever **(Chap. 134)**. The chapter is not intended to be all-inclusive, but it covers the most important and most common diseases associated with fever and rash. Rashes are classified herein on the basis of lesion morphology and distribution. For practical purposes, this classification system is based on the most typical disease presentations. However, morphology may vary as rashes evolve, and the presentation of diseases with rashes is subject to many variations **(Chap. 61)**. For instance, the classic petechial rash of Rocky Mountain spotted fever **(Chap. 192)** may initially consist of blanchable erythematous macules distributed peripherally; at times, however, the rash associated with this disease may not be predominantly acral, or no rash may develop at all.

Diseases with fever and rash may be classified by type of eruption: centrally distributed maculopapular, peripheral, confluent desquamative erythematous, vesiculobullous, urticaria-like, nodular, purpuric, ulcerated, or with eschars. Diseases are listed by these categories in **Table 21-1**, and many are highlighted in the text. However, for a more detailed discussion of each disease associated with a rash, the reader is referred to the chapter dealing with that specific disease. **(Reference chapters are cited in the text and listed in Table 21-1.)**

■ CENTRALLY DISTRIBUTED MACULOPAPULAR ERUPTIONS

Centrally distributed rashes, in which lesions are primarily truncal, are the most common type of eruption. The rash of *rubeola* (measles) starts at the hairline 2–3 days into the illness and moves down the body, typically sparing the palms and soles **(Fig. 21-1; see also Fig. A1-3) (Chap. 211)**. It begins as discrete erythematous lesions, which become confluent as the rash spreads. Koplik's spots (1- to 2-mm white or bluish lesions with an erythematous halo on the buccal mucosa) **(Fig. A1-2)** are pathognomonic for measles and are generally seen during the first 2 days of symptoms. They should not be confused with Fordyce's spots (ectopic sebaceous glands), which have no erythematous halos and are found in the mouth of healthy individuals. Koplik's spots may briefly overlap with the measles exanthem.

Rubella (German measles) **(Fig. A1-4)** also spreads from the hairline downward; unlike that of measles, however, the rash of rubella tends to clear from originally affected areas as it migrates, and it may be pruritic **(Chap. 212)**. Forchheimer spots (palatal petechiae) may develop but are nonspecific because they also develop in *infectious mononucleosis* **(Chap. 199)**, *scarlet fever* **(Chap. 153)**, and *Zika virus infection* **(Chap. 215) (Fig. A1-51*A-D*)**. Postauricular and suboccipital adenopathy and arthritis are common among adults with rubella. Exposure of pregnant women to ill individuals should be avoided, as rubella causes severe congenital abnormalities. Numerous strains of *enteroviruses* **(Chap. 210)**, primarily echoviruses and coxsackieviruses, cause nonspecific syndromes of fever and eruptions that may mimic rubella or measles. Patients with *infectious mononucleosis* caused by Epstein-Barr virus **(Chap. 199)** or with *primary HIV infection* **(Fig. A1-6; see also**

TABLE 21-1 Diseases Associated with Fever and Rash

DISEASE	ETIOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Centrally Distributed Maculopapular Eruptions					
Acute meningococcemia[a]	—	—	—	—	160
Drug reaction with eosinophilia and systemic symptoms (DRESS); also termed drug-induced hypersensitivity syndrome (DIHS)[b]; Chikungunya[c]; COVID-19[c]	—	—	—	—	63
Rubeola (measles, first disease) **(Fig. 21-1, Fig. A1-2, Fig. A1-3)**	Paramyxovirus	Discrete lesions that become confluent as rash spreads from hairline downward, usually sparing palms and soles; lasts ≥3 days; Koplik's spots	Nonimmune individuals	Cough, conjunctivitis, coryza, severe prostration	211
Rubella (German measles, third disease) **(Fig. A1-4)**	Togavirus	Spreads from hairline downward, clearing as it spreads; Forchheimer spots (palatal petechiae)	Nonimmune individuals	Adenopathy, arthritis	212
Erythema infectiosum (fifth disease) **(Fig. A1-1)**	Human parvovirus B19	Bright-red "slapped-cheeks" appearance followed by lacy reticular rash that waxes and wanes over 3 weeks; rarely, papular-purpuric "gloves-and-socks" syndrome on hands and feet	Most common among children 3–12 years old; occurs in winter and spring	Mild fever; arthritis in adults; rash following resolution of fever	202
Exanthem subitum (roseola, sixth disease) **(Fig. A1-5)**	Human herpesvirus 6 or, less commonly, the closely related human herpesvirus 7	Diffuse maculopapular eruption over trunk and neck; resolves within 2 days	Usually affects children <3 years old	Rash following resolution of fever; similar to Boston exanthem (echovirus 16); febrile seizures may occur	200
Primary HIV infection **(Fig. A1-6)**	HIV	Nonspecific diffuse macules and papules most commonly on upper thorax, face, collar region; less commonly, urticarial or vesicular lesions; oral or genital ulcers	Individuals recently infected with HIV	Pharyngitis, adenopathy, arthralgias	208
Infectious mononucleosis	Epstein-Barr virus	Diffuse maculopapular eruption historically in ~5% of cases increasing to ~90% of cases when antibiotics, particularly ampicillin, given, but recent observed rates of ~20% without antibiotics and little increase with antibiotics; urticaria, petechiae in some cases; periorbital edema (50%); palatal petechiae (25%)	Adolescents, young adults	Hepatosplenomegaly, pharyngitis, cervical lymphadenopathy, atypical lymphocytosis, heterophile antibody	199
Other viral exanthems	Echoviruses 2, 4, 9, 11, 16, 19, 25; coxsackieviruses A9, B1, B5; etc.	Wide range of skin findings that may mimic rubella or measles	Affects children more commonly than adults	Nonspecific viral syndromes	210
Exanthematous drug-induced eruption **(Fig. A1-7)**	Drugs (antibiotics, anticonvulsants, diuretics, etc.)	Intensely pruritic, bright-red macules and papules, symmetric on trunk and extremities; may become confluent	Occurs 2–3 days after exposure in previously sensitized individuals; otherwise, after 2–3 weeks (but can occur anytime, even shortly after drug is discontinued)	Variable findings: fever and eosinophilia	63
Epidemic typhus	*Rickettsia prowazekii*	Maculopapular eruption appearing in axillae, spreading to trunk and later to extremities; usually spares face, palms, soles; evolves from blanchable macules to confluent eruption with petechiae; rash evanescent in recrudescent typhus (Brill-Zinsser disease)	Exposure to body lice; occurrence of recrudescent typhus as relapse after 30–50 years	Headache, myalgias; mortality rates 10–40% if untreated; milder clinical presentation in recrudescent form	192
Endemic (murine) typhus	*Rickettsia typhi*	Maculopapular eruption, usually sparing palms, soles	Exposure to rat or cat fleas	Headache, myalgias	192
Scrub typhus	*Orientia tsutsugamushi*	Diffuse macular rash starting on trunk; eschar at site of mite bite	Endemic in South Pacific, Australia, Asia; transmitted by mites	Headache, myalgias, regional adenopathy; mortality rates up to 30% if untreated	192

(*Continued*)

TABLE 21-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETIOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Rickettsial spotted fevers **(Fig. 21-8)**	*Rickettsia conorii* (boutonneuse fever), *Rickettsia australis* (North Queensland tick typhus), *Rickettsia sibirica* (Siberian tick typhus), *Rickettsia africae* (African tick-bite fever), and others	Eschar common at bite site; maculopapular (rarely, vesicular and petechial) eruption on proximal extremities, spreading to trunk and face	Exposure to ticks; *R. conorii* in Mediterranean region, India, Africa; *R. australis* in Australia; *R. sibirica* in Siberia, Mongolia; *R. africae* in Africa, Caribbean	Headache, myalgias, regional adenopathy	**192**
Human monocytotropic ehrlichiosis[d]	*Ehrlichia chaffeensis*	Maculopapular eruption (40% of cases), involves trunk and extremities; may be petechial	Tick-borne; most common in U.S. Southeast, southern Midwest, and mid-Atlantic regions	Headache, myalgias, leukopenia	**192**
Leptospirosis	*Leptospira interrogans* and other *Leptospira* species	Maculopapular eruption; conjunctivitis; scleral hemorrhage in some cases	Exposure to water contaminated with animal urine	Myalgias; aseptic meningitis; *fulminant form*: icterohemorrhagic fever (Weil's disease)	**189**
Lyme disease **(Fig. A1-8)**	*Borrelia burgdorferi* (sole cause in U.S.), *Borrelia afzelii, Borrelia garinii*	Papule expanding to erythematous annular lesion with central clearing (erythema migrans; average diameter, 15 cm), sometimes with concentric rings, sometimes with indurated or vesicular center; multiple secondary erythema migrans lesions in some cases	Bite of *Ixodes* tick vector	Headache, myalgias, chills, photophobia occurring acutely; CNS disease, myocardial disease, arthritis weeks to months later in some cases	**191**
Southern tick-associated rash illness (STARI, Master's disease)	Unknown (possibly *Borrelia lonestari* or other *Borrelia* spirochetes)	Similar to erythema migrans of Lyme disease with several differences, including: multiple secondary lesions less likely; lesions tending to be smaller (average diameter, ~8 cm); central clearing more likely	Bite of tick vector *Amblyomma americanum* (Lone Star tick); often found in regions where Lyme disease is uncommon, including southern United States	Compared with Lyme disease: fewer constitutional symptoms, tick bite more likely to be recalled; other Lyme disease sequelae lacking	**191**
Typhoid fever **(Fig. A1-9)**	*Salmonella typhi*	Transient, blanchable erythematous macules and papules, 2–4 mm, usually on trunk (rose spots)	Ingestion of contaminated food or water (rare in U.S.)	Variable abdominal pain and diarrhea; headache, myalgias, hepatosplenomegaly	**171**
Dengue fever[e] **(Fig. A1-53)**	Dengue virus (4 serotypes; flaviviruses)	Rash in 50% of cases; initially diffuse flushing; midway through illness, onset of maculopapular rash, which begins on trunk and spreads centrifugally to extremities and face; pruritus, hyperesthesia in some cases; after defervescence, petechiae on extremities may occur	Occurs in tropics and subtropics; transmitted by mosquito	Headache; musculoskeletal pain ("breakbone fever"); leukopenia; occasionally biphasic ("saddleback") fever	**215**
Rat-bite fever (sodoku)	*Spirillum minus*	Eschar at bite site; then blotchy violaceous or red-brown rash involving trunk and extremities	Rat bite; primarily found in Asia; rare in U.S.	Regional adenopathy; recurrent fevers if untreated	**146**
Relapsing fever	*Borrelia* species	Central rash at end of febrile episode; petechiae in some cases	Exposure to ticks or body lice	Recurrent fever, headache, myalgias, hepatosplenomegaly	**190**
Erythema marginatum (rheumatic fever)	Group A *Streptococcus*	Erythematous annular papules and plaques occurring as polycyclic lesions in waves over trunk, proximal extremities; evolving and resolving within hours	Patients with rheumatic fever	Pharyngitis preceding polyarthritis, carditis, subcutaneous nodules, chorea	**153**
Systemic lupus erythematosus (SLE) **(Fig. A1-10, Fig. A1-11, Fig. A1-12)**	Autoimmune disease	Macular and papular erythema, often in sun-exposed areas; discoid lupus lesions (local atrophy, scale, pigmentary changes); periungual telangiectasis; malar rash; vasculitis sometimes causing urticaria, palpable purpura; oral erosions in some cases	Most common in young to middle-aged women; flares precipitated by sun exposure	Arthritis; cardiac, pulmonary, renal, hematologic, and vasculitic disease	**368**
Still's disease **(Fig. A1-13)**	Autoimmune disease	Transient 2- to 5-mm erythematous papules appearing at height of fever on trunk, proximal extremities; lesions evanescent	Children and young adults	High spiking fever, polyarthritis, splenomegaly; erythrocyte sedimentation rate >100 mm/h	—

(*Continued*)

TABLE 21-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETIOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
African trypanosomiasis **(Fig. A1-47)**	*Trypanosoma brucei rhodesiense/gambiense*	Blotchy or annular erythematous macular and papular rash (trypanid), primarily on trunk; pruritus; chancre at site of tsetse fly bite may precede rash by several weeks	Tsetse fly bite in eastern (*T. brucei rhodesiense*) or western (*T. brucei gambiense*) Africa	Hemolymphatic disease followed by meningoencephalitis; Winterbottom's sign (posterior cervical lymphadenopathy) (*T. brucei gambiense)*	**234**
Arcanobacterial pharyngitis	*Arcanobacterium (Corynebacterium) haemolyticum*	Diffuse, erythematous, maculopapular eruption involving trunk and proximal extremities; may desquamate	Children and young adults	Exudative pharyngitis, lymphadenopathy	**155**
West Nile virus infection	West Nile virus	Maculopapular eruption involving the trunk, extremities, and head or neck; rash in 20–50% of cases	Mosquito bite; rarely, blood transfusion or transplanted organ	Headache, weakness, malaise, myalgia, neuroinvasive disease (encephalitis, meningitis, flaccid paralysis)	**215**
Zika virus infection **(Fig. A1-51)**	Zika virus	Pruritic macular and papular erythema; rash may begin on trunk and descend to lower body; conjunctival injection; palatal petechiae may occur	Mosquito bite; sexual transmission or blood transfusion less common	Arthralgia (especially of small joints), myalgia, lymphadenopathy, headache, low-grade fever; illness in pregnancy may cause severe birth defects, including microcephaly; neurologic complications, including Guillain-Barré, may occur	**215**
Peripheral Eruptions					
Chronic meningococcemia, disseminated gonococcal infection,[a] human parvovirus B19 infection,[f] RIME[g]	—	—	—	—	**160, 161, 202**
Rocky Mountain spotted fever **(Fig. 21-2, Fig. A1-16)**	*Rickettsia rickettsii*	Rash beginning on wrists and ankles and spreading centripetally; appears on palms and soles later in disease; lesion evolution from blanchable macules to petechiae	Tick vector; widespread but more common in southeastern and southwest-central U.S.	Headache, myalgias, abdominal pain; mortality rates up to 40% if untreated	**192**
Secondary syphilis **(Fig. A1-18, Fig. A1-19, Fig. A1-20, Fig. A1-21)**	*Treponema pallidum*	Coincident primary chancre in 10% of cases; copper-colored, scaly papular eruption, diffuse but prominent on palms and soles; rash never vesicular in adults; condyloma latum, mucous patches, and alopecia in some cases	Sexually transmitted	Fever, constitutional symptoms	**187**
Chikungunya fever **(Fig. A1-54)**	Chikungunya virus	Maculopapular eruption; typically occurs on trunk, but also occurs on extremities and face	*Aedes aegypti* and *A. albopictus* mosquito bites; tropical and subtropical regions	Severe polyarticular, migratory arthralgias, especially involving small joints (e.g., hands, wrists, ankles)	**215**
Hand-foot-and-mouth disease **(Fig. A1-22)**	Coxsackievirus A16 and enterovirus 71 most common causes; coxsackievirus A6 associated with atypical syndrome	Tender vesicles, erosions in mouth; 0.25- to 1-cm papules on hands and feet with rim of erythema evolving into tender vesicles; shedding of nails (onychomadesis) can occur 1–2 months after acute illness; coxsackievirus A6 lesions may also be maculopapular, petechial, purpuric, or erosive; atypical form often extends to perioral area, extremities, trunk, buttocks, genitals, and areas affected by eczema (eczema coxsackium)	Summer and fall; primarily children <10 years old; multiple family members; coxsackievirus A6 infection also occurs in young adults	Transient fever; enterovirus 71 can be associated with brainstem encephalitis, flaccid paralysis resembling polio, or aseptic meningitis	**210**
Erythema multiforme (EM) **(Fig. A1-24)**	Infection, drugs, idiopathic causes	Target lesions (central erythema surrounded by area of clearing and another rim of erythema) up to 2 cm; symmetric on knees, elbows, palms, soles; spreads centripetally; papular, sometimes vesicular; when extensive and involving mucous membranes, termed *EM major*	Herpes simplex virus or *Mycoplasma pneumoniae* infection; drug intake (i.e., sulfa, phenytoin, penicillin)	50% of patients <20 years old; fever more common in most severe form, EM major, which can be confused with Stevens-Johnson syndrome (but EM major lacks prominent skin sloughing)	—[h]

(*Continued*)

TABLE 21-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETIOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Rat-bite fever (Haverhill fever)	*Streptobacillus moniliformis*	Maculopapular eruption over palms, soles, and extremities; tends to be more severe at joints; eruption sometimes becoming generalized; may be purpuric; may desquamate	Rat bite, ingestion of contaminated food	Myalgias; arthritis (50%); fever recurrence in some cases	**146**
Bacterial endocarditis **(Fig. A1-23)**	*Streptococcus*, *Staphylococcus*, etc.	*Subacute course* (e.g., viridans streptococci): Osler's nodes (tender pink nodules on finger or toe pads); petechiae on skin and mucosa; splinter hemorrhages. *Acute course* (e.g., *Staphylococcus aureus*): Janeway lesions (painless erythematous or hemorrhagic macules, usually on palms and soles)	Abnormal heart valve (e.g., viridans streptococci), intravenous drug use	New or changing heart murmur	**133**
COVID-19 **(Fig. A1-57)**	SARS-CoV-2	*Mild or asymptomatic COVID-19:* Pernio (macules, papules, or plaques that are tender, erythematous/violaceous; acral, feet more common than hands). *Moderate/severe COVID-19:* Vesicles, urticaria, maculopapular erythema; often pruritic; occur on trunk, extremities. *Severe COVID-19:* Retiform purpura (net-like, purple patches/plaques often with necrosis); lesions often asymptomatic; occur on extremities, buttocks. *Multisystem inflammatory syndrome in children (MIS-C):* Findings *similar to Kawasaki disease*	Infection with SARS-CoV-2; MIS-C in older children/adolescents	Ranging from asymptomatic to mild/moderate with loss of taste/smell, pharyngitis, cough, fever, to severe with dyspnea, ARDS; complications include thrombosis, especially with retiform purpura; lesions may be delayed compared to other COVID-19 symptoms; MIS-C occurs ~2-6 weeks following acute (often asymptomatic) infection	
Confluent Desquamative Erythemas					
Scarlet fever (second disease) **(Fig. A1-25)**	Group A *Streptococcus* (pyrogenic exotoxins A, B, C)	Diffuse blanchable erythema beginning on face and spreading to trunk and extremities; circumoral pallor; "sandpaper" texture to skin; accentuation of linear erythema in skin folds (Pastia's lines); enanthem of white evolving into red "strawberry" tongue; desquamation in second week	Most common among children 2–10 years old; usually follows group A streptococcal pharyngitis	Fever, pharyngitis, headache	**153**
Kawasaki disease **(Fig. A1-29)**	Idiopathic	Rash similar to scarlet fever (scarlatiniform) or EM; fissuring of lips, strawberry tongue; conjunctivitis; edema of hands, feet; desquamation later in disease	Children <8 years old	Cervical adenopathy, pharyngitis, coronary artery vasculitis	**61, 375**
Streptococcal toxic shock syndrome	Group A *Streptococcus* (associated with pyrogenic exotoxin A and/or B or certain M types)	When present, rash often scarlatiniform	May occur in setting of severe group A streptococcal infections (e.g., necrotizing fasciitis, bacteremia, pneumonia)	Multiorgan failure, hypotension; mortality rate 30%	**153**
Staphylococcal toxic shock syndrome	*S. aureus* (toxic shock syndrome toxin 1, enterotoxins B and others)	Diffuse erythema involving palms; pronounced erythema of mucosal surfaces; conjunctivitis; desquamation 7–10 days into illness	Colonization with toxin-producing *S. aureus*	Fever >39°C (>102°F), hypotension, multiorgan dysfunction	**152**
Staphylococcal scalded-skin syndrome **(Fig. A1-28)**	*S. aureus*, phage group II	Diffuse tender erythema, often with bullae and desquamation; Nikolsky's sign	Colonization with toxin-producing *S. aureus*; occurs in children <10 years old (termed *Ritter's disease* in neonates) or adults with renal dysfunction	Irritability; nasal or conjunctival secretions	**152**
Exfoliative erythroderma syndrome **(Fig. A1-27)**	Underlying psoriasis, eczema, drug eruption, mycosis fungoides	Diffuse erythema (often scaling) interspersed with lesions of underlying condition	Usually occurs in adults over age 50; more common among men	Fever, chills (i.e., difficulty with thermoregulation); lymphadenopathy	**61, 63**

(*Continued*)

TABLE 21-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETIOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
DRESS (drug reaction with eosinophilia and systemic symptoms; also known as drug-induced hypersensitivity syndrome [DIHS]) **(Fig. A1-48)**	Aromatic anticonvulsants; other drugs, including sulfonamides, minocycline	Maculopapular eruption (mimicking exanthematous drug rash), sometimes progressing to exfoliative erythroderma; profound edema, especially facial; pustules may occur	Individuals genetically unable to detoxify arene oxides (anticonvulsant metabolites), patients with slow *N*-acetylating capacity (sulfonamides)	Lymphadenopathy, multiorgan failure (especially hepatic), eosinophilia, atypical lymphocytes; HHV6 viremia; mimics sepsis	**63**
Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN) **(Fig. 21-3, Fig. A1-26)**	Drugs (80% of cases; often allopurinol, anticonvulsants, antibiotics), infection, idiopathic factors	Erythematous and purpuric macules, sometimes targetoid, or diffuse erythema progressing to bullae, with sloughing and necrosis of entire epidermis; Nikolsky's sign; involves mucosal surfaces; TEN (>30% epidermal necrosis) is maximal form; SJS involves <10% of epidermis; SJS/TEN overlap involves 10–30% of epidermis	Uncommon among children; more common among people living with HIV systemic lupus erythematosus, certain HLA types, or slow acetylators	Dehydration, sepsis sometimes resulting from lack of normal skin integrity; mortality rates up to 30%	**63**
Vesiculobullous or Pustular Eruptions					
Hand-foot-and-mouth syndrome[c]; staphylococcal scalded-skin syndrome[b]; TEN[b]; DRESS/DIHS[b]; COVID-19[c]	—	—	—	—	—[h]
Varicella (chickenpox) **(Fig. 21-4, Fig. A1-30)**	Varicella-zoster virus (VZV)	Macules (2–3 mm) evolving into papules, then vesicles (sometimes umbilicated), on an erythematous base ("dewdrops on a rose petal"); pustules then forming and crusting; lesions appearing in crops; may involve scalp, mouth; intensely pruritic	Usually affects children; 10% of adults susceptible; most common in late winter and spring; incidence down by 90% in U.S. as a result of varicella vaccination	Malaise; generally mild disease in healthy children; more severe disease with complications in adults and immunocompromised children	**198**
Pseudomonas "hot-tub" folliculitis **(Fig. A1-55)**	*Pseudomonas aeruginosa*	Pruritic erythematous follicular, papular, vesicular, or pustular lesions that may involve axillae, buttocks, abdomen, and especially areas occluded by bathing suits; can manifest as tender isolated nodules on palmar or plantar surfaces (the latter designated "*Pseudomonas* hot-foot syndrome")	Bathers in hot tubs or swimming pools; occurs in outbreaks	Earache, sore eyes and/or throat; fever may be absent; generally self-limited	**170**
Variola (smallpox) **(Fig. A1-50)**	Variola major virus	Red macules on tongue and palate evolving to papules and vesicles; skin macules evolving to papules, then vesicles, then pustules over 1 week, with subsequent lesion crusting; lesions initially appearing on face and spreading centrifugally from trunk to extremities; differs from varicella in that (1) skin lesions in any given area are at same stage of development and (2) there is a prominent distribution of lesions on face and extremities (including palms, soles)	Nonimmune individuals exposed to smallpox	Prodrome of fever, headache, backache, myalgias; vomiting in 50% of cases	**S4**
Mpox **(Fig. A1-59)**	Monkeypox virus	*Classically,* lesions similar in morphology and distribution to those of variola (smallpox); *2022 outbreak:* macules evolve to papules, vesicles, and pustules, with subsequent crusting over 1–2 weeks; often umbilicated; unlike traditional Mpox, lesions may be in different stages of development and typically number fewer than 20; lesions often painful; mucosal lesions (anorectal, oropharyngeal, ocular) may occur at sites of inoculation; disease may be severe and progressive in advanced HIV	Nonimmune individuals exposed to monkeypox; virus endemic to Central and West Africa; 2022 outbreak: among nonendemic countries including Europe, U.S.; most cases in men who have sex with men	*Classically,* similar to smallpox, though with lymphadenopathy and typically milder; 2022 outbreak: fever, lymphadenopathy, headache, proctitis, or pharyngitis	**201**

(*Continued*)

TABLE 21-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETIOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Primary herpes simplex virus (HSV) infection (Fig. A1-58)	HSV	Erythema rapidly followed by hallmark painful *grouped vesicles* that may evolve into pustules that ulcerate, especially on mucosal surfaces; lesions at site of inoculation: commonly gingivostomatitis for HSV-1 and genital lesions for HSV-2; recurrent disease milder (e.g., herpes labialis typically does not involve oral mucosa)	Primary infection most common among children and young adults for HSV-1 and among sexually active young adults for HSV-2; no fever in recurrent infection	Regional lymphadenopathy	197
Disseminated herpesvirus infection (Fig. A1-31, Fig. A1-58D)	VZV or HSV	Generalized nongrouped vesicles that can evolve to pustules and ulcerations; individual lesions similar for VZV and HSV. *Zoster cutaneous dissemination*: >25 lesions extending outside involved dermatome. *HSV*: extensive, progressive mucocutaneous lesions may occur in absence of dissemination; HSV may disseminate in eczematous skin (eczema herpeticum); HSV visceral dissemination may occur with only localized mucocutaneous disease; in disseminated HSV neonatal disease, skin lesions diagnostically helpful when present, but rash absent in a substantial minority of cases	Patients with immunosuppression, eczema; neonates	Visceral organ involvement (e.g., liver, lungs) in some cases; neonatal disease particularly severe	143, 197, 198
Rickettsialpox (Fig. A1-33)	*Rickettsia akari*	Eschar found at site of mite bite; generalized rash involving face, trunk, extremities; may involve palms and soles; <100 papules and plaques (2–10 mm); centers of papules develop vesicles or pustules	Seen in urban settings; transmitted by mouse mites	Headache, myalgias, regional adenopathy; mild disease	192
Acute generalized exanthematous pustulosis (Fig. A1-49)	Drugs (mostly anticonvulsants or antimicrobials); also viral	Tiny, sterile, nonfollicular pustules on erythematous, edematous skin; begins on face and in body folds, then becomes generalized	Appears 2–21 days after start of drug therapy, depending on whether patient has been sensitized	Acute fever, pruritus, leukocytosis	63
Disseminated *Vibrio vulnificus* infection	*V. vulnificus*	Erythematous lesions evolving into hemorrhagic bullae and then into necrotic ulcers	Patients with cirrhosis, diabetes, renal failure; exposure by ingestion of contaminated saltwater, seafood	Hypotension; mortality rate 50%	173
Ecthyma gangrenosum (Fig. A1-34)	*P. aeruginosa*, other gram-negative rods, fungi	Indurated plaque evolving into hemorrhagic bulla or pustule that sloughs, resulting in eschar formation; erythematous halo; most common in axillary, groin, perianal regions	Usually affects neutropenic patients; occurs in up to 28% of individuals with *Pseudomonas* bacteremia	Clinical signs of sepsis	170
Reactive infectious mucocutaneous eruption (RIME) (this generalized term encompasses multiple infectious etiologies, and includes *Mycoplasma pneumoniae*–induced rash and mucositis [MIRM])	*M. pneumoniae, Chlamydia pneumoniae,* human metapneumovirus, parainfluenzavirus 2, rhinovirus, influenza B virus, SARS-CoV-2	Severe mucositis of at least two sites (e.g., oropharynx, ocular, genital) with nearly universal hemorrhagic crusting of lips; sparse, vesiculobullous, or atypical targetoid rash over <10% of body; lesions typically on extremities but can be truncal; rash sometimes absent	More common in males; usually children (mean age 11–12 years old)	Cough, respiratory infection symptoms often precede rash by ~1 week, good prognosis; distinct from SJS/TEN	
Urticaria-Like Eruptions					
COVID-19^c					
Urticarial vasculitis (Fig. 21-5, Fig. A1-35)	Serum sickness, often due to infection (including acute hepatitis B, enteroviral, parasitic), drugs; connective tissue disease	Erythematous, edematous "urticaria-like" plaques, pruritic or burning; unlike urticaria: typical lesion duration >24 h (up to 5 days) and lack of complete lesion blanching with compression due to hemorrhage	Patients with serum sickness (including acute hepatitis B), connective tissue disease	Fever variable; arthralgias/arthritis	375^h

(*Continued*)

TABLE 21-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETIOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Nodular Eruptions					
Disseminated infection (**Fig. 21-6**, **Fig. A1-36**, **Fig. A1-37**, **Fig. A1-38**)	Fungal infections (e.g., candidiasis, histoplasmosis, cryptococcosis, sporotrichosis, coccidioidomycosis); mycobacteria	Subcutaneous nodules (up to 3 cm); fluctuance, draining common with mycobacteria; necrotic nodules (extremities, periorbital or nasal regions) common with *Aspergillus*, *Mucor*	Immunocompromised hosts (e.g., bone marrow transplant recipients, patients undergoing chemotherapy, HIV-infected patients)	Features vary with organism	—[h]
Erythema nodosum (septal panniculitis) (**Fig. A1-39**)	Infections (e.g., streptococcal, fungal, mycobacterial, yersinial); drugs (e.g., sulfas, penicillins, oral contraceptives); sarcoidosis; idiopathic causes	Large, violaceous, nonulcerative, subcutaneous nodules; exquisitely tender; usually on lower legs but also on upper extremities	More common among females 15–30 years old	Arthralgias (50%); features vary with associated condition	—[h]
Sweet syndrome (acute febrile neutrophilic dermatosis) (**Fig. A1-40**)	*Yersinia* infection; upper respiratory infection; inflammatory bowel disease; pregnancy; malignancy (usually hematologic); drugs (G-CSF)	Tender red or blue edematous nodules giving impression of vesiculation; usually on face, neck, upper extremities; when on lower extremities, may mimic erythema nodosum	More common among women and among persons 30–60 years old; 20% of cases associated with malignancy (men and women equally affected in this group)	Headache, arthralgias, leukocytosis	**61**
Bacillary angiomatosis	*Bartonella henselae*, *B. quintana*	Many forms, including erythematous, smooth vascular nodules; friable, exophytic lesions; erythematous plaques (may be dry, scaly); subcutaneous nodules (may be erythematous)	Immunosuppressed individuals, especially those with advanced HIV infection	Peliosis of liver and spleen in some cases; lesions sometimes involving multiple organs; bacteremia	**177**
Purpuric Eruptions					
Rocky Mountain spotted fever, rat-bite fever, endocarditis[c]; epidemic typhus[f]; dengue fever[e,f]; human parvovirus B19 infection[f]; COVID-19[c]	—	—	—	—	—[h]
Acute meningococcemia	*Neisseria meningitidis*	Initially pink maculopapular lesions evolving into petechiae; petechiae rapidly becoming numerous, sometimes enlarging and becoming vesicular; trunk, extremities most commonly involved; may appear on face, hands, feet; may include purpura fulminans (see below) reflecting DIC	Most common among children, individuals with asplenia or terminal complement component deficiency (C5–C8)	Hypotension, meningitis (sometimes preceded by upper respiratory infection)	**160**
Purpura fulminans (**Fig. A1-41**)	Severe DIC	Large ecchymoses with sharply irregular shapes evolving into hemorrhagic bullae and then into black necrotic lesions	Individuals with sepsis (e.g., involving *N. meningitidis*), malignancy, or massive trauma; asplenic patients at high risk for sepsis	Hypotension	**160, 315**
Chronic meningococcemia (**Fig. A1-42**)	*N. meningitidis*	Variety of recurrent eruptions, including pink maculopapular; nodular (usually on lower extremities); petechial (sometimes developing vesicular centers); purpuric areas with pale blue-gray centers	Individuals with complement deficiencies	Fevers, sometimes intermittent; arthritis, myalgias, headache	**160**
Disseminated gonococcal infection (**Fig. A1-43**)	*Neisseria gonorrhoeae*	Papules (1–5 mm) evolving over 1–2 days into hemorrhagic pustules with gray necrotic centers; hemorrhagic bullae occurring rarely; lesions (usually <40) distributed peripherally near joints (more commonly on upper extremities)	Sexually active individuals (more often females), some with complement deficiency	Low-grade fever, tenosynovitis, arthritis	**161**
Enteroviral petechial rash	Usually echovirus 9 or coxsackievirus A9	Disseminated petechial lesions (may also be maculopapular, vesicular, or urticarial)	Often occurs in outbreaks	Pharyngitis, headache; aseptic meningitis with echovirus 9	**210**
Viral hemorrhagic fever	Arenaviruses, bunyaviruses, filoviruses (including Ebola), flaviviruses (including dengue)	Petechial rash	Residence in or travel to endemic areas, other virus exposure	Triad of fever, shock, hemorrhage from mucosa or gastrointestinal tract	**215, 216**

(*Continued*)

TABLE 21-1 Diseases Associated with Fever and Rash (Continued)

DISEASE	ETIOLOGY	DESCRIPTION	GROUP AFFECTED/ EPIDEMIOLOGIC FACTORS	CLINICAL SYNDROME	CHAPTER
Thrombotic thrombocytopenic purpura/hemolytic-uremic syndrome	Idiopathic, bloody diarrhea caused by Shiga toxin–generating bacteria (e.g., *Escherichia coli* O157:H7), deficiency in ADAMTS13 (cleaves von Willebrand factor), drugs (e.g., quinine, chemotherapy, immunosuppression)	Petechiae	Individuals with Shiga toxin producing bacterial (commonly *E. coli* O157:H7) gastroenteritis (especially children), cancer chemotherapy, HIV infection, autoimmune diseases, pregnant/postpartum women, those with ADAMTS13 deficiency	Fever (not always present), microangiopathic hemolytic anemia, thrombocytopenia, renal dysfunction, neurologic dysfunction; coagulation studies normal	**61, 105, 120, 166, 172**
Cutaneous small-vessel vasculitis (leukocytoclastic vasculitis) **(Fig. 21-7, Fig. A1-44)**	Infections (including group A streptococcal infection, hepatitis B or C), drugs, idiopathic factors	Palpable purpuric lesions appearing in crops on legs or other dependent areas; may become vesicular or ulcerative	Occurs in a wide spectrum of diseases, including connective tissue disease, cryoglobulinemia, malignancy, Henoch-Schönlein purpura (HSP); more common among children	Fever (not always present), malaise, arthralgias, myalgias; systemic vasculitis in some cases; renal, joint, and gastrointestinal involvement common in HSP	**61**
Eruptions with Ulcers and/or Eschars					
Scrub typhus, rickettsial spotted fevers, rat-bite fever, African trypanosomiasis[f]; rickettsialpox, ecthyma gangrenosum[g]	—	—	—	—	—[h]
Tularemia **(Fig. A1-45, Fig. A1-46)**	*Francisella tularensis*	Ulceroglandular form: erythematous, tender papule evolves into necrotic, tender ulcer with raised borders; in 35% of cases, eruptions (maculopapular, vesiculopapular, acneiform, or urticarial; erythema nodosum; or EM) may occur	Exposure to ticks, biting flies, infected animals	Fever, headache, lymphadenopathy	**175**
Anthrax **(Fig. A1-52)**	*Bacillus anthracis*	Pruritic papule enlarging and evolving into a 1- by 3-cm painless ulcer surrounded by vesicles and then developing a central eschar with edema; residual scar	Exposure to infected animals or animal products, other exposure to anthrax spores	Lymphadenopathy, headache	**S4**

[a] See "Purpuric Eruptions." [b] See "Confluent Desquamative Erythemas." [c] See "Peripheral Eruptions." [d] Rash is rare in human granulocytotropic ehrlichiosis or anaplasmosis (caused by *Anaplasma phagocytophilum*; most common in the upper midwestern and northeastern United States). [e] See "Viral hemorrhagic fever" under "Purpuric Eruptions" for dengue hemorrhagic fever/dengue shock syndrome. [f] See "Centrally Distributed Maculopapular Eruptions." [g] See "Vesiculobullous or Pustular Eruptions." [h] See etiology-specific chapters.

Abbreviations: ARDS, acute respiratory distress syndrome; CNS, central nervous system; DIC, disseminated intravascular coagulation; G-CSF, granulocyte colony-stimulating factor; HLA, human leukocyte antigen.

Chapter 208) may exhibit pharyngitis, lymphadenopathy, and a nonspecific maculopapular exanthem.

The rash of *erythema infectiosum* (fifth disease), which is caused by human parvovirus B19, primarily affects children 3–12 years old; it develops after fever has resolved as a bright blanchable erythema on the cheeks ("slapped cheeks") **(Fig. A1-1*A*)** with perioral pallor **(Chap. 202)**. A more diffuse rash (often pruritic) appears the next day on the trunk and extremities and then rapidly develops into a lacy reticular eruption **(Fig. A1-1*B*)** that may wax and wane (especially with temperature change) over 3 weeks. Adults with fifth disease often have arthritis, and fetal hydrops can develop in association with this condition in pregnant women.

Exanthem subitum (roseola) is caused by human herpesvirus 6, or less commonly by the closely related human herpesvirus 7, and is most common among children <3 years of age **(Chap. 200)**. As in erythema infectiosum, the rash usually appears after fever has subsided. It consists of 2- to 3-mm rose-pink macules and papules that coalesce only rarely, occur initially on the trunk **(Fig. A1-5)** and sometimes on the extremities (sparing the face), and fade within 2 days.

Although drug reactions have many manifestations, including urticaria, exanthematous *drug-induced eruptions* **(Chap. 63) (Fig. A1-7)** are most common and are often difficult to distinguish from viral exanthems. Eruptions elicited by drugs are usually more intensely erythematous and pruritic than viral exanthems, but this distinction is not reliable. A history of new medications and an absence of prostration may help to distinguish a drug-related rash from an eruption of another etiology. Rashes may persist for up to 2 weeks after administration of the offending agent is discontinued. Certain populations are more prone than others to drug rashes. Of people living with HIV, 50–60% develop a rash in response to sulfa drugs.

Rickettsial illnesses **(Chap. 192)** should be considered in the evaluation of individuals with centrally distributed maculopapular eruptions. The usual setting for *epidemic typhus* is a site of war or natural disaster in which people are exposed to body lice. Endemic typhus or *leptospirosis* (the latter caused by a spirochete) **(Chap. 189)** may be seen in urban environments where rodents proliferate. Outside the United States, other rickettsial diseases cause a spotted-fever syndrome and should be considered in residents of or travelers to endemic areas. Similarly, *typhoid fever*, a nonrickettsial disease caused by *Salmonella typhi* **(Chap. 171) (Fig. A1-9)**, is usually acquired during travel outside the United States. *Dengue fever* **(Fig. A1-53)**, caused by a mosquito-transmitted flavivirus, occurs in tropical and subtropical regions of the world **(Chap. 215)**.

Some centrally distributed maculopapular eruptions have distinctive features. Erythema migrans **(Fig. A1-8)**, the rash of *Lyme disease*

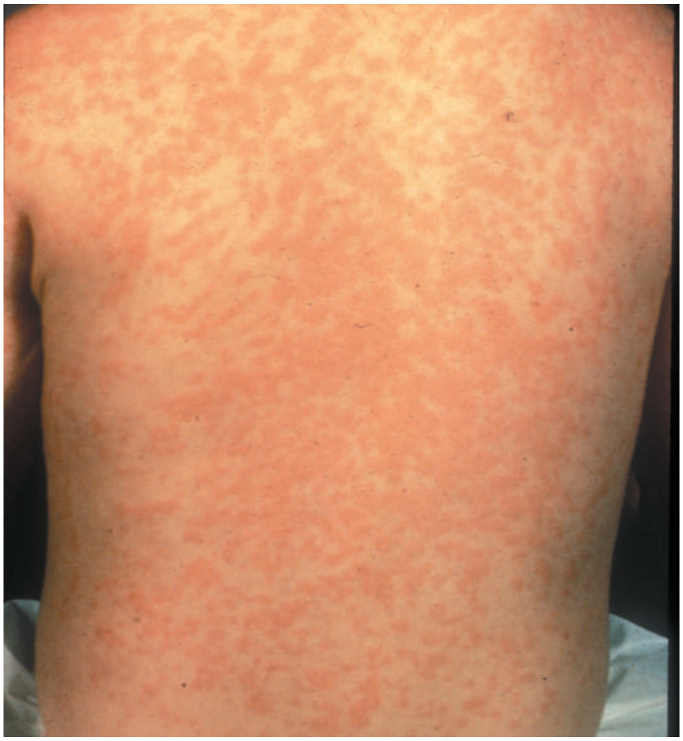

FIGURE 21-1 Centrally distributed, maculopapular eruption on the trunk in a patient with measles. *(From EJ Mayeaux Jr et al: Measles, in Usatine RP et al [eds]: Color Atlas and Synopsis of Family Medicine, 3rd ed. New York, McGraw-Hill, 2019, p. 797, Figure 132-2. Reproduced with permission from Richard P. Usatine, MD.)*

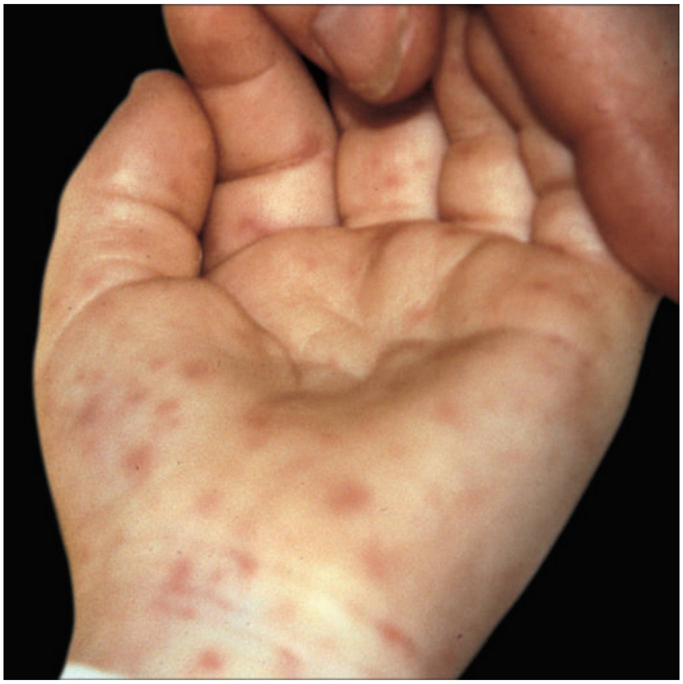

FIGURE 21-2 Peripheral eruption on the wrist and palm exhibiting erythematous macules in the process of evolving into petechial lesions in a patient with Rocky Mountain spotted fever. *(From K Wolff et al [eds]: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 8th ed. New York, McGraw-Hill, 2017, p. 562, Figure 25-50; with permission.)*

(Chap. 191), typically manifests as single or multiple annular lesions. Untreated erythema migrans lesions usually fade within a month but may persist for more than a year. *Southern tick-associated rash illness* (STARI) **(Chap. 191)** has an erythema migrans–like rash but is less severe than Lyme disease and often occurs in regions where Lyme is not endemic. Erythema marginatum, the rash of *acute rheumatic fever* **(Chap. 371)**, has a distinctive pattern of enlarging and shifting transient annular lesions.

Collagen vascular diseases may cause fever and rash. Patients with *systemic lupus erythematosus* **(Chap. 368)** typically develop a sharply defined, erythematous eruption in a butterfly distribution on the cheeks (malar rash) **(Fig. A1-10)** as well as many other skin manifestations **(Figs. A1-11, A1-12)**. *Still's disease* presents as an evanescent, salmon-colored rash on the trunk and proximal extremities that coincides with fever spikes **(Fig. A1-13)**.

Hemophagocytic lymphohistiocytosis may be familial or triggered by infection, autoimmunity, or neoplasia. Cutaneous manifestations are protean and can present as an erythematous maculopapular eruption, pyoderma gangrenosum, purpura, panniculitis, or Stevens-Johnson syndrome.

Zika virus is a mosquito-transmitted flavivirus that is associated with severe birth defects **(Chap. 215)**. Zika is widespread among tropical and subtropical regions of the world. The eruption of Zika virus infection **(Fig. A1-51*A*, A1-51*B*)** is typically pruritic and often accompanied by conjunctival injection **(Fig. A1-51*C*)**.

■ PERIPHERAL ERUPTIONS

These rashes are alike in that they are most prominent peripherally or begin in peripheral (acral) areas before spreading centripetally. Early diagnosis and therapy are critical in *Rocky Mountain spotted fever* **(Chap. 192)** because of its grave prognosis if untreated. Lesions **(Fig. 21-2; see also Fig. A1-16)** evolve from macular to petechial, start on the wrists and ankles, spread centripetally, and appear on the palms and soles only later in the disease. The rash of *secondary syphilis* **(Chap. 187)**, which may be generalized **(Fig. A1-18)** but is prominent on the palms and soles **(Fig. A1-19)**, should be considered in the differential diagnosis of pityriasis rosea, especially in sexually active patients. *Chikungunya fever* **(Chap. 215)**, which is transmitted by mosquito bite in tropical and subtropical regions, is associated with a maculopapular eruption **(Fig. A1-54)** and severe polyarticular small-joint arthralgias. *Hand-foot-and-mouth disease* **(Chap. 210)**, most commonly caused by coxsackievirus A16 or enterovirus 71, is distinguished by tender vesicles distributed on the hands and feet and in the mouth **(Fig. A1-22)**; coxsackievirus A6 causes an atypical syndrome with more extensive lesions. The classic target lesions of *erythema multiforme* **(Fig. A1-24*A*)** appear symmetrically on the elbows, knees, palms, soles, and face. In severe cases, these lesions spread diffusely and involve mucosal surfaces **(Fig. A1-24*B, C*)**. Lesions may develop on the hands and feet in *endocarditis* **(Fig. A1-23) (Chap. 133)**. Pernio, tender violaceous lesions that are acral **(Fig. A1-57)**, occur most commonly on the feet in asymptomatic or mild COVID-19. Vesicles, urticaria, or maculopapular eruptions, often pruritic, may occur on the trunk and extremities in moderate or severe disease, whereas retiform purpura occurs on the extremities and buttocks in severe COVID-19.

■ CONFLUENT DESQUAMATIVE ERYTHEMAS

These eruptions consist of diffuse erythema frequently followed by desquamation. The eruptions caused by group A *Streptococcus* or *Staphylococcus aureus* are toxin-mediated. *Scarlet fever* **(Chap. 153) (Fig. A1-25)** usually follows pharyngitis; patients have a facial flush, a "strawberry" tongue, and accentuated petechiae in body folds (Pastia's lines). *Kawasaki disease* **(Fig. A1-29) (Chaps. 61 and 375)** presents in the pediatric population as fissuring of the lips, a strawberry tongue, conjunctivitis, adenopathy, and sometimes cardiac abnormalities. *Streptococcal toxic shock syndrome* **(Chap. 153)** manifests with hypotension, multiorgan failure, and, often, a severe group A streptococcal infection (e.g., necrotizing fasciitis). *Staphylococcal toxic shock syndrome* **(Chap. 152)** also presents with hypotension and multiorgan failure, but usually only *S. aureus* colonization—not a severe *S. aureus* infection—is documented. *Staphylococcal scalded-skin syndrome* **(Fig. A1-28) (Chap. 152)** is seen primarily in children and in immunocompromised adults. Generalized erythema is often evident during the prodrome of fever and malaise; profound tenderness of the skin is distinctive. In the exfoliative stage, the skin can be induced to form bullae with light lateral pressure (Nikolsky's sign) **(Fig. A1-28*B*)**. In a mild form, a scarlatiniform eruption mimics scarlet fever, but the patient

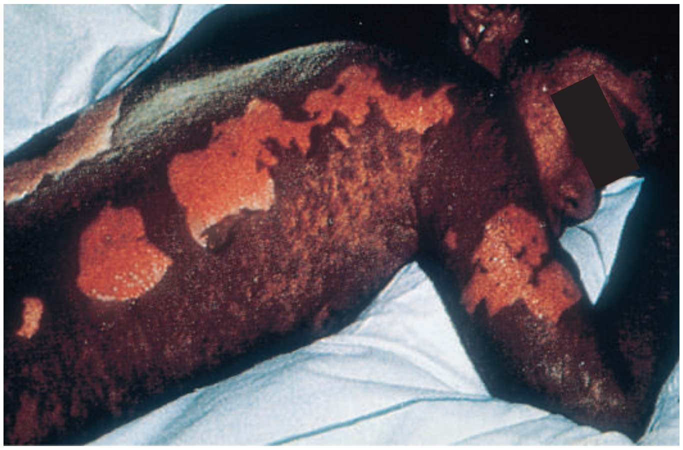

FIGURE 21-3 Confluent desquamation in a patient with toxic epidermal necrolysis. *(From KS-M Kane et al: Color Atlas & Synopsis of Pediatric Dermatology, 3rd ed. New York, McGraw Hill, 2017, Figure 15-6; with permission.)*

does not exhibit a strawberry tongue or circumoral pallor. In contrast to the staphylococcal scalded-skin syndrome, in which the cleavage plane is superficial in the epidermis, *toxic epidermal necrolysis* **(Chap. 63)**, a maximal variant of *Stevens-Johnson syndrome*, involves sloughing of the entire epidermis **(Fig. 21-3, see also Fig. A1-26)**, resulting in severe disease. *Exfoliative erythroderma syndrome* **(Chaps. 61 and 63)** is a serious reaction associated with systemic toxicity that is often due to eczema, psoriasis **(Fig. A1-27)**, a drug reaction, or mycosis fungoides. *Drug reaction with eosinophilia and systemic symptoms (DRESS)* (also termed *drug-induced hypersensitivity syndrome [DIHS]*), often due to antiepileptic or antibiotic agents **(Chap. 63)**, initially appears similar to an exanthematous drug reaction **(Fig. A1-48)** but may progress to exfoliative erythroderma; it is accompanied by multiorgan failure and has an associated mortality rate of ~10%.

■ VESICULOBULLOUS OR PUSTULAR ERUPTIONS

Varicella **(Chap. 198)** is highly contagious, often occurring in winter or spring, and is characterized by pruritic lesions that, within a given region of the body, are in different stages of development at any point in time **(Fig. 21-4; see also Fig. A1-30)**. In immunocompromised hosts, varicella vesicles may lack the characteristic erythematous base or may appear hemorrhagic. Lesions of *Pseudomonas "hot-tub" folliculitis* **(Chap. 170)** are also pruritic and may appear similar to those of varicella **(Fig. A1-55)**. However, hot-tub folliculitis generally occurs in outbreaks after bathing in hot tubs or swimming pools, and lesions occur in regions occluded by bathing suits. Lesions of *variola (smallpox)* **(Chap. S4)** also appear similar to those of varicella but are all at the same stage of development in a given region of the body **(Figs. A1-50*B* and A1-50*C*)**. Variola lesions are most prominent on the face **(Fig. A1-50*A*)** and extremities, while varicella lesions are most prominent on the trunk. *Mpox*, endemic to Africa, can present similarly to variola, although with lower mortality. A 2022 Mpox outbreak in nonendemic nations was characterized by presence of fewer lesions **(Fig. A1-59)** and milder disease. *Herpes simplex virus infection* **(Chap. 197)** is characterized by hallmark grouped vesicles on an erythematous base. Primary herpes infection **(Fig. A1-58*A*, *B*)** is accompanied by fever and toxicity, while recurrent disease **(Fig. A1-58C)** is milder. *Rickettsialpox* **(Chap. 192)** is often documented in urban settings and is characterized by vesicles followed by pustules **(Figs. A1-33*B*, A1-33*C*)**. It can be distinguished from varicella by an eschar at the site of the mouse-mite bite **(Fig. A1-33*A*)** and the papule/plaque base of each vesicle. *Acute generalized exanthematous pustulosis* **(Fig. A1-49)** should be considered in individuals who are acutely febrile and are taking new medications, especially anticonvulsant or antimicrobial agents **(Chap. 63)**. Disseminated *Vibrio vulnificus* infection **(Chap. 173)** or *ecthyma gangrenosum* due to *Pseudomonas aeruginosa* **(Fig. A1-34) (Chap. 170)** should be considered in immunosuppressed individuals with sepsis and hemorrhagic bullae. In children, *reactive infectious mucocutaneous eruption (RIME, encompasses MIRM)* can occur with *Mycoplasma pneumoniae* or other respiratory pathogen infection **(Fig. A1-56)** and is characterized by a sparse, often vesiculobullous eruption with prominent oral, ocular, or urogenital mucositis.

■ URTICARIA-LIKE ERUPTIONS

Individuals with classic urticaria ("hives") **(Fig. 21-5; see also Fig. A1-35)** usually have a hypersensitivity reaction without associated fever. In the presence of fever, urticaria-like eruptions are most often due to *urticarial vasculitis* **(Chap. 375)**. Unlike individual lesions of classic urticaria, which last up to 24 h, these lesions may last 3–5 days. Etiologies include serum sickness (often induced by drugs such as penicillins, sulfas, salicylates, or

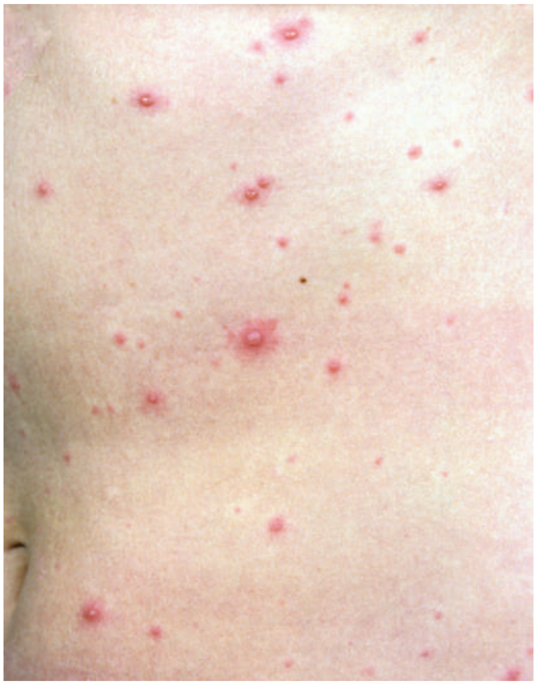

FIGURE 21-4 Vesicular and pustular lesions on the chest in a patient with varicella. *(From K Wolff et al [eds]: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 8th ed. New York, McGraw-Hill, 2017, p. 695, Figure 27-48; with permission.)*

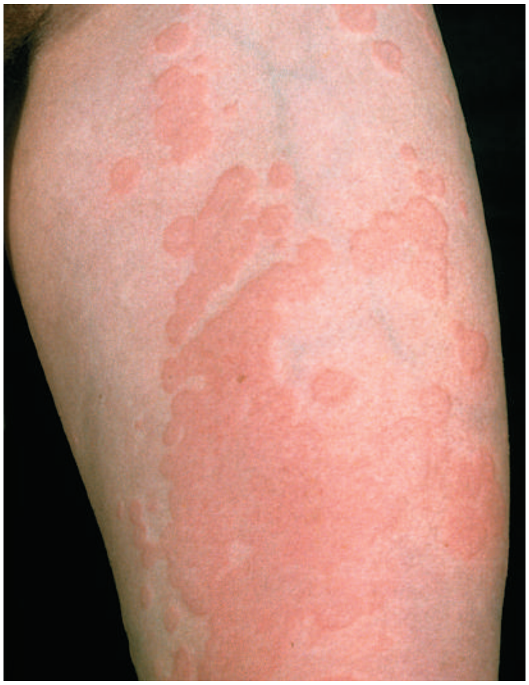

FIGURE 21-5 Urticarial eruption. *(From K Wolff et al [eds]: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 8th ed. New York, McGraw-Hill, 2017, p. 299, Figure 14-2; with permission.)*

barbiturates), connective-tissue disease (e.g., systemic lupus erythematosus or Sjögren's syndrome), and infection (e.g., with hepatitis B virus, enteroviruses, or parasites). Malignancy, especially lymphoma, may be associated with fever and chronic urticaria **(Chap. 61)**.

■ NODULAR ERUPTIONS

In immunocompromised hosts, nodular lesions often represent disseminated infection. Patients with disseminated *candidiasis* **(Fig. A1-37)** (often due to *Candida tropicalis*) may have a triad of fever, myalgias, and eruptive nodules **(Chap. 222)**. Disseminated *cryptococcosis* lesions **(Fig. 21-6; see also Fig. A1-36) (Chap. 221)** may resemble molluscum contagiosum **(Chap. 201)**. Necrosis of nodules should raise the suspicion of *aspergillosis* **(Fig. A1-38) (Chap. 223)** or *mucormycosis* **(Chap. 224)**. *Erythema nodosum* presents with exquisitely tender nodules on the lower extremities **(Fig. A1-39)**. *Sweet syndrome* **(Chap. 61)** should be considered in individuals with multiple nodules and plaques, often so edematous **(Fig. A1-40)** that they give the appearance of vesicles or bullae. Sweet syndrome may occur in individuals with infection, inflammatory bowel disease, or malignancy and can also be induced by drugs.

■ PURPURIC ERUPTIONS

Acute meningococcemia **(Chap. 160)** classically presents in children as a petechial eruption, but initial lesions may appear as blanchable macules or urticaria. Rocky Mountain spotted fever should be considered in the differential diagnosis of acute meningococcemia. *Echovirus 9 infection* **(Chap. 210)** may mimic acute meningococcemia; patients should be treated as if they have bacterial sepsis because prompt differentiation of these conditions may be impossible. Large ecchymotic areas of *purpura fulminans* **(Fig. A1-41) (Chaps. 160 and 315)** reflect severe underlying disseminated intravascular coagulation, which may be due to infectious or noninfectious causes. The lesions of *chronic meningococcemia* **(Fig. A1-42) (Chap. 160)** may have a variety of morphologies, including petechial. Purpuric nodules may develop on the legs and resemble erythema nodosum but lack its exquisite tenderness. Lesions of *disseminated gonococcemia* **(Chap. 161)** are distinctive, sparse, countable hemorrhagic pustules **(Fig. A1-43)**, usually located near joints. The lesions of chronic meningococcemia and those of gonococcemia may be indistinguishable in terms of appearance and distribution. *Viral hemorrhagic fever* **(Chaps. 215 and 216)** should be considered in patients with an appropriate travel history and a petechial rash. *Thrombotic thrombocytopenic purpura* **(Chaps. 61, 105, and 120)** and *hemolytic-uremic syndrome* **(Chaps. 120, 166, and 172)** are closely related and are noninfectious causes of fever and petechiae. *Cutaneous small-vessel vasculitis* (*leukocytoclastic vasculitis*) typically manifests as palpable purpura **(Fig. 21-7, see also Fig. A1-44)** and has a wide variety of causes **(Chap. 61)**.

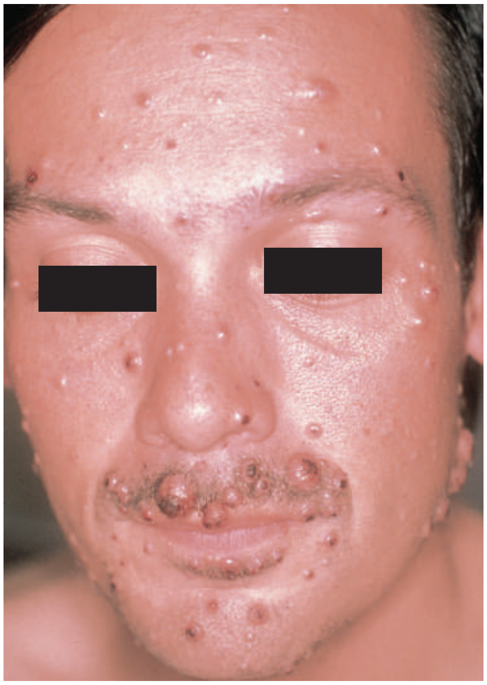
[redacted]
[redacted]

FIGURE 21-6 Nodular eruption on the face due to disseminated Cryptococcus in a patient living with HIV. *(From K Wolff et al [eds]: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 8th ed. New York, McGraw-Hill, 2017, p. 641, Figure 26-57. Used with permission from Loïc Vallant, MD.)*

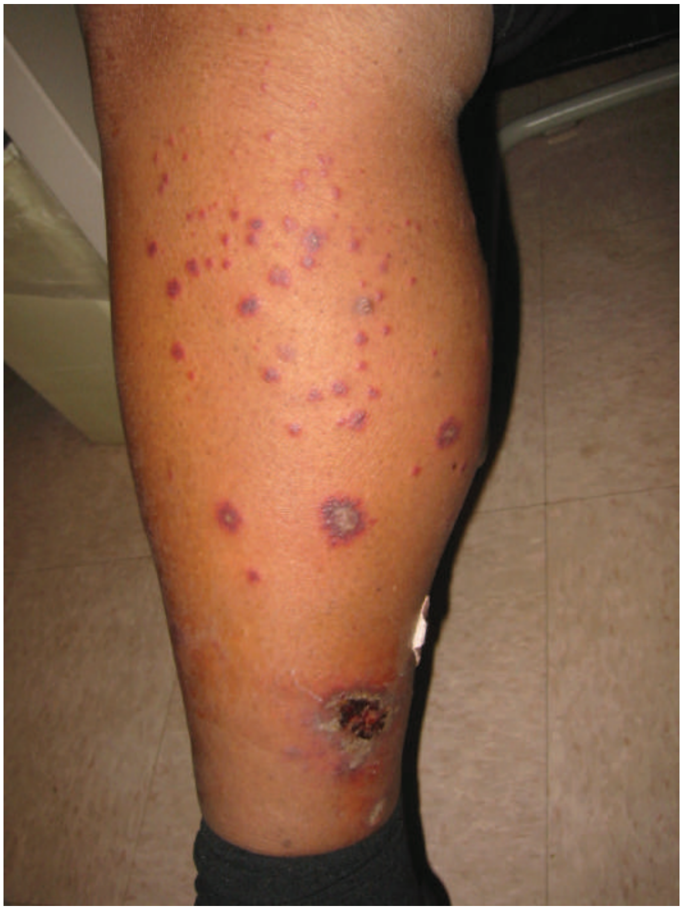

FIGURE 21-7 Purpuric lesions of cutaneous small vessel vasculitis in a patient with Henoch-Schonlein purpura. *(Courtesy of Peter Lio, MD; with permission)*

■ ERUPTIONS WITH ULCERS OR ESCHARS

The presence of an ulcer or eschar **(Fig. 21-8)** in the setting of a more widespread eruption can provide an important diagnostic clue. For

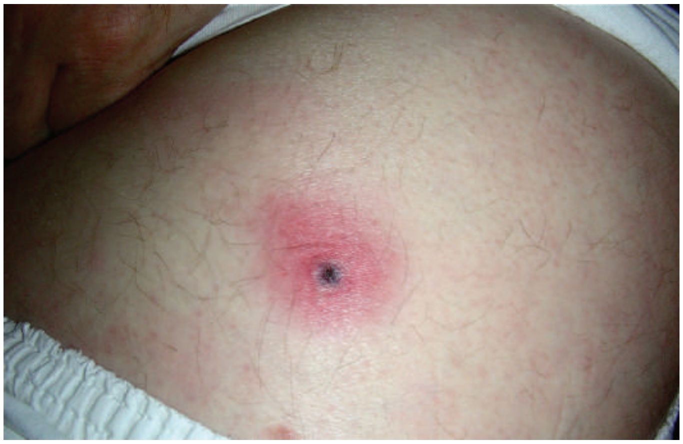

FIGURE 21-8 Eschar with surrounding erythema at the site of a tick bite in a patient with African tick-bite fever. *(From K Wolff et al [eds]: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 8th ed. New York, McGraw-Hill, 2017, p. 561, Figure 25-49; with permission.)*

example, an eschar may suggest the diagnosis of *scrub typhus* or *rickettsialpox* **(Fig. A1-33A) (Chap. 192)** in the appropriate setting. In other illnesses (e.g., anthrax) **(Fig. A1-52) (Chap. S4)**, an ulcer or eschar may be the only skin manifestation.

FURTHER READING

Cherry JD: Cutaneous manifestations of systemic infections, in *Feigin and Cherry's Textbook of Pediatric Infectious Diseases*, 8th ed. JD Cherry et al (eds). Philadelphia, Elsevier, 2019, pp 539–559.

Juliano JJ et al: The acutely ill patient with fever and rash, in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, vol 1, 10th ed. MJ Blaser et al (eds). Philadelphia, Elsevier, 2025.

Kang S et al (eds): *Fitzpatrick's Dermatology*, 9th ed. New York, McGraw-Hill, 2019.

Saavedra AP et al: *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, 9th ed. New York, McGraw Hill, 2023.

22

Fever of Unknown Origin

Catharina M. Mulders-Manders,
Chantal P. Rovers

DEFINITION

Clinicians commonly refer to any febrile illness without an initially obvious etiology as *fever of unknown origin* (FUO). Most febrile illnesses either resolve before a diagnosis can be made or develop distinguishing characteristics that lead to a diagnosis. The term *FUO* should be reserved for prolonged febrile illnesses without an established etiology despite intensive evaluation and diagnostic testing. This chapter focuses on FUO in the adult patient.

FUO was originally defined by Petersdorf and Beeson in 1961 as an illness of at least 3 weeks' duration with fever of ≥38.3°C (≥101°F) on two occasions and an uncertain diagnosis despite 1 week of inpatient evaluation. Since then, several modifications of the definition have been proposed to better reflect outpatient-based health care and to exclude immunocompromised patients, who require a different and more aggressive approach. To reduce heterogeneity between cohorts, inclusion of qualitative criteria of minimal diagnostic testing are necessary. Accordingly, FUO is now defined as:

1. Fever ≥38.3°C (≥101°F) on at least two occasions
2. Illness duration of at least 3 weeks
3. No known immunocompromised state
4. Diagnosis that remains uncertain after a thorough history-taking, physical examination, and the following obligatory investigations: determination of erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) level; platelet count; leukocyte count and differential; measurement of levels of hemoglobin, electrolytes, creatinine, total protein, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, lactate dehydrogenase, creatine kinase, ferritin, antinuclear antibodies, and rheumatoid factor; protein electrophoresis; urinalysis; blood cultures ($n = 3$); urine culture; chest x-ray; abdominal ultrasonography; and tuberculin skin test (TST) or interferon γ release assay (IGRA)

Closely related to FUO is *inflammation of unknown origin (IUO)*, which has the same definition as FUO, except for the body temperature criterion: IUO is defined as the presence of elevated inflammatory parameters (CRP or ESR) on multiple occasions for a period of at least 3 weeks in an immunocompetent patient with normal body temperature, for which a final explanation is lacking despite history-taking, physical examination, and the obligatory tests listed above. It has been shown in multiple cohorts that the causes for IUO and FUO are the same and that their workup should be identical. Therefore, for convenience, the term FUO will refer to both FUO and IUO within the remainder of this chapter.

ETIOLOGY AND EPIDEMIOLOGY

The prevalence of FUO is largely unknown and will likely vary between different geographic regions and different health care settings. In the 1990s, Japan reported a prevalence of 2.9% in admitted patients, and a recent Danish registry study reported 6220 patients in a 10-year time frame.

Table 22-1 summarizes the findings of large studies on FUO conducted over the past 20 years.

Since the 1960s, the range of FUO etiologies has evolved as a result of changes in the definition, local disease epidemiology, the widespread use of antibiotics, and especially the availability of new diagnostic techniques. The proportion of cases caused by intraabdominal abscesses and tumors, for example, has decreased because of earlier detection by CT and ultrasound. In addition, infective endocarditis is a less frequent cause because blood culture and echocardiographic techniques have improved. Conversely, some diagnoses, such as acute HIV infection or autoinflammatory diseases, were unknown six decades ago.

Data on the outcome of FUO are difficult to compare between studies because of variability in patient selection and variability in workup between cohorts. Study outcome strongly depends on study design and the origin of the cohort. The chance of a final diagnosis is higher when (older) time-based criteria are used versus when (newer) criteria including a standard basic workup are used for patient selection. The chance of remaining without a final diagnosis is two to five times higher in patients from Europe compared to patients from Asia.

TABLE 22-1 Etiology of FUO: Pooled Results of Large Studies Published in the Past 20 Years (2003–2023)

GEOGRAPHIC AREA	NO. OF COHORTS (INCLUSION PERIOD)	NO. OF PATIENTS	INFECTIONS, MEDIAN % (RANGE)	NONINFECTIOUS INFLAMMATORY DISEASES, MEDIAN % (RANGE)	MALIGNANCY, MEDIAN % (RANGE)	MISCELLANEOUS, MEDIAN % (RANGE)	NO DIAGNOSIS, MEDIAN % (RANGE)
Western Europe	12 (1995–2020)	2100	15.5 (4–36)	25 (17–33)	11 (3–30)	7.5 (0–16)	39.5 (26–54)
Other European and Turkey	15 (1984–2019)	1615	42 (26–74)	23 (12–38)	14 (4–19)	5 (2–18)	16 (0–35)
Middle East	3 (2009–2010)[a]	1289	66 (42–79)	15 (7–17)	7 (1–30)	1 (0–12)	8 (2–12)
Asia	39 (1994–2021)[a]	7191	42 (3–58)	19 (7–57)	13 (6–23)	6.5 (0–15)	17 (0–81)

Note: No studies from the Americas, Africa, or Oceania have been reported. Studies aimed at diagnostic methods were excluded.

[a] Some studies did not report an inclusion period.

For references, see supplementary material at *accessmedicine.com/harrisons*.

Roughly comparable to 60 years ago, infections remain the most common cause of FUO in non-Western cohorts. Compared to Europe, the risk of having an infection is over four times higher in patients from Southern Asia and three times higher in far East Asia. Up to half of all infections in patients with FUO outside Western nations are caused by *Mycobacterium tuberculosis*, which is a less common cause in Western Europe and probably also in the United States. Recent data on FUO from the latter, however, have not been reported. In Western cohorts, noninfectious inflammatory diseases (NIIDs), including autoimmune, autoinflammatory, and granulomatous diseases, as well as vasculitides, form the most common cause of FUO. Up to one-third of Western patients with FUO have a diagnosis that falls within the category of NIIDs. The number of FUO patients diagnosed with NIIDs probably will not decrease in the near future, as inflammation may precede more typical manifestations or laboratory evidence of these diseases by months. Moreover, NIIDs may only be diagnosed after prolonged observation and exclusion of other diseases, and an increasing number of genetic inflammatory syndromes have been described in more recent years.

Especially in Western cohorts, where the proportion of undiagnosed patients can be up to >50%, the proportion of patients remaining without a final diagnosis is much higher than 60 years ago despite technical advantages. This is called "the FUO paradox" and can be explained by the fact that in patients with fever, a diagnosis is often established before 3 weeks have elapsed because these patients tend to seek medical advice earlier and because diagnostic techniques such as CT, MRI, and positron emission tomography (PET)/CT, are now widely available. Therefore, only the cases that are most difficult to diagnose continue to meet the criteria for FUO. Furthermore, most patients who have FUO without a diagnosis currently do well. A less aggressive diagnostic approach may be used in clinically stable patients once diseases with immediate therapeutic or prognostic consequences have been ruled out.

No clinical parameters or tools that can help identify patients with a high risk of remaining undiagnosed have been identified. Recurrent inflammation (defined as repeated episodes of fever or inflammation interspersed with fever-free intervals of at least 2 weeks and apparent remission of the underlying disease) reduces the chance of finding a final explanation and is associated with a lower, but not zero, prevalence of infections. A symptom duration of >12 months is significantly associated with a lower chance of finding a final explanation, and infections and malignancy are less common, but still a possibility, in patients with such extensive symptom duration. It is important to note that even in patients with symptom duration >3 months, more common infections such as tuberculosis, spondylodiscitis, and endocarditis have been described and that lymphoma has been reported in patients with symptom duration >12 months. Normal PET/CT has also been associated with a lower chance of a final diagnosis. Artificial intelligence, using computer models to predict the final diagnosis in patients with FUO, may be able to shorten diagnostic delay and may guide specific diagnostic testing, but only limited data are available. Its exact value on a larger scale, especially in different geographic regions, is difficult to estimate at this moment.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis for FUO is extensive. It is important to note that patients with FUO most often suffer from an atypical presentation of a more common disease, rather than from a rare disease. Unsurprisingly, the distribution of causes of FUO is influenced by local epidemiology. **Table 22-2** presents an overview of possible causes of FUO. Endocarditis, diverticulitis, vertebral osteomyelitis, and extrapulmonary tuberculosis are the more common infectious disease diagnoses. Q fever (*Coxiella burnetii*) and Whipple's disease (*Tropheryma whipplei* infection) are quite rare but should always be kept in mind as causes of FUO since the presenting symptoms can be nonspecific. Serologic testing for Q fever, which results from exposure to animals or animal products, should be performed by immunofluorescence assay (IFA) when the patient lives in a rural area or has a history of heart valve disease, an aortic aneurysm, or a vascular prosthesis. In patients with unexplained symptoms localized to the central nervous system, gastrointestinal tract, or joints, polymerase chain reaction testing for *T. whipplei* should be performed on stool and blood early in the diagnostic process. Travel to or (former) residence in tropical countries or the American Southwest should lead to consideration of endemic infectious diseases such as malaria, leishmaniasis, histoplasmosis, or coccidioidomycosis. Fever with signs of endocarditis and negative blood culture results poses a special problem. Culture-negative endocarditis **(Chap. 133)** may be due to difficult-to-culture bacteria such as nutritionally variant bacteria, HACEK organisms (including *Haemophilus parainfluenzae*, *H. paraphrophilus*, *Aggregatibacter actinomycetemcomitans*, *A. aphrophilus*, *A. paraphrophilus*, *Cardiobacterium hominis*, *C. valvarum*, *Eikenella corrodens*, and *Kingella kingae*), *Coxiella burnetii*, *T. whipplei*, and *Bartonella* species. Marantic endocarditis is a sterile thrombotic disease that occurs as a paraneoplastic phenomenon, especially with adenocarcinomas. Sterile endocarditis is also seen in the context of systemic lupus erythematosus and antiphospholipid syndrome. The use of next-generation sequencing for rapid detection of infecting microbes may be valuable in patients with FUO, but its exact value and exact application are yet to be determined.

Of the NIIDs, adult-onset Still's disease, large-vessel vasculitis, polymyalgia rheumatica, systemic lupus erythematosus (SLE), and sarcoidosis are rather common diagnoses in patients with FUO. Autoinflammatory syndromes are rare (with the exception of familial Mediterranean fever in specific geographic regions), and most can only be diagnosed on clinical criteria. Schnitzler's syndrome, which can present at any age, can often be diagnosed easily in a patient with FUO who presents with urticaria, bone pain, and monoclonal gammopathy. A new autoinflammatory syndrome called vacuoles, E1 enzyme, X-linked autoinflammatory syndrome (VEXAS) has been described in >300 patients since 2020. VEXAS is an adult-onset multisystem autoinflammatory syndrome with predominance in middle-aged to elderly males and is characterized, among others, by recurrent inflammation, skin lesions, chondritis, lung disease, venous thrombosis, arthritis, and myelodysplastic syndrome. The presence of vacuoles in myeloid precursors in bone marrow biopsy is very suspicious for VEXAS in a patient with inflammatory symptoms. The disease is caused by somatic mosaicism for pathogenic variants in *UBA1*.

Although most malignancies can present with fever, malignant lymphoma is by far the most common diagnosis of FUO among the neoplasms. Sometimes the fever even precedes lymphadenopathy detectable by physical examination.

Apart from drug-induced fever and exercise-induced or benign hyperthermia, none of the miscellaneous causes of fever is found very frequently in patients with FUO. Virtually all drugs can cause fever, even after long-term use. *Drug-induced fever*, including DRESS (drug reaction with eosinophilia and systemic symptoms; now more generally referred to as drug-induced hypersensitivity syndrome [DIHS]; **Fig. A1-48**), is often accompanied by eosinophilia and also by lymphadenopathy, which can be extensive. More common causes of drug-induced fever are allopurinol, carbamazepine, lamotrigine, phenytoin, sulfasalazine, furosemide, and antimicrobial drugs (especially sulfonamides, minocycline, vancomycin, β-lactam antibiotics, and isoniazid). Benign hyperthermia or *exercise-induced hyperthermia* **(Chaps. 20 and 478)** is characterized by an elevated body temperature without an increase in CRP or ESR or other signs of inflammation; these may be postinfectious in origin. Infectious triggers for benign hyperthermia are possibly the same as for chronic fatigue syndrome, such as Epstein-Barr virus, Q-fever, and COVID-19. *Factitious fever* (fever artificially induced by the patient—for example, by IV injection of contaminated water) should be considered in all patients, especially those with easy access to medical equipment. In *fraudulent fever*, the patient is normothermic but manipulates the thermometer. Simultaneous measurements at different body sites (rectum, ear, mouth) should rapidly identify this diagnosis. Another clue to fraudulent fever is dissociation between pulse rate and temperature.

TABLE 22-2 Reported Causes of Fever of Unknown Origin (FUO)[a]

Infections	
Bacterial, nonspecific	Abdominal abscess, adnexitis, aortitis, apical granuloma, appendicitis, bacterial translocation, bronchiectasis, cholangitis, cholecystitis, diverticulitis, endocarditis, endometritis, epididymitis, epidural abscess, infected joint prosthesis, infected vascular catheter, infected vascular prosthesis, infectious arthritis, infective myonecrosis, intracranial abscess, liver abscess, lung abscess, malakoplakia, mastitis, mastoiditis, mediastinitis, muscle abscess, mycotic aneurysm, osteomyelitis, pelvic inflammatory disease, prostatitis, pyelonephritis, pylephlebitis, pyomyoma, renal abscess, septic arthritis, septic phlebitis, sinusitis, spondylodiscitis, xanthogranulomatous urinary tract infection
Bacterial, specific	Actinomycosis, *Bacteroides*, bartonellosis, brucellosis, Bacillus Calmette-Guérin infection, *Burkholderia* spp., *Campylobacter* infection, *Chlamydia pneumoniae* infection, chronic meningococcemia, ehrlichiosis, enterococci, *Escherichia coli*, *Gemella* infection, gonococcus, *Klebsiella* spp., *Lactobacillus*, legionellosis, leptospirosis, listeriosis, louse-borne relapsing fever (*Borrelia recurrentis*), Lyme disease (*Borrelia burgdorferi*), melioidosis (*Burkholderia pseudomallei*), *Mycoplasma* infection, nocardiosis, nontuberculous mycobacteria, *Prevotella* infection, *Pseudomonas* spp., psittacosis, Q fever (*Coxiella burnetii)*, rickettsiosis, *Rhodococcus* spp., *Salmonella* spp., *Spirillum minor* infection, *Sphingobacterium* infection, *Staphylococcus aureus*, *Streptobacillus moniliformis* infection, syphilis, tick-borne relapsing fever (*Borrelia duttonii, Borrelia hermsi*), tuberculosis, tularemia, typhoid fever and other salmonelloses, Whipple disease (*Tropheryma whipplei*), yersiniosis
Fungal	Aspergillosis, blastomycosis, candidiasis, coccidioidomycosis, cryptococcosis, histoplasmosis, *Malassezia furfur* infection, mucormycosis, paracoccidioidomycosis, *Penicillium* spp., *Pneumocystis jirovecii* pneumonia, sporotrichosis, *Talaromyces* infection
Parasitic	Amebiasis, babesiosis, echinococcosis, fascioliasis, malaria, schistosomiasis, strongyloidiasis, toxocariasis, toxoplasmosis, trichinellosis, trypanosomiasis, visceral leishmaniasis
Viral	Colorado tick fever, coxsackievirus infection, cytomegalovirus infection, chikungunya, dengue, Epstein-Barr virus infection, hantavirus infection, hepatitis (A, B, C, D, E), herpes simplex, HIV infection, human herpesvirus 6 infection, parvovirus infection, West Nile virus infection
Noninfectious Inflammatory Diseases	
Systemic rheumatic and autoimmune diseases	Ankylosing spondylitis, antiphospholipid syndrome, autoimmune hemolytic anemia, autoimmune hepatitis, Behçet's disease, cryoglobulinemia, dermatomyositis, Felty syndrome, gout, mixed connective tissue disease, neuromyelitis optica, polymyositis, pseudogout, reactive arthritis, psoriatic arthritis, relapsing polychondritis, rheumatic fever, rheumatoid arthritis, Sjögren's syndrome, systemic lupus erythematosus, systemic sclerosis, Vogt-Koyanagi-Harada syndrome
Vasculitis	Allergic vasculitis, antineutrophil cytoplastic antibody (ANCA) vasculitis, Cogan's syndrome, eosinophilic granulomatosis with polyangiitis, giant cell vasculitis/polymyalgia rheumatica, granulomatosis with polyangiitis, hypersensitivity vasculitis, IgA vasculitis, Kawasaki disease, polyarteritis nodosa, Takayasu arteritis, urticarial vasculitis
Granulomatous diseases	Idiopathic granulomatous hepatitis, sarcoidosis
Autoinflammatory syndromes	Blau syndrome, CAPS[b] (cryopyrin-associated periodic syndromes), Crohn's disease, DIRA (deficiency of the interleukin 1 receptor antagonist), deficiency of adenosine deaminase 2 (DADA2), familial Mediterranean fever, hyper-IgD syndrome (HIDS, due to mevalonate kinase deficiency), juvenile idiopathic arthritis, macrophage activation syndrome, NLRP12-related disease, PAPA syndrome (pyogenic sterile arthritis, pyoderma gangrenosum, and acne), PFAPA syndrome (periodic fever, aphthous stomatitis, pharyngitis, adenitis), recurrent idiopathic pericarditis, SAPHO (synovitis, acne, pustulosis, hyperostosis, osteomyelitis), Schnitzler's syndrome, Still's disease, TRAPS (tumor necrosis factor receptor–associated periodic syndrome), VEXAS (vacuoles, E1 enzyme, X-linked autoinflammatory syndrome)
Neoplasms	
Hematologic malignancies	Amyloidosis, angioimmunoblastic lymphoma, Castleman's disease, hemophagocytic syndrome (hemophagocytic lymphohistiocytosis), Hodgkin's disease, hypereosinophilic syndrome, leukemia, lymphomatoid granulomatosis, malignant histiocytosis, multiple myeloma, myelodysplastic syndrome, myelofibrosis, non-Hodgkin's lymphoma, plasmacytoma, systemic mastocytosis, vaso-occlusive crisis in sickle cell disease
Solid tumors	Most solid tumors and metastases can cause fever. Those most common causing FUO are breast, colon, hepatocellular, lung, pancreatic, and renal cell carcinomas.
Benign tumors	Angiomyolipoma, cavernous hemangioma of the liver, craniopharyngioma, necrosis of dermoid tumor in Gardner's syndrome, tumoral calcinosis
Miscellaneous Causes	
	ADEM (acute disseminated encephalomyelitis), adrenal insufficiency, alcoholic steatohepatitis, allergic purpura, aneurysms, anal fistula, anti-NMDA encephalitis, antisynthetase syndrome, anomalous thoracic duct, aortic dissection, aortic-enteral fistula, aseptic meningitis (Mollaret's syndrome), atrial myxoma, brewer's yeast ingestion, calcium pyrophosphate deposition, Caroli's disease, cholesterol emboli, cirrhosis, complex partial status epilepticus, crowned dense syndrome, cyclic neutropenia, cryptogenic organizing pneumonia, drug fever, Erdheim-Chester disease, extrinsic allergic alveolitis, Fabry's disease, factitious disease, fire-eater's lung, fraudulent fever, Dressler's syndrome, ganglioneuroma, extrinsic allergic alveolitis, Gaucher's disease, hemolytic uremic syndrome, Hamman-Rich syndrome (acute interstitial pneumonia), Hashimoto's encephalopathy, hematoma, heparin-induced thrombocytopenia, histiocytic necrotizing lymphadenitis, hypersensitivity pneumonitis, hypertriglyceridemia, hypogammaglobulinemia/subclass deficiency, hypothalamic hypopituitarism, idiopathic inflammatory myopathy, idiopathic normal-pressure hydrocephalus, IgG4 disease, inflammatory pseudotumor, interstitial nephritis, Kikuchi's disease, limbic encephalitis, linear IgA dermatosis, mesenteric fibromatosis, metal fume fever, microaspiration, milk protein allergy, Mollaret's meningitis, myotonic dystrophy, nonbacterial osteitis, obstructive suppurative pancreatic ductitis, organic dust toxic syndrome, panniculitis, POEMS (polyneuropathy, organomegaly, endocrinopathy, monoclonal protein, skin changes), obstructive sleep apnea syndrome, organizing pneumonia, polymer fume fever, post–cardiac injury syndrome, primary biliary cirrhosis, primary hyperparathyroidism, pseudomembranous colitis, pulmonary embolism, pulmonary nodular lymphoid hyperplasia, pyoderma gangrenosum, recurrent large bowel ischemia, retroperitoneal fibrosis, Rosai-Dorfman disease, sclerosing mesenteritis, silicone embolization, subacute thyroiditis (de Quervain's), Sweet's syndrome (acute febrile neutrophilic dermatosis), thrombosis, TAFRO syndrome (thrombocytopenia, anasarca, fever, reticulin fibrosis, and organomegaly) tubulointerstitial nephritis and uveitis syndrome (TINU), ulcerative colitis, XIAP deficiency (X-linked inhibitor of apoptosis deficiency)
Thermoregulatory Disorders	
Central	Brain tumor, cerebrovascular accident, encephalitis, hypothalamic dysfunction
Peripheral	Anhidrotic ectodermal dysplasia, benign hyperthermia, exercise-induced hyperthermia, heat stroke, hyperthyroidism, paroxysmal sympathetic hyperactivity, pheochromocytoma

[a] This table includes causes of FUO that have been described in the literature. [b] CAPS includes chronic infantile neurologic cutaneous and articular syndrome (CINCA, also known as neonatal-onset multisystem inflammatory disease, or NOMID), familial cold autoinflammatory syndrome (FCAS), and Muckle-Wells syndrome.

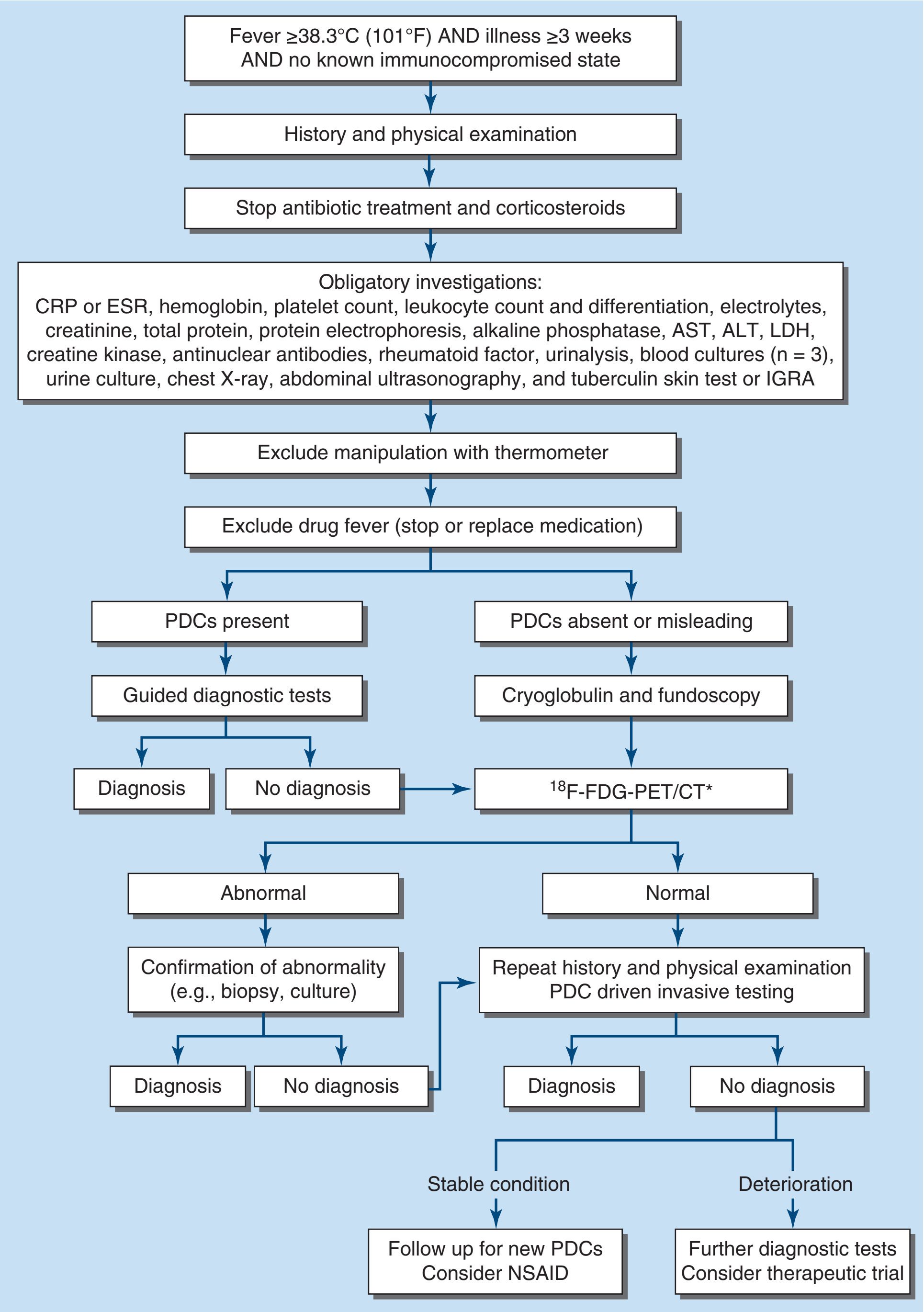
Fever ≥38.3°C (101°F) AND illness ≥3 weeks AND no known immunocompromised state
History and physical examination
Stop antibiotic treatment and corticosteroids
Obligatory investigations: CRP or ESR, hemoglobin, platelet count, leukocyte count and differentiation, electrolytes, creatinine, total protein, protein electrophoresis, alkaline phosphatase, AST, ALT, LDH, creatine kinase, antinuclear antibodies, rheumatoid factor, urinalysis, blood cultures (n = 3), urine culture, chest X-ray, abdominal ultrasonography, and tuberculin skin test or IGRA
Exclude manipulation with thermometer
Exclude drug fever (stop or replace medication)
PDCs present
PDCs absent or misleading
Guided diagnostic tests
Cryoglobulin and fundoscopy
Diagnosis
No diagnosis
18F-FDG-PET/CT*
Abnormal
Normal
Confirmation of abnormality (e.g., biopsy, culture)
Repeat history and physical examination PDC driven invasive testing
Diagnosis
No diagnosis
Diagnosis
No diagnosis
Stable condition
Deterioration
Follow up for new PDCs Consider NSAID
Further diagnostic tests Consider therapeutic trial

FIGURE 22-1 Structured approach to patients with FUO. ALT, alanine aminotransferase; AST, aspartate aminotransferase; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; ^{18}F-FDG-PET/CT, ^{18}F-fluorodeoxyglucose positron emission tomography combined with low-dose CT; IGRA, interferon γ release assay; LDH, lactate dehydrogenase; NSAID, nonsteroidal anti-inflammatory drug; PDCs, potentially diagnostic clues (all localizing signs, symptoms, and abnormalities potentially pointing toward a diagnosis). *If not available, gallium scintigraphy or labeled leukocyte scintigraphy is an acceptable alternative.

APPROACH TO THE PATIENT

A Structured Diagnostic Approach

Figure 22-1 shows a structured approach to patients presenting with FUO. The most important step in the diagnostic workup is the search for potentially diagnostic clues (PDCs) through complete and repeated history-taking and physical examination and the obligatory investigations listed above and in the figure. PDCs are defined as all localizing signs, symptoms, and abnormalities potentially pointing toward a diagnosis. Although PDCs are often misleading, only with their help can a concise list of probable diagnoses be made. The history should include information about the fever pattern (continuous or recurrent) and duration, previous medical history, present and recent drug use, family history, sexual history, country of origin, recent and remote travel, environmental exposures associated with travel or hobbies, and animal contacts. A complete physical examination should be performed, with special attention to the eyes, lymph nodes, temporal arteries, liver, spleen, sites of previous surgery, entire skin surface, and mucous membranes. Before further diagnostic tests are initiated, antibiotic and glucocorticoid treatment, which can mask many diseases, should be stopped. For example, sterile blood and other cultures are not completely reliable when samples are obtained during antibiotic treatment, and the size of enlarged lymph nodes usually decreases during glucocorticoid treatment, regardless of the cause of lymphadenopathy. Despite the high percentage of false-positive ultrasounds and the relatively low sensitivity of chest x-rays, the performance of these simple, low-cost diagnostic tests remains obligatory in all patients with FUO in order to separate cases that are caused by easily diagnosed diseases from those that are not. Abdominal ultrasound and conventional chest radiography have been preferred to thoracic and abdominal CT as an obligatory test because of their relatively low cost, lower radiation burden, and absence of side effects.

After identification of all PDCs retrieved from the history, physical examination, and obligatory tests, a limited list of the most probable diagnoses should be made. Since most investigations are helpful only for patients who have PDCs for the diagnoses sought, further diagnostic procedures should be limited to specific investigations aimed at confirming or excluding diseases on this list. In FUO, the diagnostic pointers are numerous and diverse but may be missed on initial examination, often being detected only by subsequent careful examinations. In the absence of PDCs, the history and physical examination should be repeated regularly. In patients with recurrent inflammation, the diagnostic workup is the same as for patients with continuous inflammation. Patients should be asked to return during a febrile episode so that the history, physical examination, and laboratory tests can be repeated during a symptomatic phase. Further diagnostic tests should be performed only during an inflammatory episode because abnormalities may be absent between episodes.

One of the first steps should be to rule out benign hyperthermia or factitious or fraudulent fever, particularly in patients without signs of inflammation in laboratory tests. All medications, including nonprescription drugs and nutritional supplements, should be discontinued early in the evaluation to exclude drug fever. If fever persists beyond 72 h after discontinuation of the suspected drug, it is unlikely that this drug is the cause.

Only rarely do biochemical tests (beyond the obligatory tests needed to classify a patient's fever as FUO) lead directly to a definitive diagnosis in the absence of PDCs. The diagnostic yield of immunologic serologies other than those included in the obligatory tests is relatively low. These tests more often yield false-positive rather than true-positive results and are of little use without PDCs pointing to specific disorders. Given the absence of specific symptoms in many patients and the relatively low cost of the test, investigation of cryoglobulins appears to be a valuable screening test in patients with FUO.

Multiple blood samples should be cultured in the laboratory long enough to ensure ample time for any fastidious organisms, such as those of the HACEK group. It is critical to inform the laboratory of the intent to test for unusual organisms. Specialized media should be used when the history suggests uncommon microorganisms, such as *Histoplasma* or *Legionella*. Performing more than three blood cultures or more than one urine culture is useless in patients with FUO in the absence of PDCs (e.g., a high level of clinical suspicion of endocarditis). Repeating blood or urine cultures is useful only when previously cultured samples were collected during antibiotic treatment or within 1 week after its discontinuation. FUO with headache should prompt microbiologic examination of cerebrospinal fluid (CSF) for organisms including herpes simplex virus (especially type 2), *Cryptococcus neoformans* and *C. gattii*, and *Mycobacterium tuberculosis*.

Microbiologic serology should not be included in the diagnostic workup of patients without PDCs for specific infections. A tuberculin skin test (TST) or interferon γ (IFNγ) release assay (IGRA) is included in the obligatory investigations, but it may yield indeterminate results in patients with miliary tuberculosis, with malnutrition, receiving immunosuppression, or with anti-IFNγ autoantibodies. Although the IGRA is less influenced by prior vaccination with bacille Calmette-Guérin (BCG) or by infection with most nontuberculous mycobacteria, its sensitivity is similar to that of the TST; a negative TST or IGRA does not exclude a diagnosis of tuberculosis. Miliary tuberculosis is especially difficult to diagnose. Granulomatous disease in liver or bone marrow biopsy samples, for example, should always lead to a (re)consideration of this diagnosis. If miliary tuberculosis is suspected, liver biopsy for acid-fast smear, culture, and polymerase chain reaction probably still has the highest diagnostic yield; however, biopsies of bone marrow, lymph nodes, or other involved organs also can be considered. Mycobacterial polymerase chain reaction (PCR) testing together with cultures may shorten the time to diagnosis compared to mycobacterial cultures alone.

The diagnostic yield of echocardiography, sinus radiography, radiologic or endoscopic evaluation of the gastrointestinal tract, bone marrow biopsy, liver biopsy, and bronchoscopy is very low in the absence of PDCs.

In patients without PDCs or with only misleading PDCs, fundoscopy by an ophthalmologist may be useful early in the diagnostic workup to exclude retinal vasculitis.

Several studies have shown a high prevalence of giant cell arteritis among patients with FUO, with rates up to 17% among elderly patients. Giant cell arteritis often involves large arteries and, in most cases, can be diagnosed by ^{18}F-fluordeoxyglucose (^{18}F-FDG) positron emission tomography (PET) with combined CT. However, temporal artery biopsy is still recommended for patients ≥55 years. ^{18}F-FDG-PET/CT may be confounded in vasculitis limited to the temporal arteries because of the small diameter of these vessels and the high levels of FDG uptake in the brain.

When PDC-guided diagnostic tests do not lead to a diagnosis, ^{18}F-FDG-PET/CT should be performed, preferably at a time point with elevated ESR or CRP to increase the chance of positive findings. This is especially important in patients with episodic FUO.

^{18}F-FLUORDEOXYGLUCOSE POSITRON EMISSION TOMOGRAPHY

^{18}F-FDG-PET has become an established imaging procedure in FUO. FDG accumulates in tissues with a high rate of glycolysis, which occurs not only in malignant cells but also in activated leukocytes and thus permits the imaging of acute and chronic inflammatory processes. Compared with conventional scintigraphy (see below), ^{18}F-FDG-PET/CT offers the advantages of higher resolution, greater sensitivity in chronic low-grade infections, and a high degree of accuracy in the central skeleton. Furthermore, vascular uptake of FDG is increased in patients with vasculitis **(Fig. 22-2)**. The mechanisms responsible for FDG uptake do not allow differentiation among infection, sterile inflammation, and malignancy, but because all of these can cause FUO, this is an overall advantage. It is important to realize that physiologic uptake or concentration of FDG may obscure pathologic foci in the brain, heart, bowel, kidneys, and bladder. FDG uptake in the heart, which obscures endocarditis, may be prevented by consumption of a low-carbohydrate diet before the PET investigation. FDG uptake in the brain may obscure local temporal large vessel vasculitis. In patients with fever, bone marrow uptake is frequently increased in a nonspecific way due to cytokine activation, which upregulates glucose transporters in bone marrow cells.

In recent years, many cohort studies and several meta-analyses have focused on the diagnostic yield of PET/CT in FUO. These studies are highly variable in terms of the selection of patients, the follow-up, and the selection of a gold-standard reference. Indirect comparisons of test performance suggested that ^{18}F-FDG-PET/CT outperformed gallium scintigraphy and leukocyte scintigraphy. Similarly, indirect comparisons of diagnostic yields suggested that ^{18}F-FDG-PET/CT was more likely than alternative tests to correctly identify the cause of FUO. Meta-analyses report a high diagnostic yield of PET/CT in the workup of FUO patients, with total diagnostic yield (i.e., the proportion of scans helpful in finding the final explanation for the inflammation) of ~50% for PET/CT.

As many patients with FUO present with periodic fever, the correct timing of PET/CT increases its diagnostic value. Few studies on the use of biomarkers, such as elevated CRP or ESR, as contributors to outcome of PET/CT have been performed. When both CRP and ESR are normal at the time of ^{18}F-FDG-PET/CT, the outcome may only be contributory if a patient does have fever at the time of the scan.

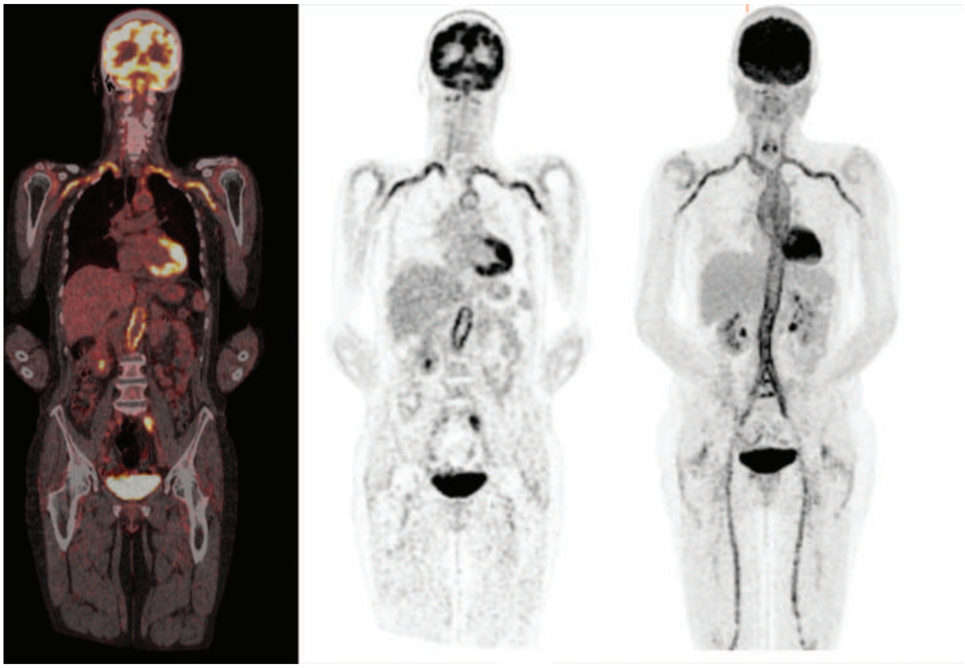

FIGURE 22-2 **^{18}F-FDG-PET/CT in a patient with FUO.** This 72-year-old woman presented with a low-grade fever and severe fatigue of almost 3 months' duration. An extensive history was taken, but the patient had no specific complaints and had not traveled recently. Her previous history was unremarkable, and she did not use any drugs. Physical examination, including palpation of the temporal arteries, yielded completely normal results. Laboratory examination showed normocytic anemia, a C-reactive protein level of 43 mg/L, an erythrocyte sedimentation rate of 87 mm/h, and mild hypoalbuminemia. Results of the other obligatory tests were all normal. ^{18}F-FDG-PET/CT showed increased FDG uptake in all major arteries (carotid, jugular, and subclavian arteries; thoracic and abdominal aorta; iliac, femoral, and popliteal arteries) and in the soft tissue around the shoulders, hips, and knees—findings compatible with large-vessel vasculitis and polymyalgia rheumatica. Within 1 week after the initiation of treatment with prednisone (60 mg once daily), the patient completely recovered. After 1 month, the prednisone dose was slowly tapered.

Although PET/CT and other scintigraphic techniques do not directly provide a definitive diagnosis (with the exception of some patients with, for instance, large vessel vasculitis), they often identify the anatomic location of a particular ongoing metabolic process. With the help of other techniques such as biopsy and culture, a timely diagnosis and treatment can be facilitated. Pathologic FDG uptake is quickly eradicated by treatment with glucocorticoids in many diseases, including vasculitis and lymphoma; therefore, glucocorticoid use should be stopped or postponed until after ^{18}F-FDG-PET/CT is performed.

^{18}F-FDG-PET/CT is a relatively expensive procedure whose availability is still limited in some regions. Nevertheless, ^{18}F-FDG-PET/CT can be cost-effective in the FUO diagnostic workup if used early, helping to establish an early diagnosis, reducing days of hospitalization for diagnostic purposes, and obviating unnecessary and unhelpful tests. When ^{18}F-FDG-PET/CT has been made under the right conditions (i.e., during active inflammation) but has not contributed to the final diagnosis, repeating PET/CT is probably of little value, unless new signs or symptoms appear.

ALTERNATIVES TO POSITRON EMISSION TOMOGRAPHY/COMPUTED TOMOGRAPHY

When PET/CT is unavailable, other whole-body imaging modalities should be considered. Compared to CT, scintigraphy provides the location of active inflammation. Conventional scintigraphic methods in clinical practice are ^{67}Ga-citrate scintigraphy and ^{111}In- or ^{99m}Tc-labeled leukocyte scintigraphy. The diagnostic yields of these conventional scintigraphic studies are lower than for PET/CT: the diagnostic yield of gallium scintigraphy ranges from 21% to 54%, and on average, the location of a source of fever can correctly be localized in approximately one-third of patients. The diagnostic value of leukocyte scintigraphy ranges from 8% to 31%, and overall, the cause of FUO can correctly be identified in one-fifth of patients. Recently, one study of full-body MRI as an alternative for PET/CT reported a diagnostic yield comparable to PET/CT. Further studies are needed to confirm this finding.

LATER-STAGE DIAGNOSTIC TESTS

Later-stage diagnostic testing should always be PDC driven and may come from repeated history-taking, physical examination, or PET/CT. Biopsies for pathology and/or microbiology evaluation will often be necessary.

Diagnostic delay frequently results from a failure to recognize PDCs in the available information. When all PDCs have been thoroughly evaluated but no final explanation has been found, waiting for new PDCs to appear probably is better than ordering more screening investigations. Only when a patient's condition deteriorates without providing new PDCs should further diagnostic workup be performed.

SECOND OPINION IN AN EXPERT CENTER

When no explanation for FUO is found despite the workup described above, a second opinion from an FUO expert center should be considered. The single study on the value of second opinion in FUO reported that in over half of patients with unexplained FUO, a diagnosis could be found in an expert center. Additionally, of all patients who remained without a diagnosis after second opinion, 1 in 10 became fever-free upon empirical treatment, adding up to a beneficial outcome in a little under 70% of patients. One of the reasons for the higher diagnostic rate in expert centers is the fact that FUO is often an atypical presentation of a common disease, making pattern recognition informed by high exposure an important diagnostic tool.

TREATMENT

Fever of Unknown Origin

Rational treatment is based on the final diagnosis. In FUO patients remaining with unexplained FUO, the large majority may spontaneously become symptom-free. Empirical therapeutic trials should be avoided, except in cases in which a patient's condition is rapidly deteriorating. Drugs that are commonly used for therapeutic trials are antibiotics, antituberculars, nonsteroidal anti-inflammatory drugs, colchicine, and interleukin-1 inhibitors.

ANTIBIOTICS AND ANTITUBERCULOUS THERAPY

Antibiotic or antituberculous therapy may irrevocably diminish the ability to culture bacteria. However, hemodynamic instability or neutropenia is a good indication for empirical antibiotic therapy. If the TST or IGRA is positive, or if granulomatous disease is present with anergy and sarcoidosis seems unlikely, or in any patient coming from an endemic region with a clinical picture fitting extrapulmonary tuberculosis, a trial of antituberculous therapy may be started, but not before mycobacterial cultures and, if available, mycobacterial PCR testing have been performed on material collected from the suspected location of inflammation. Especially in miliary tuberculosis, it may be very difficult to obtain a rapid diagnosis. If the fever does not respond after 6 weeks of empirical antituberculous treatment, another diagnosis should be considered.

COLCHICINE, NONSTEROIDAL ANTI-INFLAMMATORY DRUGS, AND GLUCOCORTICOIDS

If the fever persists and the source remains elusive after completion of investigations, supportive treatment with acetaminophen or nonsteroidal anti-inflammatory drugs (NSAIDs) can be helpful. The response of Still's disease to NSAIDs is dramatic in some cases. Colchicine is highly effective in preventing attacks of familial Mediterranean fever (FMF) but is not always effective once an attack is underway. When FMF is suspected, the response to colchicine is not a completely reliable diagnostic tool in the acute phase, but within weeks to months of continuous colchicine treatment, most patients show remarkable improvements in the frequency and severity of subsequent febrile episodes. Therefore, colchicine may be tried in patients with features compatible with FMF, especially when these patients originate from a high-prevalence region. In patients suffering from pericarditis as one of the main associated symptoms, colchicine may also be effective to prevent recurrent attacks. If Behçet's disease is considered likely, colchicine may also have favorable effect.

The effects of glucocorticoids on giant cell arteritis and polymyalgia rheumatica are impressive. Early empirical trials with glucocorticoids, however, decrease the chances of reaching a diagnosis for which more specific, less morbid, and sometimes life-saving treatment might be more appropriate, such as malignant lymphoma. The ability of glucocorticoids to mask fever while permitting the spread of infection or lymphoma dictates that their use should be avoided unless infectious diseases and malignant lymphoma have been sufficiently ruled out and inflammatory disease is probable and is likely to be debilitating or threatening.

INTERLEUKIN-1 INHIBITION

Interleukin (IL) 1 is a key cytokine in local and systemic inflammation and the febrile response. The availability of specific IL-1-targeting agents has revealed a pathologic role of IL-1-mediated inflammation in a growing list of diseases. Anakinra, a recombinant form of the naturally occurring IL-1 receptor antagonist (IL-1Ra), blocks the activity of both IL-1α and IL-1β. Anakinra is extremely effective in the treatment of many autoinflammatory syndromes, such as FMF, cryopyrin-associated periodic syndrome, tumor necrosis factor receptor–associated periodic syndrome, mevalonate kinase deficiency (hyper-IgD syndrome), Schnitzler's syndrome, and Still's disease. There are many other chronic inflammatory disorders in which anti-IL-1 therapy is highly effective. A therapeutic trial with anakinra can be considered in patients whose FUO has not been diagnosed after later-stage diagnostic tests and show signs of IL-1-driven inflammation, such as serositis, elevated CRP, and elevated ferritin. When autoinflammation is considered in the differential diagnosis, IL-1 inhibition is preferred over corticosteroids to prevent the metabolic, immunologic, and gastrointestinal side effects of glucocorticoid administration, and because IL-1 inhibition has superior efficacy.

PROGNOSIS

The prognosis of patients with FUO mostly depends on the underlying disease. In patients in whom FUO remains unexplained despite extensive evaluation, the prognosis is favorable. The risk of FUO-related mortality is probably highest during the early phases of the diagnostic process: in a cohort study including 168 patients without a final diagnosis, all four patients who died did so during the index admission; in two of them, diagnoses were made upon autopsy (intravascular lymphoma and bilateral pneumonia).

Large cohort studies in patients remaining without a diagnosis report high percentages of spontaneous resolution of fever and a mortality of 8% or less during several years of follow-up. ^{18}F-FDG-PET/CT may be helpful to predict which patients will resolve because normal ^{18}F-FDG-PET/CT scans are associated with higher rates of spontaneous resolution.

FURTHER READING

Betrains A et al: update on imaging in fever and inflammation of unknown origin: Focus on infectious disorders. Clin Microbiol Infect 18:S1198, 2023.

Erdem H et al: Classical fever of unknown origin in 21 countries with different economic development: An international ID-IRI study. Eur J Clin Microbiol Infect Dis 42:387, 2023.

Mulders-Manders C et al: Fever of unknown origin. Clin Med 15:280, 2015.

van Rijsewijk N et al: Molecular imaging of fever of unknown origin: An update. Semin Nucl Med 53:4, 2023.

Wright WF et al: Fever of unknown origin (FUO): A call for new research standards and updated clinical management. Am J Med 135:173, 2022.

Section 3 Nervous System Dysfunction

23

Syncope

Roy Freeman, Satish R. Raj

Syncope is a transient, self-limited loss of consciousness due to acute global impairment of cerebral blood flow. The onset is rapid, duration brief, and recovery spontaneous and complete. Other causes of transient loss of consciousness need to be distinguished from syncope; these include seizures, vertebrobasilar ischemia, hypoxemia, and hypoglycemia. A syncopal prodrome (*presyncope*) is common, although loss of consciousness may occur without any warning symptoms. Typical presyncopal symptoms include lightheadedness or faintness, dizziness, weakness, fatigue, and visual and auditory disturbances. The causes of syncope can be divided into three general categories: (1) neurally mediated syncope (also called *reflex or vasovagal syncope*), (2) orthostatic hypotension, and (3) cardiac syncope.

Neurally mediated syncope comprises a heterogeneous group of disorders that are characterized by a transient change in the reflexes responsible for maintaining cardiovascular homeostasis. Episodic vasodilation (or loss of vasoconstrictor tone), decreased cardiac output, and bradycardia occur in varying combinations, resulting in temporary failure of blood pressure control. In contrast, in patients with orthostatic hypotension due to autonomic failure, these cardiovascular homeostatic reflexes are chronically impaired. Cardiac syncope may be due to arrhythmias or structural cardiac diseases that cause a decrease in cardiac output. The clinical features, underlying pathophysiologic mechanisms, therapeutic interventions, and prognoses differ markedly among these three causes.

EPIDEMIOLOGY AND NATURAL HISTORY

Syncope is a common presenting problem, accounting for ~3% of all emergency department (ED) visits and 1% of all hospital admissions. The annual cost for syncope-related hospitalization in the United States is ~$2.4 billion. Syncope has a lifetime cumulative incidence of up to 40% in the general population. A bimodal age distribution exists; the peak incidence in the young occurs between ages 10 and 30 years, with a median peak around 15 years. Neurally mediated syncope is the etiology in the vast majority of these cases. In older adults, there is a sharp rise in the incidence of syncope after 70 years of age.

In population-based studies, neurally mediated syncope is the most common cause of syncope. The incidence is higher in women than men. In young subjects, there is often a family history in first-degree relatives. Cardiovascular disease due to structural disease or arrhythmias is the next most common cause in most series, particularly in ED settings and in older patients. Orthostatic hypotension also increases in prevalence with age because of the reduced baroreflex responsiveness, decreased cardiac compliance, and attenuation of the vestibulosympathetic reflex associated with aging. Other contributors are reduced fluid intake and vasoactive medications, also more likely in this age group. In the elderly, orthostatic hypotension is more common in institutionalized than community-dwelling individuals, most likely explained by a greater prevalence of predisposing neurologic disorders, physiologic impairment, and vasoactive medication use among institutionalized patients.

Syncope of noncardiac and unexplained origin in younger individuals has an excellent prognosis; life expectancy is unaffected. By contrast, syncope due to a cardiac cause, either structural heart disease or a primary arrhythmic disorder, is associated with an increased risk of sudden cardiac death and mortality from other causes. Similarly, the mortality rate is increased in individuals with syncope due to orthostatic hypotension related to age and the associated comorbid conditions **(Table 23-1)**. The likelihood of hospitalization and mortality risk are higher in older adults.

PATHOPHYSIOLOGY

The upright posture imposes a unique physiologic stress upon humans; most, although not all, syncopal episodes occur from a standing position. Standing results in pooling of 500–1000 mL of blood in the lower extremities, buttocks, and splanchnic circulation. The dependent pooling leads to a decrease in venous return to the heart and reduced ventricular filling that result in diminished cardiac output and blood pressure. These hemodynamic changes provoke a compensatory reflex response, initiated by the baroreceptors in the carotid sinus and aortic arch, resulting in increased sympathetic outflow and decreased vagal nerve activity **(Fig. 23-1)**. The reflex increases peripheral resistance, venous return to the heart, and cardiac output and thus limits the fall in blood pressure. If this response fails, as is the case chronically in orthostatic hypotension and transiently in neurally mediated syncope, hypotension and cerebral hypoperfusion occur.

TABLE 23-1 High-Risk Features Indicating Hospitalization or Intensive Evaluation of Syncope

Chest pain suggesting coronary ischemia
Features of congestive heart failure
Moderate or severe valvular disease
Moderate or severe structural cardiac disease
Electrocardiographic features of ischemia
History of ventricular arrhythmias
Prolonged QT interval (>500 ms)
Repetitive sinoatrial block or sinus pauses
Persistent sinus bradycardia
Bi- or trifascicular block or intraventricular conduction delay with QRS duration ≥120 ms
Atrial fibrillation
Nonsustained ventricular tachycardia
Family history of sudden death
Preexcitation syndromes
Brugada pattern on electrocardiogram
Palpitations at time of syncope
Syncope at rest or during exercise

Syncope is a form of transient loss of consciousness (TLOC) that is a consequence of global cerebral hypoperfusion. It represents a failure of cerebral blood flow autoregulatory mechanisms. Myogenic factors, local metabolites, and to a lesser extent autonomic neurovascular control are responsible for the autoregulation of cerebral blood flow **(Chap. 318)**. The latency of the autoregulatory response is 5–10 s. Typically, cerebral blood flow ranges from 50–60 mL/min per 100 g brain tissue and remains relatively constant over perfusion pressures ranging from 50–150 mmHg. Cessation of blood flow for 6–8 s will result in loss of consciousness, while impairment of consciousness ensues when blood flow decreases to 25 mL/min per 100 g brain tissue.

From the clinical standpoint, a fall in systemic systolic blood pressure to ~50 mmHg or lower will usually result in syncope. A decrease in cardiac output and/or systemic vascular resistance—the determinants of blood pressure—thus underlies the pathophysiology of syncope. Common causes of impaired cardiac output include decreased effective circulating blood volume, increased thoracic pressure, massive pulmonary embolus, cardiac brady- and tachyarrhythmias, valvular heart disease, and myocardial dysfunction. Systemic vascular resistance may be decreased by central and peripheral autonomic nervous system diseases, sympatholytic medications, and transiently during neurally mediated syncope. Increased cerebral vascular resistance, most frequently due to hypocarbia induced by hyperventilation, may also contribute to the pathophysiology of syncope.

Two patterns of electroencephalographic (EEG) changes occur in syncopal subjects. The first is a "slow-flat-slow" pattern in which normal background activity is replaced with high-amplitude slow delta waves. This is followed by sudden flattening of the EEG—a cessation or attenuation of cortical activity—followed by the return of slow waves, and then normal activity. A second pattern, the "slow pattern," is characterized by increasing and decreasing slow wave activity only. The EEG flattening that occurs in the slow-flat-slow pattern is a marker of more severe cerebral hypoperfusion. Despite the presence of myoclonic movements and other motor activity during some syncopal events, EEG seizure discharges are not detected.

CLASSIFICATION

NEURALLY MEDIATED SYNCOPE

Neurally mediated (reflex; vasovagal) syncope is the final pathway of a complex central and peripheral nervous system reflex arc. There is a transient change in autonomic efferent activity with increased parasympathetic outflow, plus sympathoinhibition, resulting in bradycardia, vasodilation, and/or reduced vasoconstrictor tone (the vasodepressor response) and reduced cardiac output. The resulting fall in systemic blood pressure can then reduce cerebral blood flow to below the compensatory limits of autoregulation **(Fig. 23-2)**. In order to develop neurally mediated syncope, a functioning autonomic nervous system is necessary, in contrast to syncope resulting from autonomic failure (discussed below).

Multiple triggers of the afferent limb of the reflex arc can result in neurally mediated syncope. In some situations, these can be clearly defined, e.g., orthostatic stress, and stimulation of the carotid sinus, the gastrointestinal tract, or the bladder. Often, however, the trigger is less easily recognized and the cause is multifactorial. Under these circumstances, it is likely that different afferent pathways converge on the central autonomic network within the medulla that integrates the neural impulses and mediates the vasodepressor-bradycardic response.

Classification of Neurally Mediated Syncope Neurally mediated syncope may be subdivided based on the afferent pathway and provocative trigger. Vasovagal syncope (the common faint) is provoked by intense emotion, pain, and/or orthostatic stress, whereas the situational reflex syncopes have specific localized stimuli that provoke the reflex vasodilation and bradycardia that leads to syncope. The underlying mechanisms have been identified and pathophysiology delineated for most of these situational reflex syncopes. The afferent trigger may originate in the pulmonary system, gastrointestinal system, urogenital system, heart, and carotid sinus in the carotid artery **(Table 23-2)**.

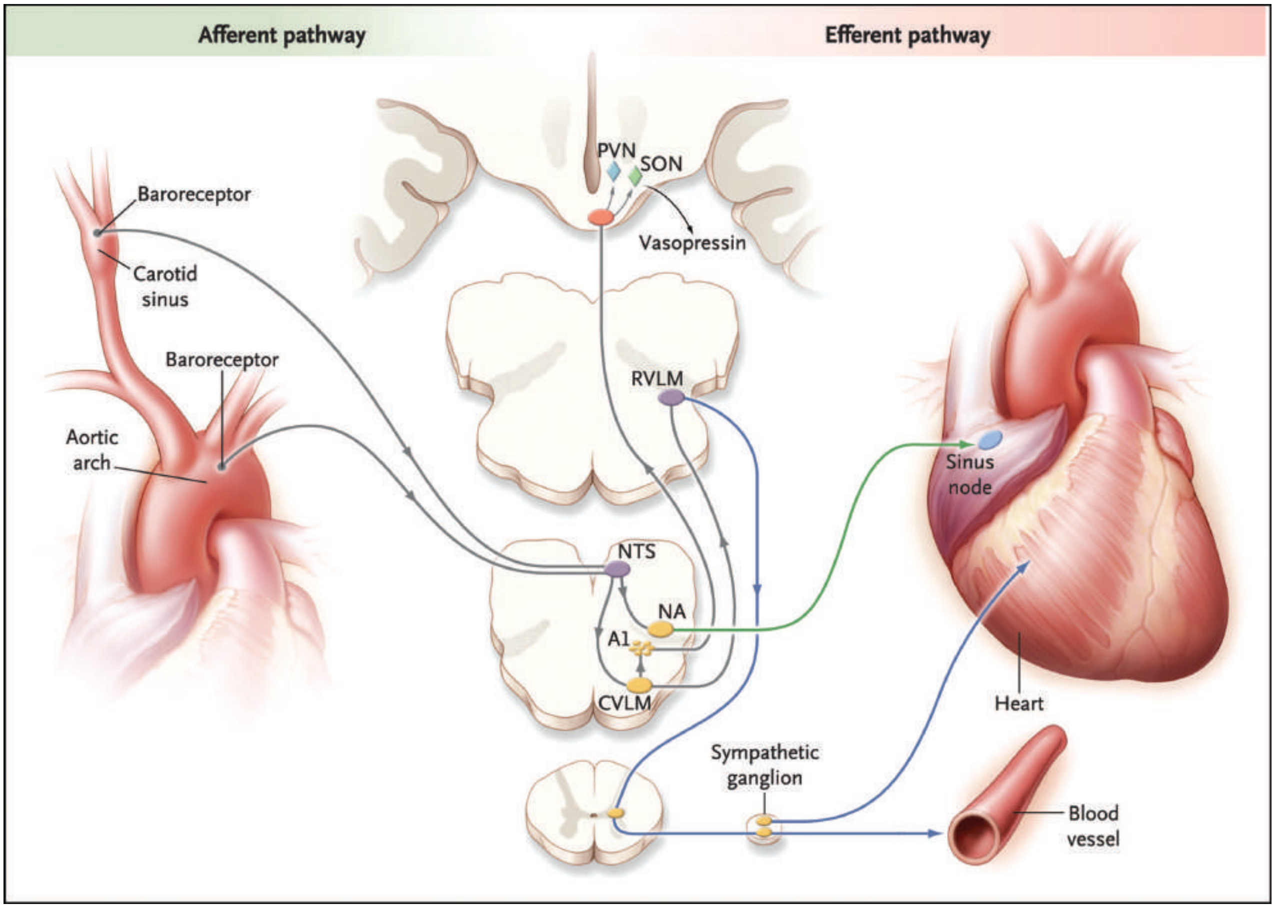

FIGURE 23-1 The baroreflex. A decrease in arterial pressure unloads the baroreceptors—the terminals of afferent fibers of the glossopharyngeal and vagus nerves—that are situated in the carotid sinus and aortic arch. This leads to a reduction in the afferent impulses that are relayed from these mechanoreceptors through the glossopharyngeal and vagus nerves to the nucleus of the tractus solitarius (NTS) in the dorsomedial medulla. The reduced baroreceptor afferent activity produces a decrease in vagal nerve input to the sinus node that is mediated via connections of the NTS to the nucleus ambiguus (NA). There is an increase in sympathetic efferent activity that is mediated by the NTS projections to the caudal ventrolateral medulla (CVLM) (an excitatory pathway) and from there to the rostral ventrolateral medulla (RVLM) (an inhibitory pathway). The activation of RVLM presympathetic neurons in response to hypotension is thus predominantly due to disinhibition. In response to a sustained fall in blood pressure, vasopressin release is mediated by projections from the A1 noradrenergic cell group in the ventrolateral medulla. This projection activates vasopressin-synthesizing neurons in the magnocellular portion of the paraventricular nucleus (PVN) and the supraoptic nucleus (SON) of the hypothalamus. *Blue* denotes sympathetic neurons, and *green* denotes parasympathetic neurons. *(From R Freeman: Neurogenic orthostatic hypotension. N Engl J Med 358:615, 2008. Copyright © 2008 Massachusetts Medical Society. Reprinted with permission.)*

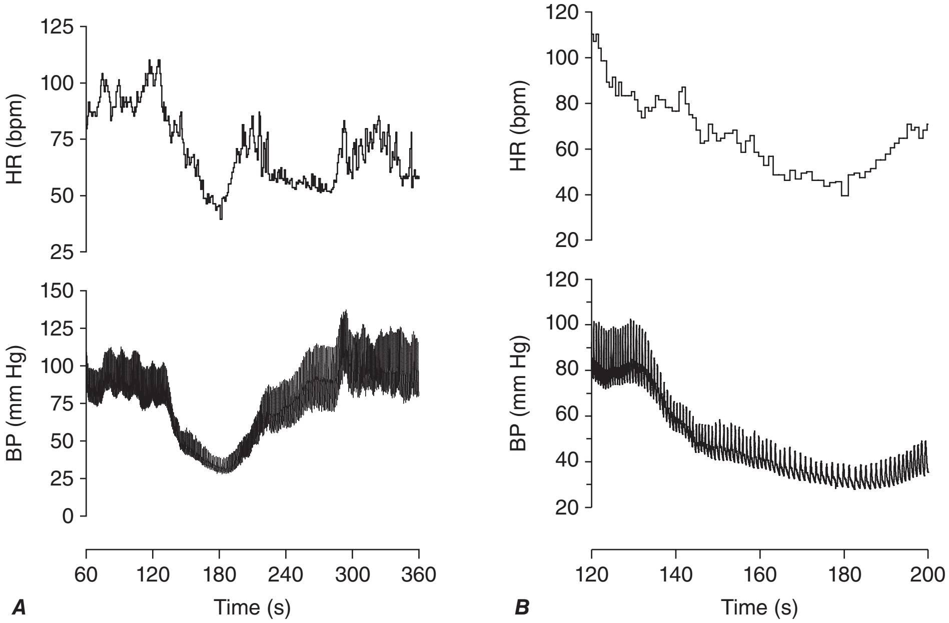

FIGURE 23-2 ***A.*** The paroxysmal hypotensive-bradycardic response that is characteristic of neurally mediated syncope. Noninvasive beat-to-beat blood pressure and heart rate are shown >5 min (from 60 to 360 s) of an upright tilt on a tilt table. ***B.*** The same tracing expanded to show 80 s of the episode (from 80 to 200 s). BP, blood pressure; bpm, beats per minute; HR, heart rate.

TABLE 23-2 Causes of Syncope

A. Neurally Mediated Syncope

- Vasovagal syncope
  - Provoked fear, pain, anxiety, intense emotion, sight of blood, unpleasant sights and odors, orthostatic stress
- Situational reflex syncope
  - Pulmonary
    - Cough syncope, wind instrument player's syncope, weightlifter's syncope, "mess trick"[a] and "fainting lark,"[b] sneeze syncope, airway instrumentation
  - Urogenital
    - Postmicturition syncope, urogenital tract instrumentation, prostatic massage
  - Gastrointestinal
    - Swallow syncope, glossopharyngeal neuralgia, esophageal stimulation, gastrointestinal tract instrumentation, rectal examination, defecation syncope
  - Cardiac
    - Bezold-Jarisch reflex, cardiac outflow obstruction
  - Carotid sinus
    - Carotid sinus sensitivity, carotid sinus massage
  - Ocular
    - Ocular pressure, ocular examination, ocular surgery

B. Orthostatic Hypotension

- Primary autonomic failure due to idiopathic central and peripheral neurodegenerative diseases—the "synucleinopathies"
  - Lewy body diseases
    - Parkinson's disease
    - Lewy body dementia
    - Pure autonomic failure
  - Multiple system atrophy (Shy-Drager syndrome)
- Secondary autonomic failure due to autonomic peripheral neuropathies
  - Diabetes
  - Hereditary amyloidosis (familial amyloid polyneuropathy)
  - Primary amyloidosis (AL amyloidosis; immunoglobulin light chain associated)
  - Hereditary sensory and autonomic neuropathies (HSAN) (especially type III—familial dysautonomia)
  - Idiopathic immune-mediated autonomic neuropathy
  - Autoimmune autonomic ganglionopathy
  - Sjögren's syndrome
  - Paraneoplastic autonomic neuropathy
  - HIV neuropathy
- Postprandial hypotension
- Iatrogenic (drug-induced)
- Volume depletion

C. Cardiac Syncope

- Arrhythmias
  - Sinus node dysfunction
  - Atrioventricular dysfunction
  - Supraventricular tachycardias
  - Ventricular tachycardias
  - Inherited channelopathies
- Cardiac structural disease
  - Valvular disease
  - Myocardial ischemia
  - Obstructive and other cardiomyopathies
  - Atrial myxoma
  - Pericardial effusions and tamponade

[a] Hyperventilation for ~1 min, followed by sudden chest compression.
[b] Hyperventilation (~20 breaths) in a squatting position, rapid rise to standing, then Valsalva maneuver.

Hyperventilation, leading to hypocarbia and cerebral vasoconstriction, and raised intrathoracic pressure that impairs venous return to the heart play a central role in many of the situational reflex syncopes. The afferent pathway of the reflex arc differs among these disorders, but the efferent response via the vagus and sympathetic pathways is similar.

Alternately, neurally mediated syncope may be subdivided based on the predominant efferent pathway. Vasodepressor syncope describes syncope predominantly due to efferent, sympathetic, vasoconstrictor failure; cardioinhibitory syncope describes syncope predominantly associated with bradycardia or asystole due to increased vagal outflow; and mixed response syncope describes syncope in which there are both vagal and sympathetic reflex changes.

Features of Neurally Mediated Syncope In addition to symptoms of orthostatic intolerance such as dizziness, lightheadedness, and fatigue, premonitory features of autonomic activation may be present in patients with neurally mediated syncope. These include diaphoresis, pallor, palpitations, nausea, hyperventilation, and yawning. During the syncopal event, proximal and distal myoclonus (typically arrhythmic and multifocal) may occur, raising the possibility of a seizure. The eyes typically remain open and usually deviate upward. Pupils are usually dilated. Roving eye movements may occur. Grunting, moaning, snorting, and stertorous breathing may be present. Urinary incontinence may occur. Fecal incontinence is very rare, however. Postictal confusion is also rare, although visual and auditory hallucinations and near-death and out-of-body experiences are sometimes reported.

Although some predisposing factors and provocative stimuli are well established (for example, motionless upright posture, warm ambient temperature, intravascular volume depletion, alcohol ingestion, hypoxemia, anemia, pain, the sight of blood, venipuncture, and intense emotion), the underlying basis for the widely different thresholds for syncope among individuals exposed to the same provocative stimulus is not known. A genetic basis for neurally mediated syncope may exist. Several studies have reported an increased incidence of syncope in first-degree relatives of fainters, and some candidate genes have been identified with sex-specific associations. These have not been reproduced in other large cohorts. It is likely that environmental, social, and cultural factors play a large role.

TREATMENT

Neurally Mediated Syncope

Reassurance, education, avoidance of provocative stimuli, and plasma volume expansion with fluid and salt are the cornerstones of the management of neurally mediated syncope. Isometric counterpressure maneuvers of the limbs (tensing of the abdominal and leg muscles, handgrip and arm tensing, and leg crossing) may raise blood pressure by increasing central blood volume and cardiac output. Of these, abdominal and leg muscle tensing is the most effective. By maintaining pressure in the autoregulatory zone, these maneuvers, which may be particularly helpful in patients with a long prodrome, avoid or delay the onset of syncope. A randomized controlled trial supports this intervention.

Fludrocortisone, vasoconstricting agents, and β-adrenoreceptor antagonists are widely used by experts to treat refractory patients. Of these, only midodrine has been shown to be effective in international, multicenter randomized controlled trials. Because vasodilation, decreased central blood volume, decreased stroke volume, and cardiac output are the dominant pathophysiologic syncopal mechanisms in most patients, use of a cardiac pacemaker is rarely beneficial. In patients with a cardioinhibitory syncope response during tilt-table testing, however, recent sham-controlled randomized clinical trial data have shown that a dual-chamber pacemaker with a closed-loop stimulation algorithm can decrease syncope recurrence. These studies restricted enrollment to older patients (>40 years) with frequent recurrence of syncope and a cardioinhibitory response on tilt-table test. In these patients, dual-chamber pacing may be helpful, although this continues to be an area of uncertainty.

ORTHOSTATIC HYPOTENSION

Orthostatic hypotension, defined as a reduction in systolic blood pressure of at least 20 mmHg or diastolic blood pressure of at least 10 mmHg after 3 min of standing or head-up tilt on a tilt table, is a manifestation of sympathetic vasoconstrictor (autonomic) failure **(Fig. 23-3)**. In many (but not all) cases, there is no compensatory increase in heart rate despite hypotension; with partial autonomic failure, heart rate may increase to some degree but is insufficient to maintain cardiac output. A variant of orthostatic hypotension is "delayed" orthostatic hypotension, which occurs beyond 3 min of standing; this may reflect a mild or early form of sympathetic adrenergic dysfunction. In some cases, orthostatic hypotension can occur within 15 s of standing with full resolution within 45 s (so-called initial orthostatic hypotension), a finding that may reflect a transient mismatch between cardiac output and peripheral vascular resistance and does not represent autonomic failure.

Characteristic symptoms of orthostatic hypotension include lightheadedness, dizziness, and presyncope (near-faintness) occurring in response to sudden postural change. However, symptoms may be absent or nonspecific such as generalized weakness, fatigue, cognitive slowing, leg buckling, or headache. Visual blurring may occur, likely due to retinal or occipital lobe ischemia. Neck pain, typically in the suboccipital, posterior cervical, and shoulder region (the "coat-hanger headache"), most likely due to neck muscle ischemia, may be the only symptom. Patients may report orthostatic dyspnea (thought to reflect ventilation-perfusion mismatch due to inadequate perfusion of ventilated lung apices) or angina (attributed to impaired myocardial perfusion even with normal coronary arteries). Symptoms may be exacerbated by exertion, prolonged standing, increased ambient temperature, or meals. Syncope is usually preceded by warning symptoms, but may occur suddenly, suggesting the possibility of a seizure or cardiac cause. Some patients have profound decreases in blood pressure, sometimes without symptoms but placing them at risk for falls and injuries if the autoregulatory threshold is crossed with ensuing cerebral hypoperfusion.

Supine hypertension is common in patients with orthostatic hypotension due to autonomic failure, affecting >50% of patients in some series. Orthostatic hypotension may present after initiation of therapy for hypertension, and supine hypertension may follow treatment of orthostatic hypotension. However, in other cases, the association of the two conditions is unrelated to therapy; it may in part be explained by baroreflex dysfunction in the presence of residual sympathetic outflow, particularly in patients with central autonomic degeneration.

Causes of Neurogenic Orthostatic Hypotension Causes of neurogenic orthostatic hypotension include central and peripheral autonomic nervous system dysfunction **(Chap. 451)**. Autonomic dysfunction of other organ systems (including the bladder, bowels, sexual organs, and sudomotor system) of varying severity frequently accompanies orthostatic hypotension in these disorders (Table 23-2).

The primary autonomic degenerative disorders are multiple system atrophy (Shy-Drager syndrome; **Chap. 451**), Parkinson's disease **(Chap. 446)**, dementia with Lewy bodies **(Chap. 445)**, and pure autonomic failure **(Chap. 451)**. These are often grouped together as "synucleinopathies" due to the presence of α-synuclein, a protein that aggregates predominantly in the cytoplasm of neurons in the Lewy body disorders (Parkinson's disease, dementia with Lewy bodies, and pure autonomic failure) and in the glia in multiple system atrophy.

Peripheral autonomic dysfunction may also accompany small-fiber peripheral neuropathies such as those associated with diabetes mellitus, acquired and hereditary amyloidosis, immune-mediated neuropathies, and hereditary sensory and autonomic neuropathies (HSAN; particularly HSAN type III, familial dysautonomia) **(Chaps. 457 and 458)**. Less frequently, orthostatic hypotension is associated with the peripheral neuropathies that accompany vitamin B_{12} deficiency, neurotoxin exposure, HIV and other infections, and porphyria.

Patients with autonomic failure and the elderly are susceptible to falls in blood pressure associated with meals. The magnitude of the blood pressure fall is exacerbated by large meals, meals high in carbohydrate, and alcohol intake. The mechanism of postprandial hypotension, and resultant syncope, is not fully elucidated.

Orthostatic hypotension is often iatrogenic. Drugs from several classes may lower peripheral resistance (e.g., α-adrenoreceptor antagonists used to treat hypertension and prostatic hypertrophy; diuretics, nitrates and other venodilators and vasodilators; other antihypertensive agents of several classes; tricyclic agents and phenothiazines). Iatrogenic volume depletion due to diuresis and volume depletion due to medical causes (hemorrhage, vomiting, diarrhea, or decreased fluid intake) may also result in decreased effective circulatory volume, orthostatic hypotension, and syncope.

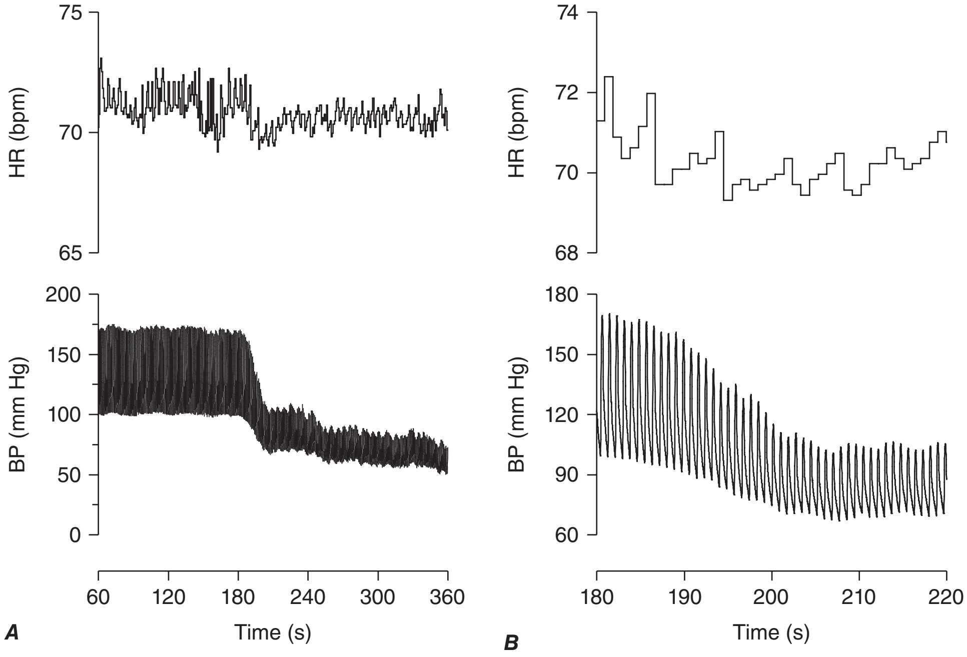

FIGURE 23-3 ***A.*** The gradual fall in blood pressure without a compensatory heart rate increase that is characteristic of orthostatic hypotension due to autonomic failure. Blood pressure and heart rate are shown >5 min (from 60 to 360 s) of an upright tilt on a tilt table. ***B.*** The same tracing expanded to show 40 s of the episode (from 180 to 220 s). BP, blood pressure; bpm, beats per minute; HR, heart rate.

TREATMENT

Orthostatic Hypotension

The first step is to remove reversible causes—usually vasoactive medications **(see Table 451-6)**. Next, nonpharmacologic interventions should be introduced. These include patient education regarding staged moves from supine to upright; warnings about the hypotensive effects of large meals; instructions about the isometric counterpressure maneuvers that increase intravascular pressure (see above); and raising the head of the bed to reduce supine hypertension and nocturnal diuresis. Intravascular volume should be expanded by increasing dietary fluid and salt. The rapid ingestion of 500 mL of plain water can often effect a short-term pressor response in these patients. If these nonpharmacologic measures fail, pharmacologic intervention with fludrocortisone acetate and vasoconstricting agents such as midodrine or L-dihydroxyphenylserine (droxidopa) should be introduced. Some patients with intractable symptoms require additional therapy with supplementary agents that include pyridostigmine, atomoxetine, yohimbine, octreotide, desmopressin acetate (DDAVP), and erythropoietin **(Chap. 451)**.

CARDIAC SYNCOPE

Cardiac (or cardiovascular) syncope is caused by arrhythmias and structural heart disease. These may occur in combination because structural disease renders the heart more vulnerable to abnormal electrical activity.

Arrhythmias Bradyarrhythmias that cause syncope include those due to severe sinus node dysfunction (e.g., sinus arrest or sinoatrial block) and atrioventricular (AV) block (e.g., Mobitz type II, high-grade, and complete AV block). The bradyarrhythmias due to sinus node dysfunction are often associated with an atrial tachyarrhythmia, a disorder known as the tachycardia-bradycardia syndrome. A prolonged pause following the termination of a tachycardic episode is a frequent cause of syncope in patients with the tachycardia-bradycardia syndrome. Medications of several classes may also cause bradyarrhythmias of sufficient severity to cause syncope. Syncope due to bradycardia or asystole has been referred to as a Stokes-Adams attack.

Ventricular tachyarrhythmias frequently cause syncope. The likelihood of syncope with ventricular tachycardia is in part dependent on the ventricular rate (rates <200 beats/min are less likely to cause syncope) and ventricular function (a patient with poor ventricular function is more likely to have syncope at a given ventricular rate). The compromised hemodynamic function during ventricular tachycardia is caused by ineffective ventricular contraction, reduced diastolic filling due to abbreviated filling periods, loss of AV synchrony, and concurrent myocardial ischemia.

Several disorders associated with cardiac electrophysiologic instability and arrhythmogenesis are due to mutations in ion channel subunit genes. These include the long QT syndrome, Brugada syndrome, and catecholaminergic polymorphic ventricular tachycardia. The long QT syndrome is a genetically heterogeneous disorder associated with prolonged cardiac repolarization and a predisposition to ventricular arrhythmias. Syncope and sudden death in patients with long QT syndrome result from a unique polymorphic ventricular tachycardia called *torsades des pointes* that may degenerate into ventricular fibrillation. The long QT syndrome has been linked to genes encoding K^+ channel α-subunits, K^+ channel β-subunits, voltage-gated Na^+ channel, and a scaffolding protein, ankyrin B (ANK2). Brugada syndrome is characterized by syncope, polymorphic ventricular tachycardia, and idiopathic ventricular fibrillation in association with right ventricular electrocardiogram (ECG) abnormalities without structural heart disease. This disorder is also genetically heterogeneous, although it is most frequently linked to mutations in the Na^+ channel α-subunit, SCN5A. Catecholaminergic polymorphic tachycardia is an inherited, genetically heterogeneous disorder often involving cardiac calcium handling that is associated with exercise- or stress-induced ventricular arrhythmias, syncope, or sudden death. Acquired QT interval prolongation, most commonly due to drugs, may also result in ventricular arrhythmias and syncope. **These disorders are discussed in detail in Chap. 262.**

Structural Disease Structural heart disease (e.g., valvular disease, myocardial ischemia, hypertrophic and other cardiomyopathies, cardiac masses such as atrial myxoma, and pericardial effusions) may lead to syncope by compromising cardiac output. Structural disease may also contribute to other pathophysiologic mechanisms of syncope. For example, cardiac structural disease may predispose to arrhythmogenesis; aggressive treatment of cardiac failure with diuretics and/or vasodilators may lead to orthostatic hypotension; and inappropriate reflex vasodilation may occur with structural disorders such as aortic stenosis and hypertrophic cardiomyopathy, possibly provoked by increased ventricular contractility.

TREATMENT

Cardiac Syncope

Treatment of cardiac disease depends on the underlying disorder. Therapies for arrhythmias consist of cardiac pacing for bradycardia, including sinus node disease and AV block, and ablation, antiarrhythmic drugs, and cardioverter-defibrillators for atrial and ventricular tachyarrhythmias. These disorders are best managed by physicians with specialized skills in this area.

APPROACH TO THE PATIENT

Syncope

DIFFERENTIAL DIAGNOSIS

Syncope is easily diagnosed when the characteristic features are present; however, several disorders with real or apparent transient loss of consciousness may create diagnostic confusion.

Generalized and partial seizures **(Chap. 436)** may be confused with syncope; however, there are a number of differentiating features. Whereas tonic-clonic movements are the hallmark of a generalized seizure, myoclonic and other movements also may occur in up to 90% of syncopal episodes. Myoclonic jerks associated with syncope may be multifocal or generalized. They are typically arrhythmic and of short duration (<30 s). Mild flexor and extensor posturing also may occur. Partial or partial-complex seizures with secondary generalization are usually preceded by an aura, commonly an unpleasant smell; fear; anxiety; abdominal discomfort; or other visceral sensations. These phenomena should be differentiated from the premonitory features of syncope.

Autonomic manifestations of seizures (autonomic epilepsy) may provide a more difficult diagnostic challenge. Autonomic seizures have cardiovascular, gastrointestinal, pulmonary, urogenital, pupillary, and cutaneous manifestations that are similar to the premonitory features of syncope. Furthermore, the cardiovascular manifestations of autonomic epilepsy include clinically significant tachycardias and bradycardias that may be of sufficient magnitude to cause loss of consciousness. The presence of accompanying nonautonomic auras may help differentiate these episodes from syncope.

Loss of consciousness associated with a seizure usually lasts >5 min and is associated with prolonged postictal drowsiness and disorientation, whereas reorientation occurs almost immediately after a syncopal event. Muscle aches may occur after both syncope and seizures, although they tend to last longer and be more severe following a seizure. Seizures, unlike syncope, are rarely provoked by emotions or pain. Incontinence of urine may occur with both seizures and syncope; however, fecal incontinence occurs only very rarely with syncope.

Hypoglycemia may cause transient loss of consciousness, typically in individuals with type 1 or type 2 diabetes **(Chap. 415)** treated with insulin. The clinical features associated with impending or actual hypoglycemia include tremor, palpitations, anxiety, diaphoresis, hunger, and paresthesias. These symptoms are due to autonomic activation to counter the falling blood glucose. Hunger, in particular, is not a typical premonitory feature of syncope. Hypoglycemia also impairs neuronal function, leading to fatigue, weakness, dizziness, and cognitive and behavioral symptoms. Diagnostic difficulties may occur in individuals in strict glycemic control; repeated hypoglycemia impairs the counterregulatory response and leads to a loss of the characteristic warning symptoms that are the hallmark of hypoglycemia.

Patients with cataplexy **(Chap. 33)** experience an abrupt partial or complete loss of muscular tone triggered by strong emotions, typically anger or laughter. Unlike syncope, consciousness is maintained throughout the attacks, which typically last between 30 s and 2 min. There are no premonitory symptoms. Cataplexy occurs in 60–75% of patients with narcolepsy.

The clinical interview and interrogation of eyewitnesses (and ancillary cellphone video of the spell, when available) usually allow differentiation of syncope from falls due to vestibular dysfunction, cerebellar disease, extrapyramidal system dysfunction, and other gait disorders. A diagnosis of syncope can be particularly challenging in patients with dementia who experience repeated falls and are unable to provide a clear history of the episodes. If the fall is accompanied by head trauma, a postconcussive syndrome, amnesia for the precipitating events, and/or a loss or alteration of consciousness, this may also contribute to diagnostic difficulty.

Apparent loss of consciousness can be a manifestation of psychiatric disorders such as generalized anxiety, panic disorders, major depression, and somatization disorder. These possibilities should be considered in individuals who faint frequently without prodromal symptoms. Such patients are rarely injured despite numerous falls. There are no clinically significant hemodynamic changes concurrent with these episodes. In contrast, transient loss of consciousness due to vasovagal syncope precipitated by fear, stress, anxiety, and emotional distress is accompanied by hypotension and sometimes bradycardia.

INITIAL EVALUATION

The goals of the initial evaluation are to determine whether the transient loss of consciousness was due to syncope; to identify the cause; and to assess risk for future episodes and serious harm (Table 23-1). The initial evaluation should include a detailed history, thorough questioning of eyewitnesses, and a complete physical and neurologic examination. Blood pressure and heart rate should be measured in the supine position and after 3 min of standing to determine whether orthostatic hypotension is present. High-risk features on history include the new onset of chest discomfort, abdominal pain, shortness of breath, or headache; syncope during exertion or while supine; sudden onset of palpitations followed by syncope; and severe coronary artery or structural heart disease.

High-risk features on examination include an unexplained systolic blood pressure of <90 mmHg; suggestion of gastrointestinal hemorrhage; persistent bradycardia (<40 beats/min); and an undiagnosed systolic murmur.

An ECG should be performed if there is suspicion of syncope due to an arrhythmia or underlying cardiac disease. Relevant electrocardiographic abnormalities include bradyarrhythmias or tachyarrhythmias, AV block, acute myocardial ischemia, old myocardial infarction, long QT, and bundle branch block. This initial assessment will lead to the identification of a cause of syncope in ~50% of patients and also allows stratification of patients at risk for cardiac mortality.

Laboratory Tests Baseline laboratory blood tests are rarely helpful in identifying the cause of syncope. Blood tests should be performed when specific disorders, e.g., myocardial infarction, anemia, and secondary autonomic failure, are suspected (Table 23-2).

Autonomic Nervous System Testing (Chap. 451) Autonomic testing, including tilt-table testing, can be performed in specialized centers. Autonomic testing is helpful to uncover objective evidence of autonomic failure and also to demonstrate a predisposition to neurally mediated syncope. Autonomic testing includes assessments of parasympathetic autonomic nervous system function (e.g., heart rate variability to deep respiration and a Valsalva maneuver), sympathetic cholinergic function (e.g., thermoregulatory sweat response and quantitative sudomotor axon reflex test), and sympathetic adrenergic function (e.g., blood pressure response to a Valsalva maneuver and a tilt-table test with beat-to-beat blood pressure measurement). The hemodynamic abnormalities demonstrated on the tilt-table test (Figs. 23-2 and 23-3) may be useful in distinguishing orthostatic hypotension due to autonomic failure from the hypotensive bradycardic response of neurally mediated syncope. Similarly, the tilt-table test may help identify patients with syncope due to immediate or delayed orthostatic hypotension.

Carotid sinus massage should be considered in patients with symptoms suggestive of carotid sinus syncope and in patients >40 years with recurrent syncope of unknown etiology. This test should ideally be carried out under continuous ECG and blood pressure monitoring and should be avoided in patients with carotid bruits, possible or known plaques, or stenosis.

Cardiac Evaluation ECG monitoring is indicated for patients with a high pretest probability of arrhythmia causing syncope. Patients should be monitored in the hospital if the likelihood of a life-threatening arrhythmia is high, e.g., patients with severe coronary artery or structural heart disease, nonsustained ventricular tachycardia, supraventricular tachycardia, paroxysmal atrial fibrillation, trifascicular heart block, prolonged QT interval, Brugada syndrome ECG pattern, syncope during exertion, syncope while seated or supine, and family history of sudden cardiac death (Table 23-1). Continuous ambulatory electrocardiographic (Holter) monitoring is recommended for patients who experience frequent syncopal episodes (e.g., daily or almost daily), whereas loop recorders, which continually record and erase cardiac rhythm, are indicated for patients with suspected arrhythmias with low risk of sudden cardiac death. Loop recorders may be external (e.g., for evaluation of episodes that occur at a frequency of >1 per month) or implantable (e.g., if syncope occurs less frequently). The monitoring duration should ideally be at least twice the interspell duration.

Echocardiography should be performed in patients with a history of cardiac disease or if abnormalities are found on physical examination or the ECG. Echocardiographic diagnoses that may be responsible for syncope include aortic stenosis, hypertrophic cardiomyopathy, cardiac tumors, aortic dissection, and pericardial tamponade. Echocardiography also has a role in risk stratification for sudden cardiac death based on the left ventricular ejection fraction.

Treadmill exercise testing with ECG and blood pressure monitoring should be performed in patients who have experienced syncope during or shortly after exercise. Treadmill testing may help identify exercise-induced arrhythmias (e.g., tachycardia-related AV block) and exercise-induced exaggerated vasodilation.

Invasive electrophysiologic studies are indicated in patients with structural heart disease and ECG abnormalities in whom noninvasive investigations have failed to yield a diagnosis. Electrophysiologic studies have low sensitivity and specificity and should only be performed when a high pretest probability exists. Currently, these tests are rarely performed to evaluate patients with syncope.

Psychiatric Evaluation Screening for psychiatric disorders may be appropriate in patients with recurrent unexplained syncope episodes. Tilt-table testing, with demonstration of symptoms in the absence of hemodynamic change, may be useful in reproducing syncope in patients with suspected psychogenic syncope.

FURTHER READING

Brignole M et al: 2018 ESC Guidelines for the diagnosis and management of syncope. Eur Heart J 39:1883, 2018.

Brignole M et al: Cardiac pacing in severe recurrent reflex syncope and tilt-induced asystole. Eur Heart J 42:508, 2021.

Cheshire WP et al: Electrodiagnostic assessment of the autonomic nervous system: A consensus statement endorsed by the American Autonomic Society, American Academy of Neurology, and the International Federation of Clinical Neurophysiology. Clin Neurophysiol 132:666, 2021.

Freeman R et al: Consensus statement on the definition of orthostatic hypotension, neurally mediated syncope and the postural tachycardia syndrome. Auton Neurosci 161:46, 2011.

Freeman R et al: Orthostatic Hypotension: JACC state-of-the-art review. J Am Coll Cardiol 72:1294, 2018.

Gibbons CH et al: The recommendations of a consensus panel for the screening, diagnosis, and treatment of neurogenic orthostatic hypotension and associated supine hypertension. J Neurol 264:1567, 2017.

Sheldon RS, Raj SR: Pacing and vasovagal syncope: Back to our physiologic roots. Clin Auton Res 27:213, 2017.

Sheldon R et al: Midodrine for the prevention of vasovagal syncope: A randomized clinical trial. Ann Intern Med 74:1349, 2021.

Shen WK et al: 2017 ACC/AHA/HRS guideline for the evaluation and management of patients with syncope: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. Circulation 136:e60, 2017.

24

Dizziness and Vertigo

Mark F. Walker, Robert B. Daroff[*]

Dizziness is an imprecise symptom used to describe a variety of common sensations that include vertigo, light-headedness, faintness, and imbalance. *Vertigo* refers to a sense of spinning or other motion that may be physiological, occurring during or after a sustained head rotation, or pathological, due to vestibular dysfunction. The term *light-headedness* is classically applied to presyncopal sensations resulting from brain hypoperfusion but as used by patients has little specificity, as it may also refer to other symptoms such as disequilibrium and imbalance. A challenge to diagnosis is that patients often have difficulty distinguishing among these various symptoms, and the words they choose do not reliably indicate the underlying etiology.

There are many causes of dizziness. Vestibular dizziness (vertigo or imbalance) may be due to peripheral disorders that affect the labyrinths or vestibular nerves, or it may result from disruption of central vestibular pathways. It may be paroxysmal or due to a fixed unilateral or bilateral vestibular deficit. Acute unilateral lesions cause vertigo due to a sudden imbalance in vestibular inputs from the two labyrinths. Bilateral lesions cause imbalance and instability of vision when the head moves (*oscillopsia*) due to loss of normal vestibular reflexes.

Presyncopal dizziness occurs when cardiac dysrhythmia, orthostatic hypotension, medication effects, or another cause leads to brain hypoperfusion. Such presyncopal sensations vary in duration; they may increase in severity until loss of consciousness occurs, or they may resolve before loss of consciousness if the cerebral ischemia is corrected. Faintness and syncope, which are discussed in detail in Chap. 23, should always be considered when one is evaluating patients with brief episodes of dizziness or dizziness that occurs with upright posture. Other causes of dizziness include nonvestibular balance and gait disorders (e.g., loss of proprioception from sensory neuropathy, parkinsonism) and anxiety.

When evaluating patients with dizziness, questions to consider include the following: (1) Is it dangerous (e.g., arrhythmia, transient ischemic attack/stroke)? (2) Is it vestibular? (3) If vestibular, is it peripheral or central? A careful history and examination often provide sufficient information to answer these questions and determine whether additional studies or referral to a specialist is necessary.

APPROACH TO THE PATIENT

HISTORY

When a patient presents with dizziness, the first step is to delineate more precisely the nature of the symptom. In the case of vestibular disorders, the physical symptoms depend on whether the lesion is unilateral or bilateral, and whether it is acute or chronic. Vertigo, an illusion of self or environmental motion, implies an acute asymmetry of vestibular inputs from the two labyrinths or in their central pathways. Symmetric bilateral vestibular hypofunction causes imbalance but no vertigo. Because of the ambiguity in patients' descriptions of their symptoms, diagnosis based simply on symptom characteristics is typically unreliable. Thus, the history should focus closely on other features, including whether this is the first attack, the duration of this and any prior episodes, provoking factors, and accompanying symptoms.

Timing Dizziness can be divided into episodes that last for seconds, minutes, hours, or days. Common causes of brief dizziness (seconds) include benign paroxysmal positional vertigo (BPPV) and orthostatic hypotension, both of which typically are provoked by changes in head and/or body position relative to gravity. Attacks of vestibular migraine and Ménière's disease often last hours. When episodes are of intermediate duration (minutes), transient ischemic attacks of the posterior circulation should be considered, although migraine and other causes are also possible.

Associated Symptoms Symptoms that accompany vertigo may be helpful in distinguishing peripheral vestibular lesions from central causes. Unilateral hearing loss and other acute aural symptoms (ear pain, pressure, fullness, new tinnitus) typically point to a peripheral cause. Because the auditory pathways quickly become bilateral upon entering the brainstem, central lesions are unlikely to cause unilateral hearing loss unless the lesion lies near the root entry zone of the auditory nerve. Symptoms such as double vision, numbness, and limb ataxia suggest a brainstem or cerebellar lesion.

EXAMINATION

Because dizziness and imbalance can be a manifestation of a variety of neurologic disorders, the neurologic examination is essential in the evaluation of these patients. Focus should be given to assessment of eye movements, vestibular function, and hearing.

Ocular Motility The range of eye movements and whether they are equal in each eye should be observed. Peripheral eye movement disorders (e.g., cranial neuropathies, eye muscle weakness) are usually disconjugate (different in the two eyes). One should check pursuit (the ability to follow a smoothly moving target) and saccades (the ability to look back and forth accurately between two targets). Poor pursuit or inaccurate (dysmetric) saccades usually indicate central pathology, often involving the cerebellum. Alignment of the two eyes can be checked with a cover test: while the patient is looking at a target, alternately cover the eyes and observe for corrective saccades. A vertical misalignment may indicate a brainstem or cerebellar lesion. Finally, one should look for spontaneous nystagmus, an involuntary back-and-forth movement of the eyes. Nystagmus is most often of the jerk type, in which a slow drift (slow phase) in one direction alternates with a rapid saccadic movement (quick phase or fast phase) in the opposite direction that resets the position of the eyes in the orbits. Except in the case of acute vestibulopathy (e.g., vestibular neuritis), if nystagmus is

---

[*] Deceased.

TABLE 24-1 Features of Peripheral and Central Vertigo

- Nystagmus from an acute peripheral lesion is unidirectional, with fast phases beating away from the ear with the lesion. Nystagmus that changes direction with gaze is due to a central lesion.
- Transient mixed vertical-torsional nystagmus occurs in benign paroxysmal positional vertigo (BPPV), but pure vertical or pure torsional nystagmus is a central sign.
- Nystagmus from a peripheral lesion may be inhibited by visual fixation, whereas central nystagmus is not suppressed.
- Absence of a head impulse sign in a patient with acute prolonged vertigo should suggest a central cause.
- Unilateral hearing loss suggests peripheral vertigo. Findings such as diplopia, dysarthria, and limb ataxia suggest a central disorder.

easily seen in the light, it is probably due to a central cause. Two forms of nystagmus that are characteristic of lesions of the cerebellar pathways are vertical nystagmus with downward fast phases (downbeat nystagmus) and horizontal nystagmus that changes direction with gaze (gaze-evoked nystagmus). By contrast, peripheral lesions typically cause unidirectional horizontal nystagmus. Use of Frenzel eyeglasses (self-illuminated goggles with convex lenses that blur the patient's vision but allow the examiner to see the eyes greatly magnified) or infrared video goggles can aid in the detection of peripheral vestibular nystagmus, because they reduce the patient's ability to use visual fixation to suppress nystagmus. **Table 24-1** outlines key findings that help distinguish peripheral from central causes of vertigo.

Head Impulse Test The most useful bedside test of peripheral vestibular function is the head impulse test, in which the vestibulo-ocular reflex (VOR) is assessed with small-amplitude (~20 degrees) rapid head rotations. While the patient fixates on a target, the head is rotated quickly to the left or right. If the VOR is deficient, the rotation is followed by a catch-up saccade in the opposite direction (e.g., a leftward saccade after a rightward rotation). The head impulse test can identify both unilateral (catch-up saccades after rotations toward the weak side) and bilateral (catch-up saccades after rotations in both directions) vestibular hypofunction.

Positioning Maneuvers All patients with episodic dizziness, especially if provoked by positional change, should be tested with the Dix-Hallpike maneuver. The patient begins in a sitting position with the head turned 45 degrees; holding the back of the head, the examiner then lowers the patient into a supine position with the head extended backward by about 20 degrees while watching the eyes. Posterior canal BPPV can be diagnosed confidently if transient upbeating-torsional nystagmus is seen. If no nystagmus is observed after 15–20 s, the patient is raised to the sitting position, and the procedure is repeated with the head turned to the other side. Again, Frenzel goggles may improve the sensitivity of the test.

Dynamic Visual Acuity This is a functional test that can be useful in assessing vestibular function. Visual acuity is measured with the head still and when the head is rotated back and forth by the examiner (about 1–2 Hz). A drop in visual acuity during head motion of more than one line on a near card or Snellen chart is abnormal and indicates vestibular dysfunction.

ANCILLARY TESTING

The choice of ancillary tests should be guided by the history and examination findings. Audiometry should be performed whenever a vestibular disorder is suspected. Unilateral sensorineural hearing loss supports a peripheral disorder (e.g., vestibular schwannoma). Predominantly low-frequency hearing loss is characteristic of Ménière's disease. Videonystagmography includes recordings of spontaneous nystagmus (if present) and measurement of positional nystagmus. Caloric testing compares the responses of the two horizontal semicircular canals, while video head-impulse testing measures the integrity of each of the six semicircular canals. Vestibular evoked potentials assess otolith reflexes. The test battery often includes recording of saccades and pursuit to evaluate central ocular motor function. Neuroimaging is important if a central vestibular disorder is suspected. In addition, patients with unexplained unilateral hearing loss or vestibular hypofunction should undergo MRI of the internal auditory canals, including administration of gadolinium, to rule out a schwannoma.

DIFFERENTIAL DIAGNOSIS AND TREATMENT

Treatment of vestibular symptoms should be driven by the underlying diagnosis. Simply treating dizziness with vestibular suppressant medications is often not helpful and may prolong recovery. The diagnostic and specific treatment approaches for the most commonly encountered vestibular disorders are discussed below.

ACUTE PROLONGED VERTIGO (VESTIBULAR NEURITIS)

An acute unilateral vestibular lesion causes constant vertigo, nausea, vomiting, oscillopsia (motion of the visual scene), and imbalance. These symptoms are due to a sudden asymmetry of inputs from the two labyrinths or in their central connections, simulating a continuous rotation of the head. Unlike BPPV, continuous vertigo persists even when the head remains still.

History and Examination When a patient presents with an acute vestibular syndrome, the most important question is whether the lesion is central (e.g., a cerebellar or brainstem infarct or hemorrhage), which may be life-threatening, or peripheral, affecting the vestibular nerve or labyrinth (vestibular neuritis). Attention should be given to any symptoms or signs that point to central dysfunction (diplopia, weakness or numbness, dysarthria). The pattern of spontaneous nystagmus, if present, may be helpful (Table 24-1). If the head impulse test is normal, an acute peripheral vestibular lesion is unlikely. A central lesion cannot always be excluded with certainty based on symptoms and examination alone; thus, older patients with vascular risk factors who present with an acute vestibular syndrome should be evaluated for the possibility of stroke even when there are no specific findings that indicate a central lesion.

Treatment Most patients with vestibular neuritis recover spontaneously, although chronic dizziness, motion sensitivity, and disequilibrium may persist. The role of early glucocorticoid therapy is uncertain, as studies have yielded disparate results. Antiviral medications are of no proven benefit and are not typically given unless there is evidence to suggest herpes zoster oticus (Ramsay Hunt syndrome). Vestibular suppressant medications may reduce acute symptoms but should be avoided after the first several days because they may impede central compensation and recovery. Patients should be encouraged to resume a normal level of activity as soon as possible, and directed vestibular rehabilitation therapy may accelerate improvement.

BENIGN PAROXYSMAL POSITIONAL VERTIGO

BPPV is a common cause of recurrent vertigo. Episodes are brief (<1 min and typically 15–20 s) and are always provoked by changes in head position relative to gravity, such as lying down, rising from a supine position, and extending the head to look upward. Rolling over in bed is a common trigger that may help to distinguish BPPV from orthostatic hypotension. The attacks are caused by free-floating otoconia (calcium carbonate crystals) that have been dislodged from the utricular macula and have moved into one of the semicircular canals, usually the posterior canal. When head position changes, gravity causes the otoconia to move within the canal, producing vertigo and nystagmus. With posterior canal BPPV, the nystagmus beats upward and torsionally (the upper poles of the eyes beat toward the affected lower ear). Less commonly, the otoconia enter the horizontal canal, resulting in a horizontal nystagmus when the patient is lying with either ear down. Superior (also called anterior) canal involvement is rare. BPPV is treated with repositioning maneuvers that use gravity to remove the otoconia from the semicircular canal. For posterior canal BPPV, the Epley maneuver **(Fig. 24-1)** is the most commonly used procedure. For

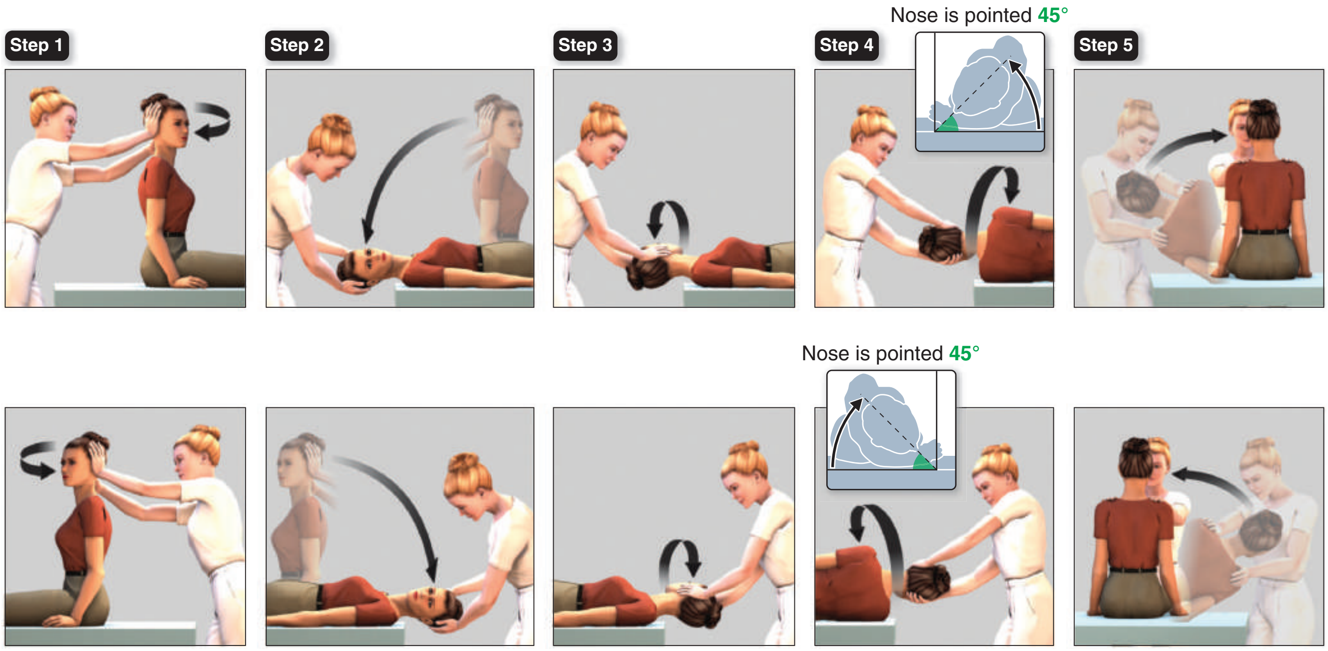

FIGURE 24-1 Modified Epley maneuver for treatment of benign paroxysmal positional vertigo of the right (*top panels*) and left (*bottom panels*) posterior semicircular canals. **Step 1.** With the patient seated, turn the head 45 degrees toward the affected ear. **Step 2.** Keeping the head turned, lower the patient to the head-hanging position and hold for at least 30 s and until nystagmus disappears. **Step 3.** Without lifting the head, turn it 90 degrees toward the other side. Hold for another 30 s. **Step 4.** Rotate the patient onto their side while turning the head another 90 degrees, so that the nose is pointed down 45 degrees. Hold again for 30 s. **Step 5.** Have the patient sit up on the side of the table. After a brief rest, the maneuver should be repeated to confirm successful treatment. *(Reproduced with permission from Chicago Dizziness and Hearing (CDH). Figure adapted from http://www.dizziness-and-balance.com/disorders/bppv/movies/Epley-480x640.avi.)*

more refractory cases of BPPV, patients can be taught a variant of this maneuver that they can perform alone at home. A demonstration of the Epley maneuver is available online (*http://www.dizziness-and-balance.com/disorders/bppv/bppv.html*).

VESTIBULAR MIGRAINE

Vestibular migraine is a common yet underdiagnosed cause of episodic vertigo. Vertigo sometimes precedes a typical migraine headache but more often occurs without headache or with only a mild headache. Some patients who have had frequent migraine headaches in the past present later in life with vestibular migraine as the predominant problem. In vestibular migraine, the duration of vertigo may be from minutes to hours, and some migraineurs also experience more prolonged periods of disequilibrium (lasting days to weeks). Motion sensitivity and sensitivity to visual motion (e.g., movies) are common. Even in the absence of headache, other migraine features may be present, such as photophobia, phonophobia, or a visual aura. Although data from controlled studies are generally lacking, vestibular migraine typically is treated with medications that are used for prophylaxis of migraine headaches **(Chap. 441)**. Antiemetics may be helpful to relieve symptoms at the time of an attack.

MÉNIÈRE'S DISEASE

Attacks of Ménière's disease consist of vertigo and hearing loss, as well as pain, pressure, and/or fullness in the affected ear. Low-frequency hearing loss and aural symptoms are key features that distinguish Ménière's disease from other peripheral vestibulopathies and from vestibular migraine. Audiometry at the time of an attack shows a characteristic asymmetric low-frequency hearing loss; hearing commonly improves between attacks, although permanent hearing loss may eventually occur. Ménière's disease is associated with excess endolymph fluid in the inner ear; hence the term *endolymphatic hydrops*. The exact pathophysiological mechanism, however, remains unclear. Patients suspected of having Ménière's disease should be referred to an otolaryngologist for further evaluation. Diuretics and sodium restriction are typically the initial treatments. If attacks persist, injections of glucocorticoids or gentamicin into the middle ear may be considered. Nonablative surgical options include decompression and shunting of the endolymphatic sac. Full ablative procedures (vestibular nerve section, labyrinthectomy) are seldom required.

VESTIBULAR SCHWANNOMA

Vestibular schwannomas (sometimes termed *acoustic neuromas*) and other tumors at the cerebellopontine angle cause slowly progressive unilateral sensorineural hearing loss and vestibular hypofunction. These patients typically do not have vertigo because the gradual vestibular deficit is compensated centrally as it develops. The diagnosis often is not made until there is sufficient hearing loss to be noticed. The vestibular examination will show a deficient response to the head impulse test when the head is rotated toward the affected side, but nystagmus will not be prominent. As noted above, patients with unexplained unilateral sensorineural hearing loss or vestibular hypofunction require MRI of the internal auditory canals to look for a schwannoma.

BILATERAL VESTIBULAR HYPOFUNCTION

Patients with bilateral loss of vestibular function also typically do not have vertigo, because vestibular function is lost on both sides simultaneously, and there is no asymmetry of vestibular input. Symptoms include loss of balance, particularly in the dark, where vestibular input is most critical, and oscillopsia during head movement, such as while walking or riding in a car. Bilateral vestibular hypofunction may be (1) idiopathic and progressive, (2) part of a neurodegenerative disorder, or (3) iatrogenic due to medication ototoxicity (most commonly gentamicin or other aminoglycoside antibiotics). Other causes include bilateral vestibular schwannomas (neurofibromatosis type 2), autoimmune disease, superficial siderosis, and meningeal-based infection or tumor. It also may occur in patients with peripheral polyneuropathy; in these patients, both vestibular loss and impaired proprioception may contribute to poor balance. Finally, unilateral processes such as vestibular neuritis and Ménière's disease may involve both ears sequentially, resulting in bilateral vestibulopathy.

Examination findings include diminished *dynamic visual acuity* (see above) due to loss of stable vision when the head is moving, abnormal head impulse responses in both directions, and a Romberg sign. Responses to caloric testing are reduced. Patients with bilateral vestibular hypofunction should be referred for vestibular rehabilitation therapy. Vestibular suppressant medications should not be used, as they will increase the imbalance. Evaluation by a neurologist is important not only to confirm the diagnosis but also to consider any other associated neurologic abnormalities that may clarify the etiology.

CENTRAL VESTIBULAR DISORDERS

Central lesions causing vertigo typically involve vestibular pathways in the brainstem and/or cerebellum. They may be due to discrete lesions, such as from ischemic or hemorrhagic stroke **(Chaps. 437–439)**, demyelination **(Chap. 455)**, or tumors **(Chap. 95)**, or they may be due to neurodegenerative conditions that include the vestibulocerebellum **(Chaps. 442–445)**. Subacute cerebellar degeneration may be due to immune, including paraneoplastic, processes **(Chaps. 99 and 450)**. Table 24-1 outlines important features of the history and examination that help to identify central vestibular disorders. Acute central vertigo is a medical emergency, due to the possibility of life-threatening stroke or hemorrhage. All patients with suspected central vestibular disorders should undergo brain MRI, and the patient should be referred for full neurologic evaluation.

FUNCTIONAL DIZZINESS

Psychological factors play an important role in chronic dizziness. First, dizziness may be a somatic manifestation of a psychiatric condition such as major depression, anxiety, or panic disorder **(Chap. 463)**. Second, patients may develop anxiety and autonomic symptoms as a consequence or comorbidity of an independent vestibular disorder. One particular form of this has been termed variously phobic postural vertigo, psychophysiologic vertigo, or chronic subjective dizziness, but is now referred to as persistent postural-perceptual dizziness (PPPD). These patients have a chronic feeling (3 months or longer) of fluctuating dizziness and disequilibrium that is present at rest but worse while standing. There is an increased sensitivity to self-motion and visual motion (e.g., watching movies) and a particular intensification of symptoms when moving through complex visual environments such as supermarkets. Although there may be a past history of an acute vestibular disorder (e.g., vestibular neuritis), the neuro-otologic examination and vestibular testing are normal or indicative of a compensated vestibular deficit, indicating that the ongoing subjective dizziness cannot be explained by a primary vestibular pathology. Anxiety disorders are particularly common in patients with chronic dizziness; when present, they contribute substantially to the morbidity. Treatment approaches for PPPD include pharmacological therapy with selective serotonin reuptake inhibitors (SSRIs), cognitive-behavioral psychotherapy, and vestibular rehabilitation. Vestibular suppressant medications generally should be avoided.

TABLE 24-2 Treatment of Vertigo

AGENT[a]	DOSE[b]
Antihistamines	
Meclizine	25–50 mg 3 times daily
Dimenhydrinate	50 mg 1–2 times daily
Promethazine	25 mg 2–3 times daily (also can be given rectally and IM)
Benzodiazepines	
Diazepam	2.5 mg 1–3 times daily
Clonazepam	0.25 mg 1–3 times daily
Anticholinergic	
Scopolamine transdermal[c]	Patch
Physical therapy	
Repositioning maneuvers[d]	
Vestibular rehabilitation	
Other	
Diuretics and/or low-sodium (1000 mg/d) diet[e]	
Antimigrainous drugs[f]	
Selective serotonin reuptake inhibitors[g]	

[a]All listed drugs are approved by the U.S. Food and Drug Administration, but most are not approved for the treatment of vertigo. [b]Usual oral (unless otherwise stated) starting dose in adults; a higher maintenance dose can be reached by a gradual increase. [c]For motion sickness only. [d]For benign paroxysmal positional vertigo. [e]For Ménière's disease. [f]For vestibular migraine. [g]For persistent postural-perceptual vertigo and anxiety.

TREATMENT

Table 24-2 provides a list of commonly used medications for suppression of vertigo. As noted, these medications should be reserved for short-term control of active vertigo, such as during the first few days of acute vestibular neuritis, or for acute attacks of Ménière's disease. They are less helpful for chronic dizziness and, as previously stated, may hinder central compensation. An exception is that benzodiazepines may attenuate psychosomatic dizziness and the associated anxiety, although SSRIs are generally preferable in such patients.

Vestibular rehabilitation therapy promotes central adaptation processes that compensate for vestibular loss and also may help habituate motion sensitivity and other symptoms of perceptual dizziness. The general approach is to use a graded series of exercises that progressively challenge gaze stabilization and balance. For patients with bilateral vestibular hypofunction, an implanted vestibular prosthesis has shown promise as a future option.

FURTHER READING

Altissimi G et al: Drugs inducing hearing loss, tinnitus, dizziness and vertigo: An updated guide. Eur Rev Med Pharmacol Sci 24:7946, 2020.

Kim JS, Zee DS: Benign paroxysmal positional vertigo. N Engl J Med 370:1138, 2014.

Smyth D et al: Vestibular migraine treatment: A comprehensive practical review. Brain 145:3741, 2022.

Staab JP: Persistent postural-perceptual dizziness. Neurol Clin 41:647, 2023.

25

Fatigue

Jeffrey M. Gelfand, Vanja C. Douglas

Fatigue is one of the most common symptoms in clinical medicine. It is a prominent manifestation of a number of systemic, neurologic, and psychiatric syndromes, although a precise cause will not be identified in a substantial minority of patients. Fatigue refers to the subjective experience of physical and mental weariness, sluggishness, low energy, and exhaustion. In the context of clinical medicine, fatigue is most practically defined as difficulty initiating or maintaining voluntary mental or physical activity. Nearly everyone who has ever been ill with a self-limited infection has experienced this near-universal symptom, and fatigue is usually brought to medical attention only when it is either of unclear cause, fails to remit, or the severity is out of proportion with what would be expected for the associated trigger.

Fatigue should be distinguished from *muscle weakness*, a reduction of neuromuscular power **(Chap. 26)**; most patients complaining of fatigue are not truly weak when direct muscle power is tested. Fatigue is also distinct from *somnolence*, which refers to sleepiness in the context of disturbed sleep-wake physiology **(Chap. 33)**, and from *dyspnea on exertion*, although patients may use the word fatigue to describe any of these symptoms. The task facing clinicians when a patient presents with fatigue is to identify the underlying cause and develop a therapeutic alliance, the goal of which is to spare patients expensive and fruitless diagnostic workups and steer them toward effective therapy.

EPIDEMIOLOGY AND GLOBAL CONSIDERATIONS

Variability in the definitions of fatigue and the survey instruments used in different studies makes it difficult to arrive at precise figures about the global burden of fatigue. The point prevalence of fatigue was 6.7% and the lifetime prevalence was 25% in a large National Institute of Mental Health survey of the U.S. general population. In primary care clinics in Europe and the United States, between 10 and 25% of patients surveyed endorsed symptoms of prolonged (present for >1 month) or chronic (present for >6 months) fatigue, but in only a minority was fatigue the primary reason for seeking medical attention. In a community survey of women in India, 12% reported chronic fatigue. By contrast, the prevalence of myalgic encephalomyelitis/chronic fatigue syndrome **(Chap. 461)**, as defined by the U.S. Centers for Disease Control and Prevention, is low.

DIFFERENTIAL DIAGNOSIS

Psychiatric Disease Fatigue is a common somatic manifestation of many major psychiatric syndromes, including depression, anxiety, and somatoform disorders **(Chap. 463)**. Psychiatric symptoms are reported in more than three-quarters of patients with unexplained chronic fatigue. Even in patients with systemic or neurologic disorders in which fatigue is independently recognized as a symptom, comorbid psychiatric disease may still be an important contributor.

Neurologic Disease Patients with fatigue often say they feel weak, but upon careful examination, objective muscle weakness is rarely discernible. If found, muscle weakness must then be localized to the central nervous system, peripheral nervous system, neuromuscular junction, or muscle, and appropriate follow-up studies obtained **(Chap. 26)**. *Fatigability* of muscle power is a cardinal manifestation of some neuromuscular disorders such as myasthenia gravis and is distinguished from *fatigue* by finding clinically evident diminution of the amount of force that a muscle generates upon repeated contraction **(Chap. 459)**. Fatigue is one of the most common and bothersome symptoms reported in multiple sclerosis (MS) **(Chap. 455)**, affecting nearly 90% of patients; fatigue in MS can persist between MS attacks and does not necessarily correlate with magnetic resonance imaging (MRI) disease activity. Fatigue is also increasingly identified as a troublesome feature of many neurodegenerative diseases, including Parkinson's disease **(Chap. 446)**, amyotrophic lateral sclerosis **(Chap. 448)**, and central nervous system dysautonomias **(Chap. 451)**. Fatigue after stroke **(Chap. 437)** is a well-described but poorly understood entity with a widely varying prevalence. Episodic fatigue can be a premonitory symptom of migraine **(Chap. 441)**. Fatigue is also a frequent consequence of traumatic brain injury **(Chap. 454)**, often occurring in association with depression and sleep disorders.

Sleep Disorders Obstructive sleep apnea is an important cause of excessive daytime sleepiness in association with fatigue and should be investigated using overnight polysomnography, particularly in those with prominent snoring, obesity, or other predictors of obstructive sleep apnea **(Chap. 308)**. Whether the cumulative sleep deprivation that is common in modern society contributes to clinically apparent fatigue is not known **(Chap. 33)**.

Endocrine Disorders Fatigue, sometimes in association with true muscle weakness, can be a heralding symptom of hypothyroidism **(Chap. 395)**, particularly in the context of hair loss, dry skin, cold intolerance, constipation, and weight gain. Fatigue associated with heat intolerance, sweating, and palpitations is typical of hyperthyroidism **(Chap. 396)**. Adrenal insufficiency **(Chap. 398)** can also manifest with unexplained fatigue as a primary or prominent symptom, often with anorexia, weight loss, nausea, myalgias, and arthralgias; hyponatremia, hyperkalemia, and hyperpigmentation may be present at time of diagnosis. Mild hypercalcemia can cause fatigue, which may be relatively vague, whereas severe hypercalcemia can lead to lethargy, stupor, and coma **(Chap. 422)**. Both hypoglycemia and hyperglycemia can cause lethargy, often in association with confusion; diabetes mellitus, in particular type 1 diabetes, is also associated with fatigue independent of glucose levels **(Chap. 415)**. Fatigue may also accompany Cushing's disease, hypoaldosteronism, and hypogonadism. Low vitamin D status has also been associated with fatigue.

Liver and Kidney Disease Both chronic liver failure and chronic kidney disease can cause fatigue. Over 80% of hemodialysis patients complain of fatigue, which makes it one of the most common symptoms reported by patients in chronic kidney disease **(Chap. 322)**.

Obesity Obesity **(Chap. 413)** is associated with fatigue and sleepiness independent of the presence of obstructive sleep apnea. Obese patients undergoing bariatric surgery experience improvement in daytime sleepiness sooner than would be expected if the improvement were solely the result of weight loss and resolution of sleep apnea. A number of other factors common in obese patients are likely contributors as well, including physical inactivity, diabetes, and depression.

Physical Inactivity Physical inactivity is associated with fatigue, and increasing physical activity can improve fatigue in some patients.

Malnutrition Although fatigue can be a presenting feature of malnutrition **(Chap. 345)**, nutritional status may also be an important comorbidity and contributor to fatigue in other chronic illnesses, including cancer-associated fatigue.

Infection Both acute and chronic infections commonly lead to fatigue as part of the broader infectious syndrome. Evaluation for undiagnosed infection as the cause of unexplained fatigue, and particularly prolonged or chronic fatigue, should be guided by the history, physical examination, and infectious risk factors, with particular attention to risk for tuberculosis, HIV, chronic hepatitis, and endocarditis. Infectious mononucleosis may cause prolonged fatigue that persists for weeks to months following the acute illness, but infection with the Epstein-Barr virus is only very rarely the cause of unexplained chronic fatigue. Postinfectious fatigue may also occur following a variety of acute infections. For example, a substantial minority of patients who have recovered from SARS-CoV-1, SARS-CoV-2, Dengue, and Ebola virus experience persistent fatigue. Almost one-third of patients report fatigue 3 or more months following SARS-CoV-2 (COVID-19) diagnosis.

Drugs Many medications, drugs, drug withdrawal, and chronic alcohol use can all lead to fatigue. Medications that are more likely to be causative include antidepressants, antipsychotics, anxiolytics, opiates, antispasticity agents, antiseizure agents, and beta blockers.

Cardiovascular and Pulmonary Disorders Fatigue is one of the most taxing symptoms reported by patients with congestive heart failure and chronic obstructive pulmonary disease and negatively affects quality of life. In a population-based cohort study in Norfolk, United Kingdom, fatigue was associated with an increased hazard of all-cause mortality in the general population, but particularly for deaths related to cardiovascular disease.

Malignancy Fatigue, particularly in association with unexplained weight loss, can be a sign of occult malignancy, but cancer is rarely identified in patients with unexplained chronic fatigue in the absence of other telltale signs or symptoms. Cancer-related fatigue is experienced by 40% of patients at the time of diagnosis and by >80% at some time in the disease course.

Hematologic Disorders Chronic or progressive anemia may present with fatigue, sometimes in association with exertional tachycardia and breathlessness. Anemia may also contribute to fatigue in chronic illness. Low serum ferritin in the absence of anemia may also cause fatigue that is reversible with iron replacement.

Immune-Mediated Disorders Fatigue is a prominent complaint in many chronic inflammatory disorders, including systemic lupus erythematosus, polymyalgia rheumatica, rheumatoid arthritis, inflammatory bowel disease, antineutrophil cytoplasmic antibody (ANCA)–associated vasculitis, sarcoidosis, and Sjögren's syndrome, but is not usually an isolated symptom. Fatigue is also associated with primary immunodeficiency diseases.

Pregnancy Fatigue is very commonly reported by women during all stages of pregnancy and postpartum.

Disorders of Unclear Cause Myalgic encephalomyelitis (ME)/chronic fatigue syndrome (CFS) **(Chap. 461)** and fibromyalgia **(Chap. 385)** incorporate chronic fatigue as part of the syndromic definition when fatigue is present in association with other criteria, as discussed in the respective chapters. Chronic multisymptom illness, also known as Gulf-War syndrome, is another symptom complex with prominent fatigue; it is most commonly, although not exclusively, observed in veterans of the 1991 Gulf War conflict **(Chap. S8)**. Idiopathic chronic fatigue is used to describe the syndrome of unexplained chronic fatigue in the absence of enough additional clinical features to meet the diagnostic criteria for ME/CFS.

APPROACH TO THE PATIENT

Fatigue

A detailed history focusing on the quality, pattern, time course, associated symptoms, and alleviating factors of fatigue is necessary to define the syndrome and help direct further evaluation and treatment. It is important to determine if fatigue is the appropriate designation, whether symptoms are acute or chronic, and if the impairment is primarily mental, physical, or a combination of the two. The review of systems should attempt to distinguish fatigue from excessive sleepiness, dyspnea on exertion, exercise intolerance, and muscle weakness. The presence of fever, chills, night sweats, or weight loss should raise suspicion for an occult infection or malignancy. A careful review of prescription, over-the-counter, herbal, and recreational drug and alcohol use is required. Circumstances surrounding the onset of symptoms and potential triggers should be investigated. The social history is important, with attention paid to life stressors and adverse experiences, workhours, the social support network, and domestic affairs including a screen for intimate partner violence. Sleep habits and sleep hygiene should be questioned. The impact of fatigue on daily functioning is important to understand the patient's experience and gauge recovery and the success of treatment.

The physical examination of patients with fatigue is guided by the history and differential diagnosis. A detailed mental status examination should be performed with particular attention to symptoms of depression and anxiety. A formal neurologic examination is required to determine whether objective muscle weakness is present. This is usually a straightforward exercise, although occasionally patients with fatigue have difficulty sustaining effort against resistance and sometimes report that generating full power requires substantial mental effort. On confrontational testing, full power may be generated for only a brief period before the patient suddenly gives way to the examiner. This type of weakness is often referred to as *breakaway weakness* and may or may not be associated with pain. This is contrasted with weakness due to lesions in the motor tracts or lower motor unit, in which the patient's resistance can be overcome in a smooth and steady fashion and full power can never be generated. Occasionally, a patient may demonstrate fatigable weakness, in which power is full when first tested but becomes weak upon repeat evaluation without interval rest. Fatigable weakness, which usually indicates a problem of neuromuscular transmission, never has the sudden breakaway quality that one occasionally observes in patients with fatigue. If the presence or absence of muscle weakness cannot be determined with the physical examination, electromyography with nerve conduction studies can be a helpful ancillary test.

The general physical examination should screen for signs of cardiopulmonary disease, malignancy, lymphadenopathy, organomegaly, infection, liver failure, kidney disease, malnutrition, endocrine abnormalities, and connective tissue disease. In patients with associated widespread musculoskeletal pain, assessment of tender points may help to reveal fibromyalgia. Although the diagnostic yield of the general physical examination may be relatively low in the context of evaluation of unexplained chronic fatigue, elucidating the cause of only 2% of cases in one prospective analysis, the yield of a detailed neuropsychiatric and mental status evaluation is likely to be much higher, revealing a potential explanation for fatigue in up to 75–80% of patients in some series. Furthermore, a complete physical examination demonstrates a serious and systematic approach to the patient's complaint and helps build trust and a therapeutic alliance.

Laboratory testing is likely to identify the cause of chronic fatigue in only about 5% of cases. Beyond a few standard screening tests, laboratory evaluation should be guided by the history and physical examination; extensive testing is likely to lead to incidental findings that require explanation and unnecessary follow-up investigation and should be avoided in lieu of frequent clinical follow-up. A reasonable approach to screening includes a complete blood count with differential (to screen for anemia, infection, and malignancy), electrolytes (including sodium, potassium, and calcium), glucose, renal function, liver function, and thyroid function. Testing for HIV and adrenal function can also be considered. Published guidelines for ME/CFS also recommend an erythrocyte sedimentation rate (ESR) as part of the evaluation for mimics, but unless the value is very high, such nonspecific testing in the absence of other features is unlikely to clarify the situation. Routine screening with an antinuclear antibody (ANA) test is also unlikely to be informative in isolation and is frequently positive at low titers in otherwise healthy adults. Additional unfocused studies, such as whole-body imaging scans, are usually not indicated; in addition to their inconvenience, potential risk, and cost, they often reveal unrelated incidental findings that can prolong the workup unnecessarily.

TREATMENT

Fatigue

The first priority is to address the underlying disorder or disorders that account for fatigue, because this can be curative in select contexts and palliative in others. Unfortunately, in many chronic illnesses, fatigue may be refractory to traditional disease-modifying therapies, but it is nevertheless important in such cases to evaluate for other potential contributors because the cause may be multifactorial. Antidepressants **(Chap. 463)** may be helpful for treatment of chronic fatigue when symptoms of depression are present and are generally most effective as part of a multimodal approach. However, antidepressants can also cause fatigue and should be discontinued if they are not clearly effective. Cognitive-behavioral therapy has also been demonstrated to be helpful in ME/CFS as well as cancer-associated fatigue. Both cognitive-behavioral therapy and graded exercise therapy, in which physical exercise, most typically walking, is gradually increased with attention to target heart rates to avoid overexertion, were shown to modestly improve walking times and self-reported fatigue measures when compared to standard medical care in patients in the United Kingdom with chronic fatigue. These benefits were maintained after a median follow-up of 2.5 years. Exercise as an intervention has also demonstrated some benefit for patients with fatigue related to cancer, MS, and diabetes, among other conditions. Psychostimulants such as amphetamines, modafinil, and armodafinil can help increase alertness and concentration and reduce excessive daytime sleepiness in certain clinical contexts, which may in turn help with symptoms of fatigue in a minority of patients, but they have generally proven to be unhelpful in randomized trials for treating fatigue in posttraumatic brain injury, Parkinson's disease, cancer, and MS. In patients with low vitamin D status, vitamin D replacement may lead to improvement in fatigue.

Development of more effective therapy for fatigue is hampered by limited knowledge of the biologic basis of this symptom, including how fatigue is detected and registered in the nervous system.

Proinflammatory cytokines, such as interleukin 1α and 1β and tumor necrosis factor α, might mediate fatigue in some patients. While preliminary studies of biologic therapies that inhibit cytokines have suggested a benefit against fatigue in some patients with inflammatory conditions, this approach has largely not led to improvement in clinical trials that focused on fatigue as the primary endpoint. Nonetheless, specific targeting with cytokine antagonists could represent a possible future approach for some patients.

PROGNOSIS

Acute fatigue significant enough to require medical evaluation is more likely to lead to an identifiable medical, neurologic, or psychiatric cause than is unexplained chronic fatigue. Evaluation of unexplained chronic fatigue most commonly leads to diagnosis of a psychiatric condition or remains unexplained. Identification of a previously undiagnosed serious or life-threatening culprit etiology is rare, even with longitudinal follow-up of patients with unexplained chronic fatigue. Complete resolution is uncommon, at least over the short term, but multidisciplinary treatment approaches can lead to symptomatic improvements that substantially improve quality of life.

FURTHER READING

Basu N et al: Fatigue is associated with excess mortality in the general population: Results from the EPIC-Norfolk study. BMC Med 14:122, 2016.

Ceban F et al: Fatigue and cognitive impairment in post-COVID-19 syndrome: A systematic review and meta-analysis. Brain Behav Immun 101:93, 2022.

Dukes JC et al: Approach to fatigue: Best practice. Med Clin North Am 105:137, 2021.

Roerink ME et al: Interleukin-1 as a mediator of fatigue in disease: A narrative review. J Neuroinflammation 14:16, 2017.

Sharpe M et al: Rehabilitative treatments for chronic fatigue syndrome: Long-term follow-up from the PACE trial. Lancet Psychiatry 2:1067, 2015.

26

Neurologic Causes of Weakness and Paralysis

Stephen L. Hauser

Normal motor function involves integrated muscle activity that is modulated by the activity of the cerebral cortex, basal ganglia, cerebellum, red nucleus, brainstem reticular formation, lateral vestibular nucleus, and spinal cord. Motor system dysfunction leads to weakness or paralysis, discussed in this chapter, or to ataxia **(Chap. 450)** or abnormal movements **(Chap. 447)**. *Weakness* is a reduction in the power that can be exerted by one or more muscles. It must be distinguished from increased *fatigability* (i.e., the inability to sustain the performance of an activity that should be normal for a person of the same age, sex, and size), limitation in function due to pain or articular stiffness, or impaired motor activity because severe *proprioceptive sensory loss* prevents adequate feedback information about the direction and power of movements. It is also distinct from *bradykinesia* (in which increased time is required for full power to be exerted) and *apraxia*, a disorder of planning and initiating a skilled or learned movement unrelated to a significant motor or sensory deficit **(Chap. 32)**.

Paralysis or the suffix "-plegia" indicates weakness so severe that a muscle cannot be contracted at all, whereas *paresis* refers to less severe weakness. The prefix "hemi-" refers to one-half of the body, "para-" to both legs, and "quadri-" to all four limbs.

The *distribution* of weakness helps to localize the underlying lesion. Weakness from involvement of upper motor neurons occurs particularly in the extensors and abductors of the upper limb and the flexors of the lower limb. Lower motor neuron weakness depends on whether involvement is at the level of the anterior horn cells, nerve root, limb plexus, or peripheral nerve—only muscles supplied by the affected structure are weak. Myopathic weakness is generally most marked in proximal muscles. Weakness from impaired neuromuscular transmission has no specific pattern of involvement.

Weakness often is accompanied by other neurologic abnormalities that help indicate the site of the responsible lesion **(Table 26-1)**.

Tone is the resistance of a muscle to passive stretch. Increased tone may be of several types. *Spasticity* is the increase in tone associated with disease of upper motor neurons. It is velocity dependent, has a sudden release after reaching a maximum (the "clasp-knife" phenomenon), and predominantly affects the antigravity muscles (i.e., upper-limb flexors and lower-limb extensors). *Rigidity* is hypertonia that is present throughout the range of motion (a "lead pipe" or "plastic" stiffness) and affects flexors and extensors equally; it sometimes has a cogwheel quality that is enhanced by voluntary movement of the contralateral limb (reinforcement). Rigidity occurs with extrapyramidal disorders such as Parkinson's disease. *Paratonia* (or *gegenhalten*) is increased tone that varies irregularly in a manner seemingly related to the degree of relaxation, is present throughout the range of motion, and affects flexors and extensors equally; it usually results from disease of the frontal lobes. Weakness with *decreased tone (flaccidity)* or normal tone occurs with disorders of *motor units*. A motor unit consists of a single lower motor neuron and all the muscle fibers that it innervates.

Muscle bulk generally is not affected by upper motor neuron lesions, although mild disuse atrophy eventually occurs. By contrast, atrophy is often conspicuous when a lower motor neuron lesion is responsible for weakness and also may occur with advanced muscle disease.

Muscle stretch (tendon) reflexes are usually increased with upper motor neuron lesions but may be decreased or absent for a variable period immediately after onset of an acute lesion. Hyperreflexia is usually—but not invariably—accompanied by loss of *cutaneous reflexes* (such as superficial abdominals; **Chap. 433**) and, in particular, by an extensor plantar (Babinski) response. The muscle stretch reflexes are depressed with lower motor neuron lesions involving specific reflex arcs. They generally are preserved in patients with myopathic weakness, except in advanced stages when they sometimes are attenuated. In disorders of the neuromuscular junction, reflex responses may be affected by preceding voluntary activity of affected muscles; such

TABLE 26-1 Signs That Distinguish the Origin of Weakness

SIGN	UPPER MOTOR NEURON	LOWER MOTOR NEURON	MYOPATHIC	PSYCHOGENIC
Atrophy	None	Severe	Mild	None
Fasciculations	None	Common	None	None
Tone	Spastic	Decreased	Normal/decreased	Variable/paratonia
Distribution of weakness	Pyramidal/regional	Distal/segmental	Proximal	Variable/inconsistent with daily activities
Muscle stretch reflexes	Hyperactive	Hypoactive/absent	Normal/hypoactive	Normal
Babinski sign	Present	Absent	Absent	Absent

activity may lead to enhancement of initially depressed reflexes in Lambert-Eaton myasthenic syndrome and, conversely, to depression of initially normal reflexes in myasthenia gravis **(Chap. 459)**.

The distinction of *neuropathic* (lower motor neuron) from *myopathic* weakness is sometimes difficult clinically, although distal weakness is likely to be neuropathic, and symmetric proximal weakness myopathic. *Fasciculations* (visible or palpable twitches within a muscle due to the spontaneous discharge of a motor unit) and early atrophy indicate that weakness is neuropathic.

PATHOGENESIS

Upper Motor Neuron Weakness Lesions of the upper motor neurons or their descending axons to the spinal cord **(Fig. 26-1)** produce weakness through decreased activation of lower motor neurons. In general, distal muscle groups are affected more severely than proximal ones, and axial movements are spared unless the lesion is severe and bilateral. Spasticity is typical but may not be present acutely. Rapid repetitive movements are slowed and coarse, but normal rhythmicity is maintained. With corticobulbar involvement, weakness occurs in the lower face and tongue; extraocular, upper facial, pharyngeal, and jaw muscles are typically spared. Bilateral corticobulbar lesions produce a *pseudobulbar palsy*: dysarthria, dysphagia, dysphonia, and emotional lability accompany bilateral facial weakness and a brisk jaw jerk. Electromyogram (EMG) **(Chap. 457)** shows that with weakness of the upper motor neuron type, motor units have a diminished maximal discharge frequency.

Lower Motor Neuron Weakness This pattern results from disorders of lower motor neurons in the brainstem motor nuclei and the anterior horn of the spinal cord or from dysfunction of the axons of these neurons as they pass to skeletal muscle **(Fig. 26-2)**. Weakness is due to a decrease in the number of muscle fibers that can be activated through a loss of α motor neurons or disruption of their connections to muscle. Loss of γ motor neurons does not cause weakness but decreases tension on the muscle spindles, which decreases muscle tone and attenuates the stretch reflexes. An absent stretch reflex suggests involvement of spindle afferent fibers.

When a motor unit becomes diseased, especially in anterior horn cell diseases, it may discharge spontaneously, producing *fasciculations*. When α motor neurons or their axons degenerate, the denervated muscle fibers also may discharge spontaneously. These single muscle fiber discharges, or *fibrillation potentials,* cannot be seen but can be recorded with EMG. Weakness leads to delayed or reduced recruitment of motor units, with fewer than normal activated at a particular discharge frequency.

Neuromuscular Junction Weakness Disorders of the neuromuscular junction produce weakness of variable degree and distribution. The number of muscle fibers that are activated varies over time, depending on the state of rest of the neuromuscular junctions. Strength is influenced by preceding activity of the affected muscle. In myasthenia gravis, for example, sustained or repeated contractions of affected muscle decline in strength despite continuing effort **(Chap. 459)**. Thus, fatigable weakness is suggestive of disorders of the neuromuscular junction, which cause functional loss of muscle fibers due to failure of their activation.

Myopathic Weakness Myopathic weakness is produced by a decrease in the number or contractile force of muscle fibers activated within motor units. With muscular dystrophies, inflammatory myopathies, or myopathies with muscle fiber necrosis, the number of muscle fibers is reduced within many motor units. On EMG, the size of each motor unit action potential is decreased, and motor units must be recruited more rapidly than normal to produce the desired power. Some myopathies produce weakness through loss of contractile force of muscle fibers or through relatively selective involvement of type II (fast) fibers. These myopathies may not affect the size of individual motor unit action potentials and are detected by a discrepancy between the electrical activity and force of a muscle.

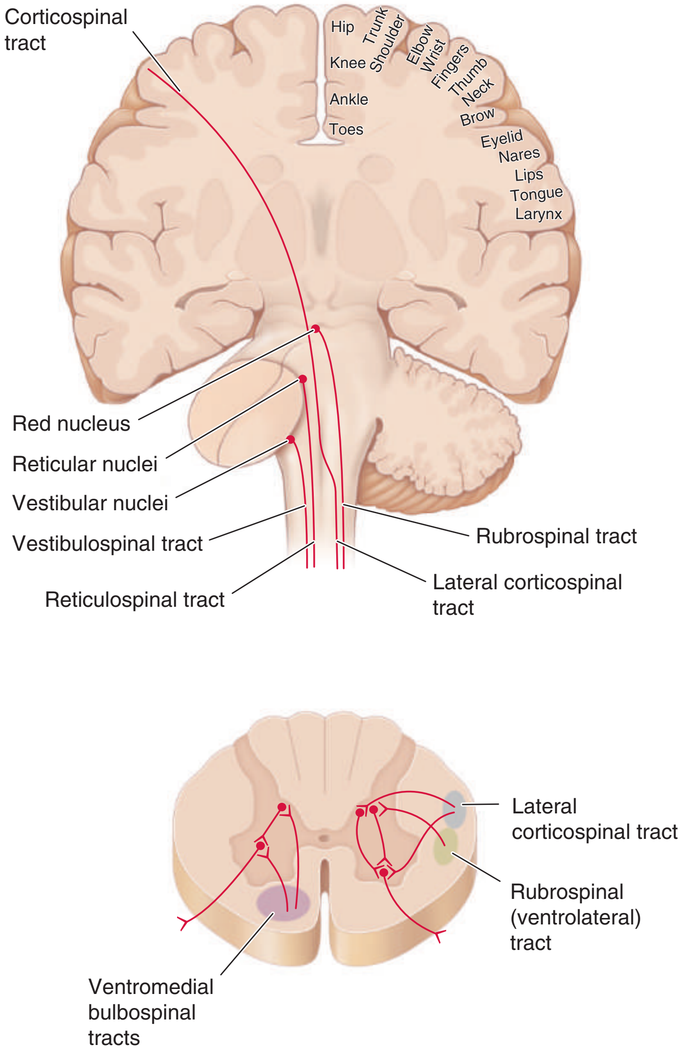

FIGURE 26-1 The corticospinal and bulbospinal upper motor neuron pathways. Upper motor neurons have their cell bodies in layer V of the primary motor cortex (the precentral gyrus, or Brodmann area 4) and in the premotor and supplemental motor cortex (area 6). The upper motor neurons in the primary motor cortex are somatotopically organized (*right side of figure*). Axons of the upper motor neurons descend through the subcortical white matter and the posterior limb of the internal capsule. Axons of the pyramidal or corticospinal system descend through the brainstem in the cerebral peduncle of the midbrain, the basis pontis, and the medullary pyramids. At the cervicomedullary junction, most corticospinal axons decussate into the contralateral corticospinal tract of the lateral spinal cord, but 10–30% remain ipsilateral in the anterior spinal cord. Corticospinal neurons synapse on premotor interneurons, but some—especially in the cervical enlargement and those connecting with motor neurons to distal limb muscles—make direct monosynaptic connections with lower motor neurons. They innervate most densely the lower motor neurons of hand muscles and are involved in the execution of learned, fine movements. Corticobulbar neurons are similar to corticospinal neurons but innervate brainstem motor nuclei. Bulbospinal upper motor neurons influence strength and tone but are not part of the pyramidal system. The descending ventromedial bulbospinal pathways originate in the tectum of the midbrain (tectospinal pathway), the vestibular nuclei (vestibulospinal pathway), and the reticular formation (reticulospinal pathway). These pathways influence axial and proximal muscles and are involved in the maintenance of posture and integrated movements of the limbs and trunk. The descending ventrolateral bulbospinal pathways, which originate predominantly in the red nucleus (rubrospinal pathway), facilitate distal limb muscles. The bulbospinal system sometimes is referred to as the extrapyramidal upper motor neuron system. In all figures, nerve cell bodies and axon terminals are shown, respectively, as *closed circles* and *forks*.

Psychogenic Weakness Weakness may occur without a recognizable organic basis. It tends to be variable, inconsistent, and with a pattern of distribution that cannot be explained on a neuroanatomic basis. On formal testing, antagonists may contract when the patient is supposedly activating the agonist muscle. The severity of weakness is out of keeping with the patient's daily activities.

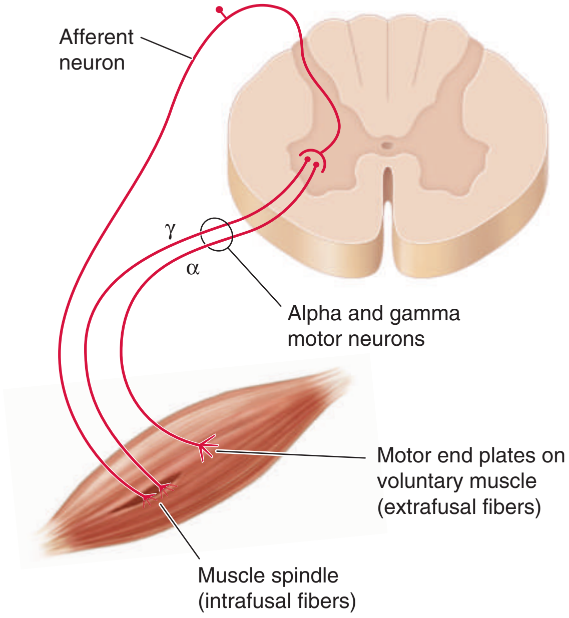
Afferent neuron
γ
α
Alpha and gamma motor neurons
Motor end plates on voluntary muscle (extrafusal fibers)
Muscle spindle (intrafusal fibers)

FIGURE 26-2 Lower motor neurons are divided into α and γ types. The larger α motor neurons are more numerous and innervate the extrafusal muscle fibers of the motor unit. Loss of α motor neurons or disruption of their axons produces lower motor neuron weakness. The smaller, less numerous γ motor neurons innervate the intrafusal muscle fibers of the muscle spindle and contribute to normal tone and stretch reflexes. The α motor neuron receives direct excitatory input from corticomotoneurons and primary muscle spindle afferents. The α and γ motor neurons also receive excitatory input from other descending upper motor neuron pathways, segmental sensory inputs, and interneurons. The α motor neurons receive direct inhibition from Renshaw cell interneurons, and other interneurons indirectly inhibit the α and γ motor neurons. A muscle stretch (tendon) reflex requires the function of all the illustrated structures. A tap on a tendon stretches muscle spindles (which are tonically activated by γ motor neurons) and activates the primary spindle afferent neurons. These neurons stimulate the α motor neurons in the spinal cord, producing a brief muscle contraction, which is the familiar tendon reflex.

DISTRIBUTION OF WEAKNESS

Hemiparesis Hemiparesis results from an upper motor neuron lesion above the midcervical spinal cord; most such lesions are above the foramen magnum. The presence of other neurologic deficits helps localize the lesion. Thus, language disorders, for example, point to a cortical lesion. Homonymous visual field defects reflect either a cortical or a subcortical hemispheric lesion. A "pure motor" hemiparesis of the face, arm, and leg often is due to a small, discrete lesion in the posterior limb of the internal capsule, cerebral peduncle in the midbrain, or upper pons. Some brainstem lesions produce "crossed paralyses," consisting of ipsilateral cranial nerve signs and contralateral hemiparesis **(Chap. 437)**. The absence of cranial nerve signs or facial weakness suggests that a hemiparesis is due to a lesion in the high cervical spinal cord, especially if associated with Brown-Séquard syndrome, consisting of loss of joint position and vibration sense on the side of the weakness, and loss of pain and temperature sense on the opposite side **(Chap. 453)**.

Acute or episodic hemiparesis usually results from focal structural lesions, particularly vascular etiologies, rapidly expanding lesions, or an inflammatory process. *Subacute hemiparesis* that evolves over days or weeks may be caused by subdural hematoma, infectious or inflammatory disorders (e.g., cerebral abscess, fungal granuloma or meningitis, parasitic infection, multiple sclerosis, sarcoidosis), or primary or metastatic neoplasms. AIDS may present with subacute hemiparesis due to toxoplasmosis or primary central nervous system (CNS) lymphoma. *Chronic hemiparesis* that evolves over months usually is due to a neoplasm or vascular malformation, chronic subdural hematoma, or a degenerative disease.

Investigation of hemiparesis **(Fig. 26-3)** of acute origin usually starts with a CT scan of the brain and laboratory studies. If the CT is normal, or in subacute or chronic cases of hemiparesis, MRI of the brain and/or cervical spine (including the foramen magnum) is performed, depending on the clinical accompaniments.

Paraparesis ***Acute paraparesis*** is caused most commonly by an intraspinal lesion, but its spinal cord origin may not be recognized initially if the legs are flaccid and areflexic. Usually, however, there is sensory loss in the legs with an upper level on the trunk; a dissociated sensory loss (loss of pain and temperature but not touch, position, and vibration sense) suggestive of a central cord syndrome; or hyperreflexia in the legs with normal reflexes in the arms **(Chap. 453)**. Imaging the spinal cord (Fig. 26-3) may reveal compressive lesions, infarction (proprioception usually is spared), arteriovenous fistulas or other vascular anomalies, or transverse myelitis **(Chap. 453)**.

Diseases of the cerebral hemispheres that produce acute paraparesis include anterior cerebral artery ischemia (shoulder shrug also is affected), superior sagittal sinus or cortical venous thrombosis, and acute hydrocephalus.

Paraparesis may also result from a cauda equina syndrome, for example, after trauma to the low back, a midline disk herniation, or an intraspinal tumor. The sphincters are commonly affected, whereas hip flexion often is spared, as is sensation over the anterolateral thighs. Rarely, paraparesis is caused by a rapidly evolving anterior horn cell disease (such as poliovirus or West Nile virus infection), peripheral neuropathy (such as Guillain-Barré syndrome; **Chap. 458**), or myopathy **(Chap. 460)**.

Subacute or chronic spastic paraparesis is caused by upper motor neuron disease. When associated with lower-limb sensory loss and sphincter involvement, a chronic spinal cord disorder should be considered **(Chap. 453)**. If hemispheric signs are present, a parasagittal meningioma or chronic hydrocephalus is likely. The absence of spasticity in a long-standing paraparesis suggests a lower motor neuron or myopathic etiology.

Investigations typically begin with spinal MRI, but when upper motor neuron signs are associated with drowsiness, confusion, seizures, or other hemispheric signs, brain MRI should also be performed. Electrophysiologic studies are diagnostically helpful when clinical findings suggest an underlying neuromuscular disorder.

Quadriparesis or Generalized Weakness Generalized weakness may be due to disorders of the CNS or the motor unit. Although the terms often are used interchangeably, *quadriparesis* is commonly used when an upper motor neuron cause is suspected, and *generalized weakness* is used when a disease of the motor units is likely. Weakness from CNS disorders usually is associated with changes in consciousness or cognition and accompanied by spasticity, hyperreflexia, and sensory disturbances. Most neuromuscular causes of generalized weakness are associated with normal mental function, hypotonia, and hypoactive muscle stretch reflexes. The major causes of intermittent weakness are listed in **Table 26-2**. A patient with generalized fatigability without objective weakness may have myalgic encephalomyelitis/chronic fatigue syndrome **(Chap. 461)**.

ACUTE QUADRIPARESIS Quadriparesis with onset over minutes may result from disorders of upper motor neurons (such as from anoxia, hypotension, brainstem or cervical cord ischemia, trauma, and systemic metabolic abnormalities) or muscle (electrolyte disturbances, certain inborn errors of muscle energy metabolism, toxins, and periodic paralyses). Onset over hours to weeks may, in addition to these disorders, be due to lower motor neuron disorders such as Guillain-Barré syndrome **(Chap. 458)**.

In obtunded patients, evaluation begins with a CT or MRI scan of the brain. If upper motor neuron signs are present but the patient is alert, the initial test is usually an MRI of the cervical cord. If weakness is lower motor neuron, myopathic, or uncertain in origin, the clinical approach begins with blood studies to determine the level of muscle enzymes and electrolytes and with EMG and nerve conduction studies.

SUBACUTE OR CHRONIC QUADRIPARESIS Quadriparesis due to upper motor neuron disease may develop over weeks to years from chronic myelopathies, multiple sclerosis, brain or spinal tumors, chronic subdural hematomas, and various metabolic, toxic, and infectious disorders. It may also result from lower motor neuron disease, a chronic

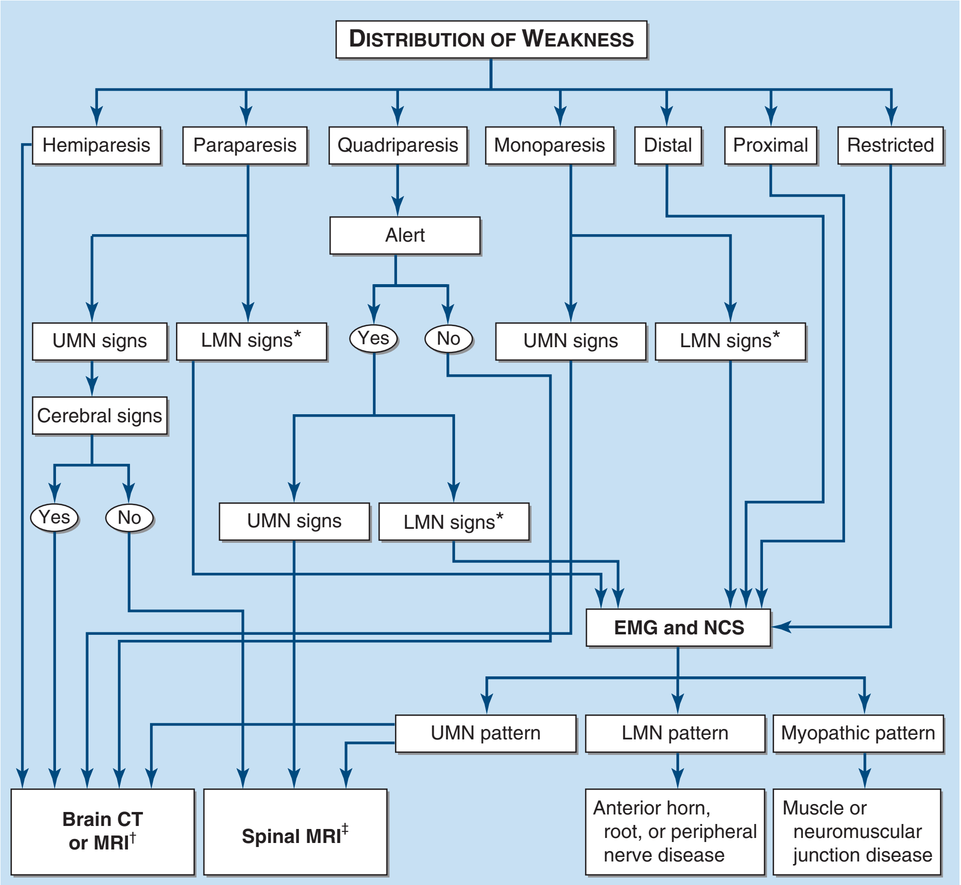

*or signs of myopathy

† If no abnormality detected, consider spinal MRI.

‡ If no abnormality detected, consider myelogram or brain MRI.

FIGURE 26-3 An algorithm for the initial workup of a patient with weakness. CT, computed tomography; EMG, electromyography; LMN, lower motor neuron; MRI, magnetic resonance imaging; NCS, nerve conduction studies; UMN, upper motor neuron.

neuropathy (in which weakness is often most profound distally), or myopathic weakness (typically proximal).

When quadriparesis develops acutely in obtunded patients, evaluation begins with a CT scan of the brain. If upper motor neuron signs have developed acutely but the patient is alert, the initial test is usually an MRI of the cervical cord. When onset has been gradual, disorders of the cerebral hemispheres, brainstem, and cervical spinal cord can usually be distinguished clinically, and imaging is directed first at the clinically suspected site of pathology. If weakness is lower motor neuron, myopathic, or uncertain in origin, laboratory studies can determine the levels of muscle enzymes and electrolytes, and EMG and nerve conduction studies help to localize the pathologic process **(Chap. 460)**.

TABLE 26-2 Causes of Episodic Generalized Weakness

1. Electrolyte disturbances, e.g., hypokalemia, hyperkalemia, hypercalcemia, hypernatremia, hyponatremia, hypophosphatemia, hypermagnesemia
2. Muscle disorders
    a. Channelopathies (periodic paralyses)
    b. Metabolic defects of muscle (impaired carbohydrate or fatty acid utilization; abnormal mitochondrial function)
3. Neuromuscular junction disorders
    a. Myasthenia gravis
    b. Lambert-Eaton myasthenic syndrome
4. Central nervous system disorders
    a. Transient ischemic attacks of the brainstem
    b. Transient global cerebral ischemia
    c. Multiple sclerosis
5. Lack of voluntary effort
    a. Anxiety
    b. Pain or discomfort
    c. Somatization disorder

Monoparesis Monoparesis usually is due to lower motor neuron disease, with or without associated sensory involvement. Upper motor neuron weakness occasionally presents as a monoparesis of distal and nonantigravity muscles. Myopathic weakness rarely is limited to one limb.

ACUTE MONOPARESIS If weakness is predominantly distal and of upper motor neuron type and is not associated with sensory impairment or pain, focal cortical ischemia is likely **(Chap. 438)**; diagnostic possibilities are similar to those for acute hemiparesis. Sensory loss and pain usually accompany acute lower motor neuron weakness; the weakness commonly localizes to a single nerve root or peripheral nerve, but occasionally reflects plexus involvement. If lower motor neuron weakness is likely, evaluation begins with EMG and nerve conduction studies.

SUBACUTE OR CHRONIC MONOPARESIS Weakness and atrophy that develop over weeks or months are usually of lower motor neuron origin. When associated with sensory symptoms, a peripheral cause (nerve, root, or plexus) is likely; otherwise, anterior horn cell disease should be considered. In either case, an electrodiagnostic study is indicated. If weakness is of the upper motor neuron type, a discrete cortical (precentral gyrus) or cord lesion may be responsible, and appropriate imaging is performed.

Distal Weakness Involvement of two or more limbs distally suggests lower motor neuron or peripheral nerve disease. Acute distal lower-limb weakness results occasionally from an acute toxic polyneuropathy or cauda equina syndrome. Distal symmetric weakness usually develops over weeks, months, or years and, when associated with numbness, is due to peripheral neuropathy **(Chap. 457)**. Anterior horn cell disease may begin distally but is typically asymmetric and without accompanying numbness **(Chap. 448)**. Rarely, myopathies present with distal weakness **(Chap. 460)**. Electrodiagnostic studies help localize the disorder (Fig. 26-3).

Proximal Weakness Myopathy often produces symmetric weakness of the pelvic or shoulder girdle muscles **(Chap. 460)**. Diseases of the neuromuscular junction, such as myasthenia gravis **(Chap. 459)**, may present with symmetric proximal weakness often associated with ptosis, diplopia, or bulbar weakness and fluctuate in severity during the day. In anterior horn cell disease, proximal weakness is usually asymmetric, but it may be symmetric, especially in genetic forms. Numbness does not occur with any of these diseases. The evaluation usually begins with determination of the serum creatine kinase level and electrophysiologic studies.

Weakness in a Restricted Distribution Weakness may not fit any of these patterns, being limited, for example, to the extraocular, hemifacial, bulbar, or respiratory muscles. If it is unilateral, restricted weakness usually is due to lower motor neuron or peripheral nerve disease, such as in a facial palsy. Weakness of part of a limb is commonly due to a peripheral nerve lesion such as an entrapment neuropathy. Relatively symmetric weakness of extraocular or bulbar muscles frequently is due to a myopathy **(Chap. 460)** or neuromuscular junction disorder **(Chap. 459)**. Bilateral facial palsy with areflexia suggests Guillain-Barré syndrome **(Chap. 458)**. Worsening of relatively symmetric weakness with fatigue is characteristic of neuromuscular junction disorders. Asymmetric bulbar weakness usually is due to motor neuron disease. Weakness limited to respiratory muscles is uncommon and usually is due to motor neuron disease, myasthenia gravis, or polymyositis/dermatomyositis **(Chap. 377)**.

FURTHER READING

Brazis P et al: *Localization in Clinical Neurology*, 8th ed. Philadelphia, Lippincott William & Wilkins, 2021.

Campbell WW, Barohn RJ: *DeJong's The Neurological Examination*, 8th ed. Philadelphia, Lippincott William & Wilkins, 2019.

O'Brien M on behalf of the Guarantors of Brain: *Aids to the Examination of the Peripheral Nervous System*, 6th ed. Amsterdam, Elsevier, 2023.

27

Numbness, Tingling, and Sensory Loss

Stephen L. Hauser

Normal somatic sensation reflects a continuous monitoring process, little of which reaches consciousness under ordinary conditions. By contrast, disordered sensation, particularly when experienced as painful, is alarming and dominates the patient's attention. Physicians should be able to recognize abnormal sensations by how they are described, know their type and likely site of origin, and understand their implications. **Pain is considered separately in Chap. 14.**

POSITIVE AND NEGATIVE SYMPTOMS

Abnormal sensory symptoms can be divided into two categories: positive and negative. The prototypical positive symptom is tingling (pins and needles); other positive sensory phenomena include itch and altered sensations that are described as pricking, bandlike, lightning-like shooting feelings (lancinations), aching, knifelike, twisting, drawing, pulling, tightening, burning, searing, electrical, or raw feelings. Such symptoms are often painful.

Positive phenomena usually result from trains of impulses generated at sites of lowered threshold or heightened excitability along a peripheral or central sensory pathway. The nature and severity of the abnormal sensation depend on the number, rate, timing, and distribution of ectopic impulses and the type and function of nervous tissue in which they arise. Because positive phenomena represent excessive activity in sensory pathways, they are not necessarily associated with a sensory deficit (loss) on examination.

Negative phenomena represent loss of sensory function and are characterized by diminished or absent feeling that often is experienced as numbness and by abnormal findings on sensory examination. In disorders affecting peripheral sensation, at least one-half of the afferent axons innervating a particular site are probably lost or functionless before a sensory deficit can be demonstrated by clinical examination. If the rate of loss is slow, however, lack of cutaneous feeling may be unnoticed by the patient and difficult to demonstrate on examination, even though few sensory fibers are functioning; if it is rapid, both positive and negative phenomena are usually conspicuous. Subclinical degrees of sensory dysfunction may be revealed by sensory nerve conduction studies or somatosensory-evoked potentials.

Whereas sensory symptoms may be either positive or negative, sensory signs on examination are always a measure of negative phenomena.

TERMINOLOGY

Paresthesias and dysesthesias are general terms used to denote positive sensory symptoms. The term *paresthesias* typically refers to tingling or pins-and-needles sensations but may include a wide variety of other abnormal sensations, except pain; it sometimes implies that the abnormal sensations are perceived spontaneously. The more general term *dysesthesias* denotes all types of abnormal sensations, including painful ones, regardless of whether a stimulus is evident.

Another set of terms refers to sensory abnormalities found on examination. *Hypesthesia* or *hypoesthesia* refers to a reduction of cutaneous sensation to a specific type of testing such as pressure, light touch, and warm or cold stimuli; *anesthesia,* to a complete absence of skin sensation to the same stimuli plus pinprick; and *hypalgesia* or *analgesia*, to reduced or absent pain perception (nociception). *Hyperesthesia* means pain or increased sensitivity in response to touch. Similarly, *allodynia* describes the situation in which a nonpainful stimulus, once perceived, is experienced as painful, even excruciating. An example is elicitation of a painful sensation by application of a vibrating tuning fork. *Hyperalgesia* denotes severe pain in response to a mildly noxious stimulus, and *hyperpathia*, a broad term, encompasses all the phenomena described by hyperesthesia, allodynia, and hyperalgesia. With hyperpathia, the threshold for a sensory stimulus is increased and perception is delayed, but once felt, it is unduly painful.

Disorders of deep sensation arising from muscle spindles, tendons, and joints affect proprioception (position sense). Manifestations include imbalance (particularly with eyes closed or in the dark), clumsiness of precision movements, and unsteadiness of gait, which are referred to collectively as *sensory ataxia*. Other findings on examination usually, but not invariably, include reduced or absent joint position and vibratory sensation and absent deep tendon reflexes in the affected limbs. The Romberg sign is positive, which means that the patient sways markedly or topples when asked to stand with feet close together and eyes closed. In severe states of deafferentation involving deep sensation, the patient cannot walk or stand unaided or even sit unsupported. Continuous involuntary movements (*pseudoathetosis*) of the outstretched hands and fingers occur, particularly with eyes closed.

ANATOMY OF SENSATION

Cutaneous receptors are classified by the type of stimulus that optimally excites them. They consist of naked nerve endings (nociceptors, which respond to tissue-damaging stimuli, and thermoreceptors, which respond to noninjurious thermal stimuli) and encapsulated terminals (several types of mechanoreceptor, activated by physical deformation of the skin or stretch of muscles). Each type of receptor has its own set of sensitivities to specific stimuli, size and distinctness of receptive fields, and adaptational qualities.

Afferent peripheral nerve fibers conveying somatosensory information from the limbs and trunk traverse the dorsal roots and enter the dorsal horn of the spinal cord **(Fig. 27-1)**; the cell bodies of first-order neurons are located in the dorsal root ganglia (DRG). In an analogous fashion, sensations from the face and head are conveyed through the trigeminal system **(Fig. 452-2)**. Once fiber tracts enter the spinal cord, the polysynaptic projections of the smaller fibers (unmyelinated and small myelinated), which subserve mainly nociception, itch, temperature sensibility, and touch, cross and ascend in the opposite anterior and lateral columns of the spinal cord, through the brainstem, to the ventral posterolateral (VPL) nucleus of the thalamus and ultimately project to the postcentral gyrus of the parietal cortex and other cortical areas **(Chap. 14)**. This is the *spinothalamic pathway* or *anterolateral system*. The larger fibers, which subserve tactile and position sense and kinesthesia, project rostrally in the posterior and posterolateral columns on the same side of the spinal cord and make their first synapse in the gracile or cuneate nucleus of the lower medulla. Axons of

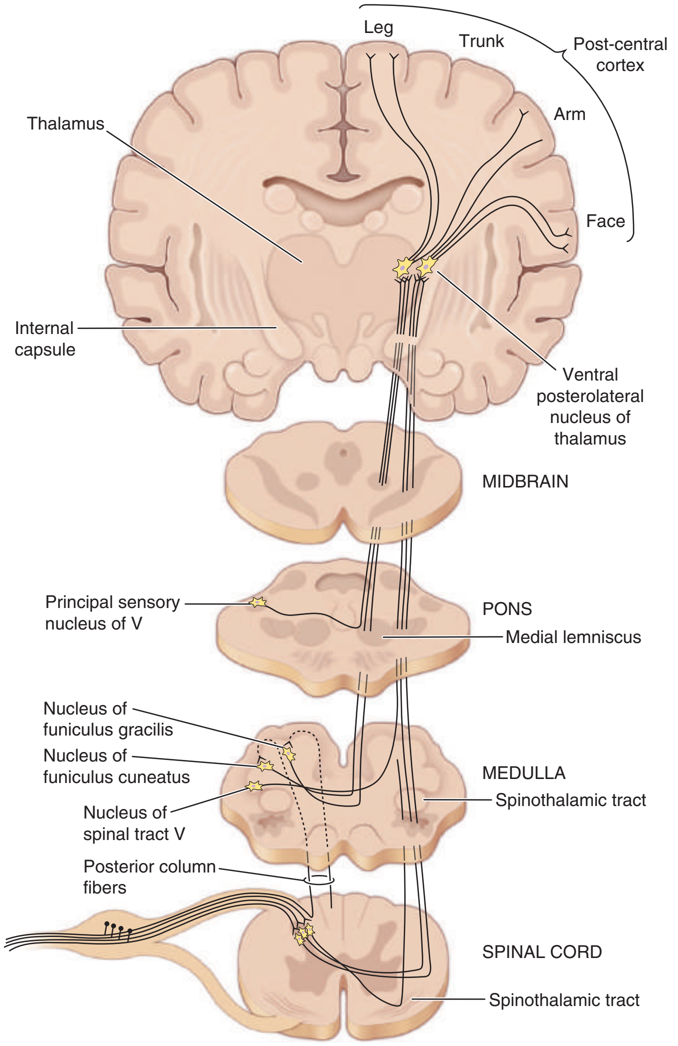

FIGURE 27-1 The main somatosensory pathways. The spinothalamic tract (pain, thermal sense) and the posterior column–lemniscal system (touch, pressure, joint position) are shown. Offshoots from the ascending anterolateral fasciculus (spinothalamic tract) to nuclei in the medulla, pons, and mesencephalon and nuclear terminations of the tract are indicated. *(Reproduced with permission from AH Ropper, MA Samuels: Adams and Victor's Principles of Neurology, 9th ed. New York, McGraw-Hill, 2009.)*

second-order neurons decussate and ascend in the medial lemniscus located medially in the medulla and in the tegmentum of the pons and midbrain and synapse in the VPL nucleus; third-order neurons project to parietal cortex as well as to other cortical areas. This large-fiber system is referred to as the *posterior column–medial lemniscal pathway* (lemniscal, for short). Although the fiber types and functions that make up the spinothalamic and lemniscal systems are relatively well known, many other fibers, particularly those associated with touch, pressure, and position sense, ascend in a diffusely distributed pattern both ipsilaterally and contralaterally in the anterolateral quadrants of the spinal cord. This explains why a complete lesion of the posterior columns of the spinal cord may be associated with little sensory deficit on examination.

APPROACH TO THE PATIENT

Clinical Examination of Sensation

The main components of the sensory examination are tests of primary sensation (pain, touch, vibration, joint position, and thermal sensation) **(Table 27-1)**. The examiner must depend on patient responses, and this complicates interpretation. Further, examination may be limited in some patients. In a stuporous patient, for example, sensory examination is reduced to observing the briskness of withdrawal in response to a pinch or another noxious stimulus. Comparison of responses on the two sides of the body is essential. In an alert but uncooperative patient, it may not be possible to examine cutaneous sensation, but some idea of proprioceptive function may be gained by noting the patient's best performance of movements requiring balance and precision.

In patients with sensory complaints, testing should begin in the center of the affected region and proceed radially until sensation is perceived as normal. The distribution of any abnormality is defined and compared to root and peripheral nerve territories **(Figs. 27-2 and 27-3)**. Some patients present with sensory symptoms that do not fit an anatomic localization and are accompanied by either no abnormalities or gross inconsistencies on examination. The examiner should consider in such cases the possibility of a psychological cause (see "Psychogenic Symptoms," below). Sensory examination of a patient who has no neurologic complaints can be brief and consist of pinprick, touch, and vibration testing in the hands and feet plus evaluation of stance and gait, including the Romberg maneuver **(Chap. V6)**. Evaluation of stance and gait also tests the integrity of motor and cerebellar systems.

PRIMARY SENSATION

The sense of pain usually is tested with a clean pin, which is then discarded. The patient is asked to close the eyes and focus on the pricking or unpleasant quality of the stimulus, not just the pressure or touch sensation elicited. Areas of hypalgesia should be mapped by proceeding radially from the most hypalgesic site. Temperature sensation to both hot and cold is best tested with small containers filled with water of the desired temperature. An alternative way to test cold sensation is to touch a metal object, such as a tuning fork at room temperature, to the skin. For testing warm temperatures, the tuning fork or another metal object may be held under warm water of the desired temperature and then used. The appreciation of both cold and warmth should be tested because different receptors respond to each. Touch usually is tested with a wisp of cotton, minimizing pressure on the skin. In general, it is better to avoid testing touch on hairy skin because of the profusion of the sensory endings that surround each hair follicle. The patient is tested with the eyes closed and should respond as soon as the stimulus is perceived, indicating its location.

Joint position testing is a measure of proprioception. With the patient's eyes closed, joint position is tested in the distal interphalangeal joint of the great toe and fingers. The digit is held by its sides, distal to the joint being tested, and moved passively while more proximal joints are stabilized—the patient indicates the change in position or direction of movement. If errors are made, more proximal joints are tested. A test of proximal joint position sense, primarily at the shoulder, is performed by asking the patient to bring the two index fingers together with arms extended and eyes closed. Normal individuals can do this accurately, with errors of 1 cm or less.

The sense of vibration is tested with an oscillating tuning fork that vibrates at 128 Hz. Vibration is tested over bony points, beginning distally; in the feet, it is tested over the dorsal surface of the distal phalanx of the big toes and at the malleoli of the ankles, and in the hands, it is tested dorsally at the distal phalanx of the fingers. If abnormalities are found, more proximal sites should be examined. Vibratory thresholds at the same site in the patient and the examiner may be compared for control purposes.

CORTICAL SENSATION

The most commonly used tests of cortical function are two-point discrimination, touch localization, and bilateral simultaneous stimulation, and tests for graphesthesia and stereognosis. Abnormalities of these sensory tests, in the presence of normal primary sensation in an alert cooperative patient, signify a lesion of the parietal cortex or thalamocortical projections. If primary sensation is altered, these

TABLE 27-1 Testing Primary Sensation

SENSE	TEST DEVICE	ENDINGS ACTIVATED	FIBER SIZE MEDIATING	CENTRAL PATHWAY
Pain	Pinprick	Cutaneous nociceptors	Small	SpTh, also D
Temperature, heat	Warm metal object	Cutaneous thermoreceptors for hot	Small	SpTh
Temperature, cold	Cold metal object	Cutaneous thermoreceptors for cold	Small	SpTh
Touch	Cotton wisp, fine brush	Cutaneous mechanoreceptors, also naked endings	Large and small	Lem, also D and SpTh
Vibration	Tuning fork, 128 Hz	Mechanoreceptors, especially pacinian corpuscles	Large	Lem, also D
Joint position	Passive movement of specific joints	Joint capsule and tendon endings, muscle spindles	Large	Lem, also D

Abbreviations: D, diffuse ascending projections in ipsilateral and contralateral anterolateral columns; Lem, posterior column and lemniscal projection, ipsilateral; SpTh, spinothalamic projection, contralateral.

cortical discriminative functions usually will be abnormal also. Comparisons should always be made between analogous sites on the two sides of the body because the deficit with a specific parietal lesion is likely to be unilateral.

Two-point discrimination can be tested with calipers, the points of which may be set from 2 mm to several centimeters apart and then applied simultaneously to the test site. On the fingertips, a normal individual can distinguish about a 3-mm separation of points.

Touch localization is performed by light pressure for an instant with the examiner's fingertip or a wisp of cotton wool; the patient, whose eyes are closed, is asked to identify the site of touch. *Bilateral simultaneous stimulation* at analogous sites (e.g., the dorsum of both hands) can be carried out to determine whether the perception of touch is extinguished consistently on one side (*extinction* or *neglect*). *Graphesthesia* refers to the capacity to recognize, with eyes closed, letters or numbers drawn by the examiner's fingertip on the palm of the hand. Once again, side-to-side comparison is of prime importance. Inability to recognize numbers or letters is termed *agraphesthesia*.

Stereognosis refers to the ability to identify common objects by palpation, recognizing their shape, texture, and size. Common standard objects such as keys, paper clips, and coins are best used. Patients with normal stereognosis should be able to distinguish a dime from a penny and a nickel from a quarter without looking.

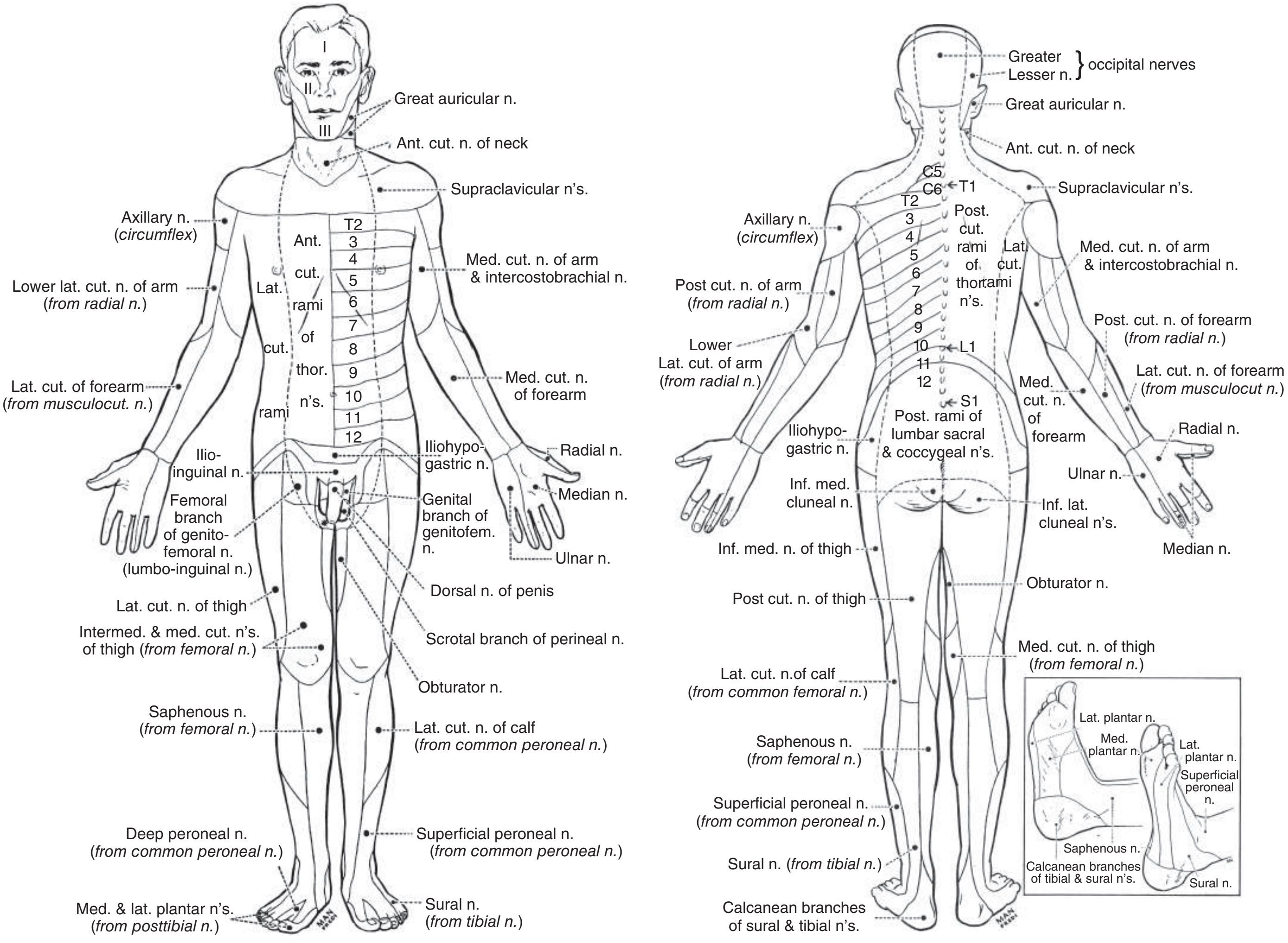
I
II
III
Great auricular n.
Ant. cut. n. of neck
Supraclavicular n's.
Axillary n. (circumflex)
Ant. cut. rami of thor. n's.
Lat. cut. rami
T2
3
4
5
6
7
8
9
10
11
12
Med. cut. n. of arm & intercostobrachial n.
Lower lat. cut. n. of arm (from radial n.)
Lat. cut. of forearm (from musculocut. n.)
Med. cut. n. of forearm
Ilio-inguinal n.
Iliohypogastric n.
Radial n.
Median n.
Femoral branch of genito-femoral n. (lumbo-inguinal n.)
Genital branch of genitofem. n.
Ulnar n.
Dorsal n. of penis
Lat. cut. n. of thigh
Intermed. & med. cut. n's. of thigh (from femoral n.)
Scrotal branch of perineal n.
Obturator n.
Saphenous n. (from femoral n.)
Lat. cut. n. of calf (from common peroneal n.)
Deep peroneal n. (from common peroneal n.)
Superficial peroneal n. (from common peroneal n.)
Med. & lat. plantar n's. (from posttibial n.)
Sural n. (from tibial n.)
MAN
Greater
Lesser n.
occipital nerves
Great auricular n.
Ant. cut. n. of neck
C5
C6
T1
T2
3
4
5
6
7
8
9
10
11
12
L1
S1
Supraclavicular n's.
Axillary n. (circumflex)
Post. cut. rami of thor. n's.
Lat. cut. rami
Med. cut. n. of arm & intercostobrachial n.
Post cut. n. of arm (from radial n.)
Post. cut. n. of forearm (from radial n.)
Lower Lat. cut. of arm (from radial n.)
Lat. cut. n. of forearm (from musculocut n.)
Med. cut. n. of forearm
Radial n.
Iliohypogastric n.
Post. rami of lumbar sacral & coccygeal n's.
Ulnar n.
Inf. med. cluneal n.
Inf. lat. cluneal n's.
Median n.
Inf. med. n. of thigh
Obturator n.
Post cut. n. of thigh
Med. cut. n. of thigh (from femoral n.)
Lat. cut. n.of calf (from common femoral n.)
Lat. plantar n.
Med. plantar n.
Lat. plantar n.
Superficial peroneal n.
Saphenous n. (from femoral n.)
Saphenous n.
Superficial peroneal n. (from common peroneal n.)
Calcanean branches of tibial & sural n's.
Sural n.
Sural n. (from tibial n.)
Calcanean branches of sural & tibial n's.
MAN

FIGURE 27-2 The cutaneous fields of peripheral nerves. *(Reproduced with permission from W Haymaker, B Woodhall: Peripheral Nerve Injuries, 2nd ed. Philadelphia, Saunders, 1953.)*

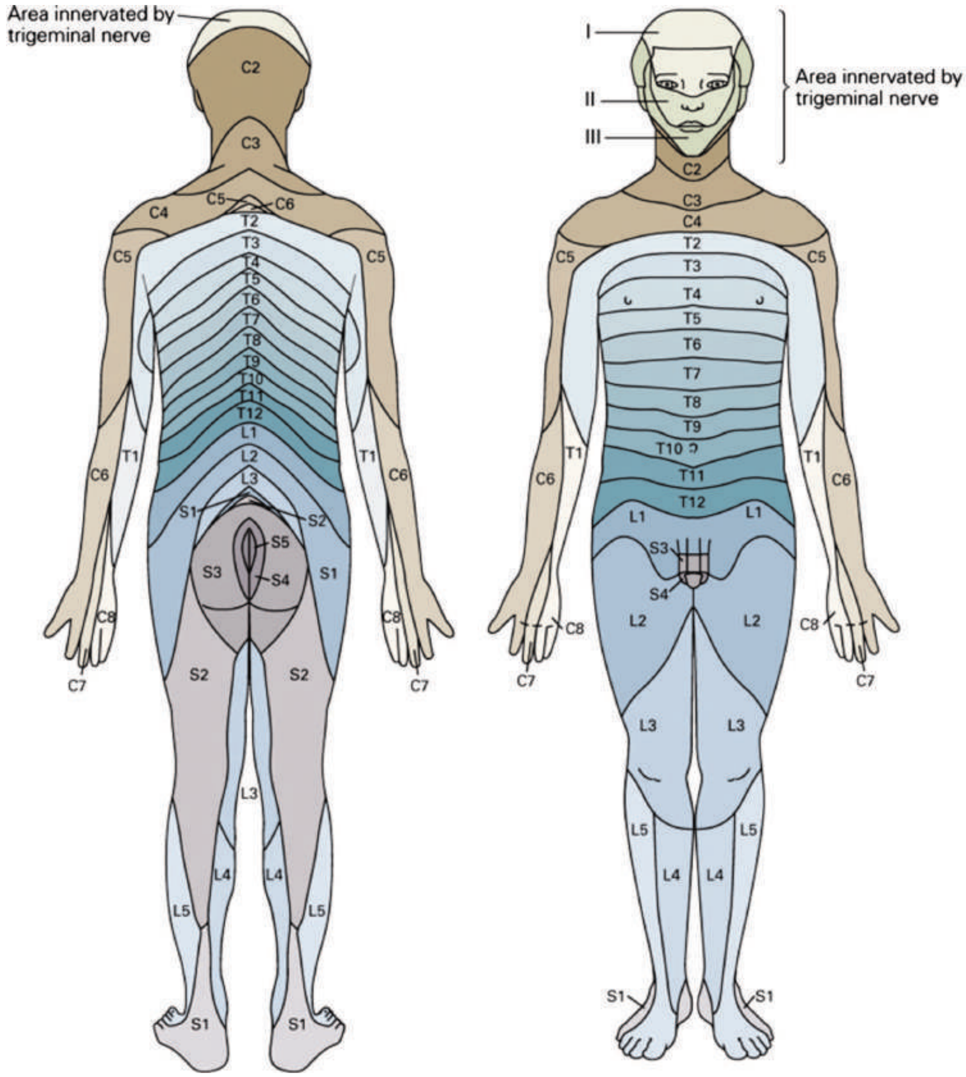
Area innervated by trigeminal nerve
C2
C3
C4
C5
C6
T2
T3
T4
T5
T6
T7
T8
T9
T10
T11
T12
L1
L2
L3
S1
S2
S5
S3
S4
T1
C8
C7
L4
L5
I
II
III
Area innervated by trigeminal nerve

FIGURE 27-3 Distribution of the sensory spinal roots on the surface of the body (dermatomes). *(Reproduced with permission from Receptors of the somatosensory system. In ER Kandel, JD Koester, SH Mack, SA Siegelbaum. Principles of Neural science, 6th ed. New York: McGraw Hill; 2021.)*

Patients should feel the object with only one hand at a time. If they are unable to identify it in one hand, it should be placed in the other for comparison. Individuals who are unable to identify common objects and coins in one hand but can do so in the other are said to have *astereognosis* of the abnormal hand.

QUANTITATIVE SENSORY TESTING

Effective sensory testing devices are commercially available. Quantitative sensory testing is particularly useful for serial evaluation of cutaneous sensation in clinical trials. Threshold testing for touch and vibratory and thermal sensation is the most widely used application.

ELECTRODIAGNOSTIC STUDIES AND NERVE BIOPSY

Nerve conduction studies and nerve biopsy are important means of investigating the peripheral nervous system, but they do not evaluate the function or structure of cutaneous receptors and free nerve endings or of unmyelinated or thinly myelinated nerve fibers in the nerve trunks. Skin biopsy can be used to evaluate these structures in the dermis and epidermis.

LOCALIZATION OF SENSORY ABNORMALITIES

Sensory symptoms and signs can result from lesions at many different levels of the nervous system from the parietal cortex to the peripheral sensory receptor. Noting their distribution and nature is the most important way to localize their source. Their extent, configuration, symmetry, quality, and severity are the key observations.

Dysesthesias without sensory findings by examination may be difficult to interpret. To illustrate, tingling dysesthesias in an acral distribution (hands and feet) can be systemic in origin, for example, secondary to hyperventilation, or induced by a medication such as acetazolamide. Distal dysesthesias can also be an early event in an evolving polyneuropathy or may herald a myelopathy, such as from vitamin B_{12} deficiency. Sometimes, distal dysesthesias have no definable basis. In contrast, dysesthesias that correspond in distribution to that of a particular peripheral nerve structure denote a lesion at that site. For instance, dysesthesias restricted to the fifth digit and the adjacent one-half of the fourth finger on one hand reliably point to disorder of the ulnar nerve, most commonly at the elbow.

Nerve and Root In focal nerve trunk lesions, sensory abnormalities are readily mapped and generally have discrete boundaries (Figs. 27-2 and 27-3). Root ("radicular") lesions frequently are accompanied by deep, aching pain along the course of the related nerve trunk. With compression of a fifth lumbar (L5) or first sacral (S1) root, as from a ruptured intervertebral disk, sciatica (radicular pain relating to the sciatic nerve trunk) is a common manifestation **(Chap. 18)**. With a lesion affecting a single root, sensory deficits may be minimal or absent because adjacent root territories overlap extensively.

Isolated mononeuropathies may cause symptoms beyond the territory supplied by the affected nerve, but abnormalities on examination typically are confined to expected anatomic boundaries. In multiple mononeuropathies, symptoms and signs occur in discrete territories supplied by different individual nerves and—as more nerves are affected—may simulate a polyneuropathy if deficits become confluent. With polyneuropathies, sensory deficits are generally graded, distal, and symmetric in distribution (Chap. 457). Dysesthesias, followed by numbness, begin in the toes and ascend symmetrically. When dysesthesias reach the knees, they usually also have appeared in the fingertips. The process is nerve length–dependent, and the deficit is often described as "stocking glove" in type. Involvement of both hands and feet also occurs with lesions of the upper cervical cord or the brainstem, but an upper level of the sensory disturbance may then be found on the trunk and other evidence of a central lesion may be present, such as sphincter involvement or signs of an upper motor neuron lesion (Chap. 26). Although most polyneuropathies are pansensory and affect all modalities of sensation, selective sensory dysfunction according to nerve fiber size may occur. Small-fiber polyneuropathies are characterized by burning, painful dysesthesias with reduced pinprick and thermal sensation but with sparing of proprioception, motor function, and deep tendon reflexes. Touch is involved variably; when it is spared, the pattern is referred to as a *dissociated sensory loss*. Importantly, sensory dissociation may occur also with spinal cord lesions (Chap. 453, and below). Large-fiber polyneuropathies are characterized by vibration and position sense deficits, imbalance, absent tendon reflexes, and variable motor dysfunction but preservation of most cutaneous sensation. Dysesthesias, if present at all, tend to be tingling or bandlike in quality.

Sensory neuronopathy (or ganglionopathy) is characterized by widespread but asymmetric sensory loss occurring in a non-length-dependent manner so that it may occur proximally or distally, and in the arms, legs, or both. Pain and numbness progress to sensory ataxia and impairment of all sensory modalities over time. This condition is usually paraneoplastic or idiopathic in origin (Chaps. 99 and 457) or related to an autoimmune disease, particularly Sjögren's syndrome (Chap. 373).

Spinal Cord (See also Chap. 453) If the spinal cord is transected, all sensation is lost below the level of transection. Bladder and bowel function also are lost, as is motor function. Lateral hemisection of the spinal cord produces the Brown-Séquard syndrome, with absent pain and temperature sensation contralaterally and loss of proprioceptive sensation and power ipsilaterally below the lesion (see Figs. 27-1 and 453-1); ipsilateral pain or hyperesthesia may also occur.

Numbness or paresthesias in both feet may arise from a spinal cord lesion; this is especially likely when the upper level of the sensory loss extends to the trunk. When all extremities are affected, the lesion is probably in the cervical region or brainstem unless a peripheral neuropathy is responsible. The presence of upper motor neuron signs (Chap. 26) supports a central lesion; a hyperesthetic band on the trunk may suggest the level of involvement.

A dissociated sensory loss can reflect spinothalamic tract involvement in the spinal cord, especially if the deficit is unilateral and has an upper level on the torso. Bilateral spinothalamic tract involvement occurs with lesions affecting the center of the spinal cord, such as in syringomyelia. There is a dissociated sensory loss with impairment of pinprick and temperature appreciation but relative preservation of light touch, position sense, and vibration appreciation.

Dysfunction of the posterior columns in the spinal cord or of the posterior root entry zone may lead to a bandlike sensation around the trunk or a feeling of tight pressure in one or more limbs. Flexion of the neck sometimes leads to an electric shock–like sensation that radiates down the back and into the legs (Lhermitte's sign) in patients with a cervical lesion affecting the posterior columns, such as from multiple sclerosis, cervical spondylosis, or following irradiation to the cervical region.

Brainstem Crossed patterns of sensory disturbance, in which one side of the face and the opposite side of the body are affected, localize to the lateral medulla. Here a small lesion may damage both the ipsilateral descending trigeminal tract and trigeminal nucleus, plus the ascending spinothalamic fibers subserving the opposite arm, leg, and hemitorso (see "Lateral medullary syndrome" in Fig. 437-7). A lesion in the tegmentum of the pons and midbrain, where the lemniscal and spinothalamic tracts merge, causes pansensory loss contralaterally.

Thalamus Hemisensory disturbance with tingling numbness from head to foot is often thalamic in origin but also can arise from the anterior parietal region. If abrupt in onset, the lesion is likely to be due to a small stroke (lacunar infarction), particularly if localized to the thalamus. Occasionally, with lesions affecting the VPL nucleus or adjacent white matter, a syndrome of thalamic pain, also called *Déjerine-Roussy syndrome*, may ensue. The severe, persistent, unrelenting unilateral pain often is described in dramatic terms.

Cortex With lesions of the parietal lobe involving either the cortex or subjacent white matter, the most prominent symptoms are contralateral hemineglect, hemi-inattention, and a tendency not to use the affected hand and arm. On cortical sensory testing (e.g., two-point discrimination, graphesthesia), abnormalities are often found, but primary sensation is usually intact. Anterior parietal infarction may present as a pseudothalamic syndrome with contralateral loss of primary sensation from head to toe. Dysesthesias or a sense of numbness and, rarely, a painful state may also occur.

Focal Sensory Seizures These seizures generally are due to lesions in the area of the postcentral or precentral gyrus. The principal symptom of focal sensory seizures is tingling, but additional, more complex sensations may occur, such as a rushing feeling, a sense of warmth, or a sense of movement without detectable motion. Symptoms typically are unilateral; commonly begin in the arm or hand, face, or foot; and often spread in a manner that reflects the cortical representation of different bodily parts, a pattern referred to as a *Jacksonian march*. Their duration is variable; seizures may be transient, lasting only for seconds, or persist for an hour or more. Focal motor features may supervene, often becoming generalized with loss of consciousness and tonic-clonic jerking.

Psychogenic Symptoms Sensory symptoms can have a psychogenic basis. Such symptoms may be generalized or have an anatomic boundary that is difficult to explain neurologically, for example, circumferentially at the groin or shoulder or around a specific joint. Pain is common, but the nature and intensity of any sensory disturbances are variable. The diagnosis should not be one of exclusion but based on suggestive findings that are otherwise difficult to explain, such as midline splitting of impaired vibration, pinprick, or light touch appreciation; variability or poor reproducibility of sensory deficits; or normal performance of tasks requiring sensory input that is seemingly abnormal on formal testing, such as good performance with eyes closed of the finger-to-nose test despite an apparent loss of position sense in the upper limb. The side with abnormal sensation may be confused when the limbs are placed in an unusual position, such as crossed behind the back. Sensory complaints should not be regarded as psychogenic simply because they are unusual.

■ TREATMENT

Management is based on treatment of the underlying condition. Symptomatic treatment of acute and chronic pain is discussed in Chap. 14. Dysesthesias, when severe and persistent, may respond to anticonvulsants (carbamazepine, 100–1000 mg/d; gabapentin, 300–3600 mg/d; or pregabalin, 50–300 mg/d), antidepressants (amitriptyline, 25–150 mg/d; nortriptyline, 25–150 mg/d; desipramine, 100–300 mg/d; or venlafaxine, 75–225 mg/d).

■ FURTHER READING

Brazis P et al: *Localization in Clinical Neurology*, 8th ed. Philadelphia, Lippincott William & Wilkins, 2021.

Campbell WW, Barohn RJ: *DeJong's The Neurological Examination*, 8th ed. Philadelphia, Lippincott William & Wilkins, 2019.

Waxman S: *Clinical Neuroanatomy*, 30th ed. New York, McGraw Hill Education, 2024.

28

Gait Disorders, Imbalance, and Falls

Jessica M. Baker, Gerald Pankratz

PREVALENCE, MORBIDITY, AND MORTALITY

Gait and balance problems are common in the elderly and contribute to the risk of falls and injury. Gait disorders have been described in 15% of individuals aged >65. By age 80, one person in four will use a mechanical aid to assist with ambulation. Among those aged ≥85, the prevalence of gait abnormality approaches 40%. In epidemiologic studies, gait disorders are consistently identified as a major risk factor for falls and injury.

ANATOMY AND PHYSIOLOGY

An upright bipedal gait depends on the successful integration of postural control and locomotion. These functions are widely distributed in the central nervous system. The biomechanics of bipedal walking are complex, and the performance is easily compromised by a neurologic deficit at any level. Command and control centers in the brainstem, cerebellum, and forebrain modify the action of spinal pattern generators to promote stepping. While a form of "fictive locomotion" can be elicited from quadrupedal animals after spinal transection, this capacity is limited in primates. Step generation in primates is dependent on locomotor centers in the pontine tegmentum, midbrain, and subthalamic region. Locomotor synergies are executed through the reticular formation and descending pathways in the ventromedial spinal cord. Cerebral control provides a goal and purpose for walking and is involved in avoidance of obstacles and adaptation of locomotor programs to context and terrain.

Postural control requires the maintenance of the center of mass over the base of support through the gait cycle. Unconscious postural adjustments maintain standing balance: long latency responses are measurable in the leg muscles, beginning 110 milliseconds after a perturbation. Forward motion of the center of mass provides propulsive force for stepping, but failure to maintain the center of mass within stability limits results in falls. The anatomic substrate for dynamic balance has not been well defined, but the vestibular nucleus and midline cerebellum contribute to balance control in animals. Patients with damage to these structures have impaired balance while standing and walking.

Standing balance depends on good-quality sensory information about the position of the body center with respect to the environment, support surface, and gravitational forces. Sensory information for postural control is primarily generated by the visual system, the vestibular system, and proprioceptive receptors in the muscle spindles and joints. A healthy redundancy of sensory afferent information is generally available, but loss of two of the three pathways is sufficient to compromise standing balance. Balance disorders in older individuals sometimes result from multiple insults in the peripheral sensory systems (e.g., visual loss, vestibular deficit, peripheral neuropathy) that critically degrade the quality of afferent information needed for balance stability.

Older patients with cognitive impairment appear to be particularly prone to falls and injury. There is a growing body of literature on the use of attentional resources to manage gait and balance. Walking is generally considered to be unconscious and automatic, but the ability to walk while attending to a cognitive task (*dual-task walking*) may be compromised in the elderly. Older patients with deficits in executive function may have particular difficulty in managing the attentional resources needed for dynamic balance when distracted. Decline in gait speed and memory are increasingly associated with risk for dementia in older adults.

DISORDERS OF GAIT

Disorders of gait may be attributed to neurologic and nonneurologic causes, although significant overlap often exists. The *antalgic gait* results from avoidance of pain associated with weight bearing and is commonly seen in osteoarthritis. Asymmetry is a common feature of gait disorders due to contractures and other orthopedic deformities. Impaired vision rounds out the list of common nonneurologic causes of gait disorders.

Neurologic gait disorders are disabling and equally important to address. The heterogeneity of gait disorders observed in clinical practice reflects the large network of neural systems involved in the task. Walking is vulnerable to neurologic disease at every level. Gait disorders have been classified descriptively on the basis of abnormal physiology and biomechanics. One problem with this approach is that many failing gaits look fundamentally similar. This overlap reflects common patterns of adaptation to threatened balance stability and declining performance. *The gait disorder observed clinically must be viewed as the product of a neurologic deficit and a functional adaptation*. Unique features of the failing gait are often overwhelmed by the adaptive response. Some common patterns of abnormal gait are summarized next. Gait disorders can also be classified by etiology **(Table 28-1)**. Analysis of gait remains primarily observational in clinical settings, though

TABLE 28-1 Prevalence of Neurologic Gait Disorders

NEUROLOGIC GAIT DISORDER	NO. (%)[a]	TOTAL NUMBER[b]	CAUSES (NO.)
Single neurologic gait disorder	81 (69%)		
Sensory ataxic	22 (18%)	46	Peripheral sensory neuropathy (46)
Parkinsonian	19 (16%)	34	Parkinson's disease (18), drug-induced parkinsonism (8), dementia with parkinsonism (4), parkinsonism (4)
Higher level	9 (8%)	31	Vascular encephalopathy (20), normal pressure hydrocephalus (1), severe dementia (7), hypoxic ischemic encephalopathy (1), unknown (1)
Cerebellar ataxic	7 (6%)	10	Cerebellar stroke (3), cerebellar lesion due to multiple sclerosis (1), severe essential tremor (3), postvaccinal cerebellitis (1), chronic alcohol abuse (1), multiple system atrophy (1)
Cautious	7 (6%)	7	Idiopathic, associated fear of falling (7)
Paretic/hypotonic	6 (5%)	14	Neurogenic claudication (7), diabetic neuropathy (1), nerve lesion due to trauma or surgery (4), distal paraparesis after Guillain-Barré syndrome (1), unknown (2)
Spastic	6 (5%)	7	Ischemic stroke (3), intracerebral hemorrhage (3), congenital (1)
Vestibular ataxic	4 (3%)	6	Bilateral vestibulopathy (3), recent vestibular neuronitis (1), recent Ménière's attack (1), acoustic neuroma with surgery (1)
Dyskinetic	1 (1%)	4	Levodopa-induced dyskinesia (3), chorea (1)
Multiple neurologic gait disorders	36 (30%)		
Total	117		

[a] Percentage of individuals with a single gait disorder. [b] Includes individuals with multiple gait disorders.

Note: Of 117 patients with a neurologic gait disorder, 81 had a single neurologic gait disorder; the remainder (36) had multiple neurologic gait disorders.

Source: Reproduced from P Mahlknecht et al: PLoS One 8:e69627, 2013.

quantitative gait analysis, with devices such as instrumented walkways, motion capture systems and wearables, is standard in research settings and has the potential to inform diagnosis and treatment of gait disorders when more widely adopted in clinical settings.

■ CAUTIOUS GAIT

The term *cautious gait* is used to describe the patient who walks with an abbreviated stride, widened base, and lowered center of mass, as if walking on a slippery surface. Arms are often held abducted. This disorder is both common and nonspecific. It is, in essence, an adaptation to a perceived postural threat. There may be an associated fear of falling. This disorder can be observed in more than one-third of older patients with gait impairment. Physical therapy often improves walking to the degree that follow-up observation may reveal a more specific underlying disorder.

■ STIFF-LEGGED GAIT

Spastic gait is characterized by stiffness in the legs, an imbalance of muscle tone, and a tendency to circumduct and scuff the feet. The disorder reflects compromise of corticospinal command and overactivity of spinal reflexes. The patient may walk on the toes. In extreme instances, the legs cross due to increased tone in the adductors ("scissoring" gait). Upper motor neuron signs are present on physical examination. The disorder may be cerebral or spinal in origin.

Myelopathy from cervical spondylosis is a common cause of spastic or spastic-ataxic gait in the elderly. Demyelinating disease and trauma are the leading causes of myelopathy in younger patients. In chronic progressive myelopathy of unknown cause, a workup with laboratory and imaging tests may establish a diagnosis. A structural lesion, such as a tumor or a spinal vascular malformation, should be excluded with appropriate testing. **Spinal cord disorders are discussed in detail in Chap. 453.**

With cerebral spasticity, asymmetry is common, the upper extremities are usually involved, and dysarthria is often an associated feature. Common causes include vascular disease (stroke), multiple sclerosis, motor neuron disease, and perinatal nervous system injury (cerebral palsy).

Other stiff-legged gaits include dystonia **(Chap. 447)** and stiff-person syndrome **(Chap. 99)**. Dystonia is a disorder characterized by sustained muscle contractions resulting in repetitive twisting movements and abnormal posture. It often has a genetic basis, though asymmetric dystonia, particularly involving gait, may be a sign of Parkinson's disease in younger adults. Dystonic spasms can produce plantar flexion and inversion of the feet, sometimes with torsion of the trunk. In autoimmune stiff-person syndrome, exaggerated lordosis of the lumbar spine and overactivation of antagonist muscles restrict trunk and lower-limb movement and result in a wooden or fixed posture.

■ PARKINSONISM, FREEZING GAIT, AND OTHER MOVEMENT DISORDERS

Parkinson's disease **(Chap. 446)** is common, affecting 1% of the population >65 years of age. The stooped posture, shuffling gait, and decreased arm swing are characteristic and distinctive features. Patients sometimes accelerate (festinate) with walking, display retropulsion, or exhibit a tendency to turn en bloc. Falls are a major source of morbidity, particularly later in the disease course, as balance becomes progressively impaired. Dopamine replacement improves step length, arm swing, turning speed, and gait initiation. There is increasing evidence that deficits in cholinergic circuits in the pedunculopontine nucleus and cortex contribute to the gait disorder of Parkinson's disease. Cholinesterase inhibitors such as donepezil and rivastigmine may decrease fall frequency, even in the absence of cognitive impairment, perhaps through improvement in attention.

Freezing is defined as a brief, episodic absence of forward progression of the feet, despite the intention to walk. Freezing may be triggered by approaching a narrow doorway or crowd and contributes to fall risk. Gait freezing is present in approximately one-quarter of Parkinson's patients within 5 years of onset, and its frequency increases further over time. In patients with motor fluctuations, freezing of gait occurring in the "on" state often fails to respond to changes in levodopa; auditory and visual cueing strategies may be used to overcome freezing. End-of-dose gait freezing may improve with optimization of dopaminergic drugs or with use of monoamine oxidase type B inhibitors such as rasagiline or selegiline **(Chap. 446)**.

Freezing of gait is also common in other neurodegenerative disorders associated with parkinsonism, including progressive supranuclear palsy (PSP) **(Chap. 443)**, multiple-system atrophy **(Chap. 451)**, and corticobasal degeneration **(Chap. 443)**. Patients with these disorders frequently present with axial stiffness, postural instability, and a shuffling, freezing gait while lacking the characteristic pill-rolling tremor of Parkinson's disease. The gait of PSP is typically more erect compared with the stooped posture of typical Parkinson's disease, and falls within the first year also suggest the possibility of PSP. The gait of vascular parkinsonism tends to be broad-based and shuffling with reduced arm swing bilaterally; disproportionate involvement of gait early in the disease course differentiates this entity from Parkinson's disease.

Hyperkinetic movement disorders also produce characteristic and recognizable disturbances in gait. In Huntington's disease **(Chap. 447)**, the unpredictable occurrence of choreic movements gives the gait a dancing quality. Tardive dyskinesia is the cause of many stereotypic gait disorders seen in patients chronically exposed to antipsychotics and other drugs that block the D_2 dopamine receptor. *Orthostatic tremor* is a high-frequency, low-amplitude tremor predominantly involving the lower extremities. Patients often report shakiness or unsteadiness on standing and improvement with sitting or walking. Falls are common. The tremor is often only appreciable by palpating the legs while standing.

■ FRONTAL GAIT DISORDER

Frontal gait disorder, also known as higher-level gait disorder, is common in the elderly and has a variety of causes. The term is used to describe a shuffling, freezing gait with imbalance, and other signs of higher cerebral dysfunction. Typical features include a wide base of support, a short stride, shuffling along the floor, and difficulty with starts and turns. Many patients exhibit a difficulty with gait initiation that is descriptively characterized as the "slipping clutch" syndrome or gait ignition failure. The term *lower-body parkinsonism* is also used to describe such patients. Strength is generally preserved, and patients are able to make stepping movements when not standing and maintaining their balance at the same time. This disorder is best considered a higher-level motor control disorder, as opposed to an apraxia **(Chap. 32)**, though the term *gait apraxia* persists in the literature.

The most common cause of frontal gait disorder is vascular disease, particularly subcortical small-vessel disease in the deep frontal white matter and centrum ovale. Over three-quarters of patients with subcortical vascular dementia demonstrate gait abnormalities; decreased arm swing and a stooped posture are particularly prevalent features. The clinical syndrome also includes dysarthria, pseudobulbar affect (emotional disinhibition), increased tone, and hyperreflexia in the lower limbs **(Chap. 444)**.

Normal pressure (communicating) hydrocephalus (NPH) in adults also presents with a similar gait disorder **(Chap. 442)**. Other features of the diagnostic triad (mental changes, incontinence) may be absent in a substantial number of patients. Magnetic resonance imaging (MRI) demonstrates ventricular enlargement, an enlarged flow void about the aqueduct, periventricular white matter change, and high-convexity tightness (disproportionate widening of the sylvian fissures versus the cortical sulci). A lumbar puncture or dynamic test is necessary to confirm a diagnosis of NPH. Neurodegenerative dementias and mass lesions of the frontal lobes cause a similar clinical picture and can be differentiated from vascular disease and hydrocephalus by neuroimaging.

■ CEREBELLAR GAIT ATAXIA

Disorders of the cerebellum **(Chap. 450)** have a dramatic impact on gait and balance. Cerebellar gait ataxia is characterized by a wide base of support, lateral instability of the trunk, erratic foot placement, and decompensation of balance when attempting to walk on a narrow base. Difficulty maintaining balance when turning is often an early feature. Patients are unable to walk tandem heel to toe and display truncal sway

TABLE 28-2 Features of Cerebellar Ataxia, Sensory Ataxia, and Frontal Gait Disorders

FEATURE	CEREBELLAR ATAXIA	SENSORY ATAXIA	FRONTAL GAIT
Base of support	Wide-based	Wide-based, looks down	Wide-based
Velocity	Variable	Slow	Very slow
Stride	Irregular, lurching	Regular with path deviation	Short, shuffling
Romberg test	+/–	Unsteady, falls	+/–
Heel → shin	Abnormal	+/–	Normal
Initiation	Normal	Normal	Hesitant
Turns	Unsteady	+/–	Hesitant, multistep
Postural instability	+	+++	++++ Poor postural synergies rising from a chair
Falls	Late event	Frequent	Frequent

in narrow-based or tandem stance. They show considerable variation in their tendency to fall in daily life.

Causes of cerebellar ataxia in older patients include stroke, trauma, tumor, and neurodegenerative disease such as multiple-system atrophy **(Chap. 451)** and various forms of hereditary cerebellar degeneration **(Chap. 450)**. A short expansion at the site of the fragile X mutation (*fragile X premutation*) has been associated with gait ataxia in older men. Alcohol causes acute and chronic cerebellar ataxia. In patients with ataxia due to cerebellar degeneration, MRI demonstrates the extent and topography of cerebellar atrophy.

SENSORY ATAXIA

As reviewed earlier in this chapter, balance depends on high-quality afferent information from the visual and the vestibular systems and proprioception. When this information is lost or degraded, balance during locomotion is impaired and instability results. The sensory ataxia of tabetic neurosyphilis is a classic example. The contemporary equivalent is the patient with neuropathy affecting large fibers. Vitamin B_{12} deficiency is a treatable cause of large-fiber sensory loss in the spinal cord and peripheral nervous system. Joint position and vibration sense are diminished in the lower limbs. The stance in such patients is destabilized by eye closure; they often look down at their feet when walking and do poorly in the dark. **Table 28-2** compares sensory ataxia with cerebellar ataxia and frontal gait disorder.

NEUROMUSCULAR DISEASE

Patients with neuromuscular disease often have an abnormal gait, occasionally as a presenting feature. With distal weakness (peripheral neuropathy), the step height is increased to compensate for foot drop, and the sole of the foot may slap on the floor during weight acceptance, termed the *steppage gait*. Patients with myopathy or muscular dystrophy more typically exhibit proximal weakness. Weakness of the hip girdle may result in some degree of excess pelvic sway during locomotion. The stooped posture of lumbar spinal stenosis ameliorates pain from the compression of the cauda equina occurring with a more upright posture while walking and may mimic early parkinsonism.

TOXIC AND METABOLIC DISORDERS

Chronic toxicity from medications and metabolic disturbances can impair motor function and gait. Examination may reveal mental status changes, asterixis, or myoclonus. Static equilibrium is disturbed, and such patients are easily thrown off balance. Disequilibrium is particularly evident in patients with chronic renal disease and those with hepatic failure, in whom asterixis may impair postural support. Sedative drugs, especially neuroleptics and long-acting benzodiazepines, affect postural control and increase the risk for falls. These disorders are especially important to recognize because they are often treatable.

FUNCTIONAL GAIT DISORDER

Functional neurologic disorders (formerly “psychogenic”) are common in practice, and the presentation often involves gait. Sudden onset, inconsistent deficits, waxing and waning course, incongruence of symptoms with an organic lesion, and improvement with distraction are key features. Phenomenology is variable; extreme slow motion, an inappropriately overcautious gait, gyrations of posture with wastage of muscular energy, astasia–abasia (inability to stand and walk), bouncing, and foot stiffness (dystonia) have been described. Falls are rare, and there are often discrepancies between examination findings and the patient’s functional status. Preceding stress or trauma is variably present, and its absence does not preclude the diagnosis of a functional gait disorder. Functional gait disorders may be challenging to diagnose and should be differentiated from the slowness and psychomotor retardation seen in certain patients with major depression.

APPROACH TO THE PATIENT

Slowly Progressive Disorder of Gait

When reviewing the history, it is helpful to inquire about the onset and progression of disability. Initial awareness of an unsteady gait often follows a fall. Stepwise evolution or sudden progression suggests vascular disease. Gait disorder may be associated with urinary urgency and incontinence, particularly in patients with cervical spine disease or hydrocephalus. It is always important to review the use of alcohol and medications that affect gait and balance. Information on localization derived from the neurologic examination can be helpful in narrowing the list of possible diagnoses.

Gait observation provides an immediate sense of the patient’s degree of disability. Arthritic and antalgic gaits are recognized by observation, although neurologic and orthopedic problems may coexist. Characteristic patterns of abnormality are sometimes seen, although, as stated previously, failing gaits often look fundamentally similar. Cadence (steps per minute), velocity, and stride length can be recorded by timing a patient over a fixed distance. Watching the patient rise from a chair provides a good functional assessment of balance.

Brain imaging studies may be informative in patients with an undiagnosed disorder of gait. MRI is sensitive for cerebral lesions of vascular or demyelinating disease and is a good screening test for occult hydrocephalus. Patients with recurrent falls are at risk for subdural hematoma. As mentioned earlier, many elderly patients with gait and balance difficulty have white matter abnormalities in the periventricular region and centrum semiovale. While these lesions may be an incidental finding, a substantial burden of white matter disease will ultimately impact cerebral control of locomotion.

DISORDERS OF BALANCE

DEFINITION, ETIOLOGY, AND MANIFESTATIONS

Balance is the ability to maintain equilibrium—a dynamic state in which one’s center of mass is controlled with respect to the lower extremities, gravity, and the support surface despite external perturbations. The reflexes required to maintain upright posture require input from cerebellar, vestibular, and somatosensory systems; the premotor cortex and corticospinal and reticulospinal tracts mediate output to axial and proximal limb muscles. These responses are physiologically complex, and the anatomic representation they entail is not well understood. Failure can occur at any level and presents as difficulty maintaining posture while standing and walking.

The history and physical examination may differentiate underlying causes of imbalance. Patients with *cerebellar* ataxia do not generally complain of dizziness, although balance is visibly impaired. Neurologic examination reveals a variety of cerebellar signs. Postural compensation may prevent falls early on, but falls are inevitable with disease progression. The progression of neurodegenerative ataxia is often measured by the number of years to loss of stable ambulation.

Vestibular disorders **(Chap. 24)** have symptoms and signs that fall into three categories: (1) vertigo (the subjective inappropriate perception or illusion of movement); (2) nystagmus (involuntary eye movements); and (3) impaired standing balance. Not every patient has all manifestations. Patients with vestibular deficits related to ototoxic drugs may lack vertigo or obvious nystagmus, but their balance is impaired on standing and walking, and they cannot navigate in the dark. Laboratory testing is available to investigate vestibular deficits.

Somatosensory deficits also produce imbalance and falls. There is often a subjective sense of insecure balance and fear of falling. Postural control is compromised by eye closure (*Romberg's sign*); these patients also have difficulty navigating in the dark. A dramatic example is provided by the patient with autoimmune subacute sensory neuronopathy, which is sometimes a paraneoplastic disorder **(Chap. 99)**. Compensatory strategies enable such patients to walk in the virtual absence of proprioception, but the task requires active visual monitoring.

Patients with *higher-level disorders of equilibrium* have difficulty maintaining balance in daily life and may present with falls. Their awareness of balance impairment may be reduced. Patients taking sedating medications are in this category.

FALLS

Falls are common in the elderly. Over one-third of people aged >65 who are living in the community fall each year, and this number is even higher in nursing homes and hospitals. Elderly people are not only at higher risk for falls but are also more likely to suffer serious complications due to medical comorbidities such as osteoporosis. Hip fractures result in hospitalization, can lead to nursing home admission, and are associated with a substantially increased mortality risk in the subsequent year. Falls may result in brain or spinal injury, the history of which may be difficult for the patient to provide. The proportion of spinal cord injuries due to falls in individuals aged >65 years has doubled in the past decade, perhaps due to increasing activity in this age group. Some falls result in a prolonged time lying on the ground increasing risk for dehydration and rhabdomyolysis.

For each person who is physically disabled, there are others whose functional independence is limited by anxiety and fear of falling. Nearly one in five elderly individuals voluntarily restricts their activity because of fear of falling. With loss of ambulation, the quality of life diminishes, and rates of morbidity and mortality increase.

RISK FACTORS FOR FALLS

Risk factors for falls may be *intrinsic* (e.g., gait and balance disorders, visual impairment) or *extrinsic* (e.g., wet surfaces, improper footwear); some risk factors are modifiable. The presence of multiple risk factors is associated with a substantially increased risk of falls. Polypharmacy (use of four or more prescription medications) has also been identified as an important risk factor. **Table 28-3** summarizes the principal risk factors for falls.

TABLE 28-3 Common Risk Factors for Falls in Older Adults

RISK FACTOR
History of falls
Gait and/or balance disorder
Visual deficits
Physical disability
• Includes muscle weakness, use of assistive devices, osteoarthritis of the knee
Orthostatic hypotension
Depression
Cognitive impairment
Medications
• Polypharmacy, use of antipsychotics, antidepressants, benzodiazepines, anticholinergics, antihypertensives, and diuretics

ASSESSMENT OF THE PATIENT WITH FALLS

The most productive approach is to identify the high-risk patient prospectively, before there is a serious injury. All community-dwelling adults should be asked annually about falls and whether or not fear of falling limits daily activities. The Timed Up and Go (TUG) test involves timing a patient as they stand up from a chair, walk 10 feet, turn, and then sit down. Patients with a history of falls or those requiring >12 seconds to complete the TUG test are at high risk for falls and should undergo further assessment.

History The history surrounding a fall is often problematic or incomplete, and the underlying mechanism or cause may be difficult to establish in retrospect. A report from an attentive observer to the fall can be invaluable. If a sudden drop without provocation is reported, suspicion for syncope, seizure, or other neurologic event should rise. Gait freezing and festination should raise concern for parkinsonism. Falling after arising from a chair or after a minor perturbation suggests a muscular strength issue. Falls occurring in changing footing situations or in poor lighting suggest somatosensory, visual, or vestibular deficits as causes.

Patients should be queried about any provoking factors (including head turn, standing, carrying an object) or prodromal symptoms, such as dizziness, vertigo, presyncopal symptoms, or focal weakness. A history of baseline mobility and medical comorbidities should be elicited. Patients at particular risk include those with mental status changes or dementia. Medications should be reviewed, with particular attention to benzodiazepines, opioids, antipsychotics, antiepileptics, antidepressants, antiarrhythmics, and diuretics, all of which are associated with an increased risk of falls. It is equally important to distinguish *mechanical falls* (those caused by tripping or slipping) due to purely extrinsic or environmental factors from those in which a possibly modifiable intrinsic factor contributes. *Recurrent falls* suggest an underlying gait or balance disorder. Falls associated with loss of consciousness (syncope, seizure) may require appropriate cardiac **(Chap. 243)** or neurologic **(Chaps. 23 and 436)** evaluation and intervention, although a patient's report of change in consciousness may be unreliable.

Physical Examination Examination of the patient with falls should include a cardiac examination looking for rhythm irregularities or murmurs. Supine blood pressure should be compared both to standing pressures shortly after arising and after 3 minutes to evaluate for immediate or delayed orthostasis **(Chap. 451)**. A systolic drop of >20 mmHg, a diastolic drop >10 mmHg, or a marked increase in heart rate should be investigated.

Visual acuity can be assessed with a pocket or wall eye chart and performed with eyewear that the patient typically wears when walking. Peripheral fields can be assessed if suspicion for a visual field cut is present. Bilateral hearing assessments can be pursued, especially if a vestibular problem is suspected.

The patient's standing posture should be evaluated for spinal deformity (kyphosis, lordosis, scoliosis). Orthopedic complications such as arthritis can cause knee alignment deformity and overpronation at the ankle/hindfoot. Feet should be inspected for calluses, great toe deformation like hallux valgus, hammertoes, and loss of the arch. Footwear should be evaluated for fit to the top of the mid foot, attachment behind the calcaneus, and depth of tread and wear pattern on the sole.

Mental status is easily assessed while obtaining a history from the patient. Motor function in the lower extremities should be done with particular attention paid to hip abductors and flexors, knee extensors and flexors, and ankle dorsiflexors and plantar flexors. Tone assessment in these muscle groups can be useful when neuromuscular disease is suspected. A sensory exam is useful to determine if the patient has developed a neuropathy impacting how they feel the surface they are standing and walking upon. Monofilament testing, during which pressure applied to a small nylon filament is used to assess sensation over the plantar aspect of the foot, can be used to screen for neuropathy in patients with diabetes. Sensory testing should also include proprioceptive awareness at the great toe and ankle. A heel-to-shin test can assess cerebellar input on the lower extremity. The remainder of the neurologic examination should assess gait and balance as described earlier in this chapter.

Fall Patterns The description of a fall event may provide further clues to the underlying etiology. While there is no standard nosology of falls, some common clinical patterns may emerge and provide a clue.

DROP ATTACKS AND COLLAPSING FALLS Drop attacks and collapsing falls are associated with a sudden loss of postural tone. Patients may report that their legs just "gave out" underneath them or that they "collapsed in a heap." Syncope or orthostatic hypotension may be a factor in some such falls. Neurologic causes are relatively rare but include atonic seizures, myoclonus, and intermittent obstruction of the foramen of Monro by a colloid cyst of the third ventricle causing acute obstructive hydrocephalus. An emotional trigger suggests cataplexy. While collapsing falls are more common among older patients with vascular risk factors, drop attacks should not be confused with vertebrobasilar ischemic attacks.

TOPPLING FALLS Some patients maintain tone in antigravity muscles but fall over like a tree trunk, as if postural defenses had disengaged. Causes include cerebellar pathology and lesions of the vestibular system. There may be a consistent direction to such falls. Toppling falls are an early feature of PSP, and a late feature of Parkinson's disease, once postural instability has developed. Thalamic lesions causing truncal instability (*thalamic astasia*) may also contribute to this type of fall.

FALLS DUE TO GAIT FREEZING Freezing of gait is seen in Parkinson's disease and related disorders. The feet stick to the floor and the center of mass keeps moving, resulting in a disequilibrium from which the patient has difficulty recovering, resulting in a forward fall. Similarly, patients with Parkinson's disease and festinating gait may find their feet unable to keep up and may thus fall forward.

FALLS RELATED TO SENSORY LOSS Patients with somatosensory, visual, or vestibular deficits are prone to falls. These patients have particular difficulty dealing with poor illumination or walking on uneven ground. They often report subjective imbalance, apprehension, and fear of falling. These patients may be especially responsive to a rehabilitation-based intervention.

FALLS RELATED TO WEAKNESS Patients who lack strength in antigravity muscles have difficulty rising from a chair or maintaining their balance after a perturbation. These patients are often unable to get up after a fall and may have to remain on the floor for a prolonged period until help arrives. If due to deconditioning, this is often treatable. Resistance strength training can increase muscle mass and leg strength, even for people in their eighties and nineties.

TREATMENT

Interventions to Reduce the Risk of Falls and Injury

Efforts should be made to define the mechanism underlying falls in a given patient, as specific treatment may be possible once a diagnosis is established. Exercise should be recommended for everyone. Home- and group-based exercise programs focusing on leg strength and balance, physical therapy, and use of assistive devices reduce fall risk in individuals with a history of falls or disorders of gait and balance. Rehabilitative interventions aim to improve muscle strength and balance stability and to make the patient more resistant to injury and fracture. Resistance training with weights and machines is useful to improve muscle mass, even in frail older patients. Improvements realized in posture and gait should translate to reduced risk of falls and injury. Sensory balance training is another approach to improving balance stability. Measurable gains can be made in a few weeks of training, and benefits can be maintained over 6 months by a 10- to 20-min home exercise program. This strategy is particularly successful in patients with vestibular and somatosensory balance disorders. The Centers for Disease Control and Prevention's STEADI program (Stopping Elderly Accidents, Deaths and Injuries) is freely available online and contains tools for use by clinicians, patients, and their families to screen, assess, and offer interventions to minimize the risk of falls in older adults.

Orthostatic changes in blood pressure and pulse should be recorded. Contributory medications should be scrutinized and behavioral modifications initiated. Medications (including over-the-counter) should be reviewed, reevaluating benefits and burdens of medications that might increase fall risk. Treatment of cataracts and avoidance of multifocal lenses could be considered for patients whose falls may result from vision impairment. A home visit (often by an occupational therapist) to look for environmental hazards can be helpful. A variety of modifications may be recommended to improve safety, including improved lighting, installation of grab bars and nonslip surfaces, and use of adaptive equipment like walkers. Cognitive training, including dual-task training, may improve mobility in older adults with cognitive impairment. Adequate supervision appears to be most essential in this high-risk population.

FURTHER READING

American Geriatrics Society, British Geriatrics Society, American Academy of Orthopedic Surgeons Panel on Falls Prevention: Guideline for the prevention of falls in older persons. J Am Geriatr Soc 49:664, 2001.

Centers for Disease Control and Prevention: STEADI: Older Adult Fall Prevention. Available from *https://www.cdc.gov/steadi/index.html*. Accessed January 2, 2024.

Colón-Emeric CS et al: Risk assessment and prevention of falls in older community-dwelling adults: A review. JAMA 331:1397, 2024.

Nutt JG: Classification of gait and balance disorders. Adv Neurol 87:135, 2001.

Pirker W, Katzenschlager R: Gait disorders in adults and the elderly. Wien Klin Wochenschr 129:81, 2017.

29

Confusion and Delirium

S. Andrew Josephson, Bruce L. Miller

Confusion, a mental and behavioral state of reduced comprehension, coherence, and capacity to reason, is one of the most common problems encountered in medicine, accounting for a large number of emergency department visits, hospital admissions, and inpatient consultations. *Delirium*, a term used to describe an acute confusional state, remains a major cause of morbidity and mortality, costing billions of dollars yearly in health care costs in the United States alone. Despite increased efforts targeting awareness of this condition, delirium often goes unrecognized despite evidence that it is often the cognitive manifestation of serious underlying medical or neurologic illness.

CLINICAL FEATURES OF DELIRIUM

A multitude of terms are used to describe patients with delirium, including *encephalopathy, acute brain failure, acute altered mental status, acute confusional state*, and *postoperative* or *intensive care unit (ICU) psychosis*. Delirium has many clinical manifestations, but it is defined as a relatively acute decline in cognition that fluctuates over hours or days. The hallmark of delirium is a deficit of attention, although all cognitive domains—including memory, executive function, visuospatial tasks, and language—are variably involved. Associated symptoms that may be present in some cases include altered sleep-wake cycles, perceptual disturbances such as hallucinations or delusions, affect changes, and autonomic findings that include heart rate and blood pressure instability.

Delirium is a clinical diagnosis that is made only at the bedside. Two subtypes have been described—hyperactive and hypoactive—based on differential psychomotor features. The cognitive syndrome associated with severe alcohol withdrawal (i.e., "delirium tremens") remains the classic example of the hyperactive subtype, featuring prominent hallucinations, agitation, and hyperarousal, often accompanied by life-threatening autonomic instability. In striking contrast is the hypoactive subtype, exemplified by benzodiazepine intoxication, in which patients are withdrawn and quiet, with prominent apathy and psychomotor slowing.

This dichotomy between subtypes of delirium is a useful construct, but patients often fall somewhere along a spectrum between the hyperactive and hypoactive extremes, sometimes fluctuating from one to the other. Therefore, clinicians must recognize this broad range of presentations of delirium to identify all patients with this potentially reversible cognitive disturbance. Hyperactive patients are often easily recognized by their characteristic severe agitation, tremor, hallucinations, and autonomic instability. Patients who are quietly hypoactive are more often overlooked and underdiagnosed on the medical wards and in the ICU.

The reversibility of delirium is emphasized because many etiologies, such as infection and medication effects, can be treated easily. The long-term cognitive consequences of delirium remain an area of active research. Some episodes of delirium continue for weeks, months, or even years. The persistence of delirium in some patients and its high recurrence rate may be due to inadequate initial treatment of the underlying etiology, including neurodegenerative diseases. In other instances, delirium appears to cause permanent neuronal damage and long-term cognitive decline. Therefore, prevention strategies are important to implement. Even if an episode of delirium completely resolves, there may be lingering effects of the disorder; a patient's recall of events after delirium varies widely, ranging from complete amnesia to repeated reexperiencing of the frightening period of confusion, similar to what is seen in patients with posttraumatic stress disorder.

■ RISK FACTORS

An effective primary prevention strategy for delirium begins with identification of high-risk patients. Some hospital systems have initiated comprehensive delirium programs that screen most or all patients upon admission or before elective surgery; positive screens trigger a host of focused prevention measures. Multiple validated scoring systems have been developed as a screen for asymptomatic patients, many of which emphasize well-established risk factors for delirium.

The two most consistently identified risk factors are older age and baseline cognitive dysfunction. Individuals who are aged >65 or exhibit low scores on standardized tests of cognition develop delirium upon hospitalization at a rate approaching 50%. Whether age and baseline cognitive dysfunction are truly independent risk factors or are related is uncertain. Other predisposing factors include sensory deprivation, such as preexisting hearing and visual impairment, as well as indices for poor overall health, including baseline immobility, malnutrition, and underlying medical or neurologic illness.

In-hospital risks for delirium include the use of physical restraints, bladder catheterization, sleep and sensory deprivation, and the addition of three or more new medications. Avoiding such risks remains a key component of delirium prevention as well as treatment. Surgical and anesthetic risk factors for the development of postoperative delirium include procedures such as those involving cardiopulmonary bypass, inadequate or excessive treatment of pain in the immediate postoperative period, and perhaps specific agents such as inhalational anesthetics or benzodiazepines.

The relationship between delirium and dementia **(Chap. 31)** is complicated by significant overlap between the two conditions, and it is not always simple to distinguish between them. Dementia and preexisting cognitive dysfunction serve as major risk factors for delirium, and at least two-thirds of cases of delirium occur in patients with coexisting underlying dementia. A form of dementia with parkinsonism, *dementia with Lewy bodies* **(Chap. 445)**, is characterized by a fluctuating course, prominent visual hallucinations, parkinsonism, and an attentional deficit that clinically resembles hyperactive delirium; patients with this condition are particularly vulnerable to delirium. Delirium in the elderly often reflects an insult to a brain that is vulnerable due to an underlying neurodegenerative condition. Therefore, the development of delirium sometimes heralds the onset of a previously unrecognized brain disorder, and after the acute delirious episode has cleared, careful screening for an underlying condition should occur in the outpatient setting.

■ EPIDEMIOLOGY

Delirium is common, but its reported incidence has varied widely with the criteria used to define this disorder. Estimates of delirium in hospitalized patients range from 10% to >50%, with higher rates reported for elderly patients and patients undergoing hip surgery. Older patients in the ICU have especially high rates of delirium that approach 75%. The condition is not recognized in up to one-third of delirious inpatients, and the diagnosis is especially problematic in the ICU environment, where cognitive dysfunction is often difficult to appreciate in the setting of serious systemic illness and sedation. Delirium in the ICU should be viewed as an important manifestation of organ dysfunction not unlike liver, kidney, or heart failure. Outside the acute hospital setting, delirium occurs in nearly one-quarter of patients in nursing homes and in 50–80% of those at the end of life. These estimates emphasize the remarkably high frequency of this cognitive syndrome in older patients, a population that continues to grow.

An episode of delirium was previously viewed as a transient condition that carried a benign prognosis. It is now recognized as a disorder with substantial morbidity and mortality that often represents the first manifestation of a serious underlying illness. Estimates of in-hospital mortality rates among delirious patients range from 25% to 33%, similar to mortality rates due to sepsis. Patients with an in-hospital episode of delirium have a fivefold higher mortality rate in the months after their illness compared with age-matched nondelirious hospitalized patients. Delirious hospitalized patients also have a longer length of stay, are more likely to be discharged to a nursing home, have a higher frequency of readmission, and are more likely to experience subsequent episodes of delirium and cognitive decline; as a result, this condition has an enormous economic cost.

■ PATHOGENESIS

The pathogenesis and anatomy of delirium are incompletely understood. The attentional deficit that serves as the neuropsychological hallmark of delirium has a diffuse localization within the brainstem, thalamus, prefrontal cortex, and parietal lobes. Rarely, focal lesions such as ischemic strokes have led to delirium in otherwise healthy persons; right parietal and medial dorsal thalamic lesions have been reported most commonly, pointing to the importance of these areas in delirium pathogenesis. In most cases, however, delirium results from widespread disturbances in cortical and subcortical regions of the brain. Electroencephalogram (EEG) often reveals symmetric slowing, a nonspecific finding that supports diffuse cerebral dysfunction.

Multiple neurotransmitter abnormalities, proinflammatory factors, and specific genes likely play a role in the pathogenesis of delirium. Deficiency of acetylcholine may play a key role, and medications with anticholinergic properties can commonly precipitate delirium. As noted earlier, patients with preexisting dementia are particularly susceptible to episodes of delirium. Alzheimer's disease **(Chap. 442)**, dementia with Lewy bodies **(Chap. 445)**, and Parkinson's disease dementia **(Chap. 446)** are all associated with cholinergic deficiency due to degeneration of acetylcholine-producing neurons in the basal forebrain. In addition, other neurotransmitters are also likely to be involved in this diffuse cerebral disorder. For example, increases in dopamine can lead to delirium, and patients with Parkinson's disease treated with dopaminergic medications can develop a delirium-like state that features visual hallucinations, fluctuations, and confusion.

Not all individuals exposed to the same insult will develop signs of delirium. A low dose of an anticholinergic medication may have no cognitive effects on a healthy young adult but produce a florid delirium in an elderly person with known underlying dementia, although even healthy young persons develop delirium with very high doses of anticholinergic medications. This concept of delirium developing as the result of an insult in predisposed individuals is currently the most widely accepted pathogenic construct. Therefore, if a previously healthy individual with no known history of cognitive illness develops delirium in the setting of a relatively minor insult such as elective surgery, a urinary tract infection, or hospitalization, an unrecognized underlying neurologic illness such as a neurodegenerative disease, multiple previous strokes, or another diffuse cerebral cause should be considered. In this context, delirium can be viewed as a "stress test for the brain" whereby exposure to known inciting factors such as systemic infection and offending drugs can unmask a decreased cerebral reserve and herald a serious underlying and potentially treatable illness. New blood-based biomarkers for specific dementias may soon be available to help predict people at risk for delirium prior to surgical procedures or hospitalization.

APPROACH TO THE PATIENT

Delirium

Because the diagnosis of delirium is clinical and is made at the bedside, a careful history and physical examination are necessary in evaluating patients with possible confusional states. Screening tools can aid physicians and nurses in identifying patients with delirium, including the Confusion Assessment Method (CAM); the Nursing Delirium Screening Scale (NuDESC); the Organic Brain Syndrome Scale; the Delirium Rating Scale; and, in the ICU, the ICU version of the CAM and the Delirium Detection Score. Using the well-validated CAM, a diagnosis of delirium is made if there is (1) an acute onset and fluctuating course and (2) inattention accompanied by either (3) disorganized thinking or (4) an altered level of consciousness. These scales may not identify the full spectrum of patients with delirium, and all patients who are acutely confused should be presumed delirious regardless of their presentation due to the wide variety of possible clinical features. A course that fluctuates over hours or days and may worsen at night (termed *sundowning*) is typical but not essential for the diagnosis. Observation will usually reveal an altered level of consciousness or a deficit of attention. Other features that are only sometimes present include alteration of sleep-wake cycles, thought disturbances such as hallucinations or delusions, autonomic instability, and changes in affect.

HISTORY

It may be difficult to elicit an accurate history in delirious patients who have altered levels of consciousness or impaired attention. Information from a collateral source such as a spouse or another family member is therefore invaluable. The three most important pieces of history are the patient's baseline cognitive function, the time course of the present illness, and current medications.

Premorbid cognitive function can be assessed through the collateral source or, if needed, via a review of outpatient records. Delirium by definition represents a change from a cognitive baseline that is relatively acute and usually develops over hours to days. An acute confusional state is nearly impossible to diagnose without some knowledge of baseline cognitive function. Without this information, many patients with dementia or longstanding depression may be mistaken as delirious during a single initial evaluation. Patients with a more hypoactive, apathetic presentation with psychomotor slowing may be identified as being different from baseline only through conversations with family members. A number of validated instruments have been shown to diagnose cognitive dysfunction accurately using a collateral source, including the modified Blessed Dementia Rating Scale and the Clinical Dementia Rating (CDR). Baseline cognitive impairment is common in patients with delirium. Even when no such history of cognitive impairment is elicited, there should still be a high suspicion for a previously unrecognized underlying neurologic disorder.

Establishing the time course of cognitive change is important not only to make a diagnosis of delirium but also to correlate the onset of the illness with potentially treatable etiologies such as recent medication changes or systemic infection.

Medications remain a common cause of delirium, especially compounds with anticholinergic or sedative properties. It is estimated that nearly one-third of all cases of delirium are secondary to medications, especially in the elderly. Medication histories should include all prescription as well as over-the-counter and herbal substances taken by the patient and any recent changes in dosing or formulation, including substitution of generics for brand-name medications.

Other important elements of the history include screening for symptoms of organ failure or systemic infection, which often contributes to delirium in the elderly. A history of illicit drug use, alcoholism, or toxin exposure is important, particularly in younger delirious patients. Finally, asking the patient and collateral source about other symptoms that may accompany delirium, such as depression or anxiety, may help identify potential therapeutic targets.

PHYSICAL EXAMINATION

The general physical examination in a delirious patient should include careful screening for signs of infection such as fever, tachypnea, pulmonary consolidation, heart murmur, and meningismus. The patient's fluid status should be assessed; both dehydration and fluid overload with resultant hypoxemia have been associated with delirium, and each is usually easily rectified. The appearance of the skin can be helpful, showing jaundice in hepatic encephalopathy, cyanosis in hypoxemia, or needle tracks in patients using intravenous drugs.

The neurologic examination requires a careful assessment of mental status. Patients with delirium often present with a fluctuating course; therefore, the diagnosis can be missed when one relies on a single time point of evaluation. For patients who worsen in the evening (sundowning), assessment only during morning rounds may be falsely reassuring.

An altered level of consciousness ranging from hyperarousal to lethargy to coma is present in most patients with delirium and can be assessed easily at the bedside. In a patient with a relatively normal level of consciousness, a screen for an attentional deficit is in order, because this deficit is the classic neuropsychological hallmark of delirium. Attention can be assessed while taking a history from the patient. Tangential speech, a fragmentary flow of ideas, or inability to follow complex commands often signifies an attentional problem. There are formal neuropsychological tests to assess attention, but a simple bedside test of digit span forward is quick and fairly sensitive. In this task, patients are asked to repeat successively longer random strings of digits beginning with two digits in a row, said to the patient at one per second intervals. Healthy adults can repeat a string of five to seven digits before faltering; a digit span of four or less usually indicates an attentional deficit unless hearing or language barriers are present, and many patients with delirium have forward digit spans of three or fewer digits.

More formal neuropsychological testing can be helpful in assessing a delirious patient, but it is usually too cumbersome and time-consuming in the inpatient setting. A Mini-Mental State Examination (MMSE) provides information regarding orientation, language, and visuospatial skills **(Chap. 31)**; however, performance of many tasks on the MMSE, including the spelling of "world" backward and serial subtraction of digits, will be impaired by delirious patients' attentional deficits, rendering the test unreliable.

The remainder of the screening neurologic examination should focus on identifying new focal neurologic deficits. Focal strokes or mass lesions in isolation are rarely the cause of delirium, but patients with underlying extensive cerebrovascular disease or neurodegenerative conditions may not be able to cognitively tolerate even relatively small new insults. Patients should be screened for other signs of neurodegenerative conditions such as parkinsonism, which is seen not only in idiopathic Parkinson's disease but also in other dementing conditions including Alzheimer's disease, dementia with Lewy bodies, and progressive supranuclear palsy. The presence of multifocal myoclonus or asterixis on the motor examination is nonspecific but usually indicates a metabolic or toxic etiology of the delirium.

ETIOLOGY

Some etiologies can be easily discerned through a careful history and physical examination, whereas others require confirmation with laboratory studies, imaging, or other ancillary tests. A large, diverse group of insults can lead to delirium, and the cause in many patients is multifactorial. Common etiologies are listed in **Table 29-1**.

Prescribed, over-the-counter, and herbal medications all can precipitate delirium. Drugs with anticholinergic properties, narcotics, and benzodiazepines are particularly common offenders, but nearly any compound can lead to cognitive dysfunction in a predisposed patient. Whereas an elderly patient with baseline dementia may become delirious upon exposure to a relatively low dose of a medication, in less susceptible individuals, delirium occurs only with very high doses of the same medication. This observation emphasizes the importance of correlating the timing of recent medication changes, including dose and formulation, with the onset of cognitive dysfunction.

In younger patients, illicit drugs and toxins are common causes of delirium. In addition to more classic drugs of abuse, the availability of fentanyl **(Chap. 467)**, synthetic cannabis **(Chap. 466)**, "bath salts," methylenedioxymethamphetamine (MDMA, ecstasy), γ-hydroxybutyrate (GHB), and the phencyclidine (PCP)-like agent ketamine has led to an increase in delirious young persons presenting to acute care settings **(Chap. 468)**. Many common prescription drugs such as oral narcotics and benzodiazepines are often abused and readily available on the street. Alcohol abuse leading to high serum levels causes confusion, but more commonly, it is withdrawal from alcohol that leads to a hyperactive delirium **(Chap. 464)**. Alcohol and benzodiazepine withdrawal should be considered in all cases of delirium, including in the elderly, because even patients who drink only a few servings of alcohol every day can experience relatively severe withdrawal symptoms upon hospitalization.

Metabolic abnormalities such as electrolyte disturbances of sodium, calcium, magnesium, or glucose can cause delirium, and mild derangements can lead to substantial cognitive disturbances in susceptible individuals. Other common metabolic etiologies include liver and renal failure, hypercarbia and hypoxemia, vitamin deficiencies of thiamine and B_{12}, autoimmune disorders including central nervous system (CNS) vasculitis, and endocrinopathies such as thyroid and adrenal disorders.

Systemic infections often cause delirium, especially in the elderly. A common scenario involves the development of an acute cognitive decline in the setting of a urinary tract infection in a patient with baseline dementia. Pneumonia, skin infections such as cellulitis, and frank sepsis also lead to delirium. This so-called septic encephalopathy, often seen in the ICU, is probably due to the release of proinflammatory cytokines and their diffuse cerebral effects. CNS infections such as meningitis, encephalitis, and abscess are less common etiologies of delirium, as are cases of autoimmune or paraneoplastic encephalitis; however, in light of the high morbidity and mortality rates associated with these conditions when they are not treated, clinicians must always maintain a high index of suspicion.

In some susceptible individuals, exposure to the unfamiliar environment of a hospital itself can contribute to delirium. This etiology usually occurs as part of a multifactorial delirium and should be considered a diagnosis of exclusion after all other causes have been thoroughly investigated. Many primary prevention and treatment strategies for delirium involve relatively simple methods to address the aspects of the inpatient setting that are most confusing.

TABLE 29-1 Differential Diagnosis of Delirium

Toxins
Prescription medications: especially those with anticholinergic properties, narcotics, and benzodiazepines
Drugs of abuse: alcohol intoxication and alcohol withdrawal, opiates, ecstasy, LSD, GHB, PCP, ketamine, cocaine, "bath salts," marijuana and its synthetic forms
Poisons: inhalants, carbon monoxide, ethylene glycol, pesticides
Metabolic Conditions
Electrolyte disturbances: hypoglycemia, hyperglycemia, hyponatremia, hypernatremia, hypercalcemia, hypocalcemia, hypomagnesemia
Hypothermia and hyperthermia
Pulmonary failure: hypoxemia and hypercarbia
Liver failure/hepatic encephalopathy
Renal failure/uremia
Cardiac failure
Vitamin deficiencies: B_{12}, thiamine, folate, niacin
Dehydration and malnutrition
Anemia
Infections
Systemic infections: urinary tract infections, pneumonia and other respiratory infections, skin and soft tissue infections, sepsis
CNS infections: meningitis, encephalitis, brain abscess
Endocrine Conditions
Hyperthyroidism, hypothyroidism
Hyperparathyroidism
Adrenal insufficiency
Cerebrovascular Disorders
Global hypoperfusion states
Hypertensive encephalopathy
Focal ischemic strokes and hemorrhages (rare): especially nondominant parietal and thalamic lesions
Autoimmune Disorders
CNS vasculitis
Cerebral lupus
Neurologic paraneoplastic and autoimmune encephalitis
Seizure-Related Disorders
Nonconvulsive status epilepticus
Intermittent seizures with prolonged postictal states
Neoplastic Disorders
Diffuse metastases to the brain
Diffuse glioma
Carcinomatous meningitis
CNS lymphoma
Hospitalization
Terminal end-of-life delirium

Abbreviations: CNS, central nervous system; GHB, γ-hydroxybutyrate; LSD, lysergic acid diethylamide; PCP, phencyclidine.

Cerebrovascular etiologies of delirium are usually due to global hypoperfusion in the setting of systemic hypotension from heart failure, septic shock, dehydration, or anemia. Focal strokes in the right parietal lobe and medial dorsal thalamus rarely can lead to a delirious state. A more common scenario involves a new focal stroke or hemorrhage causing confusion in a patient who has decreased cerebral reserve. In these individuals, it is sometimes difficult to distinguish between cognitive dysfunction resulting from the new neurovascular insult itself and delirium due to the infectious, metabolic, and pharmacologic complications that can accompany hospitalization after stroke.

Because a fluctuating course often is seen in delirium, intermittent seizures may be overlooked when one is considering potential etiologies. Both nonconvulsive status epilepticus and recurrent focal or generalized seizures followed by postictal confusion can cause delirium; EEG remains essential for this diagnosis and should be considered whenever the etiology of delirium remains unclear following initial workup. Seizure activity spreading from an electrical focus in a mass or infarct can explain global cognitive dysfunction caused by relatively small lesions.

It is extremely common for patients to experience delirium at the end of life in palliative care settings. This condition must be identified and treated aggressively because it is an important cause of patient and family discomfort at the end of life. It should be remembered that these patients also may be suffering from more common etiologies of delirium such as systemic infection.

LABORATORY AND DIAGNOSTIC EVALUATION

A cost-effective approach allows the history and physical examination to guide further tests after initial screening laboratory studies are obtained. No single algorithm will fit all delirious patients due to the staggering number of potential etiologies, but one stepwise approach is detailed in **Table 29-2**. If a clear precipitant such as an offending medication is identified, further testing may not be required. If, however, no likely etiology is uncovered with initial evaluation, an aggressive search for an underlying cause should be initiated.

TABLE 29-2 Stepwise Evaluation of a Patient with Delirium

Initial Evaluation
History with special attention to medications (including over-the-counter and herbals)
General physical examination and neurologic examination
Complete blood count
Electrolyte panel including calcium, magnesium, phosphorus
Liver function tests, including albumin
Renal function tests
First-Tier Further Evaluation Guided by Initial Evaluation
Systemic infection screen
Urinalysis and culture
Chest radiograph and tests for respiratory pathogens
Blood cultures
Electrocardiogram
Arterial blood gas
Serum and/or urine toxicology screen (perform earlier in young persons)
Brain imaging with MRI with diffusion and gadolinium (preferred) or CT
Suspected CNS infection or other inflammatory disorder: lumbar puncture after brain imaging
Suspected seizure-related etiology: electroencephalogram (EEG) (if high suspicion, should be performed immediately)
Second-Tier Further Evaluation
Vitamin levels: B_{12}, folate, thiamine
Endocrinologic laboratories: thyroid-stimulating hormone (TSH) and free T_4; cortisol
Serum ammonia
Sedimentation rate
Autoimmune serologies: antinuclear antibodies (ANA), complement levels; p-ANCA, c-ANCA, consider paraneoplastic/autoimmune encephalitis testing in the serum and CSF
Infectious serologies: rapid plasmin reagin (RPR); fungal and viral serologies if high suspicion; HIV antibody
Lumbar puncture (if not already performed)
Brain MRI with and without gadolinium (if not already performed)

Abbreviations: c-ANCA, cytoplasmic antineutrophil cytoplasmic antibody; CNS, central nervous system; CSF, cerebrospinal fluid; CT, computed tomography; MRI, magnetic resonance imaging; p-ANCA, perinuclear antineutrophil cytoplasmic antibody.

Basic screening lab tests, including a complete blood count, electrolyte panel, and tests of liver and renal function, should be obtained in all patients with delirium. In elderly patients, screening for systemic infection, including chest radiography, urinalysis and culture, and possibly blood cultures, is important. In younger individuals, serum and urine drug and toxicology screening may be appropriate earlier in the workup. Additional laboratory tests addressing other autoimmune, endocrinologic, metabolic, and infectious etiologies should be reserved for patients in whom the diagnosis remains unclear after initial testing.

Multiple studies have demonstrated that brain imaging in patients with delirium is often unhelpful. If, however, the initial workup is unrevealing, most clinicians quickly move toward imaging of the brain to exclude structural causes. A noncontrast computed tomography (CT) scan can identify large masses and hemorrhages but is otherwise unlikely to help determine an etiology of delirium. The ability of magnetic resonance imaging (MRI) to identify nearly all acute ischemic strokes as well as to provide neuroanatomic detail that gives clues to possible infectious, inflammatory, neurodegenerative, and neoplastic conditions makes it the test of choice. Because MRI techniques are limited by availability, speed of imaging, patient's cooperation, and contraindications, many clinicians begin with CT scanning and proceed to MRI if the etiology of delirium remains elusive.

Lumbar puncture (LP) must be obtained immediately after neuroimaging for all patients in whom CNS infection is suspected. Spinal fluid examination can also be useful in identifying autoimmune, other inflammatory, and neoplastic conditions. As a result, LP should be considered in any delirious patient with a negative workup. EEG remains invaluable if seizures are considered or if there is no cause readily identified.

TREATMENT

Delirium

Management of delirium begins with treatment of the underlying inciting factor (e.g., patients with systemic infections should be given appropriate antibiotics, and underlying electrolyte disturbances should be judiciously corrected). These treatments often lead to prompt resolution of delirium. Blindly targeting the symptoms of delirium pharmacologically only serves to prolong the time patients remain in the confused state and may mask important diagnostic information.

Relatively simple methods of supportive care can be highly effective **(Fig. 29-1)**. Reorientation by the nursing staff and family combined with visible clocks, calendars, and outside-facing windows can reduce confusion. Sensory isolation should be prevented by providing glasses and hearing aids to patients who need them. Sundowning can be addressed to a large extent through vigilance to appropriate sleep-wake cycles. During the day, a well-lit room should be accompanied by activities or exercises to prevent napping. At night, a quiet, dark environment with limited interruptions by staff can assure proper rest; melatonin can be considered before bed to promote sleep. These sleep-wake cycle interventions are especially important in the ICU setting as the usual constant 24-h activity commonly provokes delirium. Attempting to mimic the home environment as much as possible also has been shown to help treat and even prevent delirium. Visits from friends and family throughout the day minimize the anxiety associated with the constant flow of new faces of staff and physicians. Allowing hospitalized patients to have access to home bedding, clothing, and nightstand objects makes the hospital environment less foreign and therefore less confusing. Simple standard nursing practices such as maintaining proper nutrition and volume status as well as managing pain, incontinence, and skin breakdown also help alleviate discomfort and resulting confusion.

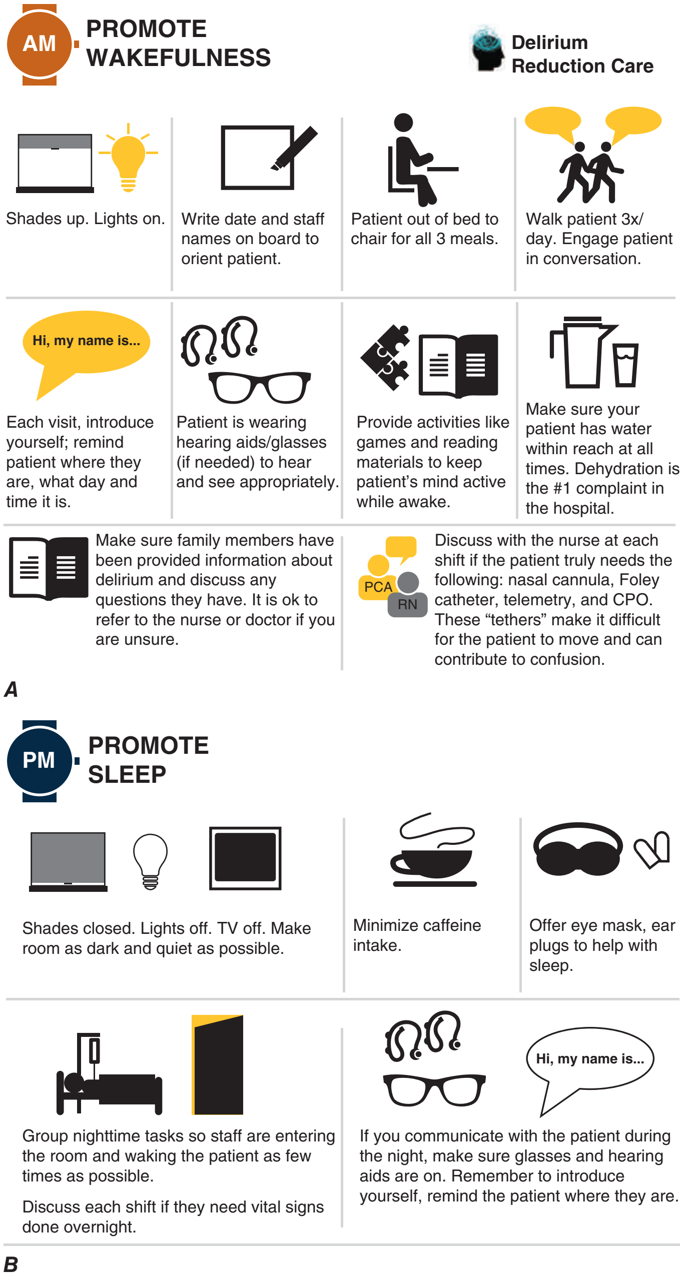

Shades up. Lights on.

Write date and staff names on board to orient patient.

Patient out of bed to chair for all 3 meals.

Walk patient 3x/day. Engage patient in conversation.

Each visit, introduce yourself; remind patient where they are, what day and time it is.

Patient is wearing hearing aids/glasses (if needed) to hear and see appropriately.

Provide activities like games and reading materials to keep patient's mind active while awake.

Make sure your patient has water within reach at all times. Dehydration is the #1 complaint in the hospital.

Make sure family members have been provided information about delirium and discuss any questions they have. It is ok to refer to the nurse or doctor if you are unsure.

Discuss with the nurse at each shift if the patient truly needs the following: nasal cannula, Foley catheter, telemetry, and CPO. These "tethers" make it difficult for the patient to move and can contribute to confusion.

Shades closed. Lights off. TV off. Make room as dark and quiet as possible.

Minimize caffeine intake.

Offer eye mask, ear plugs to help with sleep.

Group nighttime tasks so staff are entering the room and waking the patient as few times as possible.

Discuss each shift if they need vital signs done overnight.

If you communicate with the patient during the night, make sure glasses and hearing aids are on. Remember to introduce yourself, remind the patient where they are.

FIGURE 29-1 Delirium management and prevention: a checklist for hospitalized patients. Effective management of delirium relies on broad efforts to promote wakefulness (***A***) and sleep (***B***). CPO, continuous pulse oximetry.

In some instances, patients pose a threat to their own safety or to the safety of staff members, and acute management is required. Bed alarms and personal sitters are more effective and much less disorienting than physical restraints. Chemical restraints should be avoided, but when necessary, very-low-dose typical or atypical antipsychotic medications administered on an as-needed basis can be used, recognizing that multiple clinical trials have consistently shown that these medications are ineffective in preventing or treating delirium. Therefore, they should be reserved for patients who display severe agitation and significant potential to harm themselves or staff. The association of antipsychotic drug use in the elderly with increased mortality rates underscores the importance of using these medications judiciously and only as a last resort. Benzodiazepines often worsen confusion through their sedative properties. Although many clinicians use benzodiazepines to treat acute confusion, their use should be limited to cases in which delirium is caused by alcohol or benzodiazepine withdrawal.

PREVENTION

In light of the high morbidity associated with delirium and the tremendously increased health care costs that accompany it, development of an effective strategy to prevent delirium in hospitalized patients is extremely important. Successful identification of high-risk patients is the first step, followed by initiation of appropriate interventions. Increasingly, hospitals are using nursing or physician-administered tools to screen for high-risk individuals, triggering simple standardized protocols used to manage risk factors for delirium, including sleep-wake cycle reversal, immobility, visual impairment, hearing impairment, sleep deprivation, and dehydration. No specific medications have been definitively shown to be effective for delirium prevention, including trials of cholinesterase inhibitors and antipsychotic agents. Melatonin and its agonist ramelteon have shown some promising results in small preliminary trials. Recent studies in the ICU have focused both on identifying sedatives, such as dexmedetomidine, that are less likely to lead to delirium in critically ill patients and on developing protocols for daily awakenings in which infusions of sedative medications are interrupted and the patient is reorientated by the staff. All hospitals and health care systems should work toward decreasing the incidence of delirium and promptly recognizing and treating the disorder when it occurs.

FURTHER READING

Goldberg TE et al: Association of delirium with long-term cognitive decline: A meta-analysis. JAMA Neurol 77:1, 2020.

Kunicki ZJ et al: Six-year cognitive trajectory in older adults following major surgery and delirium. JAMA Int Med 183:442, 2023.

Livingston G et al: Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 396:413, 2020.

Ng KT et al: The effect of dexmedetomidine on delirium and agitation in patients in intensive care: Systematic review and meta-analysis with trial sequential analysis. Anesthesia 74:380, 2019.

Smit L et al: Efficacy of haloperidol to decrease the burden of delirium in adult critically ill patients: The EuRIDICE randomized clinical trial. Crit Care 27:413, 2023.

30

Coma

S. Andrew Josephson, Allan H. Ropper, Stephen L. Hauser

Coma is among the most common neurologic emergencies encountered in general medicine and requires an organized approach. It accounts for a substantial portion of emergency department visits and occurs on all services in the hospital.

There exists a continuum of states of reduced alertness, the most severe form being coma, defined as a deep sleeplike state with eyes closed, from which the patient cannot be aroused. Stupor refers to a lower threshold for arousability, in which the patient can be transiently awakened by vigorous stimuli, accompanied by motor behavior that leads to avoidance or withdrawal from noxious stimuli. Drowsiness simulates light sleep and is characterized by easy arousal that may persist for brief periods. Stupor and drowsiness are usually accompanied by some degree of confusion when the patient is alerted **(Chap. 29)**. A precise narrative description of the level of arousal and of the type of responses evoked by various stimuli as observed at the bedside is preferable to use of more ambiguous terms such as lethargy, semicoma, or obtundation.

Several conditions that render patients unresponsive and simulate coma are considered separately because of their special significance. The *vegetative state* signifies an awake-appearing but nonresponsive state, usually encountered in a patient who has emerged from coma. In the vegetative state, the eyelids typically open periodically, giving the appearance of wakefulness. Respiratory and autonomic functions are retained. Yawning, coughing, swallowing, and limb and head movements persist, but there are few, if any, meaningful responses to the external and internal environment. There are typically accompanying signs that indicate extensive damage in both cerebral hemispheres, e.g., decerebrate or decorticate limb posturing and absent responses to visual stimuli (see below). In the related but less severe *minimally conscious state*, the patient displays rudimentary vocal or motor behaviors, often spontaneous, but sometimes in response to touch, visual stimuli, or command. Cardiac arrest with cerebral hypoperfusion and head trauma are the most common causes of the vegetative and minimally conscious states **(Chap. 318)**.

The prognosis for regaining meaningful mental faculties once the vegetative state has supervened for several months is poor, and after a year, almost nil; hence the term *persistent vegetative state*. Most reports of dramatic recovery, when investigated carefully, are found to yield to the usual rules for prognosis, but there have been rare instances in which recovery has occurred to a severely disabled condition and, particularly in childhood cases, to an even better state. Patients in the minimally conscious state carry a better prognosis for some recovery compared to those in a persistent vegetative state, but even in these patients, dramatic recovery after 12 months is unusual.

The possibility of incorrectly attributing meaningful behavior to patients in the vegetative and minimally conscious states creates problems and anguish for families and physicians. The question of whether some of these patients have the capability for cognition has been investigated by functional MRI and electroencephalogram (EEG) studies that have demonstrated cerebral activation that is temporally consistent in response to verbal and other stimuli, as discussed in more detail below. This finding suggests at a minimum that some of these patients could in the future be able to communicate using technological advances and that further research could shed light on treatment approaches targeting areas of the brain and their connections that appear to be preserved in such individuals.

Several syndromes that affect alertness are prone to be misinterpreted as stupor or coma, and clinicians should be aware of these pitfalls when diagnosing coma at the bedside. Akinetic mutism refers to a partially or fully awake state in which the patient remains virtually immobile and mute but can form impressions and think, as demonstrated by later recounting of events. This condition results from damage in the regions of the medial thalamic nuclei or the frontal lobes (particularly lesions situated deeply or on the orbitofrontal surfaces) or from extreme hydrocephalus. The term *abulia* describes a milder form of akinetic mutism characterized by mental and physical slowness and diminished ability to initiate activity. It is also usually the result of damage to the medial frontal lobes and their connections **(Chap. 32)**.

Catatonia is a hypomobile and mute syndrome that occurs usually as part of a major psychiatric disorder, typically schizophrenia or major depression. Catatonic patients make few voluntary or responsive movements, although they blink, swallow, and may not appear distressed. There are nevertheless signs that the patient is responsive, although it takes a careful examination to demonstrate these features. For example, eyelid elevation is often actively resisted, blinking occurs in response to a visual threat, and the eyes move concomitantly with head rotation, all of which are inconsistent with the presence of a brain lesion causing unresponsiveness. The limbs may retain postures in which they have been placed by the examiner ("waxy flexibility," or catalepsy). There may be remarkable reversal, albeit sometimes brief, of the signs of catatonia when lorazepam is administered, validating the diagnosis. With appropriate medical therapy and recovery from catatonia, patients often have some memory of events that occurred during their stupor. Catatonia is superficially similar to akinetic mutism, but clinical evidence of cerebral damage such as hyperreflexia and hypertonicity of the limbs is lacking in the former.

The locked-in state describes a type of pseudocoma in which an awake but paralyzed patient has no means of producing speech or volitional limb movement but retains voluntary vertical eye movements and lid elevation, thus allowing the patient to communicate. The pupils are normally reactive. The usual cause is an infarction (e.g., basilar artery thrombosis) or hemorrhage of the ventral pons bilaterally that transects all descending motor (corticospinal and corticobulbar) pathways and those that control horizontal eye movements. Another awake but de-efferented state occurs as a result of total paralysis of the musculature in severe cases of neuromuscular weakness such as in forms of Guillain-Barré syndrome **(Chap. 458)**, critical illness neuropathy **(Chap. 318)**, or pharmacologic neuromuscular blockade.

THE ANATOMY AND PHYSIOLOGY OF COMA

Almost all instances of coma can be traced to either (1) widespread abnormalities of both cerebral hemispheres or (2) reduced activity of the thalamocortical alerting system, the reticular activating system (RAS), which is an assemblage of neurons located diffusely in the upper brainstem and thalamus. The proper functioning of this system, its ascending projections to the cortex, and the cortex itself are required to maintain alertness and coherence of thought. In addition to structural damage to either or both of these systems, suppression of reticulocerebral function commonly occurs by drugs, toxins, or metabolic derangements such as hypoglycemia, anoxia, uremia, and hepatic failure, or by seizures; these types of metabolic causes of coma are far more common than structural injuries.

Coma Due to Cerebral Mass Lesions and Herniation Syndromes The skull prevents outward expansion of the brain, and infoldings of the dura create compartments that restrict displacement of brain tissue within the cranium. The two cerebral hemispheres are separated by the falx and the anterior and posterior fossae by the tentorium. Herniation refers to displacement of brain tissue by an intracerebral or overlying mass into a contiguous compartment that it normally does not occupy. Coma from mass lesions, and many of its associated signs, are attributable to these tissue shifts, and certain clinical features are characteristic of specific configurations of herniation **(Fig. 30-1)**.

In the most common form of herniation, brain tissue is displaced from the supratentorial to the infratentorial compartment through the tentorial opening, referred to as transtentorial herniation. The cause is often a mass hemispheral lesion that is evident by contralateral hemiparesis and other cerebral signs. Uncal transtentorial herniation refers to impaction of the anterior medial temporal gyrus (the uncus) into the tentorial opening anterior to and adjacent to the midbrain **(Fig. 30-1*A*)**. The uncus can compress the third nerve as the nerve traverses the subarachnoid space, causing enlargement of the ipsilateral pupil as the first sign (the fibers subserving parasympathetic pupillary function are located peripherally in the nerve). The

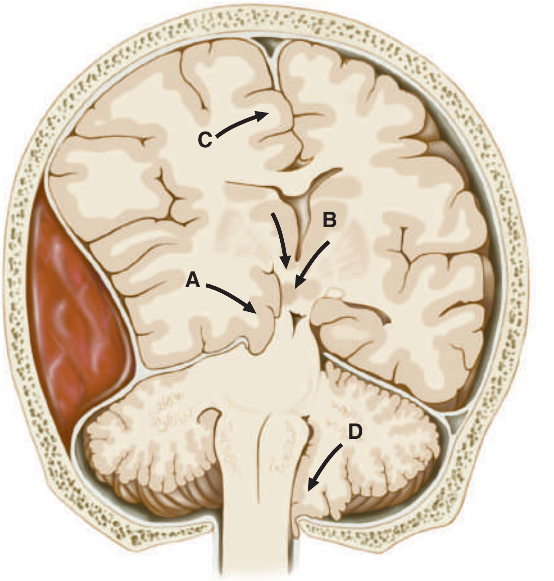

FIGURE 30-1 Types of cerebral herniation: (***A***) uncal; (***B***) central; (***C***) transfalcial; and (***D***) foraminal.

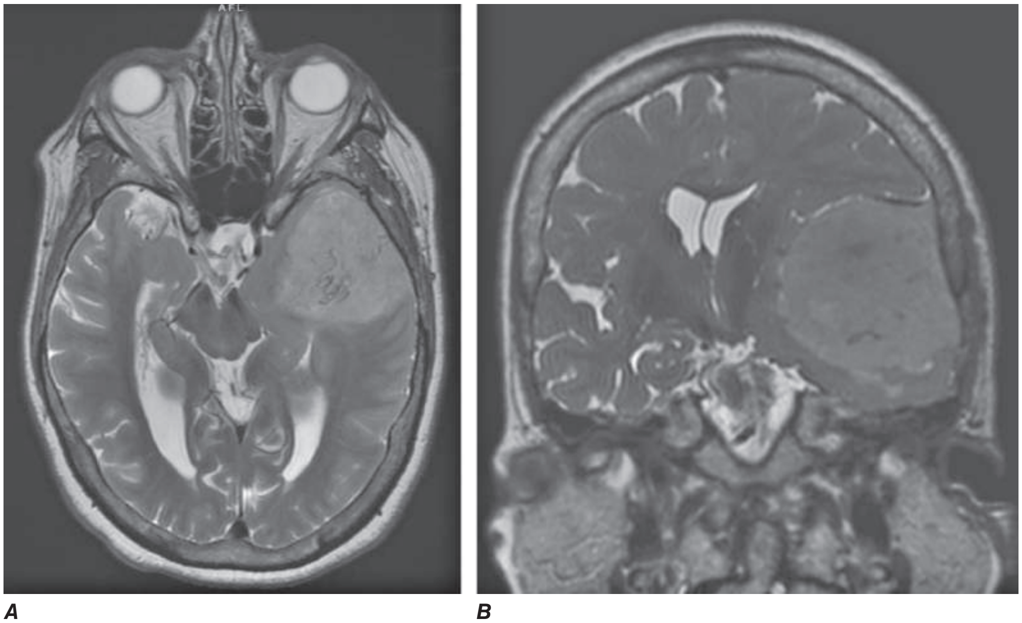

FIGURE 30-2 Axial (A) and coronal (B) T2-weighted magnetic resonance images from a stuporous patient with a left third nerve palsy from a large left-sided meningioma. **A.** The upper midbrain is compressed and displaced horizontally away from the mass, and there is transtentorial herniation of the medial temporal lobe structures, including the uncus. **B.** The lateral ventricle opposite to the mass has become enlarged as a result of compression of the third ventricle.

coma that typically follows is due to vertical and lateral displacement of the midbrain (and therefore the RAS) against the opposite tentorial edge by the displaced parahippocampal gyrus **(Fig. 30-2)**. An additional feature of lateral displacement can be compression of the opposite cerebral peduncle, producing a Babinski sign and hemiparesis ipsilateral to the supratentorial mass (the Kernohan-Woltman sign). Herniation may also compress the anterior and posterior cerebral arteries as they pass over the tentorial reflections, with resultant brain infarction. These distortions may also entrap portions of the ventricular system, causing hydrocephalus.

Central transtentorial herniation denotes a symmetric downward movement of the thalamic structures through the tentorial opening with compression of the upper midbrain **(Fig. 30-1*B*)**. Miotic pupils and drowsiness are the heralding signs, in contrast to a unilaterally enlarged pupil of the uncal syndrome. Both uncal and central transtentorial herniations cause progressive compression of the brainstem and RAS, with initial damage to the rostral structures in midbrain, then the pons, and finally the most caudal medulla. The result is an approximate sequence of neurologic signs that corresponds to each affected level, with respiratory centers in the brainstem often spared until late in the herniation syndrome. This orderly sequential appearance of signs is not always respected, and rapid deterioration reflecting disruption of many brainstem functions can occur. Other forms of herniation include transfalcial herniation (displacement of the cingulate gyrus under the falx and across the midline, **Fig. 30-1*C***) and foraminal herniation (downward forcing of the cerebellar tonsils into the foramen magnum, **Fig. 30-1*D***), which causes early compression of the medulla and respiratory arrest.

Coma Due to Metabolic, Drug, and Toxic Disorders Many systemic metabolic abnormalities cause coma by interrupting the delivery of energy substrates (e.g., oxygen, glucose) or by altering neuronal excitability (drugs and alcohol, anesthetics, and epilepsy). These are the most common causes of coma in general practice and in large case series. The metabolic abnormalities that produce coma in milder forms may induce a confusional state (metabolic encephalopathy).

Cerebral neurons are dependent on cerebral blood flow (CBF) and the delivery of oxygen and glucose. Brain stores of glucose are able to provide energy for ~2 min after blood flow is interrupted, and oxygen stores last 8–10 s after the cessation of blood flow. Simultaneous hypoxia and ischemia exhaust glucose more rapidly. The EEG rhythm in these circumstances becomes diffusely slowed, typical of metabolic encephalopathies, and as substrate delivery worsens, eventually brain electrical activity ceases.

Unlike systemic hypoxia-ischemia, which first causes a metabolic encephalopathy due to reduced energy substrate but ultimately causes neuronal destruction, most metabolic disorders such as hypoglycemia, hyponatremia, hyperosmolarity, hypercapnia, hypercalcemia, and hepatic and renal failure cause no or only minor neuropathologic changes in the brain. The reversible effects of these conditions are not fully understood but may result from impaired energy supplies, changes in ion fluxes across neuronal membranes, and neurotransmitter abnormalities. In hepatic encephalopathy (HE), high ammonia concentrations have been proposed to lead to increased synthesis of glutamine in astrocytes and osmotic swelling of the cells, mitochondrial energy failure, production of reactive nitrogen and oxygen species, increases in the inhibitory neurotransmitter γ-aminobutyric acid (GABA), and synthesis of putative "false" neurotransmitters, none of which has provided a fully adequate explanation for confusion and stupor. Over time, development of a diffuse astrocytosis is typical of chronic HE. Which, if any, of these is mainly responsible for hepatic coma is not known.

The mechanism of the encephalopathy of renal failure is also uncertain and likely to be multifactorial; unlike ammonia, urea does not produce central nervous system (CNS) depression when infused into normal persons. Contributors to uremic encephalopathy may include accumulation of neurotoxic substances such as creatinine, guanidine, and related compounds; depletion of catecholamines; altered glutamate and GABA tone; increases in brain calcium; and inflammation with disruption of the blood-brain barrier.

Coma and seizures are common accompaniments of large shifts in sodium and water balance in the brain. These changes in osmolarity arise from systemic medical disorders, including diabetic ketoacidosis, the nonketotic hyperosmolar state, and hyponatremia from any cause (e.g., water intoxication, excessive secretion of antidiuretic hormone, or atrial natriuretic peptides). Sodium levels <125 mmol/L, especially if achieved quickly, induce confusion, and levels <119 mmol/L, when arrived at acutely, are typically associated with coma and convulsions. In hyperosmolar coma, the serum osmolarity is generally >350 mosmol/L. Hypercapnia depresses the level of consciousness in proportion to the rise in carbon dioxide (CO_2) in the blood. In all of these metabolic encephalopathies, the degree of neurologic change depends on the rapidity with which the serum changes occur. The pathophysiology of other metabolic encephalopathies such as those due to hypercalcemia, hypothyroidism, vitamin B_{12} deficiency, and hypothermia are incompletely understood but must reflect derangements of synaptic function, cellular biochemistry, membrane function, or neurotransmitters.

Coma due to drugs and toxins is typically reversible and leaves no residual damage provided there has not been hypoxia or severe hypotension. Many drugs and toxins are capable of depressing nervous system function. Some produce coma by affecting both the RAS and the cerebral cortex. The combination of cortical and brainstem signs, which occurs occasionally in certain drug overdoses, may lead to an incorrect diagnosis of structural brainstem disease. Overdose of medications that have atropinic actions produces signs such as dilated pupils, tachycardia, and dry skin; opiate overdose produces pinpoint pupils <1 mm in diameter. Some drug intoxications, typified by barbiturates, can mimic all of the signs of brain death; thus, toxic etiologies must be excluded prior to making a diagnosis of brain death.

Epileptic Coma Generalized electrical seizures are associated with coma, even in the absence of motor convulsions (a condition termed "nonconvulsive status epilepticus"). As a result, EEG monitoring is often used in the evaluation of unexplained coma to exclude this potentially treatable etiology. The self-limited coma that follows a seizure, the postictal state, may be due to exhaustion of energy reserves or effects of locally toxic molecules that are the by-product of seizures. The postictal state produces continuous, generalized slowing of the background EEG activity similar to that of metabolic encephalopathies. It typically lasts for a few minutes but in some cases can be prolonged for hours or, even rarely, for days.

Coma Due to Widespread Structural Damage to the Cerebral Hemispheres This category, comprising several unrelated disorders, results from extensive bilateral structural cerebral damage. The clinical appearance simulates a metabolic encephalopathy. Hypoxia-ischemia is perhaps the best characterized form of this type of injury, in which it is not possible initially to distinguish the acute reversible effects of oxygen deprivation of the brain from the subsequent effects of neuronal damage. Similar cerebral damage may be produced by disorders that occlude widespread small blood vessels throughout the brain; examples include thrombotic thrombocytopenic purpura, hyperviscosity, and cerebral malaria. Diffuse white matter damage from cranial trauma, the delayed effects of some opioid intoxications, or inflammatory demyelinating diseases can cause a similar coma syndrome.

APPROACH TO THE PATIENT

Coma

Acute respiratory and cardiovascular problems should be attended to prior to neurologic assessment. In most instances, a complete medical evaluation, except for vital signs, fundoscopy, systemic trauma survey, and examination for nuchal rigidity, may be deferred until the neurologic evaluation has established the severity and nature of coma. **A video examination of the comatose patient is shown in Chap. V4. The approach to the patient with coma from cranial trauma is discussed in Chap. 454.**

HISTORY

The cause of coma may be immediately evident as in cases of trauma, cardiac arrest, or observed drug ingestion. In the remainder, certain points are useful: (1) the circumstances and rapidity with which neurologic symptoms developed; (2) antecedent symptoms (confusion, weakness, headache, fever, seizures, dizziness, double vision, or vomiting); (3) the use of medications, drugs, or alcohol; and (4) chronic liver, kidney, lung, heart, or other medical disease. Direct interrogation of family, observers, and emergency medical technicians on the scene, in person or by telephone, is an important part of the evaluation when possible.

GENERAL PHYSICAL EXAMINATION

Signs of head trauma raise the possibility of coexisting spinal cord injury, and in such cases, immobilization of the cervical spine is essential to prevent further injury. Fever suggests a systemic infection, bacterial meningitis, encephalitis, heat stroke, neuroleptic malignant syndrome, malignant hyperthermia due to anesthetics, or anticholinergic drug intoxication. Only rarely is fever attributable to a lesion that has disturbed hypothalamic temperature-regulating centers ("central fever"), and this diagnosis should only be considered after an exhaustive search for other causes fails to reveal an alternative explanation. A slight elevation in temperature may follow vigorous convulsions. Hypothermia is observed with exposure to cold ambient temperature, drowning, alcohol, barbiturate, sedative, or phenothiazine intoxication; hypoglycemia; peripheral circulatory failure; or extreme hypothyroidism. Hypothermia itself causes coma when the temperature is <31°C (87.8°F) regardless of the underlying etiology; less dramatically, low body temperatures can also cause coma in some instances. Tachypnea may indicate systemic acidosis or pneumonia. Aberrant respiratory patterns that reflect brainstem disorders are discussed below. Marked hypertension suggests hypertensive encephalopathy, cerebral hemorrhage, large cerebral infarction, or head injury. Hypotension is characteristic of coma from alcohol or barbiturate intoxication, internal hemorrhage or myocardial infarction causing poor delivery of blood to the brain, sepsis, profound hypothyroidism, or Addisonian crisis. The funduscopic examination can detect increased intracranial pressure (ICP) (papilledema), subarachnoid hemorrhage (subhyaloid hemorrhages), and hypertensive encephalopathy (exudates, hemorrhages, vessel-crossing changes, papilledema). Cutaneous petechiae suggest thrombotic thrombocytopenic purpura, meningococcemia, or a bleeding diathesis associated with an intracerebral hemorrhage. Cyanosis and reddish or anemic skin coloration are other indications of an underlying systemic disease or carbon monoxide as responsible for the coma.

NEUROLOGIC EXAMINATION

The patient is ideally first observed without intervention by the examiner. Spontaneously moving about the bed, reaching up toward the face, crossing legs, yawning, swallowing, coughing, and moaning reflect a drowsy state that is close to normal awakeness. Lack of restless movements on one side or an externally rotated leg suggests hemiplegia (or hip fracture). Subtle, intermittent twitching movements of a foot, finger, or facial muscle may be the only sign of seizures. Multifocal myoclonus usually indicates a metabolic disorder, particularly uremia, hypoxemia, drug intoxication, or rarely a prion disease **(Chap. 449)**. In a drowsy and confused patient, bilateral asterixis is a sign of metabolic encephalopathy or drug intoxication.

Decorticate rigidity and decerebrate rigidity, or "posturing," describe stereotyped arm and leg movements occurring spontaneously or elicited by sensory stimulation. Flexion of the elbows and wrists and supination of the arm (decorticate posturing) classically suggest bilateral damage rostral to the midbrain, whereas extension of the elbows and wrists with pronation (decerebrate posturing) indicates damage to motor tracts caudal to the midbrain. However, these localizations have been adapted from animal work and cannot be applied with precision to coma in humans. In fact, acute and widespread cerebral disorders of any type, regardless of location, frequently cause limb extension.

LEVEL OF AROUSAL

A sequence of increasingly intense stimuli is first used to determine the threshold for arousal and the motor response of each side of the body. The results of testing may vary from minute to minute, and serial examinations are useful. Tickling the nostrils with a cotton wisp is a moderate stimulus to arousal—all but deeply stuporous and comatose patients will move the head away and arouse to some degree. An even greater degree of responsiveness is present if the patient uses the hand to remove an offending stimulus. Pressure on bony prominences and pinprick stimulation, when necessary, are humane forms of noxious stimuli; pinching the skin causes ecchymoses and is generally not performed but may be useful in eliciting abduction withdrawal movements of the limbs. Posturing in response to noxious stimuli indicates severe damage to the corticospinal system, whereas abduction-avoidance movement of a limb is usually purposeful and denotes an intact corticospinal system. Posturing may also be unilateral and coexist with purposeful limb movements, reflecting incomplete damage to the motor system.

BRAINSTEM REFLEXES

Assessment of brainstem function allows localization of the lesion in coma **(Fig. 30-3)**. Patients with preserved brainstem reflexes typically have a bihemispheric localization to coma, including toxic or drug intoxication, whereas patients with abnormal brainstem reflexes either have a lesion in the brainstem or a herniation syndrome from a cerebral mass lesion impacting the brainstem secondarily. The most important brainstem reflexes are pupillary size and reaction to light, spontaneous and elicited eye movements, corneal responses, and the respiratory pattern.

Pupillary Signs Pupillary reactions are examined with a bright, diffuse light. Reactive and round pupils of midsize (2.5–5 mm) essentially exclude upper midbrain damage, either primary or secondary to compression from herniation. A response to light may be difficult to appreciate in pupils <2 mm in diameter, and bright room lighting may mute pupillary reactivity. One enlarged (>6 mm) and poorly reactive pupil signifies compression of the third nerve from the effects of a cerebral mass above. Enlargement of the pupil contralateral to a hemispheral mass may occur but is infrequent. An oval and slightly eccentric pupil is a transitional sign that accompanies early midbrain–third nerve compression. The most extreme pupillary sign, bilaterally dilated and unreactive pupils, indicates severe midbrain damage, usually from compression by a supratentorial mass. Ingestion of drugs with anticholinergic activity, the use of mydriatic eye drops, nebulizer treatments that inadvertently spray into the eye, and direct ocular trauma are other causes of pupillary enlargement.

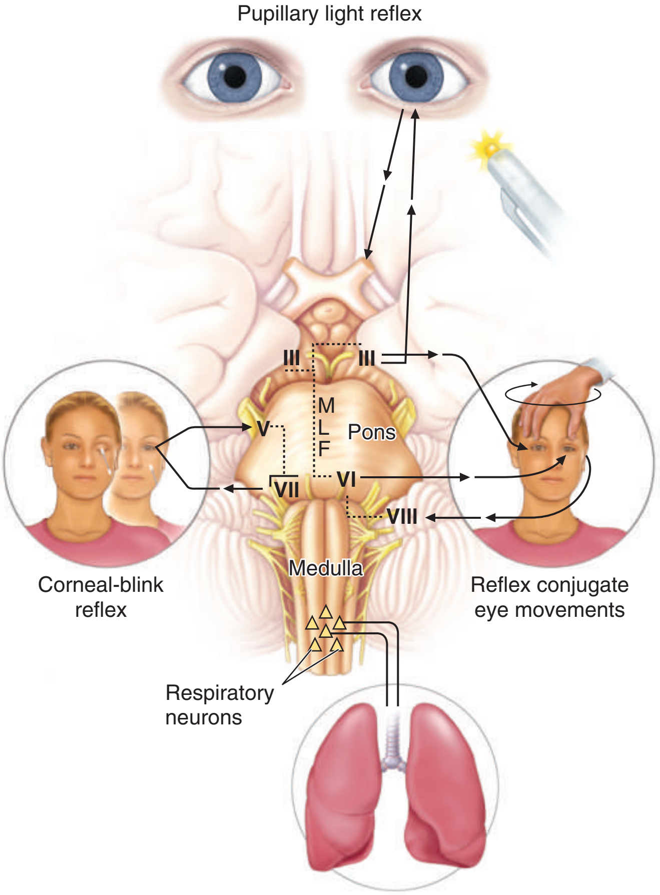

FIGURE 30-3 Examination of brainstem reflexes in coma. Midbrain and third nerve function are tested by pupillary reaction to light, pontine function by spontaneous and reflex eye movements and corneal responses, and medullary function by respiratory and pharyngeal responses. Reflex conjugate, horizontal eye movements are dependent on the medial longitudinal fasciculus (MLF) interconnecting the sixth and contralateral third nerve nuclei. Head rotation (oculocephalic reflex) or caloric stimulation of the labyrinths (oculovestibular reflex) elicits contraversive eye movements (for details, see text).

Reactive and bilaterally small (1–2.5 mm) but not pinpoint pupils are seen in metabolic encephalopathies or in deep bilateral hemispheral lesions such as hydrocephalus or thalamic hemorrhage. Even smaller reactive pupils (<1 mm) characterize opioid overdoses but also occur with extensive pontine hemorrhage. The response to naloxone and the presence of reflex eye movements (see below) assist in distinguishing between these. Unilateral miosis in coma has been attributed to dysfunction of sympathetic efferents originating in the posterior hypothalamus and descending in the tegmentum of the brainstem to the cervical cord. It is an occasional finding in patients with a large cerebral hemorrhage that affects the thalamus.

Ocular Movements The eyes are first observed by elevating the lids and observing the resting position and spontaneous movements of the globes. Horizontal divergence of the eyes at rest is normal in drowsiness. As coma deepens, the ocular axes may become parallel again.

Spontaneous eye movements in coma often take the form of conjugate horizontal roving. This finding alone exonerates extensive damage in the midbrain and pons and has the same significance as normal reflex eye movements (see below). Conjugate horizontal ocular deviation to one side indicates damage to the frontal lobe on the same side or less commonly the pons on the opposite side. This phenomenon is summarized by the following maxim: *The eyes look toward a hemispheral lesion and away from a brainstem lesion*. Seizures involving the frontal lobe drive the eyes to the opposite side, simulating a pontine destructive lesion. The eyes may occasionally turn paradoxically away from the side of a deep hemispheral or thalamic lesion ("wrong-way eyes"). The eyes more often turn down and inward with thalamic and upper midbrain lesions, typically thalamic hemorrhage. "Ocular bobbing" describes brisk downward and slow upward movements of the eyes associated with loss of horizontal eye movements and is diagnostic of bilateral pontine damage, usually from thrombosis of the basilar artery. "Ocular dipping" is a slower, arrhythmic downward movement followed by a faster upward movement in patients with normal reflex horizontal gaze; it usually indicates diffuse cortical anoxic damage.

The oculocephalic reflex, elicited after establishing there is no cervical spinal cord injury, is examined by moving the head from side to side or vertically and observing eye movements in the direction opposite to the head movement. The reflex depends on the integrity of the ocular motor nuclei and their interconnecting tracts that extend from the midbrain to the pons and medulla (Fig. 30-3). The movements, called somewhat inaccurately "doll's eyes," are normally suppressed in the awake patient with intact frontal lobes. The ability to elicit an oculocephalic reflex therefore reflects both reduced cortical influence on the brainstem and intact brainstem pathways. The opposite, an absence of reflex eye movements, usually signifies damage within the brainstem but can result from overdoses of certain drugs. In this circumstance, normal pupillary size and light reaction distinguishes most drug-induced comas from structural brainstem damage.

Thermal, or "caloric," stimulation of the vestibular apparatus (oculovestibular response) provides a more intense stimulus for the oculocephalic reflex but provides essentially the same information. The test is performed by irrigating the external auditory canal with cold water in order to induce convection currents in the labyrinths. After a brief latency, the result is tonic deviation of both eyes to the side of cold-water irrigation. In comatose patients, nystagmus in the opposite direction may not occur. The acronym "COWS" has been used to remind generations of medical students of the direction of nystagmus—cold water opposite, warm water same—but since nystagmus is often absent in the opposite direction due to frontal lobe dysfunction in coma, this mnemonic does not often hold true in this situation.

The corneal reflex, elicited by touching the cornea with a wisp of cotton and observing bilateral lid closure, depends on the integrity of pontine pathways between the fifth (afferent) and both seventh (efferent) cranial nerves; it is a useful test of pontine function. CNS-depressant drugs diminish or eliminate the corneal responses soon after reflex eye movements are paralyzed but before the pupils become unreactive to light. The corneal response may be lost for a time on the side of an acute hemiplegia.

Respiratory Patterns These are of less localizing value in comparison to other brainstem signs. Shallow, slow, but regular breathing suggests metabolic or drug-induced depression of the medullary respiratory centers. Cheyne-Stokes respiration in its typical cyclic form, ending with a brief apneic period, signifies bihemispheric damage or metabolic suppression and commonly accompanies light coma. Rapid, deep (Kussmaul) breathing usually implies metabolic acidosis but may also occur with pontomesencephalic lesions. Agonal gasps are the result of lower brainstem (medullary) damage and are recognized as the terminal respiratory pattern of severe brain damage. Other cyclic breathing patterns have been described but are of lesser significance.

LABORATORY STUDIES AND IMAGING

The studies that are most useful in the diagnosis of coma are chemical-toxicologic analyses of blood and urine, cranial CT or MRI, EEG, and cerebrospinal fluid (CSF) examination. Arterial blood gas analysis is helpful in patients with lung disease and acid-base disorders. The metabolic aberrations commonly encountered in clinical practice are usually revealed by measurement of electrolytes, glucose, calcium, magnesium, osmolarity, and renal (blood urea nitrogen) and hepatic (including NH_3) function. Toxicologic analysis may be necessary in cases of acute coma when the diagnosis is not immediately clear. However, the presence of exogenous drugs or toxins, especially alcohol, does not exclude the possibility that other factors, including head trauma, are contributing to the clinical state. An ethanol level of 43 mmol/L (0.2 g/dL) in nonhabituated patients generally causes impaired mental activity; a level of >65 mmol/L (0.3 g/dL) is associated with stupor. The development of tolerance may allow some individuals who chronically ingest alcohol to remain awake at levels >87 mmol/L (0.4 g/dL).

The availability of cranial CT and MRI has focused attention on causes of coma that are detectable by imaging (e.g., hemorrhage, tumor, or hydrocephalus). Resorting primarily to this approach, although at times expedient, is imprudent because most cases of coma (and confusion) are metabolic or toxic in origin. Furthermore, a normal CT scan does not exclude an anatomic lesion as the cause of coma; for example, early bilateral hemisphere infarction, acute brainstem infarction, encephalitis, meningitis, mechanical shearing of axons as a result of closed head trauma, sagittal sinus thrombosis, hypoxic injury, and subdural hematoma isodense to adjacent brain are some of the disorders that may not be detected. Sometimes imaging results can be misleading such as when small subdural hematomas or old strokes are found, but the patient's coma is due to intoxication. Additional imaging with CT angiography or MRI can be obtained if acute posterior circulation stroke is considered.

The EEG **(Chap. 436)** provides clues in metabolic or drug-induced states but is rarely diagnostic in these disorders. However, it is the essential test to reveal coma due to nonconvulsive seizures and shows fairly characteristic patterns in some conditions including herpesvirus encephalitis and prion disease. The EEG may be further helpful in disclosing generalized slowing of the background activity, a reflection of the severity of an encephalopathy. Predominant high-voltage slowing (δ or triphasic waves) in the frontal regions is typical of metabolic coma, as from hepatic failure, and widespread fast (β) activity implicates overdose with sedative drugs (e.g., benzodiazepines). A special pattern of "alpha coma," defined by widespread, variable, 8- to 12-Hz activity, superficially resembles the normal α rhythm of waking but, unlike normal α activity, is not altered by environmental stimuli. Alpha coma results from pontine or diffuse cortical damage and is associated with a poor prognosis. A unique EEG pattern in adults of "extreme delta brush" is characteristic of the anti–*N*-methyl-D-aspartate (NMDA) receptor form of autoimmune encephalitis. Normal α activity on the EEG, which is suppressed by stimulating the patient, also alerts the clinician to the locked-in syndrome, functional disorder, or catatonia.

Lumbar puncture should be performed if no cause is readily apparent, as examination of the CSF remains indispensable in the diagnosis of various forms of meningitis and encephalitis. An imaging study is generally performed prior to lumbar puncture to exclude a large intracranial mass lesion, which could lead to herniation with lumbar puncture. Blood cultures and administration of antimicrobials should precede the imaging study if infectious meningitis is suspected **(Chap. 143)**.

DIFFERENTIAL DIAGNOSIS OF COMA

(Table 30-1) The causes of coma can be divided into three broad categories: those without focal neurologic signs (e.g., metabolic and toxic encephalopathies); those with prominent focal signs (e.g., stroke, cerebral hemorrhage); and meningitis syndromes, characterized by fever or stiff neck and an excess of cells in the spinal fluid (e.g., bacterial meningitis, subarachnoid hemorrhage, infectious or

TABLE 30-1 Differential Diagnosis of Coma

1. Diseases that cause no focal brainstem or lateralizing neurologic signs (CT scan is often normal)
    a. Intoxications: alcohol, sedative drugs, opiates, etc.
    b. Metabolic disturbances: anoxia, hyponatremia, hypernatremia, hypercalcemia, diabetic acidosis, nonketotic hyperosmolar hyperglycemia, hypoglycemia, uremia, hepatic coma, hypercarbia, Addisonian crisis, hypo- and hyperthyroid states, profound nutritional deficiency
    c. Severe systemic infections: pneumonia, septicemia, typhoid fever, malaria, Waterhouse-Friderichsen syndrome
    d. Shock from any cause
    e. Status epilepticus, nonconvulsive status epilepticus, postictal states
    f. Hyperperfusion syndromes including hypertensive encephalopathy, eclampsia, posterior reversible encephalopathy syndrome (PRES)
    g. Severe hyperthermia, hypothermia
    h. Concussion
    i. Acute hydrocephalus
2. Diseases that cause focal brainstem or lateralizing cerebral signs (CT scan is typically abnormal)
    a. Hemispheral hemorrhage (basal ganglionic, thalamic) or infarction (large middle cerebral artery territory) with secondary brainstem compression
    b. Brainstem infarction due to basilar artery thrombosis or embolism
    c. Brain abscess, subdural empyema
    d. Epidural and subdural hemorrhage, brain contusion
    e. Brain tumor with surrounding edema
    f. Cerebellar and pontine hemorrhage and infarction
    g. Widespread traumatic brain injury
    h. Metabolic coma (see above) in the setting of preexisting focal damage
3. Diseases that cause meningeal irritation with or without fever, and with an excess of white blood cells or red blood cells in the CSF
    a. Subarachnoid hemorrhage from ruptured aneurysm, arteriovenous malformation, trauma
    b. Infectious meningitis and meningoencephalitis
    c. Paraneoplastic and autoimmune encephalitis
    d. Carcinomatous and lymphomatous meningitis

autoimmune/paraneoplastic encephalitis). Causes of sudden coma include drug ingestion, cerebral hemorrhage, trauma, cardiac arrest, epilepsy, and basilar artery occlusion. Coma that appears subacutely is usually related to a preexisting medical or neurologic problem or, less often, to secondary brain swelling surrounding a mass such as tumor or cerebral infarction.

The diagnosis of coma due to cerebrovascular disease can be difficult but is generally disclosed by cerebral imaging **(Chap. 437)**. The most common diseases in this category are (1) basal ganglia and thalamic hemorrhage (acute but not instantaneous onset, vomiting, headache, hemiplegia, and characteristic eye signs); (2) pontine hemorrhage (sudden onset, pinpoint pupils, loss of reflex eye movements and corneal responses, ocular bobbing, posturing, and hyperventilation); (3) cerebellar hemorrhage (occipital headache, vomiting, gaze paresis, and inability to stand and walk); (4) basilar artery thrombosis (neurologic prodrome or transient ischemic attack warning spells, diplopia, dysarthria, vomiting, eye movement and corneal response abnormalities, and asymmetric limb paresis); and (5) subarachnoid hemorrhage (precipitous coma after sudden severe headache and vomiting). The most common large-vessel stroke, unilateral infarction in the territory of the middle cerebral artery, does not cause coma, but edema surrounding large infarctions may expand over several days and cause coma from mass effect.

The syndrome of acute hydrocephalus accompanies many intracranial diseases, particularly subarachnoid hemorrhage. It is characterized by headache and sometimes vomiting that may progress quickly to coma with extensor posturing of the limbs, bilateral Babinski signs, small unreactive pupils, and impaired oculocephalic movements in the vertical direction. At times, the coma may be featureless without lateralizing signs, although papilledema is often present.

■ BRAIN DEATH

Brain death is a state of irreversible cessation of all cerebral and brainstem function with preservation of cardiac activity and maintenance of respiratory and somatic function by artificial means. It is the only type of brain damage recognized as morally, ethically, and legally equivalent to death. Criteria have been advanced for the diagnosis of brain death, and it is essential to adhere to consensus standards as multiple studies have shown variability in local practice. Given the implications of the diagnosis, clinicians must be thorough and precise in determining brain death. It is advisable to delay clinical testing for at least 24 h if a cardiac arrest has caused brain death or if the inciting disease is not known. Some centers advocate a brief period of observation between two examiners' tests during which the clinical signs of brain death are sustained.

Established criteria contain two essential elements, after assuring that no confounding factors (e.g., hypothermia, drug intoxication) are present: (1) widespread cortical destruction that is reflected by deep coma and unresponsiveness to all forms of stimulation; and (2) global brainstem damage as demonstrated by absent pupillary light reaction, absent corneal reflexes, loss of oculovestibular reflexes, and destruction of the medulla, manifested by complete and irreversible apnea. Diabetes insipidus is often present but may only develop hours or days after the other clinical signs of brain death appear and is not used as a criterion. The pupils are usually midsized but may be enlarged. Loss of deep tendon reflexes is not required because the spinal cord remains functional. Occasionally, other reflexes that originate from the spine may be present and should not preclude a diagnosis of brain death.

Demonstration that apnea is due to medullary damage requires that the P_{CO_2} be high enough to stimulate respiration during a test of spontaneous breathing. Apnea testing can be done by the use of preoxygenation with 100% oxygen prior to and following removal of the ventilator. CO_2 tension increases ~0.3–0.4 kPa/min (2–3 mmHg/min) during apnea. Apnea is confirmed if no respiratory effort has been observed in the presence of a sufficiently elevated P_{CO_2}. The apnea test is usually stopped if there is cardiovascular instability and alternative means of testing can be employed.

An isoelectric EEG may be used as an optional confirmatory test for total cerebral damage. Radionuclide brain scanning, cerebral angiography, or transcranial Doppler measurements may be used to demonstrate the absence of blood flow when a confirmatory study is desired.

It is largely accepted in Western society that the ventilator can be disconnected from a brain-dead patient and that organ donation is subsequently possible. Good communication between the physician and the family is important with appropriate preparation of the family for brain death testing and diagnosis.

TREATMENT

Coma

The immediate goal in a comatose patient is prevention of further nervous system damage. Hypotension, hypoglycemia, hypercalcemia, hypoxia, hypercapnia, and hyperthermia should be corrected rapidly. Hyponatremia should be corrected slowly to avoid injury from osmotic demyelination **(Chap. 318)**. An oropharyngeal airway is adequate to keep the pharynx open in a drowsy patient who is breathing normally. Tracheal intubation is indicated if there is apnea, upper airway obstruction, hypoventilation, or emesis, or if the patient is at risk for aspiration. Mechanical ventilation is required if there is hypoventilation or a need to induce hypocapnia in order to lower ICP. The management of raised ICP is discussed in **Chap. 318**. IV access is established and naloxone and dextrose are administered if opioid overdose or hypoglycemia are possibilities; thiamine is given along with glucose to avoid provoking Wernicke's encephalopathy in malnourished patients. In cases of suspected ischemic stroke including basilar thrombosis with brainstem ischemia, IV tissue plasminogen activator or mechanical embolectomy is often used after cerebral hemorrhage has been excluded and when the patient presents within established time windows for these interventions **(Chap. 438)**. Physostigmine may awaken patients with anticholinergic-type drug overdose but should be used only with careful monitoring; many physicians believe that it should only be used to treat anticholinergic overdose–associated cardiac arrhythmias. The use of benzodiazepine antagonists offers some prospect of improvement after overdose; however, these drugs are not commonly used empirically in part due to their tendency to provoke seizures. Certain other toxic and drug-induced comas have specific treatments such as fomepizole for ethylene glycol ingestion.

Administration of hypotonic IV solutions should be monitored carefully in any serious acute brain illness because of the potential for exacerbating brain swelling. Cervical spine injuries must not be overlooked, particularly before attempting intubation or evaluation of oculocephalic responses. Fever and meningismus indicate an urgent need for examination of the CSF to diagnose meningitis. Whenever acute bacterial meningitis is suspected, antibiotics including at least vancomycin and a third-generation cephalosporin are typically administered rapidly along with dexamethasone **(see Chap. 143)**.

■ PROGNOSIS

Some patients, especially children and young adults, may have ominous early clinical findings such as abnormal brainstem reflexes and yet recover; early prognostication outside of brain death therefore is unwise. Metabolic comas have a far better prognosis than traumatic or ischemic ones. Systems for estimating prognosis in adults should be taken as approximations, and medical judgments must be tempered by factors such as age, underlying systemic disease, and general medical condition. In an attempt to collect prognostic information from large numbers of patients with head injury, the Glasgow Coma Scale was devised; it has predictive value in cases of brain trauma **(Chap. 454)**, but not in most other causes of coma. For anoxic coma, clinical signs such as the pupillary and motor responses after 1 day, 3 days, and 1 week have predictive value; however, in the setting of therapeutic hypothermia, these prediction rules are less reliable, and therefore, serial examinations and multimodal prognostication approaches are advised. For example, the absence of the cortical responses of the somatosensory evoked potentials has been shown to be a strong indicator of poor outcome following hypoxic injury, as has high elevations of serum neuron-specific enolase drawn at established intervals after anoxia.

The poor outcome of persistent vegetative and minimally conscious states has already been mentioned, but reports of a small number of patients displaying cortical activation on functional MRI in response to salient stimuli have begun to alter the perception of such individuals. In one series, about 10% of vegetative patients (mainly following traumatic brain injury) could activate their frontal or temporal lobes in response to requests by an examiner to imagine certain visuospatial tasks. Another series demonstrated that up to 15% of patients with various forms of acute brain injury and absence of behavioral responses to motor commands showed EEG activation in response to these commands. It is prudent to avoid generalizations from these findings, but the need for future studies of novel techniques to help communication and possibly recovery is needed.

■ FURTHER READING

Claasen J et al: Detection of brain activation in unresponsive patients with acute brain injury. N Engl J Med 380:2497, 2019.

Edlow JA et al: Recovery from disorders of consciousness: Mechanisms, prognosis and emerging therapies. Nat Rev Neurol 17:135, 2021.

Greer DM et al: Pediatric and adult brain death/death by neurologic criteria consensus guideline: Report of the AAN guidelines subcommittee, AAP, CNS, and SCCM. Neurology 101:1112, 2023.

Posner JB et al: *Plum and Posner's Diagnosis of Stupor and Coma*, 5th ed. New York, Oxford University Press, 2019.

Wijdicks EFM: Predicting the outcome of a comatose patient at the bedside. Pract Neurol 20:26, 2020.

31

Dementia

Gil D. Rabinovici, William W. Seeley,
Bruce L. Miller

Dementia, a syndrome with many causes, affects over 6 million people in the United States and results in a total annual health care cost in excess of $300 billion. Dementia is defined as an acquired deterioration in cognitive abilities that impairs the successful performance of activities of daily living. Episodic memory, the ability to recall events specific in time and place, is the cognitive function most commonly lost; 10% of persons age >70 years and 20–40% of individuals age >85 years have clinically identifiable memory loss. In addition to memory, dementia may erode other mental faculties, including language, visuospatial, praxis, calculation, judgment, and problem-solving abilities. Neuropsychiatric and social deficits also arise in many dementia syndromes, manifesting as depression, apathy, anxiety, hallucinations, delusions, agitation, insomnia, sleep disturbances, compulsions, or disinhibition. The clinical course may be slowly progressive, as in *Alzheimer's disease (AD)*; static, as in anoxic encephalopathy; or may fluctuate from day to day or minute to minute, as in *dementia with Lewy bodies (DLB)*. Most patients with AD, the most prevalent form of dementia, begin with episodic memory impairment, but in other dementias, such as *frontotemporal dementia (FTD)*, memory loss is not typically a presenting feature. When dementia is caused by a progressive neurodegenerative disease, it is preceded by a prodromal clinical stage called *mild cognitive impairment (MCI),* in which individuals experience cognitive decline but remain independent in most daily activities. Increasingly, a *preclinical* stage is recognized for AD and other dementing illnesses, in which brain pathology is present but clinical symptoms are not yet manifest. **Focal cerebral disorders are discussed in Chap. 32 and illustrated in a video library in Chap. V2; detailed discussions of AD can be found in Chap. 442; FTD and related disorders in Chap. 443; vascular dementia in Chap. 444; DLB in Chap. 445; Huntington's disease (HD) in Chap. 447; and prion diseases in Chap. 449.**

FUNCTIONAL ANATOMY OF THE DEMENTIAS

Dementia syndromes result from the disruption of specific large-scale neuronal networks by initially focal brain lesions, including neurodegenerative changes and vascular injury. Ultimately, the location and severity of synaptic and neuronal loss combine to produce the clinical features **(Chap. 32)**. Behavior, mood, and attention are also modulated by ascending noradrenergic, serotonergic, and dopaminergic pathways, whereas cholinergic signaling is critical for attention and memory functions. The dementias differ in the underlying molecular pathology and relative neurotransmitter deficit profiles; accordingly, accurate diagnosis guides effective therapy.

AD typically begins in the entorhinal region of the medial temporal lobe, spreads to the hippocampus and other limbic structures, moves through the basal temporal areas, and then into lateral and posterior temporal and parietal neocortex, eventually causing a more widespread degeneration. Vascular dementia is associated with focal damage in a variable patchwork of cortical and subcortical regions or white matter tracts that disconnects nodes within distributed networks. In keeping with its anatomy, AD typically presents with episodic memory loss accompanied later by aphasia, executive dysfunction, or navigational problems. In contrast, dementias that begin in frontal or subcortical regions, such as FTD or HD, are less likely to begin with memory problems and more likely to present with difficulties with judgment, mood, executive control, movement, and behavior.

Lesions of frontal-striatal[1] pathways produce specific and predictable effects on behavior. The dorsolateral prefrontal cortex has connections with a central band of the caudate nucleus. Lesions of either the caudate or dorsolateral prefrontal cortex, or their connecting white matter pathways, may result in executive dysfunction, manifesting as poor organization and planning, decreased cognitive flexibility, and impaired working memory. The lateral orbital frontal cortex connects with the ventromedial caudate, and lesions of this system cause impulsiveness, distractibility, and disinhibition. The anterior cingulate cortex and adjacent medial prefrontal cortex project to the nucleus accumbens, and interruption of this system produces apathy, poverty of speech, emotional blunting, or even akinetic mutism. All corticostriatal systems also include topographically organized projections through the globus pallidus and thalamus, and damage to these nodes can likewise reproduce the clinical syndrome associated with the corresponding cortical or striatal injuries. Lesions in nodes of the dominant hemisphere speech and language networks can present as a primary progressive aphasia, with deficits in naming, word retrieval, motor speech, grammar, and comprehension of single words or more complex phrases **(Chap. 443)**. Involvement of brainstem nuclei and cerebellar structures can further contribute to cognitive, behavioral, motor and autonomic manifestations.

■ THE CAUSES OF DEMENTIA

The single strongest risk factor for dementia is increasing age. The prevalence of disabling memory loss increases with each decade over age 50 and is usually associated with the microscopic changes of AD at autopsy. Yet some centenarians have intact memory function and no evidence of clinically significant dementia. Whether dementia is an inevitable consequence of normal human aging remains controversial, although the prevalence increases with every decade of life.

The many causes of dementia are listed in **Table 31-1**. The frequency of each condition depends on the age group under study, access of the

---

[1] The striatum comprises the caudate/putamen.

TABLE 31-1 Differential Diagnosis of Dementia

Most Common Causes of Dementia

- Alzheimer's disease
- Vascular dementia
  - Multi-infarct
  - Diffuse white matter disease (Binswanger's)
- Alcoholism[a]
- PDD/LBD spectrum
- Drug/medication intoxication[a]
- Limbic-predominant age-related TDP-43 encephalopathy

Less Common Causes of Dementia

- Vitamin deficiencies
  - Thiamine (B_1): Wernicke's encephalopathy[a]
  - B_{12} (subacute combined degeneration)[a]
  - Nicotinic acid (pellagra)[a]
- Endocrine and other organ failure
  - Hypothyroidism[a]
  - Adrenal insufficiency and Cushing's syndrome[a]
  - Hypo- and hyperparathyroidism[a]
  - Renal failure[a]
  - Liver failure[a]
  - Pulmonary failure[a]
- Chronic infections
  - HIV
  - Neurosyphilis[a]
  - Papovavirus (JC virus) (progressive multifocal leukoencephalopathy)
  - Tuberculosis, fungal, and protozoal[a]
  - Whipple's disease[a]
- Head trauma and diffuse brain damage
  - Chronic traumatic encephalopathy
  - Chronic subdural hematoma[a]
  - Postanoxia
  - Postencephalitis
  - Normal-pressure hydrocephalus[a]
- Intracranial hypotension
- Neoplastic
  - Primary brain tumor[a]
  - Metastatic brain tumor[a]
  - Paraneoplastic/autoimmune limbic encephalitis[a]
- Toxic disorders
  - Drug, medication, and narcotic poisoning[a]
  - Heavy metal intoxication[a]
  - Organic toxins
- Psychiatric
  - Depression (pseudodementia)[a]
  - Schizophrenia[a]
  - Conversion disorder[a]
- Degenerative disorders
  - Huntington's disease
  - Multisystem atrophy
  - Hereditary ataxias (some forms)
  - Frontotemporal lobar degeneration spectrum
  - Multiple sclerosis
  - Adult Down's syndrome with Alzheimer's disease
  - ALS-parkinsonism-dementia complex of Guam
  - Prion (Creutzfeldt-Jakob and Gerstmann-Sträussler-Scheinker diseases)
- Miscellaneous
  - Sarcoidosis[a]
  - Vasculitis[a]
  - CADASIL, etc.
  - Acute intermittent porphyria[a]
  - Recurrent nonconvulsive seizures[a]
- Additional conditions in children or adolescents
  - Pantothenate kinase–associated neurodegeneration
  - Subacute sclerosing panencephalitis
  - Metabolic disorders (e.g., Wilson's and Leigh's diseases, leukodystrophies, lipid storage diseases, mitochondrial mutations)

[a] Potentially reversible dementia.

Abbreviations: ALS, amyotrophic lateral sclerosis; CADASIL, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy; LBD, Lewy body disease; PDD, Parkinson's disease dementia.

group to medical care, country of origin, and perhaps racial or ethnic background. AD is the most common cause of dementia in Western countries, accounting for more than half of all patients. Vascular disease is the second most frequent cause for dementia and is particularly common in elderly patients or populations with limited access to medical care, where vascular risk factors are undertreated. Often, vascular brain injury is mixed with neurodegenerative disorders, particularly AD, making it difficult, even for the neuropathologist, to estimate the contribution of cerebrovascular disease to the cognitive disorder in an individual patient. Dementias associated with Parkinson's disease (PD) are common and may develop years after onset of a parkinsonian disorder, as seen with PD-related dementia (PDD), or can occur concurrently with or preceding the motor syndrome, as in DLB. A recently characterized dementia is limbic-predominant aging-related TDP-43 encephalopathy (LATE), which is common after age 70 and has been linked to declining episodic memory function. Chronic traumatic encephalopathy (CTE), a unique disease found in individuals with high exposure to repetitive head impacts (e.g., professional athletes in collision or fighting sports, military veterans exposed to multiple blasts), presents with changes in cognition, mood, behavior, or motor function. Mixed pathology is common, especially in older individuals. In patients under the age of 65, FTD rivals AD as the most common cause of dementia. Chronic intoxications, including those resulting from alcohol and prescription drugs, are an important and often treatable cause of dementia. Other disorders listed in Table 31-1 are uncommon but important because many are reversible. The classification of dementing illnesses into reversible and irreversible disorders is a useful approach to differential diagnosis.

In a study of 1000 persons attending a memory disorders clinic, 19% had a potentially reversible cause of the cognitive impairment and 23% had a potentially reversible concomitant condition that may have contributed to the patient's impairment. The three most common potentially reversible diagnoses were depression, normal pressure hydrocephalus (NPH), and alcohol dependence; medication side effects are also common and should be considered in every patient (Table 31-1).

The term *rapidly progressive dementia (RPD)* is applied to illnesses that progress from initial symptom onset to dementia within a year or less; confusional states related to toxic/metabolic conditions are excluded. Although the prion proteinopathy Creutzfeldt-Jakob disease (CJD) **(Chap. 449)** is the classic cause of a rapidly progressive dementia, especially when associated with myoclonus, more often cases of RPD are due to AD or another neurodegenerative disorder, or to an autoimmune encephalitis **(Chap. 99)**.

Subtle cumulative decline in episodic memory is a common part of aging. This frustrating experience, often the source of jokes and humor, has historically been referred to as *benign forgetfulness of the elderly*. *Benign* means that it is not so progressive or serious that it impairs successful and productive daily functioning, although the distinction between benign and significant memory loss can be subtle. At age 85, the average person is able to learn and recall approximately one-half of the items (e.g., words on a list) that they could at age 18. The term *subjective cognitive decline* is used to refer to individuals who experience a subjective decline from their cognitive baseline but perform within normal limits for their age and educational attainment on formal neuropsychological testing. As noted earlier, MCI is defined as a decline in cognition that is confirmed on objective cognitive testing but does not disrupt normal daily activities. MCI can be further subcategorized based on the presenting complaints and deficits (e.g., amnestic MCI, dysexecutive MCI). Factors that predict progression from MCI to an AD dementia include a prominent memory deficit, family history of dementia, presence of an apolipoprotein ε4 (Apo ε4) allele, small hippocampal volumes on brain imaging, and positive AD biofluid or imaging biomarkers (see below). The term *mild behavioral impairment (MBI)* refers to the emergence of sustained and impactful neuropsychiatric symptoms in older adults (e.g., apathy, emotional dysregulation, impulse control, social inappropriateness, hallucinations, or delusions). Like its cognitive counterpart (MCI), MBI can reflect a neuropsychiatric prodrome to a neurodegenerative dementia.

The major degenerative dementias include AD, DLB, LATE, FTD and related disorders, HD, and prion diseases, including CJD. These disorders are all associated with the abnormal aggregation of a specific protein: Aβ and tau in AD; α-synuclein in DLB; TAR DNA-binding protein of 43 kDa (TDP-43) in LATE; tau, TDP-43, or the FET family of proteins (*f*used in *s*arcoma [FUS], Ewing sarcoma [EWS], and TBP-associated factor 15 [TAF15]) in FTD; huntingtin in HD; and misfolded prion protein (PrP^{sc}) in CJD **(Table 31-2)**.

The risk of developing dementia in late life is associated with numerous exposures that can happen across the lifespan. Modifiable risk factors based on large-scale epidemiologic studies include low education, hearing loss, social isolation, traumatic brain injury, hypertension, diabetes mellitus, obesity, heavy alcohol use, smoking, depression, physical inactivity, and air pollution exposure. Improved

TABLE 31-2 The Molecular Basis for Degenerative Dementia

DEMENTIA	MOLECULAR BASIS	CAUSAL GENES (CHROMOSOME)	SUSCEPTIBILITY GENES	PATHOLOGIC FINDINGS
AD	Aβ/tau	*APP* (21), *PS-1* (14), *PS-2* (1) (<2% carry these mutations, most often in *PS-1*)	*Apo ε4* (19)	Amyloid plaques, neurofibrillary tangles, and neuropil threads
DLB	α-Synuclein	Very rare *SNCA* (4)	Unknown	α-Synuclein neuronal inclusions (Lewy bodies)
LATE	TDP-43	None identified	*TMEM106B, GRN*	TDP-43 neuronal "inclusion bodies" and neurites in neurons and glia, with or without hippocampal sclerosis
FTD	Tau	*MAPT* exon and intron mutations (17) (~10% of familial cases)	H1 *MAPT* haplotype	Tau neuronal and glial inclusions varying in morphology and distribution
	TDP-43	*GRN* (10% of familial cases), *C9ORF72* (20–30% of familial cases), rare *VCP*, very rare *TARDBP, TBK1, TIA1*		TDP-43 neuronal and glial inclusions varying in morphology and distribution
	FUS	Very rare *FUS*		FUS neuronal and glial inclusions varying in morphology and distribution
CJD	PrP^{SC}	*PRNP* (20) (up to 15% of patients carry these dominant mutations)	Codon 129 homozygosity for methionine or valine	PrP^{SC} deposition, panlaminar spongiosis

Abbreviations: AD, Alzheimer's disease; CJD, Creutzfeldt-Jakob disease; DLB, dementia with Lewy bodies; FTD, frontotemporal dementia; LATE, limbic-predominant age-related TDP-43 encephalopathy.

management of mid-life cardiovascular risk factors has been credited with a decreasing incidence of dementia noted in North America and western European countries.

APPROACH TO THE PATIENT

Dementias

Three major issues should be kept at the forefront: (1) What is the clinical diagnosis? (2) What component of the dementia syndrome is treatable or reversible? (3) Can the physician help to alleviate the burden on caregivers? A broad overview of the approach to dementia is shown in **Table 31-3**. The major degenerative dementias can usually be distinguished by the initial symptoms; neuropsychological, neuropsychiatric, and neurologic findings; neuroimaging features; and other biomarkers **(Table 31-4)**.

HISTORY

The history should concentrate on the onset, duration, and tempo of progression. An acute or subacute onset of confusion may be due to delirium **(Chap. 29)** and should trigger a search for intoxication, infection, or metabolic derangement. An elderly person with slowly progressive memory loss over several years is likely to suffer from AD and/or LATE. Nearly 75% of patients with AD begin with memory symptoms, but other early symptoms include anxiety or depression and difficulty with managing money, driving, shopping, following instructions, finding words, or navigating. Personality change, disinhibition, and weight gain or compulsive eating suggest FTD, not AD. FTD is also suggested by prominent apathy, compulsivity, loss of empathy for others, or progressive loss of speech fluency or single-word comprehension with relative sparing of memory and visuospatial abilities. The diagnosis of DLB is suggested by early visual hallucinations; parkinsonism; proneness to delirium or sensitivity to psychoactive medications; rapid eye movement (REM) behavior disorder (RBD; dramatic, sometimes violent, limb movements during dreaming **[Chap. 33]**); or Capgras syndrome, the delusion that a familiar person has been replaced by an impostor.

A history of stroke with irregular stepwise progression suggests vascular dementia. Vascular dementia is also commonly seen in the setting of hypertension, atrial fibrillation, peripheral vascular disease, smoking, and diabetes. In patients suffering from cerebrovascular disease, it can be difficult to determine whether the dementia is due to AD, vascular disease, or a mixture of the two because many of the risk factors for vascular dementia, including diabetes, high cholesterol, elevated homocysteine, and low exercise, are also

TABLE 31-3 Evaluation of the Patient with Dementia

ROUTINE EVALUATION	OPTIONAL FOCUSED TESTS	OCCASIONALLY HELPFUL TESTS
History Physical examination Laboratory tests Thyroid function (TSH) Vitamin B_{12} Complete blood count Complete metabolic panel CT/MRI	Psychometric testing HIV, RPR, or VDRL Lumbar puncture PET (FDG, amyloid, tau) Chest x-ray Urine toxin screen Apolipoprotein E Blood-based AD biomarkers	EEG Parathyroid function Adrenal function Urine heavy metals RBC sedimentation rate Lab screen for autoantibodies Angiogram Brain biopsy
Diagnostic Categories		
REVERSIBLE CAUSES	**IRREVERSIBLE/DEGENERATIVE DEMENTIAS**	**PSYCHIATRIC DISORDERS**
Examples Hypothyroidism Thiamine deficiency Vitamin B_{12} deficiency Normal-pressure hydrocephalus Subdural hematoma Chronic infection Brain tumor Drug intoxication Autoimmune encephalopathy	Examples Alzheimer's Frontotemporal dementia Huntington's Dementia with Lewy bodies Vascular Leukoencephalopathies Parkinson's LATE	Depression Schizophrenia Conversion reaction
Associated Treatable Conditions		
	Depression Seizures Insomnia	Agitation Caregiver "burnout" Drug side effects

Abbreviations: CT, computed tomography; EEG, electroencephalogram; LATE, limbic-predominant age-related TDP-43 encephalopathy; MRI, magnetic resonance imaging; PET, positron emission tomography; RBC, red blood cell; RPR, rapid plasma reagin (test); TSH, thyroid-stimulating hormone; VDRL, Venereal Disease Research Laboratory (test for syphilis).

TABLE 31-4 Clinical Differentiation of the Major Dementias

DISEASE	FIRST SYMPTOM	MENTAL STATUS	NEUROPSYCHIATRY	NEUROLOGY	IMAGING
AD	Memory loss	Episodic memory loss Executive, language, and visuospatial functions variably affected	Irritability, anxiety, depression	Initially normal	Entorhinal cortex and hippocampal atrophy; posterior-predominant cortical atrophy
Vascular	Often but not always sudden; variable; apathy, falls, focal weakness	Frontal/executive, cognitive slowing; can spare memory	Apathy, delusions, anxiety	Usually motor slowing, spasticity; can be normal	Cortical and/or subcortical infarctions, confluent white matter disease
DLB	Visual hallucinations, REM sleep behavior disorder, delirium, Capgras syndrome, parkinsonism	Drawing and frontal/executive; spares memory; delirium-prone	Visual hallucinations, depression, sleep disorder, delusions	Parkinsonism	Posterior parietal atrophy; hippocampi larger than in AD
LATE	Memory loss	Episodic memory loss Mild semantic deficits	None described	Normal	Medial temporal and hippocampal atrophy, anterior predominant
FTD	Apathy; poor judgment/insight, speech/language; hyperorality	Frontal/executive and/or language; spares drawing	Apathy, disinhibition, overeating, compulsivity	May have vertical gaze palsy, axial rigidity, dystonia, alien hand, or MND	Frontal, insular, and/or temporal atrophy; usually spares posterior parietal lobe
CJD	Dementia, mood, anxiety, movement disorders	Variable, frontal/executive, focal cortical, memory	Depression, anxiety, psychosis in some	Myoclonus, rigidity, parkinsonism	Cortical ribboning and basal ganglia or thalamus hyperintensity on diffusion/FLAIR MRI

Abbreviations: AD, Alzheimer's disease; CBD, cortical basal degeneration; CJD, Creutzfeldt-Jakob disease; DLB, dementia with Lewy bodies; FLAIR, fluid-attenuated inversion recovery; FTD, frontotemporal dementia; LATE, limbic-predominant age-related TDP-43 encephalopathy; MND, motor neuron disease; MRI, magnetic resonance imaging; PSP, progressive supranuclear palsy; REM, rapid eye movement.

risk factors for AD. Moreover, many patients with a major vascular contribution to their dementia lack a history of stepwise decline.

The age at symptom onset can also aid in the differential diagnosis of dementia. AD and FTD are the most common neurodegenerative causes of "early-onset" (age at symptom onset <65) dementia. The most common causes of "late-onset" dementia (age at symptom onset >65) are AD, DLB, and vascular dementia. LATE neuropathological changes are increasingly common with older age and are found in ~20% of individuals with dementia who die at age <70 versus >50% of individuals with dementia who die at age >90. Most late-onset dementia is associated with multiple pathological entities; it is common for individuals who suffered from dementia to show three or four different pathologies at autopsy. In one large community-based autopsy cohort of individuals who presented with an amnestic dementia during life (mean age at death 89.7 years), 39% of the attributable risk for dementia was explained by AD neuropathology, 25% by cerebrovascular disease, 17% by LATE, and 12% by Lewy body disease.

Rapid progression with motor rigidity and myoclonus suggests CJD **(Chap. 449)**. Seizures may indicate strokes or neoplasm but also occur in AD, particularly early-age-of-onset AD. Gait disturbance is common in vascular dementia, PD/DLB, or NPH. A history of high-risk sexual behaviors or intravenous drug use should trigger a search for central nervous system (CNS) infection, especially HIV or syphilis. A history of recurrent head trauma could indicate chronic subdural hematoma, CTE, intracranial hypotension, or NPH. Subacute onset of severe amnesia and psychosis with mesial temporal T2/fluid-attenuated inversion recovery (FLAIR) hyperintensities on magnetic resonance imaging (MRI) should raise concern for autoimmune (paraneoplastic) encephalitis, sometimes in long-term smokers or other patients at risk for cancer. The spectrum of autoimmune etiologies producing RPD has rapidly expanded and includes antibodies targeting leucine-rich glioma-inactivated 1 (LGI1; faciobrachial dystonic seizures); contactin-associated protein-like 2 (Caspr2; insomnia, ataxia, myotonia); *N*-methyl-D-aspartate (NMDA) receptor (psychosis, insomnia, dyskinesias); and α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) receptor (limbic encephalitis with relapses), among others **(Chap. 99)**. Alcohol abuse creates risk for malnutrition and thiamine deficiency. Pernicious anemia, veganism, bowel irradiation, a remote history of gastric surgery, and chronic therapy with histamine H_2-receptor antagonists for dyspepsia or gastroesophageal reflux are causes of B_{12} deficiency. Certain occupations, such as working in a battery or chemical factory, might indicate heavy metal intoxication. Careful review of medication intake, especially for sedatives and analgesics, may raise the issue of chronic drug intoxication. An autosomal dominant family history is found in HD and in familial forms of AD, FTD, DLB, or prion disorders. A history of mood disorder, the recent death of a loved one, or depressive signs such as insomnia or weight loss raise the possibility of depression-related cognitive impairment.

PHYSICAL AND NEUROLOGIC EXAMINATION

A thorough general and neurologic examination is essential, in the setting of dementia, to look for signs of nervous system involvement and to search for clues suggesting a systemic disease that might be responsible for the cognitive disorder. Typical AD spares motor systems until later in the course. In contrast, patients with FTD often develop axial rigidity, supranuclear gaze palsy, or a motor neuron disease reminiscent of amyotrophic lateral sclerosis (ALS). In DLB, the initial symptoms may include the new onset of a parkinsonian syndrome (resting tremor, cogwheel rigidity, bradykinesia, festinating gait), but DLB often starts with visual hallucinations or cognitive impairment, and symptoms referable to the lower brainstem (RBD, gastrointestinal or autonomic problems) may arise years or even decades before parkinsonism or dementia. Corticobasal syndrome (CBS) features asymmetric akinesia and rigidity, dystonia, myoclonus, alien limb phenomena, pyramidal signs, and prefrontal deficits such as nonfluent aphasia with or without motor speech impairment, executive dysfunction, apraxia, or a behavioral disorder. Progressive supranuclear palsy (PSP) is associated with unexplained falls, axial rigidity, dysphagia, and vertical gaze deficits. CJD is suggested by the presence of diffuse rigidity, an akinetic-mute state, and prominent, often startle-sensitive myoclonus.

Hemiparesis or other focal neurologic deficits suggest vascular dementia or brain tumor. Dementia with a myelopathy and peripheral neuropathy suggests vitamin B_{12} deficiency. Peripheral neuropathy could also indicate another vitamin deficiency, heavy metal intoxication, thyroid dysfunction, Lyme disease, or vasculitis. Dry cool skin, hair loss, and bradycardia suggest hypothyroidism.

Fluctuating confusion associated with repetitive stereotyped movements may indicate ongoing limbic, temporal, or frontal seizures. In the elderly, hearing impairment or visual loss may produce confusion and disorientation misinterpreted as dementia. Profound bilateral sensorineural hearing loss in a younger patient with short stature or myopathy, however, should raise concern for a mitochondrial disorder.

COGNITIVE AND NEUROPSYCHIATRIC EXAMINATION

Brief screening tools such as the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MOCA) can be used to capture dementia and follow progression. None of these tests is highly sensitive to early-stage dementia or reliably discriminates between dementia syndromes. The MMSE is a 30-point test of cognitive function, with each correct answer being scored as 1 point. It includes tests of orientation (e.g., identify season/date/month/year/floor/hospital/town/state/country); registration (e.g., name and restate three objects); recall (e.g., remember the same three objects 5 min later); and language (e.g., name pencil and watch; repeat “no ifs, ands, or buts”; follow a three-step command; obey a written command; and write a sentence and copy a design). In most patients with MCI and some with clinically apparent AD, bedside screening tests may be normal, and a more challenging and comprehensive set of neuropsychological tests will be required. When the etiology for the dementia syndrome remains in doubt, a specially tailored evaluation should be performed that includes tasks of working and episodic memory, executive function, language, and visuospatial and perceptual abilities. In AD, the early deficits involve episodic memory, category generation (“Name as many animals as you can in 1 minute”), and visuoconstructive ability. Usually deficits in verbal or visual episodic memory are the first neuropsychological abnormalities detected, and tasks that require the patient to recall a long list of words or a series of pictures after a predetermined delay will demonstrate deficits in most patients. Patients with LATE also present with prominent deficits in episodic memory and can be identified by sparing of other cognitive domains and relatively slow progression. Patients with PDD or DLB have more severe deficits in executive and visuospatial function but do better on episodic memory tasks than patients with AD. Patients with vascular dementia often demonstrate a mixture of executive and visuospatial deficits, with prominent psychomotor slowing. In delirium, the most prominent deficits involve attention, working memory, and executive function, making the assessment of other cognitive domains challenging and often uninformative. In FTD, the earliest deficits on cognitive testing involve executive control or language (speech or naming) function, but some patients lack either finding despite profound social-emotional deficits.

A functional assessment should also be performed to help the physician determine the day-to-day impact of the disorder on the patient’s memory, community affairs, hobbies, judgment, dressing, and eating. Knowledge of the patient’s functional abilities will help the clinician and the family to organize a therapeutic approach.

Neuropsychiatric assessment is important for diagnosis, prognosis, and treatment. In the early stages of AD, mild depressive features, social withdrawal, and irritability or anxiety are the most prominent psychiatric changes, but patients often maintain core social graces into the middle or late stages, when delusions, agitation, and sleep disturbance may emerge. In FTD, dramatic personality changes with apathy, overeating, compulsions, disinhibition, and loss of empathy are early and common. DLB is associated with visual hallucinations, delusions related to person or place identity, RBD, and excessive daytime sleepiness. Dramatic fluctuations occur not only in cognition but also in arousal. Vascular dementia can present with psychiatric symptoms such as depression, anxiety, delusions, disinhibition, or apathy.

LABORATORY TESTS

The choice of laboratory tests in the evaluation of dementia is complex and should be tailored to the individual patient. The physician must take measures to avoid missing a reversible or treatable cause, yet no single treatable etiology is common; thus, a screen must use multiple tests, each of which has a low yield. Cost/benefit ratios are difficult to assess, and many laboratory screening algorithms for dementia discourage multiple tests. Nevertheless, even a test with only a 1–2% positive rate is worth undertaking if the alternative is missing a treatable cause of dementia. Table 31-3 lists most screening tests for dementia. The American Academy of Neurology recommends the routine measurement of a complete blood count; electrolytes; glucose; renal, liver, and thyroid functions; a vitamin B_{12} level; and a structural neuroimaging study (MRI or computed tomography [CT]).

Neuroimaging studies, especially MRI, help to rule out primary and metastatic neoplasms, locate areas of infarction or inflammation, detect subdural hematomas, and suggest NPH or diffuse white matter disease. They also help to establish a regional pattern of atrophy. Support for the diagnosis of AD includes hippocampal atrophy in addition to posterior-predominant cortical atrophy **(Fig. 31-1)**. Marked hippocampal and medial temporal lobe atrophy is also the MRI signature of LATE. Focal frontal, insular, and/or anterior temporal atrophy suggests FTD **(Chap. 443)**. DLB often features less prominent atrophy, with greater involvement of amygdala than hippocampus. In CJD, magnetic resonance diffusion-weighted imaging reveals restricted diffusion within the cortical ribbon and/or basal ganglia in most patients. Extensive multifocal white matter abnormalities suggest a vascular etiology **(Fig. 31-2)**. Communicating hydrocephalus with vertex effacement (crowding of dorsal convexity gyri/sulci), gaping Sylvian fissures despite minimal cortical atrophy, and additional features shown in **Fig. 31-3** suggest NPH. ^{18}F-fluorodeoxyglucose positron emission tomography (FDG-PET) scanning shows temporal-parietal or hypometabolism in AD, often with early and prominent involvement of the posterior cingulate cortex and precuneus. Conversely, patients with FTD show hypometabolism in frontal and anterior temporal cortices. The FDG signature of DLB features hypometabolism in the occipital cortex and precuneus with sparing of the posterior cingulate (“cingulate island sign”), while LATE is characterized by severe medial temporal hypometabolism with sparing of association cortices. Single-photon emission computed tomography (SPECT) demonstrates spatial patterns of hypoperfusion that mirror the FDG hypometabolic patterns described above.

Amyloid and tau PET imaging can support the diagnosis of AD by directly detecting amyloid plaques and neurofibrillary tangles, the neuropathological lesions that define the disease. There are currently three amyloid PET ligands (^{18}F-florbetapir, ^{18}F-florbetaben, ^{18}F-flutametamol) and one tau PET ligand (^{18}F-flortaucipir) approved by the U.S. Food and Drug Administration (FDA) for clinical use. Amyloid PET ligands bind to diffuse and neuritic amyloid plaques as well as to vascular amyloid deposits (cerebral amyloid angiopathy), while tau PET ligands bind to the paired helical filaments of tau characteristic of neurofibrillary tangles in AD. Current tau PET ligands do not reliably detect tau deposits in non-AD conditions. Because amyloid plaques are also commonly found in cognitively normal older persons (~25% of individuals at age 70), the main clinical value of amyloid imaging is to exclude AD as the likely cause of dementia in patients who have negative scans. In older patients presenting with a progressive amnestic disorder and hippocampal atrophy, a negative amyloid PET scan strongly suggests LATE as the underlying neuropathology. The spread of tau is more tightly linked to cognitive state **(Chap. 442)**, and thus tau PET may be more useful for “ruling in” AD, as well as for disease staging. Amyloid PET is also useful to identify candidates for novel anti-Aβ monoclonal antibodies (e.g., lecanemab, donanemab) that reduce amyloid plaque load and slow cognitive decline in patients in early clinical stages of AD. Amyloid and tau PET can also assist with prognosis, as patients who are positive on both modalities show the most rapid decline in cognition and

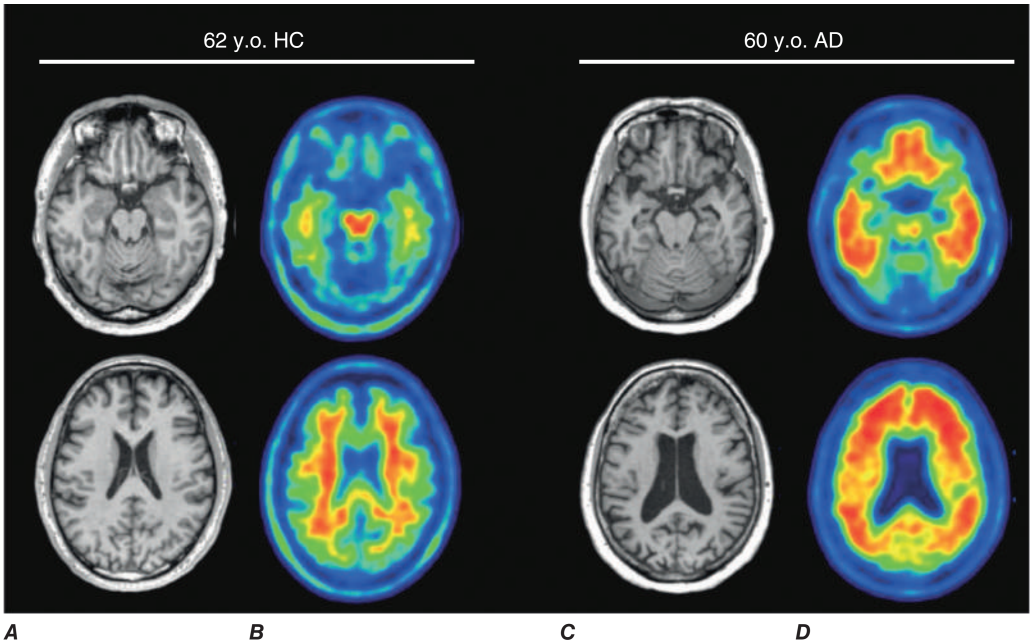

FIGURE 31-1 Alzheimer's disease (AD). Axial T1-weighted MRI and 18F-florbetapir amyloid PET images from a 62-year-old healthy control (*left*) and 60-year-old with dementia due to AD (*right panels*). Note the reduction in medial temporal volumes and prominent sulci on MRI in the patient with AD. Amyloid PET demonstrates a white matter only binding pattern in the healthy control (*negative scan*), whild the patient with AD demonstrates diffuse neocortical binding and blurring of gray/white matter contrast (*positive scan*). *(Images courtesy of Gil Rabinovici, University of California, San Francisco.)*

function. Use of amyloid and tau PET in cognitively unimpaired older adults should for now be restricted to research studies and clinical trials testing interventions aimed at reducing the risk of MCI and dementia in asymptomatic individuals who are positive for AD biomarkers.

Lumbar puncture is indicated when CNS infection or inflammation is a credible diagnostic possibility or to assess molecular biomarkers for AD in lieu of PET imaging. A cerebrospinal fluid (CSF) pattern that shows low levels of Aβ_{42} (or a low Aβ_{42}/Aβ_{40} ratio), mild to moderately elevated CSF total tau, and elevated CSF phosphorylated tau (p-Tau at residues 181 or 217) is highly suggestive of AD and sufficient for selecting patients for anti-Aβ antibody treatment. Novel fully automated CSF Aβ and tau assays perform comparably to amyloid PET, with the Aβ_{42}/Aβ_{40} or p-Tau181/Aβ_{42} ratios showing higher concordance with amyloid PET and neuropathology than any single CSF AD biomarker. Blood-based AD biomarkers, such as plasma Aβ_{42}/Aβ_{40}, p-Tau181 and p-Tau217 measured with mass spectrometry or highly sensitive immunoassays, are evolving rapidly and are likely to be approved for clinical use in the near future, greatly enhancing the scalability and cost effectiveness of biomarker testing in patients with suspected AD **(Chap. 442)**.

Recently, there has been significant progress in developing biomarkers of α-synuclein pathology, enabling a molecular diagnosis of PD or DLB. A CSF seed amplification assay for α-synuclein shows high sensitivity and specificity in clinically diagnosed PD patients and also detects pathology in a subset of individuals at risk for PD based on the presence of anosmia or RBD. α-Synuclein can also be detected with high sensitivity and specificity in living patients with PD and DLB with skin biopsies immunostained for phosphorylated α-synuclein colocalizing with nerve fiber bundles. While still in early stages of validation, these biomarkers hold

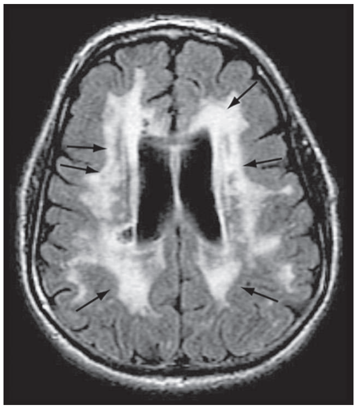

FIGURE 31-2 Diffuse white matter disease. Axial fluid-attenuated inversion recovery (FLAIR) magnetic resonance image through the lateral ventricles reveals multiple areas of hyperintensity (*arrows*) involving the periventricular white matter as well as the corona radiata and striatum. Although seen in some individuals with normal cognition, this appearance is more pronounced in patients with dementia of a vascular etiology.

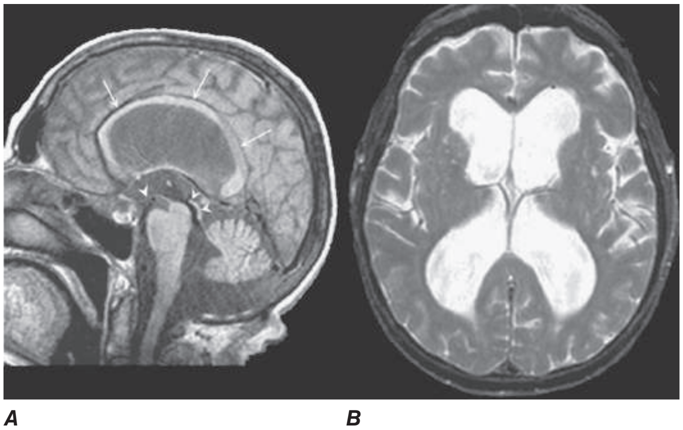

FIGURE 31-3 Normal-pressure hydrocephalus. ***A.*** Sagittal T1-weighted magnetic resonance image (MRI) demonstrates dilation of the lateral ventricle and stretching of the corpus callosum (*arrows*), depression of the floor of the third ventricle (*single arrowhead*), and enlargement of the aqueduct (*double arrowheads*). Note the diffuse dilation of the lateral, third, and fourth ventricles with a patent aqueduct, typical of communicating hydrocephalus. ***B.*** Axial T2-weighted MRIs demonstrate dilation of the lateral ventricles. This patient underwent successful ventriculoperitoneal shunting.

great promise for enhancing diagnostic accuracy and accelerating drug development for α-synuclein disorders. Active work is being done to develop PET tracers for α-synuclein and to discover both imaging and fluid-based biomarkers for other aggregated disease proteins underlying neurodegenerative dementias, such as non-AD tau and TDP-43.

Electroencephalogram (EEG) is not routinely used but can help to suggest CJD (repetitive bursts of diffuse high-amplitude sharp waves, or "periodic complexes") or an underlying nonconvulsive seizure disorder (epileptiform discharges). Brain biopsy (including meninges) is not advised except to diagnose vasculitis, potentially treatable neoplasms, or unusual infections when the diagnosis is uncertain. Systemic disorders with CNS manifestations, such as sarcoidosis, can often be confirmed through biopsy of lymph node or solid organ rather than brain. Magnetic resonance angiography should be considered when cerebral vasculitis or cerebral venous thrombosis is a possible cause of the dementia.

GLOBAL CONSIDERATIONS

Vascular dementia **(Chap. 444)** is more common in Asian countries, due to the higher prevalence of intracranial atherosclerosis. Rates of vascular dementia are also on the rise in developing countries as vascular risk factors such as hypertension, hypercholesterolemia, and diabetes mellitus become more widespread. CNS infections, HIV (and associated opportunistic infections), syphilis, cysticercosis, and tuberculosis likewise represent major contributors to dementia in the developing world. Systemic infection with SARS-CoV-2 may, in some individuals, have lasting effects on cognition due to involvement of brain microvasculature or due to immunologically mediated white matter injury (acute disseminated encephalomyelitis [ADEM]) **(Chap. 456)**. Some individuals infected with SARS-CoV-2 complain of lasting fatigue, changes in mood, and cognitive difficulties, but the long-term prognosis for SARS-CoV-2–related cognitive impairment remains unknown. Isolated populations have also contributed to our understanding of neurodegenerative dementia. Kuru, the cannibalism-associated rapidly progressive dementia seen in tribal New Guinea, played a role in the discovery of human prion disease. ALS-parkinsonism-dementia complex of Guam (or lytico-bodig disease) is a polyproteinopathy, often with tau, TDP-43, and α-synuclein aggregation. The root cause of the disease remains uncertain, but its incidence has declined sharply over the past 60 years.

TREATMENT

Dementia

The major goals of dementia management are to treat reversible causes and to provide comfort and support to the patient and their caregivers. Treatment of underlying causes includes thyroid replacement for hypothyroidism; vitamin therapy for thiamine or B_{12} deficiency or for elevated serum homocysteine; antimicrobials for opportunistic infections or antiretrovirals for HIV; ventricular shunting for NPH; or surgical, radiation, and/or chemotherapeutic treatment for CNS neoplasms. Removal of cognition-impairing drugs or medications is critical when appropriate. If the patient's cognitive complaints stem from a psychiatric disorder, vigorous treatment of the condition should be sought to eliminate the cognitive complaint or to confirm that it persists despite adequate resolution of the mood or anxiety symptoms. Patients with degenerative diseases may also be depressed or anxious, and those aspects of their condition often respond to therapy, while not necessarily improving cognition. Antidepressants, such as selective serotonin reuptake inhibitors (SSRIs) or serotonin-norepinephrine reuptake inhibitors (SNRIs) **(Chap. 463)**, which feature anxiolytic properties but few cognitive side effects, provide the mainstay of treatment when necessary. Anticonvulsants are used to control seizures, with levetiracetam and lamotrigine being the preferred agents based on their efficacy in animal models of AD and favorable cognitive side effect profiles. Furthermore, a small clinical trial found a potential cognitive benefit for levetiracetam over placebo in AD patients found to have epileptiform activity on electroencephalogram or magnetoencephalography.

Agitation, hallucinations, delusions, and confusion are difficult to treat. These behavioral problems represent major causes for nursing home placement and institutionalization. Before treating these behaviors with medications, the clinician should aggressively seek out modifiable environmental or metabolic factors. Hunger, lack of exercise, toothache, constipation, urinary tract or respiratory infection, electrolyte imbalance, and drug toxicity all represent easily correctable causes that can be remedied without psychoactive drugs. Drugs such as phenothiazines and benzodiazepines may ameliorate the behavior problems but have untoward side effects such as sedation, rigidity, dyskinesia, and occasionally paradoxical disinhibition (benzodiazepines). Despite their unfavorable side effect profile, second-generation antipsychotics such as quetiapine (starting dose, 12.5–25 mg daily) can be used for patients with agitation, aggression, and psychosis, although the risk profile for these compounds is significant, including increased mortality in patients with dementia. Brexpiprazole, an atypical antipsychotic that acts on noradrenergic, serotonergic, and dopaminergic neurotransmitter systems, recently became the first FDA-approved drug for the treatment of agitation in patients with AD dementia based on evidence of short-term efficacy and safety in a 12-week, double-blind, placebo-controlled randomized clinical trial. When patients do not respond to treatment, it is usually a mistake to advance to higher doses or to use anticholinergic drugs (e.g., diphenhydramine) or sedatives (e.g., barbiturates or benzodiazepines). It is important to recognize and treat depression; treatment can begin with a low dose of an SSRI (e.g., escitalopram, starting dose 5 mg daily, target dose 5–10 mg daily) while monitoring for efficacy and toxicity. Sometimes apathy, visual hallucinations, depression, and other psychiatric symptoms respond to the cholinesterase inhibitors, especially in DLB, obviating the need for other more toxic therapies.

Cholinesterase inhibitors are being used to treat AD (donepezil, rivastigmine, galantamine) and PDD (rivastigmine). Memantine, which acts on N-methyl-D-aspartate (NMDA) glutamate receptors, proves useful when treating some patients with moderate to severe AD; its major benefit relates to decreasing caregiver burden, most likely by decreasing resistance to dressing and grooming support. In moderate to severe AD, the combination of memantine and a cholinesterase inhibitor delayed nursing home placement in several studies, although other studies have not supported the efficacy of adding memantine to the regimen. Memantine should be used with great caution, or not at all, in patients with DLB, due to risk of worsening agitation and confusion.

In recent years, a novel class of drugs, monoclonal antibodies that target Aβ, have been approved for the treatment of MCI and mild dementia due to AD. These drugs reduce amyloid plaque burden as measured by PET, and some have been found to modestly slow clinical decline. In 2021, aducanumab became the first anti-Aβ monoclonal antibody to receive accelerated FDA approval, based on strong biomarker evidence of amyloid plaque lowering on PET. However, clinical benefits were questionable based on discordant results in two identically designed phase 3 randomized clinical trials (RCTs). In 2023, lecanemab received traditional FDA approval based on evidence of clinical efficacy in a phase 3 RCT, with treated patients showing 27% less decline over 18 months compared to placebo on the Clinical Dementia Rating–Sum of Boxes, a clinical scale that measures changes in cognition and function. A third antibody, donanemab, reported similar positive clinical results in a phase 3 RCT, and was approved by the FDA in 2024. All three drugs are infused intravenously every 2 weeks (lecanemab) or monthly (aducanumab and donanemab). Significant side effects of this class of medications include infusion reactions and amyloid-related imaging abnormalities (ARIA), which manifest as edema or sulcal effusions

(ARIA-E) or hemorrhage (ARIA-H; microbleeds, superficial siderosis, or rarely macrohemorrhages) on MRI. While most cases of ARIA are either asymptomatic (and detected by surveillance MRI) or mildly symptomatic (e.g., associated with headache, dizziness, increased confusion), ARIA can rarely cause serious adverse effects, including seizures, stroke-like episodes, and (very rarely) death. The risk of ARIA is higher in patients carrying one or two copies of APOε4, and genotyping is therefore recommended to allow fully informed counseling of patients about risks in advance of treatment. Given the modest clinical benefits and potential risks, clinicians should be judicious in identifying candidates for this new class of therapy. Appropriate patients are in early clinical stages of AD (MCI or mild dementia) based on assessment of their daily function and cognitive testing, have confirmation of amyloid pathology by PET or CSF biomarkers, and are free of comorbidities or medications that may enhance the risks of treatment. Use of thrombolytic agents is contraindicated in patients treated with anti-Aβ monoclonal antibodies based on reports of catastrophic intracerebral hemorrhages leading to death (**see Chap. 442** for additional details). Other therapies targeting the production, aggregation, and spread of Aβ, tau, and other misfolded proteins associated with dementia are under development, as are drugs targeting disease-associated neuroinflammation, metabolic changes, synaptic loss, and neurotransmitter changes.

Nondrug behavior therapy has an important place in dementia management. The primary goals are to make the patient's life comfortable, uncomplicated, and safe. Preparing lists, schedules, calendars, and labels can be helpful in the early stages. It is also useful to stress familiar routines, walks, and simple physical exercises. For many patients with dementia, memory for events is worse than their ability to carry out routine activities, and they may still be able to take part in their favorite hobbies, sports, and social activities. Patients with dementia often object to losing control over familiar tasks such as driving, cooking, and handling finances. Attempts to help may be greeted with complaints, depression, or anger. Hostile responses on the part of the caregiver are counterproductive and sometimes even harmful. Reassurance, distraction, and calm positive statements are more productive when resistance is present. Eventually, tasks such as finances and driving must be assumed by others, and the patient will conform and adjust. Safety is an important issue that includes not only driving but controlling the kitchen, bathroom, and sleeping area environments, as well as stairways. These areas need to be monitored, supervised, and made as safe as possible. A move to a retirement complex, assisted-living center, or nursing home can initially increase confusion and agitation. Repeated reassurance, reorientation, and careful introduction to the new personnel will help to smooth the process. Providing activities that are known to be enjoyable to the patient can also help.

The clinician must pay special attention to frustration and depression among family members and caregivers. Caregiver guilt and burnout are common. Family members often feel overwhelmed and helpless and may vent their frustrations on the patient, each other, and health care providers. Caregivers should be encouraged to take advantage of day-care facilities and respite services. Education and counseling about dementia are important. Local and national support groups, such as the Alzheimer's Association (*www.alz.org*), can provide considerable help.

FURTHER READING

Ayhan Y et al: Management of psychiatric symptoms in dementia. Neurol Clin 41:123, 2023.

Boyle PA et al: Attributable risk of Alzheimer's dementia attributed to age-related neuropathologies. Ann Neurol 85:114, 2019.

Self WK, Holtzman DM: Emerging diagnostics and therapeutics for Alzheimer disease. Nat Med 29:2187, 2023.

32

Aphasia, Memory Loss, and Other Cognitive Disorders

M.-Marsel Mesulam

The cerebral cortex of the human brain contains ~20 billion neurons spread over an area of 2.5 m^2. The primary sensory and motor areas constitute 10% of the cerebral cortex. The rest is subsumed by modality-selective, heteromodal, paralimbic, and limbic areas collectively known as the *association cortex* **(Fig. 32-1)**. The association cortex mediates the integrative processes that subserve cognition, emotion, and comportment. A systematic testing of these mental functions is necessary for the effective clinical assessment of association cortex and its diseases. According to current thinking, there are no centers for "hearing words," "perceiving space," or "storing memories." Cognitive and behavioral functions (domains) are coordinated by intersecting

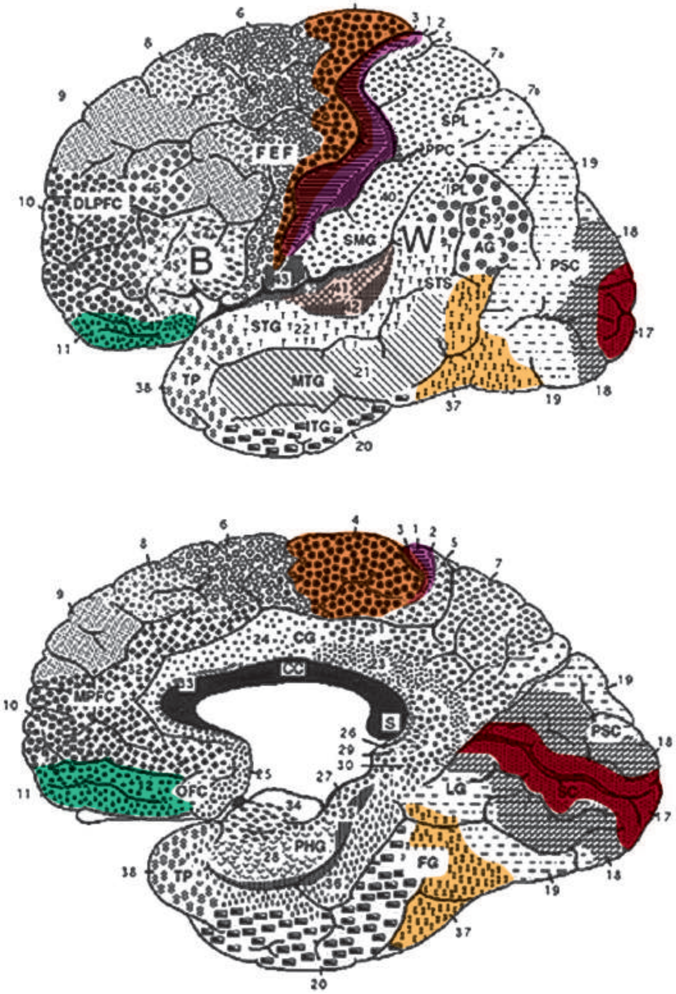

FIGURE 32-1 Lateral (*top*) and medial (*bottom*) views of the cerebral hemispheres. The numbers refer to the Brodmann cytoarchitectonic designations. Area 17 corresponds to the primary visual cortex, 41–42 to the primary auditory cortex, 1–3 to the primary somatosensory cortex, and 4 to the primary motor cortex. The rest of the cerebral cortex contains association areas. AG, angular gyrus; B, Broca's area; CC, corpus callosum; CG, cingulate gyrus; DLPFC, dorsolateral prefrontal cortex; FEF, frontal eye fields (premotor cortex); FG, fusiform gyrus; IPL, inferior parietal lobule; ITG, inferior temporal gyrus; LG, lingual gyrus; MPFC, medial prefrontal cortex; MTG, middle temporal gyrus; OFC, orbitofrontal cortex; PHG, parahippocampal gyrus; PPC, posterior parietal cortex; PSC, peristriate cortex; SC, striate cortex; SMG, supramarginal gyrus; SPL, superior parietal lobule; STG, superior temporal gyrus; STS, superior temporal sulcus; TP, temporopolar cortex; W, Wernicke's area.

large-scale neural networks that contain interconnected cortical and subcortical components. Five anatomically defined *large-scale networks* are most relevant to clinical practice: (1) a left-dominant perisylvian network for language, (2) a right-dominant parietofrontal network for spatial orientation, (3) an occipitotemporal network for face and object recognition, (4) a limbic network for episodic memory and emotional modulation, and (5) a prefrontal network for comportment and executive control of cognition. Investigations based on functional imaging have also identified a *default mode network*, which becomes activated when the person is not engaged in a specific task requiring attention to external events. The clinical consequences of damage to this network are not yet fully defined.

THE LEFT PERISYLVIAN NETWORK FOR LANGUAGE AND APHASIAS IN STROKE AND DEMENTIA

The production and comprehension of words and sentences is dependent on the integrity of a distributed network located along the perisylvian region of the language-dominant (usually left) hemisphere. One hub, situated in the inferior frontal gyrus, is known as *Broca's area*. Damage to this region impairs verbal fluency and the grammatical structure of sentences. The second hub, *Wernicke's area,* is traditionally sited within the temporoparietal junction and considered critical for language comprehension based on the examination of patients with stroke. However, patients with the neurodegenerative syndrome of primary progressive aphasia (PPA) have shown that sentence and word comprehension are dissociable, that sentence comprehension (i.e., the decoding of syntax and morphology) is a widely distributed faculty jointly subserved by Broca's and Wernicke's areas, and that the areas critical for word comprehension are more closely associated with the anterior temporal lobe than with Wernicke's area. All components of the language network are interconnected with each other and with surrounding parts of the frontal, parietal, and temporal lobes. An important pathway interconnecting components of this network is the *arcuate fasciculus*. Damage to the language network gives rise to aphasias. Aphasia should be diagnosed only when there are deficits in the formal aspects of language, such as word finding, repetition, word choice, comprehension, spelling, or grammar. Dysarthria, apraxia of speech, and mutism do not by themselves lead to a diagnosis of aphasia. In ~90% of right-handers and 60% of left-handers, aphasia occurs only after lesions of the left hemisphere.

■ CLINICAL EXAMINATION

The clinical examination of language should include the assessment of naming, spontaneous speech, comprehension, repetition, reading, and writing. A deficit of naming (*anomia*) is the single most common finding in aphasic patients. When asked to name a common object, the patient may fail to come up with the appropriate word, may provide a circumlocutious description of the object ("the thing for writing"), or may come up with the wrong word (*paraphasia*). If the patient offers an incorrect but related word ("elevator" for "escalator"), the error is known as a *semantic paraphasia*; if the word approximates the correct answer but is phonetically inaccurate ("plentil" for "pencil"), it is known as a *phonemic paraphasia*. In most anomias, the patient cannot retrieve the appropriate name when shown an object but can point to the appropriate object when the name is provided by the examiner. This is known as a one-way (or retrieval-based) naming deficit. A two-way (comprehension-based or semantic) naming deficit exists if the patient can neither provide, define, or recognize the correct noun. *Spontaneous speech* is described as "fluent" if it maintains appropriate output volume, phrase length, and melody or as "nonfluent" if it is sparse and halting and average utterance length is below four words. The examiner also should note the integrity of *grammar* as manifested by word order (syntax), tenses, suffixes, prefixes, plurals, and possessives. *Comprehension* can be tested by assessing the patient's ability to follow conversation, answer yes-no questions ("Can a dog fly?" "Does it snow in summer?"), point to appropriate objects ("Where is the source of illumination in this room?"), decode passive voice sentences, or define the meaning of words. *Repetition* is assessed by asking the patient to repeat single words, short sentences, or strings of words such as "No ifs, ands, or buts." The testing of repetition with tongue twisters such as "hippopotamus" and "Irish constabulary" provides a better assessment of dysarthria and apraxia of speech than of aphasia. It is important to make sure that the number of words does not exceed the patient's attention span. Otherwise, the failure of repetition becomes a reflection of the narrowed attention span (auditory working memory) rather than an indication of an aphasic deficit caused by dysfunction of a hypothetical *phonological loop* in the language network. *Reading* should be assessed for deficits in reading aloud as well as comprehension. *Alexia* describes an inability to either read aloud or comprehend written words and sentences; *agraphia* (or dysgraphia) is used to describe an acquired deficit in spelling.

Aphasias caused by stroke start suddenly and display maximal deficits at the onset **(Table 32-1)**. These are the "classic" aphasias described in the literature. Damage in these cases encompasses cerebral cortex as well as underlying deep white matter pathways interconnecting otherwise unaffected cortical areas. Progressive neurodegenerative diseases of the cerebral cortex have much greater cellular, laminar, and regional selectivity of peak atrophy sites, giving rise to different clinic-anatomical patterns of dysfunction from strokes centered in similar locations. For example, the full syndrome of Wernicke's aphasia is rarely, if ever, encountered in neurodegenerative diseases affecting Wernicke's area. Therefore, the diagnosis of Wernicke's aphasia implies a cause that is most commonly cerebrovascular.

■ STROKE APHASIAS

Wernicke's Aphasia Comprehension is impaired for spoken and written words and sentences. Language output is fluent but is highly

TABLE 32-1 Clinical Features of Aphasias and Related Conditions Commonly Seen in Cerebrovascular Accidents

	COMPREHENSION	REPETITION OF SPOKEN LANGUAGE	NAMING	FLUENCY
Wernicke's	Impaired	Impaired	Impaired	Preserved or increased
Broca's	Preserved (except grammar)	Impaired	Impaired	Decreased
Global	Impaired	Impaired	Impaired	Decreased
Conduction	Preserved	Impaired	Impaired	Preserved
Nonfluent (anterior) transcortical	Preserved	Preserved	Impaired	Impaired
Fluent (posterior) transcortical	Impaired	Preserved	Impaired	Preserved
Isolation	Impaired	Echolalia	Impaired	No purposeful speech
Anomic	Preserved	Preserved	Impaired	Preserved except for word-finding pauses
Pure word deafness	Impaired only for spoken language	Impaired	Preserved	Preserved
Pure alexia	Impaired only for reading	Preserved	Preserved	Preserved

paraphasic and circumlocutious. Paraphasic errors may lead to strings of neologisms, a finding known as "jargon aphasia." Speech contains few substantive nouns. The output is therefore voluminous but uninformative. For example, a patient attempts to describe how his wife accidentally threw away something important, perhaps his dentures: "We don't need it anymore, she says. And with it when that was downstairs was my teeth-tick … a … den … dentith … my dentist. And they happened to be in that bag … see? …Where my two … two little pieces of dentist that I use … that I … all gone. If she throws the whole thing away … visit some friends of hers and she can't throw them away."

Gestures and pantomime do not improve communication. The patient may not realize that their language is incomprehensible and may appear angry and impatient when the examiner fails to decipher the meaning of a severely paraphasic statement. In some patients, this type of aphasia can be associated with severe agitation and paranoia. The ability to follow commands aimed at axial musculature may be preserved. The dissociation between the failure to understand simple questions ("What is your name?") in a patient who rapidly closes their eyes, sits up, or rolls over when asked to do so is characteristic of Wernicke's aphasia and helps differentiate it from deafness, psychiatric disease, or malingering. Patients with Wernicke's aphasia cannot express their thoughts in meaning-appropriate words and cannot decode the meaning of words in any modality of input. This aphasia therefore has expressive as well as receptive components. Repetition, naming, reading, and writing also are impaired.

The lesion site commonly associated with Wernicke's aphasia caused by cerebrovascular accidents (CVAs) is in the posterior portion of the language network. An embolus to the inferior division of the middle cerebral artery (MCA), to the posterior temporal or angular branches in particular, is the most common etiology **(Chap. 437)**. Intracerebral hemorrhage, head trauma, and neoplasm are other causes of Wernicke's aphasia. A coexisting right hemianopia or superior quadrantanopia is common, and mild right nasolabial flattening may be found, but otherwise, the examination is often unrevealing. The paraphasic, neologistic speech in an agitated patient with an otherwise unremarkable neurologic examination may lead to the suspicion of a primary psychiatric disorder such as schizophrenia or mania, but the other components characteristic of acquired aphasia and the absence of prior psychiatric disease usually settle the issue. Prognosis for recovery of language function is guarded.

Broca's Aphasia Speech is nonfluent, labored, interrupted by many word-finding pauses, and usually dysarthric. It is impoverished in function words and action verbs but enriched in meaning-appropriate nouns. Abnormal word order and the inappropriate deployment of *bound morphemes* (word endings used to denote tenses, possessives, or plurals) lead to a characteristic agrammatism. Speech is telegraphic and pithy but quite informative. In the following passage, a patient with Broca's aphasia describes his medical history: "I see … the dotor, dotor sent me … Bosson. Go to hospital. Dotor … kept me beside. Two, tee days, doctor send me home."

Output may be reduced to a grunt or single word ("yes" or "no"), which is emitted with different intonations to express approval or disapproval. In addition to fluency, naming and repetition are impaired. Comprehension of spoken language is intact except for syntactically difficult sentences with a passive voice structure or embedded clauses (if Sue is killed by George, who remains alive?). Broca's aphasia is therefore more than an "expressive" or "motor" disorder and includes a comprehension deficit in decoding syntax. Patients with Broca's aphasia can be tearful, easily frustrated, and profoundly distressed. Insight into their condition is preserved, in contrast to Wernicke's aphasia. Even when spontaneous speech is severely dysarthric, the patient may be able to display a relatively normal articulation of words when singing. This dissociation has been used to develop specific therapeutic approaches (melodic intonation therapy) for Broca's aphasia. Additional neurologic deficits are invariably present and include right facial weakness, hemiparesis, or hemiplegia, and a buccofacial apraxia characterized by an inability to carry out motor commands involving oropharyngeal and facial musculature (e.g., patients are unable to demonstrate how to blow out a match or suck through a straw). The cause is most often infarction of Broca's area (the inferior frontal convolution; "B" in Fig. 32-1) and surrounding anterior perisylvian and insular cortex due to occlusion of the superior division of the MCA **(Chap. 437)**. Mass lesions, including tumor, intracerebral hemorrhage, and abscess, also may be responsible. When the cause of Broca's aphasia is stroke, recovery of language function generally peaks within 2–6 months, after which time further progress is limited. Speech therapy is more successful than in Wernicke's aphasia.

Conduction Aphasia Speech output is fluent but contains many phonemic paraphasias, comprehension of spoken language is intact, and repetition is severely impaired. Naming elicits phonemic paraphasias, and spelling is impaired. Reading aloud is impaired, but reading comprehension is preserved. Reading aloud may elicit semantically related substitutions ("bite" for "poodle"), a pattern known as deep dyslexia. The responsible lesion, usually a CVA in the temporoparietal or dorsal perisylvian region, is thought to interfere with the function of the arcuate fasciculus, which interconnects Broca's with Wernicke's area. Occasionally, a stroke-related Wernicke's aphasia may rapidly resolve into a conduction aphasia. The paraphasic and circumlocutious output in conduction aphasia interferes with the ability to express meaning, but this deficit is not nearly as severe as the one displayed by patients with classic Wernicke's aphasia. Associated neurologic signs in conduction aphasia vary according to the primary lesion site.

Transcortical Aphasias: Fluent and Nonfluent Clinical features of *fluent (posterior) transcortical aphasia* are similar to those of Wernicke's aphasia, but repetition is intact. The lesion site disconnects the intact core of the language network from other temporoparietal association areas. Associated neurologic findings may include hemianopia. Cerebrovascular lesions (e.g., infarctions in the posterior watershed zone) and neoplasms that involve the temporoparietal cortex posterior to Wernicke's area are common causes. The features of *nonfluent (anterior) transcortical aphasia* are similar to those of Broca's aphasia, but repetition is intact and agrammatism is less pronounced. The neurologic examination may be otherwise intact, but a right hemiparesis also can exist. The lesion site disconnects the intact language network from prefrontal areas of the brain and usually involves the anterior watershed zone between anterior and MCA territories or the supplementary motor cortex in the territory of the anterior cerebral artery.

Global and Isolation Aphasias *Global aphasia* represents the combined dysfunction of Broca's and Wernicke's areas and usually results from strokes that involve the entire MCA distribution in the left hemisphere. Speech output is nonfluent, and comprehension of language is severely impaired. Related signs include right hemiplegia, hemisensory loss, and homonymous hemianopia. *Isolation aphasia* represents a combination of the two transcortical aphasias. Comprehension is severely impaired, and there is no purposeful speech output. The patient may parrot fragments of heard conversations (*echolalia*), indicating that the neural mechanisms for repetition are at least partially intact. This condition represents the dysfunction of the language network when it is isolated from other regions of the brain. Broca's and Wernicke's areas tend to be spared, but there is damage to the surrounding frontal, parietal, and temporal cortex. Lesions are patchy and can be associated with anoxia, carbon monoxide poisoning, or complete watershed zone infarctions.

Anomic Aphasia This form of aphasia may be considered the "minimal dysfunction" syndrome of the language network. Articulation, comprehension, and repetition are intact, but confrontation naming, word finding, and spelling are impaired. Word-finding pauses or failures elicit paraphasic, circumlocutious, and uninformative speech. The lesion sites can be anywhere within the left hemisphere language network, including the middle and inferior temporal gyri. *Anomic aphasia is the single most common language disturbance seen in head trauma, metabolic encephalopathy, and Alzheimer's disease.*

Subcortical Aphasia Damage to subcortical components of the language network (e.g., the striatum and thalamus of the left hemisphere) also can lead to aphasia. The resulting syndromes contain combinations of deficits in the various aspects of language but rarely fit the specific patterns described in Table 32-1. In a patient with a CVA, an anomic aphasia accompanied by dysarthria or a fluent aphasia with hemiparesis should raise the suspicion of a subcortical lesion site.

Pure Word Deafness The most common causes are either bilateral or left-sided MCA strokes affecting the superior temporal gyrus. The net effect of the underlying lesion is to interrupt the flow of information from the auditory association cortex to the language network. Patients have no difficulty understanding written language and can express themselves well in spoken or written language. They have no difficulty interpreting and reacting to environmental sounds if the primary auditory cortex and auditory association areas of the right hemisphere are spared. Because auditory information cannot be conveyed to the language network, however, it cannot be decoded into neural word representations, and the patient reacts to speech as if it were in an alien tongue that cannot be deciphered. Patients cannot repeat spoken language but have no difficulty naming objects. In time, patients with pure word deafness teach themselves lipreading and may appear to have improved. There may be no additional neurologic findings, but agitated paranoid reactions are common in the acute stages. Cerebrovascular lesions are the most common cause.

Pure Alexia Without Agraphia This is the visual equivalent of pure word deafness. The lesions (usually a combination of damage to the left occipital cortex and to a posterior sector of the corpus callosum—the splenium) interrupt the flow of visual input into the language network. There is usually a right hemianopia. The patient can understand and produce spoken language, name objects in the left visual hemifield, repeat, and write. However, the patient acts as if illiterate when asked to read even the simplest sentence because the visual information from the written words (presented to the intact left visual hemifield) cannot reach the language network. Objects in the left hemifield may be named accurately because they activate nonvisual associations in the right hemisphere, which in turn can access the language network through transcallosal pathways anterior to the splenium. Patients with this syndrome also may lose the ability to name colors, although they can match colors. This is known as a *color anomia*. The most common etiology of pure alexia is a vascular lesion in the territory of the posterior cerebral artery or an infiltrating neoplasm in the left occipital cortex that involves the optic radiations as well as the crossing fibers of the splenium. Because the posterior cerebral artery also supplies medial temporal components of the limbic system, a patient with pure alexia also may experience an amnesia, but this is usually transient because the limbic lesion is unilateral.

Apraxia and Aphemia *Apraxia* designates a complex motor deficit that cannot be attributed to pyramidal, extrapyramidal, cerebellar, or sensory dysfunction and that does not arise from the patient's failure to understand the nature of the task. *Apraxia of speech* is used to designate articulatory abnormalities in the duration, fluidity, and stress of syllables that make up words. It can arise with CVAs in the posterior part of Broca's area or in frontotemporal lobar degenerations (FTLD) with tauopathy affecting the left frontal lobe. *Aphemia* is a severe form of acute speech deficit that presents with impaired fluency (often mutism). Recovery is the rule and involves an intermediate stage of hoarse whispering. Writing, reading, and comprehension are intact, and so this is not a true aphasic syndrome. CVAs in parts of Broca's area or subcortical lesions that undercut its connections with other parts of the brain may be present. Occasionally, the lesion site is on the medial aspects of the frontal lobes and may involve the supplementary motor cortex of the left hemisphere. *Ideomotor limb apraxia* is diagnosed when instructions to perform a specific motor act or pantomime the use of a common tool (a comb, hammer, or toothbrush) in the absence of the real object cannot be followed. The patient's ability to comprehend the command is ascertained by demonstrating multiple movements and establishing that the correct one can be recognized. Some patients with this type of apraxia can imitate the appropriate movement when it is demonstrated by the examiner and show no impairment when handed the real object, indicating that the sensorimotor mechanisms necessary for the movement are intact. This clinical pattern indicates a disconnection of the language network from pyramidal motor systems so that commands to execute complex movements are understood but cannot be conveyed to the appropriate motor areas. It is caused by lesions in the left hemisphere and is commonly associated with aphasic syndromes, especially Broca's aphasia and conduction aphasia. Because the handling of real objects is not impaired, ideomotor apraxia by itself causes no major limitation of daily living activities. *Buccofacial apraxia* involves apraxic deficits in movements of the face and mouth, elicited by asking the patient to cough or blow a match. Patients with lesions of the anterior corpus callosum can display ideomotor apraxia confined to the left side of the body, a sign known as *sympathetic dyspraxia*. A severe form of sympathetic dyspraxia, known as the *alien hand* syndrome, is characterized by additional features of motor disinhibition on the left hand. *Ideational apraxia* refers to a deficit in the sequencing of goal-directed movements in patients who have no difficulty executing the individual components of the sequence. For example, when the patient is asked to pick up a pen and write, the sequence of uncapping the pen, placing the cap at the opposite end, turning the point toward the writing surface, and writing may be disrupted, and the patient may be seen trying to write with the wrong end of the pen or even with the removed cap. These motor sequencing problems usually are seen in the context of confusional states and dementias rather than focal lesions associated with aphasic conditions. *Limb-kinetic apraxia* involves clumsiness in the use of tools or objects that cannot be attributed to sensory, pyramidal, extrapyramidal, or cerebellar dysfunction. Different forms of apraxia as well as the alien hand syndrome can also occur in the syndrome of *corticobasal degeneration*.

Gerstmann's Syndrome The combination of *acalculia* (impairment of simple arithmetic), *dysgraphia* (impaired writing), *finger anomia* (an inability to name individual fingers such as the index and thumb), and *right-left confusion* (an inability to tell whether a hand, foot, or arm of the patient or examiner is on the right or left side of the body) is known as Gerstmann's syndrome. In making this diagnosis, it is important to establish that the finger and left-right naming deficits are not part of a more generalized anomia and that the patient is not otherwise aphasic. When Gerstmann's syndrome arises acutely and in isolation, it is commonly associated with damage to the inferior parietal lobule (especially the angular gyrus) in the left hemisphere.

Pragmatics and Prosody *Pragmatics* refers to aspects of language that communicate attitude, affect, and the figurative rather than literal aspects of a message (e.g., "green thumb" does not refer to the actual color of the finger). One component of pragmatics, *prosody,* refers to variations of intonation that modify the nature of the message that is being conveyed. For example, the two statements "He *is* clever." and "He is *clever*?" contain identical word choice and syntax but convey vastly different messages because of differences in stress and intonation. Damage to right hemisphere regions corresponding to Broca's area impairs the ability to introduce meaning-appropriate prosody into spoken language (expressive aprosodia). The patient produces grammatically correct language with accurate word choice, but the statements are uttered in a monotone that gives the erroneous impression of depression or indifference. Other aspects of pragmatics, especially the ability to infer the figurative aspect of a message or the emotion expressed by intonation and body language, become impaired by damage to the right hemisphere or frontal lobes.

NEURODEGENERATIVE PROGRESSIVE APHASIAS

Several neurodegenerative syndromes, such as typical Alzheimer-type (amnestic; **Chap. 442**) and frontotemporal (behavioral; **Chap. 443**) dementias, can also include language impairments as the disease progresses. In these cases, the aphasia is an ancillary component of the overall syndrome. A diagnosis of PPA is justified only if the language disorder (i.e., aphasia) arises in relative isolation, becomes the primary concern that brings the patient to medical attention, and remains the most salient deficit for 1–2 years. PPA can be caused by either FTLD (tauopathy or TDP-43 proteinopathy) or Alzheimer's disease (AD) neuropathology. Rarely, an identical syndrome can be caused by the prion disorder Creutzfeldt-Jacob disease (CJD) but with a more rapid progression **(Chap. 449)**. The common denominator for all entities causing PPA is a selective asymmetric predilection for the language-dominant (usually left) hemisphere.

Four major subtypes of PPA can be recognized.

Agrammatic PPA The *agrammatic variant* is characterized by consistently low fluency and impaired grammar but intact word comprehension. It most closely resembles Broca's aphasia or anterior transcortical aphasia but lacks right hemiparesis, is much less likely to be associated with dysarthria, and tends to have word order (syntax) impairments that are much more conspicuous than in CVAs. Peak sites of neuronal loss (gray matter atrophy) include Broca's area as well as adjacent parts of the frontal lobe. The neuropathology is usually an FTLD with tauopathy but can also be an atypical form of AD pathology.

Semantic PPA The *semantic variant* is characterized by preserved fluency and syntax but poor single-word comprehension and profound two-way naming impairments. This kind of aphasia is not seen with CVAs. It differs from Wernicke's aphasia because speech is usually informative, and repetition is intact. The word comprehension deficit is much more severe than in stroke-related posterior transcortical aphasia. The most severe comprehension deficit is seen for nouns rather than action verbs. Comprehension of sentences made up of highly familiar words is relatively preserved and patients are quite good at surmising the gist of conversations through contextual cues. Such patients may appear unimpaired during casual small talk but become puzzled upon encountering words such as "pumpkin" or "umbrella." Peak atrophy sites are in the left anterior temporal lobe. Wernicke's area may remain intact, indicating that the cortex of the anterior temporal lobe rather than the traditional site of Wernicke's area plays a critical role in the comprehension of words. This region was not included within the classic language network, probably because it is not a common site for focal CVAs. The neuropathology is frequently an FTLD with abnormal precipitates of the 43-kDa transactive response DNA-binding protein TDP-43 of type C.

Logopenic PPA The *logopenic variant* is characterized by preserved syntax and comprehension but frequent and severe word-finding pauses, anomia, circumlocutions, and simplifications during spontaneous speech. Repetition is usually impaired. Peak atrophy sites are located in the temporoparietal junction and posterior temporal lobe, overlapping with Wernicke's area. However, the comprehension impairment of *Wernicke's aphasia* is absent probably because the underlying deep white matter, frequently damaged by CVAs, remains relatively intact in PPA. The repetition impairment suggests that parts of Wernicke's area are critical for auditory working memory. In contrast to Broca's aphasia or agrammatic PPA, speech may appear fluent if the patient is allowed to engage in small talk. Logopenic PPA resembles the anomic aphasia of Table 32-1 but usually has longer and more frequent word-finding pauses. When repetition is impaired, the aphasia resembles the *conduction aphasia* in Table 32-1. Of all PPA subtypes, this is the one most commonly associated with the pathology of AD, but FTLD can also be the cause.

In addition to these three major subtypes, there is also a *mixed* type of PPA where grammar, fluency, and word comprehension are jointly impaired. This is most like the global aphasia of Table 32-1. Rarely, PPA can present with patterns reminiscent of *pure word deafness* or *Gerstmann's syndrome*.

THE PARIETOFRONTAL NETWORK FOR SPATIAL ORIENTATION

Adaptive spatial orientation is subserved by a large-scale network containing three major cortical components. The *cingulate cortex* provides access to a motivational mapping of the extrapersonal space, the *posterior parietal cortex* to a sensorimotor representation of salient extrapersonal events, and the *frontal eye fields* to motor strategies for attentional behaviors **(Fig. 32-2)**. Subcortical components of this network include the striatum and the thalamus. Damage to this network can undermine the distribution of attention within the extrapersonal space, giving rise to hemispatial neglect, simultanagnosia, navigation errors, and object finding failures.

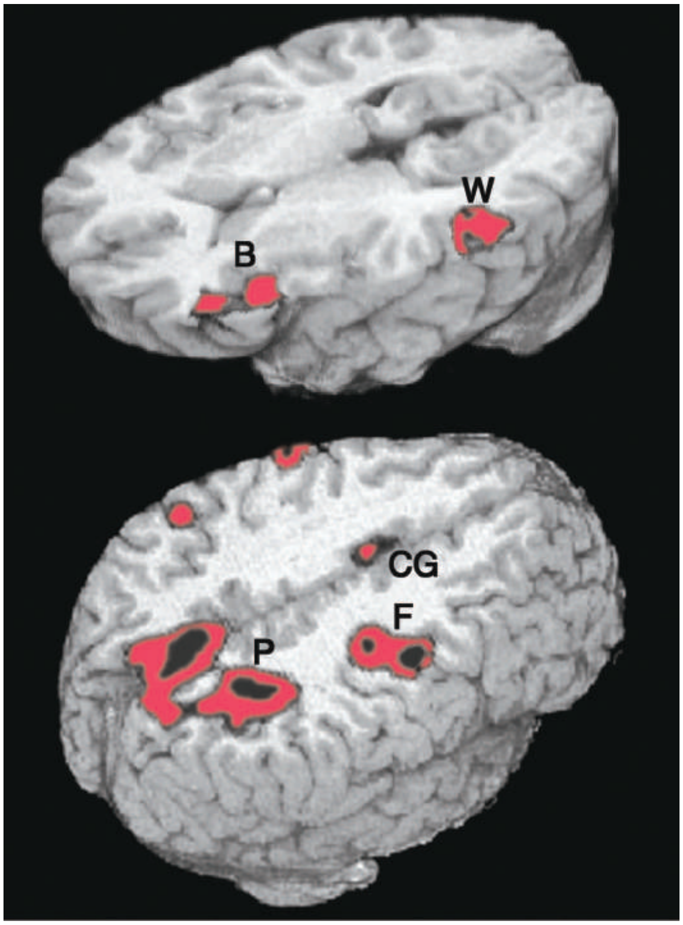

FIGURE 32-2 Functional magnetic resonance imaging of language and spatial attention in neurologically intact subjects. The red and black areas show regions of task-related significant activation. (*Top*) The subjects were asked to determine if two words were synonymous. This language task led to the simultaneous activation of the two components of the language network, Broca's area (B) and Wernicke's area (W). The activations are exclusively in the left hemisphere. (*Bottom*) The subjects were asked to shift spatial attention to a peripheral target. This task led to the simultaneous activation of the three epicenters of the attentional network: the posterior parietal cortex (P), the frontal eye fields (F), and the cingulate gyrus (CG). The activations are predominantly in the right hemisphere. *(Courtesy of Darren Gitelman, MD.)*

■ HEMISPATIAL NEGLECT

Contralesional hemispatial neglect represents one outcome of damage to the cortical or subcortical components of this network. *The traditional view that hemispatial neglect always denotes a parietal lobe lesion is inaccurate*. According to one model of spatial cognition, the right hemisphere directs attention within the *entire* extrapersonal space, whereas the left hemisphere directs attention mostly within the contralateral right hemispace. Consequently, left hemisphere lesions do not give rise to much contralesional neglect because the global attentional mechanisms of the right hemisphere can compensate for the loss of the *contralaterally* directed attentional functions of the left hemisphere. Right hemisphere lesions, however, give rise to severe contralesional left hemispatial neglect because the unaffected left hemisphere does not contain ipsilateral attentional mechanisms. This model is consistent with clinical experience, which shows that contralesional neglect is more common, more severe, and longer lasting after damage to the right hemisphere than after damage to the left hemisphere. Severe neglect for the right hemispace is rare, even in left-handers with left hemisphere lesions.

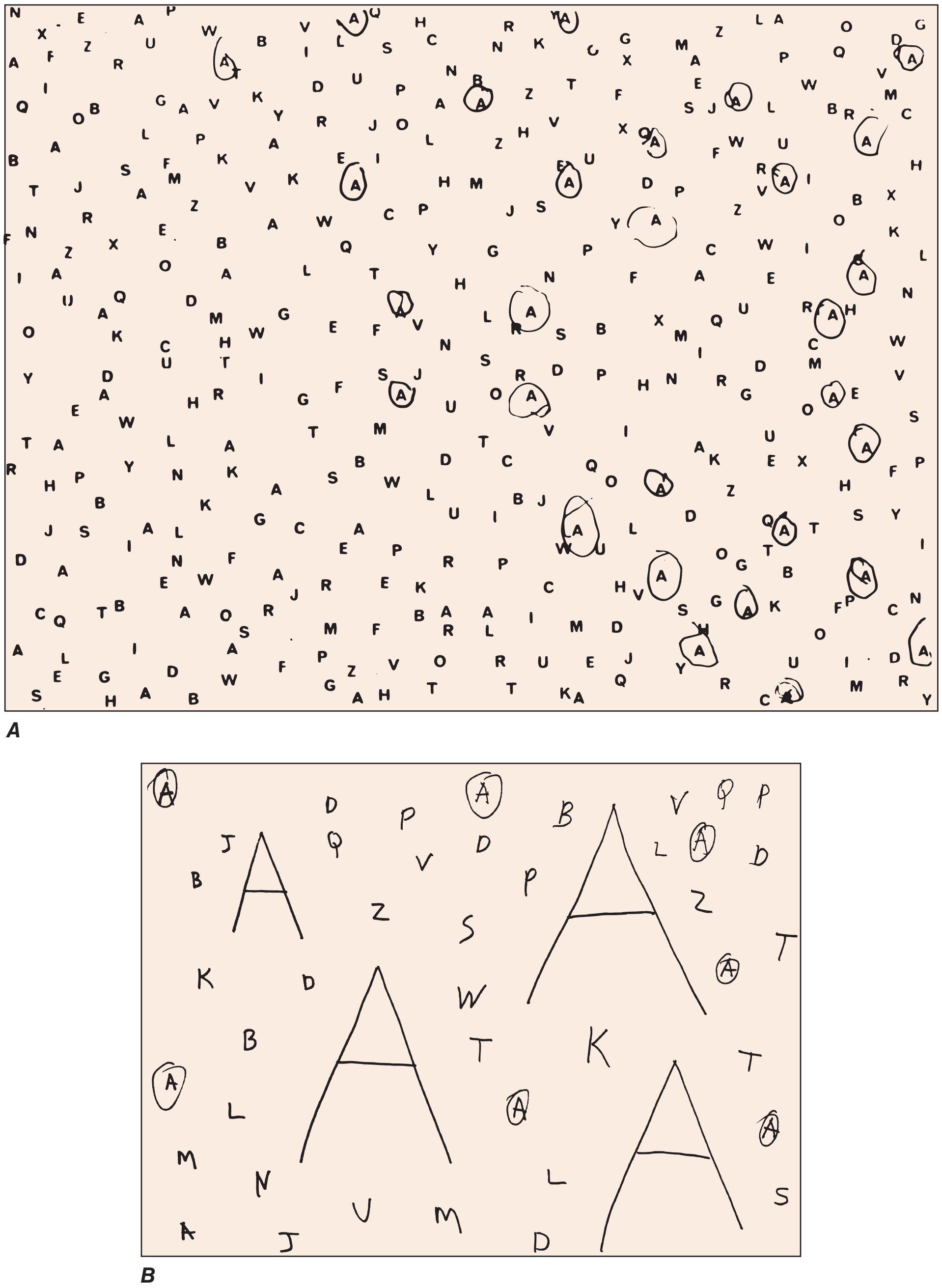

FIGURE 32-3 ***A.*** A 47-year-old man with a large frontoparietal lesion in the right hemisphere was asked to circle all the A's. Only targets on the right are circled. This is a manifestation of left hemispatial neglect. ***B.*** A 70-year-old woman with a 2-year history of degenerative dementia was able to circle most of the small targets but ignored the larger ones. This is a manifestation of simultanagnosia.

Clinical Examination Patients with severe neglect may fail to dress, shave, or groom the left side of the body; fail to eat food placed on the left side of the tray; and fail to read the left half of sentences. When asked to copy a simple line drawing, the patient fails to copy detail on the left, and when the patient is asked to write, there is a tendency to leave an unusually wide margin on the left. Two bedside tests that are useful in assessing neglect are *simultaneous bilateral stimulation* and *visual target cancellation*. In the former, the examiner provides either unilateral or simultaneous bilateral stimulation in the visual, auditory, and tactile modalities. After right hemisphere injury, patients who have no difficulty detecting unilateral stimuli on either side experience the bilaterally presented stimulus as coming only from the right. This phenomenon is known as *extinction* and reflects the sensory-representational aspect of hemispatial neglect. In the target detection task, targets (e.g., A's) are interspersed with foils (e.g., other letters of the alphabet) on a 21.5- to 28.0-cm (8.5–11 in.) sheet of paper, and the patient is asked to circle all the targets. A failure to detect targets on the left is a manifestation of the exploratory (motor) deficit in hemispatial neglect **(Fig. 32-3*A*)**. Hemianopia is not by itself sufficient to cause the target detection failure because the patient is free to turn the head and eyes to the left. Target detection failures therefore reflect a distortion of spatial attention, not just of sensory input. Some patients with neglect also may deny the existence of hemiparesis and may even deny ownership of the paralyzed limb, a condition known as *anosognosia*. Hemispatial neglect is almost never seen in neurodegenerative diseases that cause selective atrophy within components of the right hemisphere attentional network.

■ BÁLINT'S SYNDROME, SIMULTANAGNOSIA, DRESSING APRAXIA, CONSTRUCTION APRAXIA, AND ROUTE-FINDING IMPAIRMENTS

Bilateral involvement of the network for spatial attention, especially its parietal components, leads to a state of severe spatial disorientation known as *Bálint's syndrome*. Bálint's syndrome involves deficits in the orderly visuomotor scanning of the environment (*oculomotor apraxia*), accurate manual reaching toward visual targets (*optic ataxia*), and the ability to integrate visual information in the center of gaze with more peripheral information (*simultanagnosia*). A patient with simultanagnosia "misses the forest for the trees." For example, a patient who is shown a table lamp and asked to name the object may look at its circular base and call it an ashtray. Some patients with simultanagnosia report that objects they look at may vanish suddenly, probably indicating an inability to compute the oculomotor return to the original point of gaze after brief saccadic displacements. Movement and distracting stimuli greatly exacerbate the difficulties. Simultanagnosia can occur without the other two components of Bálint's syndrome, especially in association with the syndrome of *posterior cortical atrophy*, an atypical presentation of AD.

A modification of the letter cancellation task described above can be used for the bedside diagnosis of simultanagnosia. In this modification, some of the targets (e.g., A's) are made to be much larger than the others (7.5–10 cm vs 2.5 cm [3–4 in. vs 1 in.] in height), and all targets are embedded among foils. Patients with simultanagnosia display a counterintuitive but characteristic tendency to miss the larger targets **(Fig. 32-3*B*)**. This occurs because the information needed for the identification of the larger targets cannot be confined to a single foveation and requires the integration of visual information across multiple fixation points. The greater difficulty in the detection of the larger targets also indicates that poor acuity is not responsible for the impairment of visual function and that the problem is central rather than peripheral. The test shown in Fig. 32-3*B* is not by itself sufficient to diagnose simultanagnosia as a frontal network syndrome may lead to the omission of the oversized letters because they appear incongruous with the size of the paper and the patient cannot realize that the two types of target are symbolically identical despite being superficially different.

Bilateral parietal lesions can impair the integration of egocentric (self-centered) with allocentric (object centered) spatial coordinates. One manifestation is *dressing apraxia*. Patients with this condition are unable to align the body axis with the axis of the garment and can be seen struggling as they hold a coat from its bottom or extend their arm into a fold of the garment rather than into its sleeve. Lesions that involve the posterior parietal cortex also lead to severe difficulties in copying simple line drawings. This is known as a *construction apraxia* and is much more severe if the lesion is in the right hemisphere. In some patients with right hemisphere lesions, the drawing difficulties are confined to the left side of the figure and represent a manifestation of hemispatial neglect; in others, there is a more universal deficit in reproducing contours and three-dimensional perspective. Impairments of route finding can be included in this group of disorders and reflect the inability to orient the self with respect to external objects and landmarks.

Causes of Spatial Disorientation and the Posterior Cortical Atrophy Syndrome CVAs, head injury, and neoplasms in the right hemisphere are common causes of hemispatial neglect. Additional settings include watershed infarction, hypoglycemia, and sagittal sinus thrombosis. Depending on the site of the lesion, a patient with neglect also may have hemiparesis, hemihypesthesia, and hemianopia on the left, but these are not invariant findings. The majority of CVA patients display considerable improvement of hemispatial neglect, usually within the first several weeks. Bálint's syndrome, dressing apraxia, and route-finding impairments are more likely to result from bilateral dorsal parietal lesions.

A progressive form of spatial disorientation, known as *posterior cortical atrophy* (PCA) usually represents a variant of AD with unusual concentrations of neurofibrillary degeneration in the parieto-occipital cortex and the superior colliculus **(Fig. 32-4)**. Lewy body disease (LBD), CJD, and FTLD (corticobasal type) are other possible causes. The patient displays progressive hemispatial neglect, Bálint's syndrome, and route-finding impairments, usually accompanied by dressing and construction apraxia.

THE OCCIPITOTEMPORAL NETWORK FOR FACE AND OBJECT RECOGNITION

A patient with *prosopagnosia* cannot recognize familiar faces, including, sometimes, photographs of their own face. This is not a perceptual deficit because prosopagnosic patients easily can tell whether two faces are identical. Furthermore, a prosopagnosic patient who cannot recognize a familiar face by visual inspection alone can use auditory cues to reach appropriate recognition if allowed to listen to the person's voice. The deficit in prosopagnosia is therefore modality-specific and reflects the inability of the visual percept to activate the multimodal associations that are necessary for concept formation. The responsible lesion is in the ventral temporal lobe. Prosopagnosic patients characteristically have no difficulty with the generic identification of a face as a face or a car as a car but may not recognize the identity of an individual face or the make of an individual car. This reflects a visual recognition deficit for proprietary (unique) features that characterize individual members of an object class. When recognition problems become more generalized and extend to the generic identification of common objects, the condition is known as *visual object agnosia*. A third level of severity occurs when the deficit becomes multimodal and impairs the recognition of objects by their characteristic sound or vocalization (siren of ambulance or barking

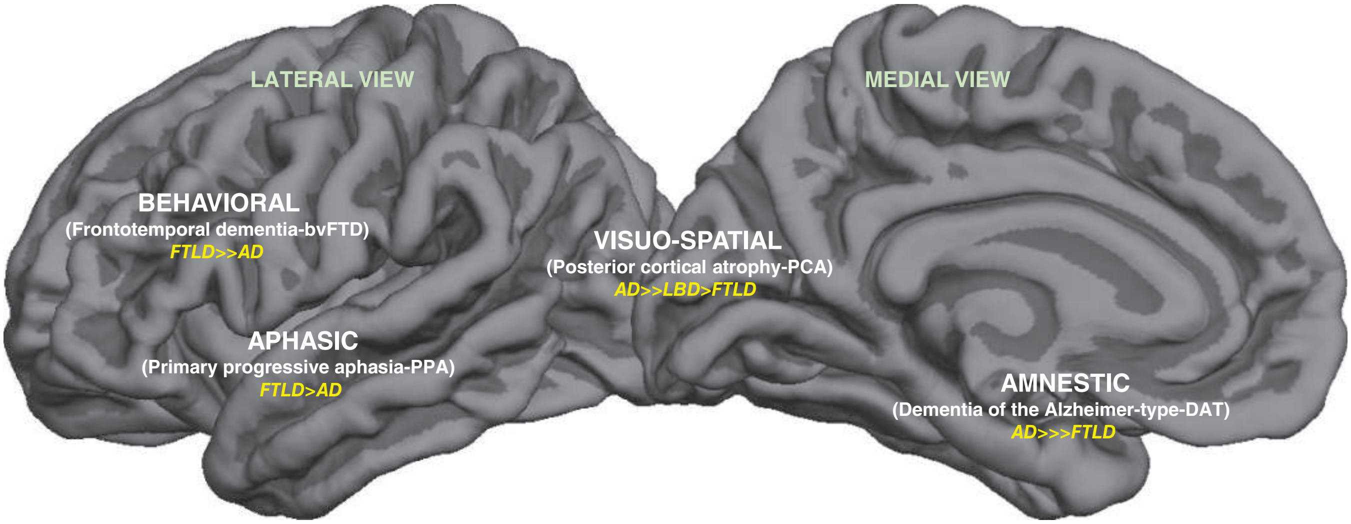
LATERAL VIEW
MEDIAL VIEW
BEHAVIORAL
(Frontotemporal dementia-bvFTD)
FTLD>>AD
VISUO-SPATIAL
(Posterior cortical atrophy-PCA)
AD>>LBD>FTLD
APHASIC
(Primary progressive aphasia-PPA)
FTLD>AD
AMNESTIC
(Dementia of the Alzheimer-type-DAT)
AD>>>FTLD

FIGURE 32-4 Four focal dementia syndromes and their most likely neuropathologic correlates. AD, Alzheimer's disease; bvFTD, behavioral variant frontotemporal dementia; DAT, amnestic dementia of the Alzheimer type; FTLD, frontotemporal lobar degeneration (tau or TDP-43 type); LBD, Lewy body disease; PCA, posterior cortical atrophy syndrome; PPA, primary progressive aphasia.

of a dog). A patient with anomia or semantic PPA cannot name the object but can describe its use. In contrast, a patient with visual agnosia is unable either to name objects or describe their uses. Face and object recognition disorders also can result from the simultanagnosia of Bálint's syndrome, in which case they are known as *apperceptive* agnosias as opposed to the *associative* agnosias that result from inferior temporal lobe lesions.

CAUSES AND RELATION TO SEMANTIC DEMENTIA

The characteristic lesions in prosopagnosia and visual object agnosia of acute onset consist of bilateral or right-sided infarctions in the territory of the posterior cerebral arteries that involve the fusiform gyrus. Associated deficits can include visual field defects (especially superior quadrantanopias) and a centrally based color blindness known as achromatopsia. Degenerative diseases of anterior and inferior temporal cortex of the right hemisphere can also cause associative prosopagnosia and multimodal object agnosia, whereas degenerative diseases of the left hemisphere, as in semantic PPA, cause word comprehension impairments. The combination of the two is known as *semantic dementia* and is usually seen in FTLD-TDP type C or Pick's disease. Patients with semantic dementia fail to recognize faces and objects and cannot understand the meaning of words denoting objects. This can be differentiated from semantic PPA where there is severe impairment in understanding words that denote objects and in naming faces and objects but a relative preservation of face and object recognition. Acute onset of the semantic dementia syndrome can be associated with herpes simplex encephalitis.

LIMBIC NETWORK FOR EXPLICIT MEMORY AND AMNESIA

Limbic areas (e.g., the hippocampus, amygdala, and entorhinal cortex), paralimbic areas (e.g., the cingulate gyrus, insula, temporopolar cortex, and parts of orbitofrontal regions), the anterior and medial nuclei of the thalamus, the medial and basal parts of the striatum, and the hypothalamus collectively constitute a distributed network known as the *limbic system*. The behavioral affiliations of this network can be classified into two groups. One includes the coordination of emotion, motivation, affiliative behaviors, autonomic tone, and endocrine function. These functions are under the influence of the amygdala and anterior paralimbic areas. They overlap with the *salience network*. The two neurologic syndromes associated with disruptions in these anterior limbic functions are temporal lobe epilepsy and behavioral variant frontotemporal dementia (bvFTD). An additional area of specialization for the limbic network and the one that is of most relevance to clinical practice is that of declarative (explicit) memory for recent episodes and experiences. This function is under the influence of the hippocampus, entorhinal cortex, posterior paralimbic areas, and limbic nuclei of the thalamus. This posterior part of the limbic system is also known as the Papez circuit. A disturbance of explicit memory is known as an *amnestic state*. In the absence of deficits in motivation, attention, language, or visuospatial function, the clinical diagnosis of a persistent global amnestic state is always associated with bilateral damage to the limbic network, usually within the hippocampo-entorhinal complex or the thalamus. Damage to the limbic network does not necessarily destroy memories but interferes with their conscious recall in coherent form. The individual fragments of information remain preserved despite the limbic lesions and can sustain what is known as *implicit memory*. For example, patients with amnestic states can acquire new motor or perceptual skills even though they may have no conscious knowledge of the experiences that led to the acquisition of these skills.

The memory disturbance in the amnestic state is multimodal and includes retrograde and anterograde components. The *retrograde amnesia* involves an inability to recall experiences that occurred before the onset of the amnestic state. Relatively recent events are more vulnerable to retrograde amnesia than more remote and more extensively consolidated events. A patient who comes to the emergency room complaining that they cannot remember their identity but can remember the events of the previous day does not have a neurologic cause of memory disturbance. The second and most important component of the amnestic state is the *anterograde amnesia*, which indicates an inability to store, retain, and recall new knowledge. Patients with amnestic states cannot remember what they ate a few hours ago or the details of an important event they may have experienced in the recent past. In the acute stages, there also may be a tendency to fill in memory gaps with inaccurate, fabricated, and often implausible information. This is known as *confabulation*. Confabulation is more common in cases where the underlying lesion also interferes with parts of the frontothalamic network, as in the case of the Wernicke-Korsakoff syndrome or traumatic head injury. Patients with the amnestic syndrome forget that they forget and tend to deny the existence of a memory problem when questioned.

CLINICAL EXAMINATION

A patient with an amnestic state is almost always disoriented, especially to time, and has little knowledge of current news. The anterograde component of an amnestic state can be tested with a list of four to five words read aloud by the examiner up to five times or until the patient can immediately repeat the entire list without an intervening delay. The next phase of the recall occurs after a period of 5–10 min during which the patient is engaged in other tasks. Amnestic patients fail this phase of the task and may even forget that they were given a list of words to remember. Accurate recognition of the words by multiple choice in a patient who cannot recall them indicates a less severe memory disturbance that affects mostly the retrieval stage of memory. The retrograde component of an amnesia can be assessed with questions related to autobiographical or historic events. The anterograde component of amnestic states is usually much more prominent than the retrograde component. In rare instances, occasionally associated with temporal lobe epilepsy or herpes simplex encephalitis, the retrograde component may dominate. Confusional states caused by toxic-metabolic encephalopathies and some types of frontal lobe damage lead to secondary memory impairments, especially at the stages of encoding and retrieval, even in the absence of limbic lesions. This sort of memory impairment can be differentiated from the amnestic state by the presence of additional impairments in the attention-related tasks described below in the section on the frontal lobes.

CAUSES, INCLUDING ALZHEIMER'S DISEASE

Neurologic diseases that give rise to an amnestic state include tumors (of the sphenoid wing, posterior corpus callosum, thalamus, or medial temporal lobe), infarctions (in the territories of the anterior or posterior cerebral arteries), head trauma, herpes simplex encephalitis, Wernicke-Korsakoff encephalopathy, autoimmune limbic encephalitis (e.g., antibodies to AMPAR, LGI1), paraneoplastic encephalitis (e.g., with Ma antibodies), and degenerative dementias such as AD and Pick's disease. The one common denominator of all these diseases is the presence of bilateral lesions within one or more components in the limbic network. Occasionally, unilateral left-sided hippocampal lesions can give rise to an amnestic state, but the memory disorder tends to be transient. Depending on the nature and distribution of the underlying neurologic disease, the patient also may have visual field deficits, eye movement limitations, or cerebellar findings.

The most common cause of progressive memory impairments in the elderly is AD. This is why a predominantly amnestic dementia is also known as a dementia of the Alzheimer type (DAT). A prodromal stage of DAT, when daily living activities are generally preserved, is known as amnestic mild cognitive impairment (aMCI). The predilection of the entorhinal cortex and hippocampus for early neurofibrillary degeneration by typical AD pathology is responsible for the initially selective impairment of episodic memory. In time, additional impairments in language, attention, and visuospatial skills emerge as the neurofibrillary degeneration spreads to additional neocortical areas. Less frequently, amnestic dementias can also be caused by FTLD.

Transient global amnesia is a distinctive syndrome usually seen in late middle age. Patients become acutely disoriented and repeatedly ask who they are, where they are, and what they are doing. The spell is characterized by anterograde amnesia (inability to retain new information) and a retrograde amnesia for relatively recent events that occurred before the onset. The syndrome usually resolves within 24–48 h and is followed by the filling in of the period affected by the retrograde amnesia, although there is persistent loss of memory for the events that occurred during the ictus. Recurrences are noted in ~20% of patients. Migraine, sudden autonomic activation (cold bath, intercourse), temporal lobe seizures, and perfusion abnormalities in the posterior cerebral territory have been postulated as causes of transient global amnesia. In many cases, no specific etiology is identified. The absence of associated neurologic findings occasionally may lead to the incorrect diagnosis of a psychiatric disorder.

THE PREFRONTAL NETWORK FOR EXECUTIVE FUNCTION AND BEHAVIOR

The frontal lobes can be subdivided into motor-premotor, dorsolateral prefrontal, medial prefrontal, and orbitofrontal components. The terms *frontal lobe syndrome* and *prefrontal cortex* refer only to the last three of these four components. These are the parts of the cerebral cortex that show the greatest phylogenetic expansion in primates, especially in humans. The dorsolateral prefrontal, medial prefrontal, and orbitofrontal areas, along with the subcortical structures with which they are interconnected (i.e., the head of the caudate, amygdala, and the dorsomedial nucleus of the thalamus), collectively make up a large-scale network that coordinates exceedingly complex aspects of human cognition and behavior. The prefrontal network overlaps with the salience network through parts of the orbitofrontal region. Impairments of social conduct and empathy seen in neurodegenerative frontal dementias (i.e., bvFTD) are attributed to pathology of the ventromedial parts of prefrontal cortex, whereas the dorsolateral parts are more closely related to the speed of information processing, working memory, and other aspects of executive functions. A nearly identical syndrome of bvFTD with or without additional anomia and prosopagnosia can arise in patients with right-sided anterior temporal lobe degeneration, reflecting the anatomical continuity of the anterior temporal lobe with orbitofrontal cortex.

The prefrontal network plays an important role in behaviors that require multitasking, the integration of thought with emotion, cognitive control, focused attention, mental flexibility, empathy, and the guidance of social conduct by context. In the *frontal abulic syndrome*, the patient shows a loss of initiative, creativity, and curiosity and displays a pervasive emotional blandness, apathy, and lack of empathy. In the *frontal disinhibition syndrome*, the patient becomes socially disinhibited and shows severe impairments of judgment, insight, foresight, and the ability to mind rules of conduct. The dissociation between intact intellectual function and a total lack of even rudimentary common sense is striking. Despite the preservation of all essential memory functions, the patient cannot learn from experience and continues to display inappropriate behaviors without appearing to feel emotional pain, guilt, or regret when those behaviors repeatedly lead to disastrous consequences. The impairments may emerge only in real-life situations when behavior is under minimal external control and may not be apparent within the structured environment of the medical office. Testing judgment by asking patients what they would do if they detected a fire in a theater or found a stamped and addressed envelope on the road is not very informative because patients who answer these questions wisely in the office may still act very foolishly in real-life settings. The physician must therefore be prepared to make a diagnosis of frontal lobe disease based on historic information alone even when the mental state is quite intact in the office examination.

■ CLINICAL EXAMINATION

The emergence of developmentally primitive reflexes, also known as frontal release signs, such as grasping (elicited by stroking the palm) and sucking (elicited by stroking the lips) are seen primarily in patients with large structural lesions that extend into the premotor components of the frontal lobes or in the context of metabolic encephalopathies. The majority of patients with prefrontal lesions and frontal lobe behavioral syndromes do not display these reflexes. Damage to the frontal lobe disrupts a variety of attention-related functions, including working memory (the transient online holding and manipulation of information), concentration span, the effortful scanning and retrieval of stored information, the inhibition of immediate but inappropriate responses, and mental flexibility. Digit span (which is on average seven forward and five reverse) is decreased, reflecting poor working memory; the recitation of the months of the year in reverse order (which should take <15 s) is slowed as another indication of poor working memory; and the fluency in producing words starting with the letter a, f, or s that can be generated in 1 min (normally ≥12 per letter) is diminished even in nonaphasic patients, indicating an impairment in the ability to search and retrieve information from long-term stores. In "go–no go" tasks (where the instruction is to raise the finger upon hearing one tap but keep it still upon hearing two taps), the patient shows a characteristic inability to inhibit the response to the "no go" stimulus. Mental flexibility (tested by the ability to shift from one criterion to another in sorting or matching tasks) is impoverished; distractibility by irrelevant stimuli is increased; and there is a pronounced tendency for impersistence and perseveration. The ability for abstracting similarities and interpreting proverbs is also undermined.

The attentional deficits disrupt the orderly registration and retrieval of new information and lead to *secondary* deficits of explicit memory. This is worsened by the low effort seen in abulic forms of the syndrome. The distinction of the underlying neural mechanisms is illustrated by the observation that severely amnestic patients (as in AD) who cannot remember events that occurred a few minutes ago may have intact if not superior working memory capacity as shown in tests of digit span. The use of the term *memory* to designate two completely different mental faculties is confusing. Working memory is a frontal network function that depends on the on-line holding of information for brief periods of time, whereas explicit memory depends on the hippocampus-dependent off-line storage and subsequent retrieval of the information.

■ CAUSES: TRAUMA, NEOPLASM, AND FRONTOTEMPORAL DEMENTIA

The abulic/apathetic syndrome tends to be associated with damage in dorsolateral or dorsomedial prefrontal cortex, and the disinhibition syndrome with damage in orbitofrontal or ventromedial cortex. When behavioral disinhibition arises in conjunction with asymmetric disease, the lesion tends to be predominantly on the right side of the brain. Common settings for frontal lobe syndromes include head trauma (traumatic brain injury), ruptured aneurysms, hydrocephalus, tumors (metastases, glioblastoma, meningioma), and focal neurodegenerations such as Pick's disease and other forms of FTLD. Progressive isolated impairment of comportment associated with degenerative diseases is known as the bvFTD syndrome. In many patients with bvFTD, the atrophy includes orbitofrontal cortex, anterior temporal lobe, insula, and anterior cingulate cortex. The behavioral changes in bvFTD can range from apathy to shoplifting, inappropriate conviviality, compulsive gambling, sexual indiscretions, remarkable lack of common sense, new ritualistic behaviors, and alterations in dietary preferences, usually leading to increased taste for sweets or rigid attachment to specific food items. Rarely, the bvFTD syndrome can arise in the context of an atypical form of AD pathology with a greater proportion of neurofibrillary tangles in prefrontal cortex.

Lesions in the caudate nucleus or in the dorsomedial nucleus of the thalamus (subcortical components of the prefrontal network) also can produce a frontal lobe syndrome affecting mostly executive functions. This is one reason why the changes in mental state associated with degenerative basal ganglia diseases such as Parkinson's disease and Huntington's disease display components of the frontal lobe syndrome. Bilateral multifocal lesions of the cerebral hemispheres, none of which are individually large enough to cause specific cognitive deficits such as aphasia and neglect, can collectively interfere with the connectivity and therefore integrative (executive) function of the prefrontal cortex. A frontal lobe syndrome, usually of the abulic form, is therefore the single most common behavioral profile associated with a variety of bilateral multifocal brain diseases, including metabolic encephalopathy, multiple sclerosis, and vitamin B_{12} deficiency, among others. To avoid making a diagnosis of "frontal lobe syndrome" in a patient with no evidence of frontal cortex disease, it is advisable to use the diagnostic term *frontal network syndrome*, with the understanding that the responsible lesions can lie anywhere within this distributed network. A patient with frontal lobe disease raises potential dilemmas in differential diagnosis, especially if the cause is neurodegenerative: the abulia and blandness may be misinterpreted as depression, and the disinhibition as idiopathic mania or acting out.

CARING FOR PATIENTS WITH DEFICITS OF HIGHER CEREBRAL FUNCTION

Spontaneous improvement of cognitive deficits following stroke or trauma is common. It is most rapid in the first few weeks but may continue for up to 2 years, especially in young individuals with single brain lesions. Some of the initial deficits in such cases appear to arise from remote dysfunction (diaschisis) in brain regions that are interconnected with the site of initial injury. Improvement in these patients may reflect, at least in part, a normalization of the remote dysfunction. Other mechanisms may involve functional reorganization in surviving neurons adjacent to the injury or the compensatory use of homologous structures. In contrast, neurodegenerative diseases show a progression of impairment but at rates that vary greatly from patient to patient.

Pharmacologic and Nonpharmacologic Interventions Some of the deficits described in this chapter are so complex that they may bewilder not only the patient and family but also the physician. The care of patients with such deficits requires a careful evaluation of the history, cognitive test results, and diagnostic procedures. Each piece of information needs to be interpreted cautiously and placed in context. A complaint of "poor memory," for example, may reflect an anomia; poor scores on a learning task may reflect a weakness of attention rather than explicit memory; a report of depression or indifference may reflect impaired prosody rather than a change in mood or empathy; jocularity may arise from poor insight rather than good mood. Treatment plans should encompass two levels: a symptomatic level that can be addressed by pharmacologic or nonpharmacologic means and a disease level that needs to be addressed through pharmacologic or molecular interventions. Although there are few well-controlled studies, several nonpharmacologic interventions have been used to treat higher cortical deficits. These include speech therapy for aphasias, behavioral modification for compartmental disorders, and cognitive training for visuospatial disorientation and amnestic syndromes. More practical interventions, usually delivered through occupational therapy, aim to improve daily living activities through assistive devices and modifications of the home environment. Determining driving competence is challenging, especially in the early stages of dementing diseases. An on-the-road driving test and reports from family members may help time decisions related to this very important activity. In neurodegenerative conditions such as PPA, transcranial magnetic (or direct current) stimulation has had mixed success in eliciting symptomatic improvement. The goal is to activate remaining neurons at sites of atrophy or in unaffected regions of the contralateral hemisphere. Depression and sleep disorders can intensify the cognitive disorders and should be treated with appropriate modalities. If neuroleptics become necessary for the control of agitation, atypical neuroleptics are preferable because of their lower extrapyramidal side effects. Treatment with neuroleptics in elderly patients with dementia requires weighing the potential benefits against the potentially serious side effects. This is especially relevant to the case of patients with Lewy body dementia, who can be unusually sensitive to side effects.

As in all other branches of medicine, a crucial step in patient care is to identify the underlying cause of the impairment. This is easily done in cases of CVA, head trauma, or encephalitis but becomes particularly challenging in neurodegenerations because the same progressive clinical syndrome can be caused by one of several neuropathologic entities. The advent of imaging, blood, and cerebrospinal fluid biomarkers now makes it possible to address this question with reasonable success and to make specific diagnoses of AD, LBD, CJD, and FTLD. A specific etiologic diagnosis allows the physician to recommend medications or clinical trials that are the most appropriate for the underlying disease process. A clinical assessment that identifies the principal domain of behavioral and cognitive impairment followed by the judicious use of biomarker information to surmise the nature of the underlying disease allows a personalized approach to patients with higher cognitive impairment.

■ FURTHER READING

Carretero RG et al: Behavioral changes as the first manifestation of a silent frontal lobe stroke. BMJ Case Rep 12:bcr-2018-227617, 2019.

Mesulam M-M: Behavioral neuroanatomy: Large-scale networks, association cortex, frontal syndromes, the limbic system and hemispheric specialization, in *Principles of Behavioral and Cognitive Neurology*, M-M Mesulam (ed). New York, Oxford University Press, 2000, pp 1–120.

Mesulam M-M et al: Frontotemporal degeneration with transactive response DNA-binding protein type C at the anterior temporal lobe. Ann Neurol 94:1, 2023.

Ricken G et al: Autoimmune global amnesia as manifestation of AMPAR encephalitis and neuropathologic findings. Neurol Neuroimmunol Neuroinflamm 8:e1019, 2021.

Singh NR, Leff AP: Advances in the rehabilitation of hemispatial inattention. Curr Neurol Neurosci Rep 23:33, 2023.

Ulugut H, Pijnenburg YAL: Frontotemporal dementia: Past, present, and future. Alzheimers Dement 19:5253, 2023.

Ulugut H et al: Right temporal variant frontotemporal dementia is pathologically heterogeneous: A case-series and systematic review. Acta Neuropathol Commun 9:131, 2021.

33

Sleep Disorders

Thomas E. Scammell, Clifford B. Saper, Charles A. Czeisler

Disturbed sleep is one of the most common health complaints that physicians encounter. More than one-half of adults in the United States experience at least intermittent sleep disturbance, and only 30% of adult Americans report consistently obtaining a sufficient amount of sleep. The National Academy of Medicine has estimated that 50–70 million Americans suffer from a chronic disorder of sleep and wakefulness, which can adversely affect daytime functioning as well as physical and mental health. A high prevalence of sleep disorders across all cultures is also now increasingly recognized, and these problems are expected to further increase in the years ahead as the global population ages. Over the past 40 years, the field of sleep medicine has emerged as a distinct specialty in response to the impact of sleep disorders and sleep deficiency on overall health. Nonetheless, >80% of patients with sleep disorders remain undiagnosed and untreated—costing the U.S. economy >$400 billion annually in increased health care costs, lost productivity, accidents, and injuries, and leading to the development of workplace-based sleep health education and sleep disorders screening programs designed to address this unmet medical need.

PHYSIOLOGY OF SLEEP AND WAKEFULNESS

Most adults need 7–9 h of sleep per night to promote optimal health, although the timing, duration, and internal structure of sleep vary among individuals. In the United States, adults tend to have one consolidated sleep episode each night, although in some cultures, sleep may be divided into a mid-afternoon nap and a shortened night sleep. This pattern changes considerably over the life span, as infants and young children sleep considerably more than older people, while individuals >70 years of age sleep on average about an hour less than young adults.

The stages of human sleep are defined on the basis of characteristic patterns in the electroencephalogram (EEG), the electrooculogram (EOG—a measure of eye-movement activity), and the surface electromyogram (EMG) measured on the chin and legs. The continuous recording of these electrophysiologic parameters to define sleep and wakefulness is termed *polysomnography*.

Polysomnographic profiles define two basic states of sleep: (1) rapid eye movement (REM) sleep and (2) non–rapid eye movement (NREM) sleep. NREM sleep is further subdivided into three stages: N1, N2, and N3, characterized by an increasing threshold for arousal and slowing of the cortical EEG. REM sleep is distinguished by a low-amplitude, mixed-frequency EEG, similar to NREM stage N1 sleep, and an EOG pattern of REMs that tend to occur in flurries or bursts. EMG activity is absent in nearly all skeletal muscles except those involved in respiration, reflecting the brainstem-mediated muscle paralysis that is characteristic of REM sleep.

ORGANIZATION OF HUMAN SLEEP

Normal nocturnal sleep in adults displays a consistent organization from night to night **(Fig. 33-1)**. After sleep onset, sleep usually progresses through NREM stages N1–N3 sleep within 45–60 min. NREM stage N3 sleep (also known as slow-wave sleep) predominates in the first third of the night and comprises 15–25% of total nocturnal sleep time in young adults. Sleep deprivation increases the rapidity of sleep onset and both the intensity and amount of slow-wave sleep.

The first REM sleep episode usually occurs in the second hour of sleep. NREM and REM sleep alternate through the night with an average period of 60–160 min (the "ultradian" sleep cycle). Overall, in a healthy young adult, REM sleep constitutes 20–25% of total sleep, and NREM stages N1 and N2 constitute 50–60%.

Age has a profound impact on sleep state organization (Fig. 33-1). N3 sleep is most intense and prominent during childhood, decreasing with puberty and across the second and third decades of life. In older adults, N3 sleep may be completely absent, and the remaining NREM sleep typically becomes more fragmented, with frequent awakenings from NREM sleep. Older people spend more time awake during their sleep episode, mainly due to more awakenings, rather than a decreased ability to fall back asleep. While REM sleep may account for 50% of total sleep time in infancy, the percentage falls off sharply over the first postnatal year as a mature REM-NREM cycle develops; in young adults, REM sleep occupies about 25% of total sleep time.

Sleep deprivation degrades cognitive performance, particularly on tests that require continual vigilance. Young adults are particularly susceptible to slowed reaction times and difficulty maintaining vigilance during sleep deprivation, which may in part account for the large number of motor vehicle accidents in this age group late at night.

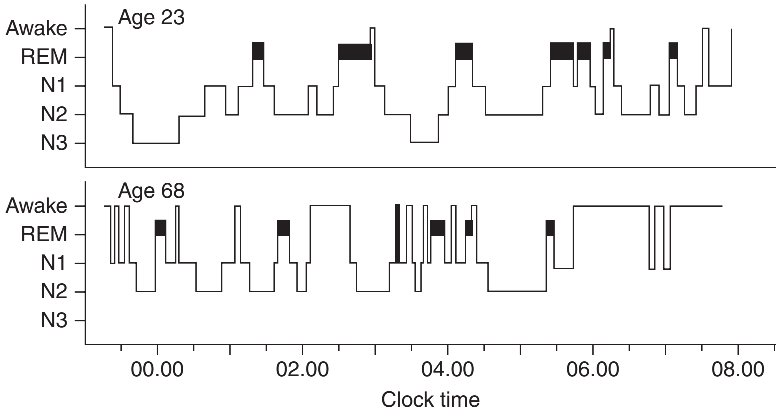

FIGURE 33-1 Wake-sleep architecture. Alternating stages of wakefulness, the three stages of non–rapid eye movement sleep (N1–N3), and rapid eye movement (REM) sleep (*solid bars*) occur over the course of the night for representative young and older adult men. Characteristic features of sleep in older people include reduction of N3 slow-wave sleep, frequent spontaneous awakenings, early sleep onset, and early morning awakening.

After sleep deprivation, NREM sleep generally recovers first, followed by REM sleep. However, because REM sleep tends to be most prominent in the second half of the night, sleep truncation (e.g., by an alarm clock) results in selective REM sleep deprivation. This may increase REM sleep pressure to the point where the first REM sleep may occur much earlier in the nightly sleep episode. To avoid these influences, it is important that the patient have sufficient sleep opportunity (at least 8 h per night) for several nights prior to a diagnostic polysomnogram.

Beyond impaired cognition, chronic sleep deficiency is associated with glucose intolerance that may contribute to the development of diabetes, obesity, and the metabolic syndrome, as well as impaired immune responses, accelerated atherosclerosis, and increased risk of cardiac disease, cognitive impairment, Alzheimer's disease, and stroke. For these reasons, the National Academy of Medicine declared sleep deficiency and sleep disorders "an unmet public health problem."

WAKE AND SLEEP ARE REGULATED BY BRAIN CIRCUITS

Two principal neural systems govern the expression of sleep and wakefulness. The ascending arousal system, illustrated in *green* in **Fig. 33-2**, consists of clusters of nerve cells extending from the upper pons to the hypothalamus and basal forebrain that activate the cerebral cortex, thalamus (which is necessary to relay sensory information to the cortex), and other forebrain regions. The ascending arousal neurons use monoamines (norepinephrine, dopamine, serotonin, and histamine), glutamate, or acetylcholine as neurotransmitters to activate their target neurons. Some basal forebrain neurons use γ-aminobutyric acid (GABA) to inhibit cortical inhibitory interneurons, thus promoting arousal. Additional wake-promoting neurons in the hypothalamus use the peptide neurotransmitter orexin (also known as hypocretin, shown in Fig. 33-2 in *blue*) to reinforce activity in the other arousal-promoting cell groups.

Damage to the arousal system at the level of the rostral pons and lower midbrain causes coma, indicating that the ascending arousal influence from this level is critical in maintaining wakefulness. Injury to the arousal system in the midbrain or hypothalamus causes profound sleepiness. Specific loss of the orexin neurons produces the sleep disorder narcolepsy (see below). Isolated damage to the thalamus causes loss of the content of wakefulness, known as a persistent vegetative state, but wake-sleep cycles are largely preserved.

The arousal system is turned off during sleep by inhibitory inputs from cell groups in the sleep-promoting system, shown in Fig. 33-2 in *red*. These neurons in the preoptic area and pons use GABA to inhibit the arousal system. Additional neurons in the lateral hypothalamus containing the peptide melanin-concentrating hormone promote REM sleep. Many sleep-promoting neurons are themselves inhibited by inputs from the arousal system. This mutual inhibition between the arousal- and sleep-promoting systems forms a neural circuit akin to what electrical engineers call a "flip-flop switch." A switch of this type tends to promote rapid transitions between the on (wake) and off (sleep) states, while avoiding intermediate states. The relatively rapid transitions between waking and sleeping states, as seen in the EEG of humans and animals, is consistent with this model.

Neurons in the ventrolateral preoptic nucleus, one of the key sleep-promoting sites, are lost during normal human aging, correlating with reduced ability to maintain sleep (sleep fragmentation). The ventrolateral preoptic neurons are also injured in Alzheimer's disease, which may in part account for the poor sleep quality in those patients.

Transitions between NREM and REM sleep appear to be governed by a similar switch in the brainstem. GABAergic REM-Off neurons have been identified in the lower midbrain that inhibit REM-On neurons in the upper pons. The REM-On group contains both GABAergic neurons that inhibit the REM-Off group (thus satisfying the conditions for a REM sleep flip-flop switch) as well as glutamatergic neurons that project widely in the central nervous system

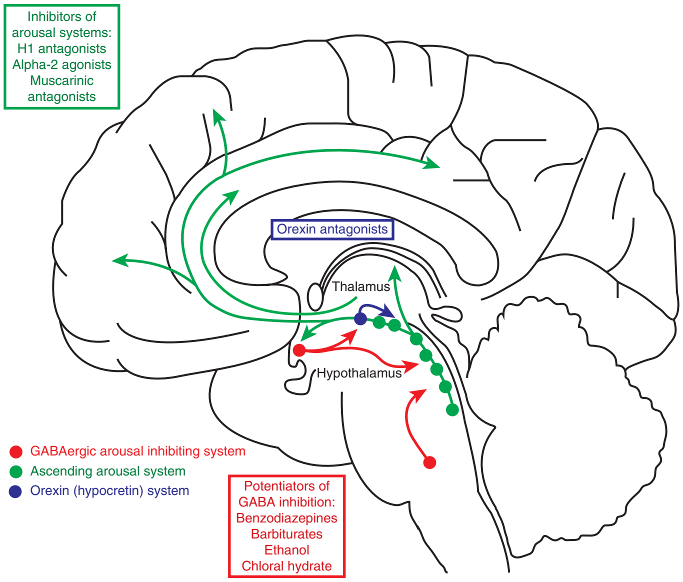
Inhibitors of arousal systems:
H1 antagonists
Alpha-2 agonists
Muscarinic antagonists
Orexin antagonists
Thalamus
Hypothalamus
● GABAergic arousal inhibiting system
● Ascending arousal system
● Orexin (hypocretin) system
Potentiators of GABA inhibition:
Benzodiazepines
Barbiturates
Ethanol
Chloral hydrate

FIGURE 33-2 Relationship of drugs for insomnia with wake-sleep systems. The arousal system in the brain (*green*) includes monoaminergic, glutamatergic, and cholinergic neurons in the brainstem that activate neurons in the hypothalamus, thalamus, basal forebrain, and cerebral cortex. Orexin neurons (*blue*) in the hypothalamus, which are lost in narcolepsy, reinforce and stabilize arousal by activating other components of the arousal system. The sleep-promoting system (*red*) consists of GABAergic neurons in the preoptic area and brainstem that inhibit the components of the arousal system, thus allowing sleep to occur. Drugs used to treat insomnia include those that block the effects of arousal system neurotransmitters (*green* and *blue*) and those that enhance the effects of γ-aminobutyric acid (GABA) produced by the sleep system (*red*).

to cause the key phenomena associated with REM sleep. REM-On neurons that project to the medulla and spinal cord activate inhibitory (GABA and glycine-containing) interneurons, which in turn hyperpolarize the motor neurons, producing the paralysis of REM sleep. REM-On neurons that project to the forebrain may be important in producing dreams.

The REM sleep switch receives cholinergic input, which favors transitions to REM sleep, and monoaminergic (norepinephrine and serotonin) input that prevents REM sleep. As a result, drugs that increase monoamine tone (e.g., serotonin or norepinephrine reuptake inhibitors) tend to reduce the amount of REM sleep. Damage to the neurons that promote REM sleep paralysis can produce REM sleep behavior disorder, a condition in which patients act out their dreams (see below).

SLEEP-WAKE CYCLES ARE DRIVEN BY HOMEOSTATIC, ALLOSTATIC, AND CIRCADIAN INPUTS

The gradual increase in sleep drive with prolonged wakefulness, followed by deeper slow-wave sleep and prolonged sleep episodes, demonstrates that there is a *homeostatic* mechanism that regulates sleep. The neurochemistry of sleep homeostasis is only partially understood, but with prolonged wakefulness, adenosine levels rise in parts of the brain. Adenosine may act through A1 receptors to directly inhibit many arousal-promoting brain regions. In addition, adenosine promotes sleep through A2a receptors; blockade of these receptors by caffeine is one of the chief ways in which people fight sleepiness. Other humoral factors, such as prostaglandin D_2, have also been implicated in this process. Both adenosine and prostaglandin D_2 activate the sleep-promoting neurons in the ventrolateral preoptic nucleus.

Allostasis is the physiologic response to a challenge such as physical danger or psychological threat that cannot be managed by homeostatic mechanisms. These stress responses can severely impact the need for and ability to sleep. For example, insomnia is very common in patients with anxiety and other psychiatric disorders. Intermittent stress-induced insomnia is even more common, affecting most people at some time in their lives. Positron emission tomography (PET) studies in patients with chronic insomnia show hyperactivation of components of the ascending arousal system, as well as their limbic system targets in the forebrain (e.g., cingulate cortex and amygdala). The limbic areas are not only targets for the arousal system, but they also send excitatory outputs back to the arousal system, which contributes to a vicious cycle of anxiety about insomnia that makes it more difficult to sleep. Approaches to treating insomnia may employ drugs that either inhibit the output of the ascending arousal system (*green* and *blue* in Fig. 33-2) or potentiate the output of the sleep-promoting system (*red* in Fig. 33-2). However, behavioral approaches (cognitive behavioral therapy [CBT] and sleep hygiene) that may reduce forebrain limbic activity at bedtime are often the best long-term treatment.

Wakefulness and sleep are also regulated by a strong *circadian* timing signal, driven by the suprachiasmatic nuclei (SCN) of the hypothalamus, as described below. The SCN sends outputs to key sites in the hypothalamus, which impose 24-h rhythms on a wide range of behaviors and body systems, including the wake-sleep cycle.

PHYSIOLOGY OF CIRCADIAN RHYTHMICITY

The wake-sleep cycle is the most evident of many 24-h rhythms in humans. Prominent daily variations also occur in endocrine, thermoregulatory, cardiac, pulmonary, renal, immune, gastrointestinal, and neurobehavioral functions. In evaluating daily rhythms in humans, it is important to distinguish between diurnal components passively evoked by periodic environmental or behavioral changes (e.g., the increase in blood pressure and heart rate that occurs upon assumption of the upright posture) and circadian rhythms actively driven by an endogenous oscillatory process (e.g., the circadian variations in adrenal cortisol and pineal melatonin secretion that persist across a variety of environmental and behavioral conditions).

At the cellular level, endogenous circadian rhythmicity is driven by self-sustaining molecular genetic feedback loops. These clock gene feedback loops of approximately 24 h duration are found in most if not all cells in the body and regulate diverse physiologic processes. However, when cells in most tissues are placed in isolation, they soon fall out of synchrony with each other and can no longer produce useful 24-h rhythms of tissue function. The only tissue that maintains this rhythm in isolation is the SCN, whose neurons are interconnected with one another in such a way as to produce a near-24-h synchronous rhythm of neural activity even in prolonged slice culture. SCN neurons are located just above the optic chiasm in the hypothalamus, from which they receive visual input to synchronize them with the external world, and they have outputs to transmit circadian timing signals to the rest of the body. Bilateral destruction of the SCN results in a loss of most endogenous circadian rhythms including wake-sleep behavior and rhythms in endocrine and metabolic systems. The genetically determined period of this endogenous neural oscillator, which averages ~24.15 h in humans, is normally synchronized to the 24-h period of the environmental light-dark cycle through direct input to the SCN from intrinsically photosensitive ganglion cells in the retina. Humans are exquisitely sensitive to the resetting effects of light, particularly the shorter wavelengths (~460–500 nm) in the blue part of the visible spectrum. Small differences in circadian period contribute to variations in diurnal preference. For example, people with short circadian cycles (e.g., 23.5 h) due to mutations of circadian clock genes prefer an early bedtime and wake up in the early morning hours (known as advanced sleep phase disorder).

The timing and internal architecture of sleep are directly coupled to the output of the endogenous circadian pacemaker. Paradoxically, the endogenous circadian rhythm for wake propensity peaks just before the habitual bedtime, whereas that of sleep propensity peaks near the habitual wake time. These rhythms are thus timed to oppose the rise of homeostatic sleep tendency throughout the usual waking day and the decline of sleep propensity during the habitual sleep episode, respectively, thus promoting consolidated sleep and wakefulness. Misalignment of the endogenous circadian pacemaker with the desired wake-sleep cycle can, therefore, induce insomnia (especially difficulty initiating sleep or waking earlier than desired in the morning), decrease alertness, and impair performance, posing health problems for night-shift workers and airline travelers. In addition, mounting evidence indicates that sleep regularity may be as important as sleep duration in terms of physical and mental health outcomes.

Participants awakened from REM sleep recall vivid dream imagery >80% of the time, especially later in the night. Less vivid imagery may also be reported after NREM sleep interruptions.

Certain disorders may occur during specific sleep stages and are described below under "Parasomnias." These include sleepwalking, night terrors, and enuresis (bed wetting), which occur most commonly in children during deep (N3) NREM sleep. In contrast, REM sleep behavior disorder occurs mainly among older people who fail to maintain full paralysis during REM sleep, and often call out, thrash around, or even act out fragments of dreams.

All major physiologic systems are influenced by sleep. Blood pressure and heart rate decrease during NREM sleep, particularly during N3 sleep. During REM sleep, bursts of eye movements are associated with large variations in both blood pressure and heart rate mediated by the autonomic nervous system. Cardiac dysrhythmias may occur selectively during REM sleep. Respiratory function also changes. In comparison to relaxed wakefulness, respiratory rate becomes slower and more regular during NREM sleep (especially N3 sleep) and becomes irregular during bursts of eye movements in REM sleep. Decreases in minute ventilation during sleep are out of proportion to the decrease in metabolic rate, resulting in a slightly higher $PCO_{2.}$

Within the brain itself, neurotransmission is supported by ion gradients across the cell membranes of neurons and astrocytes. These ion flows are accompanied by increases in intracellular volume, so that during wake there is very little extracellular space in the brain. During sleep, intracellular volume is reduced, resulting in increased extracellular space, which has higher calcium and lower potassium concentrations, supporting hyperpolarization and reduced firing of neurons. This expansion of the extracellular space during sleep increases diffusion of substances that accumulate extracellularly, like β-amyloid peptide, enhancing their clearance from the brain via cerebrospinal fluid (CSF) flow. Recent evidence suggests that lack of adequate sleep may contribute to extracellular accumulation of β-amyloid peptide, a key step in the pathogenesis of Alzheimer's disease.

Endocrine function also varies with sleep. N3 sleep is associated with secretion of growth hormone in men, while sleep in general is associated with augmented secretion of prolactin in both men and women. Sleep has a complex effect on the secretion of luteinizing hormone (LH): during puberty, sleep is associated with increased LH secretion, whereas sleep in postpubertal women inhibits LH secretion in the early follicular phase of the menstrual cycle. Sleep onset (and probably N3 sleep) is associated with inhibition of thyroid-stimulating hormone and of the adrenocorticotropic hormone–cortisol axis, an effect that is superimposed on the prominent circadian rhythms in the two systems.

The hormone melatonin is secreted from the pineal gland predominantly at night in both day- and night-active species, under the control of the SCN. Melatonin secretion does not require sleep, but melatonin secretion is inhibited by ambient light, an effect mediated by the neural connection from the retina to the SCN and then on to the pineal gland via the sympathetic nervous system. In humans, sleep efficiency is highest when sleep coincides with endogenous melatonin secretion. When endogenous melatonin levels are low, such as during the biological day or at the desired bedtime in people with delayed sleep-wake phase disorder (DSWPD), administration of exogenous melatonin can hasten sleep onset and increase sleep efficiency, but it does not increase sleep efficiency if administered when endogenous melatonin levels are elevated. This may explain why melatonin is often ineffective in the treatment of patients with primary insomnia. On the other hand, patients with sympathetic denervation of the pineal gland, such as occurs in cervical spinal cord injury or in patients with Parkinson's disease, often have low melatonin levels, and administration of melatonin (3 mg 30 min before bedtime) may help them sleep.

Sleep is accompanied by alterations of thermoregulatory function. Warming the skin is associated with an increase in the firing of warm-responsive neurons in the preoptic area, which cause a fall in body temperature and promote onset of NREM sleep. REM sleep is associated with reduced thermoregulatory responsiveness.

DISORDERS OF SLEEP AND WAKEFULNESS

APPROACH TO THE PATIENT

Sleep Disorders

Patients may seek help from a physician because of: (1) sleepiness or tiredness during the day; (2) difficulty initiating or maintaining sleep at night (insomnia); or (3) unusual behaviors during sleep itself (parasomnias).

Obtaining a careful history is essential. In particular, the duration, severity, and consistency of the symptoms are important, along with the patient's estimate of the consequences of the sleep disorder on waking function. Information from a bed partner or family member is often helpful because some patients may be unaware of symptoms such as heavy snoring or may underreport symptoms such as falling asleep at work or while driving. Physicians should inquire about when the patient typically goes to bed, when they fall asleep and wake up, whether they awaken during sleep, whether they feel rested in the morning, and whether they nap during the day. Depending on the primary complaint, it may be useful to ask about snoring, witnessed apneas, restless sensations in the legs, movements during sleep, depression, anxiety, and behaviors around the sleep episode. The physical examination may provide evidence of a small airway, large tonsils, or a neurologic or medical disorder that contributes to the main complaint.

It is important to remember that, rarely, seizures may occur exclusively during sleep, mimicking a primary sleep disorder; such sleep-related seizures typically occur during episodes of NREM sleep and may take the form of generalized tonic-clonic movements (sometimes with urinary incontinence or tongue biting) or stereotyped movements in partial complex epilepsy **(Chap. 436)**.

It is often helpful for the patient to complete a daily sleep log for 1–2 weeks to define the timing and amounts of sleep. When relevant, the log can also include information on levels of alertness, work times, and drug and alcohol use, including caffeine and hypnotics.

Polysomnography is necessary for the diagnosis of several disorders such as sleep apnea, narcolepsy, and periodic limb movement disorder (PLMD). A conventional polysomnogram performed in a clinical sleep laboratory allows measurement of sleep stages, respiratory effort and airflow, oxygen saturation, limb movements, heart rhythm, and additional parameters. A home sleep test usually focuses on just respiratory measures and is helpful in patients with a moderate to high likelihood of having obstructive sleep apnea. The multiple sleep latency test (MSLT) is used to measure a patient's propensity to sleep during the day and can provide crucial evidence for diagnosing narcolepsy and some other causes of sleepiness. The maintenance of wakefulness test is used to measure a patient's ability to sustain wakefulness during the daytime and can provide important evidence for evaluating the efficacy of therapies for improving sleepiness in conditions such as narcolepsy and obstructive sleep apnea.

EVALUATION OF DAYTIME SLEEPINESS

Up to 25% of the adult population has persistent daytime sleepiness that impairs an individual's ability to perform optimally in school, at work, while driving, and in other conditions that require alertness. Sleepy students often have trouble staying alert and performing well in school, and sleepy adults struggle to stay awake and focused on their work. More than half of Americans have fallen asleep while driving. An estimated 1.2 million motor vehicle crashes per year are due to drowsy drivers, causing about 20% of all serious crash injuries and deaths. One need not fall asleep to have a motor vehicle crash, as the inattention and slowed responses of drowsy drivers are major contributors.

Twenty-four hours of continuous wakefulness impairs reaction time as much as a blood alcohol concentration of 0.10 g/dL (which is legally drunk in all 50 states).

Identifying and quantifying sleepiness can be challenging. First, patients may describe themselves as "sleepy," "fatigued," or "tired," and the meanings of these words may differ between patients. For clinical purposes, it is best to use the term "sleepiness" to describe a propensity to fall asleep, whereas "fatigue" is best used to describe a feeling of low physical or mental energy but without a tendency to actually sleep. Sleepiness is usually most evident when the patient is sedentary, whereas fatigue may interfere with more active pursuits. Sleepiness generally occurs with disorders that reduce the quality or quantity of sleep or that interfere with the neural mechanisms of arousal, whereas fatigue is more common in inflammatory disorders such as cancer, multiple sclerosis **(Chap. 455)**, post-COVID syndrome **(Chap. 205)**, fibromyalgia **(Chap. 385)**, myalgic encephalomyelitis/chronic fatigue syndrome **(Chap. 461)**, or endocrine deficiencies such as hypothyroidism **(Chap. 395)** or Addison's disease **(Chap. 398)**. Second, sleepiness can affect judgment in a manner analogous to ethanol, such that patients may have limited insight into the condition and the extent of their functional impairment. Finally, patients may be reluctant to admit that sleepiness is a problem because they may have become unfamiliar with feeling fully alert, and because sleepiness is sometimes viewed pejoratively as reflecting poor motivation or bad sleep habits.

Table 33-1 outlines the diagnostic and therapeutic approach to the patient with a complaint of excessive daytime sleepiness.

To determine the extent and impact of sleepiness on daytime function, it is helpful to ask patients about the occurrence of sleepiness and sleep episodes during normal waking hours, both intentional and unintentional. Specific areas to be addressed include the occurrence of inadvertent sleep episodes while driving or in other safety-related settings, sleepiness while at work or school (and its impact on performance), and the effect of sleepiness on social and family life. Standardized questionnaires such as the Epworth Sleepiness Scale are often used clinically to measure sleepiness.

Eliciting a history of daytime sleepiness is usually adequate, but objective quantification is sometimes necessary. The MSLT measures a patient's propensity to sleep under quiet conditions. An overnight polysomnogram should precede the MSLT to establish that the patient has had an adequate amount of good-quality nighttime sleep. The MSLT consists of five 20-min nap opportunities every 2 h across the day. The patient is instructed to try to fall asleep, and the major endpoints are the average latency to sleep and the occurrence of REM sleep during the naps. An average sleep latency across the naps of <8 min is considered objective evidence of excessive daytime sleepiness. REM sleep normally occurs only during nighttime sleep, and the occurrence of REM sleep in two or more of the MSLT daytime naps provides support for the diagnosis of narcolepsy.

For the safety of the individual and the general public, physicians have a responsibility to help manage issues around driving in patients with sleepiness. Legal reporting requirements vary between states and countries, but at a minimum, physicians should inform sleepy patients about their increased risk of having an accident and advise such patients not to drive a motor vehicle until their sleepiness has been treated effectively. This discussion is especially important for commercial drivers, and it should be documented in the patient's medical record.

■ INSUFFICIENT SLEEP

Insufficient sleep is probably the most common cause of excessive daytime sleepiness. The average adult needs 7.5–8 h of sleep, but on weeknights, the average U.S. adult obtains only 6.75 h of sleep. Only 30% of the U.S. adult population reports consistently obtaining sufficient sleep. Insufficient sleep is especially common among shift workers, individuals working multiple jobs, people in lower socioeconomic groups, and historically minority populations. Most teenagers need ≥9 h of sleep, but many fail to get enough sleep because of circadian phase delay, plus social pressures to stay up late coupled with early school start times. Late evening light exposure, homework, television viewing, video-gaming, social media, texting, and smartphone use often delay bedtimes, despite the fixed early wake times required for work or school. As is typical with any disorder that causes sleepiness, individuals with chronically insufficient sleep may feel inattentive, irritable, unmotivated, and depressed, and have difficulty with school, work, and driving. Individuals differ in their optimal amount of sleep, and it can be helpful to ask how much sleep the patient obtains on a quiet vacation when he or she can sleep without restrictions. Some patients may think that a short amount of sleep is normal or advantageous, and they may not appreciate their biological need for more sleep, especially if coffee and other stimulants mask the sleepiness. A 2-week sleep log documenting the timing of sleep and daily level of alertness is diagnostically useful and provides helpful feedback for the patient. Extending sleep to the optimal amount on a regular basis can resolve the sleepiness and other symptoms. As with any lifestyle change, extending sleep requires commitment and adjustments, but the improvements in daytime alertness make this change worthwhile.

■ SLEEP APNEA SYNDROMES

Respiratory dysfunction during sleep is a common, serious cause of excessive daytime sleepiness as well as of disturbed nocturnal sleep. At least 24% of middle-aged men and 9% of middle-aged women in the United States have a reduction or cessation of breathing dozens or more times each night during sleep, with 9% of men and 4% of women doing so more than a hundred times per night. These episodes may be due to an occlusion of the airway (*obstructive sleep apnea*), absence of respiratory effort (*central sleep apnea*), or a combination

TABLE 33-1 Evaluation of the Patient with Excessive Daytime Sleepiness

FINDINGS ON HISTORY AND PHYSICAL EXAMINATION	DIAGNOSTIC EVALUATION	DIAGNOSIS	THERAPY
Difficulty waking in the morning, rebound sleep on weekends and vacations with improvement in sleepiness	Sleep log	Insufficient sleep	Sleep education and behavioral modification to increase amount of sleep
Obesity, snoring, hypertension	Polysomnogram or home sleep test	Obstructive sleep apnea **(Chap. 308)**	Continuous positive airway pressure; upper airway surgery (e.g., uvulopalatopharyngoplasty); dental appliance; weight loss
Cataplexy, hypnagogic hallucinations, sleep paralysis	Polysomnogram and multiple sleep latency test	Narcolepsy	Stimulants (e.g., modafinil, methylphenidate); rapid eye movement (REM) sleep-suppressing antidepressants (e.g., venlafaxine); pitolisant; solriamfetol; sodium oxybate
Restless legs, kicking movements during sleep	Assessment for predisposing medical conditions (e.g., iron deficiency or renal failure)	Restless legs syndrome with or without periodic limb movements	Treatment of predisposing condition; dopamine agonists (e.g., pramipexole, ropinirole); gabapentin; pregabalin; opiates
Sedating medications, stimulant withdrawal, head trauma, systemic inflammation, Parkinson's disease and other neurodegenerative disorders, hypothyroidism, encephalopathy	Thorough medical history and examination including detailed neurologic examination	Sleepiness due to a drug or medical condition	Change medications, treat underlying condition, consider stimulants

of these factors. Failure to recognize and treat these conditions appropriately may reduce daytime alertness and increase the risk of sleep-related motor vehicle crashes, depression, hypertension, myocardial infarction, diabetes, stroke, and mortality. Sleep apnea is particularly prevalent in overweight men and in the elderly, yet it is estimated to go undiagnosed in most affected individuals. This is unfortunate because several effective treatments are available. **Readers are referred to Chap. 308 for a comprehensive review of the diagnosis and treatment of sleep apnea.**

NARCOLEPSY

Narcolepsy is characterized by difficulty sustaining wakefulness, poor regulation of REM sleep, and disturbed nocturnal sleep. All patients with narcolepsy have excessive daytime sleepiness. This sleepiness is usually moderate to severe, and in contrast to patients with disrupted sleep (e.g., sleep apnea), people with narcolepsy usually feel well rested upon awakening and then feel tired throughout much of the day. They may fall asleep at inappropriate times, but then feel refreshed again after a nap. In addition, they often experience symptoms related to an intrusion of REM sleep characteristics into wakefulness. REM sleep is characterized by dreaming and muscle paralysis, and people with narcolepsy can have: (1) sudden muscle weakness without a loss of consciousness, which is usually triggered by strong emotions (cataplexy; **Video 33-1**); (2) dream-like hallucinations at sleep onset (hypnagogic hallucinations) or upon awakening (hypnopompic hallucinations); and (3) muscle paralysis upon awakening (sleep paralysis). With severe cataplexy, an individual may be laughing at a joke and then suddenly collapse to the ground, immobile but awake for 1–2 min. With milder episodes, patients may have partial weakness of the face or neck. Narcolepsy is one of the more common causes of chronic sleepiness and affects about 1 in 2000 people in the United States. Narcolepsy typically begins between age 10 and 20; once established, the disease persists for life.

Narcolepsy is caused by loss of the hypothalamic neurons that produce the orexin neuropeptides (also known as hypocretins). Research in mice and dogs first demonstrated that a loss of orexin signaling due to null mutations of either the orexin neuropeptides or one of the orexin receptors causes sleepiness and cataplexy nearly identical to that seen in people with narcolepsy. Although genetic mutations rarely cause human narcolepsy, researchers soon discovered that patients with narcolepsy with cataplexy (now called type 1 narcolepsy) have very low or undetectable levels of orexins in their CSF, and autopsy studies showed a nearly complete loss of the orexin-producing neurons in the hypothalamus. The orexins normally promote long episodes of wakefulness and suppress REM sleep, and thus loss of orexin signaling results in frequent intrusions of sleep during the usual waking episode, with REM sleep and elements of REM sleep at any time of day **(Fig. 33-3)**. Patients with narcolepsy but no cataplexy (type 2 narcolepsy) usually have normal orexin levels and may have partial loss of the orexin neurons or other yet uncharacterized causes of their excessive daytime sleepiness.

Extensive evidence suggests that an autoimmune process likely causes this selective loss of the orexin-producing neurons. Certain human leukocyte antigens (HLAs) can increase the risk of autoimmune disorders **(Chap. 361)**, and narcolepsy has the strongest known HLA association. HLA DQB1*06:02 is found in >90% of people with type 1 narcolepsy, whereas it occurs in only 12–25% of the general population. Researchers now hypothesize that in people with DQB1*06:02, an immune response against influenza, *Streptococcus*, or other infections may also damage the orexin-producing neurons through a process of molecular mimicry. This mechanism may account for the eight- to twelvefold increase in new cases of narcolepsy among children in Europe who received a particular brand of H1N1 influenza A vaccine (Pandemrix). In support of this hypothesis, people with type 1 narcolepsy have heightened T-cell responses against orexin peptides.

On rare occasions, narcolepsy can occur with other neurologic disorders such as anti-Ma2 paraneoplastic antibodies **(Chap. 99)**, severe traumatic brain injury, tumors, or strokes that directly damage the orexin-producing neurons in the hypothalamus or their projections.

Diagnosis Narcolepsy is most commonly diagnosed by relatively abrupt onset in a previously healthy individual of chronic sleepiness plus cataplexy or other symptoms. Narcoleptic patients report an overwhelming desire to sleep, and often awake refreshed after a brief nap. Cataplexy is distinguished from many disorders that can cause feelings of weakness by sudden onset of postural weakness (e.g., slurred speech, dropping a cup, slumping into a chair) that is often triggered by strong emotions such as laughing at a joke, happy surprise at unexpectedly seeing a friend, or intense anger. Cataplexy occurs in about half of all narcolepsy patients, who are distinguished as narcolepsy type 1; when it occurs, cataplexy is diagnostically very helpful because it occurs in almost no other disorder. In contrast, hypnagogic hallucinations and sleep paralysis occur in both type 1 and type 2 narcolepsy patients, and occasionally in about 20% of the general population, and so are not as diagnostically specific.

When narcolepsy is suspected, the diagnosis should be firmly established with a polysomnogram followed the next day by an MSLT. The polysomnogram helps rule out other causes of daytime sleepiness such as sleep apnea and establishes that the patient had adequate sleep the night before, and the MSLT provides essential, objective evidence of sleepiness plus REM sleep dysregulation. Across the five naps of the MSLT, most patients with narcolepsy will fall asleep in <8 min on average, and they will have episodes of REM sleep in at least two of the naps. Abnormal regulation of REM sleep is also manifested by the appearance of REM sleep within 15 min of sleep onset at night, which is rare in healthy individuals sleeping at their habitual bedtime. Stimulants should be stopped 1 week before the MSLT, and antidepressants should be stopped 3 weeks prior, because these medications can suppress REM sleep. In addition, patients should be encouraged to obtain a fully adequate amount of sleep each night for the week prior to the test to eliminate any effects of insufficient sleep.

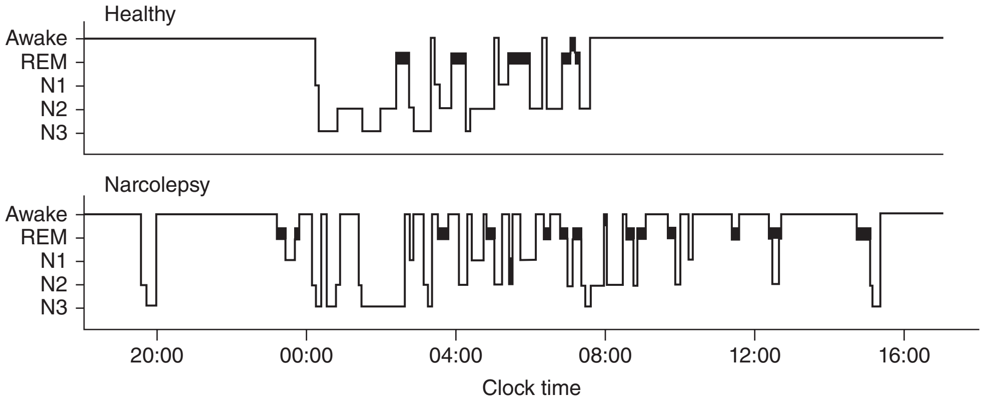

FIGURE 33-3 Polysomnographic recordings of a healthy individual and a patient with narcolepsy. The healthy individual has a long period or NREM sleep before entering REM sleep, but the individual with narcolepsy enters rapid eye movement (REM) sleep quickly at night and has moderately fragmented sleep. During the day, the healthy participant stays awake from 8:00 A.M. until midnight, but the patient with narcolepsy dozes off frequently, with many daytime naps that include REM sleep.

TREATMENT

Narcolepsy

The treatment of narcolepsy is symptomatic. Most patients with narcolepsy feel more alert after sleep, and they should be encouraged to get adequate sleep each night and to take a 15- to 20-min nap in the afternoon. This nap may be sufficient for occasional patients with mild narcolepsy, but most also require treatment with wake-promoting medications. Modafinil is often used because it has fewer side effects than amphetamines and a relatively long half-life; for most patients, 200–400 mg each morning is very effective. Methylphenidate (10–20 mg bid) and dextroamphetamine (10 mg bid) are also effective, but sympathomimetic side effects, anxiety, and the potential for abuse can be concerns. These medications are available in slow-release formulations, extending their duration of action and allowing easier dosing. Solriamfetol, a norepinephrine–dopamine reuptake inhibitor (75–150 mg daily), and pitolisant, a selective histamine 3 (H_3) receptor antagonist (8.9–35.6 mg daily), also improve sleepiness and have relatively few side effects.

Cataplexy is usually much improved with antidepressants that increase noradrenergic or serotonergic tone because these neurotransmitters strongly suppress REM sleep and cataplexy. Venlafaxine (37.5–150 mg each morning) and fluoxetine (10–40 mg each morning) are often quite effective. The tricyclic antidepressants, such as protriptyline (10–40 mg/d) or clomipramine (25–50 mg/d), are potent suppressors of cataplexy, but their anticholinergic effects, including sedation and dry mouth, make them less attractive.[1]

People with narcolepsy often have fragmented sleep at night, and sodium oxybate (gamma hydroxybutyrate), typically given at bedtime and 3–4 h later, promotes more continuous slow wave sleep. Oxybates are also available in low-sodium and once-nightly versions. Oxybates are often very valuable in improving alertness and reducing cataplexy during the day, but at too high a dosage, they can produce excessive sedation, nausea, and confusion.

---

[1] No antidepressant has been approved by the U.S. Food and Drug Administration (FDA) for treating narcolepsy.

EVALUATION OF INSOMNIA

Insomnia is the complaint of poor sleep and usually presents as difficulty initiating and/or maintaining sleep. People with insomnia are dissatisfied with their sleep and feel that it impairs their ability to function well in work, school, and social situations. Affected individuals often experience fatigue, decreased mood, irritability, malaise, and cognitive impairment.

Chronic insomnia, lasting >3 months, occurs in about 10% of adults and is more common in women, older adults, people of lower socioeconomic status, and individuals with medical, psychiatric, and substance abuse disorders. Acute or short-term insomnia affects >30% of adults and is often precipitated by stressful life events such as a major illness or loss, change of occupation, medications, and substance abuse. If the acute insomnia triggers maladaptive behaviors such as increased nocturnal light exposure, frequently checking the clock, or attempting to sleep more by napping, it can lead to chronic insomnia.

Insomnia typically begins in adulthood, but many patients may be predisposed and report easily disturbed sleep predating the insomnia, suggesting that their sleep is lighter than usual. Clinical studies and animal models indicate that insomnia is associated with activation during sleep of brain areas normally active only during wakefulness. The polysomnogram is rarely used in the evaluation of insomnia, as it typically confirms the patient's subjective report of long latency to sleep and numerous awakenings but usually adds little new information. Many patients with insomnia have more fast (beta) activity in the EEG during sleep; this fast activity is normally present only during wakefulness, which may explain why some patients report feeling awake for much of the night. The MSLT is rarely used in the evaluation of insomnia because, despite their feelings of low energy, most people with insomnia do not easily fall asleep during the day, and on the MSLT, their average sleep latencies are usually longer than normal.

Many factors can contribute to insomnia, and obtaining a careful history is essential so one can select therapies targeting the underlying factors. The assessment should focus on identifying predisposing, precipitating, and perpetuating factors.

Psychophysiological Factors Many patients with insomnia have negative expectations and conditioned arousal that interfere with sleep. These individuals may worry about their insomnia during the day and have increasing anxiety as bedtime approaches if they anticipate a poor night of sleep. While attempting to sleep, they may frequently check the clock, which only heightens anxiety and frustration. They may find it easier to sleep in a new environment rather than their bedroom, as it lacks the negative associations.

Inadequate Sleep Hygiene Patients with insomnia sometimes develop counterproductive behaviors that contribute to their insomnia. These can include daytime napping that reduces sleep drive at night; an irregular sleep-wake schedule that disrupts their circadian rhythms; use of wake-promoting substances (e.g., caffeine, tobacco) too close to bedtime; engaging in alerting or stressful activities close to bedtime (e.g., arguing with a partner, work-related emailing and texting while in bed, sleeping with a smartphone or tablet at the bedside); and routinely using the bedroom for activities other than sleep or sex (e.g., email, television, work), so the bedroom becomes associated with arousing or stressful feelings.

Psychiatric Conditions About 80% of patients with psychiatric disorders have sleep complaints, and about half of all chronic insomnia occurs in association with a psychiatric disorder **(Chap. 463)**. Depression is classically associated with early morning awakening, but it can also interfere with the onset and maintenance of sleep. Mania and hypomania can disrupt sleep and often are associated with substantial reductions in the total amount of sleep. Anxiety disorders can lead to racing thoughts and rumination that interfere with sleep and can be very problematic if the patient's mind becomes active midway through the night. Panic attacks can arise from sleep and need to be distinguished from other parasomnias. Insomnia is common in schizophrenia and other psychoses, often resulting in fragmented sleep, less deep NREM sleep, and sometimes reversal of the day-night sleep pattern.

Medications and Drugs of Abuse A wide variety of psychoactive drugs can interfere with sleep. Caffeine, which has a half-life of 6–9 h, can disrupt sleep for up to 8–14 h, depending on the dose, variations in metabolism, and an individual's caffeine sensitivity. Insomnia can also result from use of prescription medications too close to bedtime (e.g., antidepressants, stimulants, glucocorticoids, theophylline). Conversely, withdrawal of sedating medications such as alcohol, narcotics, or benzodiazepines can cause insomnia. Alcohol consumed just before bed can shorten sleep latency, but it often produces rebound insomnia 2–3 h later as it wears off. This same problem with sleep maintenance can occur with short-acting medications such as alprazolam or zolpidem.

Medical Conditions A large number of medical conditions disrupt sleep. Pain from rheumatologic disorders or a painful neuropathy commonly disrupts sleep. Some patients may sleep poorly because of respiratory conditions such as asthma, chronic obstructive pulmonary disease, cystic fibrosis, congestive heart failure, or restrictive lung disease, and some of these disorders are worse at night due to circadian variations in airway resistance and postural changes in bed that can result in nocturnal dyspnea. Obstructive sleep apnea is an example of a respiratory disorder that only becomes a problem with loss of airway muscle tone during sleep. Many women experience poor sleep with the hormonal changes of menopause. Gastroesophageal reflux is also a common cause of difficulty sleeping. Approximately 60% of patients with long COVID report symptoms of insomnia.

Neurologic Disorders Dementia **(Chap. 31)** is often associated with poor sleep, probably due to a variety of factors, including napping during the day, altered circadian rhythms, and perhaps a weakened output of the brain's sleep-promoting mechanisms. In fact, insomnia and nighttime wandering are some of the most common causes for institutionalization of patients with dementia, because they place a large burden on caregivers. Conversely, in cognitively intact elderly men, fragmented sleep and poor sleep quality are associated with subsequent cognitive decline. Patients with Parkinson's disease may sleep poorly due to rigidity, dementia, urinary frequency, REM sleep behavior disorder, restless legs syndrome, and other factors. Fatal familial insomnia is a very rare neurodegenerative condition caused by mutations in the prion protein gene **(Chap. 449)**, and although insomnia is a common early symptom, most patients present with other obvious neurologic signs such as dementia, myoclonus, dysarthria, or autonomic dysfunction.

TREATMENT

Insomnia

Treatment of insomnia improves quality of life and can promote long-term health. With improved sleep, patients often report less daytime fatigue, improved cognition, and more energy. Treating the insomnia can also improve comorbid disease. For example, management of insomnia at the time of diagnosis of major depression often improves the response to antidepressants and reduces the risk of relapse. Sleep loss can heighten the perception of pain, so a similar approach is warranted in acute and chronic pain management. Gabapentin (100–300 mg before bedtime) can often reduce the pain and improve sleep in patients with chronic pain.

The treatment plan should target all putative contributing factors: establish good sleep hygiene, treat medical disorders, and use behavioral therapies for anxiety and negative conditioning and pharmacotherapy and/or psychotherapy for psychiatric disorders. Behavioral therapies should be the first-line treatment, followed by judicious use of sleep-promoting medications if needed.

TREATMENT OF MEDICAL AND PSYCHIATRIC DISEASE

If the history suggests that a medical or psychiatric disease contributes to the insomnia, then it should be addressed by, for example, treating the pain or depression, improving breathing, and switching or adjusting the timing of medications.

IMPROVE SLEEP HYGIENE

Attention should be paid to improving sleep hygiene and avoiding counterproductive, arousing behaviors before bedtime. Patients should establish a regular bedtime and wake time, even on weekends, to help synchronize their circadian rhythms and sleep patterns. The amount of time allocated for sleep should not be more than their actual total amount of sleep. In the 30 min before bedtime, patients should establish a relaxing "wind-down" routine that can include a warm bath, listening to music, meditation, or other relaxation techniques. The bedroom should be off-limits to computers, televisions, radios, smartphones, videogames, and tablets. If an e-reader is used, the light should be adjusted for evening use (dimmer and reduced blue light) if possible, because light itself, especially in the blue spectrum, suppresses melatonin secretion and is arousing. Once in bed, patients should try to avoid thinking about anything stressful or arousing such as problems with relationships or work. If they cannot fall asleep within 20 min, it often helps to get out of bed and read or listen to relaxing music in dim light as a form of distraction from any anxiety, but "blue light," especially from a cell phone, computer, or television, should be avoided.

Table 33-2 outlines some of the key aspects of good sleep hygiene to improve insomnia.

TABLE 33-2 Methods to Improve Sleep Hygiene in Insomnia Patients

HELPFUL BEHAVIORS	BEHAVIORS TO AVOID
Use the bed only for sleep and sex • If you cannot sleep within 20 min, get out of bed and read or do other relaxing activities in dim light before returning to bed	Avoid behaviors that interfere with sleep physiology, including: • Napping, especially after 3:00 PM • Attempting to sleep too early • Caffeine after lunchtime
Make quality sleep a priority • Go to bed and get up at the same time each day • Ensure a restful environment (comfortable bed, bedroom quiet and dark)	In the 2–3 h before bedtime, avoid: • Heavy eating • Smoking or alcohol • Vigorous exercise
Develop a consistent bedtime routine. For example: • Prepare for sleep with 20–30 min of relaxation (e.g., soft music, meditation, yoga, pleasant reading) • Take a warm bath	When trying to fall asleep, avoid: • Solving problems • Thinking about life issues • Reviewing events of the day

COGNITIVE BEHAVIORAL THERAPY

Cognitive behavioral therapy (CBT) uses a combination of the techniques above plus additional methods to improve insomnia. A trained therapist may use cognitive psychology techniques to reduce excessive worrying about sleep and to reframe faulty beliefs about the insomnia and its daytime consequences. The therapist may also teach the patient relaxation techniques, such as progressive muscle relaxation or meditation, to reduce autonomic arousal, intrusive thoughts, and anxiety. While sleep restriction may improve sleep continuity, chronic exposure to sleep restriction may have adverse effects on daytime performance.

MEDICATIONS FOR INSOMNIA

If insomnia persists after treatment of these contributing factors, pharmacotherapy is often used on a nightly or intermittent basis. A variety of sedatives can improve sleep.

Antihistamines, such as diphenhydramine, are the primary active ingredient in most over-the-counter sleep aids. These may be of benefit when used intermittently but can produce tolerance and anticholinergic side effects such as dry mouth and constipation, which limit their use, particularly in the elderly.

Benzodiazepine receptor agonists (BzRAs) are an effective and well-tolerated class of medications for insomnia **(Chap. 463)**. BzRAs bind to the GABA$_A$ receptor and potentiate the postsynaptic response to GABA. GABA$_A$ receptors are found throughout the brain, and BzRAs may globally reduce neural activity and enhance the activity of specific sleep-promoting GABAergic pathways. Classic BzRAs include lorazepam, triazolam, and clonazepam, whereas newer agents such as zolpidem and zaleplon have more selective affinity for the α_1 subunit of the GABA$_A$ receptor.

Specific BzRAs are often chosen based on the desired duration of action. The most commonly prescribed agents in this family are zaleplon (5–20 mg), with a half-life of 1–2 h; zolpidem (5–10 mg) and triazolam (0.125–0.25 mg), with half-lives of 2–4 h; eszopiclone (1–3 mg), with a half-life of 5–8 h; and temazepam (15–30 mg), with a half-life of 8–20 h. Generally, side effects are uncommon when the dose is kept low and the serum concentration is minimized during the waking hours (by using the shortest-acting effective agent). For chronic insomnia, intermittent use is recommended, unless the consequences of untreated insomnia outweigh concerns regarding chronic use.

The heterocyclic antidepressants (trazodone, amitriptyline,[2] and doxepin) are the most commonly prescribed alternatives to BzRAs due to their lack of abuse potential and low cost **(Chap. 463)**. Trazodone (25–100 mg) is used more commonly than the tricyclic antidepressants, because it has a much shorter half-life (5–9 h) and less anticholinergic activity.

The orexin receptor antagonists suvorexant (10–20 mg), lemborexant (5–10 mg), and daridorexant (25–50 mg) can also improve insomnia by blocking the wake-promoting effects of the orexin neuropeptides. These have medium to long half-lives and can produce morning sedation, and as they reduce orexin signaling, they can rarely produce hypnagogic hallucinations and sleep paralysis (see narcolepsy section above).

Medications for insomnia are now among the most commonly prescribed medications, but they should be used cautiously. All sedatives increase the risk of injurious falls and confusion in the elderly, and therefore, if needed, these medications should be used at the lowest effective dose. Morning sedation can interfere with driving and judgment, so when selecting a medication, one should consider the duration of action. Benzodiazepines carry a risk of addiction and abuse, especially in patients with a history of alcohol or sedative abuse. In patients with depression, all sedatives can worsen the depression. Like alcohol, some sleep-promoting medications can worsen sleep apnea. Sedatives can also produce complex behaviors during sleep, such as sleepwalking and sleep eating, especially at higher doses.

---

[2] Trazodone and amitriptyline have not been approved by the FDA for treating insomnia.

RESTLESS LEGS SYNDROME

Patients with restless legs syndrome (RLS) report an irresistible urge to move the legs. Many patients report a creepy-crawly, tingly, or unpleasant deep ache within the thighs or calves, and those with more severe RLS may have discomfort in the arms as well. For most patients with RLS, these dysesthesias and restlessness are much worse in the evening and first half of the night. The symptoms appear with inactivity and can make sitting still when traveling or watching a movie a miserable experience. The sensations are temporarily relieved by movement, stretching, or massage. This nocturnal discomfort usually interferes with sleep, and patients may report daytime sleepiness as a consequence. RLS is very common, affecting 5–10% of adults, and is more common in women and older adults.

A variety of factors can cause RLS. Iron deficiency is the most common treatable cause, and iron replacement should be considered if the ferritin level is <75 ng/mL. RLS can also occur with peripheral neuropathies and uremia and can be worsened by pregnancy, caffeine, alcohol, antidepressants, lithium, neuroleptics, and antihistamines. Genetic factors contribute to RLS, and polymorphisms in a variety of genes (*BTBD9, MEIS1, MAP2K5/LBXCOR*, and *PTPRD*) have been linked to RLS, although as yet, the mechanism through which they cause RLS remains unknown. Roughly one-third of patients (particularly those with an early age of onset) have multiple affected family members.

RLS is treated by addressing the underlying cause such as iron deficiency if present. Otherwise, treatment is symptomatic, and alpha-2-delta calcium channel ligands or dopamine agonists are used most frequently. Alpha-2-delta calcium channel ligands such as gabapentin (300–900 mg q7PM), gabapentin enacarbil (300–600 mg q5PM), and pregabalin (150–450 mg q7PM) are quite effective, and because they are also sedating and analgesic, they can be especially helpful in patients with concomitant pain, neuropathy, or anxiety. Agonists of dopamine $D_{2/3}$ receptors such as pramipexole (0.25–0.5 mg q7PM) or ropinirole (0.5–4 mg q7PM) are usually quite effective, but about 25% of patients taking dopamine agonists develop augmentation, a worsening of RLS such that symptoms begin earlier in the day and can spread to other body regions. Other possible side effects of dopamine agonists include nausea, morning sedation, and increases in rewarding behaviors such as sex and gambling. Opioids and benzodiazepines may also be of therapeutic value in refractory patients. Most patients with restless legs also experience PLMD, although the reverse is not the case.

PERIODIC LIMB MOVEMENT DISORDER

PLMD involves rhythmic twitches of the legs that disrupt sleep. The movements resemble a triple flexion reflex with extensions of the great toe and dorsiflexion of the foot for 0.5–5.0 s, which recur every 20–40 s during NREM sleep, in episodes lasting from minutes to hours. PLMD is diagnosed by a polysomnogram that includes recordings of the anterior tibialis and sometimes additional muscles. The EEG shows that the movements of PLMD frequently cause brief arousals that disrupt sleep, sometimes resulting in insomnia and daytime sleepiness. PLMD can be caused by the same factors that cause RLS (see above), and the frequency of leg movements improves with the same medications used for RLS, including dopamine agonists. Genetic studies identified polymorphisms associated with both RLS and PLMD, suggesting that they may have a common pathophysiology.

PARASOMNIAS

Parasomnias are abnormal behaviors or experiences that arise from or occur during sleep. A variety of parasomnias can occur during NREM sleep, from brief confusional arousals to sleepwalking and night terrors. The presenting complaint is usually related to the behavior itself, but the parasomnias can disturb sleep continuity or lead to mild impairments in daytime alertness. Two main parasomnias occur in REM sleep: REM sleep behavior disorder (RBD) and nightmares.

Sleepwalking (Somnambulism) Patients affected by this disorder carry out automatic motor activities that range from simple to complex. Individuals may walk, urinate inappropriately, eat, exit the house, or drive a car with minimal awareness. It may be difficult to arouse the patient to wakefulness, and some individuals may respond to attempted awakening with agitation or violence. In general, it is safest to lead the patient back to bed, at which point they will often fall back asleep. Sleepwalking arises from NREM stage N3 sleep, usually in the first few hours of the night, and the EEG initially shows the slow cortical activity of deep NREM sleep even when the patient is moving about. Sleepwalking is most common in children and adolescents, when deep NREM sleep is most abundant. About 15% of children have occasional sleepwalking, and it persists in about 1% of adults. Episodes are usually isolated but may be recurrent in 1–6% of patients. The cause is unknown, although it has a familial basis in roughly one-third of cases. Sleepwalking can be worsened by stress, alcohol, and insufficient sleep, which subsequently causes an increase in deep NREM sleep. These should be addressed if present. Small studies have shown some efficacy of antidepressants and benzodiazepines; relaxation techniques and hypnosis can also be helpful. Patients and their families should improve home safety (e.g., replace glass doors, remove low tables to avoid tripping) to minimize the chance of injury if sleepwalking occurs.

Sleep Terrors This disorder occurs primarily in young children during the first few hours of sleep during NREM stage N3 sleep. The child often sits up during sleep and screams, exhibiting autonomic arousal with sweating, tachycardia, large pupils, and hyperventilation. The individual may be difficult to arouse and rarely recalls the episode on awakening in the morning. Treatment usually consists of reassuring parents that the condition is self-limited and benign, and like sleepwalking, it may improve by avoiding insufficient sleep.

Sleep Enuresis Bedwetting, like sleepwalking and night terrors, is another parasomnia that occurs during sleep in the young. Before age 5 or 6 years, nocturnal enuresis should be considered a normal feature of development. The condition usually improves spontaneously by puberty, persists in 1–3% of adolescents, and is rare in adulthood. Treatment consists of bladder training exercises and behavioral therapy. Symptomatic pharmacotherapy is usually accomplished in adults with desmopressin (0.2 mg qhs), oxybutynin chloride (5 mg qhs), or imipramine (10–25 mg qhs). Important causes of nocturnal enuresis in patients who were previously continent for 6–12 months include urinary tract infections or malformations, cauda equina lesions, emotional disturbances, epilepsy, sleep apnea, and certain medications.

Sleep Bruxism Bruxism is an involuntary, forceful grinding of teeth during sleep that affects 10–20% of the population. The patient is usually unaware of the problem. The typical age of onset is 17–20 years, and spontaneous remission usually occurs by age 40. In many cases, the diagnosis is made during dental examination, damage is minor, and no treatment is indicated. In more severe cases, treatment with a mouth guard is necessary to prevent tooth injury. Stress management, benzodiazepines, and biofeedback can be useful when bruxism is a manifestation of psychological stress.

REM Sleep Behavior Disorder (RBD) RBD **(Video 33-2)** is distinct from other parasomnias in that it occurs during REM sleep. The patient or the bed partner usually reports that the patient calls out and has agitated or violent behavior during sleep. If awakened during or by the event, the patient can often report a dream that matches the accompanying movements. During normal REM sleep, nearly all nonrespiratory skeletal muscles are paralyzed, but in patients with RBD, dramatic limb movements such as punching or kicking lasting seconds to minutes occur during REM sleep, and it is not uncommon for the patient or the bed partner to be injured while enacting dream behaviors.

The prevalence of RBD increases with age, afflicting about 2% of adults aged >70, and is reported more often in men. Within 12 years of disease onset, half of RBD patients develop a synucleinopathy such as Parkinson's disease **(Chap. 446)**, dementia with Lewy bodies **(Chap. 445)**, or occasionally multiple system atrophy **(Chap. 451)**, and >90% develop a synucleinopathy by 25 years. Patients with a latent synucleinopathy can be distinguished from those with other causes of symptomatic RBD by detection of alpha-synuclein aggregates in spinal fluid or in peripheral nerves in skin biopsy. RBD can occur in patients taking antidepressants, and in some, these medications may unmask this early indicator of neurodegeneration. Synucleinopathies probably cause neuronal loss in brainstem regions that regulate muscle paralysis during REM sleep, and loss of these neurons permits movements to break through during REM sleep. RBD also occurs in about 30% of people with narcolepsy, but the underlying cause is probably different, as they seem to be at no increased risk of a neurodegenerative disorder.

Many patients with RBD have sustained improvement with melatonin at doses up to 10 mg nightly. Clonazepam (0.5–2.0 mg qhs)[3] also prevents attacks, but as with all benzodiazepines, it can increase the risk of falls and confusion at night.

CIRCADIAN RHYTHM SLEEP DISORDERS

A subset of patients presenting with either insomnia or hypersomnia may have a disorder of sleep *timing* rather than sleep *generation*. Disorders of sleep timing can be either inherent (i.e., due to an abnormality of circadian pacemaker[s]) or environmental/behavioral (i.e., due to a disruption of environmental synchronizers). The therapeutic goal is to entrain the circadian rhythm of sleep propensity to the appropriate behavioral phase.

Delayed Sleep-Wake Phase Disorder DSWPD is characterized by: (1) sleep onset and wake times persistently later than desired; (2) actual sleep times at nearly the same clock hours daily; and (3) if conducted at the habitual delayed sleep time, essentially normal sleep on polysomnography (except for delayed sleep onset). About half of patients with DSWPD exhibit an abnormally delayed endogenous circadian phase, which can be assessed by measuring the onset of secretion of melatonin in either the blood or saliva; this is best done in a dimly lit environment as light suppresses melatonin secretion. In healthy people, dim-light melatonin onset (DLMO) typically occurs about 8:00–9:00 P.M. (i.e., about 1–2 h before habitual bedtime), but in DSWPD patients, DLMO occurs later in the evening than normal, which helps distinguish DSWPD from other forms of sleep-onset insomnia. Patients tend to be young adults. The delayed circadian phase could be due to: (1) an abnormally long, genetically determined intrinsic period of the endogenous circadian pacemaker; (2) reduced phase-advancing capacity of the pacemaker; (3) slower buildup of homeostatic sleep drive during wakefulness; or (4) an irregular prior sleep-wake schedule, characterized by frequent nights when the patient chooses to remain awake while exposed to artificial light well past midnight (for personal, social, school, or work reasons). In most cases, it is difficult to distinguish among these factors, as both patients with either a behaviorally induced or biologically driven circadian phase delay may exhibit a similar circadian phase delay in DLMO, and both factors make it difficult to fall asleep at the desired hour. DSWPD is a chronic condition that can persist for years and may not respond to attempts to reestablish normal bedtime hours. Typical treatment is phototherapy with blue-enriched light during the morning hours and/or melatonin administration in the evening hours, although the relapse rate is high.

Advanced Sleep-Wake Phase Disorder Advanced sleep-wake phase disorder (ASWPD) is the converse of DSWPD. Most commonly, this syndrome occurs in older people, 15% of whom report that they cannot sleep past 5:00 A.M., with twice that number complaining that they wake up too early at least several times per week. Patients with ASWPD are sleepy during the evening hours, even in social settings. Sleep-wake timing in ASWPD patients can interfere with a normal social life. Patients with this circadian rhythm sleep disorder can be distinguished from those who have early wakening due to insomnia because ASWPD patients show early onset of dim-light melatonin secretion.

In addition to age-related ASWPD, an early-onset familial variant of this condition has also been reported. In two families in which ASWPD was inherited in an autosomal dominant pattern, the syndrome was due to missense mutations in a circadian clock component (in the casein kinase binding domain of *PER2* in one family, and in casein kinase I delta in the other) that shortens the circadian period. Patients with ASWPD may benefit from bright light and/or blue-enriched phototherapy during the evening hours to reset the circadian pacemaker to a later hour.

Non-24-h Sleep-Wake Rhythm Disorder Non-24-h sleep-wake rhythm disorder (N24SWD) most commonly occurs when the primary synchronizing input (i.e., the light-dark cycle) from the environment to the circadian pacemaker is lost (as occurs in many blind people with no light perception), and the maximal phase-advancing capacity of the circadian pacemaker in response to nonphotic cues cannot accommodate the difference between the 24-h geophysical day and the intrinsic period of the patient's circadian pacemaker, resulting in loss of entrainment to the 24-h day. The sleep of most blind patients with N24SWD is restricted to the nighttime hours due to social or occupational demands. Despite this regular sleep-wake schedule, affected patients with N24SWD are nonetheless unable to maintain a stable phase relationship between the output of the nonentrained circadian pacemaker and the 24-h day. Therefore, most blind patients with no light perception present with intermittent bouts of insomnia. When the blind patient's endogenous circadian rhythms are out of phase with the local environment, nighttime insomnia coexists with excessive daytime sleepiness. Conversely, when the endogenous circadian rhythms of those same patients are in phase with the local environment, symptoms remit. The interval between symptomatic phases may last several weeks to several months in blind patients with N24SWD, depending on the difference between the period of the underlying nonentrained rhythm and the 24-h day. Nightly administration of low-dose (0.5 mg) melatonin or a melatonin receptor agonist may improve sleep and, in some cases, induce synchronization of the circadian pacemaker. In sighted patients, N24SWD can be caused by self-selected exposure to artificial light that inadvertently entrains the circadian pacemaker to a >24-h schedule, and these individuals present with an incremental pattern of successive delays in sleep timing, progressing in and out of phase with local time—a clinical presentation that is seldom seen in blind patients with N24SWD.

Shift-Work Disorder More than 7 million people in the United States regularly work at night, either on a permanent or rotating schedule. Many more begin the commute to work or school between 4:00 A.M. and 7:00 A.M., requiring them to commute and then work during a time of day that they would otherwise be asleep. In addition, each week, millions of "day" workers and students elect to remain awake at night or awaken very early in the morning to work or study to meet work or school deadlines, drive long distances, compete in sporting events, or participate in recreational activities. Such schedules can result in both sleep loss and misalignment of circadian rhythms with respect to the sleep-wake cycle.

---

[3] No medications have been approved by the FDA for the treatment of RBD.

The circadian timing system usually fails to adapt successfully to the inverted schedules required by overnight work or the phase advance required by early morning (4:00 A.M. to 7:00 A.M.) start times, particularly if the shift worker reverts to a normal day-night schedule on days off. This leads to a misalignment between the desired work-rest schedule and the output of the pacemaker, resulting in disturbed daytime sleep in most such individuals. Excessive work hours (per day or per week), insufficient time off between consecutive days of work or school, and frequent travel across time zones may be contributing factors. Sleep deficiency, increased length of time awake prior to work, and misalignment of circadian phase impair alertness and performance, increase reaction time, and increase risk of performance lapses, thereby resulting in greater safety hazards among night workers and other sleep-deprived individuals. Sleep disturbance nearly doubles the risk of a fatal work accident. Similarly, the risk of sleep-related, fatal-to-the-driver highway crashes is highest in the early morning and late afternoon hours, coincident with bimodal peaks in the daily rhythm of sleep tendency. An expert consensus panel has concluded that individuals who have slept <2 h in the prior 24 h are unfit to drive a motor vehicle. In addition, long-term night-shift workers may have higher rates of breast, colorectal, and prostate cancer and of cardiac, gastrointestinal, metabolic, and reproductive disorders. The World Health Organization has added night-shift work to its list of probable carcinogens. Both circadian rhythms and sleep deficiency contribute to this risk.

Physicians who work prolonged shifts, especially intermittent overnight shifts, constitute another group of workers at greater risk for accidents and other adverse consequences of lack of sleep and misalignment of the circadian rhythm. Recurrent scheduling of resident physicians to work shifts of ≥24 consecutive hours impairs psychomotor performance to a degree that is comparable to alcohol intoxication, doubles the risk of attentional failures among intensive care unit resident physicians working at night, and increases the risk of serious medical errors in intensive care units, including a fivefold increase in the risk of serious diagnostic mistakes. Some 20% of hospital resident physicians report making a fatigue-related mistake that injured a patient, and 5% admit making a fatigue-related mistake that resulted in the death of a patient. Moreover, working for >24 consecutive hours increases the risk of percutaneous injuries and more than doubles the risk of motor vehicle crashes during the commute home. For these reasons, a National Academy of Medicine report concluded that the practice of scheduling resident physicians to work for >16 consecutive hours without sleep is hazardous for both resident physicians and their patients.

Of individuals scheduled to work at night or in the early morning hours, 5–15% have much greater-than-average difficulties remaining awake during night work and sleeping during the day; these individuals are diagnosed with chronic and severe shift-work disorder (SWD). Patients with this disorder have a level of excessive sleepiness during work at night or in the early morning and insomnia during day sleep that the physician judges to be clinically significant; the condition is associated with an increased risk of sleep-related accidents and with some of the illnesses associated with night-shift work. Patients with chronic and severe SWD are profoundly sleepy at work. In fact, their sleep latencies during night work average just 2 min, comparable to mean daytime sleep latency durations of patients with narcolepsy or severe sleep apnea.

TREATMENT

Shift-Work Disorder

Caffeine is frequently used by night workers to promote wakefulness. However, it cannot forestall sleep indefinitely, and it does not shield users from sleep-related performance lapses. Postural changes, exercise, and strategic placement of nap opportunities can sometimes temporarily reduce the risk of fatigue-related performance lapses. Properly timed exposure to blue-enriched light or bright white light can directly enhance alertness and facilitate more rapid adaptation to night-shift work.

Modafinil (200 mg) or armodafinil (150 mg) 30–60 min before the start of an 8-h overnight shift is an effective treatment for the excessive sleepiness during night work in patients with SWD. Although treatment with modafinil or armodafinil significantly improves performance and reduces sleep propensity and the risk of lapses of attention during night work, affected patients remain excessively sleepy.

Fatigue risk management programs for night-shift workers should promote education about sleep, increase awareness of the hazards associated with sleep deficiency and night work, and screen for common sleep disorders. Work schedules should be designed to minimize: (1) exposure to night work; (2) the frequency of shift rotations; (3) the number of consecutive night shifts; and (4) the duration of night shifts.

Jet Lag Disorder Each year, >60 million people fly from one time zone to another, often resulting in excessive daytime sleepiness, sleep-onset insomnia, or frequent arousals from sleep, particularly in the latter half of the night. The syndrome is transient, typically lasting 2–14 d depending on the number of time zones crossed, the direction of travel, and the traveler's age and phase-shifting capacity. Travelers who spend more time outdoors at their destination reportedly adapt more quickly than those who remain in hotel or seminar rooms, presumably due to brighter (outdoor) light exposure. Avoidance of antecedent sleep loss or napping in the afternoon prior to overnight travel can reduce the difficulties associated with extended wakefulness. Laboratory studies suggest that low doses of melatonin can enhance sleep efficiency, but only if taken when endogenous melatonin concentrations are low (i.e., during the biologic daytime).

In addition to jet lag associated with travel across time zones, many patients report a behavioral pattern that has been termed *social jet lag*, in which bedtimes and wake times on weekends or days off occur 3–4 h or more later than during the week. Such recurrent displacement of the timing of the sleep-wake cycle is common in adolescents and young adults and is associated with delayed circadian phase, sleep-onset insomnia, excessive daytime sleepiness, poorer academic performance, and increased risk of both obesity and depressive symptoms.

MEDICAL IMPLICATIONS OF CIRCADIAN RHYTHMICITY

Prominent circadian variations have been reported in the incidence of acute myocardial infarction, sudden cardiac death, and stroke, the leading causes of death in the United States. Platelet aggregability is increased in the early morning hours, coincident with the peak incidence of these cardiovascular events. Recurrent circadian disruption combined with chronic sleep deficiency, such as occurs during night-shift work, is associated with increased plasma glucose concentrations after a meal due to inadequate pancreatic insulin secretion. Night-shift workers with elevated fasting glucose have an increased risk of progressing to diabetes. Blood pressure of night workers with sleep apnea is higher than that of day workers. A better understanding of the possible role of circadian rhythmicity in the acute destabilization of a chronic condition such as atherosclerotic disease could improve the understanding of its pathophysiology.

Diagnostic and therapeutic procedures may also be affected by the time of day at which data are collected. Examples include blood pressure, body temperature, the dexamethasone suppression test, and plasma cortisol levels. The timing of administration of drugs such as chemotherapy can affect both their toxicity and effectiveness. Anesthetic agents are particularly sensitive to time-of-day effects. Finally, the physician must be aware of the public health risks associated with the ever-increasing demands made by the 24/7 schedules in our round-the-clock society.

FURTHER READING

Aquino G et al: Towards the neurobiology of insomnia: A systematic review of neuroimaging studies. Sleep Med Rev 73:101878, 2024.

Cash RE et al: Association between sleep duration and ideal cardiovascular health among US adults, National Health and Nutrition Examination Survey. Prev Chronic Dis 17:E43, 2020.

Chinoy ED et al: Unrestricted evening use of light-emitting tablet computers delays self-selected bedtime and disrupts circadian timing and alertness. Physiol Rep 6:e13692, 2018.

Cribb L et al: Sleep regularity and mortality: A prospective analysis in the UK Biobank. Elife 12:RP88359, 2023.

Holth JK et al: The sleep-wake cycle regulates brain interstitial fluid tau in mice and CSF tau in humans. Science 363:880, 2019.

Landrigan CP et al: Effect on patient safety of a resident physician schedule without 24-hour shifts. N Engl J Med 382:2514, 2020.

Lee ML et al: High risk of near-crash driving events following night-shift work. Proc Natl Acad Sci USA 113:176, 2016.

Liblau RS et al: The immunopathogenesis of narcolepsy type 1. Nat Rev Immunol 24:33, 2024.

Scammell TE: Narcolepsy. N Engl J Med 373:2654, 2015.

Scammell TE et al: Neural circuitry of wakefulness and sleep. Neuron 93:747, 2017.

Sletten TL et al: The importance of sleep regularity: A consensus statement of the National Sleep Foundation sleep timing and variability panel. Sleep Health 9:801, 2023.

VIDEO 33-1 A typical episode of severe cataplexy. The patient is joking and then falls to the ground with an abrupt loss of muscle tone. The electromyogram recordings (*four lower traces on the right*) show reductions in muscle activity during the period of paralysis. The electroencephalogram (*top two traces*) shows wakefulness throughout the episode. *(Video courtesy of Giuseppe Plazzi, University of Bologna.)*

VIDEO 33-2 Typical aggressive movements in rapid eye movement (REM) sleep behavior disorder. *(Video courtesy of Dr. Carlos Schenck, University of Minnesota Medical School.)*

Section 4 Disorders of Eyes, Ears, Nose, and Throat

34

Disorders of the Eye

Jonathan C. Horton

THE HUMAN VISUAL SYSTEM

The visual system provides a supremely efficient means for the rapid assimilation of information from the environment to aid in the guidance of behavior. The act of seeing begins with the capture of images focused by the cornea and lens on a light-sensitive membrane in the back of the eye called the *retina*. The retina is actually part of the brain, banished to the periphery to serve as a transducer for the conversion of patterns of light energy into neuronal signals. Light is absorbed by pigment in two types of photoreceptors: rods and cones. In the human retina, there are 100 million rods and 5 million cones. The rods operate in dim (scotopic) illumination. The cones function under daylight (photopic) conditions. The cone system is specialized for color perception and high spatial resolution. The majority of cones are within the macula, the portion of the retina that serves the central 10° of vision. In the middle of the macula, a small pit termed the *fovea*, packed exclusively with cones, provides the best visual acuity.

Photoreceptors hyperpolarize in response to light, activating bipolar, amacrine, and horizontal cells in the inner nuclear layer. After processing of photoreceptor responses by this complex retinal circuit, the flow of sensory information ultimately converges on a final common pathway: the ganglion cells. These cells translate the visual image impinging on the retina into a continuously varying barrage of action potentials that propagates along the primary optic pathway to visual centers within the brain. There are a million ganglion cells in each retina and hence a million fibers in each optic nerve.

Ganglion cell axons sweep along the inner surface of the retina in the nerve fiber layer, exit the eye at the optic disc, and travel through the optic nerve, optic chiasm, and optic tract to reach targets in the brain. The majority of fibers synapse on cells in the lateral geniculate nucleus, a thalamic relay station. Cells in the lateral geniculate nucleus project in turn to the primary visual cortex. This afferent retinogeniculocortical sensory pathway provides the neural substrate for visual perception. Separate classes of ganglion cells project to subcortical visual nuclei involved in other functions. Pupillary constriction and circadian rhythms are governed by ganglion cells that are intrinsically light sensitive, owing to a pigment named melanopsin. Pupil reflexes are mediated by a projection to the pretectal olivary nuclei. Their output is supplied to the Edinger-Westphal nuclei, which provide parasympathetic innervation to the iris sphincter via an interneuron in the ciliary ganglion. Circadian rhythms are timed by melanopsin ganglion cells that project to the suprachiasmatic nucleus. Visual orientation and eye movements are served by retinal input to the superior colliculus. Gaze stabilization and optokinetic reflexes are governed by a cluster of small retinal targets known collectively as the *brainstem accessory optic system*.

The eyes must be rotated constantly within their orbits to place and maintain targets of visual interest on the fovea. This activity, called *foveation*, or looking, is governed by an elaborate efferent motor system. Each eye is moved by six extraocular muscles that are supplied by cranial nerves from the oculomotor (III), trochlear (IV), and abducens (VI) nuclei. Activity in these ocular motor nuclei is coordinated by pontine and midbrain mechanisms for smooth pursuit, saccades, and gaze stabilization during head and body movements. Large regions of the frontal and parietooccipital cortex control these brainstem eye movement centers by providing descending supranuclear input.

CLINICAL ASSESSMENT OF VISUAL FUNCTION

REFRACTIVE STATE

In approaching a patient with reduced vision, the first step is to decide whether refractive error is responsible. In *emmetropia*, parallel rays from infinity are focused perfectly on the retina. Sadly, this condition is enjoyed by only a minority of the population. In *myopia*, the globe is too long, and light rays come to a focal point in front of the retina. Near objects can be seen clearly, but distant objects require a diverging lens in front of the eye. In *hyperopia*, the globe is too short, and hence, a converging lens is used to supplement the refractive power of the eye. In *astigmatism*, the corneal surface is not spherical, necessitating a cylindrical corrective lens. Most patients elect to wear eyeglasses or contact lenses to neutralize refractive error. An alternative is to permanently alter the refractive properties of the cornea by performing laser in situ keratomileusis (LASIK) or photorefractive keratectomy (PRK).

With the onset of middle age, *presbyopia* develops as the lens within the eye becomes unable to increase its refractive power to accommodate on near objects. To compensate for presbyopia, an emmetropic patient must use reading glasses. A patient already wearing glasses for distance correction usually switches to bifocals. The only exception is a myopic patient, who may achieve clear vision at near simply by removing glasses containing the distance prescription.

Refractive errors usually develop slowly and remain stable after adolescence, except in unusual circumstances. For example, the acute onset of diabetes mellitus can produce sudden myopia because of lens edema induced by hyperglycemia. Testing vision through a pinhole aperture is a useful way to screen quickly for refractive error. If acuity is improved by viewing through a pinhole, the patient needs a refraction to obtain best corrected visual acuity.

VISUAL ACUITY

The Snellen chart is used to test acuity at a distance of 6 m (20 ft). A portable scale version of the Snellen chart called the Rosenbaum card is held at 36 cm (14 in.) from the patient (**eFig. 34-1**: available at *accessmedicine.com/harrisons*). All subjects should be able to read the 6/6 m (20/20 ft) line with each eye using their refractive correction, if any. Patients who need reading glasses because of presbyopia must wear them for accurate testing with the Rosenbaum card. If 6/6 (20/20) acuity is not present in each eye, the deficiency in vision must be explained. If it is worse than 6/240 (20/800), acuity should be recorded in terms of counting fingers, hand motions, light perception, or no light perception. Legal blindness is defined by the Internal Revenue Service as a best corrected acuity of 6/60 (20/200) or less in the better eye or a binocular visual field subtending 20° or less. Loss of vision in one eye only does not constitute legal blindness. For driving, the laws vary by state, but most require a corrected acuity of 6/12 (20/40) in at least one eye for unrestricted privileges. Patients who develop a homonymous hemianopia should not drive.

PUPILS

The pupils should be tested individually in dim light with the patient fixating on a distant target. There is no need to check the near response if the pupils respond briskly to light, because isolated loss of constriction (miosis) to accommodation does not occur. For this reason, the ubiquitous abbreviation PERRLA (pupils equal, round, and reactive to light and accommodation) implies a wasted effort with the last step. However, it is important to test the near response if the light response is poor or absent. Light-near dissociation occurs with neurosyphilis (Argyll Robertson pupil), with lesions of the dorsal midbrain (*Parinaud's syndrome*), and after aberrant regeneration (oculomotor nerve palsy, Adie's tonic pupil).

An eye with no light perception has no pupillary response to direct light stimulation. If the retina or optic nerve is only partially injured, the direct pupillary response will be weaker than the consensual pupillary response evoked by shining a light into the healthy fellow eye. A *relative afferent pupillary defect* (Marcus Gunn pupil) is elicited with the swinging flashlight test. It is an extremely useful sign in retrobulbar optic neuritis and other optic nerve diseases, in which it may be the sole objective evidence for disease. In bilateral optic neuropathy, no afferent pupil defect is present if the optic nerves are affected equally.

Subtle inequality in pupil size, up to 0.5 mm, is a fairly common finding in normal persons. The diagnosis of essential or physiologic anisocoria is secure as long as the relative pupil asymmetry remains constant as ambient lighting varies. Anisocoria that increases in dim light indicates a sympathetic paresis of the iris dilator muscle. The triad of miosis with ipsilateral ptosis and anhidrosis constitutes *Horner's syndrome*, although anhidrosis is an inconstant feature. A drop of 1% apraclonidine produces no effect on the normal pupil, but the miotic pupil dilates because of denervation hypersensitivity. Brainstem stroke, carotid dissection, and neoplasm impinging on the sympathetic chain occasionally are identified as the cause of Horner's syndrome, but most cases are idiopathic.

Anisocoria that increases in bright light suggests a parasympathetic palsy. The first concern is an oculomotor nerve paresis. This possibility is excluded if the eye movements are full and the patient has no ptosis or diplopia. Acute pupillary dilation (mydriasis) can result from damage to the ciliary ganglion in the orbit. Common mechanisms are infection (herpes zoster, influenza), trauma (blunt, penetrating, surgical), and ischemia (diabetes, temporal arteritis). After denervation of the iris sphincter, the pupil does not respond well to light, but the response to near is often relatively intact. When the near stimulus is removed, the pupil redilates very slowly compared with the normal pupil, hence the term *tonic pupil*. In *Adie's syndrome*, a tonic pupil is present, sometimes in conjunction with weak or absent tendon reflexes in the lower extremities. This benign disorder, which occurs predominantly in healthy young women, is assumed to represent a mild dysautonomia. Tonic pupils are also associated with multiple system atrophy, segmental hypohidrosis, diabetes, and amyloidosis. Occasionally, a tonic pupil is discovered incidentally in an otherwise completely normal, asymptomatic individual. The diagnosis is confirmed by placing a drop of dilute (0.125%) pilocarpine into each eye. Denervation hypersensitivity produces pupillary constriction in a tonic pupil, whereas the normal pupil shows no response. Pharmacologic dilatation from accidental or deliberate instillation of anticholinergic (atropine, scopolamine) drops can produce pupillary mydriasis. Gardener's pupil refers to mydriasis induced by exposure to tropane alkaloids, contained in plants such as deadly nightshade, jimsonweed, or angel's trumpet. When an anticholinergic agent is responsible for pupil dilation, 1% pilocarpine causes no constriction.

Both pupils are affected equally by systemic medications. They are small with opiate use and large with anticholinergics (scopolamine). Parasympathetic agents (pilocarpine) used to treat glaucoma produce miosis. In any patient with an unexplained pupillary abnormality, a slit-lamp examination is helpful to exclude an occult foreign body, perforating injury, intraocular inflammation, adhesions (synechia), angle-closure glaucoma, and iris sphincter rupture from blunt trauma.

EYE MOVEMENTS AND ALIGNMENT

Eye movements are tested by asking the patient, with both eyes open, to pursue a small target such as a pen tip into the cardinal fields of gaze. Normal ocular versions are smooth, symmetric, full, and maintained in all directions. Saccades, or quick refixation eye movements, are assessed by having the patient look back and forth between two stationary targets. The eyes should move rapidly and accurately in a single jump to their target. Ocular alignment can be judged by holding a penlight directly in front of the patient at about 1 m. If the eyes are straight, the corneal light reflex will be centered in the middle of each pupil. To test eye alignment more precisely, the cover test is useful. The patient is instructed to look at a small fixation target in the distance. One eye is occluded with a paddle or hand, while the other eye is observed. If the viewing eye shifts position to take up fixation on the target, it was misaligned. If it remains motionless, the first eye is uncovered and the test is repeated on the second eye. If neither eye moves, the eyes are aligned orthotropically. If the eyes are orthotropic in primary gaze but the patient complains of diplopia, the cover test should be performed with the head tilted or turned in whatever direction elicits diplopia. With practice, the examiner can detect an ocular deviation (heterotropia) as small as 1° with the cover test. In a patient with vertical diplopia, a small deviation can be difficult to detect and easy to dismiss. The magnitude of the deviation can be measured by placing a prism in front of the misaligned eye to determine the power required to neutralize the fixation shift evoked by covering the other eye. Temporary press-on plastic Fresnel prisms, prism eyeglasses, or eye muscle surgery can be used to restore binocular alignment.

STEREOPSIS

Stereoacuity is determined by presenting targets with retinal disparity separately to each eye by using polarized images. The most popular office tests measure a range of thresholds from 2000 to 40 s of arc. Normal stereoacuity is 40 s of arc. If a patient achieves this level of stereoacuity, one is assured that the eyes are aligned orthotropically and that vision is intact in each eye. Random dot stereograms have no monocular depth cues and provide an excellent screening test for strabismus.

COLOR VISION

The retina contains three classes of cones, with visual pigments of differing peak spectral sensitivity: red (560 nm), green (530 nm), and blue (430 nm). The red and green cone pigments are encoded on the X chromosome, and the blue cone pigment on chromosome 7. Mutations of the blue cone pigment are exceedingly rare. Mutations of the red and green pigments cause congenital X-linked color blindness in 8% of males. Affected individuals are not truly color blind; rather, they differ from normal subjects in the way they perceive color and how they combine primary monochromatic lights to match a particular color. Anomalous trichromats have three cone types, but a mutation in one cone pigment (usually red or green) causes a shift in peak spectral sensitivity, altering the proportion of primary colors required to achieve a color match. Dichromats have only two cone types and therefore will accept a color match based on only two primary colors. Anomalous trichromats and dichromats have 6/6 (20/20) visual acuity, but their hue discrimination is impaired. Ishihara color plates can be used to detect red-green color blindness. The test plates contain a hidden number that is visible only to subjects with color confusion from red-green color blindness. Because color blindness is almost exclusively X-linked, it is worthwhile screening only male children.

The Ishihara plates often are used to detect acquired defects in color vision, although they are intended as a screening test for congenital color blindness. Acquired defects in color vision frequently result from disease of the macula or optic nerve. For example, patients with a history of optic neuritis often complain of color desaturation long after their visual acuity has returned to normal. Color blindness also can result from bilateral strokes involving the ventral portion of the occipital lobe (cerebral achromatopsia). Such patients can perceive only shades of gray and also may have difficulty recognizing faces (prosopagnosia) **(Chap. 32)**. Infarcts of the dominant occipital lobe sometimes give rise to color anomia. Affected patients can discriminate colors but cannot name them.

■ VISUAL FIELDS

Vision can be impaired by damage to the visual system anywhere from the eyes to the occipital lobes. One can localize the site of the lesion with considerable accuracy by mapping the visual field deficit by finger confrontation and then correlating it with the topographic anatomy of the visual pathway **(Fig. 34-1)**. Quantitative visual field mapping is performed by computer-driven perimeters that present a target of variable intensity at fixed positions in the visual field **(Fig. 34-1*A*)**. By generating an automated printout of light thresholds, these static perimeters provide a sensitive means of detecting scotomas in the visual field. They are exceedingly useful for serial assessment of visual function in chronic diseases such as glaucoma and pseudotumor cerebri.

The crux of visual field analysis is to decide whether a lesion is before, at, or behind the optic chiasm. If a scotoma is confined to one eye, it must be due to a lesion anterior to the chiasm, involving either the optic nerve or the retina. Retinal lesions produce scotomas that correspond optically to their location in the fundus. For example, a superior-nasal retinal detachment results in an inferior-temporal field cut. Damage to the macula causes a central scotoma **(Fig. 34-1*B*)**.

Optic nerve disease produces characteristic patterns of visual field loss. Glaucoma selectively destroys axons that enter the superotemporal or inferotemporal poles of the optic disc, resulting in arcuate scotomas shaped like a Turkish scimitar, which emanate from the blind spot and curve around fixation to end flat against the horizontal meridian **(Fig. 34-1*C*)**. This type of field defect mirrors the arrangement of the nerve fiber layer in the temporal retina. Arcuate or nerve fiber layer scotomas also result from optic neuritis, ischemic optic neuropathy, optic disc drusen, and branch retinal artery or vein occlusion.

Damage to the entire upper or lower pole of the optic disc causes an altitudinal field cut that follows the horizontal meridian **(Fig. 34-1*D*)**. This pattern of visual field loss is typical of ischemic optic neuropathy but also results from retinal vascular occlusion, advanced glaucoma, and optic neuritis.

About half the fibers in the optic nerve originate from ganglion cells serving the macula. Damage to papillomacular fibers causes a cecocentral scotoma that encompasses the blind spot and macula **(Fig. 34-1*E*)**. If the damage is irreversible, pallor eventually appears in the temporal portion of the optic disc. Temporal pallor from a cecocentral scotoma may develop in optic neuritis, nutritional optic neuropathy, toxic optic neuropathy, Leber's hereditary optic neuropathy, Kjer's dominant optic atrophy, and compressive optic neuropathy. It is worth mentioning that the temporal side of the optic disc is slightly paler than the nasal side in most normal individuals. Therefore, it sometimes can be difficult to decide whether temporal pallor visible on fundus examination represents a pathologic change. Pallor of the nasal rim of the optic disc is a less equivocal sign of optic atrophy.

At the optic chiasm, fibers from nasal ganglion cells decussate into the contralateral optic tract. Crossed fibers are damaged more by compression than are uncrossed fibers. As a result, mass lesions of the sellar region cause a temporal hemianopia in each eye. Tumors anterior to the optic chiasm, such as meningiomas of the tuberculum sella, produce a junctional scotoma characterized by an optic neuropathy in one eye and a superior-temporal field cut in the other eye **(Fig. 34-1*G*)**. More symmetric compression of the optic chiasm by a pituitary adenoma **(see Fig. 392-1)**, meningioma, craniopharyngioma, glioma, or aneurysm results in a bitemporal hemianopia **(Fig. 34-1*H*)**. The insidious development of a bitemporal hemianopia often goes unnoticed by the patient and will escape detection by the physician unless each eye is tested separately.

It is difficult to localize a postchiasmal lesion accurately, because injury anywhere in the optic tract, lateral geniculate nucleus, optic radiations, or visual cortex can produce a homonymous hemianopia (i.e., a temporal hemifield defect in the contralateral eye and a matching nasal hemifield defect in the ipsilateral eye) **(Fig. 34-1*I*)**. A unilateral postchiasmal lesion leaves the visual acuity in each eye unaffected, although the patient may read the letters on only the left or right half of the eye chart. Lesions of the optic radiations tend to cause poorly matched or incongruous field defects in each eye. Damage to the optic radiations in the temporal lobe (Meyer's loop) produces a superior quadrantic homonymous hemianopia **(Fig. 34-1*J*)**, whereas injury to the optic radiations in the parietal lobe results in an inferior quadrantic homonymous hemianopia **(Fig. 34-1*K*)**. Lesions of the primary visual cortex give rise to dense, congruous hemianopic field defects. Occlusion of the posterior cerebral artery supplying the occipital lobe is a common cause of total homonymous hemianopia. Some patients have macular sparing, because the central field representation at the tip of the occipital lobe is supplied by collaterals from the middle cerebral artery **(Fig. 34-1*L*)**. Destruction of both occipital lobes produces cortical blindness. This condition can be distinguished from bilateral prechiasmal visual loss by noting that the pupil responses and optic fundi remain normal.

Partial recovery of homonymous hemianopia has been reported through computer-based rehabilitation therapy. During daily training sessions, patients fixate a central target while visual stimuli are presented within the blind region. The premise of vision restoration programs is that extra stimulation can promote recovery of partially damaged tissue located at the fringe of a cortical lesion. When fixation is controlled rigorously, however, no improvement of the visual fields can be demonstrated. No effective treatment exists for homonymous hemianopia caused by permanent brain damage.

DISORDERS

■ RED OR PAINFUL EYE

Corneal Abrasions Corneal abrasions are seen best by placing a drop of fluorescein in the eye and looking with the slit lamp, using a cobalt-blue light. A penlight with a blue filter will suffice if a slit lamp is not available. Damage to the corneal epithelium is revealed by yellow fluorescence of the basement membrane exposed by loss of the overlying epithelium. It is important to check for foreign bodies. To search the conjunctival fornices, the lower lid should be pulled down and the upper lid everted. A foreign body can be removed with a moistened cotton-tipped applicator after a drop of a topical anesthetic such as proparacaine has been placed in the eye. Alternatively, it may be possible to flush the foreign body from the eye by irrigating copiously with saline or artificial tears. If the corneal epithelium has been abraded, antibiotic ointment and a patch may be applied to the eye. A drop of an intermediate-acting cycloplegic such as cyclopentolate hydrochloride 1% helps reduce pain by relaxing the ciliary body. The eye should be reexamined the next day. Minor abrasions may not require patching, antibiotics, or cycloplegia.

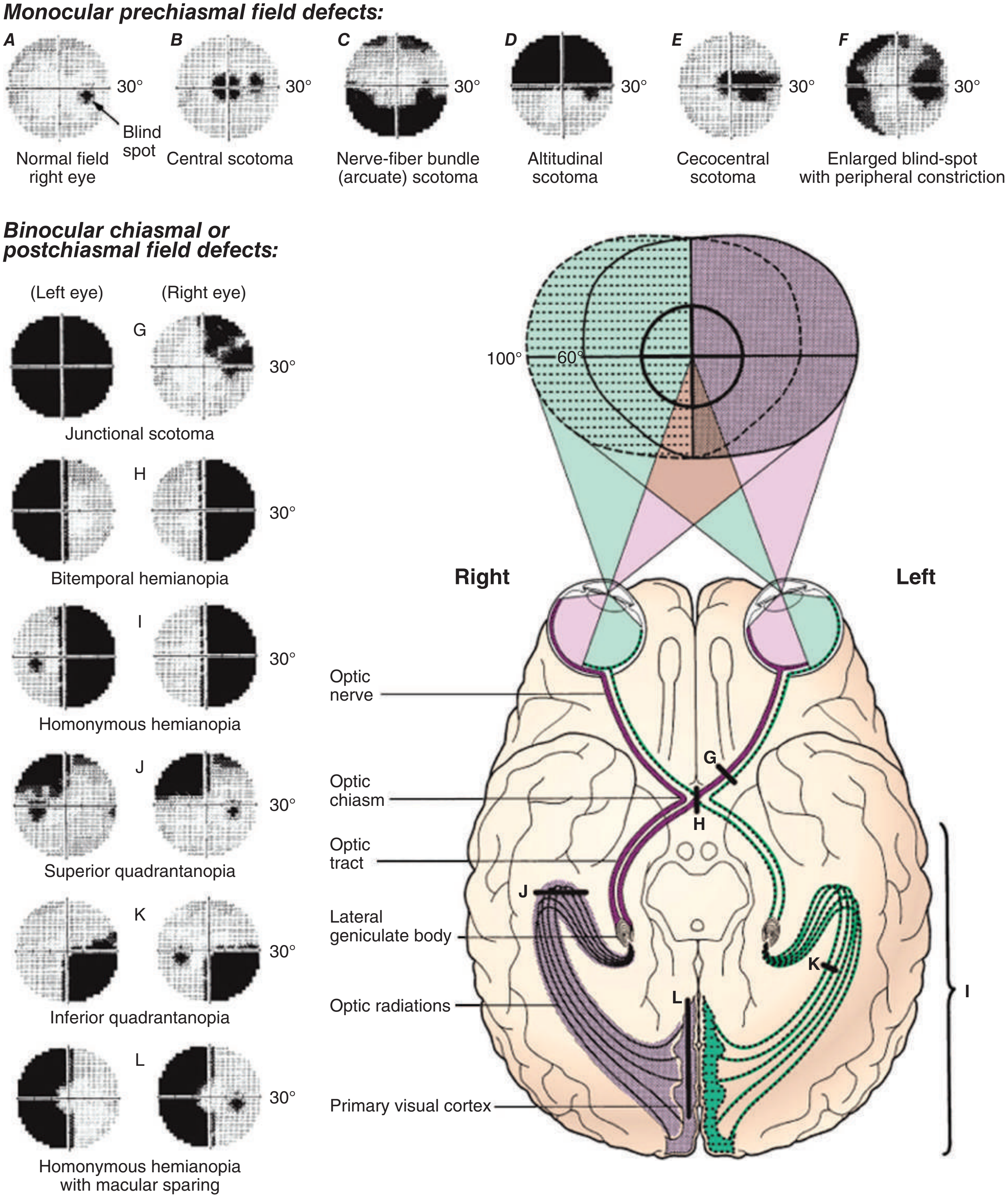
Monocular prechiasmal field defects:
A
B
C
D
E
F
30°
Blind spot
Normal field right eye
Central scotoma
Nerve-fiber bundle (arcuate) scotoma
Altitudinal scotoma
Cecocentral scotoma
Enlarged blind-spot with peripheral constriction
Binocular chiasmal or postchiasmal field defects:
(Left eye)
(Right eye)
G
Junctional scotoma
H
Bitemporal hemianopia
I
Homonymous hemianopia
J
Superior quadrantanopia
K
Inferior quadrantanopia
L
Homonymous hemianopia with macular sparing
100°
60°
Right
Left
Optic nerve
Optic chiasm
Optic tract
Lateral geniculate body
Optic radiations
Primary visual cortex
G
H
J
K
L
I

FIGURE 34-1 Ventral view of the brain, correlating patterns of visual field loss with the sites of lesions in the visual pathway. The visual fields overlap partially, creating 120° of central binocular field flanked by a 40° monocular crescent on either side. The visual field maps in this figure were done with a computer-driven perimeter (Humphrey Instruments, Carl Zeiss, Inc.). It plots the retinal sensitivity to light in the central 30° by using a gray scale format. Areas of visual field loss are shown in black. The examples of common monocular, prechiasmal field defects are all shown for the right eye. By convention, the visual fields are always recorded with the left eye's field on the left and the right eye's field on the right, just as the patient sees the world.

Subconjunctival Hemorrhage This results from rupture of small vessels bridging the potential space between the episclera and the conjunctiva. Blood dissecting into this space can produce an impressive red eye, but vision is not affected and the hemorrhage resolves without treatment. Subconjunctival hemorrhage is usually spontaneous but can result from blunt trauma, eye rubbing, or vigorous coughing. Occasionally, it is a clue to an underlying bleeding disorder.

Pinguecula Pinguecula is a small, raised conjunctival nodule, usually at the nasal limbus. In adults such lesions are extremely common and have little significance unless they become inflamed (pingueculitis). They are more apt to occur in workers with outdoor exposure. A *pterygium* resembles a pinguecula but has crossed the limbus to encroach on the corneal surface. Removal is justified when symptoms of irritation or blurring develop, but recurrence is common.

Blepharitis This refers to inflammation of the eyelids. The most common form occurs in association with acne rosacea or seborrheic dermatitis. The eyelid margins usually are colonized heavily by staphylococci. Upon close inspection, they appear greasy, ulcerated, and crusted with scaling debris that clings to the lashes. Treatment consists of strict eyelid hygiene, applying warm compresses, and eyelash scrubs with a cleansing agent. An external *hordeolum* (stye) is caused by staphylococcal infection of the superficial accessory glands of Zeis or Moll located in the eyelid margins. An internal hordeolum occurs after suppurative infection of the oil-secreting meibomian glands within the tarsal plate of the eyelid. Topical antibiotics such as bacitracin/polymyxin B ophthalmic ointment can be applied. Systemic antibiotics, usually tetracyclines or azithromycin, sometimes are necessary for treatment of meibomian gland inflammation (meibomitis) or chronic, severe blepharitis. A *chalazion* is a painless, chronic granulomatous inflammation of a meibomian gland that produces a pealike nodule within the eyelid. It can be incised and drained, but injection with glucocorticoids is equally effective. Basal cell, squamous cell, or meibomian gland carcinoma should be suspected with any nonhealing ulcerative lesion of the eyelids.

Dacryocystitis An inflammation of the lacrimal drainage system, dacryocystitis can produce epiphora (tearing) and ocular injection. Gentle pressure over the lacrimal sac evokes pain and reflux of mucus or pus from the tear puncta. Dacryocystitis usually occurs after obstruction of the lacrimal system. It is treated with topical and systemic antibiotics, followed by probing, silicone stent intubation, or surgery to reestablish patency. *Entropion* (inversion of the eyelid) or *ectropion* (sagging or eversion of the eyelid) can also lead to epiphora and ocular irritation.

Conjunctivitis Conjunctivitis is the most common cause of a red, irritated eye. Pain is minimal, and visual acuity is reduced only slightly. The most common viral etiology is adenovirus infection. It causes a watery discharge, a mild foreign-body sensation, and photophobia. Bacterial infection tends to produce a more mucopurulent exudate. Mild cases of infectious conjunctivitis usually are treated empirically with broad-spectrum topical ocular antibiotics such as sulfacetamide 10%, polymyxin-bacitracin, or a trimethoprim-polymyxin combination. Smears and cultures usually are reserved for severe, resistant, or recurrent cases of conjunctivitis. To prevent contagion, patients should be admonished to wash their hands frequently, not to touch their eyes, and to avoid direct contact with others.

Allergic Conjunctivitis This condition is extremely common and often is mistaken for infectious conjunctivitis. Itching, redness, and epiphora are typical. The palpebral conjunctiva may become hypertropic with giant excrescences called cobblestone papillae. Irritation from contact lenses or any chronic foreign body also can induce formation of cobblestone papillae. *Atopic conjunctivitis* occurs in subjects with atopic dermatitis or asthma. Symptoms caused by allergic conjunctivitis can be alleviated with cold compresses, topical vasoconstrictors (naphazoline), antihistamines (olopatadine), and mast cell stabilizers (cromolyn). Topical glucocorticoid solutions provide dramatic relief of immune-mediated forms of conjunctivitis, but their long-term use is ill advised because of the complications of glaucoma, cataract, and secondary infection. Topical nonsteroidal anti-inflammatory drugs (NSAIDs; ketorolac) are better alternatives.

Keratoconjunctivitis Sicca Also known as dry eye, this produces a burning foreign-body sensation, injection, and photophobia. In mild cases, the eye appears surprisingly normal, but tear production measured by wetting of a filter paper (Schirmer strip) is deficient. A variety of systemic drugs, including antihistaminic, anticholinergic, and psychotropic medications, result in dry eye by reducing lacrimal secretion. Disorders that involve the lacrimal gland directly, such as sarcoidosis and Sjögren's syndrome, also cause dry eye. Patients may develop dry eye after radiation therapy if the treatment field includes the orbits. Ocular drying is also common after lesions affecting cranial nerve V or VII. Corneal anesthesia is particularly dangerous, because the absence of a normal blink reflex exposes the cornea to injury without pain to warn the patient. Dry eye is managed by frequent and liberal application of artificial tears and ocular lubricants. In severe cases, the tear puncta can be plugged or cauterized to reduce lacrimal outflow.

Keratitis Keratitis is a threat to vision because of the risk of corneal clouding, scarring, and perforation. Worldwide, the two leading causes of blindness from keratitis are trachoma from chlamydial infection and vitamin A deficiency related to malnutrition. In the United States, contact lenses play a major role in corneal infection and ulceration. They should not be worn by anyone with an active eye infection. In evaluating the cornea, it is important to differentiate between a superficial infection (*keratoconjunctivitis*) and a deeper, more serious ulcerative process. The latter is accompanied by greater visual loss, pain, photophobia, redness, and discharge. Slit-lamp examination shows disruption of the corneal epithelium, a cloudy infiltrate or abscess in the stroma, and an inflammatory cellular reaction in the anterior chamber. In severe cases, pus settles at the bottom of the anterior chamber, giving rise to a hypopyon. Immediate empirical antibiotic therapy should be initiated after corneal scrapings are obtained for Gram's stain, Giemsa stain, potassium hydroxide (KOH) prep, and cultures. Fortified topical antibiotics are most effective, supplemented with subconjunctival antibiotics as required. A fungal etiology should always be considered in a patient with keratitis. Fungal infection is common in warm humid climates, especially after penetration of the cornea by plant or vegetable material. Acanthamoeba keratitis is associated with improper disinfection of contact lenses.

Herpes Simplex The *herpesviruses* are a major cause of blindness from keratitis. Most adults in the United States have serum antibodies to herpes simplex, indicating prior viral infection **(Chap. 197)**. Primary ocular infection generally is caused by herpes simplex type 1 rather than type 2. It manifests as a unilateral follicular blepharoconjunctivitis that is easily confused with adenoviral conjunctivitis, unless telltale vesicles are present on the eyelids or conjunctiva. Recurrent ocular infection arises from reactivation of latent herpesvirus. A dendritic pattern of corneal epithelial ulceration revealed by fluorescein staining is pathognomonic for herpes infection but often not present. Involvement of both eyes is extremely rare. Corneal stromal inflammation produces edema, vascularization, and iridocyclitis. Herpes keratitis is treated with cycloplegia and either a topical antiviral (trifluridine, ganciclovir) or an oral antiviral (acyclovir, valacyclovir) agent. Topical glucocorticoids are effective in mitigating corneal scarring but generally are reserved for cases involving stromal damage. Risks include corneal melting, perforation, prolonged infection, and glaucoma.

Herpes Zoster Herpes zoster from reactivation of latent varicella (chickenpox) virus causes a dermatomal pattern of painful vesicular dermatitis **(Chap. 198)**. Ocular symptoms can occur after zoster eruption in any branch of the trigeminal nerve but are particularly common when vesicles form on the nose, reflecting nasociliary (V1) nerve involvement (Hutchinson's sign). Herpes zoster ophthalmicus produces corneal dendrites, which can be difficult to distinguish from those seen in herpes simplex. Stromal keratitis, anterior uveitis, raised intraocular pressure, ocular motor nerve palsies, acute retinal necrosis, and postherpetic scarring and neuralgia are other common sequelae. Herpes zoster ophthalmicus is treated with antiviral agents and cycloplegics. In severe cases, topical steroids may be added to reduce corneal scarring. Shingles should be prevented by vaccination of all healthy adults aged 50 years and older.

Episcleritis This is an inflammation of the episclera, a thin layer of connective tissue between the conjunctiva and the sclera. Episcleritis resembles conjunctivitis, but it is a more localized process and discharge is absent. Most cases of episcleritis are idiopathic, but some occur in the setting of an autoimmune disease. *Scleritis* refers to a deeper, more severe inflammatory process that frequently is associated with a connective tissue disease such as rheumatoid arthritis, lupus erythematosus, polyarteritis nodosa, granulomatosis with polyangiitis, or relapsing polychondritis. The inflammation and thickening of the sclera can be diffuse or nodular. In anterior forms of scleritis, the globe assumes a violet hue and the patient complains of severe ocular tenderness and pain. With posterior scleritis, the pain and redness may be less marked, but there is often proptosis, choroidal effusion, reduced motility, and visual loss. Episcleritis and scleritis should be treated with NSAIDs. If these agents fail, topical or even systemic glucocorticoid therapy may be necessary, especially if an underlying autoimmune process is active.

Anterior Uveitis Involving the anterior structures of the eye, uveitis was previously called *iritis* or *iridocyclitis*. The diagnosis requires slit-lamp examination to identify inflammatory cells floating in the aqueous humor or deposited on the corneal endothelium (keratic precipitates). Anterior uveitis develops in sarcoidosis, ankylosing spondylitis, juvenile idiopathic arthritis, inflammatory bowel disease, psoriasis, reactive arthritis, and Behçet's disease. It also is associated with herpes infections, syphilis, Lyme disease, onchocerciasis, tuberculosis, and leprosy. Although anterior uveitis can occur in conjunction with many diseases, no cause is found to explain the majority of cases. For this reason, laboratory evaluation usually is reserved for patients with recurrent or severe anterior uveitis. Treatment is aimed at reducing inflammation and scarring by judicious use of topical glucocorticoids. Dilatation of the pupil reduces pain and prevents the formation of synechiae.

Posterior Uveitis This diagnosis is made by observing inflammation of the vitreous, retina, or choroid on fundus examination. It is more likely than anterior uveitis to be associated with an identifiable systemic disease. Some patients have panuveitis, or inflammation of both the anterior and posterior segments of the eye. Posterior uveitis is a manifestation of autoimmune diseases such as sarcoidosis, Behçet's disease, Vogt-Koyanagi-Harada syndrome, and inflammatory bowel disease. It also accompanies diseases such as toxoplasmosis, onchocerciasis, cysticercosis, coccidioidomycosis, toxocariasis, and histoplasmosis; infections caused by organisms such as *Candida*, *Pneumocystis carinii*, *Cryptococcus*, *Aspergillus*, herpes, and cytomegalovirus **(see Fig. 200-1)**; and other diseases, such as syphilis, Lyme disease, tuberculosis, cat-scratch disease, Whipple's disease, and brucellosis. In multiple sclerosis, chronic inflammatory changes can develop in the extreme periphery of the retina (pars planitis or intermediate uveitis). Glucocorticoids have been the mainstay of treatment for noninfectious uveitis. Biologic agents that target proinflammatory cytokines, such as the tumor necrosis factor alpha (TNF-α) inhibitor adalimumab, are effective at preventing vision loss in chronic uveitis.

Acute Angle-Closure Glaucoma This is an unusual but frequently misdiagnosed cause of a red, painful eye. Asian populations have a particularly high risk of angle-closure glaucoma. Susceptible eyes have a shallow anterior chamber because the eye has either a short axial length (hyperopia) or a lens enlarged by the gradual development of cataract. When the pupil becomes mid-dilated, the peripheral iris blocks aqueous outflow via the anterior chamber angle and the intraocular pressure rises abruptly, producing pain, injection, corneal edema, obscurations, and blurred vision. In some patients, ocular symptoms are overshadowed by nausea, vomiting, or headache, prompting a fruitless workup for abdominal or neurologic disease. The diagnosis is made by measuring the intraocular pressure during an acute attack or by performing gonioscopy, a procedure that allows one to observe a narrow chamber angle with a mirrored contact lens. Acute angle closure is treated with acetazolamide (PO or IV), topical beta blockers, prostaglandin analogues, α_2-adrenergic agonists, and pilocarpine to induce miosis. If these measures fail, a laser can be used to create a hole in the peripheral iris to relieve pupillary block. Many physicians are reluctant to dilate patients routinely for fundus examination because they fear precipitating an angle-closure glaucoma. The risk is actually remote and more than outweighed by the potential benefit to patients of discovering a hidden fundus lesion visible only through a fully dilated pupil. Moreover, a single attack of angle closure after pharmacologic dilatation rarely causes any permanent damage to the eye and serves as an inadvertent provocative test to identify patients with narrow angles who would benefit from prophylactic laser iridectomy.

Endophthalmitis This results from bacterial, viral, fungal, or parasitic infection of the internal structures of the eye. It usually is acquired by hematogenous seeding from a remote site. Chronically ill, diabetic, or immunosuppressed patients, especially those with a history of indwelling IV catheters or positive blood cultures, are at greatest risk for endogenous endophthalmitis. Although most patients have ocular pain and injection, visual loss is sometimes the only symptom. Septic emboli from a diseased heart valve or a dental abscess that lodge in the retinal circulation can give rise to endophthalmitis. White-centered retinal hemorrhages known as Roth's spots **(Fig. 34-2)** are considered pathognomonic for subacute bacterial endocarditis, but they also appear in leukemia, diabetes, and many other conditions. Endophthalmitis occurs as a complication of ocular surgery, especially glaucoma filtering, occasionally months or even years after the operation. An occult penetrating foreign body or unrecognized trauma to the globe should be considered in any patient with unexplained intraocular infection or inflammation.

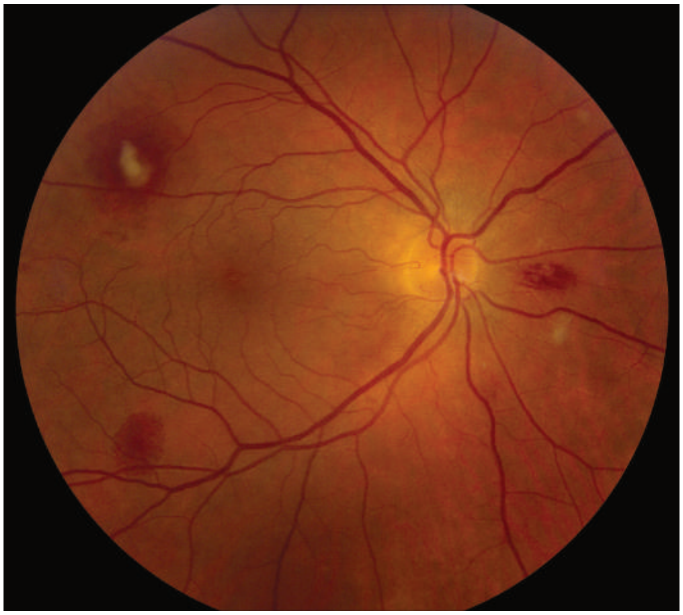

FIGURE 34-2 Roth's spot, cotton-wool spot, and retinal hemorrhages in a 48-year-old liver transplant patient with candidemia from immunosuppression.

TRANSIENT OR SUDDEN VISUAL LOSS

Amaurosis Fugax This term refers to a transient ischemic attack of the retina **(Chap. 438)**. Because neural tissue has a high rate of metabolism, interruption of blood flow to the retina for more than a few seconds results in *transient monocular blindness*, a term used interchangeably with amaurosis fugax. Patients describe a rapid fading of vision like a curtain descending, sometimes affecting only a portion of the visual field. Amaurosis fugax usually results from an embolus that becomes stuck within a retinal arteriole **(Fig. 34-3)**. If the embolus breaks up or passes, flow is restored and vision returns quickly to normal without permanent damage. With prolonged interruption of blood flow, the inner retina suffers infarction. Ophthalmoscopy reveals zones of whitened, edematous retina following the distribution of branch retinal arterioles. Complete occlusion of the central retinal artery produces arrest of blood flow and a milky retina with a cherry-red fovea **(Fig. 34-4)**. Emboli are composed of cholesterol (Hollenhorst plaque), calcium, or platelet-fibrin debris. The most common source is an atherosclerotic plaque in the carotid artery or aorta, although emboli can also arise from the heart, especially in patients with diseased valves, atrial fibrillation, or wall motion abnormalities. Urgent evaluation is appropriate, because of the risk of stroke.

In rare instances, amaurosis fugax results from low central retinal artery perfusion pressure in a patient with a critical stenosis of the ipsilateral carotid artery and poor collateral flow via the circle of Willis. In this situation, amaurosis fugax develops when there is a dip in systemic blood pressure or a slight worsening of the carotid stenosis. Sometimes there is contralateral motor or sensory loss, indicating concomitant hemispheric cerebral ischemia.

Retinal arterial occlusion also occurs rarely in association with retinal migraine, lupus erythematosus, anticardiolipin antibodies, anticoagulant deficiency states (protein S, protein C, and antithrombin

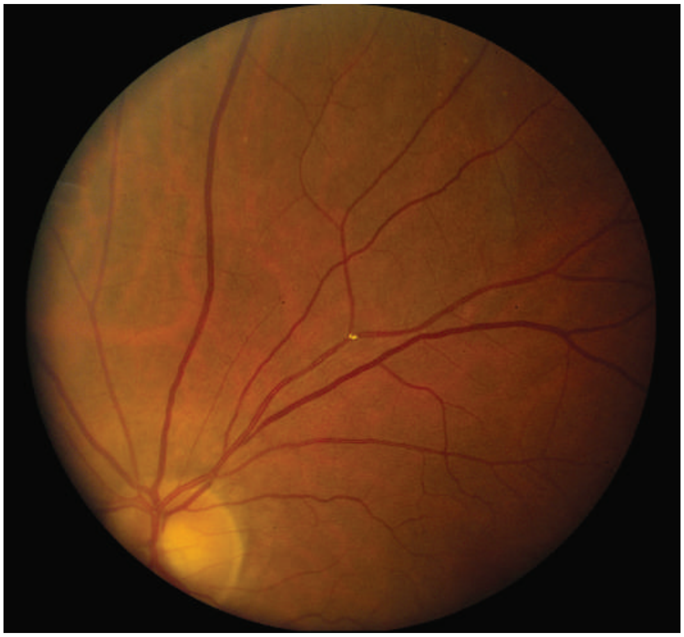

FIGURE 34-3 Hollenhorst plaque lodged at the bifurcation of a retinal arteriole proves that a patient is shedding emboli from the carotid artery, great vessels, or heart.

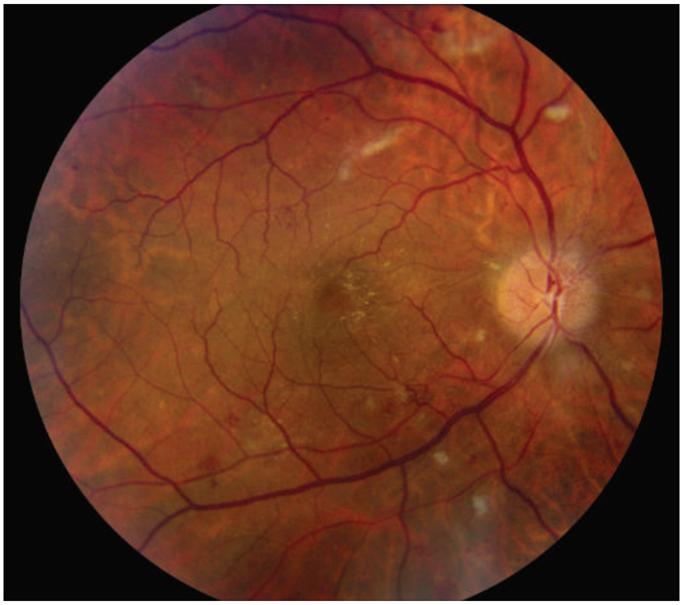

FIGURE 34-5 Hypertensive retinopathy with blurred optic disc, scattered hemorrhages, cotton-wool spots (nerve fiber layer infarcts), and foveal exudate in a 62-year-old man with chronic renal failure and a systolic blood pressure of 220.

deficiency), Susac's syndrome, pregnancy, IV drug abuse, blood dyscrasias, dysproteinemias, and temporal arteritis.

Marked *systemic hypertension* causes sclerosis of retinal arterioles, splinter hemorrhages, focal infarcts of the nerve fiber layer (cotton-wool spots), and leakage of lipid and fluid (hard exudate) into the macula **(Fig. 34-5)**. In hypertensive crisis, sudden visual loss can result from ischemia induced by vasospasm of retinal arterioles. In addition, visual loss can occur from ischemic optic disc swelling. Patients with acute hypertensive retinopathy should be treated by lowering the blood pressure. However, the blood pressure should not be reduced precipitously, because there is a danger of optic disc infarction from sudden hypoperfusion.

Impending *branch* or *central retinal vein occlusion* can produce prolonged visual obscurations that resemble those described by patients with amaurosis fugax. The veins appear engorged and phlebitic, with numerous retinal hemorrhages **(Fig. 34-6)**. In some patients, venous blood flow recovers spontaneously, whereas others evolve a frank obstruction with extensive retinal bleeding ("blood and thunder" appearance), infarction, and visual loss. Venous occlusion of the retina is often idiopathic, but hypertension, diabetes, and glaucoma are prominent risk factors. Polycythemia, thrombocythemia, or other factors leading to an underlying hypercoagulable state should be corrected; aspirin treatment may be beneficial.

Anterior Ischemic Optic Neuropathy (AION) This is caused by insufficient blood flow through the posterior ciliary arteries that supply the optic disc. It produces painless monocular visual loss that is sudden in onset, followed sometimes by stuttering progression. The optic disc is edematous and usually bordered by nerve fiber layer splinter hemorrhages **(Fig. 34-7)**. AION is divided into two forms: arteritic and nonarteritic. The nonarteritic form is most common. No specific cause is known, although diabetes, renal failure, and hypertension are

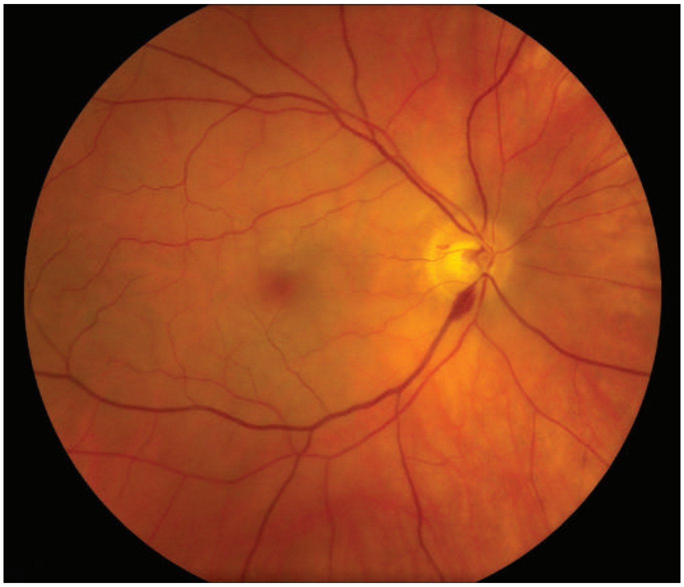

FIGURE 34-4 Central retinal artery occlusion in a 78-year-old man reducing acuity to counting fingers in the right eye. Note the splinter hemorrhage on the optic disc and the slightly milky appearance to the macula with a cherry-red fovea.

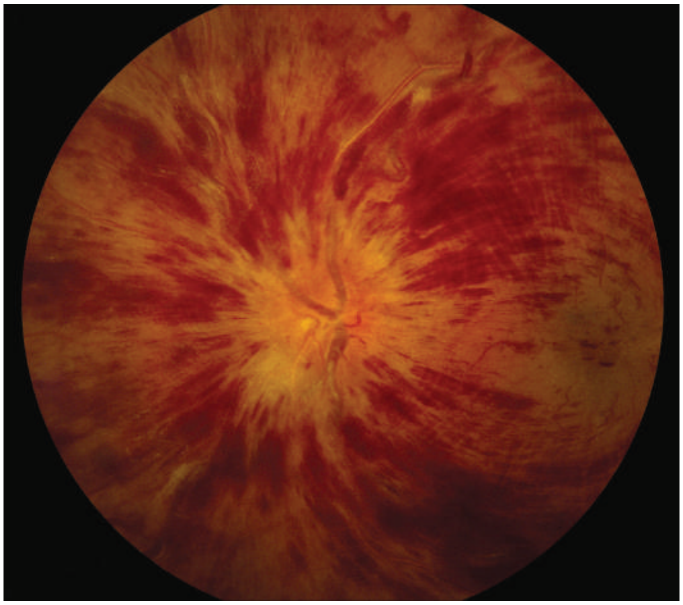

FIGURE 34-6 Central retinal vein occlusion can produce massive retinal hemorrhage ("blood and thunder"), ischemia, and vision loss.

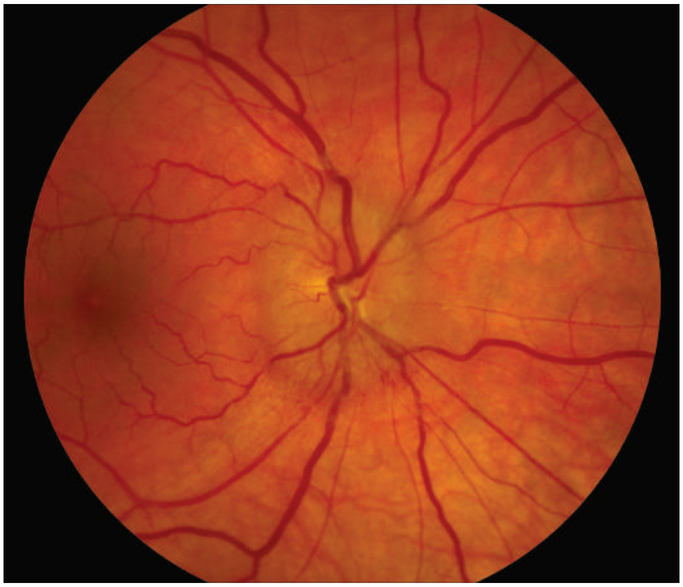

FIGURE 34-7 Anterior ischemic optic neuropathy from temporal arteritis in a 64-year-old woman with acute disc swelling, splinter hemorrhages, visual loss, and an erythrocyte sedimentation rate of 60 mm/h.

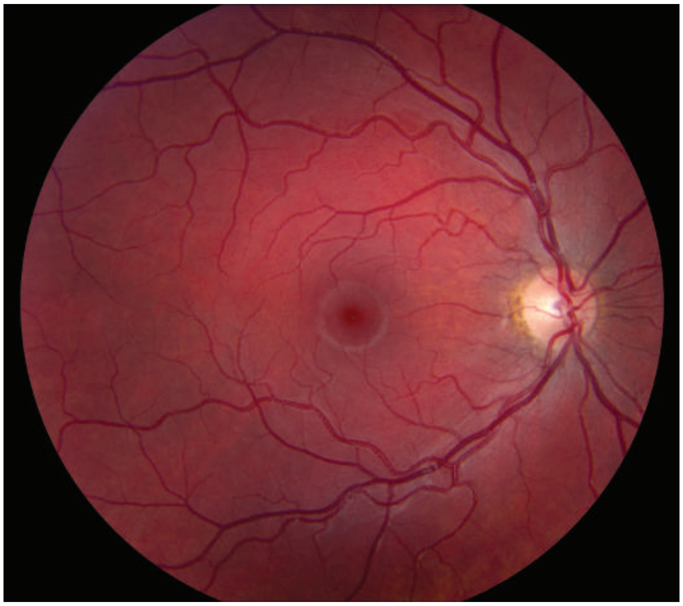

FIGURE 34-8 Retrobulbar optic neuritis is characterized by a normal fundus examination initially, hence the rubric "the doctor sees nothing, and the patient sees nothing." Optic atrophy develops after severe or repeated attacks.

common risk factors. Case reports have linked erectile dysfunction drugs to AION, but a causal association is doubtful. Evidence is strong that a crowded disc architecture and small optic cup predispose to the development of nonarteritic AION. In patients with such a "disc-at-risk," the advent of AION in one eye increases the likelihood of the same event occurring in the other eye. No treatment is available for nonarteritic AION; glucocorticoids should not be prescribed.

About 5% of patients, especially Caucasian females aged >60, have the arteritic form of AION in conjunction with giant cell (temporal) arteritis **(Chap. 375)**. It is urgent to recognize arteritic AION so that high doses of glucocorticoids can be instituted immediately to prevent blindness in the second eye. Tocilizumab, a monoclonal antibody against interleukin 6 receptor, is an effective alternative to glucocorticoids for sustained suppression of giant cell arteritis. Symptoms of polymyalgia rheumatica may be present; the sedimentation rate and C-reactive protein level are usually elevated. In a patient with visual loss from suspected arteritic AION, temporal artery biopsy is mandatory to confirm the diagnosis. Administer glucocorticoids immediately, without waiting for the biopsy to be completed. The biopsy should be obtained as soon as practical, because prolonged glucocorticoid treatment can hide inflammatory changes. It is important to harvest a long arterial segment and to examine a sufficient number of tissue sections. The histologic features of granulomatous inflammation are often quite subtle in temporal artery specimens. If an adequate biopsy is declared negative by an experienced pathologist, the diagnosis of arteritic AION is highly unlikely and glucocorticoids should usually be discontinued.

Posterior Ischemic Optic Neuropathy This is an uncommon cause of acute visual loss, induced by the combination of severe anemia and hypotension. Cases have been reported after major blood loss during surgery (especially in patients undergoing cardiac or lumbar spine operation), shock, gastrointestinal bleeding, and renal dialysis. The fundus usually appears normal, although optic disc swelling develops if the process extends anteriorly far enough to reach the globe. Vision can be salvaged in some patients by immediate blood transfusion and reversal of hypotension.

Optic Neuritis This is a common inflammatory disease of the optic nerve. In the Optic Neuritis Treatment Trial (ONTT), the mean age of patients was 32 years, 77% were female, 92% had ocular pain (especially with eye movements), and 35% had optic disc swelling. In most patients, the demyelinating event was retrobulbar and the ocular fundus appeared normal on initial examination **(Fig. 34-8)**, although optic disc pallor slowly developed over subsequent months.

Virtually all patients experience a gradual recovery of vision after a single episode of optic neuritis, even without treatment. This rule is so reliable that failure of vision to improve after a first attack of optic neuritis casts doubt on the original diagnosis. Treatment with high-dose IV methylprednisolone (250 mg every 6 h for 3 days) followed by oral prednisone (1 mg/kg per day for 11 days) makes no difference in ultimate acuity 6 months after the attack, but the recovery of visual function occurs more rapidly. Therefore, when visual loss is severe (worse than 20/100), IV followed by PO glucocorticoids are often recommended.

For some patients, optic neuritis remains an isolated event. However, the ONTT showed that the 15-year cumulative probability of developing clinically definite multiple sclerosis after optic neuritis is 50%. A brain magnetic resonance (MR) scan is advisable in every patient with a first attack of optic neuritis. If two or more plaques are present on initial imaging, treatment should be considered to prevent the development of additional demyelinating lesions **(Chap. 455)**.

A particularly severe optic neuritis, often involving a long segment of nerve, occurs in neuromyelitis optica (NMO); it may be bilateral and associated with myelitis. NMO can occur as a primary disorder, in the setting of systemic autoimmune disease, or rarely, as a paraneoplastic condition. Detection of circulating antibodies directed against aquaporin-4 or myelin oligodendrocyte glycoprotein (MOG) is diagnostic. Treatment for acute episodes consists of glucocorticoids followed by satralizumab, eculizumab, or inebilizumab to prevent relapse. **Neuromyelitis optica is discussed in detail in Chap. 456.**

■ LEBER'S HEREDITARY OPTIC NEUROPATHY

This disease usually affects young men, causing progressive, painless, severe central visual loss in one eye, followed weeks to years later by the same process in the other eye. Acutely, the optic disc appears mildly plethoric with surface capillary telangiectasias but no vascular leakage on fluorescein angiography. Eventually, optic atrophy ensues. Leber's optic neuropathy is caused by a point mutation at codon 11778 in the mitochondrial gene encoding nicotinamide adenine dinucleotide dehydrogenase (NADH) subunit 4. Additional mutations responsible for the disease have been identified, most in mitochondrial genes that encode proteins involved in electron transport. Mitochondrial mutations that cause Leber's neuropathy are maternally inherited by all children, but for unknown reasons, only 10% of cases occur in females. Clinical trials of gene therapy for this condition have been unsuccessful.

Toxic Optic Neuropathy This can result in acute visual loss with bilateral optic disc swelling and cecocentral scotomas. Cases have been

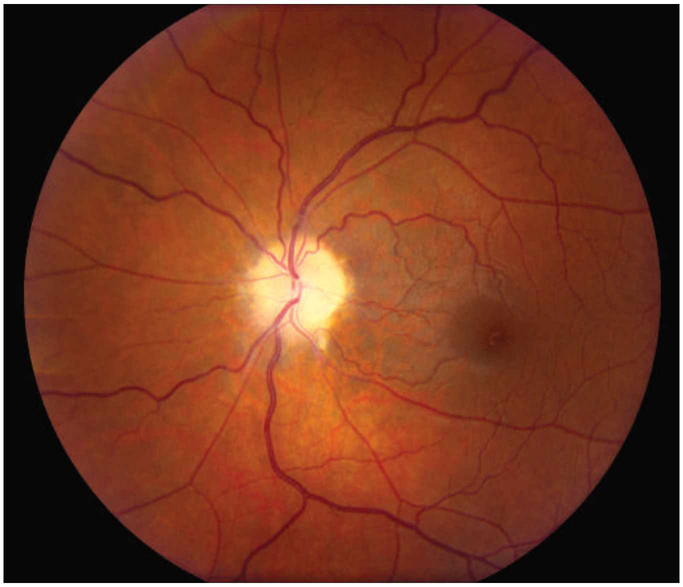

FIGURE 34-9 Optic atrophy is not a specific diagnosis but refers to the combination of optic disc pallor, arteriolar narrowing, and nerve fiber layer destruction produced by a host of eye diseases, especially optic neuropathies.

reported from exposure to methyl alcohol (moonshine) and ethylene glycol (antifreeze). More commonly, visual loss develops gradually and produces optic atrophy (Fig. 34-9) without a phase of acute optic disc edema. Ethambutol causes a dose-dependent toxic optic neuropathy in 2% of patients. Other agents have been implicated in toxic optic neuropathy, but supporting evidence is often weak. The following is a partial list of potential offending drugs or toxins: disulfiram, carbon monoxide, ethchlorvynol, chloramphenicol, amiodarone, monoclonal anti-CD3 antibody, ciprofloxacin, digitalis, streptomycin, lead, arsenic, thallium, D-penicillamine, isoniazid, emetine, and sulfonamides. Metallosis (chromium, cobalt, nickel) from hip implant failure is a rare cause of toxic optic neuropathy. Deficiency states induced by starvation, malabsorption, alcoholism, or gastric bypass can lead to insidious visual loss. Thiamine, vitamin B_{12}, and folate levels should be checked in any patient with unexplained bilateral central scotomas and optic pallor.

Papilledema This connotes bilateral optic disc swelling from raised intracranial pressure (Fig. 34-10). Headache is a common but not invariable accompaniment. All other forms of optic disc swelling (e.g., from optic neuritis or ischemic optic neuropathy) should be called "optic disc edema." This convention is arbitrary but serves to avoid confusion. Often it is difficult to differentiate papilledema from other forms of optic disc edema by fundus examination alone. Transient

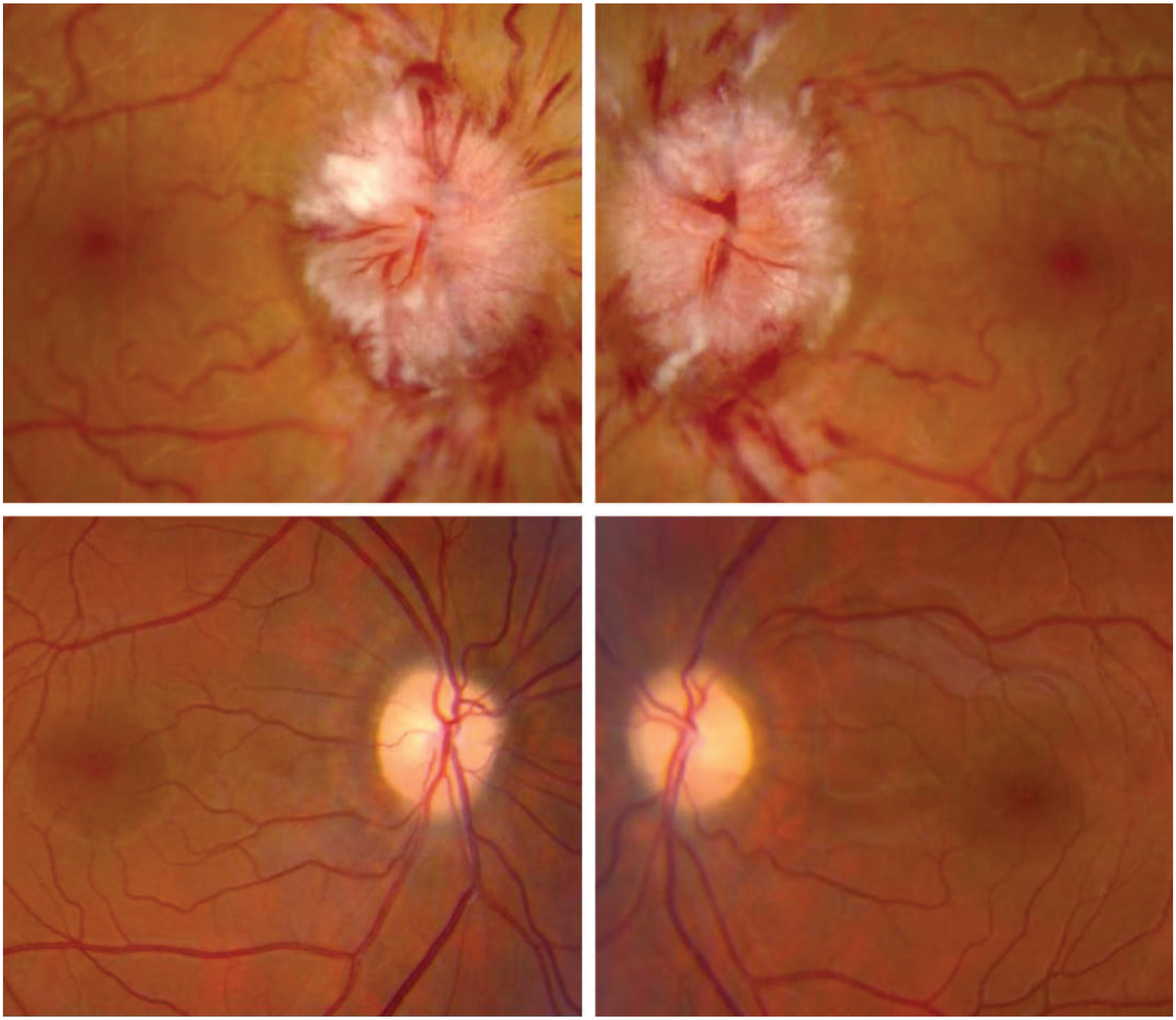

FIGURE 34-10 Papilledema in a young, obese woman with idiopathic intracranial hypertension (top), showing resolution after placement of a lumboperitoneal shunt (bottom).

visual obscurations are a classic symptom of papilledema. They occur in only one eye or simultaneously in both eyes. They usually last seconds but can persist longer. Obscurations follow abrupt shifts in posture or happen spontaneously. When obscurations are prolonged or spontaneous, the papilledema is more threatening. Visual acuity is not affected by papilledema unless the papilledema is severe, longstanding, or accompanied by macular edema and hemorrhage. Visual field testing shows enlarged blind spots and peripheral constriction **(Fig. 34-1*F*)**. With unremitting papilledema, peripheral visual field loss progresses in an insidious fashion while the optic nerve develops atrophy. In this setting, reduction of optic disc swelling is an ominous sign of a dying nerve rather than an encouraging indication of resolving papilledema.

Evaluation of papilledema requires neuroimaging to exclude an intracranial lesion. Noninvasive MR vascular imaging may be useful in selected cases to search for a dural venous sinus thrombosis or an arteriovenous shunt. If neuroradiologic studies are negative, the subarachnoid opening pressure should be measured in the lateral decubitus position by lumbar puncture. Inaccurate pressure readings are a common pitfall. An elevated pressure, with normal cerebrospinal fluid, points by exclusion to the diagnosis of *pseudotumor cerebri* (idiopathic intracranial hypertension). Almost all patients are female, and most are obese. Treatment with a carbonic anhydrase inhibitor such as acetazolamide lowers intracranial pressure by reducing the production of cerebrospinal fluid and improves the visual fields. Weight reduction is vital: treatment with a glucagon-like peptide-1 receptor agonist is recommended in patients who cannot lose weight by diet control. If vision loss is severe or progressive, a shunt (preferably lumboperitoneal) should be performed without delay to prevent blindness. Endovascular placement of a stent across the junction of the transverse and sigmoid dural sinuses, where stenosis is usually present, has emerged as a new treatment option. Optic nerve sheath fenestration is a less effective approach and does not address other neurologic symptoms, such as headache. Occasionally, fulminant papilledema produces rapid onset of blindness. In such patients, emergency surgery should be performed to install a shunt.

Optic Disc Drusen These are refractile, glittering particles within the substance of the optic nerve head **(Fig. 34-11)**. They are unrelated to drusen of the retina, which occur in age-related macular degeneration. Optic disc drusen are most common in people of northern European descent. Their diagnosis is obvious when they are visible on the surface of the optic disc. However, in many patients, they are hidden beneath the surface, producing pseudopapilledema. It is important to recognize optic disc drusen to avoid an unnecessary evaluation for papilledema. When optic disc drusen are buried, B-ultrasound is the most sensitive way to detect them. They appear hyperechoic because they contain calcium. They are also visible on computed tomography (CT) or optical coherence tomography (OCT), a technique for acquiring cross-section images of the retina. In most patients, optic disc drusen are an incidental, innocuous finding, but they can produce visual obscurations. On perimetry, they give rise to enlarged blind spots and arcuate scotomas from damage to the optic disc. With increasing age, drusen tend to become more exposed on the disc surface as optic atrophy develops. Hemorrhage, choroidal neovascular membrane, and AION are more likely to occur in patients with optic disc drusen. No treatment is available.

Vitreous Degeneration This occurs in all individuals with advancing age, leading to visual symptoms. Opacities develop in the vitreous, casting annoying shadows on the retina. As the eye moves, these distracting "floaters" move synchronously, with a slight lag caused by inertia of the vitreous gel. Vitreous traction on the retina causes mechanical stimulation, resulting in perception of flashing lights. Photopsia is brief and is confined to one eye, in contrast to the bilateral, prolonged scintillations of cortical migraine. Contraction of the vitreous can result in sudden separation from the retina, heralded by an alarming shower of floaters and photopsia. This process, known as *vitreous detachment*, is a common involutional event in the elderly. It is not harmful unless it damages the retina. A careful examination of the dilated fundus is important in any patient complaining of floaters or photopsia to search for peripheral tears or holes. If such a lesion is found, laser application can forestall a retinal detachment. Occasionally a tear ruptures a retinal blood vessel, causing vitreous hemorrhage and sudden loss of vision. On attempted ophthalmoscopy the fundus is hidden by a dark haze of blood. Ultrasound is required to examine the interior of the eye for a retinal tear or detachment. If the hemorrhage does not resolve spontaneously, the vitreous can be removed surgically. Vitreous hemorrhage also results from the fragile neovascular vessels that proliferate on the surface of the retina in diabetes, sickle cell anemia, and other ischemic ocular diseases.

Retinal Detachment This produces symptoms of floaters, flashing lights, and a scotoma in the peripheral visual field corresponding to the detachment **(Fig. 34-12)**. If the detachment includes the fovea, there is an afferent pupil defect and the visual acuity is reduced. In most eyes, retinal detachment starts with a hole, flap, or tear in the peripheral retina (rhegmatogenous retinal detachment). Patients with peripheral

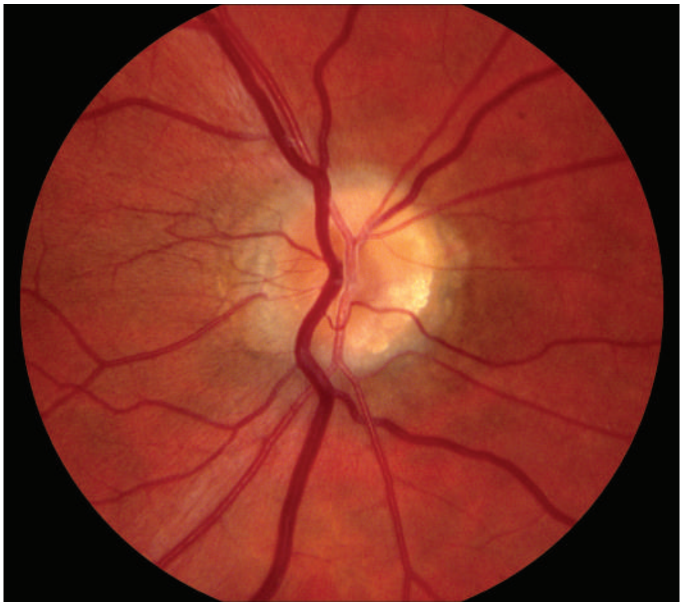

FIGURE 34-11 Optic disc drusen are calcified, mulberry-like deposits of unknown etiology within the optic disc, giving rise to "pseudopapilledema."

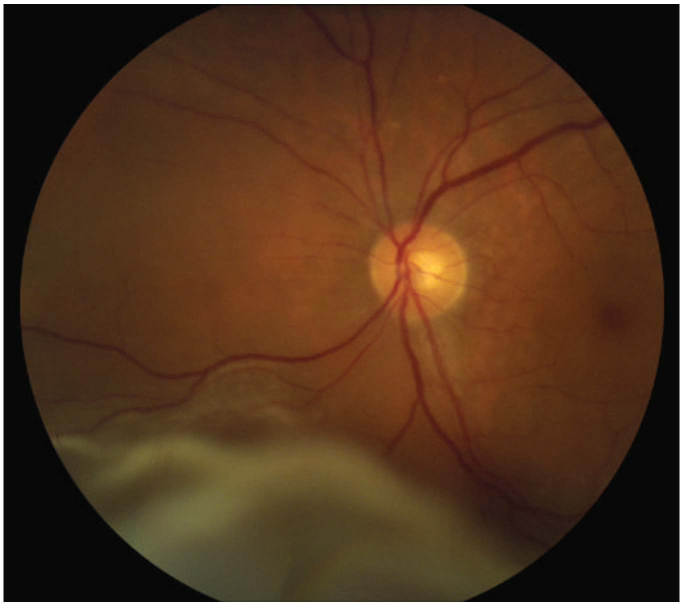

FIGURE 34-12 Retinal detachment appears as an elevated sheet of retinal tissue with folds. In this patient, the fovea was spared, so acuity was normal, but an inferior detachment produced a superior scotoma.

retinal thinning (lattice degeneration) are particularly vulnerable to this process. Once a break has developed in the retina, liquefied vitreous is free to enter the subretinal space, separating the retina from the pigment epithelium. The combination of vitreous traction on the retinal surface and passage of fluid behind the retina leads inexorably to detachment. Patients with a history of myopia, trauma, or prior cataract extraction are at greatest risk for retinal detachment. The diagnosis is confirmed by ophthalmoscopic examination of the dilated eye.

Classic Migraine **(See also Chap. 441)** This usually occurs with a visual aura lasting about 20 min. In a typical attack, a small central disturbance in the field of vision marches toward the periphery, leaving a transient scotoma in its wake. The expanding border of migraine scotoma has a scintillating, dancing, or zigzag edge, resembling the bastions of a fortified city, hence the term *fortification spectra*. Descriptions of fortification spectra vary widely and can be confused with amaurosis fugax. Migraine patterns usually last longer and are perceived in both eyes, whereas amaurosis fugax is briefer and occurs in only one eye. Migraine phenomena also remain visible in the dark or with the eyes closed. Generally, they are confined to either the right or the left visual hemifield, but sometimes, both fields are involved simultaneously. Patients often have a long history of stereotypic attacks. After the visual symptoms recede, headache develops in most patients.

Transient Ischemic Attacks Vertebrobasilar insufficiency may result in acute homonymous visual symptoms. Many patients mistakenly describe symptoms in the left or right eye when, in fact, symptoms are occurring in the left or right hemifield of both eyes. Interruption of blood supply to the visual cortex causes a sudden fogging or graying of vision, occasionally with flashing lights or other positive phenomena that mimic migraine. Cortical ischemic attacks are briefer in duration than migraine, occur in older patients, and are not followed by headache. There may be associated signs of brainstem ischemia, such as diplopia, vertigo, numbness, weakness, and dysarthria.

Stroke Permanent vision loss occurs when interruption of blood supply from the posterior cerebral artery to the visual cortex is prolonged. The only finding on examination is a homonymous visual field defect that stops abruptly at the vertical meridian. Occipital lobe stroke usually is due to thrombotic occlusion of the vertebrobasilar system, embolus, or dissection. Lobar hemorrhage, tumor, abscess, and arteriovenous malformation are other common causes of hemianopic cortical visual loss.

Factitious (Functional, Nonorganic) Visual Loss This is claimed by hysterics or malingerers. The latter account for the vast majority, seeking sympathy, special treatment, or financial gain by feigning loss of sight. The diagnosis is suspected when the history is atypical, physical findings are lacking or contradictory, inconsistencies emerge on testing, and a secondary motive can be identified. In our litigious society, the fraudulent pursuit of recompense often drives factitious visual symptoms.

CHRONIC VISUAL LOSS

Cataract Cataract is a clouding of the lens sufficient to affect vision. Most cataracts develop slowly as a result of aging, leading to gradual impairment. The formation of cataract occurs more rapidly in patients with a history of uveitis, diabetes mellitus, ocular trauma, or vitrectomy. Cataracts are acquired in a variety of genetic diseases, such as myotonic dystrophy, neurofibromatosis type 2, and galactosemia. Radiation therapy and glucocorticoid treatment can induce cataract as a side effect. Such cataracts typically have a posterior subcapsular location. Cataract can be detected by noting an impaired red reflex when viewing light reflected from the fundus with an ophthalmoscope or by examining the dilated eye with the slit lamp.

The only treatment for cataract is surgical extraction of the opacified lens. Thirty million cataract operations are performed each year around the globe. The operation generally is done under local anesthesia on an outpatient basis. A plastic or silicone intraocular lens is placed within the empty lens capsule in the posterior chamber, substituting for the natural lens and leading to rapid recovery of sight. More than 95% of patients who undergo cataract extraction can expect an improvement in vision. In some patients, the lens capsule remaining in the eye after cataract extraction eventually turns cloudy, causing secondary loss of vision. A small opening, called a posterior capsulotomy, is made in the lens capsule with a laser to restore clarity.

Glaucoma Glaucoma is a slowly progressive, insidious optic neuropathy that usually is associated with chronic elevation of intraocular pressure. After cataract, it is the most common cause of blindness in the world. It is especially prevalent in people of African descent. The mechanism by which raised intraocular pressure injures the optic nerve is not understood. Axons entering the inferotemporal and superotemporal aspects of the optic disc are damaged first, producing typical nerve fiber bundle defects called arcuate scotomas. As fibers are destroyed, the neural rim of the optic disc shrinks and the physiologic cup within the optic disc enlarges **(Fig. 34-13)**. This process is referred to as pathologic "cupping." The cup-to-disc diameter is expressed as a fraction (e.g., 0.2). The cup-to-disc ratio ranges widely in normal individuals, making it difficult to diagnose glaucoma reliably simply by observing an unusually large or deep optic cup. Careful documentation of serial examinations is helpful. In a patient with physiologic cupping, the large cup remains stable, whereas in a patient with glaucoma, it expands relentlessly over the years. Observation of progressive cupping and detection of an arcuate scotoma or a nasal step on computerized visual field testing is sufficient to establish the diagnosis of glaucoma. OCT reveals corresponding loss of fibers along the arcuate pathways in the nerve fiber layer and thinning of the ganglion cell complex.

Most patients with glaucoma have open anterior chamber angles. The cause of elevated intraocular pressure is usually unknown. Although a positive family history is a risk factor, genetic mutations impairing aqueous filtration from the eye have been identified in only a minority of cases. Surprisingly, a third of patients with open-angle glaucoma have an intraocular pressure within the normal range of 10–20 mmHg. Among patients with this normal-tension form of glaucoma, high myopia is more common.

Chronic angle-closure glaucoma and chronic open-angle glaucoma are usually asymptomatic. Only acute angle-closure glaucoma causes a red or painful eye, from abrupt elevation of intraocular pressure. In all forms of glaucoma, foveal acuity is spared until end-stage disease is reached. For these reasons, severe and irreversible damage can occur

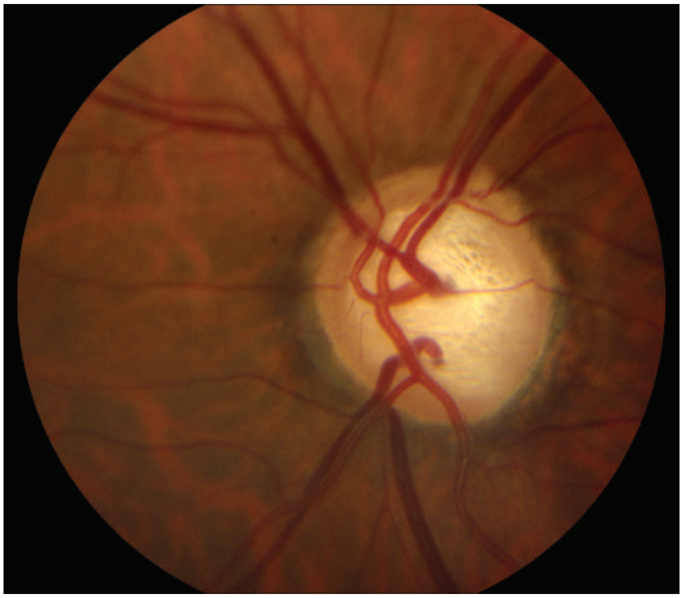

FIGURE 34-13 Glaucoma results in "cupping" as the neural rim is destroyed and the central cup becomes enlarged and excavated. The cup-to-disc ratio is about 0.8 in this patient.

before either the patient or the physician recognizes the diagnosis. Screening of patients for glaucoma by noting the cup-to-disc ratio on ophthalmoscopy and by measuring the intraocular pressure is vital. Glaucoma is treated with topical adrenergic agonists, cholinergic agonists, beta blockers, prostaglandin analogues, and carbonic anhydrase inhibitors. Occasionally, systemic absorption of beta blocker from eyedrops can be sufficient to cause side effects of bradycardia, hypotension, heart block, bronchospasm, or depression. Laser treatment of the trabecular meshwork in the anterior chamber angle improves aqueous outflow from the eye. If medical or laser treatments fail to halt optic nerve damage from glaucoma, a filter must be constructed surgically (trabeculectomy) or a drainage device placed to release aqueous from the eye in a controlled fashion.

Macular Degeneration This is a major cause of gradual, painless, bilateral central visual loss in the elderly. It occurs in a nonexudative (dry) form and an exudative (wet) form. Inflammation may be important in both forms of macular degeneration; susceptibility is associated with variants in the gene for complement factor H, an inhibitor of the alternative complement pathway. The nonexudative process begins with the accumulation of extracellular deposits called drusen underneath the retinal pigment epithelium. On ophthalmoscopy, they are pleomorphic but generally appear as small discrete yellow lesions clustered in the macula **(Fig. 34-14)**. With time, they become larger, more numerous, and confluent. The retinal pigment epithelium becomes focally detached and atrophic, causing visual loss by interfering with photoreceptor function. Treatment with vitamins C and E, beta-carotene, and zinc may slightly retard dry macular degeneration.

Exudative macular degeneration, which develops in only a minority of patients, occurs when neovascular vessels from the choroid grow through defects in Bruch's membrane and proliferate underneath the retinal pigment epithelium or the retina. Leakage from these vessels produces elevation of the retina, with distortion (metamorphopsia) and blurring of vision. Although the onset of these symptoms is usually gradual, bleeding from a subretinal choroidal neovascular membrane sometimes causes acute visual loss. Neovascular membranes can be difficult to see on fundus examination because they are located beneath the retina. Fluorescein angiography and OCT are extremely useful for their detection. Major or repeated hemorrhage under the retina from neovascular membranes results in fibrosis, development of a round (disciform) macular scar, and permanent loss of central vision.

Exudative macular degeneration can be treated with intraocular injection of antagonists to vascular endothelial growth factor. Bevacizumab, ranibizumab, aflibercept, or brolucizumab is administered by direct injection into the vitreous cavity, beginning on a monthly basis. These antibodies cause the regression of neovascular membranes by blocking the action of vascular endothelial growth factor, thereby improving visual acuity.

Central Serous Chorioretinopathy This primarily affects males between the ages of 20 and 50 years. Leakage of serous fluid from the choroid causes small, localized detachment of the retinal pigment epithelium and the neurosensory retina. These detachments produce acute or chronic symptoms of metamorphopsia and blurred vision when the macula is involved. They are difficult to visualize with a direct ophthalmoscope because the detached retina is transparent and only slightly elevated. OCT shows fluid beneath the retina, and fluorescein angiography shows dye streaming into the subretinal space. The cause of central serous chorioretinopathy is unknown. Symptoms may resolve spontaneously if the retina reattaches, but recurrent detachment is common. Laser photocoagulation has benefited some patients with this condition.

Diabetic Retinopathy A rare disease until 1921, when the discovery of insulin resulted in a dramatic improvement in life expectancy for patients with diabetes mellitus, diabetic retinopathy is now a leading cause of blindness in the United States. The retinopathy takes years to develop but eventually appears in nearly all cases. Regular surveillance of the dilated fundus is crucial for any patient with diabetes. In advanced diabetic retinopathy, the proliferation of neovascular vessels leads to blindness from vitreous hemorrhage, retinal detachment, and glaucoma **(Fig. 34-15)**. These complications can be avoided in most patients by administration of panretinal laser photocoagulation at the appropriate point in the evolution of the disease. Antivascular endothelial growth factor antibody treatment is equally effective, but intraocular injections must be given repeatedly. **For further discussion of the manifestations and management of diabetic retinopathy, see Chaps. 415–417.**

Retinitis Pigmentosa This is a general term for a disparate group of rod-cone dystrophies characterized by progressive night blindness, visual field constriction with a ring scotoma, loss of acuity, and an abnormal electroretinogram (ERG). It occurs sporadically or in an autosomal recessive, dominant, or X-linked pattern. Irregular black deposits of clumped pigment in the peripheral retina, called *bone spicules* because of their vague resemblance to the spicules of

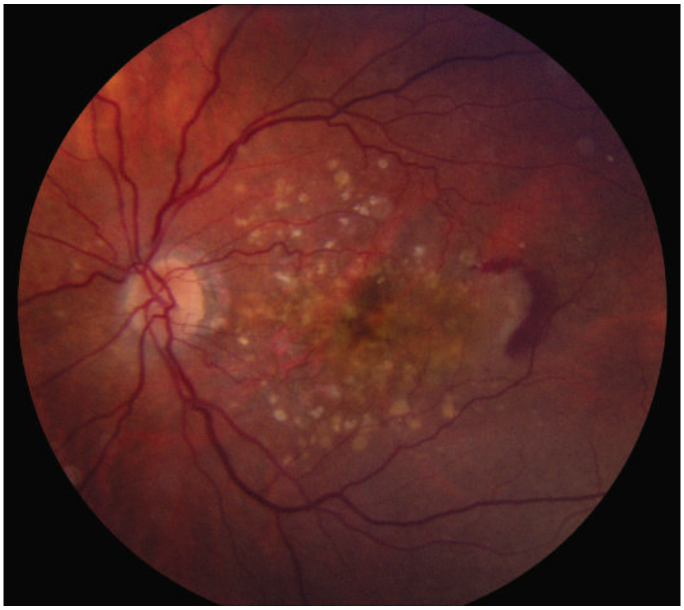

FIGURE 34-14 Age-related macular degeneration consisting of scattered yellow drusen in the macula (dry form) and a crescent of fresh hemorrhage temporal to the fovea from a subretinal neovascular membrane (wet form).

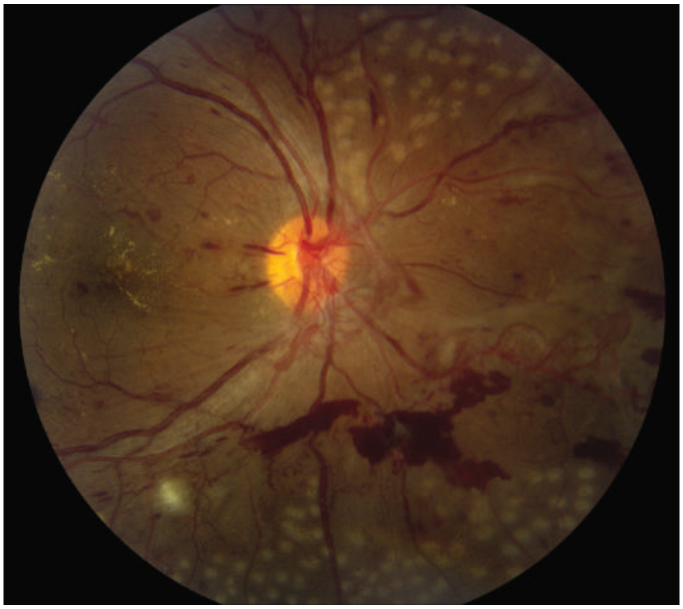

FIGURE 34-15 Proliferative diabetic retinopathy in a 25-year-old man with an 18-year history of diabetes, showing neovascular vessels emanating from the optic disc, retinal and vitreous hemorrhage, cotton-wool spots, and macular exudate. Round spots in the periphery represent recently applied panretinal photocoagulation.

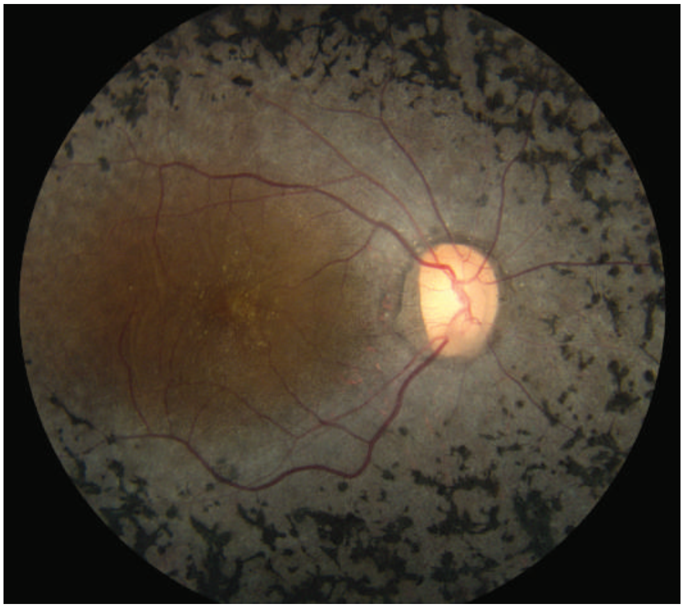

FIGURE 34-16 Retinitis pigmentosa with black clumps of pigment known as "bone spicules." The patient had peripheral visual field loss with sparing of central (macular) vision.

cancellous bone, give the disease its name (Fig. 34-16). The name is actually a misnomer because retinitis pigmentosa is not an inflammatory process. Genetic testing usually identifies a mutation in the gene for rhodopsin, the rod photopigment, or in the gene for peripherin, a glycoprotein located in photoreceptor outer segments.

Leber's congenital amaurosis, a rare cone dystrophy, has been treated by replacement of the missing RPE65 protein through gene therapy, resulting in slight improvement in visual function. Some forms of retinitis pigmentosa occur in association with rare, hereditary systemic diseases (olivopontocerebellar degeneration, Bassen-Kornzweig disease, Kearns-Sayre syndrome, Refsum's disease). Chronic treatment with chloroquine, hydroxychloroquine, and phenothiazines (especially thioridazine) can produce visual loss from a toxic retinopathy that resembles retinitis pigmentosa. Patients receiving long-term treatment with hydroxychloroquine require regular eye examinations and screening by OCT to monitor for potential development of a bull's eye maculopathy.

Epiretinal Membrane This is a fibrocellular tissue that grows across the inner surface of the retina, causing metamorphopsia and reduced visual acuity from distortion of the macula. A crinkled, cellophane-like membrane is visible on the retinal examination. Epiretinal membrane is most common in patients aged >50 years and is usually unilateral. Most cases are idiopathic, but some occur as a result of hypertensive retinopathy, diabetes, retinal detachment, or trauma. When visual acuity is reduced to the level of about 6/24 (20/80), vitrectomy and surgical peeling of the membrane to relieve macular puckering are recommended. Contraction of an epiretinal membrane sometimes gives rise to a *macular hole*. Most macular holes, however, are caused by local vitreous traction within the fovea. Vitrectomy can improve acuity in selected cases.

Melanoma and Other Tumors Melanoma is the most common primary intraocular tumor (Fig. 34-17). Approximately 3500 cases occur annually in the United States. It causes photopsia, an enlarging scotoma, and loss of vision. A small melanoma is often difficult to differentiate from a benign choroidal nevus. Serial examinations are required to document a malignant pattern of growth. Risk factors include light skin, hair, and eyes. Uveal origin accounts for 85% of cases. *GNAQ* and *GNA11* mutations are common. About half metastasize, mainly to the liver. Small and medium-sized tumors may be treated with radiation therapy; enucleation is the best treatment for large tumors. *Metastatic tumors* to the eye outnumber primary tumors.

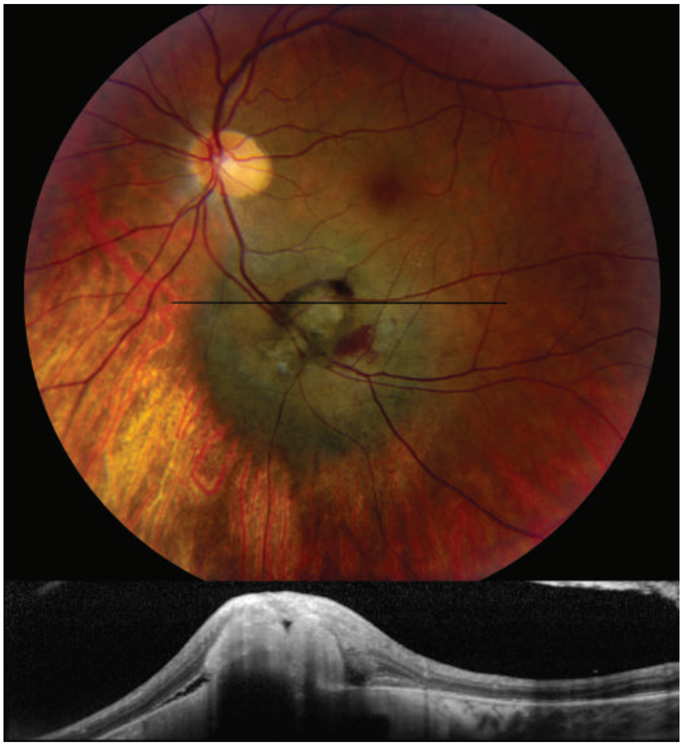

FIGURE 34-17 Melanoma of the choroid, appearing as an elevated dark mass in the inferior fundus, with overlying hemorrhage. The black line denotes the plane of the optical coherence tomography scan (*below*) showing the subretinal tumor.

Breast and lung carcinomas have a special propensity to spread to the choroid or iris. Leukemia and lymphoma also commonly invade ocular tissues. Sometimes the only sign on eye examination is cellular debris in the vitreous, which can masquerade as a chronic posterior uveitis.

In a patient with vision loss, CT or MR scanning should be considered if the cause remains unknown after careful review of the history, visual fields, and thorough examination of the eye. Optic nerve sheath meningioma is a common retrobulbar tumor. It produces the classic triad of optociliary shunt vessels, optic atrophy, and progressive visual loss. Optic disc swelling and proptosis are also frequent signs. Optic nerve glioma in young patients is usually a pilocytic astrocytoma and has a good prognosis for preservation of vision, especially in neurofibromatosis type 1 (Chap. 95). In adults, optic nerve glioma is rare and highly malignant. Chiasmal tumors (pituitary adenoma, meningioma, craniopharyngioma) produce visual loss with few objective findings except for optic disc pallor. Loss of the temporal visual field in each eye is typical, but usually patients complain of vision loss in just one eye. OCT shows loss of the retinal nerve fiber layer entering the nasal and temporal sides of the optic discs, as well as thinning of the ganglion cell complex in each nasal retina (Fig. 34-18). A high degree of vigilance is necessary to avoid missing chiasmal tumors. Although symptoms progress gradually, in rare instances, the sudden expansion of a pituitary adenoma from infarction and bleeding (*pituitary apoplexy*) causes acute severe retrobulbar visual loss, with headache, nausea, and ocular motor nerve palsies.

PROPTOSIS

When the globes appear asymmetric, the clinician must first decide which eye is abnormal. Is one eye recessed within the orbit (*enophthalmos*), or is the other eye protuberant (*exophthalmos*, or *proptosis*)? A small globe or Horner's syndrome can give the appearance of enophthalmos. True enophthalmos occurs commonly after trauma, from atrophy of retrobulbar fat, or from fracture of the orbital floor. The position of the eyes within the orbits is measured by using a Hertel exophthalmometer, a handheld instrument that records the position of the anterior corneal surface relative to the lateral orbital rim. If this instrument is not available, relative eye position can be judged by

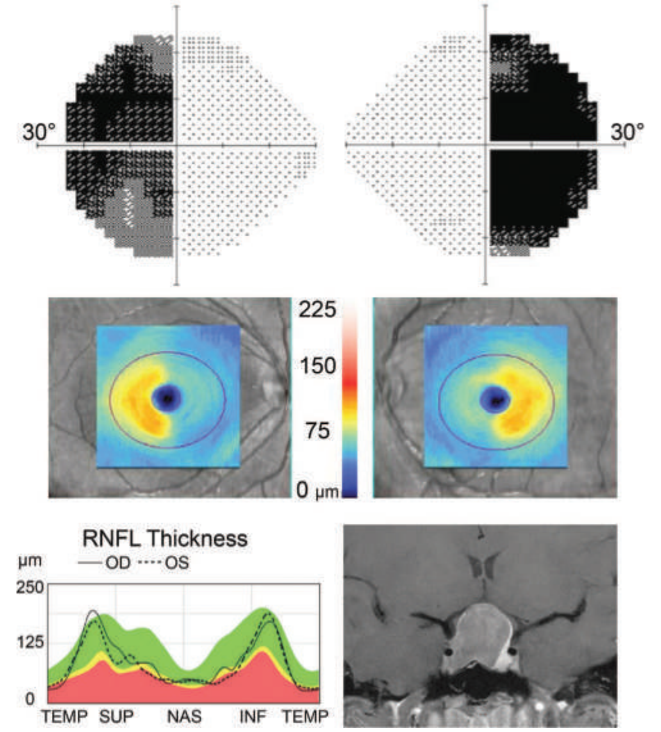

FIGURE 34-18 Bitemporal hemianopia (top), with corresponding thinning of the ganglion cell complex in the nasal maculae bilaterally (middle) and reduced retinal nerve fiber layer thickness along the temporal and nasal edges of the optic discs (red zone <1%) from compression of the optic chiasm by a pituitary adenoma (bottom).

bending the patient's head forward and looking down upon the orbits. A proptosis of only 2 mm in one eye is detectable from this perspective. The development of proptosis implies a space-occupying lesion in the orbit and may warrant CT or MR imaging.

Graves' Ophthalmopathy This is the leading cause of proptosis in adults **(Chap. 394)**. The proptosis is often asymmetric and can even appear to be unilateral. Orbital inflammation and engorgement of the extraocular muscles, particularly the medial rectus and the inferior rectus, account for protrusion of the globe. Corneal exposure, lid retraction, lid lag on downgaze, conjunctival injection, restriction of gaze, diplopia, and visual loss from optic nerve compression are cardinal symptoms. Graves' eye disease is a clinical diagnosis, but laboratory testing can be useful. The serum level of thyroid-stimulating immunoglobulins is often elevated. Orbital imaging usually reveals enlarged extraocular eye muscles, but not always. Topical lubricants, taping the eyelids closed at night, and moisture chambers are helpful to limit exposure of ocular tissues. Graves' ophthalmopathy can be treated with oral prednisone (60 mg/d) for 1 month, followed by a taper over several months, but worsening of symptoms upon glucocorticoid withdrawal is common. Infusions of teprotumumab, an inhibitor of the insulin-like growth factor I receptor, reduce proptosis and diplopia. Radiation therapy is not effective. Orbital decompression should be performed for severe, symptomatic exophthalmos or if visual function is reduced by optic nerve compression. In patients with diplopia, prisms or eye muscle surgery can be used to restore ocular alignment in primary gaze.

Orbital Pseudotumor Also known as idiopathic orbital inflammatory syndrome, orbital pseudotumor is distinguished from Graves' ophthalmopathy by the prominent complaint of pain. Other symptoms include diplopia, ptosis, proptosis, and orbital congestion. Evaluation for sarcoidosis, granulomatosis with polyangiitis, IgG4-related disease, and other types of orbital vasculitis or collagen-vascular pathology is negative. Imaging often shows swollen eye muscles (orbital myositis) with enlarged tendons. By contrast, in Graves' ophthalmopathy, the tendons of the eye muscles usually are spared. The Tolosa-Hunt syndrome **(Chap. 452)** may be regarded as an extension of orbital pseudotumor through the superior orbital fissure into the cavernous sinus. The diagnosis of orbital pseudotumor is difficult. Biopsy of the orbit frequently yields nonspecific evidence of fat infiltration by lymphocytes, plasma cells, and eosinophils. A dramatic response to a therapeutic trial of systemic glucocorticoids indirectly provides the best confirmation of the diagnosis.

Orbital Cellulitis This causes pain, lid erythema, proptosis, conjunctival chemosis, restricted motility, decreased acuity, afferent pupillary defect, fever, and leukocytosis. It often arises from the paranasal sinuses, especially by contiguous spread of infection from the ethmoid sinus through the lamina papyracea of the medial orbit. A history of recent upper respiratory tract infection, chronic sinusitis, thick mucus secretions, or dental disease is significant in any patient with suspected orbital cellulitis. Blood cultures should be obtained, but they are usually negative. Most patients respond to empirical therapy with broad-spectrum IV antibiotics. Occasionally, orbital cellulitis follows an overwhelming course, with massive proptosis, blindness, septic cavernous sinus thrombosis, and meningitis. To avert this disaster, orbital cellulitis should be managed aggressively in the early stages, with immediate imaging of the orbits and antibiotic therapy that includes coverage of methicillin-resistant *Staphylococcus aureus* (MRSA). Prompt surgical drainage of an orbital abscess or paranasal sinusitis is indicated if optic nerve function deteriorates despite antibiotics.

Tumors Tumors of the orbit cause painless, progressive proptosis. The most common primary tumors are cavernous hemangioma, lymphangioma, neurofibroma, schwannoma, dermoid cyst, adenoid cystic carcinoma, optic nerve glioma, optic nerve meningioma, and benign mixed tumor of the lacrimal gland. Metastatic tumor to the orbit occurs frequently in breast carcinoma, lung carcinoma, and lymphoma. Diagnosis by fine-needle aspiration followed by urgent radiation therapy sometimes can preserve vision.

Carotid Cavernous Fistulas With anterior drainage through the orbit, these fistulas produce proptosis, diplopia, glaucoma, and corkscrew, arterialized conjunctival vessels. Direct fistulas usually result from trauma. They are easily diagnosed because of the prominent signs produced by high-flow, high-pressure shunting. Indirect fistulas, or dural arteriovenous malformations, are more likely to occur spontaneously, especially in older women. The signs are more subtle, and the diagnosis frequently is missed. The combination of slight proptosis, diplopia, enlarged muscles, and an injected eye often is mistaken for thyroid ophthalmopathy. A bruit heard upon auscultation of the head or reported by the patient is a valuable diagnostic clue. Imaging shows an enlarged superior ophthalmic vein in the orbits. Carotid cavernous shunts can be eliminated by endovascular embolization.

■ PTOSIS

Blepharoptosis This is an abnormal drooping of the eyelid. Unilateral or bilateral ptosis can be congenital, from dysgenesis of the levator palpebrae superioris, or from abnormal insertion of its aponeurosis into the eyelid. Acquired ptosis can develop so gradually that the patient is unaware of the problem. Inspection of old photographs is helpful in dating the onset. A history of prior trauma, eye surgery, contact lens use, diplopia, systemic symptoms (e.g., dysphagia or peripheral muscle weakness), or a family history of ptosis should be sought. Fluctuating ptosis that worsens late in the day is typical of myasthenia gravis. Ptosis evaluation should focus on evidence for proptosis, eyelid masses or deformities, inflammation, pupil inequality, or limitation of motility. The width of the palpebral fissures and distance from the upper eyelid margin to corneal light reflex are measured in primary gaze to determine the degree of ptosis. The ptosis will be underestimated if the patient compensates by lifting the brow with the frontalis muscle.

Mechanical Ptosis This occurs in many elderly patients from stretching and redundancy of eyelid skin and subcutaneous fat (dermatochalasis). The extra weight of these sagging tissues causes the lid to droop. Enlargement or deformation of the eyelid from infection, tumor, trauma, or inflammation also results in ptosis on a purely mechanical basis.

Aponeurotic Ptosis This is an acquired dehiscence or stretching of the aponeurotic tendon, which connects the levator muscle to the tarsal plate of the eyelid. It occurs commonly in older patients, presumably from loss of connective tissue elasticity. Aponeurotic ptosis is also a common sequela of eyelid swelling from infection or blunt trauma to the orbit, cataract surgery, or contact lens use.

Myogenic Ptosis The causes of *myogenic ptosis* include myasthenia gravis **(Chap. 459)** and a number of rare myopathies that manifest with ptosis. The term *chronic progressive external ophthalmoplegia* refers to a spectrum of systemic diseases caused by mutations of mitochondrial DNA. As the name implies, the most prominent findings are symmetric, slowly progressive ptosis and limitation of eye movements. In general, diplopia is a late symptom because all eye movements are reduced equally. In the *Kearns-Sayre* variant, retinal pigmentary changes and abnormalities of cardiac conduction develop. Peripheral muscle biopsy shows characteristic "ragged-red fibers." *Oculopharyngeal dystrophy* is a distinct autosomal dominant disease with onset in middle age, characterized by ptosis, limited eye movements, and trouble swallowing. *Myotonic dystrophy*, another autosomal dominant disorder, causes ptosis, ophthalmoparesis, cataract, and pigmentary retinopathy. Patients have muscle wasting, myotonia, frontal balding, and cardiac abnormalities.

Neurogenic Ptosis This results from a lesion affecting the innervation to either of the two muscles that open the eyelid: Müller's muscle or the levator palpebrae superioris. Examination of the pupil helps distinguish between these two possibilities. In Horner's syndrome, the eye with ptosis has a smaller pupil and the eye movements are full. In an oculomotor nerve palsy, the eye with the ptosis has a larger or a normal pupil. If the pupil is normal but there is limitation of adduction, elevation, and depression, a pupil-sparing oculomotor nerve palsy is likely (see next section). Rarely, a lesion affecting the small, central subnucleus of the oculomotor complex will cause bilateral ptosis with normal eye movements and pupils.

■ DOUBLE VISION (DIPLOPIA)

The first point to clarify is whether diplopia persists in either eye after the opposite eye is covered. If it does, the diagnosis is monocular diplopia. The cause is usually intrinsic to the eye and therefore has no dire implications for the patient. Corneal aberrations (e.g., keratoconus, pterygium), uncorrected refractive error, cataract, or foveal traction may give rise to monocular diplopia. Occasionally, it is a symptom of malingering or psychiatric disease. Diplopia alleviated by covering one eye is binocular diplopia and is caused by disruption of ocular alignment. Inquiry should be made into the nature of the double vision (purely side-by-side vs partial vertical displacement of images), mode of onset, duration, intermittency, diurnal variation, and associated neurologic or systemic symptoms. If the patient has diplopia while being examined, motility testing should reveal a deficiency corresponding to the patient's symptoms. However, subtle limitation of ocular excursions is often difficult to detect. For example, a patient with a slight left abducens nerve paresis may appear to have full eye movements despite a complaint of horizontal diplopia upon looking to the left. In this situation, the cover test provides a more sensitive method for demonstrating the ocular misalignment. It should be conducted in primary gaze and then with the head turned and tilted in each direction while the patient fixates a central, distant target. In the above example, a cover test with the head turned to the right bringing the eyes into left gaze will maximize the fixation shift evoked by the cover test.

Occasionally, a cover test performed in an asymptomatic patient during a routine examination will reveal an ocular deviation. If the eye movements are full and the ocular misalignment is equal in all directions of gaze (comitant deviation), the diagnosis is strabismus. In this condition, which affects about 1% of the population, fusion is disrupted in infancy or early childhood. To avoid diplopia, retinal input from the nonfixating eye may be partially suppressed. In some children, this leads to impaired vision (amblyopia, or "lazy" eye) in the deviated eye.

Binocular diplopia results from a wide range of processes: infectious, neoplastic, metabolic, degenerative, inflammatory, and vascular. One must decide whether the diplopia is neurogenic in origin or is due to restriction of globe rotation by local disease in the orbit. Orbital pseudotumor, myositis, infection, tumor, thyroid disease, and muscle entrapment (e.g., from a blowout fracture) cause restrictive diplopia. The diagnosis of restriction is usually made by recognizing other associated signs and symptoms of local orbital disease. Dedicated, high-resolution orbital imaging with fat saturation and gadolinium enhancement is helpful when the cause of diplopia is not evident.

Myasthenia Gravis **(See also Chap. 459)** This is a major cause of painless diplopia. The diplopia is often intermittent, variable, and not confined to any single ocular motor nerve distribution. The pupils are always normal. Serial observation of a fatigable ptosis, often accompanied by diplopia from fluctuating ocular misalignment, establishes the diagnosis. Many patients have a purely ocular form of the disease, with no evidence of systemic muscular weakness. Classically, the diagnosis was confirmed by an IV edrophonium injection, which produces a transient reversal of eyelid or eye muscle weakness, but this drug is discontinued in the United States. Blood tests for antibodies against the acetylcholine receptor or the MuSK protein are frequently negative in the purely ocular form of myasthenia gravis. *Botulism* from food or wound poisoning can mimic ocular myasthenia.

If restrictive orbital disease and myasthenia gravis are excluded, a lesion of a cranial nerve supplying innervation to the extraocular muscles is the most likely cause of binocular diplopia.

Oculomotor Nerve The third cranial nerve innervates the medial, inferior, and superior recti; inferior oblique; levator palpebrae superioris; and the iris sphincter. Total palsy of the oculomotor nerve causes ptosis, a dilated pupil, and leaves the eye "down and out" because of the unopposed action of the lateral rectus and superior oblique. This combination of findings is obvious. More challenging is the diagnosis of early or partial oculomotor nerve palsy. In this setting, any combination of ptosis, pupil dilation, and weakness of the eye muscles supplied by the oculomotor nerve may be encountered. Frequent serial examinations during the rapidly evolving phase of the palsy help ensure that the diagnosis is not missed. The advent of an oculomotor nerve palsy with pupil involvement, especially when accompanied by pain, suggests a compressive lesion, such as a tumor or circle of Willis aneurysm. Urgent neuroimaging should be obtained, along with a CT or MR angiogram. The resolution of these noninvasive techniques has advanced to the point that catheter angiography is rarely necessary to exclude an aneurysm.

A lesion of the oculomotor nucleus in the rostral midbrain produces signs that differ from those caused by a lesion of the nerve itself. There is bilateral ptosis because the levator muscle is innervated by a single central subnucleus. There is also weakness of the contralateral superior rectus, because it is supplied by the oculomotor nucleus on the other side. Occasionally both superior recti are weak. Isolated nuclear oculomotor palsy is rare. Usually, neurologic examination reveals additional signs that suggest brainstem damage from infarction, hemorrhage, tumor, or infection.

Injury to structures surrounding fascicles of the oculomotor nerve descending through the midbrain has given rise to a number of classic eponymic designations. In *Nothnagel's syndrome*, injury to the superior cerebellar peduncle causes ipsilateral oculomotor palsy and contralateral cerebellar ataxia. In *Benedikt's syndrome*, injury to the red nucleus results in ipsilateral oculomotor palsy and contralateral tremor, chorea, and athetosis. *Claude's syndrome* incorporates features of both of these syndromes, by injury to both the red nucleus and the superior cerebellar peduncle. Finally, in *Weber's syndrome*, injury to the cerebral peduncle causes ipsilateral oculomotor palsy with contralateral hemiparesis.

In the subarachnoid space, the oculomotor nerve is vulnerable to aneurysm, meningitis, tumor, infarction, and compression. In cerebral herniation, the nerve becomes trapped between the edge of the tentorium and the uncus of the temporal lobe. Oculomotor palsy also can result from midbrain torsion and hemorrhage during herniation. In the cavernous sinus, oculomotor palsy arises from carotid aneurysm, carotid cavernous fistula, cavernous sinus thrombosis, tumor (pituitary adenoma, meningioma, metastasis), herpes zoster infection, and the Tolosa-Hunt syndrome.

The etiology of an isolated, pupil-sparing oculomotor palsy often remains an enigma even after neuroimaging and extensive laboratory testing. Most cases are thought to result from microvascular ischemia of the nerve somewhere along its course from the brainstem to the orbit. Usually, the patient complains of pain. Diabetes, hypertension, and vascular disease are major risk factors. Spontaneous recovery over a period of months is the rule. If this fails to occur or if new findings develop, the diagnosis of microvascular oculomotor nerve palsy should be reconsidered. Aberrant regeneration is common when the oculomotor nerve is injured by trauma or compression (tumor, aneurysm). Miswiring of sprouting fibers to the levator muscle and the rectus muscles results in elevation of the eyelid upon downgaze or adduction. The pupil also constricts upon attempted adduction, elevation, or depression of the globe. Aberrant regeneration is not seen after oculomotor palsy from microvascular infarct and hence vitiates that diagnosis.

Trochlear Nerve The fourth cranial nerve originates in the midbrain, just caudal to the oculomotor nerve complex. Fibers exit the brainstem dorsally and cross to innervate the contralateral superior oblique. The principal actions of this muscle are to depress and intort the globe. A palsy therefore results in hypertropia and excyclotorsion. The cyclotorsion seldom is noticed by patients. Instead, they complain of vertical diplopia, especially upon reading or looking down. Vertical diplopia is exacerbated by tilting the head toward the side with the muscle palsy and alleviated by tilting it away. This “head tilt test” is a cardinal diagnostic feature. Review of old photographs will sometimes reveal a habitual head tilt, signifying a patient with a decompensated, congenital trochlear nerve palsy.

New, isolated trochlear nerve palsy results from all the causes listed above for the oculomotor nerve except aneurysm. The trochlear nerve is particularly apt to suffer injury after closed head trauma. The free edge of the tentorium impinges on the nerve during a concussive blow. Most isolated trochlear nerve palsies are idiopathic and hence are diagnosed by exclusion as “microvascular.” Spontaneous improvement occurs over a period of months in most patients. A base-down prism (conveniently applied to the patient’s glasses as a stick-on Fresnel lens) may serve as a temporary measure to alleviate diplopia. If the palsy does not resolve, the eyes can be realigned by weakening the inferior oblique muscle.

Abducens Nerve The sixth cranial nerve innervates the lateral rectus muscle. A palsy produces horizontal diplopia, worse on gaze to the side of the lesion. A nuclear lesion has different consequences, because the abducens nucleus contains interneurons that project via the medial longitudinal fasciculus to the medial rectus subnucleus of the contralateral oculomotor complex. Therefore, an abducens nuclear lesion produces a complete lateral gaze palsy from weakness of both the ipsilateral lateral rectus and the contralateral medial rectus. *Foville’s syndrome* after dorsal pontine injury includes lateral gaze palsy, ipsilateral facial palsy, and contralateral hemiparesis incurred by damage to descending corticospinal fibers. *Millard-Gubler syndrome* from ventral pontine injury is similar except for the eye findings. There is lateral rectus weakness only, instead of gaze palsy, because the abducens fascicle is injured rather than the nucleus. Infarct, tumor, hemorrhage, vascular malformation, and multiple sclerosis are the most common etiologies of brainstem abducens palsy.

After leaving the ventral pons, the abducens nerve runs forward along the clivus to pierce the dura at the petrous apex, where it enters the cavernous sinus. Along its subarachnoid course, it is susceptible to meningitis, tumor (meningioma, chordoma, carcinomatous meningitis), subarachnoid hemorrhage, trauma, and compression by aneurysm or dolichoectatic vessels. At the petrous apex, mastoiditis can produce deafness, pain, and ipsilateral abducens palsy (*Gradenigo’s syndrome*). In the cavernous sinus, the nerve can be affected by carotid aneurysm, carotid cavernous fistula, tumor (pituitary adenoma, meningioma, nasopharyngeal carcinoma), herpes infection, and Tolosa-Hunt syndrome.

Unilateral or bilateral abducens palsy is a classic sign of raised intracranial pressure. The diagnosis can be confirmed if papilledema is observed on fundus examination. The mechanism is still debated but probably is related to rostral-caudal displacement of the brainstem. The same phenomenon accounts for abducens palsy from Chiari malformation or low intracranial pressure (e.g., after lumbar puncture, spinal anesthesia, or spontaneous dural cerebrospinal fluid leak).

Treatment of abducens palsy is aimed at prompt correction of the underlying cause. However, the cause remains obscure in many instances despite diligent evaluation. As was mentioned above for isolated trochlear or oculomotor palsy, most cases are assumed to represent microvascular infarcts because they often occur in the setting of diabetes or other vascular risk factors. Some cases may develop as a postinfectious mononeuritis (e.g., after a viral flu). Patching one eye, occluding one eyeglass lens with tape, or applying a temporary prism will provide relief of diplopia until the palsy resolves. If recovery is incomplete, eye muscle surgery nearly always can realign the eyes, at least in primary position. A patient with an abducens palsy that fails to improve should be reevaluated for an occult etiology (e.g., chordoma, carcinomatous meningitis, carotid cavernous fistula, myasthenia gravis). Skull base tumors are easily missed even on contrast-enhanced neuroimaging studies.

Multiple Ocular Motor Nerve Palsies These should not be attributed to spontaneous microvascular events affecting more than one cranial nerve at a time. This remarkable coincidence does occur, especially in diabetic patients, but the diagnosis is made only in retrospect after all other diagnostic alternatives have been exhausted. Neuroimaging should focus on the cavernous sinus, superior orbital fissure, and orbital apex, where all three ocular motor nerves are in close proximity. In a diabetic or immunocompromised host, fungal infection (*Aspergillus*, Mucorales, *Cryptococcus*) is a common cause of multiple nerve palsies. In a patient with systemic malignancy, carcinomatous meningitis is a likely diagnosis. Cytologic examination may be negative despite repeated sampling of the cerebrospinal fluid. The cancer-associated Lambert-Eaton myasthenic syndrome also can produce ophthalmoplegia. Giant cell (temporal) arteritis occasionally manifests as diplopia from ischemic palsies of extraocular muscles. Fisher’s syndrome, an ocular variant of Guillain-Barré, produces ophthalmoplegia with areflexia and ataxia. Often the ataxia is mild, and the reflexes are normal. Antiganglioside antibodies (GQ1b) can be detected in about 50% of cases.

Supranuclear Disorders of Gaze These are often mistaken for multiple ocular motor nerve palsies. For example, Wernicke’s encephalopathy can produce nystagmus and a partial deficit of horizontal and vertical gaze that mimics a combined abducens and oculomotor nerve palsy. The disorder occurs in patients who are malnourished, alcoholic, or following bariatric surgery, and can be reversed by thiamine. Infarct, hemorrhage, tumor, multiple sclerosis, encephalitis, vasculitis, and Whipple’s disease are other important causes of supranuclear gaze palsy. Disorders of vertical gaze, especially downward saccades, are an early feature of progressive supranuclear palsy. Smooth pursuit is affected later in the course of the disease. Parkinson’s disease, Huntington’s disease, and olivopontocerebellar degeneration also can affect vertical gaze.

The *frontal eye field* of the cerebral cortex is involved in generation of saccades to the contralateral side. After hemispheric stroke, the eyes usually deviate toward the lesioned side because of the unopposed action of the frontal eye field in the normal hemisphere. With time, this deficit resolves. Seizures generally have the opposite effect: the eyes deviate conjugately away from the irritative focus. *Parietal lesions* disrupt smooth pursuit of targets moving toward the side of the lesion. Bilateral parietal lesions produce *Bálint’s syndrome*, which is characterized by impaired eye-hand coordination (optic ataxia), difficulty initiating voluntary eye movements (ocular apraxia), and visuospatial disorientation (simultanagnosia).

Horizontal Gaze Descending cortical inputs mediating horizontal gaze ultimately converge at the level of the pons. Neurons in the paramedian pontine reticular formation are responsible for controlling conjugate gaze toward the same side. They project directly to the ipsilateral abducens nucleus. A lesion of either the paramedian pontine reticular formation or the abducens nucleus causes an ipsilateral conjugate gaze palsy. Lesions at either locus produce nearly identical clinical syndromes, with the following exception: vestibular stimulation (oculocephalic maneuver or caloric irrigation) will succeed in driving the eyes conjugately to the side in a patient with a lesion of the paramedian pontine reticular formation but not in a patient with a lesion of the abducens nucleus.

INTERNUCLEAR OPHTHALMOPLEGIA This results from damage to the medial longitudinal fasciculus ascending from the abducens nucleus in the pons to the oculomotor nucleus in the midbrain (hence, "internuclear"). Damage to fibers carrying the conjugate signal from abducens interneurons to the contralateral medial rectus motoneurons results in a failure of adduction on attempted lateral gaze. For example, a patient with a left internuclear ophthalmoplegia (INO) will have slowed or absent adducting movements of the left eye **(Fig. 34-19)**. A patient with bilateral injury to the medial longitudinal fasciculus will have bilateral INO. Multiple sclerosis is the most common cause, although tumor, stroke, trauma, or any brainstem process may be responsible. *One-and-a-half syndrome* is due to a lesion of the medial longitudinal fasciculus combined with a lesion of either the abducens nucleus or the paramedian pontine reticular formation on the same side. The patient's only horizontal eye movement is abduction of the eye on the other side.

Vertical Gaze Midbrain lesions of the rostral interstitial nucleus of the medial longitudinal fasciculus and the interstitial nucleus of Cajal cause supranuclear paresis of upgaze, downgaze, or all vertical eye movements. Distal basilar artery ischemia is the most common etiology. *Skew deviation* refers to a vertical misalignment of the eyes, usually constant in all positions of gaze. The finding has poor localizing value because skew deviation has been reported after lesions in widespread regions of the brainstem and cerebellum.

PARINAUD'S SYNDROME Also known as dorsal midbrain syndrome, this is a distinct supranuclear vertical gaze disorder caused by damage to the posterior commissure. It is a classic sign of hydrocephalus from aqueductal stenosis. Pineal region or midbrain tumors, cysticercosis, and stroke also cause Parinaud's syndrome. Features include loss of upgaze (and sometimes downgaze), convergence-retraction nystagmus on attempted upgaze, downward ocular deviation ("setting sun" sign), lid retraction (Collier's sign), skew deviation, pseudoabducens palsy, and light-near dissociation of the pupils.

Nystagmus This is a rhythmic oscillation of the eyes, occurring physiologically from vestibular and optokinetic stimulation or pathologically in a wide variety of diseases **(Chap. 24)**. Abnormalities of the eyes or optic nerves, present at birth or acquired in childhood, can produce a complex, searching nystagmus with irregular pendular (sinusoidal) and jerk features. Examples are albinism, Leber's congenital amaurosis, and bilateral cataract. This nystagmus is commonly referred to as *congenital sensory nystagmus*. This is a poor term because even in children with congenital lesions, the nystagmus does not appear until weeks after birth. *Congenital motor nystagmus*, which looks similar to congenital sensory nystagmus, develops in the absence of any abnormality of the sensory visual system. Visual acuity usually is also reduced, probably by the nystagmus itself, but seldom below a level of 20/200.

JERK NYSTAGMUS This is characterized by a slow drift off the target, followed by a fast corrective saccade. By convention, the nystagmus is named after the quick phase. Jerk nystagmus can be downbeat, upbeat, horizontal (left or right), and torsional. The pattern of nystagmus may vary with gaze position. Some patients will be oblivious to their nystagmus. Others will complain of blurred vision or a subjective to-and-fro movement of the environment (oscillopsia) corresponding to the nystagmus. Fine nystagmus may be difficult to see on gross examination of the eyes. Observation of nystagmoid movements of the optic disc on ophthalmoscopy is a sensitive way to detect subtle nystagmus.

GAZE-EVOKED NYSTAGMUS This is the most common form of jerk nystagmus. When the eyes are held eccentrically in the orbits, they have a natural tendency to drift back to primary position. The subject compensates by making a corrective saccade to maintain the deviated eye position. Many normal patients have mild gaze-evoked nystagmus.

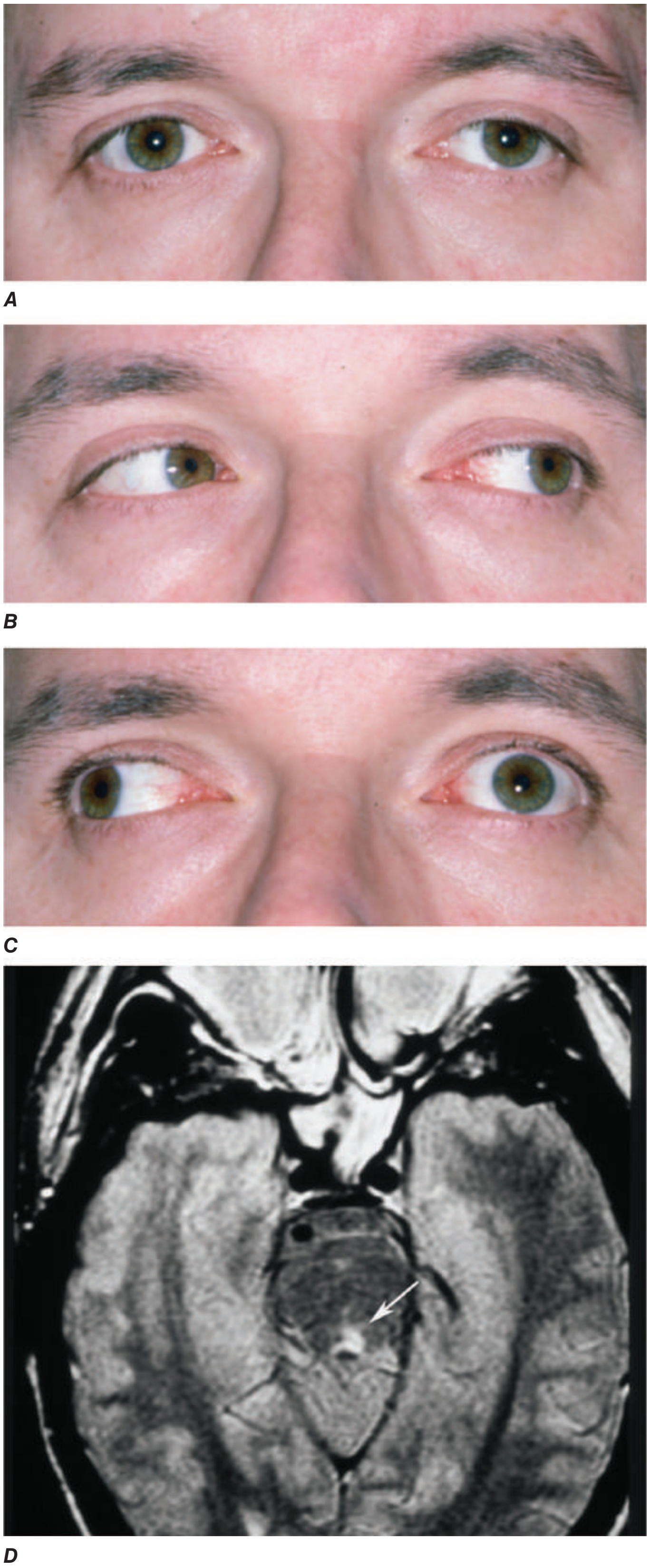

FIGURE 34-19 Left internuclear ophthalmoplegia (INO). ***A.*** In primary position of gaze, the eyes appear normal. ***B.*** Horizontal gaze to the left is intact. ***C.*** On attempted horizontal gaze to the right, the left eye fails to adduct. In mildly affected patients, the eye may adduct partially or more slowly than normal. Nystagmus is usually present in the abducted eye. ***D.*** T2-weighted axial magnetic resonance image through the pons showing a demyelinating plaque in the left medial longitudinal fasciculus (*arrow*).

Exaggerated gaze-evoked nystagmus can be induced by drugs (sedatives, anticonvulsants, alcohol); muscle paresis; myasthenia gravis; demyelinating disease; and cerebellopontine angle, brainstem, and cerebellar lesions.

VESTIBULAR NYSTAGMUS *Vestibular nystagmus* results from dysfunction of the labyrinth (Ménière's disease, benign paroxysmal positional vertigo), vestibular nerve, or vestibular nucleus in the brainstem. Peripheral vestibular nystagmus often occurs in discrete attacks, with symptoms of nausea and vertigo. There may be associated tinnitus and hearing loss. Sudden shifts in head position may provoke or exacerbate symptoms.

DOWNBEAT NYSTAGMUS *Downbeat nystagmus* results from lesions near the craniocervical junction (Chiari malformation, basilar invagination). It also has been reported in brainstem or cerebellar stroke, lithium or anticonvulsant intoxication, alcoholism, and multiple sclerosis. *Upbeat nystagmus* is associated with damage to the pontine tegmentum from stroke, demyelination, or tumor.

Opsoclonus This rare, dramatic disorder of eye movements consists of bursts of consecutive saccades (saccadomania). When the saccades are confined to the horizontal plane, the term *ocular flutter* is preferred. It can result from viral encephalitis, trauma, or a paraneoplastic effect of neuroblastoma, breast carcinoma, and other malignancies. It has also been reported as a benign, transient phenomenon in otherwise healthy patients.

FURTHER READING

Albert DM et al: *Albert and Jakobiec's Principles and Practice of Ophthalmology*, 4th ed. New York, Springer Link, 2022.

Apte RS: Age-related macular degeneration. N Engl J Med 385:540, 2021.

Chen X et al: Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex. Science 370:1191, 2020.

Durand ML et al: Infectious keratitis. JAMA 326:1319, 2021.

Heier JS et al: Pegcetacoplan for the treatment of geographic atrophy secondary to age-related macular degeneration (OAKS and DERBY): Two multicentre, randomised, double-masked, sham-controlled, phase 3 trials. Lancet 402:1434, 2023.

Jhaveri CD: Aflibercept monotherapy or bevacizumab first for diabetic macular edema. N Engl J Med 387:692, 2022.

Sahel JA: Partial recovery of visual function in a blind patient after optogenetic therapy. Nature Med 27:1223, 2021.

Toth CA: Optical coherence tomography and eye care. N Engl J Med 389:1526, 2023.

Yanoff M, Duker J: *Ophthalmology*, 6th ed. Singapore, Elsevier, 2023.

35

Disorders of Smell and Taste

Richard L. Doty, Steven M. Bromley

All environmental chemicals necessary for life enter the body by the nose and mouth. The senses of smell (olfaction) and taste (gustation) monitor such chemicals, determine the flavor and palatability of foods and beverages, and warn of dangerous environmental conditions, including fire, air pollution, leaking natural gas, and bacteria-laden foodstuffs. These senses contribute significantly to quality of life and, when dysfunctional, can have untoward physical and psychological consequences. A longitudinal study of 1162 nondemented elderly persons found, even after controlling for confounders, that those with the lowest baseline olfactory test scores had a 45% mortality rate over a 4-year period, compared to an 18% mortality rate for those with the highest olfactory test scores. A basic understanding of these senses in health and disease is critical for the physician because thousands of patients present to doctors' offices each year with complaints of chemosensory dysfunction. Among the more important recent developments in neurology is the discovery that decreased smell function is among the first signs of such neurodegenerative diseases as Parkinson's disease (PD) and Alzheimer's disease (AD), signifying their "presymptomatic" phase.

ANATOMY AND PHYSIOLOGY

Olfactory System Odorous chemicals enter the front of nose during inhalation and active sniffing, as well as the back of the nose (nasopharynx) during deglutition. After reaching the highest recesses of the nasal cavity, they dissolve in the olfactory mucus and diffuse or are actively transported by specialized proteins to receptors located on the cilia of olfactory receptor cells. The cilia, dendrites, cell bodies, and proximal axonal segments of these bipolar cells are located within a unique neuroepithelium covering the cribriform plate, the superior nasal septum, superior turbinate, and sectors of the middle turbinate **(Fig. 35-1)**. Nearly 400 types of G-protein-coupled odor receptors (GPCRs) are expressed on the cilia of the receptor cells, with only one type of GPCR being expressed on a given cell. Other receptors, including trace amine-associated receptors and members of the non-GPCR membrane-spanning 4-domain family, subfamily A (MS4A) protein family, are also present on some receptor cells. Such a plethora of receptor cell types does not exist in any other sensory system. Importantly, when damaged, the receptor cells can be replaced by stem cells near the basement membrane, although such replacement is often incomplete.

After coalescing into bundles surrounded by glia-like ensheathing cells (termed fila), the receptor cell axons pass through the cribriform plate to the olfactory bulbs, where they synapse with dendrites of other cell types within the glomeruli **(Fig. 35-2)**. These spherical structures, which make up a distinct layer of the olfactory bulb, are a site of convergence of information, because many more fibers enter than leave them. Receptor cells that express the same type of receptor project to the same glomeruli, effectively making each glomerulus a functional unit. The major projection neurons of the olfactory system—the mitral and tufted cells—send primary dendrites into the glomeruli, connecting not only with the incoming receptor cell axons, but with dendrites of periglomerular cells. The activity of the mitral/tufted cells is modulated by the periglomerular cells, secondary dendrites from other mitral/tufted cells, and granule cells, the most numerous cells of the bulb. The latter cells, which are largely GABAergic, receive inputs from central brain structures and modulate the output of the mitral/tufted cells. Interestingly, like the olfactory receptor cells, some cells within the bulb undergo replacement. Thus, neuroblasts formed within the anterior subventricular zone of the brain migrate along the rostral migratory stream, ultimately becoming granule and periglomerular cells.

The axons of the mitral and tufted cells synapse within secondary olfactory structures, which largely compose the primary olfactory cortex (POC) **(Fig. 35-3)**. The POC is defined as those cortical structures that receive direct projections from the olfactory bulb, most notably the piriform and entorhinal cortices. Although olfaction is unique in that its initial afferent projections bypass the thalamus, persons with damage to the thalamus can exhibit olfactory deficits, particularly ones of odor identification. Such deficits likely reflect the involvement of thalamic connections between the POC and the orbitofrontal cortex (OFC), where odor identification largely occurs. The close anatomic ties between the olfactory system and the amygdala, hippocampus, and hypothalamus help to explain the intimate associations between odor perception and cognitive functions such as memory, motivation, arousal, autonomic activity, digestion, and sex.

Taste System Tastants are sensed by specialized receptor cells present within taste buds—small grapefruit-like segmented structures located on the lateral margins and dorsum of the tongue, roof of the mouth, pharynx, larynx, and superior esophagus **(Fig. 35-4)**. Lingual taste buds are embedded in well-defined protuberances, termed

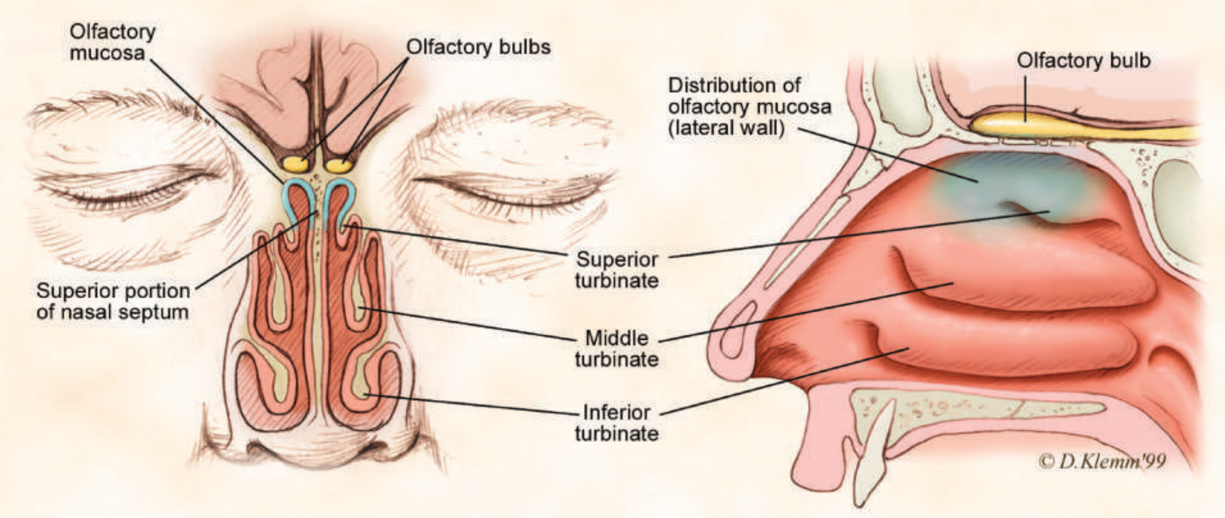
Olfactory mucosa
Olfactory bulbs
Distribution of olfactory mucosa (lateral wall)
Olfactory bulb
Superior portion of nasal septum
Superior turbinate
Middle turbinate
Inferior turbinate
© D. Klemm'99

FIGURE 35-1 Anatomy of the nose, showing the distribution of olfactory receptors in the roof of the nasal cavity. *(Copyright David Klemm, Faculty and Curriculum Support [FACS], Georgetown University Medical Center.)*

fungiform, foliate, and circumvallate papillae. After dissolving in a liquid, tastants enter the opening of the taste bud—the taste pore—and bind to receptors on microvilli, small extensions of receptor cells within each taste bud. Such binding changes the electrical potential across the taste cell, resulting in neurotransmitter release onto the first-order taste neurons. Although humans have ~7500 taste buds, not all harbor taste-sensitive cells; some contain only one class of receptor (e.g., cells responsive only to sugars), whereas others contain cells sensitive to more than one class. The number of taste receptor cells per taste bud ranges from zero to well over 100. A small family of three GPCRs, namely T1R1, T1R2, and T1R3, mediate sweet and umami taste sensations. Bitter sensations, on the other hand, depend on T2R receptors, a family of ~30 GPCRs expressed on cells different from those that express the sweet and umami receptors. T2Rs sense a wide range of bitter substances but do not distinguish among them. Sour tastants are sensed by the PKD2L1 receptor, a member of the transient receptor potential protein (TRP) family. Perception of salty sensations, such as induced by sodium chloride, arises from the entry of Na^+ ions into the cells via specialized membrane channels, such as the amiloride-sensitive Na^+ channel.

It is now well established that both bitter and sweet taste-related receptors are also present elsewhere in the body, most notably in the alimentary and respiratory tracts. This important discovery generalizes the concept of taste-related chemoreception to areas of the body beyond the mouth and throat, with α-gustducin, the taste-specific G-protein α-subunit, expressed in so-called brush cells found specifically within the human trachea, lung, pancreas, and gallbladder. These brush cells are rich in nitric oxide (NO) synthase, known to defend against xenobiotic organisms, protect the mucosa from acid-induced

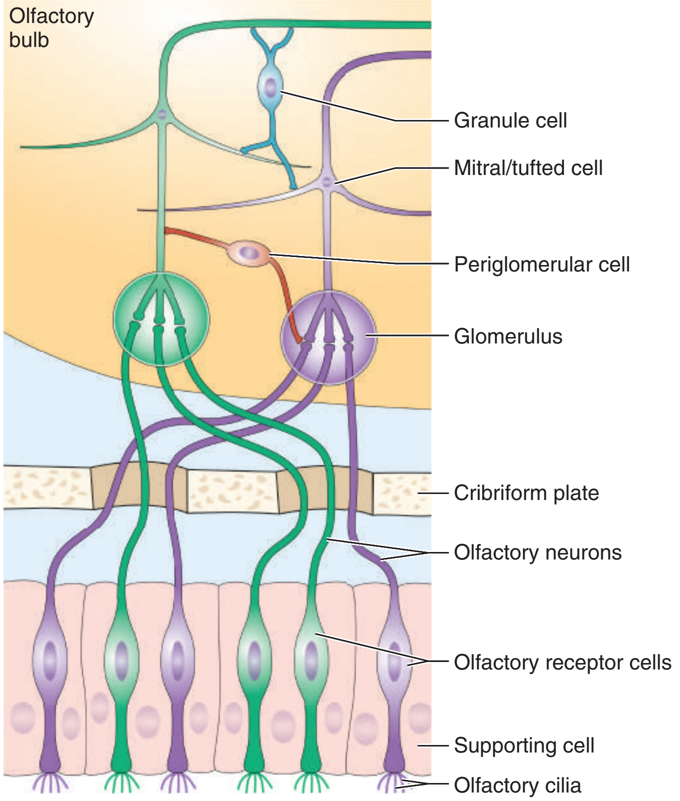
Olfactory bulb
Granule cell
Mitral/tufted cell
Periglomerular cell
Glomerulus
Cribriform plate
Olfactory neurons
Olfactory receptor cells
Supporting cell
Olfactory cilia

FIGURE 35-2 Schematic of the layers and wiring of the olfactory bulb. Each receptor type (*red, green, blue*) projects to a common glomerulus. The neural activity within each glomerulus is modulated by periglomerular cells. The activity of the primary projection cells, the mitral and tufted cells, is modulated by granule cells, periglomerular cells, and secondary dendrites from adjacent mitral and tufted cells. *(Adapted from https://medicine.yale.edu/.)*

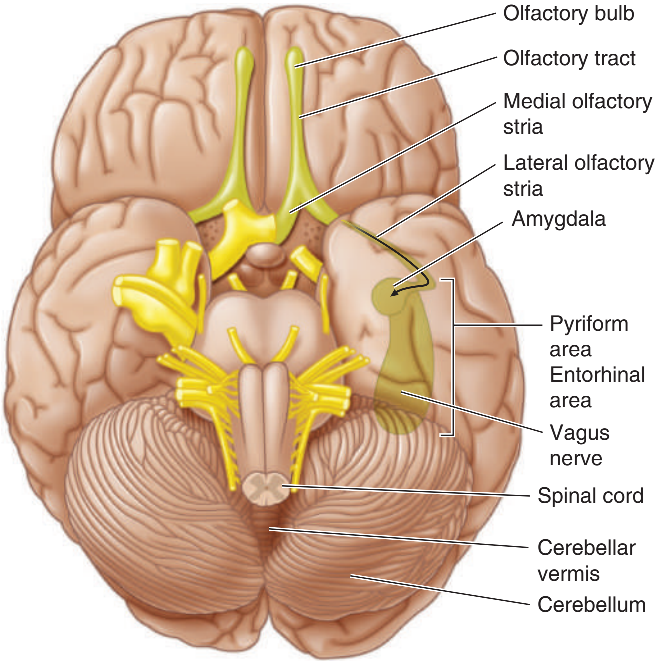
Olfactory bulb
Olfactory tract
Medial olfactory stria
Lateral olfactory stria
Amygdala
Pyriform area
Entorhinal area
Vagus nerve
Spinal cord
Cerebellar vermis
Cerebellum

FIGURE 35-3 Anatomy of the base of the brain showing the primary olfactory cortex.

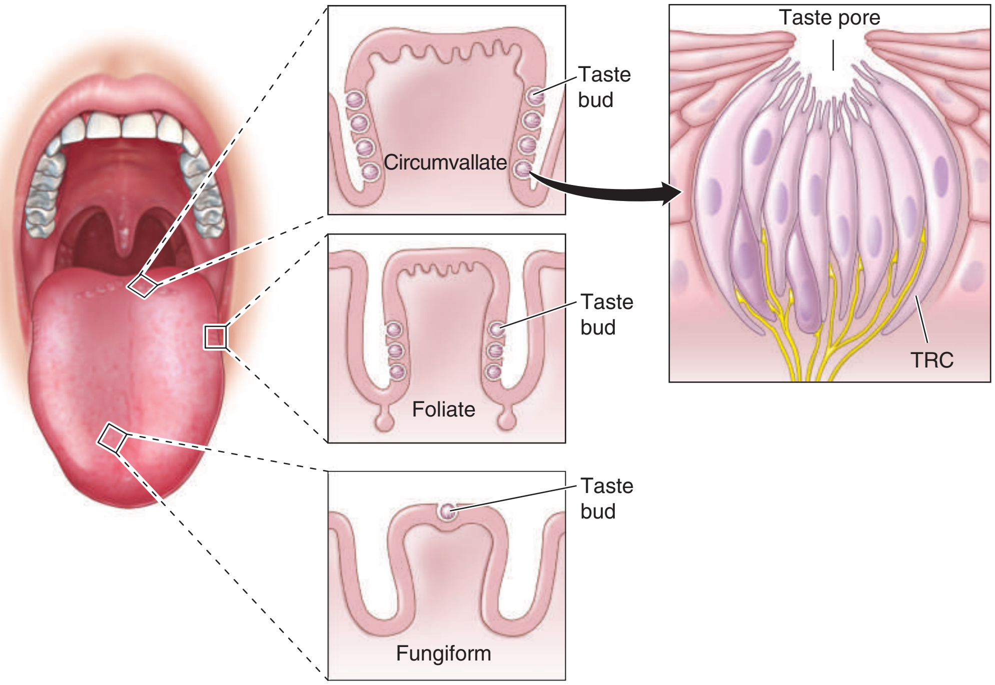

FIGURE 35-4 Schematic of the taste bud and its opening (pore), as well as the location of buds on the three major types of papillae: fungiform (anterior), foliate (lateral), and circumvallate (posterior). TRC, taste receptor cell.

lesions, and, in the case of the gastrointestinal tract, stimulate vagal and splanchnic afferent neurons. NO further acts on nearby cells, including enteroendocrine cells, absorptive or secretory epithelial cells, mucosal blood vessels, and cells of the immune system. Members of the T2R family of bitter receptors and the sweet receptors of the T1R family have been identified within the gastrointestinal tract and in enteroendocrine cell lines. In some cases, these receptors are important for metabolism, with the T1R3 receptors and gustducin playing decisive roles in the sensing and transport of dietary sugars from the intestinal lumen into absorptive enterocytes via a sodium-dependent glucose transporter and in regulation of hormone release from gut enteroendocrine cells. In other cases, these receptors may be important for airway protection, with a number of T2R bitter receptors in the motile cilia of the human airway that respond to bitter compounds by increasing their beat frequency. One specific T2R38 taste receptor is expressed in human upper respiratory epithelia and responds to acyl-monoserine lactone quorum-sensing molecules secreted by *Pseudomonas aeruginosa* and other gram-negative bacteria. Differences in T2R38 functionality, as related to TAS2R38 genotype, correlate with susceptibility to upper respiratory infections in humans.

Taste information is sent to the brain via three cranial nerves (CNs): CN VII (the *facial nerve*, which involves the intermediate nerve with its branches, the greater petrosal and chorda tympani nerves), CN IX (the *glossopharyngeal nerve*), and CN X (the *vagus nerve*) **(Fig. 35-5)**. CN VII innervates the anterior tongue and all of the soft palate, CN IX innervates the posterior tongue, and CN X innervates the laryngeal surface of the epiglottis, larynx, and proximal portion of the esophagus. The mandibular branch of CN V (V_3) conveys somatosensory information (e.g., touch, burning, cooling, irritation) to the brain. Although not technically a gustatory nerve, CN V shares primary nerve routes with many of the gustatory nerve fibers and adds temperature, texture, pungency, and spiciness to the taste experience. The chorda tympani nerve is famous for taking a recurrent course through the facial canal in the petrosal portion of the temporal bone, passing through the middle ear, and then exiting the skull via the petrotympanic fissure, where it joins the lingual nerve (a division of CN V) near the tongue. This nerve also carries parasympathetic fibers to the submandibular and sublingual glands, whereas the greater petrosal nerve supplies the palatine glands, thereby influencing saliva production.

The axons of the projection cells, which synapse with taste buds, enter the rostral portion of the nucleus of the solitary tract (NTS) within the medulla of the brainstem (Fig. 35-5). From the NTS, neurons then project to a division of the ventroposteromedial thalamic nucleus (VPM) via the medial lemniscus. From here, projections are made to the rostral part of the frontal operculum and adjoining insula, a brain region considered the *primary taste cortex* (PTC). Projections from the PTC then go to the *secondary taste cortex*, namely the caudolateral OFC. This brain region is involved in the conscious recognition of taste qualities. Moreover, because it contains cells that are activated by several sensory modalities, it is likely a center for establishing "flavor."

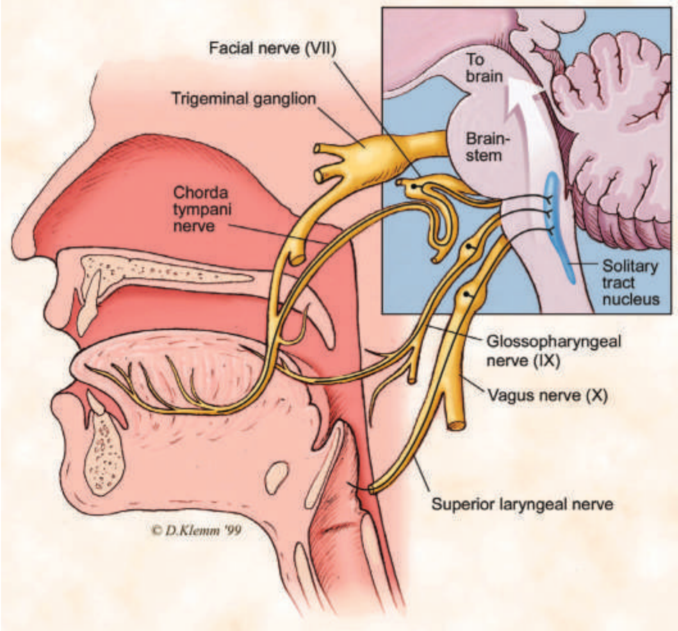

FIGURE 35-5 Schematic of the cranial nerves (CNs) that mediate taste function, including the chorda tympani nerve (CN VII), the glossopharyngeal nerve (CN IX), and the vagus nerve (CN X). *(Copyright David Klemm, Faculty and Curriculum Support [FACS], Georgetown University Medical Center.)*

DISORDERS OF OLFACTION

The ability to smell is influenced, in everyday life, by such factors as age, gender, general health, nutrition, smoking, and reproductive state. Women typically outperform men on tests of olfactory function and retain normal smell function to a later age than do men.

Estimates of the prevalence of olfactory dysfunction in the general population vary; a cross-sectional analysis from the National Health and Nutrition Examination Survey (NHANES 2013–2014) found an overall prevalence of 13.5%. However, it is apparent that significant decrements in the ability to smell are present in >50% of the population between 65 and 80 years of age and in 75% of those aged ≥80 years **(Fig. 35-6)**. Such presbyosmia helps to explain why many elderly patients report that food has little flavor, a problem that can result in nutritional disturbances. This also helps to explain why a disproportionate number of elderly people die in accidental gas poisonings. A relatively complete listing of conditions and disorders that have been associated with olfactory dysfunction is presented in **Table 35-1**.

Aside from aging, the three most common identifiable causes of long-lasting or permanent smell loss seen in the clinic are, in order of frequency, severe upper respiratory infections, head trauma, and chronic rhinosinusitis. The physiologic basis for most head trauma–related losses is the shearing and subsequent scarring of the olfactory fila as they pass from the nasal cavity into the brain cavity. The cribriform plate does not have to be fractured or show pathology for smell loss to be present. Severity of trauma, as indexed by a poor Glasgow Coma Scale score on presentation and the length of posttraumatic amnesia, is associated with higher risk of olfactory impairment. Less than 10% of posttraumatic anosmic patients will recover age-related normal function over time. This increases to nearly 25% of those with less-than-total loss. Respiratory infections, such as those associated with the common cold, influenza, pneumonia, HIV, and COVID-19, can directly and permanently damage the olfactory epithelium, decreasing receptor cell number, damaging cilia on remaining receptor cells, and inducing the replacement of sensory epithelium with respiratory epithelium. The smell loss associated with chronic rhinosinusitis is related to disease severity, with most loss occurring in cases where rhinosinusitis and polyposis are both present. Smell loss is among the first signs of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, which is responsible for COVID-19. In most cases, this loss is independent of nasal inflammation. Many of those afflicted are unaware of their deficit until objectively tested. Failure to regain normal olfactory function occurs in up to 30% even a year after

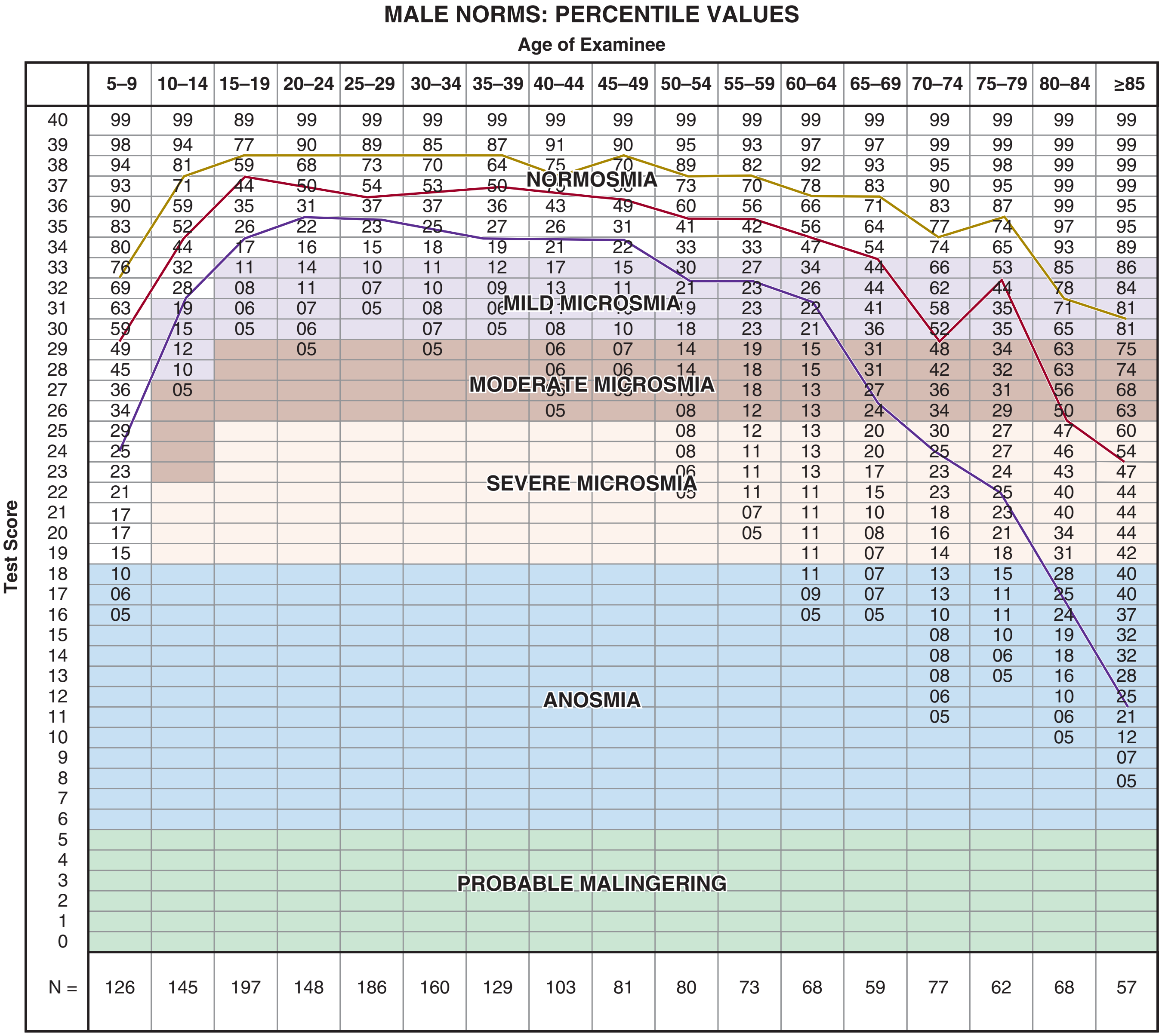

MALE NORMS: PERCENTILE VALUES

Age of Examinee

Test Score	5–9	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	≥85
40	99	99	89	99	99	99	99	99	99	99	99	99	99	99	99	99	99
39	98	94	77	90	89	85	87	91	90	95	93	97	97	99	99	99	99
38	94	81	59	68	73	70	64	75	70	89	82	92	93	95	98	99	99
37	93	71	44	50	54	53	50	75	55	73	70	78	83	90	95	99	99
36	90	59	35	31	37	37	36	43	49	60	56	66	71	83	87	99	95
35	83	52	26	22	23	25	27	26	31	41	42	56	64	77	74	97	95
34	80	44	17	16	15	18	19	21	22	33	33	47	54	74	65	93	89
33	76	32	11	14	10	11	12	17	15	30	27	34	44	66	53	85	86
32	69	28	08	11	07	10	09	13	11	21	23	26	44	62	44	78	84
31	63	19	06	07	05	08	06	11	10	19	23	22	41	58	35	71	81
30	59	15	05	06		07	05	08	10	18	23	21	36	52	35	65	81
29	49	12		05		05		06	07	14	19	15	31	48	34	63	75
28	45	10						06	06	14	18	15	31	42	32	63	74
27	36	05						05	05	10	18	13	27	36	31	56	68
26	34							05		08	12	13	24	34	29	50	63
25	29									08	12	13	20	30	27	47	60
24	25									08	11	13	20	25	27	46	54
23	23									06	11	13	17	23	24	43	47
22	21									05	11	11	15	23	25	40	44
21	17										07	11	10	18	23	40	44
20	17										05	11	08	16	21	34	44
19	15											11	07	14	18	31	42
18	10											11	07	13	15	28	40
17	06											09	07	13	11	25	40
16	05											05	05	10	11	24	37
15														08	10	19	32
14														08	06	18	32
13														08	05	16	28
12														06		10	25
11														05		06	21
10																05	12
9																	07
8																	05
7																	
6																	
5																	
4																	
3																	
2																	
1																	
0																	
N =	126	145	197	148	186	160	129	103	81	80	73	68	59	77	62	68	57

FIGURE 35-6 Scores on the University of Pennsylvania Smell Identification Test (UPSIT) as a function of subject age and sex. Numbers by each data point indicate sample sizes. Women identify odorants better than men at all ages. *(Reproduced with permission from RL Doty. Measurement of chemosensory function. WJOHNS 4:11-28, 2018.)*

TABLE 35-1 Disorders and Conditions Associated with Compromised Olfactory Function, as Measured by Olfactory Testing

Endocrine and Metabolic Conditions

- Adrenal cortical insufficiency (Addison's disease)
- Chromatin-negative gonadal dysgenesis (Turner's syndrome)
- Cushing's syndrome
- Diabetes
- Hypertension
- Hypothyroidism
- Idiopathic hypogonadotropic hypogonadism
- Kallmann's syndrome
- Liver disease
- Renal disease/kidney failure
- Pregnancy
- Pseudohypoparathyroidism
- Wilson's disease

Immune-Related Diseases

- Acute disseminated encephalomyelitis
- Allergic rhinitis
- Asthma
- Autoimmune pancreatitis
- Behçet's disease
- Churg-Strauss syndrome
- Cystic fibrosis
- Fibromyalgia
- Giant cell arteritis
- Hereditary angioedema
- Idiopathic inflammatory myopathies
- Inflammatory bowel diseases
- Lupus
- Mikulicz's disease
- Multiple sclerosis
- Myasthenia gravis
- Neuromyelitis optica
- Pemphigus vulgaris
- Psoriasis vulgaris
- Rheumatoid arthritis
- Sjögren's syndrome
- Systemic sclerosis (scleroderma)
- Wegener's granulomatosis

Nasosinus Disorders

- Adenoid hypertrophy
- Bacterial and viral upper respiratory infections
- Laryngopharyngeal reflux disease
- Rhinosinusitis/polyposis

Neurologic Diseases/Disorders

- Alzheimer's disease
- Amyotrophic lateral sclerosis (ALS)
- Bell's palsy
- Degenerative ataxias
- Down's syndrome
- Epilepsy
- Facial paralysis
- Fibromyalgia
- Frontotemporal lobe degeneration
- Guamanian ALS/Parkinson's disease/dementia syndrome
- Head trauma
- Huntington's disease
- Idiopathic inflammatory myopathies
- Korsakoff psychosis
- Lubag disease
- Migraine
- Multi-infarct dementia
- Narcolepsy with cataplexy
- Neoplasms, cranial/nasal
- Orthostatic tremor
- Parkinson's disease
- Pick's disease
- Rapid eye movement behavioral sleep disorder
- Stroke

Psychiatric-Related Diseases/Disorders

- Anorexia nervosa
- Asperger's syndrome
- Attention deficit/hyperactivity disorder
- Depression
- Obsessive compulsive disorder
- Panic disorder
- Posttraumatic stress disorder
- Psychopathy
- Schizophrenia
- Seasonal affective disorder
- 22q11 deletion syndrome

Viral, Bacterial, and Fungal Infections

- Candidiasis
- COVID-19
- Hepatitis C
- Herpetic meningoencephalitis
- Human immunodeficiency virus
- Legionnaires' disease
- Leprosy (Hansen's disease)
- Lyme disease
- Poliomyelitis
- Rhinosinusitis
- Upper respiratory infections

Other Disorders or Factors

- Alcoholism
- Bardet-Biedl syndrome
- Chemical exposure
- Congenital
- Iatrogenesis, including chemotherapy and radiation
- Nutritional deficiencies
- Obesity
- Tobacco smoking
- Toxic chemical exposures
- Vitamin B_{12} deficiency

Note: These disease/disorder classifications are not necessarily mutually exclusive.

diagnosis, with 5–10% experiencing total loss. Although in rhinosinusitis cases systemic glucocorticoid therapy can usually induce short-term functional improvement, it does not, on average, return smell test scores to normal, implying that chronic permanent neural loss is present and/or that short-term administration of systemic glucocorticoids does not completely mitigate the inflammation. It is well established that microinflammation in an otherwise seemingly normal epithelium can influence smell function.

A number of neurodegenerative diseases are accompanied by olfactory impairment, including PD, AD, Huntington's disease, parkinsonism-dementia complex of Guam, dementia with Lewy bodies (DLB), multiple system atrophy, corticobasal degeneration, frontotemporal dementia, and Down's syndrome; smell loss can also occur in idiopathic rapid eye movement (REM) behavioral sleep disorder (iRBD), as well as in multiple sclerosis (MS) related to lesions within olfaction-related structures. Olfactory impairment in PD often predates the clinical diagnosis by a number of years. In staged cases, studies of the sequence of formation of abnormal α-synuclein aggregates and Lewy bodies suggest that the olfactory bulbs may be, along with the dorsomotor nucleus of the vagus, the first site of neural damage in PD. In postmortem studies of patients with very mild "presymptomatic" signs of AD, poorer smell function has been associated with higher levels of AD-related pathology. Smell loss is more marked in patients with early clinical manifestations of DLB than in those with mild AD. Interestingly, smell loss is minimal or nonexistent in progressive supranuclear palsy and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)-induced parkinsonism. The relative contributions of disease-specific pathology or differential damage to forebrain neuromodulator/neurotransmitter systems in explaining different degrees of olfactory dysfunction among the various neurodegenerative diseases are presently unknown.

The smell loss seen in iRBD is of the same magnitude as that found in PD. This is of particular interest because patients with iRBD frequently develop PD and hyposmia. REM behavior disorder is not only seen in its idiopathic form but can also be associated with narcolepsy **(Chap. 33)**. A study of narcoleptic patients with and without REM behavior disorder demonstrated that narcolepsy, independent of REM behavior disorder, was associated with impairments in olfactory function. Loss of hypothalamic neurons expressing orexin (also known as hypocretin) neuropeptides is believed to be responsible for narcolepsy and cataplexy. Orexin-containing neurons project throughout the entire olfactory system (from the olfactory epithelium to the olfactory cortex), and damage to these projections may be one underlying mechanism for impaired olfactory performance in narcoleptic patients. Administration of intranasal orexin A (hypocretin-1) improved olfactory function, supporting the notion that mild olfactory impairment is not only a primary feature of narcolepsy with cataplexy, but that orexin deficiency may be directly responsible for the loss of smell in this condition.

DISORDERS OF TASTE

The majority of patients who present with taste dysfunction exhibit olfactory, not taste, loss. This is because most flavors attributed to taste actually depend on retronasal stimulation of the olfactory receptors during deglution. As noted earlier, taste buds only mediate basic tastes such as sweet, sour, bitter, salty, and umami. Significant impairment of whole-mouth gustatory function is rare outside of generalized metabolic disturbances or systemic use of some medications, because taste bud regeneration occurs and peripheral damage alone would require the involvement of multiple CN pathways. Taste function can be influenced by age, diet, smoking behavior, use of medications, and other subject-related factors including (1) the release of foul-tasting materials from the oral cavity from oral medical conditions (e.g., gingivitis, purulent sialadenitis) or appliances; (2) transport problems of tastants to the taste buds (e.g., drying, infections, or inflammatory conditions of the orolingual mucosa), (3) damage to the taste buds themselves (e.g., local trauma, invasive carcinomas), (4) damage to the neural pathways innervating the taste buds (e.g., middle ear infections), (5) damage to central structures (e.g., multiple sclerosis, tumor, epilepsy, stroke), and (6) systemic disturbances of metabolism (e.g., diabetes, thyroid disease, medications).

Unlike CN VII, CN IX is relatively protected along its path, although iatrogenic interventions such as tonsillectomy, bronchoscopy, laryngoscopy, endotracheal intubation, and radiation therapy can result in selective injury. CN VII damage commonly results from mastoidectomy, tympanoplasty, and stapedectomy, in some cases inducing persistent metallic sensations. Bell's palsy **(Chap. 452)** is one of the most common causes of CN VII injury that results in taste disturbance. On rare occasions, migraine **(Chap. 441)** is associated with a gustatory prodrome or aura, and in some cases, tastants can trigger a migraine attack. Interestingly, dysgeusia occurs in some cases of *burning mouth syndrome* (also termed *glossodynia* or *glossalgia*), as does dry mouth and thirst. Burning mouth syndrome is likely associated with dysfunction of the trigeminal nerve (CN V). Some of the etiologies suggested for this poorly understood syndrome are amenable to treatment, including (1) nutritional deficiencies (e.g., iron, folic acid, B vitamins, zinc), (2) diabetes mellitus (possibly predisposing to oral candidiasis), (3) denture allergy, (4) mechanical irritation from dentures or oral devices, (5) repetitive movements of the mouth (e.g., tongue thrusting, teeth grinding, jaw clenching), (6) tongue ischemia as a result of temporal arteritis, (7) periodontal disease, (8) reflux esophagitis, and (9) geographic tongue.

Although both taste and smell can be adversely influenced by drugs, taste alterations are more common. Indeed, >250 medications have been reported to alter the ability to taste. Major offenders include antineoplastic agents, antirheumatic drugs, antibiotics, and blood pressure medications. Terbinafine, a commonly used antifungal, has been linked to taste disturbance lasting up to 3 years. In a recent controlled trial, nearly two-thirds of individuals taking eszopiclone (Lunesta) for insomnia experienced a bitter dysgeusia that was stronger in women, systematically related to the time since drug administration, and positively correlated with both blood and saliva levels of the drug. Intranasal use of nasal gels and sprays containing zinc, which are common over-the-counter prophylactics for upper respiratory viral infections, has been implicated in loss of smell function. Whether their efficacy in preventing such infections, which are the most common cause of anosmia and hyposmia, outweighs their potential detriment to smell function requires study. Dysgeusia occurs commonly in the context of drugs used to treat or minimize symptoms of cancer, with a weighted prevalence from 56 to 76% depending on the type of cancer treatment. Attempts to prevent taste problems from such drugs using prophylactic zinc sulfate or amifostine have proven to be minimally beneficial. Although antiepileptic medications are occasionally used to treat smell or taste disturbances, the use of topiramate has been reported to result in a reversible loss of an ability to detect and recognize tastes and odors during treatment.

As with olfaction, a number of systemic disorders can affect taste. These include, but are not limited to, chronic renal failure, end-stage liver disease, vitamin and mineral deficiencies, diabetes mellitus, and hypothyroidism. In diabetes, there appears to be a progressive loss of taste beginning with glucose and then extending to other sweeteners, salty stimuli, and then all stimuli. Psychiatric conditions can be associated with chemosensory alterations (e.g., depression, schizophrenia, bulimia). A recent review of tactile, gustatory, and olfactory hallucinations demonstrated that no one type of hallucinatory experience is pathognomonic to any given diagnosis.

Pregnancy is a unique condition with regard to taste function. There appears to be an increase in dislike and intensity of bitter tastes during the first trimester that may help to ensure that pregnant women avoid poisons during a critical phase of fetal development. Similarly, a relative increase in the preference for salt and bitter in the second and third trimesters may support the ingestion of much-needed electrolytes to expand fluid volume and support a varied diet.

CLINICAL EVALUATION

In most cases, a careful clinical history will establish the probable etiology of a chemosensory problem, including questions about its nature, onset, duration, and pattern of fluctuations. *Sudden loss* suggests the possibility of head trauma, ischemia, infection, or a psychiatric condition. *Gradual loss* can reflect the development of a progressive obstructive lesion, although gradual loss can also follow head trauma. *Intermittent loss* suggests the likelihood of an inflammatory process. The patient should be asked about potential precipitating events, such as cold or flu infections, prior to symptom onset, because these often go underappreciated. Information regarding head trauma, smoking habits, drug and alcohol abuse (e.g., intranasal cocaine, chronic alcoholism), exposures to pesticides and other toxic agents, and medical interventions is also informative. A determination of all the medications that the patient was taking before and at the time of symptom onset is important because many can cause chemosensory disturbances. Comorbid medical conditions associated with smell impairment, such as renal failure, liver disease, hypothyroidism, diabetes, or dementia, should be assessed. Delayed puberty in association with anosmia (with or without midline craniofacial abnormalities, deafness, and renal anomalies) suggests the possibility of Kallmann's syndrome. Recollection of epistaxis, discharge (clear, purulent, or bloody), nasal obstruction, allergies, and somatic symptoms, including headache or irritation, may have localizing value. Questions related to memory, parkinsonian symptoms, and seizure activity (e.g., automatisms, blackouts, auras, déjà vu) should be posed. Pending litigation and the possibility of malingering should be considered. Modern forced-choice olfactory tests can detect malingering from improbable responses.

Neurologic and otorhinolaryngologic (ORL) examinations, along with appropriate brain and nasosinus imaging, aid in the evaluation of patients with olfactory or gustatory complaints. The neural evaluation should focus on CN function, with particular attention to possible skull base and intracranial lesions. Visual acuity, field, and optic disc examinations aid in detection of intracranial mass lesions that produce raised intracranial pressure (papilledema) and optic atrophy. Foster Kennedy syndrome refers to raised intracranial pressure plus a compressive optic neuropathy; typical causes are olfactory groove meningiomas or other frontal lobe tumors. The ORL examination should thoroughly assess the intranasal architecture and mucosal surfaces. Polyps, masses, and adhesions of the turbinates to the septum may compromise the flow of air to the olfactory receptors, because less than a fifth of the inspired air traverses the olfactory cleft in the unobstructed state. Blood tests may be helpful to identify such conditions as diabetes, infection, heavy metal exposure, nutritional deficiency (e.g., vitamin B_6 or B_{12}), allergy, and thyroid, liver, and kidney disease.

As with other sensory disorders, quantitative sensory testing is advised. Self-reports of patients can be misleading, and some patients who complain of chemosensory dysfunction have normal function for their age and gender. Quantitative smell and taste testing provides objective information for worker's compensation and other legal claims, as well as a way to accurately assess the effects of treatment interventions. A number of standardized olfactory and taste tests are commercially available. The most widely used olfactory test, the 40-item University of Pennsylvania Smell Identification Test (UPSIT), uses norms based on over 10,000 normal subjects. A determination is made of both absolute dysfunction (i.e., mild loss, moderate loss, severe loss, total loss, probable malingering) and relative dysfunction (percentile rank for age and gender). Although electrophysiologic testing is available at some smell and taste centers (e.g., odor event-related potentials), they require complex stimulus presentation and recording equipment and rarely provide additional diagnostic information. With the exception of electrogustometers, commercially available taste tests have only recently become available. Most use filter paper strips or similar materials impregnated with tastants, so no stimulus preparation is required.

TREATMENT AND MANAGEMENT

Given the various mechanisms by which olfactory and gustatory disturbance can occur, management of patients tends to be condition-specific. For example, patients with hypothyroidism, diabetes, or infections often benefit from specific treatments to correct the underlying disease process that is adversely influencing chemoreception. For most patients who present primarily with obstructive/transport loss affecting the nasal and paranasal regions (e.g., allergic rhinitis, polyposis, intranasal neoplasms, nasal deviations), medical and/or surgical intervention is often beneficial. Antifungal and antibiotic treatments may reverse taste problems secondary to candidiasis or other oral infections. Chlorhexidine mouthwash mitigates some salty or bitter dysgeusias, conceivably as a result of its strong positive charge. Excessive dryness of the oral mucosa is a problem with many medications and conditions, and artificial saliva (e.g., Xerolube) or oral pilocarpine treatments may prove beneficial. Other methods to improve salivary flow include the use of mints, lozenges, or sugarless gum. Flavor enhancers may make food more palatable (e.g., monosodium glutamate), but caution is advised to avoid overusing ingredients containing sodium or sugar, particularly in circumstances when a patient also has underlying hypertension or diabetes. Medications that induce distortions of taste can often be discontinued and replaced with other types of medications or modes of therapy. As mentioned earlier, pharmacologic agents result in taste disturbances much more frequently than smell disturbances. It is important to note, however, that many drug-related effects are long lasting and not reversed by short-term drug discontinuance.

A study of endoscopic sinus surgery in patients with chronic rhinosinusitis and hyposmia revealed that patients with severe olfactory dysfunction prior to the surgery had a more dramatic and sustained improvement over time compared to patients with more mild olfactory dysfunction prior to intervention. In the case of intranasal and sinus-related inflammatory conditions, such as seen with allergy, viruses, and traumas, the use of intranasal or systemic glucocorticoids may also be helpful. One common approach is to use a tapering course of oral prednisone. Topical intranasal administration of glucocorticoids was found to be less effective in general than systemic administration; however, the effects of different nasal administration techniques were not analyzed. For example, intranasal glucocorticoids are more effective if administered in the Moffett's position (head in the inverted position such as over the edge of the bed with the bridge of the nose perpendicular to the floor). After head trauma, an initial trial of glucocorticoids may help to reduce local edema and the potential deleterious deposition of scar tissue around olfactory fila at the level of the cribriform plate.

Treatments are limited for patients with chemosensory loss or primary injury to neural pathways. Nonetheless, spontaneous recovery can occur. In a follow-up study of 542 patients presenting to our center with smell loss from a variety of causes, modest improvement occurred over an average time period of 4 years in about half of the participants. However, only 11% of the anosmic and 23% of the hyposmic patients regained normal age-related function. Interestingly, the amount of dysfunction at the time of presentation, not etiology, was the best predictor of prognosis. Other predictors were age and the duration of dysfunction prior to initial testing.

Several studies have reported that patients with hyposmia may benefit from repeated smelling of odors over the course of weeks or months, although it remains to be determined how much improvement, if any, occurs over that known to occur spontaneously. The usual paradigm is to smell odors such as eucalyptol, citronella, eugenol, and phenyl ethyl alcohol before going to bed and immediately upon awakening each day. The rationale for such an approach comes from animal studies demonstrating that prolonged exposure to odorants can induce increased neural activity within the olfactory bulb. There is also limited evidence that α-lipoic acid (400 mg/d), an essential cofactor for many enzyme complexes with possible antioxidant effects, may be beneficial in mitigating smell loss following viral infection of the upper respiratory tract. However, double-blind studies are needed to confirm this observation. α-Lipoic acid has also been suggested to be useful in some cases of hypogeusia and burning mouth syndrome.

The use of zinc and vitamin A in treating olfactory disturbances is controversial, and there does not appear to be much benefit beyond replenishing established deficiencies. However, zinc has been shown to improve taste function secondary to hepatic deficiencies, and retinoids (bioactive vitamin A derivatives) are known to play an essential role in the survival of olfactory neurons. One protocol in which zinc was infused with chemotherapy treatments suggested a possible protective effect against developing taste impairment. Diseases of the alimentary tract can not only influence chemoreceptive function but also occasionally influence vitamin B_{12} absorption. This can result in a relative deficiency of vitamin B_{12}, theoretically contributing to olfactory nerve disturbance. Vitamin B_2 (riboflavin) and magnesium supplements are reported in the alternative literature to aid in the management of migraine that, in turn, may be associated with smell dysfunction. Because vitamin D deficiency is a cofactor of chemotherapy-induced mucocutaneous toxicity and dysgeusia, adding vitamin D_3, 1000–2000 units per day, may benefit some patients with smell and taste complaints during or following chemotherapy.

A number of medications have reportedly been used with success in ameliorating olfactory symptoms, although strong scientific evidence for efficacy is generally lacking. A report that theophylline improved smell function was uncontrolled and failed to account for the fact that some meaningful improvement occurs without treatment; indeed, the percentage of responders was about the same (~50%) as that noted by others to show spontaneous improvement over a similar time period. Antiepileptics and some antidepressants (e.g., amitriptyline) have been used to treat dysosmias and smell distortions, particularly following head trauma. Ironically, amitriptyline is also frequently on the list of medications that can ultimately distort smell and taste function, possibly from its anticholinergic effects. One study suggested that the centrally acting acetylcholinesterase inhibitor donepezil in AD resulted in improvements on smell identification measures that correlated with overall clinician-based impressions of change in dementia severity scores.

Alternative therapies, such as acupuncture, meditation, cognitive-behavioral therapy, and yoga, can help patients manage uncomfortable experiences associated with chemosensory disturbance and oral pain syndromes and to cope with the psychosocial stressors surrounding the impairment. Additionally, modification of diet and eating habits is also important. By accentuating the other sensory experiences of a meal, such as food texture, aroma, temperature, and color, one can optimize the overall eating experience for a patient. In some cases, a flavor enhancer like monosodium glutamate (MSG) can be added to foods to increase palatability and encourage intake.

Proper oral and nasal hygiene and routine dental care are extremely important ways for patients to protect themselves from disorders of the mouth and nose that can ultimately result in chemosensory disturbance. Patients should be warned not to overcompensate for their taste loss by adding excessive amounts of sugar or salt. Smoking cessation and the discontinuance of oral tobacco use are essential in the management of any patient with smell and/or taste disturbance and should be repeatedly emphasized.

A major and often overlooked element of therapy comes from chemosensory testing itself. Confirmation or lack of conformation of loss is beneficial to patients who come to believe, in light of unsupportive family members and medical providers, that they may be "crazy." In cases where the loss is minor, patients can be informed of the likelihood of a more positive prognosis. Importantly, quantitative testing places the patient's problem into overall perspective. Thus, it is often therapeutic for an older person to know that, while their smell function is not what it used to be, it still falls above the average of their peer group. Without testing, many such patients are simply told that they are getting old and nothing can be done for them, leading in some cases to depression and decreased self-esteem.

■ FURTHER READING

Devanand DP et al: Olfactory identification deficits are associated with increased mortality in a multiethnic urban community. Ann Neurol 78:401, 2015.

Doty RL: Olfactory dysfunction in COVID-19: Pathology and long-term implications for brain health. Trends Mol Med 28:781, 2022.

Doty RL: Olfactory dysfunction in neurodegenerative diseases: Is there a common pathological substrate? Lancet Neurol 16:478, 2017.

Doty RL et al: Systemic diseases and disorders. Handbook Clin Neurol 164:361, 2019.

Doty RL et al: Taste function in early stage treated and untreated Parkinson's disease. J Neurol 262:547, 2015.

Doty RL et al: Treatments for smell and taste disorders: A critical review. Handbook Clin Neurol 164:455, 2019.

Finlay JB et al: Persistent post-COVID-19 smell loss is associated with immune cell infiltration and altered gene expression in olfactory epithelium. Sci Transl Med 21:676, 2022.

Fornazieri MA et al: Adherence and efficacy of olfactory training as a treatment for persistent olfactory loss. Am J Rhinol Allergy 34:238, 2020.

Hawkes CH, Doty RL: *Smell and Taste Disorders*. Cambridge, Cambridge University Press, 2018.

Liu G et al: Prevalence and risk factors of taste and smell impairment in a nationwide sample of the US population: A cross-sectional study. BMJ Open 6:e013246, 2016.

London B et al: Predictors of prognosis in patients with olfactory disturbance. Ann Neurol 63:159, 2008.

Moein ST et al: Smell dysfunction: A biomarker for COVID-19. Int Forum Allergy Rhinol 10:944, 2020.

Perricone C et al: Smell and autoimmunity: A comprehensive review. Clin Rev Allergy Immunol 45:87, 2013.

36

Disorders of Hearing

Anil K. Lalwani

Hearing loss can present at any age and is one of the most common sensory disorders in humans. Nearly 10% of the adult population has some hearing loss, and one-third of individuals age >65 years have a hearing loss of sufficient magnitude to require a hearing aid.

PHYSIOLOGY OF HEARING

The function of the external and middle ear is to amplify sound to facilitate conversion of the mechanical energy of the sound wave into an electrical signal by the inner-ear hair cells, a process called mechanotransduction **(Fig. 36-1)**. Sound waves enter the external auditory canal and set the tympanic membrane (eardrum) in motion, which in turn moves the malleus, incus, and stapes of the middle ear. Movement of the footplate of the stapes causes pressure changes in the fluid-filled inner ear, eliciting a traveling wave in the basilar membrane of the cochlea. The tympanic membrane and the ossicular chain in the middle ear serve as an impedance-matching mechanism, improving the efficiency of energy transfer from air to the fluid-filled inner ear. In its absence, nearly 99.9% of the acoustical energy would be reflected and thus not heard. Instead, the eardrum and the ossicles boost the sound energy nearly 200-fold by the time it reaches the inner ear.

Within the cochlea of the inner ear, there are two types of hair cells that aid in hearing: inner and outer. The inner and outer hair cells of the organ of Corti have different innervation patterns, but both are mechanoreceptors; they detect the mechanical energy of the acoustic signal and aid its conversion to an electrical signal that travels by the auditory nerve. The afferent innervation relates principally to the inner hair cells while the efferent innervation relates principally to the outer hair cells. The outer hair cells outnumber the inner hair cells by nearly 6:1 (20,000 vs 3500). The motility of the outer hair cells alters the micromechanics of the inner hair cells, creating a cochlear amplifier, which explains the exquisite sensitivity and frequency selectivity of the cochlea.

Stereocilia of the hair cells of the organ of Corti, which rests on the basilar membrane, are in contact with the tectorial membrane and are deformed by the traveling wave. The deformation stretches tiny filamentous connections (tip links) between stereocilia, leading to opening of ion channels, influx of potassium, and hair cell depolarization and subsequent neurotransmission. A point of maximal displacement of the basilar membrane is determined by the frequency of the stimulating tone. High-frequency tones cause maximal displacement of the basilar membrane near the base of the cochlea, whereas for low-frequency sounds, the point of maximal displacement is toward the apex of the cochlea.

Beginning in the cochlea, the frequency specificity is maintained at each point of the central auditory pathway: dorsal and ventral cochlear nuclei, trapezoid body, superior olivary complex, lateral lemniscus, inferior colliculus, medial geniculate body, and auditory cortex. At low frequencies, individual auditory nerve fibers can respond more or less synchronously with the stimulating tone. At higher frequencies, phase-locking occurs so that neurons alternate in response to particular phases of the cycle of the sound wave. Intensity is encoded by the amount of neural activity in individual neurons, the number of neurons that are active, and the specific neurons that are activated.

There is evidence that the right and left ears as well as the central nervous system may process speech asymmetrically. Generally, a sound is processed symmetrically from the peripheral to the central auditory system. However, a "right ear advantage" exists for dichotic listening tasks, in which subjects are asked to report on competing sounds presented to each ear. In most individuals, a perceptual right ear advantage for consonant-vowel syllables, stop consonants, and words also exists. Similarly, whereas central auditory processing for sounds is symmetric with minimal lateral specialization for the most part, speech processing is lateralized. There is specialization of the left auditory cortex for speech recognition and production, and of the right hemisphere for emotional and tonal aspects of speech. Left hemisphere dominance for speech is found in 95–98% of right-handed persons and 70–80% of left-handed persons.

■ DISORDERS OF THE SENSE OF HEARING

Hearing loss can result from disorders of the auricle, external auditory canal, middle ear, inner ear, or central auditory pathways. In general, lesions in the auricle, external auditory canal, or middle ear that impede the transmission of sound from the external environment to the inner ear cause conductive hearing loss, whereas lesions that impair mechanotransduction in the inner ear or transmission of the electrical signal along the eighth nerve to the brain cause sensorineural hearing loss **(Table 36-1)**.

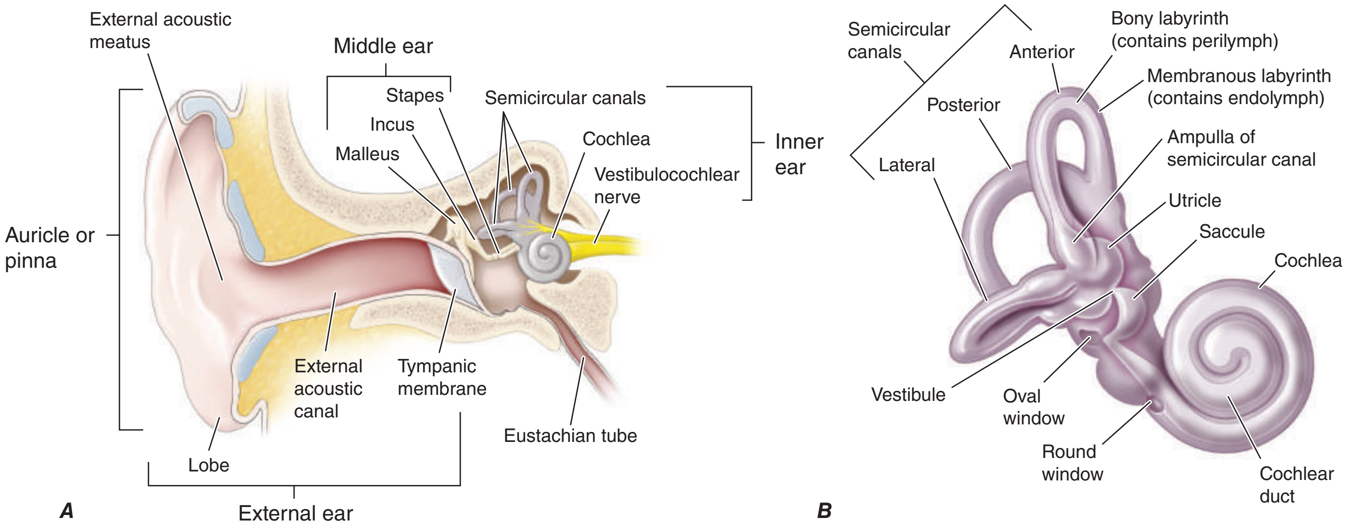

FIGURE 36-1 Ear anatomy. ***A.*** Drawing of modified coronal section through external ear and temporal bone, with structures of the middle and inner ear demonstrated. ***B.*** High-resolution view of inner ear.

Conductive Hearing Loss The external ear, the external auditory canal, and the middle-ear apparatus are designed to collect and amplify sound and efficiently transfer the mechanical energy of the sound wave to the fluid-filled cochlea. Factors that obstruct the transmission of sound or dampen the acoustic energy result in conductive hearing loss. Conductive hearing loss can occur from obstruction of the external auditory canal by cerumen, debris, and foreign bodies; swelling of the lining of the canal; atresia or neoplasms of the canal; perforations of the tympanic membrane; disruption of the ossicular chain, as occurs with necrosis of the long process of the incus in trauma or infection; otosclerosis; or fluid, scarring, or neoplasms in the middle ear. Rarely, inner-ear malformations or pathologies that create a "third window" in the inner ear such as superior semicircular canal dehiscence, lateral semicircular canal dysplasia, incomplete partition of the inner ear, and large vestibular aqueduct, are also associated with conductive hearing loss. This pathologic third window is associated with loss of mechanical energy associated with the sound wave leading to conductive hearing loss (see below).

Eustachian tube dysfunction is extremely common in adults and may predispose to acute otitis media (AOM) or serous otitis media (SOM). Balloon dilation has been shown to relieve acquired inflammatory obstruction of the Eustachian tube orifice and improve symptoms due to Eustachian tube dysfunction. Trauma, AOM, and chronic otitis media are the usual factors responsible for tympanic membrane perforation. While small perforations often heal spontaneously, larger defects usually require surgical intervention. Tympanoplasty is highly effective (>90%) in the repair of tympanic membrane perforations. Otoscopy is usually sufficient to diagnose AOM, SOM, chronic otitis media, cerumen impaction, tympanic membrane perforation, and Eustachian tube dysfunction; tympanometry and Eustachian tube function testing can be useful to confirm the clinical suspicion of these conditions.

Cholesteatoma, a benign tumor composed of stratified squamous epithelium in the middle ear or mastoid, occurs frequently in adults, often in the setting of severe Eustachian tube dysfunction. This is a slowly growing lesion that destroys bone and normal ear tissue. Theories of pathogenesis include traumatic immigration and invasion of squamous epithelium through a retraction pocket of the tympanic membrane, implantation of squamous epithelia in the middle ear through a perforation or surgery, and metaplasia following chronic infection and irritation. A chronically draining ear that fails to respond to appropriate antibiotic therapy should raise suspicion of a cholesteatoma. On examination, there is often a perforation of the tympanic membrane filled with cheesy white squamous debris. The presence of an aural polyp obscuring the tympanic membrane is highly suggestive of an underlying cholesteatoma. Conductive hearing loss secondary to ossicular erosion is common. Bony destruction visualized on computed tomography (CT) of the temporal bone is also highly suggestive of cholesteatoma. Surgery is required to remove this destructive process and reconstruct the ossicles.

Conductive hearing loss with a normal ear canal and intact tympanic membrane suggests either ossicular pathology or the presence of a "third window" in the inner ear (see below). Fixation of the stapes from *otosclerosis* is a common cause of low-frequency conductive hearing loss. It occurs equally in men and women and is inherited as an autosomal dominant trait with incomplete penetrance; in some cases, it may be a manifestation of osteogenesis imperfecta. Hearing impairment usually presents between the late teens and the forties. In women, the otosclerotic process is accelerated during pregnancy, and the hearing loss is often first noticeable at this time. A hearing aid or a simple outpatient surgical procedure (stapedectomy) can provide excellent auditory rehabilitation. Extension of otosclerosis beyond the stapes footplate to involve the cochlea (cochlear otosclerosis) can lead to mixed or sensorineural hearing loss. Fluoride therapy to prevent hearing loss from cochlear otosclerosis is of uncertain value.

Disorders that lead to the formation of a pathologic "third window" in the inner ear can be associated with conductive hearing loss. There are normally two major openings, or windows, that connect the inner ear with the middle ear and serve as conduits for transmission of sound; these are, respectively, the oval and round windows. A third window is formed where the normally hard otic bone surrounding the inner ear is eroded; dissipation of the acoustic energy at the third window is responsible for the "inner-ear conductive hearing loss." The superior semicircular canal dehiscence syndrome resulting from erosion of the otic bone over the superior circular canal can present with conductive hearing loss that mimics otosclerosis. A common symptom is vertigo evoked by loud sounds (Tullio phenomenon), by Valsalva maneuvers that change middle-ear pressure, or by applying positive pressure on the tragus (the cartilage anterior to the external opening of the ear canal). Patients with this syndrome also complain of fullness of the ear, pulsatile tinnitus, and being able to hear the movement of their eyes and neck. A large jugular bulb or jugular bulb diverticulum can create a "third window" by eroding into the vestibular aqueduct or posterior semicircular canal; the symptoms are similar to those of the superior semicircular canal dehiscence syndrome. Other inner-ear malformations such as lateral semicircular canal dysplasia, large vestibular aqueduct, or incomplete partition seen in stapes gusher syndrome can also be associated with inner-ear conductive hearing loss as a result of the third window. Low activation threshold on the vestibular-evoked myogenic potential test (VEMP test, see below) and inner-ear erosion on CT are diagnostic. Recalcitrant vertigo and dizziness may respond to surgical repair of the dehiscence.

TABLE 36-1 Common Causes of Conductive, Mixed, and Sensorineural Hearing Loss

Conductive Hearing Loss
Acute otitis media
Cerumen impaction
Cholesteatoma
Eustachian tube dysfunction
Inner ear dehiscence or "third window"
Middle ear tumors (facial nerve tumors, glomus tympanicum, glomus jugulare)
Ossicular discontinuity
Ossicular fixation
Otosclerosis
Serous otitis media
Temporal bone trauma
Tympanic membrane abnormalities
Tympanic membrane perforation
Mixed Hearing Loss
Acute otitis media
Cholesteatoma
Inner ear dehiscence or "third window"
Inner ear malformation
Middle ear tumors (facial nerve tumors, glomus tympanicum, glomus jugulare)
Otosclerosis
Stapes gusher syndrome
Temporal bone trauma
Sensorineural Hearing Loss
CNS infection (meningitis)
CNS tumors
Cerebellopontine angle tumors (vestibular schwannoma, meningioma)
Endolymphatic hydrops (Ménière's disease)
Endolymphatic sac tumor
Inner ear malformation
Perilymphatic fistula
Labyrinthitis
Multiple sclerosis
Noise-induced hearing loss
Presbycusis
Radiation therapy
Sudden hearing loss
Stroke
Temporal bone or head trauma

Sensorineural Hearing Loss Sensorineural hearing loss results from either damage to the mechanotransduction apparatus of the cochlea or disruption of the electrical conduction pathway from the inner ear to the brain. Thus, injury to hair cells, supporting cells, auditory neurons, or the central auditory pathway can cause sensorineural hearing loss. Damage to the hair cells of the organ of Corti may be caused by intense noise, viral infections, ototoxic drugs (e.g., salicylates, quinine and its synthetic analogues, aminoglycoside antibiotics, loop diuretics such as furosemide and ethacrynic acid, and cancer chemotherapeutic agents such as cisplatin), fractures of the temporal bone, meningitis, cochlear otosclerosis (see above), Ménière's disease, and aging. Congenital malformations of the inner ear may be the cause of hearing loss in some adults. Genetic predisposition alone or in concert with environmental exposures may also be responsible (see below).

Noise-Induced Hearing Loss Exposure to loud noise, either a short burst or over a more prolonged period of time, can lead to noise-induced hearing loss. Acute exposure to noise can lead to either temporary or permanent threshold shifts, depending on the intensity and duration of sound, due to hair cell injury and/or death. Typically, with permanent hearing loss, there is a "noise notch" with elevated hearing thresholds at 3000–4000 Hz. More recently, loud noise exposure has also been associated with cochlear synaptopathy or "hidden hearing loss"—hidden, because routine audiometry shows the pure tone hearing to be normal. Patients usually complain of not being able to hear clearly and are more bothered by the presence of background noise. In contrast to hair cell loss, hidden hearing loss is thought to be due to loss of auditory synapses on hair cells following noise exposure. In an increasingly noisy world, avoiding acoustic trauma with earplugs or earmuffs is highly recommended to prevent noise-induced or hidden hearing loss.

Presbycusis (age-associated hearing loss) is the most common cause of sensorineural hearing loss in adults. It is estimated to affect over half of adults aged >75 years in the United States, a population that is expected to double in size over the next 40 years. In the early stages, it is characterized by symmetric, gentle to sharply sloping, high-frequency hearing loss (**Fig. 36-2**). With progression, the hearing loss involves all frequencies. More importantly, the hearing impairment is associated with significant loss in clarity. There is a loss of discrimination for phonemes, recruitment (abnormal growth of loudness), and particular difficulty in understanding speech in noisy environments such as at restaurants and social events. Poor hearing is also associated with an increased incidence of cognitive impairment, rate of cognitive decline, and falls. In the elderly, left untreated, hearing loss leads to diminished quality of life, and has been shown to increase overall morbidity and mortality through falls and accidents. Hearing aids are helpful in enhancing the signal-to-noise ratio by amplifying sounds that are close to the listener. Hearing aid use has been shown to reduce cognitive decline and risk of falls. Although hearing aids are able to amplify sounds, they cannot restore the clarity of hearing. Thus, amplification with hearing aids may provide only limited rehabilitation once the word recognition score deteriorates below 50%. Cochlear implants are the treatment of choice when hearing aids prove inadequate, even when hearing loss is incomplete (see below).

Ménière's disease is characterized by episodic vertigo, fluctuating sensorineural hearing loss, tinnitus, and aural fullness. An absence of vertigo is inconsistent with the diagnosis of Ménière's disease, and the presence of fluctuating sensorineural hearing loss, tinnitus, and fullness without vertigo is more suggestive of cochlear hydrops. Tinnitus and/or deafness may be absent during the initial attacks of vertigo, but invariably appear as the disease progresses and increases in severity during acute attacks. The annual incidence of Ménière's disease is 0.5–7.5 per 1000; onset is most frequently in the fifth decade of life but may also occur in young adults or the elderly. Histologically, there

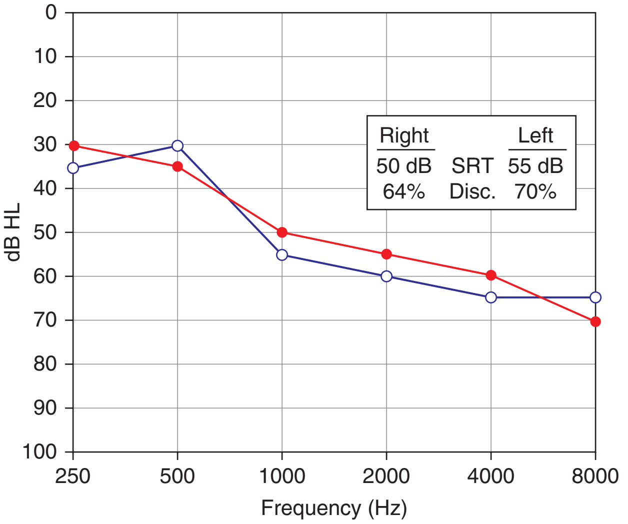

FIGURE 36-2 Presbycusis or age-related hearing loss. The audiogram shows a moderate to severe downsloping sensorineural hearing loss typical of presbycusis. The loss of high-frequency hearing is associated with a decreased speech discrimination score; consequently, patients complain of lack of clarity of hearing, especially in a noisy background. HL, hearing threshold level; SRT, speech reception threshold.

is distention of the endolymphatic system (endolymphatic hydrops) leading to degeneration of vestibular and cochlear hair cells. This may result from endolymphatic sac dysfunction secondary to infection, trauma, autoimmune disease, inflammatory causes, or tumor; an idiopathic etiology constitutes the largest category and is most accurately referred to as Ménière's disease. Endolymphatic sac tumors, often associated with von Hippel Lindau disease, may clinically mimic Ménière's disease. Although any pattern of hearing loss can be observed, typically, low-frequency, unilateral sensorineural hearing impairment is present. An abnormal VEMP test (see below) may be helpful in detecting Ménière's disease in a clinically unaffected contralateral ear. Magnetic resonance imaging (MRI) should be obtained to exclude retrocochlear pathology such as a cerebellopontine angle tumor, endolymphatic sac tumor, or demyelinating disorder. Therapy is directed toward the control of vertigo. A 2-g/d low-salt diet is the mainstay of treatment for control of rotatory vertigo. Diuretics, a short course of oral glucocorticoids, intratympanic glucocorticoids, or intratympanic gentamicin may also be useful adjuncts in recalcitrant cases. Surgical therapy of vertigo is reserved for unresponsive cases and includes endolymphatic sac decompression, labyrinthectomy, and vestibular nerve section. Both labyrinthectomy and vestibular nerve section abolish rotatory vertigo in >90% of cases. Unfortunately, there is no effective therapy for hearing loss, tinnitus, or aural fullness from Ménière's disease.

Sensorineural hearing loss may also result from any neoplastic, vascular, demyelinating, infectious, degenerative disease, or trauma affecting the central auditory pathways. Characteristically, in hearing loss due to central nervous system pathology, a reduction in clarity of hearing and speech comprehension is much greater than the loss of the ability to hear pure tone. Auditory testing is consistent with an auditory neuropathy; normal otoacoustic emissions (OAEs) and an abnormal auditory brainstem response (ABR) are typical (see below). Hearing loss can accompany hereditary sensorimotor neuropathies and inherited disorders of myelin. Tumors of the cerebellopontine angle such as vestibular schwannoma and meningioma **(Chap. 95)** usually present with asymmetric sensorineural hearing loss with greater deterioration of speech understanding than pure tone hearing. Multiple sclerosis **(Chap. 455)** may present with acute unilateral or bilateral hearing loss; typically, pure tone testing remains relatively stable while speech understanding fluctuates. Isolated labyrinthine infarction can present with acute hearing loss and vertigo due to a cerebrovascular accident involving the posterior circulation, usually the anterior inferior cerebellar artery; it may also be the heralding sign of impending catastrophic basilar artery infarction **(Chap. 437)**. HIV **(Chap. 208)**, which can produce both peripheral and central auditory system pathology, is another consideration in the evaluation of sensorineural hearing impairment.

A finding of conductive and sensorineural hearing loss in combination is termed *mixed hearing loss*. Mixed hearing losses can result from pathology of both the middle and inner ear, as can occur in otosclerosis involving the ossicles and the cochlea, head trauma, chronic otitis media, cholesteatoma, middle-ear tumors, and some inner-ear malformations.

Trauma resulting in temporal bone fractures may be associated with conductive, sensorineural, or mixed hearing loss. If the fracture spares the inner ear, there may simply be conductive hearing loss due to rupture of the tympanic membrane or disruption of the ossicular chain. These abnormalities can be surgically corrected. Profound hearing loss and severe vertigo are associated with temporal bone fractures involving the inner ear. A perilymphatic fistula associated with leakage of inner-ear fluid into the middle ear can occur and may require surgical repair. An associated facial nerve injury is not uncommon. Computed tomography (CT) is best suited to assess fracture of the traumatized temporal bone, evaluate the ear canal, and determine the integrity of the ossicular chain and involvement of the inner ear. Cerebrospinal fluid leaks that accompany temporal bone fractures are usually self-limited; the value of prophylactic antibiotics is uncertain.

Central Auditory Processing Disorders (CAPD) Following the reception of sound by the inner ear hair cells, the generated electrical signal from mechanotransduction is carried by the central auditory pathway—beginning with the auditory nerve to dorsal and ventral cochlear nuclei, trapezoid body, superior olivary complex, lateral lemniscus, inferior colliculus, medial geniculate body, and the auditory cortex—to refine, analyze, modify, organize, and interpret peripheral auditory input via central processing to facilitate speech comprehension, binaural processing, and interpret temporal cues. Loss of auditory nerve synapses due to loud noise exposure, tumors of the auditory/vestibular nerve, central nervous system (CNS) demyelinating disorders, diseases of the brainstem, cerebral stroke, and aging can all be associated with CAPDs. CAPDs can result in poor sound localization, lateralization, and speech comprehension, and deterioration of auditory performance in background noise. Poor understanding of speech in background noise is a common complaint in the elderly; this may be due to peripheral hearing loss as well as deterioration of central auditory processing. CAPD is best treated by enhancing the signal to noise ratio through amplifying sounds of interest while reducing/minimizing background noise; other intervention may include auditory training and strengthening compensatory strategies to minimize the impact of CAPD.

Tinnitus Tinnitus is defined as the perception of a sound when there is no sound in the environment. It can have a buzzing, roaring, or ringing quality and may be pulsatile (synchronous with the heartbeat). Tinnitus is often associated with either a conductive or sensorineural hearing loss. The pathophysiology of tinnitus is not well understood. The cause of the tinnitus can usually be determined by finding the cause of the associated hearing loss. Tinnitus may be the first symptom of a serious condition such as a vestibular schwannoma. Pulsatile tinnitus requires evaluation of the vascular system of the head to exclude vascular tumors such as glomus jugulare tumors, aneurysms, dural arteriovenous fistulas, and stenotic arterial lesions; it may also occur with SOM, superior semicircular dehiscence, and inner-ear dehiscence. It is most commonly associated with some abnormality of the jugular bulb such as a large jugular bulb or jugular bulb diverticulum. In absence of demonstrated pathology on magnetic resonance angiography (MRA)/magnetic resonance venography (MRV) or CT angiography, pulsatile tinnitus is usually attributed to turbulent venous blood flow through the transverse sinus, sigmoid sinus, and the jugular bulb.

GENETIC CAUSES OF HEARING LOSS

More than half of childhood hearing impairment is thought to be hereditary; hereditary hearing impairment (HHI) can also manifest later in life. HHI may be classified as either nonsyndromic, when hearing loss is the only clinical abnormality, or syndromic, when hearing loss is associated with anomalies in other organ systems. Nearly two-thirds of HHIs are nonsyndromic. Between 70 and 80% of nonsyndromic HHI is inherited in an autosomal recessive manner and designated DFNB; another 15–20% is autosomal dominant (DFNA). Less than 5% is X-linked (DFNX) or maternally inherited via the mitochondria.

More than 150 loci harboring genes for nonsyndromic HHI have been mapped, with recessive loci outnumbering dominant ones; numerous genes have now been identified **(eTable 36-1**: available at *accessmedicine.com/harrisons***)**. The hearing genes fall into the categories of structural proteins (*MYH9, MYO7A, MYO15, TECTA, DIAPH1*), transcription factors (*POU3F4, POU4F3*), ion channels (*KCNQ4, SLC26A4*), and gap junction proteins (*GJB2, GJB3, GJB6*). Several of these genes, including *GJB2*, *TECTA*, and *TMC1*, cause both autosomal dominant and recessive forms of nonsyndromic HHI. In general, the hearing loss associated with dominant genes has its onset in adolescence or adulthood, varies in severity, and progresses with age, whereas the hearing loss associated with recessive inheritance is congenital and profound. Connexin 26, a product of the *GJB2* gene, is particularly important because it is responsible for nearly 20% of all cases of childhood deafness; half of genetic deafness in children is *GJB2* related. Two frameshift mutations, 35delG and 167delT, account for >50% of the cases; however, screening for these two mutations alone is insufficient, and sequencing of the entire gene is required to fully capture *GJB2*-related recessive deafness. The 167delT mutation is highly prevalent in Ashkenazi Jews; ~1 in 1765 individuals in this population is homozygous and affected. *GJB2* hearing loss can also vary among the members of the same family, suggesting that other genes or factors influence the auditory phenotype. A single mutation in *GJB2* in combination with a single mutation in *GJB6* (connexin 30) can also lead to hearing loss and is an example of digenic inheritance of hearing loss.

In addition to *GJB2*, several other nonsyndromic genes are associated with hearing loss that progresses with age. The contribution of genetics to presbycusis is also becoming better understood and likely reflects a combination of genetic susceptibility impacted by environmental exposure to sound. Sensitivity to aminoglycoside ototoxicity can be maternally transmitted through a mitochondrial mutation. Susceptibility to noise-induced hearing loss may also be genetically determined.

There are >400 syndromic forms of hearing loss. These include Usher's syndrome (retinitis pigmentosa and hearing loss), Waardenburg's syndrome (pigmentary abnormality and hearing loss), Pendred's syndrome (thyroid organification defect and hearing loss), Alport's syndrome (renal disease and hearing loss), Jervell and Lange-Nielsen syndrome (prolonged QT interval and hearing loss), neurofibromatosis type 2 (bilateral vestibular schwannoma), and mitochondrial disorders (mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes [MELAS]; myoclonic epilepsy and ragged red fibers [MERRF]; and progressive external ophthalmoplegia [PEO]) (**eTable 36-2**: available at *accessmedicine.com/harrisons*).

APPROACH TO THE PATIENT

Disorders of the Sense of Hearing

The goal in the evaluation of a patient with auditory complaints is to determine (1) the nature of the hearing impairment (conductive vs sensorineural vs mixed), (2) the severity of the impairment (mild, moderate, severe, or profound), (3) the anatomy of the impairment (external ear, middle ear, inner ear, or central auditory pathway), and (4) the etiology. The presence of signs and symptoms associated with hearing loss should be ascertained **(Table 36-2)**. The history should elicit characteristics of the hearing loss, including the duration of deafness, unilateral versus bilateral involvement, nature of onset (sudden vs insidious), and rate of progression (rapid vs slow). Symptoms of tinnitus, vertigo, imbalance, aural fullness, otorrhea, headache, facial nerve dysfunction, and head and neck paresthesias should be noted. Information regarding head trauma, exposure to ototoxins, occupational or recreational noise exposure, and family history of hearing impairment may also be important. A sudden onset of unilateral hearing loss, with or without tinnitus, may represent a viral infection of the inner ear, vestibular schwannoma, or a stroke. Patients with unilateral hearing loss (sensory or conductive) usually complain of reduced hearing, poor sound localization, and difficulty hearing clearly in the presence of background noise. Gradual progression of a hearing deficit is common with otosclerosis, noise-induced hearing loss, vestibular schwannoma, or Ménière's disease. Small vestibular schwannomas typically present with asymmetric hearing impairment, tinnitus, and imbalance (rarely vertigo); cranial neuropathy, in particular of the trigeminal or facial nerve, may accompany larger tumors.

In addition to hearing loss, Ménière's disease may be associated with episodic vertigo, tinnitus, and aural fullness. Sound-induced vertigo, autophony, and being able to hear one's own neck or eye movement are highly suggestive of superior semicircular canal dehiscence. Hearing loss with otorrhea is most likely due to chronic otitis media or cholesteatoma.

Examination should include the auricle, external ear canal, and tympanic membrane. In the elderly, the external ear canal is often dry and fragile; it is preferable to clean cerumen with wall-mounted suction or cerumen loops and to avoid irrigation. Irrigation should also be avoided when a tympanic membrane perforation is present or the integrity of the eardrum cannot be established. In examining the eardrum, the topography of the tympanic membrane is more important than the presence or absence of the light reflex. In addition to the pars tensa (the lower two-thirds of the tympanic membrane), the pars flaccida (upper one-third of the tympanic membrane) above the short process of the malleus should also be examined for retraction pockets that may be evidence of chronic Eustachian tube dysfunction or cholesteatoma. Insufflation of the ear canal is necessary to assess tympanic membrane mobility and compliance. Careful inspection of the nose, nasopharynx, and upper respiratory tract is important. Unilateral serous effusion or unexplained otalgia should prompt a fiberoptic examination of the nasopharynx and larynx to exclude neoplasms. Cranial nerves should be evaluated with special attention to facial and trigeminal nerves, which are commonly affected with tumors involving the cerebellopontine angle.

The Rinne and Weber tuning fork tests, with a 512-Hz tuning fork, are used to screen for hearing loss, differentiate conductive from sensorineural hearing losses, and confirm the findings of audiologic evaluation. The Rinne test compares the ability to hear by air conduction with the ability to hear by bone conduction. The tines of a vibrating tuning fork are held near the opening of the external auditory canal, and then the stem is placed on the mastoid process; for direct contact, it may be placed on teeth or dentures. The patient is asked to indicate whether the tone is louder by air conduction or bone conduction. Normally, and in the presence of sensorineural hearing loss, a tone is heard louder by air conduction than by bone conduction; however, with conductive hearing loss of ≥30 dB (see "Audiologic Assessment," below), the bone-conduction stimulus is perceived as louder than the air-conduction stimulus. For the Weber test, the stem of a vibrating tuning fork is placed on the head in the midline and the patient is asked whether the tone is heard in both ears or better in one ear than in the other. With a unilateral conductive hearing loss, the tone is perceived in the affected ear. With a unilateral sensorineural hearing loss, the tone is perceived in the unaffected ear. A 5-dB difference in hearing between the two ears is required for lateralization.

TABLE 36-2 Signs and Symptoms Suggestive of Hearing Loss

Saying "huh" a great deal
Reduced clarity of hearing
Difficulty understanding conversations in background noise
Family complaining of hearing loss
Tinnitus
Turning the volume up on radio or television
Sensitivity to noises
Fullness in the ear
Avoiding social settings

LABORATORY ASSESSMENT OF HEARING

Audiologic Assessment The minimum audiologic assessment for hearing loss should include the measurement of pure tone air-conduction and bone-conduction thresholds, speech reception threshold, word recognition score, tympanometry, acoustic reflexes, and acoustic-reflex decay. This test battery provides a screening evaluation of the entire auditory system and allows one to determine whether further differentiation of a sensory (cochlear) from a neural (retrocochlear) hearing loss is indicated.

Pure tone audiometry assesses hearing acuity for pure tones. The test is administered by an audiologist and is performed in a sound-attenuated chamber. The pure tone stimulus is delivered with an audiometer, an electronic device that allows the presentation of specific frequencies (generally between 250 and 8000 Hz) at specific intensities. Air- and bone-conduction thresholds are established for each ear. Air-conduction thresholds are determined by presenting the stimulus in air with the use of headphones. Bone-conduction thresholds are determined by placing the stem of a vibrating tuning fork or an oscillator of an audiometer in contact with the head. In the presence of a hearing loss, broad-spectrum noise is presented to the nontest ear for masking purposes so that responses are based on perception from the ear under test.

The responses are measured in decibels (dBs). An *audiogram* is a plot of intensity in dBs of hearing threshold versus frequency. A dB is equal to 20 times the logarithm of the ratio of the sound pressure required to achieve threshold in the patient to the sound pressure required to achieve threshold in a normal-hearing person. Therefore, a change of 6 dB represents doubling of sound pressure, and a change of 20 dB represents a tenfold change in sound pressure. Loudness, which depends on the frequency, intensity, and duration of a sound, doubles with approximately each 10-dB increase in sound pressure level. Pitch, on the other hand, does not directly correlate with frequency. The perception of pitch changes slowly in the low and high frequencies. In the middle tones, which are important for human speech, pitch varies more rapidly with changes in frequency.

Pure tone audiometry establishes the presence and severity of hearing impairment, unilateral versus bilateral involvement, and the type of hearing loss. Conductive hearing losses with a large mass component, as is often seen in middle-ear effusions, produce elevation of thresholds that predominate in the higher frequencies. Conductive hearing losses with a large stiffness component, as in fixation of the footplate of the stapes in early otosclerosis, produce threshold elevations in the lower frequencies. Often, the conductive hearing loss involves all frequencies, suggesting involvement of both stiffness and mass. In general, sensorineural hearing losses such as presbycusis affect higher frequencies more than lower frequencies **(Fig. 36-3)**. An exception is Ménière's disease, which is characteristically associated with low-frequency sensorineural hearing loss (though any frequency can be affected). Noise-induced hearing loss has an unusual pattern of hearing impairment in which the loss at 3000–4000 Hz is greater than at higher frequencies. Vestibular schwannomas characteristically affect the higher frequencies, but any pattern of hearing loss can be observed.

Speech recognition requires greater synchronous neural firing than is necessary for appreciation of pure tones. *Speech audiometry* tests the clarity with which one hears. The *speech reception threshold* (*SRT*) is defined as the intensity at which speech is recognized as a meaningful symbol and is obtained by presenting two-syllable words with an equal accent on each syllable. The intensity at which the patient can repeat 50% of the words correctly is the SRT. Once the SRT is determined, discrimination or word recognition ability is tested by presenting one-syllable words at 25–40 dB above the SRT. The words are phonetically balanced in that the phonemes (speech sounds) occur in the list of words at the same frequency that they occur in ordinary conversational English. An individual with normal hearing or conductive hearing loss can repeat 88–100% of the phonetically balanced words correctly. Patients with a sensorineural hearing loss have variable loss of discrimination. As a general rule, neural lesions produce greater deficits in discrimination than do cochlear lesions. For example, in a patient with mild asymmetric sensorineural hearing loss, a clue to the diagnosis of vestibular schwannoma is the presence of greater than expected deterioration in discrimination ability. Deterioration in discrimination ability when words are presented at higher intensities above the SRT ("rollover phenonmena") also suggests a lesion in the eighth nerve or central auditory pathways.

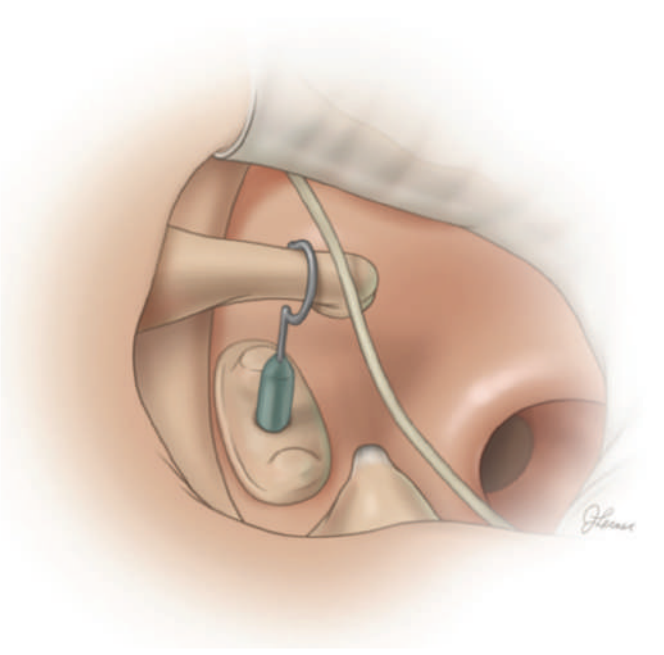

FIGURE 36-3 Stapedectomy. The fixation of stapes in otosclerosis can be addressed by placing a piston prosthesis from the incus to a small fenestration in the stapes footplate. This operation is successful in addressing the conductive hearing loss associated with otosclerosis. *(Reproduced with permission from Julia Lerner.)*

Tympanometry measures the impedance of the middle ear to sound and is useful in diagnosis of middle-ear effusions. A *tympanogram* is the graphic representation of change in impedance or compliance as the pressure in the ear canal is changed. Normally, the middle ear is most compliant at atmospheric pressure, and the compliance decreases as the pressure is increased or decreased (type A); this pattern is seen with normal hearing or in the presence of sensorineural hearing loss. Compliance that does not change with change in pressure suggests middle-ear effusion (type B). With a negative pressure in the middle ear, as with Eustachian tube obstruction, the point of maximal compliance occurs with negative pressure in the ear canal (type C). A tympanogram in which no point of maximal compliance can be obtained is most commonly seen with discontinuity of the ossicular chain (type A_d). A reduction in the maximal compliance peak can be seen in otosclerosis (type A_s).

During tympanometry, an intense tone elicits contraction of the stapedius muscle, also known as the *acoustic reflex*. The change in compliance of the middle ear with contraction of the stapedius muscle can be detected. The presence or absence of the acoustic reflex is important in determining the etiology of hearing loss as well as in the anatomic localization of facial nerve paralysis. The acoustic reflex can help differentiate between conductive hearing loss due to otosclerosis and that caused by an inner-ear "third window": it is absent in otosclerosis and present in inner-ear conductive hearing loss. Normal or elevated acoustic reflex thresholds in an individual with sensorineural hearing impairment suggest a cochlear hearing loss. An absent acoustic reflex in the setting of sensorineural hearing loss is not helpful in localizing the site of lesion. Assessment of *acoustic reflex decay* helps differentiate sensory from neural hearing losses. In neural hearing loss, such as with vestibular schwannoma, the reflex adapts or decays with time.

OAEs generated by outer hair cells only can be measured with microphones inserted into the external auditory canal. The emissions may be spontaneous or evoked with sound stimulation. The presence of OAEs indicates that the outer hair cells of the organ of Corti are intact and can be used to assess auditory thresholds and to distinguish sensory from neural hearing losses.

Evoked Responses *Electrocochleography* measures the earliest evoked potentials generated in the cochlea and the auditory nerve. Receptor potentials recorded include the cochlear microphonic, generated by the outer hair cells of the organ of Corti, and the summating potential, generated by the inner hair cells in response to sound. The whole nerve action potential representing the composite firing of the first-order neurons can also be recorded during electrocochleography. Clinically, the test is useful in the diagnosis of Ménière's disease, in which an elevation of the ratio of summating potential to action potential is seen.

Brainstem auditory-evoked responses (*BAERs*), also known as ABRs, are useful in differentiating the site of sensorineural hearing loss. In response to sound, five distinct electrical potentials arising from different stations along the peripheral and central auditory pathway (eighth nerve, cochlear nucleus, superior olivary complex, lateral lemniscus, and inferior colliculus) can be identified using computer averaging from scalp surface electrodes. BAERs are valuable in situations in which patients cannot or will not give reliable voluntary thresholds.

They are also used to assess the integrity of the auditory nerve and brainstem in various clinical situations, including diagnosis of retrocochlear pathology such as vestibular schwannoma, intraoperative monitoring of hearing, and determination of brain death.

The *VEMP test* investigates otolith and vestibular nerve function by presenting a high-level acoustic stimulus and evoking a short-latency electromyographic potential; cVEMP (or cervical VEMP) and oVEMP (or ocular VEMP) have been described. The cVEMP elicits a vestibulocollic reflex whose afferent limb arises from acoustically sensitive cells in the saccule, with signals conducted via the inferior vestibular nerve. cVEMP is a biphasic, short-latency response recorded from the tonically contracted sternocleidomastoid muscle in response to loud auditory clicks or tones. cVEMPs may be diminished or absent in patients with early and late Ménière's disease, vestibular neuritis, benign paroxysmal positional vertigo, and vestibular schwannoma. On the other hand, the threshold for VEMPs may be lower in cases of superior canal dehiscence, other inner-ear dehiscence ("third window"), and perilymphatic fistula. The oVEMP, in contrast, is a response involving the utricle primarily and superior vestibular nerve. The oVEMP excitatory response is recorded from the extraocular muscle. The oVEMP is abnormal in superior vestibular neuritis.

Imaging Studies The choice of radiologic tests is largely determined by whether the goal is to evaluate the bony anatomy of the external, middle, and inner ear or to image the auditory nerve and brain. Axial and coronal CT of the temporal bone with fine 0.3-mm cuts is ideal for determining the caliber of the external auditory canal, integrity of the ossicular chain, and presence of middle-ear or mastoid disease; it can also detect inner-ear malformations. CT is also ideal for the detection of bone erosion with chronic otitis media and cholesteatoma. Pöschl reformatting in the plane of the superior semicircular canal is required for the identification of dehiscence or absence of bone over the superior semicircular canal. MRI is superior to CT for imaging of retrocochlear pathology such as vestibular schwannoma, meningioma, other lesions of the cerebellopontine angle, demyelinating lesions of the brainstem, and brain tumors. Both CT and MRI are equally capable of identifying inner-ear malformations and assessing cochlear patency for preoperative evaluation of patients for cochlear implantation.

TREATMENT

Disorders of the Sense of Hearing

In general, conductive hearing losses are amenable to surgical correction, whereas sensorineural hearing losses are usually managed medically. Atresia of the ear canal can be surgically repaired, often with significant improvement in hearing. Alternatively, the conductive hearing loss associated with atresia can be addressed with a bone-anchored hearing aid (BAHA). Tympanic membrane perforations due to chronic otitis media or trauma can be repaired with an outpatient tympanoplasty. Likewise, conductive hearing loss associated with otosclerosis can be treated by stapedectomy, which is successful in >95% of cases (Fig. 36-3). Tympanostomy tubes allow the prompt return of normal hearing in individuals with middle-ear effusions. Hearing aids are effective and well tolerated in patients with conductive hearing losses.

Patients with mild, moderate, and severe sensorineural hearing losses are regularly rehabilitated with hearing aids of varying configuration and strength. Hearing aids have been improved to provide greater fidelity and have been miniaturized. The current generation of hearing aids is nearly invisible, thus reducing stigma associated with their use. In general, the more severe the hearing impairment, the larger the hearing aid required for auditory rehabilitation. Digital hearing aids lend themselves to individual programming, and multiple and directional microphones at the ear level may be helpful in noisy surroundings. Because all hearing aids amplify noise as well as speech, the only absolute solution to the problem of noise is to place the microphone closer to the speaker than the noise source. This arrangement is not possible with a self-contained, cosmetically acceptable device. A significant limitation of rehabilitation with a hearing aid is that although it is able to enhance detection of sound with amplification, it cannot restore clarity of hearing that is lost with presbycusis.

The cost of a single hearing aid is a significant obstacle for many hearing-impaired individuals, and usually bilateral amplification is recommended. To reduce cost and spur innovation, a new category of over-the-counter amplification devices that can be purchased similar to reading eyeglasses by simply walking into a store has been approved by the U.S. Food and Drug Administration. By reducing the cost of amplification devices to consumers, promoting innovation, and increasing competition, this new class of devices is making hearing rehabilitation more accessible.

Patients with unilateral deafness have difficulty with sound localization and reduced clarity of hearing in background noise. They may benefit from a contralateral routing of signal (CROS) hearing aid in which a microphone is placed on the hearing-impaired side, and the sound is transmitted to the receiver placed on the contralateral ear. The same result may be obtained with a BAHA, in which a hearing aid clamps to a screw integrated into the skull on the hearing-impaired side. Like the CROS hearing aid, the BAHA transfers the acoustic signal to the contralateral hearing ear, but it does so by vibrating the skull. Patients with profound deafness on one side and some hearing loss in the better ear are candidates for a BICROS hearing aid; it differs from the CROS hearing aid in that the patient wears a hearing aid, and not simply a receiver, in the better ear. Unfortunately, while CROS and BAHA devices provide benefit, they do not restore hearing in the deaf ear. Only cochlear implants can restore hearing (see below). Increasingly, cochlear implants are being used for the treatment of patients with single-sided deafness; in this setting, the cochlear implant restores hearing, reduces tinnitus, improves sound localization and performance in background noise, and reduces fatigue.

In many situations, including lectures and the theater, hearing-impaired persons benefit from assistive devices that are based on the principle of having the speaker closer to the microphone than any source of noise. Assistive devices include infrared and frequency-modulated (FM) transmission as well as an electromagnetic loop around the room for transmission to the individual's hearing aid. Hearing aids with telecoils can also be used with properly equipped telephones in the same way. Bluetooth technology has revolutionized connectivity between hearing aids and other devices such as smart phones.

In the event that the hearing aid provides inadequate rehabilitation, cochlear implants may be appropriate **(Fig. 36-4)**. Criteria for implantation include severe to profound hearing loss with open-set sentence cognition of ≤40% under best-aided conditions. Worldwide, >600,000 hearing-impaired individuals have received cochlear implants. Cochlear implants are neural prostheses that convert sound energy to electrical energy and can be used to stimulate the auditory division of the eighth nerve directly. In most cases of profound hearing impairment, the auditory hair cells are lost but the ganglionic cells of the auditory division of the eighth nerve are preserved. Cochlear implants consist of electrodes that are inserted into the cochlea through the round window, speech processors that extract acoustic elements of speech for conversion to electrical currents, and a means of transmitting the electrical energy through the skin. Patients with implants experience sound that helps with speech reading, allows open-set word recognition, and helps in modulating the person's own voice. Usually, within the first 3–6 months after implantation, adult patients can understand speech without visual cues. With the current generation of multichannel cochlear implants, nearly 75% of patients are able to converse on the telephone. Bilateral cochlear implantations are commonly performed, especially in children; these patients perform better in background noise, have better sound localization, and are less fatigued by the "work" compared to monaural hearing.

Hybrid cochlear implants are indicated for the treatment of high-frequency hearing loss in patients who do not have profound

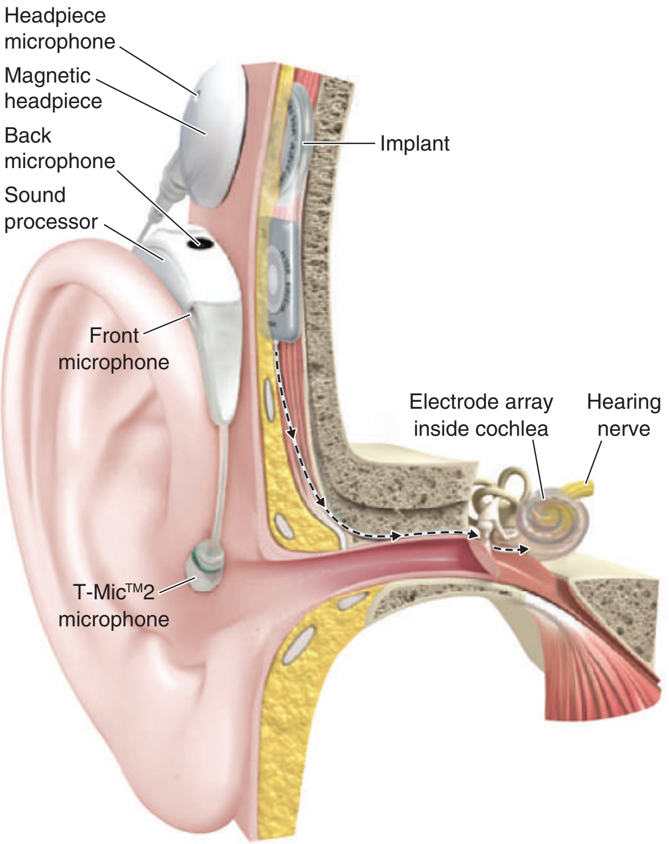

FIGURE 36-4 A cochlear implant is composed of an external microphone and speech processor worn on the ear and a receiver implanted underneath the temporalis muscle. The internal receiver is attached to an electrode that is placed surgically in the cochlea.

hearing loss and yet do not benefit from hearing aids. Patients with presbycusis typically have normal low-frequency hearing while suffering from high-frequency hearing loss associated with loss of clarity that cannot always be adequately rehabilitated with a hearing aid. However, these patients are not candidates for conventional cochlear implants because they have too much residual hearing. The hybrid implant has been specifically designed for this patient population; it has a shorter electrode than a conventional cochlear implant and can be introduced into the cochlea atraumatically, thus preserving low-frequency hearing. Individuals with a hybrid implant use their own natural low-frequency “acoustic” hearing and rely on the implant for providing “electrical” high-frequency hearing. Patients who have received the hybrid implant perform better on speech discrimination tests in both quiet and noisy backgrounds.

For individuals who were born without cochlea or have had both eighth nerves destroyed by trauma or bilateral vestibular schwannomas (e.g., neurofibromatosis type 2), brainstem auditory implants placed near the cochlear nucleus may provide auditory rehabilitation. Currently, brainstem implants provide sound awareness, but unfortunately, speech understanding remains elusive.

Understanding of the genetic basis of hearing loss, the molecular signals for hair cell regeneration, and the pathways involved in environmental and pharmaceutical ototoxicity has opened opportunities for molecular therapies. Molecular therapies, including gene therapy, stem cell therapy, RNA-based therapies, and CRISPR/Cas9 gene editing, hold promise for prevention and treatment of genetic and nongenetic causes of hearing loss. In one exciting recent development, adeno-associated virus–mediated delivery of otoferlin protein restored hearing in congenitally deaf children. Thus, hearing loss may be treated with gene replacement, gene suppression, and genome editing technology, to protect, preserve, or regenerate delicate inner ear structures.

Tinnitus often accompanies hearing loss. Similar to background noise, tinnitus can degrade speech comprehension in individuals with hearing impairment. Patients with tinnitus should be advised to minimize caffeine ingestion, avoid high dosages of nonsteroidal anti-inflammatory drugs (NSAIDs), and reduce stress. Therapy for tinnitus is usually directed toward minimizing the appreciation of tinnitus. Relief of the tinnitus may be obtained by masking it with background music or white noise. Hearing aids are also helpful in tinnitus suppression, as are tinnitus maskers, devices that present a sound to the affected ear that is more pleasant to listen to than the tinnitus. The use of a tinnitus masker is often followed by several hours of inhibition of the tinnitus. Antidepressants have also been shown to be beneficial in helping patients cope with tinnitus.

Hard-of-hearing individuals often benefit from a reduction in unnecessary noise in the environment (e.g., radio or television) to enhance the signal-to-noise ratio. Speech comprehension is aided by lip reading; therefore, the impaired listener should be seated so that the face of the speaker is well illuminated and easily seen. Although speech should be in a loud, clear voice, one should be aware that in sensorineural hearing losses in general and in hard-of-hearing elderly in particular, recruitment (abnormal perception of loud sounds) may be troublesome. Above all, optimal communication cannot take place without both parties giving it their full and undivided attention.

PREVENTION

Conductive hearing losses may be prevented by prompt antibiotic therapy of adequate duration for AOM and by ventilation of the middle ear with tympanostomy tubes in middle-ear effusions lasting ≥12 weeks. Loss of vestibular function and deafness due to aminoglycoside antibiotics can largely be prevented by careful monitoring of serum peak and trough levels.

Some 10 million Americans have noise-induced hearing loss, and 20 million are exposed to hazardous noise in their employment. Noise-induced hearing loss can be prevented by avoidance of exposure to loud noise or by regular use of earplugs or fluid-filled ear muffs to attenuate intense sound. Table 36-3 lists loudness levels for a variety of environmental sounds. High-risk activities for noise-induced hearing loss include use of electrical equipment for wood- and metalworking, and target practice or hunting with small firearms. All internal-combustion and electric engines, including snow and leaf blowers, snowmobiles, outboard motors, and chainsaws, require protection of the user with hearing protectors. Virtually all noise-induced hearing loss is preventable through education, which should begin before the teenage years. Programs for conservation of hearing in the workplace are required by the Occupational Safety and Health Administration (OSHA) whenever the exposure over an 8-h period averages 85 dB. OSHA mandates that

TABLE 36-3 Decibel (Loudness) Level of Common Environmental Noise

SOURCE	DECIBEL (dB)
Weakest sound heard	0
Whisper	30
Normal conversation	55–65
City traffic inside car	85
OSHA Monitoring Requirement Begins	**90**
Jackhammer	95
Subway train at 200 ft	95
Power mower	107
Power saw	110
Painful Sound	**125**
Jet engine at 100 ft	140
12-gauge shotgun blast	165
Loudest sound that can occur	194

Abbreviation: OSHA, Occupational Safety and Health Administration.

TABLE 36-4 OSHA Daily Permissible Noise Level Exposure

SOUND LEVEL (dB)	DURATION PER DAY (h)
90	8
92	6
95	4
97	3
100	2
102	1.5
105	1
110	0.5
115	≤0.25

Note: Exposure to impulsive or impact noise should not exceed 140-dB peak sound pressure level.

Source: From *https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=standards&p_id=9735*.

workers in such noisy environments have hearing monitoring and protection programs that include a preemployment screen, an annual audiologic assessment, and the mandatory use of hearing protectors. Exposure to loud sounds above 85 dB in the work environment is restricted by OSHA, with halving of allowed exposure time for each increment of 5 dB above this threshold; for example, exposure to 90 dB is permitted for 8 h; 95 dB for 4 h, and 100 dB for 2 h **(Table 36-4)**.

FURTHER READING

Carlson ML: Cochlear implantation in adults. N Engl J Med 382:1531, 2020.

Jarach C et al: Global prevalence and incidence of tinnitus: A systematic review and meta-analysis. JAMA Neurol 79:888, 2022.

Lee HA, Chung JH: Contemporary review of idiopathic sudden sensorineural hearing loss: Management and prognosis. J Audiol Otol 28:10, 2024.

Lin CY et al: Salvage therapy for refractory sudden sensorineural hearing loss (RSSNHL): A systematic review and network meta-analysis. Int J Audiol 22:1, 2024.

Lopez-Escamez JA, Liu Y: Epidemiology and genetics of Meniere's disease. Curr Opin Neurol 37:88, 2024.

Lv J et al: AAV1-hOTOF gene therapy for autosomal recessive deafness 9: A single-arm trial. Lancet 24:S0140, 2024.

Ruan Q et al: Which came first, age-related hearing loss with tinnitus or cognitive impairment? What are the potential pathways? J Integr Neurosci 22:109, 2023.

37

Upper Respiratory Symptoms, Including Earache, Sinus Symptoms, and Sore Throat

Rachel L. Amdur, Jeffrey A. Linder

Upper respiratory symptoms are most commonly caused by viral infection but can also be caused by other infectious, inflammatory, allergic, autoimmune, and neoplastic conditions. This chapter will discuss ambulatory antibiotic prescribing and review the most common causes of upper respiratory symptoms, including nonspecific upper respiratory infections.

Ear pain is most commonly caused by otitis externa, acute otitis media (AOM), otitis media with effusion (OME), and acute mastoiditis. Sinus symptoms can be caused by acute sinusitis, invasive fungal sinusitis, nosocomial sinusitis, and chronic sinusitis. Sore throat and neck pain can be caused by streptococcal pharyngitis, nonstreptococcal pharyngitis, acute infectious mononucleosis, other types of bacterial pharyngitis, Lemierre syndrome, gonococcal pharyngitis, diphtheria, acute HIV infection, head and neck abscesses, epiglottitis, and laryngitis. At the time of presentation, upper respiratory symptoms of most common viral and bacterial etiologies have generally lasted from hours up to a few days.

UPPER RESPIRATORY INFECTIONS

Upper respiratory infections (URIs) are acute respiratory infections that occur above the vocal cords. URIs, including nonspecific upper respiratory tract infection, otitis media, sinusitis, and pharyngitis, are collectively the most common symptomatic reason for seeking care in the United States. In terms of etiology, symptoms, and signs, URIs overlap with lower acute respiratory infections that occur below the vocal cords, such as influenza **(Chap. 206)**, acute bronchitis, and pneumonia **(Chap. 131)**, as well as with noninfectious cough **(Chap. 40)**. The average adult has 2–4 URIs per year; children can have 6–10 URIs annually. URIs can be prevented by hand washing or sanitization, physical distancing, use of facial masks, isolation of persons who are ill, and environmental cleaning **(Chap. 204)**.

SARS-CoV-2, the pathogen that causes COVID-19, can cause virtually any upper respiratory symptom **(Chap. 205)**. COVID-19 symptoms appear 2–14 days after exposure and may include fever, chills, cough, shortness of breath, fatigue, myalgias, headaches, rhinorrhea, sore throat, nausea, vomiting, or diarrhea. New loss of taste or smell appears to be specific for COVID-19. Any respiratory symptom occurring in areas where SARS-CoV-2 is circulating should be considered a potential manifestation of COVID-19.

IMPROVING AMBULATORY ANTIBIOTIC PRESCRIBING

The only common acute respiratory infections that should be treated with antibiotics are AOM, sinusitis, streptococcal pharyngitis, and pneumonia. Even for AOM, sinusitis, and pharyngitis, only a minority of cases meet the criteria for antibiotic prescribing. Common respiratory viruses **(Chap. 204)** cause the overwhelming majority of acute respiratory infections, and these infections are generally self-limited; antibiotics neither speed resolution nor prevent complications for the majority of acute respiratory infections. Unfortunately, for this reason, at least half of ambulatory antibiotic prescriptions for acute respiratory infections in the United States are inappropriate. Internationally, population rates of antibiotic prescribing vary nearly threefold, with no differences in infectious complications. Antibiotics cause adverse drug effects, alter the microbiome, cause *Clostridioides difficile* infection **(Chap. 139)**, increase health care costs, and increase the prevalence of antibiotic-resistant bacteria **(Chap. 150)**.

Clinicians prescribe inappropriate antibiotics because of time pressure; fear of missing a rare bacterial diagnosis; concern about preventing a rare bacterial complication; a lack of salience of adverse antibiotic effects; or a mistaken belief that most patients expect, demand, or will not be satisfied without an antibiotic prescription.

AMBULATORY ANTIBIOTIC STEWARDSHIP

Antibiotic stewardship has traditionally been an inpatient concern **(Chap. 149)**, but ambulatory antibiotic use accounts for ~85% of antibiotic use by patients in most developed countries. In 2016, the Centers for Disease Control and Prevention published the "Core Elements of Outpatient Antibiotic Stewardship." The core elements include (1) committing to improving antibiotic prescribing; (2) implementing at least one policy or practice to improve antibiotic prescribing and assessing its effectiveness; (3) monitoring antibiotic prescribing and providing feedback; and (4) providing educational resources to clinicians and patients on antibiotic prescribing. Effective interventions to decrease inappropriate ambulatory antibiotic prescribing include peer comparison, accountable justification, precommitment, clinical decision support, patient education, and multifaceted interventions. Communication training has been particularly effective when it includes making a clear diagnosis, focusing on positive actions patients can take to feel better, reviewing the expected course of illness, and informing patients about concerning symptoms (red flags) for which they should seek or reconnect with care. Telemedicine—synchronous telephone or video or asynchronous electronic messaging—has the potential to improve patient convenience and reduce inappropriate antibiotic prescribing.

Several techniques that seemed promising for the reduction of ambulatory antibiotic prescribing remain unproven, have been ineffective (e.g., procalcitonin testing), or are not durable (e.g., C-reactive protein testing). The practice of delayed antibiotic prescription—i.e., a prescription given to a patient who is asked not to fill it unless symptoms do not improve in a few days—is conceptually flawed and should be avoided. Delayed antibiotic prescriptions are usually given for antibiotic-inappropriate diagnoses (e.g., viral infections); they ignore the natural history of acute respiratory infections, which are self-limited and generally last from 5 to 14 days; they put the burden of clinical decision-making on patients; and they send a confusing, mixed message to patients about the appropriateness of antibiotics for respiratory infections.

NONSPECIFIC UPPER RESPIRATORY INFECTION ("THE COMMON COLD")

DEFINITION AND ETIOLOGY

Nonspecific URI, or the common cold, is a respiratory tract infection in which no single symptom predominates. Nonspecific URI is most commonly caused by respiratory viruses that are acquired through direct contact with infected individuals, contaminated surfaces, and large and small respiratory droplets. The most common viral causes of nonspecific URIs are rhinoviruses (well over 100 serotypes), coronaviruses, parainfluenza virus, respiratory syncytial virus, influenza virus, adenovirus (57 serotypes), metapneumovirus, and bocavirus **(all covered in Chap. 204)**. Making a specific viral diagnosis is not practical, cost-effective, or necessary. Multiplex panels of reverse transcription polymerase chain reaction are available but may be overly sensitive, as prior recent infection can cause false-positive results. Although the diagnosis is usually obvious, clinicians diagnosing a nonspecific URI should also consider influenza **(Chap. 206)**, measles (cough, coryza, and conjunctivitis; **Chap. 211**), acute HIV infection (in which sore throat and rash often predominate; see below and **Chap. 208**), and COVID-19 **(Chap. 205)**.

Individual susceptibility to nonspecific URIs depends on prior exposure, immunity, general health, genetics, microbiome-related factors, and mental health and social factors, including stress. Prior exposure leads to immunity to specific rhinoviruses and adenoviruses, but the number of serotypes makes reinfection likely. Immunity to non-COVID-19 coronaviruses, parainfluenza virus, respiratory syncytial virus, and metapneumoviruses is generally weak or of short duration.

SYMPTOMS AND SIGNS

Common respiratory viruses have incubation periods of 2–8 days after exposure. Symptoms generally begin gradually and include nasal fullness or obstruction, rhinorrhea, sore throat, laryngitis, lymphadenopathy, cough, and low-grade fever. Patients may have myalgias, but this feature usually is not as prominent as it is in influenza. Epistaxis is common with frequent nose blowing.

Findings vary on physical examination, but patients may have conjunctivitis, pharyngeal erythema, pharyngeal exudates, or pharyngeal cobblestoning. Depending on the phase of illness, the nasal mucosa may be pale, boggy, or red and swollen. Nasal mucus can range from watery to purulent. On auscultation, the lungs may be clear, or the patient may have diffuse wheezing or bronchial breath sounds consistent with a viral infection. Symptoms usually last 5–10 days but often last up to 14 days.

TREATMENT

Nonspecific Upper Respiratory Infection

For adults and older children, treatment of nonspecific URI is symptom-based. Fever, myalgias, and sore throat can be treated with acetaminophen or a nonsteroidal anti-inflammatory drug (NSAID) such as ibuprofen. Rhinorrhea can be treated with ipratropium bromide. Nasal congestion can be managed with nasal decongestants such as oxymetazoline (two sprays into each nostril twice a day for up to 5 days) or systemic decongestants such as pseudoephedrine. Products that combine a decongestant with analgesics, antihistamines, or both help relieve symptoms. Although supporting data are weak, cough may be relieved with dextromethorphan or benzonatate (Tessalon Perles). Opioids, while effective at relieving cough, are associated with somnolence, dysphoria, constipation, and dependence.

For children <6 years old, cough and cold medicines should not be prescribed, recommended, or used because of the risks of adverse effects. Honey can help soothe a sore throat for children >1 year old. Cool-mist humidifiers may help with breathing, and saline nasal drops and bulb suctioning can help with nasal congestion.

Patients need to be informed that symptoms generally peak early but can last for up to 14 days; that they are infectious as long as they have symptoms; and that they should rest and drink plenty of fluids to avoid dehydration. Red flags for which patients should seek care include a fever of >102°F, chest pain (other than from a pulled muscle), shortness of breath, dizziness, confusion, new ear or sinus pain, and symptoms lasting >14 days. Although nonspecific URI can be complicated by otitis media and bacterial sinusitis, for an individual patient, an antibiotic is more likely to cause an adverse reaction than to prevent complications.

Other remedies that are ineffective, of questionable benefit, or associated with significant adverse effects include echinacea, zinc, inhaled steam, vitamin C, vitamin D, garlic, antihistamines, Chinese medicinal herbs, intranasal corticosteroids, *Pelargonium sidoides* herbal extract, saline nasal irrigation, and antiviral drugs.

EAR PAIN

Ear pain is most commonly caused by otitis externa and otitis media. In adults, otologic disease is almost always associated with hearing changes. At >50 years of age, temporal arteritis should be considered in patients who have headache, malaise, weight loss, fever, anorexia, and a normal ear exam. Head and neck cancers should be considered in persons with a history of smoking and alcohol use. In children, the presence of a foreign body should be considered.

Ear pain can also result from other causes of local infection, inflammation, trauma, or tumors or can be referred. Innervation of the ear and surrounding areas includes cranial nerves V, VII, IX, and X and cervical nerves C2 and C3. Neuropathic and myopathic pain syndromes (e.g., trigeminal neuralgia) can cause ear pain. Ramsay Hunt syndrome (herpes zoster oticus; **Chap. 452)** and Bell's palsy **(Chap. 452)** are both associated with ear pain.

Dental pathology can cause pain that radiates to the ear; caries and abscesses are most common. Bruxism, malocclusion, and temporomandibular disorder may be associated with tenderness in muscular attachments and the temporomandibular joint. Salivary gland pathology and cervical adenopathy can cause pain that radiates to the ear.

Sinusitis, tonsillitis, and pharyngitis cause pain that can radiate to the ear via cranial nerve IX. Gastroesophageal reflux disease **(Chap. 332)** is often associated with ear symptoms. Myocardial infarction can cause ear pain via cranial nerve X.

Relapsing polychondritis **(Chap. 378)** is a rare condition associated with recurrent, sometimes bilateral, erythematous, or violaceous swelling of the auricle (sparing the earlobe). Inflammation from relapsing polychondritis can involve nasal septal, laryngeal, or respiratory cartilage and can cause ocular inflammation, audiovestibular damage, and nonerosive seronegative inflammatory arthritis.

OTITIS EXTERNA

Etiology and Clinical Manifestations Otitis externa is an inflammation or infection of the external auditory canal manifesting as pain, redness, swelling, aural discharge, and hearing impairment. It is often associated with bacterial infection (frequently by *Pseudomonas aeruginosa* or *Staphylococcus aureus*), but fungi like *Aspergillus* or *Candida* can be implicated.

Otitis externa is most common among preteen and teenage children. Risk factors for otitis externa include swimming (with the resulting condition referred to as "swimmer's ear," which is more common in the summer), mechanical trauma (from cotton swabs or hearing aids), narrow ear canals, cerumen obstruction, eczema, and psoriasis. Classic swimmer's ear is associated with bacterial infection. Physical exam is notable for pain on movement of the auricle or tragus and an external auditory canal that is erythematous, edematous, inflamed, and sometimes coated with exudate on otoscopy. In contrast, fungal otitis externa often manifests with pruritus and ear discharge but without much pain.

Otitis externa can co-occur with otitis media. Preauricular, mastoid, parotid, or cervical lymphadenopathy may be present. AOM with tympanic membrane rupture (see below) can be associated with ear discharge and debris in the ear canal but (unlike otitis externa) without sensitivity to movement of the auricle.

Malignant Otitis Externa Malignant otitis externa is a potentially life-threatening form of otitis externa that involves the temporal bone and occurs in patients with diabetes or other types of immunosuppression, often in older adults. Patients may have fever. Progression of malignant otitis externa can affect cranial nerve VII, IX, XI, or XII.

TREATMENT
Otitis Externa

Analgesia should be provided with acetaminophen or an NSAID. The mainstay of treatment is one or more topical antibacterial drugs with a glucocorticoid for 7–10 days. Polymyxin B–neomycin–hydrocortisone is often used but should be avoided in patients with tympanic membrane perforation because of ototoxicity. Ciprofloxacin–hydrocortisone is an alternative.

Topical aluminum acetate may be as effective as a topical antibacterial–glucocorticoid regimen. For patients whose condition does not improve within 2–4 days with topical treatment, ear wicks or gauze impregnated with or soaked in anti-infective agents can be placed. Ineffective treatments include oral antibiotics and topical antifungals. Otitis externa frequently recurs; its recurrence may be prevented with periodic acetic acid or aluminum acetate drops.

For malignant otitis externa, oral antipseudomonal antibiotics are often prescribed. Patients sometimes require IV pain medication, fluids, or other antimicrobials.

ACUTE OTITIS MEDIA

Epidemiology and Etiology AOM—for which patients almost always present within days—is predominantly a disease of children, with incidence peaking at 6–24 months of age. By age 6, ~60% of children will have had an episode of AOM. Younger children appear to be susceptible because of a shorter, more horizontal eustachian tube that more easily accumulates fluid than it does in older children and adults and because their immune system is still developing.

AOM is caused by a viral URI leading to edema and inflammation of the nasopharynx and eustachian tube, collection of fluid, and infection by bacteria that colonize the nasopharynx. Viruses isolated include respiratory syncytial virus, rhinoviruses, enteroviruses, coronaviruses, influenza virus, adenoviruses, and human metapneumovirus. The bacteria most commonly isolated are *Streptococcus pneumoniae*, nontypeable *Haemophilus influenzae*, and *Moraxella catarrhalis*.

Symptoms and Signs Symptoms of AOM include ear pain, fever, irritability, otorrhea, and anorexia. Physical examination may be notable for a bulging, inflamed, cloudy tympanic membrane, with obscured landmarks, and immobility of the membrane on pneumotoscopy, the Valsalva maneuver, or swallowing while holding the nose shut. (An immobile tympanic membrane is also indicative of perforation, old middle-ear adhesions, a blocked auditory tube, or the presence of middle-ear fluid.) Patients have conductive hearing loss. Severe signs and symptoms include moderate to severe otalgia, otalgia lasting at least 2 days, and a temperature of >102.2°F.

AOM should be diagnosed in children with moderate to severe bulging of the tympanic membrane or new-onset otorrhea (not due to otitis externa). With mild bulging of the tympanic membrane, AOM can also be diagnosed if the patient has had symptoms for <48 h or if there is intense erythema of the tympanic membrane. AOM should *not* be diagnosed in children who do not have middle-ear effusion.

TREATMENT
Acute Otitis Media

Pain from AOM should be treated with NSAIDs or acetaminophen, which are effective for mild to moderate pain. Topical agents like benzocaine, procaine, or lidocaine may provide some additional, brief benefit.

In up to 80% of children, AOM resolves without antibiotics. Indications for antibiotic treatment in children include an age of <6 months, bilateral ear findings in children 6 months to 2 years old, otorrhea in children >6 months old, and—in children of all ages—ear findings with severe otalgia, ear pain for >48 h, or a fever of >102.2°F **(Table 37-1)**.

The benefits of antibiotics are modest and are offset by adverse effects. Antibiotics do not result in early resolution of pain but do decrease pain by day 2 or 3 (number needed to treat, 20 patients treated with antibiotics for 1 patient to have decreased pain by day 2 or 3). More children who receive antibiotics have vomiting, diarrhea, and rash (number needed to harm, 14 patients treated with antibiotics for 1 to have vomiting, diarrhea, or rash). Severe complications like mastoiditis are rare, and the number needed to treat to prevent a case of mastoiditis is ~5000 (i.e., 5000 otitis media patients treated with antibiotics to prevent 1 case of mastoiditis). The American Academy of Family Physicians recommends not routinely prescribing antibiotics for otitis media in children 2–12 years old who have nonsevere symptoms and for whom the observation option is reasonable.

The antibiotic of choice for AOM is high-dose amoxicillin (90 mg/kg per d, up to 3 g). Alternatives include cefdinir, cefuroxime, cefpodoxime, or IM ceftriaxone. If the patient has received amoxicillin in the prior 30 days, clinicians should prescribe amoxicillin/clavulanate (90/6.4 mg/kg per d) in two divided doses. The duration of antibiotic treatment is 10 days for children <2 years old or children with severe symptoms; 5–7 days for children 2–5 years old with mild to moderate AOM; and 5 days for children ≥6 years old with mild or moderate symptoms.

If a patient's condition is not better after 48–72 h of treatment, the antibiotic regimen should be changed to amoxicillin/clavulanate, a second- or third-generation oral cephalosporin, or IM ceftriaxone for 3 days. If, despite a change in antibiotics, the patient's condition still does not improve, that patient should be referred to a specialist. Middle-ear effusions are present in 60–70% of children with AOM; these should resolve over 3 months. Tympanostomy tubes should be

TABLE 37-1 Indications for Antibiotic Treatment of Acute Otitis Media

AGE	INDICATION
<6 months	Antibiotic treatment reasonable for all
6 months to 2 years	Bilateral ear findings
≥6 months	Otorrhea
>2 years	Symptoms worsening or not improving within 48–72 h
All ages	Ear findings with severe otalgia, otalgia lasting at least 2 days, or temperature of >102.2°F

considered for recurrent AOM (i.e., three episodes in 6 months or four episodes in 1 year). Mastoiditis is a rare complication of AOM that is suggested by postauricular tenderness, a postauricular mass, or protrusion of the ear lobe.

In adults, AOM is rare and there is little high-quality evidence to guide treatment. For adults, it remains important to differentiate AOM from OME, but AOM is generally treated with antibiotics, regardless of bilaterality or otorrhea. Amoxicillin is the drug of choice. Adults should also be treated with decongestants and analgesics. Adults with more than two episodes in a year or persistent effusion should be referred to an otolaryngologist.

OTITIS MEDIA WITH EFFUSION

Definition and Etiology OME, also called serous otitis media, occurs when there is fluid in the middle ear but no acute infection. Most patients with OME are young children; >60% of cases occur in children <2 years old. Many children have recurrent episodes.

OME is most often a sequela of a viral infection causing AOM, but it can also be caused by allergies. In addition to allergies, predisposing factors include craniofacial abnormalities, gastroesophageal reflux, and enlarged adenoids.

Symptoms and Signs The most common symptoms are decreases in sound conduction and hearing. Children with OME may exhibit impaired language development or communication difficulties. More rarely, patients complain of intermittent ear fullness or earache, tinnitus, or balance problems. On examination, the tympanic membrane may be translucent or gray with fluid (often colorless or amber), air-fluid levels, or bubbles behind the membrane. There is a loss of the light reflex. The tympanic membrane has decreased mobility on pneumatic otoscopy. The evaluation may include audiometry, tympanometry, and, in infants, measurement of auditory brainstem responses.

OME usually resolves spontaneously within 4–6 weeks. If it persists for >3 months, the condition is referred to as chronic OME or chronic serous otitis media.

Cholesteatomas are accumulations of epithelium or keratin in the middle ear that can enlarge, perforate the tympanic membrane, envelop the ossicles, or destroy surrounding tissue. Cholesteatomas can cause labyrinthitis, hearing loss, cranial nerve palsies, vertigo, meningitis, extradural or brain abscess, and lateral sinus thrombophlebitis.

TREATMENT

Otitis Media with Effusion

OME is treated with myringotomy with tympanostomy tube insertion. For young children with nasal obstruction or recurrent infection, adenoidectomy may be considered. Medications, including antihistamines, glucocorticoids, or antibiotics, do not reliably help. Children at risk for speech or language delay may need earlier referral for more aggressive treatment.

ACUTE MASTOIDITIS

Etiology Acute mastoiditis is a serious infection with significant morbidity despite antibiotic and surgical treatment. This condition is most common among children <2 years old but can occur at any age. Acute mastoiditis is often a complication of AOM but may develop without clinically apparent, prior AOM. In older children with acute mastoiditis, clinicians should suspect cholesteatoma.

The pathogenesis of mastoiditis involves spread of organisms from the middle-ear spaces through the aditus ad antrum to the mastoid air cells. *Incipient* mastoiditis consists of fluid within the mastoid air cells, without bony destruction of the bony septa, and can progress to *coalescent* mastoiditis, with destruction of the bony septa. Acute mastoiditis often causes subperiosteal abscess laterally. The organisms most commonly involved in mastoiditis are *S. pneumoniae*, *Streptococcus pyogenes*, *H. influenzae*, *S. aureus* (including methicillin-resistant *S. aureus* [MRSA] strains), and *P. aeruginosa*.

Symptoms and Signs Symptoms of acute mastoiditis include ear pain, fever, lethargy, or fussiness despite adequate treatment of AOM. Patients—especially those with subperiosteal abscess—may have postauricular erythema, tenderness, warmth, fluctuance, and protrusion of the auricle. Otoscopic examination most often yields findings of AOM and may show superoposterior protrusion of the external auditory canal. Complications of mastoiditis include facial nerve palsy, labyrinthitis, skull osteomyelitis, temporal lobe abscess, cerebellar abscess, meningitis, epidural abscess, subdural abscess, venous sinus thrombosis, or Bezold's abscess (an abscess medial to the sternocleidomastoid that tracks into the deep cervical fascia).

Evaluation Laboratory evaluation reveals elevation of inflammatory markers and white blood cells with neutrophilia. Imaging is not necessary in children with a classic history and presentation but may be required if there is concern about complications or severity. CT may show disruption of bony septations, fluid, mucosal thickening, periosteal thickening, disruption of the periosteum, or subperiosteal abscess. MRI with gadolinium permits better visualization of abscesses and vascular problems.

Differential Diagnosis The differential diagnosis of acute mastoiditis includes cellulitis, otitis externa, postauricular lymphadenopathy, perichondritis, and tumors, including rhabdomyosarcoma, Ewing sarcoma, and myofibroblastic tumor.

TREATMENT

Mastoiditis

Patients with mastoiditis should be admitted to the hospital and treated with IV antibiotics and myringotomy, with or without tympanostomy tubes; if there is no improvement within 48 h, mastoidectomy should be undertaken. Tympanostomy or myringotomy samples or subperiosteal abscess drainage should be sent for culture and sensitivity testing. Depending on complications, additional drainage and surgical procedures may be necessary.

Empirical IV antibiotic therapy for children without recurrent AOM or recent antibiotic treatment consists of vancomycin (if there is concern about antibiotic-resistant *S. pneumoniae* or MRSA) or a cephalosporin (e.g., cefepime or ceftazidime). Patients with recurrent AOM or recent antibiotic treatment should be given vancomycin plus an antipseudomonal penicillin. Culture and sensitivity results will guide antibiotic changes. IV antibiotic therapy should be continued for 7–10 days, and patients should complete a 4-week course of oral antibiotics.

SINUS SYMPTOMS

Sinus symptoms are commonly due to respiratory viruses. These symptoms are considered acute if they last <4 weeks, subacute if they last 4–12 weeks, and chronic if they last ≥12 weeks. Beyond sinus infection, the differential diagnosis of rhinitis includes the common cold, allergic rhinitis **(Chap. 363)**, vasomotor rhinitis, rhinitis medicamentosa due to topical decongestants, drug-induced rhinitis (e.g., due to aspirin, ibuprofen, or beta blockers), autoimmune disease (e.g., granulomatosis with polyangiitis), and cerebrospinal fluid leak. Pain over the sinuses can be caused by headaches **(Chap. 441)**, facial pain syndromes, temporomandibular disorder **(Chap. 38)**, and dental pathology. Gastroesophageal reflux can cause referral of symptoms to the sinuses. Patients who have uncontrolled diabetes or are otherwise immunocompromised can have rapidly progressing invasive fungal infections **(Chap. 217)**. More indolent fungal infections should be considered in the event of recurrent or nonresolving sinusitis. In children, it is important to consider the presence of a foreign body as a cause of sinus symptoms.

ACUTE SINUSITIS

Definition and Etiology *Sinusitis* is an inflammation of the paranasal sinuses; *rhinosinusitis* also involves the nasal passages. The majority of acute sinusitis cases are caused by respiratory viruses. A diagnosis of sinusitis is a major reason for unnecessary antibiotic prescribing in adults: although <2% of sinusitis episodes are due to bacteria (most often *S. pneumoniae*, *H. influenzae*, or *M. catarrhalis*), antibiotics are prescribed at >70% of office visits for sinusitis. According to guideline criteria, no more than 50% of adults—and probably closer to 20%—meet the criteria for antibiotic prescribing.

Symptoms and Signs Sinusitis symptoms commonly include purulent nasal discharge, facial congestion or fullness, and facial pain or pressure. Other symptoms include fever; hyposmia or anosmia; ear pain, pressure, or fullness; postnasal drip; halitosis; maxillary toothache; cough; and fatigue. Risk factors for developing sinusitis include an age of 45–65 years, smoking, asthma, air travel, and allergies.

On physical examination, direct rhinoscopy reveals excess mucus or purulence. Patients may have tenderness over the maxillary sinuses and, in severe cases, erythema and swelling of the maxilla. Sinus transillumination is not accurate in diagnosing sinusitis.

Complications Complications from sinusitis can be dramatic but are extremely rare. These complications may include orbital cellulitis, osteomyelitis, meningitis, intracranial abscesses, and cavernous sinus thrombosis. New symptoms that might indicate a sinusitis complication include confusion, unilateral weakness, proptosis, limited ocular movements, and acute vision changes.

RECURRENT ACUTE SINUSITIS Patients who have four or more episodes of acute sinusitis in a year, without signs or symptoms between episodes, are said to have recurrent acute sinusitis.

INVASIVE FUNGAL SINUSITIS Invasive fungal sinusitis may develop in immunocompromised patients, such as those with uncontrolled diabetes or transplant recipients, and should be considered an emergency. Invasive fungal sinusitis is caused by Mucorales fungi or *Aspergillus* **(Chap. 223)**. Patients may appear to have a rapidly progressive case of rhinosinusitis, with facial pain and pressure, headaches, and fever followed within days by cranial nerve involvement, orbital swelling, cellulitis, proptosis, chemosis, and ophthalmoplegia. Patients may be critically ill. Evaluation should include nasal endoscopy with biopsy and imaging with gadolinium-enhanced MRI as the preferred modality.

NOSOCOMIAL SINUSITIS Nosocomial sinusitis occurs in critically ill patients, often those who are nasotracheally intubated. Nosocomial sinusitis should be suspected in hospitalized patients who have fever without another identifiable cause.

TREATMENT

Acute Sinusitis

All patients with acute sinusitis should be counseled about symptom-based treatments, which may include decongestants, analgesic/antipyretics, nasal saline, or intranasal glucocorticoids. Intranasal decongestants (e.g., oxymetazoline, two sprays in each nostril twice a day for no more than 5 days) and oral decongestants (e.g., 12-h pseudoephedrine [120 mg] during the day) relieve pain, pressure, and rhinorrhea. Analgesics and antipyretics like acetaminophen or NSAIDs (e.g., ibuprofen), nasal saline spray, and nasal washes provide relief. Intranasal glucocorticoids may help, particularly for patients with an allergic cause of sinusitis. Because patients may be accustomed to receiving antibiotics, provision of a clear explanation, symptom-based treatments, and reasons for reconsultation are important. Red flags for which patients should reconsult include recurrent fever of >102°F, sinus symptoms that worsen after initial improvement, and rapid worsening of facial pain that becomes persistent, as well as any other concerning symptoms.

Antibiotic prescribing criteria for sinusitis are based on symptoms **(Table 37-2)**. Only patients with persistent, severe, or worsening symptoms, especially those who have already used decongestants and analgesics for 2–4 days, meet the criteria for antibiotic prescribing. The antibiotic of choice is amoxicillin/clavulanate (875 mg/125 mg bid for 7 days). Amoxicillin (875 mg PO bid for 7 days) is an alternative. For patients with mild penicillin allergies, cefuroxime is a reasonable choice. For those with severe penicillin allergies, doxycycline is a reasonable alternative. Macrolides are specifically not recommended for sinusitis because of high rates of macrolide-resistant *S. pneumoniae.*

TABLE 37-2 Indications for Antibiotic Treatment of Acute Sinusitis

INDICATION	DEFINITION
Persistent	Symptoms lasting ≥10 days
Severe	Fever of >102°F and either purulent nasal discharge or nasal pain for at least 3–4 consecutive days
Worsening	New fever, headache, or increase in nasal discharge following an upper respiratory tract infection that lasted for 5–6 days and was initially improving

Note: In typical populations, roughly 20% and no more than 50% of adults with sinusitis will meet the criteria for antibiotic prescribing.

Patients who meet the criteria for antibiotic prescribing should show signs of improvement after 3–5 days of therapy. If not, second-line regimens include amoxicillin/clavulanate (2000 mg/125 mg bid for 7 days) or levofloxacin, although fluoroquinolones are associated with dysglycemia, neuropathy, and tendon and aortic rupture. For patients whose condition still is not improving after 3–5 days of treatment with a second-line antibiotic or in whom a complication or an alternative diagnosis is suspected, clinicians should consider referral to an otorhinolaryngologist and/or the performance of imaging tests. The imaging modality of choice is noncontrast CT. Patients with recurrent acute sinusitis may benefit from nasal culture during episodes; imaging between episodes to identify predisposing anatomic abnormalities; and allergic or immunologic evaluation.

Patients with acute fungal sinusitis should be treated with IV antifungal agents and often require surgical debridement. Patients with nosocomial sinusitis should have precipitating factors (e.g., nasotracheal intubation) addressed and should be empirically treated with broad-spectrum antibiotics until culture and susceptibility results are available.

CHRONIC SINUSITIS

Definition and Etiology Chronic sinusitis is defined as inflammation of the paranasal sinuses that lasts >12 weeks. Chronic sinusitis is primarily an inflammatory disease and can also be associated with acute or chronic infection or allergic, structural (e.g., deviated nasal septum or polyps), and immunologic etiologies. Repeated viral infections may lead to chronic sinusitis. Bacterial colonization or chronic infection plays a role in some cases of chronic sinusitis. *S. aureus* and gram-negative bacteria are commonly identified. Commonly involved allergens and irritants are dust mites, mold, tobacco smoke, occupational factors, and other airborne toxins. Functional or immunologic problems can include impaired mucociliary clearance (e.g., due to cystic fibrosis) or immunodeficiency due to acquired conditions or medications. Chronic sinusitis often coexists with allergic rhinitis and asthma.

Symptoms and Signs Cardinal symptoms of chronic sinusitis are facial pain or pressure, nasal discharge or postnasal drip, congestion, and hyposmia or anosmia. Associated symptoms may include fatigue, malaise, ear pressure, hoarseness, and cough. The diagnosis of sinus inflammation must be confirmed with anterior rhinoscopy, nasal endoscopy, or imaging because up to 40% of patients with chronic sinus symptoms do not have mucosal changes evidencing disease.

In practical terms, chronic sinusitis can be divided into three main types (in decreasing order of frequency): (1) chronic sinusitis without polyps, (2) chronic sinusitis with polyps, and (3) allergic fungal sinusitis. In general, chronic sinusitis without polyps is more common among women, develops in childhood and young adulthood, is characterized by presentations with facial pain, and is often due to T_H1 lymphocyte predominance associated with bacterial infection or colonization. Chronic sinusitis with polyps is more common among men; develops in adulthood; is characterized by presentations with decrease or loss of smell, asthma, or aspirin sensitivity **(Chap. 298)**; and is often due to T_H2 lymphocyte predominance associated with eosinophilic inflammation, asthma, or aspirin sensitivity. Allergic fungal rhinosinusitis is also associated with polyp formation; typically occurs in patients in their 20s and 30s who are from warm, humid regions and who have other atopic diseases; and is associated with IgE-mediated allergy and eosinophils **(Chap. 223)**. The mucus in allergic fungal rhinosinusitis is classically greenish-brown, has a peanut butter–like consistency, and includes viable hyphae from *Aspergillus* or other fungal species. Allergic fungal rhinosinusitis is resistant to medical treatments.

Evaluation On anterior rhinoscopy, polyps are seen as white, gray, tan, or yellow translucent growths in the middle meatus. The imaging modality of choice is noncontrast CT. Allergic fungal rhinosinusitis may be unilateral; however, unilateral symptoms or polyps on exam or imaging, especially if associated with bloody discharge, should raise concern about tumors.

TREATMENT

Chronic Sinusitis

Treatment includes avoidance of identifiable triggers such as allergens, smoke, and irritants. Saline sprays and washes provide symptom relief, and higher-volume saline washes are probably more effective. Intranasal glucocorticoids, including mometasone and fluticasone sprays or higher-potency and higher-volume budesonide rinses, are mainstays of treatment, especially for chronic sinusitis with polyps. Intranasal glucocorticoids reduce polyp size. Oral administration of glucocorticoids for 2–3 weeks is sometimes effective against chronic sinusitis that is unresponsive to intranasal steroids—again, especially for patients with polyps. Intranasal or systemic antihistamines may help patients whose illness has an allergic component. Likewise, leukotriene antagonists like montelukast may help.

Although antibiotics are frequently prescribed for 2–4 weeks to patients with chronic sinusitis, there is little evidence that these drugs are effective. Evidence of modest quality supports the use of 3 months of macrolide treatment for patients who have chronic sinusitis without polyps. Antifungal agents have not shown benefit against any subtype of chronic sinusitis. Decongestants should be used only sparingly and briefly.

Endoscopic sinus surgery improves quality of life in patients who have had inadequate responses to medical therapy. Patients with more limited, focal disease may more reliably have better results. The goals of surgery are to remove polyps from the nasal cavity and paranasal sinuses. For patients with allergic fungal rhinosinusitis, medical therapy is classically ineffective, surgery produces good results, and patients should be treated with perioperative glucocorticoids. In children, adenoidectomy may be effective in some cases. In the future, immune endotyping may allow selection of more individualized biologic treatments.

SORE THROAT AND NECK PAIN

Sore throat is not synonymous with pharyngitis and can also be caused by submandibular space, retropharyngeal and peritonsillar abscesses, thyroiditis, gastroesophageal reflux, tumors, and postnasal drainage.

Acute pharyngitis, in which symptoms are generally present for days, is most often caused by respiratory viruses; is often caused by group A β-hemolytic streptococci (GAS); and can be caused by other bacteria (including *Neisseria gonorrhoeae*), Epstein-Barr virus (EBV), and HIV. On physical examination, pharyngeal erythema is associated most commonly with viral infections, including the common cold and influenza. Pharyngeal exudate should not be confused with *Candida* infection, which looks like cottage cheese, can be scraped off, and leaves a bleeding surface, or leukoplakia, that cannot be scraped off. History and exam findings may help differentiate sore throat and pharyngitis of various etiologies **(Table 37-3)**.

TABLE 37-3 Clinical Findings That Suggest Various Forms of Nonstreptococcal Pharyngitis

CLINICAL FINDING(S) OR BEHAVIORAL FACTOR	SUSPECTED DIAGNOSIS
Scarlatiniform rash	Group A β-hemolytic streptococci or *Arcanobacterium haemolyticum*
Cough and otitis media	*Haemophilus influenzae*
Sex between men with associated urogenital symptoms, fellatio between a woman and a man who has current urogenital symptoms, persistent sore throat unresponsive to penicillin	*Neisseria gonorrhoeae*
Travel to endemic areas, pseudomembrane on examination	*Corynebacterium diphtheriae*
Persistent sore throat with bronchopulmonary symptoms	*Mycoplasma pneumoniae*
Marked adenopathy (especially that involving posterior cervical or auricular nodes), splenomegaly, palatine petechiae, gelatinous uvula	Acute infectious mononucleosis
New sexual partner in the previous month; fever, rash, myalgias, headache	Acute HIV infection

STREPTOCOCCAL PHARYNGITIS

GAS is the only common cause of sore throat that should be treated with antibiotics. The principal goal in the evaluation of adults with sore throat is to identify patients likely to have GAS pharyngitis, or “strep throat.” Prompt antibiotic treatment of adults likely to have strep throat has the potential to reduce symptoms, prevent the spread of disease, and reduce suppurative complications (e.g., peritonsillar abscess). Nonsuppurative complications are rare. In developed countries, the prevalence of rheumatic fever **(Chap. 153)** is extremely low, and antibiotic treatment does not prevent poststreptococcal glomerulonephritis **(Chap. 153)**.

Most patients with non-GAS pharyngitis have various forms of viral pharyngitis and do not require antibiotics. Nevertheless, clinicians prescribe antibiotics to a majority of adults with sore throats. By using a simple clinical scoring algorithm, clinicians can predict the presence or absence of GAS with sufficient accuracy and avoid prescribing antibiotics to patients who are unlikely to have strep throat. Although there is a role for testing (see “Evaluation,” below), most adults with sore throat do not need to have a GAS test.

About 10% of adults with sore throat are infected with GAS. Among children with sore throat, the prevalence of GAS can be as high as 35%, with rates peaking from 5 to 15 years of age. The prevalence of GAS is higher in winter and early spring. The risk of streptococcal pharyngitis is elevated among health care and child care workers, teachers, parents of young children, and patients exposed to individuals with strep throat. Clinicians need to be aware of local outbreaks of GAS infection, particularly in military and institutional settings, where the prevalence of GAS and the risk of acute rheumatic fever may be elevated.

Evaluation The Centor criteria consist of four findings, each of which is assigned 1 point: (1) history of fever, (2) absence of cough, (3) tender anterior cervical lymphadenopathy, and (4) tonsillar exudate or swelling. The Centor criteria are easy to assess and accurately stratify adult patients with suspected streptococcal pharyngitis. Patients with no points have a 2% probability of being infected with GAS, whereas those with 4 points have a probability of 41% **(Table 37-4)**. The Centor criteria have an area under the curve of 0.79. Other clinical decision algorithms similar to the Centor criteria may not perform as well, are not as simple, or have not been as rigorously evaluated.

If the test/no treatment threshold is set at 5%, for a GAS prevalence of ~10%, adults meeting no criteria or only one Centor criterion have a probability of GAS pharyngitis so low that they should neither be tested nor be treated with an antibiotic. Adults meeting two or three Centor criteria have an intermediate probability of GAS pharyngitis; they should have a rapid antigen test performed, and the results should

TABLE 37-4 The Centor Criteria and the Probability of Streptococcal Pharyngitis for Adults[a]

NO. OF CRITERIA MET[b]	POSTEVALUATION PROBABILITY (%)	RECOMMENDATION
0	2	No test, no antibiotic
1	3	No test, no antibiotic
2	8	Rapid test
3	19	Rapid test
4	41	Empirical antibiotic treatment or rapid test

[a]Assuming a pretest probability of strep throat for adults of 10%. [b]The criteria are (1) a history of fever, (2) an absence of cough, (3) tender anterior cervical lymphadenopathy, and (4) tonsillar swelling or exudate. Each criterion gets 1 point. Roughly 40–60% of adults will meet no criteria or one criterion; ~20% will meet the criteria for antibiotic prescribing.

guide antibiotic treatment. For adults meeting four Centor criteria, it is reasonable either to perform a rapid test or to institute empirical antibiotic treatment. However, some guidelines recommend—and some ambulatory quality measures require—a GAS test to be associated with antibiotic prescribing in adults, regardless of the number of Centor criteria met.

In children, the Centor criteria are less specific, and streptococcal pharyngitis should be confirmed with testing. Children who have signs of pharyngitis without signs of viral infection (conjunctivitis, runny nose, cough, hoarseness, nonexudative oral lesions) should have testing performed.

Outside of the United States, because complications are rare and even streptococcal pharyngitis is self-limited in the vast majority of cases, some guidelines do not recommend use of rapid GAS testing or routine antibiotic treatment of sore throat.

Clinicians should have a lower threshold for diagnosing and treating GAS pharyngitis in patients with a history of acute rheumatic fever, patients with documented streptococcal exposure in the past week, patients who live in a community with a current strep throat epidemic, and patients who are diabetic or otherwise immunocompromised.

RAPID STREP TESTS Rapid GAS-specific antigen tests have a sensitivity of ~80% and a specificity of ~95%. Results are available within minutes and can be used to make therapeutic decisions before the patient leaves the office. Improper collection technique can adversely affect the sensitivity of rapid strep tests: clinicians should rub the tonsils and pharynx, touching any areas where exudate or ulceration are present.

THROAT CULTURES A single-swab throat culture has a sensitivity of ~85–90%, as defined by isolation of GAS on a second swab. A throat culture can also be falsely positive for true infection: some patients with a culture positive for GAS may be only uninfected carriers, as defined by their failure to exhibit a fourfold increase in antibodies to GAS—the gold standard test. Among adults and children seeking medical care for a sore throat, test specificity may be as low as 50–70% because of patients who do not exhibit serologic evidence of infection. Throat cultures are not recommended for the routine evaluation of adults with sore throat. The modest gain in sensitivity over rapid testing is outweighed by the 24- to 48-h delay in test results, with a consequent delay in the symptomatic relief associated with antibiotic treatment.

Indiscriminate strep testing in adults with sore throat or respiratory symptoms should be discouraged. Rapid strep tests and culture do not differentiate between patients who have true infection and those who are carriers of GAS (with carriage rates as high as 20% among schoolchildren and ~5% among adolescents and young adults). In adults who meet no Centor criteria or only one criterion—40–60% of adults with pharyngitis—a positive test is highly likely to be falsely positive and/or to represent GAS carriage.

Complications Complications of streptococcal pharyngitis are rare but include acute rheumatic fever, poststreptococcal glomerulonephritis, scarlet fever (all covered in Chap. 153), sinusitis, peritonsillar abscess, and other invasive GAS infections.

TREATMENT

Streptococcal Pharyngitis

All patients with pharyngitis—nonstreptococcal and streptococcal—should receive analgesics (acetaminophen or NSAIDs). Saline gargles, humidification, soft foods, and tea with honey soothe a painful throat.

Penicillin is the antibiotic of choice for streptococcal pharyngitis (Table 37-5). Penicillin is a narrow-spectrum, low-cost, and well-tolerated drug to which no GAS isolate has been resistant. Amoxicillin is an acceptable alternative in children as it comes in a palatable liquid form. For patients with mild penicillin allergy, cephalexin and cefadroxil are good alternatives. For patients with severe penicillin allergies, clinicians should prescribe erythromycin, clarithromycin, or clindamycin. Unlike other infections for which emerging evidence supports progressively shorter antibiotic courses, streptococcal pharyngitis requires longer courses (7–10 days), which are more effective.

Glucocorticoids (e.g., dexamethasone, 10 mg as a single oral dose) have so far been poorly studied as an adjunctive treatment for sore throat and strep throat and are not recommended. These drugs may result in decreased pain within 24 h but do not decrease school or work absenteeism or relapse rates. Even short courses of steroids are associated with increased rates of sepsis, gastrointestinal bleeding, congestive heart failure, venous thromboembolism, and fracture within 30 days.

Streptococcal and nonstreptococcal pharyngitis should resolve in 3–5 days. Symptoms that should lead patients to seek further care include shaking chills (rigors), neck swelling (beyond lymphadenopathy), trouble swallowing, drooling, or symptoms that persist for >5 days without improvement.

NONSTREPTOCOCCAL PHARYNGITIS

Acute Infectious Mononucleosis New EBV infection may be the cause of pharyngitis in 1–6% of young adults (Chap. 199). EBV is rarely the cause of pharyngitis in adults >40 years of age. The full-blown acute syndrome, which is present in only about one-fourth of patients with infectious mononucleosis ("mono"), is characterized by a triad of clinical, hematologic, and serologic findings. The clinical presentation is typified by the development over several days of malaise, fever, sore throat, and marked adenopathy that is particularly evident in the cervical lymph nodes. On physical examination, marked adenopathy is virtually always documented and is most specific for mononucleosis when the posterior cervical or posterior auricular nodes are involved. Splenomegaly and exudative pharyngitis with prominent tonsillar swelling, palatine petechiae, and a gelatinous uvula are often noted. The classic hematologic findings are an absolute lymphocyte count of >4000/μL or a relative lymphocyte count of >50% with "atypical"

TABLE 37-5 Antibiotic Treatment of Group A Streptococcal Pharyngitis

ANTIBIOTIC	DOSING
Antibiotic of Choice	
Penicillin	500 mg PO qid or 1000 mg PO bid × 10 days
Alternative for Non-Penicillin-Allergic Patients	
Amoxicillin	500 mg PO bid or 1000 mg PO qd × 10 days
Alternatives for Non-Anaphylactic Penicillin-Allergic Patients	
Cephalexin	500 mg PO bid × 10 days
Cefadroxil	1 g PO qd × 10 days
Alternatives for Patients with Severe Penicillin Allergy	
Erythromycin	250–500 mg PO qid or 500–1000 mg PO bid × 5 days
Clarithromycin	500 mg PO bid × 5 days
Clindamycin	300 mg PO tid × 10 days

morphologic features in >10% of the lymphocytes. The characteristic serologic finding is the heterophil antibody, which is detectable in only 40% of patients during the first week of illness but in 80–90% of patients by the third week.

Other Bacterial Pharyngitis Non–group A streptococci (especially group C and group G streptococci), *Mycoplasma pneumoniae*, *Chlamydia pneumoniae*, *N. gonorrhoeae*, and *H. influenzae* have all been associated with sore throat in some studies. Although antibacterial treatment has not been proven to speed the resolution of symptoms and signs of any of these types of nonstreptococcal pharyngitis, antibiotic treatment is indicated if throat cultures from a patient with persistent sore throat yield group C or group G streptococci.

Lemierre Syndrome Lemierre syndrome consists of septic thrombophlebitis of the internal jugular vein accompanied by metastatic infections, most commonly of the lung but with possible involvement of the joints, bones, liver, meninges, and brain. Lemierre syndrome is most commonly caused by *Fusobacterium necrophorum*, although it can also be caused by species of *Bacteroides*, *Eikenella*, *Streptococcus*, *Peptostreptococcus*, or other bacterial genera. This syndrome probably occurs predominantly in male patients. Clinicians should consider Lemierre syndrome in a teenage or young adult patient who has non-GAS pharyngitis that is not resolving, particularly if it is accompanied by rigors, neck pain or swelling, or other extrapharyngeal symptoms **(see also Chaps. 127, 132, and 182)**.

Gonococcal Pharyngitis *N. gonorrhoeae* may be the cause of pharyngitis in 1% of adult patients seeking primary care for a sore throat, although gonococcal infection of the pharynx is more often asymptomatic. When symptomatic, pharyngeal gonorrhea may range from mild to severe, with protracted pharyngitis characterized by pain, fever, and pharyngeal exudate. Gonococcal pharyngitis should be suspected in men who have sex with men with associated symptoms of urogenital infection, women who have practiced fellatio with a man with genital gonorrhea, and anyone who has persistent sore throat that has been unresponsive to treatment for presumptive streptococcal pharyngitis.

Diphtheria Diphtheria, caused by *Corynebacterium diphtheriae,* is endemic in developing countries **(Chap. 155)**. Diphtheria produces only mild pharyngitis beneath its characteristic grayish pseudomembrane.

Acute HIV Infection Clinicians should consider acute HIV infection in patients with sore throat, particularly when it is associated with headache, fever, myalgias, lymphadenopathy, anorexia, and rash **(Chap. 208)**. Of patients with acute HIV infection, roughly half have a sore throat. However, in most settings in the United States, only ~1% of patients with viral or mononucleosis-like symptoms have acute HIV infection.

■ HEAD AND NECK ABSCESSES

Head and neck abscesses are more common among patients with diabetes, those who are immunocompromised, and older adults. Such abscesses are often a complication of infections of the teeth and gums, throat, or salivary ducts; lymphadenitis; ear infections; sinus infections; congenital cysts; and IV drug use. Prompt recognition is important, as head and neck abscesses can cause airway compromise due to edema or mass effect. Head and neck abscesses can follow fascial planes and spread to the mediastinum (where they can cause mediastinitis, pleural effusions, empyema, or pericarditis), the carotid sheath, the skull base, and the meninges. Head and neck abscesses have also been associated with aspiration pneumonia, necrotizing fasciitis, Lemierre syndrome, and toxic shock syndrome.

Submandibular abscesses generally result from an infected or extracted tooth and can cause Ludwig angina, a swelling of the floor of the mouth that can enlarge and displace the tongue posteriorly.

Peritonsillar abscesses, which may occur predominantly in male patients, generally result from complicated bacterial pharyngitis and present with fever, dysphagia, profound throat pain (necessitating drooling to avoid swallowing saliva), trismus, and "hot potato voice" (inability to articulate, as if patients have hot food in their mouths). Patients are likely to have unilateral palate bulging, often with uvular deviation. Peritonsillar abscesses are caused by viridans group streptococci, β-hemolytic streptococci, *F. necrophorum*, *S. aureus*, *Prevotella*, and *Bacteroides*.

Retropharyngeal abscesses often present after an antecedent URI in children with sore throat, dysphagia, deep neck pain, neck stiffness, trismus, and drooling. The pharyngeal wall may be displaced, but swelling or abscess may not be apparent on examination. In severe cases, patients may have dyspnea and stridor.

Patients with suspected head and neck abscesses, with the possible exception of patients who have obvious peritonsillar abscesses, should undergo imaging by CT.

TREATMENT

Head and Neck Abscesses

The mainstays of treatment for head and neck abscesses are securing the airway, surgical drainage, and IV antibiotic administration. To secure the airway, mask ventilation or oral intubation may not be effective, and oral fiberoptic intubation or tracheotomy may be necessary. Peritonsillar abscess may be managed with needle aspiration and/or tonsillectomy. Other head and neck abscesses require incision and drainage. The selected IV antibiotics should cover streptococci, anaerobes, and possibly *S. aureus*. Frequently used antibiotics include ampicillin/sulbactam, clindamycin plus ceftriaxone, or meropenem. For some abscesses with adequate source control with incision and drainage, penicillin may be as effective as broader-spectrum agents.

■ EPIGLOTTITIS

Along with associated dysphagia, odynophagia, hoarseness, and stridor or tachypnea, supraglottitis or epiglottitis must be considered in adults presenting with sore throat. The inflamed and enlarged epiglottis protrudes up into the oropharynx. Patients may extend their neck or lean forward and drool oral secretions to avoid swallowing. Epiglottitis can cause "hot potato voice." Attempts to examine or swab the posterior pharynx or obtain a culture can provoke laryngospasm and should only be done carefully in a controlled setting. Because obstruction of the airway may become acutely life-threatening, the patient with epiglottitis must be observed in a hospital setting; examination in an operating room, where an airway can be established immediately by an experienced operator, should be strongly considered. Although not necessary for the diagnosis, a lateral neck radiograph can demonstrate epiglottal swelling referred to as the "thumb sign."

In adults, conservative therapy under observation is sufficient in most cases, but intubation by an experienced clinician or tracheostomy may become necessary. Treatments also include humidification with nebulized normal saline or humidified oxygen and administration of glucocorticoids, IV antibiotics, and nebulized epinephrine.

H. influenzae, the most common cause of supraglottitis in children, is less common in adults. Other responsible organisms in adults are *S. pneumoniae*, *S. pyogenes*, and *S. aureus*. The *H. influenzae* type b vaccine has led to a dramatic decrease in epiglottitis overall, with large reductions in young children; however, the incidence of supraglottitis and epiglottitis in adults may be increasing.

■ LARYNGITIS

Laryngitis—inflammation of the larynx and surrounding structures—is most commonly caused by viral URIs. In children, parainfluenza virus can cause croup, or laryngotracheobronchitis, which is characterized by a "barking" cough but can also include laryngitis.

Beyond viruses, laryngitis can be caused in rare cases by bacteria and fungi. Bacterial laryngitis can be a complication of viral laryngitis, occurring about 7 days into the illness. The most common bacteria involved are *S. pneumoniae*, *H. influenzae*, and *M. catarrhalis.* Fungal laryngitis is probably rarer but should be considered in patients who are immunosuppressed or who have recently been treated with antibacterial drugs.

Noninfectious causes of laryngitis include vocal trauma (e.g., due to yelling, screaming, or loud singing), inhalation injuries, allergies, gastroesophageal reflux disease (laryngopharyngeal reflux), asthma, and pollution. Immunosuppressed patients are at risk for infections with herpesvirus, HIV, and coxsackievirus. Smokers are at elevated risk for malignancy and other infections.

Laryngitis is characterized by a raspy, hoarse, or breathy voice, sometimes progressing to a complete loss of voice. Laryngitis can have associated dry cough and anterior throat pain; patients often feel a need to clear their throats. The physical examination in patients who may have laryngitis should focus on the head, neck, and lungs, but the diagnosis of laryngitis is generally based on history. If visualization of the vocal cords is necessary, indirect examination with a mirror or flexible laryngoscopy usually shows erythema and edema of the vocal cords and surrounding structures.

TREATMENT

Laryngitis

Laryngitis is generally self-limited, usually lasting 3–7 days, but may last up to 14 days. Vocal rest is crucial. Airway humidification and hydration should help. Patients likely to have laryngopharyngeal reflux should avoid gastroesophageal reflux–inducing foods and behaviors and should take antireflux medications. In randomized controlled trials, antibiotics were not effective in decreasing objective symptoms of laryngitis.

Red flags for emergency evaluation and monitoring include shortness of breath, stridor, dysphagia, odynophagia, drooling, and posturing that could indicate epiglottitis. Referral to an otolaryngologist should be considered for patients who rely on their voice for work, such as singers and teachers. A history of smoking or weight loss should raise suspicion of malignancy. Symptoms lasting >3 weeks should prompt referral to an otolaryngologist or speech specialist.

FURTHER READING

Centor RM, Linder JA: Web exclusive. Annals on call—*Fusobacterium* pharyngitis debate. Ann Intern Med 171:OC1, 2019.

Chua KP et al: Appropriateness of outpatient antibiotic prescribing among privately insured US patients: ICD-10-CM based cross sectional study. BMJ 364:k5092, 2019.

LiebertHal AS et al: Clinical practice guideline: The diagnosis and management of acute otitis media. Pediatrics 131:e964, 2013.

Rowe TA, Linder JA: Novel approaches to decrease inappropriate ambulatory antibiotic use. Expert Rev Anti Infect Ther 17:511, 2019.

Sanchez GV et al: Antibiotic stewardship in outpatient telemedicine: Adapting Centers for Disease Control and Prevention core elements to optimize antibiotic use. Telemed J E Health 30:951, 2024.

38

Oral Manifestations of Disease

Samuel C. Durso

Internists are often asked to evaluate patients with disease of the oral soft tissues, teeth, and pharynx. Knowledge of the oral milieu and its unique structures is necessary to guide preventive services and recognize oral manifestations of local or systemic disease **(Chap. A3)**. Furthermore, internists frequently collaborate with dentists in the care of patients who have a variety of medical conditions that affect oral health or who undergo dental procedures that increase their risk of medical complications.

DISEASES OF THE TEETH AND PERIODONTAL STRUCTURES

Tooth formation begins during the sixth week of embryonic life and continues through 17 years of age. Teeth start to develop in utero and continue to develop until after the tooth erupts. Normally, all 20 deciduous teeth have erupted by age 3 and have been shed by age 13. Permanent teeth, eventually totaling 32, begin to erupt by age 6 and have completely erupted by age 14, though third molars ("wisdom teeth") may erupt later.

The erupted tooth consists of the visible *crown* covered with enamel and the root submerged below the gum line and covered with bonelike *cementum*. *Dentin*, a material that is denser than bone and exquisitely sensitive to pain, forms the majority of the tooth substance, surrounding a core of myxomatous *pulp* containing the vascular and nerve supply. The tooth is held firmly in the alveolar socket by the *periodontium*, supporting structures that consist of the gingivae, alveolar bone, cementum, and periodontal ligament. The periodontal ligament tenaciously binds the tooth's cementum to the alveolar bone. Above this ligament is a collar of attached gingiva just below the crown. A few millimeters of unattached or free gingiva (1–3 mm) overlap the base of the crown, forming a shallow sulcus along the gum-tooth margin.

Dental Caries, Pulpal and Periapical Disease, and Complications Dental caries usually begin asymptomatically as a destructive infectious process of the enamel. Bacteria—principally *Streptococcus mutans*—colonize the organic buffering biofilm (*plaque*) on the tooth surface. If not removed by brushing or by the natural cleansing and antibacterial action of saliva, bacterial acids can demineralize the enamel. Fissures and pits on the occlusal surfaces are the most frequent sites of early decay. Surfaces between the teeth, adjacent to tooth restorations and exposed roots, are also vulnerable, particularly as individuals age. Over time, dental caries extend to the underlying dentin, leading to cavitation of the enamel. Without management, the caries will penetrate to the tooth pulp, producing *acute pulpitis*. At this stage, when the pulp infection is limited, the tooth may become sensitive to percussion and to hot or cold, and pain resolves immediately when the irritating stimulus is removed. Should the infection spread throughout the pulp, *irreversible pulpitis* occurs, leading to *pulp necrosis*. At this later stage, pain can be severe and has a sharp or throbbing visceral quality that may be worse when the patient lies down. Once pulp necrosis is complete, pain may be constant or intermittent, but cold sensitivity is lost.

Treatment of caries involves removing the softened and infected hard tissue and restoration of the tooth structure with silver amalgam, glass ionomer, composite resin, or gold. Once irreversible pulpitis occurs, root canal therapy becomes necessary; removal of the contents of the pulp chamber and root canal is followed by thorough cleaning and filling with an inert material. Alternatively, the tooth may be extracted.

Pulpal infection leads to *periapical abscess* formation, which can produce pain on chewing. If the infection is mild and chronic, a *periapical granuloma* or eventually a *periapical cyst* forms, either of which produces radiolucency at the root apex. When unchecked, a periapical abscess can erode into the alveolar bone, producing osteomyelitis; penetrate and drain through the gingivae, producing a parulis (gumboil); or track along deep fascial planes, producing virulent cellulitis (Ludwig's angina) involving the submandibular space and floor of the mouth **(Chap. 182)**. Elderly patients, patients with diabetes mellitus, and patients taking glucocorticoids may experience little or no pain or fever as these complications develop.

Periodontal Disease Periodontal disease and dental caries are the primary causes of tooth loss. Like dental caries, chronic infection of the gingiva and anchoring structures of the tooth begins with formation of bacterial plaque. The process begins at the gum line. Plaque and *calculus* (calcified plaque) are preventable by appropriate daily oral hygiene, including periodic professional cleaning. Left undisturbed, chronic inflammation can ensue and produce hyperemia of the free and attached gingivae (*gingivitis*), which then typically bleed with brushing. If this issue is ignored, severe *periodontitis* can develop, leading to deepening of the physiologic sulcus and destruction of the periodontal ligament. Gingival pockets develop around the teeth. As the periodontium (including the supporting bone) is destroyed, the teeth loosen. A role for chronic inflammation due to chronic periodontal disease in promoting coronary heart disease and stroke has been proposed. Epidemiologic studies have demonstrated a moderate but significant association between chronic periodontal inflammation and atherogenesis, though a causal role remains unproven.

Acute and aggressive forms of periodontal disease are less common than the chronic forms described above. However, if the host is stressed or exposed to a new pathogen, rapidly progressive and destructive disease of the periodontal tissue can occur. A virulent example is *acute necrotizing ulcerative gingivitis*. The presentation includes sudden gingival inflammation, ulceration, bleeding, interdental gingival necrosis, and fetid halitosis. *Localized juvenile periodontitis*, which is seen in adolescents, is particularly destructive and appears to be associated with impaired neutrophil chemotaxis. *AIDS-related periodontitis* resembles acute necrotizing ulcerative gingivitis in some patients and a more destructive form of adult chronic periodontitis in others. It may also produce a gangrene-like destructive process of the oral soft tissues and bone that resembles *noma*, an infectious condition seen in severely malnourished children in developing nations.

Prevention of Tooth Decay and Periodontal Infection Despite the reduced prevalences of dental caries and periodontal disease in the United States (due in large part to water fluoridation and improved dental care, respectively), both diseases constitute a major public health problem worldwide. The internist should promote preventive dental care and hygiene as part of health maintenance. Populations at high risk for dental caries and periodontal disease include those with hyposalivation and/or xerostomia, diabetics, alcoholics, tobacco users, persons with Down syndrome, and those with gingival hyperplasia. Furthermore, patients lacking access to dental care (e.g., due to low socioeconomic status) and patients with a reduced ability to provide self-care (e.g., individuals with disabilities, nursing home residents, and persons with dementia or upper-extremity disability) suffer disproportionately. It is important to provide counseling (to patients and/or their caregivers) regarding regular dental hygiene and professional cleaning, use of fluoride-containing toothpaste, professional fluoride treatments, and (for patients with limited dexterity) use of electric toothbrushes. Cost, fear of dental care, and differences in language and culture create barriers that prevent some people from seeking preventive dental services.

Developmental and Systemic Disease Affecting the Teeth and Periodontium In addition to posing cosmetic issues, *malocclusion*, the most common developmental oral problem, can interfere with mastication unless corrected through orthodontic and surgical techniques. Impacted third molars are common and can become infected or erupt into an insufficient space. Acquired prognathism due to *acromegaly* may also lead to malocclusion, as may deformity of the maxilla and mandible due to *Paget's disease* of the bone. Delayed tooth eruption, a receding chin, and a protruding tongue are occasional features of *cretinism* and *hypopituitarism*. Congenital syphilis produces tapering, notched (*Hutchinson's*) incisors and finely nodular (*mulberry*) molar crowns. *Enamel hypoplasia* results in crown defects ranging from pits to deep fissures of primary or permanent teeth. Intrauterine infection (syphilis, rubella), vitamin deficiency (A, C, or D), disorders of calcium metabolism (malabsorption, vitamin D–resistant rickets, hypoparathyroidism), prematurity, high fever, and rare inherited defects (*amelogenesis imperfecta*) are all causes. Tetracycline, given in sufficiently high doses during the first 8 years of life, may produce enamel hypoplasia and discoloration. Doxycycline does not cause permanent tooth staining in children despite warnings included for all tetracycline-class antibiotics. Worn enamel is seen with age, bruxism, or excessive acid exposure (e.g., chronic gastric reflux or bulimia). Celiac disease is associated with nonspecific enamel defects in children but not in adults.

Total or partial tooth loss resulting from periodontitis is seen with cyclic neutropenia, Papillon-Lefévre syndrome, Chédiak-Higashi syndrome, and leukemia. Rapid focal tooth loosening is most often due to infection, but rarer causes include Langerhans cell histiocytosis, Ewing's sarcoma, osteosarcoma, and Burkitt's lymphoma. Early loss of primary teeth is a feature of *hypophosphatasia*, a rare congenital error of metabolism.

Pregnancy may produce gingivitis and localized *pyogenic granulomas*. Severe periodontal disease occurs in uncontrolled diabetes mellitus. *Drug-induced gingival overgrowth* may be caused by anticonvulsants, calcium channel blockers, and immunosuppressants, although excellent daily oral care can prevent or reduce its occurrence. *Idiopathic familial gingival fibromatosis* and several syndrome-related disorders cause similar conditions. Discontinuation of the medication may reverse the drug-induced form, although surgery may be needed to control both of the latter entities. *Linear gingival erythema* is variably seen in patients with advanced HIV infection and probably represents immune deficiency and decreased neutrophil activity. Diffuse or focal gingival swelling may be a feature of early or late acute myelomonocytic leukemia as well as of other lymphoproliferative disorders. A rare but pathognomonic sign of granulomatosis with polyangiitis is a red-purplish, granular gingivitis (*strawberry gums*).

DISEASES OF THE ORAL MUCOSA

Infections Most oral mucosal diseases involve microorganisms (**Table 38-1**).

Pigmented Lesions See **Table 38-2**.

Dermatologic Diseases See Tables 38-1, 38-2, and **38-3** and **Chaps. 59–64**.

Diseases of the Tongue See **Table 38-4**.

HIV Disease and AIDS See Tables 38-1, 38-2, 38-3, and **38-5**; **Chap. 208**.

Ulcers Ulceration is the most common oral mucosal lesion. Although there are many causes, the host and the pattern of lesions, including the presence of organ system features, narrow the differential diagnosis (Table 38-1). Most acute ulcers are painful and self-limited. Recurrent aphthous ulcers and herpes simplex account for the majority. Persistent and deep aphthous ulcers can be idiopathic or can accompany HIV/AIDS. Aphthous lesions are often the presenting symptom in *Behçet's syndrome* **(Chap. 376)**. Similar-appearing, though less painful, lesions may occur in reactive arthritis, and aphthous ulcers are occasionally present during phases of *discoid* or *systemic lupus erythematosus* **(Chap. 372)**. Aphthous-like ulcers are seen in *Crohn's disease* **(Chap. 337)**, but, unlike the common aphthous variety, they may exhibit granulomatous inflammation on histologic examination. Recurrent aphthae are more prevalent in patients with *celiac disease* and have been reported to remit with elimination of gluten.

Of major concern are chronic, relatively painless ulcers and mixed red/white patches (erythroplakia and leukoplakia) of >2 weeks' duration. Squamous cell carcinoma and premalignant dysplasia should be considered early and a diagnostic biopsy performed. This awareness and this procedure are critically important because early-stage malignancy is vastly more treatable than late-stage disease. High-risk sites include the lower lip, floor of the mouth, ventral and lateral tongue, and soft palate–tonsillar pillar complex. Significant risk factors for oral cancer in Western countries include sun exposure (lower lip), tobacco and alcohol use, and human papillomavirus infection. In India and some other Asian countries, smokeless tobacco mixed with betel nut, slaked lime, and spices is a common cause of oral cancer. Rarer causes of chronic oral ulcer, such as tuberculosis, fungal infection, granulomatosis with polyangiitis, and midline granuloma, may look identical to carcinoma. Making the correct diagnosis depends on recognizing other clinical features and performing a biopsy of the lesion. The syphilitic

TABLE 38-1 Vesicular, Bullous, or Ulcerative Lesions of the Oral Mucosa

CONDITION	USUAL LOCATION	CLINICAL FEATURES	COURSE
Viral Diseases			
Primary acute herpetic gingivostomatitis (HSV type 1; rarely type 2)	Lip and oral mucosa (buccal, gingival, lingual mucosa)	Labial vesicles that rupture and crust, and intraoral vesicles that quickly ulcerate; extremely painful; acute gingivitis, fever, malaise, foul odor, and cervical lymphadenopathy; occurs primarily in infants, children, and young adults	Heals spontaneously in 10–14 days; unless secondarily infected, lesions lasting >3 weeks are not due to primary HSV infection
Recurrent herpes labialis	Mucocutaneous junction of lip, perioral skin	Eruption of groups of vesicles that may coalesce, then rupture and crust; painful to pressure or spicy foods	Lasts ~1 week, but condition may be prolonged if secondarily infected; if severe, topical or oral antiviral treatment may reduce healing time
Recurrent intraoral herpes simplex	Palate and gingiva	Small vesicles on keratinized epithelium that rupture and coalesce; painful	Heals spontaneously in ~1 week; if severe, topical or oral antiviral treatment may reduce healing time
Chickenpox (VZV)	Gingiva and oral mucosa	Skin lesions may be accompanied by small vesicles on oral mucosa that rupture to form shallow ulcers; may coalesce to form large bullous lesions that ulcerate; mucosa may have generalized erythema	Lesions heal spontaneously within 2 weeks
Herpes zoster (VZV reactivation)	Cheek, tongue, gingiva, or palate	Unilateral vesicular eruptions and ulceration in linear pattern following sensory distribution of trigeminal nerve or one of its branches	Gradual healing without scarring unless secondarily infected; postherpetic neuralgia is common; oral acyclovir, famciclovir, or valacyclovir reduces healing time and postherpetic neuralgia
Infectious mononucleosis (Epstein-Barr virus)	Oral mucosa	Fatigue, sore throat, malaise, fever, and cervical lymphadenopathy; numerous small ulcers usually appear several days before lymphadenopathy; gingival bleeding and multiple petechiae at junction of hard and soft palates	Oral lesions disappear during convalescence; no treatment is given, though glucocorticoids are indicated if tonsillar swelling compromises the airway
Herpangina (coxsackievirus A; also possibly coxsackievirus B and echovirus)	Oral mucosa, pharynx, tongue	Sudden onset of fever, sore throat, and oropharyngeal vesicles, usually in children <4 years old, during summer months; diffuse pharyngeal congestion and vesicles (1–2 mm), grayish-white surrounded by red areola; vesicles enlarge and ulcerate	Incubation period of 2–9 days; fever for 1–4 days; recovery uneventful
Hand-foot-and-mouth disease (most commonly coxsackievirus A16)	Oral mucosa, pharynx, palms, and soles	Fever, malaise, headache with oropharyngeal vesicles that become painful, shallow ulcers; highly infectious; usually affects children under age 10	Incubation period 2–18 days; lesions heal spontaneously in 2–4 weeks
Primary HIV infection	Gingiva, palate, and pharynx	Acute gingivitis and oropharyngeal ulceration, associated with febrile illness resembling mononucleosis and including lymphadenopathy	Followed by HIV seroconversion, asymptomatic HIV infection, and usually ultimately by HIV disease
Bacterial or Fungal Diseases			
Acute necrotizing ulcerative gingivitis (“trench mouth”)	Gingiva	Painful, bleeding gingiva characterized by necrosis and ulceration of gingival papillae and margins plus lymphadenopathy and foul breath	Debridement and diluted (1:3) peroxide lavage provide relief within 24 h; antibiotics in acutely ill patients; relapse may occur
Prenatal (congenital) syphilis	Palate, jaws, tongue, and teeth	Gummatous involvement of palate, jaws, and facial bones; Hutchinson’s incisors, mulberry molars, glossitis, mucous patches, and fissures at corner of mouth	Tooth deformities in permanent dentition irreversible
Primary syphilis (chancre)	Lesion appearing where organism enters body; may occur on lips, tongue, or tonsillar area	Small papule developing rapidly into a large, painless ulcer with indurated border; unilateral lymphadenopathy; chancre and lymph nodes containing spirochetes; serologic tests positive by third to fourth weeks	Healing of chancre in 1–2 months, followed by secondary syphilis in 6–8 weeks
Secondary syphilis	Oral mucosa frequently involved with mucous patches, which occur primarily on palate and also at commissures of mouth	Maculopapular lesions of oral mucosa, 5–10 mm in diameter with central ulceration covered by grayish membrane; eruptions occurring on various mucosal surfaces and skin, accompanied by fever, malaise, and sore throat	Lesions may persist from several weeks to a year
Tertiary syphilis	Palate and tongue	Gummatous infiltration of palate or tongue followed by ulceration and fibrosis; atrophy of tongue papillae produces characteristic bald tongue and glossitis	Gumma may destroy palate, causing complete perforation
Gonorrhea	Lesions may occur in mouth at site of inoculation or secondarily by hematogenous spread from a primary focus	Most pharyngeal infection is asymptomatic; may produce burning or itching sensation; oropharynx and tonsils may be ulcerated and erythematous; saliva viscous and fetid	More difficult to eradicate than urogenital infection, though pharyngitis usually resolves with appropriate antimicrobial treatment
Tuberculosis	Tongue, tonsillar area, soft palate	Painless, solitary, 1- to 5-cm, irregular ulcer covered with persistent exudate; ulcer has firm undermined border	Autoinoculation from pulmonary infection is usual; lesions resolve with appropriate antimicrobial therapy
Cervicofacial actinomycosis	Swellings in region of face, neck, and floor of mouth	Infection may be associated with extraction, jaw fracture, or eruption of molar tooth; in acute form, resembles acute pyogenic abscess, but contains yellow “sulfur granules” (gram-positive mycelia and their hyphae)	Typically, swelling is hard and grows painlessly; multiple abscesses with draining tracts develop; penicillin first choice; surgery usually necessary

(*Continued*)

TABLE 38-1 Vesicular, Bullous, or Ulcerative Lesions of the Oral Mucosa (Continued)

CONDITION	USUAL LOCATION	CLINICAL FEATURES	COURSE
Bacterial or Fungal Diseases (Continued)			
Histoplasmosis	Any area of the mouth, particularly tongue, gingiva, or palate	Nodular, verrucous, or granulomatous lesions; ulcers are indurated and painful; usual source hematogenous or pulmonary, but may be primary	Systemic antifungal therapy necessary
Candidiasis[a]			
Dermatologic Diseases			
Mucous membrane pemphigoid	Typically produces marked gingival erythema and ulceration; other areas of oral cavity, esophagus, and vagina may be affected	Painful, grayish-white collapsed vesicles or bullae of full-thickness epithelium with peripheral erythematous zone; gingival lesions desquamate, leaving ulcerated area	Protracted course with remissions and exacerbations; involvement of different sites develops slowly; glucocorticoids may temporarily reduce symptoms but do not control disease
EM minor and EM major (Stevens-Johnson syndrome)	Primarily oral mucosa and skin of hands and feet	Intraoral ruptured bullae surrounded by inflammatory area; lips may show hemorrhagic crusts; “iris” or “target” lesion on skin is pathognomonic; patient may have severe signs of toxicity	Onset very rapid; usually idiopathic, but may be associated with trigger such as drug reaction; condition may last 3–6 weeks; mortality rate for untreated EM major is 5–15%
Pemphigus vulgaris	Oral mucosa and skin; sites of mechanical trauma (soft/hard palate, frenulum, lips, buccal mucosa)	Usually (>70%) presents with oral lesions; fragile, ruptured bullae and ulcerated oral areas; mostly in older adults	With repeated occurrence of bullae, toxicity may lead to cachexia, infection, and death within 2 years; often controllable with oral glucocorticoids
Lichen planus	Oral mucosa and skin	White striae in mouth; purplish nodules on skin at sites of friction; occasionally causes oral mucosal ulcers and erosive gingivitis	White striae alone usually asymptomatic; erosive lesions often difficult to treat, but may respond to glucocorticoids
Other Conditions			
Recurrent aphthous ulcers	Usually on nonkeratinized oral mucosa (buccal and labial mucosa, floor of mouth, soft palate, lateral and ventral tongue)	Single or clustered painful ulcers with surrounding erythematous border; lesions may be 1–2 mm in diameter in crops (herpetiform), 1–5 mm (minor), or 5–15 mm (major)	Lesions heal in 1–2 weeks but may recur monthly or several times a year; protective barrier with benzocaine and topical glucocorticoids relieve symptoms; systemic glucocorticoids may be needed in severe cases
Behçet’s syndrome	Oral mucosa, eyes, genitalia, gut, and CNS	Multiple aphthous ulcers in mouth; inflammatory ocular changes, ulcerative lesions on genitalia; inflammatory bowel disease and CNS disease	Oral lesions often first manifestation; persist several weeks and heal without scarring
Traumatic ulcers	Anywhere on oral mucosa; dentures frequently responsible for ulcers in vestibule	Localized, discrete ulcerated lesions with red border; produced by accidental biting of mucosa, penetration by foreign object, or chronic irritation by dentures	Lesions usually heal in 7–10 days when irritant is removed, unless secondarily infected
Squamous cell carcinoma	Any area of mouth, most commonly on lower lip, lateral borders of tongue, and floor of mouth	Red, white, or red and white ulcer with elevated or indurated border; failure to heal; pain not prominent in early lesions	Invades and destroys underlying tissues; frequently metastasizes to regional lymph nodes
Acute myeloid leukemia (usually monocytic)	Gingiva	Gingival swelling and superficial ulceration followed by hyperplasia of gingiva with extensive necrosis and hemorrhage; deep ulcers may occur elsewhere on mucosa, complicated by secondary infection	Usually responds to systemic treatment of leukemia; occasionally requires local irradiation
Lymphoma	Gingiva, tongue, palate, and tonsillar area	Elevated, ulcerated area that may proliferate rapidly, giving appearance of traumatic inflammation	Fatal if untreated; may indicate underlying HIV infection
Chemical or thermal burns	Any area in mouth	White slough due to contact with corrosive agents (e.g., aspirin, hot cheese) applied locally; removal of slough leaves raw, painful surface	Lesion heals in several weeks if not secondarily infected

[a] See Table 38-3.

Abbreviations: CNS, central nervous system; EM, erythema multiforme; HSV, herpes simplex virus; VZV, varicella-zoster virus.

chancre is typically painless and therefore easily missed. Regional lymphadenopathy is invariably present.

Disorders of mucosal fragility often produce painful oral ulcers that fail to heal within 2 weeks. *Mucous membrane pemphigoid* and *pemphigus vulgaris* are the major acquired disorders. While their clinical features are often distinctive, a biopsy or immunohistochemical examination should be performed to diagnose these entities and to distinguish them from *lichen planus* and drug reactions.

Hematologic and Nutritional Disease Internists are more likely to encounter patients with acquired, rather than congenital, bleeding disorders. Bleeding should stop 15 min after minor trauma and within an hour after tooth extraction if local pressure is applied. More prolonged bleeding, if not due to continued injury or rupture of a large vessel, should lead to investigation for a clotting abnormality. In addition to bleeding, petechiae and ecchymoses are prone to occur at the vibrating line between the soft and hard palates in patients with platelet dysfunction or thrombocytopenia.

All forms of leukemia, but particularly *acute myelomonocytic leukemia*, can produce gingival bleeding, ulcers, and gingival enlargement. Oral ulcers are a feature of agranulocytosis, and ulcers and mucositis are often severe complications of chemotherapy and radiation therapy

TABLE 38-2 Pigmented Lesions of the Oral Mucosa

CONDITION	USUAL LOCATION	CLINICAL FEATURES	COURSE
Oral melanotic macule	Any area of mouth	Discrete or diffuse, localized, brown to black macule	Remains indefinitely; no growth
Diffuse melanin pigmentation	Any area of mouth	Diffuse pale to dark-brown pigmentation; may be physiologic ("racial") or due to smoking	Remains indefinitely
Nevi	Any area of mouth	Discrete, localized, brown to black pigmentation	Remains indefinitely
Malignant melanoma	Any area of mouth	Can be flat and diffuse, painless, brown to black; or can be raised and nodular	Expands and invades early; metastasis leads to death
Addison's disease	Any area of mouth, but mostly buccal mucosa	Blotches or spots of bluish-black to dark-brown pigmentation occurring early in disease, accompanied by diffuse pigmentation of skin; other symptoms of adrenal insufficiency	Condition controlled by adrenal steroid replacement
Peutz-Jeghers syndrome	Any area of mouth	Dark-brown spots on lips, buccal mucosa, with characteristic distribution of pigment around lips, nose, and eyes and on hands; concomitant intestinal polyposis	Oral pigmented lesions remain indefinitely; gastrointestinal polyps may become malignant
Drug ingestion (neuroleptics, oral contraceptives, minocycline, zidovudine, quinine derivatives)	Any area of mouth	Brown, black, or gray areas of pigmentation	Gradually disappears following cessation of drug intake
Amalgam tattoo	Gingiva and alveolar mucosa	Small blue-black pigmented areas associated with embedded amalgam particles in soft tissues; may show up on radiographs as radiopaque particles in some cases	Remains indefinitely
Heavy metal pigmentation (bismuth, mercury, lead)	Gingival margin	Thin blue-black pigmented line along gingival margin; rarely seen except in children exposed to lead-based paint	Indicative of systemic absorption; no significance for oral health
Black hairy tongue	Dorsum of tongue	Elongation of filiform papillae of tongue, which become stained by coffee, tea, tobacco, or pigmented bacteria	Improves within 1–2 weeks with gentle brushing of tongue or (if due to bacterial overgrowth) discontinuation of antibiotic
Fordyce spots	Buccal and labial mucosa	Numerous small yellowish spots just beneath mucosal surface; no symptoms; due to hyperplasia of sebaceous glands	Benign; remains without apparent change
Kaposi's sarcoma	Palate most common, but may occur at any other site	Red or blue plaques of variable size and shape; often enlarge, become nodular, and may ulcerate	Usually indicative of HIV infection or non-Hodgkin's lymphoma; rarely fatal, but may require treatment for comfort or cosmesis
Mucous retention cysts	Buccal and labial mucosa	Bluish, clear fluid–filled cyst due to extravasated mucus from injured minor salivary gland	Benign; painless unless traumatized; may be removed surgically

TABLE 38-3 White Lesions of Oral Mucosa

CONDITION	USUAL LOCATION	CLINICAL FEATURES	COURSE
Lichen planus	Buccal mucosa, tongue, gingiva, and lips; skin	Striae, white plaques, red areas, ulcers in mouth; purplish papules on skin; may be asymptomatic, sore, or painful; lichenoid drug reactions may look similar	Protracted; responds to topical glucocorticoids
White sponge nevus	Oral mucosa, vagina, anal mucosa	Painless white thickening of epithelium; adolescence/early adulthood onset; familial	Benign and permanent
Smoker's leukoplakia and smokeless tobacco lesions	Any area of oral mucosa, sometimes related to location of habit	White patch that may become firm, rough, or red-fissured and ulcerated; may become sore and painful but is usually painless	May or may not resolve with cessation of habit; 2% of patients develop squamous cell carcinoma; early biopsy essential
Erythroplakia with or without white patches	Floor of mouth commonly affected in men; tongue and buccal mucosa in women	Velvety, reddish plaque; occasionally mixed with white patches or smooth red areas	High risk of squamous cell cancer; early biopsy essential
Candidiasis	Any area in mouth	*Pseudomembranous type* ("thrush"): creamy white curdlike patches that reveal a raw, bleeding surface when scraped; found in sick infants, debilitated elderly patients receiving high-dose glucocorticoids or broad-spectrum antibiotics, and patients with AIDS	Responds favorably to antifungal therapy and correction of predisposing causes where possible
		Erythematous type: flat, red, sometimes sore areas in same groups of patients	Course same as for pseudomembranous type
		Candidal leukoplakia: nonremovable white thickening of epithelium due to *Candida*	Responds to prolonged antifungal therapy
		Angular cheilitis: sore fissures at corner of mouth	Responds to topical antifungal therapy
Hairy leukoplakia	Usually on lateral tongue, rarely elsewhere on oral mucosa	White areas ranging from small and flat to extensive accentuation of vertical folds; found in HIV carriers (all risk groups for AIDS)	Due to Epstein-Barr virus; responds to high-dose acyclovir but recurs; rarely causes discomfort unless secondarily infected with *Candida*
Warts (human papillomavirus)	Anywhere on skin and oral mucosa	Single or multiple papillary lesions with thick, white, keratinized surfaces containing many pointed projections; cauliflower lesions covered with normal-colored mucosa or multiple pink or pale bumps (focal epithelial hyperplasia)	Lesions grow rapidly and spread; squamous cell carcinoma must be ruled out with biopsy; excision or laser therapy; may regress in HIV-infected patients receiving antiretroviral therapy

TABLE 38-4 Alterations of the Tongue

TYPE OF CHANGE	CLINICAL FEATURES
Size or Morphology	
Macroglossia	Enlarged tongue that may be part of a syndrome found in developmental conditions such as Down syndrome, Simpson-Golabi-Behmel syndrome, or Beckwith-Wiedemann syndrome; may be due to tumor (hemangioma or lymphangioma), metabolic disease (e.g., primary amyloidosis), or endocrine disturbance (e.g., acromegaly or cretinism); may occur when all teeth are removed
Fissured ("scrotal") tongue	Dorsal surface and sides of tongue covered by painless shallow or deep fissures that may collect debris and become irritated
Median rhomboid glossitis	Congenital abnormality with ovoid, denuded area in median posterior portion of tongue; may be associated with candidiasis and may respond to antifungal treatment
Color	
"Geographic" tongue (benign migratory glossitis)	Asymptomatic inflammatory condition of tongue, with rapid loss and regrowth of filiform papillae leading to appearance of denuded red patches "wandering" across surface
Hairy tongue	Elongation of filiform papillae of medial dorsal surface area due to failure of keratin layer of papillae to desquamate normally; brownish-black coloration may be due to staining by tobacco, food, or chromogenic organisms
"Strawberry" and "raspberry" tongue	Appearance of tongue during scarlet fever due to hypertrophy of fungiform papillae as well as changes in filiform papillae
"Bald" tongue	Atrophy may be associated with xerostomia, pernicious anemia, iron-deficiency anemia, pellagra, or syphilis; may be accompanied by painful burning sensation; may be an expression of erythematous candidiasis and respond to antifungal treatment

TABLE 38-5 Oral Lesions Associated with HIV Infection

LESION MORPHOLOGY	ETIOLOGIES
Papules, nodules, plaques	Candidiasis (hyperplastic and pseudomembranous)[a] Condyloma acuminatum (human papillomavirus infection) Squamous cell carcinoma (preinvasive and invasive) Non-Hodgkin's lymphoma[a] Hairy leukoplakia[a]
Ulcers	Recurrent aphthous ulcers[a] Angular cheilitis Squamous cell carcinoma Acute necrotizing ulcerative gingivitis[a] Necrotizing ulcerative periodontitis[a] Necrotizing ulcerative stomatitis Non-Hodgkin's lymphoma[a] Viral infection (herpes simplex, herpes zoster, cytomegalovirus infection) Infection caused by *Mycobacterium tuberculosis* or *Mycobacterium avium-intracellulare* Fungal infection (histoplasmosis, cryptococcosis, candidiasis, geotrichosis, aspergillosis) Bacterial infection (*Escherichia coli, Enterobacter cloacae, Klebsiella pneumoniae, Pseudomonas aeruginosa*) Drug reactions (single or multiple ulcers)
Pigmented lesions	Kaposi's sarcoma[a] Bacillary angiomatosis (skin and visceral lesions more common than oral) Zidovudine pigmentation (skin, nails, and occasionally oral mucosa) Addison's disease
Miscellaneous	Linear gingival erythema[a]

[a] Strongly associated with HIV infection.

for hematologic and other malignancies. *Plummer-Vinson syndrome* (iron deficiency, angular stomatitis, glossitis, and dysphagia) raises the risk of oral squamous cell cancer and esophageal cancer at the postcricoidal tissue web. Atrophic papillae and a red, burning tongue may occur with pernicious anemia. Deficiencies in B-group vitamins produce many of these same symptoms, as well as oral ulceration and cheilosis. Consequences of *scurvy* include swollen, bleeding gums; ulcers; and loosening of the teeth.

NONDENTAL CAUSES OF ORAL PAIN

Most, but not all, oral pain emanates from inflamed or injured tooth pulp or periodontal tissues. Nonodontogenic causes are often overlooked. In most instances, toothache is predictable and proportional to the stimulus applied, and an identifiable condition (e.g., caries, abscess) is found. Local anesthesia eliminates pain originating from dental or periodontal structures, but not referred pains. The most common nondental source of pain is myofascial pain referred from muscles of mastication, which become tender and ache with increased use. Many sufferers exhibit *bruxism* (grinding of the teeth) secondary to stress and anxiety. *Temporomandibular joint disorder* is closely related. Features include pain, limited mandibular movement, and temporomandibular joint sounds. The etiologies are complex; malocclusion does not play the primary role once attributed to it. *Osteoarthritis* is a common cause of masticatory pain. Anti-inflammatory medication, jaw rest, soft foods, and heat provide relief. The temporomandibular joint is involved in 50% of patients with *rheumatoid arthritis* and is usually a late feature of severe disease.

Migrainous neuralgia may be localized to the mouth. Episodes of pain and remission without an identifiable cause and a lack of relief with local anesthesia are important clues. *Trigeminal neuralgia* (*tic douloureux*) can involve the entire branch or part of the mandibular or maxillary branch of the fifth cranial nerve and can produce pain in one or a few teeth. Pain may occur spontaneously or may be triggered by touching the lip or gingiva, brushing the teeth, or chewing. *Glossopharyngeal neuralgia* produces similar acute neuropathic symptoms in the distribution of the ninth cranial nerve. Swallowing, sneezing, coughing, or pressure on the tragus of the ear triggers pain that is felt in the base of the tongue, pharynx, and soft palate and may be referred to the temporomandibular joint. *Neuritis* involving the maxillary and mandibular divisions of the trigeminal nerve (e.g., maxillary sinusitis, neuroma, and leukemic infiltrate) is distinguished from ordinary toothache by the neuropathic quality of the pain. Occasionally, *phantom pain* follows tooth extraction. Pain and hyperalgesia behind the ear and on the side of the face in the day or so before facial weakness develops often constitute the earliest symptom of *Bell's palsy*. Likewise, similar symptoms may precede visible lesions of herpes zoster infecting the seventh nerve (*Ramsy-Hunt syndrome*) or trigeminal nerve. *Postherpetic neuralgia* may follow either condition. *Coronary ischemia* may produce pain exclusively in the face and jaw; as in typical angina pectoris, this pain is usually reproducible with increased myocardial demand. Aching in several upper molar or premolar teeth that is unrelieved by anesthetizing the teeth may point to *maxillary sinusitis*.

Giant cell arteritis is notorious for producing headache, but it may also produce facial pain or sore throat without headache. Jaw and tongue claudication with chewing or talking is relatively common. Tongue infarction is rare. Patients with subacute thyroiditis often experience pain referred to the face or jaw before the tenderness of the thyroid gland and transient hyperthyroidism are appreciated.

"Burning mouth syndrome" (*glossodynia*) occurs in the absence of an identifiable cause (e.g., vitamin B_{12} deficiency, iron deficiency, diabetes mellitus, low-grade *Candida* infection, food sensitivity, or subtle xerostomia). The etiology may be neuropathic. Clonazepam, α-lipoic acid, and cognitive-behavioral therapy benefit some patients. Some cases associated with an angiotensin-converting enzyme inhibitor have remitted when the drug was discontinued.

DISEASES OF THE SALIVARY GLANDS

Saliva is essential to oral health. Its absence leads to dental caries, periodontal disease, and difficulties in wearing dental prostheses, masticating, and speaking. Its major components, water and mucin, serve as a cleansing solvent and lubricating fluid. In addition, saliva contains antimicrobial factors (e.g., lysozyme, lactoperoxidase, secretory IgA), epidermal growth factor, minerals, and buffering systems. The major salivary glands secrete intermittently in response to autonomic stimulation, which is high during a meal. Hundreds of minor glands in the lips and cheeks secrete mucus continuously throughout the day and night. Consequently, oral function becomes impaired when salivary function is reduced. The sensation of a dry mouth (*xerostomia*) is perceived when salivary flow is reduced by 50%. The most common etiology is medication, especially drugs with anticholinergic properties but also alpha and beta blockers, calcium channel blockers, and diuretics. Other causes include Sjögren's syndrome, chronic parotitis, salivary duct obstruction, diabetes mellitus, HIV/AIDS, and radiation therapy that includes the salivary glands in the field. Management involves eliminating or limiting drying medications, preventive dental care, and supplementation with oral liquid or salivary substitutes. Sugarless mints or chewing gum may stimulate salivary secretion if dysfunction is mild. When sufficient exocrine tissue remains, pilocarpine or cevimeline can increase secretions. Commercial saliva substitutes or gels relieve dryness. Fluoride supplementation is critical to prevent caries.

Sialolithiasis presents most often as painful swelling but in some instances as only swelling or only pain. Conservative therapy consists of local heat, massage, and hydration. Promotion of salivary secretion with mints or lemon drops may flush out small stones. Antibiotic treatment is necessary when bacterial infection is suspected. In adults, *acute bacterial parotitis* is typically unilateral and most commonly affects postoperative, dehydrated, and debilitated patients. *Staphylococcus aureus* (including methicillin-resistant strains) and anaerobic bacteria are the most common pathogens. Chronic bacterial *sialadenitis* results from lowered salivary secretion and recurrent bacterial infection. When suspected bacterial infection is not responsive to therapy, the differential diagnosis should expand to include benign and malignant neoplasms, lymphoproliferative disorders, Sjögren's syndrome, sarcoidosis, tuberculosis, lymphadenitis, actinomycosis, and granulomatosis with polyangiitis. Bilateral nontender parotid enlargement occurs with diabetes mellitus, cirrhosis, bulimia, HIV/AIDS, and drugs (e.g., iodide, propylthiouracil).

Pleomorphic adenoma composes two-thirds of all salivary neoplasms. The parotid is the principal salivary gland affected, and the tumor presents as a firm, slow-growing mass. Although this tumor is benign, its recurrence is common if resection is incomplete. Malignant tumors such as mucoepidermoid carcinoma, adenoid cystic carcinoma, and adenocarcinoma tend to grow relatively fast, depending upon grade. They may ulcerate and invade nerves, producing numbness and facial paralysis. Surgical resection is the primary treatment. Radiation therapy (particularly neutron-beam therapy) is used when surgery is not feasible and after resection for certain histologic types with a high risk of recurrence. Malignant salivary gland tumors have a 5-year survival rate of 94% when the stage is local, 70% with regional spread, and 43% when distant.

Dental Care for Medically Complex Patients Routine dental care (e.g., uncomplicated extraction, scaling and cleaning, tooth restoration, and root canal) is remarkably safe. The most common concerns regarding care of dental patients with medical disease are excessive bleeding for patients taking anticoagulants, infection of the heart valves and prosthetic devices from hematogenous seeding by the oral flora, and cardiovascular complications resulting from vasopressors used with local anesthetics during dental treatment, although the risk of any of these complications is very low.

Patients undergoing tooth extraction or alveolar and gingival surgery rarely experience uncontrolled bleeding when warfarin anticoagulation is maintained within the therapeutic range currently recommended for prevention of venous thrombosis, atrial fibrillation, or mechanical heart valve. Embolic complications and death, however, have been reported during subtherapeutic anticoagulation. Therapeutic anticoagulation should be confirmed before and continued through the procedure. Likewise, low-dose aspirin (e.g., 81–325 mg) can safely be continued. For patients taking aspirin and another antiplatelet medication (e.g., clopidogrel), continuation of the second antiplatelet medication should be based on individual consideration of the risks of thrombosis and bleeding. Target-specific oral anticoagulants (dabigatran, apixaban, rivaroxaban, and edoxaban) are in increasingly common use. Simple extractions of one to three teeth, periodontal surgery, abscess drainage, and implant positioning do not typically require interruption of therapy. More extensive surgery may necessitate delaying or holding a dose of the anticoagulant or more elaborate measures to manage the risk of thrombosis and bleeding.

Patients at risk for bacterial endocarditis **(Chap. 133)** should maintain optimal oral hygiene, including flossing, and regular professional cleanings. Currently, guidelines recommend that prophylactic antibiotics be restricted to patients at high risk for bacterial endocarditis who undergo dental and oral procedures involving significant manipulation of gingival or periapical tissue or penetration of the oral mucosa. If unexpected bleeding occurs, antibiotics given within 2 h after the procedure provide effective prophylaxis.

Hematogenous bacterial seeding from oral infection can produce late prosthetic-joint infection and therefore requires removal of the infected tissue (e.g., drainage, extraction, root canal) and appropriate antibiotic therapy. However, evidence that late prosthetic-joint infection follows routine dental procedures is lacking. For this reason, antibiotic prophylaxis is generally not recommended before oral surgery or oral mucosal manipulation for patients who have undergone joint replacement surgery. Exceptions to this may be considered for patients who have experienced joint replacement complications.

Concern often arises regarding the use of vasoconstrictors to treat patients with hypertension and heart disease. Vasoconstrictors enhance the depth and duration of local anesthesia, thus reducing the anesthetic dose and potential toxicity. If intravascular injection is avoided, 2% lidocaine with 1:100,000 epinephrine (limited to a total of 0.036 mg of epinephrine) can be used safely in patients with controlled hypertension and stable coronary heart disease, arrhythmia, or congestive heart failure. Precautions should be taken with patients taking tricyclic antidepressants and nonselective beta blockers because these drugs may potentiate the effect of epinephrine.

Elective dental treatments should be postponed for at least 1 month and preferably for 6 months after myocardial infarction, after which the risk of reinfarction is low provided the patient is medically stable (e.g., stable rhythm, stable angina, and no heart failure). Patients who have suffered a stroke should have elective dental care deferred for 9 months. In both situations, effective stress reduction requires good pain control, including the use of the minimal amount of vasoconstrictor necessary to provide good hemostasis and local anesthesia.

Bisphosphonate therapy is associated with *osteonecrosis* of the jaw. However, the risk with oral bisphosphonate therapy is very low. Most patients affected have received high-dose aminobisphosphonate therapy for multiple myeloma or metastatic breast cancer and have undergone tooth extraction or dental surgery. Intraoral lesions, of which two-thirds are painful, appear as exposed yellow-white hard bone involving the mandible or maxilla. Screening tests for determining risk of osteonecrosis are unreliable. Patients slated for aminobisphosphonate therapy should receive preventive dental care that reduces the risk of infection and the need for future dentoalveolar surgery.

Halitosis Halitosis typically emanates from the oral cavity or nasal passages. Bacterial decay of food and cellular debris account for the malodor. Periodontal disease, caries, acute forms of gingivitis, poorly fitting dentures, oral abscess, and tongue coating are common causes. Treatment includes correcting poor hygiene, treating infection, and tongue brushing. Hyposalivation can produce and exacerbate halitosis. Pockets of decay in the tonsillar crypts, esophageal diverticulum, esophageal stasis (e.g., achalasia, stricture), sinusitis, and lung abscess account for some instances. A few systemic diseases produce distinctive odors: renal failure and *Helicobacter pylori* gastritis (ammoniacal), hepatic (fishy), and ketoacidosis (fruity). If a patient presents because of concern about halitosis but no odor is detectable, then pseudohalitosis or halitophobia must be considered.

Aging and Oral Health While tooth loss and dental disease are not normal consequences of aging, a complex array of structural and functional changes that occur with age can affect oral health. Subtle changes in tooth structure (e.g., diminished pulp space and volume, sclerosis of dentinal tubules, and altered proportions of nerve and vascular pulp content) result in the elimination or diminution of pain sensitivity and a reduction in the reparative capacity of the teeth. In addition, age-associated fatty replacement of salivary acini may reduce physiologic reserve, thus increasing the risk of hyposalivation. In healthy older adults, there is minimal, if any, reduction in salivary flow.

Poor oral hygiene often results when general health fails or when patients lose manual dexterity and upper-extremity flexibility. This situation is particularly common among frail older adults and nursing home residents, and regular oral cleaning and dental care reduce the incidence of pneumonia, oral disease, the mortality risk in this population. Other risks for dental decay include limited lifetime fluoride exposure. Without assiduous care, decay can become quite advanced yet remain asymptomatic. Consequently, much of a tooth—or the entire tooth—can be destroyed before the patient is aware of the process.

Periodontal disease, a leading cause of tooth loss, is indicated by loss of alveolar bone height. More than 90% of the U.S. population has some degree of periodontal disease by age 50. Healthy adults who have not had significant alveolar bone loss by the sixth decade of life do not typically experience significant worsening with advancing age.

Complete edentulousness in the United States is becoming less common in the elderly and is increasingly restricted to impoverished populations. When it is present, speech, mastication, and facial contours are dramatically affected. Edentulousness may also exacerbate obstructive sleep apnea, particularly in asymptomatic individuals who wear dentures. Dentures can improve verbal articulation and restore diminished facial contours. Mastication can also be restored; however, patients expecting dentures to facilitate oral intake are often disappointed. Accommodation to dentures requires a period of adjustment. Pain can result from friction or traumatic lesions produced by loose dentures. Poor fit and poor oral hygiene may permit the development of candidiasis. This fungal infection may be either asymptomatic or painful and is suggested by erythematous smooth or granular tissue conforming to an area covered by the appliance. Individuals with dentures and no natural teeth need regular (annual) professional oral examinations.

FURTHER READING

Chavez EM et al: Dental care for geriatric and special needs populations. Dent Clin N Am 62:245, 2018.

Durso SC: Interaction with other health team members in caring for elderly patients. Dent Clin North Am 49:377, 2005.

Kaplovitch E, Dounaevskaia V: Treatment in the dental practice of the patient receiving anticoagulant therapy. J Am Dent Assoc 150:602, 2019.

Section 5 Alterations in Circulatory and Respiratory Functions

39 Dyspnea

Rebecca M. Baron

DYSPNEA

DEFINITION

The American Thoracic Society consensus statement defines *dyspnea* as a "subjective experience of breathing discomfort that consists of qualitatively distinct sensations that vary in intensity. The experience derives from interactions among multiple physiological, psychological, social, and environmental factors and may induce secondary physiological and behavioral responses." Dyspnea, a symptom, can be perceived only by the person experiencing it and, therefore, must be self-reported. In contrast, signs of increased work of breathing, such as tachypnea, accessory muscle use, and intercostal retraction, can be measured and reported by clinicians.

EPIDEMIOLOGY

Dyspnea is common. It has been reported that up to one-half of inpatients and one-quarter of ambulatory patients experience dyspnea, with a prevalence of 9–13% in the community that increases to as high as 37% for adults aged ≥70 years. Dyspnea is a frequent cause of emergency room visits, accounting for as many as 3–4 million visits per year. Furthermore, it is increasingly appreciated that the degree of dyspnea may better predict outcomes in chronic obstructive pulmonary disease (COPD) than does the forced expiratory volume in 1 s (FEV_1), and formal measures of dyspnea have been incorporated into the Global Initiative for Chronic Obstructive Lung Disease (GOLD) COPD severity assessment guidelines. Dyspnea may also predict outcomes in other chronic heart and lung diseases as well. Dyspnea can arise from a diverse array of pulmonary, cardiac, and neurologic underlying causes, and elucidation of particular symptoms may point toward a specific etiology and/or mechanism driving dyspnea (although additional diagnostic testing is often required, as will be further discussed below). There has been an increased focus on dyspnea given its increased incidence in the setting of the SARS-CoV-2 pandemic, and post-COVID persistent symptoms in many patients **(Chap. 205)**.

MECHANISMS UNDERLYING DYSPNEA

The mechanisms underlying dyspnea are complex, as it can arise from different contributory respiratory sensations. Although a large body of research has increased our understanding of mechanisms underlying particular respiratory sensations such as "chest tightness" or "air hunger," it is likely that a given disease state might produce the sensation of dyspnea via more than one underlying mechanism. Dyspnea can arise from a variety of pathways, including generation of *afferent* signals from the respiratory system to the central nervous system (CNS), *efferent* signals from the CNS to the respiratory muscles, and particularly when there is a mismatch in the integrative signaling between these two pathways, termed *efferent-reafferent mismatch* **(Fig. 39-1)**.

Afferent signals trigger the CNS (brainstem and/or cortex) and include primarily: (1) peripheral chemoreceptors in the carotid body and aortic arch and central chemoreceptors in the medulla that are activated by hypoxemia, hypercapnia, or acidemia, and might produce a sense of "air hunger"; and (2) mechanoreceptors in the upper airways, lungs (including stretch receptors, irritant receptors, and J receptors), and chest wall (including muscle spindles as stretch receptors and tendon organs that monitor force generation) that are activated in the setting of an increased workload from a disease state producing an increase in airway resistance that may be associated with symptoms of chest tightness (e.g., asthma or COPD) or decreased lung or chest wall

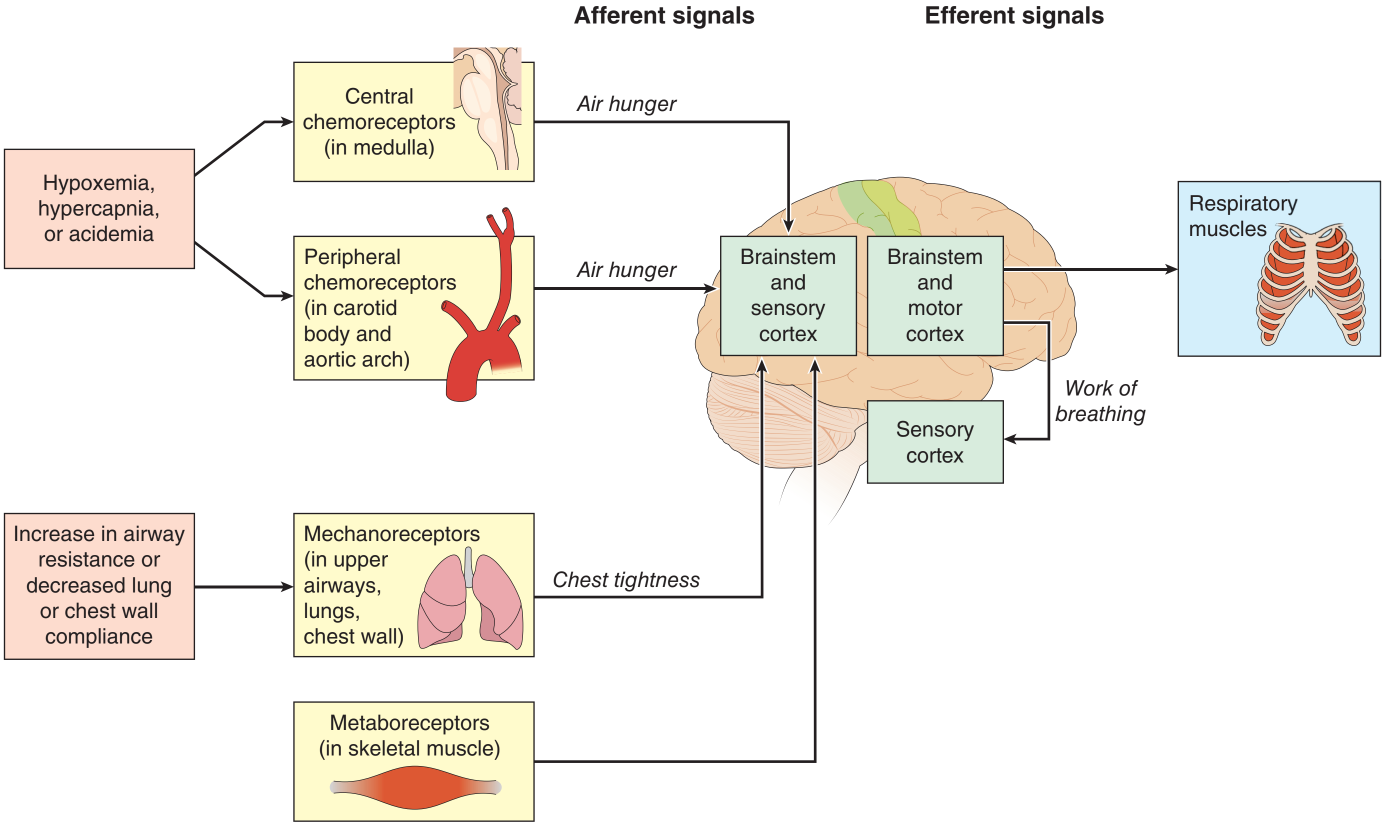
Afferent signals
Efferent signals
Hypoxemia, hypercapnia, or acidemia
Central chemoreceptors (in medulla)
Peripheral chemoreceptors (in carotid body and aortic arch)
Air hunger
Air hunger
Brainstem and sensory cortex
Brainstem and motor cortex
Respiratory muscles
Work of breathing
Sensory cortex
Increase in airway resistance or decreased lung or chest wall compliance
Mechanoreceptors (in upper airways, lungs, chest wall)
Chest tightness
Metaboreceptors (in skeletal muscle)

FIGURE 39-1 Potential signaling pathways underlying the sensation of dyspnea. Dyspnea is a complex sensation that can arise from a variety of physiologic stimuli (examples in pink boxes include hypoxemia, hypercapnia, acidemia, increase in airway resistance, and decrease in lung and/or chest wall compliance). Physiologic signals are mediated via relevant receptors (examples in yellow boxes include central chemoreceptors in the medulla; peripheral chemoreceptors in the carotid body and aortic arch; mechanoreceptors in upper airways, lungs, and chest wall; and metaboreceptors in skeletal muscle) that send afferent signals to the brainstem and sensory cortex. Efferent signals are sent from the brainstem and motor cortex to the respiratory muscles and sensory cortex. Potential symptoms associated with dyspnea that may be generated by these signals are indicated in italics.

compliance (e.g., pulmonary fibrosis). Other afferent signals that trigger dyspnea within the respiratory system can arise from pulmonary vascular receptor responses to changes in pulmonary artery pressure and skeletal muscle (termed metaboreceptors) that are believed to sense changes in the biochemical environment.

Efferent signals are sent from the CNS (motor cortex and brainstem) to the respiratory muscles and are also transmitted by corollary discharge to the sensory cortex; they are believed to underlie sensations of respiratory effort (or "work of breathing") and perhaps contribute to sensations of "air hunger," especially in response to an increased ventilatory load in a disease state such as COPD. In addition, fear or anxiety may heighten the sense of dyspnea by exacerbating the underlying physiologic disturbance in response to an increased respiratory rate or disordered breathing pattern.

■ ASSESSING DYSPNEA

While it is well appreciated that dyspnea is a difficult quality to measure reliably owing to multiple relevant possible domains that can be measured (e.g., sensory-perceptual experience, affective distress, and symptom impact or burden) and that there are no uniformly agreed upon tools for dyspnea assessment, consensus opinion holds that dyspnea should be formally assessed in a context most relevant and beneficial for patient management and, furthermore, that the specific domains being measured are adequately described. There are a number of emerging tools that have been developed for formal dyspnea assessment. As an example, the GOLD criteria advocate use of a dyspnea assessment tool such as the Modified Medical Research Council Dyspnea Scale **(Table 39-1)** to assess symptom/impact burden in COPD.

■ DIFFERENTIAL DIAGNOSIS

This chapter focuses largely on chronic dyspnea, which is defined as symptoms lasting longer than 1 month and can arise from a broad array of different underlying conditions, most commonly attributable to pulmonary or cardiac conditions that account for as many as 85% of the underlying causes of dyspnea. However, as many as one-third of patients may have multifactorial reasons underlying dyspnea, with an increasing number of individuals suffering from dyspnea as part of a post-COVID syndrome **(Chap. 205)**. Examples of a wide array of conditions that underlie dyspnea with possible mechanisms underlying the presenting symptoms are described in **Table 39-2**.

Respiratory system causes include diseases of the airways (e.g., asthma and COPD), diseases of the parenchyma (more commonly, interstitial lung diseases are seen in the setting of chronic dyspnea,

TABLE 39-1 An Example of a Clinical Method for Rating Dyspnea: The Modified Medical Research Council Dyspnea Scale[a]

GRADE OF DYSPNEA	DESCRIPTION
0	Not troubled by breathlessness, except with strenuous exercise
1	Shortness of breath walking on level ground or with walking up a slight hill
2	Walks slower than people of similar age on level ground due to breathlessness, or has to stop to rest when walking at own pace on level ground
3	Stops to rest after walking 100 m or after walking a few minutes on level ground
4	Too breathless to leave the house, or breathless with activities of daily living (e.g., dressing/undressing)

[a]This scale has been integrated into the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines clinical classification scheme.

Source: Reproduced with permission from DA Mahler, CK Wells: Evaluation of clinical methods for rating dyspnea. Chest 93:580, 1988.

TABLE 39-2 Differential Diagnosis of Disease Processes Underlying Dyspnea

SYSTEM	TYPE OF PROCESS	EXAMPLE OF DISEASE PROCESS	POSSIBLE PRESENTING DYSPNEA SYMPTOMS	POSSIBLE PHYSICAL FINDINGS	POSSIBLE MECHANISMS UNDERLYING DYSPNEA	INITIAL DIAGNOSTIC STUDIES (AND POSSIBLE FINDINGS)
Pulmonary	Airways disease	Asthma, COPD, upper airway obstruction	Chest tightness, tachypnea, increased WOB, air hunger, inability to get a deep breath	Wheezing, accessory muscle use, exertional hypoxemia (especially with COPD)	Increased WOB, hypoxemia, hypercapnia, stimulation of pulmonary receptors	Peak flow (reduced); spirometry (OVD); CXR (hyperinflation; loss of lung parenchyma in COPD), chest CT and airway examination for upper airway obstruction
	Parenchymal disease	Interstitial lung disease[a]	Air hunger, inability to get a deep breath	Dry end-inspiratory crackles, clubbing, exertional hypoxemia	Increased WOB, increased respiratory drive, hypoxemia, hypercapnia, stimulation of pulmonary receptors	Spirometry and lung volumes (RVD); CXR and chest CT (interstitial lung disease)
	Chest wall disease	Kyphoscoliosis, neuromuscular (NM) weakness	Increased WOB, inability to get a deep breath	Decreased diaphragm excursion; atelectasis	Increased WOB; stimulation of pulmonary receptors (if atelectasis is present)	Spirometry and lung volumes (RVD); MIP and MEPs (reduced in NM weakness)
Pulmonary and cardiac	Pulmonary vasculature	Pulmonary hypertension	Tachypnea	Elevated right heart pressures, exertional hypoxemia	Increased respiratory drive, hypoxemia, stimulation of vascular receptors	Diffusion capacity (reduced); ECG; ECHO, RHC (to evaluate pulmonary artery pressures)[b]
Cardiac	Left heart failure Pericardial disease	Coronary artery disease, cardiomyopathy[c] Constrictive pericarditis; cardiac tamponade	Chest tightness, air hunger	Elevated left heart pressures; wet crackles on lung examination; pulsus paradoxus (pericardial disease)	Increased WOB and drive, hypoxemia, stimulation of vascular and pulmonary receptors[d]	Consider BNP testing, especially in the acute setting; ECG, ECHO, may need stress testing and/or LHC
Other	Variable	Anemia Deconditioning Psychological Metabolic disturbances Gastrointestinal (e.g., gastroesophageal reflux disease [GERD], aspiration pneumonitis) Post-COVID syndrome	Exertional breathlessness Poor fitness Anxiety	Variable	Metaboreceptors (anemia, poor fitness); chemoreceptors (anaerobic metabolism from poor fitness); some subjects may have increased sensitivity to hypercapnia	Hematocrit for anemia; laboratory studies (e.g., metabolic panel, thyroid hormone testing for metabolic disturbances); consider upper gastrointestinal endoscopy and/or esophageal pH probe testing for GERD and concerns for aspiration; consider referral to post-COVID care center for persistent symptoms after COVID infection; exclude other causes

[a]Differential diagnosis of interstitial lung disease includes idiopathic pulmonary fibrosis, collagen vascular disease, drug- or occupation-induced pneumonitis, lymphangitic spread of malignancy; processes that are more alveolar rather than interstitial in nature can also less commonly contribute to parenchymal lung disease underlying chronic dyspnea and include entities such as hypersensitivity pneumonitis, bronchiolitis obliterans organizing pneumonia, etc. [b]Would additionally consider these patients for CT angiography to evaluate for presence of thromboemboli and ventilation/perfusion scanning to evaluate for the presence of chronic thromboembolic disease. [c]Diastolic dysfunction in the setting of a stiff left ventricle is often seen and contributes significantly to insidious dyspnea that can be difficult to treat. [d]May stimulate metaboreceptors if cardiac output is sufficiently reduced to result in a lactic acidosis.

Abbreviations: BNP, brain natriuretic peptide; COPD, chronic obstructive pulmonary disease; COVID, coronavirus disease; CT, computed tomography; CXR, chest x-ray; ECG, electrocardiogram; ECHO, echocardiogram; GERD, gastroesophageal reflux disease; LHC, left heart catheterization; MIP/MEP, maximal inspiratory and maximal expiratory pressures (obtained in the pulmonary function testing laboratory); OVD, obstructive ventilatory defect; RVD, restrictive ventilatory defect; RHC, right heart catheterization; WOB, work of breathing.

but alveolar filling processes, such as hypersensitivity pneumonitis or bronchiolitis obliterans organizing pneumonia [BOOP], can also present with similar symptoms), diseases affecting the chest wall (e.g., bony abnormalities such as kyphoscoliosis, or neuromuscular weakness conditions such as amyotrophic lateral sclerosis), and diseases affecting the pulmonary vasculature (e.g., pulmonary hypertension that can arise from a variety of underlying causes, or chronic thromboembolic disease). Diseases affecting the cardiovascular system that can present with dyspnea include processes affecting left heart function, such as coronary artery disease and cardiomyopathy, as well as disease processes affecting the pericardium, including constrictive pericarditis and cardiac tamponade. Other conditions underlying dyspnea that might not directly emanate from the pulmonary or cardiovascular systems include anemia (thereby potentially affecting oxygen-carrying capacity), deconditioning, and psychological processes such as anxiety. Distinguishing among the myriad of underlying processes that might present with dyspnea can be challenging. A graded approach that begins with a history and physical examination, followed by selected laboratory testing that might then advance to additional diagnostics and potentially subspecialty referral, may help elucidate the underlying cause of dyspnea. However, a substantial proportion of patients may have persistent dyspnea despite treatment for an underlying process or may not have a specific underlying process identified that is driving the dyspnea.

APPROACH TO THE PATIENT

Dyspnea (See Fig. 39-2)

OVERALL

For patients with a known prior pulmonary, cardiac, or neuromuscular condition and worsening dyspnea, the initial focus of the evaluation will usually address determining whether the known

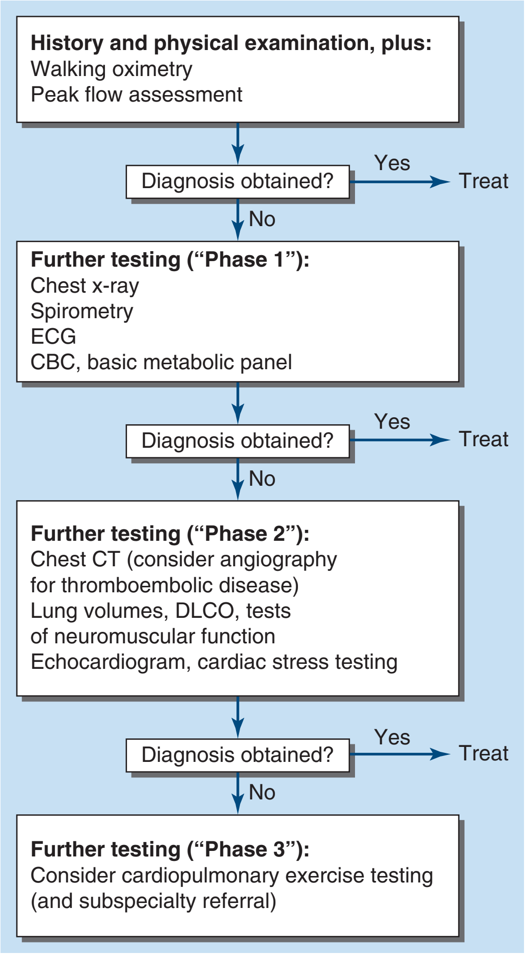
History and physical examination, plus:
Walking oximetry
Peak flow assessment
Diagnosis obtained?
Yes
Treat
No
Further testing ("Phase 1"):
Chest x-ray
Spirometry
ECG
CBC, basic metabolic panel
Diagnosis obtained?
Yes
Treat
No
Further testing ("Phase 2"):
Chest CT (consider angiography for thromboembolic disease)
Lung volumes, DLCO, tests of neuromuscular function
Echocardiogram, cardiac stress testing
Diagnosis obtained?
Yes
Treat
No
Further testing ("Phase 3"):
Consider cardiopulmonary exercise testing (and subspecialty referral)

FIGURE 39-2 Possible algorithm for the evaluation of the patient with dyspnea. As described in the text, the approach should begin with a detailed history and physical examination, followed by progressive testing and ultimately more invasive testing and subspecialty referral as is indicated to determine the underlying cause of dyspnea. CBC, complete blood count; DLCO, diffusing capacity of the lungs for carbon monoxide; ECG, electrocardiogram. *(Adapted from NG Karnani et al: Am Fam Physician 71:1529, 2005.)*

condition has progressed or whether a new process has developed that is causing dyspnea. For patients without a prior known potential cause of dyspnea, the initial evaluation will focus on determining an underlying etiology. Determining the underlying cause, if possible, is extremely important, as the treatment may vary dramatically based on the predisposing condition. An initial history and physical examination remain fundamental to the evaluation followed by initial diagnostic testing as indicated that might prompt subspecialty referral (e.g., pulmonary, cardiology, neurology, sleep, and/or specialized dyspnea clinic) if the cause of dyspnea remains elusive (Fig. 39-2). As many as two-thirds of patients will require diagnostic testing beyond the initial clinical presentation.

HISTORY

The patient should be asked to describe in their own words what the discomfort feels like as well as the effect of position, infections, and environmental stimuli on the dyspnea, as descriptors may be helpful in pointing toward an etiology. For example, symptoms of chest tightness might suggest the possibility of bronchoconstriction, and the sensation of inability to take a deep breath may correlate with dynamic hyperinflation from COPD. Orthopnea is a common indicator of congestive heart failure (CHF), mechanical impairment of the diaphragm associated with obesity, or asthma triggered by esophageal reflux. Nocturnal dyspnea suggests CHF or asthma. Acute, intermittent episodes of dyspnea are more likely to reflect episodes of myocardial ischemia, bronchospasm, or pulmonary embolism, whereas chronic persistent dyspnea is more typical of COPD, interstitial lung disease, and chronic thromboembolic disease. Information on risk factors for drug-induced or occupational lung disease and for coronary artery disease should be elicited. Left atrial myxoma or hepatopulmonary syndrome should be considered when the patient complains of *platypnea*—i.e., dyspnea in the upright position with relief in the supine position.

PHYSICAL EXAMINATION

Initial vital signs might be helpful in pointing toward an underlying etiology in the context of the remainder of the evaluation. For example, the presence of fever might point toward an underlying infectious or inflammatory process; the presence of hypertension in the setting of a heart failure might point toward diastolic dysfunction; the presence of tachycardia might be associated with many different underlying processes including fever, cardiac dysfunction, and deconditioning; and the presence of resting hypoxemia suggests processes involving hypercapnia, ventilation-perfusion mismatch, shunt, or impairment in diffusion capacity might be involved. An exertional oxygen saturation should also be obtained as described below. The physical examination should begin during the interview of the patient. Inability of the patient to speak in full sentences before stopping to get a deep breath suggests a condition that leads to stimulation of the controller or impairment of the ventilatory pump with reduced vital capacity. Evidence of increased work of breathing (supraclavicular retractions; use of accessory muscles of ventilation; and the tripod position, characterized by sitting with the hands braced on the knees) is indicative of increased airway resistance or stiffness of the lungs and the chest wall. When measuring the vital signs, the physician should accurately assess the respiratory rate and measure the pulsus paradoxus **(Chap. 281)**; if the systolic pressure decreases by >10 mmHg on inspiration, the presence of COPD, acute asthma, or pericardial disease should be considered. During the general examination, signs of anemia (pale conjunctivae), cyanosis, and cirrhosis (spider angiomata, gynecomastia) should be sought. Examination of the chest should focus on symmetry of movement; percussion (dullness is indicative of pleural effusion; hyperresonance is a sign of pneumothorax and emphysema); and auscultation (wheezes, rhonchi, prolonged expiratory phase, and diminished breath sounds are clues to disorders of the airways; rales suggest interstitial edema or fibrosis). The cardiac examination should focus on signs of elevated right heart pressures (jugular venous distention, edema, accentuated pulmonic component to the second heart sound); left ventricular dysfunction (S3 and S4 gallops); and valvular disease (murmurs). When examining the abdomen with the patient in the supine position, the physician should note whether there is paradoxical movement of the abdomen as well as the presence of increased respiratory distress in the supine position: inward motion during inspiration is a sign of diaphragmatic weakness, and rounding of the abdomen during exhalation is suggestive of pulmonary edema. Clubbing of the digits may be an indication of interstitial pulmonary fibrosis or bronchiectasis, and joint swelling or deformation as well as changes consistent with Raynaud's disease may be indicative of a collagen-vascular process that can be associated with pulmonary disease.

Patients should be asked to walk under observation with oximetry in order to reproduce the symptoms. The patient should be examined during and at the end of exercise for new findings that were not present at rest (e.g., presence of wheezing) and for changes in oxygen saturation.

CHEST IMAGING

After the history elicitation and the physical examination, a chest radiograph should be obtained if the diagnosis remains elusive. The lung volumes should be assessed: hyperinflation is consistent with obstructive lung disease, whereas low lung volumes suggest interstitial edema or fibrosis, diaphragmatic dysfunction, or impaired chest wall motion. The pulmonary parenchyma should be examined for evidence of interstitial disease, infiltrates, and emphysema. Prominent pulmonary vasculature in the upper zones indicates pulmonary venous hypertension, while enlarged central pulmonary arteries may suggest pulmonary arterial hypertension. An enlarged cardiac silhouette can point toward dilated cardiomyopathy or valvular disease. Bilateral pleural effusions are typical of CHF and some forms of collagen-vascular disease. Unilateral effusions raise the specter of carcinoma and pulmonary embolism but may also occur in heart failure or in the case of a parapneumonic effusion. CT of the chest is generally reserved for further evaluation of the lung parenchyma (e.g., interstitial lung disease) and possible pulmonary embolism (with CT angiography) if diagnostic uncertainty remains.

LABORATORY STUDIES

Initial laboratory testing should include a hematocrit to exclude occult anemia as an underlying cause of reduced oxygen-carrying capacity contributing to dyspnea, and a basic metabolic panel may be helpful to exclude a significant underlying metabolic acidosis (and conversely, an elevated bicarbonate might point toward the possibility of carbon dioxide retention that might be seen in chronic respiratory failure—in such a setting, an arterial blood gas may provide useful additional information). Additional laboratory studies should include electrocardiography to seek evidence of ventricular hypertrophy and prior myocardial infarction, and spirometry, which can be diagnostic of the presence of an obstructive ventilatory defect and suggest the possibility of a restrictive ventilatory defect (that then might prompt additional pulmonary function laboratory testing, including lung volumes, diffusion capacity, and possible tests of neuromuscular function). Echocardiography is indicated when systolic dysfunction, pulmonary hypertension, or valvular heart disease is suspected. Bronchoprovocation testing and/or home peak-flow monitoring may be useful in patients with intermittent symptoms suggestive of asthma who have a normal physical examination and spirometry; up to one-third of patients with the clinical diagnosis of asthma do not have reactive airways disease when formally tested. Measurement of brain natriuretic peptide levels in serum is increasingly used to assess for CHF in patients presenting with acute dyspnea but may be elevated in the presence of right ventricular strain as well.

DISTINGUISHING CARDIOVASCULAR FROM RESPIRATORY SYSTEM DYSPNEA

If a patient has evidence of both pulmonary and cardiac disease that is not responsive to treatment or it remains unclear what factors are primarily driving the dyspnea, a cardiopulmonary exercise test (CPET) can be conducted to determine which system is responsible for the exercise limitation. CPET includes incremental symptom-limited exercise (cycling or treadmill) with measurements of ventilation and pulmonary gas exchange and, in some cases, includes noninvasive and invasive measures of pulmonary vascular pressures and cardiac output. If, at peak exercise, the patient achieves predicted maximal ventilation, demonstrates an increase in dead space or hypoxemia, or develops bronchospasm, the respiratory system may be the cause of the problem. Alternatively, if the heart rate is >85% of the predicted maximum, if the anaerobic threshold occurs early, if the blood pressure becomes excessively high or decreases during exercise, if the O_2 pulse (O_2 consumption/heart rate, an indicator of stroke volume) falls, or if there are ischemic changes on the electrocardiogram, an abnormality of the cardiovascular system is likely the explanation for the breathing discomfort. Additionally, a CPET may also help point toward a peripheral extraction deficit or metabolic/neuromuscular disease as potential underlying processes driving dyspnea.

TREATMENT

Dyspnea

The first goal is to correct the underlying condition(s) driving dyspnea and address potentially reversible causes with appropriate treatment for the particular condition. Multiple different interventions may be necessary, given that dyspnea often arises from multifactorial causes. If relief of dyspnea with treatment of the underlying condition(s) is not fully possible, an effort is made to lessen the intensity of the symptom and its effect on the patient's quality of life. More recent work at the consensus conference level has sought to define an identifiable entity of persistent dyspnea in order to develop an approach to improving efforts to address symptom management for this condition. In 2017, an international group of experts defined "chronic breathlessness syndrome" as "the experience of breathlessness that persists despite optimal treatment of the underlying pathophysiology and results in disability for the patient." Despite an increased understanding of the mechanisms underlying dyspnea, there has been limited progress in treatment strategies for dyspnea. Supplemental O_2 should be administered if the resting O_2 saturation is ≤88% or if the patient's saturation drops to these levels with activity or sleep. In particular, for patients with COPD, supplemental oxygen for those with hypoxemia has been shown to improve mortality, and pulmonary rehabilitation programs (including some home and community-based exercise programs such as yoga and Tai Chi) have demonstrated positive effects on dyspnea, exercise capacity, and rates of hospitalization. More recent studies have suggested that supervised exercise programs similarly improved outcomes in post-COVID conditions. Opioids have been shown in some studies to reduce symptoms of dyspnea, largely through reducing air hunger, thus likely suppressing respiratory drive and influencing cortical activity. However, a recent study did not support the benefit of two different doses of daily low-dose extended-release morphine in COPD, for which reason opioids should be considered for each patient individually based on the risk-benefit profile in regard to respiratory depression. Studies of anxiolytics for dyspnea have not demonstrated consistent benefit.

FURTHER READING

Benzo R et al: Promoting chronic obstructive pulmonary disease wellness through remote monitoring and health coaching: A clinical trial. Ann Am Thorac Soc 19:1808, 2022.

Ekström M et al: Effect of regular, low-dose, extended-release morphine on chronic breathlessness in chronic obstructive pulmonary disease. The BEAMS randomized clinical trial. JAMA 328:2022, 2022.

Johnson M et al: Toward an expert consensus to delineate a clinical syndrome of chronic breathlessness: Chronic breathlessness syndrome. Eur Respir J 49:1602277, 2017.

Müller A et al: Prevalence of dyspnea in general adult populations: A systematic review and meta-analysis. Respir Med 218:107379, 2023.

O'Donnell DE et al: Unraveling the causes of unexplained dyspnea. Clin Chest Med 40:471, 2019.

Parshall MB et al: An Official American Thoracic Society Statement: Update on the mechanisms, assessment, and management of dyspnea. Am J Respir Crit Care Med 185:435, 2012.

Pouliopoulou DV et al: Rehabilitation interventions for physical capacity and quality of life in adults with post-COVID-19 condition. A systematic review and meta-analysis. JAMA Netw Open 6:e2333838, 2023.

Ratarasarn K et al: Yoga and Tai Chi: A mind-body approach in managing respiratory symptoms in obstructive lung diseases. Curr Opin Pulm Med 26:186, 2020.

40 Cough

Bruce D. Levy

Cough performs an essential protective function for human airways and lungs. Without an effective cough reflex, we are at risk for retained airway secretions and aspirated material predisposing to infection, atelectasis, and respiratory compromise. At the other extreme, excessive coughing can be exhausting; can be complicated by emesis, syncope, muscular pain, or rib fractures; can aggravate low back pain, abdominal or inguinal hernias, and urinary incontinence; and can be a major impediment to social interactions. Cough is often a clue to the presence of respiratory disease. In many instances, cough is an expected and accepted manifestation of disease, as in acute respiratory tract infection; however, persistent cough in the absence of other respiratory symptoms commonly causes patients to seek medical attention.

COUGH MECHANISM

Both chemical (e.g., capsaicin) and mechanical (e.g., mucus, particulates in air pollution) stimuli can initiate a cough. Cationic channels (e.g., transient receptor potential channels) and adenosine triphosphate–activated ion channels (P2X3) function as sensory neuronal receptors, with signals transmitted centrally via Aδ (mechanosensory) and C fibers (chemosensory). Afferent nerve endings richly innervate the pharynx, larynx, and airways to the level of the terminal bronchioles and extend into the lung parenchyma. They are also located in the external auditory canal (the auricular branch of the vagus nerve, or Arnold's nerve) and in the esophagus. Sensory signals travel via the vagus and superior laryngeal nerves to a region of the brainstem in the nucleus tractus solitarius. Integrated neural networks process this input into a conscious sensation referred to as the "urge to cough." The efferent limb of the cough reflex involves a highly orchestrated series of involuntary muscular actions, with the potential for input from cortical pathways as well, making possible voluntary cough. The vocal cords adduct, leading to transient upper-airway occlusion. Expiratory muscles contract, generating positive intrathoracic pressures as high as 300 mmHg. With sudden release of the laryngeal contraction, rapid expiratory flows are generated, exceeding the normal "envelope" of maximal expiratory flow seen on the flow-volume curve **(Fig. 40-1)**. Bronchial smooth-muscle contraction together with dynamic compression of airways narrow airway lumens and maximize the velocity of exhalation. The kinetic energy available to dislodge mucus from the inside of airway walls is directly proportional to the square of the velocity of expiratory airflow. A deep breath preceding a cough optimizes the function of the expiratory muscles; a series of repetitive coughs at successively lower lung volumes sweeps the point of maximal expiratory velocity progressively further into the lung periphery.

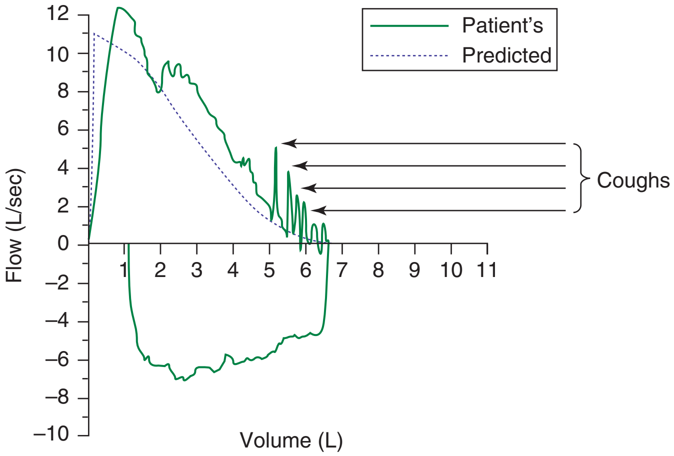

FIGURE 40-1 Flow-volume curve shows spikes of high expiratory flow achieved with cough.

TABLE 40-1 Causes of Impaired Cough and Airway Clearance
Respiratory muscle weakness
Chest wall or abdominal pain
Chest wall deformity (e.g., severe kyphoscoliosis)
Impaired glottic closure or tracheostomy
Central respiratory depression (e.g., anesthesia, sedation, or neurologic disease)
Abnormal airway secretions
Ciliary dysfunction
Tracheobronchomalacia
Bronchiectasis
Tracheal or bronchial stenoses

IMPAIRED COUGH

Weak or ineffective cough compromises the ability to clear lower respiratory tract secretions, predisposing to more serious infections and their sequelae. Weakness or paralysis of the expiratory (abdominal and intercostal) muscles and pain in the chest wall or abdomen are among the most common causes of impaired cough **(Table 40-1)**. Cough strength is generally assessed qualitatively; peak expiratory flow or maximal expiratory pressure at the mouth can be used as a surrogate marker for cough strength. A variety of assistive devices and techniques have been developed to improve cough efficacy, running the gamut from simple (splinting of the abdominal muscles with a tightly held pillow to reduce postoperative pain while coughing) to complex (a mechanical cough-assist device supplied via face mask or tracheal tube that applies a cycle of positive pressure followed rapidly by negative pressure). Cough may fail to clear secretions completely despite a preserved ability to generate normal expiratory velocities; such failure may be due to abnormal airway secretions (e.g., abnormally viscous secretions of cystic fibrosis), ciliary dysfunction (e.g., primary ciliary dyskinesia), or structural abnormalities of the airways (e.g., tracheomalacia with excessive expiratory collapse of the trachea during cough).

SYMPTOMATIC COUGH

Cough may occur in the context of other respiratory symptoms that together point to a diagnosis; for example, cough accompanied by wheezing, shortness of breath, and chest tightness after exposure to a cat or other sources of allergens suggests asthma. At times, cough is the dominant or sole symptom of disease, and it may be of sufficient duration and severity that relief is sought. The duration of cough is a clue to its etiology, at least retrospectively. Acute cough (<3 weeks) is most commonly due to a respiratory tract infection, aspiration, or inhalation of noxious chemicals or smoke. Subacute cough (3–8 weeks in duration) is a common residuum of tracheobronchitis, as in pertussis or "postviral tussive syndrome." Chronic cough (>8 weeks) may be caused by a wide variety of cardiopulmonary diseases, including those of inflammatory, infectious, neoplastic, and cardiovascular etiologies. When initial assessment with chest examination and radiography is normal, cough-variant asthma, gastroesophageal reflux, rhinosinusitis with excessive nasopharyngeal drainage, and medications (angiotensin-converting enzyme [ACE] inhibitors) are the most common identifiable causes of chronic cough. In a long-time cigarette smoker, an early-morning, productive cough suggests chronic bronchitis. A dry, irritative cough that lingers for >2 months following one or more respiratory tract infections ("postbronchitic cough") is a very common cause of chronic cough, especially in the winter months. Chronic cough in the absence of identifiable etiology has been recognized with increasing frequency, is thought to be due to exaggerated neurologic signaling via sensory cough-reflex pathways, and is referred to as "chronic cough hypersensitivity syndrome."

ASSESSMENT OF CHRONIC COUGH

Except for our ability to detect the sound of excess airway secretions, details as to the resonance of the cough, its time of occurrence during the day, and the pattern of coughing (e.g., occurring in paroxysms) infrequently provide useful etiologic clues. Regardless of cause, cough often worsens upon first lying down at night, with talking, or with the hyperpnea of exercise; it frequently improves with sleep. An exception may involve the cough that occurs only with certain allergic exposures or exercise in cold air, as in asthma. Useful historical questions include what circumstances surrounded the onset of cough, what makes the cough better or worse, and whether the cough produces sputum.

The physical examination seeks clues suggesting the presence of cardiopulmonary disease, including findings such as wheezing or crackles on chest examination. Examination of the auditory canals and tympanic membranes (for irritation of the latter resulting in stimulation of Arnold's nerve), the nasal passageways (for rhinitis or polyps), and the nails (for clubbing) may also provide etiologic clues. Because cough can be a manifestation of a systemic disease such as sarcoidosis or vasculitis, a thorough general examination is likewise important.

In virtually all instances, evaluation of chronic cough merits a chest radiograph. The list of diseases that can cause persistent cough without other symptoms and without detectable abnormalities on physical examination is long. It includes serious illnesses such as sarcoidosis or Hodgkin's disease in young adults, lung cancer in older patients, and (worldwide) pulmonary tuberculosis. An abnormal chest film prompts an evaluation aimed at explaining the radiographic abnormality. In a patient with chronic productive cough, examination of expectorated sputum is warranted, because determining the cause of mucus hypersecretion is a crucial clue to etiology. Purulent-appearing sputum should be sent for routine bacterial culture and, in certain circumstances, mycobacterial culture as well. Cytologic examination of mucoid sputum may be useful to assess for malignancy and oropharyngeal aspiration and to distinguish neutrophilic from eosinophilic bronchitis. Expectoration of blood—whether streaks of blood, blood mixed with airway secretions, or pure blood—deserves a special approach to assessment and management **(Chap. 41)**.

CHRONIC COUGH WITH A NORMAL CHEST RADIOGRAPH

It is commonly held that (alone or in combination) the use of an ACE inhibitor; postnasal drainage; gastroesophageal reflux; and asthma account for >90% of cases of chronic cough with a normal or noncontributory chest radiograph; however, clinical experience does not support this contention. Strict adherence to this concept discourages the search for alternative explanations by both clinicians and researchers. In recent years, the concept of a distinct "cough hypersensitivity syndrome" has emerged, emphasizing the putative role of sensitized sensory nerve endings and afferent neural pathways in causing chronic refractory cough, akin to chronic neuropathic pain. It presents with a dry or minimally productive cough and a tickle or sensitivity in the throat and is made worse with talking, laughing, or exertion. It is more common in women than men and can last for years. Specific diagnostic criteria are lacking; the diagnosis is suspected when alternative etiologies are excluded by diagnostic testing or failed therapeutic trials. It is uncertain whether persistent daily coughing elicits an inflammatory response and is thereby self-perpetuating.

ACE inhibitor–induced cough occurs in 5–30% of patients taking these agents and is not dose-dependent. ACE metabolizes bradykinin and other tachykinins, such as substance P. The mechanism of ACE inhibitor–associated cough may involve sensitization of sensory nerve endings due to accumulation of bradykinin. Any patient with chronic unexplained cough who is taking an ACE inhibitor should have a trial period off the medication, regardless of the timing of the onset of cough relative to the initiation of ACE inhibitor therapy. In most instances, a safe alternative is available; angiotensin receptor blockers do not cause cough. Failure to observe a decrease in cough after 1 month off medication argues strongly against this etiology.

Postnasal drainage of any etiology can cause cough as a response to stimulation of sensory receptors of the cough-reflex pathway in the hypopharynx or aspiration of draining secretions into the trachea. The term *upper airway cough syndrome* has been coined to encompass the concept that chronic inflammation in the nose and sinuses can cause cough even in the absence of physical drainage into the pharynx. Historical clues suggesting this etiology include a sensation of postnasal drip, frequent throat clearing, and sneezing and rhinorrhea. On speculum examination of the nose, excess mucoid or purulent secretions, inflamed and edematous nasal mucosa, and/or polyps may be seen; in addition, secretions or a cobblestoned appearance of the mucosa along the posterior pharyngeal wall may be noted. Unfortunately, there is no means by which to quantitate postnasal drainage. In many instances, this diagnosis must rely on subjective information provided by the patient. Furthermore, this assessment must also be counterbalanced by the fact that many people who have chronic postnasal drainage do not experience cough.

Linking gastroesophageal reflux to chronic cough poses similar challenges. It is thought that reflux of gastric contents into the lower esophagus may trigger cough via reflex pathways initiated in the esophageal mucosa. Reflux to the level of the pharynx (laryngopharyngeal reflux), with consequent aspiration of gastric contents, causes a chemical bronchitis and possibly pneumonitis that can elicit cough for days afterward, but it is an uncommon finding among persons with chronic cough. Retrosternal burning after meals or on recumbency, frequent eructation, hoarseness, and throat pain may be indicative of gastroesophageal reflux. Nevertheless, reflux may also elicit minimal or no symptoms. Glottic inflammation detected on laryngoscopy may be a manifestation of recurrent reflux to the level of the throat, but it is a nonspecific finding. Quantification of the frequency and level of reflux requires a somewhat invasive procedure to measure esophageal pH (either nasopharyngeal placement of a catheter with a pH probe into the esophagus for 24 h or endoscopic placement of a radiotransmitter capsule into the esophagus) and, with newer techniques, esophageal pressures (manometry) and nonacid reflux. The precise interpretation of test results that permits an etiologic linking of reflux events and cough remains debated. Again, assigning the cause of cough to gastroesophageal reflux must be weighed against the observation that many people with symptomatic reflux do not experience chronic cough.

Cough alone as a manifestation of asthma is common among children but not among adults. Cough due to asthma in the absence of wheezing, shortness of breath, and chest tightness is referred to as "cough-variant asthma." A history suggestive of cough-variant asthma ties the onset of cough to an exposure that typically triggers asthma and the resolution of cough to discontinuation of the exposure. Objective testing can establish the diagnosis of asthma (airflow obstruction on spirometry that varies over time or reverses in response to a bronchodilator) or exclude it with certainty (a negative response to a bronchoprovocation challenge—e.g., with methacholine). In a patient capable of taking reliable measurements, home expiratory peak flow monitoring can be a cost-effective method to support or discount a diagnosis of asthma.

Eosinophilic bronchitis causes chronic cough with a normal chest radiograph. This uncommon condition is characterized by sputum eosinophilia in excess of 3% without airflow obstruction or bronchial hyperresponsiveness and is successfully treated with inhaled glucocorticoids. Measurement of an elevated concentration of nitric oxide in exhaled breath has the potential to detect eosinophilic airway inflammation (in asthma or eosinophilic bronchitis) and predict a favorable response to inhaled steroids in persons with chronic cough.

Treatment of chronic cough in a patient with a normal chest radiograph is often empirical and is targeted at the most likely cause(s) of cough as determined by history, physical examination, and possibly pulmonary function testing. Therapy for postnasal drainage depends on the presumed etiology (infection, allergy, or vasomotor rhinitis) and may include systemic antihistamines; decongestants; antibiotics; nasal saline irrigation; and nasal pump sprays with glucocorticoids, antihistamines, or anticholinergics. Antacids, histamine type 2 (H_2) receptor antagonists, and proton pump inhibitors are used to neutralize or decrease the production of gastric acid in gastroesophageal reflux disease; dietary changes, elevation of the head and torso during sleep, and medications to improve gastric emptying or impede the flow of refluxate (e.g., alginates) are additional therapeutic measures. Cough-variant asthma typically responds well to inhaled glucocorticoids and intermittent use of inhaled β_2-agonist bronchodilators.

Patients who fail to respond to treatment targeting the common causes of chronic cough or who have had these causes excluded by appropriate diagnostic testing should, in the opinion of the author, undergo chest CT. Diseases causing cough that may be missed on chest x-ray include tumors, early interstitial lung disease, bronchiectasis, and atypical mycobacterial pulmonary infection. On the other hand, patients with chronic cough who have normal findings on chest examination, lung function testing, oxygenation assessment, and chest CT can be reassured as to the absence of serious pulmonary pathology.

GLOBAL CONSIDERATIONS

Regular exposure to air pollution can cause chronic cough and throat clearing, as well as lower respiratory tract disease. Smoke from cooking and heating fuels in poorly ventilated homes; toxic exposures in work settings lacking implementation of occupational safety standards; and ambient chemicals and particulates in highly polluted outdoor air are all forms of air pollution causing cough. Limited therapeutic options are available; treatment focuses on improving environmental air quality (e.g., use of a stove chimney in the home), removal from the exposure, and use of an appropriate face mask.

In areas of the world where tuberculosis is endemic, chronic cough conjures the possibility of active pulmonary tuberculosis and mandates appropriate evaluation, including chest imaging and sputum analysis.

SYMPTOM-BASED TREATMENT OF COUGH

Empiric treatment of chronic idiopathic cough with inhaled corticosteroids, inhaled anticholinergic bronchodilators, and macrolide antibiotics has been tried without consistent success. Currently available cough suppressants are only modestly effective. Most potent are narcotic cough suppressants, such as codeine, hydrocodone, or morphine, which are thought to act in the “cough center” in the brainstem. The tendency of narcotic cough suppressants to cause drowsiness and constipation and their potential for addictive dependence limit their appeal for long-term use. Dextromethorphan is an over-the-counter, centrally acting cough suppressant with fewer side effects and less efficacy than the narcotic cough suppressants. Dextromethorphan is thought to have a different site of action than narcotic cough suppressants and can be used in combination with them if necessary. Benzonatate is thought to inhibit neural activity of sensory nerves in the cough-reflex pathway. It is generally free of side effects; however, its effectiveness in suppressing cough is variable and unpredictable. Inhaled lidocaine, an inhibitor of voltage-gated sodium channels, provides transient cough suppression, but because of associated oropharyngeal anesthesia, it poses an increased risk of aspiration.

Attempts to treat cough hypersensitivity syndrome have focused on inhibition of neural pathways. Small case series and randomized clinical trials have indicated benefit from off-label use of gabapentin, pregabalin, or amitriptyline. Recent studies suggest a role for behavioral modification using specialized speech therapy techniques, but widespread application of this modality is currently not practical. Novel cough suppressants without the limitations of currently available agents are greatly needed. Approaches that are being explored include the development of neurokinin-1 receptor antagonists, transient receptor protein vanilloid-1 (TRPV1) channel antagonists, P2X3 channel antagonists, and novel opioid and opioid-like receptor agonists.

ACKNOWLEDGMENT

Christopher H. Fanta contributed to this chapter in the 21st edition and some material from that chapter has been retained here.

FURTHER READING

Brightling CE et al: Eosinophilic bronchitis as an important cause of chronic cough. Am J Respir Crit Care Med 160:406, 1999.

Carroll TL (ed): *Chronic Cough*. San Diego, Plural Publishing, Inc., 2019.

Gibson P et al: Treatment of unexplained chronic cough: CHEST guideline and expert panel report. Chest 149:27, 2016.

Kahrilas PJ et al: Chronic cough due to gastroesophageal reflux in adults: CHEST Guideline and Expert Panel Report. Chest 150:1381, 2016.

Mazzone SB et al: Chronic cough and cough hypersensitivity: From mechanistic insights to novel antitussives. Lancet Respir Med 10:1113, 2022.

Morice AH et al: ERS guidelines on the diagnosis and treatment of chronic cough in adults and children. Eur Respir J 55:1901136, 2020.

Smith JA, Woodcock A: Chronic cough. N Engl J Med 375:1544, 2016.

41

Hemoptysis

Erin M. DeBiasi, Carolyn D’Ambrosio

Hemoptysis is the expectoration of blood originating from the lower respiratory tract. It can be confused initially with bleeding from the gastrointestinal tract (hematemesis) or nasal cavities (epistaxis). The amount of blood that is being expectorated (volume and frequency) is the most important information to gather as massive or life-threatening hemoptysis (variable definitions but commonly expectorating >150 mL in 24 h or a bleeding rate of ≥100 mL/h) requires emergent intervention. This chapter focuses on non-life-threatening hemoptysis, which is more common.

ANATOMY AND PHYSIOLOGY OF HEMOPTYSIS

Hemoptysis originates in the lower respiratory tract, anywhere from the glottis to the alveolus. The bleeding most commonly arises from the bronchi or medium-sized airways, but a thorough evaluation of the entire respiratory tree is important. The blood supply to the lungs is from both the pulmonary and bronchial circulations. The pulmonary circulation is a low-pressure system that is essential for gas exchange at the alveoli; in contrast, the bronchial circulation originates from the aorta and, thus, is a higher-pressure system. The bronchial arteries supply the airways and can neovascularize tumors, dilated airways in bronchiectasis, and cavitary lesions. Most hemoptysis originates from the bronchial circulation, the higher-pressure system, which can make it difficult to control.

ETIOLOGY

Infection, malignancy, and vascular disease are some of the common causes of hemoptysis, but the differential is quite broad. In the United States, the most common causes remain viral bronchitis, bronchiectasis, or malignancy. In other parts of the world, infections such as tuberculosis are the most common causes.

INFECTIONS

Although most small-volume hemoptysis cases are due to viral bronchitis, patients with chronic bronchitis are at risk for bacterial superinfection. *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis* are the more common bacteria involved, and these infections can increase airway inflammation that leads to bleeding. Similarly, patients with bronchiectasis, including those with cystic fibrosis, can have hemoptysis during exacerbations. Due to recurrent bacterial infection, bronchiectatic airways are dilated, inflamed, and highly vascular, supplied by the bronchial circulation. This can cause bronchiectasis to also be a significant cause of massive hemoptysis and subsequent death.

Tuberculosis used to be the most common cause of hemoptysis worldwide, but in industrialized countries, bronchitis and bronchiectasis are more common. In patients with tuberculosis, development of cavitary disease is frequently the source of bleeding, but rarer complications such as erosion of a pulmonary artery aneurysm into a preexisting cavity (i.e., Rasmussen’s aneurysm) can also be the source.

Other infectious agents such as endemic fungi, *Nocardia*, and nontuberculous mycobacteria can present as cavitary lung disease complicated by hemoptysis. In addition, *Aspergillus* species can develop into mycetomas within preexisting cavities, with neovascularization to these inflamed spaces leading to bleeding. Pulmonary abscesses and necrotizing pneumonia can cause bleeding by devitalizing lung parenchyma. Common responsible organisms include *Staphylococcus aureus, Klebsiella pneumoniae*, and oral anaerobes.

Paragonimiasis can mimic tuberculosis and is another significant cause of hemoptysis seen globally; it is common in Southeast Asia and China, although cases have been reported in North America from raw crayfish ingestion. It should be considered as a cause of hemoptysis in recent immigrants from endemic areas.

■ VASCULAR

Hemoptysis from a vascular cause can be associated with cardiac disease, pulmonary embolism, arteriovenous malformation, or diffuse alveolar hemorrhage (DAH). While the classic description of the sputum expectorated in pulmonary edema (from elevated left end-diastolic pressure) is "pink and frothy," a spectrum of hemoptysis including frank blood can be seen. This observation is particularly true now with the more widespread use of anticoagulants and antiplatelet medications.

Pulmonary embolism with parenchymal infarction can present with hemoptysis, but pulmonary emboli do not commonly cause hemoptysis. An ectatic vessel in an airway or a pulmonary arteriovenous malformation can be a source of bleeding. A rare vascular cause of hemoptysis is the rupture of an aortobronchial fistula; these fistulae arise in the setting of aortic pathology such as aneurysm or pseudoaneurysm and can cause small bleeding episodes that result in massive hemoptysis.

DAH causes significant bleeding into the lung parenchyma but, interestingly, is not often associated with hemoptysis. DAH typically presents with diffuse ground-glass opacities on chest imaging. A range of insults cause DAH, including immune-mediated capillaritis from diseases such as systemic lupus erythematosus, toxicity from cocaine and other inhalants, and stem cell transplantation. The so-called "pulmonary-renal" syndromes, including granulomatosis with polyangiitis and anti–glomerular basement membrane (anti-GBM) disease, may lead to both hemoptysis and hematuria (though one manifestation may be present without the other). A recently identified cause of hemoptysis and DAH is vaping-induced lung injury.

■ MALIGNANCY

Bronchogenic carcinoma of any histology is a common cause of hemoptysis (both massive and nonmassive). Hemoptysis can indicate airway involvement of the tumor and can be a presenting symptom of carcinoid tumors, vascular lesions that frequently arise in the proximal airways. Small-cell and squamous cell carcinomas are frequently central in nature and more likely to erode into major pulmonary vessels, resulting in massive hemoptysis. Pulmonary metastases from distant tumors (e.g., melanoma, sarcoma, adenocarcinomas of the breast and colon) can also cause bleeding. Kaposi's sarcoma, seen in advanced acquired immunodeficiency syndrome, is very vascular and can develop anywhere along the respiratory tract, from the bronchi to the oral cavity.

■ MECHANICAL AND OTHER CAUSES

In addition to infection, vascular disease, and malignancy, other insults to the pulmonary system can cause hemoptysis. Pulmonary endometriosis causes cyclical bleeding known as catamenial hemoptysis. Foreign body aspiration can lead to airway irritation and bleeding. Diagnostic and therapeutic procedures are also potential offenders: pulmonary vein stenosis can result from left atrial procedures, such as pulmonary vein isolation, and pulmonary artery catheters can lead to rupture of the pulmonary artery if the distal balloon is kept inflated. Finally, in the setting of thrombocytopenia, coagulopathy, anticoagulation, or antiplatelet therapy, even minor insults can cause hemoptysis.

EVALUATION AND MANAGEMENT

■ HISTORY

The initial history should be directed at assessing the pattern, severity, and quantity of hemoptysis. An approach to management of hemoptysis is outlined in **Fig. 41-1**. A patient's description of the sputum (e.g., flecks of blood, pink-tinged, frank blood or clot) is helpful if you cannot examine it. Quantification is often challenging for patients, so using references like cups (one U.S. cup is 236 mL) can be helpful. Life-threatening hemoptysis is defined by the presence of significantly abnormal gas exchange, hemodynamic compromise, or threat for airway obstruction. Patients rarely die of exsanguination but, rather, are at risk of death due to asphyxiation from blood filling the airways and airspaces. This can occur with blood loss of >400 mL within 24 h or >150 mL at one time. Fortunately, life-threatening hemoptysis only accounts for 5–15% of cases of hemoptysis.

Further history may help define the etiology of hemoptysis. Smoking history and/or unintentional weight loss may point to possible malignancy. Preceding fevers, cough, and/or sputum production may suggest infection. A history of prior diagnosed chronic lung conditions, especially cystic fibrosis or other chronic bronchiectatic diseases, is important to note. Screening for causes of pseudohemoptysis (i.e., other upper airway or gastrointestinal) is also helpful.

■ PHYSICAL EXAMINATION

Patients should initially be assessed for signs of life-threatening hemoptysis including hypoxemia, tachycardia, and hemodynamic instability. Examination should include possible sites of extrapulmonary bleeding such as the nasal and oral cavities. Auscultation of the lungs may suggest a laterality. Other relevant physical findings may suggest other etiologies of the hemoptysis and include clubbing, signs of a bleeding diathesis (e.g., skin or mucosal ecchymoses and petechiae), telangiectasias, or skin rash.

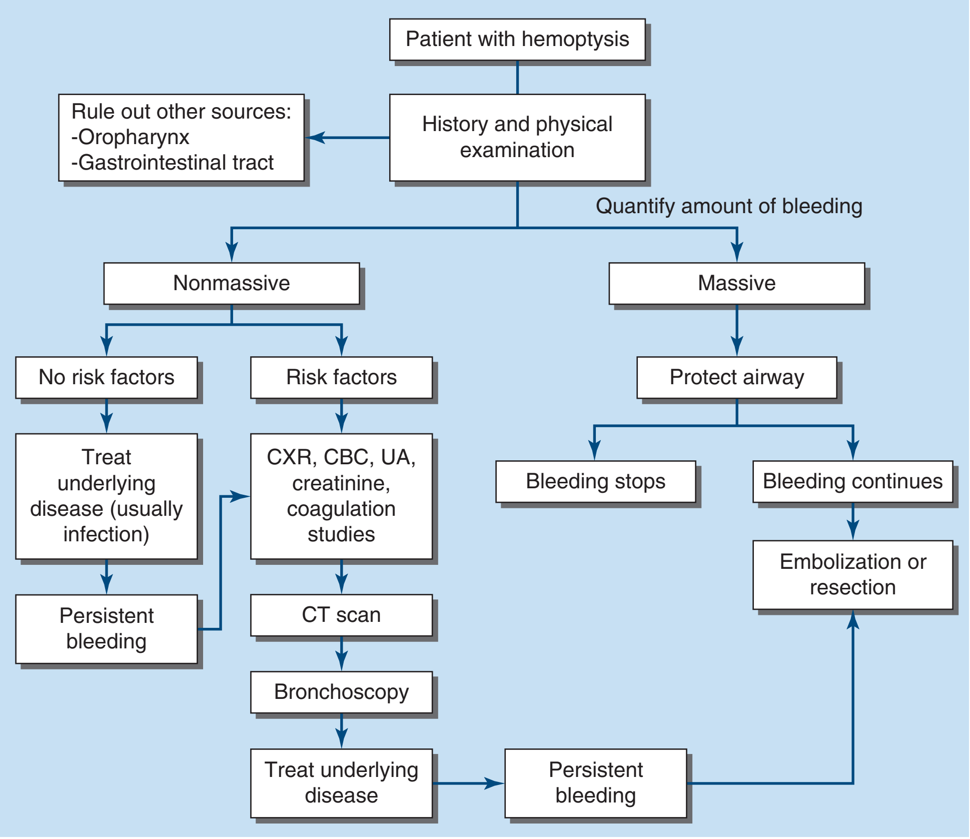

FIGURE 41-1 Approach to the management of hemoptysis. CBC, complete blood count; CT, computed tomography; CXR, chest x-ray; UA, urinalysis.

DIAGNOSTIC STUDIES

Laboratory studies for initial evaluation should include a complete blood count, coagulation studies, basic metabolic panel, and urinalysis (to evaluate for pulmonary-renal disease). The absence of anemia in life-threatening hemoptysis is not uncommon, as asphyxiation can occur without large blood volume loss. Infectious workup (including bacterial sputum culture, acid-fast bacillus culture, and respiratory viral panel) and serologic workup (including antinuclear antibody, antineutrophilic cytoplasmic antibody, and anti-GBM) can be obtained if indicated.

Imaging of the chest is essential. While a chest radiograph is often first obtained, aside from large masses, it often does not localize the site and can be normal. In most patients without risk factors for malignancy or other abnormalities in the initial evaluation and with a normal chest radiograph, treating for bronchitis and ensuring close follow-up is a reasonable strategy, with further diagnostic workup.

In contrast, patients with risk factors for malignancy (i.e., age >40, especially those with risk factors for malignancy or significant hemoptysis) should have a chest computed tomography (CT). A chest CT identifies the site and cause of hemoptysis in the majority of patients. A CT angiogram can be considered to identify a source for possible embolization if needed. A diagnostic bronchoscopy can be a complimentary study performed to evaluate for subtle mucosal lesions that are often missed on CT scans, perform interventions as per below, or localize the general location and laterality of bleeding. This is best performed during an acute episode of hemoptysis to increase the yield.

INTERVENTIONS

When the amount of hemoptysis is life-threatening, there are three simultaneous goals: first, protect the nonbleeding lung; second, locate the site of bleeding; and third, control the bleeding.

Protecting the airway and nonbleeding lung is paramount in the management of life-threatening hemoptysis because asphyxiation can happen quickly. If the side of bleeding is known, the patient should be positioned with the bleeding side down to use gravitational advantage to keep blood out of the nonbleeding lung. A strong cough reflex can often be highly effective at clearing clot from the airway, but intubation is sometimes necessary if the patient cannot protect their airway. Intubation with a large endotracheal tube (size 8.5) is preferred to allow for suctioning and interventions. If needed, this can be advanced into the mainstem airway of the nonbleeding lung for isolation. Double-lumen endotracheal tubes can also be used, but the lumen of each is often too small for suctioning and therapeutic interventions.

Imaging obtained during initial evaluation can indicate the source of the bleeding. Flexible bronchoscopy during active bleeding is most useful. Even if a bleeding lesion is not identified, flexible bronchoscopy can help identify a side and lobe. Finally, proceeding directly to CT angiography is also a reasonable strategy given that it has both diagnostic and therapeutic capabilities.

Controlling the bleeding during an episode of life-threatening hemoptysis can be accomplished in one of three ways: from the airway lumen, from the involved blood vessel, or by surgical resection of both airway and vessel involved. Any known bleeding diathesis should be corrected first. As above, bronchoscopic measures can often be diagnostic, but also therapeutic if the bleeding is from a central airway lesion. Initially, a flexible bronchoscope can be used to help stabilize a patient via suctioning clot and insertion of a balloon catheter or bronchial blocker to occlude the involved airway. Additionally, instillation of iced saline or epinephrine can help temporize bleeding. Tranexamic acid (TXA), an antifibrinolytic agent (500 mg/5 mL), can be instilled during bronchoscopy or administered via nebulizer. Rigid bronchoscopy, done by an interventional pulmonologist or thoracic surgeon, allows therapeutic interventions of bleeding airway lesions such as argon plasma coagulation and cautery, which can be definitive. However, because most life-threatening cases of hemoptysis arise from the bronchial circulation, bronchial artery embolization is the procedure of choice for control of the bleeding. It is generally successful in the short term, with >80% success rate at controlling bleeding immediately, although bleeding can recur if the underlying disease (e.g., a mycetoma) is not treated. Surgical resection has a high mortality rate (up to 15–40%) and should not be pursued unless initial measures have failed and bleeding is ongoing. Ideal candidates for surgery have localized disease but otherwise normal lung parenchyma.

Acknowledgment

Anna K. Brady and Patricia A. Kritek contributed to this chapter in the 20th edition and some material from that chapter has been retained here.

FURTHER READING

Bellam BL et al: Efficacy of tranexamic acid in haemoptysis: A randomized, controlled pilot study. Pulm Pharmacol Ther 40:80, 2016.

Davidson K, Shojaee S: Managing massive hemoptysis. Chest 157:77, 2020.

Flume PA et al: CF pulmonary guidelines. Pulmonary complications: Hemoptysis and pneumothorax. Am J Respir Crit Care Med 182:298, 2010.

Layden JE et al: Pulmonary illness related to e-cigarette use in Illinois and Wisconsin-final report. N Engl J Med 382:903, 2020.

Lim RK et al: Evaluating hemoptysis hospitalizations among patients with bronchiectasis in the United States: A population-based cohort study. BMC Pulm Med 21:392, 2021.

Mondoni M et al: Observational, multicentre study on the epidemiology of haemoptysis. Eur Resp J 51:1701813, 2018.

Mondoni M et al: Bronchoscopy to assess patients with hemoptysis: Which is the optimal timing? BMC Pulm Med 19:36, 2019.

42

Hypoxia and Cyanosis

Joseph Loscalzo

HYPOXIA

The fundamental purpose of the cardiorespiratory system is to deliver O_2 and nutrients to cells and to remove CO_2 and other metabolic products from them. Proper maintenance of this function depends not only on intact cardiovascular and respiratory systems, but also on an adequate number of red blood cells and hemoglobin, and a supply of inspired gas containing adequate O_2.

RESPONSES TO HYPOXIA

Decreased O_2 availability to cells typically results in an inhibition of oxidative phosphorylation and increased anaerobic glycolysis. This switch from aerobic to anaerobic metabolism, the Pasteur effect, reduces the yield of adenosine 5′-triphosphate (ATP) produced per mole of glucose. In severe hypoxia, when ATP production is inadequate to meet the energy requirements of ionic and osmotic equilibrium, cell membrane depolarization leads to uncontrolled Ca^{2+} influx and activation of Ca^{2+}-dependent phospholipases and proteases. These events, in turn, cause cell swelling, activation of apoptotic pathways, and, ultimately, cell death.

The adaptations to hypoxia are mediated, in part, by the upregulation of genes encoding a variety of proteins, including glycolytic enzymes, such as phosphoglycerate kinase and phosphofructokinase, as well as the glucose transporters Glut-1 and Glut-2; and by growth factors, such as vascular endothelial growth factor (VEGF) and erythropoietin, which enhance erythrocyte production. The hypoxia-induced increase in expression of these and other key proteins is largely governed by the hypoxia-sensitive transcription factor, hypoxia-inducible factor-1 (HIF-1).

During hypoxia, systemic arterioles dilate, at least in part, by opening of K_{ATP} channels in vascular smooth-muscle cells due to the hypoxia-induced reduction in ATP concentration. By contrast, in pulmonary vascular smooth-muscle cells, inhibition of K^+ channels causes depolarization, which, in turn, activates voltage-gated Ca^{2+} channels, raising the cytosolic $[Ca^{2+}]$ and causing smooth-muscle cell contraction. Hypoxia-induced pulmonary arterial constriction shunts blood away from poorly ventilated portions toward better ventilated portions of the lung (i.e., improves ventilation-perfusion mismatch); however, it also increases pulmonary vascular resistance and right ventricular afterload.

Effects on the Central Nervous System Changes in the central nervous system (CNS) function, particularly the higher centers, are especially important consequences of hypoxia. Acute hypoxia causes impaired judgment, motor incoordination, and a clinical picture resembling acute alcohol intoxication. High-altitude illness is characterized by headache secondary to cerebral vasodilation, gastrointestinal symptoms, dizziness, insomnia, fatigue, or somnolence. Pulmonary arterial and sometimes venous constriction causes capillary leakage and high-altitude pulmonary edema (HAPE) **(Chap. 39)**, which intensifies hypoxia, further promoting vasoconstriction. Rarely, high-altitude cerebral edema (HACE) develops, which is manifest by severe headache and papilledema, and can cause coma. As hypoxia becomes more severe, the regulatory centers of the brainstem are affected, and death usually results from respiratory failure.

Effects on the Cardiovascular System Acute hypoxia stimulates the chemoreceptor reflex arc to induce venoconstriction and systemic arterial vasodilation. These acute changes are accompanied by transiently increased myocardial contractility, which is followed by depressed myocardial contractility with prolonged hypoxia.

CAUSES OF HYPOXIA

Respiratory Hypoxia When hypoxia occurs from respiratory failure, $Pa{O_2}$ declines, and when respiratory failure is persistent, the hemoglobin-oxygen (Hb-O_2) dissociation curve **(see Fig. 103-2)** is displaced to the right, with greater quantities of O_2 released at any level of tissue $P{O_2}$. Arterial hypoxemia, that is, a reduction of O_2 saturation of arterial blood ($Sa{O_2}$), and consequent cyanosis are likely to be more marked when such depression of $Pa{O_2}$ results from pulmonary disease than when the depression occurs as the result of a decline in the fraction of oxygen in inspired air ($F{I}{O_2}$). In this latter situation, $Pa{CO_2}$ falls secondary to anoxia-induced hyperventilation and the Hb-O_2 dissociation curve is displaced to the left, limiting the decline in $Sa{O_2}$ at any level of $Pa{O_2}$.

The most common cause of respiratory hypoxia is *ventilation-perfusion mismatch* resulting from perfusion of poorly ventilated alveoli. Respiratory hypoxemia may also be caused by *hypoventilation*, in which case it is associated with an elevation of $Pa{CO_2}$ **(Chap. 296)**. These two forms of respiratory hypoxia are usually correctable by inspiring 100% O_2 for several minutes. A third cause of respiratory hypoxia is shunting of blood across the lung from the pulmonary arterial to the venous bed (*intrapulmonary right-to-left shunting*) by perfusion of nonventilated portions of the lung, as in pulmonary atelectasis or through pulmonary arteriovenous connections. The low $Pa{O_2}$ in this situation is only partially corrected by an $F{I}{O_2}$ of 100%.

Hypoxia Secondary to High Altitude As one ascends rapidly to 3000 m (~10,000 ft), the reduction of the O_2 content of inspired air ($F{I}{O_2}$) leads to a decrease in alveolar $P{O_2}$ to ~60 mmHg, and a condition termed *high-altitude illness* develops (see above). At higher altitudes, arterial saturation declines rapidly and symptoms become more severe; and at 5000 m, unacclimated individuals usually cease to be able to function normally owing to the changes in CNS function described above.

Hypoxia Secondary to Right-to-Left Extrapulmonary Shunting From a physiologic viewpoint, this cause of hypoxia resembles intrapulmonary right-to-left shunting but is caused by congenital cardiac malformations, such as tetralogy of Fallot, transposition of the great arteries, atrial or ventricular septal defect, patent ductus arteriosus, and Eisenmenger's syndrome **(Chap. 280)**. As in pulmonary right-to-left shunting, the $Pa{O_2}$ cannot be restored to normal with inspiration of 100% O_2.

Anemic Hypoxia A reduction in hemoglobin concentration of the blood is accompanied by a corresponding decline in the O_2-carrying capacity of the blood. Although the $Pa{O_2}$ is normal in anemic hypoxia, the absolute quantity of O_2 transported per unit volume of blood is diminished. As the anemic blood passes through the capillaries and the usual quantity of O_2 is removed from it, the $P{O_2}$ and saturation in the venous blood decline to a greater extent than normal.

Carbon Monoxide (CO) Intoxication **(See also Chap. 476)** Hemoglobin that binds CO (carboxy-hemoglobin [COHb]) is unavailable for O_2 transport. In addition, the presence of COHb shifts the Hb-O_2 dissociation curve to the left **(see Fig. 103-2)** so that O_2 is unloaded only at lower tensions, further contributing to tissue hypoxia.

Circulatory Hypoxia As in anemic hypoxia, the $Pa{O_2}$ is usually normal, but venous and tissue $P{O_2}$ values are reduced as a consequence of reduced tissue perfusion and greater tissue O_2 extraction. This pathophysiology leads to an increased arterial-mixed venous O_2 difference (a-v-O_2 difference), or gradient. Generalized circulatory hypoxia occurs in heart failure **(Chap. 264)** and in most forms of shock **(Chap. 314)**.

Specific Organ Hypoxia Localized circulatory hypoxia may occur as a result of decreased perfusion secondary to arterial obstruction, as in localized atherosclerosis in any vascular bed, or as a consequence of vasoconstriction, as observed in Raynaud's phenomenon **(Chap. 292)**. Localized hypoxia may also result from venous obstruction and the resultant expansion of interstitial fluid causing arteriolar compression and, thereby, reduction of arterial inflow. Edema, which increases the distance through which O_2 must diffuse before it reaches cells, can also cause localized hypoxia. In an attempt to maintain adequate perfusion to more vital organs in patients with reduced cardiac output secondary to heart failure or hypovolemic shock, vasoconstriction may reduce perfusion in the limbs and skin, causing hypoxia of these regions.

Increased O_2 Requirements If the O_2 consumption of tissues is elevated without a corresponding increase in perfusion, tissue hypoxia ensues and the $P{O_2}$ in venous blood declines. Ordinarily, the clinical picture of patients with hypoxia due to an elevated metabolic rate, as in fever or thyrotoxicosis, is quite different from that in other types of hypoxia: the skin is warm and flushed owing to increased cutaneous blood flow that dissipates the excessive heat produced, and cyanosis is usually absent.

Exercise is a classic example of increased tissue O_2 requirements. These increased demands are normally met by several mechanisms operating simultaneously: (1) an increase in the cardiac output and ventilation and, thus, O_2 delivery to the tissues; (2) a preferential shift in blood flow to the exercising muscles by changing vascular resistances in the circulatory beds of exercising tissues, directly and/or reflexly; (3) an increase in O_2 extraction from the delivered blood and a widening of the arteriovenous O_2 difference; and (4) a reduction in the pH of the tissues and capillary blood, shifting the Hb-O_2 curve to the right **(see Fig. 103-2)** and thereby unloading more O_2 from hemoglobin. If the capacity of these mechanisms is exceeded, then hypoxia, especially of the exercising muscles, will result.

Improper Oxygen Utilization Cyanide **(Chap. 470)** and several other similarly acting poisons cause cellular hypoxia by impairing electron transport in mitochondria, thereby limiting oxidative phosphorylation and ATP production. The tissues are unable to use O_2, and as a consequence, the venous blood tends to have a high O_2 tension. This condition has been termed *histotoxic hypoxia*.

ADAPTATION TO HYPOXIA

An important component of the respiratory response to hypoxia originates in special chemosensitive cells in the carotid and aortic bodies and in the respiratory center in the brainstem. The stimulation of these cells by hypoxia increases ventilation, with a loss of CO_2, and can lead to respiratory alkalosis. When combined with the metabolic acidosis resulting from the production of lactic acid, the serum bicarbonate level declines **(Chap. 58)**.

With the reduction of Pao$_2$, cerebrovascular resistance decreases and cerebral blood flow increases in an attempt to maintain O$_2$ delivery to the brain. However, when the reduction of Pao$_2$ is accompanied by hyperventilation and a reduction of Paco$_2$, cerebrovascular resistance rises, cerebral blood flow falls, and tissue hypoxia intensifies.

The diffuse, systemic vasodilation that occurs in generalized hypoxia increases the cardiac output. In patients with underlying heart disease, the requirements of peripheral tissues for an increase of cardiac output with hypoxia may precipitate congestive heart failure. In patients with ischemic heart disease, a reduced Pao$_2$ may intensify myocardial ischemia and further impair left ventricular function.

One of the important compensatory mechanisms for chronic hypoxia is an increase in the hemoglobin concentration and in the number of red blood cells in the circulating blood, that is, the development of polycythemia induced by erythropoietin production **(Chap. 108)**. In persons with chronic hypoxemia secondary to prolonged residence at a high altitude (>13,000 ft, 4200 m), a condition termed *chronic mountain sickness* develops. This disorder is characterized by a blunted respiratory drive, reduced ventilation, erythrocytosis, cyanosis, weakness, right ventricular enlargement secondary to pulmonary hypertension, and even stupor.

CYANOSIS

Cyanosis refers to a bluish color of the skin and mucous membranes resulting from an increased quantity of reduced hemoglobin (i.e., deoxygenated hemoglobin) or of hemoglobin derivatives (e.g., methemoglobin or sulfhemoglobin) in the small blood vessels of those tissues. It is usually most marked in the lips, nail beds, ears, and malar eminences. Cyanosis, especially if developed recently, is more commonly detected by a family member than the patient. A cherry-colored flush, rather than cyanosis, is caused by COHb **(Chap. 470)**.

The degree of cyanosis is modified by the color of the cutaneous pigment and the thickness of the skin, as well as by the state of the cutaneous capillaries. The accurate clinical detection of the presence and degree of cyanosis is difficult, as proved by oximetric studies. In some instances, central cyanosis can be detected reliably when the Sao$_2$ has fallen to 85%; in others, particularly in dark-skinned persons, it may not be detected until it has declined to 75%. In the latter case, examination of the mucous membranes in the oral cavity and the conjunctivae rather than examination of the skin is more helpful in the detection of cyanosis.

The increase in the quantity of reduced hemoglobin in the mucocutaneous vessels that produces cyanosis may be brought about either by an increase in the quantity of venous blood as a result of dilation of the venules (including precapillary venules) or by a reduction in the Sao$_2$ in the capillary blood. In general, cyanosis becomes apparent when the concentration of reduced hemoglobin in capillary blood exceeds 40 g/L (4 g/dL).

It is the *absolute*, rather than the *relative*, quantity of reduced hemoglobin that is important in producing cyanosis. Thus, in a patient with severe anemia, the *relative* quantity of reduced hemoglobin in the venous blood may be very large when considered in relation to the total quantity of hemoglobin in the blood. However, since the concentration of the latter is markedly reduced, the *absolute* quantity of reduced hemoglobin may still be low, and therefore, patients with severe anemia and even *marked* arterial desaturation may not display cyanosis. Conversely, the higher the total hemoglobin content, the greater is the tendency toward cyanosis; thus, patients with marked polycythemia tend to be cyanotic at higher levels of Sao$_2$ than patients with normal hematocrit values. Likewise, local passive congestion, which causes an increase in the total quantity of reduced hemoglobin in the vessels in a given area, may cause cyanosis. Cyanosis is also observed when nonfunctional hemoglobin, such as methemoglobin (consequential or acquired) or sulfhemoglobin **(Chap. 103)**, is present in blood.

Cyanosis may be subdivided into central and peripheral types. In *central* cyanosis, the Sao$_2$ is reduced or an abnormal hemoglobin derivative is present, and the mucous membranes and skin are both affected. *Peripheral* cyanosis is due to a slowing of blood flow and abnormally great extraction of O$_2$ from normally saturated arterial blood; it results from vasoconstriction and diminished peripheral blood flow, such as occurs in cold exposure, shock, congestive failure, and peripheral vascular disease. Often in these conditions, the mucous membranes of the oral cavity, including the sublingual mucosa, may be spared. Clinical differentiation between central and peripheral cyanosis may not always be straightforward, and in conditions such as cardiogenic shock with pulmonary edema, there may be a mixture of both types.

DIFFERENTIAL DIAGNOSIS

Central Cyanosis (Table 42-1) Decreased Sao$_2$ results from a marked reduction in the Pao$_2$. This reduction may be brought about by a decline in the Fio$_2$ without sufficient compensatory alveolar hyperventilation to maintain alveolar Po$_2$. Cyanosis usually becomes manifest in an ascent to an altitude of 4000 m (13,000 ft).

Seriously *impaired pulmonary function*, through perfusion of unventilated or poorly ventilated areas of the lung or alveolar hypoventilation, is a common cause of central cyanosis **(Chap. 296)**. This condition may occur acutely, as in extensive pneumonia or pulmonary edema, or chronically, with chronic pulmonary diseases (e.g., emphysema). In the latter situation, secondary polycythemia is generally present and clubbing of the fingers (see below) may occur. Another cause of reduced Sao$_2$ is *shunting of systemic venous blood into the arterial circuit*. Certain forms of congenital heart disease are associated with cyanosis on this basis (see above and **Chap. 280**).

Pulmonary arteriovenous fistulae may be congenital or acquired, solitary or multiple, and microscopic or massive. The severity of cyanosis produced by these fistulae depends on their size and number. They occur with some frequency in hereditary hemorrhagic telangiectasia. Sao$_2$ reduction and cyanosis may also occur in some patients with cirrhosis, presumably as a consequence of pulmonary arteriovenous fistulae or portal vein–pulmonary vein anastomoses.

In patients with cardiac or pulmonary right-to-left shunts, the presence and severity of cyanosis depend on the size of the shunt relative to the systemic flow and on the Hb-O$_2$ saturation of the venous blood. With increased extraction of O$_2$ from the blood by the exercising muscles, the venous blood returning to the right side of the heart is more unsaturated than at rest, and shunting of this blood intensifies the cyanosis. Secondary polycythemia occurs frequently in patients in this setting and contributes to the cyanosis.

TABLE 42-1 Causes of Cyanosis

Central Cyanosis
Decreased arterial oxygen saturation
Decreased atmospheric pressure—high altitude
Impaired pulmonary function
Alveolar hypoventilation
Inhomogeneity in pulmonary ventilation and perfusion (perfusion of hypoventilated alveoli)
Impaired oxygen diffusion
Anatomic shunts
Certain types of congenital heart disease
Pulmonary arteriovenous fistulas
Multiple small intrapulmonary shunts
Hemoglobin with low affinity for oxygen
Hemoglobin abnormalities
Methemoglobinemia—hereditary, acquired
Sulfhemoglobinemia—acquired
Carboxyhemoglobinemia (not true cyanosis)
Peripheral Cyanosis
Reduced cardiac output
Cold exposure
Redistribution of blood flow from extremities
Arterial obstruction
Venous obstruction

Cyanosis can be caused by small quantities of circulating methemoglobin (Hb Fe^{3+}) and by even smaller quantities of sulfhemoglobin **(Chap. 103)**; both of these hemoglobin derivatives impair oxygen delivery to the tissues. Although they are uncommon causes of cyanosis, these abnormal hemoglobin species should be sought by spectroscopy when cyanosis is not readily explained by malfunction of the circulatory or respiratory systems. Generally, digital clubbing does not occur with these conditions.

Peripheral Cyanosis Probably the most common cause of peripheral cyanosis is the normal vasoconstriction resulting from exposure to cold air or water. When cardiac output is reduced, cutaneous vasoconstriction occurs as a compensatory mechanism so that blood is diverted from the skin to more vital areas such as the CNS and heart, and cyanosis of the extremities may result even though the arterial blood is normally saturated.

Arterial obstruction to an extremity, as with an embolus, or arteriolar constriction, as in cold-induced vasospasm (Raynaud's phenomenon) **(Chap. 292)**, generally results in pallor and coldness, and there may be associated cyanosis. Venous obstruction, as in thrombophlebitis or deep venous thrombosis, dilates the subpapillary venous plexuses and thereby intensifies cyanosis.

APPROACH TO THE PATIENT

Cyanosis

Certain features are important in arriving at the cause of cyanosis:

1. It is important to ascertain the time of onset of cyanosis. Cyanosis present since birth or infancy is usually due to congenital heart disease.
2. Central and peripheral cyanosis must be differentiated. Evidence of disorders of the respiratory or cardiovascular systems is helpful. Massage or gentle warming of a cyanotic extremity will increase peripheral blood flow and abolish peripheral, but not central, cyanosis.
3. The presence or absence of clubbing of the digits (see below) should be ascertained. The combination of cyanosis and clubbing is frequent in patients with congenital heart disease and right-to-left shunting and is seen occasionally in patients with pulmonary disease, such as lung abscess or pulmonary arteriovenous fistulae. In contrast, peripheral cyanosis or acutely developing central cyanosis is *not* associated with clubbed digits.
4. Pa_{O_2} and Sa_{O_2} should be determined, and in patients with cyanosis in whom the mechanism is obscure, spectroscopic examination of the blood should be performed to look for abnormal types of hemoglobin (critical in the differential diagnosis of cyanosis).

CLUBBING

The selective bulbous enlargement of the distal segments of the fingers and toes due to proliferation of connective tissue, particularly on the dorsal surface, is termed *clubbing*; there is also increased sponginess of the soft tissue at the base of the clubbed nail. Clubbing may be hereditary, idiopathic, or acquired and associated with a variety of disorders, including cyanotic congenital heart disease (see above), infective endocarditis, and a variety of pulmonary conditions (among them primary and metastatic lung cancer, bronchiectasis, asbestosis, sarcoidosis, lung abscess, cystic fibrosis, tuberculosis, and mesothelioma), as well as with some gastrointestinal diseases (including inflammatory bowel disease and hepatic cirrhosis). In some instances, it is occupational, for example, in jackhammer operators.

Clubbing in patients with primary and metastatic lung cancer, mesothelioma, bronchiectasis, or hepatic cirrhosis may be associated with *hypertrophic osteoarthropathy*. In this condition, the subperiosteal formation of new bone in the distal diaphyses of the long bones of the extremities causes pain and symmetric arthritis-like changes in the shoulders, knees, ankles, wrists, and elbows. The diagnosis of hypertrophic osteoarthropathy may be confirmed by bone radiograph or magnetic resonance imaging (MRI). Although the mechanism of clubbing is unclear, it appears to be secondary to humoral substances that cause dilation of the vessels of the distal digits as well as growth factors released from platelet precursors in the digital circulation. In certain circumstances, clubbing is reversible, such as following lung transplantation for cystic fibrosis.

■ FURTHER READING

Essouma M et al: Epidemiology of digital clubbing and hypertrophic osteoarthropathy: A systematic review and meta-analysis. J Clin Rheumatol 28:104, 2022.

Gao H et al: Acquired methemoglobinemia: A systematic review of reported cases. Transfus Apher Sci 61:103299, 2022.

MacIntyre NR: Tissue hypoxia: Implications for the respiratory clinician. Respir Care 59:1590, 2014.

43

Edema

Joseph Loscalzo

PLASMA AND INTERSTITIAL FLUID EXCHANGE

Approximately two-thirds of total body water is intracellular and one-third is extracellular. One-fourth of the latter is in the plasma, and the remainder comprises the interstitial fluid. Edema represents an excess of interstitial fluid that has become evident clinically.

There is constant interchange of fluid between the two compartments of the extracellular fluid. The hydrostatic pressure within the capillaries and the colloid oncotic pressure in the interstitial fluid promote the movement of water and diffusible solutes from plasma to the interstitium. This movement is most prominent at the arterial origin of the capillary and falls progressively with the decline in intracapillary pressure and the rise in oncotic pressure toward the venular end. Fluid is returned from the interstitial space into the vascular system largely through the lymphatic system. These interchanges of fluids are normally balanced so that the volumes of the intravascular and interstitial compartments remain constant. However, a net movement of fluid from the intravascular to the interstitial spaces takes place and may be responsible for the development of edema under the following conditions: (1) an increase in intracapillary hydrostatic pressure; (2) inadequate lymphatic drainage; (3) reductions in the oncotic pressure of plasma; (4) damage to or dysfunction of the capillary endothelial barrier; and (5) increases in the oncotic pressure of the interstitial space.

■ REDUCTION OF EFFECTIVE ARTERIAL VOLUME

In many forms of edema, the effective arterial blood volume, a parameter that represents the filling of the arterial tree and that effectively perfuses the tissues, is reduced. Underfilling of the arterial tree may be caused by a reduction of cardiac output and/or systemic vascular resistance, by the pooling of blood in the splanchnic veins (as in cirrhosis), and by hypoalbuminemia **(Fig. 43-1*A*)**. As a consequence of this underfilling, a series of physiological responses designed to restore the effective arterial volume to normal are set into motion. A key element of these responses is the renal retention of sodium and, therefore, water, thereby restoring effective arterial volume, but sometimes also leading to the development or intensification of edema.

■ RENAL FACTORS AND THE RENIN-ANGIOTENSIN-ALDOSTERONE SYSTEM

The diminished renal blood flow characteristic of states in which the effective arterial blood volume is reduced is translated by the renal

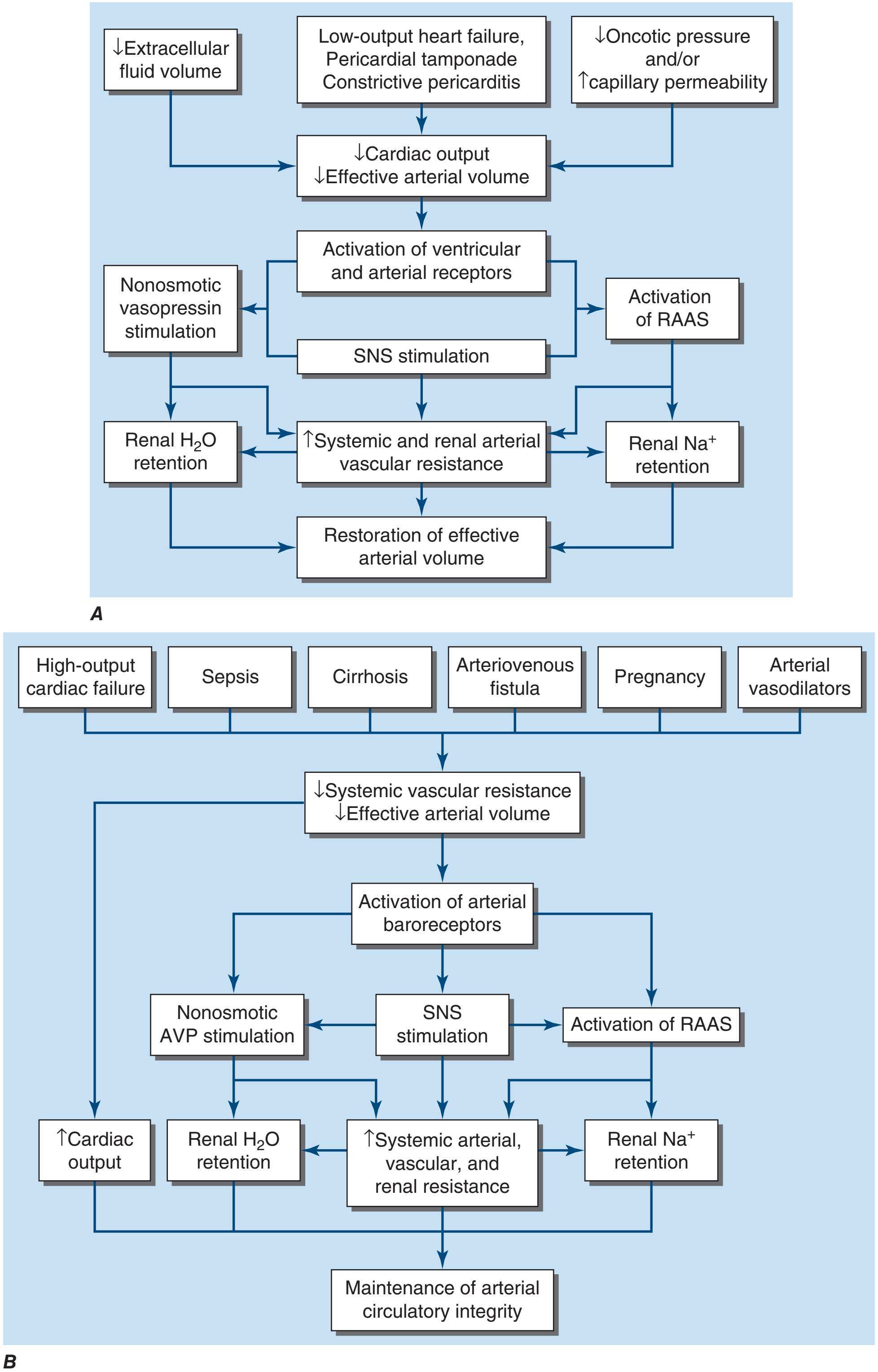
↓Extracellular fluid volume
Low-output heart failure, Pericardial tamponade Constrictive pericarditis
↓Oncotic pressure and/or ↑capillary permeability
↓Cardiac output ↓Effective arterial volume
Activation of ventricular and arterial receptors
Nonosmotic vasopressin stimulation
Activation of RAAS
SNS stimulation
Renal H_2O retention
↑Systemic and renal arterial vascular resistance
Renal Na^+ retention
Restoration of effective arterial volume
A
High-output cardiac failure
Sepsis
Cirrhosis
Arteriovenous fistula
Pregnancy
Arterial vasodilators
↓Systemic vascular resistance ↓Effective arterial volume
Activation of arterial baroreceptors
Nonosmotic AVP stimulation
SNS stimulation
Activation of RAAS
↑Cardiac output
Renal H_2O retention
↑Systemic arterial, vascular, and renal resistance
Renal Na^+ retention
Maintenance of arterial circulatory integrity
B

FIGURE 43-1 Clinical conditions in which a decrease in cardiac output (*A*) and systemic vascular resistance (*B*) cause arterial underfilling with resulting neurohumoral activation and renal sodium and water retention. In addition to activating the neurohumoral axis, adrenergic stimulation causes renal vasoconstriction and enhances sodium and fluid transport by the proximal tubule epithelium. AVP, arginine vasopressin; RAAS, renin-angiotensin aldosterone system; SNS, sympathetic nervous system. *(From Annals of Internal Medicine, RW Schrier: Body fluid volume regulation in health and disease: A unifying hypothesis. 113(2):155-159, 1990. Copyright © 1990, American College of Physicians. All Rights Reserved. Reprinted with the permission of American College of Physicians, Inc.)*

juxtaglomerular cells (specialized myoepithelial cells surrounding the afferent arteriole) into a signal for increased renin release. Renin is an enzyme with a molecular mass of about 40,000 Da that acts on its substrate, angiotensinogen, an α_2-globulin synthesized by the liver, to release angiotensin I, a decapeptide, which in turn is converted to angiotensin II (AII), an octapeptide. AII has generalized vasoconstrictor properties, particularly on the renal efferent arterioles. This action reduces the hydrostatic pressure in the peritubular capillaries, whereas the increased filtration fraction raises the colloid osmotic pressure in these vessels, thereby enhancing salt and water reabsorption in the proximal tubule as well as in the ascending limb of the loop of Henle.

The renin-angiotensin-aldosterone system (RAAS) operates as both a hormonal and paracrine system. Its activation causes sodium and water retention and thereby contributes to edema formation. Blockade of the conversion of angiotensin I to AII and blockade of the AII receptors enhance sodium and water excretion and reduce many forms of edema. AII that enters the systemic circulation stimulates the production of aldosterone by the zona glomerulosa of the adrenal cortex. Aldosterone in turn enhances sodium reabsorption (and potassium excretion) by the collecting tubule, further favoring edema formation. Blockade of the action of aldosterone by spironolactone or eplerenone (aldosterone antagonists) or by amiloride (a blocker of epithelial sodium channels) often induces a moderate diuresis in edematous states.

ARGININE VASOPRESSIN

(See also Chap. 393) The secretion of arginine vasopressin (AVP) by the posterior pituitary gland occurs in response to increased intracellular osmolar concentration; by stimulating V_2 receptors, AVP increases the reabsorption of free water in the distal tubules and collecting ducts of the kidney, thereby increasing total body water. Circulating AVP is elevated in many patients with heart failure secondary to a nonosmotic stimulus associated with decreased effective arterial volume and reduced compliance of the left atrium. Such patients fail to show the normal reduction of AVP with a reduction of osmolality, contributing to edema formation and hyponatremia.

ENDOTHELIN-1

This potent peptide vasoconstrictor is released by endothelial cells. Its concentration in the plasma is elevated in patients with severe heart failure and contributes to renal vasoconstriction, sodium retention, and edema.

NATRIURETIC PEPTIDES

Atrial distention causes release into the circulation of atrial natriuretic peptide (ANP), a 3000-Da polypeptide. A high-molecular-weight precursor of ANP is stored in secretory granules within atrial myocytes. A closely related natriuretic peptide (pre-prohormone brain natriuretic peptide [BNP]) is stored primarily in ventricular myocytes and is released when ventricular diastolic pressure rises. Released ANP and BNP (which is derived from its precursor) bind to the natriuretic receptor-A, which causes (1) excretion of sodium and water by augmenting glomerular filtration, inhibiting sodium reabsorption in the proximal tubule, and inhibiting release of renin and aldosterone; and (2) dilation of arterioles and venules by antagonizing the vasoconstrictor actions of AII, AVP, and sympathetic stimulation. Thus, elevated levels of natriuretic peptides have the capacity to oppose sodium retention in hypervolemic and edematous states.

Although circulating levels of ANP and BNP are elevated in heart failure and in cirrhosis with ascites, these natriuretic peptides are not sufficiently potent to prevent edema formation. Indeed, in edematous states, resistance to the actions of natriuretic peptides may be increased, further reducing their effectiveness.

Further discussion of the control of sodium and water balance is found in **Chap. S1**.

CLINICAL CAUSES OF EDEMA

A weight gain of several kilograms usually precedes overt manifestations of generalized edema. *Anasarca* refers to gross, generalized edema. *Ascites* **(Chap. 53)** and *hydrothorax* refer to accumulation of excess fluid in the peritoneal and pleural cavities, respectively, and are considered special forms of edema.

Edema is recognized by the persistence of an indentation of the skin after pressure, known as "pitting" edema. In its more subtle form, edema may be detected by noting that after the stethoscope is removed from the chest wall, the rim of the bell leaves an indentation on the skin of the chest for a few minutes. Edema may be present when the ring on a finger fits more snugly than in the past or when a patient complains of difficulty putting on shoes, particularly in the evening when dependent edema is greatest. Edema may also be recognized by puffiness of the face (most prominent upon awakening), which is most readily apparent in the periorbital areas owing to relative tissue laxity.

GENERALIZED EDEMA

The differences among the major causes of generalized edema are shown in **Table 43-1**. Cardiac, renal, hepatic, or nutritional disorders are responsible for a large majority of patients with generalized edema. Consequently, the differential diagnosis of generalized edema should be directed toward identifying or excluding these several conditions.

Heart Failure **(See also Chap. 264)** In heart failure, the impaired systolic emptying of the ventricle(s) and/or the impairment of ventricular relaxation promotes an accumulation of blood in the venous circulation at the expense of the effective arterial volume. In addition, activation of the sympathetic nervous system and the RAAS (see above) acts in concert to cause renal vasoconstriction and reduction of glomerular filtration and salt and water retention. Sodium and water retention continue, and the increment in blood volume accumulates in the venous circulation, raising venous and intracapillary pressure and resulting in edema (Fig. 43-1).

The presence of overt cardiac disease, as manifested by cardiac enlargement and/or ventricular hypertrophy, together with clinical evidence of cardiac failure, such as dyspnea, basilar rales, venous distention, and hepatomegaly, usually indicates that edema results from heart failure. Noninvasive tests such as electrocardiography, echocardiography, and measurements of BNP (or N-terminal proBNP [NT-proBNP]) are helpful in establishing the diagnosis of heart disease. The edema of heart failure typically occurs in the dependent portions of the body.

Edema of Renal Disease **(See also Chap. 326)** The edema that occurs during the acute phase of glomerulonephritis is characteristically associated with hematuria, proteinuria, and hypertension. In most instances, the edema results from primary retention of sodium and water by the kidney owing to renal dysfunction. This state differs from most forms of heart failure in that it is characterized by a normal (or sometimes even increased) cardiac output. Patients with *chronic* renal failure may also develop edema due to primary renal retention of sodium and water.

Nephrotic Syndrome and Other Hypoalbuminemic States The primary alteration in the nephrotic syndrome is a diminished colloid oncotic pressure due to losses of large quantities (≥3.5 g/d) of protein into the urine and hypoalbuminemia (<3.0 g/dL). As a result of the reduced colloid osmotic pressure, the sodium and water that are retained cannot be confined within the vascular compartment, and total and effective arterial blood volumes decline. This process initiates the edema-forming sequence of events described above, including activation of the RAAS. The nephrotic syndrome may occur during the course of a variety of kidney diseases, including glomerulonephritis, diabetic glomerulosclerosis, and hypersensitivity reactions. The edema is diffuse, symmetric, and most prominent in the dependent areas; periorbital edema is most prominent in the morning.

Hepatic Cirrhosis **(See also Chap. 355)** This condition is characterized, in part, by hepatic venous outflow obstruction, which in turn expands the splanchnic blood volume, and hepatic lymph formation. Intrahepatic hypertension acts as a stimulus for renal sodium retention

TABLE 43-1 Principal Causes of Generalized Edema: History, Physical Examination, and Laboratory Findings

ORGAN SYSTEM	HISTORY	PHYSICAL EXAMINATION	LABORATORY FINDINGS
Cardiac	Dyspnea with exertion prominent—often associated with orthopnea—or paroxysmal nocturnal dyspnea	Elevated jugular venous pressure, ventricular (S_3) gallop; occasionally with displaced or dyskinetic apical impulse; peripheral cyanosis, cool extremities, low pulse pressure when severe	Elevated urea nitrogen-to-creatinine ratio common; serum sodium often decreased; elevated natriuretic peptides
Hepatic	Dyspnea uncommon, except if associated with significant degree of ascites; most often a history of ethanol abuse	Frequently associated with ascites; jugular venous pressure normal or low; blood pressure lower than in renal or cardiac disease; one or more additional signs of chronic liver disease (jaundice, palmar erythema, Dupuytren's contracture, spider angiomata, male gynecomastia; asterixis and other signs of encephalopathy) may be present	If severe, reductions in serum albumin, cholesterol, other hepatic proteins (transferrin, fibrinogen); liver enzymes elevated, depending on the cause and acuity of liver injury; tendency toward hypokalemia, respiratory alkalosis; macrocytosis from folate deficiency
Renal (CRF)	Usually chronic: may be associated with uremic signs and symptoms, including decreased appetite, altered (metallic or fishy) taste, altered sleep pattern, difficulty concentrating, restless legs, or myoclonus; dyspnea can be present, but generally less prominent than in heart failure	Elevated blood pressure; hypertensive retinopathy; uremic fetor; pericardial friction rub in advanced cases with uremia	Elevation of serum creatinine and cystatin C; albuminuria; hyperkalemia, metabolic acidosis, hyperphosphatemia, hypocalcemia, anemia (usually normocytic)
Renal (NS)	Childhood diabetes mellitus; plasma cell dyscrasias	Periorbital edema; hypertension	Proteinuria (≥3.5 g/d); hypoalbuminemia; hypercholesterolemia; microscopic hematuria

Abbreviations: CRF, chronic renal failure; NS, nephrotic syndrome.

Source: Reproduced with permission from GM Chertow, E Braunwald, L Goldman (eds). Approach to the patient with edema, in Primary Cardiology, 2nd ed. Philadelphia, Saunders, 2003.

and causes a reduction of effective arterial blood volume. These alterations are frequently complicated by hypoalbuminemia secondary to reduced hepatic synthesis, as well as peripheral arterial vasodilation. These effects reduce the effective arterial blood volume, leading to activation of the sodium- and water-retaining mechanisms described above **(Fig. 43-1*B*)**. The concentration of circulating aldosterone often is elevated by the failure of the liver to metabolize this hormone. Initially, the excess interstitial fluid is localized preferentially proximal (upstream) to the congested portal venous system, causing ascites **(Chap. 53)**. In later stages, particularly when there is severe hypoalbuminemia, peripheral edema may develop. A sizable accumulation of ascitic fluid may increase intraabdominal pressure and impede venous return from the lower extremities, thereby contributing to the accumulation of the edema.

Drug-Induced Edema A large number of widely used drugs can cause edema **(Table 43-2)**. Mechanisms include renal vasoconstriction (nonsteroidal anti-inflammatory drugs and cyclosporine), arteriolar dilation (vasodilators), augmented renal sodium reabsorption (steroid hormones), and capillary damage.

Edema of Nutritional Origin A diet grossly deficient in calories and particularly in protein over a prolonged period may produce hypoproteinemia and edema. The latter may be intensified by the development of beriberi heart disease, which also is of nutritional origin, in which multiple peripheral arteriovenous fistulae result in reduced effective systemic perfusion and effective arterial blood volume, thereby enhancing edema formation **(Chap. 344)** (Fig. 43-1*B*). Edema develops or becomes intensified when famished subjects are first provided with an adequate diet. The ingestion of more food may increase the quantity of sodium ingested, which is then retained along with water. So-called refeeding edema also may be linked to increased release of insulin, which directly increases tubular sodium reabsorption. In addition to hypoalbuminemia, hypokalemia and caloric deficits may be involved in the edema of starvation.

LOCALIZED EDEMA

In thrombophlebitis, varicose veins, and primary venous valve failure, the hydrostatic pressure in the capillary bed upstream (proximal) of the obstruction increases so that an abnormal quantity of fluid is transferred from the vascular to the interstitial space, which may give rise to localized edema. The latter may also occur in lymphatic obstruction caused by chronic lymphangitis, resection of regional lymph nodes, filariasis, and genetic (frequently called primary) lymphedema. The latter is particularly intractable because restriction of lymphatic flow results in both an increase in intracapillary pressure and increased protein concentration in the interstitial fluid, which act in concert to aggravate fluid retention.

TABLE 43-2 Drugs Associated with Edema Formation

Nonsteroidal anti-inflammatory drugs
Antihypertensive agents
Direct arterial/arteriolar vasodilators
Hydralazine
Clonidine
Methyldopa
Guanethidine
Minoxidil
Calcium channel antagonists
α-Adrenergic antagonists
Thiazolidinediones
Steroid hormones
Glucocorticoids
Anabolic steroids
Estrogens
Progestins
Cyclosporine
Growth hormone
Immunotherapies
Interleukin 2
OKT3 monoclonal antibody

Source: Reproduced with permission from GM Chertow, in E Braunwald, L Goldman (eds): Approach to the patient with edema, in Primary Cardiology, 2nd ed. Philadelphia, Saunders, 2003.

Other Causes of Edema These causes include hypothyroidism (myxedema) due to tissue deposition of hyaluronic acid; hyperthyroidism (pretibial myxedema secondary to Graves' disease), in which edema is typically nonpitting; hypercortisolism; pregnancy; and administration of estrogens and vasodilators, particularly calcium channel blockers.

DISTRIBUTION OF EDEMA

The distribution of edema is an important guide to its cause. Edema associated with heart failure tends to be more extensive in the legs and to be accentuated in the evening, a feature also determined largely by posture. When patients with heart failure are confined to bed, edema may be most prominent in the presacral region.

Edema resulting from hypoproteinemia, as occurs in the nephrotic syndrome, characteristically is generalized, but it is especially evident in the very soft tissues of the eyelids and face and tends to be most pronounced in the morning owing to the recumbent posture assumed during the night. Less common causes of facial edema include trichinosis, allergic reactions, and myxedema. Edema limited to one leg or to one or both arms is usually the result of venous and/or lymphatic obstruction. Unilateral paralysis reduces lymphatic and venous drainage on the affected side and may also be responsible for unilateral edema. In patients with obstruction of the superior vena cava, edema is confined to the face, neck, and upper extremities in which the venous pressure is elevated compared with that in the lower extremities.

APPROACH TO THE PATIENT

Edema

An important first question is whether the edema is localized or generalized. If it is localized, the local phenomena that may be responsible should be identified. If the edema is generalized, one should determine if there is serious hypoalbuminemia, e.g., serum albumin <3.0 g/dL. If so, the history, physical examination, urinalysis, and other laboratory data will help evaluate the question of cirrhosis, severe malnutrition, or the nephrotic syndrome as the underlying disorder. If hypoalbuminemia is not present, one should determine if there is evidence of heart failure severe enough to promote generalized edema. Finally, it should be ascertained as to whether or not the patient has an adequate urine output or if there is significant oliguria or anuria. **These abnormalities are discussed in Chaps. 55, 321, and 322.**

FURTHER READING

Clark AL, Cleland JG: Causes and treatment of oedema in patients with heart failure. Nature Rev Cardiol 10:156, 2013.

Frison S et al: Omitting edema measurement: How much acute malnutrition are we missing? Am J Clin Nutr 102:1176, 2015.

Koirala A et al: Etiology and management of edema: A review. Adv Kidney Dis Health 30:110, 2023.

Levick JR, Michel CC: Microvascular fluid exchange and the revised Starling principle. Cardiovascular Res 87:198, 2010.

Telinius N, Hjortdal VE: Role of the lymphatic vasculature in cardiovascular medicine. Heart 105:1777, 2019.

44 Approach to the Patient with a Heart Murmur

Patrick T. O'Gara, Joseph Loscalzo

The differential diagnosis of a heart murmur begins with a careful assessment of its major attributes and response to bedside maneuvers. The history, clinical context, and associated physical examination findings provide additional clues to help establish its significance. Accurate bedside identification of a heart murmur can inform decisions regarding the indications for noninvasive testing and the need for referral to a cardiovascular specialist. Preliminary discussions can be held with the patient regarding natural history, antibiotic use for infective endocarditis or rheumatic fever prophylaxis, the need to restrict various forms of physical activity, and the potential role for family screening.

Heart murmurs are caused by audible vibrations that are due to increased turbulence from accelerated blood flow through normal or abnormal orifices; flow through a narrowed or irregular orifice into a dilated vessel or chamber; or backward flow through an incompetent valve, ventricular septal defect, or patent ductus arteriosus. They traditionally are defined by their timing within the cardiac cycle **(Fig. 44-1)**. *Systolic murmurs* begin with or after the first heart sound (S_1) and terminate at or before the component (A_2 or P_2) of the second heart sound (S_2) that corresponds to their site of origin (left or right, respectively). *Diastolic murmurs* begin with or after the associated component of S_2 and end at or before the subsequent S_1. *Continuous murmurs* are not confined to either phase of the cardiac cycle but instead begin in early systole and proceed through S_2 into all or part of diastole. The accurate timing of heart murmurs is the first step in their identification. The distinction between S_1 and S_2, and therefore systole and diastole, is usually a straightforward process but can be difficult in the setting of a tachyarrhythmia, in which case the heart sounds can be distinguished by simultaneous palpation of the carotid upstroke, which should closely follow S_1.

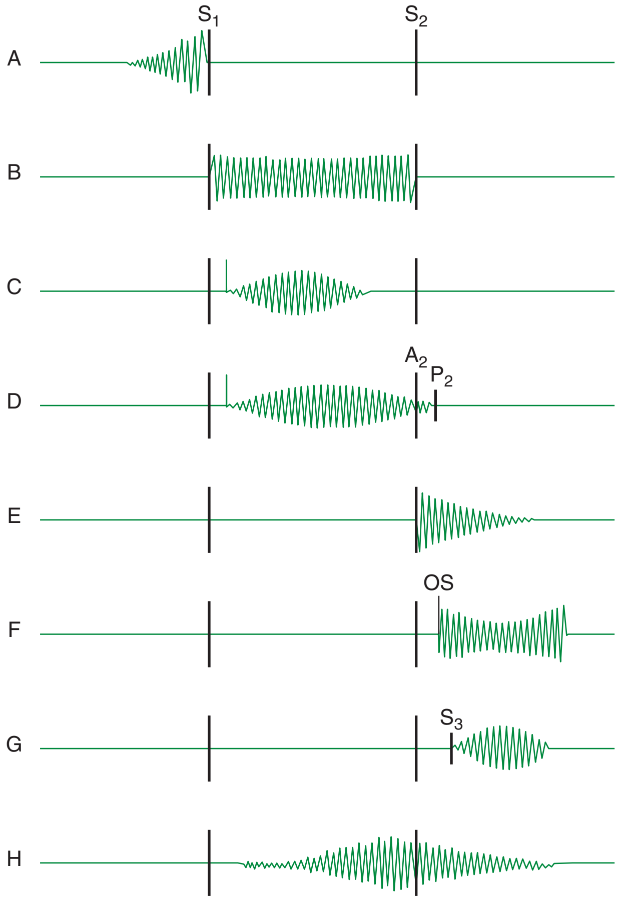

FIGURE 44-1 Diagram depicting principal heart murmurs. ***A.*** Presystolic murmur of mitral or tricuspid stenosis in a patient with sinus rhythm. ***B.*** Holosystolic (pansystolic) murmur of mitral or tricuspid regurgitation or of ventricular septal defect. ***C.*** Aortic ejection murmur beginning with an ejection click and fading before the second heart sound. ***D.*** Systolic murmur in pulmonic stenosis spilling through the aortic second sound, pulmonic valve closure being delayed. ***E.*** Aortic or pulmonary diastolic murmur. ***F.*** Long diastolic murmur of mitral stenosis after the opening snap (OS). ***G.*** Short mid-diastolic inflow murmur after a third heart sound. ***H.*** Continuous murmur of patent ductus arteriosus. *(Courtesy of Antony and Julie Wood.)*

Duration and Character The duration of a heart murmur depends on the length of time over which a pressure difference exists between two cardiac chambers, the left ventricle and the aorta, the right ventricle and the pulmonary artery, or the great vessels. The magnitude and variability of this pressure difference, coupled with the geometry and compliance of the involved chambers or vessels, dictate the velocity of flow; the degree of turbulence; and the resulting frequency, configuration, and intensity of the murmur. The diastolic murmur of chronic aortic regurgitation (AR) is a blowing, high-frequency event, whereas the murmur of mitral stenosis (MS), indicative of the left atrial–left ventricular diastolic pressure gradient, is a low-frequency event, heard as a rumbling sound with the bell of the stethoscope. The frequency components of a heart murmur may vary at different sites of auscultation. The coarse systolic murmur of aortic stenosis (AS) may sound higher pitched and more acoustically pure at the apex, a phenomenon eponymously referred to as the *Gallavardin effect*. Some murmurs may have a distinct or unusual quality, such as the "honking" sound appreciated in some patients with mitral regurgitation (MR) due to mitral valve prolapse (MVP).

The configuration of a heart murmur may be described as crescendo, decrescendo, crescendo-decrescendo, or plateau. The decrescendo configuration of the murmur of chronic AR **(Fig. 44-1*E*)** can be understood in terms of the progressive decline in the diastolic pressure gradient between the aorta and the left ventricle. The crescendo-decrescendo configuration of the murmur of AS reflects the changes in the systolic pressure gradient between the left ventricle and the aorta as ejection occurs, whereas the plateau configuration of the murmur of chronic MR **(Fig. 44-1*B*)** is consistent with the large and nearly constant pressure difference between the left ventricle and the left atrium.

Intensity The intensity of a heart murmur is graded on a scale of 1–6 (or I–VI). A grade 1 murmur is very soft and is heard only with great effort. A grade 2 murmur is easily heard but not particularly loud. A grade 3 murmur is loud but is not accompanied by a palpable thrill over the site of maximal intensity. A grade 4 murmur is very loud and accompanied by a thrill. A grade 5 murmur is loud enough to be heard with only the edge of the stethoscope touching the chest, whereas a grade 6 murmur is loud enough to be heard with the stethoscope slightly off the chest. Murmurs of grade 3 or greater intensity usually signify important structural heart disease and indicate high blood flow velocity at the site of murmur production. Small, restrictive ventricular septal defects (VSDs), for example, are accompanied by loud, usually grade 4 or greater, systolic murmurs as blood is ejected at high velocity from the left ventricle to the right ventricle through a narrow passage. Low-velocity events, such as left-to-right shunting across an atrial septal defect (ASD), are usually silent. The intensity of a heart murmur may be diminished by any process that increases the distance between the intracardiac source and the stethoscope on the chest wall, such as obesity, obstructive lung disease, or a large pericardial effusion. The intensity of a murmur also may be misleadingly soft when cardiac output is reduced significantly or when the pressure gradient between the involved cardiac structures is low.

Location and Radiation Recognition of the location and radiation of the murmur helps facilitate its accurate identification **(Fig. 44-2)**. Adventitious sounds, such as a systolic click or diastolic snap, or abnormalities of S_1 or S_2 may provide additional clues. Careful attention to the characteristics of the murmur and other heart sounds during the respiratory cycle and the performance of simple bedside maneuvers

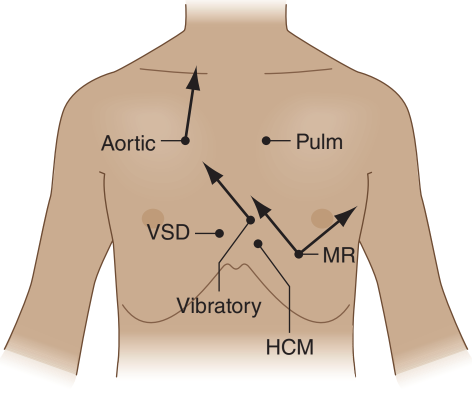

FIGURE 44-2 Maximal intensity and radiation of six isolated systolic murmurs. Aortic, aortic stenosis; HCM, hypertrophic obstructive cardiomyopathy; MR, mitral regurgitation; Pulm, pulmonic stenosis; VSD, ventricular septal defect. *(From JB Barlow: Perspectives on the Mitral Valve. Philadelphia, FA Davis, 1987, p 140.)*

complete the auscultatory examination. These features, along with recommendations for further testing, are discussed below in the context of specific systolic, diastolic, and continuous heart murmurs **(Table 44-1)**.

SYSTOLIC HEART MURMURS

Early Systolic Murmurs Early systolic murmurs begin with S$_1$ and extend for a variable period, ending well before S$_2$. Their causes are relatively few. *Acute, severe MR* into a normal-sized, relatively noncompliant left atrium results in an early, decrescendo systolic murmur best heard at or just medial to the apical impulse. These characteristics reflect the progressive attenuation of the pressure gradient between the left ventricle and the left atrium during systole owing to the rapid rise in left atrial pressure caused by the sudden volume load into an unprepared, noncompliant chamber, and contrast sharply with the auscultatory features of chronic MR. Clinical settings in which acute, severe MR occur include (1) papillary muscle rupture complicating acute myocardial infarction (MI) **(Chap. 286)**, (2) rupture of chordae tendineae in the setting of myxomatous mitral valve disease (MVP, **Chap. 276**), (3) infective endocarditis **(Chap. 133)**, and (4) blunt chest wall trauma.

Acute, severe MR from papillary muscle rupture usually accompanies an inferior, posterior, or lateral MI and occurs 2–7 days after presentation. It often is signaled by chest pain, hypotension, and pulmonary edema, but a murmur may be absent in up to 50% of cases. The posteromedial papillary muscle is involved 6–10 times more frequently than the anterolateral papillary muscle. The murmur is to be distinguished from that associated with post-MI ventricular septal rupture, which is accompanied by a systolic thrill at the left sternal border in nearly all patients and is holosystolic in duration. A new heart murmur after an MI is an indication for transthoracic echocardiography (TTE) **(Chap. 248)**, which allows bedside delineation of its etiology and pathophysiologic significance. The distinction between acute MR and ventricular septal rupture also can be achieved with right-sided heart catheterization, sequential determination of oxygen saturations, and analysis of the pressure waveforms (tall *v* wave in the pulmonary artery wedge pressure in MR). Post-MI mechanical complications of this nature mandate aggressive medical stabilization and prompt referral for surgical repair.

Spontaneous chordal rupture can complicate the course of myxomatous mitral valve disease (MVP) and result in new-onset or "acute on chronic" severe MR. MVP may occur as an isolated phenomenon, or the lesion may be part of a more generalized connective tissue disorder as seen, for example, in patients with Marfan syndrome. Acute, severe MR as a consequence of infective endocarditis results from destruction of leaflet tissue, chordal rupture, or both. Blunt chest wall trauma is usually self-evident but may be disarmingly trivial; it can result in papillary muscle contusion and rupture, chordal detachment, or leaflet avulsion. TTE is indicated in all cases of suspected acute, severe MR to

TABLE 44-1 Principal Causes of Heart Murmurs

Systolic Murmurs

- Early systolic
  - Mitral
    - Acute MR
  - VSD
    - Muscular
    - Nonrestrictive with pulmonary hypertension
  - Tricuspid
    - TR with normal pulmonary artery pressure
- Midsystolic
  - Aortic
    - Obstructive
      - Supravalvular–supravalvular AS, coarctation of the aorta
      - Valvular–AS and aortic sclerosis
      - Subvalvular–discrete, tunnel or HOCM
    - Increased flow, hyperkinetic states, AR, complete heart block
    - Dilation of ascending aorta, atheroma, aortitis
  - Pulmonary
    - Obstructive
      - Supravalvular–pulmonary artery stenosis
      - Valvular–pulmonic valve stenosis
      - Subvalvular–infundibular stenosis (dynamic)
    - Increased flow, hyperkinetic states, left-to-right shunt (e.g., ASD)
    - Dilation of pulmonary artery
- Late systolic
  - Mitral
    - MVP, acute myocardial ischemia
  - Tricuspid
    - TVP
- Holosystolic
  - Atrioventricular valve regurgitation (MR, TR)
  - Left-to-right shunt at ventricular level (VSD)

Early Diastolic Murmurs

- AR
  - Valvular: congenital (bicuspid valve), rheumatic deformity, endocarditis, prolapse, trauma, post-valvotomy
  - Dilation of valve ring: aortic dissection, annuloaortic ectasia, medial degeneration, hypertension, ankylosing spondylitis
  - Widening of commissures: syphilis
- Pulmonic regurgitation
  - Valvular: post-valvotomy, endocarditis, rheumatic fever, carcinoid
  - Dilation of valve ring: pulmonary hypertension; Marfan syndrome
  - Congenital: isolated or associated with tetralogy of Fallot, VSD, pulmonic stenosis

Mid-Diastolic Murmurs

- Mitral
  - MS
  - Carey-Coombs murmur (mid-diastolic apical murmur in acute rheumatic fever)
  - Increased flow across nonstenotic mitral valve (e.g., MR, VSD, PDA, high-output states, and complete heart block)
- Tricuspid
  - Tricuspid stenosis
  - Increased flow across nonstenotic tricuspid valve (e.g., TR, ASD, and anomalous pulmonary venous return)
- Left and right atrial tumors (myxoma)
- Severe AR (Austin Flint murmur)

Continuous Murmurs

Patent ductus arteriosus	Proximal coronary artery stenosis
Coronary AV fistula	Mammary souffle of pregnancy
Ruptured sinus of Valsalva aneurysm	Pulmonary artery branch stenosis
Aortic septal defect	Bronchial collateral circulation
Cervical venous hum	Small (restrictive) ASD with MS
Anomalous left coronary artery	Intercostal AV fistula

Abbreviations: AR, aortic regurgitation; AS, aortic stenosis; ASD, atrial septal defect; AV, arteriovenous; HOCM, hypertrophic obstructive cardiomyopathy; MR, mitral regurgitation; MS, mitral stenosis; MVP, mitral valve prolapse; PDA, patent ductus arteriosus; TR, tricuspid regurgitation; TVP, tricuspid valve prolapse; VSD, ventricular septal defect.

Source: E Braunwald, JK Perloff, in D Zipes et al (eds): *Braunwald's Heart Disease*, 7th ed. Philadelphia, Elsevier, 2005; PJ Norton, RA O'Rourke, in E Braunwald, L Goldman (eds): *Primary Cardiology*, 2nd ed. Philadelphia, Elsevier, 2003.

define its mechanism and severity, delineate left ventricular size and systolic function, and provide an assessment of suitability for valve repair.

A congenital, small muscular VSD **(Chap. 280)** may be associated with an early systolic murmur. The defect closes progressively during septal contraction, and thus the murmur is confined to early systole. It is localized to the left sternal border (Fig. 44-2) and is usually of grade 4 or 5 intensity. Signs of pulmonary hypertension or left ventricular volume overload are absent. Anatomically large and uncorrected VSDs, which usually involve the membranous portion of the septum, may lead to pulmonary hypertension. The murmur associated with the left-to-right shunt, which earlier may have been holosystolic, becomes limited to the first portion of systole as the elevated pulmonary vascular resistance leads to an abrupt rise in right ventricular pressure and an attenuation of the interventricular pressure gradient during the remainder of the cardiac cycle. In such instances, signs of pulmonary hypertension (right ventricular lift, loud and single or closely split S_2) may predominate. The murmur is best heard along the left sternal border but is softer. Suspicion of a VSD is an indication for TTE.

Tricuspid regurgitation (TR) with normal pulmonary artery pressures, as may occur with infective endocarditis, may produce an early systolic murmur. The murmur is soft (grade 1 or 2), is best heard at the lower left sternal border, and may increase in intensity with inspiration (Carvallo's sign). Regurgitant *c-v* waves may be visible in the jugular venous pulse. TR in this setting is not associated with signs of right heart failure, such as ascites or lower extremity edema.

Midsystolic Murmurs Midsystolic murmurs begin at a short interval after S_1, end before S_2 **(Fig. 44-1*C*)** and are usually crescendo-decrescendo in configuration. AS is the most common cause of a midsystolic murmur in an adult. The murmur of AS is usually loudest to the right of the sternum in the second intercostal space (aortic area, Fig. 44-2) and radiates into the carotids. Transmission of the midsystolic murmur to the apex, where it becomes higher-pitched, is common (Gallavardin effect; see above).

Differentiation of this apical systolic murmur from MR can be difficult. The murmur of AS will increase in intensity or become louder, in the beat after a premature beat, whereas the murmur of MR will have constant intensity from beat to beat. The intensity of the AS murmur also varies directly with the cardiac output. With a normal cardiac output, a systolic thrill at the second right intercostal space and a grade 4 or higher murmur suggest severe AS. The murmur is softer in the setting of heart failure and low cardiac output. Other auscultatory findings of severe AS include a soft or absent A_2, paradoxical splitting of S_2, an apical S_4, and a late-peaking systolic murmur. In children, adolescents, and young adults with congenital valvular AS, an early ejection sound (click) is usually audible, more often along the left sternal border than at the base. Its presence signifies a flexible, noncalcified bicuspid valve (or one of its variants) and localizes the left ventricular outflow obstruction to the valvular (rather than sub- or supravalvular) level.

Assessment of the volume and rate of rise of the carotid pulse can provide additional information. A small and delayed upstroke (*parvus et tardus*) is consistent with severe AS. The carotid pulse examination is less discriminatory, however, in older patients with stiffened arteries. The electrocardiogram (ECG) shows signs of left ventricular hypertrophy (LVH) as the severity of the stenosis increases. TTE is indicated to assess the anatomic features of the aortic valve, the severity of the stenosis, left ventricular size, wall thickness and function, and the size and contour of the aortic root and proximal ascending aorta.

The obstructive form of hypertrophic cardiomyopathy (HOCM) is associated with a midsystolic murmur that is usually loudest along the left sternal border or between the left lower sternal border and the apex (**Chaps. 266–270**, Fig. 44-2). The murmur is produced by both dynamic left ventricular outflow tract obstruction and MR, and thus, its configuration is a hybrid between ejection and regurgitant phenomena. The intensity of the murmur may vary from beat to beat and after provocative maneuvers but usually does not exceed grade 3. The murmur classically will increase in intensity with maneuvers that result in increasing degrees of outflow tract obstruction, such as a reduction in preload or afterload (Valsalva, standing, vasodilators), or with an augmentation of contractility (inotropic stimulation). Maneuvers or medications that increase preload (squatting, passive leg raising, volume administration) or afterload (squatting, vasopressors) or that reduce contractility (β-adrenoreceptor blockers) decrease the intensity of the murmur. In rare patients, there may be reversed splitting of S_2. A sustained left ventricular apical impulse and an S_4 may be appreciated. In contrast to AS, the carotid upstroke is rapid and of normal volume. Rarely, it is bisferiens or bifid in contour **(see Fig. 246-2*D*)** due to midsystolic closure of the aortic valve. LVH is present on the ECG, and the diagnosis is confirmed by TTE. Although the systolic murmur associated with MVP behaves similarly to that due to HOCM in response to the Valsalva maneuver and to standing/squatting **(Fig. 44-3)**, these two lesions can be distinguished on the basis of their associated findings, such as the presence of LVH in HOCM or a nonejection click in MVP.

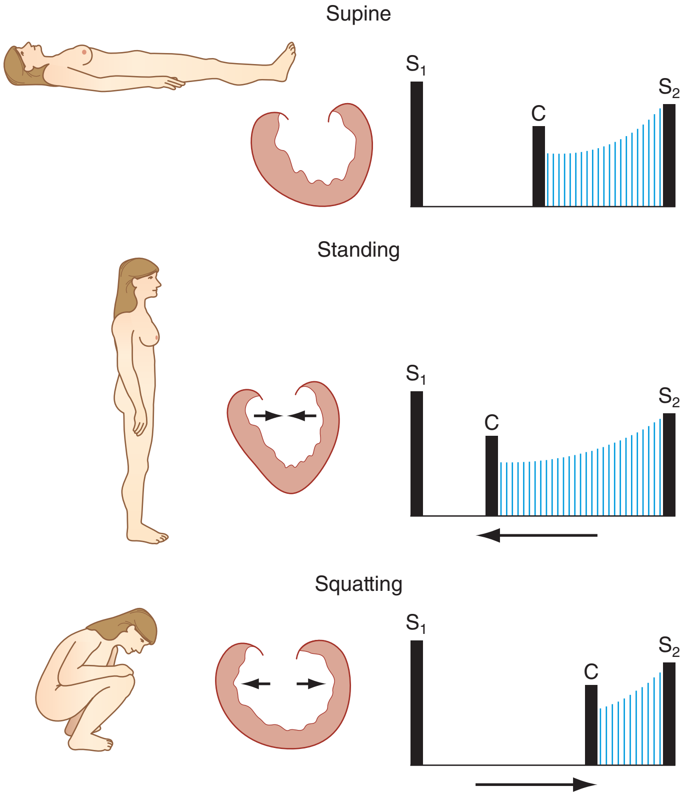

FIGURE 44-3 A midsystolic nonejection sound (C) occurs in mitral valve prolapse and is followed by a mid-late systolic murmur that crescendos to the second heart sound (S_2). Standing decreases venous return; the heart becomes smaller; C moves closer to the first heart sound (S_1), and the mitral regurgitant murmur has an earlier onset and longer duration. With prompt squatting, venous return and afterload increase; the heart becomes larger; C moves toward S_2; and the duration of the murmur shortens. The systolic murmur of hypertrophic obstructive cardiomyopathy behaves similarly. *(Reprinted with permission from Examination of the Heart, Part IV: Auscultation of the Heart © American Heart Association, Inc.)*

The midsystolic, crescendo-decrescendo murmur of congenital pulmonic stenosis (PS; **Chap. 280**) is best appreciated in the second and third left intercostal spaces (pulmonic area) (**Figs.** 44-2 and **44-4**). The duration of the murmur lengthens and the intensity of P_2 diminishes with increasing degrees of valvular stenosis **(Fig. 44-1*D*)**. An early ejection sound, the intensity of which *decreases* with inspiration, is heard in younger patients. A parasternal lift and ECG evidence of right ventricular hypertrophy indicate severe pressure overload. If obtained, the chest x-ray may show poststenotic dilation of the main pulmonary artery. TTE is recommended for complete characterization.

Significant left-to-right intracardiac shunting due to an ASD **(Chap. 280)** leads to an increase in pulmonary blood flow and a grade 2–3 midsystolic murmur at the middle to upper left sternal border attributed to increased flow rates across the pulmonic valve with fixed splitting of S_2. Ostium secundum ASDs are the most common cause of

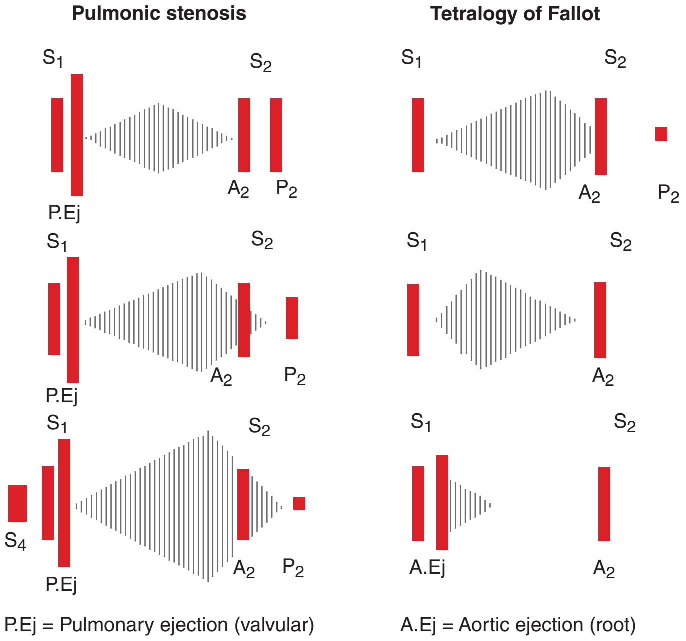

FIGURE 44-4 ***Left***. In valvular pulmonic stenosis with intact ventricular septum, right ventricular systolic ejection becomes progressively longer, with increasing obstruction to flow. As a result, the murmur becomes longer and louder, enveloping the aortic component of the second heart sound (A_2). The pulmonic component (P_2) occurs later, and splitting becomes wider but more difficult to hear because A_2 is lost in the murmur and P_2 becomes progressively fainter and lower pitched. As the pulmonic gradient increases, the isometric contraction phase shortens until the pulmonic valve ejection sound fuses with the first heart sound (S_1). In severe pulmonic stenosis with concentric hypertrophy and decreasing right ventricular compliance, a fourth heart sound appears. ***Right***. In tetralogy of Fallot with increasing obstruction at the pulmonic infundibular area, an increasing amount of right ventricular blood is shunted across the silent ventricular septal defect and flow across the obstructed outflow tract decreases. Therefore, with increasing obstruction, the murmur becomes shorter, earlier, and fainter. P_2 is absent in severe tetralogy of Fallot. A large aortic root receives almost all cardiac output from both ventricular chambers, and the aorta dilates and is accompanied by a root ejection sound that does not vary with respiration. *(Reprinted with permission from Examination of the Heart, Part IV: Auscultation of the Heart © American Heart Association, Inc.)*

these shunts in adults. Features suggestive of a primum ASD include the coexistence of MR due to a cleft anterior mitral valve leaflet and left axis deviation of the QRS complex on the ECG. With sinus venosus ASDs, the left-to-right shunt is usually not large enough to result in a systolic murmur, although the ECG may show abnormalities of sinus node function. A grade 2 or 3 midsystolic murmur may also be heard best at the upper left sternal border in patients with idiopathic dilation of the pulmonary artery; a pulmonary ejection sound is also present in these patients. TTE is indicated to evaluate a grade 2 or 3 midsystolic murmur when there are other signs of cardiac disease.

An isolated grade 1 or 2 midsystolic murmur, heard in the absence of symptoms or signs of heart disease, is most often a benign finding for which no further evaluation, including TTE, is necessary. The most common example of a murmur of this type in an older adult patient is the crescendo-decrescendo murmur of aortic valve sclerosis, heard at the second right interspace (Fig. 44-2). Aortic sclerosis is defined as focal thickening and calcification of the aortic valve to a degree that does not interfere with leaflet opening. The carotid upstrokes are normal, and electrocardiographic LVH is not present. A grade 1 or 2 midsystolic murmur often can be heard at the left sternal border with pregnancy, hyperthyroidism, or anemia, physiologic states that are associated with accelerated blood flow. *Still's murmur* refers to a benign grade 2, vibratory or musical midsystolic murmur at the mid or lower left sternal border in normal children and adolescents, best heard in the supine position (Fig. 44-2).

Late Systolic Murmurs A late systolic murmur that is best heard at the left ventricular apex is usually due to MVP **(Chap. 276)**. Often, this murmur is introduced by one or more nonejection clicks. The radiation of the murmur can help identify the specific mitral leaflet involved in the process of prolapse or flail. The term *flail* refers to the movement made by an unsupported portion of the leaflet (usually the tip) after loss of its chordal attachment(s). With posterior leaflet prolapse or flail, the resultant jet of MR is directed anteriorly and medially, as a result of which the murmur radiates to the base of the heart and masquerades as AS. Anterior leaflet prolapse or flail results in a posteriorly directed MR jet that radiates to the axilla or left infrascapular region. Leaflet flail is associated with a murmur of grade 3 or 4 intensity that can be heard throughout the precordium in thin-chested patients. The presence of an S_3 or a short, rumbling mid-diastolic murmur due to enhanced flow signifies severe MR.

Bedside maneuvers that decrease left ventricular preload, such as standing, will cause the click and murmur of MVP to move closer to the first heart sound, as leaflet prolapse occurs earlier in systole. Standing also causes the murmur to become louder and longer. With squatting, left ventricular preload and afterload are increased abruptly, leading to an increase in left ventricular volume, and the click and murmur move away from the first heart sound as leaflet prolapse is delayed; the murmur becomes softer and shorter in duration (Fig. 44-3). As noted above, these responses to standing and squatting are directionally similar to those observed in patients with HOCM.

A late, apical systolic murmur indicative of MR may be heard transiently in the setting of acute myocardial ischemia; it is due to apical tethering and malcoaptation of the leaflets in response to structural and functional changes of the ventricle and mitral annulus. The intensity of the murmur varies as a function of left ventricular afterload and will increase in the setting of hypertension. TTE is recommended for assessment of late systolic murmurs.

Holosystolic Murmurs (**Figs.** 44-1*B* and **44-5**) Holosystolic murmurs begin with S_1 and continue through systole to S_2. They are usually indicative of chronic mitral or tricuspid valve regurgitation or a VSD and warrant TTE for further characterization. The holosystolic murmur of chronic MR is best heard at the left ventricular apex and radiates to the axilla (Fig. 44-2); it is usually high-pitched and plateau in configuration because of the wide difference between left ventricular and left atrial pressure throughout systole. In contrast to acute MR, left atrial compliance is normal or even increased in chronic MR. As a result, there is only a small increase in left atrial pressure for any increase in regurgitant volume.

Several conditions are associated with chronic MR and an apical holosystolic murmur, including rheumatic scarring of the leaflets, mitral annular calcification, postinfarction left ventricular remodeling, and severe left ventricular chamber enlargement in the setting of a dilated cardiomyopathy **(Chaps. 266–270)**. The severity of the MR is worsened by any contribution from apical displacement of the papillary muscles and leaflet tethering (remodeling). Because the mitral annulus is contiguous with the left atrial endocardium, gradual enlargement of the left atrium from chronic MR will result in further stretching of the annulus and more MR; thus, "MR begets MR." Chronic severe MR results in enlargement and leftward displacement of the left ventricular apex beat and, in some patients, a diastolic filling complex, as described previously **(Fig. 44-1G)**.

The holosystolic murmur of chronic TR is generally softer than that of MR, is loudest at the left lower sternal border, and usually increases in intensity with inspiration (Carvallo's sign). Associated signs include *c-v* waves in the jugular venous pulse, an enlarged and pulsatile liver, ascites, and peripheral edema. The abnormal jugular venous waveforms are the predominant finding and seen very often in the absence of an audible murmur despite Doppler echocardiographic verification of TR. Causes of *primary* TR include myxomatous disease (prolapse), endocarditis, rheumatic disease, radiation, carcinoid, Ebstein's anomaly, and chordal detachment as a complication of right ventricular endomyocardial biopsy. TR is much more commonly a passive process that results secondarily from annular enlargement due to right ventricular dilation in the face of volume or pressure overload or adverse right ventricular remodeling. The leads from cardiac implanted electronic

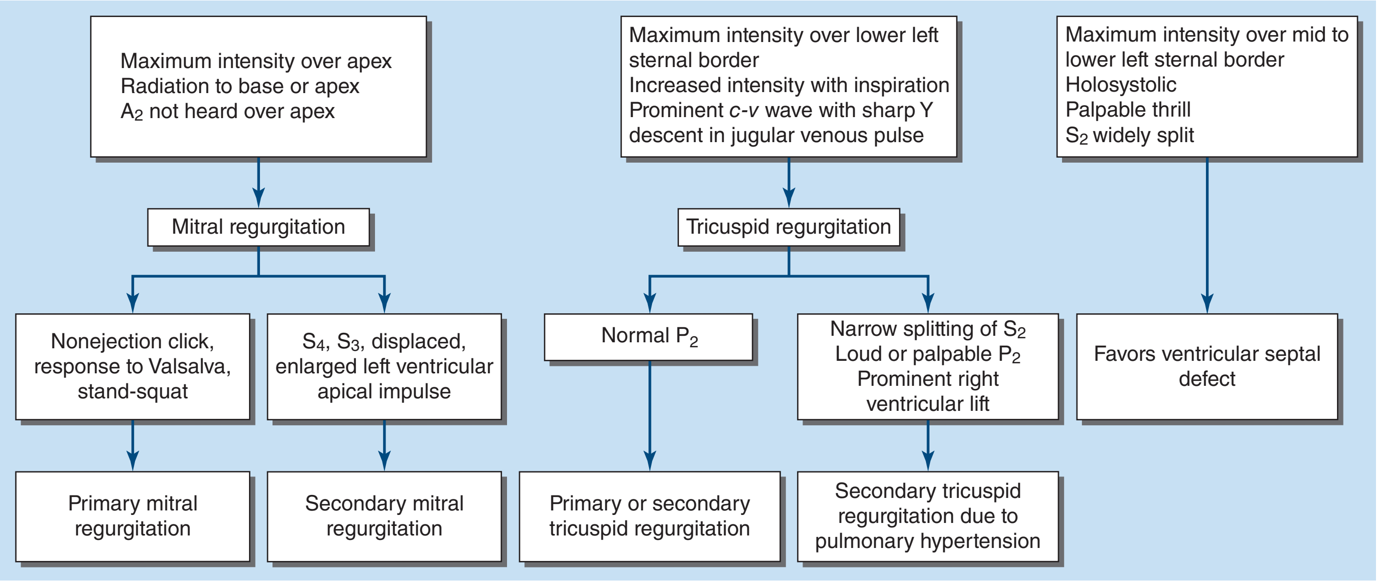
Maximum intensity over apex
Radiation to base or apex
A_2 not heard over apex
Maximum intensity over lower left sternal border
Increased intensity with inspiration
Prominent *c-v* wave with sharp Y descent in jugular venous pulse
Maximum intensity over mid to lower left sternal border
Holosystolic
Palpable thrill
S_2 widely split
Mitral regurgitation
Tricuspid regurgitation
Nonejection click, response to Valsalva, stand-squat
S_4, S_3, displaced, enlarged left ventricular apical impulse
Normal P_2
Narrow splitting of S_2
Loud or palpable P_2
Prominent right ventricular lift
Favors ventricular septal defect
Primary mitral regurgitation
Secondary mitral regurgitation
Primary or secondary tricuspid regurgitation
Secondary tricuspid regurgitation due to pulmonary hypertension

FIGURE 44-5 Differential diagnosis of a holosystolic murmur. The murmur of mitral regurgitation is best heard over the left ventricular apex. The radiation of the murmur depends on the direction in which the jet of mitral regurgitation enters into the left atrium. Differentiation of primary and secondary causes of mitral regurgitation is usually accomplished with transthoracic echocardiography, although the presence of a nonejection click and a mid-late apical systolic murmur, for example, can establish a bedside diagnosis of mitral valve prolapse (primary mitral regurgitation). Secondary mitral regurgitation can occur as a result of left ventricular remodeling. The murmur may be soft and difficult to hear. Other signs of left ventricular dysfunction may be present. Alternatively, mitral regurgitation may occur in the context of long-standing atrial fibrillation with left atrial enlargement and annular dilation (atrial functional mitral regurgitation). Ventricular size may be normal and an S_4 would not be present. Greater than 80% of the tricuspid regurgitation encountered clinically is due to a secondary cause. Severe pulmonary hypertension can be appreciated by a loud, single P_2. Tricuspid regurgitation may be present in the setting of pacemaker leads (see text) or in patients with carcinoid syndrome (primary tricuspid regurgitation) who usually have signs of liver involvement. A ventricular septal defect is usually manifested by a holosystolic murmur with a palpable thrill along the mid- to lower left sternal edge.

devices can cause TR because of leaflet trauma or impingement. Pacing from the right ventricular apex results in dyssynchronous biventricular activation, which may contribute to altered tricuspid valve function.

The holosystolic murmur of a VSD is loudest at the mid- to lower-left sternal border (Fig. 44-2) and radiates widely. A thrill is present at the site of maximal intensity in the majority of patients. There is no change in the intensity of the murmur with inspiration. The intensity of the murmur varies as a function of the anatomic size of the defect. Small, restrictive VSDs, as exemplified by the *maladie de Roger*, create a very loud murmur due to the significant and sustained systolic pressure gradient between the left and right ventricles. With large defects, the ventricular pressures tend to equalize, shunt flow is balanced, and a murmur may not be appreciated. The distinction between post-MI ventricular septal rupture and MR has been reviewed previously.

■ DIASTOLIC HEART MURMURS

Early Diastolic Murmurs (Fig. 44-1*E*) Chronic AR results in a high-pitched, blowing, decrescendo, early- to mid-diastolic murmur that begins after the aortic component of S_2 (A_2) and is best heard at the second right interspace and along the left sternal border. The murmur may be soft and difficult to hear unless auscultation is performed with the patient leaning forward at end expiration. This maneuver brings the aortic root closer to the anterior chest wall. Radiation of the murmur may provide a clue to the cause of the AR. With primary valve disease, such as that due to congenital bicuspid disease, prolapse, or endocarditis, the diastolic murmur tends to radiate along the left sternal border, where it is often louder than appreciated in the second right interspace. When AR is caused by aortic root disease, the diastolic murmur may radiate along the right sternal border. Diseases of the aortic root cause dilation or distortion of the aortic annulus and failure of leaflet coaptation. Causes include Marfan syndrome with aneurysm formation, annuloaortic ectasia, ankylosing spondylitis, and aortic dissection.

Chronic, severe AR also may produce a lower-pitched mid to late, grade 1 or 2 diastolic murmur at the apex (Austin Flint murmur), which is thought to reflect turbulence at the mitral inflow area from the admixture of regurgitant (aortic) and forward (mitral) blood flow. This lower-pitched, apical diastolic murmur can be distinguished from that due to MS by the absence of an opening snap and the response of the murmur to a vasodilator challenge. Lowering afterload with an agent such as amyl nitrite will decrease the duration and magnitude of the aortic–left ventricular diastolic pressure gradient, and thus, the Austin Flint murmur of severe AR will become shorter and softer. The intensity of the diastolic murmur of MS **(Fig. 44-6)** may either remain constant or increase with afterload reduction because of the reflex increase in cardiac output and mitral valve flow.

Although AS and AR may coexist, a grade 2 or 3 crescendo-decrescendo midsystolic murmur frequently is heard at the base of the heart in patients with isolated, severe AR and is due to an increased

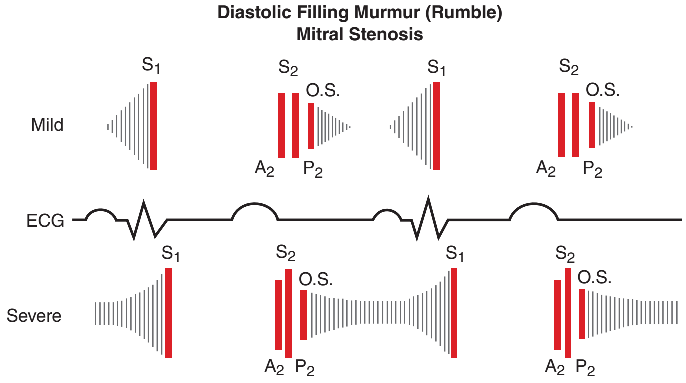
Diastolic Filling Murmur (Rumble)
Mitral Stenosis
Mild
S_1
S_2
O.S.
A_2
P_2
ECG
Severe

FIGURE 44-6 Diastolic filling murmur (rumble) in mitral stenosis. In mild mitral stenosis, the diastolic gradient across the valve is limited to the phases of rapid ventricular filling in early diastole and presystole. The rumble may occur during either or both periods. As the stenotic process becomes severe, a large pressure gradient exists across the valve during the entire diastolic filling period, and the rumble persists throughout diastole. As the left atrial pressure becomes greater, the interval between A_2 (or P_2) and the opening snap (O.S.) shortens. In severe mitral stenosis, secondary pulmonary hypertension develops and results in a loud P_2 and the splitting interval usually narrows. ECG, electrocardiogram. *(Reprinted with permission from Examination of the Heart, Part IV: Auscultation of the Heart. © American Heart Association, Inc.)*

volume and rate of systolic flow. Accurate bedside identification of coexistent AS can be difficult unless the carotid pulse examination is abnormal or the midsystolic murmur is of grade 4 or greater intensity. In the absence of heart failure, chronic severe AR is accompanied by several peripheral signs of significant diastolic runoff, including a wide pulse pressure, a "water-hammer" carotid upstroke (Corrigan's pulse), and Quincke's pulsations of the nail beds. The diastolic murmur of *acute, severe AR* is notably shorter in duration and lower pitched than the murmur of chronic AR. It can be very difficult to appreciate in the presence of a rapid heart rate. These attributes reflect the abrupt rate of rise of diastolic pressure within the unprepared and noncompliant left ventricle and the correspondingly rapid decline in the aortic–left ventricular diastolic pressure gradient. Left ventricular diastolic pressure may increase sufficiently to result in premature closure of the mitral valve and a soft first heart sound. Peripheral signs of significant diastolic runoff are generally not present.

Pulmonic regurgitation (PR) results in a decrescendo, early to mid-diastolic murmur (*Graham Steell murmur*) that begins after the pulmonic component of S_2 (P_2), is best heard at the second left interspace, and radiates along the left sternal border. The intensity of the murmur may increase with inspiration. PR is most commonly due to dilation of the valve annulus from chronic elevation of the pulmonary artery pressure. Signs of pulmonary hypertension, including a right ventricular lift and a loud, single or narrowly split S_2, are present. These features also help distinguish PR from AR as the cause of a decrescendo diastolic murmur heard along the left sternal border. PR in the absence of pulmonary hypertension can occur with endocarditis or a congenitally deformed valve. It is usually present after repair of tetralogy of Fallot in childhood. When pulmonary hypertension is not present, the diastolic murmur is softer and lower pitched than the classic Graham Steell murmur, and the severity of the PR can be difficult to appreciate.

TTE is indicated for the further evaluation of a patient with an early to mid-diastolic murmur. Longitudinal assessment of lesion severity, ventricular size, and systolic function helps guide a potential decision for surgical management. TTE also can provide anatomic information regarding the root and proximal ascending aorta, although computed tomographic or magnetic resonance angiography may be indicated for more precise characterization **(Chap. 248)**.

Mid-Diastolic Murmurs (**Figs. 44-1*F*** and 44-1*G*) Mid-diastolic murmurs result from obstruction and/or augmented flow at the level of the mitral or tricuspid valve. Rheumatic fever is the most common cause of MS (Fig. 44-6). In younger patients with pliable valves, S_1 is loud and the murmur begins after an opening snap, which is a high-pitched sound that occurs shortly after S_2. The interval between the pulmonic component of the second heart sound (P_2) and the opening snap is inversely related to the magnitude of the left atrial–left ventricular pressure gradient. The murmur of MS is low-pitched and thus is best heard with the bell of the stethoscope. It is loudest at the left ventricular apex and often is appreciated only when the patient is turned in the left lateral decubitus position. It is usually of grade 1 or 2 intensity but may be absent when the cardiac output is severely reduced despite significant obstruction. The intensity of the murmur increases during maneuvers that increase cardiac output and mitral valve flow, such as exercise. The duration of the murmur reflects the length of time over which left atrial pressure exceeds left ventricular diastolic pressure. An increase in the intensity of the murmur just before S_1, a phenomenon known as *presystolic accentuation* (**Figs. 44-1*A*** and 44-6), occurs in patients in sinus rhythm and is due to a late increase in transmitral flow with atrial contraction. Presystolic accentuation does not occur in patients with atrial fibrillation.

The mid-diastolic murmur associated with tricuspid stenosis is best heard at the lower left sternal border and increases in intensity with inspiration. A prolonged *y* descent may be visible in the jugular venous waveform. This murmur is very difficult to hear and most often is obscured by left-sided acoustical events.

There are several other causes of mid-diastolic murmurs. Large left atrial myxomas may prolapse across the mitral valve and cause variable degrees of obstruction to left ventricular inflow **(Chap. 282)**. The murmur associated with an atrial myxoma may change in duration and intensity with changes in body position. An opening snap is not present, and there is no presystolic accentuation. Augmented mitral diastolic flow can occur with isolated severe MR or with a large left-to-right shunt at the ventricular or great vessel level and produce a soft, rapid filling sound (S_3) followed by a short, low-pitched mid-diastolic apical murmur (Fig. 44-1*G*). The Austin Flint murmur of severe, chronic AR has already been described.

A short, mid-diastolic murmur is rarely heard during an episode of acute rheumatic fever (Carey-Coombs murmur) and probably is due to flow through an edematous mitral valve. An opening snap is not present in the acute phase, and the murmur dissipates with resolution of the acute attack. Complete heart block with asynchronous atrial and ventricular activation may be associated with intermittent mid- to late diastolic murmurs if atrial contraction occurs when the mitral valve is partially closed. Mid-diastolic murmurs indicative of increased tricuspid valve flow can occur with severe, isolated TR and with large ASDs and significant left-to-right shunting. Other signs of an ASD are present **(Chap. 280)**, including fixed splitting of S_2 and a midsystolic murmur at the mid- to upper left sternal border. TTE is indicated for evaluation of a patient with a mid- to late diastolic murmur. Findings specific to the diseases discussed above will help guide management.

CONTINUOUS MURMURS

(Figs. 44-1*H* and 44-7) Continuous murmurs begin in systole, peak near the second heart sound, and continue into all or part of diastole. Their presence throughout the cardiac cycle implies a pressure gradient between two chambers or vessels during both systole and diastole. The continuous murmur associated with a patent ductus arteriosus is best heard lateral to the upper left sternal border. Large, uncorrected shunts may lead to pulmonary hypertension, attenuation or obliteration of the diastolic component of the murmur, reversal of shunt flow, and differential cyanosis of the lower extremities. A ruptured sinus of Valsalva aneurysm creates a continuous murmur of abrupt onset at the upper right sternal border. Rupture typically occurs into a right heart chamber, and the murmur is indicative of a continuous pressure difference between the aorta and either the right atrium or the right ventricle. A continuous murmur also may be audible along the left sternal border with a coronary arteriovenous fistula and at the site of an arteriovenous fistula used for hemodialysis access. Enhanced flow through enlarged intercostal collateral arteries in patients with aortic coarctation may produce a continuous murmur along the course of one or more ribs. A cervical bruit with both systolic and diastolic components (a to-fro murmur, Fig. 44-7) usually indicates a high-grade carotid artery stenosis.

Not all continuous murmurs are pathologic. A continuous venous hum can be heard in healthy children and young adults, especially

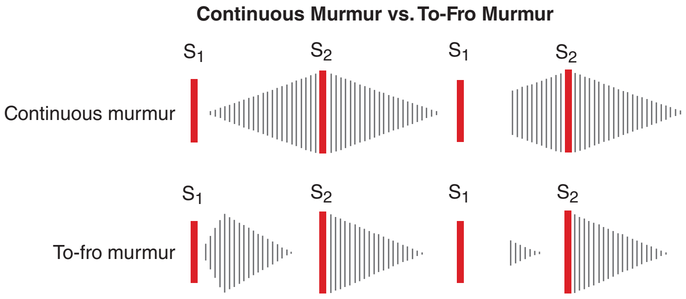

FIGURE 44-7 Comparison of the continuous murmur and the to-fro murmur. During abnormal communication between high-pressure and low-pressure systems, a large pressure gradient exists throughout the cardiac cycle, producing a continuous murmur. A classic example is patent ductus arteriosus. At times, this type of murmur can be confused with a to-fro murmur, which is a combination of systolic ejection murmur and a murmur of semilunar valve incompetence. A classic example of a to-fro murmur is aortic stenosis and regurgitation. A continuous murmur crescendos to near the second heart sound (S_2), whereas a to-fro murmur has two components. The midsystolic ejection component decrescendos and disappears as it approaches S_2. *(Reprinted with permission from Examination of the Heart, Part IV: Auscultation of the Heart. © American Heart Association, Inc.)*

TABLE 44-2 Dynamic Auscultation: Bedside Maneuvers That Can Be Used to Change the Intensity of Cardiac Murmurs (See Text)

1. Respiration
2. Isometric exercise (handgrip)
3. Transient arterial occlusion
4. Pharmacologic manipulation of preload and/or afterload
5. Valsalva maneuver
6. Rapid standing/squatting
7. Passive leg raising
8. Post-premature beat

during pregnancy; it is best appreciated in the right supraclavicular fossa and can be obliterated by pressure over the right internal jugular vein or by having the patient turn his or her head toward the examiner. The continuous mammary souffle of pregnancy is created by enhanced arterial flow through engorged breasts and usually appears during the late third trimester or early puerperium. The murmur is louder in systole. Firm pressure with the diaphragm of the stethoscope can eliminate the diastolic portion of the murmur.

DYNAMIC AUSCULTATION

(See Tables 44-2 and 246-1) Careful attention to the behavior of heart murmurs during simple maneuvers that alter cardiac hemodynamics can provide important clues to their cause and significance.

Respiration Auscultation should be performed during quiet respiration or with a modest increase in inspiratory effort, as more forceful movement of the chest tends to obscure the heart sounds. Left-sided murmurs may be best heard at end expiration, when lung volumes are minimized, and the heart and great vessels are brought closer to the chest wall. This phenomenon is characteristic of the murmur of AR. Murmurs of right-sided origin, such as tricuspid or pulmonic regurgitation, increase in intensity during inspiration. The intensity of left-sided murmurs either remains constant or decreases with inspiration.

Bedside assessment also should evaluate the behavior of S_2 with respiration and the dynamic relationship between the aortic and pulmonic components (Fig. 44-8). Reversed splitting can be a feature of severe AS, HOCM, left bundle branch block, right ventricular pacing, or acute myocardial ischemia. Fixed splitting of S_2 in the presence of a grade 2 or 3 midsystolic murmur at the mid- or upper left sternal border indicates an ASD. Physiologic but wide splitting during the respiratory cycle implies either premature aortic valve closure, as can occur with severe MR, or delayed pulmonic valve closure due to PS or right bundle branch block.

Alterations of Systemic Vascular Resistance Murmurs can change characteristics after maneuvers that alter systemic vascular resistance and left ventricular afterload. The systolic murmurs of MR and VSD become louder during sustained handgrip, simultaneous inflation of blood pressure cuffs on both upper extremities to pressures 20–40 mmHg above systolic pressure for 20 s, or infusion of a vasopressor agent. The murmurs associated with AS or HOCM will become softer or remain unchanged with these maneuvers. The diastolic murmur of AR becomes louder in response to interventions that raise systemic vascular resistance.

Opposite changes in systolic and diastolic murmurs may occur with the use of pharmacologic agents that lower systemic vascular resistance. Inhaled amyl nitrite is now rarely used for this purpose but can help distinguish the murmur of AS or HOCM from that of either MR or VSD, if necessary. The former two murmurs increase in intensity, whereas the latter two become softer after exposure to amyl nitrite. As noted previously, the Austin Flint murmur of severe AR becomes softer, but the mid-diastolic rumble of MS becomes louder, in response to the abrupt lowering of systemic vascular resistance with amyl nitrite and enhanced transmitral valve flow.

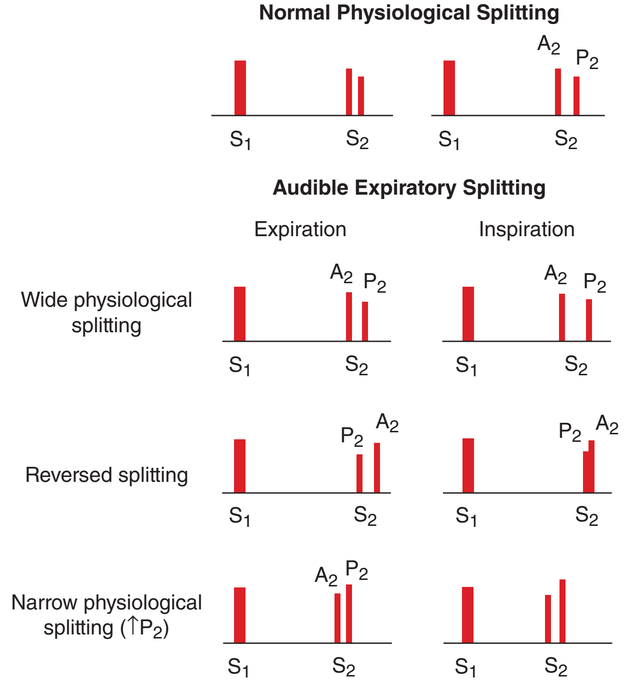
Normal Physiological Splitting
A2
P2
S1
S2
S1
S2
Audible Expiratory Splitting
Expiration
Inspiration
Wide physiological splitting
A2
P2
A2
P2
S1
S2
S1
S2
Reversed splitting
P2
A2
P2
A2
S1
S2
S1
S2
Narrow physiological splitting (↑P2)
A2
P2
S1
S2
S1
S2

FIGURE 44-8 ***Top.*** Normal physiologic splitting of the second heart sound. During expiration, the aortic (A_2) and pulmonic (P_2) components of the second heart sound are separated by <30 ms and are appreciated as a single sound. During inspiration, the splitting interval widens, and A_2 and P_2 are clearly separated into two distinct sounds. ***Bottom.*** Audible expiratory splitting. Wide physiologic splitting is caused by a delay in P_2 (as, for example, with right bundle branch block) or by early closure of the aortic valve (A_2, as for example with severe mitral regurgitation). Reversed splitting is caused by a delay in A_2, resulting in paradoxical movement, i.e., with inspiration P_2 moves toward A_2, and the splitting interval narrows. Narrow physiologic splitting occurs in pulmonary hypertension, and both A_2 and P_2 are heard during expiration at a narrow splitting interval because of the increased intensity and high-frequency composition of P_2. *(Reprinted with permission from Examination of the Heart, Part IV: Auscultation of the Heart © American Heart Association, Inc.)*

Changes in Venous Return The Valsalva maneuver results in an increase in intrathoracic pressure, followed by a decrease in venous return, ventricular filling, and cardiac output. The majority of murmurs decrease in intensity during the strain phase of the maneuver. Two notable exceptions are the murmurs associated with MVP and HOCM, both of which become louder during the Valsalva maneuver. The murmur of MVP may also become longer as leaflet prolapse occurs earlier in systole at smaller ventricular volumes. These murmurs behave in a similar and parallel fashion with standing. Both the click and the murmur of MVP move closer in timing to S_1 on rapid standing from a squatting position (Fig. 44-3). The increase in the intensity of the murmur of HOCM is predicated on the augmentation of the dynamic left ventricular outflow tract gradient that occurs with reduced ventricular filling. Squatting results in abrupt increases in both venous return (preload) and left ventricular afterload that increase ventricular volume, changes that predictably cause a decrease in the intensity and duration of the murmurs associated with MVP and HOCM; the click and murmur of MVP move away from S_1 with squatting. Passive leg raising can be used to increase venous return in patients who are unable to squat and stand. This maneuver may lead to a decrease in the intensity of the murmur associated with HOCM but has less effect in patients with MVP.

Post-Premature Ventricular Contraction A change in the intensity of a systolic murmur in the first beat after a premature beat, or in the beat after a long cycle length in patients with atrial fibrillation, can help distinguish AS from MR, particularly in an older patient in whom the murmur of AS is well transmitted to the apex. Systolic murmurs due to left ventricular outflow obstruction, including that due to AS, increase in intensity in the beat after a premature beat because of the combined effects of enhanced left ventricular filling and post-extrasystolic potentiation of contractile function. Forward flow accelerates, causing an increase in the gradient and a louder murmur. The intensity of the murmur of MR does not change in the post-premature beat as there is relatively little further increase in mitral valve flow or change in the left ventricular–left atrial gradient.

THE CLINICAL CONTEXT

Additional clues to the etiology and importance of a heart murmur can be gleaned from the history and other physical examination findings. Symptoms suggestive of cardiovascular, neurologic, or pulmonary disease help focus the differential diagnosis, as do findings relevant to the jugular venous pressure and waveforms, the arterial pulses, other heart sounds, the lungs, the abdomen, the skin, and the extremities. In many instances, laboratory studies, an ECG, and/or a chest x-ray may have been obtained earlier and may contain valuable information. A patient with suspected infective endocarditis, for example, may have a murmur in the setting of fever, chills, anorexia, fatigue, dyspnea, splenomegaly, petechiae, and positive blood cultures. A new systolic murmur in a patient with a marked fall in blood pressure after a recent MI suggests myocardial rupture. By contrast, an isolated grade 1 or 2 midsystolic murmur at the left sternal border in a healthy, active, and asymptomatic young adult is most likely a benign finding for which no further evaluation is indicated. The context in which the murmur is appreciated often dictates the need for further testing and the pace of the evaluation.

ECHOCARDIOGRAPHY

(Please see **Chap. 42, Fig. 42-9** on page 286 in HPIM 21e; **Chaps. 246 and 248**) Echocardiography with color flow and spectral Doppler is a valuable tool for the assessment of cardiac murmurs. Information regarding valve structure and function, chamber size, wall thickness, ventricular function, estimated pulmonary artery pressures, intracardiac shunt flow, pulmonary and hepatic vein flow, and aortic flow can be ascertained readily. It is important to note that Doppler signals of trace or mild valvular regurgitation of no clinical consequence can be detected with structurally normal tricuspid, pulmonic, and mitral valves. Such signals are not likely to generate enough turbulence to create an audible murmur.

Echocardiography is indicated for the evaluation of patients with early, late, or holosystolic murmurs and patients with grade 3 or louder midsystolic murmurs. Patients with grade 1 or 2 midsystolic murmurs but other symptoms or signs of cardiovascular disease, including those from ECG or chest x-ray, should also undergo echocardiography. Echocardiography is also indicated for the evaluation of any patient with a diastolic murmur and for patients with continuous murmurs not due to a venous hum or mammary souffle. Echocardiography should be considered when there is a clinical need to verify normal cardiac structure and function in a patient whose symptoms and signs are probably noncardiac in origin. The performance of serial echocardiography to follow the course of asymptomatic individuals with valvular heart disease is a central feature of their longitudinal assessment and provides valuable information that may have an impact on decisions regarding the timing of surgery. Routine echocardiography is *not* recommended for asymptomatic patients with a grade 1 or 2 midsystolic murmur without other signs of heart disease. For this category of patients, referral to a cardiovascular specialist could be considered if there is doubt about the significance of the murmur after the initial examination.

The selective use of echocardiography outlined above has not been subjected to rigorous analysis of its cost-effectiveness. For some clinicians, handheld or miniaturized cardiac ultrasound devices have replaced the stethoscope. Although several reports attest to the improved sensitivity of such devices for the detection of valvular heart disease (e.g., rheumatic heart disease in susceptible populations), accuracy is highly operator-dependent, and incremental cost considerations and outcomes have not been addressed adequately for most patient scenarios. The use of electronic or digital stethoscopes with spectral display capabilities has also been proposed as a method to improve the characterization of heart murmurs and the mentored teaching of cardiac auscultation.

OTHER CARDIAC TESTING

(**Chap. 248,** Fig. 44-9) In relatively few patients, clinical assessment and TTE do not adequately characterize the origin and significance of a heart murmur. Transesophageal echocardiography (TEE) can be considered for further evaluation, especially when the TTE windows are limited by body size, chest configuration, or intrathoracic pathology. TEE offers enhanced sensitivity for the detection of a wide range of structural cardiac disorders. Electrocardiographically gated cardiac magnetic resonance (CMR) imaging can provide quantitative information regarding valvular function, regurgitant fraction, regurgitant volume, shunt flow, chamber and great vessel size, ventricular function, and myocardial perfusion. CMR imaging has largely supplanted the need for cardiac catheterization and invasive hemodynamic assessment when there is a discrepancy between the clinical and echocardiographic findings in patients with regurgitant heart valve disease, such as MR or AR. Both CMR and cardiac CT can provide assessment of aortic valve leaflet number when there is uncertainty by TTE regarding whether the valve is bi- or tricuspid, as well as provide information on aortic root and ascending aortic anatomy. The use of coronary CT angiography to exclude coronary artery disease in selected patients with a low pretest probability of disease before valve surgery has gained wide acceptance. Invasive angiography and hemodynamic assessment may be required for a more complete preoperative evaluation.

INTEGRATED APPROACH

The accurate identification of a heart murmur begins with a systematic approach to cardiac auscultation. Characterization of its major attributes, as reviewed above, allows the examiner to construct a preliminary differential diagnosis, which is then refined by integration of information available from the history, associated cardiac findings, the general physical examination, and the clinical context. The need for and urgency of further testing follow sequentially. Correlation of the findings on auscultation with the noninvasive data provides an educational feedback loop and an opportunity for improving physical examination skills. Cost considerations mandate that noninvasive imaging be justified on the basis of its incremental contribution to diagnosis, treatment, and outcome. Cardiac auscultation using a stethoscope remains a time-honored tradition in medicine, the benefits of which extend beyond accurate recognition of heart sounds. Selective augmentation with, rather than wholesale replacement by, handheld ultrasound and newer technologies such as artificial intelligence may improve diagnostic accuracy and better guide therapeutic decisions.

FURTHER READING

Chamsi-Pasha MA et al: Handheld echocardiography: Current state and future perspectives. Circulation 136:2178, 2017.

Chorba JS et al: Deep learning algorithm for automated cardiac murmur detection via a digital stethoscope platform. J Am Heart Assoc 10: e019905, 2021.

Fang JC et al: History, physical examination and the virtual visit. An evidence-based approach, in *Braunwald's Heart Disease. A Textbook of Cardiovascular Medicine*, 12th ed, RO Bonow et al (eds). Philadelphia, Elsevier/Saunders, 2024, pp. 123–140.

Jani V et al: The discerning ear: Cardiac auscultation in the era of artificial intelligence and telemedicine. Eur Heart J Digit Health 2:456, 2021.

Mehta M et al: Handheld ultrasound versus physical examination in patients referred for transthoracic echocardiography for a suspected cardiac condition. J Am Coll Cardiol Imaging 7:983, 2014.

Otto CM et al: 2020 AHA/ACC guideline for the management of patients with valvular heart disease. J Am Coll Cardiol 143:e72, 2021.

Stout KK et al: 2018 AHA/ACC guidelines for the management of patients with congenital heart disease. J Am Coll Cardiol 73:e81, 2019.

Thomas F et al: Accuracy and efficacy of hand-held echocardiography in diagnosing valve disease: A systematic review. Am J Med 131:1155, 2018.

45

Palpitations

Joseph Loscalzo

Palpitations are extremely common among patients who present to their internists and can best be defined as a "thumping," "pounding," or "fluttering" sensation in the chest. This sensation can be either intermittent or sustained and either regular or irregular. Most patients interpret palpitations as an unusual awareness of the heartbeat and become especially concerned when they sense that they have had "skipped" or "missing" heartbeats. Palpitations are often noted when the patient is quietly resting, during which time other stimuli are minimal. Palpitations that are positional generally reflect a structural process within (e.g., atrial myxoma) or adjacent to (e.g., mediastinal mass) the heart.

Palpitations are brought about by cardiac (43%), psychiatric (31%), miscellaneous (10%), and unknown (16%) causes, according to one large series. Among the cardiovascular causes are premature atrial and ventricular contractions, supraventricular and ventricular arrhythmias, mitral valve prolapse (with or without associated arrhythmias), aortic insufficiency, atrial myxoma, myocarditis, and pulmonary embolism. Intermittent palpitations are commonly caused by premature atrial or ventricular contractions: the post-extrasystolic beat is sensed by the patient owing to the increase in ventricular end-diastolic dimension following the pause in the cardiac cycle and the increased strength of contraction (post-extrasystolic potentiation) of that beat. Regular, sustained palpitations can be caused by regular supraventricular and ventricular tachycardias. Irregular, sustained palpitations can be caused by atrial fibrillation. It is important to note that most arrhythmias are not associated with palpitations. In those that are, it is often useful either to ask the patient to "tap out" the rhythm of the palpitations or to take their pulse during palpitations. In general, hyperdynamic cardiovascular states caused by catecholaminergic stimulation from exercise, stress, or pheochromocytoma can lead to palpitations. Palpitations are common among athletes, especially older endurance athletes. In addition, the enlarged ventricle of aortic regurgitation and accompanying hyperdynamic precordium frequently lead to the sensation of palpitations. Other factors that enhance the strength of myocardial contraction, including tobacco, caffeine, aminophylline, atropine, thyroxine, cocaine, amphetamines, and cannabis, can cause palpitations.

Psychiatric causes of palpitations include panic attacks or disorders, anxiety states, and somatization, alone or in combination. Patients with psychiatric causes for palpitations more commonly report a longer duration of the sensation (>15 min) and other accompanying symptoms than do patients with other causes. Among the miscellaneous causes of palpitations are thyrotoxicosis, drugs (see above) and ethanol, spontaneous skeletal muscle contractions of the chest wall, pheochromocytoma, systemic mastocytosis, and post-COVID syndrome.

APPROACH TO THE PATIENT

Palpitations

The principal goal in assessing patients with palpitations is to determine whether the symptom is caused by a life-threatening arrhythmia. Patients with preexisting coronary artery disease (CAD) or risk factors for CAD are at greatest risk for ventricular arrhythmias (Chap. 253) as a cause for palpitations. In addition, the association of palpitations with other symptoms suggesting hemodynamic compromise, including syncope or lightheadedness, supports this diagnosis. Palpitations caused by sustained tachyarrhythmias in patients with CAD can be accompanied by angina pectoris or dyspnea, and, in patients with ventricular dysfunction (systolic or diastolic), aortic stenosis, hypertrophic cardiomyopathy, or mitral stenosis (with or without CAD), can be accompanied by dyspnea from increased left atrial and pulmonary venous pressure.

Key features of the physical examination that will help confirm or refute the presence of an arrhythmia as a cause for palpitations (as well as its adverse hemodynamic consequences) include measurement of the vital signs, assessment of the jugular venous pressure and pulse, and auscultation of the chest and precordium. A resting electrocardiogram can be used to document the arrhythmia. If exertion is known to induce the arrhythmia and accompanying palpitations, exercise electrocardiography can be used to make the diagnosis. If the arrhythmia is sufficiently infrequent, other methods must be used, including continuous electrocardiographic (Holter) monitoring; telephonic monitoring, through which the patient can transmit an electrocardiographic tracing during a sensed episode; loop recordings (external or implantable), which can capture the electrocardiographic event for later review; and mobile (self-monitoring) cardiac outpatient telemetry. Data suggest that Holter monitoring is of limited clinical utility, while the implantable loop recorder and mobile cardiac outpatient telemetry are safe and possibly more cost-effective in the assessment of patients with (infrequent) recurrent, unexplained palpitations. The use of a diary or an electronic marker to indicate the timing of palpitations sensed by the patient is essential for appropriate interpretation of these studies.

Most patients with palpitations do not have serious arrhythmias or underlying structural heart disease. If sufficiently troubling to the patient, occasional benign atrial or ventricular premature contractions can often be managed with beta-blocker therapy. Palpitations incited by alcohol, tobacco, or illicit drugs need to be managed by abstention, while those caused by pharmacologic agents should be addressed by considering alternative therapies when appropriate or possible. Psychiatric causes of palpitations may benefit from cognitive therapy or pharmacotherapy. The physician should note that palpitations are at the very least bothersome and, on occasion, frightening to the patient. Once serious causes for the symptom have been excluded, the patient should be reassured that the palpitations will not adversely affect overall prognosis.

■ FURTHER READING

Crossland S, Berkin L: Problem based review: The patient with palpitations. Acute Med 11:169, 2012.

Jamshed N et al: Emergency management of palpitations in the elderly: Epidemiology, diagnostic approaches, and therapeutic options. Clin Geriatr Med 29:205, 2013.

Martson HR et al: Mobile self-monitoring ECG devices to diagnose arrhythmias that coincide with palpitations: A scoping review. Healthcare (Basel) 7:E96, 2019.

Sakh R et al: Insertable cardiac monitors: current indications and devices. Expert Rev Med Devices 16:45, 2019.

Weinstock C et al: Evidence-based approach to palpitations. Med Clin North Am 105:93, 2021.

46

Exercise Intolerance

Joseph Loscalzo, William M. Oldham

Exercise intolerance is defined as the inability to perform physical activity at a level expected for a person of a given age, sex, body mass, and muscle mass. Reduced exercise tolerance is a common symptom of many chronic diseases, including ischemic heart disease, valvular heart disease, heart failure, chronic obstructive pulmonary disease, interstitial lung disease, cystic fibrosis, pulmonary hypertension, stroke, neuromuscular disorders, and postinfection syndromes, and it reduces quality of life. While not all patients with these disorders necessarily manifest exercise intolerance, those who do often have an increased rate of disease progression, as well as mortality.

PATHOBIOLOGY

The pathobiology of exercise intolerance depends on the specific underlying cause but can be physiologically defined by measuring exercise capacity in terms of the rate of oxygen consumption ($\dot{V}O_2$) at peak exercise ($\dot{V}O_{2max}$). Any factor that impairs O_2 delivery or utilization can reduce $\dot{V}O_{2max}$ and cause exercise intolerance. In natural sequence, reduced inspired O_2 concentration, reduced alveolar ventilation, impaired lung diffusion from alveolus to capillary, reduced hemoglobin concentration or transport (loading or release) of O_2, decreased cardiac output, impaired diffusion of O_2 into (skeletal) muscle, and impaired mitochondrial respiration can each lead to decreased exercise capacity and exercise intolerance **(Fig. 46-1)**. In the absence of clear evidence for a singular cause among these specific pathophysiologic abnormalities (such as decreased cardiac output in a patient with heart failure with reduced ejection fraction) or in the setting of multiple causes, cardiopulmonary exercise testing (CPET) can serve as an important diagnostic test. CPET typically involves measurement of breath-by-breath O_2 consumption and CO_2 production and continuously recording the electrocardiogram during stationary cycle ergometry. In more advanced types of CPET, intracardiac (right-heart) pressures are also measured during exercise, as are arterial blood gases and lactate before, during, and after exercise, providing additional diagnostic information to identify the underlying physiologic etiology of exercise limitation.

SIGNS AND SYMPTOMS

The signs and symptoms of exercise intolerance can vary depending on its severity and etiology. For example, patients with heart failure with reduced ejection fraction following an acute myocardial infarction may be unable to walk up a flight of stairs yet feel comfortable at rest, or patients with severe pulmonary arterial hypertension may be short of breath with minimal exertion. In some cases, patients who are quite intolerant of exercise may note unusual breathlessness or dyspnea at rest accompanied by tachypnea, (sinus) tachycardia, muscle fatigue, weakness, or frank myalgias.

SPECIFIC DISORDERS

Exercise intolerance in patients with heart failure with preserved ejection fraction (HFpEF) is a well-recognized feature of the syndrome. The great majority of patients with HFpEF manifest multiple causes, including (in decreasing order of frequency) impaired skeletal muscle diffusion of O_2, reduced cardiac output (with exercise), decreased alveolar ventilation, reduced lung diffusing capacity, and anemia. Internal work, a body mass index–related measure of the cost of initiating movement, is higher in patients with HFpEF compared with controls and is associated with rapid increases in cardiac filling pressures (and the pulmonary capillary wedge pressure) early in exercise. Other associated mechanisms for these abnormalities include increased large artery stiffness, chronotropic incompetence, and microvascular dysfunction (both endothelium-dependent and endothelium-independent) **(Fig. 46-2)**. Accompanying comorbid conditions may also contribute to these multifactorial drivers of exercise intolerance in HFpEF, including inflammation, adipokine signaling, and insulin resistance in obese diabetics. Ascertaining the set of causes in any individual patient serves as the basis for a precision medicine approach to the treatment of this common symptom in HFpEF, as is the case for other diseases associated with exercise intolerance.

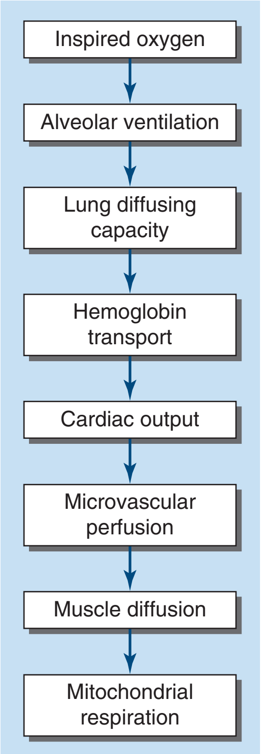

FIGURE 46-1 Oxygen delivery and utilization pathway.

A variety of (post)viral syndromes are also associated with exercise intolerance, including Epstein-Barr virus infection and post-COVID-19 (long COVID) syndrome. Patients with long COVID often have symptoms that are similar to those of patients with myalgic encephalomyelitis; however, the latter more typically include postexertional malaise. In patients following recovery from acute COVID, the inability to return to normal activity levels often heralds the development of long COVID. In patients with long COVID, a significant number noted exertional dyspnea, fatigue, and anxiety for up to 3 months after the acute infection, with fatigue remaining as the most common symptom thereafter. Orthostatic intolerance and positional tachycardia are common and reflect autonomic dysfunction. Among the mechanisms for long COVID–associated fatigue and exercise intolerance are oxidative stress, altered energy metabolism, and dysbiosis of the gut microbiome. Cardiac deconditioning is also believed to play a role in the exercise intolerance of long COVID. Thus, focused exercise prescriptions guided by patient-specific pathophysiology have been shown to relieve symptoms in many individuals.

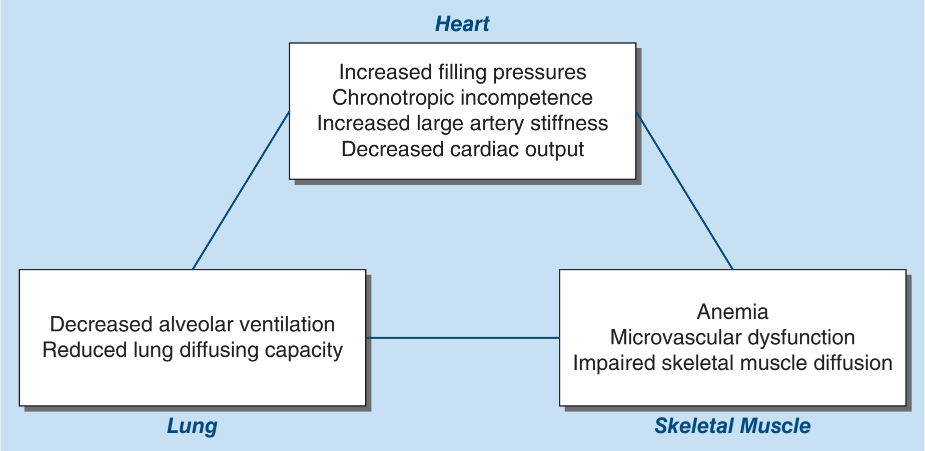
Heart
Increased filling pressures
Chronotropic incompetence
Increased large artery stiffness
Decreased cardiac output
Decreased alveolar ventilation
Reduced lung diffusing capacity
Lung
Anemia
Microvascular dysfunction
Impaired skeletal muscle diffusion
Skeletal Muscle

FIGURE 46-2 Determinants of exercise intolerance in heart failure with preserved ejection fraction (HFpEF).

FURTHER READING

Bunsawat K et al: Exercise intolerance in heart failure with preserved ejection fraction: Causes, consequences and the journey toward a cure. Exp Physiol 109:502, 2023.

Edwards JA et al: Characteristics and treatment of exercise intolerance in patients with long COVID. J Cardiopulm Rehabil Prev 43:400, 2023.

Houstis NE et al: Exercise intolerance in heart failure with preserved ejection fraction. Circulation 137:148, 2018.

McCoy J et al: Pathophysiology of exercise intolerance in chronic diseases: The role of diminished cardiac performance in mitochondrial and heart failure patients. Open Heart 4:000632, 2017.

Ramirez MF et al: Obesity-related biomarkers are associated with exercise intolerance in HFpEF. Circ Heart Fail 16:010618, 2023.

Shah RV et al: Metabolic cost of exercise initiation in patients with heart failure with preserved ejection fraction vs community dwelling adults. JAMA Cardiol 6:653, 2021.

Section 6 Alterations in Gastrointestinal Function

47

Dysphagia

Ikuo Hirano[*], Peter J. Kahrilas

Dysphagia—difficulty with swallowing—refers to problems with the transit of food or liquid from the mouth to the hypopharynx or through the esophagus. Severe dysphagia can compromise nutrition, cause aspiration, and reduce quality of life. Additional terminology pertaining to swallowing dysfunction is as follows. *Aphagia* (inability to swallow) typically denotes complete esophageal obstruction, most commonly encountered in the acute setting of a food bolus or foreign body impaction. *Odynophagia* refers to painful swallowing, typically resulting from mucosal ulceration within the oropharynx or esophagus. It commonly is accompanied by dysphagia, but the converse is not true. *Globus pharyngeus* is a foreign body sensation localized in the neck that does not interfere with swallowing and sometimes is relieved by swallowing. *Transfer dysphagia* frequently results in nasal regurgitation or pulmonary aspiration during swallowing and is characteristic of oropharyngeal dysphagia. *Phagophobia* (fear of swallowing) and *refusal to swallow* may be psychogenic or related to anticipatory anxiety about food bolus obstruction, odynophagia, or aspiration.

PHYSIOLOGY OF SWALLOWING

Swallowing begins with a voluntary (oral) phase that includes preparation during which food is masticated and mixed with saliva. This is followed by a transfer phase during which the bolus is pushed into the pharynx by the tongue. Bolus entry into the hypopharynx initiates the pharyngeal swallow response, which is centrally mediated and involves a complex series of actions, the net result of which is to propel food through the pharynx into the esophagus while preventing its entry into the airway. To accomplish this, the larynx is elevated and pulled forward, actions that also facilitate upper esophageal sphincter (UES) opening. Tongue pulsion then propels the bolus through the UES, followed by a peristaltic contraction that clears residue from the pharynx and through the esophagus. The lower esophageal sphincter (LES) relaxes as the food enters the esophagus and remains relaxed until the peristaltic contraction has delivered the bolus into the stomach. Peristaltic contractions elicited in response to a swallow are called *primary peristalsis* and involve sequenced inhibition followed by contraction of the musculature along the entire length of the esophagus. The inhibition that precedes the peristaltic contraction is called *deglutitive inhibition*. Local distention of the esophagus anywhere along its length, as may occur with gastroesophageal reflux, activates *secondary peristalsis* that begins at the point of distention and proceeds distally. Tertiary esophageal contractions are nonperistaltic, disordered esophageal contractions that may be observed to occur spontaneously during fluoroscopic observation.

The musculature of the oral cavity, pharynx, UES, and cervical esophagus is striated and directly innervated by lower motor neurons carried in cranial nerves **(Fig. 47-1)**. Oral cavity muscles are innervated by the fifth (trigeminal) and seventh (facial) cranial nerves; the tongue, by the twelfth (hypoglossal) cranial nerve. Pharyngeal muscles are innervated by the ninth (glossopharyngeal) and tenth (vagus) cranial nerves.

Physiologically, the UES consists of the cricopharyngeus muscle, the adjacent inferior pharyngeal constrictor, and the proximal portion of the cervical esophagus. UES innervation is derived from the vagus nerve, whereas the innervation to the musculature acting on the UES to facilitate its opening during swallowing comes from the fifth, seventh, and twelfth cranial nerves. The UES remains closed at rest owing to both its inherent elastic properties and neurogenically mediated contraction of the cricopharyngeus muscle. UES opening during swallowing involves both cessation of vagal excitation to the cricopharyngeus and simultaneous contraction of the suprahyoid and geniohyoid muscles that pull open the UES in conjunction with the upward and forward displacement of the larynx.

The neuromuscular apparatus for peristalsis is distinct in proximal and distal parts of the esophagus. The cervical esophagus, like the pharyngeal musculature, consists of striated muscle and is directly innervated by lower motor neurons of the vagus nerve. Peristalsis in the proximal esophagus is governed by the sequential activation of the vagal motor neurons in the nucleus ambiguus. In contrast, the distal esophagus and LES are composed of smooth muscle and are controlled by excitatory and inhibitory neurons within the esophageal myenteric plexus. Medullary preganglionic neurons from the dorsal motor nucleus of the vagus trigger peristalsis via these ganglionic neurons during primary peristalsis. Neurotransmitters of the excitatory ganglionic neurons are acetylcholine and substance P; those of the inhibitory neurons are vasoactive intestinal peptide and nitric oxide. Peristalsis results from the patterned activation of inhibitory followed by excitatory ganglionic neurons, with progressive dominance of the inhibitory neurons distally. Similarly, LES relaxation occurs with the onset of deglutitive inhibition and persists until the peristaltic sequence is complete. At rest, the LES is contracted because of excitatory ganglionic stimulation and its intrinsic myogenic tone, a property that distinguishes it from the adjacent esophagus. The function of the LES is supplemented by the surrounding muscle of the right diaphragmatic crus, which acts as an external sphincter during inspiration, cough, or abdominal straining.

PATHOPHYSIOLOGY OF DYSPHAGIA

Dysphagia can be subclassified both by location and by the circumstances in which it occurs. With respect to location, distinct considerations apply to oral, pharyngeal, or esophageal dysphagia. Normal transport of an ingested bolus depends on the consistency and size of the bolus, the caliber of the lumen, the integrity of peristaltic contraction, and deglutitive inhibition of both the UES and the LES. Dysphagia caused by an oversized bolus or a narrow lumen is called *structural dysphagia*, whereas dysphagia due to abnormalities of peristalsis or impaired sphincter relaxation after swallowing is called *propulsive* or *motor dysphagia*. More than one mechanism may be operative in a patient with dysphagia. Scleroderma commonly presents with absent peristalsis as well as a weakened LES that predisposes patients to peptic stricture formation. Likewise, radiation therapy for head and neck cancer may compound the functional deficits in the oropharyngeal swallow attributable to the tumor and cause cervical esophageal stenosis. It is worth noting that in addition to bolus transit, symptom reporting of dysphagia is dependent upon intact sensory innervation and central nervous system perception. Chronicity of motor dysfunction in

---

[*] Deceased.

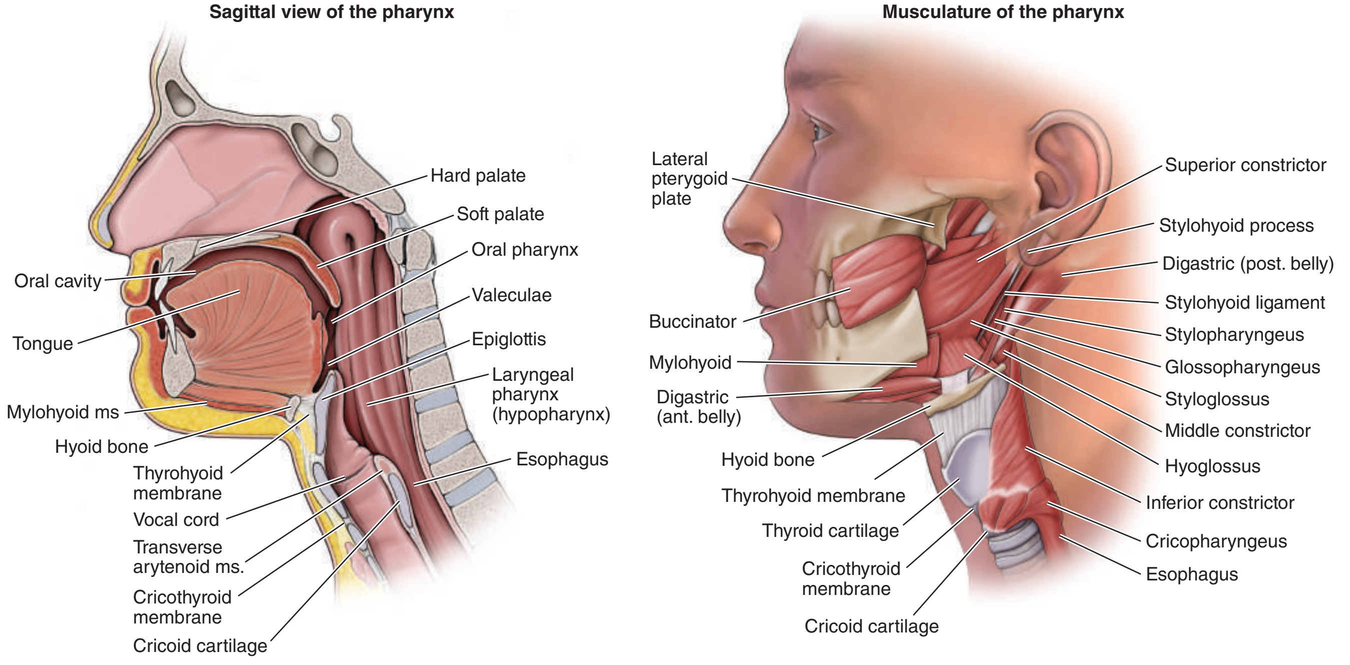
Sagittal view of the pharynx
Hard palate
Soft palate
Oral pharynx
Oral cavity
Valeculae
Tongue
Epiglottis
Laryngeal pharynx (hypopharynx)
Mylohyoid ms
Hyoid bone
Esophagus
Thyrohyoid membrane
Vocal cord
Transverse arytenoid ms.
Cricothyroid membrane
Cricoid cartilage
Musculature of the pharynx
Lateral pterygoid plate
Superior constrictor
Stylohyoid process
Digastric (post. belly)
Stylohyoid ligament
Buccinator
Stylopharyngeus
Mylohyoid
Glossopharyngeus
Digastric (ant. belly)
Styloglossus
Middle constrictor
Hyoid bone
Hyoglossus
Thyrohyoid membrane
Inferior constrictor
Thyroid cartilage
Cricopharyngeus
Cricothyroid membrane
Esophagus
Cricoid cartilage

FIGURE 47-1 Sagittal and diagrammatic views of the musculature involved in enacting oropharyngeal swallowing. Note the dominance of the tongue in the sagittal view and the intimate relationship between the entrance to the larynx (airway) and the esophagus. In the resting configuration illustrated, the esophageal inlet is closed. This is transiently reconfigured such that the esophageal inlet is open and the laryngeal inlet closed during swallowing. *(Adapted from PJ Kahrilas, in DW Gelfand and JE Richter [eds]: Dysphagia: Diagnosis and Treatment. New York, Igaku-Shoin Medical Publishers, 1989, pp. 11–28.)*

combination with defects in peripheral sensory perception may reduce patient-reported dysphagia in patients with achalasia.

Oral and Pharyngeal (Oropharyngeal) Dysphagia Oral-phase dysphagia is associated with poor bolus formation and control so that food has prolonged retention within the oral cavity and may seep out of the mouth. Drooling and difficulty in initiating swallowing are characteristic signs. Poor bolus control also may lead to premature spillage of food into the hypopharynx with resultant aspiration into the trachea, evident as swallow-induced cough or regurgitation into the nasal cavity. Pharyngeal-phase dysphagia is associated with retention of food in the pharynx due to poor tongue or pharyngeal propulsion or obstruction at the UES. Signs and symptoms of concomitant hoarseness or cranial nerve dysfunction may be associated with oropharyngeal dysphagia that is associated with neurologic disorders.

Oropharyngeal dysphagia may be due to neurologic, muscular, structural, iatrogenic, infectious, and metabolic causes. Iatrogenic, neurologic, and structural pathologies are most common. Iatrogenic causes include surgery and radiation, often in the setting of head and neck cancer. Neurogenic dysphagia resulting from cerebrovascular accidents, Parkinson's disease, and amyotrophic lateral sclerosis is a major source of morbidity related to aspiration and malnutrition. Medullary nuclei directly innervate the oropharynx. Lateralization of pharyngeal dysphagia implies either a structural pharyngeal lesion or a neurologic process that selectively targets the ipsilateral brainstem nuclei or cranial nerve. Advances in functional brain imaging have elucidated an important role of the cerebral cortex in swallow function and dysphagia. Asymmetry in the cortical representation of the pharynx provides an explanation for the dysphagia that occurs as a consequence of unilateral cortical cerebrovascular accidents.

Oropharyngeal structural lesions causing dysphagia include Zenker's diverticulum, cricopharyngeal bar, and neoplasia. Zenker's diverticulum typically is encountered in elderly patients. In addition to dysphagia, patients may present with regurgitation of particulate food debris, aspiration, and halitosis. The pathogenesis is related to stenosis of the cricopharyngeus that causes diminished opening of the UES and results in increased hypopharyngeal pressure during swallowing with development of a pulsion diverticulum immediately above the cricopharyngeus in a region of potential weakness known as Killian's dehiscence. A cricopharyngeal bar, appearing as a prominent indentation behind the lower third of the cricoid cartilage, is related to Zenker's diverticulum in that it involves limited distensibility of the cricopharyngeus and can lead to the formation of a Zenker's diverticulum. However, a cricopharyngeal bar is a common radiographic finding, and most patients with transient cricopharyngeal bars are asymptomatic, making it important to rule out alternative etiologies of dysphagia before treatment. Furthermore, cricopharyngeal bars may be secondary to other neuromuscular disorders that impair opening of the UES.

Since the pharyngeal phase of swallowing occurs in less than a second, rapid-sequence fluoroscopy is necessary to evaluate functional abnormalities. Adequate fluoroscopic examination requires that the patient be conscious and cooperative. The study incorporates recordings of swallow sequences during ingestion of food and liquids of varying consistencies. The pharynx is examined to detect bolus retention, regurgitation into the nose, or aspiration into the trachea. Timing and integrity of pharyngeal contraction and opening of the UES with a swallow are analyzed to assess both aspiration risk and the potential for swallow therapy. Structural abnormalities of the oropharynx, especially those that may require biopsies, also should be assessed by direct laryngoscopic examination.

Esophageal Dysphagia The adult esophagus measures 18–26 cm in length and is anatomically divided into the cervical esophagus, extending from the pharyngoesophageal junction to the suprasternal notch, and the thoracic esophagus, which continues to the diaphragmatic hiatus. When distended, the esophageal lumen has internal dimensions of about 2 cm in the anteroposterior plane and 3 cm in the lateral plane. Solid food dysphagia becomes common when the lumen is narrowed to <13 mm, but also can occur with larger diameters in the setting of poorly masticated food or motor dysfunction. Circumferential lesions are more likely to cause dysphagia than are lesions that involve only a partial circumference of the esophageal wall. The most common structural causes of dysphagia are Schatzki's rings, eosinophilic esophagitis, and peptic strictures. Dysphagia also occurs in the setting of gastroesophageal reflux disease without a stricture, perhaps on the basis of altered esophageal sensation, reduced esophageal mural distensibility, or motor dysfunction.

Propulsive disorders leading to esophageal dysphagia result from abnormalities of peristalsis and/or deglutitive inhibition, potentially affecting the cervical or thoracic esophagus. Since striated muscle pathology usually involves both the oropharynx and the cervical esophagus, the clinical manifestations usually are dominated by oropharyngeal dysphagia. Diseases affecting smooth muscle involve both the thoracic esophagus and the LES. A dominant manifestation of this, absent peristalsis, refers to either the complete absence of swallow-induced contraction (absent contractility) or the presence of nonperistaltic, disordered contractions. Absent peristalsis and failure of deglutitive LES relaxation are the defining features of achalasia. In distal esophageal spasm (DES), LES function is normal, with the disordered motility restricted to the esophageal body. Absent contractility combined with severe weakness of the LES is a pattern commonly found in patients with scleroderma.

APPROACH TO THE PATIENT

Dysphagia

Figure 47-2 shows an algorithm for the approach to a patient with dysphagia.

HISTORY

The patient history is extremely valuable in making a presumptive diagnosis or at least substantially limiting the differential diagnoses in most patients. Key elements of the history are the localization of dysphagia, the circumstances in which dysphagia is experienced, other symptoms associated with dysphagia, and progression. Dysphagia that localizes to the suprasternal notch may indicate either an oropharyngeal or an esophageal etiology as distal dysphagia is referred proximally about 30% of the time. Dysphagia that localizes to the chest is esophageal in origin. Nasal regurgitation and tracheobronchial aspiration manifest by coughing with swallowing are hallmarks of oropharyngeal dysphagia. Severe cough with swallowing may also be a sign of a tracheoesophageal fistula. The presence of hoarseness may be another important diagnostic clue. When hoarseness precedes dysphagia, the primary lesion is usually laryngeal; hoarseness that occurs after the development of dysphagia may result from compromise of the recurrent laryngeal nerve by a malignancy. The type of food causing dysphagia is an important consideration. Intermittent dysphagia that occurs only with solid food implies structural dysphagia, whereas constant dysphagia with both liquids and solids strongly suggests an esophageal motor abnormality. Two caveats to this pattern are that despite having a motor abnormality, patients with scleroderma generally develop mild dysphagia for solids only and that patients with oropharyngeal dysphagia often have greater difficulty managing liquids than solids. Dysphagia that is progressive over the course of weeks to months raises concern for neoplasia. Episodic dysphagia to solids that is unchanged or slowly progressive over years indicates a benign disease process such as a Schatzki ring or eosinophilic esophagitis. Food impaction with a prolonged inability to pass an ingested bolus even with ingestion of liquid is typical of structural dysphagia. Chest pain may accompany dysphagia whether it is related to motor disorders, structural disorders, or reflux disease. A prolonged history of heartburn preceding the onset of dysphagia is suggestive of peptic stricture and, infrequently, esophageal adenocarcinoma. A history of prolonged nasogastric intubation, esophageal or head and neck surgery, ingestion of caustic agents or pills, previous radiation or chemotherapy, or associated mucocutaneous diseases may help isolate the cause of dysphagia. With accompanying odynophagia, which usually is indicative of ulceration, infectious or pill-induced esophagitis should be suspected. In patients with AIDS or other immunocompromised states, esophagitis due to opportunistic infections such as *Candida*, herpes simplex virus, or cytomegalovirus and to tumors such as Kaposi's sarcoma and lymphoma should be considered. A history of atopy increases concerns for eosinophilic esophagitis, which is most prevalent in Caucasian male patients between the ages of 20 and 40 years. Medication use should identify agents associated with pill esophagitis and narcotics that are associated with opioid-induced esophageal dysmotility.

PHYSICAL EXAMINATION

Physical examination is important in the evaluation of oral and pharyngeal dysphagia because dysphagia is usually only one of many manifestations of a more global disease process. Signs of bulbar or pseudobulbar palsy, including dysarthria, dysphonia, ptosis, and

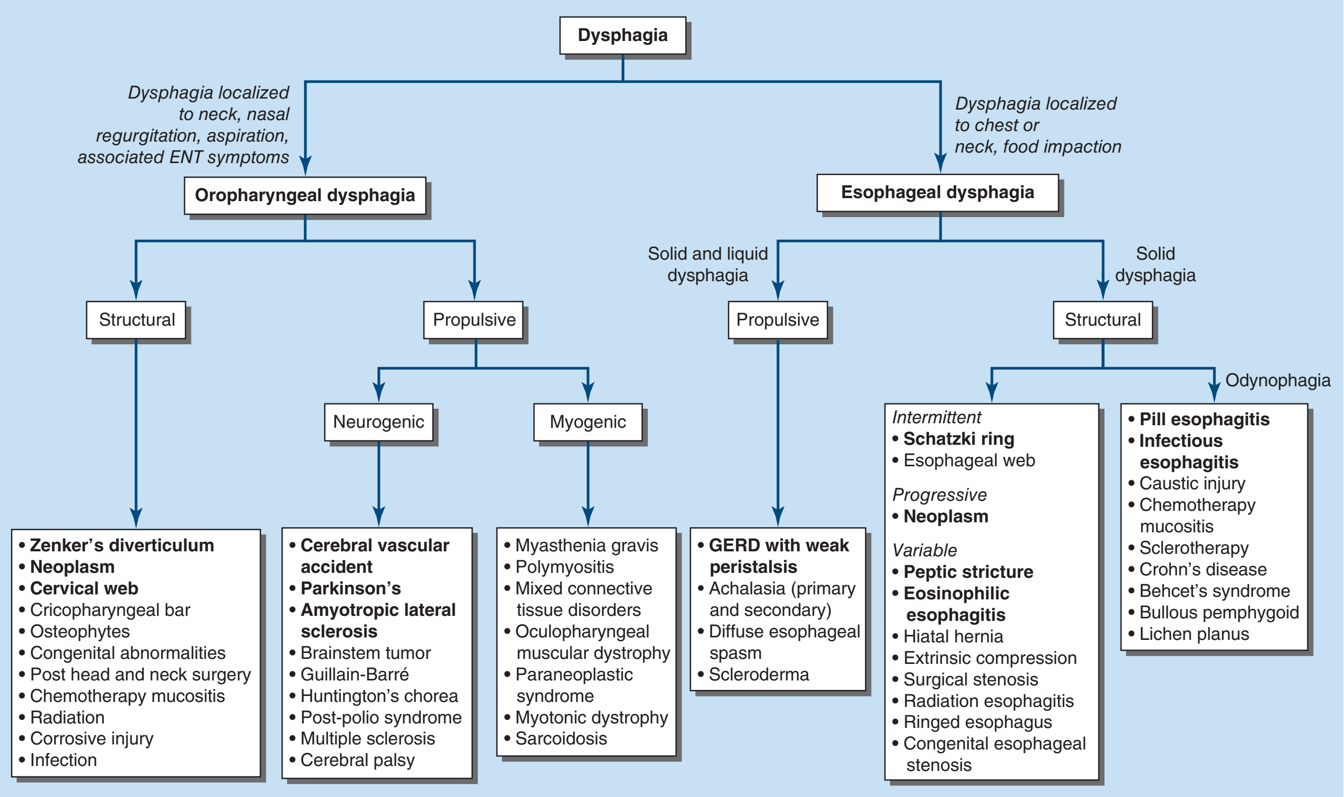
Dysphagia
Dysphagia localized to neck, nasal regurgitation, aspiration, associated ENT symptoms
Dysphagia localized to chest or neck, food impaction
Oropharyngeal dysphagia
Esophageal dysphagia
Structural
Propulsive
Solid and liquid dysphagia
Solid dysphagia
Propulsive
Structural
Neurogenic
Myogenic
Odynophagia
• **Zenker's diverticulum**
• **Neoplasm**
• **Cervical web**
• Cricopharyngeal bar
• Osteophytes
• Congenital abnormalities
• Post head and neck surgery
• Chemotherapy mucositis
• Radiation
• Corrosive injury
• Infection
• **Cerebral vascular accident**
• **Parkinson's**
• **Amyotropic lateral sclerosis**
• Brainstem tumor
• Guillain-Barré
• Huntington's chorea
• Post-polio syndrome
• Multiple sclerosis
• Cerebral palsy
• Myasthenia gravis
• Polymyositis
• Mixed connective tissue disorders
• Oculopharyngeal muscular dystrophy
• Paraneoplastic syndrome
• Myotonic dystrophy
• Sarcoidosis
• **GERD with weak peristalsis**
• Achalasia (primary and secondary)
• Diffuse esophageal spasm
• Scleroderma
Intermittent
• **Schatzki ring**
• Esophageal web
Progressive
• **Neoplasm**
Variable
• **Peptic stricture**
• **Eosinophilic esophagitis**
• Hiatal hernia
• Extrinsic compression
• Surgical stenosis
• Radiation esophagitis
• Ringed esophagus
• Congenital esophageal stenosis
• **Pill esophagitis**
• **Infectious esophagitis**
• Caustic injury
• Chemotherapy mucositis
• Sclerotherapy
• Crohn's disease
• Behcet's syndrome
• Bullous pemphygoid
• Lichen planus

FIGURE 47-2 Approach to the patient with dysphagia. Etiologies in bold print are the most common. ENT, ear, nose, and throat; GERD, gastroesophageal reflux disease.

tongue atrophy, in addition to evidence of generalized neuromuscular disease, should be elicited. The neck should be examined for thyromegaly or lymphadenopathy. A careful inspection of the mouth and pharynx should disclose inflammatory or infectious lesions. Missing dentition can interfere with mastication and exacerbate an existing cause of dysphagia. Physical examination is less helpful in the evaluation of esophageal dysphagia as the most relevant pathology is restricted to the esophagus. The notable exception is skin disease. Changes in the skin and oral mucosa may suggest a diagnosis of scleroderma or mucocutaneous diseases such as pemphigoid, lichen planus, and epidermolysis bullosa, all of which can involve the esophagus.

DIAGNOSTIC PROCEDURES

Although most instances of dysphagia are attributable to benign disease processes, dysphagia is also a cardinal symptom of several malignancies, making it an important symptom to evaluate. Cancer may result in dysphagia, most commonly as the result of intraluminal obstruction (esophageal or proximal gastric cancer, metastatic deposits) and less commonly due to extrinsic compression (lymphoma, lung cancer) or paraneoplastic syndromes. Even when not attributable to malignancy, dysphagia is usually a manifestation of an identifiable and treatable disease entity, making its evaluation beneficial to the patient and gratifying to the practitioner. The specific diagnostic algorithm to pursue is guided by the details of the history (Fig. 47-2). If oral or pharyngeal dysphagia is suspected, a fluoroscopic swallow study, usually done by a swallow therapist, is the procedure of choice. Otolaryngoscopic and neurologic evaluation also can be important, depending on the circumstances. For suspected esophageal dysphagia, upper endoscopy is the single most useful test. Endoscopy allows better visualization of mucosal lesions than does barium radiography and also allows for procurement of mucosal biopsies. Endoscopic or histologic abnormalities are evident in the leading causes of esophageal dysphagia: Schatzki's ring, gastroesophageal reflux disease, and eosinophilic esophagitis. Furthermore, therapeutic intervention with esophageal dilation can be done as part of the procedure if it is deemed necessary. The emergence of eosinophilic esophagitis as a leading cause of dysphagia in both children and adults has led to the recommendation that esophageal mucosal biopsies be obtained routinely in the evaluation of unexplained dysphagia even if characteristic, endoscopically identified esophageal mucosal features are absent. For cases of suspected esophageal motility disorders, endoscopy is still the appropriate initial evaluation as neoplastic and inflammatory conditions can secondarily produce patterns of either achalasia or esophageal spasm. Esophageal manometry is done if dysphagia is not adequately explained by endoscopy or to confirm the diagnosis of a suspected esophageal motor disorder. Barium radiography can provide useful adjunctive information in cases of subtle or complex esophageal strictures, prior esophageal surgery, esophageal diverticula, or paraesophageal herniation. Use of a barium tablet in conjunction with fluoroscopy can identify strictures and esophageal motility disorders that may be overlooked with liquid barium. In specific cases, computed tomography (CT) examination, esophageal manometry with solid meal challenge, and endoscopic ultrasonography may be useful. Impedance planimetry using the functional lumen imaging probe (FLIP) device is increasingly used in the evaluation of dysphagia, particularly for disorders of the esophagogastric junction (esophagogastric junction outflow obstruction and achalasia) **(see Chap. 334)**. Advantages of this technology include patient tolerance, as the procedure is done at the time of upper endoscopy with sedation, and, more importantly, the information regarding the dynamic opening characteristics of the esophagogastric junction in response to distension that complements esophageal manometry.

TREATMENT

Treatment of dysphagia depends on both the locus and the specific etiology. Oropharyngeal dysphagia most commonly results from functional deficits caused by neurologic disorders. In such circumstances, the treatment focuses on utilizing postures or maneuvers devised to reduce pharyngeal residue and enhance airway protection learned under the direction of a swallow therapist. Aspiration risk may be reduced by altering the consistency of ingested food and liquid. Dysphagia resulting from a cerebrovascular accident usually, but not always, spontaneously improves within the first few weeks after the event. More severe and persistent cases may require consideration of gastrostomy and enteral feeding. Patients with myasthenia gravis **(Chap. 459)** and polymyositis **(Chap. 377)** may respond to medical treatment of the primary neuromuscular disease. Surgical intervention with cricopharyngeal myotomy is usually not helpful, with the exception of specific disorders such as symptomatic cricopharyngeal bar, Zenker's diverticulum, and oculopharyngeal muscular dystrophy. Chronic neurologic disorders such as Parkinson's disease and amyotrophic lateral sclerosis may manifest with severe oropharyngeal dysphagia. Feeding by a nasogastric tube or an endoscopically placed gastrostomy tube may be considered for nutritional support; however, these maneuvers do not provide protection against aspiration of salivary secretions or refluxed gastric contents.

Treatment of esophageal dysphagia is covered in detail in **Chap. 334**. The majority of causes of structural, esophageal dysphagia are effectively managed by means of esophageal dilation using bougie or balloon dilators. Cancer and achalasia are often managed surgically, although endoscopic techniques are available for both palliation and primary therapy, respectively. Infectious etiologies respond to antimicrobial medications or treatment of the underlying immunosuppressive state. Finally, eosinophilic esophagitis is an important and increasingly recognized cause of dysphagia that is amenable to treatment by elimination of dietary allergens, proton pump inhibition, swallowed, topically acting glucocorticoids, and biologic therapies targeting cytokines involved in type 2 inflammation. While diet and medical therapies are effective at reducing dysphagia, esophageal dilation is used adjunctively for persistent strictures.

FURTHER READING

Hirano I: Esophagus: Anatomy and structural anomalies, in *Yamada Atlas of Gastroenterology*, 7th ed. New York, Wiley-Blackwell Publishing Co., 2022, pp 42–59.

Kim JP, Kahrilas PJ: How I approach dysphagia. Curr Gastroenterol Rep 21:49, 2019.

Pandolfino JE, Kahrilas PJ: Esophageal neuromuscular function and motility disorders, in *Sleisenger and Fordtran's Gastrointestinal and Liver Disease*, 11th ed, Feldman M, Friedman LS, Brandt LJ (eds). Philadelphia, Elsevier, 2020, pp 638–660.

Shaker R et al (eds): *Principles of Deglutition: A Multidisciplinary Text for Swallowing and Its Disorders*. New York, Springer, 2016.

Yadlapati R et al: The Chicago Classification of esophageal motility disorders, v4.0. Neurogastroenterol Motil 33:e14058, 2021.

48

Nausea, Vomiting, and Indigestion

William L. Hasler

Nausea is the feeling of a need to vomit. *Vomiting* (emesis) is the oral expulsion of gastrointestinal contents resulting from gut and thoracoabdominal wall contractions. Vomiting is contrasted with *regurgitation*, the effortless passage of gastric contents into the mouth. *Rumination* is the repeated regurgitation of food residue. In contrast to emesis, these phenomena exhibit volitional control. *Indigestion* broadly encompasses complaints including nausea, vomiting, heartburn, regurgitation, and dyspepsia (defined as symptoms that are thought to originate in the gastroduodenal region). Some individuals with dyspepsia experience postprandial fullness, early satiety (inability to complete a meal due to premature fullness), bloating, belching, and anorexia. Others report predominant epigastric pain or burning.

NAUSEA AND VOMITING

MECHANISMS

Vomiting is coordinated by the brainstem and is effected by responses in the gut, pharynx, and somatic musculature. Mechanisms underlying nausea likely involve the cerebral cortex, as nausea requires cognitive and emotional input and is associated with autonomic responses (diaphoresis, pallor, altered heart rate). Functional brain imaging studies support this idea showing activation of cerebral regions including the insula, anterior cingulate cortex, and amygdala during nausea.

Coordination of Vomiting Brainstem nuclei—including the nucleus tractus solitarius; dorsal vagal and phrenic nuclei; medullary nuclei regulating respiration; and nuclei that control pharyngeal, facial, and tongue movements—coordinate vomiting involving neurokinin NK$_1$, serotonin 5-HT$_3$, endocannabinoid, and vasopressin pathways.

Somatic and visceral muscles respond stereotypically during emesis. Inspiratory thoracic and abdominal wall muscles contract, increasing intrathoracic and intraabdominal pressures to help gastric evacuation. During vomiting, propulsive gastroduodenal motor activity is replaced by orally propagating retrograde contractions that facilitate expulsion of gut contents.

Activators of Emesis Emetic stimuli act at several sites. Emesis evoked by unpleasant thoughts or smells originates in the brain. Motion sickness and inner ear disorders act on labyrinthine pathways. Gastric irritants and cytotoxic agents like cisplatin stimulate vagal afferent nerves. Nongastric afferents are activated by bowel obstruction and mesenteric ischemia. The area postrema, in the medulla, responds to bloodborne stimuli (emetogenic drugs, bacterial toxins, uremia, ketoacidosis) and is termed the *chemoreceptor trigger zone*.

Neurotransmitters mediating vomiting are selective for different sites. Labyrinthine disorders stimulate vestibular muscarinic M$_1$ and histaminergic H$_1$ receptors. Vagal afferent stimuli activate 5-HT$_3$ receptors. The area postrema is served by 5-HT$_3$, M$_1$, H$_1$, and dopamine D$_2$ pathways. Central nervous system (CNS) NK$_1$ receptors mediate both nausea and vomiting. Cannabinoid CB$_1$ pathways participate in the cerebral cortex and brainstem. These receptor-mediated pathways are targets for many agents that treat nausea and vomiting.

DIFFERENTIAL DIAGNOSIS

Nausea and vomiting are caused by conditions within and outside the gut, medications, and circulating toxins (Table 48-1). In an epidemiologic study, nausea alone at least weekly was reported by 1.9% while nausea plus vomiting was noted by 1.1% of the population.

Intraperitoneal Disorders Obstruction and inflammation of hollow and solid viscera may elicit vomiting. Ulcers and malignancy cause gastric obstruction, while adhesions, masses, volvulus, intussusception, or inflammatory diseases like Crohn's disease cause small intestinal and colonic obstruction. The superior mesenteric artery syndrome, occurring after weight loss or prolonged immobilization, results when the duodenum is compressed by the overlying superior mesenteric artery. Median arcuate ligament syndrome, with celiac artery compression, is a rare cause of vomiting. Enteric infectious causes of vomiting include viruses (norovirus, rotavirus), bacteria (*Staphylococcus aureus*, *Bacillus cereus*), and opportunistic organisms like cytomegalovirus or herpes simplex in immunocompromised individuals. Abdominal irradiation impairs intestinal motility and induces strictures. Biliary colic causes nausea by acting on afferent nerves. Vomiting with pancreatitis, cholecystitis, and appendicitis results from visceral irritation and ileus.

TABLE 48-1 Causes of Nausea and Vomiting

INTRAPERITONEAL

- Obstructing disorders
  - Pyloric obstruction
  - Small-bowel obstruction
  - Colonic obstruction
  - Superior mesenteric artery syndrome
- Enteric infections
  - Viral
  - Bacterial
- Inflammatory diseases
  - Cholecystitis
  - Pancreatitis
  - Appendicitis
  - Hepatitis
- Altered sensorimotor function
  - Gastroparesis
  - Intestinal pseudoobstruction
  - Gastroesophageal reflux
  - Chronic nausea vomiting syndrome (CNVS)
  - Gastroparesis-like symptoms (GPLS)
  - Cyclic vomiting syndrome (CVS)
  - Cannabinoid hyperemesis syndrome (CHS)
  - Rumination syndrome
- Mesenteric insufficiency
  - Celiac artery stenosis
  - Median arcuate ligament syndrome
- Biliary colic
- Abdominal irradiation

EXTRAPERITONEAL

- Cardiopulmonary disease
  - Cardiomyopathy
  - Myocardial infarction
- Labyrinthine disease
  - Motion sickness
  - Labyrinthitis
  - Malignancy
- Intracerebral disorders
  - Malignancy
  - Hemorrhage
  - Abscess
  - Hydrocephalus
- Psychiatric illness
  - Anorexia and bulimia nervosa
  - Depression
- Postoperative vomiting

MEDICATIONS/ METABOLIC DISORDERS

- Drugs
  - Cancer chemotherapy
  - Opioids
  - Analgesics
  - Glucagon-like peptide-1 (GLP-1) receptor agonists
  - Oral hypoglycemics
  - Parkinson's disease/ restless legs therapies
  - Antidepressants
  - Smoking cessation agents
  - Antibiotics
  - Cardiac antiarrhythmics/ antihypertensives
  - Oral contraceptives
- Endocrine/metabolic disease
  - Pregnancy
  - Uremia
  - Ketoacidosis
  - Thyroid and parathyroid disease
  - Adrenal insufficiency
- Toxins
  - Liver failure
  - Ethanol

Gut motor and sensory dysfunction often causes nausea and vomiting. *Gastroparesis* presents most often with nausea and is documented by demonstrating delayed gastric emptying. Idiopathic gastroparesis occurring in the absence of systemic illness is the most prevalent etiology and follows a viral illness in ~15–20% of cases. Gastroparesis also occurs after vagotomy or with neoplasm, mesenteric vascular insufficiency, or organic diseases like diabetes, connective tissue diseases including scleroderma, Parkinson's disease, and amyloidosis. Rapid gastric emptying is associated with nausea and vomiting in some conditions. *Intestinal pseudoobstruction* is characterized by disrupted intestinal motility with retention of food residue and secretions; bacterial overgrowth; nutrient malabsorption; and symptoms of nausea, vomiting, bloating, pain, and altered defecation. Intestinal pseudoobstruction may be idiopathic, inherited, related to a mitochondrial disorder, result from systemic disease like scleroderma or an infiltrative process like amyloidosis, or occur as a paraneoplastic consequence of malignancy. Patients with gastroesophageal reflux disease (GERD), irritable bowel syndrome (IBS), or chronic constipation often report nausea and vomiting.

Other gastroduodenal disorders of gut-brain interaction (DGBIs) without organic abnormalities have been characterized. *Chronic nausea vomiting syndrome* is defined as bothersome nausea and/or one or more vomiting episodes at least weekly. A syndrome termed *gastroparesis-like symptoms (GPLS)* presents with symptoms indistinguishable from gastroparesis but with normal gastric emptying. *Cyclic vomiting syndrome (CVS)* presents with discrete episodes of relentless vomiting, has a prevalence of 1.4%, and is associated with migraines, autonomic dysfunction, and menstrual cycling. Some cases exhibit rapid gastric emptying. A related condition, *cannabinoid hyperemesis syndrome (CHS)*, presents with cyclical vomiting in individuals with long-standing (>1 year) use of large quantities of cannabis at least 4 days weekly and resolves with its discontinuation for ≥6 months. *Rumination syndrome* is often misdiagnosed as refractory vomiting.

Extraperitoneal Disorders Myocardial infarction and congestive heart failure may cause nausea and vomiting. Postoperative emesis occurs after 25% of surgeries. Increased intracranial pressure from tumors, bleeding, abscess, or blockage of cerebrospinal fluid outflow produces vomiting with or without nausea. Patients with anorexia nervosa, bulimia nervosa, anxiety, and depression often report nausea associated with delayed gastric emptying.

Medications and Metabolic Disorders Many medications cause nausea and vomiting including opioids, nonsteroidal anti-inflammatory drugs (NSAIDs), glucagon-like peptide-1 receptor agonists, oral hypoglycemics, antiparkinsonian drugs, agents for restless legs, antidepressants (especially selective serotonin norepinephrine reuptake inhibitors), smoking cessation drugs, antibiotics, cardiac antiarrhythmics, antihypertensives, and contraceptives. Cancer chemotherapy causes acute (within hours of administration), delayed (after ≥1 day), or anticipatory vomiting. Acute emesis from highly emetogenic agents is mediated by 5-HT$_3$ pathways while delayed emesis is dependent on NK$_1$ mechanisms. Anticipatory nausea responds better to anxiolytic therapy than antiemetics.

Metabolic disorders elicit nausea and vomiting. Nausea affects 70% of women in the first trimester of pregnancy. Hyperemesis gravidarum is a severe form of nausea of pregnancy that produces dehydration and electrolyte disturbances and may result from excessive amounts of a blood protein—growth differentiation factor 15. Uremia, ketoacidosis, adrenal insufficiency, and parathyroid and thyroid disease are other etiologies.

Circulating toxins evoke emesis via effects on the area postrema. Endogenous toxins are generated in fulminant liver failure, whereas exogenous enterotoxins may be produced by enteric bacterial infection. Ethanol intoxication is a common toxic etiology of nausea and vomiting.

APPROACH TO THE PATIENT

Nausea and Vomiting

HISTORY AND PHYSICAL EXAMINATION

The history helps define the etiology of nausea and vomiting. Drugs, toxins, and infections often cause acute symptoms, whereas established illnesses evoke chronic complaints. Gastroparesis and pyloric obstruction elicit vomiting within an hour of eating. Emesis from intestinal blockage occurs later. Vomiting occurring minutes after eating characterizes rumination syndrome. With severe gastric emptying delays, vomitus may contain food residue ingested days before. Intense episodic emesis with intervening intervals with much less severe symptoms suggests CVS or CHS. Hematemesis raises suspicion of ulcer, malignancy, or Mallory-Weiss tear. Feculent emesis is noted with distal intestinal or colonic obstruction. Bilious vomiting excludes gastric obstruction. Emesis of undigested food is consistent with a Zenker's diverticulum or achalasia. Vomiting can relieve abdominal pain from a bowel obstruction but has no effect in pancreatitis or cholecystitis. Weight loss raises concern about malignancy or ischemia. Taking prolonged hot baths or showers is associated with CHS and CVS but is less common with CNVS or gastroparesis. Intracranial sources are considered if there are headaches or visual changes. Vertigo or tinnitus indicates labyrinthine disease.

The physical examination complements the history. Orthostatic hypotension and reduced skin turgor indicate fluid loss. Pulmonary abnormalities raise concern for aspiration of vomitus. Bowel sounds are absent with ileus. High-pitched rushes suggest bowel obstruction. A succussion splash is found with gastroparesis or pyloric obstruction. Involuntary guarding raises suspicion of inflammation. Fecal blood suggests ulcer, ischemia, or tumor. Neurologic disease presents with papilledema or focal neural abnormalities. Neoplasm is suggested by palpable masses or adenopathy.

DIAGNOSTIC TESTING

For intractable symptoms or an elusive diagnosis, screening testing directs care. Electrolyte replacement is indicated for hypokalemia or metabolic alkalosis. Iron-deficiency anemia mandates exclusion of mucosal causes. Abnormal pancreatic or liver biochemistries are found with pancreaticobiliary disease. Endocrinologic, rheumatologic, or paraneoplastic etiologies are suggested by hormone or serologic abnormalities. Small-bowel obstruction is indicated by intestinal air-fluid levels and reduced colonic air on abdominal radiography, while ileus is characterized by diffusely dilated air-filled bowel loops.

Anatomic studies are performed if initial testing is nondiagnostic. Upper endoscopy detects ulcers, malignancy, and food retention in gastroparesis. Computed tomography (CT) can diagnose partial bowel obstruction. CT and magnetic resonance imaging (MRI) enterography provide detailed definition of bowel wall thickening or inflammation as seen with Crohn's disease. Ultrasound is helpful for biliary etiologies. Mesenteric angiography, CT, or MRI is useful for suspected ischemia. Brain CT or MRI delineates intracranial disease.

Gastrointestinal motility testing can detect underlying motor disorders. Gastroparesis commonly is diagnosed by gastric scintigraphy, which measures emptying of a radiolabeled meal. A nonradioactive ^{13}C-labeled gastric emptying breath test is an alternative to scintigraphy. Intestinal pseudoobstruction is suggested by luminal dilation on imaging or abnormal transit on intestinal scintigraphy or contrast radiography. Small-intestinal manometry confirms a diagnosis of pseudoobstruction and discriminates between neuropathic or myopathic disease. Nausea as a manifestation of atypical GERD can be diagnosed by esophageal pH monitoring. Combined esophageal pH/impedance testing with high-resolution manometry facilitates diagnosis of rumination syndrome. Impedance planimetry detects reduced pyloric distensibility and diameter in some cases of gastroparesis.

TREATMENT

Nausea and Vomiting

GENERAL PRINCIPLES

Therapy of vomiting is tailored to correct remediable abnormalities if possible. Patients with severe dehydration should be hospitalized if oral replenishment is unsustainable. Once oral intake is tolerated, low-fat liquid nutrients are initially restarted. Low-residue, small-particle diets have shown durable efficacy in gastroparesis. Glycemic control should be optimized in diabetic gastroparesis patients. If feasible, medications deemed to contribute to a patient's nausea should be discontinued or their doses reduced.

ANTIEMETIC MEDICATIONS

Most antiemetic agents act on CNS sites and have been evaluated for specific indications **(Table 48-2)**. Antihistamines like dimenhydrinate and meclizine and anticholinergics like scopolamine act on vestibular pathways to treat motion sickness and labyrinthine disorders. D$_2$ antagonists treat emesis evoked by area postrema stimuli including medications, toxins, and metabolic disturbances. 5-HT$_3$ antagonists like ondansetron prevent postoperative vomiting, radiation–induced symptoms, and cancer chemotherapy–induced emesis. NK$_1$ antagonists such as aprepitant and cannabinoids like dronabinol are approved for chemotherapy-induced vomiting.

TABLE 48-2 Treatment of Nausea and Vomiting

TREATMENT	MECHANISM	EXAMPLES	CLINICAL INDICATIONS
Antiemetic agents	Antihistaminergic	Dimenhydrinate, meclizine	Motion sickness, inner ear disease
	Anticholinergic	Scopolamine	Motion sickness, inner ear disease
	Antidopaminergic	Prochlorperazine, thiethylperazine, haloperidol	Medication-, toxin-, or metabolic-induced emesis, chemotherapy-induced emesis,? cannabinoid hyperemesis syndrome
	5-HT$_3$ antagonist	Ondansetron, granisetron	Chemotherapy- and radiation-induced emesis, postoperative emesis, opioid-induced nausea and vomiting
	Cannabinoids	Tetrahydrocannabinol, cannabidiol	Chemotherapy-induced emesis, gastroparesis
	Tricyclic antidepressant	Amitriptyline, nortriptyline	Chronic nausea vomiting syndrome, cyclic vomiting syndrome,? gastroparesis
	Other antidepressant/atypical antipsychotic	Mirtazapine, olanzapine	Functional dyspepsia, chemotherapy-induced emesis,? gastroparesis
	Neuropathic modulator	Gabapentin	Chemotherapy-induced emesis
	Neurokinin (NK$_1$) receptor antagonists	Aprepitant, fosaprepitant, netupitant, rolapitant	Chemotherapy-induced emesis
Prokinetic agents	5-HT$_4$ agonist and antidopaminergic	Metoclopramide	Gastroparesis
	Motilin agonist	Erythromycin	Gastroparesis,? intestinal pseudoobstruction
	Peripheral antidopaminergic	Domperidone	Gastroparesis
	Pure 5-HT$_4$ agonist	Prucalopride	Idiopathic gastroparesis
	Somatostatin analogue	Octreotide	Intestinal pseudoobstruction
	Acetylcholinesterase inhibitor	Pyridostigmine	?Small-intestinal dysmotility/pseudoobstruction
Special settings	Benzodiazepines	Lorazepam	Anticipatory nausea and vomiting with chemotherapy, cyclic vomiting syndrome
	5-HT$_{1A}$ agonist	Buspirone, tandospirone	Functional dyspepsia
	Glucocorticoids	Methylprednisolone, dexamethasone	Chemotherapy-induced emesis
	Anticonvulsants	Topiramate, zonisamide, levetiracetam	Cyclic vomiting syndrome
	Antimigraine agents	Sumatriptan	Cyclic vomiting syndrome
	Topical analgesic	Capsaicin cream	Cannabinoid hyperemesis syndrome

Note:? , indication is uncertain.

Although these drug classes have divergent mechanisms of action, they are broadly employed in a range of settings for their antinausea and antiemetic actions. Aprepitant can reduce symptoms in CVS and gastroparesis. The cannabinoid agent cannabidiol was beneficial in a controlled trial in gastroparesis.

Tricyclic antidepressants can reduce symptoms in CVS and functional causes of vomiting but did not show benefits in a controlled trial in gastroparesis. Other neuromodulators with antiemetic action in some settings include the antidepressant mirtazapine, the atypical antipsychotic olanzapine, and the pain-modulating agent gabapentin.

GASTROINTESTINAL MOTOR STIMULANTS

Drugs that stimulate gastric emptying are used for gastroparesis (Table 48-2). Metoclopramide, a 5-HT$_4$ agonist and D$_2$ antagonist, is effective in gastroparesis. Erythromycin increases gastroduodenal motility by action on receptors for motilin, a transmitter that regulates fasting motility. Erythromycin may be useful for short-term use, but its long-term benefits are limited by development of tolerance. Domperidone, a D$_2$ antagonist not available in the United States, exhibits prokinetic and antiemetic effects but does not penetrate most brain regions. Prucalopride, a 5-HT$_4$ agonist, accelerates gastric emptying and improves symptoms in idiopathic gastroparesis.

Refractory motility disorders pose challenges. Intestinal pseudoobstruction may respond to the somatostatin analogue octreotide, which induces propagative small-intestinal contractions. Acetylcholinesterase inhibitors like pyridostigmine benefit some patients with small-bowel dysmotility. Pyloric botulinum toxin injections reduced gastroparesis symptoms in uncontrolled studies, but small controlled trials observed benefits no greater than sham treatments. Surgical pyloroplasty and gastric peroral endoscopic myotomy (G-POEM) of the pylorus improved symptoms in case series and one sham-controlled trial. Enteral feedings through a jejunostomy reduce hospitalizations and improve overall health in some patients with refractory gastroparesis. The utilities of surgical gastric bypass and sleeve gastrectomy for gastroparesis are unproven. Implanted gastric electrical stimulators may reduce symptoms and health care expenditures in medication-refractory gastroparesis. A controlled trial confirmed greater improvements in vomiting during gastric electrical stimulation versus sham treatment.

SAFETY CONSIDERATIONS

Safety concerns have been raised about selected antiemetics and prokinetics. Dopamine antagonists that cross the blood-brain barrier cause anxiety, mood disturbances, movement disorders, and hyperprolactinemic effects (galactorrhea, sexual dysfunction). Metoclopramide causes irreversible movement disorders like tardive dyskinesia, particularly in older patients. This risk should be explained and documented in the medical record. Domperidone rarely causes dystonias but can induce hyperprolactinemic side effects by penetrating pituitary regions with a porous blood-brain barrier. Domperidone, erythromycin, tricyclic antidepressants, and 5-HT$_3$ antagonists increase risks of cardiac arrhythmias and sudden cardiac death in those with QTc interval prolongation on electrocardiography (ECG). Surveillance ECG testing is advocated for some agents.

OTHER CLINICAL SETTINGS

Combining a 5-HT$_3$ antagonist, an NK$_1$ antagonist, and a glucocorticoid can control acute and delayed vomiting after highly emetogenic cancer chemotherapy **(Chaps. 74 and 78)**. Anticipatory nausea and vomiting is managed with benzodiazepines like lorazepam or behavioral therapy. Other therapies that benefit chemotherapy-induced emesis include cannabinoids, olanzapine, metoclopramide, gabapentin, and alternative therapies like ginger.

Clinicians should exercise caution in managing nausea of pregnancy. Studies of teratogenic effects of antiemetic agents provide conflicting results. Antihistamines like meclizine and doxylamine, antidopaminergics like prochlorperazine, and antiserotonergics like ondansetron demonstrate limited efficacy. Some obstetricians recommend alternative therapies including pyridoxine, acupressure, or ginger.

Managing CVS and CHS is challenging. Prophylaxis with tricyclic agents or anticonvulsants (topiramate, zonisamide, levetiracetam) reduces the severity and frequency of CVS attacks in uncontrolled reports. Combining intravenous 5-HT$_3$ or NK$_1$ antagonists with the sedating effects of lorazepam is a mainstay for aborting acute flares in the emergency department. Studies report benefits with aprepitant and injectable or intranasal forms of the 5-HT$_1$ agonist sumatriptan to manage acute CVS episodes. These treatments are less effective for CHS, but intravenous or intramuscular haloperidol, topical capsaicin cream, or benzodiazepines may reduce acute CHS attacks.

INDIGESTION

MECHANISMS

Several mechanisms contribute to indigestion, including acid reflux, altered gut motility or sensation, inflammation, microbial processes, and other factors.

Gastroesophageal Reflux Gastroesophageal reflux results from many defects. Reduced lower esophageal sphincter (LES) tone causes reflux in scleroderma and pregnancy and may also be a factor in some patients without systemic illness. Other cases exhibit frequent transient LES relaxations (TLESRs). Reductions in esophageal body motility or saliva production prolong esophageal acid clearance. Increased intragastric pressure promotes gastroesophageal reflux with obesity. Large hiatal hernias can increase symptomatic acid reflux.

Gastric Motor Dysfunction Disturbed gastric motility may contribute to gastroesophageal reflux in up to one-third of cases. Delayed gastric emptying is also found in ~30% of functional dyspeptics, while rapid gastric emptying affects 5%. Impairment of gastric fundus relaxation after eating (i.e., accommodation) may underlie selected dyspeptic symptoms like bloating, nausea, and early satiety in ~40% of patients and may predispose to TLESRs and acid reflux.

Visceral Afferent Hypersensitivity Disturbed gastric sensation is another pathogenic factor in functional dyspepsia. Approximately 30% of dyspeptic patients note discomfort with gastric or duodenal distention to lower pressures than in healthy controls. Other individuals with dyspepsia exhibit hypersensitivity to chemical stimulation of the stomach and duodenum with capsaicin or with duodenal acid or lipid perfusion. Some patients with heartburn without increased reflux of acidic or nonacidic fluid exhibit heightened perception of normal esophageal acidity and are conferred a diagnosis of esophageal hypersensitivity.

Immune Activation Increases in duodenal epithelial permeability in functional dyspepsia may relate to increases in eosinophils and mast cells adjacent to submucosal neurons, most prominently in the 20% of patients who report symptom onset after a viral illness. Increased activation of these cells may contribute to gastric emptying delays and altered sensory function in functional dyspepsia and may elicit early satiety and epigastric pain. Populations of selected duodenal bacteria are altered in functional dyspepsia and correlate with symptom severity, suggesting a role for microbiome alterations. Food antigens, gluten, and fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAPs) increase duodenal inflammation.

Other Factors *Helicobacter pylori* has a proven etiologic role in peptic ulcer disease but is a minor factor in functional dyspepsia pathogenesis. Anxiety, depression, and stress play contributing roles in some functional dyspepsia cases. Other studies observe hypothalamic-pituitary-adrenal axis dysregulation. Parasympathetic and sympathetic autonomic nervous system abnormalities also have been found. Functional MRI and positron emission tomography studies show increased activation of several brain regions, emphasizing CNS contributions. Bile salt composition is abnormal in functional dyspepsia and relates to dyspeptic symptoms and gastric emptying. Analgesics cause dyspepsia, whereas nitrates, calcium channel blockers, theophylline, and progesterone promote gastroesophageal reflux. Ethanol, tobacco, and caffeine induce LES relaxation and reflux. Genetic factors can predispose to development of reflux and dyspepsia.

DIFFERENTIAL DIAGNOSIS

Gastroesophageal Reflux Disease Heartburn or regurgitation is reported weekly by 18–28% of the population, highlighting GERD prevalence. Symptoms of heartburn and regurgitation confer 70% sensitivity and specificity for a diagnosis of GERD. Other causes of these symptoms include esophageal hypersensitivity, hypervigilance, and regurgitation of nonacidic fluid.

Functional Dyspepsia Functional dyspepsia is the cause of symptoms in 70–80% of dyspeptic patients and has a population prevalence of 7.2%. The disorder is defined as bothersome postprandial fullness, early satiety, or epigastric pain or burning with symptom onset ≥6 months before diagnosis in the absence of organic cause. Functional dyspepsia is subdivided into postprandial distress syndrome (PDS) (prevalence 6.1%), characterized by meal-induced fullness and early satiety, and epigastric pain syndrome (EPS) (prevalence 2.4%), with epigastric pain or burning that may or may not be meal-related. The overlap of PDS and EPS has a prevalence of 1.3%. Functional dyspepsia is associated with other DGBIs including IBS and nongastrointestinal disorders like fibromyalgia, chronic fatigue, and anxiety.

Ulcer Disease Most GERD patients do not exhibit esophageal injury, but 5% develop esophageal ulcers. Symptoms cannot distinguish nonerosive from erosive or ulcerative esophagitis. A minority of cases of dyspepsia stem from gastric or duodenal ulcers. The most common causes of ulcers are *H. pylori* infection and NSAID use. Other rare causes of gastroduodenal ulcers include Crohn's disease **(Chap. 337)** and Zollinger-Ellison syndrome **(Chap. 335)**, resulting from gastrin overproduction by an endocrine tumor.

Malignancy Dyspeptic patients may seek care because of fear of cancer, but few cases result from malignancy. Esophageal adenocarcinoma usually complicates prolonged acid reflux. Eight to 20% of GERD patients exhibit esophageal intestinal metaplasia, termed *Barrett's metaplasia,* which predisposes to esophageal adenocarcinoma **(Chap. 85)**. Esophageal squamous cell carcinoma occurs most often with long-standing tobacco or ethanol intake. Other risks include prior caustic ingestion, achalasia, and the hereditary disorder tylosis. Gastric malignancies include adenocarcinoma, which is prevalent in certain Asian societies, and lymphoma.

Other Causes Opportunistic fungal or viral esophageal infections may produce heartburn but more often cause odynophagia. Other causes of esophageal inflammation include eosinophilic esophagitis and pill esophagitis. Biliary colic can cause unexplained chronic upper abdominal pain, but most patients report discrete acute episodes of right upper quadrant or epigastric pain rather than chronic burning or fullness. Twenty percent of gastroparesis patients note predominance of pain rather than nausea and vomiting. Intestinal lactase deficiency may cause gas, bloating, and discomfort, more commonly in blacks and Asians. Intolerance of other carbohydrates (e.g., fructose, sorbitol) produces similar symptoms. Small-intestinal bacterial overgrowth may cause dyspepsia, as well as bowel dysfunction, distention, and malabsorption. Celiac disease, nonceliac gluten sensitivity, pancreatic disease (chronic pancreatitis, malignancy), hepatocellular carcinoma, Ménétrier's disease, infiltrative diseases (sarcoidosis, mastocytosis, eosinophilic gastroenteritis), mesenteric ischemia, thyroid and parathyroid disease, and abdominal wall strain cause dyspepsia. Extraperitoneal etiologies of indigestion include congestive heart failure and tuberculosis.

APPROACH TO THE PATIENT

Indigestion

HISTORY AND PHYSICAL EXAMINATION

Managing indigestion requires a thorough interview. GERD classically produces heartburn, a substernal warmth that moves toward the neck. Heartburn often is exacerbated by meals and may awaken the patient. Associated symptoms include regurgitation of acid or nonacidic fluid and water brash, the reflex release of salty saliva into the mouth. Atypical symptoms include pharyngitis, asthma, cough, bronchitis, hoarseness, and chest pain. Some patients with acid reflux on esophageal pH testing note abdominal pain instead of heartburn. Dyspeptic patients report symptoms referable to the upper abdomen that may be meal-related or independent of food ingestion. The history in functional dyspepsia may also report symptoms of GERD or IBS.

The physical exam with GERD and functional dyspepsia usually is normal. In atypical GERD, pharyngeal erythema and wheezing may be noted. Recurrent regurgitation may cause poor dentition. Dyspeptics may exhibit epigastric tenderness or distention.

Discriminating functional from structural causes of indigestion mandates excluding certain historic and exam features. Odynophagia suggests esophageal infection. Dysphagia is concerning for esophageal blockage. Other alarm features include unexplained weight loss, recurrent vomiting, dysphagia, occult or gross bleeding, nocturnal symptoms, jaundice, palpable mass or adenopathy, fever, and a family history of gastrointestinal neoplasm. Patients with an abdominal wall source of upper abdominal pain may exhibit a positive Carnett's sign of increased tenderness with tensing of abdominal muscles upon lifting the head from the exam table.

DIAGNOSTIC TESTING

Because indigestion is prevalent and most cases result from GERD or functional dyspepsia, it is generally recommended to perform no more than limited and directed diagnostic testing in most individuals.

After excluding alarm factors **(Table 48-3)**, patients with typical GERD do not need further evaluation and are treated empirically for 4–8 weeks with single or double dosing of a proton pump inhibitor (PPI) acid suppressant. Upper endoscopy is indicated for cases with persistent symptoms, atypical presentations, or alarm factors. For heartburn >5 years in duration, especially in patients >50 years old, endoscopy is advocated to screen for Barrett's metaplasia. Endoscopy is not needed in low-risk patients who respond to acid suppressants. Up to one-third of patients with presumed GERD do not respond to PPI therapy. Ambulatory esophageal pH measurement for 48–96 hours off acid-suppressing medications using a wireless capsule endoscopically attached to the esophageal wall is considered for drug-refractory symptoms. Combined esophageal pH and impedance testing using a transnasal catheter while on PPI therapy can define if a patient with persistent or atypical symptoms has esophageal hypersensitivity or regurgitation of nonacidic or incompletely controlled acid fluid. High-resolution esophageal manometry is ordered when fundoplication is considered for treatment of GERD. Poor esophageal body peristalsis raises concern about postoperative dysphagia and directs the choice of surgical technique.

In the absence of alarm features and for patients <60 years old, assessment of *H. pylori* status by fecal antigen or urea breath testing should be performed as initial diagnostic testing for uninvestigated dyspepsia. Those who are *H. pylori* positive are given therapy to eradicate the infection. Confirmation of *H. pylori* eradication should be conducted 4–6 weeks after completing therapy. Empiric PPI therapy is reserved for those who are negative for infection or who fail to respond to *H. pylori* treatment. Upper endoscopy is advocated to exclude malignancy for patients with unexplained dyspepsia who are >60 years old, who report alarm symptoms, or who fail to respond to these therapies.

Further testing is indicated in some settings. For associated bleeding, a blood count can exclude anemia. Thyroid chemistries or calcium levels screen for metabolic etiologies. Serologies may suggest celiac disease. Pancreatic and liver chemistries are obtained for suspected pancreaticobiliary causes, which are further investigated with ultrasound, CT, or MRI. Gastric emptying testing is considered to exclude gastroparesis in patients who report symptoms resembling PDS when therapy fails. Breath testing after carbohydrate ingestion detects lactase deficiency, intolerance to other carbohydrates, or small-intestinal bacterial overgrowth.

TABLE 48-3 Alarm Symptoms in Gastroesophageal Reflux Disease

- Odynophagia or dysphagia
- Unexplained weight loss
- Recurrent vomiting
- Occult or gross gastrointestinal bleeding
- Jaundice
- Palpable mass or adenopathy
- Family history of gastroesophageal malignancy

TREATMENT

Indigestion

LIFESTYLE, DIET, AND NONMEDICATION RECOMMENDATIONS

Patients with mild indigestion are reassured that careful evaluation revealed no serious disease and are offered no other intervention. If possible, drugs that cause GERD or dyspepsia should be stopped. GERD patients should limit ethanol, caffeine, chocolate, and tobacco use and can ingest smaller low-fat meals with no snacks before bedtime, avoid tight clothing, and elevate the head of the bed. Functional dyspepsia patients are advised to reduce intake of fat, spicy foods, caffeine, and alcohol. Dietary lactose restriction is appropriate for lactase deficiency, while gluten exclusion is indicated for celiac disease. Small studies suggest benefits of low FODMAP, six-food elimination, and gluten-free diets. These findings warrant confirmation in larger functional dyspepsia cohorts.

ACID-SUPPRESSING OR -NEUTRALIZING MEDICATIONS

Drugs that reduce or neutralize gastric acid are commonly prescribed for GERD. Histamine H_2 antagonists like cimetidine and famotidine are useful in mild to moderate GERD. For severe symptoms or for erosive or ulcerative esophagitis, PPIs like omeprazole, lansoprazole, rabeprazole, pantoprazole, esomeprazole, or dexlansoprazole are needed. These drugs inhibit gastric H^+, K^+-ATPase and are more potent than H_2 antagonists. Heartburn responds better to PPI therapy than regurgitation or atypical GERD symptoms. Some individuals respond to doubling of the PPI dose or adding an H_2 antagonist. Complications of long-term PPI therapy include diarrhea (*Clostridioides difficile* infection, microscopic colitis), small-intestinal bacterial overgrowth, nutrient deficiency (vitamin B_{12}, iron, calcium), hypomagnesemia, bone demineralization, interstitial nephritis, and impaired medication absorption (clopidogrel). Many patients started on a PPI can be stepped down to an H_2 antagonist or switched to on-demand use. Normal acid exposure on 96-hour esophageal pH testing predicts successful PPI withdrawal. Vonoprazan is a new potassium-competitive acid blocker, more potent than PPI medications, that was recently approved for erosive esophagitis.

Acid suppressants are also effective for both the PDS and EPS subtypes of functional dyspepsia as initial therapy in *H. pylori*–negative patients or for persisting symptoms after *H. pylori* eradication. A meta-analysis of 18 controlled trials calculated a risk ratio of 0.88, with a 95% confidence interval of 0.82–0.94, favoring PPI therapy over placebo in functional dyspepsia. A 4- to 8-week PPI trial is advocated and should be discontinued if no benefit is observed. H_2 antagonists also improve symptoms in functional dyspepsia, but a guideline advocated PPIs over H_2 antagonists as first-line therapy.

Antacids are useful for short-term control of mild GERD but have less benefit in severe cases unless given at high doses that cause side effects (diarrhea and constipation with magnesium- and aluminum-containing agents, respectively). Alginic acid combined with antacids forms a floating barrier to reflux in patients with postprandial or nocturnal symptoms. Sucralfate, a salt of aluminum hydroxide and sucrose octasulfate that buffers acid and binds pepsin and bile salts, shows efficacy in GERD similar to H_2 antagonists.

HELICOBACTER PYLORI ERADICATION

The benefits of *H. pylori* eradication in functional dyspepsia are limited but are statistically significant with a 9% risk reduction compared to placebo. A systematic review of 25 controlled trials calculated a pooled risk ratio of 1.24, with a 95% confidence interval of 1.12–1.37, favoring *H. pylori* eradication over placebo. Most drug combinations **(Chaps. 168 and 335)** include 7–14 days of a PPI plus two or three antibiotics with or without bismuth products. *H. pylori* infection is associated with reduced prevalence of GERD. However, *H. pylori* eradication does not worsen GERD symptoms.

AGENTS THAT MODIFY GASTROINTESTINAL MOTOR ACTIVITY

The γ-aminobutyric acid B (GABA-B) agonist baclofen reduces esophageal exposure to refluxed gastric contents by reducing TLESRs by 40% and can be used in patients with refractory regurgitation of acid or nonacid fluid. Several studies promote the efficacy of agents that stimulate gastric emptying in functional dyspepsia with 33% relative risk reductions, but publication bias and small sample sizes raise questions about the true benefits of these agents. The PDS subtype may respond preferentially to such prokinetic drugs. Although the 5-HT_4 agonist prucalopride can reduce symptoms in idiopathic gastroparesis, no similar studies have been conducted in functional dyspepsia. The 5-HT_{1A} agonists buspirone and tandospirone may improve symptoms in functional dyspepsia and gastroparesis by enhancing meal-induced gastric accommodation. Acotiamide stimulates gastric emptying and augments accommodation by enhancing acetylcholine release via muscarinic receptor antagonism and acetylcholinesterase inhibition. This agent is approved for functional dyspepsia in Japan and India.

CENTRALLY ACTING NEUROMODULATORS

Some patients with refractory functional heartburn respond to tricyclic antidepressants. Their mechanism of action may involve blunting of visceral pain processing in the brain. In one controlled trial in functional dyspepsia, the tricyclic drug amitriptyline produced symptom reductions, whereas the selective serotonin reuptake inhibitor (SSRI) escitalopram had no benefit in a three-way comparison with placebo. A second study reported greater improvement with a different tricyclic agent imipramine versus placebo. In a controlled trial of a different antidepressant, mirtazapine produced superior symptom reductions in functional dyspepsia compared to placebo. In contrast, a meta-analysis of 13 trials of SSRIs and serotonin-norepinephrine reuptake inhibitors observed no benefits in this disorder.

OTHER OPTIONS

Gas and bloating may be difficult to control in patients with indigestion. Simethicone, activated charcoal, α-galactosidase, and other over-the-counter herbal products provide benefits in some cases. One trial suggested possible benefits of the nonabsorbable antibiotic rifaximin in functional dyspepsia, while another reported improvement with the probiotic *Lactobacillus gasseri*. Herbal remedies like STW 5 (a mixture of nine herbal agents) and formulations of caraway oil and menthol show efficacy in some dyspeptic patients in controlled studies. Acupuncture has shown benefits in sham-controlled trials. Psychological treatments (e.g., behavioral therapy, psychotherapy, hypnotherapy, diaphragmatic breathing, mindfulness) may be offered for refractory heartburn due to esophageal hypersensitivity or functional dyspepsia; a meta-analysis of four trials reported benefits in patients with persistent dyspepsia.

Antireflux surgery (fundoplication) to enhance the barrier function of the LES may be offered to GERD patients who are young and require lifelong therapy, have typical heartburn, are responsive to PPIs, and show acid reflux on pH monitoring. Surgery also is effective for some cases with prominent regurgitation. Factors associated with poor response to fundoplication include atypical reflux symptoms, no pathologic acid reflux on pH testing, esophageal body motor disturbances on manometry, and delayed gastric emptying. Dysphagia, gas-bloat syndrome, and gastroparesis are long-term complications of fundoplication; ~60% develop recurrent GERD symptoms over time. Magnetic sphincter augmentation may be appropriate for GERD treatment, while endoscopic transoral incisionless fundoplication can be considered for some patients.

FURTHER READING

Camilleri M et al: ACG Clinical Guideline: gastroparesis. Am J Gastroenterol 117:1197, 2022.

Frazier R et al: Diagnosis and management of cyclic vomiting syndrome: A critical review. Am J Gastroenterol 118:1157, 2023.

Hungin AP et al: Management advice for patients with reflux-like symptoms: An evidence-based consensus. Eur J Gastroenterol Hepatol 36:13, 2024.

Moshiree B, Talley NJ: Functional dyspepsia: A critical appraisal of the European consensus from a global perspective. Neurogastroenterol Motil 33:e14216, 2021.

Sperber AD et al: Worldwide prevalence and burden of functional gastrointestinal disorders, results of Rome Foundation Global Study. Gastroenterology 160:99, 2021.

49

Diarrhea and Constipation

Michael Camilleri, Joseph A. Murray

Diarrhea and constipation are exceedingly common and, together, exact an enormous toll in terms of mortality, morbidity, social inconvenience, loss of work productivity, and consumption of medical resources. Worldwide, >1 billion individuals suffer one or more episodes of acute diarrhea each year. Among the 100 million persons affected annually by acute diarrhea in the United States, nearly half must restrict activities, 10% consult physicians, ~250,000 require hospitalization, and ~5000 die (primarily the elderly). Updated 2021 annual disease burden data from the United States show 4.05 million annual clinic or emergency department visits for constipation or diarrhea or irritable bowel syndrome (IBS), and annual median economic burden to society for functional intestinal disorders of $145 million. Acute infectious diarrhea is the second leading cause of death in children, and the World Health Organization reports around 525,000 children under age 5 years die each year from acute diarrhea. Recurrent, acute diarrhea in children in tropical countries results in environmental enteropathy with long-term impacts on physical (stunted growth) and intellectual development.

Constipation, by contrast, is rarely associated with mortality and is exceedingly common in developed countries, leading to frequent self-medication and, in a third of those, to medical consultation. Annual disease burden data for 2016 show about 5 million clinic or emergency department (ED) visits for constipation or hemorrhoids, and over 190,000 ED visits for a primary diagnosis of hemorrhoids in 2018.

Population statistics on chronic diarrhea and constipation have been the subject of global studies, and in U.S. population surveys, the prevalence rates were 5.0% (95% confidence interval [CI], 4.1%–6.0%) for chronic diarrhea and 8.8% (95% CI, 7.5%–10.0%) for chronic constipation, with women affected 1.5 times as often as men. Diarrhea and constipation are among the most common patient complaints presenting in primary care and account for nearly 50% of referrals to gastroenterologists.

Although diarrhea and constipation may present as mere nuisance symptoms at one extreme, they can be severe or life threatening at the other. Even mild symptoms may signal a serious underlying gastrointestinal (GI) lesion, such as colorectal cancer, or systemic disorder, such as thyroid disease. Given the heterogeneous causes and potential severity of these common complaints, it is imperative for clinicians to appreciate the pathophysiology, etiologic classification, diagnostic strategies, and principles of management of diarrhea and constipation so that rational and cost-effective care can be delivered.

NORMAL PHYSIOLOGY

While the primary function of the small intestine is the digestion and assimilation of nutrients from food, the small intestine and colon together perform important functions that regulate the secretion and absorption of water and electrolytes, the storage and subsequent transport of intraluminal contents aborally, and the salvage of some nutrients that are not absorbed in the small intestine after bacterial metabolism of carbohydrate allows salvage of short-chain fatty acids. The main motor functions are summarized in **Table 49-1**. Alterations in fluid and electrolyte handling contribute significantly to diarrhea. Alterations in motor and sensory functions of the colon result in highly prevalent syndromes such as IBS, chronic diarrhea, and chronic constipation.

NEURAL CONTROL

The small intestine and colon have intrinsic and extrinsic innervation. The *intrinsic innervation*, also called the enteric nervous system, comprises myenteric, submucosal, and mucosal neuronal layers. The function of these layers is modulated by interneurons through the actions of neurotransmitter amines or peptides, including acetylcholine, vasoactive intestinal peptide (VIP), opioids, norepinephrine, serotonin, adenosine triphosphate (ATP), and nitric oxide (NO). The myenteric plexus regulates smooth-muscle function through intermediary pacemaker-like cells called the interstitial cells of Cajal, and the submucosal plexus affects secretion, absorption, and mucosal blood flow. The enteric nervous system receives input from the extrinsic nerves, but it is capable of independent control of these functions and peristalsis.

The *extrinsic innervations* of the small intestine and colon are part of the autonomic nervous system and also modulate motor and secretory functions. The parasympathetic nerves convey visceral sensory pathways from and excitatory pathways to the small intestine and colon. Parasympathetic fibers via the vagus nerve reach the small intestine and proximal colon along the branches of the superior mesenteric artery. The distal colon is supplied by sacral parasympathetic nerves (S_{2-4}) via the pelvic plexus; these fibers course through the wall of the colon as ascending intracolonic fibers as far as the proximal colon. The chief excitatory neurotransmitters controlling motor function are acetylcholine and the tachykinins, such as substance P. The sympathetic nerves reach the small intestine and colon along their arterial vessels. Sympathetic input to the gut is generally excitatory to sphincters and inhibitory to nonsphincteric muscle. Visceral afferents convey sensation from the gut to the central nervous system (CNS). Some afferent fibers synapse in the prevertebral ganglia and reflexly modulate intestinal motility, blood flow, and secretion.

TABLE 49-1 Normal Gastrointestinal Motility: Functions at Different Anatomic Levels

Stomach and Small Bowel
Synchronized MMC in fasting
Accommodation, trituration, mixing, transit
Stomach ~3 h
Small bowel ~3 h
Ileal reservoir empties boluses
Colon: Irregular Mixing, Fermentation, Absorption, Transit
Cecum, ascending, transverse: reservoirs
Descending: conduit
Sigmoid/rectum: volitional reservoir

Abbreviation: MMC, migrating motor complex.

INTESTINAL FLUID ABSORPTION AND SECRETION

On an average day, 9 L of fluid enter the GI tract, ~1 L of residual fluid reaches the colon, and the stool excretion of fluid constitutes about 0.2 L/d. The colon has a large capacitance and functional reserve and may recover up to four times the usual recovered volume of 0.8 L/d, provided the rate of flow permits reabsorption to occur. Thus, the colon can partially compensate for excess fluid delivery to the colon that may result from intestinal absorptive or secretory disorders.

In the small intestine and colon, sodium absorption is predominantly electrogenic (i.e., it can be measured as an ionic current across the membrane because there is not an equivalent loss of a cation from the cell), and uptake takes place at the apical membrane; it is compensated for by the export functions of the basolateral sodium pump. Several active transport proteins are at the apical membrane, especially in the small intestine, whereby sodium ion entry is coupled to monosaccharides (e.g., glucose through the transporter SGLT1 or fructose through GLUT-5). Glucose then exits the basal membrane through a specific transport protein, GLUT-2, creating a glucose concentration and osmotic gradient between the lumen and the intercellular space, drawing water and electrolytes passively from the lumen. Several channels mediate the secretion of chloride ions in diarrheal diseases or in response to medications administered for the treatment of constipation. The diverse ion channels (chloride channels and cystic fibrosis transmembrane regulator), transporters (SGLT1, GLUT-2), and receptors (e.g., guanylate cyclase C receptor) are summarized in **Fig. 49-1**.

A variety of neural and nonneural mediators regulate colonic fluid and electrolyte balance, including cholinergic, adrenergic, and serotonergic mediators. Angiotensin and aldosterone also influence colonic absorption, reflecting the common embryologic development of the distal colonic epithelium and the renal tubules.

SMALL-INTESTINAL MOTILITY

During the fasting period, the motility of the small intestine is characterized by a cyclical event called the migrating motor complex (MMC), which serves to clear nondigestible residue from the small intestine (the intestinal "housekeeper"). This organized, propagated series of contractions lasts, on average, 4 min, occurs every 60–90 min, and usually involves the entire small intestine. After food ingestion, the small intestine produces irregular, mixing contractions of relatively low amplitude, except in the distal ileum where more powerful contractions occur intermittently and empty the ileum by bolus transfers.

ILEOCOLONIC STORAGE AND SALVAGE

The distal ileum acts as a reservoir, emptying intermittently by bolus movements. This action allows time for salvage of fluids, electrolytes, and nutrients. Segmentation by haustra compartmentalizes the colon and facilitates mixing, retention of residue, and formation of solid stools. There is increased appreciation of the intimate interaction between the colonic function and the luminal ecology. The resident microorganisms, predominantly anaerobic bacteria, in the colon are necessary for the digestion of unabsorbed carbohydrates that reach the colon even in health, thereby providing a vital source of nutrients to the mucosa. Normal intestinal flora also keeps pathogens at bay by

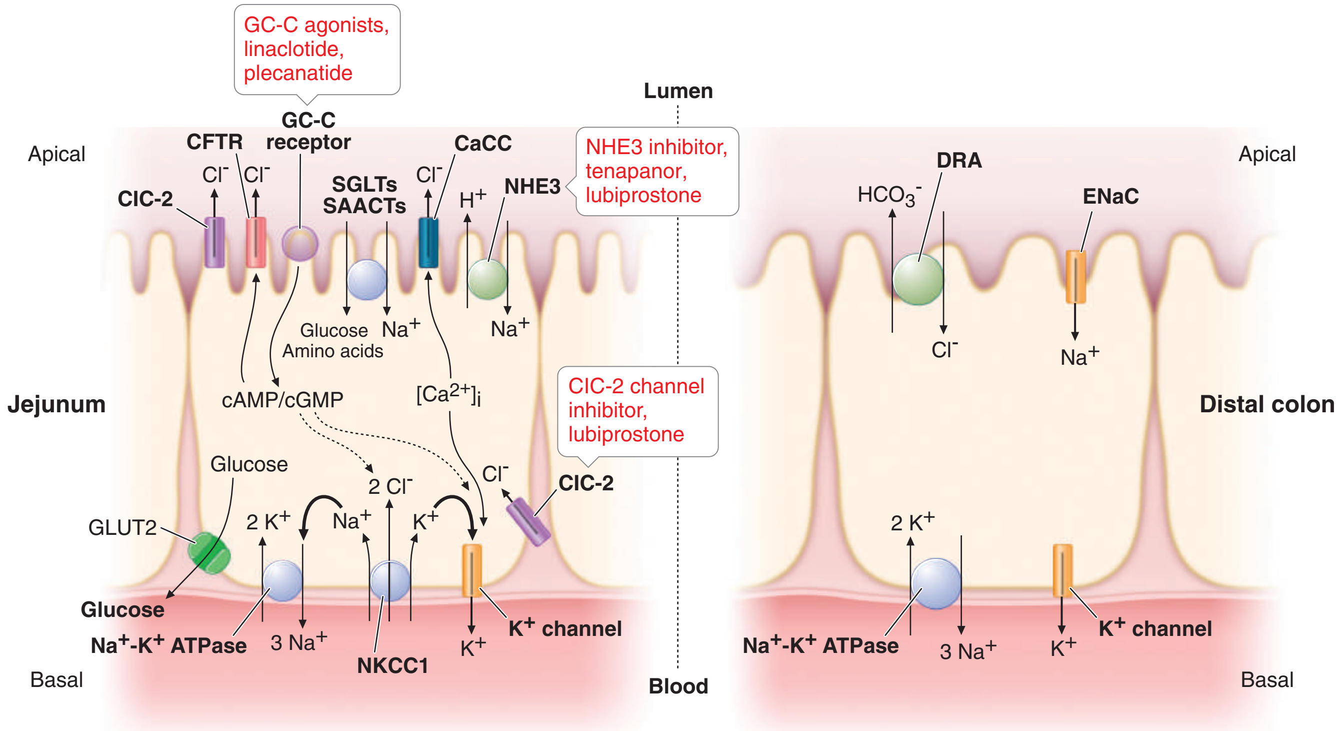
GC-C agonists,
linaclotide,
plecanatide
Lumen
GC-C
receptor
CFTR
CaCC
NHE3 inhibitor,
tenapanor,
lubiprostone
Apical
Cl⁻
Cl⁻
SGLTs
SAACTs
Cl⁻
H+
NHE3
ClC-2
DRA
HCO3⁻
ENaC
Apical
Glucose
Amino acids
Na+
Na+
Cl⁻
Na+
Jejunum
cAMP/cGMP
[Ca2+]i
ClC-2 channel
inhibitor,
lubiprostone
Distal colon
Glucose
2 Cl⁻
Cl⁻
ClC-2
GLUT2
2 K+
Na+
K+
2 K+
Glucose
Na+-K+ ATPase
3 Na+
NKCC1
K+
K+ channel
Na+-K+ ATPase
3 Na+
K+
K+ channel
Basal
Blood
Basal

FIGURE 49-1 Important ion transport mechanisms in the jejunum and colon, and the site of action of medications used as secretagogues in the treatment of chronic constipation. CFTR, cystic fibrosis transmembrane regulator; ClC2, type 2 chloride channel, DRA, downregulated in adenoma (also called SLC26A3); ENaC, epithelial sodium channel; GC-C, guanylate cyclase C; Na^+-K^+ ATPase, sodium-potassium adenosine triphosphatase; NHE3, sodium-hydrogen exchanger; NKCC1, Na-K-Cl cotransporter; SAACT, sodium amino acid co-transporters; SGLT, sodium glucose transporters.

a variety of mechanisms including a crucial role in the development and maintenance of a potent but well-regulated immune response capacity to pathogens and tolerance to normal ingesta. In health, the ascending and transverse regions of colon function as reservoirs (average transit time, 15 h), and the descending colon acts as a conduit (average transit time, 3 h). The colon is efficient at conserving sodium and water, a function that is particularly important in sodium-depleted patients in whom the small intestine alone is unable to maintain sodium balance. Diarrhea or constipation may result from alteration in the reservoir function of the proximal colon or the propulsive function of the left colon. Constipation may also result from disturbances of the rectal or sigmoid reservoir, typically as a result of dysfunction of the pelvic floor, the anal sphincters, and the coordination of defecation.

COLONIC MOTILITY AND TONE

The small-intestinal MMC only rarely continues into the colon. However, short duration or phasic antegrade and retrograde contractions mix colonic contents, and high-amplitude (>75 mmHg) propagated contractions (HAPCs) are sometimes associated with mass movements through the colon and normally occur approximately five times per day, usually on awakening in the morning and postprandially. Increased frequency of HAPCs may result in diarrhea or urgency.

Colonic tone refers to the background contractility upon which phasic contractile activity is superimposed. It is an important cofactor in the colon's capacitance (volume accommodation) and sensation.

COLONIC MOTILITY AFTER MEAL INGESTION

After meal ingestion, colonic phasic and tonic contractility increase for a period of ~2 h. The initial phase (~10 min) is mediated by the vagus nerve in response to mechanical distention of the stomach. The subsequent response of the colon requires caloric stimulation (e.g., intake of at least 500 kcal) and is mediated, at least in part, by hormones (e.g., motilin and cholecystokinin).

DEFECATION

Tonic contraction of the puborectalis muscle, which forms a sling around the rectoanal junction, is important to maintain continence; during defecation, sacral parasympathetic nerves relax this muscle, facilitating the straightening of the rectoanal angle (Fig. 49-2).

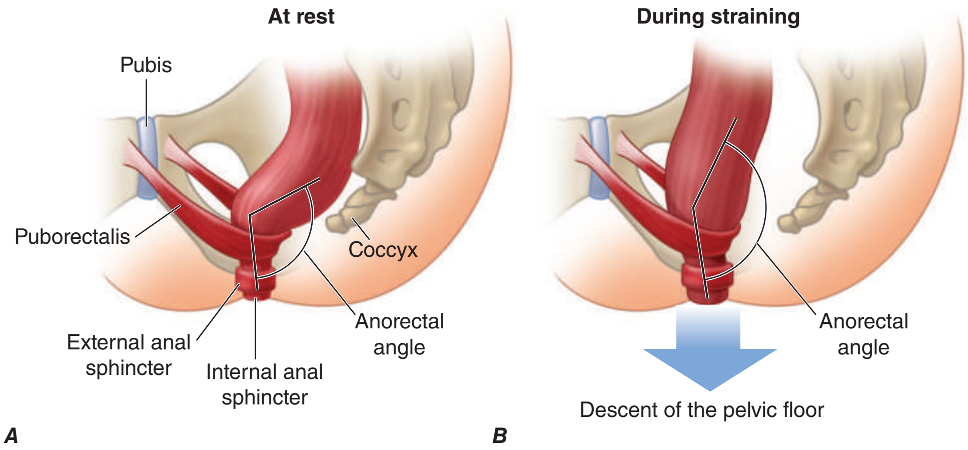
At rest
During straining
Pubis
Puborectalis
Coccyx
External anal
sphincter
Internal anal
sphincter
Anorectal
angle
Anorectal
angle
Descent of the pelvic floor
A
B

FIGURE 49-2 Sagittal view of the anorectum (*A*) at rest and (*B*) during straining to defecate. Continence is maintained by normal rectal sensation and tonic contraction of the internal anal sphincter and the puborectalis muscle, which wraps around the anorectum, maintaining an anorectal angle between 80° and 110°. During defecation, the pelvic floor muscles (including the puborectalis) relax, allowing the anorectal angle to straighten by at least 15°, and the perineum descends by 1–3.5 cm. The external anal sphincter also relaxes and reduces pressure on the anal canal. *(From The New England Journal of Medicine, A Lembo, M Camilleri; Chronic constipation, 349:1360, Copyright © 2003, Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

Distention of the rectum results in transient relaxation of the internal anal sphincter via intrinsic and reflex sympathetic innervation. As sigmoid and rectum contract, and straining (Valsalva maneuver) increases intraabdominal pressure and pressure within the rectum, the rectosigmoid angle opens by >15°. Voluntary relaxation of the external anal sphincter (striated muscle innervated by the pudendal nerve) in response to the sensation produced by distention permits the evacuation of feces. Defecation can also be delayed voluntarily by contraction of the external anal sphincter.

DIARRHEA

DEFINITION

Diarrhea is loosely defined as passage of abnormally liquid or unformed stools at an increased frequency. For adults on a typical Western diet, stool weight >200 g/d can generally be considered diarrheal. Diarrhea may be further defined as *acute* if <2 weeks, *persistent* if 2–4 weeks, and *chronic* if >4 weeks in duration.

Two common conditions, usually associated with the passage of stool totaling <200 g/d, must be distinguished from diarrhea, because diagnostic and therapeutic algorithms differ. *Pseudodiarrhea*, or the frequent passage of small volumes of stool, is often associated with rectal urgency, tenesmus, or a feeling of incomplete evacuation and accompanies IBS or proctitis. *Fecal incontinence* is the involuntary discharge of rectal contents and is most often caused by neuromuscular disorders or structural anorectal problems. Diarrhea and urgency, especially if severe, may aggravate or cause incontinence. Pseudodiarrhea and fecal incontinence occur at prevalence rates comparable to or higher than that of chronic diarrhea and should always be considered in patients complaining of "diarrhea." Overflow diarrhea may occur in nursing home patients due to fecal impaction that is readily detectable by rectal examination. A careful history and physical examination generally allow these conditions to be discriminated from true diarrhea.

ACUTE DIARRHEA

More than 90% of cases of acute diarrhea are caused by infectious agents; these cases are often accompanied by vomiting, fever, and abdominal pain. The remaining 10% or so are caused by medications, toxic ingestions, ischemia, food indiscretions, and other conditions.

Infectious Agents Domestically, infections predominate, whereas bacterial infections may be more important in travel-acquired disease. Most infectious diarrheas are acquired by fecal-oral transmission or, more commonly, via ingestion of food or water contaminated with pathogens from human or animal feces. In the immunocompetent person, the resident fecal microflora, containing >500 taxonomically distinct species, are rarely the source of diarrhea and may actually play a role in suppressing the growth of ingested pathogens. Disturbances of flora by antibiotics can lead to diarrhea by reducing the digestive function or by allowing the overgrowth of pathogens, such as *Clostridioides difficile* **(Chap. 139)**. Acute infection or injury occurs when the ingested agent overwhelms or bypasses the host's mucosal immune and nonimmune (gastric acid, digestive enzymes, mucus secretion, peristalsis, and suppressive resident flora) defenses. Established clinical associations with specific enteropathogens may offer diagnostic clues. Diarrhea occasionally is an early symptom of infection such as SARS-CoV-2 and *Legionella*.

In the United States, five high-risk groups are recognized:

1. *Travelers*. Up to 40% of U.S. tourists to endemic regions of Latin America, Africa, and Asia develop so-called traveler's diarrhea, most commonly due to enterotoxigenic or enteroaggregative *Escherichia coli* as well as to *Campylobacter*, *Shigella*, *Aeromonas*, norovirus, *Coronavirus*, and *Salmonella*. Other causative organisms in traveler's diarrhea among campers, backpackers, and swimmers are *Giardia* and *Cyclospora*. Cruise ships may be affected by outbreaks of gastroenteritis caused by agents such as norovirus.
2. *Consumers of certain foods*. Diarrhea closely following food consumption at a picnic, banquet, or restaurant may suggest infection with *Salmonella*, *Campylobacter*, or *Shigella* from chicken; enterohemorrhagic *E. coli* (O157:H7) from undercooked hamburger; *Bacillus cereus* from fried rice or other reheated food; *Staphylococcus aureus* or *Salmonella* from mayonnaise or creams; *Salmonella* from eggs; *Listeria* from fresh or frozen uncooked foods, mushrooms, or dairy products; and *Vibrio* species, *Salmonella*, or acute hepatitis A from seafood, especially if raw. State departments of public health and the National Outbreak Reporting System issue communications regarding domestic and foreign food-related illnesses, often identified by rapid DNA typing (PulseNet), that cause epidemics in the United States (e.g., the *Listeria* epidemic of 2020 from imported enoki mushrooms). Diverse foods including meat, seafood, and fresh produce and drinking water may be contaminated with *C. difficile* and the same ribotypes that actually cause the human disease. However, it has not been conclusively demonstrated that food or water actually causes community-acquired *C. difficile* infection other than a waterborne outbreak in Finland associated with contamination of a public tap water distribution system.
3. *Immunodeficient persons*. Individuals at risk for diarrhea include those with either primary immunodeficiency (e.g., IgA deficiency, common variable hypogammaglobulinemia, chronic granulomatous disease) or secondary immunodeficiency states (e.g., AIDS, senescence, pharmacologic suppression). Common enteric pathogens often cause a more severe and protracted diarrheal illness. In persons with AIDS, opportunistic infections, such as by *Mycobacterium* species, certain viruses (cytomegalovirus, adenovirus, and herpes simplex), and protozoa (*Cryptosporidium*, *Isospora belli*, microsporidia, and *Blastocystis hominis*) may also play a role **(Chap. 208)**. In patients with AIDS, agents transmitted venereally per rectum or by extension from vaginal infection (e.g., *Neisseria gonorrhoeae*, *Treponema pallidum*, *Chlamydia*) may contribute to proctocolitis. Persons with hemochromatosis are especially prone to invasive, even fatal, enteric infections with *Vibrio* species and *Yersinia* infections and should avoid raw fish and exposing open wounds to seawater.
4. *Daycare attendees and their family members*. Infections with *Shigella*, *Giardia*, *Cryptosporidium*, rotavirus, and other agents are very common.
5. *Institutionalized persons*. Infectious diarrhea is one of the most frequent nosocomial infections in hospitals and long-term care facilities; the causes are a variety of microorganisms but most commonly *C. difficile*. *C. difficile* can affect those with no history of antibiotic use and is often community acquired.

The pathophysiology underlying acute diarrhea by infectious agents results in specific clinical features that may be helpful in diagnosis **(Table 49-2)**. Profuse, watery diarrhea secondary to small-bowel hypersecretion occurs with ingestion of preformed bacterial toxins, enterotoxin-producing bacteria, and enteroadherent pathogens. Abrupt onset of diarrhea associated with marked vomiting and minimal or no fever within a few hours after ingestion suggests a toxin etiology; vomiting is usually less, abdominal cramping or bloating is greater, and fever is higher with enteroadherent pathogens. Cytotoxin-producing and invasive microorganisms all cause high fever and abdominal pain. Invasive bacteria and *Entamoeba histolytica* often cause bloody diarrhea (referred to as *dysentery*). *Yersinia* invades the terminal ileal and proximal colon mucosa and may cause especially severe abdominal pain with tenderness mimicking acute appendicitis.

Finally, infectious diarrhea may be associated with systemic manifestations. Reactive arthritis or the combination of arthritis, urethritis, and conjunctivitis (Reiter's syndrome) may accompany or follow infections by *Salmonella*, *Campylobacter*, *Shigella*, and *Yersinia*. Yersiniosis may also lead to an autoimmune-type thyroiditis, pericarditis, and glomerulonephritis. Both enterohemorrhagic *E. coli* (O157:H7) and *Shigella* can lead to the *hemolytic-uremic syndrome* with an attendant

TABLE 49-2 Association Between Pathobiology of Causative Agents and Clinical Features in Acute Infectious Diarrhea

PATHOBIOLOGY/AGENTS	INCUBATION PERIOD	VOMITING	ABDOMINAL PAIN	FEVER	DIARRHEA
Toxin producers					
Preformed toxin					
Bacillus cereus, Staphylococcus aureus, *Clostridium perfringens*	1–8 h 8–24 h	3–4+	1–2+	0–1+	3–4+, watery
Enterotoxin					
Vibrio cholerae, enterotoxigenic *Escherichia coli*, *Klebsiella pneumoniae*, *Aeromonas* species	8–72 h	2–4+	1–2+	0–1+	3–4+, watery
Enteroadherent					
Enteropathogenic and enteroadherent *E. coli*, *Giardia* organisms, cryptosporidiosis, helminths	1–8 d	0–1+	1–3+	0–2+	1–2+, watery, mushy
Cytotoxin producers					
Clostridioides difficile	1–3 d	0–1+	3–4+	1–2+	1–3+, usually watery, occasionally bloody
Hemorrhagic *E. coli*	12–72 h	0–1+	3–4+	1–2+	1–3+, initially watery, quickly bloody
Invasive organisms					
Minimal inflammation					
Rotavirus and norovirus	1–3 d	1–3+	2–3+	3–4+	1–3+, watery
Variable inflammation					
Salmonella, *Campylobacter*, and *Aeromonas* species, *Vibrio parahaemolyticus*, *Yersinia*	12 h–11 d	0–3+	2–4+	3–4+	1–4+, watery or bloody
Severe inflammation					
Shigella species, enteroinvasive *E. coli*, *Entamoeba histolytica*	12 h–8 d	0–1+	3–4+	3–4+	1–2+, bloody

Source: Adapted from DW Powell, in T Yamada (ed): *Textbook of Gastroenterology and Hepatology*, 4th ed. Philadelphia, Lippincott Williams & Wilkins, 2003.

high mortality rate. The syndrome of postinfectious IBS has now been recognized as a complication of infectious diarrhea, particularly *Campylobacter jejuni* infection. Similarly, acute gastroenteritis may precede the diagnosis of celiac disease or Crohn's disease, but the risk of developing these diseases is low, and there is no indication to screen for these diseases (e.g., by serology or endoscopy) during the episode of acute gastroenteritis. Acute diarrhea can also be a major symptom of several systemic infections including *viral hepatitis*, *listeriosis*, *legionellosis*, and *toxic shock syndrome*.

Other Causes Side effects from *medications* are probably the most common noninfectious causes of acute diarrhea, and etiology may be suggested by a temporal association between use and symptom onset. Although innumerable medications may produce diarrhea, some of the more frequently incriminated include antibiotics, cardiac antidysrhythmics, antihypertensives, nonsteroidal anti-inflammatory drugs (NSAIDs), certain antidepressants, weight loss agents, chemotherapeutic agents, bronchodilators, antacids, and laxatives.

Occlusive or nonocclusive *ischemic colitis* typically occurs in persons aged >50 years; often presents as acute lower abdominal pain preceding watery, then bloody diarrhea; and generally results in acute inflammatory changes in the sigmoid or left colon while sparing the rectum. Acute diarrhea may accompany colonic diverticulitis and graft-versus-host disease.

Acute diarrhea, often associated with systemic compromise, can follow *ingestion of toxins* including organophosphate insecticides, amanita and other mushrooms, arsenic, and preformed toxins in seafood such as ciguatera (from algae that the fish eat) and scombroid (an excess of histamine due to inadequate refrigeration). Acute allergic reaction to food ingestion, including galactose alpha-1,3-galactose (alpha gal) contained in meat, can have a similar presentation including systemic collapse, which should not be mistaken for simple dehydration secondary to diarrhea. *Conditions causing chronic diarrhea* can also be confused with acute diarrhea early in their course. This confusion may occur with inflammatory bowel disease (IBD) and some of the other inflammatory chronic diarrheas that may have an abrupt rather than insidious onset and exhibit features that mimic infection.

APPROACH TO THE PATIENT

Acute Diarrhea

The decision to evaluate acute diarrhea depends on its severity and duration and on various host factors (Fig. 49-3). Most episodes of acute diarrhea are mild and self-limited and do not justify the cost and potential morbidity rate of diagnostic or pharmacologic interventions. Indications for evaluation (stool microbiological studies) include profuse diarrhea with dehydration, overtly bloody stools, fever ≥38.5°C (≥101°F), duration >48 h without improvement, recent antibiotic use, new community outbreaks, associated severe abdominal pain in patients aged >50 years, and elderly (≥70 years) or immunocompromised patients. In some cases of moderately severe febrile diarrhea associated with fecal leukocytes (or increased fecal levels of the leukocyte proteins, such as calprotectin) or with overt blood, a diagnostic evaluation might be avoided in favor of an empirical antibiotic trial (see below).

The cornerstone of diagnosis in those suspected of severe acute infectious diarrhea acquired domestically is microbiologic analysis of the stool. Workup for a microbiologic cause now starts with culture-independent methods. These methods, relying on the identification of unique DNA sequences, are more rapid, sensitive, specific, and cost-effective. The less sensitive culture-based methods for bacteria, direct inspection for ova and parasites, and immunoassays for certain bacterial toxins (*C. difficile*), viral antigens (rotavirus), and protozoal antigens (*Giardia*, *E. histolytica*) are now primarily used when multiplexed tests are not available or for outbreak detection that still relies on isolate cultures and should be done when an outbreak is suspected or in domestically acquired severe diarrhea. Alternatively, the aforementioned clinical and epidemiologic associations may assist in focusing the evaluation. If a particular pathogen or set of possible pathogens is so implicated, either the whole panel of routine studies may not be necessary or, in some instances, special cultures may be appropriate, as for enterohemorrhagic and other types of *E. coli*, *Vibrio* species, and *Yersinia*. There is no indication for serologic or endoscopic evaluation to exclude underlying intestinal diseases in the context of acute diarrhea.

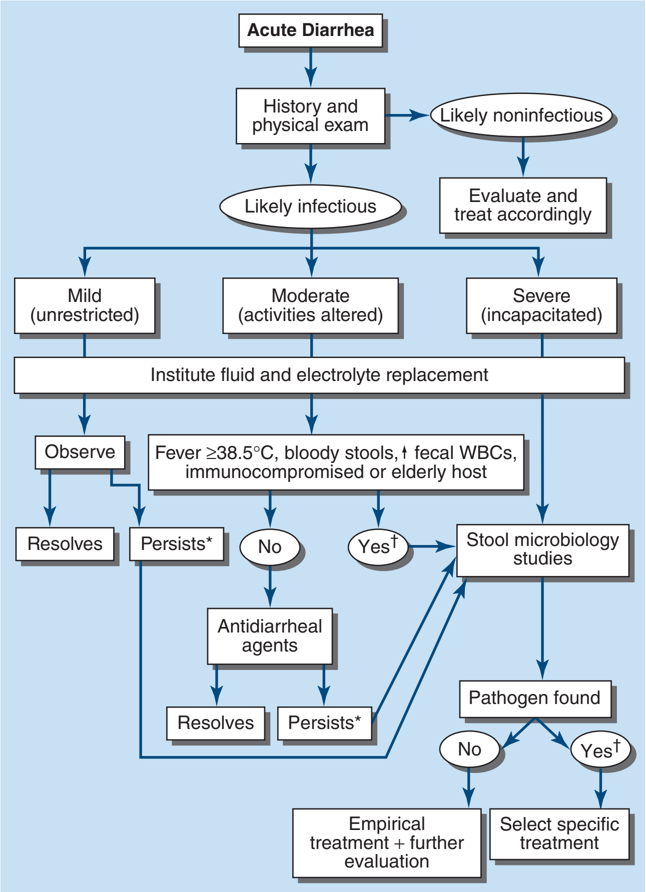
Acute Diarrhea
History and physical exam
Likely noninfectious
Evaluate and treat accordingly
Likely infectious
Mild (unrestricted)
Moderate (activities altered)
Severe (incapacitated)
Institute fluid and electrolyte replacement
Observe
Fever ≥38.5°C, bloody stools, ↑ fecal WBCs, immunocompromised or elderly host
Resolves
Persists*
No
Yes†
Stool microbiology studies
Antidiarrheal agents
Resolves
Persists*
Pathogen found
No
Yes†
Empirical treatment + further evaluation
Select specific treatment

FIGURE 49-3 Algorithm for the management of acute diarrhea. Consider empirical treatment before evaluation with (*) metronidazole and with (†) quinolone. WBCs, white blood cells.

Persistent diarrhea is commonly due to *Giardia* **(Chap. 230)**, but additional causative organisms that should be considered include *C. difficile* (especially if antibiotics had been administered), *E. histolytica*, *Cryptosporidium*, *Campylobacter*, and others. If stool studies are unrevealing and if the diarrhea continues to become chronic (>30 days), flexible sigmoidoscopy with biopsies and upper endoscopy with duodenal aspirates and biopsies may be indicated to exclude celiac or Crohn's disease, but these should not be performed during or early after acute enteritis because of risk of false-positive results. Brainerd diarrhea is an increasingly recognized entity characterized by an abrupt-onset diarrhea that persists for at least 4 weeks, but may last 1–3 years, and is thought to be of infectious origin. It may be associated with subtle inflammation of the distal small intestine or proximal colon.

Structural examination by sigmoidoscopy, colonoscopy, or abdominal computed tomography (CT) scanning (or other imaging approaches) should only be used for patients with uncharacterized persistent diarrhea when other symptoms suggest IBD or as an initial approach in patients with suspected noninfectious acute diarrhea such as might be caused by ischemic colitis, diverticulitis, or partial bowel obstruction. However, acute diarrhea may be associated with finding of air-fluid levels on abdominal radiograph, necessitating CT imaging to exclude organic obstruction.

TREATMENT

Acute Diarrhea

Fluid and electrolyte replacement are of central importance to all forms of acute diarrhea. Fluid replacement alone may suffice for mild cases. Oral sugar-electrolyte solutions (iso-osmolar sport drinks or designed formulations) should be instituted promptly with severe diarrhea to limit dehydration or in elderly or vulnerable patients, which is the major cause of death. Profoundly dehydrated patients, especially infants and the elderly, require IV rehydration.

In moderately severe nonfebrile and nonbloody diarrhea, antimotility and antisecretory agents such as loperamide can be useful adjuncts to control symptoms. Such agents should be avoided with febrile dysentery, which may be prolonged by them, and should be used with caution with drugs that increase levels due to cardiac arrhythmia (due to effect on cardiac transmembrane ion channels). Bismuth subsalicylate may reduce symptoms of vomiting and diarrhea but should not be used to treat immunocompromised patients or those with renal impairment because of the risk of bismuth encephalopathy.

Judicious use of antibiotics is appropriate in selected instances of acute diarrhea, especially in traveler's diarrhea, and may reduce its severity and duration (Fig. 49-3). Many physicians treat moderately to severely ill patients with febrile diarrheal illness empirically without diagnostic evaluation using a macrolide such as azithromycin (single dose of 500 mg) or a quinolone such as ciprofloxacin (500 mg bid for 3–5 d). Empirical treatment can also be considered for suspected giardiasis with metronidazole (250 mg qid for 7 d). Selection of antibiotics and dosage regimens are otherwise dictated by specific pathogens, geographic patterns of resistance, and conditions found **(Chaps. 138, 166, and 170–176)**. Because of resistance to first-line treatments, newer agents such as nitazoxanide (500 mg bid for 3 d) may be required for *Giardia* and *Cryptosporidium* infections. Antibiotic coverage is indicated, whether or not a causative organism is discovered, in patients who are immunocompromised, have mechanical heart valves or recent vascular grafts, or are elderly. Bismuth subsalicylate may reduce the frequency of traveler's diarrhea. Antibiotic prophylaxis is only indicated for certain patients traveling to high-risk countries in whom the likelihood or seriousness of acquired diarrhea would be especially high, including those with immunocompromise, IBD, hemochromatosis, or gastric achlorhydria. Use of ciprofloxacin, azithromycin, or rifaximin may reduce bacterial diarrhea by 90% in such travelers. Rifaximin is not suitable for invasive disease but rather as treatment for uncomplicated traveler's diarrhea. There is little role for endoscopic evaluation in most circumstances except in immunocompromised patients. Finally, physicians should be vigilant to identify if an outbreak of diarrheal illness and to alert the public health authorities promptly in order to reduce further spread.

■ CHRONIC DIARRHEA

Diarrhea lasting >4 weeks warrants evaluation to exclude serious underlying pathology. In contrast to acute diarrhea, most of the causes of chronic diarrhea are noninfectious. The classification of chronic diarrhea by pathophysiologic mechanism facilitates a rational approach to management, although many diseases cause diarrhea by more than one mechanism **(Table 49-3)**.

Secretory Causes Secretory diarrheas are due to derangements in fluid and electrolyte transport across the enterocolonic mucosa. They are characterized by watery, large-volume fecal outputs that are typically painless and persist with fasting. Because there is no malabsorbed solute, stool osmolality shows no fecal osmotic gap.

MEDICATIONS Side effects from regular ingestion of drugs and toxins are the most common secretory causes of chronic diarrhea. Hundreds of prescription and over-the-counter medications (see earlier section, "Acute Diarrhea, Other Causes") may produce diarrhea. Surreptitious or habitual use of stimulant laxatives (e.g., senna, cascara, bisacodyl, ricinoleic acid [castor oil]) must also be considered. Chronic ethanol consumption may cause a secretory-type diarrhea due to enterocyte injury with impaired sodium and water absorption, rapid transit, and other alterations. Inadvertent ingestion of certain environmental toxins (e.g., arsenic) may lead to chronic rather than acute diarrhea. Certain bacterial infections may occasionally persist and be associated with a secretory-type diarrhea. The oral angiotensin receptor blockers, including olmesartan, are associated with diarrhea due to sprue-like

TABLE 49-3 Major Causes of Chronic Diarrhea According to Predominant Pathophysiologic Mechanism

Secretory Causes
Exogenous stimulant laxatives
Chronic ethanol ingestion
Other drugs and toxins
Endogenous laxatives (dihydroxy bile acids)
Idiopathic secretory diarrhea or bile acid diarrhea
Certain bacterial infections
Bowel resection, disease, or fistula (↓ absorption)
Partial bowel obstruction or fecal impaction
Hormone-producing tumors (carcinoid, VIPoma, medullary cancer of thyroid, mastocytosis, gastrinoma, colorectal villous adenoma)
Addison's disease
Congenital electrolyte absorption defects
Osmotic Causes
Osmotic laxatives (Mg^{2+}, PO_4^{-3}, SO_4^{-2})
Lactase and other disaccharide deficiencies
Nonabsorbable carbohydrates (fructose, sorbitol, lactulose, polyethylene glycol)
Wheat and FODMAP intolerance
Steatorrheal Causes
Intraluminal maldigestion (pancreatic exocrine insufficiency, bacterial overgrowth, bariatric surgery, liver disease)
Mucosal malabsorption (celiac disease, Whipple's disease, infections, abetalipoproteinemia, ischemia, drug-induced enteropathy)
Postmucosal obstruction (1° or 2° lymphatic obstruction), amyloidosis
Inflammatory Causes
Idiopathic inflammatory bowel disease (Crohn's, chronic ulcerative colitis)
Lymphocytic and collagenous colitis
Immune-related mucosal disease (1° or 2° immunodeficiencies, food allergy, eosinophilic gastroenteritis, C1-esterase inhibitor deficiency, graft-versus-host disease)
Infections (invasive bacteria, viruses, and parasites, Brainerd diarrhea)
Radiation injury
Gastrointestinal malignancies
Dysmotile Causes
Irritable bowel syndrome (including postinfectious IBS)
Visceral neuromyopathies
Hyperthyroidism
Drugs (prokinetic agents) and poisons
Postvagotomy
Factitial Causes
Munchausen
Eating disorders
Iatrogenic Causes
Cholecystectomy
Ileal resection
Bariatric surgery
Vagotomy, fundoplication

Abbreviations: FODMAP, fermentable oligosaccharides, disaccharides, monosaccharides, and polyols; IBS, irritable bowel syndrome.

enteropathy, and more rarely, angiotensin-converting enzyme inhibitors are associated with intestinal angioedema. Diarrhea is also a relatively infrequent adverse effect of treatment with glucagon-like peptide 1 receptor agonists (e.g., semaglutide), although the mechanism in humans has not been elucidated.

BOWEL RESECTION OR BYPASS, MUCOSAL DISEASE, OR ENTEROCOLIC FISTULA These conditions may result in a secretory-type diarrhea because of inadequate surface for reabsorption of secreted fluids and electrolytes. Unlike other secretory diarrheas, this subset of conditions tends to worsen with eating. With disease (e.g., Crohn's ileitis) or resection of <100 cm of terminal ileum, dihydroxy bile acids may escape absorption and stimulate colonic secretion (cholerheic diarrhea). This mechanism may contribute to so-called *idiopathic secretory diarrhea or bile acid diarrhea (BAD)*, in which bile acids are functionally malabsorbed from a normal-appearing terminal ileum. This may account for an average of 40% of unexplained chronic diarrhea and about 1% of the general population (same as celiac disease). Reduced negative feedback regulation of bile acid synthesis in hepatocytes by fibroblast growth factor 19 (FGF-19) produced by ileal enterocytes results in a degree of bile-acid synthesis that exceeds the normal capacity for ileal reabsorption, producing BAD. BAD is characterized by severe diarrhea, urgency, and impaired quality of life in part due to the need to be close to toilet facilities. It can be positively diagnosed by elevated fasting serum $7\alpha C4$ or fecal primary or total bile acids.

Partial bowel obstruction, ostomy stricture, or fecal impaction may paradoxically lead to increased fecal output due to fluid hypersecretion.

HORMONES Although uncommon, secretory diarrhea may be mediated by hormones. *Metastatic gastrointestinal carcinoid tumors* or, rarely, *primary bronchial carcinoids* may produce watery diarrhea alone or as part of the carcinoid syndrome that comprises episodic flushing, wheezing, dyspnea, and right-sided valvular heart disease. Diarrhea is due to the release into the circulation of potent intestinal secretagogues including serotonin, histamine, prostaglandins, and various kinins. Pellagra-like skin lesions may rarely occur as the result of serotonin overproduction with niacin depletion. *Gastrinoma* typically presents with refractory peptic ulcers, but diarrhea occurs in up to one-third of cases and may be the only clinical manifestation in 10%. Diarrhea most often results from fat maldigestion owing to pancreatic enzyme inactivation by low intraduodenal pH secondary to gastric acid hypersecretion. The watery diarrhea, hypokalemia, achlorhydria syndrome, also called *pancreatic cholera*, is due to secretion of VIP and a host of other peptide hormones by a non-β cell pancreatic adenoma, referred to as a *VIPoma*. The secretory diarrhea is often massive, with stool volumes >3 L/d; daily volumes as high as 20 L have been reported. Life-threatening dehydration; neuromuscular dysfunction from associated hypokalemia, hypomagnesemia, or hypercalcemia; flushing; and hyperglycemia may accompany a VIPoma. *Medullary carcinoma of the thyroid* may present with watery diarrhea caused by calcitonin, other secretory peptides, or prostaglandins. Prominent diarrhea is often associated with metastatic disease and poor prognosis. *Systemic mastocytosis*, which may be associated with the skin lesion urticaria pigmentosa, may cause diarrhea that is either secretory and mediated by histamine or inflammatory due to intestinal infiltration by mast cells. Large *colorectal villous adenomas* may rarely be associated with a secretory diarrhea that may cause hypokalemia; these effects may be mediated by prostaglandins and can be inhibited by NSAIDs.

CONGENITAL DEFECTS IN ION ABSORPTION Rarely, defects in specific carriers associated with ion absorption cause watery diarrhea from birth. These disorders include defective Cl^-/HCO_3^- exchange (*congenital chloridorrhea*) with alkalosis (which results from a mutated *DRA* [down-regulated in adenoma] gene) and defective Na^+/H^+ exchange (*congenital sodium diarrhea*), which results from a mutation in the *NHE3* (sodium-hydrogen exchanger) gene and results in acidosis.

Some hormone deficiencies may be associated with watery diarrhea, such as occurs with adrenocortical insufficiency (Addison's disease).

Osmotic Causes Osmotic diarrhea occurs when ingested, poorly absorbable, osmotically active solutes draw enough fluid into the lumen to exceed the reabsorptive capacity of the colon. Fecal water output increases in proportion to such a solute load. Osmotic diarrhea characteristically ceases with fasting or with discontinuation of the causative agent.

OSMOTIC LAXATIVES Ingestion of magnesium-containing antacids, health supplements, or laxatives may induce osmotic diarrhea typified by a stool osmotic gap (>50 mosmol/L), which is calculated by serum osmolarity (typically 290 mosmol/kg) − (2 × [fecal sodium + potassium concentration]). Measurement of fecal osmolarity may be erroneous because carbohydrates are metabolized by colonic bacteria, causing an increase in osmolarity.

CARBOHYDRATE MALABSORPTION Carbohydrate malabsorption due to acquired or congenital defects in brush-border disaccharidases and other enzymes leads to osmotic diarrhea with a low fecal pH. One of the most common causes of chronic diarrhea in adults is *lactase deficiency*, which affects three-fourths of nonwhites worldwide and 5–30% of persons in the United States; the total lactose load at any one time influences the symptoms experienced. Most patients learn to avoid milk products without requiring treatment with enzyme supplements. Sugars such as sorbitol, lactulose, or fructose ingested with medications, gum, or high fructose–containing beverages or candies may be incompletely absorbed, and diarrhea ensues.

FODMAP AND WHEAT INTOLERANCE Chronic diarrhea, bloating, and abdominal pain are recognized as symptoms of nonceliac wheat intolerance (which is associated with impaired intestinal or colonic barrier function) and intolerance of fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAPs). The latter's effects represent the interaction between the GI microbiome and the nutrients.

Steatorrheal Causes Fat malabsorption may lead to greasy, foul-smelling, difficult-to-flush diarrhea often associated with weight loss and nutritional deficiencies due to concomitant malabsorption of amino acids and vitamins. Increased fecal output is caused by the osmotic effects of fatty acids, especially after bacterial hydroxylation, and, to a lesser extent, by the neutral fat. Whereas the normal fecal fat is <7 g/d on a 100 g/d fat diet, rapid-transit diarrhea may result in fecal fat up to 14 g/d of fat; daily fecal fat averages 15–25 g with small-intestinal diseases and is often >32 g with pancreatic exocrine insufficiency. Intraluminal maldigestion, mucosal malabsorption, or lymphatic obstruction may produce steatorrhea.

INTRALUMINAL MALDIGESTION This condition most commonly results from pancreatic exocrine insufficiency, which occurs when >90% of pancreatic secretory function is lost. *Chronic pancreatitis*, usually a sequel of ethanol abuse, most frequently causes pancreatic insufficiency. Other causes include medications, *cystic fibrosis*, *pancreatic duct obstruction*, and, rarely, *somatostatinoma*. Bacterial overgrowth in the small intestine may deconjugate bile acids and alter micelle formation, impairing fat digestion; it occurs with stasis from a blind-loop, multiple small-bowel diverticula, or dysmotility with intestinal dilatation. Finally, cirrhosis or biliary obstruction may lead to mild steatorrhea due to deficient intraluminal bile acid concentration.

MUCOSAL MALABSORPTION Mucosal malabsorption occurs from a variety of enteropathies, most commonly *celiac disease*. This gluten-sensitive enteropathy affects all ages and is characterized by villous atrophy and crypt hyperplasia in the proximal small bowel and can present with steatorrhea with multiple nutritional deficiencies of varying severity. Celiac disease, which is more frequent than previously thought, affects ~1% of the population, frequently presents without steatorrhea, can mimic IBS, and has many other GI and extraintestinal manifestations. *Tropical sprue* may produce a similar histologic and clinical syndrome but occurs in residents of or travelers to tropical climates; abrupt onset and response to antibiotics suggest an infectious etiology. *Whipple's disease*, due to the bacillus *Tropheryma whipplei* and macrophage infiltration of the small-bowel mucosa, is a far less common cause of steatorrhea that typically occurs in young or middle-aged men; it is frequently associated with arthralgias, fever, lymphadenopathy, and extreme fatigue, and it may affect the CNS and endocardium. A similar clinical and histologic picture results from *Mycobacterium avium-intracellulare* infection in patients with AIDS. *Abetalipoproteinemia* is a rare defect of chylomicron formation and fat malabsorption in children. It is associated with acanthocytic erythrocytes, ataxia, and retinitis pigmentosa. Several other conditions may cause mucosal malabsorption including infections, especially with protozoa such as *Giardia*, numerous medications (e.g., olmesartan, mycophenolate mofetil, colchicine, cholestyramine, neomycin), idiopathic enteropathies, lymphoproliferative disorders, amyloidosis, and chronic ischemia.

POSTMUCOSAL LYMPHATIC OBSTRUCTION This may result from rare *congenital intestinal lymphangiectasia* or from *acquired lymphatic obstruction* secondary to trauma, tumor, cardiac disease, or infection, and leads to the unique fat malabsorption with enteric losses of protein (often causing edema) and lymphocytopenia. Carbohydrate and amino acid absorption are preserved.

Inflammatory Causes Inflammatory diarrheas are generally accompanied by pain, fever, bleeding, or other manifestations of inflammation. The mechanism of diarrhea may not only be exudation but, depending on lesion site, may include fat malabsorption, disrupted fluid/electrolyte absorption, and hypersecretion or hypermotility from release of cytokines and other inflammatory mediators. Stool analysis reveals leukocytes or leukocyte-derived proteins such as calprotectin. With severe inflammation, exudative protein loss can lead to anasarca (generalized edema). Any middle-aged or older person with chronic inflammatory-type diarrhea, especially with blood, should be carefully evaluated (e.g., with colonoscopy) to exclude a colorectal tumor.

IDIOPATHIC INFLAMMATORY BOWEL DISEASE *Crohn's disease* and *chronic ulcerative colitis* are among the most common organic causes of chronic diarrhea in adults and range in severity from mild to fulminant and life-threatening. They may be associated with uveitis, polyarthralgias, oral ulceration, cholestatic liver disease (primary sclerosing cholangitis), and skin lesions (erythema nodosum, pyoderma gangrenosum). *Microscopic colitis*, including both lymphocytic and *collagenous colitis*, may cause chronic watery diarrhea, especially in middle-aged women and those on NSAIDs, statins, proton pump inhibitors (PPIs), and selective serotonin reuptake inhibitors (SSRIs); biopsy of an otherwise normal-appearing colon is required for histologic diagnosis. It may coexist with symptoms suggesting IBS or with celiac sprue or drug-induced enteropathy. It may be associated with BAD. It typically responds well to anti-inflammatory drugs (e.g., bismuth), the opioid agonist loperamide, or budesonide or bile acid sequestrants.

PRIMARY OR SECONDARY FORMS OF IMMUNODEFICIENCY Immunodeficiency may lead to prolonged infectious diarrhea. With selective IgA deficiency or common variable *hypogammaglobulinemia*, diarrhea is particularly prevalent and often the result of giardiasis, bacterial overgrowth, or sprue or *Strongyloides stercoralis* acquired in the tropics or subtropics.

EOSINOPHILIC GASTROENTERITIS Eosinophil infiltration of the mucosa, muscularis, or serosa at any level of the GI tract may cause diarrhea, pain, vomiting, or ascites. Affected patients often have an atopic history, Charcot-Leyden crystals due to extruded eosinophil contents may be seen on microscopic inspection of stool, and peripheral eosinophilia is present in 50–75% of patients. While hypersensitivity to certain foods occurs in adults, true food allergy causing chronic diarrhea is rare.

OTHER CAUSES Chronic inflammatory diarrhea may be caused by *radiation enterocolitis*, *chronic graft-versus-host disease*, autoimmune or idiopathic enteropathies, *Behçet's syndrome*, and *Cronkhite-Canada syndrome*, among others.

Dysmotility Causes Rapid transit may accompany many diarrheas as a secondary or contributing phenomenon, but primary dysmotility is an unusual etiology of true diarrhea. Stool features often suggest a secretory diarrhea, but mild steatorrhea of up to 14 g of fat per day can be produced by maldigestion from rapid transit alone. *Hyperthyroidism*, *carcinoid syndrome*, and certain drugs (e.g., prostaglandins, prokinetic agents, cholinesterase inhibitors) may produce hypermotility with resultant diarrhea. Primary visceral neuromyopathies or idiopathic acquired intestinal pseudoobstruction may lead to stasis with secondary bacterial overgrowth causing diarrhea. *Diabetic diarrhea*, often accompanied by peripheral and generalized autonomic neuropathies, may occur in part because of intestinal dysmotility.

The exceedingly common IBS (10% point prevalence, 1–2% per year incidence) is characterized by disturbed intestinal and colonic motor and sensory responses to various stimuli. Symptoms of increased stool frequency typically cease at night, alternate with periods of constipation, are accompanied by abdominal pain relieved with defecation, and rarely result in weight loss.

Factitial Causes Factitial diarrhea accounts for up to 15% of unexplained diarrheas referred to tertiary care centers. Either as a form of *Munchausen syndrome* (deception or self-injury for secondary gain) or *eating disorders*, some patients covertly self-administer laxatives alone or in combination with other medications (e.g., diuretics) or surreptitiously add water or urine to stool sent for analysis. Such patients are typically women, often with histories of psychiatric illness and disproportionately involved in careers in health care. Hypotension and hypokalemia are common co-presenting features. The evaluation of such patients may be difficult: contamination of the stool with water or urine is suggested by very low or high stool osmolarity, respectively. Such patients often deny this possibility when confronted, but they do benefit from psychiatric counseling when they acknowledge their behavior.

APPROACH TO THE PATIENT

Chronic Diarrhea

The laboratory tools available to evaluate the very common problem of chronic diarrhea are extensive, and many are costly and invasive. As such, the diagnostic evaluation must be rationally directed by a careful history, including medications, and physical examination **(Fig. 49-4)**. When this strategy is unrevealing, simple triage tests are often warranted to direct the choice of more complex investigations

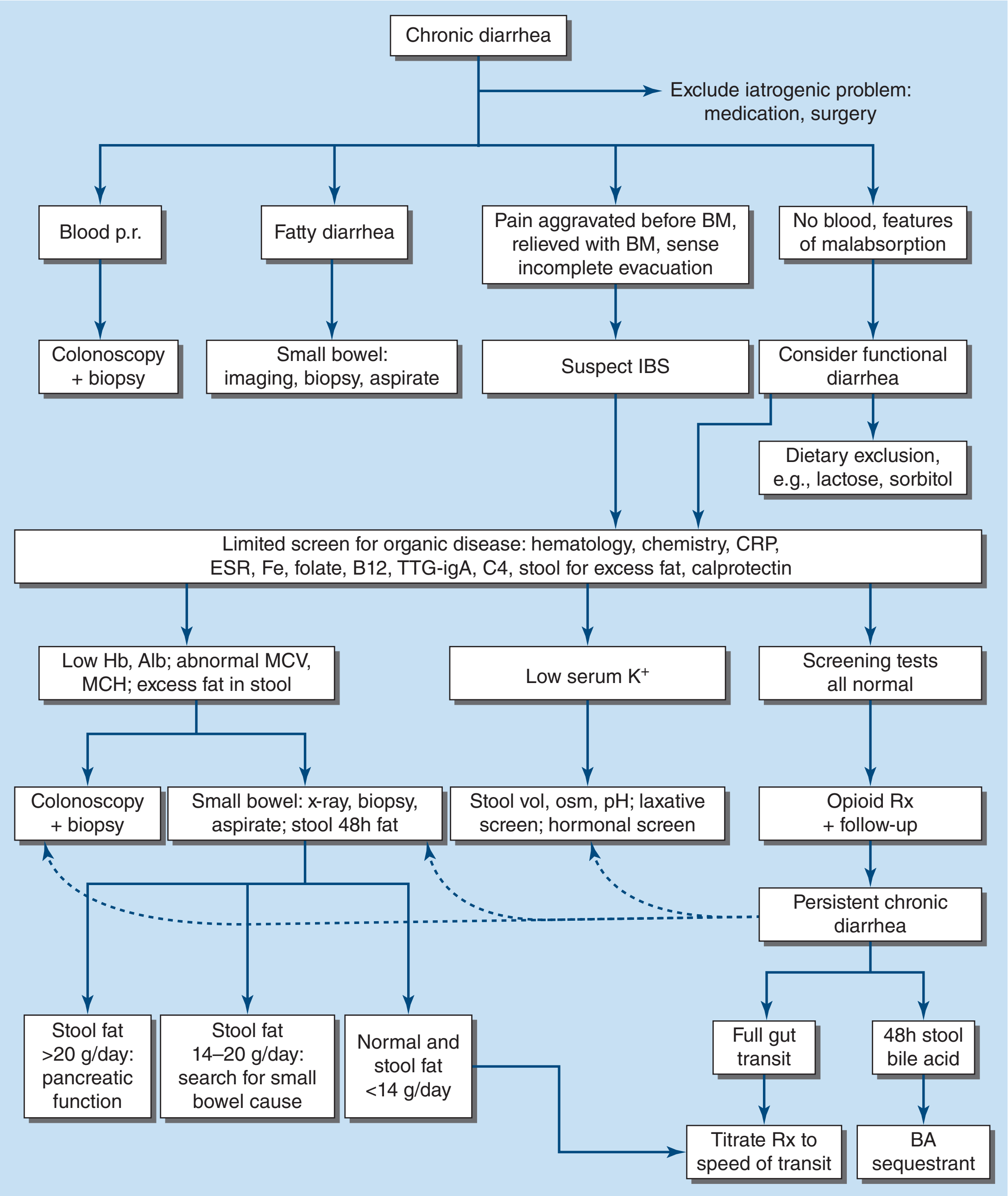
Chronic diarrhea
Exclude iatrogenic problem: medication, surgery
Blood p.r.
Fatty diarrhea
Pain aggravated before BM, relieved with BM, sense incomplete evacuation
No blood, features of malabsorption
Colonoscopy + biopsy
Small bowel: imaging, biopsy, aspirate
Suspect IBS
Consider functional diarrhea
Dietary exclusion, e.g., lactose, sorbitol
Limited screen for organic disease: hematology, chemistry, CRP, ESR, Fe, folate, B12, TTG-igA, C4, stool for excess fat, calprotectin
Low Hb, Alb; abnormal MCV, MCH; excess fat in stool
Low serum K+
Screening tests all normal
Colonoscopy + biopsy
Small bowel: x-ray, biopsy, aspirate; stool 48h fat
Stool vol, osm, pH; laxative screen; hormonal screen
Opioid Rx + follow-up
Persistent chronic diarrhea
Stool fat >20 g/day: pancreatic function
Stool fat 14–20 g/day: search for small bowel cause
Normal and stool fat <14 g/day
Full gut transit
48h stool bile acid
Titrate Rx to speed of transit
BA sequestrant

FIGURE 49-4 Algorithm for management of chronic diarrhea. Patients undergo an initial evaluation based on different symptom presentations, leading to selection of patients for imaging, biopsy analysis, and limited screens for organic diseases. Alb, albumin; BA, bile acid; BM, bowel movement; C4, 7 α-hydroxy-4-cholesten-3-one; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; Hb, hemoglobin; Hx, history; IBS, irritable bowel syndrome; MCH, mean corpuscular hemoglobin; MCV, mean corpuscular volume; osm, osmolality; p.r., per rectum; Rx, treatment; TTG, tissue transglutaminase. *(Reproduced with permission from M Camilleri, JH Sellin, KE Barrett: Pathophysiology, evaluation, and management of chronic watery diarrhea. Gastroenterology 152:515, 2017.)*

TABLE 49-4 Physical Examination in Patients with Chronic Diarrhea

1. Are there general features to suggest malabsorption or inflammatory bowel disease (IBD) such as anemia, arthritis, dermatitis herpetiformis, edema, or clubbing?
2. Are there features to suggest underlying autonomic neuropathy or collagen-vascular disease in the pupils, orthostasis, skin, hands, or joints?
3. Is there an abdominal mass or tenderness?
4. Are there any abnormalities of rectal mucosa, rectal defects, or altered anal sphincter functions?
5. Are there any mucocutaneous manifestations of systemic disease such as dermatitis herpetiformis (celiac disease), erythema nodosum (ulcerative colitis), flushing (carcinoid), or oral ulcers for IBD or celiac disease?

(Fig. 49-4). The history, physical examination **(Table 49-4)**, and routine blood studies should attempt to characterize the mechanism of diarrhea, identify diagnostically helpful associations, and assess the patient's fluid/electrolyte and nutritional status. Patients should be questioned about the onset, duration, pattern, aggravating (especially diet) and relieving factors, and stool characteristics of their diarrhea (such as relationship to food ingestion, presence of nocturnal episodes, undigested food or fat in the stool, and consistency based on Bristol Stool Form Scale). The presence or absence of fecal incontinence, fever, weight loss, pain, certain exposures (travel, medications, contacts with diarrhea), and common extraintestinal manifestations (skin changes, arthralgias, oral aphthous ulcers) should be noted. A family history of IBD or celiac disease may indicate those possibilities. Physical findings may offer clues such as a skin rash, thyroid mass, wheezing, heart murmurs, edema, hepatomegaly, abdominal masses, lymphadenopathy, mucocutaneous abnormalities, perianal fistulas, or anal sphincter laxity. Peripheral blood leukocytosis, elevated sedimentation rate, or C-reactive protein suggests inflammation; anemia reflects blood loss or nutritional deficiencies; or eosinophilia may occur with parasitoses, neoplasia, collagen-vascular disease, allergy, or eosinophilic gastroenteritis. Blood chemistries may demonstrate electrolyte, hepatic, or other metabolic disturbances. Measuring IgA tissue transglutaminase antibodies while eating gluten may help detect celiac disease. Bile acid diarrhea is confirmed by a scintigraphic radiolabeled bile acid retention test; where unavailable, a screening blood test (serum 7αC4 or FGF-19), measurement of fecal bile acids, or a therapeutic trial with a bile acid sequestrant (e.g., cholestyramine, colestipol or colesevelam) is indicated.

A therapeutic trial is often appropriate, definitive, and highly cost-effective when a specific diagnosis is suggested on the initial physician encounter. For example, chronic watery diarrhea, which ceases with fasting in an otherwise healthy young adult, may justify a trial of a lactose-restricted diet; bloating and diarrhea persisting since a mountain backpacking trip may warrant a trial of metronidazole for likely giardiasis; and postprandial diarrhea persisting following resection of terminal ileum might be due to bile acid malabsorption and be treated with cholestyramine, colestipol, or colesevelam before further evaluation. Persistent symptoms require additional investigation.

Certain diagnoses may be suggested on the initial encounter (e.g., idiopathic IBD); however, additional focused evaluations may be necessary to confirm the diagnosis and characterize the severity or extent of disease so that treatment can be best guided. Patients suspected of having IBS should be initially evaluated with flexible sigmoidoscopy with colorectal biopsies to exclude IBD, or particularly microscopic colitis, which is clinically indistinguishable from IBS with diarrhea or functional diarrhea. Patients with normal findings might be reassured and, as indicated, treated empirically with antispasmodics, antidiarrheals, or antidepressants (e.g., tricyclic agents). Any patient who presents with chronic diarrhea and hematochezia should be evaluated with stool microbiologic studies and colonoscopy.

In an estimated two-thirds of cases, the cause for chronic diarrhea remains unclear after the initial encounter, and further testing is required. Quantitative stool collection and analyses can yield important objective data that may establish a diagnosis or characterize the type of diarrhea as a triage for focused additional studies (Fig. 49-4). If stool weight is >200 g/d, additional stool analyses should be performed that might include electrolyte concentration, pH, occult blood testing, leukocyte inspection (or leukocyte protein assay), fat and bile acid quantitation, and laxative screens.

For secretory diarrheas (watery, with normal osmotic gap), possible medication-related side effects or surreptitious laxative use should be reconsidered. Stool microbiologic studies should be done with a multiplex panel initially or, if negative and diarrhea is persistent, fecal bacterial cultures (including media for *Aeromonas* and *Plesiomonas*), inspection for ova and parasites, and *Giardia* antigen assay (the most sensitive test for giardiasis). Small-bowel bacterial overgrowth can be excluded by intestinal aspirates with quantitative cultures or with glucose or lactulose breath tests involving measurement of breath hydrogen, methane, or other metabolite with an early peak in breath excretion. Note, interpretation of these breath tests may be confounded by disturbances of intestinal transit, e.g., if the gas peak is beyond 60 min. Upper endoscopy and colonoscopy with biopsies and small-bowel x-rays (formerly barium, but increasingly CT with enterography or magnetic resonance with enteroclysis) are helpful to rule out structural or occult inflammatory disease. When suggested by history or other findings, screens for peptide hormones should be pursued (e.g., serum gastrin, VIP, calcitonin, thyroid hormone/thyroid-stimulating hormone, urinary 5-hydroxyindolacetic acid, histamine).

Further evaluation of osmotic diarrhea should include tests for lactose and fructose intolerance and magnesium ingestion; intolerance of fructose is usually the result of ingestion of high fructose corn sweetener in which the fructose is ingested in the absence of sufficient glucose, which normally enhances fructose uptake. Low fecal pH suggests carbohydrate malabsorption; lactose malabsorption can be confirmed by lactose breath testing or by a therapeutic trial with lactose exclusion and observation of the effect of lactose challenge (e.g., a liter of milk). Lactase determination on small-bowel biopsy is not generally available. If fecal magnesium or laxative levels are elevated, inadvertent or surreptitious ingestion should be considered and psychiatric help should be sought.

For those with proven fatty diarrhea, endoscopy with small-bowel biopsy (including aspiration for quantitative cultures, if available) should be performed; if this procedure is unrevealing, a small-bowel radiograph is often an appropriate next step. If small-bowel studies are negative or if pancreatic disease is suspected, pancreatic exocrine insufficiency (PEI) should be excluded if possible with direct tests, such as the secretin-cholecystokinin stimulation test or a variation that could be performed endoscopically. In general, indirect tests for PEI such as assay of fecal elastase or chymotrypsin activity or a bentiromide test may be performed, but they have low sensitivity and specificity.

Chronic inflammatory-type diarrheas should be suspected by the presence of blood or leukocytes in the stool. Such findings warrant stool pathogen testing; colonoscopy with biopsies; and, if indicated, small-bowel imaging studies.

TREATMENT

Chronic Diarrhea

Treatment of chronic diarrhea depends on the specific etiology and may be curative, suppressive, or empirical. If the cause can be eradicated, treatment is curative as with resection of a colorectal cancer, antibiotic administration for Whipple's disease or tropical sprue, or discontinuation of a drug. For many chronic conditions, diarrhea can be controlled by suppression of the underlying mechanism. Examples include restriction of dietary lactose for lactase deficiency or gluten for celiac sprue, use of anti-inflammatory agents for idiopathic IBDs, bile acid sequestrants for bile acid malabsorption, PPIs for the gastric acid hypersecretion of gastrinomas, somatostatin analogues such as octreotide for malignant carcinoid syndrome, prostaglandin inhibitors such as indomethacin for medullary carcinoma of the thyroid, and pancreatic enzyme replacement for pancreatic insufficiency. When the specific cause or mechanism of chronic diarrhea evades diagnosis, empirical therapy may be beneficial. Mild opiates, such as diphenoxylate or loperamide, are often helpful in mild or moderate watery diarrhea. For those with more severe diarrhea, codeine or tincture of opium may be beneficial. Such antimotility agents should be avoided with severe IBD because toxic megacolon may be precipitated. Clonidine, an α_2-adrenergic agonist, may allow control of diabetic diarrhea, although the medication may be poorly tolerated because it causes postural hypotension; an alternative to oral administration is a clonidine skin patch. The 5-HT$_3$ receptor antagonists (e.g., alosetron, ondansetron) may relieve diarrhea and urgency in patients with IBS diarrhea. Other medications approved for the treatment of diarrhea associated with IBS are the nonabsorbed antibiotic, rifaximin, and the mixed μ-opioid receptor (OR) and κ-OR agonist and δ-OR antagonist, eluxadoline. The latter may induce sphincter of Oddi spasm and subsequent acute pancreatitis, usually in patients with prior cholecystectomy. Crofelemer, an antisecretory agent, has been used for AIDS-associated diarrhea. For all patients with chronic diarrhea, fluid and electrolyte repletion is an important component of management (see “Acute Diarrhea,” earlier). Replacement of fat-soluble vitamins may also be necessary in patients with chronic steatorrhea.

CONSTIPATION

DEFINITION

Constipation is a common complaint in clinical practice and usually refers to persistent, difficult, infrequent, or seemingly incomplete defecation. Because of the wide range of normal bowel habits, constipation is difficult to define precisely. Most persons have at least three bowel movements per week; however, low stool frequency alone is not the sole criterion for the diagnosis of constipation. Many constipated patients have a normal frequency of defecation but complain of excessive straining, hard stools, lower abdominal fullness, or a sense of incomplete evacuation. The individual patient’s symptoms must be analyzed in detail to ascertain what is meant by “constipation” or “difficulty” with defecation.

Stool form and consistency (which can be assessed using the Bristol Stool Form Scale) are well correlated with the time elapsed from the preceding defecation. Hard, pellety stools occur with slow transit, whereas loose, watery stools are associated with rapid transit. Both small pellety or very large stools are more difficult to expel than normal stools.

The perception of hard stools or excessive straining is more difficult to assess objectively, and the need for enemas or digital disimpaction is a clinically useful way to corroborate the patient’s perceptions of difficult defecation.

Psychosocial or cultural factors may also be important. A person whose parents attached great importance to daily defecation will become greatly concerned when they miss a daily bowel movement; some children withhold stool to gain attention or because of fear of pain from anal irritation; and some adults habitually ignore or delay the call to have a bowel movement.

CAUSES

Pathophysiologically, chronic constipation generally results from sedentary lifestyle, habitually ignoring the call to stool, inadequate fiber or fluid intake, or disordered colonic transit or anorectal function. These result from neurogastroenterologic disturbance, certain drugs, advancing age, or in association with a large number of systemic diseases that affect the GI tract (**Table 49-5**). Constipation of recent onset may be a symptom of significant organic disease such as colorectal cancer, anorectal irritation, or stricture. In *idiopathic constipation*, a subset of patients exhibits delayed emptying of the ascending and transverse colon with prolongation of transit (often in the proximal colon) and a reduced frequency of propulsive HAPCs. *Outlet obstruction to defecation* (also called *evacuation disorders or dyssynergic defecation*) accounts for at least a quarter of cases presenting with constipation in tertiary care and may cause delayed colonic transit, which is usually corrected by biofeedback retraining of the disordered defecation. Constipation of any cause may be exacerbated by hospitalization or chronic illnesses that lead to physical or mental impairment and result in inactivity or physical immobility.

TABLE 49-5 Causes of Constipation in Adults

TYPES OF CONSTIPATION AND CAUSES	EXAMPLES
Recent Onset	
Colonic obstruction	Neoplasm; stricture: ischemic, diverticular, inflammatory
Anal sphincter spasm	Anal fissure, painful hemorrhoids
Medications	
Chronic	
Irritable bowel syndrome	Constipation-predominant, alternating
Medications	Ca^{2+} blockers, antidepressants
Colonic pseudoobstruction	Slow-transit constipation, megacolon (rare Hirschsprung’s, Chagas’ diseases)
Disorders of rectal evacuation	Pelvic floor dysfunction; anismus; descending perineum syndrome; rectal mucosal prolapse; rectocele
Endocrinopathies	Hypothyroidism, hypercalcemia, pregnancy
Psychiatric disorders	Depression, eating disorders, drugs
Neurologic disease	Parkinsonism, multiple sclerosis, spinal cord injury
Generalized muscle disease	Progressive systemic sclerosis

APPROACH TO THE PATIENT

Constipation

A careful history should explore the patient’s symptoms and confirm whether they are indeed constipated based on frequency (e.g., fewer than three bowel movements per week), consistency (lumpy/hard), excessive straining, prolonged defecation time, or need to support the perineum or digitate the anorectum to facilitate stool evacuation. These latter items identified in the history suggest the presence of a rectal evacuation disorder. In the vast majority of cases (probably >90%), there is no underlying cause (e.g., cancer, depression, or hypothyroidism), and constipation responds to ample hydration, exercise, and supplementation of dietary fiber (15–25 g/d). A good diet and medication history and attention to psychosocial issues are key. Physical examination and, particularly, rectal examination are mandatory and should exclude fecal impaction and most of the important diseases that present with constipation and possibly indicate features suggesting an evacuation disorder (e.g., high anal sphincter tone, failure of perineal descent, or paradoxical puborectalis contraction or puborectalis tenderness during straining to simulate stool evacuation).

The presence of weight loss, rectal bleeding, or anemia with constipation mandates either flexible sigmoidoscopy plus barium enema or colonoscopy alone, to exclude structural diseases such as colorectal cancer (CRC) or strictures; however, screening should start earlier in patients with a known or suspected predisposing hereditary CRC syndrome and/or a family history of CRC. Colonoscopy alone is most cost-effective in this setting because it provides an opportunity to biopsy mucosal lesions, perform polypectomy, or dilate strictures. Barium enema has advantages over colonoscopy in the patient with isolated constipation because it is less costly and identifies colonic dilation and all significant mucosal lesions or strictures that are likely to present with constipation. Melanosis coli, or pigmentation of the colon mucosa, indicates the use of anthraquinone laxatives such as cascara or senna. An unexpected

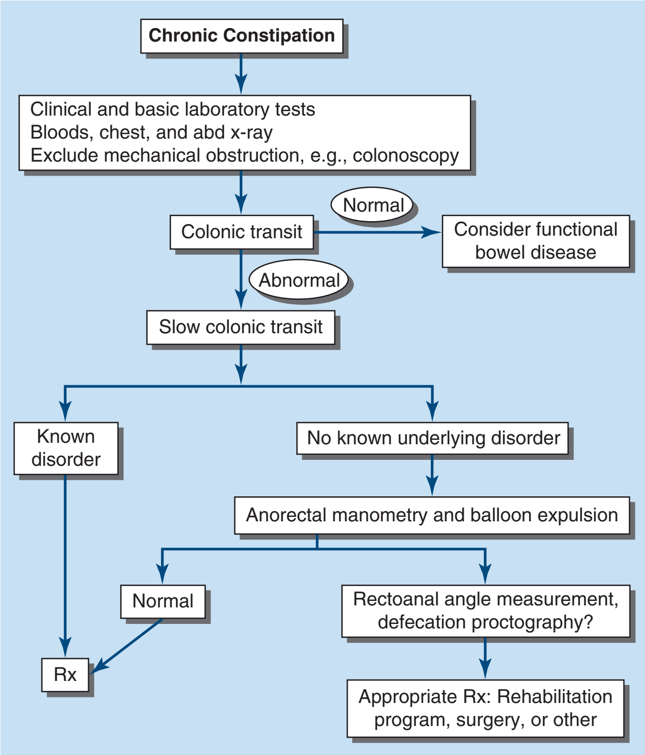
Chronic Constipation
Clinical and basic laboratory tests
Bloods, chest, and abd x-ray
Exclude mechanical obstruction, e.g., colonoscopy
Colonic transit
Normal
Consider functional bowel disease
Abnormal
Slow colonic transit
Known disorder
No known underlying disorder
Anorectal manometry and balloon expulsion
Normal
Rectoanal angle measurement, defecation proctography?
Rx
Appropriate Rx: Rehabilitation program, surgery, or other

FIGURE 49-5 Algorithm for the management of constipation. abd, abdominal; Rx, treatment.

disorder such as megacolon or cathartic colon may also be detected by colonic radiographs. Measurement of serum calcium, potassium, and thyroid-stimulating hormone levels will identify rare patients with metabolic disorders.

Patients with more troublesome constipation may not respond to fiber alone and may be helped by a bowel-training regimen, which involves taking an osmotic laxative (e.g., polyethylene glycol magnesium salts, lactulose, sorbitol) and evacuating with enema or suppository (e.g., glycerin or bisacodyl) as needed. After breakfast, a distraction-free 15–20 min on the toilet without straining is encouraged. The use of squat position may help. Excessive straining may lead to development of hemorrhoids and, if there is weakness of the pelvic floor or injury to the pudendal nerve, may result in obstructed defecation from descending perineum syndrome several years later. Those few who do not benefit from the simple measures delineated above or require long-term treatment or fail to respond to potent laxatives should undergo further investigation (Fig. 49-5). Novel agents that induce secretion (e.g., lubiprostone, a chloride channel activator, or linaclotide or plecanatide, both guanylate cyclase C agonists that activate chloride secretion, or tenapanor, an NHE3 inhibitor) are also available.

INVESTIGATION OF SEVERE CONSTIPATION

A small minority (probably <5%) of patients have severe or "intractable" constipation; at least 25% (of patients seen in gastroenterology clinics) have evacuation disorders. These are the patients most likely to require evaluation by gastroenterologists or in referral centers. Further observation of the patient may occasionally reveal a previously unrecognized cause, such as an evacuation disorder, laxative abuse, malingering, or psychological disorder. In these patients, evaluations of the physiologic function of the colon and pelvic floor and of psychological status aid in the rational choice of treatment. Even among these highly selected patients with severe constipation, a cause can be identified in only about one-third of tertiary referral patients, with the others being diagnosed with normal transit constipation. Since evacuation disorders also retard colonic transit through the left colon or the entire colon, anorectal and pelvic floor testing should precede transit measurements if there is clinical suspicion of an evacuation disorder. If an evacuation disorder is identified on testing, colonic transit may be unnecessary.

Measurement of Colonic Transit Radiopaque marker transit tests are easy, repeatable, generally safe, inexpensive, reliable, and highly applicable in evaluating constipated patients in clinical practice. Several validated methods are very simple. For example, radiopaque markers are ingested; an abdominal flat film taken 5 days later should indicate passage of 80% of the markers out of the colon without the use of laxatives or enemas. This test does not provide useful information about the transit profile of the stomach and small bowel. An alternative approach involves ingestion of 24 radiopaque markers on 3 successive days and an abdominal radiograph on the fourth day. The number of markers counted in the radiograph is an estimate of the colonic transit in hours. The collection of gas in the rectum between the level of the ischial spines and the lower border of the sacroiliac joints may suggest the presence of a rectal evacuation disorder as the cause of constipation.

Radioscintigraphy has been used to noninvasively characterize normal, accelerated, or delayed colonic function over 24–72 h with low radiation exposure. These approaches simultaneously assess gastric, small bowel (which may be important in ~20% of patients with delayed colonic transit because they reflect a more generalized GI motility disorder), and colonic transit. The disadvantages are the greater cost and the need for specific materials prepared in a nuclear medicine laboratory. Newer approaches consisting of extracorporeal detector plate, ingestible electromagnetic capsules, and analysis software are becoming available and may obviate need for approaches involving radiation.

Anorectal and Pelvic Floor Tests Pelvic floor dysfunction is suggested by the inability to completely evacuate the rectum, a feeling of persistent rectal fullness, rectal pain, the need to extract stool from the rectum digitally, application of pressure on the posterior wall of the vagina (suggestive of anterior rectocele), support of the perineum during straining, and excessive straining. These significant symptoms contrast with simple sense of incomplete rectal evacuation, which is common in IBS.

Formal psychological evaluation may identify eating disorders including ARFID (avoidant/restrictive food intake disorder), "control issues," depression, or posttraumatic stress disorders that may respond to cognitive or other intervention and may be important in restoring quality of life to patients who might present with chronic constipation.

A simple clinical test in the office to document a nonrelaxing puborectalis muscle is to have the patient strain to expel the index finger during a digital rectal examination. Motion of the puborectalis posteriorly with descent during straining indicates proper coordination of the pelvic floor muscles. Motion anteriorly with paradoxical puborectalis contraction or limited perineal descent (<1.5 cm) during simulated evacuation indicates pelvic floor dysfunction.

Measurement of perineal descent is relatively easy to gauge clinically by placing the patient in the left decubitus position and watching the perineum to detect inadequate descent (<1.5 cm, a sign of pelvic floor dysfunction) or perineal ballooning during straining relative to bony landmarks (>4 cm, suggesting excessive perineal descent).

A useful overall test of evacuation is the balloon expulsion test. A balloon-tipped urinary catheter is placed and inflated with 50 mL of water. Normally, a patient can expel it while seated on a toilet or in the left lateral decubitus position. In the lateral position, the weight needed to facilitate expulsion of the balloon is determined; normally, expulsion occurs unaided within 1 min.

Anorectal manometry, when used in the evaluation of patients with severe constipation, may find an excessively high resting (>80 mmHg) or squeeze anal sphincter tone, suggesting anismus (anal sphincter spasm). The rectoanal pressure difference along with a balloon expulsion time >60 s is consistently identified as the most useful indicators of evacuation disorders. This test also identifies rare syndromes, such as Hirschsprung's disease, by absence of the rectoanal inhibitory reflex (RAIR). However, the RAIR may be absent in patients with enlarged rectum due to prolonged retention of stool.

Defecography (a dynamic barium enema including lateral views obtained during barium expulsion or a magnetic resonance defecogram) measures changes in rectoanal angle, degree of pelvic floor relaxation and perineal descent, as well as anatomic defects of the rectum such as internal mucosal prolapse, intussusception, and enteroceles or rectoceles. Surgically remediable conditions are identified in only a few patients with severe, whole-thickness intussusception or extremely large rectocele that fills preferentially during attempts at defecation instead of expulsion through the anus. Magnetic resonance defecography provides more information about the structure of the pelvic floor, distal colorectum, and anal sphincters and some information about the function during defecatory effort; however, it is performed in supine rather than seated position.

Neurologic testing (electromyography) is rarely performed in the evaluation of patients with constipation. The absence of neurologic signs in the lower extremities suggests that any documented denervation of the puborectalis results from pelvic (e.g., obstetric) injury or from stretching of the pudendal nerve by chronic, long-standing straining. Constipation is common among patients with spinal cord injuries, neurologic diseases such as Parkinson's disease, multiple sclerosis, and diabetic neuropathy.

Spinal-evoked responses during electrical rectal stimulation or stimulation of external anal sphincter contraction by applying magnetic stimulation over the lumbosacral cord identify patients with limited sacral neuropathies with sufficient residual nerve conduction to attempt biofeedback training.

In summary, a balloon expulsion test is an important component of anorectal manometry as screening tests for anorectal dysfunction. An anatomic evaluation of the rectum or anal sphincters is rarely required when evaluating patients in whom obstructed defecation is suspected. Indications for anatomic evaluation are symptoms of rectal mucosal prolapse, pressure of the posterior wall of the vagina to facilitate defecation (suggestive of anterior rectocele), or prior pelvic surgery that may be complicated by enterocele.

TREATMENT

Constipation

After the cause of constipation is characterized, a treatment decision can be made. Slow-transit constipation requires aggressive medical or surgical treatment; conversely, anismus or pelvic floor dysfunction usually responds to biofeedback management (Fig. 49-5). The remaining ~60% of patients with constipation have normal colonic transit and can be treated symptomatically. Patients with spinal cord injuries or other neurologic disorders require a dedicated bowel regimen that often includes rectal stimulation, enema therapy, and carefully timed laxative therapy.

Patients with constipation are treated with bulk (fiber, psyllium), osmotic (milk of magnesia, lactulose, polyethylene glycol), secretory (linaclotide, lubiprostone, plecanatide, tenapanor), and prokinetic or stimulant laxatives (including diphenyl methanes such as bisacodyl and sodium picosulfate, senna alkaloids, and the 5-HT_4 agonist prucalopride). If a 3- to 6-month trial of medical therapies fails, unassociated with obstructed defecation, the patient should be considered for laparoscopic colectomy with ileorectostomy; however, this should not be undertaken for pain or in the presence of continued evidence of an evacuation disorder or a generalized GI dysmotility. Referral to a specialized center for further tests of colonic motor function is warranted. The decision to resort to surgery is facilitated by the presence of megacolon and megarectum or documented sigmoid or cecal volvulus. The complications after surgery include small-bowel obstruction (11%) and fecal soiling, particularly at night during the first postoperative year. Frequency of defecation is 3–8 per day during the first year, dropping to 1–3 per day from the second year after surgery.

Patients who have a combined (evacuation and transit/motility) disorder should first pursue pelvic floor retraining (biofeedback and muscle relaxation), psychological counseling, and dietetic advice. If symptoms are intractable despite biofeedback and optimized medical therapy, colectomy and ileorectostomy could be considered as long as the evacuation disorder is resolved and optimized medical therapy is unsuccessful. In patients with pelvic floor dysfunction alone, biofeedback training for 3 months has a 70–80% success rate, measured by the acquisition of comfortable stool habits. Attempts to manage pelvic floor dysfunction with operations (internal anal sphincter or puborectalis muscle division) or injections with botulinum toxin have achieved only mediocre success and have been largely abandoned. Persistence of the evacuation disorder despite physical therapy with biofeedback may require regular enemas or consideration of a loop- or end-ileostomy to relieve the intractable constipation.

FURTHER READING

Assi R et al: Sexually transmitted infections of the anus and rectum. World J Gastroenterol 20:15262, 2014.

Blackett JW et al: Comparison of anorectal manometry, rectal balloon expulsion test, and defecography for diagnosing defecatory disorders. Gastroenterology 163:1582, 2022.

Boeckxstaens G et al: Fundamentals of neurogastroenterology: Physiology/motility—sensation. Gastroenterology 150:1292, 2016.

BouSaba J et al: Impact of bile acid diarrhea in patients with diarrhea-predominant irritable bowel syndrome on symptoms and quality of life. Clin Gastro Hepatol 20:2083, 2022.

Camilleri M et al: Chronic constipation. Nat Rev Dis Primers 3:17095, 2017.

Camilleri M et al: Pathophysiology, evaluation, and management of chronic watery diarrhea. Gastroenterology 152:515, 2017.

Issaka RB et al: AGA clinical practice update on risk stratification for colorectal cancer screening and post-polypectomy surveillance: Expert review. Gastroenterology 165:1280, 2023.

Peery AF et al: Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: Update 2021. Gastroenterology 162:621, 2022.

Palsson OS et al: Prevalence and associated factors of disorders of gut-brain interaction in the United States: Comparison of two nationwide internet surveys. Neurogastro Motil 35:e14564, 2023.

Riddle MS et al: ACG Clinical Guideline: Diagnosis, treatment, and prevention of acute diarrheal infections in adults. Am J Gastroenterol 111:602, 2016.

Rubio-Tapia A et al: American College of Gastroenterology guidelines update: Diagnosis and management of celiac disease. Am J Gastroenterol 118:59, 2023.

Sanchez DA et al: Characterization of infectious and non-infectious gastrointestinal disease in common variable immunodeficiency: Analysis of 114 patient cohort. Front Immunol 14:1209570, 2023.

Smalley W et al: AGA Clinical Practice Guidelines on the laboratory evaluation of functional diarrhea and diarrhea-predominant irritable bowel syndrome in adults (IBS-D). Gastroenterology 157:851, 2019.

50

Unintentional Weight Loss

J. Larry Jameson

Involuntary or unintentional weight loss (UWL) is frequently insidious and can have important implications, often serving as a harbinger of serious underlying disease. Clinically important weight loss is defined as the loss of >5% of body weight over a period of 6–12 months. UWL is not uncommon in individuals aged ≥65 years. It can be challenging to recognize in patients with preexisting obesity or inadequate documentation of previous weights. There is no identifiable cause in up to one-quarter of patients despite extensive investigation. People with no known cause of weight loss generally have a better prognosis than do those with known causes, particularly when the source is neoplastic. Weight loss in older persons is associated with a variety of deleterious effects, including falls and fractures, pressure ulcers, impaired immune function, and decreased functional status. Not surprisingly, significant weight loss is associated with increased mortality within 1–2 years.

■ PHYSIOLOGY OF WEIGHT REGULATION WITH AGING

(See also Chaps. 413 and 488) Among healthy aging people, total body weight peaks in the sixth decade of life and generally remains stable until the ninth decade, after which it gradually falls. In contrast, lean body mass (fat-free mass) begins to decline at a rate of 0.3 kg per year in the third decade, and the rate of decline increases further beginning at age 60 in men and age 65 in women. These changes in lean body mass largely reflect the age-dependent decline in growth hormone secretion and, consequently, circulating levels of insulin-like growth factor type I (IGF-I) that occur with normal aging. Loss of sex steroids, at menopause in women and more gradually in men, also contributes to these changes in body composition. In the healthy elderly, an increase in fat tissue balances the loss in lean body mass until very old age, when loss of both fat and skeletal muscle occurs. Age-dependent changes also occur at the cellular level. Telomeres shorten, and body cell mass—the fat-free portion of cells—declines steadily with aging.

Between ages 20 and 80, mean energy intake is reduced by up to 1200 kcal/d in men and 800 kcal/d in women. Decreased hunger is a reflection of reduced physical activity and loss of lean body mass, producing lower demand for calories and food intake. Several important age-associated physiologic changes also predispose elderly persons to weight loss, such as declining chemosensory function (smell and taste), reduced efficiency of chewing, slowed gastric emptying, and alterations in the neuroendocrine axis, including changes in levels of leptin, cholecystokinin, neuropeptide Y, and other hormones and peptides. These changes are associated with early satiety and a decline in both appetite and the hedonistic appreciation of food. Collectively, they contribute to the "anorexia of aging." As noted below, these physiologic changes with aging may be accompanied by social isolation, poverty, and immobility, further contributing to undernutrition.

■ CAUSES OF UNINTENTIONAL WEIGHT LOSS

Most causes of UWL belong to one of four categories: (1) malignant neoplasms, (2) chronic inflammatory or infectious diseases, (3) metabolic disorders (e.g., hyperthyroidism and diabetes), or (4) psychiatric disorders **(Table 50-1)**. Not infrequently, more than one of these causes can be responsible for UWL. Depending upon patient populations, UWL is caused by malignant disease in a quarter of patients and by organic disease in one-third, with the remainder due to psychiatric disease, medications, or uncertain causes. Risk factors for undiagnosed cancer include a history of smoking, particularly for men, localizing symptoms, and abnormal laboratory tests.

The most common malignant causes of UWL are gastrointestinal, hepatobiliary, hematologic, lung, breast, genitourinary, ovarian, and prostate. Half of all patients with cancer lose some body weight; one-third lose more than 5% of their original body weight, and up to 20% of all cancer deaths are caused directly by cachexia (through immobility and/or cardiac/respiratory failure). The greatest incidence of weight loss is seen among patients with solid tumors. Malignancy that reveals itself through significant weight loss usually has a very poor prognosis.

In addition to malignancies, gastrointestinal diseases are among the most prominent causes of UWL. Peptic ulcer disease, inflammatory bowel disease, dysmotility syndromes, chronic pancreatitis, celiac disease, constipation, and atrophic gastritis are some of the more common entities. Oral and dental problems are easily overlooked and may manifest with halitosis, poor oral hygiene, xerostomia, inability to chew, reduced masticatory force, nonocclusion, temporomandibular joint syndrome, edentulousness, and pain due to caries or abscesses.

TABLE 50-1 Causes of Involuntary Weight Loss

Cancer
- Upper gastrointestinal
- Lung
- Colon
- Hepatobiliary
- Hematologic
- Breast
- Genitourinary
- Ovarian
- Prostate

Gastrointestinal disorders
- Difficulty swallowing
- Malabsorption
- Peptic ulcer
- Inflammatory bowel disease
- Pancreatitis
- Obstruction/constipation
- Pernicious anemia

Endocrine and metabolic
- Hyperthyroidism
- Diabetes mellitus
- Pheochromocytoma
- Adrenal insufficiency

Cardiac disorders
- Chronic ischemia
- Chronic congestive heart failure

Respiratory disorders
- Emphysema
- Chronic obstructive pulmonary disease

Renal insufficiency

Rheumatologic disease

Infections
- HIV
- Tuberculosis
- Parasitic infection
- Subacute bacterial endocarditis

Medications
- Sedatives
- Antibiotics
- Nonsteroidal anti-inflammatory drugs
- Serotonin reuptake inhibitors
- Metformin
- Levodopa
- Angiotensin-converting enzyme inhibitors
- Other drugs

Disorders of the mouth and teeth
- Dental caries
- Dysgeusia

Age-related factors
- Physiologic changes
- Visual impairment
- Decreased taste and smell
- Functional disabilities

Neurologic
- Stroke
- Parkinson's disease
- Neuromuscular disorders
- Dementia

Social
- Isolation
- Poverty

Psychiatric and behavioral
- Depression
- Anxiety
- Paranoia
- Bereavement
- Alcoholism
- Eating disorders
- Increased activity or exercise

Idiopathic

Tuberculosis, fungal diseases, parasites, subacute bacterial endocarditis, and HIV are well-documented causes of UWL. Cardiovascular and pulmonary diseases cause UWL through increased metabolic demand and decreased appetite and caloric intake. Repeated surgeries may lead to weight loss because of reduced caloric intake and increased metabolic demands resulting from a systemic inflammatory response. Uremia produces nausea, anorexia, and vomiting. Connective tissue diseases may increase metabolic demand and disrupt nutritional balance. As the incidence of diabetes mellitus increases with aging, the associated glucosuria can contribute to weight loss. Hyperthyroidism in the elderly may have less prominent sympathomimetic features and may present as "apathetic hyperthyroidism" or T_3 toxicosis **(Chap. 394)**.

Neurologic injuries such as stroke, quadriplegia, and multiple sclerosis may lead to visceral and autonomic dysfunction that can impair caloric intake. Dysphagia from these neurologic insults is a common mechanism. Functional disability that compromises activities of daily living (ADLs) is a common cause of undernutrition in the elderly. Visual impairment from ophthalmic or central nervous system disorders such as a tremor can limit the ability of people to prepare and eat meals. UWL may be one of the earliest manifestations of Alzheimer's dementia.

Isolation and depression are significant causes of UWL that may manifest as an inability to care for oneself, including nutritional needs. A cytokine-mediated inflammatory metabolic cascade can be both a cause of and a manifestation of depression. Bereavement can be a cause of UWL and is often more pronounced in men. More intense forms of mental illness such as paranoid disorders may lead to delusions about food and cause weight loss. Alcoholism can be a significant source of weight loss and malnutrition.

Elderly persons living in poverty may have to choose whether to purchase food or use the money for other expenses, including medications. Screening questions can probe whether patients have run out of food or whether they routinely purchase less than they need. Institutionalization is an independent risk factor, as up to 30–50% of nursing home patients have inadequate food intake.

Medications can cause anorexia, nausea, vomiting, gastrointestinal distress, diarrhea, dry mouth, and changes in taste. This is particularly an issue in the elderly, many of whom take five or more medications.

ASSESSMENT

The four major manifestations of UWL are (1) anorexia (loss of appetite), (2) sarcopenia (loss of muscle mass), (3) cachexia (a syndrome that combines weight loss, loss of muscle and adipose tissue, anorexia, and weakness), and (4) dehydration. The current obesity epidemic adds complexity, as excess adipose tissue can mask the development of sarcopenia and delay awareness of the development of cachexia. If it is not possible to measure weight directly, a change in clothing size, corroboration of weight loss by a relative or friend, and a numeric estimate of weight loss provided by the patient are suggestive of true weight loss.

Initial assessment includes a comprehensive history and physical, a complete blood count, tests of liver enzyme levels, C-reactive protein, erythrocyte sedimentation rate, renal function studies, thyroid function tests, chest radiography, and an abdominal ultrasound **(Table 50-2)**. Age-, sex-, and risk factor–specific cancer screening tests, such as fecal occult blood, colonoscopy, or mammography, should be performed **(Chap. 75)**. Patients at risk should have HIV testing. All elderly patients with weight loss should undergo screening for dementia and depression by using instruments such as the Mini-Mental State Examination and the Geriatric Depression Scale, respectively **(Chap. 489)**. The Mini Nutritional Assessment (*www.mna-elderly.com*) and the Nutrition Screening Initiative (*http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1694757/*) are also available for the nutritional assessment of elderly patients. Almost all patients with a malignancy and >90% of those with other organic diseases have at least one laboratory abnormality. In patients presenting with substantial UWL, major organic and malignant diseases are unlikely when a baseline evaluation is completely normal. Careful follow-up rather than additional undirected testing is advised because the prognosis of weight loss of undetermined cause is generally favorable.

TABLE 50-2 Assessment and Testing for Involuntary Weight Loss

Indications	Laboratory
5% weight loss in 6 mo	Complete blood count
	Comprehensive electrolyte and metabolic panel, including liver and renal function tests
Body mass index <21	Thyroid function tests
25% of food left uneaten after 7 d	Erythrocyte sedimentation rate
Change in fit of clothing	C-reactive protein
Change in appetite, smell, or taste	Ferritin
Abdominal pain, nausea, vomiting, diarrhea, constipation, dysphagia	HIV testing, if indicated
Assessment	**Radiology**
Complete physical examination, including dental evaluation	Chest x-ray Abdominal ultrasound
Medication review	
Recommended cancer screening	
Mini-Mental State Examination[a]	
Mini-Nutritional Assessment[a]	
Nutrition Screening Initiative[a]	
Simplified Nutritional Assessment Questionnaire[a]	
Observation of eating[a]	
Activities of daily living[a]	
Instrumental activities of daily living[a]	

[a] May be more specific to assess weight loss in the elderly.

TREATMENT

Unintentional Weight Loss

The first priority in managing weight loss is to identify and treat the underlying causes. Treatment of underlying metabolic, social, psychiatric, dental, infectious, or other systemic disorders may be sufficient to restore weight and functional status gradually. Medications that cause nausea or anorexia should be withdrawn or changed, if possible. For those with unexplained UWL, oral nutritional supplements such as high-energy drinks sometimes reverse weight loss. Advising patients to choose appealing foods and to consume supplements between meals rather than with a meal may help minimize appetite suppression and facilitate increased overall intake. Orexigenic, anabolic, and anticytokine agents are not generally recommended. In selected patients, the antidepressant mirtazapine significantly increases body weight, body fat mass, and leptin concentration. However, side effects, including dizziness, fatigue, and somnolence, occur in about 10% of patients. Patients with wasting conditions who can comply with an appropriate exercise program gain muscle protein mass, strength, and endurance and may be more capable of performing ADLs.

FURTHER READING

Gaddey HL, Holder KK: Unintentional weight loss in older adults. Am Fam Physician 104:34, 2021.

McMinn J et al: Investigation and management of unintentional weight loss in older adults. BMJ 342:d1732, 2011.

Nicholson BD et al: Prioritising primary care patients with unexpected weight loss for cancer investigation. BMJ 370:m2651, 2020.

Perera LAM et al: Approach to patients with unintentional weight loss. Med Clin North Am 105:175, 2021.

Wong CJ: Involuntary weight loss. Med Clin North Am 98:625, 2014.

51

Gastrointestinal Bleeding

Loren Laine

Gastrointestinal bleeding (GIB) presents as either overt or occult bleeding. *Overt GIB* is manifested by *hematemesis*, vomitus of red blood or "coffee-grounds" material; *melena*, black, tarry stool; and/or *hematochezia*, passage of red or maroon blood from the rectum. In the absence of overt bleeding, *occult GIB* may present with *symptoms of blood loss or anemia* such as lightheadedness, syncope, angina, or dyspnea; with iron-deficiency anemia; or a positive fecal occult blood test on colorectal cancer screening. GIB is also categorized by the site of bleeding as upper, from the esophagus, stomach, or duodenum; lower, from the colon; small intestinal; or obscure GIB if the source is unclear.

GIB is the most common gastrointestinal condition leading to hospitalization in the United States, accounting for ~530,000 admissions annually and a case fatality of ~2%. Patients generally die from decompensation of other underlying illnesses rather than exsanguination.

SOURCES OF GASTROINTESTINAL BLEEDING

Upper Gastrointestinal Sources of Bleeding • **PEPTIC ULCERS** Peptic ulcers are the most common cause of upper GIB (UGIB), accounting for ~50% of UGIB hospitalizations. Features of an ulcer at endoscopy provide important prognostic information that guides subsequent management decisions **(Fig. 51-1)**. Approximately 20% of patients with bleeding ulcers have the highest-risk findings of active bleeding or a nonbleeding visible vessel; one-third of such patients have further bleeding that requires urgent surgery if they are treated conservatively. These patients benefit from endoscopic therapy such as bipolar electrocoagulation, heater probe, injection therapy (e.g., absolute alcohol, 1:10,000 epinephrine), clips, and/or topical hemostatic powder with reductions in bleeding, hospital stay, mortality, and costs. In contrast, patients with clean-based ulcers have rates of serious rebleeding approaching zero. If stable with no other reason for hospitalization, such patients may be discharged home after endoscopy.

Randomized controlled trials document that high-dose, proton pump inhibitor (PPI), given to reduce intragastric acid and thereby enhance clot stability, decreases further bleeding and mortality in patients with high-risk ulcers (active bleeding, nonbleeding visible vessel, adherent clot) when given after endoscopic therapy. Meta-analysis of randomized trials indicates that outcomes are comparable with high-dose PPIs given as a constant infusion or intermittently. Patients with lower-risk findings (flat pigmented spot, clean base) do not require endoscopic therapy and receive standard doses of oral PPI.

Approximately 10–50% of patients with bleeding ulcers rebleed within the next year if no preventive strategies are employed. Prevention of recurrent bleeding focuses on the three main factors in ulcer pathogenesis, *Helicobacter pylori*, nonsteroidal anti-inflammatory drugs (NSAIDs), and acid. Eradication of *H. pylori* in patients with bleeding ulcers decreases rebleeding rates to <5%. If a bleeding ulcer develops in a patient taking NSAIDs, the NSAIDs should be discontinued. If NSAIDs must be given, a cyclooxygenase-2 selective NSAID plus a PPI is recommended, based on results of a randomized trial. Patients with established cardiovascular disease who develop bleeding ulcers while taking low-dose aspirin for secondary prevention should not discontinue aspirin and, if aspirin is held, should restart aspirin once hemostasis is confirmed. A randomized trial showed that immediate reinstitution of aspirin was associated with a lower 8-week mortality compared to not restarting aspirin (1% vs 13%; hazard ratio, 0.2; 95% CI, 0.1–0.6). In contrast, aspirin should be discontinued in patients taking aspirin for primary prevention of cardiovascular events who develop UGIB. Patients with bleeding ulcers unrelated to *H. pylori* or NSAIDs should remain on PPI therapy indefinitely given a 42% incidence of rebleeding at 7 years without protective therapy. **Peptic ulcers are discussed in Chap. 335.**

MALLORY-WEISS TEARS Mallory-Weiss tears account for ~2–10% of UGIB hospitalizations. The classic history is vomiting, retching, or coughing preceding hematemesis. Bleeding from these tears, which are usually on the gastric side of the gastroesophageal junction, stops spontaneously in ~80–90% of patients and recurs in only 0–10%. Endoscopic therapy is indicated for actively bleeding Mallory-Weiss tears. **Mallory-Weiss tears are discussed in Chap. 334.**

ESOPHAGEAL VARICES The proportion of UGIB hospitalizations due to varices varies widely from ~2–40%, depending on the population. Patients with variceal hemorrhage have poorer outcomes than patients with other sources of UGIB. Esophageal varices are treated with endoscopic ligation and an IV vasoactive medication (octreotide, somatostatin, vapreotide, terlipressin) for 2–5 days. Combination of endoscopic and medical therapy is superior to either therapy alone in decreasing rebleeding. Over the long term, treatment with nonselective beta blockers plus endoscopic ligation is recommended because the combination is more effective than either alone in reduction of recurrent esophageal variceal bleeding. Transjugular intrahepatic portosystemic shunt (TIPS) is recommended in patients who have persistent or recurrent bleeding despite endoscopic and medical therapy. TIPS also is suggested for acute variceal bleeding in patients with advanced liver disease (Child-Pugh class B with score 8–9 and active bleeding at endoscopy, Child-Pugh class C with score 10–13), because randomized trials show significant decreases in rebleeding and mortality compared with standard endoscopic and medical therapy.

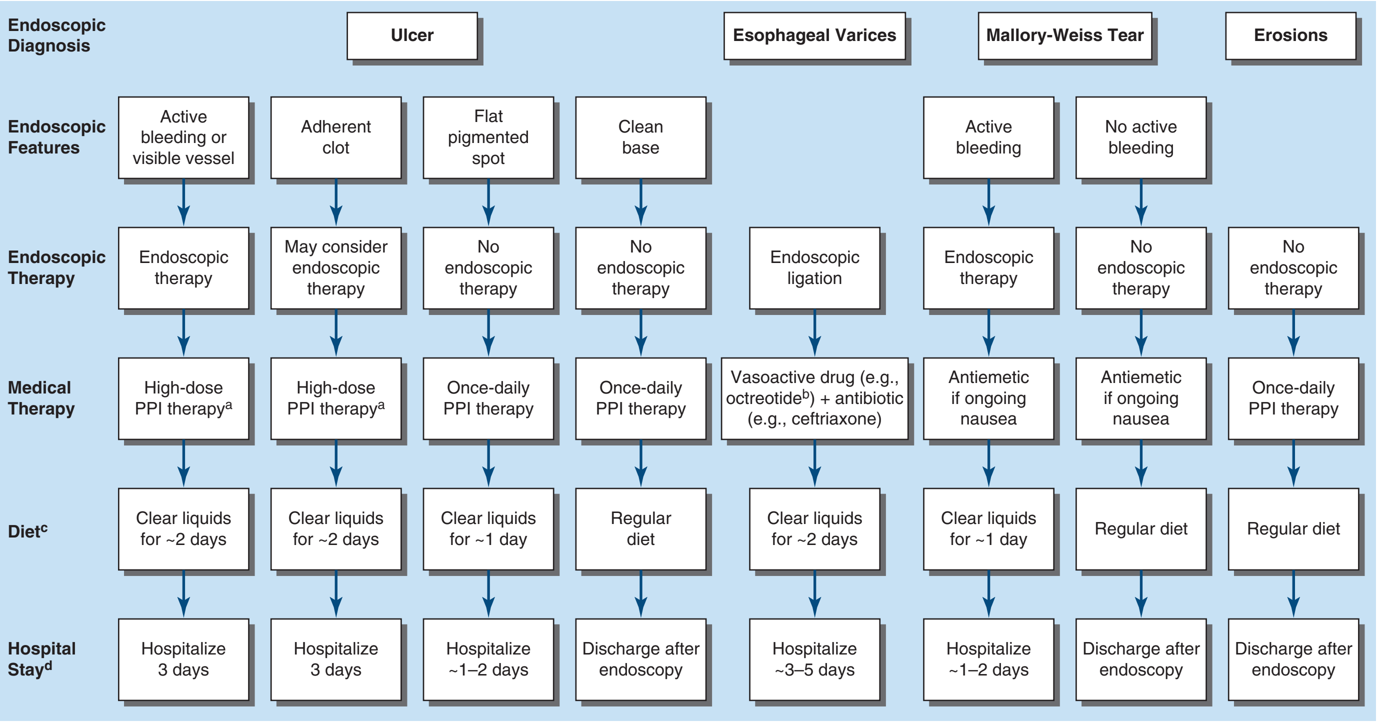

Endoscopic Diagnosis	Ulcer				Esophageal Varices	Mallory-Weiss Tear		Erosions
Endoscopic Features	Active bleeding or visible vessel	Adherent clot	Flat pigmented spot	Clean base		Active bleeding	No active bleeding	
Endoscopic Therapy	Endoscopic therapy	May consider endoscopic therapy	No endoscopic therapy	No endoscopic therapy	Endoscopic ligation	Endoscopic therapy	No endoscopic therapy	No endoscopic therapy
Medical Therapy	High-dose PPI therapy[a]	High-dose PPI therapy[a]	Once-daily PPI therapy	Once-daily PPI therapy	Vasoactive drug (e.g., octreotide[b]) + antibiotic (e.g., ceftriaxone)	Antiemetic if ongoing nausea	Antiemetic if ongoing nausea	Once-daily PPI therapy
Diet[c]	Clear liquids for ~2 days	Clear liquids for ~2 days	Clear liquids for ~1 day	Regular diet	Clear liquids for ~2 days	Clear liquids for ~1 day	Regular diet	Regular diet
Hospital Stay[d]	Hospitalize 3 days	Hospitalize 3 days	Hospitalize ~1–2 days	Discharge after endoscopy	Hospitalize ~3–5 days	Hospitalize ~1–2 days	Discharge after endoscopy	Discharge after endoscopy

FIGURE 51-1 Suggested algorithm for patients with acute upper gastrointestinal bleeding (hematemesis, melena) based on endoscopic findings. [a]Intravenous bolus (80 mg) followed by infusion (8 mg/h) or by intermittent oral or intravenous doses (e.g., 40 mg 2–4 times per day) for 3 days. [b]Intravenous 50 µg bolus followed by 50 µg/h infusion for 2–5 days. [c]Diet after endoscopy, assuming no nausea or vomiting. [d]Duration after endoscopy assuming patient stable without further bleeding or concurrent medical conditions requiring hospitalization. PPI, proton pump inhibitor.

Portal hypertension is also responsible for bleeding from gastric varices, varices in the small and large intestine, and portal hypertensive gastropathy and enterocolopathy. Bleeding gastric varices are treated with endoscopic injection of tissue adhesive (e.g., *n*-butyl cyanoacrylate), TIPS, or retrograde transvenous obliteration.

EROSIVE DISEASE Erosions are endoscopically visualized breaks that are confined to the mucosa and do not cause major bleeding because arteries and veins are not present in the mucosa. Erosions in the esophagus, stomach, or duodenum commonly cause mild UGIB, with erosive esophagitis (primarily due to gastroesophageal reflux disease), gastritis, and duodenitis accounting for up to perhaps ~30% of UGIB hospitalizations. The most important cause of gastric and duodenal erosions is NSAID use: up to ~50% of patients who chronically ingest NSAIDs may have gastric erosions. Other potential causes of gastric erosions include alcohol intake, *H. pylori* infection, and stress-related mucosal injury.

Stress-related gastric mucosal injury occurs only in extremely ill patients, such as those with serious trauma, major surgery, burns covering more than one-third of the body surface area, major intracranial disease, or severe medical illness. Severe bleeding should not develop unless ulceration occurs. The mortality rate in these patients is high because of their serious underlying illnesses.

The incidence of bleeding from stress-related gastric mucosal injury has decreased dramatically in recent years, most likely due to better care of critically ill patients. Meta-analysis of randomized trials indicates prophylaxis with PPIs reduces clinically important GIB more than no prophylaxis or H_2-receptor antagonists without impacting mortality or risk of infection (e.g., pneumonia, *Clostridioides difficile*). A guideline suggested PPI prophylaxis in critically ill patients at high risk (≥4%) of bleeding, defined as mechanical ventilation without enteral nutrition, portal hypertension, cirrhosis, platelets $<50 \times 10^9$/L, international normalized ratio >1.5, or two of the following: mechanical ventilation with enteral nutrition, acute kidney injury, sepsis, or shock.

OTHER CAUSES Less common causes of UGIB include neoplasms, vascular ectasias (including hereditary hemorrhagic telangiectasias and gastric antral vascular ectasia), Dieulafoy's lesion (in which an aberrant vessel in the mucosa bleeds from a pinpoint mucosal defect), prolapse gastropathy (prolapse of proximal stomach into esophagus with retching), aortoenteric fistulas, and hemobilia or hemosuccus pancreaticus (bleeding from the bile duct or pancreatic duct).

Small-Intestinal Sources of Bleeding Patients without a source of GIB identified on upper endoscopy and colonoscopy were previously labeled as having obscure GIB. With the advent of improved diagnostic modalities, ~75% of GIB previously labeled obscure is now estimated to originate in the small intestine beyond the extent of a standard upper endoscopic exam. Small-intestinal GIB may account for ~5% of GIB cases. The most common causes in adults include vascular ectasias, neoplasm (e.g., gastrointestinal stromal tumor, carcinoid, adenocarcinoma, lymphoma, metastases), and NSAID-induced erosions and ulcers. Meckel's diverticulum is the most common cause of significant small-intestinal GIB in children, decreasing in frequency as a cause of bleeding with age. Other less common causes of small-intestinal GIB include Crohn's disease, infection, ischemia, vasculitis, small-bowel varices, diverticula, intussusception, Dieulafoy's lesions, aortoenteric fistulas, and duplication cysts.

Small-intestinal vascular ectasias are treated initially with endoscopic therapy, based on observational studies suggesting short-term efficacy. However, rebleeding is common, with pooled rebleeding rates of ~45% over a mean follow-up of ~2 years in systematic reviews, leading guidelines to suggest medical therapy if further bleeding occurs after endoscopic therapy. The best available evidence supports use of thalidomide, with a multicenter double-blind randomized trial showing marked reductions in bleeding episodes, transfusions, and hospitalizations. Monthly intramuscular injection of octreotide long-acting release also is suggested based on observational studies and a small open-label randomized trial. Other isolated lesions, such as tumors, generally require surgical resection.

Colonic Sources of Bleeding Hemorrhoids are probably the most common cause of lower GIB (LGIB); anal fissures also cause minor bleeding and pain. If these local anal processes, which rarely require hospitalization, are excluded, the most common cause of LGIB in adults is diverticulosis. Other causes include vascular ectasias (especially in the proximal colon of patients >70 years), neoplasms (primarily adenocarcinoma), colitis (ischemic, infectious, Crohn's or ulcerative colitis, NSAID-induced colitis or ulcers), postpolypectomy bleeding, and radiation proctopathy. Rarer causes include solitary rectal ulcer syndrome, varices (most commonly rectal), lymphoid nodular hyperplasia, vasculitis, trauma, and aortocolic fistulas. In children and adolescents, the most common colonic causes of significant GIB are inflammatory bowel disease and juvenile polyps.

Diverticular bleeding is abrupt in onset, usually painless, sometimes massive, and often from the right colon; chronic or occult bleeding is not characteristic. Case series from the United States and Europe suggest colonic diverticula stop bleeding spontaneously in ≥90% of patients, with rebleeding on long-term follow-up as low as ~15% over 4–5 years. Rebleeding is substantially higher in reports from Asia. Case series suggest endoscopic therapy may decrease recurrent bleeding in the uncommon case when colonoscopy identifies the specific bleeding diverticulum. When diverticular bleeding is found at angiography, transcatheter arterial embolization by superselective technique prevents further bleeding in most patients. Segmental surgical resection is recommended for refractory diverticular bleeding.

Bleeding from colonic vascular ectasias may be overt or occult; it tends to be chronic and only occasionally hemodynamically significant. Endoscopic hemostatic therapy may be used in the treatment of vascular ectasias, as well as discrete bleeding ulcers and postpolypectomy bleeding. Transcatheter arterial embolization also may be attempted for persistent bleeding from vascular ectasias and other lesions, although rebleeding is higher in nondiverticular LGIB at ~45%. Surgical therapy is generally required for major persistent or recurrent bleeding from colonic sources that cannot be treated medically, endoscopically, or angiographically. Patients with Heyde's syndrome (bleeding vascular ectasias and aortic stenosis) appear to benefit from aortic valve replacement.

APPROACH TO THE PATIENT

Gastrointestinal Bleeding

INITIAL ASSESSMENT

Measurement of the heart rate and blood pressure is the best way to initially assess a patient with GIB. Clinically significant bleeding leads to postural changes in heart rate or blood pressure, tachycardia, and, finally, recumbent hypotension. In contrast, hemoglobin does not fall immediately with acute GIB, due to proportionate reductions in plasma and red cell volumes ("people bleed whole blood"). Thus, hemoglobin may be normal or only minimally decreased at initial presentation of a severe bleeding episode. As extravascular fluid enters the vascular space to restore volume, the hemoglobin falls, but this process may take up to 72 h. Transfusion is recommended when the hemoglobin drops below 7 g/dL, based on a large randomized trial showing this restrictive transfusion strategy decreases rebleeding and death in acute UGIB compared with a transfusion threshold of 9 g/dL. Patients with slow, chronic GIB may have very low hemoglobin values despite normal blood pressure and heart rate. With the development of iron-deficiency anemia, the mean corpuscular volume is low and red-blood-cell distribution width is increased.

DIFFERENTIATION OF UGIB FROM LGIB

Hematemesis indicates an UGIB source. Melena indicates blood has been present in the gastrointestinal (GI) tract for ≥14 h and as long as 3–5 days. The more proximal the bleeding site, the more likely melena will occur. Hematochezia usually represents a lower GI source of bleeding, although an upper GI lesion may bleed so briskly that blood transits the bowel before melena develops. When hematochezia is the presenting symptom of UGIB, it is associated with hemodynamic instability and dropping hemoglobin. Bleeding lesions of the small bowel may present as melena or hematochezia. Other clues to UGIB include hyperactive bowel sounds and elevated blood urea nitrogen (due to volume depletion and blood proteins absorbed in the small intestine).

A nonbloody nasogastric aspirate may be seen in ~15% of patients with UGIB who present with clinically serious hematochezia. A bile-stained appearance does not exclude UGIB because reports of bile in the aspirate are incorrect in ~50% of cases. Testing of aspirates that are not grossly bloody for occult blood is not useful.

EVALUATION AND MANAGEMENT OF UGIB (FIG. 51-1)

Initial Risk Assessment Baseline characteristics predictive of rebleeding and death include hemodynamic compromise (tachycardia or hypotension), increasing age, and comorbidities. Risk assessment tools may be used to identify patients with very low risk. Discharge from the emergency room with outpatient management is suggested for patients with a Glasgow-Blatchford score (possible range 0–23, **Table 51-1**) of 0–1 because only ~1% of patients who require transfusion, require hemostatic intervention, or die have a score of 0–1.

Pre-Endoscopic Medications PPI infusion may be considered at presentation; it modestly decreases need for endoscopic therapy due to a reduction in high-risk ulcer stigmata (e.g., active bleeding) but does not improve clinical outcomes such as further bleeding or death. The promotility agent erythromycin, 250 mg intravenously ~30–90 min before endoscopy, is suggested to improve visualization at endoscopy, thereby reducing the need for repeat endoscopy and hospital stay. Cirrhotic patients presenting with UGIB should be given an antibiotic (e.g., ceftriaxone) and IV vasoactive medication (e.g., octreotide) upon presentation. Antibiotics decrease bacterial infections, rebleeding, and mortality, and vasoactive medications improve control of bleeding in the 12 h after presentation.

TABLE 51-1 Glasgow-Blatchford Score

RISK FACTORS AT ADMISSION	SCORE
Blood urea nitrogen (mg/dL)	
18.2 to <22.4	2
22.4 to <28.0	3
28.0 to <70.0	4
≥70.0	6
Hemoglobin (g/dL)	
12.0 to <13.0 (men); 10.0 to <12.0 (women)	1
10.0 to <12.0 (men)	3
<10.0	6
Systolic blood pressure (mmHg)	
100–109	1
90–99	2
<90	3
Heart rate (beats per minute)	
≥100	1
Melena	1
Syncope	2
Hepatic disease	2
Cardiac failure	2

Endoscopy Upper endoscopy should be performed within 24 h in most patients hospitalized with UGIB whether they have clinical features predicting low or high risk of further bleeding and death. Even in high-risk patients, more urgent endoscopy (performed within 6–12 h of gastroenterology consultation) does not improve clinical outcomes, and some observational studies suggest increased mortality with endoscopy within 6–12 h in high-risk patients. Early endoscopy in low-risk patients (e.g., hemodynamically stable without severe comorbidities) identifies low-risk findings (e.g., clean-based ulcers, erosions, nonbleeding Mallory-Weiss tears) that allow discharge in ≥40% of patients, thereby reducing hospital stay

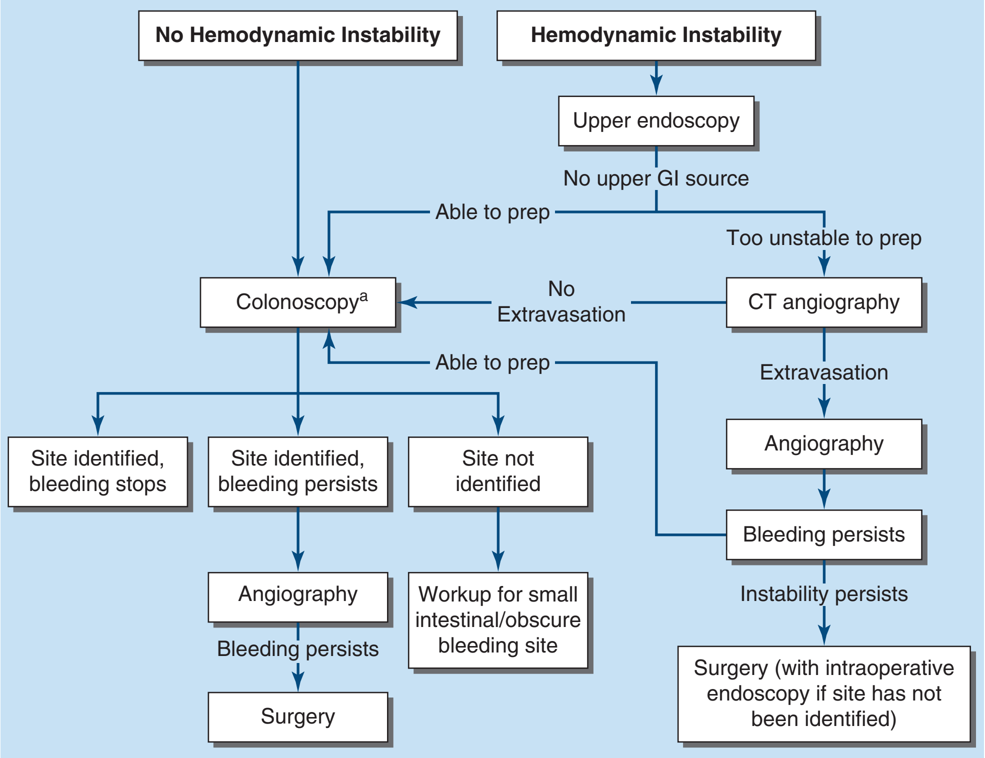
No Hemodynamic Instability
Hemodynamic Instability
Upper endoscopy
No upper GI source
Able to prep
Too unstable to prep
Colonoscopy[a]
No Extravasation
CT angiography
Able to prep
Extravasation
Angiography
Site identified, bleeding stops
Site identified, bleeding persists
Site not identified
Bleeding persists
Angiography
Workup for small intestinal/obscure bleeding site
Instability persists
Bleeding persists
Surgery
Surgery (with intraoperative endoscopy if site has not been identified)

FIGURE 51-2 Suggested algorithm for patients with acute lower gastrointestinal (GI) bleeding (hematochezia). [a]For patients <35 years old with minor bleeding (e.g., blood on toilet paper), normal blood pressure and heart rate, normal hemoglobin and iron panel, and no family history of colorectal cancer, some would suggest flexible sigmoidoscopy is adequate.

and costs. Patients with high-risk endoscopic findings (e.g., varices, ulcers with active bleeding or a visible vessel) benefit from hemostatic therapy at endoscopy.

EVALUATION AND MANAGEMENT OF LGIB (FIG. 51-2)

Patients with hematochezia and hemodynamic instability should have upper endoscopy to rule out an upper GI source before evaluation of the lower GI tract.

Colonoscopy after an oral lavage solution is the procedure of choice in most patients admitted with LGIB. If ongoing hemodynamically significant hematochezia precludes bowel preparation and colonoscopy, computed tomography (CT) angiography is suggested. If extravasation is seen, angiography is performed, allowing treatment with transcatheter arterial embolization. If no extravasation is seen, colonoscopy is subsequently performed.

EVALUATION AND MANAGEMENT OF SMALL-INTESTINAL OR OBSCURE GIB

In patients with severe bleeding suspected to be from the small intestine, CT angiography or angiography is the initial test. For others, video capsule endoscopy is generally the next diagnostic test suggested, although repeat upper and lower endoscopy may be considered because second-look procedures in some observational studies identify a source in up to ~25% of cases (a push enteroscopy, usually performed with a pediatric colonoscope to inspect the entire duodenum and proximal jejunum, may be substituted for a repeat standard upper endoscopy). Systematic reviews report a diagnostic yield with capsule endoscopy of ~55%. Limitations of capsule endoscopy include the inability to fully visualize the small intestinal mucosa, sample tissue, or apply therapy.

If capsule endoscopy is positive, management is dictated by the finding. If capsule endoscopy is negative, clinically stable patients may be observed and treated with iron if iron deficiency is present. In those with further bleeding (e.g., need for transfusions), additional testing is performed. A second capsule endoscopy may be considered because some observational studies report it identifies a source in up to 50% or more of cases. CT enterography also may follow a negative capsule endoscopy, given its higher sensitivity for small-intestinal masses (CT enterography also may be used initially instead of video capsule in patients with concern for small bowel narrowing [e.g., stricture, prior surgery or radiation, Crohn's disease]).

"Deep" enteroscopy (double-balloon, single-balloon, or spiral enteroscopy) is commonly the next test after capsule endoscopy for clinically important GIB documented or suspected to be from the small intestine because it allows the endoscopist to examine, obtain specimens from, and provide therapy to much or all of the small intestine. Other imaging techniques sometimes used in evaluation of obscure GIB include CT angiography, angiography, ^{99m}Tc-labeled red-blood-cell scintigraphy, and ^{99m}Tc-pertechnetate scintigraphy for Meckel's diverticulum (especially in young patients). If all tests are unrevealing, intraoperative endoscopy is indicated in patients with severe recurrent or persistent bleeding requiring repeated transfusions.

POSITIVE FECAL OCCULT BLOOD TEST

Fecal occult blood testing is recommended only for colorectal cancer screening, beginning at age 45 years in average-risk adults. A positive test necessitates colonoscopy. If evaluation of the colon is negative, further workup is not recommended unless iron-deficiency anemia or GI symptoms are present.

■ FURTHER READING

Chen H et al: Thalidomide for recurrent bleeding due to small-intestinal angiodysplasia. N Engl J Med 389:1649, 2023.

Kaplan DE et al: AASLD practice guidance on risk stratification and management of portal hypertension and varices in cirrhosis. Hepatology 79:1180, 2024.

Karuppasamy K et al: ACR Appropriateness criteria® radiologic management of lower gastrointestinal tract bleeding: 2021 update. J Am Coll Radiol 18:S139, 2021.

Laine L et al: ACG clinical guideline: Upper gastrointestinal and ulcer bleeding. Am J Gastroenterol 116:899, 2021.

Lau JYW et al: Timing of endoscopy for acute upper gastrointestinal bleeding. N Engl J Med 382:1299, 2020.

Pennazio M et al: Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) guideline—update 2022. Endoscopy 55:58, 2023.

Sengupta N et al: Management of patients with acute lower gastrointestinal bleeding: An updated ACG guideline. Am J Gastroenterol 118:208, 2023.

Villaneuva C et al: Transfusion strategies for acute upper gastrointestinal bleeding. N Engl J Med 368:11, 2013.

Ye Z et al: Gastrointestinal bleeding prophylaxis for critically ill patients: A clinical practice guideline. BMJ 368:16722, 2020.

52

Jaundice

Savio John, Daniel S. Pratt

Jaundice is a yellowish discoloration of body tissues resulting from the deposition of bilirubin. Tissue deposition of bilirubin occurs only in the presence of serum hyperbilirubinemia and is a sign of either liver disease or, less often, a hemolytic disorder or disorder of bilirubin metabolism. The degree of serum bilirubin elevation can be estimated by physical examination. Slight increases in serum bilirubin level are best detected by examining the sclerae for icterus. Sclerae have a particular affinity for bilirubin due to their high elastin content, and the presence of scleral icterus indicates a serum bilirubin level of at least 51 μmol/L (3 mg/dL). The ability to detect scleral icterus is made more difficult if the examining room has fluorescent lighting. If the examiner suspects scleral icterus, a second site to examine is underneath the tongue. As serum bilirubin levels rise, the skin will eventually become yellow in light-skinned patients and even green if the process is long-standing; the green color is produced by oxidation of bilirubin to biliverdin.

The differential diagnosis for yellowing of the skin is limited. In addition to jaundice, it includes carotenoderma; the use of drugs including quinacrine, sunitinib, and sorafenib; and excessive exposure to phenols. Carotenoderma, a yellow coloring of the skin, is associated with diabetes, hypothyroidism, and anorexia nervosa, but most commonly, it is caused by the ingestion of an excessive amount of vegetables and fruits such as carrots, leafy vegetables, squash, peaches, and oranges that contain carotene. In jaundice, the yellow coloration of the skin is uniformly distributed over the body, whereas in carotenoderma, the pigment is concentrated on the palms, soles, forehead, and nasolabial folds. Carotenoderma can be distinguished from jaundice by the sparing of the sclerae. Quinacrine causes a yellow discoloration of the skin in 4–37% of patients treated with it. It has also been reported with the use of the tyrosine kinase inhibitors sunitinib and sorafenib.

Another sensitive indicator of increased serum bilirubin is darkening of the urine, which is due to the renal excretion of conjugated bilirubin. Patients often describe their urine as tea- or cola-colored. Bilirubinuria indicates an elevation of the direct serum bilirubin fraction and, therefore, the presence of liver or biliary disease.

Serum bilirubin levels increase when an imbalance exists between bilirubin production and clearance. A logical evaluation of the patient who is jaundiced requires an understanding of bilirubin production and metabolism.

■ PRODUCTION AND METABOLISM OF BILIRUBIN

(See Chap. 349) Bilirubin, a tetrapyrrole pigment, is a breakdown product of heme (ferroprotoporphyrin IX). About 80–85% of the

4 mg/kg body weight of bilirubin produced each day is derived from the breakdown of hemoglobin in senescent red blood cells. The remainder comes from prematurely destroyed erythroid cells in bone marrow and from the turnover of hemoproteins such as myoglobin and cytochromes found in tissues throughout the body.

The formation of bilirubin occurs in reticuloendothelial cells, primarily in the spleen and liver. The first reaction, catalyzed by the microsomal enzyme heme oxygenase, oxidatively cleaves the α bridge of the porphyrin group and opens the heme ring. The end products of this reaction are biliverdin, carbon monoxide, and iron. The second reaction, catalyzed by the cytosolic enzyme biliverdin reductase, reduces the central methylene bridge of biliverdin and converts it to bilirubin. Bilirubin formed in the reticuloendothelial cells is virtually insoluble in water due to tight internal hydrogen bonding between the water-soluble moieties of bilirubin—that is, the bonding of the propionic acid carboxyl groups of one dipyrrolic half of the molecule with the imino and lactam groups of the opposite half. This configuration blocks solvent access to the polar residues of bilirubin and places the hydrophobic residues on the outside. To be transported in blood, bilirubin must be solubilized. Solubilization is accomplished by the reversible, noncovalent binding of bilirubin to albumin. Unconjugated bilirubin bound to albumin is transported to the liver. There, the bilirubin—but not the albumin—is taken up by hepatocytes via a process that at least partly involves carrier-mediated membrane transport. Remarkably, a full understanding of this process remains elusive as no specific bilirubin transporter has yet been identified **(Chap. 349, Fig. 349-1)**.

After entering the hepatocyte, unconjugated bilirubin is bound in the cytosol to several proteins including proteins in the glutathione-S-transferase superfamily. These proteins serve both to reduce efflux of bilirubin back into the serum and to present the bilirubin for conjugation. In the endoplasmic reticulum, bilirubin is made aqueous soluble by conjugation to glucuronic acid, a process that disrupts the hydrophobic internal hydrogen bonds and yields bilirubin monoglucuronide and diglucuronide. The conjugation of glucuronic acid to bilirubin is catalyzed by bilirubin uridine diphosphate-glucuronosyl transferase (UDPGT). The now-hydrophilic bilirubin conjugates diffuse from the endoplasmic reticulum to the canalicular membrane, where bilirubin monoglucuronide and diglucuronide are actively transported into canalicular bile by an energy-dependent mechanism involving the multidrug resistance–associated protein 2 (MRP2). A portion of bilirubin glucuronides is transported into the sinusoids and portal circulation by MRP3 and is subjected to reuptake into the hepatocyte by the sinusoidal organic anion transport protein 1B1 (OATP1B1) and OATP1B3. The conjugated bilirubin excreted into bile drains into the duodenum and passes unchanged through the proximal small bowel. Conjugated bilirubin is not reabsorbed by the intestinal mucosa due to its hydrophilicity and increased molecular size. When the conjugated bilirubin reaches the distal ileum and colon, it is hydrolyzed to unconjugated bilirubin by bacterial β-glucuronidases. The unconjugated bilirubin is reduced by normal gut bacteria to form a group of colorless tetrapyrroles called *urobilinogens* and other products, the nature and relative amounts of which depend on the bacterial flora. About 80–90% of these products are excreted in feces, either unchanged or oxidized to orange derivatives called *urobilins*. The remaining 10–20% of the urobilinogens undergo enterohepatic cycling. A small fraction (usually <3 mg/dL) escapes hepatic uptake, filters across the renal glomerulus, and is excreted in urine. Increased urinary excretion of urobilinogen can be due to increased bilirubin production, increased hepatic reabsorption of urobilinogen from the colon, or decreased hepatic clearance of urobilinogen.

■ MEASUREMENT OF SERUM BILIRUBIN

The terms *direct* and *indirect* bilirubin—that is, conjugated and unconjugated bilirubin, respectively—are based on the original van den Bergh reaction. This assay, or a variation of it, is still used in most clinical chemistry laboratories to determine the serum bilirubin level. In this assay, bilirubin is exposed to diazotized sulfanilic acid and splits into two relatively stable dipyrrylmethene azopigments that absorb maximally at 540 nm, allowing photometric analysis. The direct fraction is that which reacts with diazotized sulfanilic acid in the absence of an accelerator substance such as alcohol. The direct fraction provides an approximation of the conjugated bilirubin level in serum. The *total* serum bilirubin is the amount that reacts after the addition of alcohol to allow the release of unconjugated bilirubin from albumin binding sites. The indirect fraction is the difference between the total and the direct bilirubin levels and provides an estimate of the unconjugated bilirubin in serum. Unconjugated bilirubin also reacts with diazo reagents, albeit slowly, even when the accelerator is absent. Thus, the calculated indirect bilirubin may underestimate the true amount of unconjugated bilirubin in circulation.

With the van den Bergh method, the normal serum bilirubin concentration usually is between 17 and 26 μmol/L (1 and 1.5 mg/dL). Total serum bilirubin concentrations are between 3.4 and 15.4 μmol/L (0.2 and 0.9 mg/dL) in 95% of a normal population. Unconjugated hyperbilirubinemia is present when the direct fraction is <15% of the total serum bilirubin. The presence of even limited amounts of true conjugated bilirubin in serum suggests significant hepatobiliary pathology. As conjugated hyperbilirubinemia is always associated with bilirubinuria (except in the presence of delta bilirubin in prolonged cholestasis when jaundice is overt), detection of bilirubin in urine via dipstick test is extremely helpful to confirm the presence of conjugated hyperbilirubinemia in a patient with a mildly elevated direct fraction.

Other techniques, although less convenient to perform, have added considerably to our understanding of bilirubin metabolism. First, studies using these methods demonstrate that, in normal persons or those with Gilbert's syndrome, almost 100% of the serum bilirubin is unconjugated; <3% is monoconjugated bilirubin. Second, in jaundiced patients with hepatobiliary disease, the total serum bilirubin concentration measured by these more accurate methods is lower than the values found with diazo methods. This finding suggests that there are diazo-positive compounds distinct from bilirubin in the serum of patients with hepatobiliary disease. Third, these studies indicate that, in jaundiced patients with hepatobiliary disease, monoglucuronides of bilirubin predominate over diglucuronides. Fourth, part of the direct-reacting bilirubin fraction includes conjugated bilirubin that is covalently linked to albumin. This albumin-linked fraction of conjugated bilirubin (*delta fraction, delta bilirubin*, or *biliprotein*) represents an important fraction of total serum bilirubin in patients with cholestasis and hepatobiliary disorders. The delta bilirubin is formed in serum when hepatic excretion of bilirubin glucuronides is impaired and the glucuronides accumulate in serum. By virtue of its tight binding to albumin, the clearance rate of delta bilirubin from serum approximates the longer half-life of albumin (18–20 days) rather than the short half-life of bilirubin (about 4 h).

The prolonged half-life of albumin-bound conjugated bilirubin accounts for two previously unexplained enigmas in jaundiced patients with liver disease: (1) that some patients with conjugated hyperbilirubinemia do not exhibit bilirubinuria during the recovery phase of their disease because the delta bilirubin, although conjugated, is covalently bound to albumin and therefore not filtered by the renal glomeruli, and (2) that the elevated serum bilirubin level declines more slowly than expected in some patients who otherwise appear to be recovering satisfactorily. Late in the recovery phase of hepatobiliary disorders, all the conjugated bilirubin may be in the albumin-linked form.

■ MEASUREMENT OF URINE BILIRUBIN

Unconjugated bilirubin is always bound to albumin in the serum, is not filtered by the kidney, and is not found in the urine. Conjugated bilirubin is filtered at the glomerulus, and the majority is reabsorbed by the proximal tubules; a small fraction is excreted in the urine. Any bilirubin found in the urine is conjugated bilirubin. The presence of bilirubinuria on urine dipstick test (Ictotest) indicates an elevation of the conjugated bilirubin fraction that cannot be excreted from the liver and implies the presence of hepatobiliary disease. A false-negative result is possible in patients with prolonged cholestasis due to the predominance of delta bilirubin, which is covalently bound to albumin and therefore not filtered by the renal glomeruli.

APPROACH TO THE PATIENT

Jaundice

The goal of this chapter is not to provide an encyclopedic review of every condition that causes jaundice. Rather, the chapter is intended to offer a framework that helps a physician to evaluate the patient with jaundice in a logical way **(Fig. 52-1)**.

The initial step is to perform appropriate blood tests in order to determine whether the patient has an isolated elevation of serum bilirubin. If so, is the bilirubin elevation due to an increased unconjugated or conjugated fraction? If the hyperbilirubinemia is accompanied by other liver test abnormalities, is the disorder hepatocellular or cholestatic? If cholestatic, is the cause intra- or extrahepatic? These questions can all be answered with a thoughtful history, physical examination, and interpretation of appropriate laboratory and radiologic tests and procedures.

The bilirubin present in serum represents a balance between input from the production of bilirubin and hepatic/biliary removal of the pigment. Hyperbilirubinemia may result from (1) overproduction of bilirubin; (2) impaired uptake, conjugation, or excretion of bilirubin; or (3) regurgitation of unconjugated or conjugated bilirubin from damaged hepatocytes or bile ducts. An increase in unconjugated bilirubin in serum results from overproduction, impaired uptake, or conjugation of bilirubin. An increase in conjugated bilirubin is due to decreased excretion into the bile ductules or backward leakage of the pigment. The initial steps in evaluating the patient with jaundice are to determine (1) whether the hyperbilirubinemia is predominantly conjugated or unconjugated in nature and (2) whether other biochemical liver tests are abnormal. The thoughtful interpretation of limited data permits a rational evaluation of the patient (Fig. 52-1). The following discussion will focus solely on the evaluation of the adult patient with jaundice.

ISOLATED ELEVATION OF SERUM BILIRUBIN

Unconjugated Hyperbilirubinemia The differential diagnosis of isolated unconjugated hyperbilirubinemia is limited **(Table 52-1)**. The critical determination is whether the patient is suffering from a hemolytic process resulting in an overproduction of bilirubin (hemolytic disorders and ineffective erythropoiesis) or from impaired hepatic uptake/conjugation of bilirubin (drug effect or genetic disorders).

Hemolytic disorders that cause excessive heme production may be either inherited or acquired. Inherited disorders include

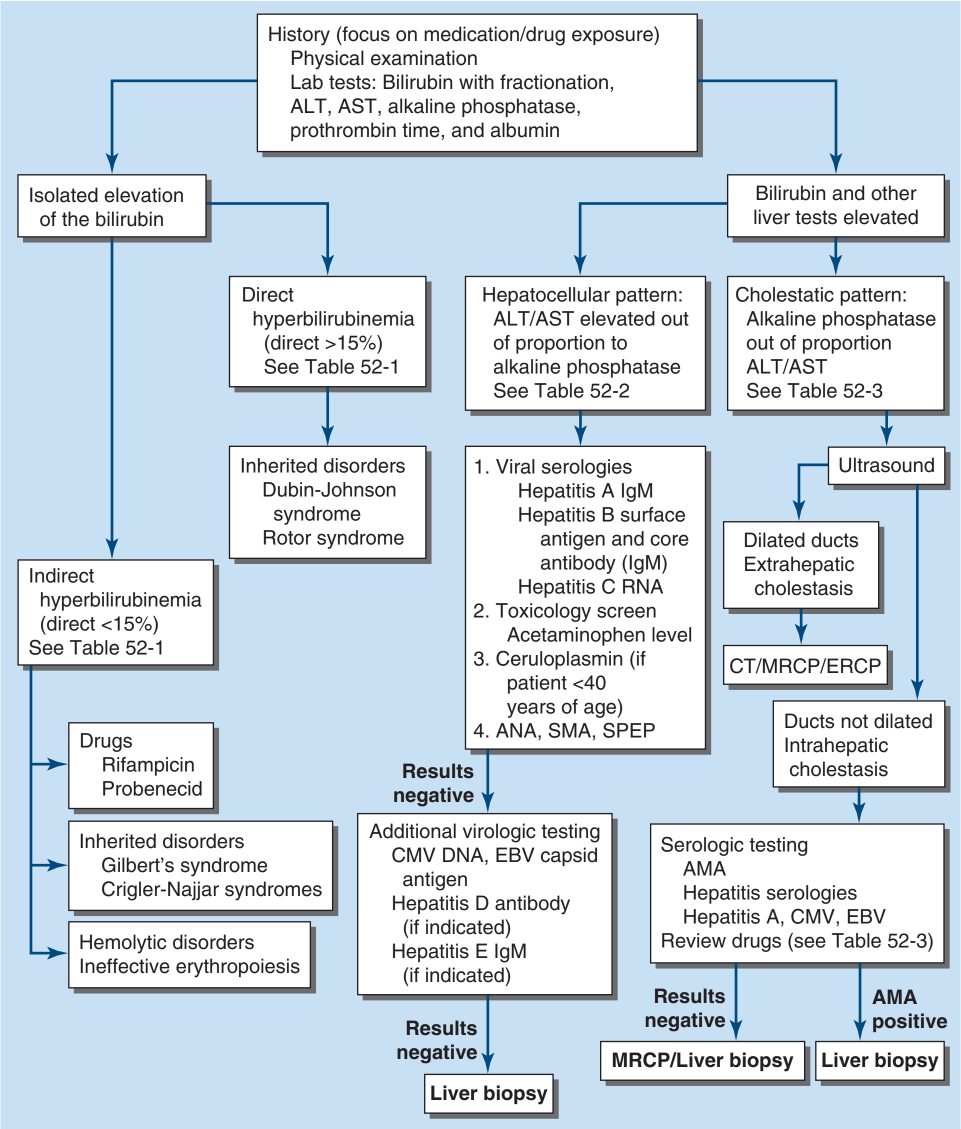
History (focus on medication/drug exposure)
Physical examination
Lab tests: Bilirubin with fractionation, ALT, AST, alkaline phosphatase, prothrombin time, and albumin
Isolated elevation of the bilirubin
Bilirubin and other liver tests elevated
Direct hyperbilirubinemia (direct >15%) See Table 52-1
Hepatocellular pattern: ALT/AST elevated out of proportion to alkaline phosphatase See Table 52-2
Cholestatic pattern: Alkaline phosphatase out of proportion ALT/AST See Table 52-3
Inherited disorders Dubin-Johnson syndrome Rotor syndrome
1. Viral serologies Hepatitis A IgM Hepatitis B surface antigen and core antibody (IgM) Hepatitis C RNA
2. Toxicology screen Acetaminophen level
3. Ceruloplasmin (if patient <40 years of age)
4. ANA, SMA, SPEP
Ultrasound
Dilated ducts Extrahepatic cholestasis
CT/MRCP/ERCP
Ducts not dilated Intrahepatic cholestasis
Indirect hyperbilirubinemia (direct <15%) See Table 52-1
Drugs Rifampicin Probenecid
Inherited disorders Gilbert's syndrome Crigler-Najjar syndromes
Hemolytic disorders Ineffective erythropoiesis
Results negative
Additional virologic testing CMV DNA, EBV capsid antigen Hepatitis D antibody (if indicated) Hepatitis E IgM (if indicated)
Results negative
Liver biopsy
Serologic testing AMA Hepatitis serologies Hepatitis A, CMV, EBV Review drugs (see Table 52-3)
Results negative
AMA positive
MRCP/Liver biopsy
Liver biopsy

FIGURE 52-1 Evaluation of the patient with jaundice. ALT, alanine aminotransferase; AMA, antimitochondrial antibody; ANA, antinuclear antibody; AST, aspartate aminotransferase; CMV, cytomegalovirus; EBV, Epstein-Barr virus; ERCP, endoscopic retrograde cholangiopancreatography; LKM, liver-kidney microsomal antibody; MRCP, magnetic resonance cholangiopancreatography; SMA, smooth-muscle antibody; SPEP, serum protein electrophoresis.

TABLE 52-1 Causes of Isolated Hyperbilirubinemia

I. Indirect hyperbilirubinemia
  A. Hemolytic disorders
  B. Ineffective erythropoiesis
  C. Increased bilirubin production
    1. Massive blood transfusion
    2. Resorption of hematoma
  D. Drugs
    1. Rifampin
    2. Probenecid
    3. Atazanavir
    4. Antibiotics—cephalosporins and penicillins
  E. Inherited conditions
    1. Crigler-Najjar types I and II
    2. Gilbert's syndrome
II. Direct hyperbilirubinemia (inherited conditions)
  A. Dubin-Johnson syndrome
  B. Rotor syndrome

spherocytosis, sickle cell anemia, thalassemia, and deficiency of red cell enzymes such as pyruvate kinase and glucose-6-phosphate dehydrogenase. In these conditions, the serum bilirubin level rarely exceeds 86 μmol/L (5 mg/dL). Higher levels may occur when there is coexistent renal or hepatocellular dysfunction or in acute hemolysis, such as a sickle cell crisis. In evaluating jaundice in patients with chronic hemolysis, it is important to remember the high incidence of pigmented (calcium bilirubinate) gallstones found in these patients, which increases the likelihood of choledocholithiasis as an alternative explanation for hyperbilirubinemia.

Acquired hemolytic disorders include microangiopathic hemolytic anemia (e.g., hemolytic-uremic syndrome), paroxysmal nocturnal hemoglobinuria, spur cell anemia, immune hemolysis, and parasitic infections (e.g., malaria and babesiosis). Ineffective erythropoiesis occurs in cobalamin, folate, and iron deficiencies. Resorption of hematomas and massive blood transfusions both can result in increased hemoglobin release and overproduction of bilirubin.

In the absence of hemolysis, the physician should consider a problem with the hepatic uptake or conjugation of bilirubin. Certain drugs, including rifampin and probenecid, may cause unconjugated hyperbilirubinemia by diminishing hepatic uptake of bilirubin. Impaired bilirubin conjugation occurs in three genetic conditions: Crigler-Najjar syndrome types I and II and Gilbert's syndrome. *Crigler-Najjar type I* is an exceptionally rare condition found in neonates and characterized by severe jaundice (bilirubin >342 μmol/L [>20 mg/dL]) and neurologic impairment due to kernicterus, frequently leading to death in infancy or childhood. These patients have a complete absence of bilirubin UDPGT activity; are totally unable to conjugate bilirubin; and hence cannot excrete it.

Crigler-Najjar type II is somewhat more common than type I. Patients live into adulthood with serum bilirubin levels of 103–428 μmol/L (6–25 mg/dL). In these patients, mutations in the bilirubin *UDPGT* gene cause the reduction—typically ≤10%—of the enzyme's activity. Bilirubin UDPGT activity can be induced by the administration of phenobarbital, which can reduce serum bilirubin levels in these patients. Despite marked jaundice, these patients usually survive into adulthood, although they may be susceptible to kernicterus under the stress of concurrent illness or surgery.

Gilbert's syndrome is also marked by the impaired conjugation of bilirubin due to reduced bilirubin UDPGT activity (typically 10–35% of normal). Patients with Gilbert's syndrome have mild unconjugated hyperbilirubinemia, with serum levels almost always <103 μmol/L (6 mg/dL). The serum levels may fluctuate, and jaundice is often identified only during periods of stress, concurrent illness, alcohol use, or fasting. Unlike both Crigler-Najjar syndromes, Gilbert's syndrome is very common. The reported incidence is 3–7% of the population, with males predominating over females by a ratio of 1.5–7:1. The mildly elevated indirect serum hyperbilirubinemia seen in Gilbert's syndrome is generally of no clinical consequence and may actually have protective effects.

Conjugated Hyperbilirubinemia Elevated conjugated hyperbilirubinemia is found in two rare inherited conditions: *Dubin-Johnson syndrome* and *Rotor syndrome* (Table 52-1). Patients with either condition present with asymptomatic jaundice. The defect in Dubin-Johnson syndrome is the presence of mutations in the gene for MRP2. These patients have altered excretion of bilirubin into the bile ducts. Rotor syndrome is caused by a deficiency of the major hepatic drug reuptake transporters OATP1B1 and OATP1B3. Differentiating between these syndromes is possible but is clinically unnecessary due to their benign nature.

ELEVATION OF SERUM BILIRUBIN WITH OTHER LIVER TEST ABNORMALITIES

The remainder of this chapter will focus on the evaluation of patients with conjugated hyperbilirubinemia in the setting of other liver test abnormalities. This group of patients can be divided into those with a primary hepatocellular process and those with intra- or extrahepatic cholestasis. This distinction, which is based on the history and physical examination as well as the pattern of liver test abnormalities, guides the clinician's evaluation (Fig. 52-1).

History A complete medical history is perhaps the single most important part of the evaluation of the patient with unexplained jaundice. Important considerations include the use of or exposure to any chemical or medication, whether physician-prescribed, over-the-counter, complementary, or alternative medicines (e.g., herbal and vitamin preparations) or other drugs such as anabolic steroids. The patient should be carefully questioned about possible parenteral exposures, including transfusions, intravenous and intranasal drug use, tattooing, and sexual activity. Other important points include recent travel history; exposure to people with jaundice; exposure to possibly contaminated foods; occupational exposure to hepatotoxins; alcohol consumption; the duration of jaundice; and the presence of any accompanying signs and symptoms, such as arthralgias, myalgias, rash, anorexia, weight loss, abdominal pain, fever, pruritus, and changes in the urine and stool. While none of the latter manifestations is specific for any one condition, any of them can suggest a diagnosis. A history of arthralgias and myalgias predating jaundice suggests hepatitis, either viral or drug related. Jaundice associated with the sudden onset of severe right-upper-quadrant pain and shaking chills suggests choledocholithiasis and ascending cholangitis.

Physical Examination The general assessment should include evaluation of the patient's nutritional status. Temporal and proximal muscle wasting suggests long-standing disease such as pancreatic cancer or cirrhosis. Stigmata of chronic liver disease, including spider nevi, palmar erythema, gynecomastia, caput medusae, Dupuytren's contractures, parotid gland enlargement, and testicular atrophy, are commonly seen in advanced alcohol-related cirrhosis and occasionally in other types of cirrhosis. An enlarged left supraclavicular node (Virchow's node) or a periumbilical nodule (Sister Mary Joseph's nodule) suggests an abdominal malignancy. Jugular venous distention, a sign of right-sided heart failure, and/or a pulsatile liver suggest hepatic congestion. Right pleural effusion even in the absence of clinically apparent ascites may be seen in advanced cirrhosis.

The abdominal examination should focus on the size and consistency of the liver, on whether the spleen is palpable and hence enlarged, and on whether ascites is present. Patients with cirrhosis may have an enlarged left lobe of the liver, which is felt below the xiphoid, and an enlarged spleen. A grossly enlarged nodular liver or an obvious abdominal mass suggests malignancy. An enlarged tender liver could signify viral or alcoholic hepatitis; an infiltrative process such as amyloidosis; or, less often, an acutely congested liver secondary to right-sided heart failure. Severe right-upper-quadrant tenderness with respiratory arrest on inspiration (Murphy's sign) suggests cholecystitis. Ascites in the presence of jaundice suggests either cirrhosis or malignancy with peritoneal spread.

Laboratory Tests A battery of tests are helpful in the initial evaluation of a patient with unexplained jaundice. These include total and direct serum bilirubin measurement; determination of serum aminotransferase, alkaline phosphatase, and albumin concentrations; and prothrombin time tests. Enzyme tests (alanine aminotransferase [ALT], aspartate aminotransferase [AST], and alkaline phosphatase [ALP]) are helpful in differentiating between a hepatocellular process and a cholestatic process (Table 348-1; Fig. 52-1)—a critical step in determining what additional workup is indicated. Patients with a hepatocellular process generally have a rise in the aminotransferases that is disproportionate to that in ALP, whereas patients with a cholestatic process have a rise in ALP that is disproportionate to that of the aminotransferases. The serum bilirubin can be prominently elevated in both hepatocellular and cholestatic conditions and therefore is not necessarily helpful in differentiating between the two.

In addition to enzyme tests, all jaundiced patients should have additional blood tests—specifically, an albumin level and a prothrombin time—to assess liver function. A low albumin level suggests a chronic process such as cirrhosis or cancer. A normal albumin level is suggestive of a more acute process such as viral hepatitis or choledocholithiasis. An elevated prothrombin time indicates either vitamin K deficiency due to prolonged jaundice and malabsorption of vitamin K or significant hepatocellular dysfunction. The failure of the prothrombin time to correct with parenteral administration of vitamin K indicates severe hepatocellular injury.

The results of the bilirubin, enzyme, albumin, and prothrombin time tests will usually indicate whether a jaundiced patient has a hepatocellular or a cholestatic disease and offer some indication of the duration and severity of the disease. The causes and evaluations of hepatocellular and cholestatic diseases are quite different.

Hepatocellular Conditions Hepatocellular diseases that can cause jaundice include viral hepatitis, drug or environmental toxicity, alcohol, and end-stage cirrhosis from any cause (Table 52-2). Wilson's disease occurs primarily in young adults and usually between the ages of 3 and 55. Autoimmune hepatitis is typically seen in young to middle-aged women but may affect men and women of any age. Alcoholic hepatitis can be differentiated from viral and toxin-related hepatitis by the pattern of the aminotransferases: patients with alcoholic hepatitis typically have an AST-to-ALT ratio of at least 2:1, and the AST level rarely exceeds 300 U/L. Patients with acute viral hepatitis and toxin-related injury severe enough to produce jaundice typically have aminotransferase levels >500 U/L, with the ALT greater than or equal to the AST. While ALT and AST values <8 times normal may be seen in either hepatocellular or cholestatic liver disease, values 25 times normal or higher are seen primarily in acute hepatocellular diseases. Patients with jaundice from cirrhosis can have normal or only slightly elevated aminotransferase levels.

When the clinician determines that a patient has a hepatocellular disease, appropriate testing for acute viral hepatitis includes a hepatitis A IgM antibody assay, a hepatitis B surface antigen and core IgM antibody assay, a hepatitis C viral RNA test, and, depending on the circumstances, a hepatitis E IgM antibody assay. The hepatitis C antibody can take up to 6 weeks to become detectable, making it an unreliable test if acute hepatitis C is suspected. Studies for hepatitis D, Epstein-Barr virus (EBV), and cytomegalovirus (CMV) may also be indicated. Ceruloplasmin is the initial screening test for Wilson's disease. Testing for autoimmune hepatitis usually includes antinuclear antibody and anti–smooth muscle antibody assays and measurement of specific immunoglobulins.

Drug-induced hepatocellular injury can be classified as either predictable or unpredictable. Predictable drug reactions are dose-dependent and affect all patients who ingest a toxic dose of the drug in question. The classic example is acetaminophen hepatotoxicity. Unpredictable or idiosyncratic drug reactions are not dose-dependent and occur in a minority of patients. A great number of drugs can cause idiosyncratic hepatic injury. Environmental toxins are also an important cause of hepatocellular injury. Examples include industrial chemicals such as vinyl chloride, herbal preparations containing pyrrolizidine alkaloids (Jamaica bush tea) or kava, and the mushrooms *Amanita phalloides* and *A. verna*, which contain highly hepatotoxic amatoxins.

TABLE 52-2 Hepatocellular Conditions That May Produce Jaundice

- Viral hepatitis
  - Hepatitis A, B, C, D, and E
  - Epstein-Barr virus
  - Cytomegalovirus
  - Herpes simplex virus
- Alcoholic hepatitis
- Chronic liver disease and cirrhosis
- Drug toxicity
  - Predictable, dose-dependent (e.g., acetaminophen)
  - Unpredictable, idiosyncratic (e.g., isoniazid)
- Environmental toxins
  - Vinyl chloride
  - Jamaica bush tea—pyrrolizidine alkaloids
  - Kava kava
  - Wild mushrooms—*Amanita phalloides, A. verna*
- Wilson's disease
- Autoimmune hepatitis

Cholestatic Conditions When the pattern of the liver tests suggests a cholestatic disorder, the first step is to determine whether it is intra- or extrahepatic cholestasis (Fig. 52-1). Distinguishing intrahepatic from extrahepatic cholestasis may be difficult. History, physical examination, and laboratory tests often are not helpful. The next appropriate test is an ultrasound. The ultrasound is inexpensive, does not expose the patient to ionizing radiation, and can detect dilation of the intra- and extrahepatic biliary tree with a high degree of sensitivity and specificity. The absence of biliary dilation suggests intrahepatic cholestasis, while its presence indicates extrahepatic cholestasis. False-negative results occur in patients with partial obstruction of the common bile duct or in patients with cirrhosis or primary sclerosing cholangitis (PSC), in which scarring prevents the intrahepatic ducts from dilating.

Although ultrasonography may indicate extrahepatic cholestasis, it rarely identifies the site or cause of obstruction. The distal common bile duct is a particularly difficult area to visualize by ultrasound because of overlying bowel gas. Appropriate next tests include computed tomography (CT), magnetic resonance cholangiopancreatography (MRCP), endoscopic retrograde cholangiopancreatography (ERCP), percutaneous transhepatic cholangiography (PTC), and endoscopic ultrasound (EUS). CT and MRCP are better than ultrasonography for assessing the head of the pancreas and for identifying choledocholithiasis in the distal common bile duct, particularly when the ducts are not dilated. ERCP is the "gold standard" for identifying choledocholithiasis. Beyond its diagnostic capabilities, ERCP allows therapeutic interventions, including the removal of common bile duct stones and the placement of stents. PTC can provide the same information as ERCP and also allows for intervention in patients in whom ERCP is unsuccessful due to proximal biliary obstruction or altered gastrointestinal anatomy. MRCP has replaced ERCP as the initial diagnostic test in most cases. EUS displays sensitivity and specificity comparable to that of MRCP in the detection of bile duct obstruction and allows biopsy of suspected malignant lesions.

In patients with apparent *intrahepatic cholestasis*, the diagnosis is often made by serologic testing in combination with a liver biopsy. The list of possible causes of intrahepatic cholestasis is long and varied **(Table 52-3)**. A number of conditions that typically cause a hepatocellular pattern of injury can also present as a cholestatic variant. Both hepatitis B and C viruses can cause cholestatic hepatitis (fibrosing cholestatic hepatitis). This disease variant has been reported in patients who have undergone solid organ transplantation. The availability of direct-acting antiviral drugs has reduced the incidence of this condition. Hepatitis A and E, alcoholic hepatitis, and EBV or CMV infections may also present as cholestatic liver disease.

TABLE 52-3 Cholestatic Conditions That May Produce Jaundice

I. Intrahepatic
    A. Viral hepatitis
        1. Fibrosing cholestatic hepatitis—hepatitis B and C
        2. Hepatitis A, Epstein-Barr virus infection, cytomegalovirus infection
    B. Alcoholic hepatitis
    C. Drug toxicity
        1. Pure cholestasis—anabolic and contraceptive steroids
        2. Cholestatic hepatitis—chlorpromazine, erythromycin estolate
        3. Chronic cholestasis—chlorpromazine and prochlorperazine
    D. Primary biliary cholangitis
    E. Sclerosing cholangitis, primary and secondary
    F. Vanishing bile duct syndrome
        1. Chronic rejection of liver transplants
        2. Sarcoidosis
        3. Drugs
    G. Congestive hepatopathy and ischemic hepatitis
    H. Inherited conditions
        1. Progressive familial intrahepatic cholestasis
        2. Benign recurrent intrahepatic cholestasis
    I. Cholestasis of pregnancy
    J. Total parenteral nutrition
    K. Nonhepatobiliary sepsis
    L. Benign postoperative cholestasis
    M. Paraneoplastic syndrome
    N. Veno-occlusive disease
    O. Graft-versus-host disease
    P. Infiltrative disease
        1. Tuberculosis
        2. Lymphoma
        3. Amyloidosis
    Q. Infections
        1. Malaria
        2. Leptospirosis
II. Extrahepatic
    A. Malignant
        1. Cholangiocarcinoma
        2. Pancreatic cancer
        3. Gallbladder cancer
        4. Ampullary cancer
        5. Malignant involvement of the porta hepatis lymph nodes
    B. Benign
        1. Choledocholithiasis
        2. Postoperative biliary strictures
        3. Primary sclerosing cholangitis
        4. Chronic pancreatitis
        5. AIDS cholangiopathy
        6. Mirizzi's syndrome
        7. Parasitic disease (ascariasis)

Drugs may cause intrahepatic cholestasis that is usually reversible after discontinuation of the offending agent, although it may take many months for cholestasis to resolve. Drugs most commonly associated with cholestasis are the anabolic and contraceptive steroids. Cholestatic hepatitis has been reported with chlorpromazine, imipramine, tolbutamide, sulindac, cimetidine, and erythromycin estolate. It also occurs in patients taking trimethoprim; sulfamethoxazole; and penicillin-based antibiotics such as ampicillin, dicloxacillin, and clavulanic acid. Rarely, cholestasis may be chronic and associated with progressive fibrosis despite early discontinuation of the offending drug. Chronic cholestasis has been associated with chlorpromazine and prochlorperazine.

Primary biliary cholangitis is an autoimmune disease predominantly affecting women and characterized by progressive destruction of interlobular bile ducts. The diagnosis is made by the detection of antimitochondrial antibody, which is found in 95% of patients. *PSC* is characterized by the destruction and fibrosis of larger bile ducts. The diagnosis of PSC is made with cholangiography (MRCP), which demonstrates the pathognomonic segmental strictures. Approximately 75% of patients with PSC also have inflammatory bowel disease.

The *vanishing bile duct syndrome* and *adult bile ductopenia* are rare conditions in which a decreased number of bile ducts are seen in liver biopsy specimens. This histologic picture is also seen in patients who develop chronic rejection after liver transplantation and in those who develop graft-versus-host disease after bone marrow transplantation. Vanishing bile duct syndrome also occurs in rare cases of sarcoidosis, in patients taking certain drugs (including chlorpromazine), and idiopathically.

There are also familial forms of intrahepatic cholestasis. The familial intrahepatic cholestatic syndromes include *progressive familial intrahepatic cholestasis* (PFIC) *types 1–3* and *benign recurrent intrahepatic cholestasis* (BRIC) *types 1 and 2*. BRIC is characterized by episodic attacks of pruritus, cholestasis, and jaundice beginning at any age, which can be debilitating but do not lead to chronic liver disease. Serum bile acids are elevated during episodes, but serum γ-glutamyltransferase (γ-GT) activity is normal. PFIC disorders begin at childhood and are progressive in nature. All three types of PFIC are associated with progressive cholestasis, elevated levels of serum bile acids, and similar phenotypes but different genetic mutations. Only type 3 PFIC is associated with high levels of γ-GT. *Cholestasis of pregnancy* occurs in the second and third trimesters and resolves after delivery. Its cause remains unknown, but the condition is probably inherited, and cholestasis can be triggered by estrogen administration.

Other causes of intrahepatic cholestasis include total parenteral nutrition (TPN); nonhepatobiliary sepsis; benign postoperative cholestasis; and a paraneoplastic syndrome associated with a number of different malignancies, including Hodgkin's disease, medullary thyroid cancer, renal cell cancer, renal sarcoma, T-cell lymphoma, prostate cancer, and several gastrointestinal malignancies. In patients developing cholestasis in the intensive care unit, the major considerations should be sepsis, ischemic hepatitis ("shock liver"), and TPN-related jaundice. Jaundice occurring after bone marrow transplantation is most likely due to veno-occlusive disease (also called sinusoidal obstruction syndrome) or graft-versus-host disease. In addition to hemolysis, sickle cell disease may cause intrahepatic and extrahepatic cholestasis. Jaundice is a late finding in heart failure caused by hepatic congestion and hepatocellular hypoxia. Ischemic hepatitis is a distinct entity of acute hypoperfusion characterized by an acute and dramatic elevation in the serum aminotransferases followed by a gradual peak in serum bilirubin.

Jaundice with associated liver dysfunction can be seen in severe cases of *Plasmodium falciparum* malaria. The jaundice in these cases is due to a combination of indirect hyperbilirubinemia from hemolysis and both cholestatic and hepatocellular jaundice. Weil's disease, a severe presentation of leptospirosis, is marked by jaundice with renal failure, fever, headache, and muscle pain.

Causes of *extrahepatic cholestasis* can be split into malignant and benign (Table 52-3). Malignant causes include pancreatic, gallbladder, and ampullary cancers as well as cholangiocarcinoma. This last malignancy is most commonly associated with PSC and is exceptionally difficult to diagnose because its appearance is often identical to that of PSC. Pancreatic and gallbladder tumors as well as cholangiocarcinoma are rarely resectable at the time of diagnosis and have poor prognoses. Ampullary carcinoma has the highest surgical cure rate of all the tumors that present as painless jaundice. Hilar lymphadenopathy due to metastases from other cancers may cause obstruction of the extrahepatic biliary tree.

Choledocholithiasis is the most common cause of extrahepatic cholestasis. The clinical presentation can range from mild right-upper-quadrant discomfort with only minimal elevations of enzyme test values to ascending cholangitis with jaundice, sepsis, and circulatory collapse. PSC may occur with clinically important strictures limited to the extrahepatic biliary tree. IgG4-associated cholangitis is marked by stricturing of the biliary tree. It is critical that the clinician differentiate this condition from PSC as it is responsive to glucocorticoid therapy. In rare instances, chronic pancreatitis causes strictures of the distal common bile duct, where it passes through the head of the pancreas. AIDS cholangiopathy is a condition that is usually due to infection of the bile duct epithelium with CMV or cryptosporidia and has a cholangiographic appearance similar to that of PSC. The affected patients usually present with greatly elevated serum alkaline phosphatase levels (mean, 800 IU/L), but the bilirubin level is often near normal. These patients do not typically present with jaundice. Its incidence has dropped dramatically since the introduction of potent antiretrovirals in the 1990s.

GLOBAL CONSIDERATIONS

While extrahepatic biliary obstruction and drugs are common causes of new-onset jaundice in developed countries, infections remain the leading cause in developing countries. Liver involvement and jaundice are observed with numerous infections, particularly malaria, babesiosis, severe leptospirosis, infections due to *Mycobacterium tuberculosis* and the *Mycobacterium avium* complex, typhoid fever, infection with hepatitis viruses A–E, EBV, CMV, viral hemorrhagic fevers including Ebola virus, late phases of yellow fever, dengue fever, schistosomiasis, fascioliasis, clonorchiasis, opisthorchiasis, ascariasis, echinococcosis, hepatosplenic candidiasis, disseminated histoplasmosis, cryptococcosis, coccidioidomycosis, ehrlichiosis, chronic Q fever, yersiniosis, brucellosis, syphilis, and leprosy. Bacterial infections that do not necessarily involve the liver and bile ducts may also lead to jaundice, as in cholestasis of sepsis. The presence of fever or abdominal pain suggests concurrent infection, sepsis, or complications from gallstones. The development of encephalopathy and coagulopathy in a jaundiced patient with no preexisting liver disease signifies acute liver failure, which warrants urgent liver transplant evaluation.

FURTHER READING

Erlinger S et al: Inherited disorders of bilirubin transport and conjugation: New insights into molecular mechanisms and consequences. Gastroenterology 146:1625, 2014.

Wolkoff AW et al: Bilirubin metabolism and jaundice, in *Schiff's Diseases of the Liver*, 11th ed, Schiff ER et al (eds). Oxford, UK, John Wiley & Sons, Ltd, 2012, pp 120–150.

53

Abdominal Swelling and Ascites

Lawrence S. Friedman

ABDOMINAL SWELLING

Abdominal swelling is a manifestation of numerous diseases. Patients may complain of *bloating* or abdominal fullness and may note increasing abdominal girth on the basis of increased clothing or belt size. Abdominal discomfort is often reported, but pain is less frequent. When abdominal pain does accompany swelling, it is frequently the result of an intraabdominal infection, peritonitis, or pancreatitis. Patients with abdominal distention from *ascites* (fluid in the abdomen) may report the new onset of an inguinal or umbilical hernia. Dyspnea may result from pressure against the diaphragm and the inability to expand the lungs fully.

CAUSES

The causes of abdominal swelling can be remembered conveniently as the *six Fs*: flatus, fat, fluid, fetus, feces, or a "fatal growth" (often a neoplasm).

Flatus Abdominal swelling may be the result of increased intestinal gas. The normal small intestine contains ~200 mL of gas made up of nitrogen, oxygen, carbon dioxide, hydrogen, and methane. Nitrogen and oxygen are consumed (swallowed), whereas carbon dioxide, hydrogen, and methane are produced intraluminally by bacterial fermentation. Increased intestinal gas can occur in a number of conditions. *Aerophagia*, the swallowing of air, can result in increased amounts of oxygen and nitrogen in the small intestine and lead to abdominal swelling. Aerophagia typically results from gulping food; chewing gum; smoking; or as a response to anxiety, which can lead to repetitive belching. Celiac disease, or gluten-sensitive enteropathy, and gastroparesis may be associated with bloating and distention. In some cases, increased intestinal gas is the consequence of bacterial metabolism of excess fermentable substances such as lactose, fructose, and other fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAPs), which can lead to production of hydrogen, carbon dioxide, or methane, as occurs also in small intestinal bacterial overgrowth or intestinal methanogen overgrowth. In some persons, particularly those with irritable bowel syndrome and bloating, the subjective sense of abdominal pressure is attributable to impaired intestinal transit of gas (rather than increased gas volume) or to visceral hypersensitivity. Abdominal *distention*—an objective increase in girth—may result from a lack of coordination between diaphragmatic contraction and anterior abdominal wall relaxation ("abdominophrenic dyssynergia"), a response in some cases to intraluminal bowel stimuli; dietary alterations, manipulation of the intestinal microbiota, or biofeedback may be effective therapy. Occasionally, increased lumbar lordosis accounts for apparent abdominal distention.

Fat Weight gain with an increase in abdominal fat can result in an increase in abdominal girth and can be perceived as abdominal swelling. Abdominal fat may be caused by an imbalance between caloric intake and energy expenditure associated with a poor diet and sedentary lifestyle; it also can be a manifestation of certain diseases, such as Cushing's syndrome. Excess abdominal fat has been associated with an increased risk of insulin resistance and cardiovascular disease.

Fluid The accumulation of fluid within the abdominal cavity (ascites) often results in abdominal distention and is discussed in detail below. Grade 1 ascites is detectable only by ultrasonography; grade 2 ascites is detectable by physical examination; and grade 3 ascites results in marked abdominal distention.

Fetus Pregnancy results in increased abdominal girth. Typically, an increase in abdominal size is first noted at 12–14 weeks of gestation, when the uterus moves from the pelvis into the abdomen. Abdominal distention may be seen before this point as a result of fluid retention and relaxation of the abdominal muscles.

Feces In the setting of severe constipation or intestinal obstruction, increased stool in the colon leads to increased abdominal girth. These conditions are often accompanied by abdominal discomfort or pain, nausea, and vomiting and can be diagnosed by imaging studies.

Fatal Growth An abdominal mass can result in abdominal swelling. Neoplasms (including ovarian cancer in women), abscesses, or cysts can grow to sizes that lead to increased abdominal girth. Enlargement of the intraabdominal organs, specifically the liver (hepatomegaly) or spleen (splenomegaly), or an abdominal aortic aneurysm can result in abdominal distention. Bladder distention also may result in abdominal swelling.

APPROACH TO THE PATIENT

Abdominal Swelling

HISTORY

Determining the etiology of abdominal swelling begins with history-taking and a physical examination. Patients should be questioned regarding the onset and timing of bloating and distention, the relationship to food or bowel movements, prior surgery, dietary habits, and medications, as well as symptoms suggestive of malignancy, including weight loss, night sweats, and anorexia. Inability to pass stool or flatus together with nausea or vomiting suggests bowel obstruction, severe constipation, or an ileus (lack of peristalsis). Increased eructation and flatus may point toward aerophagia or increased intestinal production of gas. Patients should be questioned about risk factors for or symptoms of chronic liver disease, including excessive alcohol use and jaundice, which suggest ascites. Patients should also be asked about symptoms of other medical conditions, including heart failure and tuberculosis, which may cause ascites.

PHYSICAL EXAMINATION

Physical examination should include an assessment for signs of systemic disease. The presence of lymphadenopathy, especially supraclavicular lymphadenopathy (*Virchow's node*), suggests metastatic abdominal malignancy. Care should be taken during the cardiac examination to evaluate for elevation of jugular venous pressure (JVP); *Kussmaul's sign* (elevation of the JVP during inspiration); a pericardial knock, which may be seen in constrictive pericarditis and heart failure; or a murmur of tricuspid regurgitation. Spider angiomas, palmar erythema, dilated superficial veins around the umbilicus (*caput medusae*), and gynecomastia suggest liver disease.

The abdominal examination should begin with inspection for the presence of uneven distention or an obvious mass. Auscultation should follow. The absence of bowel sounds or the presence of high-pitched localized bowel sounds points toward an intestinal obstruction or ileus. An umbilical venous hum may suggest the presence of portal hypertension, and a harsh bruit over the liver is heard rarely in patients with hepatocellular carcinoma or alcohol-associated hepatitis. Abdominal swelling caused by intestinal gas can be differentiated from swelling caused by fluid or a solid mass by percussion; an abdomen filled with gas is tympanic, whereas an abdomen containing a mass or fluid is dull to percussion. The absence of abdominal dullness, however, does not exclude ascites, because a minimum of 1500 mL of ascitic fluid is required for detection on physical examination. The abdomen should be palpated to assess for tenderness, a mass, enlargement of the spleen or liver, or presence of a nodular liver suggesting cirrhosis or tumor. Light palpation of the liver may detect pulsations suggesting retrograde vascular flow from the heart in patients with right-sided heart failure, particularly tricuspid regurgitation. A rectal examination may help identify an evacuation disorder in patients with constipation.

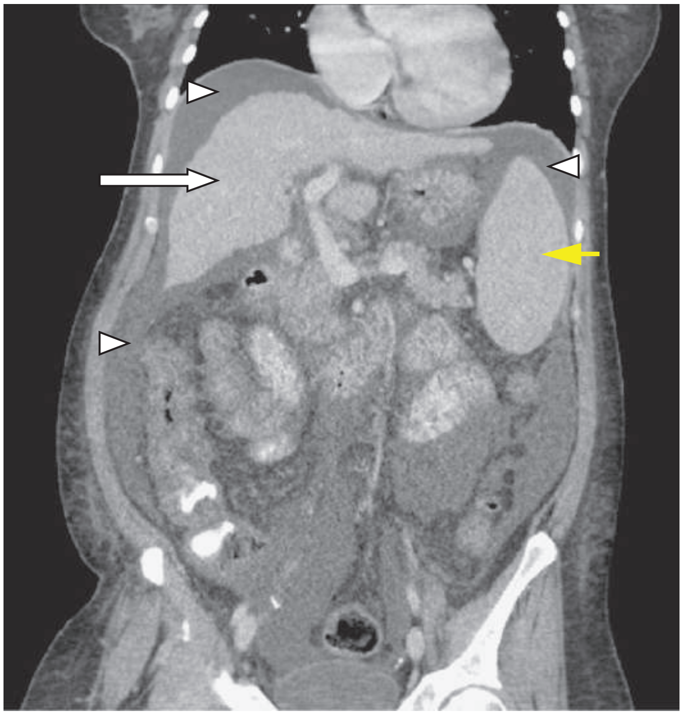

FIGURE 53-1 CT of a patient with a cirrhotic, nodular liver (*white arrow*), splenomegaly (*yellow arrow*), and ascites (*arrowheads*).

IMAGING AND LABORATORY EVALUATION

Abdominal x-rays can be used to detect dilated loops of bowel suggesting intestinal obstruction or ileus. Motility studies may be considered in patients with severe constipation. Abdominal ultrasonography can detect as little as 100 mL of ascitic fluid, hepatosplenomegaly, a nodular liver, or a mass. Ultrasonography is often inadequate to detect retroperitoneal lymphadenopathy or a pancreatic lesion because of overlying bowel gas. If malignancy or pancreatic disease is suspected, CT can be performed. CT may also detect changes associated with advanced cirrhosis and portal hypertension **(Fig. 53-1)**.

Laboratory evaluation should include liver biochemical testing, serum albumin level measurement, and prothrombin time determination (international normalized ratio) to assess hepatic function as well as a complete blood count to evaluate for the presence of cytopenias that may result from portal hypertension or of leukocytosis, anemia, and thrombocytosis that may result from systemic infection. Serum amylase and lipase levels should be checked to evaluate the patient for acute pancreatitis. Urinary protein quantitation is indicated when nephrotic syndrome, which may cause ascites, is suspected. Hydrogen and methane absorbed from the intestine are not metabolized by the host and are excreted in expired air, and detection of increased amounts of these gases in expired breath is the basis for tests used to diagnose carbohydrate (e.g., lactose) malabsorption and small intestinal bacterial overgrowth, although the reliability of the test results for the diagnosis of small intestinal bacterial overgrowth has been questioned.

In selected cases, the hepatic venous pressure gradient (pressure across the liver between the portal and hepatic veins) can be measured via cannulation of the hepatic vein to confirm that ascites is caused by cirrhosis. In some cases, noninvasive tests for liver fibrosis, including liver stiffness measurement by elastography, or a liver biopsy may be necessary to confirm cirrhosis **(Chap. 355)**.

ASCITES

PATHOGENESIS IN THE PRESENCE OF CIRRHOSIS

Ascites in patients with cirrhosis is the result of portal hypertension and renal salt and water retention. Similar mechanisms contribute to ascites formation in heart failure. Portal hypertension signifies elevation of the pressure within the portal vein. According to Ohm's law, pressure is the product of resistance and flow. Increased hepatic resistance occurs by several mechanisms. First, the development of hepatic fibrosis, which defines cirrhosis, disrupts the normal architecture of the hepatic sinusoids and impedes normal blood flow through the liver. Second, activation of hepatic stellate cells, which mediate fibrogenesis, leads to smooth-muscle contraction and fibrosis. Finally, cirrhosis is associated with a decrease in endothelial nitric oxide synthetase (eNOS) production, which results in decreased nitric oxide production and increased intrahepatic vasoconstriction.

The development of cirrhosis is also associated with increased systemic circulating levels of nitric oxide (in contrast to the decrease seen intrahepatically), as well as increased levels of vascular endothelial growth factor and tumor necrosis factor, that result in splanchnic arterial vasodilation. Vasodilation of the splanchnic circulation results in pooling of blood and a decrease in the effective circulating volume, which is perceived by the kidneys as hypovolemia. Compensatory vasoconstriction via release of antidiuretic hormone ensues; the consequences are free water retention and activation of the sympathetic nervous system and the renin-angiotensin-aldosterone system, which lead in turn to renal sodium and water retention.

PATHOGENESIS IN THE ABSENCE OF CIRRHOSIS

Ascites in the absence of cirrhosis generally results from peritoneal carcinomatosis, peritoneal infection, pancreatic disease, or marked hypoalbuminemia. Peritoneal carcinomatosis can result from primary peritoneal malignancies such as mesothelioma or sarcoma, abdominal malignancies such as gastric or colonic adenocarcinoma, or metastatic disease from breast or lung carcinoma or melanoma **(Fig. 53-2)**. The tumor cells lining the peritoneum produce a protein-rich fluid that contributes to the development of ascites. Fluid from the extracellular space is drawn into the peritoneum, further contributing to the development of ascites. Tuberculous peritonitis causes ascites via a similar mechanism; tubercles deposited on the peritoneum exude a proteinaceous fluid. Pancreatic ascites results from leakage of pancreatic enzymes into the peritoneum. Marked hypoalbuminemia may result from nephrotic syndrome, protein-losing enteropathy, or malnutrition.

CAUSES

Cirrhosis accounts for 84% of cases of ascites. Cardiac ascites, peritoneal carcinomatosis, and "mixed" ascites resulting from cirrhosis and a second disease account for 10–15% of cases. Less common causes of ascites include massive hepatic metastasis, infection (tuberculosis, *Chlamydia* infection), pancreatitis, and renal disease (nephrotic syndrome). Rare causes of ascites include hypothyroidism and familial Mediterranean fever.

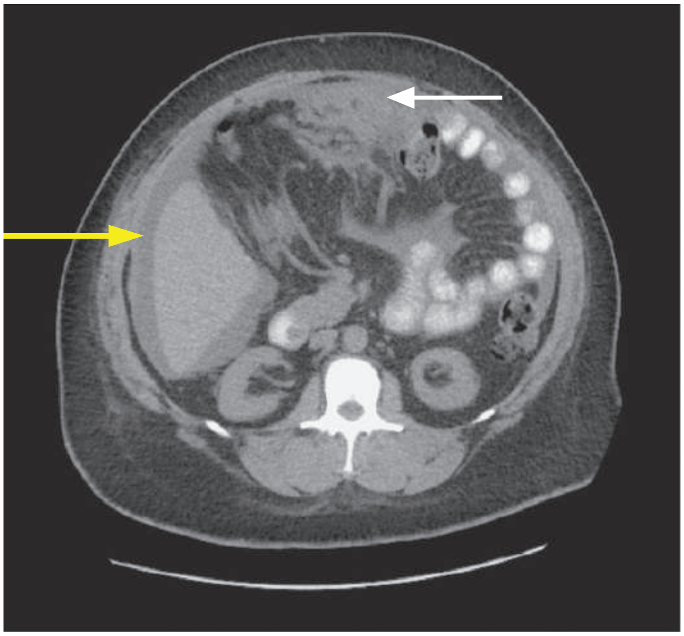

FIGURE 53-2 CT of a patient with peritoneal carcinomatosis (*white arrow*) and ascites (*yellow arrow*).

EVALUATION

Once the presence of ascites has been confirmed, the etiology of the ascites is best determined by *paracentesis*, a bedside procedure in which a needle or small catheter is passed transcutaneously to extract ascitic fluid from the peritoneum. The lower quadrants are the most frequent sites for paracentesis. The left lower quadrant is preferred because of the greater depth of ascites and the thinner abdominal wall. Paracentesis is a safe procedure even in patients with coagulopathy; complications, including abdominal wall hematomas, hypotension, hepatorenal syndrome, and infection, are infrequent.

Once ascitic fluid has been extracted, its gross appearance should be examined. Turbid fluid can result from the presence of infection or tumor cells. White, milky fluid indicates a triglyceride level >200 mg/dL (and often >1000 mg/dL), which is the hallmark of *chylous ascites*. Chylous ascites results from lymphatic disruption that may occur with trauma, cirrhosis, tumor, tuberculosis, or certain congenital abnormalities. Dark brown fluid can reflect a high bilirubin concentration and indicates biliary tract perforation. Black fluid may indicate the presence of pancreatic necrosis or metastatic melanoma.

The ascitic fluid should be sent for measurement of albumin and total protein levels, cell and differential counts, and, if infection is suspected, Gram's stain and culture, with inoculation into blood culture bottles at the patient's bedside to maximize the yield. A serum albumin level should be measured simultaneously to permit calculation of the *serum-ascites albumin gradient* (SAAG).

The SAAG is useful for distinguishing ascites caused by portal hypertension from nonportal hypertensive ascites **(Fig. 53-3)**. The SAAG reflects the pressure within the hepatic sinusoids and correlates with the hepatic venous pressure gradient. The SAAG is calculated by subtracting the ascitic albumin concentration from the serum albumin level and does not change with diuresis. A SAAG ≥1.1 g/dL reflects the presence of portal hypertension and indicates that the ascites is due to increased pressure in the hepatic sinusoids. According to Starling's law, a high SAAG reflects the oncotic pressure that counterbalances the portal pressure. Possible causes include cirrhosis, cardiac ascites, hepatic vein thrombosis (Budd-Chiari syndrome), sinusoidal obstruction syndrome (veno-occlusive disease), or massive liver metastases. A SAAG <1.1 g/dL indicates that the ascites is not related to portal hypertension, as in tuberculous peritonitis, peritoneal carcinomatosis, pancreatic ascites, or nephrotic syndrome.

For high-SAAG (≥1.1) ascites, the ascitic protein level can provide further clues to the etiology (Fig. 53-3). An ascitic protein level of ≥2.5 g/dL indicates that the hepatic sinusoids are normal and are allowing passage of protein into the ascites, as occurs in cardiac ascites, early Budd-Chiari syndrome, or sinusoidal obstruction syndrome. An ascitic protein level <2.5 g/dL indicates that the hepatic sinusoids have been damaged and scarred and no longer allow passage of protein, as occurs with cirrhosis, late Budd-Chiari syndrome, or massive liver metastases. Pro-brain-type natriuretic peptide (BNP) is a natriuretic hormone released by the heart as a result of increased volume and ventricular wall stretch. High levels of BNP in serum occur in heart failure and may be useful in identifying heart failure as the cause of high-SAAG ascites.

Further tests are indicated only in specific clinical circumstances. When secondary peritonitis resulting from a perforated hollow viscus is suspected, ascitic glucose and lactate dehydrogenase (LDH) levels can be measured. In contrast to "spontaneous" bacterial peritonitis, which may complicate cirrhotic ascites (see "Complications," below), secondary peritonitis is suggested by an ascitic glucose level <50 mg/dL, an ascitic LDH level higher than the serum LDH level, and the detection of multiple pathogens on ascitic fluid culture. When pancreatic ascites is suspected, the ascitic amylase level should be measured and is typically >1000 mg/dL. Cytology can be useful in the diagnosis of peritoneal carcinomatosis. At least 50 mL of fluid should be obtained and sent for immediate processing. Tuberculous peritonitis is typically associated with ascitic fluid lymphocytosis but can be difficult to

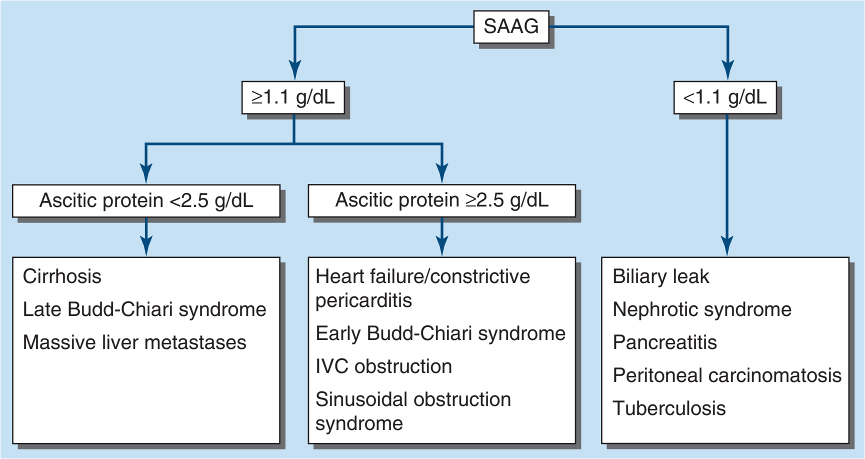
SAAG
≥1.1 g/dL
<1.1 g/dL
Ascitic protein <2.5 g/dL
Ascitic protein ≥2.5 g/dL
Cirrhosis
Late Budd-Chiari syndrome
Massive liver metastases
Heart failure/constrictive pericarditis
Early Budd-Chiari syndrome
IVC obstruction
Sinusoidal obstruction syndrome
Biliary leak
Nephrotic syndrome
Pancreatitis
Peritoneal carcinomatosis
Tuberculosis

FIGURE 53-3 Algorithm for the diagnosis of ascites according to the serum-ascites albumin gradient (SAAG). IVC, inferior vena cava.

diagnose by paracentesis. A smear for acid-fast bacilli has a diagnostic sensitivity of only 0–3%; a culture increases the sensitivity to 35–50%. In patients without cirrhosis, an elevated ascitic adenosine deaminase level has a sensitivity of >90% for tuberculous ascites when a cut-off value of 30–45 U/L is used. When the cause of ascites remains uncertain, laparotomy or laparoscopy with peritoneal biopsies for histology and culture remains the gold standard.

TREATMENT

Ascites

The initial treatment for cirrhotic ascites is moderate restriction of sodium intake to 2 g/d. When sodium restriction alone is inadequate to control ascites, oral diuretics—typically the combination of spironolactone and furosemide—are used to increase urinary sodium excretion. Spironolactone is an aldosterone antagonist that inhibits sodium resorption in the distal convoluted tubule of the kidney. Use of spironolactone may be limited by hyponatremia, hyperkalemia, and painful gynecomastia. If the gynecomastia is distressing, amiloride (5–40 mg/d) may be substituted for spironolactone. Furosemide is a loop diuretic that is generally combined with spironolactone in a ratio of 40:100; maximal daily doses of spironolactone and furosemide are generally 400 mg and 160 mg, respectively. Fluid intake may be restricted in patients with hyponatremia (serum sodium <125 mEq/L). Nonsteroidal anti-inflammatory drugs, angiotensin-converting enzyme inhibitors, and angiotensin receptor blockers should be avoided in patients with cirrhosis and ascites.

Refractory cirrhotic ascites is defined by the persistence of ascites despite sodium restriction and maximum diuretic use (diuretic resistant) or the development of side effects of diuretics that preclude the use of maximum doses (diuretic intractable). Pharmacologic therapy for refractory ascites includes the addition of midodrine, an α_1-adrenergic agonist, or clonidine, an α_2-adrenergic agonist, to diuretic therapy. These agents act as vasoconstrictors, counteracting splanchnic vasodilation. Midodrine alone or in combination with clonidine improves systemic hemodynamics and control of ascites over that obtained with diuretics alone. Although β-adrenergic blocking agents (beta blockers) are often prescribed to prevent variceal hemorrhage in patients with cirrhosis, the use of beta blockers in patients with refractory ascites may be associated with decreased survival rates.

When medical therapy alone is insufficient, refractory cirrhotic ascites can be managed by repeated large-volume (>5 L) paracentesis (LVP) or a transjugular intrahepatic peritoneal shunt (TIPS)—a radiologically placed portosystemic shunt that decompresses the hepatic sinusoids. Intravenous (IV) infusion of albumin accompanying LVP decreases the risk of "postparacentesis circulatory dysfunction" and death. Patients undergoing LVP should receive IV albumin infusions of 6–8 g/L of ascitic fluid removed. TIPS placement is superior to LVP in reducing the reaccumulation of ascites but is associated with an increased frequency of hepatic encephalopathy, with no difference in mortality rates. The Alfapump system, which consists of an automated pump and tunneled peritoneal catheter that transports ascites from the peritoneal cavity to the urinary bladder, has shown promise in the management of refractory ascites but is associated with a high frequency of technical difficulties and renal dysfunction.

Malignant ascites does not respond to sodium restriction or diuretics. Patients must undergo serial LVPs, transcutaneous drainage catheter placement, or, rarely, creation of a peritoneovenous shunt (a shunt from the abdominal cavity to the vena cava) or placement of the Alfapump system, if available.

Ascites caused by tuberculous peritonitis is treated with standard antituberculosis therapy. Noncirrhotic ascites of other causes is treated by correction of the precipitating condition.

COMPLICATIONS

Spontaneous bacterial peritonitis (SBP; **Chap. 137**) is a common and potentially lethal complication of cirrhotic ascites. Occasionally, SBP also complicates ascites caused by nephrotic syndrome, heart failure, acute hepatitis, and acute liver failure but is rare in malignant ascites. Patients with SBP generally note an increase in abdominal girth; however, abdominal tenderness is found in only 40% of patients, and rebound tenderness is uncommon. Patients may present with fever, nausea, vomiting, or the new onset or an exacerbation of preexisting hepatic encephalopathy.

In hospitalized patients with ascites, paracentesis within 12 hours of admission reduces mortality because of early detection of SBP. SBP is defined by a polymorphonuclear neutrophil (PMN) count of ≥250/μL in the ascitic fluid. Cultures of ascitic fluid should be performed in blood culture bottles and typically reveal one bacterial pathogen. The presence of multiple pathogens in the setting of an elevated ascitic PMN count suggests *secondary peritonitis* from a ruptured viscus or abscess **(Chap. 137)**. The presence of multiple pathogens without an elevated PMN count suggests bowel perforation from the paracentesis needle. SBP is generally the result of enteric bacteria that have translocated across an edematous bowel wall. The most common pathogens are gram-negative rods, including *Escherichia coli* and *Klebsiella*, as well as streptococci and enterococci.

Treatment of SBP has traditionally been with a third-generation cephalosporin such as IV cefotaxime, 2 g every 12 hours, and is generally effective against gram-negative and gram-positive aerobes. A 5-day course of treatment is sufficient if the patient improves clinically. Repeat paracentesis is recommended after 48 hours of antibiotic therapy to confirm that the ascitic PMN count has decreased by at least 25% from baseline. Increasingly, SBP, particularly if nosocomial or health care–acquired, is caused by multidrug-resistant bacteria, and initial antibiotic therapy should be guided by the local bacterial epidemiology. Initial use of a carbapenem is often considered in patients with nosocomial SBP or those who are critically ill.

Cirrhotic patients with a history of SBP, an ascitic fluid total protein concentration <1 g/dL, or active gastrointestinal bleeding should receive prophylactic antibiotics to prevent SBP; oral daily ciprofloxacin or, where available, norfloxacin is commonly used. IV ceftriaxone may be used in hospitalized patients. Diuresis increases the activity of ascitic fluid protein opsonins and may decrease the risk of SBP.

Hepatic hydrothorax occurs when ascites, often caused by cirrhosis, migrates via fenestrae in the diaphragm into the pleural space; a serum-to-pleural fluid albumin gradient ≥1.1 g/dL is suggestive of hepatic hydrothorax. This condition can result in shortness of breath, hypoxia, and infection (*spontaneous bacterial empyema*). Treatment is similar to that for cirrhotic ascites and includes sodium restriction, diuretics, and, if needed, thoracentesis or TIPS placement. Chest tube placement should be avoided.

Abdominal wall and inguinal hernias are common in patients with cirrhosis and ascites. *Hepatorenal syndrome* is a potential complication of cirrhotic ascites **(Chap. 355)**.

■ FURTHER READING

Adebayo D et al: Refractory ascites in liver cirrhosis. Am J Gastroenterol 114:40, 2019.

Aithal GP et al: Guidelines on the management of ascites in cirrhosis. Gut 70:9, 2021.

Barba E et al: Correction of abdominal distention by biofeedback-guided control of abdominothoracic muscular activity in a randomized, placebo-controlled trial. Clin Gastroenterol Hepatol 15:1922, 2017.

Bernardi M et al: Albumin infusion in patients undergoing large-volume paracentesis: A meta-analysis of randomized trials. Hepatology 55:1172, 2012.

Biggins SW et al: Diagnosis, evaluation, and management of ascites, spontaneous bacterial peritonitis and hepatorenal syndrome: 2021 Practice Guidance by the American Association for the Study of Liver Diseases. Hepatology 74:1014, 2021.

Damianos JA et al: Abdominophrenic dyssynergia: A narrative review. Am J Gastroenterol 118:41, 2023.

Farias AQ et al: Serum B-type natriuretic peptide in the initial workup of patients with new onset ascites: A diagnostic accuracy study. Hepatology 59:1043, 2014.

Fernandez J et al: Prevalence and risk factors of infections by multiresistant bacteria in cirrhosis: A prospective study. Hepatology 55:1551, 2012.

John S, Friedman LS: Portal hypertensive ascites: Current status. Curr Hepatol Rep 19:226, 2020.

John S, Thuluvath PJ: Hyponatremia in cirrhosis: Pathophysiology and management. World J Gastroenterol 21:3197, 2015.

Lacy BE et al: Management of chronic abdominal distension and bloating. Clin Gastroenterol Hepatol 19:219, 2021.

Lepida A et al: Systematic review with meta-analysis: Automated low-flow ascites pump therapy for refractory ascites. Aliment Pharmacol Ther 50:978, 2019.

Lizaola B et al: Review article: The diagnostic approach and current management of chylous ascites. Aliment Pharmacol Ther 46:816, 2017.

Orman ES et al: Paracentesis is associated with reduced mortality in patients hospitalized with cirrhosis and ascites. Clin Gastroenterol Hepatol 12:496, 2014.

Runyon BA et al: The serum-ascites albumin gradient is superior to the exudate-transudate concept in the differential diagnosis of ascites. Ann Intern Med 117:215, 1992.

Trebicka J: Role of albumin in the treatment of decompensated liver cirrhosis. Curr Opin Gastroenterol 38:200, 2022.

Williams JW Jr, Simel DL: The rational clinical examination. Does this patient have ascites? How to divine fluid in the abdomen. JAMA 267:2645, 1992.

Section 7 Alterations in Renal and Urinary Tract Function

54

Interstitial Cystitis/ Bladder Pain Syndrome

R. Christopher Doiron, J. Curtis Nickel

DEFINITION

A condition associated with bladder inflammation and pain, with what were thought to be discrete bladder ulcerations, was first described in 1887. The description of the classic bladder-wall ulcer—now referred to as a *Hunner lesion*—became known as *interstitial cystitis* (IC). The first generally accepted definition of IC was derived from a National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) consensus of experts in the field in 1998. The NIDDK criteria used to define IC included typical cystoscopic findings such as glomerulations (submucosal petechial hemorrhages of the urothelium) or Hunner lesions. However, over time, the syndrome experienced by patients, including bladder and/or pelvic pain with associated urinary storage symptoms of urinary frequency and urgency, negative urine cultures, and no specific identifiable causes, became known as *interstitial cystitis/bladder pain syndrome* (IC/BPS).

The nomenclature and definitions have evolved, but the contemporary definitions accepted by the American Urological Association, the Canadian Urological Association, the International Continence Society, the Society for Urodynamics and Female Urology, and the European Society for the Study of IC/BPS, although they all differ somewhat in language and specifics, generally reflect several fundamental concepts common in the disease: (1) it is chronic in nature; (2) it causes pain perceived to be attributable to the bladder; (3) this pain occurs in the presence of lower urinary tract symptoms (LUTS); and (4) pain outside the bladder—in the pelvis, perineum, genitals, abdomen, and beyond—is common.

The following definition incorporates the major descriptions by all international groups interested in the diagnosis and management of IC/BPS: an unpleasant sensation (pain, pressure, discomfort) perceived to be related to the urinary bladder, associated with LUTS of >6 weeks' duration, in the absence of infection or other identifiable causes.

A generalized urologic chronic pelvic pain syndrome (UCPPS) encompasses two distinct urologic chronic pain disorders: IC/BPS, which may be present in men and women, and chronic prostatitis/chronic pelvic pain syndrome (CP/CPPS), which is present only in men. The latter refers to a urologic pain disorder with pain localized to the perineum and/or male genitals, with or without LUTS. IC/BPS can exist independent of CP/CPPS in men. In reality, the urologic chronic pain disorders often have overlapping symptom presentations and may share common etiologic and pathophysiologic origins, but the focus of the current chapter will be on IC/BPS.

ETIOLOGY AND PATHOGENESIS

Pinning a single etiology to a diagnosis of IC/BPS has been an endeavor fraught with uncertainty that ultimately has failed thus far. Instead, it is much more likely and widely accepted that IC/BPS represents a syndrome or constellation of interrelated disease processes that manifest in a spectrum of disease that reaches beyond the bladder. While the search for a single etiology soldiers on, we will review here a collection of proposed theories.

■ INFECTION AND THE URINARY MICROBIOTA

Bacterial infection of the urothelium has long been regarded as a major suspect in the etiology of IC/BPS but has never been definitively shown to cause the disease. It is not uncommon for patients presenting with

IC/BPS to describe a long history of “urinary tract infections” (UTIs); these patients often have undergone multiple courses of treatment with one or multiple antibiotics prescribed by their physicians. Often, however, in patients with IC/BPS, the benefit of antibiotic treatment is short-lived, urine culture results are negative, and the return of symptoms is inevitable.

Although studies examining the role of microbiologic organisms in this patient population are numerous and the results conflicting, far more studies have yielded negative results rather than positive findings. Furthermore, our understanding of the urinary microbiota continues to expand, rendering older studies using outdated and insensitive cultivation techniques less relevant.

Using state-of-the-art, culture-independent techniques for microorganism identification, investigators observe subtle differences between the urinary microbiota of IC/BPS patients and that of healthy controls and between IC/BPS patients experiencing symptom flares and IC/BPS patients not in flare. The clinical relevance of these findings is still not fully understood. As the study of the urinary microbiota continues to unfold, researchers and clinicians believe that although a single causative microbe is unlikely, dysbiosis or disturbance in the microbial ecology of the lower urinary tract may be responsible for flares or symptom patterns experienced by IC/BPS patients.

AUTOIMMUNITY

The consideration of IC/BPS as a disorder of the immune system stems from the observation of a significant prevalence of autoimmune disorders in IC/BPS patients; several historic studies have identified antiurothelial antibodies and B cell expansion within the bladder mucosa of IC/BPS patients. It has been recognized that the prevalence of B cell–mediated autoimmune conditions (e.g., rheumatoid arthritis, systemic lupus erythematosus, Sjögren’s syndrome) amongst IC/BPS patients is significantly increased compared with healthy controls. Furthermore, a shared B cell clonotype amongst a cohort of IC/BPS patients with Hunner lesions has recently been identified, representing the most robust evidence of antigen-specific B cells as a possible pathoetiology of disease in some patients.

Although IC/BPS is not a pathologic diagnosis, there are widely accepted, recognizable patterns of inflammatory infiltration in the bladder mucosa of this patient population, including lymphoplasmacytic infiltrates, stromal edema and fibrosis, urothelial denudation, and detrusor mastocytosis. Thus, although it is likely that immune disturbances cause the condition in a subset of patients (for example, in those with associated Sjögren’s syndrome), researchers and clinicians have been unable to leverage this knowledge in describing a common mechanism, and its clinical relevance is not fully understood.

INFLAMMATION

It is well established that a subset of patients suffering from IC/BPS clearly have associated bladder inflammation of unknown etiology (in contrast to the autoimmune-based mechanisms of inflammation discussed above). The best described of these patients are those with Hunner lesions—discrete inflammatory lesions, previously believed to be ulcers, that have a well-characterized inflammatory profile on histologic and pathologic analysis. While Hunner lesions are easily identified under direct vision by cystoscopy, a spectrum of other, less obvious inflammatory patterns in the bladder is associated with infiltration of acute and chronic inflammatory cells and mast cells. This inflammation observed on histologic analysis can be so subtle that it cannot be recognized under direct visual examination of the bladder with cystoscopy.

Investigators in the Multidisciplinary Approach to the Study of Chronic Pelvic Pain (MAPP) Research Network have found that, among patients with UCPPS, women exhibit more robust inflammatory responses to stimulation of Toll-like receptor 2 (TLR2) and TLR4. Furthermore, an increased response to stimulation of TLR4 predicts more severe symptoms, widespread pain (vs pelvic/bladder pain only), and a higher number of chronic overlapping pain conditions (COPCs). Further studies aimed at a better understanding of these findings are underway.

UROTHELIAL DYSFUNCTION

Urothelial Permeability and the Glycosaminoglycan Layer

The stratified epithelium of the bladder—the urothelium—is composed of basal precursor cells, intermediate cells, and a layer of specialized, superficial epithelial cells called *umbrella cells*. Collectively, these layers are responsible for the various functions of the bladder lining. One important function of the urothelium is to provide a robust barrier layer. This function is fulfilled by the dense layering of glycosaminoglycans (GAGs) on the luminal surface of the urothelium along with a complex arrangement of numerous intercellular tight junctions among urothelial cells that protects the underlying bladder interstitium from the constituents of the urine resting in the bladder.

Defects in this barrier function—either disruptions in the GAG layer or disruptions in the epithelial layer itself or its cellular junctions—have been proposed as a possible mechanism for bladder pain in IC/BPS patients. This theory, while still popular, lacks definitive evidence supporting it as the etiology of this disease.

Antiproliferative Factor The discovery that urothelial cells from IC/BPS patients appear to grow far more slowly than urothelial cells from a healthy control population led to the identification of antiproliferative factor (APF). Although APF initially showed promise as a sensitive and specific urine biomarker for IC/BPS, this idea has not been widely adopted, and the etiologic role of APF is not yet fully understood.

PELVIC ORGAN CROSSTALK

The observation of dysfunction and symptoms in multiple organ systems, including gastrointestinal, gynecologic, and genital organs, in patients with IC/BPS is so common that it might be considered the norm. Mechanisms of neural sensitization in patients with chronic pain have been reported, and abnormalities in the autonomic nervous system have been observed among IC/BPS patients. Again, although these observations apply in a subset of patients, their broader application to the heterogeneous IC/BPS patient population as a clear cause of disease is not yet warranted.

NEUROBIOLOGIC CONTRIBUTIONS AND CENTRAL SENSITIZATION

One breakthrough by the MAPP Research Network is an investigation of the role of structural and functional alterations in the brains of patients with UCPPS. The network’s innovative methods of correlating clinical and deep phenotyping data with functional MRI data identified such structural and functional differences. These differences were later shown to successfully predict the progression of symptoms in a cohort of patients with UCPPS. Although the relevant study did not differentiate between IC/BPS and CP/CPPS patients, the findings are nevertheless informative, and further longitudinal studies are ongoing.

In addition to the novel MAPP-led findings using neuroimaging, quantitative sensory testing (QST) methods have been used to investigate the sensory processing mechanisms in UCPPS patients. The findings—generalized pain hypersensitivity and altered endogenous inhibitory pain control systems among UCPPS patients—further support a hypothesis of a central sensitization phenotype in urologic chronic pelvic pain. The clinical implications of observed neural alterations and multisensory hypersensitivity remain under investigation.

Although a single etiology for this clinically heterogeneous pain syndrome may never be identified, efforts to do so have revealed much about its pathogenesis in subsets of patients and have provided valuable insight into specific patient phenotypes. The challenge for researchers and clinicians moving forward will be to unify clinical phenotypes with these proposed underlying mechanisms of disease and to integrate this knowledge into clinically actionable interventions that may provide meaningful outcomes.

EPIDEMIOLOGY

The prevalence of IC/BPS has been difficult to determine because definitions and diagnostic criteria (in the absence of a definitive diagnostic test or biomarker) are constantly evolving. In addition, the various methods used in attempting to describe the syndrome's epidemiology (patient self-reports, symptom-based surveys, physician visits, population-based databases) have been problematic and have made comparisons of results challenging. Many studies have historically been performed only in female populations. Currently, it is estimated that 2.7–6.5% of North American women experience symptoms consistent with a diagnosis of IC/BPS. Fewer than 10% of women who experience these symptoms actually have a diagnosis of IC/BPS. The syndrome does occur in men, with a reported 10:1 female-to-male ratio, but it is thought that the condition is dramatically underreported in men.

Some predictors of the development of IC/BPS have been suggested through an analysis of retrospective observational studies of childhood disorders and adverse childhood experiences (ACEs), including childhood UTI, childhood bowel and bladder dysfunction, and childhood sexual trauma. Furthermore, it has been well established that IC/BPS patients exhibit a remarkable prevalence of COPCs such as fibromyalgia, irritable bowel syndrome (IBS), chronic back pain, and chronic fatigue syndrome (CFS). Recent MAPP-led studies have shown that more than one-third of IC/BPS patients have one COPC (IBS, fibromyalgia, or CFS), while up to 10% have multiple COPCs. Thus, these conditions might be considered risk factors for the development of IC/BPS.

CLINICAL MANIFESTATIONS

Patients with IC/BPS, both female and male, present with varying degrees of discomfort and/or pain perceived to be related to the bladder and associated with urinary storage symptoms, including daytime and nighttime urinary frequency and urinary urgency. For some patients, urinary symptoms (the most common complaint after bladder pain) are the most bothersome, while for most patients, bladder pain causes the most distress and most significantly affects quality of life. Unfortunately, most patients with IC/BPS present with both types of symptoms, as patients void frequently to relieve pain (or because of fear of bladder pain). Typically, this combination of bladder pain and urinary frequency severely impacts patients' quality of life, social interactions, and physical activities.

Pain-mapping studies have been used to identify different pain phenotypes within the disease. Nickel and colleagues (2024) first described a bladder-only phenotype present in 20% of a cohort of female IC/BPS patients, whereas up to 80% of patients described pain in the pelvis and at least one site beyond. Common associated conditions include IBS (40%), pelvic floor dysfunctional pain syndrome (40–60%), vulvodynia (17%), fibromyalgia (36%), CFS (10%), and chronic back pain (47%). As described above, these multiorgan symptoms may be due to central nervous system sensitization and associated spinal crosstalk, which may promote phenotypic progression as patients with one pain syndrome slowly progress to another. Subsequent MAPP-led studies among a more heterogeneous UCPPS cohort of men and women have supported this concept of specific pain phenotypes, reporting a pelvic-pain-only phenotype in 25% of participants and pain in the pelvis and beyond in up to 75%.

Another important finding from the MAPP investigations is their identification of not just a pelvic-pain-only phenotype but also a bladder-focused phenotype. The latter phenotype was identified by patients' responses to two RAND Interstitial Cystitis Epidemiology (RICE) survey questions: whether they had "painful bladder filling" and/or "painful urinary urgency." Most female UCPPS patients (88%) responded "yes" to at least one of these questions. The bladder-focused phenotype was associated with more severe urologic symptoms and worse quality of life.

Patients present with unique pain trajectories. Some initially have mild discomfort that progresses over many years to pain with bladder filling and finally to chronic unremitting pelvic pain with only short periods of relief with urination. Other patients begin with UTI-like symptoms and acute bladder and urethral pain with urinary frequency and urgency; these manifestations persist as a chronic cystitis–like syndrome despite negative cultures and no benefit from antimicrobial therapy. Still other patients report a waxing and waning of pain over time, with flares exacerbated by diet, anxiety/stress, infection, or hormone cycle (typically with increased pain prior to menses). In a longitudinal study of UCPPS patients followed over a 12-month period during routine care for their disease, MAPP investigators described 60% of patients' symptoms as stable, 20% as improved, and 20% as worsened.

APPROACH TO THE PATIENT

Interstitial Cystitis/Bladder Pain Syndrome

Patients with IC/BPS present to their family physician or internist with pelvic pain that typically increases in severity with bladder filling, with associated pain within and/or outside the pelvis, and with various degrees of urinary symptomatology. The course that should be followed by primary care practitioners or general internists during the patient's workup and before referral to a specialist is outlined in **Table 54-1**. Most of these physicians will not move beyond a suspected diagnosis and conservative advice; that is acceptable. Patients with IC/BPS can often represent diagnostic challenges, and referral to an appropriate subspecialist is warranted if any diagnostic uncertainty remains.

A diagnosis of IC/BPS is often missed and can be delayed for many years because physicians tend to silo patients into various medical-specialty streams on the basis of the most bothersome symptom. For example, patients presenting with pelvic pain in which flares are associated with monthly menstrual cycles may be referred to gynecologists. Patients with abdominal/pelvic pain associated with diarrhea and/or constipation tend to be referred to gastroenterologists, while those with generalized muscle and joint pain, perhaps associated with fatigue, may be referred to rheumatologists. Patients with urinary symptoms and bladder pain are treated for UTIs (even with negative urine cultures) or overactive bladder—a common bladder condition associated with urinary frequency and urgency, but not pain.

It would be simple if the approach to patients presenting with pelvic pain was only to determine the actual pelvic organ and/or disease causing the symptoms. However, spinal crosstalk, phenotype progression over time, central sensitization, and COPCs complicate the picture. Since only 20–25% of patients eventually diagnosed with IC/BPS have bladder-only disease, one must not be bladder-centric in approach but rather must consider the entire patient. The provider must determine the patient's "clinical picture"—that is, the patient's unique presenting clinical phenotype.

Several symptom questionnaires are available for use in IC/BPS patients but would be considered beyond the scope of a general pracititioner. Approach to history, though, in a patient suspected of IC/BPS should separate the assessment of a patient's urinary

TABLE 54-1 Workup of Patients by a Primary Care Practitioner or General Internist

STEPS IN WORKUP	SPECIFICS
History/physical examination	Conduct a pelvic exam (recommended), including evaluation of the pelvic floor. Categorize symptoms as bladder/pelvis focused and/or extending beyond the pelvis.
Urine studies	Perform a microscopic urinalysis and urine culture with sensitivity testing.
Consideration of patient-centered treatment options if satisfied with diagnosis[a]	Begin with conservative measures. Introduce further symptom-specific treatments as needed.
Referral to an appropriate specialist under certain conditions	Referral should follow if: • The diagnosis is unclear • Microscopic or gross hematuria is present • The condition is refractory to treatment • Symptoms are severe • The presentation is complex

[a] See text.

symptoms from their pain symptoms and evaluate the degree of bother attributable to each. This distinction is supported by evidence from the MAPP network's symptom studies and will help guide patient-specific treatment approaches.

Sexual dysfunction is common amongst this patient population. This should be inquired about and can play a significant role in quality of life of IC/BPS patients. Symptom flares are another common feature of IC/BPS patients' symptomology, and their frequency, severity, and strategy for mitigation should be evaluated.

A further simplified clinical approach to the assessment of patients with symptoms of IC/BPS is to classify patients with perceived bladder pain (a mandatory criterion for diagnosis) into one of two categories: (1) a "pelvic pain–only" category, which would include the "bladder pain–only," pelvic floor dysfunctional pain, and associated gynecologic pain groups; or (2) a "widespread pain" category, which would include patients with associated COPCs (such as IBS and fibromyalgia). This approach has been supported by recent observations from the MAPP investigators.

The contribution of psychosocial parameters, such as depression, catastrophizing, anxiety, and stress, and their impact on pain and disability cannot be overlooked and are important to ascertain in all cases. This approach to clinical phenotyping will let the physician tailor a unique treatment plan for each individual patient, using combinations of local bladder, pelvic floor, or more general systemic therapies.

DIAGNOSIS

IC/BPS is a clinically heterogeneous condition whose lack of a clear etiopathogenesis presents difficulties in diagnosis. In making a diagnosis of exclusion, clinicians must rule out other confusable diseases and identify to the best of their ability the phenotypic presentation of the presenting patient. Although attempts have been made to establish a set of diagnostic criteria in the past, the specified criteria have proved overly stringent and too exclusive to be clinically useful. Furthermore, although several guidelines exist to aid in decision-making in diagnostic investigations, most investigations serve merely to rule out other pathology. In contrast, history and physical examination, along with some simple laboratory testing, are the most reliable tools with which to establish a diagnosis of IC/BPS. Details of relevant investigations, some of which may be beyond the scope of the general practitioner or internist, are presented here. Table 54-1 offers an approach for the general practitioner, and Table 54-2 provides a more complete summary of diagnostic recommendations.

HISTORY AND PHYSICAL (INCLUDING FREQUENCY/VOLUME CHARTS)

A thorough history and physical examination are of utmost importance in diagnosing IC/BPS. A history of the patient's pain symptoms is a logical place to start. The nature, intensity, and timing of the pain are all significant factors. Some patients will be less explicit than others in describing their pain and may instead describe a sense of pressure, burning, or vague fullness in the pelvis or bladder area.

TABLE 54-2 Recommendations for Investigations in Patients with Suspected Interstitial Cystitis/Bladder Pain Syndrome

MANDATORY	RECOMMENDED	OPTIONAL	NOT RECOMMENDED
History	Frequency/volume chart	Ultrasound/pelvic imaging	Potassium sensitivity test
Physical examination (including pelvic exam and examination of pelvic floor)	Urinalysis	Postvoid residual	Urodynamics
	Urine culture	Urine cytology	Bladder biopsy
	Symptom scores	Intravesical anesthetic bladder challenge	
	Cystoscopy	Hydrodistension	

Source: Adapted from Cox A et al: CUA guideline: Diagnosis and treatment of interstitial cystitis/bladder pain syndrome. Can Urol Assoc J 10:E136, 2016.

All aspects of the patient's pain should be explored, as many patients' pain will not be limited to the pelvis or bladder but will be associated with the genitals, anus or rectum, perineum, abdomen, and beyond. Furthermore, although pain is commonly experienced with bladder filling, patients may also have suprapubic tenderness or pressure with voiding or burning or pain in the bladder, urethra, or perineum, with radiation into the vagina for women or the prostate, penis, and testicles for men. In male patients, distinguishing IC/BPS from CP/CPPS can be challenging. Physicians must assess for more widespread pain locations outside the pelvis; screening for COPCs, particularly IBS, fibromyalgia, CFS, back pain, and headache, is important in adequately addressing the clinical impact of IC/BPS.

Eliciting and understanding associated LUTS—specifically urinary frequency, urgency, and nocturia—should be another focus of the history. While several confusable diseases can present with LUTS, the manifestation of IC/BPS as voiding dysfunction can help guide treatment decisions and is often a significant focus of bother for the patient. Having patients complete frequency/volume charts, noting the time and volume of each urination over a 24-h period, can help provide objective evidence of a LUTS history and facilitate follow-up during and after treatment.

Physical examination should focus on the abdomen, pelvis, genitals, and pelvic floor. The degree of pelvic floor relaxation (i.e., degree of muscle tension and/or spasm) during examination is important to note. Trigger points in the pelvic floor musculature and any areas of localized spasticity should be identified. In women, an examination of the vulva, vaginal mucosa, and urethral meatus is essential to identify the presence of vulvodynia (vulvar mucosal pain with no identifiable cause) or any signs of genitourinary syndrome of menopause. In men, an examination of the external genitalia and a digital rectal (prostate) examination as well as a similar pelvic floor examination should be included to rule out related pathology.

SYMPTOM SCORES

The quality of a history can be elevated by an accompanying validated, objective measurement of the patient's symptoms. Although several relevant tools exist, the Interstitial Cystitis Symptom Index (ICSI) and the Interstitial Cystitis Problem Index (ICPI) are the most widely used, are commonly employed in research trials as outcome measures, and are straightforward enough for the practitioner to perform in an outpatient setting. These short questionnaires document pain severity, urinary frequency, urgency, and nocturia as well as the bother experienced from each of these symptoms.

More recently, MAPP investigators have suggested that pain and urinary symptoms should be assessed independently using two separate questionnaires: the Genitourinary Pain Index (GUPI) to assess pain and the ICSI to separately assess urinary symptoms. This suggestion is based on their finding of variable effects of urologic pain versus urinary symptoms on quality of life and mental health. Although symptom scores should not be relied on as diagnostic tools, their utility in establishing objective baseline measures to monitor response to treatment and symptoms over time can be valuable to the patient and the practitioner.

URINE STUDIES

Urine studies (urinalysis, culture, sensitivity, and cytology) should be included in the workup of a patient in whom IC/BPS is suspected. However, their role is mostly in ruling out other confusable disease rather than in aiding in the diagnosis of IC/BPS. A microscopic examination of the urine can reveal abnormalities attributable to the kidney that may warrant referral to a nephrologist; microscopic hematuria may trigger cystoscopic examination and referral to a urologist. The presenting symptoms of IC/BPS often mimic those of UTI, which must be ruled out by urine cultures. It is important to recognize that IC/BPS patients are subject to at least as great a risk of UTIs as the general population and that UTI should be considered when a flare in symptoms is reported. Finally, urine cytology should be considered if a diagnosis of bladder cancer is suspected or if there is a history of hematuria.

IMAGING, CYSTOSCOPY, AND URODYNAMICS

More intensive investigations and imaging studies should be considered in specific scenarios—but other than cystoscopy, most should not be routinely performed. Abdominal and pelvic imaging studies in selected patients can help identify anatomic abnormalities of the upper or lower urinary tracts, diagnose urolithiasis or masses in the upper urinary tract, and rule out hydronephrosis, which may suggest obstructive uropathy. Furthermore, brain imaging with functional MRI and quantitative sensory testing may prove beneficial in establishing a central sensitization phenotype; however, this is an emerging field of investigation, and routine brain imaging and sensory testing are currently not recommended.

Cystoscopy is used to rule out bladder pathology—most importantly, bladder cancer. Moreover, cystoscopy plays an important role in phenotyping IC/BPS and is required for identification of Hunner lesions. The authors and others advocate for routine cystoscopic evaluation when IC/BPS is suspected, given the potential therapeutic implications and the ability of this measure to make phenotype-directed therapies possible. Finally, urodynamics testing should be reserved for specific scenarios—for example, cases in which complex voiding dysfunction may be contributing to the presentation.

INTRAVESICAL ANESTHETIC BLADDER CHALLENGE AND HYDRODISTENSION

An intravesical anesthetic bladder challenge (using intravesical lidocaine) can be done in the outpatient setting and can help distinguish bladder-focused pain from pelvic pain of other causes. It can further be harnessed as a therapeutic strategy if the patient experiences an improvement in symptoms. Similarly, hydrodistension, which requires a general or regional anesthetic, can play a diagnostic or therapeutic role. Bladder capacities of <400 mL under general anesthesia have correlated with worse pain and poor prognosis. The diagnostic role of post-hydrodistension inspection for bladder glomerulations has been suggested as a possible important clinical differentiation, although the utility of identifying and grading glomerulations is debated.

TREATMENT

CLINICAL PHENOTYPING

Clinically phenotyping patients has long been recognized as an important tool in the approach to UCPP patients—the UPOINT phenotyping tool was introduced in 2009. Given the heterogeneity of presentation, especially amongst patients with IC/BPS, approaching treatment with a patient-centric approach and a multimodal strategy is imperative for improving pain, urinary symptoms, and quality of life. The MAPP network has expanded on patient phenotyping through their deep phenotyping studies and has identified several important clinical phenotypes.

Pelvic Pain Only and Widespread Pain The most clear distinction in clinical phenotyping IC/BPS patients is between those with widespread pain and those with pelvic pain only. Patients with widespread pain can be a management challenge, as it is thought they likely have an element of central sensitization and often have poor response to local therapies. Trials evaluating treatment strategies specific to this phenotype—e.g., transcranial magnetic stimulation—are ongoing, but to date management relies primarily on systemic oral therapies. Although it may appear that pelvic pain–only patients have a host of therapies available to them, it is important to recognize that the MAPP network's phenotyping studies have identified a bladder-focused pain subset within this phenotype who appear to suffer worse pain, more urinary symptoms, and lower quality of life and may respond poorly to available therapies.

Hunner Lesions Another distinct phenotype to be considered is those patients with Hunner lesions. Some believe this subset of patients may represent a unique disease process and should be separated from IC/BPS patients. However, the science to support this distinction is not yet there and we continue to rely on Hunner lesion–specific treatment strategies for this phenotype.

Pelvic Floor Tenderness Patients with pelvic floor tenderness, pelvic floor muscle spasm, and/or trigger point pain on pelvic examination represent another distinct clinical phenotype. Pelvic floor physiotherapy among this phenotype of IC/BPS patients is supported with level 1 evidence and is an important part of management for these patients.

Infection Interstitial cystitis/bladder pain syndrome patients can also present with concomitant urinary tract infections. Some patients present with a history of a UTI as an index event, while others present with episodes of bona fide recurrent UTIs that trigger symptom flares despite their infection having been appropriately treated. It is important to approach the possibility of UTI among IC/BPS patients as one would with the general population, obtain urine culture, and treat the UTI accordingly. UTI vaccination has emerged as an important preventative therapy in patients with recurrent UTI and should be considered for recurrent UTI patients with IC/BPS.

Psychosocial It has long been recognized that patients with mental health diagnoses (e.g., depression, anxiety, posttraumatic stress disorder), psychosocial challenges, and maladaptive coping mechanisms (e.g., pain catastrophization) can have worse outcomes and suffer worse quality of life than those IC/BPS patients without. See **Fig. 54-1** and "Psychological Interventions," below, for treatment options for this phenotype.

Applying an individualized multimodal treatment approach has proved beneficial in clinical practice, and options for each phenotype should continue to expand as the importance of phenotype-directed clinical trials has been recognized. A collection of treatment options that might be directed at existing clinical phenotypes is presented in Fig. 54-1.

Although many of these treatments would be considered outside the scope of a general practitioner or general internist, it is important for the practitioner to be aware of them. In general, treatment should begin with more conservative measures, moving on to oral regimens or more invasive procedures if the patient's condition does not improve. A patient-centered approach is paramount in considering treatment escalation, and a multimodal strategy, implementing various treatments that address the individual patient's presenting clinical phenotype or "clinical picture," is essential.

CONSERVATIVE MEASURES

Conservative measures should be implemented for all patients with a diagnosis of IC/BPS. These therapies tend to be simple and inexpensive to introduce, pose little risk of significant side effects, and can be intensified or abandoned based on the patient's response.

Patient Education Patient education and empowerment are paramount in this chronic pain disorder. Patients have often seen multiple practitioners prior to their diagnosis of IC/BPS. Acknowledging their suffering while educating them about their disease can go a long way in terms of relieving stress and anxiety related to an unknown and poorly understood problem. This acknowledgment also helps to develop a therapeutic patient–provider relationship. Setting realistic expectations and understanding that cure is not the goal constitute an important first step. Several resources are available for patients to explore at their own leisure.

Dietary Modifications Although limited evidence supports the role of specific dietary modifications, it has long been recognized that certain foods can trigger flares in IC/BPS patients and that simple dietary modifications can result in meaningful improvements in symptoms. Common dietary triggers include acidic and spicy foods and/or drinks, caffeinated or alcoholic beverages, artificial sweeteners, and/or gluten products; this list is by no means exhaustive, and dietary modifications should be made on an individual level using an elimination diet approach.

Pelvic Floor Physiotherapy Involvement of the pelvic floor in the pain syndrome can be ascertained on physical examination. Randomized

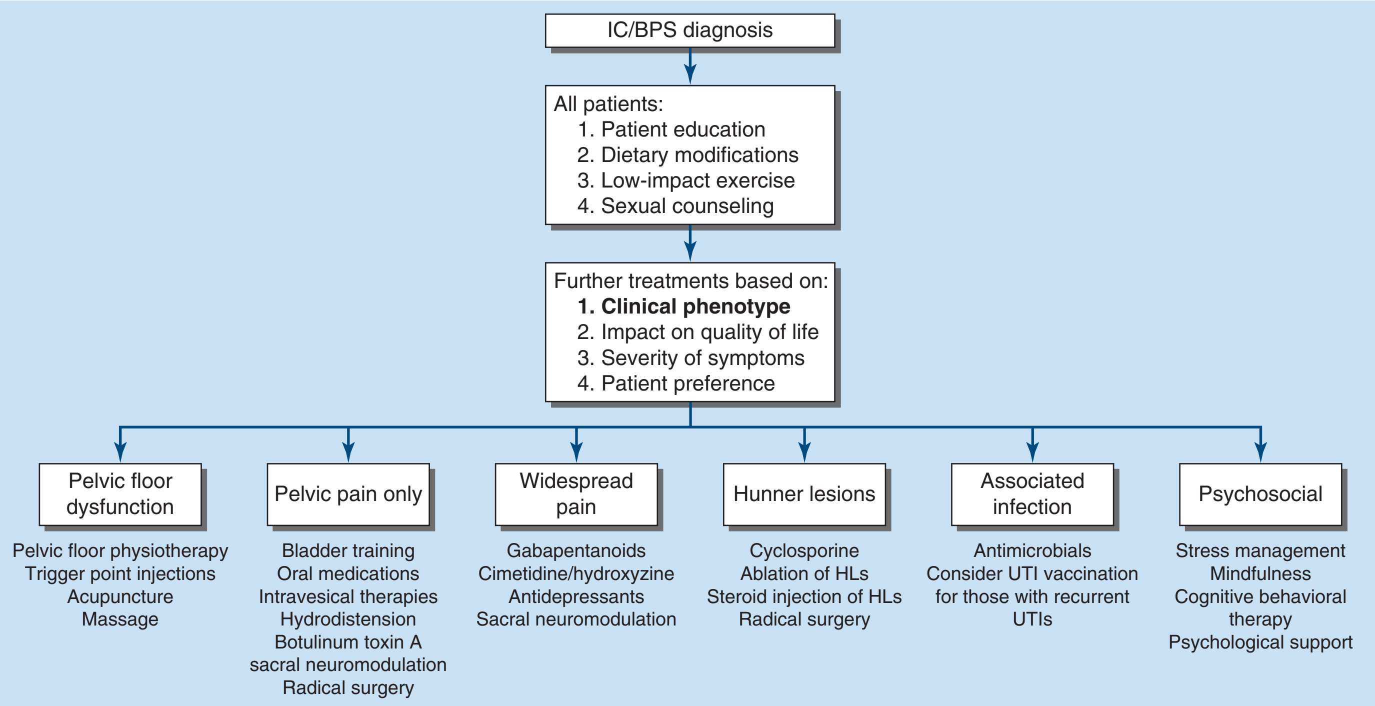
IC/BPS diagnosis
All patients:
1. Patient education
2. Dietary modifications
3. Low-impact exercise
4. Sexual counseling
Further treatments based on:
1. Clinical phenotype
2. Impact on quality of life
3. Severity of symptoms
4. Patient preference
Pelvic floor dysfunction
Pelvic floor physiotherapy
Trigger point injections
Acupuncture
Massage
Pelvic pain only
Bladder training
Oral medications
Intravesical therapies
Hydrodistension
Botulinum toxin A
sacral neuromodulation
Radical surgery
Widespread pain
Gabapentanoids
Cimetidine/hydroxyzine
Antidepressants
Sacral neuromodulation
Hunner lesions
Cyclosporine
Ablation of HLs
Steroid injection of HLs
Radical surgery
Associated infection
Antimicrobials
Consider UTI vaccination for those with recurrent UTIs
Psychosocial
Stress management
Mindfulness
Cognitive behavioral therapy
Psychological support

FIGURE 54-1 Proposed management paradigm for the treatment of interstitial cystitis (IC)/bladder pain syndrome (BPS). HLs, Hunner lesions.

studies have shown that, for patients who are found to have dysfunctional pelvic floors—muscle spasm, trigger points, or tenderness—contributing to their pain syndrome, pelvic floor physiotherapy may be beneficial. The musculoskeletal anatomy of the pelvic floor is complex; finding a provider with training specifically on the pelvic floor can be difficult but is crucial. Because accessing this resource may be financially burdensome for the patient, working together to find a way to obtain this helpful adjunctive therapy is important.

Psychological Interventions Mental health and psychosocial factors have long been identified as significantly prevalent in the IC/BPS population and can impact disease and quality-of-life outcomes. There is some indication that, in IC/BPS and other related chronic pain conditions, mindfulness and cognitive behavioral therapy may improve outcomes. Challenges in accessing these therapies are a major barrier, and there is a general lack of consensus on which specific interventions are best suited to individual patients.

MEDICAL THERAPIES

Only two medications are currently approved by the U.S. Food and Drug Administration (FDA) for the treatment of IC/BPS: pentosan polysulfate sodium (PPS) given orally and dimethyl sulfoxide (DMSO) given intravesically. However, a collection of medications, administered orally or intravesically, are commonly used (albeit off-label) for this purpose.

Oral Therapies • **PPS** The only FDA-approved oral medication for IC/BPS has recently come under scrutiny because of reports regarding its association with vision-threatening maculopathy. Although causation has yet to be established, given its marginal benefit in the treatment of IC/BPS, the authors recommend against the long-term use of this medication. For patients currently taking PPS, the risks and benefits of treatment must be weighed. Consideration of a trial of weaning off the medication may be in the best interest of the patient. Any patients experiencing vision-related complaints while taking PPS should undergo immediate ophthalmologic assessment.

Antibiotics IC/BPS is not an infectious condition, and thus antibiotics should have no role in treatment, unless the patient has developed a confirmed culture-positive urinary tract infection. Furthermore, the overwhelming majority of IC/BPS patients will have received at least one course, if not several courses, of antibiotics at some point in the course of their disease. Nevertheless, it is not unreasonable to administer a single course of antibiotics (after obtaining a sample for urine culture and sensitivity testing) if the patient has never previously received such therapy.

Amitriptyline Amitriptyline's pharmacologic activity is attributable primarily to its anticholinergic properties, its serotonin and norepinephrine uptake–inhibiting activity, and its sedative effects, which may include an antihistaminic pathway. Amitriptyline has been used to treat IC/BPS and other chronic pain syndromes. Studies support the use of amitriptyline in IC/BPS patients while recognizing that the benefits can be marginal and associated with significant side effects.

Cimetidine and Hydroxyzine Early nonrandomized studies of hydroxyzine in the treatment of IC/BPS yielded promising results. As with amitriptyline, hydroxyzine's mechanism of action in treating IC/BPS is not fully understood and is likely multifactorial, owing largely to its antihistaminic effect via H_1-receptor antagonism but perhaps also its anticholinergic properties, its anxiolytic and sedative effects, and its inhibition of mast cell secretion and activation. An underpowered randomized study found no significant difference in symptom improvement between hydroxyzine and placebo.

After the modest success reported for hydroxyzine, cimetidine—an H_2-receptor antagonist—was investigated as another possible IC/BPS treatment. Only two observational studies and a single, small, randomized clinical trial were completed and showed improvement in suprapubic pain and LUTS, particularly urinary frequency.

Gabapentinoids Although no randomized studies of gabapentinoids have been performed in the IC/BPS population, these agents have been shown to improve symptoms in related chronic pain conditions. Furthermore, observational studies have shown some efficacy in IC/BPS. In properly selected patients in whom neuropathic pain is suspected, this medication class may have some success.

Cyclosporine Despite its significant side effect profile, cyclosporine has been investigated as treatment for IC/BPS refractory to other, more standard therapies. Because of its potent anti-inflammatory properties (it is used extensively in organ transplant recipients), this drug is particularly effective in IC/BPS patients with Hunner lesions, although the improvement in symptoms is modest. Side effects, including hypertension and nephrotoxicity, must be carefully monitored for, and the medication is typically reserved for patients in whom standard therapies have failed.

Intravesical Therapies Intravesical instillations remain a mainstay in the management of bladder-specific pain. Although this treatment modality typically requires an office visit, able and motivated patients can be trained to administer the medication at home. Patient responses are variable, and the treatment is not curative, but it can significantly change the trajectory of disease in some patients and can rescue those experiencing symptom flares. Intravesical instillations can be administered as induction therapy; maintenance strategies have been proposed and can be effective for properly selected patients. A plethora of agents—most of them used off-label for this indication—have been investigated. The best-studied options are reviewed here.

DMSO DMSO, a solvent with anti-inflammatory properties, has been used in intravesical treatment for IC/BPS for several years. Despite a lack of high-quality evidence (with efficacy documented in only one placebo-controlled randomized clinical trial), DMSO remains the only FDA-approved intravesical medication for IC/BPS treatment. Its use has fallen out of favor, however, largely because of its unpleasant side effect of halitosis (its elimination via the lungs being associated with a garlic-like odor). Although the degree of improvement in symptoms is highly variable (60–95%), DMSO remains in the armamentarium of intravesical therapies.

Heparin Heparin, a glycosaminoglycan, was first investigated as a treatment for IC/BPS in light of the glycosaminoglycan layer deficiency theory of IC/BPS etiology; in animal models, heparin was shown to restore areas of damaged urothelium. Although there are no randomized clinical trials showing its efficacy, several observational studies have suggested benefit. Furthermore, in current practice, heparin is commonly administered with other medications as part of an intravesical "cocktail." Systemic absorption is minimal, and it appears not to affect coagulation parameters.

Lidocaine Lidocaine is commonly used as a local anesthetic and has been investigated as an option for intravesical treatment for IC/BPS. This agent works by blocking sensory nerves in the urothelium. Its absorption and efficacy increase by alkalinization, commonly through coadministration with sodium bicarbonate.

Trigger Point Injection Injection of a local anesthetic into myofascial trigger points in the pelvic floor (identified on physical examination) is a minimally invasive, practical therapy that can be administered during an office visit and can provide relief in properly selected patients. As with all therapies for this chronic pain condition, patient selection is paramount. The evidence supporting this treatment is largely anecdotal and based on expert opinion. A small, nonblind, observational trial found a 72% success rate among women diagnosed with chronic pelvic pain and trigger points on physical examination; 33% of women were completely pain free after the injection. Although robust prospective trials are needed, this modality adds to the clinician's armamentarium in the treatment of IC/BPS.

SURGICAL THERAPIES

Treatment of Hunner Lesions A unique subset of IC/BPS patients have Hunner lesions. Direct treatment of these lesions through ablation with cauterization or laser treatment or, alternatively, injection of the lesion with a glucocorticoid improves symptoms in 70–90% of these patients. Hunner lesions tend to be recurrent, however; thus, treated patients still require follow-up, with consideration of a multimodal approach to their disease.

Hydrodistension Recent reports confirm that one of the oldest therapies for IC/BPS, hydrodistension under general anesthesia, provides some benefit in up to 54% of patients. Short- and long-term adverse effects (including bladder perforation and long-term bladder wall fibrosis) and the temporary nature of the benefit (with symptoms typically recurring within 3–12 months) mean that repeated bladder distension may not be an ideal long-term management strategy.

Onabotulinum Toxin A Onabotulinum toxin A (i.e., Botox) has been used to treat IC/BPS. Several randomized studies have evaluated this agent's efficacy and have shown improvements in symptoms; unfortunately, these trials often use onabotulinum toxin A in combination with hydrodistension, lack a placebo arm, and do not control for LUTS. Attributing benefit to this treatment is thus challenging. Furthermore, the side effect of acute urinary retention can be catastrophic in IC/BPS patients, whose pain is often secondary to bladder filling. A multicenter, randomized, placebo-controlled trial is ongoing and should help clarify its efficacy in treatment of this pain disorder.

Sacral Neuromodulation In properly selected patients, sacral neuromodulation (SNM) can provide symptom relief in IC/BPS. Its use for this purpose is off-label, and it is not FDA approved for pain therapy. Nevertheless, SNM is FDA approved for treatment of bladder overactivity—a common symptom in IC/BPS patients—that is refractory to standard therapies. Although studies evaluating improvements in pain have shown variable results, a recent meta-analysis examining 17 observational trials (but no randomized clinical trials) of the use of SNM for IC/BPS does support its efficacy, with a statistically significant pooled treatment success rate of up to 84%. Side effects related to SNM must be considered, and the procedure carries with it a high rate of revision surgery, which is needed in up to half of patients undergoing this treatment.

Radical Surgery Radical surgery for the treatment of IC/BPS is reserved as a modality of desperation for the most refractory patients. Options range from substitution cystoplasty to cystectomy with urinary diversion. Although improved symptoms and quality of life may result, such surgery is a potentially morbid operation and patient selection must be very specific.

COMPLICATIONS AND PROGNOSIS

IC/BPS is not unlike other chronic pain conditions in that, although a clear link has not been established with higher mortality, this condition is certainly associated with significant disability, decreased quality of life, and significant mental health morbidity. The economic impact of the disability associated with IC/BPS is similar to that of fibromyalgia, low back pain, rheumatoid arthritis, and peripheral neuropathy. Suicidal ideation is a reality in this patient population, with a reported prevalence as high as 11–23%.

For most patients, IC/BPS onset is subacute, with continuous development of the classic symptom complex over a short time and rapid (within 5 years) progression to its final stage. Symptoms then continue to wax and wane without significant overall change in symptomatology for the majority of patients. However, spontaneous improvement and/or resolution occurs in some patients, while a small subset experience subsequent deterioration to a small-capacity, fibrotic, noncompliant bladder ("end-stage bladder"). A multimodal approach to therapy, interdisciplinary involvement in patient care, particular attention to psychosocial parameters, and check-ins on mental health are important aspects of ongoing care.

GLOBAL CONSIDERATIONS

Significant challenges have been encountered in confirming the prevalence of IC/BPS, particularly globally. Prevalence estimates have ranged widely from as low as 3.5 per 100,000 women in a study of a Japanese population to as high as 20,000 per 100,000 in a self-report questionnaire study of a U.S. population. Despite these challenges, it has been recognized that IC/BPS is not simply a disease of the global West. Although robust epidemiologic studies outside North America, Europe, and some regions of Asia are lacking, it is presumed that this disease occurs at similar rates globally. This presumption may be extrapolated from epidemiologic studies of a related population and its male counterpart: CP/CPPS. These studies have shown rates of CP/CPPS in African and Asian populations that are similar to rates in North American populations.

There is no evidence to suggest that IC/BPS is phenotypically distinct in various geographic regions. Thus, this condition should be diagnosed and treated in the same ways globally. Given that its diagnosis of exclusion is based largely on history and physical examination and its treatment is based on a minimally invasive algorithm, with the focus on the patient's clinical phenotype and the initial implementation of conservative therapeutic measures, IC/BPS can be well managed even in resource-poor settings. As with many poorly understood and difficult-to-treat conditions, the greatest barrier to its diagnosis and treatment may perhaps be its recognition.

FURTHER READING

Clemens JQ et al: Urologic chronic pelvic pain syndrome: Insights from the MAPP Research Network. Nat Rev Urol 16:187, 2019.

Clemens JQ et al: AUA guideline for the diagnosis and treatment of interstitial cystitis/bladder pain syndrome. J Urol 208:34, 2022.

Cox A et al: CUA guideline: Diagnosis and treatment of interstitial cystitis/bladder pain syndrome. Can Urol Assoc J 10:E136, 2016.

Moldwin RM et al: Interstitial cystitis/bladder pain syndrome and related disorders, in *Campbell-Walsh-Wein Urology*, 12th ed. AW Partin et al (eds). Philadelphia, Elsevier, 2021.

Nickel JC et al: MV140 sublingual vaccine reduces recurrent urinary tract infection in women. Results from the first North American clinical experience study. Can Urol Assoc J 18:25, 2024.

55

Azotemia and Urinary Abnormalities

David B. Mount

Normal kidney functions occur through numerous cellular processes to maintain body homeostasis. Disturbances in any of these functions can lead to abnormalities that may be detrimental to survival. Clinical manifestations of these disorders depend on the pathophysiology of renal injury and often are identified as a complex of symptoms, abnormal physical findings, and laboratory changes that constitute specific syndromes. These renal syndromes **(Table 55-1)** may arise from systemic illness or as primary renal disease. Nephrologic syndromes usually consist of several elements that reflect the underlying pathologic processes, typically including one or more of the following: (1) reduction in glomerular filtration rate (GFR), (2) abnormalities of urine sediment (red blood cells [RBCs], white blood cells [WBCs], casts, and crystals), (3) abnormal urinary excretion of serum proteins (proteinuria), (4) disturbances in urine volume (oliguria, anuria, polyuria), (5) presence of hypertension and/or expanded total body fluid volume (edema), (6) electrolyte abnormalities, and (7) in some syndromes, fever/pain. The specific combination of these findings should permit identification of one of the major nephrologic syndromes (Table 55-1) and allow differential diagnoses to be narrowed so that the appropriate diagnostic and therapeutic course can be determined. All these syndromes and their associated diseases are discussed in more detail in subsequent chapters. This chapter focuses on several aspects of renal abnormalities that are critically important for distinguishing among those processes: (1) reduction in GFR, (2) alterations of the urinary sediment and/or protein excretion, and (3) abnormalities of urinary volume.

AZOTEMIA

ASSESSMENT OF GFR

Monitoring the GFR is important in both hospital and outpatient settings, and several different methodologies are available. GFR is the primary metric for kidney "function," and its direct measurement involves administration of a radioactive isotope (such as inulin or iothalamate) that is filtered at the glomerulus into the urinary space but is neither reabsorbed nor secreted throughout the tubule. GFR—i.e., the clearance of inulin or iothalamate in milliliters per minute—is calculated from the rate of appearance of the isotope in the urine over several hours. In most clinical circumstances, direct GFR measurement is not feasible, and the plasma creatinine level is used as a surrogate to estimate GFR. Plasma creatinine (P_{Cr}) is the most widely used marker for GFR, which is related directly to urine creatinine (U_{Cr}) excretion and inversely to P_{Cr}. On the basis of this relationship (with some important caveats, as discussed below), GFR will fall in roughly inverse proportion to the rise in P_{Cr}. Failure to account for GFR reductions in drug dosing can lead to significant morbidity and death from drug toxicities (e.g., digoxin, imipenem). In the outpatient setting, P_{Cr} serves as an estimate for GFR (although much less accurate; see below). In patients with chronic progressive renal disease, there is an approximately linear relationship between $1/P_{Cr}$ (y axis) and time (x axis). The slope of that line will remain constant for an individual; when values deviate, an investigation for a superimposed acute process (e.g., volume depletion, drug reaction) should be initiated. Signs and symptoms of uremia, the clinical symptom complex associated with renal failure, develop at significantly different levels of P_{Cr}, depending on the patient (size, age, and sex), underlying renal disease, existence of concurrent diseases, and true GFR. Generally, patients do not develop symptomatic uremia until renal insufficiency is severe (GFR <15 mL/min).

A significantly reduced GFR (either acute or chronic) is usually reflected in a rise in P_{Cr}, leading to retention of nitrogenous waste products (defined as azotemia) such as urea. Azotemia may result from reduced renal perfusion, intrinsic renal disease, or postrenal processes (ureteral obstruction; see below and **Fig. 55-1**). Precise determination of GFR is problematic, as both commonly measured indices (urea and creatinine) have characteristics that affect their accuracy as markers of clearance. Urea clearance may underestimate GFR significantly because of urea reabsorption by the tubule. In contrast, creatinine is derived from muscle metabolism of creatine, and its generation varies little from day to day.

Creatinine clearance (CrCl), an approximation of GFR, is measured from plasma and urinary creatinine excretion rates for a defined period (usually 24 h) and is expressed in milliliters per minute: $CrCl = (U_{vol} \times U_{Cr})/(P_{Cr} \times T_{min})$. The "adequacy" or "completeness" of the urinary collection is estimated by the urinary volume and creatinine content; creatinine is produced from muscle and excreted at a relatively constant rate. For a 20- to 50-year-old man, creatinine excretion should be 18.5–25.0 mg/kg body weight; for a woman of the same age, it should be 16.5–22.4 mg/kg body weight. For example, an 80-kg man should excrete between ~1500 and 2000 mg of creatinine in an "adequate" collection. Creatinine is useful for estimating GFR because it is a small, freely filtered solute that is not reabsorbed by the tubules. P_{Cr} levels can increase acutely from dietary ingestion of cooked meat, however, and creatinine can be secreted into the proximal tubule through an organic cation pathway (especially in advanced progressive chronic kidney disease [CKD]), leading to overestimation of GFR. When a timed collection for CrCl is not available, decisions about drug dosing must be based on P_{Cr} alone. Two formulas are used widely to estimate kidney function from P_{Cr}: (1) Cockcroft-Gault and (2) four-variable MDRD (Modification of Diet in Renal Disease).

Cockcroft-Gault:

$$\textbf{CrCl(mL/min)} = \frac{(140-\text{age})\times\text{Lean Body Weight (kg)}}{\text{Serum Creatinine (mg/dL)}\times 72}(\times 0.85\text{ if female})$$

$$\text{MDRD: eGFR (mL/min per 1.73 m}^2) = 186.3 \times P_{Cr}\ (e^{-1.154}) \times \text{age}\ (e^{-0.203}) \times (0.742\text{ if female}) \times (1.21\text{ if black}).$$

Numerous websites are available to assist with these calculations (*www.kidney.org/professionals/kdoqi/gfr_calculator.cfm*). A newer Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) estimated GFR (eGFR), which was developed by pooling several cohorts

TABLE 55-1 Initial Clinical and Laboratory Database for Defining Major Syndromes in Nephrology

SYNDROME	IMPORTANT CLUES TO DIAGNOSIS	COMMON FINDINGS	CHAP(S). DISCUSSING DISEASE-CAUSING SYNDROME
Acute or rapidly progressive renal failure	Anuria	Hypertension, hematuria	321, 326, 328, 331
	Oliguria	Proteinuria, pyuria	
	Documented recent decline in GFR	Casts, edema	
Acute nephritis	Hematuria, RBC casts	Proteinuria	326
	Azotemia, reduced GFR, oliguria	Pyuria	
	Edema, hypertension	Circulatory congestion	
Chronic renal failure	Azotemia for >3 months	Proteinuria, casts	322
	Symptoms or signs of uremia, (late manifestation), casts	Hypocalcemia, hyperphosphatemia, hyperparathyroidism	
	Symptoms or signs of renal osteodystrophy	Polyuria, nocturia	
	Kidneys reduced in size bilaterally	Edema, hypertension	
	Broad casts in urinary sediment	Hyperkalemia, metabolic acidosis	
Nephrotic syndrome	Proteinuria, with >3.5 g/24 h per 1.73 m^2	Casts	326
	Hypoalbuminemia	Lipiduria	
	Edema	Hypercoagulable state	
	Hyperlipidemia		
Asymptomatic urinary abnormalities	Hematuria		326
	Proteinuria (below nephrotic range)		
	Sterile pyuria, casts		
Urinary tract infection/pyelonephritis	Bacteriuria, with >10^5 cfu/mL	Hematuria	140
	Other infectious agent documented in urine	Mild azotemia and reduced GFR	
	Pyuria, leukocyte casts	Mild proteinuria	
	Frequency, urgency	Fever	
	Bladder tenderness, flank tenderness		
Renal tubular defects	Electrolyte disorders	Hematuria	327, 328
	Polyuria, nocturia	"Tubular" proteinuria (<1 g/24 h)	
	Renal calcification	Enuresis	
	Large kidneys	Electrolyte and/or acid-base abnormalities	
	Renal transport defects	Other electrolyte issues, e.g., hypomagnesemia	
Hypertension	Systolic/diastolic hypertension	Proteinuria	288, 329
		Casts	
		Azotemia	
Nephrolithiasis	Previous history of stone passage or removal	Hematuria	330
	Previous history of stone seen by x-ray	Pyuria	
	Renal colic	Frequency, urgency	
Urinary tract obstruction	Azotemia, oliguria, anuria	Hematuria	331
	Polyuria, nocturia, urinary retention	Pyuria	
	Slowing of urinary stream	Enuresis, dysuria	
	Large prostate, large kidneys		
	Flank tenderness, full bladder after voiding		

Abbreviations: cfu, colony-forming units; GFR; glomerular filtration rate; RBC, red blood cell.

with and without kidney disease who had data on directly measured GFR, appears to be more accurate:

$$\text{CKD-EPI: eGFR} = 141 \times \min(P_{Cr}/k, 1)^a \times \max(P_{Cr}/k, 1)^{-1.209} \times 0.993^{\text{Age}} \times 1.018\ [\text{if female}] \times 1.159\ [\text{if black}],$$

where P_{Cr} is plasma creatinine, k is 0.7 for females and 0.9 for males, a is –0.329 for females and –0.411 for males, *min* indicates the minimum of P_{Cr}/k or 1, and *max* indicates the maximum of P_{Cr}/k or 1 (*https://www.mdcalc.com/ckd-epi-equations-glomerular-filtration-rate-gfr*).

There are limitations to all creatinine-based estimates of GFR. Each equation, along with 24-h urine collection for measurement of creatinine clearance, is based on the assumption that the patient is in *steady state*, without daily increases or decreases in P_{Cr} as a result of rapidly changing GFR. The MDRD equation is better correlated with true GFR when the GFR is <60 mL/min per 1.73 m^2. The gradual loss of muscle from chronic illness, chronic use of glucocorticoids, or malnutrition can mask significant changes in GFR with small or imperceptible changes in P_{Cr}.

The coefficient of 1.159 in the CKD-EPI equation to adjust for self-reported black race reflects that measured GFR was 16% higher in blacks than nonblacks with similar age, sex, and creatinine in the data set used to develop the equation. Race is a social rather than a biological construct, for which reason the use of the "race modifier" in calculating eGFR using CKD-EPI and other equations has come under scrutiny. In particular, given the implications of utilizing self-reported race to modify clinical laboratory results, many medical centers have recently stopped reporting eGFRs that have been calculated using a race modifier. This change is projected to have positive consequences, in particular, improved access to waitlisting for renal transplantation in black patients at an earlier stage of CKD. Potential

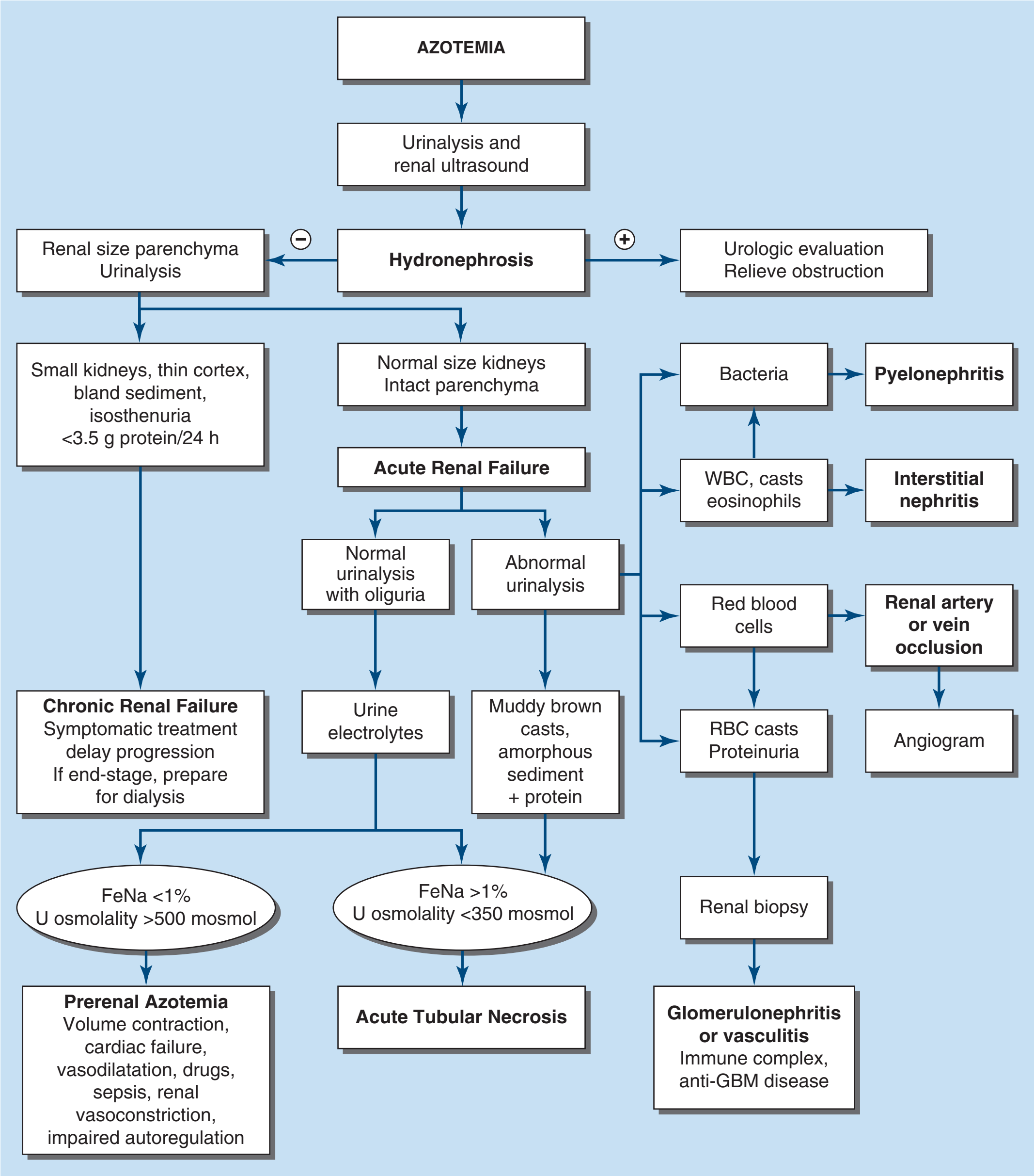
AZOTEMIA
Urinalysis and renal ultrasound
Hydronephrosis
⊖
Renal size parenchyma Urinalysis
⊕
Urologic evaluation Relieve obstruction
Small kidneys, thin cortex, bland sediment, isosthenuria <3.5 g protein/24 h
Normal size kidneys Intact parenchyma
Acute Renal Failure
Normal urinalysis with oliguria
Abnormal urinalysis
Bacteria
Pyelonephritis
WBC, casts eosinophils
Interstitial nephritis
Red blood cells
Renal artery or vein occlusion
RBC casts Proteinuria
Angiogram
Chronic Renal Failure
Symptomatic treatment delay progression If end-stage, prepare for dialysis
Urine electrolytes
Muddy brown casts, amorphous sediment + protein
FeNa <1% U osmolality >500 mosmol
FeNa >1% U osmolality <350 mosmol
Renal biopsy
Prerenal Azotemia
Volume contraction, cardiac failure, vasodilatation, drugs, sepsis, renal vasoconstriction, impaired autoregulation
Acute Tubular Necrosis
Glomerulonephritis or vasculitis
Immune complex, anti-GBM disease

FIGURE 55-1 Approach to the patient with azotemia. FeNa, fractional excretion of sodium; GBM, glomerular basement membrane; RBC, red blood cell; U, urine; WBC, white blood cell.

negative consequences include "overdiagnosis" of CKD, inadequate or inaccurate dosing of drugs that are eliminated through the kidney (e.g., metformin), reduced access to imaging modalities for black patients with CKD with a lower reported eGFR, and reductions in living kidney donation among blacks. These and the other limitations in creatinine-based eGFR have led to the development of alternative methods for estimating GFR.

Cystatin C, a member of the cystatin superfamily of cysteine protease inhibitors, is produced at a relatively constant rate from all nucleated cells. Serum cystatin C has been proposed to be a more sensitive marker of early GFR decline than is P_{Cr}, with lesser effects of muscle mass on circulating levels; however, cystatin C levels are influenced by the patient's sex and the presence of diabetes mellitus, smoking, and inflammation. To the extent that cystatin C–based calculation of eGFR is less affected by self-reported race and muscle mass, it is an increasingly important adjunct to creatinine-based eGFR. Recently, eGFR equations that include *both* creatinine and cystatin C have been shown to be more accurate than the single-measurement equations. Clinical judgement and clinical assessment also play an important role in interpreting eGFR values. For example, a bodybuilder may have an elevated creatinine level due to increased muscle mass, with an underestimate of GFR based on a creatinine-based eGFR; in that case, the cystatin C eGFR may be more accurate.

APPROACH TO THE PATIENT

Azotemia

Once GFR reduction has been established, the physician must decide if it represents acute or chronic renal injury. The clinical circumstances, history, and laboratory data often make this an easy distinction. However, the laboratory abnormalities characteristic of chronic renal failure, including anemia, hypocalcemia, and hyperphosphatemia, are also often present in patients presenting with acute renal failure. Radiographic evidence of renal osteodystrophy **(Chap. 322)** can be seen only in chronic renal failure but is a very late finding, typically in patients with end-stage renal disease (ESRD) maintained on dialysis. The urinalysis and renal ultrasound can facilitate distinguishing acute from chronic renal failure. An approach to the evaluation of azotemic patients is shown in Fig. 55-1. Patients with advanced chronic renal insufficiency often have some proteinuria, nonconcentrated urine (isosthenuria; isosmotic with plasma), and small kidneys on ultrasound, characterized by increased echogenicity and cortical thinning. Treatment should be directed toward slowing the progression of renal disease and providing symptomatic relief for edema, acidosis, anemia, and hyperphosphatemia, as discussed in **Chap. 322**. Acute renal failure **(Chap. 321)** can result from processes that affect blood flow and glomerular perfusion (prerenal azotemia), intrinsic renal diseases (affecting small vessels, glomeruli, or tubules), or postrenal processes (obstruction of urine flow in ureters, bladder, or urethra) **(Chap. 331)**.

PRERENAL FAILURE

Decreased renal perfusion accounts for 40–80% of cases of acute renal failure and, if appropriately treated, is readily reversible. The etiologies of prerenal azotemia include any cause of decreased circulating blood volume (gastrointestinal hemorrhage, burns, diarrhea, diuretics), volume sequestration (pancreatitis, peritonitis, rhabdomyolysis), or decreased effective arterial volume (cardiogenic shock, sepsis). Renal and glomerular perfusion also can be affected by reductions in cardiac output from peripheral vasodilation (sepsis, drugs) or profound renal vasoconstriction (severe heart failure, hepatorenal syndrome, agents such as nonsteroidal anti-inflammatory drugs [NSAIDs]). True or "effective" arterial hypovolemia leads to a fall in mean arterial pressure, which in turn triggers a series of neural and humoral responses, including activation of the sympathetic nervous and renin-angiotensin-aldosterone systems and vasopressin (AVP) release. GFR is maintained by prostaglandin-mediated dilatation of afferent arterioles and angiotensin II–mediated constriction of efferent arterioles. Once the mean arterial pressure falls below 80 mmHg, GFR declines steeply.

Blockade of prostaglandin production by NSAIDs can result in severe vasoconstriction and acute renal failure. Blocking angiotensin action with angiotensin-converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARBs) decreases efferent arteriolar tone and in turn decreases glomerular capillary perfusion pressure. Patients taking NSAIDs and/or ACE inhibitors/ARBs are most susceptible to hemodynamically mediated acute renal failure when blood volume or arterial perfusion pressure is reduced for any reason; under these circumstances, preservation of GFR is dependent on afferent vasodilation due to prostaglandins and efferent vasoconstriction due to angiotensin II. Patients with bilateral renal artery stenosis (or stenosis in a solitary kidney) can also be dependent on efferent arteriolar vasoconstriction for maintenance of glomerular filtration pressure and are particularly susceptible to a precipitous decline in GFR when given ACE inhibitors or ARBs.

Prolonged renal hypoperfusion may lead to acute tubular necrosis (ATN), an intrinsic renal disease that is discussed below. The urinalysis and urinary electrolyte measurements can be useful in distinguishing prerenal azotemia from ATN **(Table 55-2)**. The urine Na and osmolality of patients with prerenal azotemia can be predicted from the stimulatory actions of norepinephrine, angiotensin II, AVP, aldosterone, and low tubule fluid flow rate. In prerenal conditions, the tubules are intact, leading to a concentrated urine (>500 mosmol), avid Na retention (urine Na concentration, <20 mmol/L; fractional excretion of Na [FE_{Na}], <1%), and U_{Cr}/P_{Cr} >40 (Table 55-2). The FE_{Na} is typically >1% in ATN but may be <1% in patients with milder, nonoliguric ATN (e.g., from rhabdomyolysis) and in patients with underlying "prerenal" disorders, such as congestive heart failure (CHF) or cirrhosis or hepatorenal syndrome. The prerenal urine sediment is usually normal or has hyaline and granular casts, whereas the sediment of ATN usually is filled with cellular debris, tubular epithelial casts, and dark (muddy brown) granular casts. Microscopic examination of a urine sediment is a key test in AKI, since the presence of dark granular casts and/or tubular epithelial cells in the urine is highly predictive of ATN. The measurement of urinary biomarkers associated with tubular injury is a promising technique to detect subclinical ATN and/or help further diagnose the exact cause of acute renal failure.

TABLE 55-2 Laboratory Findings in Acute Renal Failure

INDEX	PRERENAL AZOTEMIA	OLIGURIC ACUTE RENAL FAILURE
BUN/P_{Cr} ratio	>20:1	10–15:1
Urine sodium U_{Na}, meq/L	<20	>40
Urine osmolality, mosmol/L H_2O	>500	<350
Fractional excretion of sodium[a]	<1%	>2%
Urine/plasma creatinine U_{Cr}/P_{Cr}	>40	<20
Urinalysis (casts)	None or hyaline/granular	Muddy brown

$$^{a}FE_{Na} = \frac{U_{Na} \times P_{Cr} \times 100}{P_{Na} \times U_{cr}}$$

Abbreviations: BUN, blood urea nitrogen; P_{Cr}, plasma creatinine concentration; P_{Na}, plasma sodium concentration; U_{Cr}, urine creatinine concentration; U_{Na}, urine sodium concentration.

POSTRENAL AZOTEMIA

Urinary tract obstruction accounts for <5% of cases of acute renal failure but is usually reversible and must be ruled out early in the evaluation (Fig. 55-1). Since a single kidney is capable of adequate clearance, complete obstructive acute renal failure requires obstruction at the urethra or bladder outlet, bilateral ureteral obstruction, or unilateral obstruction in a patient with a single functioning kidney. Obstruction is usually diagnosed by the presence of ureteral and renal pelvic dilation on renal ultrasound. However, early in the course of obstruction or if the ureters are unable to dilate (e.g., encasement by pelvic or periureteral tumors or by retroperitoneal fibrosis), the ultrasound examination may be negative. Other imaging, such as a furosemide renogram (MAG3 nuclear medicine study), may be required to better define the presence or absence of obstructive uropathy. The specific urologic conditions that cause obstruction are discussed in **Chap. 331**.

INTRINSIC RENAL DISEASE

When prerenal and postrenal azotemia have been excluded as etiologies of renal failure, an intrinsic parenchymal renal disease is present. Intrinsic renal disease can arise from processes involving large renal vessels, intrarenal microvasculature and glomeruli, or the tubulointerstitium. Ischemic and toxic ATN account for ~90% of cases of acute intrinsic renal failure. As outlined in Fig. 55-1, the clinical setting and urinalysis are helpful in separating the possible etiologies. Prerenal azotemia and ATN are part of a spectrum of renal hypoperfusion; evidence of structural tubule injury is present in ATN, whereas prompt reversibility occurs with prerenal azotemia upon restoration of adequate renal perfusion. Thus, ATN often can be distinguished from prerenal azotemia by urinalysis and urine electrolyte composition (Table 55-2 and Fig. 55-1). Ischemic ATN is observed most frequently in patients who have undergone major surgery, trauma, severe hypovolemia, overwhelming sepsis, or extensive burns. Nephrotoxic ATN complicates the administration of many common medications, usually by inducing a combination of intrarenal vasoconstriction, direct tubule toxicity, and/or tubular obstruction. The kidney is vulnerable to toxic injury by virtue of its rich blood supply (25% of cardiac output) and its ability to concentrate and metabolize toxins. A diligent search for hypotension and nephrotoxins usually uncovers the specific etiology of ATN. Discontinuation of nephrotoxins and stabilization of blood pressure often suffice without the need for dialysis, with ongoing regeneration of tubular cells. **An extensive list of potential drugs and toxins implicated in ATN is found in Chap. 321.**

Processes involving the tubules and interstitium can lead to acute kidney injury (AKI), a subtype of acute renal failure. These processes include drug-induced interstitial nephritis (especially by antibiotics, NSAIDs, and proton pump inhibitors), severe infections (both bacterial and viral), systemic diseases (e.g., systemic lupus erythematosus), and systemic disorders (e.g., sarcoidosis, Sjögren's syndrome, lymphoma, or leukemia). A list of drugs associated with allergic interstitial nephritis is found in **Chap. 328**. Urinalysis usually shows mild to moderate proteinuria, hematuria, and pyuria (~75% of cases) and occasionally WBC casts. The finding of RBC casts in interstitial nephritis has been reported but should prompt a search for glomerular diseases (Fig. 55-1). Renal biopsy will be needed to distinguish among these possibilities. The classic sediment finding in allergic interstitial nephritis is a predominance (>10%) of urinary eosinophils with Wright's or Hansel's stain; however, urinary eosinophils can be increased in several other causes of AKI, such that measurement of urine eosinophils has no diagnostic utility in renal disease. This test is no longer recommended in the workup of AKI.

Occlusion of large renal vessels, including arteries and veins, is an uncommon cause of acute renal failure. A significant reduction in GFR by this mechanism suggests bilateral processes or, in a patient with a single functioning kidney, a unilateral process. In patients with preexisting renal artery stenosis, a substantial renal collateral circulation can develop over time and sustain renal perfusion—typically not enough to sustain glomerular filtration, but enough to maintain tissue viability—in the event of total renal artery occlusion. Renal arteries can be occluded with atheroemboli, thromboemboli, in situ thrombosis, aortic dissection, or vasculitis. Atheroembolic renal failure can occur spontaneously but most often is associated with recent aortic instrumentation. The emboli are cholesterol-rich and lodge in medium and small renal arteries, with a consequent eosinophil-rich inflammatory reaction. Patients with atheroembolic acute renal failure often have a normal urinalysis, but the urine may contain eosinophils and casts. The diagnosis can be confirmed by renal biopsy, but this procedure is often unnecessary when other stigmata of atheroemboli are present (livedo reticularis, distal peripheral infarcts, eosinophilia). Renal artery thrombosis may lead to mild proteinuria and hematuria, whereas renal vein thrombosis typically occurs in the context of heavy proteinuria and hematuria. These vascular complications often require angiography for confirmation and are discussed in **Chap. 329**.

Diseases of the glomeruli (glomerulonephritis and vasculitis) and the renal microvasculature (hemolytic-uremic syndromes, thrombotic thrombocytopenic purpura, and malignant hypertension) usually present with various combinations of glomerular injury: proteinuria, hematuria, reduced GFR, and alterations of sodium excretion that lead to hypertension, edema, and circulatory congestion (acute nephritic syndrome). These findings may occur as primary renal diseases or as renal manifestations of systemic diseases. The clinical setting and other laboratory data help distinguish primary renal diseases from systemic diseases. The finding of RBC casts in the urine is an indication for early renal biopsy (Fig. 55-1), as the pathologic pattern has important implications for diagnosis, prognosis, and treatment. Hematuria without RBC casts can also be an indication of glomerular disease, since RBC casts are highly specific but very insensitive for glomerulonephritis. The specificity of urine microscopy can be enhanced by examining urine with a phase contrast microscope capable of detecting dysmorphic red cells ("acanthocytes") that are associated with glomerular disease. This evaluation is summarized in **Fig. 55-2**. A detailed discussion of glomerulonephritis and diseases of the microvasculature is found in **Chap. 328**.

OLIGURIA AND ANURIA

Oliguria refers to a 24-h urine output <400 mL, and *anuria* is the complete absence of urine formation (<100 mL). Anuria can be caused by complete bilateral urinary tract obstruction; a vascular catastrophe (dissection or arterial occlusion); renal vein thrombosis; acute cast nephropathy in myeloma; renal cortical necrosis; severe ATN; severe rapidly progressive glomerulonephritis; combined therapy with NSAIDs, ACE inhibitors, and/or ARBs; and hypovolemic, cardiogenic, or septic shock. Oliguria is never normal, since at least 400 mL of maximally concentrated urine must be produced to excrete the obligate daily osmolar load. *Nonoliguria* refers to urine output >400 mL/d in patients with acute or chronic azotemia. With nonoliguric ATN, disturbances of potassium and hydrogen balance are less severe than in oliguric patients, and recovery to normal renal function is usually more rapid.

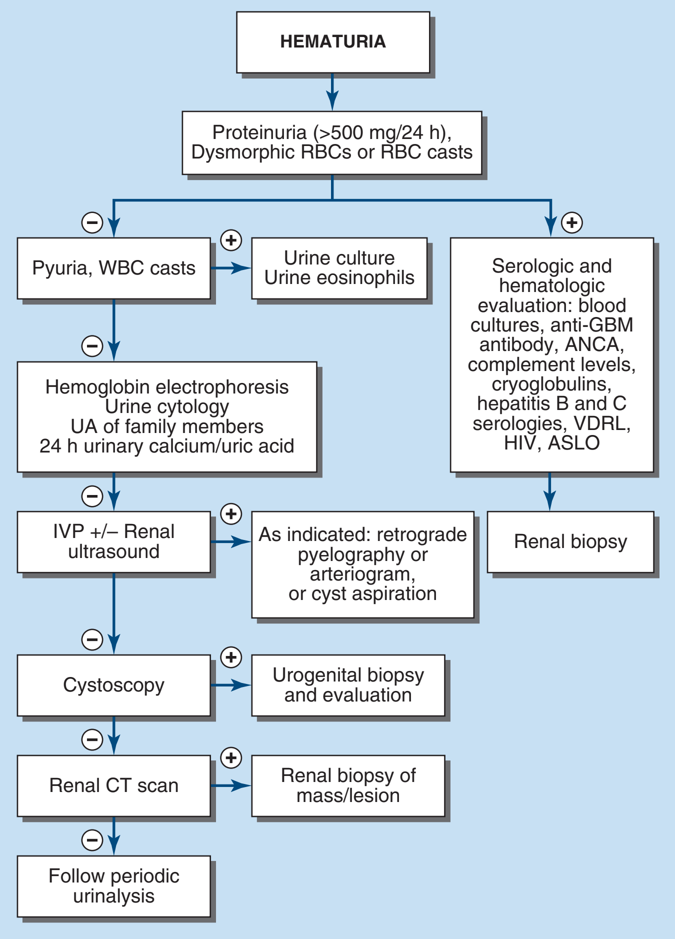
HEMATURIA
Proteinuria (>500 mg/24 h), Dysmorphic RBCs or RBC casts
−
+
Pyuria, WBC casts
+
Urine culture Urine eosinophils
Serologic and hematologic evaluation: blood cultures, anti-GBM antibody, ANCA, complement levels, cryoglobulins, hepatitis B and C serologies, VDRL, HIV, ASLO
−
Hemoglobin electrophoresis Urine cytology UA of family members 24 h urinary calcium/uric acid
−
IVP +/– Renal ultrasound
+
As indicated: retrograde pyelography or arteriogram, or cyst aspiration
Renal biopsy
−
Cystoscopy
+
Urogenital biopsy and evaluation
−
Renal CT scan
+
Renal biopsy of mass/lesion
−
Follow periodic urinalysis

FIGURE 55-2 Approach to the patient with hematuria. ANCA, antineutrophil cytoplasmic antibody; ASLO, antistreptolysin O; CT, computed tomography; GBM, glomerular basement membrane; IVP, intravenous pyelography; RBC, red blood cell; UA, urinalysis; VDRL, Venereal Disease Research Laboratory; WBC, white blood cell.

ABNORMALITIES OF THE URINE

PROTEINURIA

The evaluation of proteinuria is shown schematically in **Fig. 55-3** and typically is initiated after detection of proteinuria by dipstick examination. The dipstick measurement detects only albumin and gives false-positive results at pH >7.0 or when the urine is very concentrated or contaminated with blood. Because the dipstick relies on urinary albumin concentration, a very dilute urine may obscure significant proteinuria on dipstick examination. Quantification of urinary albumin on a spot urine sample (ideally from a first morning void) by measurement of an albumin-to-creatinine ratio (ACR) is helpful in approximating a 24-h albumin excretion rate (AER), where ACR (mg/g) ≈ AER (mg/24 h). Furthermore, proteinuria that is not predominantly due to albumin will be missed by dipstick screening. This information is particularly important for the detection of Bence-Jones proteins in the urine of patients with multiple myeloma. Tests to measure total urine protein concentration accurately rely on precipitation with sulfosalicylic or trichloroacetic acid (Fig. 55-3). As with albuminuria, the ratio of

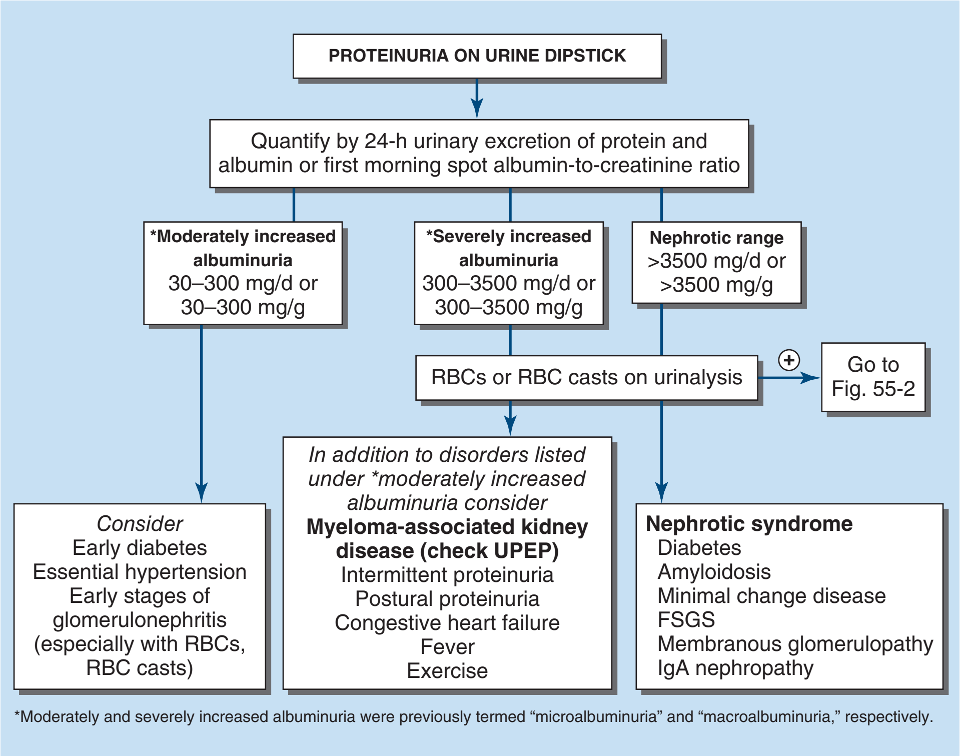
PROTEINURIA ON URINE DIPSTICK
Quantify by 24-h urinary excretion of protein and albumin or first morning spot albumin-to-creatinine ratio
*Moderately increased albuminuria 30–300 mg/d or 30–300 mg/g
*Severely increased albuminuria 300–3500 mg/d or 300–3500 mg/g
Nephrotic range >3500 mg/d or >3500 mg/g
RBCs or RBC casts on urinalysis
+
Go to Fig. 55-2
Consider
Early diabetes
Essential hypertension
Early stages of glomerulonephritis (especially with RBCs, RBC casts)
In addition to disorders listed under *moderately increased albuminuria consider
Myeloma-associated kidney disease (check UPEP)
Intermittent proteinuria
Postural proteinuria
Congestive heart failure
Fever
Exercise
Nephrotic syndrome
Diabetes
Amyloidosis
Minimal change disease
FSGS
Membranous glomerulopathy
IgA nephropathy

*Moderately and severely increased albuminuria were previously termed “microalbuminuria” and “macroalbuminuria,” respectively.

FIGURE 55-3 Approach to the patient with proteinuria. Investigation of proteinuria is often initiated by a positive dipstick on routine urinalysis. Conventional dipsticks detect predominantly albumin and provide a semiquantitative assessment (trace, 1+, 2+, or 3+), which is influenced by urinary concentration as reflected by urine specific gravity (minimum, <1.005; maximum, 1.030). However, more exact determination of proteinuria should employ a spot morning protein/creatinine ratio (mg/g) or a 24-h urine collection (mg/24 h). FSGS, focal segmental glomerulosclerosis; RBC, red blood cell; UPEP, urine protein electrophoresis.

protein to creatinine in a random “spot” urine can also provide a rough estimate of protein excretion; for example, a protein/creatinine ratio of 3.0 correlates to ~3.0 g of proteinuria per day. Formal assessment of urinary protein excretion requires a 24-h urine protein collection (see “Assessment of GFR,” above).

The magnitude of proteinuria and its composition in the urine depend on the mechanism of renal injury that leads to protein losses. Both charge and size selectivity normally prevent virtually all plasma albumin, globulins, and other high-molecular-weight proteins from crossing the glomerular wall; however, if this barrier is disrupted, plasma proteins may leak into the urine (glomerular proteinuria; Fig. 55-3). Smaller proteins (<20 kDa) are freely filtered but are readily reabsorbed by the proximal tubule. Typically, healthy individuals excrete <150 mg/d of total protein and <30 mg/d of albumin. However, even at albuminuria levels <30 mg/d, risk for progression to overt nephropathy or subsequent cardiovascular disease is increased. The remainder of the protein in the urine is secreted by the tubules (Tamm-Horsfall, IgA, and urokinase) or represents small amounts of filtered β_2-microglobulin, apoproteins, enzymes, and peptide hormones. Another mechanism of proteinuria entails excessive production of an abnormal protein that exceeds the capacity of the tubule for reabsorption. This situation most commonly occurs with plasma cell dyscrasias, such as multiple myeloma, amyloidosis, and lymphomas, that are associated with monoclonal production of immunoglobulin light chains. Other causes include lysozyme-associated nephropathy, a rare cause of kidney injury in patients with chronic myelomonocytic leukemia (CMML); overproduction of lysozyme results in excessive reabsorption of the enzyme by the proximal tubule, resulting in a severe tubulopathy with intracytoplasmic, membrane-bound vacuoles containing homogenous or granular electron dense material on electron microscopy.

The normal glomerular endothelial cell forms a barrier composed of pores of ~100 nm that retain blood cells but offer little impediment to passage of most proteins. The glomerular basement membrane traps most large proteins (>100 kDa), and the foot processes of epithelial cells (podocytes) cover the urinary side of the glomerular basement membrane and produce a series of narrow channels (slit diaphragms) to allow molecular passage of small solutes and water but not proteins. Some glomerular diseases, such as minimal change disease, cause fusion of glomerular epithelial cell foot processes, resulting in predominantly “selective” (Fig. 55-3) loss of albumin. Other glomerular diseases can present with disruption of the basement membrane and slit diaphragms (e.g., by immune complex deposition), resulting in losses of albumin and other plasma proteins. The fusion of foot processes causes increased pressure across the capillary basement membrane, resulting in areas with larger pore sizes (and more severe “nonselective” proteinuria) (Fig. 55-3).

When the total daily urinary excretion of protein is >3.5 g, hypoalbuminemia, hyperlipidemia, and edema (nephrotic syndrome; Fig. 55-3) are often present as well. However, total daily urinary protein excretion >3.5 g can occur without the other features of the nephrotic syndrome in a variety of other renal diseases, including diabetes (Fig. 55-3). Plasma cell dyscrasias (multiple myeloma) can be associated with large amounts of excreted light chains in the urine, which may not be detected by dipstick. The light chains are filtered by the glomerulus and overwhelm the reabsorptive capacity of the proximal tubule. Renal failure from these disorders occurs through a variety of mechanisms, including but not limited to proximal tubule injury, tubule obstruction (cast nephropathy), amyloid deposition, and light chain deposition **(Chap. 328)**. The specific renal lesion is dictated by the sequence and structural characteristics of the monoclonal light chain; however, not all excreted light chains are nephrotoxic.

Hypoalbuminemia in nephrotic syndrome occurs through excessive urinary losses and increased proximal tubule catabolism of filtered albumin. Edema results from renal sodium retention and reduced plasma oncotic pressure, which favors fluid movement from capillaries to interstitium. To compensate for the perceived decrease in effective intravascular volume, activation of the renin-angiotensin system, stimulation of AVP, and activation of the sympathetic nervous system take place, promoting continued renal salt and water reabsorption and progressive edema. Filtered proteases, normally retained by the glomerular filtration barrier, can also directly activate sodium reabsorption by the epithelial Na channels in principal cells (ENaC) in nephrotic syndrome. Despite these changes, hypertension is uncommon in primary kidney diseases resulting in the nephrotic syndrome (Fig. 55-3 and **Chap. 326**). The urinary loss of regulatory proteins and changes in hepatic synthesis contribute to the other manifestations of the nephrotic syndrome. A hypercoagulable state may arise from urinary losses of antithrombin III, reduced serum levels of proteins S and C, hyperfibrinogenemia, and enhanced platelet aggregation. Hypercholesterolemia may be severe and results from increased hepatic lipoprotein synthesis. Loss of immunoglobulins contributes to an increased risk of infection. Many diseases (some listed in Fig. 55-3) and drugs can cause the nephrotic syndrome; a complete list is found in **Chap. 326**.

HEMATURIA, PYURIA, AND CASTS

Isolated hematuria without proteinuria, other cells, or casts is often indicative of bleeding from the urinary tract. Hematuria is defined as two to five RBCs per high-power field (HPF) and can be detected by dipstick. A false-positive dipstick for hematuria (where no RBCs are seen on urine microscopy) may occur when myoglobinuria is present, often in the setting of rhabdomyolysis. Common causes of isolated hematuria include stones, neoplasms, tuberculosis, trauma, and prostatitis. Gross hematuria with blood clots usually is not an intrinsic renal process; rather, it suggests a postrenal source in the urinary collecting system. Evaluation of patients presenting with microscopic hematuria is outlined in Fig. 55-2. A single urinalysis with hematuria is common and can result from menstruation, viral illness, allergy, exercise, or mild trauma. Persistent or significant hematuria (>3 RBCs/HPF on three urinalyses, a single urinalysis with >100 RBCs, or gross hematuria) is associated with significant renal or urologic lesions in 9.1% of cases. The level of suspicion for urogenital neoplasms in patients with isolated painless hematuria and nondysmorphic RBCs increases with age. Neoplasms are rare in the pediatric population, and isolated hematuria is more likely to be "idiopathic" or associated with a congenital anomaly. Hematuria with pyuria and bacteriuria is typical of infection and should be treated with antibiotics after appropriate cultures. Acute cystitis or urethritis in women can cause gross hematuria. Hypercalciuria and hyperuricosuria are also risk factors for unexplained isolated hematuria in both children and adults. In some of these patients (50–60%), reducing calcium and uric acid excretion through dietary interventions can eliminate the microscopic hematuria.

Isolated microscopic hematuria can be a manifestation of glomerular diseases. The RBCs of glomerular origin are often dysmorphic when examined by phase-contrast microscopy. Irregular shapes of RBCs may also result from pH and osmolarity changes produced along the distal nephron. Observer variability in detecting dysmorphic RBCs is common. The most common etiologies of isolated glomerular hematuria are IgA nephropathy, hereditary nephritis, and thin basement membrane disease. IgA nephropathy and hereditary nephritis can lead to episodic gross hematuria. A family history of renal failure is often present in hereditary nephritis, and patients with thin basement membrane disease often have family members with microscopic hematuria. A renal biopsy is needed for the definitive diagnosis of these disorders, which are discussed in more detail in Chap. 326. Hematuria with dysmorphic RBCs, RBC casts, and protein excretion >500 mg/d is virtually diagnostic of glomerulonephritis. RBC casts form as RBCs that enter the tubule fluid and become trapped in a cylindrical mold of gelled Tamm-Horsfall protein. Even in the absence of azotemia, these patients should undergo serologic evaluation and renal biopsy as outlined in Fig. 55-2.

Isolated pyuria is unusual since inflammatory reactions in the kidney or collecting system also are associated with hematuria. The presence of bacteria suggests infection, and WBC casts with bacteria are indicative of pyelonephritis; "sterile pyuria" with negative urinary bacterial cultures can be seen in urogenital tuberculosis. WBCs and/or WBC casts also may be seen in acute glomerulonephritis as well as in tubulointerstitial processes such as interstitial nephritis and transplant rejection.

Casts can be seen in chronic renal diseases. Degenerated cellular casts called *waxy casts* or *broad casts* (arising in the dilated tubules that have undergone compensatory hypertrophy in response to reduced renal mass) may be seen in the urine.

ABNORMALITIES OF URINE VOLUME

POLYURIA

By history, it is often difficult for patients to distinguish urinary frequency (often of small volumes) from true polyuria (>3 L/d), and a quantification of volume by 24-h urine collection may be needed (Fig. 55-4). Polyuria results from two potential mechanisms: (1) excretion of nonabsorbable solutes (such as glucose) or (2) excretion of water (usually from a defect in AVP production or renal responsiveness). To distinguish a solute diuresis from a water diuresis and to determine whether the diuresis is appropriate for the clinical circumstances, urine osmolality is measured. The average person excretes between 600 and 800 mosmol of solutes per day, primarily as urea and electrolytes. If the urine output is >3 L/d and the urine is dilute (<250 mosmol/L), total osmolar excretion is normal and a water diuresis is present. This circumstance could arise from polydipsia, inadequate secretion of AVP (*central diabetes insipidus*), or failure of renal tubules to respond to AVP (*nephrogenic diabetes insipidus*). If the urine volume is >3 L/d and urine osmolality is >300 mosmol/L, a solute diuresis is clearly present and a search for the responsible solute(s) is mandatory.

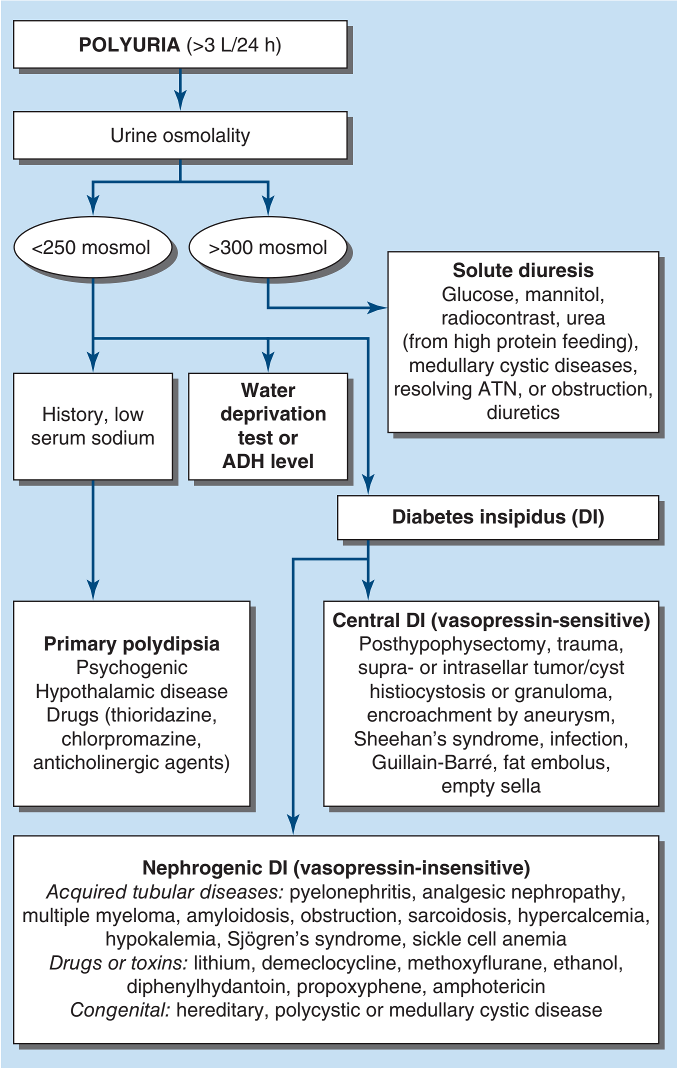

POLYURIA (>3 L/24 h)
Urine osmolality
<250 mosmol
>300 mosmol
Solute diuresis
Glucose, mannitol, radiocontrast, urea (from high protein feeding), medullary cystic diseases, resolving ATN, or obstruction, diuretics
History, low serum sodium
Water deprivation test or ADH level
Diabetes insipidus (DI)
Primary polydipsia
Psychogenic
Hypothalamic disease
Drugs (thioridazine, chlorpromazine, anticholinergic agents)
Central DI (vasopressin-sensitive)
Posthypophysectomy, trauma, supra- or intrasellar tumor/cyst histiocystosis or granuloma, encroachment by aneurysm, Sheehan's syndrome, infection, Guillain-Barré, fat embolus, empty sella
Nephrogenic DI (vasopressin-insensitive)
Acquired tubular diseases: pyelonephritis, analgesic nephropathy, multiple myeloma, amyloidosis, obstruction, sarcoidosis, hypercalcemia, hypokalemia, Sjögren's syndrome, sickle cell anemia
Drugs or toxins: lithium, demeclocycline, methoxyflurane, ethanol, diphenylhydantoin, propoxyphene, amphotericin
Congenital: hereditary, polycystic or medullary cystic disease

FIGURE 55-4 Approach to the patient with polyuria. ADH, antidiuretic hormone; ATN, acute tubular necrosis.

Excessive filtration of a poorly reabsorbed solute such as glucose or mannitol can depress reabsorption of NaCl and water in the proximal tubule and lead to enhanced excretion in the urine. Poorly controlled diabetes mellitus with glucosuria is the most common cause of a solute diuresis, leading to volume depletion and serum hypertonicity. Since the urine sodium concentration is less than that of blood, more water than sodium is lost, causing hypernatremia and hypertonicity. Common iatrogenic solute diuresis occurs in association with mannitol administration, radiocontrast media, and high-protein feedings (enteral or parenteral), leading to increased urea production and excretion. Less commonly, excessive sodium loss may result from cystic renal diseases or Bartter's syndrome or may develop during a tubulointerstitial process (such as resolving ATN). In these so-called salt-wasting disorders, the tubule damage results in direct impairment of sodium reabsorption and indirectly reduces the responsiveness of the tubule to aldosterone. Usually, the sodium losses are mild, and the obligatory urine output is <2 L/d; resolving ATN and postobstructive diuresis are exceptions and may be associated with significant natriuresis and polyuria.

Formation of large volumes of dilute urine is usually due to polydipsic states or diabetes insipidus. Primary polydipsia can result from habit, psychiatric disorders, neurologic lesions, or medications. During deliberate polydipsia, extracellular fluid volume is normal or expanded and plasma AVP levels are reduced because serum osmolality tends to be near the lower limits of normal. Urine osmolality is also maximally dilute at 50 mosmol/L.

Central diabetes insipidus may be idiopathic in origin or secondary to a variety of conditions, including hypophysectomy, trauma, neoplastic, inflammatory, vascular, or infectious hypothalamic diseases. Idiopathic central diabetes insipidus is associated with selective destruction of the AVP-secreting neurons in the supraoptic and paraventricular nuclei and can either be inherited as an autosomal dominant trait or occur spontaneously. Nephrogenic diabetes insipidus can occur in a variety of clinical situations, as summarized in Fig. 55-4.

A plasma AVP level is recommended as the best method for distinguishing between central and nephrogenic diabetes insipidus. Assays for circulating copeptin, a peptide that is cleaved from pre-pro-AVP during axonal transport in the posterior pituitary, are also now available in many centers. A water deprivation test plus exogenous desmopressin (DDAVP) may distinguish primary polydipsia from central and nephrogenic diabetes insipidus. Measurement of hypertonic saline–stimulated plasma copeptin, if available, can substitute for water deprivation testing. **For a detailed discussion, see Chap. 393.**

ACKNOWLEDGMENT

Julie Lin and Brad Denker contributed to this chapter in the 19th edition and some material from that chapter has been retained here.

■ FURTHER READING

Emmett M et al: Approach to the patient with kidney disease, in *Brenner and Rector's The Kidney*, 10th ed, K Skorecki et al (eds). Philadelphia, W.B. Saunders & Company, 2016, pp. 754–779.

Eneanya ND et al: Reconsidering the consequences of using race to estimate kidney function. JAMA 322:113, 2019.

Köhler H et al: Acanthocyturia—a characteristic marker for glomerular bleeding. Kidney Int 40:115, 1991.

Inker LA et al: New creatinine- and cystatin C-based equations to estimate GFR without race. N Engl J Med 385:1737, 2021.

Perazella MA: The urine sediment as a biomarker of kidney disease. Am J Kidney Dis 66:748, 2015.

Weisord SD et al: Prevention and management of acute kidney injury in *Brenner and Rector's The Kidney*, 11th ed, ASL Yu et al (eds). Philadelphia, W.B. Saunders & Company, 2020, pp. 940–977.

56

Fluid and Electrolyte Disturbances

David B. Mount

SODIUM AND WATER

■ COMPOSITION OF BODY FLUIDS

Water is the most abundant constituent in the body, comprising ~50% of body weight in women and 60% in men. Total-body water is distributed in two major compartments: 55–75% is intracellular (intracellular fluid [ICF]), and 25–45% is extracellular (extracellular fluid [ECF]). The ECF is further subdivided into intravascular (plasma water) and extravascular (interstitial) spaces in a ratio of 1:3. Fluid movement between the intravascular and interstitial spaces occurs across the capillary wall and is determined by Starling forces, i.e., capillary hydraulic pressure and colloid osmotic pressure. The transcapillary hydraulic pressure gradient exceeds the corresponding oncotic pressure gradient, thereby favoring the movement of plasma ultrafiltrate into the extravascular space. The return of fluid into the intravascular compartment occurs via lymphatic flow.

The solute or particle concentration of a fluid is known as its osmolality, expressed as milliosmoles per kilogram of water (mOsm/kg). Water easily diffuses across most cell membranes to achieve osmotic equilibrium (ECF osmolality = ICF osmolality). Notably, the extracellular and intracellular solute compositions differ considerably owing to the activity of various transporters, channels, and ATP-driven membrane pumps. The major ECF particles are Na^+ and its accompanying anions Cl^- and HCO_3^-, whereas K^+ and organic phosphate esters (ATP, creatine phosphate, and phospholipids) are the predominant ICF osmoles. Solutes that are restricted to the ECF or the ICF determine the "tonicity" or effective osmolality of that compartment. Certain solutes, particularly urea, do not contribute to water shifts across most membranes and are thus known as *ineffective osmoles*.

Water Balance Vasopressin secretion, water ingestion, and renal water transport collaborate to maintain human body fluid osmolality between 280 and 295 mOsm/kg. Vasopressin (AVP) is synthesized in magnocellular neurons within the hypothalamus; the distal axons of these neurons project to the posterior pituitary or neurohypophysis, from which AVP is released into the circulation. A network of central "osmoreceptor" neurons, which includes the AVP-expressing magnocellular neurons themselves, sense circulating osmolality via nonselective, stretch-activated cation channels. These osmoreceptor neurons are activated or inhibited by modest increases and decreases in circulating osmolality, respectively; activation leads to AVP release and thirst.

AVP secretion is stimulated as systemic osmolality increases above a threshold level of ~285 mOsm/kg, above which there is a linear relationship between osmolality and circulating AVP **(Fig. 56-1)**. Thirst and thus water ingestion are also activated at ~285 mOsm/kg, beyond which there is an equivalent linear increase in the perceived intensity of thirst as a function of circulating osmolality. Changes in blood volume and blood pressure are also direct stimuli for AVP release and thirst, albeit with a less sensitive response profile. Of perhaps greater clinical relevance to the pathophysiology of water homeostasis, ECF volume strongly modulates the relationship between circulating osmolality and AVP release, such that hypovolemia reduces the osmotic threshold and increases the slope of the response curve to osmolality; *hypervolemia* has an opposite effect, increasing the osmotic threshold and reducing the slope of the response curve (Fig. 56-1). Notably, AVP has a half-life in the circulation of only 10–20 min; thus, changes in ECF volume and/or circulating osmolality can rapidly affect water homeostasis. In addition to volume status, a number of other "nonosmotic" stimuli have potent activating effects on

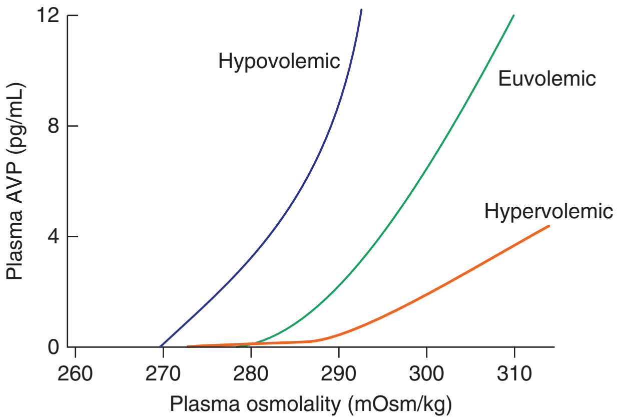

FIGURE 56-1 Circulating levels of vasopressin (AVP) in response to changes in osmolality. Plasma AVP becomes detectable in euvolemic, healthy individuals at a threshold of ~285 mOsm/kg, above which there is a linear relationship between osmolality and circulating AVP. The AVP response to osmolality is modulated strongly by volume status. The osmotic threshold is thus slightly lower in hypovolemia, with a steeper response curve; hypervolemia reduces the sensitivity of circulating AVP levels to osmolality.

osmosensitive neurons and AVP release, including nausea, intracerebral angiotensin II, serotonin, and multiple drugs.

The excretion or retention of electrolyte-free water by the kidney is modulated by circulating AVP. AVP acts on renal, V_2-type receptors in the thick ascending limb of Henle and principal cells of the collecting duct (CD), increasing intracellular levels of cyclic AMP and activating protein kinase A (PKA)–dependent phosphorylation of multiple transport proteins. The AVP- and PKA-dependent activation of Na^+-Cl^- and K^+ transport by the thick ascending limb of the loop of Henle (TALH) is a key participant in the countercurrent mechanism **(Fig. 56-2)**. The countercurrent mechanism ultimately increases the interstitial osmolality in the inner medulla of the kidney, driving water absorption across the renal CD. However, water, salt, and solute transport by both proximal and distal nephron segments participates in the renal concentrating mechanism (Fig. 56-2). Water transport across apical and basolateral aquaporin-1 water channels in the descending thin limb of the loop of Henle is thus involved, as is passive absorption of Na^+-Cl^- by the thin ascending limb, via apical and basolateral CLC-K1 chloride channels and paracellular Na^+ transport. Renal urea transport in turn plays important roles in the generation of the medullary osmotic gradient and the ability to excrete solute-free water under conditions of both high and low protein intake (Fig. 56-2).

AVP-induced, PKA-dependent phosphorylation of the aquaporin-2 water channel in principal cells stimulates the insertion of active water channels into the lumen of the CD, resulting in transepithelial water absorption down the medullary osmotic gradient **(Fig. 56-3)**. Under "antidiuretic" conditions, with increased circulating AVP, the kidney reabsorbs water filtered by the glomerulus, equilibrating the osmolality across the CD epithelium to excrete a hypertonic, "concentrated" urine (osmolality of up to 1200 mOsm/kg). In the absence of circulating AVP, insertion of aquaporin-2 channels and water absorption across the CD is essentially abolished, resulting in secretion of a hypotonic, dilute urine (osmolality as low as 30–50 mOsm/kg). Abnormalities in this "final common pathway" are involved in most disorders of water homeostasis, e.g., a reduced or absent insertion of active aquaporin-2 water channels into the membrane of principal cells in diabetes insipidus (DI).

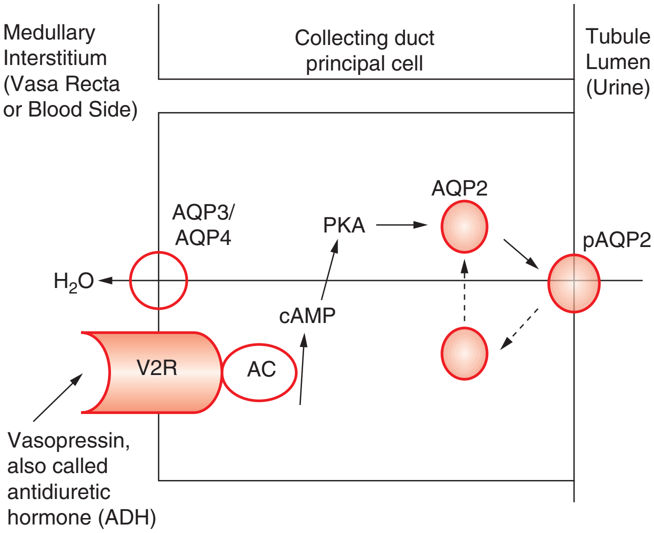
Medullary Interstitium (Vasa Recta or Blood Side)
Collecting duct principal cell
Tubule Lumen (Urine)
AQP3/ AQP4
H_2O
PKA
AQP2
pAQP2
cAMP
V2R
AC
Vasopressin, also called antidiuretic hormone (ADH)

FIGURE 56-3 Vasopressin and the regulation of water permeability in the renal collecting duct. Vasopressin binds to the type 2 vasopressin receptor (V2R) on the basolateral membrane of principal cells, activates adenylyl cyclase (AC), increases intracellular cyclic adenosine monophosphatase (cAMP), and stimulates protein kinase A (PKA) activity. Cytoplasmic vesicles carrying aquaporin-2 (AQP) water channel proteins are inserted into the luminal membrane in response to vasopressin, thereby increasing the water permeability of this membrane. When vasopressin stimulation ends, water channels are retrieved by an endocytic process and water permeability returns to its low basal rate. The AQP3 and AQP4 water channels are expressed on the basolateral membrane and complete the transcellular pathway for water reabsorption. pAQP2, phosphorylated aquaporin-2. *(From Annals of Internal Medicine JM Sands, DG Bichet: Nephrogenic diabetes insipidus. 144:186, 2006. Copyright © 2006 American College of Physicians. All Rights Reserved. Reprinted with the permission of American College of Physicians, Inc.)*

Maintenance of Arterial Circulatory Integrity Sodium is actively pumped out of cells by the Na^+/K^+-ATPase membrane pump. In consequence, 85–90% of body Na^+ is extracellular, and the ECF volume (ECFV) is a function of total-body Na^+ content. Arterial perfusion and circulatory integrity are, in turn, determined by renal Na^+ retention or excretion, in addition to the modulation of systemic arterial resistance. Within the kidney, Na^+ is filtered by the glomeruli and then sequentially reabsorbed by the renal tubules. The Na^+ cation is typically reabsorbed with the chloride anion (Cl^-), and thus, chloride homeostasis also affects the ECFV. On a quantitative level, at a glomerular filtration rate (GFR) of 180 L/d and serum Na^+ of ~140 m*M*, the kidney filters some 25,200 mmol/d of Na^+. This is equivalent to ~1.5 kg of salt, which would occupy roughly 10 times the extracellular space; 99.6% of filtered Na^+-Cl^- must be reabsorbed to excrete 100 m*M* per day. Minute changes in renal Na^+-Cl^- excretion will thus have significant effects on the ECFV, leading to edema syndromes or hypovolemia.

Approximately two-thirds of filtered Na^+-Cl^- is reabsorbed by the renal proximal tubule, via both paracellular and transcellular mechanisms. The TALH subsequently reabsorbs another 25–30% of filtered Na^+-Cl^- via the apical, furosemide-sensitive Na^+-K^+-$2Cl^-$ cotransporter. The adjacent aldosterone-sensitive distal nephron, comprising the distal convoluted tubule (DCT), connecting tubule (CNT), and CD, accomplishes the "fine-tuning" of renal Na^+-Cl^- excretion. The thiazide-sensitive apical Na^+-Cl^- cotransporter (NCC) reabsorbs 5–10% of filtered Na^+-Cl^- in the DCT. Principal cells in the CNT and CD reabsorb Na^+ via electrogenic, amiloride-sensitive epithelial Na^+ channels (ENaC); Cl^- ions are primarily reabsorbed by adjacent intercalated cells, via apical Cl^- exchange (Cl^--OH^- and Cl^--HCO_3^- exchange, mediated by the SLC26A4 anion exchanger) **(Fig. 56-4)**.

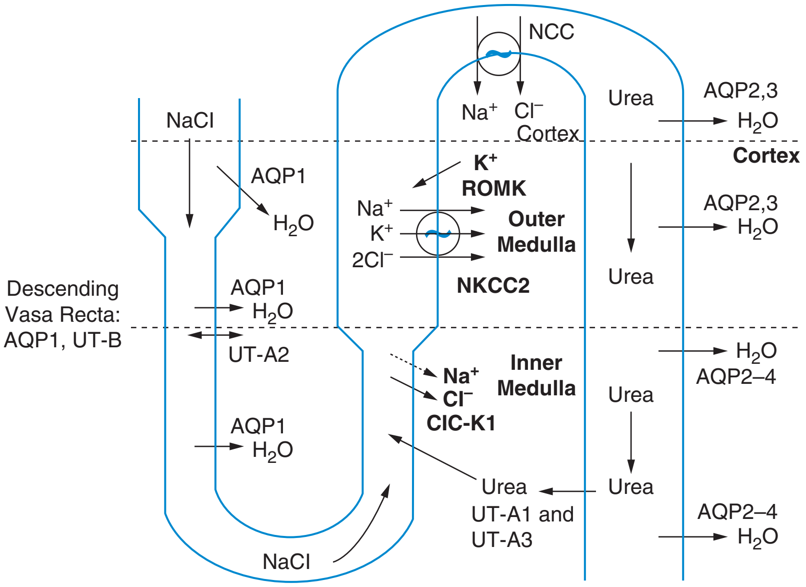

FIGURE 56-2 The renal concentrating mechanism. Water, salt, and solute transport by both proximal and distal nephron segments participates in the renal concentrating mechanism (see text for details). Diagram showing the location of the major transport proteins involved; a loop of Henle is depicted on the *left,* collecting duct on the *right.* AQP, aquaporin; CLC-K1, chloride channel; NKCC2, Na-K-2Cl cotransporter; ROMK, renal outer medullary K^+ channel; UT, urea transporter. *(Reproduced with permission from JM Sands: Molecular approaches to urea transporters. J Am Soc Nephro 13(11):2795, 2002.)*

Renal tubular reabsorption of filtered Na^+-Cl^- is regulated by multiple circulating and paracrine hormones, in addition to the activity of renal nerves. Angiotensin II activates proximal Na^+-Cl^- reabsorption, as do adrenergic receptors under the influence of renal sympathetic

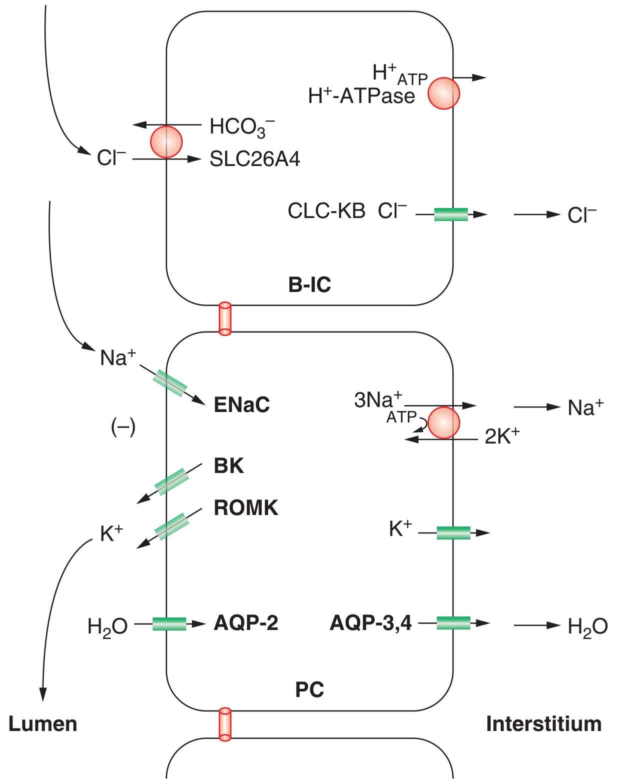

FIGURE 56-4 Sodium, water, and potassium transport in principal cells (PC) and adjacent α-intercalated cells (B-IC). The absorption of Na^+ via the amiloride-sensitive epithelial sodium channel (ENaC) generates a lumen-negative potential difference, which drives K^+ excretion through the apical secretory K^+ channel ROMK (renal outer medullary K^+ channel) and/or the flow-dependent BK channel. Transepithelial Cl^- transport occurs in adjacent β-intercalated cells, via apical Cl^--HCO_3^- and Cl^--OH^- exchange (SLC26A4 anion exchanger, also known as pendrin) basolateral CLC chloride channels. Water is absorbed down the osmotic gradient by principal cells, through the apical aquaporin-2 (AQP-2) and basolateral aquaporin-3 and aquaporin-4 (Fig. 56-3).

innervation; locally generated dopamine, in contrast, has a *natriuretic* effect. Aldosterone primarily activates Na^+-Cl^- reabsorption within the aldosterone-sensitive distal nephron. In particular, aldosterone activates the ENaC channel in principal cells, inducing Na^+ absorption and promoting K^+ excretion (Fig. 56-4).

Circulatory integrity is critical for the perfusion and function of vital organs. “Underfilling” of the arterial circulation is sensed by ventricular and vascular pressure receptors, resulting in a neurohumoral activation (increased sympathetic tone, activation of the renin-angiotensin-aldosterone axis, and increased circulating AVP) that synergistically increases renal Na^+-Cl^- reabsorption, vascular resistance, and renal water reabsorption. This occurs in the context of decreased cardiac output, as occurs in hypovolemic states, low-output cardiac failure, decreased oncotic pressure, and/or increased capillary permeability. Alternatively, excessive arterial vasodilation results in *relative* arterial underfilling, leading to neurohumoral activation in the defense of tissue perfusion. These physiologic responses play important roles in many of the disorders discussed in this chapter. In particular, it is important to appreciate that AVP functions in the defense of circulatory integrity, inducing vasoconstriction, increasing sympathetic nervous system tone, increasing renal retention of both water and Na^+-Cl^-, and modulating the arterial baroreceptor reflex. Most of these responses involve activation of systemic V_{1A} AVP receptors, but concomitant activation of V_2 receptors in the kidney can result in renal water retention and hyponatremia.

HYPOVOLEMIA

Etiology True volume depletion, or hypovolemia, generally refers to a state of combined salt and water loss, leading to contraction of the ECFV. The loss of salt and water may be renal or nonrenal in origin.

RENAL CAUSES Excessive urinary Na^+-Cl^- and water loss is a feature of several conditions. A high filtered load of endogenous solutes, such as glucose and urea, can impair tubular reabsorption of Na^+-Cl^- and water, leading to an osmotic diuresis. Exogenous mannitol, often used to decrease intracerebral pressure, is filtered by glomeruli but not reabsorbed by the proximal tubule, thus causing an osmotic diuresis. Pharmacologic diuretics selectively impair Na^+-Cl^- reabsorption at specific sites along the nephron, leading to increased urinary Na^+-Cl^- excretion. Other drugs can induce natriuresis as a side effect. For example, acetazolamide can inhibit proximal tubular Na^+-Cl^- absorption via its inhibition of carbonic anhydrase; other drugs, such as the antibiotics trimethoprim (TMP) and pentamidine, inhibit distal tubular Na^+ reabsorption through the amiloride-sensitive ENaC channel, leading to urinary Na^+-Cl^- loss. Hereditary defects in renal transport proteins are also associated with reduced reabsorption of filtered Na^+-Cl^- and/or water. Alternatively, mineralocorticoid deficiency, mineralocorticoid resistance, or inhibition of the mineralocorticoid receptor (MLR) can reduce Na^+-Cl^- reabsorption by the aldosterone-sensitive distal nephron. Finally, tubulointerstitial injury, as occurs in interstitial nephritis, acute tubular injury, or obstructive uropathy, can reduce distal tubular Na^+-Cl^- and/or water absorption.

Excessive excretion of free water, i.e., water without electrolytes, can also lead to hypovolemia. However, the effect on ECFV is usually less marked, given that two-thirds of the water volume is lost from the ICF. Excessive renal water excretion occurs in the setting of decreased circulating AVP or renal resistance to AVP (central and nephrogenic DI, respectively).

EXTRARENAL CAUSES Nonrenal causes of hypovolemia include fluid loss from the gastrointestinal tract, skin, and respiratory system. Accumulations of fluid within specific tissue compartments, typically the interstitium, peritoneum, or gastrointestinal tract, can also cause hypovolemia.

Approximately 9 L of fluid enter the gastrointestinal tract daily, 2 L by ingestion and 7 L by secretion; almost 98% of this volume is absorbed, such that daily fecal fluid loss is only 100–200 mL. Impaired gastrointestinal reabsorption or enhanced secretion of fluid can cause hypovolemia. Because gastric secretions have a low pH (high H^+ concentration), whereas biliary, pancreatic, and intestinal secretions are alkaline (high HCO_3^- concentration), vomiting and diarrhea are often accompanied by metabolic alkalosis and acidosis, respectively.

Evaporation of water from the skin and respiratory tract (so-called “insensible losses”) constitutes the major route for loss of solute-free water, which is typically 500–650 mL/d in healthy adults. This evaporative loss can increase during febrile illness or prolonged heat exposure. Hyperventilation can also increase insensible losses via the respiratory tract, particularly in ventilated patients; the humidity of inspired air is another determining factor. In addition, increased exertion and/or ambient temperature will increase insensible losses via sweat, which is hypotonic to plasma. Profuse sweating without adequate repletion of water and Na^+-Cl^- can thus lead to both hypovolemia and hypertonicity. Alternatively, replacement of these insensible losses with a surfeit of free water, without adequate replacement of electrolytes, may lead to hypovolemic hyponatremia.

Excessive fluid accumulation in interstitial and/or peritoneal spaces can also cause intravascular hypovolemia. Increases in vascular permeability and/or a reduction in oncotic pressure (hypoalbuminemia) alter Starling forces, resulting in excessive “third spacing” of the ECFV. This occurs in sepsis syndrome, burns, pancreatitis, nutritional hypoalbuminemia, and peritonitis. Alternatively, distributive hypovolemia can occur due to accumulation of fluid within specific compartments, for example, within the bowel lumen in gastrointestinal obstruction or ileus. Hypovolemia can also occur after extracorporeal hemorrhage or after significant hemorrhage into an expandable space, for example, the retroperitoneum.

Diagnostic Evaluation A careful history will usually determine the etiologic cause of hypovolemia. Symptoms of hypovolemia are nonspecific and include fatigue, weakness, thirst, and postural dizziness; more severe symptoms and signs include oliguria, cyanosis, abdominal and chest pain, and confusion or obtundation. Associated electrolyte disorders may cause additional symptoms, for example, muscle weakness in patients with hypokalemia. On examination, diminished skin turgor and dry oral mucous membranes are less than ideal markers of a decreased ECFV in adult patients; more reliable signs of hypovolemia include a decreased jugular venous pressure (JVP), orthostatic tachycardia (an increase of >15–20 beats/min upon standing), and orthostatic hypotension (a >10–20 mmHg drop in blood pressure on standing). More severe fluid loss leads to hypovolemic shock, with hypotension, tachycardia, peripheral vasoconstriction, and peripheral hypoperfusion; these patients may exhibit peripheral cyanosis, cold extremities, oliguria, and altered mental status.

Routine chemistries may reveal an increase in blood urea nitrogen (BUN) and creatinine, reflective of a decrease in GFR. Creatinine is the more dependable measure of GFR, because BUN levels may be influenced by an increase in tubular reabsorption ("prerenal azotemia"), an increase in urea generation in catabolic states, hyperalimentation, or gastrointestinal bleeding, and/or a decreased urea generation in decreased protein intake. In hypovolemic shock, liver function tests and cardiac biomarkers may show evidence of hepatic and cardiac ischemia, respectively. Routine chemistries and/or blood gases may reveal evidence of acid-base disorders. For example, bicarbonate loss due to diarrheal illness is a very common cause of metabolic acidosis; alternatively, patients with severe hypovolemic shock may develop lactic acidosis with an elevated anion gap.

The neurohumoral response to hypovolemia stimulates an increase in renal tubular Na^+ and water reabsorption. Therefore, the urine Na^+ concentration is typically <20 m*M* in nonrenal causes of hypovolemia, with a urine osmolality of >450 mOsm/kg. The reduction in both GFR and distal tubular Na^+ delivery may cause a defect in renal potassium excretion, with an increase in plasma K^+ concentration. Of note, patients with hypovolemia and a hypochloremic alkalosis due to vomiting, diarrhea, or diuretics will typically have a urine Na^+ concentration >20 m*M* and urine pH of >7.0, due to the increase in filtered HCO_3^-; the urine Cl^- concentration in this setting is a more accurate indicator of volume status, with a level <25 m*M* suggestive of hypovolemia. The urine Na^+ concentration is often >20 m*M* in patients with *renal* causes of hypovolemia, such as acute tubular necrosis; similarly, patients with DI will have an inappropriately dilute urine.

TREATMENT

Hypovolemia

The therapeutic goals in hypovolemia are to restore normovolemia and replace ongoing fluid losses. Mild hypovolemia can usually be treated with oral hydration and resumption of a normal maintenance diet. More severe hypovolemia requires intravenous hydration, tailoring the choice of solution to the underlying pathophysiology. Isotonic, "normal" saline (0.9% NaCl, 154 m*M* Na^+) is the most appropriate resuscitation fluid for normonatremic or hyponatremic patients with severe hypovolemia; colloid solutions such as intravenous albumin are not demonstrably superior for this purpose. Hypernatremic patients should receive a hypotonic solution, 5% dextrose if there has only been water loss (as in DI), or hypotonic saline (1/2 or 1/4 normal saline) if there has been water and Na^+-Cl^- loss; changes in free water administration should be made if necessary, based on frequent measuring of serum chemistries. Patients with bicarbonate loss and metabolic acidosis, as occur frequently in diarrhea, should receive intravenous bicarbonate, either an isotonic solution (150 meq of Na^+-HCO_3^- in 5% dextrose) or a more hypotonic bicarbonate solution in dextrose or dilute saline. Patients with severe hemorrhage or anemia should receive red cell transfusions, without increasing the hematocrit beyond 35%.

SODIUM DISORDERS

Disorders of serum Na^+ concentration are caused by abnormalities in water homeostasis, leading to changes in the relative ratio of Na^+ to body water. Water intake and circulating AVP constitute the two key effectors in the defense of serum osmolality; defects in one or both of these two defense mechanisms cause most cases of hyponatremia and hypernatremia. In contrast, abnormalities in sodium homeostasis per se lead to a deficit or surplus of whole-body Na^+-Cl^- content, a key determinant of the ECFV and circulatory integrity. Notably, volume status also modulates the release of AVP by the posterior pituitary, such that hypovolemia is associated with higher circulating levels of the hormone at each level of serum osmolality. Similarly, in "hypervolemic" causes of arterial underfilling, e.g., heart failure and cirrhosis, the associated neurohumoral activation encompasses an increase in circulating AVP, leading to water retention and hyponatremia. Therefore, a key concept in sodium disorders is that the absolute plasma Na^+ concentration tells one nothing about the volume status of a given patient, which furthermore must be taken into account in the diagnostic and therapeutic approach.

HYPONATREMIA

Hyponatremia, which is defined as a plasma Na^+ concentration <135 m*M*, is a very common disorder, occurring in up to 22% of hospitalized patients. This disorder is almost always the result of an increase in circulating AVP and/or increased renal sensitivity to AVP, combined with an intake of free water; a notable exception is hyponatremia due to low solute intake (see below). The underlying pathophysiology for the exaggerated or "inappropriate" AVP response differs in patients with hyponatremia as a function of their ECFV. Hyponatremia is thus subdivided diagnostically into three groups, depending on clinical history and volume status, i.e., "hypovolemic," "euvolemic," and "hypervolemic" **(Fig. 56-5)**.

Hypovolemic Hyponatremia Hypovolemia causes a marked neurohumoral activation, increasing circulating levels of AVP. The increase in circulating AVP helps preserve blood pressure via vascular and baroreceptor V_{1A} receptors and increases water reabsorption via renal V_2 receptors; activation of V_2 receptors can lead to hyponatremia in the setting of increased free water intake. Nonrenal causes of hypovolemic hyponatremia include gastrointestinal loss (e.g., vomiting, diarrhea, tube drainage) and insensible loss (sweating, burns) of Na^+-Cl^- and water, in the absence of adequate oral replacement; urine Na^+ concentration is typically <20 m*M*. Notably, these patients may be clinically classified as euvolemic, with only the reduced urinary Na^+ concentration to indicate the cause of their hyponatremia. Indeed, a urine Na^+ concentration <20 m*M*, in the absence of a cause of hypervolemic hyponatremia, predicts a rapid increase in plasma Na^+ concentration in response to intravenous normal saline; saline therapy thus induces a water diuresis in this setting, as circulating AVP levels plummet.

The *renal* causes of hypovolemic hyponatremia share an inappropriate loss of Na^+-Cl^- in the urine, leading to volume depletion and an increase in circulating AVP; urine Na^+ concentration is typically >20 m*M* (Fig. 56-5). A deficiency in circulating aldosterone and/or its renal effects can lead to hyponatremia in primary adrenal insufficiency and other causes of hypoaldosteronism; hyperkalemia and hyponatremia in a hypotensive and/or hypovolemic patient with high urine Na^+ concentration (much greater than 20 m*M*) should strongly suggest this diagnosis. Salt-losing nephropathies may lead to hyponatremia when sodium intake is reduced, due to impaired renal tubular function; typical causes include reflux nephropathy, interstitial nephropathies, postobstructive uropathy, medullary cystic disease, and the recovery phase of acute tubular necrosis. Thiazide diuretics cause hyponatremia via a number of mechanisms, including polydipsia and diuretic-induced volume depletion; presentations can mimic the syndrome of inappropriate antidiuresis (SIAD). Notably, thiazides do not inhibit the renal concentrating mechanism, such that circulating AVP retains a full effect on renal water retention. In contrast, loop diuretics, which are less frequently associated with hyponatremia, inhibit Na^+-Cl^- and K^+ absorption by the TALH, blunting the countercurrent mechanism and reducing the ability to concentrate the urine. Increased excretion of an osmotically active nonreabsorbable or poorly reabsorbable solute can also lead to volume depletion and hyponatremia; important causes include glycosuria, ketonuria (e.g., in starvation or in diabetic

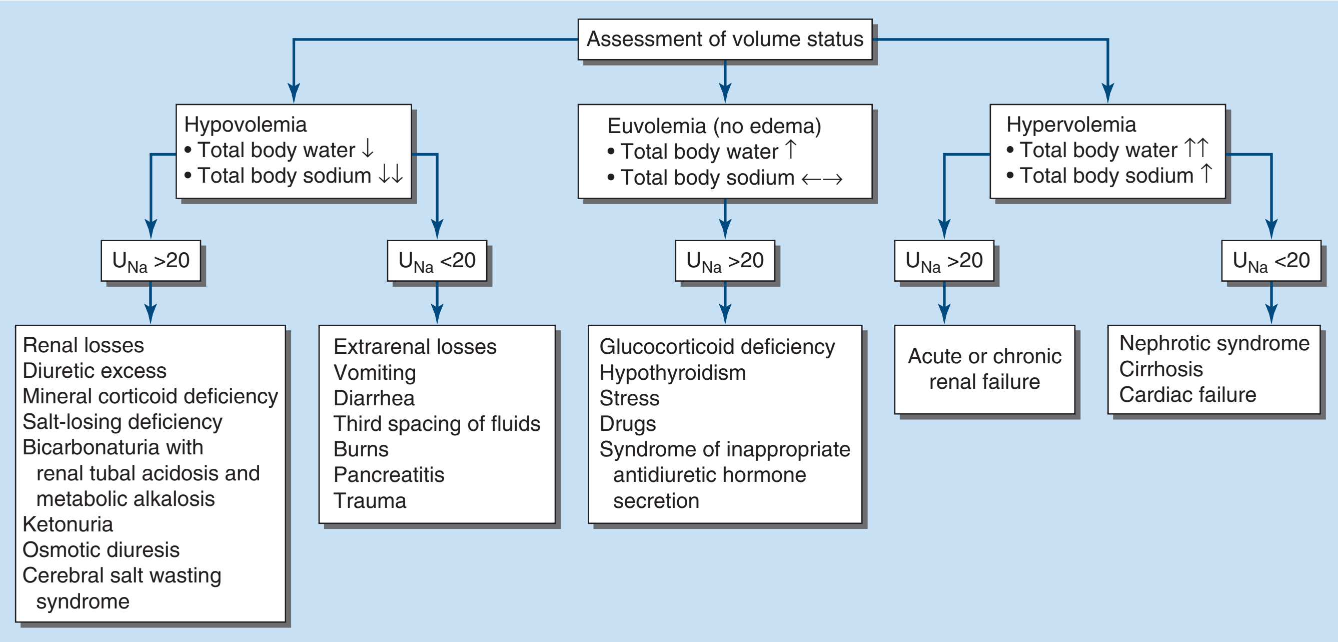
Assessment of volume status
Hypovolemia
• Total body water ↓
• Total body sodium ↓↓
Euvolemia (no edema)
• Total body water ↑
• Total body sodium ←→
Hypervolemia
• Total body water ↑↑
• Total body sodium ↑
U_{Na} >20
U_{Na} <20
U_{Na} >20
U_{Na} >20
U_{Na} <20
Renal losses
Diuretic excess
Mineral corticoid deficiency
Salt-losing deficiency
Bicarbonaturia with renal tubal acidosis and metabolic alkalosis
Ketonuria
Osmotic diuresis
Cerebral salt wasting syndrome
Extrarenal losses
Vomiting
Diarrhea
Third spacing of fluids
Burns
Pancreatitis
Trauma
Glucocorticoid deficiency
Hypothyroidism
Stress
Drugs
Syndrome of inappropriate antidiuretic hormone secretion
Acute or chronic renal failure
Nephrotic syndrome
Cirrhosis
Cardiac failure

FIGURE 56-5 The diagnostic approach to hyponatremia. *(Reproduced with permission from S Kumar, T Berl: Diseases of water metabolism, in RW Schrier [ed], Atlas of Diseases of the Kidney, Philadelphia, Current Medicine, Inc, 1999.)*

or alcoholic ketoacidosis), and bicarbonaturia (e.g., in renal tubular acidosis or metabolic alkalosis, where the associated bicarbonaturia leads to loss of Na^+).

Finally, the syndrome of "cerebral salt wasting" is a rare cause of hypovolemic hyponatremia, encompassing hyponatremia with clinical hypovolemia and inappropriate natriuresis in association with intracranial disease; associated disorders include subarachnoid hemorrhage, traumatic brain injury, craniotomy, encephalitis, and meningitis. Distinction from the more common syndrome SIAD is critical because cerebral salt wasting will typically respond to aggressive Na^+-Cl^- repletion.

Hypervolemic Hyponatremia Patients with hypervolemic hyponatremia develop an increase in total-body Na^+-Cl^- that is accompanied by a proportionately *greater* increase in total-body water, leading to a reduced plasma Na^+ concentration. As in hypovolemic hyponatremia, the causative disorders can be separated by the effect on urine Na^+ concentration, with acute or chronic renal failure uniquely associated with an increase in urine Na^+ concentration (Fig. 56-5). The pathophysiology of hyponatremia in the sodium-avid edematous disorders (congestive heart failure [CHF], cirrhosis, and nephrotic syndrome) is similar to that in hypovolemic hyponatremia, except that arterial filling and circulatory integrity is decreased due to the specific etiologic factors (e.g., cardiac dysfunction in CHF, peripheral vasodilation in cirrhosis). Urine Na^+ concentration is typically very low, i.e., <10 m*M*, even after hydration with normal saline; this Na^+-avid state may be obscured by diuretic therapy. The degree of hyponatremia provides an indirect index of the associated neurohumoral activation and is an important prognostic indicator in hypervolemic hyponatremia.

Euvolemic Hyponatremia Euvolemic hyponatremia can occur in moderate to severe hypothyroidism, with correction after achieving a euthyroid state. Severe hyponatremia can also be a consequence of secondary adrenal insufficiency due to pituitary disease; whereas the deficit in circulating aldosterone in primary adrenal insufficiency causes *hypovolemic* hyponatremia, the predominant glucocorticoid deficiency in secondary adrenal failure is associated with *euvolemic* hyponatremia. Glucocorticoids exert a negative feedback on AVP release by the posterior pituitary such that hydrocortisone replacement in these patients can rapidly normalize the AVP response to osmolality, reducing circulating AVP.

SIAD is the most frequent cause of euvolemic hyponatremia **(Table 56-1)**. The generation of hyponatremia in SIAD requires an intake of free water, with persistent intake at serum osmolalities that are lower than the usual threshold for thirst; as one would expect, the osmotic threshold and osmotic response curves for the sensation of thirst are shifted downward in patients with SIAD. Four distinct patterns of AVP secretion have been recognized in patients with SIAD, independent for the most part of the underlying cause. Unregulated, erratic AVP secretion is seen in about a third of patients, with no obvious correlation between serum osmolality and circulating AVP levels. Other patients fail to suppress AVP secretion at lower serum osmolalities, with a normal response curve to hyperosmolar conditions; others have a "reset osmostat," with a lower threshold osmolality and a left-shifted osmotic response curve. Finally, the fourth subset of patients have essentially no detectable circulating AVP, suggesting either a gain in function in renal water reabsorption or a circulating antidiuretic substance that is distinct from AVP. Gain-in-function mutations of a single specific residue in the V_2 AVP receptor have been described in some of these patients, leading to constitutive activation of the receptor in the absence of AVP and "nephrogenic" SIAD.

Strictly speaking, patients with SIAD are not euvolemic but are subclinically volume-expanded, due to AVP-induced water and Na^+-Cl^- retention; "AVP escape" mechanisms invoked by sustained increases in AVP serve to limit distal renal tubular transport, preserving a modestly hypervolemic steady state. Serum uric acid is often low (<4 mg/dL) in patients with SIAD, consistent with suppressed proximal tubular transport in the setting of increased distal tubular Na^+-Cl^- and water transport; in contrast, patients with hypovolemic hyponatremia will often be hyperuricemic due to a shared activation of proximal tubular Na^+-Cl^- and urate transport.

Common causes of SIAD include pulmonary disease (e.g., pneumonia, tuberculosis, pleural effusion) and central nervous system (CNS) diseases (e.g., tumor, subarachnoid hemorrhage, meningitis). SIAD also occurs with malignancies, most commonly with small-cell lung carcinoma (75% of malignancy-associated SIAD); ~10% of patients with this tumor will have a plasma Na^+ concentration of <130 m*M* at presentation. SIAD is also a frequent complication of certain drugs, most commonly the selective serotonin reuptake inhibitors (SSRIs). Other drugs can potentiate the renal effect of AVP, without exerting direct effects on circulating AVP levels (Table 56-1).

Low Solute Intake and Hyponatremia Hyponatremia can occasionally occur in patients with a very low intake of dietary solutes. Classically, this occurs in alcoholics whose sole nutrient is beer, hence

TABLE 56-1 Causes of the Syndrome of Inappropriate Antidiuresis (SIAD)

MALIGNANT DISEASES	PULMONARY DISORDERS	DISORDERS OF THE CENTRAL NERVOUS SYSTEM	DRUGS	OTHER CAUSES
Carcinoma	Infections	Infection	Drugs that stimulate release of AVP or enhance its action	Hereditary (gain-of-function mutations in the vasopressin V_2 receptor)
Lung	Bacterial pneumonia	Encephalitis		
Small cell	Viral pneumonia	Meningitis	Chlorpropamide	Idiopathic
Mesothelioma	Pulmonary abscess	Brain abscess	SSRIs	Transient
Oropharynx	Tuberculosis	Rocky Mountain spotted fever	Tricyclic antidepressants	Endurance exercise
Gastrointestinal tract	Aspergillosis	AIDS	Clofibrate	General anesthesia
Stomach	Asthma	Bleeding and masses	Carbamazepine	Nausea
Duodenum	Cystic fibrosis	Subdural hematoma	Vincristine	Pain
Pancreas	Respiratory failure associated with positive-pressure breathing	Subarachnoid hemorrhage	Nicotine	Stress
Genitourinary tract		Cerebrovascular accident	Narcotics	
Ureter		Brain tumors	Antipsychotic drugs	
Bladder		Head trauma	Ifosfamide	
Prostate		Hydrocephalus	Cyclophosphamide	
Endometrium		Cavernous sinus thrombosis	Nonsteroidal anti-inflammatory drugs	
Endocrine thymoma		Other	MDMA (“Ecstasy”, “Molly”)	
Lymphomas		Multiple sclerosis	AVP analogues	
Sarcomas		Guillain-Barré syndrome	Desmopressin	
Ewing’s sarcoma		Shy-Drager syndrome	Oxytocin	
		Delirium tremens	Vasopressin	
		Acute intermittent porphyria		

Abbreviations: AVP, vasopressin; MDMA; 3,4-methylenedioxymethamphetamine; SSRI, selective serotonin reuptake inhibitor.

Source: From DH Ellison, T Berl: The syndrome of inappropriate antidiuresis. N Engl J Med 356:2064, 2007. Copyright © 2007 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

the diagnostic label of *beer potomania*; beer is very low in protein and salt content, containing only 1–2 m*M* of Na^+. The syndrome has also been described in nonalcoholic patients with highly restricted solute intake due to nutrient-restricted diets, e.g., extreme vegetarian diets. Patients with hyponatremia due to low solute intake typically present with a very low urine osmolality (<100–200 mOsm/kg) with a urine Na^+ concentration that is <10–20 m*M*. The fundamental abnormality is the inadequate dietary intake of solutes; the reduced urinary solute excretion limits water excretion such that hyponatremia ensues after relatively modest polydipsia. AVP levels have not been reported in patients with beer potomania but are expected to be suppressed or rapidly suppressible with saline hydration; this fits with the overly rapid correction in plasma Na^+ concentration that can be seen with saline hydration. Resumption of a normal diet and/or saline hydration will also correct the causative deficit in urinary solute excretion, such that patients with beer potomania typically correct their plasma Na^+ concentration promptly after admission to the hospital.

Clinical Features of Hyponatremia Hyponatremia induces generalized cellular swelling, a consequence of water movement down the osmotic gradient from the hypotonic ECF to the ICF. The symptoms of hyponatremia are primarily neurologic, reflecting the development of cerebral edema within a rigid skull. The initial CNS response to acute hyponatremia is an increase in interstitial pressure, leading to shunting of ECF and solutes from the interstitial space into the cerebrospinal fluid and then on into the systemic circulation. This is accompanied by an efflux of the major intracellular ions, Na^+, K^+, and Cl^-, from brain cells. Acute hyponatremic encephalopathy ensues when these volume regulatory mechanisms are overwhelmed by a rapid decrease in tonicity, resulting in acute cerebral edema. Early symptoms can include nausea, headache, and vomiting. However, severe complications can rapidly evolve, including seizure activity, brainstem herniation, coma, and death. A key complication of acute hyponatremia is normocapneic or hypercapneic respiratory failure; the associated hypoxia may amplify the neurologic injury. Normocapneic respiratory failure in this setting is typically due to noncardiogenic, “neurogenic” pulmonary edema, with a normal pulmonary capillary wedge pressure.

Acute symptomatic hyponatremia is a medical emergency, occurring in a number of specific settings **(Table 56-2)**. Women, particularly before menopause, are much more likely than men to develop encephalopathy and severe neurologic sequelae. Acute hyponatremia often has an iatrogenic component, e.g., when hypotonic intravenous fluids are given to postoperative patients with an increase in circulating AVP. Exercise-associated hyponatremia, an important clinical issue at marathons and other endurance events, has similarly been linked to both a “nonosmotic” increase in circulating AVP and excessive free water intake. The recreational drugs Molly and Ecstasy, which share an active ingredient (MDMA, 3,4-methylenedioxymethamphetamine), cause a rapid and potent induction of both thirst and AVP, leading to severe acute hyponatremia.

Persistent, chronic hyponatremia results in an efflux of organic osmolytes (creatine, betaine, glutamate, myoinositol, and taurine) from brain cells; this response reduces intracellular osmolality and the osmotic gradient favoring water entry. This reduction in intracellular osmolytes is largely complete within 48 h, the time period that clinically defines chronic hyponatremia; this temporal definition has considerable relevance for the treatment of hyponatremia (see below). The cellular response to chronic hyponatremia does not fully protect patients from symptoms, which can include vomiting, nausea, confusion, and seizures, usually at plasma Na^+ concentration <125 m*M*. Even patients who are judged “asymptomatic” can manifest subtle gait and cognitive defects that reverse with correction of hyponatremia;

TABLE 56-2 Causes of Acute Hyponatremia

- Iatrogenic
  - Postoperative: premenopausal women
  - Hypotonic fluids with cause of ↑ vasopressin
  - Glycine irrigation: TURP, uterine surgery
  - Colonoscopy preparation
- Recent institution of thiazides
- Polydipsia
- MDMA (“Ecstasy,” “Molly”) ingestion
- Exercise induced
- Multifactorial, e.g., thiazide and polydipsia

Abbreviations: MDMA, 3,4-methylenedioxymethamphetamine; TURP, transurethral resection of the prostate.

notably, chronic "asymptomatic" hyponatremia increases the risk of falls. Chronic hyponatremia also increases the risk of bony fractures owing to the associated neurologic dysfunction and to a hyponatremia-associated reduction in bone density. Therefore, every attempt should be made to safely correct the plasma Na^+ concentration in patients with chronic hyponatremia, even in the absence of overt symptoms (see the section on treatment of hyponatremia below).

The management of chronic hyponatremia is complicated significantly by the asymmetry of the cellular response to correction of plasma Na^+ concentration. Specifically, the *reaccumulation* of organic osmolytes by brain cells is attenuated and delayed as osmolality increases after correction of hyponatremia, sometimes resulting in degenerative loss of oligodendrocytes and an osmotic demyelination syndrome (ODS). Overly rapid correction of hyponatremia (>8–10 m*M* in 24 h or 18 m*M* in 48 h) causes hypertonic stress in astrocytes within brain regions prone to ODS, leading to generalized protein ubiquitination and endoplasmic reticulum stress due to activation of the unfolded protein response; this is accompanied by apoptotic and autophagic cell death. Rapid correction of hyponatremia also causes a disruption in integrity of the blood-brain barrier, allowing the entry of immune mediators that may contribute to demyelination. The lesions of ODS classically affect the pons, a neuroanatomic structure wherein the delay in the reaccumulation of osmotic osmolytes is particularly pronounced; clinically, patients with central pontine myelinolysis can present 1 or more days after overcorrection of hyponatremia with paraparesis or quadriparesis, dysphagia, dysarthria, diplopia, a "locked-in syndrome," and/or loss of consciousness. Other regions of the brain can also be involved in ODS, most commonly in association with lesions of the pons but occasionally in isolation; in order of frequency, the lesions of extrapontine myelinolysis can occur in the cerebellum, lateral geniculate body, thalamus, putamen, and cerebral cortex or subcortex. Clinical presentation of ODS can, therefore, vary as a function of the extent and localization of extrapontine myelinolysis, with the reported development of ataxia, mutism, parkinsonism, dystonia, and catatonia. Relowering of plasma Na^+ concentration after overly rapid correction can prevent or attenuate ODS (see the section on treatment of hyponatremia below). However, even appropriately slow correction can be associated with ODS, particularly in patients with additional risk factors; these include alcoholism, malnutrition, hypokalemia, and liver transplantation.

Diagnostic Approach to Hyponatremia Clinical assessment of hyponatremic patients should focus on the underlying cause; a detailed drug history is particularly crucial (Table 56-1). A careful clinical assessment of volume status is obligatory for the classical diagnostic approach to hyponatremia (Fig. 56-5). Hyponatremia is frequently multifactorial, particularly when severe; clinical evaluation should consider *all* the possible causes for excessive circulating AVP, including volume status, drugs, and the presence of nausea and/or pain. Radiologic imaging may also be appropriate to assess whether patients have a pulmonary or CNS cause for hyponatremia. A screening chest x-ray may fail to detect a small-cell carcinoma of the lung; computed tomography (CT) scanning of the thorax should be considered in patients at high risk for this tumor (e.g., patients with a smoking history).

Laboratory investigation should include a measurement of serum osmolality to exclude pseudohyponatremia, which is defined as the coexistence of hyponatremia with a normal or increased plasma tonicity. Most clinical laboratories measure plasma Na^+ concentration by testing diluted samples with automated ion-sensitive electrodes, correcting for this dilution by assuming that plasma is 93% water. This correction factor can be inaccurate in patients with pseudohyponatremia due to extreme hyperlipidemia and/or hyperproteinemia, in whom serum lipid or protein makes up a greater percentage of plasma volume. The measured osmolality should also be converted to the effective osmolality (tonicity) by subtracting the measured concentration of urea (divided by 2.8, if in mg/dL); patients with hyponatremia have an effective osmolality of <275 mOsm/kg.

Elevated BUN and creatinine in routine chemistries can also indicate renal dysfunction as a potential cause of hyponatremia, whereas hyperkalemia may suggest adrenal insufficiency or hypoaldosteronism. Serum glucose should also be measured; plasma Na^+ concentration falls by ~1.6–2.4 m*M* for every 100-mg/dL increase in glucose, due to glucose-induced water efflux from cells; this "true" hyponatremia resolves after correction of hyperglycemia. Measurement of serum uric acid should also be performed; whereas patients with SIAD-type physiology will typically be hypouricemic (serum uric acid <4 mg/dL), volume-depleted patients will often be hyperuricemic. In the appropriate clinical setting, thyroid, adrenal, and pituitary function should also be tested; hypothyroidism and secondary adrenal failure due to pituitary insufficiency are important causes of euvolemic hyponatremia, whereas primary adrenal failure causes hypovolemic hyponatremia. A cosyntropin stimulation test is necessary to assess for primary adrenal insufficiency.

Urine electrolytes and osmolality are crucial tests in the initial evaluation of hyponatremia. A urine Na^+ concentration <20–30 m*M* is consistent with hypovolemic hyponatremia, in the clinical absence of a hypervolemic, Na^+-avid syndrome such as CHF (Fig. 56-5). In contrast, patients with SIAD will typically excrete urine with an Na^+ concentration that is >30 m*M*. However, there can be substantial overlap in urine Na^+ concentration values in patients with SIAD and hypovolemic hyponatremia, particularly in the elderly; the ultimate "gold standard" for the diagnosis of hypovolemic hyponatremia is the demonstration that plasma Na^+ concentration corrects after hydration with normal saline. Patients with thiazide-associated hyponatremia may also present with higher than expected urine Na^+ concentration and other findings suggestive of SIAD; one should defer making a diagnosis of SIAD in these patients until 1–2 weeks after discontinuing the thiazide. A urine osmolality <100 mOsm/kg is suggestive of polydipsia; urine osmolality >400 mOsm/kg indicates that AVP excess is playing a more dominant role, whereas intermediate values are more consistent with multifactorial pathophysiology (e.g., AVP excess with a significant component of polydipsia). Patients with hyponatremia due to decreased solute intake (beer potomania) typically have urine Na^+ concentration <20 m*M* and urine osmolality in the range of <100 to the low 200s. Finally, the measurement of urine K^+ concentration is required to calculate the urine-to-plasma electrolyte ratio, which is useful to predict the response to fluid restriction (see the section on treatment of hyponatremia below).

TREATMENT

Hyponatremia

Three major considerations guide the therapy of hyponatremia. First, the presence and/or severity of symptoms determine the urgency and goals of therapy. Patients with acute hyponatremia (Table 56-2) present with symptoms that can range from headache, nausea, and/or vomiting, to seizures, obtundation, and central herniation; patients with chronic hyponatremia, present for >48 h, are less likely to have severe symptoms. Second, patients with chronic hyponatremia are at risk for ODS if plasma Na^+ concentration is corrected by >8–10 m*M* within the first 24 h and/or by >18 m*M* within the first 48 h. Third, the response to interventions such as hypertonic saline, isotonic saline, or AVP antagonists can be highly unpredictable, such that frequent monitoring of plasma Na^+ concentration during corrective therapy is imperative.

Once the urgency in correcting the plasma Na^+ concentration has been established and appropriate therapy instituted, the focus should be on treatment or withdrawal of the underlying cause. Patients with euvolemic hyponatremia due to SIAD, hypothyroidism, or secondary adrenal failure will respond to successful treatment of the underlying cause, with an increase in plasma Na^+ concentration. However, not all causes of SIAD are immediately reversible, necessitating pharmacologic therapy to increase the plasma Na^+ concentration (see below). Hypovolemic hyponatremia will respond to intravenous hydration with isotonic normal saline, with a rapid reduction in circulating AVP and a brisk water diuresis; it may be necessary to reduce the rate of correction if the history suggests that hyponatremia has been chronic, i.e., present for >48 h (see below). Hypervolemic hyponatremia due to CHF will often respond to improved therapy of the underlying cardiomyopathy, e.g., following the institution or intensification of angiotensin-converting enzyme (ACE) inhibition. Finally, patients with hyponatremia due to beer potomania and low solute intake will respond very rapidly to intravenous saline and the resumption of a normal diet. Notably, patients with beer potomania have a very high risk of developing ODS, due to the associated hypokalemia, alcoholism, malnutrition, and high risk of overcorrecting the plasma Na^+ concentration.

Water deprivation has long been a cornerstone of the therapy of chronic hyponatremia. However, patients who are excreting minimal electrolyte-free water will require aggressive fluid restriction; this can be very difficult for patients with SIAD to tolerate, given that their thirst is also inappropriately stimulated. The urine-to-plasma electrolyte ratio (urinary $[Na^+]$ + $[K^+]$/plasma $[Na^+]$) can be exploited as a quick indicator of electrolyte-free water excretion **(Table 56-3)**; patients with a ratio of >1 should be more aggressively restricted (<500 mL/d) if possible, those with a ratio of ~1 should be restricted to 500–700 mL/d, and those with a ratio <1 should be restricted to <1 L/d. In hypokalemic patients, potassium replacement will serve to increase plasma Na^+ concentration, given that the plasma Na^+ concentration is a function of both exchangeable Na^+ and exchangeable K^+ divided by total-body water; a corollary is that aggressive repletion of K^+ has the potential to overcorrect the plasma Na^+ concentration even in the absence of hypertonic saline. Plasma Na^+ concentration will also tend to respond to an increase in dietary solute intake, which increases the ability to excrete free water; this can be accomplished with oral salt tablets and with newly available, palatable preparations of oral urea.

Patients in whom therapy with fluid restriction, potassium replacement, and/or increased solute intake fails may merit pharmacologic therapy to increase their plasma Na^+ concentration. Some patients with SIAD initially respond to combined therapy with oral furosemide, 20 mg twice a day (higher doses may be necessary in renal insufficiency), and oral salt tablets; furosemide serves to inhibit the renal countercurrent mechanism and blunt urinary concentrating ability, whereas the salt tablets counteract diuretic-associated natriuresis. The risk of hypokalemia and/or renal dysfunction limits enthusiasm for this approach, which requires careful titration of diuretic and salt tablets. Demeclocycline is a potent inhibitor of principal cells and can be used in patients whose Na levels do not increase in response to furosemide and salt tablets. However, this agent can be associated with a reduction in GFR, due to excessive natriuresis and/or direct renal toxicity; it should be avoided in cirrhotic patients in particular, who are at higher risk of nephrotoxicity due to drug accumulation. If available, palatable preparations of oral urea can also be used to manage SIAD, with comparable efficacy to AVP antagonists (vaptans); the increase in solute excretion with oral urea ingestion increases free water excretion, thus reducing the plasma Na^+.

TABLE 56-3 Management of Hypernatremia

Water Deficit
1. Estimate total-body water (TBW): 50% of body weight in women and 60% in men
2. Calculate free-water deficit: $[(Na^+ - 140)/140] \times TBW$
3. Administer deficit over 48–72 h, without decrease in plasma Na^+ concentration by >10 mM/24 h
Ongoing Water Losses
4. Calculate free-water clearance, C_eH_2O: $C_eH_2O = V \times \left(1 - \frac{U_{Na} + U_k}{P_{Na}}\right)$ where V is urinary volume, U_{Na} is urinary $[Na^+]$, U_K is urinary $[K^+]$, and P_{Na} is plasma $[Na^+]$
Insensible Losses
5. ~10 mL/kg per day: less if ventilated, more if febrile
Total
6. Add components to determine water deficit and ongoing water loss; correct the water deficit over 48–72 h and replace daily water loss. Avoid correction of plasma $[Na^+]$ by >10 mM/d.

AVP antagonists (vaptans) are highly effective in SIAD and in hypervolemic hyponatremia due to heart failure or cirrhosis, reliably increasing plasma Na^+ concentration due to their "aquaretic" effects (augmentation of free water clearance). Most of these agents specifically antagonize the V_2 AVP receptor; tolvaptan is currently the only oral V_2 antagonist to be approved by the U.S. Food and Drug Administration. Conivaptan, the only available intravenous vaptan, is a mixed V_{1A}/V_2 antagonist, with a modest risk of hypotension due to V_{1A} receptor inhibition. Therapy with vaptans must be initiated in a hospital setting, with a liberalization of fluid restriction (>2 L/d) and close monitoring of plasma Na^+ concentration. Although approved for the management of all but hypovolemic hyponatremia and acute hyponatremia, the clinical indications are limited. Oral tolvaptan is perhaps most appropriate for the management of significant and persistent SIAD (e.g., in small-cell lung carcinoma) that has not responded to water restriction and/or oral furosemide and salt tablets. Abnormalities in liver function tests have been reported with chronic tolvaptan therapy; hence, the use of this agent should be restricted to <1–2 months.

Treatment of acute symptomatic hyponatremia should include hypertonic 3% saline (513 m*M*) to acutely increase plasma Na^+ concentration by 1–2 m*M*/h to a total of 4–6 m*M*; this modest increase is typically sufficient to alleviate severe acute symptoms, after which corrective guidelines for chronic hyponatremia are appropriate (see below). A bolus of 100 mL of hypertonic saline is more effective than an infusion, rapidly improving both serum sodium and mental status. For ongoing infusions, a number of equations have been developed to estimate the required rate of hypertonic saline, which has an Na^+-Cl^- concentration of 513 m*M*. The traditional approach is to calculate an Na^+ deficit, where the Na^+ deficit = 0.6 × body weight × (target plasma Na^+ concentration – starting plasma Na^+ concentration), followed by a calculation of the required rate. Regardless of the method used to determine the rate of administration, the increase in plasma Na^+ concentration can be highly unpredictable during treatment with hypertonic saline, due to rapid changes in the underlying physiology; plasma Na^+ concentration should be monitored every 2–4 h during treatment, with appropriate changes in therapy based on the observed rate of change. The administration of supplemental oxygen and ventilatory support is also critical in acute hyponatremia, in the event that patients develop acute pulmonary edema or hypercapneic respiratory failure. Intravenous loop diuretics will help treat acute pulmonary edema and will also increase free water excretion, by interfering with the renal countercurrent multiplication system. AVP antagonists do *not* have an approved role in the management of acute hyponatremia.

The rate of correction should be comparatively slow in *chronic* hyponatremia (<6–8 m*M* in the first 24 h and <6 m*M* each subsequent 24 h) so as to avoid ODS; lower target rates are appropriate in patients at particular risk for ODS, such as alcoholics or hypokalemic patients. Overcorrection of the plasma Na^+ concentration can occur when AVP levels rapidly normalize, for example, following the treatment of patients with chronic hypovolemic hyponatremia with intravenous saline or following glucocorticoid replacement of patients with hypopituitarism and secondary adrenal failure. Approximately 10% of patients treated with vaptans will overcorrect; the risk is increased if water intake is not liberalized. In the event that the plasma Na^+ concentration overcorrects following therapy, hyponatremia should be reinduced or stabilized by the administration of the AVP *agonist* desmopressin acetate (DDAVP) and/or the administration of free water, typically intravenous D_5W; the goal is to prevent or reverse the development of ODS. Alternatively, the treatment of patients with marked hyponatremia can be initiated with the twice-daily administration of DDAVP to maintain constant AVP bioactivity, combined with the administration of hypertonic saline to slowly correct the serum sodium in a more controlled fashion, thus reducing upfront the risk of overcorrection.

HYPERNATREMIA

Etiology Hypernatremia is defined as an increase in the plasma Na$^+$ concentration to >145 m*M*. Considerably less common than hyponatremia, hypernatremia is nonetheless associated with mortality rates of as high as 40–60%, mostly due to the severity of the associated underlying disease processes. Hypernatremia is usually the result of a combined water and electrolyte deficit, with losses of H_2O in excess of Na$^+$. Less frequently, the ingestion or iatrogenic administration of excess Na$^+$ can be causative, for example, after IV administration of excessive hypertonic Na$^+$-Cl$^-$ or Na$^+$-HCO_3^- **(Fig. 56-6)**.

Elderly individuals with reduced thirst and/or diminished access to fluids are at the highest risk of developing hypernatremia. Patients with hypernatremia may rarely have a central defect in hypothalamic osmoreceptor function, with a mixture of both decreased thirst and reduced AVP secretion. Causes of this adipsic DI include primary or metastatic tumor, occlusion or ligation of the anterior communicating artery, trauma, hydrocephalus, and inflammation such as sarcoidosis.

Hypernatremia can develop following the loss of water via both renal and nonrenal routes. Insensible losses of water may increase in the setting of fever, exercise, heat exposure, severe burns, or mechanical ventilation. Diarrhea is, in turn, the most common gastrointestinal cause of hypernatremia. Notably, osmotic diarrhea and viral gastroenteritides typically generate stools with Na$^+$ and K$^+$ <100 m*M*, thus leading to water loss and hypernatremia; in contrast, secretory diarrhea typically results in isotonic stool and thus hypovolemia with or without hypovolemic hyponatremia.

Common causes of renal water loss include osmotic diuresis secondary to hyperglycemia, excess urea, postobstructive diuresis, or mannitol; these disorders share an increase in urinary solute excretion and urinary osmolality (see "Diagnostic Approach," below). Hypernatremia due to a water diuresis occurs in central or nephrogenic DI (NDI).

NDI is characterized by renal resistance to AVP, which can be partial or complete (see "Diagnostic Approach," below). Genetic causes include loss-of-function mutations in the X-linked V_2 receptor; mutations in the AVP-responsive aquaporin-2 water channel can cause autosomal recessive and autosomal dominant NDI, whereas recessive deficiency of the aquaporin-1 water channel causes a more modest concentrating defect (Fig. 56-2). Hypercalcemia can also cause polyuria and NDI; calcium signals directly through the calcium-sensing receptor to downregulate Na$^+$, K$^+$, and Cl$^-$ transport by the TALH and water transport in principal cells, thus reducing renal concentrating ability in hypercalcemia. Another common acquired cause of NDI is hypokalemia, which inhibits the renal response to AVP and downregulates aquaporin-2 expression. Several drugs can cause acquired NDI, in particular, lithium, ifosfamide, and several antiviral agents. Lithium causes NDI by multiple mechanisms, including direct inhibition of renal glycogen synthase kinase-3 (GSK3), a kinase thought to be the pharmacologic target of lithium in bipolar disease; GSK3 is required for the response of principal cells to AVP. The entry of lithium through the amiloride-sensitive Na$^+$ channel ENaC (Fig. 56-4) is required for the effect of the drug on principal cells, such that combined therapy within lithium and amiloride can mitigate lithium-associated NDI. However, lithium causes chronic tubulointerstitial scarring and chronic kidney disease after prolonged therapy, such that patients may have a persistent NDI long after stopping the drug, with a reduced therapeutic benefit from amiloride.

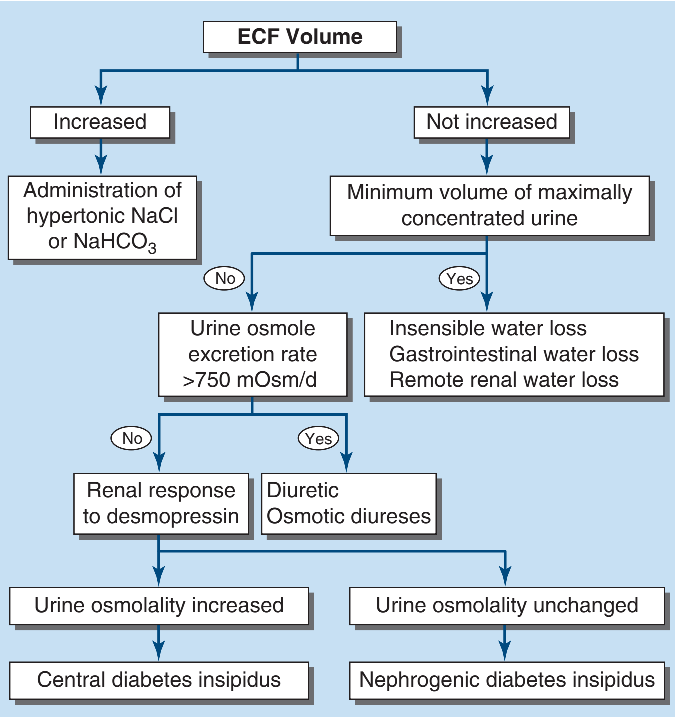
ECF Volume
Increased
Not increased
Administration of hypertonic NaCl or NaHCO3
Minimum volume of maximally concentrated urine
No
Yes
Urine osmole excretion rate >750 mOsm/d
Insensible water loss
Gastrointestinal water loss
Remote renal water loss
No
Yes
Renal response to desmopressin
Diuretic
Osmotic diureses
Urine osmolality increased
Urine osmolality unchanged
Central diabetes insipidus
Nephrogenic diabetes insipidus

FIGURE 56-6 The diagnostic approach to hypernatremia. ECF, extracellular fluid.

Finally, gestational DI is a rare complication of late-term pregnancy wherein increased activity of a circulating placental protease with "vasopressinase" activity leads to reduced circulating AVP and polyuria, often accompanied by hypernatremia. DDAVP is an effective therapy for this syndrome, given its resistance to the vasopressinase enzyme.

Clinical Features Hypernatremia increases osmolality of the ECF, generating an osmotic gradient between the ECF and ICF, an efflux of intracellular water, and cellular shrinkage. As in hyponatremia, the symptoms of hypernatremia are predominantly neurologic. Altered mental status is the most frequent manifestation, ranging from mild confusion and lethargy to deep coma. The sudden shrinkage of brain cells in acute hypernatremia may lead to parenchymal or subarachnoid hemorrhages and/or subdural hematomas; however, these vascular complications are primarily encountered in pediatric and neonatal patients. Rarely, osmotic demyelination may occur in acute hypernatremia. Osmotic damage to muscle membranes can also lead to hypernatremic rhabdomyolysis. Brain cells accommodate to a chronic increase in ECF osmolality (>48 h) by activating membrane transporters that mediate influx and intracellular accumulation of organic osmolytes (creatine, betaine, glutamate, myoinositol, and taurine); this results in an increase in ICF water and normalization of brain parenchymal volume. In consequence, patients with *chronic* hypernatremia are less likely to develop severe neurologic compromise. However, the cellular response to chronic hypernatremia predisposes pediatric patients with hypernatremia, particularly infants, to the development of cerebral edema and seizures during overly rapid hydration (overcorrection of plasma Na$^+$ concentration by >10 m*M*/d).

Diagnostic Approach The history should focus on the presence or absence of thirst, polyuria, and/or an extrarenal source for water loss, such as diarrhea. The physical examination should include a detailed neurologic exam and an assessment of the ECFV; patients with a particularly large water deficit and/or a combined deficit in electrolytes and water may be hypovolemic, with reduced JVP and orthostasis. Accurate documentation of daily fluid intake and daily urine output is also critical for the diagnosis and management of hypernatremia.

Laboratory investigation should include a measurement of serum and urine osmolality, in addition to urine electrolytes. The appropriate response to hypernatremia and a serum osmolality >295 mOsm/kg is an increase in circulating AVP and the excretion of low volumes (<500 mL/d) of maximally concentrated urine, i.e., urine with osmolality >800 mOsm/kg; should this be the case, then an extrarenal source of water loss is primarily responsible for the generation of hypernatremia. Many patients with hypernatremia are polyuric; should an osmotic diuresis be responsible, with excessive excretion of Na$^+$-Cl$^-$, glucose, and/or urea, then daily solute excretion will be >750–1000 mOsm/d (>15 mOsm/kg body water per day) (Fig. 56-6). More commonly, patients with hypernatremia and polyuria will have a predominant water diuresis, with excessive excretion of hypotonic, dilute urine.

Adequate differentiation between nephrogenic and central causes of DI requires the measurement of the response in urinary osmolality to DDAVP, combined with measurement of circulating AVP in the setting of hypertonicity. If measurement of serum copeptin is available, an "indirect water deprivation" test can be performed in patients with hypotonic polyuria without hypernatremia; if an infusion of hypertonic saline increases the level of circulating copeptin, a peptide co-secreted with AVP, then the patient suffers from polydipsia rather than central DI. By definition, patients with baseline hypernatremia are hypertonic, with an adequate stimulus for AVP by the posterior pituitary. Therefore, in contrast to polyuric patients with a normal or reduced baseline plasma Na^+ concentration and osmolality, a water deprivation test **(Chap. 55)** is unnecessary in hypernatremia; indeed, water deprivation is absolutely contraindicated in this setting, given the risk for worsening the hypernatremia. Hypernatremic patients with NDI will have high serum levels of AVP and copeptin. Their low urine osmolality will also fail to respond to DDAVP, increasing by <50% or <150 mOsm/kg from baseline; patients with central DI will respond to DDAVP, with a reduced circulating AVP and copeptin. Patients may exhibit a partial response to DDAVP, with a >50% rise in urine osmolality that nonetheless fails to reach 800 mOsm/kg; the level of circulating AVP will help differentiate the underlying cause, i.e., NDI versus central DI. In pregnant patients, AVP assays should be drawn in tubes containing the protease inhibitor 1,10-phenanthroline to prevent in vitro degradation of AVP by placental vasopressinase.

For patients with hypernatremia due to renal loss of water, it is critical to quantify *ongoing* daily losses using the calculated electrolyte-free water clearance, in addition to calculation of the baseline water deficit (the relevant formulas are discussed in Table 56-3). This requires daily measurement of urine electrolytes, combined with accurate measurement of daily urine volume.

TREATMENT

Hypernatremia

The underlying cause of hypernatremia should be withdrawn or corrected, be it drugs, hyperglycemia, hypercalcemia, hypokalemia, or diarrhea. The approach to the correction of hypernatremia is outlined in Table 56-3. It is imperative to correct hypernatremia slowly to avoid cerebral edema, typically replacing the calculated free water deficit over 48 h. Ideally, the plasma Na^+ concentration should be corrected by no more than 10 m*M*/d, which may take longer than 48 h in patients with severe hypernatremia (>160 m*M*). In critically ill adults, however, recent evidence does not indicate that rapid correction of hypernatremia is associated with a higher risk for mortality, seizure, alteration of consciousness, and/or cerebral edema. Given that restricting the rate of correction to <10 m*M*/d has no physiologic sequelae, it seems prudent to restrict correction in adults to this rate; however, should that rate be exceeded, hypernatremia does not need to be reinduced.

Water should ideally be administered by mouth or by nasogastric tube, as the most direct way to provide free water, i.e., water without electrolytes. Alternatively, patients can receive free water in dextrose-containing IV solutions, such as 5% dextrose (D_5W); blood glucose should be monitored in case hyperglycemia occurs. Depending on the history, blood pressure, or clinical volume status, it may be appropriate to initially treat with hypotonic saline solutions (1/4 or 1/2 normal saline); normal saline is usually inappropriate in the absence of very severe hypernatremia, where normal saline is proportionally more hypotonic relative to plasma, or frank hypotension. Calculation of urinary electrolyte-free water clearance (Table 56-3) is required to estimate daily, ongoing loss of free water in patients with NDI or central DI, which should be replenished daily.

Additional therapy may be feasible in specific cases. Patients with central DI should respond to the administration of intravenous, intranasal, or oral DDAVP. Patients with NDI due to lithium may reduce their polyuria with amiloride (2.5–10 mg/d), which decreases entry of lithium into principal cells by inhibiting ENaC (see above); in practice, however, most patients with lithium-associated DI are able to compensate for their polyuria by simply increasing their daily water intake. Thiazides may reduce polyuria due to NDI, ostensibly by inducing hypovolemia and increasing proximal tubular water reabsorption. Occasionally, nonsteroidal anti-inflammatory drugs (NSAIDs) have been used to treat polyuria associated with NDI, reducing the negative effect of intrarenal prostaglandins on urinary concentrating mechanisms; however, this assumes the risks of NSAID-associated gastric and/or renal toxicity. Furthermore, it must be emphasized that thiazides, amiloride, and NSAIDs are only appropriate for *chronic* management of polyuria from NDI and have *no* role in the acute management of associated hypernatremia, where the focus is on replacing free water deficits and ongoing free water loss.

POTASSIUM DISORDERS

Homeostatic mechanisms maintain plasma K^+ concentration between 3.5 and 5.0 m*M*, despite marked variation in dietary K^+ intake. In a healthy individual at steady state, the entire daily intake of potassium is excreted, ~90% in the urine and 10% in the stool; thus, the kidney plays a dominant role in potassium homeostasis. However, >98% of total-body potassium is intracellular, chiefly in muscle; buffering of extracellular K^+ by this large intracellular pool plays a crucial role in the regulation of plasma K^+ concentration. Changes in the exchange and distribution of intra- and extracellular K^+ can thus lead to marked hypo- or hyperkalemia. A corollary is that massive necrosis and the attendant release of tissue K^+ can cause severe hyperkalemia, particularly in the setting of acute kidney injury and reduced excretion of K^+.

Changes in whole-body K^+ content are primarily mediated by the kidney, which *reabsorbs* filtered K^+ in hypokalemic, K^+-deficient states and *secretes* K^+ in hyperkalemic, K^+-replete states. Although K^+ is transported along the entire nephron, it is the principal cells of the connecting segment (CNT) and cortical CD that play a dominant role in renal K^+ secretion, whereas alpha-intercalated cells of the outer medullary CD function in renal tubular reabsorption of filtered K^+ in K^+-deficient states. In principal cells, apical Na^+ entry via the amiloride-sensitive ENaC generates a lumen-negative potential difference, which drives passive K^+ exit through apical K^+ channels (Fig. 56-4). Two major K^+ channels mediate distal tubular K^+ secretion: the secretory K^+ channel ROMK (renal outer medullary K^+ channel; also known as Kir1.1 or KcnJ1) and the flow-sensitive "big potassium" (BK) or maxi-K K^+ channel. ROMK is thought to mediate the bulk of constitutive K^+ secretion, whereas increases in distal flow rate and/or genetic absence of ROMK activate K^+ secretion via the BK channel.

An appreciation of the relationship between ENaC-dependent Na^+ entry and distal K^+ secretion (Fig. 56-4) is required for the bedside interpretation of potassium disorders. For example, decreased distal delivery of Na^+, as occurs in hypovolemic, prerenal states, tends to blunt the ability to excrete K^+, leading to hyperkalemia; on the other hand, an *increase* in distal delivery of Na^+ and distal flow rate, as occurs after treatment with thiazide and loop diuretics, can enhance K^+ secretion and lead to hypokalemia. Hyperkalemia is also a predictable consequence of drugs that directly inhibit ENaC, due to the role of this Na^+ channel in generating a lumen-negative potential difference. Aldosterone in turn has a major influence on potassium excretion, increasing the activity of ENaC channels and thus amplifying the driving force for K^+ secretion across the luminal membrane of principal cells. Abnormalities in the renin-angiotensin-aldosterone system can thus cause both hypokalemia and hyperkalemia. Notably, however, potassium excess and potassium restriction have opposing, aldosterone-independent effects on the density and activity of apical K^+ channels in the distal nephron, i.e., factors other than aldosterone modulate the renal capacity to secrete K^+. In addition, potassium restriction and hypokalemia activate aldosterone-independent distal *reabsorption* of filtered K$^+$, activating apical H$^+$/K$^+$-ATPase activity in intercalated cells within the outer medullary CD. Reflective perhaps of this physiology, changes in plasma K$^+$ concentration are not universal in disorders associated with changes in aldosterone activity.

■ HYPOKALEMIA

Hypokalemia, defined as a plasma K$^+$ concentration of <3.5 m*M*, occurs in up to 20% of hospitalized patients. Hypokalemia is associated with a tenfold increase in in-hospital mortality, due to adverse effects on cardiac rhythm, blood pressure, and cardiovascular morbidity. Mechanistically, hypokalemia can be caused by redistribution of K$^+$ between tissues and the ECF or by renal and nonrenal loss of K$^+$ **(Table 56-4)**. Systemic hypomagnesemia can also cause treatment-resistant hypokalemia, due to a combination of reduced cellular uptake of K$^+$ and exaggerated renal secretion. Spurious hypokalemia or "pseudohypokalemia" can occasionally result from in vitro cellular uptake of K$^+$ after venipuncture, for example, due to profound leukocytosis in acute leukemia.

TABLE 56-4 Causes of Hypokalemia

I. Decreased intake
  A. Starvation
  B. Clay ingestion
II. Redistribution into cells
  A. Acid-base
    1. Metabolic alkalosis
  B. Hormonal
    1. Insulin
    2. Increased β$_2$-adrenergic sympathetic activity: post–myocardial infarction, head injury
    3. β$_2$-Adrenergic agonists—bronchodilators, tocolytics
    4. α-Adrenergic antagonists
    5. Thyrotoxic periodic paralysis
    6. Downstream stimulation of Na$^+$/K$^+$-ATPase: theophylline, caffeine
  C. Anabolic state
    1. Vitamin B$_{12}$ or folic acid administration (red blood cell production)
    2. Granulocyte-macrophage colony-stimulating factor (white blood cell production)
    3. Total parenteral nutrition
  D. Other
    1. Pseudohypokalemia
    2. Hypothermia
    3. Familial hypokalemic periodic paralysis
    4. Barium toxicity: systemic inhibition of "leak" K$^+$ channels
III. Increased loss
  A. Nonrenal
    1. Gastrointestinal loss (diarrhea)
    2. Integumentary loss (sweat)
  B. Renal
    1. Increased distal flow and distal Na+ delivery: diuretics, osmotic diuresis, salt-wasting nephropathies
    2. Increased secretion of potassium
      a. Mineralocorticoid excess: primary hyperaldosteronism (aldosterone-producing adenomas, primary or unilateral adrenal hyperplasia, idiopathic hyperaldosteronism due to bilateral adrenal hyperplasia, and adrenal carcinoma), genetic hyperaldosteronism (familial hyperaldosteronism types I/II/III, congenital adrenal hyperplasias), secondary hyperaldosteronism (malignant hypertension, renin-secreting tumors, renal artery stenosis, hypovolemia), Cushing's syndrome, Bartter's syndrome, Gitelman's syndrome
      b. Apparent mineralocorticoid excess: genetic deficiency of 11β-dehydrogenase-2 (syndrome of apparent mineralocorticoid excess), inhibition of 11β-dehydrogenase-2 (glycyrrhetinic/glycyrrhizinic acid and/or carbenoxolone; itraconazole and posaconazole; licorice, food products, drugs), Liddle's syndrome (genetic activation of epithelial Na$^+$ channels)
      c. Distal delivery of nonreabsorbed anions: vomiting, nasogastric suction, proximal renal tubular acidosis, diabetic ketoacidosis, glue-sniffing (toluene abuse), penicillin derivatives (penicillin, nafcillin, dicloxacillin, ticarcillin, oxacillin, and carbenicillin)
    3. Magnesium deficiency

Redistribution and Hypokalemia Insulin, β$_2$-adrenergic activity, thyroid hormone, and alkalosis promote Na$^+$/K$^+$-ATPase-mediated cellular uptake of K$^+$, leading to hypokalemia. Inhibition of the passive *efflux* of K$^+$ can also cause hypokalemia, albeit rarely; this typically occurs in the setting of systemic inhibition of K$^+$ channels by toxic barium ions. Exogenous insulin can cause iatrogenic hypokalemia, particularly during the management of K$^+$-deficient states such as diabetic ketoacidosis. Alternatively, the stimulation of *endogenous* insulin can provoke hypokalemia, hypomagnesemia, and/or hypophosphatemia in malnourished patients given a carbohydrate load. Alterations in the activity of the endogenous sympathetic nervous system can cause hypokalemia in several settings, including alcohol withdrawal, hyperthyroidism, acute myocardial infarction, and severe head injury. β$_2$ agonists, including both bronchodilators and tocolytics (ritodrine), are powerful activators of cellular K$^+$ uptake; "hidden" sympathomimetics, such as pseudoephedrine and ephedrine in cough syrup or dieting agents, may also cause unexpected hypokalemia. Finally, xanthine-dependent activation of cAMP-dependent signaling, downstream of the β$_2$ receptor, can lead to hypokalemia, usually in the setting of overdose (theophylline) or marked overingestion (dietary caffeine).

Redistributive hypokalemia can also occur in the setting of hyperthyroidism, with periodic attacks of hypokalemic paralysis (thyrotoxic periodic paralysis [TPP]). Similar episodes of hypokalemic weakness in the absence of thyroid abnormalities occur in *familial* hypokalemic periodic paralysis, usually caused by missense mutations of voltage sensor domains within the α$_1$ subunit of L-type calcium channels or the skeletal Na$^+$ channel; these mutations generate an abnormal gating pore current activated by hyperpolarization. TPP develops more frequently in patients of Asian or Latin American origin; this shared predisposition has been linked to genetic variation in Kir2.6, a muscle-specific, thyroid hormone–responsive K$^+$ channel. Genome-wide association studies have also implicated variation in the *KCNJ2* gene, which encodes a related muscle K$^+$ channel, Kir 2.1, in predisposition to TPP. Patients with TPP typically present with weakness of the extremities and limb girdles, with paralytic episodes that occur most frequently between 1 and 6 A.M. Signs and symptoms of hyperthyroidism are not invariably present. Hypokalemia is usually profound and almost invariably accompanied by hypophosphatemia and hypomagnesemia. The hypokalemia in TPP is also attributed to both direct and indirect activation of the Na$^+$/K$^+$-ATPase, resulting in increased uptake of K$^+$ by muscle and other tissues. Increases in β-adrenergic activity play an important role in that high-dose propranolol (3 mg/kg) rapidly reverses the associated hypokalemia, hypophosphatemia, and paralysis. Outward-directed inward-rectifying K$^+$ current, mediated by KIR channels (primarily Kir2.1 and Kir2.2 tetramers), is also reduced in skeletal muscles of patients with TPP, providing an additional mechanism for hypokalemia. Together with increased Na$^+$/K$^+$-ATPase activity and increased circulating insulin, this reduced KIR current may trigger a "feedforward" cycle of hypokalemia leading to inactivation of muscle Na$^+$ channels, paradoxical depolarization, and paralysis.

Nonrenal Loss of Potassium The loss of K$^+$ in sweat is typically low, except under extremes of physical exertion. Direct gastric losses of K$^+$ due to vomiting or nasogastric suctioning are also minimal; however, the ensuing hypochloremic alkalosis results in persistent kaliuresis due to secondary hyperaldosteronism and bicarbonaturia, i.e., a *renal* loss of K$^+$. Diarrhea is a globally important cause of hypokalemia, given the worldwide prevalence of infectious diarrheal disease. Noninfectious gastrointestinal processes such as celiac disease, ileostomy, villous adenomas, inflammatory bowel disease, colonic pseudo-obstruction (Ogilvie's syndrome), VIPomas, and chronic laxative abuse can also cause significant hypokalemia; an exaggerated intestinal secretion of potassium by upregulated colonic BK channels has been directly implicated in the pathogenesis of hypokalemia in some of these disorders.

Renal Loss of Potassium Drugs can increase renal K^+ excretion by a variety of different mechanisms. Diuretics are a particularly common cause, due to associated increases in distal tubular Na^+ delivery and distal tubular flow rate, in addition to secondary hyperaldosteronism. Thiazides have a greater effect on plasma K^+ concentration than loop diuretics, despite their lesser natriuretic effect. The diuretic effect of thiazides is largely due to inhibition of the Na^+-Cl^- cotransporter NCC in DCT cells. This leads to a direct increase in the delivery of luminal Na^+ to the principal cells immediately downstream in the CNT and cortical CD, which augments Na^+ entry via ENaC, increases the lumen-negative potential difference, and amplifies K^+ secretion. The higher propensity of thiazides to cause hypokalemia may also be secondary to thiazide-associated hypocalciuria, versus the *hypercalciuria* seen with loop diuretics; the increases in downstream luminal calcium in response to loop diuretics inhibit ENaC in principal cells, thus reducing the lumen-negative potential difference and attenuating distal K^+ excretion. High doses of penicillin-related antibiotics (nafcillin, dicloxacillin, ticarcillin, oxacillin, and carbenicillin) can increase obligatory K^+ excretion by acting as nonreabsorbable anions in the distal nephron. Finally, several renal tubular toxins cause renal K^+ and magnesium wasting, leading to hypokalemia and hypomagnesemia; these drugs include aminoglycosides, amphotericin, foscarnet, cisplatin, and ifosfamide (see also "Magnesium Deficiency and Hypokalemia," below).

Aldosterone activates the ENaC channel in principal cells via multiple synergistic mechanisms, thus increasing the driving force for K^+ excretion. In consequence, increases in aldosterone bioactivity and/or gains in function of aldosterone-dependent signaling pathways are associated with hypokalemia. Increases in circulating aldosterone (hyperaldosteronism) may be primary or secondary. Increased levels of circulating renin in secondary forms of hyperaldosteronism lead to increased angiotensin II and thus aldosterone; renal artery stenosis is perhaps the most frequent cause (Table 56-4). Primary hyperaldosteronism may be genetic or acquired. Hypertension and hypokalemia, due to increases in circulating 11-deoxycorticosterone, occur in patients with congenital adrenal hyperplasia caused by defects in either steroid 11β-hydroxylase or steroid 17α-hydroxylase; deficient 11β-hydroxylase results in associated virilization and other signs of androgen excess, whereas reduced sex steroids in 17α-hydroxylase deficiency lead to hypogonadism.

The major forms of *isolated* primary genetic hyperaldosteronism are familial hyperaldosteronism type I (FH-I, also known as glucocorticoid-remediable hyperaldosteronism [GRA]) and familial hyperaldosteronism types II and III (FH-II and FH-III), in which aldosterone production is not repressible by exogenous glucocorticoids. FH-I is caused by a chimeric gene duplication between the homologous 11β-hydroxylase (*CYP11B1*) and aldosterone synthase (*CYP11B2*) genes, fusing the adrenocorticotropic hormone (ACTH)–responsive 11β-hydroxylase promoter to the coding region of aldosterone synthase; this chimeric gene is under the control of ACTH and thus repressible by glucocorticoids. FH-III is caused by mutations in the *KCNJ5* gene, which encodes the G protein–activated inward rectifier K^+ channel 4 (GIRK4); these mutations lead to the acquisition of sodium permeability in the mutant GIRK4 channels, causing an exaggerated membrane depolarization in adrenal glomerulosa cells and the activation of voltage-gated calcium channels. The resulting calcium influx is sufficient to produce aldosterone secretion and cell proliferation, leading to adrenal adenomas and hyperaldosteronism.

Acquired causes of primary hyperaldosteronism include aldosterone-producing adenomas (APAs), primary or unilateral adrenal hyperplasia (PAH), idiopathic hyperaldosteronism (IHA) due to bilateral adrenal hyperplasia, and adrenal carcinoma; APA and IHA account for close to 60% and 40%, respectively, of diagnosed hyperaldosteronism. Acquired somatic mutations in *KCNJ5* or less frequently in the *ATP1A1* (an Na^+/K^+ ATPase α subunit) and *ATP2B3* (a Ca^{2+} ATPase) genes can be detected in APAs; as in FH-III (see above), the exaggerated depolarization of adrenal glomerulosa cells caused by these mutations is implicated in the excessive adrenal proliferation and the exaggerated release of aldosterone.

Random testing of plasma renin activity (PRA) and aldosterone is a helpful screening tool in hypokalemic and/or hypertensive patients, with an aldosterone:PRA ratio of >50 suggestive of primary hyperaldosteronism. Hypokalemia and multiple antihypertensive drugs may alter the aldosterone:PRA ratio by suppressing aldosterone or increasing PRA, leading to a ratio of <50 in patients who do in fact have primary hyperaldosteronism; therefore, the clinical context should always be considered when interpreting these results. Additionally, drugs that modulate the renin-angiotensin-aldosterone axis should be stopped during workup of hyperaldosteronism; the only antihypertensives that do not interfere with the workup of hyperaldosteronism are verapamil, alpha blockers, hydralazine, and low-dose amiloride.

The glucocorticoid cortisol has equal affinity for the MLR to that of aldosterone, with resultant "mineralocorticoid-like" activity. However, cells in the aldosterone-sensitive distal nephron are protected from this "illicit" activation by the enzyme 11β-hydroxysteroid dehydrogenase-2 (11βHSD-2), which converts cortisol to cortisone; cortisone has minimal affinity for the MLR. Recessive loss-of-function mutations in the *11βHSD-2* gene are thus associated with cortisol-dependent activation of the MLR and the syndrome of apparent mineralocorticoid excess (SAME), encompassing hypertension, hypokalemia, hypercalciuria, and metabolic alkalosis, with suppressed PRA and suppressed aldosterone. A similar syndrome is caused by biochemical inhibition of 11βHSD-2 by glycyrrhetinic/glycyrrhizinic acid and/or carbenoxolone. Glycyrrhizinic acid is a natural sweetener found in licorice root, typically encountered in licorice and its many guises or as a flavoring agent in tobacco and food products. More recently, the antifungals itraconazole and posaconazole have been shown to inhibit 11βHSD-2, leading to hypertension and hypokalemia.

Finally, hypokalemia may also occur with systemic increases in glucocorticoids. In Cushing's syndrome caused by increases in pituitary ACTH **(Chap. 398)**, the incidence of hypokalemia is only 10%, whereas it is 60–100% in patients with ectopic secretion of ACTH, despite a similar incidence of hypertension. Indirect evidence suggests that the activity of renal 11βHSD-2 is reduced in patients with ectopic ACTH compared with Cushing's syndrome, resulting in SAME.

Finally, defects in multiple renal tubular transport pathways are associated with hypokalemia. For example, loss-of-function mutations in subunits of the acidifying H^+-ATPase in alpha-intercalated cells cause hypokalemic distal renal tubular acidosis, as do many acquired disorders of the distal nephron. Liddle's syndrome is caused by autosomal dominant gain-in-function mutations of ENaC subunits. Disease-associated mutations either activate the channel directly or abrogate aldosterone-inhibited retrieval of ENaC subunits from the plasma membrane; the end result is increased expression of activated ENaC channels at the plasma membrane of principal cells. Patients with Liddle's syndrome classically manifest severe hypertension with hypokalemia, unresponsive to spironolactone yet sensitive to amiloride. Hypertension and hypokalemia are, however, variable aspects of the Liddle's phenotype; more consistent features include a blunted aldosterone response to ACTH and reduced urinary aldosterone excretion.

Loss of the transport functions of the TALH and DCT nephron segments causes hereditary hypokalemic alkalosis and Bartter's syndrome (BS) and Gitelman's syndrome (GS), respectively. Patients with classic BS typically suffer from polyuria and polydipsia, due to the reduction in renal concentrating ability. They may have an increase in urinary calcium excretion, and 20% are hypomagnesemic. Other features include marked activation of the renin-angiotensin-aldosterone axis. Patients with antenatal BS suffer from a severe systemic disorder characterized by marked electrolyte wasting, polyhydramnios, and hypercalciuria with nephrocalcinosis; renal prostaglandin synthesis and excretion are significantly increased, accounting for much of the systemic symptoms. There are five disease genes for BS, all of them functioning in some aspect of regulated Na^+, K^+, and Cl^- transport by the TALH. In contrast, GS is genetically homogeneous, caused almost exclusively by loss-of-function mutations in the thiazide-sensitive Na^+-Cl^- cotransporter of the DCT. Patients with GS are uniformly hypomagnesemic and exhibit marked hypocalciuria, rather than the hypercalciuria typically seen in

BS; urinary calcium excretion is thus a critical diagnostic test in GS. GS is a milder phenotype than BS; however, patients with GS may suffer from chondrocalcinosis, an abnormal deposition of calcium pyrophosphate dihydrate (CPPD) in joint cartilage (Chap. 327).

Magnesium Deficiency and Hypokalemia Magnesium depletion has inhibitory effects on muscle Na^+/K^+-ATPase activity, reducing influx into muscle cells and causing a secondary kaliuresis. In addition, magnesium depletion causes exaggerated K^+ secretion by the distal nephron; this effect is attributed to a reduction in the magnesium-dependent, intracellular block of K^+ efflux through the secretory K^+ channel of principal cells (ROMK; Fig. 56-4). In consequence, hypomagnesemic patients are clinically refractory to K^+ replacement in the absence of Mg^{2+} repletion. Notably, magnesium deficiency is also a common concomitant of hypokalemia because many disorders of the distal nephron may cause both potassium and magnesium wasting (Chap. 327).

Clinical Features Hypokalemia has prominent effects on cardiac, skeletal, and intestinal muscle cells. In particular, hypokalemia is a major risk factor for both ventricular and atrial arrhythmias. Hypokalemia predisposes to digoxin toxicity by a number of mechanisms, including reduced competition between K^+ and digoxin for shared binding sites on cardiac Na^+/K^+-ATPase subunits. Electrocardiographic changes in hypokalemia include broad flat T waves, ST depression, and QT prolongation; these are most marked when serum K^+ is <2.7 mmol/L. Hypokalemia can thus be an important precipitant of arrhythmia in patients with additional genetic or acquired causes of QT prolongation. Hypokalemia also results in hyperpolarization of skeletal muscle, thus impairing the capacity to depolarize and contract; weakness and even paralysis may ensue. It also causes a skeletal myopathy and predisposes to rhabdomyolysis. Finally, the paralytic effects of hypokalemia on intestinal smooth muscle may cause intestinal ileus.

The functional effects of hypokalemia on the kidney can include Na^+-Cl^- and HCO_3^- retention, polyuria, phosphaturia, hypocitraturia, and an activation of renal ammoniagenesis. Bicarbonate retention and other acid-base effects of hypokalemia can contribute to the generation of metabolic alkalosis. Hypokalemic polyuria is due to a combination of central polydipsia and an AVP-resistant renal concentrating defect. Structural changes in the kidney due to hypokalemia include a relatively specific vacuolizing injury to proximal tubular cells, interstitial nephritis, and renal cysts. Hypokalemia also predisposes to acute kidney injury and can lead to end-stage renal disease (ESRD) in patients with long-standing hypokalemia due to eating disorders and/or laxative abuse.

Hypokalemia and/or reduced dietary K^+ are implicated in the pathophysiology and progression of hypertension, heart failure, vascular disease, and stroke. For example, short-term K^+ restriction in healthy humans and patients with essential hypertension induces Na^+-Cl^- retention and hypertension. Correction of hypokalemia is particularly important in hypertensive patients treated with diuretics, in whom blood pressure improves with potassium supplementation and the establishment of normokalemia.

Diagnostic Approach The cause of hypokalemia is usually evident from history, physical examination, and/or basic laboratory tests. The history should focus on medications (e.g., laxatives, diuretics, antibiotics), diet and dietary habits (e.g., licorice), and/or symptoms that suggest a particular cause (e.g., periodic weakness, diarrhea). The physical examination should pay particular attention to blood pressure, volume status, and signs suggestive of specific hypokalemic disorders, e.g., hyperthyroidism and Cushing's syndrome. Initial laboratory evaluation should include electrolytes, BUN, creatinine, serum osmolality, Mg^{2+}, Ca^{2+}, a complete blood count, and urinary pH, osmolality, creatinine, and electrolytes (Fig. 56-7). The presence of a non–anion gap acidosis suggests a distal, hypokalemic renal tubular acidosis or diarrhea; calculation of the urinary anion gap can help differentiate these two diagnoses. Renal K^+ excretion can be assessed with a 24-h urine collection; a 24-h K^+ excretion of <15 mmol is indicative of an extrarenal cause of hypokalemia (Fig. 56-7). If only a random, spot urine sample is available, serum and urine osmolality can be used to calculate the transtubular K^+ gradient (TTKG), which should be <3 in the presence of hypokalemia (see also "Hyperkalemia"). Alternatively, a urinary K^+-to-creatinine ratio of >13 mmol/g creatinine (>1.5 mmol/mmol creatinine) is compatible with excessive renal K^+ excretion. Urine Cl^- is usually decreased in patients with hypokalemia from a nonreabsorbable anion, such as antibiotics or HCO_3^-. The most common causes of chronic hypokalemic alkalosis are surreptitious vomiting, diuretic abuse, and GS; these can be distinguished by the pattern of urinary electrolytes. Hypokalemic patients with vomiting due to bulimia will thus typically have a urinary Cl^- <10 mmol/L; urine Na^+, K^+, and Cl^- are persistently elevated in GS, due to loss of function in the thiazide-sensitive Na^+-Cl^- cotransporter, but less elevated in diuretic abuse and with greater variability. Urine diuretic screens for loop diuretics and thiazides may be necessary to further exclude diuretic abuse.

Other tests, such as urinary Ca^{2+}, thyroid function tests, and/or PRA and aldosterone levels, may also be appropriate in specific cases. A plasma aldosterone:PRA ratio of >50, due to suppression of circulating renin and an elevation of circulating aldosterone, is suggestive of hyperaldosteronism. Patients with hyperaldosteronism or apparent mineralocorticoid excess may require further testing, for example, adrenal vein sampling (Chap. 398) or the clinically available testing for specific genetic causes (e.g., FH-I, SAME, Liddle's syndrome). Patients with primary aldosteronism should thus be tested for the chimeric FH-I/GRA gene (see above) if they are younger than 20 years of age or have a family history of primary aldosteronism or stroke at a young age (<40 years). Preliminary differentiation of Liddle's syndrome due to mutant ENaC channels from SAME due to mutant 11βHSD-2 (see above), both of which cause hypokalemia and hypertension with aldosterone suppression, can be made on a clinical basis and then confirmed by genetic analysis; patients with Liddle's syndrome should respond to amiloride (ENaC inhibition) but not spironolactone, whereas patients with SAME will respond to spironolactone.

TREATMENT
Hypokalemia

The goals of therapy in hypokalemia are to prevent life-threatening and/or serious chronic consequences, to replace the associated K^+ deficit, and to correct the underlying cause and/or mitigate future hypokalemia. The urgency of therapy depends on the severity of hypokalemia, associated clinical factors (e.g., cardiac disease, digoxin therapy), and the rate of decline in serum K^+. Patients with a prolonged QT interval and/or other risk factors for arrhythmia should be monitored by continuous cardiac telemetry during repletion. Urgent but cautious K^+ replacement should be considered in patients with severe redistributive hypokalemia (plasma K^+ concentration <2.5 m*M*) and/or when serious complications ensue; however, this approach has a risk of rebound hyperkalemia following acute resolution of the underlying cause. When excessive activity of the sympathetic nervous system is thought to play a dominant role in redistributive hypokalemia, as in TPP, theophylline overdose, and acute head injury, high-dose propranolol (3 mg/kg) should be considered; this nonspecific β-adrenergic blocker will correct hypokalemia without the risk of rebound hyperkalemia.

Oral replacement with K^+-Cl^- is the mainstay of therapy in hypokalemia. Potassium phosphate, oral or IV, may be appropriate in patients with combined hypokalemia and hypophosphatemia. Potassium bicarbonate or potassium citrate should be considered in patients with concomitant metabolic acidosis. Notably, hypomagnesemic patients are refractory to K^+ replacement alone, such that concomitant Mg^{2+} deficiency should *always* be corrected with oral or intravenous repletion. The deficit of K^+ and the rate of correction should be estimated as accurately as possible; renal function, medications, and comorbid conditions such as diabetes should also be considered, so as to gauge the risk of overcorrection. In the absence of abnormal K^+ redistribution, the total deficit correlates with serum K^+, such that serum K^+ drops by ~0.27 m*M* for every

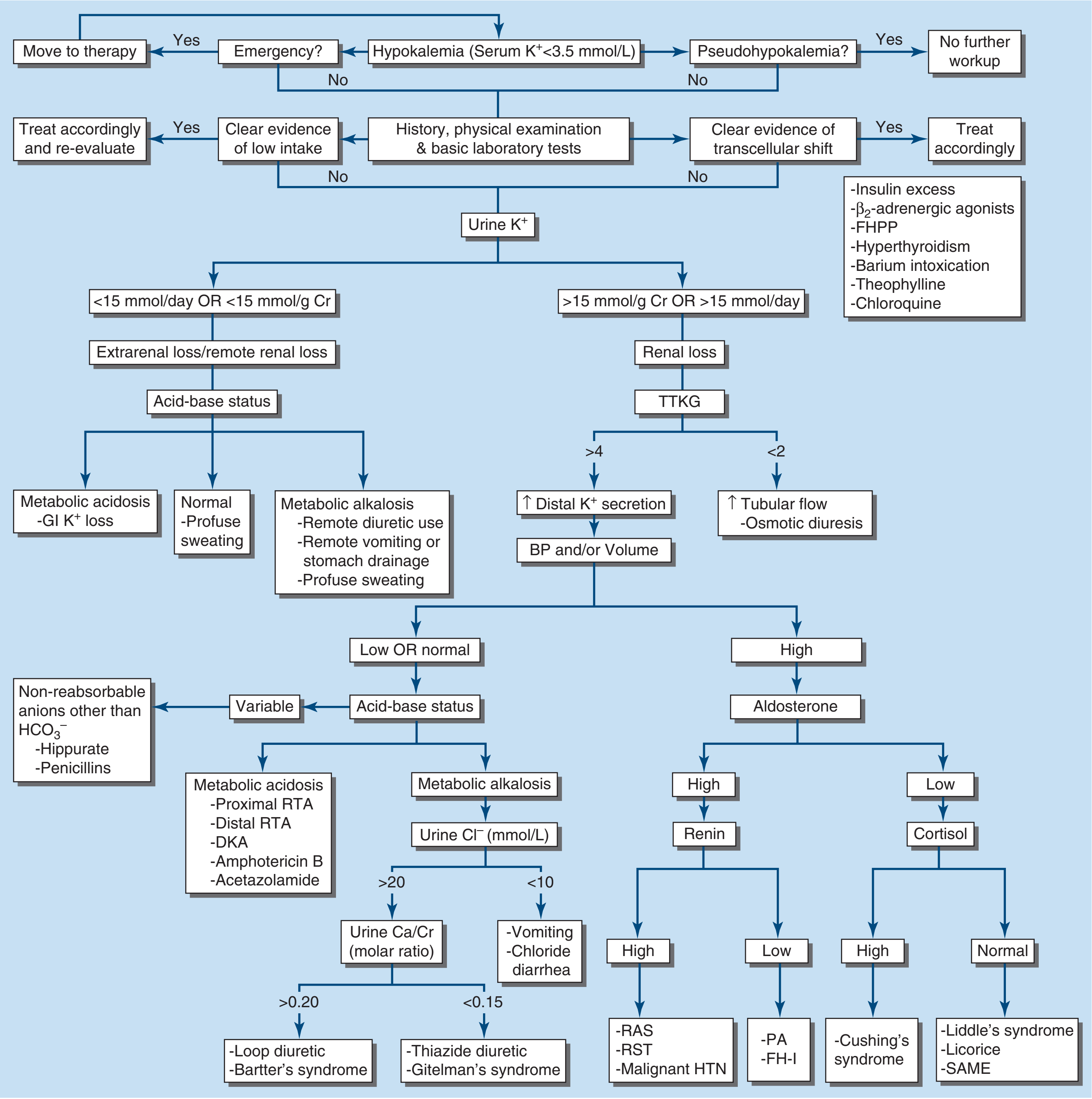
Hypokalemia (Serum K+<3.5 mmol/L)
Emergency?
Yes
Move to therapy
No
Pseudohypokalemia?
Yes
No further workup
No
History, physical examination & basic laboratory tests
Clear evidence of low intake
Yes
Treat accordingly and re-evaluate
No
Clear evidence of transcellular shift
Yes
Treat accordingly
No
-Insulin excess
-β2-adrenergic agonists
-FHPP
-Hyperthyroidism
-Barium intoxication
-Theophylline
-Chloroquine
Urine K+
<15 mmol/day OR <15 mmol/g Cr
Extrarenal loss/remote renal loss
Acid-base status
Metabolic acidosis -GI K+ loss
Normal -Profuse sweating
Metabolic alkalosis -Remote diuretic use -Remote vomiting or stomach drainage -Profuse sweating
>15 mmol/g Cr OR >15 mmol/day
Renal loss
TTKG
>4
↑ Distal K+ secretion
BP and/or Volume
<2
↑ Tubular flow -Osmotic diuresis
Low OR normal
Acid-base status
Variable
Non-reabsorbable anions other than HCO3− -Hippurate -Penicillins
Metabolic acidosis -Proximal RTA -Distal RTA -DKA -Amphotericin B -Acetazolamide
Metabolic alkalosis
Urine Cl− (mmol/L)
>20
Urine Ca/Cr (molar ratio)
>0.20
-Loop diuretic -Bartter's syndrome
<0.15
-Thiazide diuretic -Gitelman's syndrome
<10
-Vomiting -Chloride diarrhea
High
Aldosterone
High
Renin
High
-RAS -RST -Malignant HTN
Low
-PA -FH-I
Low
Cortisol
High
-Cushing's syndrome
Normal
-Liddle's syndrome -Licorice -SAME

FIGURE 56-7 The diagnostic approach to hypokalemia. See text for details. AME, apparent mineralocorticoid excess; BP, blood pressure; CCD, cortical collecting duct; DKA, diabetic ketoacidosis; FH-I, familial hyperaldosteronism type I; FHPP, familial hypokalemic periodic paralysis; GI, gastrointestinal; GRA, glucocorticoid remediable aldosteronism; HTN, hypertension; PA, primary aldosteronism; RAS, renal artery stenosis; RST, renin-secreting tumor; RTA, renal tubular acidosis; SAME, syndrome of apparent mineralocorticoid excess; TTKG, transtubular potassium gradient. *(Reproduced with permission from DB Mount, K Zandi-Nejad: Disorders of potassium balance, in BM Brenner [ed], Brenner and Rector's The Kidney, 8th ed, Philadelphia, W.B. Saunders & Company, 2008.)*

100-mmol reduction in total-body stores; loss of 400–800 mmol of total-body K$^+$ results in a reduction in serum K$^+$ by ~2.0 m*M*. Notably, given the delay in redistributing potassium into intracellular compartments, this deficit must be replaced gradually over 24–48 h, with frequent monitoring of plasma K$^+$ concentration to avoid transient overrepletion and transient hyperkalemia.

The use of intravenous administration should be limited to patients unable to use the enteral route or in the setting of severe complications (e.g., paralysis, arrhythmia). Intravenous K$^+$-Cl$^-$ should always be administered in saline solutions, rather than dextrose, because the dextrose-induced increase in insulin can acutely exacerbate hypokalemia. The peripheral intravenous dose is usually 20–40 mmol of K$^+$-Cl$^-$ per liter; higher concentrations can cause localized pain from chemical phlebitis, irritation, and sclerosis. If hypokalemia is severe (<2.5 mmol/L) and/or critically symptomatic, intravenous K$^+$-Cl$^-$ can be administered through a central vein with cardiac monitoring in an intensive care setting, at rates of 10–20 mmol/h; higher rates should be reserved for acutely life-threatening complications. The absolute amount of administered K$^+$ should be restricted (e.g., 20 mmol in 100 mL of saline solution) to prevent inadvertent infusion of a large dose.

Strategies to minimize K$^+$ losses should also be considered. These measures may include minimizing the dose of non-K$^+$-sparing diuretics, restricting Na$^+$ intake, and using clinically appropriate combinations of non-K$^+$-sparing and K$^+$-sparing medications (e.g., loop diuretics with ACE inhibitors).

HYPERKALEMIA

Hyperkalemia is defined as a plasma potassium level of 5.5 m*M*, occurring in up to 10% of hospitalized patients; severe hyperkalemia (>6.0 m*M*) occurs in ~1%, with a significantly increased risk of mortality. Although redistribution and reduced tissue uptake can acutely cause hyperkalemia, a decrease in renal K^+ excretion is the most frequent underlying cause **(Table 56-5)**. Excessive intake of K^+ is a rare cause, given the adaptive capacity to increase renal secretion; however, dietary intake can have a major effect in susceptible patients, e.g., diabetics with hyporeninemic hypoaldosteronism and chronic kidney disease. Drugs that impact on the renin-angiotensin-aldosterone axis are also a major cause of hyperkalemia.

TABLE 56-5 Causes of Hyperkalemia

I. Pseudohyperkalemia
  A. Cellular efflux; thrombocytosis, erythrocytosis, leukocytosis, in vitro hemolysis
  B. Hereditary defects in red cell membrane transport
II. Intra- to extracellular shift
  A. Acidosis
  B. Hyperosmolality; radiocontrast, hypertonic dextrose, mannitol
  C. β_2-Adrenergic antagonists (noncardioselective agents)
  D. Digoxin and related glycosides (yellow oleander, foxglove, bufadienolide)
  E. Hyperkalemic periodic paralysis
  F. Lysine, arginine, and ε-aminocaproic acid (structurally similar, positively charged)
  G. Succinylcholine; thermal trauma, neuromuscular injury, disuse atrophy, mucositis, or prolonged immobilization
  H. Rapid tumor lysis
III. Inadequate excretion
  A. Inhibition of the renin-angiotensin-aldosterone axis; ↑ risk of hyperkalemia when used in combination
    1. Angiotensin-converting enzyme (ACE) inhibitors
    2. Renin inhibitors; aliskiren (in combination with ACE inhibitors or angiotensin receptor blockers [ARBs])
    3. ARBs
    4. Blockade of the mineralocorticoid receptor: spironolactone, eplerenone, drospirenone
    5. Blockade of the epithelial sodium channel (ENaC): amiloride, triamterene, trimethoprim, pentamidine, nafamostat
  B. Decreased distal delivery
    1. Congestive heart failure
    2. Volume depletion
  C. Hyporeninemic hypoaldosteronism
    1. Tubulointerstitial diseases: systemic lupus erythematosus (SLE), sickle cell anemia, obstructive uropathy
    2. Diabetes, diabetic nephropathy
    3. Drugs: nonsteroidal anti-inflammatory drugs (NSAIDs), cyclooxygenase 2 (COX2) inhibitors, β blockers, cyclosporine, tacrolimus
    4. Chronic kidney disease, advanced age
    5. Pseudohypoaldosteronism type II: defects in WNK1 or WNK4 kinases, Kelch-like 3 (KLHL3), or Cullin 3 (CUL3)
  D. Renal resistance to mineralocorticoid
    1. Tubulointerstitial diseases: SLE, amyloidosis, sickle cell anemia, obstructive uropathy, post–acute tubular necrosis
    2. Hereditary: pseudohypoaldosteronism type I; defects in the mineralocorticoid receptor *or* the epithelial sodium channel (ENaC)
  E. Advanced renal insufficiency
    1. Chronic kidney disease
    2. End-stage renal disease
    3. Acute oliguric kidney injury
  F. Primary adrenal insufficiency
    1. Autoimmune: Addison's disease, polyglandular endocrinopathy
    2. Infectious: HIV, cytomegalovirus, tuberculosis, disseminated fungal infection
    3. Infiltrative: amyloidosis, malignancy, metastatic cancer
    4. Drug-associated: heparin, low-molecular-weight heparin
    5. Hereditary: adrenal hypoplasia congenita, congenital lipoid adrenal hyperplasia, aldosterone synthase deficiency
    6. Adrenal hemorrhage or infarction, including in antiphospholipid syndrome

Pseudohyperkalemia Hyperkalemia should be distinguished from factitious hyperkalemia or "pseudohyperkalemia," an artifactual increase in serum K^+ due to the release of K^+ during or after venipuncture. Pseudohyperkalemia can occur in the setting of excessive muscle activity during venipuncture (e.g., fist clenching), a marked increase in cellular elements (thrombocytosis, leukocytosis, and/or erythrocytosis) with in vitro efflux of K^+, and acute anxiety during venipuncture with respiratory alkalosis and redistributive hyperkalemia. Cooling of blood following venipuncture is another cause, due to reduced cellular uptake; the converse is the increased uptake of K^+ by cells at high ambient temperatures, leading to normal values for hyperkalemic patients and/or to spurious hypokalemia in normokalemic patients. Finally, there are multiple genetic subtypes of hereditary pseudohyperkalemia, caused by increases in the passive K^+ permeability of erythrocytes. For example, causative mutations have been described in the red cell anion exchanger (AE1, encoded by the *SLC4A1* gene), leading to reduced red cell anion transport, hemolytic anemia, the acquisition of a novel AE1-mediated K^+ leak, and pseudohyperkalemia.

Redistribution and Hyperkalemia Several different mechanisms can induce an efflux of intracellular K^+ and hyperkalemia. Acidemia is associated with cellular uptake of H^+ and an associated efflux of K^+; it is thought that this effective K^+-H^+ exchange serves to help maintain extracellular pH. Notably, this effect of acidosis is limited to non–anion gap causes of metabolic acidosis and, to a lesser extent, respiratory causes of acidosis; hyperkalemia due to an acidosis-induced shift of potassium from the cells into the ECF does *not* occur in the anion gap acidoses lactic acidosis and ketoacidosis. Hyperkalemia due to hypertonic mannitol, hypertonic saline, and intravenous immune globulin is generally attributed to a "solvent drag" effect, as water moves out of cells along the osmotic gradient. Diabetics are also prone to osmotic hyperkalemia in response to intravenous hypertonic glucose, when given without adequate insulin. Cationic amino acids, specifically lysine, arginine, and the structurally related drug epsilon-aminocaproic acid, cause efflux of K^+ and hyperkalemia, through an effective cation-K^+ exchange of unknown identity and mechanism. Digoxin inhibits Na^+/K^+-ATPase and impairs the uptake of K^+ by skeletal muscle, such that digoxin overdose predictably results in hyperkalemia. Structurally related glycosides are found in specific plants (e.g., yellow oleander, foxglove) and in the cane toad, *Bufo marinus* (bufadienolide); ingestion of these substances and extracts thereof can also cause hyperkalemia. Finally, fluoride ions also inhibit Na^+/K^+-ATPase, such that fluoride poisoning is typically associated with hyperkalemia.

Succinylcholine depolarizes muscle cells, causing an efflux of K^+ through acetylcholine receptors (AChRs). The use of this agent is contraindicated in patients who have sustained thermal trauma, neuromuscular injury, disuse atrophy, mucositis, or prolonged immobilization. These disorders share a marked increase and redistribution of AChRs at the plasma membrane of muscle cells; depolarization of these upregulated AChRs by succinylcholine leads to an exaggerated efflux of K^+ through the receptor-associated cation channels, resulting in acute hyperkalemia.

Hyperkalemia Caused by Excess Intake or Tissue Necrosis Increased intake of even small amounts of K^+ may provoke severe hyperkalemia in patients with predisposing factors; hence, an assessment of dietary intake is crucial. Foods rich in potassium include tomatoes, bananas, and citrus fruits; occult sources of K^+, particularly K^+-containing salt substitutes, may also contribute significantly. Iatrogenic causes include simple overreplacement with K^+-Cl^- or the administration of a potassium-containing medication (e.g., K^+-penicillin) to a susceptible patient. Red cell transfusion is a well-described cause of hyperkalemia, typically in the setting of massive transfusions. Finally, severe tissue necrosis, as in acute tumor lysis syndrome and rhabdomyolysis, will predictably cause hyperkalemia from the release of intracellular K^+.

Hypoaldosteronism and Hyperkalemia Aldosterone release from the adrenal gland may be reduced by hyporeninemic hypoaldosteronism, medications, primary hypoaldosteronism, or isolated deficiency of ACTH (secondary hypoaldosteronism). Primary hypoaldosteronism may be genetic or acquired **(Chap. 398)** but is commonly caused by autoimmunity, either in Addison's disease or in the context of a polyglandular endocrinopathy. HIV is a particularly important infectious cause of adrenal insufficiency. The adrenal involvement in HIV disease is usually subclinical; however, adrenal insufficiency may be precipitated by stress, drugs such as ketoconazole that inhibit steroidogenesis, or the acute withdrawal of steroid agents such as megestrol. Among medications associated with hyperkalemia, heparin preparations can cause selective inhibition of aldosterone synthesis by zona glomerulosa cells, leading to hyperreninemic hypoaldosteronism.

Hyporeninemic hypoaldosteronism is a very common predisposing factor in several overlapping subsets of hyperkalemic patients: diabetics, the elderly, and patients with renal insufficiency. Classically, patients should have suppressed PRA and aldosterone; ~50% have an associated acidosis, with a reduced renal excretion of NH_4^+, a positive urinary anion gap, and urine pH <5.5. Most patients are volume expanded, with secondary increases in circulating atrial natriuretic peptide (ANP) that inhibit both renal renin release and adrenal aldosterone release.

Renal Disease and Hyperkalemia Chronic kidney disease and end-stage kidney disease are very common causes of hyperkalemia, due to the associated deficit or absence of functioning nephrons. Hyperkalemia is more common in oliguric acute kidney injury; distal tubular flow rate and Na^+ delivery are less limiting factors in nonoliguric patients. Hyperkalemia out of proportion to GFR can also be seen in the context of tubulointerstitial disease that affects the distal nephron, such as amyloidosis, sickle cell anemia, interstitial nephritis, and obstructive uropathy.

Hereditary renal causes of hyperkalemia have overlapping clinical features with hypoaldosteronism, hence the diagnostic label *pseudohypoaldosteronism* (PHA). PHA type I (PHA-I) has both an autosomal recessive and an autosomal dominant form. The autosomal dominant form is due to loss-of-function mutations in the MLR; the recessive form is caused by various combinations of mutations in the three subunits of ENaC, resulting in impaired Na^+ channel activity in principal cells and other tissues. Patients with recessive PHA-I suffer from lifelong salt wasting, hypotension, and hyperkalemia, whereas the phenotype of autosomal dominant PHA-I due to MLR dysfunction improves in adulthood. PHA type II (PHA-II; also known as *hereditary hypertension with hyperkalemia*) is in every respect the mirror image of GS caused by loss of function in NCC, the thiazide-sensitive Na^+-Cl^- cotransporter (see above); the clinical phenotype includes hypertension, hyperkalemia, hyperchloremic metabolic acidosis, suppressed PRA and aldosterone, hypercalciuria, and reduced bone density. PHA-II thus behaves like a gain of function in NCC, and treatment with thiazides results in resolution of the entire clinical phenotype. However, the NCC gene is not directly involved in PHA-II, which is caused by mutations in the WNK1 and WNK4 serine-threonine kinases or the upstream Kelch-like 3 (KLHL3) and Cullin 3 (CUL3) proteins, two components of an E3 ubiquitin ligase complex that regulates these kinases; these proteins collectively regulate NCC activity, with PHA-II-associated activation of the transporter.

Medication-Associated Hyperkalemia Most medications associated with hyperkalemia cause inhibition of some component of the renin-angiotensin-aldosterone axis. ACE inhibitors, angiotensin receptor blockers, renin inhibitors, and MLRs are predictable and common causes of hyperkalemia, particularly when prescribed in combination. The oral contraceptive agent Yasmin-28 contains the progestin drospirenone, which inhibits the MLR and can cause hyperkalemia in susceptible patients. Cyclosporine, tacrolimus, NSAIDs, and cyclooxygenase 2 (COX2) inhibitors cause hyperkalemia by multiple mechanisms, but share the ability to cause hyporeninemic hypoaldosteronism. Notably, most drugs that affect the renin-angiotensin-aldosterone axis also block the local adrenal response to hyperkalemia, thus attenuating the *direct* stimulation of aldosterone release by increased plasma K^+ concentration.

Inhibition of apical ENaC activity in the distal nephron by amiloride and other K^+-sparing diuretics results in hyperkalemia, often with a voltage-dependent hyperchloremic acidosis and/or hypovolemic hyponatremia. Amiloride is structurally similar to the antibiotics TMP and pentamidine, which also block ENaC; risk factors for TMP-associated hyperkalemia include the administered dose, renal insufficiency, and hyporeninemic hypoaldosteronism. Indirect inhibition of ENaC at the plasma membrane is also a cause of drug-associated hyperkalemia; nafamostat, a protease inhibitor used in some countries for anticoagulation and for the management of pancreatitis, inhibits aldosterone-induced renal proteases that activate ENaC by proteolytic cleavage.

Clinical Features Hyperkalemia is a medical emergency due to its effects on the heart. Cardiac arrhythmias associated with hyperkalemia include sinus bradycardia, sinus arrest, slow idioventricular rhythms, ventricular tachycardia, ventricular fibrillation, and asystole. Mild increases in extracellular K^+ affect the repolarization phase of the cardiac action potential, resulting in changes in T-wave morphology; further increase in plasma K^+ concentration depresses intracardiac conduction, with progressive prolongation of the PR and QRS intervals. Severe hyperkalemia results in loss of the P wave and a progressive widening of the QRS complex; development of a sine-wave sinoventricular rhythm suggests impending ventricular fibrillation or asystole. Hyperkalemia can also cause a type I Brugada pattern in the electrocardiogram (ECG), with a pseudo–right bundle branch block and persistent coved ST-segment elevation in at least two precordial leads. This hyperkalemic Brugada's sign occurs in critically ill patients with severe hyperkalemia and can be differentiated from genetic Brugada's syndrome by an absence of P waves, marked QRS widening, and an abnormal QRS axis. Classically, the ECG manifestations in hyperkalemia progress from tall peaked T waves (5.5–6.5 m*M*), to a loss of P waves (6.5–7.5 m*M*), to a widened QRS complex (7.0–8.0 m*M*), and, ultimately, a to a sine wave pattern (>8.0 m*M*). However, these changes are notoriously insensitive, particularly in patients with chronic kidney disease or ESRD.

Hyperkalemia from a variety of causes can also present with ascending paralysis, denoted *secondary hyperkalemic paralysis* to differentiate it from familial hyperkalemic periodic paralysis (HYPP). The presentation may include diaphragmatic paralysis and respiratory failure. Patients with familial HYPP develop myopathic weakness during hyperkalemia induced by increased K^+ intake or rest after heavy exercise. Depolarization of skeletal muscle by hyperkalemia unmasks an inactivation defect in skeletal Na^+ channel; autosomal dominant mutations in the *SCN4A* gene encoding this channel are the predominant cause.

Within the kidney, hyperkalemia has negative effects on the ability to excrete an acid load, such that hyperkalemia per se can contribute to metabolic acidosis. This defect appears to be due in part to competition between K^+ and NH_4^+ for reabsorption by the TALH and subsequent countercurrent multiplication, ultimately reducing the medullary gradient for NH_3/NH_4 excretion by the distal nephron. Regardless of the underlying mechanism, restoration of normokalemia can, in many instances, correct hyperkalemic metabolic acidosis.

Diagnostic Approach The first priority in the management of hyperkalemia is to assess the need for emergency treatment, followed by a comprehensive workup to determine the cause **(Fig. 56-8)**. History and physical examination should focus on medications, diet and dietary supplements, risk factors for kidney failure, reduction in urine output, blood pressure, and volume status. Initial laboratory tests should include electrolytes, BUN, creatinine, serum osmolality, Mg^{2+} and Ca^{2+}, a complete blood count, and urinary pH, osmolality, creatinine, and electrolytes. A urine Na^+ concentration of <20 m*M* indicates that distal Na^+ delivery is a limiting factor in K^+ excretion; volume

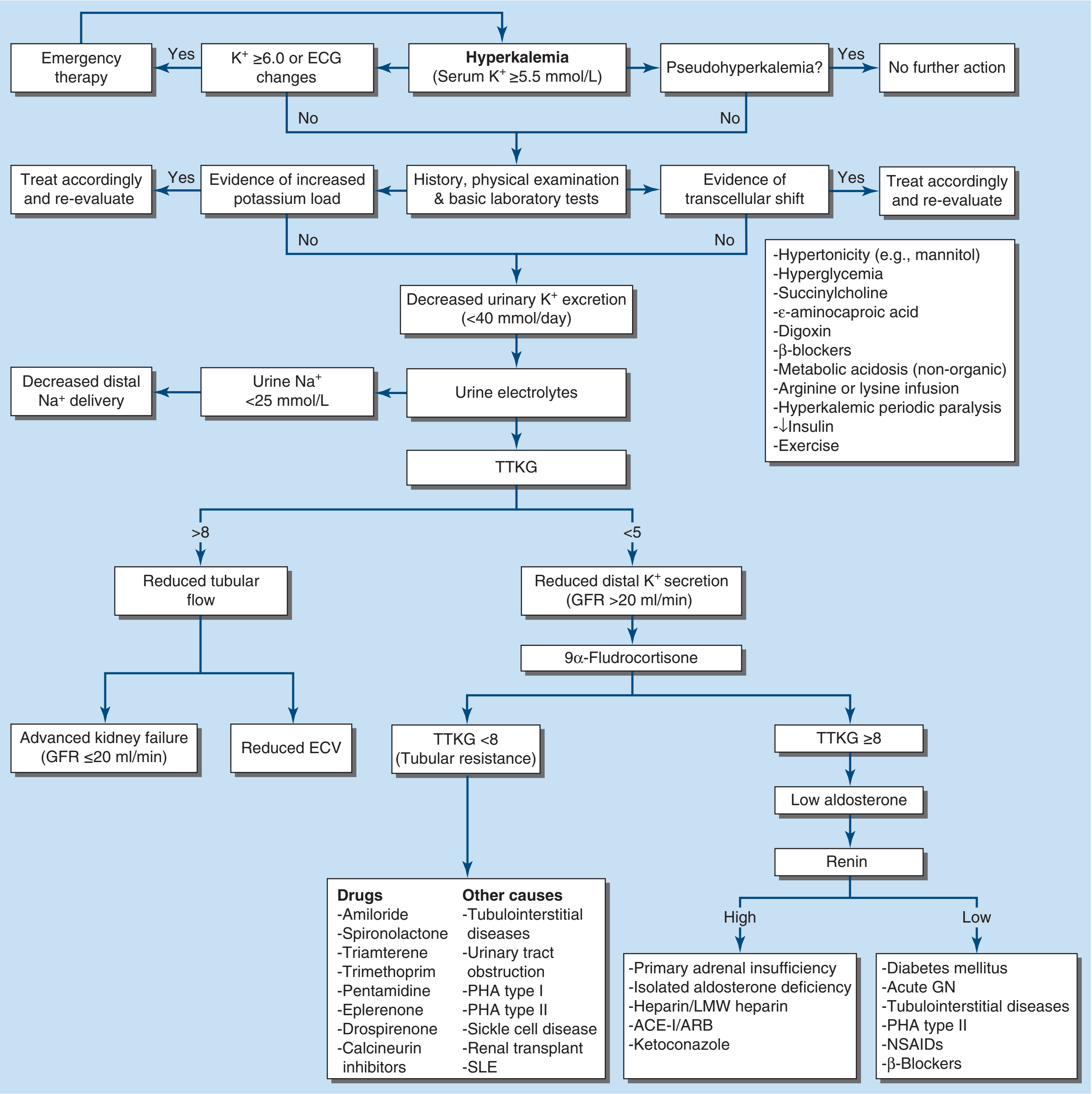
Emergency therapy
Yes
K+ ≥6.0 or ECG changes
Hyperkalemia (Serum K+ ≥5.5 mmol/L)
Pseudohyperkalemia?
Yes
No further action
No
No
Treat accordingly and re-evaluate
Yes
Evidence of increased potassium load
History, physical examination & basic laboratory tests
Evidence of transcellular shift
Yes
Treat accordingly and re-evaluate
No
No
-Hypertonicity (e.g., mannitol)
-Hyperglycemia
-Succinylcholine
-ε-aminocaproic acid
-Digoxin
-β-blockers
-Metabolic acidosis (non-organic)
-Arginine or lysine infusion
-Hyperkalemic periodic paralysis
-↓Insulin
-Exercise
Decreased urinary K+ excretion (<40 mmol/day)
Decreased distal Na+ delivery
Urine Na+ <25 mmol/L
Urine electrolytes
TTKG
>8
<5
Reduced tubular flow
Reduced distal K+ secretion (GFR >20 ml/min)
9α-Fludrocortisone
Advanced kidney failure (GFR ≤20 ml/min)
Reduced ECV
TTKG <8 (Tubular resistance)
TTKG ≥8
Low aldosterone
Renin
High
Low
Drugs
-Amiloride
-Spironolactone
-Triamterene
-Trimethoprim
-Pentamidine
-Eplerenone
-Drospirenone
-Calcineurin inhibitors
Other causes
-Tubulointerstitial diseases
-Urinary tract obstruction
-PHA type I
-PHA type II
-Sickle cell disease
-Renal transplant
-SLE
-Primary adrenal insufficiency
-Isolated aldosterone deficiency
-Heparin/LMW heparin
-ACE-I/ARB
-Ketoconazole
-Diabetes mellitus
-Acute GN
-Tubulointerstitial diseases
-PHA type II
-NSAIDs
-β-Blockers

FIGURE 56-8 The diagnostic approach to hyperkalemia. See text for details. ACE-I, angiotensin-converting enzyme inhibitor; ARB, angiotensin II receptor blocker; CCD, cortical collecting duct; ECG, electrocardiogram; ECV, effective circulatory volume; GFR, glomerular filtration rate; GN, glomerulonephritis; HIV, human immunodeficiency virus; LMW heparin, low-molecular-weight heparin; NSAIDs, nonsteroidal anti-inflammatory drugs; PHA, pseudohypoaldosteronism; SLE, systemic lupus erythematosus; TTKG, transtubular potassium gradient. *(Reproduced with permission from DB Mount, K Zandi-Nejad: Disorders of potassium balance, in BM Brenner [ed], Brenner and Rector's The Kidney, 8th ed, Philadelphia, W.B. Saunders & Company, 2008.)*

repletion with 0.9% saline or treatment with furosemide may be effective in reducing plasma K^+ concentration. Serum and urine osmolality are required for calculation of the transtubular K^+ gradient (TTKG) (Fig. 56-8). The expected values of the TTKG are largely based on historical data, and are <3 in the presence of hypokalemia and >7–8 in the presence of hyperkalemia. Notably, some authors have opined that the TTKG does not consider the effects of distal tubular urea reabsorption on potassium excretion, concluding that the TTKG is, thus, an unreliable test in the assessment of hyperkalemia. These criticisms are theoretical and not supported by animal experiments; the TTKG remains a helpful bedside test of urinary potassium excretion in hyperkalemia.

$$\text{TTKG} = \frac{[K^+]_{\text{urine}} \times \text{Osm}_{\text{serum}}}{[K^+]_{\text{serum}} \times \text{Osm}_{\text{urine}}}$$

TREATMENT

Hyperkalemia

ECG manifestations of hyperkalemia should be considered a medical emergency and treated urgently. However, patients with significant hyperkalemia (plasma K^+ concentration ≥6.5 m*M*) in the absence of ECG changes should also be aggressively managed, given the limitations of ECG changes as a predictor of cardiac toxicity. Urgent management of hyperkalemia includes admission to the hospital, continuous cardiac monitoring, and immediate treatment. The treatment of hyperkalemia is divided into three stages:

1. *Immediate antagonism of the cardiac effects of hyperkalemia.* Intravenous calcium serves to protect the heart, whereas other

measures are taken to correct hyperkalemia. Calcium raises the action potential threshold and reduces excitability, without changing the resting membrane potential. By restoring the difference between resting and threshold potentials, calcium reverses the depolarization blockade due to hyperkalemia. The recommended dose is 10 mL of 10% calcium gluconate (3–4 mL of calcium chloride), infused intravenously over 2–3 min with cardiac monitoring. The effect of the infusion starts in 1–3 min and lasts 30–60 min; the dose should be repeated if there is no change in ECG findings or if they recur after initial improvement. Hypercalcemia potentiates the cardiac toxicity of digoxin; hence, intravenous calcium should be used with extreme caution in patients taking this medication; if judged necessary, 10 mL of 10% calcium gluconate can be added to 100 mL of 5% dextrose in water and infused over 20–30 min to avoid acute hypercalcemia.

2. *Rapid reduction in plasma K^+ concentration by redistribution into cells.* Insulin lowers plasma K^+ concentration by shifting K^+ into cells. The recommended dose is 10 units of intravenous regular insulin followed immediately by 50 mL of 50% dextrose ($D_{50}W$, 25 g of glucose total); the effect begins in 10–20 min, peaks at 30–60 min, and lasts for 4–6 h. Bolus $D_{50}W$ without insulin is *never* appropriate, given the risk of acutely worsening hyperkalemia due to the osmotic effect of hypertonic glucose. Hypoglycemia is common with insulin plus glucose; hence, this should be followed by an infusion of 10% dextrose at 50–75 mL/h, with close monitoring of plasma glucose concentration. In hyperkalemic patients with glucose concentrations of ≥200–250 mg/dL, insulin should be administered *without* glucose, again with close monitoring of glucose concentrations.

   β_2-Agonists, most commonly albuterol, are effective but underused agents for the acute management of hyperkalemia. Albuterol and insulin with glucose have an additive effect on plasma K^+ concentration; however, ~20% of patients with ESRD are resistant to the effect of β_2-agonists; hence, these drugs should not be used without insulin. The recommended dose for inhaled albuterol is 10–20 mg of nebulized albuterol in 4 mL of normal saline, inhaled over 10 min; the effect starts at about 30 min, reaches its peak at about 90 min, and lasts for 2–6 h. Hyperglycemia is a side effect, along with tachycardia. β_2-Agonists should be used with caution in hyperkalemic patients with known cardiac disease.

   Intravenous bicarbonate has no role in the acute treatment of hyperkalemia but may slowly attenuate hyperkalemia with sustained administration over several hours. It should not be given repeatedly as a hypertonic intravenous bolus of undiluted ampules, given the risk of associated hypernatremia and hypertonicity, but should instead be infused in an isotonic or hypotonic fluid (e.g., 150 milliequivalents of sodium bicarbonate in 1 L of D_5W). In patients with metabolic acidosis, a delayed drop in plasma K^+ concentration can be seen after 4–6 h of isotonic bicarbonate infusion.

3. *Removal of potassium.* This is typically accomplished using cation exchange resins, diuretics, and/or dialysis. The cation exchange resin sodium polystyrene sulfonate (SPS) exchanges Na^+ for K^+ in the gastrointestinal tract and increases the fecal excretion of K^+. The recommended dose of SPS is 15–30 g of powder, almost always given in a premade suspension with 33% sorbitol. The effect of SPS on plasma K^+ concentration is slow; the full effect may take up to 24 h and usually requires repeated doses every 4–6 h. Intestinal necrosis, typically of the colon or ileum, is a rare but usually fatal complication of SPS. Intestinal necrosis is more common in patients with reduced intestinal motility (e.g., in the postoperative state or after treatment with opioids). The coadministration of SPS with sorbitol appears to increase the risk of intestinal necrosis; however, this complication can also occur with SPS alone, and in animal models, SPS is the causative agent. The low but real risk of intestinal necrosis with SPS, which can sometimes be the only available or appropriate therapy for the removal of potassium, must be weighed against the delayed onset of efficacy. Whenever possible, alternative therapies for the acute management of hyperkalemia (i.e., alternative potassium binders, aggressive redistributive therapy, isotonic bicarbonate infusion, diuretics, and/or hemodialysis) should be used instead of SPS.

   Two other intestinal potassium binders are available for the management of hyperkalemia. These agents lack the intestinal toxicity of SPS and are preferred over SPS for the management of hyperkalemia. Patiromer is a nonabsorbed polymer provided as a powder for suspension, which binds K^+ in exchange for Ca^{2+}. In healthy adults, patiromer causes a decrease in urinary potassium, magnesium, and sodium excretion, suggesting the binding of the polymer to these cations in the intestine; notably, a major side effect of the medication is hypomagnesemia. Sodium zirconium cyclosilicate is an inorganic, nonabsorbable crystalline compound that exchanges both Na^+ and H^+ ions in exchange for K^+ and NH_4^+ in the intestine. These agents have revolutionized the management of both chronic and acute hyperkalemia. In particular, the availability of safe, well-tolerated potassium binders allows for greater intensity of renin-angiotensin-aldosterone system inhibition in both renal and cardiac disease.

   Therapy with intravenous saline may be beneficial in hypovolemic patients with oliguria and decreased distal delivery of Na^+, with the associated reductions in renal K^+ excretion. Loop and thiazide diuretics can be used to reduce plasma K^+ concentration in volume-replete or hypervolemic patients with sufficient renal function for a diuretic response; this may need to be combined with intravenous saline or isotonic bicarbonate to achieve or maintain euvolemia.

   Hemodialysis is the most effective and reliable method to reduce plasma K^+ concentration; peritoneal dialysis is considerably less effective. Patients with acute kidney injury require temporary, urgent venous access for hemodialysis, with the attendant risks; in contrast, patients with ESRD or advanced chronic kidney disease may have a preexisting venous access. The amount of K^+ removed during hemodialysis depends on the relative distribution of K^+ between ICF and ECF (potentially affected by prior therapy for hyperkalemia), the type and surface area of the dialyzer used, dialysate and blood flow rates, dialysate flow rate, dialysis duration, and the plasma-to-dialysate K^+ gradient.

FURTHER READING

Choi M et al: K^+ channel mutations in adrenal aldosterone-producing adenomas and hereditary hypertension. Science 331:768, 2011.

Clase KM et al: Potassium homeostasis and management of dyskalemia in kidney diseases: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. Kidney Int 97:42, 2020.

Fenske W et al: A copeptin-based approach in the diagnosis of diabetes insipidus. N Engl J Med 379:428, 2018.

Gankam-Kengne F et al: Osmotic stress–induced defective glial proteostasis contributes to brain demyelination after hyponatremia treatment. J Am Soc Nephrol 28:1802, 2017.

Mount DB: Disorders of potassium balance, in *Brenner and Rector's The Kidney*, 11th ed, ASL Yu et al: (eds). Philadelphia, W.B. Saunders & Company, 2020, pp. 537–579.

Packham DK et al: Sodium zirconium cyclosilicate in hyperkalemia. N Engl J Med 372:222, 2015.

Perianayagam A et al: DDAVP is effective in preventing and reversing inadvertent overcorrection of hyponatremia. Clin J Am Soc Nephrol 3:331, 2008.

Rondon-Berrios H, Sterns RH: Hypertonic saline for hyponatremia: Meeting goals and avoiding harm. Am J Kidney Dis 79:890, 2022.

Soupart A et al: Efficacy and tolerance of urea compared with vaptans for long-term treatment of patients with SIADH. Clin J Am Soc Nephrol 7:742, 2012.

Turcu AF et al: Primary aldosteronism: A multidimensional syndrome. Nat Rev Endocrinol 18:665, 2022.

57 Hypercalcemia and Hypocalcemia

Dwight A. Towler

Calcium is the most abundant mineral in the human body, fulfilling numerous physiological functions. These functions include providing fundamental structure and strength of the skeleton [as the solid hydroxyapatite salt $Ca_{10}(PO_4)_6(OH)_2$]; cellular signaling and hormone secretion; cardiac pacemaker rhythmicity and contractility; skeletal muscle contractility; immune function; and neural signaling, among others. Given these and other vital biological functions, intracellular and extracellular calcium concentrations are tightly regulated on timescales of milliseconds (intracellular) to minutes–hours (extracellular), with the former very dynamic and the latter very stable as necessary for health.

Serum calcium concentrations are maintained within an exquisitely narrow range through feedback mechanisms that involve parathyroid hormone (PTH) and the vitamin D metabolites 1,25-dihydroxyvitamin D [$1,25(OH)_2D$] and 25-hydroxyvitamin D [25(OH)D]. PTH increases osteocyte production of RANKL, a potent stimulus for osteoclast-mediated bone resorption that mobilizes skeletal calcium. These feedback mechanisms integrate signals between the parathyroid glands, kidney, intestine, and bone to control circulating extracellular calcium **(Fig. 57-1; Chap. 421)**.

Disorders of serum calcium concentration are very common, arising from primary or secondary perturbations in the calciotropic functions. Disease manifestations reflect not only the impact of acute calcium changes on organ function, but also the impact of chronic compensatory responses that come with clinical consequences (e.g., osteoporosis, nephrocalcinosis). This chapter provides a brief summary of the approach to patients with altered serum calcium levels. **See Chap. 422 for a detailed discussion of this topic.**

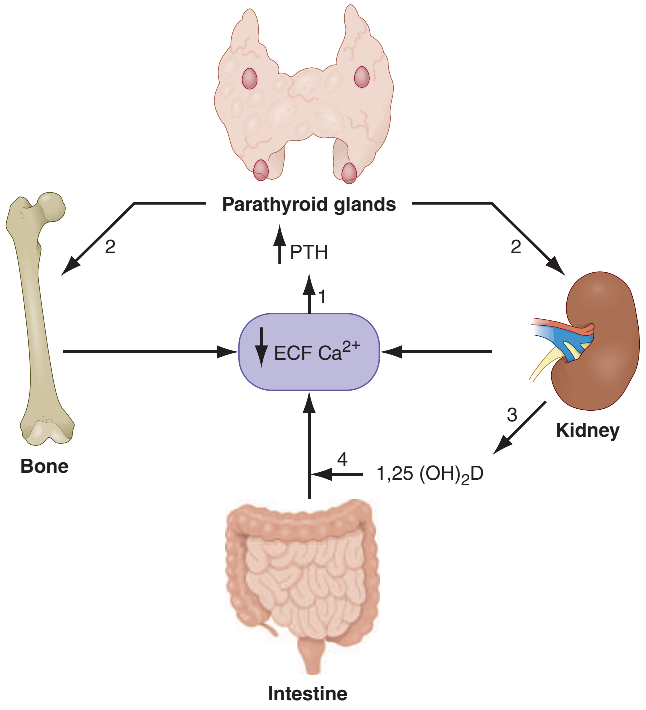

FIGURE 57-1 Homeostatic feedback mechanisms maintain extracellular calcium concentrations within a narrow physiologic range, typically 8.5–10.5 mg/dL (2.1–2.6 m*M*) (reference values vary slightly by lab). A decrease in extracellular (ECF) calcium (Ca^{2+}) triggers an increase in parathyroid hormone (PTH) secretion (1) via the calcium sensor receptor (CaSR) on parathyroid chief cells. PTH, in turn, increases resorption of calcium from bone (2, left) and distal renal tubule reabsorption of filtered urinary calcium (2, right), and also stimulates renal proximal tubule $1,25(OH)_2D$ production (3). Circulating $1,25(OH)_2D$ acts on the small intestine to increase dietary calcium absorption (4). Collectively, these homeostatic mechanisms defend against life-threatening hypocalcemia and serve to restore serum calcium levels to normal.

HYPERCALCEMIA

ETIOLOGY

The causes of hypercalcemia can be understood based on derangements in the normal feedback mechanisms that regulate serum calcium **(Table 57-1)**. Primary hyperparathyroidism and malignancy are the two most common causes of hypercalcemia, although other diseases must be considered **(Table 57-2)**. Excess PTH production, which is not appropriately suppressed by increased serum calcium concentrations, occurs in primary hyperparathyroidism (pHPT). pHPT is almost always due to a nonmalignant neoplastic disorder of the parathyroid glands (a parathyroid adenoma or parathyroid hyperplasia) that is associated with increased parathyroid cell mass and impaired feedback

TABLE 57-1 Causes of Hypercalcemia

Excessive PTH production
Primary hyperparathyroidism
Sporadic or familial (e.g., MEN) parathyroid adenoma or hyperplasia
Rarely parathyroid carcinoma
Tertiary hyperparathyroidism (long-term stimulation of PTH secretion in renal insufficiency)
Ectopic PTH secretion (very rare)
FHH (mutations reducing CaSR signaling)
Nongenetic alterations in CaSR function (lithium therapy, rare CaSR inhibitory autoantibodies)
Hypercalcemia of malignancy
Overproduction of PTHrP (many solid tumors)
Lytic skeletal metastases (breast, myeloma)
Excessive $1,25(OH)_2D$ production (by constitutive macrophage CYP27B1 activity)
Granulomatous diseases (many causes: sarcoidosis, tuberculosis, leprosy, fungal infections, foreign body reaction)
Lymphomas (B cell, cutaneous T cell/Sezary syndrome)
Vitamin D intoxication
Ingestion of excessive vitamin D: 25(OH)D
Ingestion of excessive $1,25(OH)_2D$ (calcitriol) or 1-alpha-hydroxylated vitamin D analogues
Reduced vitamin D catabolism due to CYP24A1 deficiency
Primary increase in bone resorption
Hyperthyroidism
Paget's disease of bone
Immobilization
Ketogenic diet (in treatment of children with refractory epilepsy, uncommon)
Excessive calcium intake
Milk-alkali syndrome
Total parenteral nutrition
Other causes
Endocrine disorders (adrenal insufficiency, pheochromocytoma, acromegaly, VIPoma)
Medications (thiazides, vitamin A, lithium, foscarnet, teriparatide, aromatase inhibitors)
Discontinuation of denosumab with rebound resorption
Recovery phase of rhabdomyolysis with renal failure
Williams-Beuren syndrome
Excessive mammary PTHrP production with pregnancy or lactation (rare)
Pseudohypercalcemia (e.g., calcium-binding IgM in Waldenström's macroglobulinemia)

Abbreviations: CaSR, calcium sensor receptor; FHH, familial hypocalciuric hypercalcemia; MEN, multiple endocrine neoplasia; PTH, parathyroid hormone; PTHrP, PTH-related peptide.

TABLE 57-2 Causes of Hypocalcemia

Low Parathyroid Hormone Levels (Hypoparathyroidism)

- Parathyroid agenesis
  - Isolated (genetic: autosomal recessive *GCM2* loss of function; X-linked mutations)
  - Syndromic (e.g., DiGeorge's, HDR, Kearns-Sayre)
- Parathyroid destruction
  - Surgical
  - Autoimmune (genetic, e.g., APS-1; or acquired, including immune checkpoint inhibitors)
  - Radiation
  - Iron or copper overload (hemochromatosis, transfusion-dependent thalassemia, Wilson's disease)
  - Infiltration by metastases or systemic diseases (sarcoid, IgG4-related disease, Riedel's struma)
- Reduced bioactive PTH secretion
  - Hypomagnesemia (genetic, drug-induced, anti–claudin-16 antibodies)
  - Severe hypermagnesemia (e.g., obstetric magnesium infusions)
  - Autosomal dominant hypocalcemia (CaSR activating mutations)
  - Autoimmune (CaSR activating antibodies; immune checkpoint inhibitor therapy)
  - Familial isolated hypoparathyroidism (rare PTH gene mutations)

High Parathyroid Hormone Levels (Secondary Hyperparathyroidism)

- Vitamin D deficiency or impaired 1,25(OH)$_2$D production/action
  - Nutritional vitamin D deficiency (poor intake or absorption)
  - Renal insufficiency [phosphate retention, impaired 1,25(OH)2D production]
  - Excessive intravenous phosphate infusion
  - Vitamin D resistance, including receptor defects
- Intrinsic small intestine disorders with chronic calcium malabsorption
  - Roux-en-Y bariatric surgery, other causes of short bowel syndrome
  - Celiac disease, Crohn's disease (often with concomitant vitamin D deficiency)
- Parathyroid hormone resistance
  - Pseudohypoparathyroidism (*GNAS* mutations or imprinting defects)
  - Congenital PTH receptor mutation syndromes
  - PTH receptor blocking autoantibodies
- Drugs
  - Alcohol (both acute and chronic ingestion)
  - Calcium chelators (including citrate from massive blood transfusion or FFP)
  - Inhibitors of bone resorption (bisphosphonates, denosumab, plicamycin)
  - Increased vitamin D catabolism (phenytoin, phenobarbital)
  - Phosphate-containing enemas (excessive or with renal disease)
- Miscellaneous causes
  - Acute pancreatitis
  - Acute rhabdomyolysis
  - Hungry bone syndrome after parathyroidectomy
  - Osteoblastic metastases with marked stimulation of bone formation (prostate cancer)
  - Respiratory or metabolic alkalosis (increased calcium binding to albumin)
  - Refeeding syndrome and its management
- Pseudohypocalcemia (e.g., low serum albumin with malnutrition, nephrotic syndrome, cirrhosis; gadoversetamide MRI contrast)

Abbreviations: APS-1, autoimmune polyglandular syndrome type 1; CaSR, calcium sensing receptor; FFP, fresh frozen plasma; HDR, hypoparathyroidism, sensorineural deafness, and renal disease; MEN, multiple endocrine neoplasia; MRI, magnetic resonance imaging; PTH, parathyroid hormone.

inhibition by calcium. When investigated, genetic causes can be identified in approximately 30% of pHPT cases even in nonfamilial sporadic pHPT, primarily affecting cyclin D1 or menin expression. Moreover, reduced calcium-sensing receptor (CaSR) expression due to *CASR* chromatin inhibitory H3K27me3 and H3K9me3 hypermethylation is observed in about 50% of sporadic parathyroid adenomas. Reduced CASR protein expression increases both adenoma growth and the basal point for PTH secretion. Inappropriate PTH secretion for the ambient level of serum calcium also occurs in familial hypocalciuric hypercalcemia (FHH). FHH is an autosomal dominant syndrome most commonly involving inactivating mutations in the *CASR* gene (FHH type 1), with rare families having mutations in other genes encoding proteins necessary for normal CaSR signal transduction. All FHH mutations impair extracellular calcium sensing in both the parathyroid gland and the kidneys, leading to concomitant inappropriate PTH secretion and decreased urinary calcium excretion. In hypercalcemic disorders with normal parathyroid function, PTH levels are suppressed by high serum calcium levels that activate the CaSR to inhibit parathyroid gland PTH production (Fig. 57-1). This is frequently encountered in patients with advanced malignancy. Many solid tumors produce PTH-related peptide (PTHrP). Once processed and secreted, PTHrP shares homology with the first 13 amino acids of PTH and binds the PTH receptor, thus mimicking effects of PTH on bone and kidney to cause hypercalcemia. The hypercalcemia associated with lymphoma or with granulomatous disease (e.g., sarcoidosis, tuberculosis, foreign body reaction) is caused by extrarenal conversion of 25(OH)D to the more potent pro-calcific hormone 1,25(OH)$_2$D by macrophage CYP27B1 activity. The latter is a feature of local innate immune regulation by vitamin D that operates independent of the normal negative feedback loops outlined in Figure 57-1. In granulomatous disorders, this chronic 1,25(OH)$_2$D production enhances intestinal calcium absorption, resulting in hypercalcemia and suppressed PTH. Reduced 1,25(OH)$_2$D catabolism due to genetic deficiency in the inactivating 24-hydroxylase *CYP24A1* can also cause hypercalcemia. Unlike the direct actions of 1,25(OH)$_2$D intoxication, mechanisms of acute vitamin D$_3$ (cholecalciferol) or vitamin D$_2$ (ergocalciferol) intoxication relate to acute cellular bioavailability of the 25(OH)D prohormone that is normally stored bound to plasma vitamin D binding protein (DBP). Excessive vitamin D ingestion rapidly displaces this large pool of 25(OH)D normally sequestered by plasma DBP. Although the potency of 25(OH)D for the cellular vitamin D receptor is 500-fold less than 1,25(OH)$_2$D, acute displacement of 25(OH)D from DBP by excessive exogenous vitamin D exposes tissues to 25(OH)D concentrations sufficient to directly activate the receptor in small intestine, increase transcellular calcium absorption, and drive hypercalcemia. While 1,25(OH)$_2$D is normally metabolized within a day or two, the longer half-lives of vitamin D and 25(OH)D can prolong hypercalcemia from vitamin D intoxication for several weeks. Disorders that directly increase calcium mobilization from bone, such as thyrotoxicosis, Paget's disease of bone, or osteolytic metastases, also lead to hypercalcemia with suppressed PTH secretion, as does exogenous calcium overload. The latter is most commonly observed in excessive oral ingestion (e.g., milk-alkali syndrome) or during total parenteral nutrition with excessive calcium supplementation. Ketogenic diets used to mitigate refractory seizures in children also increase bone resorption and can cause hypercalcemia. Hypercalcemia associated with prolonged immobilization arises from reduced mechanical loading of bone that increases osteocyte RANKL production and osteoclast-mediated bone resorption. In this setting, particularly with older patients, a previously unrecognized contributor to bone resorption (e.g., pHPT, Paget's disease, or myeloma) may be present and should be considered. Immune checkpoint inhibitors have been rarely associated with either hypercalcemia or hypocalcemia, dependent upon whether the autoimmune immunoglobulin response elicited inhibits or activates the CaSR, respectively.

■ CLINICAL MANIFESTATIONS

Mild hypercalcemia (10.5–11.9 mg/dL) is largely asymptomatic and usually recognized only on routine calcium measurements. Some patients may complain of vague neuropsychiatric symptoms, including trouble concentrating, personality changes, or depression. Other common presenting symptoms may include fatigue, proximal muscle weakness, constipation, hyporeflexia, or nephrolithiasis, and bone mineral density may be reduced. Moderate (≥12–14 mg/dL) to severe (≥14 mg/dL) hypercalcemia may result in progressive lethargy, stupor, or coma, as well as prominent gastrointestinal symptoms (nausea, vomiting, anorexia, peptic ulcer, or pancreatitis) and acute renal failure. Hypercalcemia decreases renal concentrating ability in part by downregulating renal collecting duct aquaporin 2, which causes polyuria and volume depletion with frequent prerenal failure. With long-standing hyperparathyroidism, patients may present with bone pain or pathologic fractures and hypertension. Finally, hypercalcemia can result in significant electrocardiographic changes, including bradycardia, atrioventricular block, QRS widening, and short QTc interval.

DIAGNOSTIC APPROACH

The first step in the diagnostic evaluation of hypercalcemia or hypocalcemia is to ensure that the reported alteration in serum calcium is not due to an abnormal albumin concentration or a similar confounder (pseudohypercalcemia or pseudohypocalcemia; Tables 57-1 and 57-2). About 50% of total serum calcium is ionized, with the remainder bound principally to albumin. Although direct measurements of ionized calcium are possible, these are easily influenced by collection methods that alter calcium binding to albumin (e.g., prolonged tourniquet application during phlebotomy). Thus, it is generally preferable to measure total serum calcium and albumin to "correct" the serum calcium. When serum albumin concentrations are reduced, a corrected calcium concentration is calculated by adding 0.8 mg/dL to the total calcium level for every decrement in serum albumin of 1.0 g/dL below the albumin reference value of 4.0 g/dL. Conversely, when albumin concentrations are increased, calcium is corrected by subtracting 0.8 mg/dL from the total serum calcium for every increase in serum albumin of 1.0 g/dL above 4.0 g/dL.

A detailed history provides important clues regarding the etiology of hypercalcemia (Table 57-1). Chronic hypercalcemia (many months to several years), often fluctuating between high normal to just above the limits of normal, is most commonly caused by pHPT or by FHH, and occasionally by multiple myeloma. Because of the phosphaturic actions of PTH, the hypercalcemia of pHPT is often accompanied by low serum phosphate or frank hypophosphatemia. Most malignancy-associated hypercalcemia presents with a more acute course. The history should include review of medications (e.g., thiazides, recent initiation of an aromatase inhibitor), over-the-counter supplements or special diets, personal or family history of nephrolithiasis or other endocrine/metabolic disorders, any previous neck surgery, prior or current malignancy diagnosis, and systemic symptoms suggestive of sarcoidosis, lymphoma, or hyperthyroidism that may be causative or contributory.

Once true hypercalcemia is established, the second most important laboratory test in the diagnostic evaluation is a PTH level using a two-site assay for the intact hormone. Serum creatinine should be measured to assess renal function; hypercalcemia can impair renal function, and reduced clearance of inactive PTH fragments may still be detected in some third-generation two-site PTH assays. However, if the PTH level is increased (or "inappropriately normal") in the setting of elevated calcium and low/low-normal phosphorus, the diagnosis is almost always pHPT, particularly if the 24-hour urinary fractional excretion of calcium is elevated (>0.02; described below). Genetic analysis is important if early-onset or dominant hereditary pHPT is present with (1) other endocrinopathy suggesting multiple endocrine neoplasia or (2) a lytic-sclerotic tumor of the jaw. In the latter, the inactivating mutations in the tumor suppressor gene *CDC73* that cause this syndromic pHPT profoundly increase the lifetime risk for parathyroid carcinoma (~30%). Because individuals with FHH may also present with mildly elevated PTH levels and hypercalcemia, FHH should be considered and excluded because parathyroid surgery is ineffective in this condition. Due to impaired CaSR function in parathyroids and kidney, a calcium/creatinine clearance ratio (calculated as urine calcium/serum calcium divided by urine creatinine/serum creatinine) of <0.01 is suggestive of FHH, particularly when there is a family history of mild, asymptomatic hypercalcemia or recurrent unsuccessful surgeries for "presumed" pHPT. Sequencing of the genes that control CaSR signaling (*CASR*, *GNA11*, *AP2S1*) is now commonly performed for the definitive diagnosis of FHH, particularly if urine calcium studies are inconclusive. Prolonged lithium therapy increases the CaSR set point for negative feedback by calcium and can cause hyperparathyroidism due to parathyroid hyperplasia or adenoma. Ectopic PTH secretion is extremely rare but has been reported in carcinomas with PTH gene rearrangements; unlike pHPT or parathyroid carcinoma, ectopic PTH secretion does not fluctuate in response to therapeutic correction of hypercalcemia since it is not controlled by a CaSR-sensitive tissue.

A suppressed PTH level in the face of hypercalcemia indicates that the hypercalcemia is independent of parathyroid gland function; while many other causes must be excluded (Table 57-1), this is most often due to an underlying malignancy. Although a tumor that causes hypercalcemia is generally overt, a PTHrP level may be needed to establish the diagnosis of hypercalcemia of malignancy. Multiple myeloma must be evaluated in this setting as well, including measurement of serum free light chains since ~3% of myeloma may be undetected by serum or urine protein electrophoresis with immunofixation. Measurement of serum 25(OH)D is useful for evaluation of both hypercalcemia and hypocalcemia. Hypercalcemia is one of the few settings where a serum 1,25$(OH)_2$D measurement is useful. In hypercalcemia caused by granulomatous disorders or lymphomas, serum 1,25$(OH)_2$D levels are either frankly increased or inappropriately normal, in the setting of a low serum PTH. With vitamin D– or 1,25$(OH)_2$D-mediated hypercalcemia, serum phosphate is frequently elevated or high normal due to increased gastrointestinal absorption and reduced renal excretion. In the vast majority of cases, clinical evaluation in combination with laboratory testing should provide a diagnosis. The very useful mnemonic **"VITAMINS TRAP"** captures much of the differential for the various disorders listed in Table 57-1: vitamin D, vitamin A intoxication; immobilization; thyrotoxicosis, thiazides, 24-hydroxylase deficiency; Addison's disease, acromegaly, acquired autoimmune; milk-alkali syndrome; inflammation (chronic); neoplasia including multiple myeloma; sarcoidosis; tuberculosis or another granulomatous disease; rhabdomyolysis recovery phase, rebound from denosumab discontinuation; AIDS; PTH (hyperparathyroidism), Paget's disease, parenteral sources, pheochromocytoma, pregnancy-associated PTHrP, and pseudohypercalcemia.

TREATMENT

Hypercalcemia

Mild, asymptomatic hypercalcemia does not require immediate therapy beyond assiduous maintenance of adequate oral hydration with close diagnostic follow-up; management is dictated by the underlying cause. By contrast, significant, symptomatic hypercalcemia requires acute intervention to restore normocalcemia independent of the etiology, in addition to treatment of the underlying cause. Initial therapy of significant hypercalcemia begins with volume expansion because hypercalcemia invariably leads to dehydration, as noted above; 4–6 L of intravenous saline may be required over the first 24 h, keeping in mind that underlying comorbidities (e.g., congestive heart failure) will impact the clinical euvolemic set-point. Saline infusion improves renal perfusion, urine output, and thus natriuresis, which thereby increases calciuresis. Although loop diuretics can be transiently used to further augment sodium and thus calcium excretion by the distal tubule, loop diuretics should not be initiated until the volume status has been restored to normal. Following intravenous hydration, which helps to most rapidly address hypercalcemia, the mainstays of therapy are intravenous aminobisphosphonates. Aminobisphosphonates are potent inhibitors of osteoclast-mediated bone resorption, the major contributor to severe hypercalcemia in most malignancies and hyperparathyroidism. Zoledronic acid (4 mg intravenously over ~30 min) and pamidronate (60–90 mg intravenously over 2–4 h) are commonly used for the treatment of hypercalcemia of malignancy and parathyroid crisis in adults and are occasionally deployed at lower doses in life-threatening hypercalcemia in children. The onset of intravenous aminobisphosphonate action is within 1–2 days, with normalization of serum calcium levels occurring in 60–90% of patients. To achieve a more rapid albeit transient reduction in bone resorption, calcitonin is routinely implemented as well (4–8 IU/kg given intramuscularly or subcutaneously every 6 h) during the first 48 hours following admission in conjunction with hydration and aminobisphosphonate infusion. Denosumab (dosed 120 mg subcutaneously on days 1, 8, 15, and 29, then monthly thereafter), an antibody to RANKL, is a potent inhibitor of bone resorption and has been shown to be effective in treating hypercalcemia of malignancy refractory to aminobisphosphonates. Lower doses of denosumab (e.g., 0.3 mg/kg) can be used to treat hypercalcemia patients in the setting of advanced kidney disease, since intravenous aminobisphosphonates are contraindicated with glomerular filtration rate <35 mL/min/1.73 m^2 and denosumab is cleared by the reticuloendothelial system. In some instances, dialysis may be required to control severe hypercalcemia, particularly in the settings of anuric acute renal failure or congestive heart failure. When diagnosed, parathyroid carcinoma, a rare cause of hypercalcemia of malignancy driven by PTH, is treated with the CaSR agonist cinacalcet, starting with 30 mg orally twice daily and titrated to normalize serum calcium. Of note, cinacalcet is also used to treat patients with pHPT who are not candidates for parathyroid gland surgery.

In patients with vitamin D– or 1,25(OH)$_2$D-mediated hypercalcemia, glucocorticoids represent foundational pharmacotherapy, as they directly reduce calcium absorption from the small intestine and decrease 1,25(OH)$_2$D production. Intravenous hydrocortisone (200–400 mg daily) for 3–5 days or oral prednisone (40–60 mg daily) for 7–10 days is used most often, with prolonged taper as guided by serum calcium. However, for severe vitamin D–induced intoxication requiring hospitalization, intravenous hydration and aminobisphosphonates are frequently necessary to rapidly restore normocalcemia. Other drugs, such as ketoconazole, chloroquine, and hydroxychloroquine, decrease 1,25(OH)$_2$D production but are used infrequently due to side effects. However, for individuals with genetic CYP24A1 deficiency, inhibiting 1,25(OH)$_2$D synthesis with ketoconazole or fluconazole can help control hypercalcemia and mitigate the toxicity of chronic glucocorticoid administration. For hypercalcemia of thyroid storm, the aggressive management of the underlying thyrotoxicosis, which also includes intravenous glucocorticoid administration in addition to thionamides, usually restores normocalcemia.

HYPOCALCEMIA

ETIOLOGY

The causes of hypocalcemia can be differentiated according to whether serum PTH levels are low (hypoparathyroidism) or high (secondary hyperparathyroidism). Although there are many potential causes of hypocalcemia, impaired PTH production and severely impaired intestinal calcium absorption with or without profound vitamin D deficiency are the most common etiologies (Table 57-2) **(Chap. 422)**. Because PTH is the main minute-by-minute defense against hypocalcemia, disorders associated with deficient PTH production may be associated with profound, debilitating, and life-threatening hypocalcemia. In adults, hypoparathyroidism most commonly results from inadvertent damage to all four parathyroid glands during thyroid or parathyroid gland surgery. Hypoparathyroidism is a cardinal feature of autoimmune endocrinopathies **(Chap. 400)**, responsible for about a quarter of hypoparathyroidism. Even late-onset adult idiopathic hypoparathyroidism—unrelated to neck surgery or autoimmune polyglandular syndrome type 1 (APS-1)—frequently exhibits autoimmunity that targets CaSR and other parathyroid antigens. In patients with APS-1, anti-interferon and anti-NALP5 antibodies are frequently present. Hypoparathyroidism may be associated with infiltrative diseases such sarcoidosis or IgG4-related disease, iron overload (hemochromatosis, transfusion-dependent thalassemia), copper overload (Wilson's syndrome), genetic disorders (e.g., DiGeorge's syndrome), or parathyroid involution following neck irradiation. During the COVID-19 pandemic, via unknown mechanisms, transient hypoparathyroidism and even spontaneous resolution of pHPT were reported. Rare, congenital hypoparathyroidism has been well documented due to inactivating mutations in *GCM2*, *GATA3*, or even the *PTH* gene itself. Conversely, activating mutations in the CaSR, or in the GNA11 protein that mediates CaSR signaling (autosomal dominant hypocalcemia), also cause congenital hypoparathyroidism. More commonly, impaired PTH secretion may be secondary to magnesium deficiency (via intracellular actions). Hypomagnesemia with hypocalcemia can arise from drug-induced renal loss including chronic alcohol abuse, oncotherapeutics (cisplatin, anti–epidermal growth factor [EGF] antibodies), proton pump inhibitors, or rare autoimmunity that causes tubulointerstitial nephritis with claudin-16 inhibition. Because extracellular magnesium can bind the CaSR with an affinity about one-third that of calcium, severe hypermagnesemia arising from obstetric magnesium infusions for preeclampsia or preterm labor can also suppress PTH release; this rarely causes symptomatic hypoparathyroidism that must be managed until the magnesium infusion is discontinued.

Vitamin D deficiency, impaired 1,25(OH)$_2$D production due to chronic renal insufficiency, or vitamin D resistance also cause hypocalcemia. However, the degree of hypocalcemia in these disorders is most often not as severe as that seen with hypoparathyroidism; this is because the parathyroids are capable of mounting a compensatory increase in PTH secretion that buffers the extent of hypocalcemia to the detriment of bone. By increasing calcium binding to albumin, both respiratory alkalosis (e.g., hyperventilation) and metabolic alkalosis (e.g., intractable vomiting) can also cause hypocalcemia. When tetany arises with alkalosis, another cause of impaired calcium homeostasis is frequently present. Hypocalcemia may also occur in conditions associated with severe tissue injury such as burns, rhabdomyolysis, tumor lysis, or pancreatitis. The cause of hypocalcemia in these settings may include a combination of hyperphosphatemia, tissue deposition of calcium, and impaired PTH secretion.

CLINICAL MANIFESTATIONS

Patients with hypocalcemia may be asymptomatic if the decreases in serum calcium are relatively mild and chronic, or they may present with life-threatening complications. Moderate to severe hypocalcemia is associated with paresthesias, usually of the fingers, toes, and circumoral regions, and is caused by increased neuromuscular irritability. On physical examination, a Chvostek's sign (twitching of the ipsilateral perioral levator and zygomaticus muscles in response to gentle tapping of the facial nerve, just anterior to the ear or between the corner of the mouth and zygomatic arch) may be elicited, although it is also present in ~10% of normal individuals. As a more definitive sign of neuromuscular irritability and impending tetany, carpal spasm may be induced by inflation of an upper arm blood pressure cuff to 20 mmHg above the patient's systolic blood pressure for 3 min (Trousseau's sign; ipsilateral thumb apposition with wrist flexion). With profound vitamin D insufficiency or hypomagnesemia, proximal myopathy is frequently evident with difficulty rising from a chair while sitting with arms crossed over the chest. Severe hypocalcemia can induce seizures, carpopedal spasm, bronchospasm, laryngospasm, and prolongation of the QTc interval with risk of ventricular tachyarrhythmia (cardiac Torsade) and sudden death.

DIAGNOSTIC APPROACH

In addition to measuring serum calcium, it is important to determine albumin, phosphorus, and magnesium levels. As for the evaluation of hypercalcemia, determining the PTH level is central to the evaluation of hypocalcemia. A suppressed (or "inappropriately low") PTH level in the setting of hypocalcemia establishes impaired PTH secretion (hypoparathyroidism) as the cause of the hypocalcemia. The history will often elicit the underlying cause (e.g., prior neck surgery with parathyroid gland destruction). By contrast, an elevated PTH level (secondary hyperparathyroidism, a compensatory change in response to low calcium) should direct attention to the vitamin D axis or primary disease of the small intestine (e.g., short bowel syndrome due to Roux-en-Y bariatric surgery) as the cause of hypocalcemia. These latter patients are at great risk for hypocalcemia with even modest hypoparathyroidism following neck surgery and can be very difficult to manage. Nutritional vitamin D deficiency is best assessed by obtaining serum

25(OH)D levels, which reflect vitamin D stores. Urinary calcium levels in 24-hour collections are low with both vitamin D deficiency and primary intestinal disease-causing severe calcium malabsorption. In the setting of nonnutritional rickets, with suspected vitamin D resistance, serum 1,25(OH)$_2$D levels are informative.

TREATMENT

Hypocalcemia

The approach to treatment depends on the severity of the hypocalcemia, the rapidity with which it develops, and the accompanying complications (e.g., tetany, seizures, QTc prolongation). Acute, symptomatic hypocalcemia is initially managed with calcium gluconate, a 10-mL ampule of 10% wt/vol (90 mg or 2.2 mmol) diluted in 50 mL of 5% dextrose or 0.9% sodium chloride and given intravenously over 5–10 min with telemetry. After one or two additional ampules are given at 10-60 minute intervals as needed to initially resolve symptoms, while transitioning to oral medication symptomatic hypocalcemia often requires a period of continuous intravenous calcium infusion (often starting at 1 mg/min elemental calcium for adults) that is titrated to symptoms, ECG, and blood calcium levels targeting the lower limit of normal. Telemetry with serial exams and blood calcium levels every 4–6 hours should be closely monitored. Accompanying hypomagnesemia, if present, should be treated with appropriate magnesium supplementation, with initiation of chronic-phase treatment once tetany has resolved, the patient is stable and able to safely take oral medication, and the QTc has normalized.

Chronic hypocalcemia due to hypoparathyroidism is treated with oral calcium supplements (1000–3000 mg/d elemental calcium in divided doses), with careful titration of oral calcitriol [1,25(OH)$_2$D, 0.25–1 μg/d] to achieve albumin-corrected low-normal serum calcium levels. Because calcium resorption in the distal convoluted tubule (about 10% of the filtered load) is dependent upon PTH, normalization of calcium to above the low-normal range in the absence of PTH replacement increases hypercalciuria with risk of nephrolithiasis and nephrocalcinosis. Urine 24-hour calcium targets are under 250 mg (women) to 300 mg (men). Adequate vitamin D$_3$ nutrition (1000–2000 IU daily for most adults) must still be maintained and monitored every 6–12 months, measuring serum 25(OH)D levels.

PTH (1-84) (Natpara), approved by the Food and Drug Administration for treatment of hypoparathyroidism, has now been discontinued by the manufacturer. PTH (1-34), also known as teriparatide, is approved for osteoporosis, but has been successfully used off-label (20 μg subcutaneously twice daily) for treating refractory hypoparathyroidism. Dosing may need to be decreased in patients with chronic renal insufficiency to avoid hypercalcemia. Palopegteriparatide, a pegylated prodrug form of teriparatide with pharmacokinetics that enable daily subcutaneous dosing for PTH replacement, is now under consideration for regulatory approval. Routine daily vitamin D (~1000 IU D$_3$) and calcium (~1000 mg elemental calcium) nutrition suffices, and hypercalciuria should be absent, with PTH replacement strategies. Hypocalcemia from vitamin D deficiency is best treated using vitamin D supplementation, with the dose depending on the severity of the deficit and the underlying cause. Thus, nutritional vitamin D deficiency generally responds to moderate doses of oral vitamin D (e.g., ergocalciferol at 50,000 IU, 2–3 times per week for several months), whereas vitamin D deficiency due to severe enteric malabsorption may require much higher doses (e.g., ergocalciferol carefully titrated up to 100,000 IU/d), as is often the case following Roux-en-Y bariatric surgery. In the setting of intestinal disease, calcium supplementation as calcium citrate is preferred because it is better absorbed, and this formulation increases urinary citrate to mitigate the risk of nephrolithiasis in this setting due to increased dietary oxalate absorption. Serum calcium, phosphate, and PTH should be monitored initially every 4 weeks when treating the hypocalcemia of severe vitamin D insufficiency or gastrointestinal calcium malabsorption, with the goal of normalizing serum biochemistries and 24-hour urine calcium levels to low normal.

GLOBAL CONSIDERATIONS

In countries with limited access to health care or screening laboratory testing of serum calcium levels, pHPT often presents in its advanced form with severe skeletal complications (osteitis fibrosa cystica), in contrast to the incidental finding of asymptomatic hypercalcemia common in developed countries. Climate change increases the negative impacts of hypercalcemia on risks for nephrolithiasis and heat-related illness with dehydration in those with previously asymptomatic disease. In addition, vitamin D deficiency is paradoxically common in some countries despite extensive sunlight (e.g., India) due to poor dietary vitamin D intake and avoidance of sun exposure.

Acknowledgment

The author gratefully acknowledges the contributions of Dr. Sundeep Khosla to this chapter in previous editions of Harrison's.

FURTHER READING

El-Hajj Fuleihan G: Treatment of hypercalcemia of malignancy in adults: An Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 108:507, 2023.

Hannan FM et al: The calcium-sensing receptor in physiology and in calcitropic and noncalcitropic diseases. Nat Rev Endocrinol 15:33, 2018.

Kahn AA et al: Evaluation and management of hypoparathyroidism. Summary statement and guidelines from the Second International Workshop. J Bone Miner Res 37:2568, 2022.

Minisola S et al: Epidemiology, pathophysiology, and genetics of primary hyperparathyroidism. J Bone Miner Res 37:2315, 2022.

Motlaghzadeh Y et al: Rare causes of hypercalcemia: 2021 update. J Clin Endocrinol Metab 106:3113, 2021.

Walker MD, Shane E: Hypercalcemia. A review. JAMA 328:1624, 2022.

58

Acidosis and Alkalosis

Thomas D. DuBose, Jr.

NORMAL ACID-BASE HOMEOSTASIS

Systemic arterial pH is maintained between 7.35 and 7.45 by extracellular and intracellular chemical buffering together with respiratory and renal regulatory mechanisms. The control of arterial CO$_2$ tension (PaCO_2) by the central nervous system (CNS) and respiratory system and the control of plasma bicarbonate by the kidney stabilize the arterial pH by excretion or retention of acid or alkali. The metabolic and respiratory components that regulate systemic pH are described by the Henderson-Hasselbalch equation and solved for pH when the solubility of CO$_2$ is considered (dissolved CO$_2$ in mmol/L = 0.03 × PaCO_2 in mmHg), at a pK' of 6.1:

$$\text{pH} = \text{p}K' + \log_{10}\frac{[\text{HCO}_3^-]}{\alpha_{\text{CO}_2}\text{PCO}_2}$$

Under most circumstances, CO$_2$ production and excretion are matched, and the usual steady-state PaCO_2 is maintained at ~40 mmHg. Underexcretion of CO$_2$ produces hypercapnia, and overexcretion causes hypocapnia. Nevertheless, production and excretion are again matched at a new steady-state Paco$_2$. Therefore, the Paco$_2$ is regulated primarily by neural respiratory factors and is not subject to regulation by the rate of CO$_2$ production. Hypercapnia is usually the result of hypoventilation rather than of increased CO$_2$ production. Increases or decreases in Paco$_2$ represent derangements of neural respiratory control or are due to compensatory changes in response to a primary alteration in the plasma [HCO$_3^-$].

DIAGNOSIS OF GENERAL TYPES OF DISTURBANCES

The most common clinical disturbances are simple acid-base disorders; i.e., metabolic acidosis or alkalosis or respiratory acidosis or alkalosis occurring individually. Recognition of simple acid-base disorders requires appreciation of the limits of physiologic compensation for a primary disturbance.

SIMPLE ACID-BASE DISORDERS

Primary respiratory disturbances (primary changes in Paco$_2$) invoke compensatory metabolic responses (secondary changes in [HCO$_3^-$]), and primary metabolic disturbances elicit predictable compensatory respiratory responses (secondary changes in Paco$_2$). Physiologic compensation can be predicted from the relationships displayed in **Table 58-1**. In general, with one exception, compensatory responses return the pH toward, but not to, the normal value. Chronic respiratory alkalosis when prolonged is an exception to this rule and may return the pH to a normal value. Metabolic acidosis due to an increase in endogenous acid production (e.g., ketoacidosis or lactic acid acidosis) lowers extracellular fluid [HCO$_3^-$] and decreases extracellular pH. This change stimulates the medullary chemoreceptors to increase ventilation and to return the ratio of [HCO$_3^-$] to Paco$_2$, and, thus, pH, toward, but not typically to, the normal value. The degree of respiratory compensation expected in a metabolic acidosis can be predicted from the relationship: Paco$_2$ = (1.5 × [HCO$_3^-$]) + 8 ± 2 (Winter's equation). For example, applying this equation, a patient with metabolic acidosis and [HCO$_3^-$] of 12 mmol/L would be expected to have a Paco$_2$ of approximately 26 mmHg. Therefore, if values for Paco$_2$ were <24 or >28 mmHg, values that exceed the boundaries for compensation for a simple disorder, a *mixed* disturbance should be recognized (metabolic acidosis plus respiratory alkalosis or metabolic acidosis plus respiratory acidosis, respectively). Compensatory responses for primary metabolic disorders move the Paco$_2$ in the same direction as the change in [HCO$_3^-$], while compensation for primary respiratory disorders moves the [HCO$_3^-$] in the same direction as the primary change in Paco$_2$ (Table 58-1). Therefore, changes in Paco$_2$ and [HCO$_3^-$] in **opposite directions** (i.e., Paco$_2$ or [HCO$_3^-$] is increased, but the accompanying value is decreased) indicate a **mixed acid-base disturbance**. Another way to judge the appropriateness of the response in [HCO$_3^-$] or Paco$_2$ is to use an acid-base nomogram **(Fig. 58-1)**. While the shaded areas of the nomogram show the 95% confidence limits for physiologic compensation in simple disturbances, finding acid-base values within the shaded area does not necessarily rule out a mixed disturbance. Imposition of one disorder over another may result in values lying within the area of a third. Thus, the nomogram, while convenient, is not a substitute for the equations in Table 58-1.

TABLE 58-1 Prediction of Compensatory Responses to Simple Acid-Base Disturbances and Pattern of Changes

		RANGE OF VALUES		
DISORDER	**PREDICTION OF COMPENSATION**	**pH**	**HCO$_3^-$**	**Paco$_2$**
Metabolic acidosis	Paco$_2$ = (1.5 × HCO$_3^-$) + 8 ± 2 *or* Paco$_2$ will ↓ 1.25 mmHg per mmol/L ↓ in [HCO$_3^-$] *or* Paco$_2$ = [HCO$_3^-$] + 15	Low	Low	Low
Metabolic alkalosis	Paco$_2$ will ↑ 0.75 mmHg per mmol/L ↑ in [HCO$_3^-$] *or* Paco$_2$ will ↑ 6 mmHg per 10 mmol/L ↑ in [HCO$_3^-$] *or* Paco$_2$ = [HCO$_3^-$] + 15	High	High	High
Respiratory alkalosis		High	Low	Low
Acute	[HCO$_3^-$] will ↓ 0.2 mmol/L per mmHg ↓ in Paco$_2$			
Chronic	[HCO$_3^-$] will ↓ 0.4 mmol/L per mmHg ↓ in Paco$_2$			
Respiratory acidosis		Low	High	High
Acute	[HCO$_3^-$] will ↑ 0.1 mmol/L per mmHg ↑ in Paco$_2$			
Chronic	[HCO$_3^-$] will ↑ 0.4 mmol/L per mmHg ↑ in Paco$_2$			

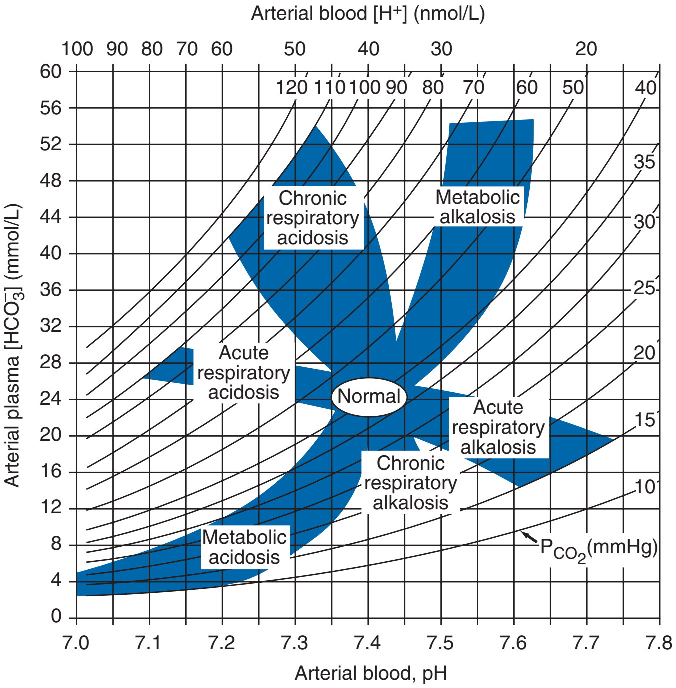

FIGURE 58-1 Acid-base nomogram. Shown are the 90% confidence limits (range of values) of the normal respiratory and metabolic compensations for primary acid-base disturbances. *(Reproduced with permission from LL Hamm and TD DuBose Jr, in Alan S.L. Yu, et al (eds): Brenner and Rector's The Kidney, 11th ed. Philadelphia, Elsevier, 2020.)*

MIXED ACID-BASE DISORDERS

Acid-base disorders in this category are defined as independently coexisting disorders, not merely compensatory responses. These types of disturbances are often seen in critically ill patients and can lead to dangerous extremes of pH **(Table 58-2)**. The diagnosis of mixed acid-base disorders requires consideration of the anion gap (AG). To be accurate, the AG requires the presence of, or correction to, a normal serum albumin of 4.5 g/dL (see below, "Evaluate the Anion Gap"). If a patient with diabetic ketoacidosis (metabolic acidosis) and a high AG has an independent and concomitant respiratory disorder (e.g., pneumonia), the latter may lead to a superimposed respiratory acidosis or alkalosis and the Paco$_2$ will deviate from the predicted value for the response to a pure high-AG metabolic acidosis (Table 58-2). Patients with underlying chronic obstructive pulmonary disease may not respond to metabolic acidosis with an appropriate ventilatory response owing to insufficient respiratory reserve (Table 58-2). The combined presence of respiratory acidosis and metabolic acidosis can lead to severe acidemia. In contrast, when metabolic acidosis and metabolic alkalosis coexist in the same patient, the pH may be in the normal range. In this

TABLE 58-2 Clinical Examples of Mixed Acid-Base Disorders

Mixed Metabolic and Respiratory
Metabolic acidosis—respiratory alkalosis
Key: High-AG metabolic acidosis; prevailing Paco_2 *below* predicted value (Table 58-1)
Example: Na^+, 140; K^+, 4.0; Cl^-, 106; HCO_3^-, 14; AG, 20; Paco_2, 24; pH, 7.39 (etiology: lactic acidosis, sepsis in ICU)
Metabolic acidosis—respiratory acidosis
Key: High-AG metabolic acidosis; prevailing Paco_2 *above* predicted value (Table 58-1)
Example: Na^+, 140; K^+, 4.0; Cl^-, 102; HCO_3^-, 18; AG, 20; Paco_2, 42; pH, 7.25 (etiology: severe pneumonia or pulmonary edema)
Metabolic alkalosis—respiratory alkalosis
Key: Paco_2 does not increase as predicted; pH higher than expected
Example: Na^+, 140; K^+, 4.0; Cl^-, 91; HCO_3^-, 33; AG, 16; Paco_2, 38; pH, 7.56 (end-stage liver disease with ascites in patient receiving diuretics)
Metabolic alkalosis—respiratory acidosis
Key: Paco_2 higher than predicted; pH normal although both Paco_2 and HCO_3^- abnormal
Example: Na^+, 140; K^+, 3.5; Cl^-, 88; HCO_3^-, 42; AG, 10; Paco_2, 67; pH, 7.42 (COPD in patient receiving diuretics)
Mixed Metabolic Disorders
Metabolic acidosis—metabolic alkalosis
Key: Only detectable if in patient with high-AG acidosis; ΔAG (10) >> ΔHCO_3 (0)
Example: Na^+, 140; K^+, 3.0; Cl^-, 95; HCO_3^-, 25; AG, 20; Paco_2, 40; pH, 7.42 (uremia with vomiting)
Metabolic acidosis—metabolic acidosis
Key: Mixed high-AG—normal-AG acidosis; ΔHCO_3^- accounted for by combined change in ΔAG and ΔCl^-
Example: Na^+, 135; K^+, 3.0; Cl^-, 110; HCO_3^-, 10; AG, 15; Paco_2, 25; pH, 7.20 (diarrhea and lactic acidosis, toluene toxicity, treatment of diabetic ketoacidosis)

Abbreviations: AG, anion gap; COPD, chronic obstructive pulmonary disease; ICU, intensive care unit.

circumstance, it is the recognition of an elevated AG (see below) that denotes the existence of an accompanying metabolic acidosis. Assuming a normal value for the AG of 10 mmol/L, incongruity in the ΔAG (existing AG minus normal AG) and the ΔHCO_3^- (normal value of 25 mmol/L minus abnormal HCO_3^- in the patient) indicates the presence of a mixed high-gap acidosis—metabolic alkalosis (see example below). A diabetic patient with ketoacidosis may have acute or chronic kidney failure resulting in a combination of metabolic acidoses from accumulation of both ketoacids and uremic acids. Patients who have ingested an overdose of drug combinations such as sedatives and salicylates may have mixed disturbances as a result of the acid-base response to the individual drugs (metabolic acidosis mixed with respiratory acidosis or respiratory alkalosis, respectively). Triple acid-base disturbances are more complex. For example, patients with metabolic acidosis due to alcoholic ketoacidosis may develop metabolic alkalosis due to vomiting and superimposed respiratory alkalosis due to the hyperventilation of hepatic dysfunction or alcohol withdrawal.

APPROACH TO THE PATIENT

Acid-Base Disorders

The accurate diagnosis of acid-base disorders requires adherence to a stepwise approach **(Table 58-3)**. Blood for plasma electrolytes and arterial blood gases should be drawn simultaneously, prior to therapy. An increase in $[\text{HCO}_3^-]$ occurs with either metabolic alkalosis or respiratory acidosis. Conversely, a decrease in $[\text{HCO}_3^-]$ occurs with either metabolic acidosis or respiratory alkalosis. In the determination of arterial blood gases by the clinical laboratory, both pH and Paco_2 are measured, and the $[\text{HCO}_3^-]$ is calculated from the Henderson-Hasselbalch equation. This *calculated* value should be compared with the *measured* $[\text{HCO}_3^-]$ (or total CO_2) on the electrolyte panel. These two values should agree within ±2 mmol/L. If the values do not agree, the blood samples may not have been drawn simultaneously, or a laboratory error may be present. After verifying the blood acid-base values, the precise acid-base disorder can then be classified.

TABLE 58-3 Steps in Accurate Diagnosis of Acid-Base Disorders

1. Obtain arterial blood gas (ABG) and venous electrolytes simultaneously.
2. Calculated $[\text{HCO}_3^-]$ on ABG and measured value on electrolyte panel should be approximately same; if not, suspect lab error or sampling error.
3. Assess anion gap (AG); correct to albumin concentration of 4.5 g/dL if hypoalbuminemia; high AG present if AG >10 mEq/L.
4. Known causes of high-AG acidosis (Table 58-4; ketoacidosis, lactic acid acidosis, advanced kidney disease, or toxic alcohol ingestion).
5. Known causes of nongap acidosis (Table 58-5; bicarbonate loss from gastrointestinal tract, renal tubular acidosis).
6. Estimate predicted compensatory response (Table 58-1).
7. Compare delta values (ΔAG and ΔHCO_3^-).
8. Compare change in $[\text{Cl}^-]$ with change in $[\text{Na}^+]_s$ on the electrolyte panel.

EVALUATE THE ANION GAP

Evaluations of acid-base disorders should begin with appreciation of the patient's AG. The AG is calculated, either by the clinical laboratory or the clinician, as follows: $\text{AG} = \text{Na}^+ - (\text{Cl}^- + \text{HCO}_3^-)$. The value for plasma $[\text{K}^+]$ is typically omitted from the calculation of the AG in the United States. The "normal" value for the AG reported by clinical laboratories has declined with improved methodology for measuring plasma electrolytes and ranges from 6–12 mmol/L, with an average of approximately 10 mmol/L. The unmeasured anions normally present in plasma include anionic proteins (e.g., albumin), phosphate, sulfate, and organic anions. When acid anions, such as acetoacetate and lactate, accumulate in extracellular fluid, the AG increases, causing a **high-AG acidosis**. An increase in the AG is most often due to an increase in unmeasured anions but, less commonly, may be due to a decrease in unmeasured cations (calcium, magnesium, potassium). In addition, the AG may increase with an increase in anionic albumin (e.g., severe dehydration). A decrease in the AG can be due to (1) an increase in unmeasured cations; (2) the addition to the blood of abnormal cations, such as lithium (lithium intoxication) or cationic immunoglobulins (plasma cell dyscrasias); (3) a reduction in the plasma anion albumin concentration (nephrotic syndrome, liver disease, or malabsorption); or (4) hyperviscosity and severe hyperlipidemia, which can lead to an underestimation of sodium and chloride concentrations. Since a normal AG of the normal AG of approximately 10 mmol/L assumes that the serum albumin is normal if hypoalbuminemia is present, the value for the calculated AG must be corrected. For each g/dL of serum albumin below the normal value (4.5 g/dL), 2.5 mmol/L should be added to the reported (uncorrected) AG. Therefore, in a patient with a serum albumin of 2.5 g/dL (2 g/dL below the normal value) and an uncorrected AG of 15, the corrected AG is calculated by adding 5 mmol/L (2.5 × 2 = 5); thus, adding this value to the calculated AG (5 + 15), a corrected AG of 20 mmol/L is appreciated. Since clinical laboratories do not correct the AG for coexisting hypoalbuminemia and typically report the uncorrected value, the attention of the clinician to the prevailing serum albumin concentration is necessary. The clinical disorders that may cause a high-AG acidosis are displayed in Table 58-4.

A high AG is usually due to accumulation of non–chloride-containing acids that contain inorganic (phosphate, sulfate), organic (ketoacids, lactate, uremic organic anions), exogenous (salicylate or ingested toxins with organic acid production), or unidentified anions. The high AG is meaningful even if the $[\text{HCO}_3^-]$ or pH is normal. Simultaneous metabolic acidosis of the high-AG variety plus either chronic respiratory acidosis or metabolic alkalosis represents a situation in which [HCO_3^-] may be normal or even high (Table 58-3). In cases of high-AG metabolic acidosis, it is valuable to compare the decline in [HCO_3^-] from the normal value (ΔHCO_3^-: 25 − patient's [HCO_3^-]) with the increase in the AG (ΔAG: patient's AG − 10).

Similarly, normal values for [HCO_3^-], Pa_{CO_2}, and pH do not ensure the absence of an acid-base disturbance. For example, an alcoholic who has been vomiting prior to admission may develop a metabolic alkalosis with a pH of 7.55, Pa_{CO_2} of 47 mmHg, [HCO_3^-] of 40 mmol/L, [Na^+] of 135, [Cl^-] of 80, and [K^+] of 2.8. If such a patient were then to develop a superimposed alcoholic ketoacidosis with a β-hydroxybutyrate concentration of 15 mmol/L, the arterial pH would fall to 7.40, the [HCO_3^-] to 25 mmol/L, and the Pa_{CO_2} to 40 mmHg. Although these values are normal, the AG is significantly elevated at 30 mmol/L, documenting that a mixed metabolic alkalosis and metabolic acidosis coexist. A mixture of high-gap acidosis and metabolic alkalosis is recognized easily by comparing the differences (Δ values) in the normal to prevailing patient values. In this example, the ΔHCO_3^- is 0 (25 − 25 mmol/L), but the ΔAG is 20 (30 − 10 mmol/L). Therefore, 20 mmol/L is unaccounted for in the Δ/Δ value (ΔAG to ΔHCO_3^-).

METABOLIC ACIDOSIS

Metabolic acidosis can occur because of an increase in endogenous acid production (such as lactate and ketoacids), loss of bicarbonate (as in diarrhea), or accumulation of endogenous acids because of inappropriately low excretion of net acid by the kidney (as in chronic kidney disease). Metabolic acidosis has profound effects on the respiratory, cardiac, and nervous systems. The fall in blood pH is accompanied by a characteristic increase in ventilation. Intrinsic cardiac contractility may be depressed, but inotropic function can be normal because of catecholamine release. Both peripheral arterial vasodilation and central venoconstriction may be present; accordingly, the decrease in central and pulmonary vascular compliance predisposes to pulmonary edema with even minimal volume overload. CNS function is depressed, with headache, lethargy, stupor, and, in some cases, coma. Glucose intolerance may also occur.

There are two major categories of clinical metabolic acidosis: high-AG and non-AG acidosis (Tables 58-3 and **58-4**). The presence of metabolic acidosis, a normal AG, and hyperchloremia denotes the presence of a non-AG metabolic acidosis.

TREATMENT
Metabolic Acidosis

Treatment of metabolic acidosis with alkali should be reserved for severe acidemia except when the patient has no "potential HCO_3^-" in plasma. The potential [HCO_3^-] can be estimated from the increment (Δ) in the AG (ΔAG = patient's AG − 10), only if the acid anion that has accumulated in plasma is metabolizable (i.e., β-hydroxybutyrate, acetoacetate, and lactate). Conversely, nonmetabolizable anions that may accumulate in advanced-stage chronic kidney disease or after toxin ingestion are not metabolizable and do not represent "potential" HCO_3^-. In patients with acute kidney failure or acute-on-chronic kidney failure, improvement in kidney function after volume resuscitation may improve the serum [HCO_3^-], but this is a slow and unpredictable process. Consequently, patients with a non-AG acidosis (hyperchloremic acidosis) or an AG acidosis attributable to a nonmetabolizable anion due to advanced kidney failure ("uremic" acidosis) should receive alkali therapy, either PO ($NaHCO_3$ tablets or Shohl's solution) or IV ($NaHCO_3$), in an amount necessary to slowly increase the plasma [HCO_3^-] to a target value of 22 mmol/L. Importantly, overcorrection should be avoided.

Bicarbonate therapy in diabetic ketoacidosis (DKA) is reserved for adult patients with severe acidemia (pH <7.00) and/or evidence of shock. In such circumstances, bicarbonate may be administered IV, as a slow infusion of 50 meq of $NaHCO_3$ diluted in 300 mL of a saline solution, over 30–45 min, during the initial 1–2 h of therapy. Bolus administration should be avoided. Administration of $NaHCO_3$ requires careful monitoring of plasma electrolytes during the course of therapy because of the risk for hypokalemia as urine output is reestablished. A reasonable initial goal in DKA is to increase the [HCO_3^-] to a target of 10–12 mmol/L and the pH to approximately 7.20, but definitely not to increase these values to normal.

TABLE 58-4 Causes of High-Anion Gap Metabolic Acidosis

Lactic acidosis	Toxins
Ketoacidosis	Ethylene glycol
Diabetic	Methanol
Alcoholic	Salicylates
Starvation	Propylene glycol
	Pyroglutamic acid (5-oxoproline)
	Kidney failure (acute and chronic)

HIGH-ANION GAP ACIDOSES

APPROACH TO THE PATIENT
High-Anion Gap Acidoses

There are four principal causes of a high-AG acidosis: (1) lactic acidosis, (2) ketoacidosis, (3) ingested toxins, and (4) acute and chronic kidney failure (Table 58-4). Initial screening to differentiate the high-AG acidoses should include (1) a careful history of whether drug or toxin ingestion is present and measurement of arterial blood gas to detect coexistent respiratory alkalosis (e.g., salicylate intoxication); (2) a history of diabetes mellitus (DKA); (3) evidence of alcohol abuse or increased levels of β-hydroxybutyrate (alcoholic ketoacidosis); (4) a history of progressive chronic kidney disease (CKD) and an increase in the patient's baseline blood urea nitrogen (BUN) and creatinine values (uremic acidosis); (5) inspection of the urine for oxalate crystals (ethylene glycol ingestion); and (6) recognition of the numerous clinical settings in which lactate levels may be increased (hypotension, shock, cardiac failure, leukemia, cancer, and drug or toxin ingestion).

Lactic Acidosis An increase in plasma L-lactate may be secondary to poor tissue perfusion ("type A" lactic acidosis)—circulatory insufficiency (shock, cardiac failure), severe anemia, mitochondrial enzyme defects, and inhibitors (carbon monoxide, cyanide)—or to aerobic disorders ("type B" lactic acidosis)—malignancies, nucleoside analogue reverse transcriptase inhibitors in HIV, diabetes mellitus, kidney or hepatic failure, thiamine deficiency, severe infections (cholera, malaria), seizures, or drugs/toxins (biguanides, ethanol, and the toxic alcohols: ethylene glycol, methanol, or propylene glycol). Unrecognized bowel ischemia or infarction in a patient with severe atherosclerosis or cardiac decompensation receiving vasopressors is a relatively common cause of lactic acidosis in elderly patients. Pyroglutamic acidemia may occur in critically ill patients receiving acetaminophen, because of depletion of glutathione and accumulation of 5-oxyprolene. D-Lactic acid acidosis, which may be associated with jejunoileal bypass, short bowel syndrome, or intestinal obstruction, is due to formation of D-lactate by gut bacteria.

APPROACH TO THE PATIENT
L-Lactic Acid Acidosis

The overarching goal of treatment in lactic acidosis is to correct the underlying condition that disrupts lactate metabolism; e.g., tissue perfusion should be restored when inadequate, but vasoconstrictors should be avoided, if possible, or used cautiously, because they may worsen tissue perfusion. Alkali therapy is generally advocated for acute, severe acidemia (pH <7.00) to improve cardiovascular function. However, $NaHCO_3$ therapy may paradoxically depress cardiac performance and exacerbate acidosis by enhancing lactate production (HCO_3^- stimulates phosphofructokinase). While the use of alkali in moderate lactic acidosis is controversial, it is generally agreed that attempts to return the pH or $[HCO_3^-]$ to normal by administration of exogenous $NaHCO_3$ are deleterious. A reasonable approach with severe acidemia is to infuse sufficient $NaHCO_3$ to raise arterial pH to no more than 7.2 or the $[HCO_3^-]$ to no more than 12 mmol/L.

$NaHCO_3$ therapy can cause fluid overload, hypercapnia, and hypertension because the amount required can be massive when accumulation of lactic acid is relentless. Fluid administration is poorly tolerated, especially in the oliguric patient, when central venoconstriction coexists. If the underlying cause of the lactic acidosis can be remedied, blood lactate will be converted to HCO_3^- and may result in an overshoot alkalosis if exogenous $NaHCO_3$ has been administered excessively.

Ketoacidosis • DIABETIC KETOACIDOSIS (DKA) This condition is caused by increased fatty acid metabolism and the accumulation of ketoacids (acetoacetate and β-hydroxybutyrate). DKA usually occurs in insulin-dependent diabetes mellitus in association with cessation of insulin administration or an intercurrent illness such as an infection, gastroenteritis, pancreatitis, or myocardial infarction, which increases insulin requirements temporarily and acutely. DKA is characterized by hyperglycemia, ketonemia, and a high-AG acidosis. Nevertheless, the plasma glucose may be normal or only slightly elevated in the setting of starvation ketoacidosis or in diabetics receiving an agent that inhibits the proximal tubule sodium-glucose co-transporter 2 (SGLT2 inhibitors) (euglycemic DKA [eDKA]). These agents cause glycosuria, an osmotic diuresis, volume depletion, and decreased plasma glucose. Although the accumulation of ketoacids in plasma accounts for the increment in the AG in both classical DKA and eDKA, the plasma glucose is elevated in classical DKA but is typically in the normal range in eDKA. Measurement of urine ketones (by the dipstick nitroprusside reaction) does not detect β-hydroxybutyrate accurately and may underestimate the degree of ketosis (see below). Excretion of ketoacids obligates the excretion of cations, such as Na^+ and K^+, contributing to volume depletion and Cl^- retention. In some circumstances, a mixed non-AG–high-AG acidosis may occur simultaneously and is recognized when the ΔHCO_3^- exceeds the ΔAG. It should be noted that bicarbonate therapy is rarely necessary in DKA in adults, except with extreme acidemia (pH <7.00) or if the patient is in shock. If administered, $NaHCO_3$ should be given in only limited amounts because of the risk for cerebral edema. Patients with DKA are typically volume depleted and require fluid resuscitation with isotonic saline. Volume overexpansion with isotonic saline should be avoided, however, because aggressive saline administration may cause overt volume overload and/or hyperchloremic acidosis during or following treatment of DKA. Regular insulin should be administered IV as an initial bolus of 0.1 U/kg followed by an infusion of 0.1 U/kg/h until the AG returns to normal; see Chap. 417 for more detail.

ALCOHOLIC KETOACIDOSIS (AKA) AKA is usually associated with chronic alcoholism, binge drinking, vomiting, abdominal pain, poor nutrition, and volume depletion. The glucose concentration is variable, and acidosis may be severe because of elevated ketones, predominantly β-hydroxybutyrate. The presence of a high-AG acidosis, in the absence of hyperglycemia, in a patient with chronic alcoholism suggests the diagnosis of AKA. Mixed acid-base disorders are common in AKA. Hypoperfusion may enhance lactic acid production (mixed high-AG acidosis), chronic respiratory alkalosis may accompany liver disease (mixed high-AG acidosis and respiratory alkalosis), and metabolic alkalosis can result from vomiting (mixed high-AG acidosis and metabolic alkalosis: ΔAG exceeds ΔHCO_3^-). As the circulation is restored by administration of IV fluids, the preferential accumulation of β-hydroxybutyrate is then shifted to acetoacetate. This explains the common clinical observation of an increasingly positive nitroprusside reaction (ketones) as the circulation is restored. The nitroprusside reaction can detect acetoacetic acid but not β-hydroxybutyrate, so that the degree of ketosis and ketonuria can not only change with therapy but can also be underestimated initially. Therefore, the plasma β-hydroxybutyrate level should be measured specifically. Patients with AKA usually present with relatively normal kidney function, as opposed to DKA, where kidney function is often compromised because of volume depletion (osmotic diuresis) or diabetic nephropathy. The AKA patient with normal kidney function may excrete relatively large quantities of ketoacids and retain Cl^- and, therefore, may have a mixed high-AG–non-AG metabolic acidosis (ΔHCO_3^- exceeds ΔAG).

TREATMENT

Alcoholic Ketoacidosis

Extracellular fluid deficits almost always accompany AKA and should be repaired by IV administration, initially, of saline and glucose (5% dextrose in 0.9% NaCl). Hypophosphatemia, hypokalemia, and hypomagnesemia may coexist and should be monitored carefully and corrected when indicated. Hypophosphatemia typically emerges 12–24 h after admission and may be severe. Hypophosphatemia is exacerbated by glucose infusion, and, if severe, may induce marked muscle weakness, hemolysis, rhabdomyolysis, or respiratory arrest. Upper gastrointestinal hemorrhage, pancreatitis, and pneumonia may accompany this disorder.

Drug- and Toxin-Induced Acidosis • SALICYLATES (See also Chap. 469) Salicylate intoxication in adults usually causes respiratory alkalosis or a mixture of high-AG metabolic acidosis and respiratory alkalosis. Only a portion of the AG is due to salicylates. Lactic acid production is also often increased.

TREATMENT

Salicylate-Induced Acidosis

Vigorous gastric lavage with isotonic saline (not $NaHCO_3$) should be initiated immediately. All patients should receive at least one round of activated charcoal per nasogastric tube (1 g/kg up to 50 g). To facilitate excretion of salicylate in the acidotic patient, IV $NaHCO_3$ is administered in amounts adequate to alkalinize the urine (urine pH >7.5) and to maintain urine output. Raising urine pH from 6.5 to 7.5 increases salicylate clearance fivefold. Patients with coexisting respiratory alkalosis may also receive $NaHCO_3$, but if given, it should be administered cautiously to avoid excessive alkalemia. Acetazolamide may be administered with coexisting alkalemia, when an alkaline diuresis cannot be achieved, or to ameliorate volume overload associated with $NaHCO_3$ administration. Caution is needed because acetazolamide may cause systemic metabolic acidosis if the excreted HCO_3^- is not replaced, a circumstance that can markedly reduce salicylate clearance. **Hypokalemia should be anticipated** with vigorous bicarbonate therapy and should be treated promptly and aggressively. Glucose-containing fluids should be administered because of the danger of hypoglycemia. Excessive insensible fluid losses may cause severe volume depletion and hypernatremia. If acute kidney injury prevents rapid clearance of salicylate, hemodialysis should be performed against a standard bicarbonate dialysate ($[HCO_3^-]$ = 30–35 meq/L).

ALCOHOLS Under most physiologic conditions, sodium, urea, and glucose generate the osmotic pressure of blood. Plasma osmolality is calculated according to the following expression: $P_{osm} = 2Na^+ + Glu + BUN$ (all in mmol/L), or using conventional laboratory values in which glucose and BUN are expressed in mg/dL: $P_{osm} = 2Na^+ + Glu/18 + BUN/2.8$. The calculated and determined osmolality should agree within 10–15 mmol/kg H_2O. When the measured osmolality exceeds the calculated osmolality by >10–15 mmol/kg H_2O, one of two circumstances prevails. Either the serum sodium is spuriously low, as with hyperlipidemia or hyperproteinemia (pseudohyponatremia), or osmolytes other than sodium salts, glucose, or urea have accumulated in plasma. Examples of such osmolytes include mannitol, radiocontrast media, ethanol, isopropyl alcohol, ethylene glycol, propylene glycol, methanol, and acetone. In this situation, the difference between the calculated osmolality and the measured osmolality (*osmolar gap*) is proportional to the concentration of the unmeasured solute. With an appropriate clinical history and index of suspicion, identification of a serum osmolar gap is helpful in identifying the presence of toxic alcohol-associated AG acidosis. Three alcohols may cause fatal intoxications: ethylene glycol, methanol, and isopropyl alcohol. All cause an elevated osmolar gap, but only the first two cause a high-AG acidosis. Isopropyl alcohol ingestion does not typically elevate the AG unless extreme overdose causes hypotension and lactic acid acidosis.

ETHYLENE GLYCOL (See also Chap. 469) Ethylene glycol (EG) (commonly used in antifreeze, but also in brake fluid and windshield washer fluid deicers) is metabolized by alcohol dehydrogenase. Ingestion of EG leads to metabolic acidosis and severe damage to the CNS, heart, lungs, and kidneys. The combination of both a high AG and osmolar gap is highly suspicious for EG or methanol intoxication. The osmolar gap is determined by comparing the calculated and measured serum osmolality (Measured S_{osm} – Calculated S_{osm}). The serum osmolality is calculated as follows:

$$\text{Osmolality} = 2\,[\text{Na+}] + [\text{BUN}]/2.8 + [\text{Glucose}]/18$$

The serum osmolality is measured in the clinical laboratory most accurately by freezing point depression. The combination of a high AG and high osmolar gap in a patient suspected of EG ingestion should be taken as evidence of EG toxicity prior to measurement of EG levels, especially when the history is suspicious or highly suggestive of EG ingestion. Most importantly, in the face of an elevated osmolar gap and anion gap, treatment should not be delayed while awaiting return of ethylene glycol or methanol levels from the laboratory. The osmolar gap is typically elevated earlier than the AG, and as the osmolar gap declines, the AG usually increases. The increased AG and osmolar gap in EG intoxication are attributable to accumulation of EG and its metabolites, glycolate, oxalate, and other organic acids. Lactic acid production (L-lactate) increases secondary to inhibition of the tricarboxylic acid cycle and an altered intracellular redox state, thus contributing to the high AG. Acute tubule injury is caused initially by glycolate and later is amplified by tubule obstruction from oxalate crystals.

TREATMENT

Ethylene Glycol Intoxication

Therapy requires prompt institution of IV isotonic fluids, thiamine and pyridoxine supplements, fomepizole, and usually, hemodialysis. Ethanol is of historic interest and is no longer recommended as initial therapy unless fomepizole is not available. Both fomepizole and ethanol compete with EG for metabolism by alcohol dehydrogenase. Fomepizole (4-methylpyrazole; 15 mg/kg IV over 30 min as a loading dose, then 10 mg/kg for four doses every 12 h) is the agent of choice and offers the advantage of a predictable decline in EG levels without excessive obtundation, as commonly seen during ethyl alcohol infusion. Fomepizole should be continued until blood pH is normal or the osmolar gap is <10 mOsm/kg H_2O. Hemodialysis is indicated when the arterial pH is <7.3, a high-AG acidosis is present, the osmolar gap exceeds 20 mOsm/kg H_2O, or there is evidence of end organ damage such as CNS manifestations and kidney failure.

METHANOL (See also Chap. 469) The ingestion of methanol (wood alcohol) causes metabolic acidosis, and its metabolites formaldehyde and formic acid cause severe optic nerve and CNS damage. Lactic acid, ketoacids, and other unidentified organic acids contribute to the acidosis. Due to its low molecular mass (32 Da), an osmolar gap is present and may precede the elevation of the AG.

TREATMENT

Methanol Intoxication

Treatment of methanol intoxication is similar to that for EG intoxication, including general supportive measures, fomepizole, and hemodialysis.

PROPYLENE GLYCOL Propylene glycol is the vehicle used in the IV preparation of diazepam, lorazepam, phenobarbital, nitroglycerine, etomidate, enoximone, and phenytoin. Propylene glycol is generally safe for limited use in these IV preparations, but toxicity has been reported in the setting of the intensive care unit in patients receiving frequent or continuous administration, because propylene glycol may accumulate in plasma. This form of high-gap acidosis should be considered in patients with unexplained high-gap acidosis, hyperosmolality, and clinical deterioration, especially in the setting of treatment for alcohol withdrawal. Propylene glycol, like EG and methanol, is metabolized by alcohol dehydrogenase. With intoxication by propylene glycol, the first response is to stop the offending infusion. Additionally, fomepizole may be administered in severely acidotic patients.

ISOPROPYL ALCOHOL Ingested isopropanol is absorbed rapidly and may be fatal when as little as 150 mL of rubbing alcohol, solvent, or deicer is consumed. A plasma level >400 mg/dL is life-threatening. Isopropyl alcohol is metabolized by alcohol dehydrogenase to acetone. The characteristic features differ significantly from EG and methanol intoxication in that the parent compound (isopropyl alcohol), not its metabolites, causes toxicity, and a high-AG acidosis is ***not*** present because acetone is rapidly excreted. Both isopropyl alcohol and acetone increase the osmolar gap, and hypoglycemia is common. Alternative diagnoses should be considered if the patient does not improve significantly within a few hours. Patients with hemodynamic instability and/or plasma levels above 400 mg/dL should be considered for acute hemodialysis.

TREATMENT

Isopropyl Alcohol Toxicity

Isopropanol alcohol toxicity is treated by supportive therapy, IV fluids, pressors, ventilatory support if needed, and acute hemodialysis for prolonged coma, hemodynamic instability, or levels >400 mg/dL.

PYROGLUTAMIC ACID Acetaminophen-induced high-AG metabolic acidosis is being recognized more frequently and is observed in patients with acetaminophen overdose and in malnourished or critically ill patients receiving acetaminophen in standard dosage. 5-Oxoproline accumulation after acetaminophen should be suspected in the setting of an unexplained high-AG acidosis in the absence of an elevated osmolar gap in patients receiving acetaminophen. The first step in treatment is to immediately discontinue acetaminophen. Additionally, sodium bicarbonate should be given IV. Although *N*-acetylcysteine has been suggested, it has not been demonstrated unequivocally that it hastens the metabolism of 5-oxoproline by increasing intracellular glutathione concentrations in this setting, as assumed.

Chronic Kidney Disease (See also Chap. 322) The hyperchloremic acidosis of moderate CKD (stage 3B) is eventually converted to the high-AG acidosis of advanced renal failure (stages 4 and 5 CKD). Poor filtration and reabsorption of organic anions contribute to the pathogenesis. As renal disease progresses, the number of functioning nephrons eventually becomes insufficient to keep pace with net acid production. Uremic acidosis in advanced CKD is characterized by a reduced rate of NH_4^+ production and excretion. Alkaline salts from bone buffer the acid retained in CKD. Despite significant retention of acid (up to 20 mmol/d), the serum $[HCO_3^-]$ does not typically decrease further, indicating participation of buffers outside the extracellular compartment. Therefore, a recognized trade-off in untreated chronic metabolic acidosis of CKD stages 3 and 4 is significant loss of bone mass due to reduction in bone calcium carbonate. Chronic acidosis also contributes significantly to muscle wasting and disability in advancing CKD. Evidence has been advanced recently that the high anion gap acidosis of chronic kidney disease contributes significantly per se to the progressive loss of kidney function.

TREATMENT

Metabolic Acidosis of Chronic Kidney Disease

Because chronic metabolic acidosis in advanced CKD is clearly associated with muscle catabolism, bone disease, and more rapid progression of CKD, both the "uremic acidosis" of end-stage renal disease and the non-AG metabolic acidosis of stages 3 and 4 CKD require oral alkali replacement to increase and maintain the $[HCO_3^-]$ to a value >22–24 mmol/L. This can be accomplished with relatively modest amounts of alkali (1.0–1.5 mmol/kg body weight per day) and has been shown to slow the progression of CKD. Either $NaHCO_3$ tablets (650-mg tablets contain 7.8 meq) or oral sodium citrate (Shohl's solution) is effective. The addition of fruits and vegetables (citrate) to the diet increases the plasma $[HCO_3^-]$ and slows progression of CKD safely and is well tolerated. Hyperkalemia is not a common complication of increasing the dietary intake of fruits and vegetables.

NON-ANION GAP METABOLIC ACIDOSES

Alkali can be lost from the gastrointestinal tract as a result of diarrhea or from the kidneys due to renal tubular abnormalities (e.g., renal tubular acidosis [RTA]). In these disorders **(Table 58-5)**, reciprocal changes in $[Cl^-]$ and $[HCO_3^-]$ maintain a normal AG. In non-AG acidosis the increase in $[Cl^-]$ above the normal value approximates the decrease in $[HCO_3^-]$. The absence of such a relationship (disparity in the Δ values) suggests a mixed disturbance.

Stool contains a higher concentration of HCO_3^- and decomposed HCO_3^- than plasma so that metabolic acidosis develops in diarrhea. Instead of an acid urine pH (as anticipated with systemic acidosis), urine pH is usually >6 because metabolic acidosis and hypokalemia increase renal synthesis and excretion of NH_4^+, thus providing a urinary buffer that increases urine pH. Metabolic acidosis due to gastrointestinal losses with a high urine pH can be differentiated from RTA because urinary NH_4^+ excretion is typically low in RTA and high with diarrhea. Urinary NH_4^+ levels are not routinely measured by clinical laboratories, but can be estimated by calculating the urine anion gap (UAG): UAG = $[Na^+ + K^+]_u - [Cl^-]_u$. When $[Cl^-]_u > [Na^+ + K^+]_u$, the UAG is negative by definition. This suggests that the urine ammonium level is appropriately increased, supporting an extrarenal cause of the acidosis. Conversely, when the UAG is positive, the urine ammonium level is predictably low, suggesting a renal tubular origin of the acidosis. Recent studies have shown a poor correlation between the UAG and the measured urine ammonium, thus calling the estimation of urine ammonium by calculation of the UAG into question. Therefore, clinical laboratories should be encouraged to measure urine ammonium by adaptation of automated plasma ammonium assays, using the enzymatic method. This is easily accomplished if the urine sample is diluted 1:200 in normal saline.

Proximal RTA (type 2 RTA) **(Chap. 327)** is often due to generalized proximal tubular dysfunction manifested by glycosuria, generalized aminoaciduria, and phosphaturia (Fanconi syndrome). When the plasma $[HCO_3^-]$ is low, the urine pH is acid (pH <5.5) but exceeds 5.5 with exogenous alkali therapy. The fractional excretion of $[HCO_3^-]$ may exceed 10–15% when the serum HCO_3^- is >20 mmol/L. Because of the defect in HCO_3^- reabsorption by the proximal tubule, therapy with $NaHCO_3$ will enhance delivery of HCO_3^- to the distal nephron and enhance renal potassium secretion, thereby causing hypokalemia. For this reason, potassium supplementation is often added to alkali therapy and may be accomplished by administration of potassium citrate–citric acid solution (Polycitra K or Cytra-K), as discussed below.

Distal RTA (type 1 RTA) may be seen as an acquired or inherited disorder. The features of classical distal RTA (type 1 RTA) include hypokalemia, a non-AG metabolic acidosis, low urinary NH_4^+ excretion (positive UAG, low urine $[NH_4^+]$), and an inappropriately high urine pH (pH >5.5) for the prevailing metabolic acidosis. Most patients have hypocitraturia and hypercalciuria. Therefore, nephrolithiasis, nephrocalcinosis, and bone disease are common. In contrast, in **generalized distal RTA** (type 4 RTA), hyperkalemia is disproportionate to the accompanying reduction in glomerular filtration rate (GFR) because of coexisting impairment of potassium and acid secretion. Urinary ammonium excretion is invariably depressed, and kidney function may be compromised secondary to diabetic nephropathy, obstructive uropathy, or chronic tubulointerstitial disease.

TABLE 58-5 Causes of Non–Anion Gap Acidosis

- I. Gastrointestinal bicarbonate loss
    - A. Diarrhea
    - B. External pancreatic or small-bowel drainage/fistula
    - C. Diversion of ureter: ureterosigmoidostomy, jejunal loop, ileal loop
    - D. Drugs
        - 1. Calcium chloride (acidifying agent)
        - 2. Magnesium sulfate (diarrhea)
        - 3. Cholestyramine (bile acid diarrhea)
- II. Renal acidosis
    - A. Hypokalemia
        - 1. Proximal RTA (type 2 RTA)

          Drug-induced: acetazolamide, topiramate

          Inherited: (a) autosomal recessive missense mutation of *SLCA4* (encodes for basolateral NBCe1) (accompanied by ocular abnormalities); (b) autosomal dominant mutation of *NHE3* (apical Na^+/H^+ exchanger) (rare; associated with short stature)
        - 2. Distal (classic) RTA (type 1 RTA)

          Drug-induced: amphotericin B, ifosfamide

          Inherited: defect of *ATP6V1B1* (encodes for basolateral HCO_3^-/Cl^- exchanger of distal tubule and collecting duct)
    - B. Hyperkalemia
        - 1. Generalized distal nephron dysfunction (type 4 RTA)
            - a. Selective aldosterone deficiency
            - b. Mineralocorticoid resistance (PHA I, autosomal dominant)
            - c. Voltage defect (PHA I, autosomal recessive, and PHA II)
            - d. Hyporeninemic hypoaldosteronism
            - e. Tubulointerstitial disease
    - C. Normokalemia
        - 1. Chronic progressive kidney disease
- III. Drug-induced hyperkalemia (with CKD)
    - A. Potassium-sparing diuretics (amiloride, triamterene, spironolactone, eplerenone)
    - B. Trimethoprim
    - C. Pentamidine
    - D. ACE-Is and ARBs
    - E. Nonsteroidal anti-inflammatory drugs
    - F. Calcineurin inhibitors
    - G. Heparin in critically ill patients
- IV. Other
    - A. Acid loads (ammonium chloride, IV hyperalimentation [uncommon])
    - B. Loss of potential bicarbonate: ketosis with ketone excretion
    - C. Expansion acidosis (rapid saline administration)
    - D. Hippurate
    - E. Cation exchange resins

Abbreviations: ACE-I, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; CKD, chronic kidney disease; PHA, pseudohypoaldosteronism; RTA, renal tubular acidosis.

Hyporeninemic hypoaldosteronism typically presents as a non-AG metabolic acidosis in older adults with diabetes mellitus or tubulointerstitial disease and stage 3 or 4 CKD. These patients typically have hyperkalemia ($[K^+]$ 5.2–6.0 mmol/L), hypertension, and congestive heart failure. Both the metabolic acidosis and the hyperkalemia are out of proportion to the reduction in GFR. Nonsteroidal anti-inflammatory drugs, trimethoprim, pentamidine, angiotensin-converting enzyme (ACE) inhibitors, and angiotensin receptor blockers (ARBs) may also increase the risk for hyperkalemia and a non-AG metabolic acidosis in patients with CKD, especially from diabetic nephropathy (Table 58-5).

TREATMENT

Non–Anion Gap Metabolic Acidoses

For non-AG acidosis due to gastrointestinal losses of bicarbonate, $NaHCO_3$ may be administered intravenously or orally, as indicated by the severity of both the acidosis and the accompanying volume depletion. Proximal RTA is the most challenging of the RTAs to treat if the goal is to restore the serum $[HCO_3^-]$ to normal (as recommended to support normal growth in children with isolated bicarbonate wasting), because administration of oral alkali increases urinary excretion of bicarbonate and potassium. In patients with proximal RTA (type 2), potassium administration is typically necessary. An oral solution of sodium and potassium citrate (citric acid 334 mg, sodium citrate 500 mg, and potassium citrate 550 mg per 5 mL) may be prescribed for this purpose (Virtrate or Cytra-3). In classical distal RTA (type 1), hypokalemia should be corrected initially. When accomplished, alkali therapy with either sodium citrate (Shohl's solution) or $NaHCO_3$ tablets (650-mg tablets contain 7.8 meq) should be initiated to correct and maintain the serum $[HCO_3^-]$ in the range of 24–26 meq/L. Type 1 RTA patients typically respond to chronic alkali therapy readily. The long-term benefits of adequate alkali therapy in distal (type 1) RTA include a decrease in the frequency of nephrolithiasis, improvement in bone density, resumption of normal growth patterns in children, and preservation of kidney function in both adults and children. For type 4 RTA, it is necessary to correct the metabolic acidosis, using the same approach as for classical distal RTA (type 1 RTA), and also to correct the plasma $[K^+]$. Hyperkalemia directly reduces ammoniagenesis and net acid excretion. Therefore, restoration of normokalemia increases urinary net acid excretion and consequently can greatly improve the metabolic acidosis. Historically, hyperkalemia was treated by chronic administration of oral sodium polystyrene sulfonate (15 g of power prepared as an oral solution, without sorbitol, once daily 2–3 times per week). However, this preparation is often unpalatable and patient compliance is low. The nonabsorbed, calcium-potassium cation exchange polymer, patiromer, may be considered for type 4 RTA patients with hyperkalemia because it is more palatable and more effective but has very few side effects. Patiromer is administered as 8.4-g packets of powder for suspension PO twice daily with dose adjustment at weekly intervals, based on the plasma $[K^+]$, not to exceed 25.2 g/d. Additionally, the diet should be low in potassium-containing foods or supplements (e.g., salt substitute). Potassium-retaining medications should be discontinued. Finally, patients with documented isolated hypoaldosteronism should receive fludrocortisone, but the dose varies with the cause of the hormone deficiency. This agent should be administered very cautiously and in combination with furosemide in patients with edema and hypertension because of potential aggravation of these conditions.

METABOLIC ALKALOSIS

Metabolic alkalosis is established by an elevated arterial pH, an increase in the serum $[HCO_3^-]$, and an increase in $Paco_2$ as a result of compensatory alveolar hypoventilation (Table 58-1). It is often accompanied by hypochloremia and hypokalemia. The elevation in arterial pH establishes the diagnosis because pH is decreased in respiratory acidosis, even though both have an elevated $Paco_2$. Metabolic alkalosis may present as a mixed acid-base disorder in association with either respiratory acidosis, respiratory alkalosis, or metabolic acidosis.

ETIOLOGY AND PATHOGENESIS

Metabolic alkalosis occurs as a result of net gain of $[HCO_3^-]$ or loss of nonvolatile acid (usually HCl by vomiting) from the extracellular fluid. When vomiting causes loss of HCl from the stomach, HCO_3^- secretion cannot be initiated in the small bowel, so that HCO_3^- is retained in the extracellular fluid. Thus, vomiting or nasogastric suction is an example of the *generation stage* of metabolic alkalosis, in which the loss of acid typically causes alkalosis. Upon cessation of vomiting, the *maintenance stage* ensues because secondary factors prevent the kidneys from excreting HCO_3^- appropriately.

Maintenance of metabolic alkalosis, therefore, represents a failure of the kidneys to eliminate excess HCO_3^- from the extracellular compartment. The kidneys will retain, rather than excrete, the excess alkali and maintain the alkalosis if (1) volume deficiency, chloride deficiency, and K^+ deficiency exist in combination with a reduced GFR (associated with a low urine $[Cl^-]$) or (2) hypokalemia exists because of autonomous hyperaldosteronism (normal urine $[Cl^-]$). In the first example, saline-responsive metabolic alkalosis is corrected by extracellular fluid volume (ECFV) restoration (IV administration of NaCl and KCl), whereas, in the latter, it may be necessary to repair the alkalosis by pharmacologic or surgical intervention, not with saline administration (saline-unresponsive metabolic alkalosis).

DIFFERENTIAL DIAGNOSIS

To establish the cause of metabolic alkalosis **(Table 58-6)**, it is necessary to assess the status of the patient's ECFV. It is important to measure the recumbent and upright blood pressure and pulse (to determine if orthostasis is present) and to obtain a serum $[K^+]$ and a urine $[Cl^-]$. When hyperreninemia or isolated hyperaldosteronism is suspected, renin and aldosterone should be measured. For example, the presence of chronic hypertension and chronic hypokalemia in an alkalotic patient suggests either mineralocorticoid excess or that the hypertensive patient is receiving diuretics. Low plasma renin activity and values for urine $[Cl^-]$ >20 meq/L in a patient not receiving diuretics suggest primary mineralocorticoid excess. The combination of hypokalemia and alkalosis in a normotensive, nonedematous patient may be due to Bartter's or Gitelman's syndrome, magnesium deficiency, vomiting, exogenous alkali, or diuretic ingestion. Measurement of urine electrolytes (especially the urine $[Cl^-]$) is recommended, and occasionally, screening of the urine for diuretics may be necessary if surreptitious diuretic abuse is suspected. If the urine is alkaline, with an elevated $[Na^+]_u$ and $[K^+]_u$ but low $[Cl^-]_u$, the diagnosis of either vomiting (overt or surreptitious) or alkali ingestion should be considered. If the urine is relatively acid with low concentrations of Na^+, K^+, and Cl^-, the most likely possibilities are prior vomiting, the posthypercapnic state, or prior diuretic ingestion. If the urine sodium, potassium, and chloride concentrations are not depressed, magnesium deficiency, Bartter's or Gitelman's syndrome, or current diuretic ingestion should be considered. Bartter's syndrome is distinguished from Gitelman's syndrome by the presence of hypocalciuria in the latter disorder.

Alkali Administration Chronic administration of alkali to individuals with normal renal function rarely causes alkalosis. However, in patients with coexistent hemodynamic disturbances associated with effective ECFV depletion (e.g., congestive heart failure), alkalosis can develop because of diminished capacity to excrete HCO_3^- or enhanced reabsorption of HCO_3^-. Such patients include those who receive $NaHCO_3$ (PO or IV), citrate loads IV (transfusions of whole blood or therapeutic apheresis), or antacids plus cation-exchange resins (aluminum hydroxide and sodium polystyrene sulfonate [uncommon]). Nursing home patients receiving enteral tube feedings have a higher incidence of metabolic alkalosis than nursing home patients receiving regular diets.

METABOLIC ALKALOSIS ASSOCIATED WITH ECFV CONTRACTION, K^+ DEPLETION, AND SECONDARY HYPERRENINEMIC HYPERALDOSTERONISM

Gastrointestinal Origin Gastrointestinal loss of H^+ from vomiting or gastric aspiration causes simultaneous addition of HCO_3^- into the extracellular fluid. During active vomiting, the filtered load of bicarbonate reaching the kidneys is acutely increased and will exceed the reabsorptive capacity of the proximal tubule for HCO_3^- absorption.

TABLE 58-6 Causes of Metabolic Alkalosis

- I. Exogenous HCO_3^- loads
  - A. Acute alkali administration
  - B. Milk-alkali syndrome
- II. Effective ECFV depletion, normal or low BP (with orthostasis), K^+ deficiency, and secondary hyperreninemic hyperaldosteronism
  - A. Gastrointestinal origin
    - 1. Vomiting
    - 2. Gastric aspiration
    - 3. Congenital chloridorrhea
    - 4. Gastrocystoplasty
    - 5. Villous adenoma
  - B. Renal origin
    - 1. Diuretic use (thiazides and loop diuretics)
    - 2. Posthypercapnic state
    - 3. Hypercalcemia/hypoparathyroidism
    - 4. Recovery from lactic acidosis or ketoacidosis
    - 5. Nonreabsorbable anion administration (e.g., IV penicillin, carbenicillin)
    - 6. Mg^{2+} deficiency
    - 7. K^+ depletion
    - 8. Bartter's syndrome (loss-of-function mutations of transporters and ion channels in TALH)
    - 9. Gitelman's syndrome (loss-of-function mutation of Na^+-Cl^- cotransporter in DCT and collecting duct)
- III. ECFV expansion, hypertension, K^+ deficiency, and mineralocorticoid excess
  - A. High renin
    - 1. Renal artery stenosis
    - 2. Accelerated hypertension
    - 3. Renin-secreting tumor
    - 4. Estrogen therapy
  - B. Low renin
    - 1. Primary aldosteronism
      - a. Adenoma
      - b. Hyperplasia
      - c. Carcinoma
    - 2. Adrenal enzyme defects
      - a. 11β-Hydroxylase deficiency
      - b. 17α-Hydroxylase deficiency
    - 3. Cushing's syndrome or disease
    - 4. Other
      - a. Licorice
      - b. Carbenoxolone
      - c. Chewer's tobacco
- IV. Gain-of-function mutation of sodium channel in DCT (ENaC) with ECFV expansion, hypertension, K^+ deficiency, and hyporeninemic-hypoaldosteronism
  - A. Liddle's syndrome

Abbreviations: DCT, distal convoluted tubule; ECFV, extracellular fluid volume; TALH, thick ascending limb of Henle's loop.

Subsequently, enhanced delivery of HCO_3^- to the distal nephron, where the capacity for HCO_3^- reabsorption is lower, will result in excretion of alkaline urine that stimulates potassium secretion. When vomiting ceases, the persistence of volume, potassium, and chloride depletion triggers maintenance of the alkalosis because these conditions promote HCO_3^- reabsorption. Correction of the contracted ECFV with NaCl and repair of K^+ deficits with KCl correct the acid-base disorder by restoring the ability of the kidney to excrete the excess bicarbonate.

Renal Origin • DIURETICS (See also Chap. 265) Diuretics such as thiazides and loop diuretics (furosemide, bumetanide, torsemide) increase excretion of salt and acutely diminish the ECFV without altering the total body bicarbonate content. The serum $[HCO_3^-]$ increases because the reduced ECFV "contracts" around the $[HCO_3^-]$ in plasma (contraction alkalosis). The chronic administration of diuretics tends to generate an alkalosis by increasing distal salt delivery so that both K^+ and H^+ secretion are stimulated. The alkalosis is maintained by persistence of the contraction of the ECFV, secondary hyperaldosteronism, K^+ deficiency, and the direct effect of the diuretic (as long as diuretic administration continues). Discontinuing the diuretic and providing isotonic saline to correct the ECFV deficit will repair the alkalosis.

SOLUTE LOSING DISORDERS: BARTTER'S AND GITELMAN'S SYNDROMES
See Chap. 327.

NON-REABSORBABLE ANIONS AND MAGNESIUM DEFICIENCY Administration of large quantities of the penicillin derivatives carbenicillin or ticarcillin causes their non-reabsorbable anions to appear in the distal tubule, causing an increase in the transepithelial potential difference in the collecting tubule. The more negative potential difference increases both H^+ and K^+ secretion across the apical membrane. Mg^{2+} deficiency may occur with chronic administration of thiazide diuretics, alcoholism, and malnutrition. In Gitelman's syndrome, the development of hypokalemic alkalosis occurs through stimulation of renin and aldosterone secretion to enhance distal acidification.

POTASSIUM DEPLETION Chronic K^+ depletion, as a result of extreme dietary potassium restriction, diuretics, or alcohol abuse, may initiate metabolic alkalosis by increasing urinary net acid excretion. Potassium depletion often occurs concurrent with magnesium deficiency in alcoholics with malnutrition. The renal generation of NH_4^+ (ammoniagenesis) is upregulated directly by hypokalemia. Chronic K^+ deficiency also upregulates the H^+, K^+-ATPases in the distal tubule and collecting duct to increase K^+ absorption while simultaneously increasing H^+ secretion. Alkalosis associated with severe K^+ depletion is resistant to salt administration, but repair of the K^+ deficiency corrects the alkalosis.

AFTER TREATMENT OF LACTIC ACIDOSIS OR KETOACIDOSIS When an underlying stimulus for the generation of lactic acid or ketoacid is corrected, such as correction of shock or severe volume depletion by volume restoration, or with insulin therapy for DKA, the lactate or ketones are metabolized to yield an equivalent amount of HCO_3^-. If there have been exogenous sources of HCO_3^-, this additional HCO_3^- will be additive to that amount generated by organic anion metabolism that together may create a surfeit of HCO_3^- ("rebound alkalosis").

POSTHYPERCAPNIA Prolonged CO_2 retention with chronic respiratory acidosis enhances renal HCO_3^- absorption and the generation of new HCO_3^- (increased net acid excretion). Metabolic alkalosis occurs when the elevated $Paco_2$ is abruptly returned toward normal because of residual stimulation of HCO_3^- absorption by the prior chronic increase in $Paco_2$.

■ METABOLIC ALKALOSIS ASSOCIATED WITH ECFV EXPANSION, HYPERTENSION, AND MINERALOCORTICOID EXCESS

An increase in aldosterone may be the result of autonomous primary adrenal overproduction or of secondary aldosterone secretion in response to overproduction of renin by the kidney. Mineralocorticoid excess increases net acid excretion and may result in metabolic alkalosis, which is typically exacerbated by associated K^+ deficiency. The characteristic salt retention and hypertension are due to upregulation of the epithelial Na^+ channel (ENaC) in the collecting tubule in response to aldosterone. The kaliuresis persists because of mineralocorticoid excess and stimulation of ENaC, causing an increase in transepithelial voltage that enhances K^+ secretion by the collecting duct. Persistent K^+ depletion may cause polydipsia and polyuria.

Liddle's syndrome (Chap. 327) results from an inherited gain-of-function mutation of genes that regulate the collecting duct Na^+ channel, ENaC. This rare monogenic form of hypertension is the result of volume expansion that secondarily suppresses aldosterone elaboration. Patients typically present with hypertension, hypokalemia, and metabolic alkalosis.

Symptoms With metabolic alkalosis, changes in CNS and peripheral nervous system function are similar to those of hypocalcemia

(Chap. 421); symptoms include mental confusion; obtundation; and a predisposition to seizures, paresthesias, muscular cramping, tetany, aggravation of arrhythmias, and hypoxemia in chronic obstructive pulmonary disease. Related electrolyte abnormalities include hypokalemia and hypophosphatemia.

TREATMENT

Metabolic Alkalosis

The first goal of therapy is to correct the underlying stimulus for HCO_3^- generation. If primary aldosteronism or Cushing's syndrome is present, correction of the underlying cause will reverse the hypokalemia and alkalosis. $[\text{H}^+]$ loss by the stomach or kidneys can be mitigated by the use of proton pump inhibitors or the discontinuation of diuretics, respectively. The second aspect of treatment is to eliminate factors that sustain the inappropriate increase in HCO_3^- reabsorption, such as ECFV contraction or K^+ deficiency. K^+ deficits should always be repaired. Isotonic saline is recommended to reverse the alkalosis when ECFV contraction is present. If associated conditions, such as congestive heart failure, preclude infusion of isotonic saline, renal HCO_3^- loss can be accelerated by administration of acetazolamide (125–250 mg IV), a carbonic anhydrase inhibitor, which is usually effective in patients with adequate kidney function. Close monitoring is required since acetazolamide triggers urinary K^+ losses and may cause hypokalemia that should be corrected promptly. Dilute hydrochloric acid IV (0.1 *N* HCl) has been advocated in extreme cases of metabolic alkalosis but causes hemolysis and must be delivered slowly in a central vein. This preparation is not available generally and requires preparation by the pharmacy. Because serious errors or harm may occur with dilute HCl infusion, its use is not advised. Therapy in Liddle's syndrome should include a potassium-sparing diuretic (amiloride or triamterene) to inhibit ENaC and correct both the hypertension and the hypokalemia.

RESPIRATORY ACIDOSIS

Respiratory acidosis occurs as a result of severe pulmonary disease, respiratory muscle fatigue, or abnormalities in ventilatory control. It is characterized by an elevated Paco_2 and reduced pH **(Table 58-7)**. In acute respiratory acidosis, there is a compensatory elevation in HCO_3^- (due to cellular buffering mechanisms) that increases the serum $[\text{HCO}_3^-]$ 1 mmol/L for every 10-mmHg increase in Paco_2. In chronic respiratory acidosis (>24 h), renal adaptation increases the $[\text{HCO}_3^-]$ by 4 mmol/L for every 10-mmHg increase in Paco_2. The serum HCO_3^- usually does not increase above 38 mmol/L in respiratory acidosis.

The clinical features of respiratory acidosis vary according to the severity and duration of the disorder, the underlying disease, and whether there is accompanying hypoxemia. A rapid increase in Paco_2 (acute hypercapnia) may cause anxiety, dyspnea, confusion, psychosis, and hallucinations and may progress to coma. Chronic hypercapnia may cause sleep disorders; loss of memory; daytime somnolence; personality changes; impairment of coordination; and motor disturbances such as tremor, myoclonic jerks, and asterixis. Headaches and other signs that mimic raised intracranial pressure, such as papilledema, abnormal reflexes, and focal muscle weakness, may also occur.

Depression of the respiratory center by a variety of drugs, injury, or disease can produce respiratory acidosis. This may occur acutely with general anesthetics, sedatives, and head trauma or chronically with sedatives, alcohol, intracranial tumors, and the syndromes of sleep-disordered breathing including the primary alveolar and obesity-hypoventilation syndromes **(Chaps. 296 and 308)**. Abnormalities or disease in the motor neurons, neuromuscular junction, and skeletal muscle can cause hypoventilation via respiratory muscle fatigue. Mechanical ventilation, when not properly adjusted, may result in respiratory acidosis, particularly if CO_2 production suddenly rises (because of fever, agitation, sepsis, or overfeeding) or alveolar ventilation decreases because of worsening pulmonary function. High levels of positive end-expiratory pressure in the presence of reduced cardiac output may cause hypercapnia as a result of large increases in alveolar dead space **(Chap. 296)**. Permissive hypercapnia may be used to minimize intrinsic positive end-expiratory pressure in respiratory distress syndrome, but the consequential respiratory acidosis may require administration of NaHCO_3 to increase the arterial pH to approximately 7.20. The pH should not be increased to the normal value by NaHCO_3 infusion, however.

TABLE 58-7 Respiratory Acid-Base Disorders

- I. Alkalosis
  - A. Central nervous system stimulation
    - 1. Pain
    - 2. Anxiety
    - 3. Fever
    - 4. Cerebrovascular accident
    - 5. Meningitis, encephalitis
    - 6. Tumor
    - 7. Trauma
  - B. Hypoxemia or tissue hypoxia
    - 1. High-altitude acclimatization
    - 2. Pneumonia, pulmonary edema
    - 3. Aspiration
    - 4. Severe anemia
  - C. Drugs or hormones
    - 1. Pregnancy, progesterone
    - 2. Salicylates
    - 3. Cardiac failure
  - D. Stimulation of chest receptors
    - 1. Hemothorax
    - 2. Flail chest
    - 3. Cardiac failure
    - 4. Pulmonary embolism
  - E. Miscellaneous
    - 1. Septicemia
    - 2. Hepatic failure
    - 3. Mechanical hyperventilation
    - 4. Heat exposure
    - 5. Recovery from metabolic acidosis
- II. Acidosis
  - A. Central
    - 1. Drugs (anesthetics, morphine, sedatives)
    - 2. Stroke
    - 3. Infection
  - B. Airway
    - 1. Obstruction
    - 2. Asthma
  - C. Parenchyma
    - 1. Emphysema
    - 2. Pneumoconiosis
    - 3. Bronchitis
    - 4. Adult respiratory distress syndrome
    - 5. Barotrauma
  - D. Neuromuscular
    - 1. Poliomyelitis
    - 2. Kyphoscoliosis
    - 3. Myasthenia
    - 4. Muscular dystrophies
  - E. Miscellaneous
    - 1. Obesity
    - 2. Hypoventilation
    - 3. Permissive hypercapnia

Acute hypercapnia follows sudden occlusion of the upper airway or generalized bronchospasm as in severe asthma, anaphylaxis, inhalational burn, or toxin injury. Chronic hypercapnia and respiratory acidosis occur in end-stage obstructive lung disease. Restrictive disorders involving both the chest wall and the lungs can cause respiratory acidosis because the high metabolic cost of respiration initiates ventilatory muscle fatigue. Advanced stages of intrapulmonary and extrapulmonary restrictive defects present as chronic respiratory acidosis.

The diagnosis of respiratory acidosis requires the measurement of $\mathrm{Pa{CO}_2}$ and arterial pH. A detailed history and physical examination will typically identify the cause. Pulmonary function studies **(Chap. 297)**, including spirometry, diffusion capacity for carbon monoxide, lung volumes, and arterial $\mathrm{Pa{CO}_2}$ and O_2 saturation, usually make it possible to determine if respiratory acidosis is secondary to lung disease. The workup for nonpulmonary causes should include a detailed drug history, measurement of hematocrit, and assessment of upper airway, chest wall, pleura, and neuromuscular function.

TREATMENT

Respiratory Acidosis

The management of respiratory acidosis depends on the severity and rate of onset. Acute respiratory acidosis can be life-threatening, and measures to reverse the underlying cause should be undertaken simultaneously with restoration of adequate alveolar ventilation. This may necessitate tracheal intubation and assisted mechanical ventilation. Oxygen administration should be titrated carefully in patients with severe obstructive pulmonary disease and chronic CO_2 retention who are breathing spontaneously **(Chap. 307)**. When oxygen is used injudiciously, these patients may experience progression of the respiratory acidosis, causing severe acidemia. Aggressive and rapid correction of hypercapnia should be avoided, because the falling $\mathrm{Pa{CO}_2}$ may provoke the same complications noted with acute respiratory alkalosis (i.e., cardiac arrhythmias, reduced cerebral perfusion, and seizures). The $\mathrm{Pa{CO}_2}$ should be lowered gradually in chronic respiratory acidosis, aiming to restore the $\mathrm{Pa{CO}_2}$ to baseline levels and to provide sufficient Cl^- and K^+ to enhance the renal excretion of HCO_3^-.

Chronic respiratory acidosis is frequently difficult to correct, but the primary goal is to institute measures that may improve lung function **(Chap. 303)**.

RESPIRATORY ALKALOSIS

Alveolar hyperventilation decreases $\mathrm{Pa{CO}_2}$ and increases the $HCO_3^-/\mathrm{Pa{CO}_2}$ ratio, thus increasing pH (Table 58-7). Nonbicarbonate cellular buffers respond by consuming HCO_3^-. Hypocapnia develops when a sufficiently strong ventilatory stimulus causes CO_2 output in the lungs to exceed its metabolic production by tissues. Plasma pH and $[HCO_3^-]$ appear to vary proportionately with $\mathrm{Pa{CO}_2}$ over a range from 40–15 mmHg. The relationship between arterial $[H^+]$ concentration and $\mathrm{Pa{CO}_2}$ is ~0.7 mmol/L per mmHg (or 0.01 pH unit/mmHg), and that for plasma $[HCO_3^-]$ is 0.2 mmol/L per mmHg. Hypocapnia sustained for >2–6 h is further compensated by a decrease in renal ammonium and titratable acid excretion and a reduction in filtered HCO_3^- reabsorption. Full renal adaptation to respiratory alkalosis may take several days and requires normal volume status and renal function. The kidneys appear to respond directly to the lowered $\mathrm{Pa{CO}_2}$ rather than to alkalosis per se. In chronic respiratory alkalosis, a 1-mmHg decrease in $\mathrm{Pa{CO}_2}$ causes a 0.4- to 0.5-mmol/L decrease in $[HCO_3^-]$ and a 0.3-mmol/L decrease in $[H^+]$ (or 0.003 unit increase in pH).

The effects of respiratory alkalosis vary according to duration and severity but are primarily those of the underlying disease. Reduced cerebral blood flow as a consequence of a rapid decline in $\mathrm{Pa{CO}_2}$ may cause dizziness, mental confusion, and seizures, even in the absence of hypoxemia. The cardiovascular effects of acute hypocapnia in the conscious human are generally minimal, but in the anesthetized or mechanically ventilated patient, cardiac output and blood pressure may fall because of the depressant effects of anesthesia and positive-pressure ventilation on heart rate, systemic resistance, and venous return. Cardiac arrhythmias may occur in patients with heart disease as a result of changes in oxygen unloading by blood from a left shift in the hemoglobin-oxygen dissociation curve (Bohr effect). Acute respiratory alkalosis causes intracellular shifts of Na^+, K^+, and PO_4^{2-} and reduces free $[Ca^{2+}]$ by increasing the protein-bound fraction. Hypocapnia-induced hypokalemia is usually minor.

Chronic respiratory alkalosis is the most common acid-base disturbance in critically ill patients and, when severe, portends a poor prognosis. Many cardiopulmonary disorders manifest respiratory alkalosis in their early to intermediate stages. Normocapnia and hypoxemia in a patient with hyperventilation may herald the onset of rapid respiratory failure. Therefore, prompt assessment is necessary to determine if the patient is becoming fatigued. Respiratory alkalosis is also common during mechanical ventilation.

The hyperventilation syndrome may be disabling. Paresthesia; circumoral numbness; chest wall tightness or pain; dizziness; inability to take an adequate breath; and, rarely, tetany may be sufficiently stressful to perpetuate the disorder. Arterial blood-gas analysis demonstrates an acute or chronic respiratory alkalosis, often with hypocapnia in the range of 15–30 mmHg, but without hypoxemia. CNS diseases or injury can produce several patterns of hyperventilation and sustained $\mathrm{Pa{CO}_2}$ levels of 20–30 mmHg. Hyperthyroidism, high caloric loads, and exercise raise the basal metabolic rate, but ventilation usually increases in proportion so that arterial blood gases are unchanged and respiratory alkalosis does not develop. Salicylates are the most common cause of drug-induced respiratory alkalosis because of direct stimulation of the medullary chemoreceptor by salicylates **(Chap. 469)**. In addition, the methylxanthines, theophylline and aminophylline stimulate ventilation and increase the ventilatory response to CO_2. Progesterone increases ventilation and lowers arterial $\mathrm{Pa{CO}_2}$ by as much as 5–10 mmHg. Therefore, chronic respiratory alkalosis is a common feature of pregnancy. Respiratory alkalosis is also prominent in hepatic failure, and the severity correlates with the degree of hepatic insufficiency. Respiratory alkalosis may be an early finding in gram-negative septicemia, often occurring before fever, hypoxemia, or hypotension develops.

The diagnosis of respiratory alkalosis depends on measurement of arterial pH and $\mathrm{Pa{CO}_2}$. The plasma $[K^+]$ is often reduced and the $[Cl^-]$ increased. In the acute phase, respiratory alkalosis is not associated with increased renal HCO_3^- excretion, but within hours, net acid excretion is reduced. In general, the HCO_3^- concentration falls by 2.0 mmol/L for each 10-mmHg decrease in $\mathrm{Pa{CO}_2}$. Chronic respiratory alkalosis occurs when hypocapnia persists for greater than 3–5 days. The decline in $\mathrm{Pa{CO}_2}$ reduces the serum $[HCO_3^-]$ by 4.0–5 mmol/L for each 10-mmHg decrease in $\mathrm{Pa{CO}_2}$. It is unusual to observe a plasma HCO_3^- <12 mmol/L as a result of a pure respiratory alkalosis. The compensatory reduction in plasma $[HCO_3^-]$ is so effective in chronic respiratory alkalosis that the pH may not decline significantly from the normal value. Therefore, chronic respiratory alkalosis is the only acid-base disorder for which compensation can return the pH to the normal value.

When the diagnosis of respiratory alkalosis is made, its cause should be investigated. The diagnosis of hyperventilation syndrome is made by exclusion. In difficult cases, it may be important to rule out other conditions such as pulmonary embolism, coronary artery disease, and hyperthyroidism.

TREATMENT

Respiratory Alkalosis

The management of respiratory alkalosis is directed toward alleviation of the underlying disorder. If respiratory alkalosis complicates ventilator management, changes in dead space and tidal volume can minimize the hypocapnia. Patients with the hyperventilation syndrome may benefit from reassurance, rebreathing from a paper bag during symptomatic attacks, and attention to underlying psychological stress. Antidepressants and sedatives are not recommended. β-Adrenergic blockers may ameliorate peripheral manifestations of the hyperadrenergic state.

REFERENCES

Banerjee T et al: High dietary acid load predicts ESRD among adults with CKD. J Am Soc Nephrol 26:1693, 2015.

Berend K et al: Physiological approach to assessment of acid-base disturbances. N Engl J Med 371:1434, 2014.

Hamm LL, DuBose TD: Disorders of acid-base balance. In *Brenner and Rector's The Kidney*, 11th ed. Yu A et al (eds). Philadelphia, Elsevier, 2020, pp 496–536.

Kraut JA, Madias NE: Metabolic acidosis of CKD: An update. Am J Kidney Dis 67:307, 2016.

Kraut JA, Madias NE: Re-evaluation of the normal range of serum total CO$_2$ concentration. Clin J Am Soc Nephrol 13:343, 2018.

Palmer BF, Clegg DJ: Electrolyte and acid–base disturbances in patients with diabetes mellitus. N Engl J Med 373:548, 2015.

Wesson DE et al: Mechanisms of metabolic acidosis-induced kidney injury in chronic kidney disease. J Am Soc Nephrol 31:469, 2020.

Section 8 Alterations in the Skin

59

Approach to the Patient with a Skin Disorder

Kim B. Yancey, Thomas J. Lawley

The challenge of examining the skin lies in distinguishing normal from abnormal findings, distinguishing significant findings from trivial ones, and integrating pertinent signs and symptoms into an appropriate differential diagnosis. The fact that the largest organ in the body is visible is both an advantage and a disadvantage to those who examine it. It is advantageous because no special instrumentation is necessary and because the skin can be biopsied with little morbidity. However, the casual observer can be misled by a variety of stimuli and overlook important, subtle signs of skin or systemic disease. For instance, the sometimes minor differences in color and shape that distinguish a melanoma **(Fig. 59-1)** from a benign nevomelanocytic nevus **(Fig. 59-2)** can be difficult to recognize. A variety of descriptive terms have been developed that characterize cutaneous lesions **(Tables 59-1, 59-2, and 59-3; Fig. 59-3)**, thereby aiding in their interpretation and in the formulation of a differential diagnosis **(Table 59-4)**. For example, the

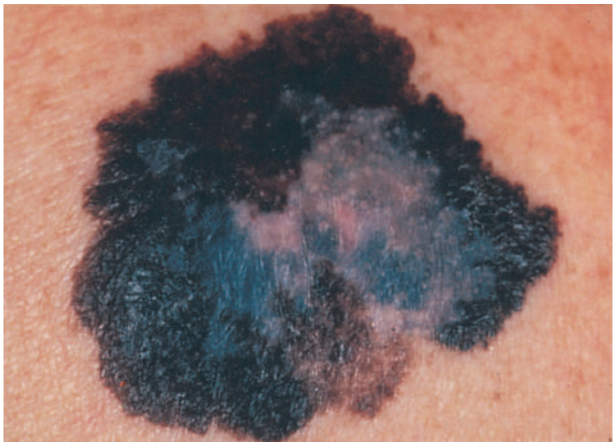

FIGURE 59-1 Superficial spreading melanoma. This is the most common type of melanoma. Such lesions usually demonstrate asymmetry, border irregularity, color variegation (black, blue, brown, pink, and white), a diameter >6 mm, and a history of change (e.g., an increase in size or development of associated symptoms such as pruritus or pain).

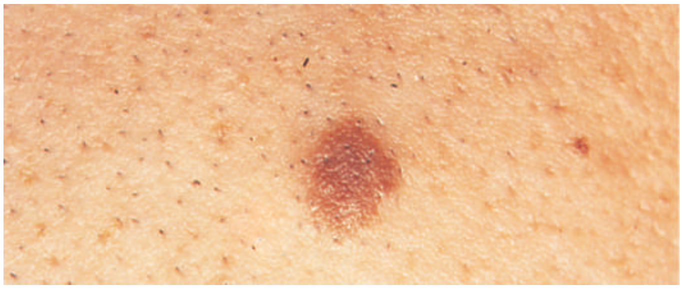

FIGURE 59-2 Nevomelanocytic nevus. Nevi are benign proliferations of nevomelanocytes characterized by regularly shaped hyperpigmented macules or papules of a uniform color.

TABLE 59-1 Description of Primary Skin Lesions
Macule: A flat, colored lesion, <2 cm in diameter, not raised above the surface of the surrounding skin. A "freckle," or ephelid, is a prototypical pigmented macule.
Patch: A large (>2 cm) flat lesion with a color different from the surrounding skin. This differs from a macule only in size.
Papule: A small, solid lesion, <0.5 cm in diameter, raised above the surface of the surrounding skin and thus palpable (e.g., a closed comedone, or whitehead, in acne).
Nodule: A larger (0.5–5.0 cm), firm lesion raised above the surface of the surrounding skin. This differs from a papule only in size (e.g., a large dermal nevomelanocytic nevus).
Tumor: A solid, raised growth >5 cm in diameter.
Plaque: A large (>1 cm), flat-topped, raised lesion; edges may either be distinct (e.g., in psoriasis) or gradually blend with surrounding skin (e.g., in eczematous dermatitis).
Vesicle: A small, fluid-filled lesion, <0.5 cm in diameter, raised above the plane of surrounding skin. Fluid is often visible, and the lesions are translucent (e.g., vesicles in allergic contact dermatitis caused by *Toxicodendron* [poison ivy]).
Pustule: A vesicle filled with leukocytes. Note: The presence of pustules does not necessarily signify the existence of an infection.
Bulla: A fluid-filled, raised, often translucent lesion >0.5 cm in diameter.
Wheal: A raised, erythematous, edematous papule or plaque, usually representing short-lived vasodilation and vasopermeability.
Telangiectasia: A dilated, superficial blood vessel.

TABLE 59-2 Description of Secondary Skin Lesions
Lichenification: A distinctive thickening of the skin that is characterized by accentuated skinfold markings.
Scale: Excessive accumulation of stratum corneum.
Crust: Dried exudate of body fluids that may be either yellow (i.e., serous crust) or red (i.e., hemorrhagic crust).
Erosion: Loss of epidermis without an associated loss of dermis.
Ulcer: Loss of epidermis and at least a portion of the underlying dermis.
Excoriation: Linear, angular erosions that may be covered by crust and are caused by scratching.
Atrophy: An acquired loss of substance. In the skin, this may appear as a depression with intact epidermis (i.e., loss of dermal or subcutaneous tissue) or as sites of shiny, delicate, wrinkled lesions (i.e., epidermal atrophy).
Scar: A change in the skin secondary to trauma or inflammation. Sites may be erythematous, hypopigmented, or hyperpigmented depending on their age or character. Sites on hair-bearing areas may be characterized by destruction of hair follicles.

TABLE 59-3 Common Dermatologic Terms

Alopecia: Hair loss, partial or complete.

Annular: Ring-shaped.

Cyst: A soft, raised, encapsulated lesion filled with semisolid or liquid contents.

Herpetiform: In a grouped configuration.

Lichenoid eruption: Violaceous to purple, polygonal lesions that resemble those seen in lichen planus.

Milia: Small, firm, white papules filled with keratin.

Morbilliform rash: Generalized, small erythematous macules and/or papules that resemble lesions seen in measles.

Nummular: Coin-shaped.

Poikiloderma: Skin that displays variegated pigmentation, atrophy, and telangiectasias.

Polycyclic lesions: A configuration of skin lesions formed from coalescing rings or incomplete rings.

Pruritus: A sensation that elicits the desire to scratch. Pruritus is often the predominant symptom of inflammatory skin diseases (e.g., atopic dermatitis, allergic contact dermatitis); it is also commonly associated with xerosis and aged skin. Systemic conditions that can be associated with pruritus include chronic renal disease, cholestasis, pregnancy, malignancy, thyroid disease, polycythemia vera, and delusions of parasitosis.

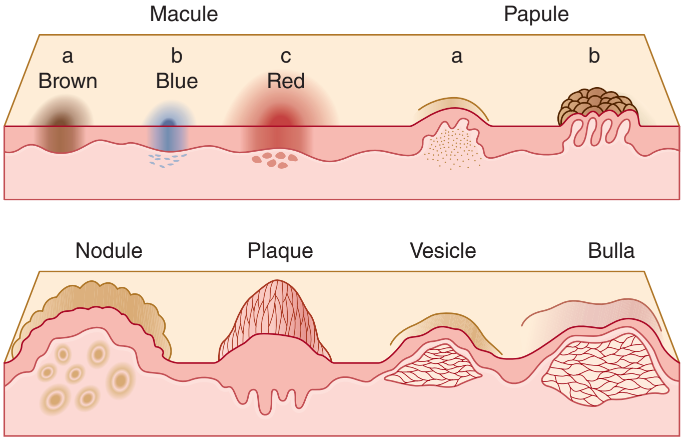

FIGURE 59-3 A schematic representation of several common primary skin lesions (see Table 59-1).

TABLE 59-4 Selected Common Dermatologic Conditions

DIAGNOSIS	COMMON DISTRIBUTION	USUAL MORPHOLOGY	DIAGNOSIS	COMMON DISTRIBUTION	USUAL MORPHOLOGY
Acne vulgaris	Face, upper back, chest	Open and closed comedones, erythematous papules, pustules, cysts	Seborrheic keratosis	Trunk, face, extremities	Brown plaques with adherent, greasy scale; “stuck on” appearance
Rosacea	Blush area of cheeks, nose, forehead, chin	Erythema, telangiectasias, papules, pustules	Folliculitis Impetigo	Any hair-bearing area Anywhere	Follicular pustules Papules, vesicles, pustules, often with honey-colored crusts
Seborrheic dermatitis	Scalp, eyebrows, perinasal areas	Erythema with greasy yellow-brown scale	Herpes simplex	Lips, genitalia	Grouped vesicles progressing to crusted erosions
Atopic dermatitis	Antecubital and popliteal fossae; may be widespread	Patches and plaques of erythema, scaling, and lichenification; pruritus	Herpes zoster	Dermatomal, usually trunk but may be anywhere	Vesicles limited to a dermatome (often painful)
Stasis dermatitis	Ankles, lower legs over medial malleoli	Patches of erythema and scaling on background of hyperpigmentation associated with signs of venous insufficiency	Varicella	Face, trunk, relative sparing of extremities	Lesions arise in crops and quickly progress from erythematous macules, to papules, to vesicles, to pustules, to crusted sites
Dyshidrotic eczema	Palms, soles, sides of fingers, and toes	Deep vesicles	Pityriasis rosea	Trunk (Christmas tree pattern); herald patch followed by multiple smaller lesions	Symmetric erythematous papules and plaques with a collarette of scale
Allergic contact dermatitis	Anywhere	Localized erythema, vesicles, scale, and pruritus (e.g., fingers, earlobes—nickel; dorsal aspect of foot—shoe; exposed surfaces—poison ivy)	Tinea versicolor	Chest, back, abdomen, proximal extremities	Scaly hyper- or hypopigmented macules
Psoriasis	Elbows, knees, scalp, lower back, fingernails (may be generalized)	Papules and plaques covered with silvery scale; nails have pits	Candidiasis	Groin, beneath breasts, vagina, oral cavity	Erythematous macerated areas with satellite pustules; white, friable patches on mucous membranes
Lichen planus	Wrists, ankles, mouth (may be widespread)	Violaceous flat-topped papules and plaques	Dermatophytosis	Feet, groin, beard, or scalp	Varies with site (e.g., tinea corporis—scaly annular plaque)
Keratosis pilaris	Extensor surfaces of arms and thighs, buttocks	Keratotic follicular papules with surrounding erythema	Scabies	Groin, axillae, between fingers and toes, beneath breasts	Excoriated papules, burrows, pruritus
Melasma	Forehead, cheeks, temples, upper lip	Tan to brown patches	Insect bites	Anywhere	Erythematous papules with central puncta
Vitiligo	Periorificial, trunk, extensor surfaces of extremities, flexor wrists, axillae	Chalk-white macules	Cherry angioma Keloid Dermatofibroma	Trunk Anywhere (site of previous injury) Anywhere	Red, blood-filled papules Firm tumor, pink, purple, or brown Firm red to brown nodule that shows dimpling of overlying skin with lateral compression

(*Continued*)

TABLE 59-4 Selected Common Dermatologic Conditions (Continued)

DIAGNOSIS	COMMON DISTRIBUTION	USUAL MORPHOLOGY	DIAGNOSIS	COMMON DISTRIBUTION	USUAL MORPHOLOGY
Actinic keratosis	Sun-exposed areas	Skin-colored or red-brown macule or papule with dry, rough, adherent scale	Acrochordons (skin tags)	Groin, axilla, neck	Fleshy papules
Basal cell carcinoma	Face	Papule with pearly, telangiectatic border on sun-damaged skin	Urticaria	Anywhere	Wheals, sometimes with surrounding flare; pruritus
Squamous cell carcinoma	Face, especially lower lip, ears	Indurated and possibly hyperkeratotic lesions often showing ulceration and/or crusting	Transient acantholytic dermatosis Xerosis	Trunk, especially anterior chest Extensor extremities, especially legs	Erythematous papules Dry, erythematous, scaling patches; pruritus

finding of scaling papules, which are present in psoriasis or atopic dermatitis, places the patient in a different diagnostic category than would hemorrhagic papules, which may indicate vasculitis or sepsis **(Figs. 59-4 and 59-5, respectively)**. It is also important to differentiate primary from secondary skin lesions. If the examiner focuses on linear erosions overlying an area of erythema and scaling, they may incorrectly assume that the erosion is the primary lesion and that the redness and scale are secondary, whereas the correct interpretation would be that the patient has a pruritic eczematous dermatitis with erosions caused by scratching.

APPROACH TO THE PATIENT

Skin Disorder

In examining the skin, it is usually advisable to assess the patient before taking an extensive history. This approach ensures that the entire cutaneous surface will be evaluated, and objective findings can be integrated with relevant historical data. Four basic features of a skin problem must be noted and considered during a physical examination: the *distribution* of the eruption, the *types* of primary and secondary lesions, the *shape* of individual lesions, and the *arrangement* of the lesions. An ideal skin examination includes evaluation of the skin, hair, and nails as well as the mucous membranes of the mouth, eyes, nose, nasopharynx, and anogenital region. In the initial examination, it is important that the patient be disrobed as completely as possible to minimize chances of missing important individual skin lesions and permit accurate assessment of the distribution of the eruption. The patient should first be viewed from a distance of about 1.5–2 m (4–6 ft) so that the general character of the skin and the distribution of lesions can be evaluated. Indeed, the distribution of lesions often correlates highly with diagnosis **(Fig. 59-6)**. For example, a hospitalized patient with a generalized erythematous exanthem is more likely to have a drug eruption than is a patient with a similar rash limited to the sun-exposed portions of the face. Once the distribution of the lesions has been established, the nature of the primary lesion must be determined. Thus, when lesions are distributed on elbows, knees, and scalp, the most likely possibility based solely on distribution is psoriasis or dermatitis herpetiformis **(Figs. 59-7 and 59-8, respectively)**. The primary lesion in psoriasis is a scaly papule that soon forms erythematous plaques covered with a white scale, whereas that of dermatitis herpetiformis is an urticarial papule that quickly becomes a small vesicle. In this manner, identification of the primary lesion directs the examiner toward the proper diagnosis. Secondary changes in skin can also be quite helpful. For example, scale represents excessive epidermis, while crust is the result of a discontinuous epithelial cell layer. Palpation of skin lesions can yield insight into the character of an eruption. Thus, red papules on the lower extremities that blanch with pressure can be a manifestation of many different diseases, but hemorrhagic red papules that do not blanch with pressure indicate palpable purpura characteristic of necrotizing vasculitis (Fig. 59-4).

The shape of lesions is also an important feature. Flat, round, erythematous papules and plaques are common in many cutaneous diseases. However, target-shaped lesions that consist in part of erythematous plaques are specific for erythema multiforme **(Fig. 59-9)**. Likewise, the arrangement of individual lesions is important. Erythematous papules and vesicles can occur in many conditions, but their arrangement in a specific linear array suggests an external etiology such as allergic contact dermatitis **(Fig. 59-10)** or primary irritant dermatitis. In contrast, lesions

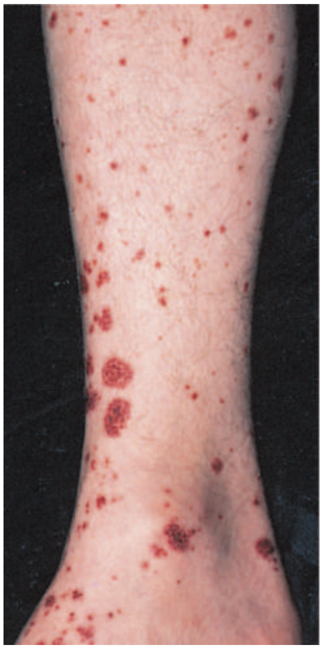

FIGURE 59-4 Necrotizing vasculitis. Palpable purpuric papules on the lower legs are seen in this patient with cutaneous small-vessel vasculitis. *(Courtesy of Robert Swerlick, MD; with permission.)*

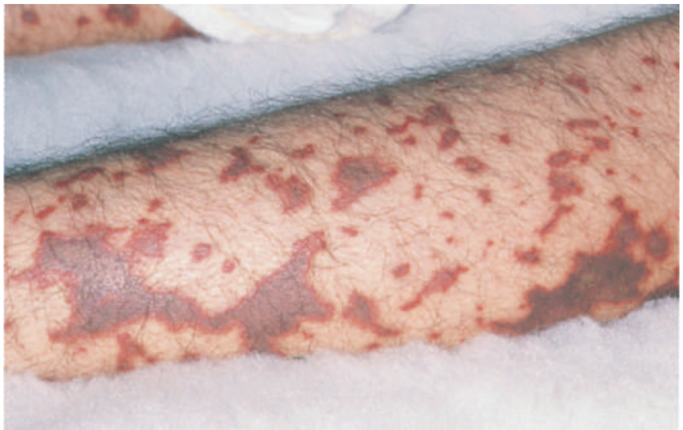

FIGURE 59-5 Meningococcemia. An example of fulminant meningococcemia with extensive angular purpuric patches. *(Courtesy of Stephen E. Gellis, MD; with permission.)*

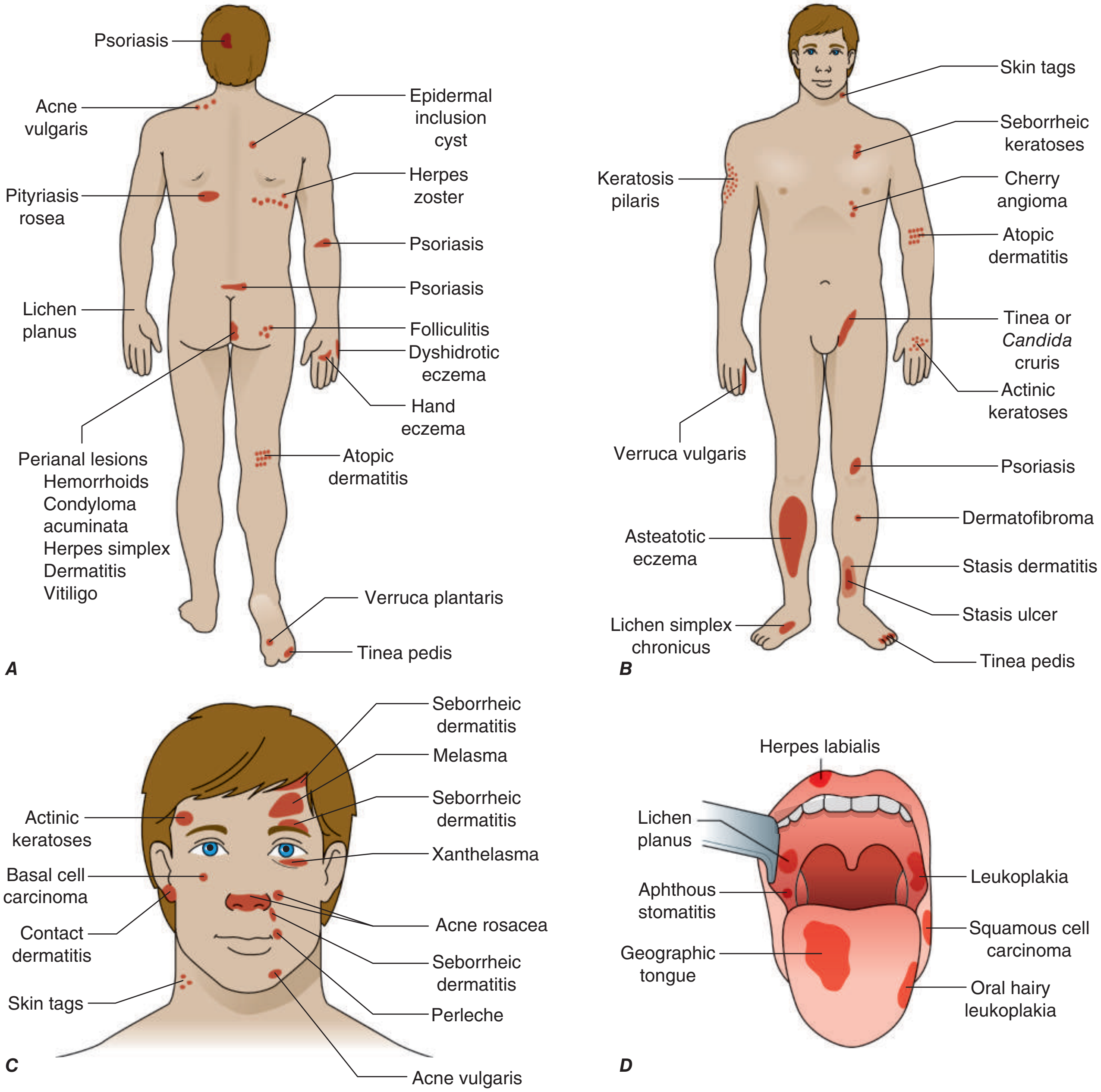
Psoriasis
Acne vulgaris
Epidermal inclusion cyst
Pityriasis rosea
Herpes zoster
Psoriasis
Psoriasis
Lichen planus
Folliculitis
Dyshidrotic eczema
Hand eczema
Perianal lesions
Hemorrhoids
Condyloma acuminata
Herpes simplex
Dermatitis
Vitiligo
Atopic dermatitis
Verruca plantaris
Tinea pedis
A
Skin tags
Seborrheic keratoses
Keratosis pilaris
Cherry angioma
Atopic dermatitis
Tinea or *Candida* cruris
Actinic keratoses
Verruca vulgaris
Psoriasis
Dermatofibroma
Asteatotic eczema
Stasis dermatitis
Stasis ulcer
Lichen simplex chronicus
Tinea pedis
B
Seborrheic dermatitis
Melasma
Actinic keratoses
Seborrheic dermatitis
Xanthelasma
Basal cell carcinoma
Acne rosacea
Contact dermatitis
Seborrheic dermatitis
Skin tags
Perleche
Acne vulgaris
C
Herpes labialis
Lichen planus
Leukoplakia
Aphthous stomatitis
Squamous cell carcinoma
Geographic tongue
Oral hairy leukoplakia
D

FIGURE 59-6 Distribution of some common dermatologic diseases and lesions.

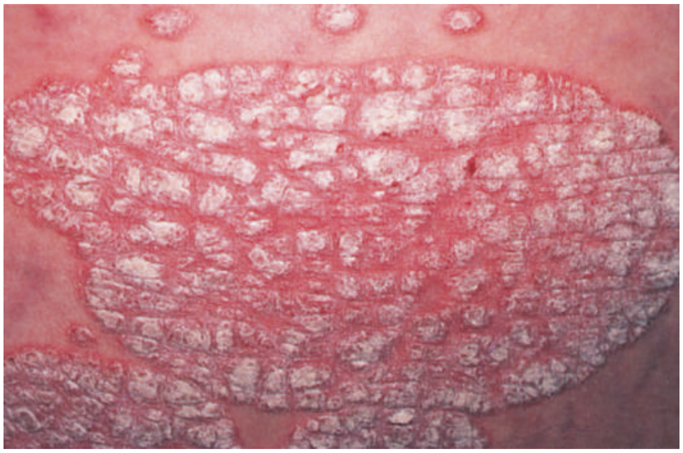

FIGURE 59-7 Psoriasis. This papulosquamous skin disease is characterized by small and large erythematous papules and plaques with overlying adherent silvery scale.

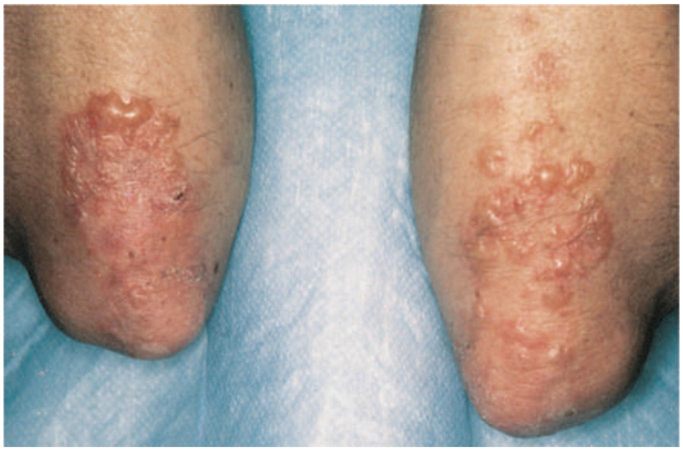

FIGURE 59-8 Dermatitis herpetiformis. This disorder typically displays pruritic, grouped papulovesicles on elbows, knees, buttocks, and posterior scalp. Vesicles are often excoriated due to associated pruritus.

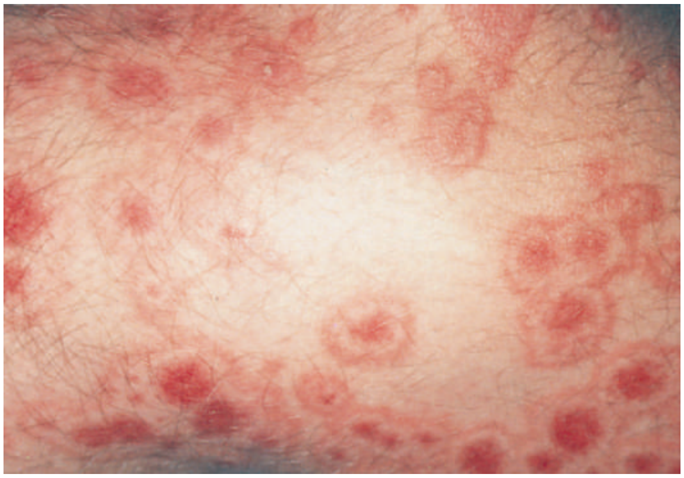

FIGURE 59-9 Erythema multiforme. This eruption is characterized by multiple erythematous plaques with a target or iris morphology. It usually represents a hypersensitivity reaction to drugs (e.g., sulfonamides) or infections (e.g., herpes simplex virus). *(Courtesy of the Yale Resident's Slide Collection; with permission.)*

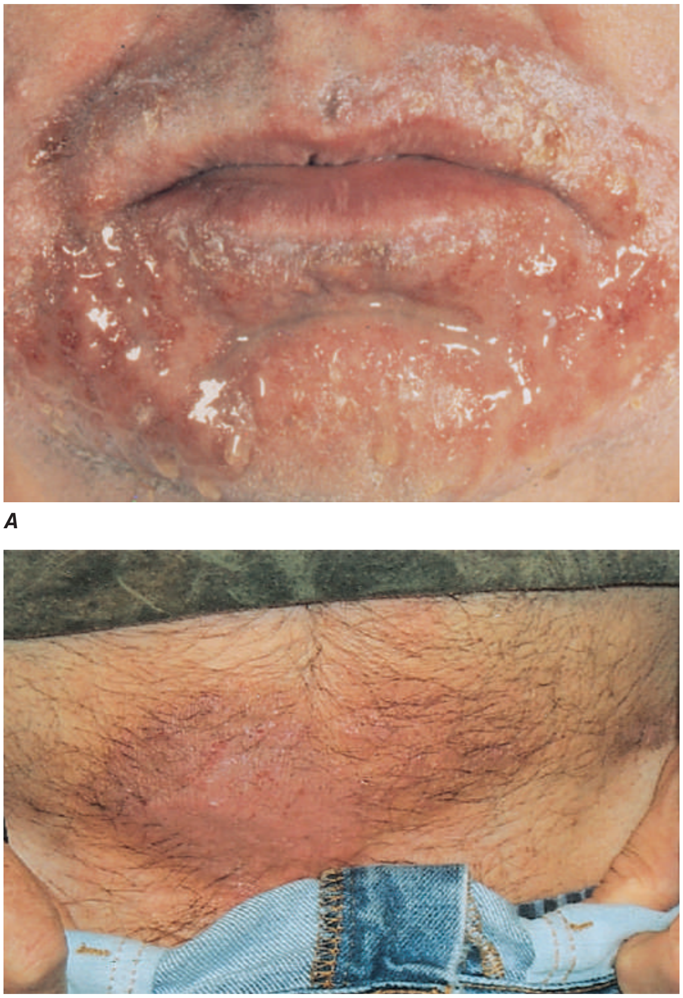

FIGURE 59-10 Allergic contact dermatitis (ACD). ***A.*** An example of ACD in its acute phase, with sharply demarcated, weeping, eczematous plaques in a perioral distribution. ***B.*** ACD in its chronic phase, with an erythematous, lichenified, weeping plaque on skin chronically exposed to nickel in a metal snap. *(B, Courtesy of Robert Swerlick, MD; with permission.)*

with a generalized arrangement are common and suggest a systemic etiology.

As in other branches of medicine, a complete history should be obtained to emphasize the following features:

1. Evolution of lesions
    a. Site of onset
    b. Manner in which the eruption progressed or spread
    c. Duration
    d. Periods of resolution or improvement in chronic eruptions
2. Symptoms associated with the eruption
    a. Itching, burning, pain, numbness
    b. What, if anything, has relieved symptoms
    c. Time of day when symptoms are most severe
3. Current or recent medications (prescribed as well as over-the-counter)
4. Associated systemic symptoms (e.g., malaise, fever, arthralgias)
5. Ongoing or previous illnesses
6. History of allergies
7. Presence of photosensitivity
8. Review of systems
9. Family history (particularly relevant for patients with melanoma, atopy, psoriasis, or acne)
10. Social, sexual, or travel history

DIAGNOSTIC TECHNIQUES

Many skin diseases can be diagnosed on the basis of gross clinical appearance, but sometimes relatively simple diagnostic procedures can yield valuable information. In most instances, they can be performed at the bedside with a minimum of equipment.

Skin Biopsy A skin biopsy is a straightforward minor surgical procedure; however, it is important to biopsy a lesion that is most likely to yield diagnostic findings. This decision may require expertise in skin diseases and knowledge of superficial anatomic structures in selected areas of the body. In this procedure, a small area of skin is anesthetized with 1% lidocaine with or without epinephrine. The skin lesion in question can be excised or saucerized with a scalpel or removed by punch biopsy. In the latter technique, a punch is pressed against the surface of the skin and rotated with downward pressure until it penetrates to the subcutaneous tissue. The circular biopsy is then lifted with forceps, and the bottom is cut with iris scissors. Biopsy sites may or may not need suture closure, depending on size and location.

KOH Preparation A potassium hydroxide (KOH) preparation is performed on scaling skin lesions where a fungal infection is suspected. The edge of such a lesion is scraped gently with a no. 15 scalpel blade. The removed scale is collected on a glass microscope slide, treated with 1 or 2 drops of a solution of 10–20% KOH, and placement of a cover slip. KOH dissolves keratin and allows easier visualization of fungal elements. Brief heating of the slide accelerates dissolution of keratin. When the preparation is viewed under the microscope, the refractile hyphae are seen more easily when the light intensity is reduced and the condenser is lowered. This technique can be used to identify hyphae in dermatophyte infections, pseudohyphae and budding yeasts in *Candida* infections, and "spaghetti and meatballs" yeast forms in tinea versicolor. The same sampling technique can be used to obtain scale for culture of selected pathogenic organisms.

Tzanck Smear A Tzanck smear is a cytologic technique most often used in the diagnosis of herpesvirus infections (herpes simplex virus [HSV] or varicella-zoster virus [VZV]) **(see Figs. 198-1 and 198-3)**. An early vesicle, not a pustule or crusted lesion, is unroofed, and the base of the lesion is scraped gently with a scalpel blade. The material is placed on a glass slide, air-dried, and stained with Giemsa or Wright's stain. Multinucleated epithelial giant cells suggest the presence of HSV or VZV; culture, immunofluorescence microscopy, or genetic testing must be performed to identify the specific virus.

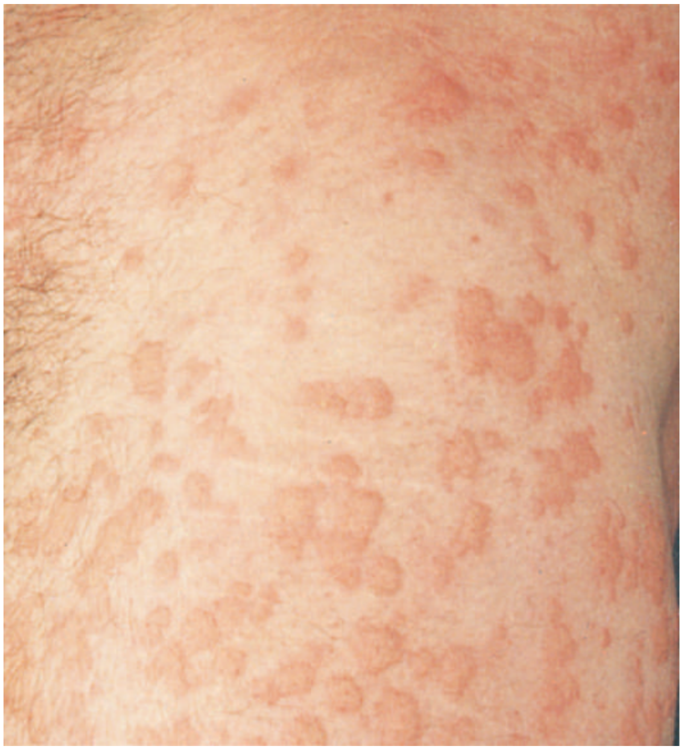

FIGURE 59-11 Urticaria. Discrete and confluent, edematous, erythematous papules and plaques are characteristic of this whealing eruption.

Diascopy Diascopy is designed to assess whether a skin lesion will blanch with pressure as, for example, in determining whether a red lesion is hemorrhagic or simply blood-filled. Urticaria (Fig. 59-11) will blanch with pressure, whereas a purpuric lesion caused by necrotizing vasculitis (Fig. 59-4) will not. Diascopy is performed by pressing a microscope slide or magnifying lens against a lesion and noting the amount of blanching that occurs. Granulomas often have an opaque to transparent, brown-pink "apple jelly" appearance on diascopy.

Dermoscopy Dermoscopy is a noninvasive method of examining the skin surface that uses a high-quality magnifying lens and a specialized light source (i.e., a dermatoscope). Dermoscopy identifies skin structures, colors, and patterns that are not visible to the naked eye. It is particularly useful in the evaluation of pigmented skin lesions.

Wood's Light A Wood's lamp generates 360-nm ultraviolet ("black") light that can be used to aid the evaluation of certain skin disorders. For example, a Wood's lamp will cause erythrasma (a superficial, intertriginous infection caused by *Corynebacterium minutissimum*) to show a characteristic coral pink color, and wounds colonized by *Pseudomonas* will appear pale blue. Tinea capitis caused by certain dermatophytes (e.g., *Microsporum canis* or *M. audouinii*) exhibits a yellow fluorescence. Pigmented lesions of the epidermis such as freckles are accentuated, while dermal pigment such as postinflammatory hyperpigmentation fades under a Wood's light. Vitiligo (Fig. 59-12) appears totally white under a Wood's lamp, and previously unsuspected areas of involvement often become apparent. A Wood's lamp may also aid in the demonstration of tinea versicolor, detection of sites of depigmentation within and/or surrounding melanomas, and recognition of ash leaf spots in patients with tuberous sclerosis.

Patch Tests Patch testing is designed to document sensitivity to a specific antigen. In this procedure, a battery of suspected allergens is applied to the patient's back under occlusive dressings and allowed to remain in contact with the skin for 48 h. The dressings are removed, and the area is examined for evidence of delayed hypersensitivity reactions (e.g., erythema, edema, or papulovesicles). This test is best performed by physicians with special expertise in patch testing and is often helpful in the evaluation of patients with chronic dermatitis.

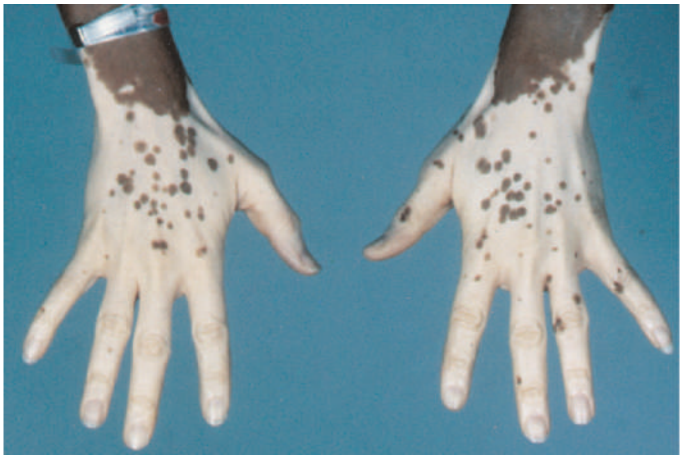

FIGURE 59-12 Vitiligo. Characteristic lesions display an acral distribution and striking depigmentation as a result of loss of melanocytes.

FURTHER READING

Bolognia JL et al (eds): *Dermatology*, 4th ed. Philadelphia, Elsevier, 2018.

James WD: *Andrews' Diseases of the Skin: Clinical Dermatology*, 13th ed. Philadelphia, Elsevier, 2019.

Kang S et al (eds): *Fitzpatrick's Dermatology in General Medicine*, 9th ed. New York, McGraw-Hill, 2019.

60

Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders

Justin T. Cheeley, Leslie P. Lawley

ECZEMA AND DERMATITIS

Eczema is a type of dermatitis, and these terms are often used synonymously (e.g., atopic eczema or atopic dermatitis [AD]). Eczema is a reaction pattern that presents with variable clinical findings and the common histologic finding of *spongiosis* (intercellular edema of the epidermis). Eczema is the final common expression for a number of disorders, including those discussed in the following sections. Primary lesions may include erythematous macules, papules, and vesicles, which can coalesce to form patches and plaques. In severe eczema, secondary lesions from infection or excoriation, marked by weeping and crusting, may predominate. In chronic eczematous conditions, *lichenification* (cutaneous hypertrophy and accentuation of normal skin markings) may alter the characteristic appearance of eczema.

ATOPIC DERMATITIS

AD is the cutaneous expression of the atopic state, characterized by a family history of asthma, allergic rhinitis, or eczema. The prevalence of AD is increasing worldwide. Some of its features are shown in Table 60-1.

The etiology of AD is only partially defined, but there is a clear genetic predisposition. When both parents are affected by AD, >80% of their children manifest the disease. When only one parent is affected, the prevalence drops to slightly >50%. A characteristic defect in AD that contributes to the pathophysiology is an impaired epidermal barrier. In many patients, a mutation in the gene encoding

TABLE 60-1 Clinical Features of Atopic Dermatitis

1. Pruritus and scratching
2. Course marked by exacerbations and remissions
3. Lesions typical of eczematous dermatitis
4. Personal or family history of atopy (asthma, allergic rhinitis, food allergies, or eczema)
5. Clinical course lasting >6 weeks
6. Lichenification of skin
7. Presence of dry skin

filaggrin, a structural protein in the stratum corneum, is responsible. Patients with AD may display a variety of immunoregulatory abnormalities, including increased IgE synthesis; increased serum IgE levels; and impaired, delayed-type hypersensitivity reactions.

The clinical presentation often varies with age. Half of patients with AD present within the first year of life, and 80% present by 5 years of age. About 80% ultimately coexpress allergic rhinitis or asthma. The infantile pattern is characterized by weeping inflammatory patches and crusted plaques on the face, neck, and extensor surfaces. The childhood and adolescent patterns are typified by dermatitis of flexural skin, particularly in the antecubital and popliteal fossae **(Fig. 60-1)**. AD may resolve spontaneously, but approximately 40% of all individuals affected as children will have dermatitis in adult life. The distribution of lesions in adults may be similar to those seen in childhood; however, adults frequently have localized disease manifesting as lichen simplex chronicus or hand eczema (see below). In patients with localized disease, AD may be suspected because of a typical personal or family history or the presence of cutaneous stigmata of AD such as perioral pallor, an extra fold of skin beneath the lower eyelid (Dennie-Morgan folds), increased palmar skin markings, and an increased incidence of cutaneous infections, particularly with *Staphylococcus aureus*. Regardless of other manifestations, pruritus is a prominent characteristic of AD in all age groups and is exacerbated by dry skin. Many of the cutaneous findings in affected patients, such as lichenification, are secondary to rubbing and scratching.

TREATMENT

Atopic Dermatitis

Therapy for AD should include avoidance of cutaneous irritants, adequate moisturization through the application of emollients, judicious use of topical anti-inflammatory agents, and prompt treatment of secondary infection. Patients should be instructed to bathe no more often than daily, using warm or cool water, and to use only mild bath soap. Immediately after bathing, while the skin is still moist, a topical anti-inflammatory agent in a cream or ointment base should be applied to areas of dermatitis, and all other skin areas should be lubricated with a moisturizer. Approximately 30 g of a topical agent is required to cover the entire body surface of an average adult.

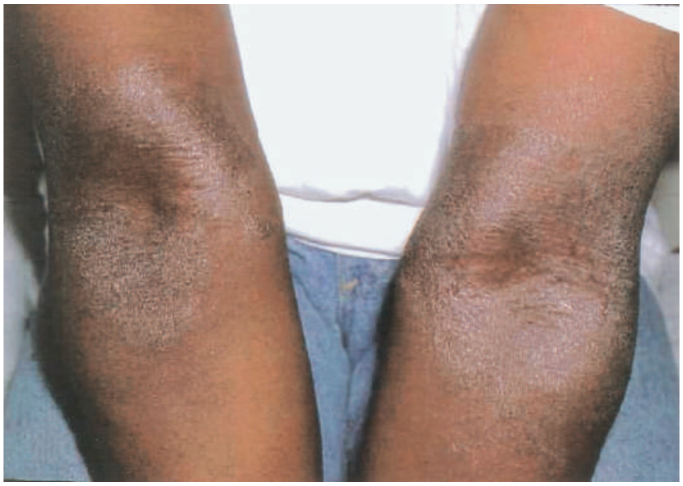

FIGURE 60-1 Atopic dermatitis. Hyperpigmentation, lichenification, and scaling in the antecubital fossae are seen in this patient with atopic dermatitis. *(Courtesy of Robert Swerlick, MD.)*

Low- to mid-potency topical glucocorticoids are employed in most treatment regimens for AD. Skin atrophy and the potential for systemic absorption are constant concerns, especially with more potent agents. Low-potency topical glucocorticoids or nonglucocorticoid anti-inflammatory agents should be selected for use on the face and in intertriginous areas to minimize the risk of skin atrophy. Several nonglucocorticoid anti-inflammatory agents approved by the U.S. Food and Drug Administration (FDA) are available for topical use in AD: tacrolimus ointment, pimecrolimus cream, crisaborole ointment, and ruxolitinib cream. These agents do not cause skin atrophy, nor do they suppress the hypothalamic-pituitary-adrenal axis. The first two agents are topical calcineurin inhibitors (TCIs), crisaborole is a phosphodiesterase-4 inhibitor, and ruxolitinib is a Janus kinase (JAK) inhibitor. Concerns regarding the potential for lymphomas in patients treated with TCIs have largely been unfounded. Currently, all of these agents are more costly than topical glucocorticoids. Barrier-repair products that attempt to restore the impaired epidermal barrier are also nonglucocorticoid agents and are gaining popularity in the treatment of AD.

Secondary infection of eczematous skin may lead to exacerbation of AD. Crusted and weeping skin lesions may be infected with *S. aureus*. When secondary infection is suspected, eczematous lesions should be cultured, and patients treated with systemic antibiotics active against *S. aureus*. The initial use of penicillinase-resistant penicillins or cephalosporins is preferable. Dicloxacillin or cephalexin (250 mg qid for 7–10 days) is generally adequate for adults; however, antibiotic selection must be directed by culture results and clinical response. More than 50% of *S. aureus* isolates are now methicillin resistant in some communities. Current recommendations for the treatment of infection with these community-acquired methicillin-resistant *S. aureus* (CA-MRSA) strains in adults include trimethoprim-sulfamethoxazole (one double-strength tablet bid), minocycline (100 mg bid), doxycycline (100 mg bid), or clindamycin (300–450 mg qid). Duration of therapy should be 7–10 days. Inducible resistance may limit clindamycin's usefulness. Such resistance can be detected by the double-disk diffusion test, which should be ordered if the isolate is erythromycin resistant and clindamycin sensitive. As an adjunct, antibacterial washes or dilute sodium hypochlorite baths (0.005% bleach) and intermittent nasal mupirocin may be useful.

Control of pruritus is essential for treatment, as AD often represents "an itch that rashes." Topical anti-inflammatories may reduce itch, but when inadequate, antihistamines are most often used to control pruritus. Diphenhydramine (25 mg every 4–6 h), hydroxyzine (10–25 mg every 6 h), and doxepin (10–25 mg at bedtime) are useful primarily due to their sedating action. Higher doses of these agents may be required, but sedation can become bothersome. Patients need to be counseled about driving or operating heavy equipment after taking these medications. When used at bedtime, sedating antihistamines may improve the patient's sleep. Although they are effective in urticaria, nonsedating antihistamines and selective H_2 blockers are of little use in controlling the pruritus of AD.

Treatment with systemic glucocorticoids should be limited to severe exacerbations unresponsive to topical therapy. In the patient with chronic AD, therapy with systemic glucocorticoids will generally clear the skin only briefly, and cessation of the systemic therapy will invariably be accompanied by a return, if not a worsening, of the dermatitis. For chronic severe AD poorly responsive to standard topical regimens, systemic agents may be considered. Cyclosporine is approved for treatment of severe recalcitrant AD in some European countries. Monitoring of renal function and secondary infections is required. Dupilumab, an interleukin (IL)-4 receptor blocker, is FDA approved for use in patients 6 months of age and older.

Tralokinumab, an IL-13 inhibitor, is approved for use in adults. These agents provide more targeted immunomodulation and a better safety profile than cyclosporine. Abrocitinib and upadacitinib are JAK inhibitors FDA approved to treat moderate to severe AD. Patients who do not respond to conventional therapies should be considered for patch testing to rule out allergic contact dermatitis (ACD). The role of dietary allergens in AD is controversial, and there is little evidence that they play any role outside of infancy, during which a small percentage of patients with AD may be affected by food allergens.

LICHEN SIMPLEX CHRONICUS

Lichen simplex chronicus may represent the end stage of a variety of pruritic and eczematous disorders, including AD. It consists of a circumscribed plaque or plaques of lichenified skin due to chronic scratching or rubbing. Common areas involved include the posterior nuchal region, dorsum of the feet, and ankles. Treatment of lichen simplex chronicus centers on breaking the cycle of chronic itching and scratching. High-potency topical glucocorticoids are helpful in most cases, but in recalcitrant cases, application of topical glucocorticoids under occlusion or intralesional injection of glucocorticoids may be required.

CONTACT DERMATITIS

Contact dermatitis is an inflammatory skin process caused by an exogenous agent or agents that directly or indirectly injure the skin. In *irritant* contact dermatitis (ICD), this injury is caused by an inherent characteristic of a compound—for example, a concentrated acid or base. Agents that cause *allergic* contact dermatitis (ACD) induce an antigen-specific immune response (e.g., poison ivy dermatitis). The clinical lesions of contact dermatitis may be acute (wet and edematous) or chronic (dry, thickened, and scaly), depending on the persistence of the insult **(see Chap. 59, Fig. 59-10)**.

Irritant Contact Dermatitis ICD is generally well demarcated and often localized to areas of thin skin (eyelids, intertriginous areas) or areas where the irritant was occluded. Lesions may range from minimal skin erythema to areas of marked edema, vesicles, and ulcers. Prior exposure to the offending agent is not necessary, and the reaction develops in minutes to a few hours. Chronic low-grade irritant dermatitis is the most common type of ICD, and the most common area of involvement is the hands (see below). The most common irritants encountered are chronic wet work, soaps, and detergents. Treatment should be directed toward the avoidance of irritants and the use of protective gloves or clothing.

Allergic Contact Dermatitis ACD is a manifestation of delayed-type hypersensitivity mediated by memory T lymphocytes in the skin. Prior exposure to the offending agent is necessary to develop the hypersensitivity reaction, which may take as little as 12 h or as long as 72 h to develop. The most common cause of ACD is exposure to plants, especially to members of the family Anacardiaceae, including the genus *Toxicodendron*. Poison ivy, poison oak, and poison sumac are members of this genus and cause an allergic reaction marked by erythema, vesiculation, and severe pruritus. The eruption is often linear or angular, corresponding to areas where plants have touched the skin. The sensitizing antigen common to these plants is urushiol, an oleoresin containing the active ingredient pentadecylcatechol. The oleoresin may adhere to skin, clothing, tools, and pets, and contaminated articles may cause dermatitis even after prolonged storage. Blister fluid does not contain urushiol and is not capable of inducing skin eruption in exposed subjects.

TREATMENT
Contact Dermatitis

If contact dermatitis is suspected and an offending agent is identified and removed, the eruption will resolve. Usually, treatment with high-potency topical glucocorticoids is enough to relieve symptoms while the dermatitis runs its course. For patients who require systemic therapy, daily oral prednisone—beginning at 1 mg/kg, but usually ≤60 mg/d—is sufficient. The dose should be tapered over 2–3 weeks, and each daily dose should be taken in the morning with food.

Identification of a contact allergen can be a difficult and time-consuming task. ACD should be suspected in patients with dermatitis unresponsive to conventional therapy or with an unusual and patterned distribution. Patients should be questioned carefully regarding occupational exposures and topical medications. Common sensitizers include preservatives in topical preparations, nickel sulfate, potassium dichromate, thimerosal, neomycin sulfate, fragrances, formaldehyde, and rubber-curing agents. Patch testing is helpful in identifying these agents but should not be attempted when patients have widespread active dermatitis or are taking systemic glucocorticoids.

HAND ECZEMA

Hand eczema is a very common, chronic skin disorder in which both exogenous and endogenous factors play important roles. It may be associated with other cutaneous disorders such as AD and contact with various agents may be involved. Hand eczema represents a large proportion of cases of occupation-associated skin disease. Chronic, excessive exposure to water and detergents, harsh chemicals, or allergens may initiate or aggravate this disorder. It may present with dryness and cracking of the skin of the hands as well as with variable amounts of erythema and edema. Often, the dermatitis will begin under rings, where water and irritants are trapped. *Dyshidrotic* eczema, a variant of hand eczema, presents with multiple, intensely pruritic, small papules and vesicles on the thenar and hypothenar eminences and the sides of the fingers **(Fig. 60-2)**. Lesions tend to occur in crops that slowly form crusts and then heal.

The evaluation of a patient with hand eczema should include an assessment of potential occupation-associated exposures. The history should be directed to identifying possible irritant or allergen exposures.

TREATMENT
Hand Eczema

Therapy for hand eczema is directed toward avoidance of irritants, identification of possible contact allergens, treatment of coexistent infection, and application of topical glucocorticoids. Whenever

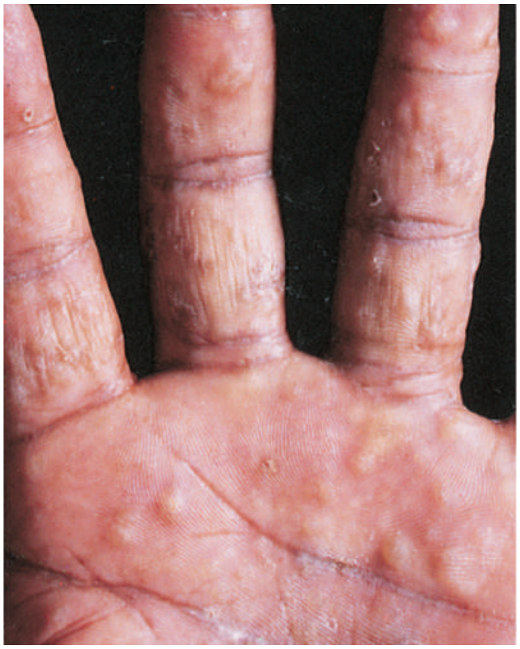

FIGURE 60-2 Dyshidrotic eczema. Deep-seated vesicles and papules on the lateral fingers. The disease is often associated with an atopic diathesis. *(Courtesy of Justin T. Cheeley, MD.)*

possible, the hands should be protected by gloves, preferably vinyl. The use of rubber gloves (latex) to protect dermatitic skin is sometimes associated with the development of hypersensitivity reactions to components of the gloves, which could be either a type I hypersensitivity reaction to the latex (manifested by the development of hives, itching, angioedema, and possibly anaphylaxis within minutes to hours of exposure) or a type IV hypersensitivity reaction to rubber accelerators (with worsening of eczematous eruptions days after exposure). Patients can be treated with cool moist compresses followed by application of a mid- to high-potency topical glucocorticoid in a cream or ointment base. As in AD, treatment of secondary infection is essential for good control. In addition, patients with hand eczema should be examined for dermatophyte infection by potassium hydroxide (KOH) preparation and culture (see below).

NUMMULAR ECZEMA

Nummular eczema is characterized by circular or oval "coinlike" lesions, beginning as small edematous papules that become crusted and scaly. The etiology of nummular eczema is unknown, but dry skin is a contributing factor. Common locations are the trunk or the extensor surfaces of the extremities, particularly on the pretibial areas or dorsum of the hands. Nummular eczema occurs more frequently in men and is most common in middle age. The treatment of nummular eczema is similar to that for AD.

ASTEATOTIC ECZEMA

Asteatotic eczema, also known as *xerotic eczema* or "winter itch," is a mildly inflammatory dermatitis that develops in areas of extremely dry skin, especially during the dry winter months. Clinically, there may be considerable overlap with nummular eczema. This form of eczema accounts for many physician visits because of the associated pruritus. Fine cracks and scale, with or without erythema, characteristically develop in areas of dry skin, especially on the anterior surfaces of the lower extremities in elderly patients. Asteatotic eczema responds well to topical moisturizers and the avoidance of cutaneous irritants. Overbathing and the use of harsh soaps exacerbate asteatotic eczema.

STASIS DERMATITIS AND STASIS ULCERATION

Stasis dermatitis develops on the lower extremities secondary to venous incompetence and chronic edema. Patients may give a history of deep venous thrombosis and may have evidence of vein removal or varicose veins. Early findings in stasis dermatitis consist of mild erythema and scaling associated with pruritus. The typical initial site of involvement is the medial aspect of the ankle, often over a distended vein **(Fig. 60-3)**.

Stasis dermatitis may become acutely inflamed, with crusting and exudate. In this state, it is easily confused with cellulitis. Of note, symmetrical and bilateral involvement is more likely stasis dermatitis, whereas unilateral involvement may represent cellulitis. Chronic stasis dermatitis is often associated with dermal fibrosis that is recognized clinically as brawny edema of the skin. As the disorder progresses, the dermatitis becomes progressively pigmented due to chronic erythrocyte extravasation leading to cutaneous hemosiderin deposition. Stasis dermatitis may be complicated by secondary infection and contact dermatitis. Severe stasis dermatitis may precede the development of stasis ulcers.

TREATMENT

Stasis Dermatitis and Stasis Ulceration

Patients with stasis dermatitis and stasis ulceration benefit greatly from leg elevation and the routine use of compression stockings with a gradient of at least 30–40 mmHg. Stockings providing less compression, such as antiembolism hose, are poor substitutes. Use of emollients and/or mid-potency topical glucocorticoids and avoidance of irritants are also helpful in treating stasis dermatitis. Protection of the legs from injury, including scratching, and control of chronic edema are essential to prevent ulcers. Diuretics may be required to adequately control chronic edema.

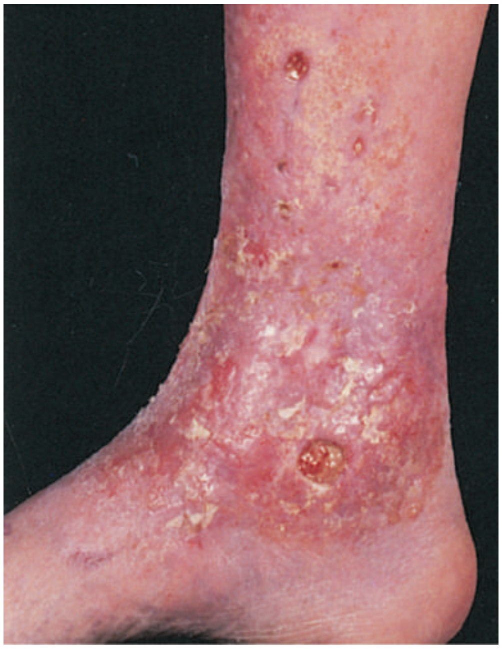

FIGURE 60-3 Stasis dermatitis. Brawny red patches overlaid with yellow scale primarily on the distal anteromedial leg. Phlebectasia around the medial malleolus is visible. *(Courtesy of Justin T. Cheeley, MD.)*

Stasis ulcers are difficult to treat, and resolution is slow. It is extremely important to elevate the affected limb as much as possible. The ulcer should be kept clear of necrotic material by gentle debridement and covered with a semipermeable dressing and a compression dressing or compression stocking. Glucocorticoids should not be applied to ulcers, because they may retard healing; however, they may be applied to the surrounding skin to control itching, scratching, and additional trauma. Superficial bacterial cultures of chronic stasis ulcers often yield polymicrobial colonizers and are of little utility in determination of secondary infection. Care must be taken to exclude treatable causes of leg ulcers (hypercoagulation, vasculitis, arterial insufficiency) before beginning the chronic management outlined above.

SEBORRHEIC DERMATITIS

Seborrheic dermatitis is a common, chronic disorder characterized by greasy scales overlying erythematous patches or plaques. Induration and scale are generally less prominent than in psoriasis, but clinical overlap exists between these diseases ("sebopsoriasis"). The most common location is in the scalp, where it may be recognized as severe dandruff. On the face, seborrheic dermatitis affects the eyebrows, eyelids, glabella, and nasolabial folds **(Fig. 60-4)**. Scaling of the external auditory canal is common in seborrheic dermatitis. In addition, the postauricular areas often become macerated and tender. Seborrheic dermatitis may also develop in the central chest, axilla, groin, submammary folds, and gluteal cleft. Rarely, it may cause widespread generalized dermatitis. Pruritus is variable.

Seborrheic dermatitis may be evident within the first few weeks of life, and within this context, it typically occurs in the scalp ("cradle cap"), face, or groin. It is rarely seen in children beyond infancy but becomes evident again during adolescent and adult life. Although it is frequently seen in patients with Parkinson's disease, in those who have had cerebrovascular accidents, and in those with HIV infection, the overwhelming majority of individuals with seborrheic dermatitis have no underlying disorder.

TREATMENT

Seborrheic Dermatitis

Treatment with low-potency topical glucocorticoids in conjunction with a topical antifungal agent, such as ketoconazole cream

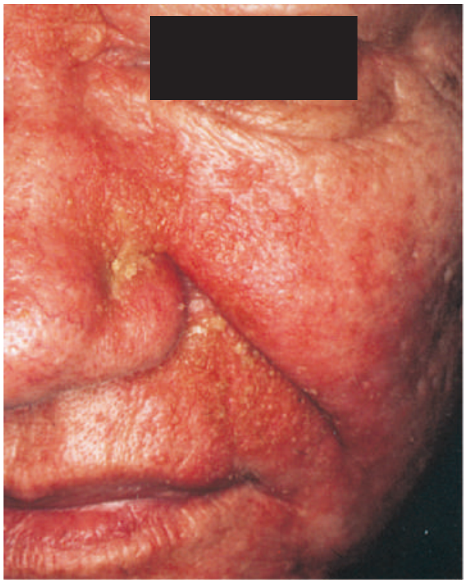
[redacted]

FIGURE 60-4 Seborrheic dermatitis. Central facial erythema with overlying greasy, yellowish scale is seen in this patient. *(Courtesy of Jean Bolognia, MD; with permission.)*

or ciclopirox cream, is often effective. The scalp and beard areas may benefit from antidandruff shampoos, which should be left in place 3–5 min before rinsing. High-potency topical glucocorticoid solutions (betamethasone or clobetasol) are effective for control of severe scalp involvement. High-potency glucocorticoids should not be used on the face because this treatment is often associated with steroid-induced rosacea or atrophy.

PAPULOSQUAMOUS DISORDERS (TABLE 60-2)

PSORIASIS

Psoriasis is one of the most common dermatologic diseases, affecting up to 2% of the world's population. It is an immune-mediated disease clinically characterized by erythematous, sharply demarcated papules and rounded plaques covered by silvery micaceous scale. The skin lesions of psoriasis are variably pruritic. Traumatized areas often develop lesions of psoriasis (the *Koebner* or isomorphic phenomenon). In addition, other external factors may exacerbate psoriasis, including infections, stress, and medications (lithium, beta blockers, and antimalarial drugs).

The most common variety of psoriasis is called *plaque-type*. Patients with plaque-type psoriasis have stable, slowly enlarging plaques, which remain basically unchanged for long periods of time. The most commonly involved areas are the elbows, knees, gluteal cleft, and scalp. Involvement tends to be symmetric. Plaque psoriasis generally develops slowly and runs an indolent course. It rarely remits spontaneously. *Inverse psoriasis* affects the intertriginous regions, including the axilla, groin, submammary region, and navel; it also tends to affect the scalp, palms, and soles. The individual lesions are sharply demarcated plaques **(see Chap. 59, Fig. 59-7)**, but they may be moist and without scale due to their locations.

Guttate psoriasis (eruptive psoriasis) is most common in children and young adults. It develops acutely in individuals without psoriasis or in those with chronic plaque psoriasis. Patients present with many small erythematous, scaling papules, frequently after upper respiratory tract infection with β-hemolytic streptococci. The differential diagnosis should include pityriasis rosea and secondary syphilis.

In *pustular psoriasis*, patients may have disease localized to the palms and soles, or the disease may be generalized. Regardless of the extent of disease, the skin is erythematous, with pustules and variable scale. Localized to the palms and soles, it is easily confused with dyshidrotic eczema. When it is generalized, episodes are characterized by fever (39°–40°C [102.2°–104.0°F]) lasting several days, an accompanying generalized eruption of sterile pustules, and a background of intense erythema; patients may become erythrodermic. Episodes of fever and pustules are recurrent. Local irritants, pregnancy, medications, infections, and systemic glucocorticoid withdrawal can precipitate this form of psoriasis. Oral retinoids are the treatment of choice in nonpregnant patients.

Fingernail involvement, appearing as punctate pitting, onycholysis, nail thickening, or subungual hyperkeratosis, may be a clue to the diagnosis of psoriasis when the clinical presentation is not classic.

According to the National Psoriasis Foundation, up to 30% of patients with psoriasis have psoriatic arthritis (PsA). It develops most commonly between the ages of 30 and 50 years. There are five subtypes of PsA: symmetric PsA, asymmetric PsA, distal PsA, spondylitis, and arthritis mutilans. Approximately 50% of PsA is classified as symmetric, which may resemble rheumatoid arthritis. Asymmetric arthritis comprises about 35% of cases. It can involve any joint and may present as "sausage digits." Distal PsA is the classic form; however, it occurs in only about 5% of patients with PsA. It can involve fingers and toes; fingernails and toenails are often dystrophic, including nail pitting. Spondylitis also occurs in ~5% of patients with PsA. Arthritis mutilans is severe and deforming and affects primarily the small joints of the hands and feet. It accounts for fewer than 5% of PsA cases.

An increased risk of metabolic syndrome, including increased morbidity and mortality from cardiovascular events, has been demonstrated in psoriasis patients. Appropriate screening tests should be performed. The etiology of psoriasis is still poorly understood, but there is clearly a genetic component to the disease. In various studies, 30–50% of patients with psoriasis report a positive family history. Psoriatic lesions contain infiltrates of activated T cells that are thought

TABLE 60-2 Papulosquamous Disorders

	CLINICAL FEATURES	OTHER NOTABLE FEATURES	HISTOLOGIC FEATURES
Psoriasis	Sharply demarcated, erythematous plaques with mica-like scale; predominantly on elbows, knees, and scalp; atypical forms may localize to intertriginous areas; eruptive forms may be associated with infection	May be aggravated by certain drugs, infection; severe forms seen in association with HIV	Acanthosis, vascular proliferation
Lichen planus	Purple polygonal papules marked by severe pruritus; lacy white markings, especially associated with mucous membrane lesions	Certain drugs may induce: thiazides, antimalarial drugs	Interface dermatitis
Pityriasis rosea	Rash often preceded by herald patch; oval to round plaques with trailing scale; most often affects trunk; eruption lines up in skinfolds giving a "fir tree–like" appearance; generally spares palms and soles	Variable pruritus; self-limited, resolving in 2–8 weeks; may be imitated by secondary syphilis	Pathologic features often nonspecific
Dermatophytosis	Polymorphous appearance depending on dermatophyte, body site, and host response; sharply defined to ill-demarcated scaly plaques with or without inflammation; may be associated with hair loss	KOH preparation may show branching hyphae; culture helpful	Hyphae and neutrophils in stratum corneum

Abbreviations: HIV, human immunodeficiency virus; KOH, potassium hydroxide.

TABLE 60-3 FDA-Approved Systemic Therapy for Psoriasis

		ADMINISTRATION		
AGENT	**MEDICATION CLASS**	**ROUTE**	**FREQUENCY**	**ADVERSE EVENTS (SELECTED)**
Methotrexate	Antimetabolite	Oral	Weekly	Hepatotoxicity, pulmonary toxicity, pancytopenia, potential for increased malignancies, ulcerative stomatitis, nausea, diarrhea, teratogenicity
Acitretin	Retinoid	Oral	Daily	Teratogenicity, hepatotoxicity, hyperostosis, hyperlipidemia/pancreatitis, depression, ophthalmologic effects, pseudotumor cerebri
Cyclosporine	Calcineurin inhibitor	Oral	Twice daily	Renal dysfunction, hypertension, hyperkalemia, hyperuricemia, hypomagnesemia, hyperlipidemia, increased risk of malignancies
Apremilast	Phosphodiesterase type 4 inhibitor	Oral	Twice daily[a]	Hypersensitivity reaction, depression, nausea, diarrhea, vomiting, dyspepsia, weight loss, headache, fatigue
Deucravacitinib	Tyrosine kinase 2 inhibitor	Oral	Daily	Upper respiratory infections, elevated creatine phosphokinase, herpes simplex reactivation, folliculitis

[a] Initial dose escalation is required.

Abbreviation: FDA, Food and Drug Administration.

to elaborate cytokines responsible for keratinocyte hyperproliferation, which results in the characteristic clinical findings. Agents inhibiting T-cell activation, clonal expansion, or release of proinflammatory cytokines are often effective for the treatment of severe psoriasis (see below).

TREATMENT

Psoriasis

Treatment of psoriasis depends on the type, location, and extent of disease. All patients should be instructed to avoid excess drying or irritation of their skin and to maintain adequate cutaneous hydration. Most cases of localized, plaque-type psoriasis can be managed with mid-potency topical glucocorticoids, although their long-term use is often accompanied by loss of effectiveness (tachyphylaxis) and atrophy of the skin. A topical vitamin D analogue (calcipotriene) and a retinoid (tazarotene) are also efficacious in the treatment of limited psoriasis and have largely replaced other topical agents such as coal tar, salicylic acid, and anthralin.

Ultraviolet (UV) light, natural or artificial, is an effective therapy for many patients with widespread psoriasis. Ultraviolet B (UV-B), narrowband UV-B, and ultraviolet A (UV-A) light with either oral or topical psoralens (PUVA) are used clinically. UV light's immunosuppressive properties are thought to be responsible for its therapeutic activity in psoriasis. It is also mutagenic, potentially leading to an increased incidence of nonmelanoma and melanoma skin cancer. UV-light therapy is contraindicated in patients receiving cyclosporine and should be used with great care in all immunocompromised patients due to the increased risk of skin cancer.

Various systemic agents can be used for severe, widespread psoriatic disease **(Table 60-3)**. Oral glucocorticoids should not be used for the treatment of psoriasis due to the potential for development of life-threatening pustular psoriasis when therapy is discontinued. Methotrexate is an effective agent, especially in patients with PsA. The synthetic retinoid acitretin is useful, especially when immunosuppression must be avoided; however, teratogenicity limits its use. Apremilast inhibits phosphodiesterase type 4. It is approved for both psoriasis and PsA. It must be used cautiously in the presence of renal failure or depression. Deucravacitinib is an inhibitor of tyrosine kinase 2 that results in inhibition of IL-23, IL-12, and type 1 interferon pathways. It is approved to treat plaque psoriasis.

The evidence implicating psoriasis as a T-cell–mediated disorder has directed therapeutic efforts to immunoregulation. Cyclosporine and other immunosuppressive agents can be very effective in the treatment of psoriasis, and much attention is currently directed toward the development of biologic agents with more selective immunosuppressive properties and better safety profiles **(Table 60-4)**. These biologic agents appear to be quite efficacious in the treatment of psoriasis and are well tolerated; however, caution with certain patient comorbidities must be exercised. Use of tumor necrosis factor-α (TNF-α) inhibitors may worsen congestive heart failure (CHF), and they should be used with caution in patients at risk for or known to have CHF. Further, none of the immunosuppressive agents used in the treatment of psoriasis should be initiated if the patient has a severe infection (including tuberculosis, HIV, hepatitis B or C); patients on such therapy should be routinely screened for tuberculosis. There have been reports of progressive multifocal leukoencephalopathy and lupus

TABLE 60-4 FDA-Approved Biologics for Psoriasis or Psoriatic Arthritis

MECHANISM OF ACTION	AGENTS (INDICATION; ROUTE)	FREQUENCY	WARNINGS, SELECTED
Anti-TNF-α	Etanercept (Ps, PsA; SC) Adalimumab (Ps, PsA; SC) Certolizumab (Ps, PsA; SC) Infliximab (Ps, PsA; IV) Golimumab (PsA; SC)	Ranges from once or twice weekly[a] to every 8 weeks[a]	Serious infections, hepatotoxicity, CHF, hematologic events, hypersensitivity reactions, neurologic events, potential for increased malignancies
Anti-IL-12 and anti-IL-23	Ustekinumab (Ps, PsA; SC)	Every 12 weeks[a]	Serious infections, neurologic events, potential for increased malignancies
Anti-IL-23	Risankizumab (Ps; SC) Tildrakizumab (Ps; SC) Guselkumab (Ps; SC)	Ranges from every 8–12 weeks[a]	Serious infections, headaches
Anti-IL-17	Secukinumab (Ps, PsA; SC) Ixekizumab (Ps; SC) Brodalumab (Ps; SC)	Ranges from every 2–4 weeks[a]	Serious infections, hypersensitivity reaction, inflammatory bowel disease

[a] Initial dose modifications required.

Abbreviations: CHF, congestive heart failure; IL, interleukin; IV, intravenous; Ps, psoriasis; PsA, psoriatic arthritis; SC, subcutaneous; TNF-α, tumor necrosis factor-α.

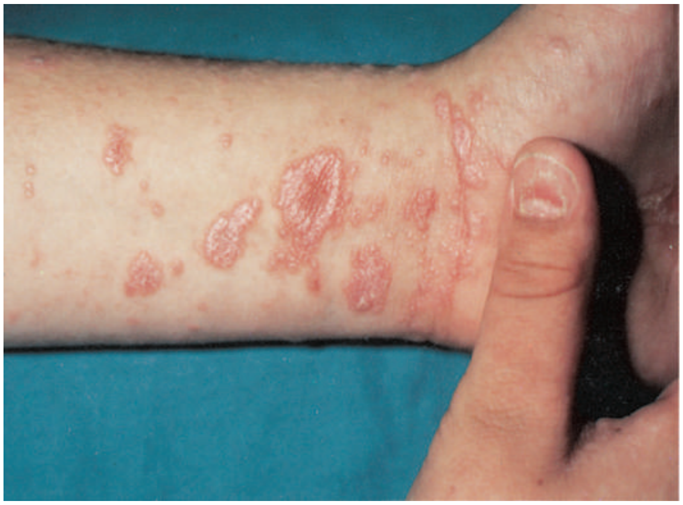

FIGURE 60-5 Lichen planus. An example of lichen planus showing multiple flat-topped, violaceous papules and plaques. Nail dystrophy, as seen in this patient's thumbnail, may also be a feature. *(Courtesy of Robert Swerlick, MD; with permission.)*

erythematosus in association with treatment with the TNF-α inhibitors. Malignancies, including a risk or history of certain cancers, may limit the use of these systemic agents. In general, immunosuppressive agents have also been linked to an increased risk of skin cancer, and patients receiving these agents should be monitored for the development of skin cancer.

LICHEN PLANUS

Lichen planus (LP) is a papulosquamous disorder that may affect the skin, scalp, nails, and mucous membranes. The primary cutaneous lesions are pruritic, polygonal, flat-topped, violaceous papules. Close examination of the surface of these papules often reveals a network of gray lines (*Wickham's striae*). The skin lesions may occur anywhere but have a predilection for the wrists, shins, lower back, and genitalia (Fig. 60-5). Involvement of the scalp (*lichen planopilaris*) may lead to scarring alopecia, and nail involvement may lead to permanent deformity or loss of fingernails and toenails. LP commonly involves mucous membranes, particularly the buccal mucosa, where it can present on a spectrum ranging from a mild, white, reticulate eruption of the mucosa to a severe, erosive stomatitis. Erosive stomatitis may persist for years and may be linked to an increased risk of oral squamous cell carcinoma. Cutaneous eruptions clinically resembling LP have been observed after administration of numerous drugs, including thiazide diuretics, gold, antimalarial agents, penicillamine, and phenothiazines, and in patients with skin lesions of chronic graft-versus-host disease. In addition, LP may be associated with hepatitis C infection. The course of LP is variable, but most patients have spontaneous remissions 6 months to 2 years after the onset of disease. Topical glucocorticoids are the mainstay of therapy.

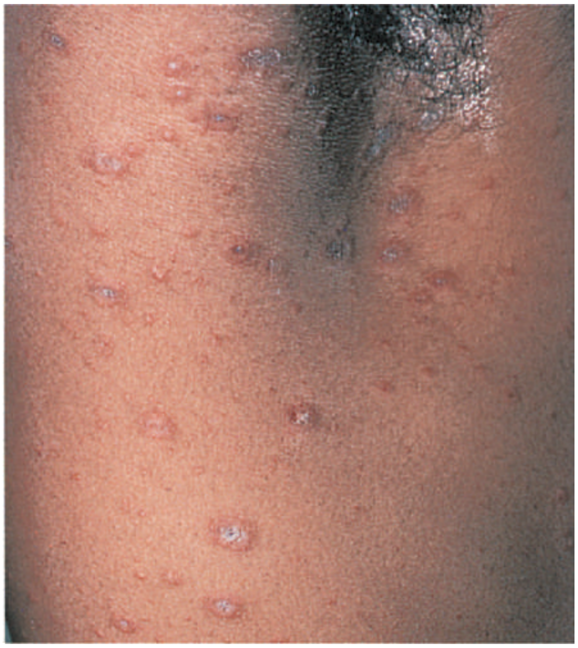

FIGURE 60-6 Pityriasis rosea. Multiple, round to oval, pink papules with fine central scale and hyperpigmentation are distributed along skin tension lines on the trunk. *(Courtesy of Justin T. Cheeley, MD.)*

PITYRIASIS ROSEA

Pityriasis rosea (PR) is a papulosquamous eruption of unknown etiology occurring more commonly in the spring and fall. Its first manifestation is the development of a 2- to 6-cm annular lesion (the herald patch). This is followed in a few days to a few weeks by the appearance of many smaller annular or papular lesions with a predilection to occur on the trunk (Fig. 60-6). The lesions are generally oval, with their long axis parallel to the skinfold lines. Individual lesions may range in color from red to brown and have a trailing scale. PR shares many clinical features with the eruption of secondary syphilis, but palm and sole lesions are extremely rare in PR and common in secondary syphilis. The eruption tends to be moderately pruritic and lasts 3–8 weeks. Treatment is directed at alleviating pruritus and consists of oral antihistamines; mid-potency topical glucocorticoids; and, in some cases, UV-B phototherapy.

CUTANEOUS INFECTIONS (TABLE 60-5)

IMPETIGO, ECTHYMA, AND FURUNCULOSIS

Impetigo is a common superficial bacterial infection of the skin caused most often by *S. aureus* (Chap. 152) and in some cases by group A β-hemolytic streptococci (Chap. 153). The primary lesion is a superficial pustule that ruptures and forms a characteristic yellow-brown honey-colored crust (see Chap. 153, Fig. 153-3). Lesions may occur

TABLE 60-5 Common Skin Infections

	CLINICAL FEATURES	ETIOLOGIC AGENT	TREATMENT
Impetigo	Honey-colored crusted papules, plaques, or bullae	Group A *Streptococcus* and *Staphylococcus aureus*	Systemic or topical antistaphylococcal and antistreptococcal antibiotics
Dermatophytosis	Inflammatory or noninflammatory annular scaly plaques; may involve hair loss; groin involvement spares scrotum; hyphae on KOH preparation	*Trichophyton*, *Epidermophyton*, or *Microsporum* spp.	Topical azoles, systemic griseofulvin, terbinafine, or azoles
Candidiasis	Inflammatory papules and plaques with satellite pustules, frequently in intertriginous areas; may involve scrotum; pseudohyphae on KOH preparation	*Candida albicans* and other *Candida* spp.	Topical nystatin or azoles; systemic azoles for resistant disease
Tinea versicolor	Hyper- or hypopigmented scaly patches on trunk; characteristic mixture of hyphae and spores ("spaghetti and meatballs") on KOH preparation	*Malassezia furfur*	Topical selenium sulfide lotion or azoles

Abbreviation: KOH, potassium hydroxide.

on normal skin (primary infection) or in areas already affected by another skin disease (secondary infection). Lesions caused by staphylococci may be tense, clear bullae, and this less common form of the disease is called *bullous impetigo*. Blisters are caused by the production of exfoliative toxin by *S. aureus* phage type II. This is the same toxin responsible for staphylococcal scalded-skin syndrome, often resulting in dramatic loss of the superficial epidermis due to blistering. The latter syndrome is much more common in children than in adults; however, it should be considered along with toxic epidermal necrolysis and severe drug eruptions in patients with widespread blistering of the skin. *Ecthyma* is a deep nonbullous variant of impetigo that causes punched-out ulcerative lesions. It is more often caused by a primary or secondary infection with *Streptococcus pyogenes*. Ecthyma is a deeper infection than typical impetigo and resolves with scars. Treatment of both ecthyma and impetigo involves gentle debridement of adherent crusts, facilitated by using soaks and topical antibiotics in conjunction with appropriate oral antibiotics.

Furunculosis is also caused by *S. aureus*, and this disorder has gained prominence in the past few decades because of CA-MRSA. A furuncle, or boil, is a painful, erythematous nodule that can occur on any cutaneous surface. The lesions may be solitary but are most often multiple. Patients frequently believe they have been bitten by spiders or insects. Family members or close contacts may also be affected. Furuncles can rupture and drain spontaneously or may need incision and drainage, which may be adequate therapy for small solitary furuncles without cellulitis or systemic symptoms. Whenever possible, lesional material should be sent for culture. Current recommendations for methicillin-sensitive infections are β-lactam antibiotics. Therapy for CA-MRSA is discussed previously (see "Atopic Dermatitis"). Warm compresses and nasal mupirocin are helpful therapeutic additions. Severe infections may require IV antibiotics.

■ ERYSIPELAS AND CELLULITIS

See Chap. 134.

■ DERMATOPHYTOSIS

Dermatophytes are fungi that infect skin, hair, and nails and include members of the genera *Trichophyton*, *Microsporum*, and *Epidermophyton* (Chap. 225). *Tinea corporis*, or infection of the relatively hairless skin of the body (glabrous skin), may have a variable appearance depending on the extent of the associated inflammatory reaction. Typical infections consist of erythematous, scaly plaques, with an annular appearance that accounts for the common name "ringworm." Deep inflammatory nodules or granulomas occur in some infections, most often those inappropriately treated with mid- to high-potency topical glucocorticoids. Involvement of the groin (*tinea cruris*) is more common in males than in females. It presents as a scaling, erythematous eruption sparing the scrotum. Infection of the foot (*tinea pedis*) is the most common dermatophyte infection and is often chronic; it is characterized by variable erythema, edema, scaling, pruritus, and occasionally vesiculation. The infection may be widespread or localized but generally involves the web space between the fourth and fifth toes. Infection of the nails (*tinea unguium* or *onychomycosis*) occurs in many patients with tinea pedis and is characterized by opacified, thickened nails and subungual debris. The distal-lateral variant is most common. Proximal subungual onychomycosis may be a marker for HIV infection or other immunocompromised states. Dermatophyte infection of the scalp (*tinea capitis*) continues to be common, particularly affecting inner-city children but also affecting immunocompromised adults. The predominant organism is *Trichophyton tonsurans*, which can produce a relatively noninflammatory infection with mild scale and hair loss that is diffuse or localized. *T. tonsurans* and *Microsporum canis* can also cause a markedly inflammatory dermatosis with edema and nodules. This latter presentation is a *kerion*.

The diagnosis of tinea can be made from skin scrapings, nail scrapings, or hair by culture or direct microscopic examination with KOH. Nail clippings may be sent for histologic examination with periodic acid–Schiff (PAS) stain.

TREATMENT

Dermatophytosis

Both topical and systemic therapies may be used in dermatophyte infections. Treatment depends on the site involved and the type of infection. Topical therapy is generally effective for uncomplicated tinea corporis, tinea cruris, and limited tinea pedis. Topical agents are not effective as monotherapy for tinea capitis or onychomycosis (see below), and nystatin is not active against dermatophytes. Topicals are generally applied twice daily, and treatment should continue for 1 week beyond clinical resolution of the infection. Tinea pedis often requires longer treatment courses and frequently relapses. Oral antifungal agents may be required for recalcitrant tinea pedis or tinea corporis.

For dermatophyte infections involving the hair and nails and for other infections unresponsive to topical therapy, oral antifungal agents are often used. Markedly inflammatory tinea capitis may result in scarring and hair loss, and a systemic antifungal agent plus systemic or topical glucocorticoids may be helpful in preventing these sequelae. Fungal etiology should be confirmed by direct microscopic examination or by culture before oral antifungal agents are prescribed for any infection. All oral agents may cause hepatotoxicity. They should not be used in women who are pregnant or breast-feeding.

Griseofulvin is approved in the United States for dermatophyte infections involving the skin, hair, or nails. Common side effects of griseofulvin include gastrointestinal distress, headache, and urticaria.

Two other oral antifungal agents, itraconazole and terbinafine, are sometimes prescribed "off-label" for superficial fungal infections. Oral itraconazole is approved for onychomycosis. Itraconazole has the potential for serious interactions with other drugs requiring the P450 enzyme system for metabolism. Itraconazole should not be administered to patients with evidence of ventricular dysfunction or patients with known CHF.

Terbinafine is also approved for onychomycosis, and the granule version is approved for treatment of tinea capitis. Terbinafine has fewer interactions with other drugs than itraconazole; however, caution should be used with patients who are on multiple medications. The risk/benefit ratio should be considered when an asymptomatic toenail infection is treated with systemic agents.

The FDA has limited the use of a third oral agent due to potential hepatotoxicity and published the following: "Nizoral [ketoconazole] oral tablets should not be a first-line treatment for any fungal infection." The topical form of ketoconazole is not affected by this action.

■ TINEA (PITYRIASIS) VERSICOLOR

Tinea versicolor is caused by a nondermatophytic, dimorphic fungus, *Malassezia furfur*, a normal inhabitant of the skin. The expression of infection is promoted by heat and humidity. The typical lesions consist of oval scaly macules, papules, and patches concentrated on the chest, shoulders, and back but only rarely on the face or distal extremities. On dark skin, the lesions often appear as hypopigmented areas, whereas on light skin, they are slightly erythematous or hyperpigmented. A KOH preparation from scaling lesions will demonstrate a confluence of short hyphae and round spores ("spaghetti and meatballs"). Lotions or shampoos containing sulfur, salicylic acid, or selenium sulfide are the treatments of choice and will clear the infection if used daily for 1–2 weeks and then weekly thereafter. These preparations are irritating if left on the skin for >10 min; thus, they should be washed off completely. Treatment with some oral antifungal agents is also effective, but they do not provide lasting results and are not FDA approved for this indication.

■ CANDIDIASIS

Candidiasis is a fungal infection caused by a related group of yeasts whose manifestations may be localized to the skin and mucous membranes or, rarely, may be systemic and life-threatening **(Chap. 222)**. The causative organism is usually *Candida albicans*. These organisms are normal saprophytic inhabitants of the gastrointestinal tract but may overgrow due to broad-spectrum antibiotic therapy, diabetes mellitus, or immunosuppression and cause disease. Candidiasis is a very common infection in HIV-infected individuals **(Chap. 208)**. The oral cavity is commonly involved. Lesions may occur on the tongue or buccal mucosa (*thrush*) and appear as white plaques. Fissured, macerated lesions at the corners of the mouth (*perlèche*) are often seen in individuals with poorly fitting dentures and may also be associated with candidal infection. In addition, candidal infections have an affinity for sites that are chronically wet and macerated, including the skin around nails (onycholysis and paronychia), and in intertriginous areas. Intertriginous lesions are characteristically edematous, erythematous, and scaly, with scattered "satellite pustules." In males, there is often involvement of the penis and scrotum as well as the inner aspect of the thighs. In contrast to dermatophyte infections, candidal infections are frequently painful and accompanied by a marked inflammatory response. Diagnosis of candidal infection is based on the clinical pattern and demonstration of yeast on KOH preparation or culture.

TREATMENT
Candidiasis

Treatment involves removal of any predisposing factors such as antibiotic therapy or chronic moisture and the use of appropriate topical or systemic antifungal agents. Effective topicals include nystatin or azoles (miconazole, clotrimazole, econazole, or ketoconazole). The associated inflammatory response accompanying candidal infection on glabrous skin can be treated with a mild glucocorticoid lotion or cream (2.5% hydrocortisone). Systemic therapy is usually reserved for immunosuppressed patients or individuals with chronic or recurrent disease who fail to respond to appropriate topical therapy. Oral fluconazole is most commonly prescribed for cutaneous candidiasis. Oral nystatin is effective only for candidiasis of the gastrointestinal tract.

WARTS

Warts are cutaneous neoplasms caused by papillomaviruses. More than 100 different human papillomaviruses (HPVs) have been described. A typical wart, *verruca vulgaris*, is sessile, dome-shaped, and usually about a centimeter in diameter. Its surface is hyperkeratotic, consisting of many small filamentous projections. HPV also causes typical plantar warts, flat warts (*verruca plana*), and filiform warts. Plantar warts are endophytic and are covered by thick keratin. Paring of the wart will generally reveal a central core of keratinized debris and punctate bleeding points. Filiform warts are commonly seen on the face, neck, and skinfolds and present as papillomatous lesions on a narrow base. Flat warts are only slightly elevated and have a velvety, nonverrucous surface. They have a propensity for the face, arms, and legs, and are often spread by shaving.

Genital warts begin as small papillomas that may grow to form large, fungating lesions. In women, they may involve the labia, perineum, or perianal skin. In addition, the mucosa of the vagina, urethra, and anus can be involved as well as the cervical epithelium. In men, the lesions often occur initially in the coronal sulcus but may be seen on the shaft of the penis, the scrotum, or the perianal skin or in the urethra.

Appreciable evidence has accumulated indicating that HPV plays a role in the development of neoplasia of the uterine cervix and anogenital skin **(Chap. 94)**. HPV types 16 and 18 have been most intensely studied and are the major risk factors for intraepithelial neoplasia and squamous cell carcinoma of the cervix, anus, vulva, and penis. The risk is higher among patients immunosuppressed after solid organ transplantation and among those infected with HIV. Recent evidence also implicates other HPV types. Histologic examination of biopsied samples from affected sites may reveal changes associated with typical warts and/or features typical of intraepidermal carcinoma (Bowen's disease). Squamous cell carcinomas associated with HPV infections have also been observed in extragenital skin **(Chap. 81)**, most commonly in patients immunosuppressed after organ transplantation. Patients on long-term immunosuppression should be monitored for the development of squamous cell carcinoma and other cutaneous malignancies.

TREATMENT
Warts

Treatment of warts, other than anogenital warts, should be tempered by the observation that most warts in normal individuals resolve spontaneously within 1–2 years. There are many modalities available to treat warts, but no single therapy is universally effective. Factors that influence the choice of therapy include the location of the wart, the extent of disease, the age and immunologic status of the patient, and the patient's desire for therapy. Perhaps the most useful and convenient method for treating warts in almost any location is cryotherapy with liquid nitrogen. Equally effective for nongenital warts, but requiring much more patient compliance, is the use of keratolytic agents such as salicylic acid plasters or solutions. For genital warts, in-office application of a podophyllin solution is moderately effective but may be associated with marked local reactions. Prescription preparations of dilute, purified podophyllin are available for home use. Topical imiquimod, a potent inducer of local cytokine release, has been approved for treatment of genital warts. A topical compound composed of green tea extracts (sinecatechins) is also available. Conventional and laser surgical procedures may be required for recalcitrant warts. Recurrence of warts appears to be common with all these modalities. A highly effective vaccine for selected types of HPV has been approved by the FDA, and its use reduces the incidence of anogenital and cervical carcinoma.

HERPES SIMPLEX

See Chap. 197.

HERPES ZOSTER

See Chap. 198.

ACNE

ACNE VULGARIS

Acne vulgaris is a self-limited disorder primarily of teenagers and young adults, although perhaps 10–20% of adults may continue to experience some form of the disorder. The permissive factor for the expression of the disease in adolescence is the increase in sebum production by sebaceous glands with puberty. Small cysts, called *comedones*, form in hair follicles due to blockage of the follicular orifice by retention of keratinous material and sebum. The activity of bacteria (*Cutibacterium acnes*) within the comedones releases free fatty acids from sebum, causes inflammation within the cyst, and results in rupture of the cyst wall. An inflammatory foreign-body reaction develops as result of extrusion of oily and keratinous debris from the cyst.

The clinical hallmark of acne vulgaris is the comedone, which may be closed (*whitehead*) or open (*blackhead*). Closed comedones appear as 1- to 2-mm pebbly white papules, which are accentuated when the skin is stretched. They are the precursors of inflammatory lesions of acne vulgaris. The contents of closed comedones are not easily expressed. Open comedones, which rarely result in inflammatory acne lesions, have a dilated follicular orifice and are filled with easily expressible oxidized, darkened, oily debris. Comedones are usually accompanied by inflammatory lesions: papules, pustules, or nodules.

The earliest lesions seen in adolescence are generally mildly inflamed or noninflammatory comedones on the forehead. Subsequently, more typical inflammatory lesions develop on the cheeks, nose, and chin **(Fig. 60-7)**. The most common location for acne is the face, but involvement of the chest and back is common. Most disease remains mild and

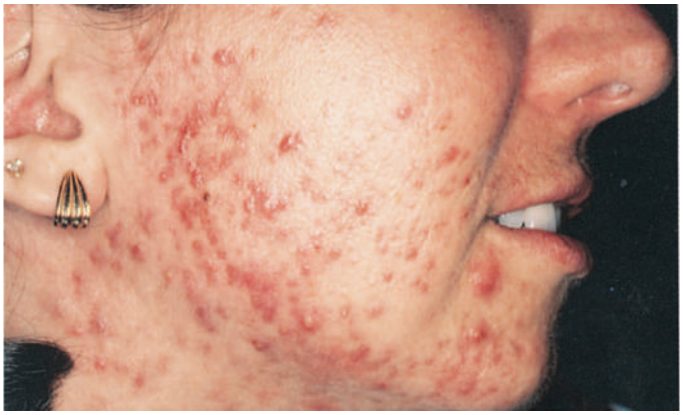

FIGURE 60-7 Acne vulgaris. An example of acne vulgaris with inflammatory papules, pustules, and comedones. *(Courtesy of Kalman Watsky, MD; with permission.)*

does not lead to scarring. A small number of patients develop large inflammatory cysts and nodules, which may drain and result in significant scarring. Regardless of the severity, acne may affect a patient's quality of life. With adequate treatment, this effect may be transient. In the case of severe, scarring acne, the effects can be permanent and profound. Early therapeutic intervention in severe acne is essential.

Exogenous and endogenous factors can alter the expression of acne vulgaris. Friction and trauma (from headbands or chin straps of athletic helmets), application of comedogenic topical agents (cosmetics or hair preparations), or chronic topical exposure to certain industrial compounds may elicit or aggravate acne. Glucocorticoids, topical or systemic, may also elicit acne. Other systemic medications such as progestin-only contraception, lithium, isoniazid, androgenic steroids, halogens, phenytoin, and phenobarbital may produce acneiform eruptions or aggravate preexisting acne. Genetic factors and polycystic ovary disease may also play a role.

TREATMENT

Acne Vulgaris

Treatment of acne vulgaris is directed toward elimination of comedones by normalizing follicular keratinization and decreasing sebaceous gland activity, the population of *C. acnes*, and inflammation. Minimal to moderate pauci-inflammatory disease may respond adequately to local therapy alone. Although areas affected with acne should be kept clean, overly vigorous scrubbing may aggravate acne due to mechanical rupture of comedones. Topical agents such as retinoic acid, benzoyl peroxide, or salicylic acid may alter the pattern of epidermal desquamation, preventing the formation of comedones and aiding in the resolution of preexisting cysts. Topical antibacterial agents (such as benzoyl peroxide, azelaic acid, erythromycin, clindamycin, or dapsone) are also useful adjuncts to therapy. Topical antibiotics (erythromycin and clindamycin) should be used in combination with benzoyl peroxide to prevent development of bacterial resistance. Topical antiandrogen receptor agents such as clascoterone cream are also FDA approved to treat acne.

Patients with moderate to severe acne with a prominent inflammatory component will benefit from the addition of systemic therapy, such as minocycline or doxycycline in doses of 100 mg bid or in lower dose, extended-release preparations. Such antibiotics appear to have anti-inflammatory effects independent of their antibacterial effects. Adequate response is expected at 3 months of therapy. Female patients may benefit from hormonal therapy. Several oral contraceptives are approved by the FDA for use in the treatment of acne vulgaris. Spironolactone is a safe, effective, and durable antiandrogen treatment in women.

Patients with severe nodulocystic acne unresponsive to the therapies discussed above may benefit from treatment with the synthetic retinoid isotretinoin. Dosing is weight-based and cumulative, with duration of therapy dictated by summative dose or acne lesion remission. Results are excellent in appropriately selected patients. Its use is highly regulated due to its potential for teratogenicity. Many patients receiving this medication develop dry skin and cheilitis, but severe extracutaneous and laboratory derangements are rare. At present, prescribers must enroll in a program designed to prevent pregnancy and adverse events while patients are taking isotretinoin. These measures are imposed to ensure that all prescribers are familiar with the risks of isotretinoin, that all female patients have two negative pregnancy tests prior to initiation of therapy and a negative pregnancy test prior to each refill, and that all patients have been warned about the risks associated with isotretinoin.

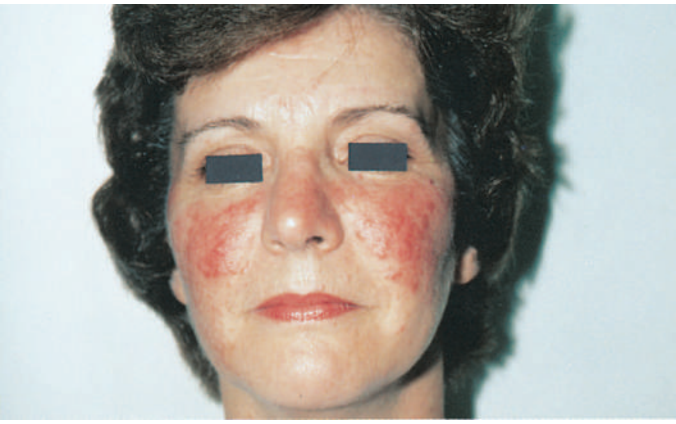

FIGURE 60-8 Acne rosacea. Prominent facial erythema, telangiectasia, scattered papules, and small pustules are seen in this patient with acne rosacea. *(Courtesy of Robert Swerlick, MD; with permission.)*

ACNE ROSACEA

Acne rosacea, commonly referred to simply as *rosacea*, is an inflammatory disorder predominantly affecting the central face. Persons most often affected are Caucasians of northern European background, but rosacea also occurs in patients with dark skin. Rosacea is seen almost exclusively in adults, only rarely affecting patients <30 years old. Rosacea is more common in women, but those most severely affected are men. It is characterized by the presence of erythema, telangiectasias, and superficial pustules (Fig. 60-8) but is not associated with the presence of comedones. Rosacea rarely involves the chest or back.

There is a relationship between the tendency for facial flushing and the subsequent development of acne rosacea. Often, individuals with rosacea initially demonstrate a pronounced flushing reaction. This may be in response to heat, emotional stimuli, alcohol, hot drinks, or spicy foods. As the disease progresses, the flush persists longer and longer and may eventually become permanent. Papules, pustules, and telangiectasias can become superimposed on the persistent flush. Rosacea of very long standing may lead to connective tissue overgrowth, particularly of the nose (*rhinophyma*). Rosacea may also be complicated by various inflammatory disorders of the eye, including keratitis, blepharitis, iritis, and recurrent chalazion. These ocular problems are potentially sight-threatening and warrant ophthalmologic evaluation.

TREATMENT

Acne Rosacea

Acne rosacea can be treated topically or systemically. Mild disease often responds to topical preparations of metronidazole, sodium sulfacetamide, azelaic acid, ivermectin, brimonidine, or oxymetazoline. More severe disease requires oral tetracyclines in subantimicrobial, modified-release preparations or low-dose isotretinoin. Residual telangiectasia may respond to laser therapy. Topical glucocorticoids, especially potent agents, should be avoided because chronic use of these preparations may elicit rosacea. Application of topical agents to the skin is not an effective treatment for ocular disease.

FURTHER READING

Bolognia JL et al (eds): *Dermatology*, 4th ed. Philadelphia, Elsevier, 2018.

James WD et al (eds): *Andrew's Diseases of the Skin Clinical Dermatology*, 13th ed. Philadelphia, Elsevier, 2020.

Kang S et al (eds): *Fitzpatrick's Dermatology*, 9th ed. New York, McGraw-Hill, 2019.

Wolff K et al (eds): *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, 8th ed. New York, McGraw-Hill, 2017.

61

Skin Manifestations of Internal Disease

Jean L. Bolognia, Jonathan S. Leventhal, Irwin M. Braverman

It is a generally accepted concept in medicine that the skin can develop signs of internal disease. Therefore, in textbooks of medicine, one finds a chapter describing in detail the major systemic disorders that can be identified by cutaneous signs. The underlying assumption of such a chapter is that the clinician has been able to identify the specific disorder in the patient and needs only to read about it in the textbook. In reality, concise differential diagnoses and the identification of these disorders are actually difficult for nondermatologists because they are not well-versed in the recognition of cutaneous lesions or their spectrum of presentations. Therefore, this chapter covers this particular topic of cutaneous medicine not by simply focusing on individual diseases, but by describing the various presenting clinical signs and symptoms that point to specific disorders. Concise differential diagnoses will be generated in which the significant diseases will be distinguished from the more common cutaneous disorders that have minimal or no significance with regard to associated internal disease. The latter disorders are reviewed in table form and always need to be excluded when considering the former. For a detailed description of individual diseases, the reader should consult a dermatologic text.

PAPULOSQUAMOUS SKIN LESIONS

(Table 61-1) When an eruption is characterized by elevated lesions, either papules (<1 cm) or plaques (>1 cm), in association with scale, it is referred to as *papulosquamous*. The most common papulosquamous diseases—*tinea*, *psoriasis*, *pityriasis rosea*, and *lichen planus*—are primary cutaneous disorders (Chap. 60). When psoriatic lesions are accompanied by arthritis, the possibility of psoriatic arthritis or reactive arthritis should be considered. A history of oral ulcers, conjunctivitis, uveitis, and/or urethritis points to the latter diagnosis. Lithium, beta blockers, anti-PD-1/PD-L1 antibodies, HIV or streptococcal infections, and a rapid taper of systemic glucocorticoids are known to exacerbate psoriasis; despite being used to treat psoriasis, tumor necrosis factor (TNF) inhibitors can paradoxically induce psoriatic lesions. Comorbidities in patients with psoriasis include cardiovascular disease and metabolic syndrome.

Whenever the clinical diagnosis of pityriasis rosea or lichen planus is made, it is important to review the patient's medications because the eruption may resolve by simply discontinuing the offending agent. Pityriasis rosea–like drug eruptions are seen most commonly with angiotensin-converting enzyme (ACE) inhibitors, beta blockers, and metronidazole, whereas the drugs that can produce a lichenoid eruption include thiazides, antimalarials, quinidine, beta blockers, TNF inhibitors, anti-PD-1/PD-L1 antibodies, and ACE inhibitors. In some populations (e.g., Europeans), there is a higher prevalence of hepatitis C viral infection in patients with oral lichen planus. Lichen planus–like lesions are also observed in chronic graft-versus-host disease.

TABLE 61-1 Selected Causes of Papulosquamous Skin Lesions

1. Primary cutaneous disorders
    a. Tinea[a]—widespread disease may be sign of immunosuppression
    b. Psoriasis[a]—widespread or resistant disease may be sign of HIV infection
    c. Pityriasis rosea[a]
    d. Lichen planus[a]
    e. Parapsoriasis, small plaque and large plaque[b]
    f. Bowen's disease (squamous cell carcinoma *in situ*)[c]
2. Drugs
3. Systemic diseases
    a. Lupus erythematosus, primarily subacute or chronic (discoid) lesions[d]
    b. Cutaneous T-cell lymphoma, in particular, mycosis fungoides[e]
    c. Secondary syphilis
    d. Reactive arthritis
    e. Sarcoidosis[f]—with scale less common than without scale
    f. Bazex syndrome (acrokeratosis paraneoplastica)[g]

[a]Discussed in detail in Chap. 60; cardiovascular disease and the metabolic syndrome are comorbidities in psoriasis; primarily in Europe, hepatitis C virus is associated with oral lichen planus. [b]Some authors view large plaque parapsoriasis as a form of early mycosis fungoides. [c]Associated with chronic sun exposure more often than exposure to arsenic; usually one or a few lesions. [d]See also Red Lesions in "Papulonodular Skin Lesions." [e]Also cutaneous lesions of HTLV-1-associated adult T-cell leukemia/lymphoma. [f]See also Red-Brown Lesions in "Papulonodular Skin Lesions." [g]Psoriasiform lesions of the helices, nose, and acral sites; squamous cell carcinoma of the upper aerodigestive tract most common underlying malignancy.

Abbreviation: HIV, human immunodeficiency virus.

In its early stages, the mycosis fungoides (MF) form of *cutaneous T-cell lymphoma* (CTCL) may be confused with eczema or psoriasis, but it often eventually fails to respond to appropriate therapy for those inflammatory diseases. MF can develop within lesions of large-plaque parapsoriasis and is suggested by an increase in the thickness of the lesions. The diagnosis of MF is established by skin biopsy in which collections of atypical T lymphocytes are found in the epidermis and dermis and is supported by *TCR* gene rearrangements. As the disease progresses, cutaneous tumors and lymph node involvement may appear.

In *secondary syphilis*, there are scattered pink to red-brown papules with thin scale. The eruption often involves the palms and soles and can resemble pityriasis rosea. Associated findings are helpful in making the diagnosis and include nonscarring alopecia, annular plaques on the face, mucous patches, condyloma lata (broad-based and moist), and lymphadenopathy, as well as malaise, fever, headache, and myalgias. The interval between the primary chancre and the secondary stage is usually 4–8 weeks, and spontaneous resolution without appropriate therapy occurs.

ERYTHRODERMA

(Table 61-2) *Erythroderma* is the term used when the majority of the skin surface is erythematous (red in color). Erythema may be more difficult to appreciate in darkly pigmented skin. There may be associated scale, erosions, or pustules as well as shedding of the hair and nails. Potential systemic manifestations include fever, chills, hypothermia, reactive lymphadenopathy, peripheral edema, hypoalbuminemia, and high-output cardiac failure. The major etiologies of erythroderma are (1) cutaneous diseases such as psoriasis and dermatitis (Table 61-3); (2) drugs; (3) systemic diseases, most commonly CTCL; and (4) idiopathic. In the first three groups, the location and description of the initial lesions, prior to the development of the erythroderma, aid in the diagnosis. For example, a history of red scaly plaques on the elbows and knees would point to psoriasis. It is also important to examine the skin carefully for a migration of the erythema and associated secondary changes such as pustules or erosions. Migratory waves of erythema studded with superficial pustules are seen in *pustular psoriasis*.

TABLE 61-2 Causes of Erythroderma

1. Primary cutaneous disorders
    a. Psoriasis[a]
    b. Dermatitis (atopic > contact >> stasis [with autosensitization] or seborrheic [primarily infants])[a]
    c. Pityriasis rubra pilaris
2. Drugs
3. Systemic diseases
    a. Cutaneous T-cell lymphoma (Sézary syndrome, erythrodermic mycosis fungoides)
    b. Other lymphomas
    c. Rarely, late-stage solid tumors, autoimmune bullous diseases
4. Idiopathic (usually older men)

[a] Discussed in detail in **Chap. 60**.

Drug-induced erythroderma may begin as an exanthematous (morbilliform) eruption **(Chap. 63)** or may arise as diffuse erythema. A number of drugs can produce an erythroderma, including penicillins, sulfonamides, aromatic anticonvulsants (e.g., carbamazepine, phenytoin), and allopurinol. Fever and peripheral eosinophilia often accompany the eruption, and there may also be facial swelling, hepatitis, allergic interstitial nephritis, myocarditis, and delayed thyroiditis; this constellation is frequently referred to as *drug reaction with eosinophilia and systemic symptoms* (DRESS) or *drug-induced hypersensitivity syndrome* (DIHS). In addition, these reactions, especially to aromatic anticonvulsants, can lead to a pseudolymphoma syndrome with adenopathy and circulating atypical lymphocytes, while reactions to allopurinol may be accompanied by gastrointestinal bleeding.

The most common malignancy that is associated with erythroderma is CTCL; in some series, up to 25% of the cases of erythroderma were

TABLE 61-3 Erythroderma (Primary Cutaneous Disorders)

	INITIAL LESIONS	LOCATION OF INITIAL LESIONS	OTHER FINDINGS	DIAGNOSTIC AIDS	TREATMENT
Psoriasis[a]	Pink-red, silvery scale, sharply demarcated	Elbows, knees, scalp, presacral area, intergluteal fold	Nail dystrophy (e.g., pits, oil drop sign), arthritis, pustules, SAPHO syndrome[b]	Skin biopsy	Topical glucocorticoids, vitamin D analogues, aryl hydrocarbon receptor agonist, PDE4 inhibitor; UV-B (narrowband) > PUVA; oral retinoids; MTX; anti-TNF agents, anti-IL-12/23 Ab, anti-IL-23 Ab, anti-IL-17A, anti-IL-17F, or anti-IL-17 receptor A Ab; TYK2 inhibitor; apremilast; cyclosporine; anti-IL-36 receptor Ab for generalized pustular psoriasis
Dermatitis[a]					
Atopic	Acute: Erythema, fine scale, crust, indistinct borders, excoriations Chronic: Lichenification (increased skin markings), excoriations	Antecubital and popliteal fossae, neck, hands, eyelids	Pruritus Personal and/or family history of atopy, including asthma, allergic rhinitis or conjunctivitis, and atopic dermatitis Exclude secondary infection with *Staphylococcus aureus* or HSV Exclude superimposed irritant or allergic contact dermatitis	Skin biopsy	Topical glucocorticoids, tacrolimus, pimecrolimus, crisaborole, JAK inhibitors, and antipruritics; oral antihistamines for sedation; open wet dressings; UV-B ± UV-A > PUVA; anti-IL-4R alpha Ab, anti-IL-13 Ab; JAK inhibitors, e.g. abrocitinib, upadacitinib; oral/IM glucocorticoids (short-term); MTX; mycophenolate mofetil; azathioprine; cyclosporine Topical or oral antibiotics
Contact	Local: Erythema, crusting, vesicles, and bullae	Depends on offending agent	Irritant—onset often within hours Allergic—delayed-type hypersensitivity; lag time of 48 h with rechallenge	Patch testing; repeat open application test	Remove irritant or allergen; topical glucocorticoids; oral antihistamines; oral/IM glucocorticoids (short-term)
	Systemic: Erythema, fine scale, crust	Generalized vs major intertriginous zones (especially groin)	Patient has history of allergic contact dermatitis to topical agent and then receives systemic medication that is structurally related, e.g., formaldehyde (skin), aspartame (oral)	Patch testing	Same as local
Seborrheic (rare in adults)	Pink-red to pink-orange, greasy scale	Scalp, nasolabial folds, eyebrows, intertriginous zones	Flares with stress, HIV infection Associated with Parkinson's disease	Skin biopsy	Topical glucocorticoids and imidazoles
Stasis (with autosensitization)	Erythema, crusting, excoriations	Lower extremities	Pruritus, lower extremity edema, varicosities, hemosiderin deposits, lipodermatosclerosis History of venous ulcers, thrombophlebitis, and/or cellulitis Exclude cellulitis Exclude superimposed contact dermatitis, e.g., topical neomycin	Skin biopsy	Topical glucocorticoids; open wet dressings; leg elevation; pressure stockings; pressure wraps if associated ulcers
Pityriasis rubra pilaris	Orange-red (salmon-colored), perifollicular papules	Scalp When generalized, characteristic "skip" areas of uninvolved skin	Wax-like palmoplantar keratoderma Exclude cutaneous T-cell lymphoma	Skin biopsy	Isotretinoin or acitretin; MTX; anti-IL-12/23 Ab, anti-IL-23 Ab, anti-TNF agents, anti-IL-17A, anti-IL-17F, or -IL-17 receptor A Ab

[a] Discussed in detail in **Chap. 60**. [b] SAPHO syndrome occurs more commonly in patients with palmoplantar pustulosis than in those with erythrodermic psoriasis.

Abbreviations: Ab, antibody; HSV, herpes simplex virus; IL, interleukin; IM, intramuscular; MTX, methotrexate; PUVA, psoralens plus ultraviolet A irradiation; SAPHO, *s*ynovitis, *a*cne, *p*ustulosis, *h*yperostosis, and *o*steitis (a subtype is chronic recurrent multifocal osteomyelitis); TNF, tumor necrosis factor; UV-A, ultraviolet A irradiation; UV-B, ultraviolet B irradiation.

TABLE 61-4 Causes of Alopecia

- I. Nonscarring alopecia
  - A. Primary cutaneous disorders
    1. Androgenetic alopecia (female pattern, male pattern)
    2. Telogen effluvium
    3. Alopecia areata
    4. Tinea capitis
    5. Traumatic alopecia[a]
    6. Psoriasiform alopecia, including tumor necrosis factor (TNF) inhibitor–induced
  - B. Drugs
    1. Telogen effluvium—see text for most common causes
    2. Anagen effluvium—chemotherapeutic agents (e.g., anthracyclines)
  - C. Systemic diseases
    1. Systemic lupus erythematosus
    2. Secondary syphilis
    3. Hypothyroidism
    4. Hyperthyroidism
    5. Hypopituitarism
    6. Deficiencies of protein, biotin, zinc, and perhaps iron
- II. Scarring alopecia
  - A. Primary cutaneous disorders
    1. Cutaneous lupus (chronic discoid lesions)[b]
    2. Lichen planus, including frontal fibrosing alopecia
    3. Central centrifugal cicatricial alopecia
    4. Folliculitis decalvans
    5. Dissecting cellulitis
    6. Linear morphea (linear scleroderma)[c]
  - B. Drugs
    1. Chemotherapeutic agents (e.g., taxanes, busulfan)
  - C. Systemic diseases
    1. Discoid lesions in the setting of systemic lupus erythematosus[b]
    2. Sarcoidosis
    3. Cutaneous metastases

[a] Most patients with trichotillomania or early stages of traction alopecia and some patients with pressure-induced alopecia. [b] While the majority of patients with discoid lesions have only cutaneous disease, these lesions do represent one of the criteria in the European League Against Rheumatism (EULAR)/American College of Rheumatology (ACR) [2019] and ACR [1982] classification schemes for systemic lupus erythematosus. [c] Can involve underlying muscles and osseous structures, and rarely in linear morphea of the frontal scalp (*en coup de sabre*), there is involvement of the meninges and brain.

due to CTCL. The patient may progress from isolated plaques and tumors, but more commonly, the erythroderma is present throughout the course of the disease (Sézary syndrome). In Sézary syndrome, there are circulating clonal atypical T lymphocytes, pruritus, and lymphadenopathy. In cases of erythroderma where there is no apparent cause (idiopathic), longitudinal evaluation is mandatory to monitor for the possible development of CTCL.

ALOPECIA

(Table 61-4) The two major forms of alopecia are scarring and nonscarring. *Scarring alopecia* is associated with fibrosis, inflammation, and loss of hair follicles. A smooth scalp with a decreased number of follicular openings is usually observed clinically, but in some patients, the changes are seen only in biopsy specimens from affected areas. In *nonscarring alopecia*, the hair shafts are absent or miniaturized, but the hair follicles are preserved, explaining the reversible nature of nonscarring alopecia.

The most common causes of nonscarring alopecia include *androgenetic alopecia*, *telogen effluvium*, *alopecia areata*, *tinea capitis*, and the early phase of *traumatic alopecia* **(Table 61-5)**. In women with androgenetic alopecia, an elevation in circulating levels of androgens may be seen as a result of ovarian or adrenal gland dysfunction or neoplasm. When there are signs of virilization, such as a deepened voice and/or enlarged clitoris, the possibility of an ovarian or adrenal gland tumor should be considered.

Exposure to various drugs can also cause diffuse hair loss, usually by inducing a telogen effluvium. An exception is the anagen effluvium observed with chemotherapeutic agents such as daunorubicin. Alopecia is a side effect of the following drugs: warfarin, heparin, propylthiouracil, carbimazole, isotretinoin, acitretin, lithium, beta blockers, interferons, colchicine, and amphetamines. Fortunately, spontaneous regrowth usually follows discontinuation of the offending agent.

Less commonly, nonscarring alopecia is associated with *lupus erythematosus* and *secondary syphilis.* In systemic lupus, there are two forms of alopecia—one is scarring secondary to discoid lesions (see below), and the other is nonscarring. The latter form coincides with flares of systemic disease and may involve the entire scalp or just the frontal scalp, with the appearance of multiple short hairs ("lupus hairs") as a sign of initial regrowth. Scattered, poorly circumscribed patches of alopecia with a "moth-eaten" appearance are a manifestation of the secondary stage of syphilis. Diffuse thinning of the hair is also associated with hypothyroidism and hyperthyroidism (Table 61-4).

Scarring alopecia is more frequently the result of a primary cutaneous disorder such as *lichen planus*, *chronic cutaneous (discoid) lupus*, *central centrifugal cicatricial alopecia*, *folliculitis decalvans*, or *linear scleroderma* (*morphea*) than it is a sign of systemic disease. Although the scarring lesions of *discoid lupus* can be seen in patients with systemic lupus, in the majority of patients, the disease process is limited to the skin. Less common causes of scarring alopecia include *sarcoidosis* (see "Papulonodular Skin Lesions," below), chemotherapeutic agents, and *cutaneous metastases*.

In the early phases of discoid lupus, lichen planus, and folliculitis decalvans, there are circumscribed areas of alopecia. Fibrosis and subsequent loss of hair follicles are observed primarily in the center of these alopecic patches, whereas the inflammatory process is most prominent at the periphery. The areas of active inflammation in discoid lupus are erythematous with scale, whereas the areas of previous inflammation are often hypopigmented with a rim of hyperpigmentation. In lichen planus, perifollicular macules at the periphery are usually violet-colored. A complete examination of the skin and oral mucosa combined with a biopsy and direct immunofluorescence microscopy of inflamed skin will aid in distinguishing these two entities. The peripheral active lesions in folliculitis decalvans are follicular pustules; these patients can develop a reactive arthritis.

FIGURATE SKIN LESIONS

(Table 61-6) In *figurate eruptions*, the lesions form rings and arcs that are usually erythematous but can be skin-colored to brown. Most commonly, they are due to primary cutaneous diseases such as *tinea*, *urticaria*, *granuloma annulare*, and *erythema annulare centrifugum* **(Chaps. 60 and 62)**. An underlying systemic illness is found in a second, less common group of migratory annular erythemas. It includes *erythema migrans*, *erythema gyratum repens*, *erythema marginatum*, and *necrolytic migratory erythema*.

In erythema gyratum repens, one sees numerous mobile concentric arcs and wavefronts that resemble the grain in wood. A search for an underlying malignancy is mandatory in a patient with this eruption. Erythema migrans is the cutaneous manifestation of Lyme disease, which is caused by the spirochete *Borrelia burgdorferi*. In the initial stage (3–30 days after the tick bite, which may go unnoticed), a single annular lesion is usually seen that can expand to ≥10 cm in diameter. Primary EM, as defined by the CDC, has a diameter of ≥5 cm. Within several days, up to half of the patients develop multiple smaller erythematous lesions at sites distant from the bite. Associated symptoms include fever, headache, photophobia, myalgias, arthralgias, and malar rash. Erythema marginatum is seen in patients with rheumatic fever, primarily on the trunk. Lesions are pink-red in color, flat to minimally elevated, and transient.

There are additional cutaneous diseases that present as annular eruptions but lack an obvious migratory component. Examples include *CTCL*, *subacute cutaneous lupus*, *secondary syphilis*, and *sarcoidosis* (see "Papulonodular Skin Lesions," below).

TABLE 61-5 Nonscarring Alopecia (Primary Cutaneous Disorders)

	CLINICAL CHARACTERISTICS	PATHOGENESIS	TREATMENT
Telogen effluvium	Diffuse shedding of normal hairs Follows major stress (high fever, severe infection) or change in hormone levels (postpartum) Reversible without treatment	Stress causes more of the asynchronous growth cycles of individual hairs to become synchronous; therefore, larger numbers of growing (anagen) hairs simultaneously enter the dying (telogen) phase	Observation; discontinue any drugs that have alopecia as a side effect; must exclude underlying metabolic causes, e.g., hypothyroidism, hyperthyroidism
Androgenetic alopecia (male pattern; female pattern)	Miniaturization of hairs along the midline of the scalp Recession of the anterior scalp line in men and some women	Increased sensitivity of affected hairs to the effects of androgens—most common Increased levels of circulating androgens (ovarian or adrenal source in women)—less common	If no evidence of hyperandrogenemia, then topical minoxidil; low-dose oral minoxidil; finasteride[a]; spironolactone (women); hair transplant
Alopecia areata	Well-circumscribed, circular areas of hair loss, 2–5 cm in diameter In extensive cases, coalescence of lesions and/or involvement of other hair-bearing surfaces of the body Pitting or sandpapered appearance of the nails	The germinative zones of the hair follicles are surrounded by T lymphocytes Occasional associated diseases: hyperthyroidism, hypothyroidism, vitiligo, Down syndrome	Intralesional glucocorticoids; topical anthralin or tazarotene; topical contact sensitizers; JAK inhibitors; pulse prednisone (e.g., 300 mg orally once a month for 4–6 doses)
Tinea capitis	Varies from scaling with minimal hair loss to discrete patches with "black dots" (sites of broken infected hairs) to boggy plaque with pustules (kerion)[b]	Invasion of hairs by dermatophytes, most commonly *Trichophyton tonsurans*	Oral griseofulvin or terbinafine plus 2.5% selenium sulfide or ketoconazole shampoo; examine family members
Traumatic alopecia[c]	Broken hairs, often of varying lengths Irregular outline in trichotillomania and traction alopecia Fringe sign in traction alopecia	Traction with curlers, rubber bands, tight braiding Exposure to heat or chemicals (e.g., hair straighteners) Mechanical pulling (trichotillomania)	Discontinuation of offending hair style or chemical treatments; diagnosis of trichotillomania may require observation of shaved hairs (for growth) or biopsy, possibly followed by psychotherapy

[a] To date, Food and Drug Administration–approved for men. [b] Scarring alopecia can occur at sites of kerions. [c] May also be scarring, especially late-stage traction alopecia.

ACNE

(Table 61-7) In addition to *acne vulgaris* and *acne rosacea*, the two major forms of acne **(Chap. 60)**, there are drugs and systemic diseases that can lead to acneiform eruptions.

Patients with the *carcinoid syndrome* have episodes of flushing of the head, neck, and sometimes the trunk. Resultant skin changes of the face, in particular telangiectasias, may mimic the clinical appearance of erythematotelangiectatic acne rosacea.

TABLE 61-6 Causes of Figurate Skin Lesions

- I. Primary cutaneous disorders
  - A. Tinea
  - B. Urticaria (primary in ≥90% of patients)
  - C. Granuloma annulare
  - D. Erythema annulare centrifugum
  - E. Psoriasis, annular pustular psoriasis
  - F. Reactive granulomatous dermatitis, which includes interstitial granulomatous drug reaction
- II. Systemic diseases
  - A. Migratory
    - 1. Erythema migrans (CDC case definition is ≥5 cm in diameter)
    - 2. Urticaria (≤10% of patients)
    - 3. Erythema gyratum repens
    - 4. Erythema marginatum
    - 5. Pustular psoriasis (generalized and annular forms)
    - 6. Necrolytic migratory erythema (glucagonoma syndrome)[a]
  - B. Nonmigratory (may slowly expand)
    - 1. Subacute cutaneous LE, LE tumidus
    - 2. Sarcoidosis
    - 3. Leprosy (borderline, tuberculoid)
    - 4. Secondary syphilis (especially the face)
    - 5. Cutaneous T-cell lymphoma (especially mycosis fungoides)
    - 6. Interstitial granulomatous dermatitis[b]
    - 7. Annular erythema of Sjögren's syndrome

[a] Migratory erythema with erosions; favors lower extremities and girdle area.
[b] Underlying diseases include rheumatoid arthritis, LE, granulomatosis with polyangiitis, and hematologic disorders, in particular myelodysplasia and chronic myelomonocytic leukemia.

Abbreviations: CDC, Centers for Disease Control and Prevention; LE, lupus erythematosus.

PUSTULAR LESIONS

Acneiform eruptions (see "Acne," above) and *folliculitis* represent the most common pustular dermatoses. An important consideration in the evaluation of follicular pustules is a determination of the associated pathogen, for example, normal flora (culture-negative), *Staphylococcus aureus*, *Pseudomonas aeruginosa* ("hot tub" folliculitis), *Malassezia*, dermatophytes (Majocchi's granuloma), and *Demodex* spp. Noninfectious forms of folliculitis include HIV- or immunosuppression-associated eosinophilic folliculitis and folliculitis secondary to drugs such as glucocorticoids, lithium, and epidermal growth factor receptor (EGFR) or MEK inhibitors. Administration of high-dose systemic glucocorticoids can result in a widespread eruption of follicular pustules on the trunk, characterized by lesions in the same stage of development. With regard to underlying systemic diseases, nonfollicular-based pustules are a characteristic component of pustular psoriasis (sterile) and can be seen in septic emboli of bacterial or fungal origin (see "Purpura," below). In patients with acute generalized exanthematous pustulosis (AGEP) due primarily to medications (e.g., cephalosporins), there are large

TABLE 61-7 Causes of Acneiform Eruptions

- I. Primary cutaneous disorders
  - A. Acne vulgaris
  - B. Acne rosacea
- II. Drugs, e.g., anabolic steroids, glucocorticoids, lithium, EGFR inhibitors, HER2 inhibitors, MEK inhibitors, iodides
- III. Systemic diseases
  - A. Increased androgen production
    - 1. Adrenal origin, e.g., Cushing's disease, 21-hydroxylase deficiency
    - 2. Ovarian origin, e.g., polycystic ovary syndrome, ovarian hyperthecosis
  - B. Cryptococcosis, disseminated
  - C. Dimorphic fungal infections
  - D. Behçet's disease

Abbreviations: EGFR, epidermal growth factor receptor; HER2, human epidermal growth factor receptor 2; MEK, MAP (mitogen-activated protein) kinase.

areas of erythema studded with multiple sterile pustules in addition to neutrophilia.

TELANGIECTASIAS

(Table 61-8) To distinguish the various types of telangiectasias, it is important to examine the shape and configuration of the dilated blood vessels. *Linear telangiectasias* are seen on the face of patients with *actinically damaged skin* and *acne rosacea,* and they are found on the legs of patients with *venous hypertension* and first appear on the legs in *generalized essential telangiectasia*. Patients with an unusual form of *mastocytosis* (telangiectasia macularis eruptiva perstans) and the *carcinoid syndrome* (see "Acne," above) also have linear telangiectasias. Lastly, linear telangiectasias are found in areas of cutaneous inflammation. For example, longstanding lesions of discoid lupus frequently have telangiectasias within them.

Poikiloderma is a term used to describe a patch of skin with: (1) reticulated hypo- and hyperpigmentation, (2) wrinkling secondary to epidermal atrophy, and (3) telangiectasias. Poikiloderma does not imply a single disease entity—although it is becoming less common, it is seen in skin damaged by *ionizing radiation* as well as in patients with autoimmune connective tissue diseases, primarily *dermatomyositis* (DM), and rare genodermatoses (e.g., Kindler syndrome).

In *systemic sclerosis (scleroderma)*, the dilated blood vessels have a unique configuration and are known as *mat telangiectasias*. The lesions are broad macules that usually measure 2–7 mm in diameter but occasionally are larger. Mats have a polygonal or oval shape, and their erythematous color may appear uniform, but, upon closer inspection, the erythema is the result of delicate telangiectasias. The most common locations for mat telangiectasias are the face, oral mucosa, and hands—peripheral sites that are prone to intermittent ischemia. The limited form of systemic sclerosis, also referred to as the CREST (*c*alcinosis cutis, *R*aynaud's phenomenon, *e*sophageal dysmotility, *s*clerodactyly, and *t*elangiectasia) variant **(Chap. 372)**, is associated with a chronic course and anticentromere antibodies. Mat telangiectasias are an important clue to the diagnosis of this variant as well as the diffuse form of systemic sclerosis because they may be the only cutaneous finding.

TABLE 61-8 Causes of Telangiectasias

- I. Primary cutaneous disorders
  - A. Linear/branching
    - 1. Acne rosacea (face)
    - 2. Actinically damaged skin (face, neck, V of chest)
    - 3. Venous hypertension (legs)
    - 4. Generalized essential telangiectasia
    - 5. Cutaneous collagenous vasculopathy
    - 6. Within basal cell carcinomas or cutaneous lymphoma
  - B. Poikiloderma
    - 1. Ionizing radiation[a]
  - C. Spider angioma
    - 1. Idiopathic
    - 2. Pregnancy
- II. Systemic diseases
  - A. Linear/branching
    - 1. Carcinoid (head, neck, upper trunk)
    - 2. Ataxia-telangiectasia (bulbar conjunctivae, head and neck)
    - 3. Mastocytosis (within lesions)
  - B. Poikiloderma
    - 1. Dermatomyositis, lupus erythematosus
    - 2. Mycosis fungoides, patch stage
    - 3. Genodermatoses, e.g., xeroderma pigmentosum, Kindler syndrome
  - C. Mat
    - 1. Systemic sclerosis (scleroderma)
  - D. Nailfold
    - 1. Lupus erythematosus
    - 2. Systemic sclerosis (scleroderma)
    - 3. Dermatomyositis
    - 4. Hereditary hemorrhagic telangiectasia
  - E. Papular
    - 1. Hereditary hemorrhagic telangiectasia
  - F. Spider angioma
    - 1. Cirrhosis[b]
    - 2. Trastuzumab emtansine (may also involve mucosae)

[a] Becoming less common. [b] Due to hyperestrogenic state.

Proximal nailfold telangiectasias are pathognomonic signs of the three major autoimmune connective tissue diseases: *lupus erythematosus*, *systemic sclerosis*, and *DM*. They are easily visualized by the naked eye and occur in at least two-thirds of these patients. In both DM and lupus, there is associated nailfold erythema, and in DM, the erythema is often accompanied by "ragged" cuticles and fingertip tenderness. Under 10× magnification or by dermoscopy, the blood vessels in the nailfolds of lupus patients are tortuous and resemble "glomeruli," whereas in systemic sclerosis and DM, there is a loss of capillary loops and those that remain are markedly dilated.

In *hereditary hemorrhagic telangiectasia* (Osler-Rendu-Weber disease), the lesions usually appear during adolescence (mucosal) and adulthood (cutaneous) and are most commonly seen on the mucous membranes (nasal, orolabial), face, and distal extremities, including under the nails. They represent arteriovenous (AV) malformations of the dermal microvasculature, are dark red in color, and are usually slightly elevated. When the skin is stretched over an individual lesion, an eccentric punctum with radiating legs is seen. Although the degree of systemic involvement varies in this autosomal dominant disease (due primarily to mutations in either the endoglin or activin receptor–like kinase gene), the major symptoms are recurrent epistaxis and gastrointestinal bleeding. The fact that these mucosal telangiectasias are actually AV communications helps to explain their tendency to bleed.

HYPOPIGMENTATION

(Table 61-9) Disorders of hypopigmentation are often classified as either diffuse or localized. The classic example of *diffuse hypopigmentation* is *oculocutaneous albinism* (OCA). The most common forms are due to mutations in the tyrosinase gene (type I) or the *OCA2* (previously *P*) gene (type II); patients with type IA OCA have a total lack of enzyme activity. At birth, different forms of OCA can appear similar—white hair, gray-blue eyes, and pink-white skin. However, the patients with no tyrosinase activity maintain this phenotype, whereas those with decreased activity will acquire some pigmentation of the eyes, hair, and skin as they age. The degree of pigment formation is also a function of racial background, and the pigmentary dilution is more readily apparent when patients are compared to their first-degree relatives. The ocular findings in OCA correlate with the degree of hypopigmentation and include decreased visual acuity, nystagmus, photophobia, strabismus, and a lack of normal binocular vision.

The differential diagnosis of *localized hypomelanosis* includes the following primary cutaneous disorders: *postinflammatory hypopigmentation*, *idiopathic guttate hypomelanosis*, *pityriasis* (*tinea*) *versicolor*, *vitiligo*, *chemical- or drug-induced leukoderma*, *nevus depigmentosus* (see below), *progressive macular hypomelanosis*, and *piebaldism* **(Table 61-10)**. In this group of diseases, the areas of involvement are macules or patches with a decrease or absence of pigmentation. Patients with vitiligo also have an increased incidence of several autoimmune disorders, including Hashimoto's thyroiditis, Graves' disease, pernicious anemia, Addison's disease, uveitis, alopecia areata, chronic mucocutaneous candidiasis, and the autoimmune polyendocrine syndromes (types I and II). Diseases of the thyroid gland are the most frequently associated disorders, occurring in up to 30% of patients with vitiligo. Circulating autoantibodies are often found, and the most common ones are antithyroglobulin, antimicrosomal, and antithyroid-stimulating hormone receptor antibodies.

There are four systemic diseases that should be considered in a patient with skin findings suggestive of vitiligo—*systemic sclerosis*,

TABLE 61-9 Causes of Hypopigmentation

I. Primary cutaneous disorders
  A. Diffuse
    1. Generalized vitiligo[a]
  B. Localized
    1. Postinflammatory
    2. Idiopathic guttate hypomelanosis
    3. Pityriasis (tinea) versicolor
    4. Vitiligo[a]
    5. Chemical- or drug-induced leukoderma, e.g., topical imiquimod, oral imatinib
    6. Nevus depigmentosus and Blaschko-linear hypopigmentation (pigmentary mosaicism)[b]
    7. Progressive macular hypomelanosis
    8. Piebaldism[a]
II. Systemic diseases
  A. Diffuse
    1. Oculocutaneous albinism[b]
    2. Hermansky-Pudlak syndrome[b,c]
    3. Chédiak-Higashi syndrome[b,d]
    4. Phenylketonuria
  B. Localized
    1. Systemic sclerosis (scleroderma)[e]
    2. Melanoma-associated vitiligo-like leukoderma, immunotherapy-induced or spontaneous[e]
    3. Sarcoidosis
    4. Cutaneous T-cell lymphoma (especially mycosis fungoides)
    5. Tuberculoid and indeterminate leprosy
    6. Onchocerciasis[e]
    7. Blaschko-linear hypopigmentation (pigmentary mosaicism)[b,f]
    8. Incontinentia pigmenti (stage IV)
    9. Tuberous sclerosis
    10. Waardenburg syndrome and Shah-Waardenburg syndrome
    11. Vogt-Koyanagi-Harada syndrome[e]

[a] Absence of melanocytes in areas of leukoderma; congenital in piebaldism. [b] Normal number of melanocytes. [c] Platelet storage defect and restrictive lung disease secondary to deposits of ceroid-like material or immunodeficiency; due to mutations in β or δ subunit of adaptor-related protein complex 3 as well as subunits of *b*iogenesis of *l*ysosome-related *o*rganelles *c*omplex (BLOC)-1, -2, and -3. [d] Giant lysosomal granules and recurrent infections. [e] Can resemble vitiligo due to acquired complete loss of pigment. [f] Minority of patients in a nonreferral setting have systemic abnormalities (musculoskeletal, central nervous system, ocular), previously referred to as hypomelanosis of Ito.

melanoma-associated leukoderma, *onchocerciasis*, and *Vogt-Koyanagi-Harada syndrome*. The vitiligo-like leukoderma seen in patients with systemic sclerosis has a clinical resemblance to idiopathic vitiligo that has begun to repigment as a result of treatment; that is, perifollicular macules of normal pigmentation are seen within areas of depigmentation. The basis of this leukoderma is unknown; there is no evidence of inflammation in areas of involvement, but it can resolve if the underlying connective tissue disease becomes inactive. In contrast to idiopathic vitiligo, melanoma-associated vitiligo-like leukoderma often begins on the upper trunk and extensor forearms, and its appearance, if spontaneous, should prompt a search for metastatic disease. It is also seen in patients undergoing immunotherapy for melanoma, including immune checkpoint-blocking antibodies (checkpoint inhibitors), with cytotoxic T lymphocytes presumably recognizing cell surface antigens common to melanoma cells and melanocytes, and is associated with a greater likelihood of a clinical response. A history of aseptic meningitis, nontraumatic uveitis, tinnitus, hearing loss, and/or dysacousia points to the diagnosis of the Vogt-Koyanagi-Harada syndrome. In these patients, the face and scalp are the most common locations of pigment loss.

There are two systemic disorders (neurocristopathies) that may have the cutaneous findings of piebaldism (Table 61-9). They are *Shah-Waardenburg syndrome* and *Waardenburg syndrome*. A possible explanation for both disorders is an abnormal embryonic migration or survival of two neural crest–derived elements, one of them being melanocytes and the other myenteric ganglion cells (leading to Hirschsprung disease in Shah-Waardenburg syndrome) or auditory nerve cells (Waardenburg syndrome). The latter syndrome is characterized by congenital sensorineural hearing loss, dystopia canthorum (lateral displacement of the inner canthi but normal interpupillary distance), heterochromic irises, and a broad nasal root, in addition to the piebaldism. The facial dysmorphism can be explained by the neural crest origin of the connective tissues of the head and neck. Patients with Waardenburg syndrome have been shown to have mutations in four genes, including *PAX-3* and *MITF*, three of which encode transcription factors, whereas patients with Hirschsprung disease plus white spotting have mutations in one of three genes—endothelin 3, endothelin B receptor, and the transcription factor *SOX10*.

In *tuberous sclerosis*, the earliest cutaneous sign is macular hypomelanosis, referred to as an ash leaf spot. These lesions are often present at birth and are usually multiple; however, detection may require Wood's lamp examination, especially in lightly pigmented individuals. The pigment within them is reduced, but not absent. The average size is 1–3 cm, and the common shapes are polygonal and lance-ovate. Examination of the patient for additional cutaneous signs such as multiple angiofibromas of the face (adenoma sebaceum), ungual and intraoral fibromas, fibrous cephalic plaques, and connective tissue nevi (shagreen patches) is recommended. It is important to remember that an ash leaf spot on the scalp will result in a circumscribed patch of lightly pigmented hair. Internal manifestations include seizures, intellectual impairment, central nervous system (CNS) and retinal hamartomas, pulmonary lymphangioleiomyomatosis (women), bilateral renal angiomyolipomas, and myocardial rhabdomyomas. The latter can be detected in up to 60% of children (<18 years) with tuberous sclerosis by echocardiography.

Nevus depigmentosus is a stable, well-circumscribed hypomelanosis that is present at birth. There is usually a single oval or rectangular lesion, but when there are multiple lesions, the possibility of tuberous sclerosis needs to be considered. In Blaschko-linear hypopigmentation, also referred to as linear nevoid hypopigmentation and pigmentary mosaicism, streaks and swirls of hypopigmentation are observed; these terms have replaced hypomelanosis of Ito. Up to one-third of patients in a tertiary care setting had associated abnormalities involving the musculoskeletal system (asymmetry), the CNS (seizures and intellectual disability), and the eyes (strabismus and hypertelorism). Genetic mosaicism has been detected in these patients (e.g. activating mutations in *MTOR*), lending support to the hypothesis that the cutaneous pattern is the result of the migration of two clones of primordial melanocytes, each with a different pigment potential.

Localized areas of decreased pigmentation are commonly seen as a result of cutaneous inflammation (Table 61-10) and have been observed in the skin overlying active lesions of sarcoidosis (see "Papulonodular Skin Lesions," below) as well as in CTCL. Cutaneous infections also present as disorders of hypopigmentation, and in *tuberculoid leprosy*, there are a few asymmetric patches of hypomelanosis that have associated anesthesia, anhidrosis, and alopecia. Biopsy specimens of the palpable border show dermal granulomas that contain rare, if any, *Mycobacterium leprae* organisms.

HYPERPIGMENTATION

(Table 61-11) Disorders of hyperpigmentation are also divided into two major groups—localized and diffuse. The localized forms are due to an epidermal alteration, a proliferation of melanocytes, or an increase in pigment production. Both acanthosis nigricans and seborrheic keratoses belong to the first group. *Acanthosis nigricans* can be a reflection of an internal malignancy, most commonly of the gastrointestinal tract, and it appears as velvety hyperpigmentation, primarily in flexural areas. However, in the majority of patients, acanthosis nigricans is associated with obesity and insulin resistance, although it may be a reflection of an endocrinopathy such as acromegaly, Cushing's syndrome, polycystic ovary syndrome, or insulin-resistant diabetes mellitus (type A, type B, and lipodystrophic forms). *Seborrheic*

TABLE 61-10 Hypopigmentation (Primary Cutaneous Disorders, Localized)

	CLINICAL CHARACTERISTICS	WOOD'S LAMP EXAMINATION (UV-A; PEAK = 365 NM)	SKIN BIOPSY SPECIMEN	PATHOGENESIS	TREATMENT
Postinflammatory hypopigmentation	Can develop within active lesions, as in subacute cutaneous lupus, or after the lesion fades, as in atopic dermatitis	Depends on particular disease Usually less enhancement than in vitiligo	Type of inflammatory infiltrate depends on specific disease	Block in transfer of melanin from melanocytes to keratinocytes could be secondary to edema or decrease in contact time Destruction of melanocytes if inflammatory cells attack basal layer of epidermis	Treat underlying inflammatory disease
Idiopathic guttate hypomelanosis	Common; acquired; usually 2–4 mm in diameter Shins and extensor forearms	Less enhancement than vitiligo	Abrupt decrease in epidermal melanin content	Possible somatic mutations as a reflection of aging or UV exposure	None
Pityriasis (tinea) versicolor[a]	Common disorder Upper trunk and neck (shawl-like distribution), body folds Young adults Macules have fine white scale when scratched	Golden fluorescence	Hyphal forms and budding yeast in stratum corneum	Invasion of stratum corneum by the yeast *Malassezia* Yeast is lipophilic and produces C_9 and C_{11} dicarboxylic acids, which in vitro inhibit tyrosinase	Selenium sulfide 2.5% shampoo; topical imidazoles; oral fluconazole
Vitiligo	Acquired; progressive Symmetric areas of complete pigment loss Periorificial—around mouth, nose, eyes, nipples, umbilicus, anus Other areas—flexor wrists, extensor distal extremities Segmental form is less common—unilateral, dermatomal-like	More apparent Chalk-white	Absence of melanocytes in well-developed lesions Mild inflammation	Autoimmune phenomenon that results in destruction of melanocytes—primarily cellular (circulating skin-homing autoreactive T cells)	Topical glucocorticoids, calcineurin inhibitors, JAK inhibitors; UV-B (narrowband) > PUVA; oral JAK inhibitors; transplants, if stable; depigmentation (topical MBEH), if widespread and treatment-resistant
Chemical- or drug-induced leukoderma	Similar appearance to vitiligo Often begins on hands when associated with chemical exposure Satellite lesions in areas not exposed to chemicals	More apparent Chalk-white	Decreased number or absence of melanocytes	Exposure to chemicals that selectively destroy melanocytes, in particular phenols and catechols (germicides; rubber products) or ingestion of drugs such as imatinib Release of cellular antigens and activation of circulating lymphocytes may explain satellite phenomenon Possible inhibition of KIT receptor	Avoid exposure to offending agent, then treat as vitiligo Drug-induced variant may undergo repigmentation when medication is discontinued
Piebaldism	Autosomal dominant Congenital, stable White forelock Areas of amelanosis contain normally pigmented and hyperpigmented macules of various sizes Symmetric involvement of central forehead, ventral trunk, and mid regions of upper and lower extremities	Enhancement of leukoderma and hyperpigmented macules	Amelanotic areas—few to no melanocytes	Defect in migration of melanoblasts from neural crest to involved skin or failure of melanoblasts to survive or differentiate in these areas Mutations within the *KIT* protooncogene that encodes the tyrosine kinase receptor for stem cell growth factor (kit ligand)	None; occasionally transplants

[a] If potassium hydroxide (KOH) examination of scale is negative, consider the possibility of progressive macular hypomelanosis.

Abbreviations: MBEH, monobenzylether of hydroquinone; PUVA, psoralens plus ultraviolet A irradiation; UV-B, ultraviolet B irradiation.

keratoses are common lesions, but in one rare clinical setting, they are a sign of systemic disease, and that setting is the sudden appearance of multiple lesions, often with an inflammatory base and in association with acrochordons (skin tags) and acanthosis nigricans. This is termed the *sign of Leser-Trélat* and alerts the clinician to search for an internal malignancy.

A proliferation of melanocytes results in the following pigmented lesions: *lentigo*, *melanocytic nevus*, and *melanoma* **(Chap. 81)**. In an adult, the majority of lentigines are related to sun exposure, which explains their distribution. However, in the Peutz-Jeghers and LEOPARD (*l*entigines; *E*CG abnormalities, primarily conduction defects; *o*cular hypertelorism; *p*ulmonary stenosis and subaortic valvular stenosis; *a*bnormal genitalia [cryptorchidism, hypospadias]; *r*etardation of growth; and *d*eafness [sensorineural]) syndromes, lentigines do serve as a clue to systemic disease. In *LEOPARD/Noonan with multiple lentigines syndrome*, hundreds of lentigines develop during childhood and are scattered over the entire surface of the body. The lentigines in patients with *Peutz-Jeghers syndrome* (PJS) are located primarily around the nose and mouth, on the hands and feet, and within the oral cavity. While the pigmented macules on the face may

TABLE 61-11 Causes of Hyperpigmentation

- I. Primary cutaneous disorders
  - A. Localized
    - 1. Epidermal alteration
      - a. Seborrheic keratosis
      - b. Pigmented actinic keratosis
    - 2. Proliferation of melanocytes
      - a. Lentigo
      - b. Melanocytic nevus (mole)
      - c. Melanoma
    - 3. Increased pigment production
      - a. Ephelide (freckle)
      - b. Café au lait macule
      - c. Postinflammatory hyperpigmentation (also dermal)
      - d. Melasma (also dermal)
    - 4. Dermal pigmentation
      - a. Fixed drug eruption
  - B. Localized and diffuse
    - 1. Drugs (e.g., minocycline, hydroxychloroquine, bleomycin)
- II. Systemic diseases
  - A. Localized
    - 1. Epidermal alteration
      - a. Acanthosis nigricans (insulin resistance > other endocrine disorders, paraneoplastic)
      - b. Seborrheic keratoses (sign of Leser-Trélat)
    - 2. Proliferation of melanocytes
      - a. Lentigines (Peutz-Jeghers and LEOPARD/Noonan with multiple lentigines syndromes; xeroderma pigmentosum)
      - b. Melanocytic nevi (Carney complex [LAMB and NAME syndromes])[a]
    - 3. Increased pigment production
      - a. Café au lait macules (neurofibromatosis, Legius syndrome, McCune-Albright syndrome[b])
      - b. Urticaria pigmentosa[c]
    - 4. Dermal pigmentation
      - a. Incontinentia pigmenti (stage III)
      - b. Dyskeratosis congenita
    - 5. Dermal deposits
      - a. Exogenous ochronosis
      - b. Localized argyria
  - B. Diffuse
    - 1. Endocrinopathies
      - a. Addison's disease
      - b. Nelson's syndrome
      - c. Ectopic ACTH syndrome
      - d. Hyperthyroidism
    - 2. Metabolic
      - a. Porphyria cutanea tarda
      - b. Hemochromatosis
      - c. Vitamin B_{12}, folate deficiency
      - d. Pellagra
      - e. Malabsorption, including Whipple's disease
    - 3. Melanosis secondary to metastatic melanoma
    - 4. Autoimmune
      - a. Primary biliary cholangitis
      - b. Systemic sclerosis (scleroderma)
      - c. POEMS syndrome
      - d. Eosinophilia-myalgia syndrome[d]
    - 5. Drugs (e.g., cyclophosphamide) and metals (e.g., silver)

[a] Also lentigines. [b] Polyostotic fibrous dysplasia. [c] See also "Papulonodular Skin Lesions." [d] Late 1980s.

Abbreviations: ACTH, adrenocorticotropic hormone; LAMB, *l*entigines, *a*trial *m*yxomas, *m*ucocutaneous myxomas, and *b*lue nevi; LEOPARD, *l*entigines, *E*CG abnormalities, *o*cular hypertelorism, *p*ulmonary stenosis and subaortic valvular stenosis, *a*bnormal genitalia, *r*etardation of growth, and *d*eafness (sensorineural); NAME, *n*evi, *a*trial myxoma, *m*yxoid neurofibroma, and *e*phelides (freckles); POEMS, *p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathies, *M*-protein, and *s*kin changes.

fade with age, the oral lesions persist. However, similar intraoral lesions are also seen in Addison's disease, in Laugier-Hunziker syndrome (no internal manifestations), and as a normal finding in darkly pigmented individuals. Patients with PJS, an autosomal dominant syndrome due to mutations in a novel serine threonine kinase gene, have multiple benign polyps of the gastrointestinal tract, testicular or ovarian tumors, and an increased risk of developing gastrointestinal (primarily colon) and pancreatic cancers.

In the *Carney complex,* numerous lentigines are also seen, but they are in association with cardiac myxomas. This autosomal dominant disorder is also known as the *LAMB* (*l*entigines, *a*trial myxomas, *m*ucocutaneous myxomas, and *b*lue nevi) *syndrome* or *NAME* (*n*evi, *a*trial myxoma, *m*yxoid neurofibroma, and *e*phelides [freckles]) *syndrome*. These patients can also have evidence of endocrine overactivity in the form of Cushing's syndrome (primary pigmented nodular adrenocortical disease) and acromegaly.

The third type of localized hyperpigmentation is due to a local increase in pigment production, and it includes *ephelides* and *café au lait macules* (CALMs). While a single CALM can be seen in up to 10% of the normal population, the presence of multiple or large-sized CALMs raises the possibility of an associated genodermatosis, for example, neurofibromatosis (NF) or McCune-Albright syndrome. *CALMs* are flat, uniformly brown in color (usually two shades darker than uninvolved skin), and can vary in size from 0.5 to 12+ cm. More than 90% of adult patients with *type I NF* will have six or more CALMs measuring ≥1.5 cm in diameter. Additional findings are discussed in the section on neurofibromas (see "Papulonodular Skin Lesions," below). In comparison with NF, the CALMs in patients with *McCune-Albright syndrome* (polyostotic fibrous dysplasia with precocious puberty in females due to mosaicism for an activating mutation in a G protein [$G_s\alpha$] gene) are usually larger, are more irregular in outline, and tend to respect the midline.

In incontinentia pigmenti, dyskeratosis congenita, and bleomycin pigmentation, the areas of localized hyperpigmentation form a pattern—swirls and streaks in the first, reticulated in the second, and flagellate in the third. In *dyskeratosis congenita*, atrophic reticulated hyperpigmentation is seen on the neck, trunk, and thighs and is accompanied by nail dystrophy, pancytopenia, and leukoplakia of the oral and anal mucosae. The latter often develops into squamous cell carcinoma. In addition to the flagellate pigmentation (linear streaks) on the trunk, patients receiving bleomycin often have hyperpigmentation overlying the elbows, knees, and small joints of the hand.

Localized hyperpigmentation is seen as a side effect of several other *systemic medications*, including those that produce fixed drug reactions (nonsteroidal anti-inflammatory drugs [NSAIDs], sulfonamides, and tetracyclines) and those that can complex with melanin or iron (antimalarials and minocycline). Fixed drug eruptions recur in the exact same location as circular areas of erythema that can become bullous and then resolve as macular brown circles. The eruption usually appears within hours of readministration of the offending agent, and common locations include the genitalia, distal extremities, and perioral region. Chloroquine and hydroxychloroquine produce gray-brown to blue-black discoloration of the shins, hard palate, and face, while blue macules (often misdiagnosed as bruises) can be seen most commonly on the lower extremities and in sites of inflammation with prolonged minocycline administration. Estrogen in oral contraceptives can induce melasma—symmetric brown patches on the face, especially the cheeks, upper lip, and forehead. Similar changes are seen in pregnancy and in patients receiving phenytoin.

In the diffuse forms of hyperpigmentation, the darkening of the skin may be of equal intensity over the entire body or may be accentuated in sun-exposed areas. The causes of diffuse hyperpigmentation can be divided into four major groups—endocrine, metabolic, autoimmune, and drugs. The endocrinopathies that frequently have associated hyperpigmentation include *Addison's disease*, *Nelson's syndrome*, and *ectopic adrenocorticotropic hormone (ACTH) syndrome*. In these diseases, the increased pigmentation is diffuse but is accentuated in sun-exposed areas, as well as in the palmar creases, sites of friction, and scars. An overproduction of the pituitary hormones α-MSH (melanocyte-stimulating hormone) and ACTH can lead to an increase in melanocyte activity. These peptides are products of the proopiomelanocortin gene and exhibit homology; for example, α-MSH and ACTH share 13 amino acids. A minority of patients with Cushing's disease or hyperthyroidism have generalized hyperpigmentation.

The metabolic causes of hyperpigmentation include *porphyria cutanea tarda* (PCT), *hemochromatosis*, *vitamin* B_{12} *deficiency*, *folic acid deficiency*, *pellagra*, and *malabsorption*, including *Whipple's disease*. In patients with PCT (see "Vesicles/Bullae," below), the skin darkening is seen in sun-exposed areas and is a reflection of the photoreactive properties of porphyrins. The increased level of iron in the skin of patients with types 1 and 2 hemochromatosis stimulates melanin pigment production and leads to the classic bronze color. Patients with pellagra have a brown discoloration of the skin, especially in sun-exposed areas, as a result of nicotinic acid (niacin) deficiency. In the areas of increased pigmentation, there is a thin, varnish-like scale. These changes are also seen in patients who are vitamin B_6 deficient, have functioning carcinoid tumors (increased consumption of niacin), or take isoniazid. Approximately 50% of the patients with Whipple's disease have an associated generalized hyperpigmentation in association with diarrhea, weight loss, arthritis, and lymphadenopathy. A diffuse, slate-blue to gray-brown color is seen in patients with *melanosis secondary to metastatic melanoma*. The color reflects widespread deposition of melanin within the dermis as a result of the high concentration of circulating melanin precursors.

Of the autoimmune diseases associated with diffuse hyperpigmentation, *primary biliary cholangitis* and *systemic sclerosis* are the most common, and occasionally, both disorders are seen in the same patient. The skin is dark brown in color, especially in sun-exposed areas. In primary biliary cholangitis, the hyperpigmentation is accompanied by pruritus, jaundice, and xanthomas, whereas in systemic sclerosis, it is accompanied by sclerosis of the extremities, face, and, less commonly, the trunk. Additional clues to the diagnosis of systemic sclerosis are mat and cuticular telangiectasias, calcinosis cutis, Raynaud's phenomenon, and distal ulcerations (see "Telangiectasias," above). The differential diagnosis of cutaneous sclerosis with hyperpigmentation includes POEMS (*p*olyneuropathy; *o*rganomegaly [liver, spleen, lymph nodes]; *e*ndocrinopathies [impotence, gynecomastia]; *M*-protein; and *s*kin changes) syndrome. The skin changes include hyperpigmentation, induration, hypertrichosis, angiomas, clubbing, and facial lipoatrophy.

Diffuse hyperpigmentation that is due to drugs or metals can result from one of several mechanisms—induction of melanin pigment formation, complexing of the drug or its metabolites to melanin, and deposits of the drug in the dermis. Busulfan, cyclophosphamide, 5-fluorouracil, and inorganic arsenic induce pigment production. Complexes containing melanin or iron plus the drug or its metabolites are seen in patients receiving minocycline, and a diffuse, brown-gray, muddy appearance within sun-exposed areas may develop, in addition to pigmentation of the mucous membranes, teeth, nails, bones, and thyroid. Administration of amiodarone can result in both a phototoxic eruption (exaggerated sunburn) and/or a slate-gray to violaceous discoloration of sun-exposed skin. Biopsy specimens of the latter show yellow-brown granules in dermal macrophages, which represent intralysosomal accumulations of lipids, amiodarone, and its metabolites. Actual deposits of a particular drug or metal in the skin are seen with silver (argyria), where the skin appears blue-gray in color; gold (chrysiasis), where the skin has a brown to blue-gray color; and clofazimine, where the skin appears reddish brown. The associated pigmentation is accentuated in sun-exposed areas, and discoloration of the eye is seen with gold (sclerae) and clofazimine (conjunctivae).

VESICLES/BULLAE

(Table 61-12) Depending on their size, cutaneous blisters are referred to as *vesicles* (<1 cm) or *bullae* (>1 cm). The primary autoimmune blistering disorders include *pemphigus vulgaris*, *pemphigus foliaceus*, *paraneoplastic pemphigus*, *bullous pemphigoid*, *gestational pemphigoid*, *cicatricial pemphigoid*, *epidermolysis bullosa acquisita*, *linear IgA bullous dermatosis (LABD)*, and *dermatitis herpetiformis* **(Chap. 62)**.

TABLE 61-12 Causes of Vesicles/Bullae

I. Primary mucocutaneous diseases
    A. Primary blistering diseases (autoimmune)
        1. Pemphigus, foliaceus and vulgaris[a]
        2. Bullous pemphigoid[b]
        3. Gestational pemphigoid[b]
        4. Cicatricial pemphigoid[b]
        5. Dermatitis herpetiformis[b,c]
        6. Linear IgA bullous dermatosis[b]
        7. Epidermolysis bullosa acquisita[b,d]
    B. Secondary blistering diseases
        1. Contact dermatitis[a,b]
        2. Erythema multiforme[e]
        3. Stevens-Johnson syndrome[e]
        4. Toxic epidermal necrolysis[e]
        5. Bullous fixed drug eruption, including generalized variant[e]
        6. Pseudoporphyria, drug- or tanning booth–induced
    C. Infections
        1. Varicella-zoster virus[a,f]
        2. Herpes simplex virus[a,f]
        3. Enteroviruses, e.g., hand-foot-and-mouth disease[f]
        4. SARS-CoV-2
        5. Staphylococcal scalded-skin syndrome[a,g]
        6. Bullous impetigo[a]
        7. Bullous tinea
II. Systemic diseases
    A. Autoimmune
        1. Paraneoplastic pemphigus[a] (bronchiolitis obliterans)
        2. Bullous systemic lupus erythematosus
    B. Infections
        1. Cutaneous emboli[b]
    C. Metabolic
        1. Diabetic bullae[a,b]
        2. Porphyria cutanea tarda[b]
        3. Porphyria variegata[b]
        4. Bullous dermatosis of hemodialysis[b] (less often associated with peritoneal dialysis and also referred to as pseudoporphyria)
    D. Ischemia
        1. Coma bullae
    E. Secondary blistering diseases
        1. Toxic epidermal necrolysis[e] (respiratory and gastrointestinal tracts can be involved)
        2. Edema bullae (venous hypertension, congestive heart failure)

[a]Intraepidermal. [b]Subepidermal. [c]Associated with gluten enteropathy. [d]Associated with inflammatory bowel disease. [e]Degeneration of cells within the basal layer of the epidermis can give impression split is subepidermal. [f]Also systemic. [g]In adults, associated with renal failure and immunocompromised state.

Vesicles and bullae are also seen in *contact dermatitis*, both allergic and irritant forms **(Chap. 60)**. When there is a linear arrangement of vesicular lesions, an exogenous cause or herpes zoster should be suspected. Bullous disease secondary to the ingestion of drugs can take one of several forms, including phototoxic eruptions, isolated bullae, Stevens-Johnson syndrome (SJS), and toxic epidermal necrolysis (TEN) **(Chap. 63)**. Clinically, phototoxic eruptions resemble an exaggerated sunburn with diffuse erythema and bullae in sun-exposed areas. The most commonly associated drugs are doxycycline, quinolones, voriconazole, thiazides, NSAIDs, vemurafenib, and psoralens. The development of a phototoxic eruption is dependent on the doses of both the drug and ultraviolet (UV)-A irradiation.

Toxic epidermal necrolysis is characterized by bullae that arise on widespread areas of tender erythema and then slough. This results in large areas of denuded skin. The associated morbidity, such as sepsis, and mortality rates are relatively high and are a function of the extent of epidermal necrosis. In addition, these patients may also have involvement of the mucous membranes and respiratory and intestinal tracts. Drugs are the primary cause of TEN, and the most common offenders are aromatic anticonvulsants (phenytoin, barbiturates, carbamazepine), sulfonamides, aminopenicillins, allopurinol, and NSAIDs. Generalized bullous fixed drug eruption, severe acute graft-versus-host disease (grade 4), vancomycin-induced LABD, and flares of cutaneous lupus can also resemble TEN.

In *erythema multiforme* (EM), the primary lesions are pink-red macules and edematous papules, the centers of which may become vesicular. In contrast to a morbilliform exanthem, the clue to the diagnosis of EM, and especially SJS, is the development of a "dusky" violet color in the center of the lesions. Target lesions are also characteristic of EM and arise as a result of active centers and borders in combination with centrifugal spread. However, target lesions need not be present to make the diagnosis of EM.

EM has been subdivided into two major groups: (1) EM minor due to herpes simplex virus (HSV); and (2) EM major due to HSV, *Mycoplasma pneumonia*, or, occasionally, other viruses, *Chlamydia*, or drugs. Involvement of the mucous membranes (ocular, nasal, oral, and genital) is seen more commonly in the latter form, and in patients with *Mycoplasma pneumoniae*–induced rash and mucositis (MIRM)/reactive infectious mucocutaneous eruption (RIME), there may be minimal cutaneous involvement. Hemorrhagic crusts of the lips are characteristic of EM major and SJS as well as herpes simplex, pemphigus vulgaris, and paraneoplastic pemphigus. Fever, malaise, myalgias, sore throat, and cough may precede or accompany the eruption. The lesions of EM usually resolve over 2–4 weeks but may be recurrent, especially when due to HSV. In addition to HSV (in which lesions usually appear 7–12 days after the viral eruption), EM can also follow vaccinations, radiation therapy, and exposure to environmental toxins, including the oleoresin in poison ivy.

Induction of SJS is most often due to drugs, especially sulfonamides, aromatic anticonvulsants, lamotrigine, aminopenicillins, and nonnucleoside reverse transcriptase inhibitors (e.g., nevirapine). Widespread dusky macules and significant mucosal involvement are characteristic of SJS, and the cutaneous lesions may or may not develop epidermal detachment. If the latter occurs, by definition, it is limited to <10% of the body surface area (BSA). Greater involvement leads to the diagnosis of SJS/TEN overlap (10–30% BSA) or TEN (>30% BSA).

In addition to primary blistering disorders and hypersensitivity reactions, bacterial and viral infections can lead to vesicles and bullae. The most common infectious agents are HSV **(Chap. 197)**, varicella-zoster virus **(Chap. 198)**, and *S. aureus* **(Chap. 152)**.

Staphylococcal scalded-skin syndrome (SSSS) and *bullous impetigo* are two blistering disorders associated with staphylococcal (phage group II) infection. In SSSS, the initial findings are redness and tenderness of the central face, neck, trunk, and intertriginous zones. This is followed by short-lived flaccid bullae and a slough or exfoliation of the superficial epidermis. Crusted areas then develop, characteristically around the mouth in a radial pattern. SSSS is distinguished from TEN by the following features: younger age group (primarily infants and toddlers), more superficial site of blister formation, no oral lesions, shorter course, lower morbidity and mortality rates, and an association with staphylococcal exfoliative toxin ("exfoliatin"), not drugs. A rapid diagnosis of SSSS versus TEN can be made by a frozen section of the blister roof or exfoliative cytology of the blister contents. In SSSS, the site of staphylococcal infection is usually extracutaneous (conjunctivitis, rhinorrhea, otitis media, pharyngitis, tonsillitis), and the cutaneous lesions are sterile, whereas in bullous impetigo, the skin lesions are the site of infection. Impetigo is more localized than SSSS and usually presents with honey-colored crusts. Occasionally, superficial purulent blisters also form. *Cutaneous emboli* from gram-negative infections may present as isolated bullae, but the base of the lesion is purpuric or necrotic, and it may develop into an ulcer (see "Purpura," below).

Several metabolic disorders are associated with blister formation, including diabetes mellitus, renal failure, and porphyria. Local hypoxemia secondary to decreased cutaneous blood flow can also produce blisters, which explains the presence of bullae over pressure points in comatose patients (coma bullae). In *diabetes mellitus*, tense bullae with clear sterile viscous fluid arise on normal skin. The lesions can be as large as 6 cm in diameter and are located on the distal extremities. There are several types of porphyria, but the most common form with cutaneous findings is *porphyria cutanea tarda* (PCT). In sun-exposed areas (primarily the hands), the skin is very fragile, with trauma leading to erosions mixed with tense vesicles. These lesions then heal with scarring and formation of milia; the latter are firm, 1- to 2-mm white or yellow papules that represent epidermoid cysts. Associated findings can include hypertrichosis of the lateral malar region (men) or face (women) and, in sun-exposed areas, hyperpigmentation and firm sclerotic plaques. An elevated level of urinary uroporphyrins confirms the diagnosis and is due to a decrease in uroporphyrinogen decarboxylase activity. PCT can be exacerbated by alcohol, hemochromatosis and other forms of iron overload, chlorinated hydrocarbons, hepatitis C virus and HIV infections, and hepatomas.

The differential diagnosis of PCT includes (1) *porphyria variegata*—the skin signs of PCT plus the systemic findings of acute intermittent porphyria; it has a diagnostic plasma porphyrin fluorescence emission at 626 nm; (2) *drug-induced pseudoporphyria*—the clinical and histologic findings are similar to PCT, but porphyrins are normal; etiologic agents include naproxen and other NSAIDs, furosemide, tetracycline, and voriconazole; (3) *bullous dermatosis of hemodialysis*—the same appearance as PCT, but porphyrins are usually normal or occasionally borderline elevated; patients have chronic renal failure and are on hemodialysis; (4) *PCT associated with hepatomas and hemodialysis;* and (5) *epidermolysis bullosa acquisita* **(Chap. 62)**.

EXANTHEMS

(Table 61-13) Exanthems are characterized by an acute generalized eruption. The most common presentation is erythematous macules and papules (morbilliform) and less often confluent blanching erythema (scarlatiniform). *Morbilliform* eruptions are usually due to either drugs or viral infections. For example, up to 5% of patients receiving penicillins, sulfonamides, phenytoin, or nevirapine will develop a maculopapular eruption. Accompanying signs may include pruritus, fever, eosinophilia, transaminitis, and transient lymphadenopathy **(Chap. 63)**.

TABLE 61-13 Causes of Exanthems

I. Morbilliform
    A. Drugs
    B. Viral
        1. Rubeola (measles)
        2. Rubella
        3. Erythema infectiosum (erythema of cheeks; reticulated on extremities)
        4. Epstein-Barr virus, echovirus, coxsackievirus, CMV, adenovirus, HHV-6/HHV-7[a], SARS-CoV-2, dengue, chikungunya, and West Nile virus infections
        5. HIV seroconversion exanthem (plus mucosal ulcerations)
    C. Bacterial
        1. Typhoid fever
        2. Early secondary syphilis
        3. Early *Rickettsia* infections
        4. Early meningococcemia
        5. Ehrlichiosis
    D. Acute graft-versus-host disease
    E. Kawasaki disease
II. Scarlatiniform
    A. Scarlet fever
    B. Toxic shock syndrome
    C. Kawasaki disease
    D. Early staphylococcal scalded-skin syndrome

[a] Primary infection in infants and reactivation in the setting of immunosuppression.

Abbreviations: CMV, cytomegalovirus; HHV, human herpesvirus; HIV, human immunodeficiency virus; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Similar maculopapular eruptions are seen in the classic childhood viral exanthems, including (1) *rubeola* (measles)—a prodrome of coryza, cough, and conjunctivitis followed by Koplik's spots on the buccal mucosa; the eruption begins behind the ears, at the hairline, and on the forehead and then spreads down the body, often becoming confluent; (2) *rubella*—the eruption begins on the forehead and face and then spreads down the body; it resolves in the same order and is associated with retroauricular and suboccipital lymphadenopathy; and (3) *erythema infectiosum* (fifth disease)—erythema of the cheeks is followed by a reticulated pattern on the extremities; it is secondary to a parvovirus B19 infection, and an associated arthritis is seen in adults.

Both measles and rubella can occur in unvaccinated adults, and an atypical form of measles is seen in adults immunized with either killed measles vaccine or killed vaccine followed in time by live vaccine. In contrast to classic measles, the eruption of atypical measles begins on the palms, soles, wrists, and ankles, and the lesions may become purpuric. The patient with atypical measles can have pulmonary involvement and be quite ill. Rubelliform and roseoliform eruptions are also associated with *Epstein-Barr virus* (5–15% of patients), *echovirus*, *coxsackievirus*, *cytomegalovirus*, *adenovirus*, *SARS-CoV-2*, *dengue*, *chikungunya*, and *West Nile virus* infections. While detection of specific IgM antibodies or fourfold elevations in IgG antibodies often allows the proper diagnosis, polymerase chain reaction (PCR) has gradually replaced serologic assays. Occasionally, a maculopapular drug eruption is a reflection of an underlying viral infection. For example, ~95% of the patients with infectious mononucleosis who are given ampicillin will develop a rash.

Of note, early in the course of infections with *Rickettsia* and meningococcus, prior to the development of petechiae and purpura, the lesions may be erythematous macules and papules. This is also the case in chickenpox prior to the development of vesicles. Maculopapular eruptions are associated with early *HIV infection*, early *secondary syphilis*, *typhoid fever*, and *acute graft-versus-host disease*. In the last, lesions frequently begin on the dorsal hands, forearms, and upper trunk; the macular rose spots of typhoid fever involve primarily the anterior trunk.

The prototypic *scarlatiniform* eruption is seen in *scarlet fever* and is due to an erythrogenic toxin produced by bacteriophage-containing group A β-hemolytic streptococci, most commonly in the setting of pharyngitis. This eruption is characterized by diffuse erythema, which begins on the neck and upper trunk, and red follicular puncta. Additional findings include a white strawberry tongue (white coating with red papillae) followed by a red strawberry tongue (red tongue with red papillae); petechiae of the palate; a facial flush with circumoral pallor; linear petechiae in the antecubital fossae; and desquamation of the involved skin, palms, and soles 5–20 days after onset of the eruption. A similar desquamation of the palms and soles is seen in toxic shock syndrome (TSS), in Kawasaki disease, and after severe febrile illnesses. Certain strains of staphylococci also produce an erythrotoxin that leads to the same clinical findings as in streptococcal scarlet fever, except that the anti-streptolysin O or DNase B titers are not elevated.

In *toxic shock syndrome*, staphylococcal (phage group I) infections produce an exotoxin (TSST-1) that causes the fever and rash as well as enterotoxins. Initially, the majority of cases were reported in menstruating women who were using tampons. However, other sites of infection, including wounds and nasal packing, can lead to TSS. The diagnosis of TSS is based on clinical criteria **(Chap. 152)**, and three of these involve mucocutaneous sites (diffuse erythema of the skin, desquamation of the palms and soles 1–2 weeks after onset of illness, and involvement of the mucous membranes). The latter is characterized as hyperemia of the vagina, oropharynx, or conjunctivae. Similar systemic findings have been described in *streptococcal toxic shock syndrome* **(Chap. 153)**, and although an exanthem is seen less often than in TSS due to a staphylococcal infection, the underlying infection is often in the soft tissue (e.g., cellulitis).

The cutaneous eruption in *Kawasaki disease* **(Chap. 375)** is polymorphous, but the two most common forms are morbilliform and scarlatiniform. Additional mucocutaneous findings include bilateral conjunctival injection; erythema and edema of the hands and feet followed by desquamation; and diffuse erythema of the oropharynx, red strawberry tongue, and dry fissured lips. This clinical picture can resemble TSS and scarlet fever, but clues to the diagnosis of Kawasaki disease are cervical lymphadenopathy, cheilitis, and thrombocytosis. The most serious associated systemic finding in this disease is coronary aneurysms secondary to arteritis. Seen primarily in children, SARS-CoV-2–associated multisystem inflammatory syndrome must be distinguished from Kawasaki disease. Scarlatiniform eruptions are also seen in the early phase of SSSS (see "Vesicles/Bullae," above), in young adults with *Arcanobacterium haemolyticum* infection, and as reactions to drugs.

URTICARIA

(Table 61-14) *Urticaria* (hives) are transient lesions that are composed of a central wheal surrounded by an erythematous halo or flare. Individual lesions are round, oval, or figurate and are often pruritic. Acute and chronic urticarias have a wide variety of allergic etiologies and reflect edema in the dermis. Urticarial lesions can also be seen in patients with mastocytosis (urticaria pigmentosa), hypo- or hyperthyroidism, Schnitzler's syndrome, and systemic-onset juvenile idiopathic arthritis (Still's disease). In both juvenile- and adult-onset Still's disease, the lesions coincide with the fever spike, are transient, and are due to dermal infiltrates of neutrophils; the latter is also referred to as neutrophilic urticarial dermatosis.

The common *physical urticarias* include dermographism, solar urticaria, cold urticaria, and cholinergic urticaria. Patients with *dermographism* exhibit linear wheals following minor pressure or scratching of the skin and may be a contributing factor to pruritic dermatoses. It is a common disorder, affecting ~5% of the population. *Solar urticaria* characteristically occurs within minutes of sun exposure and is a skin sign of one systemic disease—erythropoietic protoporphyria. In addition to the urticaria, these patients have subtle pitted scarring of the nose and hands. *Cold urticaria* is precipitated by exposure to the cold, and therefore, exposed areas are usually affected. In occasional patients, the disease is associated with abnormal circulating proteins—more commonly cryoglobulins and less commonly cryofibrinogens. Additional systemic symptoms include wheezing and syncope, thus explaining the need for these patients to avoid swimming in cold water. Autosomal dominantly inherited cold urticaria is associated with dysfunction of cryopyrin. *Cholinergic urticaria* is precipitated by heat, exercise, or emotion and is characterized by small wheals with relatively large flares. It is occasionally associated with wheezing.

Whereas urticarias are the result of dermal edema, subcutaneous edema leads to the clinical picture of *angioedema*. Sites of involvement include the eyelids, lips, tongue, larynx, and gastrointestinal tract as well as the subcutaneous tissue. Angioedema occurs alone or in combination with urticaria, including urticarial vasculitis and the physical urticarias. Both acquired and hereditary (autosomal dominant) forms

TABLE 61-14 Causes of Urticaria and Angioedema

I. Primary cutaneous disorders
    A. Acute and chronic urticaria[a]
    B. Physical urticaria
        1. Dermographism
        2. Solar urticaria[b]
        3. Cold urticaria[b]
        4. Cholinergic urticaria[b]
    C. Angioedema (hereditary and acquired)[b,c]
II. Systemic diseases
    A. Urticarial vasculitis
    B. Hepatitis B or C viral infection, SARS-CoV-2 infection
    C. Serum sickness
    D. Angioedema (hereditary and acquired)

[a]A small minority develop anaphylaxis. [b]Also systemic. [c]Acquired angioedema can be idiopathic, associated with a lymphoproliferative disorder, or due to a drug, e.g., angiotensin-converting enzyme (ACE) inhibitors.

Abbreviation: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

of angioedema occur **(Chap. 366)**, and in the latter, urticaria is rarely, if ever, seen.

Urticarial vasculitis is an immune complex disease that may be confused with simple urticaria. In contrast to simple urticaria, individual lesions tend to last longer than 24 h and usually develop central petechiae that can be observed even after the urticarial phase has resolved. The patient may also complain of burning rather than pruritus. On biopsy, there is a leukocytoclastic vasculitis of the small dermal blood vessels. Although urticarial vasculitis may be idiopathic in origin, it can be a reflection of an underlying systemic illness such as lupus erythematosus, Sjögren's syndrome, or hereditary complement deficiency. There is a spectrum of urticarial vasculitis that ranges from purely cutaneous to multisystem involvement. The most common systemic signs and symptoms are arthralgias and/or arthritis, nephritis, and crampy abdominal pain, with asthma and chronic obstructive lung disease seen less often. Hypocomplementemia occurs in one- to two-thirds of patients, even in the idiopathic cases. Urticarial vasculitis can also be seen in patients with *hepatitis B* and *hepatitis C* infections and *serum sickness* but is usually not seen in *serum sickness–like illnesses* (e.g., due to cefaclor, minocycline).

PAPULONODULAR SKIN LESIONS

(Table 61-15) In the *papulonodular diseases*, the lesions are elevated above the surface of the skin and may coalesce to form larger plaques. The location, consistency, and color of the lesions are the keys to their diagnosis; this section is organized on the basis of color.

WHITE LESIONS

In *calcinosis cutis*, there are firm white to white-yellow papules with an irregular surface. When the contents are expressed, a chalky white material is seen. *Dystrophic calcification* is seen at sites of previous inflammation or damage to the skin. It develops in acne scars as well as on the distal extremities of patients with systemic sclerosis and in the subcutaneous tissue and intermuscular fascial planes in DM. The latter is more extensive and is more commonly seen in children. A previous or current elevated calcium phosphate product, most commonly due to secondary hyperparathyroidism in the setting of renal failure, can lead to nodules of *metastatic calcinosis cutis*, which tend to be subcutaneous and periarticular. These patients can also develop calcification of muscular arteries and subsequent ischemic necrosis (calciphylaxis). *Osteoma cutis*, in the form of small papules, most commonly occurs on the face of individuals with a history of acne vulgaris, whereas plate-like lesions occur in rare genetic syndromes.

SKIN-COLORED LESIONS

There are several types of skin-colored lesions, including epidermoid cysts, lipomas, rheumatoid nodules, neurofibromas, angiofibromas, neuromas, and adnexal tumors such as tricholemmomas. Both *epidermoid cysts* and *lipomas* are very common mobile subcutaneous nodules—the former are rubbery and drain cheeselike material (sebum and keratin) if incised. Lipomas are also rubbery and somewhat lobulated on palpation. When extensive facial epidermoid cysts develop during childhood or there is a family history of such lesions, the patient should be examined for other signs of Gardner syndrome, including osteomas and desmoid tumors. *Rheumatoid nodules* are firm 0.5- to 4-cm nodules that favor the extensor aspect of joints, especially the elbows. They are seen in ~20% of patients with rheumatoid arthritis and ~5% of patients with Still's disease. Biopsies of the nodules show palisading granulomas. Similar lesions that are smaller and shorter-lived are seen in rheumatic fever. Subcutaneous granuloma annulare is sometimes misdiagnosed as rheumatoid nodules.

Neurofibromas (benign Schwann cell tumors) are soft papules or nodules that exhibit the "button-hole" sign; that is, they invaginate into the skin with pressure in a manner similar to a hernia. Single lesions are seen in normal individuals, but multiple neurofibromas, usually in combination with six or more CALMs measuring >1.5 cm (see "Hyperpigmentation," above), axillary freckling, and multiple Lisch nodules, are seen in von Recklinghausen's disease (NF type I) **(Chap. 95)**. In some patients, the neurofibromas are localized and unilateral due to somatic mosaicism.

TABLE 61-15 Papulonodular Skin Lesions According to Color Groups

- I. White
  - A. Calcinosis cutis
  - B. Osteoma cutis (also skin-colored or blue)
- II. Skin-colored
  - A. Rheumatoid nodules
  - B. Neurofibromas (von Recklinghausen's disease [NF1])
  - C. Angiofibromas (tuberous sclerosis, MEN syndrome, type 1; also pink-red)
  - D. Neuromas (MEN syndrome, type 2b)
  - E. Adnexal tumors
    - 1. Basal cell carcinomas (basal cell nevus syndrome)
    - 2. Tricholemmomas (Cowden disease)
    - 3. Fibrofolliculomas (Birt-Hogg-Dubé syndrome)
  - F. Osteomas (arise in skull and jaw in Gardner syndrome)
  - G. Primary cutaneous disorders
    - 1. Epidermal inclusion cysts[a]
    - 2. Lipomas
- III. Pink/translucent[b]
  - A. Amyloidosis, primary systemic
  - B. Papular mucinosis/scleromyxedema
  - C. Multicentric reticulohistiocytosis
- IV. Yellow
  - A. Xanthomas
  - B. Tophi
  - C. Necrobiosis lipoidica
  - D. Pseudoxanthoma elasticum
  - E. Sebaceous adenomas (Muir-Torre syndrome)
- V. Red[b]
  - A. Papules
    - 1. Angiokeratomas (Fabry disease and related lysosomal storage diseases)[c]
    - 2. Bacillary angiomatosis (primarily in AIDS)
  - B. Papules/plaques
    - 1. Cutaneous lupus erythematosus
    - 2. Lymphoma cutis
    - 3. Leukemia cutis
    - 4. Sweet syndrome
  - C. Nodules
    - 1. Panniculitis
    - 2. Medium-sized vessel vasculitis (e.g., cutaneous polyarteritis nodosa/cutaneous arteritis)
  - D. Primary cutaneous disorders
    - 1. Arthropod bites
    - 2. Cherry hemangiomas
    - 3. Infections, e.g., streptococcal cellulitis, sporotrichosis
    - 4. Polymorphous light eruption
    - 5. Cutaneous lymphoid hyperplasia (lymphocytoma cutis, pseudolymphoma)
- VI. Red-brown[b]
  - A. Sarcoidosis
  - B. Urticaria pigmentosa
  - C. Erythema elevatum diutinum (chronic leukocytoclastic vasculitis)
  - D. Lupus vulgaris
- VII. Blue[b]
  - A. Venous malformations (e.g., blue rubber bleb syndrome)
  - B. Primary cutaneous disorders
    - 1. Venous lake
    - 2. Blue nevus
- VIII. Violaceous
  - A. Lupus pernio (sarcoidosis)
  - B. Lymphoma cutis
  - C. Cutaneous lupus erythematosus
- IX. Purple
  - A. Kaposi's sarcoma, acral angiodermatitis (pseudo-Kaposi's sarcoma)
  - B. Angiosarcoma
  - C. Palpable purpura (see Table 61-16)
  - D. Primary cutaneous disorders
    - 1. Angiokeratomas of the scrotum and vulva
- X. Brown-black[d]
- XI. Any color
  - A. Metastases

[a] If multiple with childhood onset, consider Gardner syndrome. [b] May have darker hue in more darkly pigmented individuals. [c] More widespread, especially lower trunk and girdle region, and often red-purple in color. [d] See also "Hyperpigmentation."

Abbreviations: MEN, multiple endocrine neoplasia; NF1, neurofibromatosis type 1.

Angiofibromas are firm pink-red to skin-colored papules that measure from 3 mm to 1.5 cm in diameter. When multiple lesions are located on the central cheeks (adenoma sebaceum), the patient should be evaluated for tuberous sclerosis or multiple endocrine neoplasia (MEN) syndrome, type 1. The former is an autosomal disorder due to mutations in two different genes, and the associated findings are discussed in the section on ash leaf spots as well as in **Chap. 95**.

Neuromas (benign proliferations of nerve fibers) are also firm, skin-colored papules. They are more commonly found at sites of amputations and in rudimentary polydactyly. However, when there are multiple neuromas on the eyelids, lips, distal tongue, and/or oral mucosa, the patient should be investigated for other signs of MEN syndrome, type 2b. Associated findings include marfanoid habitus, protuberant lips, intestinal ganglioneuromas, and medullary thyroid carcinoma (>75% of patients; **Chap. 400**).

Adnexal tumors are derived from pluripotent cells of the epidermis that can differentiate toward hair, sebaceous, apocrine or eccrine glands, or remain undifferentiated. *Basal cell carcinomas* (BCCs) are examples of adnexal tumors that have little or no evidence of differentiation. Clinically, they are translucent papules with rolled borders, telangiectasias, and central erosion. BCCs commonly arise in sun-damaged skin of the head and neck as well as the upper trunk. When a patient has multiple BCCs, especially prior to age 30, the possibility of the basal cell nevus syndrome should be raised. It is inherited as an autosomal dominant trait and is associated with jaw cysts, palmar and plantar pits, frontal bossing, medulloblastomas, and calcification of the falx cerebri and diaphragma sellae. *Tricholemmomas* are also skin-colored adnexal tumors but differentiate toward hair follicles and can have a wartlike appearance. The presence of multiple tricholemmomas on the face and cobblestoning of the oral mucosa points to the diagnosis of Cowden disease (PTEN hamartoma tumor syndrome) due to mutations in the phosphatase and tensin homolog (*PTEN*) gene. Internal organ involvement (in decreasing order of frequency) includes fibrocystic disease and carcinoma of the breast, adenomas and carcinomas of the thyroid, genitourinary carcinomas, and gastrointestinal polyposis. Keratoses of the palms, soles, and dorsal aspect of the hands are also seen. *Fibrofolliculomas* are skin-colored to white, smooth papules that favor the face, ears, and neck and, when multiple, are associated with Birt-Hogg-Dubé syndrome, which is associated with renal lesions including cancer as well as pulmonary cysts **(Chap. 90)**.

PINK LESIONS

The cutaneous lesions associated with *primary systemic amyloidosis* are often pink to pink-orange in color and translucent. Common locations are the face, especially the periorbital and perioral regions, and flexural areas. On biopsy, homogeneous deposits of amyloid are seen in the dermis and in the walls of blood vessels; the latter lead to an increase in vessel wall fragility. As a result, petechiae and purpura develop in clinically normal skin as well as in lesional skin following minor trauma, hence the term *pinch purpura*. Amyloid deposits are also seen in the striated muscle of the tongue and result in macroglossia.

Even though specific mucocutaneous lesions are present in only ~30% of the patients with primary systemic (AL) amyloidosis, the diagnosis can be made via histologic examination of abdominal subcutaneous fat, in conjunction with a serum free light chain assay. By special staining, amyloid deposits can be seen around blood vessels or individual fat cells. There are also three forms of amyloidosis that are limited to the skin and that should not be construed as cutaneous lesions of systemic amyloidosis. They are macular amyloidosis (upper back), lichen amyloidosis (usually lower extremities), and nodular amyloidosis. In macular and lichen amyloidosis, lesions are hyperpigmented and the deposits are composed of altered epidermal keratin. Early-onset macular and lichen amyloidosis have been associated with MEN syndrome, type 2a.

Patients with *multicentric reticulohistiocytosis* also have pink-colored papules and nodules on the face and mucous membranes as well as on the extensor surface of the hands and forearms. They have a polyarthritis that can mimic rheumatoid arthritis clinically. On histologic examination, the papules have characteristic giant cells that are not seen in biopsies of rheumatoid nodules. Pink to skin-colored papules that are firm, 2–5 mm in diameter, and often in a linear arrangement are seen in patients with *papular mucinosis*. This disease is also referred to as *scleromyxedema*. The latter name comes from the induration of the face and extremities that may accompany the papular eruption. Biopsy specimens of the papules show localized mucin deposition, and serum protein electrophoresis plus immunofixation electrophoresis demonstrates a monoclonal spike of IgG, usually with a λ light chain.

YELLOW LESIONS

Several systemic disorders are characterized by yellow-colored cutaneous papules or plaques—hyperlipidemia (xanthomas), gout (tophi), diabetes (necrobiosis lipoidica), pseudoxanthoma elasticum, and Muir-Torre syndrome (sebaceous tumors). Eruptive xanthomas are the most common form of *xanthomas* and are associated with hypertriglyceridemia (primarily hyperlipoproteinemia types I, IV, and V). Crops of yellow papules with erythematous halos occur primarily on the extensor surfaces of the extremities and the buttocks, and they spontaneously involute with a fall in serum triglycerides. Types II and III result in one or more of the following types of xanthoma: xanthelasma, tendon xanthomas, and plane xanthomas. Xanthelasma are found on the eyelids, whereas tendon xanthomas are frequently associated with the Achilles and extensor finger tendons; plane xanthomas are flat and favor the palmar creases and flexural folds. Tuberous xanthomas are frequently associated with hypercholesterolemia; however, they are also seen in patients with hypertriglyceridemia and are found most frequently over the large joints or hand. Biopsy specimens of xanthomas show collections of lipid-containing macrophages (foam cells).

Patients with several disorders, including biliary cirrhosis, can have a secondary form of hyperlipidemia with associated tuberous and plane xanthomas. However, patients with plasma cell dyscrasias have *normolipemic plane xanthomas*. This latter form of xanthoma may be ≥12 cm in diameter and is most frequently seen on the neck, upper trunk, and flexural folds. It is important to note that the most common setting for eruptive xanthomas is uncontrolled diabetes mellitus. The least specific sign for hyperlipidemia is xanthelasma, because at least 50% of the patients with this finding have normal lipid profiles.

In *tophaceous gout*, there are deposits of monosodium urate in the skin around the joints, particularly those of the hands and feet. Additional sites of *tophi* formation include the helix of the ear and the olecranon and prepatellar bursae. The lesions are firm, yellow to yellow-white in color, and occasionally discharge a chalky material. Their size varies from 1 mm to 7 cm, and the diagnosis can be established by polarized light microscopy of the aspirated contents of a tophus. Lesions of *necrobiosis lipoidica* are found primarily on the shins (90%), and patients can have diabetes mellitus or develop it subsequently. Characteristic findings include a central yellow color, atrophy (transparency), telangiectasias, and a red to red-brown border. Ulcerations can also develop within the plaques. Biopsy specimens show necrobiosis of collagen and granulomatous inflammation.

In *pseudoxanthoma elasticum* (PXE), due to mutations in the gene *ABCC6*, there is an abnormal deposition of calcium on the elastic fibers of the skin, eye, and blood vessels. In the skin, the flexural areas such as the neck, axillae, antecubital fossae, and inguinal area are the primary sites of involvement. Yellow papules coalesce to form reticulated plaques that have an appearance similar to that of plucked chicken skin. In severely affected skin, hanging, redundant folds develop. Biopsy specimens of involved skin show swollen and irregularly clumped elastic fibers with deposits of calcium. In the eye, the calcium deposits in Bruch's membrane lead to angioid streaks and choroiditis; in the arteries of the heart, kidney, gastrointestinal tract, and extremities, the deposits lead to angina, hypertension, gastrointestinal bleeding, and claudication, respectively.

Adnexal tumors that have differentiated toward sebaceous glands include sebaceous adenoma, sebaceous carcinoma, and sebaceous hyperplasia. Except for sebaceous hyperplasia, which is commonly seen on the face, these tumors are fairly rare. Patients with Muir-Torre syndrome have one or more *sebaceous adenoma(s)*, and they can also have sebaceous carcinomas and sebaceous hyperplasia as well as keratoacanthomas. As a variant of Lynch syndrome, the internal manifestations of Muir-Torre syndrome include *multiple* carcinomas of the gastrointestinal tract (primarily colon) as well as cancers of the genitourinary tract.

RED LESIONS

Cutaneous lesions that are red in color have a wide variety of etiologies; in an attempt to simplify their identification, they will be subdivided into papules, papules/plaques, and subcutaneous nodules. Common red papules include *arthropod bites* and *cherry hemangiomas*; the latter are small, bright-red, dome-shaped papules that represent a benign proliferation of capillaries. In patients with AIDS **(Chap. 208)**, the development of multiple red hemangioma-like lesions points to bacillary angiomatosis, and biopsy specimens show clusters of bacilli that stain positively with the Warthin-Starry stain; the pathogens have been identified as *Bartonella henselae* and *Bartonella quintana*. Disseminated visceral disease is seen primarily in immunocompromised hosts but can occur in immunocompetent individuals.

Multiple *angiokeratomas* are seen in Fabry disease, an X-linked recessive lysosomal storage disease that is due to a deficiency of α-galactosidase A. The lesions are red to red-purple in color and can be quite small in size (1–3 mm), with the most common location being the lower trunk. Associated findings include chronic renal disease, peripheral neuropathy, and corneal opacities (cornea verticillata). While electron photomicrographs demonstrate lamellar lipid deposits in dermal fibroblasts, pericytes, and endothelial cells, nowadays, genetic analysis is more frequently performed for diagnosis. Widespread acute eruptions of erythematous papules are discussed in the section on exanthems.

There are several infectious diseases that present as erythematous papules or nodules in a lymphocutaneous or sporotrichoid pattern, that is, in a linear arrangement along the lymphatic channels. The two most common etiologies are *Sporothrix schenckii* (sporotrichosis) and the atypical mycobacterium *Mycobacterium marinum*. The organisms are introduced as a result of trauma, and a primary inoculation site is often seen in addition to the lymphatic nodules. Additional causes include *Nocardia*, *Leishmania*, and other atypical mycobacteria and dimorphic fungi; culture or PCR of lesional tissue will aid in the diagnosis.

The diseases that are characterized by erythematous plaques with scale are reviewed in the papulosquamous section, and the various forms of dermatitis are discussed in the section on erythroderma. Additional disorders in the differential diagnosis of red papules/plaques include *cellulitis*, *polymorphous light eruption* (PMLE), *cutaneous lymphoid hyperplasia* (lymphocytoma cutis), *cutaneous lupus*, *lymphoma cutis*, and *leukemia cutis*. The first three diseases represent primary cutaneous disorders, although cellulitis may be accompanied by a bacteremia. PMLE is characterized by erythematous papules and plaques in a primarily sun-exposed distribution—dorsum of the hand, extensor forearm, and upper trunk. Lesions follow exposure to UV-B and/or UV-A, and in higher latitudes, PMLE is most severe in the late spring and early summer. A process referred to as "hardening" occurs with continued UV exposure, and the eruption fades spontaneously, but in temperate climates, it recurs the next spring. PMLE must be differentiated from cutaneous lupus, and this is accomplished by observation of the natural history, histologic examination, and sometimes direct immunofluorescence of the lesions. Cutaneous lymphoid hyperplasia (pseudolymphoma) is a *benign* polyclonal proliferation of lymphocytes within the skin that presents as infiltrated pink-red to red-purple papules and plaques; it must be distinguished from lymphoma cutis.

Several types of red plaques are seen in patients with systemic *lupus*, including (1) erythematous urticarial plaques across the cheeks and nose in the classic butterfly rash; (2) erythematous discoid lesions with fine or "carpet-tack" scale, telangiectasias, central hypopigmentation, peripheral hyperpigmentation, follicular plugging, and atrophy located on the scalp, face, external ears, arms, and upper trunk; and (3) psoriasiform or annular lesions of subacute cutaneous lupus with hypopigmented centers located primarily on the extensor arms and upper trunk. Additional mucocutaneous findings include (1) a violaceous flush on the face and V of the neck; (2) photosensitivity; (3) urticarial vasculitis (see "Urticaria," above); (4) lupus panniculitis (see below); (5) diffuse alopecia; (6) alopecia secondary to discoid lesions; (7) proximal nailfold telangiectasias and erythema; (8) EM- or TEN-like lesions that may become bullous; (9) oral or nasal ulcers; (10) livedo reticularis; and (11) distal ulcerations secondary to Raynaud's phenomenon, vasculitis, or livedoid vasculopathy. Patients with only discoid lesions usually have the form of lupus that is limited to the skin. However, up to 10–15% of these patients eventually develop systemic lupus. Direct immunofluorescence of involved skin, in particular discoid lesions, shows deposits of IgG or IgM and C3 in a granular distribution along the dermal-epidermal junction.

In *lymphoma cutis*, there is a clonal proliferation of malignant lymphocytes within the skin, and the clinical appearance resembles that of cutaneous lymphoid hyperplasia—infiltrated pink-red to red-purple papules and plaques. Lymphoma cutis can occur anywhere on the surface of the skin, whereas the sites of predilection for lymphocytomas include the malar ridge, tip of the nose, and earlobes. Patients with non-Hodgkin's lymphomas have specific cutaneous lesions more often than those with Hodgkin's lymphoma, and, occasionally, the skin nodules precede the development of extracutaneous non-Hodgkin's lymphoma or represent the only site of involvement (e.g., primary cutaneous B-cell lymphoma). Arcuate lesions are sometimes seen in lymphoma and lymphocytoma cutis as well as in CTCL. *Adult T-cell leukemia/lymphoma* that develops in association with HTLV-1 infection is characterized by cutaneous plaques, hypercalcemia, and circulating CD25+ lymphocytes. *Leukemia cutis* has the same appearance as lymphoma cutis, and specific lesions are seen more commonly in monocytic leukemias than in lymphocytic or granulocytic leukemias. Cutaneous chloromas (granulocytic sarcomas) may precede the appearance of circulating blasts in acute myelogenous leukemia and, as such, represent a form of aleukemic leukemia cutis.

Sweet syndrome is characterized by pink-red to red-brown edematous plaques that are frequently painful and occur primarily on the head, neck, and upper extremities. The patients also have fever, neutrophilia, and a dense dermal infiltrate of neutrophils in the lesions. In ~10% of the patients, there is an associated malignancy, most commonly acute myelogenous leukemia. Sweet syndrome has also been reported with myelodysplasia, inflammatory bowel disease, systemic lupus erythematosus, and solid tumors (primarily of the genitourinary tract) as well as drugs (e.g., granulocyte colony-stimulating factor [G-CSF], hypomethylating agents, all-*trans*-retinoic acid). The differential diagnosis includes neutrophilic eccrine hidradenitis; bullous forms of pyoderma gangrenosum; and, occasionally, cellulitis. Extracutaneous sites of involvement include joints, muscles, eyes, kidneys (proteinuria, occasionally glomerulonephritis), and lungs (neutrophilic infiltrates). The idiopathic form of Sweet syndrome is seen more often in women, following a respiratory tract infection.

Common causes of erythematous subcutaneous nodules include inflamed epidermoid cysts, acne cysts, and furuncles. *Panniculitis*, an inflammation of the fat, also presents as subcutaneous nodules and is frequently a sign of systemic disease. There are several forms of panniculitis, including erythema nodosum, erythema induratum/nodular vasculitis, lupus panniculitis, lipodermatosclerosis, α_1-antitrypsin deficiency, factitial, and fat necrosis secondary to pancreatic disease. Except for erythema nodosum, these lesions may break down and ulcerate or heal with a scar. The shin is the most common location for the nodules of erythema nodosum, whereas the calf is the most common location for lesions of erythema induratum. In erythema nodosum, the nodules are initially red but then develop a blue bruise-like color as they resolve. Patients with erythema nodosum but no underlying systemic illness can still have fever, malaise, leukocytosis, arthralgias, and/or arthritis. However, the possibility of an underlying illness should be excluded, and the most common associations are streptococcal infections, upper respiratory viral infections, sarcoidosis, and inflammatory bowel disease, in addition to drugs (oral contraceptives, sulfonamides, penicillins, bromides, iodides, BRAF inhibitors). Less common associations include bacterial gastroenteritis (*Yersinia*, *Salmonella*) and coccidioidomycosis followed by pregnancy, Sweet syndrome, tuberculosis, histoplasmosis, brucellosis, and infections with *Chlamydophilia pneumoniae*, *Chlamydia trachomatis, Mycoplasma pneumoniae,* or hepatitis B virus.

Erythema induratum and nodular vasculitis have overlapping features clinically and histologically, and whether they represent two separate entities or the ends of a single disease spectrum is a point of debate; in general, the latter is usually idiopathic and the former is associated with the presence of *Mycobacterium tuberculosis* DNA by PCR within skin lesions. The lesions of lupus panniculitis are found primarily on the cheeks, upper arms, and buttocks (sites of abundant fat) and are seen in both the cutaneous and systemic forms of lupus. The overlying skin may be normal, erythematous, or have the changes of discoid lupus. The subcutaneous fat necrosis that is associated with pancreatic disease is presumably secondary to circulating lipases and is seen in patients with pancreatic carcinoma as well as in patients with acute and chronic pancreatitis. In this disorder, there may be an associated arthritis, fever, and inflammation of visceral fat. Histologic examination of deep incisional biopsy specimens will aid in the diagnosis of the particular type of panniculitis.

Subcutaneous erythematous nodules are also seen in cutaneous polyarteritis nodosa and as a manifestation of *systemic vasculitis* when there is involvement of medium-sized vessels, for example, systemic polyarteritis nodosa, eosinophilic granulomatosis with polyangiitis, or granulomatosis with polyangiitis **(Chap. 375)**. Cutaneous polyarteritis nodosa, more recently referred to as cutaneous arteritis, presents with painful subcutaneous nodules and ulcers within a red-purple, netlike pattern of livedo reticularis. The latter is due to slowed blood flow through the superficial horizontal venous plexus. The majority of lesions are found on the lower extremities, and while arthralgias and myalgias may accompany cutaneous polyarteritis nodosa, there is no evidence of systemic involvement. In both the cutaneous and systemic forms of vasculitis, skin biopsy specimens of the associated nodules will show the changes characteristic of a necrotizing vasculitis and/or granulomatous inflammation.

■ RED-BROWN LESIONS

The cutaneous lesions in *sarcoidosis* **(Chap. 379)** are classically red to red-brown in color, and with diascopy (pressure with a glass slide), a yellow-brown residual color is observed that is secondary to the granulomatous infiltrate. The waxy papules and plaques may be found anywhere on the skin, but the face is the most common location. Usually there are no surface changes, but occasionally, the lesions will have scale. Biopsy specimens of the papules show "naked" granulomas in the dermis, that is, granulomas surrounded by a minimal number of lymphocytes. Other cutaneous findings in sarcoidosis include annular lesions with an atrophic or scaly center, papules within scars, hypopigmented papules and plaques, subcutaneous plaques, alopecia, acquired ichthyosis, erythema nodosum, and lupus pernio (see below).

The differential diagnosis of sarcoidosis includes foreign-body granulomas produced by chemicals such as beryllium and zirconium, late secondary syphilis, reactive granulomatous dermatitis, and *lupus vulgaris*. Lupus vulgaris is a form of cutaneous tuberculosis that is seen in previously infected and sensitized individuals. There is often underlying active tuberculosis elsewhere, usually in the lungs or lymph nodes. Lesions occur primarily in the head and neck region and are red-brown plaques with a yellow-brown color on diascopy. Secondary scarring can develop within the central portion of the plaques. Cultures or PCR analysis of the lesions should be performed, along with an interferon γ release assay of peripheral blood, because it is rare for the acid-fast stain to show bacilli within the dermal granulomas.

A generalized distribution of red-brown macules and papules is seen in the maculopapular form of mastocytosis, also known as *urticaria pigmentosa* **(Chap. 366)**. Each lesion represents a collection of mast cells in the dermis, with hyperpigmentation of the overlying epidermis. Stimuli such as rubbing cause these mast cells to degranulate, and this leads to the formation of localized urticaria (Darier's sign). Additional symptoms can result from mast cell degranulation and include headache, flushing, diarrhea, and pruritus. Mast cells also infiltrate various organs such as the liver, spleen, and gastrointestinal tract, and accumulations of mast cells in the bones may produce either osteosclerotic or osteolytic lesions on radiographs. In the majority of these patients, however, the internal involvement remains indolent. A subtype of chronic cutaneous small-vessel vasculitis, *erythema elevatum diutinum* (EED), also presents with papules that are red-brown in color. The papules coalesce into plaques on the extensor surfaces of knees, elbows, and the small joints of the hand. Flares of EED have been associated with streptococcal infections.

■ BLUE LESIONS

Lesions that are blue in color are the result of vascular ectasias, hyperplasias, and tumors or melanin pigment within the dermis. *Venous lakes* (ectasias) are compressible dark-blue lesions that are found commonly in the head and neck region. *Venous malformations* are also compressible blue papulonodules and plaques that can occur anywhere on the body, including the oral mucosa. When there are multiple papulonodules rather than a single congenital lesion, the patient may have the blue rubber bleb syndrome or Maffucci's syndrome due to mutations in *TEK* or *IDH1*, respectively. Patients with the blue rubber bleb syndrome also have vascular anomalies of the gastrointestinal tract that may bleed, whereas patients with Maffucci's syndrome have associated osteochondromas. *Blue nevi* (moles) are seen when there are collections of pigment-producing nevus cells in the dermis. These benign papular lesions are dome-shaped and occur most commonly on the dorsum of the hand or foot or in the head and neck or presacral region.

■ VIOLACEOUS LESIONS

Violaceous papules and plaques are seen in *lupus pernio*, *lymphoma cutis*, and *cutaneous lupus*. Lupus pernio is a particular type of sarcoidosis that involves the tip and alar rim of the nose as well as the earlobes, with lesions that are violaceous in color rather than red-brown. This form of sarcoidosis is associated with involvement of the upper respiratory tract. The plaques of lymphoma cutis and cutaneous lupus may be red or violaceous in color and were discussed above.

■ PURPLE LESIONS

Purple-colored papules and plaques are seen in vascular tumors, such as *Kaposi's sarcoma* **(Chap. 208)** and *angiosarcoma*, and when there is extravasation of red blood cells into the skin in association with inflammation, as in *palpable purpura* (see "Purpura," below). Patients with congenital or acquired AV fistulas and venous hypertension can develop purple papules on the lower extremities that can resemble Kaposi's sarcoma clinically and histologically; this condition is referred to as pseudo-Kaposi's sarcoma (acral angiodermatitis). Angiosarcoma is found most commonly on the scalp and face of elderly patients or within areas of chronic lymphedema and presents as purple papules and plaques. In the head and neck region, the tumor often extends beyond the clinically defined borders and may be accompanied by facial edema.

■ BROWN AND BLACK LESIONS

Brown- and black-colored papules are reviewed in "Hyperpigmentation," above.

■ CUTANEOUS METASTASES

These are discussed last because they can have a wide range of colors. Most commonly, they present as either firm, skin-colored subcutaneous nodules or firm, red to red-brown papulonodules, whereas metastatic melanoma can be pink, blue, or black in color. Cutaneous metastases develop from hematogenous or lymphatic spread and are most often due to the following primary carcinomas: in men, melanoma, oropharynx, lung, and colon; and in women, breast, melanoma, and ovary. These metastatic lesions may be the initial presentation of the carcinoma, especially when the primary site is the lung.

PURPURA

(Table 61-16) *Purpura* are seen when there is an extravasation of red blood cells into the dermis and, as a result, the lesions do not blanch with pressure. This is in contrast to those erythematous or violet-colored

TABLE 61-16 Causes of Purpura

I. Primary cutaneous disorders
  A. Nonpalpable
    1. Trauma
    2. Solar (actinic, senile) purpura
    3. Steroid purpura
    4. Stasis purpura due to venous hypertension
    5. Capillaritis
    6. Livedoid vasculopathy in the setting of venous hypertension[a]
II. Drugs (e.g., antiplatelet agents, anticoagulants)
III. Systemic diseases
  A. Nonpalpable
    1. Clotting disturbances
      a. Thrombocytopenia (including ITP)
      b. Abnormal platelet function
      c. Clotting factor defects
    2. Vascular fragility
      a. Amyloidosis (within normal-appearing skin)
      b. Ehlers-Danlos syndrome
      c. Scurvy
    3. Thrombi
      a. Disseminated intravascular coagulation, purpura fulminans
      b. Warfarin (Coumadin®)-induced necrosis
      c. Heparin-induced thrombocytopenia and thrombosis
      d. Antiphospholipid antibody syndrome
      e. Monoclonal cryoglobulinemia
      f. Vasculopathy induced by levamisole-adulterated cocaine[b]
      g. SARS-CoV-2 infection
      h. Thrombotic thrombocytopenic purpura
      i. Thrombocytosis
      j. Homozygous protein C or protein S deficiency
    4. Emboli
      a. Cholesterol
      b. Fat
    5. Possible immune complex
      a. Gardner-Diamond syndrome (autoerythrocyte sensitivity)
      b. Waldenström's hypergammaglobulinemic purpura
    6. Calciphylaxis
  B. Palpable
    1. Vasculitis
      a. Cutaneous small-vessel vasculitis, including in the setting of systemic vasculitides
    2. Emboli[c]
      a. Acute meningococcemia
      b. Disseminated gonococcal infection
      c. Rocky Mountain spotted fever
      d. Ecthyma gangrenosum

[a] Also associated with underlying disorders that lead to hypercoagulability/thrombophilia, e.g., factor V Leiden, protein C dysfunction/deficiency. [b] Combined vasculopathy/vasculitis can be seen. [c] Bacterial (including rickettsial), fungal, or parasitic.

Abbreviations: ITP, idiopathic thrombocytopenic purpura; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

lesions that are due to localized vasodilatation—they do blanch with pressure. Purpura (≥3 mm) and petechiae (≤2 mm) are divided into two major groups: palpable and nonpalpable (macular). The most frequent causes of *nonpalpable* purpura and petechiae are primary cutaneous disorders such as *trauma*, *solar* (*actinic*) *purpura*, *stasis purpura*, and *capillaritis*. Less common causes are *steroid purpura* and *livedoid vasculopathy* (see "Ulcers," below). Solar purpura are seen primarily on the extensor forearms, whereas steroid purpura secondary to potent topical glucocorticoids or endogenous or exogenous Cushing's syndrome can be more widespread. In both cases, there is alteration of the supporting connective tissue that surrounds the dermal blood vessels. In contrast, the petechiae that result from capillaritis are found primarily on the lower extremities. In capillaritis, there is an extravasation of erythrocytes as a result of perivascular lymphocytic inflammation. The petechiae are bright red, 1–2 mm in size, and scattered within yellow-brown patches. The yellow-brown color is caused by hemosiderin deposits within the dermis.

Systemic causes of nonpalpable purpura fall into several categories, and those secondary to clotting disturbances and vascular fragility will be discussed first. The former group includes *thrombocytopenia* **(Chap. 120)**, *abnormal platelet function* as is seen in uremia, and *clotting factor defects*. The initial site of presentation for thrombocytopenia-induced petechiae is the distal lower extremity. Capillary fragility leads to nonpalpable purpura in patients with systemic *amyloidosis* (see "Papulonodular Skin Lesions," above), disorders of collagen production such as *Ehlers-Danlos syndrome*, and *scurvy*. In scurvy, there are flattened corkscrew hairs with surrounding perifollicular hemorrhage on the lower extremities, in addition to gingivitis. Vitamin C is a cofactor for lysyl hydroxylase, an enzyme involved in the posttranslational modification of procollagen that is necessary for cross-link formation.

In contrast to the previous group of disorders, the noninflammatory purpura seen in the following group of diseases are associated with thrombi formation within vessels and have a retiform configuration. It is important to note that these thrombi are demonstrable in skin biopsy specimens. This group of disorders includes disseminated intravascular coagulation (DIC), monoclonal cryoglobulinemia, thrombocytosis, thrombotic thrombocytopenic purpura, antiphospholipid antibody syndrome, and reactions to warfarin and heparin (heparin-induced thrombocytopenia and thrombosis). DIC is triggered by several types of infection (gram-negative, gram-positive, viral, and rickettsial) as well as by tissue injury and neoplasms. Widespread purpura and hemorrhagic infarcts of the distal extremities are seen. Similar lesions are found in purpura fulminans, which is a form of DIC associated with fever and hypotension that occurs more commonly in children following an infectious illness such as varicella, scarlet fever, or an upper respiratory tract infection. In both disorders, hemorrhagic bullae can develop in involved skin.

Monoclonal cryoglobulinemia is associated with plasma cell dyscrasias, chronic lymphocytic leukemia, and lymphoma. Purpura, primarily of the lower extremities, and hemorrhagic infarcts of the fingers, toes, nose, and ears are seen in these patients. Exacerbations of disease activity can follow cold exposure or an increase in serum viscosity. Biopsy specimens show precipitates of the cryoglobulin within dermal vessels. Similar deposits have been found in the lung, brain, and renal glomeruli. Patients with *thrombotic thrombocytopenic purpura* can also have hemorrhagic infarcts as a result of intravascular thromboses. Additional signs include microangiopathic hemolytic anemia and fluctuating neurologic abnormalities, especially headaches and confusion.

Administration of *warfarin* can result in painful areas of erythema that become purpuric and then necrotic with an adherent black eschar; the condition is also referred to as Coumadin®-induced necrosis. This reaction is seen more often in women and in areas with abundant subcutaneous fat—breasts, abdomen, buttocks, thighs, and calves. The erythema and purpura develop between the third and tenth day of therapy, most likely as a result of a transient imbalance in the levels of anticoagulant and procoagulant vitamin K–dependent factors. Continued therapy does not exacerbate preexisting lesions, and patients with an inherited or acquired deficiency of protein C are at increased risk for this particular reaction as well as for purpura fulminans and calciphylaxis. The latter can have a similar clinical appearance.

Purpura secondary to *cholesterol emboli* are usually seen on the lower extremities of patients with atherosclerotic vascular disease. They often follow anticoagulant therapy or an invasive vascular procedure such as an arteriogram but also occur spontaneously from disintegration of atheromatous plaques. Associated findings include livedo reticularis, gangrene, cyanosis, ischemic ulcerations, and peripheral eosinophilia. Multiple step sections of the biopsy specimen may be necessary to demonstrate the cholesterol clefts within the vessels. Petechiae are also an important sign of *fat embolism* and occur primarily on the upper body 2–3 days after a major injury. By using special fixatives, the emboli can be demonstrated in biopsy specimens of the petechiae. Rarely, emboli of tumor or thrombus are seen in patients with atrial myxomas and marantic endocarditis.

In the *Gardner-Diamond syndrome* (autoerythrocyte sensitivity), female patients develop large ecchymoses within areas of painful, warm erythema. Intradermal injections of autologous erythrocytes or phosphatidyl serine derived from the red cell membrane can reproduce the lesions in some patients; however, there are instances where a reaction is seen at an injection site of the forearm but not in the midback region. The latter has led some observers to view Gardner-Diamond syndrome as a cutaneous manifestation of severe emotional stress. More recently, the possibility of platelet dysfunction (as assessed via aggregation studies) has been raised. *Waldenström's hypergammaglobulinemic purpura*, more recently referred to as *recurrent macular vasculitis in hypergammaglobulinemia*, is a chronic disorder characterized by recurrent crops of petechiae and larger purpuric macules on the lower extremities. There are circulating complexes of IgG or IgA rheumatoid factor, and exacerbations are associated with prolonged standing or walking. Patients may have an underlying autoimmune connective tissue disease, e.g., Sjögren's syndrome.

Palpable purpura are further subdivided into vasculitic and embolic. In the group of vasculitic disorders, cutaneous small-vessel vasculitis, also known as *leukocytoclastic vasculitis* (LCV), is the one most commonly associated with palpable purpura **(Chap. 375)**. Underlying etiologies include drugs (e.g., antibiotics), infections (e.g., hepatitis C virus), and autoimmune connective tissue diseases (e.g., rheumatoid arthritis, Sjögren's syndrome, lupus). *Henoch-Schönlein purpura* (HSP) is a subtype of acute LCV that is seen more commonly in children and adolescents following an upper respiratory infection. The majority of lesions are found on the lower extremities and buttocks. Systemic manifestations include fever, arthralgias (primarily of the knees and ankles), abdominal pain, gastrointestinal bleeding, and nephritis. Direct immunofluorescence examination shows deposits of IgA within dermal blood vessel walls. Renal disease is of particular concern in adults with IgA vasculitis.

Several types of infectious emboli can give rise to palpable purpura. These embolic lesions are usually *irregular* in outline as opposed to the lesions of LCV, which are *circular* in outline. The irregular outline is indicative of a cutaneous infarct, and the size corresponds to the area of skin that received its blood supply from that particular arteriole or artery. The palpable purpura in LCV are circular because the erythrocytes simply diffuse out evenly from the postcapillary venules as a result of inflammation. Infectious emboli are most commonly due to gram-negative cocci (meningococcus, gonococcus), gram-negative rods (Enterobacteriaceae), and gram-positive cocci (*Staphylococcus*). Additional causes include *Rickettsia* and, in immunocompromised patients, *Aspergillus* and other opportunistic fungi.

The embolic lesions in *acute meningococcemia* are found primarily on the trunk, lower extremities, and sites of pressure, and a gunmetal-gray color often develops within them. Their size varies from a few millimeters to several centimeters, and the organisms can be cultured from the lesions. Associated findings include a preceding upper respiratory tract infection; fever; meningitis; DIC; and, in some patients, a deficiency of the terminal components of complement. In *disseminated gonococcal infection* (arthritis-dermatitis syndrome), a small number of inflammatory papules and vesicopustules, often with central purpura or hemorrhagic necrosis, are found on the distal extremities. Additional symptoms include arthralgias, tenosynovitis, and fever. To establish the diagnosis, a Gram stain of these lesions should be performed. *Rocky Mountain spotted fever* is a tick-borne disease that is caused by *Rickettsia rickettsii*. A several-day history of fever, chills, severe headache, and photophobia precedes the onset of the cutaneous eruption. The initial lesions are erythematous macules and papules on the wrists, ankles, palms, and soles. With time, the lesions spread centripetally and become purpuric.

Lesions of *ecthyma gangrenosum* begin as edematous, erythematous papules or plaques and then develop central purpura and necrosis. Bullae formation also occurs in these lesions, and they are frequently found in the girdle region. The organism that is classically associated with ecthyma gangrenosum is *Pseudomonas aeruginosa*, but other gram-negative rods such as *Klebsiella*, *Escherichia coli*, and *Serratia* can produce similar lesions. In immunocompromised hosts, the list of potential pathogens is expanded to include *Candida* and other opportunistic fungi (e.g., *Aspergillus*, *Fusarium*).

ULCERS

The approach to the patient with a cutaneous ulcer is outlined in **Table 61-17**. Peripheral vascular diseases of the extremities are reviewed in **Chap. 292**, as is Raynaud's phenomenon.

Livedoid vasculopathy (livedoid vasculitis; atrophie blanche) represents a combination of a vasculopathy plus intravascular thrombosis. Purpuric lesions and livedo reticularis are found in association with *painful* ulcerations of the lower extremities. These ulcers are often

TABLE 61-17 Causes of Mucocutaneous Ulcers

I. Primary cutaneous disorders
   - A. Peripheral vascular disease **(Chap. 292)**
      - 1. Venous
      - 2. Arterial[a]
   - B. Livedoid vasculopathy in the setting of venous hypertension[b]
   - C. Squamous cell carcinoma (e.g., within scars), basal cell carcinomas
   - D. Infections, e.g., ecthyma caused by *Streptococcus* **(Chap. 153)**
   - E. Physical, e.g., trauma, pressure
   - F. Drugs, e.g., hydroxyurea

II. Systemic diseases
   - A. Lower legs
      - 1. Small-vessel and medium-vessel vasculitis[c]
      - 2. Hemoglobinopathies **(Chap. 103)**
      - 3. Cryoglobulinemia,[c] cryofibrinogenemia
      - 4. Cholesterol emboli[a,c]
      - 5. Necrobiosis lipoidica[d]
      - 6. Antiphospholipid syndrome **(Chap. 121)**
      - 7. Neuropathic[e] **(Chap. 415)**
      - 8. Panniculitis
      - 9. Kaposi's sarcoma, acral angiodermatitis (pseudo-Kaposi's sarcoma)
      - 10. Diffuse dermal angiomatosis
   - B. Hands and feet
      - 1. Raynaud's phenomenon **(Chap. 292)**
      - 2. Buerger disease
   - C. Generalized
      - 1. Pyoderma gangrenosum, but most commonly legs
      - 2. Calciphylaxis **(Chap. 422)**
      - 3. Infections, e.g., dimorphic fungi, leishmaniasis
      - 4. Lymphoma
   - D. Face, especially perioral, and anogenital
      - 1. Chronic herpes simplex[f]

III. Mucosal
   - A. Aphthae
   - B. Drug-induced mucositis
   - C. Behçet's disease **(Chap. 376)**
   - D. Erythema multiforme major, Stevens-Johnson syndrome, TEN
   - E. Primary blistering disorders **(Chap. 62)**
   - F. Lupus erythematosus, lichen planus, lichenoid GVHD
   - G. Inflammatory bowel disease
   - H. Acute HIV infection
   - I. Reactive arthritis

[a]Underlying atherosclerosis. [b]Also associated with underlying disorders that lead to hypercoagulability/thrombophilia, e.g., factor V Leiden, protein C dysfunction/deficiency, antiphospholipid antibodies. [c]Reviewed in section on purpura. [d]Reviewed in section on papulonodular skin lesions. [e]Favors plantar surface of the foot. [f]Sign of immunosuppression.

Abbreviations: GVHD, graft versus host disease; HIV, human immunodeficiency virus; TEN, toxic epidermal necrolysis.

slow to heal, but when they do, irregularly shaped white scars form. The majority of cases are secondary to venous hypertension, but possible underlying illnesses include disorders of hypercoagulability, for example, antiphospholipid syndrome and factor V Leiden **(Chaps. 122 and 369)**.

In *pyoderma gangrenosum*, the border of untreated active ulcers has a characteristic appearance consisting of an undermined necrotic violaceous edge and a peripheral erythematous halo. The ulcers often begin as pustules that then expand rather rapidly to a size as large as 20 cm. Although these lesions are most commonly found on the lower extremities, they can arise anywhere on the surface of the body, including at sites of trauma (pathergy). An estimated 30–50% of cases are idiopathic, and the most common associated disorder is inflammatory bowel disease. Less commonly, pyoderma gangrenosum is associated with seropositive rheumatoid arthritis, myelodysplasia, acute myelogenous leukemia, a monoclonal gammopathy (usually IgA), or an autoinflammatory disorder. Because the histology of pyoderma gangrenosum may be nonspecific (dermal infiltrate of neutrophils when in untreated state), the diagnosis requires clinicopathologic correlation, in particular, the exclusion of similar-appearing ulcers such as vasculitis, Meleney's ulcer (synergistic infection at a site of trauma or surgery), dimorphic fungi, cutaneous amebiasis, spider bites, and factitial. In the hematologic disorders, the ulcers may be more superficial with a pustulobullous border, and these lesions provide a connection between classic pyoderma gangrenosum and acute febrile neutrophilic dermatosis (Sweet syndrome).

FEVER AND RASH

The major considerations in a patient with a fever and a rash are inflammatory diseases versus infectious diseases. In the hospital setting, the most common scenario is a patient who has a drug rash plus a fever secondary to an underlying infection. However, it should be emphasized that a drug reaction can lead to both a cutaneous eruption and a fever ("drug fever"), especially in the setting of DRESS, AGEP, or serum sickness–like reaction. Additional inflammatory diseases that are often associated with a fever include pustular psoriasis, erythroderma, and Sweet syndrome. Lyme disease, secondary syphilis, and viral and bacterial exanthems (see "Exanthems," above) are examples of infectious diseases that produce a rash and a fever. Lastly, it is important to determine whether or not the cutaneous lesions represent septic emboli (see "Purpura," above). Such lesions usually have evidence of ischemia in the form of purpura, necrosis, or impending necrosis (gunmetal-gray color). In the patient with thrombocytopenia, however, purpura can be seen in inflammatory reactions such as morbilliform drug eruptions and infectious lesions. In addition, because of venous hypertension, lesions of a morbilliform drug eruption that are below the knee can be purpuric.

FURTHER READING

Bologna JL, Schaffer JV, Cerroni L (eds): *Dermatology*, 5th ed. Philadelphia, Elsevier, 2024.

Callen JP et al (eds): *Dermatological Signs of Systemic Disease*, 5th ed. Edinburgh, Elsevier, 2017.

Fazel N (ed): *Oral Signs of Systemic Disease*. Switzerland, Springer, 2019.

Kurtzman D: Rheumatologic dermatology. Clin Dermatol 36:439, 2018.

Taylor SC et al (eds): *Taylor and Kelly's Dermatology for Skin of Color*, 2nd ed. New York, McGraw-Hill, 2016.

62

Immunologically Mediated Skin Diseases

Kim B. Yancey, Benjamin F. Chong,
Thomas J. Lawley

A number of immunologically mediated skin diseases and immunologically mediated systemic disorders with cutaneous manifestations are now recognized as distinct entities with consistent clinical, histologic, and immunopathologic findings. Clinically, these disorders are characterized by morbidity (pain, pruritus, disfigurement) and, in some instances, risk of mortality (largely due to loss of epidermal barrier function and/or secondary infection). The major features of the more common immunologically mediated skin diseases are summarized in this chapter **(Table 62-1)**, as are autoimmune systemic disorders with cutaneous manifestations.

AUTOIMMUNE CUTANEOUS DISEASES

PEMPHIGUS VULGARIS

Pemphigus refers to a group of autoantibody-mediated intraepidermal blistering diseases characterized by loss of cohesion between epidermal cells (a process termed *acantholysis*). Manual pressure to the skin of these patients may elicit the separation of the epidermis (*Nikolsky's sign*). This finding, while characteristic of pemphigus, is not specific to this group of disorders and is also seen in toxic epidermal necrolysis, Stevens-Johnson syndrome, and a few other skin diseases.

Pemphigus vulgaris (PV) is a mucocutaneous blistering disease that predominantly occurs in patients >40 years of age. PV typically begins on mucosal surfaces and often progresses to involve the skin. This disease is characterized by fragile, flaccid blisters that rupture to produce extensive denudation of mucous membranes and skin **(Fig. 62-1)**. The mouth, scalp, face, neck, axilla, groin, and trunk are typically involved. PV may be associated with severe skin pain; some patients experience pruritus as well. Lesions usually heal without scarring except at sites complicated by secondary infection or mechanically induced dermal wounds. Postinflammatory hyperpigmentation is usually present for some time at sites of healed lesions.

Biopsies of early lesions demonstrate intraepidermal vesicle formation secondary to loss of cohesion between epidermal cells (i.e., acantholytic blisters). Blister cavities contain acantholytic epidermal cells, which appear as round homogeneous cells containing hyperchromatic nuclei. Basal keratinocytes remain attached to the epidermal basement membrane; hence, blister formation takes place within the suprabasal portion of the epidermis. Lesional skin may contain focal collections of intraepidermal eosinophils within blister cavities; dermal alterations are slight, often limited to an eosinophil-predominant leukocytic infiltrate. Direct immunofluorescence microscopy of lesional or intact patient skin shows deposits of IgG on the surface of keratinocytes; deposits of complement components are typically found in lesional but not in uninvolved skin. Deposits of IgG on keratinocytes are derived from circulating autoantibodies to cell-surface autoantigens. Such circulating autoantibodies can be demonstrated in 80–90% of PV patients by indirect immunofluorescence microscopy; monkey esophagus is the optimal substrate for these studies. Patients with PV have IgG autoantibodies to *desmogleins* (Dsgs), transmembrane desmosomal glycoproteins that belong to the cadherin family of calcium-dependent adhesion molecules. Such autoantibodies can be precisely quantitated by enzyme-linked immunosorbent assay (ELISA). Patients with early PV (i.e., mucosal disease) have IgG autoantibodies to Dsg3; patients with advanced PV (i.e., mucocutaneous disease) have IgG autoantibodies to both Dsg3 and Dsg1. Experimental studies have shown that autoantibodies from patients with PV are pathogenic (i.e., responsible for blister formation) and that their titer correlates with disease activity.

TABLE 62-1 Immunologically Mediated Blistering Diseases

DISEASE	CLINICAL MANIFESTATIONS	HISTOLOGY	IMMUNOPATHOLOGY	AUTOANTIGENS[a]
Pemphigus vulgaris	Oromucosal lesions, flaccid blisters, denuded skin	Acantholytic blister formed in suprabasal layer of epidermis	Cell surface deposits of IgG on keratinocytes	Dsg3 (plus Dsg1 in patients with skin involvement)
Pemphigus foliaceus	Crusts and shallow erosions on scalp, central face, upper chest, and back	Acantholytic blister formed in superficial layer of epidermis	Cell surface deposits of IgG on keratinocytes	Dsg1
Paraneoplastic pemphigus	Painful stomatitis with papulosquamous or lichenoid eruptions that may progress to blisters	Acantholysis, keratinocyte necrosis, and vacuolar interface dermatitis	Cell surface deposits of IgG and C3 on keratinocytes and (variably) similar immunoreactants in epidermal BMZ	Plakin protein family members and desmosomal cadherins (see text for details)
Bullous pemphigoid	Large tense blisters on flexor surfaces and trunk	Subepidermal blister with eosinophil-rich infiltrate	Linear band of IgG and/or C3 in epidermal BMZ	BPAG1, BPAG2
Pemphigoid gestationis	Pruritic, urticarial plaques rimmed by vesicles and bullae on the trunk and extremities	Teardrop-shaped, subepidermal blisters in dermal papillae; eosinophil-rich infiltrate	Linear band of C3 in epidermal BMZ	BPAG2 (plus BPAG1 in some patients)
Dermatitis herpetiformis	Extremely pruritic small papules and vesicles on elbows, knees, buttocks, and posterior neck	Subepidermal blister with neutrophils in dermal papillae	Granular deposits of IgA in dermal papillae	Epidermal transglutaminase
Linear IgA disease	Pruritic papulovesicles on extensor surfaces; occasionally larger, arciform blisters	Subepidermal blister with neutrophil-rich infiltrate	Linear band of IgA in epidermal BMZ	BPAG2 (see text for specific details)
Epidermolysis bullosa acquisita	Blisters, erosions, scars, and milia on sites exposed to trauma; widespread, inflammatory, tense blisters may be seen initially	Subepidermal blister that may or may not include a leukocytic infiltrate	Linear band of IgG and/or C3 in epidermal BMZ	Type VII collagen
Mucous membrane pemphigoid	Erosive and/or blistering lesions of mucous membranes and possibly the skin; scarring of some sites	Subepidermal blister that may or may not include a leukocytic infiltrate	Linear band of IgG, IgA, and/or C3 in epidermal BMZ	BPAG2, laminin-332, or others

[a] Autoantigens bound by these patients' autoantibodies are defined as follows: Dsg1, desmoglein 1; Dsg3, desmoglein 3; BPAG1, bullous pemphigoid antigen 1; BPAG2, bullous pemphigoid antigen 2.

Abbreviation: BMZ, basement membrane zone.

Recent studies have shown that the anti-Dsg autoantibody profile in these patients' sera as well as the tissue distribution of Dsg3 and Dsg1 determine the site of blister formation in patients with PV. Coexpression of Dsg3 and Dsg1 by epidermal cells protects against pathogenic IgG antibodies to either of these cadherins but not against pathogenic autoantibodies to both.

PV can be life-threatening. Prior to the availability of glucocorticoids, mortality rates ranged from 60% to 90%; the current figure is ~5%. Common causes of morbidity and death are infection and complications of treatment. Bad prognostic factors include advanced age, widespread involvement, and the requirement for high doses of glucocorticoids (with or without other immunosuppressive agents) for control of disease. The course of PV in individual patients is variable and difficult to predict. Some patients experience remission, while others may require long-term treatment or succumb to complications of their disease or its treatment. The mainstay of treatment is systemic glucocorticoids alone or in combination with other immunosuppressive agents. Patients with moderate to severe PV are usually started on prednisone at doses ≤1 mg/kg per day (single morning dose). If new lesions continue to appear after 1–2 weeks of treatment, the dose of prednisone may need to be increased and/or combined with another immunosuppressive agent. Among these, rituximab in combination with prednisone often achieves remission (though maintenance therapy may be required to prevent relapse). Other immunosuppressive agents sometimes combined with prednisone to treat PV include azathioprine, mycophenolate mofetil, or cyclophosphamide. Patients with severe, treatment-resistant disease may derive benefit from plasmapheresis (six high-volume exchanges [i.e., 2–3 L per exchange] over ~2 weeks) and/or IV immunoglobulin (IVIg). It is important to bring severe or progressive disease under control quickly in order to lessen the severity and/or duration of this disorder. Increasingly, rituximab and daily glucocorticoids are used early in PV patients to avert the development of advanced and/or treatment-resistant disease.

PEMPHIGUS FOLIACEUS

Pemphigus foliaceus (PF) is distinguished from PV by several features. In PF, acantholytic blisters are located high within the epidermis, usually just beneath the stratum corneum. Hence, PF is a more superficial blistering disease than PV. The distribution of lesions in the two disorders is much the same, except that in PF mucous membranes are almost always spared. Patients with PF rarely have intact blisters but rather exhibit shallow erosions associated with erythema, scale, and crust formation. Mild cases of PF can resemble severe seborrheic dermatitis; severe PF may cause extensive exfoliation. Sun exposure (ultraviolet irradiation) may be an aggravating factor.

PF has immunopathologic features in common with PV. Specifically, direct immunofluorescence microscopy of perilesional skin demonstrates IgG on the surface of keratinocytes. Similarly, patients with PF have circulating IgG autoantibodies directed against the surface of keratinocytes. In PF, autoantibodies are directed against Dsg1, a 160-kDa desmosomal cadherin. These autoantibodies can be quantitated by ELISA. As noted for PV, the autoantibody profile in patients with PF (i.e., anti-Dsg1 IgG) and the tissue distribution of this autoantigen (i.e., expression in oral mucosa that is compensated by coexpression of Dsg3) are thought to account for the distribution of lesions in this disease.

Endemic forms of PF are found in south-central rural Brazil, where the disease is known as *fogo salvagem* (FS), as well as in selected sites in Latin America and Tunisia. Endemic PF, like other forms of this disease, is mediated by IgG autoantibodies to Dsg1. Clusters of FS overlap with those of leishmaniasis, a disease transmitted by bites of the sand fly *Lutzomyia longipalis*. Studies have shown that sand fly salivary antigens (specifically, the LJM11 salivary protein) are recognized by IgG autoantibodies from FS patients (as well as by monoclonal antibodies to Dsg1 derived from these patients). The demonstration that mice immunized with LJM11 produce antibodies to Dsg1 suggests that insect bites may deliver salivary antigens, initiate a cross-reactive humoral immune response, and lead to FS in genetically susceptible individuals.

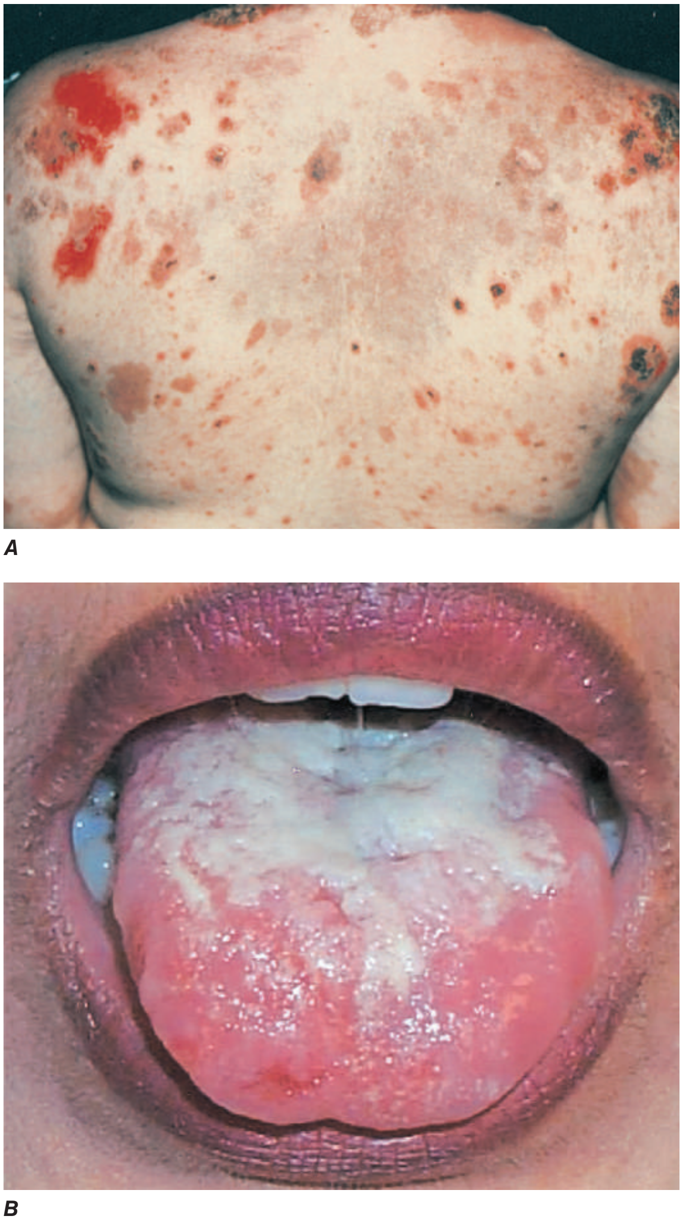

FIGURE 62-1 Pemphigus vulgaris. ***A.*** Flaccid bullae are easily ruptured, resulting in multiple erosions and crusted plaques. ***B.*** Involvement of the oral mucosa, which is almost invariable, may present with erosions on the gingiva, buccal mucosa, palate, posterior pharynx, or tongue. *(Courtesy of Robert Swerlick, MD; with permission.)*

Although pemphigus has been associated with several autoimmune diseases, its association with thymoma and/or myasthenia gravis is particularly notable. To date, >30 cases of thymoma and/or myasthenia gravis have been reported in association with pemphigus, usually with PF. Patients may also develop pemphigus as a consequence of drug exposure; drug-induced pemphigus usually resembles PF rather than PV. Drugs containing a thiol group in their chemical structure (e.g., penicillamine, captopril, enalapril) are most commonly associated with drug-induced pemphigus. Nonthiol drugs linked to pemphigus include penicillins, cephalosporins, and piroxicam. Some cases of drug-induced pemphigus are durable and require treatment with systemic glucocorticoids and/or immunosuppressive agents.

PF is generally a less severe disease than PV and usually carries a better prognosis. Localized disease can sometimes be treated with topical or intralesional glucocorticoids; more active cases can usually be controlled with systemic glucocorticoids either alone or in combination with other immunosuppressive agents. Patients with severe, treatment-resistant disease may require more aggressive interventions, as described above for patients with PV.

PARANEOPLASTIC PEMPHIGUS

Paraneoplastic pemphigus (PNP) is an autoimmune acantholytic mucocutaneous disease associated with an occult or confirmed neoplasm. Patients with PNP typically have painful stomatitis in association with papulosquamous and/or lichenoid eruptions that often progress to blisters. Palm and sole involvement are common in these patients and raise the possibility that prior reports of neoplasia-associated erythema multiforme may have represented unrecognized cases of PNP. Biopsies of lesional skin from these patients show varying combinations of acantholysis, keratinocyte necrosis, and vacuolar-interface dermatitis. Direct immunofluorescence microscopy of a patient's skin shows deposits of IgG and complement on the surface of keratinocytes and (variably) similar immunoreactants in the epidermal basement membrane zone. Patients with PNP have IgG autoantibodies to cytoplasmic proteins that are members of the plakin family (e.g., desmoplakins I and II, bullous pemphigoid antigen [BPAG] 1, envoplakin, periplakin, and plectin) and to other proteins (e.g., Dsg1, Dsg3, and α_2-macroglobulin like-1). Passive transfer studies have shown that autoantibodies from patients with PNP are pathogenic in animal models.

The predominant neoplasms associated with PNP are non-Hodgkin's lymphoma, chronic lymphocytic leukemia, thymoma, spindle cell tumors, Waldenström's macroglobulinemia, and Castleman's disease; the last-mentioned neoplasm is particularly common among children with PNP. Rare cases of seronegative PNP have been reported in patients with B-cell malignancies previously treated with rituximab. In addition to severe skin lesions, many patients with PNP develop life-threatening bronchiolitis obliterans. PNP is generally resistant to conventional therapies (i.e., those used to treat PV); rarely, a patient's disease may ameliorate or even remit following ablation or removal of underlying neoplasms.

BULLOUS PEMPHIGOID

Bullous pemphigoid (BP) is a polymorphic autoimmune subepidermal blistering disease usually seen in the elderly. Initial lesions may consist of urticarial plaques; most patients eventually display tense vesicles or blisters on either normal-appearing or erythematous skin **(Fig. 62-2)**. The lesions are usually distributed over the lower abdomen, groin, and flexor surface of the extremities; oral mucosal lesions are found in some patients. Pruritus may be nonexistent or severe. As lesions evolve, tense blisters tend to rupture and be replaced by erosions with or without surmounting crust. Nontraumatized blisters heal without scarring.

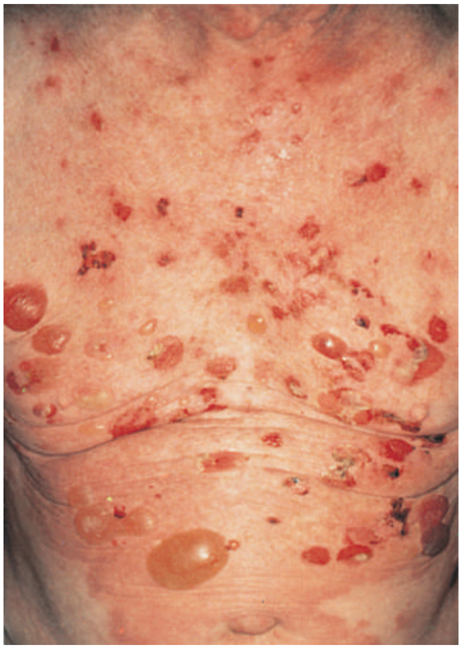

FIGURE 62-2 Bullous pemphigoid with tense vesicles and bullae on erythematous, urticarial bases. *(Courtesy of the Yale Resident's Slide Collection; with permission.)*

The major histocompatibility complex class II allele HLA-DQβ1*0301 is prevalent in patients with BP. Though most cases occur sporadically, BP can be triggered by medications (e.g., furosemide, dipeptidyl peptidase-4 inhibitors, immune checkpoint inhibitors), ultraviolet light, or ionizing radiation. Several studies have shown that BP is associated with neurologic diseases (e.g., stroke, dementia, Parkinson's disease, and multiple sclerosis).

Biopsies of early lesional skin demonstrate subepidermal blisters and histologic features that roughly correlate with the clinical character of the lesion under study. Lesions on normal-appearing skin generally contain a sparse perivascular leukocytic infiltrate with some eosinophils; conversely, biopsies of inflammatory lesions typically show an eosinophil-rich infiltrate at sites of vesicle formation and in perivascular areas. In addition to eosinophils, cell-rich lesions also contain mononuclear cells and neutrophils. It is not possible to distinguish BP from other subepidermal blistering diseases by routine histologic studies alone.

Direct immunofluorescence microscopy of normal-appearing perilesional skin from patients with BP shows linear deposits of IgG and/or C3 in the epidermal basement membrane. The sera of ~70% of these patients contain circulating IgG autoantibodies that bind the epidermal basement membrane of normal human skin in indirect immunofluorescence microscopy. IgG from an even higher percentage of patients reacts with the epidermal side of 1 *M* NaCl split skin (an alternative immunofluorescence microscopy test substrate used to distinguish circulating IgG autoantibodies to the basement membrane in patients with BP from those in patients with similar, yet different, subepidermal blistering diseases; see below). In BP, circulating autoantibodies recognize 230- and 180-kDa hemidesmosome-associated proteins in basal keratinocytes (i.e., BPAG1 and BPAG2, respectively). Autoantibodies to BPAG2 are thought to deposit in situ, activate complement, produce dermal mast-cell degranulation, and generate granulocyte-rich infiltrates that cause tissue damage and blister formation.

BP may persist for months to years, with exacerbations or remissions. Extensive involvement may result in widespread erosions and compromise cutaneous integrity; elderly and/or debilitated patients may die. Local or minimal disease can sometimes be controlled with potent topical glucocorticoids alone; more extensive lesions generally respond to systemic glucocorticoids either alone or in combination with other agents. Adjuncts to systemic glucocorticoids include doxycycline, azathioprine, mycophenolate mofetil, and rituximab.

PEMPHIGOID GESTATIONIS

Pemphigoid gestationis (PG), also known as *herpes gestationis*, is a rare, nonviral, subepidermal blistering disease of pregnancy and the puerperium. PG may begin during any trimester of pregnancy or present shortly after delivery. Lesions are usually distributed over the abdomen, trunk, and extremities; mucous membrane lesions are rare. Skin lesions in these patients may be quite polymorphic and consist of erythematous urticarial papules and plaques, vesiculopapules, and/or frank bullae. Lesions are almost always extremely pruritic. Severe exacerbations of PG frequently follow delivery, typically within 24–48 h. PG tends to recur in subsequent pregnancies, often beginning earlier during such gestations. Brief flare-ups of disease may occur with resumption of menses and may develop in patients later exposed to oral contraceptives. Occasionally, infants of affected mothers have transient skin lesions.

Biopsies of early lesional skin show teardrop-shaped subepidermal vesicles forming in dermal papillae in association with an eosinophil-rich leukocytic infiltrate. Differentiation of PG from other subepidermal bullous diseases by light microscopy is difficult. However, direct immunofluorescence microscopy of perilesional skin from PG patients reveals the immunopathologic hallmark of this disorder: linear deposits of C3 in the epidermal basement membrane. These deposits develop as a consequence of complement activation produced by low-titer IgG anti–basement membrane autoantibodies directed against BPAG2, the same hemidesmosome-associated protein that is targeted by autoantibodies in patients with BP—a subepidermal bullous disease that resembles PG clinically, histologically, and immunopathologically.

The goals of therapy in patients with PG are to prevent the development of new lesions, relieve intense pruritus, and care for erosions at sites of blister formation. Many patients require treatment with moderate doses of daily glucocorticoids (i.e., 20–40 mg of prednisone) at some point in their course. Mild cases (or brief flare-ups) may be controlled by vigorous use of potent topical glucocorticoids. Infants born of mothers with PG appear to be at increased risk of being born slightly premature or "small for dates." Current evidence suggests that there is no difference in the incidence of uncomplicated live births between PG patients treated with systemic glucocorticoids and those managed more conservatively. If systemic glucocorticoids are administered, newborns are at risk for development of reversible adrenal insufficiency.

DERMATITIS HERPETIFORMIS

Dermatitis herpetiformis (DH) is an intensely pruritic, papulovesicular skin disease characterized by lesions symmetrically distributed over extensor surfaces (i.e., elbows, knees, buttocks, back, scalp, and posterior neck) **(see Fig. 59-8)**. Primary lesions in this disorder consist of papules, papulovesicles, or urticarial plaques. Because pruritus is prominent, patients may present with excoriations and crusted papules but no observable primary lesions. Patients sometimes report that their pruritus has a distinctive burning or stinging component; the onset of such local symptoms reliably heralds the development of distinct clinical lesions 12–24 h later. Almost all DH patients have associated, usually subclinical, gluten-sensitive enteropathy **(Chap. 336)**, and >90% express the HLA-B8/DRw3 and HLA-DQw2 haplotypes. DH may present at any age, including in childhood; onset in the second to fourth decades is most common. The disease is typically chronic.

Biopsy of early lesional skin reveals neutrophil-rich infiltrates within dermal papillae. Neutrophils, fibrin, edema, and microvesicle formation at these sites are characteristic of early disease. Older lesions may demonstrate nonspecific features of a subepidermal bulla or an excoriated papule. Because the clinical and histologic features of this disease can be variable and resemble those of other subepidermal blistering disorders, the diagnosis is confirmed by direct immunofluorescence microscopy of normal-appearing perilesional skin. Such studies demonstrate granular deposits of IgA (with or without complement components) in the papillary dermis and along the epidermal basement membrane zone. IgA deposits in the skin are unaffected by control of disease with medication; however, these immunoreactants diminish in intensity or disappear in patients maintained for long periods on a strict gluten-free diet (see below). Patients with DH have granular deposits of IgA in their epidermal basement membrane zone and should be distinguished from individuals with linear IgA deposits at this site (see below).

Although most DH patients do not report overt gastrointestinal symptoms or have laboratory evidence of malabsorption, biopsies of the small bowel usually reveal blunting of intestinal villi and a lymphocytic infiltrate in the lamina propria. As is true for patients with celiac disease, this gastrointestinal abnormality can be reversed by a gluten-free diet. Moreover, if maintained, this diet alone may control the skin disease and eventuate in clearance of IgA deposits from these patients' epidermal basement membrane zones. Subsequent gluten exposure in such patients alters the morphology of their small bowel, elicits a flare-up of their skin disease, and is associated with the reappearance of IgA in their epidermal basement membrane zones. As in patients with celiac disease, dietary gluten sensitivity in patients with DH is associated with IgA anti-endomysial autoantibodies that target tissue transglutaminase. Studies indicate that patients with DH also have high-avidity IgA autoantibodies to epidermal transglutaminase and that the latter is co-localized with granular deposits of IgA in the papillary dermis of DH patients. Patients with DH also have an increased incidence of thyroid abnormalities, achlorhydria, atrophic gastritis, and autoantibodies to gastric parietal cells. These associations likely relate to the high frequency of the HLA-B8/DRw3 haplotype in these patients, since this marker is commonly linked to autoimmune disorders. The mainstay of treatment of DH is dapsone, a sulfone. Patients respond rapidly (24–48 h) to dapsone but require careful pretreatment evaluation (e.g., screening for glucose-6-phosphate dehydrogenase deficiency) and close follow-up to ensure that complications are avoided or controlled. All patients taking dapsone at ≥100 mg/d will have some hemolysis and methemoglobinemia, which are expected pharmacologic side effects of this agent. Gluten restriction can control DH and lessen dapsone requirements; this diet must rigidly exclude gluten to be of maximal benefit. Many months of dietary restriction may be necessary before a beneficial result is achieved. Good dietary counseling by a trained dietitian is essential.

LINEAR IGA DISEASE

Linear IgA disease, once considered a variant form of DH, is actually a separate and distinct entity. Clinically, patients with linear IgA disease may resemble individuals with DH, BP, or other subepidermal blistering diseases. Lesions typically consist of papulovesicles, bullae, and/or urticarial plaques that develop predominantly on central or flexural sites. Oral mucosal involvement occurs in some patients. Severe pruritus resembles that seen in patients with DH. Patients with linear IgA disease do not have an increased frequency of the HLA-B8/DRw3 haplotype or an associated enteropathy and therefore are not candidates for treatment with a gluten-free diet.

Histologic alterations in early lesions may be virtually indistinguishable from those in DH. However, direct immunofluorescence microscopy of normal-appearing perilesional skin reveals a linear band of IgA (and often C3) in the epidermal basement membrane zone. Most patients with linear IgA disease have circulating IgA anti-basement membrane autoantibodies directed against neoepitopes in the proteolytically processed extracellular domain of BPAG2. These patients generally respond to treatment with dapsone (50–200 mg/d) alone or in combination with low daily doses of prednisone.

EPIDERMOLYSIS BULLOSA ACQUISITA

Epidermolysis bullosa acquisita (EBA) is a rare, noninherited, polymorphic, chronic, subepidermal blistering disease. (The inherited form is discussed in **Chap. 425**.) Patients with classic or noninflammatory EBA have blisters on noninflamed skin, atrophic scars, milia, nail dystrophy, hair loss, and oral lesions. Because lesions generally occur at sites exposed to minor trauma, classic EBA is considered a mechanobullous disease. Other patients with EBA have widespread inflammatory scarring and bullous lesions that resemble severe BP. Inflammatory EBA may evolve into the classic, noninflammatory form of this disease. Rarely, patients present with lesions that predominate on mucous membranes. The HLA-DR2 haplotype is found with increased frequency in EBA patients. Studies suggest that EBA is sometimes associated with inflammatory bowel disease (especially Crohn's disease).

The histology of lesional skin varies with the character of the lesion being studied. Noninflammatory bullae are subepidermal, feature a sparse leukocytic infiltrate, and resemble the lesions in patients with porphyria cutanea tarda. Inflammatory lesions consist of neutrophil-rich subepidermal blisters. EBA patients have continuous deposits of IgG (and frequently C3) in a linear pattern within the epidermal basement membrane zone. Ultrastructurally, these immunoreactants are found in the sublamina densa region in association with anchoring fibrils. Approximately 50% of EBA patients have demonstrable circulating IgG anti-basement membrane autoantibodies directed against type VII collagen—the collagen species that makes up anchoring fibrils. Such IgG autoantibodies bind the dermal side of 1 *M* NaCl split skin (in contrast to IgG autoantibodies in patients with BP). Studies have shown that passive transfer of experimental or patient IgG against type VII collagen can produce lesions in mice that clinically, histologically, and immunopathologically resemble those in patients with EBA.

Treatment of EBA is generally unsatisfactory. Some patients with inflammatory EBA may respond to systemic glucocorticoids, either alone or in combination with immunosuppressive agents. Other patients (especially those with neutrophil-rich inflammatory lesions) may respond to dapsone. The chronic, noninflammatory form of EBA is largely resistant to treatment, although some patients may respond to prednisone in combination with rituximab, cyclosporine, mycophenolate mofetil, azathioprine, or IVIg.

MUCOUS MEMBRANE PEMPHIGOID

Mucous membrane pemphigoid (MMP) is a rare, acquired, subepithelial immunobullous disease characterized by erosive lesions of mucous membranes and skin that result in scarring of at least some sites of involvement. Common sites include the oral mucosa (especially the gingiva) and conjunctiva; other sites that may be affected include the nasopharyngeal, laryngeal, esophageal, and anogenital mucosa. Skin lesions (present in about one-third of patients) tend to predominate on the scalp, face, and upper trunk and generally consist of a few scattered erosions or tense blisters on an erythematous or urticarial base. MMP is typically a chronic and progressive disorder. Serious complications may arise as a consequence of ocular, laryngeal, esophageal, or anogenital lesions. Erosive conjunctivitis may result in shortened fornices, symblepharon, ankyloblepharon, entropion, corneal opacities, and (in severe cases) blindness. Similarly, erosive lesions of the larynx may cause hoarseness, pain, and tissue loss that, if unrecognized and untreated, may eventuate in complete destruction of the airway. Esophageal lesions may result in stenosis and/or strictures that could place patients at risk for aspiration. Strictures may also complicate anogenital involvement.

Biopsies of lesional tissue generally show subepithelial vesiculobullae and a mononuclear leukocytic infiltrate. Neutrophils and eosinophils may be seen in biopsies of early lesions; older lesions may demonstrate a scant leukocytic infiltrate and fibrosis. Direct immunofluorescence microscopy of perilesional tissue typically reveals deposits of IgG, IgA, and/or C3 in the epidermal basement membrane. Because many patients with MMP exhibit no evidence of circulating anti-basement membrane autoantibodies, testing of perilesional skin is important diagnostically. Although MMP was once thought to be a single nosologic entity, it is now largely regarded as a disease phenotype that may develop as a consequence of an autoimmune reaction to a variety of molecules in the epidermal basement membrane (e.g., BPAG2, laminin-332, type VII collagen, $\alpha_6\beta_4$ integrin) and other antigens yet to be completely defined. Studies suggest that MMP patients with autoantibodies to laminin-332 have an increased relative risk for cancer. Treatment of MMP is largely dependent upon the sites of involvement. Due to potentially severe complications, patients with ocular, laryngeal, esophageal, and/or anogenital involvement require aggressive systemic treatment with dapsone, prednisone, or the latter in combination with other immunosuppressive agents (e.g., rituximab, azathioprine, mycophenolate mofetil, or cyclophosphamide), or IVIg. Less threatening forms of the disease may be managed with topical or intralesional glucocorticoids.

AUTOIMMUNE SYSTEMIC DISEASES WITH PROMINENT CUTANEOUS FEATURES

DERMATOMYOSITIS

The cutaneous manifestations of dermatomyositis **(Chap. 377)** are often distinctive but at times may resemble those of systemic lupus erythematosus (SLE) **(Chap. 368)**, scleroderma **(Chap. 372)**, or other overlapping connective tissue diseases **(Chap. 372)**. The extent and severity of cutaneous disease may or may not correlate with the extent and severity of the myositis. The cutaneous manifestations of dermatomyositis are similar, whether the disease appears in children or in the elderly, except that calcification of subcutaneous tissue is a common late sequela in childhood dermatomyositis. Dermatomyositis may be associated with interstitial lung disease or cancer.

The cutaneous signs of dermatomyositis may precede or follow the development of myositis by weeks to years. Cases lacking muscle involvement (i.e., *dermatomyositis sine myositis* or *amyopathic dermatomyositis*) have been reported. The most common manifestation is a purple-red discoloration of the upper eyelids, sometimes associated with scaling ("heliotrope" erythema; **Fig. 62-3**) and periorbital edema. Erythema on the cheeks and nose in a "butterfly" distribution may resemble the malar eruption of SLE. Erythematous or violaceous thin,

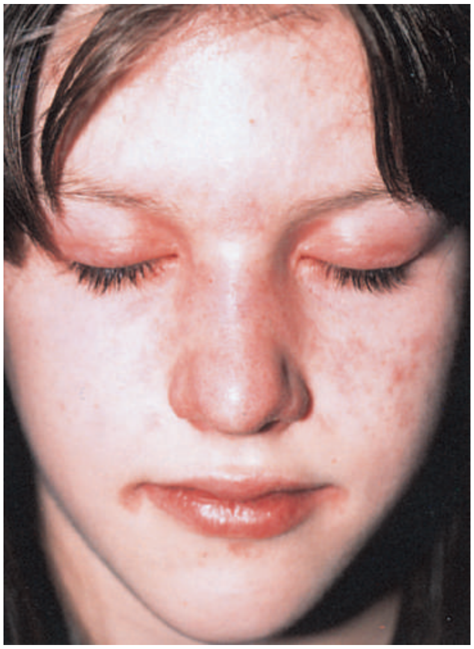

FIGURE 62-3 Dermatomyositis. Periorbital violaceous erythema characterizes the classic heliotrope rash. *(Courtesy of James Krell, MD; with permission.)*

scaly plaques are common on the upper trunk and neck (shawl sign), the scalp, lateral aspects of the thighs (holster sign), and the extensor surfaces of the forearms and hands (tendon streaking). Approximately one-third of patients have violaceous, flat-topped papules over the dorsal interphalangeal joints that are pathognomonic of dermatomyositis (Gottron's papules) **(Fig. 62-4)**. Thin violaceous papules and plaques on the elbows and knees of patients with dermatomyositis are referred to as *Gottron's sign*. These lesions can be contrasted with the erythema and scaling on the dorsum of the fingers that spares the skin over the interphalangeal joints of some SLE patients. Periungual telangiectasias and edema may be prominent in patients with dermatomyositis. Other patients, particularly those with long-standing disease, develop areas of hypopigmentation, hyperpigmentation, mild atrophy, and telangiectasia known as *poikiloderma*. Poikiloderma is rare in both SLE and scleroderma and thus can serve as a clinical sign that distinguishes dermatomyositis from these two diseases. Cutaneous changes may be similar in dermatomyositis and various overlap syndromes where thickening and binding down of the skin of the hands (*sclerodactyly*) as well as Raynaud's phenomenon can be seen. However, the presence of severe muscle disease, Gottron's papules, heliotrope erythema, and poikiloderma serves to distinguish patients with dermatomyositis. Skin biopsy of the erythematous, scaling lesions of dermatomyositis may reveal only mild nonspecific inflammation, but sometimes may show changes indistinguishable from those found in cutaneous lupus erythematosus (LE), including epidermal atrophy, hydropic degeneration of basal keratinocytes, and dermal changes consisting of interstitial mucin deposition and a mild mononuclear cell perivascular infiltrate. Direct immunofluorescence microscopy of lesional skin is usually negative, although granular deposits of immunoglobulin(s) and complement in the epidermal basement membrane zone have been described in some patients. Treatment should be stratified based on the relative severity of disease. Topical treatments include glucocorticoids, sunscreens, and aggressive photoprotective measures. Treatment of systemic disease includes antimalarials (though some patients may develop a drug eruption upon initiation of therapy) or systemic glucocorticoids in conjunction with methotrexate, mycophenolate mofetil, azathioprine, rituximab, or IVIg.

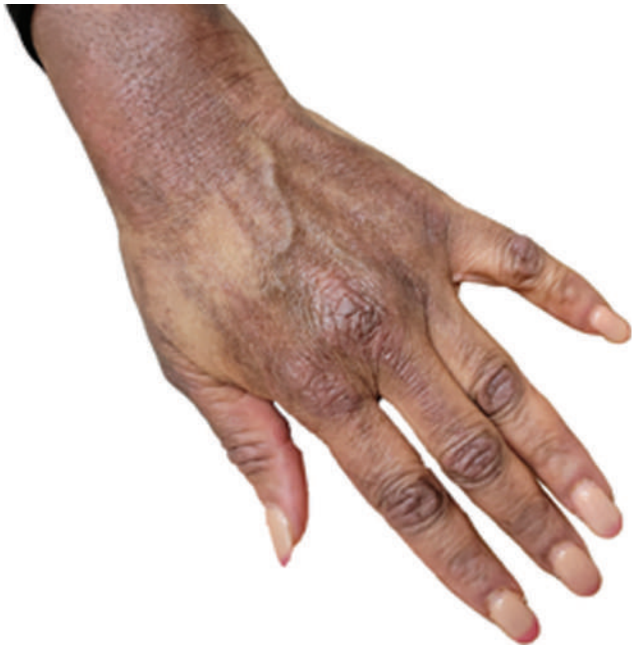

FIGURE 62-4 Dermatomyositis showing involvement of the hands with erythematous flat-topped papules over the joints (i.e., Gottron's papules) as well as periungual erythema and telangiectasias. *(Courtesy of Justin Cheeley, MD; with permission.)*

LUPUS ERYTHEMATOSUS

The cutaneous manifestations of LE **(Chap. 368)** can be divided into acute, subacute, and chronic types. *Acute cutaneous LE* is characterized by erythema of the nose and malar eminences in a "butterfly" distribution **(Fig. 62-5*A*)**. The erythema is often sudden in onset, accompanied by edema and fine scale, and correlated with systemic involvement. Patients may have widespread involvement of the face as well as erythema and scaling of the extensor surfaces of the extremities and upper chest **(Fig. 62-5*B*)**. These acute lesions, while sometimes evanescent, usually last for days and are often associated with exacerbations of systemic disease. Skin biopsy of acute lesions typically shows hydropic degeneration of basal keratinocytes, dermal edema, and (in some cases) a sparse perivascular infiltrate of mononuclear cells in the upper dermis as well as dermal mucin. Direct immunofluorescence microscopy of lesional skin frequently reveals deposits of immunoglobulin(s) and complement in the epidermal basement membrane zone. Treatment of cutaneous disease includes topical glucocorticoids, aggressive photoprotection, antimalarials, and control of systemic disease. Treatment of systemic disease associated with acute cutaneous LE includes systemic glucocorticoids in conjunction with antimalarials and other immunosuppressive agents.

Subacute cutaneous lupus erythematosus (SCLE) is characterized by a widespread photosensitive, nonscarring eruption. In most patients, renal and central nervous system involvement is mild or absent. SCLE may present as a papulosquamous eruption that resembles psoriasis or as annular polycyclic lesions. In the papulosquamous form, discrete erythematous papules arise on the upper back, upper chest, shoulders, extensor surfaces of the arms, and dorsum of the hands; lesions are uncommon on the central face and the flexor surfaces of the arms as well as below the waist. These slightly scaling papules tend to merge into plaques. The annular form involves the same areas and presents with erythematous papules that evolve into oval, circular, or polycyclic lesions. The lesions of SCLE are more widespread than those in patients with discoid LE and have less tendency for scarring. In many patients with SCLE, drugs (e.g., hydrochlorothiazide, calcium channel blockers, antifungals, proton pump inhibitors) may induce or exacerbate disease. Skin biopsy typically reveals epidermal changes that include atrophy, hydropic degeneration of basal keratinocytes, and apoptosis accompanied by an infiltrate of mononuclear cells in the upper dermis. Direct immunofluorescence microscopy of lesional skin reveals deposits of immunoglobulin(s) in the epidermal basement membrane zone in about one-half of these cases. A particulate pattern of IgG deposition throughout the epidermis has been associated with SCLE. Most SCLE patients have anti-Ro autoantibodies. Local therapy alone is usually unsuccessful. Most patients require treatment with

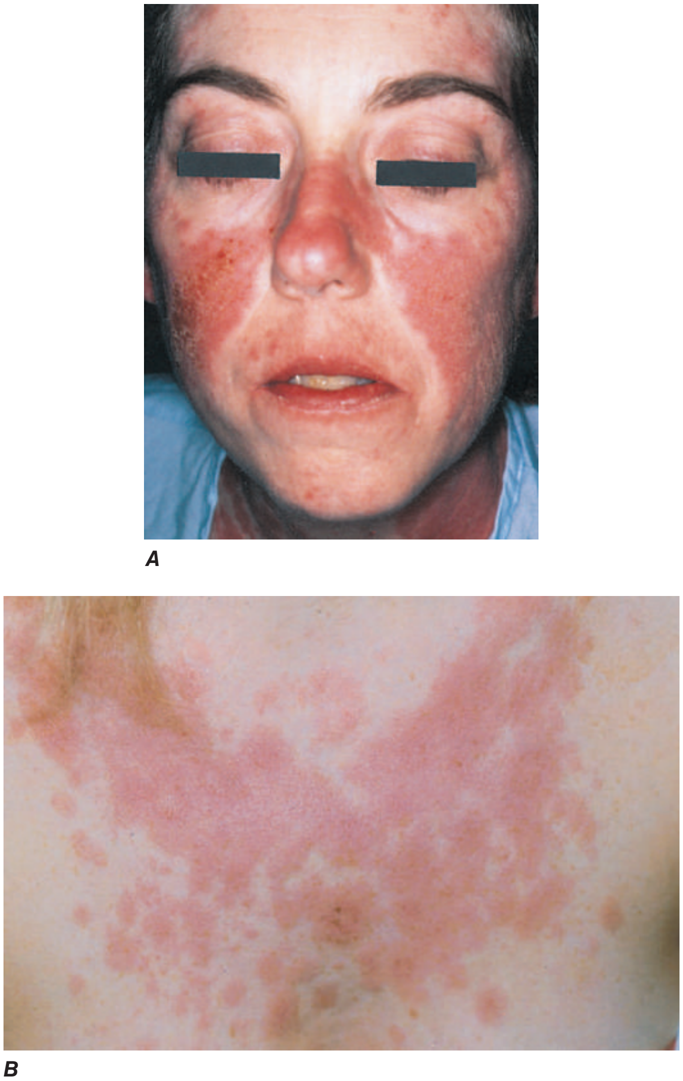

FIGURE 62-5 Acute cutaneous lupus erythematosus (LE). ***A.*** Acute cutaneous LE on the face, showing prominent, scaly, malar erythema. Involvement of other sun-exposed sites is also common. ***B.*** Acute cutaneous LE on the upper chest, demonstrating brightly erythematous and slightly edematous papules and plaques. *(Courtesy of Robert Swerlick, MD; with permission.)*

aminoquinoline antimalarial drugs. Low-dose therapy with oral glucocorticoids is sometimes necessary. Photoprotective measures against both ultraviolet B and ultraviolet A wavelengths are very important.

Chronic cutaneous LE has multiple subtypes; *discoid LE* (DLE) is the most common. DLE is characterized by discrete lesions, most often found on the face, scalp, and/or external ears. It is unusual to see DLE lesions below the neck without lesions above the neck. The lesions are erythematous papules or plaques with a thick, adherent scale that occludes hair follicles (follicular plugging). When the scale is removed, its underside shows small excrescences that correlate with the openings of hair follicles (so-called "carpet tacking"), a finding relatively specific for DLE. Long-standing lesions develop central atrophy, scarring, and hypopigmentation but frequently have erythematous, sometimes raised borders **(Fig. 62-6)**. These lesions persist for years and tend to expand slowly. Up to 20% of patients with DLE eventually meet the American College of Rheumatology criteria for SLE, but lower percentages of DLE patients develop SLE under the Systemic Lupus International Collaborating Clinics classification criteria and European League Against Rheumatism and American College of Rheumatology classification criteria. Typical discoid lesions are frequently seen in patients with SLE. Biopsy of DLE lesions shows hyperkeratosis, follicular plugging, atrophy of the epidermis, hydropic degeneration of basal keratinocytes, thickening of the epidermal basement membrane zone, dermal mucin, and a mononuclear cell infiltrate adjacent to epidermal, adnexal, and microvascular basement membranes. Direct immunofluorescence microscopy demonstrates immunoglobulin(s) and complement deposits at the basement membrane zone in ~90% of cases. Treatment is focused on control of local cutaneous disease and consists mainly of photoprotection and topical or intralesional glucocorticoids. If local therapy is ineffective, use of aminoquinoline antimalarial agents may be indicated.

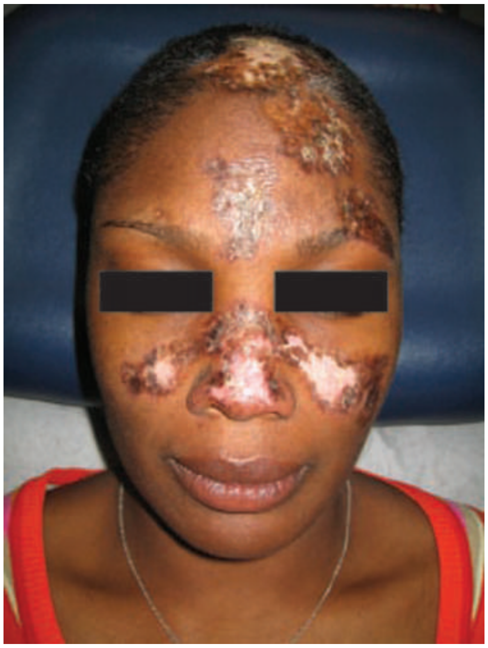

FIGURE 62-6 Discoid lupus erythematosus (DLE). Erythematous to violaceous, dyspigmented, atrophic plaques with scarring of the face and scalp. *(Courtesy of Robert Swerlick, MD; with permission.)*

■ SCLERODERMA AND MORPHEA

The skin changes of scleroderma **(Chap. 372)** may be limited or diffuse. In both instances, disease usually begins on the fingers, hands, toes, feet, and face, with episodes of recurrent nonpitting edema. Sclerosis of the skin commences distally on the fingers (sclerodactyly) and spreads proximally, usually accompanied by resorption of bone of the fingertips, which may have punched out ulcers, stellate scars, or areas of hemorrhage **(Fig. 62-7)**. The fingers may shrink and become sausage-shaped, and, because the fingernails are usually unaffected, they may curve over the end of the fingertips. Periungual telangiectasias are usually present, but periungual erythema is rare. In diffuse disease, the extremities show contractures and calcinosis cutis; facial involvement includes a smooth, unwrinkled brow, taut skin over the nose, shrinkage of tissue around the mouth, and perioral radial

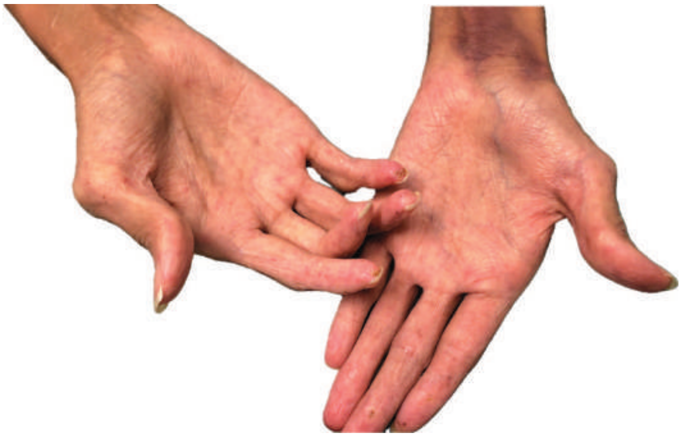

FIGURE 62-7 Scleroderma showing acral sclerosis, matlike telangiectasias, and focal digital ulcers. *(Courtesy of Justin Cheeley, MD; with permission.)*

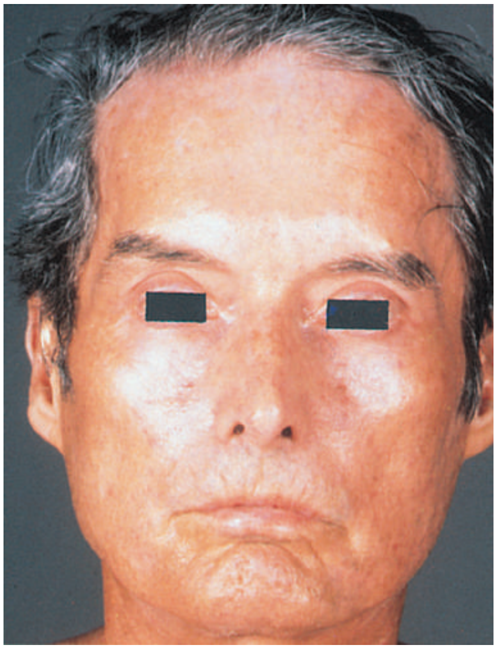

FIGURE 62-8 Development of an expressionless, masklike facies in a patient with scleroderma. *(Courtesy of Thomas J. Lawley, MD; with permission.)*

furrowing **(Fig. 62-8)**. Matlike telangiectasias are often present, particularly on the face and hands. Involved skin feels indurated, smooth, and bound to underlying structures; hyper- and hypopigmentation are common as well. *Raynaud's phenomenon* (i.e., cold-induced blanching, cyanosis, and reactive hyperemia) is documented in almost all patients and can precede development of scleroderma by many years. The combination of calcinosis cutis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasias has been termed as the *CREST syndrome*. Anti-centromere autoantibodies have been reported in a very high percentage of patients with CREST syndrome but in only a small minority of patients with scleroderma. Skin biopsy reveals thickening of the dermis, homogenization of collagen bundles, atrophic pilosebaceous and eccrine glands, and a sparse mononuclear cell infiltrate in the dermis and subcutaneous fat. Direct immunofluorescence microscopy of lesional skin is usually negative. Treatments for cutaneous disease include emollients, antipruritics, and phototherapy (UVA1 [ultraviolet A1 irradiation] or PUVA [psoralens + ultraviolet A irradiation]). Treatment of systemic disease includes vascular modifying agents, immunosuppressives, and antifibrotics.

Morphea is characterized by localized thickening and sclerosis of skin; it dominates on the trunk. This disorder may affect children or adults. Morphea begins as erythematous or flesh-colored plaques that become sclerotic, develop central hypopigmentation, and have an erythematous border. In most cases, patients have one or a few lesions, and the disease is termed *circumscribed morphea*. In some patients, widespread cutaneous lesions may occur without systemic involvement (*generalized morphea*). Many adults with generalized morphea have concomitant rheumatic or other autoimmune disorders. Skin biopsy of morphea is generally indistinguishable from that of scleroderma. Scleroderma and morphea are usually quite resistant to therapy. For this reason, physical therapy to prevent joint contractures and to maintain function is employed and is often helpful. Treatment options for early, rapidly progressive disease include phototherapy (UVA1 or PUVA) or methotrexate alone or in combination with daily glucocorticoids.

Diffuse fasciitis with eosinophilia is a clinical entity that can sometimes be confused with scleroderma. There is usually a sudden onset of swelling, induration, and erythema of the extremities, frequently following significant physical exertion, initiation of hemodialysis, exposure to certain medications, or other triggers. The proximal portions of the extremities (upper arms, forearms, thighs, calves) are more often involved than are the hands and feet. While the skin is indurated, it usually displays a woody, dimpled, or "pseudocellulite" appearance rather than being bound down as in scleroderma; contractures may occur early secondary to fascial involvement. The latter may also cause muscle groups to be separated and veins to appear depressed (i.e., the "groove sign"). These skin findings are accompanied by peripheral-blood eosinophilia, increased erythrocyte sedimentation rate, and sometimes hypergammaglobulinemia. Deep biopsy of affected areas of skin reveals inflammation and thickening of the deep fascia overlying muscle. An inflammatory infiltrate composed of eosinophils and mononuclear cells is usually found. Patients with eosinophilic fasciitis appear to be at increased risk for developing bone marrow failure or other hematologic abnormalities. While the ultimate course of eosinophilic fasciitis is variable, most patients respond favorably to treatment with prednisone. Relapses may occur and require treatment with prednisone in combination with other immunosuppressive or immunomodulatory agents.

The *eosinophilia-myalgia syndrome*, a disorder with epidemic numbers of cases reported in 1989 and linked to ingestion of L-tryptophan manufactured by a single company in Japan, is a multisystem disorder characterized by debilitating myalgias and absolute eosinophilia in association with varying combinations of arthralgias, pulmonary symptoms, and peripheral edema. In a later phase (3–6 months after initial symptoms), these patients often develop localized sclerodermatous skin changes, weight loss, and/or neuropathy **(Chap. 372)**.

FURTHER READING

Bolognia JL et al (eds): *Dermatology*, 4th ed. Philadelphia, Elsevier, 2018.

Ellebrecht CT et al: Pemphigus and pemphigoid. J Invest Dermatol 142:907, 2022.

Hammers CM, Stanley JR: Mechanisms of disease: Pemphigus and bullous pemphigoid. Annu Rev Pathol 11:175, 2016.

Kang S et al (eds): *Fitzpatrick's Dermatology*, 9th ed. New York, McGraw-Hill, 2019.

63

Cutaneous Drug Reactions

Robert G. Micheletti, Misha Rosenbach,
Elizabeth J. Phillips, Bruce U. Wintroub,
Kanade Shinkai

Cutaneous reactions are the most frequent adverse reactions to medications, representing 10–15% of reported adverse drug reactions. Most are benign, but a few can be life threatening. Prompt recognition of severe reactions, drug withdrawal, and appropriate therapeutic interventions can minimize toxicity. This chapter focuses on adverse cutaneous reactions to systemic medications; it covers their incidence, patterns, and pathogenesis, and provides some practical guidelines on treatment, assessment of causality, and future use of drugs.

USE OF PRESCRIPTION DRUGS IN THE UNITED STATES

In the United States, more than 6.7 billion prescriptions for >20,000 drug products are dispensed annually. Hospital inpatients alone annually receive about 120 million courses of drug therapy, and approximately two-thirds of adult Americans receive prescription drugs on a regular outpatient basis. Adverse effects of a prescription medication may result in 4.5 million urgent or emergency care visits and over 7000 deaths each year in the United States. Many patients use over-the-counter medicines that may cause adverse cutaneous reactions.

INCIDENCE OF CUTANEOUS REACTIONS

Several prospective studies reported that acute cutaneous reactions to drugs affect between 2.2 and 10 per 1000 hospitalized patients. Reactions usually occur a few days to 4 weeks after initiation of therapy.

In a series of 48,005 inpatients over a 20-year period, morbilliform rash (91%) and urticaria (6%) were the most frequent skin reactions, and antimicrobials, radiocontrast, and nonsteroidal anti-inflammatory drugs (NSAIDs) were the most common drug associations. Severe hypersensitivity reactions to medications have been reported to occur in between 1 in 1000 and 2 per million users, depending on the reaction type. Although rare, severe cutaneous reactions to drugs have an important impact on health because of significant sequelae; in addition, they may require hospitalization, increase the duration of hospital stay, or be life-threatening. Some populations are at increased risk of drug reactions, including elderly patients, patients with autoimmune disease, hematopoietic stem cell transplant recipients, and those with acute Epstein-Barr virus (EBV) or human immunodeficiency virus (HIV) infection. The pathophysiology underlying this association may be related to immune dysregulation. Individuals with advanced HIV disease that is not virologically suppressed (e.g., CD4+ T lymphocyte count <200 cells/μL) have a 40- to 50-fold increased risk of adverse reactions to sulfamethoxazole **(Chap. 208)** and increased risk of severe hypersensitivity reactions to medications overall.

In addition to acute eruptions, a variety of skin diseases can be induced or exacerbated by medications (e.g., pruritus, pigmentation, nail or hair disorders, psoriasis, bullous pemphigoid, photosensitivity, and even cutaneous neoplasms). These drug reactions are not frequent; however, neither their incidence nor their impact on public health has been evaluated.

PATHOGENESIS OF DRUG REACTIONS

Adverse cutaneous responses to drugs can arise as a result of immunologic or nonimmunologic mechanisms.

NONIMMUNOLOGIC DRUG REACTIONS

Examples of nonimmunologic cutaneous drug reactions are pigmentary changes due to dermal accumulation of medications or their metabolites or alteration of hair follicles by antimetabolites and signaling inhibitors. These side effects are predictable and sometimes can be prevented.

IMMUNOLOGIC DRUG REACTIONS

Evidence suggests a direct immunologic basis for most acute drug eruptions. Drug reactions such as urticaria or anaphylaxis may result from immediate release of preformed mediators. Such reactions may be antibody-mediated through an adaptive immune response. Alternatively, they may result from complement activation or direct activation of a mast cell receptor such as the mast-related G-protein–coupled receptor-2 (MRGPRX2) that is not associated with immune memory. T-cell–mediated reactions typically manifest with a delayed exanthem. Drug-specific CD4+ and CD8+ T-cell clones can be derived from the blood or from skin lesions of such patients, suggesting these T cells mediate drug allergy in an antigen-specific manner. For severe cutaneous adverse reactions (SCARs) such as drug-induced hypersensitivity syndrome (DIHS), also known as drug reaction with eosinophilia and systemic symptoms (DRESS), and Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN), drug presentation to T cells is major histocompatibility complex (MHC)-restricted and likely involves drug-peptide complex recognition by specific T-cell receptors (TCRs).

Once a drug has induced an immune response, the phenotype of the reaction is determined by the nature of effectors: cytotoxic (CD8+) T cells in blistering and certain hypersensitivity reactions, chemokines for reactions mediated by neutrophils or eosinophils, and B-cell collaboration for production of specific antibodies for urticarial reactions. There has been consideration for reclassification of hypersensitivity reactions into antibody and cell-mediated processes that are directly driven by inflammation and the immune system, tissue driven mechanisms, and those mediated through a direct response to chemicals **(Table 63-1 and Fig. 63-1)**.

Immediate Reactions Immediate reactions depend on the release of mediators of inflammation by tissue mast cells or circulating basophils. These mediators include histamine, leukotrienes, prostaglandins, bradykinins, platelet-activating factor, enzymes, and proteoglycans. Drugs can trigger mediator release either directly through non–IgE-mediated mast cell activation (formerly “pseudoallergic” or

TABLE 63-1 Revised Classification of Adverse Drug Reactions Based on Immune Pathway

TYPE	KEY PATHWAY	KEY IMMUNE MEDIATORS	ADVERSE DRUG REACTION TYPE
Antibody Mediated			
Type I	IgE, immediate	IgE, B cells, T_H2, ILC2 (IL-4,IL-5, IL-9, IL-13) Mast cells, basophils	Acute urticaria, angioedema, anaphylaxis
Type II	IgG-mediated cytotoxicity	IgG, B cells, IgM Phagocytes, neutrophils, macrophages Complement-dependent cytotoxicity, NK (antibody-dependent cellular cytotoxicity)	Drug-induced cytopenia (e.g. hemolysis, thrombocytopenia secondary to penicillin)
Type III	Immune complexes	IgG + antigen (immune complexes) B cells, IgM, IgG Complement, basophils, mast cells, platelets, Neutrophils, monocytes, macrophages	Serum sickness, Arthus reaction, drug-induced vasculitis and lupus
Cell-Mediated (T-cell mediated) (Delayed Hypersensitivity)			
Type IVa (T1)	T lymphocyte–mediated macrophage inflammation and cytotoxic functions of T cells	T_H1 cells, ILC1, Tc1, NK (IFN-γ, TNF-α, granzyme B, perforins, granulysin) Macrophages (granulomas)	SJS/TEN, erythema multiforme, allergic contact dermatitis, fixed drug eruption, drug induced liver injury
Type IVb (T2)	T lymphocyte–mediated eosinophil inflammation	T_H2 cells, ILC2, Tc2, NK-T (IL-4, IL-5, IL-9, IL-13, IL-31) Eosinophils, B cells, mast cells/basophils	DIHS/DRESS Morbilliform eruption
Type IVc (T3)	T lymphocyte–mediated neutrophil inflammation	TH17, ILC3, Tc17 (IL17, IL-22, IL-23, CXCL8, GM-CSF) Neutrophils	AGEP

Abbreviations: AGEP, acute generalized exanthematous pustulosis; DIHS, drug-induced hypersensitivity syndrome; DRESS, drug reaction with eosinophilia and systemic symptoms. GM-CSF, granulocyte-macrophage colony-stimulating factor; IFN, interferon; IL, interleukin; SJS, Stevens-Johnson syndrome; TEN, toxic epidermal necrolysis; TNF, tumor necrosis factor; Ig, immunoglobulin; ILC1/2/3, innate lymphoid cells type 1/2/3; NK, natural killer cell, NK-T, natural killer T cell. T1/T2/T3, type 1/2/3 immune response; Tc1/2/17, cytotoxic lymphocyte type 1/2/17. Th, T helper lymphocytes. *Source:* Adapted from M Jutel, et al. Allergy 78:2851, 2023.

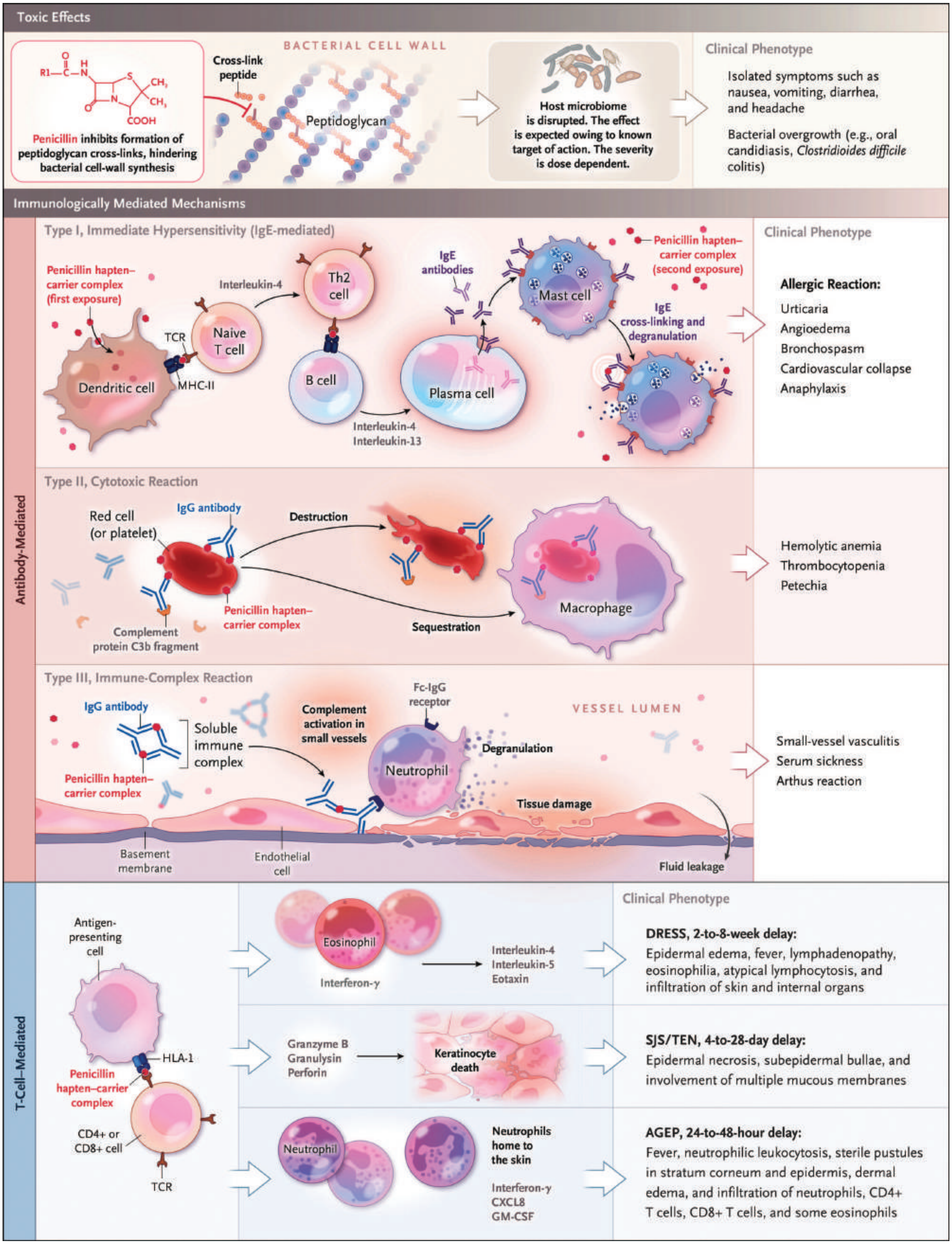

Toxic Effects
R1—C—N H O
CH3
CH3
COOH
Penicillin inhibits formation of peptidoglycan cross-links, hindering bacterial cell-wall synthesis
Cross-link peptide
BACTERIAL CELL WALL
Peptidoglycan
Host microbiome is disrupted. The effect is expected owing to known target of action. The severity is dose dependent.
Clinical Phenotype
Isolated symptoms such as nausea, vomiting, diarrhea, and headache
Bacterial overgrowth (e.g., oral candidiasis, Clostridioides difficile colitis)
Immunologically Mediated Mechanisms
Antibody-Mediated
Type I, Immediate Hypersensitivity (IgE-mediated)
Penicillin hapten–carrier complex (first exposure)
Interleukin-4
Th2 cell
IgE antibodies
Penicillin hapten–carrier complex (second exposure)
Mast cell
IgE cross-linking and degranulation
TCR
Naive T cell
Dendritic cell
MHC-II
B cell
Plasma cell
Interleukin-4
Interleukin-13
Clinical Phenotype
Allergic Reaction:
Urticaria
Angioedema
Bronchospasm
Cardiovascular collapse
Anaphylaxis
Type II, Cytotoxic Reaction
Red cell (or platelet)
IgG antibody
Destruction
Penicillin hapten–carrier complex
Complement protein C3b fragment
Sequestration
Macrophage
Hemolytic anemia
Thrombocytopenia
Petechia
Type III, Immune-Complex Reaction
IgG antibody
Soluble immune complex
Complement activation in small vessels
Fc-IgG receptor
VESSEL LUMEN
Penicillin hapten–carrier complex
Neutrophil
Degranulation
Tissue damage
Basement membrane
Endothelial cell
Fluid leakage
Small-vessel vasculitis
Serum sickness
Arthus reaction
T-Cell–Mediated
Antigen-presenting cell
HLA-1
Penicillin hapten–carrier complex
CD4+ or CD8+ cell
TCR
Eosinophil
Interferon-γ
Interleukin-4
Interleukin-5
Eotaxin
Granzyme B
Granulysin
Perforin
Keratinocyte death
Neutrophil
Neutrophils home to the skin
Interferon-γ
CXCL8
GM-CSF
Clinical Phenotype
DRESS, 2-to-8-week delay:
Epidermal edema, fever, lymphadenopathy, eosinophilia, atypical lymphocytosis, and infiltration of skin and internal organs
SJS/TEN, 4-to-28-day delay:
Epidermal necrosis, subepidermal bullae, and involvement of multiple mucous membranes
AGEP, 24-to-48-hour delay:
Fever, neutrophilic leukocytosis, sterile pustules in stratum corneum and epidermis, dermal edema, and infiltration of neutrophils, CD4+ T cells, CD8+ T cells, and some eosinophils

FIGURE 63-1 Clinical manifestations and mechanisms of immune-mediated cutaneous drug reactions. Antibody-mediated reactions include IgE-mediated reactions, cytotoxic reactions, and immune complex reactions. IgE-mediated reactions, also classified as type I immediate reactions, require a period of prior exposure and sensitization to a drug (e.g., penicillin). Drug-specific IgE is bound to high-affinity IgE receptors on mast cells and then cross-linked by the drug. This leads to activation of mast cells and release of mediators including tryptase and histamine, which are responsible for clinical symptoms such as urticaria, hypotension, bronchospasm, and anaphylaxis. In type II cytotoxic reactions, antibodies target the cell membrane of red or white blood cells and platelets, leading to cell destruction and cytopenia. In type III immune complex reactions, antibodies react with a drug protein carrier forming soluble immune complexes. This leads to complement activation and deposition, ultimately leading to tissue damage and small vessel vasculitis. Cell- (T-cell) mediated reactions generally occur more than 6 hours after the first exposure to a drug. In these reactions, an antigen presenting cell processes peptides modified by the drug and presents them in the antigen binding groove of HLA for recognition by the T-cell receptor on CD8+ or CD4+ T cells, which leads to activation of T cells and release of specific cytokines (dependent on the specific reaction). Specific cytokine and cellular features defining these reactions are shown in Table 63-1, and specific clinical features and causative drugs are listed in Table 63-3. *(From M Castells et al: Penicillin allergy. N Engl J Med 381:2338. Copyright © (2019) Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

"anaphylactoid" reaction) or through IgE-specific antibodies. These reactions usually manifest in the skin and gastrointestinal, respiratory, and cardiovascular systems (Chap. 364). Primary symptoms and signs include pruritus, urticaria, nausea, vomiting, abdominal cramps, bronchospasm, laryngeal edema, and, occasionally, anaphylactic shock with hypotension and death. They occur within minutes of drug exposure. Aspirin, NSAIDs, radiocontrast media, opiates, and some antibiotics (fluoroquinolones, vancomycin) are frequent causes of non–IgE-mediated mast cell activation, which can occur on first exposure, is dose-related, and can be variable in occurrence with multiple exposures over time. Penicillins and cephalosporins (e.g., cefazolin) are the most frequent causes of IgE-dependent reactions to drugs, requiring prior sensitization. When IgE-mediated reactions occur, they typically occur secondary to a haptenated product of the parent drug or metabolite but occasionally can be due to an IgE-mediated response against an inactive excipient of the drug (e.g., polyethylene glycol 3350 in injectable methylprednisolone acetate). Release of mediators is triggered when polyvalent drug protein conjugates cross-link IgE molecules fixed to sensitized cells. Certain routes of administration favor different clinical patterns (e.g., gastrointestinal effects from oral route, circulatory effects from intravenous route).

Immune Complex–Dependent Reactions Serum sickness is produced by tissue deposition of circulating immune complexes with consumption of complement. It is characterized by the triad of fever, arthritis, and urticarial, papular, or purpuric rash with accompanying hypocomplementemia and nephritis (Chap. 375). First described following administration of nonhuman sera, it currently occurs in the setting of monoclonal antibodies and similar medications. In classic serum sickness, symptoms develop 6 or more days after drug exposure, the latent period representing the time needed to synthesize antibody. Vasculitis, a relatively rare complication of drugs, may also be a result of immune complex deposition (Chap. 375). Clinically similar "serum sickness–like" reactions (SSLRs) may be associated with penicillin, cefaclor, amoxicillin, trimethoprim/sulfamethoxazole, and monoclonal antibodies such as infliximab, rituximab, and omalizumab. The mechanism of this reaction is unknown but is unrelated to immune complex formation and complement activation, and systemic involvement is rare. Whereas serum sickness most commonly occurs in adults, SSLR is more frequently observed in children in the setting of a concurrent viral illness. It is now recognized that most children who experience SSLR in association with a β-lactam antibiotic will not have a repeat reaction on rechallenge with the same drug.

Delayed Hypersensitivity While not completely understood, delayed hypersensitivity directed by drug antigen–driven T cells is an important mechanism underlying the most common drug eruptions, specifically morbilliform eruptions and also rare and severe cutaneous and systemic forms, such as DIHS/DRESS, acute generalized exanthematous pustulosis (AGEP), and SJS/TEN (Table 63-1). Drug-specific T cells have been detected in some of these reactions. Drugs themselves are not independently antigenic, and the specific drug-altered antigens mediating delayed hypersensitivity reactions have not been identified. In SJS/TEN, the immune response is triggered by the presentation of a drug-antigen by a keratinocyte or other antigen presenting cell (APC) expressing a risk HLA class I allele. A specific T-cell receptor (TCR) expressed by CD8+ T cells enriched in the skin recognizes this drug-HLA complex, leading to activation and release of cytolytic peptides such as granulysin that are thought to be the main mediators of keratinocyte death. In the case of carbamazepine-related SJS/TEN, studies have identified cytotoxic T lymphocytes (CTLs) reactive to carbamazepine that use highly restricted V-alpha and V-beta TCR repertoires in affected patients. This TCR restriction is shared among patients with carbamazepine SJS/TEN restricted by HLA-B*15:02 and other B75 serotypes but is not found in carbamazepine-tolerant individuals or individuals with carbamazepine-related DIHS/DRESS.

The specific mechanisms by which medications activate T cells and by which a memory T-cell response is maintained despite the absence of the drug are unknown. The hapten-prohapten model suggests that antigens driving these reactions may be the drug or a metabolite that covalently binds to endogenous proteins, presented in association with HLA molecules to T cells through the classic antigen presentation pathway. The pharmacologic interaction (p-i) model proposes that the drug/metabolite-altered antigen binds noncovalently to the risk HLA or TCR without undergoing classical antigen processing. Compelling evidence exists that abacavir activates CD8+ T cells through an altered peptide repertoire model whereby abacavir binds within the antigen binding cleft of HLA-B*57:01 and alters the repertoire of self-peptides that are seen as immunogenic. Functional data and x-ray crystallography solving the crystal structure of abacavir bound to synthetic or natural peptide and HLA-B*57:01 support this model. It is still puzzling how memory T-cell responses to small-molecule drugs are preserved over time in the absence of reexposure. Independent of existing models of T-cell activation, a heterologous immunity model proposes that a cross-reactive memory T-cell response occurs between HLA class I restricted immunity to a prevalent pathogen early in life (such as a common virotope) and response to a drug-altered self-peptide restricted by the same HLA class I allele.

GENETIC FACTORS AND CUTANEOUS DRUG REACTIONS

Current knowledge suggests that genetic determinants may predispose individuals to severe drug reactions by affecting immune responses to drugs and, in some cases, through altered drug metabolism. Associations between T-cell–mediated delayed drug hypersensitivities and specific class I HLA alleles suggest a key role for immune mechanisms, especially those leading to severe skin or systemic organ involvement, such as SJS/TEN, DIHS/DRESS, abacavir hypersensitivity (AHS), drug-induced liver injury, pancreatitis, or agranulocytosis. Hypersensitivity to the anti-HIV medication abacavir is strongly associated with HLA-B*57:01. HLA-B*57:01 has 100% negative predictive value for the development of AHS, and routine preprescription screening for HLA-B*57:01 has eliminated AHS as a clinical entity (Chap. 208). In Taiwan, within a homogeneous Han Chinese population, a strong association was observed between SJS/TEN (but not DIHS/DRESS) related to carbamazepine and HLA-B*15:02. In the same population, a strong association was found between HLA-B*58:01 and both DIHS/DRESS and SJS/TEN associated with allopurinol. Other notable HLA associations include HLA-B*13:01 and dapsone or sulfonamide antibiotic-induced DIHS/DRESS and SJS/TEN, as well as HLA-A*32:01 and vancomycin-induced DIHS/DRESS. These associations are drug and phenotype specific; that is, HLA-specific T-cell stimulation by medications leads to distinct reactions. However, while these HLA associations are extremely strong, the presence of a risk HLA allele alone is not sufficient to cause severe drug hypersensitivity reactions. Other factors outside of the MHC, such as drug metabolism (e.g., CYP2C9*3 slow metabolism status and risk of phenytoin-induced morbilliform drug eruption, DIHS/DRESS, and SJS/TEN), are also important. The entirety of genetic and ecological factors that define why the majority of individuals carrying an HLA risk allele remain drug-tolerant and those factors that drive risk for specific drug hypersensitivity reactions are not known at this time.

GLOBAL CONSIDERATIONS

The PREDICT-1 study was a randomized, double-blind clinical trial of real-time preprescription HLA-B*57:01 screening versus the previous standard of care of no screening that provided the basis for the utility of HLA-B*57:01 screening to prevent AHS. If an HLA risk allele is present, it seems to have the same implication regardless of the population. This generalizability across populations, the fact that self-identified race is a social rather than biological construct, and greater availability of less costly single-HLA assays with a shorter turnaround time argue against targeted genetic screening on the basis of self-identified race. The American College of Rheumatology has recommended HLA-B*58:01 screening of Han Chinese patients prescribed allopurinol; however, HLA-B*58:01 does not have 100% negative predictive value for allopurinol DIHS/DRESS or SJS/TEN in other populations, and safety monitoring and risk counseling remain imperative. To date, screening for a single HLA (but not multiple HLA haplotypes) in specific populations has been determined to be cost-effective (e.g., HLA-B*1301 screening in Chinese patients with leprosy treated with dapsone). Genetic testing for specific HLA haplotypes is becoming more widely available, and its utility will increase as mechanistic factors driving drug hypersensitivity and tolerance are elucidated. Further data are needed to clarify the role of genetic and other testing beyond screening (e.g., in diagnosis) in order to optimize cost-effectiveness, utility, and benefit.

CLINICAL PRESENTATION OF CUTANEOUS DRUG REACTIONS

NONIMMUNE CUTANEOUS REACTIONS

Exacerbation or Induction of Dermatologic Diseases A variety of drugs can exacerbate preexisting diseases or induce—or unmask—a disease that may or may not disappear after withdrawal of the inducing medication. For example, NSAIDs, lithium, beta blockers, tumor necrosis factor (TNF) antagonists, interferon (IFN) α, and angiotensin-converting enzyme (ACE) inhibitors can exacerbate plaque psoriasis, whereas antimalarials and withdrawal of systemic glucocorticoids can worsen pustular psoriasis. The situation of TNF-α inhibitors is unusual, as this class of medications is used to treat psoriasis; however, they may paradoxically induce psoriasis (especially palmoplantar) in patients being treated for other conditions. Acne may be induced by glucocorticoids, androgens, lithium, and antidepressants. Follicular papular or pustular eruptions of the face and trunk resembling acne frequently occur with epidermal growth factor receptor (EGFR) antagonists, mitogen-activated protein kinase (MEK) inhibitors, tyrosine kinase (TK) inhibitors, and other targeted inhibitors. With EGFR antagonists, the severity of the eruption correlates with a better anticancer effect. This rash is typically responsive to and prevented by tetracycline antibiotics.

Several medications induce or exacerbate autoimmune disease. Checkpoint inhibitors induce a wide array of systemic autoimmune reactions, including in skin (see below). Interleukin (IL) 2, IFN-α, and anti-TNF-α are associated with new-onset systemic lupus erythematosus (SLE). Drug-induced lupus is classically marked by antinuclear and antihistone antibodies and, in some cases, anti-double-stranded DNA (D-penicillamine, anti-TNF-α) or perinuclear antineutrophil cytoplasmic antibodies (p-ANCA) (minocycline). Subacute cutaneous lupus erythematosus (SCLE) can be induced by a growing list of drugs, including thiazide diuretics, proton pump inhibitors, TNF inhibitors, terbinafine, and minocycline **(Fig. 63-2)**. Drug-induced dermatomyositis may rarely occur with TNF inhibitors or capecitabine, and flares have been reported in association with the herbal supplement spirulina. Hydroxyurea can induce skin findings of dermatomyositis. IFN and TNF inhibitors, as well as BRAF inhibitors and checkpoint inhibitors, can induce granulomatous disease and sarcoidosis. Autoimmune blistering diseases may be drug induced as well: pemphigus by D-penicillamine and ACE inhibitors; bullous pemphigoid by furosemide and, increasingly, by DPP4 inhibitors and PD-1 inhibitors; and linear IgA bullous dermatosis by vancomycin. Other medications may cause highly specific cutaneous reactions. Gadolinium contrast has been associated with nephrogenic systemic fibrosis, a condition of sclerosing skin and rare internal organ involvement, in the setting of renal compromise. Granulocyte colony-stimulating factor, azacitidine, all-*trans*-retinoic acid, the *FLT3* inhibitor class of drugs, and (rarely) levamisole-contaminated cocaine may induce neutrophilic dermatoses. The hypothesis that a drug may be responsible should always be considered, even after treatment is complete. In addition, reactions may develop in cases of long-term medication therapy due to changes in dosing or host metabolism. Resolution of the cutaneous reaction may be delayed upon discontinuation of the medication.

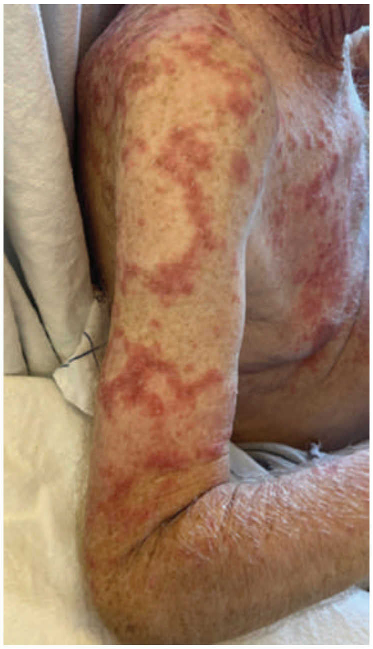

FIGURE 63-2 Subcutaneous lupus erythematosus due to medication.

Photosensitivity Eruptions Photosensitivity eruptions are usually most marked in sun-exposed areas, but they may extend to sun-protected sites. The mechanism is almost always phototoxicity. Phototoxic reactions resemble sunburn and can occur with first exposure to a drug. Blistering may occur in drug-related pseudoporphyria, most commonly with NSAIDs. The severity of the reaction depends on the tissue level of the drug, its efficiency as a photosensitizer, and the extent of exposure to the activating wavelengths of ultraviolet (UV) light **(Chap. 64)**.

Common orally administered photosensitizing drugs include fluoroquinolones, tetracycline antibiotics, and trimethoprim/sulfamethoxazole. Other drugs less frequently implicated are chlorpromazine, thiazides, NSAIDs, and BRAF inhibitors. Voriconazole may result in severe photosensitivity, accelerated photoaging, and cutaneous carcinogenesis. Hydrochlorothiazide may be associated with increased nonmelanoma skin cancer in some populations.

Because UV-A and visible light, which trigger these reactions, are not easily absorbed by nonopaque sunscreens and are transmitted through window glass, photosensitivity reactions may be difficult to prevent. Photosensitivity reactions abate with removal of either the drug or UV radiation, use of sunscreens that block UV-A light, and treatment of the reaction as one would a sunburn. Rarely, individuals develop persistent reactivity to light, necessitating long-term avoidance of sun exposure. Some chemotherapeutic agents, such as methotrexate, can induce a UV-recall reaction characterized by an erythematous, slightly scaly eruption at sites of prior severe photosensitivity reactions.

Pigmentation Changes Drugs, either systemic or topical, may cause a variety of pigmentary changes in the skin by triggering melanocyte production of melanin (as in the case of oral contraceptives causing melasma) or due to deposition of drug or drug metabolites. Long-term minocycline or amiodarone exposure may cause blue-gray pigmentation. Phenothiazine, gold, and bismuth result in gray-brown pigmentation of sun-exposed areas. Numerous cancer chemotherapeutic agents may be associated with characteristic patterns of pigmentation (e.g., bleomycin, busulfan, daunorubicin, cyclophosphamide, hydroxyurea, fluorouracil, and methotrexate). Clofazimine causes a drug-induced lipofuscinosis with characteristic red-brown coloration. Hyperpigmentation of the face, mucous membranes, and pretibial and subungual areas occurs with antimalarials. Quinacrine causes generalized yellow discoloration. Pigmentation changes may also occur in mucous membranes (busulfan, bismuth), conjunctiva (chlorpromazine, thioridazine, imipramine, clomipramine), nails (zidovudine, doxorubicin, cyclophosphamide, bleomycin, fluorouracil, hydroxyurea), hair, and teeth (tetracyclines).

Warfarin Necrosis of Skin This rare reaction (0.01–0.1%) usually occurs between the third and tenth days of therapy with warfarin, usually in women. Common sites are breasts, thighs, and buttocks **(Fig. 63-3)**. Lesions are sharply demarcated, erythematous,

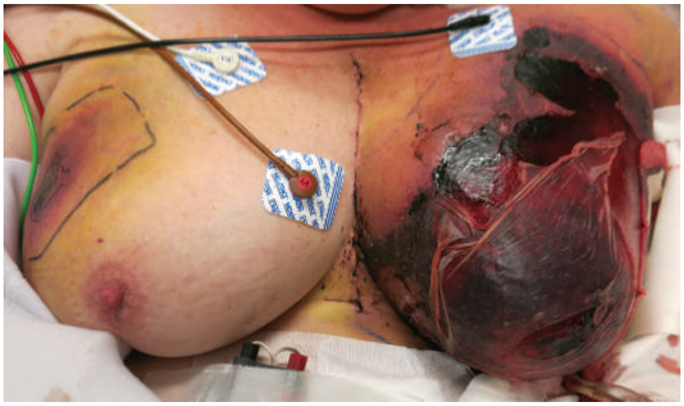

FIGURE 63-3 **Warfarin necrosis** involving the breasts.

or purpuric, and may progress to form large, hemorrhagic bullae with necrosis and eschar formation.

Warfarin anticoagulation in protein C or S deficiency causes an additional reduction in already low circulating levels of endogenous anticoagulants, permitting temporary hypercoagulability and thrombosis in the cutaneous microvasculature, with consequent areas of necrosis. Heparin-induced necrosis may have clinically similar features but is probably due to heparin-induced platelet aggregation with subsequent occlusion of blood vessels; it can affect areas adjacent to the injection site or more distant sites if infused. Levamisole-tainted cocaine (and more recently, opiates) can induce similar skin necrosis; however, the distribution tends to involve the ears and cheeks predominantly, with stellate or retiform purpura. Patients may have abnormal white blood cell counts and may be dual P- and C-ANCA positive.

Drug-Induced Hair Disorders • **DRUG-INDUCED HAIR LOSS** Medications may affect hair follicles at two different phases of their growth cycle: anagen (growth) or telogen (resting). *Anagen effluvium* occurs within days of drug administration, especially with antimetabolite or other chemotherapeutic drugs. In contrast, in *telogen effluvium*, the delay is 2–4 months following initiation of a new medication. Both present as diffuse, nonscarring alopecia, most often reversible after discontinuation of the responsible agent.

A considerable number of drugs have been associated with hair loss. These include antineoplastic agents (alkylating agents, bleomycin, vinca alkaloids, platinum compounds), anticonvulsants (carbamazepine, valproate), beta blockers, antidepressants, antithyroid drugs, IFNs, oral contraceptives, and cholesterol-lowering agents.

DRUG-INDUCED HAIR GROWTH Medications may also cause hair growth. Hirsutism is an excessive growth of terminal hair with masculine hair growth pattern in a female, most often on the face and trunk, due to androgenic stimulation of hormone-sensitive hair follicles (anabolic steroids, oral contraceptives, testosterone, corticotropin). Hypertrichosis is a distinct pattern of hair growth, not in a masculine pattern, typically located on the forehead and temporal regions of the face. Drugs responsible for hypertrichosis include anti-inflammatory drugs, glucocorticoids, vasodilators (diazoxide, minoxidil), diuretics (acetazolamide), anticonvulsants (phenytoin), immunosuppressive agents (cyclosporine A), psoralens, and zidovudine.

Changes in hair color or structure are uncommon adverse effects from medications. Hair discoloration may occur with chloroquine, IFN-α, chemotherapeutic agents, and tyrosine kinase inhibitors. Changes in hair structure have been observed in patients given EGFR inhibitors, BRAF inhibitors, tyrosine kinase inhibitors, and acitretin.

Drug-Induced Nail Disorders Drug-related nail disorders usually involve all 20 nails and need months to resolve after withdrawal of the medication. The pathogenesis is most often toxic. Drug-induced nail changes include Beau lines (transverse depression of the nail plate), onycholysis (detachment of the distal part of the nail plate), onychomadesis (detachment of the proximal part of the nail plate), pigmentation, and paronychia (inflammation of periungual skin).

ONYCHOLYSIS Onycholysis occurs with tetracyclines, fluoroquinolones, retinoids, NSAIDs, and others, including many chemotherapeutic agents, and may be triggered by exposure to sunlight.

ONYCHOMADESIS Onychomadesis is caused by temporary arrest of nail matrix mitotic activity. Common drugs reported to induce onychomadesis include carbamazepine, lithium, retinoids, and chemotherapeutic agents such as taxanes.

PARONYCHIA Paronychia and multiple pyogenic granulomas with progressive and painful periungual abscess of fingers and toes are side effects of systemic retinoids, lamivudine, indinavir, and EGFR inhibitors.

NAIL DISCOLORATION Some drugs—including anthracyclines, taxanes, fluorouracil, psoralens, and zidovudine—may induce nail bed hyperpigmentation through melanocyte stimulation. It appears to be reversible and dose dependent.

Toxic Erythema of Chemotherapy and Other Chemotherapy Reactions Because many agents used in cancer chemotherapy inhibit cell division, rapidly proliferating elements of the skin, including hair, mucous membranes, and appendages, are sensitive to their effects. A broad spectrum of chemotherapy-related skin toxicities has been reported, including neutrophilic eccrine hidradenitis, sterile cellulitis, exfoliative dermatitis, and flexural erythema. These reactions are best characterized under the unifying term "toxic erythema of chemotherapy" (TEC) **(Fig. 63-4)**. Acral erythema is marked by dysesthesia and an erythematous, edematous eruption of the palms and soles. Common causes include cytarabine, doxorubicin, methotrexate, hydroxyurea, fluorouracil, and capecitabine.

The introduction of many new monoclonal antibody and small-molecule signaling inhibitors for the treatment of cancer has been accompanied by numerous reports of skin and hair toxicity; only the most common of these are mentioned here. EGFR antagonists induce follicular eruptions and nail toxicity after a mean interval of 10 days in a majority of patients. Xerosis, eczematous eruptions, acneiform eruptions, and pruritus are common. Erlotinib is associated with marked hair textural changes. Sorafenib, a tyrosine kinase inhibitor, may result in follicular eruptions and focal bullous eruptions at palmoplantar, flexural sites, or areas of pressure or friction. BRAF inhibitors are associated with photosensitivity, palmoplantar hyperkeratosis, hair curling, dyskeratotic (Grover-like) rash, hyperkeratotic benign cutaneous neoplasms, and keratoacanthoma-like squamous cell carcinomas.

The immune checkpoint inhibitor (ICI) class of drugs (including anti-CTLA-4, anti-PD-1, and anti-PD-L1 agents) can induce a wide range of cutaneous eruptions, including lichenoid, eczematous, granulomatous, papulosquamous, bullous, vitiligo-like, alopecic, and panniculitic eruptions. Mucocutaneous immune-related adverse events (irAEs) to ICIs typically occur early and are the most common of potentially multiorgan reactions, occurring in 30–60% of patients. The onset may be as early as a few weeks after initiating the ICI and

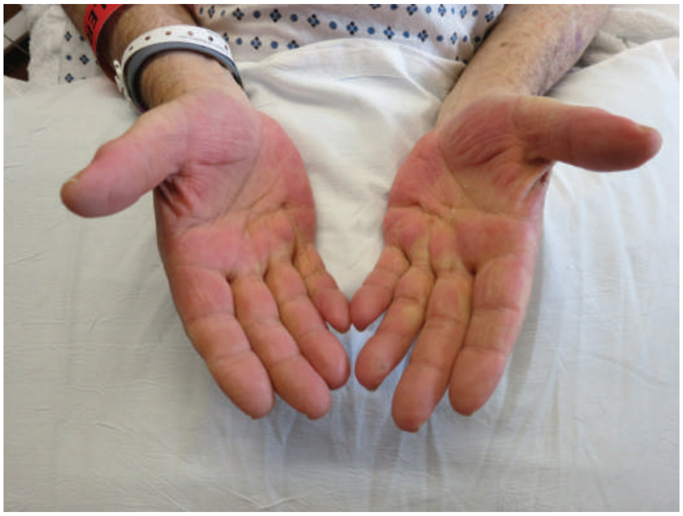

FIGURE 63-4 **Toxic erythema of chemotherapy.**

as late as 1 year or more after initiation. The incidence of irAEs differs between the distinct ICI medications and is most frequent with combination therapy with anti-CTLA-4 and anti-PD-1 agents. Severe irAEs (grade 3 or 4, including DIHS/DRESS, SJS/TEN-like, and bullous reactions involving over 30% body surface area) are rare, occurring in 2–3% of patients with ICI monotherapy and up to 15% with combined ICI therapy. The rates of irAEs may also differ between patient populations depending on the type of malignancy being treated. These reactions are largely T-cell–mediated, which informs the approach to management. Whereas milder reactions may be managed with topical corticosteroids, more severe reactions may require oral corticosteroids or other systemic immunomodulatory or immunosuppressive medications, including biologic agents. In some cases of severe reactions, including lichenoid bullous and SJS/TEN-like reactions, rechallenge of the ICI may be possible. A clear history of all new medications introduced within the few days to 2 months preceding the reaction should be taken, as case reports of ICIs increasing the risk of small-molecule SCAR have emerged. There are some data, largely retrospective studies, supporting the notion that mucocutaneous irAEs may be associated with improved prognosis and survival.

IMMUNE CUTANEOUS REACTIONS: COMMON

Morbilliform Drug Eruption Morbilliform or maculopapular drug eruptions (Fig. 63-5) are the most common of all drug-induced reactions, often start on the trunk or intertriginous areas, and consist of blanching erythematous macules and papules that are symmetric and confluent. Nonblanching, dusky, or bright-red macules, as well as mucosal involvement, should raise concern for a more severe reaction. Facial involvement in morbilliform eruptions is also uncommon, and the presence of extensive facial lesions with facial edema suggests DIHS/DRESS. Diagnosis of morbilliform eruptions is rarely assisted by laboratory testing or skin biopsy.

Morbilliform drug eruptions may be associated with moderate to severe pruritus and fever. A viral exanthem is a differential diagnostic consideration, especially in children, and graft-versus-host disease is also a consideration in the proper clinical setting. Absence of enanthems; absence of ear, nose, throat, and upper respiratory tract symptoms; and polymorphism of the skin lesions support a drug rather than a viral eruption. Common offenders include aminopenicillins, cephalosporins, antibacterial sulfonamides, allopurinol, and antiepileptic drugs. Beta blockers, calcium channel blockers, and ACE inhibitors are rarely the culprit; however, any drug can cause a morbilliform exanthem. Certain medications carry very high rates of morbilliform eruption, including nevirapine and lamotrigine, even in the absence of DIHS/DRESS reactions. Lamotrigine morbilliform rash is associated with higher starting doses, rapid dose escalation, concomitant use of valproate (which increases lamotrigine levels and half-life), and use in children.

Morbilliform drug eruptions usually develop within 1 week of initiation of therapy and last less than 2 weeks. Occasionally, these eruptions resolve despite continued use of the responsible drug. Because the eruption may also worsen, the suspect drug should be discontinued unless it is essential. It is important to note that the rash may continue to progress for a few days up to 1 week following medication discontinuation. Oral antihistamines and emollients may help relieve pruritus. Short courses of potent topical glucocorticoids can reduce inflammation and symptoms. Systemic glucocorticoid treatment is rarely indicated.

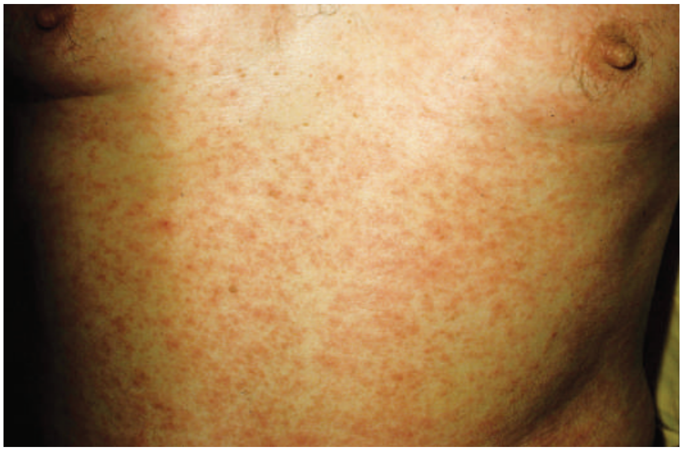

FIGURE 63-5 Morbilliform drug eruption.

Pruritus Pruritus is associated with almost all drug eruptions and, in some cases, may represent the only symptom of the adverse cutaneous reaction. It may be alleviated by antihistamines such as hydroxyzine or diphenhydramine. Pruritus stemming from specific medications may require distinct treatment, such as selective opiate antagonists for opiate-related pruritus.

Urticaria/Angioedema/Anaphylaxis Urticaria, the second most frequent type of cutaneous reaction to drugs, is characterized by pruritic, red wheals of varying size rarely lasting more than 24 hours. It has been observed in association with nearly all drugs, most frequently aspirin, NSAIDs, penicillin, and blood products. However, medications account for no more than 10–20% of acute urticaria cases. Deep edema within dermal and subcutaneous tissues is known as angioedema and may involve respiratory and gastrointestinal mucous membranes. Urticaria and angioedema may be part of a life-threatening anaphylactic reaction.

Drug-induced urticaria may be caused by three mechanisms: an IgE-dependent mechanism, circulating immune complexes (serum sickness), and nonimmunologic activation of effector pathways. IgE-dependent urticarial reactions usually occur within 36 hours of drug exposure but can occur within minutes. Immune complex–induced urticaria associated with serum sickness reactions usually occurs 6–12 days after first exposure. In this syndrome, the urticarial eruption (typically polycyclic plaques over distal joints) may be accompanied by fever, hematuria, arthralgias, hepatic dysfunction, and neurologic symptoms. Certain drugs, such as NSAIDs, ACE inhibitors, angiotensin II antagonists, radiographic dye, and opiates, may induce urticarial reactions, angioedema, and anaphylaxis in the absence of drug-specific antibodies (non–IgE-mediated mast cell activation) through direct mast cell degranulation. Non–IgE-mediated mast cell activation associated with fluoroquinolones, vancomycin, radiocontrast dye, opiates, and other medications that share a tetrahydroisoquinolone (THIQ) moiety is now thought to be mediated through their interaction and activation of the MRGPRX2 receptor on mast cells. Expression of MRGPRX2 is also increased in the skin of patients with severe chronic spontaneous urticaria (CSU), and patients with CSU are more likely to be labeled as allergic to multiple drugs.

Radiocontrast agents are a common cause of urticaria through non–IgE-mediated mast cell activation and, in rare cases, can cause anaphylaxis. High-osmolality radiocontrast media are about five times more likely to induce urticaria (1%) or anaphylaxis than are newer low-osmolality media. About one-third of those with mild reactions to previous exposure react on reexposure. Pretreatment with prednisone and diphenhydramine reduces reaction rates.

The treatment of urticaria or angioedema depends on the severity of the reaction. In severe cases with respiratory or cardiovascular compromise, epinephrine and intravenous glucocorticoids are the mainstay of therapy. For patients with urticaria without symptoms of angioedema or anaphylaxis, drug withdrawal and oral antihistamines are usually sufficient. Future drug avoidance is recommended; rechallenge, especially in individuals with severe reactions, should only occur in an intensive care setting. Vancomycin is a small-molecule ligand for MRGPRX2 and is also associated with an infusional form of non–IgE-mediated mast cell activation now termed vancomycin flushing syndrome (VFS) or vancomycin infusion reaction (VIR). This infusion syndrome is a rate- and dose-dependent histamine-related reaction characterized by flushing, diffuse urticarial or morbilliform eruption, tachycardia, and hypotension. In rare cases, cardiac arrest may be associated with rapid intravenous (IV) infusion of the medication. It is prevented by slowing the infusion and premedicating with antihistamines.

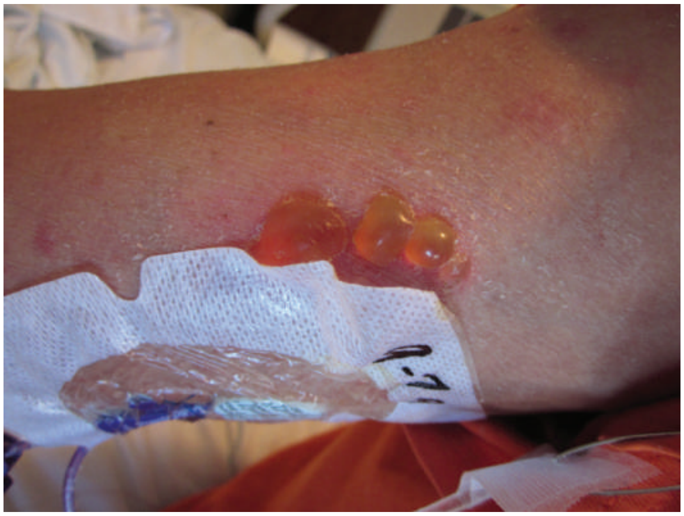

FIGURE 63-6 Allergic contact dermatitis (bullous) due to adhesive tape.

Irritant/Allergic Contact Dermatitis Patients using topical medications may develop an irritant or allergic contact dermatitis to the medication itself or to a preservative or other component of the formulation. Reactions to neomycin sulfate, bacitracin, polymyxin B, and benzalkonium are common, and co-sensitization is common due to their shared presence in multiple products such as ointments and eyedrops. Contact dermatitis may be seen to adhesive tapes, leading to irritation or blisters around ports and IV sites (Fig. 63-6). Harsh disinfectant skin cleansers may lead to localized irritant dermatitis. Systemic contact dermatitis can occur when an individual is sensitized to a contact allergen and later exposed to that same medication or a potentially cross-reactive medication systemically. The most common associated medications are corticosteroids, aminoglycoside antibiotics, and aminophylline. Symmetric drug-related intertriginous and flexural exanthema (SDRIFE) causes well-demarcated erythema in the intertriginous and flexural areas in patients without a previous history of cutaneous sensitization. It is commonly associated with aminopenicillins, β-lactams, and chemotherapeutic agents.

Fixed Drug Eruption This uncommon reaction is characterized by one or more sharply demarcated, dull red-to-brown lesions, sometimes with central dusky violaceous erythema and central bulla (Fig. 63-7). Hyperpigmentation during healing is characteristic, often persists after resolution of the acute inflammation, and can be more pronounced in patients with more richly pigmented skin types. With rechallenge, the process recurs in the same (fixed) locations but may spread to new areas as well. Lesions often involve the lips, hands, legs, face, genitalia, and oral mucosa, and cause a burning sensation. Most patients have multiple flat lesions. Fixed drug eruption has been associated with pseudoephedrine (frequently a nonpigmenting reaction), phenolphthalein (in laxatives), sulfonamides, tetracyclines, fluconazole, NSAIDs, barbiturates, and others. A severe and uncommon variant of fixed drug eruption, generalized bullous fixed drug eruption, involves the development of multiple bullous lesions comprising at least 10% body surface area (BSA). It is considered a SCAR due to its high BSA involvement and mortality that can exceed 20%.

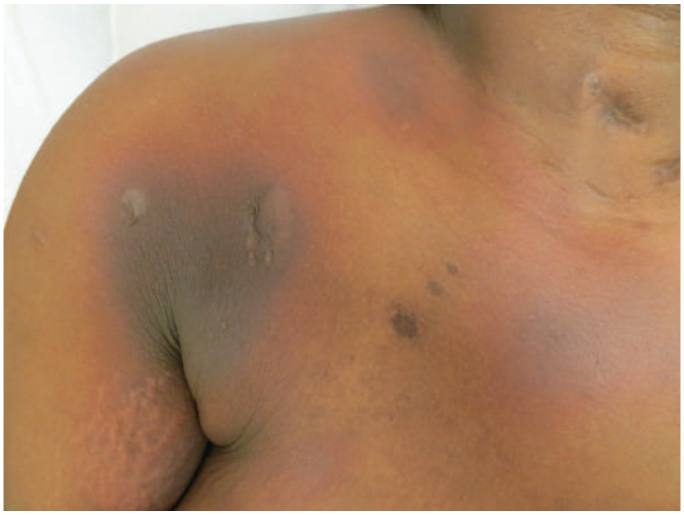

FIGURE 63-7 Fixed drug eruption.

IMMUNE CUTANEOUS REACTIONS: RARE AND SEVERE

Drug-Induced Hypersensitivity Syndrome/Drug Reaction with Eosinophilia and Systemic Symptoms DIHS and DRESS are used interchangeably; however, eosinophilia is an inconsistent and sometimes late laboratory finding. The histopathologic features of DIHS/DRESS are nonspecific and can be indistinguishable from morbilliform drug eruption. There are mixtures of CD4 and CD8 T cells in the skin, and T_H2 cells express CD134 and produce cytokines such as IL-4, IL-5, IL-10, IL-13, and eotaxin. Type 2 innate lymphocytes (ILC2) produce IL-5 that promotes activation and migration of eosinophils. Tregs and dendritic cells produce IFN-γ, TNF-α, and CCL17 (serum thymus and activation-regulated chemokine [TARC]), which fuels the systemic presentation.

Clinically, DIHS/DRESS presents with a prodrome of fever and flu-like symptoms for several days, followed by the appearance of an extensive rash usually involving the face (Fig. 63-8). The rash of DIHS/DRESS is polymorphic. It is usually morbilliform, but can also be exfoliative, lichenoid, urticarial, eczematous, pustular, or a combination of multiple morphologies. Increased severity of DIHS/DRESS is associated with purpuric and erythema multiforme-like cutaneous presentations. Facial swelling and hand/foot swelling are often present and are also associated with severity. Systemic manifestations include lymphadenopathy, fever, and leukocytosis (often with eosinophilia or atypical lymphocytosis), as well as hepatitis, nephritis, pneumonitis, myositis, and gastroenteritis, in descending order. Distinct patterns of timing of onset and organ involvement may exist for specific medications. For example, allopurinol classically induces DIHS/DRESS with renal involvement; cardiac and lung involvement are more common

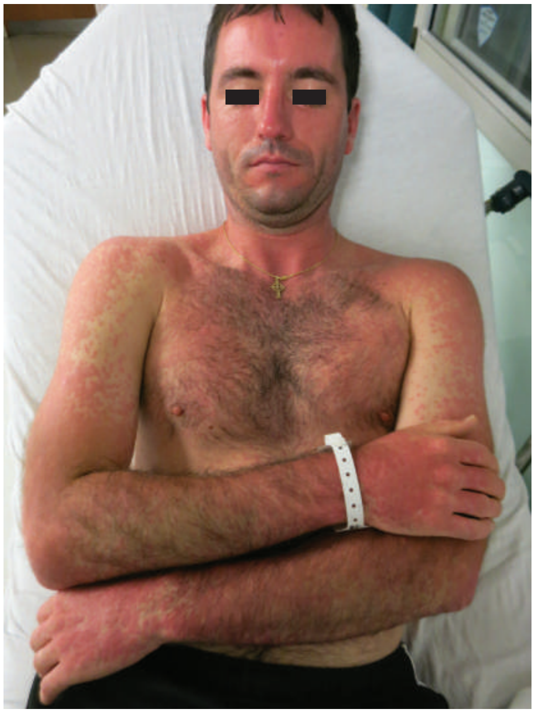

FIGURE 63-8 Drug-induced hypersensitivity syndrome/drug rash with eosinophilia and systemic symptoms (DIHS/DRESS).

with minocycline; and some medications typically do not induce eosinophilia (dapsone, lamotrigine). The cutaneous reaction usually begins 2–8 weeks after the drug is started and persists after drug cessation, with a prolonged resolution period. Signs and symptoms may continue for several weeks, especially those associated with hepatitis, while disease relapse or recrudescence can occur, particularly upon weaning corticosteroids. The eruption recurs with rechallenge, and cross-reactions among aromatic anticonvulsants, including phenytoin, carbamazepine, and phenobarbital, are common. Other drugs causing DIHS/DRESS include vancomycin, antibacterial sulfonamides, and other antibiotics. Hypersensitivity to reactive drug metabolites, hydroxylamine for sulfamethoxazole and arene oxide for aromatic anticonvulsants, may be involved in the pathogenesis of DIHS. Research suggests inciting drugs may reactivate quiescent human herpes viruses, including herpesviruses 6 and 7, EBV, and cytomegalovirus (CMV), resulting in expansion of viral-specific CD8+ T lymphocytes and subsequent end-organ damage. Viral reactivation, particularly with CMV, may be associated with a worse clinical prognosis. Mortality rates as high as 10% have been reported, with most fatalities resulting from liver failure; however, clinical experience and results of a large, unpublished cohort suggest mortality is likely closer to 2%. Systemic glucocorticoids (1.5–2 mg/kg/d prednisone equivalent) should be started and tapered slowly over 8–12 weeks, during which time clinical symptoms and labs (including complete blood count with differential, basic metabolic panel, and liver function tests) should be followed carefully. A steroid-sparing agent such as mycophenolate mofetil, IV immunoglobulin, or cyclosporine may be indicated in cases of rapid recurrence upon steroid taper. A role for newer, targeted biologic therapies, such as inhibitors of IL-5 and the JAK-STAT pathway, is under investigation. In all cases, immediate withdrawal of all potential culprit drugs is essential. Given the severe complications and potentially delayed presentation of myocarditis, patients should undergo cardiac evaluation in cases of severe DIHS/DRESS or if heart involvement is suspected due to hypotension or arrhythmia. Patients should be closely monitored for resolution of organ dysfunction and for development of late-onset autoimmune sequelae, such as thyroiditis, lupus, and diabetes.

Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis

SJS/TEN is a disease spectrum characterized by blisters and mucosal/epidermal detachment resulting from full-thickness epidermal necrosis. The term *Stevens-Johnson syndrome* (SJS) describes cases in which the total BSA of blistering and eventual detachment is <10% **(Figs. 63-9 to 63-11)**. The term *Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) overlap* is used to describe cases with 10–30% BSA epidermal detachment **(Fig. 63-12)**, and the term *toxic epidermal necrolysis (TEN)* is used to describe cases with >30% BSA detachment **(Figs. 63-13 and 63-14)**.

Other blistering eruptions with concomitant mucositis may be confused with SJS/TEN. Erythema multiforme (EM) associated with herpes simplex virus is characterized by painful mucosal erosions and target lesions, typically with an acral distribution and limited skin detachment. *Mycoplasma* and other respiratory infections in children cause a clinically distinct presentation with prominent mucositis and limited cutaneous involvement. The term *reactive infectious mucocutaneous eruption (RIME)* is used to help differentiate this clinical entity.

Patients with SJS/TEN initially present with fever >39°C (102.2°F); sore throat; conjunctivitis; and acute onset of painful dusky, atypical,

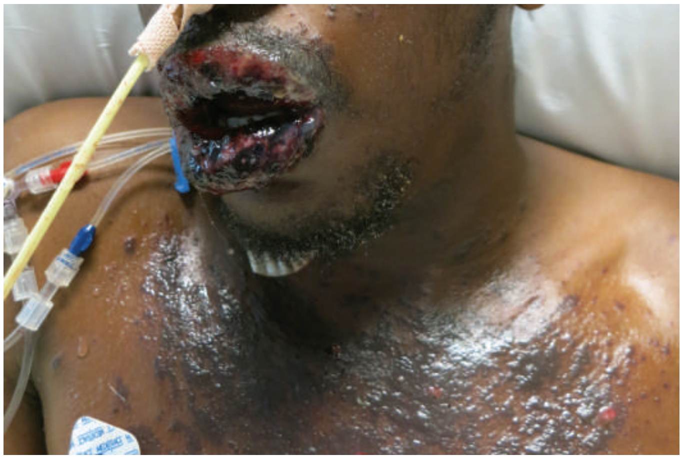

FIGURE 63-9 Stevens-Johnson syndrome (SJS).

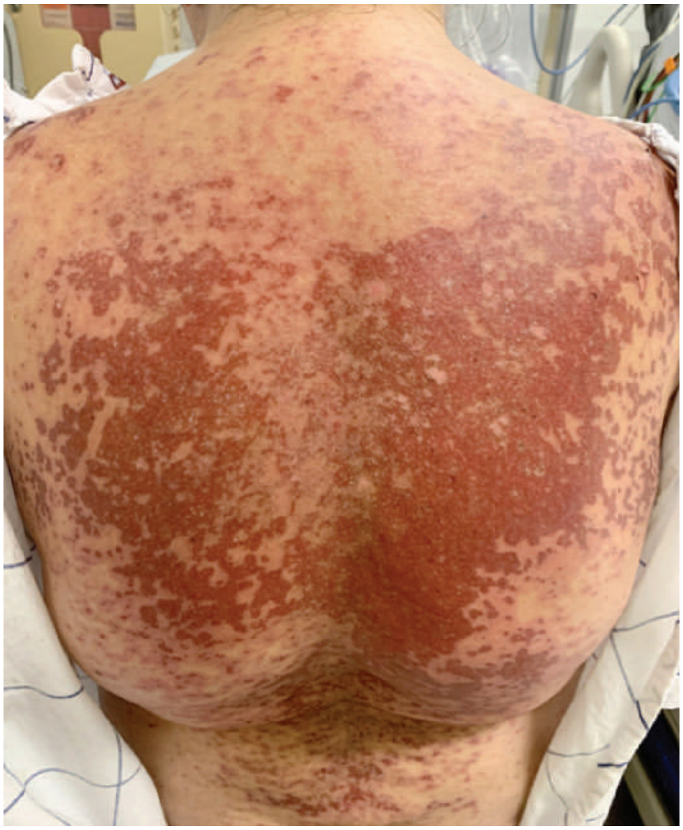

FIGURE 63-10 SJS-TEN overlap.

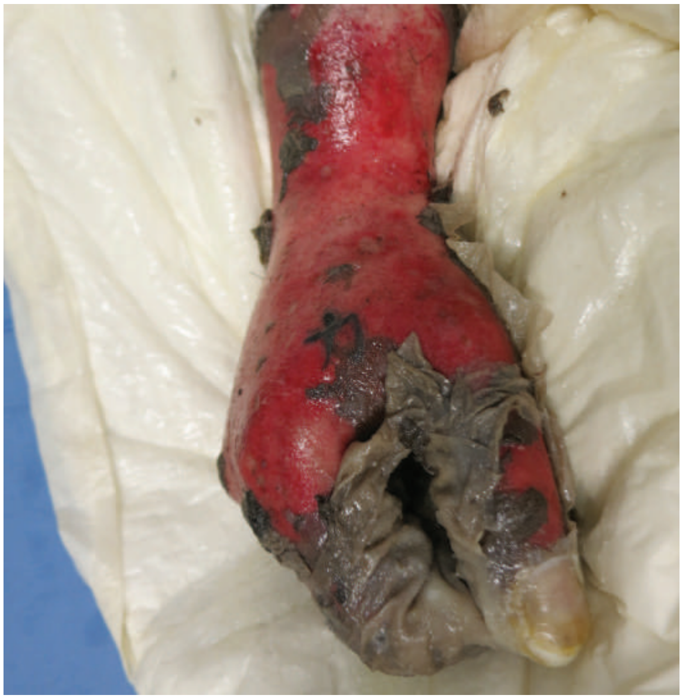

FIGURE 63-11 Toxic epidermal necrolysis, hand.

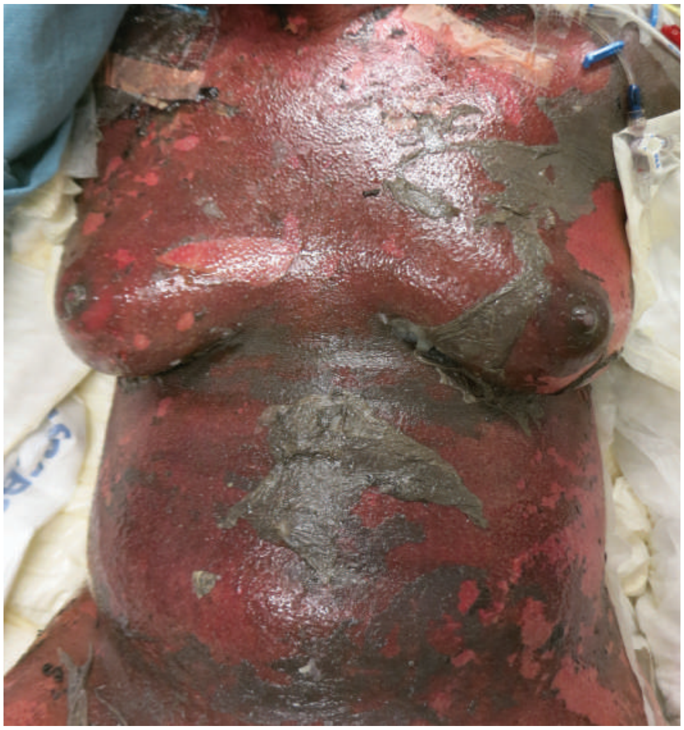

FIGURE 63-12 Toxic epidermal necrolysis.

target-like lesions (Figs. 63-13 and 63-14). Cancer and renal insufficiency are associated with a poor prognosis, as are older age and greater extent of epidermal detachment. At least 10% of those with SJS and 30% of those with TEN die from the disease. Drugs that most commonly cause SJS/TEN are sulfonamide antibiotics, allopurinol, antiepileptics (e.g., lamotrigine, phenytoin, carbamazepine), oxicam NSAIDs, β-lactam and other antibiotics, and nevirapine. Frozen-section skin biopsy may aid in rapid diagnosis.

At this time, there is no consensus on the most effective treatment for SJS/TEN. The best outcomes stem from early diagnosis, immediate discontinuation of the suspected drug, and meticulous supportive therapy in an intensive care or burn unit. Fluid management, atraumatic wound care, infection prevention and treatment, and ophthalmologic and respiratory support are critical. Early administration of systemic glucocorticoids, intravenous immunoglobulin, cyclosporine, or etanercept may improve disease outcomes, but randomized studies to evaluate potential therapies are lacking and difficult to perform.

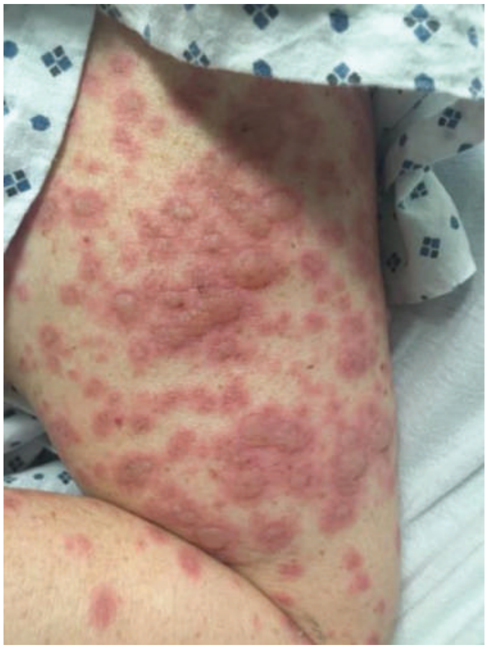

FIGURE 63-13 Target-like lesion in SJS.

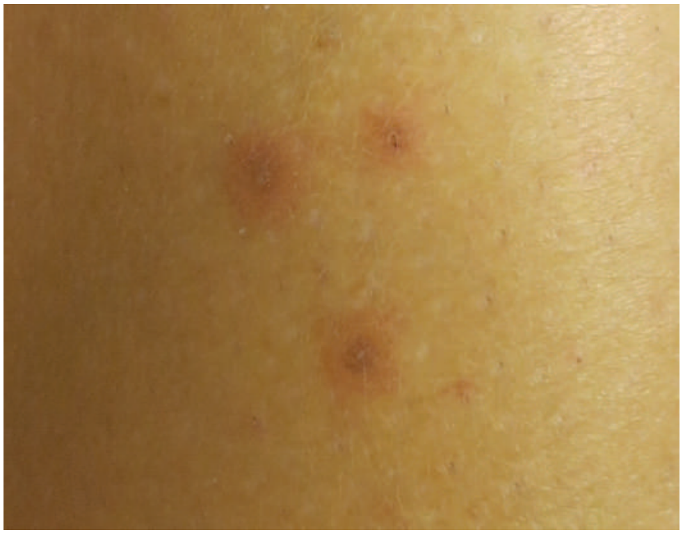

FIGURE 63-14 Atypical target-like lesion in SJS.

Pustular Eruptions AGEP is a rare reaction pattern affecting 3–5 people per million per year. It is thought to be secondary to medication exposure in >90% of cases (Fig. 63-15). Patients typically present with diffuse erythema or erythroderma, as well as fever and neutrophilia. One to two days later, innumerable pinpoint pustules develop overlying the erythema. The pustules are most pronounced in flexural and body fold areas; however, they may become generalized and, when coalescent, can lead to superficial erosion. In such cases, differentiating the eruption from SJS in its initial stages may be difficult, although in AGEP any erosions are more superficial, and prominent mucosal involvement is lacking.

In AGEP, both keratinocytes and T cells produce cytokines such as IL-8, IL-36a, and granulocyte-macrophage colony-stimulating factor (GM-CSF) that attract neutrophils to the skin, culminating in sterile pustule formation in the epidermis and neutrophilia in the peripheral blood. Skin biopsy shows collections of neutrophils and sparse necrotic keratinocytes in the upper part of the epidermis, unlike the full-thickness epidermal necrosis that characterizes SJS. Before the pustules appear, AGEP may also mimic DIHS/DRESS due to the prominent fever and erythroderma.

The principal differential diagnosis for AGEP is acute pustular psoriasis, which has an identical clinical and histologic appearance. Many patients with AGEP have a personal or family history of psoriasis. AGEP classically begins within 24–48 hours of drug exposure, although it may occur as much as 1–3 weeks later particularly with medications such as hydroxychloroquine or terbinafine. AGEP has

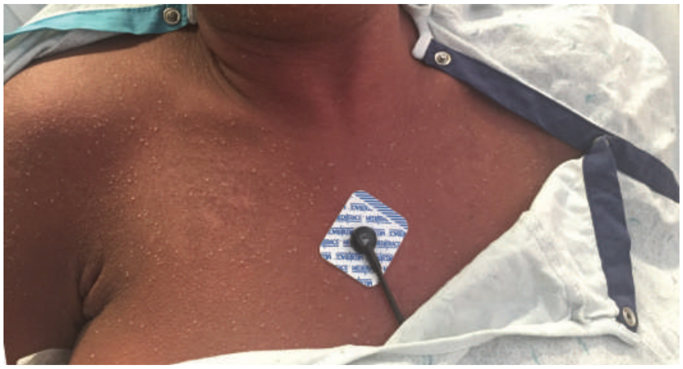

FIGURE 63-15 Acute generalized exanthematous pustulosis.

been associated with β-lactam antibiotics, calcium channel blockers, macrolide antibiotics, and other inciting agents (including radiocontrast and dialysates). Patch testing with the responsible drug often results in a localized pustular eruption mimicking the original reaction.

Overlap Hypersensitivity Syndromes An important concept in the clinical approach to severe drug eruptions is the existence of "overlap syndromes," most notably DIHS with TEN-like features, DIHS/DRESS with pustular eruption (AGEP-like), and AGEP with TEN-like features. In case series of AGEP, 50% of cases had TEN-like or DIHS/DRESS-like features, and 20% of cases had mucosal involvement resembling SJS/TEN. In one study, up to 20% of all severe drug eruptions had overlap features, suggesting that AGEP, DIHS/DRESS, and SJS/TEN can represent a clinical spectrum with some common pathophysiologic mechanisms. Designation of a single diagnosis based on cutaneous and extracutaneous involvement may not always be possible in cases of hypersensitivity; in such instances, treatment should be geared toward addressing the dominant clinical features as they evolve. The timing of rash onset with respect to drug administration, which is usually much more delayed in DIHS/DRESS than AGEP or SJS/TEN, and the presence of systemic manifestations such as hepatitis are helpful clues to that diagnosis.

Vasculitis Cutaneous small-vessel vasculitis (CSVV) typically presents with purpuric papules and macules involving the lower extremities and other dependent areas **(Fig. 63-16) (Chap. 375)**. Pustular and hemorrhagic vesicles as well as rounded ulcers also occur. Importantly, vasculitis may involve other organs, including the kidneys, joints, gastrointestinal tract, and lungs, necessitating a thorough clinical evaluation for systemic involvement. Drugs are implicated as a cause of roughly 15% of all cases of CSVV. Antibiotics, particularly β-lactams, are commonly implicated; however, almost any drug can cause vasculitis. Vasculitis may also be idiopathic or due to underlying infection, connective tissue disease, or (rarely) malignancy.

Rare but important types of drug-induced vasculitis include drug-induced ANCA vasculitis. Such patients commonly present with cutaneous manifestations but can develop the full range of symptoms associated with ANCA-associated vasculitis, including crescentic glomerulonephritis and alveolar hemorrhage. Propylthiouracil, methimazole, and hydralazine are common culprits. Drug-induced polyarteritis nodosa has been associated with long-term exposure to minocycline. The presence of perivascular eosinophils on skin biopsy can be a clue to possible drug etiology.

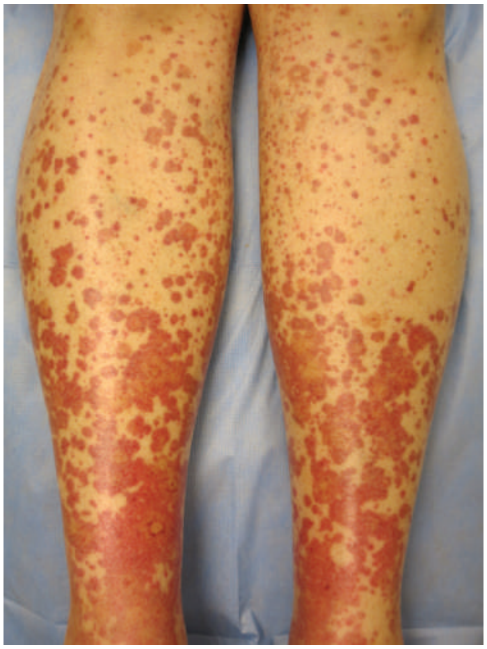

FIGURE 63-16 Cutaneous small-vessel vasculitis (CSVV, leukocytoclastic vasculitis).

MANAGEMENT OF THE PATIENT WITH SUSPECTED DRUG ERUPTION

There are four main questions to answer regarding a suspected drug eruption:

1. Is the observed rash caused by a medication?
2. Is the reaction severe or evolving with systemic manifestations?
3. Which drug or drugs are suspected, and should they be withdrawn?
4. What recommendation can be made for future medication use?

EARLY DIAGNOSIS OF SEVERE ERUPTIONS

Rapid recognition of potentially serious or life-threatening reactions is paramount. In this regard, a suspected drug eruption is best defined initially by whether or not it is one of these entities (e.g., SJS/TEN, DIHS/DRESS). **Table 63-2** lists clinical and laboratory features that, if present, suggest the presence of a severe reaction. **Table 63-3** lists the most important of these reactions, along with their key features and commonly associated medications. Any concern for a serious reaction should prompt immediate consultation with the appropriate specialist (e.g., a dermatologist or allergist and immunologist) and/or referral of the patient to a specialized center.

CONFIRMATION OF DRUG REACTION

The probability of drug etiology varies with the timing and pattern of the reaction. The latency period, or time from starting a drug until reaction onset, provides important context regarding the clinical phenotype. Immediate reactions typically occur within 6 hours of medication exposure, while delayed reactions generally occur days after the first exposure but can occur within hours following rechallenge. Morbilliform eruptions are usually viral in children and drug-induced in adults. Among severe reactions, drugs account for 10–20% of anaphylaxis and vasculitis and between 70% and 90% of AGEP, DIHS/DRESS, and SJS/TEN. Skin biopsy helps characterize the reaction but does not inform drug causality. Blood counts and liver and renal function tests are important for evaluating organ involvement. The findings of mild elevation of liver enzymes and eosinophil count are frequent but not specific for a drug reaction. Blood tests that could identify an alternative cause, serologic tests (to rule out drug-induced lupus), and serology or polymerase chain reaction for infections may be of great diagnostic importance.

TABLE 63-2 Clinical and Laboratory Findings Suggestive of Severe Cutaneous Adverse Drug Reaction

Cutaneous
Generalized rash
Dusky or target-like lesions
Purpura
Blisters or epidermal detachment
Positive Nikolsky sign
Skin pain
Skin necrosis
Erosions of the mucous membranes
Facial or acral edema
Swelling of the lips or tongue
General
Fever
Enlarged lymph nodes
Arthralgias or arthritis
Tachycardia, hypotension
Shortness of breath, hoarseness, wheezing
Laboratory Results
Eosinophil count >1000/μL
Leukopenia or leukocytosis with atypical lymphocytes
Abnormal liver or kidney function tests

From The New England Journal of Medicine, JC Roujeau, RS Stern: Severe adverse cutaneous reactions to drugs. 331, 1272-1285. Copyright © (1994) Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

TABLE 63-3 Clinical Features of Severe Cutaneous Drug Reactions

DIAGNOSIS	MUCOSAL LESIONS	TYPICAL SKIN LESIONS	FREQUENT SIGNS AND SYMPTOMS	MOST COMMON CULPRIT DRUGS
Stevens-Johnson syndrome (SJS)	Erosions usually of two or more sites	Dusky macules or atypical targets evolve into small blisters; rare areas of confluence; detachment ≤10% body surface area	Most cases involve fever	Trimethoprim-sulfamethoxazole, allopurinol, anticonvulsants, β-lactam antibiotics, nonsteroidal anti-inflammatory drugs (NSAIDs)
Toxic epidermal necrolysis (TEN)[a]	Erosions usually of two or more sites	Individual lesions like those in SJS evolving into confluent dusky erythema; large sheets of necrotic epidermis; total detachment of >30% body surface area	Fever, leukopenia	Same as for SJS
Drug-induced hypersensitivity syndrome/drug rash with eosinophilia and systemic symptoms (DIHS/DRESS)	Mucositis reported in up to 30%	Diffuse, deep red morbilliform eruption with facial involvement; facial and acral swelling	Fever, lymphadenopathy; eosinophilia, atypical lymphocytosis, hepatitis, nephritis, myocarditis	Anticonvulsants, sulfonamides, allopurinol, minocycline, vancomycin
Acute generalized exanthematous pustulosis (AGEP)	Oral erosions in perhaps 20%	Diffuse erythematous eruption; innumerable pinpoint pustules with preference for body folds; superficial peeling or erosions	High fever, leukocytosis (neutrophilia), hypocalcemia	β-Lactam antibiotics, macrolide antibiotics, calcium channel blockers, IV contrast
Serum sickness or serum sickness–like reaction	Absent	Urticarial rash, often serpiginous or polycyclic; purpuric eruption along the sides of the feet and hands is characteristic	Fever, arthralgias, lymphadenopathy	Antithymocyte globulin, anti-toxins, rituximab, other monoclonal antibodies; cefaclor, penicillin, amoxicillin, trimethoprim-sulfamethoxazole
Anticoagulant-induced necrosis	Infrequent	Purpura and necrosis, especially of central, fatty areas	Ischemic pain in affected areas	Warfarin, heparin
Angioedema	Often involved	Swelling, erythema, and discomfort of skin and subcutaneous tissue, generally associated with urticaria	Stridor, respiratory distress, cardiovascular collapse	Angiotensin-converting enzyme (ACE) inhibitors, NSAIDs, contrast dye

Note: SJS-TEN overlap is characterized by detachment of 10–30% of body surface area.

From The New England Journal of Medicine, JC Roujeau, RS Stern: Severe adverse cutaneous reactions to drugs. 331, 1272-1285. Copyright © (1994) Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

WHAT DRUG(S) TO SUSPECT AND WITHDRAW

Most drug eruptions occur during the first course of treatment with a new medication. A notable exception is IgE-mediated urticaria and anaphylaxis that require presensitization and develop a few minutes to hours after rechallenge. Characteristic timing of onset of delayed hypersensitivity reactions following drug administration is as follows: 4–14 days for morbilliform drug eruption, 2–4 days for AGEP, 1.5 days to 3 weeks for AHS, 5–28 days for SJS/TEN, and 14–48 days for DIHS **(Fig. 63-17)**. A drug timeline, compiling comprehensive information of all current and past prescription and nonprescription medications/supplements and the timing of administration relative to the rash, is a key diagnostic tool for identifying the inciting drug. Medications introduced for the first time in the relevant time frame are prime suspects. Two other important elements to suspect causality at this stage are (1) previous experience with the drug (or related members of the same pharmacologic class) and (2) alternative etiologic candidates.

The decision to continue or discontinue any medication depends on the severity of the reaction, the severity of the primary disease undergoing treatment, the degree of suspicion of causality, and the feasibility of finding an alternative safer treatment. In any potentially fatal drug reaction, elimination of all possible suspect drugs and unnecessary medications should be immediately attempted. Some rashes may resolve when "treating through" a benign drug-related eruption. The decision to treat through an eruption should, however, remain the exception based on a higher benefit-risk ratio and withdrawal of every suspect drug the general rule. On the other hand, drugs that are associated with a low pretest probability of a hypersensitivity reaction and are important for the patient (e.g., antihypertensive agents), or have been tolerated by the patient for more than 12 weeks, generally should not be quickly withdrawn. This approach may permit judicious use of these agents in the future.

RECOMMENDATION FOR FUTURE USE OF DRUGS

The aims are to (1) prevent the recurrence of the drug eruption and (2) avoid compromising future treatment by inaccurately excluding otherwise useful medications.

A thorough assessment of drug causality is based on timing of the reaction, evaluation of other possible causes, and effect of drug withdrawal or continuation. The RegiSCAR group has proposed the Algorithm of Drug Causality for Epidermal Necrolysis (ALDEN) to rank likelihood of drug causality in SJS/TEN; validation of this and other instruments, such as the Naranjo adverse drug reaction probability scale, is limited. Medication(s) with a "definite" or "probable" causality should be contraindicated, a warning card or medical alert tag (e.g., wristband) should be given to the patient, and the drugs should be listed in the patient's medical chart as serious allergic reactions warranting permanent drug avoidance.

CROSS-SENSITIVITY

Because of possible cross-sensitivity among chemically related drugs, many physicians recommend avoidance of not only the medication that induced the reaction but also all drugs of the same pharmacologic class.

There are two types of cross-sensitivity. Reactions that depend on a pharmacologic interaction may occur with all drugs that target the same pathway, whether the drugs are structurally similar or not. This is the case with angioedema caused by COX-1 inhibitors (aspirin/NSAIDs) and ACE inhibitors. In this situation, the risk of recurrence varies from drug to drug in a particular class; however, avoidance of all drugs in the class is usually recommended. Immune recognition of structurally related drugs is the second mechanism by which cross-sensitivity occurs. A classic example is hypersensitivity to aromatic antiepileptics (barbiturates, phenytoin, carbamazepine) with up to 50% of patients who reacted to one drug reacting to a second. For other drugs, in vitro and in vivo data have suggested that cross-reactivity exists only between compounds with very similar chemical structures. Sulfamethoxazole-specific lymphocytes may be activated by other antibacterial sulfonamides but not diuretics, antidiabetic drugs, or anti-COX2 NSAIDs that have a sulfonamide group (SO_2-NH_2) but do not share the N4 substituted aryl-amine or the aromatic ring at N1 characteristic of antibacterial sulfonamides. Though it has been previously reported that 10% of patients with penicillin allergies will also develop allergic reactions to cephalosporin class antibiotics, the cross-reactivity is now thought to be mainly driven by the shared R1

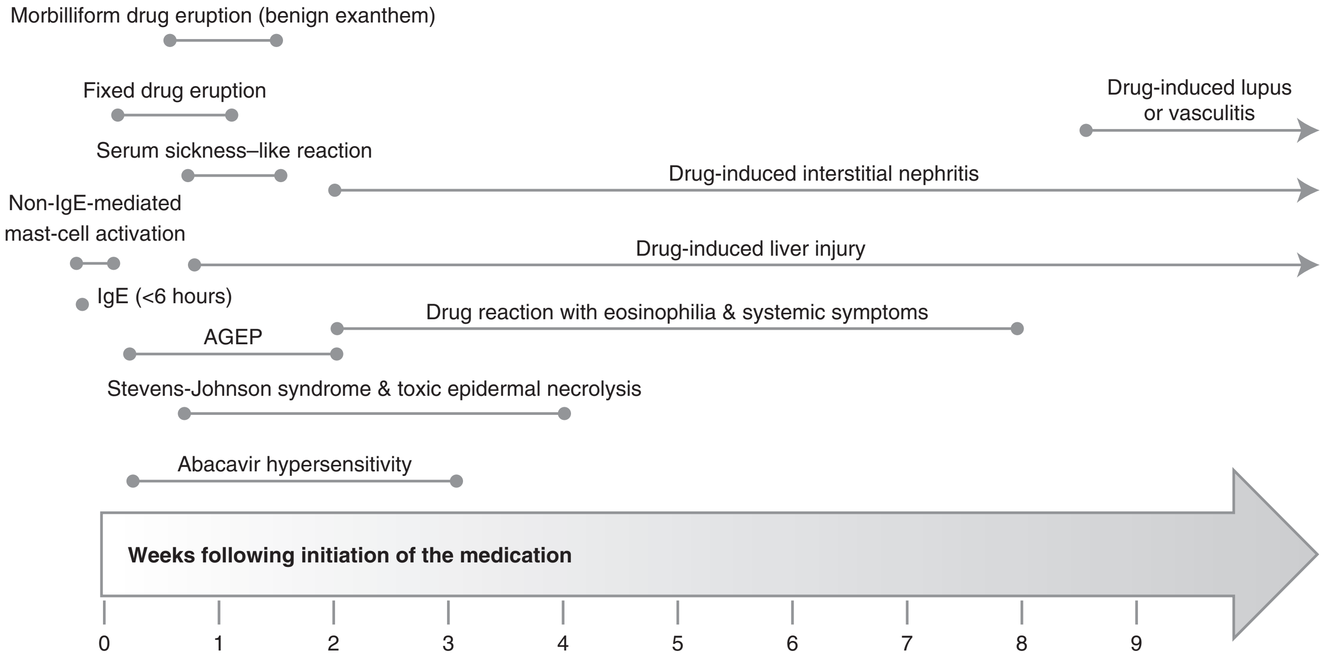

FIGURE 63-17 Timeline of drug hypersensitivity reactions. Drug reactions differ in their latency period, which is the time from first ingestion of a medication to the time a reaction occurs. For IgE- and non–IgE-mediated reactions, these are typically immediate and occur less than 6 hours from ingestion. Delayed reactions typically occur several hours to days or even weeks following first ingestion. The latency period can be a valuable clue to the clinical phenotype of the drug reaction and the culprit drug. For example, acute generalized exanthematous pustulosis (AGEP) associated with antibiotics typically occurs within 24–48 hours, whereas when AGEP is associated with hydroxychloroquine, it may take up to 3 weeks. Drug-induced hypersensitivity syndrome (DIHS)/drug reaction with eosinophilia and systemic symptoms (DRESS) to most drugs has a latency period of 2–3 weeks, although shorter latencies associated with β-lactam antibiotics have been described. Because multiple drugs are frequently started together in a complex medical patient, a timeline outlining all medications taken at the first time symptoms of a reaction occur and documentation of the evolution of these symptoms in relation to initiation of specific medications are valuable in clinical phenotype and drug causality assessment. *(Reproduced with permission from DA Khan et al: Drug allergy: A 2022 practice parameter update. J Allerg Clin Immun 150:1333, 2022, Figure 1.)*

side chain between specific aminopenicillins and aminocephalosporins. The fact that more than 95% of patients labeled with a penicillin allergy do not have a true allergy and that cross-reactivity between penicillins and unrelated cephalosporin is less than 2% means severe reactions are very rare.

Data suggest that although the risk of developing a drug eruption to another drug is increased in persons with a prior reaction, "cross-sensitivity" is probably not the explanation. For example, those with a history of an allergic-like reaction to penicillin are at greater risk of developing a reaction to antibacterial sulfonamides than to cephalosporins.

These data suggest the list of drugs to avoid after a drug reaction should be limited to the causative one(s) and to a few very similar medications.

Because of growing evidence that some severe cutaneous reactions to drugs are associated with HLA genes that are inherited co-dominantly (one set of alleles is active from each parent), consideration should be given to counsel first-degree family members of patients with severe cutaneous reactions about genetic testing and potential medication avoidance.

ROLE OF TESTING FOR CAUSALITY AND DRUG RECHALLENGE

The usefulness of laboratory tests, skin-prick, or patch testing to determine causality is highly dependent on the specific medication and clinical phenotype of the drug hypersensitivity reaction. Many in vitro immunologic assays have been developed for research purposes; however, the predictive value of these tests has not been validated in large series of affected patients. In some cases, where the benefit of treatment outweighs the risk of rechallenge, diagnostic rechallenge may be appropriate, even for drugs with high rates of adverse reactions. Supportive data from resource-poor settings suggest careful sequential additive rechallenge can be achieved with anti-tuberculous medications, with immediate high-dose methylprednisolone rescue used successfully to abort any reaction.

Skin-prick testing has clinical value in specific settings. In patients with a history suggesting immediate IgE-mediated reactions to penicillin, skin-prick testing with the specific culprit penicillins and major and minor determinants of penicillins or cephalosporins has proven useful for identifying patients at risk of anaphylactic reactions to these agents. Negative skin tests do not totally rule out IgE-mediated reactivity; however, the risk of anaphylaxis in response to penicillin administration in patients with negative skin tests is about 1%. In contrast, two-thirds of patients with a positive skin test experience an allergic response upon rechallenge. The skin tests themselves carry a small risk of anaphylaxis in patients who have had recent IgE-mediated reactions. Many patients with childhood reactions to antibiotics such as penicillin were never allergic, and even true IgE-mediated responses are thought to wane over time. For low-risk penicillin allergic adults and children with remote histories or whose history consisted of cutaneous symptoms only (e.g., urticaria) and where symptoms resolved without treatment, evidence currently supports direct oral challenge with a penicillin without preceding skin testing. This approach of direct oral challenge has also been applied to remove the allergy label to low-risk reactions associated with sulfonamides and other antibiotics.

For patients with delayed-type hypersensitivity, the clinical utility of skin tests is dependent on the clinical phenotype of the reaction. Evidence and standardization of approaches including the concentration of medications used in skin and patch testing are needed as the lower concentrations used in prick and intradermal testing for immediate IgE-mediated reactions may have limited utility for dose-dependent T-cell–mediated reactions. At least one of a combination of several tests (prick, patch, and intradermal) is positive in 50–70% of patients with a delayed reaction "definitely" attributed to a single medication. This low sensitivity corresponds to the observation that readministration of drugs with negative skin testing results in eruptions in 17% of cases.

Desensitization can be considered in those with a history of reaction to a medication that must be used again. Efficacy of such procedures has been demonstrated in cases of immediate reaction to penicillin and positive skin tests, anaphylactic reactions to platinum chemotherapy, and delayed reactions to sulfonamides in patients with AIDS. Desensitization is often successful in HIV-infected patients with morbilliform eruptions to sulfonamides but is not recommended in HIV-infected patients who developed erythroderma or a bullous reaction in response to prior sulfonamide exposure. Various protocols are available, including oral and parenteral approaches. Oral desensitization appears to have a lower risk of serious anaphylactic reaction. Desensitization carries the risk of anaphylaxis regardless of how it is performed and should be conducted in monitored clinical settings such as an intensive care unit in these higher-risk patients. After desensitization, many patients experience non-life-threatening reactions during therapy with the culprit drug.

REPORTING

Any severe reaction to a drug should be reported to a regulatory agency or to pharmaceutical companies. Because severe reactions may be too rare to be detected in premarketing clinical trials, spontaneous reports are of critical importance for early detection of unexpected life-threatening events. To be useful, the report should contain enough details to permit ascertainment of severity and drug causality.

FURTHER READING

Alfirevic A et al: Genetic testing for prevention of severe drug-induced skin rash. Cochrane Database Syst Rev 7:CD010891, 2019.

Blumenthal KG et al: Antibiotic allergy. Lancet 393:183, 2019.

Fay C et al: Diagnosis and management of dermatologic adverse events from systemic melanoma therapies. Am J Clin Dermatol 24:765, 2023.

Gibson A et al: Updates on the immunopathology and genomics of severe cutaneous adverse drug reactions. J Allergy Clin Immunol 151:289, 2023.

Jacobsen A et al: Systemic interventions for treatment of SJS, TEN, and SJS/TEN overlap syndrome. Cochrane Database Sys Rev 3:CD013130, 2022.

Jutel M et al: Nomenclature of allergic diseases and hypersensitivity reactions: Adapted to modern needs: An EAACI position paper. Allergy 78:2851, 2023.

Khan DA et al: Drug allergy: A 2022 practice parameter update. J Allergy Clin Immunol 150:1333, 2022.

Liu S et al: Association between medication use and bullous pemphigoid. JAMA Dermatol 156:891, 2020.

Seminario-vidal L et al: Society of Dermatology Hospitalists supportive care guidelines for the management of Stevens-Johnson syndrome/toxic epidermal necrolysis in adults. J Am Acad Dermatol 82:1553, 2020.

Wang CW et al: An updated review of genetic associations with severe adverse drug reactions: Translation and implementation of pharmacogenomic testing in clinical practice. Front Pharmacol 13:886377, 2022.

Watanabe T, Yamaguchi Y: Cutaneous manifestations associated with immune checkpoint inhibitors. Front Immunol 14:1071983, 2023.

64

Photosensitivity and Other Reactions to Sunlight

Alexander G. Marneros, David R. Bickers

SOLAR RADIATION

Sunlight is the most visible and obvious source of comfort in the environment. The sun provides the beneficial effects of warmth and vitamin D synthesis. However, acute and chronic sun exposure also has pathologic consequences. Cutaneous exposure to sunlight is a major cause of human skin cancer and can have immunosuppressive effects as well.

The sun's energy reaching the Earth's surface is limited to components of the ultraviolet (UV) spectrum, the visible spectrum, and portions of the infrared spectrum. The cutoff at the short end of the UV spectrum at ~290 nm is due primarily to stratospheric ozone—formed by highly energetic ionizing radiation—that prevents penetration to the earth's surface of the shorter, more energetic, potentially more harmful wavelengths of solar radiation. Indeed, concern about destruction of the ozone layer by chlorofluorocarbons released into the atmosphere has led to international agreements to reduce production of those chemicals.

Measurements of solar flux showed a 20-fold regional variation in the amount of energy at 300 nm that reaches the earth's surface. This variability relates to seasonal effects, the path that sunlight traverses through ozone and air, the altitude (a 4% increase for each 300 m of elevation), the latitude (increasing intensity with decreasing latitude), and the amount of cloud cover, fog, and pollution.

The major components of the photobiologic action spectrum that can affect human skin include the UV and visible wavelengths between 290 and 700 nm. In addition, the wavelengths beyond 700 nm in the infrared spectrum primarily emit heat and in certain circumstances may exacerbate the pathologic effects of energy in the UV and visible spectra.

The UV spectrum reaching the Earth represents <10% of total incident solar energy and is arbitrarily divided into two major segments, UV-B and UV-A, which constitute the wavelengths from 290 to 400 nm. UV-B consists of wavelengths between 290 and 320 nm. This portion of the photobiologic action spectrum is the most efficient in producing redness or erythema in human skin and thus is sometimes known as the "sunburn spectrum." UV-A includes wavelengths between 320 and 400 nm and is ~1000-fold less efficient in producing skin erythema than is UV-B.

The wavelengths between 400 and 700 nm are visible to the human eye. The photon energy in the visible spectrum is incapable of damaging human skin in the absence of a photosensitizing chemical. Without energy absorption by a molecule, there can be no photosensitivity. Thus, the *absorption spectrum* of a molecule is defined as the range of wavelengths it absorbs, whereas the *action spectrum* for an effect of incident radiation is defined as the range of wavelengths that evoke the response.

Photosensitivity occurs when a photon-absorbing chemical (*chromophore*) present in the skin absorbs incident energy, becomes excited, and transfers the absorbed energy to various structures or to molecular oxygen.

UV RADIATION (UVR) AND SKIN STRUCTURE AND FUNCTION

Human skin consists of two major compartments: the outer epidermis, which is a stratified squamous epithelium, and the underlying dermis, which is rich in matrix proteins such as collagens and elastin. Both compartments are susceptible to damage from sun exposure. The epidermis and the dermis contain multiple chromophores capable of absorbing incident solar energy, including nucleic acids, proteins, and lipids. The outermost epidermal layer, the stratum corneum, is a major absorber of UV-B, and <10% of incident solar UV-B wavelengths penetrate from the epidermis to the dermis. Approximately 3% of radiation below 300 nm, 20% of radiation below 360 nm, and 33% of short visible radiation reach the basal cell layer in untanned human skin. UV-A readily penetrates to the dermis and is capable of altering structural and matrix proteins that contribute to photoaging of chronically sun-exposed skin, particularly in individuals of light complexion. In essence, longer wavelengths can penetrate more deeply into the skin.

Molecular Targets for UVR-Induced Skin Effects Epidermal DNA—predominantly in keratinocytes and in Langerhans cells (dendritic antigen-presenting cells)—absorbs UV-B and undergoes structural changes between adjacent pyrimidine bases (thymine or cytosine), including the formation of cyclobutane dimers and 6,4-photoproducts. These structural changes are potentially mutagenic and are found in nonmelanoma skin cancers (NMSCs), including basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and Merkel cell carcinoma (MCC). They can be repaired by cellular mechanisms that recognize, repair, and restore normal base sequences. The efficient repair of these structural entities is crucial, since individuals with defective DNA repair are at high risk for the development of cutaneous cancer. For example, patients with xeroderma pigmentosum, an autosomal recessive disorder, have a variably deficient repair of UV-induced photoproducts. The skin of these patients often shows the dry, leathery appearance of prematurely photoaged skin, and these patients have an increased frequency of skin cancer already in the first two decades of life. Studies in transgenic mice have verified the importance of functional genes that regulate these repair pathways in preventing the development of UV-induced skin cancer. DNA damage to Langerhans cells may also contribute to the known immunosuppressive effects of UV-B (see "Photoimmunology," later).

In addition to DNA, molecular oxygen is a major target for incident solar UVR, leading to the generation of reactive oxygen species (ROS) capable of damaging skin components through DNA oxidative damage, oxidation of polyunsaturated fatty acids in lipids (lipid peroxidation), or oxidation of amino acids in proteins, or they can lead to oxidative deactivation of specific enzymes. UVR can also promote increased cross-linking and degradation of dermal matrix proteins and accumulation of abnormal dermal elastin, leading to photoaging changes known as *solar elastosis*.

Cutaneous Optics and Chromophores *Chromophores* are endogenous or exogenous chemicals that can absorb physical energy. Endogenous chromophores are of two types: (1) normal components of skin, including nucleic acids, proteins, lipids, and 7-dehydrocholesterol (the precursor of vitamin D); and (2) components that are synthesized elsewhere in the body such as porphyrins that circulate in the bloodstream and diffuse into the skin. Normally, only trace amounts of porphyrins are present in the skin, but, in selected diseases known as the *porphyrias* **(Chap. 428)**, porphyrins are released into the circulation in increased amounts from the bone marrow and/or the liver and are transported to the skin, where they absorb incident energy both in the Soret band (~400 nm; short visible) and, to a lesser extent, in the red portion of the visible spectrum (580–660 nm). This energy absorption results in the generation of ROS that can mediate structural damage to the skin, manifested as erythema, edema, urticaria, or blister formation. It is of interest that photoexcited porphyrins are currently used in the treatment of BCCs and SCCs and their precursor lesions, actinic keratoses. Known as *photodynamic therapy* (PDT), this modality generates ROS in the skin, leading to cell death. Topical photosensitizers used in PDT include the porphyrin precursors 5-aminolevulinic acid and methyl aminolevulinate, which are readily converted to porphyrins in the skin. It is believed that PDT targets tumor cells for destruction more selectively than it targets adjacent nonneoplastic cells. The efficacy of such therapy requires appropriate timing of the application of methyl aminolevulinate or 5-aminolevulinic acid to the affected skin followed by exposure to artificial sources of visible light. High-intensity blue light has been used successfully for PDT of thin actinic keratoses. Red light PDT penetrates more deeply into the skin and is more beneficial in the treatment of superficial BCCs.

Acute Effects of Sun Exposure The acute effects of skin exposure to sunlight include sunburn and vitamin D synthesis.

SUNBURN This painful skin condition is an acute inflammatory response of the skin, predominantly to UV-B. Generally, an individual's ability to tolerate sunlight is inversely proportional to that individual's degree of melanin pigmentation. Melanin, a complex polymer of tyrosine derivatives, is synthesized in specialized epidermal dendritic cells known as *melanocytes* and is packaged into *melanosomes* that are transferred via dendritic processes into *keratinocytes*, thereby providing photoprotection (dissipating the vast majority of absorbed UVR in the skin) and simultaneously darkening the skin. Sun-induced melanogenesis is a consequence of increased tyrosinase activity in melanocytes. Central to the suntan response is the melanocortin-1 receptor (*MC1R*), and mutations in this gene contribute to genetic variations in human skin and hair color; individuals with red hair and fair skin typically have low MC1R activity. In the skin, there are two main types of melanin: eumelanin (providing brown and black pigmentation associated with high MC1R activity) and pheomelanin (providing red pigmentation associated with low MC1R activity). Pheomelanin is a cysteine-containing red polymer of benzothiazine units and has much weaker shielding capacity against UVR compared to eumelanin. This may explain why individuals with a higher proportion of pheomelanin (red hair/fair skin appearance) have an increased risk of melanoma formation. In addition, pheomelanin may also promote melanoma formation through induction of oxidative damage by amplifying UV-A–induced ROS but also through UVR-independent mechanisms.

The human *MC1R* gene encodes a G protein–coupled receptor that binds α-melanocyte-stimulating hormone (α-MSH), which is secreted in the skin mainly by keratinocytes in response to UVR. The UV-induced expression of this hormone is controlled by the tumor suppressor p53, and absence of functional p53 attenuates the tanning response. Activation of the melanocortin receptor leads to increased intracellular cyclic adenosine 5′-monophosphate (cAMP) and protein kinase A activation, resulting in increased transcription of the microphthalmia-associated transcription factor (MITF), which stimulates melanogenesis. Since the precursor of α-MSH, proopiomelanocortin produced by keratinocytes, is also the precursor of β-endorphins, UVR may result in not only increased pigmentation but also increased β-endorphin production in the skin, an effect that has been hypothesized to promote sun-seeking behaviors and even mediate addiction to tanning.

The Fitzpatrick classification of human skin phototypes is based on the efficiency of the epidermal-melanin unit, which usually can be ascertained by asking an individual two questions relating to sun exposure: (1) Do you burn after sun exposure? (2) Do you tan after sun exposure? The answers to these questions permit division of the population into six skin types, varying from type I (always burn, never tan) to type VI (never burn, always tan) **(Table 64-1)**.

Sunburn erythema is due to vasodilation of dermal blood vessels. There is a lag time (usually 4–12 h) between skin exposure to sunlight and the development of visible redness. The action spectrum for sunburn erythema includes UV-B and UV-A, although UV-B is much more efficient than UV-A in evoking the response. However, UV-A may contribute to sunburn erythema at midday, when much more UV-A than UV-B is present in the solar spectrum. The erythema that accompanies the inflammatory response induced by UVR results from the orchestrated release of cytokines along with growth factors and the generation of ROS. Furthermore, UV-induced activation of nuclear factor κB–dependent gene transcription augments release of several proinflammatory cytokines and vasoactive mediators. These cytokines and mediators accumulate locally in sunburned skin, providing chemotactic factors that attract neutrophils, macrophages, and T lymphocytes, which promote the inflammatory response. UVR also stimulates infiltration of inflammatory cells through induced expression of adhesion molecules such as E-selectin and intercellular adhesion molecule 1 on endothelial cells and keratinocytes. UVR has been shown to activate phospholipase A_2, resulting in increases in eicosanoids such as prostaglandin E_2, which is known to be a potent inducer of sunburn erythema. The role of eicosanoids in this reaction has been verified by studies showing that nonsteroidal anti-inflammatory drugs (NSAIDs) can reduce sunburn erythema.

TABLE 64-1 Skin Type and Sunburn Sensitivity (Fitzpatrick Classification)

TYPE	DESCRIPTION
I	Always burn, never tan
II	Always burn, sometimes tan
III	Sometimes burn, sometimes tan
IV	Sometimes burn, always tan
V	Never burn, sometimes tan
VI	Never burn, always tan

Epidermal changes in sunburn include the induction of "sunburn cells," which are keratinocytes undergoing p53-dependent apoptosis as a defense against skin cancer by eliminating cells that harbor UV-B–induced structural DNA damage.

VITAMIN D SYNTHESIS AND PHOTOCHEMISTRY Cutaneous exposure to UV-B causes photolysis of epidermal 7-dehydrocholesterol, converting it to pre–vitamin D_3, which then undergoes temperature-dependent isomerization to form the stable hormone vitamin D_3. This compound diffuses to the dermal vasculature and circulates to the liver and kidney, where it is converted to the dihydroxylated functional hormone 1,25-dihydroxyvitamin D_3. Vitamin D metabolites from the circulation and those produced in the skin itself can augment epidermal differentiation signaling and inhibit keratinocyte proliferation. These effects are exploited therapeutically in psoriasis with the topical application of synthetic vitamin D analogues. In addition, vitamin D is increasingly thought to have beneficial effects in several other inflammatory conditions, and some evidence suggests that—besides its classic physiologic effects on calcium metabolism and bone homeostasis—it is associated with a reduced risk of various internal malignancies. There is controversy regarding the risk-to-benefit ratio of sun exposure for vitamin D homeostasis. At present, it is important to emphasize that no clear-cut evidence suggests that the use of sunscreens substantially diminishes vitamin D levels. Since chronologic aging also substantially decreases the ability of human skin to photocatalytically produce vitamin D_3, the widespread use of sunscreens that filter out UV-B has led to concerns that the elderly might be unduly susceptible to vitamin D deficiency. However, the amount of sunlight needed to produce sufficient vitamin D is small and does not justify the risks of skin cancer and other types of photodamage linked to increased sun exposure or tanning behavior. Nutritional supplementation of vitamin D is a preferable strategy for patients with vitamin D deficiency.

Chronic Effects of Sun Exposure: Nonmalignant The clinical features of photoaging (*dermatoheliosis*) include wrinkling, blotchiness, and telangiectasia, as well as a roughened, irregular, "weather-beaten" leathery appearance.

UVR is important in the pathogenesis of photoaging in human skin, and ROS are likely involved. The dermis and its connective tissue matrix are major targets for sun-associated chronic damage that manifests as solar elastosis, a massive increase in thickened irregular masses of abnormal-appearing elastic fibers. Collagen fibers are also abnormally clumped in the deeper dermis of sun-damaged skin. The chromophores, the action spectra, and the specific biochemical events orchestrating these changes are only partially understood, although more deeply penetrating UV-A seems to be primarily involved. Chronologically aged sun-protected skin and photoaged skin share important molecular features, including connective tissue damage and elevated levels of matrix metalloproteinases (MMPs). These are enzymes involved in the degradation of extracellular structural matrix proteins. UV-A induces expression of some MMPs, including MMP-1 and MMP-3, leading to increased collagen breakdown. In addition, UV-A reduces type I procollagen messenger RNA (mRNA) expression. Thus, chronic UVR alters the structure and function of dermal collagen both by inhibiting its synthesis and enhancing its breakdown. Based on these observations, it could explain why high-dose UV-A phototherapy may have beneficial effects in some patients with localized fibrotic diseases of the skin, such as localized scleroderma.

Chronic Effects of Sun Exposure: Malignant One of the major known consequences of chronic excessive skin exposure to sunlight is NMSC, including SCCs, BCCs, and MCCs **(Chap. 81)**. A model for skin cancer induction involves three major steps: initiation, promotion, and progression. Exposure of human skin to sunlight results in *initiation*, a step by which structural (mutagenic) changes in DNA evoke an irreversible mutagenic alteration in the target cell (keratinocyte) that begins the tumorigenic process. Exposure to a tumor initiator such as UV-B is believed to be a necessary but not a sufficient step in the malignant process, since initiated skin cells not exposed to tumor promoters generally do not develop into tumors. The second stage in tumor development is *promotion*, a multistep process by which chronic exposure to sunlight evokes further changes that culminate in the clonal expansion of initiated cells and cause the development of premalignant growths known as *actinic keratoses*, which may progress to form SCCs. As a result of extensive studies, it seems clear that UV-B is a *complete carcinogen*, meaning that it can act as both a tumor initiator and a tumor promoter. The third and final step in the malignant process is *malignant conversion* of benign precursors into cancers, a process thought to promote genetic instability.

On a molecular level, skin carcinogenesis results from the accumulation of gene mutations that cause inactivation of tumor suppressors, activation of oncogenes, or reactivation of cellular signaling pathways that normally are expressed only during embryologic epidermal development that drive aberrant cell proliferation. Interestingly, a large number of UV-induced oncogenic driver mutations that are present in SCCs are readily detectable in aged sun-exposed normal skin, leading to a growth advantage and innumerable precancerous clones carrying cancer-causing mutations. These mutations occur particularly often in genes that affect proliferation of epidermal stem cells (e.g., NOTCH receptor genes). The pattern of oncogenic gene mutations in aged sun-exposed skin shows considerable overlap with the mutations identified in SCCs, while there is little overlap with the mutations identified in BCCs or melanomas. For example, ~20% of normal aged sun-exposed skin cells and ~60% of SCCs carry driver mutations in *NOTCH1*. Additionally, the accumulation of mutations in the tumor-suppressor gene *p53* can also promote skin carcinogenesis. Indeed, the majority of both human and murine UV-induced skin cancers have characteristic UVR-induced *p53* mutations (C → T and CC → TT transitions). Studies in mice have shown that sunscreens can substantially reduce the frequency of these signature mutations in *p53* and inhibit the induction of tumors. The comparison of UVR-induced gene mutations between aged sun-exposed normal skin and SCCs supports the hypothesis of a progressive accumulation of additional oncogenic mutations that eventually lead to the transition from precancerous cell clones to SCCs. It has been estimated that SCCs harbor ~10 times more oncogenic driver mutations per cell than cells in aged sun-exposed normal skin. Furthermore, while aged sun-exposed skin and SCCs carry similar UVR-induced mutations in *p53* or NOTCH receptors, oncogenic mutations in other genes (e.g., *CDKN2A*) were mainly found in SCCs and not in aged sun-exposed skin, which are thus likely to play a critical role in malignant progression.

Compared to SCCs, BCCs carry a distinct mutational profile in specific genes. BCCs harbor inactivating mutations particularly in the tumor-suppressor gene *patched* or activating mutations in the oncogene *smoothened*, which result in the constitutive activation of the sonic hedgehog signaling pathway and increased cell proliferation. There is also evidence linking alterations in the Wnt/β-catenin signaling pathway, which is known to be critical for hair follicle development and skin cancer as well. Thus, interactions between this pathway and the hedgehog signaling pathway appear to be involved in both skin carcinogenesis and embryologic development of the skin and hair follicles.

Clonal analysis in mouse models of BCC revealed that tumor cells arise from stem cells of the interfollicular epidermis and the upper infundibulum of the hair follicle. These BCC-initiating cells are reprogrammed to resemble embryonic hair follicle progenitors, whose tumor-initiating ability depends on activation of the Wnt/β-catenin signaling pathway.

SCC initiation occurs both in the interfollicular epidermis and in the hair follicle bulge stem cell populations. In mouse models, the combination of mutant K-Ras and p53 is sufficient to induce invasive SCCs from these cell populations. The transcription factor Myc is important for stem cell maintenance in the skin, and oncogenic activation of Myc has been implicated in the development of both BCCs and SCCs.

The third NMSC is MCC, which is named after its resemblance to Merkel cells in the skin. The incidence of MCCs has been increasing in recent years for unknown reasons. The age-adjusted global incidence is about 1 in 100,000. Just like SCC and BCC, patients with MCCs are usually fair-skinned males in the sixth to eighth decades of life who are living in geographic regions with greater solar UVR. These tumors occur predominantly on the head and neck in older individuals and on the trunk in younger people. MCCs also have a higher incidence among immunosuppressed patients. MCCs are aggressive, life-threatening, poorly differentiated neuroendocrine carcinomas. Overall survival at 5 years is around 50% for local disease, 35% for nodal disease, and 15% for metastatic disease. While the majority of MCCs present locally, nodal and metastatic disease can occur simultaneously. The pathogenesis of MCCs is closely connected to the Merkel cell polyoma virus (MCPyV). It is now recognized that MCCs can either be MCPyV positive or MCPyV negative. MCPyV-negative MCCs manifest high levels of classic UV-induced signature mutations (C to T or CC to TT) and inactivation of tumor suppressor genes, which could explain the growth of these viral-negative lesions. MCPyV-positive tumors are thought to grow secondary to viral integration into the host genome and acquisition of a truncating mutation of the large T antigen that results in the production of viral oncoproteins. The growth of MCPyV-positive tumors may be further promoted by UVR-induced local immunosuppression. Both forms of MCC are immunogenic, and metastatic MCC has been treated in some patients successfully with PD-1/PD-L1 immune checkpoint inhibitors.

In summary, NMSC involves mutations and alterations in multiple genes and pathways that occur as a result of their chronic accumulation driven by exposure to environmental factors such as solar UVR.

Epidemiologic studies have linked excessive sun exposure to an increased risk of NMSCs and melanoma of the skin; the evidence is far more direct for NMSCs (BCCs, SCCs, and MCCs) than for melanoma. Approximately 80% of NMSCs develop on sun-exposed body areas, including the face, neck, and hands. Major risk factors include male sex, childhood sun exposures, older age, fair skin, and residence at latitudes relatively close to the equator. Individuals with darker-pigmented skin have a lower risk of skin cancer than do fair-skinned individuals. More than 2 million individuals in the United States develop NMSC annually, and the lifetime risk that a fair-skinned individual will develop such a neoplasm is estimated at ~15%. The incidence of NMSC in the population is increasing at a rate of 2–3% per year, likely due to earlier detection and increased opportunities for outdoor activities linked to greater sun exposure.

The relationship of sun exposure to melanoma development is less direct, but strong evidence supports an association. Clear-cut risk factors include a positive family or personal history of melanoma and multiple dysplastic nevi. Melanomas can occur during adolescence; the implication is that the latent period for tumor growth is shorter than that for NMSC. For reasons that are only partially understood, melanomas are among the most rapidly increasing human malignancies **(Chap. 81)**. One potential explanation is the widespread use of indoor tanning. Furthermore, epidemiologic studies suggest that life in a sunny climate from birth or early childhood may increase the risk of melanoma development. In general, risk does not correlate with cumulative sun exposure but may be related to the duration and extent of exposure in childhood.

However, in contrast to NMSCs, melanoma frequently develops in non–sun-exposed skin, and oncogenic mutations in melanoma may also not be UVR-signature mutations. These observations suggest that UVR-independent factors may contribute to melanoma risk, which is consistent with findings in mouse models showing that pheomelanin is less efficient in protecting against melanoma than is eumelanin and may promote melanoma through UVR-independent mechanisms.

Importantly, mutations in BRAF and NRAS that lead to activation of a growth-promoting signaling cascade are frequently found in melanoma (but not in SCCs or BCCs), which has led to the development of specific inhibitors of this pathway for the treatment of BRAF-mutant melanoma. However, a high mutational load in melanoma may not be equated with a more unfavorable prognosis. Tumor-specific missense mutations in melanomas can result in neoantigens that facilitate an immune response to the tumor cell. A major advance in treating melanoma, termed immune checkpoint blockade, targets inhibitors of cytotoxic T effector function. For example, the PD-1/PD-L1 interaction inhibits tumor cell apoptosis, promotes peripheral T effector cell exhaustion, and induces conversion of T effector cells to regulatory T cells. Checkpoint inhibitor treatment (e.g., with antibodies that inhibit PD-1 or PD-L1) disrupts this interaction and has resulted in a durable and potent immune destruction of melanoma cells in a subset of patients. This has resulted in prolonged survival of patients with locally advanced or metastatic melanoma. It has recently been shown that a high mutational load in melanomas correlates with improved therapeutic outcome to immune checkpoint blockade, consistent with the hypothesis that acquired missense mutations in the tumor cells lead to neoantigens that increase the vulnerability of these melanoma cells to attack by activated T cells.

GLOBAL CONSIDERATIONS The frequency of skin cancer shows strong geographic variation, depending on the skin phototype of the majority of the population in these geographic areas, but also depending on the intensity of UVR. For example, both melanoma and NMSCs are particularly common in Australia.

Photoimmunology

Exposure to solar radiation causes both local and systemic immunosuppression and involves both the innate and adaptive immune systems. Local immunosuppression is defined as inhibition of immune responses to antigens applied at the irradiated site, whereas systemic immunosuppression is defined as inhibition of immune responses to antigens applied at remote, unirradiated sites. An example of local immunosuppression is that human skin exposure to modest doses of UV-B can deplete the epidermal antigen-presenting Langerhans cells, thereby reducing the degree of allergic sensitization to topical application of the potent contact allergen dinitrochlorobenzene at the irradiated skin site. An example of the systemic immunosuppressive effects of higher doses of UVR is the diminished immunologic response to antigens introduced either epicutaneously or intracutaneously at sites remote from the irradiated site.

The major chromophores in the upper epidermis that are known to initiate UV-mediated immunosuppression include DNA, *trans*-urocanic acid, and membrane components. The action spectrum for UV-induced immunosuppression closely mimics the absorption spectrum of DNA. UVR-induced cyclobutane pyrimidine dimers in Langerhans cells may inhibit antigen presentation. The absorption spectrum of epidermal urocanic acid closely mimics the action spectrum for UV-B–induced immunosuppression. Urocanic acid is a metabolic product of the essential amino acid histidine and accumulates in the upper epidermis through breakdown of the histidine-rich protein filaggrin due to the absence of its catabolizing enzyme in keratinocytes. Urocanic acid is synthesized as a *trans*-isomer, and UV-induced *trans-cis* isomerization of urocanic acid in the stratum corneum drives immunosuppression. *Cis*-urocanic acid may exert its immunosuppressive effects through a variety of mechanisms, including inhibition of antigen presentation by Langerhans cells.

Various additional immunomodulatory factors and cytokines have been implicated in UVR-induced systemic immunosuppression, including tumor necrosis factor-α, interleukin 4 (IL-4), interleukin 10 (IL-10), and eicosanoids. Keratinocytes can release multiple immunomodulators as a response to UVR-induced cell damage that result in an immunosuppressive environment. Induction of IL-4–producing natural killer T cells and of regulatory T cells and B cells has been linked to cell-mediated and humoral immunosuppression as a consequence of UVR damage to skin. Moreover, UVR-induced formation of damage-associated molecular patterns (DAMPs) from necrotic keratinocytes can lead to a type I interferon innate immune response via activation of Toll-like receptor signaling.

One important consequence of chronic sun exposure and associated immunosuppression is an enhanced risk of skin cancer. In part, UV-B activates regulatory T cells that suppress antitumor immune responses via IL-10 expression, whereas in the absence of high UV-B exposure, epidermal Langerhans cells present tumor-associated antigens and induce protective immunity, thereby inhibiting skin tumorigenesis. UV-induced DNA damage is a major molecular trigger of this immunosuppressive effect.

Perhaps the most graphic demonstration of the role of long-term immunosuppression in enhancing the risk of NMSC comes from studies of organ transplant recipients who require lifelong immunosuppressive/antirejection drug regimens. More than 50% of organ transplant recipients develop BCCs and SCCs, and these skin cancers are the most common types of malignancies arising in these patients. The important contributory role of UVR for the formation of these skin cancers in immunosuppressed individuals is highlighted by the observation that nonwhite transplant recipients develop these skin cancers far less often than do white transplant recipients. Rates of BCC and SCC increase with the duration and degree of immunosuppression. Transplant recipients ideally should be screened prior to organ transplantation, be monitored closely thereafter, and adhere to rigorous photoprotection measures, including the use of sunscreens and protective clothing as well as sun avoidance. Notably, immunosuppressive drugs that target the mTOR pathway, such as sirolimus and everolimus, may reduce the risk of NMSC in organ transplant recipients compared to that associated with the use of calcineurin inhibitors (cyclosporine and tacrolimus). The latter may contribute to NMSC formation not only through their immunosuppressive effects but also through suppression of p53-dependent cancer cell senescence pathways independent of host immunity.

Whereas the immunosuppressive effects of UVR contribute to skin cancer, UVR can also exacerbate autoimmune and inflammatory diseases of the skin, including systemic lupus erythematosus (SLE). It has been proposed that in SLE UVR-induced damage to DNA may promote autoantibody formation.

PHOTOSENSITIVITY DISEASES

The diagnosis of photosensitivity requires elicitation of a careful history to define the duration of signs and symptoms, the length of time between exposure to sunlight and the development of subjective symptoms, and visible changes in the skin. The age of onset can also be a helpful diagnostic clue. For example, the acute photosensitivity of erythropoietic protoporphyria (EPP) almost always begins in infancy or early childhood, whereas the chronic photosensitivity of porphyria cutanea tarda (PCT) typically begins in the fourth and fifth decades of life. A patient's history of exposure to topical and systemic drugs and chemicals may provide important diagnostic clues. Many classes of drugs can cause photosensitivity on the basis of either phototoxicity or photoallergy.

Examination of the skin may offer important clues. Anatomic areas that are naturally protected from direct sunlight, such as the hairy scalp, the upper eyelids, the retroauricular areas, and the infranasal and submental regions, may be relatively spared, whereas exposed areas show the characteristic features of the cutaneous photoaging pathologic process. These anatomic localization patterns are often helpful, but not infallible, in making the diagnosis. For example, airborne contact sensitizers that are blown onto the skin may produce dermatitis that can be difficult to distinguish from photosensitivity despite the fact that such material may trigger skin reactivity in areas shielded from direct sunlight.

Many dermatologic conditions may be caused or aggravated by sunlight (Table 64-2). The role of light in evoking these responses may be dependent on genetic abnormalities ranging from well-described defects in DNA repair that occur in xeroderma pigmentosum to the inherited abnormalities in heme synthesis that characterize the porphyrias.

TABLE 64-2 Classification of Photosensitivity Diseases

TYPE	DISEASE
Genetic	Congenital erythropoietic porphyria (CEP)
	Erythropoietic protoporphyria (EPP)
	X-linked protoprophyria (XLP)
	Porphyria cutanea tarda (PCT) —familial
	Variegate porphyria (VP)
	Hepatoerythropoietic porphyria (HEP)
	Albinism
	Xeroderma pigmentosum
	Rothmund-Thomson syndrome
	Bloom syndrome
	Cockayne syndrome
	Kindler syndrome
	Phenylketonuria
Metabolic	Porphyria cutanea tarda—sporadic
	Hartnup disease
	Kwashiorkor
	Pellagra
	Carcinoid syndrome
Phototoxic	
Internal	Drugs
External	Drugs, plants, food
Photoallergic	
Immediate	Solar urticaria
Delayed	Drug photoallergy
	Persistent light reaction/chronic actinic dermatitis
Neoplastic and degenerative	Photoaging
	Actinic keratosis
	Melanoma and nonmelanoma skin cancer
Idiopathic	Polymorphous light eruption/ Juvenile spring eruption
	Hydroa vacciniforme
	Actinic prurigo
Photoaggravated	Lupus erythematosus
	Systemic
	Subacute cutaneous
	Discoid
	Dermatomyositis
	Herpes simplex
	Lichen planus actinicus
	Acne vulgaris (aestivale)
	Pseudoporphyria

Polymorphous Light Eruption The most common type of photosensitivity disease is *polymorphous light eruption* (PMLE). It is more common in women than in men. *Juvenile spring eruption* is considered a variant of PMLE and presents in childhood or early adolescence (erythematous scaly papules or bullae on ears, face, or dorsal hands). Many affected individuals may never seek medical attention because the condition is often transient, becoming manifest in the spring with initial sun exposure but then subsiding spontaneously with continuing exposure, a phenomenon known as "hardening." The major manifestations of PMLE include (often intensely) pruritic erythematous papules that may coalesce into plaques in a patchy distribution on exposed areas of the trunk and forearms. The face is usually less affected. Pruritic skin lesions may occur within a few hours of sun exposure and may persist for weeks before PMLE resolves without scarring. Whereas the morphologic skin findings remain similar for each patient with subsequent recurrences, significant interindividual variations in skin findings are characteristic (hence the term *polymorphous*).

A skin biopsy and phototest procedures in which skin is exposed to multiple erythemal doses of UV-A and UV-B may aid in the diagnosis. The action spectrum for PMLE is usually within these portions of the solar spectrum.

Whereas the treatment of an acute flare of PMLE may require topical or systemic glucocorticoids, approaches to preventing PMLE are important and include the use of high–sun protection factor (SPF) broad-spectrum sunscreens as well as the induction of "hardening" by the cautious administration of artificial UV-B (broadband or narrowband) and/or UV-A radiation or the use of psoralen plus UV-A (PUVA) photochemotherapy for ~4 weeks before initial sun exposure. Such prophylactic phototherapy or photochemotherapy at the beginning of spring may prevent the occurrence of PMLE throughout the summer.

Actinic prurigo is a photo-induced pruritic eruption that shares similarities with PMLE and often occurs in the spring; however, it can persist throughout the summer and extend into the winter months. Actinic cheilitis is often observed. A familial predisposition is common. Photoprotective measures and treatment approaches are similar as those used for PMLE and include corticosteroids or prophylactic phototherapy.

Hydroa vacciniforme is a photodermatosis of childhood and manifests with vesicular lesions on sun-exposed skin (commonly face and dorsal hands) during the summer, usually healing with depressed scars. A role of latent Epstein-Barr virus (EBV) infection in its pathogenesis has been suggested. Photoprotective measures are important to prevent its manifestation. Treatment options include photodesensitization with narrowband (NB) UVB, antimalarials, or intermittent systemic corticosteroids.

Phototoxicity and Photoallergy These photosensitivity disorders are related to the topical or systemic administration of drugs and other chemicals that act as chromophores. Both reactions require energy absorption by a drug or chemical with consequent production of an excited-state photosensitizer that can transfer its absorbed energy to bystander molecules or to molecular oxygen, thereby generating tissue-destructive chemical species, including ROS.

Phototoxicity is a nonimmunologic reaction that can be caused by a broad range of drugs and chemicals, some of which are listed in Table 64-3. The usual clinical manifestations include erythema resembling a sunburn reaction that quickly desquamates, or "peels," within several days. In addition, edema, vesicles, and bullae may occur. A common phototoxic reaction that occurs after contact with plant-derived psoralens and exposure to UV-A radiation is called phytophotodermatitis.

Photoallergy is much less common and is distinct in that it is an immunopathologic process. The excited-state photosensitizer may create highly unstable haptenic free radicals that bind covalently to macromolecules to form a functional antigen (photoallergen) capable of evoking a delayed-type hypersensitivity response. Most photoallergic reactions are initiated by UV-A rather than UV-B exposure. Some drugs and chemicals that can produce photoallergy are listed in Table 64-4. The clinical manifestations typically differ from those of phototoxicity in that an intensely pruritic eczematous dermatitis tends to predominate and evolves into lichenified, thickened, "leathery" changes in sun-exposed areas. A small subset (perhaps 5–10%) of patients with photoallergy may develop a persistent exquisite hypersensitivity to light even when the offending drug or chemical is identified and eliminated, a condition known as *persistent light reaction*.

An uncommon type of persistent photosensitivity is known as *chronic actinic dermatitis*. The affected patients are typically elderly men with a long history of preexisting allergic contact dermatitis or photosensitivity. Lichenified eczematous patches and plaques on sun-exposed areas are typical clinical manifestations of chronic actinic dermatitis, although in severe cases, skin areas not exposed to sun can be involved as well. Common photoallergens associated with this condition are sunscreen ingredients (e.g., avobenzone and oxybenzone) and plant photoallergens (commonly *Tanacetum vulgare* [common tansy] and *Taraxacum officinale* [common dandelion]). These individuals are usually exquisitely sensitive to UV-B, UV-A, and visible wavelengths. Strict photoprotection and avoidance of known contact and photocontact allergens or photosensitizing drugs are important. Immunosuppressive topical or systemic therapies are often needed to achieve symptom improvement.

TABLE 64-3 Drugs That May Cause a Phototoxic Reaction

DRUG	TOPICAL	SYSTEMIC
Amiodarone		+
Dacarbazine		+
Fluoroquinolones		+
5-Fluorouracil	+	+
Furosemide		+
Nalidixic acid		+
Phenothiazines		+
Psoralens	+	+
Retinoids	+/–	+
Sulfonamides		+
Sulfonylureas		+
Tetracyclines		+
Thiazides		+
Vinblastine		+

TABLE 64-4 Drugs That May Cause a Photoallergic Reaction

DRUG	TOPICAL	SYSTEMIC
6-Methylcoumarin	+	
Aminobenzoic acid and esters	+	
Bithionol	+	
Chlorpromazine		+
Diclofenac		+
Fluoroquinolones		+
Halogenated salicylanilides	+	
Hypericin (St. John's wort)	+	+
Musk ambrette	+	
Piroxicam		+
Promethazine		+
Sulfonamides		+
Sulfonylureas		+

Solar urticaria represents an immediate histamine-mediated immunologic reaction to sunlight. Pruritic wheals occur within minutes of sun exposure and usually resolve within 24 hours. A neutralizing antibody to IgE, omalizumab, has shown therapeutic benefit in some patients.

Phototoxicity and photoallergy often can be diagnostically confirmed by phototest procedures. In patients with suspected phototoxicity, determining the minimal erythemal dose (MED) while the patient is exposed to a suspected agent and then repeating the MED after discontinuation of the agent may provide a clue to the causative drug or chemical. Photopatch testing can be performed to confirm the diagnosis of photoallergy. In this simple variant of ordinary patch testing, a series of known photoallergens is applied to the skin in duplicate, and one set is irradiated with a suberythemal dose of UV-A. The development of eczematous changes at sites exposed to sensitizer and light is a positive result. The characteristic abnormality in patients with persistent light reaction is a diminished threshold to erythema evoked by UV-B. Patients with chronic actinic dermatitis usually manifest a broad spectrum of UV hyperresponsiveness and require meticulous photoprotection, including avoidance of sun exposure, use of high-SPF (>30) sunscreens, and, in severe cases, systemic immunosuppression, such as with azathioprine.

The management of drug photosensitivity involves first and foremost the elimination of exposure to the chemical agents responsible for the reaction and the minimization of sun exposure. The acute symptoms of phototoxicity may be ameliorated by cool moist compresses, topical glucocorticoids, and systemically administered NSAIDs. In severely affected individuals, a tapered course of systemic glucocorticoids may be useful. Judicious use of analgesics may be necessary.

Photoallergic reactions require a similar management approach. Furthermore, patients with persistent light reaction and chronic actinic dermatitis must be meticulously protected from light exposure. In selected patients for whom chronic systemic high-dose glucocorticoids pose unacceptable risks, it may be necessary to employ an immunosuppressive drug such as azathioprine, cyclophosphamide, cyclosporine, or mycophenolate mofetil.

Porphyria The porphyrias (Chap. 428) are a group of diseases that have in common inherited or acquired derangements in the synthesis of heme. Heme is an iron-chelated tetrapyrrole or porphyrin, and only the nonmetal chelated porphyrins are potent photosensitizers that absorb light intensely in both the short (400–410 nm) and the long (580–650 nm) portions of the visible spectrum.

Heme cannot be reutilized and must be synthesized continuously. The two body compartments with the largest capacity for its production are the bone marrow and the liver. Accordingly, the porphyrias originate in one or the other of these compartments, with an end result of excessive endogenous production of potent photosensitizing porphyrins. The porphyrins circulate in the bloodstream and diffuse into the skin, where they absorb solar energy, become photoexcited, generate ROS, and evoke cutaneous photosensitivity. The mechanism of porphyrin photosensitization is known to be photodynamic, or oxygen-dependent, and is mediated by ROS such as singlet oxygen and superoxide anions.

The group of cutaneous porphyrias can be classified as causing either (1) chronic blistering photosensitivity or (2) acute nonblistering photosensitivity. Chronic cutaneous porphyrias include PCT, congenital erythropoietic porphyria (CEP), hepatoerythropoietic porphyria (HEP), hereditary coproporphyria (HCP), and variegate porphyria (VP). CEP, HEP, and PCT manifest only with cutaneous symptoms, whereas HCP and VP have acute neurovisceral symptoms in addition to the skin photosensitivity. Acute, cutaneous, nonblistering porphyrias include EPP and X-linked protoporphyria (XLP). EPP is caused by a deficiency of ferrochelatase (FECH), whereas XLP is caused by a gain-of-function mutation of the erythroid-specific form of delta-aminolevulinic acid (ALAS2). In both, accumulation of protoporphyrins in the bone marrow and then in erythrocytes and plasma leads to cutaneous photosensitivity. Total and metal-free erythrocyte protoporphyrin (PROTO) are increased in these protoporphyrias. Representative examples of chronic and acute cutaneous porphyrias are discussed below.

Porphyria cutanea tarda (PCT) is the most common type of human porphyria and is associated with decreased activity of the heme pathway enzyme uroporphyrinogen decarboxylase (UROD) to <20% of normal. Increased iron and various acquired factors (e.g., alcohol consumption, estrogens, smoking, hepatitis C or HIV infection) can reduce UROD activity. There are two basic types of PCT: (1) the sporadic or acquired type, generally seen in individuals ingesting ethanol or receiving estrogens; and (2) the inherited type, in which there is autosomal dominant transmission of deficient enzyme activity (resulting in heterozygosity for UROD with a reduction to 50% of UROD enzymatic activity and, thus, predisposing the individual to PCT). Both forms are associated with increased hepatic iron stores.

In both types of PCT, the predominant feature is chronic photosensitivity characterized by increased fragility of sun-exposed skin, particularly areas subject to repeated trauma such as the dorsa of the hands, the forearms, the face, and the ears. The predominant skin lesions are vesicles and bullae that rupture, producing moist erosions (often with a hemorrhagic base) that heal slowly, with crusting and purplish discoloration of the affected skin. Hypertrichosis, mottled pigmentary change, and scleroderma-like induration are associated features. The diagnosis can be confirmed biochemically by measurement of urinary porphyrin excretion, plasma porphyrin assay, and assay of erythrocyte and/or hepatic UROD. Multiple mutations of the *UROD* gene have been identified in human populations. Some patients with PCT have associated mutations in the *HFE* gene, which is linked to hemochromatosis and leads to increased oral iron absorption by reducing hepcidin expression; these mutations could contribute to the iron overload precipitating PCT, although iron status as measured by serum ferritin, iron levels, and transferrin saturation is no different from that in PCT patients without *HFE* mutations.

Treatment of PCT consists of repeated phlebotomies to diminish the excessive hepatic iron stores and/or intermittent (twice weekly) low doses of orally administered hydroxychloroquine. This treatment is highly effective for PCT but not suited for treatment of other porphyrias. Long-term remission of the disease can often be achieved if the patient eliminates exposure to porphyrinogenic agents, such as ethanol or estrogens, and avoids sun exposure.

Pseudoporphyria is a term used for a photodermatosis that clinically resembles PCT but that does not have abnormalities in porphyrin metabolism. It has been associated with chronic kidney disease and hemodialysis, tanning bed use, or certain medications (e.g., naproxen).

Erythropoietic protoporphyria (EPP) is an acute nonblistering cutaneous porphyria, originates in the bone marrow, and is due to genetic mutations that in most cases decrease the activity of the mitochondrial enzyme ferrochelatase. The major clinical features include acute photosensitivity characterized by painful burning and stinging of exposed skin that often develops during or just after sun exposure. There may be associated skin swelling and, after repeated episodes, a waxlike scarring. Rigorous sunlight protection is essential in the management of EPP. Notably, the U.S. Food and Drug Administration (FDA) has approved a synthetic peptide analogue of α-MSH, afamelanotide, in patients with EPP. This drug increases skin pigmentation through melanogenesis, and patients receiving it tolerate sun exposure without pain for longer periods of time and have an improved quality of life as compared to untreated patients. Interestingly, initial studies suggest that afamelanotide may also be beneficial when combined with NB-UV-B in the treatment of patients with vitiligo (in patients with skin phototypes IV–VI). Some studies reported that patients with EPP had a moderate increase in tolerance to sunlight after taking oral β-carotene, which may provide this effect by quenching oxygen free radicals.

An algorithm for managing patients with photosensitivity is presented in Fig. 64-1.

PHOTOPROTECTION

Since photosensitivity of the skin results from exposure to sunlight, it follows that absolute avoidance of sunlight will eliminate these disorders. However, contemporary lifestyles make this approach impractical for most individuals. Thus, better approaches to photoprotection have been sought. Natural photoprotection is provided by structural proteins in the epidermis, particularly keratins and melanin. The amount of melanin and its distribution in cells are genetically regulated, and individuals of darker complexion (skin types IV–VI) are at decreased risk for the development of acute sunburn and cutaneous malignancy. Other forms of photoprotection include clothing and sunscreens. Clothing constructed of tightly woven sun-protective fabrics, irrespective of color, affords substantial protection. Wide-brimmed hats, long sleeves, and trousers all reduce direct exposure.

Sunscreens are now considered over-the-counter drugs, and a monograph from the FDA has recognized category I ingredients as safe and effective. Those ingredients are listed in Table 64-5. Sunscreens are rated for their photoprotective effect by their SPF. The SPF is simply a ratio of the time required to produce sunburn erythema with and without sunscreen application. The SPF of most sunscreens reflects protection from UV-B but not from UV-A. The FDA monograph stipulates that sunscreens must be rated on a scale ranging from minimal (SPF ≥2 and <12) to moderate (SPF ≥12 and <30) to high (SPF ≥30, labeled as 30+).

Broad-spectrum sunscreens contain both UV-B–absorbing and UV-A–absorbing chemicals (organic filters). These chemicals absorb UVR and transfer the absorbed energy to surrounding cells. Among these sunscreen ingredients, cinnamates, PABA derivatives, and salicylates absorb UV-B. Benzophenones or ecamsule (terephthalylidene dicamphor sulfonic acid) offer protection against UV-B and UV-A2, whereas avobenzone protects mainly against UV-A1. In contrast, physical UV blockers (zinc oxide and titanium dioxide) absorb or reflect UVR and offer broad-spectrum protection against UV-B and UV-A. In addition to light absorption, a critical determinant of the sustained photoprotective effect of sunscreens is their water resistance. Sunscreen products with an SPF of 30 or higher, broad-spectrum coverage, and water or sweat resistance are recommended for adequate sun protection.

Some degree of photoprotection can be achieved by limiting the time of sun exposure during the day. Since a large part of an individual's total lifetime sun exposure may occur by age 18, it is important

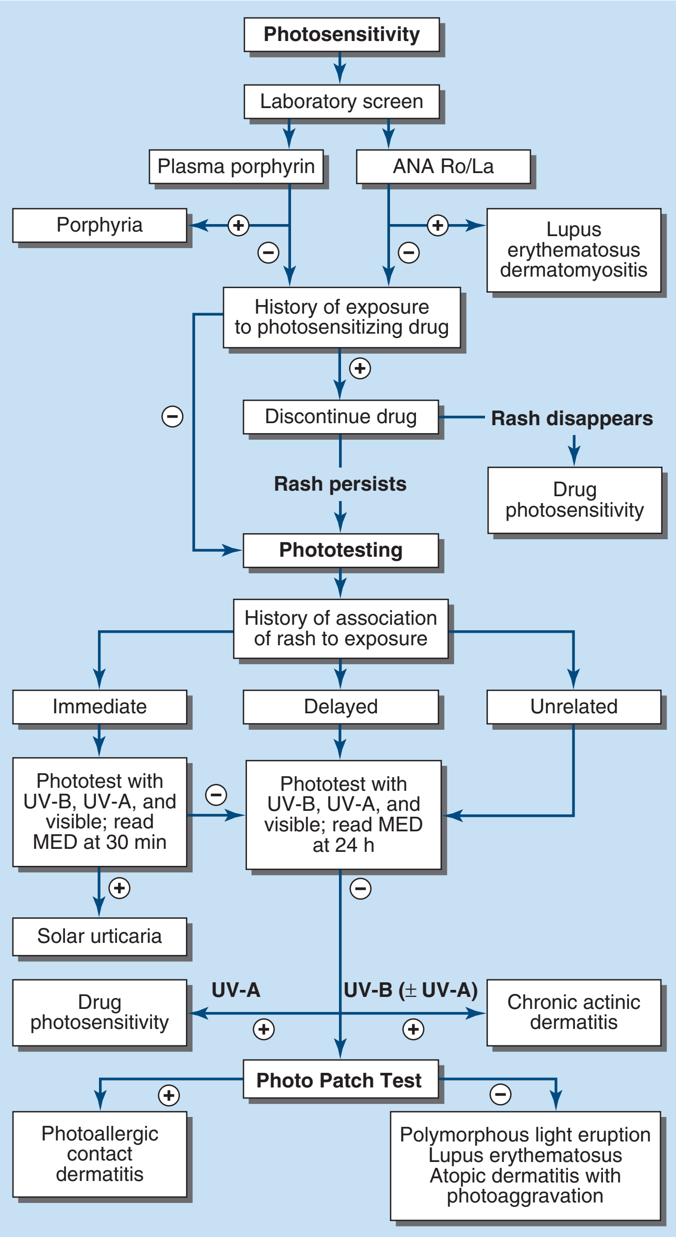
Photosensitivity
Laboratory screen
Plasma porphyrin
ANA Ro/La
Porphyria
+
−
+
−
Lupus erythematosus dermatomyositis
History of exposure to photosensitizing drug
+
−
Discontinue drug
Rash disappears
Rash persists
Drug photosensitivity
Phototesting
History of association of rash to exposure
Immediate
Delayed
Unrelated
Phototest with UV-B, UV-A, and visible; read MED at 30 min
−
Phototest with UV-B, UV-A, and visible; read MED at 24 h
+
−
Solar urticaria
Drug photosensitivity
UV-A
UV-B (± UV-A)
+
+
Chronic actinic dermatitis
Photo Patch Test
+
−
Photoallergic contact dermatitis
Polymorphous light eruption
Lupus erythematosus
Atopic dermatitis with photoaggravation

FIGURE 64-1 Algorithm for the diagnosis of a patient with photosensitivity. ANA, antinuclear antibody; MED, minimal erythemal dose; UV-A and UV-B, ultraviolet spectrum segments including wavelengths of 320–400 nm and 290–320 nm, respectively.

TABLE 64-5 FDA Category I Monographed Sunscreen Active Ingredients

INGREDIENTS	MAXIMUM CONCENTRATION, %
p-Aminobenzoic acid (PABA)	15
Avobenzone	3
Cinoxate	3
Dioxybenzone (benzophenone-8)	3
Ecamsule	15
Homosalate	15
Meradimate	5
Octocrylene	10
Octinoxate	7.5
Octisalate	5
Oxybenzone (benzophenone-3)	6
Padimate O (octyl dimethyl PABA)	8
Ensulizole	4
Sulisobenzone (benzophenone-4)	10
Titanium dioxide	25
Trolamine salicylate	12
Zinc oxide	25

Abbreviation: FDA, U.S. Food and Drug Administration.

to educate parents and young children about the hazards of sunlight. Eliminating exposure at midday will substantially reduce lifetime UVR exposure.

PHOTOTHERAPY AND PHOTOCHEMOTHERAPY

UVR can be used therapeutically. The administration of UV-B alone or in combination with topically applied agents can induce remissions of many dermatologic diseases, including psoriasis, atopic dermatitis, and vitiligo. In particular, NB-UV-B treatments (with fluorescent bulbs emitting radiation at ~311 nm) have enhanced efficacy over that obtained with broadband UV-B in the treatment of psoriasis.

Photochemotherapy in which topically applied or systemically administered psoralens are combined with UV-A (PUVA) is effective in treating psoriasis and the early stages of cutaneous T-cell lymphoma and vitiligo. Psoralens are tricyclic furocoumarins that, when intercalated into DNA and exposed to UV-A, form adducts with pyrimidine bases and eventually form DNA cross-links. These structural changes are thought to decrease DNA synthesis and to be related to the amelioration of psoriasis. Why PUVA photochemotherapy is effective in cutaneous T-cell lymphoma is only partially understood, but it has been shown to induce apoptosis of atypical T lymphocyte populations in the skin. Consequently, direct treatment of circulating atypical lymphocytes by extracorporeal photochemotherapy (photopheresis) has been used in Sézary syndrome as well as in other severe systemic diseases with circulating atypical lymphocytes, such as graft-versus-host disease.

In addition to its effects on DNA, PUVA photochemotherapy stimulates epidermal thickening and melanin synthesis; the latter property, together with its anti-inflammatory effects, provides the rationale for use of PUVA in the depigmenting disease vitiligo. However, NB-UVB is more commonly used as a phototherapy option for the treatment of vitiligo. NB-UVB phototherapy treatments over a period of 6–12 months are often required to achieve repigmentation. Topical calcineurin inhibitors and the topical JAK inhibitor ruxolitinib have shown therapeutic benefit for localized vitiligo.

Not surprisingly, the major side effects of long-term UV-B phototherapy and PUVA photochemotherapy mimic those seen in individuals with chronic sun exposure. Despite these risks, the therapeutic index of these modalities continues to be excellent. It is important to choose the most appropriate phototherapeutic approach for a specific dermatologic disease. For example, NB-UV-B has been reported in several studies to be as effective as PUVA photochemotherapy in the treatment of psoriasis but to pose a lower risk of skin cancer development than PUVA.

FURTHER READING

Bernard JJ et al: Photoimmunology: How ultraviolet radiation affects the immune system. Nat Rev Immunol 11:688, 2019.

Harms PW et al: The biology and treatment of Merkel cell carcinoma: Current understanding and research priorities. Nat Rev Clin Oncol 15:763, 2018.

Jansen R et al: Photoprotection: Part II. Sunscreen: Development, efficacy, and controversies. J Am Acad Dermatol 69:867, 2013.

Lo JA et al: The melanoma revolution: From UV carcinogenesis to a new era in therapeutics. Science 346:945, 2014.

Martincorena I et al: Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science 348:880, 2015.

Sanchez-Danes A et al: Defining the clonal dynamics leading to mouse skin tumour initiation. Nature 536:298, 2016.

Section 9 Hematologic Alterations

65

Interpreting Peripheral Blood Smears

Dan L. Longo

Some of the relevant findings in peripheral blood, enlarged lymph nodes, and bone marrow are illustrated in this chapter. Systematic histologic examination of the bone marrow and lymph nodes is beyond the scope of a general medicine textbook. However, every internist should know how to examine a peripheral blood smear.

The examination of a peripheral blood smear is one of the most informative exercises a physician can perform. Although advances in automated technology have made the examination of a peripheral blood smear by a physician seem less important, the technology is not a completely satisfactory replacement for a blood smear interpretation by a trained medical professional who also knows the patient's clinical history, family history, social history, and physical findings. It is useful to ask the laboratory to generate a Wright's-stained peripheral blood smear and examine it. A normal peripheral blood smear is shown in **Figure 65-1**.

The best place to examine blood cell morphology is the feathered edge of the blood smear where red cells lie in a single layer, side by side, just barely touching one another but not overlapping. The author's approach is to look at the smallest cellular elements, the platelets, first and work his way up in size to red cells and then white cells.

Using an oil immersion lens that magnifies the cells 100-fold, one counts the platelets in five to six fields, averages the number per field, and multiplies by 20,000 to get a rough estimate of the platelet count. The platelets are usually 1–2 μm in diameter and have a blue granulated appearance. There is usually 1 platelet for every 20 or so red cells. Of course, the automated counter is much more accurate, but gross disparities between the automated and manual counts should be assessed. Large platelets may be a sign of rapid platelet turnover, as young platelets are often larger than old ones; alternatively, certain rare inherited syndromes can produce large platelets. If the platelet count is low, the absence of large (young) platelets may be an indicator of marrow production problems. Platelet clumping visible on the smear can be associated with falsely low automated platelet counts. Clumping may be caused by the anticoagulant into which the blood is drawn. Similarly, neutrophil fragmentation can be a source of falsely elevated automated platelet counts. The absence of platelet granules may be an artifact of the handling of the blood or may indicate marrow disease or a rare congenital anomaly, gray platelet syndrome. Elevated platelet counts usually signify a myeloproliferative disorder or a reaction to systemic inflammation.

Next one examines the red blood cells. One can gauge their size by comparing the red cell to the nucleus of a small lymphocyte. Both are normally about 8-μm wide. Red cells that are smaller than the small lymphocyte nucleus may be microcytic; those larger than the small lymphocyte nucleus may be macrocytic. Macrocytic cells also tend to be more oval than spherical in shape and are sometimes called macroovalocytes. The automated mean corpuscular volume (MCV) can assist in making a classification. However, some patients may have both iron and vitamin B_{12} deficiency, which will produce an MCV in the normal range but wide variation in red cell size. When the red cells vary greatly in size, *anisocytosis* is said to be present. When the red cells vary greatly in shape, *poikilocytosis* is said to be present. The electronic cell counter provides an independent assessment of variability in red cell size. It measures the range of red cell volumes and reports the results as "red cell distribution width" (RDW). This value is calculated from the MCV; thus, cell width is not being measured but cell volume is. The term is derived from the curve displaying the frequency of cells at each volume, also called the distribution. The width of red cell volume distribution curve is what determines the RDW. The RDW is calculated as follows: RDW = (standard deviation of MCV ÷ mean MCV) × 100. In the presence of morphologic anisocytosis, RDW (normally 11–14%) increases to 15–18%. The RDW is useful in at least two clinical settings. In patients with microcytic anemia, the differential diagnosis is generally between iron deficiency and thalassemia. In thalassemia, the small red cells are generally of uniform size with a normal small RDW. In iron deficiency, the size variability and the RDW are large. In addition, a large RDW can suggest a dimorphic anemia when a chronic atrophic gastritis can produce both vitamin B_{12} malabsorption to produce macrocytic anemia and blood loss to produce iron deficiency. In such settings, RDW is also large. An elevated RDW also has been reported as a risk factor for all-cause mortality in population-based studies, a finding that is unexplained currently.

After red cell size is assessed, one examines the hemoglobin content of the cells. They are either normal in color (*normochromic*) or pale in color (*hypochromic*). They are never "hyperchromic." If more than the normal amount of hemoglobin is made, the cells get larger—they do not become darker. In addition to hemoglobin content, the red cells are examined for inclusions. Red cell inclusions are the following:

1. *Basophilic stippling*—diffuse fine or coarse blue dots in the red cell usually representing RNA residue—especially common in lead poisoning
2. *Howell-Jolly bodies*—dense blue circular inclusions that represent nuclear remnants—their presence implies defective splenic function
3. *Nuclei*—red cells may be released or pushed out of the marrow prematurely before nuclear extrusion—often implies a myelophthisic process or a vigorous narrow response to anemia, usually hemolytic anemia
4. *Parasites*—red cell parasites include malaria and babesia **(Chaps. A2 and A6)**
5. *Polychromatophilia*—the red cell cytoplasm has a bluish hue, reflecting the persistence of ribosomes still actively making hemoglobin in a young red cell

Vital stains are necessary to see precipitated hemoglobin called *Heinz bodies*.

Red cells can take on a variety of different shapes. All abnormally shaped red cells are *poikilocytes*. Small red cells without the central pallor are *spherocytes*; they can be seen in hereditary spherocytosis, hemolytic anemias of other causes, and clostridial sepsis. *Dacrocytes* are teardrop-shaped cells that can be seen in hemolytic anemias, severe iron deficiency, thalassemias, myelofibrosis, and myelodysplastic syndromes. *Schistocytes* are helmet-shaped cells that reflect microangiopathic hemolytic anemia or fragmentation on an artificial heart valve. *Echinocytes* are spiculated red cells with the spikes evenly spaced; they can represent an artifact of abnormal drying of the blood smear or reflect changes in stored blood. They also can be seen in renal failure and malnutrition and are often reversible. *Acanthocytes* are spiculated red cells with the spikes irregularly distributed. This process tends to be irreversible and reflects underlying renal disease, abetalipoproteinemia, or splenectomy. *Elliptocytes* are elliptical-shaped red cells that can reflect an inherited defect in the red cell membrane, but they also are seen in iron deficiency, myelodysplastic syndromes, megaloblastic anemia, and thalassemias. *Stomatocytes* are red cells in which the area of central pallor takes on the morphology of a slit instead of the usual round shape. Stomatocytes can indicate an inherited red cell membrane defect and also can be seen in alcoholism. *Target cells* have an area of central pallor that contains a dense center, or bull's eye. These cells are seen classically in thalassemia, but they are also present in iron deficiency, cholestatic liver disease, and some hemoglobinopathies. They also can be generated artifactually by improper slide making.

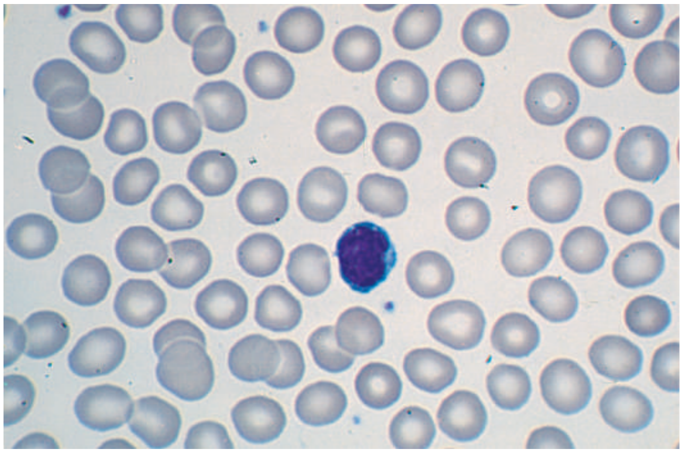

FIGURE 65-1 Normal peripheral blood smear. Small lymphocyte in center of field. Note that the diameter of the red blood cell is similar to the diameter of the small lymphocyte nucleus. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault, Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

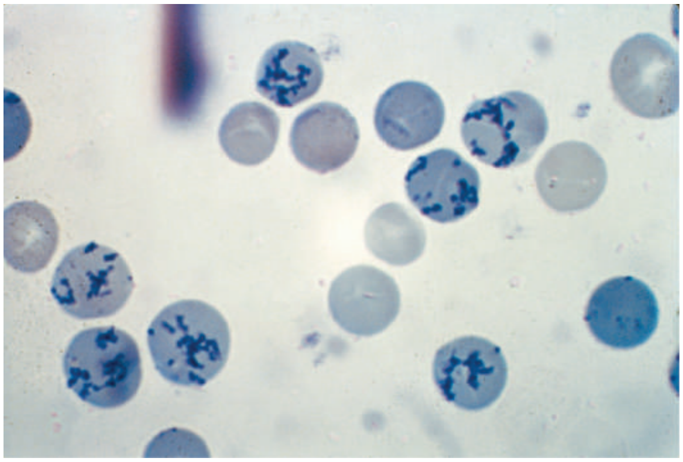

FIGURE 65-2 Reticulocyte count preparation. This new methylene blue–stained blood smear shows large numbers of heavily stained reticulocytes (the cells containing the dark blue–staining RNA precipitates). *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

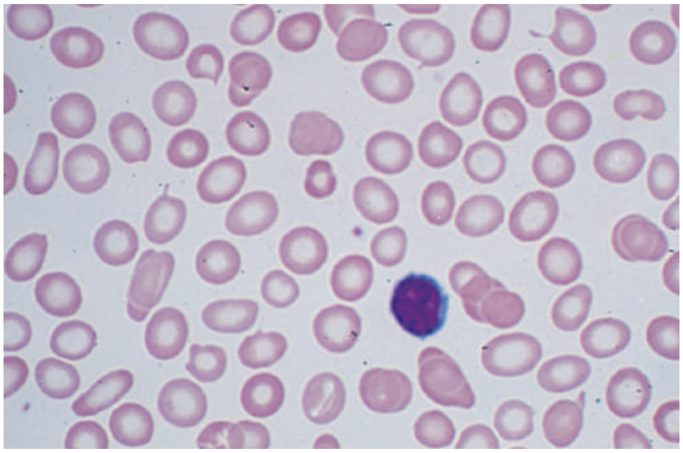

FIGURE 65-3 Hypochromic microcytic anemia of iron deficiency. Small lymphocyte in field helps assess the red blood cell size. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

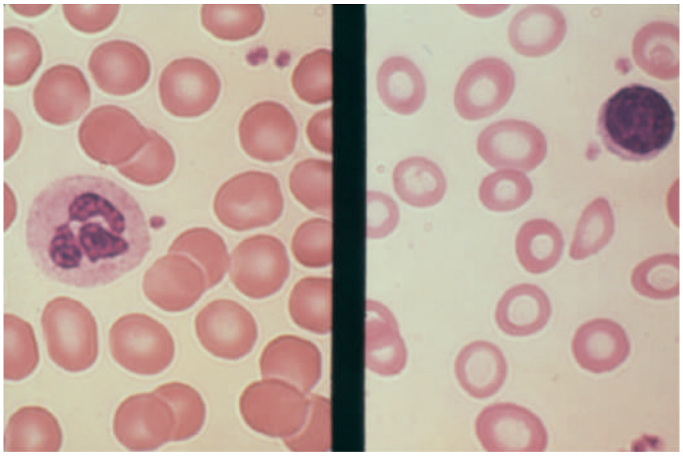

FIGURE 65-4 Iron deficiency anemia next to normal red blood cells. Microcytes (*right panel*) are smaller than normal red blood cells (cell diameter <7 μm) and may or may not be poorly hemoglobinized (hypochromic). *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault, Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

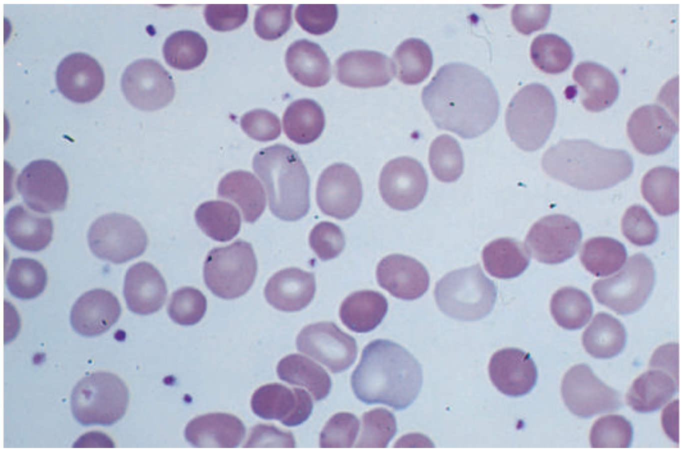

FIGURE 65-5 Polychromatophilia. Note large red cells with light purple coloring. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

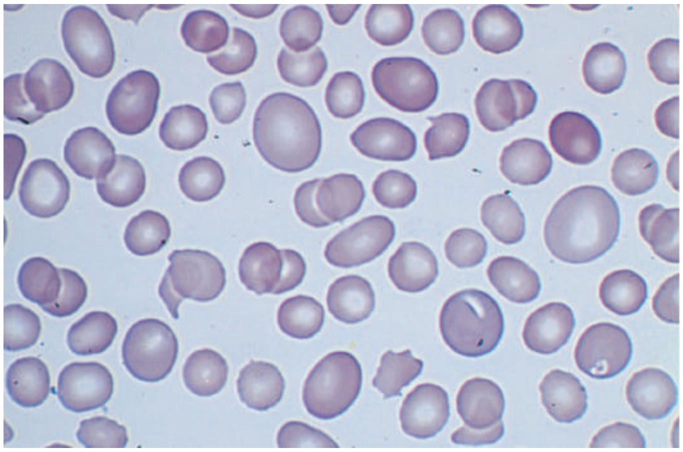

FIGURE 65-6 Macrocytosis. These cells are both larger than normal (mean corpuscular volume >100) and somewhat oval in shape. Some morphologists call these cells macroovalocytes. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

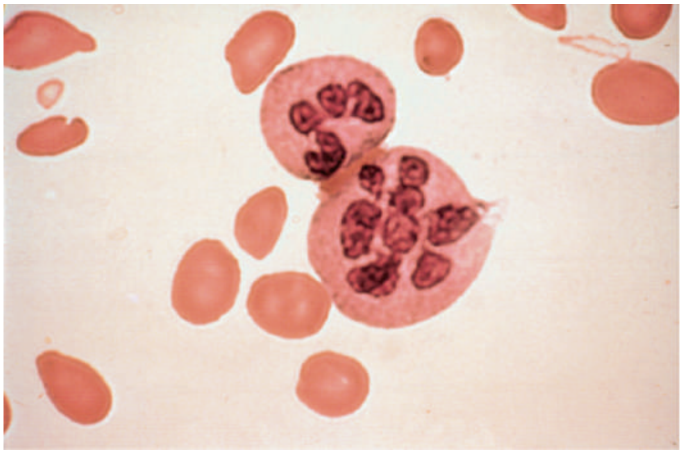

FIGURE 65-7 Hypersegmented neutrophils. Hypersegmented neutrophils (multilobed polymorphonuclear leukocytes) are larger than normal neutrophils with five or more segmented nuclear lobes. They are commonly seen with folic acid or vitamin B_{12} deficiency. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

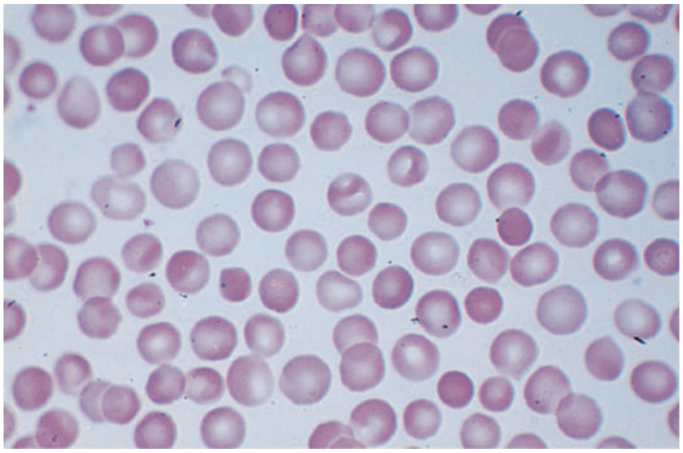

FIGURE 65-8 Spherocytosis. Note small hyperchromatic cells without the usual clear area in the center. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

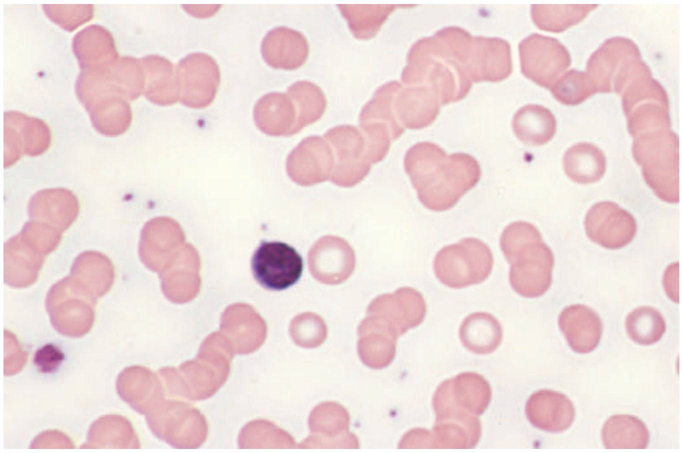

FIGURE 65-9 Rouleaux formation. Small lymphocyte in center of field. These red cells align themselves in stacks and are related to increased serum protein levels. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

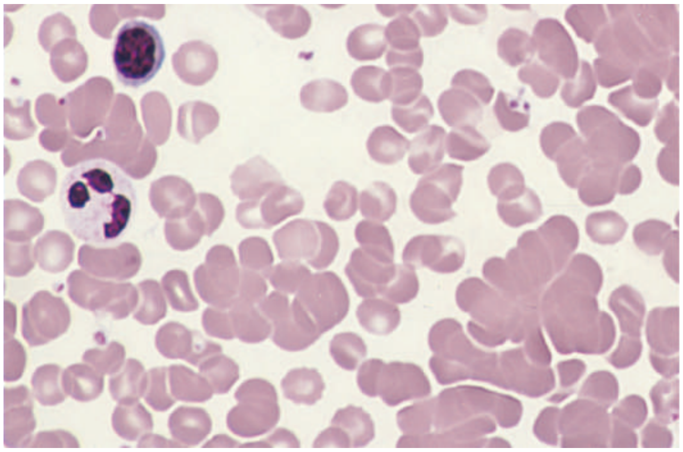

FIGURE 65-10 Red cell agglutination. Small lymphocyte and segmented neutrophil in upper left center. Note irregular collections of aggregated red cells. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

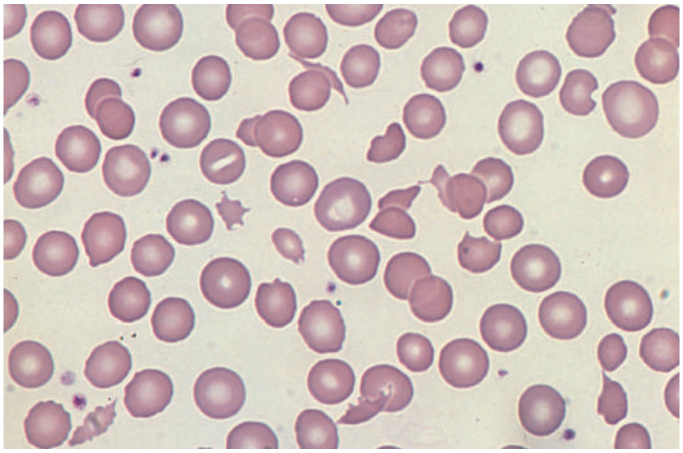

FIGURE 65-11 Fragmented red cells. Heart valve hemolysis. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

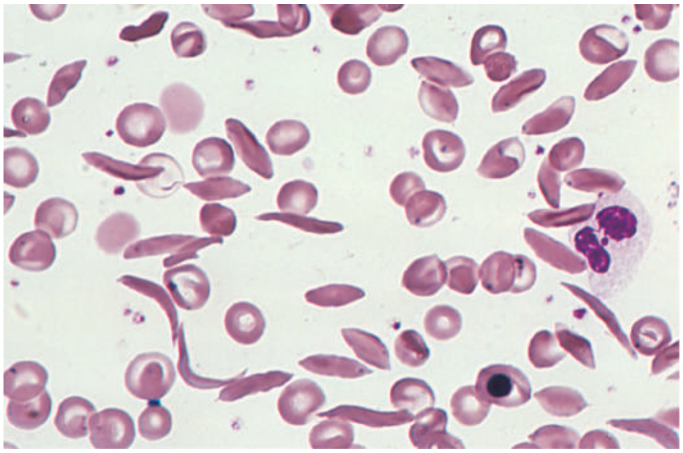

FIGURE 65-12 Sickle cells. Homozygous sickle cell disease. A nucleated red cell and neutrophil are also in the field. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

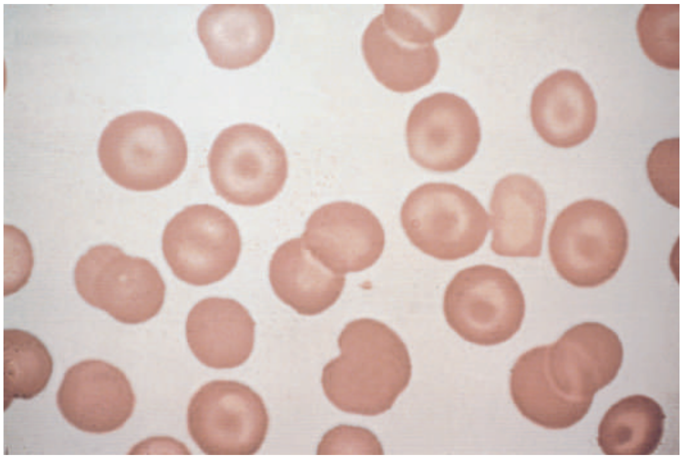

FIGURE 65-13 Target cells. Target cells are recognized by the bull's-eye appearance of the cell. Small numbers of target cells are seen with liver disease and thalassemia. Larger numbers are typical of hemoglobin C disease. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

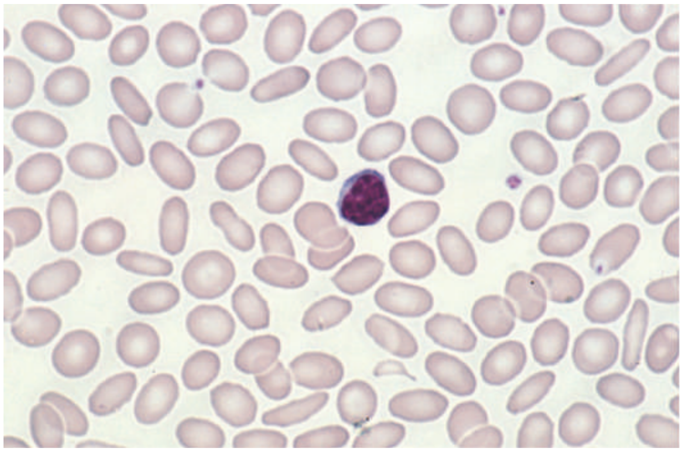

FIGURE 65-14 Elliptocytosis. Small lymphocyte in center of field. Elliptical shape of red cells related to weakened membrane structure, usually due to mutations in spectrin. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

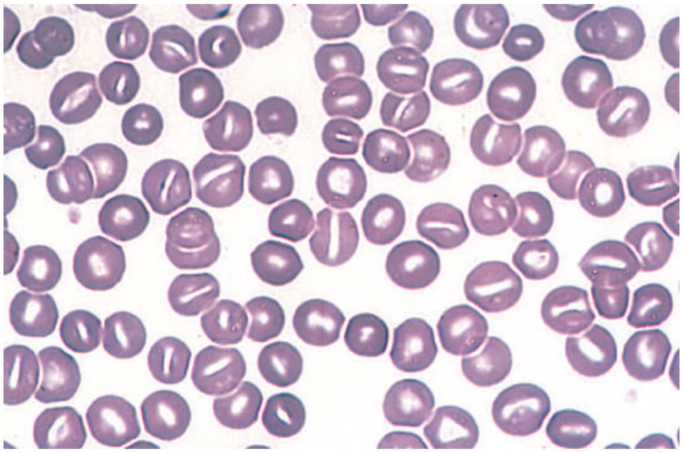

FIGURE 65-15 Stomatocytosis. Red cells characterized by a wide transverse slit or stoma. This often is seen as an artifact in a dehydrated blood smear. These cells can be seen in hemolytic anemias and in conditions in which the red cell is overhydrated or dehydrated. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

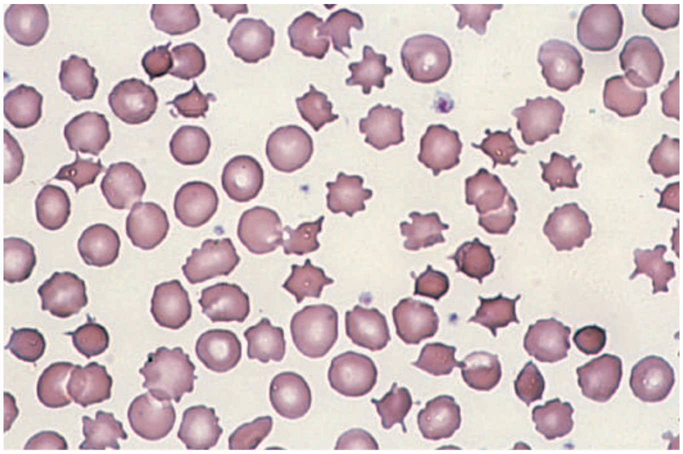

FIGURE 65-16 Acanthocytosis. Spiculated red cells are of two types: *acanthocytes* are contracted dense cells with irregular membrane projections that vary in length and width; *echinocytes* have small, uniform, and evenly spaced membrane projections. Acanthocytes are present in severe liver disease, in patients with abetalipoproteinemia, and in rare patients with McLeod blood group. Echinocytes are found in patients with severe uremia, in glycolytic red cell enzyme defects, and in microangiopathic hemolytic anemia. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

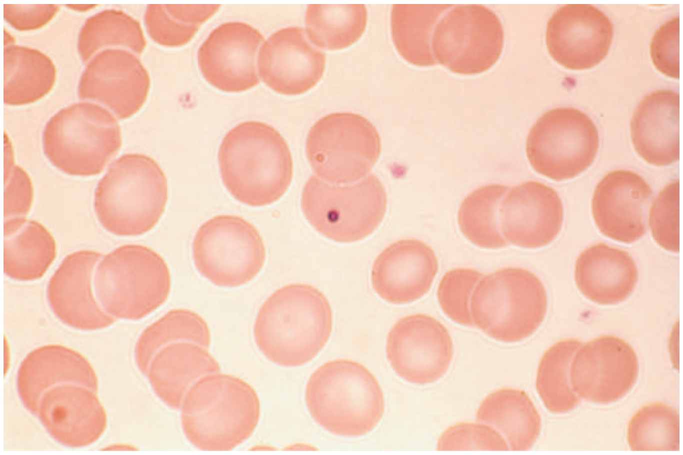

FIGURE 65-17 Howell-Jolly bodies. Howell-Jolly bodies are tiny nuclear remnants that normally are removed by the spleen. They appear in the blood after splenectomy (defect in removal) and with maturation/dysplastic disorders (excess production). *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

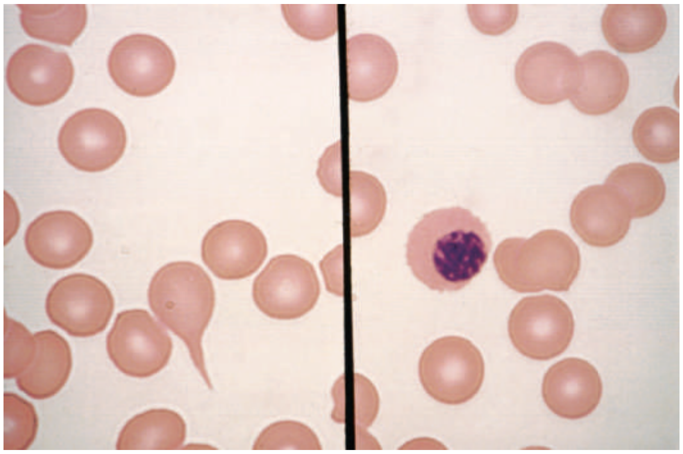

FIGURE 65-18 Teardrop cells and nucleated red blood cells characteristic of myelofibrosis. A teardrop-shaped red blood cell (*left panel*) and a nucleated red blood cell (*right panel*) as typically seen with myelofibrosis and extramedullary hematopoiesis. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

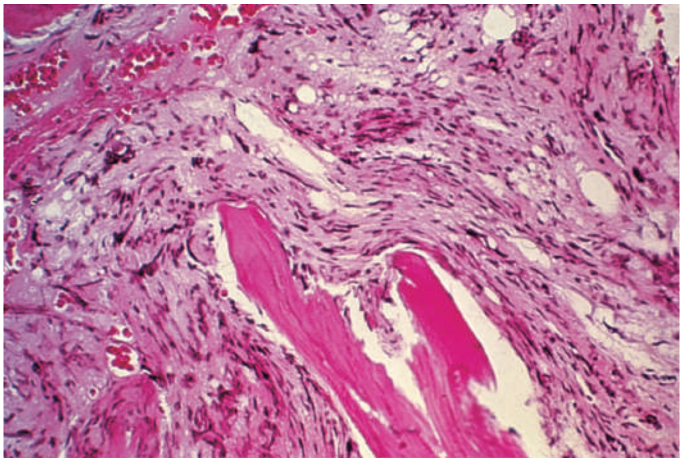

FIGURE 65-19 Myelofibrosis of the bone marrow. Total replacement of marrow precursors and fat cells by a dense infiltrate of reticulin fibers and collagen (hematoxylin and eosin stain). *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

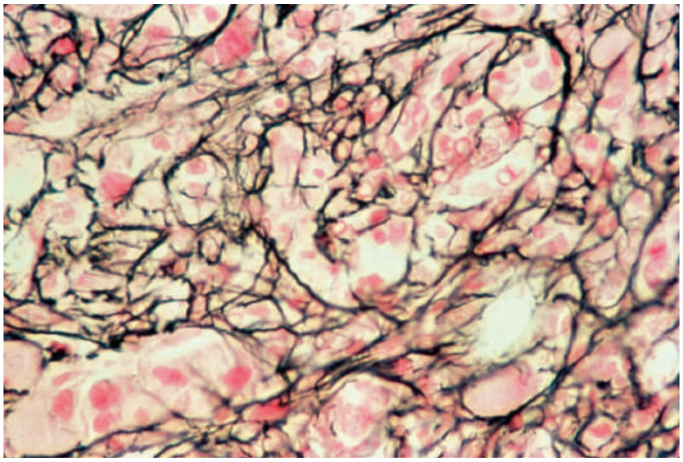

FIGURE 65-20 Reticulin stain of marrow myelofibrosis. Silver stain of a myelofibrotic marrow showing an increase in reticulin fibers (black-staining threads). *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

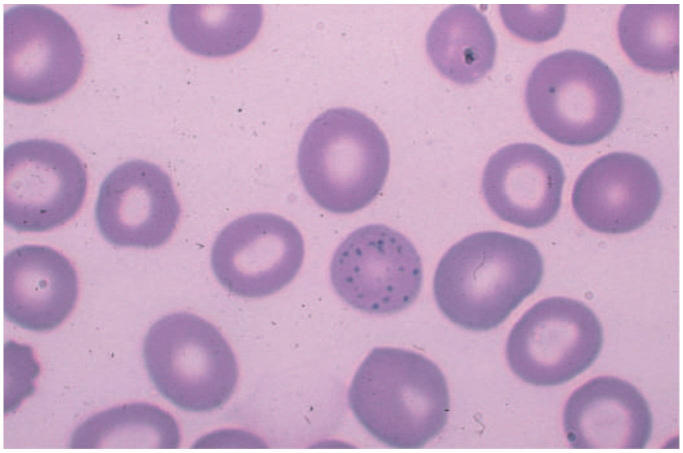

FIGURE 65-21 Stippled red cell in lead poisoning. Mild hypochromia. Coarsely stippled red cell. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

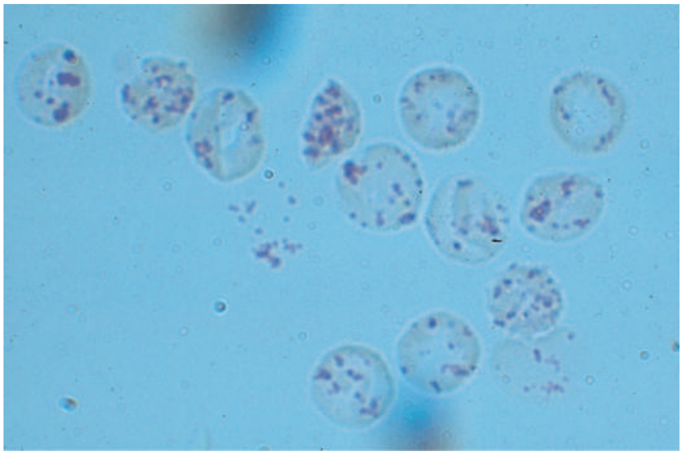

FIGURE 65-22 Heinz bodies. Blood mixed with hypotonic solution of crystal violet. The stained material is precipitates of denatured hemoglobin within cells. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

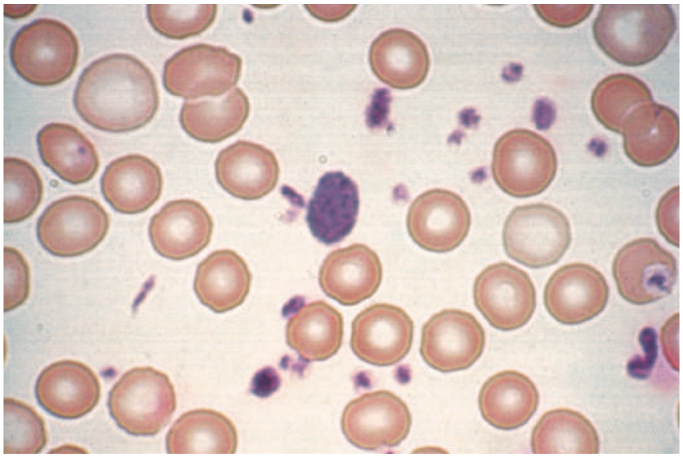

FIGURE 65-23 Giant platelets. Giant platelets, together with a marked increase in the platelet count, are seen in myeloproliferative disorders, especially primary thrombocythemia. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

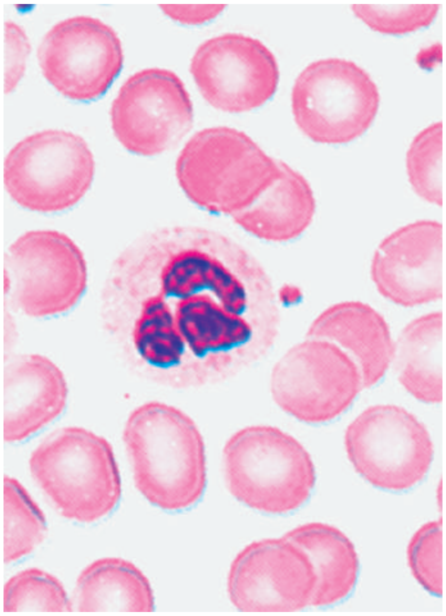

FIGURE 65-24 Normal granulocytes. The normal granulocyte has a segmented nucleus with heavy, clumped chromatin; fine neutrophilic granules are dispersed throughout the cytoplasm. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

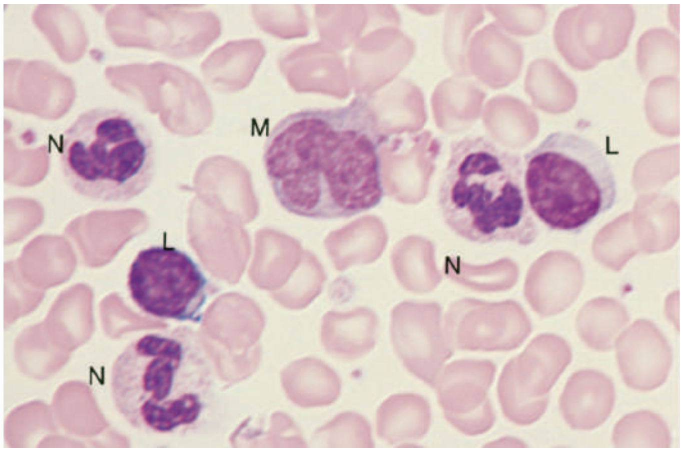

FIGURE 65-25 Normal monocytes. The film was prepared from the buffy coat of the blood from a normal donor. L, lymphocyte; M, monocyte; N, neutrophil. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

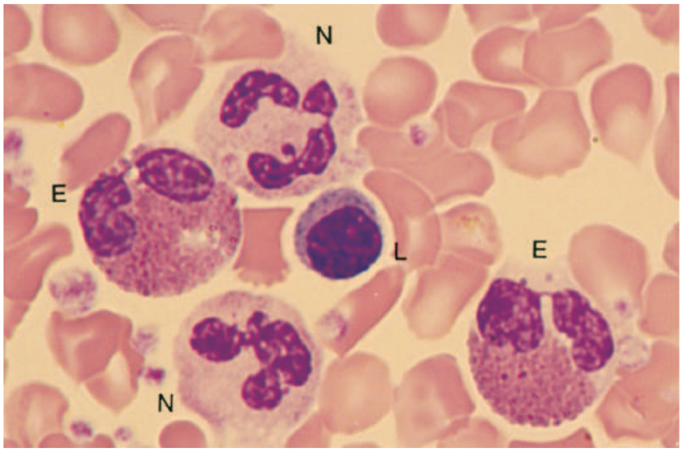

FIGURE 65-26 Normal eosinophils. The film was prepared from the buffy coat of the blood from a normal donor. E, eosinophil; L, lymphocyte; N, neutrophil. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

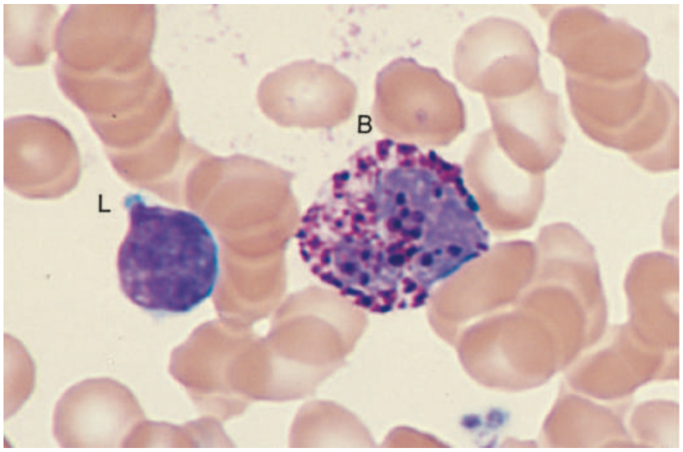

FIGURE 65-27 Normal basophil. The film was prepared from the buffy coat of the blood from a normal donor. B, basophil; L, lymphocyte. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

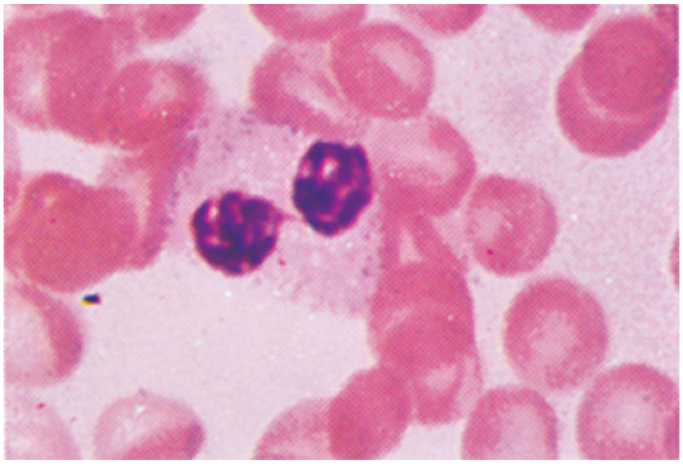

FIGURE 65-28 Pelger-Hüet anomaly. In this benign disorder, the majority of granulocytes are bilobed. The nucleus frequently has a spectacle-like, or "pince-nez," configuration. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

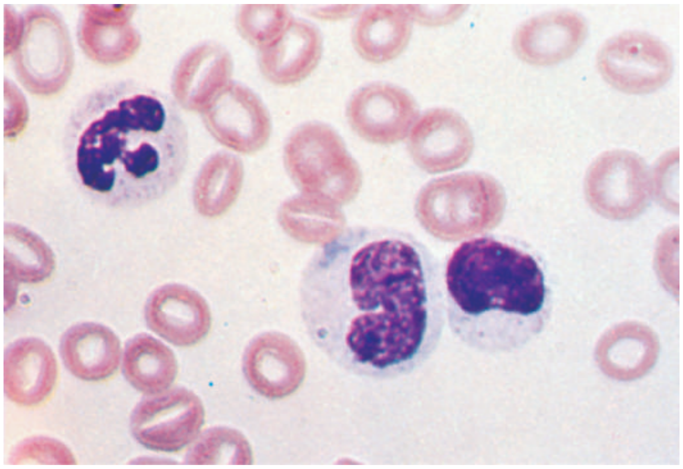

FIGURE 65-29 Döhle body. Neutrophil band with Döhle body. The neutrophil with a sausage-shaped nucleus in the center of the field is a band form. Döhle bodies are discrete, blue-staining nongranular areas found in the periphery of the cytoplasm of the neutrophil in infections and other toxic states. They represent aggregates of rough endoplasmic reticulum. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

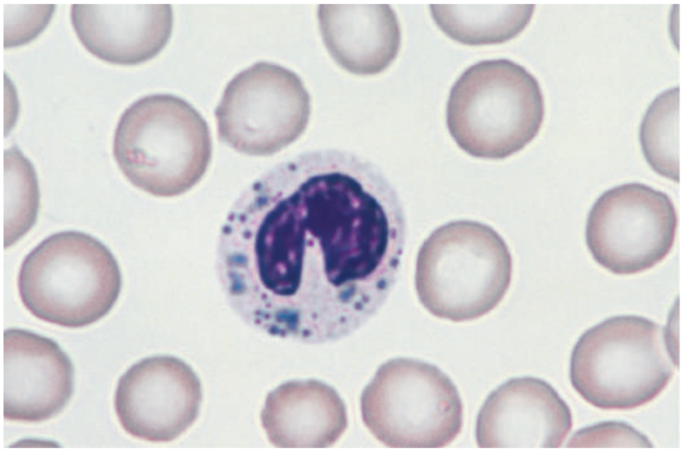

FIGURE 65-30 Chédiak-Higashi disease. Note giant granules in neutrophil. *(Source: From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

One last feature of the red cells to assess before moving to the white blood cells is the distribution of the red cells on the smear. In most individuals, the cells lie side by side in a single layer. Some patients have red cell clumping (called *agglutination*) in which the red cells pile upon one another; it is seen in certain paraproteinemias and autoimmune hemolytic anemias. Another abnormal distribution involves red cells lying in single cell rows on top of one another like stacks of coins. This is called *rouleaux formation* and reflects abnormal serum protein levels.

Finally, one examines the white blood cells. Three types of granulocytes are usually present: neutrophils, eosinophils, and basophils, in decreasing frequency. Neutrophils are generally the most abundant white cell. They are round, are 10–14 μm wide, and contain a lobulated nucleus with two to five lobes connected by a thin chromatin thread. Bands are immature neutrophils that have not completed nuclear condensation and have a U-shaped nucleus. Bands reflect a left shift in neutrophil maturation in an effort to make more cells more rapidly. Neutrophils can provide clues to a variety of conditions. Vacuolated neutrophils may be a sign of bacterial sepsis. The presence of 1- to 2-μm blue cytoplasmic inclusions, called *Döhle bodies*, can reflect infections, burns, or other inflammatory states. If the neutrophil granules are larger than normal and stain a darker blue, "toxic granulations" are said to be present, and they also suggest a systemic inflammation. The presence of neutrophils with more than five nuclear lobes suggests megaloblastic anemia. Large misshapen granules may reflect the inherited Chédiak-Higashi syndrome.

Eosinophils are slightly larger than neutrophils, have bilobed nuclei, and contain large red granules. Diseases of eosinophils are associated with too many of them rather than any morphologic or qualitative change. They normally total less than one-thirtieth the number of neutrophils. Basophils are even more rare than eosinophils in the blood. They have large dark blue granules and may be increased as part of chronic myeloid leukemia.

Lymphocytes can be present in several morphologic forms. Most common in healthy individuals are small lymphocytes with a small dark nucleus and scarce cytoplasm. In the presence of viral infections, more of the lymphocytes are larger, about the size of neutrophils, with abundant cytoplasm and a less condensed nuclear chromatin. These cells are called *reactive lymphocytes*. About 1% of lymphocytes are larger and contain blue granules in a light blue cytoplasm; they are called *large granular lymphocytes*. In chronic lymphoid leukemia, the small lymphocytes are increased in number, and many of them are ruptured in making the blood smear, leaving a smudge of nuclear material without a surrounding cytoplasm or cell membrane; they are called *smudge cells* and are rare in the absence of chronic lymphoid leukemia.

Monocytes are the largest white blood cells, ranging from 15 to 22 μm in diameter. The nucleus can take on a variety of shapes but usually appears to be folded; the cytoplasm is gray.

Abnormal cells may appear in the blood. Most often, the abnormal cells originate from neoplasms of bone marrow–derived cells, including lymphoid cells, myeloid cells, and occasionally red cells. More rarely, other types of tumors can get access to the bloodstream, and rare epithelial malignant cells may be identified. The chances of seeing such abnormal cells are increased by examining blood smears made from buffy coats, the layer of cells that is visible on top of sedimenting red cells when blood is left in the test tube for an hour. Smears made from finger sticks may include rare endothelial cells.

ACKNOWLEDGMENT

Figures in this chapter were borrowed from Williams Hematology, 7th edition, M Lichtman et al (eds). New York, McGraw-Hill, 2005; Hematology in General Practice, 4th edition, RS Hillman, KA Ault. New York, McGraw-Hill, 2005.

66

Anemia and Polycythemia

Thomas G. DeLoughery, Dan L. Longo

Anemia is one of the most common medical problems in the world, affecting almost 2 billion people, and it is a significant source of morbidity and reduced quality of life. Many causes of anemia are treatable, and some forms of anemia can be a clue to underlying disorders.

NORMAL RED CELL PRODUCTION

Hematopoiesis is the process by which the formed elements of blood are produced. The process is regulated through a series of steps beginning with the hematopoietic stem cell. Stem cells are capable of producing red cells, all classes of granulocytes, monocytes, platelets, and the cells of the immune system. The precise molecular mechanism by which the stem cell becomes committed to a given lineage is not fully defined. However, experiments in mice suggest that erythroid cells come from a common erythroid/megakaryocytic progenitor that does not develop in the absence of expression of the GATA-1 and FOG-1 (friend of GATA-1) transcription factors **(Chap. 101)**. Following lineage commitment, hematopoietic progenitor and precursor cells become increasingly under the regulatory influence of growth factors and hormones. For red cell production, erythropoietin (EPO) is the primary regulatory hormone. EPO is required for the maintenance of committed erythroid progenitor cells that, in the absence of the hormone, undergo programmed cell death (apoptosis). The regulated process of red cell production is erythropoiesis, and its key elements are illustrated in **Fig. 66-1**.

In the bone marrow, the first morphologically recognizable erythroid precursor is the pronormoblast. This cell can undergo four to five cell divisions, which result in the production of 16–32 mature red cells. With increased EPO production, or the administration of EPO as a drug, early progenitor cell numbers are amplified and, in turn, give rise to increased numbers of erythrocytes. The regulation of EPO production is linked to tissue oxygenation.

In mammals, oxygen is transported to tissues bound to the hemoglobin contained within circulating red cells. The mature red cell is 8 μm in diameter, anucleate, discoid in shape, and extremely pliable in order to traverse the microcirculation successfully; its membrane integrity is maintained by the intracellular generation of ATP. The biconcave disk provides the greatest amount of surface area for a given volume; this maximizes oxygen delivery. Normal red cell production results in the daily replacement of 0.8–1% of all circulating red cells in the body, since the average red cell lives 100–120 days. The organ responsible for red cell production is called the erythron. The erythron is a dynamic organ made up of a rapidly proliferating pool of marrow erythroid precursor cells and a large mass of mature circulating red

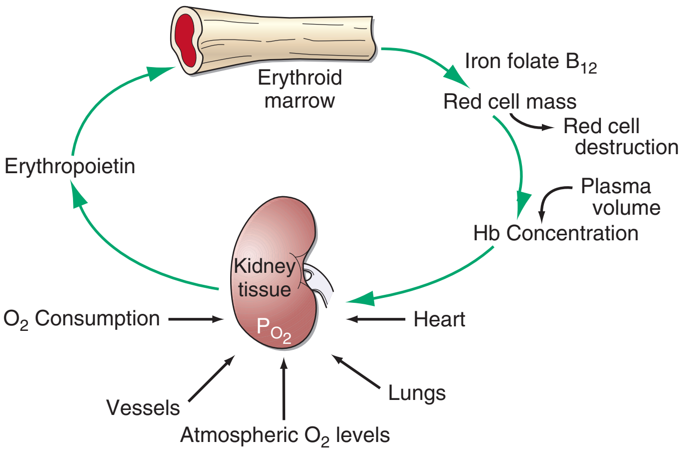

FIGURE 66-1 The physiologic regulation of red cell production by tissue oxygen tension. Hb, hemoglobin.

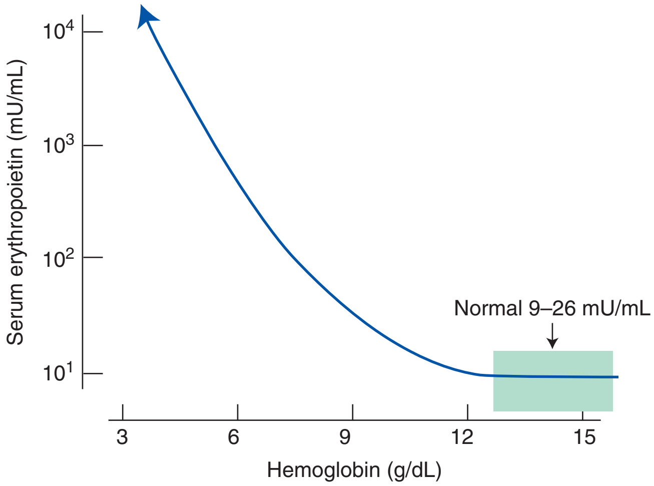

FIGURE 66-2 Erythropoietin (EPO) levels in response to anemia. When the hemoglobin level falls to 120 g/L (12 g/dL), plasma EPO levels increase logarithmically. In the presence of chronic kidney disease or chronic inflammation, EPO levels are typically lower than expected for the degree of anemia. As individuals age, the level of EPO needed to sustain normal hemoglobin levels appears to increase. *(Reproduced with permission from RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)*

blood cells. The size of the red cell mass reflects the balance of red cell production and destruction. The physiologic basis of red cell production and destruction provides an understanding of the mechanisms that can lead to anemia.

The physiologic regulator of red cell production, the glycoprotein hormone EPO, is produced and released by peritubular capillary lining cells within the kidney. These cells are highly specialized epithelial-like cells. A small amount of EPO is produced by hepatocytes. The fundamental stimulus for EPO production is the availability of oxygen (O_2) for tissue metabolic needs. Key to EPO gene regulation is hypoxia-inducible factor (HIF)-1α. In the presence of O_2, HIF-1α is hydroxylated at a key proline, allowing HIF-1α to be ubiquitinated and degraded via the proteasome pathway. If O_2 becomes limiting, this critical hydroxylation step does not occur, allowing HIF-1α to partner with other proteins, translocate to the nucleus, and upregulate the expression of the EPO gene, among others.

Impaired O_2 delivery to the kidney can result from a decreased red cell mass (anemia), impaired O_2 loading of the hemoglobin molecule, a high O_2 affinity mutant hemoglobin (hypoxemia), or, rarely, impaired blood flow to the kidney (e.g., renal artery stenosis). EPO governs the day-to-day production of red cells, and ambient levels of the hormone can be measured in the plasma by sensitive immunoassays with the normal level being 10–25 U/L. When the hemoglobin concentration falls below 100–120 g/L (10–12 g/dL), plasma EPO levels increase in proportion to the severity of the anemia **(Fig. 66-2)**. In circulation, EPO has half-life time of 6–9 h. EPO acts by binding to specific receptors on the surface of marrow erythroid precursors, inducing them to proliferate and to mature. With EPO stimulation, red cell production can increase four- to fivefold within a 1- to 2-week period, but only in the presence of adequate nutrients, especially iron. The functional capacity of the erythron, therefore, requires normal renal production of EPO, a functioning erythroid marrow, and an adequate supply of substrates for hemoglobin synthesis. A defect in any of these key components can lead to anemia.

APPROACH TO THE DIAGNOSIS OF ANEMIA

There are four initial steps in the diagnosis of anemia. The first step is a good history; one should ask about previous episodes of anemia, any previous therapy such as iron pills or transfusions, family history of anemia, and being a blood donor. Attention should also be paid to symptoms of disease that can lead to anemia. For example, diarrhea can be a sign of celiac disease or inflammatory bowel disease. In people who menstruate, a good menstrual history should be taken, including duration of periods, number of pads/tampons used, and passing large clots.

The physical exam should be focused on the consequences of anemia such as a cardiac flow murmur and clues to the cause of the anemia such as the presence of splenomegaly or blood in the stool.

A review of the blood smear is a crucial part of any evaluation for anemia as changes in red cell morphology can point to specific causes of anemia. Measuring the number of new red cells—the reticulocyte count—assesses the function of the bone marrow. Both of these tests are discussed in more detail below.

SIGNS AND SYMPTOMS OF ANEMIA

Patients who gradually develop anemia over months can tolerate amazingly low hemoglobin levels due to compensatory mechanisms. The overall health status of an individual will also determine their response to anemia. Since blood delivers oxygen, many of the clinical signs are related to lack of oxygen delivery, such as tiredness and shortness of breath. On exam, this is manifested by paleness of the mucosa/conjunctiva and resting tachycardia. If patients have atherosclerosis, they may suffer ischemic symptoms such as angina or transient ischemic attacks/strokes. Also, in cases of nutritional deficiency, there may be symptoms related to that such as pica in iron deficiency or neuropathy in B_{12} deficiency. In general, the signs and symptoms of anemia are very unreliable in predicting the patient's hematocrit.

COMPENSATION FOR ANEMIA

The body has a tremendous ability to compensate for anemia. This compensation improves over time, so inherited anemias or those that occur gradually are better tolerated than acute anemias. There are three physiologic compensatory mechanisms for anemia, which are described below.

The first compensatory mechanism is an increase in cardiac output. Oxygen delivery to tissues is a function of cardiac output times hemoglobin, so if the hemoglobin is lower, the cardiac output rises to compensate. Therefore, patients with a limited cardiac reserve will have symptoms of anemia at a higher hematocrit than patients with normal cardiac function. This compensatory mechanism can occur within minutes.

The second compensatory mechanism is via increased levels of 2,3-disphosphoglyceric acid (2,3-DPG). 2,3-DPG decreases oxygen affinity for hemoglobin by increasing the stability of deoxygenated hemoglobin. While this may seem counterintuitive, this decreased affinity leads to more oxygen delivery in the tissues. The high ambient oxygen tension in the alveoli results in full oxygenation of hemoglobin despite this decreased oxygen affinity. This mechanism takes place over hours to days.

Finally, over time, plasma volume increases which preserves cardiac output and maintains blood pressure and in theory may lower blood viscosity. In some cases, this increased plasma volume can overwhelm the heart, leading to edema and other signs of heart failure. This compensation for anemia occurs over weeks.

LABORATORY TESTING

Because most anemia develops slowly, anemia is most often detected by finding low hemoglobin and hematocrit levels on a complete blood count. Current testing of the blood count is performed by electronic cell counters that can directly measure red cell size (mean corpuscular volume [MCV]), number of red cells, and hemoglobin levels with a variety of measurements derived from these values. For example:

$$\text{Hematocrit} = \frac{\text{Red cell number/liter} \times \text{MCV}}{10}$$

The classic red cell indices reflect red cell size and hemoglobin concentration **(Table 66-1)**. The index most commonly used is the MCV because the mean corpuscular hemoglobin and mean corpuscular hemoglobin concentration will trend with the MCV.

Normal variations in hemoglobin and hematocrit with age are shown in **Table 66-2**. A normal peripheral blood smear is shown in **Fig. 66-3**.

TABLE 66-1 Red Cell Indices

INDEX	FORMULA	NORMAL VALUES	COMMENT
Mean corpuscular volume (MCV)	Hct/RBC count × 10	85–95 fL	RBC size
Mean corpuscular hemoglobin concentration	Hgb/Hct × 100	33.8–34.2 g/dL	Changes very little in most cases of anemia; therefore, of limited value
Mean corpuscular hemoglobin (MCH)	Hgb/RBC count × 10	28.5–32.3 pg	Varies linearly with MCV; therefore, of limited additional value

Abbreviations: Hct, hematocrit; Hgb, hemoglobin; RBC, red blood cell.

BLOOD SMEAR

Blood is one of the few tissues of the body that allows direct visualization. The morphology of the red cell on the blood smear can be a diagnostic clue to a variety of anemias. As a complement to the red cell indices, the blood smear also reveals variations in cell size (*anisocytosis*) and shape (*poikilocytosis*). The degree of anisocytosis usually correlates with increases in the red cell distribution width (RDW) or the range of cell sizes. Poikilocytosis suggests a defect in the maturation of red cell precursors in the bone marrow or fragmentation of circulating red cells. The blood smear may also reveal *polychromasia*—red cells that are slightly larger than normal and grayish blue in color on the Wright-Giemsa stain. These cells are reticulocytes that have been released prematurely from the bone marrow, and their color represents residual amounts of ribosomal RNA. These cells appear in circulation in response to EPO stimulation or to architectural damage of the bone marrow (e.g., fibrosis, infiltration of the marrow by malignant cells) that results in their disordered release from the marrow. The appearance of nucleated red cells, Howell-Jolly bodies, target cells, sickle cells, and other red cell morphology changes may provide clues to specific disorders **(Figs. 66-4 to 66-12)**. (See also **Table 66-3** and the Atlas of Hematology, **Chap. A6**.)

BURR CELLS

Also called echinocytes, these cells have multiple small projections. This can be a laboratory artifact but is most often seen with liver disease or uremia.

ELLIPTOCYTES

These elongated red cells may be seen most commonly in heredity elliptocytosis or with severe iron deficiency.

HOWELL-JOLLY BODIES

These are small remnants of the red cell nucleus. Usually these are rapidly cleared by the spleen, but if the spleen is missing or not functioning, these red cell inclusions may be seen.

HYPOCHROMIA/MICROCYTOSIS

These are cells that have larger areas of central pallor than normal. This is a sign of inadequate hemoglobinization of the red cell. This can be seen in iron deficiency, thalassemia, or in the rare sideroblastic anemia.

TABLE 66-2 Changes in Normal Hemoglobin/Hematocrit Values with Age, Sex, and Pregnancy

AGE/SEX	HEMOGLOBIN, g/dL	HEMATOCRIT, %
At birth	17	52
Childhood	12	36
Adolescence	13	40
Adult man	16 (±2)	47 (±6)
Adult woman (menstruating)	13 (±2)	40 (±6)
Adult woman (postmenopausal)	14 (±2)	42 (±6)
During pregnancy	12 (±2)	37 (±6)

Source: From RS Hillman et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.

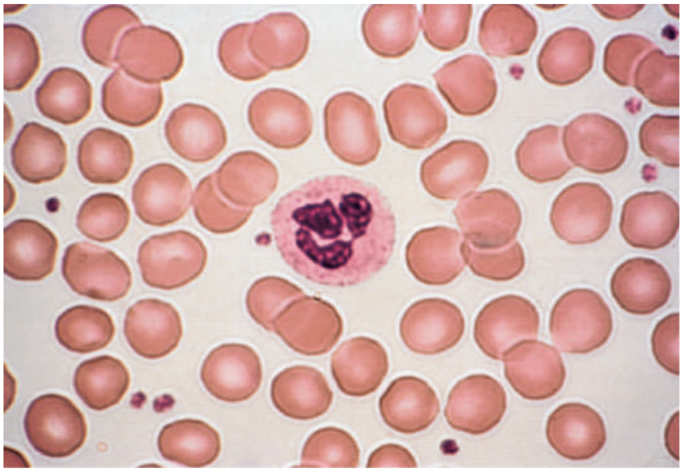

FIGURE 66-3 Normal blood smear (Wright stain). High-power field showing normal red cells, a neutrophil, and a few platelets. *(From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)*

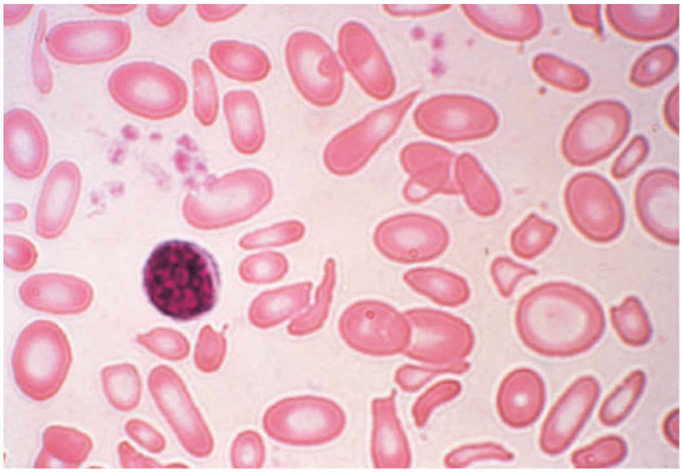

FIGURE 66-4 Severe iron-deficiency anemia. Microcytic and hypochromic red cells smaller than the nucleus of a lymphocyte associated with marked variation in size (anisocytosis) and shape (poikilocytosis). *(From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)*

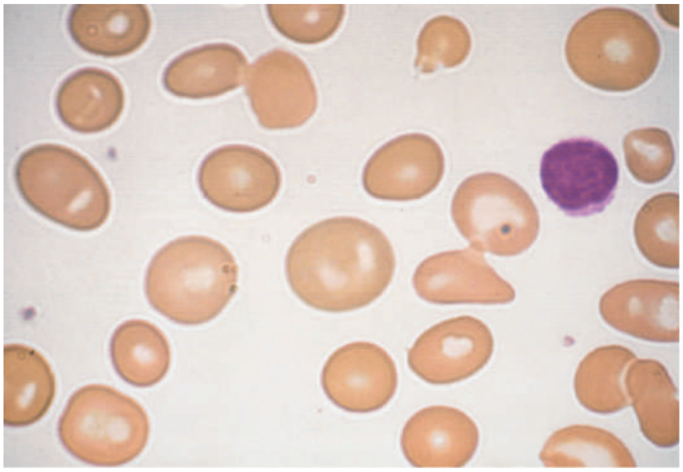

FIGURE 66-5 Macrocytosis. Red cells are larger than a small lymphocyte and well hemoglobinized. Often macrocytes are oval shaped (macro-ovalocytes). *(From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)*

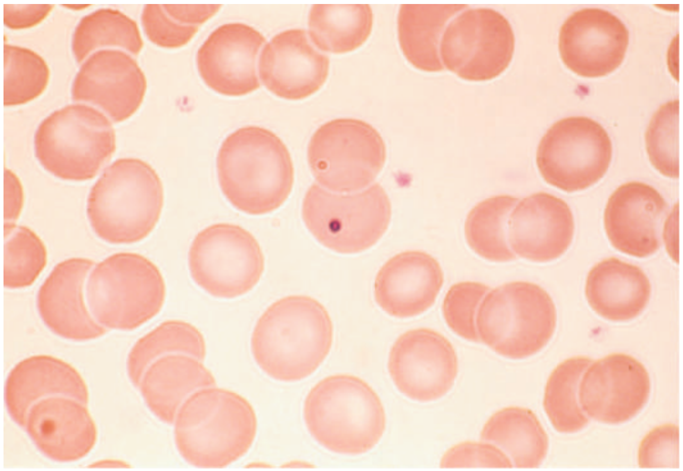

FIGURE 66-6 Howell-Jolly bodies. In the absence of a functional spleen, nuclear remnants are not culled from the red cells and remain as small homogeneously staining blue inclusions on Wright stain. *(From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

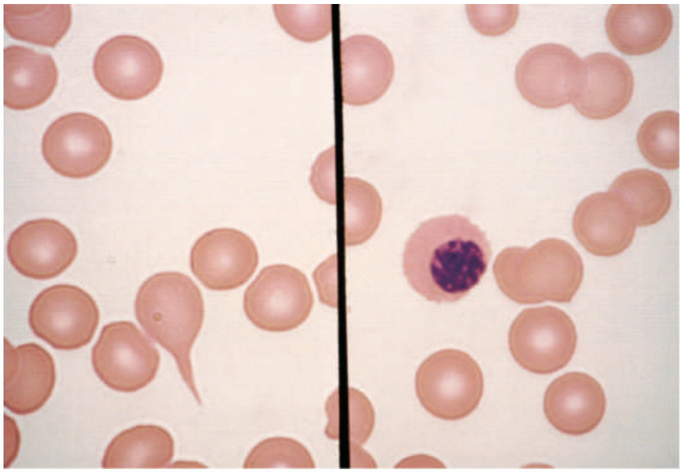

FIGURE 66-7 Red cell changes in myelofibrosis. The left panel shows a teardrop-shaped cell. The right panel shows a nucleated red cell. These forms can be seen in myelofibrosis. *(From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)*

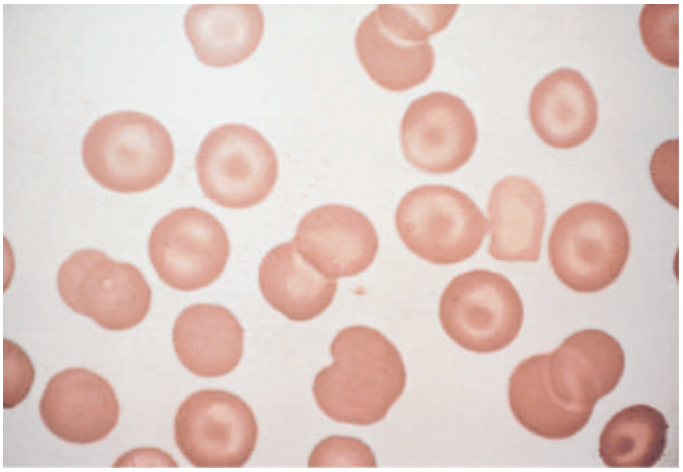

FIGURE 66-8 Target cells. Target cells have a bull's-eye appearance and are seen in thalassemia and in liver disease. *(From M Lichtman et al (eds): Williams Hematology, 7th ed. New York, McGraw-Hill, 2005; RS Hillman, KA Ault: Hematology in General Practice, 4th ed. New York, McGraw-Hill, 2005.)*

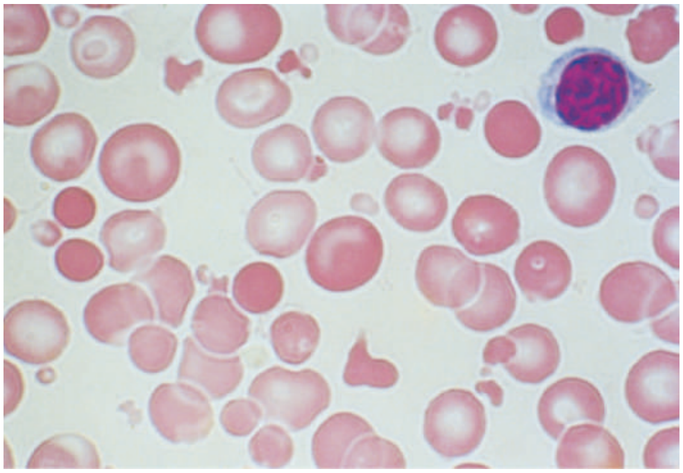

FIGURE 66-9 Red cell fragmentation. Red cells may become fragmented in the presence of foreign bodies in the circulation, such as mechanical heart valves, or in the setting of thermal injury. *(From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)*

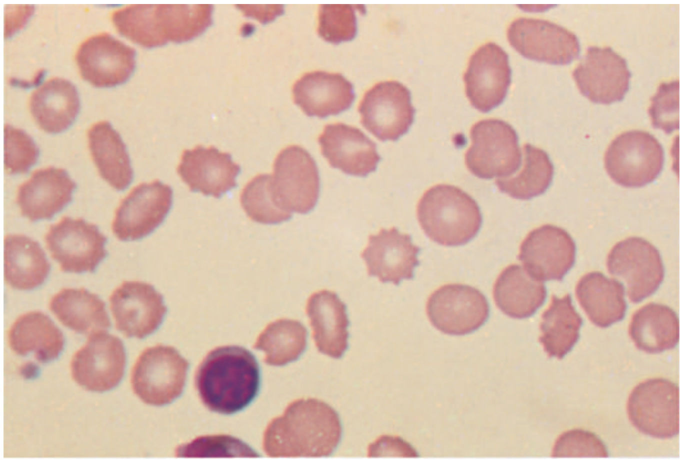

FIGURE 66-10 Uremia. The red cells in uremia may acquire numerous regularly spaced, small, spiny projections. Such cells, called burr cells or echinocytes, are readily distinguishable from irregularly spiculated acanthocytes shown in Fig. 66-11. *(From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)*

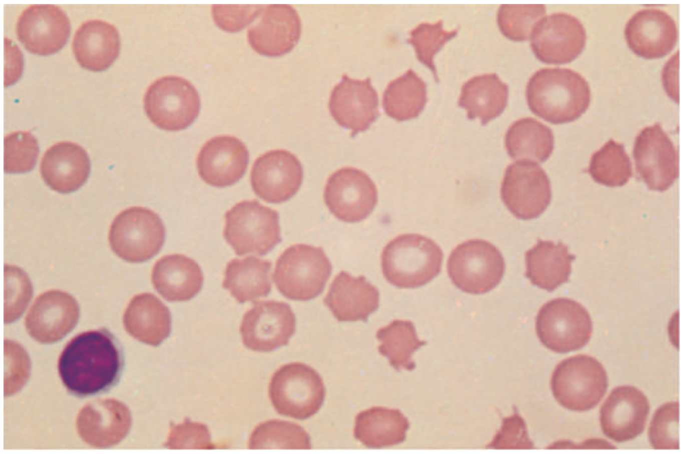

FIGURE 66-11 Spur cells. Spur cells are recognized as distorted red cells containing several irregularly distributed thorn-like projections. Cells with this morphologic abnormality are also called acanthocytes. *(From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)*

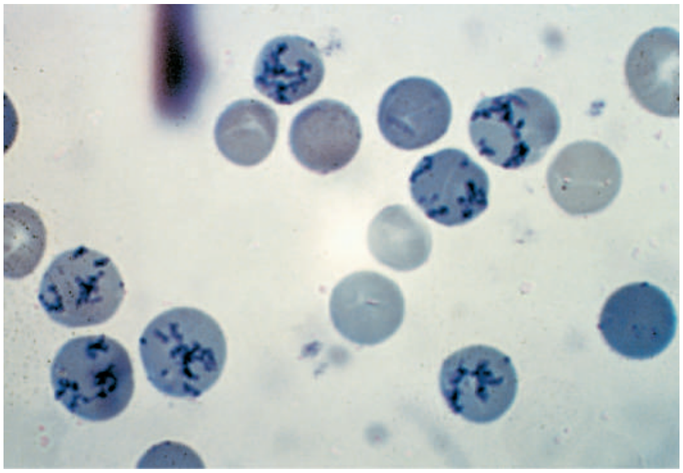

FIGURE 66-12 Reticulocytes. Methylene blue stain demonstrates residual RNA in newly made red cells. *(From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)*

MACRO-OVALOCYTES

These are red cells that are larger in size and slightly oval-shaped. This is caused by disruption of DNA synthesis in the developing red cell from B_{12}/folate deficiency, certain drugs, or bone marrow disorders.

NUCLEATED RED BLOOD CELLS

The nucleus of the red cell is cleared by the marrow before being released; any remaining are cleared by the spleen. The presence of nucleated red cells can be seen with asplenia, severe marrow stress such as massive bleeding or hemolysis, and many bone marrow disorders, especially those that disrupt the marrow architecture such as marrow fibrosis.

SCHISTOCYTES

These are fragments of red cells that have been physically disrupted by blockage in the blood vessels by platelets or fibrin strands or by external forces. They can be seen in a wide variety of processes such as thrombotic thrombocytopenic purpura or march hemoglobinuria.

SICKLE CELLS

As the name implies, the blood cells are sickle-shaped. This can be seen in hemoglobin defects, most commonly in sickle cell anemia.

TABLE 66-3 Red Cell Morphology

	PATHOPHYSIOLOGY	DISEASE STATES
Macro-ovalocytes	Hemoglobin excess (nuclear-cytoplasmic dyssynchrony)	B_{12} and/or folate deficiencies Myelodysplasia
Spherocytes	Loss of membrane	Hereditary spherocytosis Immune hemolytic anemia
Hypochromia	Hemoglobin deficiency (corresponds to low mean corpuscular hemoglobin concentration)	Iron deficiency Thalassemia Sideroblastic anemia Anemia of chronic disease
Schistocyte	Red blood cell fragmentation	Microangiopathic hemolysis Heart-valve hemolysis
Sickle cell	Hemoglobin polymerization	Sickle cell disease
Target cell	Relative membrane excess	Liver disease Thalassemia Hemoglobinopathy
Polychromatophilia	Persistence of polyribosomes (corresponds to high reticulocyte count)	Hemolytic anemia
Basophilic stippling	Pathologic precipitation of polyribosomes	Thalassemia Lead poisoning

SPHEROCYTES

These are cells that on the blood smear lack central pallor as they are spherical and not biconcave disks. This can be caused by any process that leads to the loss of red cell membrane. This membrane loss turns the biconcave disk into a sphere as this is the shape with the least amount of surface area for a given volume. Spherocytes are most commonly seen in autoimmune hemolytic anemia and hereditary spherocytosis.

SPUR CELLS

As opposed to burr cells, spur cells have fewer longer projections from the red cell and are most commonly seen in severe liver disease. They result from cholesterol crystal formation in the red cell membrane due to the abnormal lipid metabolism in liver disease. These crystals get caught in narrow passages of the spleen, resulting in these projections. They can also be seen in asplenic patients and those with McLeod blood group.

TARGET CELLS

These can be seen where there is a redundant red cell membrane. In thalassemia, there is less hemoglobin filling the red cell, and in liver disease, there is excess red cell membrane.

TEARDROP CELLS

These are cells that look like a teardrop and are most often seen with disruption of marrow architecture or severe iron deficiency.

THE RETICULOCYTE COUNT

Red cells still contain mRNA for about 24 h after being released by the marrow. This mRNA can be detected by staining, and these cells are called "reticulocytes." The number of reticulocytes present is a measure of red cell production and is helpful in separating increased destruction from anemias due to impaired red cell production.

The oldest method of determining reticulocyte counts is to stain the blood smear with new methylene blue and determine the percentage of red cells that take up the stain; this is the reticulocyte count. However, this needs to be adjusted for the hematocrit as the reticulocyte percentage will appear to increase with decreasing blood counts when the absolute count has not actually increased. For example, a reticulocyte count of 1% will increase to 2% with a hematocrit of 23% (Fig. 66-13). Therefore, the reticulocyte count needs to be corrected for the hematocrit:

$$\text{Corrected reticulocyte count} = \text{Measured reticulocyte count} \\ \times (\text{Patient hematocrit}/45\%)\ (\text{Normal hematocrit})$$

Many newer complete blood count machines can directly quantitate the reticulocytes present in a given volume of blood, and this is the **absolute reticulocyte count**. Since this is per volume, no adjustment for anemia is required.

$$\text{Absolute reticulocyte number} = \text{Reticulocyte count}/100 \\ \times \text{Red blood cell number}$$

BONE MARROW EXAM

The bone marrow may be easily sampled in a bedside procedure and can be helpful in the diagnosis of anemia or other hematologic processes. With bone marrow aspiration, the marrow is removed from the posterior iliac spine and stained with Wright's or Wright-Giemsa stains (photomicrographs of normal marrow [Fig. 66-14], marrow showing

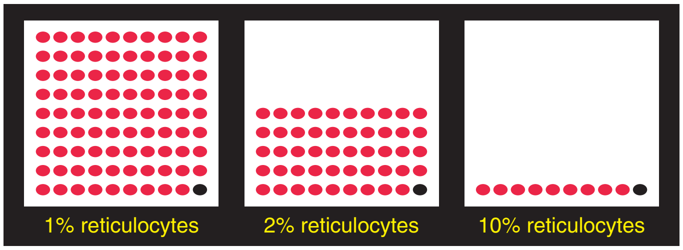

FIGURE 66-13 Interpretation of reticulocyte counts. If the absolute reticulocyte count is not provided, the percent reticulocytes must be adjusted based on the hematocrit.

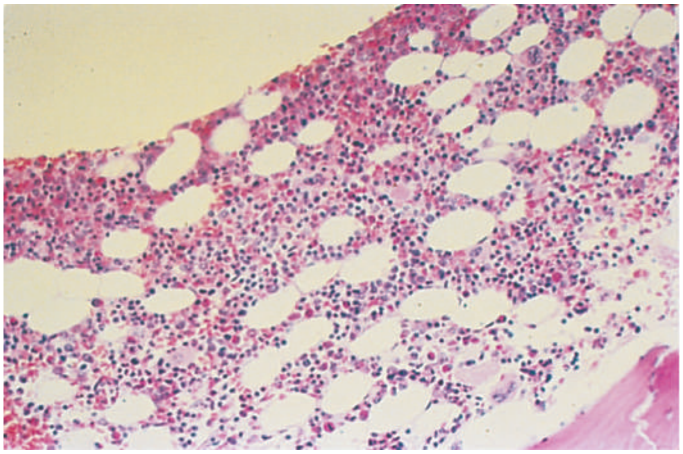

FIGURE 66-14 Normal bone marrow. This is a low-power view of a section of a normal bone marrow biopsy stained with hematoxylin and eosin (H&E). Note that the nucleated cellular elements account for ~40–50% and the fat (clear areas) accounts for ~50–60% of the area. *(From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)*

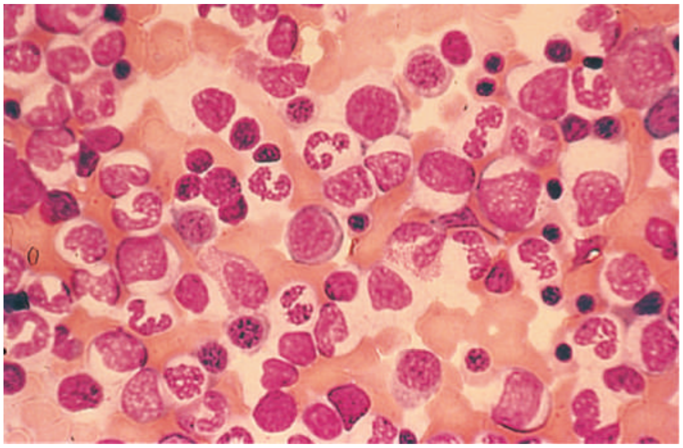

FIGURE 66-16 Myeloid hyperplasia. This marrow shows an increase in the fraction of cells in the myeloid or granulocytic lineage as might be seen in a normal marrow responding to infection. The myeloid/erythroid (M/E) ratio is >3:1. *(From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)*

erythroid hyperplasia **[Fig. 66-15]**, and marrow showing myeloid hyperplasia **[Fig. 66-16]** are shown). The biopsy disturbs the marrow architecture but allows for examination of individual cell morphology, determination of a differential cell count, and determination of the myeloid-to-erythroid ratio. This ratio is normally 2.5:1 (range 2:1–5:1). Bone marrow iron stores can be examined by use of an appropriate stain. Samples can also be drawn for specialized testing such as flow cytometry, genetic testing, and/or microbiological testing.

For a marrow biopsy, a "core" sample of marrow is removed intact from the iliac spine and then decalcified, sectioned, and stained with hematoxylin and eosin. This technique reveals the undisturbed marrow architecture and is useful for the determination of cellularity and the presence of abnormal marrow infiltrates or fibrosis. The normal bone marrow is ~50% cellular (i.e., half hematopoietic cells and half fat cells). In general, the marrow fat percentage is roughly equivalent to the patient's age.

A bone marrow examination is a necessary diagnostic study in situations where the blood and clinical findings suggest the possibility of marrow infiltration by abnormal or nonhematopoietic elements or the likelihood of generalized marrow dysfunction (e.g., aplastic anemia or myelodysplasia). In most situations, it is necessary to perform both an aspiration and a biopsy. The biopsy should precede the aspiration. If the aspiration is done first, the subsequent biopsy tends to be distorted by the bleeding induced by the aspiration.

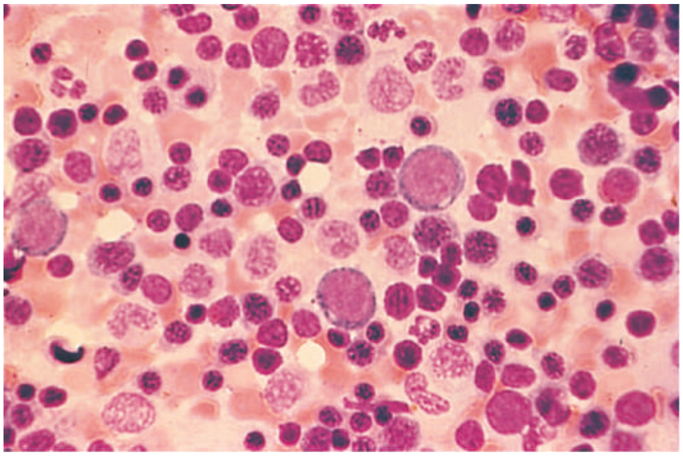

FIGURE 66-15 Erythroid hyperplasia. This marrow shows an increase in the fraction of cells in the erythroid lineage as might be seen when a normal marrow compensates for acute blood loss or hemolysis. The myeloid/erythroid (M/E) ratio is about 1:1. *(From RS Hillman et al: Hematology in Clinical Practice, 5th ed. New York, McGraw-Hill, 2010.)*

General indications for bone marrow examination are:

1. Circulating immature cells (e.g., blasts)
2. Severe pancytopenia
3. Very low reticulocyte counts (<0.1%)
4. Circulating nucleated red blood cells
5. Evidence of marrow infiltration (teardrop red cells, nucleated red blood cells)
6. Staging of certain malignancies (e.g., lymphoma)
7. Unexplained severe anemia

ANEMIA DEFINITION AND CLASSIFICATION

Anemia is simply defined as blood counts below normal for a given population.

There are two general ways of classifying anemia **(Table 66-4)**. A time-honored and still practical way is by the size of the red cell, as this can help guide diagnostic workup. Another method is by mechanism of anemia.

Anemia classification by red cell size was pioneered by the hematologist Max Wintrobe. In this classification, anemia is grouped by MCV, with those with smaller MCV called microcytic; normal MCV, normocytic; and larger MCV, macrocytic.

Microcytic anemia is due to any process that interferes with hemoglobin production; less hemoglobin leads to smaller red cells. This would include:

- Thalassemia: defects of hemoglobin protein synthesis
- Iron deficiency: unable to make heme
- Anemia of chronic disease: lack of iron delivery to red cell (note: the anemia of chronic disease is often associated with normal red cell volume; the mechanism remains the inhibition of iron reutilization based on cytokine inhibition)
- Sideroblastic anemias: defects of heme synthesis

Macrocytic anemia may be due to defects in DNA synthesis and has two subdivisions: round macrocytes and oval macrocytes. Oval macrocytes are due to defects in DNA synthesis, while round macrocytes are caused by membrane defects. Unlike the limited differential of microcytosis, many processes can lead to macrocytosis. The classic macrocytic anemias are due to B_{12} and folate deficiency, but these account for a minority of macrocytic anemias.

Normocytic anemia includes all other causes of anemia; the broad differential for this classification of anemia is a drawback of this anemia classification scheme.

An alternative method of classifying anemias is by mechanism. The absolute reticulocyte count is an important first indicator of mechanism. If reticulocytes are elevated, the mechanism of anemia is increased loss or destruction of red cells. If reticulocytes are low or

TABLE 66-4 Classification of Anemia

By Size of Red Cell
Microcytic
Anemia of inflammation
Iron deficiency
Sideroblastic
Thalassemia
Macrocytic
Oval Macrocytes
Vitamin B_{12} deficiency
Folate deficiency
Medications (chemotherapy, some antiseizure medications)
Myelodysplasia
Round Macrocytes
Alcohol use
Dysproteinemia
Hypothyroidism
Hypoxia
Liver disease
Reticulocytosis
Smoking
Normocytic
Aplastic anemia
Endocrinopathies
Marrow invasion
Myeloma
Pure red cell aplasia
Renal disease
By Mechanism
Hyperproduction
Bleeding
Hemolysis
Acquired
Autoimmune
Mechanical
Congenital
Hemibrain defect
Hemoglobinopathies
Enzyme defects
Underproduction
Nutritional Deficiency
Vitamin B_{12}
Copper
Folate
Iron
Vitamin C
Absence of Red Cell Precursors
Aplastic anemia
Pure red cell aplasia
Lack of Erythropoietin
Anemia of inflammation
Renal disease
Anemia of aging
Marrow Replacement
Granulomatous disease
Infection
Neoplasm
Stem Cell Defects
Acute leukemia
Chronic leukemia
Myelodysplasia

inappropriately low for the level of anemia, the mechanism is impaired production of red cells. Impaired production includes:

- Nutritional: deficiencies such as iron or B_{12}
- Marrow replacement by infection, cancer, granulomas
- Absence of red cell precursors: aplastic anemia, pure red cell aplasia
- Stem cell defects: myelodysplasia
- Lack of EPO: renal disease, anemia of inflammation, anemia of aging

HYPERPRODUCTION: HEMOLYSIS/BLEEDING

In general, the finding of an elevated reticulocyte count should raise the suspicion of hemolysis. Hemolysis is where red cell breakdown is accelerated. This can be due to extrinsic causes such as autoimmune hemolytic anemia or processes intrinsic to the red cell such as enzyme defects. Laboratories will show evidence of red cell breakdown including high lactate dehydrogenase (LDH; an enzyme found in abundance in red cells), low haptoglobin (a serum protein that salvages free hemoglobin), and increased blood/urine free hemoglobin. Often the blood smear will show evidence of hemolysis with abnormal cells such as schistocytes or spherocytes. Testing is then focused on finding the particular cause of hemolysis. For example, autoimmune hemolytic anemia will present with spherocytes, high LDH, and low haptoglobin; testing will show the presence of autoantibodies on red cells.

UNDERPRODUCTION

At least 75% of all cases of anemia are hypoproliferative in nature. The most common cause is mild to moderate iron deficiency or inflammation.

NUTRITIONAL

Production of red cells is dependent on a consistent supply of nutrients. The most vital nutrient is iron; four atoms are required for every hemoglobin molecule, with about 1 billion iron atoms in every red cell. Iron deficiency is the most common nutritional deficiency worldwide and is very common in premenopausal women due to obligate menstrual losses. The laboratory measurements that reflect the availability of iron for hemoglobin synthesis include the serum iron, the total iron-binding capacity (TIBC), and the percent transferrin saturation. The percent transferrin saturation is derived by dividing the serum iron level (× 100) by the TIBC. The normal serum iron ranges from 9 to 27 μmol/L (50–150 μg/dL), whereas the normal TIBC is 54–64 μmol/L (300–360 μg/dL); the normal transferrin saturation ranges from 25 to 50%. A diurnal variation in the serum iron leads to a variation in the percent transferrin saturation. The serum ferritin is used to evaluate total-body iron stores. Adult males have serum ferritin levels that average ~100 μg/L, corresponding to iron stores of ~1 g. Adult premenopausal females have lower serum ferritin levels averaging 30 μg/L, reflecting lower iron stores (~300 mg). A serum ferritin less than 30 μg/L indicates depletion of body iron stores. However, ferritin is also an acute-phase reactant and, in the presence of acute or chronic inflammation, may rise several-fold above baseline levels. As a rule, a serum ferritin >200 μg/L means there is at least some iron in tissue stores.

The classic finding of iron deficiency in the blood is a microcytic anemia, but this can be absent in early iron deficiency or when other confounding issues are present like liver disease. Laboratory findings show a low serum ferritin and may show a low iron saturation.

Vitamin B_{12} is another crucial nutrient. The red cell requires vitamin B_{12} in order to synthesize DNA, so deficiency results in impaired DNA synthesis. This leads to the classic hematologic finding of B_{12} deficiency—macrocytosis. In addition, the neutrophils will show hypersegmentation and the marrow will be remarkably hypercellular with large abnormal red cell precursors.

Folate is also required for red cell DNA synthesis; deficiencies present similar to B_{12} deficiency, and these two are grouped together as "megaloblastic anemia." In countries that supplement flour with folate, deficiencies are rare but may still be seen in those with a very poor diet or severe malabsorption.

Other nutrient deficiencies lead to anemia, but these tend to be more unusual. Copper deficiency is often associated with neurologic disease and neutropenia. Vitamin C deficiency can be associated with severe anemia as well as the classic clinical findings of scurvy.

MARROW REPLACEMENT

Several processes, most notably infections and neoplasms, can invade the marrow and crowd out hematopoietic elements. Many infections can occupy the marrow including infections such as histoplasmosis or tuberculosis. Metastasis to the marrow from any tumor can also lead to anemia. For example, patients with prostate cancer may have their marrow replaced by tumor, leading to severe anemia. More commonly in older patients, multiple myeloma can present with anemia.

In many cases of marrow replacement, the blood smear can show a myelophthisic picture with the presence of nucleated red cells, teardrop cells, and immature white cells. These findings mandate marrow examination.

LACK OF ERYTHROPOIETIN

As noted earlier, EPO is essential for red cell production. Since the kidney is the primary source of this hormone, renal disease is often associated with anemia. While anemia is almost always present with a glomerular filtration rate of <30 mL/min/1.73 m^2, some patients will be anemic with higher levels of renal function. Certain medications, most notably angiotensin-converting enzyme inhibitors, can also suppress EPO production.

Inflammatory cytokines such as tumor necrosis factor can decrease production of EPO. One of the hallmarks of the anemia of inflammation is a low EPO level for any given degree of anemia. In addition, inflammation will increase levels of hepcidin, which then blocks iron absorption and its release from stores.

Anemia is common in people over age 65 years; 11% of community-dwelling and up to 40% of nursing home residents over age 65 are anemic. Anemia is associated with increased risk of death, hospitalization, and frailty. EPO levels tend to be lower than expected for the degree of anemia, and administration of EPO can increase the hemoglobin level, but it is unclear whether the adverse consequences of anemia are reduced by treatment.

ABSENCE OF RED CELL PRECURSORS

Aplastic anemia is a disease where the marrow is very hypocellular. This can be caused by autoimmune processes, reactions to certain medications, or as a result of toxins/radiation. The bone marrow shows markedly reduced levels of all precursors, and other blood elements such as platelets and neutrophils will be reduced.

Pure red cell aplasia is defined as only the red cell precursors in the marrow being reduced. Autoimmune processes that target only the red cell can lead to this, as well as infections with parvovirus B19. The patient will present with a very low reticulocyte count, and the bone marrow will show absent or markedly reduced erythroid cells.

STEM CELL DEFECTS

Finally, processes intrinsic to the marrow can lead to anemia. Primary neoplasms of the marrow such as acute myelogenous leukemia or chronic myelogenous leukemia will often have a component of severe anemia. Patients most often will have elevated white cell counts with immature cells present; rare patients will present only with anemia.

Myelodysplastic syndromes are caused by stem cell defects that lead to impaired marrow function. Patients can present with only a macrocytic anemia but can also present with pancytopenia. The natural history can vary; some patients' courses are measured in years while others rapidly evolve into acute leukemia.

Diagnosis of leukemia and myelodysplasia is most often made by bone marrow testing. Currently, the pathologic examination is augmented by molecular testing for mutations that are diagnostic of leukemia or myelodysplasia.

EVALUATION OF ANEMIA

The results of the complete blood count and reticulocyte count together can guide further testing **(Fig. 66-17)**. If the reticulocyte count is high and bleeding has been ruled out, then specific testing for hemolysis can be performed, including LDH, haptoglobin, and direct antibody testing. If there are clues present on the blood smear such as sickle cells, then focused testing can be done.

If the reticulocyte count is not elevated, workup is guided by the MCV. If the MCV is low, a ferritin should be performed, and if normal, then the patient should be assessed for thalassemia. Another clue to distinguishing iron deficiency from thalassemia is the RDW. Iron deficiency is characterized by a high RDW because of anisocytosis. The red cells in thalassemia are more homogeneous in volume and have a low RDW. Anemia of inflammation is a diagnosis of exclusion and usually occurs in the presence of an inflammatory disorder such as cancer or infection.

The workup of macrocytosis is guided by the blood smear. If signs of megaloblastic anemia are seen (hypersegmented neutrophils and macro-ovalocytes), then B$_{12}$ and folate levels should be assessed. Some experts distinguish large round red cells from large oval red cells on smear. If macro-ovalocytes are seen and nutritional tests are normal, this raises the concern for myelodysplasia, and bone marrow testing is the next step. With round macrocytosis, often the history can provide clues, such as presence of liver disease or alcoholism, for example.

Workup of normocytic anemia involves assessment of renal function, EPO levels, and other disease processes that can lead to anemia and consideration of marrow exam if no other cause is apparent.

Additional laboratory tests may be of value in confirming specific diagnoses. **For details of these tests and how they are applied in specific disorders, see Chaps. 102 to 106.**

TREATMENT

Anemia

While definitive treatment of anemia requires knowing the cause, some patients may need treatment with transfusions to support them during the diagnostic workup. Clear indications for blood transfusion are hypotension or signs of cardiac compromise such as angina. For some patients such as those with bone marrow failure, transfusion may be needed to support them while definitive therapy is being performed.

Often, the cause of the anemia is multifactorial. For example, a patient with severe rheumatoid arthritis who has been taking anti-inflammatory drugs may have a hypoproliferative anemia associated with chronic inflammation as well as chronic blood loss associated with intermittent gastrointestinal bleeding. In every circumstance, it is important to evaluate the patient's iron status fully before and during the treatment of any anemia. **Transfusion is discussed in Chap. 118; iron therapy is discussed in Chap. 102; treatment of megaloblastic anemia is discussed in Chap. 104; treatment of other entities is discussed in their respective chapters (sickle cell anemia, Chap. 103; hemolytic anemias, Chap. 105; aplastic anemia and myelodysplasia, Chap. 107).**

Therapeutic options for the treatment of anemias have expanded dramatically during the past 30 years. Blood component therapy is available and safe. Recombinant EPO as an adjunct to anemia management has transformed the lives of patients with chronic renal failure on dialysis and reduced transfusion needs of anemic cancer patients receiving chemotherapy. Transforming growth factor β inhibitors (anemia associated with myelodysplastic syndrome), complement inhibitors (associated with paroxysmal nocturnal hemoglobinuria), and other therapies are making a difference in the quality of life in selected types of anemia. Eventually, patients with inherited disorders of globin synthesis or mutations in the globin gene, such as sickle cell disease, may benefit from the successful introduction of targeted genetic therapy **(Chap. 483)**.

POLYCYTHEMIA

Polycythemia is defined as an increase in the hemoglobin above normal. This increase may be real or only apparent because of a decrease in plasma volume (spurious or relative polycythemia). The term erythrocytosis may be used interchangeably with polycythemia, but some draw a distinction between them: erythrocytosis implies documentation of

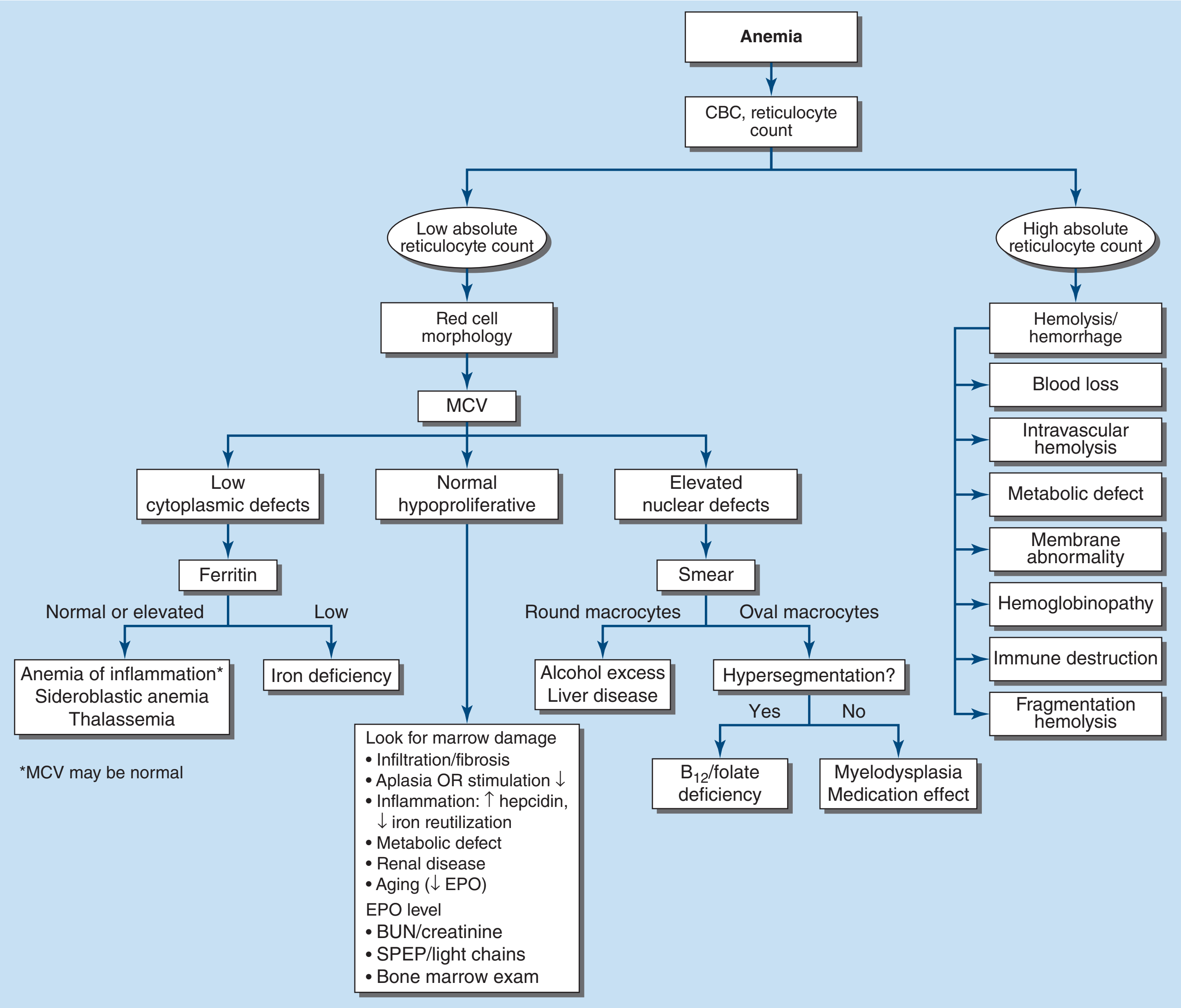
Anemia
CBC, reticulocyte count
Low absolute reticulocyte count
High absolute reticulocyte count
Red cell morphology
MCV
Low cytoplasmic defects
Normal hypoproliferative
Elevated nuclear defects
Ferritin
Normal or elevated
Low
Anemia of inflammation* Sideroblastic anemia Thalassemia
Iron deficiency
*MCV may be normal
Smear
Round macrocytes
Oval macrocytes
Alcohol excess Liver disease
Hypersegmentation?
Yes
No
B12/folate deficiency
Myelodysplasia Medication effect
Look for marrow damage
• Infiltration/fibrosis
• Aplasia OR stimulation ↓
• Inflammation: ↑ hepcidin, ↓ iron reutilization
• Metabolic defect
• Renal disease
• Aging (↓ EPO)
EPO level
• BUN/creatinine
• SPEP/light chains
• Bone marrow exam
Hemolysis/ hemorrhage
Blood loss
Intravascular hemolysis
Metabolic defect
Membrane abnormality
Hemoglobinopathy
Immune destruction
Fragmentation hemolysis

FIGURE 66-17 The physiologic classification of anemia. BUN, blood urea nitrogen; CBC, complete blood count; EPO, erythropoietin level; MCV, mean corpuscular volume; SPEP, serum protein electrophoresis.

increased red cell mass, whereas polycythemia refers to any increase in red cells. Often patients with polycythemia are detected through an incidental finding of elevated hemoglobin or hematocrit levels. Concern that the hemoglobin level may be abnormally high is usually triggered at 17 g/dL (170 g/L) for men and 15 g/dL (150 g/L) for women. Hematocrit levels of >50% in men or >45% in women may be abnormal. Hematocrit levels of >60% in men and >55% in women are almost invariably associated with an increased red cell mass. Given that the machine that quantitates red cell parameters actually measures hemoglobin concentrations and calculates hematocrits, hemoglobin levels may be a better index.

Features of the clinical history that are useful in the differential diagnosis include smoking, currently living at high altitude, a history of diuretic use, congenital heart disease, sleep apnea, or chronic lung disease. It is also useful to inquire about the use of testosterone, EPO, and SGLT2 (sodium-glucose transport protein 2) inhibitors (gliflozins).

Patients with polycythemia may be asymptomatic or experience symptoms related to the increased red cell mass or the underlying disease process that leads to the increased red cell mass. The dominant symptoms from an increased red cell mass are related to hyperviscosity and thrombosis (both venous and arterial), because the blood viscosity increases logarithmically at hematocrit levels of >55%. Manifestations include neurologic symptoms such as vertigo, tinnitus, headache, and visual disturbances. Hypertension is often present. Patients with *polycythemia vera* may have aquagenic pruritus, symptoms related to hepatosplenomegaly, easy bruising, epistaxis, or bleeding from the gastrointestinal tract. Peptic ulcer disease is common. Such patients also may present with digital ischemia, Budd-Chiari syndrome, or hepatic or splenic/mesenteric vein thrombosis. Patients with hypoxemia may develop cyanosis on minimal exertion or have headache, impaired mental acuity, and fatigue.

The physical examination usually reveals a ruddy complexion. Splenomegaly favors polycythemia vera as the diagnosis **(Chap. 108)**. The presence of cyanosis or evidence of a right-to-left shunt suggests congenital heart disease presenting in the adult, particularly tetralogy of Fallot or Eisenmenger's syndrome **(Chap. 280)**. Increased blood viscosity raises pulmonary artery pressure; hypoxemia can lead to increased pulmonary vascular resistance. Together, these factors can produce cor pulmonale.

Polycythemia can be spurious (related to a decrease in plasma volume; Gaisbock's syndrome), primary, or secondary in origin. The secondary causes are nearly all mediated by EPO: either a physiologically adapted appropriate level based on tissue hypoxia (lung disease, high altitude, carbon monoxide [CO] poisoning, high-affinity hemoglobinopathy) or an abnormal overproduction (renal cysts, renal artery stenosis, tumors with ectopic EPO production). A rare familial form of

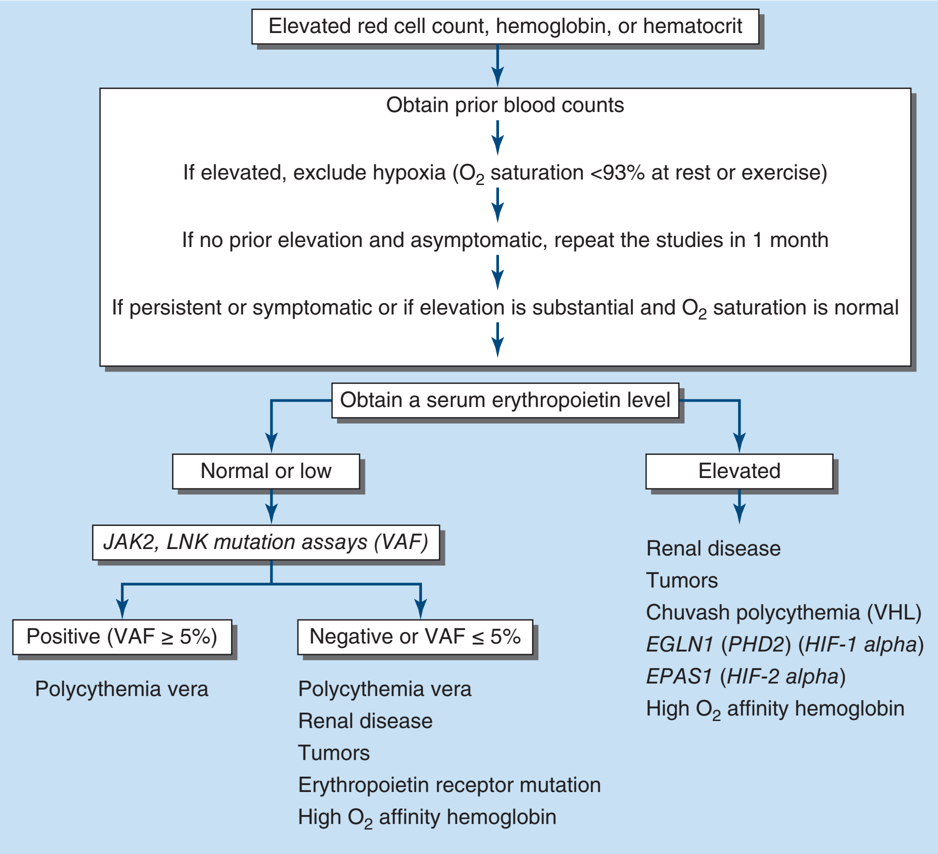
Elevated red cell count, hemoglobin, or hematocrit
Obtain prior blood counts
If elevated, exclude hypoxia (O2 saturation <93% at rest or exercise)
If no prior elevation and asymptomatic, repeat the studies in 1 month
If persistent or symptomatic or if elevation is substantial and O2 saturation is normal
Obtain a serum erythropoietin level
Normal or low
Elevated
JAK2, LNK mutation assays (VAF)
Positive (VAF ≥ 5%)
Negative or VAF ≤ 5%
Polycythemia vera
Polycythemia vera
Renal disease
Tumors
Erythropoietin receptor mutation
High O2 affinity hemoglobin
Renal disease
Tumors
Chuvash polycythemia (VHL)
EGLN1 (PHD2) (HIF-1 alpha)
EPAS1 (HIF-2 alpha)
High O2 affinity hemoglobin

FIGURE 66-18 An approach to the differential diagnosis of patients with an elevated hemoglobin (possible polycythemia). VAF, variant allele frequency; VHL, von Hippel-Lindau syndrome. *(Reproduced with permission from Jerry L. Spivak.)*

polycythemia is associated with normal EPO levels but hyperresponsive EPO receptors due to mutations. Rare hemochromatosis (*HFE* mutations) may have elevated hemoglobin levels.

APPROACH TO THE PATIENT

Polycythemia

As shown in **Fig. 66-18**, the first step is to search for earlier blood counts that might suggest the chronicity of the increase. Ideally one would document the presence of an increased red cell mass; however, the dependence of this technique on radiolabeled red cells has led to its abandonment. If the red cell mass is normal (<36 mL/kg in men, <32 mL/kg in women), the patient has spurious or relative polycythemia. If the red cell mass is increased (>36 mL/kg in men, >32 mL/kg in women), serum EPO levels should be measured. It must be acknowledged that measurement of red cell mass is a physiologic approach to distinguishing polycythemia, and because of the use of radionuclide-labeled red cells, it is rarely performed. It is more common to measure EPO levels in a person with an elevated hemoglobin level or hematocrit once it has been documented that the patient is not hypoxic (i.e., O_2 saturation is >93%). If EPO levels are low or unmeasurable, the patient most likely has polycythemia vera. A mutation in *JAK2* (Val617Phe), a key member of the cytokine intracellular signaling pathway, can be found in 90–95% of patients with polycythemia vera. Many of those without this particular *JAK2* mutation have mutations in exon 12. If EPO levels are low, check for *JAK2* mutation(s), and perform an abdominal ultrasound to assess spleen size. Tests that support the diagnosis of polycythemia vera include elevated white blood cell count, increased absolute basophil count, and thrombocytosis. In practice, many physicians order EPO levels and assessment for *JAK2* mutations at the same time.

If serum EPO levels are elevated, one needs to distinguish whether the elevation is a physiologic response to hypoxia or related to autonomous EPO production. Patients with low arterial O_2 saturation (<92%) should be further evaluated for the presence of heart or lung disease, if they are not living at high altitude. Patients with normal O_2 saturation who are smokers may have elevated EPO levels because of CO displacement of O_2. If carboxyhemoglobin (COHb) levels are high, the diagnosis is "smoker's polycythemia." Such patients should be urged to stop smoking. Those who cannot stop smoking require phlebotomy to control their polycythemia. Patients with normal O_2 saturation who do not smoke either have an abnormal hemoglobin that does not deliver O_2 to the tissues (evaluated by finding elevated O_2–hemoglobin affinity) or have a source of EPO production that is not responding to the normal feedback inhibition. Further workup is dictated by the differential diagnosis of EPO-producing neoplasms. Hepatoma, uterine leiomyoma, and renal cancer or cysts are all detectable with abdominopelvic computed tomography scans. Cerebellar hemangiomas may produce EPO, but they present with localizing neurologic signs and symptoms rather than polycythemia-related symptoms.

FURTHER READING

Camaschella C: Iron deficiency. Blood 133:30, 2019.

Hillman RS et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010.

McMullin MF et al: Guidelines for the diagnosis, investigation and management of polycythaemia/erythrocytosis. Br J Haematol 130:174, 2005.

Sankaran VG, Weiss MJ: Anemia: Progress in molecular mechanisms and therapies. Nat Med 21:221, 2015.

Spivak JL: How I manage polycythemia vera. Blood 134:341, 2019.

67

Disorders of Granulocytes and Monocytes

Steven M. Holland, John I. Gallin*

Leukocytes, the major cells comprising inflammatory and immune responses, include neutrophils, T and B lymphocytes, natural killer (NK) cells, mononuclear phagocytes (blood monocytes and tissue macrophages), eosinophils, and basophils. These cells have specific functions, such as antibody production by B lymphocytes or destruction of bacteria by neutrophils, but in no single infectious disease is the exact role of the cell types completely established. Thus, whereas neutrophils are classically thought to be critical to host defense against bacteria, they may also play important roles in defense against viral infections.

The blood delivers leukocytes to the various tissues from the bone marrow, where they are produced. Normal blood leukocyte counts are 4.3–10.8 × 10^9/L, with neutrophils representing 45–74% of the cells, bands 0–4%, lymphocytes 16–45%, monocytes 4–10%, eosinophils 0–7%, and basophils 0–2%. Variation among individuals and among different ethnic groups can be substantial, with lower leukocyte numbers for certain African-American ethnic groups. Lower granulocyte numbers in African Americans are often in the 1500–2000/μL range and are generally without sequelae, a condition termed benign ethnic neutropenia. The lower number of granulocytes is associated with null expression of the Duffy antigen receptor for cytokines (*DARC*) gene, a receptor for *Plasmodium vivax*, the absence of which conveys resistance to this form of malaria. The various leukocytes are derived from a common stem cell in the bone marrow. Three-fourths of the nucleated cells of bone marrow are committed to the production of leukocytes. Leukocyte maturation in the marrow is under the regulatory control of a number of different factors, known as colony-stimulating factors (CSFs) and interleukins (ILs). Because an alteration in the number and type of leukocytes is often associated with disease processes, total white blood cell (WBC) count (cells per μL) and differential counts are informative. This chapter focuses on neutrophils, monocytes, and eosinophils. **Lymphocytes and basophils are discussed in Chaps. 360 and 364, respectively.**

NEUTROPHILS

MATURATION

Important events in neutrophil life are summarized in **Fig. 67-1**. In normal humans, neutrophils are produced only in the bone marrow. The minimum number of stem cells necessary to support hematopoiesis is estimated to be 400–500 at any one time. Human blood monocytes, tissue macrophages, and stromal cells produce CSFs, hormones required for the growth of monocytes and neutrophils in the bone marrow. The hematopoietic system not only produces enough neutrophils (~1.3 × 10^{11} cells per 80-kg person per day) to carry out physiologic functions but also has a large reserve stored in the marrow, which can be mobilized in response to inflammation or infection. An increase in the number of blood neutrophils is called *neutrophilia*, and the presence of immature cells is termed a *shift to the left*. A decrease in the number of blood neutrophils is called *neutropenia*.

Neutrophils and monocytes evolve from pluripotent stem cells under the influence of cytokines and CSFs **(Fig. 67-2)**. The proliferation phase through the metamyelocyte takes about 1 week, while the maturation phase from metamyelocyte to mature neutrophil takes another week. The myeloblast is the first recognizable precursor cell and is followed by the *promyelocyte*. The promyelocyte evolves when the classic lysosomal granules, called the *primary*, or *azurophil, granules* are produced. The primary granules contain hydrolases, elastase, myeloperoxidase, cathepsin G, cationic proteins, and bactericidal/permeability-increasing protein, which is important for killing gram-negative bacteria. Azurophil granules also contain *defensins*, a family of cysteine-rich polypeptides with broad antimicrobial activity against bacteria, fungi, and certain enveloped viruses. The promyelocyte divides to produce the *myelocyte*, a cell responsible for the synthesis of the *specific*, or *secondary, granules*, which contain unique (specific) constituents such as lactoferrin, vitamin B_{12}–binding protein, membrane components of the reduced nicotinamide-adenine dinucleotide phosphate (NADPH) oxidase (NOX2) required for hydrogen peroxide production, histaminase, and receptors for certain chemoattractants and adherence-promoting factors (e.g., CR3) as well as receptors for the basement membrane component, laminin. The secondary granules do not contain acid hydrolases and therefore are not classic lysosomes. Packaging of secondary granule contents during myelopoiesis is controlled by CCAAT/enhancer binding protein-ε (encoded by *CEBPE*). Secondary granule contents are readily released extracellularly, and their mobilization is important in modulating inflammation. During the final stages of maturation, no cell division occurs, and the cell passes through the metamyelocyte stage and then to the band neutrophil with a sausage-shaped nucleus **(Fig. 67-3)**. As the band cell matures, the nucleus assumes a lobulated configuration. The nucleus of neutrophils normally contains up to four segments **(Fig. 67-4)**. Excessive segmentation (>5 nuclear lobes) may be a manifestation of folate or vitamin B_{12} deficiency or the congenital neutropenia syndrome of warts, hypogammaglobulinemia, infections, and myelokathexis (WHIM) described below. The Pelger-Hüet anomaly **(Fig. 67-5)**, an infrequent dominant benign inherited trait caused by heterozygous mutations in the lamin B receptor, results in neutrophils with distinctive bilobed nuclei that must be distinguished from band forms. Acquired bilobed nuclei, pseudo-Pelger-Hüet anomaly, can occur with acute infections or in myelodysplastic syndromes. The physiologic role of the normal multilobed nucleus of neutrophils is unknown, but it may allow great deformation of neutrophils during migration into tissues at sites of inflammation and facilitate production of neutrophil nets.

In severe acute bacterial infection, prominent neutrophil cytoplasmic granules, called *toxic granulations*, are occasionally seen. Toxic granulations are immature or abnormally staining azurophil granules. Cytoplasmic inclusions, also called *Döhle bodies* (Fig. 67-3), can be seen during infection and are fragments of ribosome-rich endoplasmic reticulum. Large neutrophil vacuoles are often present in acute bacterial infection in some viral infections such as COVID-19 and probably represent pinocytosed (internalized) membrane **(Fig. 67-6)**.

Neutrophils are heterogeneous in function. Monoclonal antibodies have been developed that recognize only a subset of mature neutrophils. The meaning of neutrophil heterogeneity is not known.

The morphology of eosinophils is shown in **Fig. 67-7**.

MARROW RELEASE AND CIRCULATING COMPARTMENTS

Specific signals, including IL-1, tumor necrosis factor α (TNF-α), the CSFs, complement fragments, and chemokines, mobilize leukocytes from the bone marrow and deliver them to the blood in an unstimulated state. Under normal conditions, ~90% of the neutrophil pool is in the bone marrow, 2–3% in the circulation, and the remainder in the tissues **(Fig. 67-8)**.

The circulating pool exists in two dynamic compartments: one freely flowing and one marginated. The freely flowing pool is about one-half the neutrophils in the basal state and is composed of those cells that are in the blood and not in contact with the endothelium. Marginated leukocytes are those that are in close physical contact with the endothelium **(Fig. 67-9)**. In the pulmonary circulation, where an extensive capillary bed (~1000 capillaries per alveolus) exists, margination occurs because the capillaries are about the same size as a mature neutrophil. Therefore, neutrophil fluidity and deformability are necessary to make the transit through the pulmonary bed. Increased neutrophil rigidity and decreased deformability lead to augmented neutrophil trapping and margination in the lung. In contrast, in the systemic postcapillary venules, margination is mediated by the interaction of specific

---

*Deceased

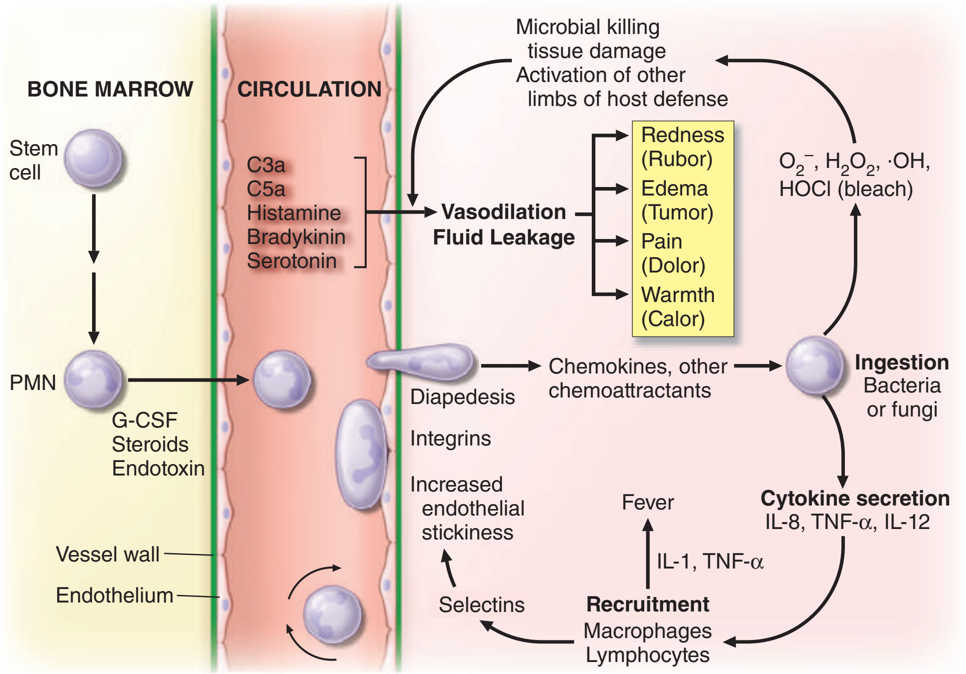
BONE MARROW
CIRCULATION
Stem cell
PMN
G-CSF
Steroids
Endotoxin
Vessel wall
Endothelium
C3a
C5a
Histamine
Bradykinin
Serotonin
Vasodilation Fluid Leakage
Redness (Rubor)
Edema (Tumor)
Pain (Dolor)
Warmth (Calor)
Microbial killing tissue damage Activation of other limbs of host defense
O_2^-, H_2O_2, ·OH, HOCl (bleach)
Diapedesis
Integrins
Increased endothelial stickiness
Selectins
Chemokines, other chemoattractants
Ingestion Bacteria or fungi
Cytokine secretion IL-8, TNF-α, IL-12
Fever
IL-1, TNF-α
Recruitment Macrophages Lymphocytes

FIGURE 67-1 Schematic events in neutrophil production, recruitment, and inflammation. The four cardinal signs of inflammation (rubor, tumor, calor, dolor) are indicated, as are the interactions of neutrophils with other cells and cytokines. G-CSF, granulocyte colony-stimulating factor; IL, interleukin; PMN, polymorphonuclear leukocyte; TNF-α, tumor necrosis factor α.

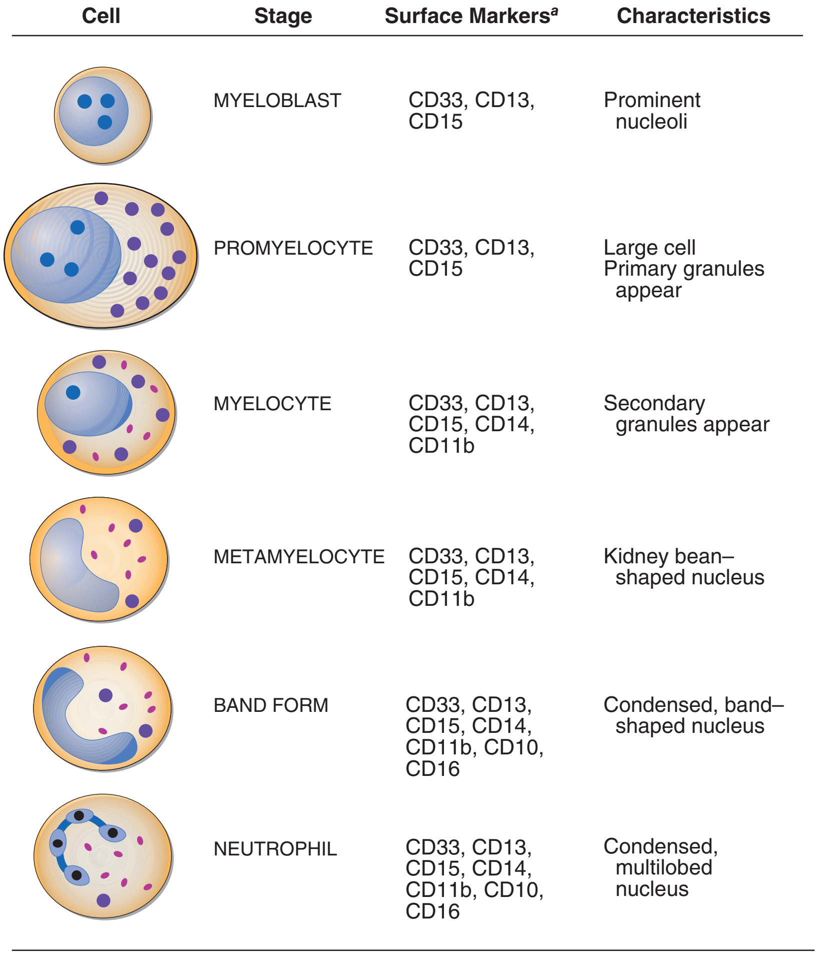

Cell	Stage	Surface Markers[a]	Characteristics
	MYELOBLAST	CD33, CD13, CD15	Prominent nucleoli
	PROMYELOCYTE	CD33, CD13, CD15	Large cell Primary granules appear
	MYELOCYTE	CD33, CD13, CD15, CD14, CD11b	Secondary granules appear
	METAMYELOCYTE	CD33, CD13, CD15, CD14, CD11b	Kidney bean–shaped nucleus
	BAND FORM	CD33, CD13, CD15, CD14, CD11b, CD10, CD16	Condensed, band–shaped nucleus
	NEUTROPHIL	CD33, CD13, CD15, CD14, CD11b, CD10, CD16	Condensed, multilobed nucleus

[a]CD = Cluster Determinant; ● Nucleolus; ● Primary granule; • Secondary granule.

FIGURE 67-2 Stages of neutrophil development shown schematically. Granulocyte colony-stimulating factor (G-CSF) and granulocyte-macrophage colony-stimulating factor (GM-CSF) are critical to this process. Identifying cellular characteristics and specific cell-surface markers are listed for each maturational stage.

cell-surface molecules called *selectins*. Selectins are glycoproteins expressed on neutrophils and endothelial cells, among others, that cause a low-affinity interaction, resulting in "rolling" of the neutrophil along the endothelial surface. On neutrophils, the molecule L-selectin (cluster determinant [CD] 62L) binds to glycosylated proteins on endothelial cells (e.g., glycosylation-dependent cell adhesion molecule [GlyCAM1] and CD34). Glycoproteins on neutrophils, most importantly sialyl-Lewis^x (SLe^x, CD15s), are targets for binding of selectins expressed on endothelial cells (E-selectin [CD62E] and P-selectin [CD62P]) and other leukocytes. In response to chemotactic stimuli from injured tissues (e.g., complement product C5a, leukotriene B$_4$, IL-8) or bacterial products (e.g., *N*-formylmethionylleucylphenylalanine [f-met-leu-phe of fMLF]), neutrophil adhesiveness increases through mobilization of intracellular adhesion proteins stored in specific granules to the cell surface, and the cells "stick" to the endothelium through *integrins*. The integrins are leukocyte glycoproteins that exist as complexes of a common CD18 β chain with CD11a (LFA-1), CD11b (called Mac-1, CR3, or the C3bi receptor), and CD11c (called p150,95 or CR4). CD11a/CD18 and CD11b/CD18 bind to specific endothelial receptors (intercellular adhesion molecules [ICAM] 1 and 2).

On cell stimulation, L-selectin is shed from neutrophils, and E-selectin increases in the blood, presumably because it is shed from endothelial cells; receptors for chemoattractants and opsonins are mobilized; and the phagocytes orient toward the chemoattractant source in the extravascular space, increase their motile activity (chemokinesis), and migrate directionally (chemotaxis) into tissues. The process of migration into tissues is called *diapedesis* and involves the crawling of neutrophils between postcapillary endothelial cells that open junctions between adjacent cells to permit leukocyte passage. Diapedesis involves platelet/endothelial cell adhesion molecule (PECAM) 1 (CD31), which is expressed on both the emigrating leukocyte and the

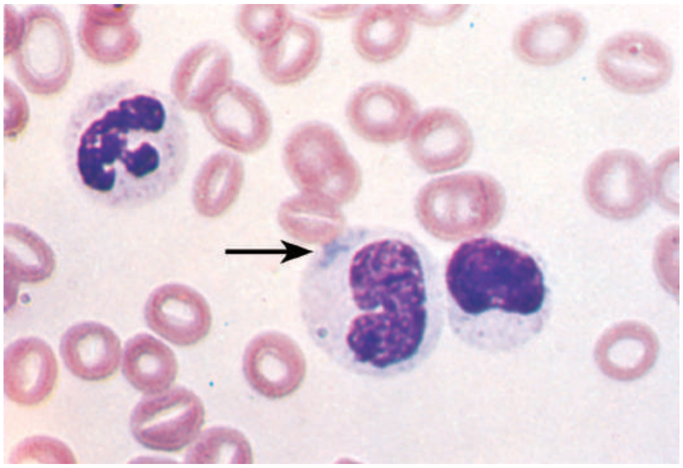

FIGURE 67-3 Neutrophil band with Döhle body. The neutrophil with a sausage-shaped nucleus in the center of the field is a band form. Döhle bodies (*arrow*) are discrete, blue-staining, nongranular areas found in the periphery of the cytoplasm of the neutrophil in infections and other toxic states. They represent aggregates of rough endoplasmic reticulum.

endothelial cells. The endothelial responses (increased blood flow from increased vasodilation and permeability) are mediated by anaphylatoxins (e.g., C3a and C5a) as well as vasodilators such as histamine, bradykinin, serotonin, nitric oxide, vascular endothelial growth factor (VEGF), and prostaglandins E and I. Cytokines regulate some of these processes (e.g., TNF-α induction of VEGF, interferon [IFN] γ inhibition of prostaglandin E).

In the healthy adult, most neutrophils leave the body by migration through the mucous membrane of the oral cavity and especially the gastrointestinal tract. Normally, neutrophils spend a short time in the circulation (half-life, 6–7 h). Senescent neutrophils are cleared from the circulation by macrophages in the lung and spleen. Once in the tissues, neutrophils release enzymes, such as collagenase and elastase, which may help establish abscess cavities. Neutrophils ingest pathogenic materials that have been opsonized by IgG and C3b. Fibronectin and the tetrapeptide tuftsin also facilitate phagocytosis.

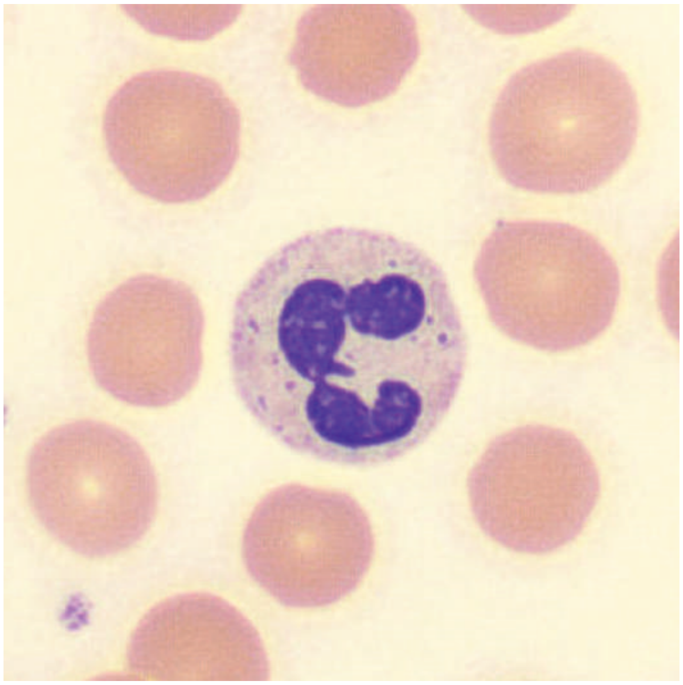

FIGURE 67-4 Normal granulocyte. The normal granulocyte has a segmented nucleus with heavy, clumped chromatin; fine neutrophilic granules are dispersed throughout the cytoplasm.

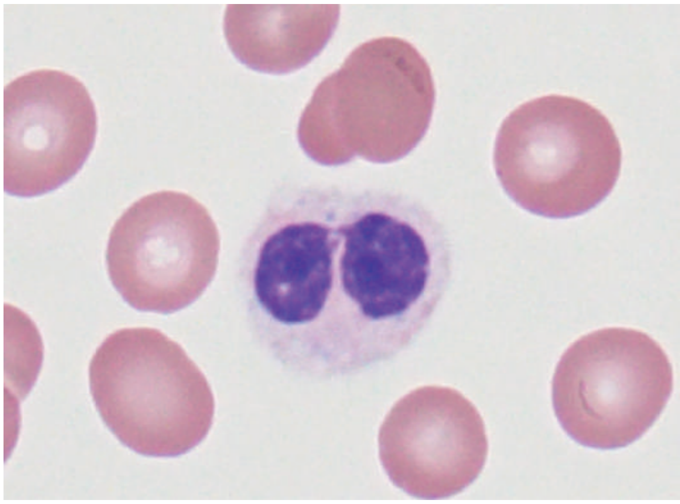

FIGURE 67-5 Pelger-Hüet anomaly. In this disorder, granulocytes are bilobed. In the benign, genetic form (due to mutations in *LBR*), most granulocytes have this appearance. In the acquired form, which is associated with myelodysplastic syndrome, it may be more occasional. The nucleus frequently has a spectacle-like, or "pince-nez," configuration. *(Courtesy of Dr. Katherine Calvo, Hematopathology Laboratory, Department of Laboratory Medicine, Clinical Center, NIH.)*

With phagocytosis comes a burst of oxygen consumption and activation of the hexose-monophosphate shunt. A membrane-associated NADPH oxidase called NOX2, consisting of membrane and cytosolic components, is assembled and catalyzes the univalent reduction of oxygen to superoxide anion, which is then converted by superoxide dismutase to hydrogen peroxide and other toxic oxygen products (e.g., hydroxyl radical). Hydrogen peroxide + chloride + neutrophil myeloperoxidase generates hypochlorous acid (bleach), hypochlorite, and chlorine. These products oxidize and halogenate microorganisms and tumor cells and, when uncontrolled, can damage host tissue. Strongly cationic proteins, defensins, elastase, cathepsins, and probably nitric oxide also participate in microbial killing. Lactoferrin chelates iron, an important growth factor for microorganisms, especially fungi. Other enzymes, such as lysozyme and acid proteases, help digest microbial debris. After 1–4 days in tissues, neutrophils die. The apoptosis of neutrophils is also cytokine-regulated; granulocyte colony-stimulating factor (G-CSF) and IFN-γ prolong their life span. Neutrophil extracellular traps (NETs) consisting of a DNA scaffold decorated with neutrophil-granule derived proteins, such as enzymatically active proteases and antimicrobial peptides, are thought to be formed as a defense mechanism to immobilize invading microorganisms. Under certain conditions, such as in delayed-type hypersensitivity, monocyte accumulation occurs within 6–12 h of initiation of inflammation. Neutrophils, monocytes, microorganisms in various states of digestion, and altered local tissue cells make up the inflammatory exudate, pus. Myeloperoxidase confers the characteristic green color to pus and may participate in turning off the inflammatory process by inactivating chemoattractants and immobilizing phagocytic cells.

Neutrophils respond to certain cytokines (IFN-γ, granulocyte-macrophage colony-stimulating factor [GM-CSF], IL-8) and produce cytokines and chemotactic signals (TNF-α, IL-8, macrophage inflammatory protein [MIP] 1) that modulate the inflammatory response. In the presence of fibrinogen, fMLF, or leukotriene B_4, IL-8 production by neutrophils is induced, providing autocrine amplification of inflammation. *Chemokines* (*chemo*attractant cyto*kines*) are small proteins produced by many different cell types, including endothelial cells, fibroblasts, epithelial cells, neutrophils, and monocytes, that regulate neutrophil, monocyte, eosinophil, and lymphocyte recruitment and activation. Chemokines transduce their signals through heterotrimeric G protein–linked receptors that have seven cell membrane–spanning domains, the same type of cell-surface receptor that mediates the response to the classic chemoattractants fMLF and C5a. Four major groups of chemokines are recognized based on the cysteine structure near the N terminus: C, CC, CXC, and CXXXC. The C chemokine

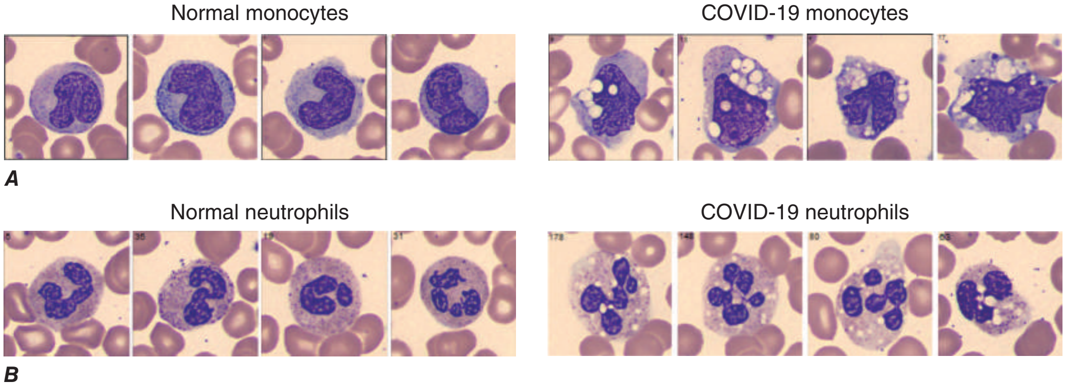
Normal monocytes
COVID-19 monocytes
A
Normal neutrophils
COVID-19 neutrophils
B

FIGURE 67-6 COVID-19: Vacuolization in peripheral blood monocytes and neutrophils of COVID-19 patients. Peripheral blood smear showing vacuolization in (***A***) monocytes and (***B***) neutrophils from hospitalized hypoxemic COVID-19 patients relative to healthy volunteers. Increased vacuoles were noted in ~80% of monocytes and ~50% of neutrophils in each COVID-19 patient throughout their hospitalization.

lymphotactin is T-cell tropic; CC chemokines such as MIP-1 attract lymphocytes, monocytes, eosinophils, and basophils; CXC cytokines such as IL-8 mainly attract neutrophils; while the CXXXC chemokine fractalkine attracts neutrophils, monocytes, and T cells. Not only do these molecules and their receptors regulate the trafficking and activation of inflammatory cells, but also specific chemokine receptors serve as co-receptors for HIV infection **(Chap. 208)**, while others have roles in other viral infections (e.g., West Nile virus), susceptibility and response to *Candida*, and atherogenesis.

NEUTROPHIL ABNORMALITIES

Defects in the neutrophil life cycle can lead to dysfunction and compromised host defenses. When inflammation is severely depressed, the clinical result is often recurrent, severe bacterial and fungal infections. Aphthous ulcers of mucous membranes (gray ulcers without pus) and gingivitis and periodontal disease suggest a phagocytic cell disorder. Patients with congenital phagocyte defects can have infections within the first few days of life. Skin, ear, upper and lower respiratory tract, and bone infections are common. Sepsis and meningitis are rare. In some disorders, the frequency of infection is variable, and patients can go for months or even years without major infection. Aggressive management of these congenital diseases, including hematopoietic stem cell transplantation and gene therapy, has extended the life span of patients well into adulthood.

Neutropenia The consequences of absent neutrophils are dramatic. Susceptibility to infectious diseases increases sharply when neutrophil counts fall to <1000 cells/μL. When the absolute neutrophil count (ANC; band forms and mature neutrophils combined) falls to <500 cells/μL, control of endogenous microbial flora (e.g., mouth, gut) is impaired; when the ANC is <200/μL, the local inflammatory process is absent. Neutropenia can be due to depressed production, increased peripheral destruction, or excessive peripheral pooling. A falling neutrophil count or a significant decrease in the number of neutrophils below steady-state levels, together with a failure to increase neutrophil counts in the setting of infection or other challenge, requires investigation. Acute neutropenia, such as that caused by cancer chemotherapy, is more likely to be associated with increased risk of infection than chronic neutropenia (months to years) that reverses in response to infection or carefully controlled administration of endotoxin (see "Laboratory Diagnosis and Management," below).

Some causes of inherited and acquired neutropenia are listed in **Table 67-1**. The most common neutropenias are iatrogenic, resulting from the use of cytotoxic or immunosuppressive therapies for malignancy or control of autoimmune disorders. These drugs cause

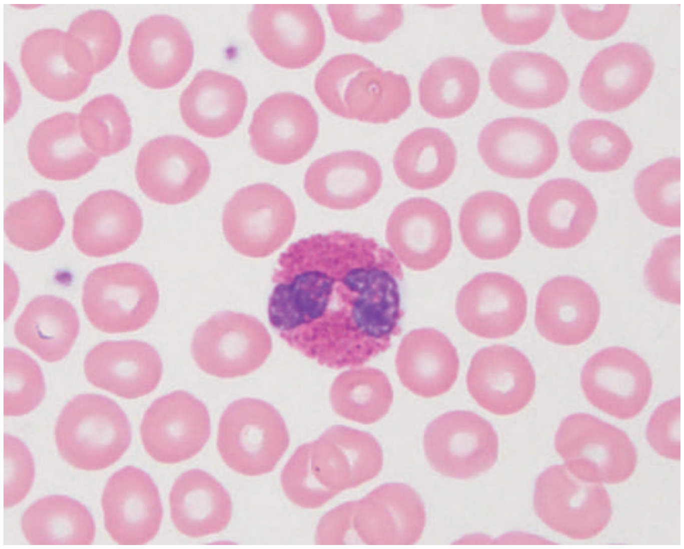

FIGURE 67-7 The normal eosinophil seen here contains large, bright orange granules and usually a bilobed nucleus. *(Courtesy of Dr. Katherine Calvo, Hematopathology Laboratory, Department of Laboratory Medicine, Clinical Center, NIH.)*

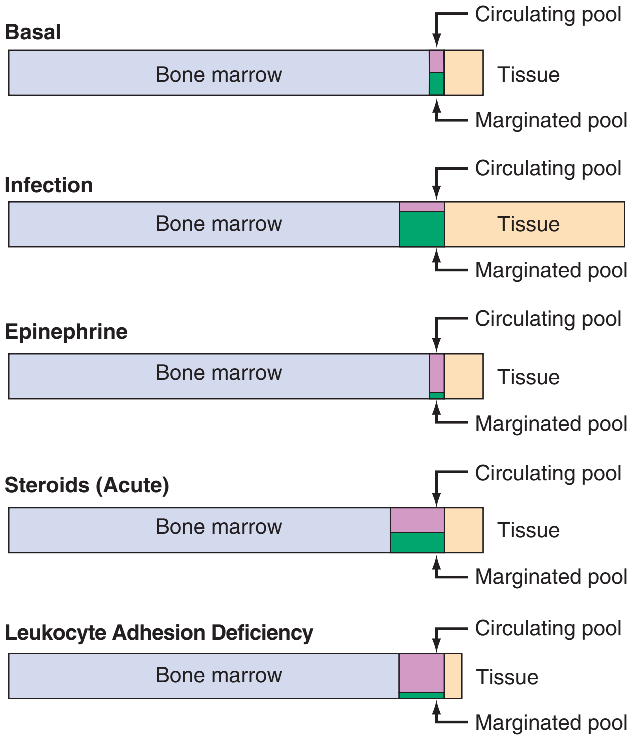
Basal
Circulating pool
Bone marrow
Tissue
Marginated pool
Infection
Circulating pool
Bone marrow
Tissue
Marginated pool
Epinephrine
Circulating pool
Bone marrow
Tissue
Marginated pool
Steroids (Acute)
Circulating pool
Bone marrow
Tissue
Marginated pool
Leukocyte Adhesion Deficiency
Circulating pool
Bone marrow
Tissue
Marginated pool

FIGURE 67-8 Schematic neutrophil distribution and kinetics between the different anatomic and functional pools.

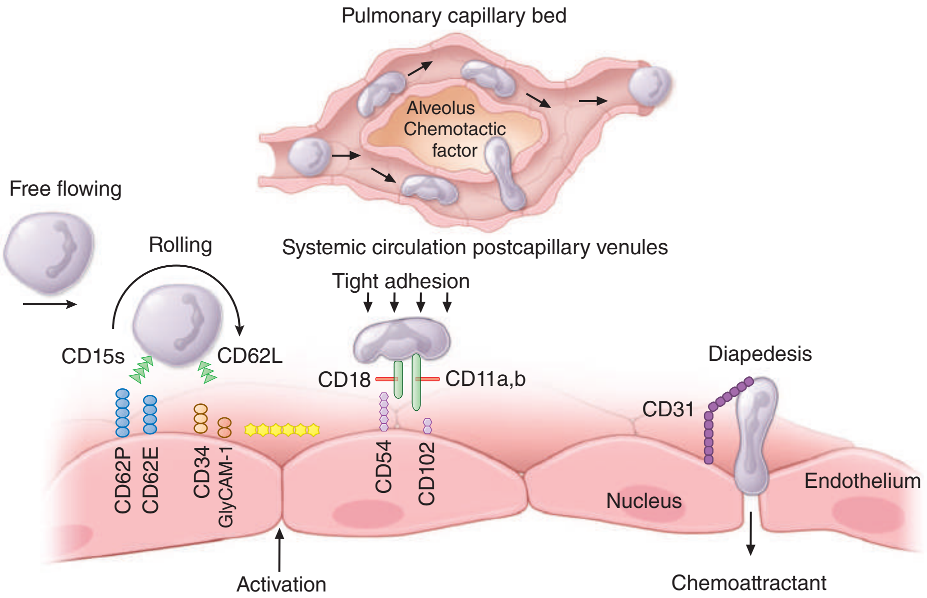

FIGURE 67-9 Neutrophil travel through the pulmonary capillaries is dependent on neutrophil deformability. Neutrophil rigidity (e.g., caused by C5a) enhances pulmonary trapping and response to pulmonary pathogens in a way that is not so dependent on cell-surface receptors. Intraalveolar chemotactic factors, such as those caused by certain bacteria (e.g., *Streptococcus pneumoniae*), lead to diapedesis of neutrophils from the pulmonary capillaries into the alveolar space. Neutrophil interaction with the endothelium of the systemic postcapillary venules is dependent on molecules of attachment. The neutrophil "rolls" along the endothelium using selectins: neutrophil CD15s (sialyl-Lewis^x) binds to CD62E (E-selectin) and CD62P (P-selectin) on endothelial cells; CD62L (L-selectin) on neutrophils binds to CD34 and other molecules (e.g., GlyCAM-1) expressed on endothelium. Chemokines or other activation factors stimulate integrin-mediated "tight adhesion": CD11a/CD18 (LFA-1) and CD11b/CD18 (Mac-1, CR3) bind to CD54 (ICAM-1) and CD102 (ICAM-2) on the endothelium. Diapedesis occurs between endothelial cells: CD31 (PECAM-1) expressed by the emigrating neutrophil interacts with CD31 expressed at the endothelial cell-cell junction. CD, cluster determinant; GlyCAM, glycosylation-dependent cell adhesion molecule; ICAM, intercellular adhesion molecule; PECAM, platelet/endothelial cell adhesion molecule.

neutropenia because they result in decreased production of rapidly growing progenitor (stem) cells of the marrow. Certain antibiotics such as chloramphenicol, trimethoprim-sulfamethoxazole, flucytosine, vidarabine, and the antiretroviral drug zidovudine may cause neutropenia by inhibiting proliferation of myeloid precursors. Azathioprine and 6-mercaptopurine are metabolized by the enzyme thiopurine methyltransferase (TMPT); hypofunctional polymorphisms that are found in 11% of whites can lead to accumulation of 6-thioguanine and profound marrow toxicity, which is why pharmacogenomic testing is recommended before initiating these drugs. The marrow suppression is generally dose-related and resolves after cessation of the drug.

Another important mechanism for iatrogenic neutropenia is the effect of drugs that serve as immune haptens and sensitize neutrophils or neutrophil precursors to immune-mediated peripheral destruction. This form of drug-induced neutropenia can be seen within 7 days of exposure to the drug; with previous drug exposure, resulting in preexisting antibodies, neutropenia may occur a few hours after administration of the drug. Although any drug can cause this form of neutropenia, the most frequent causes are commonly used antibiotics, such as sulfa-containing compounds, penicillins, and cephalosporins. Fever and eosinophilia may also be associated with drug reactions, but often these signs are not present. Drug-induced neutropenia can be severe, but discontinuation of the sensitizing drug is sufficient for recovery, which is usually seen within 5–7 days and is complete by 10 days. Readministration of the sensitizing drug should be avoided, because abrupt neutropenia will often result. For this reason, diagnostic challenge should be avoided.

TABLE 67-1 Causes of Neutropenia

Decreased Production
Drug-induced—alkylating agents (nitrogen mustard, busulfan, chlorambucil, cyclophosphamide); antimetabolites (methotrexate, 6-mercaptopurine, 5-flucytosine); noncytotoxic agents (antibiotics [chloramphenicol, penicillins, sulfonamides], phenothiazines, tranquilizers [meprobamate], anticonvulsants [carbamazepine], antipsychotics [clozapine], certain diuretics, anti-inflammatory agents, antithyroid drugs, many others)
Hematologic diseases—idiopathic, cyclic neutropenia, Chédiak-Higashi syndrome, aplastic anemia, infantile genetic disorders (see text)
Tumor invasion, myelofibrosis
Nutritional deficiency—vitamin B_{12}, folate (especially alcoholics)
Infection—tuberculosis, typhoid fever, brucellosis, tularemia, measles, infectious mononucleosis, malaria, viral hepatitis, leishmaniasis, AIDS
Peripheral Destruction
Antineutrophil antibodies and/or splenic or lung trapping
Autoimmune disorders—Felty syndrome, rheumatoid arthritis, lupus erythematosus
Drugs as haptens—aminopyrine, α-methyldopa, phenylbutazone, mercurial diuretics, some phenothiazines
Granulomatosis with polyangiitis (Wegener)
Peripheral Pooling (Transient Neutropenia)
Overwhelming bacterial infection (acute endotoxemia)
Hemodialysis
Cardiopulmonary bypass

Autoimmune neutropenias caused by circulating antineutrophil antibodies are another form of acquired neutropenia that results in increased destruction of neutrophils. Acquired neutropenia may also be seen with viral infections, including acute infection with HIV. Acquired neutropenia may be cyclic in nature, occurring at intervals of several weeks. Acquired cyclic or stable neutropenia may be associated with an expansion of large granular lymphocytes (LGLs), which may be T cells, NK cells, or NK-like cells. Patients with large granular lymphocytosis may have moderate blood and bone marrow lymphocytosis, neutropenia, polyclonal hypergammaglobulinemia, splenomegaly, rheumatoid arthritis, and absence of lymphadenopathy. Such patients may have a chronic and relatively stable course. Recurrent bacterial infections are frequent. Benign and malignant forms of this syndrome occur. In some patients, spontaneous regression has occurred even after 11 years, suggesting an immunoregulatory defect as the basis for at least one form of the disorder. Glucocorticoids, cyclosporine, methotrexate, and monoclonals are commonly used to manage these cytopenias.

Hereditary Neutropenias Hereditary neutropenias are rare and may manifest in early childhood as a profound constant neutropenia or agranulocytosis. Congenital forms of neutropenia include Kostmann's syndrome (neutrophil count <100/μL), which is often fatal and due to mutations in the anti-apoptosis gene *HAX-1*; severe chronic neutropenia (neutrophil count of 300–1500/μL) due to mutations in neutrophil elastase (*ELANE*); hereditary cyclic neutropenia or, more appropriately, cyclic hematopoiesis, also due to mutations in neutrophil elastase (*ELANE*); the cartilage-hair hypoplasia syndrome due to mutations in the mitochondrial RNA-processing endoribonuclease *RMRP*; Shwachman-Diamond syndrome associated with pancreatic insufficiency due to mutations in the Shwachman-Bodian-Diamond syndrome gene *SBDS*; the WHIM syndrome (*w*arts, *h*ypogammaglobulinemia, *i*nfections, *m*yelokathexis [retention of WBCs in the marrow]), characterized by neutrophil hypersegmentation and bone marrow myeloid arrest due to mutations in the chemokine receptor *CXCR4*; and neutropenias associated with other immune defects, such as GATA2 deficiency, X-linked agammaglobulinemia, Wiskott-Aldrich syndrome, and CD40 ligand deficiency. Mutations in the G-CSF receptor can develop in severe congenital neutropenia and are linked to the development of leukemia. Absence of both myeloid and lymphoid cells is seen in reticular dysgenesis, due to mutations in the nuclear genome-encoded mitochondrial enzyme adenylate kinase-2 (*AK2*).

Maternal factors can be associated with neutropenia in the newborn. Transplacental transfer of IgG directed against antigens on fetal neutrophils can result in peripheral destruction. Drugs (e.g., thiazides) ingested during pregnancy can cause neutropenia in the newborn by either depressed production or peripheral destruction.

In Felty syndrome—the triad of rheumatoid arthritis, splenomegaly, and neutropenia **(Chap. 370)**—antibodies can shorten neutrophil life span, while large granular lymphocytes can attack marrow neutrophil precursors. Splenectomy may increase the neutrophil count in Felty syndrome and lower serum neutrophil-binding IgG. Some Felty syndrome patients also have autoantibodies to G-CSF, while others have increased numbers of LGLs. Splenomegaly with peripheral trapping and destruction of neutrophils is also seen in lysosomal storage diseases and commonly in portal hypertension.

Neutrophilia Neutrophilia results from increased neutrophil production, increased marrow release, or defective margination **(Table 67-2)**. The most important acute cause of neutrophilia is infection. Neutrophilia from acute infection represents both increased production and increased marrow release. Increased production is also associated with chronic inflammation and certain myeloproliferative diseases. Increased marrow release and mobilization of the marginated leukocyte pool are induced by glucocorticoids. Release of epinephrine, as with vigorous exercise, excitement, or stress, will demarginate neutrophils in the spleen and lungs and double the neutrophil count in minutes. Cigarette smoking can elevate neutrophil counts above the normal range. Leukocytosis with cell counts of 10,000–25,000/μL occurs in response to infection and other forms of acute inflammation and results from both release of the marginated pool and mobilization of marrow reserves. Persistent neutrophilia with cell counts of ≥30,000–50,000/μL is called a *leukemoid reaction*, a term often used to distinguish this degree of neutrophilia from leukemia. In a leukemoid reaction, the circulating neutrophils are usually mature and not clonally derived.

Abnormal Neutrophil Function Inherited and acquired abnormalities of phagocyte function are listed in **Table 67-3**. The resulting diseases are best considered in terms of the functional defects of adherence, chemotaxis, and microbicidal activity. The distinguishing features of the important inherited disorders of phagocyte function are shown in **Table 67-4**.

TABLE 67-2 Causes of Neutrophilia

Increased Production
Idiopathic
Drug-induced—glucocorticoids, G-CSF
Infection—bacterial, fungal, sometimes viral
Inflammation—thermal injury, tissue necrosis, myocardial and pulmonary infarction, hypersensitivity states, collagen vascular diseases
Myeloproliferative diseases—myelocytic leukemia, myeloid metaplasia, polycythemia vera
Increased Marrow Release
Glucocorticoids
Acute infection (endotoxin)
Inflammation—thermal injury
Decreased or Defective Margination
Drugs—epinephrine, glucocorticoids, nonsteroidal anti-inflammatory agents
Stress, excitement, vigorous exercise
Leukocyte adhesion deficiency type 1 (CD18); leukocyte adhesion deficiency type 2 (selectin ligand, CD15s); leukocyte adhesion deficiency type 3 (FERMT3)
Miscellaneous
Metabolic disorders—ketoacidosis, acute renal failure, eclampsia, acute poisoning
Drugs—lithium
Other—metastatic carcinoma, acute hemorrhage or hemolysis

Abbreviation: G-CSF, granulocyte colony-stimulating factor.

DISORDERS OF ADHESION Three main types of leukocyte adhesion deficiency (LAD) have been described. All are autosomal recessive and result in the impairment of neutrophil exit from the circulation to sites of infection, leading to leukocytosis and increased susceptibility to infection (Fig. 67-9). Patients with LAD 1 have mutations in *CD18*, the common component of the integrins LFA-1, Mac-1, and p150,95, leading to a defect in tight adhesion between neutrophils and the endothelium. The heterodimer formed by CD18/CD11b (Mac-1) is also the receptor for the complement-derived opsonin C3bi (CR3). The *CD18* gene is located on distal chromosome 21q. The severity of the defect determines the severity of clinical disease. Complete lack of expression of the leukocyte integrins results in a severe phenotype in which inflammatory stimuli do not increase the expression of leukocyte integrins on neutrophils or activated T and B cells. Neutrophils (and monocytes) from patients with LAD 1 adhere poorly to endothelial cells and protein-coated surfaces and exhibit defective spreading, aggregation, and chemotaxis. The inability of neutrophils to exit the vasculature to the tissue deprives the tissue macrophage of its expected neutrophil ingestion, leading to macrophage production of IL-23,

TABLE 67-3 Types of Granulocyte and Monocyte Disorders

	CAUSE OF INDICATED DYSFUNCTION		
FUNCTION	**DRUG-INDUCED**	**ACQUIRED**	**INHERITED**
Adherence-aggregation	Aspirin, colchicine, alcohol, glucocorticoids, ibuprofen, piroxicam	Neonatal state, hemodialysis	Leukocyte adhesion deficiency types 1, 2, and 3
Deformability		Leukemia, neonatal state, diabetes mellitus, immature neutrophils	
Chemokinesis-chemotaxis	Glucocorticoids (high dose), auranofin, colchicine (weak effect), phenylbutazone, naproxen, indomethacin, interleukin 2	Thermal injury, malignancy, malnutrition, periodontal disease, neonatal state, systemic lupus erythematosus, rheumatoid arthritis, diabetes mellitus, sepsis, influenza virus infection, herpes simplex virus infection, acrodermatitis enteropathica, AIDS	Chédiak-Higashi syndrome, neutrophil-specific granule deficiency, WDR1 deficiency, Job's syndrome (in some patients), Down syndrome, α-mannosidase deficiency, leukocyte adhesion deficiencies, Wiskott-Aldrich syndrome
Microbicidal activity	Colchicine, cyclophosphamide, glucocorticoids (high dose), TNF-α-blocking antibodies	Leukemia, aplastic anemia, certain neutropenias, tuftsin deficiency, thermal injury, sepsis, neonatal state, diabetes mellitus, malnutrition, AIDS	Chédiak-Higashi syndrome, neutrophil-specific granule deficiency, chronic granulomatous disease, defects in IFNγ/IL-12 axis, Anti-IFNγ autoantibodies

Abbreviations: IFNγ, interferon γ; IL, interleukin; TNF-α, tumor necrosis factor alpha.

TABLE 67-4 Inherited Disorders of Phagocyte Function: Differential Features

CLINICAL MANIFESTATIONS	CELLULAR OR MOLECULAR DEFECTS	DIAGNOSIS
Chronic Granulomatous Diseases (70% X-Linked, 30% Autosomal Recessive)		
Severe infections of skin, ears, lungs, liver, and bone with microorganisms such as *Staphylococcus aureus, Burkholderia cepacia* complex, *Aspergillus* spp., *Chromobacterium violaceum*; often hard to culture organism; excessive inflammation with granulomas, frequent lymph node suppuration; granulomas can obstruct GI or GU tracts; gingivitis, aphthous ulcers	No respiratory burst due to impaired of NADPH oxidase in neutrophils, monocytes, and eosinophils. Mutations in *CYBB, CYBA, NCF1, NCF2, NCF4,* or *CYBC1*	DHR or NBT test; no superoxide and H_2O_2 production by neutrophils; immunoblot for NADPH oxidase components; genetic detection
Chédiak-Higashi Syndrome (Autosomal Recessive)		
Recurrent pyogenic infections, especially with *S. aureus*; many patients get lymphoma-like illness during adolescence; periodontal disease; partial oculocutaneous albinism, nystagmus, progressive peripheral neuropathy, cognitive impairment in some patients	Reduced chemotaxis and phagolysosome fusion, increased respiratory burst activity, defective egress from marrow, abnormal skin window; defect in *CHS1*	Giant primary granules in neutrophils and other granule-bearing cells (Wright stain); genetic detection
Specific Granule Deficiency (Autosomal Recessive and Dominant)		
Recurrent infections of skin, ears, and sinopulmonary tract; delayed wound healing; decreased inflammation; bleeding diathesis	Abnormal chemotaxis, impaired respiratory burst and bacterial killing, failure to upregulate chemotactic and adhesion receptors with stimulation, defect in transcription of granule proteins. Mutations in *CEBPE* or *SMARCD2*	Lack of secondary (specific) granules in neutrophils (Wright stain), no neutrophil-specific granule contents (i.e., lactoferrin), no defensins, platelet α granule abnormality; genetic detection
Myeloperoxidase Deficiency (Autosomal Recessive)		
Clinically normal except in patients with underlying disease such as diabetes mellitus; then candidiasis or other fungal infections	No myeloperoxidase due to pre- and posttranslational defects in myeloperoxidase deficiency. Mutations in *MPO*	No peroxidase in neutrophils; genetic detection
Leukocyte Adhesion Deficiency		
Type 1: Delayed separation of umbilical cord, sustained neutrophilia, recurrent infections of skin and mucosa, gingivitis, periodontal disease	Impaired phagocyte adherence, aggregation, spreading, chemotaxis, phagocytosis of C3bi-coated particles; defective production of CD18 subunit common to leukocyte integrins. Mutations in *ITGB2*	Reduced phagocyte surface expression of the CD18-containing integrins with monoclonal antibodies against LFA-1 (CD18/CD11a), Mac-1 or CR3 (CD18/CD11b), p150,95 (CD18/CD11c); genetic detection
Type 2: Cognitive impairment, short stature, Bombay (hh) blood phenotype, recurrent infections, neutrophilia	Impaired phagocyte rolling along endothelium; due to defects in fucose transporter. Mutations in *SLC35C1*	Reduced phagocyte surface expression of Sialyl-Lewis^x, with monoclonal antibodies against CD15s; genetic detection
Type 3: Petechial hemorrhage, recurrent infections	Impaired signaling for integrin activation resulting in impaired adhesion. Mutations in *FERMT3*	Reduced signaling for adhesion through integrins; genetic detection
Phagocyte Activation Defects (X-Linked and Autosomal Recessive)		
NEMO deficiency: mild hypohidrotic ectodermal dysplasia; broad-based immune defect: pyogenic and encapsulated bacteria, viruses, *Pneumocystis*, mycobacteria; X-linked	Impaired phagocyte activation by IL-1, IL-18, TLR, CD40L, TNF-α leading to problems with inflammation and antibody production. Mutations in *IKBKG*	Poor in vitro response to endotoxin; impaired NF-κB activation; genetic detection
IRAK4 and MyD88 deficiency: susceptibility to pyogenic bacteria such as staphylococci, streptococci, clostridia; resistant to *Candida*; autosomal recessive	Impaired phagocyte activation by endotoxin through TLR and other pathways; TNF-α signaling preserved. Mutations in *IRAK4* or *MYD88*	Poor in vitro response to endotoxin; lack of NF-κB activation by endotoxin; genetic detection
Hyper IgE–Recurrent Infection Syndrome (Autosomal Dominant) (Job's Syndrome)		
Eczematoid or pruritic dermatitis, "cold" skin abscesses, recurrent pneumonias with *S. aureus* with bronchopleural fistulae and cyst formation, mild eosinophilia, mucocutaneous candidiasis, characteristic facies, restrictive lung disease, scoliosis, delayed primary dental deciduation	Reduced chemotaxis in some patients, reduced memory T and B cells. Mutations in *STAT3*	Somatic and immune features involving lungs, skeleton, and immune system; serum IgE >2000 IU/mL; genetic testing
DOCK8 deficiency (autosomal recessive), severe eczema, atopic dermatitis, cutaneous abscesses, HSV, HPV, and molluscum infections, severe allergies, cancer	Impaired T-cell proliferation to mitogens. Mutations in *DOCK8*	Severe allergies, viral infections, high IgE, eosinophilia, low IgM, progressive lymphopenia, genetic detection
Mycobacterial Susceptibility (Autosomal Dominant and Recessive Forms)		
Severe extrapulmonary or disseminated infections with bacille Calmette-Guérin (BCG), nontuberculous mycobacteria, salmonella, histoplasmosis, coccidioidomycosis, poor granuloma formation	Inability to kill intracellular organisms due to low IFN-γ production or response. Mutations in IFN-γ receptors, IL-12 receptors, IL-12 p40, *STAT1, NEMO, ISG15, and others*	Abnormally low or very high levels of IFN-γ receptor 1; functional assays of cytokine production and response; genetic detection
GATA2 Deficiency (Autosomal Dominant)		
Persistent or disseminated warts, disseminated mycobacterial disease, low monocytes, NK cells, B cells; hypoplastic myelodysplasia, leukemia, cytogenetic abnormalities, pulmonary alveolar proteinosis	Impaired macrophage activity, cytopenias. Mutations in *GATA2*	Profound circulating monocytopenia, NK and B-cell cytopenias; genetic detection

Abbreviations: C/EBPε, CCAAT/enhancer binding protein-ε; DHR, dihydrorhodamine (oxidation test); DOCK8, dedicator of cytokinesis 8; GI, gastrointestinal; GU, genitourinary; HPV, human papillomavirus; HSV, herpes simplex virus; IFN, interferon; IL, interleukin; IRAK4, IL-1 receptor–associated kinase 4; LFA-1, leukocyte function–associated antigen 1; MyD88, myeloid differentiation primary response gene 88; NADPH, nicotinamide–adenine dinucleotide phosphate; NBT, nitroblue tetrazolium (dye test); NEMO, NF-κB essential modulator; NF-κB, nuclear factor-κB; NK, natural killer; STAT1–3, signal transducer and activator of transcription 1–3; TLR, Toll-like receptor; TNF, tumor necrosis factor.

which induces T-cell production of IL-17, a potent proinflammatory cytokine. These processes conspire to drive inflammation in LAD 1. Patients with LAD 1 have recurrent bacterial infections involving the skin, oral and genital mucosa, and respiratory and intestinal tracts; persistent leukocytosis (resting neutrophil counts of 15,000–20,000/μL) because cells do not marginate; and, in severe cases, a history of delayed separation of the umbilical stump. Infections, especially of the skin, may become necrotic with progressively enlarging borders, slow healing, and development of dysplastic scars. The most common bacteria are *Staphylococcus aureus* and enteric gram-negative bacteria. LAD 2 is caused by an abnormality of fucosylation of SLe^x (CD15s), the ligand on neutrophils that interacts with selectins on endothelial cells and is responsible for neutrophil rolling along the endothelium. Infection susceptibility in LAD 2 appears to be less severe than in LAD 1. LAD 2 is also known as *congenital disorder of glycosylation IIc* (CDGIIc) due to mutation in a GDP-fucose transporter (*SLC35C1*). LAD 3 is characterized by infection susceptibility, leukocytosis, and petechial hemorrhage due to impaired integrin activation caused by mutations in *FERMT3*.

DISORDERS OF NEUTROPHIL GRANULES The most common neutrophil defect is myeloperoxidase deficiency, a primary granule defect inherited as an autosomal recessive trait; the incidence is ~1 in 2000 persons. Myeloperoxidase deficiency is not associated with clinically compromised defenses unless there is another contributing factor such as diabetes. In myeloperoxidase deficiency, neutrophil microbicidal activity is delayed but not absent. Patients with myeloperoxidase deficiency and diabetes are more susceptible to *Candida* infections. An acquired form of myeloperoxidase deficiency occurs in myelomonocytic leukemia and acute myeloid leukemia. Importantly, myeloperoxidase deficiency gives an abnormal dihydrorhodamine (DHR) assay (see chronic granulomatous disease below).

Chédiak-Higashi syndrome (CHS) is a rare disease with autosomal recessive inheritance due to defects in the lysosomal transport protein LYST, encoded by the gene *CHS1* at 1q42. This protein is required for normal packaging and disbursement of granules. Neutrophils (and all cells containing lysosomes) from patients with CHS characteristically have large granules **(Fig. 67-10)**, making it a systemic disease. Patients with CHS have nystagmus, partial oculocutaneous albinism, and an increased number of infections resulting from many bacterial agents. Some CHS patients develop an "accelerated phase" in childhood with a hemophagocytic syndrome and an aggressive lymphoma requiring bone marrow transplantation. CHS neutrophils and monocytes have impaired chemotaxis and abnormal rates of microbial killing due to slow rates of fusion of the lysosomal granules with phagosomes. NK cell function is also impaired. CHS patients may develop a severe disabling peripheral neuropathy in adulthood.

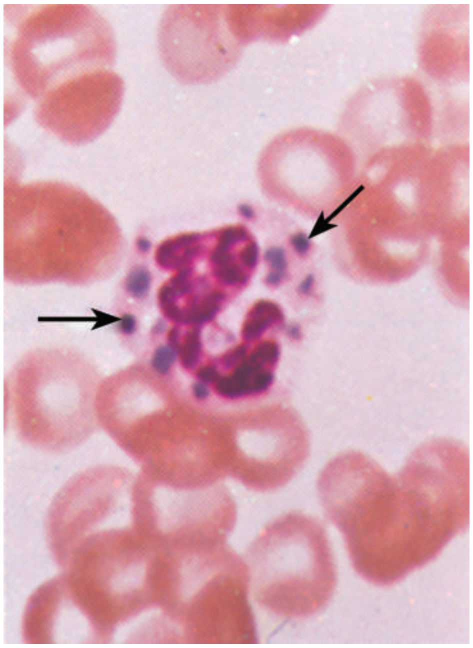

FIGURE 67-10 Chédiak-Higashi syndrome. The granulocytes contain huge cytoplasmic granules formed from aggregation and fusion of azurophilic and specific granules. Large abnormal granules are found in other granule-containing cells throughout the body.

Specific granule deficiency is a rare autosomal recessive disease in which the production of secondary granules and their contents, as well as the primary granule component defensins, is defective. The defect in killing leads to severe bacterial infections. One type of specific granule deficiency is due to a mutation in *CEBPE*, which regulates expression of granule components. A dominant mutation in *CEBPE* has also been described. Specific granule deficiency can also be caused by mutations in *SMARCD2*.

CHRONIC GRANULOMATOUS DISEASE Chronic granulomatous disease (CGD) is a group of genetic disorders of granulocyte and monocyte oxidative metabolism due to defects in the enzyme NADPH oxidase, also called NOX2. CGD has an incidence of ~1 in 100,000–200,000 individuals. In about two-thirds of patients, CGD is inherited as an X-linked recessive trait; the remainder inherit their disease in autosomal recessive patterns. Mutations in the genes encoding the six proteins that allow assembly at the plasma membrane of NOX2 account for all patients with CGD. Two proteins (a 91-kDa protein, abnormal in X-linked CGD, and a 22-kDa protein, absent in one form of autosomal recessive CGD) form the heterodimer cytochrome b-558 in the plasma membrane. The protein *e*ssential for *r*eactive *o*xidant *s*ignaling (EROS) is encoded by *CYBC1*, which is required to transport the 91- and 22-kDa proteins to the endoplasmic reticulum. Three other proteins (40, 47, and 67 kDa, abnormal in the other autosomal recessive forms of CGD) are cytoplasmic and interact with the cytochrome after cell activation to form the NADPH oxidase, which is required for phagocyte hydrogen peroxide production. Therefore, leukocytes from patients with CGD have severely diminished hydrogen peroxide production. Patients with CGD characteristically have increased numbers of infections due to a relatively narrow range of microorganisms (in North America, *S. aureus*, *Serratia marsescens*, *Burkholderia cepacia* complex, and *Nocardia* and *Aspergillus* species). Outside of North America, *Salmonella*, tuberculosis, and bacillus Calmette-Guérin (BCG) are important pathogens. When patients with CGD become infected, they often have extensive inflammatory reactions, and suppuration is common despite the administration of appropriate antibiotics. Aphthous ulcers and chronic infection of the nares are often present. Granulomas are frequent and can obstruct the gastrointestinal or genitourinary tracts. The excessive inflammation is due to failure to downregulate inflammation, reflecting a failure to inhibit the synthesis of, degradation of, or response to interleukins or chemoattractants, leading to persistent myeloid reaction. Impaired killing of intracellular microorganisms by macrophages may lead to persistent cell-mediated immune activation and granuloma formation. Autoimmune complications such as immune thrombocytopenic purpura and juvenile idiopathic arthritis are also increased in CGD. In addition, for unexplained reasons, discoid lupus is more common in X-linked carriers. Late complications, including nodular regenerative hyperplasia and portal hypertension, are increasingly recognized in adolescent and adult patients with CGD. Interestingly, patients with CGD have been reported to be protected from atherosclerosis, suggesting an important role for NADPH oxidase (NOX2) in the pathogenesis of this inflammatory disease of arteries.

DISORDERS OF PHAGOCYTE ACTIVATION Phagocytes depend on cell-surface stimulation to induce signals that evoke multiple levels of the inflammatory response, including cytokine synthesis, chemotaxis, and antigen presentation. Mutations affecting the major pathway that signals through NF-κB have been noted in patients with a variety of infection susceptibility syndromes. If the defects are at a very late stage of signal transduction, in the protein critical for NF-κB activation known as the NF-κB essential modulator (NEMO), then affected males develop ectodermal dysplasia and severe immune deficiency with susceptibility to bacteria, fungi, mycobacteria, and viruses. There are several other proteins that interact with NEMO, mutations in which are autosomal and can be phenocopies of NEMO deficiency. If the defects in NF-κB activation are closer to the cell-surface receptors, in the proteins transducing Toll-like receptor signals, IL-1 receptor–associated kinase 4 (IRAK4), and myeloid differentiation primary response gene 88 (MyD88), then children have a marked susceptibility to pyogenic infections associated with a striking blunting of the febrile and inflammatory responses early in life but develop resistance to infection later.

The small Rho-family GTPase, RAC2, is involved in actin assembly, intracellular signaling, and superoxide production. Given these multiple roles, it is not surprising that different mutations in *RAC2* can have different effects on its functions. Accordingly, different *RAC2* mutations present, ranging from early-onset sepsis (constitutively active mutations), to abscesses with impaired superoxide production (dominant negative mutations), to common variable immunodeficiency (hyperactive mutations).

MONONUCLEAR PHAGOCYTES

The mononuclear phagocyte system is composed of monoblasts, promonocytes, and monocytes, in addition to the structurally diverse tissue macrophages that make up what was previously referred to as the reticuloendothelial system. Monocytes, which leave the circulation by diapedesis more slowly than neutrophils and have a half-life of 12–24 h, migrate into tissues and differentiate into macrophages. Macrophages are long-lived phagocytic cells capable of many of the functions of neutrophils. They are also secretory cells that participate in many immunologic and inflammatory processes distinct from neutrophils.

Many tissue macrophages ("big eaters") arise in the embryonic yolk sac outside of conventional hematopoiesis and then go on to take up residence in tissues. In addition, there are macrophages derived from monocytes, which may have specialized functions suited for specific anatomic locations. Macrophages are particularly abundant in capillary walls of the lung, spleen, liver, and bone marrow, where they function to remove microorganisms and other noxious elements from the blood. Alveolar macrophages, liver Kupffer cells, splenic macrophages, peritoneal macrophages, bone marrow macrophages, lymphatic macrophages, brain microglial cells, and dendritic macrophages all have specialized functions. Macrophage-secreted products include lysozyme, neutral proteases, acid hydrolases, arginase, complement components, enzyme inhibitors (plasmin, α_2-macroglobulin), binding proteins (transferrin, fibronectin, transcobalamin II), nucleosides, and cytokines (TNF-α; IL-1, 8, 12, 18). IL-1 **(Chaps. 20 and 360)** has many functions, including initiating fever in the hypothalamus, mobilizing leukocytes from the bone marrow, and activating lymphocytes and neutrophils. TNF-α is a pyrogen that duplicates many of the actions of IL-1 and plays an important role in the pathogenesis of gram-negative shock **(Chap. 315)**. TNF-α stimulates production of hydrogen peroxide and related toxic oxygen species by macrophages and neutrophils. In addition, TNF-α induces catabolic changes that contribute to the profound wasting (cachexia) associated with many chronic diseases.

Other macrophage-secreted products include reactive oxygen and nitrogen metabolites, bioactive lipids (arachidonic acid metabolites and platelet-activating factors), chemokines, CSFs, and factors stimulating fibroblast and vessel proliferation. Macrophages help regulate the replication of lymphocytes and participate in the killing of tumors, viruses, and certain bacteria (*Mycobacterium tuberculosis* and *Listeria monocytogenes*). Macrophages are key effector cells in the elimination of intracellular microorganisms. Their ability to fuse to form giant cells that coalesce into granulomas in response to some inflammatory stimuli is important in the elimination of intracellular microbes and is under the control of IFN-γ. Nitric oxide induced by IFN-γ may be an important effector against intracellular parasites, including tuberculosis and *Leishmania*.

Macrophages play an important role in the immune response **(Chap. 360)**. They process and present antigen to lymphocytes and secrete cytokines that modulate and direct lymphocyte development and function. Macrophages participate in autoimmune phenomena by removing immune complexes and other substances from the circulation. Polymorphisms in macrophage receptors for immunoglobulin (FcγRII) determine susceptibility to some infections and autoimmune diseases. In wound healing, they dispose of senescent cells, and they also contribute to atheroma development. Macrophage elastase mediates development of emphysema from cigarette smoking.

DISORDERS OF THE MONONUCLEAR PHAGOCYTE SYSTEM

Many disorders of neutrophils extend to mononuclear phagocytes. Monocytosis is associated with tuberculosis, brucellosis, subacute bacterial endocarditis, Rocky Mountain spotted fever, malaria, and visceral leishmaniasis (kala-azar). Monocytosis also occurs with certain malignancies, leukemias, myeloproliferative syndromes, hemolytic anemias, chronic idiopathic neutropenias, and granulomatous diseases such as sarcoidosis, Crohn disease, and some collagen vascular diseases. Patients with LAD1, hyperimmunoglobulin E–recurrent infection (Job's) syndrome, CHS, and CGD all have defects in the mononuclear phagocyte system.

Monocyte cytokine production or response is impaired in some patients with disseminated nontuberculous mycobacterial infection who are not infected with HIV. Genetic defects in the pathways regulated by IFN-γ and IL-12 lead to impaired killing of intracellular bacteria, mycobacteria, salmonellae, and certain viruses **(Fig. 67-11)**. Autoantibodies to IFN-γ, IL-23, and GM-CSF can block critical signals for macrophage intracellular killing and are associated with mycobacterial, fungal, and *Nocardia* infections.

Certain viral infections impair mononuclear phagocyte function. For example, influenza virus infection causes abnormal monocyte chemotaxis. Mononuclear phagocytes can be infected by HIV using CCR5, the chemokine receptor that acts as a co-receptor with CD4 for HIV. T lymphocytes produce IFN-γ, which induces FcR expression and phagocytosis and stimulates hydrogen peroxide production by mononuclear phagocytes and neutrophils. In certain diseases, such as AIDS, IFN-γ production may be deficient, whereas in other diseases, such as T-cell lymphomas, excessive release of IFN-γ may be associated with erythrophagocytosis by splenic macrophages.

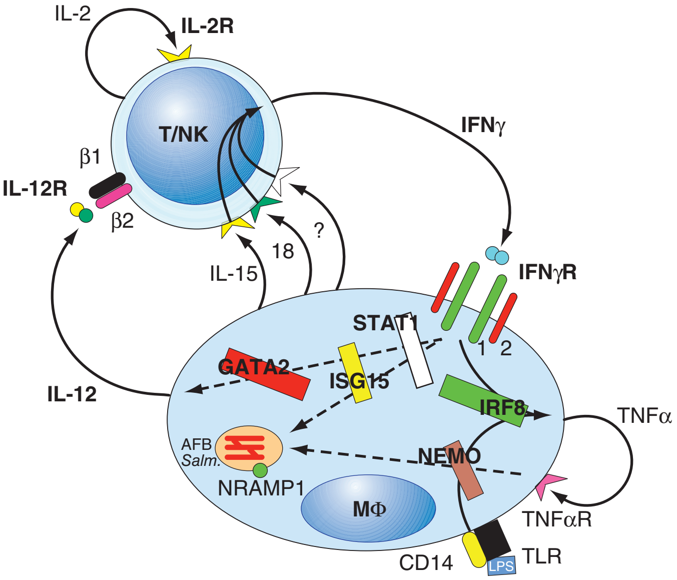

FIGURE 67-11 Lymphocyte-macrophage interactions underlying resistance to mycobacteria and other intracellular pathogens such as *Salmonella, Histoplasma,* and *Coccidioides*. Mycobacteria (and others) infect macrophages, leading to the production of IL-12, which activates T or NK cells through its receptor, leading to production of IL-2 and IFN-γ. IFN-γ acts through its receptor on macrophages to upregulate TNF-γ and IL-12 and kill intracellular pathogens. Other critical interacting molecules include signal transducer and activator of transcription 1 (STAT1), interferon regulatory factor 8 (IRF8), GATA2, and ISG15. Mutant forms of the cytokines and receptors shown in *bold type* have been found in severe cases of nontuberculous mycobacterial infection, salmonellosis, and other intracellular pathogens. AFB, acid-fast bacilli; IFN, interferon; IL, interleukin; NEMO, nuclear factor-κB essential modulator; NK, natural killer; TLR, Toll-like receptor; TNF, tumor necrosis factor.

Autoinflammatory diseases are characterized by abnormal cytokine regulation, leading to excess inflammation in the absence of infection. These diseases can mimic infectious or immunodeficient syndromes. Gain-of-function mutations in the TNF-α receptor cause TNF-α receptor–associated periodic syndrome (TRAPS), which is characterized by recurrent fever in the absence of infection, due to persistent stimulation of the TNF-α receptor **(Chap. 381)**. Diseases with abnormal IL-1 regulation leading to fever include familial Mediterranean fever due to mutations in *PYRIN*. Mutations in *cold-induced autoinflammatory syndrome 1* (*CIAS1*) lead to neonatal-onset multisystem autoinflammatory disease, familial cold urticaria, and Muckle-Wells syndrome. The syndrome of *p*yoderma gangrenosum, *a*cne, and sterile *p*yogenic *a*rthritis (PAPA syndrome) is caused by mutations in *PSTPIP1*. In contrast to these syndromes of overexpression of proinflammatory cytokines, blockade of TNF-α by the antagonists infliximab, adalimumab, certolizumab, golimumab, or etanercept has been associated with severe infections due to tuberculosis, nontuberculous mycobacteria, and fungi **(Chap. 381)**.

Monocytopenia occurs with acute infections, with stress, and after treatment with glucocorticoids. Drugs that suppress neutrophil production in the bone marrow can cause monocytopenia. Persistent severe circulating monocytopenia is seen in GATA2 deficiency, even though macrophages are found at the sites of inflammation. Monocytopenia also occurs in aplastic anemia, hairy cell leukemia, acute myeloid leukemia, and as a direct result of myelotoxic drugs.

EOSINOPHILS

Eosinophils and neutrophils share similar morphology, many lysosomal constituents, phagocytic capacity, and oxidative metabolism. Eosinophils express a specific chemoattractant receptor and respond to a specific chemokine, eotaxin, but little is known about their required role. Eosinophils are much longer lived than neutrophils, and unlike neutrophils, tissue eosinophils can recirculate. During most infections, eosinophils appear unimportant. However, in invasive helminthic infections, such as hookworm, schistosomiasis, strongyloidiasis, toxocariasis, trichinosis, filariasis, echinococcosis, and cysticercosis, the eosinophil plays a central role in host defense. Recently, eosinophils have been recognized as important in the host response to tuberculosis. Eosinophils are associated with bronchial asthma, cutaneous allergic reactions, and other hypersensitivity states.

The distinctive feature of the red-staining (Wright's stain) eosinophil granule is its crystalline core consisting of an arginine-rich protein (major basic protein) with histaminase activity, important in host defense against parasites. Eosinophil granules also contain a unique eosinophil peroxidase that catalyzes the oxidation of many substances by hydrogen peroxide and may facilitate killing of microorganisms.

Eosinophil peroxidase, in the presence of hydrogen peroxide and halide, initiates mast cell secretion in vitro and thereby promotes inflammation. Eosinophils contain cationic proteins, some of which bind to heparin and reduce its anticoagulant activity. Eosinophil-derived neurotoxin and eosinophil cationic protein are ribonucleases that can kill respiratory syncytial virus. Eosinophil cytoplasm contains Charcot-Leyden crystal protein, a hexagonal bipyramidal crystal first observed in a patient with leukemia and then in sputum of patients with asthma; this protein is lysophospholipase and may function to detoxify certain lysophospholipids.

Several factors enhance the eosinophil's function in host defense. T cell–derived factors enhance the ability of eosinophils to kill parasites. Mast cell–derived eosinophil chemotactic factor of anaphylaxis (ECFa) increases the number of eosinophil complement receptors and enhances eosinophil killing of parasites. Eosinophil CSFs (e.g., IL-5) produced by macrophages increase eosinophil production in the bone marrow and activate eosinophils to kill parasites.

■ EOSINOPHILIA

Eosinophilia is the presence of >500 eosinophils per μL of blood and is common in many settings besides parasite infection. Significant tissue eosinophilia can occur without an elevated blood count. A common cause of eosinophilia is allergic reaction to drugs (iodides, aspirin, sulfonamides, nitrofurantoin, penicillins, and cephalosporins). Allergies such as hay fever, asthma, eczema, serum sickness, allergic vasculitis, and pemphigus are associated with eosinophilia. Eosinophilia also occurs in collagen vascular diseases (e.g., rheumatoid arthritis, eosinophilic fasciitis, allergic angiitis, and periarteritis nodosa) and malignancies (e.g., Hodgkin disease; mycosis fungoides; chronic myeloid leukemia; and cancer of the lung, stomach, pancreas, ovary, or uterus), as well as in dominant negative STAT3 (Job's syndrome), DOCK8 deficiency (see below), and CGD. Eosinophilia is commonly present in helminthic infections. IL-5 is the dominant eosinophil growth factor. Therapeutic administration of the cytokines IL-2 or GM-CSF frequently leads to transient eosinophilia. The most dramatic hypereosinophilic syndromes are Loeffler's syndrome, tropical pulmonary eosinophilia, Loeffler's endocarditis, eosinophilic leukemia, and idiopathic hypereosinophilic syndrome (50,000–100,000/μL). IL-5 is the dominant eosinophil growth factor and can be specifically inhibited with monoclonal antibodies against it or its receptor.

The idiopathic hypereosinophilic syndromes are a heterogeneous group of disorders with the common feature of prolonged eosinophilia of unknown cause and organ system dysfunction, including the heart, central nervous system, kidneys, lungs, gastrointestinal tract, and skin. The bone marrow is involved in all affected individuals, but the most severe complications involve the heart and central nervous system. Clinical manifestations and organ dysfunction are highly variable. Eosinophils are found in the involved tissues and likely cause tissue damage by local deposition of toxic eosinophil proteins such as eosinophil cationic protein and major basic protein. In the heart, the pathologic changes lead to thrombosis, endocardial fibrosis, and restrictive endomyocardiopathy. The damage to tissues in other organ systems is similar. Some cases are due to mutations involving the platelet-derived growth factor receptor, and these are extremely sensitive to the tyrosine kinase inhibitor imatinib. Glucocorticoids, hydroxyurea, and IFN-α each have been used successfully, as have therapeutic antibodies against IL-5 or its receptor. Cardiovascular complications are managed aggressively.

The *eosinophilia-myalgia syndrome* is a multisystem disease, with prominent cutaneous, hematologic, and visceral manifestations, that frequently evolves into a chronic course and can occasionally be fatal. The syndrome is characterized by eosinophilia (eosinophil count >1000/μL) and generalized disabling myalgias without other recognized causes. Eosinophilic fasciitis, pneumonitis, and myocarditis; neuropathy culminating in respiratory failure; and encephalopathy may occur. The disease is thought to be caused by ingesting contaminants in L-tryptophan–containing products. Eosinophils, lymphocytes, macrophages, and fibroblasts accumulate in the affected tissues, but their role in pathogenesis is unclear. Activation of eosinophils and fibroblasts and the deposition of eosinophil-derived toxic proteins in affected tissues may contribute. IL-5 and transforming growth factor β have been implicated as potential mediators. Treatment is withdrawal of products containing L-tryptophan and the administration of glucocorticoids. Most patients recover fully, remain stable, or show slow recovery, but the disease can be fatal in up to 5% of patients.

Eosinophilic neoplasms are discussed in Chap. 115.

■ EOSINOPENIA

Eosinopenia occurs with stress, such as acute bacterial infection, and after treatment with glucocorticoids. The mechanism of eosinopenia of acute bacterial infection is unknown but is independent of endogenous glucocorticoids, because it occurs in animals after total adrenalectomy. There is no known adverse effect of eosinopenia.

HYPERIMMUNOGLOBULIN E–RECURRENT INFECTION SYNDROME

The hyperimmunoglobulin E–recurrent infection syndrome, Job's syndrome, is a rare multisystem disease in which the immune and somatic systems are affected, including neutrophils, monocytes, T cells, B cells, and osteoclasts. Autosomal dominant negative mutations in signal transducer and activator of transcription 3 (STAT3) impair normal STAT signaling with broad and profound effects. Patients have characteristic facies with broad nose, kyphoscoliosis, and eczema. Primary teeth erupt normally but do not deciduate, often requiring extraction. Recurrent sinopulmonary and cutaneous infections tend to elicit much less inflammation than appropriate for the degree of infection, leading to "cold abscesses." Pneumonias typically cavitate, leading to pneumatoceles. Coronary artery aneurysms are common, as are cerebral demyelinated plaques that accumulate with age. IL-17–producing T cells, which are thought responsible for protection against extracellular and mucosal infections, are reduced in Job's syndrome. Despite very high IgE levels, Job's patients have only mildly elevated levels of allergy. Patients with autosomal recessive defects in dedicator of cytokinesis 8 (DOCK8) also have very high IgE levels joined to severe allergy, extensive viral susceptibility, and increased rates of cancer. Autosomal dominant *gain-of-function* (GOF) mutations in STAT3 lead to a disease characterized by onset in childhood of lymphadenopathy, autoimmune cytopenias, multiorgan autoimmunity, infections, and interstitial lung disease.

LABORATORY DIAGNOSIS AND MANAGEMENT

A complete blood count (CBC) and differential are essential. Careful examination of neutrophils on peripheral blood smears can diagnose CHS and suggest other neutrophil granule abnormalities such as specific granule deficiency. Bone marrow examination, where indicated, should be accompanied by either gene panel or whole exome/genome sequencing if genetic defects are suspected. Bone marrow reserves (steroid challenge test), marginated circulating pool of cells (epinephrine challenge test), and marginating ability (endotoxin challenge test) are also doable (Fig. 67-8). In vivo assessment of inflammation is possible with a Rebuck skin window test or an in vivo skin blister assay, which measure the ability of leukocytes and inflammatory mediators to accumulate locally in the skin. In vitro tests of phagocyte aggregation, adherence, chemotaxis, phagocytosis, degranulation, and microbicidal activity may help pinpoint cellular or humoral lesions. CGD is detected with either the nitroblue tetrazolium (NBT) dye test or the DHR oxidation test. These tests are based on the ability of products of oxidative metabolism to alter the oxidation states of reporter molecules so that they can be detected microscopically (NBT) or by flow cytometry (DHR). While the DHR is very sensitive for CGD, impaired DHR responses are seen in myeloperoxidase deficiency and with acetaminophen ingestion.

Patients with leukopenias or leukocyte dysfunction often have delayed inflammatory responses. Therefore, clinical manifestations may be minimal despite overwhelming infection, and unusual infections must always be suspected. Early signs of infection demand prompt, aggressive culturing for microorganisms, use of antibiotics, and drainage of abscesses. Prolonged courses of antibiotics are often required. In patients with CGD, prophylactic antibiotics (trimethoprim-sulfamethoxazole) and antifungals (itraconazole) markedly diminish the frequency of life-threatening infections. Glucocorticoids may relieve gastrointestinal or genitourinary tract obstruction by granulomas in patients with CGD. Although TNF-α-blocking agents may markedly relieve inflammatory bowel symptoms, extreme caution must be exercised in their use in CGD inflammatory bowel disease, because it profoundly increases these patients' already heightened susceptibility to infection. Recombinant human IFN-γ, which nonspecifically stimulates phagocytic cell function, reduces the frequency of infections in patients with CGD by 70% and reduces the severity of infection. This effect of IFN-γ in CGD is additive to the effect of prophylactic antibiotics. The recommended dose is 50 μg/m^2 subcutaneously three times weekly. IFN-γ has also been used successfully in the treatment of leprosy, nontuberculous mycobacteria, and visceral leishmaniasis.

Rigorous oral hygiene reduces but does not eliminate the discomfort of gingivitis, periodontal disease, and aphthous ulcers; chlorhexidine mouthwash and tooth brushing with a hydrogen peroxide–sodium bicarbonate paste also helps many patients. Oral antifungal agents (fluconazole, itraconazole, voriconazole, posaconazole) have reduced mucocutaneous candidiasis in patients with Job's syndrome. Recombinant G-CSF is useful in the management of certain forms of neutropenia due to depressed neutrophil production, including those related to cancer chemotherapy. Patients with chronic neutropenia with evidence of a good bone marrow reserve need not receive prophylactic antibiotics. Patients with chronic or cyclic neutrophil counts <500/μL may benefit from prophylactic antibiotics and G-CSF during periods of neutropenia. Oral trimethoprim-sulfamethoxazole (160/800 mg) twice daily can prevent infection. Increased numbers of fungal infections are not seen in patients with CGD on this regimen. Oral quinolones such as levofloxacin and ciprofloxacin are alternatives.

In the setting of cytotoxic chemotherapy with severe, persistent lymphocyte dysfunction, trimethoprim-sulfamethoxazole prevents *Pneumocystis jirovecii* pneumonia. These patients, and patients with phagocytic cell dysfunction, should avoid heavy exposure to airborne soil, dust, or decaying matter (mulch, manure), which are often rich in *Nocardia* and the spores of *Aspergillus* and other fungi. Restriction of activities or social contact has no proven role in reducing risk of infection for phagocyte defects.

Although aggressive medical care for many patients with phagocytic disorders can allow them to go for years without a life-threatening infection, there may still be delayed effects of prolonged antimicrobials and other inflammatory complications. Cure of most congenital phagocyte defects is possible by bone marrow transplantation, and rates of success are improving **(Chap. 119)**. The identification of specific gene defects in patients with LAD1, CGD, and other immunodeficiencies has led to gene therapy trials in a number of genetic white cell disorders.

■ FURTHER READING

Boeltz S et al: To NET or not to NET: Current opinions and state of the science regarding the formation of neutrophil extracellular traps. Cell Death Differ 26:395, 2019.

Bousfiha A et al: The 2022 update of IUIS phenotypical classification for human inborn errors of immunity. J Clin Immunol 42:1508, 2022.

Donko A et al: Interpretation of dihydrorhodamine-1,2,3 flow cytometry in chronic granulomatous disease: An atypical exemplar. J Clin Immunol 42:986, 2022.

Donko A et al: Clinical and functional spectrum of RAC2-related immunodeficiency. Blood 143:1476, 2024.

Idol RA et al: Neutrophil and macrophage NADPH oxidase 2 differentially control responses to inflammation and to *Aspergillus fumigatus* in mice. J Immunol 209:1960, 2022.

Khoury P et al: Biologic therapy in rare eosinophil-associated disorders: Remaining questions and translational research opportunities. J Leukoc Biol 116:307, 2024.

Lazarov T et al: Physiology and diseases of tissue-resident macrophages. Nature 618:698, 2023.

Ochoa S et al: Genetic susceptibility to fungal infection in children. Curr Opin Pediatr 32:780, 2020.

Peiseler M, Kubes P: More friend than foe: the emerging role of neutrophils in tissue repair. J Clin Invest 129:2629, 2019.

Tangye SG et al: Human Inborn Errors of Immunity: 2022 Update on the Classification from the International Union of Immunological Societies Expert Committee. J Clin Immunol 42:1473, 2022.

Zerbe CS, Holland SM: Functional neutrophil disorders: Chronic granulomatous disease and beyond. Immunol Rev 322:71, 2024.

68

Primary and Secondary Hemophagocytic Lymphohistiocytosis

Jan-Inge Henter

Hemophagocytic lymphohistiocytosis (HLH) is a highly life-threatening condition of massive, uncontrolled inflammation (a hyperinflammation). It exists in a primary (genetic) form and a secondary (acquired) form, and is often triggered by infections, malignancies, and autoimmune diseases. HLH triggered by autoimmune conditions is also referred to as macrophage activation syndrome (MAS or MAS-HLH). While the underlying factors are biologically very different, the clinical picture of all forms of HLH is quite similar, with a sepsis-like inflammation often leading to multiple organ failure and death if not treated promptly and appropriately (convergent phenotype). Primary HLH mostly affects children, whereas secondary HLH is far more common in adults. The current chapter will present primary and secondary HLH separately. A classification of HLH is presented in **Table 68-1**.

TABLE 68-1 Simplified Classification of Hemophagocytic Lymphohistiocytosis (HLH)

Primary HLH: Mendelian Conditions

- Lymphocyte cytotoxic defects
  - FHL2 (*PRF1*)
  - FHL3 (*UNC13D*)
  - FHL4 (*STX11*)
  - FHL5 (*STXBP2*)
  - XLP1 (*SH2D1A*)
  - Griscelli syndrome type 2 (*RAB27A*)
  - Chédiak-Higashi syndrome (*LYST*)
- Abnormalities of inflammasome activation
  - XLP2 (*BIRC4*)
  - NLRC4
- Mendelian disorders affecting inflammation
  - Lysinuric protein intolerance (*SLC7A7*)
  - Wolman disease (*LIPA*)

Secondary HLH (Non-Mendelian)

- Infection-associated HLH
  - Virus-associated HLH
  - Bacteria-associated HLH
  - Parasite-associated HLH
  - Fungal-associated HLH
- Malignancy-associated HLH
  - Malignancy-triggered HLH
  - HLH occurring during chemotherapy
- Autoimmune-associated HLH (MAS-HLH)
  - HLH with SoJIA
  - HLH with adult-onset Still's disease
  - HLH with SLE
  - HLH with vasculitis
- Transplant-associated HLH

Abbreviations: MAS, macrophage activation syndrome; SLE, systemic lupus erythematosus; SoJIA, systemic-onset juvenile idiopathic arthritis.

Source: Reproduced with permission from JF Emile et al: Revised classification of histiocytoses and neoplasms of the macrophage-dendritic cell lineages. Blood 127: 2672, 2016.

PRIMARY HEMOPHAGOCYTIC LYMPHOHISTIOCYTOSIS

Primary HLH comprises Mendelian-inherited conditions causing HLH. The most common form, by far, is familial HLH (FHL), which is associated with lymphocyte cytotoxicity defects.

Other forms of primary HLH associated with lymphocyte cytotoxicity defects are Griscelli syndrome type 2 (associated with hypopigmentation, silver-colored hair, and immune defects due to mutation in *RAB27A*, a *RAS* gene family member), Chédiak-Higashi syndrome, and X-linked lymphoproliferative syndrome (XLP) type 1. Another group of primary HLHs are associated with abnormalities of inflammasome activation, inborn errors of immunity, and inborn errors of metabolism.

FAMILIAL HEMOPHAGOCYTIC LYMPHOHISTIOCYTOSIS

If not treated, FHL is rapidly fatal, with a median survival of 1–2 months. The incidence of FHL has been estimated to be 1 in 50,000 live births. Together with severe combined immunodeficiency, it is the most common fatal inherited human immunodeficiency. Because its inheritance is autosomal recessive, it is more common in regions where consanguinity is frequent; it is equally common in males and females. The median age at onset is between 3 and 6 months. It may be diagnosed at birth or even in utero, whereas less severe forms may result in onset during adolescence and adulthood.

Clinical and Laboratory Features A typical presentation is a fever, often in an infant, with a sepsis-like condition associated with cytopenias, in particular thrombocytopenia, and hepatosplenomegaly. If blood cultures are negative and signs of malignancy are absent, HLH should be considered and assessed for promptly.

HEMATOLOGIC AND BONE MARROW FINDINGS Thrombocytopenia is frequent, and the platelet levels may be extremely low. Anemia and neutropenia are also common, and bicytopenia (i.e., at least two of these three lineages affected) is a diagnostic criterion.

Hemophagocytosis is not always present, particularly in the early phase of the disease, and hence, serial examinations over time may be helpful. However, hemophagocytosis is not specific for HLH. Over time, the bone marrow may become hypocellular.

LIVER INVOLVEMENT AND COAGULOPATHY Ferritin and transaminase levels are almost always elevated. Ferritin, an acute-phase reactant (i.e., a marker of inflammation), can serve as a rapid and easily available screening test for HLH, although it is not specific for the disease. Icterus, with a predominantly conjugated hyperbilirubinemia, and elevated γ-glutamyl transferase levels are also common. HLH may resemble acute liver failure, and in histopathologic specimens, an image similar to chronic persistent hepatitis may be found. Disseminated intravascular coagulation, severe acute bleeding, and low levels of fibrinogen are frequent.

CENTRAL NERVOUS SYSTEM INVOLVEMENT About one a third of children with FHL have neurologic symptoms at diagnosis, which may be severe, and around half have abnormal cerebrospinal fluid with elevated cell and/or protein content. Common symptoms include seizures, decreased consciousness, and signs of meningismus, as well as irritability and hypotonia in neonates and infants. Cranial nerve palsies and ataxia may also occur. Isolated neurologic symptoms may be the first presentation of FHL, particularly in older children and adolescents. Magnetic resonance imaging findings are nonspecific.

OTHER CLINICAL AND LABORATORY FINDINGS Petechia, purpura, and a nonspecific transient maculopapular rash are frequent. Lymph node enlargement develops in about half of the children. Genetic defects in *STXBP2* may cause gastrointestinal symptoms with severe diarrhea that persists despite successful hematopoietic stem cell transplantation (HSCT). Triglyceride levels are often significantly elevated, and sodium and albumin levels decreased. Soluble interleukin (IL) 2 receptor (sIL-2r; also known as sCD25) levels are often markedly elevated, which, together with ferritin, may serve as biomarkers for monitoring of HLH severity.

Late-onset disease, such as after the age of 7 years, may be less typical, both clinically and biologically, and often represents a diagnostic challenge.

LONG-TERM COMPLICATIONS The most frequent late effects are neurologic sequelae, affecting around one-third of patients, with the most common being psychomotor retardation (25%), epilepsy (10%), attention-deficit/hyperactivity disorder (10%), ataxia, and hemiplegia/tetraplegia. Prompt treatment of HLH is essential to reduce these. Approximately 10% develop endocrinopathies.

Pathophysiology HLH is often described as a cytokine storm, a hypercytokinemia. Inflammatory cytokines such as interferon (IFN)-γ, IL-1β, IL-6, IL-18, and tumor necrosis factor (TNF)-α are involved in the development of HLH, and in particular, IFN-γ plays a pivotal role in both primary and secondary HLH. Serum levels of C-X-C motif chemokine ligand 9 (CXCL9), an IFN-γ–induced chemokine, are also elevated.

LYMPHOCYTE CYTOTOXICITY DEFECTS In FHL, the initiation of the immune response is appropriate, but there is an inability to regulate and terminate it due to defective function of natural killer (NK) cells and cytotoxic T cells. These cells are crucial for immune surveillance against virus-infected cells and tumor cells, and studies of FHL have highlighted their importance also in the regulation and termination of immune responses.

Cytotoxic lymphocytes secrete granules that contain perforin and granzymes. Perforin forms pores in the target cell membrane that facilitate entry of granzymes into the cytosol of the target cell where they promote programmed cell death (apoptosis) **(Fig. 68-1)**. If this process is defective, as in FHL, the adaptive immune system is overactivated with uncontrolled expansion and activation of polyclonal CD8+ cytotoxic T cells. Inefficient clearance of pathologic antigens, overactivation of inflammatory cells with resulting hypercytokinemia, and inability to terminate the excessive immune response lead to a vicious cycle of inflammation and subsequent tissue destruction in several organs leading to a state of multiorgan failure.

Individuals with mutations in FHL-causing genes also have a markedly increased risk of developing malignancies, but there is no apparent increase in viral infections.

GENETIC DEFECTS FHL is an autosomal recessive disease caused by genetic variants in four specific genes, *PRF1*, *UNC13D*, *STX11*, and *STXBP2*, encoding the proteins perforin, Munc13-4, syntaxin-11, and Munc18-2, respectively. A closely related disorder is Griscelli syndrome type 2, caused by variants in *RAB27A* encoding Rab27a (Fig. 68-1). The frequency of these gene variants varies in different ethnic groups, but, in general, variants in *PRF1* and *UNC13D* are most common. Some variants are associated with more severe disease and early onset, while others often cause a milder phenotype with later onset.

APPROACH TO THE PATIENT

Possible Primary Hemophagocytic Lymphohistiocytosis

Early diagnosis is crucial because of the high risk of rapid death and central nervous system (CNS) damage if adequate treatment is delayed. Ferritin levels are often elevated and suggested as a screening test in every child with suspected HLH. In adults, elevated ferritin is less indicative of HLH, but markedly elevated levels should raise suspicion of HLH.

CLINICAL DIAGNOSTIC CRITERIA

The diagnosis of FHL can be made clinically, genetically, and by evaluation of lymphocyte cytotoxicity. The gold standard is molecular confirmation of HLH-causing genetic variants, but this may take weeks. Therefore, a set of diagnostic criteria has been developed, the majority of which are readily available. In the current international HLH-2004 criteria, five of the following eight criteria need to be fulfilled for a diagnosis of FHL: fever, splenomegaly, bicytopenia (hemoglobin <90 g/L [<100 g/L in infants <4 weeks]), platelets $<100 \times 10^9$/L, neutrophils $<1.0 \times 10^9$/L, hypertriglyceridemia (fasting triglycerides ≥3.0 mmol/L; i.e., 265 mg/dL) and/or hypofibrinogenemia (fibrinogen ≤1.5 g/L), hemophagocytosis, ferritin ≥500 μg/L, low or absent NK cell activity, and soluble IL-2 receptor alpha levels ≥2400 U/mL. In 2024, the Histiocyte Society decided to remove the criterion on NK cell activity, and instead five of the remaining seven criteria are now to be fulfilled. Not all FHL patients fulfill these diagnostic criteria, particularly at onset, so in some cases, therapy has to be commenced on strong clinical suspicion of HLH before overwhelming disease activity gives irreversible damage.

GENETIC AND CELLULAR DIAGNOSTIC ASSAYS

Confirmation of pathogenic genetic variants is the cornerstone of diagnosis. Functional cytotoxicity assays are often faster and can suggest genetically determined HLH. Absent perforin expression by flow cytometric assay, quantifying intracellular perforin expression in cytotoxic cells, suggests *PRF1* genetic variants. Defective patient cell exocytosis, assayed by CD107a fluorescence at the cell membrane, is indicative of *UNC13D*, *STXBP2*, *STX11*, and *RAB27A* variants. Genetic and lymphocyte cytotoxicity studies are recommended in unaffected siblings. Prenatal testing is possible if the pathogenic gene variant in the family is known.

TREATMENT

At diagnosis, the patients are often critically ill, so the first step is to stop the hyperinflammation promptly in order for the patient to survive and to reduce CNS damage. In parallel, search for and treat triggering factors. The second step is a curative HSCT to replace the defective immune system.

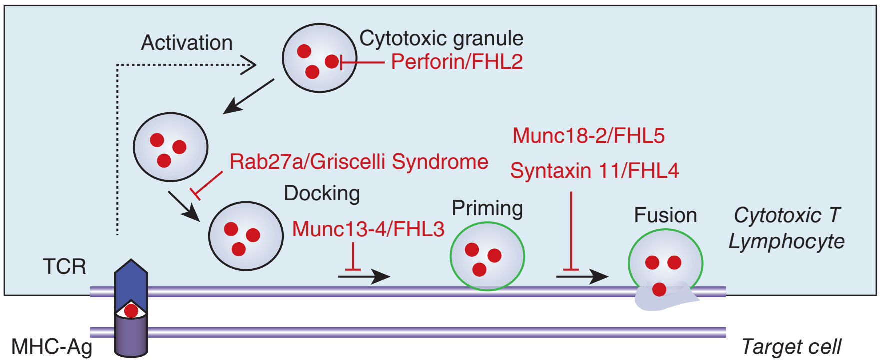

FIGURE 68-1 Inborn errors in the cytotoxic activity of lymphocytes. The genetic defects causing hemophagocytic lymphohistiocytosis (HLH) affect a precise step of the cytotoxic machinery, i.e., granule content, docking, priming, or fusion. Only the defects causing Griscelli syndrome (type 2) and familial hemophagocytic lymphohistiocytosis (FHL) are shown. *(Reproduced with permission from JP Schmid et al: Inherited defects in lymphocyte cytotoxic activity. J Immunol Rev 235:10, 2010.)*

Pretransplant Therapy The most established treatments are the HLH-94 and HLH-2004 treatment protocols, both based on dexamethasone, which penetrates well into the brain, and the cytotoxic drug etoposide, which has remarkably positive effects in FHL. In addition, cyclosporin A, which decreases the production of inflammatory cytokines by T lymphocytes may be administered. Biologically, etoposide causes a potent selective depletion of activated T cells as well as efficient suppression of inflammatory cytokine production.

Other drugs used in the treatment of primary HLH, almost always in combination with corticosteroids, include alemtuzumab, a monoclonal antibody directed against the CD52 antigen expressed on lymphocytes; emapalumab, an IFN-γ–blocking antibody; and Janus kinase inhibitors, such as ruxolitinib, which inhibits several inflammatory cytokines.

In large series of patients with verified FHL receiving first-line treatment with etoposide or alemtuzumab (anti-CD52), pre-HSCT survival is now almost 90% in experienced centers. Emapalumab (anti–IFN-γ) alone appears to be less effective in these patients, and data on ruxolitinib (JAK 1 and 2 inhibitor) are limited.

Hematopoietic Stem Cell Transplantation Allogeneic HSCT is the only curative treatment for FHL at this time. Pretransplant complete remission is beneficial but not mandatory to achieve post-HSCT survival. Survival after HSCT has reached 85–90% in experienced centers. Thus, overall survival using traditional first-line therapy with etoposide is now almost 80%. The survival is close to 100% in asymptomatic siblings.

OTHER FORMS OF PRIMARY HEMOPHAGOCYTIC LYMPHOHISTIOCYTOSIS

HLH and Partial Albinism Three conditions are characterized by both a predisposition for HLH and abnormal pigmentation, all autosomal recessive: Griscelli syndrome type 2, Chédiak-Higashi syndrome, and the less frequent Hermansky-Pudlak syndrome type II.

GRISCELLI SYNDROME TYPE 2 This disorder is characterized by a combination of hypopigmentation, characterized by silvery-gray hair and eyelashes, and a predisposition to HLH and is caused by defect gene variants in *RAB27A* encoding Rab27a. The clinical picture is very similar to FHL. Diagnosis, treatment, and late effects are also similar to FHL, with the diagnostically helpful addition that hair has irregular pigment clumps seen by ordinary light microscopy.

CHÉDIAK-HIGASHI SYNDROME In addition to HLH, this disorder is associated with increased risk of bacterial infections and a progressive neurologic disorder. Diagnosis can be made genetically (variants in the gene *LYST*), by light microscopy of the hair, and also by the presence of characteristic giant lysosomes within leukocytes. Allogeneic HSCT, curative for the HLH, does not cure the neurologic disorder, which often is the most prominent long-term problem.

X-Linked Lymphoproliferative Syndrome (XLP) XLP comes in two distinct forms, XLP1 and XLP2, both associated with an increased risk of HLH. The diagnosis is based on identification of a disease-causing gene variant (*SH2D1A* and *XIAP/BIRC4*, respectively) or deficient expression of the encoded proteins SAP and XIAP. Onset is commonly observed in males of preschool age but may also be observed in adolescence or adulthood. The HLH is frequently, but not always, triggered by an Epstein-Barr virus (EBV) infection, in which case rituximab (anti-CD20) should be considered.

XLP1 is associated with a dysgammaglobulinemia, typically a hypogammaglobulinemia of one or more immunoglobulin subclasses, which can be treated with immunoglobulin replacement therapy. Almost one-third of patients develop lymphomas; hence, allogeneic HSCT should be considered early. Because many patients do not present with all symptoms or present with varying severity, less intensive pre-HSCT therapy than HLH-94/HLH-2004 may be used.

In XLP2, also called XIAP deficiency, the most frequent clinical manifestations are HLH-like findings that may be mild to severe and an inflammatory bowel disease resembling Crohn's disease, which may be severe and treatment-resistant. The disease course is very variable. The HLH can sometimes be controlled by mild immunosuppression and sometimes also without HSCT. Myeloablative conditioning has been associated with unfavorable outcome.

Other Causes of Primary HLH Defective NLR (nucleotide-binding domain leucine-rich repeat containing) family CARD (caspase recruitment domain) domain-containing protein 4 (NLRC4), caused by variants in *NLRC4*, may cause severe, early-onset HLH with markedly elevated IL-18 levels. Life-threatening NLRC4-associated HLH can be treated with IL-18 inhibition. Other Mendelian conditions associated with HLH include inborn errors of metabolism such as lysinuric protein intolerance and Wolman disease (lysosomal acid lipase deficiency).

SECONDARY HEMOPHAGOCYTIC LYMPHOHISTIOCYTOSIS

INTRODUCTION

Clinically, primary (genetic) HLH and secondary (acquired) HLH are similar and can be difficult to distinguish. However, important differences exist. While primary HLH typically has its onset during infancy, secondary HLH is mostly seen in adults, although it can be found in all ages. Moreover, secondary HLH is not Mendelian inherited and is instead caused by various triggering factors. Both forms are associated with high mortality, but while primary HLH almost always requires HSCT for cure, HSCT is rarely indicated in secondary HLH, for which, instead, tailored HLH-directed therapy is recommended.

Triggering Factors In a review of adults, infections were associated with 50% of HLH cases, neoplasms with 48%, autoimmune diseases with 13%, and transplantations with 4%. The most common associated infections were viral (35% of all patients), with EBV (15%) and HIV (8%) being most frequent, and bacterial infections triggering in 9%. Among malignancies, hematologic malignancies were by far the most frequent (45% of all patients), mostly T- and NK-cell lymphomas (17%) and B-cell lymphomas (15%). Among autoimmune diseases, systemic lupus erythematosus (6%) and adult-onset Still's disease (2%) were most common.

Epidemiology Although the awareness and reported incidence of secondary HLH have increased dramatically, it is still underdiagnosed and, therefore, undertreated. Population-based studies have reported an annual incidence of 4.2 per million people in England and an annual incidence of malignancy-associated HLH of 6.2 per million adults in Sweden.

CLINICAL AND LABORATORY FEATURES

A patient with secondary HLH often presents as a febrile, critically ill patient with severe inflammation who is unresponsive to sepsis-directed therapy. Rapid deterioration is common, and unexpected and disproportional clinical and laboratory abnormalities are often noted. Typical findings largely resemble those in familial HLH.

Clinical Features The clinical findings partly depend on the underlying trigger. Prolonged fever is almost always present. Splenomegaly and hepatomegaly are reported in about two-thirds of patients, while pulmonary involvement, lymphadenopathy, skin rash, and neurologic involvement are reported in a quarter to half of all patients.

Laboratory Findings The inflammatory markers ferritin and soluble IL-2 receptor alpha are almost always elevated, often markedly. Both of these can be used to monitor HLH intensity. Thrombocytopenia, anemia, and elevated levels of serum transaminases, lactate dehydrogenase, and triglycerides are often found. Elevated levels of D-dimer, low levels of fibrinogen, and disseminated intravascular coagulation are each reported in about half of the patients. Importantly, hemophagocytosis is not always detected and is not a prerequisite for the diagnosis.

CNS Involvement Neurologic symptoms are heterogeneous, reported in up to one-quarter of the patients, and include coma, seizures, ataxia, meningitis, encephalitis, cranial nerve palsies, hemiplegia, or cerebral hemorrhage. Psychiatric disorders may also develop. About half of patients have abnormal findings reported on magnetic resonance imaging. The reported frequency of cerebrospinal fluid abnormalities is 15–40%.

PATHOPHYSIOLOGY

The cause of secondary HLH is multifactorial, but it often has a dominating underlying cause. According to "the threshold model," the factors triggering and unbalancing the immune system accumulate, eventually reaching a threshold at which inflammation becomes uncontrolled and fulminant HLH develops (Fig. 68-2). Diverse pathways and etiologies can thus result in the same common end stage of HLH. Although secondary HLH is not an inherited disorder, whole-genome sequencing has revealed variably pathogenic monoallelic HLH-associated genetic variants in some patients. As in primary HLH, inflammatory cytokines play a crucial role in the development of symptoms and signs in secondary HLH. The cytokine pattern is different in the various forms of secondary HLH, which can be used diagnostically.

Lymphocyte Cytotoxicity In contrast to familial HLH, which is characterized by inherited defects in cytotoxic cells but normal numbers of cytotoxic cells, the number of circulating NK cells and cytotoxic T cells is often reduced in secondary HLH. As a consequence, the overall lymphocyte cytotoxic function can be deficient, as in familial HLH.

DIAGNOSIS

Secondary HLH, despite being life-threatening and treatable, is markedly underdiagnosed. Awareness of the diagnosis is, therefore, essential. As in primary HLH, elevated ferritin levels may serve as a rapid and readily available screening test.

In contrast to primary HLH, no exact or specific diagnostic test is available for secondary HLH. For diagnostic assistance, combinations of clinical and laboratory criteria are often used, such as the HLH-2004 criteria presented above and the HScore. Although the diagnosis of HLH typically is based on fulfilment of five HLH-2004 criteria, many centers find four fulfilled criteria sufficient to establish the diagnosis in secondary HLH. The HScore is a weighted scoring system based on nine variables: fever, organomegaly, known underlying immunosuppression, number of cytopenias, ferritin, triglycerides, fibrinogen, aspartate aminotransferase, and hemophagocytosis. Both HLH-2004

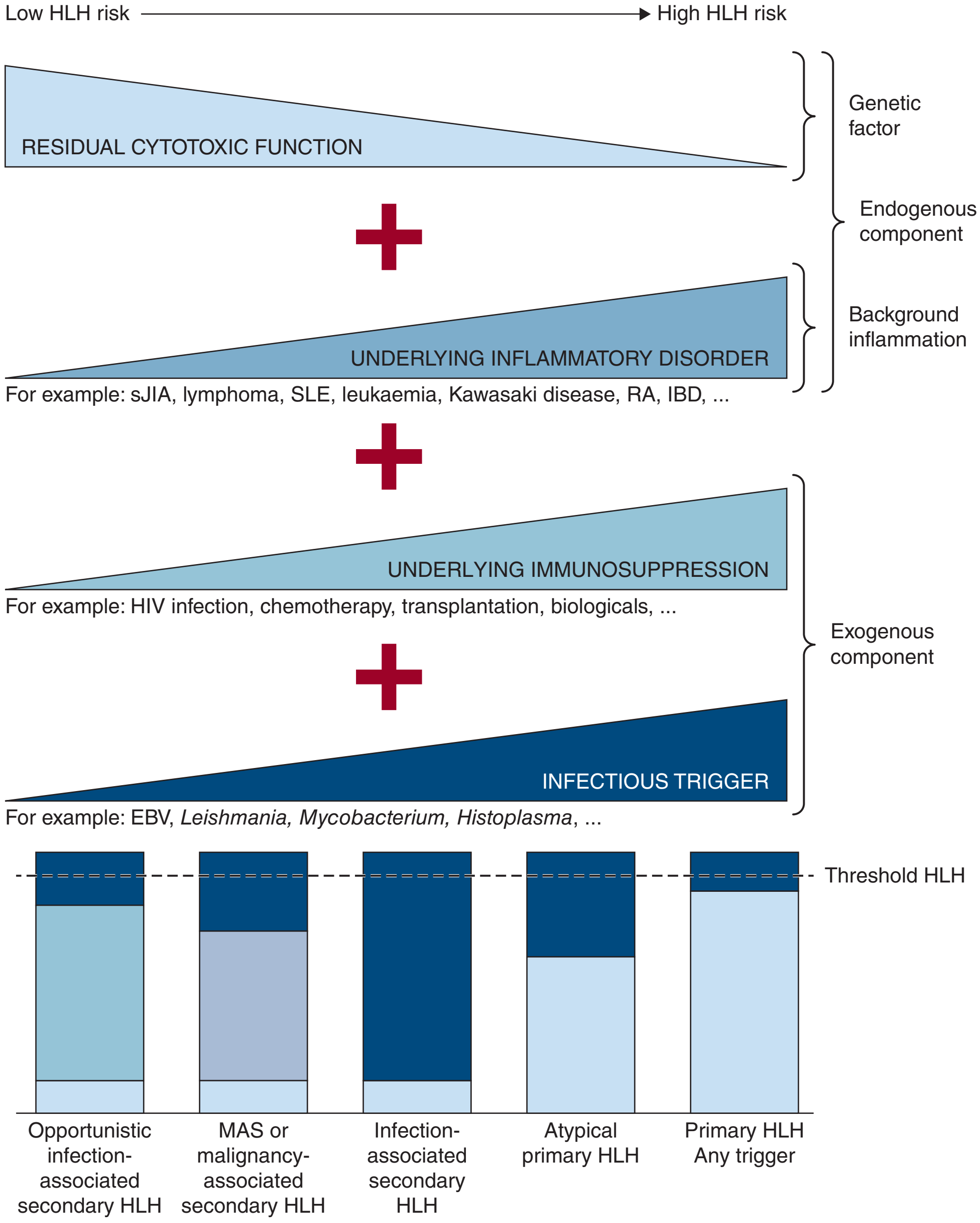
Low HLH risk
High HLH risk
RESIDUAL CYTOTOXIC FUNCTION
Genetic factor
Endogenous component
UNDERLYING INFLAMMATORY DISORDER
Background inflammation
For example: sJIA, lymphoma, SLE, leukaemia, Kawasaki disease, RA, IBD, ...
UNDERLYING IMMUNOSUPPRESSION
For example: HIV infection, chemotherapy, transplantation, biologicals, ...
Exogenous component
INFECTIOUS TRIGGER
For example: EBV, Leishmania, Mycobacterium, Histoplasma, ...
Threshold HLH
Opportunistic infection-associated secondary HLH
MAS or malignancy-associated secondary HLH
Infection-associated secondary HLH
Atypical primary HLH
Primary HLH Any trigger

FIGURE 68-2 The threshold model on the multifactorial pathogenesis of hemophagocytic lymphohistiocytosis (HLH). Genetic factors, background inflammation, underlying immunosuppression, and infectious triggers are superimposed until a certain threshold point where inflammation becomes uncontrolled and HLH develops. EBV, Epstein-Barr virus; IBD, inflammatory bowel disease; RA, rheumatoid arthritis; sJIA, systemic juvenile idiopathic arthritis; SLE, systemic lupus erythematosus. *(Reproduced with permission from E Brisse et al: Advances in the pathogenesis of primary and secondary haemophagocytic lymphohistiocytosis: differences and similarities. Br J Haematology 174:203, 2016.)*

criteria and HScore have good diagnostic accuracy when evaluated in critically ill adults.

A note of caution: The underlying disease associated with secondary HLH itself can sometimes result in several HLH-2004 criteria being fulfilled, which may complicate diagnostic reasoning.

INFECTION-ASSOCIATED HEMOPHAGOCYTIC LYMPHOHISTIOCYTOSIS

Infections are the most common drivers of secondary HLH in both adults and children. Of these, around 70% of cases are associated with viral infections, 20% with bacterial infections, and the remaining 10% with parasitic and fungal infections.

Viral Infections Herpes viruses account for around 60% of reported viral cases in adults, with 40% due to EBV and 10% to cytomegalovirus. EBV is the most common viral trigger in children as well. EBV-HLH is even more frequent in East Asia and in individuals originating from this region. Hepatitis viruses, parvovirus B19, and influenza are among the more common nonherpes viral triggers. The high mortality of about 50% in the avian influenza A H5N1 infection pandemic was associated with findings resembling secondary HLH. In a study of severe dengue fever, about 10% of patients had HLH, half of whom died. A small proportion of patients with COVID-19, caused by SARS-CoV-2, developed secondary HLH.

Other Infections Bacteria associated with HLH include *Mycobacterium tuberculosis*, *Mycoplasma pneumoniae*, *Rickettsia* spp., *Staphylococcus* spp., *Escherichia coli*, and many others. Parasitic HLH is most often caused by visceral leishmaniasis and *Plasmodium malaria*, and fungal HLH by *Histoplasma* spp.

Identifying HLH in critically ill patients is complicated. HLH may be triggered by bacterial septicemia, but because symptoms and signs of sepsis resemble those of HLH, it is difficult to identify the patients with HLH. If a patient with a sepsis-like condition does not respond to adequate sepsis-directed therapy, HLH should be considered.

In a septic-like patient, the diagnosis of HLH is supported by hyperferritinemia, splenomegaly, marked cytopenias, hypofibrinogenemia, hypertriglyceridemia, and a characteristic cytokine profile. The hypercytokinemias in sepsis and HLH have different patterns. In sepsis, IL-6 levels are markedly elevated but IFN-γ levels are moderate, whereas IFN-γ is markedly elevated in HLH in combination with moderate levels of IL-6.

TREATMENT

In all forms of secondary HLH, treatment is based on two main principles: (1) find and treat the underlying cause, which may be surprisingly difficult; and (2) employ HLH-directed therapy adapted to the severity of the HLH.

Studies on EBV-HLH pioneered in Japan showed that a combination of glucocorticoids and etoposide improved survival markedly in secondary HLH. Other immunomodulatory agents used include IL-1, IL-6, IFN-γ, and Janus kinase inhibitors and intravenous immunoglobulin (IVIG). Solid data comparing these agents are lacking. Etoposide appears to the most potent drug but is not always necessary when milder therapy is effective.

Because primary HLH is often triggered by viral infections, consider the possibility of an underlying primary HLH in patients with infection-associated HLH, particularly in children.

EBV-HLH EBV-HLH has a high mortality if untreated, but about half of the patients have a milder disease course. The outcome may be difficult to predict initially, and a severity-stratified treatment approach is recommended. For mild/moderate cases, glucocorticoids alone or glucocorticoids plus IVIG and/or cyclosporin A can be considered, and in case of insufficient response, the addition of etoposide is suggested. In severe disease, the combination of etoposide, dexamethasone, and rituximab (anti-CD20) is suggested. Importantly, when etoposide is used in secondary HLH, treatment is suggested to be less intense than in primary HLH, with less frequent dosing, shorter therapy duration, and lower doses, such as 75–100 mg/m^2 weekly in individuals aged <40 years and 50–75 mg/m^2 weekly in those ≥40 years old, and with weekly decisions on continued therapy.

Chronic active EBV (CAEBV) is a highly progressive disease with organ failure, severe hypercytokinemia, HLH, and overt lymphomatous/leukemic changes. Most affected children and adolescents die within 5–10 years, and adults die earlier. Allogeneic HSCT is the recommended treatment. Rarely, HSCT may also be required as salvage therapy for refractory EBV-HLH.

Non-EBV Viral HLH Most cases of non-EBV HLH are mild to moderate and can be managed by immunomodulatory agents alone, such as glucocorticoids and IVIG. HIV-HLH has a high mortality, and comorbidities such as lymphomas and concomitant opportunistic infections may contribute to the development of HLH. Treatment of severe influenza HLH, such as in the often-fatal avian influenza A H5N1 infection, has not been well studied. In dengue HLH, which may progress extremely rapidly, it is important to be prepared for very rapid escalation of HLH-directed treatment.

Moderate and severe COVID-19 is often associated with increased levels of inflammatory cytokines but only rarely induces full-blown HLH. Nevertheless, immunomodulatory drugs used in HLH are also valuable for COVID-19, including dexamethasone, anakinra (IL-1 inhibition), tocilizumab (IL-6 inhibition), and baricitinib (Janus kinase inhibition).

Bacterial, Parasitic, and Fungal Infections The most common bacteria so far reported to be associated with HLH is *M. tuberculosis*, with the mortality rate reported to be approximately 50%. Tuberculostatic drugs in combination with mild immunomodulatory HLH-directed therapy, mostly glucocorticoids, may be effective.

The most common parasite associated with HLH is *Leishmania*, for which liposomal amphotericin B monotherapy is recommended and effective. HLH associated with fungal infections is predominantly seen in immunocompromised individuals. *Histoplasma* is the most frequently reported fungus to trigger HLH, for which amphotericin B also is effective.

With regard to HLH in critically ill patients with sepsis, knowledge of suitable HLH-directed treatment is still limited. Mild immunomodulation may improve survival in some patients.

MALIGNANCY-ASSOCIATED HEMOPHAGOCYTIC LYMPHOHISTIOCYTOSIS

Malignancy-associated HLH (mal-HLH) is likely underdiagnosed; it is reported to be almost as common as infection-associated HLH in adults, but it is rare in children. It is most frequent in young adult males, affecting about 2.5% of their hematologic malignancies. Mal-HLH is the form of secondary HLH with the worst prognosis, with about 20–30% 2-year survival, mainly owing to poor survival associated with the underlying malignancies. One-month survival, mainly representing the HLH episode, has been reported to be around 70%.

Mal-HLH comes in two forms: "malignancy-triggered HLH," when HLH is diagnosed at the time of cancer diagnosis before initiation of cancer therapy, and HLH that occurs during chemotherapy. The latter often has an infectious trigger and may develop years after initiation of chemotherapy.

Lymphomas are found in about half of all cases, followed by leukemias (30%) and other hematologic malignancies (10%), whereas underlying solid tumors are less frequent (10%). The triggering malignancy may be very difficult to identify. If one does not find the cause of a secondary HLH, searching repeatedly for an underlying (occult) malignancy is prudent. Repeated bone marrow examinations, biopsies, and various radiologic examinations may be required.

A soluble IL-2 receptor alpha/ferritin ratio >2.0 has been reported to have a positive predictive value of 85% that a patient with HLH has lymphoma-associated HLH, whereas a low ratio is more common in patients with non–mal-HLH.

TREATMENT

Although the main principle in secondary HLH is to treat the underlying cause and provide HLH-directed therapy, a complicated question in mal-HLH is what to do first.

It is suggested that if organ damage is HLH-triggered, lympholytic agents should be considered and a two-step approach is suggested: First, target the cytokine storm and T-cell proliferation with moderately dosed etoposide (75–100 mg/m^2), glucocorticoids, and possibly IVIG, and then target the neoplastic disease by specific treatment as soon as organ function has improved to an acceptable degree. Other HLH-directed immunomodulatory agents, such as anakinra, may also be valuable, but results of studies on such therapies are limited.

MACROPHAGE ACTIVATION SYNDROME

Macrophage activation syndrome (MAS) is a life-threatening hyperinflammatory complication of rheumatic disease and other autoimmune diseases. It is characterized by an uncontrolled activation and proliferation of T lymphocytes and macrophages and classified among the secondary, acquired forms of HLH because it shares many clinical and laboratory features with both primary and secondary HLH, hence the term MAS-HLH.

MAS-HLH is the third most common form of HLH in adults and the second most common in children. In children, it occurs most frequently in individuals with systemic juvenile idiopathic arthritis (sJIA), affecting 10% of these patients, and with systemic lupus erythematosus. In adults, systemic lupus erythematosus is the most common cause, followed by adult-onset Still's disease, affecting about 5% and 10–15% of these patient groups, respectively. Other causes include systemic vasculitis and inflammatory bowel disease.

Clinically, MAS-HLH manifests as fever, liver dysfunction, cytopenia, hyperferritinemia, coagulopathy, CNS abnormalities, and, more rarely, hemophagocytosis. Fibrinogen and platelet levels are often higher than in other forms of HLH, due to the inflammatory nature of sJIA. In 2016, MAS-HLH in sJIA patients was defined as a febrile patient with ferritin >684 μg/L and any two of the following: platelet count ≤181 × 10^9/L, aspartate aminotransferase >48 U/L, fasting triglycerides >1.76 mmol/L (156 mg/dL), and fibrinogen ≤3.6 g/L.

Most MAS-HLH flares are reported to be triggered by active disease, but about a third have an infectious trigger. The cytokine pattern in MAS-HLH is characterized by high serum levels of IL-18, distinguishing it from other forms of HLH such as FHL.

The mortality rate in MAS-HLH is about 5–10% in children and 10–15% in adults. CNS involvement is frequent and may lead to irreversible neurologic damage. Early diagnosis and treatment are therefore crucial. Patients with MAS-HLH may also develop severe pulmonary disease with a high fatality rate, reported to be about 50%, for which the best treatment and prevention still is unknown. The predominant pathology is pulmonary alveolar proteinosis and/or endogenous lipoid pneumonia, but the underlying cause is unknown.

TREATMENT

A common first-line approach is glucocorticoids in high doses, such as intravenous pulse methylprednisolone 30 mg/kg per dose up to a maximum of 1000 mg/dose once daily for 3–5 days followed by high-dose oral or intravenous glucocorticoids. Cyclosporin A (2–7 mg/kg/d) can be added. IL-1–blocking therapy is also effective, such as with anakinra in a dose of 2–6 mg/kg up to 10 mg/kg per day in divided doses. Experience with other immunomodulating agents, including tocilizumab, emapalumab, and ruxolitinib, is increasing.

In patients with severe disease or CNS involvement despite glucocorticoids, cyclosporin A, and/or anakinra, a moderate dose of etoposide (50–100 mg/m^2 once weekly) can be very effective.

TRANSPLANT-RELATED AND CHIMERIC ANTIGEN RECEPTOR HEMOPHAGOCYTIC LYMPHOHISTIOCYTOSIS

Other causes of secondary HLH include transplantation, particularly kidney and hematologic transplantations, and novel drugs, such as chimeric antigen receptor (CAR) T cells, bispecific T-cell engagers, and checkpoint inhibitors.

HLH with late onset (>30 days) after HSCT is often comparable to infection-associated HLH. Based on other forms of secondary HLH, it is reasonable to start treatment with corticosteroids. As second-line treatment, favorable response after low-dose etoposide has been reported; one dose of 50–75 mg/m^2 may be sufficient.

Increasing use of CAR T-cell therapy and other immune effector cell-based therapies has led to an increasing number of cases with a clinical picture resembling secondary HLH that is distinct from cytokine release syndrome. This HLH-like complication is more frequent when using CD22 CAR T cells, affecting about a third of these patients. Data on treatment results are limited, but anakinra with or without glucocorticoids is suggested as first-line therapy. As second- and third-line therapy, ruxolitinib, emapalumab, and low-dose etoposide have been suggested.

CONCLUSION

The survival and biological understanding of primary and secondary HLH have increased dramatically over the past decade(s), but much remains to be learned. Despite being life-threatening and now also treatable, HLH is still markedly underdiagnosed. Numerous lives might be saved by increased awareness of HLH.

FURTHER READING

Daver N et al: A consensus review on malignancy-associated hemophagocytic lymphohistiocytosis in adults. Cancer 123:3229, 2017.

Ehl S et al: Recommendations for the use of etoposide-based therapy and bone marrow transplantation for the treatment of HLH: Consensus statements by the HLH Steering Committee of the Histiocyte Society. J Allergy Clin Immunol Pract 6:1508, 2018.

Hines MR et al: Consensus-based guidelines for the recognition, diagnosis, and management of hemophagocytic lymphohistiocytosis in critically ill children and adults. Crit Care Med 50:860, 2022.

La Rosée P et al: Recommendations for the management of hemophagocytic lymphohistiocytosis in adults. Blood 133:2465, 2019.

Ramos-Casals M et al: Adult haemophagocytic syndrome. Lancet 383:1503, 2014.

69

Bleeding and Thrombosis

Barbara A. Konkle

The human hemostatic system provides a natural balance between procoagulant and anticoagulant forces. The procoagulant forces include platelet adhesion and aggregation and fibrin clot formation; anticoagulant forces include the natural inhibitors of coagulation and fibrinolysis. Under normal circumstances, hemostasis is regulated to promote blood flow; however, it is also prepared to clot blood rapidly to arrest blood flow and prevent exsanguination. After bleeding is successfully halted, the system remodels the damaged vessel to restore normal blood flow. The major components of the hemostatic system, which function in concert, are (1) platelets and other formed elements of blood, such as monocytes and red cells; (2) plasma proteins (the coagulation and fibrinolytic factors and inhibitors); and (3) the vessel wall.

STEPS OF NORMAL HEMOSTASIS

PLATELET PLUG FORMATION

On vascular injury, platelets adhere to the site of injury, usually the denuded vascular intimal surface. Platelet adhesion is mediated primarily by von Willebrand factor (VWF), a large multimeric protein present in both plasma and the extracellular matrix of the subendothelial vessel wall, which serves as the primary "molecular glue," providing sufficient strength to withstand the high levels of shear stress that would tend to detach them with the flow of blood. Platelet adhesion is also facilitated by direct binding to subendothelial collagen through specific platelet membrane collagen receptors.

Platelet adhesion results in subsequent platelet activation and aggregation. This process is enhanced and amplified by humoral mediators in plasma (e.g., epinephrine, thrombin); mediators released from activated platelets (e.g., adenosine diphosphate, serotonin); and vessel wall extracellular matrix constituents that come in contact with adherent platelets (e.g., collagen, VWF). Activated platelets undergo the release reaction, during which they secrete contents that further promote aggregation and inhibit the naturally anticoagulant endothelial cell factors. During platelet aggregation (platelet-platelet interaction), additional platelets are recruited from the circulation to the site of vascular injury, leading to the formation of an occlusive platelet thrombus. The platelet plug is anchored and stabilized by the developing fibrin mesh.

The platelet glycoprotein (Gp) IIb/IIIa ($\alpha_{IIb}\beta_3$) complex is the most abundant receptor on the platelet surface. Platelet activation converts the normally inactive Gp IIb/IIIa receptor into an active receptor, enabling binding to fibrinogen and VWF. Because the surface of each platelet has about 50,000 Gp IIb/IIIa–binding sites, numerous activated platelets recruited to the site of vascular injury can rapidly form an occlusive aggregate by means of a dense network of intercellular fibrinogen bridges.

■ FIBRIN CLOT FORMATION

Plasma coagulation proteins (*clotting factors*) normally circulate in plasma in their inactive forms. The sequence of coagulation protein reactions that culminate in the formation of fibrin was originally described as a *waterfall* or a *cascade*. Two pathways of blood coagulation have been described in the past: the so-called extrinsic, or tissue factor, pathway and the so-called intrinsic, or contact activation, pathway. We now know that coagulation is normally initiated through tissue factor (TF) exposure and activation through the classic *extrinsic pathway* but with critically important amplification through elements of the classic *intrinsic pathway*, as illustrated in **Fig. 69-1**. These reactions take place on phospholipid surfaces, usually the activated platelet surface. Coagulation testing in the laboratory can reflect other influences due to the artificial nature of the in vitro systems used (see below).

The immediate trigger for coagulation is vascular damage that exposes blood to TF that is constitutively expressed on the surfaces of subendothelial cellular components of the vessel wall, such as smooth muscle cells and fibroblasts. TF is also present in circulating microparticles, presumably shed from cells including monocytes and platelets. TF binds the serine protease factor VIIa; the complex activates factor X to factor Xa. Alternatively, the complex can indirectly activate factor X by initially converting factor IX to factor IXa, which then activates factor X. The participation of factor XI in hemostasis is not dependent on its activation by factor XIIa but rather on its positive feedback activation by thrombin. Thus, factor XIa functions in the propagation and amplification, rather than in the initiation, of the coagulation cascade. The role of factor XIIa in activation of factor XI is not fully elucidated, but studies suggest it may be a mechanism to promote thrombosis.

Factor Xa can be formed through the actions of either the TF/factor VIIa complex or factor IXa (with factor VIIIa as a cofactor) and converts prothrombin to thrombin, the pivotal protease of the coagulation system. The essential cofactor for this reaction is factor Va, which is produced by thrombin-induced limited proteolysis of factor V. Thrombin is a multifunctional enzyme that converts soluble plasma fibrinogen to an insoluble fibrin matrix. Fibrin polymerization involves an orderly process of intermolecular associations **(Fig. 69-2)**. Thrombin also activates factor XIII (fibrin-stabilizing factor) to factor XIIIa, which covalently cross-links and thereby stabilizes the fibrin clot.

The assembly of the clotting factors on activated cell membrane surfaces greatly accelerates their reaction rates and also serves to localize blood clotting to sites of vascular injury. The critical cell membrane components, acidic phospholipids, are not normally exposed on resting cell membrane surfaces. However, when platelets, monocytes, and endothelial cells are activated by vascular injury or inflammatory stimuli, the procoagulant head groups of the membrane anionic phospholipids become translocated to the surfaces of these cells or released as part of microparticles, making them available to support and promote the plasma coagulation reactions.

ANTITHROMBOTIC MECHANISMS

Several physiologic antithrombotic mechanisms act in concert to prevent clotting under normal circumstances. These mechanisms operate to preserve blood fluidity and to limit blood clotting to specific focal sites of vascular injury. Endothelial cells have many antithrombotic effects. They produce prostacyclin, nitric oxide, and ectoADPase/CD39, which act to inhibit platelet binding, secretion, and aggregation. Endothelial cells produce anticoagulant factors including heparan proteoglycans, TF pathway inhibitor, and thrombomodulin. They also activate fibrinolytic mechanisms through the production of tissue plasminogen activator, urokinase, plasminogen activator inhibitors, and annexin-2.

Antithrombin is the major plasma protease inhibitor of thrombin and other clotting factors in coagulation. Antithrombin neutralizes thrombin and other activated coagulation factors by forming a complex between the active site of the enzyme and the reactive center of antithrombin. The rate of formation of these inactivating complexes increases by a factor of several thousand in the presence of heparin. Antithrombin inactivation of thrombin and other activated clotting factors occurs physiologically on vascular surfaces, where glycosaminoglycans, including heparan sulfates, are present to catalyze these reactions. Inherited quantitative or qualitative deficiencies of antithrombin lead to a lifelong predisposition to venous thromboembolism (VTE).

Protein C is a plasma glycoprotein that becomes an anticoagulant when it is activated by thrombin. The thrombin-induced activation of protein C occurs physiologically on thrombomodulin, a transmembrane proteoglycan-binding site

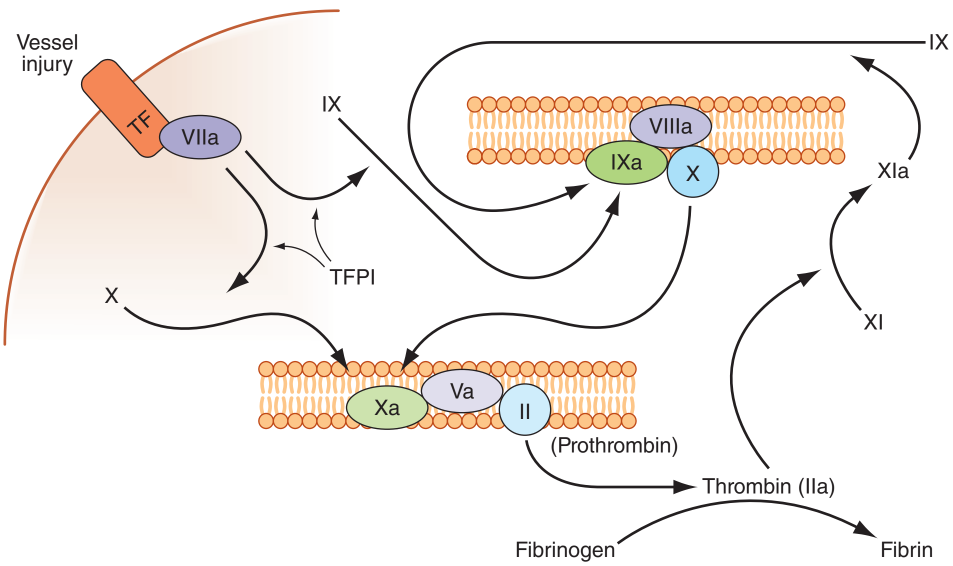

FIGURE 69-1 Coagulation is initiated by tissue factor (TF) exposure, which, with factor (F) VIIa, activates FIX and FX, which in turn, with FVIII and FV as cofactors, respectively, results in thrombin formation and subsequent conversion of fibrinogen to fibrin. Thrombin activates FXI, FVIII, and FV, amplifying the coagulation signal. Once the TF/FVIIa/FXa complex is formed, tissue factor pathway inhibitor (TFPI) inhibits the TF/FVIIa pathway, making coagulation dependent on the amplification loop through FIX/FVIII. Coagulation requires calcium (not shown) and takes place on phospholipid surfaces, usually the activated platelet membrane.

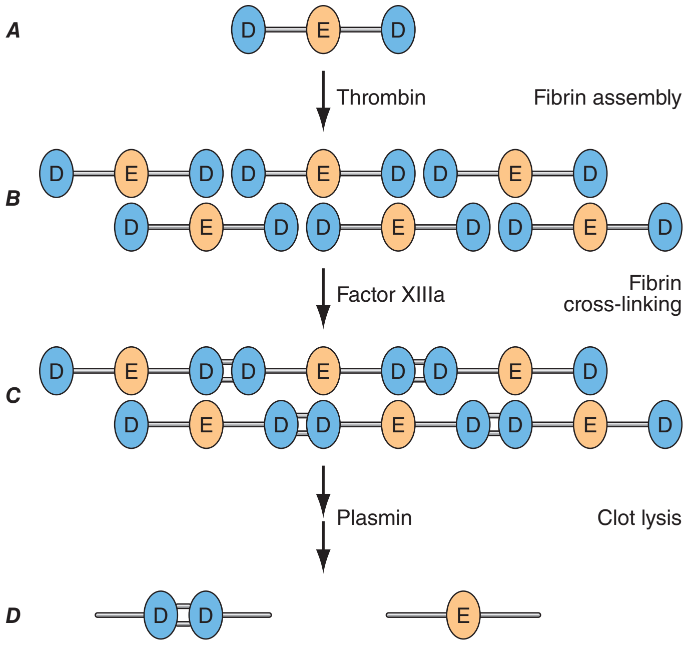

FIGURE 69-2 Fibrin formation and dissolution. (***A***) Fibrinogen is a trinodular structure consisting of two D domains and one E domain. Thrombin activation results in an ordered lateral assembly of protofibrils (***B***) with noncovalent associations. Factor XIIIa cross-links the D domains on adjacent molecules (***C***). Fibrin and fibrinogen (not shown) lysis by plasmin occurs at discrete sites and results in intermediary fibrin(ogen) degradation products (not shown). D-Dimers are the product of complete lysis of fibrin (***D***), maintaining the cross-linked D domains.

for thrombin on endothelial cell surfaces. The binding of protein C to its receptor on endothelial cells places it in proximity to the thrombin-thrombomodulin complex, thereby enhancing its activation efficiency. **(See Fig. 69-3.)** Activated protein C acts as an anticoagulant by cleaving and inactivating activated factors V and VIII. This reaction is accelerated by a cofactor, protein S, which, like protein C, is a glycoprotein that undergoes vitamin K–dependent posttranslational modification. Quantitative or qualitative deficiencies of protein C or protein S, or resistance to the action of activated protein C by a specific variant at its target cleavage site in factor Va (factor V Leiden), lead to hypercoagulable states.

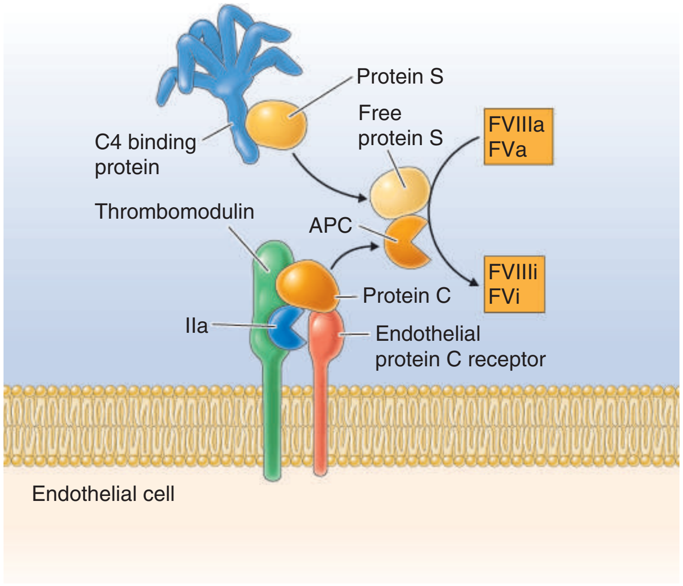

FIGURE 69-3 The activated protein C pathway in regulation of thrombosis. Thrombin generation results in protein C activation through interaction with thrombomodulin and protein C bound to the endothelial protein C receptor (EPCR). Activated protein C (APC) with free protein S converts activated factors (F) VIII and V to inactivate forms, thus in turn decreasing thrombin generation. C4BP, C4 binding protein; EC, endothelial cell; F, factor; IIa, thrombin; PC, protein C; PS, protein S; TM, thrombomodulin.

Tissue factor pathway inhibitor (TFPI) is a plasma protease inhibitor that regulates the TF-induced extrinsic pathway of coagulation. TFPI inhibits the TF/factor VIIa/factor Xa complex, essentially turning off the TF/factor VIIa initiation of coagulation, which then becomes dependent on the "amplification loop" via factor XI and factor VIII activation by thrombin. TFPI is bound to lipoprotein and can also be released by heparin from endothelial cells, where it is bound to glycosaminoglycans, and from platelets. The heparin-mediated release of TFPI may play a role in the anticoagulant effects of unfractionated and low-molecular-weight heparins.

■ THE FIBRINOLYTIC SYSTEM

Any thrombin that escapes the inhibitory effects of the physiologic anticoagulant systems is available to convert fibrinogen to fibrin. In response, the endogenous fibrinolytic system is then activated to dispose of intravascular fibrin and thereby maintain or reestablish the patency of the circulation. Just as thrombin is the key protease enzyme of the coagulation system, plasmin is the major protease enzyme of the fibrinolytic system, acting to digest fibrin to fibrin degradation products. The general scheme of fibrinolysis and its control is shown in **Fig. 69-4**.

The plasminogen activators, tissue type plasminogen activator (tPA) and the urokinase-type plasminogen activator (uPA), cleave the Arg560-Val561 bond of plasminogen to generate the active enzyme plasmin. The lysine-binding sites of plasmin (and plasminogen) permit it to bind to fibrin, so that physiologic fibrinolysis is "fibrin specific." Both plasminogen (through its lysine-binding sites) and tPA possess specific affinity for fibrin and thereby bind selectively to clots. The assembly of a ternary complex, consisting of fibrin, plasminogen, and tPA, promotes the localized interaction between plasminogen and tPA and greatly accelerates the rate of plasminogen activation to plasmin. Moreover, partial degradation of fibrin by plasmin exposes new plasminogen and tPA-binding sites in carboxy-terminus lysine residues of fibrin fragments to enhance these reactions further. This creates a highly efficient mechanism to generate plasmin locally on the fibrin clot, which then becomes plasmin's substrate for digestion to fibrin degradation products.

Plasmin cleaves fibrin at distinct sites of the fibrin molecule, leading to the generation of characteristic fibrin fragments during the process of fibrinolysis (Fig. 69-2). The sites of plasmin cleavage of fibrin are the same as those in fibrinogen. However, when plasmin acts on covalently cross-linked fibrin, D-dimers are released; hence, D-dimers

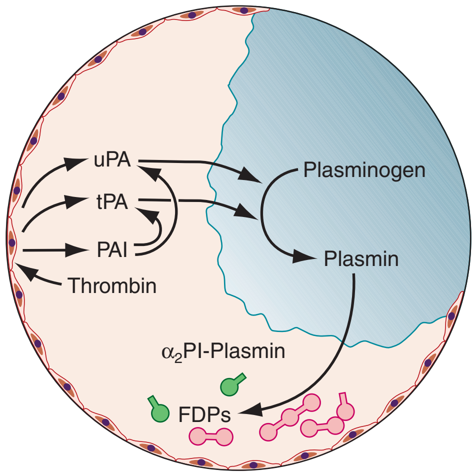

FIGURE 69-4 A schematic diagram of the fibrinolytic system. Tissue plasminogen activator (tPA) is released from endothelial cells, binds the fibrin clot, and activates plasminogen to plasmin. Release of plasminogen activator inhibitors (PAI-1 and PAI-2) inhibits tPA and urokinase (uPA). Excess fibrin is degraded by plasmin to distinct degradation products [FDPs (D-dimers)]. Any free plasmin is complexed with α_2-antiplasmin (α_2PI). PAI, plasminogen activator inhibitor; uPA, urokinase-type plasminogen activator.

can be measured in plasma as a relatively specific test of fibrin (rather than fibrinogen) degradation. D-Dimer assays can be used as sensitive markers of blood clot formation and have been validated for clinical use to exclude the diagnosis of deep venous thrombosis (DVT) and pulmonary embolism in selected populations. D-Dimer levels increase with age. Use of an age-adjusted D-dimer threshold for risk stratification results in less additional testing for VTE.

Physiologic regulation of fibrinolysis occurs primarily at three levels: (1) plasminogen activator inhibitors (PAIs), specifically PAI-1 and PAI-2, inhibit the physiologic plasminogen activators; (2) the thrombin-activatable fibrinolysis inhibitor (TAFI) limits fibrinolysis; and (3) α_2-antiplasmin inhibits plasmin. PAI-1 is the primary inhibitor of tPA and uPA in plasma. TAFI cleaves the N-terminal lysine residues of fibrin, which aid in localization of plasmin activity. α_2-Antiplasmin is the main inhibitor of plasmin in human plasma, inactivating any nonfibrin clot–associated plasmin.

APPROACH TO THE PATIENT

Bleeding and Thrombosis

CLINICAL PRESENTATION

Disorders of hemostasis may be either inherited or acquired. A detailed personal and family history is key in determining the chronicity of symptoms and the likelihood of the disorder being inherited, as well as providing clues to underlying conditions that have contributed to the bleeding or thrombotic state. In addition, the history can give clues as to the etiology by determining (1) the bleeding (mucosal and/or joint) or thrombosis (arterial and/or venous) site and (2) whether an underlying bleeding or clotting tendency was enhanced by another medical condition or the introduction of medications or dietary supplements.

History of Bleeding A history of bleeding is the most important predictor of bleeding risk. In evaluating a patient for a bleeding disorder, a history of at-risk situations, including the response to past surgeries, should be assessed. Does the patient have a history of spontaneous or trauma/surgery-induced bleeding? Spontaneous hemarthroses are a hallmark of moderate and severe factor VIII and IX deficiency and, in rare circumstances, of other clotting factor deficiencies. Mucosal bleeding symptoms are more suggestive of underlying platelet disorders or von Willebrand disease (VWD), termed *disorders of primary hemostasis or platelet plug formation.* Disorders affecting primary hemostasis are shown in **Table 69-1**.

A bleeding score has been validated as a tool to predict patients more likely to have an inherited bleeding disorder, particularly type 1 VWD (International Society on Thrombosis and Haemostasis Bleeding Assessment Tool [*www.isth.org/resource/resmgr/ssc/isth-ssc_bleeding_assessment.pdf*]), and a self-administered form has been validated. This is the most useful tool in excluding the diagnosis of a bleeding disorder, thus avoiding unnecessary testing, and is recommended by 2021 guidelines for screening for VWD in primary care. Bleeding symptoms that are more common in patients with bleeding disorders include prolonged bleeding with surgery, dental procedures and extractions, and/or trauma; heavy menstrual bleeding or postpartum hemorrhage; and large bruises (often described with lumps).

Easy bruising and heavy menstrual bleeding are common complaints in patients with and without bleeding disorders. Easy bruising can also be a sign of medical conditions in which there is no identifiable coagulopathy; instead, the conditions are caused by an abnormality of blood vessels or their supporting connective tissues. In Ehlers-Danlos syndrome, there may be posttraumatic bleeding and a history of joint hyperextensibility. Cushing's syndrome, chronic steroid use, and aging result in changes in skin and subcutaneous tissue, and subcutaneous bleeding occurs in response to minor trauma. The latter has been termed *senile purpura.*

TABLE 69-1 Primary Hemostatic (Platelet Plug) Disorders

Defects of Platelet Adhesion
von Willebrand disease
Bernard-Soulier syndrome (absence or dysfunction of platelet Gp Ib-IX-V)
Defects of Platelet Aggregation
Glanzmann's thrombasthenia (absence or dysfunction of platelet glycoprotein [Gp] IIb/IIIa)
Afibrinogenemia
Defects of Platelet Secretion
Decreased cyclooxygenase activity
Drug-induced (aspirin, nonsteroidal anti-inflammatory agents, thienopyridines)
Inherited
Granule storage pool defects
Inherited
Acquired
Nonspecific inherited secretory defects
Nonspecific drug effects
Uremia
Platelet coating (e.g., paraprotein, penicillin)
Defect of Platelet Coagulant Activity
Scott's syndrome

Epistaxis is a common symptom, particularly in children and in dry climates, and may not reflect an underlying bleeding disorder. However, it is the most common symptom in hereditary hemorrhagic telangiectasia and in boys with VWD. Clues that epistaxis is a symptom of an underlying bleeding disorder include lack of seasonal variation and bleeding that requires medical evaluation or treatment, including cauterization. Bleeding with eruption of primary teeth is seen in children with more severe bleeding disorders, such as moderate and severe hemophilia. It is uncommon in children with mild bleeding disorders. Patients with disorders of primary hemostasis (platelet adhesion) may have increased bleeding after dental cleanings and other procedures that involve gum manipulation.

Heavy menstrual bleeding is defined quantitatively as a loss of >80 mL of blood per cycle, based on the quantity of blood loss required to produce iron-deficiency anemia. A complaint of heavy menses is subjective and has a poor correlation with excessive blood loss. Predictors of heavy menstrual bleeding include bleeding resulting in iron-deficiency anemia or a need for blood transfusion, passage of clots >1 inch in diameter, and changing a pad or tampon more than hourly. Heavy menstrual bleeding is a common symptom in women with underlying bleeding disorders and is reported in the majority of women with VWD, factor XI deficiency, platelet function disorders, and hemophilia, including genetic carriers with borderline-normal factor levels. Women with underlying bleeding disorders are more likely to have other bleeding symptoms, including bleeding after dental extractions and postoperative and postpartum bleeding, and are much more likely to have heavy menstrual bleeding beginning at menarche than women with heavy menstrual bleeding due to other causes. Heavy menstrual bleeding may result in iron deficiency and is documented to have significant adverse effects on quality of life.

Postpartum hemorrhage is a common symptom in women with underlying bleeding disorders. In women with type 1 VWD or hemophilia A in whom levels of VWF and factor VIII often normalize during pregnancy, postpartum hemorrhage may be delayed. Women with a history of postpartum hemorrhage may have a higher risk of recurrence with subsequent pregnancies. Women with underlying bleeding disorders are at risk for other reproductive tract bleeding, including rupture of ovarian cysts with intraabdominal hemorrhage.

Tonsillectomy is a major hemostatic challenge, because intact hemostatic mechanisms are essential to prevent excessive bleeding from the tonsillar bed. Bleeding may occur early after surgery or after approximately 7 days postoperatively, with loss of the eschar at the operative site. Similar delayed bleeding is seen after colonic polyp resection. Gastrointestinal (GI) bleeding and hematuria are usually due to underlying pathology, and procedures to identify and treat the bleeding site should be undertaken, even in patients with known bleeding disorders. VWD, particularly types 2 and 3, is associated with angiodysplasia of the bowel and GI bleeding.

Hemarthroses and spontaneous muscle hematomas are characteristic of moderate or severe congenital factor VIII or IX deficiency. They can also be seen in moderate and severe deficiencies of fibrinogen, prothrombin, and factors V, VII, and X. Spontaneous hemarthroses occur rarely in other bleeding disorders except for severe VWD, with associated very low factor VIII levels. Muscle and soft tissue bleeds are also common in acquired factor VIII deficiency. Bleeding into a joint results in severe pain and swelling, as well as loss of function, but is rarely associated with discoloration from bruising around the joint. Life-threatening sites of bleeding include bleeding into the oropharynx, where bleeding can obstruct the airway, into the central nervous system, and into the retroperitoneum. Central nervous system bleeding is the major cause of bleeding-related deaths in patients with severe congenital factor deficiencies.

Prohemorrhagic Effects of Medications and Dietary Supplements Aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) that inhibit cyclooxygenase 1 impair primary hemostasis and may exacerbate bleeding from another cause or even unmask a previously occult mild bleeding disorder such as VWD. All NSAIDs, however, can precipitate GI bleeding, which may be more severe in patients with underlying bleeding disorders. The aspirin effect on platelet function lasts for the life of the platelet; however, in individuals with typical platelet turnover, the functional defect reverts to near-normal within 2–3 days after the last dose. The effect of other NSAIDs is shorter, as the inhibitor effect is reversed when the drug is removed. Inhibitors of the ADP P2Y$_{12}$ receptor (clopidogrel, prasugrel, and ticagrelor) inhibit ADP-mediated platelet aggregation and, like NSAIDs, can precipitate or exacerbate bleeding symptoms. The risk of bleeding with these drugs is higher than with NSAIDs.

Many herbal supplements can impair hemostatic function. Some are more convincingly associated with a bleeding risk than others. Fish oil or concentrated omega-3 fatty acid supplements impair platelet function. They alter platelet biochemistry to produce more PGI$_3$, a more potent platelet inhibitor than prostacyclin (PGI$_2$), and more thromboxane A$_3$, a less potent platelet activator than thromboxane A$_2$. In fact, diets naturally rich in omega-3 fatty acids can result in a prolonged bleeding time and abnormal platelet aggregation studies, but the actual associated bleeding risk is unclear. Many supplements have been associated with increased bleeding with surgery and anticoagulant-related bleeding. In patients with unexplained bruising or bleeding, it is prudent to review any new medications or supplements and discontinue those that have been associated with bleeding.

Underlying Systemic Diseases That Cause or Exacerbate a Bleeding Tendency Acquired bleeding disorders are commonly secondary to, or associated with, systemic disease. The clinical evaluation of a patient with a bleeding tendency must therefore include a thorough assessment for evidence of underlying disease. Bruising or mucosal bleeding may be the presenting complaint in liver disease, severe renal impairment, hypothyroidism, paraproteinemias or amyloidosis, and conditions causing bone marrow failure. All coagulation factors are synthesized in the liver, and hepatic failure results in combined factor deficiencies. This is often compounded by thrombocytopenia and portal hypertension. Coagulation factors II, VII, IX, and X and proteins C, S, and Z are dependent on vitamin K for posttranslational modification. Although vitamin K is required in both procoagulant and anticoagulant processes, the phenotype of vitamin K deficiency or the warfarin effect on coagulation is bleeding.

The normal blood platelet count is 150,000–450,000/μL. Thrombocytopenia results from decreased production, increased destruction, and/or sequestration. Although the bleeding risk varies somewhat by the reason for the thrombocytopenia, bleeding rarely occurs in isolated thrombocytopenia at counts >50,000/μL and usually not until <10,000–20,000/μL. Coexisting coagulopathies, as is seen in liver failure or disseminated coagulation; infection; platelet-inhibitory drugs; and underlying medical conditions can all increase the risk of bleeding in the thrombocytopenic patient. Most procedures can be performed in patients with a platelet count of 50,000/μL or greater.

TABLE 69-2 Some Risk Factors for Thrombosis

VENOUS	VENOUS AND ARTERIAL
Inherited	**Inherited**
Factor V Leiden	Homocystinuria
Prothrombin G20210A	Dysfibrinogenemia
Antithrombin deficiency	**Acquired**
Protein C deficiency	Malignancy
Protein S deficiency	Antiphospholipid antibody syndrome
Acquired	Hormonal therapy
Age	Polycythemia vera
Previous thrombosis	Essential thrombocythemia
Immobilization	Paroxysmal nocturnal hemoglobinuria
Major surgery	Thrombotic thrombocytopenic purpura
Pregnancy and puerperium	Heparin-induced thrombocytopenia
Hospitalization	Disseminated intravascular coagulation
Obesity	Infection
Infection	**Unknown**[a]
Smoking	Elevated factor II, VIII, IX, XI
	Elevated TAFI levels
	Low levels of TFPI

[a] Unknown whether risk is inherited or acquired.

Abbreviations: APC, activated protein C; TAFI, thrombin-activatable fibrinolysis inhibitor; TFPI, tissue factor pathway inhibitor.

HISTORY OF THROMBOSIS

The risk of thrombosis, like that of bleeding, is influenced by both genetic and environmental factors. The major risk factor for arterial thrombosis is atherosclerosis, whereas for venous thrombosis, the risk factors are immobility, surgery, underlying medical conditions such as malignancy, medications such as hormonal therapy, obesity, and genetic predispositions. Factors that increase risks for venous and for both venous and arterial thromboses are shown in **Table 69-2**.

The most important point in a history related to venous thrombosis is determining whether the thrombotic event was idiopathic (meaning there was no clear precipitating factor) or was a precipitated event. In patients without underlying malignancy, having an idiopathic event is the strongest predictor of recurrence of VTE. In patients who have a vague history of thrombosis, a history of being treated with warfarin or other anticoagulants suggests a past DVT. Age is an important risk factor for venous thrombosis—the risk of DVT increases per decade, with an approximate incidence of 1/100,000 per year in early childhood to 1/200 per year among octogenarians. Family history is helpful in determining if there is a genetic predisposition and how strong that predisposition appears to be. A genetic thrombophilia that confers a relatively small increased risk, such as being a heterozygote for the prothrombin G20210A or factor V Leiden mutation, is a minor determinant of risk in an elderly individual undergoing a high-risk surgical procedure. As illustrated in **Fig. 69-5**, a thrombotic event usually has more than one contributing factor. Predisposing factors must be carefully assessed to determine the risk of recurrent thrombosis and, with consideration of the patient's bleeding risk, determine the length of anticoagulation. Testing for inherited thrombophilias in adults should be limited to instances where results would change clinical care. Such instances are rare.

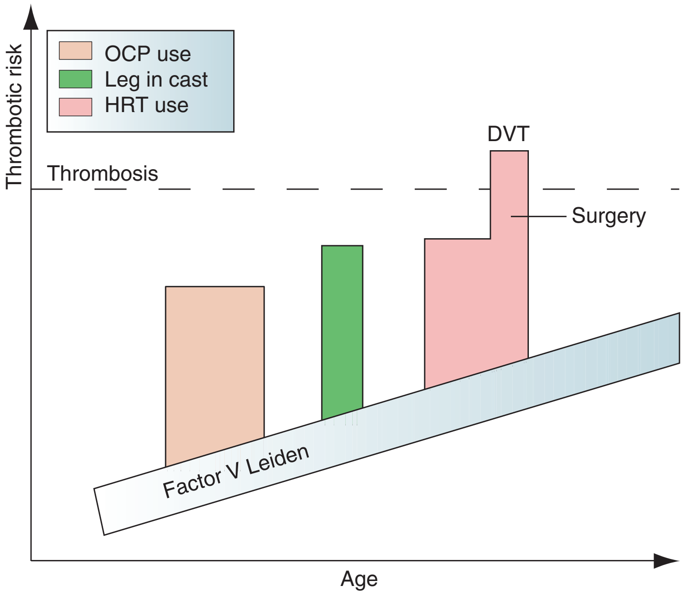

FIGURE 69-5 Thrombotic risk over time. Shown schematically is an individual's thrombotic risk over time. An underlying factor V Leiden variant provides a "theoretically" constant increased risk. The thrombotic risk increases with age and, intermittently, with oral contraceptive (OCP) or oral hormone replacement therapy (HRT) use; other events, like major surgery or illness, will increase the risk further. At some point, the cumulative risk may increase to the threshold for thrombosis and result in deep venous thrombosis (DVT). Note: The magnitude and duration of risk portrayed in the figure are meant for example only and may not precisely reflect the relative risk determined by clinical study. *(Sources: From BA Konkle, A Schafer, in DP Zipes et al [eds]: Braunwald's Heart Disease, 7th ed. Philadelphia, Saunders, 2005; from FR Rosendaal: Venous thrombosis: A multicausal disease. Lancet 353:1167, 1999.)*

LABORATORY EVALUATION

Careful history taking and clinical examination are essential components in the assessment of bleeding and thrombotic risk. The use of laboratory tests of coagulation complements, but cannot substitute for, clinical assessment. No test exists that provides a global assessment of hemostasis. Thrombin generation assays have not generally provided reproducible results across laboratories. The bleeding time does not predict bleeding risk, and it is not recommended for this indication. Thromboelastography can be useful in guiding intraoperative transfusion and is being explored in other settings but is not broadly applicable for the diagnosis of disorders of hemostasis and thrombosis.

For routine preoperative and preprocedure testing, an abnormal prothrombin time (PT) may detect liver disease or vitamin K deficiency that had not been previously appreciated. Studies have not confirmed the usefulness of an activated partial thromboplastin time (aPTT) in preoperative evaluations in patients with a negative bleeding history. The primary use of coagulation testing should be to confirm the presence and type of bleeding disorder in a patient with a suspicious clinical history.

Because of the nature of coagulation assays, proper sample acquisition and handling is critical to obtaining valid results. In patients with abnormal coagulation assays who have no bleeding history, repeat studies with attention to these factors frequently results in normal values. Most coagulation assays are performed in sodium citrate anticoagulated plasma that is recalcified for the assay. Because the anticoagulant is in liquid solution and needs to be added to blood in proportion to the plasma volume, incorrectly filled or inadequately mixed blood collection tubes will give erroneous results. These vacutainer tubes should be filled to >90% of the recommended fill, which is usually denoted by a line on the tube. An elevated hematocrit (>55%) can result in a false value due to a decreased plasma-to-anticoagulant ratio.

Screening Assays The most commonly used screening tests are the PT, aPTT, and platelet count. The PT assesses the factors I (fibrinogen), II (prothrombin), V, VII, and X **(Fig. 69-6)**. The PT measures the time for clot formation of the citrated plasma after recalcification and addition of thromboplastin, a mixture of TF and phospholipids. The sensitivity of the assay varies by the source of thromboplastin. The relationship between defects in secondary hemostasis (fibrin formation) and coagulation test abnormalities is shown in **Table 69-3**. To adjust for this variability, the overall sensitivity of different thromboplastins to reduction of the vitamin K–dependent clotting factors II, VII, IX, and X in anticoagulation patients is expressed as the International Sensitivity Index (ISI). The international normalized ratio (INR) is determined based on the formula: INR = $(PT_{patient}/PT_{normal\ mean})^{ISI}$.

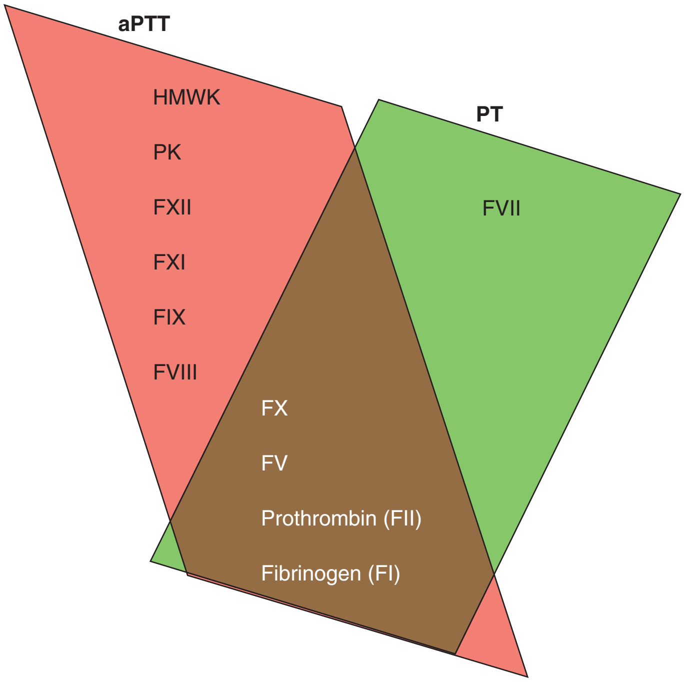

FIGURE 69-6 Coagulation factor activity tested in the activated partial thromboplastin time (aPTT) in red and prothrombin time (PT) in green, or both. F, factor; HMWK, high-molecular-weight kininogen; PK, prekallikrein.

The INR was developed to assess stable anticoagulation due to reduction of vitamin K–dependent coagulation factors; it is commonly used in the evaluation of patients with liver disease. Although it does allow comparison between laboratories, reagent sensitivity as used to determine the ISI is not the same in liver disease as with warfarin anticoagulation. In addition, progressive liver failure is associated with variable changes in coagulation factors; the degree of prolongation of either the PT or the INR only roughly predicts the bleeding risk. Thrombin generation has been shown to be normal in many patients with mild to moderate liver dysfunction. Because the PT only measures one aspect of hemostasis affected by liver dysfunction, we likely overestimate the bleeding risk of a mildly elevated INR in this setting. PT reagents have variable sensitivity to the direct Xa inhibitors, and the PT is usually normal in patients on apixaban.

The aPTT assesses the intrinsic and common coagulation pathways; factors XI, IX, VIII, X, V, and II; fibrinogen; prekallikrein; high-molecular-weight kininogen; and factor XII (Fig. 69-6). The aPTT reagent contains phospholipids derived from either animal or vegetable sources that function as a platelet substitute in the coagulation pathways and includes an activator of the intrinsic coagulation system, such as nonparticulate ellagic acid or the particulate activators kaolin, celite, or micronized silica.

The phospholipid composition of aPTT reagents varies, which influences the sensitivity of individual reagents to clotting factor deficiencies and to inhibitors such as heparin and lupus anticoagulants. Thus, aPTT results will vary from one laboratory to another, and the normal range in the laboratory where the testing occurs

TABLE 69-3 Hemostatic Disorders and Coagulation Test Abnormalities

Prolonged Activated Partial Thromboplastin Time (aPTT)
No clinical bleeding—↓ factor XII, high-molecular-weight kininogen, prekallikrein
Variable, but usually mild, bleeding—↓ factor XI, mild ↓ factor VIII and factor IX
Frequent, severe bleeding—severe deficiencies of factors VIII and IX
Heparin and direct thrombin inhibitors
Prolonged Prothrombin Time (PT)
Factor VII deficiency
Vitamin K deficiency—early
Warfarin anticoagulation
Direct Xa inhibitors (rivaroxaban, edoxaban, apixaban—note PT may be normal)
Prolonged aPTT and PT
Factor II, V, X, or fibrinogen deficiency
Vitamin K deficiency—late
Direct thrombin inhibitors
Prolonged Thrombin Time
Heparin or heparin-like inhibitors
Direct thrombin inhibitors (e.g., dabigatran, argatroban, bivalirudin)
Mild or no bleeding—dysfibrinogenemia
Frequent, severe bleeding—afibrinogenemia
Prolonged PT and/or aPTT Not Corrected with Mixing with Normal Plasma
Bleeding—specific factor inhibitor
No symptoms, or clotting and/or pregnancy loss—lupus anticoagulant
Disseminated intravascular coagulation
Heparin or direct thrombin inhibitor
Abnormal Clot Solubility
Factor XIII deficiency
Inhibitors or defective cross-linking
Rapid Clot Lysis
Deficiency of α_2-antiplasmin or plasminogen activator inhibitor 1
Treatment with fibrinolytic therapy

should be used in the interpretation. Local laboratories can relate their aPTT values to the therapeutic heparin anticoagulation by correlating aPTT values with direct measurements of heparin activity (anti-Xa or protamine titration assays) in samples from heparinized patients, although correlation between these assays is often poor. The aPTT reagent will vary in sensitivity to individual factor deficiencies and usually becomes prolonged with individual factor deficiencies of ≤30–50%.

Mixing Studies Mixing studies are used to evaluate a prolonged aPTT or, less commonly, PT to distinguish between a factor deficiency and an inhibitor. In this assay, normal plasma and patient plasma are mixed in a 1:1 ratio, and the aPTT or PT is determined immediately and after incubation at 37°C for varying times, typically 30, 60, and/or 120 min. With isolated factor deficiencies, the aPTT will correct with mixing and stay corrected with incubation. With aPTT prolongation due to a lupus anticoagulant, the mixing and incubation will show no correction. In acquired neutralizing factor antibodies, notably an acquired factor VIII inhibitor, the initial assay may or may not correct immediately after mixing but will prolong or remain prolonged with incubation at 37°C. Failure to correct with mixing can also be due to the presence of other inhibitors or interfering substances such as heparin, fibrin split products, and paraproteins.

Specific Factor Assays Decisions to proceed with specific clotting factor assays will be influenced by the clinical situation and the results of coagulation screening tests. Precise diagnosis and effective management of inherited and acquired coagulation deficiencies necessitate quantitation of the relevant factors. When bleeding is severe, specific assays are urgently required to guide appropriate therapy. Individual factor assays are usually performed as modifications of the mixing study, where the patient's plasma is mixed with plasma deficient in the factor being studied. This will correct all factor deficiencies to >50%, thus making prolongation of clot formation due to a factor deficiency dependent on the factor missing from the added plasma. Chromogenic assays may also be used.

Testing for Antiphospholipid Antibodies Antibodies to phospholipids (cardiolipin) or phospholipid-binding proteins (β_2-microglobulin and others) are detected by enzyme-linked immunosorbent assay (ELISA). When these antibodies interfere with phospholipid-dependent coagulation tests, they are termed *lupus anticoagulants*. The aPTT has variability sensitivity to lupus anticoagulants, depending in part on the aPTT reagents used. An assay using a sensitive reagent has been termed an *LA-PTT*. The dilute Russell viper venom test (dRVVT) is a modification of a standard test with the phospholipid reagent decreased, thus increasing the sensitivity to antibodies that interfere with the phospholipid component. These tests, however, are not specific for lupus anticoagulants, because factor deficiencies or other inhibitors will also result in prolongation. Documentation of a lupus anticoagulant requires not only prolongation of a phospholipid-dependent coagulation test but also lack of correction when mixed with normal plasma and correction with the addition of activated platelet membranes or certain phospholipids (e.g., hexagonal phase).

Other Coagulation Tests The thrombin time and the reptilase time measure fibrinogen conversion to fibrin and are prolonged when the fibrinogen level is low (usually <80–100 mg/dL) or qualitatively abnormal, as seen in inherited or acquired dysfibrinogenemias, or when fibrin/fibrinogen degradation products interfere. The thrombin time, but not the reptilase time, is prolonged in the presence of heparin. The thrombin time is markedly prolonged in the presence of the direct thrombin inhibitor, dabigatran; a dilute thrombin time is used to assess drug activity. Measurement of anti–factor Xa plasma inhibitory activity is a test frequently used to assess low-molecular-weight heparin (LMWH) levels, as a direct measurement of unfractionated heparin (UFH) activity, or to assess activity of the direct Xa inhibitors rivaroxaban, apixaban, and edoxaban. Drug in the patient sample inhibits the enzymatic conversion of an Xa-specific chromogenic substrate to colored product by factor Xa. Standard curves are created using multiple concentrations of the specific drug and are used to calculate the concentration of anti-Xa activity in the patient plasma.

Laboratory Testing for Thrombophilia Laboratory assays to detect thrombophilic states include molecular diagnostics and immunologic and functional assays. These assays vary in their sensitivity and specificity for the condition being tested. Furthermore, acute thrombosis, acute illnesses, inflammatory conditions, pregnancy, and medications affect levels of many coagulation factors and their inhibitors. Antithrombin is decreased by heparin and in the setting of acute thrombosis. Protein C and S levels may be increased in the setting of acute thrombosis and are decreased by warfarin. Antiphospholipid antibodies are frequently transiently positive in acute illness. Testing for genetic thrombophilias should, in general, only be performed when there is a strong family history of thrombosis and results would affect clinical decision-making.

Because thrombophilia evaluations are usually performed to assess the need to extend anticoagulation, testing, if indicated, should be performed in a steady state, remote from the acute event. Functional assays, but not genetic assays, will be affected by anticoagulants including warfarin (for vitamin K–dependent proteins) and thrombin and Xa inhibitors and cannot be interpreted in patients on those drugs. In most instances, when discontinuation of anticoagulation is being considered, drugs can be stopped after the initial 3–6 months of treatment, and testing can be performed at least 3 weeks later.

Measures of Platelet Function The bleeding time was used in the past to assess platelet function but is nonspecific and it is not recommended for use for this indication. The PFA-100 and similar instruments that measure platelet-dependent coagulation under flow conditions are generally more sensitive and specific for platelet disorders and VWD than the bleeding time; however, data are insufficient to support their use to predict bleeding risk or monitor response to therapy, and they will be normal in some patients with platelet disorders or mild VWD. When they are used in the evaluation of a patient with bleeding symptoms, abnormal results require specific testing, such as VWF assays and/or platelet aggregation studies. Because all of these "screening" assays may miss patients with mild bleeding disorders, further studies are needed to define their role in hemostasis testing.

For classic platelet aggregometry, various agonists are added to the patient's platelet-rich plasma or whole blood, and platelet aggregation is measured. Tests of platelet secretion in response to agonists can also be measured. These remain the gold standard for diagnosis of platelet function disorders. However, they are affected by many factors, including numerous medications, and the association between minor defects in these assays and bleeding risk is not clearly established.

FURTHER READING

Baker P et al: Guidelines on the laboratory aspects of assays used in haemostasis and thrombosis. Br J Haematol 191:347, 2020.

Connors JM: Thrombophilia testing and venous thrombosis. N Engl J Med 377:12, 2017.

Darzi AJ et al: Prognostic factors for VTE and bleeding in hospitalized medical patients: A systematic review and meta-analysis. Blood 135:1788, 2020.

Devreese KMJ et al: Guidance from the Scientific and Standardization Committee for lupus anticoagulant/antiphospholipid antibodies of the International Society on Thrombosis and Haemostasis update of the guidelines for lupus anticoagulant detection and interpretation. J Thromb Haemost 18:2828, 2020.

Elbaz C, Sholzberg M: An illustrated review of bleeding assessment tools and common coagulation tests. Res Pract Thromb Haemost 4:761, 2020.

Hatfield et al: Dietary supplements and bleeding. Proc (Bayl Univ Med Cent) 35:802, 2022.

James PD et al: ASH ISTH NHF WFH 2021 guidelines on the diagnosis of von Willebrand disease. Blood Adv 5:280, 2021.

Moran J, Bauer KA: Managing thromboembolic risk in patients with hereditary and acquired thrombophilias. Blood 135:344, 2020.

Risman RA et al: Fibrinolysis: an illustrated review. Res Pract Thromb Haemost 7:e100081, 2023.

Yau JW et al: Endothelial cell control of thrombosis. BMC Cardiovasc Disord 15:130, 2015.

70

Enlargement of Lymph Nodes and Spleen

Dan L. Longo

This chapter is intended to serve as a guide to the evaluation of patients who present with enlargement of the lymph nodes (*lymphadenopathy*) or the spleen (*splenomegaly*). Lymphadenopathy is a rather common clinical finding in primary care settings, whereas palpable splenomegaly is less so.

LYMPHADENOPATHY

Lymphadenopathy may be an incidental finding in patients being examined for various reasons, or it may be a presenting sign or symptom of the patient's illness. The physician must eventually decide whether the lymphadenopathy is a normal finding or one that requires further study, up to and including biopsy. Soft, flat, submandibular nodes (<1 cm) are often palpable in healthy children and young adults; healthy adults may have palpable inguinal nodes of up to 2 cm, which are considered normal. Further evaluation of these normal nodes is not warranted. In contrast, if the physician believes the node(s) to be abnormal, then pursuit of a more precise diagnosis is needed.

APPROACH TO THE PATIENT

Lymphadenopathy

Lymphadenopathy may be a primary or secondary manifestation of numerous disorders, as shown in **Table 70-1**. Many of these disorders are infrequent causes of lymphadenopathy. In primary care practice, more than two-thirds of patients with lymphadenopathy have nonspecific causes or upper respiratory illnesses (viral or bacterial) and <1% have a malignancy. In one study, 84% of patients referred for evaluation of lymphadenopathy had a "benign" diagnosis. The remaining 16% had a malignancy (lymphoma or metastatic adenocarcinoma). Of the patients with benign lymphadenopathy, 63% had a nonspecific or reactive etiology (no causative agent found), and the remainder had a specific cause demonstrated, most commonly infectious mononucleosis, toxoplasmosis, or tuberculosis. Thus, the vast majority of patients with lymphadenopathy will have a nonspecific etiology requiring few diagnostic tests.

CLINICAL ASSESSMENT

The physician will be aided in the pursuit of an explanation for the lymphadenopathy by a careful medical history, physical examination, selected laboratory tests, and perhaps an excisional lymph node biopsy.

The *medical history* should reveal the setting in which lymphadenopathy is occurring. Symptoms such as sore throat, cough, fever, night sweats, fatigue, weight loss, or pain in the nodes should be sought. The patient's age, sex, occupation, exposure to pets, sexual behavior, and use of drugs such as diphenylhydantoin are other important historic points. For example, children and young adults usually have benign (i.e., nonmalignant) disorders that account for the observed lymphadenopathy such as viral or bacterial upper respiratory infections; infectious mononucleosis; toxoplasmosis; and, in some countries, tuberculosis. In contrast, after age 50, the incidence of malignant disorders increases and that of benign disorders decreases.

The *physical examination* can provide useful clues such as the extent of lymphadenopathy (localized or generalized), size of nodes, texture, presence or absence of nodal tenderness, signs of inflammation over the node, skin lesions, and splenomegaly. A thorough ear, nose, and throat (ENT) examination is indicated in adult patients with cervical adenopathy and a history of tobacco use. Localized or regional adenopathy implies involvement of a single anatomic area. Generalized adenopathy has been defined as involvement of three or more noncontiguous lymph node areas. Many of the causes of lymphadenopathy (Table 70-1) can produce localized *or* generalized adenopathy, so this distinction is of limited utility in the differential diagnosis. Nevertheless, generalized lymphadenopathy is frequently associated with nonmalignant disorders such as infectious mononucleosis (Epstein-Barr virus [EBV] or cytomegalovirus [CMV]), toxoplasmosis, AIDS, other viral infections, systemic lupus erythematosus (SLE), and mixed connective tissue disease. Acute and chronic lymphocytic leukemias and malignant lymphomas also produce generalized adenopathy in adults.

The site of localized or regional adenopathy may provide a useful clue about the cause. Occipital adenopathy often reflects an

TABLE 70-1 Diseases Associated with Lymphadenopathy

1. Infectious diseases
    a. Viral—infectious mononucleosis syndromes (EBV, CMV), infectious hepatitis, herpes simplex, herpesvirus-6, varicella-zoster virus, rubella, measles, adenovirus, HIV, epidemic keratoconjunctivitis, vaccinia, herpesvirus-8
    b. Bacterial—streptococci, staphylococci, cat-scratch disease, brucellosis, tularemia, plague, chancroid, melioidosis, glanders, tuberculosis, atypical mycobacterial infection, primary and secondary syphilis, diphtheria, leprosy, bartonella
    c. Fungal—histoplasmosis, coccidioidomycosis, paracoccidioidomycosis
    d. Chlamydial—lymphogranuloma venereum, trachoma
    e. Parasitic—toxoplasmosis, leishmaniasis, trypanosomiasis, filariasis
    f. Rickettsial—scrub typhus, rickettsialpox, Q fever
2. Immunologic diseases
    a. Rheumatoid arthritis
    b. Juvenile rheumatoid arthritis
    c. Mixed connective tissue disease
    d. Systemic lupus erythematosus
    e. Dermatomyositis
    f. Sjögren's syndrome
    g. Serum sickness
    h. Drug hypersensitivity—diphenylhydantoin, hydralazine, allopurinol, primidone, gold, carbamazepine, etc.
    i. Angioimmunoblastic lymphadenopathy
    j. Primary biliary cirrhosis
    k. Graft-versus-host disease
    l. Silicone-associated
    m. Autoimmune lymphoproliferative syndrome
    n. IgG4-related disease
    o. Immune reconstitution inflammatory syndrome (IRIS)
3. Malignant diseases
    a. Hematologic—Hodgkin's disease, non-Hodgkin's lymphomas, acute or chronic lymphocytic leukemia, hairy cell leukemia, malignant histiocytosis, amyloidosis
    b. Metastatic—from numerous primary sites
4. Lipid storage diseases—Gaucher's, Niemann-Pick, Fabry, Tangier
5. Endocrine diseases—hyperthyroidism
6. Passive congestion
7. Other disorders
    a. Castleman's disease (giant lymph node hyperplasia)
    b. Sarcoidosis
    c. Dermatopathic lymphadenitis
    d. Lymphomatoid granulomatosis
    e. Histiocytic necrotizing lymphadenitis (Kikuchi's disease)
    f. Sinus histiocytosis with massive lymphadenopathy (Rosai-Dorfman disease)
    g. Mucocutaneous lymph node syndrome (Kawasaki's disease)
    h. Histiocytosis X
    i. Familial Mediterranean fever
    j. Severe hypertriglyceridemia
    k. Vascular transformation of sinuses
    l. Inflammatory pseudotumor of lymph node
    m. Congestive heart failure

Abbreviations: CMV, cytomegalovirus; EBV, Epstein-Barr virus.

infection of the scalp, and preauricular adenopathy accompanies conjunctival infections and cat-scratch disease. The most frequent site of regional adenopathy is the neck, and most of the causes are benign—upper respiratory infections, oral and dental lesions, infectious mononucleosis, or other viral illnesses. The chief malignant causes include metastatic cancer from head and neck, breast, lung, and thyroid primaries. Enlargement of supraclavicular and scalene nodes is always abnormal. Because these nodes drain regions of the lung and retroperitoneal space, they can reflect lymphomas, other cancers, or infectious processes arising in these areas. Virchow's node is an enlarged left supraclavicular node infiltrated with metastatic cancer from a gastrointestinal primary. Metastases to supraclavicular nodes also occur from lung, breast, testis, or ovarian cancers. Tuberculosis, sarcoidosis, and toxoplasmosis are nonneoplastic causes of supraclavicular adenopathy. Axillary adenopathy is usually due to injuries or localized infections of the ipsilateral upper extremity. Malignant causes include melanoma or lymphoma and, in women, breast cancer. Inguinal lymphadenopathy is usually secondary to infections or trauma of the lower extremities and may accompany sexually transmitted diseases such as lymphogranuloma venereum, primary syphilis, genital herpes, or chancroid. These nodes may also be involved by lymphomas and metastatic cancer from primary lesions of the rectum, genitalia, or lower extremities (melanoma).

The size and texture of the lymph node(s) and the presence of pain are useful parameters in evaluating a patient with lymphadenopathy. Nodes <1.0 cm^2 in area (1.0 cm × 1.0 cm or less) are almost always secondary to benign, nonspecific reactive causes. In one retrospective analysis of younger patients (9–25 years) who had a lymph node biopsy, a maximum diameter of >2 cm served as one discriminant for predicting that the biopsy would reveal malignant or granulomatous disease. Another study showed that a lymph node size of 2.25 cm^2 (1.5 cm × 1.5 cm) was the best size limit for distinguishing malignant or granulomatous lymphadenopathy from other causes of lymphadenopathy. Patients with node(s) ≤1.0 cm^2 should be observed after excluding infectious mononucleosis and/or toxoplasmosis unless there are symptoms and signs of an underlying systemic illness.

The texture of lymph nodes may be described as soft, firm, rubbery, hard, discrete, matted, tender, movable, or fixed. Tenderness is found when the capsule is stretched during rapid enlargement, usually secondary to an inflammatory process. Some malignant diseases such as acute leukemia may produce rapid enlargement and pain in the nodes. Nodes involved by lymphoma tend to be large, discrete, symmetric, rubbery, firm, mobile, and nontender. Nodes containing metastatic cancer are often hard, nontender, and nonmovable because of fixation to surrounding tissues. The coexistence of splenomegaly in the patient with lymphadenopathy implies a systemic illness such as infectious mononucleosis, lymphoma, acute or chronic leukemia, SLE, sarcoidosis, toxoplasmosis, cat-scratch disease, or other less common hematologic disorders. The patient's story should provide helpful clues about the underlying systemic illness.

Nonsuperficial presentations (thoracic or abdominal) of adenopathy are usually detected as the result of a symptom-directed diagnostic workup. Thoracic adenopathy may be detected by routine chest radiography or during the workup for superficial adenopathy. It may also be found because the patient complains of a cough or wheezing from airway compression; hoarseness from recurrent laryngeal nerve involvement; dysphagia from esophageal compression; or swelling of the neck, face, or arms secondary to compression of the superior vena cava or subclavian vein. The differential diagnosis of mediastinal and hilar adenopathy includes primary lung disorders and systemic illnesses that characteristically involve mediastinal or hilar nodes. In the young, mediastinal adenopathy is associated with infectious mononucleosis and sarcoidosis. In endemic regions, histoplasmosis can cause unilateral paratracheal lymph node involvement that mimics lymphoma. Tuberculosis can also cause unilateral adenopathy. In older patients, the differential diagnosis includes primary lung cancer (especially among smokers), lymphomas, metastatic carcinoma (usually lung), tuberculosis, fungal infection, and sarcoidosis.

Enlarged intraabdominal or retroperitoneal nodes are usually malignant. Although tuberculosis may present as mesenteric lymphadenitis, these masses usually contain lymphomas or, in young men, germ cell tumors.

LABORATORY INVESTIGATION

The laboratory investigation of patients with lymphadenopathy must be tailored to elucidate the etiology suspected from the patient's history and physical findings. One study from a family practice clinic evaluated 249 younger patients with "enlarged lymph nodes, not infected" or "lymphadenitis." No laboratory studies were obtained in 51%. When studies were performed, the most common were a complete blood count (CBC) (33%), throat culture (16%), chest x-ray (12%), or monospot test (10%). Only eight patients (3%) had a node biopsy, and half of those were normal or reactive. The CBC can provide useful data for the diagnosis of acute or chronic leukemias, EBV or CMV mononucleosis, lymphoma with a leukemic component, pyogenic infections, or immune cytopenias in illnesses such as SLE. Serologic studies may demonstrate antibodies specific to components of EBV, CMV, HIV, and other viruses; *Toxoplasma gondii*; *Brucella*; etc. If SLE is suspected, antinuclear and anti-DNA antibody studies are warranted.

The chest x-ray is usually negative, but the presence of a pulmonary infiltrate or mediastinal lymphadenopathy would suggest tuberculosis, histoplasmosis, sarcoidosis, lymphoma, primary lung cancer, or metastatic cancer and demands further investigation.

A variety of imaging techniques (CT, MRI, ultrasound, color Doppler ultrasonography) have been employed to differentiate benign from malignant lymph nodes, especially in patients with head and neck cancer. CT and MRI are comparably accurate (65–90%) in the diagnosis of metastases to cervical lymph nodes. Ultrasonography has been used to determine the long (L) axis, short (S) axis, and a ratio of long to short axis in cervical nodes. An L/S ratio of <2.0 has a sensitivity and a specificity of 95% for distinguishing benign and malignant nodes in patients with head and neck cancer. This ratio has greater specificity and sensitivity than palpation or measurement of either the long or the short axis alone.

The indications for lymph node biopsy are imprecise, yet it is a valuable diagnostic tool. The decision to biopsy may be made early in a patient's evaluation or delayed for up to two weeks. Prompt biopsy should occur if the patient's history and physical findings suggest a malignancy; examples include a solitary, hard, nontender cervical node in an older patient who is a chronic user of tobacco; supraclavicular adenopathy; and solitary or generalized adenopathy that is firm, movable, and suggestive of lymphoma. If a primary head and neck cancer is suspected as the basis of a solitary, hard cervical node, then a careful ENT examination should be performed. Any mucosal lesion that is suspicious for a primary neoplastic process should be biopsied first. If no mucosal lesion is detected, an excisional biopsy of the largest node should be performed. Fine-needle aspiration should not be performed as the first diagnostic procedure. Most diagnoses require more tissue than such aspiration can provide, and it often delays a definitive diagnosis. Fine-needle aspiration should be reserved for thyroid nodules and for confirmation of relapse in patients whose primary diagnosis is known. If the primary physician is uncertain about whether to proceed to biopsy, consultation with a hematologist or medical oncologist should be helpful. In primary care practices, <5% of lymphadenopathy patients will require a biopsy. That percentage will be considerably larger in referral practices, i.e., hematology, oncology, or ENT.

Two groups have reported algorithms that they claim will identify more precisely those lymphadenopathy patients who should have a biopsy. Both reports were retrospective analyses in referral practices. The first study involved patients 9–25 years of age who had a node biopsy performed. Three variables were identified that predicted those young patients with peripheral lymphadenopathy who should undergo biopsy; lymph node size >2 cm in diameter and abnormal chest x-ray had positive predictive values, whereas recent ENT symptoms had negative predictive values. The second study evaluated 220 lymphadenopathy patients in a hematology unit and identified five variables (lymph node size, location [supraclavicular or nonsupraclavicular], age [>40 years or <40 years], texture [nonhard or hard], and tenderness) that were used in a mathematical model to identify those patients requiring a biopsy. Positive predictive value was found for age >40 years, supraclavicular location, node size >2.25 cm^2, hard texture, and lack of pain or tenderness. Negative predictive value was evident for age <40 years, node size <1.0 cm^2, nonhard texture, and tender or painful nodes. Ninety-one percent of those who required biopsy were correctly classified by this model. Because both of these studies were retrospective analyses and one was limited to young patients, it is not known how useful these models would be if applied prospectively in a primary care setting.

Most lymphadenopathy patients do not require a biopsy, and at least half require no laboratory studies. If the patient's history and physical findings point to a benign cause for lymphadenopathy, careful follow-up at a 2- to 4-week interval can be employed. The patient should be instructed to return for reevaluation if there is an increase in the size of the nodes. Antibiotics are not indicated for lymphadenopathy unless strong evidence of a bacterial infection is present. Glucocorticoids should not be used to treat lymphadenopathy because their lympholytic effect obscures some diagnoses (lymphoma, leukemia, Castleman's disease) and they contribute to delayed healing or activation of underlying infections. An exception to this statement is the life-threatening pharyngeal obstruction by enlarged lymphoid tissue in Waldeyer's ring that is occasionally seen in infectious mononucleosis.

SPLENOMEGALY

STRUCTURE AND FUNCTION OF THE SPLEEN

The spleen is a reticuloendothelial organ that has its embryologic origin in the dorsal mesogastrium at about 5 weeks' gestation. It arises in a series of hillocks, migrates to its normal adult location in the left upper quadrant (LUQ), and is attached to the stomach via the gastrolienal ligament and to the kidney via the lienorenal ligament. When the hillocks fail to unify into a single tissue mass, accessory spleens may develop in around 20% of persons. The function of the spleen has been elusive. Galen believed it was the source of "black bile" or melancholia, and the word *hypochondria* (literally, beneath the ribs) and the idiom "to vent one's spleen" attest to the beliefs that the spleen had an important influence on the psyche and emotions.* In humans, its normal physiologic roles seem to be the following:

1. Maintenance of quality control over erythrocytes in the red pulp by removal of senescent and defective red blood cells. The spleen accomplishes this function through a unique organization of its parenchyma and vasculature **(Fig. 70-1)**.
2. Synthesis of antibodies in the white pulp.
3. The removal of antibody-coated bacteria and antibody-coated blood cells from the circulation.

An increase in these normal functions may result in splenomegaly.

The spleen is composed of *red pulp* and *white pulp*, which are Malpighi's terms for the red blood–filled sinuses and reticuloendothelial cell–lined cords and the white lymphoid follicles arrayed within the red pulp matrix. The spleen is in the portal circulation. The reason for this is unknown but may relate to the fact that lower blood pressure allows less rapid flow and minimizes damage to normal erythrocytes. Blood flows into the spleen at a rate of about 150 mL/min through the splenic artery, which ultimately ramifies into central arterioles. Some blood goes from the arterioles to capillaries and then to splenic veins and out of the spleen, but the majority of blood from central arterioles flows into the macrophage-lined sinuses and cords. The blood entering the sinuses reenters the circulation through the splenic venules, but the blood entering the cords is subjected to an inspection of sorts. To return to the circulation, the blood cells in the cords must squeeze through slits in the cord lining to enter the sinuses that lead to the venules. Old and damaged erythrocytes are less deformable and are retained in the cords, where

---

*Reproduced with permission from RS Hillman, KA Ault: Hematology in Clinical Practice, 4th ed. New York, McGraw-Hill; 2005.

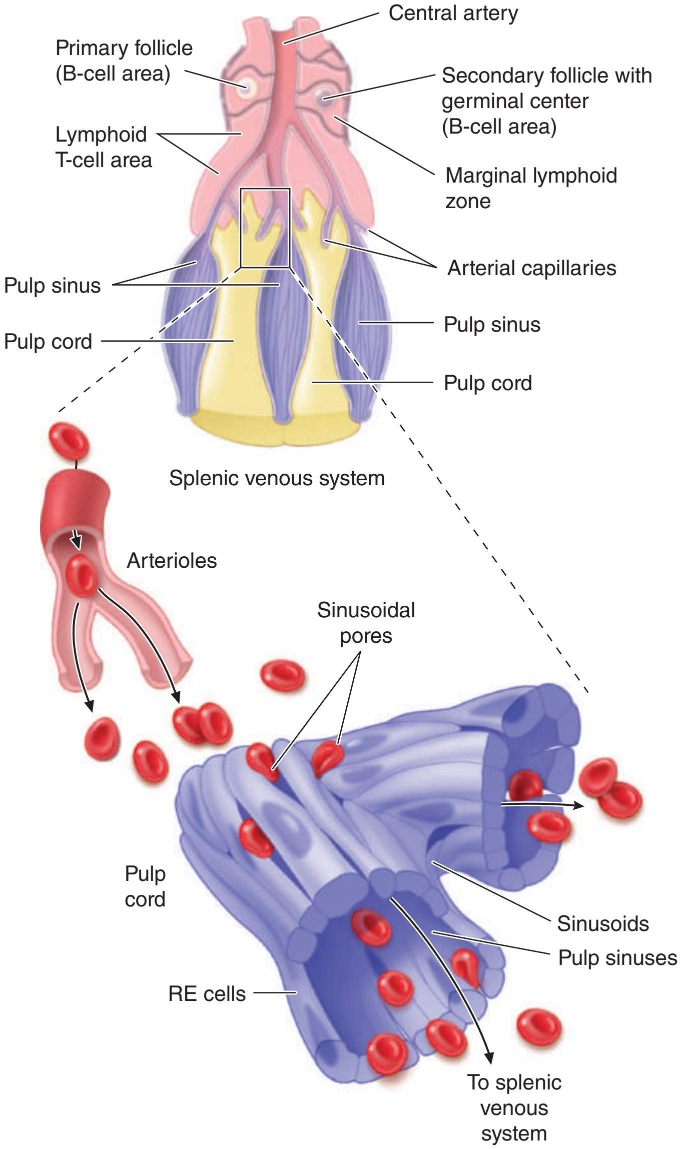

FIGURE 70-1 Schematic spleen structure. The spleen comprises many units of red and white pulp centered around small branches of the splenic artery, called *central arteries*. White pulp is lymphoid in nature and contains B-cell follicles, a marginal zone around the follicles, and T-cell–rich areas sheathing arterioles. The red pulp areas include pulp sinuses and pulp cords. The cords are dead ends. In order to regain access to the circulation, red blood cells must traverse tiny openings in the sinusoidal lining. Stiff, damaged, or old red cells cannot enter the sinuses. RE, reticuloendothelial. *(Bottom portion of figure reproduced with permission from RS Hillman, KA Ault: Hematology in Clinical Practice, 4th ed. New York, McGraw-Hill, 2005.)*

they are destroyed and their components recycled. Red cell–inclusion bodies such as parasites **(Chaps. 231, 232, and A2)**, nuclear residua (Howell-Jolly bodies, **see Fig. 66-6**), or denatured hemoglobin (Heinz bodies) are pinched off in the process of passing through the slits, a process called *pitting*. The culling of dead and damaged cells and the pitting of cells with inclusions appear to occur without significant delay because the blood transit time through the spleen is only slightly slower than in other organs.

The spleen is also capable of assisting the host in adapting to its hostile environment. It has at least three adaptive functions: (1) clearance of bacteria and particulates from the blood, (2) the generation of immune responses to certain pathogens, and (3) the generation of cellular components of the blood under circumstances in which the marrow is unable to meet the needs (i.e., extramedullary hematopoiesis). The latter adaptation is a recapitulation of the blood-forming function the spleen plays during gestation. In some animals, the spleen also serves a role in the vascular adaptation to stress because it stores red blood cells (often hemoconcentrated to higher hematocrits than normal) under normal circumstances and contracts under the influence of β-adrenergic stimulation to provide the animal with an autotransfusion and improved oxygen-carrying capacity. However, the normal human spleen does not sequester or store red blood cells and does not contract in response to sympathetic stimuli. The normal human spleen contains approximately one-third of the total body platelets and a significant number of marginated neutrophils. These sequestered cells are available when needed to respond to bleeding or infection.

APPROACH TO THE PATIENT

Splenomegaly

CLINICAL ASSESSMENT

The most common *symptoms* produced by diseases involving the spleen are pain and a heavy sensation in the LUQ. Massive splenomegaly may cause early satiety. Pain may result from acute swelling of the spleen with stretching of the capsule, infarction, or inflammation of the capsule. For many years, it was believed that splenic infarction was clinically silent, which, at times, is true. However, Soma Weiss, in his classic 1942 report of the self-observations by a Harvard medical student on the clinical course of subacute bacterial endocarditis, documented that severe LUQ and pleuritic chest pain may accompany thromboembolic occlusion of splenic blood flow. Vascular occlusion, with infarction and pain, is commonly seen in children with sickle cell crises. Rupture of the spleen, from either trauma or infiltrative disease that breaks the capsule, may result in intraperitoneal bleeding, shock, and death. The rupture itself may be painless.

A palpable spleen is the major *physical sign* produced by diseases affecting the spleen and suggests enlargement of the organ. The normal spleen weighs <250 g, decreases in size with age, normally lies entirely within the rib cage, has a maximum cephalocaudad diameter of 13 cm by ultrasonography or maximum length of 12 cm and/or width of 7 cm by radionuclide scan, and is usually not palpable. However, a palpable spleen was found in 3% of 2200 asymptomatic, male, freshman college students. Follow-up at 3 years revealed that 30% of those students still had a palpable spleen without any increase in disease prevalence. Ten-year follow-up found no evidence for lymphoid malignancies. Furthermore, in some tropical countries (e.g., New Guinea), the incidence of splenomegaly may reach 60%. Thus, the presence of a palpable spleen does not always equate with presence of disease. Even when disease is present, splenomegaly may not reflect the primary disease but rather a reaction to it. For example, in patients with Hodgkin's lymphoma, only two-thirds of the palpable spleens show involvement by the cancer.

Physical examination of the spleen uses primarily the techniques of palpation and percussion. Inspection may reveal fullness in the LUQ that descends on inspiration, a finding associated with a massively enlarged spleen. Auscultation may reveal a venous hum or friction rub.

Palpation can be accomplished by bimanual palpation, ballotment, and palpation from above (Middleton maneuver). For bimanual palpation, which is at least as reliable as the other techniques, the patient is supine with flexed knees. The examiner's left hand is placed on the lower rib cage and pulls the skin toward the costal margin, allowing the fingertips of the right hand to feel the tip of the spleen as it descends while the patient inspires slowly, smoothly, and deeply. Palpation is begun with the right hand in the left lower quadrant with gradual movement toward the left costal margin, thereby identifying the lower edge of a massively enlarged spleen. When the spleen tip is felt, the finding is recorded as centimeters below the left costal margin at some arbitrary point, i.e., 10–15 cm, from the midpoint of the umbilicus or the xiphisternal junction. This allows other examiners to compare findings or the initial examiner to determine changes in size over time. Bimanual palpation in the right lateral decubitus position adds nothing to the supine examination.

Percussion for splenic dullness is accomplished with any of three techniques described by Nixon, Castell, or Barkun:

1. *Nixon's method*: The patient is placed on the right side so that the spleen lies above the colon and stomach. Percussion begins at the lower level of pulmonary resonance in the posterior axillary line and proceeds diagonally along a perpendicular line toward the lower midanterior costal margin. The upper border of dullness is normally 6–8 cm above the costal margin. Dullness >8 cm in an adult is presumed to indicate splenic enlargement.
2. *Castell's method*: With the patient supine, percussion in the lowest intercostal space in the anterior axillary line (8th or 9th) produces a resonant note if the spleen is normal in size. This is true during expiration or full inspiration. A dull percussion note on full inspiration suggests splenomegaly.
3. *Percussion of Traube's semilunar space*: The borders of Traube's space are the sixth rib superiorly, the left midaxillary line laterally, and the left costal margin inferiorly. The patient is supine with the left arm slightly abducted. During normal breathing, this space is percussed from medial to lateral margins, yielding a normal resonant sound. A dull percussion note suggests splenomegaly.

Studies comparing methods of percussion and palpation with a standard of ultrasonography or scintigraphy have revealed sensitivity of 56–71% for palpation and 59–82% for percussion. Reproducibility among examiners is better for palpation than percussion. Both techniques are less reliable in obese patients or patients who have just eaten. Thus, the physical examination techniques of palpation and percussion are imprecise at best. It has been suggested that the examiner perform percussion first and, if positive, proceed to palpation; if the spleen is palpable, then one can be reasonably confident that splenomegaly exists. However, not all LUQ masses are enlarged spleens; gastric or colon tumors and pancreatic or renal cysts or tumors can mimic splenomegaly.

The presence of an enlarged spleen can be more precisely determined, if necessary, by liver-spleen radionuclide scan, CT, MRI, or ultrasonography. The latter technique is the current procedure of choice for routine assessment of spleen size (normal = a maximum cephalocaudad diameter of 13 cm) because it has high sensitivity and specificity and is safe, noninvasive, quick, mobile, and less costly. Equipment advances allow ultrasonography to be performed at the bedside with excellent sensitivity and specificity. Nuclear medicine scans are accurate, sensitive, and reliable but are costly, require greater time to generate data, and use immobile equipment. They have the advantage of demonstrating accessory splenic tissue. CT and MRI provide accurate determination of spleen size, but the equipment is immobile and the procedures are expensive. MRI appears to offer no advantage over CT. Changes in spleen structure such as mass lesions, infarcts, inhomogeneous infiltrates, and cysts are more readily assessed by CT, MRI, or ultrasonography. None of these techniques is very reliable in the detection of patchy infiltration (e.g., Hodgkin's lymphoma).

DIFFERENTIAL DIAGNOSIS

Many of the diseases associated with splenomegaly are listed in **Table 70-2**. They are grouped according to the presumed basic mechanisms responsible for organ enlargement:

1. Hyperplasia or hypertrophy related to a particular splenic function such as reticuloendothelial hyperplasia (work hypertrophy) in diseases such as hereditary spherocytosis or thalassemia syndromes that require removal of large numbers of defective red blood cells; immune hyperplasia in response to systemic infection (infectious mononucleosis, subacute bacterial endocarditis) or to immunologic diseases (immune thrombocytopenia, SLE, Felty's syndrome).
2. Passive congestion due to decreased blood flow from the spleen in conditions that produce portal hypertension (cirrhosis, Budd-Chiari syndrome, congestive heart failure).
3. Infiltrative diseases of the spleen (lymphomas, metastatic cancer, amyloidosis, Gaucher's disease, myeloproliferative disorders with extramedullary hematopoiesis).

The differential diagnostic possibilities are much fewer when the spleen is "massively enlarged," palpable >8 cm below the left costal margin, or its drained weight is ≥1000 g **(Table 70-3)**. The vast majority of such patients will have non-Hodgkin's lymphoma, chronic lymphocytic leukemia, hairy cell leukemia, chronic myeloid leukemia, myelofibrosis with myeloid metaplasia, or polycythemia vera.

LABORATORY ASSESSMENT

The major laboratory abnormalities accompanying splenomegaly are determined by the underlying systemic illness. Erythrocyte counts may be normal, decreased (thalassemia major syndromes, SLE, cirrhosis with portal hypertension), or increased (polycythemia vera). Granulocyte counts may be normal, decreased (Felty's syndrome, congestive splenomegaly, leukemias), or increased (infections or inflammatory disease, myeloproliferative disorders). Similarly, the platelet count may be normal, decreased when there is enhanced sequestration or destruction of platelets in an enlarged spleen (congestive splenomegaly, Gaucher's disease, immune thrombocytopenia), or increased in the myeloproliferative disorders such as polycythemia vera.

The CBC may reveal cytopenia of one or more blood cell types, which should suggest *hypersplenism*. This condition is characterized by splenomegaly, cytopenia(s), normal or hyperplastic bone marrow, and a response to splenectomy. The latter characteristic is less precise because reversal of cytopenia, particularly granulocytopenia, is sometimes not sustained after splenectomy. The cytopenias result from increased destruction of the cellular elements secondary to reduced flow of blood through enlarged and congested cords (congestive splenomegaly) or to immune-mediated mechanisms. In hypersplenism, various cell types usually have normal morphology on the peripheral blood smear, although the red cells may be spherocytic due to loss of surface area during their longer transit through the enlarged spleen. The increased marrow production of red cells should be reflected as an increased reticulocyte production index, although the value may be less than expected due to increased sequestration of reticulocytes in the spleen.

The need for additional laboratory studies is dictated by the differential diagnosis of the underlying illness of which splenomegaly is a manifestation.

SPLENECTOMY

Splenectomy is infrequently performed for diagnostic purposes, especially in the absence of clinical illness or other diagnostic tests that suggest underlying disease. More often, splenectomy is performed for symptom control in patients with massive splenomegaly, for disease control in patients with traumatic splenic rupture, or for correction of cytopenias in patients with hypersplenism or immune-mediated destruction of one or more cellular blood elements. Splenectomy is necessary for staging of patients with Hodgkin's lymphoma only in those with clinical stage I or II disease in whom radiation therapy alone is contemplated as the treatment. Noninvasive staging of the spleen in Hodgkin's lymphoma is not a sufficiently reliable basis for treatment decisions because one-third of normal-sized spleens will be involved with Hodgkin's lymphoma and one-third of enlarged spleens will be tumor-free. The widespread use of systemic therapy to test all stages of Hodgkin's lymphoma has made staging laparotomy with splenectomy unnecessary. Although splenectomy in chronic myeloid leukemia (CML) does not affect the natural history of disease, removal of the massive spleen usually makes patients significantly more comfortable and simplifies their management by significantly reducing transfusion requirements. The improvements in therapy of CML have reduced the need for splenectomy for symptom control. Splenectomy is an effective secondary or tertiary treatment for two chronic B-cell leukemias,

TABLE 70-2 Diseases Associated with Splenomegaly Grouped by Pathogenic Mechanism

Enlargement Due to Increased Demand for Splenic Function

- Reticuloendothelial system hyperplasia (for removal of defective erythrocytes)
  - Spherocytosis
  - Early sickle cell anemia
  - Ovalocytosis
  - Thalassemia major
  - Hemoglobinopathies
  - Paroxysmal nocturnal hemoglobinuria
  - Pernicious anemia
- Immune hyperplasia
  - Response to infection (viral, bacterial, fungal, parasitic)
    - Infectious mononucleosis
    - AIDS
    - Viral hepatitis
    - Cytomegalovirus
    - Subacute bacterial endocarditis
    - Bacterial septicemia
    - Congenital syphilis
    - Splenic abscess
    - Tuberculosis
    - Histoplasmosis
  - Malaria
  - Leishmaniasis
  - Trypanosomiasis
  - Ehrlichiosis
- Disordered immunoregulation
  - Rheumatoid arthritis (Felty's syndrome)
  - Systemic lupus erythematosus
  - Collagen vascular diseases
  - Serum sickness
  - Immune hemolytic anemias
  - Immune thrombocytopenias
  - Immune neutropenias
  - Drug reactions
  - Angioimmunoblastic lymphadenopathy
  - Sarcoidosis
  - Thyrotoxicosis (benign lymphoid hypertrophy)
  - Interleukin 2 therapy
- Extramedullary hematopoiesis
  - Myelofibrosis
  - Marrow damage by toxins, radiation, strontium
  - Marrow infiltration by tumors, leukemias, Gaucher's disease

Enlargement Due to Abnormal Splenic or Portal Blood Flow

- Cirrhosis
- Hepatic vein obstruction
- Portal vein obstruction, intrahepatic or extrahepatic
- Cavernous transformation of the portal vein
- Splenic vein obstruction
- Splenic artery aneurysm
- Hepatic schistosomiasis
- Congestive heart failure
- Hepatic echinococcosis
- Portal hypertension (any cause including the above): "Banti's disease"

Infiltration of the Spleen

- Intracellular or extracellular depositions
  - Amyloidosis
  - Gaucher's disease
  - Niemann-Pick disease
  - Tangier disease
  - Hurler's syndrome and other mucopolysaccharidoses
  - Hyperlipidemias
- Benign and malignant cellular infiltrations
  - Leukemias (acute, chronic, lymphoid, myeloid, monocytic)
  - Lymphomas
  - Hodgkin's lymphoma
  - Myeloproliferative syndromes (e.g., polycythemia vera, essential thrombocytosis)
  - Angiosarcomas
  - Metastatic tumors (melanoma is most common)
  - Eosinophilic granuloma
  - Histiocytosis X
  - Hamartomas
  - Hemangiomas, fibromas, lymphangiomas
  - Splenic cysts
  - Hemophagocytic lymphohistiocytosis

Unknown Etiology

- Idiopathic splenomegaly
- Berylliosis
- Iron-deficiency anemia

hairy cell leukemia and prolymphocytic leukemia, and for the very rare splenic mantle cell or marginal zone lymphoma. Splenectomy in these diseases may be associated with significant tumor regression in bone marrow and other sites of disease. Similar regressions of systemic disease have been noted after splenic irradiation in some types of lymphoid tumors, especially chronic lymphocytic leukemia and prolymphocytic leukemia. This has been termed the *abscopal effect*. Such systemic tumor responses to local therapy directed at the spleen suggest that some hormone or growth factor produced by the spleen may affect tumor cell proliferation, but this conjecture is not yet substantiated. A common therapeutic indication for splenectomy is traumatic or iatrogenic splenic rupture. In a fraction of patients with splenic rupture, peritoneal seeding of splenic fragments can lead to *splenosis*—the presence of multiple rests of spleen tissue not connected to the portal circulation. This ectopic spleen tissue may cause pain or gastrointestinal obstruction, as in endometriosis. A large number of hematologic, immunologic, and congestive causes of splenomegaly can lead to destruction of one or more cellular blood elements. In the majority of such cases, splenectomy can correct the cytopenias, particularly anemia and thrombocytopenia. In a large series of patients seen in two tertiary care centers, the indication for splenectomy was diagnostic in 10% of patients, therapeutic in 44%, staging for Hodgkin's

TABLE 70-3 Diseases Associated with Massive Splenomegaly[a]

Chronic myeloid leukemia	Gaucher's disease
Lymphomas	Chronic lymphocytic leukemia
Hairy cell leukemia	Sarcoidosis
Myelofibrosis with myeloid metaplasia	Autoimmune hemolytic anemia
Polycythemia vera	Diffuse splenic hemangiomatosis

[a] The spleen extends >8 cm below left costal margin and/or weighs >1000 g.

disease in 20%, and incidental to another procedure in 26%. Perhaps the only contraindication to splenectomy is the presence of marrow failure, in which the enlarged spleen is the only source of hematopoietic tissue.

Often the splenectomy is done by laparoscopy, which is associated with shorter hospital stays and faster recovery than the open procedure; however, concern has emerged that the laparoscopic approach is associated with a higher risk of postoperative portal venous system thrombosis and Budd-Chiari syndrome.

The absence of the spleen has minimal long-term effects on the hematologic profile. In the immediate postsplenectomy period, leukocytosis (up to 25,000/μL) and thrombocytosis (up to 1×10^6/μL) may develop, but within 2–3 weeks, blood cell counts and survival of each cell lineage are usually normal. The chronic manifestations of splenectomy are marked variation in size and shape of erythrocytes (anisocytosis, poikilocytosis) and the presence of Howell-Jolly bodies (nuclear remnants), Heinz bodies (denatured hemoglobin), basophilic stippling, and an occasional nucleated erythrocyte in the peripheral blood. When such erythrocyte abnormalities appear in a patient whose spleen has not been removed, one should suspect splenic infiltration by tumor that has interfered with its normal culling and pitting function.

The most serious consequence of splenectomy is increased susceptibility to bacterial infections, particularly those with capsules such as *Streptococcus pneumoniae*, *Haemophilus influenzae*, and some gram-negative enteric organisms. Patients aged <20 years are particularly susceptible to overwhelming sepsis with *S. pneumoniae*, and the overall actuarial risk of sepsis in patients who have had their spleens removed is about 7% in 10 years. The case-fatality rate for pneumococcal sepsis in splenectomized patients is 50–80%. About 25% of patients without spleens will develop a serious infection at some time in their life. The frequency is highest within the first 3 years after splenectomy. About 15% of the infections are polymicrobial, and lung, skin, and blood are the most common sites. No increased risk of viral infection has been noted in patients who have no spleen. The susceptibility to bacterial infections relates to the inability to remove opsonized bacteria from the bloodstream and a defect in making antibodies to T cell–independent antigens such as the polysaccharide components of bacterial capsules. Pneumococcal vaccine should be administered to all patients 2 weeks before elective splenectomy. The Advisory Committee on Immunization Practices recommends that these patients receive repeat vaccination 5 years after splenectomy. Efficacy has not been proven for this group, and the recommendation discounts the possibility that administration of the vaccine may actually lower the titer of specific pneumococcal antibodies. A more effective pneumococcal conjugate vaccine that involves T cells in the response is now available (PCV13). The vaccine to *Neisseria meningitidis* should also be given to patients in whom elective splenectomy is planned. Although efficacy data for *H. influenzae* type b vaccine are not available for older children or adults, it may be given to patients who have had a splenectomy.

Splenectomized patients should be educated to consider any unexplained fever as a medical emergency. Prompt medical attention with evaluation and treatment of suspected bacteremia may be lifesaving. Routine chemoprophylaxis with oral penicillin can result in the emergence of drug-resistant strains and is not recommended.

In addition to an increased susceptibility to bacterial infections, splenectomized patients are also more susceptible to the parasitic disease babesiosis. The splenectomized patient should avoid areas where the parasite *Babesia* is endemic (e.g., Cape Cod, Massachusetts).

Surgical removal of the spleen is an obvious cause of hyposplenism. Patients with sickle cell disease often suffer from autosplenectomy as a result of splenic destruction by the numerous infarcts associated with sickle cell crises during childhood. Indeed, the presence of a palpable spleen in a patient with sickle cell disease after age 5 suggests a coexisting hemoglobinopathy, e.g., thalassemia or hemoglobin C. In addition, patients who receive splenic irradiation for a neoplastic or autoimmune disease are also functionally hyposplenic. The term *hyposplenism* is preferred to *asplenism* in referring to the physiologic consequences of splenectomy because asplenia is a rare, specific, and fatal congenital abnormality in which there is a failure of the left side of the coelomic cavity (which includes the splenic anlagen) to develop normally. Infants with asplenia have no spleens, but that is the least of their problems. The right side of the developing embryo is duplicated on the left so there is liver where the spleen should be, there are two right lungs, and the heart comprises two right atria and two right ventricles.

Acknowledgment

Patrick H. Henry, MD, friend and mentor now deceased, contributed significantly to the chapter in past editions and much of his work remains in this chapter.

FURTHER READING

Barkun AN et al: The bedside assessment of splenic enlargement. Am J Med 91:512, 1991.

Cessford T et al: Comparing physical examination with sonographic versions of the same examination techniques for splenomegaly. J Ultrasound Med 37:1621, 2018.

Facchetti F: Tumors of the spleen. Int J Surg Pathol 18:136S, 2010.

Girard E et al: Management of splenic and pancreatic trauma. J Visc Surg 153(suppl 4):45, 2016.

Kim DK et al: Advisory committee on immunization practices recommended immunization schedule for adults aged 19 years or older—United States, 2017. MMWR 66:136, 2017.

Kraus MD et al: The spleen as a diagnostic specimen: A review of ten years' experience at two tertiary care institutions. Cancer 91:2001, 2001.

McIntyre OR, Ebaugh FG Jr: Palpable spleens: Ten-year follow-up. Ann Intern Med 90:130, 1979.

Pangalis GA et al: Clinical approach to lymphadenopathy. Semin Oncol 20:570, 1993.

Sjoberg BP et al: Splenomegaly: A combined clinical and radiologic approach to the differential diagnosis. Gastroenterol Clin North Am 47:643, 2018.

71 Principles of Clinical Pharmacology

Dan Roden

Drugs are the cornerstone of modern therapeutics. Nevertheless, it is well recognized among health care providers and the lay community that the outcome of drug therapy varies widely among individuals. While this variability has been perceived as an unpredictable, and therefore inevitable, accompaniment of drug therapy, this is not the case.

Drugs interact with specific target molecules to produce their beneficial and adverse effects. The chain of events between administration of a drug and production of these effects in the body can be divided into two components, both of which contribute to variability in drug actions. The first component comprises the processes that determine drug delivery to, and removal from, molecular targets. The resulting description of the relationship between drug concentration and time is termed *pharmacokinetics*. The second component of variability in drug action comprises the processes that determine variability in drug actions independent of variability in drug delivery to effector drug sites. This description of the relationship between drug concentration and effect is termed *pharmacodynamics*. This chapter describes how these processes can be analyzed for any drug, and presents examples from multiple areas of therapeutics, each of which is covered in more detail in other chapters.

Two important goals of clinical pharmacology are (1) to provide a description of conditions under which drug actions vary among human subjects; and (2) to determine mechanisms underlying this variability, with the goal of improving therapy with available drugs as well as pointing to mechanisms whose targeting by new drugs may be effective in the treatment of human disease. The drug development process is briefly described at the end of this chapter.

The first steps in the discipline of clinical pharmacology were empirical descriptions of the influence of disease on drug actions and of individuals or families with unusual sensitivities to adverse drug reactions (ADRs). These important descriptive findings are now being replaced by an understanding of the molecular mechanisms underlying variability in drug actions.

One useful unifying framework is to consider that the effects of disease, drug coadministration, or familial factors in modulating drug action reflect variability in expression or function of specific genes whose products determine pharmacokinetics and pharmacodynamics. This idea forms the basis for pharmacogenomic science; a few examples are cited in this chapter, and further details are addressed in Chap. 72.

A SYSTEMS BIOLOGY VIEW

The framework that this chapter, and the field, use to analyze variable drug actions rests on the premise that the processes occurring between administration of a drug and generation of a drug effect are relatively linear. However, as the complexities of how drugs interact with disease mechanisms are becoming increasingly well-defined, it is apparent that this linear framework represents a first approximation of understanding drug effects. Disease processes change over time, and drugs often exert multiple (and occasionally counterregulatory) effects. The development of complex computational models for these processes, and tools such as gene editing and multiple "omic" measurements (transcriptomics, metabolomics, proteomics) to study and refine those models, are now presenting the opportunity of understanding disease mechanisms and their variable responses to drug challenge at a much finer and more precise scale. Thus, this systems biology approach may represent the future in clinical pharmacology and new drug development.

GLOBAL CONSIDERATIONS

It is true across all cultures and diseases that factors such as adherence, genetic variants affecting pharmacokinetics or pharmacodynamics (which themselves vary by ancestry), and drug interactions contribute to drug responses; the term "adherence" is preferred to the older term "compliance" because it removes the idea that the patient is at fault. Cost issues or cultural factors may determine the likelihood that specific drugs, drug combinations, or over-the-counter (OTC) remedies are prescribed.

INDICATIONS FOR DRUG THERAPY: RISK VERSUS BENEFIT

It is self-evident that the benefits of drug therapy should outweigh the risks. Benefits fall into broad categories: alleviation of symptoms, prevention of disease progression or complications, and prolonged life. However, establishing the balance between risk and benefit for an individual patient is not always simple. In addition to variability seen even within highly controlled drug trials, patients treated in clinical settings may display responses that were not observed in trials, sometimes due to comorbidities that were trial exclusion criteria. In addition, therapies that provide symptomatic benefits but shorten life may be entertained in patients with serious and highly symptomatic diseases such as heart failure or cancer.

Adverse Effects Some adverse effects are so common and so readily associated with drug therapy that they are identified very early during clinical use of a drug. By contrast, serious ADRs may be sufficiently uncommon that they escape detection for many years after a drug begins to be widely used. Potential approaches to detect rare ADRs range from an increased understanding of the molecular and genetic basis of variability in drug actions to expanded postmarketing surveillance mechanisms.

Therapeutic Index Beneficial and adverse reactions to drug therapy can be described by a series of dose-response relations (Fig. 71-1). Well-tolerated drugs demonstrate a wide margin, termed the *therapeutic ratio*, *therapeutic index*, or *therapeutic window*, between the doses required to produce a therapeutic effect and those producing toxicity. In cases where there is an established relationship between plasma drug concentration and effects, monitoring plasma concentrations can be a highly effective aid in managing drug therapy by enabling concentrations to be maintained above the minimum required to produce an effect and below the concentration range likely to produce toxicity.

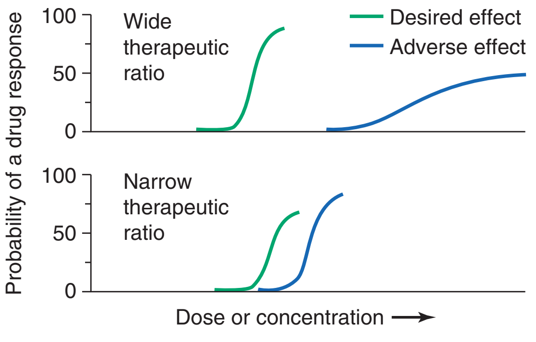

FIGURE 71-1 The concept of a therapeutic ratio. Each panel illustrates the relationship between increasing dose and cumulative probability of a desired or adverse drug effect. ***Top.*** A drug with a wide therapeutic ratio, that is, a wide separation of the two curves. ***Bottom.*** A drug with a narrow therapeutic ratio; here, the likelihood of adverse effects at therapeutic doses is increased because the curves are not well separated. Further, a steep dose-response curve for adverse effects is especially undesirable, as it implies that even small dosage increments may sharply increase the likelihood of toxicity. When there is a definable relationship between drug concentration (usually measured in plasma) and desirable and adverse effect curves, concentration may be substituted on the abscissa. Note that not all patients necessarily demonstrate a therapeutic response (or adverse effect) at any dose and that some effects (notably some adverse effects) may occur in a dose-independent fashion.

PRINCIPLES OF PHARMACOKINETICS

The processes of absorption, distribution, metabolism, and excretion—collectively termed *drug disposition*—determine the concentration of drug delivered to target effector molecules.

ABSORPTION AND BIOAVAILABILITY

When a drug is administered orally, subcutaneously, intramuscularly, rectally, sublingually, or directly into desired sites of action, the amount of drug eventually entering the systemic circulation may be less than with the intravenous route **(Fig. 71-2*A*)**. The fraction of drug available to the systemic circulation by other routes is termed *bioavailability*. Bioavailability may be <100% for two main reasons: (1) incomplete absorption, or (2) metabolism or elimination prior to entering the systemic circulation.

Compared to the same dose given intravenously, a nonintravenous dose will have a later and lower peak plasma concentration **(Fig. 71-2*B*)**. Drug absorption may be reduced because a drug is incompletely released from its dosage form, undergoes destruction at the site of administration, or has physicochemical properties such as poor solubility that prevent complete absorption from its site of administration. Slow absorption rates are deliberately designed into "slow-release" or "sustained-release" drug formulations in order to minimize variation in plasma concentrations during the interval between doses.

"First-Pass" Effect When a drug is administered orally, it must traverse the intestinal epithelium, the portal venous system, and the liver prior to entering the systemic circulation **(Fig. 71-3)**. Once a drug enters the enterocyte, it may undergo metabolism, be transported into the portal vein, or be excreted back into the intestinal lumen. Both excretion into the intestinal lumen and metabolism decrease bioavailability. Once a drug passes this enterocyte barrier, it may also be taken up into the hepatocyte, where bioavailability can be further limited by metabolism or excretion into the bile. This elimination in intestine and liver, which reduces the amount of drug delivered to the systemic circulation, is termed *presystemic elimination*, *presystemic extraction*, or *first-pass elimination*.

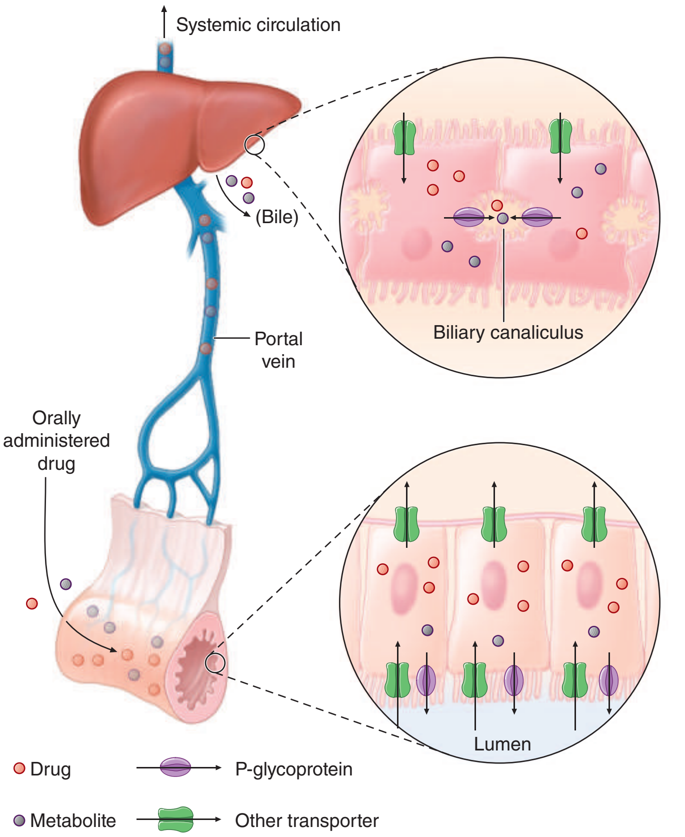

FIGURE 71-3 Mechanism of presystemic elimination. After drug enters the enterocyte, it can undergo metabolism, excretion into the intestinal lumen, or transport into the portal vein. Similarly, the hepatocyte may accomplish metabolism and biliary excretion prior to the entry of drug and metabolites to the systemic circulation. *(Reproduced with permission from DM Roden, in DP Zipes, J Jalife [eds]: Cardiac Electrophysiology: From Cell to Bedside, 4th ed. Philadelphia, Saunders, 2003.)*

DRUG TRANSPORT

Drug movement across the membrane of any cell, including enterocytes and hepatocytes, is a combination of passive diffusion and active transport, mediated by specific drug uptake and efflux molecules. One widely studied drug transport molecule is the drug efflux pump P-glycoprotein, the product of the *ABCB1* (or *MDR1*) gene. P-glycoprotein is expressed on the apical aspect of the enterocyte and on the canalicular aspect of the hepatocyte (Fig. 71-3). In both locations, it serves as an efflux pump, limiting availability of drug to the systemic circulation. P-glycoprotein–mediated drug efflux from cerebral capillaries limits drug brain penetration and is an important component of the blood-brain barrier. Other transporters mediate uptake into cells of drugs and endogenous substrates such as vitamins or nutrients.

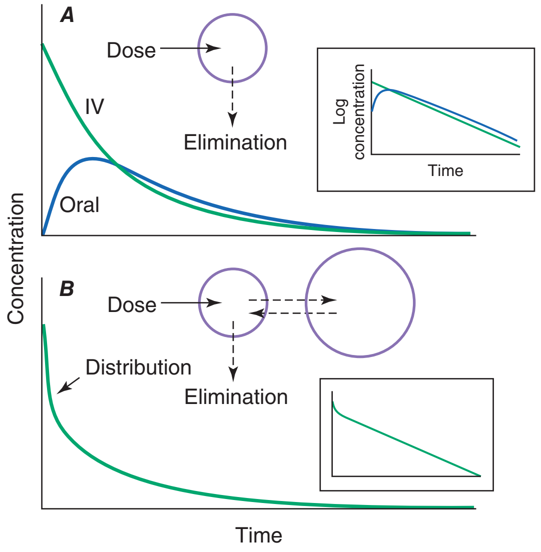

FIGURE 71-2 Idealized time-plasma concentration curves after a single dose of drug. *A.* The time course of drug concentration after an instantaneous intravenous (IV) bolus or an oral dose in the one-compartment model shown. The area under the time-concentration curve is clearly less with the oral drug than the IV drug, indicating incomplete bioavailability. Note that despite this incomplete bioavailability, concentration after the oral dose can be higher than after the IV dose at some time points. The inset shows that the decline of concentrations over time is linear on a log-linear plot, characteristic of first-order elimination, and that oral and IV drugs have the same elimination (parallel) time course. ***B.*** The decline of central compartment concentration when drug is distributed both to and from a peripheral compartment and eliminated from the central compartment. The rapid initial decline of concentration reflects not drug elimination but distribution.

DRUG METABOLISM

Drug metabolism generates compounds that are usually more polar and, hence, more readily excreted than parent drug. Metabolism takes place predominantly in the liver but can occur at other sites such as kidney, intestinal epithelium, lung, and plasma. Phase I metabolism involves chemical modification, most often oxidation accomplished by members of the cytochrome P450 (CYP) monooxygenase superfamily. CYPs and other molecules that are especially important for drug metabolism are presented in **Table 71-1**, and each drug may be a substrate for one or more of these enzymes. Phase II metabolism involves conjugation of specific endogenous compounds to drugs or their metabolites. The enzymes that accomplish phase II reactions include glucuronyl-, acetyl-, sulfo-, and methyltransferases. Drug metabolites may exert important pharmacologic activity, as discussed further below. Therapeutic antibodies are very slowly eliminated (allowing infrequent dosing, e.g., monthly injections), probably by lysosomal uptake and degradation.

Clinical Implications of Reduced Bioavailability Some drugs undergo near-complete presystemic metabolism and thus cannot be administered orally. Nitroglycerin cannot be used orally because it is

TABLE 71-1 Molecular Pathways Mediating Drug Disposition

ENZYME	SUBSTRATES[a]	INHIBITORS[a]
CYP3A	Calcium channel blockers	Amiodarone
	Antiarrhythmics (lidocaine, quinidine, mexiletine)	Ketoconazole, itraconazole
	HMG-CoA reductase inhibitors ("statins"; see text)	Erythromycin, clarithromycin
	Cyclosporine, tacrolimus	Ritonavir
	Indinavir, saquinavir, ritonavir	Gemfibrozil and other fibrates
CYP2D6[b]	Timolol, metoprolol, carvedilol	Bupropion
	Propafenone, flecainide	Quinidine (even at ultra-low doses)
	Tricyclic antidepressants	Tricyclic antidepressants
	Fluoxetine, paroxetine	Fluoxetine, paroxetine
CYP2C9[b]	Warfarin	Amiodarone
	Phenytoin	Fluconazole
	Glipizide	Phenytoin
	Losartan	
CYP2C19[b]	Omeprazole	Omeprazole
	Mephenytoin	Ritonavir
	Clopidogrel	Fluoxetine
		Fluvoxamine
CYP2B6[b]	Efavirenz	Ticlopidine
Thiopurine S-methyltransferase[b]	6-Mercaptopurine, azathioprine	
N-acetyltransferase[b]	Isoniazid	
	Procainamide	
	Hydralazine	
	Some sulfonamides	
UGT1A1[b]	Irinotecan	
Pseudocholinesterase[b]	Succinylcholine	
TRANSPORTER	**SUBSTRATES[a]**	**INHIBITORS[a]**
P-glycoprotein	Digoxin	Quinidine
	HIV protease inhibitors	Amiodarone
	Many CYP3A substrates	Verapamil
		Cyclosporine
		Itraconazole
		Erythromycin
SLCO1B1[b]	Simvastatin and some other statins	

[a]Examples are presented. Inhibitors affect the molecular pathway and thus may decrease substrate metabolism. [b]Clinically important genetic variants described; **see Chap. 72**.

Note: An extensive listing of CYP substrates, inhibitors, and inducers is maintained at *https://drug-interactions.medicine.iu.edu/MainTable.aspx*.

completely extracted prior to reaching the systemic circulation. The drug is, therefore, used by the sublingual, transdermal, or intravascular routes, which bypass presystemic metabolism.

Some drugs with very extensive presystemic metabolism can still be administered by the oral route, using much higher doses than those required intravenously. Thus, a typical intravenous dose of verapamil is 1–5 mg, compared to a usual single oral dose of 40–120 mg. Administration of low-dose aspirin can result in exposure of cyclooxygenase in platelets in the portal vein to the drug, but systemic sparing because of first-pass aspirin deacylation in the liver. This is an example of presystemic metabolism being exploited to therapeutic advantage.

PLASMA HALF-LIFE

Most pharmacokinetic processes, such as elimination, are first-order; that is, the rate of the process depends on the amount of drug present. Elimination can occasionally be zero-order (fixed amount eliminated per unit time), and this can be clinically important (see "Principles of Dose Selection," later in this chapter). In the simplest pharmacokinetic model (Fig. 71-2*A*), a drug bolus is administered instantaneously to a central compartment, from which drug elimination occurs as a first-order process. *Half-life* is the time required for 50% of a first-order process to be completed. Thus, 50% of drug elimination is achieved after one drug-elimination half-life, 75% after two, 87.5% after three, etc. In practice, first-order processes such as elimination are near-complete after four to five half-lives.

In some cases, drug is removed from the central compartment not only by elimination but also by distribution into peripheral compartments. In this case, the plot of plasma concentration versus time after a bolus may demonstrate two (or more) exponential components (Fig. 71-2*B*). In general, the initial rapid drop in drug concentration represents not elimination but drug distribution into and out of peripheral tissues (also first-order processes), while the slower component represents drug elimination; the initial precipitous decline is usually evident with administration by intravenous but not by other routes. Drug concentrations at peripheral sites are determined by a balance between drug distribution to and redistribution from those sites, as well as by elimination. Once distribution is near-complete (four to five distribution half-lives), plasma and tissue concentrations decline in parallel.

Clinical Implications of Half-Life Measurements

The elimination half-life not only determines the time required for drug concentrations to fall to near-unmeasurable levels after a single bolus, but it is also the sole determinant of the time required for steady-state plasma concentrations to be achieved after any change in drug dosing **(Fig. 71-4)**. This applies to the initiation of chronic drug therapy (whether by multiple oral doses or by continuous intravenous infusion), a change in chronic drug dose or dosing interval, or discontinuation of drug.

Steady state describes the situation during chronic drug administration when the amount of drug administered per unit time equals drug eliminated per unit time. With a continuous intravenous infusion, plasma concentrations at steady state are stable, while with chronic oral

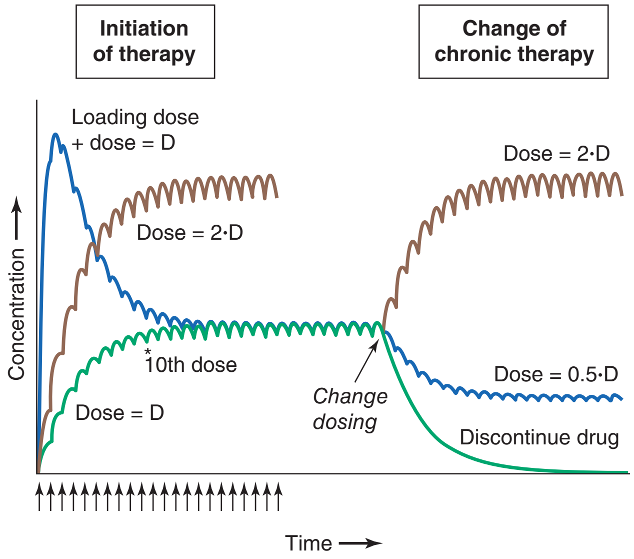

FIGURE 71-4 Drug accumulation to steady state. In this simulation, drug was administered (arrows) at intervals = 50% of the elimination half-life. Steady state is achieved during initiation of therapy after ~5 elimination half-lives, or 10 doses. A loading dose did not alter the eventual steady state achieved. A doubling of the dose resulted in a doubling of the steady state but the same time course of accumulation. Once steady state is achieved, a change in dose (increase, decrease, or drug discontinuation) results in a new steady state in ~5 elimination half-lives. *(Reproduced with permission from DM Roden, in DP Zipes, J Jalife [eds]: Cardiac Electrophysiology: From Cell to Bedside, 4th ed. Philadelphia, Saunders, 2003.)*

drug administration, plasma concentrations vary during the dosing interval, but the time-concentration profile between dosing intervals is stable (Fig. 71-4).

■ DRUG DISTRIBUTION

In some cases, pharmacologic effects require drug distribution to peripheral sites. In this instance, the time course of drug delivery to and removal from these sites determines the time course of drug effects; anesthetic uptake into the central nervous system (CNS) is an example.

Loading Doses For some drugs, the indication may be so urgent that administration of "loading" dosages is required to achieve rapid elevations of drug concentration and therapeutic effects earlier than with chronic maintenance therapy (Fig. 71-4). Nevertheless, the time required for a true steady state to be achieved is still determined only by the elimination half-life.

Rate of Intravenous Drug Administration Although the simulations in Fig. 71-2 use a single intravenous bolus, this is usually inappropriate in practice because side effects related to transiently very high concentrations can result. Rather, drugs are more usually administered orally or as a slower intravenous infusion.

Transiently high drug concentrations after rapid intravenous administration can occasionally be used to advantage. The use of midazolam for intravenous sedation, for example, depends upon its rapid uptake by the brain during the distribution phase to produce sedation quickly, with subsequent egress from the brain during the redistribution of the drug as equilibrium is achieved.

Similarly, adenosine must be administered as a rapid bolus in the treatment of reentrant supraventricular tachycardias **(Chap. 253)** to prevent elimination by very rapid ($t_{1/2}$ of seconds) uptake into erythrocytes and endothelial cells before the drug can reach its clinical site of action, the atrioventricular node.

Clinical Implications of Altered Protein Binding Many drugs circulate in the plasma partly bound to plasma proteins. Since only unbound (free) drug can distribute to sites of pharmacologic action, drug response is related to the free rather than the total circulating plasma drug concentration. In chronic kidney or liver disease, protein binding may be decreased and thus drug actions increased. In some situations (myocardial infarction, infection, surgery), acute phase reactants transiently increase binding of some drugs and thus decrease efficacy. These changes assume the greatest clinical importance for drugs that are highly protein-bound since even a small change in protein binding can result in large changes in free drug; for example, a decrease in binding from 99 to 98% doubles the free drug concentration from 1 to 2%. For some drugs (e.g., phenytoin), monitoring free rather than total drug concentrations can be useful.

■ DRUG ELIMINATION

Drug elimination reduces the amount of drug in the body over time by metabolism or excretion. An important approach to quantifying this reduction is to consider that drug concentrations at the beginning and end of a time period are unchanged, and that a specific volume of the body has been "cleared" of the drug during that time period. This defines clearance as volume/time.

Clinical Implications of Altered Clearance Genetic variants, drug interactions, or diseases that reduce the activity of drug-metabolizing enzymes or excretory mechanisms lead to decreased clearance and, hence, a requirement for a downward dose adjustment to avoid toxicity. Conversely, some drug interactions and genetic variants increase the function of drug elimination pathways, and hence, increased drug dosage is necessary to maintain a therapeutic effect. Metabolites may produce effects similar to, overlapping with, or distinct from those of the parent drug.

Prodrugs are inactive compounds that require metabolism to generate active metabolites that mediate the drug effects. Examples include many angiotensin-converting enzyme (ACE) inhibitors, the angiotensin receptor blocker losartan, the antineoplastic irinotecan, the antiestrogen tamoxifen, the analgesic codeine (whose active metabolite morphine underlies the opioid effect during codeine administration), and the antiplatelet drug clopidogrel. Drug metabolism has also been implicated in bioactivation of procarcinogens and in the generation of reactive metabolites that mediate certain ADRs.

■ THE CONCEPT OF HIGH-RISK PHARMACOKINETICS

When plasma concentrations of active drug depend exclusively on a single metabolic pathway, any condition that inhibits that pathway (be it disease related, genetic, or due to a drug interaction) can lead to dramatic changes in drug concentrations and marked variability in drug action. Two scenarios can generate highly variable drug concentrations and effects through such "high-risk pharmacokinetics." *First*, variability in bioactivation of a prodrug can lead to striking variability in drug action; examples include decreased CYP2D6 activity, which prevents analgesia by codeine, and decreased CYP2C19 activity, which reduces the antiplatelet effects of clopidogrel. The s*econd* scenario is administration of an active drug whose elimination relies on a single pathway. In this case, inhibition of the elimination pathway by genetic variants or by administration of inhibiting drugs leads to marked elevation of drug concentration and, for drugs with a narrow therapeutic window, an increased likelihood of dose-related toxicity. When drugs undergo elimination by multiple-drug metabolizing or excretory pathways, absence of one pathway (due to a genetic variant or drug interaction) is much less likely to have a large impact on drug concentrations or drug actions.

■ PRINCIPLES OF PHARMACODYNAMICS

Time Course of Drug Action Pharmacokinetic parameters, such as half-life and clearance, explain drug concentrations over time, but understanding the action of a drug over time (pharmacodynamics) often requires an understanding of its precise mechanism of action. Drugs exert therapeutic or adverse effects by interacting with drug target molecules, often in specific tissues, and with a cascade of downstream consequences. For drugs used in urgent treatment (e.g., vascular thrombosis, shock, status epilepticus), little or no delay is anticipated (or desired) between the administration of the drug, the drug-target interaction, and the development of a clinical effect.

For many conditions, the indication for therapy is less urgent, and a delay in the onset of action clinically acceptable. Delay can be due to pharmacokinetic mechanisms such as slow elimination (resulting in slow accumulation to steady state), slow uptake into the target tissue, or slow accumulation of active metabolites. In addition, effects such as platelet inhibition, relief of depression, or control of hypertension may be delayed because the drug's interaction with its receptor is only the first step in mediating clinical drug actions. Thus, for example, the elimination half-life of clopidogrel in plasma in most subjects is ~6 h, but the antiplatelet effect (which is due to irreversible binding to P2Y12 receptors) persists for the life of the platelet, 7–10 days.

Drug Effects May Be Disease Specific A drug may produce no action or a different spectrum of actions in unaffected individuals compared to patients with underlying disease. Further, concomitant disease can complicate interpretation of response to drug therapy, especially ADRs. For example, increasing dyspnea in a patient with chronic lung disease receiving amiodarone therapy could be due to the drug, underlying disease, or an intercurrent cardiopulmonary problem. As a result, alternate antiarrhythmic therapies may be preferable in patients with chronic lung disease.

While drugs interact with specific molecular receptors, drug effects may vary over time, even if stable drug and metabolite concentrations are maintained. The drug-receptor interaction occurs in a complex biologic milieu that can vary to modulate the drug effect. For example, ion channel blockade by drugs, an important anticonvulsant and antiarrhythmic effect, is often modulated by membrane potential, itself a function of factors such as extracellular potassium or local ischemia. Receptors may be up- or downregulated by disease or by the drug itself. For example, β-adrenergic blockers upregulate β-receptor density during chronic therapy. While this effect does not usually result in resistance to the therapeutic effect of the drugs, it may produce severe agonist-mediated effects (e.g., hypertension or tachycardia) if the blocking drug is abruptly withdrawn.

As molecular mechanisms of disease become better defined, drugs targeting those mechanisms have been introduced into practice. Antineoplastic agents targeting mutant kinases overexpressed in cancers and HMG-CoA reductase inhibitors (statins) and PCSK9 inhibitors for hypercholesterolemia are examples.

PRINCIPLES OF DOSE SELECTION

The desired goal of therapy with any drug is to maximize the likelihood of a beneficial effect while minimizing the risk of ADRs. Previous experience with the drug, in controlled clinical trials or in postmarketing use, defines the relationships between dose or plasma concentration and these dual effects (Fig. 71-1) and has important implications for initiation of drug therapy:

1. *The target drug effect should be defined when drug treatment is started.* With some drugs, the desired effect may be difficult to measure objectively, or the onset of efficacy can be delayed for weeks or months; drugs used in the treatment of cancer and psychiatric disease are examples. Sometimes a drug is used to treat a symptom, such as pain or palpitations, and here it is the patient who will report whether the selected dose is effective. In yet other settings, such as anticoagulation or hypertension, the desired response can be repeatedly and objectively assessed by simple clinical or laboratory tests.
2. *The nature of anticipated toxicity often dictates the starting dose.* If side effects are minor, it may be acceptable to start chronic therapy at a dose highly likely to achieve efficacy and down-titrate if side effects occur. However, this approach is rarely, if ever, justified if the anticipated toxicity is serious or life-threatening; in this circumstance, it is more appropriate to initiate therapy with the lowest dose that may produce a desired effect. In cancer chemotherapy, it is common practice to use maximally tolerated doses.
3. *The above considerations do not apply if these relationships between dose and effects cannot be defined.* This is especially relevant to some ADRs (discussed further below) whose development is not readily related to drug dose.
4. *If a drug dose does not achieve its desired effect, a dosage increase is justified only if toxicity is absent and the likelihood of serious toxicity is small.*

Failure of Efficacy Even assuming the diagnosis is correct and the correct drug and dose are prescribed, drugs may fail to be effective. A complete therapeutic response is often absent with antihypertensive or antidepressant drugs, and a major challenge in contemporary therapeutics is to identify patient-specific predictors of response to individual drugs. Other explanations for failure of efficacy include drug interactions, decreased adherence, or unexpectedly low drug concentration due to administration of expired or degraded drug. These are situations in which measurement of plasma drug concentrations, if available, can be especially useful. Adherence is an especially frequent problem in the chronic treatment of diseases such as hypertension or HIV infection. Multidrug regimens with multiple doses per day are especially prone to decreased adherence.

Monitoring response to therapy, by physiologic measures or by plasma concentration measurements, requires an understanding of the relationships between plasma concentration and anticipated effects. For example, measurement of QT interval is used during treatment with sotalol or dofetilide to avoid marked QT prolongation that can herald serious arrhythmias. In this setting, evaluating the electrocardiogram at the time of anticipated peak plasma concentration and effect (e.g., 1–2 h postdose at steady state) is most appropriate. Maintained high vancomycin levels carry a risk of nephrotoxicity, so dosages should be adjusted on the basis of plasma concentrations measured at trough (predose). Similarly, for dose adjustment of other drugs (e.g., anticonvulsants), concentration should be measured at its lowest during the dosing interval, just prior to a dose at steady state (Fig. 71-4), to ensure a maintained therapeutic effect.

Concentration of Drugs in Plasma as a Guide to Therapy Therapeutic drug monitoring can be useful (Fig. 71-1) with certain types of drugs including many anticonvulsants, antirejection agents, antiarrhythmics, and antibiotics. By contrast, if no such relationship can be established (e.g., if drug access to important sites of action outside plasma is highly variable), monitoring plasma concentration may not provide an accurate guide to therapy **(Fig. 71-5)**.

The common situation of first-order elimination implies that average, maximum, and minimum steady-state concentrations are related linearly to the dosing rate. Accordingly, the maintenance dose may be adjusted on the basis of the ratio between the desired and measured concentrations *at steady state*; for example, if a doubling of the steady-state plasma concentration is desired, the dose should be doubled. This does not apply to drugs eliminated by zero-order kinetics (fixed amount per unit time), where small dosage increases will produce disproportionate increases in plasma concentration; examples include aspirin and fluoxetine.

If an increase in dosage is needed, this is usually best achieved by increasing the drug dose and leaving the dosing interval constant (e.g., by giving 200 mg every 8 h instead of 100 mg every 8 h); occasionally the dosing interval is shortened to avoid high peak concentrations recognizing that adherence is decreased with regimens requiring frequent dosing.

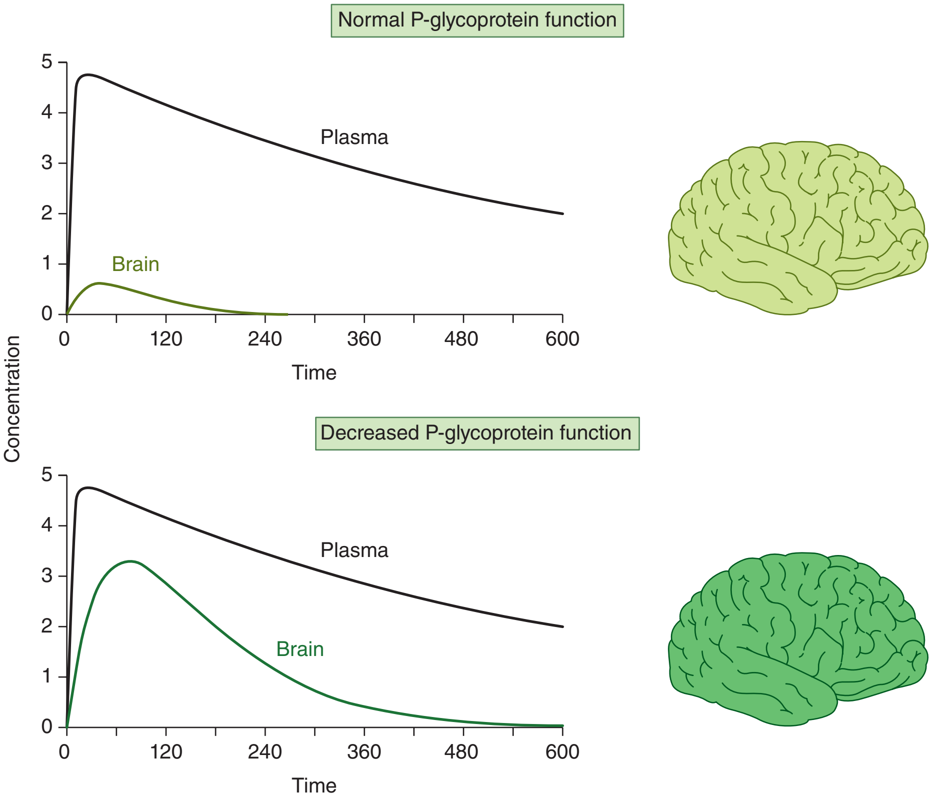

FIGURE 71-5 Drug concentrations in specific tissues may not always parallel those in plasma. For example, the efflux pump P-glycoprotein excludes drugs from the endothelium of capillaries in the brain and so constitutes a key element of the blood-brain barrier. Reduced P-glycoprotein function (e.g., due to drug interactions) can thus increase penetration of substrate drugs into the brain, even when plasma concentrations are unchanged.

EFFECTS OF DISEASE ON DRUG CONCENTRATION AND RESPONSE

RENAL DISEASE

If a drug or its metabolites are primarily excreted through the kidneys and increased drug levels are associated with ADRs (an example of "high-risk pharmacokinetics" described above), drug dosages must be reduced in patients with renal dysfunction to avoid toxicity. Digoxin is one example; another is the antiarrhythmics dofetilide and sotalol, which undergo predominant renal excretion and carry a risk of QT prolongation and arrhythmias if doses are not reduced in renal disease. At approved doses, the anticoagulant edoxaban appears to be somewhat more effective in subjects with mild renal dysfunction, possibly reflecting higher drug levels. The narcotic analgesic meperidine undergoes extensive hepatic metabolism, so that renal failure has little effect on its plasma concentration. However, its metabolite, normeperidine, does undergo renal excretion, accumulates in renal failure, and probably accounts for the signs of CNS excitation, such as irritability, twitching, and seizures, that appear when multiple doses of meperidine are administered to patients with renal disease. Protein binding of some drugs (e.g., phenytoin) may be altered in uremia, so measuring free drug concentration may be desirable.

In practice, most decisions involving dosing adjustment in patients with renal failure use published recommended adjustments in dosage or dosing interval based on the severity of renal dysfunction indicated by creatinine clearance. Any such modification of dose is a first approximation and should be followed by plasma concentration data (if available) and clinical observation to further optimize therapy for the individual patient.

LIVER DISEASE

Standard tests of liver function are not useful in adjusting doses in diseases like hepatitis or cirrhosis. First-pass metabolism may decrease, leading to increased oral bioavailability as a consequence of disrupted hepatocyte function, altered liver architecture, and portacaval shunts. The oral bioavailability for high first-pass drugs such as morphine, meperidine, midazolam, and nifedipine is almost doubled in patients with cirrhosis, compared to those with normal liver function. Therefore, the size of the oral dose of such drugs should be reduced in this setting.

HEART FAILURE AND SHOCK

Under conditions of decreased tissue perfusion, the cardiac output is redistributed to preserve blood flow to the heart and brain at the expense of other tissues **(Chap. 264)**. As a result, drugs may be distributed into a smaller volume of distribution, higher drug concentrations will be present in the plasma, and the tissues that are best perfused (the brain and heart) will be exposed to these higher concentrations, resulting in increased CNS or cardiac effects. In addition, decreased perfusion of the kidney and liver may impair drug clearance. Another consequence of severe heart failure is decreased gut perfusion, which may reduce drug absorption and thus lead to reduced or absent effects of orally administered therapies.

DRUG USE IN THE ELDERLY

In the elderly, multiple pathologies and medications used to treat them result in more drug interactions and ADRs. Aging also results in changes in organ function, especially of the organs involved in drug disposition. Initial doses should be less than the usual adult dosage and should be increased slowly. The number of medications, and doses per day, should be kept as low as possible.

ADRs are especially common in the elderly because of altered pharmacokinetics and pharmacodynamics, the frequent use of multidrug regimens, and concomitant disease. For example, use of long half-life benzodiazepines is linked to the occurrence of hip fractures in elderly patients, likely reflecting both a risk of falls from these drugs (due to increased sedation) and the increased incidence of osteoporosis in elderly patients.

DRUG USE IN CHILDREN

Drug metabolism and drug response pathways mature at different rates after birth, and the relative size of various body compartments and function of various organs change during development. However, there are few studies providing solid evidence to guide pediatric dosing, so in practice, doses are adjusted for size (weight or body surface area) as a first approximation unless age-specific data are available.

INTERACTIONS BETWEEN DRUGS

Drug interactions can complicate therapy by increasing or decreasing the action of a drug; interactions may be based on changes in drug disposition or in drug response in the absence of changes in drug levels **(Table 71-2)**. *Interactions must be considered in the differential diagnosis of any unusual response occurring during drug therapy*. Prescribers should recognize that patients often come to them with a legacy of drugs acquired during previous medical experiences, often with multiple physicians who may not be aware of all the patient's medications. A meticulous drug history should list all medications, including agents not often volunteered during questioning, such as OTC drugs, health food supplements, and topical agents such as eye drops. Lists of interactions are available from a number of electronic sources. While it is unrealistic to expect the practicing physician to memorize these, certain drugs consistently run the risk of generating interactions, often by inhibiting or inducing specific drug elimination pathways; these include CYP2D6, CYP3A, and P-glycoprotein inhibitors (Table 71-1) and CYP3A/P-glycoprotein inducers (Table 71-2). Accordingly, when these drugs are started or stopped, prescribers must be especially alert to the possibility of interactions.

ADVERSE DRUG REACTIONS

The morbidity and mortality from ADRs may present diagnostic problems because they can involve every organ and system of the body and may be mistaken for signs of underlying disease. In addition, drug therapy for chronic conditions such as psychiatric disease or hypertension does not achieve the desired goal in up to half of treated patients; thus, the most common "adverse" drug effect may be failure of efficacy.

ADRs can be classified in two broad groups. Type A reactions result from exaggeration of an intended pharmacologic action of the drug, such as increased bleeding with anticoagulants or bone marrow suppression with some antineoplastics, and these tend to be dose-dependent. Type B reactions result from toxic effects unrelated to the intended pharmacologic actions. The latter effects are often unanticipated (especially with new drugs) and frequently severe and may result from recognized (often immunologic) as well as previously undescribed mechanisms. Type B reactions may occur at low dosages and are often termed dose-independent.

Prior to regulatory approval and marketing, new drugs are tested in relatively few patients who tend to be less sick and to have fewer concomitant diseases than those patients who subsequently receive the drug therapeutically. Because of the relatively small number of patients studied in clinical trials and the selected nature of these patients, rare serious ADRs are generally not detected prior to a drug's approval; indeed, if they are detected, the new drugs are generally not approved. Therefore, physicians need to be cautious in the prescription of new drugs and alert for the appearance of previously unrecognized ADRs.

Elucidating mechanisms underlying ADRs can assist development of safer compounds or allow a patient subset at especially high risk to be excluded from drug exposure. National adverse reaction reporting systems, such as those operated by the Food and Drug Administration (suspected ADRs can be reported online at *https://www.fda.gov/safety/medwatch-fda-safety-information-and-adverse-event-reporting-program*) and the Committee on Safety of Medicines in Great Britain, can prove useful.

Occasionally, "adverse" effects may be exploited to develop an entirely new indication for a drug. Unwanted hair growth during minoxidil treatment of severely hypertensive patients led to development of the drug for hair growth. Sildenafil was initially developed as an antianginal, but its effects to alleviate erectile dysfunction not only led to a new drug indication but also to increased understanding of the role of type 5 phosphodiesterase in erectile tissue.

TABLE 71-2 Drug Interactions

MECHANISM	EXAMPLE
Pharmacokinetic Interactions Causing Decreased Drug Effect	
Decreased absorption due to drug binding in the gut	Antacids or bile acid sequestrants decrease the absorption of many drugs: Antacids/tetracyclines Cholestyramine/digoxin
Decreased solubility due to altered gastric pH	H_2 receptor blockers or proton pump inhibitors decrease solubility and absorption of weak bases: Omeprazole/ketoconazole
Induction of drug metabolism and/or drug transport: Rifampin Carbamazepine Phenytoin St. John's wort Glutethimide (also smoking, exposure to chlorinated insecticides, and chronic alcohol ingestion)	Decreased concentrations and effects of: Warfarin Quinidine Cyclosporine Losartan Oral contraceptives Methadone Dabigatran
Decreased prodrug bioactivation	Proton pump inhibitors may prevent clopidogrel bioactivation CYP2D6 inhibitors (fluoxetine, paroxetine, quinidine, and others) may prevent codeine bioactivation
Reduced delivery of drug to active sites of action	Tricyclics prevent clonidine uptake into adrenergic neurons, preventing antihypertensive effects
Pharmacokinetic Interactions Causing Increased Drug Effect	
Inhibited drug metabolism	Cimetidine (inhibits many CYPs): Warfarin Theophylline Phenytoin CYP2D6 inhibitors[a]/β blockers CYP3A inhibitors[a]: HMG-CoA reductase inhibitors Colchicine (toxicity risk) Decreased cyclosporine dose requirement
Inhibited drug transport	Amiodarone (inhibits many CYPs and P-glycoprotein): Warfarin Digoxin Dabigatran
Inhibition of drug metabolism causing accumulation of toxic metabolites	Allopurinol (xanthine oxidase inhibitor) inhibits an alternate pathway for azathioprine and 6-mercaptopurine elimination, increasing risk for toxicity
Decreased elimination due to altered renal function	Inhibitors of renal tubular transport (phenylbutazone, probenecid, salicylates) increase methotrexate toxicity
Pharmacodynamic Drug Interactions	
Combined effects on the same biologic process	Excess bleeding with combinations of antiplatelet drugs, anticoagulants, and NSAIDs Long QT–related arrhythmias with QT-prolonging antiarrhythmics plus diuretics Hyperkalemia with ACE inhibitors plus potassium Hypotension with nitrates plus sildenafil
Antagonistic effects on the same biologic process	Loss of antihypertensive drug effects with NSAIDs

[a]See Table 71-1.

Abbreviations: ACE, angiotensin-converting enzyme; CYP, cytochrome P; NSAID, nonsteroidal anti-inflammatory drug.

SCOPE OF THE ADVERSE DRUG REACTION PROBLEM

One estimate in the United Kingdom was that 6.5% of all hospital admissions are due to ADRs and that 2.3% of these patients (0.15%) died as a result. The most common culprit drugs were aspirin, nonsteroidal anti-inflammatory drugs, diuretics, warfarin, ACE inhibitors, antidepressants, opiates, digoxin, steroids, and clopidogrel. One study in the late 1990s suggested that ADRs were responsible for >100,000 in-hospital deaths in the United States, making them the fourth to sixth most common cause of in-hospital death. Another study 10 years later showed no change in this trend.

Serious ADRs are also well recognized with "herbal" remedies and OTC compounds; examples include kava-associated hepatotoxicity, L-tryptophan-associated eosinophilia-myalgia, and phenylpropanolamine-associated stroke, each of which has caused fatalities.

TOXICITY UNRELATED TO A DRUG'S PRIMARY PHARMACOLOGIC ACTIVITY

Drugs or, more commonly, reactive metabolites generated by CYPs can covalently bind to tissue macromolecules (e.g., proteins or DNA) to cause tissue toxicity. Because of the reactive nature of these metabolites, covalent binding often occurs close to the site of production, typically the liver.

Acetaminophen A common cause of drug-induced hepatotoxicity is acetaminophen overdosage (Chap. 351). Normally, reactive metabolites are detoxified by combining with hepatic glutathione. When glutathione becomes depleted, the metabolites bind instead to hepatic protein, with resultant hepatocyte damage. The hepatic necrosis produced by the ingestion of acetaminophen can be prevented or attenuated by the administration of substances such as *N*-acetylcysteine that reduce the binding of electrophilic metabolites to hepatic proteins. The risk of acetaminophen-related hepatic necrosis is increased in patients receiving drugs such as phenobarbital or phenytoin, which increase the rate of drug metabolism, or ethanol, which exhausts glutathione stores.

Immunologic Reactions Generation of an immune response to a drug often requires in vivo activation and covalent linkage to protein, carbohydrate, or nucleic acid. Drug stimulation of antibody production may mediate tissue injury by several mechanisms. The antibody may attack the drug when the drug is covalently attached to a cell and thereby destroy the cell. This occurs in penicillin-induced hemolytic anemia. Antibody-drug-antigen complexes may be passively adsorbed by a bystander cell, which is then destroyed by activation of complement; this occurs in quinine- and quinidine-induced thrombocytopenia. Heparin-induced thrombocytopenia arises when antibodies against complexes of platelet factor 4 peptide and heparin generate immune complexes that activate platelets; thus, the thrombocytopenia is accompanied by "paradoxical" thrombosis and is treated with thrombin inhibitors. Drugs or their reactive metabolites may alter a host tissue, rendering it antigenic and eliciting autoantibodies; hydralazine- or procainamide-induced lupus erythematosus is an example. Drug-induced pure red cell aplasia (Chap. 107) is due to an immune-based drug reaction. Serum sickness (Chap. 363) results from the deposition of circulating drug-antibody complexes on endothelial surfaces. Many drugs, particularly antimicrobial agents, ACE inhibitors, and aspirin, can elicit anaphylaxis with production of IgE, which binds to mast cell membranes. Contact with a drug antigen initiates a series of biochemical events in the mast cell and results in the release of mediators that can produce the characteristic urticaria, wheezing, flushing, rhinorrhea, and (occasionally) hypotension.

Drugs may also elicit cell-mediated immune responses. One serious reaction is Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN), which can result in death due to T-cell-mediated massive skin sloughing. Another probable immune-mediated drug reaction is the DRESS (drug reaction with eosinophilia and systemic symptoms) syndrome, a rare ADR with a chronic relapsing course, often triggered by antiseizure medications and possibly arising from herpes virus reactivation. As described in Chap. 72, specific genetic variants appear necessary but not sufficient to elicit SJS/TEN or DRESS.

While the use of antibodies targeting immune checkpoints is dramatically improving prognosis in many cancers, these agents have also been associated with the unpredictable development of many apparently immune-related ADRs. Some, like colitis or thyroiditis, may be self-limited or medically manageable, while others, notably myocarditis, are rarer but can be rapidly fatal.

DIAGNOSIS AND TREATMENT OF ADVERSE DRUG REACTIONS

A suspected ADR developing after introduction of a new drug naturally implicates that drug. It is also important to remember that a drug interaction may be responsible. Thus, for example, a patient on a chronic stable warfarin dose may develop a bleeding complication after introduction of amiodarone; this does not reflect a direct reaction to amiodarone but rather its effect to inhibit warfarin metabolism.

Adverse effects such as hypoglycemia with insulin or bleeding with anticoagulants are more readily related to a specific drug than are nonspecific symptoms such as rash; drug fever often escapes initial diagnosis because fever is such a common manifestation of disease.

Electronic listings of ADRs can be useful. However, exhaustive compilations often provide little sense of perspective in terms of frequency and seriousness, which can vary considerably among patients.

Abnormalities such as G6PD deficiency, serum pseudocholinesterase level, or genotyping may be useful in diagnosis **(Chap. 72)**.

Once an ADR is suspected, discontinuation of the suspected drug followed by disappearance of the reaction is presumptive evidence of a drug-induced illness. Confirming evidence may be sought by cautiously reintroducing the drug and seeing if the reaction reappears. However, rechallenge should be done only if the suspected culprit drug is critical to the patient's care. When the reaction is thought to be immunologic, challenge is generally avoided. Testing for genetic abnormalities (e.g., G6PD deficiency, low serum pseudocholinesterase level, or genotyping; **Chap. 72**) may be useful in diagnosis.

Serious immunologically mediated ADRs have been treated with high-dose steroids; other agents such as rituximab, infliximab, abatacept, or mycophenolate mofetil, as well as plasmapheresis, have been used with variable success.

If the patient is receiving many drugs when an ADR is suspected, the drugs likeliest to be responsible can usually be identified; this should include both potential culprit agents as well as drugs that alter their elimination. All drugs may be discontinued at once or, if this is not practical, discontinued one at a time, starting with the ones most suspect, and the patient observed for signs of improvement. The time needed for a concentration-dependent ADR to disappear depends on the time required for the concentration to fall below the range associated with the ADR; that, in turn, depends on the initial blood level and on the rate of elimination or metabolism of the drug. Adverse effects of drugs with long half-lives or those not directly related to serum concentration may take a considerable time to disappear.

THE DRUG DEVELOPMENT PROCESS

Drug therapy is an ancient feature of human culture. The first treatments were plant extracts discovered empirically to be effective for indications like fever, pain, or breathlessness. This symptom-based empiric approach to drug development was supplanted in the twentieth century by identification of compounds targeting more fundamental biologic processes, such as bacterial growth or elevated blood pressure. The term "magic bullet," coined by Paul Ehrlich to describe the search for effective compounds for syphilis, captures the essence of the hope that understanding basic biologic processes will lead to highly effective new therapies.

A common starting point for the development of many widely used modern therapies has been basic biologic discovery that implicates potential target molecules: examples of such target molecules include HMG-CoA reductase, a key step in cholesterol biosynthesis, or the *BRAF* V600E mutation that appears to drive the development of some malignant melanomas and other tumors. The development of compounds targeting these molecules has not only revolutionized treatment for diseases such as hypercholesterolemia or malignant melanoma but also revealed new biologic features of disease. Thus, for example, initial spectacular successes with vemurafenib (which targets *BRAF* V600E) were followed by near-universal tumor relapse, strongly suggesting that inhibition of this pathway alone would be insufficient for tumor control. This reasoning, in turn, supports a view that many complex diseases will not lend themselves to cure by targeting a single magic bullet, but rather single drugs or combinations that attack multiple pathways whose perturbation results in disease. The use of combination therapy in settings such as hypertension, tuberculosis, HIV infection, and many cancers highlights the potential for the "systems biology" view of drug therapy outlined above.

A common approach in contemporary drug development is to start with a high-throughput screening procedure to identify "lead" chemical(s) modulating the activity of a potential drug target. The next step is application of increasingly sophisticated medicinal chemistry-based modification of the "lead" to develop compounds with specificity for the chosen target, lack of "off-target" effects, and pharmacokinetic properties suitable for human use (e.g., consistent bioavailability, long elimination half-life, and no high-risk pharmacokinetic features). Drug evaluation in human subjects then proceeds from initial safety and tolerance (phase 1) to dose finding (phase 2) and then to large efficacy trials (phase 3). This is a very expensive process, and the vast majority of lead compounds fail at some point. Thus, new approaches to identify likely successes and failures early are needed. One idea, described further in **Chap. 72**, is to use genomic and other high-throughput profiling approaches in drug development. This can identify new drug targets and define disease subsets for which drugs approved for other indications might be "repurposed," thereby avoiding the costly development process. In addition, drugs whose development includes supporting human genetic data have a decreased chance of failing in the drug development process.

SUMMARY

Modern clinical pharmacology aims to replace empiricism in the use of drugs with therapy based on in-depth understanding of factors that determine an individual's response to drug treatment. Molecular pharmacology, pharmacokinetics, genetics, clinical trials, and the educated prescriber all contribute to this process. No drug response should ever be termed *idiosyncratic*; all responses have a mechanism whose understanding will help guide further therapy with that drug. This rapidly expanding understanding of variability in drug actions makes the process of prescribing drugs increasingly daunting for the practitioner. However, fundamental principles should guide this process:

- The benefits of drug therapy, however defined, should outweigh the risk.
- The smallest dosage necessary to produce the desired effect should be used.
- The number of medications and doses per day should be minimized.
- Although the literature is rapidly expanding, accessing it is becoming easier; electronic tools to search databases of literature and unbiased opinion will become increasingly commonplace.
- Genetics play a role in determining variability in drug response and is becoming a part of clinical practice **(Chap. 72)**.
- Electronic health records and pharmacy systems will increasingly incorporate prescribing advice, such as indicated medications not used; unindicated medications being prescribed; and potential dosing errors, drug interactions, or genetically determined drug responses.
- Prescribers should be particularly wary when adding or stopping specific drugs that are especially liable to provoke interactions and adverse drug reactions.
- Prescribers should use only a limited number of drugs, with which they are thoroughly familiar.

FURTHER READING

Chothe PP et al: Drug transporters in drug disposition: The year 2022 in review. Drug Metab Rev 55:343, 2023.

Holford N: Pharmacodynamic principles and the time course of immediate drug effects. Transl Clin Pharmacol 25:157, 2017.

Levy RH, Isabelle Ragueneau-Majlessi I: Past, present, and future of drug-drug interactions. Clin Pharmacol Ther 105:1286, 2019.

McColl ER et al: The age of omics-driven precision medicine. Clin Pharmacol Ther 106:477, 2019.

Sultana J et al: Clinical and economic burden of adverse drug reactions. J Pharmacol Pharmacother 4:S73, 2013.

Wang RS et al: Multiomics network medicine approaches to precision medicine and therapeutics in cardiovascular diseases. Arterioscler Thromb Vasc Biol 43:493, 2023.

72

Pharmacogenomics

Dan Roden

The previous chapter discussed mechanisms underlying variability in drug action, highlighting pharmacokinetic and pharmacodynamic pathways to beneficial and adverse drug events. Work in the past several decades has defined how genetic variation can play a prominent role in modulating these pathways. Initial studies described unusual drug responses due to single genetic variants in individual subjects, defining the field of pharmacogenetics. A more recent view extends this idea to multiple genetic variants across populations, and the term "pharmacogenomics" is often used. Understanding the role of genetic variation in drug response could improve the use of current drugs, avoid drug use in those at increased risk for adverse drug reactions (ADRs), guide development of new drugs, and even be used as a lens through which to understand mechanisms of diseases themselves. This chapter will outline the principles of pharmacogenomics, currently available evidence that genetic factors play a role in variable drug actions, and areas of controversy and ongoing work. The chapter tables are available online. They can be viewed by opening the table of contents of *Harrison's* 22nd edition at *accessmedicine.com/harrisons*.

■ PRINCIPLES OF GENETIC VARIATION AND DRUG RESPONSE (SEE ALSO CHAPS. 479 AND 480)

A goal of traditional Mendelian genetics is to identify DNA variants associated with a distinct phenotype in multiple related family members **(Chap. 480)**. However, it is unusual for a drug response phenotype to be accurately measured in more than one family member, let alone across a kindred. Some clinical studies have examined drug disposition traits (such as urinary drug excretion after a fixed test dose) in twins and have, in some instances, shown greater concordance in monozygotic compared to dizygotic pairs, supporting a genetic contribution to the trait under study. However, in general, non–family-based approaches are usually used to identify and validate DNA variants contributing to variable drug actions. Both candidate gene and genome-wide studies have been used, and as with any genomic study, results require replication before they should be accepted as valid.

Types of Genetic Variants Influencing Drug Response (Table e72-1) The most common type of genetic variant is a single nucleotide polymorphism (SNP), and nonsynonymous SNPs (i.e., those that alter primary amino acid sequence encoded by a gene) are a common cause of variant function in genes regulating drug responses, often termed *pharmacogenes*. Small insertions and deletions can similarly alter protein function or lead to functionally important splice variation. Examples of synonymous coding region variants altering pharmacogene function have also been described; postulated mechanisms include an altered rate of RNA translation and thus altered folding of the nascent protein, or altered splicing. Variation in pharmacogene promoters or in copy number (gene deletion or multiple functional copies of the same gene) is also well described.

Table e72-1 lists examples of individual types of genomic variation and the impact they can have on function of pharmacogenes. Multiple genotyping approaches may be needed to detect important variants; for example, SNP assays may fail to detect large gene duplications, and highly polymorphic regions (such as the major histocompatibility locus on chromosome 6 that includes multiple genes of the human leukocyte antigen [HLA] family) are best evaluated by sequencing.

Table e72-1 also highlights the fact that the frequency of important variation across pharmacogenes can vary strikingly by ancestry, with the result that certain ethnic groups may be at unusually high risk of displaying variant response to specific drugs.

Candidate Gene Approaches Most studies to date have used an understanding of the molecular mechanisms modulating drug action to identify candidate genes in which variants could explain variable drug responses. One very common scenario is that variable drug actions can be attributed to variability in plasma drug concentrations. When plasma drug concentrations vary widely (e.g., more than an order of magnitude), especially if their distribution is non-unimodal as in **Fig. 72-1**, variants in single genes controlling drug concentrations often contribute. In this case, the most obvious candidate genes are those responsible for drug metabolism and elimination. Other candidate genes are those encoding the target molecules with which drugs interact to produce their effects or molecules modulating that response, including those involved in disease pathogenesis.

Genome-Wide Association Studies The field has also had some success with "unbiased" approaches such as genome-wide association (GWA) **(Chap. 479)**, particularly in identifying single variants associated with high risk for certain forms of drug toxicity, and in validating the results of candidate gene studies. GWA studies have identified variants in the HLA locus that are associated with high risk for severe skin rashes during treatment with the anticonvulsant carbamazepine and hepatotoxicity with flucloxacillin, an antibiotic never marketed in the United States. A GWA study of simvastatin-associated myopathy identified a single noncoding SNP in *SLCO1B1*, encoding OATP1B1, a drug transporter known to modulate simvastatin uptake into the liver, which accounts for 60% of myopathy risk. African-American subjects are known to have higher dose requirements to achieve stable anticoagulation with warfarin, due in part to variations in *CYP2C9* and *VKORC1*, discussed below; in GWA studies, variants in these two genes account for up to 50% of variable warfarin effects.

■ GENETIC VARIANTS AFFECTING PHARMACOKINETICS

Clinically important genetic variants have been described in multiple molecular pathways of drug disposition **(Table e72-2)**. A distinct multimodal distribution of drug disposition (as shown in Fig. 72-1) argues for a predominant effect of variants in a single gene in the metabolism of that substrate. Individuals with two alleles (variants) encoding for nonfunctional protein make up one group, often termed *poor metabolizers* (PM phenotype). For most genes, many variants can produce such a loss of function, and assessing whether they are on the same or different alleles (i.e., the *diplotype*) can complicate the use of genotyping in clinical practice. Furthermore, some variants produce only partial loss of function, and the presence of more than one variant may be required to define a specific allele. Individuals with one functional allele, or multiple reduction of function alleles, make up a second group (*intermediate metabolizers*) and may or may not be distinguishable from those with two functional alleles (normal metabolizers, sometimes termed *extensive metabolizers*, EMs). *Ultrarapid metabolizers* (UMs) with especially high enzymatic activity (usually attributed to specific SNPs or to gene duplication; Table e72-1 and Fig. 72-1) have also been described for some traits. Many drugs in widespread use can inhibit specific drug disposition pathways **(see Chap. 71, Table 71-1)**, and so EM individuals receiving such inhibitors can respond like PM patients (*phenocopying*). Polymorphisms in genes encoding drug uptake or drug efflux transporters may be other contributors to variability in drug delivery to target sites and, hence, in

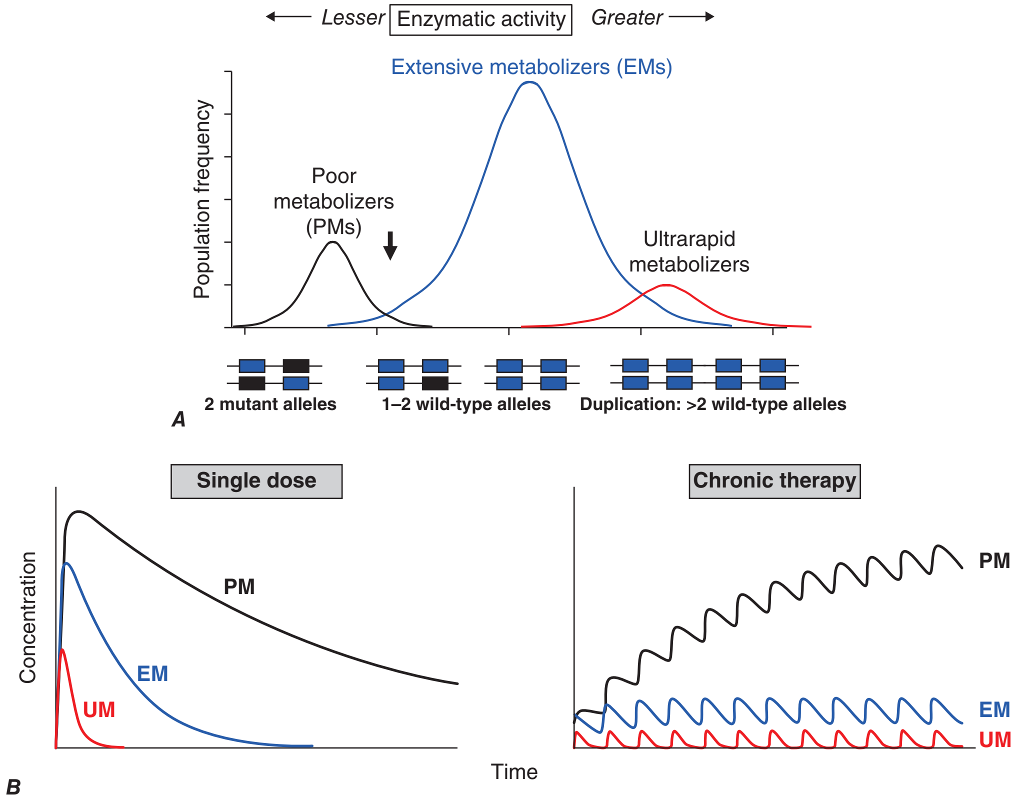

FIGURE 72-1 ***A.*** Distribution of CYP2D6 metabolic activity across a population. The heavy arrow indicates an antimode, separating poor metabolizer subjects (PMs, black), with two loss-of-function CYP2D6 alleles (black), indicated by the intron-exon structures below the chart. Individuals with one or two functional alleles are grouped together as extensive metabolizers (EMs, blue). Also shown are ultra-rapid metabolizers (UMs, red), with 2–12 functional copies of the gene, displaying the greatest enzyme activity. *(Adapted from M-L Dahl et al: J Pharmacol Exp Ther 274:516, 1995.)* ***B.*** These simulations show the predicted effects of CYP2D6 genotype on disposition of a substrate drug. With a single dose (left), there is an inverse "gene-dose" relationship between the number of active alleles and the areas under the time-concentration curves (smallest in UM subjects; highest in PM subjects); this indicates that clearance is greatest in UM subjects. In addition, elimination half-life is longest in PM subjects. The right panel shows that these single-dose differences are exaggerated during chronic therapy: steady-state concentration is much higher in PM subjects (decreased clearance), as is the time required to achieve steady state (longer elimination half-life).

drug effects. Examples of common pharmacogene polymorphisms are described here.

CYP3A Members of the CYP3A family (*CYP3A4*, *CYP3A5*) metabolize the greatest number of drugs in therapeutic use. CYP3A4 activity is highly variable (up to an order of magnitude) among individuals, but nonsynonymous coding region polymorphisms (those that change the encoded amino acid) are unusual. Thus, the underlying mechanism likely reflects genetic variation in regulatory regions.

Most subjects of European or Asian origin carry a polymorphism that disrupts splicing in the closely related *CYP3A5* gene. As a result, these individuals display less CYP3A5 activity compared to subjects of African origin in whom splicing is not disrupted. Decreased efficacy of the antirejection agent tacrolimus in subjects of African origin has been attributed to more rapid CYP3A5-mediated elimination, and a lower risk of vincristine-associated neuropathy has been reported in CYP3A5 "expressers."

CYP2D6 CYP2D6 is second to CYP3A4 in the number of commonly used drugs that it metabolizes. CYP2D6 activity is polymorphically distributed, and 5–10% of European- and African-derived populations (but few Asians) display the PM phenotype (Fig. 72-1). Dozens of loss-of-function variants in *CYP2D6* have been described; the PM phenotype arises in individuals with two such alleles. In addition, UMs with multiple functional copies of *CYP2D6* have been identified especially in East Africa, the Middle East, and Oceania. PMs have slower elimination rates and lower clearance of substrate drugs; as a consequence (Fig. 72-1*B*), steady-state concentrations are higher and the time taken to achieve steady state is longer than in EMs **(Chap. 71)**. Conversely, UMs display very low steady-state parent drug concentrations and an abbreviated time to steady state.

Codeine is biotransformed by CYP2D6 to the potent active metabolite morphine, so its effects are blunted in PMs and exaggerated in UMs. Deaths due to respiratory depression in children given codeine after tonsillectomy have been attributed to the UM trait, and the U.S. Food and Drug Administration (FDA) has revised the package insert to include a prominent "black box" warning against its use in this setting and, in fact, forbidding its use in children less than 12 years old. In the case of drugs with beta-blocking properties metabolized by CYP2D6, greater signs of beta blockade (e.g., bronchospasm, bradycardia) have been reported in PM subjects than in EMs. This can be seen not only with orally administered beta blockers such as metoprolol and carvedilol, but also with ophthalmic timolol and with the sodium channel–blocking antiarrhythmic propafenone, a CYP2D6 substrate with beta-blocking properties. UMs may require very high dosages of nortriptyline and other tricyclic antidepressants to achieve a therapeutic effect. Tamoxifen is a prodrug that undergoes CYP2D6-mediated biotransformation to active metabolites, so its efficacy may be in part related to this polymorphism. In addition, the widespread use of selective serotonin reuptake inhibitors (SSRIs) to treat tamoxifen-related hot flashes may also alter the drug's effects because many SSRIs, notably fluoxetine and paroxetine, are also CYP2D6 inhibitors **(see Table 71-2)**.

CYP2C19 The PM phenotype for CYP2C19 is common (20%) among Asians and rarer (2–3%) in other populations; the frequency of the PM trait is especially high (>50%) in Oceania. The impact of polymorphic CYP2C19-mediated metabolism has been demonstrated with the proton pump inhibitor omeprazole, where ulcer cure rates with "standard" dosages were much lower in EM patients (29%) than in PMs (100%). CYP2C19 is responsible for bioactivation of the antiplatelet drug clopidogrel, and large retrospective and prospective studies have documented decreased efficacy (e.g., increased myocardial infarction after placement of coronary stents or increased stroke or transient ischemic attacks) among subjects with one or two reduction of function alleles. In addition, some studies suggest that omeprazole and possibly other proton pump inhibitors phenocopy this effect by inhibiting CYP2C19.

CYP2C9 Relatively common loss-of-function alleles in *CYP2C9* are associated with increased rates of neurologic complications with phenytoin, hypoglycemia with glipizide, and reduced warfarin dose required to maintain stable anticoagulation. Rare patients homozygous for loss-of-function alleles may require very low warfarin dosages. Up to 50% of the variability in steady-state warfarin dose requirement is attributable to polymorphisms in *CYP2C9* and in the promoter of *VKORC1*, which encodes the warfarin target, with lesser contributions by genes such as *CYP4F2* controlling vitamin K metabolism. The angiotensin receptor blocker losartan is a prodrug that is bioactivated by CYP2C9; as a result, PMs and those receiving inhibitor drugs may display little response to therapy.

DPYD Individuals homozygous for loss-of-function alleles in dihydropyrimidine dehydrogenase, encoded by *DPYD*, are at increased risk for severe toxicity when exposed to the substrate anticancer drug 5-fluorouracil (5-FU), as well as to capecitabine and tegafur, which are metabolized to 5-FU. Dose reductions have been recommended in intermediate metabolizers.

Transferase Variants Thiopurine *S*-methyltransferase (TPMT) bioinactivates the antileukemic drug 6-mercaptopurine (6-MP), and 6-MP is itself an active metabolite of the immunosuppressive azathioprine. Homozygotes for alleles encoding inactive TPMT (1/300 individuals) predictably exhibit severe and potentially fatal pancytopenia on standard doses of azathioprine or 6-MP. On the other hand, homozygotes for fully functional alleles may display less anti-inflammatory or antileukemic effect with standard doses of the drugs. GWA studies have also identified loss-of-function variants in *NUDT15* that reduce degradation of thiopurine metabolites and, thereby, also increase risk of excessive myelosuppression.

N-acetylation is accomplished by hepatic *N*-acetyl transferase (NAT), which represents the activity of two genes, *NAT1* and *NAT2*. Both enzymes transfer an acetyl group from acetyl coenzyme A to the drug; polymorphisms in *NAT2* are thought to underlie individual differences in the rate at which drugs are acetylated and thus define "rapid acetylators" and "slow acetylators." Slow acetylators make up ~50% of European and African populations but are less common among East Asians. Slow acetylators have an increased incidence of the drug-induced lupus syndrome during procainamide and hydralazine therapy and of hepatitis with isoniazid.

Individuals homozygous for a common promoter polymorphism that reduces transcription of uridine diphosphate glucuronosyltransferase (*UGT1A1*) have benign hyperbilirubinemia (Gilbert's syndrome; **Chap. 348**). This variant has also been associated with diarrhea and increased bone marrow depression with the antineoplastic prodrug irinotecan, whose active metabolite is normally detoxified by UGT1A1-mediated glucuronidation. The antiretroviral atazanavir is a UGT1A1 inhibitor and, thus, can increase bilirubin levels especially in individuals with the Gilbert's variant. While this is benign, the hyperbilirubinemia can complicate clinical care because it may raise the question of whether coexistent hepatic injury is present.

Transporter Variants The risk for myotoxicity with simvastatin and possibly other statins appears increased with variants in *SLCO1B1*. Variants in *MDR1*, encoding the drug efflux transporter P-glycoprotein, may increase digoxin toxicity. Variants in the uptake transporters *MATE1* and *MATE2* have been reported to modulate metformin's glucose-lowering activity.

GENETIC VARIANTS AFFECTING PHARMACODYNAMICS

A variant in the *VKORC1* promoter, especially common in Asian subjects (Table e72-1), reduces transcriptional activity generating less protein and, thus, lowering warfarin dose requirement. Multiple polymorphisms identified in the β_2-adrenergic receptor appear to be linked to specific drug responses in asthma and congestive heart failure, diseases in which β_2-receptor function might be expected to determine drug response. Polymorphisms in the β_2-receptor gene have also been associated with response to inhaled β_2-receptor agonists, while those in the β_1-adrenergic receptor gene have been associated with variability in heart rate slowing and blood pressure lowering.

Drugs may also interact with genetic pathways of disease to elicit or exacerbate symptoms of the underlying conditions. In the porphyrias, CYP inducers are thought to increase the activity of enzymes proximal to the deficient enzyme, exacerbating or triggering attacks **(Chap. 428)**. Variants decreasing activity of glucose-6-phosphate dehydrogenase (G6PD), which occur most often in individuals of African, Mediterranean, or South Asian descent, increase the risk of hemolytic anemia in response to the antimalarial primaquine **(Chap. 105)** and the uric acid–lowering agent rasburicase, which does not cause hemolysis in patients with normal amounts of the enzyme. Patients with mutations in *RYR1* encoding the skeletal muscle intracellular release calcium (also termed type 1 ryanodine receptor) are asymptomatic until exposed to certain general anesthetics, which can trigger the rare syndrome of malignant hyperthermia. Certain antiarrhythmics and other drugs can produce marked QT prolongation and torsades de pointes **(Chap. 253)**, and in a minority of affected patients, this adverse effect represents unmasking of previously subclinical congenital long QT syndrome. A variant in *ACKR1* is common in African ancestry individuals and is associated with white cell counts lower than the conventional "normal range" but not associated with disease. While this has been termed "benign ethnic neutropenia," these individuals not only undergo more diagnostic testing (including bone marrow biopsies that are almost always normal) but also have chemotherapy (associated with neutropenia) withdrawn at higher rates than noncarriers.

Immunologically Mediated Drug Reactions Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) is a potentially fatal skin and systemic reaction now increasingly recognized to be linked to specific HLA alleles (Table e72-2), as have cases of drug-induced hepatotoxicity and of the drug rash with eosinophilia and systemic symptoms (DRESS) syndrome. The frequency of risk alleles often varies by ancestry (Table e72-1). The HLA risk alleles appear to be necessary but not sufficient to elicit these reactions. For example, HLA-B*57:01 is a risk allele for abacavir-related SJS/TEN and flucloxacillin-related hepatotoxicity. However, while 55% of abacavir-exposed subjects will develop a reaction, only 1/10,000 subjects exposed to flucloxacillin develop hepatotoxicity. Thus, a third factor, the nature of which has not yet been established, seems necessary.

Tumor and Infectious Agent Genomes The actions of drugs used to treat infectious or neoplastic disease may be modulated by variants in these nonhuman germline genomes. Genotyping tumors to target therapies to underlying mechanisms and to avoid potentially toxic therapy in patients who would derive no benefit is now standard in many cancers **(Chap. 76)**. Trastuzumab, which potentiates anthracycline-related cardiotoxicity, is ineffective in breast cancers that do not express the Herceptin receptor. Imatinib targets a specific tyrosine kinase, BCR-Abl1, that is generated by the translocation that creates the Philadelphia chromosome typical of chronic myelogenous leukemia (CML). Imatinib is also an inhibitor of another kinase, c-kit, and the drug is remarkably effective in c-kit–driven cancer, such as gastrointestinal stromal tumors **(Chap. 76)**. Vemurafenib does not inhibit wild-type *BRAF* but is active against the V600E mutant form of the kinase. Crizotinib is highly effective in non-small-cell lung cancers harboring anaplastic lymphoma kinase (ALK) mutations.

INCORPORATING PHARMACOGENETIC INFORMATION INTO CLINICAL PRACTICE

The discovery of common variant alleles with relatively large effects on drug response raises the prospect that these variants could be used to guide therapy. Desired outcomes could be better ways of choosing likely effective drugs and dosages, or avoiding drugs that are likely to produce severe adverse drug events or be ineffective in individual subjects. Indeed, the FDA now incorporates pharmacogenetic data into package inserts meant to guide prescribing. A decision to adopt pharmacogenetically guided dosing for a given drug depends on multiple factors. The most important are the magnitude and clinical importance of the genetic effect and the strength of evidence linking genetic variation to variable drug effects (e.g., anecdote vs post hoc analysis of clinical trial data vs randomized clinical trial [RCT]). The evidence can be strengthened if statistical arguments from clinical trial data are complemented by an understanding of underlying physiologic mechanisms. Cost versus expected benefit may also be a factor.

Point of Care Versus Preemptive Approaches Two approaches to pharmacogenetic implementation have been put in place at "early adopter" institutions and are currently being evaluated. In the first, variant-specific assays are ordered at the time of drug prescription and delivered rapidly (often within an hour or two), and the results are then used to guide therapy with that specific drug. The alternative to this "point-of-care" approach is a "preemptive" approach in which pharmacogenetic testing for large numbers of potential variants across many drugs is undertaken prior to prescription of any drug. The data are then available in electronic health record (EHR) systems and coupled to real-time clinical decision support (CDS). When a drug whose effects are known to be influenced by pharmacogenetic variants is prescribed, the EHR system looks up whether variants likely to affect response are present; if so, CDS will alert health care providers that an alternate drug or a different dose may be required.

Challenges There are multiple challenges in putting in place either system. Assay validity and reproducibility have been issues in the past but are less likely now. While common variants in genes such as those listed in Table e72-1 have been clearly associated with variable drug responses, the effect of rare variants, now readily discoverable by large-scale sequencing, remains largely unexplored. The Clinical Pharmacogenetics Implementation Consortium (CPIC) and the Dutch Pharmacogenetics Working Group have developed and published guidelines for multiple drug-gene pairs focusing on the question of what might be an appropriate drug dose adjustment given the availability of genetic data. These resources do not directly address the question of when or how such genetic testing should be undertaken.

Developing Evidence That Pharmacogenetic Testing Alters Drug Outcomes A major issue is whether pharmacogenetic testing affects important drug response outcomes. When the evidence is compelling, alternate therapies are not available, and there are clear recommendations for dosage adjustment in subjects with variants, there is a strong argument for deploying genetic testing as a guide to prescribing; HLA-B*57:01 testing for abacavir is an example described below. In other situations, the arguments are less compelling: the magnitude of the genetic effect may be smaller, the consequences may be less serious, alternate therapies may be available, or the drug effect may be amenable to monitoring by other approaches.

One school argues that the physiology and pharmacology are known and that RCTs are, therefore, unnecessary (and conceivably unethical). The analogy is sometimes drawn to well-recognized dose adjustment of renally excreted drugs in the presence of renal dysfunction. RCTs have not been conducted and the idea of such dose adjustment is well accepted in the medical community and recommended in FDA-approved drug labels. Others have argued that the effect of genetic variants is generally modest and variability in drug actions has many nongenetic sources, so genetic testing might provide marginal benefit at best.

Efforts to demonstrate the value of pharmacogenetic testing have met with mixed results. An RCT clearly showed that HLA-B*57:01 testing eliminates SJS/TEN due to abacavir. Similarly, regulatory authorities in some countries in Southeast Asia mandated HLA-B*15:02 testing prior to initiation of carbamazepine; however, in this case, an unfortunate outcome in some jurisdictions was that prescribers stopped using carbamazepine, often substituting phenytoin (another drug associated with SJS/TEN), so the incidence of the severe ADR was unchanged.

RCTs evaluating the effect of using pharmacogenetically guided therapy to optimize warfarin treatment have shown either no effect or a modest benefit of incorporating genetic information into prescribing the drug. New effective alternate therapies to clopidogrel that appear to lack important pharmacogenetic variants have emerged. One approach to therapy, therefore, is to use pharmacogenetic testing to identify subjects in whom variants are absent and, therefore, a standard dose of clopidogrel is likely to be effective and to reserve alternate more expensive therapies for subjects likely to have variant responses to warfarin or clopidogrel.

Two large trials have randomized patients with acute coronary syndromes to newer antiplatelet therapies (ticagrelor or prasugrel) or clopidogrel if *CYP2C19* variants were absent; in one, clopidogrel was superior, and in the second, there was a trend in the same direction. A meta-analysis including these trials and others suggested that therapy guided by platelet function testing and genotyping resulted in improved efficacy and decreased minor bleeding with clopidogrel.

A 6944-patient trial conducted in seven European countries reported in 2023 that genotyping for 50 pharmacogene variants in 12 genes reduced serious ADRs (associated with 42 drugs) by 30% compared to standard prescribing pharmacogenetically guided treatment. Although there was heterogeneity in the effect across countries, the result of this large trial further supports the idea of preemptive pharmacogenetic screening.

GENETICS AND DRUG DEVELOPMENT

Genetic tools are now being increasingly used to identify or validate new drug targets. Available data suggest that a new drug development program is more likely to succeed if evidence from human genetics supports the role of a possible drug target in disease pathogenesis and suggests that the risk of toxicity due to high-risk pharmacokinetics or other mechanisms is small. Furthermore, studies of the relationships between variants in genes encoding drug target molecules and a range of phenotypes (e.g., those in EHRs) are being used for drug "repurposing," identifying new indications for existing drugs.

Finding Protective Alleles Can Identify Drug Targets One example of using genetics to identify a new drug target started with the discovery that very rare gain-of-function variants in *PCSK9* are a rare cause of familial hypercholesterolemia. Subsequently, population studies showed that carriers of loss-of-function SNPs (2.5% of African Americans) had decreased low-density lipoprotein cholesterol, decreased incidence of coronary artery disease, and no deleterious consequences in other organ systems. These data triggered the development of PCSK9 monoclonal antibodies, which were marketed <10 years after the initial population studies. Other targets implicated by similar population genetic studies include HSD17B13 for prevention of chronic liver disease and ANGPTL3 for hyperlipidemia. Discovering rare protective alleles may require very large data sets (>100,000), such as EHR systems coupled to DNA biobanks as in the U.S. *All of Us* cohort or epidemiologic cohorts like the UK Biobank.

Cancer In cancer, tumor sequencing has identified new targets for drug development, often constitutively active kinases. A problem in this area has been the rapid emergence of drug resistance, often after extraordinary initial responses. For example, 40% of melanomas appear to be driven by the V600E mutant form of *BRAF*, and the specific inhibitor vemurafenib can produce clinically spectacular remission. However, durable responses are rare, and it is now apparent that combination therapy, often with inhibitors of the MEK pathway, can provide improved therapy. Another approach that is rapidly gaining wide use in cancer involves drugs that reverse immune system inhibition **(Chap. 78)**. In some patients, the release of this "brake" can provide durable remissions, whereas in others, severe adverse events, including colitis, pneumonitis, and myocarditis, have been reported. Understanding the mechanisms underlying variability to these therapies is a major emerging challenge in the field.

Using Multiple Data Types The development of methods to understand associations across multiple large data sets is another approach that is being explored in drug development. For example, a GWA study of risk of rheumatoid arthritis identified multiple risk loci, many encoding proteins that are known targets for intervention in the disease. Interestingly, others encode proteins that are targets for drugs used in other conditions, such as certain cancers, raising the question of whether such drugs could be "repurposed" for rheumatoid arthritis.

While the field has, to date, focused on individual high effect size variants (that are often common in a population), newer approaches combining many (dozens to millions) common variants into polygenic risk scores to predict drug responses are also being explored. An extension of this approach is the broader issue of systems pharmacology **(see Chap. 71)**, in which multiple sources of data are used to identify potential molecules or pathways that would be amenable to treatment, by new drugs or by existing agents, using analysis of genomic, transcriptomic, proteomic, and other large data sets.

SUMMARY

The science of pharmacogenomics has evolved from isolated examples of rare adverse drug actions to a more comprehensive view of the role of genetic variation in mediating the effects of most drugs. Current principles include:

- Genetic variants with an important effect on drug actions can be common, and their frequencies often vary by ancestry.
- One common mechanism is modulation of drug concentrations.
- No practitioner can be expected to remember all variants important for all drugs. Electronic data systems can now be accessed to describe this information. Ultimately, this information will be used by linking individual pharmacogenetic data to smart EHR systems.
- Incorporating genetic approaches into drug development projects holds the promise of more rapid development of targeted, safe, and effective therapies.

FURTHER READING

Diogo D et al: Phenome-wide association studies across large population cohorts support drug target validation. Nat Commun 9:4285, 2018.

Galli M et al: Comparative effects of guided vs. potent P2Y12 inhibitor therapy in acute coronary syndrome: A network meta-analysis of 61 898 patients from 15 randomized trials. Eur Heart J 43:959, 2022.

Osanlou O et al: Pharmacogenetics of adverse drug reactions. Adv Pharmacol 83:155, 2018.

Roden DM et al: Pharmacogenomics. Lancet 394:521, 2019.

Swen JJ et al: A 12-gene pharmacogenetic panel to prevent adverse drug reactions: An open-label, multicentre, controlled, cluster-randomised crossover implementation study. Lancet 401:10374, 2023.

This page intentionally left blank

Oncology and Hematology PART 4

Section 1 Neoplastic Disorders

73 Approach to the Patient with Cancer

Dan L. Longo

The application of current treatment techniques (surgery, radiation therapy, chemotherapy, and biologic therapy) results in the cure of nearly two of three patients diagnosed with cancer. Nevertheless, patients experience the diagnosis of cancer as one of the most traumatic and revolutionary events that has ever happened to them. Independent of prognosis, the diagnosis brings with it a change in a person's self-image and in their role in the home and workplace. The prognosis of a person who has just been found to have pancreatic cancer is the same as the prognosis of the person with aortic stenosis who develops the first symptoms of congestive heart failure (median survival, ~8 months). However, the patient with heart disease may remain functional and maintain a self-image as a fully intact person with just a malfunctioning part, a diseased organ ("a bum ticker"). By contrast, the patient with pancreatic cancer has a completely altered self-image and is viewed differently by family and anyone who knows the diagnosis. The patient is being attacked and invaded by a disease that could be anywhere in the body. Every ache or pain takes on desperate significance. Cancer is an exception to the coordinated interaction among cells and organs. In general, the cells of a multicellular organism are programmed for collaboration. Many diseases occur because the specialized cells fail to perform their assigned task. Cancer takes this malfunction one step further. Not only is there a failure of the cancer cell to maintain its specialized function, but it also strikes out on its own; the cancer cell competes to survive using natural mutability and natural selection to seek advantage over normal cells in a recapitulation of evolution. One consequence of the traitorous behavior of cancer cells is that the patient feels betrayed by their body. The cancer patient feels that they, and not just a body part, are diseased.

THE MAGNITUDE OF THE PROBLEM

No nationwide cancer registry exists; therefore, the incidence of cancer is estimated on the basis of the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) database, which tabulates cancer incidence and death figures from 13 sites, accounting for about 10% of the U.S. population, and from population data from the U.S. Census Bureau. In 2024, 2.001 million new cases of invasive cancer (1,029,080 men and 927,060 women) were diagnosed, and 611,720 persons (322,800 men and 288,920 women) died from cancer. The percent distribution of new cancer cases and cancer deaths by site for men and women is shown in **Table 73-1**. Cancer mortality continues to decline; however, 6 of the 10 most common cancers have increased in incidence by 1–3% in recent years, and troubling disparities among different racial/ethnic groups persist. Mortality is twice as high in black people than white people for cancers of the prostate, stomach, and uterine corpus. Cancer is the cause of one in four deaths in the United States.

TABLE 73-1 Distribution of Cancer Incidence and Deaths for 2021

MALE			FEMALE		
SITES	%	NUMBER	SITES	%	NUMBER
Cancer Incidence					
Prostate	29	299,010	Breast	32	310,720
Lung	11	116,310	Lung	12	118,270
Colorectal	8	81,540	Colorectal	7	71,270
Bladder	6	63,070	Endometrial	7	67,880
Melanoma	6	59,170	Melanoma	4	41,470
Kidney	5	52,380	Lymphoma	4	36,030
Lymphoma	4	44,590	Pancreas	3	31,910
Oral cavity	4	41,510	Thyroid	3	31,520
Leukemia	4	36,450	Kidney	3	29,230
Pancreas	3	34,530	Leukemia	3	26,320
All others	19	200,520	All others	21	207,440
All sites	100	1,029,080	All sites	100	972,060
Cancer Deaths					
Lung	20	65,790	Lung	21	59,280
Prostate	11	35,250	Breast	15	42,250
Colorectal	9	28,700	Pancreas	8	24,480
Pancreas	8	25,270	Colorectal	8	24,310
Liver	6	19,120	Endometrial	5	13.250
Leukemia	4	13,640	Ovary	4	12,740
Esophagus	4	12,880	Liver	3	10,720
Bladder	4	12,290	Leukemia	3	10,030
Lymphoma	4	11,780	Lymphoma	3	8,360
CNS	3	10,690	CNS	3	8,070
All others	27	87,390	All others	26	75,4330
All sites	100	322,800	All sites	100	288,920

Source: From RL Siegel et al: Cancer statistics, 2024. CA Cancer J Clin 74:12, 2024. Reproduced John Wiley & Sons Ltd.

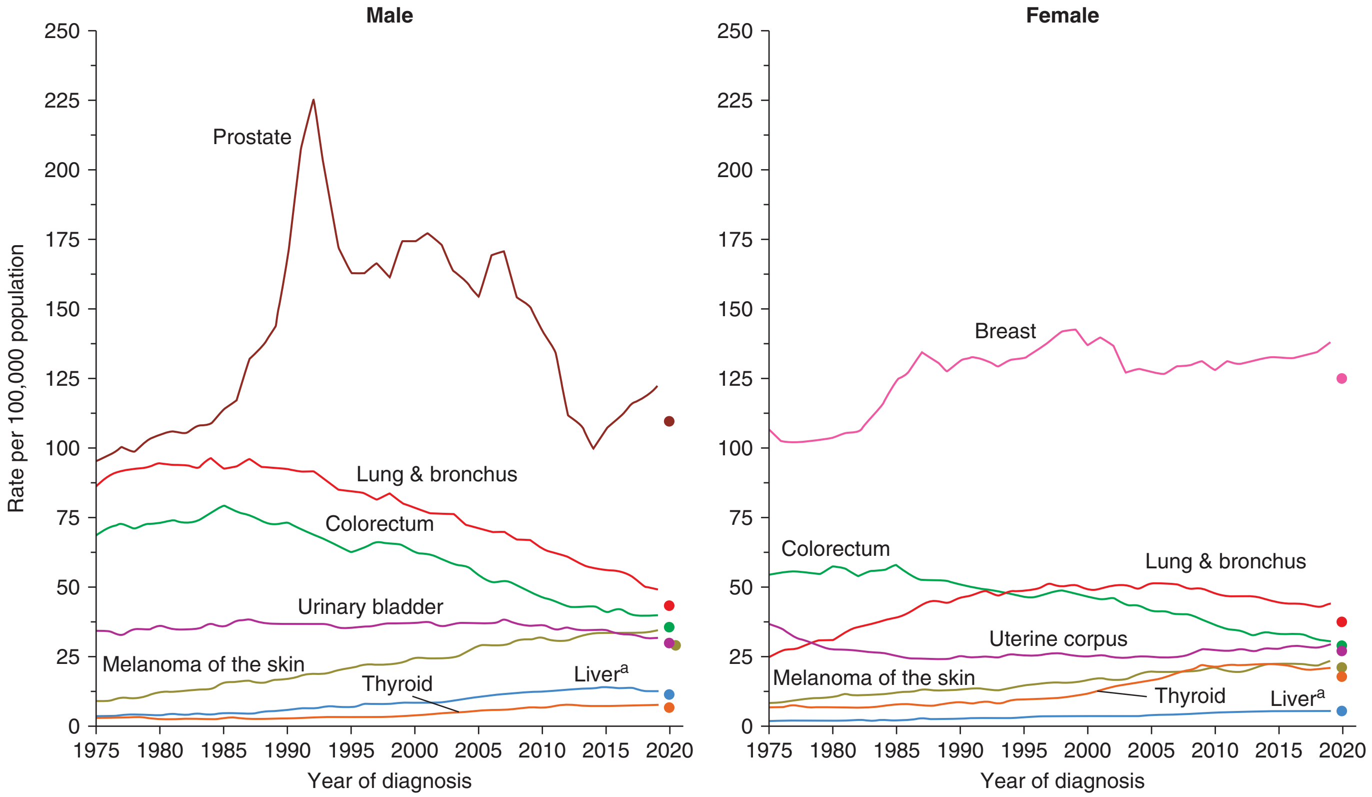

FIGURE 73-1 Trends in cancer incidence for men and women, 1975–2020. Incidence data for 2020 are indicated by color dots separate from the trend lines. *(Reproduced with permission from RL Siegel et al: Cancer statistics, 2024. CA Cancer J Clin 74:1, 2024.)*

The most significant risk factor for cancer overall is age; 58% of all cases were in those aged >65 years, down from 61% in 1995, despite an increase in this age group from 13% to 17% of the population. Cancer incidence increases as the third, fourth, or fifth power of age in different sites. For the interval between birth and age 49 years, 1 in 29 men and 1 in 19 women will develop cancer; for the interval between ages 50 and 59 years, 1 in 15 men and 1 in 17 women will develop cancer; for the interval between ages 60 and 69 years, 1 in 6 men and 1 in 10 women will develop cancer; and for people aged ≥70, 1 in 3 men and 1 in 4 women will develop cancer. Overall, men have a 40.5% risk of developing cancer at some time during their lives; women have a 38.9% lifetime risk.

Cancer is the second leading cause of death behind heart disease. Deaths from heart disease have declined 45% in the United States since 1950 and continue to decline. Cancer has overtaken heart disease as the number one cause of death in persons aged <85 years. Incidence trends over time are shown in **Fig. 73-1**. After a 70-year period of increase, cancer deaths began to decline in 1990–1991 **(Fig. 73-2)**. Between 1990 and 2010, cancer deaths decreased by 21% among men and 12.3% among women. The incidence has been steady since 2013. The magnitude of the decline is illustrated in **Fig. 73-3**. The five leading causes of cancer deaths are shown for various populations in **Table 73-2**. The 5-year survival for white patients was 39% in 1960–1963 and 68% in 2010–2016. Cancers are more often deadly in blacks; the 5-year survival was 63% for the 2010–2016 interval; however, the racial differences are narrowing over time. Incidence and mortality vary among racial and ethnic groups **(Table 73-3)**. The basis for these differences is unclear.

Advances in cancer prevention, diagnosis, and treatment since the early 1990s have averted millions of cancer deaths based on projections from the slopes of the mortality curves leading up to the 1990s **(Fig. 73-4)**.

■ CANCER AROUND THE WORLD

In 2022, nearly 20 million new cancer cases and 9.7 million cancer deaths were estimated worldwide, according to estimates of GLOBOCAN 2022, developed by the International Agency for Research on Cancer (IARC). Rates are increasing worldwide. When broken down by region of the world, almost half of cases were in Asia (which has 59.2% of the world's population), 26% in Europe (9.6% of the world's population), 13.1% in North America, 7.1% in Central/South America (the Americas, North and South, account for 13.3% of the world's population), 6% in Africa (16.6% of the world's population), and 1% in Australia/New Zealand (0.5% of the world's population) **(Fig. 73-5)**. Lung cancer is the most common cancer and the most common cause of cancer death in the world. Its incidence is highly variable, affecting only 2 per 100,000 African women but as many as 61 per 100,000 North American men. Breast cancer is the second most common cancer worldwide; however, it ranks fourth as a cause of death behind lung, stomach, and liver cancer. Among the eight most common forms of cancer, lung (2-fold), breast (3-fold), prostate (2.5-fold), and colorectal (3-fold) cancers are more common in more developed countries than in less developed countries. By contrast, liver (2-fold), cervical (2-fold), and esophageal (2- to 3-fold) cancers are more common in less developed countries. Stomach cancer incidence is similar in more and less developed countries but is much more common in Asia than North America or Africa. The most common cancers in Africa are cervical, breast, and liver cancers. It has been estimated that nine modifiable risk factors are responsible for more than one-third of cancers worldwide. These include smoking, alcohol consumption, obesity, physical inactivity, low fruit and vegetable consumption, unsafe sex, air pollution, indoor smoke from household fuels, and contaminated injections.

PATIENT MANAGEMENT

Important information is obtained from every portion of the routine history and physical examination. The duration of symptoms may reveal the chronicity of disease. The past medical history may alert the physician to the presence of underlying diseases that may affect the choice of therapy or the side effects of treatment. The social history may reveal occupational exposure to carcinogens or habits, such as smoking or alcohol consumption, that may influence the course of disease and its treatment. The family history may suggest an underlying familial cancer predisposition and point out the need to begin surveillance or other preventive therapy for unaffected siblings of the patient. The review of systems may suggest early symptoms of metastatic disease or a paraneoplastic syndrome.

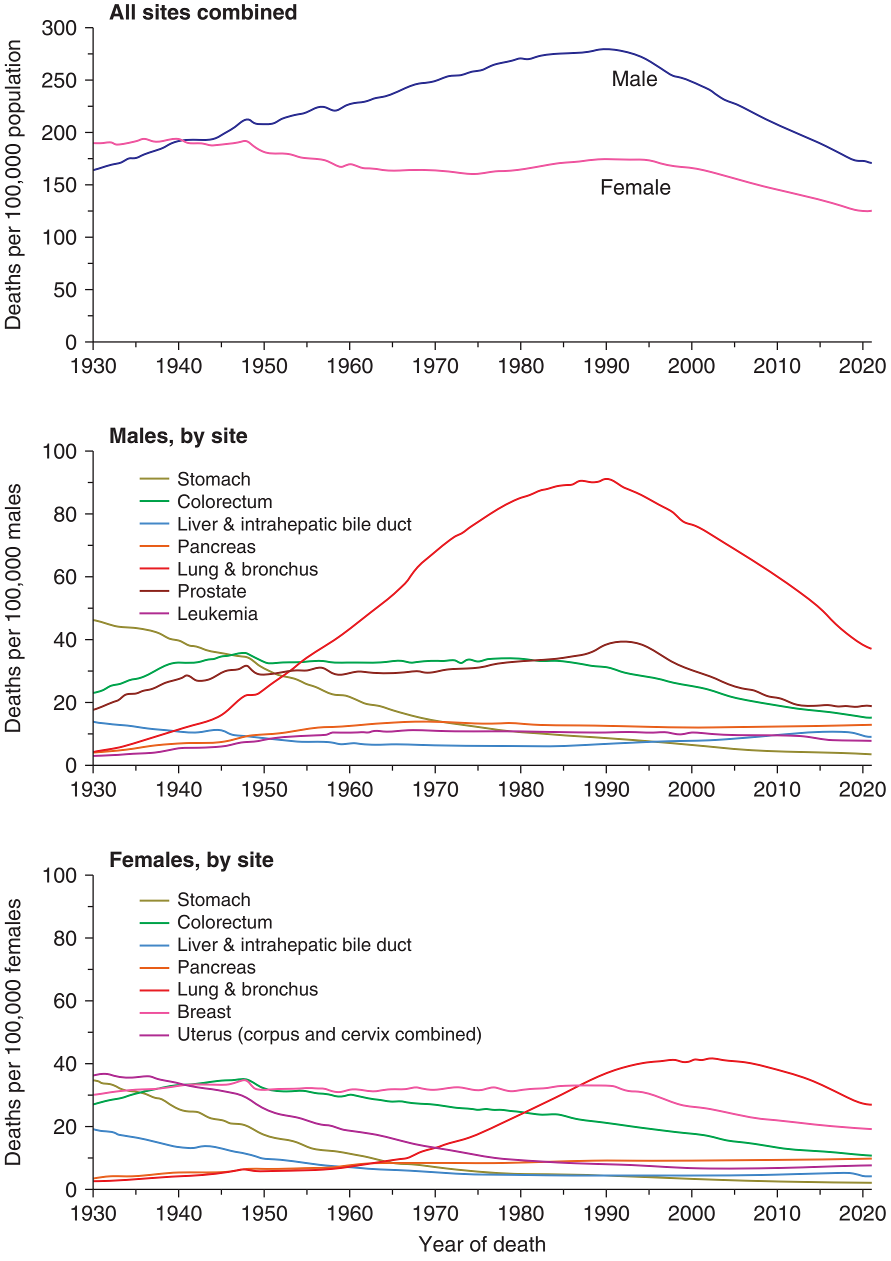

FIGURE 73-2 Trends in cancer mortality rates in men and women, 1930–2021. *(Reproduced with permission from RL Siegel et al: Cancer statistics, 2024. CA Cancer J Clin 74:1, 2024.)*

DIAGNOSIS

The diagnosis of cancer relies most heavily on invasive tissue biopsy and should never be made without obtaining tissue; no noninvasive diagnostic test is sufficient to define a disease process such as cancer (one exception to this rule may be hepatocellular carcinoma, which may be reliably diagnosed based on computed tomography [CT] or magnetic resonance imaging [MRI] showing the characteristic dynamic perfusion pattern of arterial hyperenhancement and venous or delayed phase washout). Although in rare clinical settings (e.g., thyroid nodules), fine-needle aspiration is an acceptable diagnostic procedure, the diagnosis generally depends on obtaining adequate tissue to permit careful evaluation of the histology of the tumor, its grade, and its invasiveness and to yield further molecular diagnostic information, such as the expression of cell-surface markers or intracellular proteins that typify a particular cancer, or the presence of a molecular marker, such as the t(8;14) translocation of Burkitt's lymphoma. Increasing evidence links the expression of certain genes with the prognosis and response to therapy **(Chaps. 76 and 77)**.

Occasionally, a patient will present with a metastatic disease process that is defined as cancer on biopsy but has no apparent primary site of disease. Efforts should be made to define the primary site based on age, sex, sites of involvement, histology and tumor markers, and personal and family history. Particular attention should be focused on ruling out the most treatable causes **(Chap. 97)**.

Once the diagnosis of cancer is made, the management of the patient is best undertaken as a multidisciplinary collaboration among the primary care physician, medical oncologists, surgical oncologists, radiation oncologists, oncology nurse specialists, pharmacists, social workers, rehabilitation medicine specialists, and a number of other consulting professionals working closely with each other and with the patient and family.

DEFINING THE EXTENT OF DISEASE AND THE PROGNOSIS

The first priority in patient management after the diagnosis of cancer is established and shared with the patient is to determine the extent of disease. The curability of a tumor usually is inversely proportional to the tumor burden. Ideally, the tumor will be diagnosed before symptoms develop or as a consequence of screening efforts **(Chap. 75)**. A very high proportion of such patients can be cured. However, most patients with cancer present with symptoms related to the cancer, caused either by mass effects of the tumor or by alterations associated with the production of cytokines or hormones by the tumor.

For most cancers, the extent of disease is evaluated by a variety of noninvasive and invasive diagnostic tests and procedures. This process is called *staging*. There are two types. *Clinical staging* is based on physical examination, radiographs, isotopic scans, CT scans, and other imaging procedures; *pathologic staging* takes into account information obtained during a surgical procedure, which might include intraoperative palpation, resection of regional lymph nodes and/or tissue adjacent to the tumor, and inspection and biopsy of organs commonly involved in disease spread. Pathologic staging includes histologic examination of all tissues removed during the surgical procedure. Surgical procedures performed may include a simple lymph node biopsy or more extensive procedures such as thoracotomy, mediastinoscopy, or laparotomy. Surgical staging may occur in a separate procedure or may be done at the time of definitive surgical resection of the primary tumor. A subset of pathologic staging is the examination of tissue obtained at initial surgery that occurs after the delivery of some treatment, which is called neoadjuvant therapy. Stage of disease determined after neoadjuvant therapy is designated with the prefix y.

Knowledge of the predilection of particular tumors for spreading to adjacent or distant organs helps direct the staging evaluation.

Information obtained from staging is used to define the extent of disease as localized, as exhibiting spread outside of the organ of origin to regional but not distant sites, or as metastatic to distant sites. The most widely used system of staging is the tumor, node, metastasis (TNM) system codified by the International Union Against Cancer and the American Joint Committee on Cancer. The TNM classification is an anatomically based system that categorizes the tumor on the basis of the size of the primary tumor lesion (T1–4, where a higher number indicates a tumor of larger size), the presence of nodal involvement (usually N0 and N1 for the absence and presence, respectively, of involved nodes, although some tumors have more elaborate systems of nodal grading), and the presence of metastatic disease (M0 and M1

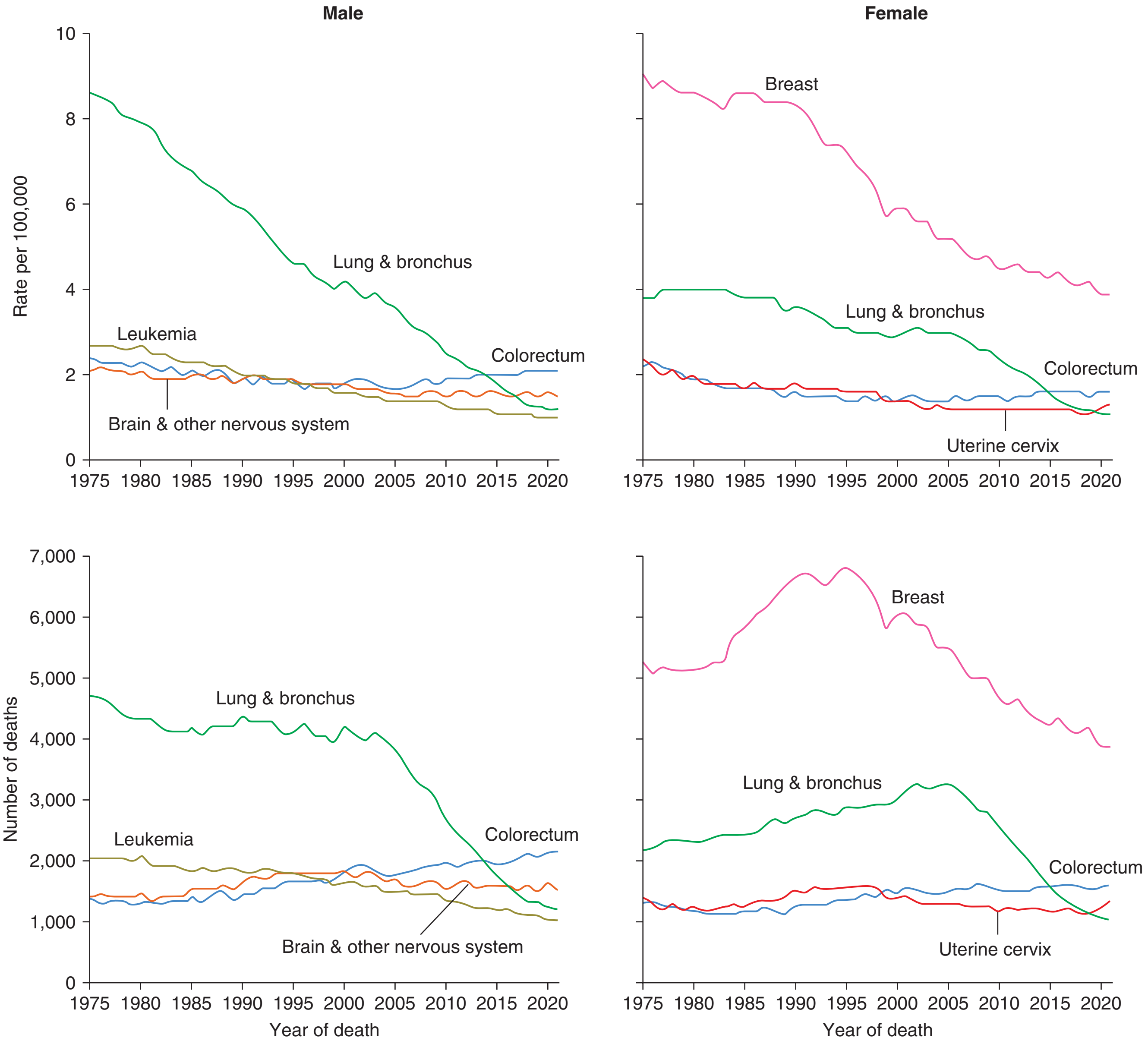

FIGURE 73-3 Trends in cancer incidence and death rates. *(Reproduced with permission from RL Siegel et al: Cancer statistics, 2024. CA Cancer J Clin 74:1, 2024.)*

TABLE 73-2 The Five Leading Primary Tumor Sites for Patients Dying of Cancer Based on Age and Sex in 2018

			AGE, YEARS					
RANK	SEX	ALL AGES	UNDER 20	20–39	40–49	50-64	65–79	>80
1	M	Lung	CNS	Colorectal	Colorectal	Lung	Lung	Lung
	F	Lung	CNS	Breast	Breast	Lung	Lung	Lung
2	M	Prostate	Leukemia	CNS	Lung	Colorectal	Prostate	Prostate
	F	Breast	Leukemia	Cervix	Colorectal	Breast	Breast	Breast
3	M	Colorectal	Bone sarcoma	Leukemia	CNS	Pancreas	Liver	Colorectal
	F	Colorectal	Soft tissue sarcoma	Colorectal	Lung	Colorectal	Pancreas	Colorectal
4	M	Pancreas	Soft tissue sarcoma	Testis	Pancreas	Liver	Colorectal	Bladder
	F	Pancreas	Bone sarcoma	CNS	Cervix	Pancreas	Colorectal	Pancreas
5	M	Liver	Lymphoma	Lymphoma	Esophagus	Esophagus	Liver	Pancreas
	F	Ovary	Kidney	Leukemia	Ovary	Ovary	Ovary	Leukemia

Abbreviations: CNS, central nervous system; F, female; M, male.

Source: From RL Siegel et al: Cancer statistics, 2024. CA Cancer J Clin 74:12, 2024.

TABLE 73-3 Cancer Incidence and Mortality in Racial and Ethnic Groups, United States, 2016–2020

SITE	SEX	WHITE	BLACK	ASIAN/PACIFIC ISLANDER	AMERICAN INDIAN[a]	HISPANIC
Incidence per 100,000 Population						
All	M	511.2	533.9	299.0	504.1	377.2
	F	499.3	409.9	307.3	465.5	351.3
Breast		134.9	129.6	104.6	115.5	100.7
Colorectal	M	40.4	48.8	33.4	57.8	38.2
	F	30.5	35.0	23.7	43.7	27.2
Kidney	M	24.3	26.4	11.6	43.9	23.5
	F	12.1	13.7	5.5	23.9	13.3
Liver	M	11.2	17.0	18.4	27.3	20.4
	F	4.2	5.5	6.7	12.3	8.4
Lung	M	765.7	72,4	40.8	67.2	34.3
	F	54.8	45.8	28.1	58.6	24.0
Prostate		110.7	186.1	60.9	91.9	90.9
Stomach	M	7.1	13.0	11.8	13.1	11.4
	F	3.4	7.4	6.9	7.8	7.7
Cervix		7.2	8.6	6.0	11.4	9.7
Endometrial		27.9	28.9	21.7	30.4	25.8
Deaths per 100,000 Population						
All	M	183.3	217.4	111.6	221.6	130.2
	F	133.6	150.2	83.7	157.9	93.5
Breast		19.7	27.8	11.8	21.1	13.7
Colorectal	M	15.5	22.4	11.0	23.1	13.6
	F	11.1	14.4	7.8	16.0	8.5
Kidney	M	5.3	5.2	2.4	9.9	4.8
	F	2.3	2.2	1.0	4.2	2.1
Liver	M	8.5	13.0	12.6	19.9	13.1
	F	3.7	4.8	5.2	8.8	6.0
Lung	M	44.9	51.3	25.9	52.3	21.-
	F	32.9	28.0	15.6	37.0	11.4
Prostate		17.9	37.9	8.7	22.5	15.4
Stomach	M	2.9	7.2	6.0	7.7	5.9
	F	1.5	3.5	3.7	4.1	3.9
Cervix		2.0	3.3	1.7	2.3	2.5
Endometrial		4.6	9.1	3.5	4.9	4.3

[a] Based on Indian Health Service delivery areas.

Abbreviations: F, female; M, male.

Source: From RL Siegel et al: Cancer statistics, 2024. CA Cancer J Clin 74:12, 2024. Reproduced of John Wiley & Sons Ltd.

for the absence and presence, respectively, of metastases). The various permutations of T, N, and M scores (sometimes including tumor histologic grade [G]) are then broken into stages, usually designated by the roman numerals I through IV. Tumor burden increases and curability decreases with increasing stage. Other anatomic staging systems are used for some tumors, e.g., the Dukes classification for colorectal cancers, the International Federation of Gynecologists and Obstetricians classification for gynecologic cancers, and the Ann Arbor classification for Hodgkin's lymphoma.

Certain tumors cannot be grouped on the basis of anatomic considerations. For example, hematopoietic tumors such as leukemia, myeloma, and lymphoma are often disseminated at presentation and do not spread like solid tumors. For these tumors, other prognostic factors have been identified **(Chaps. 109–116)**.

In addition to tumor burden, a second major determinant of treatment outcome is the physiologic reserve of the patient. Patients who are bedridden before developing cancer are likely to fare worse, stage for stage, than fully active patients. Physiologic reserve is a determinant of how a patient is likely to cope with the physiologic stresses imposed by the cancer and its treatment. This factor is difficult to assess directly. Instead, surrogate markers for physiologic reserve are used, such as the patient's age or Karnofsky performance status **(Table 73-4)** or Eastern Cooperative Oncology Group (ECOG) performance status **(Table 73-5)**. Older patients and those with a Karnofsky performance status <70 or ECOG performance status ≥3 have a poor prognosis unless the poor performance is a reversible consequence of the tumor. Some have advocated for using one of the geriatric assessment tools to gauge physiologic reserve. See ASCO's video for an example of a geriatric assessment tool (*https://www.youtube.com/watch?v=jnaQIjOz2Dw*).

Increasingly, biologic features of the tumor are being related to prognosis. The expression of particular oncogenes, drug-resistance genes, apoptosis-related genes, and genes involved in metastasis is being found to influence response to therapy and prognosis. The presence of selected cytogenetic abnormalities may influence survival. Tumors with higher growth fractions, as assessed by expression of proliferation-related markers such as proliferating cell nuclear antigen (detectable by staining with Ki67 antibody), behave more aggressively than tumors with lower growth fractions. Information obtained from studying the tumor itself will increasingly be used to influence treatment decisions. Host genes involved in drug metabolism can influence the safety and efficacy of particular treatments.

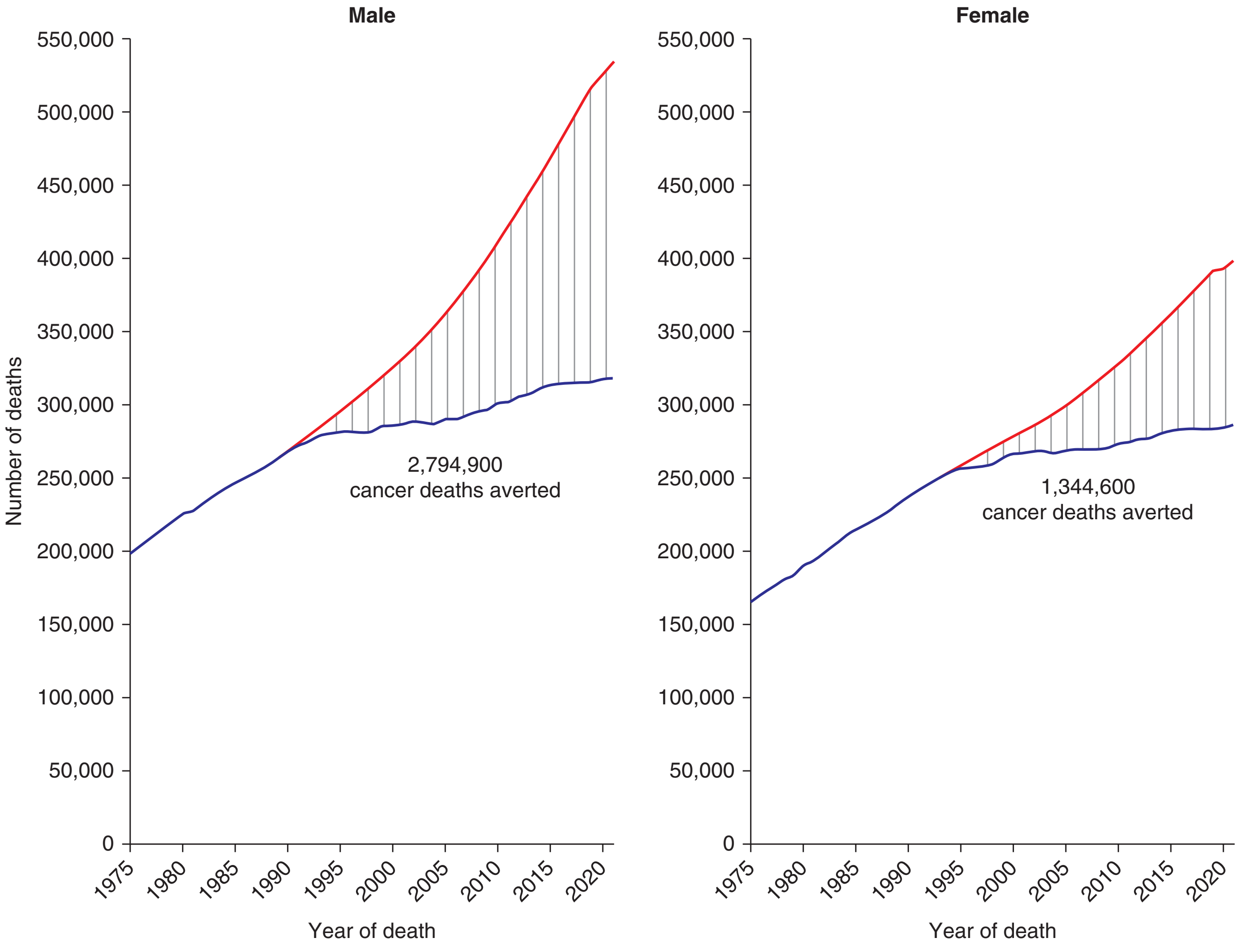

FIGURE 73-4 Cancer deaths averted in men and women since the early 1990s. Projections based on death continuing on the established trajectory. *(Reproduced with permission from RL Siegel et al: Cancer statistics, 2024. CA Cancer J Clin 74:1, 2024.)*

Enormous heterogeneity has been noted by studying tumors; we have learned that morphology is not capable of discerning certain distinct subsets of patients whose tumors have different sets of abnormalities. Tumors that look the same by light microscopy can be very different. Similarly, tumors that look quite different from one another histologically can share genetic lesions that predict responses to treatments. Furthermore, tumor cells vary enormously within a single patient even though the cells share a common origin.

MAKING A TREATMENT PLAN

From information on the extent of disease and the prognosis and in conjunction with the patient's wishes, it is determined whether the treatment approach should be curative or palliative in intent. Cooperation among the various professionals involved in cancer treatment is of the utmost importance in treatment planning. For some cancers, chemotherapy or chemotherapy plus radiation therapy delivered before the use of definitive surgical treatment (so-called neoadjuvant therapy) may improve the outcome, as seems to be the case for locally advanced breast cancer, head and neck cancers, and lung cancers, among others. In certain settings in which combined-modality therapy is intended, coordination among the medical oncologist, radiation oncologist, and surgeon is crucial to achieving optimal results. Sometimes the chemotherapy and radiation therapy need to be delivered sequentially and other times concurrently. Surgical procedures may precede or follow other treatment approaches. It is best for the treatment plan either to follow a standard protocol precisely or else to be part of an ongoing clinical research protocol evaluating new treatments. Ad hoc modifications of standard protocols are likely to compromise treatment results.

The choice of treatment approaches was formerly dominated by the local culture in both the university and the practice settings. However, it is now possible to gain access electronically to standard treatment protocols and to every approved clinical research study in North America through a personal computer interface with the Internet.[1]

The skilled physician also has much to offer the patient for whom curative therapy is no longer an option. Often a combination of guilt and frustration over the inability to cure the patient and the pressure of a busy schedule greatly limit the time a physician spends with a patient who is receiving only palliative care. Resist these forces. In addition to the medicines administered to alleviate symptoms (see below), it is important to remember the comfort that is provided by holding the patient's hand, continuing regular examinations, and taking time to talk.

MANAGEMENT OF DISEASE AND TREATMENT COMPLICATIONS

Although medicine has been guided through centuries by the aphorism "primum non nocere," *first do no harm*, it fits modern medicine poorly. As a practical matter, nearly everything we do in patient care has risk of doing harm; diagnostic tests, therapeutic interventions, and even physical diagnosis can lead to patient harm. A more relevant guide to modern medicine is "primum succerrere"; *first hasten to help*. Because cancer therapies are toxic **(Chap. 78)**, patient management involves addressing complications of both the disease and its treatment as well as the complex psychosocial problems associated with cancer. In the short term during a course of curative therapy, the patient's functional

---

[1] The National Cancer Institute maintains a database called PDQ (Physician Data Query) that is accessible on the Internet under the name CancerNet at *https://www.cancer.gov/publications/pdq*. Information can be obtained through a facsimile machine using CancerFax by dialing 301-402-5874. Patient information is also provided by the National Cancer Institute in at least three formats: on the Internet via CancerNet at *www.cancer.gov*, through the CancerFax number listed above, or by calling 1-800-4-CANCER. The quality control for the information provided through these services is rigorous.

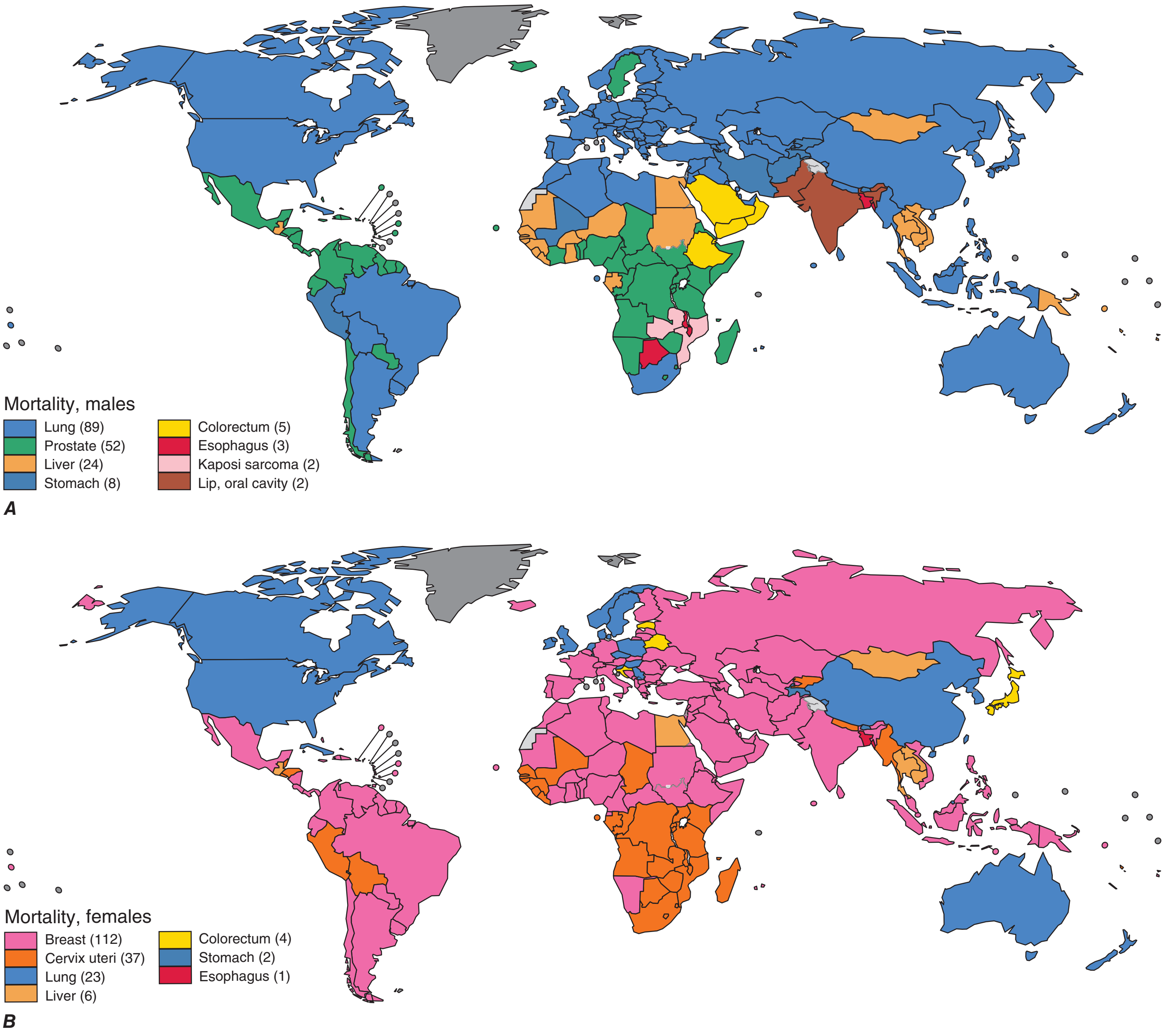

FIGURE 73-5 Global maps showing most common cause of cancer mortality by country in 2022 among (*A*) men and (*B*) women. *(Reproduced with permission from F Bray et al: Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74:229, 2024.)*

status may decline. Treatment-induced toxicity is less acceptable if the goal of therapy is palliation. The most common side effects of treatment are nausea and vomiting (see below), febrile neutropenia **(Chap. 79)**, and myelosuppression **(Chap. 78)**. Tools are now available to minimize the acute toxicity of cancer treatment.

TABLE 73-4 Karnofsky Performance Index

PERFORMANCE STATUS	FUNCTIONAL CAPABILITY OF THE PATIENT
100	Normal; no complaints; no evidence of disease
90	Able to carry on normal activity; minor signs or symptoms of disease
80	Normal activity with effort; some signs or symptoms of disease
70	Cares for self; unable to carry on normal activity or do active work
60	Requires occasional assistance but is able to care for most needs
50	Requires considerable assistance and frequent medical care
40	Disabled; requires special care and assistance
30	Severely disabled; hospitalization is indicated, although death is not imminent
20	Very sick; hospitalization is necessary; active supportive treatment is necessary
10	Moribund, fatal processes progressing rapidly
0	Dead

TABLE 73-5 The Eastern Cooperative Oncology Group (ECOG) Performance Scale

ECOG grade 0: Fully active, able to carry on all predisease performance without restriction

ECOG grade 1: Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light housework, office work

ECOG grade 2: Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about >50% of waking hours

ECOG grade 3: Capable of only limited self-care, confined to bed or chair >50% of waking hours

ECOG grade 4: Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair

ECOG grade 5: Dead

Source: Reproduced with permission from MM Oken et al: Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 5:649, 1982.

New symptoms developing in the course of cancer treatment should always be assumed to be reversible until proven otherwise. The fatalistic attribution of anorexia, weight loss, and jaundice to recurrent or progressive tumor could result in a patient dying from a reversible intercurrent cholecystitis. Intestinal obstruction may be due to reversible adhesions rather than progressive tumor. Systemic infections, sometimes with unusual pathogens, may be a consequence of the immunosuppression associated with cancer therapy. Some drugs used to treat cancer or its complications (e.g., nausea) may produce central nervous system symptoms that look like metastatic disease or may mimic paraneoplastic syndromes such as the syndrome of inappropriate antidiuretic hormone. A definitive diagnosis should be pursued and may even require a repeat biopsy.

A critical component of cancer management is assessing the response to treatment. In addition to a careful physical examination in which all sites of disease are physically measured and recorded in a flow chart by date, response assessment usually requires periodic repeating of imaging tests that were abnormal at the time of staging. If imaging tests have become normal, repeat biopsy of previously involved tissue is performed to document complete response by pathologic criteria. Biopsies are not usually required if there is macroscopic residual disease. A *complete response* is defined as disappearance of all evidence of disease, and a *partial response* as >50% reduction in the sum of the products of the perpendicular diameters of all measurable lesions. The determination of partial response may also be based on a 30% decrease in the sums of the longest diameters of lesions (Response Evaluation Criteria in Solid Tumors [RECIST]). *Progressive disease* is defined as the appearance of any new lesion or an increase of >25% in the sum of the products of the perpendicular diameters of all measurable lesions (or an increase of 20% in the sums of the longest diameters by RECIST). Tumor shrinkage or growth that does not meet any of these criteria is considered *stable disease*. Some sites of involvement (e.g., bone) or patterns of involvement (e.g., lymphangitic lung or diffuse pulmonary infiltrates) are considered unmeasurable. No response is complete without biopsy documentation of their resolution, but partial responses may exclude their assessment unless clear objective progression has occurred.

For some hematologic neoplasms, flow cytometric and genetic assays may determine the presence of residual tumor cells that escape microscopic detection. In general, these techniques can reliably detect as few as 1 tumor cell among 10,000 cells. If such tests do not detect tumor cells, the patient is said to have minimal residual disease negativity, a finding generally associated with more durable remissions. Accumulating data are defining interventions in patients with minimal residual disease positivity that can extend remission duration and survival.

Tumor markers may be useful in patient management in certain tumors. Response to therapy may be difficult to gauge with certainty. However, some tumors produce or elicit the production of markers that can be measured in the serum or urine, and in a particular patient, rising and falling levels of the marker are usually associated with increasing or decreasing tumor burden, respectively. Some clinically useful tumor markers are shown in Table 73-6. Tumor markers are not in themselves specific enough to permit a diagnosis of malignancy to be made, but once a malignancy has been diagnosed and shown to be associated with elevated levels of a tumor marker, the marker can be used to assess response to treatment.

TABLE 73-6 Tumor Markers

TUMOR MARKERS	CANCER	NONNEOPLASTIC CONDITIONS
Hormones		
Human chorionic gonadotropin	Gestational trophoblastic disease, gonadal germ cell tumor	Pregnancy
Calcitonin	Medullary cancer of the thyroid	
Catecholamines	Pheochromocytoma	
Oncofetal Antigens		
α Fetoprotein	Hepatocellular carcinoma, gonadal germ cell tumor	Cirrhosis, hepatitis
Carcinoembryonic antigen	Adenocarcinomas of the colon, pancreas, lung, breast, ovary	Pancreatitis, hepatitis, inflammatory bowel disease, smoking
Enzymes		
Prostatic acid phosphatase	Prostate cancer	Prostatitis, prostatic hypertrophy
Neuron-specific enolase	Small-cell cancer of the lung, neuroblastoma	
Lactate dehydrogenase	Lymphoma, Ewing's sarcoma	Hepatitis, hemolytic anemia, many others
Tumor-Associated Proteins		
Prostate-specific antigen	Prostate cancer	Prostatitis, prostatic hypertrophy
Monoclonal immunoglobulin	Myeloma	Infection, MGUS
CA-125	Ovarian cancer, some lymphomas	Menstruation, peritonitis, pregnancy
CA 19-9	Colon, pancreatic, breast cancer	Pancreatitis, ulcerative colitis
CD30	Hodgkin's disease, anaplastic large-cell lymphoma	—
CD25	Hairy cell leukemia, adult T-cell leukemia/lymphoma	Hemophagocytic lymphohistiocytosis

Abbreviation: MGUS, monoclonal gammopathy of uncertain significance.

The recognition and treatment of depression are important components of management. The incidence of depression in cancer patients is ~25% overall and may be greater in patients with greater debility. This diagnosis is likely in a patient with a depressed mood (dysphoria) and/or a loss of interest in pleasure (anhedonia) for at least 2 weeks. In addition, three or more of the following symptoms are usually present: appetite change, sleep problems, psychomotor retardation or agitation, fatigue, feelings of guilt or worthlessness, inability to concentrate, and suicidal ideation. Patients with these symptoms should receive therapy. Medical therapy with a serotonin reuptake inhibitor such as fluoxetine (10–20 mg/d), sertraline (50–150 mg/d), or paroxetine (10–20 mg/d) or a tricyclic antidepressant such as amitriptyline (50–100 mg/d) or desipramine (75–150 mg/d) should be tried, allowing 4–6 weeks for response. Effective therapy should be continued at least 6 months after resolution of symptoms. If therapy is unsuccessful, other classes of antidepressants may be used. In addition to medication, psychosocial interventions such as support groups, psychotherapy, and guided imagery may be of benefit.

Many patients opt for unproven or unsound approaches to treatment when it appears that conventional medicine is unlikely to be curative. Those seeking such alternatives are often well educated and may be early in the course of their disease. Unsound approaches are usually hawked on the basis of unsubstantiated anecdotes and not only cannot help the patient but may be harmful. Physicians should strive to keep communications open and nonjudgmental, so that patients are more likely to discuss with the physician what they are actually doing. The appearance of unexpected toxicity may be an indication that a supplemental therapy is being taken.[2]

LONG-TERM FOLLOW-UP/LATE COMPLICATIONS

At the completion of treatment, sites originally involved with tumor are reassessed, usually by radiography or imaging techniques, and any persistent abnormality is biopsied. If disease persists, the multidisciplinary team discusses a new salvage treatment plan. If the patient has been rendered disease-free by the original treatment, the patient is followed regularly for disease recurrence. The optimal guidelines for follow-up care are not known. For many years, a routine practice has been to follow the patient monthly for 6–12 months, then every other month for a year, every 3 months for a year, every 4 months for a year, every 6 months for a year, and then annually. At each visit, a battery of laboratory and radiographic and imaging tests was obtained on the assumption that it is best to detect recurrent disease before it becomes symptomatic. However, where follow-up procedures have been examined, this assumption has been found to be untrue. Studies of breast cancer, melanoma, lung cancer, colon cancer, and lymphoma have all failed to support the notion that asymptomatic relapses are more readily cured by salvage therapy than symptomatic relapses. In view of the enormous cost of a full battery of diagnostic tests and their manifest lack of impact on survival, new guidelines are emerging for less frequent follow-up visits, during which the history and physical examination are the major investigations performed.

---

[2] Information about unsound methods may be obtained from the National Council Against Health Fraud, Box 1276, Loma Linda, CA 92354, or from the Center for Medical Consumers and Health Care Information, 237 Thompson Street, New York, NY 10012.

As time passes, the likelihood of recurrence of the primary cancer diminishes. For many types of cancer, survival for 5 years without recurrence is tantamount to cure. However, important medical problems can occur in patients treated for cancer and must be examined **(Chap. 100)**. Some problems emerge as a consequence of the disease and some as a consequence of the treatment. An understanding of these disease- and treatment-related problems may help in their detection and management.

Despite these concerns, most patients who are cured of cancer return to normal lives.

SUPPORTIVE CARE

In many ways, the success of cancer therapy depends on the success of the supportive care. Failure to control the symptoms of cancer and its treatment may lead patients to abandon curative therapy. Of equal importance, supportive care is a major determinant of quality of life. Even when life cannot be prolonged, the physician must strive to preserve its quality. Quality-of-life measurements have become common endpoints of clinical research studies. Furthermore, palliative care has been shown to be cost-effective when approached in an organized fashion. A credo for oncology could be to cure sometimes, to extend life often, and to comfort always.

Management strategies for cancer pain, nausea, and other common side effects of cancer and its treatment are outlined in **Chap. 74**. An approach to end-of-life care is provided in **Chap. 13**.

Psychosocial Support The psychosocial needs of patients vary with their situation. Patients undergoing treatment experience fear, anxiety, and depression. Self-image is often seriously compromised by deforming surgery and loss of hair. Women who receive cosmetic advice that enables them to look better also feel better. Loss of control over how one spends time can contribute to the sense of vulnerability. Juggling the demands of work and family with the demands of treatment may create enormous stresses. Sexual dysfunction is highly prevalent and needs to be discussed openly with the patient. An empathetic health care team is sensitive to the individual patient's needs and permits negotiation where such flexibility will not adversely affect the course of treatment.

Cancer survivors have other sets of difficulties. Patients may have fears associated with the termination of a treatment they associate with their continued survival. Adjustments are required to physical losses and handicaps, real and perceived. Patients may be preoccupied with minor physical problems. They perceive a decline in their job mobility and view themselves as less desirable workers. They may be victims of job and/or insurance discrimination. Patients may experience difficulty reentering their normal past life. They may feel guilty for having survived and may carry a sense of vulnerability to colds and other illnesses. Perhaps the most pervasive and threatening concern is the ever-present fear of relapse (the Damocles syndrome).

Patients in whom therapy has been unsuccessful have other problems related to the end of life.

Death and Dying The most common causes of death in patients with cancer are infection (leading to circulatory failure), respiratory failure, hepatic failure, and renal failure. Intestinal blockage may lead to inanition and starvation. Central nervous system disease may lead to seizures, coma, and central hypoventilation. About 70% of patients develop dyspnea preterminally. However, many months usually pass between the diagnosis of cancer and the occurrence of these complications, and during this period, the patient is severely affected by the possibility of death. The path of unsuccessful cancer treatment usually occurs in three phases. First, there is optimism at the hope of cure; when the tumor recurs, there is the acknowledgment of an incurable disease, and the goal of palliative therapy is embraced in the hope of being able to live with disease; finally, at the disclosure of imminent death, another adjustment in outlook takes place. The patient imagines the worst in preparation for the end of life and may go through stages of adjustment to the diagnosis. These stages include denial, isolation, anger, bargaining, depression, acceptance, and hope. Of course, patients do not all progress through all the stages or proceed through them in the same order or at the same rate. Nevertheless, developing an understanding of how the patient has been affected by the diagnosis and is coping with it is an important goal of patient management.

It is best to speak frankly with the patient and the family regarding the likely course of disease. These discussions can be difficult for the physician as well as for the patient and family. The critical features of the interaction are to reassure the patient and family that everything that can be done to provide comfort will be done. They will not be abandoned. Many patients prefer to be cared for in their homes or in a hospice setting rather than a hospital. The American College of Physicians has published a book called *Home Care Guide for Cancer: How to Care for Family and Friends at Home* that teaches an approach to successful problem-solving in home care. With appropriate planning, it should be possible to provide the patient with the necessary medical care as well as the psychological and spiritual support that will prevent the isolation and depersonalization that can attend in-hospital death.

The care of dying patients may take a toll on the physician. A "burnout" syndrome has been described that is characterized by fatigue, disengagement from patients and colleagues, and a loss of self-fulfillment. Efforts at stress reduction, maintenance of a balanced life, and setting realistic goals may combat this disorder.

End-of-Life Decisions Unfortunately, a smooth transition in treatment goals from curative to palliative may not be possible in all cases because of the occurrence of serious treatment-related complications or rapid disease progression. Vigorous and invasive medical support for a reversible disease or treatment complication is assumed to be justified. However, if the reversibility of the condition is in doubt, the patient's wishes determine the level of medical care. These wishes should be elicited before the terminal phase of illness and reviewed periodically. Information about advance directives can be obtained from the American Association of Retired Persons, 601 E Street, NW, Washington, DC 20049, 202-434-2277, or Choice in Dying, 250 West 57th Street, New York, NY 10107, 212-366-5540. Some states allow physicians to assist patients who choose to end their lives. This subject is challenging from an ethical and a medical point of view. Discussions of end-of-life decisions should be candid and involve clear informed consent, waiting periods, second opinions, and documentation. **A full discussion of end-of-life management is provided in Chap. 13.**

FURTHER READING

Bray F et al: Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74:229, 2024.

Hesketh PJ et al: Antiemetics: ASCO guideline update. J Clin Oncol 38:2782, 2020.

Kelley AS, Morrison RS: Palliative care for the seriously ill. N Engl J Med 373:747, 2015.

Martini RS et al: Integrative approaches for cancer pain management. Curr Oncol Rep 26:691, 2024.

Samala RV et al: Frequently asked questions about managing cancer pain: An update. Cleve Clin Med J 88:183, 2021.

Siegel RL et al: Cancer statistics, 2024. CA Cancer J Clin 74:12, 2024.

74

Symptom Control in Patients with Cancer

Charles L. Loprinzi, Thomas J. Smith

THE FIRST STEP: IMPORTANCE OF PROPER SYMPTOM ASSESSMENT AND RESOLUTION

Several randomized trials have now shown that for metastatic cancer patients, prompt symptom assessment and resolution lead to substantially increased survival. Basch and colleagues at Memorial Sloan Kettering Cancer Center randomized 766 metastatic solid tumor patients to usual care or once-a-week monitoring of 12 common symptoms (the patient-reported outcomes [PRO] group); these patients checked in once a week by computer or smartphone. Nurses monitored the symptom for "spikes" such as in pain and contacted the patient or brought them in for evaluation. Quality of life was significantly better in the PRO group. Surprisingly, median overall survival (OS) was 31.2 months in the PRO group and 26 months in the usual care group ($p = .03$). There was an absolute difference in OS of 6 per 100 patients at 5 years in the PRO group compared to the control group. In a similar trial of 121 stage III and IV lung cancer patients randomized to usual care or once-weekly PRO reporting, the median OS was 22.5 months versus 13.5 months. The absolute OS difference at 2 years was 10–12 of every 100 patients. In a cluster randomized trial of 52 oncology practices and 1191 patients, PROs completed each week led to statistically significant improvements in quality of life outcomes at 3 months; the primary outcome, OS, has not been reported. This extra attention to symptom assessment and management may be part of the reason that patients randomized to concurrent palliative care in addition to their cancer care live longer.

A general principle from the above trials is that the health care provider should use some form of PRO-related scale to assess symptoms. Homsi et al. evaluated 200 patients referred to the Cleveland Clinic using a 48-item symptom checklist versus what the patients volunteered. The median number of volunteered symptoms was one; the median number using a checklist was 10. Fifty-two percent of these symptoms were rated moderate or severe, 48% mild, and 53% distressing. Fatigue and pain were the most common symptoms reported in both groups.

Not everyone has access to an electronic computer-based PRO system or the ability to integrate this into the electronic medical record (EMR). Readily available and free scales such as the Edmonton Symptom Assessment System–Revised can be done on paper or electronic tablets and pasted into the EMR. The most recent scale includes questions on spirituality and financial distress and is available in many languages **(Fig. 74-1)**. Even with a nurse assisting the patient, the average completion time was only 117 s, so time should not be a barrier.

Using pain as the most common example of cancer symptoms, many clinicians have been conditioned to use pain as the fifth vital sign and almost immediately ask for a number from 0–10. First, try to establish the generators of the symptom and classify it into a treatable subtype **(Table 74-1)**. It may be helpful to categorize how much each symptom actually bothers the patient; for instance, anorexia may bother the family but not the patient. Next, clinicians can attempt to manage the cancer symptom. We have listed the most common cancer symptoms in Table 74-1 and will review each in turn, including some uncommon ones, and provide current therapeutic modalities.

INDIVIDUAL SYMPTOMS

FATIGUE

Frequency Fatigue is the most commonly reported symptom by patients, with often an 80% or higher prevalence.

Etiology Fatigue can be due to the cancer, electrolyte abnormalities especially hypercalcemia, anemia, hypothyroidism, chemotherapy, and immunotherapy such as pembrolizumab, nivolumab, and ipilimumab. Besides general fatigue due to checkpoint inhibitors, 5–22% (average 6%) of patients will develop hypothyroidism and another 1–2% de novo hypophysitis.

Assessment Check for hypothyroidism, anemia, hypercalcemia and other electrolyte abnormalities, and renal failure, also screen for depression and/or demoralization. It is important to remember that while most cases of hypothyroidism manifest in 6–8 weeks, it can appear months after the patient has stopped the immunotherapy.

Treatment The simplest treatments consist of correcting anemia, hypothyroidism, and/or electrolyte abnormalities (hypercalcemia). There are few other modalities that work quickly or reliably. Paradoxically, structured exercise (e.g., 45 min of walking daily) has been shown to be preventative and therapeutic. Rather than prescribe a new exercise regimen, ask the patient what they have done before and urge them to resume it. One can consider a trial of methylphenidate 5 mg at 7 and 11 A.M. for 48 h, with continuation if some benefit, or modafinil or congeners. Bupropion helps; in a systematic review, six of seven studies reported significant reductions in fatigue with minimal side effects. In a recent randomized trial, 40 patients were randomized to bupropion sustained release or placebo. The bupropion group had significant improvements in fatigue and quality of life but no improvements in depression. Steroids (e.g., dexamethasone 8 mg in the morning) can benefit some patients, usually toward the end of life. Long-term use has not been studied, and patients should be observed for hyperglycemia.

PAIN

Frequency At least 80% of cancer patients will experience pain during their lifetime. While secondary to fatigue in prevalence, it is the most feared symptom.

Etiology Try to characterize the type of pain **(Table 74-2)**.

Assessment Ask for a 0–10 rating for best, worst, and average pain and what an acceptable pain score would be. List what the person has tried before, including acupuncture, massage, other alternative/complementary treatments, and medications. Ascertain if the patient is on anticoagulants given the widespread use of novel oral anticoagulants (NOACs).

Treatment • **NOCICEPTIVE PAIN** For nociceptive pain, start with acetaminophen, then nonsteroidal anti-inflammatory agents, and then opioids, which should always be used with constipation-preventing measures. Recent studies have shown that the addition of acetaminophen to high doses of opioids has no additional effect. Steroids did slightly improve pain compared to placebo in cancer patients and improved fatigue, nausea, and well-being; a Cochrane systematic review found a reduction in pain of 0.84 on a 0-10 point scale at 1 week.

VISCERAL PAIN Visceral pain is among the hardest to treat of all pains. Nonsteroidal anti-inflammatory agents or neuropathic pain medications such as gabapentin or pregabalin, combined with opioids, may be required. Octreotide in combination with opioids reduces visceral hyperalgesia and may be helpful if nothing else works for malignant bowel obstruction.

NEUROPATHIC PAIN Several different classes of drugs are useful, but the number needed to treat ranges from 3 to 7 or even 10 for all of them. This necessitates 3- to 4-week trials of alternate drugs alone or in combination. Contrary to common teaching, randomized trials show that opioids do help neuropathic pain with as much benefit as gabapentin or nortriptyline; the combination is even more effective. A variety of neuropathic pain drugs are used: gabapentin, pregabalin, duloxetine, nortriptyline, amitriptyline, carbamazepine, lamotrigine, and many others. One systematic review reported that a reduction in cancer pain intensity of >1 point was unlikely with any of these drugs. Pain relief occurred in 4–8 days, if it occurred, so trials do not need to be >2–3 weeks.

Chemotherapy-induced peripheral neuropathy (CIPN) is a prominent clinical problem that is applicable to the vast majority of patients

Edmonton Symptom Assessment Scale (ESAS–FS)

Please circle the number that best describes your symptoms:

No Pain 0 1 2 3 4 5 6 7 8 9 10 Worst Pain

No Fatigue 0 1 2 3 4 5 6 7 8 9 10 Worst Fatigue

No Nausea 0 1 2 3 4 5 6 7 8 9 10 Worst Nausea

No Depression 0 1 2 3 4 5 6 7 8 9 10 Worst Depression

No Anxiety 0 1 2 3 4 5 6 7 8 9 10 Worst Anxiety

No Drowsiness 0 1 2 3 4 5 6 7 8 9 10 Worst Drowsiness

No Shortness of Breath 0 1 2 3 4 5 6 7 8 9 10 Worst Shortness of Breath

Best Appetite 0 1 2 3 4 5 6 7 8 9 10 Worst Appetite

Best Feeling of Well-being 0 1 2 3 4 5 6 7 8 9 10 Worst Feeling of Well-being

Best Sleep 0 1 2 3 4 5 6 7 8 9 10 Worst Sleep

No Financial Distress 0 1 2 3 4 5 6 7 8 9 10 Worst Financial Distress
(Distress/suffering experienced secondary to financial issues)

No Spiritual Pain 0 1 2 3 4 5 6 7 8 9 10 Worst Spiritual Pain
(Pain deep in your soul/being that is not physical)

FIGURE 74-1 Edmonton Symptom Assessment System–Revised.

TABLE 74-1 Common Cancer Symptoms and Their Association

SYMPTOM	PREVALENCE	COMMONLY FOUND CAUSES/EXAMPLES
Fatigue	70–80%	Chemotherapy Immunotherapy
Pain, all	40–70% overall	
Nociceptive		Pancreatic cancer pain
Visceral		Intestinal obstruction
Neuropathic		Chemotherapy-induced neuropathy
Incident (movement)		Bone metastases
Oral and gastrointestinal toxicity	Dependent on agents used; estimates from 30 to 40%	Standard chemotherapy such as doxorubicin Molecularly targeted agents such as palbociclib, infigratinib, everolimus, lenvatinib Radiation therapy
Nausea due to chemotherapy	Dependent on emetogenic potential of drugs administered; 10–90%	Particularly common with cisplatin and doxorubicin
Nausea not due to chemotherapy	Limited information on incidence but common	Due to small-bowel obstruction, opioids
Anorexia/cachexia	20–80% due to cancer	Most common in lung and pancreas cancers
Dyspnea	10–80% during a lifetime, more common near end of life	Most common in lung cancer patients or those with effusions, multiple pulmonary metastases
Hot flashes	Two-thirds of breast cancer and three-quarters of prostate cancer patients with androgen deprivation	Very common but often not reported by either men or women
Nasal vestibulitis	Up to 75% of patients	Those receiving taxanes, bevacizumab, etc.

TABLE 74-2 Types of Commonly Encountered Cancer Pain

TYPE OF PAIN	CAUSE	CHARACTERISTICS	EXAMPLES
Nociceptive	Pressure on nerves	Deep, dull, aching, constant and worsening with time	Pancreatic cancer pain, deep boring, and epigastric
Visceral	Distention of a hollow viscus	Cramping, bloating pain, intermittent	Intestinal obstruction, renal colic
Neuropathic	Direct damage to the nerves from cancer, treatment, or both	Local pain, sharp shooting, burning, stabbing, often with allodynia (painful sensation with normal touch) or hyperalgesia	
	Chemotherapy-induced neuropathic pain; direct damage to the longest nerves with damaged receptors and even loss of nerve fiber density	Numbness, tingling, pain, which may be mixed together; longest nerves affected most, giving a stocking-glove neuropathy	Increasingly common and dose limiting; 40–70% of people getting modern treatments. Duloxetine only proven medication, which is only moderately effective.
Incident or movement pain	Pathologic fractures, bone damage from cancer, residual damage left after cancer	Minimal pain at rest, but excruciating pain with movement "bone on bone"	Very difficult to control

receiving chemotherapy. Some chemotherapeutic agents are much more prone to cause neuropathy than are others. Among the biggest offenders are taxanes, platinums, epothilones, eribulin, vinca alkaloids, bortezomib, and lenalidomide. At the time that this chapter was written, there were no clearly established means for preventing CIPN from any agent, other than decreasing the dose or stopping the offending drug. The most promising approach at the time this chapter was penned involved attempts to decrease the concentration of paclitaxel in the peripheral extremities using cryotherapy, compression therapy, or both. Multiple preliminary trials have suggested that these approaches are helpful. In 2024, a randomized clinical trial is ongoing to compare compression therapy versus cryo-compression therapy versus a control arm (minimal compression that is not thought to be enough to decrease paclitaxel delivery to the distal extremities). At present, American Society of Clinical Oncology (ASCO) guidelines suggest that in patients who develop substantial CIPN in the midst of a treatment course, the attending clinicians should consider stopping the offending chemotherapy drug(s) or attenuating their dose.

In terms of treating established CIPN in patients who have completed all planned neurotoxic chemotherapy, there is evidence from more than one clinical trial that supports that duloxetine can attenuate established neuropathy symptoms to a small degree (0.6–1.0 on a 0–10 scale). Currently, there are no other proven treatments for treating established CIPN.

INCIDENT (MOVEMENT) PAIN DUE TO BONE METASTASES There is minimal evidence for any one modality to treat this type of pain. In one convenience trial, six of six patients with bone metastasis–incident pain had complete or near-complete responses to low doses of gabapentin (100–200 mg three times a day) added to opioids, but this has not been repeated (trial in the works). There is one case report of complete pain remission for weeks until death after four treatments with Scrambler Therapy, a form of cutaneous nerve stimulation, and one report of "burst" ketamine with morphine.

GASTROINTESTINAL AND ORAL MUCOSITIS RELATED TO TREATMENT

Frequency Oral mucositis varies by drug but is reported in 30–40% of people receiving molecularly targeted agents such as regorafenib, sorafenib, or erdafitinib, or the breast cancer drug palbociclib. The worst offending chemotherapy drug for oral mucositis, fluorouracil, is not used frequently anymore. The most common cause of oral mucositis among commonly used chemotherapy drugs is doxorubicin.

Gastrointestinal (GI) mucositis can involve the entire GI tract, and estimates range from 50–80% of cancer patients experiencing some level of chemotherapy-induced diarrhea.

Etiology The most common cause is interference by the tyrosine kinase inhibitors with endothelial growth factor receptors (GFRs) or vascular GFRs. Drugs used for both chemotherapy and immune suppression, such as everolimus, cause both inflammation and the usual disruption of endothelial cells. At its worst, the entire GI tract can be denuded, and the intestine cannot reabsorb the 25 L of intestinal fluid made daily, producing high-output diarrhea. Most mucositis happens 7–10 days after treatment.

Assessment If the time course is right and the patient has received a drug known to cause mucositis, suspicion should be high. Other considerations for oral mucositis include thrush, aphthous stomatitis, herpes, or other infections. Exam may show denuded raw areas that are very painful.

GI mucositis is usually manifest by diarrhea, often accompanied by cramps and pain. The diarrhea can be copious, resulting in liters per day. If the time course is not right (very soon or weeks after cancer therapy initiation), other causes such as infections or too many laxatives should be sought.

Treatment For patients receiving bolus fluorouracil, sucking on ice chips during infusion (oral cryotherapy) reduces the incidence and severity by ~50%. There is conflicting evidence on the use of oral dexamethasone or other steroid rinses to prevent oral mucositis from drugs such as everolimus and palbociclib. Most studies show improvement after treatment with glucocorticoid rinses, regardless of the steroid. Viscous lidocaine 2% is commonly recommended for pain. For ulcerations, topical high-potency steroids are indicated, and a short course of high-dose steroids may be necessary.

For GI mucositis manifested as diarrhea, loperamide is the first choice, up to 24 mg a day. Octreotide up to 300 mg a day is reserved for refractory cases. Glucocorticoids are also commonly employed.

NAUSEA AND VOMITING RELATED TO CHEMOTHERAPY

Frequency and Etiology It is well established that chemotherapy can cause substantial nausea and/or vomiting, noting that this problem is strongly associated with individual chemotherapy agents. Some agents, such as cisplatin and doxorubicin, cause substantial nausea and vomiting, whereas other agents, such as fluorouracil, do not cause any substantial nausea and vomiting.

Assessment Clinical determination of whether a patient developed nausea or vomiting from chemotherapy is obtained by talking to the patient and/or their family and asking questions regarding these symptoms. This is often done when the patient comes back for another cycle of therapy. Clinical trials use established PRO data that could also be used in clinical practice. In clinical trials, data are usually obtained before chemotherapy and then daily for a few days after each chemotherapy dose.

Treatment With highly emetogenic chemotherapy regimens, antiemetic drugs are started with initiation of chemotherapy and commonly given for a few days thereafter. Dexamethasone was established as a helpful agent in the 1970s. In the 1990s, 5-HT$_3$ receptor antagonists became established agents. In the 2000s, NK1 receptor antagonists were established as helpful agents for decreasing nausea and vomiting in the days following highly emetogenic chemotherapy. These agents did not decrease nausea and vomiting in the first 24 h, as did the 5-HT$_3$ receptor antagonists; thus, NK1 receptor antagonists were added to the antiemetic cocktail as opposed to replacing 5-HT$_3$ receptor antagonists. In the 2010s, olanzapine was demonstrated to decrease nausea and vomiting when added to the three drug classes discussed above. Olanzapine appears to be the most effective of all of the drugs mentioned in this section, potentially paving the way for decreasing use of other agents in the four-drug antiemetic cocktail described above. Decreasing dexamethasone doses and/or durations might decrease long-term dexamethasone-associated toxicity.

NAUSEA AND VOMITING UNRELATED TO CHEMOTHERAPY

Frequency No recent data are available, but it is estimated that this problem affects up to 40–70% of patients.

Etiology Nausea and vomiting in patients with cancer, not associated with chemotherapy, can be associated with multiple other etiologies, such as radiation therapy, bowel obstruction, other medications, electrolyte abnormalities, post-anesthesia, and cerebral metastases.

Assessment Clinical determination of whether a patient has nausea or vomiting is obtained from talking to the patient and/or their family and asking questions regarding these symptoms. In clinical trials, data are generally obtained daily, including the day prior to the planned intervention and then for a few days following a proposed intervention.

Treatment A variety of drugs have been studied for treating nausea and vomiting in this situation, but none of them are very effective; these include diphenhydramine, metoclopramide, prochlorperazine, haloperidol, methotrimeprazine, dexamethasone, ondansetron, and dronabinol.

In 2020, a double-blind, placebo-controlled clinical trial demonstrated that a relatively low dose of olanzapine (5 mg/d) was very effective for decreasing nausea and vomiting in patients with advanced cancer who had not received chemotherapy or radiation therapy for at least 2 weeks prior to study entry. In this trial involving a total of 30 patients, nausea scores decreased from 9/10 to 2/10 on the day after olanzapine was started, in comparison to the placebo group whose nausea scores were 9/10 on both the day before and the day after the first dose of the placebo ($p < .001$). Another trial, published in 2023, demonstrated that 2.5 mg/d of olanzapine markedly reduced nausea and vomiting in patients with advanced cancer who were not receiving highly emetogenic chemotherapy.

ANOREXIA AND CACHEXIA

Frequency Up to 20% of all cancer deaths are strongly associated with cancer cachexia. Anorexia is even more common, affecting 5–25% of community-dwelling adults and twice that number of patients with cancer.

Etiology Although most anorexia is caused by release of hormones by cancer, it can be aggravated by drug treatments, radiation, mechanical difficulties in eating, and food insecurity. Cachexia is multifactorial and characterized by weight loss of skeletal muscle and adipose tissue, an imbalance in metabolic regulation, and reduced food intake. In general, cancer cachexia cannot be reversed by simply replacing calories and nutrients.

Assessment Anorexia is determined simply by asking the patient and family about appetite. Cachexia has been variously defined as 5 or 10% loss of precancer weight. More precise definitions have included the presence of fatigue, anorexia, increased inflammatory markers such as C-reactive protein, body mass index <20, and even sarcopenia on a computed tomography scan.

Treatment Treatment is often unsatisfactory for both patients and their families. A systematic review done for an ASCO guideline regarding this topic showed that dietary counseling was associated with increased weight gain in some but not all trials and can be reasonably offered. The magnitude of this approach, however, is not large. Enteral feeding tubes and intravenous nutrition are not recommended and should not be used routinely. For anorexia, olanzapine is recommended as the first-line agent and the only agent with good evidence of benefit without potentially serious harm. In the largest randomized controlled trial, 124 solid tumor patients received olanzapine 2.5 mg at bedtime versus placebo. Sixty percent of the olanzapine patients increased their weight by at least 5% versus 9% of placebo patients ($p < .001$). Mirtazapine, which is commonly prescribed, was ineffective in a large randomized trial and should not be used. Megestrol acetate and corticosteroids do increase appetite in afflicted patients and lead to some weight gain but have untoward side effects and do not appear to improve patient survival. A novel approach is to block the elevated levels of growth differentiation factor 15 (GDF-15), a serum cytokine that is elevated in cachexia with ponsegromab. This monoclonal antibody was associated with improved appetite, weight, and physical activity in a randomized phase 2 trial, but it has not been approved by the FDA.

CONSTIPATION

Frequency Constipation occurs in 20–90% of all cancer patients at some time in their trajectory and can be both painful and associated with anorexia and nausea.

Etiology Causes include commonly used drugs such as opioids, acetaminophen, ondansetron, chemotherapy (vandetanib, thalidomide, lenalidomide), and noncancer drugs such as diuretics and antidementia drugs. Inactivity, dehydration, a low-fiber diet, hypercalcemia, and hypothyroidism (including up to 20% of patients receiving checkpoint inhibitors) can all be contributing factors.

Assessment Clinicians should ask what the personal bowel habits have been and when the patient last had a bowel movement. If it was 3 or more days ago, constipation is highly likely. Using a specific scale or grading the consistency of the stool is not recommended.

Treatment Good bowel hygiene starts with prevention and using drugs that the patient can tolerate. Every opioid prescription should be accompanied by a plan to prevent constipation. Senna is the first choice, starting at 1–2 tablets a day and working up to 8, along with sufficient hydration. Some patients may have cramps from the stimulant action requiring multiple doses a day or switching drugs. Polyethylene glycol is added next, if needed and/or if senna is not tolerated. Lactulose and sorbitol are less commonly used because they commonly are less palatable. Magnesium oxide (1 oz of a 300-mL bottle hourly until movements start) can be helpful. Patients on opioids who are refractory to these measures usually have success with opioid antagonists that reverse the opioid in the periphery but not the brain. Because of cost, these agents currently may require insurance preauthorization, so they are likely not used very often.

DYSPNEA

Frequency Dyspnea (or breathlessness or air hunger) is, like pain, a subjective experience that can only be assessed by the patient. Dyspnea is common, with 10–70% of patients reporting it, especially toward the end of life. Dyspnea portends a poor prognosis in advanced cancer with an average survival of only days, weeks, or months, and should trigger serious illness conversations such as being fully truthful about prognosis and creation of an advance directive. ASCO guidelines recommend a systematic assessment for dyspnea at every inpatient and outpatient visit in patients with advanced cancer.

Etiology As with most cancer symptoms, dyspnea is multifactorial. Lung dysfunction may be due to the cancer, radiation, chemotherapy, and/or immunotherapy. With immunotherapy, the average time to onset was 52 days in one trial, and the incidence in one large series was 9.5%.

Assessment The first attempt should be to find something that is reversible (e.g., hypoxia, pneumonia, pulmonary embolism, pleural effusion, chronic obstructive pulmonary disease, asthma, or bronchial constriction due to cancer). Next should come a careful review of medication history, especially for checkpoint inhibitors, remembering that pneumonitis may develop months after the immunotherapy has stopped. The impact of the dyspnea on the patient and the family should be assessed. Management of immunotherapy-related lung disease may best be done by experts, as the therapies can be challenging and have been rapidly changing.

Treatment Treatment of the cancer, if deemed to have a good chance of success, should be a mainstay. Other contributing factors such as anxiety or chronic obstructive pulmonary disease (COPD) should be maximally treated. High-dose steroids are commonly used in the setting of lymphangitic carcinomatosis, COPD, or asthma. Opioids at low doses are generally safe and can be helpful in relieving air hunger, and low-dose benzodiazepines can help relieve anxiety. A fan blowing cold air across the face can give some comfort.

Supplemental oxygen should be available for patients with hypoxemia (i.e., SpO$_2$ ≤90% on room air). High-flow nasal oxygen (HFNO) is often used as a bridge to a next therapy, if there is one; commitment to HFNO complicates referral to hospice or home due to the cost. Referral to palliative care should always be considered, along with truthful information about prognosis and a serious illness conversation including advance directives.

HOT FLASHES

Frequency Hot flashes are common; they occur in about two-thirds of postmenopausal women treated for breast cancer, and almost half have night sweats. Of men treated with chemical or surgical orchiectomy for prostate cancer, three-quarters will have hot flashes. Common knowledge, in the not too distant past, was that hot flashes usually only lasted a couple of years. However, it is now understood that patients can have them for decades.

Etiology Lack of estrogen or testosterone is the proximate cause of hot flashes, especially when the decrease is rapid.

Assessment Asking about hot flashes, including frequency, severity, and interference with life, should be part of the appropriate care of prostate or breast cancer patients.

Treatment • FEMALES In the 1990s, four separate individuals clinically noted that patients receiving newer antidepressants, at least new at that time, appeared to have reduced hot flashes; these antidepressants were venlafaxine, paroxetine, sertraline, and fluoxetine. Such observations resulted in randomized, placebo-controlled clinical trials that showed that these and other antidepressants substantially helped women with hot flashes. The authors of this chapter, when it is decided to utilize an antidepressant for treating hot flashes, recommend citalopram, 20 mg/d, as it appears as effective as other antidepressants and is well tolerated. Gabapentin also has efficacy against hot flashes and reduces anxiety in breast cancer survivors. Oxybutynin, immediate or sustained release, also showed efficacy in randomized trials. This can be a relatively inexpensive treatment approach. Caution is recommended in older women, as there are concerns regarding mental status changes in the setting of long-term use. In refractory patients, 20–40 mg of megestrol acetate a day can control hot flashes. For women who desire a one-time intramuscular injection, 400–500 mg of medroxyprogesterone acetate can nicely decrease hot flashes for many months.

Fezolinetant, an NK3 receptor antagonist, became clinically available in 2023 as a nonhormonal agent for treating hot flashes in women in general. Although it works as well as some of the other treatments noted above, cross-study comparisons of the efficacy of this drug to the other options mentioned above suggest that hot flash reductions are similar. In a series of studies that included >1300 patients, some with a history of breast cancer (which amounted to ~75% of the subjects) and some without a history of breast cancer (the other 25% of patients), hot flash outcomes were similar. Additionally, this article revealed that hot flash reductions were similar whether or not a patient was receiving tamoxifen.

MALES Pilot trials of venlafaxine and paroxetine supported that they were useful for alleviating hot flashes in men with prostate cancer. However, a randomized trial was not able to demonstrate significant benefit from venlafaxine for controlling hot flashes in men. A prospective, randomized, placebo-controlled trial demonstrated that 900 mg/d of gabapentin works well for treating hot flashes in men, similarly to how it works in women.

Progesterone analogues also decrease hot flashes in men, similarly to how they do so in women.

Case reports support that oxybutynin is helpful in men with hot flashes; a recently completed randomized, double-blind, placebo-controlled clinical trial demonstrated that this drug nicely decreased hot flashes in this population, too.

NASAL VESTIBULITIS

Frequency Nasal vestibulitis refers to an unpleasant inflammation of the vestibule of the nose. An observational interview study of patients who had received at least 6 weeks of chemotherapy noted that >70% of those receiving taxanes and >80% of those receiving bevacizumab experienced untoward nasal symptoms, such as pain, bleeding, and/or scabbing. A prospective study of patients receiving a variety of different chemotherapy regimens illustrated that >75% of patients who received paclitaxel, nab-paclitaxel, or bevacizumab developed bothersome nasal symptoms in the following weeks to months.

Etiology Chemotherapy can damage the epithelial cells in the nose, leading to dryness, cracking, and secondary infection. Like mucositis, it generally heals in 7–10 days but is a potential source for infection.

Assessment Asking about symptoms of dryness, bleeding, crusting, or pain and an exam should confirm local tissue irritation.

Treatment Nasal sprays of saline and of rose geranium in sesame oil have been used to treat this clinical problem, with data supporting that the latter is more beneficial. This will have to be compounded; the recipe is available from the authors or the Mayo Clinic.

FURTHER READING

Basch E et al: Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. JAMA 318:197, 2017.

Correa-Morales JE et al: Cancer and non-cancer fatigue treated with bupropion: A systematic review. J Pain Symptom Manage 65:e21, 2023.

Finnerup NB et al: Neuropathic pain: From mechanisms to treatment. Physiol Rev 101:259, 2021.

Groarke JD et al: Ponsegromab for the Treatment of Cancer Cachexia. N Engl J Med 391:2291, 2024.

Homsi J et al: Symptom evaluation in palliative medicine: Patient report vs systematic assessment. Support Care Cancer 14:444, 2006.

Hui D, Bruera E: The Edmonton Symptom Assessment System 25 years later: Past, present, and future developments. J Pain Symptom Manage 53:630, 2017.

Loprinzi CL et al: Prevention and management of chemotherapy-induced peripheral neuropathy in survivors of adult cancers: ASCO guideline update. J Clin Oncol 38:3325, 2020.

Mannix KA: Palliation of nausea and vomiting, in *Oxford Textbook of Palliative Medicine*, 2nd ed. Doyle D et al (eds). Oxford, UK, Oxford University Press, 1998, p. 48.

Molinares D et al: Chemotherapy-induced peripheral neuropathy: Diagnosis, agents, general clinical presentation, and treatments. Curr Oncol Rep 25:1227, 2023.

Navari RM et al: Olanzapine for the treatment of advanced cancer-related chronic nausea and/or vomiting: A randomized pilot trial. JAMA Oncol 6:895, 2020.

Roeland EJ et al: Cancer Cachexia Expert Panel. Cancer cachexia: ASCO guideline rapid recommendation update. J Clin Oncol 41:4178, 2023.

Smith TJ, Saiki CB: Cancer pain management. Mayo Clin Proc 90:1428, 2015.

75

Prevention and Early Detection of Cancer

Jennifer M. Croswell, Otis W. Brawley,
Barnett S. Kramer

Improved understanding of carcinogenesis has allowed cancer prevention and early detection to expand beyond identification and avoidance of carcinogens. Specific interventions to reduce cancer mortality by preventing cancer in those at risk and effective screening for early detection of cancer are the goals.

Carcinogenesis is a process that usually extends over years, a continuum of discrete tissue and cellular changes over time resulting in aberrant physiologic processes. Prevention concerns the identification and manipulation of the biologic, environmental, social, and genetic factors in the causal pathway of cancer. Examination of national epidemiologic patterns can provide indicators of the relative contributions of advances in prevention, screening, and therapy in progress against cancer, but randomized trials provide the best evidence to guide practice, especially in the healthy general population.

EDUCATION AND HEALTHFUL HABITS

Public education on the avoidance of identified risk factors for cancer and encouraging healthy habits contributes to cancer prevention. The clinician is a powerful messenger in this process. The patient-provider encounter provides an opportunity to teach patients about the hazards of smoking, influence of a healthy lifestyle and other exposures, and use of proven cancer screening methods.

■ SMOKING CESSATION

Tobacco smoking is a strong, modifiable risk factor for cardiovascular disease, pulmonary disease, and cancer. Smokers have an ~1 in 3 lifetime risk of dying prematurely from a tobacco-related cancer, cardiovascular, or pulmonary disease, and cigarette smoking shortens life expectancy, on average, by a decade. Tobacco use causes more deaths from cardiovascular disease than from cancer. Lung cancer and cancers of the larynx, oropharynx, esophagus, kidney, bladder, colon, pancreas, stomach, and uterine cervix are all tobacco related.

The number of cigarettes smoked per day and the level of inhalation of cigarette smoke are correlated with risk of lung cancer mortality. Light- and low-tar cigarettes are not safer because smokers tend to inhale them more frequently and deeply.

Those who stop smoking have a 30–50% lower 10-year lung cancer mortality rate compared to those who continue smoking, despite the fact that some carcinogen-induced gene mutations persist for years after smoking cessation. Smoking cessation and avoidance would save more lives from cancer than any other public health activity.

The risk of tobacco smoke is not limited to the smoker. Environmental tobacco smoke, known as secondhand or passive smoke, is carcinogenic and associated with a variety of respiratory illnesses in exposed children.

Tobacco use prevention is a pediatric issue. More than 80% of adult American smokers began smoking before the age of 18 years. Traditional cigarette smoking has declined substantially in recent years: for example, for high school seniors between 2012 and 2019, prevalence of past 30-day cigarette smoking dropped from 19% to 7% for boys and 7% to 2% for girls. Electronic cigarettes, on the other hand, are rapidly increasing in use: in 2020, approximately 20% of high school students reported being current electronic cigarette users. Counseling of adolescents and young adults is critical to prevent all forms of tobacco use. A clinician's simple advice can be of benefit. Providers should query patients on tobacco use and offer smokers assistance in quitting or referrals to cessation programs.

Current approaches to smoking cessation recognize nicotine in tobacco as addicting **(Chap. 465)**. The smoker who is quitting goes through identifiable stages, including contemplation of quitting, an action phase in which the smoker quits, and a maintenance phase. Smokers who quit completely are more likely to be successful than those who gradually reduce the number of cigarettes smoked or change to lower-tar or lower-nicotine cigarettes. Organized cessation programs may help individual efforts. Heavy smokers may need an intensive broad-based cessation program that includes counseling, behavioral strategies, and pharmacologic adjuncts, such as nicotine replacement (gum, patches, sprays, lozenges, and inhalers), bupropion, and/or varenicline. Electronic cigarettes have been advocated as a tool to achieve smoking cessation or as a harm reduction strategy in adults, but the net effects of electronic cigarettes on health are poorly studied. Absence of strict manufacturing controls of vaping material has produced serious injury.

The health risks of cigars are similar to those of cigarettes. Smoking one or two cigars daily doubles the risk for oral and esophageal cancers; smoking three or four cigars daily increases the risk of oral cancers more than eightfold and esophageal cancer fourfold. The risks of occasional use are unknown.

Smokeless tobacco also represents a substantial health risk. Chewing tobacco is linked to dental caries, gingivitis, oral leukoplakia, and oral cancer. The systemic effects of smokeless tobacco (including snuff) may increase risks for other cancers. Esophageal cancer is linked to carcinogens in tobacco dissolved in saliva and swallowed.

■ PHYSICAL ACTIVITY

Physical activity is associated with a decreased risk of colon and breast cancer. A variety of mechanisms have been proposed. However, such studies are prone to confounding factors such as recall bias, association of exercise with other health-related practices, and effects of preclinical cancers on exercise habits (reverse causality).

■ DIET MODIFICATION

International ecologic studies suggest that diets high in fat are associated with increased risk for cancers of the breast, colon, prostate, and endometrium. Despite correlations, dietary fat has not been proven to cause cancer. Case-control and cohort epidemiologic studies give conflicting results. Diet is a highly complex exposure to many nutrients and chemicals. Low-fat diets are associated with many dietary changes beyond simple subtraction of fat. Other lifestyle factors are also associated with adherence to a low-fat diet.

In some observational studies, dietary fiber has been associated with a reduced risk of colonic polyps and invasive cancer of the colon. Two large prospective cohort studies of >100,000 health professionals showed no association between fruit and vegetable intake and risk of cancer, however. Cancer-protective effects of increasing fiber and lowering dietary fat have not been shown in the context of a prospective clinical trial. The Polyp Prevention Trial randomly assigned 2000 elderly persons, who had polyps removed, to a low-fat, high-fiber diet versus routine diet for 4 years. No differences were noted in polyp formation.

The U.S. National Institutes of Health Women's Health Initiative, launched in 1994, was a long-term clinical trial enrolling >100,000 women age 45–69 years. It placed women into 22 intervention groups. Participants received calcium/vitamin D supplementation; hormone replacement therapy; and counseling to increase exercise, eat a low-fat diet with increased consumption of fruits, vegetables, and fiber, and cease smoking. The study showed that although dietary fat intake was lower in the diet intervention group, invasive breast cancers were not reduced over an 8-year follow-up period compared to the control group. Additionally, no reduction was seen in the incidence of colorectal cancer in the dietary intervention arm. In the aggregate, cohort studies and randomized trials suggest that reduction of red meat or processed meat consumption has a small (if any) effect on cancer incidence and mortality, although the overall evidence base is weak. Evidence does not currently establish the anticarcinogenic value of vitamin, mineral, or nutritional supplements in amounts greater than those provided by a balanced diet.

ENERGY BALANCE

Risk of certain cancers appears to increase modestly (relative risks generally in the 1.0–2.0 range) as body mass index (BMI) increases beyond 25 kg/m^2. A cohort study of >5 million adults included in the U.K. Clinical Practice Research Datalink (a primary care database) found that each 5 kg/m^2 increase in BMI was linearly associated with cancers of the uterus, gallbladder, kidney, cervix, thyroid, and leukemia. High BMI appears to have an inverse association with prostate and premenopausal breast cancer.

SUN AVOIDANCE

Nonmelanoma skin cancers (basal cell and squamous cell) are induced by cumulative exposure to ultraviolet (UV) radiation. Evidence for a direct association between cumulative UV exposure and melanoma is weaker. Reduction of sun exposure through use of protective clothing and changing patterns of outdoor activities can reduce skin cancer risk. Sunscreens decrease the risk of actinic keratoses, the precursor to squamous cell skin cancer, but melanoma risk may not be reduced. Sunscreens prevent burning, but they may encourage more prolonged exposure to the sun and may not filter out wavelengths of energy that cause melanoma.

Appearance-focused behavioral interventions in young women can decrease indoor tanning use and other UV exposures and may be more effective than messages about long-term cancer risks. Those who recognize themselves as being at risk tend to be more compliant with sun-avoidance recommendations. Risk factors for melanoma include a propensity to sunburn, a large number of benign melanocytic nevi, and atypical nevi. However, about 70–80% of melanomas arise de novo, rather than from existing benign nevi.

CANCER CHEMOPREVENTION

Chemoprevention involves the use of specific natural or synthetic chemical agents to reverse, suppress, or prevent carcinogenesis before the development of invasive malignancy.

Cancer develops through an accumulation of tissue abnormalities associated with genetic and epigenetic changes, and growth regulatory pathways that are potential points of intervention to prevent cancer. The initial changes are termed *initiation*. The alteration can be inherited or acquired through the action of physical, infectious, or chemical carcinogens. Like most human diseases, cancer arises from an interaction between genetics and environmental exposures (Table 75-1). Influences that cause the initiated cell and its surrounding tissue microenvironment to progress through the carcinogenic process and change phenotypically are termed *promoters*. Promoters include hormones such as androgens, linked to prostate cancer, and estrogen, linked to breast and endometrial cancer. The difference between an initiator and promoter is indistinct; some components of cigarette smoke are "complete carcinogens," acting as both initiators and promoters. Cancer can be prevented or controlled through avoidance of the factors that cause cancer initiation, promotion, or progression. Compounds of interest in chemoprevention may have antimutagenic, hormone modulation, anti-inflammatory, antiproliferative, or proapoptotic activity (or a combination).

CHEMOPREVENTION OF CANCERS OF THE UPPER AERODIGESTIVE TRACT

Smoking causes diffuse epithelial injury in the oral cavity, esophagus, and lung. Patients cured of squamous cell cancers of the lung, esophagus, and oral cavity are at risk (as high as 5% per year) of developing second cancers of the upper aerodigestive tract. Cessation of cigarette smoking does not markedly decrease the cured cancer patient's risk of second malignancy, even though it does lower the cancer risk in those who have never developed a malignancy.

Persistent oral human papillomavirus (HPV) infection, particularly HPV-16, increases the risk for cancers of the oropharynx. This association exists even in the absence of other risk factors such as smoking or alcohol use (although the magnitude of increased risk appears greater than additive when HPV infection and smoking are both present). Oral HPV infection is believed to be largely sexually acquired. The use of the HPV vaccine is associated with a reduction in prevalence of oropharyngeal infection rates and may eventually reduce oropharyngeal cancer rates.

TABLE 75-1 Suspected Carcinogens

CARCINOGENS[a]	ASSOCIATED CANCER OR NEOPLASM
Alkylating agents	Acute myeloid leukemia, bladder cancer
Androgens	Prostate cancer
Aromatic amines (dyes)	Bladder cancer
Arsenic	Cancer of the lung, skin
Asbestos	Cancer of the lung, pleura, peritoneum
Benzene	Acute myelocytic leukemia
Chromium	Lung cancer
Diethylstilbestrol (prenatal)	Vaginal cancer (clear cell)
Epstein-Barr virus	Burkitt's lymphoma, nasal T-cell lymphoma
Estrogens	Cancer of the endometrium, liver, breast
Ethyl alcohol	Cancer of the breast, liver, esophagus, head and neck
Helicobacter pylori	Gastric cancer, gastric mucosa-associate lymphoid tissue (MALT) lymphoma
Hepatitis B or C virus	Liver cancer
Human immunodeficiency virus	Non-Hodgkin's lymphoma, Kaposi's sarcoma, squamous cell carcinomas (especially of the urogenital tract)
Human papillomavirus	Cancers of the cervix, anus, oropharynx
Human T-cell lymphotropic virus type 1 (HTLV-1)	Adult T-cell leukemia/lymphoma
Immunosuppressive agents (azathioprine, cyclosporine, glucocorticoids)	Non-Hodgkin's lymphoma
Ionizing radiation (therapeutic or diagnostic)	Breast, bladder, thyroid, soft tissue, bone, hematopoietic, and many more
Nitrogen mustard gas	Cancer of the lung, head and neck, nasal sinuses
Nickel dust	Cancer of the lung, nasal sinuses
Diesel exhaust	Lung cancer (miners)
Phenacetin	Cancer of the renal pelvis and bladder
Polycyclic hydrocarbons	Cancer of the lung, skin (especially squamous cell carcinoma of scrotal skin)
Radon gas	Lung cancer
Schistosomiasis	Bladder cancer (squamous cell)
Sunlight (ultraviolet)	Skin cancer (squamous cell and likely melanoma)
Tobacco (including smokeless)	Cancer of the upper aerodigestive tract, bladder, kidney
Vinyl chloride	Liver cancer (angiosarcoma)

[a] Agents that are thought to act as cancer initiators and/or promoters.

Oral leukoplakia, a premalignant lesion commonly found in smokers, has been used as an intermediate marker of chemopreventive activity in smaller shorter-duration, randomized, placebo-controlled trials. Although therapy with high, relatively toxic doses of isotretinoin (13-*cis*-retinoic acid) causes regression of oral leukoplakia, more tolerable doses of isotretinoin have not shown benefit in the prevention of head and neck cancer.

Several large-scale trials have assessed agents in the chemoprevention of lung cancer in patients at high risk. In the α-tocopherol/β-carotene (ATBC) Lung Cancer Prevention Trial, participants were male smokers with an average 30-plus pack-year history, age 50–69 years at entry. After median follow-up of 6 years, lung cancer incidence and mortality were statistically significantly *increased* in those receiving β-carotene. α-Tocopherol had no effect on lung cancer mortality, but patients receiving it had a higher incidence of hemorrhagic stroke.

The β-Carotene and Retinol Efficacy Trial (CARET) involved 17,000 American smokers and workers with asbestos exposure. This trial also demonstrated harm from β-carotene: a lung cancer rate of 5 per 1000 subjects per year for those taking placebo versus 6 per 1000 subjects per year for those taking β-carotene.

The ATBC and CARET results demonstrate the importance of testing chemoprevention hypotheses in randomized trials before widespread implementation because the results contradict a number of observational studies.

CHEMOPREVENTION OF COLON CANCER

No agent with sufficient evidence exists at present for the prevention of colon cancer in the general population.

A meta-analysis of four randomized controlled trials primarily designed to examine aspirin's effects on cardiovascular events found no association between low-dose aspirin use and colorectal cancer incidence at up to 10 years of follow-up. The same review reported highly variable results for colorectal cancer mortality and aspirin use and noted that, overall, studies have not been adequately powered to assess aspirin's effects on colorectal cancer deaths. As such, the evidence is insufficient that nonsteroidal anti-inflammatory drugs (NSAIDs) reduce colorectal cancer incidence or mortality. However, there is clear evidence that regular NSAID use is associated with important bleeding harms, such as major gastrointestinal and intra- and extracranial bleeding, particularly with increasing dosage and age.

Cyclooxygenase-2 (COX-2) inhibitors have been considered for colorectal cancer and polyp prevention. Trials with COX-2 inhibitors were initiated, but an increased risk of cardiovascular events in those taking the COX-2 inhibitors was noted, so these agents are not suitable for chemoprevention in the general population.

The Women's Health Initiative demonstrated that postmenopausal women taking estrogen plus progestin have a 44% lower relative risk of colorectal cancer compared to women taking placebo. Of >16,600 women randomized and followed for a median of 5.6 years, 43 invasive colorectal cancers occurred in the hormone group and 72 in the placebo group. The positive effect on colon cancer is mitigated by the modest increase in cardiovascular and breast cancer risks associated with combined estrogen plus progestin therapy, and guidelines do not recommend hormonal therapy as a standard preventive agent for colorectal cancer incidence or mortality.

Most case-control and cohort studies have not confirmed early reports of an association between regular statin use and a reduced risk of colorectal cancer. No randomized controlled trials have addressed this hypothesis. A meta-analysis of statin use showed no protective effect of statins on overall cancer incidence or death.

CHEMOPREVENTION OF BREAST CANCER

Tamoxifen is an antiestrogen with partial estrogen agonistic activity in some tissues, such as endometrium and bone. One of its actions is to upregulate transforming growth factor β, which decreases breast cell proliferation. In a randomized placebo-controlled prevention trial involving >13,000 pre- and postmenopausal women at high risk, tamoxifen decreased the risk of developing breast cancer by 49% (from 43.4 to 22 per 1000 women) after a median follow-up of nearly 6 years. Tamoxifen also reduced bone fractures; a small increase in risk of endometrial cancer, stroke, pulmonary emboli, and deep vein thrombosis was noted. The International Breast Cancer Intervention Study (IBIS-I) and the Italian Randomized Tamoxifen Prevention Trial also demonstrated reductions in breast cancer incidence with tamoxifen use. A trial comparing tamoxifen with another selective estrogen receptor modulator, raloxifene, performed in postmenopausal women showed that raloxifene is comparable to tamoxifen in cancer prevention but without the risk of endometrial cancer. Raloxifene was associated with a smaller reduction in invasive breast cancers and a trend toward more noninvasive breast cancers, but fewer thromboembolic events than tamoxifen; the drugs are similar in risks of other cancers, fractures, ischemic heart disease, and stroke. Both tamoxifen and raloxifene (the latter for postmenopausal women only) have been approved by the U.S. Food and Drug Administration (FDA) for reduction of breast cancer in women at high risk for the disease (1.66% risk at 5 years based on the Gail risk model: *http://www.cancer.gov/bcrisktool/*).

Because the aromatase inhibitors (anastrozole, exemestane, and letrozole) are even more effective than tamoxifen in adjuvant breast cancer treatment, it has been hypothesized that they would be more effective in breast cancer prevention. A randomized, placebo-controlled trial of exemestane reported a 65% relative reduction (from 5.5 to 1.9 per 1000 women) in the incidence of invasive breast cancer in women at elevated risk after a median follow-up of about 3 years. Common adverse effects included arthralgias, hot flashes, fatigue, and insomnia. No trial has directly compared aromatase inhibitors with selective estrogen receptor modulators for breast cancer chemoprevention.

CHEMOPREVENTION OF PROSTATE CANCER

Finasteride and dutasteride are 5-α-reductase inhibitors. They inhibit conversion of testosterone to dihydrotestosterone (DHT), a potent stimulator of prostate cell proliferation. The Prostate Cancer Prevention Trial (PCPT) randomly assigned men age 55 years or older at average risk of prostate cancer to finasteride or placebo. All men in the trial were being regularly screened with prostate-specific antigen (PSA) levels and digital rectal examination. After 7 years of therapy, the overall incidence of prostate cancer was 18.4% in the finasteride arm, compared with 24.4% in the placebo arm, a statistically significant reduction. However, the finasteride group had more patients with tumors of Gleason score 7 and higher compared with the placebo arm (6.4 vs 5.1%). Long-term (10–15 years) follow-up did not reveal any statistically significant differences in overall or prostate cancer–specific mortality between all men in the finasteride and placebo arms or in men diagnosed with prostate cancer, but the power to detect a difference was limited.

Dutasteride has also been evaluated as a preventive agent for prostate cancer. The Reduction by Dutasteride of Prostate Cancer Events (REDUCE) trial was a randomized double-blind trial in which ~8200 men with an elevated PSA (2.5–10 ng/mL for men age 50–60 years and 3–10 ng/mL for men age 60 years or older) and negative prostate biopsy on enrollment received daily 0.5 mg of dutasteride or placebo. The trial found a statistically significant 23% relative risk reduction in the incidence of biopsy-detected prostate cancer in the dutasteride arm at 4 years of treatment (659 cases vs 858 cases, respectively). Overall, across years 1 through 4, no difference was seen between the arms in the number of tumors with a Gleason score of 7 to 10; however, during years 3 and 4, there was a statistically significant difference in tumors with Gleason score of 8 to 10 in the dutasteride arm (12 tumors vs 1 tumor, respectively).

The finding of an apparent increased incidence of higher-grade tumors likely represents an increased sensitivity of PSA and digital rectal exam for high-grade tumors in men receiving 5-α-reductase inhibitors due to a decrease in prostatic volume. Although detection bias may have accounted for the finding, a causative role cannot be conclusively dismissed. These agents are therefore not FDA-approved for prostate cancer prevention.

Because all men in both the PCPT and REDUCE trials were being screened and because screening approximately doubles the rate of prostate cancer, it is not known if finasteride or dutasteride decreases the risk of prostate cancer in men who are not being screened or simply reduces the risk of non-life-threatening cancers detectable by screening.

Several favorable laboratory and observational studies led to the formal evaluation of selenium and α-tocopherol (vitamin E) as potential prostate cancer preventives. The Selenium and Vitamin E Cancer Prevention Trial (SELECT), after a median follow-up of 7 years, found a trend toward an *increased* risk of developing prostate cancer for men taking vitamin E alone as compared to placebo (hazard ratio 1.17; 95% confidence interval, 1.004–1.36).

VACCINES AND CANCER PREVENTION

A number of infectious agents cause cancer. Hepatitis B and C viruses are linked to liver cancer; some HPV strains are linked to cervical, anal, and oropharyngeal cancer; and *Helicobacter pylori* is associated with gastric adenocarcinoma and gastric lymphoma. Vaccines to protect against these agents may therefore reduce the risk of their associated cancers.

The hepatitis B vaccine is effective in preventing hepatitis and hepatomas due to chronic hepatitis B infection.

A nonavalent HPV vaccine (covering strains 6, 11, 16, 18, 31, 33, 45, 52, and 58) is available for use in the United States. HPV types 6 and 11 cause genital papillomas. The remaining HPV types cause cervical and anal cancer; reduction in HPV types 16 and 18 alone could theoretically prevent >70% of cervical cancers worldwide. For individuals not previously infected with these HPV strains, the vaccine demonstrates high efficacy in preventing persistent strain-specific HPV infections. Studies also confirm the vaccine's ability to prevent preneoplastic lesions (cervical or anal intraepithelial neoplasia [CIN/AIN] I, II, and III). The vaccine does not appear to impact preexisting infections. A two-dose schedule is currently recommended in the United States for females and males age 9–14 years; teens and young adults who start the series between 15 and 26 years are recommended to receive three doses of the vaccine. However, observational studies suggest similar efficacy with a single dose in young girls, and a large randomized trial is currently comparing one to two doses.

SURGICAL PREVENTION OF CANCER

Some organs in some individuals are at such high risk of developing cancer that surgical removal of the organ at risk may be considered. Women with severe cervical dysplasia are treated with laser or loop electrosurgical excision or conization. Colectomy may be used to prevent colon cancer in patients with familial polyposis or ulcerative colitis.

Prophylactic bilateral mastectomy may be chosen for breast cancer prevention among women with high-risk genetic predisposition to breast cancer. In a prospective series of 139 women with *BRCA1* and *BRCA2* mutations, at 3 years, no cases of breast cancer had been diagnosed in those opting for surgery (n = 76), but eight of 63 patients in the surveillance group had developed breast cancer. A larger retrospective cohort study reported three patients developed breast cancer after prophylactic mastectomy compared with an expected incidence of 30–53 cases. Postmastectomy breast cancer–related deaths were 81–94% lower in high-risk women compared with sister controls and 100% lower in moderate-risk women when compared with expected rates.

Prophylactic salpingo-oophorectomy may also be employed for the prevention of ovarian and breast cancers among high-risk women. A prospective cohort study evaluating the outcomes of *BRCA* mutation carriers demonstrated a statistically significant association between prophylactic salpingo-oophorectomy and a reduced incidence of ovarian or primary peritoneal cancer (36% relative risk reduction, or a 4.5% absolute difference).

Studies of prophylactic oophorectomy for prevention of breast cancer in women with genetic mutations have shown relative risks of approximately 0.50; the risk reduction may be greatest for women having the procedure at younger (i.e., <45–50) ages. The observation that most high-grade serous "ovarian cancers" actually arise in the fallopian tube fimbria raises the possibility that this lethal subtype may be prevented by ovary-sparing salpingectomy. Formal testing of this hypothesis is important because an emulated target trial (i.e., study design principles of an randomized controlled trial applied to observational data to estimate causal effects) demonstrated associations between bilateral salpingo-oophorectomy during benign hysterectomy and an increased risk for cardiovascular disease when done in premenopausal women and a trend toward increased 10-year mortality in all ages (with statistical significance in perimenopausal women).

All of the evidence concerning the use of prophylactic mastectomy and salpingo-oophorectomy for prevention of breast and ovarian cancer in high-risk women has been observational; such studies are prone to a variety of biases, including case selection bias, family relationships between patients and controls, and inadequate information about hormone use. Thus, they may give an overestimate of the magnitude of benefit.

CANCER SCREENING

Screening is a means of early detection in asymptomatic individuals, with the goal of decreasing morbidity and mortality. While screening can potentially reduce disease-specific deaths and has been shown to do so in cervical, colon, lung, and breast cancer, it is also subject to several biases that can suggest a benefit when there is none. Biases can even mask net harm. Early detection does not in itself confer benefit. Cause-specific mortality, rather than survival after diagnosis, is the preferred endpoint (see below).

Because screening is done on asymptomatic, healthy persons, it should offer substantial likelihood of benefit that outweighs harm. Screening tests and their appropriate use should be carefully evaluated before their use is widely encouraged in screening programs.

The Accuracy of Screening A screening test's accuracy or ability to discriminate disease is described by four indices: sensitivity, specificity, positive predictive value, and negative predictive value **(Table 75-2)**. *Sensitivity*, also called the true-positive rate, is the proportion of persons with the disease who test positive in the screen (i.e., the ability of the test to detect disease when it is present). *Specificity*, or 1 minus the false-positive rate, is the proportion of persons who do not have the disease who test negative in the screening test (i.e., the ability of a test to correctly indicate that the disease is not present). The *positive predictive value* is the proportion of persons who test positive and who actually have the disease. Similarly, *negative predictive value* is the proportion testing negative who do not have the disease. The sensitivity and specificity of a test are relatively independent of the underlying prevalence (or risk) of the disease in the population screened, but the predictive values depend strongly on the prevalence of the disease.

Screening is most beneficial, efficient, and economical when the target disease is common in the population being screened. Specificity is at least as important to the ultimate feasibility and success of a screening test as sensitivity.

Potential Biases of Screening Tests Common biases of screening are lead time, length-biased sampling (and related overdiagnosis), and selection. These biases can make a screening test seem beneficial when it is not (or even causes net harm). Whether beneficial or not, screening can create the false impression of an epidemic by increasing the number of cancers diagnosed. It can also produce a shift in the *proportion* of patients diagnosed at an early stage (even without a reduction in absolute incidence of late-stage disease) and inflate survival statistics without reducing mortality (i.e., the number of deaths from a given cancer relative to the number of those at risk for the cancer in

TABLE 75-2 Assessment of the Value of a Diagnostic Test[a]

	CONDITION PRESENT	CONDITION ABSENT
Positive test	*a*	*b*
Negative test	*c*	*d*
a = true positive		
b = false positive		
c = false negative		
d = true negative		
Sensitivity	The proportion of persons with the condition who test positive: $a/(a+c)$	
Specificity	The proportion of persons without the condition who test negative: $d/(b+d)$	
Positive predictive value (PPV)	The proportion of persons with a positive test who have the condition: $a/(a+b)$	
Negative predictive value	The proportion of persons with a negative test who do not have the condition: $d/(c+d)$	
Prevalence, sensitivity, and specificity determine PPV $\text{PPV} = \frac{\text{prevalence} \times \text{sensitivity}}{(\text{prevalence} \times \text{sensitivity}) + (1 - \text{prevalence})(1 - \text{specificity})}$		

[a] For diseases of low prevalence, such as cancer, poor specificity has a dramatic adverse effect on PPV such that only a small fraction of positive tests are true positives.

the total population). In such a case, the *apparent* duration of survival (measured from date of diagnosis) increases without lives being saved or life expectancy changed.

Lead-time bias occurs whether or not a test influences the natural history of the disease; the patient is merely diagnosed at an earlier date. Survival *appears* increased even if life is not prolonged. The screening test only prolongs the time the subject is aware of the disease and spends as a cancer patient.

Length-biased sampling occurs because screening tests generally can more easily detect slow-growing, less aggressive cancers than fast-growing cancers. Cancers diagnosed due to the onset of symptoms between scheduled screenings are on average more aggressive, and treatment outcomes are not as favorable. An extreme form of length bias sampling is termed *overdiagnosis*, the detection of "pseudo disease." The reservoir of some slow-growing tumors is large. Many of these tumors fulfill the histologic criteria of cancer but will never become clinically significant or cause death during the patient's remaining life span. This problem is compounded by the fact that the most common cancers appear most frequently at ages when competing causes of death are more frequent.

Selection bias occurs because the population most likely to seek screening often differs from the general population to which the screening test might be applied. In general, volunteers for studies are more health conscious and likely to have a better prognosis or lower mortality rate irrespective of the screening result. This is termed the *healthy volunteer effect*.

Potential Drawbacks of Screening Risks associated with screening include harm caused by the screening intervention itself, harm due to the further investigation of persons with positive tests (both true and false positives), and harm from the treatment of persons with a true-positive result, whether or not life is extended by treatment (e.g., even if a screening test reduces relative cause-specific mortality by 15–30%, 70–85% of those diagnosed still go on to die of the target cancer). The diagnosis and treatment of cancers that would never have caused medical problems (i.e., overdiagnosis) can lead to the harm of unnecessary treatment (i.e., overtreatment) and give patients the anxiety of a cancer diagnosis. The psychosocial impact of cancer screening can be substantial when applied to the entire population.

Assessment of Screening Tests Good clinical trial design can offset some biases of screening and directly identify the balance of risks and benefits of a screening test. A randomized controlled screening trial with cause-specific mortality as the endpoint provides the strongest support for a screening intervention. Overall mortality should also be reported to detect an adverse effect of screening and treatment on other disease outcomes (e.g., cardiovascular disease, treatment-induced cancers). In a randomized trial, two like populations are randomly established. One is given the usual standard of care (which may be no screening at all) and the other receives the screening intervention being assessed. Efficacy for the population studied is established when the group receiving the screening test has a lower cause-specific mortality rate than the control group. Studies showing a reduction in the incidence of advanced-stage disease, improved survival, or a stage shift are weaker (and possibly misleading) evidence of benefit. These latter criteria are early indicators but not sufficient to definitively establish the value of a screening test.

Although a randomized, controlled screening trial provides the strongest evidence to support a screening test, it is not perfect. Unless the trial is population-based, it does not remove the question of generalizability to the target population. Screening trials generally involve thousands of persons and last for years. Less definitive study designs are therefore used to assess the effectiveness of a screening test of proven efficacy in actual practice. However, every nonrandomized study design is subject to strong confounders. In descending order of strength, evidence may also be derived from the findings of internally controlled trials using intervention allocation methods other than randomization (e.g., allocation by birth date, date of clinic visit); the findings of analytic observational studies; or the results of multiple time series studies with or without the intervention.

Screening for Specific Cancers Screening for cervical, colon, and breast cancer has the potential to be beneficial for certain age groups. Although these tests can reduce deaths from the cancer being screened for, meta-analyses of randomized trials have not shown that they affect all-cause mortality. Lung cancer screening can also reduce deaths in specific settings, depending on age and smoking history. Special surveillance of those at high risk for a specific cancer because of a family history or a genetic risk factor may be prudent, but few studies have assessed the effect on mortality. A number of organizations have considered whether or not to endorse routine use of certain screening tests. Because their perspectives have varied, they have arrived at different recommendations. The American Cancer Society (ACS) and the U.S. Preventive Services Task Force (USPSTF) publish screening guidelines **(Table 75-3)**; the American Academy of Family Practitioners (AAFP) often follows/endorses the USPSTF recommendations; and the American College of Physicians (ACP) develops recommendations based on structured reviews of other organizations' guidelines.

BREAST CANCER Breast self-examination, clinical breast examination by a caregiver, mammography, and magnetic resonance imaging (MRI) have all been variably advocated as useful screening tools.

A number of trials suggest that annual or biennial screening with mammography in normal-risk women older than age 50 years decreases breast cancer mortality. In most, breast cancer–related mortality rates were decreased by 15–30%. Experts disagree on whether average-risk women age 40–49 years should receive regular screening. The U.K. Age Trial, the only randomized trial of breast cancer screening to specifically evaluate the impact of mammography in women age 40–49 years, found no statistically significant difference in breast cancer mortality for screened women versus controls after about 22 years of follow-up (relative risk 0.88; 95% confidence interval, 0.74–1.03); however, <70% of women received screening in the intervention arm, potentially diluting the observed effect. A meta-analysis of nine large randomized trials showed an 8% relative reduction in breast cancer mortality (relative risk 0.92; 95% confidence interval, 0.75–1.02) from mammography screening for women age 39–49 years after 11–20 years of follow-up. This is equivalent to 3 breast cancer deaths prevented per 10,000 women over 10 years (although the result is not statistically significant). At the same time, nearly half of women age 40–49 years screened annually will have false-positive mammograms necessitating further evaluation, often including biopsy. Estimates of overdiagnosis range from 10 to 50% of diagnosed invasive cancers.

In the United States, widespread screening over the past several decades has not been accompanied by a reduction in incidence of metastatic breast cancer despite a large increase in early-stage disease, suggesting a substantial amount of overdiagnosis at the population level. In addition, the substantial improvements in systemic therapy have likely decreased the impact of mammography and early detection on falling breast cancer mortality rates.

Digital breast tomosynthesis is a newer method of breast cancer screening that reconstructs multiple x-ray images of the breast into superimposed "three-dimensional" slices. Although some evidence is available concerning the test characteristics of this modality, there are currently no data on its effects on health outcomes such as breast cancer–related morbidity, mortality, or overdiagnosis rates. A large randomized trial comparing standard digital mammography to tomosynthesis is in progress.

No study of breast self-examination has shown it to decrease mortality. A randomized controlled trial of approximately 266,000 women in China demonstrated no difference in breast cancer mortality between a group that received intensive breast self-exam instruction and reinforcement/reminders and controls at 10 years of follow-up. However, more benign breast lesions were discovered and more breast biopsies were performed in the self-examination arm.

Genetic screening for *BRCA1* and *BRCA2* mutations and other markers of breast cancer risk has identified a group of women at high risk for breast cancer. Unfortunately, when to begin and the optimal frequency of screening have not been defined. Mammography is less sensitive at detecting breast cancers in women carrying *BRCA1* and

TABLE 75-3 Screening Recommendations for Asymptomatic Subjects Not Known to Be at Increased Risk for the Target Condition[a]

CANCER TYPE	TEST OR PROCEDURE	USPSTF	ACS
Breast	Self-examination	"D"[b] (Not in current recommendations; from 2009)	Women, all ages: No specific recommendation
	Clinical examination	Women ≥40 years: "I" (as a stand-alone without mammography) (Not in current recommendations; from 2009)	Women, all ages: Do not recommend
	Mammography	Women aged 40-74 years: Biennial screening mammography ("B") Women ≥75 years: "I"	Women 40–44 years: Provide the opportunity to begin annual screening Women 45–54 years: Screen annually Women ≥55 years: Transition to biennial screening or have the opportunity to continue annual screening Women ≥40 should continue screening mammography as long as their overall health is good and they have a life expectancy of 10 years or longer
	Magnetic resonance imaging (MRI)	"I"	Women with >20% lifetime risk of breast cancer: Screen with MRI plus mammography annually Women with 15–20% lifetime risk of breast cancer: Discuss option of MRI plus mammography annually Women with <15% lifetime risk of breast cancer: Do not screen annually with MRI
Cervical	Pap test (cytology)	Women <21 years: "D" Women 21–29 years: Screen with cytology alone every 3 years ("A") Women 30–65 years: Screen with cytology alone every 3 years, or with co-testing (HPV testing + cytology) every 5 years (two of three options, see HPV test below) ("A") Women >65 years, with adequate, normal prior Pap screenings: "D" Women after total hysterectomy for noncancerous causes: "D"	Women <21 years: No screening Women 21–29 years: Screen every 3 years Women 30–65 years: Screen with co-testing (HPV testing + cytology) every 5 years or cytology alone every 3 years (see HPV test below) Women >65 years: No screening following adequate negative prior screening Women after total hysterectomy for noncancerous causes: Do not screen
	HPV test	Women <30 years: Do not use HPV testing for cervical cancer screening Women 30–65 years: Screen with HPV testing alone or in combination with cytology every 5 years (two of three options, see Pap test above) ("A") Women >65 years, with adequate, normal prior Pap screenings: "D" Women after total hysterectomy for noncancerous causes: "D"	Women <30 years: Do not use HPV testing for cervical cancer screening Women 30–65 years: Preferred approach to screen with HPV and cytology co-testing every 5 years (see Pap test above) Women >65 years: No screening following adequate negative prior screening Women after total hysterectomy for noncancerous causes: Do not screen
Colorectal	Overall	Adults 50–75 years: "A" Screen for colorectal cancer Adults 45–50 years: "B" Screen for colorectal cancer Adults 76–85 years: "C" Selectively offer screening for colorectal cancer; consider the patient's overall health, prior screening history, and preferences	Adults ≥45–75 years: Screen for colorectal cancer with either a high-sensitivity stool-based test or a structural (visual) examination (≥45 years, qualified recommendation; ≥50 years, strong recommendation). Adults 76–85 years: Individualize screening based on patient preferences, life expectancy, health status, and prior screening history (qualified recommendation). Adults >85 years: Discourage screening (qualified recommendation). Every 5 years
	Sigmoidoscopy	Every 5 years; modeling suggests improved benefit if performed every 10 years in combination with annual FIT	Adults ≥45 years: Every 5 years
	Fecal occult blood testing (FOBT)	Every year	Adults ≥45 years: Every year
	Colonoscopy	Every 10 years	Adults ≥45 years: Every 10 years
	Fecal DNA testing	At least every 3 years	Adults ≥45 years: Every 3 years
	Fecal immunochemical testing (FIT)	Every year	Adults ≥45 years: Every year
	Computed tomography (CT) colonography	Every 5 years	Adults ≥45 years: Every 5 years
Lung	Low-dose CT scan	Adults 50–80 years, with a ≥20 pack-year smoking history, still smoking or have quit within past 15 years, annually: "B" Discontinue once a person has not smoked for 15 years or develops a health problem that substantially limits life expectancy or the ability to have curative lung surgery	Men and women, 50–80 years, with ≥20 pack-year smoking history, current or former smoker, annually: Discuss benefits, limitations, and harms of screening; offer smoking cessation counseling and connection to resources where relevant. Individuals with comorbid conditions that substantially limit life expectancy should not be screened.

(*Continued*)

TABLE 75-3 Screening Recommendations for Asymptomatic Subjects Not Known to Be at Increased Risk for the Target Condition[a] (Continued)

CANCER TYPE	TEST OR PROCEDURE	USPSTF	ACS
Ovarian	CA-125 Transvaginal ultrasound	Women, all ages: "D" Women with a high-risk hereditary cancer syndrome: No recommendation	Currently, there are no reliable screening tests for the early detection of ovarian cancer. For women at high risk of ovarian cancer, it has not been proven that using transvaginal ultrasound or serum CA-125 lowers their chances of dying from ovarian cancer.
Prostate	Prostate-specific antigen (PSA)	Men 55–69 years: The decision to undergo periodic PSA-based screening should be an individual one. Men should have an opportunity to discuss the potential benefits and harms of screening with their clinician. Clinicians should not screen men who do not express a preference for screening ("C") Men ≥70 years: "D"	Starting at age 50, men at average risk and with a life expectancy of ≥10 years should talk to a doctor about the uncertainties, risks, and potential benefits of screening. If African American or have a father or brother who had prostate cancer before age 65, men should have this talk starting at age 45. For men with more than one first-degree relative with prostate cancer diagnosed before age 65, have this talk starting at age 40. How often they are screened will depend on their PSA level.
	Digital rectal examination (DRE)	No individual recommendation	As for PSA; if men decide to be tested, they should have the PSA blood test with or without a rectal exam.
Skin	Complete skin examination by clinician or patient	Adults, all ages: "I"	No guidelines

[a] Summary of the screening procedures recommended for the general population by the USPSTF and the ACS. These recommendations refer to asymptomatic persons who are not known to have risk factors, other than age or gender, for the targeted condition. [b] USPSTF lettered recommendations are defined as follows: "A": The USPSTF recommends the service because there is high certainty that the net benefit is substantial; "B": The USPSTF recommends the service because there is high certainty that the net benefit is moderate or moderate certainty that the net benefit is moderate to substantial; "C": The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences; there is at least moderate certainty that the net benefit is small; "D": The USPSTF recommends against the service because there is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits; "I": The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Table only includes final, not draft, recommendations, from the USPSTF.

Abbreviations: ACS, American Cancer Society; USPSTF, U.S. Preventive Services Task Force.

BRCA2 mutations, possibly because such cancers occur in younger women, in whom mammography is known to be less sensitive. MRI screening may be more sensitive than mammography in women at high risk due to genetic predisposition or in women with very dense breast tissue, but specificity may be lower. An increase in overdiagnosis may accompany the higher sensitivity. The impact of MRI on breast cancer mortality with or without concomitant use of mammography has not been evaluated in a randomized controlled trial.

CERVICAL CANCER The cervical cancer mortality rate has fallen substantially since the widespread use of the Pap smear. Screening guidelines recommend regular Pap testing for all women who have reached the age of 21 (before this age, even in individuals that have begun sexual activity, screening may cause more harm than benefit). The recommended interval for Pap screening is 3 years. In all cases, screening more frequently adds little benefit but leads to important harms, including unnecessary procedures and overtreatment of transient lesions. With the onset of sexual activity comes the risk of sexual transmission of HPV, the fundamental etiologic factor for cervical cancer. Beginning at age 30, guidelines also include HPV testing with or without Pap smear. The screening interval for women who test normal using this approach may be lengthened to 5 years.

An upper age limit at which screening ceases to be effective is not known, but women age 65 years with no abnormal results in the previous 10 years may choose to stop screening. Screening should be discontinued in women who have undergone a hysterectomy with cervical excision for noncancerous reasons.

Although the efficacy of the Pap smear in reducing cervical cancer mortality has never been directly confirmed in a randomized, controlled setting, a clustered randomized trial in India evaluated the impact of one-time cervical visual inspection and immediate colposcopy, biopsy, and/or cryotherapy (where indicated) versus counseling on cervical cancer deaths in women age 30–59 years. After 7 years of follow-up, the age-standardized rate of death due to cervical cancer was 39.6 per 100,000 person-years in the intervention group versus 56.7 per 100,000 person-years in controls.

COLORECTAL CANCER Digital rectal examination (DRE), stool-based testing (fecal occult blood testing [FOBT], fecal immunochemical tests [FITs], multitargeted stool DNA tests), blood-based testing (*SEPT9*), and optical testing (rigid and flexible sigmoidoscopy, colonoscopy, computed tomography [CT] colonography) have been considered for colorectal cancer screening.

DRE is not an effective colorectal cancer screening test and is not recommended by any medical organization.

Stool-Based Testing A meta-analysis of five randomized controlled trials demonstrated a 22% relative reduction in colorectal cancer mortality after two to nine rounds of biennial FOBT at 30 years of follow-up; annual screening was shown to result in a greater mortality reduction in a single trial (a 32% relative reduction). However, only 2–10% of those with occult blood in the stool actually have cancer. The high false-positive rate of FOBT therefore leads to a large number of diagnostic follow-up colonoscopies. FITs have higher sensitivity for colorectal cancer than FOBT tests and have largely replaced FOBT in clinical practice. Limited evidence suggests FITs may not detect proximal colonic tumors at the same rate as distal ones. Multitargeted stool DNA testing (e.g., Cologuard) combines FIT with testing for altered DNA biomarkers that are shed into the stool. Although limited evidence demonstrates that it can have a higher single-test sensitivity for colorectal cancer than FIT alone, its specificity is lower, resulting in a higher number of false-positive tests and follow-up colonoscopies. No studies have yet evaluated its effects on colorectal cancer incidence, morbidity, or mortality.

Blood-Based Testing A blood test for the methylated *SEPT9* gene associated with colorectal cancer is available. Case-control studies suggest that it has a lower sensitivity and specificity than the FIT stool test. It also is more costly, and it is not recommended as a first-line screening test.

Optical Testing Two meta-analyses of five randomized controlled trials of sigmoidoscopy found an 18% relative reduction in colorectal cancer incidence and a 28% relative reduction in colorectal cancer mortality. Participant ages ranged from 50 to 74 years, with follow-up ranging from 6 to 13 years. Diagnosis of adenomatous polyps by sigmoidoscopy should lead to evaluation of the entire colon with colonoscopy. The most efficient interval for screening sigmoidoscopy is unknown, but an interval of 5 years is often recommended. Case-control studies suggest that intervals of up to 15 years may confer benefit; a randomized trial in the U.K. demonstrated colorectal cancer mortality reduction with one-time screening.

One-time colonoscopy detects ~25% more advanced lesions (polyps >10 mm, villous adenomas, adenomatous polyps with high-grade dysplasia, invasive cancer) than one-time FOBT with sigmoidoscopy; comparative *programmatic* performance of the two modalities over time is not known. Perforation rates are about 4/10,000 for colonoscopy and 1/10,000 for sigmoidoscopy. Debate continues on whether colonoscopy is too expensive and invasive and whether sufficient provider capacity exists to be recommended as the *preferred* screening tool in standard-risk populations. Some observational studies suggest that efficacy of colonoscopy to decrease colorectal cancer mortality is higher on the left side of the colon than the right.

CT colonography, if done at expert centers, appears to have a sensitivity for polyps ≥6 mm, comparable to colonoscopy. However, the rate of extracolonic findings of abnormalities of uncertain significance that must nevertheless be worked up is high (~5–37%); the long-term cumulative radiation risk of repeated colonography screenings is also a concern.

LUNG CANCER Chest x-ray and sputum cytology have been evaluated in several randomized lung cancer screening trials. The most recent and largest (n = 154,901) of these, a component of the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial, found that, compared with usual care, annual chest x-ray did not reduce the risk of dying from lung cancer (relative risk 0.99; 95% confidence interval, 0.87–1.22) after 13 years. However, it showed evidence of overdiagnosis associated with chest x-ray. Low-dose CT has also been evaluated in several randomized trials. The largest and longest of these, the National Lung Screening Trial (NLST), was a randomized controlled trial of screening for lung cancer in ~53,000 persons age 55–74 years with a 30+ pack-year smoking history. It demonstrated a statistically significant reduction of about 3 fewer deaths per 1000 people screened with CT compared to chest x-ray after 12 years. Harms include the potential radiation risks associated with multiple scans, the discovery of incidental findings of unclear significance, and a high rate of false-positive test results. Both incidental findings and false-positive tests can lead to invasive diagnostic procedures associated with anxiety, complications, and expense. The NLST was performed at experienced screening centers, and the balance of benefits and harms may differ in the community setting at less experienced centers.

OVARIAN CANCER Adnexal palpation, transvaginal ultrasound (TVUS), and serum CA-125 assay have been considered for ovarian cancer screening. A large randomized, controlled trial has shown that an annual screening program of TVUS and CA-125 in average-risk women does not reduce deaths from ovarian cancer (relative risk 1.21; 95% confidence interval, 0.99–1.48). Adnexal palpation was dropped early in the study because it did not detect any ovarian cancers that were not detected by either TVUS or CA-125. A second large, randomized trial used a two-stage screening approach incorporating a risk of ovarian cancer algorithm that determined whether additional testing with CA-125 or TVUS was required. At 14 years of follow-up, there was no statistically significant reduction in ovarian cancer deaths. The risks and costs associated with the high number of false-positive results are impediments to routine use of these modalities for screening. In the PLCO trial, 10% of participants had a false-positive result from TVUS or CA-125, and one-third of these women underwent a major surgical procedure; the ratio of surgeries to screen-detected ovarian cancer was approximately 20:1. In September 2016, the FDA issued a safety communication recommending against using any ovarian cancer screening test, including the algorithm for risk of the cancer.

PROSTATE CANCER The most common prostate cancer screening modalities are digital rectal exam (DRE) and serum PSA assay. An emphasis on PSA screening has made prostate cancer the most common nonskin cancer diagnosed in American males. This disease is prone to lead-time bias, length bias, and overdiagnosis, and substantial debate continues among experts as to whether screening should be offered unless the patient specifically asks to be screened. Virtually all organizations stress the importance of informing men about the uncertainty regarding screening efficacy and the associated harms. Prostate cancer screening clearly detects many asymptomatic cancers, but men older than age 50 years have a high prevalence of clinically insignificant prostate cancers (about 30–50% of men, increasing further as men age). The ability to distinguish tumors that are lethal but still curable from those that pose little or no threat to health is limited, although evidence suggests multiparametric MRI may aid decision-making.

Randomized trials indicate that the effect of PSA screening on prostate cancer mortality across a population is, at best, small. Two major trials have been published. The PLCO Cancer Screening Trial was a multicenter U.S. trial that randomized almost 77,000 men age 55–74 years to receive either annual PSA testing for 6 years or usual care. At 13 years of follow-up, no statistically significant difference in the number of prostate cancer deaths was noted between the arms (rate ratio 1.09; 95% confidence interval, 0.87–1.36). More than half of men in the control arm received at least one PSA test during the trial, which may have diluted a small effect.

The European Randomized Study of Screening for Prostate Cancer (ERSPC) was a multinational study that randomized ~182,000 men between age 50 and 74 years (with a predefined "core" screening group of men age 55–69 years) to receive PSA testing or no screening. Recruitment and randomization procedures, as well as actual frequency of PSA testing, varied by country. After a median follow-up of 15.5 years, a 20% relative reduction in the risk of prostate cancer death in the screened arm was noted in the "core" screening group. The trial found that 570 men (95% confidence interval, 380–1137 men) would need to be invited to screening, and 18 cases of prostate cancer detected, to avert 1 death from prostate cancer. There was an unexplained imbalance in treatment between the two study arms, with a higher proportion of men with clinically localized cancer receiving radical prostatectomy in the screening arm and receiving it at experienced referral centers.

Screening must be linked to effective therapy to have any benefit. Two trials conducted after the initiation of widespread PSA testing did not find a substantial decrease in prostate cancer deaths in control arms of "watchful waiting" or monitoring (i.e., no curative treatment) compared to radical prostatectomy or radiation therapy. Prostate cancer–specific survival was very good (about 99%) and nearly identical at a median follow-up of 10 years. Treatments for low-stage prostate cancer, such as surgery and radiation therapy, can cause substantial morbidity, including impotence and urinary incontinence.

SKIN CANCER Visual examination of all skin surfaces by the patient or by a health care provider is used in screening for basal and squamous cell cancers and melanoma. No prospective randomized study has been performed to look for a mortality decrease. Unfortunately, screening is associated with a substantial rate of overdiagnosis.

MULTICANCER EARLY DETECTION TESTS Multicancer early detection (MCED) tests look for multiple biomarkers in blood that are either directly released or induced by cancer cells. One potential purpose of these tests could be to screen for multiple cancers at the same time in asymptomatic people. However, most of the evidence at present for these tests comes from people who have already been diagnosed with cancer (a fundamentally different population); additionally, no data are available on health outcomes associated with using MCED tests for cancer screening.

FURTHER READING

Fenton JJ et al: Prostate-specific antigen-based screening for prostate cancer: Evidence report and systematic review for the U.S. Preventive Services Task Force. JAMA 319:1914, 2018.

Jonas DE et al: Screening for lung cancer with low-dose computed tomography: Updated Evidence report and systematic review for the US Preventive Services Task Force. JAMA 325:971, 2021.

Kramer BS, Croswell JM: Cancer screening: The clash of science and intuition. Annu Rev Med 60:125, 2009.

Lin JS et al: Screening for colorectal cancer: Updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 325:1978, 2021.

Manson JE et al: Vitamin D supplements and prevention of cancer and cardiovascular disease. N Engl J Med 380:33, 2019.

McNeil JJ et al: Effect of aspirin on all-cause mortality in the healthy elderly. N Engl J Med 379:1519, 2018.

Melnikow J et al: Screening for cervical cancer with high-risk human papillomavirus testing: Updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 320:687, 2018.

Welch HG et al: Epidemiologic signatures in cancer. N Engl J Med 384:14, 2019.

Woloshin S et al: Breast-cancer mortality trends in four countries with varied screening practices. N Engl J Med 389:1061, 2023.

Zeraatkar D et al: Effect of lower versus high red meat intake on cardiometabolic and cancer outcomes: A systematic review of randomized trials. Ann Intern Med 171:721, 2019.

76

Cancer Genetics

Fred Bunz, Bert Vogelstein

CANCER IS A GENETIC DISEASE

Cancer arises through a series of somatic alterations in DNA that result in unrestrained cellular proliferation. Most of these alterations involve subtle sequence changes in DNA (i.e., mutations). The somatic mutations may originate as a consequence of random replication errors or exposure to carcinogens (e.g., radiation) and can be exacerbated by faulty DNA repair processes. While most cancers arise sporadically, clustering of cancers occurs in families that carry a germline mutation in a cancer gene.

HISTORICAL PERSPECTIVE

The idea that cancer progression is driven by sequential somatic mutations in specific genes has only gained general acceptance in the past 30 years. Before the advent of the microscope, cancer was believed to be composed of aggregates of mucus or other noncellular matter. By the middle of the nineteenth century, it became clear that tumors were masses of cells and that these cells arose from the normal cells of the tissue from which the cancer originated. The molecular basis for the uncontrolled proliferation of cancer cells was to remain a mystery for another century. During that time, a number of theories for the origin of cancer were postulated. The great biochemist Otto Warburg proposed the combustion theory of cancer, which stipulated that cancer was due to abnormal oxygen metabolism. Others believed that all cancers were caused by viruses and that cancer was in fact a contagious disease.

In the end, observations of cancer occurring in chimney sweeps, studies of x-rays, and the overwhelming data demonstrating cigarette smoke as a causative agent in lung cancer, together with Ames's work on chemical mutagenesis, were consistent with the idea that cancer originated through changes in DNA. However, it was not until the somatic mutations responsible for cancer were identified at the molecular level that the genetic basis of cancer was definitively established. Although the viral theory of cancer did not prove to be generally accurate (with exceptions such as human papillomaviruses, which can cause cervical and other cancers), the study of retroviruses led to the discovery of the first human *oncogenes* in the late 1970s. Oncogenes are one of the two major classes of cancer driver genes. The study of families with genetic predisposition to cancer was instrumental to the discovery of the other major class of cancer driver genes, called *tumor-suppressor genes*. Current technologies permit the sequence analysis of entire cancer genomes and provide a comprehensive view of the genetic changes that cause tumors to arise and become malignant. The field that studies the various types of mutations, as well as the consequences of these mutations in tumor cells, is now known as *cancer genetics*.

THE CLONAL ORIGIN AND MULTISTEP NATURE OF CANCER

Nearly all cancers originate from a single cell; this clonal origin is a critical discriminating feature between neoplasia and hyperplasia. Multiple cumulative mutational events are invariably required for the progression of a tumor from normal to fully malignant phenotype. The process can be seen as Darwinian microevolution in which, at each successive step, the mutated cells gain a growth advantage resulting in the expansion of a neoplastic clone **(Fig. 76-1)**. Based on observations of cancer frequency increases during aging, the epidemiologists Armitage and Doll and Nordling independently proposed that cancer is a result of three discrete cellular changes. Remarkably, this early model has been validated by extensive sequencing of cancer genomes. These studies revealed that just three causal mutations are required for the development of several of the most common cancers. Overall, it is currently believed that most common solid tumors require a minimum of three mutated cancer driver genes (either oncogenes or tumor-suppressor genes) for their development. One or two mutations are sufficient for benign tumorigenesis, but not for the invasive capacity that distinguishes cancers from benign tumors. Less common tumors, such as liquid tumors (leukemias or lymphomas), sarcomas, and childhood tumors, appear to require only two driver gene alterations for malignancy. Note that a cancer driver gene is best defined as one containing a mutation that increases the selective growth advantage of the cell containing it. Normally, cell birth and cell death are in perfect equilibrium; every time a cell is born, another in the same lineage dies. Cancer driver gene mutations alter this equilibrium, so that more cells are born than die. The imbalance is often slight, so that the difference between cell birth and cell death can be less than 1%. This explains, in combination with the low rate of mutation, why tumorigenesis—the journey from a normal cell to a typical malignant, solid tumor—often takes decades.

We now know the precise nature of the genetic alterations responsible for nearly all malignancies and are beginning to understand how these alterations promote the distinct stages of tumor growth. The prototypical example is colon cancer, in which analyses of genomes from the entire spectrum of neoplastic growths—from normal colon

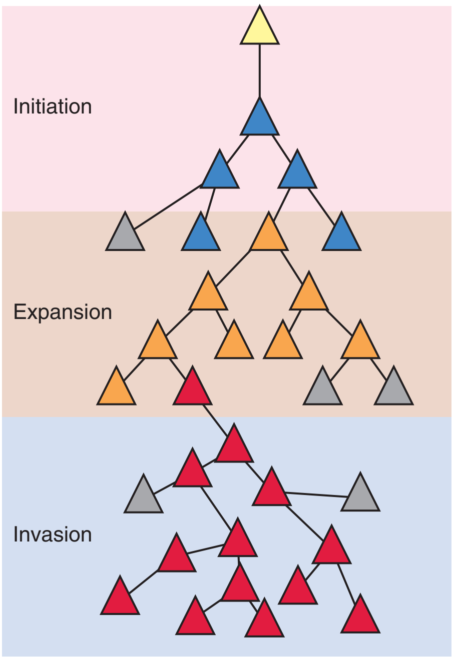

FIGURE 76-1 Multistep clonal development of malignancy. In this diagram, a series of three cumulative mutations, each with a modest growth advantage acting alone, eventually results in a malignant tumor. Note that not all such alterations result in progression. The actual number of cumulative mutations necessary to transform from the normal to the malignant state has been estimated to be three for several of the most common types of cancer. *(Adapted and modified from PC Nowell: The clonal evolution of tumor cell populations. Science 194:23, 1976.)*

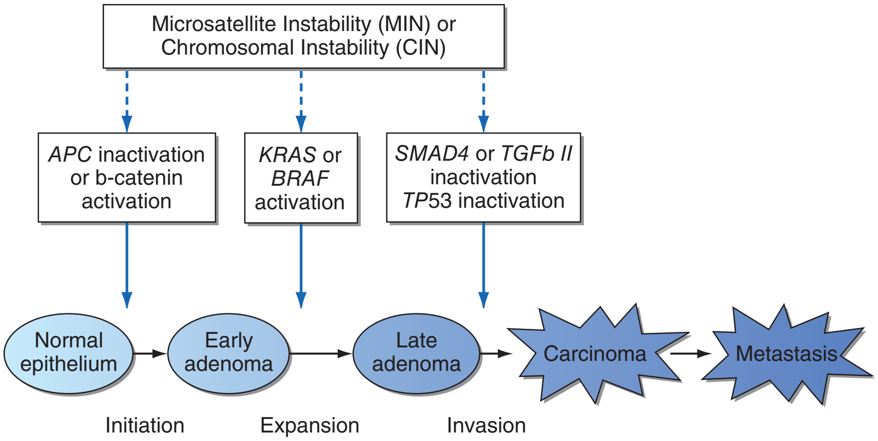
Microsatellite Instability (MIN) or
Chromosomal Instability (CIN)
APC inactivation or b-catenin activation
KRAS or *BRAF* activation
SMAD4 or *TGFb II* inactivation
TP53 inactivation
Normal epithelium
Early adenoma
Late adenoma
Carcinoma
Metastasis
Initiation
Expansion
Invasion

FIGURE 76-2 Progressive somatic mutational steps in the development of colon carcinoma. The accumulation of alterations in a number of different genes results in the progression from normal epithelium through adenoma to full-blown carcinoma. Genetic instability (microsatellite or chromosomal) accelerates the progression by increasing the likelihood of mutation at each step. Patients with familial polyposis are already one step into this pathway because they inherit a germline alteration of the *APC* gene. TGF, transforming growth factor.

epithelium through adenoma to carcinoma—have identified mutations that are highly characteristic of each type of lesion **(Fig. 76-2)**.

TWO TYPES OF CANCER GENES: ONCOGENES AND TUMOR-SUPPRESSOR GENES

Oncogenes and tumor-suppressor genes exert their effects on tumor growth through their ability to determine cell fates, influence cell survival, and contribute to genome maintenance. The underlying molecular mechanisms can be extremely complex. While tightly regulated in normal cells, oncogenes acquire mutations that typically relieve this control and lead to increased activity of the gene products. This activating mutational event occurs in a single allele. In contrast, the normal function of tumor-suppressor genes is usually to restrain cell growth, and this function is lost in cancer. Because of the diploid nature of mammalian cells, both alleles must be inactivated for a cell to completely lose the function of a tumor-suppressor gene. Thus, two genetic events are required to inactivate a tumor-suppressor gene, while only one genetic event is required to activate an oncogene.

A subset of tumor-suppressor genes controls the ability of the cell to maintain the integrity of its genome. Cells with a deficiency in these genes acquire an increased number of mutations throughout their genomes, including those in oncogenes and tumor-suppressor genes. This "mutator" phenotype was first hypothesized by Loeb to explain how the multiple rare mutational events required for tumorigenesis can occur in the lifetime of an individual. A mutator phenotype underlies several forms of cancer, such as those associated with deficiencies in DNA mismatch repair. The great majority of cancers do not harbor repair deficiencies, and their rate of mutation is similar to that observed in normal cells. Many of these cancers, however, appear to harbor a different kind of genetic instability, affecting the loss or gains of whole chromosomes or large parts thereof (as explained in more detail below).

ONCOGENES IN HUMAN CANCER

Work by Peyton Rous in the early 1900s revealed that a chicken sarcoma could be transmitted from animal to animal in cell-free extracts, suggesting that cancer could be induced by an agent acting positively to promote tumor formation. The agent responsible for the transmission of the cancer was a retrovirus (Rous sarcoma virus [RSV]), and the oncogene responsible was identified 75 years later as *V-SRC*. Other oncogenes were also discovered through their presence in the genomes of retroviruses that are capable of causing cancers in chickens, mice, and rats. The nonmutated cellular homologues of these viral genes are called proto-oncogenes and are often targets of mutation or aberrant regulation in human cancer. Whereas many oncogenes were discovered on the basis of their presence in retroviruses, other oncogenes, particularly those involved in translocations characteristic of particular leukemias and lymphomas, were identified through genomic approaches. Investigators cloned the sequences surrounding the chromosomal translocations observed cytogenetically and identified the genes activated at the breakpoints (see below). Some of these were oncogenes previously found in retroviruses (like *ABL*, involved in chronic myeloid leukemia [CML]), whereas others were new (like *BCL2*, involved in B-cell lymphoma). In the normal cellular environment, proto-oncogenes have crucial roles in cell proliferation and differentiation. **Table 76-1** is a partial list of oncogenes known to be involved in human cancer.

The normal growth and differentiation of cells is controlled by growth factors that bind to receptors on the surface of the cell. The signals generated by the membrane receptors are transmitted inside the cells through signaling cascades involving kinases, G proteins, and other regulatory proteins. Ultimately, these signals affect the activity of transcription factors in the nucleus, which regulate the expression of genes crucial in cell proliferation, cell differentiation, and cell death. Oncogene products function at critical steps in these signaling pathways **(Chap. 77)**. Inappropriate activation of these pathways can lead to tumorigenesis.

MECHANISMS OF ONCOGENE ACTIVATION

■ POINT MUTATION

Point mutation (alternatively known as single nucleotide substitution) is a common mechanism of oncogene activation. For example, point mutations in *KRAS* are present in >95% of pancreatic cancers and 40% of colon cancers. Activating *KRAS* mutations are less common in other cancer types, although they can occur at significant frequencies in leukemia, lung, and thyroid cancers. Remarkably—and in contrast

TABLE 76-1 Oncogenes Commonly Altered in Human Cancers

ONCOGENE	FUNCTION	ALTERATION IN CANCER	NEOPLASM
AKT1	Serine/threonine kinase	Point mutation	Skin
BRAF	Serine/threonine kinase	Point mutation	Melanoma, thyroid, colorectal
CCND1	Cell cycle progression	Amplification	Esophageal, head and neck
CTNNB1	Signal transduction	Point mutation	Colon, liver, uterine, melanoma
EGFR	Signal transduction	Point mutation	Lung
FLT3	Signal transduction	Point mutation	AML
IDH1	Chromatin modification	Point mutation	Glioma
MDM2	Inhibitor of p53	Amplification	Sarcoma, glioma
MDM4	Inhibitor of p53	Amplification	Breast
MYC	Transcription factor	Amplification	Prostate, ovarian, breast, liver, pancreatic
MYCL1	Transcription factor	Amplification	Ovarian, bladder
MYCN	Transcription factor	Amplification	Neuroblastoma
PIK3CA	Phosphoinositol-3-kinase	Point mutation	Multiple cancers
KRAS	GTPase	Point mutation	Pancreatic, colorectal, lung
NRAS	GTPase	Point mutation	Melanoma

Abbreviation: AML, acute myeloid leukemia.

to the diversity of mutations found in tumor-suppressor genes—most of the activated *KRAS* alleles contain point mutations in codons 12, 13, or 61. These mutations lead to constitutive activation of the mutant RAS protein. The restricted pattern of mutations observed in oncogenes compared to that of tumor-suppressor genes reflects the fact that gain-of-function mutations must occur at specific sites, while a broad variety of mutations can lead to loss of activity. Indeed, inactivation of a gene can in theory be accomplished through the introduction of a stop codon anywhere in the coding sequence, whereas activations require precise substitutions at residues that can somehow lead to an increase in the activity of the encoded protein under particular circumstances within the cell.

■ DNA AMPLIFICATION

The second mechanism for activation of oncogenes is DNA sequence amplification, leading to overexpression of the gene product. This increase in DNA copy number may cause cytologically recognizable chromosome alterations referred to as *homogeneous staining regions* (HSRs) if integrated within chromosomes, or *double minutes* (dmins) if extrachromosomal.

Numerous genes have been reported to be amplified in cancer. Several of these genes, including *NMYC* and *LMYC*, were identified through their presence within the amplified DNA sequences of a tumor and their homology to known oncogenes. Because amplified regions often include hundreds of thousands of base pairs, multiple oncogenes may be amplified in a single amplicon in some cancers. For example, *MDM2*, *GLI1*, *CDK4*, and *TPSPAN31* at chromosomal location 12q13-15 have been shown to be co-amplified in several types of sarcomas and other tumors; which of these genes play the causal role in the neoplastic process is still an active area of research. Amplification of a cellular gene is often a predictor of poor prognosis; for example, *ERBB2/HER2* and *NMYC* are often amplified in aggressive breast cancers and neuroblastoma, respectively.

■ CHROMOSOMAL REARRANGEMENT

Chromosomal alterations provide important clues to the genetic changes in cancer. The chromosomal alterations in human solid tumors such as carcinomas are heterogeneous and complex and occur as a result of the frequent chromosomal instability observed in these tumors (see below). In contrast, the chromosome alterations in myeloid and lymphoid tumors are often simple translocations, that is, reciprocal transfers of chromosome arms from one chromosome to another. The breakpoints of recurring chromosome abnormalities usually occur at the site of cellular oncogenes. **Table 76-2** lists representative examples of recurring chromosome alterations in malignancy and the associated gene(s) rearranged or deregulated by the chromosomal rearrangement. Translocations are often observed in liquid tumors in general and are particularly common in lymphoid tumors, probably because these cell types have the capability to rearrange their DNA to generate antigen receptors. Indeed, antigen receptor genes are commonly involved in the translocations, implying that an imperfect regulation of receptor gene rearrangement may be involved in their pathogenesis. In addition to transcription factors and signal transduction molecules, translocation may result in the overexpression of cell cycle regulatory proteins or proteins such as cyclins and of proteins that regulate cell death. Recurrent translocations have more recently been identified in solid tumors such as prostate cancers. For example, fusions between *TMPRSS2* and *ERG*, which are normally located in tandem on chromosome 21, contribute to more than one-third of all prostate cancers.

TABLE 76-2 Representative Oncogenes at Chromosomal Translocations

GENE (CHROMOSOME)	TRANSLOCATION	MALIGNANCY
BCR-ABL	(9;22)(q34;q11)	Chronic myeloid leukemia
BCL1 (11q13.3)–*IgH* (14q32)	(11;14)(q13;q32)	Mantle cell lymphoma
BCL2 (18q21.3)–*IgH* (14q32)	(14;18)(q32;q21)	Follicular lymphoma
FLI-EWSR1	(11;22)(q24;q12)	Ewing's sarcoma
LCK-TCRB	(1;7)(p34;q35)	T-cell acute lymphocytic leukemia
PAX3-FOXO1	(2;13)(q35;q14)	Rhabdomyosarcoma
PAX8-PPARG	(2;3)(q13;p25)	Thyroid
IL21R-BCL6	(3;16)(q27;p11)	Non-Hodgkin's lymphoma
TAL1-TCTA	(1;3)(p34;p21)	Acute T-cell leukemia
TMPRSS2-ERG	Rearrangement on Chr21q22	Prostate

The first reproducible chromosome abnormality detected in human malignancy was the Philadelphia chromosome detected in CML. This cytogenetic abnormality is generated by reciprocal translocation involving the *ABL* oncogene on chromosome 9, encoding a tyrosine kinase, being placed in proximity to the breakpoint cluster region (*BCR*) gene on chromosome 22. **Figure 76-3** illustrates the generation of the translocation and its protein product. The consequence of expression of the *BCR-ABL* gene product is the activation of signal transduction pathways leading to cell growth independent of normal external signals. Imatinib, a drug that specifically blocks the activity of Abl tyrosine kinase, has shown remarkable efficacy with little toxicity in patients with CML. The successful targeting of *BCR-ABL* by imatinib is the paradigm for molecularly targeted anticancer therapies.

CHROMOSOMAL INSTABILITY IN SOLID TUMORS

Solid tumors generally contain an abnormal number of chromosomes, a state known as aneuploidy. Chromosomes from aneuploid tumors also exhibit structural alterations such as translocations, deletions, and amplifications. These abnormalities reflect an underlying defect in cancer cells known as chromosomal instability. While aneuploidy is a striking cellular phenotype, chromosomal instability is manifest as only a small increase in the tendency of cells to gain, lose, or rearrange chromosomes during any given cell cycle. This intrinsically low rate of chromosome aberration implies that cancer cells become aneuploid only after many generations of clonal expansion. The molecular basis of aneuploidy remains incompletely understood. It is widely believed that defects in checkpoints, the quality-control mechanisms that halt the cell cycle if chromosomes are damaged or misaligned, contribute to chromosomal instability. This hypothesis emerged from experimental observations that the tumor suppressor p53 controls checkpoints that regulate the initiation of DNA replication and the onset of mitosis. These processes are therefore defective in many cancer cells. The mitotic spindle checkpoint, which ensures proper chromosome attachment to the mitotic spindle before allowing the sister chromatids to separate, is also altered in some cancers, irrespective of p53 status. The precise relationship between checkpoint deficiency, p53, and chromosomal instability remains unclear, but it is believed that even a subtle perturbation of the highly orchestrated process of cell division can impact the ability of a cell to faithfully replicate and segregate its complement of chromosomes. From a therapeutic standpoint, the checkpoint defects that are prevalent in cancers have been proposed as vulnerabilities that may be exploited by novel agents and combinatorial strategies.

In contrast to the genome-wide cytogenetic changes that are typical indications of an underlying chromosomal instability, more focal patterns of chromosomal rearrangement have been recurrently detected in many cancer types. A curious phenomenon known as *chromothripsis* causes dozens of distinct breakpoints that are localized on one or several chromosomes. These striking structural alterations are thought to reflect a single event in which a chromosome is fragmented and then imprecisely reassembled. In some cancer types, chromothripsis contributes to oncogene amplification and tumor suppressor gene inactivation in a substantial proportion of tumors. While the exact process that underlies chromothripsis remains obscure, a transient period of extreme instability stands in contrast to the gradual loss, gain, and rearrangement of chromosomes that are typically observed in serially cultured cancer cells.

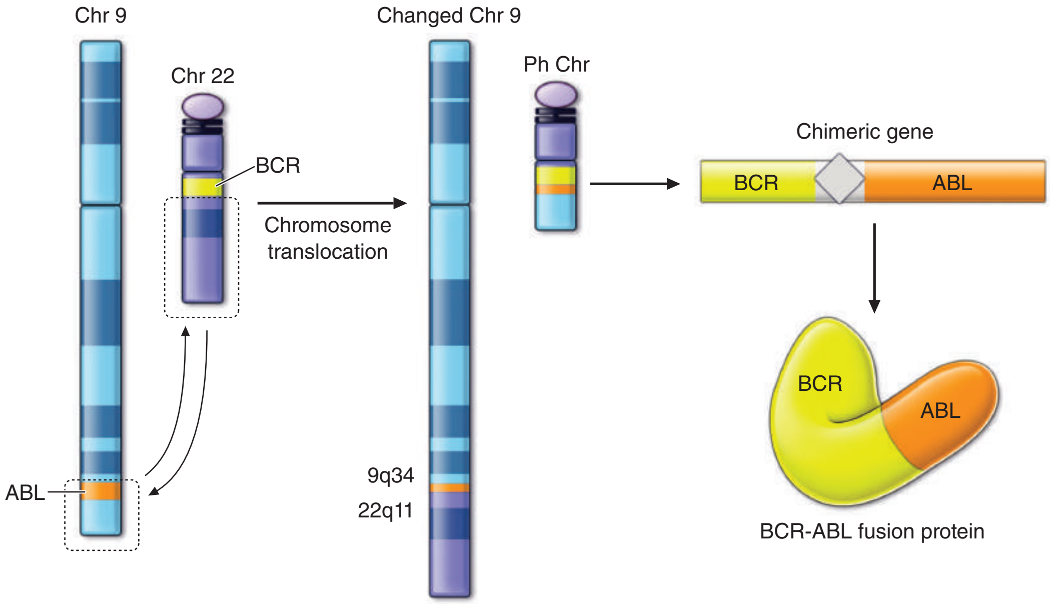

FIGURE 76-3 Specific translocation seen in chronic myeloid leukemia (CML). The Philadelphia chromosome (Ph) is derived from a reciprocal translocation between chromosomes 9 and 22 with the breakpoint joining the sequences of the *ABL* oncogene with the *BCR* gene. The fusion of these DNA sequences allows the generation of an entirely novel fusion protein with modified function.

TUMOR-SUPPRESSOR GENE INACTIVATION IN CANCER

The normal role of tumor-suppressor genes is to restrain cell growth, and the function of these genes is inactivated in cancer. The three major types of somatic lesions observed in tumor-suppressor genes during tumor development are *point mutations*, small insertions and/or deletions known as *indels*, and *large deletions*. Point mutations or indels in the coding region of tumor-suppressor genes will frequently lead to truncated protein products or allele-specific loss of RNA expression by the process of *nonsense-mediated decay*. Unlike the highly recurrent point mutations that are found in critical positions of activated oncogenes, known as mutational *hotspots*, the point mutations that cause tumor-suppressor gene inactivation tend to be distributed throughout the open reading frame. Large deletions lead to the loss of a functional product and sometimes encompass the entire gene or even the entire chromosome arm, leading to loss of heterozygosity (LOH) in the tumor DNA compared to the corresponding normal tissue DNA **(Fig. 76-4)**. Mapping regions of LOH was a useful approach in the positional cloning of many tumor-suppressor genes. The rate of LOH is increased in the presence of chromosomal instability, a relationship that would explain the selective forces leading to the high prevalence of aneuploidy in late-stage cancers.

Gene silencing, an epigenetic change that leads to the loss of gene expression, occurs in conjunction with hypermethylation of the promoter and histone deacetylation, and is another mechanism of tumor-suppressor gene inactivation. An *epigenetic modification* refers to a covalent modification of chromatin, heritable by cell progeny that may involve DNA but does not involve a change in the DNA sequence.

FAMILIAL CANCER SYNDROMES

A small fraction of cancers occurs in patients with a genetic predisposition. Based on studies of inherited and sporadic forms of retinoblastoma, Knudson and others formulated a hypothesis that explains the differences between sporadic and inherited forms of the same tumor type. In inherited forms of cancer, called *cancer predisposition syndromes*, one allele of a particular tumor-suppressor gene is inherited in mutant form. This germline mutation is not sufficient to initiate a tumor, however; the other allele, inherited from the unaffected parent, must become somatically inactivated in a normal stem cell for tumorigenesis to be initiated. In sporadic (noninherited) forms of the same disease, all cells in the body start out with two normal copies of the tumor-suppressor gene. A single cell must then sequentially acquire mutations in both alleles of the tumor-suppressor gene to initiate a tumor. Thus, biallelic mutations of the same tumor-suppressor gene are required for both inherited and noninherited forms of the disease; the only difference is that individuals with the inherited form have a "head start": they already have one allele mutated, from conception, and only need one additional mutation to initiate the process (Fig. 76-4).

This distinction explains why those with inherited forms of the disease develop more cancers, at an earlier age, than the general population. It also explains why, even though every cell in an individual with a cancer predisposition syndrome has a mutant gene, only a relatively small number of tumors arise during their lifetime. The reason is that the vast majority of cells within such individuals are functionally normal because one of the two alleles of the tumor-suppressor gene is normal. Mutations are uncommon events, and only the rare cells that develop a mutation in the remaining normal allele will exhibit uncontrolled proliferation. The same principle applies to virtually all types of cancer predisposition syndromes, though the particular genes differ. For example, inherited mutations in *RB1*, *WT1*, *VHL*, *APC*, and *BRCA1* lead to predispositions to retinoblastomas, Wilms' tumors, renal cell carcinomas, colorectal carcinomas, and breast carcinomas, respectively **(Table 76-3)**. Also note that the biallelic inactivation of any of these genes is not sufficient to develop cancer; it requires other, additional somatic alterations in other genes for the initiating cells to evolve to malignancy, as noted above.

Roughly 100 familial cancer syndromes have been reported; the great majority are very rare. Most of these syndromes exhibit an autosomal dominant pattern of inheritance, although some of those associated with DNA repair abnormalities (xeroderma pigmentosum, Fanconi's anemia, ataxia telangiectasia) are inherited in an autosomal recessive fashion. Table 76-3 shows a number of cancer predisposition syndromes and the responsible genes.

Familial adenomatous polyposis (FAP) is a dominantly inherited colon cancer syndrome caused by germline mutations in the *adenomatous polyposis coli* (*APC*) tumor-suppressor gene on chromosome 5. Affected individuals develop hundreds to thousands of adenomas in the colon. In each of these adenomas, the *APC* allele inherited from the nonaffected parent has been inactivated by virtue of a somatic mutation (Fig. 76-2). This inactivation usually occurs through a gross chromosomal event resulting in loss of all or a large part of the long arm of chromosome 5, where *APC* resides. In other cases, the remaining allele is inactivated by a subtle intragenic mutation of *APC*, which is typically a single base substitution resulting in a nonsense codon.

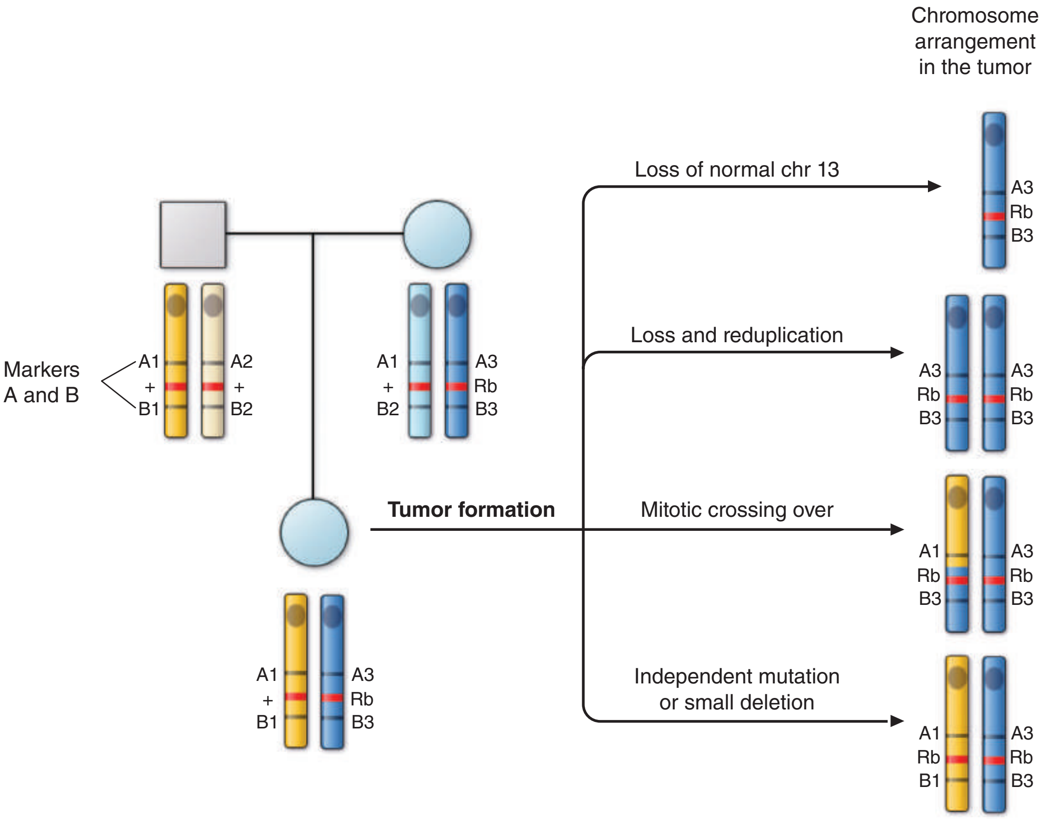

FIGURE 76-4 Diagram of possible mechanisms for tumor formation in an individual with hereditary (familial) retinoblastoma. On the left is shown the pedigree of an affected individual who has inherited the abnormal (Rb) allele from her affected mother. The normal allele is shown as a (+). The four chromosomes of her two parents are drawn to indicate their origin. Flanking the retinoblastoma locus are genetic markers (A and B) also analyzed in this family. Markers A3 and B3 are on the chromosome carrying the retinoblastoma disease gene. Tumor formation results when the normal allele, which this patient inherited from her father, is inactivated. On the right are shown four possible ways in which this could occur. In each case, the resulting chromosome 13 arrangement is shown. Note that in the first three situations, the normal allele (B1) has been lost in the tumor tissue, which is referred to as loss of heterozygosity (LOH) at this locus.

Gross chromosomal losses occur more commonly than point mutations in normal cells, explaining why chromosomal loss rather than point mutation is the predominant mechanism underlying the inactivation of the normal allele of *APC*. The same is true for other cancer predisposition syndromes caused by other inherited tumor suppressor gene mutations; gross chromosomal events are generally responsible for inactivation of the tumor-suppressor gene allele inherited from the nonaffected parent. Several thousand adenomas form in FAP patients, and a small subset of the millions of cells within an adenoma will acquire a second mutation, leading to tumor progression, that is, a larger adenoma. A third mutation in such a larger adenoma may convert it to a carcinoma. If untreated (by colectomy), at least one of the adenomas will progress to cancer by the time patients are in their mid-40s. *APC* is a gatekeeper for colon tumorigenesis in the sense that in the absence of mutation in *APC* (or a gene acting within the same pathway), a colorectal tumor simply cannot be initiated. **Figure 76-5** shows the germline and somatic mutations found in the *APC* gene. A negative regulator of a signaling pathway that determines cell fate during development, the APC protein provides differentiation and apoptotic cues to colonic epithelial cells as they migrate up the crypt. Defects in this process can lead to abnormal accumulation of cells that would otherwise differentiate and eventually undergo apoptosis.

In contrast to patients with FAP, patients with hereditary nonpolyposis colon cancer (HNPCC, or Lynch syndrome) do not develop polyposis, but instead develop only one or a small number of adenomas that rapidly progress to cancer. HNPCC is due to inherited mutations in one of four DNA mismatch repair genes (Table 76-3) that are components of a repair system responsible for correcting errors in newly replicated DNA. Germline mutations in *MSH2* and *MLH1* together account for more than 90% of HNPCC cases, and mutations in *MSH6* and *PMS2* account for the remainder. When a somatic mutation inactivates the remaining wild-type allele of a mismatch repair gene, the cell develops a hypermutable phenotype characterized by profound genomic instability that is most readily apparent in short repeated sequences called *microsatellites* and is sometimes called microsatellite instability (MSI). The high rate of mutation in such cells impacts all genes, including oncogenes and tumor-suppressor genes, and thereby accelerates the activation of the former and the inactivation of the latter (Fig. 76-2). HNPCC can be considered a disease of tumor progression; once tumors are initiated (by an inactivating mutation of *APC* or by some other gene in the APC pathway), tumors rapidly progress because of the accelerated mutation rate. Progression from a tiny adenoma to carcinoma takes only a few years in HNPCC patients instead of the two or three decades this progression takes in patients with FAP (or in patients with sporadic colorectal tumors). Approximately half of HNPCC patients develop colorectal cancers by the time they are in their mid-40s—similar to that of FAP patients. This coincidence in age of onset emphasizes that both tumor initiation (abnormal in FAP patients) and tumor progression (abnormal in HNPCC patients) are the two pillars of cancer development and are equally important for cancer development.

Another general principle is apparent from the comparison between FAP and HNPCC patients. The tumors in FAP patients, like those in patients without hereditary predisposition to cancers, exhibit chromosomal instability rather than MSI. Indeed, MSI and chromosomal instability tend to be mutually exclusive in colon cancers, suggesting that they represent alternative mechanisms for the generation of genomic instability (Fig. 76-2). Other cancer types rarely exhibit MSI. Chromosomal instability is far more prevalent than MSI among all cancer types, perhaps explaining why nearly all cancers are aneuploid.

Although most autosomal dominant inherited cancer syndromes are due to mutations in tumor-suppressor genes (Table 76-3), there are a few interesting exceptions. Multiple endocrine neoplasia type 2, a dominant disorder characterized by pituitary adenomas, medullary

TABLE 76-3 Cancer Predisposition Syndromes and Associated Genes

SYNDROME	GENE	CHROMOSOME	INHERITANCE	TUMORS
Ataxia telangiectasia	*ATM*	11q22-q23	AR	Breast
Autoimmune lymphoproliferative syndrome	*FAS* *FASL*	10q24 1q23	AD	Lymphomas
Birt-Hogg-Dubé syndrome	*FLCN*	17p11.2	AD	Kidney (hybrid oncocytic, chromophobe)
Bloom syndrome	*BLM*	15q26.1	AR	Various
Cowden syndrome	*PTEN*	10q23	AD	Breast, thyroid
Familial adenomatous polyposis	*APC* *MUTYH*	5q21 1p34.1	AD AR	Colorectal (early onset)
Familial melanoma	*CDKN2A*	9p21	AD	Melanoma, pancreatic
Familial Wilms' tumor	*WT1*	11p13	AD	Kidney (pediatric)
Hereditary breast/ovarian cancer	*BRCA1* *BRCA2*	17q21 13q12.3	AD	Breast, ovarian, prostate
Hereditary diffuse gastric cancer	*CDH1*	16q22	AD	Stomach
Hereditary multiple exostoses	*EXT1* *EXT2*	8q24 11p11-12	AD	Exostoses, chondrosarcoma
Hereditary retinoblastoma	*RB1*	13q14.2	AD	Retinoblastoma, osteosarcoma
Hereditary nonpolyposis colon cancer (HNPCC)	*MSH2* *MLH1* *MSH6* *PMS2*	2p16 3p21.3 2p16 7p22	AD	Colon, endometrial, ovarian, stomach, small bowel, ureter carcinoma
Hereditary papillary renal carcinoma	*MET*	7q31	AD	Papillary kidney
Juvenile polyposis syndrome	*SMAD4* *BMPR1A*	18q21	AD	Gastrointestinal, pancreatic
Li-Fraumeni syndrome	*TP53*	17p13.1	AD	Sarcoma, breast
Multiple endocrine neoplasia type 1	*MEN1*	11q13	AD	Parathyroid, endocrine, pancreas, and pituitary
Multiple endocrine neoplasia type 2a	*RET*	10q11.2	AD	Medullary thyroid carcinoma, pheochromocytoma
Neurofibromatosis type 1	*NF1*	17q11.2	AD	Neurofibroma, neurofibrosarcoma, brain
Neurofibromatosis type 2	*NF2*	22q12.2	AD	Vestibular schwannoma, meningioma, spine
Nevoid basal cell carcinoma syndrome (Gorlin's syndrome)	*PTCH1*	9q22.3	AD	Basal cell carcinoma, medulloblastoma, jaw cysts
Peutz-Jeghers syndrome	*STK11/LKB1*	19p13.3	AD	Gastrointestinal, breast
Tuberous sclerosis	*TSC1* *TSC2*	9q34 16p13.3	AD	Angiofibroma, renal angiomyolipoma
von Hippel–Lindau disease	*VHL*	3p25-26	AD	Kidney, cerebellum, pheochromocytoma

Abbreviations: AD, autosomal dominant; AR, autosomal recessive.

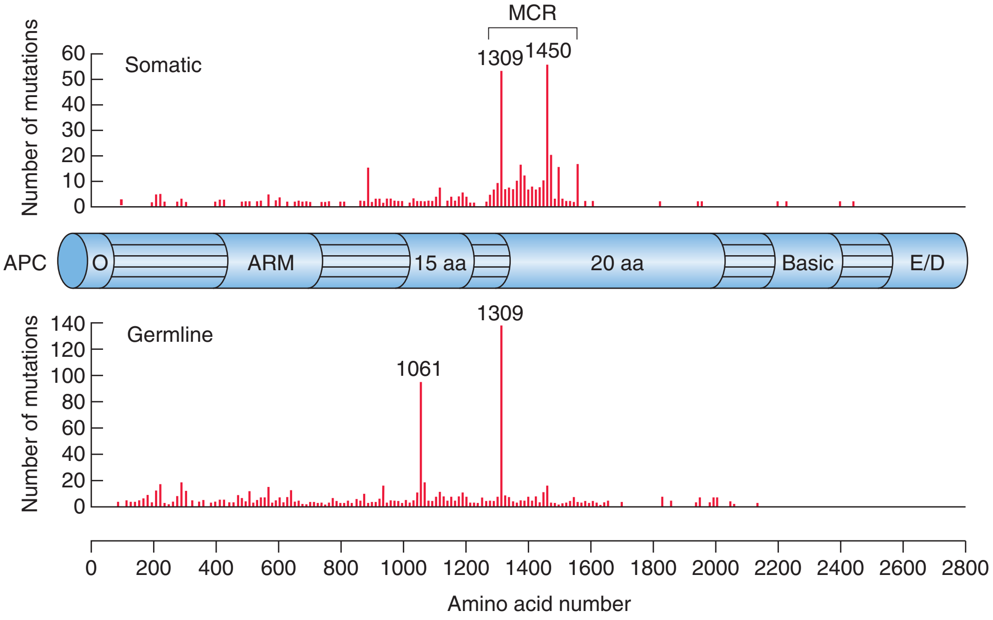

FIGURE 76-5 Germline and somatic mutations in the tumor-suppressor gene adenomatous polyposis coli (*APC*). *APC* encodes a 2843-amino-acid protein with six major domains: an oligomerization region (O), armadillo repeats (ARM), 15-amino-acid repeats (15 aa), 20-amino-acid repeats (20 aa), a basic region, and a domain involved in binding EB1 and the *Drosophila* discs large homologue (E/D). Shown are 650 somatic and 826 germline mutations representative of the mutations that occur within the *APC* gene (from the *APC* database at *www.umd.be/APC*). All known pathogenic mutations of *APC* result in the truncation of the APC protein. Germline mutations are found to be relatively evenly distributed up to codon 1600 except for two mutation hotspots surrounding amino acids 1061 and 1309, which together account for one-third of the mutations found in familial adenomatous polyposis (FAP) families.

carcinoma of the thyroid, and (in some pedigrees) pheochromocytoma, is due to gain-of-function mutations in the proto-oncogene *RET* on chromosome 10. Similarly, gain-of-function mutations in the tyrosine kinase domain of the *MET* oncogene lead to hereditary papillary renal carcinoma. Interestingly, loss-of-function mutations in the *RET* gene cause a completely different disease, Hirschsprung's disease (aganglionic megacolon **[Chaps. 339 and 400]**).

Although the heritable forms of cancer have taught us much about the mechanisms of growth control, most forms of cancer do not follow simple Mendelian patterns of inheritance. The majority of human cancers arise in a sporadic fashion, solely as a result of somatic mutation, and in the absence of any mutations in cancer-predisposing genes in their germlines.

GENETIC TESTING FOR FAMILIAL CANCER

The discovery of cancer susceptibility genes raises the possibility of DNA testing to predict the risk of cancer in individuals of affected families. An algorithm for cancer risk assessment and decision making in high-risk families using genetic testing is shown in **Fig. 76-6**. Once a mutation is discovered in a family, subsequent testing of asymptomatic family members is crucial. A negative gene test in these individuals can prevent years of anxiety, providing comfort in the knowledge that their cancer risk is no higher than that of the general population. On the other hand, a positive test may lead to alteration of clinical management, such as increased frequency of cancer screening and, when feasible and appropriate, prophylactic surgery. Potential negative consequences of a positive test result include psychological distress (anxiety, depression) and discrimination, although the Genetic Information Nondiscrimination Act (GINA) makes it illegal for predictive genetic information to be used to discriminate in health insurance or employment. Testing should therefore not be conducted without counseling before testing is administered and during and after disclosure of the test result.

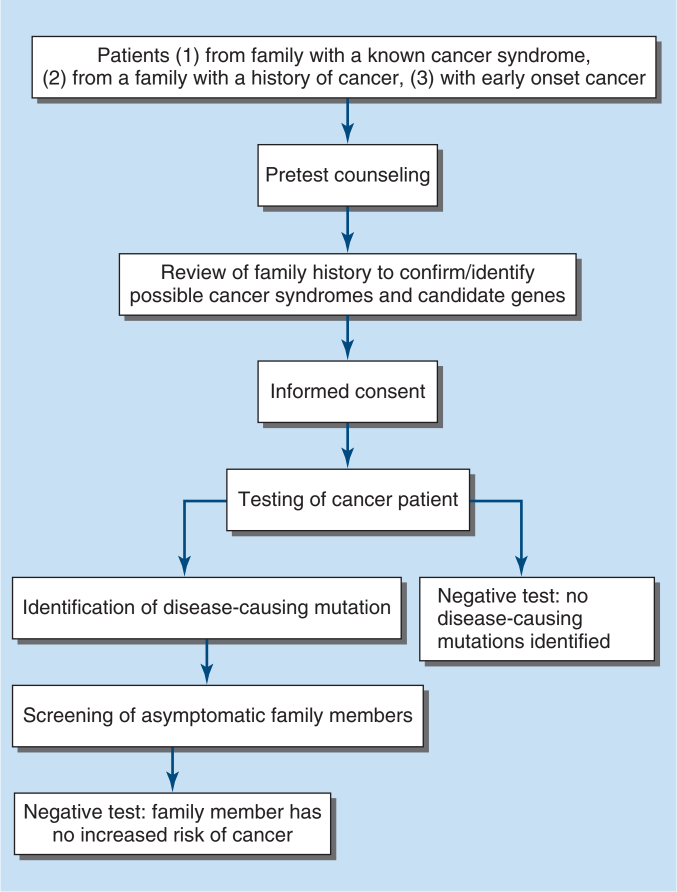
Patients (1) from family with a known cancer syndrome, (2) from a family with a history of cancer, (3) with early onset cancer
Pretest counseling
Review of family history to confirm/identify possible cancer syndromes and candidate genes
Informed consent
Testing of cancer patient
Identification of disease-causing mutation
Negative test: no disease-causing mutations identified
Screening of asymptomatic family members
Negative test: family member has no increased risk of cancer

FIGURE 76-6 Algorithm for genetic testing in a family with cancer predisposition. The key step is the identification of a disease mutation in a cancer patient, which is an indication for the testing of asymptomatic family members. Asymptomatic family members who test positive may require increased screening or surgery, whereas those who test negative are at no greater risk for cancer than the general population. It should be emphasized that no molecular assay used for this sort of testing is 100% sensitive; negative results must be interpreted with this caveat in mind.

It is now feasible to obtain high-quality sequence of all of the protein-coding DNA sequences, and even of the entire genome, in any given individual. In such studies, numerous variants in DNA sequences will inevitably be identified in every subject, but the significance of the vast majority of these DNA sequence findings will be unclear. Even mutations in tumor-suppressor genes can be difficult to interpret unless there is an obvious functional implication, such as the truncation of the open reading frame, or that particular mutation has previously been correlated with cancer in other individuals. Germline mutations associated with cancer predisposition are uncommon in individuals without a family history of cancer, though they do occur. Much more common are *variants of unknown significance* (*VUS*). VUS that are found during genetic testing cannot be used to evaluate the relative risk of cancer but may nonetheless cause anxiety because they represent a deviation from the reference allele that is established as "normal." Because of the low yield of informative mutations that modify cancer risk and the frequent identification of VUS, it is generally not appropriate to use DNA sequencing to assess cancer risk in individuals without a family history of cancer. However, there are exceptions. Testing may be appropriate in some subpopulations with a known increased risk, even without a personal family history. For example, two mutations in the breast cancer susceptibility gene *BRCA1*, 185delAG and 5382insC, exhibit a sufficiently high frequency in the Ashkenazi Jewish population that genetic testing based on ethnicity alone may be warranted.

It is important that genetic test results be communicated to families by trained genetic counselors. To ensure that the families clearly understand its advantages and disadvantages and the impact it may have on disease management and psyche, genetic testing should never be done *before* counseling. Significant expertise is needed to communicate the results of genetic testing to families.

VIRUSES IN HUMAN CANCER

Several human malignancies are associated with viruses. Examples include Burkitt's lymphoma (Epstein-Barr virus; **Chap. 199**), hepatocellular carcinoma (hepatitis viruses), cervical cancer (human papillomavirus [HPV]; **Chap. 203**), and T-cell leukemia (retroviruses; **Chap. 207**). There are several types of HPV, including the high-risk types 16 and 18 that are strongly associated with the development of cervical, vulvar, vaginal, penile, anal, and oropharyngeal cancer. The mechanisms of action of all these viruses involve inactivation of tumor-suppressor genes. For example, HPV proteins E6 and E7 bind to and inactivate cellular tumor suppressors p53 and pRB, respectively. This is the reason that HPV is such a potent initiator of cancer: infection with a virus is tantamount to having two of the three mutant driver genes required for cancer, that is, one viral oncogene inactivates p53 and the other inactivates Rb. Once these two inactivated gene products initiate tumorigenesis, only one additional mutant gene is required for these tumors to progress to malignancy.

■ CANCER GENOMES

The advent of relatively inexpensive technologies for rapid and high-throughput DNA sequencing has facilitated the comprehensive analysis of numerous genomes from many types of tumors. This unprecedented view into the genetic nature of cancer has provided remarkable insights. Most cancers do not arise in the context of a mutator phenotype, and accordingly, the number of mutations in even the most advanced cancers is relatively modest. Common solid tumors harbor 30–70 subtle mutations that are nonsynonymous (i.e., result in an amino acid change in the encoded protein). Liquid tumors such as leukemias, as well as pediatric tumors, typically have fewer than 20 mutations. The vast majority of the mutations detected in tumors are not functionally significant; they simply arose by chance in a single cell that gave rise to an expanding clone. Such mutations, which provide no selective advantage to the cell in which they occur, are known as *passenger* mutations. As noted above, only a small number of the mutations confer a selective growth advantage and thereby promote tumorigenesis. These functional mutations are known as *driver* mutations, and the genes in which they occur are called driver genes.

The frequency and distribution of driver mutations within a single tumor type can be represented as a topographical landscape. The picture that emerges from cancer genome studies reveals that most genes that are mutated in tumors are actually mutated at relatively low frequencies, as would be expected of passenger genes, whereas a small number of genes (the driver genes) are mutated in a large proportion of tumors. Only ~200 driver genes contribute to the development of solid tumors of all kinds. Driver genes that play a role in ever smaller fractions of cancers are still being discovered. The majority of the mutations in driver genes provide a direct selective growth advantage by altering the signaling pathways that mediate cell survival or the determination of cell fate. The remaining driver gene mutations indirectly provide a selective growth advantage by accelerating the mutation rate of proto-oncogenes and tumor-suppressor genes. That the same driver genes play a role in multiple cancer types was unexpected before their discovery and has important implications for the development of new "tumor-agnostic" therapeutic and diagnostic approaches. Moreover, the functions of all these driver genes can be organized into a small number of signaling pathways, as shown in **Table 76-4**.

As a consequence of the mutations they harbor, cancer cells invariably express mutant proteins that are only rarely found in normal cells. Some of these mutant proteins are processed and displayed on the cell surface in the context of major histocompatibility complexes, a process that would normally facilitate their recognition by the adaptive immune system. Thus, all cancers have the theoretical potential to be recognized as foreign, or "nonself," via the display of these tumor-specific antigens, known as mutation-associated neoantigens (MANAs). In fact, established tumors invariably prevent the activation of local T cells by inducing an intercellular suppressive mechanism known as an *immune checkpoint*. Therapeutic approaches to exploit this potential vulnerability by blocking immune checkpoints have elicited striking responses in patients with several types of cancer.

It was hypothesized that the potential immunogenicity of a tumor would be related to the total number of distinctive neoantigens it can express, which in turn is directly determined by the total number of mutations in the cancer genome. This does seem to be the case. Colorectal cancers that develop as a result of mismatch repair deficiency and smoking-related lung cancers, both of which characteristically harbor large numbers of mutations, exhibit more robust responses to therapeutic immune checkpoint blockade than most other tumor types. Notably, driver mutations as well as passenger mutations that result in the expression of mutant proteins can both contribute to the display of immunogenic neoantigens. Thus, the total number of coding mutations, a metric known as *mutational load*, is one of the determinants of potential immunogenicity.

TABLE 76-4 Signaling Pathways Altered in Cancer

PROCESS	PATHWAY	REPRESENTATIVE DRIVER GENES
Cell survival	Cell cycle regulation/ apoptosis	*RB1, BCL2*
	RAS	*KRAS, BRAF*
	PIK3CA	*PTEN, PIK3CA*
	JAK/STAT	*JAK2, FLT3*
	MAPK	*MAP3K, ERK*
	TGF-β	*BMPR1A, SMAD4*
Cell fate	Notch	*NOTCH1, FBWX7*
	Hedgehog	*PTCH1, SMO*
	WNT/APC	*APC, CTNNB1*
	Chromatin modification	*DNMT1, IDH1*
	Transcriptional regulation	*AR, KLF4*
Genome maintenance	DNA damage signaling and repair	*ATM, BRCA1*

The ability of cancer cells to evade immune-mediated cell death is an intrinsic property that is essential for their continued growth. While tumor suppressor genes and oncogenes have been intensively studied with respect to their effects on the intracellular signaling pathways that regulate cell proliferation and cell death, little is known about how these genetic alterations affect the interactions between cancer cells and neighboring immune cells. In particular, neoplastic cells containing certain mutations may be culled by the immune system because the mutations create neoantigens that can be recognized by T cells.

TUMOR HETEROGENEITY

The mutant cells that compose a single tumor are not genetically identical. Rather, cells obtained from different sites on a tumor will harbor common mutations as well as mutations that are unique to each sample. Genetic heterogeneity results from the ongoing acquisition of mutations during tumor growth. Each time a genome is replicated, there is a small but quantifiable probability that a mutation will spontaneously arise as a result of a replication error and be passed on to the cellular progeny. This is true in normal cells or in tumor cells. Any randomly chosen cell from the skin of one individual will harbor hundreds of genetic alterations that distinguish it from a different randomly chosen skin cell, and the same is true for all organs of self-renewing tissues. Tumors are actually *less* genetically heterogeneous than normal tissues; any two randomly chosen cells from a tumor of an individual will have fewer differences than any two randomly chosen cells from that individual's normal tissues. The reason for this decrease in heterogeneity is clonal expansion, the fundamental feature of tumorigenesis. Every time a clonal expansion occurs, a genetic bottleneck wipes out heterogeneity among the cells that did not expand; these unexpanded cells either die or form only a minute proportion of the total cells in the expanding tumor.

The mutations that vary between cells of a given tumor are invariably passenger mutations that arose since the last evolutionary bottleneck, that is, those mutations that arose during the expansion of the founder cell that gave rise to the final clonal expansion. In contrast, the passenger mutations that were present in the founder cell will be uniformly present in every cell in the tumor. In that respect, these passenger mutations are not heterogeneously distributed and are in fact uniformly present in virtually all cancer cells. These "clonal" mutations, i.e., present in all cells of the cancers, are the main source of MANAs that can be exploited through immune checkpoint inhibitors. The total number of mutations and their distribution within tumor cells represent a complex interplay between the age of the patient (the older the patient, the more passenger mutations will have accumulated in the founding cell of the first clonal expansion) and the evolutionary history of the cancer (its age and number of clonal expansions it experienced).

Tumor heterogeneity has been recognized for decades at the cytogenetic, biochemical, and histopathologic levels. However, it is only recently, with the advent of a deep understanding of cancer genetics, that genetic heterogeneity can be interpreted in a medically relevant fashion. The first important point to recognize about tumor heterogeneity is that it is only the variation in driver gene alterations that is important; the cellular distribution of passenger gene mutations is irrelevant except for immune-related phenomena. In this discussion of heterogeneity, we can expand the definition of "driver genes" to include those that provide a selective growth advantage in the face of therapy in addition to those that provide a selective growth advantage during tumor evolution, prior to treatment.

Type I heterogeneity refers to that among tumors of the same type from different patients **(Fig. 76-7)**. Though adenocarcinomas of the lung generally harbor mutations in three or more driver genes, the genes differ among the patients, and the precise mutations within the same gene can vary considerably. Type I heterogeneity is the basis for precision medicine, where the goal is to treat patients with drugs that target the proteins encoded by genetic alterations within their specific tumors. Type II heterogeneity refers to the genetic heterogeneity among different cells from the same primary tumor. Tumors continue to evolve as

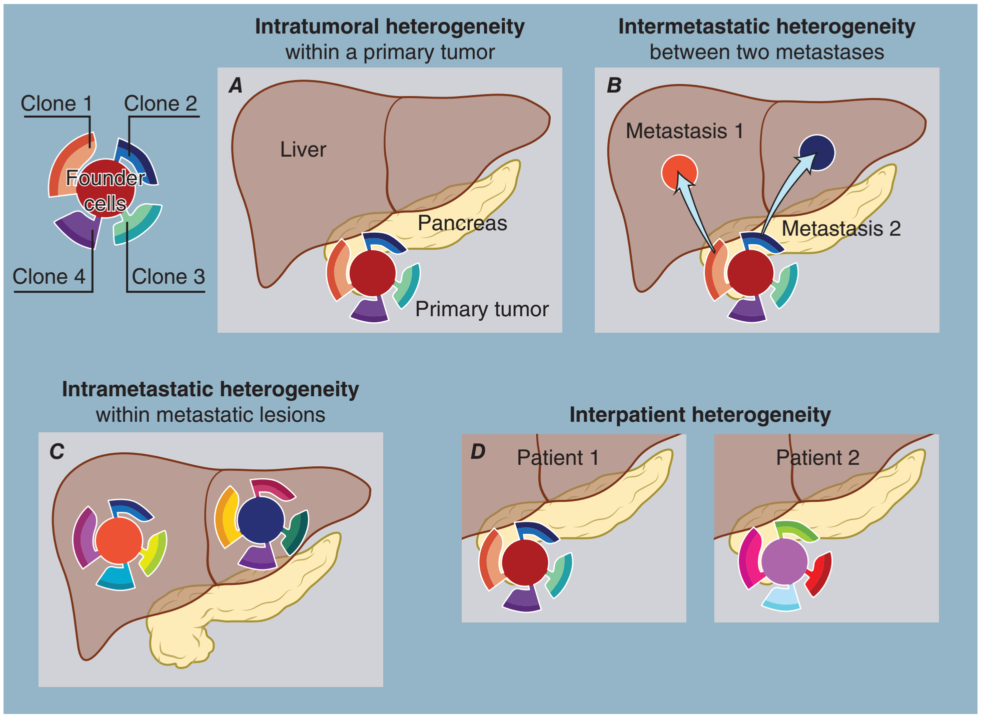

FIGURE 76-7 Four types of tumor heterogeneity. Tumor heterogeneity is the inevitable result of cell proliferation, as new mutations are introduced during clonal expansion. In a typical tumor (*upper left*), founder cells that harbor a large fraction of the total mutations give rise to subclones, which continue to evolve independently. The tumors of the founding populations are shown in the middle of each circle; the distinct subclones are shown around the periphery. ***A.*** Heterogeneity among the cells of a primary tumor is known as intratumoral heterogeneity. ***B.*** Heterogeneity among the founding cells of distinct metastatic lesions (marked as 1 and 2) that arise in the same patient is known as intermetastatic heterogeneity. ***C.*** Heterogeneity among the cells of each metastatic tumor is known as intrametastatic heterogeneity. ***D.*** Interpatient heterogeneity. The mutations in the tumors of two patients are almost completely distinct. *(Reproduced with permission from B Vogelstein et al: Cancer genome landscapes. Science 339(6127):1546, 2013.)*

they grow, and different cells of the same cancer, in its original site (e.g., the pancreas), may acquire other driver gene mutations that are not shared among the other cells of the tumor. Such a mutation can result in a small clonal expansion that may or may not be important biologically. In cases in which the primary tumor can be surgically excised, such mutations are unimportant unless they give rise to type III heterogeneity (described below). The reason they are unimportant is because all primary tumor cells, whether homogeneous or not, are removed by the surgical procedure. In primary tumors that cannot be completely excised (such as most advanced brain tumors and many pancreatic ductal adenocarcinomas), heterogeneity is biomedically important because it can give rise to drug resistance, analogously to that described for type IV heterogeneity (see below). Type III heterogeneity refers to the genetic differences among the founder cells of the metastatic lesions from the same patient. For example, a patient with melanoma may have 100 different metastases distributed throughout various organs. Only if a mutant *BRAF* is present in every founder cell of every metastasis, then the patient has a chance at a complete response to a *BRAF* inhibitor. There have been several recent detailed studies of the metastases from various tumor types. Fortunately, these studies suggest there is very little, if any, type III heterogeneity among driver genes, a necessary prerequisite for the successful implementation of current and future targeted therapies. Finally, type IV heterogeneity refers to that among cells of individual metastatic lesions. As the founder cell of each metastasis expands to become detectable, it acquires mutations, a small number of which can act as "drivers" when the patient is exposed to therapeutics. This type of heterogeneity is of major clinical importance, as it has been shown to be responsible for the development of resistance in virtually all targeted therapies. The development of such resistance is a fait accompli based simply on known mutation rates and genetic resistance mechanisms. The only way to circumvent acquired resistance is to treat metastatic tumors earlier (i.e., in adjuvant setting, before much tumor expansion has occurred) or to treat with combinations of drugs for which cross-resistance is genetically impossible.

PERSONALIZED CANCER DETECTION AND TREATMENT

High-throughput DNA sequencing has led to an unprecedented understanding of cancer at the molecular level. A comprehensive mutation profile provides a molecular history of a given tumor and insights into how it arose. Because tumor cells and tumor DNA are shed into the blood and other bodily fluids, common driver mutations can be used as highly specific biomarkers for early detection. For diagnosed tumors, tumor-specific mutations can be used to estimate tumor burden, assess treatment responses, and detect recurrence.

In some cases, information regarding specific genes and pathways that are altered provides patients and physicians with options for personalized therapy. This general approach is sometimes referred to as *precision medicine*. Because tumor behavior is highly variable, even within a tumor type, personalized information-based medicine can supplement and perhaps eventually supplant histology-based tumor assessment, especially in the case of tumors that are resistant to conventional therapeutic approaches. Conversely, molecular nosology has revealed similarities in tumors of diverse histotype. The success of the precision medicine approach in any given patient depends on the presence of tumor-associated genetic alterations that are actionable (i.e., can be targeted with a specific drug). Examples of currently actionable changes include mutations in *BRAF* (targeted by the drug vemurafenib), *RET* (targeted by sunitinib and sorafenib), *ALK* rearrangements (targeted by crizotinib), and mismatch repair genes (targetable by immune checkpoint inhibitors).

The development of new targeted agents is at present hindered by the fact that most such agents can only target activated oncogenes, while the great majority of genetic alterations in common solid tumors are those that inactivate tumor-suppressor genes. Because all drugs, whether for use in oncology or any other purpose, can only inhibit protein actions, drugs cannot be used to directly target the proteins encoded by inactivated tumor-suppressor genes; these proteins are already inactive. More information about the pathways through which tumor-suppressor genes act may provide a way around this obstacle. For example, when a tumor-suppressor gene is inactivated, some downstream component of the pathway is likely to be activated, thereby presenting a realistic target. Alternatively, the mutational inactivation of a DNA repair pathway might create a unique dependence on the repair pathways that remain intact. An example of this is provided by PARP-1 inhibitors, which have been successfully used to treat patients whose tumors have inactivating mutations of genes involved in DNA repair processes, such as *BRCA1*. Patterns of global gene expression can be used to help unravel such pathways and are already being used to predict drug sensitivities and provide prognostic information in addition to that provided by DNA sequence analysis. Evaluation of proteomic and metabolomic patterns may also prove useful for this purpose.

■ THE FUTURE

A revolution in cancer genetics has occurred in the past 30 years. Most types of cancer are now understood at the DNA sequence level, and this accomplishment has led to an increasingly refined understanding of tumorigenesis. Cancer gene mutations have proven to be reliable biomarkers for cancer detection and monitoring as well as for informing therapeutics through precision medicine approaches. Gene-based tests are already standard of care for patients with certain tumor types, such as colorectal and lung cancers, and the utility of these tests will undoubtedly be expanded in the coming years as new therapies and ways of predicting responses to therapies are developed. While effective treatment of advanced cancers remains difficult, the early promise shown by immune-based therapies notwithstanding, it is expected that breakthroughs in these areas will continue to emerge and be applicable to an ever-increasing number of cancers. Moreover, with the hoped-for advances in diagnostics, particularly in the earlier detection of cancers, the new and old therapies for cancer can be expected to have a much greater impact on reducing cancer deaths.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the past contributions of Pat J. Morin, Jeff Trent, and Francis Collins to earlier versions of this chapter.

■ FURTHER READING

Bunz F: *Principles of Cancer Genetics*, 3rd ed. Dordrecht, Springer, 2022.

Le DT et al: PD-1 Blockade in tumors with mismatch-repair deficiency. N Engl J Med 372:2509, 2015.

Vogelstein B, Kinzler KW: The path to cancer—three strikes and you're out. N Engl J Med 373:1895, 2015.

Vogelstein B et al: Cancer genome landscapes. Science 339:1546, 2013.

77

Cancer Cell Biology

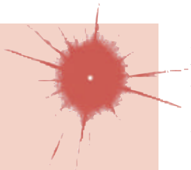

Jeffrey W. Clark, Dan L. Longo

■ CANCER CELL BIOLOGY

Cancers are characterized by unregulated cell division, avoidance of cell death, tissue invasion, and the ability to spread to other areas of the body (metastasize). A neoplasm is benign when it grows in an unregulated fashion without tissue invasion or metastasizing. The presence of unregulated growth, tissue invasion, and the ability to metastasize is characteristic of malignant neoplasms. Cancers are named based on their tissue of origin: those derived from epithelial tissue are called carcinomas, those derived from mesenchymal tissues are sarcomas, and those derived from hematopoietic tissue are leukemias, lymphomas, and plasma cell dyscrasias (including multiple myeloma).

Cancers arise as a consequence of genetic alterations, the vast majority of which begin in a single cell and therefore are monoclonal in origin. However, because a wide variety of genetic and epigenetic changes can occur in different cells within malignant tumors over time, as well as varied responses with their microenvironments and the biology of the patient, most cancers are characterized by plasticity and marked heterogeneity in the populations of cells and their composite behavior. In addition, extrinsic factors in the cancer environment (e.g., the stroma, infiltrating cells, various cell-to-cell interactions, spatial orientation, secreted factors, and availability of oxygen and nutrients) vary in different areas within the tumor or different metastases, compounding this heterogeneity. This heterogeneity significantly complicates the treatment of most cancers because it is likely that there are subsets of cells that will be resistant to therapy for a variety of reasons and will therefore survive and proliferate even if the majority of cells are killed.

A few cancers appear to, at least initially, be primarily driven by an alteration in a dominant gene that produces uncontrolled cell proliferation. Examples include chronic myeloid leukemia (*abl*), about half of melanomas (*braf*), Burkitt's lymphoma (*c-myc*), and subsets of lung adenocarcinomas (*egfr, alk, ros1, met, ret, braf*, and *ntrk*). Genes that can promote cell growth when altered are often called oncogenes. They were first identified as critical elements of viruses that cause animal tumors; it was subsequently found that the viral genes had normal counterparts with important functions in the cell and had been captured and mutated by viruses as they passed from host to host.

However, most human cancers are characterized by a multiple-step process involving many genetic abnormalities, each of which contributes to the loss of control of cell proliferation and differentiation and the acquisition of capabilities, such as tissue invasion, the ability to metastasize, angiogenesis (development of new blood vessels required for tumor growth), and alteration of the extracellular environment. These properties are not found in the normal adult cell from which the tumor is derived. Indeed, normal cells have a large number of safeguards against DNA damage (including multiple DNA repair and extensive DNA damage response mechanisms), uncontrolled proliferation, and invasion. Many cancers go through recognizable steps of progressively more abnormal phenotypes: hyperplasia, to adenoma, to dysplasia, to carcinoma in situ, to invasive cancer with the ability to metastasize **(Table 77-1)**. For most cancers, these changes occur over a prolonged period of time, usually many years.

In most organs, only primitive undifferentiated cells are capable of proliferating and cells lose the capacity to proliferate as they differentiate and acquire functional capabilities. The expansion of the primitive cells (stem cells) is linked to some functional need in the host, such as normal turnover of tissues or regeneration after acute injury, through receptors that receive signals from cells and other factors in the local tissue microenvironment or through hormonal and other influences delivered by the vascular supply. In the absence of such signals, the cells are at rest or quiescent (out of the cell cycle but capable of being activated to reenter the cell cycle). The signals that induce quiescence in primitive cells as well as those that keep the cells at rest are complex, including the process for quiescent entry, maintenance, and exit. Although much has been learned, including the importance of notch signaling, the STING pathway, other quiescent factors, and transcriptional, posttranscriptional, and epigenetic regulation in quiescent entry and maintenance, overall control of the process within the body remains incompletely understood. These signals must be, at least in part, environmental, based on the observations that a regenerating liver stops growing when it has replaced the portion that has been surgically removed after partial hepatectomy and regenerating bone marrow stops growing when the peripheral blood counts return to normal. Cancer cells clearly have lost responsiveness to such controls and do not recognize when they have overgrown the niche normally occupied by the organ from which they are derived. A better understanding of these mechanisms of growth regulation in the context of organ homeostasis continues to evolve.

TABLE 77-1 Phenotypic Characteristics of Malignant Cells

Deregulated cell proliferation: Loss of function of negative growth regulators (tumor suppressor genes, i.e., *Rb*, *p53*) and increased action of positive growth regulators (oncogenes, i.e., *Ras, Myc*). Leads to aberrant cell cycle control and includes loss of normal checkpoint responses.

Failure to differentiate: Arrest at a stage before terminal differentiation. May retain stem cell properties. (Frequently observed in leukemias due to transcriptional repression of developmental programs by the gene products of chromosomal translocations.)

Loss of normal apoptosis pathways: Inactivation of p53, increases in Bcl-2 (antiapoptotic) family members. This defect enhances the survival of cells with oncogenic mutations and genetic instability and allows clonal expansion and diversification within the tumor without activation of physiologic cell death pathways.

Genetic instability: Defects in DNA repair pathways leading to either single nucleotide or oligonucleotide mutations (as in microsatellite instability, MIN) or, more commonly, chromosomal instability (CIN) leading to aneuploidy (abnormal number of chromosomes in a cell). Caused by loss of function of a number of proteins including p53, BRCA1/2, mismatch repair genes, DNA repair enzymes, and the spindle checkpoint. Leads to accumulation of a variety of mutations in different cells within the tumor and heterogeneity.

Loss of replicative senescence: Normal cells stop dividing in vitro after 25–50 population doublings. Arrest is mediated by the Rb, p16INK4a, and p53 pathways. While most cells remain arrested, genetic and epigenetic changes in a subset of cells allow further replication, leading to telomere loss, with crisis leading to death of many cells. Cells that survive often harbor gross chromosomal abnormalities and the ability to continue to proliferate. These cells express telomerase, which maintains telomeres and is important for ongoing growth of these cells. Relevance to human in vivo cancer remains uncertain. Many human cancers express telomerase.

Nonresponsiveness to external growth-inhibiting signals: Cancer cells have lost responsiveness to signals normally present to stop proliferating when they have overgrown the niche normally occupied by the organ from which they are derived. Our understanding about this mechanism of growth regulation remains limited.

Increased angiogenesis: Due to increased gene expression of proangiogenic factors (VEGF, FGF, IL-8, angiopoietin) by tumor or stromal cells, or loss of negative regulators (endostatin, tumstatin, thrombospondin).

Invasion: Cell mobility and ability to move through extracellular matrix and into other tissues or organs. Loss of cell-cell contacts (gap junctions, cadherins) and increased production of matrix metalloproteinases (MMPs). Can take the form of epithelial-to-mesenchymal transition (EMT), with anchored epithelial cells becoming more like motile fibroblasts.

Metastasis: Spread of tumor cells to lymph nodes or distant tissue sites. Limited by the ability of tumor cells to migrate out of initial site and to survive in a foreign environment, including evading the immune system (see below).

Evasion of the immune system: Downregulation of MHC class I and II molecules; induction of T-cell tolerance; inhibition of normal dendritic cell and/or T-cell function; antigenic loss variants and clonal heterogeneity; increase in regulatory T cells.

Shift in cell metabolism: Complex changes including alterations due to tumor stress such as hypoxia and energy generation shifts from oxidative phosphorylation to aerobic glycolysis generate building blocks for malignant cell production and proliferation.

Complex interactions with the extracellular environment around the cancer cells: Induction of changes as well as complex interactions with the extracellular environment around cancer cells, including modifications to the extracellular matrix, vasculature, chemokines, mesenchymal stromal cells, fibroblasts, immune cells, other hematopoietic cells, platelets, nerves, and potentially infectious agents impacting many of the above processes.

Abbreviations: FGF, fibroblast growth factor; IL, interleukin; MHC, major histocompatibility complex; VEGF, vascular endothelial growth factor.

DIFFERENCES BETWEEN PEDIATRIC AND ADULT CANCERS

The underlying importance of genetic mutations and other molecular changes is similar for pediatric and adult cancers. However, some important differences exist. Childhood cancers have a different epidemiology (e.g., they do not have the same extent of environmental or lifestyle risk factors), are much less frequent, and have a different spectrum of frequency beginning primarily during embryogenesis in mesodermal (e.g., sarcomas or hematologic malignancies such as acute lymphocytic leukemia [ALL] or lymphomas) or ectodermal (e.g., neuronal including brain and spinal) tissues. They generally have fewer genetic changes and lower mutational burdens than adult cancers. Likely due to a number of factors including the nature of the cancers (e.g., less genetic complexity than adult cancers) and the ability of children to tolerate more intense chemotherapy regimens, childhood cancers generally tend to be much more responsive to chemotherapy than adult cancers with significantly higher rates of cure.

CANCER AS AN ORGAN THAT IGNORES ITS NICHE

The fundamental cellular defects that create a malignant neoplasm act at the cellular level, and some of these are cell autonomous. However, that is not the entire story. Cancers consist of both malignant cells as well as other cells, blood vessels, extracellular matrix, and signaling and other molecules in the cancer microenvironment. They behave as organs that have lost their specialized function and stopped responding to signals that would limit their growth in tightly regulated normal tissue homeostasis. Most human cancers usually become clinically detectable when a primary mass is approximately 1 cm in diameter—such a mass consists of about 10^9 cells. Often, patients present with tumors that are approximately 10^{10} cells. Although it varies by type of cancer and where the primary tumor and metastases are located, a lethal tumor burden is usually about 10^{12}–10^{13} cells. If all malignant cells were dividing without any cell death at the time of diagnosis, most patients would reach a lethal tumor burden in a very short time. However, human tumors grow by Gompertzian kinetics—this means that not every daughter cell produced by a cell division is actively dividing. In addition, the overall growth rate of a tumor depends on differences between growth rates of different cells within the tumor and rate of cell loss. The growth fraction of a tumor declines with time, largely due to factors in the microenvironment and accumulation of genetic damage over time. The growth fraction of the first malignant cell is 100%, and by the time a patient presents for medical care, the growth fraction is estimated to be <5%, although the fraction varies between different types of cancers and even different cancers of the same type in different individuals. This fraction is often similar to the growth fraction of normal bone marrow and normal intestinal epithelium, the most highly proliferative normal tissues in the human body, a fact that may explain the dose-limiting toxicities to these tissues of agents that target dividing cells.

The implication of these data is that the tumor is slowing its own growth over time. How does it do this? The tumor cells have multiple genetic lesions that tend to promote proliferation, yet by the time the tumor is clinically detectable, its capacity for proliferation has declined. Better understanding of how a tumor slows its own growth would provide important clues for better cancer treatment. A number of factors, including those in the tumor microenvironment, are known to contribute to the decreased proliferation of tumor cells over time in the patient. For example, normal cells and other factors in the microenvironment can contribute to slowing down the growth of cancer cells. Some cancer cells are hypoxemic and have inadequate supply of nutrients and energy. Some have sustained too much genetic damage to complete the cell cycle but have lost the capacity to undergo apoptosis and therefore survive but do not proliferate. However, an important subset is not actively dividing but retains the capacity to divide and can start dividing again under certain conditions such as when the tumor mass is reduced by treatments leading to improved conditions in the tumor microenvironment favorable for cell proliferation.

Just as the bone marrow increases its rate of proliferation in response to bone marrow–damaging agents, the tumor also seems to sense when tumor cell numbers have been reduced and can respond by increasing growth rate. However, the critical difference is that the marrow stops growing when it has reached its production goals, whereas tumors do not.

The ultimate structure and organization of an organ are based on a number of factors including growth, migration, elimination, and death of various cells; communication between cells to establish the correct architecture; competition between cells; and the composition of the extracellular matrix that is produced. In addition to normal cells stopping proliferation in an organ when that is appropriate, normal tissues have various mechanisms for eliminating cells in both the process of development as well as ongoing homeostasis of an organ. These include mechanical processes based on a number of factors including cell size, cell shape, and topology between cells that can determine cell fate as well as an active process of cell extrusion, which plays a major role in the elimination of both cells that are no longer needed by the organ and cells that are damaged and potentially dangerous (such as those with mutations that might be precursors for malignancy). The process of cell extrusion may depend on cell cycle arrest in the S phase; aberrations in this process may contribute to the metastatic process. A variety of processes, including major alterations in cell cycle control, apoptosis and other mechanisms of cell death, and uncontrolled cell signaling, all contribute to defects in appropriate cell extrusion contributing to the development of cancer.

Additional tumor cell vulnerabilities are likely to be detected when we learn more about how normal cells respond to "stop" signals from their environment, and why and how tumor cells and tissues fail to heed such signals.

CELL CYCLE CHECKPOINTS

The cell division cycle consists of four phases—G_1 (growth and preparation for DNA synthesis), S (DNA synthesis), G_2 (preparation to divide), and M (mitosis, cell division). Cells can also exit the cell cycle and be quiescent (G_0). Progression of a cell through the cell cycle is tightly regulated at a number of checkpoints (especially at the G_1/S boundary, the G_2/M boundary, and during M [spindle checkpoint]) by an array of genes that are targeted by specific genetic alterations in cancer. These checkpoints are quality-control features; at G_1, the checkpoint does not allow cells to proceed that are not ready for genome replication; at G_2/M, the cell assesses whether the genome has been appropriately duplicated and is ready to divide. Critical proteins in these control processes that are frequently mutated or otherwise inactivated in cancers are called tumor-suppressor genes because when they function normally, they inhibit the development or growth of cancer cells. Examples include *p53* and *Rb* (discussed below).

In the first phase, G_1, preparations are made to replicate the genetic material. The cell stops before entering the DNA synthesis phase, or S phase, to take inventory. Are we ready to replicate our DNA? Is the DNA repair machinery in place to fix any mutations that are detected? Are the DNA replicating enzymes available? Is there an adequate supply of nucleotides? Is there sufficient energy to proceed? The retinoblastoma protein, Rb, plays a central role in placing a brake on the process until the cell is ready. When the cell determines that it is prepared to move ahead, sequential activation of cyclin-dependent kinases (CDKs) results in the inactivation of the brake, Rb, by phosphorylation. Phosphorylated Rb releases the S phase–regulating transcription factor, E2F/DP1, and genes required for S-phase progression are expressed. If the cell determines that it is unready to move ahead with DNA replication, a number of inhibitors are capable of blocking the action of the CDKs, including p21Cip2/Waf1, p16Ink4a, and p27Kip1. Nearly every cancer has one or more defects in the G_1 checkpoint that permit progression to S phase despite abnormalities in DNA repair machinery or other deficiencies that would affect normal DNA synthesis.

At the end of the G_2 phase and before the M phase, after the cell has exactly duplicated its DNA content, a second inventory is taken at the G_2 checkpoint. Have all of the chromosomes been fully duplicated? Were all segments of DNA copied only once? Has all damaged DNA been repaired? Do we have the right number of chromosomes and the right amount of DNA? If so, the cell proceeds to G_2, in which the cell prepares for division by synthesizing mitotic spindle and other proteins needed to produce two daughter cells. If DNA damage is detected, the p53 pathway is normally activated. Called the guardian of the genome, p53 is a transcription factor that is normally present in the cell in very low levels. This level is generally regulated through its rapid turnover. Normally, p53 is bound to mdm2, a ubiquitin ligase that both inhibits p53 transcriptional activation and also targets p53 for degradation in the proteasome. When DNA damage is sensed, the ATM (ataxia-telangiectasia mutated) pathway is activated; ATM phosphorylates mdm2, releasing it from its inhibitory bond to p53. p53 then stops cell cycle progression, directs the synthesis of repair enzymes, or if the damage is too great, initiates apoptosis (programmed cell death) of the cell to prevent the propagation of a damaged cell (Fig. 77-1).

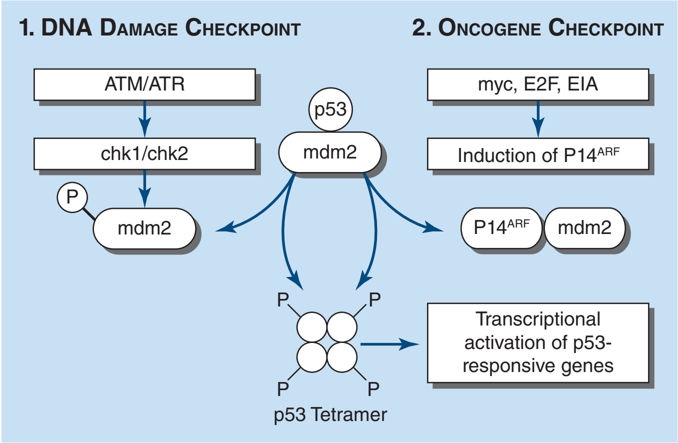

FIGURE 77-1 Induction of p53 by the DNA damage and oncogene checkpoints. In response to noxious stimuli, p53 and mdm2 are phosphorylated by the ataxia-telangiectasia mutated (ATM) and related ATR serine/threonine kinases, as well as the immediate downstream checkpoint kinases, Chk1 and Chk2. This causes dissociation of p53 from mdm2, leading to increased p53 protein levels and transcription of genes leading to cell cycle arrest ($p21^{Cip1/Waf1}$) or apoptosis (e.g., the proapoptotic Bcl-2 family members Noxa and Puma). Inducers of p53 include hypoxemia, DNA damage (caused by ultraviolet radiation, gamma irradiation, or chemotherapy), ribonucleotide depletion, and telomere shortening. A second mechanism of p53 induction is activated by oncogenes such as *Myc*, which promote aberrant G_1/S transition. This pathway is regulated by a second product of the Ink4a locus, $p14^{ARF}$ (p19 in mice), which is encoded by an *a*lternative *r*eading *f*rame (ARF) of the same stretch of DNA that codes for $p16^{Ink4a}$. Levels of ARF are upregulated by *Myc* and E2F, and ARF binds to mdm2 and rescues p53 from its inhibitory effect. This *oncogene checkpoint* leads to the death or senescence (an irreversible arrest in G_1 of the cell cycle) of renegade cells that attempt to enter S phase without appropriate physiologic signals. Senescent cells have been identified in patients whose premalignant lesions harbor activated oncogenes, for instance, dysplastic nevi that encode an activated form of BRAF (see below), demonstrating that induction of senescence is a protective mechanism that operates in humans to prevent the outgrowth of neoplastic cells.

A second method of activating p53 involves the induction of p14ARF by hyperproliferative signals from oncogenes. p14ARF competes with p53 for binding to mdm2, allowing p53 to escape the effects of mdm2 and accumulate in the cell. p53 then stops cell cycle progression by activating CDK inhibitors such as p21 and/or initiating the apoptosis pathway. Not surprisingly given its critical role in controlling cell cycle progression, mutations in the gene for p53 on chromosome 17p are among the most frequent mutations in human cancers, although percentages vary between different cancers. Most commonly these mutations are acquired in the malignant tissue in one allele and the second allele is inactivated (such as by deletion or epigenetic silencing), leaving the cell unprotected from DNA-damaging agents or activated oncogenes.

Some environmental exposures produce signature mutations in p53; for example, aflatoxin exposure leads to mutation of arginine to serine at codon 249 and leads to hepatocellular carcinoma. In rare instances, p53 mutations are in the germline (Li-Fraumeni syndrome) and produce a familial cancer syndrome. Another mechanism for inactivation of p53 in malignant cells is due to alterations in regulators such as overexpression of the inhibitory mdm2 protein. Whether inactivated by mutation or inhibited by regulatory factors, absence of normal p53 function leads to chromosomal instability and accumulation of DNA damage including acquisition of properties that give the abnormal cell a proliferative and survival advantage. Like Rb dysfunction, most cancers have mechanisms that disable the p53 pathway. Indeed, the importance of p53 and Rb in the development of cancer is underscored by the neoplastic transformation mechanism of human papillomavirus. This virus has two main oncogenes, E6 and E7. E6 acts to increase the rapid turnover of p53, and E7 acts to inhibit Rb function; inhibition of these two targets is required for transformation of epithelial cells by the virus.

Another cell cycle checkpoint exists when the cell is undergoing division (M phase); this is the spindle checkpoint, which acts to ensure that there is proper attachment of chromosomes to the mitotic spindle before progression through the cell cycle can occur. If the spindle apparatus does not properly align the chromosomes for division, if the chromosome number is abnormal (i.e., greater or less than $4n$), or if the centromeres are not properly paired with their duplicated partners, then the cell initiates a cell death pathway to prevent the production of aneuploid progeny (having an altered number of chromosomes). Abnormalities in the spindle checkpoint facilitate the development of aneuploidy, which is frequently found in cancers. In some tumors, aneuploidy is a predominant genetic feature.

In other tumors, a defect in the cells' ability to repair errors in the DNA, such as due to mutations in genes coding for the proteins critical for mismatched DNA repair, is the primary genetic lesion. Cancer cells can have defects in any of several DNA repair pathways in addition to mismatch repair, including deficient interstrand cross-link, double-strand breaks (homologous recombination or nonhomologous end joining repair), single-strand breaks, base excision, nucleotide excision, and translesional synthesis.

In general, tumors have either defects in chromosome number or defective DNA repair pathways but not both. Defects that lead to cancer include abnormal cell cycle checkpoints, inadequate DNA repair, and failure to preserve genome integrity leading to DNA damage. These defects and the stress of the resultant increased DNA damage make cancer cells more vulnerable to additional DNA damage, which can be exploited by chemotherapy, radiation therapy, targeted therapy, and immunotherapy—the major systemic therapeutic approaches effective against cancer.

Alternatively, research is ongoing in an attempt to therapeutically restore the defects in cell cycle regulation and DNA repair that characterize cancer, although this remains a challenging problem because it is much more difficult to restore normal biologic function than to inhibit abnormal function of proteins driving cell proliferation, such as occurs with activated oncogenes. Newer approaches to gene editing (e.g., clustered regularly interspaced short palindromic repeats [CRISPR]) and subsequent modifications to this approach should eventually make gene editing more clinically feasible.

■ CELLULAR SENESCENCE

The irreversible cessation of growth of normal cells while the cells remain viable has been termed cellular senescence. Senescence is important for several processes involved in normal development and homeostasis including embryogenesis and wound healing. It is also an important component of host mechanisms to prevent tumorigenesis by preventing replication of abnormal cells as well as other mechanisms including secreted substances that can stimulate an immune response against the abnormal senescent cell. However, paradoxically, senescent cells in tumors can also stimulate tumorigenesis and malignant progression, in part by other secreted substances that stimulate a harmful inflammatory response. It was initially identified by the fact that when normal cells are placed in culture in vitro, most are not capable of sustained growth. They quickly reach a point where they either undergo cell death due to excessive DNA damage or other factors or they become senescent. Fibroblasts are an exception to this rule. When they are cultured, fibroblasts may divide 30–50 times and then they undergo what has been termed a "crisis" during which the majority of cells stop dividing (usually due to an increase in p21 expression, a CDK inhibitor). This form of senescence is termed replicative senescence. Many other cells die, and a small fraction emerge that have acquired genetic and epigenetic changes that permit their uncontrolled growth. Among the cellular changes during in vitro propagation is telomere shortening. DNA polymerase is unable to replicate the tips of chromosomes, resulting in the loss of DNA at the specialized ends of chromosomes (called telomeres) with each replication cycle. At birth, human telomeres are 15- to 20-kb pairs long and are composed of tandem repeats of a six-nucleotide sequence (TTAGGG) that associate with specialized telomere-binding proteins to form a T-loop structure that protects the ends of chromosomes from being mistakenly recognized as damaged. The loss of telomeric repeats with each cell division cycle causes gradual telomere shortening, leading to growth arrest when one or more critically short telomeres trigger a p53-regulated DNA-damage checkpoint response. Cell death usually ensues when the unprotected ends of chromosomes lead to chromosome fusions or other catastrophic DNA rearrangements. Cells with certain abnormalities, such as those with nonfunctional pRb and p53, can bypass this growth arrest. The ability to bypass telomere-based growth limitations is thought to be a critical step in the evolution of most malignancies. This occurs by reactivation of telomerase expression in cancer cells. Telomerase is an enzyme that adds TTAGGG repeats onto the 3′ ends of chromosomes. It contains a catalytic subunit with reverse transcriptase activity (hTERT) and an RNA component that provides the template for telomere extension. Most normal somatic cells do not express sufficient telomerase to prevent telomere attrition with each cell division. Exceptions include stem cells (such as those found in hematopoietic tissues, gut and skin epithelium, and germ cells) that require extensive cell division to maintain tissue homeostasis. More than 90% of human cancers express high levels of telomerase that prevent telomere shortening to critical levels and allow indefinite cell proliferation. In vitro experiments indicate that inhibition of telomerase activity leads to tumor cell apoptosis. Major efforts are underway to develop methods to inhibit telomerase activity in cancer cells. For example, the protein component of telomerase (hTERT) may act as one of the most widely expressed tumor-associated antigens and can be targeted by vaccine approaches. However, a caveat to targeting telomerase for anticancer treatment is the potential for inhibiting its activity in certain normal cells (such as stem cells) required for maintaining the normal physiologic state.

Although most of the functions of telomerase relate to cell division, it also has several other effects including interfering with the differentiated functions of at least certain stem cells. However, the impact on differentiated function of normal nonstem cells is less clear. The picture is further complicated by the fact that rare genetic defects in the telomerase enzyme seem to cause dyskeratosis congenita (characterized by abnormalities in various rapidly dividing cells in the body including skin, nails, oral mucosa, hair, and bone marrow with increased risk for leukemia and certain other cancers). This can be associated with a number of other abnormalities including pulmonary fibrosis, bone marrow failure (aplastic anemia), or liver fibrosis. However, paradoxically, defects in nutrient absorption in the gastrointestinal tract, a site that should be highly sensitive to defective cell proliferation, are not seen. Much remains to be learned about how telomere shortening and telomere maintenance are related to human illness in general and cancer in particular.

A variety of other stresses on cells (both environmental and intrinsic including radiation, chemotherapy, reactive oxygen species, and oncogenic mutations) can also lead to senescence, primarily those that induce DNA damage similar to that seen in cells with shortened telomeres. This is termed replicative senescence.

■ SIGNAL TRANSDUCTION PATHWAYS IN CANCER CELLS

Signals that affect cell behavior come from adjacent cells, the stroma in which the cells are located, hormonal signals that originate remotely, and the cells themselves (autocrine signaling). These signals generally exert their influence on the receiving cell through activation of signal transduction pathways that have as their end result the induction of activated transcription factors that mediate a change in cell behavior or function or the acquisition of effector machinery to accomplish a new task. Although signal transduction pathways can lead to a wide variety of outcomes, many such pathways rely on cascades of signals that sequentially activate different proteins or glycoproteins and lipids or glycolipids, and the activation steps often involve the addition or removal of one or more phosphate groups on a downstream target.

Other chemical changes can result from signal transduction pathways, but reversible phosphorylation and dephosphorylation play a major role. Proteins that add phosphate groups to other molecules (proteins, lipids, or nucleic acids) are called kinases. Two major classes of kinases involved in signal transduction pathways important for cancer cells are tyrosine kinases that phosphorylate tyrosine and serine/threonine kinases that phosphorylate serine/threonine either directly or indirectly. However, some kinases can phosphorylate both, such as the MEK kinases that can phosphorylate both threonine and tyrosine. Phosphatases (protein tyrosine phosphatases and protein serine/threonine phosphatases) remove the phosphate groups to reverse the kinase activity.

Various kinases play critical roles in signal transduction pathways important for malignant cells. These include a number of receptor tyrosine kinases (RTKs) as well as various protein kinases (both tyrosine and serine/threonine kinases) downstream of receptors that transmit the signals within the cell **(Fig. 77-2)**. Two important signaling pathways are the RAS-RAF-MEK-ERK pathway and the phosphatidylinositol-3-kinase (PI3K) pathway (Fig. 77-2). Although pathways are depicted as distinct, complex interactions between pathways occur within cells.

Normally, kinase activity is short-lived and reversed by protein phosphatases. However, in many human cancers, RTKs or components of their downstream pathways are activated by mutation, gene amplification, or chromosomal translocations to have enhanced and/or prolonged activity. Because these pathways are important in regulating proliferation, survival, migration, and angiogenesis, they have been identified as important targets for cancer therapeutics.

Inhibition of kinase activity is effective in the treatment of a number of neoplasms. Lung cancers with mutations in the epidermal growth factor receptor are highly responsive to osimertinib as well as other inhibitors **(Table 77-2)**. Inhibitors have been developed to treat lung cancers with other tyrosine kinase–activating mutations (including anaplastic lymphoma kinase [ALK], ROS1, NTRK, MET, HER2, and RET). BRAF (a serine/threonine kinase) inhibitors are highly effective in melanomas and thyroid cancers and are also used in combination with other agents for lung and colon cancers as well as other solid tumors with *BRAF* V600E mutations. Targeting the MEK protein (which phosphorylates both threonine and tyrosine residues) downstream of BRAF also has activity against BRAF mutant melanomas, and combined inhibition of BRAF and MEK is more effective than either alone with activity that extends to BRAF-mutant lung cancer. Janus kinase (JAK) inhibitors are active in myeloproliferative syndromes in which JAK2 activation is a pathogenetic event. Imatinib (which targets a number of tyrosine kinases) is an effective agent in

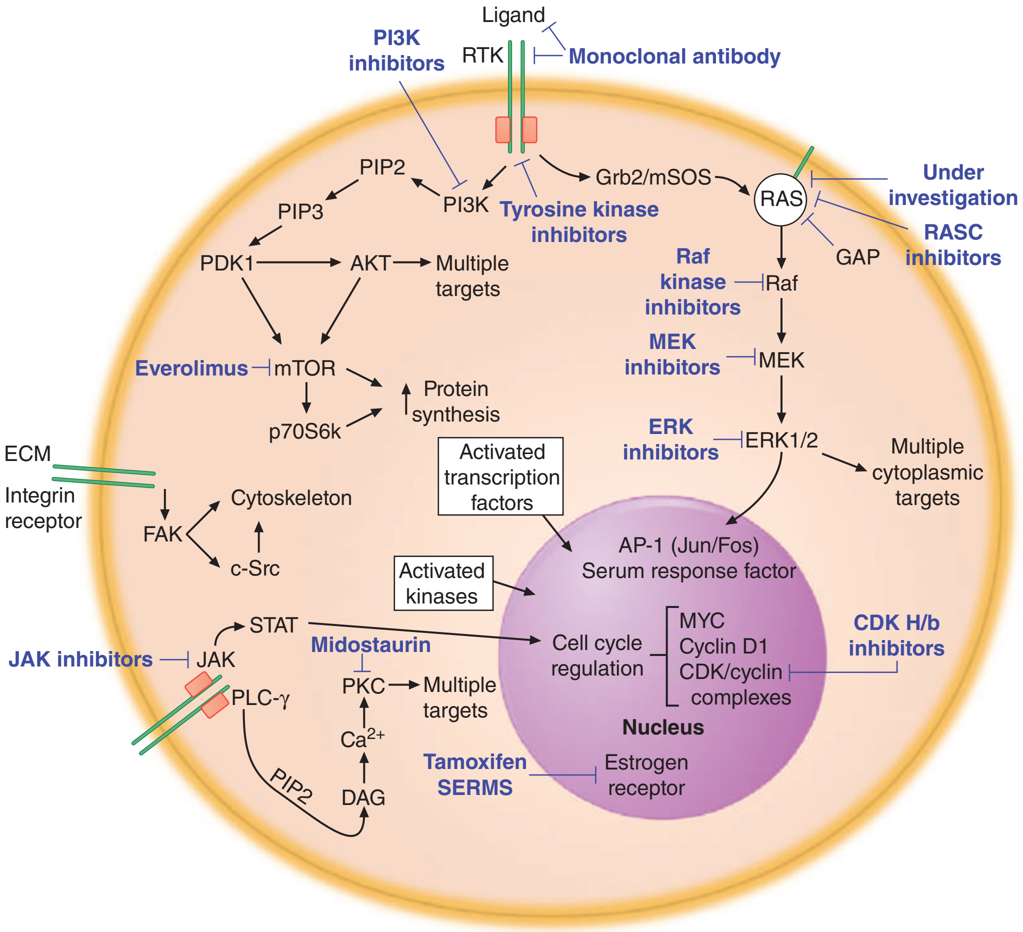
Ligand
RTK
Monoclonal antibody
PI3K inhibitors
PIP2
PIP3
PI3K
Tyrosine kinase inhibitors
Grb2/mSOS
RAS
Under investigation
RASC inhibitors
GAP
PDK1
AKT
Multiple targets
Raf kinase inhibitors
Raf
MEK inhibitors
MEK
Everolimus
mTOR
Protein synthesis
p70S6k
ERK inhibitors
ERK1/2
Multiple cytoplasmic targets
ECM
Integrin receptor
FAK
Cytoskeleton
c-Src
Activated transcription factors
AP-1 (Jun/Fos)
Serum response factor
Activated kinases
STAT
JAK inhibitors
JAK
Midostaurin
Cell cycle regulation
MYC
Cyclin D1
CDK/cyclin complexes
CDK H/b inhibitors
PLC-γ
PKC
Multiple targets
Ca2+
Nucleus
Tamoxifen SERMS
Estrogen receptor
PIP2
DAG

FIGURE 77-2 Therapeutic targeting of signal transduction pathways in cancer cells. Three major signal transduction pathways are activated by receptor tyrosine kinases (RTKs). *1.* The protooncogene Ras is activated by the Grb2/mSOS guanine nucleotide exchange factor, which induces an association with Raf and activation of downstream kinases (MEK and ERK1/2). *2.* Activated PI3K phosphorylates the membrane lipid PIP$_2$ to generate PIP$_3$, which acts as a membrane-docking site for a number of cellular proteins including the serine/threonine kinases PDK1 and Akt. PDK1 has numerous cellular targets, including Akt and mTOR. Akt phosphorylates target proteins that promote resistance to apoptosis and enhance cell cycle progression, while mTOR and its target p70S6K upregulate protein synthesis to potentiate cell growth. *3.* Activation of PLC-γ leads the formation of diacylglycerol (DAG) and increased intracellular calcium, with activation of multiple isoforms of PKC and other enzymes regulated by the calcium/calmodulin system. Other important signaling pathways involve non-RTKs that are activated by cytokine or integrin receptors. Janus kinases (JAK) phosphorylate STAT (signal transducer and activator of transcription) transcription factors, which translocate to the nucleus and activate target genes. Integrin receptors mediate cellular interactions with the extracellular matrix (ECM), inducing activation of FAK (focal adhesion kinase) and c-Src, which activate multiple downstream pathways, including modulation of the cell cytoskeleton. Many activated kinases and transcription factors migrate into the nucleus, where they regulate gene transcription, thus completing the path from extracellular signals, such as growth factors, to a change in cell phenotype, such as induction of differentiation or cell proliferation. The nuclear targets of these processes include transcription factors (e.g., Myc, AP-1, and serum response factor) and the cell cycle machinery (cyclin-dependent kinases [CDKs] and cyclins). Inhibitors of many of these pathways have been developed for the treatment of human cancers. Examples of inhibitors that are either approved or are currently being evaluated in clinical trials are shown in purple type.

TABLE 77-2 Some FDA-Approved Molecularly Targeted Agents for the Treatment of Cancer

DRUG	MOLECULAR TARGET	DISEASE	MECHANISM OF ACTION
All-*trans* retinoic acid	PML-RARα oncogene	Acute promyelocytic leukemia M3 AML, t(15;17)	Inhibits transcriptional repression by PML-RARα Adagrasib, Solorasib/KRAS12C/NSCLC/Inhibits KRAS12C
Imatinib, dasatinib, nilotinib, ponatinib, bosutinib	Bcr-Abl, c-Abl, c-Kit, PDGFR-α/β	Chronic myeloid leukemia, GIST	Blocks ATP binding to tyrosine kinase active site
Ripretinib	c-Kit, PDGFR-α	GIST	Inhibits tyrosine kinase activity
Asciminib	Bcr-Abl	Chronic myeloid leukemia	Allosteric inhibitor of BCR-ABL
Sunitinib	c-Kit, VEGFR-2, PDGFR-β, Flt-3	GIST, RCC, PNET	Inhibits activated c-Kit and PDGFR in GIST; inhibits VEGFR in RCC and probably in PNET
Sorafenib	RAF, VEGFR-2, PDGFR-α/β, Flt-3, c-Kit	RCC, hepatocellular carcinoma (HCC), differentiated thyroid cancer, desmoid	Targets VEGFR pathways in RCC and HCC. Possible activity against BRAF in thyroid cancer
Regorafenib	VEGFR1–3, TIE-2, FGFR1, KIT, RET, PDGFR	Colorectal cancer, GIST, HCC	Competitive inhibitor ATP binding site of tyrosine kinase domain multiple kinases including VEGFR
Larotrectinib, entrectinib	NTRK	Cancers with NTRK mutation	Competitive inhibitor of ATP binding site of the tyrosine kinase domain of NTRK
Axitinib	VEGFR1–3	RCC	Competitive inhibitor ATP binding site of tyrosine kinase domain VEGF receptors
Erlotinib	EGFR	NSCLC, pancreatic cancer	Competitive inhibitor of the ATP-binding site of the EGFR
Afitinib	EGFR (and other HER family)	NSCLC	Irreversible inhibitor of ATP-binding site of HER family members
Osimertinib	EGFR (T790M)	NSCLC	Inhibits EGFR mutations including T790M mutant NSCLC
Dacomitinib	EGFR	NSCLC (exon19 deletion/exon 21 L858R)	Inhibits EGFR mutant lung cancer Mobocertinib/EGFR/NSCLC/Tumors with Exon20 insertion mutations
Erdafitinib, pemigatinib, futibatinib, infigratinib	FGFR2, FGFR3	Urothelial (erdafitinib), myeloid/lymphoid neoplasms (pemigatinib) cholangiocarcinoma (pemigatinib, futibatinib)	Inhibits tyrosine kinase of FGFR
Lapatinib, tucatinib, niratinib	HER2/neu	Breast cancer, CRC (tucatinib + trastuzumab)	Competitive inhibitor of the ATP-binding site of HER2
Crizotinib, ceritinib, alectinib, brigatinib, lorlatinib	ALK	NSCLC ALK+ large cell lymphoma, inflammatory myofibroblastic tumors (crizotinib)	Inhibitor of ALK tyrosine kinase
Crizotinib, entrectinib repotrectinib	ROS1	NSCLC	Inhibitor of ROS1 tyrosine kinase
Palbociclib, ribociclib, abemaciclib	CDK4/6	Breast	Inhibitor of CDK4/6
Bortezomib, carfilzomib, ixazomib	Proteasome	Multiple myeloma	Inhibits proteolytic degradation of multiple cellular proteins
Vemurafenib, dabrafenib Encorafenib	BRAF V600E	Melanoma lung cancer, CRC (combined with Cetuximab)	Inhibitor of serine-threonine kinase domain of V600E mutant of BRAF
Trametinib, Cobimetinib, binimetinib	MEK	Melanoma	Inhibitor of serine-threonine kinase domain of MEK
Cabozantinib	RET, MET, VEGFR	MTC, RCC	Competitive inhibitor of ATP-binding site of tyrosine kinase domain of multiple kinases, including VEGFR2 and RET
Capmatinib, tepotinib	MET	NSCLC with MET exon14 deletions	
Selpercatinib, vandetinib, pralsetinib	RET	NSCLC, MTC, RET fusion thyroid cancer, RET fusion positive solid tumors	Inhibitor of RET, VEGFR1, VEGFR2 tyrosine kinases
Temsirolimus	mTOR	RCC	Competitive inhibitor of mTOR serine-threonine kinase
Everolimus	mTOR	RCC, PNET	Binds to immunophilin FK binding protein-12, which forms a complex that inhibits mTOR kinase
Vorinostat, romidepsin, belinostat	HDAC	CTCL/PTL	HDAC inhibitor, epigenetic modulation
Panobinostat	HDAC	MM	HDAC inhibitor, epigenetic modulation
Ruxolitinib	JAK-1, 2	Myelofibrosis	Competitive inhibitor of tyrosine kinase
Vismodegib	Hedgehog pathway	Basel cell cancer (skin)	Inhibits smoothened in hedgehog pathway
Lenvatinib	Multikinase inhibitor (VEGFR, FGFR, PGFR-α, others)	RCC, thyroid cancer, HCC	Competitive inhibitor of ATP-binding site of tyrosine kinase domain of multiple kinases
Olaparib, rucaparib, niraparib, talazoparib	PARP	BRCA mutant ovarian, breast, prostate, pancreas cancers; not all agents approved for all cancers	Inhibits PARP and DNA repair
Venetoclax	BCL-2	CLL (with 17p deletion)	Inhibits BCL-2 and enhances apoptosis
Ibrutinib, acalabrutinib pirtobrutinib, zanubrutinib	Bruton tyrosine kinase (BTK)	CLL, MCL, MZL, SLL, WM	Inhibitor of BTK

(*Continued*)

TABLE 77-2 Some FDA-Approved Molecularly Targeted Agents for the Treatment of Cancer (Continued)

DRUG	MOLECULAR TARGET	DISEASE	MECHANISM OF ACTION
Ivosidenib, olutasidenib	IDH1	AML, MDS, cholangiocarcinoma	IDH1 inhibitor
Gilteritinib, quizartinib	FLT3	AML	FLT3 inhibitor
Idelalisib	PI3K-delta	CLL	Inhibits PI3k-delta, preventing proliferation and inducing apoptosis
Alpelisib	PIK3CA	Breast cancer with a PIK3CA mutation	Inhibits PIK3CA
Belzutifan	Hif-2α	HIF-1α-associated RCC, pancreatic neuroendocrine, CNS hemangioblastoma	Inhibits Hif-2α
Capivasertib	AKT	Breast cancer	Inhibits AKT
Umbralisib	PI3K-delta, CK1-epsilon	MZL, FL	Inhibits PI3K-delta and CK1-epsilon
Selinexor	Exportin-1	MM, DLBCL	Induces apoptosis of cells
Tazemetostat	EZH2	FL, epithelioid sarcoma	Inhibits EZH2
Monoclonal Antibodies			
Trastuzumab, margetuximab	HER2/neu (ERBB2)	Breast cancer, gastric or GEJ cancer	Binds HER2 on tumor cell surface and induces receptor internalization
Pertuzumab	HER2/neu (ERBB2)	Breast cancer	Binds HER2 on tumor cell surface at distinct site from trastuzumab and prevents binding to other receptors
Cetuximab	EGFR	Colon cancer, squamous cell carcinoma of the head and neck	Binds extracellular domain of EGFR and blocks binding of EGF and TGF-α; induces receptor internalization. Potentiates the efficacy of chemotherapy and radiotherapy
Panitumumab	EGFR	Colon cancer	Similar to cetuximab but fully humanized rather than chimeric
Necitumumab	EGFR	Squamous NSCLC	Binds EGFR
Rituximab	CD20	B-cell lymphomas and leukemias that express CD20	Multiple potential mechanisms, including direct induction of tumor cell apoptosis and immune mechanisms
Alemtuzumab	CD52	Chronic lymphocytic leukemia and CD52-expressing lymphoid tumors	Immune mechanisms
Bevacizumab	VEGF	Colorectal, lung cancers, RCC, glioblastoma	Inhibits angiogenesis by high-affinity binding to VEGF
Ziv-aflibercept	VEGFA, VEGFB, PLGF	Colorectal cancers	Inhibits angiogenesis by high-affinity binding to VEGFA, VEGFB, and PLGF
Ramucirumab	VEGFR	Gastric, colorectal, lung cancers	Inhibits angiogenesis by binding to VEGFR
Ipilimumab	CTLA-4	Melanoma, HCC, MSI-high colorectal cancer	Blocks CTLA-4, preventing interaction with CD80/86 and T-cell inhibition
Nivolumab, pembrolizumab, dostarlimab-gxly, toripalimab, retifanlimab-dlwr, cemiplimab-rwlc	PD-1	Melanoma, head and neck cancer, NSCLC, SCLC, Hodgkin's disease, urothelial cancer, RCC, HCC, gastric cancer, esophageal cancer, cholangiocarcinoma, MSI-high cancers, endometrial cancer, cervical cancer, cutaneous squamous cell carcinoma, basal cell carcinoma, breast cancer, nasopharyngeal cancer, Merkel cell tumor	Blocks PD-1, preventing interaction with PD-L1 and T-cell inhibition
Atezolizumab, durvalumab, avelumab	PD-L1	NSCLC, urothelial cancer, SCLC (durvalumab), HCC (atezolizumab), Merkel cell cancer (avelumab)	Blocks PD-L1, preventing interaction with PD-1 and T-cell inhibition
Relatlimab	LAG3	Melanoma (combined with nivolumab)	Blocks LAG3 interaction with MHCII and other ligands inhibiting immune activation
Denosumab	Rank ligand	Breast, prostate	Inhibits Rank ligand, primary signal for bone removal
Dinutuximab	Glycolipid GD2	Neuroblastoma (pediatric)	Immune-mediated attack on GD2-expressing cells
Daratumumab, Isatuximab	CD38	MM	Binds to CD38 on MM cells causing apoptosis by antibody-dependent or compliment-mediated cytotoxicity
Elotuzumab	SLAMF7	MM	Activating NK cells to kill MM cells
Olaratumab	PDGFRα	Soft tissue sarcomas	Blocks PDGFRα activity
Naxitamab	GD2	Neuroblastoma	Immune-mediated antitumor effect
Bispecific Antibodies			
Blinatumomab	CD19 and CD3	Ph-relapsed precursor B-cell ALL	Binds CD19 on ALL cells and CD3 on T cells; immune attack on CD19-expressing cells
Glofitamab-gxbm, epcoritamab-bysp, mosunetuzumab-axgb	CD20 and CD3	DLBCL, FL	Binds CD20 on DLBCL or FL and CD3 on T cells, immune attack on CD20-expressing cells
Teclistamab-cqyv, elranatamab-bcmm	B-cell maturation antigen (BCMA) and CD3	MM	Binds BCMA on MM cells and CD3 on T cells
Talquetamab	CD3 and GPRC5D	MM	Binds CD3 T cells and GPRC5D-expressing MM cells

(*Continued*)

TABLE 77-2 Some FDA-Approved Molecularly Targeted Agents for the Treatment of Cancer (Continued)

DRUG	MOLECULAR TARGET	DISEASE	MECHANISM OF ACTION
Amivantamab-vmjw	EGFR and MET	NSCLC	Targets EGFR exon 20 insertion mutations by also inhibiting MET
Tebentafusp-tebn	GP100 and CD3	Uveal melanoma	Binds GP100 on melanoma cells and CD3 on T cells
Antibody-Chemotherapy Conjugates			
Brentuximab vedotin	CD30	Hodgkin's disease, anaplastic lymphoma	Delivers chemotherapeutic agent MMAE to CD30-expressing tumor cells
Ado-trastuzumab emtansine	HER2	Breast cancer	Delivers chemotherapeutic agent emtansine to HER2-expressing breast cancer cells
Fam-trastuzumab	HER2	Breast, NSCLC, and gastric cancers	Delivers chemotherapeutic agent deruxtecan to HER2-expressing breast cancer cells
Sacituzumab govitecan	Trop2	Breast, urothelial cancers	Delivers chemotherapy to Trop2-expressing cells
Enfortumab-vedotin	Nectin-4	Urothelial cancers	Delivers chemotherapeutic agent MMAE to Nectin-4-expressing cells
Polatuzumab-vedotin	CD79b	DLBCL or high-grade BCL	Delivers MMAE chemotherapy to B-cell lymphomas
Loncastuximab tesirine-lpyl	CD19	DLBCL	Delivers chemotherapy to CD19 expressing cells
Mirvetuximab soravtansine-gynx	Folate receptor alpha	Ovarian, fallopian, peritoneal cancers	Delivers chemotherapy to folate receptor alpha tumors
Tisotumab vedotin-tftv	Tissue factor (TF)	Cervical cancer	Delivers chemotherapy to TF-positive cells
Gemtuzumab ozogamicin	CD33	Pediatric CD33+ AML	Delivery of chemotherapy to CD33+ cells
CAR-T Cells and Tumor-Infiltrating Lymphocyte (TIL)			
Tisagenlecleucel, axicabtagene ciloleucel, brexucabtagene autoleucel, lisocabtagene maraleucel	CD19	ALL (tisagenlecleucel), DLBCL/high-grade BCL (axicabtagene ciloleucel), B-cell precursor ALL (brexucabtagene), large BCL (lisocabtagene maraleucel)	Targeted T cells to protein on surface of malignant cells
Ciltacabtagene autoleucel Idecabtagene vicleucel	BCMA	MM	Targets T cells to protein on surface of MM cells
Lifileucel	Melanoma antigens	Melanoma	Tumor-infiltrating lymphocyte therapy

Abbreviations: ALL, acute lymphocytic leukemia; AML, acute myeloid leukemia; BCL, B-cell lymphoma; CAR-T, chimeric antigen receptor T cells; CLL, chronic lymphocytic leukemia; CRC, colorectal cancer; CTCL, cutaneous T cell lymphoma; DLBCL, diffuse large B-cell lymphoma; EGFR, epidermal growth factor receptor; FDA, U.S. Food and Drug Administration; FGFR, fibroblast growth factor receptor; FL, follicular lymphoma; Flt-3, fms-like tyrosine kinase-3; GEJ, gastroesophageal junction; GIST, gastrointestinal stromal tumor; HDAC, histone deacetylases; MCL, mantle cell lymphoma; MM, multiple myeloma; MSI, microsatellite instability; MMAE, monomethyl auristatin E; MTC, medullary thyroid cancer; mTOR, mammalian target of rapamycin; MZL, mantle zone lymphoma; NK, natural killer; NSCLC, non-small-cell lung cancer; PARP, poly-ADP ribose polymerase; PDGFR, platelet-derived growth factor receptor; PLGF, placenta growth factor; PML-RARα, promyelocytic leukemia–retinoic acid receptor-alpha; PNET, pancreatic neuroendocrine tumors; PTL, peripheral T-cell lymphoma; RCC, renal cell cancer; t(15;17), translocation between chromosomes 15 and 17; SCLC, small-cell lung cancer; SLL, small lymphocytic lymphoma; TGF-α, transforming growth factor-alpha; VEGFR, vascular endothelial growth factor receptor; WM, Waldenström's macroglobulinemia.

Note: The pace of drug discovery is rapid and this list is not comprehensive.

tumors that have translocations of the c-Abl and BCR gene (such as chronic myeloid leukemia), mutant c-Kit (gastrointestinal stromal cell tumors), or mutant platelet-derived growth factor receptor (PDGFRα; gastrointestinal stromal tumors). Second-generation inhibitors of BCR-Abl, dasatinib and nilotinib, are even more effective, and the third-generation agent bosutinib has activity in some patients who have progressed on other inhibitors, while the third-generation agent ponatinib has activity against the T315I mutation, which is resistant to the other agents. Although almost all tyrosine kinase inhibitors are not entirely selective for one protein, certain inhibitors have significant activity against a broad number of proteins. These include sorafenib, regorafenib, cabozantinib, sunitinib, and lenvatinib. These have shown antitumor activity in various malignancies, including renal cell cancer (RCC) (sorafenib, sunitinib, cabozantinib, lenvatinib), hepatocellular carcinoma (sorafenib, regorafenib, lenvatinib), gastrointestinal stromal tumor (GIST) (sunitinib, regorafenib), thyroid cancer (sorafenib, cabozantinib, lenvatinib), colorectal cancer (regorafenib), and pancreatic neuroendocrine tumors (sunitinib).

Inhibitors of the PI3K pathway also have been approved for cancer therapy. The PI3K family includes three classes and several isoforms within each class. Inhibitors against different isoforms have proved effective against different types of malignancies, with inhibitors of the delta isoform (either specifically or also with inhibition of other isoforms; e.g., idelalisib) having activity against lymphoid malignancies (CLL), whereas the specific inhibitor of a mutation in the alpha isoform (alpelisib) has activity against breast cancers with this mutation. Inhibitors of mammalian target of rapamycin (mTOR; which is downstream of PI3K; e.g., everolimus, temsirolimus) are active in RCC; well-differentiated nonfunctional neuroendocrine tumors of pancreatic, gastrointestinal, or lung origin; and breast cancer. Additional inhibitors of the PI3K pathway and other phospholipid signaling pathways such as the phospholipase C-gamma pathway, which are involved in a large number of cellular processes important in cancer development and progression, are being evaluated.

The list of active agents and treatment indications is growing rapidly (Table 77-2). These agents have ushered in a new era of personalized therapy. For many cancers, it is now standard for tumor biopsies to be assessed for specific molecular changes that predict response and to have clinical decision-making guided by those results. This is now an important component of choosing therapy for metastatic lung, gastroesophageal, melanoma, breast, and colorectal cancers as well as in adjuvant therapy for breast cancer. This list will continue to evolve as both new agents against existing targets are developed and new targets are discovered.

An alternative approach to testing samples directly from tumors is to test blood for the presence of mutations or amplification in circulating tumor DNA, which has the significant advantage of being noninvasive. As cancers grow, some of the cells die, break apart, and release cellular contents, including DNA, into the circulation. Sensitive methods have been developed to detect this DNA and to identify mutations and other DNA changes in the malignant cells. This has the potential advantage over tumor biopsies of sampling all of the tumor and not being limited to one site that may not be representative of the overall tumor heterogeneity. Distinct metastatic lesions may have different genetic abnormalities that will not be detected in a biopsy of a single site. In addition to identifying potential changes that can be targeted for therapy, there is also the potential for monitoring a patient's response to therapy, identifying resistance mechanisms to therapy earlier, detecting disease recurrence before it can be detected by tumor markers or scans, monitoring bodily fluids in addition to blood, and possibly providing a means of earlier initial detection of cancer if sufficiently sensitive and specific detection methods can be developed. Optimizing the sensitivity and specificity of these tests is essential for their potential utility in patient care. Research is ongoing to determine if other cellular components specific to cancer cells (e.g., mRNA, proteins from mutant genes, or other protein modifications found in cancer cells) might also be useful for diagnosis or monitoring therapeutic response.

However, none of these targeted therapies has yet been curative by themselves for any malignancy, although prolonged periods of disease control lasting many years frequently occur in chronic myeloid leukemia (CML), including a >80% survival rate at 10 years, and antibodies to HER2 have increased survival for breast cancer patients in combination with chemotherapy. The reasons for the failure of targeted agents to cure are not completely defined, although resistance to the treatment ultimately develops in most patients. In some tumors, resistance to kinase inhibitors is related to proliferation of cells with a mutation in the target kinase that inhibits drug binding. Many of these kinase inhibitors act as competitive inhibitors of the ATP-binding pocket. ATP is the phosphate donor in these phosphorylation reactions. For example, mutation in the critical BCR-ABL kinase in the ATP-binding pocket (such as the threonine to isoleucine change at codon 315 [T315I]) can prevent imatinib binding. Other resistance mechanisms include alterations in other signal transduction pathways to bypass the inhibited pathway. As resistance mechanisms continue to be better defined, rational strategies to overcome resistance are emerging. In addition, many kinase inhibitors are less specific for an oncogenic target than was hoped, and toxicities related to off-target inhibition of kinases limit the use of the agent at a dose that would optimally inhibit the cancer-relevant kinase.

Antibodies against protein targets more highly expressed on malignant than normal cells can also be used to deliver highly toxic compounds relatively specifically to cancer cells. Examples of protein targets for currently approved antibody-drug conjugates include CD30 for Hodgkin's and anaplastic lymphomas; HER2 on breast cancer; CD33 on acute myeloid leukemias; CD22 on B-cell acute lymphocytic and hairy cell leukemias; and CD79b on diffuse large B-cell lymphomas.

Another strategy to enhance the antitumor effects of targeted agents is to use them in rational combinations with each other as well as with chemotherapy or immunotherapy agents that kill cells in ways distinct from agents targeting specific mutant or overexpressed proteins. Combinations of trastuzumab (a monoclonal antibody that targets the HER2 receptor [member of the EGFR family]) with chemotherapy have significant activity against breast, gastric, and esophageal cancers that have high levels of expression of the HER2 protein. The activity of trastuzumab and chemotherapy can be enhanced further by combinations with another targeted monoclonal antibody (pertuzumab), which prevents dimerization of the HER2 receptor with other HER family members including HER3, or in some cases with immunotherapy, such as combinations of chemotherapy, trastuzumab, and the immunotherapeutic agent pembrolizumab against HER2-positive gastric and esophageal cancers.

Although targeted therapies have not yet resulted in cures when used alone, their use in the adjuvant setting and when combined with other effective treatments has substantially increased the fraction of patients cured. For example, the addition of rituximab, an anti-CD20 antibody, to combination chemotherapy in patients with diffuse large B-cell lymphoma improves cure rates by ~15%. The addition of trastuzumab, an antibody to HER2, to combination chemotherapy in the adjuvant treatment of HER2-positive breast cancer significantly improves overall survival.

A major effort continues to develop targeted therapies for mutations in the ras family of genes, which play a critical role in transmitting signals through a number of downstream signaling pathways including the MAP (mitogen-activated protein) kinase and PI3K pathways. Mutations in ras are the most common mutations in oncogenes in cancers (especially kras) but have proved to be very difficult targets for a number of reasons related to the structure of RAS proteins as well as mechanisms of activation and inactivation (active when bound to guanosine triphosphate [GTP] and inactive when bound to guanosine diphosphate [GDP]). RAS proteins are not kinases but bind directly to the BRAF serine/threonine kinase with preferential binding when RAS is in the active GTP bound state. Agents that target one of the mutant forms of KRAS (12C), which is the most common RAS mutation in lung cancer and is also found in a subset of other cancers, have sufficient antitumor activity to now be approved for the treatment of these lung cancers. They are under active study (often in combination with other agents) for the treatment of other cancers that have the KRAS12C mutation. Agents targeting other mutations in the RAS genes (especially other KRAS-mutant proteins) are also being evaluated in clinical trials.

One strategy for new drug development is to take advantage of so-called oncogene addiction. This situation **(Fig. 77-3)** is created when a tumor cell develops an activating mutation in an oncogene that becomes a dominant pathway for survival and growth with reduced contributions from other pathways, even when there may be abnormalities in those pathways. This dependency on a single pathway creates a cell that is vulnerable to inhibitors of that oncogene pathway. For example, cells harboring mutations in BRAF are sensitive to MEK inhibitors that inhibit signaling via the BRAF pathway.

Proteins critical for transcription of other proteins essential for malignant cell survival or proliferation provide another potential target for treating cancers. The transcription factor nuclear factor (NF)-κB is a heterodimer composed of p65 and p50 subunits that associate with an inhibitor, IκB, in the cell cytoplasm. In response to growth factor or cytokine signaling, a multisubunit kinase called IKK (IκB-kinase) phosphorylates IκB and directs its degradation by the ubiquitin/proteasome system. NF-κB, free of its inhibitor, translocates to the nucleus and activates target genes, many of which promote the survival of tumor cells. One of the mechanisms by which novel drugs called proteasome inhibitors are thought to produce an anticancer effect is by blocking the proteolysis of IκB, thereby preventing NF-κB activation. For reasons that have not been fully elucidated, this has a differential toxicity effect on tumor, as compared to normal cells. Although this mechanism appears to be an important aspect of the antitumor effects of proteasome inhibitors, other effects involving the inhibition of the degradation of multiple cellular proteins important in malignant cell survival or proliferation also play a role.

Proteasome inhibitors (e.g., bortezomib, carfilzomib, ixazomib) have activity in patients with multiple myeloma, including partial and complete remissions. Inhibitors of IKK are also in development, with the hope of more selectively blocking the degradation of IκB, thus "locking" NF-κB in an inhibitory complex and rendering the cancer cell more susceptible to apoptosis-inducing agents. Many other transcription factors are activated by phosphorylation, which can be prevented by tyrosine or serine/threonine kinase inhibitors, a number of which are currently in clinical trials.

Estrogen receptors (ERs) and androgen receptors (ARs), members of the steroid hormone family of nuclear receptors, are targets of inhibition by drugs used to treat breast and prostate cancers, respectively. Selective estrogen receptor modulators (SERMs) have been developed as a treatment approach for ER-positive breast cancer. Tamoxifen, a partial agonist and antagonist of ER function, is frequently used in breast cancer, can mediate tumor regression in metastatic breast cancer, and can prevent disease recurrence in the adjuvant setting. Tamoxifen binds to the ER and modulates its transcriptional activity, inhibiting activity in the breast but promoting activity in bone but unfortunately also in uterine epithelium, leading to a small increased risk of uterine cancer. Attempts have been made to develop SERMs that would have antiestrogenic effects in both breast and uterus while maintaining protective effects on bone. However, none of these to date has been an improvement over tamoxifen. Aromatase inhibitors, which block the conversion of androgens to estrogens in breast and subcutaneous fat tissues, have demonstrated improved clinical efficacy compared with tamoxifen in postmenopausal women and are often used as first-line therapy in postmenopausal patients with ER-positive disease. They are occasionally used in premenopausal patients with ER-positive disease

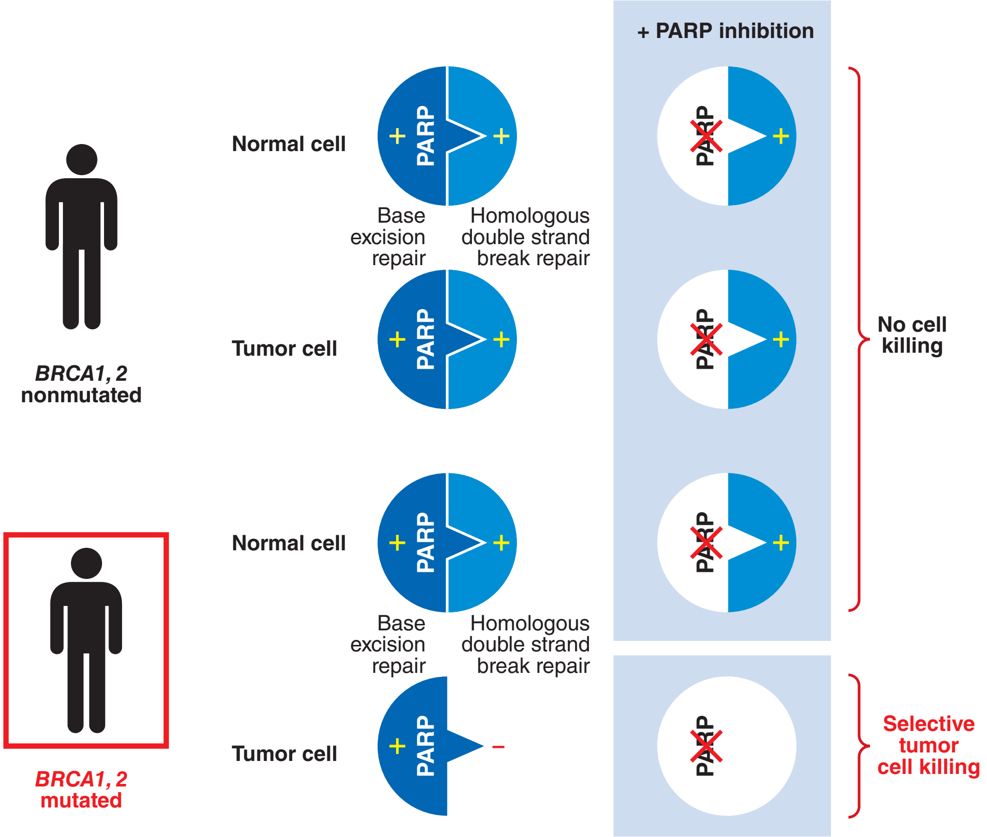

FIGURE 77-3 Synthetic lethality. Genes are said to have a synthetic lethal relationship when mutation of either gene alone is tolerated by the cell, but mutation of both genes leads to lethality, as originally noted by Bridges and later named by Dobzhansky. Thus, mutant *gene a* and *gene b* have a synthetic lethal relationship, implying that the loss of one gene makes the cell dependent on the function of the other gene. In cancer cells, loss of function of a DNA repair gene like *BRCA1*, which repairs double-strand breaks, makes the cell dependent on base excision repair mediated in part by *PARP.* If the PARP gene product is inhibited, the cell attempts to repair the break using the error-prone nonhomologous end-joining method, which results in tumor cell death. High-throughput screens can now be performed using isogenic cell line pairs in which one cell line has a defined defect in a DNA repair pathway. Compounds can be identified that selectively kill the mutant cell line; targets of these compounds have a synthetic lethal relationship to the repair pathway and are potentially important targets for future therapeutics.

in combination with ovarian suppression approaches such as luteinizing hormone–releasing hormone (LHRH) agonists.

A number of approaches have been developed for blocking androgen stimulation of prostate cancer, including decreasing production by the testicles (e.g., orchiectomy, LHRH agonists or antagonists), directly blocking actions of androgen (a number of agents have been developed to do this), or blocking production by inhibiting the enzyme CYP17, which is central in production of androgens from cholesterol.

CANCER-SPECIFIC GENETIC CHANGES AND SYNTHETIC LETHALITY

The concepts of oncogene addiction and synthetic lethality have spurred new drug development targeting oncogene- and tumor-suppressor pathways. As discussed earlier in this chapter and outlined in Fig. 77-3, cancer cells can become dependent upon signaling pathways containing activated oncogenes; this can effect proliferation (i.e., mutated KRAS, BRAF, overexpressed MYC, or activated tyrosine kinases). Additional genetic changes in malignant cells or unique features of tumors including defects in DNA repair (e.g., loss of *BRCA1* or *BRCA2* gene function), modifications in cell cycle control (e.g., changes in protein levels or mutations in cyclins and CDKs), enhanced survival mechanisms (overexpression of Bcl-2 or NF-κB), altered cell metabolism (such as occurs when mutant KRAS enhances glucose uptake and aerobic glycolysis), tumor-stromal interactions, and angiogenesis (e.g., production of vascular endothelial growth factor [VEGF] in response to HIF-2α in RCC) can also be successfully exploited to relatively specifically target cancers. However, resistance to inhibition of specific oncogenic pathways almost always eventually develops. In addition, targeting defects in tumor-suppressor genes has been much more difficult, both because the target of mutation is often deleted and because it is much more difficult to restore normal function than to inhibit abnormal function of a protein. Synthetic lethality occurs when loss of function in either of two or more genes individually has limited effects on cell survival but loss of function in both (or more) genes leads to cell death. In the case of oncogene-addicted pathways, identifying genes that have a synthetic lethal relationship with the activated pathway may allow enhanced cell killing and decreased resistance by targeting those genes or their proteins. In the case of mutant tumor-suppressor genes, identifying genes that have a synthetic lethal relationship to those mutated pathways may allow targeting by inhibiting proteins required uniquely by those cells for survival or proliferation (Fig. 77-3). This is a much more tractable approach than attempting to repair normal function of the mutant suppressor gene itself. Examples of synthetic lethality with clinical impact have been identified. For instance, cells with mutations in the *BRCA1* or *BRCA2* tumor-suppressor genes (e.g., a subset of breast and ovarian cancers) are unable to repair DNA damage by homologous recombination. Poly-ADP ribose polymerase (PARP) is a family of proteins important for single-strand break (SSB) DNA repair. PARP inhibition results in selective killing of cancer cells that have lost *BRCA1* or *BRCA2* function. A number of PARP inhibitors have been approved for treatment of ovarian, breast, and pancreatic cancers with BRCA mutations and are likely to have activity in other tumors with defective DNA repair mechanisms. The concept of synthetic lethality provides a framework for genetic screens to identify other synthetic lethal combinations involving known tumor-suppressor genes and development of novel therapeutic agents to target dependent pathways. Other unique aspects of malignant tumors, including those outlined elsewhere in the chapter, may also be vulnerable to synthetic lethal interactions.

EPIGENETIC INFLUENCES ON CANCER GENE TRANSCRIPTION

Chromatin structure regulates the hierarchical order of sequential gene transcription that governs differentiation and tissue homeostasis. Disruption of chromatin remodeling (the process of modifying chromatin structure to control exposure of specific genes to transcriptional proteins, thereby controlling the expression of those genes) leads to aberrant gene expression that can significantly alter the biology of cells including inducing proliferation or migration of cells. Epigenetic changes are those that alter the pattern of gene expression that persist across at least one cell division but are not caused by changes in the DNA code. These include alterations of chromatin structure mediated by methylation of cytosine residues of DNA (primarily in context of CpG dinucleotides in somatic cells), modification of histones by altering acetylation or methylation, or changes in higher-order chromosome structure **(Fig. 77-4)**. Appropriate control of DNA methylation is essential for normal cell function and development, and both altered methylation and hypomethylation of histones occur in cancers. Hypermethylation of DNA promoter regions is a common mechanism by which tumor-suppressor loci are epigenetically silenced in cancer cells. Thus, one allele of a tumor-suppressor gene may be inactivated by mutation or deletion, while expression of the other allele is epigenetically silenced, usually by methylation, leading to loss of gene function. Aberrant hypomethylation is also frequently found in a number of cancers consistent with the dysregulated pattern of gene transcription that is a hallmark of cancer cells, with some genes being inappropriately turned off while others are inappropriately turned on. Specific changes in DNA methylation in cancer cells provide a potentially more sensitive and specific approach to utilizing circulating tumor DNA to identify the presence of cancer than utilizing only DNA mutational analysis.

Acetylation of the amino terminus of the core histones H3 and H4 induces an open chromatin conformation that promotes transcription initiation. Histone acetylases are components of coactivator complexes recruited to promoter/enhancer regions by sequence-specific transcription factors during the activation of genes (Fig. 77-4). Histone deacetylases (HDACs; multiple HDACs are encoded in the human genome) are recruited to genes by transcriptional repressors and prevent the initiation of gene transcription. Methylated cytosine residues in promoter regions become associated with methyl cytosine–binding proteins that recruit protein complexes with HDAC activity. The balance between permissive and inhibitory chromatin structure is therefore largely determined by the activity of transcription factors in modulating the "histone code" and the methylation status of the genetic regulatory elements of genes. The pattern of gene transcription is aberrant in all human cancers, and in many cases, epigenetic events are responsible. Epigenetic events play a critical role in carcinogenesis (e.g., long-lasting changes in methylation induced by smoking) and are found in premalignant lesions. Unlike genetic events that alter DNA primary structure (e.g., deletions), epigenetic changes are potentially reversible and appear amenable to therapeutic intervention. In certain human cancers, including a subset of pancreatic cancers and multiple myeloma, the p16Ink4a promoter is inactivated by methylation, thus permitting the unchecked activity of CDK4/cyclin D and rendering pRb nonfunctional. In sporadic forms of renal, breast, and colon cancer, the von Hippel–Lindau (*VHL*), breast cancer 1 (*BRCA1*), and serine/threonine kinase 11 (*STK11*) genes, respectively, can be epigenetically silenced. Other targeted genes include the p15Ink4b CDK inhibitor, glutathione-*S*-transferase (which detoxifies reactive oxygen species [ROS]), and the E-cadherin molecule (important for junction formation between epithelial cells). Epigenetic silencing can affect genes involved in DNA repair, thus predisposing to further genetic damage. Examples include MLH1 (mutL homologue in sporadic colon cancers that have microsatellite instability) and MSH2 in a subset of hereditary nonpolyposis colon cancer patients who have a mutation in the 3′ end of epithelial cell adhesion molecule (EPCAM). These are critical genes involved in

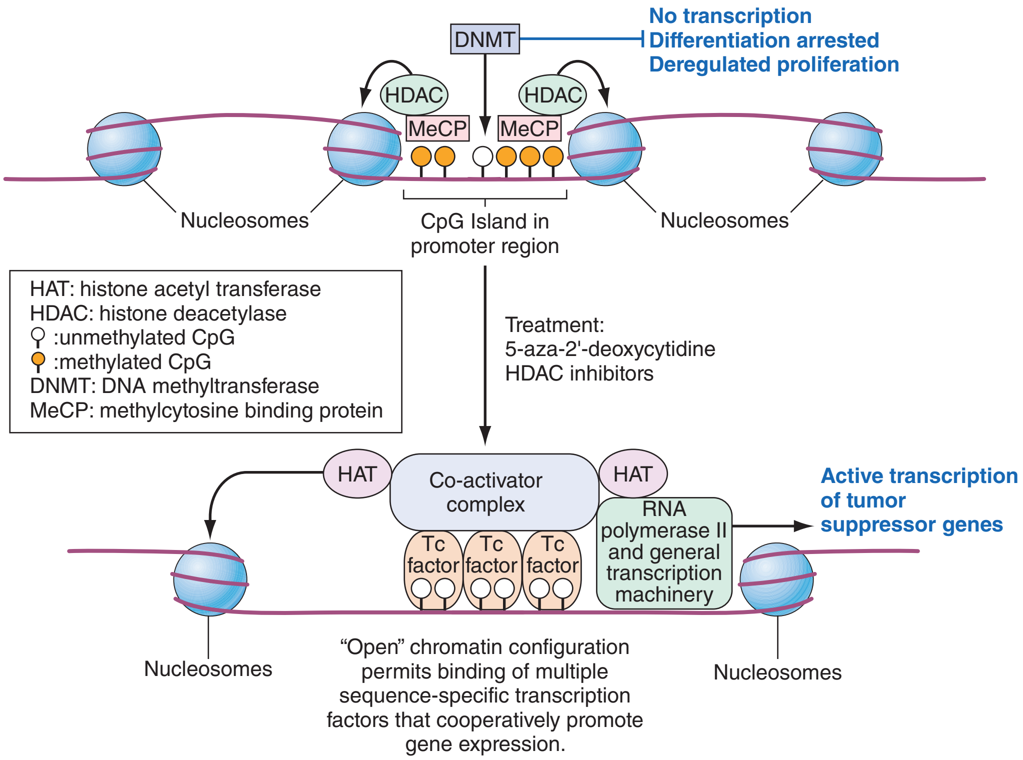
DNMT
No transcription
Differentiation arrested
Deregulated proliferation
HDAC
HDAC
MeCP
MeCP
Nucleosomes
CpG Island in promoter region
Nucleosomes
HAT: histone acetyl transferase
HDAC: histone deacetylase
:unmethylated CpG
:methylated CpG
DNMT: DNA methyltransferase
MeCP: methylcytosine binding protein
Treatment:
5-aza-2'-deoxycytidine
HDAC inhibitors
HAT
Co-activator complex
HAT
RNA polymerase II and general transcription machinery
Tc factor
Tc factor
Tc factor
Active transcription of tumor suppressor genes
Nucleosomes
"Open" chromatin configuration permits binding of multiple sequence-specific transcription factors that cooperatively promote gene expression.
Nucleosomes

FIGURE 77-4 Epigenetic regulation of gene expression in cancer cells. Tumor-suppressor genes are often epigenetically silenced in cancer cells. In the upper portion, a CpG island within the promoter and enhancer regions of the gene has been methylated, resulting in the recruitment of methyl-cytosine binding proteins (MeCP) and complexes with histone deacetylase (HDAC) activity. Chromatin is in a condensed, nonpermissive conformation that inhibits transcription. Clinical trials are under way utilizing the combination of demethylating agents such as 5-aza-2′-deoxycytidine plus HDAC inhibitors, which together confer an open, permissive chromatin structure (*lower portion*). Transcription factors bind to specific DNA sequences in promoter regions and, through protein-protein interactions, recruit coactivator complexes containing histone acetyl transferase (HAT) activity. This enhances transcription initiation by RNA polymerase II and associated general transcription factors. The expression of the tumor-suppressor gene commences, with phenotypic changes that may include growth arrest, differentiation, or apoptosis.

repair of mismatched bases that occur during DNA synthesis, and their silencing can lead to mutations in the DNA.

Human leukemias often have chromosomal translocations that code for novel fusion proteins with activities that alter chromatin structure by interacting with HDACs or histone acetyl transferases (HATs). For example, the promyelocytic leukemia–retinoic acid receptor α (PML-RARα) fusion protein, generated by the t(15;17) translocation observed in most cases of acute promyelocytic leukemia (APL), binds to promoters containing retinoic acid response elements and recruits HDACs to these promoters, effectively inhibiting gene expression. This arrests differentiation at the promyelocyte stage and promotes tumor cell proliferation and survival. Treatment with pharmacologic doses of all-*trans* retinoic acid (ATRA), the ligand for RARα, results in the release of HDAC activity and the recruitment of coactivators, which overcome the differentiation block. This induced differentiation of APL cells has improved treatment of these patients but also has led to a novel treatment toxicity when newly differentiated tumor cells infiltrate the lungs. ATRA represents a treatment paradigm for the reversal of epigenetic changes in cancer. Other leukemia-associated fusion proteins, such as Tel-acute myeloid leukemia (AML1), AML1-eight-twenty-one (ETO), and the MLL fusion proteins seen in acute myeloid leukemia (AML) and acute lymphocytic leukemia, also lead to repression through the HDAC complex. Therefore, efforts are ongoing to determine the structural basis for interactions between translocation fusion proteins and chromatin-remodeling proteins and to use this information to rationally design small molecules that will disrupt specific protein-protein associations, although this has proven to be technically difficult. Several drugs that block the enzymatic activity of HDACs (HDAC inhibitors [HDACis]) are approved for cancer treatment, and others are being tested. HDACis have demonstrated sufficient antitumor activity against cutaneous T-cell lymphoma (vorinostat, romidepsin), peripheral T-cell lymphoma (romidepsin, belinostat), and multiple myeloma (panobinostat) to be approved by the U.S. Food and Drug Administration (FDA).

HDACis have also demonstrated antitumor activity in clinical studies against some solid tumors, and additional studies are ongoing. HDACis may target cancer cells via a number of mechanisms including both epigenetic modulation via histone acetylation and effects on other proteins that are acetylated. The pleiotropic effects of some HDACis include enhancement of apoptosis by upregulation of a number of proteins that enhance apoptosis including death receptors (DR4/5, FAS, and their ligands) and downregulation of proteins that inhibit apoptosis (e.g., X-linked inhibitor of apoptosis [XIAP]); upregulation of proteins that inhibit cell cycle progression (e.g., p21Cip1/Waf1); inhibition of DNA repair and generation of ROS leading to increased DNA damage; and disruption of the chaperone protein HSP90. Efforts are also under way to modulate other epigenetic processes such as reversing the hypermethylation of CpG islands that characterizes many malignancies. Drugs that induce DNA demethylation, such as 5-aza-2-deoxycytidine, can lead to reexpression of silenced genes in cancer cells with restoration of function, and 5-aza-2-deoxycytidine is approved for use in myelodysplastic syndrome. However, 5-aza-2-deoxycytidine has limited aqueous solubility and is myelosuppressive, limiting its usefulness. Other inhibitors of DNA methyltransferases are in development. In ongoing clinical trials, inhibitors of DNA methylation are being combined with HDACis, with the idea that reversing coexisting epigenetic changes will reverse the deregulated patterns of gene transcription in cancer cells.

Epigenetic gene regulation can also occur via microRNAs or long noncoding RNAs (lncRNA). MicroRNAs (miRNA) are short (average 22 nucleotides in length) single strand RNA molecules that regulate gene expression after transcription by specifically binding to and inhibiting the translation or promoting the degradation of mRNA transcripts. It is estimated that >1000 miRNAs are encoded in the human genome. Each tissue has a distinctive repertoire of miRNA expression, and this pattern is altered in specific ways in cancers. miRNA's are involved in controlling multiple aspects of cell biology through modulating protein expression (primarily by down regulation) and thus are also involved in multiple aspects of cancer biology. Specific correlations between expression of different miRNA molecules and tumor biology and clinical behavior are continuing to emerge. Therapies targeting miRNAs are not currently at hand but represent an ongoing area of treatment development. LncRNAs are longer than 200 nucleotides and comprise the largest group of noncoding RNAs. Some of them have been shown to play important roles in gene regulation. The potential for altering these RNAs for therapeutic benefit is an area of active investigation. In addition to epigenetic changes, mutations in genes (such as enhancer and promoter regions) involved in controlling expression of other genes important in cancer cell biology can also lead to enhanced or decreased expression of the protein products of these genes.

APOPTOSIS AND OTHER MECHANISMS OF CELL DEATH

Tissue homeostasis requires a balance between the death of aged, terminally differentiated cells or severely damaged cells and their renewal by proliferation of committed progenitors. Genetic damage to growth-regulating genes of stem cells could lead to catastrophic results for the host as a whole. Thus, in normal cells, the genetic events causing activation of oncogenes or loss of tumor suppressors, which would be predicted to lead to unregulated cell proliferation unless corrected, also usually activate signal transduction pathways that block aberrant cell proliferation. These pathways can lead to forms of programmed cell death including apoptosis or autophagy (degradation of proteins and organelles by lysosomal proteases) or irreversible growth arrest (senescence). A number of other regulated cell death processes have been identified, including: pyroptosis, a caspase-1-dependent process leading to cleavage of gasdermins with subsequent formation of pores in the plasma membrane; ferroptosis (iron and reactive oxygen species dependent); and necroptosis (caspase-independent regulated cell death involving breakdown of cellular components and cell rupture, leading to inflammation and damage to surrounding tissues), which also play roles in tissue homeostasis and cell death. However, the exact roles they play in the fate of cancer cells and tissues are still being elucidated. Much as a panoply of intra- and extracellular signals impinge upon the core cell cycle machinery to regulate cell division, so too these signals are transmitted to a core enzymatic machinery that regulates cell death and survival. Cancer cells have developed mechanisms that either inhibit these processes to prevent cell death or utilize them to enhance survival.

Apoptosis is a tightly regulated process induced by two main pathways **(Fig. 77-5)**. The extrinsic pathway of apoptosis is activated by cross-linking members of the tumor necrosis factor (TNF) receptor superfamily, such as CD95 (Fas) and death receptors DR4 and DR5, by their ligands, Fas ligand or TRAIL (TNF-related apoptosis-inducing ligand), respectively. This induces the association of FADD (Fas-associated death domain) and procaspase-8 to death domain motifs of the receptors. Caspase-8 is activated and then cleaves and activates effector caspases-3 and -7, which then target cellular constituents (including caspase-activated DNase, cytoskeletal proteins, and a number of regulatory proteins), inducing the morphologic appearance characteristic of apoptosis, which pathologists term karyorrhexis (literally "nucleus bursting").

The intrinsic pathway of apoptosis is initiated by the release of cytochrome c and SMAC (second mitochondrial activator of caspases) from the mitochondrial intermembrane space in response to a variety of noxious stimuli, including DNA damage, loss of adherence to the extracellular matrix (ECM), oncogene-induced proliferation, and growth factor deprivation. Upon release into the cytoplasm, cytochrome c associates with dATP, procaspase-9, and the adaptor protein APAF-1, leading to the sequential activation of caspase-9 and effector caspases. SMAC binds to and blocks the function of inhibitor of apoptosis proteins (IAP), negative regulators of caspase activation. The release of apoptosis-inducing proteins from the mitochondria is regulated by pro- and antiapoptotic members of the Bcl-2 family. Antiapoptotic members (e.g., Bcl-2, Bcl-XL, and Mcl-1) associate with the mitochondrial outer membrane via their carboxyl termini, exposing to the cytoplasm a hydrophobic binding pocket composed of Bcl-2

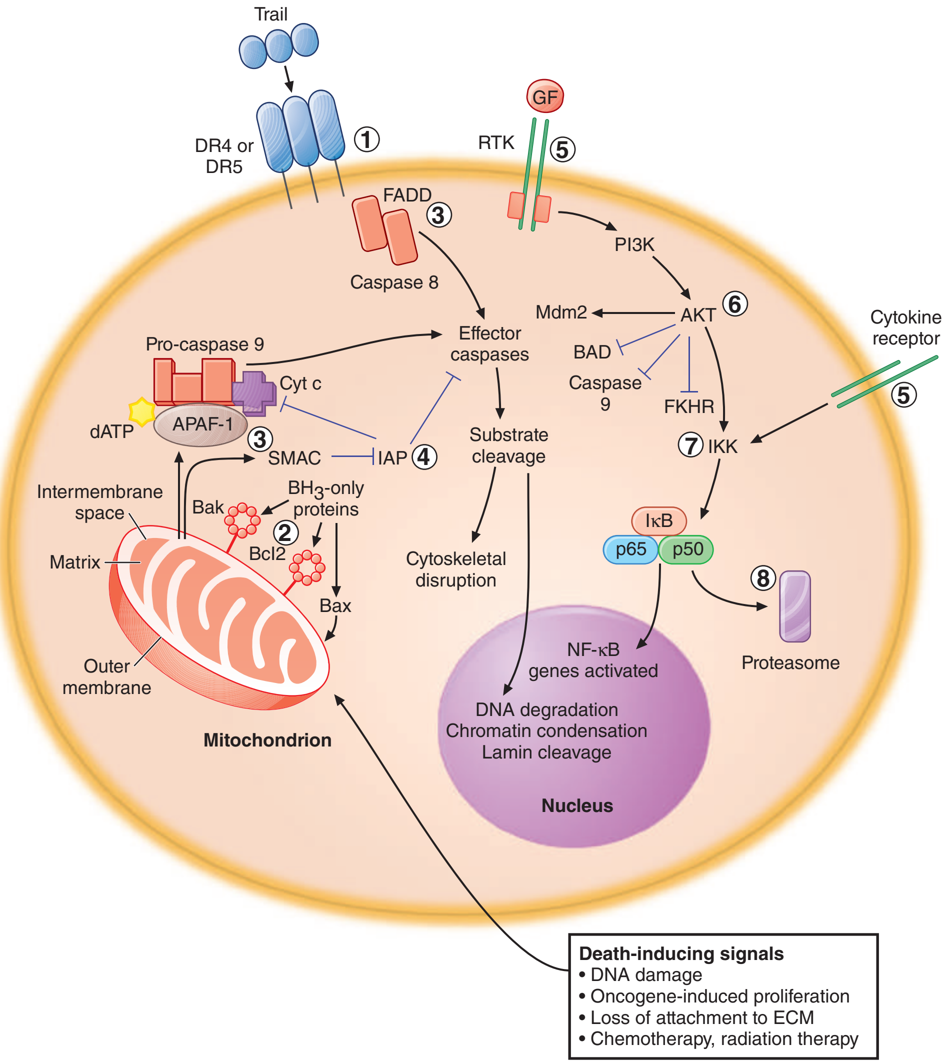
Trail
DR4 or DR5
①
GF
RTK
⑤
FADD
③
Caspase 8
PI3K
Mdm2
AKT
⑥
Effector caspases
BAD
Caspase 9
FKHR
Cytokine receptor
⑤
Pro-caspase 9
Cyt c
dATP
APAF-1
③
SMAC
IAP
④
Substrate cleavage
⑦
IKK
Intermembrane space
Bak
BH3-only proteins
②
Bcl2
Matrix
IκB
p65
p50
Cytoskeletal disruption
⑧
Bax
NF-κB genes activated
Proteasome
Outer membrane
DNA degradation
Chromatin condensation
Lamin cleavage
Mitochondrion
Nucleus
Death-inducing signals
• DNA damage
• Oncogene-induced proliferation
• Loss of attachment to ECM
• Chemotherapy, radiation therapy

FIGURE 77-5 Therapeutic strategies to overcome aberrant survival pathways in cancer cells. *1.* The extrinsic pathway of apoptosis can be selectively induced in cancer cells by TRAIL (the ligand for death receptors 4 and 5) or by agonistic monoclonal antibodies. *2.* Inhibition of antiapoptotic Bcl-2 family members with antisense oligonucleotides or inhibitors of the BH_3-binding pocket will promote formation of Bak- or Bax-induced pores in the mitochondrial outer membrane. *3.* Epigenetic silencing of APAF-1, caspase-8, and other proteins can be overcome using demethylating agents and inhibitors of histone deacetylases. *4.* Inhibitor of apoptosis proteins (IAP) blocks activation of caspases; small-molecule inhibitors of IAP function (mimicking SMAC action) should lower the threshold for apoptosis. *5.* Signal transduction pathways originating with activation of receptor tyrosine kinase receptors (RTKs) or cytokine receptors promote survival of cancer cells by a number of mechanisms. Inhibiting receptor function with monoclonal antibodies, such as trastuzumab or cetuximab, or inhibiting kinase activity with small-molecule inhibitors can block the pathway. *6.* The Akt kinase phosphorylates many regulators of apoptosis to promote cell survival; inhibitors of Akt may render tumor cells more sensitive to apoptosis-inducing signals; however, the possibility of toxicity to normal cells may limit the therapeutic value of these agents. *7* and *8.* Activation of the transcription factor NF-κB (composed of p65 and p50 subunits) occurs when its inhibitor, IκB, is phosphorylated by IκB-kinase (IKK), with subsequent degradation of IκB by the proteasome. Inhibition of IKK activity should selectively block the activation of NF-κB target genes, many of which promote cell survival. Inhibitors of proteasome function are U.S. Food and Drug Administration approved and may work in part by preventing destruction of IκB, thus blocking NF-κB nuclear localization. NF-κB is unlikely to be the only target for proteasome inhibitors.

homology (BH) domains 1, 2, and 3 that is crucial for their activity. Perturbations of normal physiologic processes in specific cellular compartments lead to the activation of BH_3-only proapoptotic family members (e.g., Bad, Bim, Bid, Puma, Noxa, and others) that can alter the conformation of the outer-membrane proteins Bax and Bak, which then oligomerize to form pores in the mitochondrial outer membrane resulting in cytochrome c release. If proteins composed only by BH_3 domains are sequestered by Bcl-2, Bcl-XL, or Mcl-1, pores do not form and apoptosis-inducing proteins are not released from the mitochondria. The ratio of levels of antiapoptotic Bcl-2 family members and the levels of proapoptotic BH_3-only proteins at the mitochondrial membrane determines the activation state of the intrinsic pathway. The mitochondrion must therefore be recognized not only as an organelle with vital roles in intermediary metabolism and oxidative phosphorylation but also as a central regulatory structure of the apoptotic process. The evolution of tumor cells to a more malignant phenotype requires the acquisition of genetic changes that subvert apoptosis pathways and promote cancer cell survival and resistance to anticancer therapies.

However, this means that cancer cells may be more vulnerable than normal cells to therapeutic interventions that target the apoptosis pathways that cancer cells depend upon. For instance, overexpression of Bcl-2 as a result of the t(14;18) translocation contributes to follicular lymphoma, and it is highly expressed in many lymphoid malignancies including chronic lymphocytic leukemia (CLL). Upregulation of Bcl-2 expression is also observed in other cancers including prostate, breast and lung cancers, and melanoma. Targeting of antiapoptotic Bcl-2 family members has been accomplished by the identification of several low-molecular-weight compounds that bind to the hydrophobic pockets of either Bcl-2 or Bcl-XL and block their ability to associate with death-inducing BH_3-only proteins. An oral BH3 mimetic inhibitor of BCL-2, venetoclax, is approved for use in patients with refractory CLL with 17p deletion, and is active in AML. Preclinical studies targeting death receptors DR4 and -5 have demonstrated that recombinant, soluble, human TRAIL or humanized monoclonal antibodies with agonist activity against DR4 or -5 can induce apoptosis of tumor cells while sparing normal cells. The mechanisms for this selectivity may include expression of decoy receptors or elevated levels of intracellular inhibitors (such as FLIP, which competes with caspase-8 for FADD) by normal cells but not tumor cells. Synergy has been shown between TRAIL-induced apoptosis and chemotherapeutic agents in some preclinical studies. However, studies have not yet shown significant clinical activity of approaches targeting the TRAIL pathway.

Many of the signal transduction pathways perturbed in cancer promote tumor cell survival (Fig. 77-5). These include activation of the PI3K/Akt pathway, increased levels of the NF-κB transcription factor, and epigenetic silencing of genes such as APAF-1 (apoptosis protease activating factor-1 involved in activating caspase-9 and essential for apoptosis) and caspase-8. Each of these pathways is a target for therapeutic agents that, in addition to affecting cancer cell proliferation or gene expression, may render cancer cells more susceptible to apoptosis, thus promoting synergy when combined with other chemotherapeutic agents.

Some tumor cells resist drug-induced apoptosis indirectly by eliminating the noxious stimulus-inducing apoptosis through expression of one or more members of the ABC (ATP-binding cassette proteins) family of ATP-dependent efflux pumps that mediate the multidrug resistance (MDR) phenotype. The prototype member of this family, P-glycoprotein (PGP), spans the plasma membrane 12 times and has two ATP-binding sites. Hydrophobic drugs (e.g., anthracyclines and vinca alkaloids) are recognized by PGP as they enter the cell and are pumped out. Numerous clinical studies have failed to demonstrate that drug resistance can be overcome using inhibitors of PGP. However, ABC transporters have different substrate specificities, and inhibition of a single family member may not be sufficient to overcome the MDR phenotype. Efforts to reverse PGP-mediated drug resistance continue.

In addition to its role in cell death, autophagy can also serve as a homeostatic mechanism to promote cell survival by recycling cellular components to provide necessary energy. The mechanisms that control the balance between enhancing survival versus leading to cell death are still not fully understood. Autophagy appears to play conflicting roles in the development and survival of cancer. Early in the carcinogenic process, it can act as a tumor suppressor by preventing the cell from accumulating abnormal proteins and organelles. However, in established tumors, it may serve as a mechanism of survival for cancer cells when they are stressed by damage such as from chemotherapy. Preclinical studies have indicated that inhibition of this process can enhance the sensitivity of cancer cells to chemotherapy or radiation therapy, and ongoing trials are evaluating inhibitors of autophagy in combination with chemotherapy and/or radiation therapy. Better understanding of the factors that control the survival-promoting versus death-inducing aspects of autophagy is required in order to know how to best manipulate it for therapeutic benefit.

METASTASIS

The metastatic process accounts for the vast majority of deaths from solid tumors, and therefore, an understanding of this process is critical for improvements in survival from cancer. The biology of metastasis is complex and requires multiple steps. The initial step involves cell migration and invasion through the ECM. The three major features of tissue invasion are cell adhesion to the basement membrane, local proteolysis of the membrane, and movement of the cell through the rent in the membrane and the ECM. Cells that lose contact with the ECM normally undergo programmed cell death (anoikis-apoptosis induced by the loss of contact), and this process has to be suppressed in cells that metastasize. Another process important for many, but not necessarily all, metastasizing epithelial cancer cells is epithelial mesenchymal transition (EMT). This is a process by which cells lose their epithelial properties and gain mesenchymal properties. This normally occurs during the developmental process in embryos, allowing cells to migrate to their appropriate destinations in the embryo. It also occurs in wound healing, tissue regeneration, and fibrotic reactions, but in all of these processes, cells stop proliferating when the process is complete. Malignant cells that metastasize often undergo EMT as an important step in that process but retain the capacity for unregulated proliferation. However, there is evidence that not all metastasizing cancer cells require EMT, and the exact role of EMT in different metastasizing cancer cells continues to be elucidated. Malignant cells that gain access to the circulation must then repeat those steps at a remote site, find a hospitable niche in a foreign tissue, avoid detection and elimination by host defenses including the immune system, and induce the growth of new blood vessels. Some metastatic cells occur as oligoclonal clusters, which appear to be more potent in establishing metastasis than single cells, perhaps, in part, through differential and cooperative effects in evading host defenses. The rate-limiting step for metastasis is the ability for tumor cells to survive and expand in the novel microenvironment of the metastatic site, and multiple host-tumor interactions determine the ultimate outcome **(Fig. 77-6)**.

As is true for cells in primary cancers, there is significant heterogeneity as well as plasticity in metastatic cancer cells. In addition to actively dividing cells, a population of quiescent cells are present that can evade the immune system as well as chemotherapy targeting dividing cells. The processes that keep metastatic cancer cells quiescent as well as lead them to divide are complex, as discussed in the introductory section. Efforts to inhibit growth of metastatic cells by modulating these pathways are being explored. Few drugs have been developed to attempt to directly target the process of metastasis, in part because the specifics of the critical steps in the process that would be potentially good targets for drugs are still being identified. However, a number of potential targets are known. HER2 can enhance the metastatic potential of breast cancer cells, and as discussed above, the monoclonal antibody trastuzumab, which targets HER2, improves survival in the adjuvant setting for HER2-positive breast cancer patients. A number of other potential targets that increase metastatic potential of cells in preclinical studies include HIF-1 and -2, transcription factors induced by hypoxia within tumors, growth factors (e.g., cMET and VEGFR), oncogenes (e.g., SRC), adhesion molecules (e.g., focal adhesion kinase [FAK]), ECM proteins (e.g., matrix metalloproteinases 1 and 2), and inflammatory molecules (e.g., COX-2).

The metastatic phenotype is likely restricted to a fraction of tumor cells (Fig. 77-6). A number of genetic and epigenetic changes are required for tumor cells to be able to metastasize, including activation of metastatic-promoting genes and inhibition of genes that suppress the metastatic ability. Given the role of microRNAs in controlling gene expression (see epigenetic section) including those critical to the metastatic process, efforts are under way to modulate these to try to inhibit metastasis. Cells with metastatic capability frequently express chemokine receptors that are likely important in the metastatic process. A number of candidate metastasis-suppressor genes have been identified, including genes coding for proteins that enhance apoptosis, suppress cell division, are involved in the interactions of cells with each other or the ECM, or suppress cell migration. The loss of function of these genes enhances metastasis. Gene expression profiling is being used to study the metastatic process and other properties of tumor cells that may predict susceptibilities.

An example of the ability of malignant cells to survive and grow in a novel microenvironment is bone metastases. Bone metastases can be extremely painful, cause fractures of weight-bearing bones, can lead to hypercalcemia, and are a major cause of morbidity for cancer patients. Osteoclasts and their monocyte-derived precursors express the surface receptor RANK (receptor activator of NF-κB), which is required for terminal differentiation and activation of osteoclasts. Osteoblasts and other stromal cells express RANK ligand (RANKL), as both a membrane-bound and soluble cytokine. Osteoprotegerin (OPG), a soluble receptor for RANKL produced by stromal cells, acts as a decoy receptor to inhibit RANK activation. The relative balance of RANKL and OPG determines the activation state of RANK on osteoclasts. Bone modulation and resorption by osteoclasts is an important component of the establishment and progression of metastases in bone. Many tumors increase osteoclast activity by secretion of substances such as parathyroid hormone (PTH), PTH-related peptide, interleukin (IL) 1, or Mip1 that perturb the homeostatic balance of bone remodeling by increasing RANK signaling. One example is multiple myeloma, where tumor cell–stromal cell interactions activate osteoclasts and inhibit osteoblasts, leading to the development of multiple lytic bone lesions.

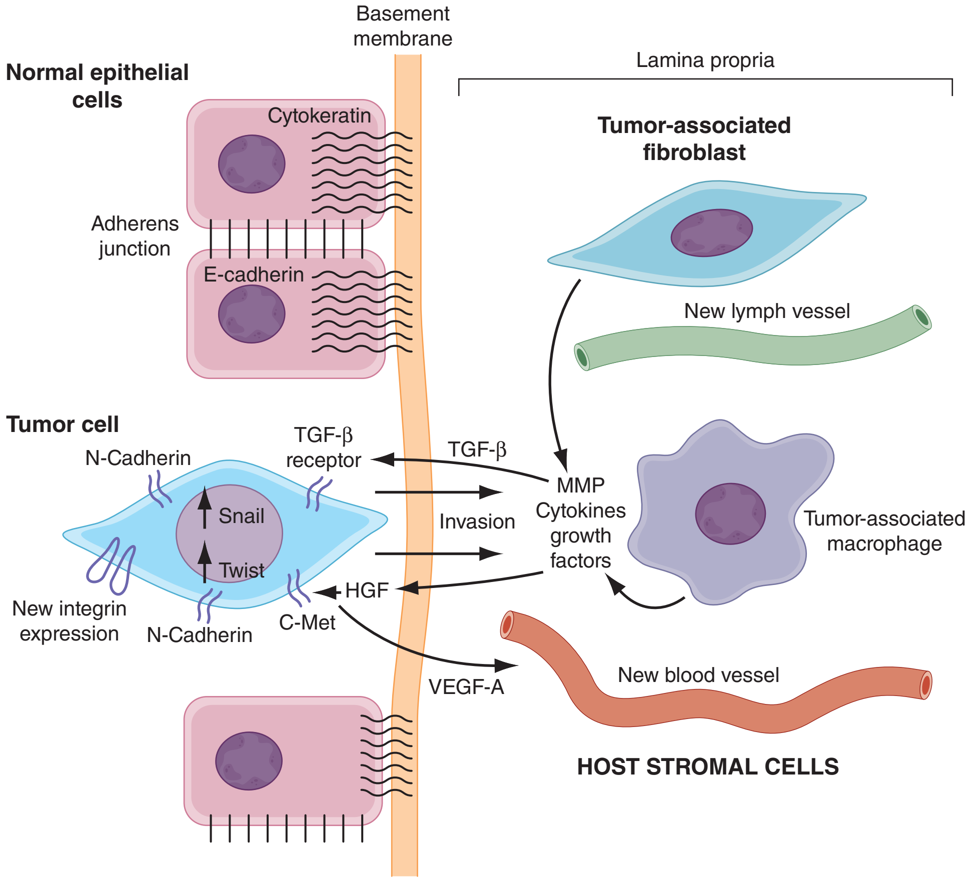

FIGURE 77-6 Oncogene signaling pathways are activated during tumor progression and promote metastatic potential. This figure shows a cancer cell that has undergone epithelial to mesenchymal transition (EMT) under the influence of several environmental signals. Critical components include activated transforming growth factor beta (TGF-β) and the hepatocyte growth factor (HGF)/c-Met pathways, as well as changes in the expression of adhesion molecules that mediate cell-cell and cell–extracellular matrix interactions. Important changes in gene expression are mediated by the Snail and Twist family of transcriptional repressors (whose expression is induced by the oncogenic pathways), leading to reduced expression of E-cadherin, a key component of adherens junctions between epithelial cells. This, in conjunction with upregulation of N-cadherin, a change in the pattern of expression of integrins (which mediate cell–extracellular matrix associations that are important for cell motility), and a switch in intermediate filament expression from cytokeratin to vimentin, results in the phenotypic change from adherent highly organized epithelial cells to motile and invasive cells with a fibroblast or mesenchymal morphology. EMT is thought to be an important step leading to metastasis in some human cancers. Host stromal cells, including tumor-associated fibroblasts and macrophages, play an important role in modulating tumor cell behavior through secretion of growth factors and proangiogenic cytokines, and matrix metalloproteinases that degrade the basement membrane. VEGF-A, -C, and -D are produced by tumor cells and stromal cells in response to hypoxemia or oncogenic signals and induce production of new blood vessels and lymphatic channels through which tumor cells metastasize to lymph nodes or tissues.

Inhibition of RANKL by an antibody (denosumab) can prevent further bone destruction. Bisphosphonates are also effective inhibitors of osteoclast function that are used in the treatment of cancer patients with bone metastases.

CANCER STEM CELLS

Normal tissues have stem cells capable of self-renewal and repairing damaged tissue, whereas the majority of cells in normal tissues do not have this capacity. Similarly, only a small proportion of the cells within a tumor are capable of initiating colonies in vitro or forming tumors at high efficiency when injected into immunocompromised NOD/SCID mice. For example, AML and CML have a small population of cells (estimated to be <1%) that have properties of stem cells, such as unlimited self-renewal and the capacity to cause leukemia when serially transplanted in mice. These cells have an undifferentiated phenotype (Thy1–CD34+CD38– and do not express other differentiation markers) and resemble normal stem cells in many ways but are no longer under homeostatic control (Fig. 77-7). Solid tumors may also contain a population of stem cells. It is not yet known how often cancers may originate within a stem cell population, although a body of evidence argues that stem cells are likely involved in the development of the majority of cancers. Cancer stem cells, like their normal counterparts, have unlimited proliferative capacity and paradoxically traverse the cell cycle at a slow rate; cancer growth occurs largely due to expansion of the stem cell pool, the unregulated proliferation of an amplifying population, and failure of apoptosis pathways (Fig. 77-7). Slow cell cycle progression and high levels of expression of antiapoptotic Bcl-2 family members and drug efflux pumps of the MDR family render cancer stem cells less vulnerable to cancer chemotherapy or radiation therapy.

Implicit in the cancer stem cell hypothesis is the idea that failure to cure most human cancers is due to the fact that current therapeutic agents are not very effective in killing stem cells. Efforts are ongoing to identify and isolate cancer stem cells from different types of malignancies, which should allow determination of the aberrant signaling pathways that distinguish these cells from normal tissue stem cells. These would serve as potential therapeutic targets. Evidence that cells with stem cell properties can arise from other epithelial cells within the cancer by processes such as epithelial mesenchymal transition also implies that it is essential to treat all of the cancer cells, and not just those with current stem cell–like properties, in order to eliminate the self-renewing cancer cell population. The exact nature of cancer stem cells remains an area of investigation. One of the unanswered questions is the exact origin of cancer stem cells for the different cancers.

PLASTICITY AND RESISTANCE

Cancer cells, and especially stem cells, have the capacity for significant plasticity, allowing them to alter multiple aspects of cell biology in response to external factors (e.g., chemotherapy, radiation therapy, inflammation, immune response). In addition, heterogeneity between

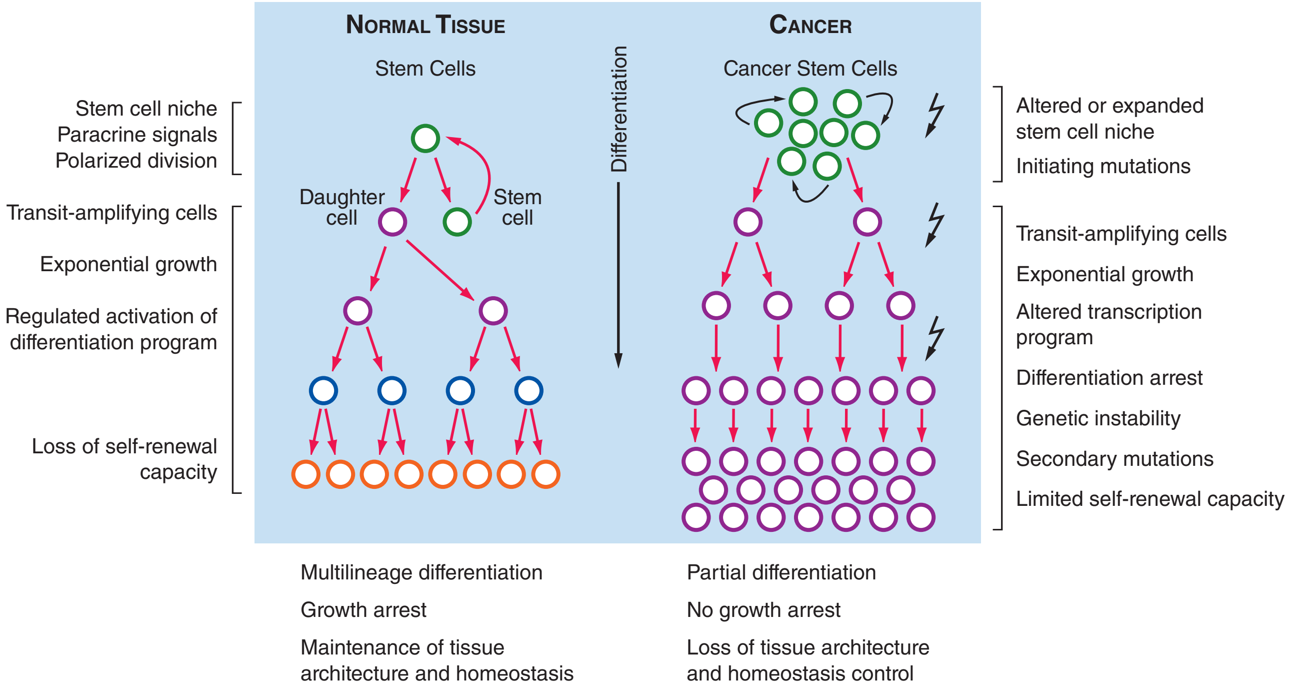

FIGURE 77-7 Cancer stem cells play a critical role in the initiation, progression, and resistance to therapy of malignant neoplasms. In normal tissues (*left*), homeostasis is maintained by asymmetric division of stem cells, leading to one progeny cell that will differentiate and one cell that will maintain the stem cell pool. This occurs within highly specific niches unique to each tissue, such as in close apposition to osteoblasts in bone marrow, or at the base of crypts in the colon. Here, paracrine signals from stromal cells, such as sonic hedgehog or Notch ligands, as well as upregulation of β-catenin and telomerase, help to maintain stem cell features of unlimited self-renewal while preventing differentiation or cell death. This occurs in part through upregulation of the transcriptional repressor Bmi-1 and inhibition of the p16$^{\text{Ink4a}}$/Arf and p53 pathways. Daughter cells leave the stem cell niche and enter a proliferative phase (referred to as *transit-amplifying*) for a specified number of cell divisions, during which time a developmental program is activated, eventually giving rise to fully differentiated cells that have lost proliferative potential. Cell renewal equals cell death, and homeostasis is maintained. In this hierarchical system, only stem cells are long-lived. The hypothesis is that cancers harbor stem cells that make up a small fraction (i.e., 0.001–1%) of all cancer cells. These cells share several features with normal stem cells, including an undifferentiated phenotype, unlimited self-renewal potential, and a capacity for some degree of differentiation; however, due to initiating mutations (mutations are indicated by lightning bolts), they are no longer regulated by environmental cues. The cancer stem cell pool is expanded, and rapidly proliferating progeny, through additional mutations, may attain stem cell properties, although most of this population is thought to have a limited proliferative capacity. Differentiation programs are dysfunctional due to reprogramming of the pattern of gene transcription by oncogenic signaling pathways. Within the cancer transit-amplifying population, genomic instability generates aneuploidy and clonal heterogeneity as cells attain a fully malignant phenotype with metastatic potential. The cancer stem cell hypothesis has led to the idea that current cancer therapies may be effective at killing the bulk of tumor cells but do not kill tumor stem cells, leading to a regrowth of tumors that is manifested as tumor recurrence or disease progression. Research is in progress to identify unique molecular features of cancer stem cells that can lead to their direct targeting by novel therapeutic agents.

the different clones of cells within the tumor population and their interactions with each other and the tumor microenvironment provides the tumor with the capacity for significant plasticity in dealing with both internal and external stresses. Thus, a major problem in cancer therapy is that malignancies have a wide spectrum of mechanisms for both initial and adaptive resistance to treatments. These include inhibiting drug delivery to the cancer cells, blocking drug uptake and retention, increasing drug metabolism, altering levels of target proteins making them less sensitive to drugs, acquiring mutations in target proteins making them no longer sensitive to the drug, modifying metabolism and cell signaling pathways, using alternate signaling pathways, adjusting the cell replication process including mechanisms by which the cell deals with DNA damage, inhibiting apoptosis, and evading the immune system. Thus, most metastatic cancers (except those curable with chemotherapy such as germ cell tumors) eventually become resistant to the therapy being utilized. Overcoming resistance is a major area of research.

CANCER METABOLISM

One of the distinguishing characteristics of cancer cells is that they have altered metabolism as compared with normal cells in supporting survival, their high rates of proliferation, and ability to metastasize. Complicating studies evaluating metabolic differences between normal and malignant cells is that there is heterogeneity in metabolism between different cells within a cancer. Malignant cells must focus a significant fraction of their energy resources into synthesis of proteins and other molecules (building blocks required for the production of new cells) while still maintaining sufficient ATP production to survive and grow. Although normal proliferating cells also have similar needs, there are differences in how cancer cells metabolize glucose and a number of other compounds including the amino acid glutamine as compared to normal cells in part because of genetic and epigenetic changes within cancer cells but also likely due to differences in the environments of cancer and normal cells. Many cancer cells utilize aerobic glycolysis (the Warburg effect) **(Fig. 77-8)** to metabolize glucose, leading to increased lactic acid production, whereas normal cells utilize oxidative phosphorylation in mitochondria under aerobic conditions, a much more efficient process for generating ATP for energy utilization but one that does not produce the same level of building blocks needed for new cells. One consequence is increased glucose uptake and utilization by cancer cells, a fact utilized in fluorodeoxyglucose (FDG)-positron emission tomography (PET) scanning to detect tumors. A number of proteins in cancer cells, including cMYC, HIF1, RAS, p53, pRB, and AKT, are involved in modulating glycolytic processes and controlling the Warburg effect. Although these pathways overall remain difficult to target therapeutically, some progress has been made in targeting HIF1 and the RAS pathways with inhibitors approved to treat cancers with mutations in HIF1α or KRAS12C. In addition, both the PI3K pathway with signaling through mTOR and the AMP-activated kinase (AMPK) pathway that inhibits mTORC1 (a protein complex that includes mTOR) are important in controlling the glycolytic process and thus provide potential targets for inhibiting this process. An inhibitor of mTOR is approved for use against RCC (temsirolimus), and another inhibitor (everolimus) has activity against breast and neuroendocrine cancer and RCC. Other mTOR inhibitors are in trials, and modulators of AMPK are being investigated. The inefficient utilization of glucose

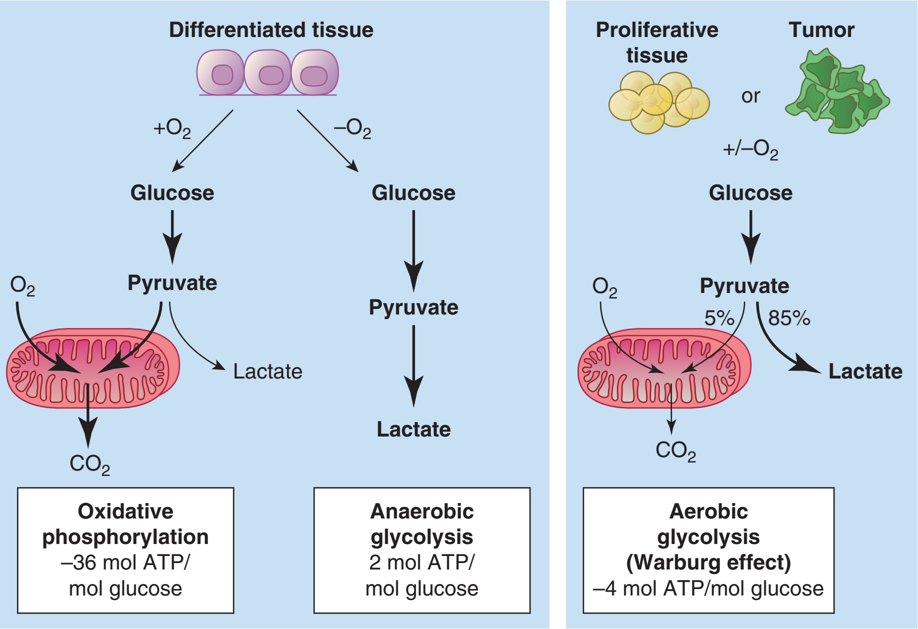
Differentiated tissue
+O₂
–O₂
Glucose
Pyruvate
O₂
Lactate
CO₂
Glucose
Pyruvate
Lactate
Proliferative tissue
or
Tumor
+/–O₂
Glucose
O₂
Pyruvate
5%
85%
Lactate
CO₂
Oxidative phosphorylation –36 mol ATP/ mol glucose
Anaerobic glycolysis 2 mol ATP/ mol glucose
Aerobic glycolysis (Warburg effect) –4 mol ATP/mol glucose

FIGURE 77-8 Warburg versus oxidative phosphorylation. In most normal tissues, the vast majority of cells are differentiated and dedicated to a particular function within the organ in which they reside. The metabolic needs are mainly for energy and not for building blocks for new cells. In these tissues, ATP is generated by oxidative phosphorylation that efficiently generates about 36 molecules of ATP for each molecule of glucose metabolized. By contrast, proliferative tumor tissues, especially in the setting of hypoxia, a typical condition within tumors, use aerobic glycolysis to generate energy for cell survival and generation of building blocks for new cells.

by malignant cells also leads to a need for alternative metabolic pathways for other compounds as well, one of which is glutamine. Similar to glucose, this provides both a source for structural molecules as well as energy production. Similarly to glucose, glutamine is also inefficiently utilized by cancer cells. Cancer cells can also take up nutrients released by surrounding cells and tissues, increasing the complexity of successfully therapeutically inhibiting metabolism in cancer.

Mutations in genes involved in the metabolic process occur in a number of cancers. Among the most frequently found to date are mutations in isocitrate dehydrogenases 1 and 2 (IDH1 and IDH2). These have been most commonly seen in gliomas, AMLs, and intrahepatic cholangiocarcinomas. These mutations lead to the production of an oncometabolite (2-hydroxyglutarate [2HG]) instead of the normal product α-ketoglutarate. Although the exact mechanisms of oncogenesis by 2HG are still being elucidated, α-ketoglutarate is a key cofactor for a number of dioxygenases involved in controlling DNA methylation. 2HG can act as a competitive inhibitor for α-ketoglutarate, leading to alterations in methylation status (primarily hypermethylation) of genes (leading to epigenetic changes) that can have profound effects on a number of cellular processes including differentiation. Inhibitors of mutant IDH1 and/or IDH2 are approved for treating IDH mutant AML and cholangiocarcinoma; and a dual IDH1/2 inhibitor is approved for treatment of low-grade gliomas and astrocytomas with IDH mutations.

Much needs to be learned about the specific differences in metabolism between cancer cells and normal cells in order to develop more effective approaches to using these differences therapeutically; however, even with the currently limited state of knowledge, modulators of metabolism are being tested clinically. One of these is the antidiabetic agent metformin, both alone and in combination with chemotherapeutic agents. Metformin inhibits gluconeogenesis and may have direct effects on tumor cells by activating AMPK, a serine/threonine protein kinase that is downstream of the LKB1 tumor suppressor, and thus inhibiting mTOR complex 1 (mTORC1). This leads to decreased protein synthesis and proliferation. Studies to date have not yet established metformin to have a clear role as an anticancer agent.

TUMOR MICROENVIRONMENT, ANGIOGENESIS, AND IMMUNE EVASION

Tumors consist not only of malignant cells but also of a complex microenvironment including many other types of cells (including lymphocytes, macrophages, myeloid cells; other inflammatory cells; vascular cells, lymphatic endothelial cells, nerve cells, fibroblasts, and fat cells), ECM, stroma, secreted factors (including growth factors and hormones), reactive oxygen and nitrogen species, mechanical factors, blood vessels, and lymphatics. There is extensive cross-talk between the cells with each other, the ECM, and the various secreted factors within the tumor microenvironment. This microenvironment is not static but rather is dynamic and continually evolving. Both the complexity and dynamic nature of the microenvironment enhance the difficulty of treating tumors. The microenvironment is involved in altered tumor metabolism, tumor maintenance, growth, phenotypic plasticity, metastasis, and immune escape, and can contribute to resistance to anticancer therapies through a number of mechanisms. These include immune evasion by a variety of mechanisms including suppression of effector T cells, increase in regulatory T cells, induction of an inflammatory environment, and altered vasculature that inhibits effector T-cell access to malignant cells. Similarly, it contributes to drug resistance through multiple mechanisms, alteration in metabolic pathways including creating a hypoxic and acidic environment, vascular and mechanical factors that limit drug access to malignant cells, various secreted factors that inhibit apoptosis or stimulate survival pathways, and generation of ROS that enhance drug resistance. Multiple additional mechanisms are also involved in enhancing resistance to immune-mediated anticancer effects and anticancer drug therapy.

OBESITY AND CANCER

Significant evidence links obesity and the increased risk of developing certain cancers including postmenopausal breast, colorectal, ovarian, endometrial, esophageal, gallbladder, thyroid, and kidney cancers, among others. Less certain are the mechanisms responsible for this risk. As outlined above, cancers arise in an environment with multiple factors, many of which can stimulate cell proliferation. Obesity impacts a variety of factors including hormonal factors, altered metabolism (especially adipose metabolism), and mediators of inflammatory response that all can impact the development of malignancy. Obesity is associated with a number of hormonal changes including high insulin, glucagon, and leptin levels that can stimulate growth of cells. It also leads to insulin resistance, which may contribute to cancer cell development, in part by increasing insulin-like growth factor-1 (IGF-1) levels. Obesity also leads to alterations in adipose, including fatty acid, metabolism with production of compounds important for energy metabolism as well as for membrane function within cells that may contribute to carcinogenic process. Obesity contributes to an inflammatory environment in a variety of ways including increased levels of inflammatory proteins such as IL-6 and TNF-α. In terms of impact on survival with cancer, data primarily from breast cancer suggest that obesity is associated with decreased survival likely due, at least in part, to the impact of obesity on hormonal factors in development of certain breast cancers, although this may be limited to subsets of breast cancer patients. Some studies have suggested, paradoxically, that obesity may be associated with improved survival in some patients such as those with advanced-stage colorectal cancer. Furthermore, immune checkpoint inhibitor therapy has appeared to be more effect in obese patients. Clearly, the biology of the association between obesity and cancer and its impact on disease outcome is complex, and additional studies are necessary to better define the mechanisms involved.

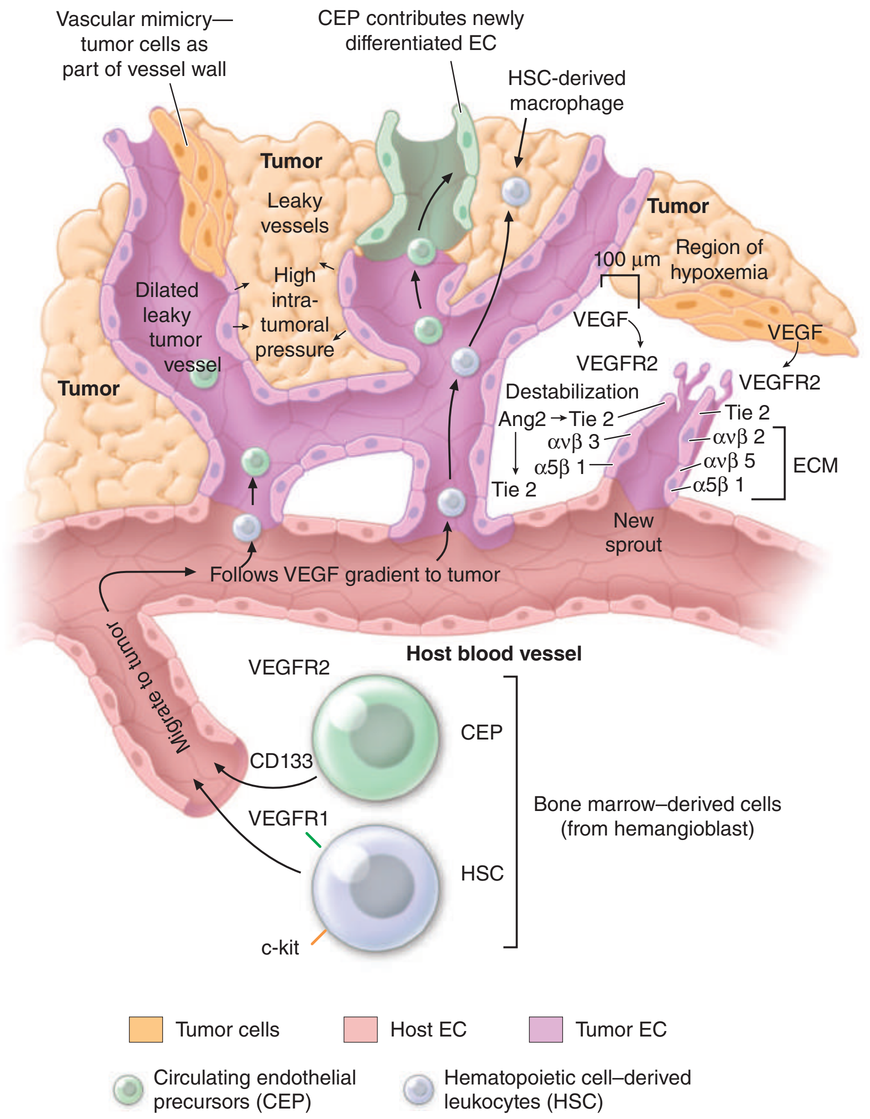

FIGURE 77-9 Tumor angiogenesis is a complex process involving many different cell types that must proliferate, migrate, invade, and differentiate in response to signals from the tumor microenvironment. Endothelial cells (ECs) sprout from host vessels in response to VEGF, bFGF, Ang2, and other proangiogenic stimuli. Sprouting is stimulated by VEGF/VEGFR2, Ang2/Tie-2, and integrin/extracellular matrix (ECM) interactions. Bone marrow–derived circulating endothelial precursors (CEPs) migrate to the tumor in response to VEGF and differentiate into ECs, while hematopoietic stem cells differentiate into leukocytes, including tumor-associated macrophages that secrete angiogenic growth factors and produce matrix metalloproteinases (MMPs) that remodel the ECM and release bound growth factors. Tumor cells themselves may directly form parts of vascular channels within tumors. The pattern of vessel formation is haphazard: vessels are tortuous, dilated, leaky, and branch in random ways. This leads to uneven blood flow within the tumor, with areas of acidosis and hypoxemia (which stimulate release of angiogenic factors) and high intratumoral pressures that inhibit delivery of therapeutic agents.

■ MECHANISMS OF TUMOR VESSEL FORMATION

One of the critical elements of tumor cell proliferation is delivery of oxygen, nutrients, and circulating factors important for growth and survival. Thus, a critical element in growth of primary tumors and formation of metastatic sites is the angiogenic switch: the ability of the tumor to promote the formation of new blood vessels, including the recruitment of vascular endothelial cells (ECs). The angiogenic switch is a phase in tumor development when the dynamic balance of pro- and antiangiogenic factors is tipped in favor of vessel formation by the effects of the tumor on its immediate environment. Stimuli for tumor angiogenesis include hypoxemia, inflammation, and genetic lesions in oncogenes or tumor suppressors that alter tumor cell gene expression. Angiogenesis consists of several steps, including the stimulation of ECs by growth factors, degradation of the ECM by proteases, proliferation and migration of ECs into the tumor, and the eventual formation of new capillary tubes. Tumors use a number of mechanisms to promote vascularization, subverting normal angiogenic processes for this purpose **(Fig. 77-9)**. Primary or metastatic tumor cells sometimes arise in proximity to host blood vessels and grow around these vessels, parasitizing nutrients by co-opting the local blood supply. However, most tumor blood vessels arise by the process of sprouting, in which tumors secrete trophic angiogenic molecules, the most potent being vascular endothelial growth factors (VEGFs), that induce the proliferation and migration of host ECs into the tumor. Sprouting in normal and pathogenic angiogenesis is regulated by three families of transmembrane RTKs expressed on ECs and their ligands (VEGFs, angiopoietins, and ephrins; **Fig. 77-10**), which are produced by tumor cells, inflammatory cells, or stromal cells in the tumor microenvironment.

Central to the angiogenic response are hypoxia-inducible factors (HIFs; especially 1 and 2), which are transcription factors that normally, in response to hypoxia, stimulate the transcription of a large number of genes responsive to hypoxia, including genes involved in metabolism as well as angiogenesis. HIF1 has a bigger role in stimulating metabolism (glycogenesis), whereas HIF2 plays a bigger role in angiogenesis. HIF protein function can also be enhanced in a number of ways in cancer not involving hypoxia, including mutations in the von Hippel–Lindau tumor suppressor gene (an E3 ubiquitin ligase that controls HIF levels by targeting it for degradation), such as occurs in some RCCs. Among the genes stimulated by HIF are VEGF and VEGF receptors. VEGFs and their receptors are required for embryonic vasculogenesis (development of new blood vessels when none preexist) and normal (wound healing, corpus luteum formation) and pathologic angiogenesis (tumor angiogenesis, inflammatory conditions such as rheumatoid arthritis). VEGF-A is a heparin-binding glycoprotein with at least four isoforms (splice variants) that regulates blood vessel formation by binding to the RTKs VEGFR1 and VEGFR2, which are expressed on all ECs in addition to a subset of hematopoietic cells (Fig. 77-9). VEGFR2 plays a more direct role in regulating EC proliferation, migration, and survival, whereas VEGFR1 appears to have more nuanced functions with a less direct role in stimulating EC processes in the normal adult (even acting as a decoy protein for VEGFA to decrease binding to VEGFR2) but with important effects during embryogenesis and on tumor angiogenesis. Tumor vessels may be more dependent on VEGFR signaling for growth and survival than normal ECs.

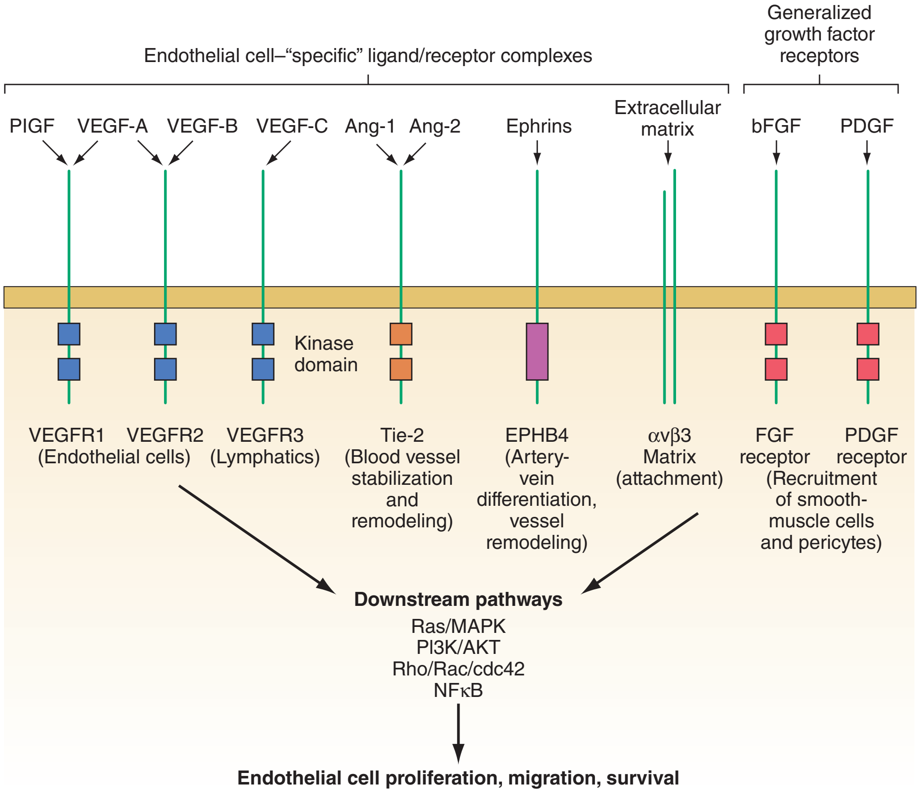
Endothelial cell–"specific" ligand/receptor complexes
Generalized growth factor receptors
PlGF
VEGF-A
VEGF-B
VEGF-C
Ang-1
Ang-2
Ephrins
Extracellular matrix
bFGF
PDGF
Kinase domain
VEGFR1
VEGFR2
(Endothelial cells)
VEGFR3
(Lymphatics)
Tie-2
(Blood vessel stabilization and remodeling)
EPHB4
(Artery-vein differentiation, vessel remodeling)
αvβ3
Matrix
(attachment)
FGF receptor
PDGF receptor
(Recruitment of smooth-muscle cells and pericytes)
Downstream pathways
Ras/MAPK
PI3K/AKT
Rho/Rac/cdc42
NFκB
Endothelial cell proliferation, migration, survival

FIGURE 77-10 Critical molecular determinants of endothelial cell biology. Angiogenic endothelium expresses a number of receptors not found on resting endothelium. These include receptor tyrosine kinases (RTKs) and integrins that bind to the extracellular matrix and mediate endothelial cell (EC) adhesion, migration, and invasion. ECs also express RTKs (i.e., the fibroblast growth factor [FGF] and platelet-derived growth factor [PDGF] receptors) that are found on many other cell types. Critical functions mediated by activated RTK include proliferation, migration, and enhanced survival of endothelial cells, as well as regulation of the recruitment of perivascular cells and bloodborne circulating endothelial precursors and hematopoietic stem cells to the tumor. Intracellular signaling via EC-specific RTK utilizes molecular pathways that may be targets for future antiangiogenic therapies.

While VEGF signaling is a critical initiator of angiogenesis, this is a complex process regulated by additional signaling pathways (Fig. 77-10). The angiopoietin, Ang1, produced by stromal cells, binds to the EC RTK Tie-2 and promotes the interaction of ECs with the ECM and perivascular cells, such as pericytes and smooth-muscle cells, to form tight, nonleaky vessels. PDGF and basic fibroblast growth factor (bFGF) help to recruit these perivascular cells. Ang1 is required for maintaining the quiescence and stability of mature blood vessels and prevents the vascular permeability normally induced by VEGF and inflammatory cytokines. For tumor cell–derived VEGF to initiate sprouting from host vessels, the stability conferred by the Ang1/Tie2 pathway must be perturbed; this occurs by the secretion of Ang2 by ECs that are undergoing active remodeling. Ang2 binds to Tie2 and is a competitive inhibitor of Ang1 action: under the influence of Ang2, preexisting blood vessels become more responsive to remodeling signals, with less adherence of ECs to stroma and associated perivascular cells and more responsiveness to VEGF. Therefore, Ang2 is required at early stages of tumor angiogenesis for destabilizing the vasculature by making host ECs more sensitive to angiogenic signals. In the presence of Ang2, there is no stabilization by the Ang1/Tie2 interaction, and tumor blood vessels are leaky, hemorrhagic, and have poor association of ECs with underlying stroma. Sprouting tumor ECs express high levels of the transmembrane protein ephrin-B2 and its receptor, the RTK EPH, whose signaling appears to work with the angiopoietins during vessel remodeling. During embryogenesis, EPH receptors are expressed on the endothelium of primordial venous vessels while the transmembrane ligand ephrin-B2 is expressed by cells of primordial arteries; the reciprocal expression may regulate differentiation and patterning of the vasculature.

A number of additional ubiquitously expressed host molecules play critical roles in normal and pathologic angiogenesis. Proangiogenic cytokines, chemokines, and growth factors secreted by stromal cells or inflammatory cells make important contributions to neovascularization, including bFGF, transforming growth factor-β (TGF-β), TNF-α, and IL-8. In contrast to normal endothelium, angiogenic endothelium overexpresses specific members of the integrin family of ECM-binding proteins that mediate EC adhesion, migration, and survival. Specifically, expression of integrins αvβ3, αvβ5, and α5β1 mediates spreading and migration of ECs and is required for angiogenesis induced by VEGF and bFGF, which in turn can upregulate EC integrin expression. The αvβ3 integrin physically associates with VEGFR2 in the plasma membrane and promotes signal transduction from each receptor to promote EC proliferation (via focal adhesion kinase, src, PI3K, and other pathways) and survival (by inhibition of p53 and increasing the Bcl-2/Bax expression ratio). In addition, αvβ3 forms cell-surface complexes with matrix metalloproteinases (MMPs), zinc-requiring proteases that cleave ECM proteins, leading to enhanced EC migration and the release of heparin-binding growth factors, including VEGF and bFGF. EC adhesion molecules can be upregulated (i.e., by VEGF, TNF-α) or downregulated (by TGF-β); this, together with chaotic blood flow, explains poor leukocyte-endothelial interactions in tumor blood vessels and may help tumor cells avoid immune surveillance.

Tumor blood vessels are not normal; they have chaotic architecture and blood flow. Due to an imbalance of angiogenic regulators such as VEGFs and angiopoietins (see below), tumor vessels are tortuous and dilated with an uneven diameter, excessive branching, and shunting. Tumor blood flow is variable, with areas of hypoxemia and acidosis leading to the selection of cancer cell variants that are resistant to hypoxemia-induced apoptosis (often involving the loss of p53 expression). Tumor vessel walls have numerous openings, widened interendothelial junctions, and discontinuous or absent basement membrane. This contributes to the high permeability of these vessels and, together with lack of functional intratumoral lymphatics, causes increased interstitial pressure within the tumor (which also interferes with the delivery of therapeutics to the tumor; Figs. 77-9, 77-10, and 77-11). Tumor blood vessels have a deficit of perivascular cells such as pericytes and smooth-muscle cells that normally regulate flow in response to tissue metabolic needs. Unlike normal blood vessels, the vascular lining of tumor vessels is not a homogeneous layer of ECs but often consists of a mosaic of ECs and tumor cells, which, because of their plasticity, can upregulate expression of genes normally only seen in ECs under hypoxic conditions. These cancer cell–derived vascular channels, which may be lined by ECM secreted by the tumor cells, are referred to as vascular mimicry. During tumor angiogenesis, ECs are highly proliferative and express a number of plasma membrane proteins that are characteristic of activated endothelium, including growth factor receptors and adhesion molecules such as integrins. These abnormalities in tumor vasculature provide potential differential sensitivities from normal vessels to approaches inhibiting the process, allowing for the use of antiangiogenic agents in cancer treatment.

Lymphatic vessels also exist within tumors. Development of tumor lymphatics is associated with expression of VEGFR3 and its ligands VEGF-C and VEGF-D. The role of these vessels in tumor cell metastasis to regional lymph nodes remains to be determined. However, VEGF-C levels correlate significantly with metastasis to regional lymph nodes in lung, prostate, and colorectal cancers.

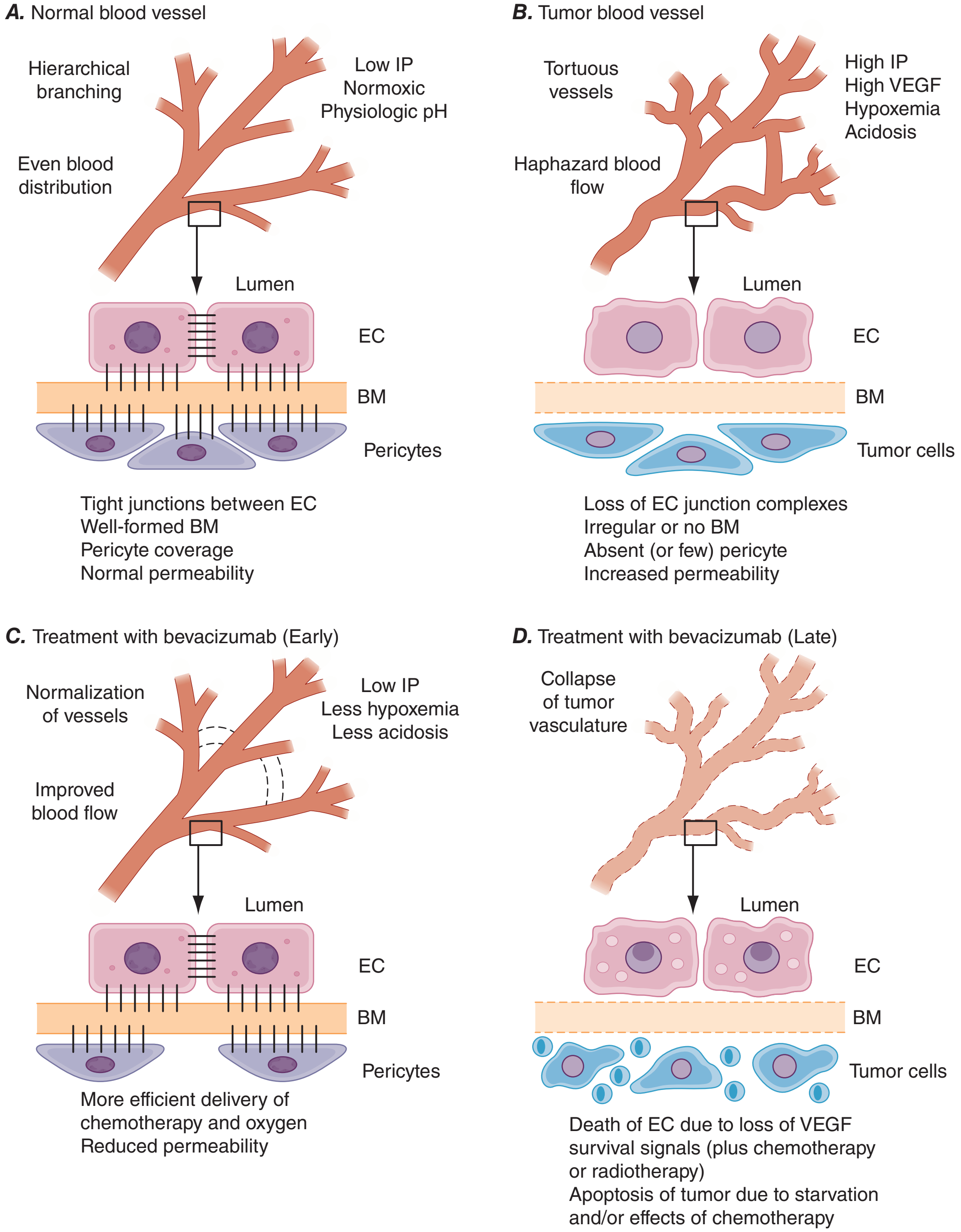
A. Normal blood vessel
Hierarchical branching
Low IP
Normoxic
Physiologic pH
Even blood distribution
Lumen
EC
BM
Pericytes
Tight junctions between EC
Well-formed BM
Pericyte coverage
Normal permeability
B. Tumor blood vessel
Tortuous vessels
High IP
High VEGF
Hypoxemia
Acidosis
Haphazard blood flow
Lumen
EC
BM
Tumor cells
Loss of EC junction complexes
Irregular or no BM
Absent (or few) pericyte
Increased permeability
C. Treatment with bevacizumab (Early)
Normalization of vessels
Low IP
Less hypoxemia
Less acidosis
Improved blood flow
Lumen
EC
BM
Pericytes
More efficient delivery of chemotherapy and oxygen
Reduced permeability
D. Treatment with bevacizumab (Late)
Collapse of tumor vasculature
Lumen
EC
BM
Tumor cells
Death of EC due to loss of VEGF survival signals (plus chemotherapy or radiotherapy)
Apoptosis of tumor due to starvation and/or effects of chemotherapy

FIGURE 77-11 Normalization of tumor blood vessels due to inhibition of VEGF signaling. ***A.*** Blood vessels in normal tissues exhibit a regular hierarchical branching pattern that delivers blood to tissues in a spatially and temporally efficient manner to meet the metabolic needs of the tissue (top). At the microscopic level, tight junctions are maintained between endothelial cells (ECs), which are adherent to a thick and evenly distributed basement membrane (BM). Pericytes form a surrounding layer that provides trophic signals to the EC and helps maintain proper vessel tone. Vascular permeability is regulated, interstitial fluid pressure (IP) is low, and oxygen tension and pH are physiologic. ***B.*** Tumors have abnormal vessels with tortuous branching and dilated, irregular interconnecting branches, causing uneven blood flow with areas of hypoxemia and acidosis. This harsh environment selects genetic events that result in resistant tumor variants, such as the loss of p53. High levels of VEGF (secreted by tumor cells) disrupt gap junction communication, tight junctions, and adherens junctions between EC via src-mediated phosphorylation of proteins such as connexin 43, zonula occludens-1, VE-cadherin, and α/β-catenins. Tumor vessels have thin, irregular BM, and pericytes are sparse or absent. Together, these molecular abnormalities result in a vasculature that is permeable to serum macromolecules, leading to high tumor interstitial pressure, which can prevent the delivery of drugs to the tumor cells. This is made worse by the binding and activation of platelets at sites of exposed BM, with release of stored VEGF and microvessel clot formation, creating more abnormal blood flow and regions of hypoxemia. ***C.*** In experimental systems, treatment with bevacizumab or blocking antibodies to VEGFR2 leads to changes in the tumor vasculature that have been termed *vessel normalization*. During the first week of treatment, abnormal vessels are eliminated or pruned (*dotted lines*), leaving a more normal branching pattern. ECs partially regain features such as cell-cell junctions, adherence to a more normal BM, and pericyte coverage. These changes lead to a decrease in vascular permeability, reduced interstitial pressure, and a transient increase in blood flow within the tumor. Note that in murine models, this normalization period lasts only for ~5–6 days. ***D.*** After continued anti-VEGF/VEGFR therapy (which is often combined with chemo- or radiotherapy), ECs die, leading to tumor cell death (either due to direct effects of the chemotherapy or lack of blood flow).

■ ANTIANGIOGENIC THERAPY

Angiogenesis inhibitors function by targeting the critical molecular pathways involved in EC proliferation, migration, and/or survival, many of which are highly expressed in the activated endothelium in tumors. Inhibition of growth factor and adhesion-dependent signaling pathways can induce EC apoptosis with concomitant inhibition of tumor growth. Different types of tumors can use distinct combinations of molecular mechanisms to activate the angiogenic switch. Therefore, it is doubtful that a single antiangiogenic strategy will suffice for all human cancers; rather, a number of agents or combinations of agents

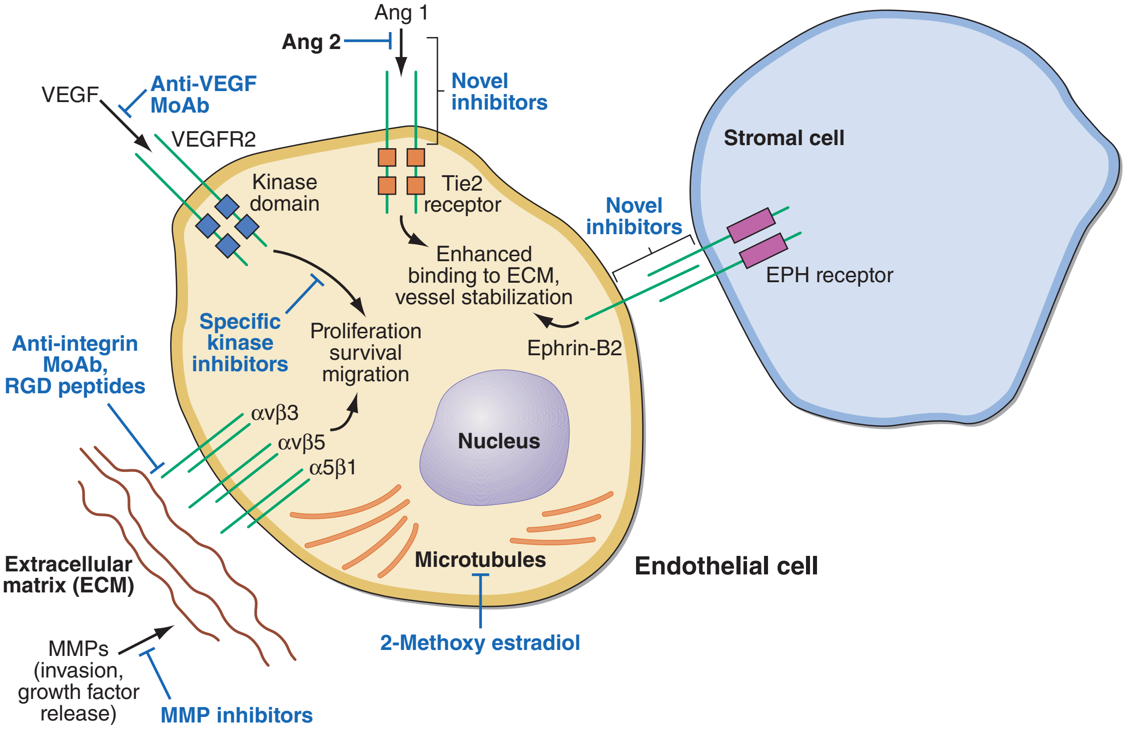

FIGURE 77-12 Knowledge of the molecular events governing tumor angiogenesis has led to a number of therapeutic strategies to block tumor blood vessel formation. The successful therapeutic targeting of VEGF and its receptors VEGFR is described in the text. Other endothelial cell (EC)–specific receptor tyrosine kinase pathways (e.g., angiopoietin/Tie2 and ephrin/EPH) are likely targets for the future. Ligation of the $\alpha_v\beta_3$ integrin is required for EC survival. Integrins are also required for EC migration and are important regulators of matrix metalloproteinase (MMP) activity, which modulates EC movement through the ECM as well as release of bound growth factors. Targeting of integrins includes development of blocking antibodies, small peptide inhibitors of integrin signaling, and arg-gly-asp–containing peptides that prevent integrin:ECM binding. Peptides derived from normal proteins by proteolytic cleavage, including endostatin and tumstatin, inhibit angiogenesis by mechanisms that include interfering with integrin function. Signal transduction pathways that are dysregulated in tumor cells indirectly regulate EC function. Inhibition of EGF-family receptors, whose signaling activity is upregulated in a number of human cancers (e.g., breast, colon, and lung cancers), results in downregulation of VEGF and IL-8, while increasing expression of the antiangiogenic protein thrombospondin-1. The Ras/MAPK, PI3K/Akt, and Src kinase pathways constitute important antitumor targets that also regulate the proliferation and survival of tumor-derived EC. The discovery that ECs from normal tissues express tissue-specific "vascular addressins" on their cell surface suggests that targeting specific EC subsets may be possible.

will be needed, depending on distinct programs of angiogenesis used by different human cancers. Despite this, experimental data indicate that for some tumor types, blockade of a single growth factor (e.g., VEGF) may inhibit tumor-induced vascular growth.

Bevacizumab, an antibody that binds circulating VEGF, modestly potentiates the effects of a number of different types of active chemotherapeutic regimens used to treat a variety of different tumor types including colon, lung, ovarian, and cervical cancers. It also has some activity in combination with immunotherapy against RCCs and alone for glioblastomas. Other protein inhibitors of the VEGF signaling pathway approved for anticancer therapy include ramucirumab (a monoclonal antibody directed against VEGFR2, approved for use against gastric/gastroesophageal, colon, and lung cancers) and ziv-aflibercept (a recombinant protein inhibitor of VEGF, approved for colorectal cancer). Hypertension is the most common side effect of inhibitors of VEGF (or its receptors) but can be treated with antihypertensive agents and uncommonly requires discontinuation of therapy. Rare but serious potential risks include arterial thromboembolic events, including stroke and myocardial infarction, hemorrhage, bowel perforation, and inhibition of wound healing.

Several small-molecule inhibitors (SMIs) that target VEGF RTK activity but are also inhibitory to other kinases have also been approved to treat certain cancers. Sunitinib (see above and Table 77-2) has activity directed against mutant c-Kit receptors (approved for GIST), but also targets VEGFR and PDGFR, and has antitumor activity against pancreatic neuroendocrine and metastatic RCCs, presumably on the basis of its antiangiogenic activity. Similarly, sorafenib, originally developed as a Raf kinase inhibitor but with potent activity against VEGFR and PDGFR, has activity against RCC, differentiated thyroid and hepatocellular cancers, and desmoid tumors. A closely related molecule to sorafenib, regorafenib, has activity against colorectal cancer, GIST, gastric, and hepatocellular cancers. Other inhibitors of the VEGF pathway approved for the treatment of various cancers include axitinib, pazopanib, lenvatinib, and cabozantinib.

Antiangiogenic agents have been particularly effective against RCC for which angiogenic factors are important for its development and growth. The modest success in targeting tumor angiogenesis against most other cancers has led to enhanced enthusiasm for the development of drugs that target other aspects of the angiogenic process; some of these therapeutic approaches are outlined in **Fig. 77-12**. An inhibitor of HIF2-α has sufficient antitumor activity against RCC, pancreatic neuroendocrine tumors, and hemangioblastomas developing in patients with germline VHL mutations to be approved for these indications.

Evidence of enhanced activity has been seen when anti-VEGF agents are used in combination with immunomodulators including immune checkpoint inhibitors. Examples of approved combinations include durvalumab plus bevacizumab for HCC and lenvatinib plus pembrolizumab for endometrial cancer and RCC.

EVASION OF THE IMMUNE SYSTEM BY CANCERS

The immune system plays a critical role in maintaining organismal integrity including by defending against pathogens as well as preventing and limiting the growth of cancers. There is a complex interaction between cancer and the host from the development of the first malignant cell to the establishment of a clinical cancer and its subsequent growth, invasion, and metastasis. The immune system plays a critical role in the prevention of cancer development. This is exemplified by the increased risk for cancer development in individuals who are significantly immunosuppressed, such as by inherited defects in mechanisms important for immune function, the immunosuppression necessary to maintain allogeneic organ transplants, and immunosuppression seen from certain infections such as human immunodeficiency virus. It also plays a critical role in inhibiting the process of cells metastasizing as well as growth of metastatic cells at the sites of metastasis.

There are two components of the immune system. The first is innate immunity (present in the organism and not dependent on prior exposure to a specific antigen, such as those present in a pathogen or malignant cell), which tends to be general and not specific and a stimulus is not remembered if encountered again. The second is the adaptive immune component, which depends on the innate immune process for activation and provides the specificity to the response with significant expansion of cells to target the specific antigens present on the pathogen or malignant cell and memory of the encounter such that exposure to the same stimulus elicits an even more rapid and vigorous response. Thus, while the innate process provides the first line of defense, the adaptive process is necessary for the specificity of response and providing memory to more rapidly attack cells should the pathogen infection recur or the malignant cells grow. The immune system has to be tightly regulated to allow for clearance of unwanted antigens while preventing an immune-mediated attack on the self. **(See Chap. 360 for details on the function of the immune system.)**

Not surprisingly, since cancers arise from normal cells within the body that have a variety of processes to prevent harm or destruction by the immune system, they have a variety of mechanisms that allow them to evade detection and elimination by the immune system. These include downregulation of cell surface proteins involved in immune recognition (including MHC proteins and tumor-specific antigens), expression of other cell surface proteins that inhibit immune function (including members of the B7 family of proteins such as PD-L1), secretion of proteins and other molecules that are immunosuppressive such as TGF-β, recruitment and expansion of immunosuppressive cells such as regulatory T cells (which are important for maintaining tolerance against self-antigens), induction of T-cell tolerance, and downregulation of death receptors. Due to the marked heterogeneity of cells within a cancer, as well as the complexity and dynamic changes in the tumor microenvironment, a variety of immune-suppressive mechanisms are continuously occurring and changing. In addition, the inflammatory effects of some of the immune mediator cells in the tumor microenvironment (including tissue-associated macrophages and myeloid-derived suppressor cells) can suppress effector T-cell responses against the tumor as well as stimulate inflammation that can enhance tumor growth.

There are marked differences in the way different malignancies respond to current immunotherapeutic approaches. For example, melanomas, RCC, Merkel cell carcinomas, cancers with defects in DNA repair associated with microsatellite instability with accumulation of gene mutations, and lymphomas (including Hodgkin's) respond well to current immunotherapeutic approaches, whereas microsatellite-stable pancreatic and colon cancers do not. While there is not a complete understanding of why these differences exist and many factors both within the cancer cells and in the microenvironment may play a role, several factors have been identified that appear to be important. These include the number of mutations present in the tumor (tumor mutational burden), presence of increased neoantigens, expression of immune checkpoint proteins (e.g., PD-L1 for anti-PD-1 or anti-PD-L1 therapy), density of tumor-infiltrating lymphocytes, and host genetic factors. One of these (PD-L1 expression by the tumor) has sufficient predictive value for certain tumors (e.g., non-small-cell lung cancer or gastroesophageal cancers) to be used in making treatment decisions regarding the use of antibodies targeting PD-1 or PD-L1. However, neither PD-L1 expression nor any other marker can predict responsiveness of most tumors to immunotherapy. Better biomarkers that define potential responsiveness of specific cancers to immunotherapy are badly needed. A major area of research is to try to identify approaches that would convert cancers that are not responsive to immunotherapy to being responsive.

Immunotherapy approaches to treat cancer can be divided into those aimed at activating the immune response and those designed to release the brakes that prevent an effective immune response against tumors. Releasing the brakes is also important for maintaining the effectiveness of approaches that activate the immune response since, given the normally tight regulation of immune function, activation induces changes in the braking system to prevent the immune system from damaging normal tissues. Approaches at activating the immune response against cancer including using immunostimulatory molecules such as interferons, IL-2, and especially monoclonal antibodies have had success in treating a number of different cancers. For example, antibodies that target molecules highly expressed on certain cancers, such as CD20 on malignant B cells or HER2 on a variety of cancers including breast and gastroesophageal cancers, which activate the immune response locally against those malignancies, have proven highly effective.

A more direct approach to enhance the activity of T cells directed against specific tumors involves isolating T cells from patients and reengineering the cells to express chimeric antigen receptors (CAR-T) that recognize antigens present on the cells of that individual's tumor. The most commonly used approach to date has been to engineer the cells to express receptors targeting the CD19 antigen on ALL, diffuse large B-cell lymphoma (DLBCL) cells, follicular lymphoma, and mantle cell lymphoma. These have been shown to have significant antitumor activity in the treatment of patients with ALL and DLBCL, including durable remissions in patients refractory to standard therapies, and are approved for these malignancies. In addition, anti-B-cell maturation antigen (BCMA) CAR-T therapies have been approved for the treatment of multiple myeloma.

However, there have also been significant issues with toxicity including cytokine release syndrome, organ toxicity felt to be due to inadvertent targeting of antigens present in the organ, neurotoxicity, and potentially an increased risk for subsequent development of T-cell malignancies. These patients often require aggressive supportive care by individuals experienced in the delivery of CAR-T therapy. In addition, as is true for most anticancer therapies, mechanisms of resistance have developed, most commonly the outgrowth of tumor cells no longer expressing the antigen. Mechanisms for preventing the development of resistant cells are being explored, including combinations targeting different antigens. In addition to potentially preventing or overcoming resistance to the targeting of a single antigen, this could potentially increase efficacy and better reflects the normal immune response to pathogens or cancers in targeting multiple different antigens. CAR-T therapies are undergoing clinical investigation against other hematologic malignancies and solid tumors. Approaches to develop allogeneic CAR-T therapies are also being explored with the aim of having an off-the-shelf product that could be used in a number of patients rather than generating each treatment specifically for one recipient.

Another approach utilizing lymphocytes to treat cancer involves utilization of autologous tumor-derived T cells. Tumor-infiltrating lymphocyte therapy in combination with IL-2 is now approved for the treatment of melanoma. Given previously demonstrated efficacy against other cancers, such as RCC, tumor-infiltrating lymphocyte therapy may eventually be approved for other cancers as well. However, technical issues of getting adequate expansion of the cells may limit this approach.

The immune response against cancers may also be able to be enhanced through targeting of proteins or cells (e.g., regulatory T cells) involved in normal homeostatic control to prevent autoimmune damage to the host but that malignant cells and their stroma can also utilize to inhibit the immune response directed against them. A component of this process includes a number of immune checkpoints that involve interaction of proteins on the surface of effector T cells with proteins on self-cells (or cancer cells that arise from normal cells) that inhibit activation of the T cells. By inhibiting the binding of the proteins involved in this process, the brake on the effector T cells are released and they can be activated. The presence of neoantigens (e.g., mutant proteins) enhances the activation of the T cells against cancer cells as compared to normal cells. Sufficient clinical antitumor activity has been seen for monoclonal antibodies targeting various proteins involved in this process, including CTLA-4, PD-1, PD-L1, and LAG3 (others continue to be explored), for them to be approved. These are

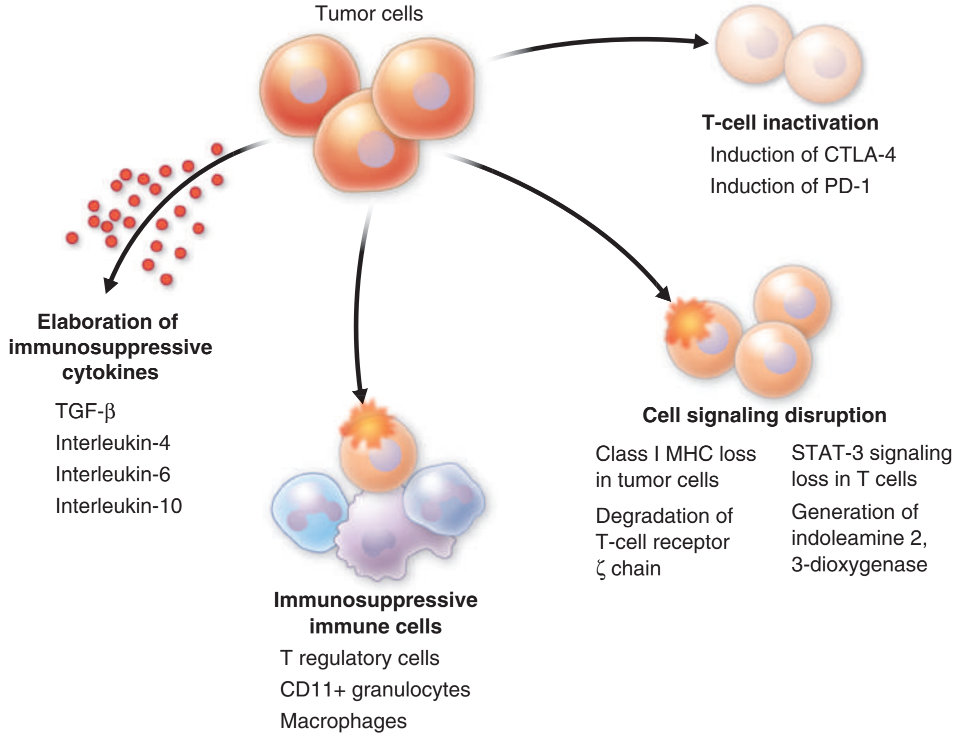
Tumor cells
T-cell inactivation
Induction of CTLA-4
Induction of PD-1
Elaboration of immunosuppressive cytokines
TGF-β
Interleukin-4
Interleukin-6
Interleukin-10
Cell signaling disruption
Class I MHC loss in tumor cells
Degradation of T-cell receptor ζ chain
STAT-3 signaling loss in T cells
Generation of indoleamine 2, 3-dioxygenase
Immunosuppressive immune cells
T regulatory cells
CD11+ granulocytes
Macrophages

FIGURE 77-13 Tumor-host interactions that suppress the immune response to the tumor.

co-inhibitory molecules that are expressed on the surface of cancer cells, and/or cells of the immune system, and/or stromal cells and are involved in inhibiting the immune response against both normal cells (their normal protective mechanism for the host) and also cancer cells that use this inhibitory process to evade immune-mediated cell death (Figs. 77-13 and 77-14). This approach has had clinical activity against a wide variety of cancers. A monoclonal antibody directed against CTLA-4 is approved for the treatment of melanoma and several other malignancies, and antibodies targeting PD-1 or PD-L1 are approved for use against melanoma, RCC, lung cancer (both non-small-cell lung and small-cell lung), head and neck cancer, nasopharyngeal cancer, urothelial cancer, cervical cancer, endometrial cancer, hepatocellular carcinoma, gastric cancer, esophageal cancer, cutaneous squamous cell carcinoma, basal cell carcinoma, Merkel cell cancer, primary B-cell mediastinal lymphoma, Hodgkin's lymphoma, and in a cancer-agnostic approach, cancers with high microsatellite instability (MSI) or high tumor mutational burden (TMB).

They continue to be evaluated against other malignancies as well. The combination of anti-CTLA-4 and anti-PD-1 antibodies has been approved for treatment of a number of malignancies, including melanoma, RCC, HCC, NSCLC, mesothelioma, and MSI-high metastatic colorectal cancers. Specific determinants of response to immune checkpoint inhibitors are still being defined, but in addition to high PD-L1 expression, the presence of increased neoantigens in the tumor, such as seen in patients with MSI-high and TMB-high cancers, may be one important determinant of better responses.

A number of other proteins are involved in controlling the immune response (both ones that enhance activity [e.g., CD27 and CD40] as well as ones involved in inhibiting response [e.g., TIM-3, TIGIT]). Antibodies have been developed to modulate function of these

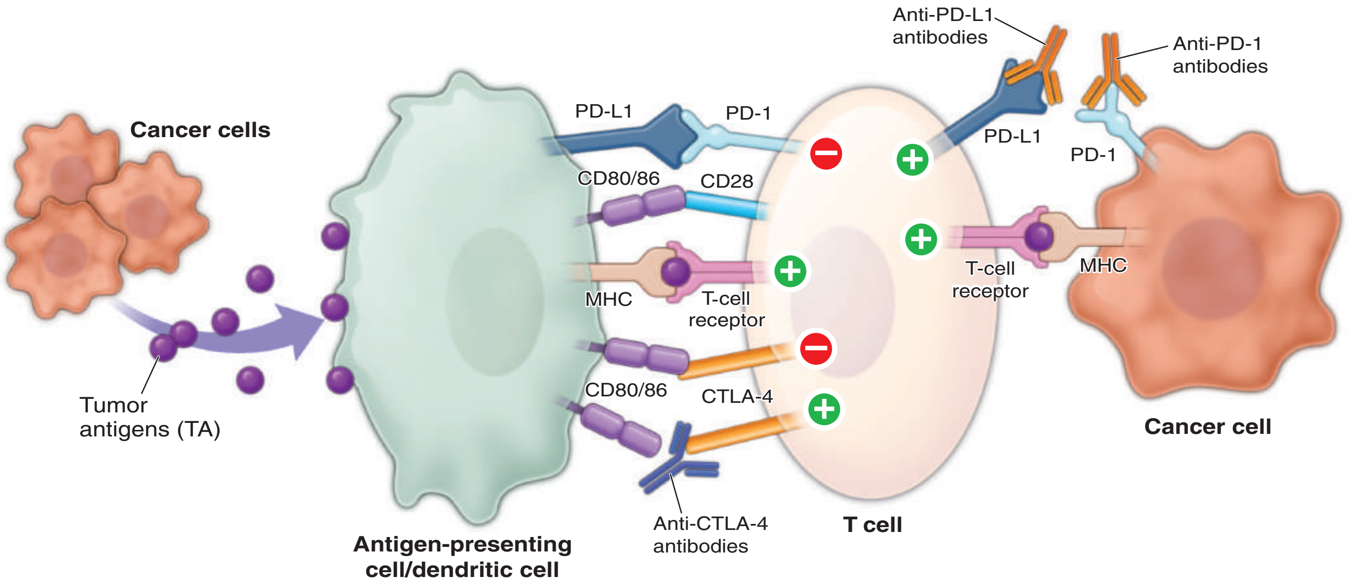
Anti-PD-L1 antibodies
Anti-PD-1 antibodies
Cancer cells
PD-L1
PD-1
PD-L1
PD-1
CD80/86
CD28
MHC
T-cell receptor
T-cell receptor
MHC
CD80/86
CTLA-4
Tumor antigens (TA)
Cancer cell
Anti-CTLA-4 antibodies
T cell
Antigen-presenting cell/dendritic cell

FIGURE 77-14 Inhibition of T-cell activation against cancer cells by engagement of co-inhibitory molecules including PD-1, PD-L1, and CTLA-4 and reversal of this inhibition by antibodies against these proteins. The red ovals in the T cell indicate inhibitory signals, and the green oval indicates stimulatory signals.

proteins, and many are in clinical development for cancer therapy. In addition, various combinations targeting more than one protein involved in modulating the immune response against cancers or with other anticancer approaches (targeted agents, chemotherapy, radiation therapy) that may lead to enhanced antitumor activity are also being explored. An important aspect of these approaches is balancing sufficient release of the negative control of the immune response to allow immune-mediated attack on the tumors while not allowing too much of an immune response against normal tissues and thus inducing severe autoimmune effects (e.g., against lung, liver, skin, thyroid, pituitary gland, gastrointestinal tract, or other organs). As is true for other immunotherapeutic approaches against cancer, major efforts are ongoing to better understand the mechanism of immune toxicity from these approaches and, therefore, ways of controlling this while not abrogating the antitumor effects. Improved knowledge of the biology of the interactions between the immune system and cancers continues to be rapid with the promise for additional significant improvements in use of immunotherapy to treat cancer.

Given the success of vaccines against multiple viruses, many efforts have been made over decades to develop anticancer vaccines that would induce a sufficient immune response to lead to killing of cancer cells or even potentially prevention. This has been largely unsuccessful for the treatment of established cancers, although three vaccines are approved for use in cancer care: bacillus Calmette-Guérin (BCG) for intravesical treatment of superficial bladder cancer, sipuleucel-T for prostate cancer, and talimogene laherparepvec for direct injection into melanoma. Except for BCG, and then only for superficial bladder cancer, these have had relatively limited effectiveness. Perhaps more promising are vaccines combined with other forms of immunotherapy. Benefit may be enhanced by using a vaccine in combination with immune checkpoint inhibition therapy to delay recurrence of resected high-risk melanoma and provides hope that this approach may be effective in the future. However, further studies are needed to confirm this.

IMPACT OF UNDERSTANDING CANCER BIOLOGY ON PREVENTION AND EARLY DETECTION

The biggest impacts on decreasing cancer mortality come from prevention followed by early detection. In addition to the critical role that new knowledge about cancer biology has in developing improved treatment approaches, it is also critical in enhancing preventative and early detection approaches. Examples in prevention include using knowledge about the roles of certain viruses in the development of some cancers to create vaccines for viruses (e.g., hepatitis B virus and human papillomavirus) to decrease the risk of developing cancer from these infections and using knowledge about how certain carcinogens cause cancer to work on controlling or eliminating carcinogenic agents (e.g., cigarette smoking and asbestos). The development of sensitive methods for detecting circulating tumor DNA holds promise for earlier detection of cancer, although this has not yet been firmly established. Imaging agents for specific tissues (such as the use of proteins highly expressed on specific cancer tissues including prostate-specific membrane antigen PET imaging to detect prostate cancer or HER2 to detect a subset of breast, gastric, and other cancers) or potentially against mutant proteins found in cancers (e.g., *KRAS* or *p53* mutations) hold promise for earlier and more specific detection of cancer. As more is learned about cancer biology, this knowledge will also continue to be applied in preventative and early detection strategies.

SUMMARY

Just as human biology is complex, cancer biology is complex. Although each of the biological aspects of cancers and examples of targeting them has been addressed individually, clearly there is complicated cross-talk between these that occurs in all cancers that needs to be better understood to optimally treat different cancers. The explosion of information on tumor cell biology, metastasis, and tumor-host interactions (including angiogenesis, other tumor-stromal interactions, and immune evasion by tumors) has ushered in a new era of rational targeted therapy for cancers as well as the potential for individualized therapy. Furthermore, it has become clear that specific molecular factors detected in individual tumors (gene mutations, gene expression profiles, miRNA expression, overexpression of specific proteins) can be used to tailor therapy and maximize antitumor effects.

Potentially of even greater impact on decreasing deaths from cancer, better understanding of the biology of early cancer development should ultimately lead to better approaches for prevention. And technologic development to improve sensitivity and specificity in detecting cancer-specific molecules (e.g., mutated genes) provide hope that approaches for earlier detection of cancer can be developed.

FURTHER READING

Agudo J et al: Targeting cancer cell dormancy. Nat Rev Cancer 24:97, 2024.

Bhullar KS et al: Kinase-targeted cancer therapies: Progress, challenges and future directions. Mol Cancer 17:48, 2018.

de Visser KE, Joyce JA: The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. Cancer Cell 41:374, 2023.

Finley LWS: What is cancer metabolism? Cell 186:1670, 2023.

Fujii M et al: Decoding the basis of histological variation in human cancer. Nat Rev Cancer 24:141, 2024.

Gacche RN: Changing landscape of anti-angiogenic therapy: Novel approaches and clinical perspectives. Biochim Biophys Acta Rev Cancer 1878:189020, 2023.

Gerstberger S et al: Metastasis. Cell 186:1564, 2023.

Hanahan D: Hallmarks of cancer: New dimensions. Cancer Discov 12:31, 2022.

Matthews HK et al: Cell cycle control in cancer. Nat Rev Mol Cell Biol 23:74, 2022.

Pottier C et al: Tyrosine kinase inhibitors in cancer: Breakthrough and challenges of targeted therapy. Cancers (Basel) 12:731, 2020.

Prager BC et al: Cancer stem cells: The architects of the tumor ecosystem. Cell Stem Cell 24:41, 2019.

Reiter JG et al: An analysis of genetic heterogeneity in untreated cancers. Nat Rev Cancer 19:639, 2019.

Romesser PB, Lowe SW: The potent and paradoxical biology of cellular senescence in cancer. Annu Rev Cancer Biol 7:207, 2023.

Sharma P, et al: Immune checkpoint therapy: Current perspectives and future directions. Cell 186:1652, 2023.

Tomuleasa C et al: Therapeutic advances of targeting receptor tyrosine kinases in cancer. Sig Transduct Target Ther 9:201, 2024.

Yang K et al: Antigen presentation in cancer: Mechanisms and clinical implications for immunotherapy. Nat Rev Clin Oncol 20:604, 2023.

Yuan J, Ofengeim D: A guide to cell death pathways. Nat Rev Mol Cell Biol 25:379, 2024.

78

Principles of Cancer Treatment

Edward A. Sausville, Dan L. Longo

CANCER PRESENTATION

Localized or systemic cancer is frequent in the differential diagnosis of a variety of common complaints. Affording patients the greatest opportunity for cure or meaningful prolongation of life is greatly aided by diagnosis of cancer early in its natural history. The spectrum of possible cancer-related interventions to make cure possible is shown in Table 78-1.

DETECTION OF A CANCER

The term *cancer*, as used here, is synonymous with the term *tumor*, whose original derivation from Latin simply meant "swelling," not

TABLE 78-1 Spectrum of Cancer-Related Interventions

- Asymptomatic patient screening (breast, cervix, colon, some lung)
- Consideration of cancer in a differential diagnosis
- Physical examination, imaging, or endoscopy to define a possible tumor
- Phlebotomy for molecular studies and circulating tumor cell characterization
- Diagnosis of cancer by biopsy or removal:
  - Routine histology
  - Specialized histology: immunohistochemistry
  - Molecular studies
  - Cytogenetic studies
- Staging the cancer: Where has it spread?
- Imaging (computed tomography, magnetic resonance imaging, positron emission tomography)
- Treatment
  - Localized (surgical removal with or without local radiation therapy and/or topical therapy; may be curative)
  - Local plus systemic, multimodality: cure advanced disease, reverse organ compromise
  - Systemic, adjuvant (cure micrometastases; all evident disease has been locally treated)
  - Systemic, neoadjuvant (before local therapy to improve local and systemic results)
  - Systemic, palliative (improve symptoms, quality of life, progression-free survival)
- Supportive care
  - During treatment: related to tumor effects on patient
  - During treatment: to counteract side effects of treatment
- After treatment: to ameliorate the adverse effects of treatment
- Palliative and end of life
  - When useful treatments are not feasible or desired

otherwise specified. Swelling reflects increased interstitial fluid pressure and increased cellular and stromal mass, compared to normal tissue. Leukemia, a cancer of the blood-forming tissues, presents in a disseminated form frequently without tumor masses. Tumors can also present by organ dysfunction, such as dyspnea on exertion from anemia caused by leukemia replacing normal marrow, cough from lung cancers, jaundice from tumors blocking bile ducts, or neurologic signs from gliomas. Hemorrhage frequently results from involvement of hollow viscera, but also may reflect altered platelet number or blood coagulation. Tumors may also present with a "paraneoplastic syndrome" owing to the effects of substances the tumor secretes. Although the fraction of patients with cancer as the basis for a presenting sign or symptom may be low, the implications of missing an early-stage tumor call for considering cancer as a basis for persistent signs or symptoms.

Evidence of a tumor's existence can come from careful physical examination, e.g., enlarged lymph nodes in lymphomas or palpable mass in a breast or soft tissue site. A mass may also be detected or confirmed by an imaging modality (e.g., x-ray, magnetic resonance imaging [MRI], or ultrasound). Endoscopy may directly visualize a tumor.

ESTABLISHING A CANCER DIAGNOSIS

Once a potential tumor is defined, establishing the diagnosis is the next step. This requires a biopsy procedure in most circumstances and pathologic confirmation that cancer is present; very rarely, where biopsy would be definitely injurious and imaging modalities are unequivocal, such as with a likely brainstem glioma, treatment might be reasonably considered based on clinical and imaging evidence without biopsy. In addition to light microscopy, biopsied tissue also allows definition of genetic abnormalities and protein expression patterns to recommend best treatment **(Table 78-2)**.

TABLE 78-2 Diagnostic Biopsy: Standard-of-Care Molecular and Special Studies to Be Considered

- All solid tumors
  - Tumor mutational burden
  - Microsatellite instability DNA repair pathway intactness
  - Homologous recombination DNA repair pathway intactness
- Breast cancer: primary and suspected metastatic
  - Breast cancer susceptibility 1 and 2 (*BRCA1/2*) gene mutations
  - Hormone receptor expression: estrogen, progesterone
  - HER2/neu oncoprotein
  - *PI3KA* mutation status
- Lung cancer: primary and suspected metastatic
  - If nonsquamous non-small-cell:
    - Epidermal growth factor receptor (*EGFR*) mutation
    - *ALK, ROS1, NRTK, NRG1* gene fusion
    - *BRAF* V600E mutation
    - Programmed cell death ligand 1 (PD-L1) expression
- Colon cancer: suspected metastatic
  - *KRAS* mutation
  - *BRAF* V600E mutation
- Gastrointestinal stromal tumor
  - *KIT* mutation
- Melanoma
  - *BRAF* mutation
  - c-kit expression and *KIT* mutation if present
- Pancreatic cancer
  - *BRCA1/2* mutation
- Prostate cancer
  - *BRCA1/2* mutation
- Thyroid cancer
  - *RET* gene alterations (mutations, translocations, amplification)
- Gliomas
  - 1p/19q co-deletion
  - Alkylguanine alkyltransferase promoter methylation
  - Isocitrate dehydrogenase 1 and 2 mutation
- Leukemia (peripheral blood mononuclear cells and/or bone marrow)
  - Cytogenetics
  - Flow cytometry
  - Treatment-defining chromosomal translocations/mutations
    - Bcr-Abl fusion protein
    - t(15;17)
    - Inversion 16
    - t(8;21)
  - FMS-associated tyrosine kinase (*FLT3*) mutation
  - Nucleophosmin gene mutational status
  - Isocitrate dehydrogenase 1 and 2 mutation
- Lymphoma
  - Immunohistochemistry for CD20, CD30, and B- and T-cell markers
  - Treatment-defining chromosomal translocations:
    - t(14;18)
    - t(8;14)
  - Translocations involving *ALK* gene

The extent of specialized testing needs to be tailored to an individual patient's case. Global DNA sequencing of genes expressed in tumors has not been shown to convey conclusive advantage in terms of survival. But the aggregate "mutational burden" present in tumors and the intactness of DNA repair genes (e.g., breast cancer susceptibility 1 and 2 [*BRCA1/2*], microsatellite instability, homologous recombination pathway–associated genes) may suggest valuable treatment courses in tumors without curative potential.

Optimally, an *excisional biopsy* occurs, in which the entire tumor mass is removed with a margin of normal tissue surrounding it. If an excisional biopsy cannot be performed, *incisional biopsy* is the procedure of second choice: a wedge of tissue is removed, trying to include the majority of the cross-sectional diameter to minimize sampling error. Biopsy techniques that involve cutting into tumor risk facilitating the spread of the tumor. Consideration with a surgeon of whether the biopsy approach is a potential prelude to a curative surgery may inform the approach taken. *Core-needle biopsy* usually obtains considerably less tissue but can provide information to plan a treatment. *Fine-needle aspiration* generally yields a suspension of cells from a mass. If positive for cancer, it may allow inception of systemic treatment, or it can provide a basis for planning a more extensive surgical procedure. A "negative" fine-needle aspiration cannot be taken as definitive evidence that a tumor is absent. In some instances, features of diagnostic imaging are sufficient to make a reliable diagnosis without obtaining tissue, usually with presence of a tumor-associated circulating diagnostic marker, e.g., α fetoprotein in hepatocellular carcinoma or prostate-specific antigen (PSA) in prostate cancer with apparent typical metastatic disease.

■ CANCER STAGING

Defining the extent of disease determines whether localized treatments, "combined-modality" approaches, or systemic treatments should initially be considered. Radiographic and other imaging tests can be helpful in defining the *clinical stage*; *pathologic staging* documents the histologic presence of tumor in tissue biopsies. Lymph node sampling in breast cancer, melanoma, lung, head and neck, colon, and other intra-abdominal cancers may provide crucial information.

Staging systems define a "T" component related to the size of the tumor or its invasion into local structures, an "N" component related to the number and nature of lymph node groups with tumor involvement, and an "M" component, based on the presence of local or distant metastatic sites. The various TNM components are then aggregated to stages, usually stage I to III or IV, depending on the anatomic site. The numerical stages reflect similar long-term survival outcomes of the respective TNM grouping stage after treatment tailored to the stage. In general, stage I tumors are T1 (reflecting small size), N0 or N1 (reflecting no or minimal node spread), and M0 (no metastases). Such early-stage tumors are usually amenable to curative approaches with local treatments. On the other hand, stage IV tumors have metastasized to distant sites or locally invaded viscera in a nonresectable way. They are treated with palliative intent, except for those diseases with exceptional sensitivity to systemic treatments such as chemotherapy or immunotherapy. Also, the TNM staging system is not useful in diseases such as leukemia, where bone marrow infiltration is never localized, or central nervous system (CNS) tumors, where tumor histology and the extent of feasible resection are more important in driving prognosis.

CANCER TREATMENT

The goal of cancer treatment is first to eradicate the cancer. If this is not possible, the goal shifts to palliation: amelioration of symptoms and preservation of quality of life while striving to extend life. When cure of cancer is possible, cancer treatments may be undertaken despite the certainty of severe toxicities, but these may produce toxicity with no benefit. Conversely, when the clinical goal is palliation, careful attention to minimizing the toxicity of treatments becomes a significant goal.

The two main types of cancer treatment are *local* and *systemic*. Local treatments include surgery, radiation therapy (including photodynamic therapy), and ablative approaches, including radiofrequency and thermal or cryosurgical approaches. Systemic treatments include chemotherapy (including hormonal therapy and molecular targeted therapy) and biologic therapy (including immunotherapy). The modalities are often used in combination. *Oncology*, the study of tumors including treatment approaches, is a multidisciplinary effort with surgical, radiation, and internal medicine–related areas of oncologic expertise.

Normal organs and cancers share the property of having a population of cells actively progressing through the cell cycle, with their division providing a basis for organ or tumor growth, and a population of cells not in cycle; these include *stem cells*, whose properties are being elucidated. Cancer stem cells serve as a basis for tumor initiating or repopulating cells. Tumors follow a Gompertzian growth curve **(Fig. 78-1)**, with the growth fraction of a neoplasm high with small tumor burdens but declining until, at the usual time of diagnosis (tumor burden of 1–5 × 10^9) the growth fraction is only 1–4% for many solid tumors. Thus the most rapid growth rate occurs before the tumor is detectable. An alternative explanation for such growth properties may also emerge from the ability of tumors at metastatic sites to recruit circulating tumor cells from the primary tumor or other metastases. Key features of tumor growth are the ability to stimulate new supporting stroma through angiogenesis and ingrowth of fibroblasts and immune cells **(Chap. 77)**.

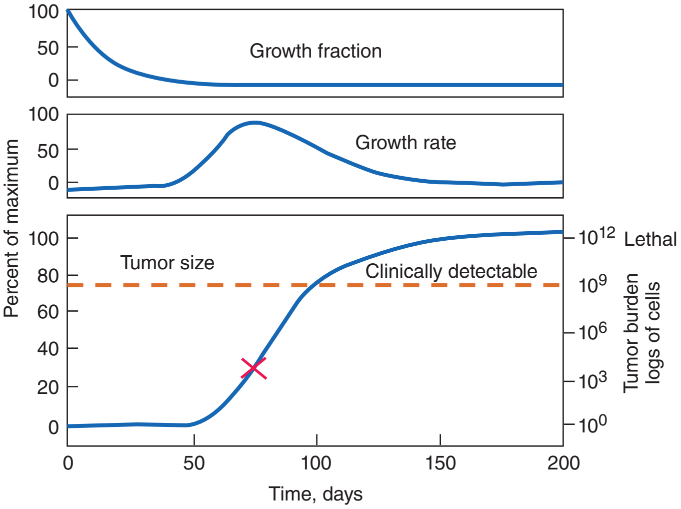

FIGURE 78-1 Gompertzian tumor growth. The growth fraction of a tumor declines exponentially over time (*top*), peaking before it is clinically detectable (*middle*). Tumor size increases slowly, goes through an exponential phase, and slows again as the tumor has limitation of nutrients or host regulatory influences occur. The maximum growth rate occurs at 1/e, the point at which the tumor is about 37% of its maximum size (*marked with an X*). Tumor becomes detectable at a burden of about 10^9 (1 cm^3) cells and kills the patient at a tumor cell burden of about 10^{12} (1 kg).

LOCALIZED CANCER TREATMENTS

■ SURGERY

Surgery is the most effective means of treating cancer. At least 40% of cancer patients are cured by surgery. Unfortunately, many patients with solid tumors have disease not practically removable. "Palliative" surgeries, however, may afford local control of tumor, preserve organ function, and achieve debulking that permits more effective subsequent therapy, while allowing more detailed staging. Cancer surgery aiming for cure is usually planned to excise the tumor completely with an adequate margin of normal tissue (the margin varies with the tumor and the anatomy). Such a resection is defined as an R0 resection. R1 and R2 resections, in contrast, are imprecisely defined pathologically as having microscopic or macroscopic, respectively, tumor at resection margins. Such outcomes may be the basis for reoperation to obtain optimal margins if feasible and of likely clinical utility. Extending the procedure to resect draining lymph nodes obtains prognostic information and may, in some anatomic locations, improve survival.

Laparoscopic approaches are being used for primary abdominal and pelvic tumors, although with certain tumors (e.g., uterine and cervix), controversy exists as to the desirability of laparoscopic tumor removal. Lymph node spread may be assessed using the sentinel node approach, in which the first draining lymph node is defined by injecting a dye or radioisotope into the tumor site at operation and then resecting the first node to turn blue or collect isotope. Sentinel node assessment appears to provide information without the risks (lymphedema, lymphangiosarcoma) associated with resection of all regional nodes. Advances in adjuvant chemotherapy (given systemically after removal of all local disease without evidence of metastatic disease) and radiation therapy following surgery have permitted a decrease in the extent of primary surgery. Thus, "lumpectomy" with radiation therapy is as effective as modified radical mastectomy for breast cancer, and limb-sparing surgery followed or preceded by adjuvant radiation therapy and chemotherapy has replaced amputation for most childhood rhabdomyosarcomas and osteosarcomas. More limited surgery spares organ function, as in larynx and bladder cancer. In some settings (e.g., bulky testicular cancer or stage III breast cancer), surgery is not the first treatment modality used. After diagnostic biopsy, chemotherapy and/or radiation therapy is delivered, followed by a surgical procedure to remove residual masses; this is called *neoadjuvant therapy*. Coordination among the surgical oncologist, radiation oncologist, and medical oncologist is crucial.

Surgery may be curative in a subset of patients with metastatic disease. Patients with limited lung metastases from osteosarcoma may be cured by resection of the lung lesions. In patients with colon cancer who have fewer than five liver metastases restricted to one lobe and no extrahepatic metastases, hepatic lobectomy may produce long-term disease-free survival in 25% of selected patients. In the setting of hormonally responsive tumors, oophorectomy may eliminate estrogen production, and orchiectomy may reduce androgen production, hormones that drive many breast and all prostate cancers, respectively. In selecting a surgeon or center for primary cancer treatment, consideration must be given to the volume of cancer surgeries undertaken by the site. Studies in a variety of cancers have shown that increased annual procedure volume appears to correlate with outcome. Surgery is used in a number of ways for palliative or supportive care of the cancer patient. These include insertion and care of central venous catheters, control of pleural and pericardial effusions and ascites, caval interruption for recurrent pulmonary emboli, stabilization of cancer-weakened weight-bearing bones, and control of hemorrhage, among others. Surgical bypass of gastrointestinal, urinary tract, or biliary tree obstruction or spinal cord decompression can alleviate symptoms and may prolong survival. Splenectomy may relieve symptoms and reverse hypersplenism. Intrathecal or intrahepatic therapy relies on surgical placement of appropriate infusion portals. Surgery may correct other treatment-related toxicities such as adhesions or strictures. Plastic and reconstructive surgery can correct the effects of disfiguring primary treatment. Surgery is also a tool valuable in the prevention of cancers in high-risk populations. Prophylactic mastectomy, colectomy, oophorectomy, and thyroidectomy are mainstays of cancer prevention in inherited cancer susceptibility syndromes.

RADIATION

Radiation Biology and Medicine Therapeutic radiation is generally ionizing, causing breaks in DNA and generation of free radicals from cell water that damage cancer cell membranes, proteins, and organelles.

$$\text{Ionizing radiation} + H_2O \rightarrow H_2O^+ + e^-$$
$$H_2O^+ + H_2O \rightarrow H_3O^+ + OH^\bullet$$
$$OH^\bullet \rightarrow \text{cell damage}$$

Radiation damage is augmented by oxygen; hypoxic cells are more resistant. X-ray and gamma-ray photons are the forms of ionizing radiation most commonly used to treat cancer. Particulate ionizing radiation using protons has also become available.

Radiation dose is quantitated based on the amount of energy absorbed by the tumor, not on radiation generated by the machine. The International System (SI) unit for radiation dose is the Gray (Gy): 1 Gy refers to 1 J/kg of tissue; 1 Gy equals 100 centigrays (cGy) of absorbed dose. A historically used unit appearing in the oncology literature, the *rad* (*r*adiation *a*bsorbed *d*ose), is defined as 100 ergs of energy absorbed per gram of tissue and is equivalent to 1 cGy. Radiation dose is measured by placing detectors at the body surface or in radiated phantoms that resemble human form and substance. The features that make a particular cell more or less sensitive to radiation involve DNA repair proteins that, in their physiologic role, protect against environmentally related DNA damage.

Localized Radiation Therapy Radiation effect is influenced by three determinants: total absorbed dose, number of fractions, and time of treatment. A typical course of radiation therapy should be described as 4500 cGy delivered to a particular target (e.g., mediastinum) over 5 weeks in 180-cGy fractions. Most curative radiation treatment programs are delivered once a day, 5 days a week, in 150- to 200-cGy fractions. Nondividing cells are more resistant than dividing cells; delivering radiation in repeated fractions is done to expose a larger number of tumor cells that have entered the division cycle. The energy of the radiation determines its ability to penetrate tissue. Low-energy x-rays (150–400 kV) scatter when they strike the body, resulting in more damage to adjacent normal tissues and less radiation delivered to the tumor. Megavoltage radiation (>1 MeV) has very low lateral scatter; this produces a skin-sparing effect, more homogeneous distribution of the radiation energy, and greater deposit of the energy in the tumor, or *target volume*. The *transit volume* includes the tissues through which the beam passes to the target volume. Computational approaches and delivery of many beams to converge on a target volume are the basis for “gamma knife” and related approaches to deliver high doses to tumor, sparing normal tissue.

Therapeutic radiation is delivered in three ways: (1) *teletherapy*, with focused beams of radiation generated at a distance and aimed at the tumor within the patient; (2) *brachytherapy*, with encapsulated sources of radiation implanted directly into or adjacent to tumor tissues; and (3) *systemic therapy*, with radionuclides administered, for example, intravenously but perhaps targeted by some means to a tumor site. For example, systemically administered isotopes of iodide have an important role in the treatment of thyroid neoplasms, owing to the selective upregulation of the iodide transporter in the tumor cell compartment. Likewise, isotopes of samarium and radium have been found useful in the palliation of bony metastases of prostate cancer owing to their selective deposition at the tumor-bone matrix interface.

Teletherapy with x-ray or gamma-ray photons is the most commonly used form of radiation therapy and also delivers particulate forms of radiation such as proton beams. The difference between photons and protons relates to volume with greatest delivery of energy: protons have a narrow range of energy deposition. Electron beams are a particulate form of radiation that, in contrast to photons and protons, have a very low tissue penetrance and are used to treat cutaneous tumors. Certain drugs used in cancer treatment may also act as radiation sensitizers. For example, compounds that incorporate into DNA (e.g., halogenated pyrimidines, cisplatin) augment radiation effects at local sites and are important adjuncts to radiation of certain tumors (e.g., squamous head and neck, uterine cervix, and rectal cancers).

Toxicity of Radiation Therapy Although radiation therapy is most often administered to a local region, systemic effects, including fatigue, anorexia, nausea, and vomiting, may develop that are related in part to the volume of tissue irradiated, dose fractionation, radiation fields, and individual susceptibility. Radiation-injured tissues release cytokines that act systemically to produce these effects. Bone is among the most radio-resistant organs, with radiation effects being manifested mainly in children through premature fusion of the epiphyseal growth plate. By contrast, the male testis, female ovary, and bone marrow are the most sensitive organs. Any bone marrow in a radiation field will be eradicated by therapeutic irradiation. Organs with less need for cell renewal, such as heart, skeletal muscle, and nerves, are more resistant to immediate radiation effects. In radiation-resistant organs, the vascular endothelium is the most sensitive component. Acute toxicities include mucositis, skin erythema (ulceration in severe cases), and bone marrow toxicity. Often these can be alleviated by periodic interruption of treatment.

Chronic toxicities are more serious. Radiation of the head and neck region produces thyroid failure; cataracts and retinal damage can lead to blindness; salivary glands stop making saliva, which leads to dental caries and poor dentition. Mediastinal irradiation increases myocardial vascular disease. Other late vascular effects include chronic constrictive pericarditis, lung fibrosis, viscus stricture, spinal cord transection, and radiation cystitis or enteritis.

A serious late toxicity is the development of hematologic tumors or second solid tumors in or adjacent to the radiation fields. Such tumors can develop in any organ or tissue and occur at a rate of ~1% per year beginning in the second decade after treatment.

OTHER LOCALIZED CANCER TREATMENTS

Endoscopy allows placement of stents to unblock viscera, palliating, for example, gastrointestinal or biliary obstructions. Radiofrequency ablation (RFA) refers to microwave nonionizing radiation to induce thermal injury within a volume of tissue. RFA can be useful in the control of metastatic lesions, particularly in liver, that may threaten biliary drainage. Cryosurgery uses extreme cold to sterilize lesions in certain sites, such as prostate and kidney, at a very early stage, eliminating the need for modalities with more side effects such as surgery.

Some chemicals (porphyrins, phthalocyanines) are preferentially taken up by cancer cells. When intense light, delivered by a laser, is shone on cells containing these compounds, free radicals are generated and the cells die. Such phototherapy is used to treat skin cancer; ovarian cancer; and cancers of the lung, colon, rectum, and esophagus. Palliation of recurrent locally advanced disease can sometimes be dramatic and last many months.

Infusion of chemotherapeutic or biologic agents or radiation-bearing delivery devices such as isotope-coated glass spheres into local sites through catheters have been used to treat disease limited to that site; in selected cases, prolonged control of truly localized disease has been possible.

SYSTEMIC CANCER TREATMENTS

The concept that systemically administered chemicals might have a useful effect on cancers was historically derived from three sets of observations. Paul Ehrlich in the nineteenth century observed that different dyes reacted with different cell and tissue components. He hypothesized the existence of “magic bullets” that might bind to tumors, owing to the affinity of the agent for the tumor. Observation of the toxic effects of certain mustard gas derivatives on the bone marrow during World War I suggested that smaller doses of these agents might be used to treat tumors of marrow-derived cells. Finally, the fact that tumors from hormone-responsive tissues, e.g., breast tumors, could shrink after oophorectomy led to the idea that endogenous or exogenous substances might modulate tumor growth by altering its regulatory biology. Chemicals achieving each of these goals are currently used as cancer chemotherapy agents.

Anecdotal reports of tumor regression following intratumoral injection of bacterial extracts raised the possibility of immune system–mediated tumor regression. Serotherapy of infectious disease in the preantibiotic era encouraged analogous efforts to develop vaccine- and antibody-based treatments for cancer. Administration of autologous immune cells obtained by pheresis procedures from a patient or purified from a patient’s removed tumor, activated by cytokines ex vivo, achieved durable disease control in a small fraction of patients with certain tumors, particularly renal cell cancers or melanoma. These observations provided the rationale for more modern efforts to treat tumors using cell-mediated immunity.

Systemic cancer treatments are of three broad types. *Cytotoxic chemotherapy agents* are “small molecules” (generally with molecular mass <1500 Da) that cause major regression of experimental tumors growing in animals. These agents mainly target DNA structure or segregation of chromosomes in mitosis. *Cancer molecular target therapies* refer to small molecules designed and developed to interact with a defined macromolecule important in maintaining the malignant state. As described in Chap. 77, successful tumors have activated biochemical pathways that lead to uncontrolled proliferation through the action of hormone receptor proteins, oncogene products, loss of cell cycle inhibitors, or loss of cell death regulation, and have acquired the capacity to replicate chromosomes indefinitely, invade, metastasize, and evade the immune system. *Cancer biologic therapies* are most frequently macromolecules, cells, or cell extracts that have a particular target (e.g., anti–growth factor receptor, cytokine, or immunomodulatory antibodies) or may have the capacity to induce a host immune response to kill tumor cells. Most recent additions to cancer biologic therapies include genetically modified cells that directly attack tumor cells and tumor-infecting viruses that can kill tumor cells but also elicit host antitumor immune responses.

SYSTEMIC CANCER THERAPY OVERVIEW

General Principles The *therapeutic index* of any drug is the degree of separation between toxic and therapeutic doses. Really useful drugs have large therapeutic indices, and this usually occurs when the drug target is expressed in the disease-causing compartment as opposed to the normal compartment. Cytotoxic chemotherapeutic agents have the unfortunate property that their main targets, DNA and microtubules, are present in both normal and tumor tissues. Therefore, they have relatively narrow therapeutic indices. Targeted agents can also cause effects on their target in normal tissues, or “off-target” effects on unrelated targets in organs experiencing damage. Biologic therapies may elicit misdirected immune responses on normal organ function. A key activity in oncology drug development is striving to administer a dose that can convey benefit with a minimal or tolerable side effect profile.

Figure 78-2 illustrates steps in cancer drug development. Following demonstration of antitumor activity in animal models, potentially useful anticancer agents are further evaluated to define an optimal schedule of administration and suitable drug formulation. Safety testing in two animal species on an analogous schedule of administration defines the starting dose for a phase 1 trial in humans, usually but not always in patients with cancer who have exhausted “standard” (already approved) treatments. The initial dose is usually one-sixth to one-tenth of the dose just causing easily reversible toxicity in the more sensitive animal species. If the agent is not intrinsically toxic, doses of drug achieving fractions of the useful plasma concentrations from model systems are studied. Escalating doses of drug are then given during the human phase 1 trial until reversible toxicity is observed or the desired drug concentration is achieved. Dose-limiting toxicity (DLT) defines a dose that conveys greater toxicity than would be acceptable in routine practice, allowing definition of a lower maximum-tolerated dose (MTD). The occurrence of toxicity is, if possible, correlated with plasma drug concentrations. The MTD or a dose just lower than the MTD is usually the dose suitable for phase 2 trials, where a fixed dose and schedule is administered to a relatively homogeneous set of patients with a particular tumor type. If no toxicity has emerged in phase 1 trials, administration of the optimal biologic dose to achieve effective drug concentrations is undertaken. A partial response historically was defined as a decrease of at least 50% in a tumor’s bidimensional area obtained by imaging; more recent response criteria (e.g., Response Evaluation Criteria in Solid Tumors [RECIST]) may use a 30% decrease in aggregate unidimensional areas of target lesions. Response criteria for immunologically directed agents may allow a substantial transient increase in tumor volume as long as a patient’s clinical status is stable, as these agents may evoke inflammatory responses in tumors with subsequent shrinkage or stabilization of lesions. A complete response connotes disappearance of all tumor; progression of disease signifies an increase in size of existing lesions by >25% from baseline or best response or development of new lesions; stable disease fits into none of the above categories.

In a phase 3 trial, evidence of improved overall survival or improvement in the time to progression of disease on the part of the new drug is sought in comparison to an appropriate control population. Data from the entire process are the basis for application to a regulatory agency to approve the new agent for commercial marketing.

Cancer drug clinical trials conventionally use a toxicity grading scale where grade 1 toxicities do not require treatment, grade 2 toxicities may require symptomatic treatment but are not life-threatening, grade 3 toxicities are potentially life-threatening if untreated, grade 4 toxicities are actually life-threatening, and grade 5 toxicities result in the patient’s death. Active efforts to quantitate effects of anticancer agents on quality of life also frequently occur in early development of oncology drugs.

Development of targeted agents may proceed differently. While phase 1–3 trials are still conducted, focus on a particular tumor type even in phase 1 may be enabled by molecular analysis to define target expression in a patient’s tumor necessary for or relevant to the drug’s action. Ideally, pharmacodynamic studies would also assess whether the drug’s target has been affected. The failure of a targeted therapy can

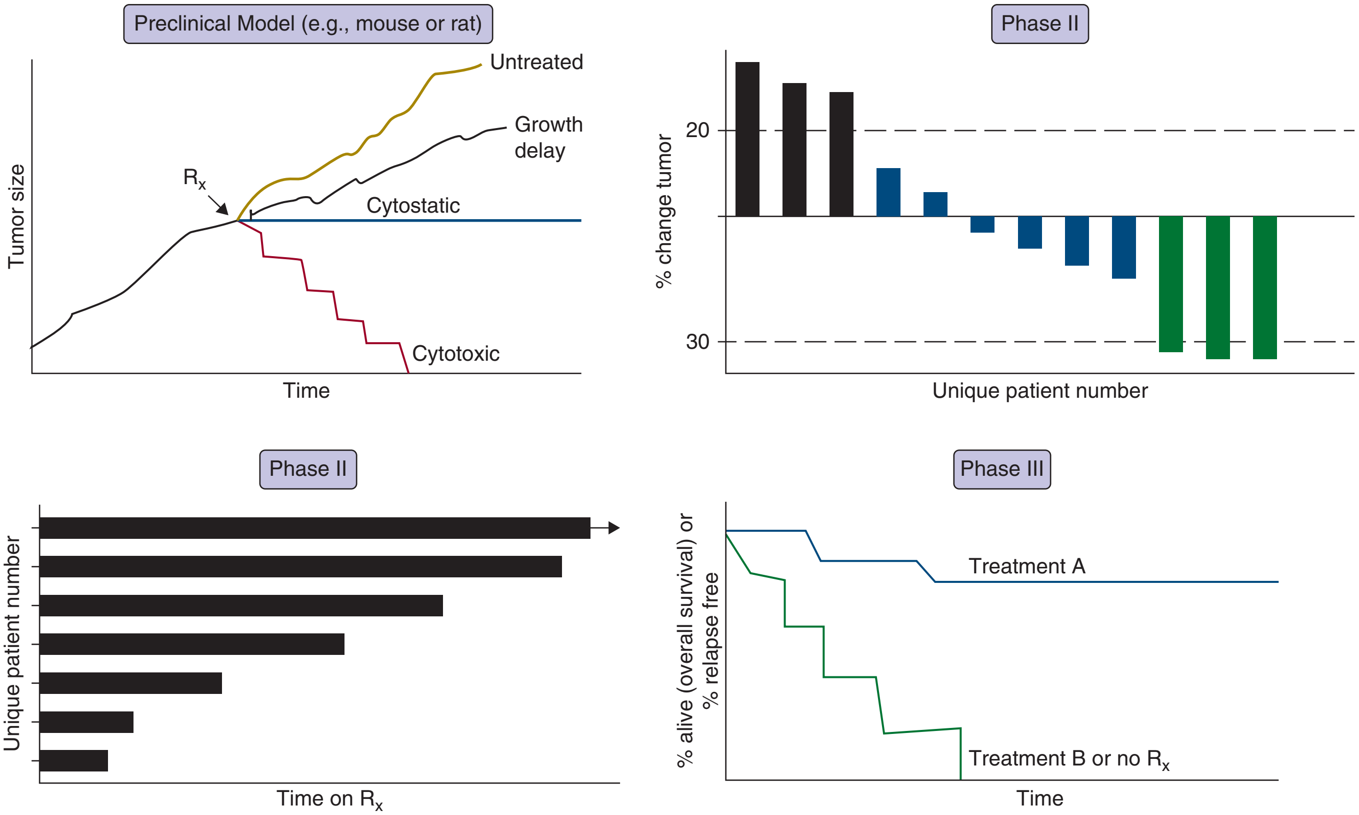

FIGURE 78-2 Steps in cancer drug discovery and development. Preclinical activity (*top left*) in animal models of cancers may be used as evidence to support the entry of the drug candidate into phase 1 trials in humans to define a correct dose and observe any clinical antitumor effect. The drug may then be advanced to phase 2 trials directed against specific cancer types, with rigorous quantitation of antitumor effects. *Waterfall plots* are a standard representation of how patients' tumor sizes change in relation to treatment, with predefined cutoffs defining progression of disease (20% increase in size) or partial response (30% decrease in size) serving as benchmarks of potential valuable effect (*top right*). *Swimmer plots* (*bottom left*) allow the delineation of patients with especially long (or short) times on treatment even without response, another basis in the former case for potential perceived clinical benefit of the treatment. *Kaplan-Meier plots* (*bottom right*) of survival indices in phase 3 comparative trials may allow definition of superiority, inferiority, or no difference of treatment effect compared to standard or no treatment.

be either because the drug missed the target or it hit the target but the target was not central to the tumor's growth and survival. Within the past decade, agents have been approved for clinical use not in relation to an originating organ site of disease but across all organ types possessing certain molecular or biologic features.

Useful cancer drug treatment strategies using conventional chemotherapy agents, targeted agents, hormonal treatments, or biologicals all have one of two valuable outcomes. They can induce cancer cell death, resulting in tumor shrinkage with corresponding clinical benefit evidenced by improvement in patient survival, or increase in time until the disease progresses. Another potential outcome is induction of cancer cell *differentiation* or *dormancy* with loss of tumor cell replicative potential and reacquisition of phenotypic properties resembling normal cells. Interaction of a chemotherapeutic drug with its target induces a "cascade" of further signaling steps. These signals ultimately lead to cell death by triggering proteases, nucleases, and endogenous regulators of the cell death pathway **(Fig. 78-3)** or differentiation by alteration of cancer genome function.

Targeted agents differ from chemotherapy agents in that they do not indiscriminately cause macromolecular lesions but regulate the action of specific driver molecules regulating processes considered "hallmarks" of cancer, including unregulated proliferation, physiologic cell death, gene expression, angiogenesis, and escape from immune mechanisms retarding tumor development.

Strategies in Systemic Cancer Management The past 35 years have witnessed a marked evolution in the systemic treatment of cancer not amenable to cure by locally applied treatments. Nonspecific cytotoxic agents of limited efficacy for most patients but highly active and curative in a minority disease types have been joined by targeted and biologic therapies. **Table 78-3, A** lists those tumors considered curable by conventionally available chemotherapeutic agents even when disseminated or metastatic. If a tumor is truly localized to a single site, consideration of surgery or primary radiation therapy should be given as well. Chemotherapy may be used as part of multimodality approaches to offer primary treatment to a clinically localized tumor **(Table 78-3, B)**, interdigitated with radiation and surgery. Chemotherapy can be administered as an *adjuvant*, i.e., in addition to surgery or radiation **(Table 78-3, C)**, even after all clinically apparent disease has been removed. This use of chemotherapy has curative potential in, for example, lung, breast, and colorectal neoplasms, as it eliminates micrometastatic clinically unapparent tumor. *Neoadjuvant* chemotherapy refers to administration of chemotherapy before any surgery or radiation to a local tumor in an effort to enhance the effect of subsequent local treatment.

Chemotherapy is routinely used in doses that produce reversible acute side effects, primarily consisting of transient myelosuppression with or without gastrointestinal toxicity (usually nausea). "High-dose" chemotherapy regimens can produce markedly increased therapeutic effect, although at the cost of potentially life-threatening complications that require intensive support, usually in the form of hematopoietic stem cell support from the patient (*autologous*) or from donors matched for histocompatibility loci (*allogeneic*), or pharmacologic "rescue" strategies to block the effect of the high-dose chemotherapy on normal tissues. High-dose regimens have curative potential in defined clinical settings **(Table 78-3, D)**.

If cure is not possible, chemotherapy may be undertaken with the goal of palliating the tumor's effect on the host **(Table 78-3, E)**. In this usage, value is perceived by the demonstration of improved symptom relief, progression-free survival, or overall survival. The best data to support these different therapeutic approaches result from clinical research protocols. In the case of a new drug, these trials might be the

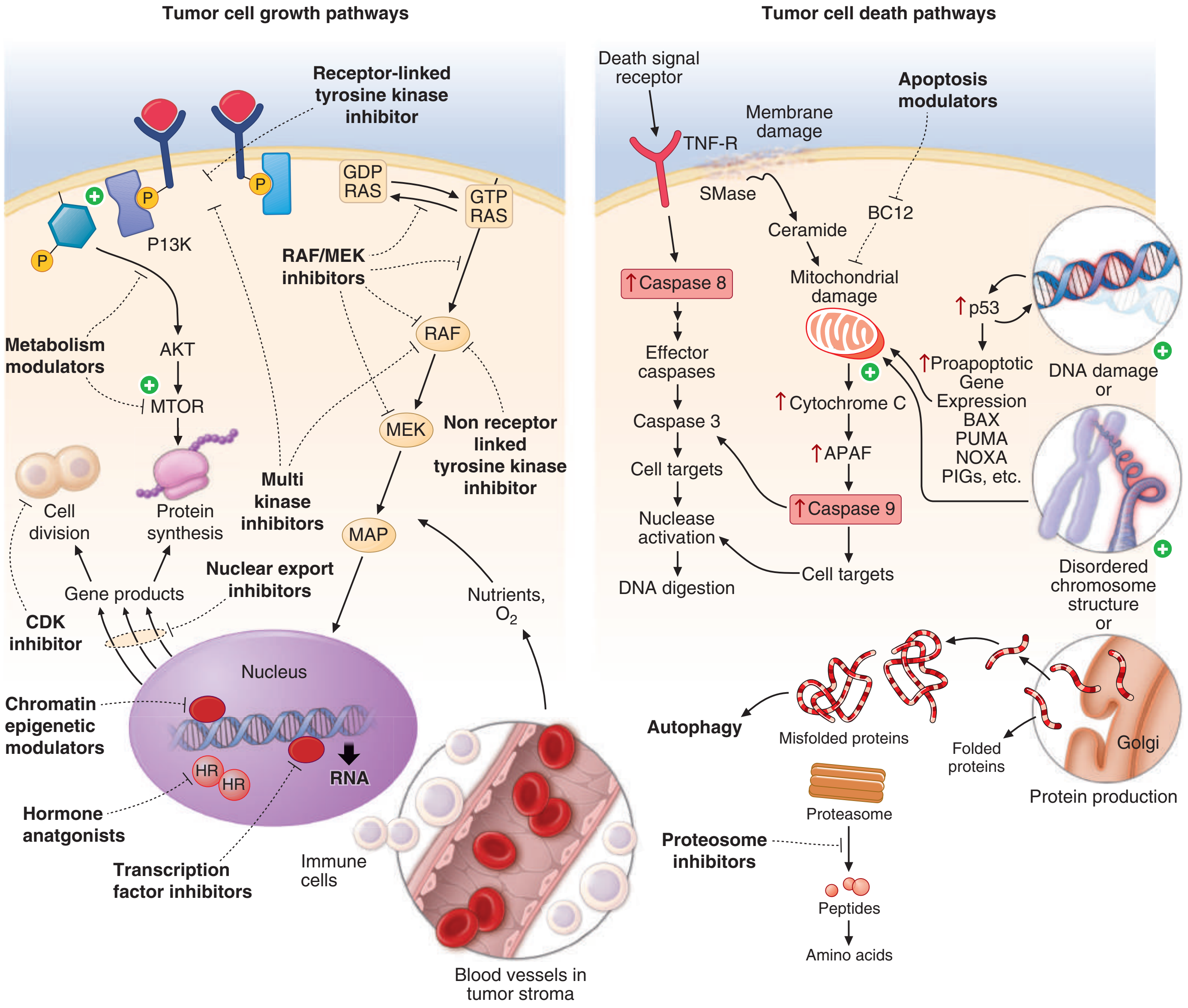
Tumor cell growth pathways
Tumor cell death pathways
Receptor-linked tyrosine kinase inhibitor
P
P
P
P13K
GDP RAS
GTP RAS
RAF/MEK inhibitors
RAF
MEK
MAP
Metabolism modulators
AKT
MTOR
Cell division
Protein synthesis
Multi kinase inhibitors
Non receptor linked tyrosine kinase inhibitor
Nuclear export inhibitors
Gene products
CDK inhibitor
Nucleus
Chromatin epigenetic modulators
HR
HR
RNA
Hormone anatgonists
Transcription factor inhibitors
Immune cells
Nutrients, O$_2$
Blood vessels in tumor stroma
Death signal receptor
TNF-R
Membrane damage
Apoptosis modulators
SMase
Ceramide
BC12
↑Caspase 8
Mitochondrial damage
↑p53
Effector caspases
↑Proapoptotic Gene Expression BAX PUMA NOXA PIGs, etc.
DNA damage or
↑Cytochrome C
Caspase 3
↑APAF
Cell targets
Nuclease activation
↑Caspase 9
Disordered chromosome structure or
DNA digestion
Cell targets
Autophagy
Misfolded proteins
Folded proteins
Golgi
Protein production
Proteasome
Proteosome inhibitors
Peptides
Amino acids

FIGURE 78-3 Tumor growth and death pathways affected by targeted and cytotoxic agents. After a growth factor binds to its receptor (*left side of figure*), in the most commonly activated cell proliferation pathway, increased tyrosine kinase activity occurs, either by autophosphorylation of receptor-linked kinases or through recruitment of non-receptor-linked tyrosine kinases, which may also be active constitutively, without requiring a growth factor. This leads to docking of "adaptor" proteins to the phosphorylated tyrosines. One important pathway activated occurs after exchange of GDP for GTP in the RAS family of proto-oncogene products. GTP-RAS activates the RAF kinase, leading to a phosphorylation cascade of MEK and MAP kinases. Mutated KRAS inhibitors have been defined, as well as RAF isoform and MEK inhibitors. The RAF/MEK/MAP kinase pathway ultimately alters gene function to activate cell cycle progression through cyclin-dependent kinases (CDKs). Another route to gene activation utilizes hormone receptors (HRs) interacting with tissue-specific growth regulators such as steroid hormones to alter gene function leading to cell cycle activation. Transcription factors acting at the level of RNA polymerase function at ensembles of gene promoters have proved difficult to target directly, but direct and indirect modulators of transcription factor activity have been defined. Receptor and non-receptor linked tyrosine phosphorylation can lead to activation of phosphatidylinositol-3-kinase (PI3K) to produce the phosphorylated lipid phosphatidylinositol-3-phosphate, which activates the AKT kinase to act downstream on the mammalian target of rapamycin kinase (mTOR), directly increasing translation of key mRNAs for gene products regulating cell growth and stimulating cell metabolism by, for example, increasing glucose transporter function. Cytotoxic agents cause cell death (*right side of figure*) through apoptosis and/or induction of autophagy. Apoptosis is also stimulated by interruption of growth factor (GF) cytokine death signals (e.g., tumor necrosis factor receptor [TNF-R]), which activate "upstream" cysteine aspartyl proteases (caspases) to directly digest cytoplasmic and nuclear proteins, resulting in activation of "downstream" caspases; these activate nucleases to cause DNA fragmentation, a hallmark of apoptosis. Chemotherapy agents that create lesions in DNA or alter mitotic spindle function activate gene function to alter mitochondrial integrity. The antiapoptotic protein BCL2 attenuates mitochondrial toxicity, whereas proapoptotic gene products such as BAX, PUMA, etc., antagonize the action of BCL2. Damaged mitochondria release cytochrome C and apoptosis-activating factor (APAF), which activate caspase 9 to cause DNA fragmentation. In addition, membrane damage with activation of sphingomyelinases results in the production of ceramides that can cause direct damage to mitochondria. Protein translation is followed by a folding process in the Golgi apparatus. Misfolded proteins are processed through the proteasome for protease digestion and recycling of amino acids. Disruption of this process can contribute to autophagy, where the cell starves for critical nutrients, or itself induce apoptosis through a distinct pathway activated by misfolded protein accumulation. Antiangiogenic agents and immune therapies work in the tumor stroma (*lower left*) on supporting elements including blood vessels and host inflammatory cells.

basis for U.S. Food and Drug Administration (FDA) approval for commercial use of the agent.

Patients treated with palliative intent should be aware of their diagnosis and the limitations of the proposed treatments, have access to supportive care, and have suitable "performance status," according to assessment algorithms such as the one developed by Karnofsky (see Table 73-4) or by the Eastern Cooperative Oncology Group (ECOG) (see Table 73-5). ECOG performance status (PS) 0 patients are without symptoms; PS1 patients are ambulatory but restricted in strenuous physical activity; PS2 patients are ambulatory and active 50% or more of the time but unable to work; PS3 patients are capable of limited self-care but are active <50% of the time; and PS4 patients are totally

TABLE 78-3 Clinical Impact on Cancers with Cytotoxic Chemotherapy

A. Advanced Cancers with Possible Cure

- Acute lymphoid and acute myeloid leukemia (pediatric/adult)
- Hodgkin's disease (pediatric/adult)
- Lymphomas—certain types (pediatric/adult)
- Germ cell neoplasms
  - Embryonal carcinoma
  - Teratocarcinoma
  - Seminoma or dysgerminoma
  - Choriocarcinoma
- Gestational trophoblastic neoplasia
- Pediatric neoplasms
  - Wilms' tumor
  - Embryonal rhabdomyosarcoma
  - Ewing's sarcoma
  - Peripheral neuroepithelioma
  - Neuroblastoma
- Small-cell lung carcinoma
- Ovarian carcinoma

B. Advanced Cancers Possibly Cured by Chemotherapy, Radiation, ± Surgery

- Squamous carcinoma (head and neck)
- Squamous carcinoma (anus)
- Bladder carcinoma
- Breast carcinoma
- Carcinoma of the uterine cervix
- Esophageal carcinoma
- Non-small-cell lung carcinoma (stage III)
- Small-cell lung carcinoma

C. Cancers Possibly Cured with Chemotherapy as Adjuvant to Surgery

- Breast carcinoma
- Colorectal carcinoma[a]
- Osteogenic sarcoma
- Soft tissue sarcoma

D. Cancers Possibly Cured with High-Dose Chemotherapy with Stem Cell Support

- Relapsed leukemias, lymphoid and myeloid
- Relapsed lymphomas, Hodgkin's and non-Hodgkin's
- Chronic myeloid leukemia
- Multiple myeloma

E. Cancers Responsive with Useful Palliation, But Not Cured, by Chemotherapy

- Bladder carcinoma
- Chronic myeloid leukemia
- Hairy cell leukemia
- Chronic lymphocytic leukemia
- Lymphoma—certain types
- Multiple myeloma
- Gastric carcinoma
- Cervix carcinoma
- Endometrial carcinoma
- Soft tissue sarcoma
- Head and neck cancer
- Adrenocortical carcinoma
- Islet cell neoplasms
- Breast carcinoma
- Colorectal carcinoma
- Glioma
- Lung cancer
  - Small-cell
  - Non-small-cell

F. Tumors Poorly Responsive in Advanced Stages to Chemotherapy

- Pancreatic carcinoma
- Biliary tract neoplasms
- Thyroid carcinoma
- Carcinoma of the vulva
- Prostate carcinoma
- Melanoma
- Hepatocellular carcinoma
- Salivary gland cancer

[a] Rectum also receives radiation therapy.

confined to bed or chair and incapable of self-care. Only PS0, PS1, and PS2 patients are generally considered suitable for palliative (noncurative) treatment. If there is curative potential, even poor-PS patients may be treated (especially if their symptoms are directly related to a cancer that may respond to treatment), but their prognosis is usually inferior to that of good-PS patients treated with similar regimens. Assessment of physiologic reserve through use of the geriatric assessment tool can be helpful, but no measure of comorbidities or physiologic reserve is considered standard.

The turn of the millennium marked the arrival of new strategies for cancer treatment. Prominent among these are *cancer biologic therapy*, which harnesses the use of immune system–derived reagents or strategies, and *cancer targeted therapies*, which are directed at specific molecular targets differentially expressed in malignant as opposed to normal tissues.

■ CANCER BIOLOGIC THERAPY

Figure 78-4 presents the landscape of cancer biologic therapy agents and actions. The goal of biologic therapy is to manipulate the host-tumor interaction in favor of the host, potentially at an optimum biologic dose that might be different than MTD. As a class, biologic therapies may be distinguished from cytotoxic and molecularly targeted agents in that biologic therapies require activity (e.g., antigen expression or internalization) on the part of the tumor cell or on the part of the host (e.g., T-cell engagement or cytokine elaboration) to allow therapeutic effect.

Antibody-Mediated Therapeutic Approaches Figure 78-4 illustrates current antibody-based strategies in cancer treatment. The ability to grow very large quantities of high-affinity monoclonal antibodies directed at specific tumor antigens produced by animals allows the grafting of animal-derived antigen-combining sequences into human immunoglobulin genes (chimerized or humanized products) or derived de novo from mice bearing human immunoglobulin gene loci. Four general strategies have emerged using antibodies. *Anti-tumor cell antibodies* target tumor cells directly to inhibit intracellular functions or attract immune or stromal cells. *Bispecific tumor engaging (BiTe) antibodies* directly bind to a tumor cell and to a host immune cell. *Immunoregulatory antibodies* target antigens expressed on host immune cells to boost the host's immune response to the tumor. Finally, *antibody conjugates* link the antibody to drugs, toxins, or radioisotopes to target these "warheads" for delivery to the tumor. *Stroma-directed antibodies* are currently available against tumor-supporting vasculature.

ANTI-TUMOR CELL ANTIBODIES (FIG. 78-4) Humanized antibodies against the CD20 molecule expressed on B-cell lymphomas (rituximab and ofatumumab) are exemplary of antibodies that affect both signaling events driving lymphomagenesis as well as activating immune responses against B-cell neoplasms. They are used as single agents and in combination with chemotherapy and radiation in the treatment of B-cell neoplasms. Obinutuzumab is an antibody with altered glycosylation that enhances its ability to activate killer cells; it is also directed against CD20 and is of value in chronic lymphocytic leukemia.

Unintended side effects of any antibody include infusion-related hypersensitivity reactions, which can be managed with glucocorticoid and/or antihistamine prophylaxis, and prolonged infusion strategies.

Anti-B-cell-directed antibodies can have the unintended effect of exacerbating immunosuppression with the emergence of increased opportunistic infections. Reactivation of latent infections may also occur; an assessment of a patient's hepatitis B and C status is conventionally done before treatment. Concomitant use of antivirals directed against hepatitis may be indicated. Patients with HIV and lymphoma need antivirals optimized to minimize interaction with anti-lymphoma treatments; consultation with infectious disease specialists is warranted. Anti-tumor cell antibodies also include approaches to activate complement and are exemplified by alemtuzumab directed against CD52; it is active in chronic lymphoid leukemia and T-cell malignancies. Tumor lysis syndrome prophylaxis may be warranted.

Epidermal growth factor receptor (EGFR)-directed antibodies (e.g., cetuximab and panitumumab) have activity in colorectal cancer refractory to chemotherapy, particularly when used to augment the activity of an additional chemotherapy program, and in the primary treatment of head and neck cancers treated with radiation therapy. Direct effects on the tumor may mediate an antiproliferative effect as well as stimulate the participation of host mechanisms involving immune cell or complement-mediated response to tumor cell–bound antibody. Anti-EGFR antibodies can cause an acneiform rash requiring topical antibiotic and glucocorticoid cream treatment; photosensitivity also occurs.

The HER2/neu receptor overexpressed on epithelial cancers, especially breast and certain gastrointestinal cancers, was initially targeted by trastuzumab, with activity in potentiating the action of chemotherapy in breast cancer as well as evidence of single-agent activity. Trastuzumab appears to interrupt intracellular signals derived from HER2/neu and to stimulate anti-tumor cell immune mechanisms. The anti-HER2 antibody pertuzumab, specifically targeting the domain of HER2/neu responsible for dimerization with other HER2 family members, is more specifically directed against HER2 signaling function and augments the action of trastuzumab. Both trastuzumab and pertuzumab can damage cardiac function, particularly in patients with prior exposure to anthracyclines, and left ventricular function should be checked before treatment and monitored during treatment.

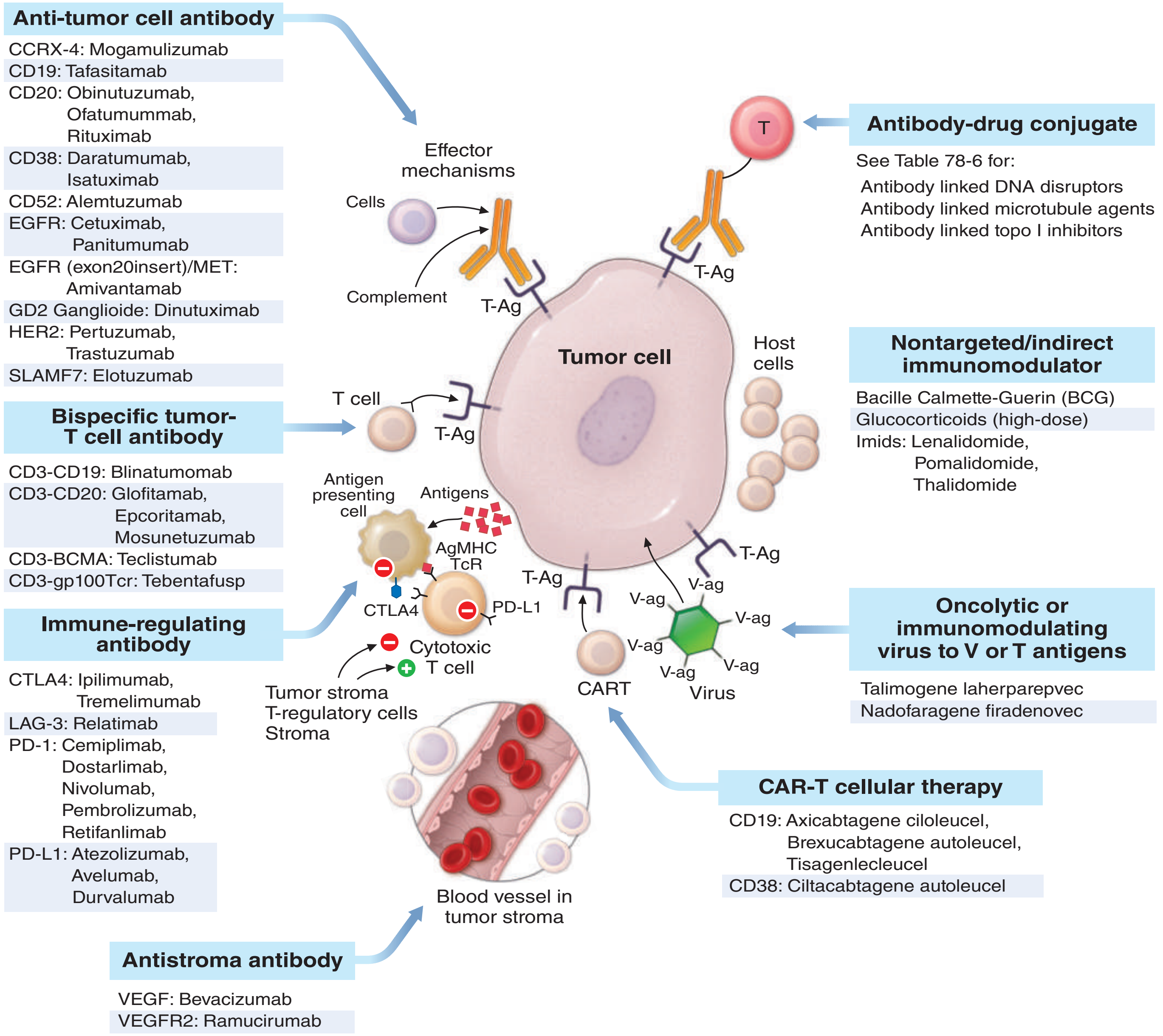
Anti-tumor cell antibody
CCRX-4: Mogamulizumab
CD19: Tafasitamab
CD20: Obinutuzumab, Ofatummmab, Rituximab
CD38: Daratumumab, Isatuximab
CD52: Alemtuzumab
EGFR: Cetuximab, Panitumumab
EGFR (exon20insert)/MET: Amivantamab
GD2 Ganglioide: Dinutuximab
HER2: Pertuzumab, Trastuzumab
SLAMF7: Elotuzumab
Bispecific tumor-T cell antibody
CD3-CD19: Blinatumomab
CD3-CD20: Glofitamab, Epcoritamab, Mosunetuzumab
CD3-BCMA: Teclistumab
CD3-gp100Tcr: Tebentafusp
Immune-regulating antibody
CTLA4: Ipilimumab, Tremelimumab
LAG-3: Relatimab
PD-1: Cemiplimab, Dostarlimab, Nivolumab, Pembrolizumab, Retifanlimab
PD-L1: Atezolizumab, Avelumab, Durvalumab
Antistroma antibody
VEGF: Bevacizumab
VEGFR2: Ramucirumab
Effector mechanisms
Cells
Complement
T-Ag
T cell
T-Ag
Tumor cell
Antigen presenting cell
Antigens
AgMHC
TcR
CTLA4
PD-L1
T-Ag
Cytotoxic T cell
Tumor stroma
T-regulatory cells
Stroma
Blood vessel in tumor stroma
T
T-Ag
Antibody-drug conjugate
See Table 78-6 for:
Antibody linked DNA disruptors
Antibody linked microtubule agents
Antibody linked topo I inhibitors
Host cells
Nontargeted/indirect immunomodulator
Bacille Calmette-Guerin (BCG)
Glucocorticoids (high-dose)
Imids: Lenalidomide, Pomalidomide, Thalidomide
T-Ag
V-ag
V-ag
V-ag
V-ag
V-ag
V-ag
CART
Virus
Oncolytic or immunomodulating virus to V or T antigens
Talimogene laherparepvec
Nadofaragene firadenovec
CAR-T cellular therapy
CD19: Axicabtagene ciloleucel, Brexucabtagene autoleucel, Tisagenlecleucel
CD38: Ciltacabtagene autoleucel

FIGURE 78-4 Immunologic treatments for cancer. Anti–tumor cell antibodies targeting the indicated antigens (T-Ag) expressed on tumor cells or antibody-derived constructs (*upper left*) can either directly interfere with tumor cell function by, e.g., inhibiting growth-promoting pathways, or recruit host immune effector cells actively (especially through bispecific tumor-engaging [BiTe] strategies; *middle left*), Fc receptors, or cytotoxic mechanisms such as complement. Endogenous T cells can be activated by *immunomodulatory cell targeting antibodies* (*lower left*). Specifically, tumor cell–derived antigens are taken up by antigen-presenting cells (APCs), also in the stroma. Antigens are processed by the APCs to peptides presented by the major histocompatibility complex (MHC) to T-cell antigen receptors (TcRs), thus providing a positive (+) activation signal for the cytotoxic tumor cells to kill tumor cells bearing that antigen. Negative (–) signals inhibiting cytotoxic T-cell action include the CTLA4 receptor (on T cells), interacting with the B7 family of negative regulatory signals from APCs, and the PD receptor (on T cells), interacting with the PD ligand-1 (PD-L1) (–) signal coming from tumor cells expressing the PD-L1. Strategies that inhibit CTLA4 and PD-1 function are a means of stimulating cytotoxic T-cell activity to kill tumor cells. Lymphocyte activating gene 3 (LAG3) has recently been shown to promote PD axis immunosuppression, and inhibition of that target has shown clinical value. *Tumor stroma-directed antibodies* cause anti–vascular endothelial cell growth factor (VEGF)–mediated antiangiogenic and tumor interstitial pressure-modulating strategies. Proceeding *clockwise* in the figure from the upper right, *antibody-drug conjugates* have recently been designed to deliver cytotoxic drugs, indicated by "T" (*upper right*). Relatively *nonspecific immunomodulators* include vaccines instilled directly into the tumor stroma, agents such as the "imids" that alter tumor and stromal cell cytokine production, and cytokines such as interferon or interleukin 2 (IL-2), which can affect tumor-infiltrating lymphocyte function or have direct antitumor effects. *Vaccines* targeting tumor cell antigens or live attenuated *oncolytic viruses* injected into tumors can cause tumor cell lysis with induction of a prominent host antitumor immune response to virus antigens and target antigens (T-Ags) (*middle right*). In the *right lower* portion of the figure, strategies to deliver activated immune cells include harvest of autologous T cells that are then infected with a lentivirus or other construct that targets antigens (T-Ags) expressed on tumor cells (*chimeric antigen receptor [CAR] T cells*), with the targeted antigen indicated.

The BiTe antibody blinatumomab was constructed to have an anti-CD19 antigen-combining site directed at a cancer cell as one valency with an anti-CD3 binding site as the other. This antibody can bring T cells (with their anti-CD3 activity) close to neoplastic B cells bearing the CD19 determinant. Blinatumomab is active in B-cell neoplasms such as acute lymphocytic leukemia. Unique toxicities include cytokine release syndrome (fever, hypotension, tachycardia) and neurologic deterioration manifest initially by deterioration in handwriting accuracy, which can proceed to more florid cortical dysfunction, suggesting a need for dose pausing and/or glucocorticoid use.

STROMA-DIRECTED ANTIBODIES (FIG. 78-4) The anti–vascular endothelial growth factor (VEGF) antibody bevacizumab shows evidence of value in renal cancers, where activation of VEGF signaling occurs as part of disabled hypoxia-induced signaling in the tumor cells. When combined with chemotherapeutic agents, it may increase responses in colorectal and nonsquamous lung cancers. The mechanism for this effect may relate to improved delivery and tumor uptake of the cytotoxic, or even immunoregulatory, antibodies used to treat hepatocellular cancer, owing to decreased tumor interstitial pressure. VEGF was originally isolated as a "tumor permeability factor" causing leakiness of tumor blood vessels. When used in gliomas, it may, by decreasing vascular permeability, allow replacement of steroids to decrease intracranial pressure. Bevacizumab has a number of side effects including hypertension, thrombosis, proteinuria, hemorrhage, and gastrointestinal perforations with or without prior surgeries; these adverse events also occur with small-molecule drugs modulating VEGF receptor (VEGFR) function.

IMMUNOREGULATORY ANTIBODIES (FIG. 78-4) Purely immunoregulatory antibodies stimulate immune responses to mediate tumor-directed cytotoxicity. An understanding of the tumor-host interface has revealed that cytotoxic tumor-directed T cells are frequently inhibited by ligands upregulated in the tumor cells. The programmed death ligand 1 (PD-L1; also known as B7-homolog 1) was initially recognized as inducing T-cell death through a receptor present on T cells, termed the programmed death (PD) receptor, which physiologically regulates the intensity of the immune response to any antigen. The PD family of ligands and receptors also regulates macrophage function, present in tumor stroma. These actions raised the hypothesis that antibodies directed against the PD signaling axis (both anti-PD-L1 and anti-PD) might be useful in cancer treatment by allowing reactivation of the immune response against tumors.

Ipilimumab, an antibody directed against the anti-CTLA4 (cytotoxic T lymphocyte antigen 4), which is expressed on T cells (not tumor cells), responds to signals from antigen-presenting cells (Fig. 78-4) and also downregulates the intensity of the T-cell proliferative response to antigens derived from tumor cells. Indeed, manipulation of the CTLA4 axis was the first demonstration that purely immunoregulatory antibody strategies directed at T-cell physiology could be safe and effective in the treatment of cancer. Ipilimumab, alone or in combination with PD-1-directed antibodies, is approved for treatment of metastatic melanoma and lung cancers.

Nivolumab, directed against the PD-1 receptor, and atezolizumab (anti-PD-L1) are exemplary of anti-PD-1-directed immunoregulatory antibodies, with clinical benefits in many cancers **(Table 78-4)**. Pembrolizumab is approved for first-line treatment of metastatic non-small-cell lung cancer tumors that express the PD-L1 ligand. This development was a milestone in cancer therapeutics, replacing chemotherapy in this patient subset. Lymphocyte activation gene 3 (*LAG3*) binds to major histocompatibility complex (MHC) on antigen-presenting cells to augment the effect of the PD-1 and CTLA4 pathways. Relatlimab is a first-in-class inhibitor of LAG3 action and augments the effect of nivolumab in melanoma.

Importantly, the clinical observation that tumors most amenable to treatment with immunoregulatory antibodies were in sites (lung, skin, genitourinary) exposed to environmental carcinogens or occurred in patients with known mutations in DNA repair pathways stimulated specific research as to whether the "mutational burden" of tumors could predict value from anti-PD strategies. Results to date in general support this hypothesis and led to the first regulatory approvals for immunomodulating antibodies in a "tissue agnostic" fashion. Specifically, detection of deficiencies in a tumor DNA mismatch repair system or with evidence of increased tumor mutation burden is a specific indication for use of certain immunoregulatory agents, irrespective of the disease site of origin. The increased efficacy in the setting of higher tumor mutational burden is thought to be due to the presence of more proteins in the tumor structurally altered by mutation that can be recognized as foreign by the immune system.

Prominent autoimmune hepatic, cardiac, endocrine, cutaneous, neurologic, and gastrointestinal adverse events can occur with the use of ipilimumab as well as the PD-1-directed antibodies. Emergency use of glucocorticoids may be required to attenuate severe toxicities. Although theoretically such glucocorticoid use can attenuate the antitumor effect, response rates do not appear to be compromised by their use to abrogate serious organ toxicity attributable to use of immunomodulatory antibodies. Importantly for the general internist, immunologic toxicities can occur late after exposure to the modulators of PD and CTLA4 action, even while the patient may have sustained control of tumor growth.

TABLE 78-4 Clinical Impact of Host T Lymphocyte–Modified Cells[a] or Host T Lymphocyte–Directed Immunoregulatory Antibodies[c]

A. Advanced Cancers with Positive Effect (at least 25% of treated patients have stable disease or progression-free survival of ≥27 weeks or better) or Frequent or Unexpected Prolonged Responders (efficacy may be limited to CD expression–dependent or PD-1 ligand–expressing subtypes)
Acute lymphoid leukemia[b]
Adrenocortical carcinoma[c]
Breast cancer, hormone receptor negative, HER2 negative (with chemotherapy)[c]
Colorectal cancer (microsatellite instability-high [MSI-H] or mismatch repair deficient, usually with fluoropyrimidine, oxaliplatin, or irinotecan)[c]
Cervix, squamous carcinoma[c]
Cutaneous, squamous carcinoma[c]
Diffuse large B-cell non-Hodgkin's lymphoma, not otherwise specified[a]
Diffuse large B-cell non-Hodgkin's lymphoma, primary mediastinal subtype[b]
Endometrial carcinoma (with lenvatinib, if microsatellite instability-stable [MSI-S] or mismatch repair wild-type)[c]
Esophageal squamous carcinoma[c]
Gastric/gastroesophageal adenocarcinoma[c]
Head and neck squamous carcinoma[c]
Hepatocellular cancer (after sorafenib)[c]
Hodgkin's disease[c]
Mantle cell lymphoma[a]
Melanoma[c]
Merkel cell carcinoma[c]
Mesothelioma[c]
MSI-H or mismatch repair–deficient solid tumors without satisfactory alternative[c]
Mycosis fungoides[c]
Multiple myeloma[a]
Non-small-cell lung carcinoma[c]
Paraganglioma/pheochromocytoma[c]
Renal cell carcinoma[c]
Sarcoma, alveolar soft part[c]
Small-cell lung carcinoma[c]
Solid tumors with high tumor mutational burden (TMB) (≥10 mutations/megabase) that have progressed following prior therapy without satisfactory alternative treatment[c]
Urothelial carcinoma[c] (including bladder, ureter)
B. Advanced Cancers with Insufficient Data to Support Host-Derived T Lymphocyte or Immunoregulatory Antibody Treatment[d]
Acute myeloid leukemia
Anus, squamous carcinoma
Breast cancer, hormone receptor positive
Breast cancer, hormone receptor negative, HER2 positive
Biliary tract cancers (if MSI-S or mismatch repair wild-type)
Chronic lymphocytic leukemia
Chronic myeloid leukemia
Gastrointestinal neuroendocrine/islet cell carcinoma
Glioma, all grades including glioblastoma
Germ cell neoplasms
Ovarian cancer
Osteogenic sarcoma
Pancreas adenocarcinoma
Pediatric tumors (Wilms', rhabdomyosarcoma, Ewing's, neuroblastoma, osteosarcoma)
Prostate adenocarcinoma
Salivary gland carcinoma
Soft tissue sarcoma (except alveolar soft part)
T-cell non-Hodgkin's lymphoma (except mycosis fungoides)
Vulva, squamous carcinoma

[a] Chimeric antigen receptor (CAR)-modified autologous T cells in relapsed or refractory cases. [b] Both CAR-modified autologous T cells or an immunoregulatory antibody. [c] T-cell-directed immunoregulatory antibody strategies including anti-PD-1 and/or anti-PD-L1 antibodies or bispecific tumor engaging (BiTe) antibodies against a particular tumor cell antigen. [d] Unless MSI-high, mismatch repair deficient, or TMB ≥10 mutations/megabase.

Nontargeted Immunomodulators (Fig. 78-4) Bacille Calmette-Guérin, a killed mycobacterial product, invokes a useful immune response when instilled locally into the bladder in the setting of preinvasive bladder cancers. The "imids" thalidomide, lenalidomide, and pomalidomide alter cytokine elaboration in the tumor microenvironment and have antiangiogenic actions. They are a cornerstone in the management of multiple myeloma. Thromboses (warranting consideration of prophylactic anticoagulation), gastrointestinal and neuropathic adverse events, and prominent teratogenicity can occur as a consequence of their use. "High-dose" glucocorticoids stimulate apoptosis in normal and neoplastic lymphoid cells and are a mainstay in the treatment of lymphoid leukemias, lymphomas, and plasma cell neoplasms.

Cytokines Only interferon α (IFN-α) and interleukin 2 (IL-2) are considered for current treatment indications. IFN is not curative for any tumor but can induce partial responses in follicular lymphoma, hairy cell leukemia, chronic myeloid leukemia, melanoma, and Kaposi's sarcoma. It produces fever, fatigue, a flulike syndrome, malaise, myelosuppression, and depression and can induce clinically significant autoimmune disease.

IL-2 exerts its antitumor effects indirectly through augmentation of immune function. Its biologic activity is to promote the growth and activity of T cells and natural killer (NK) cells. High doses of IL-2 can produce tumor regression in certain patients with metastatic melanoma and renal cell cancer. About 2–5% of patients may experience complete remissions that are durable. Patients may require blood pressure support and intensive care to manage the toxicity. However, once the agent is stopped, most of the toxicities reverse completely within 3–6 days. Use of "high-dose" IL-2 regimens has been superseded by immunoregulatory antibodies. Efforts to develop new IL-2-based strategies with less toxicity and improved efficacy are continuing.

T Cell–Mediated Therapies Three types of currently used cancer treatments take advantage of the ability of T cells to kill tumor cells.

1. *Transfer of allogeneic T cells.* This occurs in three major settings: in allogeneic bone marrow transplantation; as purified lymphocyte transfusions following bone marrow recovery after allogeneic bone marrow transplantation; and as pure lymphocyte transfusions following immunosuppressive (nonmyeloablative) therapy (also called reduced-intensity or minitransplants). In each of these settings, the effector cells are donor T cells that recognize the tumor as being foreign, probably through minor histocompatibility differences. The main risk of such therapy is the development of graft-versus-host disease because of the minimal difference between the cancer and the normal host cells. This approach has been useful in certain hematologic cancers refractory to chemotherapeutic strategies.
2. *Transfer of autologous T cells.* In this approach, the patient's own T cells are removed from the tumor-bearing host, manipulated in several ways in vitro, and given back to the patient. Tumor antigen–specific T cells can be developed after retroviral transduction of the desired T-cell antigen receptor and expanded to large numbers over many weeks ex vivo before administration. These chimeric antigen receptor (CAR) T cells (Fig. 78-4) have evidence of sustained value in patients with refractory hematopoietic neoplasms such as diffuse large B-cell lymphoma, multiple myeloma, and mantle cell lymphoma. Prominent adverse effects include *cytokine release syndrome* (fever, tachycardia, hypotension) and neurologic manifestations. Clinical investigations are seeking to develop solid-tumor antigen-directed CAR strategies, as well as to utilize different immune cell populations such as NK cells to cause useful antitumor activity in ways that may allow "off-the-shelf" products not requiring manipulation and purification of patients' autologous cells.
3. *Tumor vaccines aimed at boosting T-cell immunity.* Two types of vaccine approaches are currently approved. Purified autologous antigen-presenting cells can be pulsed with tumor, its membranes, or particular tumor antigens and delivered as a vaccine. Vaccine adjuvants such as granulocyte-macrophage colony-stimulating factor (GM-CSF) may be co-administered. One such vaccine, sipuleucel-T, is approved for use in patients with asymptomatic or minimally symptomatic metastatic hormone-independent prostate cancer. In this approach, the patient undergoes leukapheresis, wherein mononuclear cells (that include antigen-presenting cells) are removed from the patient's blood. The cells are pulsed in a laboratory with an antigenic fusion protein comprising a protein frequently expressed by prostate cancer cells, prostate acid phosphatase, fused to GM-CSF, and matured to increase their capacity to present the antigen to immune effector cells. The cells are then returned to the patient in a well-tolerated treatment. Although no objective tumor response was documented in clinical trials, median survival was increased by about 4 months.

   Another important vaccine strategy is directed at infectious agents whose action ultimately is tied to the development of human cancer. Hepatitis B vaccine in an epidemiologic sense prevents hepatocellular carcinoma, and a tetravalent human papillomavirus vaccine prevents infection by virus types currently accounting for 70% of cervical cancer. Unfortunately, these vaccines are ineffective at treating patients who have developed a virus-induced cancer.

Oncolytic or Immunomodulating Viruses (Fig. 78-4) Laboratory studies in animals have utilized viruses to destroy tumors because tumor cells lack endogenous host mechanisms, e.g., IFN elaboration or recognition strategies of viral nucleic acids, that limit virus spread. Viral infection of tumors also can stimulate a prominent host response to viral and tumor cell antigens, leading to immune effects against local tumor cells. Talimogene laherparepvec is a clinically approved attenuated herpes virus that acts to stimulate immune responses when instilled locally into melanoma deposits. Systemic effects are minimal in this application. This general strategy is being considered particularly in tumors not amenable to useful effects of currently approved immunoregulatory antibodies or in conjunction with immunoregulatory antibodies. Nadofaragene firadenovec is an adenovirus construct instilled into the bladder; it is nonreplicating but can infect urothelial cells to deliver a gene resulting in consistently produced IFN-α2b, which has a local antitumor effect. It is useful in the treatment of superficial bladder cancers. It is not established whether infection of tumor cells, normal urothelial cells, or both contribute to the antitumor effect.

■ CANCER CYTOTOXIC THERAPY

Table 78-5 lists commonly used cytotoxic cancer chemotherapy agents and pertinent clinical aspects of their use, with particular reference to adverse effects that might be encountered by the generalist in the care of patients. The drugs were initially discovered through screening of chemicals and natural product extracts for evidence of antitumor activity in animals or were designed with knowledge of biochemical pathways affecting nucleic acid synthesis. They may be usefully grouped into two general categories: those affecting DNA and those affecting microtubules.

As illustrated in Fig. 78-3, disruption of DNA or microtubule integrity is a major trigger of cellular apoptosis pathways. An additional factor in drug effect stems from recent observations that tumor cells have increased tolerance of specific types of DNA damage owing to defects in DNA repair pathways. This state is thought to facilitate the survival of the neoplastic clone as it experiences DNA mutations during the course of carcinogenesis. DNA-directed cytotoxic agents can create lesions in DNA that are poorly tolerated by cells with neoplasm-promoting DNA repair pathway mutations. This results in a "synthetic lethal" interaction of the drug with cells bearing the DNA repair pathway mutation. Examples of a potential "synthetic lethal effect" will be pointed out in relation to clinical applications below.

DNA-Interactive Agents DNA replication occurs during the synthesis or S-phase of the cell cycle, with chromosome segregation of the replicated DNA in the M, or mitosis, phase. The G_1 and G_2 "gap phases" precede S and M, respectively. Chemotherapeutic agents have been divided into "phase-nonspecific" agents, which can act in any phase of the cell cycle, and "phase-specific" agents, which require the cell to be at a particular cell cycle phase to cause greatest effect.

TABLE 78-5 Commonly Used Cytotoxic Chemotherapy Agents

DRUG	ADVERSE EVENTS	NOTES
Direct DNA-Interacting Agents[a]		
Alkylator or platinating drug		
Bendamustine	Vehicle allergy, My, Der, ↑LFTs	TLS, Ves, IR, DA-R, DA-H
Carboplatin	My, N, V, R	Dose according to CrCl: to AUC of 5–7 mg/mL per min [AUC = dose/(CrCl + 25)]
Chlorambucil	Common alkylator	
Cisplatin	N, V, Neu, My, R, Ototoxic, ↓K^+, ↓Mg^{2+}, ↓Ca^{2+}	Osmotic diuresis, N, V prophylaxis, DA-R
Cyclophosphamide	Common alkylator, cystitis, cardiac (high dose)	Liver required to activate, DA-R, DA-H, hydration ± mesna protects bladder
Dacarbazine (DTIC)	Common alkylator, Ves	DA-R
Ifosfamide	My, common alkylator, bladder, CNS	DA-R, must use concomitant mesna
Lomustine (CCNU)	Common alkylator but My, has delayed nadir, ↑LFTs	Plm ± fibrosis: PFTs prior to treatment and repeat frequently; cease if fibrosis occurs
Lurbinectedin	My, ↑LFTs, N, V	CYP3A4
Melphalan	Common alkylator but My, has delayed nadir	DA-R
Oxaliplatin	N, V, My, Neu	Reversible laryngopharyngeal spasm risk
Procarbazine	Common alkylator, CNS	Disulfiram-like effect with alcohol, MAOI, like HBP after tyramine-rich foods
Temozolomide	Common alkylator but My, has delayed nadir	*Pneumocystis* prophylaxis
Antitumor antibiotics and topoisomerase poisons		
Bleomycin	Plm (↑FIO_2 worsen), Der, Raynaud's, IR*	Monitor DLCO before/during treatment, DA-R
Dactinomycin	My, N, V, mucositis, Ves, alopecia	Radiation recall
Doxorubicin, daunorubicin, epirubicin, idarubicin	TOPOII; My, mucositis, alopecia, ↓LVEF acute/chronic, Ves	Co-administration with heparin aggregate, secondary leukemia, DA-H, radiation recall
Doxorubicin, liposomal	TOPOII; My, ↓LVEF, IR*, PPED	DA-H, radiation recall
Etoposide	TOPOII; My, alopecia, IR with rapid IV, N, V, mucositis	DA-H, DA-R
Irinotecan	TOPOI; My, D: "early onset" with cramping, flushing, vomiting: treat with atropine; "late onset" after several doses: use loperamide 4 mg with first stool then 2 mg q2h until 12 h without stool up to 16 mg/24 h; My dependent on UGT1A1 phenotype, R, ILD	Prodrug requires metabolism to active drug SN-38, which is cleared by UGT1A1 with degree of My dependent on patient UGT1A1 genotype, DA-H
Irinotecan, liposomal	TOPOI; My, D (administer loperamide for D of any severity), IR*, ILD	Consider guide dosing by UGT1A1 genotype testing, CYP3A4; NO recommended dose for T Bili >ULN
Mitoxantrone	TOPOII; Ves, blue urine, nails, and sclerae	Interacts with heparin; DA-H, alopecia, N, V, radiation recall
Topotecan	TOPOI; My, mucositis, N, V, alopecia	DA-R, rare ILD
Indirectly DNA-Interacting Agents		
Pyrimidine analogues		
Capecitabine	My, D, PPED, cardiac adverse events, R, Der	Oral prodrug of 5-FU
Cytarabine (cytosine arabinoside [ara-C])	My, mucositis, CNS (high dose), conjunctivitis (high dose; use steroid eyedrops until 72 h after last dose), noncardiogenic pulmonary edema	Metabolized in tissues by deamination but renal excretion prominent at doses >500 mg; therefore, DA-R in "high-dose" regimens
5-Fluorouracil (5-FU)	My, D, mucositis, CNS, Der, cardiac adverse events, PPED	Toxicity *enhanced* by leucovorin by increasing "ternary complex" with thymidylate synthase; metabolism in tissue; dihydropyrimidine dehydrogenase deficiency increases toxicity; enhances warfarin effect
Gemcitabine	My, N, V, ↑LFT, fever/"flu syndrome"	Rare ARDS; rare HUS, rare PRES, radiosensitization with long infusion
Trifluridine/tipiracil	My, mucositis, N, V, unusual PPED	Trifluridine directly inhibits thymidylate synthase and is incorporated into DNA; tipiracil inhibits thymidine phosphorylase, which degrades trifluridine
Purine analogues		
Fludarabine phosphate	My, mucositis, CNS, Der	Converted to F-ara ATP in cells by deoxycytidine kinase; DA-R
6-Mercaptopurine (6-MP), 6-thioguanine (6-TG)	Mv, N, ↑LFT	6-MP metabolized by xanthine oxidase, decrease dose with allopurinol; 6-MP and 6-TG increased toxicity with thiopurine methyltransferase deficiency
Antifolates		
Methotrexate	My, ↑LFT with fibrosis with chronic use, Plm, R, mucositis	Toxicity *lessened* by leucovorin, bypassing block of dihydrofolate reductase, excreted in urine; DA-R or hold; NSAIDs increase renal toxicity; sequestered in third space fluids
Pemetrexed	My	Supplement folate/B_{12}, omit for CrCl <45 mL/min
Miscellaneous Antimetabolite-Like Agents		
Asparaginase	Thrombosis by decrease of antithrombin IIII, but ↓fibrinogen can cause hemorrhage; ↑glucose; ↓albumin, hypersensitivity; CNS; pancreatitis; ↑LFTs	Decrease protein synthesis; indirect inhibition of DNA synthesis by decreased histone synthesis; blocks methotrexate action
Hydroxyurea	My, N, mucositis, rare ↓CrCl	DA-R, augments antimetabolite effect

(*Continued*)

TABLE 78-5 Commonly Used Cytotoxic Chemotherapy Agents (Continued)

DRUG	ADVERSE EVENTS	NOTES
Antimitotic agents		
Docetaxel	IR*, VLS, My, ↑LFTs, Der, Neu, stomatitis, alopecia, N, V, D	Premedicate with steroids, H_1 and H_2 blockers; DA-H; monitor for second primary malignancies; alcohol in vehicle
Eribulin	My, Neu, ↑QT	DA-H, DA-R
Ixabepilone	My, Neu, IR*, N, V, D, alopecia	Premedicate with steroids, H_1 and H_2 blockers; CYP3A4, DA-H
Nab-paclitaxel	My, Neu, Ves	DA-H, CYP3A4, CYP2C8
Paclitaxel	IR*, My, alopecia, N, V, D, mucositis, Neu, Ves	Premedicate with steroids, H_1 and H_2 blockers; DA-H, CYP3A4, CYP2C8, alcohol in vehicle
Vinblastine	My, Ves, Neu, HBP, Raynaud's, Ves, ↑LFTs, ileus/constipation (use prophylactic stool softeners)	DA-H, unusual; Plm, CYP3A4
Vincristine	Ves, Neu, SIADH, ileus/constipation (use prophylactic stool softeners)	DA-H, CYP3A4; less My than with vinblastine; unusual Plm
Vinorelbine	My, Ves, allergic bronchospasm (immediate), dyspnea/cough (subacute)	DA-H; less neurotoxic than other vincas

Note: Data abstracted in part from publicly available U.S. Food and Drug Administration label. All agents in this class have the potential for prominent embryofetal toxicity; use without contraception by female patients of childbearing potential is not recommended. Effective contraception for female partners who are of childbearing potential of patients undergoing treatment should also be considered; use during lactation is also not recommended; all DNA interacting agents have theoretical risk of late secondary neoplasms, particularly where noted. Indications and events of prominent general medical importance include the following: ARDS, acute respiratory distress syndrome; AUC, area under concentration-time curve; Common alkylator toxicities: N, V, My, alopecia, mucositis, ↓fertility, cumulative lung toxicity; CNS, central nervous system (can include altered sensorium, cortical and cerebellar signs, dysarthria, altered seizure threshold); Cor, corneal keratopathy, consider pretreatment ophthalmologic exam with prophylactic lubricating eyedrops ± ocular steroids; CrCl, creatinine clearance; CYP___, interaction with drugs metabolized by the indicated cytochrome(s) P450; DA-H, dose adjust for hepatic dysfunction; DA-R, dose adjust for renal dysfunction; Der, dermatologic toxicity; HBP, high blood pressure; HUS, hemolytic-uremic syndrome; ILD, interstitial lung disease; IR, infusion reaction; IR*, infusion reaction prophylaxis specifically recommended; LFT, liver function tests; ↓LVEF, left ventricular ejection fraction, monitor echocardiogram; My, anemia, decrease in white blood cells, platelets, with risk of neutropenic fever, hemorrhage, therefore dose reduce or hold dose for decreased neutrophil or platelet counts unless marrow infiltrated by drug-responsive tumor; N, nausea; Neu, peripheral neuropathy; Plm, pulmonary; PPED, palmar-plantar erythrodysesthesia (i.e., hand-foot syndrome); PRES, posterior reversible leukoencephalopathy syndrome; PT, prothrombin time; PTT, partial thromboplastin time; ↑QT, pre- and intratreatment electrocardiogram monitoring, normalize K^+, Mg^{2+}, ionized Ca^{2+}; R, renal injury possible; SIADH, syndrome of inappropriate antidiuretic hormone; TLS, risk of tumor lysis syndrome if brisk response; TOPO___, agent targets the indicated topoisomerase; UGT___, metabolism by UDP-glucuronosyltransferase of the indicated genotype; Ves, extravasation injury possible; V, vomiting; VLS, vascular leak syndrome; VOD, veno-occlusive liver disease.

Alkylating agents (Table 78-5) as a class are cell cycle phase–nonspecific agents. They break down, either spontaneously or after normal organ or tumor cell metabolism, to reactive intermediates that covalently modify bases in DNA. This leads to cross-linkage of DNA strands or the appearance of breaks in DNA as a result of repair efforts. Damaged DNA cannot complete normal cell division; in addition, it activates apoptosis. Alkylating agents share common toxicities: myelosuppression, alopecia, gonadal dysfunction, mucositis, and pulmonary fibrosis. They also share the capacity to cause "second" neoplasms, particularly leukemia, years after use, particularly when used in low doses for protracted periods.

Cyclophosphamide is inactive unless metabolized by the liver to 4-hydroxy-cyclophosphamide, which decomposes into an alkylating species, as well as to chloroacetaldehyde and acrolein. The latter causes chemical cystitis; therefore, excellent hydration must be maintained while using cyclophosphamide. If severe, the cystitis may be attenuated or prevented altogether (if expected from the dose of cyclophosphamide to be used) by mesna (2-*me*rcapto*e*thane*s*ulfo*na*te). Liver disease impairs cyclophosphamide activation. Sporadic interstitial pneumonitis leading to pulmonary fibrosis can accompany the use of cyclophosphamide, and high doses used in conditioning regimens for bone marrow transplant can cause cardiac dysfunction. Ifosfamide is a cyclophosphamide analogue also activated in the liver, but more slowly, and it requires co-administration of mesna to prevent bladder injury. CNS effects, including somnolence, confusion, and psychosis, can follow ifosfamide use; the incidence appears related to low body surface area or decreased creatinine clearance.

Several alkylating agents are less commonly used. Bendamustine has activity in chronic lymphocytic leukemia and certain lymphomas. It is used in transplant preparation regimens. Melphalan shows variable oral bioavailability and undergoes extensive binding to albumin and α_1-acidic glycoprotein. Mucositis appears more prominently; however, it has prominent activity in multiple myeloma. Nitrosoureas break down to carbamylating species that not only cause a distinct pattern of DNA base pair–directed reactivity but also can covalently modify proteins. Lomustine is used to treat brain tumors but causes delayed myelotoxicity and can cause lung injury. Procarbazine is metabolized in the liver and possibly in tumor cells to yield a variety of free radical and alkylating species. In addition to myelosuppression, it causes hypnotic and other CNS effects, including vivid nightmares. It can cause a disulfiram-like syndrome on ingestion of ethanol. Both procarbazine and lomustine are used in the treatment of brain tumors. Dacarbazine (DTIC) is activated in the liver to yield the highly reactive methyl diazonium cation. It causes only modest myelosuppression 21–25 days after a dose but causes prominent nausea on day 1. It is an important component of treatment regimens for Hodgkin's lymphoma and sarcomas. Temozolomide is structurally related to DTIC but is activated by nonenzymatic hydrolysis in tumors and is bioavailable orally. Brain tumors with alkylguanine alkyl transferase deficiency are selectively susceptible to temozolomide, which alkylates the O^6 position of guanine.

Cisplatin was discovered fortuitously by observing that bacteria present in electrolysis solutions with platinum electrodes could not divide. Only the *cis* diamine configuration is active as an antitumor agent. In tumor cells, a chloride is lost from each position. The resulting positively charged species is an efficient DNA interactor, forming Pt-based cross-links. Therefore "platinating agents" are considered with alkylating agents as forming related lesions in DNA. Cisplatin is administered with abundant hydration, including forced diuresis with mannitol to prevent kidney damage; even with the use of hydration, gradual decrease in kidney function is common, along with noteworthy anemia. Hypomagnesemia frequently attends cisplatin use and can lead to hypocalcemia and tetany. Other common toxicities include neurotoxicity with stocking-and-glove sensorimotor neuropathy. Hearing loss occurs in 50% of patients treated with conventional doses. Cisplatin is intensely emetogenic, requiring prophylactic antiemetics. Myelosuppression is less evident than with other alkylating agents. Chronic vascular toxicity (Raynaud's phenomenon, coronary artery disease) is a more unusual toxicity. Carboplatin displays less nephro-, oto-, and neurotoxicity. However, myelosuppression is more frequent, and because the drug is exclusively cleared through the kidney, adjustment of dose for creatinine clearance must be accomplished through use of various dosing nomograms. Oxaliplatin is a platinum analogue with noteworthy activity in colon cancers refractory to other treatments. It is prominently neurotoxic.

Lurbinectedin binds to DNA through the "DNA minor groove" with covalent interaction with the N2 position of certain guanines. Transient altered liver function can occur, as well as cytopenias. Lurbinectedin has activity in small-cell lung cancer. The first example of this agent class, trabectedin, requires a prolonged infusion schedule but is active in certain sarcomas, in part due to its modulation of transcription factor function.

Antitumor Antibiotics and Topoisomerase Poisons Antitumor antibiotics are substances produced by bacteria that provide a chemical defense against hostile microorganisms. They bind to DNA directly and can frequently undergo electron transfer reactions to generate free radicals in close proximity to DNA, leading to DNA damage in the form of single-strand breaks or cross-links. Topoisomerase poisons include natural products or semisynthetic derivatives that modify enzymes that allow DNA to unwind during replication or transcription. These include topoisomerase I, which creates single-strand breaks that then rejoin following the passage of the other DNA strand through the break. Topoisomerase II creates double-strand breaks through which another segment of DNA duplex passes before rejoining. Owing to the role of topoisomerase I in the replication fork, topoisomerase I poisons cause lethality if the topoisomerase I–induced lesions occur in S-phase.

Doxorubicin intercalates into DNA, thereby altering DNA structure, replication, and topoisomerase II function. It can also undergo reduction of its quinone ring system, with reoxidation to form reactive oxygen radicals. It causes predictable myelosuppression, alopecia, nausea, and mucositis. In addition, it can cause acute cardiotoxicity in the form of atrial and ventricular dysrhythmias, but these are rarely of clinical significance. In contrast, cumulative doses >550 mg/m^2 are associated with a 10% incidence of chronic cardiomyopathy. The incidence of cardiomyopathy appears to be related to peak serum concentration, with low-dose, frequent treatment or continuous infusions better tolerated than intermittent higher-dose exposures. Cardiotoxicity has been related to iron-catalyzed oxidation and reduction of doxorubicin in the heart. Dexrazoxane is an intracellular chelating agent that can act as a cardio-protectant. Doxorubicin's cardiotoxicity is increased when given together with trastuzumab, the anti-HER2/neu antibody. Radiation recall or interaction with concomitantly administered radiation to cause local site complications is frequent. The drug is a powerful vesicant, with necrosis of tissue apparent 4–7 days after an extravasation; therefore, it should be administered into a rapidly flowing intravenous line. Dexrazoxane also can mitigate doxorubicin extravasation. Doxorubicin is metabolized by the liver, so doses must be reduced by 50–75% in the presence of liver dysfunction. Daunorubicin is closely related to doxorubicin and is preferable to doxorubicin owing to less mucositis and colonic damage with frequent high doses used in the curative treatment of leukemia. Idarubicin is also used in leukemia treatment and may have somewhat less cardiotoxicity. Encapsulation of daunorubicin into a liposomal formulation has attenuated cardiac toxicity with antitumor activity in Kaposi's sarcoma, other sarcomas, multiple myeloma, and ovarian cancer.

Mitoxantrone is a synthetic topoisomerase II–directed agent with a mechanism similar to the anthracyclines, with less but not absent cardiotoxicity, comparing the ratio of cardiotoxic to effective doses; it is still associated with a 10% incidence of cardiotoxicity at cumulative doses of >150 mg/m^2. Etoposide binds directly to topoisomerase II and DNA in a reversible ternary complex. It stabilizes the covalent intermediate in the enzyme's action where the enzyme is covalently linked to DNA. Prominent clinical effects include myelosuppression, nausea, and transient hypotension related to the speed of administration of the agent. Etoposide is a mild vesicant but is relatively free from other large-organ toxicities.

Camptothecins target topoisomerase I. Topotecan is a camptothecin derivative approved for use in gynecologic tumors and small-cell lung cancer. Toxicity is limited to myelosuppression and mucositis. Irinotecan is a camptothecin with evidence of activity in colorectal carcinoma. Irinotecan is a prodrug, metabolized in the liver to SN-38, its active metabolite. Levels of SN-38 are particularly high in the setting of Gilbert's disease, characterized by defective uridine diphosphate glucuronosyl transferase (UGT) 1A1 and indirect hyperbilirubinemia, a condition that affects ~10% of the white population in the United States. In addition, irinotecan's myelosuppression is clearly influenced by the patient's genotype for UGT1As. Irinotecan causes a delayed (48–72 h) secretory diarrhea related to the toxicity of SN-38. The diarrhea can be treated effectively with loperamide or octreotide; immediate diarrhea when it occurs is responsive to atropine.

Bleomycin remains an important component of curative regimens for Hodgkin's lymphoma and germ cell neoplasms. It forms complexes with Fe^{2+} while also bound to DNA. Oxidation of Fe^{2+} gives rise to superoxide and hydroxyl radicals, causing DNA damage. The drug causes little, if any, myelosuppression. Bleomycin is cleared rapidly, but augmented skin and pulmonary toxicity in the presence of renal failure necessitates dose reduction in renal failure. Bleomycin is not a vesicant and can be administered intravenously, intramuscularly, or subcutaneously. Common side effects include fever and chills, facial flush, and Raynaud's phenomenon. The most feared complication of bleomycin treatment is pulmonary fibrosis, which increases in incidence at >300 cumulative units administered and is minimally responsive to treatment (e.g., glucocorticoids). The earliest indicator of an adverse effect is usually a decline in the carbon monoxide diffusing capacity (DLCO) or coughing, although cessation of drug immediately upon documentation of a decrease in DLCO may not prevent further decline in pulmonary function. Bleomycin is inactivated by a bleomycin hydrolase, which is poorly expressed in skin and lung. Because bleomycin-dependent electron transport is dependent on O_2, bleomycin toxicity may become apparent after exposure to transient very high fraction of inspired oxygen (FIO_2) even late after treatment. Thus, during surgical procedures, patients with prior exposure to bleomycin should be maintained on the lowest FIO_2 consistent with maintaining adequate tissue oxygenation.

Dactinomycin interacts directly with DNA to inhibit RNA transcription. It is important in the curative treatment of pediatric neoplasms, some of which also occur in young adults. Prominent myelosuppression, mucositis, alopecia, radiation recall, and nausea require management.

Antimetabolites A broad definition of antimetabolites would include compounds that interfere with purine or pyrimidine synthesis. Some antimetabolites also cause DNA damage indirectly, through misincorporation into DNA. They tend to convey greatest toxicity to cells in S-phase, and the degree of toxicity increases with duration of exposure. Common toxic manifestations include stomatitis, diarrhea, and myelosuppression.

Methotrexate inhibits dihydrofolate reductase, which regenerates reduced folates from the oxidized folates produced when thymidine monophosphate is formed from deoxyuridine monophosphate. Without reduced folates, cells die a "thymine-less" death. *N*5-Tetrahydrofolate or *N*5-formyltetrahydrofolate (leucovorin) can bypass this block and rescue cells from methotrexate, which is retained in cells by polyglutamylation. Methotrexate is transported into cells by a membrane carrier, and high concentrations of drug can bypass this carrier and allow diffusion of drug directly into cells. These properties have suggested the design of "high-dose" methotrexate regimens with leucovorin rescue of normal marrow and mucosa as part of curative approaches to osteosarcoma in the adjuvant setting and hematopoietic neoplasms of children and adults. Methotrexate is cleared by the kidney via both glomerular filtration and tubular secretion, and toxicity is augmented by renal dysfunction and drugs such as salicylates, probenecid, and nonsteroidal anti-inflammatory agents that undergo tubular secretion. With normal renal function, 15 mg/m^2 leucovorin will rescue 10^{-8}–10^{-6} *M* methotrexate in 3–4 doses. However, with decreased creatinine clearance, doses of 50–100 mg/m^2 are continued until methotrexate levels are $<5 \times 10^{-8}$ *M*. In addition to bone marrow suppression and mucosal irritation, methotrexate can cause renal failure itself at high doses owing to crystallization in renal tubules; therefore, high-dose regimens require alkalinization of urine with increased flow by hydration. Methotrexate can be sequestered in third-space collections and diffuse back into the general circulation, causing prolonged myelosuppression. Less frequent adverse effects include reversible increases in transaminases and hypersensitivity-like pulmonary syndrome. Chronic low-dose methotrexate can cause hepatic fibrosis. When administered to the intrathecal space, methotrexate can cause chemical arachnoiditis and CNS dysfunction.

Pemetrexed is a folate-directed antimetabolite that inhibits the activity of several enzymes, including thymidylate synthetase (TS), dihydrofolate reductase, and glycinamide ribonucleotide formyltransferase. To avoid toxicity to normal tissues, pemetrexed is given with low-dose folate and vitamin B_{12} supplementation. Pemetrexed has notable activity against certain lung cancers and, in combination with cisplatin, also against mesotheliomas.

5-Fluorouracil (5-FU) represents an early example of "rational" drug design in that tumor cells incorporate radiolabeled uracil more efficiently into DNA than normal cells. 5-FU is metabolized in cells to 5′FdUMP, which inhibits TS. In addition, misincorporation can lead to single-strand breaks, and RNA can aberrantly incorporate FUMP. 5-FU is metabolized by dihydropyrimidine dehydrogenase, and deficiency of this enzyme can lead to excessive toxicity from 5-FU. Oral bioavailability varies unreliably, but prodrugs such as capecitabine have been developed that allow at least equivalent activity to parenteral 5-FU-based approaches. Intravenous administration of 5-FU leads to bone marrow suppression after short infusions but to stomatitis after prolonged infusions. Leucovorin augments the activity of 5-FU by promoting formation of the ternary covalent complex of 5-FU, the reduced folate, and TS. Less frequent toxicities include CNS dysfunction, with prominent cerebellar signs, and endothelial toxicity manifested by thrombosis, including pulmonary embolus and myocardial infarction. Trifluridine is a fluorinated pyrimidine that as the triphosphate is directly incorporated into DNA, evoking DNA damage, and as the monophosphate can inhibit TS. It is administered as a fixed-dose combination with tipiracil, an inhibitor of trifluridine degradation by thymidine phosphorylase.

Cytosine arabinoside (ara-C) is incorporated into DNA after formation of ara-CTP, resulting in S-phase–related toxicity. Continuous infusion schedules allow maximal efficiency, with uptake maximal at 5–7 μ*M*. Ara-C can be administered intrathecally. Adverse effects include nausea, diarrhea, stomatitis, chemical conjunctivitis, and cerebellar ataxia. Gemcitabine is a cytosine derivative that is similar to ara-C in that it is incorporated into DNA after anabolism to the triphosphate, rendering DNA susceptible to breakage and repair synthesis, which differs from that in ara-C in that gemcitabine-induced lesions are very inefficiently removed. In contrast to ara-C, gemcitabine appears to have useful activity in a variety of solid tumors, with limited nonmyelosuppressive toxicities.

6-Thioguanine and 6-mercaptopurine (6-MP) are used in the treatment of acute lymphoid leukemia. Although administered orally, they display variable bioavailability. 6-MP is metabolized by xanthine oxidase and therefore requires dose reduction when used with allopurinol. 6-MP is also metabolized by thiopurine methyltransferase; genetic deficiency of thiopurine methyltransferase results in excessive toxicity.

Fludarabine phosphate is a prodrug of F-adenine arabinoside (F-ara-A). F-ara-A is incorporated into DNA and can cause delayed cytotoxicity even in cells with low growth fraction, including chronic lymphocytic leukemia and follicular B-cell lymphoma. CNS and peripheral nerve dysfunction and T-cell depletion leading to opportunistic infections can occur in addition to myelosuppression.

Agents that indirectly affect purine and pyrimidine metabolism with anti-metabolite-like effects include hydroxyurea, which reversibly inhibits ribonucleotide reductase, resulting in S-phase block. It is orally bioavailable and useful for the acute management of myeloproliferative states. Asparaginase is a bacterial enzyme that causes breakdown of extracellular asparagine required for protein synthesis in certain leukemic cells. This effectively stops tumor cell DNA synthesis, as DNA synthesis requires concurrent protein synthesis. Because asparaginase is a foreign protein, hypersensitivity reactions are common, as are effects on organs such as pancreas and liver that normally support continuing protein synthesis of secreted products. This may result in decreased insulin secretion with hyperglycemia, with or without hyperamylasemia and clotting function abnormalities. Close monitoring of clotting functions should accompany use of asparaginase. Paradoxically, owing to depletion of rapidly turning over anticoagulant factors, thromboses particularly affecting the CNS may also be seen with asparaginase.

Mitotic Spindle Inhibitors Microtubules form the mitotic spindle, and in interphase cells, they are responsible for the cellular "scaffolding" along which various motile and secretory processes occur. Microtubules are composed of repeating heterodimers of α and β isoforms of the protein tubulin. Vincristine binds to the tubulin heterodimer with the result that microtubules are disaggregated. This results in the block of growing cells in M-phase, where a structurally disordered mitotic spindle apparatus is a powerful proapoptotic signal (Fig. 78-3). Vincristine is metabolized by the liver, and dose adjustment in the presence of hepatic dysfunction is required. It is a powerful vesicant, and infiltration can be treated by local heat and infiltration of hyaluronidase. At clinically used intravenous doses, neurotoxicity in the form of glove-and-stocking neuropathy is frequent. Acute neuropathic effects include jaw pain, paralytic ileus, urinary retention, and the syndrome of inappropriate antidiuretic hormone secretion. Myelosuppression is not seen at conventional doses. Vinblastine is similar to vincristine, except that it tends to be more myelotoxic, with more frequent thrombocytopenia and also mucositis and stomatitis. Vinorelbine is a vinca alkaloid that appears to have differences in resistance patterns in comparison to vincristine and vinblastine; it may be administered orally.

The taxanes include paclitaxel and docetaxel. These agents differ from the vinca alkaloids in that the taxanes stabilize microtubules against depolymerization. The "stabilized" microtubules function abnormally and are not able to undergo the normal dynamic changes of microtubule structure and function necessary for cell cycle completion. Taxanes are among the most broadly active antineoplastic agents for use in solid tumors, with evidence of activity in ovarian, breast, prostate, and non-small cell lung cancers, and Kaposi's sarcoma. They are administered intravenously, and their vehicles cause hypersensitivity reactions. Premedication with dexamethasone (8–16 mg orally or intravenously 12 and 6 h before treatment) and diphenhydramine (50 mg) and cimetidine (300 mg), both 30 min before treatment, decreases but does not eliminate the risk of hypersensitivity reactions to the paclitaxel vehicle. A protein-bound formulation of paclitaxel (called *nab-paclitaxel*) has at least equivalent antineoplastic activity and decreased risk of hypersensitivity reactions. Paclitaxel may also cause myelosuppression, neurotoxicity in the form of glove-and-stocking numbness, and paresthesia. Docetaxel causes comparable degrees of myelosuppression and neuropathy. Docetaxel uses a different vehicle that can cause fluid retention in addition to hypersensitivity reactions; dexamethasone premedication with or without antihistamines is also generally used. Cabazitaxel is a taxane with somewhat better activity in prostate cancers than earlier generations of taxanes, perhaps due to superior delivery to sites of disease.

Epothilones represent a class of microtubule-stabilizing agents optimized for activity in taxane-resistant tumors. Ixabepilone has clear evidence of activity in breast cancers resistant to taxanes and anthracyclines such as doxorubicin. Side effects include myelosuppression and peripheral sensory neuropathy. Eribulin is a microtubule-directed agent with activity in patients who have had progression of disease on taxanes. It alters dynamics of microtubule remodeling in cells.

ANTIBODY-DRUG CONJUGATES

In an effort to improve therapeutic index and enhance antitumor effect, antibody-drug conjugates (ADCs; Fig. 78-4) have recently entered clinical practice. In this approach, humanized monoclonal antibodies are covalently linked to cytotoxic agents by "linkers." When bound to target antigens on tumor cells, the nature of the linker determines whether the cytotoxic agent is released only after the ADC is internalized by degradation in a lysosomal compartment, or release of the cytotoxin might occur extracellularly by proteases or physical properties of the tumor microenvironment, in the latter case allowing "bystander" cell killing. Of importance is selection of antibodies with sufficient tumor

TABLE 78-6 Antibody-Drug Conjugates

TOXIC MECHANISM	TARGET ANTIGEN	DISEASE ACTIVITY	NOTES
DNA Structure			
Gemtuzumab ozogamicin	CD33	Acute myeloid leukemia, CD33+	IR*, My, ↑LFTs ± VOD
Inotuzumab ozogamicin	CD22	Pre-B acute lymphoid leukemia	IR*, My, ↑LFTs ± VOD, ↑QT
Loncastuximab tesirine-lpyl	CD19	Diffuse large B-cell lymphoma	VLS, My, Der
Microtubule Structure			
Ado-trastuzumab emtansine	HER2	Breast cancer, HER2+	IR, ↑LFTs, ↓LVEF, My, Plm, Neu
Belantamab mafadotin-blm	CD38	Myeloma	IR, My, Cor
Brentuximab vedotin	CD30	Hodgkin's disease; anaplastic large-cell lymphoma; mycosis fungoides, CD30+	Neu, IR, My, Plm, R, TLS, CYP3A4, ↑LFTs, DA-H, DA-R
Enfortumab vedotin	Nectin-4	Urothelial cancer	↑Glucose ± DKA, Neu, Cor, Ves, Der
Mirvetuximab soravtansine-gynx	Folate receptor α (FRα+)	Ovary, peritoneal, fallopian tube cancer, FRα+	Cor, Neu, My
Polatuzumab vedotin	CD79b	High-grade B-cell lymphoma	Neu, IR*, My, PML, TLS, ↑LFTs
Tisotumab vedotin-tftv	Tissue factor	Uterine cervix cancer	Neu, ↑PT, ↑PTT, Cor, bleeding, Plm
Topoisomerase I			
Fam-trastuzumab deruxtecan-nxki	HER2	HER2 1+/weak BC, HER2 activating mutation+ NSCLC, HER2+ gastric	N/V prophylaxis, ILD prompt evaluate cough, dyspnea, My, ↓LVEF
Sacituzumab govitecan-bziy	Trop2	Triple-negative BC, HR+/HER2– BC, urothelial cancer	IR*, N/V prophylaxis, diarrhea consider loperamide, My esp UGT1A1*28, consider G-CSF

Note: Data abstracted in part from publicly available U.S. Food and Drug Administration label. All agents in this class have the potential for prominent embryofetal toxicity; use without contraception by female patients of childbearing potential is not recommended. Effective contraception for female partners who are of childbearing potential of patients undergoing treatment should also be considered; use during lactation is also not recommended; all DNA interacting agents have theoretical risk of late secondary neoplasms, particularly where noted. Indications and events of prominent general medical importance include the following: BC, breast cancer; Cor, corneal keratopathy, consider pretreatment ophthalmologic exam with prophylactic lubricating eyedrops ± ocular steroids; CYP___, interaction with drugs metabolized by the indicated cytochrome(s) P450; DA-H, dose adjust for hepatic dysfunction; DA-R, dose adjust for renal dysfunction; Der, dermatologic toxicity; DKA, diabetic ketoacidosis; G-CSF, granulocyte colony-stimulating factor; HER2, human epidermal growth factor receptor 2; HR, hormone receptor; ILD, interstitial lung disease; IR, infusion reaction; IR*, infusion reaction prophylaxis specifically recommended; LFT, liver function tests; ↓LVEF, left ventricular ejection fraction, monitor echocardiogram; My, anemia, decrease in white blood cells, platelets, with risk of neutropenic fever, hemorrhage, therefore dose reduce or hold dose for decreased neutrophil or platelet counts unless marrow infiltrated by drug-responsive tumor; N, nausea; Neu, peripheral neuropathy; Plm, pulmonary; PML, progressive multifocal leukoencephalopathy risk; PT, prothrombin time; PTT, partial thromboplastin time; ↑QT, pre- and intratreatment electrocardiogram monitoring, normalize K^+, Mg^{2+}, ionized Ca^{2+}; R, renal injury possible; TLS, risk of tumor lysis syndrome if brisk response; UGT___, metabolism by UDP-glucuronosyltransferase of the indicated genotype; V, vomiting; Ves, extravasation injury possible; VLS, vascular leak syndrome; VOD, veno-occlusive liver disease.

cell selectivity to avoid damage to normal tissue. In addition, the linker strategy should have low spontaneous release of the cytotoxic agent prior to reaching the tumor environment. Table 78-6 lists currently approved ADCs and features of importance relevant to their use for the general internist.

Three major groups of ADCs are currently available, based on the nature of the cytotoxic payload. ADCs targeting DNA structure include those based on calicheamicin, a DNA-interacting antitumor antibiotic too toxic for clinical use but, when used as an ADC, can be useful in the treatment of CD33+ acute myeloid leukemia (gemtuzumab ozogamicin) and CD22+ acute lymphocytic leukemia (inotuzumab ozogamicin). Patients must be monitored for hypersensitivity reactions and for hepatotoxicity due to veno-occlusive disease of hepatic veins, resulting from release of the calicheamicin or its metabolites in the liver. Likewise, the minor groove binder and DNA alkylator pyrrolobenzodiazepine dimer is the basis for CD19-directed loncastuximab tesirine-lpyl, whose use in refractory diffuse large B-cell lymphoma can be complicated by vascular leak syndrome, cytopenias, and cutaneous reactions.

ADCs targeting microtubules (in all cases with the potential for cytopenias and neuropathic adverse events) are of two types. Maytansine derivatives include ado-trastuzumab emtansine, which is an ADC employing the HER2/neu-directed trastuzumab and a highly toxic microtubule-targeted emtansine, which by itself is too toxic for human use; the ADC shows valuable activity in patients with breast cancer who have developed resistance to trastuzumab alone (still watch for cardiac and pulmonary side effects). Mirvetuximab soravtansine-gynx targets folate receptor α in refractory ovarian cancers (monitor for corneal side effects). The second microtubule-directed class of ADCs is derived from the microtubule poison monomethylauristatin E as cytotoxic agent. Brentuximab vedotin is an anti-CD30 ADC with activity in neoplasms such as Hodgkin's lymphoma where the tumor cells frequently express CD30. Polatuzumab vedotin analogously targets CD79a in B-cell lymphomas. Enfortumab vedotin uses an antibody to NECTIN4 to target the vedotin "warhead" to urothelial neoplasms expressing that target. Belantamab mafodotin targets BCMA (B-cell maturation) expressed myeloma but using a distinct microtubule toxin derived from auristatin F. Belantamab mafodotin can cause ocular keratopathy, which requires prospective monitoring.

Topoisomerase I–directed ADCs, including fam-trastuzumab deruxtecan-nxki and sacituzumab govitecan-bziy, are ADCs that allow specific targeting of camptothecin and SN-38, respectively to HER2-positive neoplasms (monitoring for interstitial pneumonitis and cardiac dysfunction) and, in the latter case, triple-negative breast cancers, which abundantly express the antibody target Trop2, a cell surface glycoprotein first discovered in trophoblast cells (but diarrhea should be vigorously addressed with loperamide, similar to what might be encountered with irinotecan). Both agents can cause cytopenias, nausea, and vomiting.

CANCER MOLECULAR TARGETED THERAPY

Agents in this class share the characteristic that they are directed at specific cancer cell molecular targets important in the proliferation of tumors. While these agents can ultimately lead to tumor cell death, this occurs by altered regulation of a specific biochemical pathway affecting tumor cell susceptibility to apoptosis or growth arrest (Fig. 78-3). Since many approved agents of this type are directed at specific mutations present at diagnosis or that arise during the course of a patient's clinical course, their use must be guided by refined diagnostic strategies. Initial molecular testing of a diagnostic biopsy, with repeated biopsies as warranted, should be considered. Alternatively liquid biopsies to examine plasma DNA shed from tumors or present in circulating tumor cells have entered into routine clinical use to assist in making decisions about best strategies to consider.

Hormone Receptor–Directed Therapy Steroid hormone receptor–related molecules were arguably the first "molecular target" classes of anticancer drugs. When bound to their ligands, these receptors can alter gene transcription in hormone-responsive tissues. While in some cases, such as breast cancer, demonstration of the target hormone receptor is necessary for their use, in other cases such prostate cancer (androgen receptor) and lymphoid neoplasms (glucocorticoid receptor), the relevant receptor is always present in the tumor.

Glucocorticoids are generally given in "pulsed" high doses in leukemias and lymphomas, where they induce cell death in tumor cells. Cushing's syndrome and inadvertent adrenal suppression on withdrawal from high-dose glucocorticoids can be significant complications, along with infections common in immunosuppressed patients, in particular *Pneumocystis* pneumonia, which classically appears a few days after completing a course of high-dose glucocorticoids.

Tamoxifen is a partial estrogen receptor (ER) antagonist; it antagonizes growth in wild-type (ER; also designated ESR1 to specify estrogen receptor α) breast tumors, mirroring its effect on breast tissue, but owing to agonistic activities in vascular and uterine tissue, side effects include increased risk of thromboembolic phenomena and a small increased incidence of endometrial carcinoma, which appears after chronic use (usually >5 years). Progestational agents—including medroxyprogesterone acetate, androgens including fluoxymesterone (Halotestin), and paradoxically, estrogens—have approximately the same degree of activity in primary hormonal treatment of breast cancers that have elevated expression of ER protein. Estrogen itself is not used often due to prominent cardiovascular and uterotropic activity.

Aromatase refers to a family of enzymes that catalyze the formation of estrogen in various tissues, including ovary, peripheral adipose tissue, and some tumor cells. Aromatase inhibitors are of two types: irreversible steroid analogues such as exemestane and the reversible inhibitors such as anastrozole and letrozole. Anastrozole is superior to tamoxifen in the adjuvant treatment of breast cancer in postmenopausal patients with ER-positive tumors. Letrozole treatment affords benefit following tamoxifen treatment. Adverse effects of aromatase inhibitors may include an increased risk of osteoporosis, fatigue, and altered serum lipids. ESR1 mutations of various types are detected at increased frequency after resistance to aromatase inhibitors occurs.

Fulvestrant is the prototype of a selective ER degrader. As a result of its binding, the ER is degraded and ER-dependent proliferation may diminish.

Metastatic prostate cancer is treated primarily by androgen deprivation. Orchiectomy causes responses in 80% of patients. If not accepted by the patient, testicular androgen suppression can also be induced by luteinizing hormone–releasing hormone (LHRH) agonists such as leuprolide and goserelin. These agents cause tonic stimulation of the LHRH receptor, with loss of normal pulsatile activation resulting in net decreased output of luteinizing hormone (LH) by the anterior pituitary. Therefore, as primary hormonal manipulation in prostate cancer, one can choose orchiectomy or an LHRH agonist, but not both. This pathway can also be blocked by relugolix, an oral gonadotropin-releasing hormone antagonist.

The addition of androgen receptor blockers, including flutamide or bicalutamide, is of uncertain additional benefit in extending overall response duration, although pretreatment with these agents before LHRH agonists is important to avoid a surge in testosterone after initial LH release. Enzalutamide also binds to the androgen receptor and antagonizes androgen action in a mechanistically distinct way. Somewhat analogous to inhibitors of aromatase, agents have been derived that inhibit testosterone and other androgen synthesis in the testis, adrenal gland, and prostate tissue. Abiraterone inhibits 17 α-hydroxylase/C17,20 lyase (CYP17A1) and has been shown to be active in prostate cancer patients experiencing disease progression despite androgen blockade.

Tumors that respond to a primary hormonal manipulation may frequently respond to second and third hormonal manipulations. Thus, breast tumors that had previously responded to tamoxifen have, on relapse, notable response rates to withdrawal of tamoxifen itself or to subsequent addition of an aromatase inhibitor or progestin. Likewise, initial treatment of prostate cancers with leuprolide plus flutamide may be followed after disease progression by response to withdrawal of flutamide. These responses may result from the removal of antagonists from mutant steroid hormone receptors that have come to depend on the presence of the antagonist as a growth-promoting influence.

Non-Receptor-Linked Tyrosine Kinase Antagonists Table 78-7 lists currently approved non–hormone receptor pathway-directed molecularly targeted chemotherapy agents, with features of their use of import to the generalist, particularly in recognizing potential drug-induced morbidities and interactions with other classes of drugs. The basis for discovery of drugs of this type was the prior knowledge of oncogene-directed pathways driving tumor growth (Fig. 78-3). In most cases, non-receptor tyrosine kinases ultimately activate signaling through the RAF/MEK/MAP kinase cascade, in common with the receptor-linked tyrosine kinases. Diagnostic demonstration of an active non-receptor tyrosine kinase may guide selection of an agent. A repeated preclinical and clinical observation in a variety of tumor types is that mutational activation of the tyrosine kinase target induces a state of "oncogene addiction" on the part of the tumor. This then is the basis for a "synthetic lethal" effect of the kinase inhibitor with respect to tumor viability.

In hematologic tumors, the prototypic agent of this type is imatinib, which targets the ATP binding site of the p210$^{bcr\text{-}abl}$ protein tyrosine kinase that is formed as the result of the chromosome 9;22 translocation producing the Philadelphia chromosome in chronic myeloid leukemia (CML). Subsequent studies have shown value of imatinib in gastrointestinal stromal tumor (GIST) and certain myeloproliferative disorders, driven by certain mutants of cKIT and platelet-derived growth factor receptor (PDGFR). Imatinib has lesser activity in the blast phase of CML, where the cells may have acquired additional mutations in p210$^{bcr\text{-}abl}$ itself or other genetic lesions. Imatinib's side effects are relatively limited in most patients and include hepatic dysfunction, diarrhea, and fluid retention. Rarely, patients receiving imatinib have decreased cardiac function, which may persist after discontinuation of the drug. The quality of response to imatinib enters into the decision about when to refer patients with CML for consideration of stem cell transplant approaches. Nilotinib is a tyrosine protein kinase inhibitor with activity against p210$^{bcr\text{-}abl}$ but with increased potency and perhaps better tolerance by certain patients. Dasatinib, another inhibitor of the p210$^{bcr\text{-}abl}$ oncoproteins, also has activity against certain mutant variants of p210$^{bcr\text{-}abl}$ that are refractory to imatinib and arise during therapy or are present de novo. Dasatinib also has inhibitory action against kinases belonging to the src tyrosine protein kinase family; this activity may contribute to its effects. The T315I mutant of p210$^{bcr\text{-}abl}$ is resistant to imatinib, nilotinib, bosutinib, and dasatinib; ponatinib has activity in patients with this T315Ip210$^{bcr\text{-}abl}$, but ponatinib has noteworthy associated thromboembolic toxicity. Use of this class of targeted agents is thus critically guided not only by the presence of the p210$^{bcr\text{-}abl}$ tyrosine kinase, but also by the presence of specific mutations in the ATP binding site. Asciminib is a first-in-class non-ATP site p210$^{bcr\text{-}abl}$ inhibitor that also has activity in BRC/ABL T315I mutant CML.

Janus kinases (JAK) 1 and 2 are mutated in certain myeloproliferative states; cytopenias and infrequent arrhythmias infrequently complicate the use of ruxolitinib, the prototypic JAK inhibitor. Bruton's tyrosine kinase (BTK) is an intrinsic component of B-cell antigen receptor signaling activated in many types of proliferating B cells and B-cell neoplasms. Irreversible inhibitors of BTK, including ibrutinib, acalabrutinib, and zanubrutinib, have noteworthy activity in certain lymphomas. Cytopenias and cardiac arrhythmias can occur, along with propensity to infection (indeed, the BTK was discovered as deficient in congenital hypogammaglobulinemia, presenting with repeated infections in childhood). Initial use of the BTK inhibitors requires consideration of prophylaxis against tumor lysis syndrome in case of a robust lympholytic effect of the agent. If and when resistance to irreversible BTK inhibitors develops (often associated with a C481S mutation), the noncovalent inhibitor pirtobrutinib may have activity.

Receptor-Linked Tyrosine Kinase Antagonists Mutated *EGFR* drives a significant fraction of non-small-cell lung cancers (NSCLCs). Erlotinib and gefitinib are the prototypic EGFR antagonists that, in early clinical trials, showed evidence of responses in a small

TABLE 78-7 Molecularly Targeted Agents

DRUG	TARGET/INDICATION	ADVERSE EVENTS	NOTES
Non-Receptor Tyrosine Kinase Antagonists			
Acalabrutinib	BTK/certain B-cell malignancies	My, atrial fibrillation/flutter, infection	CYP3A4; avoid PPIs and stagger doses with H2 blockers; irreversible BTK inhibitor
Asciminib	BCR/ABL/CML chronic phase after 2 TKIs or with BCR/ABL T315I	My, ↑lipase, ↑amylase, HBP, hypersensitivity, cardiac events	CYP3A4, CYP2C9, Pgp; not an ATP site inhibitor but inhibit by allosteric mechanism binding to myristoyl pocket in BCR/ABL
Bosutinib	BCR/ABL wild-type, some mutants/CML	My, ↑LFT, cardiac, R, fluid retention	CYP3A4, PPI (use short-acting antacids or H2 blocker staggered with dose)
Dasatinib	BCR/ABL/Ph+ CML, Ph+ ALL	My, pulmonary hypertension, fluid retention, ↓LVEF, ↑LFT, Der	CYP3A4, avoid PPIs and stagger doses with H2 blockers; caution with agents with ↑QT
Ibrutinib	As with acalabrutinib	My, stomatitis, HBP, arrhythmia, TLS	CY3A4, second primary malignancy
Imatinib	ABL, BCR/ABL wild-type, cKIT PDGFR/CML, some ALL, certain GISTs, eosinophil, mast cell, myelodysplastic syndromes	My, edema, fluid retention, ↓LVEF, ↑LFT rare liver failure, GI perforation, Der, TLS, ↓CrCl	DA-H, DA-R, CYP3A4, CYP2D6; hemorrhage at tumor site in GIST, cardiogenic shock with high eosinophil levels; monitor TFTs with thyroid replacement
Nilotinib	BCR/ABL wild-type/CML	CHF, ↑LFT, ↑QT, fluid retention, ↑lipase, TLS	CYP3A4, CYP2C8, CYP2C9, CYP2D6, CYP2B6; no food 2 h before, 1 h after dose; monitor electrolytes, TFTs
Pirtobrutinib	As with acalabrutinib	My, arrhythmia, infection, hemorrhage	DA-R, CYP3A4; CYP2C8, 2C19, Pgp, BCRP substrates; reversible BTK inhibitor, second primary malignancy
Ponatinib	T315I mutant BCR/ABL CML	Clotting, ↑LFT, TLS, ↓LVEF, pancreatitis, Neu, arrhythmia	CYP3A4, PRES possible
Ruxolitinib	JAK1,2/myeloproliferative disorders	My, dizziness, headache	DA-R, DA-H, CYP3A4 or with fluconazole >200 mg doses except with GVHD
Zanubrutinib	As with acalabrutinib	My, hemorrhage, cardiac arrhythmias	CYP3A4
Receptor-Linked Tyrosine Kinase Antagonists			
Afatinib	Nonresistant ATP site mutated *EGFR*/NSCLC	D, Der, Cor, ILD; ↑LFT	DA-R, DA-Pgp inhibitors, no food 2 h before, 1 h after dose
Alectinib	*ALK* rearranged/NSCLC	↑LFT, ILD, R, bradycardia, ↑CPK	Administer with food; can have muscle pain, tenderness, weakness
Avapritinib	*PDGFR* mutants/GIST, mastocytosis	N, edema, CNS, sleep, mood change, hallucinations	CYP3A4, monitor for intracranial hemorrhage
Capmatinib	*MET* + exon 14 skip/NSCLC	ILD, ↑LFT, photosensitivity, pancreatitis, hypersensitivity	
Ceritinib	*ALK* rearranged/NSCLC	GI, ↑LFT, ↑glucose, ↑QT, bradycardia, pancreatitis	CYP3A4, CYP2C9; if ILD, permanently discontinue
Crizotinib	*ALK* rearranged/NSCLC; inflammatory myofibroblastic tumor, anaplastic large-cell lymphoma	ILD, ↑LFT, ↑QT, bradycardia, ↓vision	CY3A4
Dacomitinib	*EGFR* exon 19 deletion or exon 21 L858R/NSCLC	D, Der: hold and/or dose reduce; ILD (permanently discontinue)	CYP2D6; avoid PPIs, use local antacids or give 6 h before or 10 h after H2 blockers
Erdafitinib	FGFR2/3 altered/urothelial cancer after Pt regimen	Der, D, stomatitis, retina (ophthalmologic exam before/during)	CYP2C9, CYP3A4, also organic cation transporter 2 and Pgp substrates, give 6 h before/after Pgp substrates; ↑PO_4 due to effect on FGFR2/3/Klotho
Erlotinib	As with afatinib; also wild-type EGFR/NSCLC 2nd line or pancreatic 1st line with gemcitabine	Der, D, R, ILD, ↑LFT, rare microangiopathic anemia	CYP3A4, give 1 h before/2 h after meals, avoid PPIs, stagger with H2 blockers; alter warfarin effect
Fruquintinib	Target VEGFR1,2,3/CRC	HBP, hemorrhage, arterial thromboses, infection, GI perforation, ↑LFT, proteinuria, ↓wound healing, allergic bronchospasm	CYP3A4; unusual, PPED, PRES
Futibatinib	*FGFR2* gene fusions/cholangiocarcinoma	Retina (ophthalmologic exam before/during)	CYP3A4, Pgp substrates; ↑PO_4 due to effect on FGFR2/3/Klotho, tissue and vascular calcification
Gefitinib	As with afatinib	Der, D, ILD, ocular keratitis, GI perforation	CYP3A4; avoid with PPIs; monitor warfarin effect
Gilteritinib	*FLT3* mutated/AML	↑LFT, ↑QT, myalgia/arthralgia, N, V, GI, Der, edema, dyspnea	Monitor for PRES, pancreatitis, ↑Cr, eye disorders
Infigratinib	*FGFR2* variants/cholangiocarcinoma	Retina (ophthalmologic exam before/during)	CYP3A4, avoid gastric acid reduction or stagger with H2 blockers or locally acting agents, ↑PO_4 due to effect on FGFR2/3/Klotho, soft tissue including myocardial mineralization
Lapatinib	HER2/HER2+ breast cancer	↓LVEF, ↑LFT, N, V, D, PPED	CYP3A4, CYP2C8, Pgp substrate; interaction, ILD, ↑QT
Larotrectinib	*NTRK* gene fusion without a resistance mutation/any solid tumor	CNS with potential cognitive impairment; ↑LFT	CYP3A4
Lorlatinib	*ALK* rearranged/NSCLC	Hyperlipidemia, AV block, CNS including seizures, mental status changes	CYP3A4 with severe ↑LFT with CYP3A interactors; monitor for ILD and discontinue if occur
Neratinib	As with lapatinib	N, V, D, abdominal pain, ↑LFT	CYP3A4, aggressive D prophylaxis with loperamide; avoid PPIs and stagger H2 blocker doses

(*Continued*)

TABLE 78-7 Molecularly Targeted Agents (Continued)

DRUG	TARGET/INDICATION	ADVERSE EVENTS	NOTES
Osimertinib	*EGFR* exon 19 altered or exon 21L858R or T790M mutations/NSCLC	ILD, ↑QT, ↓LVEF, Cor	CYP3A4
Pemigatinib	*FGFR2* fusion or other rearrangement/ cholangiocarcinoma *FGFR1* rearranged myeloid/lymphoid neoplasms	N, D, stomatitis, ↑PO_4 due to effect on FGFR2/3/Klotho	CYP3A4, retinal detachment: ophthalmologic exam with ocular tomography before and every 2–3 months during treatment
Pralsetinib	*RET* mutant or fusion/NSCLC, thyroid carcinomas	ILD, HBP, hemorrhage, ↑LFT, TLS, ↓wound healing	CYP3A4, Pgp
Quizartinib	*FLT3* mutated/AML	↑QT with risk of cardiac arrest	CYP3A4
Repotrectinib	Target *ROS1* rearranged/NSCLC	CNS, ILD, ↑LFT, myalgia, ↑CPK	CYP3A4, Pgp, OCs; also target TRKA,B,C
Selpercatinib	*RET* mutant or fusion/NSCLC, thyroid carcinomas	↑LFT, ↑QT, HBP, bleeding, ↓wound healing; hold 1 week prior and 2 weeks after surgery	Avoid with antacids, but take if not avoidable with food if PPI or stagger with other antacid, CYP3A, CYP2C8
Tepotinib	*MET* with exon 14 skipping mutations/ NSCLC	ILD, ↑LFT	CYP3A4, Pgp; DA-H
Tucatinib	As with lapatinib	↑LFT, D	CYP3A4, CYP2C8
KRAS-G12C/RAF/MEK Antagonists			
Adagrasib	*KRAS* G12C mutation/NSCLC	N, V, D, ↑LFT, ↑QT, ILD	CYP3A4, CYP2C9, CYP2D6, agents that ↑QT
Binimetinib	Targets MEK/*BRAF* V600E or V600K mutated melanoma combined with encorafenib	↓LVEF, venous thrombosis; ocular, ILD, ↑LFT, rhabdomyolysis	DA-H
Cobimetinib	Targets MEK/*BRAF* V600E or V600K melanoma combined with vemurafenib; alone in *BRAF* V600E or V600K histiocytic neoplasms	Hemorrhage, retinal ↓LVEF, Der, photosensitivity, rhabdomyolysis, ↑LFT	CYP3A4, new malignancies, cutaneous and noncutaneous
Dabrafenib	Targets *BRAF* in *BRAF* V600E or V600K mutated tumors (not CRC) with trametinib	Hemorrhage, uveitis ↓LVEF, Der, ↑glucose, hemolysis if G6PD deficient, pyrexia	CYP3A4, CYP2C8, CYP2C9, CYP2C19, CYP2B6, new malignancies, cutaneous and noncutaneous
Encorafenib	Targets *BRAF* in *BRAF* V600E or V600K mutated tumors combined with binimetinib (melanoma) or CRC with cetuximab	Hemorrhage, uveitis, ↑QT	CYP3A4, organic anion transporter 1B1, BCRP interactions, new malignancies, cutaneous and noncutaneous
Sotorasib	*KRAS* G12C mutation/NSCLC	↑LFT, ILD, N, V, D	Avoid with PPIs and H2 receptor antagonists; if necessary, administer 4 h before or 10 h after a local antacid, avoid CYP3A4, Pgp substrates
Trametinib	Targets MEK/*BRAF* V600E or V600K mutated tumors (not colorectal) combined with dabrafenib	Hemorrhage, venous thromboembolism, ↓LVEF, ILD, pyrexia, Der, ↑glucose	Colitis with GI perforation, ocular including retinal vein occlusion, new malignancies, cutaneous and noncutaneous
Vemurafenib	Targets BRAF in *BRAF* V600E or V600K mutated melanoma and Erdheim-Chester disease	Der including Stevens-Johnson, anaphylaxis and allergic hypersensitivity, ↑QT, ↑LFT, photosensitivity, radiation recall	Dupuytren's contracture and plantar fascial fibromatosis has occurred; usually combined with cobimetinib in melanoma; CYP3A4, CYP1A2, CYP2D6, new primary cutaneous and noncutaneous neoplasms
Multikinase Antagonists			
Axitinib	VEGFR, PDGFR, KIT/RCC	HBP, hemorrhage, thrombotic events; D, other GI including GI perforation, PPED, hypothyroidism, PRES, proteinuria, ↑LFT	DA-H, CYP3A4/5
Brigatinib	ALK, EGFR/NSCLC	ILD, bradycardia, HBP, visual disturbances, ↑glucose, ↑CPK	CYP3A4, decreased hormonal contraceptive effectiveness
Cabozantinib	VEGFR2, MET, AXL, RET/RCC, HCC, certain thyroid carcinoma	HBP, hemorrhage, thrombotic events, D, other GI including fistula, perforation, wound healing, PRES, PPED, proteinuria, ONJ	CYP3A4
Capivasertib	Target *PIK3CA/AKT1/PTEN*- axis; BC-HR(+) with mutation in pathway	Hypersensitivity reactions, Der, ↑glucose, pneumonitis, ILD, D	CYP3A4, CYP2C9, BCRP substrates
Entrectinib	*NRTK* gene fusion/any solid tumor; *ROS1* gene alteration/NSCLC	ILD, ↑LFT, photosensitivity, ↓LVEF, CNS effect, skeletal fractures; hyperuricemia, ↑QT,	CYP3A4, ophthalmologic exam if vision change
Fedratinib	JAK2, FLT3, RET/myeloproliferative diseases	My, N, V, D, ↑LFT pancreatitis, encephalopathy: check thiamine levels prior, replete if deficient	CYP3A4, CYP2C19
Lenvatinib	VEGFR1/2/3, FGFR1/2/3/4, PDGFRα, KIT, RET/thyroid, RC, HCC, endometrial cancer	HBP, ↓LVEF, bleeding, arterial/venous clots, RF, ONJ, proteinuria, ↑LFT, GI including D, fistula/perforation, ↓wound healing, ↑QT, ↓Ca^{2+}, PRES, ↓TFT	DA-H, DA-R
Midostaurin	FLT3 mutated/AML newly diagnosed, mast cell neoplasms	ILD; N, D	CYP3A4; many other protein kinase targets in addition to FLT3

(*Continued*)

TABLE 78-7 Molecularly Targeted Agents (Continued)

DRUG	TARGET/INDICATION	ADVERSE EVENTS	NOTES
Pazopanib	VEGFR 1/2/3, KIT, PDGFR, and FGFR/ RCC soft tissue sarcoma (not GIST or adipocytic)	D, HBP; arterial and venous thrombosis ± embolism, ↑QT, hemorrhage, ↑LFT potentially severe/fatal; GI perforation or fistula; proteinuria, ↓TFT, ↓LVEF, PRES, ILD, thrombotic microangiopathy	CYP3A4, CYP22D6, CYP2C8 interaction; use with simvastatin increases the risk of ALT elevations and should be undertaken with caution; avoid with PPIs or H2 blocker, stagger with other antacid doses
Regorafenib	VEGFR1/2/3, KIT, RET, PDGFR-α/β, FGFR1/2, TIE2, DDR2, Trk2A, Eph2A, RAF1, BRAF, BRAF V600E, SAPK2, PTK5, and ABL/CRC, GIST	↑LFT potentially severe/fatal; hemorrhage, thromboses, PPED, Der, HBP, ↓LVEF, ↑QT, GI perforation	CYP3A4, impaired TSH suppression in thyroid cancer
Sorafenib	VEGFR2, VEGFR3, PDGFRβ, Flt3, KIT, RAF1, BRAF/RCC, HCC, differentiated thyroid carcinoma	D, ↑LFT potentially severe/fatal; hemorrhage, PPED, Der, HBP, ↓LVEF, ↑QT, GI perforation	As with regorafenib
Sunitinib	VEGFRs; PDGFR, RET, KIT; other protein kinases/RCC, pancreatic neuroendocrine, GIST	HBP, ↑LFT potentially severe/fatal; hemorrhage, GI perforation, CHF, altered TFTs, ONJ, ↓wound healing, proteinuria, R	CYP3A4, rare ↑QT, rare TLS in RCC and GIST with high tumor burden
Vandetanib	VEGFR, RET, EGFR/medullary thyroid cancer	D, Der, HBP, ↑QT, thromboses, ↓LVEF, fistulas, ILD, ONJ, proteinuria, PRES	CYP3A4
Cyclin-Dependent Kinase (CDK) Inhibitors			
Abemaciclib	CDK4/6/HR+ BC	D, My, ↑LFT, venous thromboembolism, ILD	CYP3A4, avoid concomitant use of ketoconazole
Palbociclib	CDK4/6/HR+ BC	D, My, stomatitis, ILD, D	CYP3A4
Ribociclib	CDK4/6/HR+ BC	Der, ILD, ↑QT, ↑LFT, My	CYP3A4, agents known to ↑QT
Protein Homeostasis Modulators			
Bortezomib	Proteasome inhibitor/multiple myeloma, mantle cell lymphoma	Neu, N, V, D, C, ↓BP, My, ILD, ↑LFT worsening cardiac disease	Rare TLS, PRES
Carfilzomib	Proteasome inhibitor/multiple myeloma	↓LVEF, myocardial ischemia, R, TLS, pulmonary including ARDS, ILD, HBP, IR*, thrombosis, hemorrhage, PRES, thrombocytopenia, thrombotic microangiopathy, ↑LFT potentially severe	Administer after a hemodialysis procedure
Ixazomib	Proteasome inhibitor/multiple myeloma relapsed, not as maintenance	Thrombocytopenia, N, V, D, C, Neu, edema, Der, thrombotic microangiopathy, ↑LFT	DA-R, DA-H, CYP3A4
Nuclear Export Inhibitor			
Selinexor	Targets exportin 1/multiple myeloma, certain DLBCLs	My, N, V, D, anorexia, ↓Na^+, CNS, cataract development or progression	Advise against driving or dangerous equipment operation if CNS altered
Chromatin-Modifying Epigenetic Modulators			
DNA Hypomethylating Agents			
Azacytidine	Target DNA methyltransferase/ myelodysplastic syndromes, AML	My, ↑LFT, TL, CNS	Monitor if renal impairment, more N, V, D if SC administration
Decitabine	As with azacytidine	My	Combined with cedazuridine (a cytidine deaminase inhibitor) in an oral regimen
Histone Deacetylase Inhibitors			
Belinostat	Peripheral T-cell lymphoma, relapsed or refractory	My, ↑LFT, TLS, N, V	
Panobinostat	Multiple myeloma, relapsed or refractory	My, hemorrhage, ↑LFT, ↓Na^+, ↓K^+, ↓PO_4, ↑Cr	CYP3A4, CYP2D6, avoid agents that ↑QT
Romidepsin	CTCL after one systemic therapy	My, ↑QT	CYP3A4, alter warfarin effect, may ↓ effectiveness of oral contraceptives
Vorinostat	CTCL after two systemic therapies	My, N, V, D, venous thrombosis, ↑glucose	Monitor with mild or moderate liver disease
Histone Methyltransferase Inhibitors			
Tazemetostat	Target *EZH2* mutant or nonmutant/ epithelioid sarcoma, certain follicular lymphomas	N, V, C, abdominal pain	CYP3A4, increased risk of secondary malignancies (MDS, AML, lymphomas)
Transcription Factor Modulation			
Arsenic trioxide	Target PML-RARα and redox homeostasis/t(15;17) acute promyelocytic leukemia	↑QT, hypersensitivity	APL differentiation syndrome: with pulmonary dysfunction/infiltrate, pleural/pericardial effusion, fever, treat with dexamethasone ± hydroxyurea
Belzutifan	Target HIF2α/von Hippel-Lindau (VHL) disease–associated RCC, CNS hemangioblastomas, GI neuroendocrine tumors, non-VHL RCC after PD-1, VEGFR inhibitor	Anemia, hypoxia, N, ↑Cr, ↑glucose	CYP2C19, UGTB17
Glasdegib	Targets smoothened receptor in hedgehog pathway/AML	↑QT	CYP3A4, transmission to potentially pregnant partner through semen or to blood product recipient

(*Continued*)

TABLE 78-7 Molecularly Targeted Agents (Continued)

DRUG	TARGET/INDICATION	ADVERSE EVENTS	NOTES
Nirogacestat	Targets γ-secretase to inhibit Notch signaling/desmoid tumor	D (can be severe), ovarian dysfunction, ↑LFTs, PO_4 and K^+ abnormalities	CYP3A4; avoid PPIs and stagger doses with H2 blockers; nonmelanoma skin cancers; Der: monitor before and during treatment
Sonidegib	Targets smoothened receptor in hedgehog pathway/basal cell carcinoma locally advanced or metastatic	Musculoskeletal adverse events with ↑CPK, potential R, N, V, D	As with glasdegib
Tretinoin	Target PML-RARα/t(15;17) acute promyelocytic leukemia	Der including cheilitis, skin dryness; ↑ intracranial pressure; ↑lipids, ↑LFT, usually resolve	As with arsenic trioxide; also headache, visual changes may indicate ↑intracranial pressure; check for papilledema
Vismodegib	As with sonidegib	Musculoskeletal adverse events, N, V, D, C	Transmission to potentially pregnant partner through semen or to blood product recipient
Apoptosis Modulation			
Venetoclax	Targets BCL2/CLL, SLL; AML + azacytidine, decitabine, or low-dose cytarabine	My, D, TLS	CYP3A4, stagger with Pgp substrates, no live attenuated vaccines prior to, during, or after venetoclax treatment
Metabolism Modulation: mTOR Inhibitors/PI Kinase/IDH Inhibitors			
Alpelisib	*PIK3CA* mutated/HR+HER2– BC	Der, hypersensitivity, ↑glucose, ILD, D	CYP3A, CYP2C9, BCRP substrates
Copanlisib	PI3Kα,δ/FL	My, HBP, noninfectious pneumonitis, ↑glucose, Der	CYP3A4
Duvelisib	PI3Kδ,γ/CLL, SLL, FL	My, infection, D, colitis, Der, pneumonitis, ↑LFTs	CYP3A4
Enasidenib	IDH2 mutated/AML	N, V, D, ↑LFTs	AML differentiation syndrome with pulmonary dysfunction/infiltrate, pleural/pericardial effusion, fever treat with dexamethasone ± hydroxyurea
Everolimus	mTOR/RCC, tuberous sclerosis–associated neoplasms, HR+ BC, neuroendocrine, pancreatic, lung, GI NOT functional carcinoid	My, noninfectious pneumonitis, infections, hypersensitivity reactions, R, impaired wound healing, ↑glucose, ↑lipids, stomatitis	CYP3A4, Pgp substrates, angioedema with concomitant ACE inhibitors, consider alcohol-free mouthwash when starting treatment; risk of reduced efficacy of vaccination
Ivosidenib	IDH1 mutated/AML, cholangiocarcinoma, MDS	↑QT, Guillain-Barré syndrome	CYP3A4, QT-prolonging agents; AML differentiation syndrome requiring corticosteroid treatment
Idelalisib	PI3Kδ/non-first-line CLL, SLL, FL	Fatal or serious ↑LFTs, D, colitis with GI perforation, pneumonitis, infection, Der, hypersensitivity	CYP3A4, not with bendamustine or rituximab
Olutasidenib	IDH1/AML	↑LFTs	CYP3A4, differentiation syndrome requiring corticosteroid ± hydroxyurea treatment
Sirolimus protein bound particles	mTOR/perivascular epithelioid cell tumor (PEComa)	My, stomatitis, infection, ↓K^+, IR, ↑glucose, ILD, hemorrhage, male infertility	CYP3A4, avoid live vaccines
Temsirolimus	mTOR/RCC	Hypersensitivity, ↑LFTs, infection, ILD, stomatitis, thrombocytopenia, N, ↑glucose, ↑lipids, ↓wound healing, GI perforation, R ± proteinuria	DA-H, CYP3A4
Poly-ADP Ribose Polymerase (PARP) Inhibitors			
Niraparib	Ovarian, fallopian tube, or primary peritoneal cancer with good response to Pt	My, N, V, D, HBP, PRES	MDS
Olaparib	Ovarian: as with niraparib also with various *BRCA* or HRR mutations; BC mutant *BRCA,* HER2–; pancreatic mutant *BRCA* with good response to Pt; prostate with *BRCA* or HRR mutations	My, N, stomatitis, DVT ± PE, rare ILD	MDS, CY3A4
Rucaparib	As with niraparib; prostate *BRCA* mutated after hormone and after taxane	My, stomatitis, N, V, D, ↑LFTs	MDS
Talazoparib	BC *BRCA* mutated HER2–; prostate cancer, castrate-resistant HRR mutation(+), with enzalutamide	My, N, V, D, ↑LFTs	MDS
Miscellaneous			
^{177}Lu-dotatate	Target somatostatin receptor (SSR)/ gastroenterohepatic neuroendocrine tumors (SSR)+	My, R, ↑LFTs, IR,	Neuroendocrine hormonal crisis including flushing, diarrhea, hypotension, bronchoconstriction; secondary MDS, risks from radiation exposure
^{177}Lu-vipivotide tetraxetan	Target prostate-specific membrane antigen (PSMA)/refractory prostate cancer	My, R	Temporary or permanent infertility, risks from radiation exposure

(*Continued*)

TABLE 78-7 Molecularly Targeted Agents (Continued)

DRUG	TARGET/INDICATION	ADVERSE EVENTS	NOTES
Tagraxofusp-erzs	Targets CD123 (IL-3 receptor)/blastic plasmacytoid dendritic cell neoplasm	IR*, ↑LFTs, VLS	Delivers a fragment of diphtheria toxin
Ziv-aflibercept	Targets VEGF by solubilized VEGFR/CRC with chemotherapy after an oxaliplatin regimen	↓Wound healing with fistula, GI perforation, hemorrhage, HBP, DVT, arterial thromboembolism, proteinuria, PRES	Administer over 1 h, not as push or bolus

Note: Data abstracted in part from publicly available U.S. Food and Drug Administration label. All agents in this class have the potential for prominent embryofetal toxicity; use without contraception by female patients of childbearing potential is not recommended. Effective contraception for female partners who are of childbearing potential of patients undergoing treatment should also be considered; use during lactation is also not recommended. Gene products are in capital letters; genes are italicized in capitals. Indications and events of prominent general medical importance include the following: ACE, angiotensin-converting enzyme; ALK, anaplastic lymphoma kinase; ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; APL, acute promyelocytic leukemia; ARDS, acute respiratory distress syndrome; AV, atrioventricular; BC, breast cancer; BCRP, breast cancer resistance protein; BP, blood pressure; *BRCA*, breast cancer gene; BTK, Bruton's tyrosine kinase; C, constipation; CHF, congestive heart failure; CLL, chronic lymphocytic leukemia; CML, chronic myelogenous leukemia; CNS, can include altered sensorium, cortical and cerebellar signs, dysarthria, altered seizure threshold; Cor, corneal keratopathy, consider pretreatment ophthalmologic exam with prophylactic lubricating eyedrops ± ocular steroids; CRC, colorectal cancer; CrCl, creatinine clearance; CTCL, cutaneous T-cell lymphoma; CYP___, interaction with drugs metabolized by the indicated cytochrome(s) P450; D, diarrhea; DA-H, dose adjust for hepatic dysfunction; DA-R, dose adjust for renal dysfunction; Der, dermatologic toxicity; DLBCL, diffuse large B-cell lymphoma; DVT, deep vein thrombosis; EGFR, epidermal growth factor receptor; FGFR, fibroblast growth factor receptor; FL, follicular lymphoma; GIST, gastrointestinal stromal tumor; GVHD, graft-versus-host disease; H2, histamine receptor antagonist; HBP, high blood pressure; HCC, hepatocellular carcinoma; HER2, human epidermal growth factor receptor 2; HRR, homologous recombination repair; HR, hormone receptor; HUS, hemolytic-uremic syndrome; IDH, isocitrate dehydrogenase; ILD, interstitial lung disease; IR, infusion reaction; IR*, infusion reaction prophylaxis specifically recommended; LFT, liver function tests; LVEF, left ventricular ejection fraction monitor echocardiogram; MDS, myelodysplastic syndrome; MET, hepatocyte growth factor; My, anemia, decrease in white blood cells, platelets, with risk of neutropenic fever, hemorrhage, therefore dose reduce or hold dose for decreased neutrophil or platelet counts unless marrow infiltrated by drug-responsive tumor; N, nausea; Neu, peripheral neuropathy; NSCLC, non-small-cell lung cancer; OC, oral contraceptive; ONJ, osteonecrosis of the jaw; PDGFR, platelet-derived growth factor receptor; PE, pulmonary embolism; Pgp, P-glycoprotein drug resistance protein; Ph, Philadelphia chromosome; Plm, pulmonary; PML, progressive multifocal leukoencephalopathy risk; PPED, palmar-plantar erythrodysesthesia (i.e., hand-foot syndrome); PPI, proton pump inhibitor; PRES, posterior reversible leukoencephalopathy syndrome; Pt, platinating agent; PT, prothrombin time; PTT, partial thromboplastin time; ↑QT, increase of QT interval needs pre- and intratreatment electrocardiogram monitoring and normalizing K^+, Mg^{2+}, ionized Ca^{2+}; R, renal injury possible; RCC, renal cell carcinoma; ROS1, receptor tyrosine kinase 1; SLL, small lymphocytic lymphoma; TFT, thyroid function test; TKI, tyrosine kinase inhibitor; TLS, risk of tumor lysis syndrome if brisk response; TOPO___, agent targets the indicated topoisomerase; TSH, thyroid-stimulating hormone; UGT___, drug metabolism by UDP-glucuronosyltransferase of the indicated genotype; V, vomiting; VEGFR, vascular endothelial growth factor receptor; Ves, extravasation injury possible; VLS, vascular leak syndrome; VOD, veno-occlusive liver disease.

fraction of patients with NSCLC. Subsequent studies by clinical oncologists in an effort to understand the basis of these excellent responses found that the probability of response to the agents was markedly increased in patients with an activating *EGFR* mutation, and current practice now routinely profiles patients with NSCLC for the presence of sensitizing mutations of *EGFR*. Side effects were generally acceptable, consisting mostly of acneiform rash (treated with glucocorticoid creams and clindamycin gel) and diarrhea. Patients with activating mutations who initially responded to gefitinib or erlotinib but who then had progression of the disease then acquired additional mutations in the enzyme, analogous to the mutational variants responsible for imatinib resistance in CML. Subsequent generations of EGFR antagonists have activity against more uncommon mutants (osimertinib) or a biochemically irreversible mechanism (dacomitinib).

Mutated anaplastic lymphoma kinase (*ALK*) and activated *RET* oncogene likewise drive distinct fractions of NSCLCs. Alectinib, ceritinib, crizotinib, and lorlatinib target ALK, but have prominent adverse cardiac, metabolic, and, in the case of lorlatinib, pulmonary events. Selpercatinib targets RET in NSCLCs (and thyroid cancers) but also with the chance of cardiac and liver toxicity. Repotrectinib is active in *ROS1* mutant NSCLC.

HER2-driven breast cancers may be usefully treated with lapatinib; diarrhea and cardiac dysfunction can occur. Neratinib or tucatinib may also be useful in HER2-positive breast cancers after trastuzumab has ceased to be of value; diarrhea and liver toxicity also require monitoring and management.

Alteration of fibroblast growth factor (FGF) signaling can contribute to the growth of urothelial carcinomas and cholangiocarcinomas. Erdafitinib and pemigatinib, respectively, may be of utility with careful attention to ocular toxicity and hyperphosphatemia; the latter is an "on-target" toxicity of disrupting FGF receptor signaling in the kidney. Likewise, gilteritinib is active against the FMS-like tyrosine kinase-3 (FLT3) mutated in a fraction of poor-prognosis (treated by conventional chemotherapy) acute myeloid leukemias (AMLs). Cardiac, hepatic, gastrointestinal, and neurologic adverse events can occur, along with "differentiation" of the AML cells with cytokine elaboration and pulmonary side effects, requiring management with steroids and potentially hydroxyurea.

The neurotropic tyrosine kinase receptor (NRTK) undergoes translocation with fusion to a variety of different partner genes to produce a family of chimeric proteins in a small fraction of a variety of solid tumors. Larotrectinib and entrectinib may be quite useful in managing these tumors; indeed, these agents are exemplary of "histology agnostic" agents, where the utility of the drug is not tied to a particular histologic diagnosis, but to the possession of a specific *NRTK* gene alteration. Neurotoxicity, a long half-life of the agents, and hepatotoxic adverse events are of concern. Neuroregulin 1 (produced by the *NRG1* gene) undergoes gene fusions to produce activators of HER family heterodimer signaling, including HER3 and HER4; afatinib covalently binds to all HER family members and has activity in patients with *NRG1* gene fusions.

RAF/MEK Pathway Antagonists The *RAS* proto-oncogene family members (including *HRAS, KRAS,* and *NRAS*) act as "switches" to bind GTP in response to activation of receptor tyrosine kinases; RAS-GTP activates the *RAF* proto-oncogene–derived serine-threonine kinase. *RAS* mutations of various types result in persistent activation of RAS isoforms, resulting in hyperactivation of RAF and "downstream" kinases, including MEK and MAPK. Sotorasib is a first-in-class inhibitor of *KRAS* G12C signaling that in early clinical reports has evidence of effecting stable disease in patients with a variety of neoplasm histologies bearing that mutation, with fewer actual responses. Adagrasib also targets *KRAS-G12C* with a distinct pharmaceutical profile. Neither sotorasib nor adagrasib has activity against wild-type or other RAS mutants, but their discovery marked a milestone, encouraging continuing efforts to produce RAS-directed therapeutics.

The *BRAF* V600E mutation drives a substantial fraction of melanomas and certain NSCLCs and has been detected in certain thyroid tumors, colorectal tumors, hairy cell leukemias, and unusual gliomas. BRAF inhibitors such as dabrafenib, vemurafenib, and encorafenib have activity as single agents in many such tumors but are usually most active when co-administered as "doublets" with the MEK inhibitors trametinib, cobimetinib, and binimetinib, respectively, to promote "shutdown" of RAF/MEK signaling at more than pathway member. Cutaneous adverse events including generally indolent cutaneous second neoplasms, and thromboembolic, cardiac, and ocular toxicity can occur.

Multikinase Inhibitors Agents in this class also target specific macromolecules promoting the viability of tumor cells. They are "small-molecule" ATP site-directed antagonists that inhibit more than one protein kinase and may have value in the treatment of several solid tumors. Drugs of this type with prominent activity against the VEGFR tyrosine kinase have activity in renal cell carcinoma. Sorafenib is a VEGFR antagonist also with activity against the RAF serine-threonine protein kinase, and regorafenib is a closely related drug with value in relapsed advanced colon cancer. Pazopanib also prominently targets VEGFR and has activity in renal carcinoma and soft tissue sarcomas. Sunitinib has anti-VEGFR, anti-PDGFR, and anti-KIT activity. It causes prominent responses and stabilization of disease in renal cell cancers and GISTs. Side effects for agents with anti-VEGFR activity, similar to those of the anti-VEGF antibody bevacizumab, prominently include hypertension, proteinuria, and, more rarely, bleeding and clotting disorders, perforation of scarred gastrointestinal lesions, and posterior leukoencephalopathy, probably reflecting CNS vascular damage. Also encountered are fatigue, diarrhea, and hand-foot syndrome, with erythema and desquamation of the distal extremities, in some cases requiring dose modification, particularly with sorafenib.

Other agents in this class include agents such as brigatinib (clinical activity in ALK-dependent NSCLC, but also with anti-EGFR action), entrectinib (clinical activity in NTRK fusion protein diseases, but also in *ROS*-mutated NSCLC), and fedratinib (clinical activity in myeloproliferative neoplasms, but with RET activity in addition to JAK2 and FLT3 antagonism). Agents with anti-RET activity are useful in certain differentiated thyroid cancers and RET-fusion NSCLC.

Cyclin-Dependent Kinase Inhibitors Cyclin-dependent kinases (CDKs) are activated as the result of oncogene pathway activity, and CDK4 and CDK6 phosphorylate the retinoblastoma (RB) tumor-suppressor gene to allow entry into S-phase. Palbociclib, abemaciclib, and ribociclib, selective inhibitors of CDK4 and CDK6, have noteworthy activity in advanced breast cancers also expressing the ER, usually in conjunction with continued efforts to suppress ER signaling, and frequently in conjunction with mammalian target of rapamycin (mTOR) inhibitors. Further clinical investigations in other RB intact tumors may broaden their role.

Protein Homeostasis Modulators The proteasome is a macromolecular complex that degrades misfolded proteins tagged for removal by ubiquitin ligases. Proteasome inhibitors were originally designed as potential anti-inflammatory agents owing to proteasome activity to produce inflammatory cytokines but had unexpected antiproliferative activity in a variety of cell types. Proteasome inhibitors have clinical utility in myeloma and lymphoma, where unbalanced synthesis of immunoglobulin components can accumulate after proteasome inhibitor treatment and induce apoptosis or starve cells for amino acids, inducing autophagy. Boronic acid proteasome inhibitors, including bortezomib and ixazomib, cause thrombocytopenia, gastrointestinal dysfunction, and neuropathy. Carfilzomib is a distinct chemotype with attenuated neuropathy but increased incidence of infusion reactions and cytokine release, with attendant risk of cardiopulmonary adverse events.

Exportin 1 is a nuclear membrane transport protein that is responsible for normal exit and entry of a variety of nuclear proteins. Selinexor is an inhibitor of exportin action, resulting in abnormal nuclear accumulation of, for example, tumor-suppressor gene products or needed export of other products, such as oncogene products. Useful clinical activity has been seen in myeloma and diffuse large B-cell lymphomas including those arising from previously treated indolent lymphomas. Cytopenias, gastrointestinal distress, and hyponatremia are features of its clinical use.

Chromatin-Modifying Agents Gene function is altered not only by mutation of DNA structure but also by "epigenetic" mechanisms that alter the capacity of DNA to be transcribed or interact with regulatory proteins in the nucleus including transcription factors. Initial epigenetic approaches to modulate gene expression extended from the observation that 5′azacytidine and decitabine are misincorporated into DNA and then scavenge DNA methyl transferase to disable DNA methylation of cytosine near gene promoter regions and thus alter their transcription. This "hypomethylation" causes differentiation of AML cells with notably less host toxicity than higher concentrations of these agents or indeed cytosine arabinoside, which does not have prominent hypomethylation activity.

Histone deacetylase inhibitors alter the histone protein "packing" density of chromatin and induce global changes in expression of cell cycle regulatory proteins. Vorinostat, belinostat, and romidepsin are useful in cutaneous and peripheral T-cell lymphomas; panobinostat has activity in multiple myeloma. The agents are generally well tolerated but with the potential for cytopenias. The histone methyltransferase inhibitor tazemetostat is a first-in-class inhibitor of histone methyltransferase with unique activity in epithelioid sarcoma owing to its modulation of transcriptional mechanisms unique to that tumor and, recently, in certain follicular lymphomas.

Transcription Factor Modulation Distinct from hormone receptor agonists and antagonists, which as previously described modulate transcription factor activity by affecting the binding of endogenously produced ligands, such as androgens and estrogens, tretinoin (all-*trans*-retinoic acid) binds to the fusion protein PML-RARα produced as a result of the t(15,17) chromosomal translocation that underlies the pathogenesis of most cases of acute promyelocytic leukemia (APL). PML-RARα functions as transcriptional co-repressor of normal granulocyte maturation; inhibiting the repressor leads to clinical value in the curative treatment of APL. Adverse events include typical symptoms of hypervitaminosis A, including skin dryness, cheilitis, increased intracranial pressure, and the development of a "leukemia differentiation syndrome" marked by fever and lung and other organ infiltration by newly differentiated leukocytes, which usually responds to glucocorticoids and hydroxyurea but can be life-threatening. Significantly, the coagulopathy of APL is attenuated. Arsenic trioxide nonspecifically affects PML-RARα and other targets by a redox mechanism, contributing to APL treatment with tretinoin regimens, but with a narrow therapeutic index particularly related to increase of the QT interval, which must be monitored during treatment, while also being carefully attentive to K^+, Mg^{2+}, ionized Ca^{2+}, and other drugs that can alter QT.

Belzutifan is a novel agent that antagonizes directly the action of hypoxia-inducible factor 2 (HIF2) α, activated in von Hippel-Lindau (VHL) disease–related neoplasms and many sporadic clear cell renal carcinomas through loss of the VHL ubiquitin ligase tumor-suppressor gene. It is safe, but anemia and hypoxemia may occur.

The sonic hedgehog factor pathway is regulated by the WNT ligands acting on smoothened receptors, named in reference to *Drosophila* mutants in which the pathway was originally revealed. In humans, the pathway acts during embryonic and fetal life and in certain neoplasms, and mediates effects on transcription through Gli proteins. Hedgehog pathway inhibitors sonidegib and vismodegib are useful in nonsurgically treatable cutaneous basal cell carcinomas, and glasdegib is active in certain AMLs where the pathway is active (see Table 78-7). The notch pathway was likewise first noted in *Drosophila* and acts in metazoans to signal cell position and motion. Notch receptors respond to their ligands by undergoing endocytosis and processing by a γ-secretase, followed by translocation to the nucleus, and function as specific gene transcription factors. The notch pathway has long been known to be active in hematopoietic tumors, but observation has shown confirmed activity in desmoid tumors, inflammatory mesenchymal tumors with poor evidence of control by conventional cytotoxic agents, by nirogacestat, the first γ-secretase inhibitor to enter clinical practice.

Cancer Cell Metabolism Modulators Oncogenic transformation causes a "rewiring" of cellular metabolism away from oxidative phosphorylation to glycolysis (historically defined as the "Warburg effect" of aerobic glycolysis in animal and human tumors) with attendant tolerance of hypoxia and production of metabolites important for sustaining cell proliferation. Recent clinical studies have defined clinical value from inhibitors of the cell lipid membrane localized phosphoinositide-3 (PI3) kinase, mTOR, and extra-mitochondrial isocitrate dehydrogenase isoforms 1 and 2.

mTOR is a kinase whose inhibition was originally discovered as the basis for activity by the immunosuppressant rapamycin, isolated from a soil bacterium (originally obtained from Rapa Nui), which had evidence of antitumor activity in animals as well as decreased T-cell proliferation. Sirolimus as a protein-bound formulation is used for certain soft tissue tumors. Temsirolimus and everolimus are mTOR inhibitors with activity in renal cancers. They produce stomatitis and fatigue; some hyperlipidemia (10%) and myelosuppression (10%); and rare lung toxicity and immunosuppression in regimens used clinically. Everolimus is also useful in patients with hormone receptor–positive breast cancers displaying resistance to hormonal inhibition and in certain neuroendocrine and brain tumors, the latter arising in patients with sporadic or inherited mutations in the pathway activating mTOR. PI3 kinase is activated by numerous oncogenic tyrosine kinases to ultimately cause a cascade of metabolic alterations including increased glucose uptake and activation of mTOR isoforms, which selectively increase translation efficiency of key regulators of cell cycle progression and protein synthetic capacity.

Isoform-specific PI3 kinase inhibitors are of increasing importance in breast cancers with mutated *PI3K*α (alpelisib; hyperglycemia and cutaneous eruptions can occur) or owing to selective use of PI3Kδ by lymphoid tissues in lymphomas (idelalisib, copanlisib, and duvelisib).

Isocitrate dehydrogenase (IDH) inhibitors (ivosidenib specific for IDH1 and enasidenib specific for IDH2) have activity in tumors with IDH mutants (AML, cholangiocarcinomas) that generate the "oncometabolite" 2-hydroxyglutarate, which alters DNA and histone methyltransferase activity. The drugs thus function indirectly as epigenetic chromatin modulating agents through effects on cellular metabolism. Vorasidenib, an IDH1/2 inhibitor with very favorable distribution across the blood-brain barrier, is in very advanced stages of development for certain brain tumors with IDH mutations.

DNA Repair Pathway Modulators DNA repair systems act physiologically to lessen the impact of environmental genomic damaging agents and influence the susceptibility to certain chemotherapy agents. DNA repair enzyme mutations underlie inherited cancer susceptibility syndromes such as mutated *BRCA* tumor-suppressor gene–associated breast and ovarian cancers, among others.

Laboratory investigations revealed that poly-ADP ribose polymerase (PARP) acts as a synthetic lethal gene with mutations in the homologous recombination repair pathway, including the *BRCA* gene. PARP responds to detection of DNA lesions by creating chains of poly-ADP, which serve as scaffolds for the localization of DNA repair proteins still active even with mutated *BRCA* isoforms. However, without PARP activity, the scaffolds cannot form, and the DNA damage becomes lethal. This observation immediately suggested the potential utility of PARP inhibitors (e.g., olaparib) as treatments potentially useful for *BRCA*-induced tumors. Recently, PARP inhibitor utility has been extended to tumors that do not harbor *BRCA* mutations but have given evidence of responding to platinum drugs, as a way of extending the useful effect of the chemotherapy treatment. This finding underscores the likelihood that sensitivity to DNA-directed cytotoxic drugs on the part of a tumor is at least in part related to the drugs' ability to take advantage of a sensitizing effect of a tumor's endogenous DNA repair capacity.

Miscellaneous Targeted Therapies High-affinity binding to receptors on tumor cells can target toxic agents besides drugs to tumor cells, exemplified by the IL-3–diphtheria toxin fusion protein tagraxofusp-erzs, targeting the IL-3 receptor (CD123) and useful in blastic plasmacytoid dendritic cell neoplasms (Table 78-7). Capillary leak syndrome induced due to adventitious "off-target" delivery of the toxin component requires careful monitoring of fluid balance to avoid pulmonary dysfunction in particular. The somatostatin receptor conjugated to a chelate of 177-lutetium can deliver targeted radiation to gastroenterohepatic endocrine neoplasms expressing that receptor. Acutely, release of vasoactive and locally acting hormonal components from dying tumor cells and myelosuppression can occur. Renal damage and the risk of second hematopoietic tumors can complicate continuing use of the agent. Likewise, 177-lutetium can be delivered via conjugation to a ligand of prostate-specific membrane antigen (PSMA) in the treatment of hormone-resistant prostate cancer. Ziv-aflibercept is not an antibody, but a solubilized VEGF receptor VEGF binding domain, and therefore may have a distinct mechanism of action from bevacizumab but with similar side effects.

RESISTANCE TO CANCER TREATMENTS

Resistance mechanisms to the conventional cytotoxic agents were initially characterized in the late twentieth century as defects in drug uptake, metabolism, or export by tumor cells. The *multidrug resistance* (*MDR*) gene, encoding P-glycoprotein (Pgp), is prototypic of transport proteins that efficiently excrete many drugs from tumor cells; no clinically useful modulator of this process has yet emerged. Drug-metabolizing enzymes such as cytidine deaminase are upregulated in resistant tumor cells, and this is the basis for so-called "high-dose cytarabine" regimens in the treatment of leukemia. Another resistance mechanism defined during this era involved increased expression of a drug's target, exemplified by amplification of the dihydrofolate reductase gene, in patients who had lost responsiveness to methotrexate, or mutation of topoisomerase II in tumors that relapsed after topoisomerase II modulator treatment.

A second class of resistance mechanisms involves loss of the cellular apoptotic mechanism activated after the engagement of a drug's target by the drug. This occurs in a way that is heavily influenced by the biology of the particular tumor type. For example, decreased alkylguanine alkyltransferase expression defines a subset of glioblastoma patients with the prospect of enhanced benefit from treatment with temozolomide but has no value in predicting benefit from temozolomide in epithelial neoplasms. Likewise, ovarian cancers resistant to platinating agents have decreased expression of the proapoptotic gene *BAX*.

A related class of resistance mechanisms emerged from sequencing of the targets of agents directed at oncogenic kinases, revealing mutated targets, as described previously (e.g., p210$^{bcr\text{-}abl}$T315I). This reflects the phenomenon of tumor heterogeneity with distinct populations of subclones that arise in a tumor during the process of carcinogenesis. These subclones share to variable degrees mutations that may promote the growth of some subclones but that are absent or are no longer relevant to the growth of other subclones. This general mechanism of resistance is emerging as a basis for limited value of clinical responses to sotorasib, where sequencing of tumor DNA after clinical resistance has documented "treatment-emergent" mutations in *KRAS* itself as well as alterations in other proto-oncogenes and their targets. Really useful targeted therapies address a "driver mutation" present in all subclones.

Finally, other mechanisms of resistance to targeted agents include the upregulation of alternate means of activating the pathway targeted by the agent. Thus, melanomas initially responsive to *BRAF* V600E antagonists such as vemurafenib may reactivate RAF signaling by employing variant isoforms that can bypass the drug. Likewise, inhibition of HER2/neu signaling in breast cancer cells can lead to the emergence of variants with distinct ways of activating downstream effectors such as PI3 kinase.

Mechanisms of resistance to immune checkpoint inhibitors such as nivolumab and ipilimumab have not been well defined, but initial characterization of tumors resistant to these agents demonstrates alterations in antigen presentation pathways, concordant with the basis for checkpoint inhibitor action.

SUPPORTIVE CARE DURING CANCER TREATMENT

■ MYELOSUPPRESSION

Most cytotoxic chemotherapeutic agents affect bone marrow function. Polymorphonuclear leukocytes (PMNs; $t_{1/2}$ = 6–8 h), platelets ($t_{1/2}$ = 5–7 days), and red blood cells (RBCs; $t_{1/2}$ = 120 days) have most, less, and least susceptibility, respectively, to usually administered cytotoxic agents. Maximal neutropenia occurs 6–14 days after conventional doses of anthracyclines, antifolates, and antimetabolites. Alkylating agents differ from each other in the timing of cytopenias. Nitrosoureas, DTIC, temozolomide, and procarbazine can display delayed marrow toxicity, first appearing 6 weeks after dosing.

Complications of myelosuppression relate to the missing cells' function. *Febrile neutropenia* refers to the clinical presentation of fever

TABLE 78-8 Indications for the Clinical Use of G-CSF or GM-CSF

Preventive Uses	
With the first cycle of chemotherapy (so-called *primary CSF administration*)	
	Not needed on a routine basis
	Use if the probability of febrile neutropenia is ≥20%
	Use if patient has preexisting neutropenia or active infection
	Age >65 years treated for lymphoma with curative intent or other tumors treated by similar regimens
	Poor performance status
	Extensive prior chemotherapy
	Dose-dense regimens in a clinical trial or with strong evidence of benefit
With subsequent cycles if febrile neutropenia has previously occurred (so-called *secondary CSF administration*)	
	Not needed after short-duration neutropenia without fever
	Use if patient had febrile neutropenia in previous cycle
	Use if prolonged neutropenia (even without fever) delays therapy
Therapeutic Uses	
Afebrile neutropenic patients	
	No evidence of benefit
Febrile neutropenic patients	
	No evidence of benefit
	May feel compelled to use in the face of clinical deterioration from sepsis, pneumonia, or fungal infection, but benefit unclear
In bone marrow or peripheral blood stem cell transplantation	
	Use to mobilize stem cells from marrow
	Use to hasten myeloid recovery
In acute myeloid leukemia	
	G-CSF of minor or no benefit
	GM-CSF of no benefit and may be harmful
In myelodysplastic syndromes	
	Not routinely beneficial
	Use intermittently in subset with neutropenia and recurrent infection
What Dose and Schedule Should Be Used?	
G-CSF: 5 mg/kg per day subcutaneously	
GM-CSF: 250 mg/m^2 per day subcutaneously	
Pegfilgrastim: one dose of 6 mg 24 h after chemotherapy	
When Should Therapy Begin and End?	
When indicated, start 24–72 h after chemotherapy	
Continue until absolute neutrophil count is 10,000/μL	
Do not use concurrently with chemotherapy or radiation therapy	

Abbreviations: CSF, colony-stimulating factor; G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte-macrophage colony-stimulating factor.

Source: From the American Society of Clinical Oncology: J Clin Oncol 24:3187, 2006.

and <1500 granulocytes/μL. Management of febrile neutropenia is considered in **Chap. 79**. Colony-stimulating factors (CSFs) are used to augment bone marrow production of PMNs. The American Society of Clinical Oncology has developed practice guidelines for the use of granulocyte CSF (G-CSF) and GM-CSF **(Table 78-8)**.

Dangerous degrees of thrombocytopenia do not frequently complicate the management of patients with solid tumors receiving cytotoxic chemotherapy (with the possible exception of certain carboplatin-containing regimens), but they are frequent in patients with hematologic neoplasms. Severe bleeding related to thrombocytopenia occurs with increased frequency at platelet counts <20,000/μL in patients with acute leukemia and <10,000/μL in patients with solid tumors and is prevalent at counts <5000/μL.

The precise "trigger" point at which to transfuse patients has been defined as a platelet count of 10,000/μL or less in patients without medical comorbidities that may increase the risk of bleeding. This issue is important not only because of the costs of frequent transfusion but also because unnecessary platelet transfusions expose the patient to the risks of allosensitization and loss of value from subsequent transfusion, as well as the infectious and hypersensitivity risks inherent in any transfusion. Careful review of medication lists to prevent exposure to nonsteroidal anti-inflammatory agents and maintenance of clotting factor levels adequate to support near-normal prothrombin and partial thromboplastin time tests are important in minimizing the risk of bleeding in the thrombocytopenic patient.

Anemia associated with chemotherapy can be managed by transfusion of packed RBCs. Transfusion is not undertaken until the hemoglobin falls to <80 g/L (8 g/dL), compromise of end-organ function occurs, or an underlying condition (e.g., coronary artery disease) calls for maintenance of hemoglobin >90 g/L (9 g/dL). Randomized trials in certain tumors have raised the possibility that erythropoietin (EPO) use may promote tumor cell survival.

NAUSEA AND VOMITING

The most common side effect of chemotherapy administration is nausea, with or without vomiting. Nausea may be acute (within 24 h of chemotherapy), delayed (>24 h), or anticipatory of the receipt of chemotherapy. Highly emetogenic drugs (risk of emesis >90%) include DTIC, cyclophosphamide at >1500 mg/m^2, and cisplatin; moderately emetogenic drugs (30–90% risk) include carboplatin, cytosine arabinoside (>1 g/m^2), ifosfamide, conventional-dose cyclophosphamide, and anthracyclines; low-risk (10–30%) agents include 5-FU, taxanes, etoposide, and bortezomib, with minimal risk (<10%) afforded by treatment with antibodies, bleomycin, busulfan, fludarabine, and vinca alkaloids.

Serotonin antagonists (5-HT$_3$) and neurokinin 1 (NK1) receptor antagonists are useful in "high-risk" chemotherapy regimens. The combination acts at both peripheral gastrointestinal and CNS sites that control nausea and vomiting. For example, the 5-HT$_3$ blocker dolasetron, 100 mg intravenously or orally; dexamethasone, 12 mg; and the NK1 antagonist aprepitant, 125 mg orally, are combined on the day of administration of severely emetogenic regimens, with repetition of dexamethasone (8 mg) and aprepitant (80 mg) on days 2 and 3 for delayed nausea. Alternate 5-HT$_3$ antagonists include ondansetron, given as 0.15 mg/kg intravenously for three doses just before and at 4 and 8 h after chemotherapy; palonosetron at 0.25 mg over 30 s, 30 min before chemotherapy; and granisetron, given as a single dose of 0.01 mg/kg just before chemotherapy. Emesis from moderately emetic chemotherapy regimens may be prevented with a 5-HT$_3$ antagonist and dexamethasone alone for patients not receiving doxorubicin and cyclophosphamide combinations; the latter combination requires the 5-HT$_3$/dexamethasone/aprepitant on day 1, but aprepitant alone on days 2 and 3. Emesis from low-emetic-risk regimens may be prevented with 8 mg of dexamethasone alone or with non-5-HT$_3$, non-NK1 antagonist approaches including the following.

Antidopaminergic phenothiazines act directly at the chemoreceptor trigger zone (CTZ) in the brainstem medulla and include prochlorperazine (Compazine), 10 mg intramuscularly or intravenously, 10–25 mg orally, or 25 mg per rectum every 4–6 h for up to four doses; and thiethylperazine, 10 mg by potentially all of the above routes every 6 h. Haloperidol is a butyrophenone dopamine antagonist given at 1 mg intramuscularly or orally every 8 h. Metoclopramide acts on peripheral dopamine receptors to augment gastric emptying and is used in high doses for highly emetogenic regimens (1–2 mg/kg intravenously 30 min before chemotherapy and every 2 h for up to three additional doses as needed); intravenous doses of 10–20 mg every 4–6 h as needed or 50 mg orally 4 h before and 8 and 12 h after chemotherapy are used for moderately emetogenic regimens. 5-9-Tetrahydrocannabinol (Marinol) is a rather weak antiemetic compared to other available agents, but it may be useful for persisting nausea and is used orally at 10 mg every 3–4 h as needed. Olanzapine, an "atypical antipsychotic" acting at multiple neurotransmitter receptors, may be of value, most clearly in cases refractory to the measures described above. Some practice guidelines have endorsed its earlier use in adults receiving highly emetogenic chemotherapy regimens in combination with an NK1 antagonist plus a 5-HT$_3$ antagonist plus dexamethasone.

DIARRHEA

Similar to the vomiting syndromes, chemotherapy-induced diarrhea may be immediate or can occur in a delayed fashion up to 48–72 h after the drugs. Careful attention to maintained hydration and electrolyte repletion, intravenously if necessary, along with antimotility treatments such as "high-dose" loperamide (4 mg at the first occurrence of diarrhea, with 2 mg repeated every 2 h until 12 h without loose stools, not to exceed a total daily dose of 16 mg), are appropriate. Octreotide (100–150 μg), a somatostatin analogue, or intralumenally acting opiate-based preparations may be considered for patients not responding to loperamide.

MUCOSITIS

Irritation and inflammation of the mucous membranes (mucositis) particularly afflicting the oral and anal mucosa, but potentially involving the entire gastrointestinal tract, may accompany cytotoxic chemotherapy. Topical therapies, including anesthetics and barrier-creating preparations, may provide symptomatic relief in mild cases.

ALOPECIA

Chemotherapeutic agents vary widely in causing alopecia, with anthracyclines, alkylating agents, and topoisomerase inhibitors reliably causing near-total alopecia when given at therapeutic doses. Antimetabolites are more variably associated with alopecia. Psychological support and the use of cosmetic resources are to be encouraged. "Chemo caps" that reduce scalp temperature to decrease the degree of alopecia are controversial during treatment with curative intent of neoplasms, such as leukemia or lymphoma, or in adjuvant breast cancer therapy. The richly vascularized scalp can certainly harbor micrometastatic or disseminated disease.

GONADAL DYSFUNCTION AND PREGNANCY

All cancer treatments described in this chapter should be regarded as potentially injurious to the developing fetus and to newborns via lactation. However, there are gradations to the degree of reproductive harm. All agents tend to have increased risk of adverse outcomes when administered during the first trimester, and strategies to delay chemotherapy, if possible, until after this milestone should be considered if the pregnancy is to continue to term. Patients in their second or third trimester can be treated with most regimens for the common neoplasms afflicting women in their childbearing years, with the exception of antimetabolites, particularly antifolates, which have notable teratogenic or fetotoxic effects throughout pregnancy. The need for anticancer chemotherapy per se is infrequently a clear basis to recommend termination of a concurrent pregnancy, although each treatment strategy in this circumstance must be tailored to the individual needs of the patient.

Cessation of ovulation and azoospermia reliably result from regimens that contain alkylating agents and topoisomerase poison. The duration of these effects varies with age and sex. Sperm banking before treatment may be considered. Females experience amenorrhea with anovulation after alkylating agent therapy; egg preservation may be considered but may delay inception of urgent treatment. Recovery of normal menses is frequent if treatment is completed before age 30, but patients are unlikely to recover menses after age 35. Even those who regain menses usually experience premature menopause. Because the magnitude and extent of decreased fertility can be difficult to predict, patients should be counseled to maintain effective contraception, preferably by barrier means, during and after therapy. Resumption of efforts to conceive should be considered in the context of the patient's likely prognosis. Hormone replacement therapy should be undertaken in women who do not have a hormonally responsive tumor. For patients who have had a hormone-sensitive tumor primarily treated by a local modality, conventional practice would counsel against hormone replacement, but this issue is under investigation.

PALLIATIVE AND SUPPORTIVE CARE

An important perspective the primary care provider may bring to patients and their families facing incurable cancer is that, given the limited value of chemotherapeutic approaches at some point in the natural history of most metastatic cancers, *palliative care* or *hospice-based* approaches, with meticulous and ongoing attention to symptom relief and with family, psychological, and spiritual support, should receive prominent attention as a valuable therapeutic plan **(Chaps. 13 and 73)**. Optimizing the quality of life rather than attempting to extend it becomes a valued intervention. Patients facing the impending progression of disease in a life-threatening way frequently choose to undertake toxic treatments of little to no potential value, and support provided by the primary caregiver in accessing palliative and hospice-based options in contrast to receiving toxic and ineffective regimens can be critical in providing a basis for patients to make sensible choices.

Late effects of cancer and its treatment are reviewed in **Chap. 100**.

FURTHER READING

Ascione L et al: Predicting response to antibody drug conjugates: A focus on antigens' targetability. Oncologist 28:944, 2023.

Chabner BA, Longo DL: *Cancer Chemotherapy, Immunotherapy, and Biotherapy; Principles and Practice*, 7th ed. Philadelphia, Wolters Kluwer, 2025.

Emadi A, Karp JE: *Cancer Pharmacology: An Illustrated Manual of Anticancer Drugs*, 2nd ed. New York, Springer Publishing Co., 2024.

Federman N: Molecular pathogenesis of desmoid tumor and the role of γ-secretase inhibition. NPJ Precis Oncol 6:62, 2022.

Hesketh PJ et al: Antiemetics: American Society of Clinical Oncology clinical practice update. J Clin Oncol 35:3240, 2017.

Kaelin WG Jr: Von Hippel-Lindau disease: Insights into oxygen sensing, protein degradation, and cancer J Clin Invest 132:e162480, 2022.

Morad G et al: Hallmarks of response, resistance, and toxicity to immune checkpoint blockade. Cell 184:5309, 2021.

Nikanjam M et al: Liquid biopsy: Current technology and clinical applications. J Hematol Oncol 15:131, 2022.

Puzanov I et al: Managing toxicities associated with immune checkpoint inhibitors: Consensus recommendations from the Society for Immunotherapy of Cancer (SITC) Toxicity Management Working Group. J Immunother Cancer 5:95, 2017.

Rosen N, Longo DL: Targeting oncogenic RAS protein. N Engl J Med 387:184, 2022.

Sartor O et al: Lutetium-177–PSMA-617 for metastatic castration-resistant prostate cancer. N Engl J Med 385:1091, 2021.

79

Infections in Patients with Cancer

Brahm H. Segal, Juan C. Gea-Banacloche

GENERAL CONCEPTS

Infection is an important complication of cancer and cancer therapy and drives excess hospitalization and mortality. Prevention, diagnosis, and treatment of infection increases survival and improves quality of life. In this regard, the infectious diseases consultant works collaboratively with multiple stakeholders to prevent infection and to effectively diagnose and treat infections when they occur. Cornerstones of infection prevention are compliance with standard infection control guidelines, limiting exposure to pathogens, vaccination of patients and caregivers, and targeted use of antimicrobial prophylaxis. Although there have been substantial improvements in diagnostic modalities for infection including molecular and antigen-based diagnostics, the infectious diseases physician is frequently confronted with patients with suspected infection (e.g., neutropenic fever, lung lesions observed by imaging) without a definitive diagnosis. In addition, infectious and noninfectious disorders may have overlapping manifestations, such as pneumonia versus drug-related pneumonitis and fever from infection versus cancer-associated fever. Given the broad differential diagnosis, aggressive diagnostic evaluation including biopsies may be required. Sometimes, empirical therapy must be administered based on the most likely or dangerous infections; these situations are common in patients with hematologic malignancies in whom an invasive tissue diagnosis may be unsafe due to thrombocytopenia.

We provide recommendations for therapy of both documented and suspected infections. Specific treatment plans should consider several factors, such as evidence of clearance of infection, whether a persistent nidus exists (e.g., abscess or infectious phlebitis), the specific pathogen, and the immune status of the patient. A general principle of therapy is that longer courses are required in patients with persistent and severe immunocompromise (e.g., prolonged neutropenia in patients with acute leukemia or myelodysplastic syndrome). An individualized approach to the diagnosis and management of infections that is tailored to overall goals of care is recommended. For example, in the setting of uncontrolled malignancy (e.g., recurrent cholangitis from obstructive pancreatic or biliary cancers or secondary infections of tumor from bowel fistulization), source control by surgery or catheter drainage may not be feasible, and antibiotics may be used palliatively for patient comfort and to avoid unnecessary hospitalizations rather than to cure an infection.

We will address the risk factors for infections and the preventive measures that may be adopted based on those risk factors. Most specific infections have dedicated chapters, and readers are advised to access those for in-depth discussions. We will focus here on infections associated with (or caused by) the treatment of cancer. Chemotherapy-induced neutropenia is the most important etiology, and the management of neutropenic fever will be discussed in detail. We will also discuss infections associated with new treatment modalities, including biologics, immunotherapy, and cellular therapies. Infections related to hematopoietic stem cell transplant (HCT) are discussed in **Chap. 148**.

RISK FACTORS FOR INFECTION IN PATIENTS WITH CANCER

When evaluating patients with cancer and suspected infection, it is important to consider the major factors—both intrinsic and treatment-associated—that predispose to infection. This knowledge, in turn, guides the differential diagnosis, diagnostic evaluation, and initial therapy.

RISK FACTORS INTRINSIC TO THE CANCER

These are the direct consequence of the cancer. In the case of solid tumors, the most obvious risk factor for infection relates to obstruction. As examples, lung tumors that obstruct the airway predispose to postobstructive pneumonia, obstructive pancreatic and biliary tumors predispose to cholangitis, and tumors that result in obstructive uropathy predispose to urinary tract infections. Tumors of the head and neck increase the risk of local infection and aspiration pneumonia. Gastrointestinal tumors can present with obstruction and local abscess production by enteric flora and bloodstream infection (e.g., bloodstream infection by *Streptococcus gallolyticus* or *Clostridium septicum* in colon cancer). Direct invasion through the colonic mucosa is associated with local abscess formation and sepsis by enteric flora.

Some hematologic malignancies are associated with specific immune deficits. Acute leukemia and myelodysplastic syndrome can manifest with pancytopenia at diagnosis. Patients may have a high burden of circulating leukemic cells and lack normal circulating neutrophils. Some patients with myelodysplastic syndrome or acute myelogenous leukemia (AML) have mutations that are associated marrow failure and immunodeficiency. For example *GATA2* deficiency is associated with major viral, bacterial, and fungal infections as well as hematological malignancies and solid tumors. Multiple myeloma and chronic lymphocytic leukemia are associated with impaired immunoglobulin production that can manifest with recurrent sinopulmonary infections and poor antibody responses to both prior infections and vaccination. Case series suggest that patients with hairy cell leukemia are at increased risk of nontuberculous mycobacterial infections. T cell–associated leukemias and lymphomas can be associated with human T lymphotropic virus 1 (HTLV1) and have T cell impairment and opportunistic infections by *Pneumocystis jirovecii*, *Cryptococcus neoformans*, tuberculosis, or disseminated strongyloidiasis.

RISK FACTORS ASSOCIATED WITH CANCER THERAPY

There has been a dramatic expansion in cancer therapeutics, most of which influence host defense against infections **(Table 79-1)**. Broadly speaking, cancer therapy can compromise either or both the physical barriers and immune responses that protect from infection. Examples of physical barrier disruption include central venous catheters, surgical wounds that can be a portal of entry for skin microbes, and disruption of the lymphatic drainage after mastectomy and lymphadenectomy for breast cancer.

The mucosal lining of the gastrointestinal tract, respiratory tract, and other luminal surfaces constitutes the first line of host defense, both as a physical barrier and by secretion of a variety of antimicrobial products, such as immunoglobulin A (IgA), lactoferrin, and antimicrobial peptides. Standard nonselective cytotoxic agents have the combined effect of both pancytopenia and mucosal disruption that predispose to infection. Radiation is sometimes administered concurrently with chemotherapy and, depending on the location of the radiation field, can cause significant mucosal injury. Antineoplastic agents that inhibit specific pathway(s) that drive tumor cell progression are considered targeted, but they also impair specific components of the immune system whose function relies on these same pathways. For example, the anti-tumor activity of Bruton tyrosine kinase inhibitors against B-cell malignancies is by inhibiting the B-cell receptor signaling cascade, but this property can have broad immune effects that increase the risk of viral and fungal infections.

Chemotherapy-Induced Neutropenia It has been recognized for more than 50 years that the duration and degree of neutropenia drive susceptibility to infections in patients with cancer. Cytotoxic regimens that result in prolonged neutropenia also deplete other immune cells (e.g., circulating monocytes and lymphocytes) and commonly cause mucosal injury. The risk of infection is proportional to the degree of neutropenia once the absolute neutrophil count (ANC) becomes <1000/μL. For example, standard induction chemotherapy with anthracycline plus cytarabine for AML causes prolonged neutropenia and severe mucositis. The combination of prolonged neutropenia and mucositis predisposes these patients to gastrointestinal tract infections that include ulcerations of the oral mucosa, neutropenic enterocolitis (typhlitis), and perirectal infections, as well as bloodstream infections by gastrointestinal flora such as viridians group streptococci, enterococci, and Enterobacterales. Reactivation of oral mucosal herpes simplex virus (HSV) is another common complication of leukemia therapy. Reflecting the fact that *Candida* species are endogenous gastrointestinal flora, these patients are also at risk for candidemia and chronic disseminated candidiasis. Patients with more prolonged neutropenia (e.g., ANC <500/μL for ≥2 weeks) are at risk for invasive aspergillosis and other molds. Refractory and relapsed acute leukemia further increase the risk of invasive fungal disease (Table 79-1).

Antimicrobial prophylaxis tailored to the underlying immune defects has been shown to be effective in several settings, and it should be used with knowledge of the proven benefits and the direct and indirect toxicities (prevention of fever differs from prevention of infection and prolongation of survival), as well as the limitations of the available evidence. As an example, prophylaxis with a fluoroquinolone like levofloxacin should be considered in adults with prolonged neutropenia (e.g., ANC <500/μL for ≥7 days). Multiple studies have shown decreased frequency of neutropenic fever and fewer documented infections, and meta-analyses suggest a survival benefit. In patients with AML receiving induction chemotherapy, prophylaxis with posaconazole was associated with less invasive fungal disease (IFD) and a survival benefit, so posaconazole is frequently recommended as prophylaxis in patients with profound, prolonged neutropenia. It is plausible that isavuconazole may be equally effective, but its efficacy in this setting has not been demonstrated.

TABLE 79-1 Immune Defects and Associated Infections in Cancer Patients

HOST DEFENSE DEFECT	PREDOMINANT PATHOGENS	PATIENTS WITH CANCER AT GREATEST RISK
Neutropenia (ANC <500/μL)	Gram-negative and gram-positive bacteria	Cytotoxic chemotherapy, underlying hematologic malignancy (e.g., myelodysplasia, acute leukemia)
Prolonged (≥10 days), profound neutropenia (ANC <100/μL)	Increased risk of bacterial infections Candidemia Invasive aspergillosis and other molds HSV reactivation Respiratory viral infections	Induction/reinduction therapy for acute leukemia; pancytopenia from underlying hematologic malignancy; myeloablative conditioning regimens for stem cell transplantation and lymphodepletion for adoptive cellular therapy
T cell immunodeficiency	Common bacterial infections Intracellular bacteria (e.g., *Listeria monocytogenes*, *Salmonella* species) *Nocardia* species Tuberculosis and NTM Respiratory viral infections Reactivation of herpes viruses (HSV, VZV); with severe impairment: CMV, EBV-associated lymphoproliferative disease; HHV-6-associated marrow suppression or encephalopathy, and HHV-8-associated malignancies) Mucosal candidiasis *Pneumocystis jirovecii* Dimorphic fungal infections (e.g., histoplasmosis, coccidioidomycosis) *Cryptococcus neoformans* Invasive aspergillosis and other molds Toxoplasmosis *Strongyloides* hyperinfection	Underlying hematologic malignancy including primary T cell malignancies AIDS-associated malignancies Corticosteroids, TNF-α blockade Janus kinase inhibitors Purine analogues Alemtuzumab GVHD Lymphodepletion for adoptive cellular therapy
B cell immunodeficiency	Encapsulated bacteria (*Streptococcus pneumoniae*, *Haemophilus influenzae*, *Neisseria meningitidis*) Respiratory viral infections Reactivation of HSV and VZV Reactivation of HBV PML (reactivation of JC virus)	Lymphoid malignancies (e.g., chronic lymphocytic leukemia, multiple myeloma) B cell–depleting agents (e.g., rituximab, BTK inhibitors) Stem cell transplantation, particularly with chronic GVHD Bispecific antibodies and adoptive cellular therapy targeted against B cell antigens (e.g., CD19-directed CAR T cell regimens)
Splenectomy and functional asplenia	Encapsulated bacteria (can result in life-threating sepsis) Malaria, babesiosis	Functional asplenia in chronic GVHD
Mucosal injury	Localized infections by oral and GI flora (e.g., dental infections, neutropenic enterocolitis, perianal infection) Bacteremia (coliforms, oral streptococci, enterococci, anaerobes; can be polymicrobial)	Cytotoxic chemotherapy that results in both neutropenia and mucosal injury Radiation (e.g., for head and neck tumors or rectal cancer) predisposes to local tissue damage, impaired blood flow, and secondary bacterial infection
Systemic corticosteroids	Broad suppressive effect on innate and adaptive immunity related to dose and duration of treatment Increased risk of common bacterial and viral infections Prolonged high-dose corticosteroids (e.g., prednisone equivalent to ≥20 mg/day for ≥28 days) increases risk of multiple opportunistic pathogens associated with impaired T cell immunity (e.g., *Pneumocystis jirovecii*)	Corticosteroids are common components of antineoplastic regimens for hematologic cancers (e.g., for acute lymphoblastic leukemia, lymphomas, and multiple myeloma) and are administered concurrently with other immunosuppressive therapies Other common indications for corticosteroids are to reduce inflammation from central nervous system tumors and as therapy for immune-related toxicities

Abbreviations: AIDS, acquired immunodeficiency syndrome; ANC, absolute neutrophil count; BTK, Bruton tyrosine kinase; CAR, chimeric antigen receptor; CMV, cytomegalovirus; EBV, Epstein-Barr virus; GVHD, graft-versus-host disease; HBV, hepatitis B virus; HHV, human herpes virus; HSV, herpes simplex virus; MDS, myelodysplastic syndrome; NTM, nontuberculous mycobacteria; PML, progressive multifocal leukoencephalopathy; TNF, tumor necrosis factor; VZV, varicella-zoster virus.

Corticosteroids Corticosteroids are part of several chemotherapy regimens used in hematologic malignancies, such as in acute lymphoid leukemias, lymphomas, and multiple myeloma. Patients with brain tumors, both primary and metastatic, are treated with corticosteroids to control edema. In addition, corticosteroids are a mainstay of therapy to control inflammatory complications of cancer therapy. For example, immune checkpoint inhibitors are associated with pneumonitis, colitis, and autoimmune endocrinopathies that are treated with corticosteroids, and additional agents such as tumor necrosis factor α blockade when corticosteroids alone are insufficient. Corticosteroids are also used as therapy for graft-versus-host disease (GVHD) following allogeneic stem cell transplantation and for cytokine release syndrome (CRS) following bispecific antibody therapy and adoptive cellular therapy.

High-dose corticosteroids have inhibitory effects on multiple components of innate and adaptive immunity. The risk of infections is related to their dose and duration, the underlying malignancy, and other immunosuppressive agents that are used concurrently. For example, induction therapy for acute lymphoblastic leukemia includes both cytotoxic agents that result in pancytopenia and corticosteroids that cumulatively increase infection risk. Corticosteroids can also decrease signs of infection such as fever and abdominal tenderness. The inhibitory effect of corticosteroids on T cell immunity increases the risk of infections by viruses (e.g., HSV, varicella-zoster virus [VZV]), mycobacteria, *Nocardia* species, and fungal infections, including mucosal candidiasis, dimorphic fungi, and *Pneumocystis jirovecii* pneumonia (PJP) (Table 79-1). Prophylaxis against *Pneumocystis jirovecii* with trimethoprim sulfamethoxazole (TMP-SMX) (or an alternative agent if intolerant) is recommended in patients receiving the adult equivalent of prednisone ≥20 mg per day for at least 4 weeks.

Radiation Radiation causes direct tissue damage (including damage to the vasculature) and impairs wound healing. Infectious complications of radiation include local infection and fistulization. As an example, neoadjuvant radiation therapy for rectal tumors increases the risk of surgical site infections. Radiation to the head and neck increases the risk of infections related to the tumor and as a surgical complication. Osteoradionecrosis of the jaw predisposes to secondary infections of the soft tissue and bone that may require combined prolonged antibiotics and surgical removal of dead bone and reconstruction.

Splenectomy Splenectomy may be performed for diagnosis or treatment of cancer. The spleen has several key immune functions including removing of bacteria from blood, antigen presentation to T cells, and housing B cells that are activated and produce antibodies. Functional asplenia is present after splenic irradiation and with chronic GVHD. Asplenic patients are principally at risk for overwhelming sepsis by encapsulated bacteria like *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis*. Patients should be vaccinated against these pathogens prior to planned splenectomy. Antibiotic prophylaxis (e.g., with penicillin) should be considered for the first 2 years after splenectomy, especially in patients with active malignancy or receiving immunosuppressive therapy. Asplenic patients with fever should be started promptly on antibiotics active against *S. pneumoniae* (e.g., ceftriaxone or levofloxacin or moxifloxacin). Vancomycin should be added in areas where high-level penicillin or cephalosporin resistance is common. Asplenic individuals also have increased risk of babesiosis, malaria, and infection with *Capnocytophaga canimorsus* (associated with animal bites or scratches) and *Salmonella* spp.

B Cell–Depleting Agents B cell–depleting agents are used as therapy for patients with B cell malignancies, such as B cell lymphomas and chronic lymphocytic leukemia. They are commonly used in combination with other antineoplastic agents. Antibody-based drugs such as rituximab result in prolonged B cell depletion, while Bruton tyrosine kinase (BTK) inhibitors (e.g., ibrutinib) interrupt a key enzyme and B cell activation but have shorter-acting effects. The effects of B cell depletion include increased risk of encapsulated bacteria and respiratory viral infections. Importantly, these patients are likely to have reduced immune responses to vaccination against bacteria and viruses, including influenza, SARS-CoV-2, and hepatitis B. The risk of severe COVID-19 is substantially increased in patients with hematologic malignancies compared with patients with solid tumors, and B cell–depleting agents are associated with increased COVID-19 severity and lack of serologic responses to SARS-CoV-2 infection and vaccination. All patients who will receive B cell–depleting agents should be screened for hepatitis B infection and receive therapy for hepatitis B (e.g., with entecavir or tenofovir alafenamide) for active or prior hepatitis B infection (see **Chaps. 350** and **352** on hepatitis viruses and **Chap. 208** on human immunodeficiency virus [HIV] infections). Rituximab and other B cell–depleting agents have been associated with progressive multifocal leukoencephalopathy (PML) and PJP, but the magnitude of the effect is difficult to ascertain. In addition, BTK inhibitors can have off-target effects on innate immune cells and increase the risk of other infections, including invasive aspergillosis, particularly when combined with other immunosuppressive agents, such as corticosteroids.

Bispecific Antibodies Bispecific antibodies are engineered to have dual specificity for a T cell antigen (e.g., CD3, a component of the T cell receptor complex) and a tumor antigen with the goal of stimulating T cell killing of tumor cells. When bispecific antibody constructs directed against B cell antigens (e.g., blinatumomab, a bispecific antibody against CD3 and the B cell antigen CD19) are used against B cell malignancies, the overall effect is a nonselective depletion of both tumor cells and normal B cells. Importantly, bispecific antibodies are commonly used in patients with refractory hematologic malignancies who already are at high risk for infection based on the underlying cancer and prior therapy. Bispecific antibody therapy can result in CRS, characterized by fever and inflammatory organ damage. Treatment of CRS involves high-dose corticosteroids and interleukin 6 (IL-6) blockade, which further increases the risk of infections.

Adoptive Cellular Therapy Adoptive cellular therapy (ACT) involves administration of genetically engineered cells targeted to tumor antigens. The most commonly used ACT is chimeric antigen receptor (CAR) T cells, in which autologous T cells are genetically engineered to express a receptor directed against a tumor antigen. Engagement of cells expressing this antigen results in the CAR T cell activation and tumor cell killing. In contrast to bispecific antibodies, current ACT involves lymphodepletion chemotherapy (usually with fludarabine and cyclophosphamide) to deplete immune cells and allow maximal expansion of the engineered T cells. A major goal of immunotherapy is to make ACT more effective for a broad range of solid tumors; however, the current use of ACT largely involves refractory B cell malignancies: lymphoblastic leukemia, lymphomas, and multiple myeloma. Lymphodepletion regimens result in pancytopenia, and both neutropenia and lymphopenia can be prolonged (months to years). In the case of standard CD19-directed CAR T cells, B cell depletion is by design; the persistence of CAR T cells is required for antitumor immunity, but it is nonselective, and the duration of global B cell depletion can be for years. ACT can also be complicated by CRS and an immune effector cell–associated neurotoxicity syndrome that can be life-threatening and requires treatment with intensive high-dose corticosteroids often combined with IL-6 blockade. Additional hematologic complications can occur after ACT, such as hemophagocytic lymphohistiocytosis (HLH), which extends pancytopenia and entails additional immunosuppressive therapy.

Recommendations regarding prophylaxis following ACT are based mainly on extrapolation and expert opinion. These include levofloxacin during neutropenia, TMP-SMX while on systemic corticosteroids or while CD4 count is <200/μL, acyclovir, and perhaps mold-active prophylaxis (e.g., posaconazole) during periods of prolonged neutropenia. Recipients of CAR T cells constitute a heterogeneous population due to the effect of multiple previous courses of therapy, sometimes including allogeneic HCT, and prophylaxis should be tailored to the degree of immunosuppression.

Cancer and HIV Infection HIV-associated malignancies include Epstein-Barr virus–associated non-Hodgkin and Hodgkin lymphomas, HHV-8-associated Kaposi sarcoma and primary effusion lymphoma, and human papilloma virus–associated cervical and anal cancer. HIV-positive patients have higher risks of chronic hepatitis B and hepatitis C infections that increase the risk of hepatocellular carcinoma. Patients with HIV infection can also have cancers that are not HIV-associated. A mainstay of cancer management in patients who are HIV-positive is that the antineoplastic regimen must be concurrent with antiretroviral therapy and appropriate prophylactic antimicrobials. The goals of antiretroviral therapy are a nondetectable HIV viral load and tolerability of the regimen, which includes monitoring and avoiding drug–drug interactions with antineoplastic chemotherapy. In general, it is recommended to test for HIV and hepatitis B and hepatitis C infection prior to starting antineoplastic therapy.

DIAGNOSIS AND MANAGEMENT OF INFECTIONS IN PATIENTS WITH NEUTROPENIA

Chemotherapy-induced neutropenia remains the major risk factor for infection in patients with cancer. However, patients with cancer frequently have multiple risk factors for infection, both from the underlying malignancy and its treatment. When evaluating patients with neutropenia and suspected infection, it is important to consider these other risk factors in the diagnostic evaluation and therapy **(Fig. 79-1)**.

NEUTROPENIC FEVER

A high proportion of cancer patients who become neutropenic after receiving chemotherapy develop fever. The standard definition of neutropenic fever (NF) is a single oral temperature ≥38.3°C or a temperature of ≥38.0°C sustained over 1 h, with an absolute neutrophil count

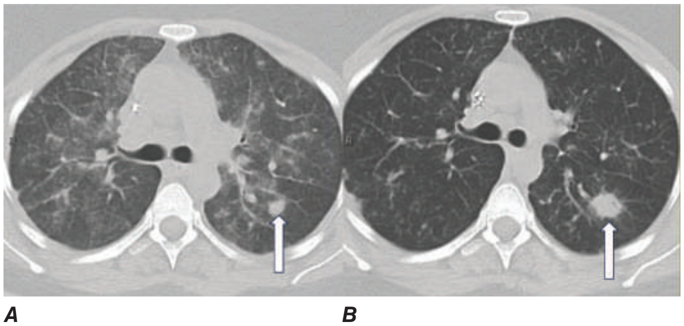

FIGURE 79-1 Multiple pulmonary infections in a patient with acute myelogenous leukemia (AML). A patient with AML in remission after reinduction therapy with cytarabine, granulocyte colony-stimulating factor, and fludarabine (FLAG) presented with fever and neutropenia. A chest CT demonstrated diffuse pulmonary infiltrates (***A***). Bronchoalveolar lavage (BAL) was unrevealing except for positive PCR for *Pneumocystis jirovecii*, and trimethoprim-sulfamethoxazole was instituted. After 11 days of appropriate treatment (***B***), a repeat CT showed resolution of the infiltrates but a conspicuous, dense, well-circumscribed pulmonary nodule that had not been appreciated initially. Targeted BAL of the left lower lobe was positive for galactomannan, providing the diagnosis of probable invasive aspergillosis. The initial BAL had been of the right middle lobe only. Immunocompromised patients may have several infections simultaneously, and cancer patients often accumulate risk factors for infection during their treatment.

(ANC) of <500 cells/μL, or an ANC that is expected to decrease to <500 cells/μL over the next 48 h.

During neutropenia localizing signs and symptoms of infection may be subtle or altogether lacking, and infections may progress very quickly. These two basic features mandate early initiation of empirical antibacterial agents in neutropenic patients whenever infection is suspected. Although fever is the most common sign of infection, it is not the only one, and similar management should be used whenever infection is suspected in a neutropenic patient, such as in the presence of otherwise unexplained pain, tenderness, or erythema potentially secondary to infection.

NF is considered infectious in origin, but an infection is identified in only a minority of cases. An infection may be documented microbiologically (e.g., *Pseudomonas aeruginosa* bacteremia identified by positive blood cultures) or only clinically (e.g., abdominal pain and bloody diarrhea with negative blood cultures, presumed to represent neutropenic enterocolitis). Using standard diagnostic methods (i.e., routine cultures, serologic tests, and imaging studies) an infection is documented in approximately 40% of episodes of NF. However, newer diagnostic modalities using plasma cell–free DNA polymerase chain reaction (PCR) identify a potential bacterial etiology most of the time. Most bacteria identified by these studies are part of the normal flora of skin and bowel (as are the bacteria isolated by standard culture methods when these are positive), since these physical barriers are often disrupted by the cancer treatments that caused the neutropenia. Infections during neutropenia are typically caused by microorganisms carried by the patient as part of their microbiome.

■ MANAGEMENT OF NEUTROPENIC FEVER

First Neutropenic Fever There is a wealth of good-quality evidence to support standard-of-care guidelines for the management of the initial episode of NF. After a swift history and physical exam (focused on potential portals of entry like the vascular catheter exit site, mouth, and perianal area), blood cultures are obtained and empirical antibiotic therapy is initiated using a single agent (monotherapy) with broad-spectrum activity, including coverage of *P. aeruginosa*. Ideally, antibiotics should be given within 1 h of the onset of NF. The diagnostic utility of chest imaging in the absence of respiratory symptoms or signs has not been established.

Ceftazidime, cefepime, imipenem, meropenem, and piperacillin-tazobactam are the best-studied antibiotics used as monotherapy for NF. The specific choice will vary from institution to institution based on the local frequency of resistant bacteria. This "backbone regimen" may be complemented with a second agent against resistant gram-negative or gram-positive bacteria depending on clinical features, prior history of resistant pathogens, or local trends of antimicrobial resistance.

It should be emphasized that standard empirical monotherapy regimens for NF apply to clinically stable patients with NF of unclear etiology. Management of neutropenic patients with localized signs of infection (e.g., respiratory, intra-abdominal, intravascular catheter–associated) are discussed below. In patients who are clinically unstable, such as those with hypotension or signs of organ injury raising concern for sepsis (e.g., impaired mental status, pulmonary edema, acute renal injury), a broader-spectrum antimicrobial regimen is warranted. Although the specific regimen for presumed or documented septic shock varies based on the patient's prior history of resistant infection and local susceptibility patterns, an example of an initial regimen is vancomycin, meropenem, and possibly an aminoglycoside; in patients at risk for candidiasis, an echinocandin may be added.

The standard monotherapy regimens lack activity against specific gram-positive pathogens. Gram-positive coverage with a glycopeptide antibiotic (vancomycin in the United States) is not routinely part of the initial regimen. Instead, the addition of a glycopeptide is reserved for clinical situations in which a gram-positive pathogen resistant to the standard regimen (e.g., methicillin-resistant *Staphylococcus aureus* [MRSA]) is more likely. The Infectious Diseases Society of America (IDSA) guidelines recommend empirical gram-positive coverage with sepsis, clinically evident soft tissue infection, clinically suspected catheter exit site infection, pneumonia, severe mucositis (only if ceftazidime is used as empirical regimen, because mucositis is a risk factor for viridans group streptococci), and known carrier status of MRSA or penicillin-resistant pneumococcus. Meta-analysis has shown that the routine inclusion of vancomycin in the initial regimen is not associated with better outcomes but does result in more nephrotoxicity.

After starting antibiotics, fever usually resolves in 24–72 h. If an infection was diagnosed, antibiotics are continued for the appropriate amount of time for that specific infection. When a specific microbe is isolated (e.g., from blood cultures), experts disagree regarding the need for continuing the broad-spectrum coverage versus narrowing based on susceptibility results.

If no microorganism is isolated and no source is identified, the optimal strategy for antibiotic management of NF that resolves with antibiotics is debated. In the past, empirical antibiotics were continued until resolution of neutropenia. However, recent studies have suggested that de-escalation from empiric therapy to prophylaxis or discontinuation are safe if specific criteria are met, including clinical stability. The advantage of such de-escalation or early discontinuation strategies relates to reduced antibiotic exposure and lower risk of selection of resistant pathogens; these benefits must be balanced against the potential for inadequately treated infection. Large, randomized, multicenter trials are required to evaluate the benefits and safety of antibiotic de-escalation or discontinuation of empirical antibiotic therapy in patients with persistent neutropenia.

Persistent Fever If the fever continues while receiving empirical antibiotic therapy without a diagnosis, evidence supports and guidelines recommend continuing the same antibiotic without addition or modification in the absence of clinical changes or new microbiologic data. The antibacterial regimen should be modified only if new clinical features arise (e.g., hypotension) of if new microbiologic data become available (e.g., a positive culture), and not just because of persistent fever. Although there is agreement on this issue between different guidelines, persistent, stable fever is a common reason why the antibacterial regimen is modified in clinical practice. The longer the neutropenic fever persists during the administration of broad-spectrum antibiotics, the higher the likelihood of IFD. Depending on prior history, use of antifungal prophylaxis, and local practices, one may choose to initiate empirical antifungal treatment after 5 days of fever (this was standard of care for decades and is called *empirical antifungal therapy*) or to perform tests focused on diagnosing an occult fungal infection and initiate antifungals only if these additional tests support the possibility of IFD. This more recent approach has been named *preemptive antifungal strategy* and seems to result in similar patient outcomes with less use of antifungals. The empirical antifungal of choice will vary depending on the antifungal prophylaxis used (if any). Caspofungin and liposomal amphotericin B are the best-studied empirical antifungal treatment of NF, but expert recommendations vary. A frequently used approach is to administer posaconazole as antifungal prophylaxis in high-risk patients with neutropenia (e.g., those receiving induction or reinduction regimens for acute myeloid leukemia) and to not modify the antifungal regimen based solely on persistent or recurrent neutropenic fever.

Recrudescent Fever and Fever at the Time of Neutrophil Recovery In addition to the initial episode of fever and persistent fever, there are two other NF scenarios that have been less well studied: recrudescent fever and fever at the time of neutrophil recovery. Recrudescent fever refers to the situation in which the initial episode of fever resolves without a diagnosis, neutropenia persists, the antibiotic regimen is continued unmodified, and fever reappears after the patient has been afebrile for 48–96 h. This is not uncommon during prolonged neutropenia during induction or reinduction therapy in acute leukemia. In this case (as opposed to the first episode of NF), an infection is identified most of the time (bacterial, fungal, or viral), and the recommended approach is to empirically modify the antimicrobial regimen that had successfully controlled the fever up to this point, aiming for coverage of resistant bacteria and fungi. Intensive diagnostic procedures, including CT imaging, should be undertaken. Where available, CT combined with positron emission tomography (CT-PET) may be considered.

Finally, fever at the time of neutrophil recovery may be infectious in origin: either a preexisting infection that was silent due to lack of neutrophils (sometimes this is considered analogous to the immune reconstitution inflammatory syndrome [IRIS] seen after initiation of antiretroviral therapy in AIDS) or, less likely, a superinfection. However, fever is frequently noninfectious at this time and is related to myeloid engraftment. In any case, the recommended response to fever that occurs simultaneously with resolution of neutropenia is not any specific empirical antimicrobial regimen, but to pursue a thorough diagnostic evaluation.

OUTPATIENT THERAPY FOR NEUTROPENIC FEVER

Outpatient therapy for neutropenic fever should be considered in patients at lower risk for infectious complications from neutropenia. These patients typically have solid rather than hematologic tumors with a short expected duration of neutropenia, usually 7 days or less. Criteria for candidates for outpatient empirical therapy include clinical stability, no localizing symptoms or signs of infection, ability to eat and drink without difficulty, absence of significant comorbidities such as chronic lung or cardiovascular disease, normal renal and liver function, and easy access to a hospital and a caregiver at home. Outpatient oral antibiotic therapy in adults typically involves a quinolone (e.g., levofloxacin or ciprofloxacin) plus amoxicillin-clavulanate; such regimens are appropriate only for patients who have not received a quinolone as prophylaxis.

DOCUMENTED INFECTIONS DURING NEUTROPENIA

BACTEREMIA

Bacteremia is the most common microbiologically documented infection in neutropenic patients with fever. NF is associated with bacteremia in only 10% of cases, but this subgroup of patients has much higher mortality than the general group of neutropenic patients with fever, particularly with gram-negative bacteremia. Most recent series show similar proportions of gram-positive (coagulase-negative *Staphylococcus*, *Streptococcus* species, *Enterococcus* species including vancomycin-resistant *Enterococcus* [VRE]) and gram-negative bacteria (Enterobacterales and *P. aeruginosa*).

Bacteremia in neutropenic patients may be secondary to another site of infection (e.g., pneumonia, neutropenic enterocolitis, cellulitis) or, most commonly, to mucosal barrier injury (MBI) caused by the antineoplastic treatment. The chemotherapy and/or radiation frequently damages the mucosa of the gastrointestinal tract, facilitating translocation of commensal bacteria during neutropenia. Clinically, one may question whether the bacteremia may be a central line–associated bloodstream infection (CLABSI). If the isolate is an intestinal bacterium, it is considered related to MBI in the absence of local signs of infection involving the venous catheter. If the isolate is part of the skin flora, like coagulase-negative *Staphylococcus*, it is more likely to be a CLABSI, but it may also represent a contaminant or colonization of the catheter. Colonization or contamination should be presumed when only one of several blood culture bottles is positive for normal skin flora, especially if the culture takes more than 24 h to turn positive. When a blood culture becomes positive for a gram-positive organism, it is important to repeat blood cultures, ideally from both central and peripheral sites, to determine whether the blood culture isolate is recovered from more than one bottle and at different time points. Skin flora growing from blood drawn from the catheter with negative peripheral blood cultures can reflect catheter colonization (i.e., not a true bloodstream infection) or CLABSI with low levels of bacteria in blood resulting in negative peripheral cultures. When both the peripheral and the central blood cultures are positive for the same organism, bacteremia is confirmed, and the differential time to positivity may allow establishing the diagnosis of catheter-related bacteremia if the central line culture grows at least 2 h earlier than the peripheral. However, the practice of always obtaining a peripheral blood culture may not be practical on an oncology ward, where patients may be febrile daily and are frequently thrombocytopenic.

More important than whether blood cultures are collected from the central catheter alone or from central and peripheral venous sources is ensuring that an adequate volume of blood (as determined by the specific blood culture platform) is collected. The optimal frequency of blood cultures in NF has not been established. Similarly, the use of surveillance blood cultures in neutropenic patients has not been adequately studied. The amplification by PCR of plasma cell–free DNA, with its superior yield compared with blood cultures, is still investigational, and its place in the routine management of NF remains to be defined.

If CLABSI is diagnosed, the preferred treatment is to remove the catheter. Again, this may not always be practical in the case of thrombocytopenic patients with limited access, and an attempt to salvage a permanent catheter (i.e., tunneled catheter or port) may be appropriate as long as the patient remains hemodynamically stable and the blood cultures become negative upon initiating appropriate antibiotics. Antibiotic lock therapy should be considered, if feasible. The likelihood of success when a catheter is colonized with *S. aureus*, mycobacteria, or *Candida* species is low, and these catheters should be removed as soon as possible. It is customary to provide some time without central access (e.g., a 48-h "line holiday") before replacing the central venous catheter, although the evidence supporting this practice is scant. Patients may require continuous central IV access, and in this case it is reasonable to use nonsurgically placed lines (e.g., a peripherally inserted central catheter) and to place a new surgically implanted port (e.g., mediport) only when there is clear evidence of blood culture clearance.

Endocarditis is rare during neutropenia, but it should be suspected with persistently positive blood cultures or with clinical findings like a new murmur or embolic phenomena. Routine echocardiogram in every case of bacteremia in neutropenic patients is not recommended, but patients with *S. aureus* bacteremia do require one.

RESPIRATORY INFECTIONS

Sinusitis Sinusitis may manifest by pressure, facial pain, rhinorrhea, and postnasal drip but occasionally has very mild symptoms that will be elicited only by targeted questioning during the physical exam. Symptomatic sinusitis in neutropenic patients may be caused by pathogens relatively uncommon in immunocompetent people, including *S. aureus*, *P. aeruginosa*, and other gram-negative bacilli. Evaluation by CT and consultation by otolaryngology to examine the integrity of the mucosa and obtain samples for culture is recommended. Invasive fungal sinusitis may be caused by *Aspergillus* species, agents of mucormycosis, *Fusarium* species, and relatively less virulent dematiaceous molds like *Alternaria*, *Bipolaris*, or *Curvularia*. CT findings suggestive of fungal etiology include marked asymmetry, heterogeneous radiodensity of the contents of the sinuses, and, later in the disease, bony erosions. The endoscopic exam may show pale, devitalized mucosa and, sometimes, necrotic ulceration of a turbinate caused by the angioinvasive process. Invasive fungal sinusitis during neutropenia mandates surgical resection. Repeated visits to the operating room in conjunction with optimal antifungal therapy are necessary to save the patient's life. The empirical antifungal agent may vary depending on the preexisting antifungal prophylaxis, but, given the possibility of mucormycosis and fusariosis, many experts would recommend a lipid formulation of amphotericin B, often considered the antifungal of choice for mucormycosis, together with posaconazole, which is noninferior to voriconazole for *Aspergillus* species and has superior activity in vitro against dematiaceous molds.

Pneumonia Pulmonary infiltrates are a common diagnostic challenge in immunocompromised hosts. A systematic approach to the differential, which includes infections and noninfectious etiologies, is mandatory. Some important noninfectious conditions to be considered include heart failure, fluid overload, transfusion reactions like transfusion-associated circulatory overload (TACO) and transfusion-associated lung injury (TRALI), chemotherapy-associated pneumonitis, organizing pneumonia, and diffuse alveolar hemorrhage, as well as dissemination of the cancer itself. The radiologic appearance is useful to separate unilateral focal processes (suggestive of bacterial or mold infection) from multifocal or diffuse infections, which may be caused in addition by viruses, atypical bacteria, and PJP. Single or multiple dense, well-circumscribed nodules suggest IFD in high-risk patients (e.g., those with leukemia and allogeneic stem cell transplant recipients), but *Nocardia* species may have the same appearance and other bacteria like *P. aeruginosa* and *Stenotrophomonas maltophilia* may cause angioinvasive infection and result in similar radiology. Unfortunately, characteristic radiologic differences may not hold true in an individual case, and every attempt should be made to obtain a respiratory sample for microbiologic diagnosis.

Early bronchoscopy with bronchoalveolar lavage (BAL) in neutropenic patients with pneumonia is recommended. The yield of BAL is highest for diffuse and multifocal processes. In the case of pulmonary nodules accessible to percutaneous or endobronchial biopsy, the choice of diagnostic procedure will often depend on institutional expertise, clinical status, and comorbidities. The empirical antimicrobial regimen may vary slightly depending on the radiology and the individual risk factors but should always provide optimal antibacterial coverage by combining one of the antipseudomonal β-lactams together with vancomycin and an agent active against *Legionella*, like azithromycin, a fluoroquinolone, or doxycycline. If the patient has risk factors for *S. maltophilia* (high prevalence in the institution, previous exposure to carbapenems), the addition of TMP-SMX or levofloxacin should be considered.

The empirical addition of TMP-SMX for PJP depends on the clinical scenario and risk factors. If the only immune defect of the patient is neutropenia, the main concerns are bacteria and molds, the likelihood of PJP is low, and empirical anti-*Pneumocystis* coverage seems unnecessary. However, patients with hematologic malignancies often have received corticosteroids or other agents that decrease cellular immunity and put them at risk for PJP. In this case, empirical coverage until PJP has been ruled out is reasonable. The current gold standard for the diagnosis of PJP is PCR of respiratory secretions, which occasionally may detect patients who are merely colonized. Serum β-D-glucan is typically elevated in patients with PJP, and a positive result increases the likelihood of PJP in a patient with appropriate risk factors and radiologic findings. However, false-positive and -negative β-D-glucan results occur, and this test is not specific for PJP; therefore, a positive serum β-D-glucan does not always obviate the need for BAL.

Invasive fungal disease (IFD) is suggested by single or multiple dense, well-circumscribed nodules, particularly larger than 2 cm. The halo sign (ground glass opacity surrounding a dense nodule) or the reverse halo sign (ground glass opacity seen inside a nodule, sometimes said to be more common in mucormycosis) may be present, but these are not diagnostic. Peripheral wedge-shaped infiltrates also may be seen. In patients at high risk for IFD, initiation or escalation of antifungal therapy is commonly begun prior to establishing a diagnosis. Such decisions must consider the antifungal prophylaxis that the patient is receiving. If the patient is on fluconazole or no antifungal prophylaxis, a mold-active azole (e.g., voriconazole, posaconazole, isavuconazole) as empirical therapy is reasonable pending a diagnosis, since *Aspergillus* is by far the most likely mold infection. If on the other hand, a suspected fungal infection occurs while receiving a mold-active azole, many experts would recommend switching the regimen to liposomal amphotericin B, while aggressively pursuing a diagnosis.

Evaluation of suspected pneumonia involves sputum and blood cultures, nasopharyngeal swab for PCR for respiratory viruses, urine *Legionella* antigen, and in patients with neutropenia, evaluation for IFD (e.g., serum galactomannan, beta-glucan) should be included. If these studies are negative and the empirical antimicrobial regimen does not lead to improvement, early BAL or image-guided percutaneous lung biopsy should be considered. If non-invasive diagnostic studies identify a pathogen, it is important to consider the potential for concurrent or secondary infection by other pathogens. For example, respiratory viral infections can be complicated by secondary bacterial pneumonias, and cases of invasive pulmonary aspergillosis following COVID-19 have been documented. It is therefore important to consider a broad differential diagnosis and to repeat imaging and diagnostic evaluation in patients with worsening pneumonia even when an initial diagnosis has been made.

INTRA-ABDOMINAL INFECTIONS

Neutropenic Enterocolitis (Typhlitis) Sometimes bacteria invade the wall of the intestine and proliferate there, causing inflammation and even necrosis of segments of the bowel. The clinical syndrome includes fever, abdominal pain, and tenderness and diarrhea, which is occasionally bloody. The term *typhlitis* (inflammation of the cecum) is sometimes used, as the cecum is most commonly involved, but the terminal ileum, ascending colon, and other segments of the intestine also may be affected, so the more general term *neutropenic enterocolitis* is preferred. Blood cultures are frequently positive (with bowel flora), but not always. Plain films of the abdomen are not sensitive or specific, but a right lower quadrant soft tissue density, distended bowel loops, ileus, or "thumbprinting" suggesting mucosal edema may sometimes be seen. A CT scan of the abdomen and pelvis usually demonstrates a right lower quadrant inflammatory mass, with thickened bowel wall and stranding of the surrounding fat. Conservative management with broad-spectrum antibiotics, bowel rest, bowel decompression, and nutritional support is usually preferred. Surgery may be required if the patient deteriorates and intestinal perforation, peritonitis, and hemodynamic instability ensue. Carbapenems and piperacillin-tazobactam are appropriate given their broader spectrum. If ceftazidime or cefepime is used, additional agents against anaerobic bacteria (e.g., metronidazole) must be added. In selecting an empirical regimen, it is important to note prior infections with antibiotic-resistant infections, such as VRE, and extended-spectrum β-lactamase (ESBL) of carbapenem-resistant (CRE) gram-negative infections.

Similar to intra-abdominal infections, perirectal and perianal infections are caused by gastrointestinal flora in the context of neutropenia and mucosal injury. CT imaging is helpful to assess the extent of infection. These infections typically respond to broad-spectrum antibiotic therapy without the need for surgery, but follow-up imaging after neutrophil recovery may be necessary if symptoms or fever persist.

***Clostridioides difficile* (*C. difficile*)–Associated Disease** Cancer patients are at increased risk of *C. difficile*–associated disease (CDAD) because of their frequent exposure to healthcare facilities, antibiotics, and chemotherapeutic drugs. Patients with a history of CDAD in the previous year are often treated prophylactically with oral vancomycin (125 mg daily) when empirical antibiotic therapy for neutropenic fever is begun. Every hospitalized neutropenic patient with diarrhea should be tested for *C. difficile*. It is unclear whether neutropenic patients with CDAD have worse outcomes, higher risk of complications, or increased frequency of neutropenic enterocolitis. First-line treatment is fidaxomicin, with oral vancomycin as an alternative. In case of ileus or overwhelming CDAD, combination of either oral fidaxomicin or vancomycin with IV metronidazole (which is secreted into the intestinal lumen) is recommended.

Cholecystitis Cholecystitis is a rare infection during neutropenia, and most cases have been seen during treatment of acute leukemia. The presentation is like that in nonneutropenic individuals, but bacteremia is more common. At least half of the reported cases have been acalculous cholecystitis. Conservative management, frequently including cholecystostomy, is used until resolution of neutropenia when cholecystectomy can be performed more safely.

■ CENTRAL NERVOUS SYSTEM INFECTIONS

Bacterial central nervous system (CNS) infections are uncommon during neutropenia and usually secondary to an episode of bacteremia (which may or may not have been detected) or to extension from the paranasal sinuses. Meningitis caused by gram-negative bacilli, including *P. aeruginosa*, may occur in this setting. Two unexpected causes of CNS infections are worth mentioning. *Rothia mucilaginosa*, a gram-positive coccus that is part of the oral flora, may seed the meninges during a transient bacteremia that seemed to be controlled easily with appropriate antibiotic treatment. This is a rare infection, and most cases treated successfully have received meropenem and vancomycin. Similarly, *Bacillus cereus*, a gram-positive bacillus ubiquitous in the environment, has been reported as a cause of difficult-to-treat meningitis and brain abscess in neutropenic patients with acute leukemia. The combination of meropenem and vancomycin also has been used successfully. *Listeria monocytogenes* meningitis and bacteremia are most common in patients with impaired cellular immunity, including patients with cancer receiving immunosuppressive regimens. Besides meningitis, *L. monocytogenes* can also cause rhombencephalitis (inflammation of the brainstem and cerebellum) and brain abscess. The combination of ampicillin and gentamicin is recommended. In cases of penicillin allergy TMP-SMX may be used. *Listeria monocytogenes* is susceptible in vitro to meropenem and linezolid, but clinical experience is limited.

In patients with hematologic cancers and prolonged neutropenia, brain lesions should raise the concern for opportunistic pathogens, usually fungi. *Aspergillus* and *Candida* brain abscesses are usually hematogenous; mucormycosis more frequently follows extension from the sinuses. Empirical treatment should be administered including antibacterial (typically meropenem plus vancomycin) and antifungal agents until the etiology is established. As mentioned for pulmonary infections, the choice of antifungal may be conditioned by preexisting antifungal prophylaxis. Voriconazole is the preferred agent for CNS aspergillosis and liposomal amphotericin B for mucormycosis. Echinocandins do not achieve therapeutic levels in the brain or cerebrospinal fluid (CSF), and they should not be relied upon for the treatment of CNS infections.

Nocardia species and *Toxoplasma gondii* are important causes of brain abscess in immunocompromised patients. The major risk factor for these infections is suppressed T cell immunity. In patients with cancer, prolonged intensive corticosteroid therapy and use of purine analogues are examples of risk factors. TMP-SMX used as prophylaxis for PJP may also confer some protection against *Listeria*, *Nocardia*, and *Toxoplasma*.

Patients with cancer are at increased risk of viral encephalitis. Herpes simplex virus (HSV) encephalitis is characterized by fever and decreased level of consciousness. Imaging shows unilateral involvement of the temporal lobe. Abnormalities in the CSF may be altered by radiation therapy or corticosteroids. The diagnosis is by PCR of the CSF, which may be falsely negative early in the course of the disease. Herpes zoster encephalitis is a rare but devastating complication in patients with defective T cell immunity. Human herpesvirus 6 (HHV-6) encephalitis is rare outside of allogeneic stem cell transplantation (allo-HCT) (see **Chap. 148**). Finding a high level of HHV-6 in blood in any other setting brings up the possibility of HHV-6 chromosomal integration, which is present in around 1% of the population.

In patients who have had neurosurgery, including those with neurosurgical devices (e.g., shunts or reservoirs), infections by gram-positive bacteria are most common, although gram-negative infections including coliforms and *P. aeruginosa* can occur. Risk factors include multiple neurosurgeries (e.g., for recurrent tumor), cranial irradiation, antiangiogenics (which disrupt blood supply), and use of immunosuppressives. Vancomycin plus an antipseudomonal cephalosporin (e.g., ceftazidime) is an appropriate initial regimen for postneurosurgical CNS infections, pending culture data. Surgical drainage and debridement of infected material and removal of hardware are usually required.

■ SKIN AND SOFT TISSUE INFECTIONS

As a rule, focal skin lesions in neutropenic, febrile patients should be biopsied, as the diagnostic possibilities are multiple, and some of the causes may be life-threatening and require immediate treatment. Cellulitis may be caused by streptococci or *S. aureus*, but also by gram-negative bacilli, and broad-spectrum coverage with an antipseudomonal β-lactam and vancomycin should be administered until an etiologic diagnosis is established. Multiple foci of erythematous tender plaques with underlying fasciitis and myositis are associated with bacteremia by *Clostridium septicum*, sometimes seen in patients with colorectal cancer. Ecthyma gangrenosum appears as a tender, erythematous papule that then becomes necrotic and ulcerated. It may be caused by local inoculation or by hematogenous seeding, typically of *P. aeruginosa* but sometimes other gram-negative bacteria or even molds like *Fusarium* species. Single or multiple lesions may be seen. Disseminated candidiasis with hematogenous cutaneous disease manifests with fever and other signs of systemic infection and multiple raised cutaneous papules. Hemorrhagic bullae are described in cellulitis caused by gram-negative bacteria (including *Vibrio vulnificus*, which may be suspected with a history of liver dysfunction and exposure to shellfish).

Many noninfectious processes can cause skin lesions, but the presence of fever is unusual except in Sweet syndrome, also known as acute neutrophilic dermatosis **(Fig. 79-2)**. This is characterized by fever and a variety of skin lesions including papules, nodules, plaques, and sometimes blisters. Skin lesions are often tender and may ulcerate. Sweet syndrome can have many causes, including underlying cancer, usually hematologic. Pathergy is characteristic, and lesions may appear at the insertion site of an intravascular catheter, mimicking infection (Fig. 79-2). Leukemia cutis may manifest as a variety of nontender papules, nodules, or plaques. Fever is unusual. Chemotherapeutic agents can cause a variety of skin reactions, including hyperpigmentation, hand-foot syndrome (erythema, edema, and blistering of palms and soles), and several rashes. Definitive diagnosis often requires biopsy of skin lesions with appropriate cultures and histopathology.

■ URINARY TRACT INFECTIONS

It is common to obtain urine cultures in patients with NF regardless of urinary symptoms, and sometimes bacteriuria is detected. The question of whether it is just asymptomatic bacteriuria (common and increasing with age in both men and women), unrelated to the fever, or reflects true urinary tract infection (UTI) necessitating specific treatment may not be immediately answerable. It is appropriate to treat bacteriuria in patients with NF, and standard empirical antibiotic regimens used for NF will cover most urinary pathogens. If UTI is believed to be the source of NF, CT imaging should be considered to

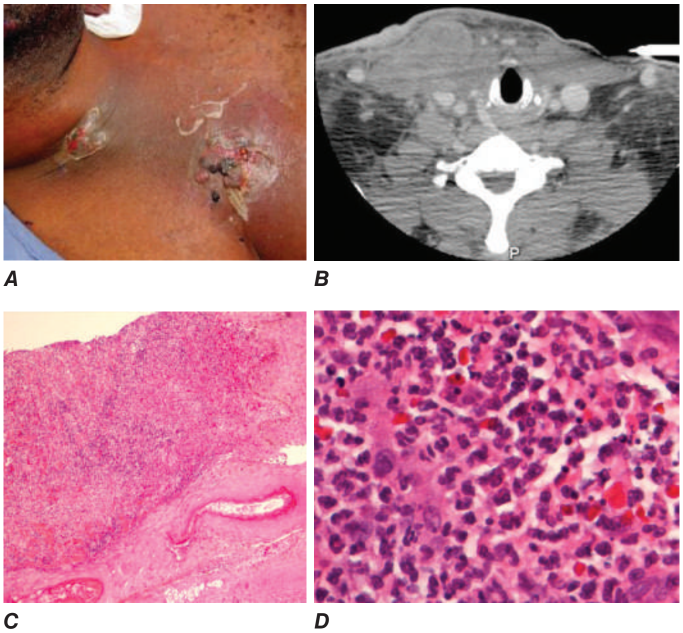

FIGURE 79-2 Sweet syndrome. A 47-year-old man admitted for acute myelogenous leukemia (AML) received a surgically implanted catheter and was started on idarubicin and cytosine arabinoside (Ara-C). He developed erythema and tenderness at the insertion site and the cuff site after 24 h, followed by fever. The catheter was removed and cultured (negative) and broad-spectrum antibiotics started, without effect. The catheter exit site worsened, with bullae formation (***A***). New skin nodules developed on his left thigh, both cheeks, and scalp. Tender induration of the sternocleidomastoid muscle was noticeable clinically and by CT (***B***). Skin biopsy showed a dense infiltrate of mature neutrophils consistent with Sweet syndrome (***C***, ***D***). The lesions and the fever resolved promptly with oral prednisone.

evaluate for complicated UTI with potential etiologies including genitourinary tract obstruction by tumor or kidney stones, pyelonephritis, and prostatitis.

PREVENTION OF INFECTIONS IN PATIENTS WITH CANCER

Prevention measures to avoid infections should be tailored to the immune status of the patient and consequent risk of infections. The most stringent measures apply to patients with hematologic malignancies with prolonged neutropenia, to stem cell transplant and adoptive cellular therapy recipients, and to other patients who may have prolonged neutropenia for other reasons (e.g., aplastic anemia) or are receiving intensive systemic corticosteroid therapy. At the other end of the spectrum are patients with solid tumors who are in remission and not receiving active chemotherapy or only receiving hormonal agents, such as antiestrogen therapy for breast cancer. These patients should adhere to guidelines to prevent infection that apply to the general population such as hand hygiene, food safety, and guideline-based vaccines. This section is focused on prevention measures for patients with cancer at high risk for infections including opportunistic infection.

ENVIRONMENTAL PRECAUTIONS

Inpatient leukemia and transplant wards are typically engineered with HEPA filters and appropriate air exchanges, and individual patient rooms are under positive pressure relative to the outside. The principal rationale for these precautions is to reduce mold spore exposure; these structural precautions were in part driven by outbreaks of aspergillosis. For similar reasons, houseplants are restricted from these wards. Construction, which can release mold spores, must be performed under stringent infection control precautions to minimize patient exposure. In addition to limiting mold exposure, these precautions may have additional benefits to limit airborne transmission of other infections.

The vast majority of chemotherapy regimens are administered to outpatients, including to patients with hematologic malignancies receiving intensive immunosuppressive regimens. The stringent environmental precautions used for inpatients are not feasible in the outpatient setting or in the home. Patients should not have construction done in their homes and should avoid exposure to construction sites during periods of significant immune compromise. Patients should also not be involved in gardening, mulching, or similar activities that result in spore exposure.

PROTECTION FROM FOOD-BORNE INFECTIONS

The old practice of implementing "low-microbial diets" for neutropenic patients did not show convincing benefit in well-designed studies and carries the disadvantages of limiting nutrition and affecting quality of life. Patients should be educated about food and water safety precautions issued by the Centers for Disease Control and Prevention (*www.cdc.gov/food-safety/foods/weakened-immune-systems.html*) and other authoritative bodies that are tailored to immunocompromised persons but are less restrictive than low-microbial diets. These guidelines stress hand hygiene, washing vegetables and fruits, not consuming raw or undercooked meat, poultry, or fish or unpasteurized dairy products, avoidance of waterborne infections, and awareness of food-borne outbreaks. Travel to regions where food and water safety is not reliable should be avoided.

Animals also are a potential source of infection, particularly via the fecal–oral route. Patients with cancer should avoid exposure to farm animals and wild animals (e.g., hunting and butchering). Acquiring a new pet while immunocompromised is not advisable. Existing pets can continue to live in the patient's house, but the patient should avoid direct exposure to animal waste, such as cleaning a cat litter box or bird cage. Patients should avoid dogs recently vaccinated for kennel cough, if possible. Contact with more exotic pets (e.g., reptiles) should be avoided because they can harbor more unusual pathogens (e.g., *Salmonella* species).

RESPIRATORY VIRAL INFECTIONS

Patients with cancer should avoid contact with persons with symptoms or signs of respiratory infection to the extent feasible. Risk of viral infection is increased in congregate settings that are indoors and involve large numbers of people in close proximity. Precautions to limit infection spread, such as choosing outdoor over indoor events and avoiding crowded settings, are advisable. For example, a dinner at a restaurant in which the table is limited to a few persons carries less risk of transmissible infections than a crowded buffet-style setting.

Patients with cancer should be aware of local patterns of viral infection. This point was amply demonstrated during the COVID-19 pandemic and applies to other viral infections, such as influenza and respiratory syncytial virus (RSV). If local viral outbreaks occur, patients with cancer should use additional precautions to limit exposure. In addition to avoiding crowds, masking may provide an additional layer of protection, especially in settings of high levels of community spread of respiratory viral infections.

SEXUAL INTIMACY

To avoid risk of sexually transmitted infections, it is recommended that sexual intercourse be in a monogamous relationship. During neutropenia, sexual intercourse and oral–genital or rectal stimulation may cause injury and predispose to infection. Once the ANC recovers, sexual activity usually can be resumed.

ANTIMICROBIAL PROPHYLAXIS

Antimicrobial prophylaxis should be tailored to infection risk. Most patients with solid tumors require no antimicrobial prophylaxis. In addition to side effects of antibiotics, increased risk of *C. difficile*, and selection of antibiotic-resistant flora, there is precedent for antibiotics diminishing the efficacy of immune checkpoint inhibitors by disruption of the microbiome.

As discussed above, prophylaxis with a quinolone should be considered in adults with prolonged neutropenia (e.g., ANC <500/μL for at least 7 days). Acyclovir prophylaxis to prevent HSV and VZV reactivation is advised in the majority of patients with hematologic malignancies on active therapy. In patients with AML receiving induction or reinduction chemotherapy, prophylaxis with posaconazole should be included. Mold-active prophylaxis is also recommended in patients with significant GVHD. PJP prophylaxis is recommended for patients with severe T cell impairment (TMP-SMX is preferred). Examples include standard induction chemotherapy for acute lymphoblastic leukemia that involves systemic corticosteroids, prolonged corticosteroids (described above), use of purine analogues (e.g., cladribine-containing regimens), and GVHD. Pneumococcal prophylaxis (e.g., with penicillin) should be considered for persons with asplenia or chronic GVHD (described above).

■ VACCINATION

Patients with cancer should receive all recommended non-live vaccines based on their approved indications. Yearly recommendations regarding COVID-19 and influenza vaccines should be followed. Patients with solid tumors in general mount greater antibody titers than patients with hematologic malignancies, especially if they have received B cell–depleting agents (e.g., rituximab, BTK inhibitors). However, vaccination is still recommended for such patients with the rationale that they may derive some protective benefit even if suboptimal. We also stress vaccination of household members, caretakers, and other close contacts to prevent them from viral infection and transmission to the patient.

Live vaccines should not be given to immunocompromised persons. However, household members including children should receive age-appropriate vaccines, including live vaccines such as MMR and varicella. If the child develops a rash following MMR or varicella vaccination, contact with the cancer patient should be limited until the rash resolves.

■ FURTHER READING

Kamboj M et al: Vaccination of adults with cancer: ASCO guideline. J Clin Oncol 42:1699, 2024.

Lehrnbecher T et al: Guideline for the management of fever and neutropenia in pediatric patients with cancer and hematopoietic cell transplantation recipients: 2023 update. J Clin Oncol 41:1774, 2023.

Maertens J et al: Empiric vs preemptive antifungal strategy in high-risk neutropenic patients on fluconazole prophylaxis: A randomized trial of the European Organization for Research and Treatment of Cancer. Clin Infect Dis 76:674, 2023.

Martinez-Nadal G et al: Inappropriate empirical antibiotic treatment in high-risk neutropenic patients with bacteremia in the era of multidrug resistance. Clin Infect Dis 70:1068, 2020.

Maschmeyer G et al: Infectious complications of targeted drugs and biotherapies in acute leukemia. Clinical practice guidelines by the European Conference on Infections in Leukemia (ECIL), a joint venture of the European Group for Blood and Marrow Transplantation (EBMT), the European Organization for Research and Treatment of Cancer (EORTC), the International Immunocompromised Host Society (ICHS) and the European Leukemia Net (ELN). Leukemia 36: 1215, 2022.

Stern A et al: Early discontinuation of antibiotics for febrile neutropenia versus continuation until neutropenia resolution in people with cancer. Cochrane Database Syst Rev 1:CD012184, 2019.

Taplitz RA et al: Antimicrobial prophylaxis for adult patients with cancer-related immunosuppression: ASCO and IDSA clinical practice guideline update. J Clin Oncol 36:3043, 2018.

Zimmer AJ et al: Bloodstream infections in hematologic malignancy patients with fever and neutropenia: Are empirical antibiotic therapies in the United States still effective? Open Forum Infect Dis 9:ofac240, 2022.

■ WEBSITE

NCCN Guidelines. Prevention and Treatment of Cancer-Related Infections. Version 2.2023. *https://www.nccn.org/guidelines/guidelines-detail?category=3&id=1457*

80

Oncologic Emergencies

Rasim Gucalp, Janice P. Dutcher

Emergencies in patients with cancer may be classified into three groups: pressure or obstruction caused by a space-occupying lesion, metabolic or hormonal problems (paraneoplastic syndromes, **Chap. 98**), and treatment-related complications.

STRUCTURAL-OBSTRUCTIVE ONCOLOGIC EMERGENCIES

■ SUPERIOR VENA CAVA SYNDROME

Superior vena cava syndrome (SVCS) is the clinical manifestation of superior vena cava (SVC) obstruction, with severe reduction in venous return from the head, neck, and upper extremities. Malignant tumors, such as lung cancer, lymphoma, and metastatic tumors, are responsible for the majority of SVCS cases. With the expanding use of intravascular devices (e.g., permanent central venous access catheters, pacemaker/defibrillator leads), the prevalence of benign causes of SVCS is now increasing, accounting for at least 40% of cases. Lung cancer, particularly of small-cell and squamous cell histologies, accounts for ~85% of all cases of malignant origin. In young adults, malignant lymphoma is a leading cause of SVCS. Hodgkin's lymphoma involves the mediastinum more commonly than other lymphomas but rarely causes SVCS. When SVCS is noted in a young man with a mediastinal mass, the differential diagnosis is lymphoma versus primary mediastinal germ cell tumor. Metastatic cancers to the mediastinal lymph nodes, such as testicular and breast carcinomas, account for a small proportion of cases. Other causes include benign tumors, aortic aneurysm, thyromegaly, thrombosis, and fibrosing mediastinitis from prior irradiation, histoplasmosis, or Behçet's syndrome. SVCS as the initial manifestation of Behçet's syndrome may be due to inflammation of the SVC associated with thrombosis.

Patients with SVCS usually present with neck and facial swelling (especially around the eyes), dyspnea, and cough. Other symptoms include hoarseness, tongue swelling, headaches, nasal congestion, epistaxis, hemoptysis, dysphagia, pain, dizziness, syncope, and lethargy. Bending forward or lying down may aggravate the symptoms. The characteristic physical findings are dilated neck veins; an increased number of collateral veins covering the anterior chest wall; cyanosis; and edema of the face, arms, and chest. Facial swelling and plethora are typically exacerbated when the patient is supine. More severe cases include proptosis, glossal and laryngeal edema, and obtundation. The clinical picture is milder if the obstruction is located above the azygos vein. Symptoms are usually progressive, but in some cases, they may improve as collateral circulation develops.

Signs and symptoms of cerebral and/or laryngeal edema, though rare, are associated with a poorer prognosis and require urgent evaluation. Seizures are more likely related to brain metastases than to cerebral edema from venous occlusion. Patients with small-cell lung cancer and SVCS have a higher incidence of brain metastases than those without SVCS.

Cardiorespiratory symptoms at rest, particularly with positional changes, suggest significant airway and vascular obstruction and limited physiologic reserve. Cardiac arrest or respiratory failure can occur, particularly in patients receiving sedatives or undergoing general anesthesia.

Rarely, esophageal varices may develop, particularly in the setting of SVC syndrome due to hemodialysis catheter. These are "downhill" varices based on the direction of blood flow from cephalad to caudad (in contrast to "uphill" varices associated with caudad to cephalad flow from portal hypertension). Variceal bleeding may be a late complication of chronic SVCS.

SVC obstruction may lead to bilateral breast edema with bilateral enlarged breasts. Unilateral breast dilation may be seen as a consequence of axillary or subclavian vein blockage.

The diagnosis of SVCS is a clinical one. The most significant chest radiographic finding is widening of the superior mediastinum, most commonly on the right side. Pleural effusion occurs in only 25% of patients, often on the right side. The majority of these effusions are exudative and occasionally chylous. However, a normal chest radiograph is still compatible with the diagnosis if other characteristic findings are present. Computed tomography (CT) provides the most reliable view of the mediastinal anatomy. The diagnosis of SVCS requires diminished or absent opacification of central venous structures with prominent collateral venous circulation. The focal hepatic hotspot on CT scan ("hot quadrate") sign suggests SVC obstruction, and it is caused by portosystemic venous shunting between the SVC and portal vein within liver. Magnetic resonance imaging (MRI) is increasingly being used to diagnose SVC obstruction with a 100% sensitivity and specificity, but dyspneic SVCS patients may have difficulty remaining supine for the entire imaging process. Invasive procedures, including bronchoscopy, percutaneous needle biopsy, mediastinoscopy, and even thoracotomy, can be performed by a skilled clinician without any major risk of bleeding. Endobronchial or esophageal ultrasound-guided needle aspiration may establish the diagnosis safely. For patients with a known cancer, a detailed workup usually is not necessary, and appropriate treatment may be started after obtaining a CT scan of the thorax. For those with no history of malignancy, a detailed evaluation is essential to rule out benign causes and determine a specific diagnosis to direct the appropriate therapy.

TREATMENT

Superior Vena Cava Syndrome

The one potentially life-threatening complication of a superior mediastinal mass is tracheal obstruction. Upper airway obstruction demands emergent therapy. Diuretics with a low-salt diet, head elevation, and oxygen may produce temporary symptomatic relief. Glucocorticoids have a limited role except in the setting of mediastinal lymphoma masses.

Radiation therapy is the primary treatment for SVCS caused by non-small-cell lung cancer and other metastatic solid tumors. Chemotherapy is effective when the underlying cancer is small-cell carcinoma of the lung, lymphoma, or germ cell tumor. SVCS recurs in 10–30% of patients; it may be palliated with the use of intravascular self-expanding stents **(Fig. 80-1)**. Endovascular therapy is more frequently used first, to provide rapid relief of clinical symptoms with reduced complications. Early stenting may be necessary in patients with severe symptoms, particularly cerebral or laryngeal edema or postural hypotension; however, the prompt increase in venous return after stenting may precipitate heart failure and pulmonary edema. Other complications of stent placement include hematoma at the insertion site, SVC perforation, stent migration in the right ventricle, stent fracture, and pulmonary embolism. Surgery may play role in treatment of SVCS secondary to nonmalignant mediastinal fibrosis.

Clinical improvement occurs in most patients, although this improvement may be due to the development of adequate collateral circulation. The mortality associated with SVCS does not relate to caval obstruction but rather to the underlying cause.

SVCS AND CENTRAL VENOUS CATHETERS IN ADULTS

The use of long-term central venous catheters has become common practice in patients with cancer. Major vessel thrombosis may occur. In these cases, catheter removal should be combined with anticoagulation to prevent embolization. SVCS in this setting, if detected early, can be treated by fibrinolytic therapy without sacrificing the catheter. When managing patients with transvenous lead-related SVC syndrome, anticoagulation, local and systemic thrombolytic therapy, and surgical intervention can be effective therapy in select patients. Endovascular stenting has also been shown to be safe and promising, with minimal procedural or clinical complications. The role of anticoagulation after SVC stent placement is controversial.

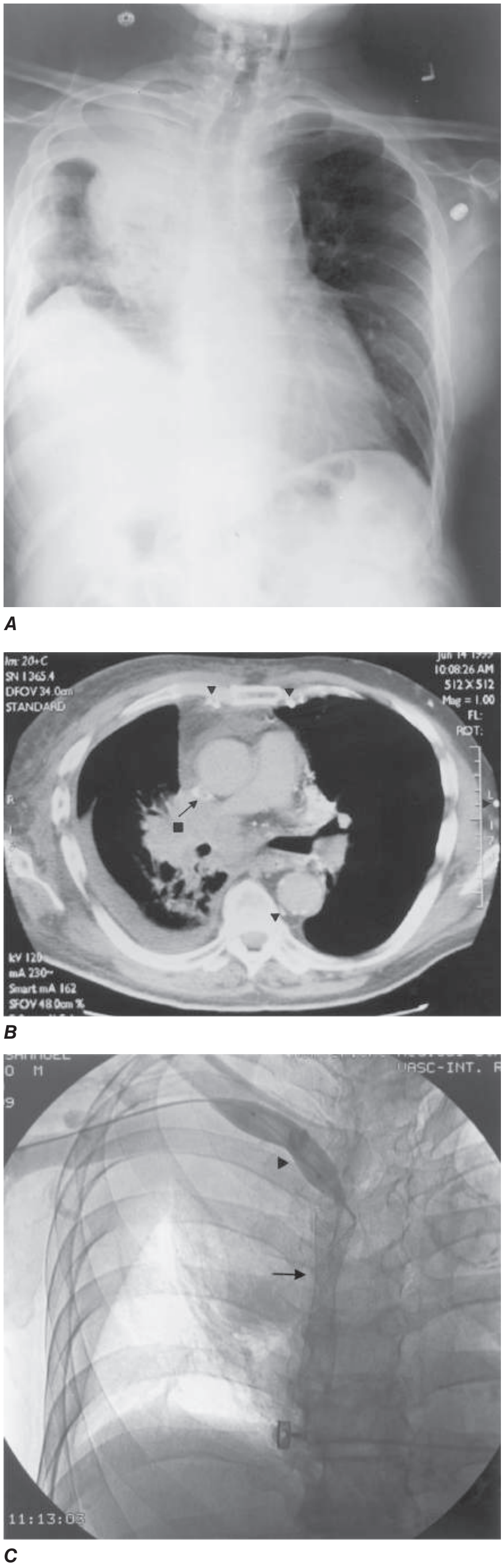

FIGURE 80-1 Superior vena cava syndrome (SVCS). ***A.*** Chest radiographs of a 59-year-old man with recurrent SVCS caused by non-small-cell lung cancer showing right paratracheal mass with right pleural effusion. ***B.*** Computed tomography of same patient demonstrating obstruction of the superior vena cava with thrombosis (*arrow*) by the lung cancer (*square*) and collaterals (*arrowheads*). ***C.*** Balloon angioplasty (*arrowhead*) with Wallstent (*arrow*) in same patient.

PERICARDIAL EFFUSION/TAMPONADE

Malignant pericardial disease is found at autopsy in 5–10% of patients with cancer, most frequently with lung cancer, breast cancer, leukemias, and lymphomas. Cardiac tamponade as the initial presentation of extrathoracic malignancy is rare. The origin is not malignancy in ~50% of cancer patients with symptomatic pericardial disease, but it can be related to irradiation; drug-induced pericarditis, including chemotherapeutic agents such as all-trans retinoic acid, arsenic trioxide, imatinib, and other abl kinase inhibitors; hypothyroidism; idiopathic pericarditis; infection; or autoimmune diseases. Pericardial disease has been associated with immune checkpoint inhibitors specifically in patients with advanced non-small-cell lung cancer. Two types of radiation pericarditis occur: an acute inflammatory, effusive pericarditis occurring within months of irradiation, which usually resolves spontaneously, and a chronic effusive pericarditis that may appear up to 20 years after radiation therapy and is accompanied by a thickened pericardium.

Most patients with pericardial metastasis are asymptomatic. However, the common symptoms are dyspnea, cough, chest pain, orthopnea, and weakness. Pleural effusion, sinus tachycardia, jugular venous distention, hepatomegaly, peripheral edema, and cyanosis are the most frequent physical findings. Relatively specific diagnostic findings, such as paradoxical pulse, diminished heart sounds, pulsus alternans (pulse waves alternating between those of greater and lesser amplitude with successive beats), and friction rub are less common than with nonmalignant pericardial disease. Chest radiographs and electrocardiogram (ECG) reveal abnormalities in 90% of patients, but half of these abnormalities are nonspecific. Echocardiography is the most helpful diagnostic test. Pericardial fluid may be serous, serosanguineous, or hemorrhagic, and cytologic examination of pericardial fluid is diagnostic in most patients. Measurements of tumor markers in the pericardial fluid are not helpful in the diagnosis of malignant pericardial fluid. Pericardioscopy with targeted pericardial and epicardial biopsy may differentiate neoplastic and benign pericardial disease. A combination of cytology, pericardial and epicardial biopsy, and guided pericardioscopy gives the best diagnostic yield. CT scan of chest may also reveal the presence of a concomitant thoracic neoplasm. Cancer patients with pericardial effusion containing malignant cells on cytology have a very poor survival.

TREATMENT

Pericardial Effusion/Tamponade

Pericardiocentesis with or without the introduction of sclerosing agents, the creation of a pericardial window, complete pericardial stripping, cardiac irradiation, and systemic chemotherapy are effective treatments. Acute pericardial tamponade with life-threatening hemodynamic instability requires immediate drainage of fluid. This can be quickly achieved by pericardiocentesis. The recurrence rate after percutaneous catheter drainage is ~20%. Sclerotherapy (pericardial instillation of bleomycin, mitomycin C, or tetracycline) may decrease recurrences. Alternatively, subxiphoid pericardiotomy can be performed in 45 min under local anesthesia. Thoracoscopic pericardial fenestration can be employed for benign causes; however, 60% of malignant pericardial effusions recur after this procedure. In a subset of patients, drainage of the pericardial effusion is paradoxically followed by worsening hemodynamic instability. This so-called "postoperative low cardiac output syndrome" occurs in up to 10% of patients undergoing surgical drainage and carries poor short-term survival.

INTESTINAL OBSTRUCTION

Intestinal obstruction and reobstruction are common problems in patients with advanced cancer, particularly colorectal or ovarian carcinoma. However, other cancers, such as lung or breast cancer and melanoma, can metastasize within the abdomen, leading to intestinal obstruction. Metastatic disease from colorectal, ovarian, pancreatic, gastric, and occasionally breast cancer can lead to peritoneal carcinomatosis, with infiltration of the omentum and peritoneal surface, thus limiting bowel motility. Typically, obstruction occurs at multiple sites in peritoneal carcinomatosis. Melanoma has a predilection to involve the small bowel; this involvement may be isolated, and resection may result in prolonged survival. Intestinal pseudoobstruction is caused by infiltration of the mesentery or bowel muscle by tumor, involvement of the celiac plexus, or paraneoplastic neuropathy in patients with small-cell lung cancer. Paraneoplastic neuropathy is associated with IgG antibodies reactive to neurons of the myenteric and submucosal plexuses of the jejunum and stomach. Ovarian cancer can lead to authentic luminal obstruction or to pseudoobstruction that results when circumferential invasion of a bowel segment arrests the forward progression of peristaltic contractions.

The onset of obstruction is usually insidious. Pain is the most common symptom and is usually colicky in nature. Pain can also be due to abdominal distention, tumor masses, or hepatomegaly. Vomiting can be intermittent or continuous. Patients with complete obstruction usually have constipation. Physical examination may reveal abdominal distention with tympany, ascites, visible peristalsis, high-pitched bowel sounds, and tumor masses. Erect plain abdominal films may reveal multiple air-fluid levels and dilation of the small or large bowel. Acute cecal dilation to >12–14 cm is considered a surgical emergency because of the high likelihood of rupture. CT scan is useful in defining the extent of disease and the exact nature of the obstruction and differentiating benign from malignant causes of obstruction in patients who have undergone surgery for malignancy. Malignant obstruction is suggested by a mass at the site of obstruction or prior surgery, adenopathy, or an abrupt transition zone and irregular bowel thickening at the obstruction site. Benign obstruction is more likely when CT shows mesenteric vascular changes, a large volume of ascites, or a smooth transition zone and smooth bowel thickening at the obstruction site. In challenging patients with obstructive symptoms, particularly low-grade small-bowel obstruction (SBO), CT enteroclysis often can help establish the diagnosis by providing distention of small-bowel loops. In this technique, water-soluble contrast is infused through a nasoenteric tube into the duodenum or proximal small bowel followed by CT images. The prognosis for the patient with cancer who develops intestinal obstruction is poor; median survival is 3–4 months. About 25–30% of patients are found to have intestinal obstruction due to causes other than cancer. Adhesions from previous operations are a common benign cause. Ileus induced by vinca alkaloids, narcotics, or other drugs is another reversible cause.

TREATMENT

Intestinal Obstruction

The management of intestinal obstruction in patients with advanced malignancy depends on the extent of the underlying malignancy, options for further antineoplastic therapy, estimated life expectancy, the functional status of the major organs, and the extent of the obstruction. The initial management should include surgical evaluation. Operation is not always successful and may lead to further complications with a substantial mortality rate (10–20%). Laparoscopy can diagnose and treat malignant bowel obstruction in some cases. Self-expanding metal stents placed in the gastric outlet, duodenum, proximal jejunum, colon, or rectum may palliate obstructive symptoms at those sites without major surgery. Patients known to have advanced intraabdominal malignancy should receive a prolonged course of conservative management, including nasogastric decompression. Percutaneous endoscopic or surgical gastrostomy tube placement is an option for palliation of nausea and vomiting, the so-called "venting gastrostomy." Treatment with antiemetics, antispasmodics, and analgesics may allow patients to remain outside the hospital. Octreotide may relieve obstructive symptoms through its inhibitory effect on gastrointestinal secretion. Glucocorticoids have anti-inflammatory effects and may help the resolution of bowel obstruction.

URINARY OBSTRUCTION

Urinary obstruction may occur in patients with prostatic or gynecologic malignancies, particularly cervical carcinoma; metastatic disease from other primary sites such as carcinomas of the breast, stomach, lung, colon, and pancreas; or lymphomas. Radiation therapy to pelvic tumors may cause fibrosis and subsequent ureteral obstruction. Bladder outlet obstruction is usually due to prostate and cervical cancers and may lead to bilateral hydronephrosis and renal failure.

Flank pain is the most common symptom. Persistent urinary tract infection, persistent proteinuria, or hematuria in patients with cancer should raise suspicion of ureteral obstruction. Total anuria and/or anuria alternating with polyuria may occur. A slow, continuous rise in the serum creatinine level necessitates immediate evaluation. Renal ultrasound is the safest and cheapest way to identify hydronephrosis. The function of an obstructed kidney can be evaluated by a nuclear scan. CT scan can reveal the point of obstruction and identify a retroperitoneal mass or adenopathy.

TREATMENT

Urinary Obstruction

Obstruction associated with flank pain, sepsis, or fistula formation is an indication for immediate palliative urinary diversion. Internal ureteral stents can be placed under local anesthesia. Percutaneous nephrostomy offers an alternative approach for drainage. The placement of a nephrostomy is associated with a significant rate of pyelonephritis. In the case of bladder outlet obstruction due to malignancy, a suprapubic cystostomy can be used for urinary drainage. An aggressive intervention with invasive approaches to improve the obstruction should be weighed against the likelihood of antitumor response, and the ability to reverse renal insufficiency should be evaluated.

MALIGNANT BILIARY OBSTRUCTION

This common clinical problem can be caused by a primary carcinoma arising in the pancreas, ampulla of Vater, bile duct, or liver or by metastatic disease to the periductal lymph nodes or liver parenchyma. The most common metastatic tumors causing biliary obstruction are gastric, colon, breast, and lung cancers. Jaundice, light-colored stools, dark urine, pruritus, and weight loss due to malabsorption are usual symptoms. Pain and secondary infection are uncommon in malignant biliary obstruction. Ultrasound, CT scan, magnetic resonance cholangiopancreatography (MRCP), or percutaneous transhepatic or endoscopic retrograde cholangiopancreatography (ERCP) will identify the site and nature of the biliary obstruction.

TREATMENT

Malignant Biliary Obstruction

Palliative intervention is indicated only in patients with disabling pruritus resistant to medical treatment, severe malabsorption, or infection. Stenting under radiographic control, surgical bypass, or radiation therapy with or without chemotherapy may alleviate the obstruction. The choice of therapy should be based on the site of obstruction (proximal vs distal), the type of tumor (sensitive to radiotherapy, chemotherapy, or neither), and the general condition of the patient. Stenting under radiographic or endoscopic control, surgical bypass, or radiation therapy with or without chemotherapy may alleviate the obstruction. Photodynamic therapy and radiofrequency ablation are promising endoscopic therapies for malignant biliary obstruction.

Endoscopic ultrasonography-guided biliary drainage is a safe and effective method of biliary drainage in patients with malignant biliary obstruction, particularly in patients whom standard ERCP failed.

SPINAL CORD COMPRESSION

Malignant spinal cord compression (MSCC) is defined as compression of the spinal cord and/or cauda equina by an extradural tumor mass. The minimum radiologic evidence for cord compression is indentation of the theca at the level of clinical features. Spinal cord compression (SCC) occurs in 5–10% of patients with cancer. Epidural tumor is the first manifestation of malignancy in ~10% of patients. The underlying cancer is usually identified during the initial evaluation; lung cancer is the most common cause of MSCC.

Metastatic tumor involves the vertebral column more often than any other part of the bony skeleton. Lung, breast, and prostate cancers are the most frequent offenders. Multiple myeloma also has a high incidence of spine involvement. Lymphomas, melanoma, renal cell cancer, and genitourinary cancers also cause cord compression. The thoracic spine is the most common site (70%), followed by the lumbosacral spine (20%) and the cervical spine (10%). Involvement of multiple sites is most frequent in patients with breast and prostate carcinoma. Cord injury develops when metastases to the vertebral body or pedicle enlarge and compress the underlying dura. Another cause of cord compression is direct extension of a paravertebral lesion through the intervertebral foramen. These cases usually involve a lymphoma, myeloma, or pediatric neoplasm. Parenchymal spinal cord metastasis due to hematogenous spread is rare. Intramedullary metastases can be seen in lung cancer, breast cancer, renal cancer, melanoma, and lymphoma, and are frequently associated with brain metastases and leptomeningeal disease.

Expanding extradural tumors induce injury through several mechanisms. Expanding extradural tumors induce mechanical injury to axons and myelin. Compression compromises blood flow, leading to ischemia and/or infarction.

The most common initial symptom in patients with SCC is localized back pain and tenderness due to involvement of vertebrae by tumor. Pain is usually present for days or months before other neurologic findings appear. It is exacerbated by movement and by coughing or sneezing. It can be differentiated from the pain of disk disease by the fact that it worsens when the patient is supine. Radicular pain is less common than localized back pain and usually develops later. Radicular pain in the cervical or lumbosacral areas may be unilateral or bilateral. Radicular pain from the thoracic roots is often bilateral and is described by patients as a feeling of tight, band-like constriction around the thorax and abdomen. Typical cervical radicular pain radiates down the arm; in the lumbar region, the radiation is down the legs. *Lhermitte's sign*, a tingling or electric sensation down the back and upper and lower limbs upon flexing or extending the neck, may be an early sign of cord compression. Loss of bowel or bladder control may be the presenting symptom but usually occurs late in the course. Occasionally, patients present with ataxia of gait without motor and sensory involvement due to involvement of the spinocerebellar tract.

On physical examination, pain induced by straight leg raising, neck flexion, or vertebral percussion may help to determine the level of cord compression. Patients develop numbness and paresthesias in the extremities or trunk. Loss of sensibility to pinprick is as common as loss of sensibility to vibration or position. The upper limit of the zone of sensory loss is often one or two vertebrae below the site of compression. Motor findings include weakness, spasticity, and abnormal muscle stretching. An extensor plantar reflex reflects significant compression. Deep tendon reflexes may be brisk. Motor and sensory loss usually precedes sphincter disturbance. Patients with autonomic dysfunction may present with decreased anal tonus, decreased perineal sensibility, and a distended bladder. The absence of the anal wink reflex or the bulbocavernosus reflex confirms cord involvement. Autonomic dysfunction is an unfavorable prognostic factor. Patients with progressive neurologic symptoms should have frequent neurologic examinations and rapid therapeutic intervention. Other illnesses that may mimic cord compression include osteoporotic vertebral collapse, disk disease, pyogenic abscess or vertebral tuberculosis, radiation myelopathy, neoplastic leptomeningitis, benign tumors, epidural hematoma, and spinal lipomatosis.

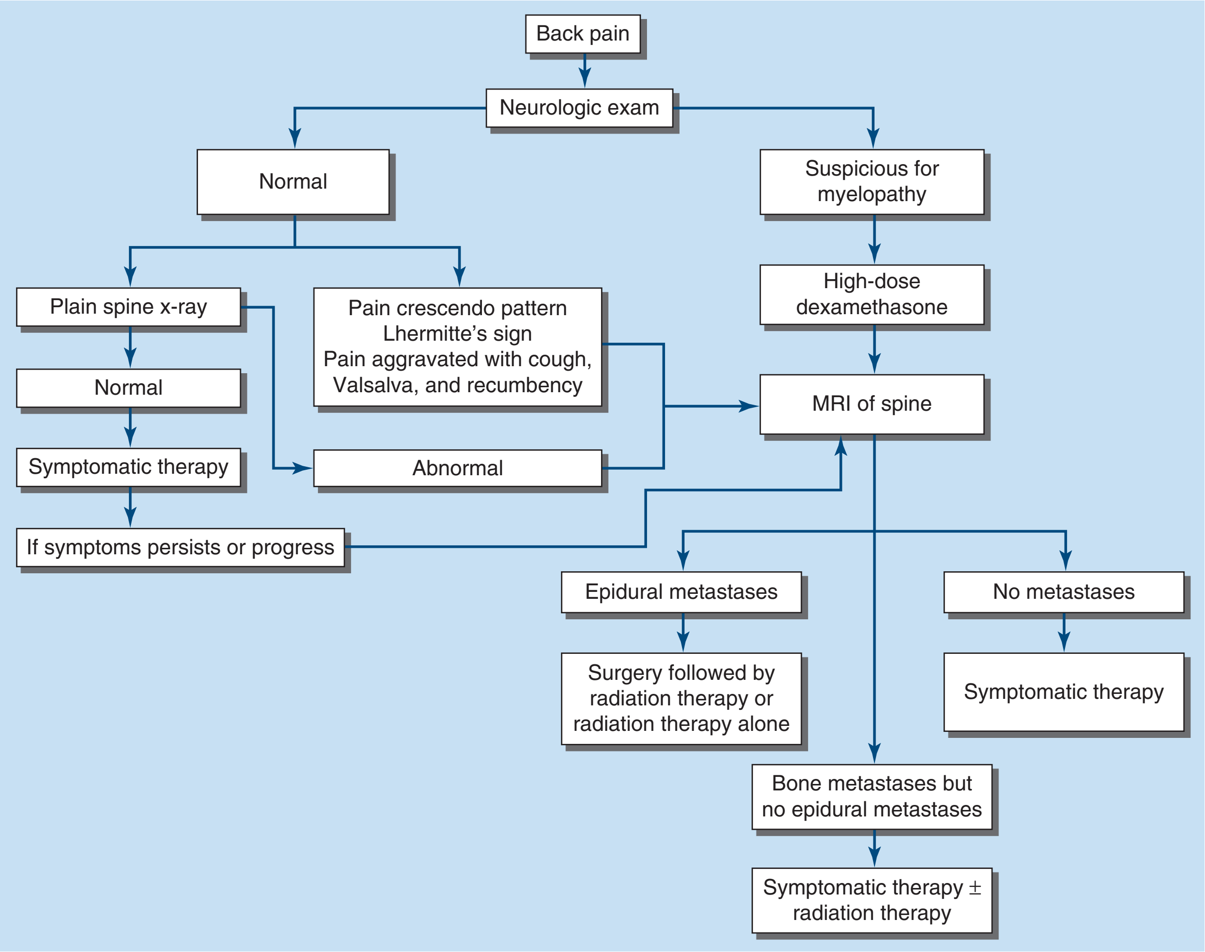
Back pain
Neurologic exam
Normal
Suspicious for myelopathy
Plain spine x-ray
Pain crescendo pattern
Lhermitte's sign
Pain aggravated with cough, Valsalva, and recumbency
High-dose dexamethasone
Normal
MRI of spine
Symptomatic therapy
Abnormal
If symptoms persists or progress
Epidural metastases
No metastases
Surgery followed by radiation therapy or radiation therapy alone
Symptomatic therapy
Bone metastases but no epidural metastases
Symptomatic therapy ± radiation therapy

FIGURE 80-2 Management of cancer patients with back pain.

Cauda equina syndrome is characterized by low back pain; diminished sensation over the buttocks, posterior-superior thighs, and perineal area in a saddle distribution; rectal and bladder dysfunction; sexual impotence; absent bulbocavernous, patellar, and Achilles' reflexes; and variable amount of lower-extremity weakness. This reflects compression of nerve roots as they form the cauda equina after leaving the spinal cord. The majority of cauda equina tumors are primary tumors of glial or nerve sheath origin; metastases are very rare.

Patients with cancer who develop back pain should be evaluated for SCC as quickly as possible **(Fig. 80-2)**. Treatment is more often successful in patients who are ambulatory and still have sphincter control at the time treatment is initiated. Patients should have a neurologic examination and plain films of the spine. Those whose physical examination suggests cord compression should receive dexamethasone starting immediately and undergo MRI imaging.

Erosion of the pedicles (the "winking owl" sign) is the earliest radiologic finding of vertebral tumor in plain films; however, plain films are insensitive. Other radiographic changes include increased intrapedicular distance, vertebral destruction, lytic or sclerotic lesions, scalloped vertebral bodies, and vertebral body collapse. Vertebral collapse is not a reliable indicator of the presence of tumor; ~20% of cases of vertebral collapse, particularly those in older patients and postmenopausal women, are due not to cancer but to osteoporosis. Also, a normal appearance on plain films of the spine does not exclude the diagnosis of cancer. The role of bone scans in the detection of cord compression is not clear; this method is sensitive but less specific than spinal radiography.

The full-length image of the cord provided by MRI is the imaging procedure of choice. Multiple epidural metastases are noted in 25% of patients with cord compression, and their presence influences treatment plans. On T1-weighted images, good contrast is noted between the cord, cerebrospinal fluid (CSF), and extradural lesions. Owing to its sensitivity in demonstrating the replacement of bone marrow by tumor, MRI can show which parts of a vertebra are involved by tumor. MRI also visualizes intraspinal extradural masses compressing the cord. T2-weighted images are most useful for the demonstration of intramedullary pathology. Gadolinium-enhanced MRI can help to delineate intramedullary disease. MRI is as good as or better than myelography plus postmyelogram CT scan in detecting metastatic epidural disease with cord compression. Myelography should be reserved for patients who have poor MRIs or who cannot undergo MRI promptly. CT scan in conjunction with myelography enhances the detection of small areas of spinal destruction.

In patients with cord compression and an unknown primary tumor, a simple workup including chest radiography, mammography, measurement of prostate-specific antigen, and abdominal CT usually reveals the underlying malignancy.

TREATMENT

Spinal Cord Compression

The treatment of patients with SCC is aimed at relief of pain and restoration/preservation of neurologic function (Fig. 80-2). Management of MSCC requires a multidisciplinary approach.

Radiation therapy plus glucocorticoids is generally the initial treatment of choice for most patients with SCC. The management decision of SCC involves assessment of neurologic (N), oncologic (O), mechanical (M), and systemic factors (S). NOMS was developed by Memorial Sloan Kettering Cancer Center (MSKCC) researchers to provide an algorithm for management of SCC. The neurologic assessment is based on the degree of epidural SCC, myelopathy, and/or functional radiculopathy. Oncologic assessment involves the radiosensitivity of the tumor type. In patients with radioresistant tumors, stereotactic body radiotherapy (SBRT) is the preferred approach if radiation is appropriate. Safe delivery of SBRT requires a 2- to 3-mm margin away from the spinal cord. Separation surgery followed by SBRT is necessary in patients with high-grade SCC due to radioresistant tumors. Separation surgery is the circumferential excision of epidural tumor to reconstitute the thecal sac and provide a 2-mm margin for safe delivery of an ablative radiation dose. In patients with mechanical instability or retropulsion of bone fragments into the spinal canal or cord, a surgical approach is the treatment of choice. Systemic factors that need to be considered are the extent of disease and medical comorbidities that determine the patient's ability to tolerate planned therapy. Chemotherapy may have a role in patients with chemosensitive tumors who have had prior radiotherapy to the same region and who are not candidates for surgery. Patients who previously received radiotherapy for MSCC with an in-field tumor progression can be treated with reirradiation with spine stereotactic radiosurgery (SRS) if they are not surgical candidates.

Patients with painful pathologic compression fractures without spinal instability may benefit from percutaneous vertebroplasty or kyphoplasty, the injection of acrylic cement into a collapsed vertebra to stabilize the fracture. Pain palliation is common, and local antitumor effects have been noted. Cement leakage may cause symptoms in ~10% of patients. Bisphosphonates and/or denosumab may be helpful in prevention of SCC in patients with bony involvement.

The histology of the tumor is an important determinant of both recovery and survival. Rapid onset and progression of signs and symptoms are poor prognostic features.

INCREASED INTRACRANIAL PRESSURE

About 25% of patients with cancer die with intracranial metastases. The cancers that most often metastasize to the brain are lung and breast cancers and melanoma. Brain metastases often occur in the presence of systemic disease, and they frequently cause major symptoms, disability, and early death. The initial presentation of brain metastases from a previously unknown primary cancer is common. Lung cancer is most commonly the primary malignancy. CT scans of the chest/abdomen and MRI of the brain as the initial diagnostic studies can identify a biopsy site in most patients.

The signs and symptoms of a metastatic brain tumor are similar to those of other intracranial expanding lesions: headache, nausea, vomiting, behavioral changes, seizures, and focal, progressive neurologic changes. Occasionally the onset is abrupt, resembling a stroke, with the sudden appearance of headache, nausea, vomiting, and neurologic deficits. This picture is usually due to hemorrhage into the metastasis. Melanoma, germ cell tumors, and renal cell cancers have a particularly high incidence of intracranial bleeding. The tumor mass and surrounding edema may cause obstruction of the circulation of CSF, with resulting hydrocephalus. Patients with increased intracranial pressure may have papilledema with visual disturbances and neck stiffness. As the mass enlarges, brain tissue may be displaced through the fixed cranial openings, producing various herniation syndromes.

MRI is superior to CT scan. Gadolinium-enhanced MRI is more sensitive than CT at revealing meningeal involvement and small lesions, particularly in the brainstem or cerebellum. The MRI of the brain shows brain metastases as multiple enhancing lesions of various sizes with surrounding areas of low-density edema.

Intracranial hypertension ("pseudotumor cerebri") secondary to tretinoin therapy for acute promyelocytic leukemia has been reported as another cause of intracranial pressure in the setting of a malignancy.

TREATMENT

Increased Intracranial Pressure

Dexamethasone is the best initial treatment for all symptomatic patients with brain metastases. The current success of immunotherapy approaches for primary and metastatic brain tumors may preclude or limit glucocorticoid use since it may decrease antitumor response. Bevacizumab should be considered in patients who are unable to wean completely off of steroids as well as those who have symptomatic brain edema and are on immunotherapy. Patients with a single brain metastasis and with controlled extracranial disease may be treated with surgical excision followed by SRS to the resection cavity. SRS is recommended in patients with a limited number of brain metastases (one to four) who have stable, systemic disease or reasonable systemic treatment options and in patients who have a small number of metastatic lesions in whom whole-brain radiation therapy has failed. The treatment of a larger number of intracranial metastases remains controversial. More patients now receive SRS because of cognitive dysfunction associated with whole-brain radiation. Some patients with increased intracranial pressure associated with hydrocephalus may benefit from shunt placement. If neurologic deterioration is not reversed with medical therapy, ventriculotomy to remove CSF or craniotomy to remove tumors or hematomas may be necessary.

Targeted agents and checkpoint inhibitors have significant activity in brain metastases from non-small-cell lung cancer, breast cancer, renal cancer, and melanoma.

NEOPLASTIC MENINGITIS

Tumor involving the leptomeninges is a complication of both primary central nervous system (CNS) tumors and tumors that metastasize to the CNS. The incidence is estimated at 3–8% of patients with cancer. Melanoma, breast and lung cancer, lymphoma (including AIDS-associated), and acute leukemia are the most common causes. The lobular or triple-negative subtypes of breast cancer, as well as tumors with expression of the mutant epidermal growth factor receptor (EGFR) or the anaplastic lymphoma kinase (ALK) rearrangement in non-small-cell lung cancer, are more likely to have CNS involvement including neoplastic meningitis and brain metastases. Synchronous intraparenchymal brain metastases are frequent in patients with neoplastic meningitis. Leptomeningeal seeding is frequent in patients undergoing resection of brain metastases or receiving stereotactic radiotherapy for brain metastases.

Patients typically present with multifocal neurologic signs and symptoms, including headache, gait abnormality, mental changes, nausea, vomiting, seizures, back or radicular pain, and limb weakness. Signs include cranial nerve palsies, extremity weakness, paresthesia, and decreased deep tendon reflexes.

Diagnosis is made by demonstrating malignant cells in the CSF; however, up to 40% of patients may have false-negative CSF cytology. An elevated CSF protein level is nearly always present. Patients with neurologic signs and symptoms consistent with neoplastic meningitis who have a negative CSF cytology should have the spinal tap repeated at least one more time for cytologic examination. MRI findings suggestive of neoplastic meningitis include leptomeningeal, subependymal, dural, or cranial nerve enhancement; superficial cerebral lesions; intradural nodules; and communicating hydrocephalus. Spinal cord imaging by MRI is a necessary component of the evaluation of nonleukemia neoplastic meningitis because ~20% of patients have cord abnormalities, including intradural enhancing nodules that are diagnostic for leptomeningeal involvement. Cauda equina lesions are common, but lesions may be seen anywhere in the spinal canal. Radiolabeled CSF flow studies are abnormal in up to 70% of patients with neoplastic meningitis; ventricular outlet obstruction, abnormal flow in the spinal canal, or impaired flow over the cerebral convexities may affect distribution of intrathecal chemotherapy, resulting in decreased efficacy or increased toxicity. Radiation therapy may correct CSF flow abnormalities before use of intrathecal chemotherapy. Neoplastic meningitis can also lead to intracranial hypertension and hydrocephalus. Placement of a ventriculoperitoneal shunt may effectively palliate symptoms in these patients.

The development of neoplastic meningitis usually occurs in the setting of uncontrolled cancer outside the CNS; thus, prognosis is poor (median survival 10–12 weeks). However, treatment of the neoplastic meningitis may successfully alleviate symptoms and control the CNS spread.

TREATMENT

Neoplastic Meningitis

Chemotherapy provided by either intrathecal injection or systemic routes is used to control leptomeningeal disease throughout the entire neuroaxis. Intrathecal chemotherapy, usually methotrexate, cytarabine, or thiotepa, is delivered by lumbar puncture or by an intraventricular reservoir (Ommaya). Among solid tumors, breast cancer responds best to therapy. Focal radiotherapy may have a role in bulky disease and in symptomatic or obstructive lesions. Targeted therapy such as systemically administered EGFR tyrosine kinase inhibitors (TKIs) in non-small-cell lung cancer may lead to improvement in some patients with leptomeningeal spread. Patients with neoplastic meningitis from either acute leukemia or lymphoma may be cured of their CNS disease if the systemic disease can be eliminated.

SEIZURES

Seizures occurring in a patient with cancer can be caused by the tumor itself, by metabolic disturbances, by radiation injury, by cerebral infarctions, by chemotherapy-related encephalopathies, or by CNS infections. Metastatic disease to the CNS is the most common cause of seizures in patients with cancer. However, seizures occur more frequently in primary brain tumors than in metastatic brain lesions. Seizures are a presenting symptom of CNS metastasis in 6–29% of cases. Approximately 10% of patients with CNS metastasis eventually develop seizures. Tumors that affect the frontal, temporal, and parietal lobes are more commonly associated with seizures than are occipital lesions. Both early and late seizures are uncommon in patients with posterior fossa and sellar lesions. Seizures are common in patients with CNS metastases from melanoma and low-grade primary brain tumors. Very rarely, cytotoxic drugs such as etoposide, busulfan, ifosfamide, and chlorambucil cause seizures. Treatment with bispecific antibodies and chimeric antigen receptor (CAR) T cells may also cause CNS toxicity including seizures and encephalopathy. Another cause of seizures related to drug therapy is reversible posterior leukoencephalopathy syndrome (RPLS). Chemotherapy, targeted therapy, and immunotherapies have been associated with the development of RPLS. RPLS occurs in patients undergoing allogeneic bone marrow or solid-organ transplantation. RPLS is characterized by headache, altered consciousness, generalized seizures, visual disturbances, hypertension, and symmetric posterior cerebral white matter vasogenic edema on CT/MRI. Seizures may begin focally but are typically generalized.

TREATMENT

Seizures

Patients in whom seizures due to CNS metastases have been demonstrated should receive anticonvulsive treatment with levetiracetam, lacosamide, or lamotrigine. Prophylactic anticonvulsant therapy is not recommended. In postcraniotomy patients, prophylactic antiepileptic drugs should be withdrawn during the first 2 weeks after surgery. Most antiseizure medications including phenytoin induce cytochrome P450 (CYP450), which alters the metabolism of many antitumor agents, including irinotecan, taxanes, and etoposide, as well as molecular targeted agents, including imatinib, gefitinib, erlotinib, tipifarnib, sorafenib, sunitinib, temsirolimus, everolimus, and vemurafenib. Levetiracetam, lacosamide, and lamotrigine are anticonvulsant agents not metabolized by the hepatic CYP450 system and do not alter the metabolism of antitumor agents. Levetiracetam has become the preferred drug. Surgical resection and other antitumor treatments such as radiotherapy and chemotherapy may improve seizure control.

PULMONARY AND INTRACEREBRAL LEUKOSTASIS

Hyperleukocytosis and the leukostasis syndrome associated with it are potentially fatal complications of acute leukemia (particularly myeloid leukemia) that can occur when the peripheral blast cell count is >100,000/mL. The frequency of hyperleukocytosis is 5–13% in acute myeloid leukemia (AML) and 10–30% in acute lymphoid leukemia; however, leukostasis is rare in lymphoid leukemia. In AML, hyperleukocytosis is more commonly seen in myelomonocytic AML (Fab-M4 and M5), those with 11q13 abnormalities involving the *MLL* gene, and those with *FLT3* mutations. At such high blast cell counts, blood viscosity is increased, blood flow is slowed by aggregates of tumor cells, and the primitive myeloid leukemic cells are capable of invading through the endothelium and causing hemorrhage. Brain and lung are most commonly affected. Patients with brain leukostasis may experience stupor, headache, dizziness, tinnitus, visual disturbances, ataxia, confusion, coma, or sudden death. On examination, papilledema, retinal vein distension, retinal hemorrhages, and focal deficit may be present. Pulmonary leukostasis may present as respiratory distress and hypoxemia and progress to respiratory failure. Chest radiographs may be normal but usually show interstitial or alveolar infiltrates. Hyperleukocytosis rarely may cause acute leg ischemia, renal vein thrombosis, myocardial ischemia, bowel infraction, and priapism. Arterial blood gas results should be interpreted cautiously. Rapid consumption of plasma oxygen by the markedly increased number of white blood cells can cause spuriously low arterial oxygen tension. Pulse oximetry is the most accurate way of assessing oxygenation in patients with hyperleukocytosis. Hydroxyurea can rapidly reduce a high blast cell count while the diagnostic workup is in progress. After the diagnosis is established, the patient should start quickly with effective induction chemotherapy. Leukapheresis should be used in patients with symptoms of hyperleukocytosis. Patients with hyperleukocytosis are also at risk for disseminated intravascular coagulation and tumor lysis syndrome. The clinician should monitor the patient for these complications and take preventive and therapeutic actions during induction therapy. Intravascular volume depletion and unnecessary blood transfusions may increase blood viscosity and worsen the leukostasis syndrome. Leukostasis is very rarely a feature of the high white cell counts associated with chronic lymphoid or chronic myeloid leukemia.

When acute promyelocytic leukemia is treated with differentiating agents like tretinoin and arsenic trioxide, cerebral or pulmonary leukostasis may occur as tumor cells differentiate into mature neutrophils. This complication can be largely avoided by using cytotoxic chemotherapy together with the differentiating agents.

HEMOPTYSIS

Hemoptysis may be caused by nonmalignant conditions, but lung cancer accounts for a large proportion of cases. Up to 20% of patients with lung cancer have hemoptysis some time in their course. Endobronchial metastases from carcinoid tumors, breast cancer, colon cancer, kidney cancer, and melanoma may also cause hemoptysis. The volume of bleeding is often difficult to gauge. Massive hemoptysis is defined as >200–600 mL of blood produced in 24 h. However, any hemoptysis should be considered massive if it threatens life. When respiratory difficulty occurs, hemoptysis should be treated emergently. The first priorities are to maintain the airway, optimize oxygenation, and stabilize the hemodynamic status. If the bleeding side is known, the patient should be placed in a lateral decubitus position, with the bleeding side down to prevent aspiration into the unaffected lung and given supplemental oxygen. If large-volume bleeding continues or the airway is compromised, the patient should be intubated and undergo emergency bronchoscopy. If the site of bleeding is detected, either the patient undergoes a definitive surgical procedure or the lesion is treated with a neodymium:yttrium-aluminum-garnet (Nd:YAG) laser, argon plasma coagulation, or electrocautery. In stable patients, multidetector CT angiography delineates bronchial and nonbronchial systemic arteries and identifies the source of bleeding and underlying pathology with high sensitivity. Massive hemoptysis usually originates from the high-pressure bronchial circulation. Bronchial artery embolization is considered a first-line definitive procedure for managing hemoptysis. Bronchial artery embolization may control brisk bleeding in 75–90% of patients, permitting the definitive surgical procedure to be done more safely if it is appropriate.

Embolization without definitive surgery is associated with rebleeding in 20–50% of patients. Recurrent hemoptysis usually responds to a second embolization procedure. A postembolization syndrome characterized by pleuritic pain, fever, dysphagia, and leukocytosis may occur; it lasts 5–7 days and resolves with symptomatic treatment. Bronchial or esophageal wall necrosis, myocardial infarction, and spinal cord infarction are rare complications. Surgery, as a salvage strategy, is indicated after failure of embolization and is associated with better survival when performed in a nonurgent setting.

Pulmonary hemorrhage with or without hemoptysis in hematologic malignancies is often associated with fungal infections, particularly *Aspergillus* spp. After granulocytopenia resolves, the lung infiltrates in aspergillosis may cavitate and cause massive hemoptysis. Thrombocytopenia and coagulation defects should be corrected, if possible. Surgical evaluation is recommended in patients with aspergillosis-related cavitary lesions. Antibodies to vascular endothelial growth factor (VEGF) including bevacizumab and ramucirumab that inhibit angiogenesis, have been associated with life-threatening hemoptysis in patients with non-small-cell lung cancer, particularly of squamous cell histology. Non-small-cell lung cancer patients with cavitary lesions or previous hemoptysis (≥2.5 mL) within the past 3 months have higher risk for pulmonary hemorrhage.

■ AIRWAY OBSTRUCTION

Airway obstruction refers to a blockage at the level of the mainstem bronchi or above. It may result either from intraluminal tumor growth or from extrinsic compression of the airway. The most common cause of malignant upper airway obstruction is invasion from an adjacent primary tumor, most commonly lung cancer, followed by esophageal, thyroid, and mediastinal malignancies including lymphomas. Extrathoracic primary tumors such as renal, colon, or breast cancer can cause airway obstruction through endobronchial and/or mediastinal lymph node metastases. Patients may present with dyspnea, hemoptysis, stridor, wheezing, intractable cough, postobstructive pneumonia, or hoarseness. Chest radiographs usually demonstrate obstructing lesions. CT scans reveal the extent of tumor. Cool, humidified oxygen, glucocorticoids, and ventilation with a mixture of helium and oxygen (Heliox) may provide temporary relief. If the obstruction is proximal to the larynx, a tracheostomy may be lifesaving. For more distal obstructions, particularly intrinsic lesions incompletely obstructing the airway, bronchoscopy with mechanical debulking and dilation or ablational treatments including laser treatment, photodynamic therapy, argon plasma coagulation, electrocautery, or stenting can produce immediate relief in most patients (Fig. 80-3). However, radiation therapy (either external-beam irradiation or brachytherapy) given together with glucocorticoids may also open the airway. Symptomatic extrinsic compression may be palliated by stenting. Patients with primary airway tumors such as squamous cell carcinoma, carcinoid tumor, adenocystic carcinoma, or non-small-cell lung cancer, if resectable, should have surgery.

METABOLIC EMERGENCIES

■ HYPERCALCEMIA

Hypercalcemia is the most common paraneoplastic syndrome. Its pathogenesis and management are discussed fully in Chaps. 98 and 422.

■ SYNDROME OF INAPPROPRIATE SECRETION OF ANTIDIURETIC HORMONE

Hyponatremia is a common electrolyte abnormality in cancer patients, and syndrome of inappropriate secretion of antidiuretic hormone (SIADH) is the most common cause among patients with cancer. SIADH is discussed fully in Chaps. 98 and 393.

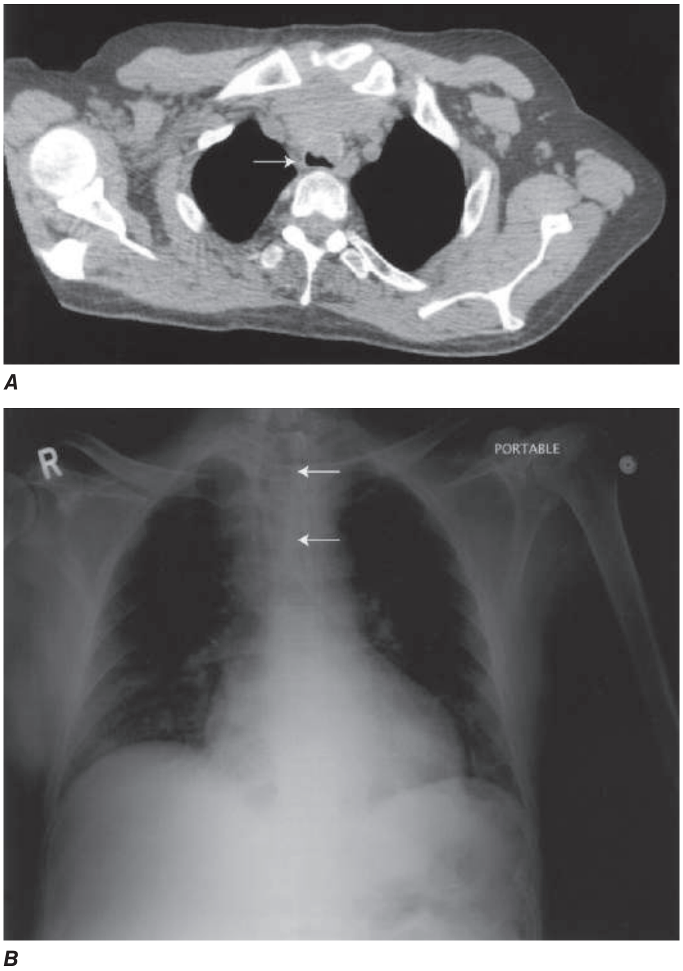

FIGURE 80-3 Airway obstruction. ***A.*** Computed tomography scan of a 62-year-old man with tracheal obstruction caused by renal carcinoma showing paratracheal mass with tracheal invasion/obstruction (*arrow*). ***B.*** Chest x-ray of same patient after stent (*arrows*) placement.

■ LACTIC ACIDOSIS

Lactic acidosis is a rare and potentially fatal metabolic complication of cancer. Lactic acidosis associated with sepsis and circulatory failure is a common preterminal event in many malignancies. Lactic acidosis in the absence of hypoxemia may occur in patients with leukemia, lymphoma, or solid tumors. In some cases, hypoglycemia also is present. Extensive involvement of the liver by tumor is often present. In most cases, decreased metabolism and increased production by the tumor both contribute to lactate accumulation. Tumor cell overexpression of certain glycolytic enzymes and mitochondrial dysfunction can contribute to its increased lactate production. HIV-infected patients have an increased risk of aggressive lymphoma; lactic acidosis that occurs in such patients may be related either to the rapid growth of the tumor or from toxicity of nucleoside reverse transcriptase inhibitors. Symptoms of lactic acidosis include tachypnea, tachycardia, change of mental status, and hepatomegaly. The serum level of lactic acid may reach 10–20 mmol/L (90–180 mg/dL). Treatment is aimed at the underlying disease. Sodium bicarbonate should be added if acidosis is very severe or if hydrogen ion production is very rapid and uncontrolled. Other treatment options include renal replacement therapy, such as hemodialysis, and thiamine replacement. The prognosis is poor regardless of the treatment offered.

■ HYPOGLYCEMIA

Persistent hypoglycemia is occasionally associated with tumors other than pancreatic islet cell tumors. Usually these tumors are large;

tumors of mesenchymal origin, hepatomas, or adrenocortical tumors may cause hypoglycemia. Mesenchymal tumors are usually located in the retroperitoneum or thorax. Obtundation, confusion, and behavioral aberrations occur in the postabsorptive period and may precede the diagnosis of the tumor. These tumors often secrete incompletely processed insulin-like growth factor II (IGF-II), a hormone capable of activating insulin receptors and causing hypoglycemia. Tumors secreting incompletely processed big IGF-II are characterized by an increased IGF-II to IGF-I ratio, suppressed insulin and C-peptide level, and inappropriately low growth hormone and β-hydroxybutyrate concentrations. Rarely, hypoglycemia is due to insulin secretion by a non–islet cell carcinoma. The development of hepatic dysfunction from liver metastases and increased glucose consumption by the tumor can contribute to hypoglycemia. If the tumor cannot be resected, hypoglycemia symptoms may be relieved by the administration of glucose, glucocorticoids, recombinant growth hormone, or glucagon.

Hypoglycemia can be artifactual; hyperleukocytosis from leukemia, myeloproliferative diseases, leukemoid reactions, or colony-stimulating factor treatment can increase glucose consumption in the test tube after blood is drawn, leading to pseudohypoglycemia.

ADRENAL INSUFFICIENCY

In patients with cancer, adrenal insufficiency may go unrecognized because the symptoms, such as nausea, vomiting, anorexia, and orthostatic hypotension, are nonspecific and may be mistakenly attributed to progressive cancer or to therapy. Primary adrenal insufficiency may develop owing to replacement of both glands by metastases (lung, breast, colon, or kidney cancer; lymphoma), to removal of both glands, or to hemorrhagic necrosis in association with sepsis or anticoagulation. Impaired adrenal steroid synthesis occurs in patients being treated for cancer with mitotane, ketoconazole, or aminoglutethimide or undergoing rapid reduction in glucocorticoid therapy. Megestrol acetate, used to manage cancer and HIV-related cachexia, may suppress plasma levels of cortisol and adrenocorticotropic hormone (ACTH). Patients taking megestrol may develop adrenal insufficiency, and even those whose adrenal dysfunction is not symptomatic may have inadequate adrenal reserve if they become seriously ill. Paradoxically, some patients may develop Cushing's syndrome and/or hyperglycemia because of the glucocorticoid-like activity of megestrol acetate. Ipilimumab, an anti-CTLA-4 antibody used for treatment of malignant melanoma and other cancers, may cause autoimmunity including autoimmune-like enterocolitis, hypophysitis (leading to secondary adrenal insufficiency), hepatitis, and, rarely, primary adrenal insufficiency. Autoimmune hypophysitis may present with headache, visual field defects, and pituitary hormone deficiencies manifesting as hypopituitarism, adrenal insufficiency (including adrenal crisis), or hypothyroidism. Ipilimumab-associated hypophysitis symptoms occur at an average of 6–12 weeks after initiation of therapy. An MRI usually shows homogenous enhancement of pituitary gland. Early glucocorticoid treatment and hormone replacement are the initial treatment. The role of high-dose glucocorticoids in the treatment of hypophysitis is not clear. High-dose glucocorticoids may not improve the frequency of pituitary function recovery. Autoimmune adrenalitis can also be observed with anti-CTLA-4 antibody. Pituitary dysfunction is usually permanent, requiring long-term hormone replacement therapy. Other checkpoint inhibitors, such as monoclonal antibodies targeting programmed cell death 1 (PD-1), an inhibitory receptor expressed by T cells or one of its ligands (PD-L1), may cause hypophysitis infrequently (~1%). Autoimmune adrenalitis is more frequent with use of PD-1/PD-L1 than with CTLA-4 inhibitors, but incidence is low. Cranial irradiation for childhood brain tumors may affect the hypothalamus-pituitary-adrenal axis, resulting in secondary adrenal insufficiency. Rarely, metastatic replacement causes primary adrenal insufficiency as the first manifestation of an occult malignancy. Metastasis to the pituitary or hypothalamus is found at autopsy in up to 5% of patients with cancer, but associated secondary adrenal insufficiency is rare.

Acute adrenal insufficiency is potentially lethal. Treatment of suspected adrenal crisis is initiated after the sampling of serum cortisol and ACTH levels **(Chap. 398)**.

TREATMENT-RELATED EMERGENCIES

TUMOR LYSIS SYNDROME

Tumor lysis syndrome (TLS) is characterized by hyperuricemia, hyperkalemia, hyperphosphatemia, and hypocalcemia and is caused by the destruction of a large number of rapidly proliferating neoplastic cells. Acidosis may also develop. Acute renal failure occurs frequently.

TLS is most often associated with the treatment of Burkitt's lymphoma, acute lymphoblastic leukemia, AML, and other rapidly proliferating lymphomas, but it also may be seen with chronic leukemias and, rarely, with solid tumors. This syndrome is increased in frequency in lymphoid neoplasms treated with venetoclax, a bcl-2 antagonist. TLS has been observed with administration of glucocorticoids, hormonal agents such as letrozole and tamoxifen, and monoclonal antibodies such as rituximab, obinutuzumab, ofatumumab, and gemtuzumab. TLS usually occurs during or shortly (1–5 days) after chemotherapy. Rarely, spontaneous necrosis of malignancies causes TLS.

Hyperuricemia may be present at the time of chemotherapy. Effective treatment kills malignant cells and leads to increased serum uric acid levels from the turnover of nucleic acids. Owing to the acidic local environment, uric acid can precipitate in the tubules, medulla, and collecting ducts of the kidney, leading to renal failure. Lactic acidosis and dehydration may contribute to the precipitation of uric acid in the renal tubules. The finding of uric acid crystals in the urine is strong evidence for uric acid nephropathy. The ratio of urinary uric acid to urinary creatinine is >1 in patients with acute hyperuricemic nephropathy and <1 in patients with renal failure due to other causes. Other events may lead to renal failure in TLS. Calcium phosphate also precipitates in the interstitium and renal microvasculature, leading to nephrocalcinosis. Both types of crystals are toxic to the tubular epithelium, inducing local active inflammatory and pro-oxidative responses. Soluble uric acid may induce hemodynamic changes, with decreased renal blood flow due to vasoconstriction and impaired autoregulation (crystal-independent pathway).

Hyperphosphatemia, which can be caused by the release of intracellular phosphate pools by tumor lysis, produces a reciprocal depression in serum calcium, which causes severe neuromuscular irritability and tetany. Deposition of calcium phosphate in the kidney and hyperphosphatemia may cause renal failure. Potassium is the principal intracellular cation, and massive destruction of malignant cells may lead to hyperkalemia. Hyperkalemia in patients with renal failure may rapidly become life threatening by causing ventricular arrhythmias and sudden death.

The likelihood that TLS will occur in patients with Burkitt's lymphoma is related to the tumor burden and renal function. Hyperuricemia and high serum levels of lactate dehydrogenase (LDH >1500 U/L), both of which correlate with total tumor burden, also correlate with the risk of TLS. In patients at risk for TLS, pretreatment evaluations should include a complete blood count, serum chemistry evaluation, and urinalysis. High leukocyte and platelet counts may artificially elevate potassium levels ("pseudohyperkalemia") due to lysis of these cells after the blood is drawn. In these cases, plasma potassium instead of serum potassium should be followed. In pseudohyperkalemia, no electrocardiographic abnormalities are present. In patients with abnormal baseline renal function, the kidneys and retroperitoneal area should be evaluated by sonography and/or CT to rule out obstructive uropathy. Urine output should be watched closely.

TREATMENT

Tumor Lysis Syndrome

Recognition of risk and prevention are the most important steps in the management of this syndrome **(Fig. 80-4)**. The standard preventive approach consists of allopurinol and aggressive hydration. Urinary alkalization with sodium bicarbonate is no longer recommended. It increases uric acid solubility, but a high pH decreases the solubility of xanthine, hypoxanthine, and calcium phosphate, potentially increasing the likelihood of intratubular crystallization. Intravenous allopurinol may be given in patients who cannot tolerate oral therapy. Febuxostat, a potent nonpurine selective xanthine

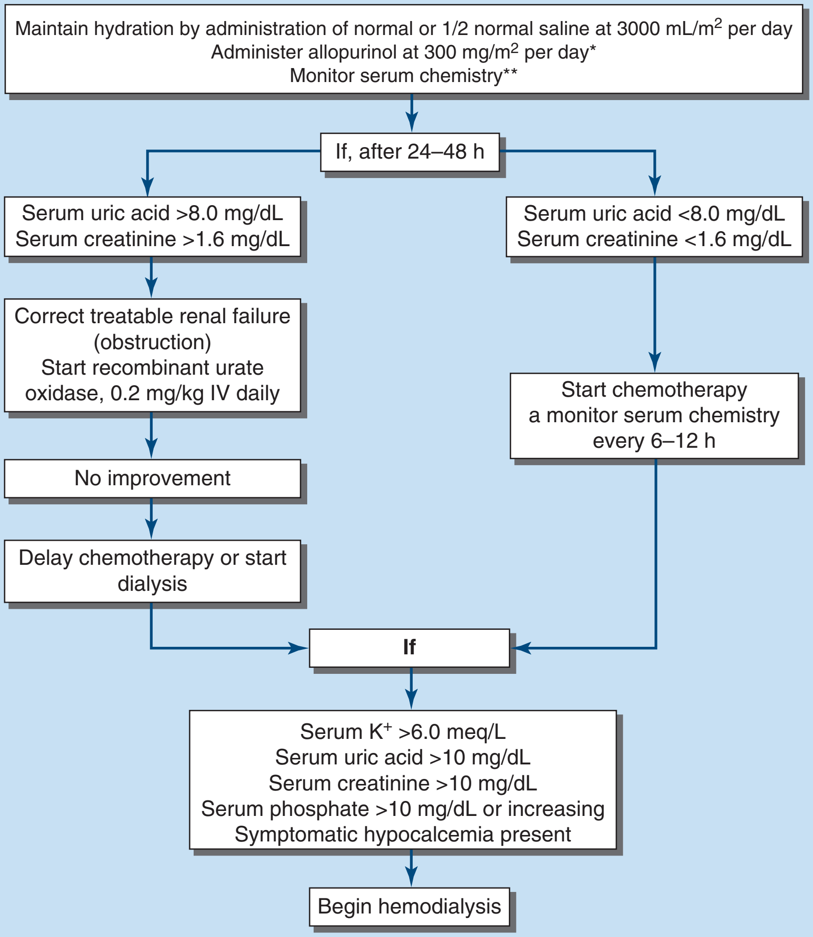
Maintain hydration by administration of normal or 1/2 normal saline at 3000 mL/m2 per day
Administer allopurinol at 300 mg/m2 per day*
Monitor serum chemistry**
If, after 24–48 h
Serum uric acid >8.0 mg/dL
Serum creatinine >1.6 mg/dL
Serum uric acid <8.0 mg/dL
Serum creatinine <1.6 mg/dL
Correct treatable renal failure (obstruction)
Start recombinant urate oxidase, 0.2 mg/kg IV daily
No improvement
Delay chemotherapy or start dialysis
Start chemotherapy
a monitor serum chemistry every 6–12 h
If
Serum K+ >6.0 meq/L
Serum uric acid >10 mg/dL
Serum creatinine >10 mg/dL
Serum phosphate >10 mg/dL or increasing
Symptomatic hypocalcemia present
Begin hemodialysis

*Use febuxostat in allopurinol allergy and/or renal failure.

**Specifically serum electrolytes, creatinine, calcium, phosphate, and uric acid.

FIGURE 80-4 Management of patients at high risk for the tumor lysis syndrome.

oxidase inhibitor, is indicated for treatment of hyperuricemia. It results in fewer hypersensitivity reactions than allopurinol. Febuxostat does not require dosage adjustment in patients with mild to moderate renal impairment. Febuxostat achieved significantly superior serum uric acid control in comparison to allopurinol in patients with hematologic malignancies at intermediate to high TLS risk. In some cases, uric acid levels cannot be lowered sufficiently with the standard preventive approach. Rasburicase (recombinant urate oxidase) can be effective in these instances, particularly when renal failure is present. Urate oxidase is missing from primates and catalyzes the conversion of poorly soluble uric acid to readily soluble allantoin. Rasburicase acts rapidly, decreasing uric acid levels within hours; however, it may cause hypersensitivity reactions such as bronchospasm, hypoxemia, and hypotension. Rasburicase should also be administered to high-risk patients for TLS prophylaxis. Rasburicase is contraindicated in patients with glucose-6-phosphate dehydrogenase deficiency who are unable to break down hydrogen peroxide, an end product of the urate oxidase reaction. Rasburicase is known to cause ex vivo enzymatic degradation of uric acid in test tube at room temperature. This leads to spuriously low uric acid levels during laboratory monitoring of the patient with TLS. Samples must be cooled immediately to deactivate the urate oxidase. Despite aggressive prophylaxis, TLS and/or oliguric or anuric renal failure may occur. Renal replacement therapy is often necessary and should be considered early in the course. Hemodialysis is preferred. Hemofiltration offers a gradual, continuous method of removing cellular by-products and fluid.

HUMAN ANTIBODY INFUSION REACTIONS

The initial infusion of human or humanized antibodies (e.g., rituximab, gemtuzumab, trastuzumab, alemtuzumab, panitumumab, brentuximab vedotin, blinatumomab) is associated with fever, chills, nausea, asthenia, and headache in up to half of treated patients. Bronchospasm and hypotension occur in 1% of patients. Severe manifestations including pulmonary infiltrates, acute respiratory distress syndrome (ARDS), and cardiogenic shock occur rarely. Laboratory manifestations include elevated hepatic aminotransferase levels, thrombocytopenia, and prolongation of prothrombin time. The pathogenesis is thought to be activation of immune effector processes (cells and complement) and release of inflammatory cytokines, such as tumor necrosis factor α, interferon γ, interleukin (IL) 6, and IL-10 (cytokine release syndrome [CRS]). Although its origins are not completely understood, CRS is believed to be due to activation of a variety of cell types including neutrophils, monocytes/macrophages, natural killer cells, and T and B lymphocytes. Hemophagocytic lymphohistiocytosis (HLH)/macrophage activation syndrome (MAS) can develop as part of CRS and usually is a manifestation of severe CRS.

Severe CRS may require intensive support for ARDS and resistant hypotension. Emerging clinical experience at several institutions has concluded that tocilizumab is an effective treatment for severe or life-threatening CRS. Tocilizumab prevents IL-6 binding to both cell-associated and soluble IL-6 receptors and therefore inhibits both classical and trans-IL-6 signaling. Other cytokine-directed therapies, such as siltuximab, a chimeric anti-IL-6 monoclonal antibody, and anakinra, an IL-1 receptor antagonist, have been used.

Adoptive transfer of CAR-engineered T cells is a promising therapy for cancers. The most common acute toxicity of CAR T cells is CRS. CAR T cell–associated CRS may be associated with cardiac dysfunction and neurotoxicity. In all cases, MAS could happen with CRS. The management includes supportive care and tocilizumab.

HEMOLYTIC-UREMIC SYNDROME

Malignancy can induce hemolytic-uremic syndrome (HUS) through a wide variety of mechanisms, including systemic microvascular metastases, extensive bone marrow invasion, or secondary necrosis. HUS syndromes have been reported with metastatic gastric and ovarian cancers, leukemias, and lymphomas. HUS and, less commonly, thrombotic thrombocytopenic purpura (TTP) **(Chap. 329)** may rarely occur after treatment with antineoplastic drugs, including mitomycin, gemcitabine, cisplatin, bleomycin, and proteasome inhibitors, and with VEGF inhibitors. Mitomycin and gemcitabine are the most common offenders. Unlike mitomycin, there is no clear-cut relationship between the cumulative dose of gemcitabine and risk of HUS. It occurs most often in patients with gastric, lung, colorectal, pancreatic, and breast carcinoma. In one series, 35% of patients were without evident cancer at the time this syndrome appeared. Secondary HUS/TTP has also been reported as a rare but sometimes fatal complication of bone marrow transplantation.

HUS usually has its onset 4–8 weeks after the last dose of chemotherapy, but it is not rare to detect it several months later. HUS is characterized by microangiopathic hemolytic anemia, thrombocytopenia, and renal failure. Dyspnea, weakness, fatigue, oliguria, and purpura are also common initial symptoms and findings. Systemic hypertension and pulmonary edema frequently occur. Severe hypertension, pulmonary edema, and rapid worsening of hemolysis and renal function may occur after a blood or blood product transfusion. Cardiac findings include atrial arrhythmias, pericardial friction rub, and pericardial effusion. Raynaud's phenomenon is part of the syndrome in patients treated with bleomycin.

Laboratory findings include severe to moderate anemia associated with red blood cell fragmentation and numerous schistocytes on peripheral smear. Reticulocytosis, decreased plasma haptoglobin, and an LDH level document hemolysis. The serum bilirubin level is usually normal or slightly elevated. The Coombs test is negative. The white cell count is usually normal, and thrombocytopenia (<100,000/μL) is almost always present. Most patients have a normal coagulation profile, although some have mild elevations in thrombin time and in levels of fibrin degradation products. The serum creatinine level is elevated at presentation and shows a pattern of subacute worsening within weeks of the initial azotemia. The urinalysis reveals hematuria, proteinuria, and granular or hyaline casts, and circulating immune complexes may be present.

The basic pathologic lesion appears to be deposition of fibrin in the walls of capillaries and arterioles, and these deposits are similar to those seen in HUS due to other causes. These microvascular abnormalities involve mainly the kidneys and rarely occur in other organs. The pathogenesis of cancer treatment–related HUS is not completely understood, but probably the most important factor is endothelial damage. Primary forms of HUS/TTP are related to a decrease in processing of von Willebrand factor by a protease called ADAMTS13.

The case-fatality rate is high; most patients die within a few months. Optimal treatment for chemotherapy-induced HUS is debated. Immunocomplex removal through plasmapheresis, plasma exchange, immunoadsorption, or exchange transfusion, antiplatelet and anticoagulant therapies, and immunosuppression have all been employed with varying degrees of success.

The outcome with plasma exchange is generally poor, as in many other cases of secondary TTP. Rituximab is successfully used in patients with chemotherapy-induced HUS as well as in ADAMTS13-deficient TTP. Eculizumab, a complement inhibitor, is now considered first-line treatment of atypical HUS. Ravulizumab, a humanized monoclonal antibody that blocks terminal complement activation at C5 and is engineered from eculizumab, has an extended half-life and is approved by the U.S. Food and Drug Administration (FDA) for atypical HUS. Vaccination against *Neisseria meningitis* is mandatory before eculizumab and ravulizumab are administered.

NEUTROPENIA AND INFECTION

These remain the most common serious complications of cancer therapy. They are covered in detail in **Chap. 79**.

PULMONARY INFILTRATES

Patients with cancer may present with dyspnea associated with diffuse interstitial infiltrates on chest radiographs. Such infiltrates may be due to progression of the underlying malignancy, treatment-related toxicities, infection, and/or unrelated diseases. The cause may be multifactorial; however, most commonly, they occur as a consequence of treatment. Infiltration of the lung by malignancy has been described in patients with leukemia, lymphoma, and breast and other solid cancers. Pulmonary lymphatics may be involved diffusely by neoplasm (pulmonary lymphangitic carcinomatosis), resulting in a diffuse increase in interstitial markings on chest radiographs. The patient is often mildly dyspneic at the onset, but pulmonary failure develops over a period of weeks. In some patients, dyspnea precedes changes on the chest radiographs and is accompanied by a nonproductive cough. This syndrome is characteristic of solid tumors. In patients with leukemia, diffuse microscopic neoplastic peribronchial and peribronchiolar infiltration is frequent but may be asymptomatic. However, some patients present with diffuse interstitial infiltrates, an alveolar capillary block syndrome, and respiratory distress. Thickening of bronchovascular bundles and prominence of peripheral arteries are CT findings suggestive of leukemic infiltration. In these situations, glucocorticoids can provide symptomatic relief, but specific chemotherapy should always be started promptly.

Several cytotoxic agents, such as bleomycin, methotrexate, busulfan, nitrosoureas, gemcitabine, mitomycin, vinorelbine, docetaxel, paclitaxel, fludarabine, pentostatin, and ifosfamide, may cause pulmonary damage. The most frequent presentations are interstitial pneumonitis, alveolitis, and pulmonary fibrosis. Some cytotoxic agents, including methotrexate and procarbazine, may cause an acute hypersensitivity reaction. Cytosine arabinoside has been associated with noncardiogenic pulmonary edema. Administration of multiple cytotoxic drugs, as well as radiotherapy and preexisting lung disease, may potentiate the pulmonary toxicity. Supplemental oxygen may potentiate the effects of drugs and radiation injury. Patients should always be managed with the lowest $\mathrm{F{\small IO}}_2$ that is sufficient to maintain hemoglobin saturation.

The onset of symptoms may be insidious, with symptoms including dyspnea, nonproductive cough, and tachycardia. Patients may have bibasilar crepitant rales, end-inspiratory crackles, fever, and cyanosis. The chest radiograph generally shows an interstitial and sometimes an intraalveolar pattern that is strongest at the lung bases and may be symmetric. A small effusion may occur. Hypoxemia with decreased carbon monoxide diffusing capacity is always present. Glucocorticoids may be helpful in patients in whom pulmonary toxicity is related to radiation therapy or to chemotherapy. Treatment is otherwise supportive.

Molecular targeted agents, imatinib, erlotinib, and gefitinib are potent inhibitors of tyrosine kinases. These drugs may cause interstitial lung disease (ILD). In the case of gefitinib, preexisting fibrosis, poor performance status, and prior thoracic irradiation are independent risk factors; this complication has a high fatality rate. In Japan, incidence of ILD associated with gefitinib was ~4.5% compared to 0.5% in the United States. Osimertinib may cause transient pulmonary opacities (TPOs). TPOs are characterized asymptomatic and localized ground-glass opacities (GGO) with or without nodular consolidation on CT. TPOs typically resolve during continued osimertinib therapy. Temsirolimus and everolimus, both esters of rapamycin (sirolimus), are agents that block the effects of mammalian target of rapamycin (mTOR), an enzyme that has an important role in regulating the synthesis of proteins that control cell division. These agents may cause GGO in the lung with or without diffuse interstitial disease and lung parenchymal consolidation. Patients may be asymptomatic with only radiologic findings or may be symptomatic. Symptoms include cough, dyspnea, and/or hypoxemia, and sometimes patients present with systemic symptoms such as fever and fatigue. The incidence of everolimus-induced ILD also appears to be higher in Japanese patients. HER2-targeting ADC trastuzumab-deruxtecan carries a known risk of ILD with fatality. Treatment includes dose reduction or withdrawal and, in some cases, the addition of glucocorticoids.

The FDA-approved immune checkpoint inhibitors (ICIs) of the PD-1 and PD-L1 pathway, including nivolumab, pembrolizumab, durvalumab, avelumab, atezolizumab, and cemiplimab, enhance antitumor activity by blocking negative regulators of T-cell function. Immune-mediated pneumonitis is rare (10%) but may be a life-threatening complication of these drugs. Pneumonitis symptoms include cough, shortness of breath, dyspnea, and fever, and often involve only asymptomatic radiographic changes. Pneumonitis shows ground-glass patchy lesions and/or disseminated nodular infiltrates, predominantly in the lower lobes. Identifying the exact cause of a pneumonitis in a patient treated with ICIs could be challenging during the current COVID-19 outbreak **(Fig. 80-5*A*)**. Chest CT manifestations of COVID-19 include an imaging pattern of pure GGO, consolidation, nodules, fibrous stripes, and mixed patterns, with the distribution slightly predominant in the lower lobe and peripheral areas of the lung. Treatment of immune-mediated pneumonitis includes temporary or permanent withdrawal of drug and the addition of high-dose glucocorticoids **(Fig. 80-5*B*)**.

Radiation pneumonitis and/or fibrosis are relatively frequent side effects of thoracic radiation therapy. It may be acute or chronic. Radiation-induced lung toxicity is a function of the irradiated lung volume, dose per fraction, and radiation dose. The larger the irradiated lung field, the higher is the risk for radiation pneumonitis. The use of concurrent chemoradiation, particularly regimens including paclitaxel, increases pulmonary toxicity. Radiation pneumonitis usually develops 2–6 months after completion of radiotherapy. The clinical syndrome, which varies in severity, consists of dyspnea, cough with scanty sputum, low-grade fever, and an initial hazy infiltrate on chest radiographs. The infiltrate and tissue damage usually are confined to the radiation field. The CT scan may show GGOs, consolidation, fibrosis, atelectatic cicatrization, pleural volume loss, or pleural thickening. The patients subsequently may develop a patchy alveolar infiltrate and air bronchograms, which may progress to acute respiratory failure that is sometimes fatal. A lung biopsy may be necessary to make the diagnosis. Asymptomatic infiltrates found incidentally after radiation therapy need not be treated. However, prednisone should be administered to patients with fever or other symptoms. The dosage should be tapered slowly after the resolution of radiation pneumonitis, because abrupt withdrawal of glucocorticoids may cause an exacerbation of pneumonia. Delayed radiation fibrosis may occur years after radiation therapy and is signaled by dyspnea on exertion. Often it is mild, but it can progress to chronic respiratory failure. Therapy is supportive.

Classic radiation pneumonitis that leads to pulmonary fibrosis is due to radiation-induced production of local cytokines such as platelet-derived growth factor β, tumor necrosis factor, interleukins, and transforming growth factor β in the radiation field.

SBRT is a radiotherapy treatment method that has been applied to the treatment of stage I lung cancers in medically inoperable patients. SBRT accurately delivers a high dose of irradiation in one or few treatment fractions to an image-defined lung mass. Most of the acute changes after SBRT occur later than 3 months after treatment, and the shape of the SBRT-induced injury conforms more tightly to the tumor.

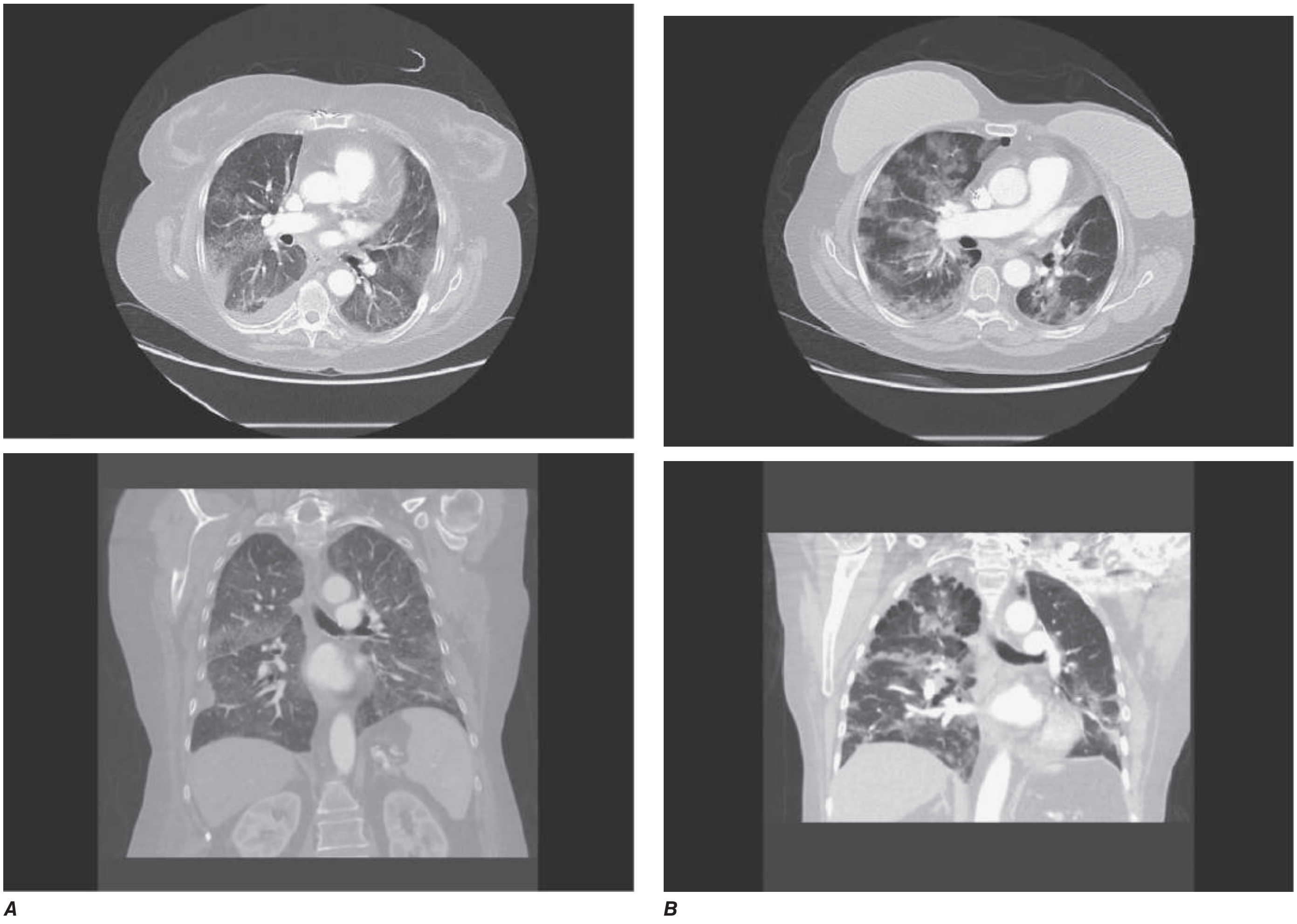

FIGURE 80-5 ***A.*** Computed tomography scan of a 63-year-old female with metastatic adenocarcinoma on nivolumab with immune check point inhibitor pneumonia showing interlobular septal thickening and diffuse ground-glass opacity to nivolumab. ***B.*** Computed tomography scan of a 68-year-old female with resected adenocarcinoma of lung and COVID-19 pneumonia showing peripheral and basilar predominant patchy ground-glass and consolidative opacity consistent with multifocal COVID pneumonia.

Pneumonia is a common problem in patients undergoing treatment for cancer **(Chap 79)**. In patients with pulmonary infiltrates who are afebrile, heart failure and multiple pulmonary emboli are in the differential diagnosis.

NEUTROPENIC ENTEROCOLITIS

Neutropenic enterocolitis (typhlitis) is the inflammation and necrosis of the cecum and surrounding tissues that may complicate the treatment of acute leukemia. Nevertheless, it may involve any segment of the gastrointestinal tract including small intestine, appendix, and colon. This complication has also been seen in patients with other forms of cancer treated with taxanes, 5-fluorouracil, irinotecan, vinorelbine, cisplatin, carboplatin, sacituzumab govitecan, and high-dose chemotherapy **(Fig. 80-6)**. It also has been reported in patients with AIDS, aplastic anemia, cyclic neutropenia, idiosyncratic drug reactions involving antibiotics, and immunosuppressive therapies. The patient develops right lower quadrant abdominal pain, often with rebound tenderness and a tense, distended abdomen, in a setting of fever and neutropenia. Watery diarrhea (often containing sloughed mucosa) and bacteremia are common, and bleeding may occur. Plain abdominal films are generally of little value in the diagnosis; CT scan may show marked bowel wall thickening, particularly in the cecum, with bowel wall edema, mesenteric stranding, and ascites, and may help to differentiate neutropenic colitis from other abdominal disorders such as appendicitis, diverticulitis, and *Clostridium difficile*–associated colitis in this high-risk population. Patients with bowel wall thickness >10 mm on ultrasonogram have higher mortality rates. However, bowel wall thickening is significantly more prominent in patients with *C. difficile* colitis. Pneumatosis intestinalis is a more specific finding, seen only in those with neutropenic enterocolitis and ischemia. The combined involvement of the small and large bowel suggests a diagnosis of neutropenic enterocolitis. Rapid institution of broad-spectrum antibiotics, bowel rest, and nasogastric suction may reverse the process. Use of myeloid growth factors improved outcome significantly. Surgical intervention is reserved for severe cases of neutropenic enterocolitis with evidence of perforation, peritonitis, gangrenous bowel, or gastrointestinal hemorrhage despite correction of any coagulopathy.

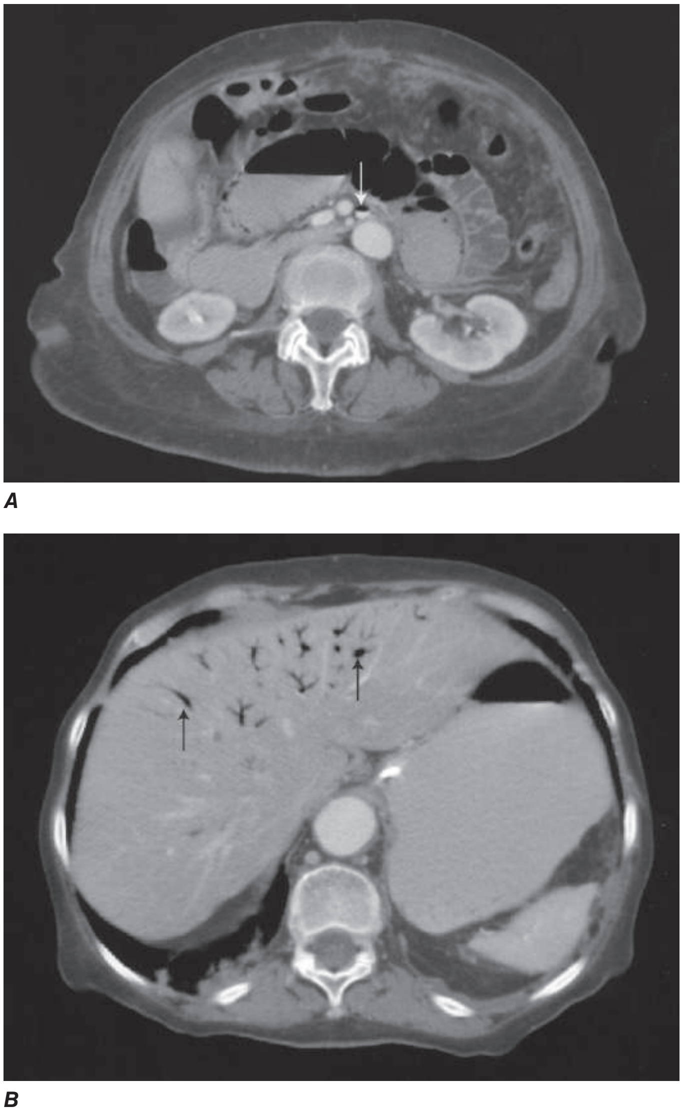

FIGURE 80-6 Abdominal computed tomography (CT) scans of a 72-year-old woman with neutropenic enterocolitis secondary to chemotherapy. ***A.*** Air in inferior mesenteric vein (*arrow*) and bowel wall with pneumatosis intestinalis. ***B.*** CT scan of upper abdomen demonstrating air in portal vein (*arrows*).

C. difficile colitis is increasing in incidence. Newer strains of *C. difficile* produce ~20 times more of toxins A and B compared to previously studied strains. *C. difficile* risk is also increased with chemotherapy. Antibiotic coverage for *C. difficile* should be added if pseudomembranous colitis cannot be excluded.

HEMORRHAGIC CYSTITIS

Hemorrhagic cystitis is characterized by diffuse bladder mucosal bleeding that develops secondary to chemotherapy (mostly cyclophosphamide or ifosfamide), radiation therapy, bone marrow transplantation (BMT), and/or opportunistic infections. Both cyclophosphamide and ifosfamide are metabolized to acrolein, which is a strong chemical irritant that is excreted in the urine. Prolonged contact or high concentrations may lead to bladder irritation and hemorrhage. Symptoms include gross hematuria, frequency, dysuria, burning, urgency, incontinence, and nocturia. The best management is prevention. Maintaining a high rate of urine flow minimizes exposure. In addition, 2-mercaptoethanesulfonate (mesna) detoxifies the metabolites and can be coadministered with the instigating drugs. Mesna usually is given three times on the day of ifosfamide administration in doses that are each 20% of the total ifosfamide dose. If hemorrhagic cystitis develops, the maintenance of a high urine flow may be sufficient supportive care. If conservative management is not effective, irrigation of the bladder with alum or formalin solution may stop the bleeding in most cases. *N*-Acetylcysteine may also be an effective irrigant. Prostaglandin (carboprost) can inhibit the process. In extreme cases, ligation of the hypogastric arteries, urinary diversion, or cystectomy may be necessary.

In the BMT setting, early-onset hemorrhagic cystitis is related to drugs in the treatment regimen (e.g., cyclophosphamide), and late-onset hemorrhagic cystitis is usually due to the polyoma virus BKV or adenovirus type 11. BKV load in urine alone or in combination with acute graft-versus-host disease correlates with development of hemorrhagic cystitis. Viral causes are usually detected by polymerase chain reaction (PCR)–based diagnostic tests. Treatment of viral hemorrhagic cystitis is largely supportive, with reduction in doses of immunosuppressive agents, if possible. No antiviral therapy is approved, although cidofovir was reported to be effective in a small series. Hyperbaric oxygen therapy has been used successfully in patients with BKV-associated and cyclophosphamide-induced hemorrhagic cystitis during hematopoietic stem cell transplantation, as well as in hemorrhagic radiation cystitis that occurs in up to 5% of patients after pelvic radiation.

HYPERSENSITIVITY REACTIONS TO ANTINEOPLASTIC DRUGS

Many antineoplastic drugs may cause hypersensitivity reaction. These reactions are unpredictable and potentially life threatening. Most reactions occur during or within hours of parenteral drug administration. Taxanes, platinum compounds, asparaginase, etoposide, procarbazine, and biologic agents, including rituximab, bevacizumab, trastuzumab, gemtuzumab, cetuximab, and alemtuzumab, are more commonly associated with acute hypersensitivity reactions than are other agents. Hypersensitivity reactions to some drugs, such as taxanes, occur during the first or second dose administered. Hypersensitivity to platinum compounds occurs after prolonged exposure. Skin testing may identify patients with high risk for hypersensitivity after carboplatin exposure.

Premedication with histamine H_1 and H_2 receptor antagonists and glucocorticoids reduces the incidence of hypersensitivity reaction to taxanes, particularly paclitaxel. Despite premedication, hypersensitivity reactions may still occur. In these cases, rapid desensitization in the intensive care unit setting or re-treatment may be attempted with care, but the use of alternative agents may be required. Skin testing is used to assess the involvement of IgE in the reaction. Tryptase levels measured at the time of the reaction help to explain the mechanism of the reaction and its severity. Increased tryptase levels indicate underlying mast cell activation. Candidate patients for desensitization include those who have mild to severe hypersensitivity type I, with mast cell–mediated and IgE-dependent reactions occurring during a chemotherapy infusion or shortly thereafter.

FURTHER READING

Azizi AH et al: Superior vena cava syndrome. JACC Cardiovasc Interv 13:2896, 2020.

Castells M et al: Hypersensitivity to antineoplastic agents: Mechanisms and treatment with rapid desensitization. Cancer Immunol Immunother 61:1575, 2012.

Castinetti F et al: Endocrine side-effects of new anticancer therapies: Overall monitoring and conclusions. Ann Endocrinol (Paris) 79:591, 2018.

Conte P et al: Drug-induced interstitial lung disease during cancer therapies: Expert opinion on diagnosis and treatment. ESMO Open 7:1, 2022.

Durani U, Hogan WJ: Emergencies in haematology: Tumour lysis syndrome. Br J Haematol 188:494, 2020.

Fajgenbaum DC, June CH: Cytokine storm. N Engl J Med 383:2255, 2020.

Gonzalez Castro LN, Milligan TA: Seizures in patients with cancer. Cancer 126:1379, 2020.

Lawton AJ et al: Assessment and management of patients with metastatic spinal cord compression: A multidisciplinary review. J Clin Oncol 37:61, 2019.

Paik WH, Park DH: Endoscopic management of malignant biliary obstruction. Gastrointest Endosc Clin N Am 34:127, 2024

Schusler R, Meyerson SL: Pericardial disease associated with malignancy. Curr Cardiol Rep 20:92, 2018.

Thomas MR, Scully M: How I treat microangiopathic hemolytic anemia in patients with cancer. Blood 137:1310, 2021.

Vogelbaum MA et al: Treatment for brain metastases: ASCO-SNO-ASTRO Guideline. J Clin Oncol 4:492, 2022.

81

Cancer of the Skin

Brendan D. Curti, John T. Vetto,
Sancy A. Leachman

MELANOMA

Pigmented lesions are among the most common findings on skin examination. The challenge for the physician is to distinguish benign lesions from cutaneous melanomas and nonmelanoma skin cancers (NMSCs), both of which are increasing in frequency. Melanoma accounts for over half of the deaths resulting from skin cancer, although recent advances in immunotherapy and targeted therapy used in the neoadjuvant, adjuvant, and advanced disease settings have significantly improved survival. Genomic analysis of melanoma has improved our understanding of prognosis, and informed treatment and surveillance strategies beyond traditional surgical staging. Melanoma is an aggressive malignancy of melanocytes, pigment-producing cells that originate from the neural crest and migrate to the skin, meninges, mucous membranes, upper esophagus, and eyes. Melanocytes in each of these locations have the potential for malignant transformation, but most melanomas arise in the skin, facilitating detection when complete surgical excision can lead to cure. Cutaneous melanoma occurs in people of all ages and all colors. Noncutaneous melanomas have a different biology and a lower probability of response to the immunotherapy agents that have transformed the care of cutaneous melanoma. Examples of malignant melanoma of the skin, mucosa, eye, and nail are shown in **Fig. 81-1**.

RISK FACTORS AND EPIDEMIOLOGY

The epidemiologic patterns seen in melanoma reflect the genetic and biologic features of melanocytes and their response to environmental ultraviolet radiation (UVR). Clinical features that confer an increased risk for melanoma include: (1) vulnerability to sun damage (light/red coloration of skin, hair, or eyes; photodamaged skin; history of exposure to natural or artificial UVR; prior history of skin cancers of any type); (2) abnormal growth of melanocytes (increased absolute number of nevi, increased size of nevi, or atypical features of moles such as multiple colors, speckles, or shapes); and (3) immunosuppression (innate, functional, or drug-induced). **Table 81-1** summarizes melanoma risk factors and the relative risk associated with these factors.

The incidence and mortality rates are strongly influenced by ethnic, geographic, and environmental factors. For instance, the incidence of melanoma is 1/100,000 per year in populations with high skin eumelanin (a brown-black pigment that absorbs ultraviolet [UV] photons efficiently as they enter the epidermis) and up to 27/100,000 per year in populations with low skin eumelanin. Men are affected slightly more than women (1.4:1), and the median age at diagnosis is 66. Melanoma is one several cancer types with increasing incidence in the United States and is now the fifth leading cancer in men (59,170 new cases in 2024; probability 1:28) and the sixth leading cancer in women (41,460 new cases in 2024; probability 1:41). Although these rankings are based on the total number of new invasive melanoma cases in 2024, estimated at 100,640, an additional 99,700 cases of melanoma in situ (MIS) occurred in 2024.

Mortality rates begin to rise at age 55, with the greatest mortality in men age >65 years. In contrast to the increasing incidence, the mortality rates for melanoma are decreasing, though this trend appears less dramatic outside of the United States. The most likely reason for the decreased mortality is the influence of immunotherapy and targeted therapy on melanoma-specific survival. After U.S. Food and Drug Administration (FDA) approval of ipilimumab and vemurafenib in 2011, the 1-year relative survival rate increased from 42% (2008–2010) to 55% (2013–2015). The mortality rate from 2013 to 2017 dropped annually by 7% in those aged 20–64 years old and dropped 5–6% per year for individuals aged ≥65 years.

GLOBAL CONSIDERATIONS

The incidence of both nonmelanoma and melanoma skin cancers around the world has been increasing. Every year, between 2 and 3 million people develop NMSC, and in 2020, there were 324,635 cases of melanoma. A disproportionate number of cases and deaths occur in North America, Europe, Australia, and New Zealand. The highly variable melanoma incidence rates in different populations are due to the interplay between risk factors, including host genetics and environmental factors, which distribute risk unevenly across these populations and account for the absolute risk in different ethnic groups and geographic areas.

Dark-skinned populations (such as those of India and Puerto Rico), blacks, and East Asians also develop melanoma but at rates 10–20 times lower than those in whites. Cutaneous melanomas in dark-skinned populations are more often diagnosed at a higher stage, and patients tend to have worse outcomes. Surveillance, Epidemiology, and End Results (SEER) data (2016–2020) reveal that whites have the highest incidence of melanoma at 37.9 (men) and 25.2 (women) per 100,000 and that the incidence drops substantially in Hispanics (4.5 [men] and 4.3 [women] per 100,000), Native Americans (8.7 [men] and 7.8 [women] per 100,000), Asians/Pacific Islanders (1.3 [men] and 1.1 [women] per 100,000), and blacks (1 [men] and 0.9 [women] per 100,000). In nonwhite (Asian and dark-skinned) populations, the

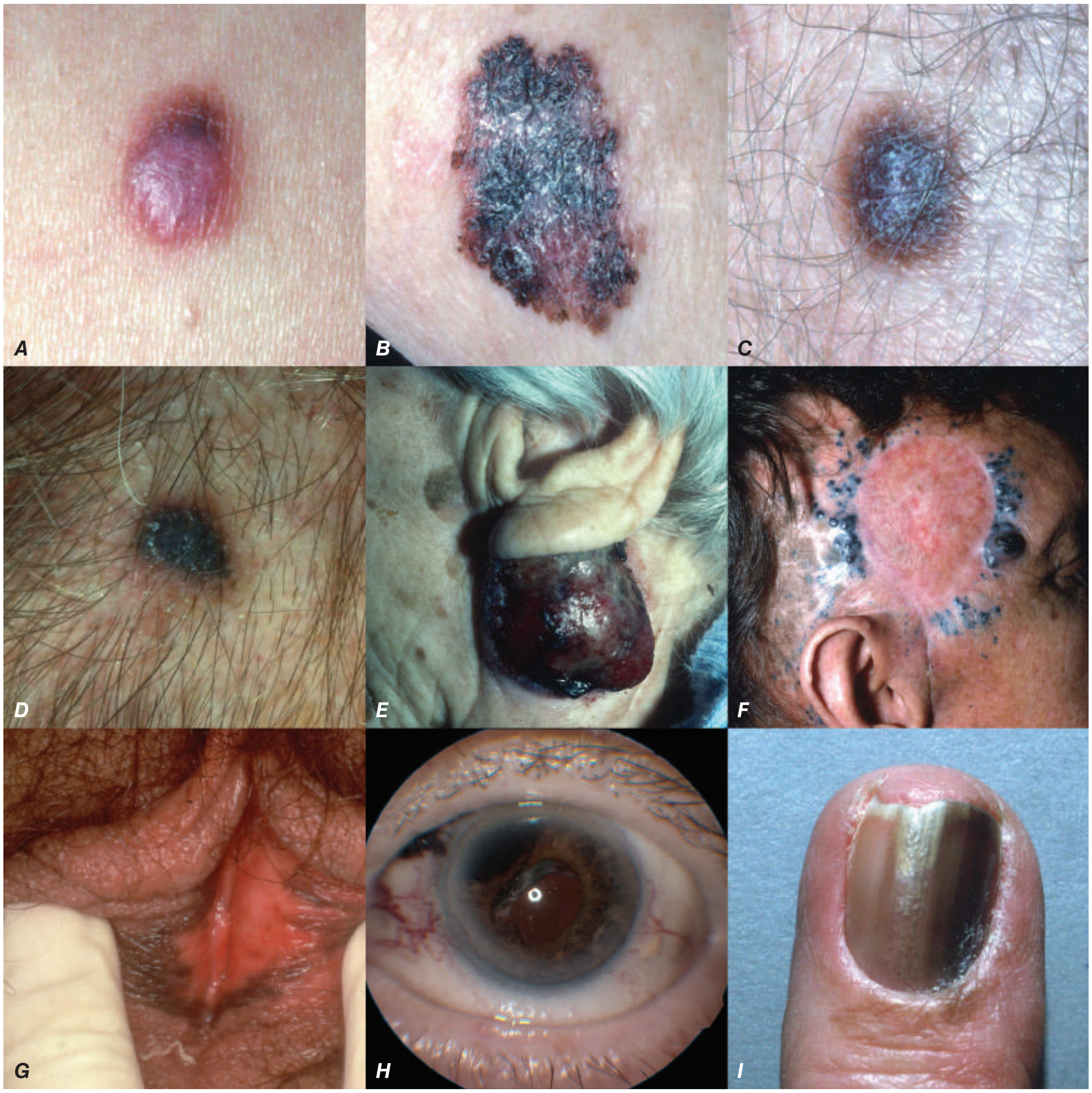

FIGURE 81-1 Types of melanoma. ***A.*** Hypomelanotic melanoma. ***B.*** Superficial spreading melanoma. ***C.*** Melanoma arising in a nevus. ***D.*** Seborrheic keratoses-like melanoma arising on the scalp. ***E.*** Nodular melanoma. **F**. Cutaneous melanoma metastases at a surgical margin (also known as melanoma satellites when <2 cm from the primary tumor and in-transit melanoma when >2 cm). ***G.*** Mucosal melanoma arising in the vulva. ***H.*** Ciliary body melanoma, note visible tumor in the pupil and areas of involvement in the iris and sclera. ***I.*** Acral melanoma with Hutchinson's sign on the proximal nail fold. *(Parts A-G and I photos courtesy of Dr. Leonard Swinyer Collection, © Copyright 2020 University of Utah and Oregon Health & Science University. Part H photo courtesy of Dr. Alison Skalet, © Copyright 2022 Oregon Health & Science University [OHSU].)*

frequency of non–sun-exposed melanomas, such as acral (subungual, plantar, palmar) and mucosal melanomas, is much higher; the incidence of melanoma in black and Hispanic populations is not associated with UV exposure. In China, ~20,000 new melanomas are reported each year, and in contrast to the United States, mortality is increasing. Non–sun-exposed melanomas have a different biology, have a lower probability of response to immunotherapy, and carry a poorer prognosis than cutaneous melanomas, thus accounting for the increase in mortality of this melanoma subgroup. Little is known about the effects of mixed ethnicity on melanoma risk.

GENETIC SUSCEPTIBILITY TO MELANOMA

Approximately 20–40% of hereditary melanomas (0.2–2% of all melanomas) are due to germline mutations in the cell cycle regulatory gene cyclin-dependent kinase inhibitor 2A (*CDKN2A*). In fact, 70% of all cutaneous melanomas have mutations or deletions affecting the *CDKN2A* locus on chromosome 9p21. This locus encodes two distinct tumor-suppressor proteins from alternate reading frames: *p16* and *ARF* (p14^{ARF}). The p16 protein inhibits CDK4/6-mediated phosphorylation and inactivation of the retinoblastoma (RB) protein, whereas *ARF* inhibits MDM2 ubiquitin-mediated degradation of p53. The loss of *CDKN2A* results in inactivation of two critical tumor-suppressor pathways, RB and p53, which control entry of cells into the cell cycle. Several studies have shown an increased risk of pancreatic cancer among melanoma-prone families with *CDKN2A* mutations. A second high-risk locus for melanoma susceptibility, *CDK4*, is located on chromosome 12q13 and encodes the cyclin-dependent kinase inhibited by p16. *CDK4* mutations, which also inactivate the RB pathway, are much rarer than *CDKN2A* mutations. Germline mutations in the melanoma lineage-specific oncogene microphthalmia-associated transcription factor (*MITF*), BRCA1-associated protein 1 (*BAP-1*), protection of telomeres 1 (*POT-1*), and telomerase reverse transcriptase (*TERT*) also predispose to familial melanoma with a not yet quantified high penetrance, based on families that have been tested.

The melanocortin-1 receptor (*MC1R*) gene is a moderate-risk inherited melanoma susceptibility factor. UVR stimulates the production of melanocortin (α-melanocyte-stimulating hormone [α-MSH]), the ligand for *MC1R*, which is a G-protein-coupled receptor that signals via cyclic AMP and regulates the amount and type of pigment produced by melanocytes. *MC1R* is highly polymorphic, and many among its ~80 variants result in partial or full loss of signaling and lead to the production of non-photoprotective red/yellow pheomelanins, rather than photoprotective brown/black eumelanins. The red hair color (RHC) phenotype produced by *MC1R* mutations includes lightly

TABLE 81-1 Melanoma Risk Factors and Relative Risk

RISK LEVEL	RISK FACTOR	RELATIVE RISK
Elevated	1 atypical nevus versus 0	1.5
	Total common nevi, 16+ versus <15	1.5
	Blue eye color versus dark	1.5
	Hazel eye color versus dark	1.5
	Green eye color versus dark	1.6
	Light brown hair versus dark	1.6
	Indoor tanning in any gender versus never	1.7
	Fitzpatrick skin type II versus IV	1.8
	Fitzpatrick skin type III versus IV	1.8
	History of sunburn versus no sunburn	2.0
	Blond hair versus dark	2.0
	2 atypical nevi versus 0	2.1
	Fitzpatrick skin type I versus IV	2.1
	High density of freckles versus none	2.1
	Total common nevi 41–60 versus <15	2.2
Moderately elevated	Family history of melanoma in 1 or more first-degree relatives	1.7–3.0
	3 atypical nevi versus 0	3.0
	Total common nevi 61–80 versus <15	3.3
	Red hair versus dark	3.6
	Chronic lymphocytic leukemia	3.9
	History of actinic keratoses and/or keratinocyte carcinoma versus not	4.3
	Indoor tanning in women aged 30–39 versus never	4.3
	4 atypical nevi versus 0	4.4
High	Transplant recipient versus not	2.2–4.6
	Indoor tanning in women aged <30 versus never	6.0
	5 atypical nevi versus 0	6.4
	Total common nevi 81–120 versus <15	6.9
	Personal history of melanoma	8.2–13.4
	CDK2NA mutation carrier	14–28

colored skin, red hair, freckles, increased sun sensitivity, and increased risk of melanoma. In addition to its weak UV-shielding capacity relative to eumelanin, increased pheomelanin production in patients with inactivating polymorphisms of *MC1R* also provides a UV-independent carcinogenic contribution to melanomagenesis via oxidative damage and reduced DNA damage repair.

Other more common, low-penetrance polymorphisms in genes related to pigmentation, nevus count, immune responses, DNA repair, metabolism, and the vitamin D receptor have small effects on melanoma susceptibility. In sum, ~50–60% of the genetic risk for hereditary melanoma can be attributed to known melanoma predisposition genes, with ~40% of the known genetic risk attributable to *CDKN2A*. The other components of inherited risk are most likely due to the presence of additional modifier genes and/or shared environmental exposures of the host.

PREVENTION AND EARLY DETECTION

Primary prevention of melanoma and NMSC is based on protection from the sun. Public health initiatives, such as the SunSmart program that started in Australia and is now operative in Europe and the United States, have demonstrated that behavioral change can decrease the incidence of NMSC and melanoma. Preventive measures should start early in life because damage from UV light begins early even though cancers develop years later. Early episodes of sunburns may be a greater risk than chronic tanning. Some individuals tan compulsively. There is now greater understanding of tanning addiction and the cutaneous-neural connections that may give rise to this behavior. Compulsive tanners exhibit differences in dopamine binding and reactivity in reward pathways in the brain, such as the basal striatum, resulting in cutaneous secretion of β-endorphins after UV exposure. Identifying individuals with tanning addiction may be another prevention method. Regular use of broad-spectrum sunscreens that block UV-A and UV-B with a sun protection factor (SPF) of at least 30 and protective clothing should be encouraged. Physical blockers such as zinc oxide and titanium dioxide have less likelihood of being absorbed or of generating an allergic reaction than chemical sunscreens. Avoidance of sunburns, tanning beds, and midday sun exposure is recommended.

Secondary prevention comprises education and screening with the goal of early detection and can be individualized based on risk factors. A full-body skin exam is warranted in populations at higher risk for melanoma such as patients with clinically atypical moles (dysplastic nevi) and those with a personal history of melanoma. Surveillance in high-risk patients should be performed by a dermatologist and include total-body photography and dermoscopy where appropriate. Individuals with three or more primary melanomas and families with at least one invasive melanoma and two or more cases of melanoma and/or pancreatic cancer, ocular melanoma, mesothelioma, or renal cell carcinoma among first- or second-degree relatives on the same side of the family may benefit from genetic testing. Atypical nevi and MIS should be completely removed with at least a 5-mm margin. Early detection of small lesions allows the use of simpler treatment modalities with higher cure rates and lower morbidity. Monthly self-screening augments provider-based screening. Patients should be taught to recognize the clinical features of melanoma and advised to report any change in a pigmented lesion. There is evidence supporting the ability of media campaigns to reduce cancer mortality in lung cancer, and results from Australia's skin cancer campaigns demonstrate improvement in attitude and behavior and a reduction in melanoma incidence. A benefit/cost analysis in Australia showed a return of $3.85 for every $1 invested. Although the U.S. Preventive Services Task Force states that there is insufficient evidence to recommend skin screening for the general population, additional research is anticipated to find best practices for skin cancer detection and prevention.

DIAGNOSIS

Early detection of melanoma before it becomes invasive and life-threatening metastases have occurred is essential and may be facilitated by applying the ABCDEs: *a*symmetry (benign lesions are usually symmetric); *b*order irregularity (most nevi have clear-cut borders); *c*olor variegation (benign lesions usually have uniform light or dark pigment); *d*iameter >6 mm (the size of a pencil eraser); and *e*volving (any change in size, shape, color, or elevation or new symptoms such as bleeding, itching, and crusting). In addition, any nevus that appears atypical and different from the rest of the nevi on that individual (an "ugly duckling") should be considered suspicious.

The entire skin surface, including the scalp and mucous membranes, as well as the nails should be examined in each patient. Bright room illumination is important, and a hand lens or dermatoscope is helpful for evaluating variation in pigment pattern. Any suspicious lesions should be biopsied, evaluated by a specialist, or recorded by chart and/or photography for follow-up. Dermoscopy employs low-level magnification of the epidermis with polarized light or water interface and permits a more precise visualization of patterns of pigmentation than is possible with the naked eye.

Biopsy Any pigmented cutaneous lesion that has changed on examination or has the other features previously discussed is a candidate for biopsy. An excisional biopsy with 1- to 3-mm margins (narrow-margin excision) is suggested. This facilitates histologic assessment of the lesion, permits accurate measurement of thickness if the lesion is melanoma, and constitutes definitive treatment if the lesion is benign. For lesions that are large or involving anatomic sites where excisional biopsy may not be feasible (such as the face, hands, and feet) or when suspicion of malignancy is low, an incisional biopsy (e.g. shave, saucerization, or punch) to include the most nodular or darkest area of the lesion is acceptable. Incisional biopsy does not appear to facilitate the spread of melanoma. For suspicious lesions, every attempt should be made to preserve the ability to assess the deep and peripheral margins and to perform immunohistochemistry. All biopsies should be deep enough to include the deepest component of the entire lesion, and any pigment at the base of the lesion should be removed and included with the biopsy specimen. Punch biopsies are more likely to clear the deep margin but more likely to be positive at the radial margins; the opposite is true for shave biopsies. The choice of biopsy type should be guided by which is most likely to remove the entire lesion for histologic evaluation.

The biopsy should be interpreted by a pathologist experienced in pigmented lesions, and the report should include Breslow thickness, mitotic rate, presence or absence of ulceration, lymphatic/vascular/neural invasion, regression, microsatellitosis, and the status of the peripheral and deep margins. Breslow thickness is the greatest thickness of a primary cutaneous melanoma measured on the slide from the top of the epidermal granular layer, or from the ulcer base, to the bottom of the tumor. To distinguish melanomas from benign nevi in challenging cases, genetic expression profiles (GEPs), fluorescence in situ hybridization with multiple probes, or comparative genomic hybridization can be helpful. GEPs have also been developed to determine prognosis.

CLASSIFICATION AND PATHOGENESIS

Clinical More recent classifications of melanoma are based on association with cumulative solar damage and nine different pathways related to genomic attributes summarized in Table 81-2. This revised classification incorporates traditional histopathologic designations such as superficial spreading, lentigo maligna, acral lentiginous, and desmoplastic, among others, but more precisely incorporates the pathophysiology and genetic drivers of melanoma subtypes.

At present, genomic alteration pathways for melanomas are not incorporated into American Joint Committee on Cancer (AJCC) staging or prognostic considerations, yet characterizing the genomic and mutational profiles of melanoma has become increasingly common in clinical practice. It is anticipated that genomic pathway characterization will become increasingly important in determination of melanoma prognosis and may influence surveillance strategies, surgical decisions, and medical therapy.

Genomic The advent of next-generation sequencing has led to whole exome sequencing of thousands of cutaneous melanomas derived from nonglabrous (hair-bearing) skin. This has revealed very complex genomic changes resulting from both germline (see "Genetic Susceptibility to Melanoma" above) and somatic mutations. Cutaneous melanomas have one of the highest somatic mutation rates (>10 mutations/Mb) among all cancers; the majority (76% of primary tumors and 84% of metastatic melanomas) exhibit mutations indicative of UVR exposure. The mutation rate varies based on body site; melanomas arising in chronic sun-damaged skin harbor substantially more mutations than melanomas from non–sun-damaged skin.

Melanomas can harbor thousands of mutations, but only a few are "driver" mutations that promote cell proliferation or inhibit normal pathways of apoptosis or DNA repair and confer a growth advantage to the neoplastic cell. Some of the driver mutations for cutaneous melanoma are depicted in Fig. 81-2 along with the clinical evolution of melanoma lesions. Driver mutations are often found in combination with mutations to germline susceptibility genes such as *p16*, which affect cell cycle arrest, and *ARF*, which result in defective apoptotic responses to genotoxic damage. As melanocytes accumulate DNA damage, they can undergo malignant transformation characterized by invasion, metastasis, and angiogenesis.

A genomic classification of cutaneous melanoma has been proposed based on the pattern of the most prevalent mutated genes, *BRAF*, *RAS*, and *NF1*, along with a triple wild type (WT), lacking mutations in these three genes. The pattern of DNA mutations can vary with the site of origin and should be determined along with the histologic subtype of the tumor. The proliferative pathways affected by the mutations include the mitogen-activated protein (MAP) kinase and phosphatidylinositol 3′ kinase/AKT pathways. *RAS* and *BRAF*, members of the MAP kinase pathway, which mediates the transcription of genes involved in cell proliferation and survival, undergo somatic mutation in melanoma and thereby represent potential therapeutic targets. *NRAS* is mutated in ~20% of melanomas, and somatic activating *BRAF* mutations are found in most benign nevi and 40–50% of cutaneous melanomas. Neither mutation by itself appears to be sufficient to cause melanoma; thus, they often are accompanied by other mutations, such as in *TERT*. The *BRAF* mutation is most commonly a T→A point mutation that results in a valine-to-glutamate amino acid substitution (V600E). V600E *BRAF* mutations are more common in younger patients and are present in most melanomas that arise on skin with intermittent sun exposure and are less common in melanomas from chronically sun-damaged skin (i.e., those of older patients).

Melanomas may harbor mutations in *AKT* (primarily in *AKT3*) and *PTEN* (phosphatase and tensin homolog). *AKT* can be amplified, and *PTEN* may be deleted or undergo epigenetic silencing that leads to constitutive activation of the PI3K/AKT pathway and enhanced cell survival by antagonizing the intrinsic pathway of apoptosis. A loss-of-function mutation in *NF1*, which can affect both the MAP kinase and PI3K/AKT pathways, has been described in 10–15% of melanomas.

TABLE 81-2 Major Histologic Subtypes of Malignant Melanoma

TYPE	SITE	APPEARANCE	ASSOCIATED MUTATIONS
Lentigo maligna	Sun-exposed surfaces, particularly malar region and temple	In flat portions, brown and tan predominate, but whitish gray sometimes present; in nodules, reddish brown, bluish gray, bluish black.	*BRAF* 28% *NRAS* 15% *PTEN*
Superficial spreading	Any (more common on upper back and, in women, lower legs)	Brown mixed with bluish red, bluish black, reddish brown, and often whitish pink. The lesion border is often visibly and/or palpably raised.	*BRAF* 57% *NRAS* 18%
Nodular	Any	Reddish blue, purple, or bluish black; can be uniform or mixed with brown and black.	*BRAF* 47% *NRAS* 33%
Acral lentiginous	Palm, sole, nail bed, mucous membrane	In flat portions, dark brown; in raised lesions (plaques), brown-black or blue-black.	*NRAS* 25% *c-KIT* 5-10% *BRAF* 10%
Desmoplastic	Any (more common on head and neck)	Highly variable; pigmentation is frequently absent. Can mimic nodular basal cell carcinoma.	*MAPK* and *PI3K* 73% High tumor mutational burden, *BRAF* and *NRAS* uncommon
Uveal	Choroid, ciliary body, iris	Dome or mushroom shaped. Display low internal reflectivity on ocular ultrasound.	BAP1, GNAQ, GNA11
Mucosal	Oral cavity, conjunctiva, sinuses, alimentary tract including rectum and anus, vulva	Can display radial growth pattern with ABCDE features associated with cutaneous melanomas. Often present with advanced tumors infiltrating local tissues	KIT, NRAS, KRAS or BRAF NF1

Abbreviation: ABCDE, asymmetry (benign lesions are usually symmetric); border irregularity (most nevi have clear-cut borders); color variegation (benign lesions usually have uniform light or dark pigment); diameter >6 mm (the size of a pencil eraser); and evolving (any change in size, shape, color, or elevation or new symptoms such as bleeding, itching, and crusting).

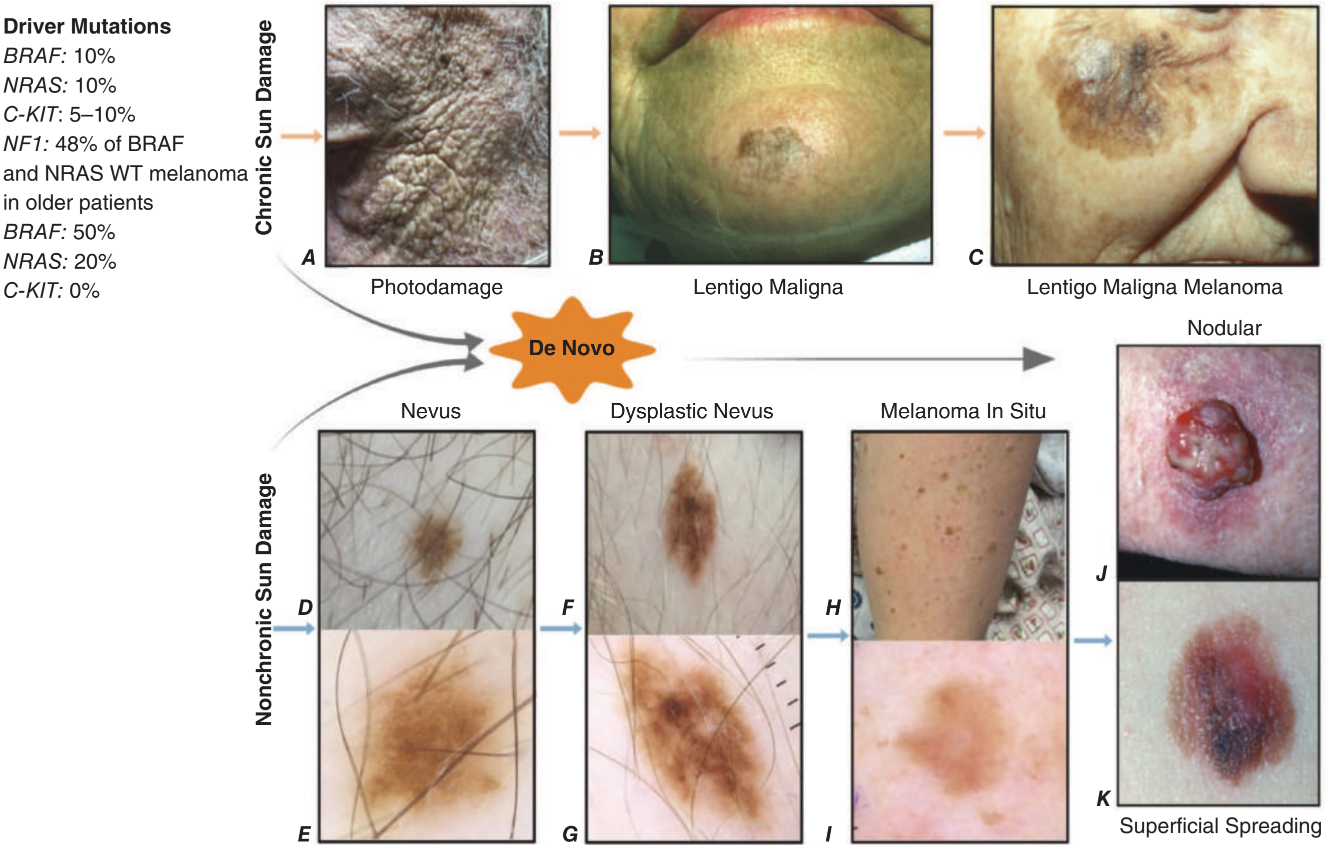

FIGURE 81-2 Cutaneous melanoma development and associated driver mutations. Chronic sun damage (with prominent solar elastosis) (***A***) predisposes to a lentigo maligna (in situ) (***B***), which can evolve into lentigo maligna melanoma (invasive) (***C***). Similarly, nonchronic sun damage can initiate melanoma de novo or in nevomelanocytes, where clinical and histologic changes of atypia may be seen prior to complete transformation. Nevi (***D, E***) can evolve into atypical lesions (***F, G***), in situ melanoma (***H, I***), and eventually invasive nodular (***J***) or superficial spreading melanomas (***K***). Images E, G, and I are dermascopic photos of images D, F, and H, respectively. *(Part A photo courtesy of Dr. Sancy Leachman, © Copyright 2022 Oregon Health & Science University [OHSU]. Parts B, C, J, and K photos courtesy of Dr. Leonard Swinyer Collection, © Copyright 2020 University of Utah and OHSU. Parts D–I photos courtesy of Dr. Elizabeth Berry, © Copyright 2022 OHSU.)*

In melanoma, these two signaling pathways (MAP kinase and PI3K/AKT) enhance tumorigenesis, chemoresistance, migration, and cell cycle dysregulation.

PROGNOSTIC FACTORS

The most important clinical prognostic factors for a newly diagnosed patient are incorporated in the AJCC staging classification. The best predictor of recurrence is Breslow thickness, followed by ulceration, which together make up the T stage for melanoma. The anatomic site of the primary tumor also influences prognosis; favorable sites are the forearm and leg, and unfavorable sites include the scalp, hands, feet, and mucous membranes. Women with stage I or II disease have better survival than men, perhaps in part because of earlier diagnosis; women frequently have melanomas on the lower leg, where self-recognition is more likely compared to the back, where melanoma is more likely in men.

Older individuals, especially men >60, have a tendency toward delayed diagnosis (and thus thicker tumors), have more head and neck and acral melanomas (which tend to have earlier vertical growth and distant metastases), and are more likely to develop melanomas in chronically UVR-damaged skin (which are more often *BRAF* wild type, with fewer options for therapy). All these factors help explain the worse prognosis in older males. Other important adverse factors include high mitotic rate and lymphatic/vascular invasion. Clinical features such as microsatellite lesions and/or in-transit metastases, evidence of nodal involvement, elevated serum lactate dehydrogenase (LDH), and certain sites of distant metastases (e.g., brain, liver, and bone) all portend a higher stage and worse prognosis.

GEPs and machine-learning algorithms that associate genomic changes with clinical outcomes have been used to estimate the prognosis of melanoma. A commercially available 31-gene GEP is available that predicts for all-site (particularly distant) relapse and incorporates the increased and decreased expression, as well as the dysregulation, of genes involved in many of the cellular processes leading to melanoma progression described earlier. Although this 31-gene GEP can estimate the probability of distant relapse, it has not supplanted the prognostic estimates derived from surgical staging. GEPs have been incorporated into management guidelines for breast, thyroid, and other cancers, but their use in cutaneous melanoma care is still under investigation.

STAGING

The purpose of staging is to estimate melanoma prognosis and determine treatment selection. The current melanoma staging criteria and estimated 10-year survival by stage are depicted in **Table 81-3**. The clinical stage is determined after the microscopic evaluation of the melanoma skin lesion and clinical and radiologic assessment. The pathologic stage incorporates the results from microscopic examination of clinically negative regional lymph nodes obtained at sentinel lymph node biopsy (SLNB), any enlarged nodes found on exam or imaging, and any suspected metastases amenable to open or image-guided biopsy.

All patients should have a complete history, with attention to symptoms that suggest metastatic disease, such as new palpable masses, malaise, weight loss, headaches, vision changes, alterations in bowel habits, hemoptysis, and pain. The provider should look for persistent melanoma at the biopsy site, dermal or subcutaneous nodules that could represent satellite or in-transit metastases, and lymphadenopathy. A complete blood count, complete metabolic panel, and LDH should be performed. Although these tests rarely lead to detection of occult metastatic disease, a microcytic anemia would raise the possibility of bowel metastases, elevated liver function tests can suggest liver metastases, and LDH is part of the AJCC system for stage IV disease. Abnormal test results should prompt a more extensive evaluation, including computed tomography (CT) scan or a positron emission tomography (PET) scan (or CT/PET combined). Magnetic resonance

TABLE 81-3 Staging and Survival

STAGE	TNM	10-YEAR MELANOMA-SPECIFIC SURVIVAL ESTIMATE
0	TisN0M0	>99%
IA	T1aN0M0, T1bN0M0	98%
IB	T2aN0M0	94%
IIA	T2b-T3aN0M0	88%
IIB	T3b-T4aN0M0	81–83%
IIC	T4bN0M0	75%
IIIA	T1a-T2aN1a-2aM0	71–88%
IIIB	T2b-T3aN1a-N2bM0	60–77%
IIIC	T3b-4bN1a-N3cM0	44–60%
IIID	T4bN3a-N3cM0	24–30%
IV M1a	Any T, any N, skin, soft tissue, or distant nodal sites	50% at 5 years
IV M1b	Any T, any N, lung + any M1a sites	35–50% at 5 years
IV M1c	Any T, any N, skin, non-CNS visceral disease, any M1a or M1b sites	~25% at 5 years
IV M1d	Any T, any N, CNS metastasis + any M1a,b,c sites	<5% at 5 years

Abbreviations: CNS, central nervous system; TNM, tumor-node-metastasis.

imaging (MRI) of the brain with contrast is recommended for the initial evaluation of patients who present with neurologic symptoms or have advanced disease on imaging or examination.

Despite all the above considerations, >80% of patients at presentation will have disease confined to the skin and a negative history and physical examination, in which case imaging is not indicated. One study suggests that imaging should be considered for node-negative low-stage melanoma with a high-risk GEP, but this is not yet standard. Imaging is sometimes done for very-high-risk primaries (e.g., >4 mm with ulceration, clinical stages IIB and IIC) in which the chance for occult distant metastases is higher than that for a positive SLNB, and the prognosis is worse compared to stage IIIA disease. Medical oncologists now routinely provide consultation for stage IIB and IIC patients to assess the potential value of adjuvant therapy (see "Treatment").

TREATMENT

Melanoma

MANAGEMENT OF CLINICALLY LOCALIZED MELANOMA (STAGE I, II)

For a newly diagnosed cutaneous melanoma, surgical wide excision (WE) of the lesion with a margin of normal skin is necessary to remove all malignant cells and minimize the probability of local recurrence. The National Comprehensive Cancer Network (NCCN), based on data from six randomized trials, recommends the following radial margins for a primary MIS, 0.5–1.0 cm; invasive up to 1 mm thick, 1 cm; >1.01–2 mm, 1–2 cm; and >2 mm, 2 cm. Smaller margins may be used for "anatomically constrained" locations such as the face, hands, feet, and genitalia due to the higher likelihood of surgical morbidity in these regions, and in some instances, Mohs with immunostaining is advantageous. In all instances, inclusion of subcutaneous fat in the surgical specimen facilitates adequate thickness measurement and assessment of surgical margins by the pathologist. When feasible, excision should go down to fascia, with fascial resection for thick (T4) lesions. Topical imiquimod, a toll-like receptor agonist, can stimulate skin macrophages to induce an immune response useful to treat lentigo maligna in cosmetically sensitive locations with narrow resection margins by promoting local immune response resulting in decreased local recurrence.

SLNB provides prognostic information to identify patients at high risk for relapse who may be candidates for adjuvant therapy. The first (sentinel) draining node(s) from the primary site is (are) located by injecting a blue dye and a gamma-emitting radioisotope around the primary site. The sentinel node(s) then is (are) identified using a handheld gamma detector brought sterilely into the operative field. The surgeon makes an incision of the area of uptake and looks for the blue-stained, "hot" node(s), which is (are) removed and subjected to histopathologic analysis with serial sectioning using hematoxylin and eosin and immunohistochemical stains (e.g., S100, HMB45, MART-1, and MelanA) to identify melanocytes.

NCCN guidelines recommend SLNB for patients with a 10% or greater chance of having tumor in the node. This includes patients with tumors >1 mm thick (T2) or T1 tumors that have ulceration (T1b). Patients with a 5–10% risk of node positivity (NCCN "Discuss and Consider" category), such as those with tumors measuring between 0.75 and 1.0 mm, transected tumors, regressed tumors, or lymphovascular invasion, should also be considered for SLNB. The NCCN does not recommend SLNB for patients with a risk of a positive SLNB ≤5% such as those with melanomas ≤0.75 mm thick and no high-risk features. In these patients, WE alone is the usual definitive therapy. There are computer nomograms that estimate the risk of sentinel lymph node involvement based on melanoma depth, clinical features (age, site), and histology (ulceration, mitotic rate, lymphovascular invasion). GEPs in combination with these other factors are being investigated as a sentinel lymph node risk assessment tool in ongoing prospective trials.

Patients with negative SLNB can either be followed or considered for adjuvant therapy if the primary lesion is considered high risk. Patients with thick and/or ulcerated stage IIB or IIC melanomas have a significant risk of recurrence after wide local excision and (negative) SLNB, estimated at 13–18% probability of death at 5 years using the AJCC melanoma database. Adjuvant anti-PD-1 immunotherapy using pembrolizumab or nivolumab for 1 year significantly reduces the risk of melanoma recurrence or death in stage IIB or IIC melanoma and has become a standard of care.

Patients with a positive sentinel lymph node should undergo CT or PET/CT imaging to rule out distant metastatic disease, and if none is found (i.e., stage III), adjuvant therapy should be offered (see next section). Complete lymphadenectomy following identification of a positive sentinel lymph node improves relapse-free but not overall survival, and therefore, it is no longer offered routinely, but should be considered in patients who cannot comply with follow-up and/or forgo adjuvant therapy. This avoids the morbidity of regional node dissection in most patients. However, patients not undergoing immediate completion node dissection should have nodal bed surveillance with physical examination and nodal bed imaging (ultrasound or CT) at 4- to 6-month intervals for approximately 3 years to rule out isolated nodal bed progression.

Mohs micrographic surgery (MMS) is an alternative to WE and is particularly useful in areas where tissue preservation is important (e.g., face, genitalia, hands) and for lesions with extensive MIS. MMS improves the probability of achieving negative margins, decreases local recurrence, and enhances cosmesis compared to WE in selected patients. It does not preclude SLNB, which can be done before the MMS procedure.

MANAGEMENT OF REGIONALLY METASTATIC MELANOMA (STAGE III)

Stage III melanoma comprises patients with a positive sentinel lymph node, resected regional nodal macrometastases, or resected locoregional disease (e.g., recurrences in the wide excision site, within 2 cm of the site ["satellite metastases"], or >2 cm from the site ["in-transit metastases"]). Even after complete resection of stage III disease, the risk of developing distant metastases (stage IV) may be high, and adjuvant systemic therapy should be offered. Melanomas may recur at the edge of the incision or graft, as satellite metastases, in-transit metastases, or most commonly, regional spread to a draining lymph node basin. Each of these presentations is managed surgically followed by postsurgical adjuvant systemic immunotherapy or targeted therapy (for *BRAF*-mutant tumors), after which there is the possibility of long-term disease-free survival. Topical therapy with imiquimod has been useful for patients with low-volume dermal lesions, but survival benefit has not been confirmed with this approach. Talimogene laherparepvec is an engineered, oncolytic herpes simplex virus type 1 that is FDA approved for injection of primary or recurrent melanomas including cutaneous and subcutaneous lesions or lymph node deposits that cannot be completely removed by surgery.

Radiotherapy can reduce the risk of local recurrence after lymphadenectomy but does not improve overall survival. Patients with large nodes (>3–4 cm), four or more involved lymph nodes, or extranodal spread on microscopic examination should be considered for radiation as local recurrence in these high-risk patients has significant morbidity. Systemic adjuvant therapy can also be considered for patients with completely resected stage IV disease.

Current options for adjuvant systemic therapy include anti-PD-1 (nivolumab or pembrolizumab) or targeted therapy with BRAF/MEK inhibitors in melanomas that harbor a *BRAF* V600 mutation. Both anti-PD-1 and targeted therapy have been shown to confer disease-free and overall survival benefits in patients with stage III and stage IV melanoma (see below for further discussion).

A subset of patients with stage III melanoma has bulky disease (usually palpable nodal involvement) at presentation (stages IIIC and IIID) that may be difficult to resect with negative margins. Even if surgery is feasible and postsurgical adjuvant immunotherapy or targeted therapy is offered, the prognosis of these patients is poor. A recent randomized phase II study comparing neoadjuvant therapy followed by resection and adjuvant pembrolizumab versus resection followed by adjuvant pembrolizumab showed significant improvements in event-free and overall survival for the neoadjuvant approach. Other randomized phase II clinical trials investigating neoadjuvant plus adjuvant ipilimumab and nivolumab or relatlimab plus nivolumab in patients with palpable nodal disease at diagnosis have demonstrated a >50% probability of achieving a pathologic complete response with neoadjuvant treatment and a low probability of recurrence at 1 and 2 years. Neoadjuvant plus adjuvant targeted therapy in patients with stage IIIC or IIID *BRAF*-mutated melanoma has demonstrated a similar high probability of achieving pathologic complete response and improved event-free survival in single-arm phase II studies. Long-term follow-up data from randomized controlled studies of neoadjuvant immunotherapy are not yet mature, and comparisons of immunotherapy to targeted therapy in the neoadjuvant setting have not yet been performed. GEP may help to identify patients with stages II or III melanoma who are at *lower* risk of recurrence and could avoid the toxicity and expense of adjuvant therapy, although prospective data on this approach are needed.

TREATMENT

Metastatic Disease

At diagnosis, 84% of patients with melanoma will have stage I or II disease and 4% will present with metastases. Many others will develop metastases after initial therapy for locoregional disease; 60% of deaths from melanoma occur in patients who were initially diagnosed as stage I or II. The probability of recurrence is related to initial stage, ranging from <5% with stage IA to >90% for subsets of patients with stage IIID disease at presentation. Patients with a history of melanoma who develop signs or symptoms suggesting recurrent disease should undergo restaging imaging as described earlier. Distant metastases (stage IV) commonly involve skin and lymph nodes as well as viscera, bone, or the brain. The prognosis is better for patients with skin and subcutaneous metastases (M1a) than for lung (M1b) and worst for those with metastases to bone or other visceral organs (M1c) or brain (M1d). An elevated serum LDH is a poor prognostic factor and places the patient in stage M1c regardless of the metastatic sites. The 15-year survival of patients with stage IV melanoma was <10% before 2010; however, the development of targeted therapy and immunotherapy has improved disease-free and overall survival, especially for patients with M1a and M1b disease, in whom the 15-year survival is nearly 50%. Even patients with M1c disease may have prolonged survival, and those who are progression-free for >2 years after immunotherapy or targeted therapy have a high probability of living >5 years from the onset of metastasis; some of these individuals may be cured.

TABLE 81-4 Treatment Options for Metastatic Melanoma

- Immunotherapy
  - Immune checkpoint blockade
    - Anti-PD-1: pembrolizumab or nivolumab
    - Anti-CTLA-4: ipilimumab
    - Combined ipilimumab and nivolumab
    - Combined relatlimab (anti-LAG-3) and nivolumab
  - T-cell engager
    - Tebentafusp (selected patients with uveal melanoma)
  - Cytokine-based immunotherapy
    - High-dose interleukin 2
- Clinical trials investigating adoptive cellular therapy with tumor-infiltrating lymphocytes for advanced disease and personalized vaccine targeting neoantigens in high-risk resected melanoma
  - Oncolytic virus
    - Talimogene laherparepvec
  - Targeted therapies
    - BRAF inhibitors: vemurafenib, dabrafenib, encorafenib
    - MEK inhibitors: trametinib, cobimetinib, binimetinib
  - Local modalities
    - Surgery
    - Stereotactic radiation

FDA-approved agents since 2011 include three immune T-cell checkpoint inhibitors (ipilimumab, nivolumab, and pembrolizumab), combination immunotherapy (ipilimumab plus nivolumab or relatlimab plus nivolumab), six oral agents that target the MAP kinase pathway (the BRAF inhibitors vemurafenib, dabrafenib, and encorafenib, and the MEK inhibitors trametinib, cobimetinib, and binimetinib), and the oncolytic virus talimogene laherparepvec. Adoptive cellular therapy using tumor-infiltrating lymphocytes (TILs) administered with interleukin 2 is undergoing FDA review and may become a new standard of care for patients with progression on checkpoint immunotherapy **(Table 81-4)**.

Local modalities, such as surgery and stereotactic radiosurgery, should be considered for patients with a limited number of metastatic sites (oligometastatic disease) because they may experience long-term disease-free survival after metastasectomy or ablative high-dose-per-fraction radiation. Patients with solitary metastases are the best candidates, but local modalities can also be offered to patients with metastases at more than one site if a complete resection or treatment of all sites can be achieved with reasonable side effects. Patients rendered free of disease can be considered for adjuvant therapy or a clinical trial because their risk of developing additional metastases remains high. Surgery can also be used as an adjunct to systemic therapy if only one or a few oligometastases remain after systemic therapy. Surgery can be used to obtain tumor for mutational profile analysis or to harvest tumor for TIL therapy.

IMMUNOTHERAPY

Immune Checkpoint Blockade Immunotherapies are based on an understanding of the control mechanisms of the normal immune response. Inhibitory receptors or checkpoints, including CTLA-4, PD-1, and lymphocyte activation gene 3 (LAG-3), are upregulated on T cells after engagement of the T-cell receptor by cognate tumor antigen in the context of the appropriate class I or II human leukocyte antigen (HLA) molecules during the interaction between a T cell and antigen-presenting cell. Immune checkpoints are needed to ensure proper regulation of a normal immune response; however, the continued expression of inhibitory receptors during chronic infection (hepatitis, HIV) and in cancer patients leads to exhausted T cells with limited potential for proliferation, cytokine production, or cytotoxicity. Ipilimumab, a fully human IgG1 antibody that binds CTLA-4 and blocks inhibitory signals, was the first drug shown in a randomized trial to improve survival in patients with metastatic melanoma. Anti-CTLA-4 monotherapy has been supplanted by anti-PD-1 monotherapy or checkpoint combinations due to higher objective response rates and longer durations of response.

The PD-1 blockers, nivolumab and pembrolizumab, have been approved to treat patients with advanced melanoma. Combination T-cell checkpoint therapy, blocking both inhibitory pathways with ipilimumab and nivolumab, leads to superior antitumor activity compared to treatment with either agent alone. Combined therapy with IV ipilimumab and nivolumab is administered in the outpatient setting every 3 weeks for four doses (induction), followed by nivolumab given every 2–4 weeks (maintenance) for up to 1 year, and is associated with an objective response rate of 56% and enhanced survival compared to ipilimumab monotherapy. Patients who have >5% expression of PD-1 on T cells in a melanoma biopsy sample derive a similar level of clinical benefit from nivolumab monotherapy, although using PD-1 expression to select therapy remains problematic as some patients whose melanoma has no detectable PD-1 expression can still respond to immunotherapy. Other elements of genomic analysis, including an estimate of tumor mutational burden (TMB), can be clinically useful as TMB is correlated with a higher probability of objective response and longer progression-free survival in patients treated with checkpoint antibody therapy in melanoma and other solid tumors. TMB appears to be a more robust predictor of response compared to PD-L1 expression in melanoma.

LAG-3 is another checkpoint present on CD4+ and CD8+ T cells and is upregulated after chronic antigen exposure. The combination of anti-LAG-3 using relatlimab and anti-PD-1 with nivolumab results in objective responses comparable to ipilimumab plus nivolumab with fewer side effects. Relatlimab plus nivolumab has been approved by the FDA for the first-line treatment of advanced melanoma, has activity after progression on other checkpoint inhibitors, and is being investigated in the neoadjuvant setting as detailed above.

T-cell checkpoint antibodies can also interfere with normal immune regulatory mechanisms, producing a novel spectrum of side effects. The most common immune-related adverse events were skin rash (discussed in depth in the Dermatology Drug Eruption section in **Chap. 63**) and diarrhea (sometimes severe, life-threatening colitis), but toxicity can involve almost any organ resulting in thyroiditis, hypophysitis, hepatitis, nephritis, pneumonitis, myocarditis, and neuritis. The severity and frequency of toxicity are greatest when anti-CTLA-4 and anti-PD-1 are combined, followed by anti-CTLA-4, anti-LAG-3 plus anti-PD-1, and then anti-PD-1 monotherapies. Vigilance, interruption of therapy, and early intervention with steroids or other immunosuppressive agents, such as anti–tumor necrosis factor antibodies or mycophenolate mofetil, can mitigate toxicity and prevent permanent organ damage. Using immunosuppressive agents to mitigate toxicity does not diminish antitumor activity, and benefit is manifest even in patients who must discontinue immunotherapy due to immune-mediated toxicity. Checkpoint immunotherapy can be administered safely to selected patients with preexisting autoimmune conditions using a multidisciplinary approach with input from endocrinology, rheumatology, and other specialty services as clinically necessary. The use of T-cell checkpoint antibodies for metastatic melanoma has become commonplace, but there is controversy about whether all patients need combined anti-CTLA-4 and anti-PD-1 and whether biomarkers can be used to select patients who may benefit from anti-PD-1 alone. There is also a significant economic impact with any anticancer therapy, which must be placed in the context of the survival benefit.

TARGETED THERAPY

The RAS-RAF-MEK-ERK pathway delivers proliferation and survival signals from the cell surface to the cytoplasm and nucleus and is mutated in approximately 50% of melanomas. Inhibitors of BRAF and MEK can induce regression of melanomas that harbor a *BRAF* mutation. Three BRAF inhibitors, vemurafenib, dabrafenib, and encorafenib, have been approved for the treatment of patients whose melanomas harbor a mutation at position 600 in *BRAF*. Monotherapy with BRAF inhibitors has been supplanted with combined BRAF and MEK inhibition to address the rapid adaptation of melanomas that use MAP kinase pathway reactivation to facilitate growth when BRAF is inhibited. Combined therapy with BRAF and MEK inhibitors (dabrafenib and trametinib, vemurafenib with cobimetinib, or encorafenib and binimetinib) improved progression-free and overall survival compared to monotherapy with a BRAF inhibitor. Long-term results of inhibition of the MAP kinase pathway confirm that some patients achieve long intervals of disease control, yet the major limitation of both monotherapy and combined therapy appears to be the acquisition of resistance. The mechanisms of resistance are diverse and reflect the genomic heterogeneity of melanoma; however, most instances involve reactivation of the MAPK pathway, often through *RAS* mutations or mutant *BRAF* amplification. Patients who develop resistance to BRAF and MEK inhibition are candidates for immunotherapy or clinical trials.

Targeted therapy is accompanied by manageable side effects that differ from those experienced during immunotherapy. Headache, pyrexia, and arthralgias are common. A class-specific side effect of BRAF inhibitor monotherapy is the development of hyperproliferative skin lesions, including well-differentiated squamous cell skin cancers (SCCs) in up to 25% of patients. Paradoxical activation of the MAP kinase pathway occurs from BRAF inhibitor–mediated changes in *BRAF* wild-type cells, and the activation is blocked by MEK inhibitor, which explains why these lesions are infrequent during combined therapy. Metastases from treatment-induced SCCs have not been reported, and BRAF and MEK inhibitors can be continued safely following simple excision of the SCCs. Cardiac and ocular toxicities, although infrequent, can occur with BRAF and MEK inhibitors and require medical evaluation, management, and usually discontinuation of targeted therapy.

Activating mutations in the c-kit receptor tyrosine kinase are found in a minority of cutaneous melanomas with chronic sun damage but are more common in mucosal and acral lentiginous subtypes. When activating mutations of *c-kit* are present, imatinib therapy can achieve clinically meaningful responses, similar to gastrointestinal stromal tumors. The probability of objective response in patients whose melanomas harbor a *c-kit* mutation is 29%, although most responses are transient. *N-RAS* mutations occur in 15–20% of melanomas. At present, there are no effective targeted agents for these patients, but *N-RAS* inhibitors are being investigated in clinical trials.

Targeting proteins that are differentially expressed on melanoma has been the basis for many clinical trials investigating vaccines and engineered biologics. An engineered bispecific fusion protein targeting gp100 on melanoma and CD3 on T cells, called tebentafusp, has garnered FDA approval for the treatment of metastatic uveal melanoma in patients who have the HLA-A*02.01 tissue type. Uveal melanoma is an aggressive melanoma subtype with a propensity for metastasis to the liver and a much lower probability of response to checkpoint immunotherapy than cutaneous melanomas. Objective response to tebentafusp is <10%, but overall survival and progression-free survival are significantly improved compared to checkpoint immunotherapy.

Other systemic therapies used to treat stage IV melanoma patients include high-dose interleukin 2, which is also associated with durable remissions in some patients. Chemotherapy using dacarbazine or taxanes is infrequently used and confers no survival benefit.

INITIAL APPROACH TO PATIENT WITH METASTATIC DISEASE

Upon diagnosis of stage IV disease, a sample of the patient's tumor should be submitted for molecular testing to determine whether a *BRAF* or *c-kit* mutation is present. Analysis of a metastatic lesion biopsy is preferred, but any sample will suffice because there is little discordance between primary and metastatic lesions. Treatment algorithms start with determining the melanoma's *BRAF* status.

For *BRAF* wild-type tumors, immunotherapy is recommended. The best sequence of targeted therapy and immunotherapy in patients with *BRAF*-mutated melanomas has been controversial. A randomized study that compared anti-CTLA-4 plus anti-PD-1 followed by BRAF/MEK targeted therapy at progression to the opposite sequence in patients with advanced melanoma showed that immunotherapy followed by targeted therapy conferred statistically significant better overall survival at 2 years and a trend toward better progression-free survival. Toxicities were similar comparing the treatment sequences. The patient's history, including sites of disease, symptom burden, and history of autoimmune conditions, influences the final recommendation for immunotherapy or targeted therapy, but survival data favor the initial use of combined checkpoint immunotherapy in the advanced disease setting. Despite improvements in therapy, most patients with metastatic melanoma will not be cured, so enrollment in a clinical trial is always an important consideration. Many will be poor candidates for therapy because of extensive disease burden, poor performance status, or concomitant illness; thus, the timely integration of palliative care and hospice remains an important element of care.

FOLLOW-UP AND SURVIVORSHIP

Skin examination and surveillance at least once a year are recommended for all patients with melanoma. Routine blood work and imaging for patients with stages IA–IIA (NCCN low risk) disease is not recommended unless symptoms are present. Surveillance diagnostic imaging can be considered in patients with stages IIB–III (NCCN high risk) disease but is mainly reserved for patients with signs or symptoms of recurrent disease or to follow response to therapy. The NCCN does not recommend surveillance imaging in asymptomatic patients who had advanced melanoma and are free of disease 5 or more years out from treatment. For stage-specific recommendations, please consult the NCCN guidelines (see "Further Reading").

The increasing incidence of melanoma has been met with more interest in advocacy and survivorship. Several national and international advocacy groups have called attention to issues such as genetic screening, sun awareness, and the care of chronic treatment-related side effects. These include, but are not limited to, skin changes (such as vitiligo), lymphedema, neuropathies, and gastrointestinal and endocrine disorders. Lymphedema can now be managed in specialty clinics that offer support, nonsurgical treatments, and newer surgical therapies such as lymphovenous bypass and vascularized lymph node transplants.

NONMELANOMA SKIN CANCERS

NMSCs (mostly SCCs and basal cell cancer [BCC]) are the most common cancers in the United States. Although tumor registries do not routinely gather data on the incidence of NMSCs, it is estimated that the annual incidence is more than 5.3 million cases in the United States; SCCs and BCCs account for 80 and 18%, respectively. While less common, the incidence of Merkel cell carcinoma (MCC) has tripled over the past 20 years. There are now an estimated 2600 cases per year with an annual increase in incidence of 8%. NMSCs can metastasize, but MCCs do this most commonly, with sentinel lymph node positivity rates of 25% (compared to 12–19% for melanoma) and mortality rates approaching 33% at 3 years. SCCs, particularly those with high-risk features, can also metastasize and account for 2400 deaths annually. Recent advances in systemic therapy using checkpoint antibodies have improved survival in patients with advanced NMSCs.

PATHOPHYSIOLOGY AND ETIOLOGY

Like melanoma, the most significant cause of NMSCs is UVR, with a dose-response relationship between tanning bed use and the incidence of NMSC. As few as four tanning bed visits per year confers a 15% increase in BCC and an 11% increase in SCC. The risk of lip or oral SCC is increased with cigarette smoking and, like SCC of the ear, has a worse prognosis than SCC found on other body sites. Human papillomaviruses and UVR may act as co-carcinogens. Inherited disorders of DNA repair, such as xeroderma pigmentosum, are associated with a greatly increased incidence of skin cancer and help to establish the link between UV-induced DNA damage, inadequate DNA repair, and skin cancer.

The genes associated with UV damage in SCC include *p53* and *N-RAS*, whereas BCC is primarily associated with damage to hedgehog signaling pathway (Hh) genes, which lead to basal cell proliferation. This is usually the result of loss of function of the tumor-suppressor patched homolog 1 (*PTCH1*), which normally inhibits the signaling of smoothened homolog (*SMO*).

Immunosuppression has also been associated with the development of NMSCs; chronically immunosuppressed solid organ transplant recipients have a 65-fold increase in SCC and a 10-fold increase in BCC. The frequency of skin cancer is proportional to the level and duration of immunosuppression and the extent of sun exposure before and after transplantation. SCCs in this population are particularly aggressive, demonstrating higher rates of local recurrence, metastasis, and mortality. Tumor necrosis factor (TNF) antagonist therapy of inflammatory bowel disease and autoimmune disorders, such as rheumatoid and psoriatic arthritis, may also confer an increased risk of NMSC.

Other risk factors for NMSCs include HIV infection, ionizing radiation, thermal burn scars, *BRAF* inhibitor monotherapy, and chronic ulcerations. Albinism, xeroderma pigmentosum, Muir-Torre syndrome, Rombo's syndrome, Bazex-Dupré-Christol syndrome, dyskeratosis congenita, and basal cell nevus syndrome (Gorlin syndrome) also increase the incidence of NMSC.

Although MCC is also clearly related to UV exposure, age, and immunosuppression, this neural crest–derived cancer also appears to have a viral etiology; an oncogenic Merkel cell polyomavirus (MCPyV) is present in 80% of tumors. In patients with MCPyV-positive tumors, there is inactivation of tumor-suppressor genes, specifically the *p53* transcription factor and retinoblastoma protein (*Rb*). In addition, the viral large T antigen is expressed on tumor cells, and many patients have detectable cellular or humoral immune responses to polyoma viral proteins, although this immune response is insufficient to eradicate the malignancy.

CLINICAL PRESENTATION

Basal Cell Carcinoma BCC arises from epidermal basal cells or the follicular bulge. The least invasive of BCC subtypes, superficial BCC, consists of often subtle, erythematous scaling plaques that slowly enlarge and are most commonly seen on the trunk and proximal extremities **(Fig. 81-3)**. This subtype may be confused with benign inflammatory dermatoses, especially nummular eczema and psoriasis or premalignant actinic keratoses. BCC also can present as a small, slowly growing, pearly nodule, often with tortuous telangiectatic vessels on its surface, rolled borders, and a central crust (nodular BCC). The occasional presence of melanin in this variant of nodular BCC (pigmented BCC) may lead to confusion with melanoma. Morpheaform (fibrosing), infiltrative, and micronodular BCC, the most invasive and potentially aggressive subtypes, manifest as solitary, flat or slightly depressed, indurated whitish, yellowish, or pink scar-like plaques. Borders are typically indistinct, and lesions can be subtle; thus, delay in treatment is common, and tumors can be more extensive than expected clinically.

Squamous Cell Carcinoma Primary *cutaneous SCC* is a malignant neoplasm of keratinizing epidermal cells that has a variable clinical course, ranging from indolent to rapid growth, with the potential to metastasize to regional and distant sites. Commonly, SCC appears as an ulcerated erythematous nodule or superficial erosion on sun-exposed skin of the head, neck, trunk, and extremities **(Fig. 81-4)**. It may also appear as a banal, firm, dome-shaped papule or rough textured plaque. It is commonly mistaken for a wart or callous when the inflammatory response to the lesion is minimal. Dotted or coiled vessels are a hallmark of SCC when viewed through a dermatoscope. The margins of this tumor may be ill defined, and fixation to underlying structures may occur ("tethering").

A very rapidly growing low-grade form of SCC, called keratoacanthoma (KA), typically appears as a large dome-shaped papule with a central keratotic crater. Some KAs regress spontaneously without therapy, but because progression to metastatic SCC has been documented,

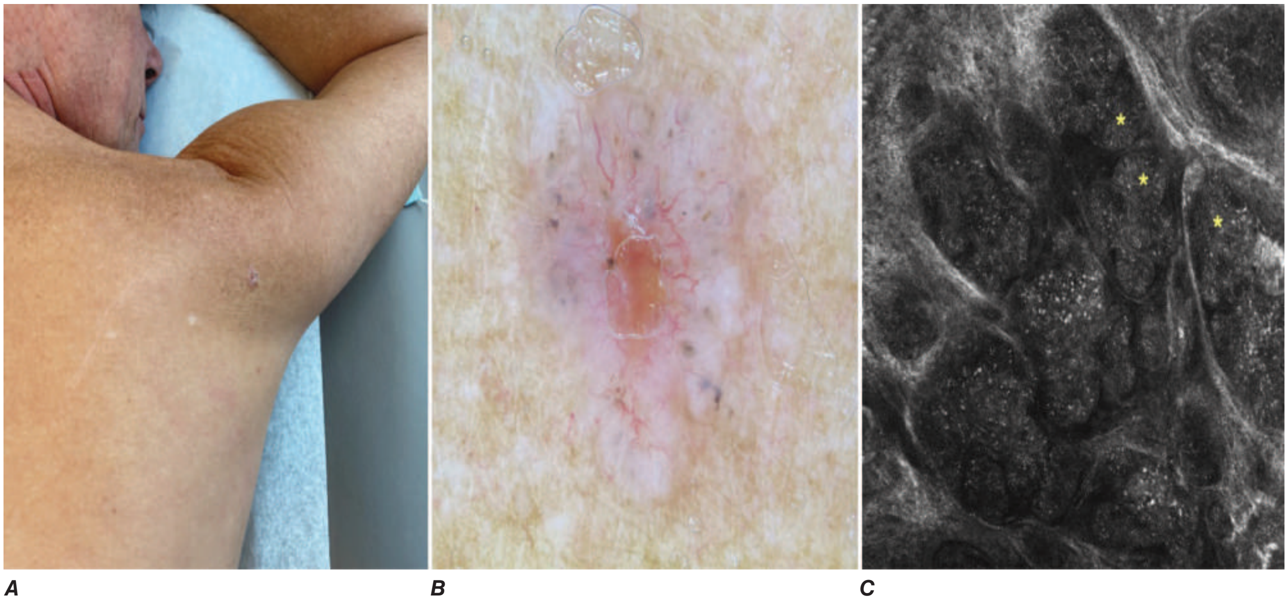

FIGURE 81-3 Clinical, dermascopic, and confocal diagnostic findings of basal cell carcinoma. ***A.*** Typical basal cell carcinoma with skin-colored, slightly translucent rolled borders and a small central erosion on chronically sun-damaged skin of the lateral posterior shoulder. ***B.*** Dermoscopic image of the same lesion as in panel *A* clearly revealing the central erosion and classic gray, nonreticular globular structures of melanophages that characterize BCC. ***C.*** In vivo reflectance confocal microscopy of the same lesion as in panel *A* showing typical nests of dermal basaloid cells (*) with classic cleft formation around the nests. *(Photos courtesy of Dr. Alexander Witkowski and Dr. Joanna Ludzik, © Copyright 2022 Oregon Health & Science University [OHSU].)*

KAs should be treated in the same manner as other types of cutaneous SCC. KAs occur in 15–25% of patients receiving monotherapy with a BRAF inhibitor.

Actinic keratoses and *cheilitis* (actinic keratoses on the lip), both premalignant forms of SCC, present as hyperkeratotic papules on sun-exposed areas. Malignant transformation occurs in 0.25–20% of untreated lesions. SCC in situ, also called *Bowen's disease*, is the intraepidermal form of SCC and usually presents as a scaling, erythematous plaque. SCC in situ most commonly arises on sun-damaged skin but can occur anywhere on the body. Bowen's disease occurring secondary to infection with human papillomavirus can arise on skin with minimal or no prior sun exposure, such as the buttock or posterior thigh. Treatment of premalignant and in situ lesions reduces the subsequent risk of invasive disease.

Merkel Cell Carcinoma MCC, also known as cutaneous apudoma, primary neuroendocrine carcinoma of the skin, primary small-cell carcinoma of the skin, and trabecular carcinoma of the skin, arises from Merkel cells, which are neuroendocrine skin cells that act as pressure receptors. Like other skin cancers, MCCs most commonly arise as visible skin lesions, usually as raised, flesh-colored nodules or masses; they can also be red or blue in color and vary in size from 0.5 to >5 cm in diameter and may enlarge rapidly. Although MCCs may arise almost anywhere on the body, they are most often found in sun-exposed areas such as the head, neck, or extremities. They can also be found around the anus and on eyelids. The common clinical features of MCC can be summarized by the acronym AEIOU: *a*symptomatic/nontender, *e*xpand rapidly, *i*mmune suppression, *o*lder than 50 years, and *u*ltraviolet-exposed site.

NATURAL HISTORY

Basal Cell Carcinoma The natural history of BCC is that of a slowly enlarging, locally invasive neoplasm. The degree of local destruction and risk of recurrence vary with the size, duration, location, and histologic subtype of the tumor. Location on the central face, ears, or scalp may portend a higher risk. Small nodular, pigmented, cystic, or superficial BCCs respond well to most treatments. Large lesions and micronodular, infiltrative, and morpheaform subtypes may be more aggressive. The metastatic potential of BCC is low (0.1%) in immunocompetent patients, but the risk of recurrence or a new primary NMSC is about 40% over 5 years.

Squamous Cell Carcinoma The natural history of SCC depends on tumor and host characteristics. Tumors arising on sun-damaged skin have a lower metastatic potential than do those on non-sun-exposed areas. Cutaneous SCC metastasizes in 0.3–5.2% of individuals, most frequently to regional lymph nodes. Tumors occurring on the lower lip and ear develop regional metastases in 13 and 11% of patients, respectively, whereas the metastatic potential of SCC arising in scars, chronic ulcerations, and genital or mucosal surfaces is higher. Recurrent SCC has a 30% probability for metastatic spread. Large, poorly differentiated, deep tumors with perineural or lymphatic invasion, multifocal tumors, and those arising in immunosuppressed patients often behave aggressively.

Merkel Cell Carcinoma MCCs usually present locally yet have a high probability of spread to regional lymph nodes and distant sites. Molecular markers of neuroendocrine origin such as synaptophysin or chromogranin A are useful to diagnose MCC. Unlike other neuroendocrine tumors, such as small-cell lung cancer (SCLC), MCCs are not associated with measurable hormone secretion or endocrine syndromes.

Survival with MCC depends on extent of disease: 90% of patients with local disease are cured, whereas 52% with nodal involvement and 10% with distant disease survive. MCC has its own tumor-node-metastasis (TNM) staging system, which incorporates tumor size (<2 cm vs >2 cm), nodal status (which can be determined by SLNB for clinically negative nodes), and the presence of distant metastases.

Independent of stage, the prognosis of MCC is improved if the tumor cells contain virus, express RB protein, and exhibit intratumoral CD8+ T lymphocyte infiltration. p63 expression, lymphovascular infiltrative pattern, and concomitant immunosuppression (e.g., organ transplant, HIV infection, and certain cancers) portend a worse prognosis.

TREATMENT

Basal Cell, Squamous Cell, and Merkel Cell Carcinoma

BASAL CELL CARCINOMA

Treatment for BCC includes electrodesiccation and curettage (ED&C), excision, cryosurgery, radiation therapy (RT), laser therapy, MMS, topical 5-fluorouracil, photodynamic therapy (PDT), and topical immunomodulators, such as imiquimod. The choice

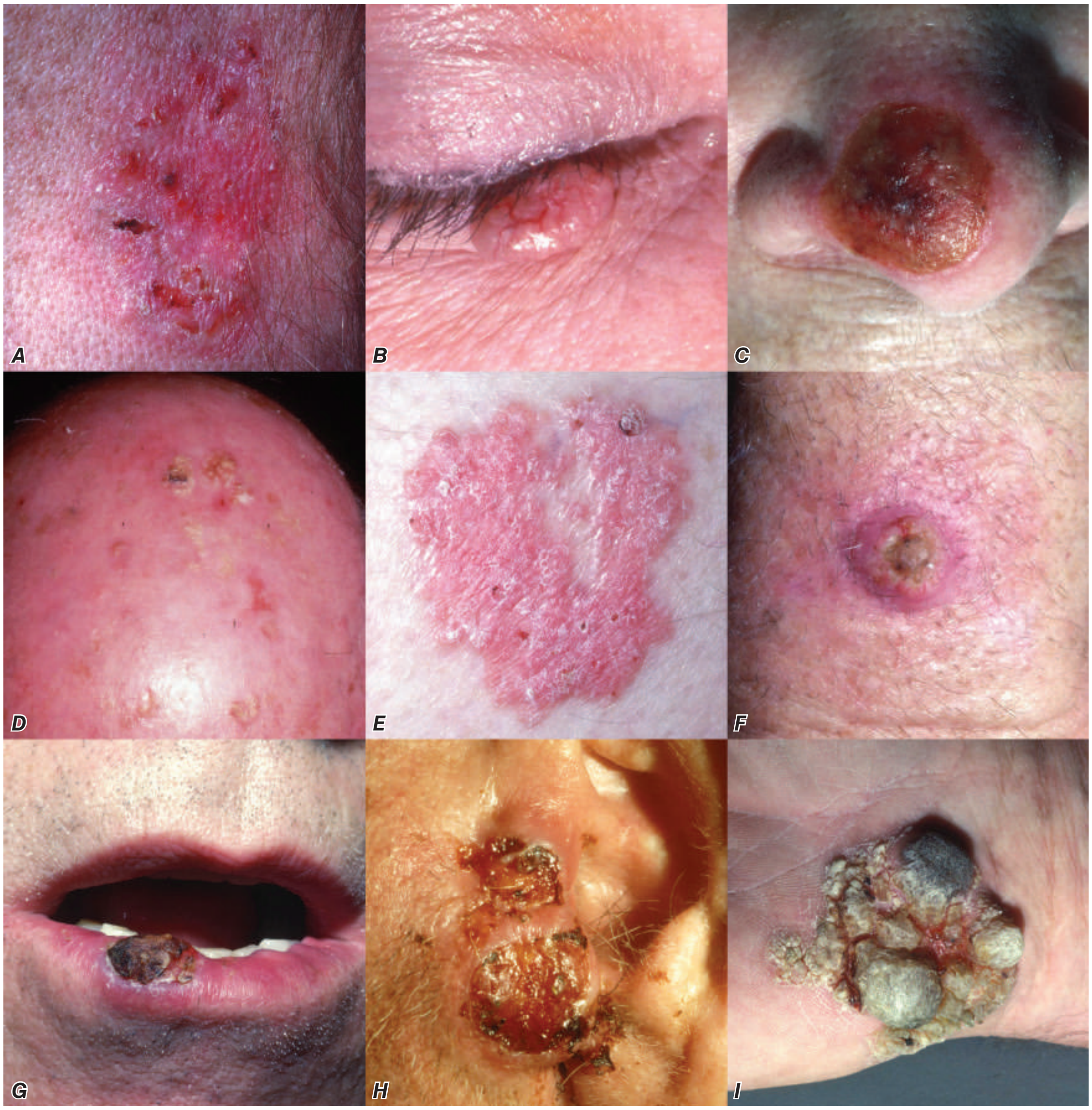

FIGURE 81-4 Progression of basal cell (BCC) and squamous cell carcinoma (SCC). ***A.*** Superficial BCC; note salmon pink color, rolled boarder, erosions, and a gray central globule. ***B.*** Nodular BCC; note shiny, slightly pearly character with prominent arborizing vessels. ***C.*** Eroded BCC with serous and sanguinous crusting. ***D.*** Multiple actinic keratoses; note flat lesions are early, and thicker lesions may require biopsy to discriminate between advanced actinic keratosis versus early SCC. ***E.*** Superficial SCC. ***F.*** Keratoacanthoma (well-differentiated SCC). ***G.*** Mucocutaneous SCC in a high-risk area on the lower lip. ***H.*** Cutaneous SCC. ***I.*** Large exophytic SCC on the wrist. *(Photos courtesy of the Dr. Leonard Swinyer Collection, © Copyright 2020 University of Utah and Oregon Health & Science University.)*

of therapy depends on tumor characteristics including depth and location, patient age, medical status, and patient preference. ED&C remains the most frequent treatment for superficial, minimally invasive nodular BCCs and low-risk tumors (e.g., a small tumor of a less aggressive subtype in a favorable location without terminal hairs). Wide local excision with standard margins is usually selected for invasive, ill-defined, and more aggressive subtypes of tumors or for cosmetic reasons. MMS, a specialized type of surgical excision that provides the best method for tumor removal while preserving uninvolved tissue, is associated with cure rates of >98%. It is the preferred modality for lesions that are recurrent, in high-risk or cosmetically sensitive locations (including recurrent tumors in these locations), and for which maximal tissue conservation is critical (e.g., the eyelids, lips, ears, nose, and digits). RT can cure patients not considered surgical candidates and can be used as a surgical adjunct in high-risk tumors. Imiquimod can be used to treat superficial and smaller nodular BCCs, although it is not FDA approved for nodular BCC. Topical 5-fluorouracil therapy should be limited to superficial BCC. PDT, which uses selective activation of a photoactive drug by visible light, has been used in patients with numerous tumors. Intralesional therapy (5-fluorouracil or interferon) can also be employed. Like RT, it remains an option for selected patients who cannot or will not undergo surgery. Systemic therapy with a targeted hedgehog pathway inhibitor, such as vismodegib or sonidegib, is indicated for patients with metastatic or advanced BCC that has recurred after local therapy and who are not candidates for surgery or RT. Targeted therapy does not cure patients with BCC but induces regression in approximately 50% of patients with a median duration of response of 9–12 months in patients with metastatic disease and ~2 years in patients with locally advanced disease. Checkpoint immunotherapy using cemiplimab can be offered to BCC patients who progress after targeted therapy.

SQUAMOUS CELL CARCINOMA

The principles for surgical management of SCC are the same as for BCC. Cemiplimab, a monoclonal antibody targeting PD-1, has become the systemic therapy of choice, inducing tumor regression in 47% of patients with advanced disease. Neoadjuvant cemiplimab has been given in stage II, III and IV SCC and is associated with a >50% probability of pathologic complete response and is becoming a standard of care in patients with very-high-risk presentations and/or who have disease that may be technically difficult to resect. SCC and KAs that develop in patients receiving BRAF-targeted therapy should be excised, after which BRAF therapy can be continued.

MERKEL CELL CARCINOMA

The epidemiology, clinical features, and treatments for MCC overlap those for melanoma and NMSC. Early-stage MCCs may be cured with wide local excision of the primary tumor and nodal staging with SLNB. Like SCLCs, MCC is sensitive to radiation, PD-1-directed immunotherapy, and platinum-based chemotherapy. RT is often used as postoperative adjuvant therapy at both the primary excision and SLNB sites, although its use may be withheld around sensitive areas such as the eyelids and hands and after a negative SLNB. For nonsensitive areas, RT may allow for primary excision margins smaller than the traditionally recommended 2-cm radial margins. When a positive sentinel node is found, adjuvant RT, close observation, and clinical trials investigating immunotherapy are favored over completion nodal dissection.

For patients with metastatic disease, immunotherapy has supplanted chemotherapy. Avelumab (anti-PD-L1) therapy led to objective responses in 33% of patients with advanced MCC; 82% of the responses were durable. Pembrolizumab has an objective response >50% in patients with MCPyV-associated and non-virus-associated advanced MCC resulting in a median duration of response approaching 2 years. Clinical trials should be offered to MCC patients who progress after checkpoint therapy and whose functional status can support additional treatment.

Follow-up of patients with MCC is based on stage and risk. Routine skin exams by a dermatologist familiar with MCC and regular examinations of the nodal basins are recommended. Antibody serum titers to MCPyV should be obtained in newly diagnosed MCC patients. The test can be used to follow patients for relapse if the titer is elevated at baseline and returns to normal after treatment. Conversely, if the titer is elevated but does not return to normal after treatment, imaging should be obtained to look for occult metastases.

PREVENTION

The principles for prevention are those described for melanoma earlier. Unique strategies for NMSC include active surveillance for patients on immunosuppressive medications or BRAF-targeted therapy. Chemoprophylaxis using synthetic retinoids and immunosuppression reduction when possible may be useful in controlling new lesions and managing patients with multiple tumors. Nicotinamide 500 mg BID may be used in patients with large numbers of actinic keratoses and SCCs to reduce the development and/or progression of disease. Field therapy with topical 5-fluorouracil (with or without calcipotriol), ingenol mebutate, or imiquimod can reduce transformation to SCC in patients with severely sun-damaged skin and numerous premalignant actinic keratoses. Older, immunosuppressed patients should be managed with the lowest doses of immunosuppression possible and encouraged to be particularly careful to minimize UV exposure. Earlier biopsy of unusual-appearing skin lesions may lead to better control of aggressive lesions.

OTHER NONMELANOMA CUTANEOUS MALIGNANCIES

Neoplasms of cutaneous adnexae and sarcomas of fibrous, mesenchymal, fatty, and vascular tissues make up the remaining 1–2% of NMSCs **(Fig. 81-5)**. Lymphomas of B- or T-cell origin can also manifest in the skin and can mimic benign conditions such as psoriasis and eczema.

Extramammary Paget's disease is an uncommon apocrine malignancy arising from stem cells of the epidermis that is characterized

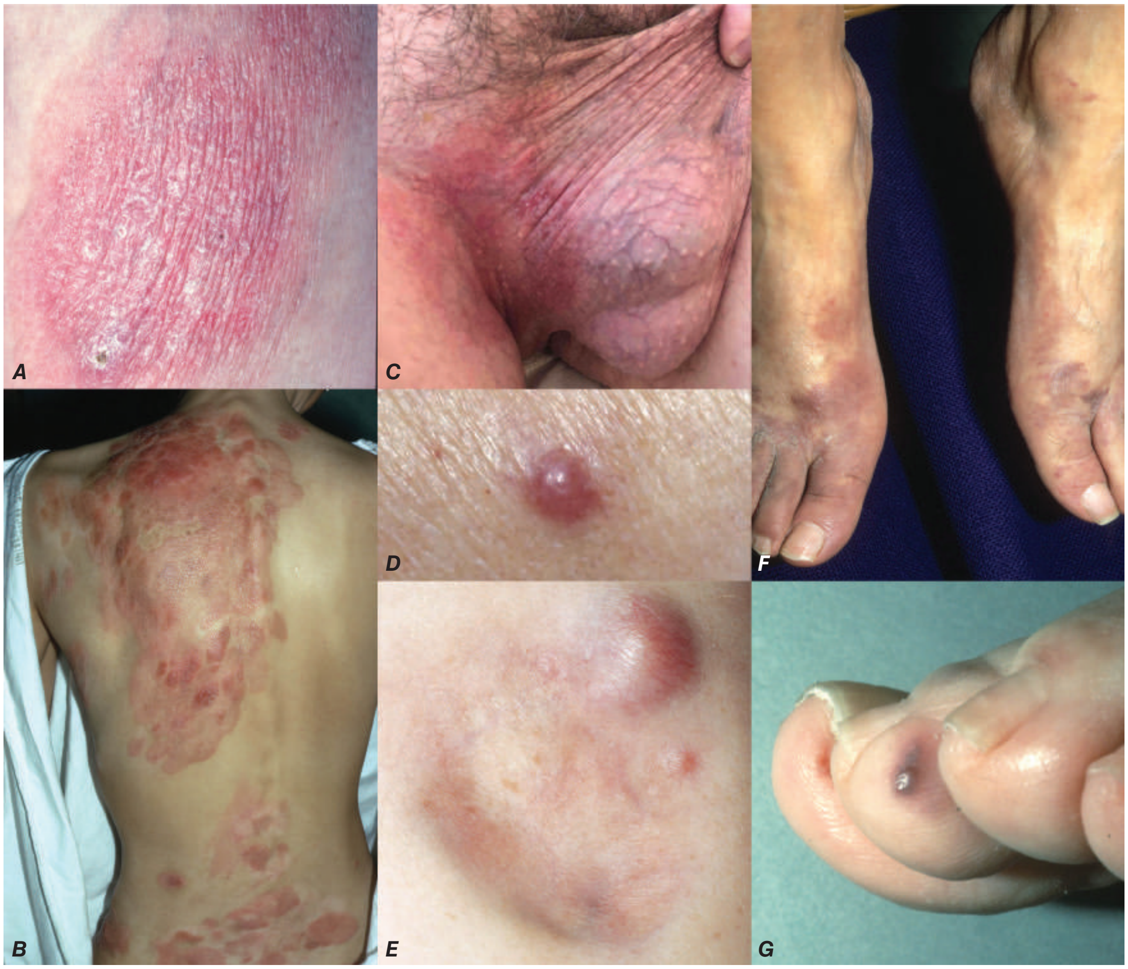

FIGURE 81-5 Other malignant cutaneous tumors. ***A.*** Patch stage mycosis fungoides (variant of cutaneous T-cell lymphoma). ***B.*** Tumor stage mycosis fungoides. ***C.*** Extramammary Paget's disease. ***D.*** Merkel cell carcinoma. ***E.*** Dermatofibrosarcoma protuberans. ***F. and G.*** Kaposi's sarcoma. *(Parts A, B, and D-G photos courtesy of the Dr. Leonard Swinyer Collection, © Copyright 2020 University of Utah and Oregon Health & Science University. Part C photo courtesy Dr. Justin Leitenberger, © Copyright 2022 Oregon Health & Science University [OHSU].)*

histologically by the presence of Paget cells. These tumors present as moist erythematous patches on anogenital or axillary skin of the elderly.

Outcomes are generally good with surgery, and 5-year disease-specific survival is 95% with localized disease. Advanced age and extensive disease at presentation confer poorer prognosis. RT or topical imiquimod can be considered for more extensive disease. Local management may be challenging because these tumors often extend far beyond clinical margins; surgical excision with MMS has the highest cure rates. Similarly, MMS is the treatment of choice in other rare cutaneous tumors with extensive subclinical extension such as *dermatofibrosarcoma protuberans*.

Kaposi's sarcoma (KS) is a soft tissue sarcoma of vascular origin that is induced by the human herpesvirus 8. The incidence of KS increased dramatically during the AIDS epidemic but has now decreased tenfold with the institution of highly active antiretroviral therapy.

Acknowledgment

Walter Urba, MD, PhD, provided valued feedback and suggested improvements to this chapter. Clinical photos were generously provided from the OHSU Swinyer Collection (Leonard Swinyer, MD) and by Drs. Elizabeth Berry, Alexander Witkowski, Joanna Ludzik, Debbie Miller, Alison Skalet, and Justin Leitenberger. Dermoscopic images were provided by Elizabeth Berry, Alexander Witkowski, Joanna Ludzik, and Debbie Miller. Reflectance confocal microscopy images were provided by Drs. Alexander Witkowski and Joanna Ludzik.

FURTHER READING

Atkins ME et al: Combination dabrafenib and trametinib versus combination nivolumab and ipilimumab for patients with advanced BRAF-mutant melanoma: The DREAMseq trial—ECOG-ACRIN EA6134. J Clin Oncol 41:186, 2022.

Elder DE et al: The 2018 World Health Organization classification of cutaneous, mucosal, and uveal melanoma. Arch Pathol Lab Med 44:500, 2020.

Faries MD et al: Completion dissection or observation for sentinel-node metastasis in melanoma. N Engl J Med 376:2211, 2017.

Harms PW et al: The biology and treatment of Merkel cell carcinoma: Current understanding and research priorities. Nat Rev Clin Oncol 15:763, 2018.

National Comprehensive Cancer Network: NCCN clinical practice guidelines in oncology (NCCN guidelines): Melanoma. Available from *https://www.nccn.org/professionals/physician_gls/pdf/melanoma.pdf*.

Tawbi HA et al: Relatlimab and nivolumab versus nivolumab in untreated advanced melanoma. N Engl J Med 386:24, 2022.

Wu YP et al: A systematic review of interventions to improve adherence to melanoma preventive behaviors for individuals at elevated risk. Prev Med 88:153, 2016.

82

Head and Neck Cancer

Everett E. Vokes

Epithelial carcinomas of the head and neck arise from the mucosal surfaces in the head and neck and typically are squamous cell in origin. This category includes tumors of the paranasal sinuses, the oral cavity, and the nasopharynx, oropharynx, hypopharynx, and larynx. Tumors of the salivary glands differ from the more common carcinomas of the head and neck in etiology, histopathology, clinical presentation, and therapy. They are rare and histologically highly heterogeneous. Thyroid malignancies are described in **Chap. 397**.

INCIDENCE AND EPIDEMIOLOGY

The number of new cases of head and neck cancers (oral cavity, pharynx, and larynx) in the United States was estimated at 66,920 in 2023, accounting for about 6% of adult malignancies; estimated deaths were 15,400. The worldwide incidence exceeds half a million cases annually. In North America and Europe, the tumors usually arise from the oral cavity, oropharynx, or larynx. The incidence of oropharyngeal cancers has been increasing in Western countries. Nasopharyngeal cancer is endemic in East Asia and some Mediterranean countries.

ETIOLOGY AND GENETICS

Alcohol and tobacco use are the most significant environmental risk factors for head and neck cancer, and when used together, they act synergistically. Smokeless tobacco is an etiologic agent for oral cancers. Other potential carcinogens include marijuana and occupational exposures such as nickel refining, exposure to textile fibers, and woodworking.

Some head and neck cancers have a viral etiology. Epstein-Barr virus (EBV) infection is frequently associated with nasopharyngeal cancer, especially in endemic areas. EBV antibody titers can be measured to screen high-risk populations and are under investigation to monitor treatment response. Nasopharyngeal cancer has also been associated with consumption of salted fish and indoor pollution.

In Western countries, the human papillomavirus (HPV) is associated with a rising incidence of tumors arising from the oropharynx, that is, the tonsillar bed and base of tongue. Over 50% of oropharyngeal tumors are caused by HPV in the United States, and in many urban centers, this proportion is higher. HPV-16 is the dominant viral subtype, although HPV-18 and other oncogenic subtypes are seen as well. Alcohol- and tobacco-related cancers, on the other hand, have decreased in incidence. HPV-related oropharyngeal cancer occurs in a younger patient population and is associated with increased numbers of sexual partners and oral sexual practices. It is associated with a better prognosis, especially for nonsmokers. Vaccination with the nine-valent HPV vaccine may prevent the disease but is not likely to result in a lower incidence for several decades due to long latency of the carcinogenic process.

Dietary factors may contribute. The incidence of head and neck cancer is higher in people with the lowest consumption of fruits and vegetables. Certain vitamins, including carotenoids, may be protective if included in a balanced diet. Supplements of retinoids, such as *cis*-retinoic acid, have not been shown to prevent head and neck cancers (or lung cancer) and may increase the risk in active smokers. No specific risk factors or environmental carcinogens have been identified for salivary gland tumors.

HISTOPATHOLOGY, CARCINOGENESIS, AND MOLECULAR BIOLOGY

Squamous cell head and neck cancers are divided into well-differentiated, moderately well-differentiated, and poorly differentiated categories. Poorly differentiated tumors have a worse prognosis than well-differentiated tumors. For nasopharyngeal cancers, the less common differentiated squamous cell carcinoma is distinguished from nonkeratinizing and undifferentiated carcinoma (lymphoepithelioma) that contains infiltrating lymphocytes and is commonly associated with EBV.

Salivary gland tumors can arise from the major (parotid, submandibular, sublingual) or minor salivary glands (located in the submucosa of the upper aerodigestive tract). Most parotid tumors are benign, but half of submandibular and sublingual gland tumors and most minor salivary gland tumors are malignant. Malignant tumors include mucoepidermoid and adenoid cystic carcinomas and adenocarcinomas.

The mucosal surface of the entire pharynx is exposed to alcohol- and tobacco-related carcinogens and is at risk for the development of a premalignant or malignant lesion. Erythroplakia (a red patch) or leukoplakia (a white patch) can be histopathologically classified as hyperplasia, dysplasia, carcinoma in situ, or carcinoma. However, most head and neck cancer patients do not present with a known history of premalignant lesions. Multiple synchronous or metachronous cancers can also be observed. In fact, over time, patients with treated early-stage tobacco- and alcohol-related head and neck cancer are at greater risk of dying from a second malignancy than from a recurrence of the primary disease.

Second head and neck malignancies are usually not therapy induced; they reflect the exposure of the upper aerodigestive mucosa to the same carcinogens that caused the first cancer. These second primaries develop in the head and neck area, the lung, or the esophagus. Thus, computed tomography (CT) screening for lung cancer in heavy smokers who have already developed a head and neck cancer is recommended. Rarely, patients can develop a radiation therapy–induced sarcoma after having undergone prior radiotherapy for a head and neck cancer.

Much progress has been made in describing the molecular features of head and neck cancer. These features have allowed investigators to describe the genetic and epigenetic alterations and the mutational spectrum of these tumors. Early reports demonstrated frequent overexpression of the epidermal growth factor receptor (EGFR). Overexpression was shown to correlate with poor prognosis. However, it has not proved to be a good predictor of tumor response to EGFR inhibitors, which are active in only about 10–15% of patients as single agents. Complex genetic analyses, including those by The Cancer Genome Atlas project, have been performed. *p53* mutations are found frequently with other major affected oncogenic driver pathways including the mitotic signaling and Notch pathways and cell cycle regulation in HPV-negative tumors. HPV oncogenes act through direct inhibition of the *p53* and *RB* tumor-suppressor genes, thereby initiating the carcinogenic process. *HRAS* mutations are detected in 4–8% of patients with recurrent head and neck cancer and may be therapeutically targetable in a small patient subset. While overall mutation rates are similar in HPV-positive and carcinogen-induced tumors, the specific mutational signature of HPV-positive tumors differs, with frequent alteration of the PI3K pathway and occasional mutations in *KRAS*. Overall, these alterations affect mitogenic signaling, genetic stability, cellular proliferation, and differentiation.

■ CLINICAL PRESENTATION AND DIFFERENTIAL DIAGNOSIS

Most tobacco-related head and neck cancers occur in patients older than age 60 years. HPV-related malignancies are frequently diagnosed in younger patients, usually in their forties or fifties, whereas EBV-related nasopharyngeal cancer can occur at all ages, including in teenagers. The manifestations vary according to the stage and primary site of the tumor. Patients with nonspecific signs and symptoms in the head and neck area should be evaluated with a thorough otolaryngologic examination, particularly if symptoms persist longer than 2–4 weeks. Males are more frequently affected than women by head and neck cancers, including HPV-positive tumors.

Cancer of the nasopharynx typically does not cause early symptoms. However, it may cause unilateral serous otitis media due to obstruction of the eustachian tube, unilateral or bilateral nasal obstruction, or epistaxis. Advanced nasopharyngeal carcinoma causes neuropathies of the cranial nerves due to skull base involvement.

Carcinomas of the oral cavity present as nonhealing ulcers, changes in the fit of dentures, or painful lesions and masses. Tumors of the tongue base or oropharynx can cause decreased tongue mobility and alterations in speech. Cancers of the oropharynx or hypopharynx rarely cause early symptoms, but they may cause sore throat and/or otalgia. HPV-related tumors frequently present with neck lymphadenopathy as the first sign.

Hoarseness may be an early symptom of laryngeal cancer, and persistent hoarseness requires referral to a specialist for indirect laryngoscopy and/or radiographic studies. If a head and neck lesion treated initially with antibiotics does not resolve in a short period, further workup is indicated; to simply continue the antibiotic treatment may be to lose the chance of early diagnosis of a malignancy.

Advanced head and neck cancers in any location can cause severe pain, otalgia, airway obstruction, cranial neuropathies, trismus, odynophagia, dysphagia, decreased tongue mobility, fistulas, skin involvement, and massive cervical lymphadenopathy, which may be unilateral or bilateral. Some patients have enlarged lymph nodes even though no primary lesion can be detected by endoscopy or biopsy; these patients are considered to have carcinoma of unknown primary **(Fig. 82-1)**. Tonsillectomy and directed biopsies of the base of tongue can identify a small primary tumor that frequently will be HPV related. If the enlarged nodes are located in the upper neck and the tumor cells are of squamous cell histology, the malignancy probably arose from a mucosal surface in the head or neck. Tumor cells in supraclavicular lymph nodes may also arise from a primary site in the chest or abdomen.

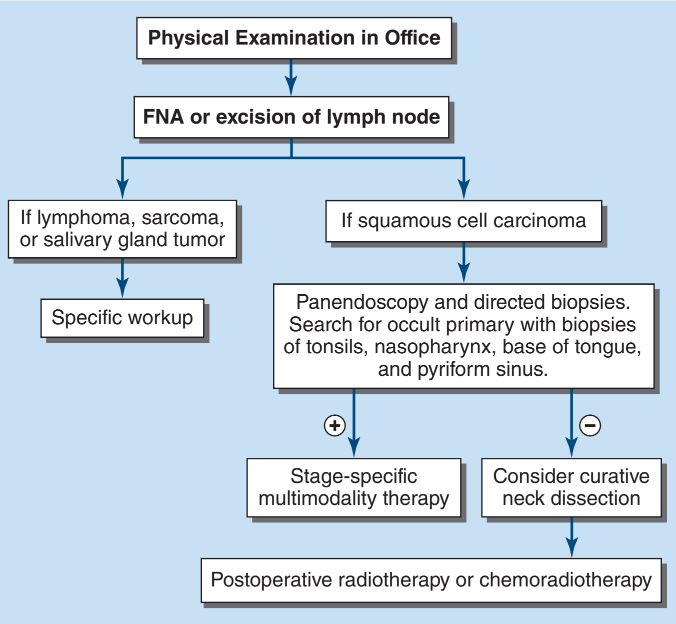
Physical Examination in Office
FNA or excision of lymph node
If lymphoma, sarcoma, or salivary gland tumor
If squamous cell carcinoma
Specific workup
Panendoscopy and directed biopsies. Search for occult primary with biopsies of tonsils, nasopharynx, base of tongue, and pyriform sinus.
+
−
Stage-specific multimodality therapy
Consider curative neck dissection
Postoperative radiotherapy or chemoradiotherapy

FIGURE 82-1 Evaluation of a patient with cervical adenopathy without a primary mucosal lesion; a diagnostic workup. FNA, fine-needle aspiration.

The physical examination should include inspection of all visible mucosal surfaces and palpation of the floor of the mouth and of the tongue and neck. In addition to a tumor, leukoplakia (a white mucosal patch) or erythroplakia (a red mucosal patch) may be observed; these "premalignant" lesions can represent hyperplasia, dysplasia, or carcinoma in situ and require biopsy. Further examination should be performed by a specialist. Additional staging procedures include CT or MRI of the head and neck to identify the extent of the disease. Patients with lymph node involvement should have CT scan of the chest and upper abdomen to screen for distant metastases. In heavy smokers, the CT scan of the chest can also serve as a screening tool to rule out a second lung primary tumor. A positron emission tomography (PET) scan can help to identify or exclude distant metastases. CT and PET scans may also be useful in evaluating response to therapy. The definitive staging procedure is an endoscopic examination under anesthesia, which may include laryngoscopy, esophagoscopy, and bronchoscopy; during this procedure, multiple biopsy samples are obtained to establish a primary diagnosis, define the extent of primary disease, and identify any additional premalignant lesions or second primaries.

Head and neck tumors are classified according to the tumor-node-metastasis (TNM) system of the American Joint Committee on Cancer (AJCC) **(Fig. 82-2)**. This classification varies according to the specific anatomic subsite. In general, primary tumors are classified as T1 to T3 by increasing size, whereas T4 usually represents invasion of another structure such as bone, muscle, or root of tongue. Lymph nodes are staged by size, number, and location (ipsilateral vs contralateral to the primary). Overt distant metastases are found in <10% of patients at initial diagnosis and are more common in patients with advanced lymph node stage; microscopic involvement of the lungs, bones, or liver is more common, particularly in patients with advanced neck lymph node disease. HPV-related oropharyngeal malignancies have consistently been shown to have a better prognosis, and in the eighth

DOI = depth of invasion

	Definition of TNM						Stage groupings
Stage I	**T1**	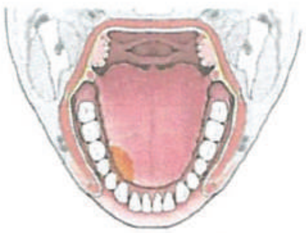	Tumor ≤2 cm in greatest dimension ≤5 mm depth of invasion (DOI)	N0		N0- No regional lymph node metastasis	T1 N0 M0
Stage II	**T2**	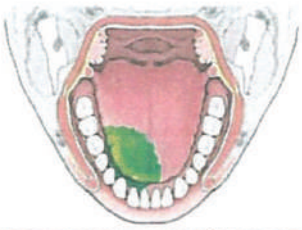	Tumor ≥2 cm but not more than 4 cm in greatest dimension OR DOI >5 mm and ≤10 mm	N0		N0- No regional lymph node metastasis	T2 N0 M0
Stage III	**T3**	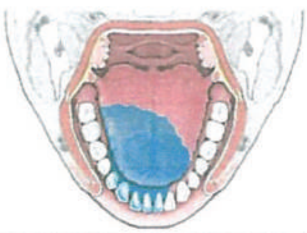	Tumor ≥4 cm OR DOI >10 mm	**N1**	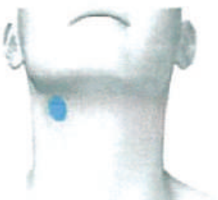 ≤3 cm	N1- Metastasis in a single ipsilateral lymph node, ≤3 cm in greatest dimension	T3 N0 M0 T1 N1 M0 T2 N1 M0 T3 N1 M0
Stage IVA	**T4a**	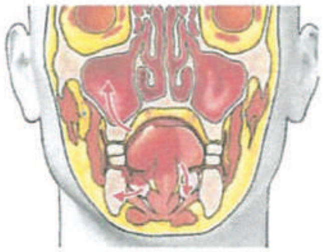	Tumor invades skin, mandible, ear canal, fascial nerve, and/or floor of mouth	**N2**	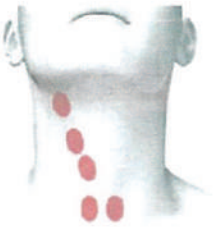 ≤6 cm	N2a- Metastasis in a single ipsilateral lymph node, >3 cm but ≤6 cm N2b- Metastasis in multiple ipsilateral lymph nodes, none >6 cm N2c- Metastasis in bilateral or contralateral lymph nodes, none >6 cm	T4a N0 M0 T4a N1 M0 T1 N2 M0 T2 N2 M0 T3 N2 M0 T4a N2 M0
Stage IVB	**T4b**	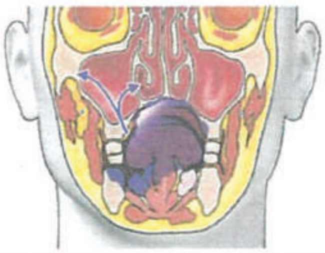	Tumor invades skull base and/or pterygoid plates and/or encases carotid artery	**N3**	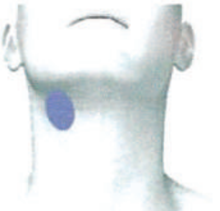 >6 cm	N3- Metastasis in a lymph node >6 cm in greatest dimension or clinically overt extranodal extension	T4b Any N M0 Any T N3 M0
Stage IVC				M1			Any T Any N M1

FIGURE 82-2 Tumor-node-metastasis (TNM) staging system. *(Figure based on the AJCC Cancer Staging Manual, 8th edition.)*

edition of the AJCC staging manual, a separate staging system that takes into account the more favorable outlook of these patients has been included. According to this system, patients with advanced nodal stage can still be considered to have the equivalent of an overall early stage (and associated good prognosis).

In patients with lymph node involvement and no visible primary, the diagnosis should be made by fine-needle aspiration or by lymph node excision (especially if only a single node appears involved) (Fig. 82-1). If the results indicate squamous cell carcinoma, a panendoscopy should be performed, with biopsy of all suspicious-appearing areas and directed biopsies of common primary sites, such as the nasopharynx, tonsil, tongue base, and pyriform sinus. HPV-positive tumors especially can have small primary tumors that spread early to locoregional lymph nodes.

TREATMENT

Head and Neck Cancer

Patients with head and neck cancer can be grossly categorized into three clinical groups: those with localized disease, those with locally or regionally advanced disease (lymph node positive), and those with recurrent and/or metastatic disease below the neck. Comorbidities associated with tobacco and alcohol abuse can affect treatment outcome and define long-term risks for patients who are cured of their disease.

LOCALIZED DISEASE

Nearly one-third of patients have localized disease, that is, T1 or T2 (stage I or stage II) lesions without detectable lymph node involvement or distant metastases. These patients are treated with curative intent by either surgery or radiation therapy. The choice of modality differs according to anatomic location and institutional expertise. Radiation therapy is often preferred for laryngeal cancer to preserve voice function, and surgery is preferred for small lesions in the oral cavity to avoid the long-term complications of radiation, such as xerostomia and osteoradionecrosis and dental decay. Randomized data have shown that a prophylactic staging neck dissection should be part of the surgical procedure to eliminate occult nodal metastatic disease. Overall 5-year survival is 60–90%. Most recurrences occur within the first 2 years following diagnosis and are usually local.

LOCALLY OR REGIONALLY ADVANCED DISEASE

Locally or regionally advanced disease—disease with a large primary tumor and/or cervical lymph node metastases—is the stage of presentation for >50% of patients. Such patients can also be treated with curative intent, but not usually with surgery or radiation therapy alone. Combined-modality therapy, including surgery and/or radiation therapy and chemotherapy, is most successful. Chemotherapy can be administered as induction chemotherapy (chemotherapy before surgery and/or radiotherapy) or as concomitant (simultaneous) chemotherapy and radiation therapy. The latter is most commonly used and supported by the best evidence. Five-year survival rates exceed 50% in many trials, but part of this increased survival may be due to an increasing fraction of study populations with HPV-related tumors who carry a better prognosis. HPV testing of newly diagnosed tumors should be performed for patients with oropharyngeal tumors at the time of diagnosis. Clinical trials for HPV-related tumors are focused on exploring reductions in treatment intensity, especially radiation dose, in order to ameliorate long-term toxicities (fibrosis, swallowing dysfunction).

In patients with intermediate-stage tumors (stage III and early stage IV), concomitant chemoradiotherapy can be administered as a primary treatment for patients with unresectable disease, to pursue an organ-preserving approach especially for patients with laryngeal cancer (omission of surgery), or in the postoperative setting for smaller resectable tumors with adverse prognostic features.

Induction Chemotherapy In this strategy, patients receive chemotherapy (current standard is a three-drug regimen of docetaxel, cisplatin, and fluorouracil [5-FU]) before surgery and radiation therapy. Most patients who receive three cycles show tumor reduction, and the response is clinically "complete" in up to half of patients. This "sequential" multimodality therapy allows for organ preservation in patients with laryngeal and hypopharyngeal cancer and results in higher cure rates compared with radiotherapy alone.

Concomitant Chemoradiotherapy With the concomitant strategy, chemotherapy and radiation therapy are given simultaneously rather than in sequence. Tumor recurrences from head and neck cancer develop most commonly locoregionally (in the head and neck area of the primary and draining lymph nodes). The concomitant approach is aimed at enhancing tumor cell killing by radiation therapy in the presence of chemotherapy (radiation enhancement) and is a conceptually attractive approach for bulky tumors. Toxicity (especially mucositis, grade 3 or 4, in 70–80%) is increased with concomitant chemoradiotherapy. However, meta-analyses of randomized trials document an improvement in 5-year survival of 8% with concomitant chemotherapy and radiation therapy. Cisplatin is preferentially given weekly during a course of daily radiotherapy over a 6- to 7-week course. In addition, concomitant chemoradiotherapy produces better laryngectomy-free survival (organ preservation) than radiation therapy alone in patients with advanced larynx cancer. For patients with advanced nasopharyngeal cancer, the addition of neoadjuvant chemotherapy before concomitant chemoradiotherapy has been adopted as standard of care leading to 5-year survival rates exceeding 80% in a Southeast Asian study. The outcome of HPV-related cancers also seems to be favorable following cisplatin-based chemoradiotherapy. However, trials substituting cisplatin with the EGFR inhibitor cetuximab in that patient population have shown inferior survival. Similarly, the investigation of immune checkpoint inhibitors in this setting has not yet led to improved outcomes.

The success of concomitant chemoradiotherapy in patients with unresectable disease has led to the testing of a similar approach in patients with resected intermediate-stage disease as a postoperative therapy. Concomitant chemoradiotherapy produces a significant improvement over postoperative radiation therapy alone for patients whose tumors demonstrate higher risk features, such as extracapsular spread beyond involved lymph nodes, involvement of multiple lymph nodes, or positive margins at the primary site following surgery.

A monoclonal antibody to EGFR (cetuximab) increases survival rates when administered during radiotherapy compared with radiotherapy alone and has been considered for patients unable to tolerate concurrent chemoradiotherapy. The addition of cetuximab to standard chemoradiotherapy regimens has failed to show further improvement in survival and is not recommended.

TREATMENT APPROACHES FOR HPV-RELATED HEAD AND NECK CANCERS

Given consistent observations of high survival rates for patients with advanced HPV-related oropharyngeal tumors using combined-modality treatment strategies, de-escalation protocols have attracted widespread interest. The goal here is to decrease the long-term morbidity resulting from high-dose radiation therapy, including extensive neck fibrosis, swallowing problems, and osteoradionecrosis of the jaw. Current studies are investigating the use of lower radiation doses, the use of induction chemotherapy and subsequent omission of chemotherapy or administration of significantly reduced chemoradiation doses in very good responders, and other strategies. In addition, interest has increased in surgical approaches using robotic surgery, which allows better visualization of the base of tongue and tonsil. While technically feasible, a large number of patients with disease involving multiple lymph nodes will still require postoperative chemoradiotherapy, thus negating the goal of treatment de-escalation. At present, treatment guidelines for HPV-related tumors are identical to carcinogen-induced tumors. It is hoped that de-escalation approaches will be validated by ongoing controlled clinical trials.

RECURRENT AND/OR METASTATIC DISEASE

Five to 10% of patients present with metastatic disease, and 30–50% of patients with locoregionally advanced disease experience recurrence, frequently outside the head and neck region. Patients with recurrent and/or metastatic disease are, with few exceptions, treated with palliative intent. Some patients may require local or regional radiation therapy for pain control, but most are given systemic therapy.

Combination chemotherapy formerly was the first-line systemic therapeutic approach to patients with recurrent disease after prior curative intent surgery and/or chemoradiotherapy or those presenting initially with metastatic disease. In particular, a combination of cisplatin with 5-FU and cetuximab (the EXTREME regimen) was frequently used.

However, immunotherapies have proven to be of value in this setting. In particular, inhibitors of the immunosuppressive lymphocyte surface receptor (PD-1) pathway have shown activity in squamous cell cancers of the head and neck. A randomized trial evaluating the PD-1 inhibitor nivolumab versus traditional chemotherapy in the second-line treatment of patients with recurrent or metastatic disease showed a significant increase in 1-year survival rates with fewer severe treatment-related toxicities. In addition, some responses were of long duration, allowing a cohort of patients to live far beyond the historical median of <1 year. The PD-1 inhibitor pembrolizumab also demonstrated activity in a similarly designed randomized trial.

Pembrolizumab was also compared as single-agent therapy and in combination with cisplatin and 5-FU with prior standard chemotherapy alone (cisplatin, 5-FU, and cetuximab). In this trial, overall survival was improved with pembrolizumab versus chemotherapy as well as with the combination of chemotherapy plus pembrolizumab with the relative benefit correlating with the expression of PD-L1 in the tumor tissue. Patients with tumors high in expression (PD-L1 score >20%; i.e., expression of PD-L1 on 20% of tumor cells) had a marked survival benefit with pembrolizumab as single agent, whereas patients with lower PD-L1 expression had a less impressive but still statistically significant survival benefit. For the group expressing lower levels of PD-L1, the combination of pembrolizumab with chemotherapy showed more substantial benefit. Current standard treatment therefore frequently consists of combination chemoimmunotherapy for patients with low PD-L1 expression, whereas those with higher expression can be treated with immunotherapy alone, especially if overall tumor burden is limited. Patients with no PD-L1 expression may still be treated with the prior EXTREME chemotherapy standard. Patients who experience progression after first-line chemoimmunotherapy or immunotherapy can then be treated with additional single-agent or combination chemotherapy.

EGFR-directed therapies, including monoclonal antibodies (e.g., cetuximab) and tyrosine kinase inhibitors (TKIs) of the EGFR signaling pathway (e.g., erlotinib or gefitinib), have single-agent activity of ~10%. Side effects are usually limited to an acneiform rash and diarrhea (for the TKIs). Drugs targeting specific mutations are under investigation, and patients with *HRAS*-driven tumors can experience shrinkage with the farnesyltransferase inhibitor tipifarnib.

COMPLICATIONS

Complications from treatment of head and neck cancer are usually correlated to the extent of surgery and exposure of normal tissue structures to radiation. The extent of surgery can be limited or surgery can be completely replaced by use of chemotherapy and radiation therapy as the primary approach. Acute complications of radiation include mucositis and dysphagia. Long-term complications include xerostomia, loss of taste, decreased tongue mobility, second malignancies, dysphagia, and neck fibrosis. The complications of chemotherapy vary with the regimen used but usually include myelosuppression, mucositis, nausea and vomiting, and nephrotoxicity (with cisplatin).

The mucosal side effects of therapy can lead to malnutrition and dehydration. Many centers address issues of dentition before starting treatment, and some place feeding tubes to ensure control of hydration and nutrition intake. About 50% of patients develop hypothyroidism from the treatment; thus, thyroid function should be monitored.

SALIVARY GLAND TUMORS

Most benign salivary gland tumors are treated with surgical excision, and patients with invasive salivary gland tumors are treated with surgery and radiation therapy. These tumors may recur regionally; adenoid cystic carcinoma has a tendency to recur along the nerve tracks. Distant metastases may occur as late as 10–20 years after the initial diagnosis. For metastatic disease, therapy is given with palliative intent, usually chemotherapy with doxorubicin and/or a platinum agent by itself or in combination with a taxane. Identification of novel agents with activity in these tumors is a high priority. It is hoped that comprehensive genomic characterization of these rare tumors will facilitate these efforts.

FURTHER READING

Agrawal N et al: Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science 333:1154, 2011.

Burtness B et al: Pembrolizumab alone or with chemotherapy for recurrent/metastatic head and neck squamous cell carcinoma in KEYNOTE-048: Subgroup analysis by programmed death ligand-1 combined positive score. J Clin Oncol 40:2321, 2022.

Chan KCA et al: Analysis of plasma Epstein-Barr virus DNA to screen for nasopharyngeal cancer. N Engl J Med 377:513, 2017.

D'Cruz AK et al: Elective versus therapeutic neck dissection in node-negative oral cancer. N Engl J Med 373:521, 2015.

Ho AL et al: Tipifarnib in head and neck squamous cell carcinoma with HRAS mutations. J Clin Oncol 39:1856, 2021.

Kang H et al: Whole-exome sequencing of salivary gland mucoepidermoid carcinoma. Clin Cancer Res 23:283, 2017.

Lechner M et al: HPV-associated oropharyngeal cancer: Epidemiology, molecular biology and clinical management. Nat Rev Clin Oncol 19:306, 2022.

Mehanna H et al: De-escalation after DE-ESCALATE and RTOG 1016: A Head and Neck Cancer Intergroup framework for future de-escalation studies. J Clin Oncol 38:2552, 2020.

Mody MD et al: Head and neck cancer. Lancet 398:2289, 2021.

Zhang Y et al: Final overall survival analysis of gemcitabine and cisplatin induction chemotherapy in nasopharyngeal carcinoma: A multicenter, randomized phase III trial. J Clin Oncol 40:2420, 2022.

83

Neoplasms of the Lung

Eric K. Singhi, Christine M. Lovly

Over 2.2 million people globally were diagnosed with lung cancer in 2020, ranking it as the second most frequently diagnosed cancer. Although maintaining its status as the leading cause of cancer-related deaths worldwide, claiming 1.8 million lives in the same year, a noteworthy decline in lung cancer fatalities has emerged. This shift is credited to advancements in screening and early detection methods, improved therapeutic strategies, and a reduction in tobacco usage.

Tobacco consumption, established as the primary cause of lung cancer in the mid-twentieth century, was solidified with the U.S. Surgeon General's 1964 report on the health effects of tobacco smoking. After the report, cigarette use declined in North America and parts of Europe, and with it, so did the incidence of lung cancer. Although tobacco smoking remains the leading global cause of lung cancer, accounting for about two-thirds of lung cancer deaths worldwide, ~65% of new lung cancer diagnoses in the United States are in individuals with a former smoking history (smoked ≥100 cigarettes per lifetime, quit ≥1 year) or individuals with no smoking history (smoked <100 cigarettes per lifetime), with one in five women and one in 12 men diagnosed having never smoked.

EPIDEMIOLOGY

Lung cancer is the most common cause of cancer death among American men and women. Approximately 235,000 individuals will be diagnosed with lung cancer in the United States in 2024, and >125,000 individuals will die from the disease. Lung cancer is less common below age 40, with rates increasing until age 80, after which the rate tapers off. The projected lifetime probability of developing lung cancer is estimated at 1 in 16 for males and 1 in 17 for females as of 2023. Since 2006, there has been a steady decline in the incidence of lung cancer, with an annual decrease of 2.5% in men and 1% in women. Despite this positive trend, disparities persist. The incidence of lung cancer varies among racial and ethnic groups, with African Americans exhibiting the highest age-adjusted rates.

While the excess in age-adjusted rates among African Americans occurs only among men, examination of age-specific rates shows that below age 50, mortality from lung cancer is >25% higher among African-American women compared to their Caucasian counterparts. Additionally, African Americans face a 19% lower likelihood of receiving definitive surgical treatment, 11% higher likelihood of not receiving any treatment, and a 16% lower likelihood of surviving 5 years compared to their Caucasian counterparts. Incidence and mortality rates among Hispanics and Native and Asian Americans are ~40–50% of those among Caucasians. Consideration of these disparities prompts reflection on potential equitable improvements in lung cancer care.

RISK FACTORS

Persons who smoke have a 10-fold or greater increased risk of developing lung cancer compared to those who have never smoked. A large-scale genomic study suggested that one genetic mutation is induced for every 15 cigarettes smoked. The risk of lung cancer is lower among persons who quit smoking than among those who continue smoking. The size of the lung cancer risk reduction increases with the length of time the person has quit smoking, although even long-term former smokers have higher risks of lung cancer than those who never smoked. Cigarette smoking has been shown to increase the risk of all major types of lung cancer. Environmental tobacco smoke (ETS) or secondhand smoke is also an established cause of lung cancer. The risk from ETS is less than from active smoking, with about a 20–30% increase in lung cancer observed among never smokers married for many years to smokers, in comparison to the 2000% increase among continuing active smokers. The impact on the development of lung cancer among users of alternate nicotine delivery devices (e-cigarettes or vaping) is undefined. While one large, randomized study demonstrated the superiority of e-cigarettes compared to traditional nicotine replacement therapy in aiding smoking cessation, e-cigarette– or vaping-associated lung injury (EVALI) is an emerging phenomenon that poses risks that may counterbalance the potential benefit in helping patients reduce traditional cigarette consumption and lung cancer risk.

Although cigarette smoking is the cause of the majority of lung cancers, several other risk factors have been identified, including occupational exposure to asbestos, arsenic, bischloromethyl ether, hexavalent chromium, mustard gas, nickel (as in certain nickel-refining processes), and polycyclic aromatic hydrocarbons.

Ionizing radiation is also an established lung carcinogen, most convincingly demonstrated from studies showing increased rates of lung cancer among survivors of the atom bombs dropped on Hiroshima and Nagasaki and large excesses among workers exposed to alpha irradiation from radon in underground uranium mining. Prolonged exposure to low-level radon in homes might impart a risk of lung cancer equal to or greater than that of ETS. Prior lung diseases such as chronic bronchitis, emphysema, and tuberculosis have been linked to increased risks of lung cancer as well. The risk of lung cancer appears to be higher among individuals with low fruit and vegetable intake during adulthood. This observation led to hypotheses that specific nutrients, in particular retinoids and carotenoids, might have chemopreventative effects for lung cancer. However, randomized trials failed to validate this hypothesis.

Smoking Cessation Given the undeniable link between cigarette smoking and lung cancer, physicians must promote complete tobacco avoidance. Stopping tobacco use before middle age avoids >90% of the lung cancer risk attributable to tobacco. Importantly, smoking cessation can even be beneficial in individuals with an established diagnosis of lung cancer, as it is associated with improved overall survival, fewer side effects from therapy, and an overall improvement in quality of life. Consequently, it is important to promote smoking cessation even *after* the diagnosis of lung cancer is established. This remains a challenge as ~30% of patients continue to smoke even after receiving a diagnosis of lung cancer.

Physicians need to understand the essential elements of smoking cessation therapy. Self-help strategies alone only marginally affect quit rates, whereas individual and combined pharmacotherapies in combination with counseling can significantly increase rates of cessation. Therapy with an antidepressant (e.g., bupropion) and nicotine replacement therapy (varenicline, a $\alpha_4\beta_2$ nicotinic acetylcholine receptor partial agonist) are approved by the U.S. Food and Drug Administration (FDA) as first-line treatments for nicotine dependence. In a randomized trial, varenicline was shown to be more efficacious than bupropion or placebo. Prolonged use of varenicline beyond the initial induction phase proved useful in maintaining smoking abstinence. Clonidine and nortriptyline are recommended as second-line treatments. A role for e-cigarettes has not been definitively established **(Chap. 465)**.

Inherited Predisposition to Lung Cancer Exposure to environmental carcinogens, such as those found in tobacco smoke, induces or facilitates the transformation from bronchoepithelial cells to a malignant phenotype. The contribution of carcinogens to transformation is modulated by polymorphic variations in genes that affect aspects of carcinogen metabolism. Certain genetic polymorphisms of the P450 enzyme system, specifically CYP1A1, and chromosome fragility are associated with the development of lung cancer. These genetic variations occur at relatively high frequency in the population, but their contribution to an individual's lung cancer risk is generally low. However, because of their population frequency, the overall impact on lung cancer risk could be high.

First-degree relatives of lung cancer probands have a two- to threefold excess risk of lung cancer and other cancers, many of which are not smoking-related. These data suggest that specific genes and/or genetic variants may contribute to susceptibility to lung cancer. However, very few such genes have yet been identified. Individuals with inherited mutations in *RB* (patients with retinoblastoma living to adulthood) and *TP53* (patients with Li-Fraumeni syndrome) genes may develop lung cancer. Common gene variants involved in lung cancer have identified three separate loci that are associated with lung cancer (5p15, 6p21, and 15q25) and include genes that regulate acetylcholine nicotinic receptors and telomerase production. A rare germline mutation (T790M) involving the epidermal growth factor receptor (EGFR) maybe be linked to lung cancer susceptibility in never smokers. Likewise, a susceptibility locus on chromosome 6q greatly increases lung cancer risk among light and never smokers. A study involving 7700 patients diagnosed with primary lung cancer who underwent germline DNA sequencing and exon-level copy number analysis revealed that 14.9% of them had one or more clinically significant pathogenic germline variants; these variants were mainly found in DNA damage repair genes, suggesting a higher prevalence of pathogenic germline mutations in patients with primary lung cancer than previously suspected. Additionally, the Taiwan National Lung Cancer Early Detection Program, launched in July 2022, provided screening for nonsmoker patients with a positive family history of lung cancer in their first-degree relatives; this program demonstrated a 1.4% cancer detection rate among individuals with a family history of lung cancer. Despite this progress in identifying heritable risk factors for lung cancer, there is still significant work to be done. Currently, no molecular criteria are suitable for selecting patients for more intense screening programs or specific chemopreventive strategies.

PATHOLOGY

The World Health Organization (WHO) defines lung cancer as tumors arising from the respiratory epithelium (bronchi, bronchioles, and alveoli). The WHO classification system divides epithelial lung cancers into four major cell types: small-cell lung cancer (SCLC), adenocarcinoma, squamous cell carcinoma, and large-cell carcinoma; the latter three types are collectively known as non-small-cell carcinomas (NSCLCs) **(Fig. 83-1)**. Small-cell carcinomas consist of small cells with scant cytoplasm, ill-defined cell borders, finely granular nuclear

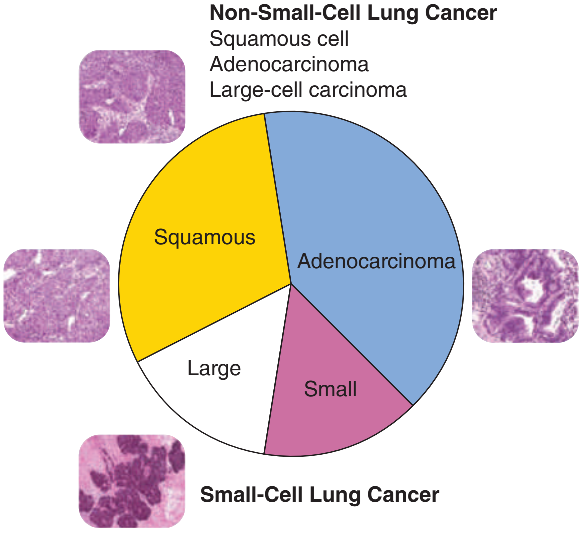

FIGURE 83-1 Histologic subsets of lung cancer.

chromatin, absent or inconspicuous nucleoli, and a high mitotic count. SCLC may be distinguished from NSCLC by the presence of neuroendocrine markers including CD56, neural cell adhesion molecule (NCAM), synaptophysin, chromogranin, and insulinoma-associated protein 1 (INSM1). Adenocarcinomas possess glandular differentiation or mucin production and may show acinar, papillary, lepidic, or solid features or a mixture of these patterns. Squamous cell carcinomas of the lung are morphologically identical to extrapulmonary squamous cell carcinomas and cannot be distinguished by immunohistochemistry alone. Squamous cell tumors show keratinization and/or intercellular bridges that arise from bronchial epithelium. The tumor consists of sheets of cells rather than the three-dimensional groups of cells characteristic of adenocarcinomas. Large-cell carcinomas compose <10% of lung carcinomas. These tumors lack the cytologic and architectural features of small-cell carcinoma and glandular or squamous differentiation. Together, these four histologic types account for ~90% of all epithelial lung cancers.

All histologic types of lung cancer can develop in current and former smokers, although squamous and small-cell carcinomas are more commonly associated with tobacco use. With the decline in cigarette consumption, adenocarcinoma has become the most frequent histologic subtype of lung cancer in the United States. In lifetime, never smokers or former light smokers (<10 pack-year history), women, and younger adults (<60 years), adenocarcinoma tends to be the most common form of lung cancer.

In addition to distinguishing between SCLC and NSCLC, because these tumors have quite different natural histories and therapeutic approaches (see below), it is necessary to classify whether NSCLC is squamous or nonsquamous. The classification system, developed jointly by the International Association for the Study of Lung Cancer, the American Thoracic Society, and the European Respiratory Society, provides an integrated approach to the classification of lung adenocarcinoma that includes clinical, molecular, radiographic, and pathologic information.

Even with advances in lung cancer screening and early detection, many lung cancers are still detected in an advanced stage and may be diagnosed based on small biopsies or cytologic specimens, rendering clear histologic distinctions difficult. In such cases, particularly in patients with advanced-stage disease, a repeat biopsy is recommended to obtain additional tissue for further clarification. The distinction between squamous and nonsquamous lung cancer is viewed as critical to optimal therapeutic decision-making, and a diagnosis of *non-small-cell carcinoma, not otherwise specified* is no longer considered acceptable. This distinction can be achieved using a single marker for adenocarcinoma (thyroid transcription factor-1 or napsin-A) plus a squamous marker (p40 or p63) and/or mucin stains. If tissue is limited and a clear morphologic pattern is evident, a diagnosis can be made without immunohistochemistry staining. In addition to determining histologic subtype, preservation of sufficient specimen material for appropriate molecular testing and programmed death ligand 1 (PD-L1) testing necessary to help guide therapeutic decision-making is recommended (see below).

The terms *adenocarcinoma in situ* and *minimally invasive adenocarcinoma* are now recommended for small solitary adenocarcinomas (≤3 cm) with either pure lepidic growth (term used to describe single-layered growth of atypical cuboidal cells coating the alveolar walls) or predominant lepidic growth with ≤5 mm invasion. Individuals with these entities experience 100% or near 100% 5-year disease-free survival with complete tumor resection. *Invasive adenocarcinomas*, representing >70–90% of surgically resected lung adenocarcinomas, are now classified by their predominant pattern: lepidic, acinar, papillary, and solid patterns. In general, lepidic-predominant subtype has a favorable prognosis, acinar and papillary have an intermediate prognosis, and solid-predominant has a poor prognosis. The terms *signet ring* and *clear cell adenocarcinoma* have been eliminated from the variants of invasive lung adenocarcinoma, whereas the term *micropapillary*, a subtype with a particularly poor prognosis, has been added. Because of prognostic implications, squamous cell carcinoma has also been modified to consist of keratinizing, nonkeratinizing, and basaloid, analogous to head and neck cancers.

IMMUNOHISTOCHEMISTRY

The diagnosis of lung cancer most often rests on the morphologic or cytologic features correlated with clinical and radiographic findings. Immunohistochemistry may be used to verify neuroendocrine differentiation within a tumor, with markers such as neuron-specific enolase (NSE), CD56 or NCAM, synaptophysin, chromogranin, and Leu7. Immunohistochemistry is also helpful in differentiating primary from metastatic adenocarcinomas; thyroid transcription factor-1 (TTF-1), identified in tumors of thyroid and pulmonary origin, is positive in >70% of pulmonary adenocarcinomas and is a reliable indicator of primary lung cancer, provided a thyroid primary has been excluded. A negative TTF-1, however, does not exclude the possibility of a lung primary. TTF-1 is also positive in neuroendocrine tumors of pulmonary and extrapulmonary origin. Napsin-A (Nap-A) is an aspartic protease that plays an important role in maturation of surfactant B7 and is expressed in cytoplasm of type II pneumocytes. In several studies, Nap-A has been reported in >90% of primary lung adenocarcinomas. Notably, a combination of Nap-A and TTF-1 is useful in distinguishing primary lung adenocarcinoma (Nap-A positive, TTF-1 positive) from primary lung squamous cell carcinoma (Nap-A negative, TTF-1 negative) and primary SCLC (Nap-A negative, TTF-1 positive). Cytokeratins (CK) 7 and 20 used in combination can help narrow the differential diagnosis; nonsquamous NSCLC, SCLC, and mesothelioma may stain positive for CK7 and negative for CK20, whereas squamous cell lung cancer often will be both CK7 and CK20 negative. p63 is a useful marker for the detection of NSCLCs with squamous differentiation when used in cytologic pulmonary samples. Mesothelioma can be easily identified ultrastructurally, but it has historically been difficult to differentiate from adenocarcinoma through morphology and immunohistochemical staining. Several markers in the past few years have proven to be more helpful including CK5/6, calretinin, and Wilms tumor gene-1 (*WT-1*), all of which show positivity in mesothelioma.

MOLECULAR PATHOGENESIS

As proposed by Hanahan and Weinberg, virtually all cancer cells acquire six hallmark capabilities: self-sufficiency in growth signals, insensitivity to antigrowth signals, evading apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis. The order in which these hallmark capabilities are acquired is variable. Events leading to acquisition of these hallmarks vary widely, although broadly, cancers arise as a result of accumulations of gain-of-function mutations in oncogenes and loss-of-function mutations in tumor-suppressor genes. Further complicating the study of lung cancer, the sequence of events that leads to disease is clearly different for the various histopathologic entities.

For cancers in general, one theory holds that a small subset of the cells within a tumor (i.e., "stem cells") are responsible for the full malignant behavior of the tumor. As part of this concept, the large bulk of the cells in a cancer are "offspring" of these cancer stem cells. While clonally related to the cancer stem cell subpopulation, most cells by themselves cannot regenerate the full malignant phenotype. The stem cell concept may explain the failure of standard medical therapies to eradicate lung cancers, even when there is a clinical complete response. Disease recurs because therapies do not eliminate the stem cell component, which may be more resistant to therapy. Precise human lung cancer stem cells have yet to be identified.

Among lung cancer histologies, adenocarcinomas have been the most extensively catalogued for recurrent genomic gains and losses as well as for somatic mutations **(Fig. 83-2, Table 83-1)**. While multiple different kinds of aberrations have been found, a major class involves "driver mutations," which are mutations that occur in genes encoding signaling proteins that, when aberrant, drive initiation and maintenance of tumor cells. Importantly, driver mutations can serve as a potential Achilles' heels for tumors, if their gene products can be targeted appropriately. Several driver oncogenes have been described in NSCLC. These include activating mutations in receptor tyrosine kinases, including EGFR, ERBB2/HER2, and MET, and activating mutations in intracellular signaling proteins, such as KRAS and BRAF. Additionally, chromosomal rearrangements may also produce driver oncogenes in lung tumors, such as those chromosomal rearrangements leading to activation of the ALK, ROS1, RET, and NTRK tyrosine kinases. Each of these so-called "driver oncogenes" represents a distinct molecular cohort of lung cancer, with differing prevalence and varying levels of evidence regarding clinical actionability. So-called "targeted therapies" directed against these aberrantly activated oncogenes are discussed below. It is worth noting that most of these driver oncogenes are enriched in nonsquamous tumors.

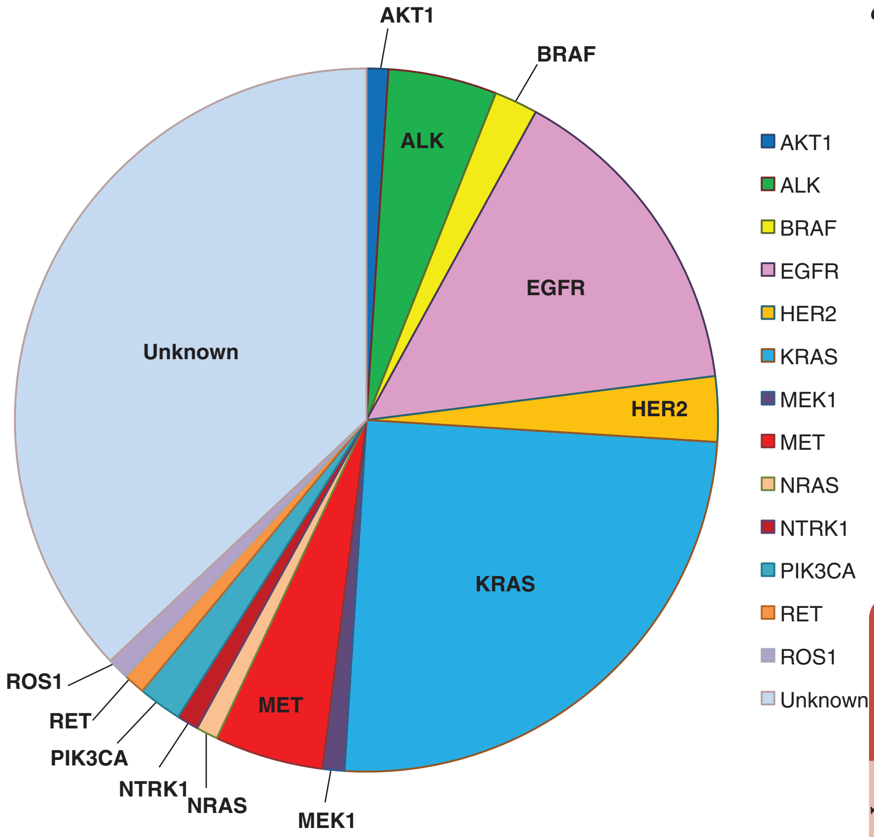

FIGURE 83-2 Driver mutations in lung adenocarcinomas.

Three potential molecular targets have been identified in squamous cell lung carcinomas: *FGFR1* amplification, *DDR2* mutations, and *PIK3CA* mutations/*PTEN* loss, as well as *BRAF* and *MET* (Table 83-1).

A large number of tumor-suppressor genes have also been identified that are inactivated during the pathogenesis of lung cancer. These include *TP53*, *RB1*, *RASSF1A*, *CDKN2A/B*, *LKB1* (*STK11*), and *FHIT*. Nearly 90% of SCLCs harbor mutations in *TP53* and *RB1*. Several tumor-suppressor genes on chromosome 3p appear to be involved in nearly all lung cancers. Allelic loss of this region occurs very early in lung cancer pathogenesis, including in histologically normal smoking-damaged lung epithelium.

EARLY DETECTION AND SCREENING

In lung cancer, clinical outcome is related to the stage at diagnosis, and hence, it is generally assumed that early detection of occult tumors will lead to improved survival. Early detection is a process that involves screening tests, surveillance, diagnosis, and early treatment. Screening refers to the use of tests across a healthy population in order to identify individuals who harbor asymptomatic disease. For a screening program to be successful, the target population must be a high-risk population; the test must be sensitive, specific, accessible, and cost effective; and effective treatment must be available that can reduce mortality. With any screening procedure, it is important to consider the possible influence of *lead-time bias* (detecting the cancer earlier without an effect on survival), *length-time bias* (indolent cancers are detected on screening and may not affect survival, whereas aggressive cancers are likely to cause symptoms earlier in patients and are less likely to be detected), and *overdiagnosis* (diagnosing cancers so slow growing that they are unlikely to cause the death of the patient).

Because a majority of lung cancer patients present with advanced disease beyond the scope of surgical resection, the value of screening for this condition is debated. Indeed, randomized controlled trials conducted in the 1960s to 1980s using screening chest x-rays (CXR), with or without sputum cytology, reported no impact on lung cancer–specific mortality in patients characterized as high risk (males age ≥45 years with a smoking history). These studies have been criticized for their design, statistical analyses, and outdated imaging modalities. In contrast to CXR, low-dose, noncontrast, thin-slice spiral chest computed tomography (LDCT) has emerged as an effective tool to screen for lung cancer. In nonrandomized studies conducted in the 1990s, LDCT scans were shown to detect more lung nodules and cancers than standard CXR in selected high-risk populations (e.g., age ≥60 years and a smoking history of ≥10 pack-years). Notably, up to 85% of the lung cancers discovered in these trials were classified as stage I disease and therefore considered potentially curable with surgical resection.

These data prompted the National Cancer Institute (NCI) to initiate the National Lung Screening Trial (NLST), a randomized study designed to determine if LDCT screening could reduce mortality

TABLE 83-1 Genetic Alterations with Existing Therapies in Non-Small-Cell Lung Cancer (NSCLC)

GENE	ALTERATION	FREQUENCY IN NSCLC	TYPICAL HISTOLOGY
ALK	Rearrangement	3–7%	Adenocarcinoma
BRAF	Mutation	1–3%	Adenocarcinoma
EGFR	Mutation	10–35%	Adenocarcinoma
HER2	Mutation	2–4%	Adenocarcinoma
KRAS	Mutation	15–25%	Adenocarcinoma
MET	Amplification	2–4%	Adenocarcinoma
NRAS	Mutation	1%	Adenocarcinoma
NTRK	Rearrangement	1–2%	Adenocarcinoma
RET	Rearrangement	1–2%	Adenocarcinoma
ROS1	Rearrangement	1–2%	Adenocarcinoma

from lung cancer in high-risk populations as compared with standard posterior anterior CXR. High-risk patients were defined as individuals between 55 and 74 years of age with a ≥30 pack-year history of cigarette smoking; former smokers must have quit within the previous 15 years. Excluded from the trial were individuals with a previous lung cancer diagnosis, a history of hemoptysis, an unexplained weight loss of >15 lb in the preceding year, or a chest computed tomography (CT) within 18 months of enrollment. A total of 53,454 persons were enrolled and randomized to annual screening yearly for 3 years (LDCT screening, n = 26,722; CXR screening, n = 26,732). Any noncalcified nodule measuring ≥4 mm in any diameter found on LDCT and CXR images with any noncalcified nodule or mass was classified as "positive." Participating radiologists had the option of not calling a final screen positive if a noncalcified nodule had been stable on the three screening examinations. Overall, 39.1% of participants in the LDCT group and 16% in the CXR group had at least one positive screening result. Of those who screened positive, the false-positive rate was 96.4% in the LDCT group and 94.5% in the CXR group. This was consistent across all three rounds. In the LDCT group, 1060 cancers were identified compared with 941 cancers in the CXR group (645 vs 572 per 100,000 person-years; relative risk [RR], 1.13; 95% confidence interval [CI], 1.03–1.23). Nearly twice as many stage IA cancers were detected in the LDCT group compared with the CXR group (40% vs 21%). The overall rates of lung cancer death were 247 and 309 deaths per 100,000 participants in the LDCT and CXR groups, respectively, representing a 20% reduction in lung cancer mortality in the LDCT-screened population (95% CI, 6.8–26.7%; $p = .004$). Compared with the CXR group, the rate of death in the LDCT group from *any* cause was reduced by 6.7% (95% CI, 1.2–13.6%; $p = .02$). The number needed to screen (NNTS) to prevent one lung cancer death was calculated to be 320.

The Nelson study was a second randomized trial comparing no screening to CT scans at baseline and in years 1, 3, and 5.5 in 13,195 men and 2594 women. Participants were 50–75 years of age and were current and former smokers with 10 years or less of cessation who smoked >15 cigarettes a day for >25 years or >10 cigarettes daily for >30 years. Participants were selected from four regions in the Netherlands or Belgium and were excluded if they were in moderate or bad self-reported health, were unable to climb two flights of stairs, had a body weight >140 kg, had a CT of the chest within the past year or a history of lung cancer <5 years ago or were still under treatment, or had current or past renal cell carcinoma, melanoma, or breast cancer. The hazard ratio for lung cancer mortality at 10 years was 0.74 (95% CI, 0.60–0.91; $p = .003$) and 0.61 (95% CI, 0.35–1.04; $p = .0543$) in men and women, respectively. These two trials have validated the use of annual CT scans for early detection of lung cancer in high-risk populations.

LDCT screening for lung cancer comes with known risks including a high rate of false-positive results, false-negative results, potential for unnecessary follow-up testing, radiation exposure, overdiagnosis, changes in anxiety level and quality of life, and substantial financial costs. One of the greatest challenges confronting the use of CT screening, in addition to its implementation, is the high false-positive rate. False positives can have a substantial impact on patients through the expense and risk of unneeded further evaluation and emotional stress. The management of these patients usually consists of serial CT scans over time to see if the nodules grow, attempted fine-needle aspirates, or surgical resection. At approximately $300 per scan (NCI estimated cost), the outlay for initial LDCT alone could run into the billions of dollars annually, an expense that only further escalates when factoring in various downstream expenditures an individual might incur in the assessment of positive findings. A formal cost-effectiveness analysis of the NLST demonstrated differences between sex, age, and current smoking status and the method of follow-up. Despite some questions, annual LDCT screening has been recommended for all patients meeting the following updated criteria from the U.S. Preventive Services Task Force as of 2021: individuals aged 50–80 years old with a 20-pack-year or greater smoking history who currently smoke or formerly smoked within the past 15 years. When discussing the option of LDCT screening, use of absolute risks rather than relative risks is helpful because studies indicate the public can process absolute terminology more effectively than relative risk projections. A useful guide has been developed by the NCI to help patients and physicians assess the benefits and harms of LDCT screening for lung cancer **(Table 83-2)**.

TABLE 83-2 The Benefits and Harms of LDCT Screening for Lung Cancer Based on NLST Data

	LDCT	CXR
Benefits: How did CT scans help compared to CXR?		
4 in 1000 fewer died from lung cancer	13 in 1000	17 in 1000
5 in 1000 fewer died from all causes	70 in 1000	75 in 1000
Harms: What problems did CT scans cause compared to CXR?		
223 in 1000 had at least 1 false alarm	365 in 1000	142 in 1000
18 in 1000 had a false alarm leading to an invasive procedure	25 in 1000	7 in 1000
2 in 1000 had a major complication from an invasive procedure	3 in 1000	1 in 1000

Abbreviations: CT, computed tomography; CXR, chest x-ray; LDCT, low-dose computed tomography; NLST, National Lung Screening Trial.

Source: From S Woloshin: Cancer screening campaigns getting past uninformative persuasion. N Engl J Med 367:1167, 2012. Copyright © (2012) Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

CLINICAL MANIFESTATIONS

Over half of all patients diagnosed with lung cancer present with locally advanced or metastatic disease at the time of diagnosis. The majority of patients present with signs, symptoms, or laboratory abnormalities that can be attributed to the primary lesion, local tumor growth, invasion or obstruction of adjacent structures, growth at distant metastatic sites, or a paraneoplastic syndrome **(Tables 83-3** and **83-4)**. A history of chronic cough with or without hemoptysis in a current or former smoker with chronic obstructive pulmonary disease (COPD) age 40 years or older should prompt a thorough investigation for lung cancer even in the face of a normal CXR. A persistent pneumonia without constitutional symptoms and unresponsive to repeated courses of antibiotics also should prompt an evaluation for the underlying cause. Lung cancer occurring in individuals who have never smoked is more prevalent in, though not limited to, women and East Asians. These patients also typically present at a younger age than their smoking counterparts at the time of diagnosis. The clinical presentation of lung cancer in never smokers tends to mirror that of current and former smokers.

Patients with central or endobronchial growth of the primary tumor may present with cough, hemoptysis, wheeze, stridor, dyspnea, or postobstructive pneumonia. Peripheral growth of the primary tumor may cause pain from pleural or chest wall involvement, dyspnea on a restrictive basis, and symptoms of a lung abscess resulting from tumor

TABLE 83-3 Presenting Signs and Symptoms of Lung Cancer

SYMPTOM AND SIGNS	RANGE OF FREQUENCY
Cough	8–75%
Weight loss	0–68%
Dyspnea	3–60%
Chest pain	20–49%
Hemoptysis	6–35%
Bone pain	6–25%
Clubbing	0–20%
Fever	0–20%
Weakness	0–10%
Superior vena cava obstruction	0–4%
Dysphagia	0–2%
Wheezing and stridor	0–2%

Source: Reproduced with permission from MA Beckles: Initial evaluation of the patient with lung cancer. Symptoms, signs, laboratory tests, and paraneoplastic syndromes. Chest 123:97, 2003.

TABLE 83-4 Clinical Findings Suggestive of Metastatic Disease

Symptoms elicited in history	• Constitutional: weight loss >10 lb • Musculoskeletal: pain • Neurologic: headaches, syncope, seizures, extremity weakness, recent change in mental status
Signs found on physical examination	• Lymphadenopathy (>1 cm) • Hoarseness, superior vena cava syndrome • Bone tenderness • Hepatomegaly (>13 cm span) • Focal neurologic signs, papilledema • Soft-tissue mass
Routine laboratory tests	• Hematocrit, <40% in men; <35% in women • Elevated alkaline phosphatase, GGT, SGOT, and calcium levels

Abbreviations: GGT, gamma-glutamyltransferase; SGOT, serum glutamic-oxaloacetic transaminase.

Source: Reproduced with permission from GA Silvestri et al: The noninvasive staging of non-small cell lung cancer. Chest 123:147S, 2003.

cavitation. Regional spread of tumor in the thorax (by contiguous growth or by metastasis to regional lymph nodes) may cause tracheal obstruction, esophageal compression with dysphagia, recurrent laryngeal nerve paralysis with hoarseness, phrenic nerve palsy with elevation of the hemidiaphragm and dyspnea, and sympathetic nerve paralysis with Horner's syndrome (i.e., enophthalmos, ptosis, miosis, and anhidrosis). Malignant pleural effusions can cause pain, dyspnea, or cough. Pancoast (or superior sulcus tumor) syndromes result from local extension of a tumor growing in the apex of the lung with involvement of the eighth cervical and first and second thoracic nerves, and present with shoulder pain that characteristically radiates in the ulnar distribution of the arm, often with radiologic destruction of the first and second ribs. Often Horner's syndrome and Pancoast syndrome coexist. Other problems of regional spread include superior vena cava syndrome from vascular obstruction; pericardial and cardiac extension with resultant tamponade, arrhythmia, or cardiac failure; lymphatic obstruction with resultant pleural effusion; and lymphangitic spread through the lungs with hypoxemia and dyspnea. In addition, lung cancer can spread transbronchially, producing tumor growth along multiple alveolar surfaces with impairment of gas exchange, respiratory insufficiency, dyspnea, hypoxemia, and sputum production. Constitutional symptoms may include anorexia, weight loss, weakness, fever, and night sweats. These parameters cannot clearly distinguish SCLC from NSCLC or even from neoplasms metastatic to lungs.

Extrathoracic metastatic disease is found at autopsy in >50% of patients with squamous carcinoma, 80% of patients with adenocarcinoma and large-cell carcinoma, and >95% of patients with SCLC. Approximately one-third of patients present with symptoms as a result of distant metastases. Lung cancer metastases may occur in virtually every organ system, and the site of metastatic involvement largely determines other symptoms. Patients with brain metastases may present with headache, nausea and vomiting, seizures, or neurologic deficits. Patients with bone metastases may present with pain, pathologic fractures, or spinal cord compression. The latter may also occur with epidural metastases. Individuals with bone marrow invasion may present with cytopenias or leukoerythroblastosis. Those with liver metastases may present with hepatomegaly, right upper quadrant pain, fever, anorexia, and weight loss. Liver dysfunction and biliary obstruction are rare. Adrenal metastases are common but rarely cause pain or adrenal insufficiency unless they are large.

Paraneoplastic syndromes are common in patients with lung cancer, especially those with SCLC, and may be the presenting finding or the first sign of recurrence. In addition, paraneoplastic syndromes may mimic metastatic disease and, unless detected, lead to inappropriate palliative rather than curative treatment. Often the paraneoplastic syndrome may be relieved with successful treatment of the tumor. In some cases, the pathophysiology of the paraneoplastic syndrome is known, particularly when a hormone with biologic activity is secreted by a tumor. However, in many cases, the pathophysiology is unknown. Systemic symptoms of anorexia, cachexia, weight loss (seen in 30% of patients), fever, and suppressed immunity are paraneoplastic syndromes of unknown etiology or at least not well defined. Weight loss >10% of total body weight is considered a bad prognostic sign. Endocrine syndromes are seen in 12% of patients; hypercalcemia resulting from ectopic production of parathyroid hormone (PTH) or, more commonly, PTH-related peptide is the most common life-threatening metabolic complication of malignancy, primarily occurring with squamous cell carcinomas of the lung. Clinical symptoms include nausea, vomiting, abdominal pain, constipation, polyuria, thirst, and altered mental status.

Hyponatremia may be caused by the syndrome of inappropriate secretion of antidiuretic hormone (SIADH) or possibly atrial natriuretic peptide (ANP) **(Chap. 98)**. SIADH resolves within 1–4 weeks of initiating chemotherapy in the vast majority of cases. During this period, serum sodium can usually be managed and maintained above 128 mEq/L via fluid restriction. Demeclocycline can be a useful adjunctive measure when fluid restriction alone is insufficient. Vasopressin receptor antagonists like tolvaptan also have been used in the management of SIADH. However, the use of tolvaptan has significant limitations including liver injury and overly rapid correction of the hyponatremia, which can lead to irreversible neurologic injury. Likewise, the cost of tolvaptan may be prohibitive. Of note, patients with ectopic ANP may have worsening hyponatremia if sodium intake is not concomitantly increased. Accordingly, if hyponatremia fails to improve or worsens after 3–4 days of adequate fluid restriction, plasma levels of ANP should be measured to determine the causative syndrome.

Ectopic secretion of ACTH by SCLC and pulmonary carcinoids usually results in additional electrolyte disturbances, especially hypokalemia, rather than the changes in body habitus that occur in Cushing's syndrome from a pituitary adenoma **(Chap. 98)**. Treatment with standard medications, such as metyrapone and ketoconazole, is largely ineffective due to extremely high cortisol levels. The most effective strategy for management of the Cushing's syndrome is effective treatment of the underlying SCLC. Bilateral adrenalectomy may be considered in extreme cases.

Skeletal–connective tissue syndromes include clubbing in 30% of cases (usually NSCLCs) and hypertrophic primary osteoarthropathy in 1–10% of cases (usually adenocarcinomas). Patients may develop periostitis, causing pain, tenderness, and swelling over the affected bones and a positive bone scan. Neurologic–myopathic syndromes are seen in only 1% of patients but are dramatic and include the myasthenic Eaton-Lambert syndrome and retinal blindness with SCLC, whereas peripheral neuropathies, subacute cerebellar degeneration, cortical degeneration, and polymyositis are seen with all lung cancer types. Many of these are caused by autoimmune responses such as the development of anti-voltage-gated calcium channel antibodies in Eaton-Lambert syndrome. Patients with this disorder present with proximal muscle weakness, usually in the lower extremities, occasional autonomic dysfunction, and rarely, cranial nerve symptoms or involvement of the bulbar or respiratory muscles. Depressed deep tendon reflexes are frequently present. In contrast to patients with myasthenia gravis, strength improves with serial effort. Some patients who respond to chemotherapy will have resolution of the neurologic abnormalities. Thus, chemotherapy is the initial treatment of choice. Paraneoplastic encephalomyelitis and sensory neuropathies, cerebellar degeneration, limbic encephalitis, and brainstem encephalitis occur in SCLC in association with a variety of antineuronal antibodies such as anti-Hu, anti-CRMP5, and ANNA-3. Paraneoplastic cerebellar degeneration may be associated with anti-Hu, anti-Yo, or P/Q calcium channel autoantibodies. Coagulation or thrombotic or other hematologic manifestations occur in 1–8% of patients and include migratory venous thrombophlebitis (Trousseau's syndrome), nonbacterial thrombotic (marantic) endocarditis with arterial emboli, and disseminated intravascular coagulation with hemorrhage, anemia, granulocytosis, and leukoerythroblastosis. Thrombotic disease complicating cancer is usually a poor prognostic sign. Cutaneous manifestations such as dermatomyositis and acanthosis nigricans are uncommon (1%), as are the renal manifestations of nephrotic syndrome and glomerulonephritis (≤1%).

DIAGNOSING LUNG CANCER

Tissue sampling is required to confirm a diagnosis in all patients with suspected lung cancer. In patients with suspected metastatic disease, a biopsy of a distant site of disease is preferred for concurrent tissue and staging confirmation. Given the greater emphasis placed on molecular and PD-L1 testing for NSCLC patients, a core biopsy is preferred to ensure adequate tissue for analysis. Tumor tissue may be obtained via minimally invasive techniques such as bronchial or transbronchial biopsy during fiberoptic bronchoscopy, by fine-needle aspiration (FNA) or percutaneous biopsy using image guidance, or via endobronchial ultrasound (EBUS)-guided biopsy. Depending on the location, lymph node sampling may occur via transesophageal endoscopic ultrasound (EUS)-guided biopsy, EBUS-guided biopsy, or blind biopsy. In patients with suspected metastatic disease, a diagnosis may be confirmed by bronchoscopy, percutaneous biopsy of a soft tissue mass, lytic bone lesion, bone marrow, pleural or liver lesion, or an adequate cell block obtained from a malignant pleural effusion. In patients with a suspected malignant pleural effusion, if the initial thoracentesis is negative, repeat thoracentesis is essential. Although the majority of pleural effusions are due to malignant disease, particularly if they are exudative or bloody, some may be parapneumonic. In the absence of distant disease, such patients should be considered for possible curative treatment.

The diagnostic yield of any biopsy depends on several factors including location (accessibility) of the tumor, tumor size, tumor type, and technical aspects of the diagnostic procedure including the experience level of the bronchoscopist and pathologist. In general, central lesions such as squamous cell carcinomas, small-cell carcinomas, or endobronchial lesions such as carcinoid tumors are more readily diagnosed by bronchoscopic examination, whereas peripheral lesions such as adenocarcinomas and large-cell carcinomas are more amenable to transthoracic biopsy.

Biomarker testing in lung cancer plays an essential role in guiding a precision-based medicine approach for patient care. Ideally performed at the time of initial diagnosis, biomarker testing is recommended for all patients with advanced stage NSCLC by most current guidelines, with consideration for those with squamous histology. Biomarker testing aims to identify specific genetic alterations, or molecular markers, within the tumor cells. Traditionally, the gold standard process involves obtaining a tissue sample from the patient, which is then subjected to advanced techniques such as next-generation sequencing (NGS) to complete multigene sequencing. By completing this testing, clinicians can discern the presence of driver oncogenes to allow for tailored treatment approaches specific to the underlying molecular profile of the tumor such as with targeted therapies. Of note, immunohistochemical staining on tumor tissue should be employed for PD-L1 testing, quantified as tumor proportion score (TPS), as the level of PD-L1 expression serves as an additional crucial biomarker in lung cancer management.

Advancements in the detection of circulating tumor DNA (ctDNA) have emerged as a noninvasive strategy to perform a "liquid biopsy" with a blood sample to assist in diagnosis and treatment monitoring. This strategy may be particularly useful to overcome tissue limitations, such as in cases where obtaining a tissue biopsy may be challenging or not feasible due to a tumor's location. The role of a liquid biopsy continues to evolve, especially as the sensitivity and specificity of this biopsy strategy continue to improve.

STAGING LUNG CANCER

Lung cancer staging consists of two parts: first, a determination of the location of the tumor and possible metastatic sites (anatomic staging), and second, an assessment of a patient's ability to withstand various antitumor treatments (physiologic staging). All patients with lung cancer should have a complete history and physical examination, with evaluation of all other medical problems, determination of performance status (i.e., standardized measure assessing a patient's physical functioning and ability to perform activities of daily living to determine their overall health and suitability for treatment), and history of weight loss. Staging with regard to a patient's potential for surgical resection is principally applicable to NSCLC.

■ ANATOMIC STAGING OF PATIENTS WITH LUNG CANCER

The accurate staging of patients with NSCLC is essential for determining the appropriate treatment in patients with resectable disease and for avoiding unnecessary surgical procedures in patients with advanced disease. All patients with NSCLC should undergo initial radiographic imaging with CT scan, positron emission tomography (PET), or preferably PET-CT. PET scanning attempts to identify sites of malignancy based on glucose metabolism by measuring the uptake of ^{18}F-fluorodeoxyglucose (FDG). Rapidly dividing cells, presumably in the lung tumors, will preferentially take up ^{18}F-FDG and appear as a "hot spot." To date, PET has been mostly used for staging and detection of metastases in lung cancer and in the detection of nodules >15 mm in diameter. Combined ^{18}F-FDG PET-CT imaging has been shown to improve the accuracy of staging in NSCLC compared to visual correlation of PET and CT or either study alone. PET-CT has been found to be superior in identifying pathologically enlarged mediastinal lymph nodes and extrathoracic metastases. A standardized uptake value (SUV) of >2.5 on PET is highly suspicious for malignancy. False negatives can be seen in diabetes, in lesions <8 mm, and in slow-growing tumors (e.g., carcinoid tumors or well-differentiated adenocarcinoma). False positives can be seen in certain infections and granulomatous disease (e.g., tuberculosis). Thus, PET should never be used alone to diagnose lung cancer, mediastinal involvement, or metastases. Confirmation with tissue biopsy is required. For brain metastases, magnetic resonance imaging (MRI) is the most effective method. MRI can also be useful in selected circumstances, such as superior sulcus tumors to rule out brachial plexus involvement, but in general, MRI does not play a major role in NSCLC staging.

If imaging is concerning for metastatic disease, biopsy of a distant site of disease is preferred for concurrent tissue diagnosis and confirmation of staging. In patients in whom distant metastatic disease has been ruled out, lymph node status needs to be assessed via minimally invasive techniques such as those mentioned above, including EBUS-guided mediastinal staging, and/or invasive techniques such as mediastinoscopy, mediastinotomy, thoracoscopy, or thoracotomy. Approximately one-quarter to one-half of patients diagnosed with NSCLC will have mediastinal lymph node metastases at the time of diagnosis. A standard nomenclature for referring to the location of lymph nodes involved with lung cancer has evolved **(Fig. 83-3)**. Lymph node sampling is recommended in all patients with enlarged nodes detected by CT or PET scan and in patients with large tumors or tumors occupying the inner third of the lung. The extent of mediastinal lymph node involvement is important in determining the appropriate definitive intent treatment strategy: neoadjuvant/perioperative therapy followed by surgical resection versus surgical resection followed by adjuvant therapy versus combined chemoradiation followed by immunotherapy consolidation (durvalumab).

In SCLC patients, current staging recommendations include a PET-CT scan and MRI of the brain (positive in 10% of asymptomatic patients). Bone marrow biopsies and aspirations are rarely performed now given the low incidence of isolated bone marrow metastases. Confirmation of metastatic disease, ipsilateral or contralateral lung nodules, or metastases beyond the mediastinum may be achieved by the same modalities recommended earlier for patients with NSCLC.

If a patient has signs or symptoms of spinal cord compression (pain, weakness, paralysis, urinary retention), a spinal CT or MRI scan should be performed. If metastases are evident on imaging, a neurosurgeon should be consulted for possible palliative surgical resection and/or a radiation oncologist should be consulted for palliative radiotherapy to the site of compression. If signs or symptoms of leptomeningeal disease develop at any time in a patient with lung cancer, an MRI of the brain and spinal cord should be performed, as well as a lumbar puncture, for evaluation and detection of malignant cells in the cerebrospinal fluid. If the lumbar puncture is negative, a repeat procedure should be considered. Leptomeningeal disease remains a clinical challenge, as there is currently no approved therapy for the specific treatment of leptomeningeal disease.

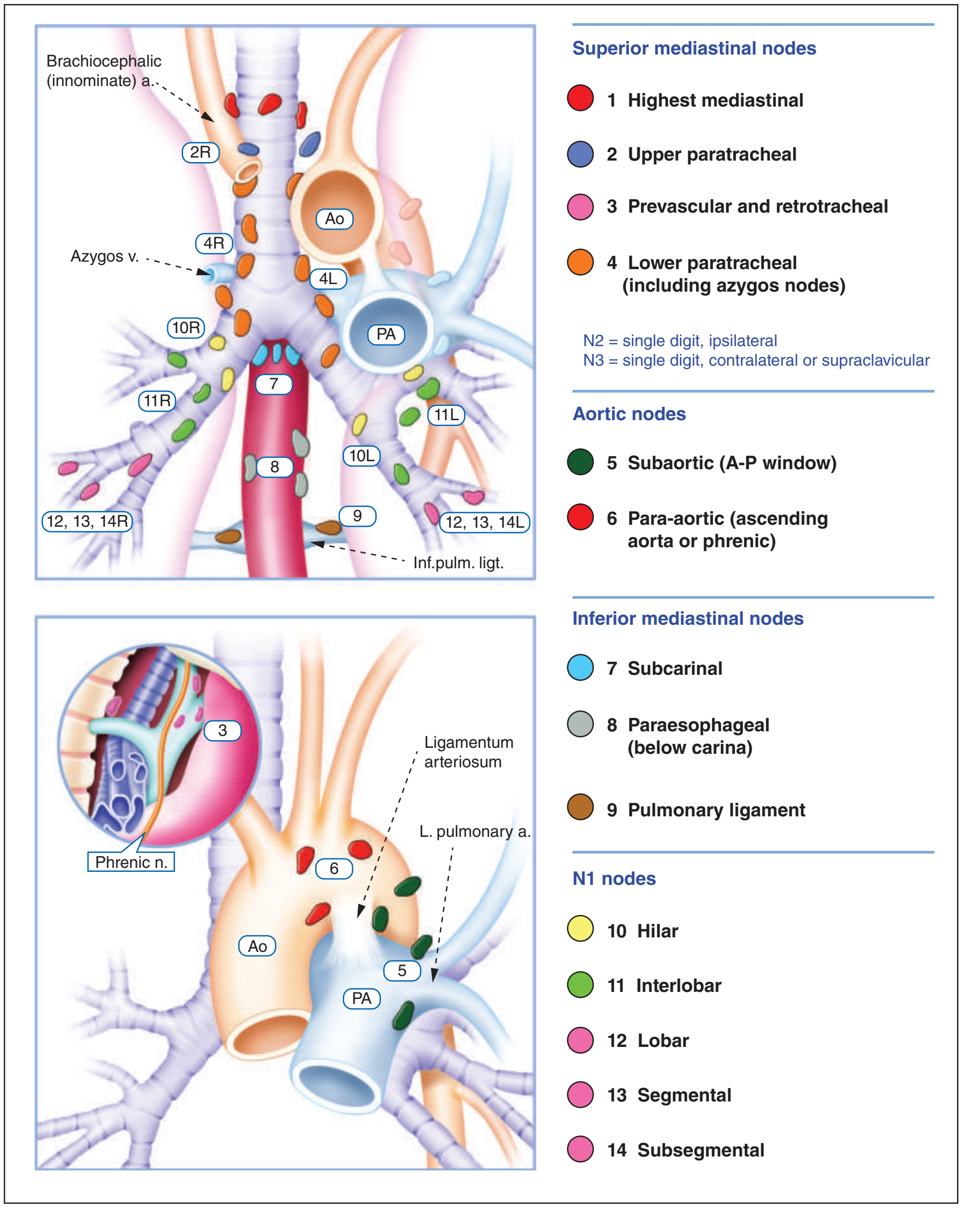
Brachiocephalic (innominate) a.
2R
Ao
4R
Azygos v.
4L
10R
PA
7
11R
11L
8
10L
12, 13, 14R
9
12, 13, 14L
Inf.pulm. ligt.
3
Ligamentum arteriosum
L. pulmonary a.
Phrenic n.
6
Ao
5
PA
Superior mediastinal nodes
1 Highest mediastinal
2 Upper paratracheal
3 Prevascular and retrotracheal
4 Lower paratracheal (including azygos nodes)
N2 = single digit, ipsilateral
N3 = single digit, contralateral or supraclavicular
Aortic nodes
5 Subaortic (A-P window)
6 Para-aortic (ascending aorta or phrenic)
Inferior mediastinal nodes
7 Subcarinal
8 Paraesophageal (below carina)
9 Pulmonary ligament
N1 nodes
10 Hilar
11 Interlobar
12 Lobar
13 Segmental
14 Subsegmental

FIGURE 83-3 Lymph node stations in staging non-small-cell lung cancer. The International Association for the Study of Lung Cancer (IASLC) lymph node map, including the proposed grouping of lymph node stations into "zones" for the purposes of prognostic analyses. a., artery; Ao, aorta; Inf. pulm. ligt., inferior pulmonary ligament; n., nerve; PA, pulmonary artery; v., vein.

■ STAGING SYSTEM FOR NON-SMALL-CELL LUNG CANCER

The tumor-node-metastasis (TNM) international staging system provides useful prognostic information and is used to stage all patients with NSCLC. The various T (tumor size), N (regional node involvement), and M (presence or absence of distant metastasis) stages are combined to form different stage groups **(Tables 83-5 and 83-6)**. The eighth edition of the TNM staging system went into effect in 2018. T1 tumors are divided into tumors ≤1 cm (T1a), >1 cm and ≤2 cm (T1b), and >2 cm and ≤3 cm (T1c). T2 tumors are those that are >3 cm but ≤5 cm, involve the visceral pleura or main bronchus, or are associated with atelectasis; T2a tumors are >3 cm and ≤4 cm, and T2b are >4 cm and ≤5 cm. T3 tumors are >5 cm and ≤7 cm. T3 tumors also include tumors with invasion into local structures such as the chest wall and diaphragm and with additional nodules in the same lobe. T4 tumors include tumors >7 cm or tumors of any size with invasion into mediastinum, heart, great vessels, trachea, esophagus, or spine or with multiple nodules in the ipsilateral lung. Lymph node staging depends on metastasis to ipsilateral pulmonary or hilar nodes (N1), mediastinal or subcarinal nodes (N2), or contralateral mediastinal, hilar, or supraclavicular nodes (N3). Patients with metastasis may be classified as M1a (malignant pleural or pericardial effusion, pleural nodules, or nodules in the contralateral lung), M1b (single distant metastasis to a single organ; e.g., bone, liver, adrenal, or brain metastasis), or M1c (multiple metastases to a single organ or metastases to multiple organs). The ninth edition of the TNM classification is scheduled to be introduced sometime in 2024.

The effect of stage on survival is illustrated in **Fig. 83-4**. Approximately 15% of patients have localized disease that can be treated with curative attempt (surgery or radiotherapy), about a quarter have local or regional disease that may or may not be amenable to a curative attempt, and half have metastatic disease at the time of diagnosis. In 10%, the extent of disease is undefined.

■ STAGING SYSTEM FOR SMALL-CELL LUNG CANCER

In patients with SCLC, it is now recommended that both the Veterans Administration system and the American Joint Committee on Cancer/International Union Against Cancer eighth edition system (TNM) be used to classify the tumor stage. The Veterans Administration system is a distinct two-stage system dividing patients into those with limited- or extensive-stage disease. Patients with limited-stage disease (LD) have cancer that is confined to the ipsilateral hemithorax and can be encompassed within a tolerable radiation port. Thus, contralateral supraclavicular nodes, recurrent laryngeal nerve involvement, and superior vena caval obstruction can all be part of LD. Patients with extensive-stage disease (ED) have overt metastatic disease by imaging or physical examination. Cardiac tamponade, malignant pleural effusion, and bilateral pulmonary parenchymal involvement generally qualify disease as ED, because the involved organs cannot be encompassed safely or effectively within a single radiation therapy port. Sixty to 70% of patients are diagnosed with ED at presentation. The TNM staging system is preferred in the rare SCLC patient presenting with what appears to be clinical stage I disease (see above).

■ PHYSIOLOGIC STAGING

Patients with lung cancer often have other comorbid conditions related to smoking including cardiovascular disease and COPD. To improve their preoperative condition, correctable problems (e.g., anemia, electrolyte and fluid disorders, infections, cardiac disease, and arrhythmias) should be addressed, appropriate chest physical therapy should be instituted, and patients should be encouraged to stop smoking. Patients with a forced expiratory volume in 1 s (FEV_1) of >2 L or >80% of predicted may tolerate a pneumonectomy, and those with an FEV_1 >1.5 L may have adequate reserve for a lobectomy. In patients with borderline lung function but a resectable tumor, cardiopulmonary exercise testing could be performed as part of the physiologic evaluation. This test allows an estimate of the maximal oxygen consumption (VO_{2max}).

TABLE 83-5 TNM Staging System for Lung Cancer (Eighth Edition)

Primary Tumor (T)	
T1	Tumor ≤3 cm diameter, surrounded by lung or visceral pleura, without invasion more proximal than lobar bronchus
T1mi	Minimally invasive adenocarcinoma (pure lepidic pattern, <3 cm in greatest dimension and <5 mm invasion)—T1a (size ≤1 cm)—T1b (1 cm < size ≤ 2 cm)—T1c (2 cm < size ≤ 3 cm)
T2	Tumor >3 cm but ≤5 cm, or tumor with any of the following features: Involves main bronchus without carina, regardless of distance from carina Invades visceral pleura Associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung
T2a	Tumor >3 cm but ≤4 cm
T2b	Tumor >4 cm but ≤5 cm
T3	Tumor >5 cm but ≤7 cm or any of the following: Directly invades any of the following: chest wall, pericardium, phrenic nerve Separate (satellite) tumor nodules in the same lobe
T4	Tumor >7 cm or any tumor with invasion of mediastinum, diaphragm, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, or carina, or with separate (satellite) tumor nodules in a different ipsilateral lobe
Nodal Stage (N)	
N0	No regional lymph node metastases
N1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Metastasis in contralateral mediastinal or hilar lymph node(s); ipsilateral/contralateral scalene/supraclavicular lymph node(s)
Metastases (M)	
M0	No distant metastasis
M1	Distant metastasis
M1a	Separate tumor nodule(s) in a contralateral lobe; pleural/pericardial nodule/malignant effusion
M1b	Single extrathoracic metastasis, including single nonregional lymph node
M1c	Multiple extrathoracic metastases in one or more organs

Abbreviation: TNM, tumor-node-metastasis.

A VO_{2max} <15 mL/(kg.min) predicts for a higher risk of postoperative complications. Patients deemed unable to tolerate lobectomy or pneumonectomy from a pulmonary functional standpoint may be candidates for more limited resections, such as wedge or anatomic segmental resection, although such procedures are associated with significantly higher rates of local recurrence and a trend toward decreased overall survival. All patients should be assessed for cardiovascular risk using American College of Cardiology and American Heart Association guidelines. A myocardial infarction within the past 3 months is a contraindication to thoracic surgery because 20% of patients will die of reinfarction. An infarction in the past 6 months is a relative contraindication. Other major contraindications include uncontrolled arrhythmias, an FEV_1 of <1 L, CO_2 retention (resting PCO_2 >45 mmHg), DL_{CO} <40%, and severe pulmonary hypertension.

TABLE 83-6 TNM Stage Groupings, Eighth Edition

T/M	Subcategory	N0	N1	N2	N3
T1	T1a	IA1	IIB	IIIA	IIIB
	T1b	IA2	IIB	IIIA	IIIB
	T1c	IA3	IIB	IIIA	IIIB
T2	T2a	IB	IIB	IIIA	IIIB
	T2b	IIA	IIB	IIIA	IIIB
T3	T3	IIB	IIIA	IIIB	IIIC
T4	T4	IIIA	IIIA	IIIB	IIIC
M1	M1a	IVA	IVA	IVA	IVA
	M1b	IVA	IVA	IVA	IVA
	M1c	IVB	IVB	IVB	IVB

TREATMENT

Non-Small-Cell Lung Cancer

OCCULT AND STAGE 0 CARCINOMAS

Patients with severe atypia on sputum cytology have an increased risk of developing lung cancer compared to those without atypia. In the uncommon circumstance where malignant cells are identified in a sputum or bronchial washing specimen but the chest imaging appears normal (TX tumor stage), the lesion must be localized. More than 90% of tumors can be localized by meticulous examination of the bronchial tree with a fiberoptic bronchoscope under general anesthesia and collection of a series of differential brushings and biopsies. Surgical resection following bronchoscopic localization has been shown to improve survival compared to no treatment. Close follow-up of these patients is indicated because of the high incidence of second primary lung cancers (5% per patient per year).

SOLITARY PULMONARY NODULE AND “GROUND-GLASS” OPACITIES

A solitary pulmonary nodule is defined as an x-ray density completely surrounded by normal aerated lung with circumscribed margins, of any shape, usually 1–6 cm in greatest diameter. The approach to a patient with a solitary pulmonary nodule is based on an estimate of the probability of cancer, determined according to the patient’s smoking history, age, and characteristics on imaging **(Table 83-7)**. Prior CXRs and CT scans should be obtained if available for comparison. A PET scan may be useful if the lesion is >7–8 mm in diameter. If no diagnosis is apparent, Mayo investigators reported that clinical characteristics (age, cigarette smoking status, and prior cancer diagnosis) and three radiologic characteristics (nodule diameter, spiculation, and upper lobe location) were independent predictors of malignancy. At present, only two radiographic criteria are thought to predict the benign nature of a solitary pulmonary nodule: lack of growth over a period >2 years and certain characteristic patterns of calcification. Calcification alone, however, does not exclude malignancy; a dense central nidus, multiple punctuate foci, and “bull’s eye” (granuloma) and “popcorn ball” (hamartoma) calcifications are highly suggestive of a benign lesion. In contrast, a relatively large lesion, lack of or asymmetric calcification, chest symptoms, associated atelectasis, pneumonitis, or growth of the lesion revealed by comparison with an old x-ray or CT scan or a positive PET scan may be suggestive of a malignant process and warrant further attempts to establish a histologic diagnosis. An algorithm for assessing these lesions is shown in **Fig. 83-5**.

Since the advent of screening CTs, small “ground-glass” opacities (GGOs) have often been observed, particularly as the increased sensitivity of CTs enables detection of smaller lesions. Many of these GGOs, when biopsied, are found to be atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), or minimally invasive adenocarcinoma (MIA). AAH is usually a nodule of <5 mm and is minimally hazy, also called nonsolid or ground glass (i.e., hazy slightly increased attenuation, no solid component, and preservation of bronchial and vascular margins). On thin-section CT, AIS is usually a nonsolid nodule and tends to be slightly more opaque than AAH. MIA is mainly solid, usually with a small (<5 mm) central solid component. However, overlap exists among the imaging features of the preinvasive and minimally invasive lesions in

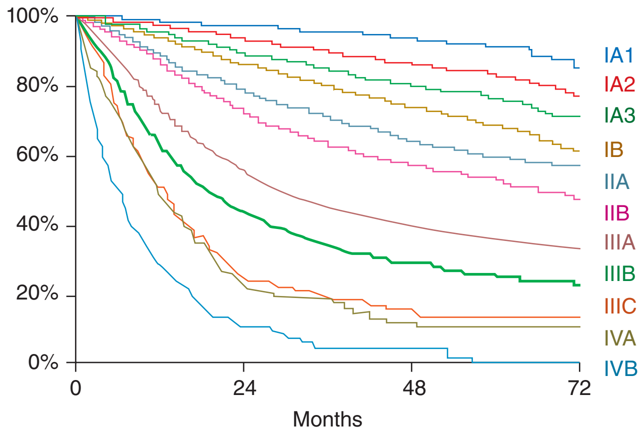

Stage	24 months	60 months
IA1 IA2 IA3	97% 94% 90%	92% 83% 77%
IB	87%	68%
IIA	79%	60%
IIB	72%	53%
IIIA	55%	36%
IIIB	44%	26%
IIIC	24%	13%
IVA	23%	10%
IVB	10%	0%

FIGURE 83-4 Influence of non-small-cell lung cancer stage on survival. Overall survival by non-small-cell lung cancer stage, according to eighth edition groupings using entire database available for the eighth edition as of 2016.

the lung adenocarcinoma spectrum. Lepidic adenocarcinomas are usually solid but may be nonsolid. Likewise, the small invasive adenocarcinomas also are usually solid but may exhibit a small nonsolid component.

PREFACE TO MANAGEMENT OF LUNG CANCER

The landscape of lung cancer management, spanning from screening and early detection to addressing early-stage, locally advanced, and distant disease, is undergoing continuous evolution. The dynamic nature of this evolution is best exemplified by the approval of 12 drugs in the United States only between 2022 and 2023. Within the confines of this single chapter, presenting the most current and comprehensive management recommendations proves challenging, if not impossible, due to the accelerated pace of the evolving treatment armamentarium. Therefore, it is recommended to consult the latest professional guidelines for the most up-to-date best practices **(Table 83-8)**.

MANAGEMENT OF STAGES I AND II NSCLC

Surgical Resection of Stage I and II NSCLC Surgical resection, ideally by an experienced thoracic surgeon, is the treatment of choice for patients with clinical stage I and II NSCLC who are able to tolerate the procedure. Operative mortality rates for patients resected by thoracic or cardiothoracic surgeons are lower compared to general surgeons. Moreover, survival rates are higher in patients who undergo resection in facilities with a high surgical volume compared to those performing fewer than 70 procedures per year, even though the higher-volume facilities often serve older and less socioeconomically advantaged populations. The improvement in survival is most evident in the immediate postoperative period. In patients with stage I NSCLC, lobectomy is superior to wedge resection with respect to rates of local recurrence. There is also a trend toward improvement in overall survival. In patients with comorbidities, compromised pulmonary reserve, and small peripheral lesions, a limited resection, wedge resection, or segmentectomy (potentially by video-assisted thoracoscopic surgery) may be reasonable surgical options. Pneumonectomy is reserved for patients with central tumors and should be performed only in patients with excellent pulmonary reserve. According to the Surveillance, Epidemiology, and End Results (SEER) data from 2022, patients with localized disease have a 5-year survival rate of 63%.

TABLE 83-7 Assessment of Risk of Cancer in Patients with Solitary Pulmonary Nodules

	RISK		
VARIABLE	**LOW**	**INTERMEDIATE**	**HIGH**
Diameter (cm)	<1.5	1.5–2.2	≥2.3
Age (years)	<45	45–60	>60
Smoking status	Never smoker	Current smoker (<20 cigarettes/d)	Current smoker (>20 cigarettes/d)
Smoking cessation status	Quit ≥7 years ago or quit	Quit <7 years ago	Never quit
Characteristics of nodule margins	Smooth	Scalloped	Corona radiata or spiculated

Source: From The New England Journal of Medicine. The Solitary Pulmonary Nodule. D Ost et al: 348:2535-2542. Copyright @ 2003 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

Accurate pathologic staging requires adequate segmental, hilar, and mediastinal lymph node sampling. Ideally, this includes a mediastinal lymph node dissection. Although there are no evidence-based guidelines mandating the number of lymph nodes to be removed at surgery for adequate staging, the International Association for the Study of Lung Cancer staging manual states that suitable nodal staging generally includes sampling/dissection of lymph nodes from stations 2R, 4R, 7, 10R and 11R on the right side, and from stations 5, 6, 7, 10L, and 11L on the left side. Hilar lymph nodes are typically resected and sent for pathologic review, although it is helpful to specifically dissect and label level 10 lymph nodes when possible. On the left side, level 2 and sometimes level 4 lymph nodes are generally obscured by the aorta. Although the therapeutic benefit of nodal dissection versus nodal sampling is controversial, a pooled analysis of three trials involving patients with stages I to IIIA NSCLC demonstrated a superior 4-year survival in patients undergoing resection and a complete mediastinal lymph node dissection compared with lymph node sampling. Moreover, complete mediastinal lymphadenectomy added little morbidity to a pulmonary resection for lung cancer when carried out by an experienced thoracic surgeon.

Radiation Therapy in Stages I and II NSCLC There is currently no role for postoperative radiation therapy in patients following resection of stage I or II NSCLC with negative margins. However, patients with stage I and II disease who either decline or are not suitable candidates for surgery should be considered for radiation therapy with *curative* intent. Stereotactic body radiation therapy (SBRT) is a technique used to treat patients with isolated pulmonary nodules (≤5 cm) who are not candidates for or refuse surgical resection. Treatment is typically administered in three to five fractions delivered over 1–2 weeks. In uncontrolled studies, disease control rates are >90%, and 5-year survival rates of up to 60% have been reported with SBRT. By comparison, survival rates typically range from 13 to 39% in patients with stage I or II NSCLC treated with standard external-beam radiotherapy. Cryoablation is another

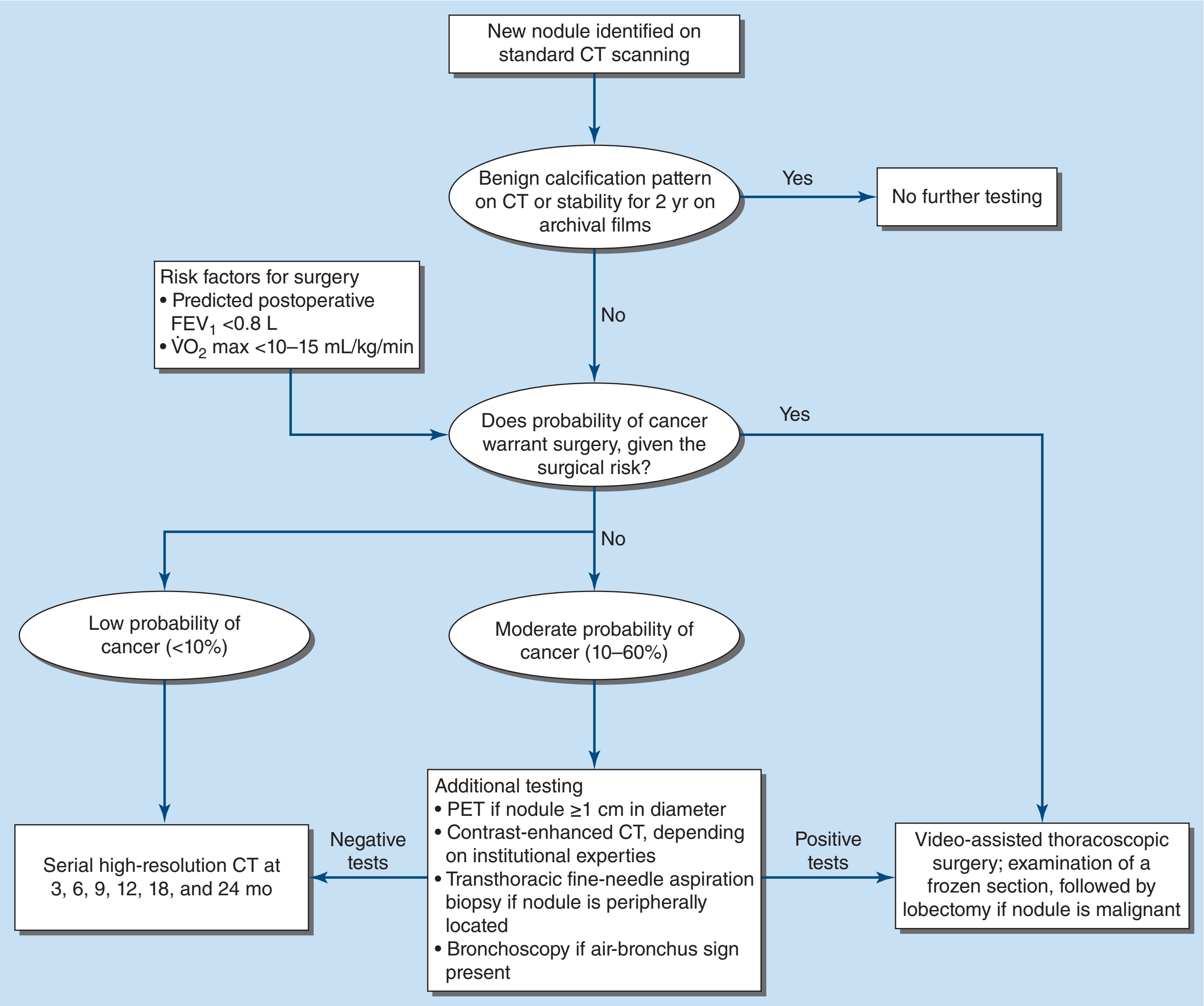
New nodule identified on standard CT scanning
Benign calcification pattern on CT or stability for 2 yr on archival films
Yes
No further testing
No
Risk factors for surgery
• Predicted postoperative FEV$_1$ <0.8 L
• $\dot{V}O_2$ max <10–15 mL/kg/min
Does probability of cancer warrant surgery, given the surgical risk?
Yes
No
Low probability of cancer (<10%)
Moderate probability of cancer (10–60%)
Additional testing
• PET if nodule ≥1 cm in diameter
• Contrast-enhanced CT, depending on institutional experties
• Transthoracic fine-needle aspiration biopsy if nodule is peripherally located
• Bronchoscopy if air-bronchus sign present
Negative tests
Positive tests
Serial high-resolution CT at 3, 6, 9, 12, 18, and 24 mo
Video-assisted thoracoscopic surgery; examination of a frozen section, followed by lobectomy if nodule is malignant

FIGURE 83-5 Approach to the solitary pulmonary nodule. FEV$_1$, forced expiratory volume in 1 s; PET, positron emission tomography.

technique occasionally used to treat small, isolated tumors (i.e., ≤3 cm). However, very little data exist on long-term outcomes with this technique.

Systemic Therapy in Stages I and II NSCLC For nearly two decades, chemotherapy has historically served as the cornerstone of adjuvant systemic therapy for patients with early-stage resected disease. The landmark meta-analysis of cisplatin-based adjuvant chemotherapy trials in patients with resected stages I to IIIA NSCLC (the Lung Adjuvant Cisplatin Evaluation [LACE] Study) published in 2008 demonstrated a 5.4% improvement in 5-year survival for cisplatin-based chemotherapy compared to surgery alone, thus becoming the standard-of-care adjuvant therapy recommendation. The survival benefit was seemingly confined to patients with stage II or III disease (Table 83-9). By contrast, survival was actually worsened in stage IA patients with the application of adjuvant therapy. In stage IB, there was a modest improvement in survival of questionable clinical significance, particularly for patients with a resected lesion ≥4 cm. Adjuvant chemotherapy was also detrimental in patients with poor performance status (Eastern Cooperative Oncology Group [ECOG] performance status = 2). These data suggest that adjuvant chemotherapy is best applied in patients with resected stage II or III NSCLC, with potential benefit for a select group of stage IB patients.

TABLE 83-8 Select References to Professional Guidelines, in Alphabetical Order

1	American Society of Clinical Oncology (ASCO), Thoracic Cancer
2.	European Society for Medical Oncology (ESMO) Clinical Practice Guidelines: Lung and Chest Tumors
3.	International Association for the Study of Lung Cancer (IASLC), Guidelines
4.	National Comprehensive Cancer Network (NCCN), Non-Small-Cell Lung Cancer Guidelines

As with any treatment recommendation, the risks and benefits of adjuvant chemotherapy should be considered on an individual patient basis. If a decision is made to proceed with adjuvant chemotherapy, in general, treatment should be initiated 6–12 weeks after surgery, assuming the patient has fully recovered, and should be administered for no more than four cycles. Although cisplatin-based chemotherapy is the preferred treatment regimen, carboplatin can be substituted for cisplatin in patients who are unlikely to tolerate cisplatin for reasons such as reduced renal function, presence of neuropathy, or hearing impairment. A large cooperative group trial compared cisplatin-based chemotherapy with vinorelbine, pemetrexed, gemcitabine, or docetaxel with or without antiangiogenic therapy. The study found similar efficacy across all treatments. Therefore, no specific chemotherapy regimen is considered superior in this setting, and treatment selection may be based on cost and patient comorbidities.

From 2020 onward, systemic therapies, initially employed in the metastatic setting—such as targeted therapy and

TABLE 83-9 Adjuvant Chemotherapy Trials in Non-Small-Cell Lung Cancer

TRIAL	STAGE	TREATMENT	NO. OF PATIENTS	5-YEAR SURVIVAL (%)	*P*
IALT	I–III	Cisplatin-based	932	44.5	< .03
		Control	835	40.4	
BR10	IB–II	Cisplatin + vinorelbine	242	69	.03
		Control	240	54	
ANITA	IB–IIIA	Cisplatin + vinorelbine	407	60	.017
		Control	433	58	
ALPI	I–III	MVP	548	50	.49
		Control	540	45	
BLT	I–III	Cisplatin-based	192	60	.90
		Control	189	58	
CALGB	IB	Carboplatin + paclitaxel	173	59	.10
			171	57	
ECOG1505	IB > 4c – IIIA	Cisplatin-based	749	NR	.90
		Cisplatin-based + bevacizumab	752	NR	

Abbreviations: ALPI, Adjuvant Lung Cancer Project Italy; ANITA, Adjuvant Navelbine International Trialist Association; BLT, Big Lung Trial; CALGB, Cancer and Lung Cancer Group B; ECOG, Eastern Cooperative Oncology Group; IALT, International Adjuvant Lung Cancer Trial; MVP, mitomycin, vindesine, and cisplatin; NR, not reported.

immunotherapy—have been incorporated into the curative approach for patients with NSCLC (**Table 83-10**).

Osimertinib, an EGFR tyrosine kinase inhibitor (TKI), demonstrated improved disease-free survival in the adjuvant setting for patients with *EGFR* mutation (exon 19 deletion or L858R)–positive NSCLC treated for 3 years with or without chemotherapy; improved overall survival benefit with adjuvant osimertinib was also reported. Regulatory authorities including the U.S. FDA, the European Commission, and China's National Medical Products Administration have granted approval for adjuvant osimertinib in resected, stages IB–IIIA *EGFR*-mutant NSCLC. Furthermore, the exploration of additional targeted therapies in the adjuvant setting is ongoing, as seen for example in the phase 3 ALINA trial investigating oral alectinib for patients with *ALK* fusion oncogene–positive NSCLC. This highlights a dynamic landscape in the pursuit of enhanced treatment outcomes for patients with early-stage NSCLC.

TABLE 83-10 Select Therapeutic Strategies Incorporating Immunotherapy or Targeted Therapy In Early-Stage, Resectable Non-Small-Cell Lung Cancer (NSCLC)

TRIAL	STAGE AND DISEASE CHARACTERISTICS	REGIMEN	APPROVAL ENDPOINT
Neoadjuvant			
CheckMate 816	IB–IIIA Irrespective of PD-L1	Nivolumab + chemotherapy × 3 cycles	EFS HR 0.63, $p = .005$
Perioperative			
KEYNOTE-671	II–IIIB (N2) Irrespective of PD-L1	Pembrolizumab + chemotherapy × 4 cycles → surgery → pembrolizumab × ~9 months	EFS HR 0.58, $p < .00001$ OS HR 0.72, $p = .00517$
Adjuvant			
IMPower010	II–IIIA PD-L1 positive (≥1%)	Adjuvant chemotherapy → atezolizumab × 1 year	DFS HR 0.66, $p = .004$
KEYNOTE 091/ PEARLS	IB–IIIA Irrespective of PD-L1	Adjuvant chemotherapy → pembrolizumab × 1 year	DFS HR 0.73
ADAURA	IB–IIIA EGFR exon 19 deletions or exon 21 L858R	Osimertinib × 3 years (irrespective of adjuvant chemotherapy)	DFS HR 0.20, $p < .0001$ OS HR 0.49, $p < .001$
ALINA	IB-IIIA ALK fusion positive	Alectinib × 2 years	DFS HR 0.24; $P<0.001$, 95% CI, 0.13 to 0.43

Abbreviations: DFS, disease-free survival; EFS, event-free survival; EGFR, epidermal growth factor receptor; HR, hazard ratio; OS, overall survival; PD-L1, programmed death ligand 1.

Approved adjuvant immunotherapy regimens, using either atezolizumab or pembrolizumab, involve approximately 1 year of treatment following adjuvant chemotherapy. This treatment approach is supported by the disease-free survival benefit observed in the IMpower010 and KEYNOTE-091 trials. Notably, the association between treatment benefit and the level of PD-L1 expression varied.

Neoadjuvant chemotherapy, which is the application of chemotherapy administered *before* an attempted surgical resection, has been advocated by some experts on the assumption that such an approach will more effectively extinguish occult micrometastases compared to postoperative chemotherapy. In addition, there is thought that preoperative chemotherapy might render an inoperable lesion resectable. A meta-analysis of 15 randomized controlled trials involving >2300 patients with stage I–III NSCLC suggested there may be a modest 5-year survival benefit (i.e., ~5%) that is virtually identical to the survival benefit achieved with postoperative chemotherapy. Accordingly, neoadjuvant therapy may prove useful in selected cases (see below). A decision to use neoadjuvant chemotherapy should always be made in consultation with an experienced surgeon. Neoadjuvant immunotherapy combined with chemotherapy has gained approval for patients with stage IB–IIIA, resectable NSCLC, with neoadjuvant nivolumab plus chemotherapy for three cycles resulting in a pathologic complete response rate of 24% and statistically significant event-free survival benefit over chemotherapy alone (hazard ratio [HR] 0.63; $p = .005$) in the CheckMate 816 trial.

Various perioperative (i.e., before and after surgical resection) treatment strategies have been explored for patients with resectable NSCLC. This involves a neoadjuvant chemoimmunotherapy approach followed by adjuvant immunotherapy. A perioperative treatment approach with pembrolizumab in combination with chemotherapy has been approved for patients with stages II–IIIB, resectable NSCLC. This approval is based on the results of the KEYNOTE-671 phase 3 trial that demonstrated a significant overall survival and event-free survival benefit over preoperative chemotherapy alone.

All patients with resected NSCLC are at high risk of developing a second primary lung cancer or recurrence, most of which occur within 18–24 months of surgery. Thus, these patients should be followed with regular physical examinations and periodic imaging studies. Given the results of the NLST, periodic CT scans appear to be the most appropriate screening modality. Professional guidelines, such as those from the National Comprehensive Cancer Network (NCCN), recommend a chest CT scan every 3–6 months for the first 3 years after surgery, then every 6 months for 2 years, and then annually to monitor for recurrence.

MANAGEMENT OF STAGE III NSCLC

Management of patients with stage III NSCLC usually requires a combined-modality approach. Patients with stage IIIA disease commonly are stratified into those with "nonbulky" or "bulky" mediastinal lymph node (N2) disease. Although the definition of "bulky" N2 disease varies somewhat in the literature, the usual criteria include the size of a dominant lymph node (i.e., >2–3 cm in short-axis diameter as measured by CT), groupings of multiple smaller lymph nodes, evidence of extracapsular nodal involvement, or involvement of more than two lymph node stations. The distinction between nonbulky and bulky stage IIIA disease is mainly used to select potential candidates for *upfront* surgical resection or for resection after neoadjuvant therapy. Many aspects of therapy of patients with stage III NSCLC remain controversial, and the optimal treatment strategy has not been clearly defined. Furthermore, because stage III disease is highly heterogeneous, no single treatment approach can be recommended for all patients. Key factors guiding treatment choices include the particular combination of tumor (T) and nodal (N) disease, the ability to achieve a complete surgical resection if indicated, and the patient's overall physical condition and preferences. For example, in carefully selected patients with limited stage IIIA disease where involved mediastinal lymph nodes can be completed resected, initial surgery followed by postoperative chemotherapy (with or without radiation therapy) may be indicated. By contrast, for patients with clinically evident bulky mediastinal lymph node involvement, the standard approach to treatment is concurrent chemoradiotherapy followed by a year of consolidation immunotherapy with durvalumab.

Absent and Nonbulky Mediastinal (N2, N3) Lymph Node Disease For the subset of stage IIIA patients initially thought to have clinical stage I or II disease (i.e., pathologic involvement of mediastinal [N2] lymph nodes is *not* detected preoperatively), surgical resection is often the treatment of choice. This is followed by adjuvant chemotherapy in patients with microscopic lymph node involvement in a resection specimen. Postoperative radiation therapy (PORT) may also have a select role for those with close or positive surgical margins. Patients with tumors exceeding 7 cm in size or involving the chest wall or proximal airways within 2 cm of the carina with hilar lymph node involvement (but not N2 disease) are classified as having T3N1 stage IIIA disease. They too are best managed with surgical resection, if technically feasible, followed by adjuvant chemotherapy if completely resected. Patients with T3N0 or T3N1 disease due to the presence of satellite nodules within the same lobe as the primary tumor are also candidates for surgery, as are patients with ipsilateral nodules in another lobe and negative mediastinal nodes (IIIA, T4N0 or T4N1). Although data regarding adjuvant chemotherapy in the latter subsets of patients are limited, it is often recommended.

Patients with T4N0-1 may have involvement of the carina, superior vena cava, or a vertebral body and yet still be candidates for surgical resection in selected circumstances. The decision to proceed with an attempted resection must be made in consultation with an experienced thoracic surgeon often in association with a vascular or cardiac surgeon and an orthopedic surgeon depending on tumor location. However, if an incomplete resection is inevitable or if there is evidence of N2 involvement (stage IIIB), surgery for T4 disease is contraindicated. Most T4 lesions are best treated with concurrent chemoradiotherapy followed by durvalumab.

The role of PORT in patients with completely resected stage III NSCLC is controversial. To a large extent, the use of PORT has historically been dictated by the presence or absence of N2 involvement and, to a lesser degree, by the biases of the treating physician. Using the SEER database, a meta-analysis of PORT identified a significant increase in survival in patients with N2 disease but not in patients with N0 or N1 disease. An earlier analysis by the PORT Meta-analysis Trialist Group using an older database produced similar results. However, two large, randomized phase 3 trials (PORT-C and Lung ART) have shown that PORT should not be routinely recommended as standard of care for patients given no improvement in disease-free survival and increased risk for cardiopulmonary toxicity.

Known Mediastinal (N2, N3) Lymph Node Disease When pathologic involvement of mediastinal lymph nodes is documented preoperatively, a combined-modality approach is recommended assuming the patient is a candidate for treatment with curative intent. These patients are at high risk for both local and distant recurrence if managed with resection alone. For patients with stage III disease who are not candidates for surgical resection, *concurrent* chemoradiotherapy is most commonly used as the initial treatment followed by durvalumab. Concurrent chemoradiotherapy has been shown to produce superior survival compared to *sequential* chemoradiotherapy; however, it also is associated with greater host toxicities (including fatigue, esophagitis, and neutropenia). Therefore, for patients with a good performance status, concurrent chemoradiotherapy is the preferred treatment approach, whereas sequential chemoradiotherapy may be more appropriate for patients with a performance status that is not as good. For patients who are *not* candidates for a combined-modality treatment approach, typically due to a poor performance status or a comorbidity that makes chemotherapy untenable, radiotherapy alone may provide a modest survival benefit in addition to symptom palliation.

For patients with potentially resectable N2 disease, it remains uncertain whether surgery after neoadjuvant chemoradiotherapy improves survival. In an NCI-sponsored Intergroup randomized trial comparing concurrent chemoradiotherapy alone to concurrent chemoradiotherapy followed by attempted surgical resection, no survival benefit was observed in the trimodality arm compared to the bimodality therapy. In fact, patients subjected to a pneumonectomy had a worse survival outcome. By contrast, those treated with a lobectomy appeared to have a survival advantage based on a retrospective subset analysis. Thus, in carefully selected, otherwise healthy patients with nonbulky mediastinal lymph node involvement, surgery may be a reasonable option if the primary tumor can be fully resected with a lobectomy. This is not the case if a pneumonectomy is required to achieve complete resection.

Advancements in neoadjuvant therapy, particularly in the realm of combination chemoimmunotherapy, have led to interest in exploring the benefits of such therapy for patients with N2 disease. The NADIM II phase 2 trial is an illustrative example that exclusively enrolled patients with stage IIIA–IIIB disease per American Joint Committee on Cancer eighth edition criteria; notably, two-thirds of these patients had N2 disease, including cases of multistation N2 disease. The trial demonstrated a pathologic complete response rate (pCR) of 36–42% with chemoimmunotherapy for patients with N2 disease, compared to 0–10% pCR rate for those who received neoadjuvant chemotherapy alone. Additionally, the study demonstrated the safety of a neoadjuvant chemoimmunotherapy approach in patients with N2 disease, without compromising the feasibility of surgical outcomes. Further investigation is warranted to better define the role of novel neoadjuvant therapeutic strategies for patients with N2 disease.

Superior Sulcus Tumors (Pancoast Tumors) Superior sulcus tumors represent a distinctive subset of stage III disease. These tumors arise in the apex of the lung and may invade the second and third ribs, the brachial plexus, the subclavian vessels, the stellate ganglion, and adjacent vertebral bodies. They also may be associated with Pancoast syndrome, characterized by pain that may arise in the shoulder or chest wall or radiate to the neck. Pain characteristically radiates to the ulnar surface of the hand. Horner's syndrome (enophthalmos, ptosis, miosis, and anhidrosis) due to invasion of the paravertebral sympathetic chain may be present as well. Patients with these tumors should undergo the same staging procedures as all patients with stage II and III NSCLC. Neoadjuvant chemotherapy or combined chemoradiotherapy followed by surgery is reserved for those without N2 involvement. This approach yields excellent survival outcomes (>50% 5-year survival in patients with an R0 resection). Patients with N2 disease are less likely to benefit from surgery and can be managed with chemoradiotherapy followed by durvalumab. Patients presenting with metastatic disease can be treated with radiation therapy (with or without chemotherapy) for symptom palliation.

SURGICAL MANAGEMENT OF NSCLC

Traditionally, in patients with NSCLC, the following have been relative contraindications to potential curative resection: extrathoracic metastases, superior vena cava syndrome, vocal cord and, in most cases, phrenic nerve paralysis, malignant pleural effusion, cardiac tamponade, tumor within 2 cm of the carina (potentially curable with combined chemoradiotherapy), metastasis to the contralateral lung, metastases to supraclavicular lymph nodes, contralateral mediastinal node metastases (potentially curable with combined chemoradiotherapy), and involvement of the main pulmonary artery. In situations where it will make a difference in treatment, abnormal scan findings should be further investigated and require tissue confirmation of malignancy so that patients are not precluded from having potentially curative therapy. The role of surgical management of NSCLC continues to evolve, given emerging neoadjuvant/perioperative treatment strategies, and warrants multidisciplinary discussion.

MANAGEMENT OF METASTATIC NSCLC

Approximately 40% of NSCLC patients present with advanced, stage IV disease at the time of diagnosis. In addition, a significant number of patients who first presented with early-stage NSCLC will eventually relapse with distant disease. Patients who have recurrent disease have a better prognosis than those presenting with metastatic disease at the time of diagnosis. Standard medical management, the appropriate use of pain medications, and the pertinent use of radiotherapy and systemic therapy—which may consist of targeted therapy, immunotherapy, and/or traditional cytotoxic chemotherapy depending on the specific diagnosis as well as PD-L1 TPS and molecular subtype—form the cornerstone of management. Systemic therapy palliates symptoms, improves quality of life, and improves survival in patients with metastatic NSCLC, particularly in patients with good performance status. Of note, the early application of palliative care in conjunction with chemotherapy in patients with advanced NSCLC is associated with both improved survival and quality of life.

Targeted Therapies for Select Molecular Cohorts of NSCLC For a cohort of NSCLC patients, the presence of an oncogenic driver mutation allows the use of oral therapies with significant antitumor activity and improved survival compared to cytotoxic chemotherapy. These driver mutations occur in genes encoding signaling proteins that, when aberrant, promote the uncontrolled growth and metastasis of tumor cells. Importantly, driver mutations can serve as Achilles' heels for tumors, if their gene products can be targeted therapeutically with small-molecule inhibitors. All patients with advanced NSCLC should ideally undergo comprehensive molecular testing (i.e., multigene testing encompassing at least all biomarkers for which there are approved biomarker-matched therapies, including less common driver mutations) with broad panel-based testing techniques such as NGS as opposed to conventional, single-gene targeted testing strategies. Point mutations, insertions/deletions, chromosomal rearrangements (sometimes called "fusions"), and copy number variants have been reported in a number of genes including *ALK*, *BRAF*, *EGFR*, *ERRB2*, *KRAS*, *MET*, *NRAS*, *NRG1*, *NTRK*, *PIK3ca*, *RET*, and *ROS1*, with varying levels of clinical evidence. As our treatment armamentarium expands, knowledge of these mutations is critical for selection of appropriate therapy.

EGFR mutations have been detected in 10–15% of North American patients diagnosed with NSCLC. *EGFR* mutations are typically (but not exclusively) associated with younger age, light (<10 pack-year) and nonsmokers, and adenocarcinoma histology. Approximately 90% of these mutations are exon 19 deletions or exon 21 L858R point mutations within the EGFR tyrosine kinase domain, resulting in hyperactivation of both EGFR kinase activity and downstream signaling. Lung tumors that harbor activating mutations within the EGFR kinase domain display high sensitivity to small-molecule EGFR TKIs. Osimertinib, erlotinib, gefitinib, afatinib, and dacomitinib are FDA-approved oral small-molecule TKIs that inhibit EGFR. Several large, international, phase 3 studies have demonstrated improved response rates and progression-free survival in patients with *EGFR* mutation–positive NSCLC treated with an EGFR TKI as compared with standard first-line chemotherapy regimens **(Table 83-11)**. Osimertinib was shown in a randomized phase 3 trial to have superior progression-free and overall survival in patients with *EGFR*-mutant NSCLC compared to earlier-generation EGFR TKIs (erlotinib or gefitinib) and to chemotherapy. Emerging combination strategies in the first-line setting, such as combining osimertinib with platinum-based chemotherapy, are also gaining traction. *EGFR* exon 20 insertion mutations are the third most common *EGFR* mutation. The anti-EGFR therapies described above are not approved for use in the context of tumors that harbor *EGFR* exon 20 insertion mutations. The current standard of care for *EGFR* exon 20–mutated lung cancer is a combination of platinum-based chemotherapy with amivantamab, which is an EGFR-MET bispecific antibody.

TABLE 83-11 Phase 3 Trials of EGFR TKIs in EGFR-Positive Non-Small-Cell Lung Cancer

TRIAL	THERAPY	NO. OF PATIENTS	ORR (%)	PFS (MONTHS)
IPASS	CbP	129	47	6.3
	Gefitinib	132	71	9.3
EURTAC	CG	87	15	5.2
	Erlotinib	86	58	9.7
OPTIMAL	CG	72	36	4.6
	Erlotinib	82	83	13.1
NEJ002	CG	114	31	5.4
	Gefitinib	114	74	10.8
WJTOG3405	CD	89	31	6.3
	Gefitinib	88	62	9.2
LUX LUNG 3	CP	115	23	6.9
	Afatinib	230	56	11.1
LUX LUNG 6	CG	122	23	5.6
	Afatinib	242	67	11.0
LUX LUNG 7	Erlotinib	159	58	10.9
	Afatinib	160	73	11.0
ARCHER 1050	Gefitinib	225	72	9.2
	Dacomitinib	227	75	14.7
FLAURA	Erlotinib or gefitinib	277	76	8.5
	Osimertinib	279	80	17.2
FLAURA2	Osimertinib/chemotherapy	279	83	25.5
	Osimertinib	278	76	16.7
MARIPOSA	Amivantamab/lazertinib	429	86	23.7
	Osimertinib	429	85	16.6

Abbreviations: CbP, carboplatin and paclitaxel; CD, cisplatin and docetaxel; CG, cisplatin and gemcitabine; CP, cisplatin and paclitaxel; EGFR, epidermal growth factor receptor; ORR, overall response rate; PFS, progression-free survival; TKI, tyrosine kinase inhibitor.

HER2/ERBB2 mutations have also been detected in ~3% of nonsquamous NSCLC. HER2, like EGFR, is a member of the ERBB family of receptor tyrosine kinases. The most common *HER2* mutations in lung cancer encompass insertion variants within exon 20, which is part of the tyrosine kinase domain (so-called "exon 20 insertions"). The anti-HER2 monoclonal antibody-drug conjugate, trastuzumab deruxtecan, is approved for patients with previously treated, metastatic NSCLC harboring a mutation in HER2/ERRB2, with an objective response rate of ~50% and median progression-free survival of ~8 months.

Chromosomal Rearrangements in NSCLC Chromosomal rearrangements are found in ~10% of patients with NSCLC. Typically, although not exclusively, chromosomal rearrangements are associated with younger age, no smoking history, and adenocarcinoma histology. However, given the potential therapeutic impact, it is strongly recommended that all patients are tested. Given that the resultant fusion proteins include the entire tyrosine kinase domain, the most effective therapeutic strategy at this time is with TKIs.

TABLE 83-12 Results of Phase 3 Trials Comparing First-Line ALK Inhibitors in ALK-Positive NSCLC

TRIAL	THERAPY	NO. OF PATIENTS	ORR (%)	MEDIAN PFS
Profile 1014	Crizotinib	172	74	10.9
	Platinum-chemotherapy	171	45	7.0
ALEX	Alectinib	152	82.9	34.8
	Crizotinib	151	75.5	10.9
ALTA1L	Brigatinib	137	71	67% at 24 months
	Crizotinib	138	60	43% at 11 months
eXalt3	Ensartinib	121	75	25.8
	Crizotinib	126	67	12.7
CROWN	Lorlatinib	149	76	64% at 60 months
	Crizotinib	147	58	9.1

Abbreviations: NSCLC, non-small-cell lung cancer; ORR, overall response rate; PFS, progression-free survival.

In NSCLC, chromosomal rearrangements involving the anaplastic lymphoma kinase (*ALK*) gene on chromosome 2 have been found in ~3–7% of patients with NSCLC. *ALK* rearrangements lead to hyperactivation of the ALK tyrosine kinase domain. Crizotinib is a first-generation ALK TKI, whereas alectinib, brigatinib, and ceritinib are second-generation ALK TKIs approved as first-line treatment options for patients with lung tumors harboring *ALK* rearrangements. Both alectinib and brigatinib have been found to have superior progression-free survival to crizotinib. Lorlatinib, a third-generation ALK TKI, is also approved for patients in the first-line setting and for patients who progress on a second-generation ALK TKIs **(Table 83-12)**. Unique adverse events of lorlatinib may include hyperlipidemia, weight gain, and cognitive effects. *ALK* testing may be performed via fluorescence in situ hybridization (FISH), immunohistochemistry (IHC), or NGS.

ROS1 fusions, detected by FISH or NGS, have been identified in ~1–2% of patients with NSCLC. Crizotinib, which inhibits both ROS1 and MET kinases, and the ROS1/TRK inhibitors entrectinib and repotrectinib have been FDA approved for patients whose tumors harbor a *ROS1* fusion. Entrectinib and repotrectinib offer improved blood-brain barrier penetration.

RET alterations typically occur as chromosomal rearrangements resulting in constitutive TKI activation. *RET* rearrangements may be detected by either FISH or NGS in ~1% of NSCLC patients. Analogous to capmatinib, pralsetinib and selpercatinib have demonstrated an excellent response rate; as many as 78–85% of treatment-naïve NSCLC patients with *RET* alterations responded.

NTRK fusions occur in members of the *NTRK* gene family (*NTRK1*, *NTRK2*, *NTRK3*) and result in constitutive protein kinase activation. *NTRK* fusions are rare, occurring in <1% of patients with NSCLC. Entrectinib and larotrectinib have demonstrated durable antitumor efficacy and are currently approved for *NTRK*-positive NSCLC.

Targeting the Mitogen-Activated Protein Kinase (MAPK) Pathway in NSCLC The MAPK pathway may be dysregulated in a subset of patients with NSCLC. For example, mutations within the *KRAS* GTPase are found in ~25% of lung adenocarcinomas, with mutations in *KRAS* G12C occurring in ~14% of adenocarcinomas. Agents targeting *KRAS* G12C such as adagrasib and sotorasib are now approved, with objective response rates of ~42 and ~37%, respectively. Defining mechanisms of acquired resistance to small-molecule inhibitors is a high research priority.

Oncogenic mutations in *BRAF* have been observed in ~2% of patients with NSCLC and, similar to *KRAS*, may serve as a key oncogene in the MAP kinase pathway. *BRAF* mutations may occur in both squamous and nonsquamous NSCLC and with an equal prevalence in patients with a history of smoking. This mutation is typically most targetable when it occurs at the 600th amino acid valine (V600). Combined inhibition with a BRAF and MEK inhibitor, dabrafenib plus trametinib or encorafenib with binimetinib, is a first-line or later therapeutic option in patients with *BRAF* V600–mutant NSCLC and appears to be superior to BRAF or MEK inhibition alone.

MET exon 14 skipping mutations have also been identified in ~3–5% of patients with NSCLC. Notably, *MET* exon 14 skipping mutations may occur in both squamous and nonsquamous NSCLC patients and those with a history of smoking. Pharmacologic inhibition of the overactive MET pathway with FDA-approved capmatinib or tepotinib resulted in response rates >70%, particularly in treatment-naïve NSCLC patients. A unique and challenging potential adverse event of oral MET inhibition with capmatinib and tepotinib is peripheral edema. All NCCN-supported targetable oncogenic driver mutations and potential therapeutic options are summarized in **Fig. 83-6**.

Immunotherapy Immune checkpoint inhibitors have significantly improved survival for a group of patients with locally advanced and metastatic NSCLC. These agents are used primarily in patients whose tumors do not express a targetable genetic lesion **(Fig. 83-7)**. Immune checkpoint inhibitors work by blocking interactions between T cells and antigen-presenting cells (APCs) or tumor cells that lead to T-cell inactivation. By inhibiting this interaction, the immune system is effectively upregulated and T cells become

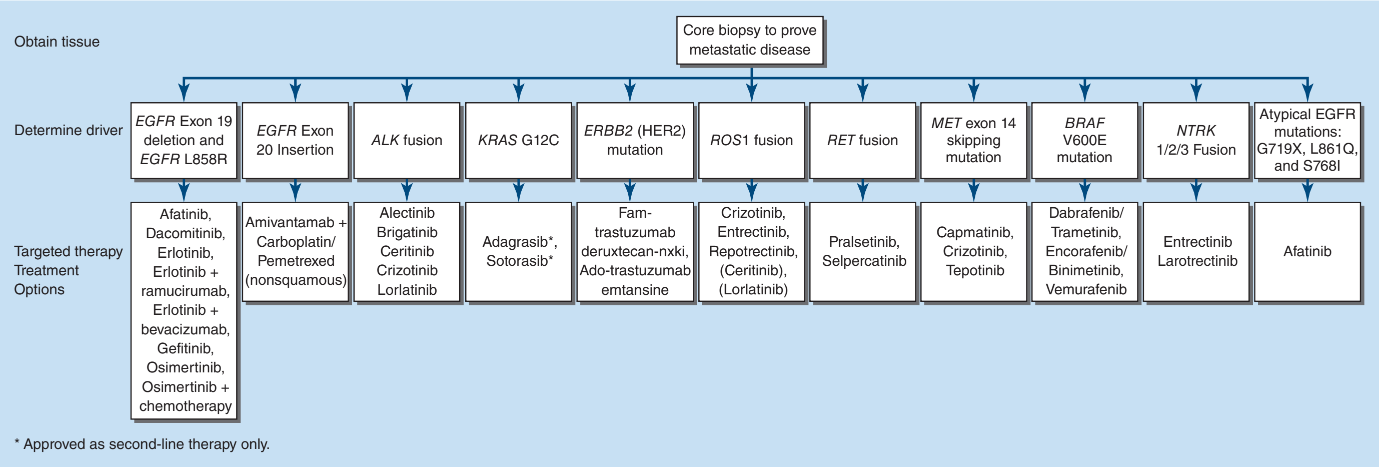
Obtain tissue
Core biopsy to prove metastatic disease
Determine driver
EGFR Exon 19 deletion and *EGFR* L858R
EGFR Exon 20 Insertion
ALK fusion
KRAS G12C
ERBB2 (HER2) mutation
*ROS*1 fusion
RET fusion
MET exon 14 skipping mutation
BRAF V600E mutation
NTRK 1/2/3 Fusion
Atypical EGFR mutations: G719X, L861Q, and S768I
Targeted therapy Treatment Options
Afatinib, Dacomitinib, Erlotinib, Erlotinib + ramucirumab, Erlotinib + bevacizumab, Gefitinib, Osimertinib, Osimertinib + chemotherapy
Amivantamab + Carboplatin/ Pemetrexed (nonsquamous)
Alectinib Brigatinib Ceritinib Crizotinib Lorlatinib
Adagrasib*, Sotorasib*
Fam-trastuzumab deruxtecan-nxki, Ado-trastuzumab emtansine
Crizotinib, Entrectinib, Repotrectinib, (Ceritinib), (Lorlatinib)
Pralsetinib, Selpercatinib
Capmatinib, Crizotinib, Tepotinib
Dabrafenib/ Trametinib, Encorafenib/ Binimetinib, Vemurafenib
Entrectinib Larotrectinib
Afatinib
* Approved as second-line therapy only.

FIGURE 83-6 United States-centric approach to targeted therapy in non-small-cell lung cancer (NSCLC) based on drug approvals. Acknowledging regional variances in approvals. Drugs are listed alphabetically, not by author preference. Current as of June 2024.

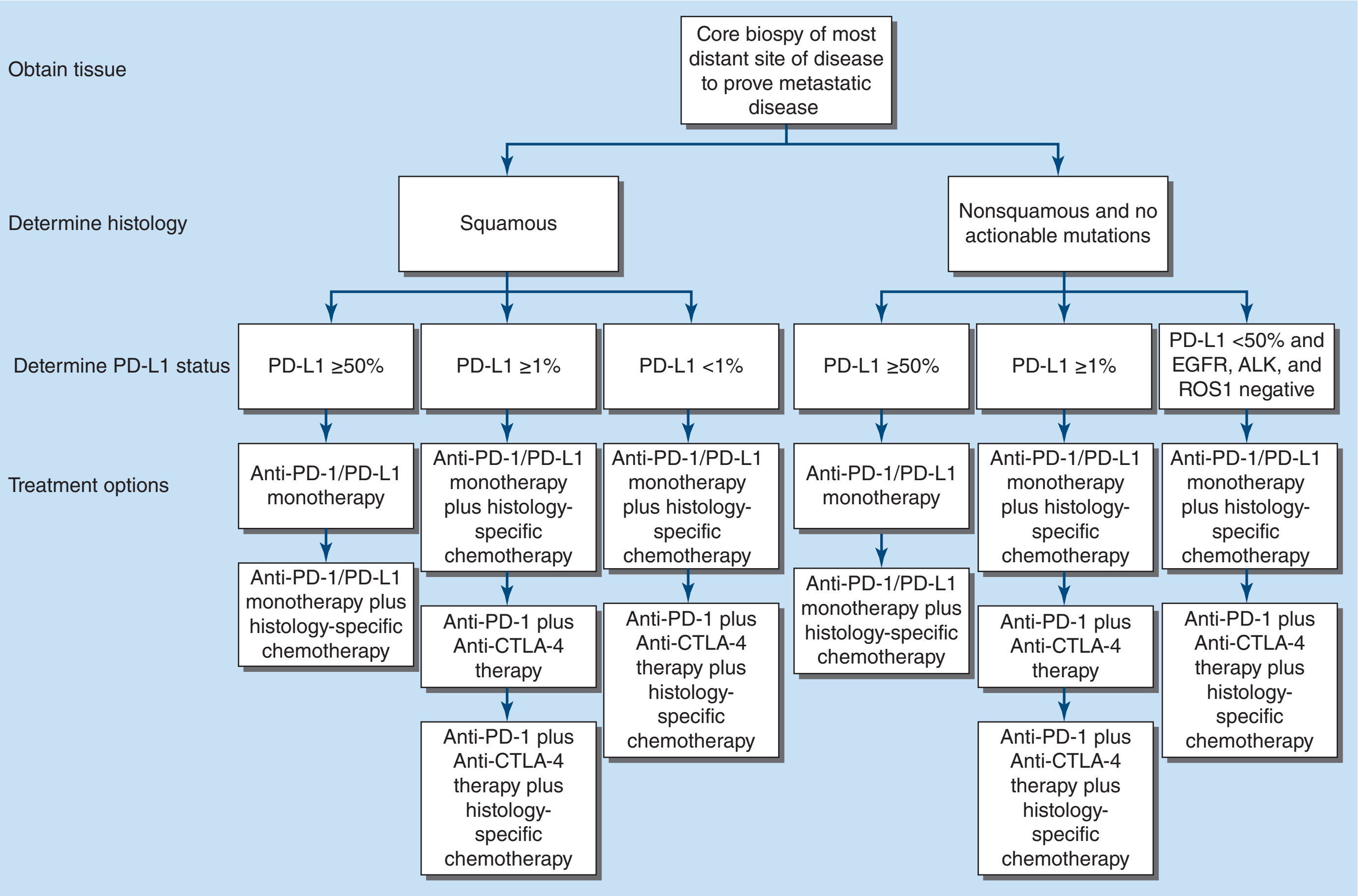
Obtain tissue
Core biospy of most distant site of disease to prove metastatic disease
Determine histology
Squamous
Nonsquamous and no actionable mutations
Determine PD-L1 status
PD-L1 ≥50%
PD-L1 ≥1%
PD-L1 <1%
PD-L1 ≥50%
PD-L1 ≥1%
PD-L1 <50% and EGFR, ALK, and ROS1 negative
Treatment options
Anti-PD-1/PD-L1 monotherapy
Anti-PD-1/PD-L1 monotherapy plus histology-specific chemotherapy
Anti-PD-1/PD-L1 monotherapy plus histology-specific chemotherapy
Anti-PD-1/PD-L1 monotherapy
Anti-PD-1/PD-L1 monotherapy plus histology-specific chemotherapy
Anti-PD-1/PD-L1 monotherapy plus histology-specific chemotherapy
Anti-PD-1/PD-L1 monotherapy plus histology-specific chemotherapy
Anti-PD-1 plus Anti-CTLA-4 therapy
Anti-PD-1 plus Anti-CTLA-4 therapy plus histology-specific chemotherapy
Anti-PD-1/PD-L1 monotherapy plus histology-specific chemotherapy
Anti-PD-1 plus Anti-CTLA-4 therapy
Anti-PD-1 plus Anti-CTLA-4 therapy plus histology-specific chemotherapy
Anti-PD-1 plus Anti-CTLA-4 therapy plus histology-specific chemotherapy
Anti-PD-1 plus Anti-CTLA-4 therapy plus histology-specific chemotherapy

FIGURE 83-7 **Approach to first-line therapy in a patient with stage IV, driver mutation–negative non-small-cell lung cancer (NSCLC).**

activated against tumor cells. Several randomized studies have demonstrated superior overall survival in patients treated with pembrolizumab or atezolizumab monotherapy or nivolumab plus ipilimumab combination immunotherapy compared to chemotherapy in patients with metastatic NSCLC with PD-L1 expression in ≥50% of tumor cells (KEYNOTE-024, IMPOWER 110) and ≥1% of tumor cells (KEYNOTE-042, CheckMate 227) **(Table 83-13)**. The evidence supporting the use of single-agent immunotherapy in patients with tumor PD-L1 <50% remains unclear; current recommendations suggest the use of chemotherapy plus immunotherapy or immunotherapy combinations as the first-line treatment strategy in patients with metastatic NSCLC with tumor PD-L1 <50%. As discussed below, specific regimens vary by tumor histology (adenocarcinoma vs squamous cell carcinoma). Although PD-L1 has been identified as a biomarker that can predict response to immune checkpoint inhibitors, responses are observed in patients who do not appear to express the biomarker, and not all PD-L1-positive patients respond to checkpoint inhibition. Importantly patients with driver mutations such as *EGFR* and *ALK* appear to derive greater benefit from targeted therapy than immunotherapy and should be treated with a TKI, even in the presence of high PD-L1 expression. Further evaluation of these agents in the neoadjuvant and perioperative settings and combined with chemoradiotherapy is ongoing.

Cytotoxic Chemotherapy for Metastatic or Recurrent NSCLC Cytotoxic chemotherapy is typically used in combination with immunotherapy as the initial treatment in patients with metastatic or recurrent NSCLC only when there is no contraindication to immunotherapy. Selected chemotherapy agents perform quite differently in squamous carcinomas versus adenocarcinomas. Patients with nonsquamous NSCLC have an improved survival when treated with cisplatin and pemetrexed compared to cisplatin and gemcitabine. By contrast, patients with squamous carcinoma have an improved survival when treated with cisplatin and gemcitabine. This survival difference is thought to be related to the differential expression between tumor types of thymidylate synthase (TS). Squamous cancers have a much higher expression of TS compared to adenocarcinomas, accounting for their lower responsiveness to pemetrexed. By contrast, the activity of gemcitabine is not impacted by the levels of TS.

Second-Line Therapy and Beyond Second-line therapy for advanced NSCLC relies on docetaxel; it improves survival compared to supportive care alone. Ramucirumab is a recombinant human IgG1 monoclonal antibody that targets VEGFR-2 and blocks the interaction of VEGF ligands and VEGFR-2. A phase 3 trial demonstrated a significant improvement in progression-free survival and overall survival when ramucirumab was combined with docetaxel as second-line therapy in patients who had progressed on platinum-based chemotherapy. Contrary to bevacizumab, ramucirumab was safe in patients with both squamous and nonsquamous NSCLC and is approved regardless of histology. Many new agents are being developed and tested in clinical trials for second-line and beyond therapy in the metastatic setting, including novel immune checkpoint agents, antibody-drug conjugates, and even cellular therapies. Improved therapies in this setting remain an area of great need for patients with lung cancer.

Supportive Care No discussion of the treatment strategies for patients with advanced lung cancer would be complete without a mention of supportive care. Coincident with advances in chemotherapy and targeted therapy was a pivotal study that demonstrated that the early integration of palliative care with standard treatment

TABLE 83-13 Results of Phase 3 Trials Comparing First-Line Immunotherapy with or without Chemotherapy Versus Chemotherapy Alone in Patients with NSCLC

STUDY	THERAPY	NO. OF PATIENTS	OS (MONTHS)	PFS (MONTHS)
KEYNOTE-024	Pembrolizumab	154	30.0	7.9
PD-L1 ≥50%	Platinum-chemotherapy	151	14.2	3.5
KEYNOTE-042	Pembrolizumab	637	16.7	5.4
PD-L1 ≥1%	Platinum-chemotherapy	637	12.1	6.5
IMPOWER 110	Atezolizumab	286	20.2	8.1
PD-L1 ≥50% TC or ≥15% IC	Platinum-chemotherapy	263	13.1	5.0
KEYNOTE-189	Pembrolizumab + platinum-chemotherapy	410	NR	8.8
Nonsquamous	Platinum-chemotherapy	206	11.3	4.9
KEYNOTE-407	Pembrolizumab + platinum-chemotherapy	278	15.9	6.4
Squamous	Platinum-chemotherapy	281	11.3	4.8
IMPOWER 150	Atezolizumab + platinum-chemotherapy	356	19.2	8.3
Nonsquamous	Platinum-chemotherapy	336	14.7	6.8
IMPOWER 130	Atezolizumab + platinum-chemotherapy	483	18.6	7.0
Nonsquamous	Platinum-chemotherapy	240	13.9	5.5
EMPOWER-Lung 3	Cemiplimab + platinum-chemotherapy	312	21.9	8.2
	Platinum-chemotherapy	154	13.0	5.0
CheckMate 227	Nivolumab + ipilimumab	583	17.1	5.1
	Platinum-chemotherapy	583	13.9	5.6
CheckMate-9LA	Nivolumab + ipilimumab plus two cycles of platinum-chemotherapy	361	14.1	6.8
	Platinum-chemotherapy	358	10.7	5
POSEIDON	Tremelimumab + durvalumab and platinum-chemotherapy	338	14	6.2
	Platinum-chemotherapy	337	11.7	4.8

Abbreviations: IC, immune cells; NR, not reported; NSCLC, non-small-cell lung cancer; OS, overall survival; PFS, progression-free survival; TC, tumor cells.

Note: Platinum-chemotherapy refers to first-line platinum doublet or triplet chemotherapy.

strategies improves both quality of life and overall survival for patients with stage IV NSCLC **(Chaps. 13 and 74)**. Aggressive pain and symptom control are important components of optimal treatment of these patients.

TREATMENT

Small-Cell Lung Cancer

The overall treatment approach to patients with SCLC is shown in **Fig. 83-8**.

SURGERY FOR LIMITED-DISEASE SCLC

SCLC is a highly aggressive disease characterized by its rapid doubling time, high growth fraction, early development of disseminated disease, and dramatic response to first-line chemotherapy and radiation. In general, surgical resection is *not* routinely recommended for patients because even patients with LD-SCLC still have occult micrometastases. However, the American College of Chest Physicians Evidence-Based Clinical Practice Guidelines recommend surgical resection over nonsurgical treatment in SCLC patients with clinical stage I disease after a thorough evaluation for distant metastases and invasive mediastinal stage evaluation (grade 2C). After resection, these patients should receive platinum-based adjuvant chemotherapy (grade 1C). If the histologic diagnosis of SCLC is made in patients on review of a resected surgical specimen, such patients should receive standard SCLC chemotherapy as well.

CHEMOTHERAPY

In patients with limited-stage SCLC, concurrent chemoradiotherapy with cisplatin-etoposide for four cycles has remained standard of care for over 4 decades. Two randomized phase 3 trials have demonstrated that chemotherapy with either cisplatin or carboplatin plus either etoposide and a PD-L1 inhibitor, atezolizumab (IMPOWER 133) or durvalumab (CASPIAN), provides superior progression-free and overall survival compared to chemotherapy alone, making combination therapy the preferred choice in appropriate patients. Despite response rates to first-line therapy as high as 80%, the median survival ranges from 12 to 20 months for patients with LD and ~12 months for patients with ED. Regardless of disease extent, the majority of patients relapse and develop chemotherapy-resistant disease. The prognosis is especially poor for patients who relapse within the first 3 months of therapy; these patients are said to have *chemotherapy-resistant disease*. Patients are said to have *sensitive disease* if they relapse >3 months after their initial therapy and are thought to have a somewhat better overall survival. These patients also are thought to have the greatest potential benefit from second-line chemotherapy. Topotecan and lurbinectedin are FDA-approved agents for second-line therapy in patients with SCLC. Topotecan has only modest activity and can be given either intravenously or orally; it appears to have more efficacy in patients with chemotherapy-sensitive disease. Lurbinectedin has a 35% response rate and progression-free survival of 3.5 months, with a greater benefit in patients with chemotherapy-sensitive disease. Other agents with similar low levels of activity in the second-line setting include irinotecan, paclitaxel, docetaxel, vinorelbine, oral etoposide, and gemcitabine. The treatment of refractory SCLC is a pressing concern, prompting the investigation of various new targets and drugs. For example, at the time this chapter is being written, one of the most exciting advancements in SCLC is the development of delta-like ligand 3 (DLL3)-targeting agents. For example, in the phase 2 DeLLphi-301 study, tarlatamab, a bispecific T-cell engager immunotherapy targeting DLL3 and CD3, demonstrated a 40% objective response rate and a median progression-free survival of 4.9 months in the 10-mg group for patients with previously treated SCLC. Confirmatory evidence from the phase 3 study remains pending.

THORACIC RADIATION THERAPY

Thoracic radiation therapy (TRT) and concurrent chemotherapy with curative intent are standard for patients with limited-stage SCLC with suitable performance status. Without TRT, virtually

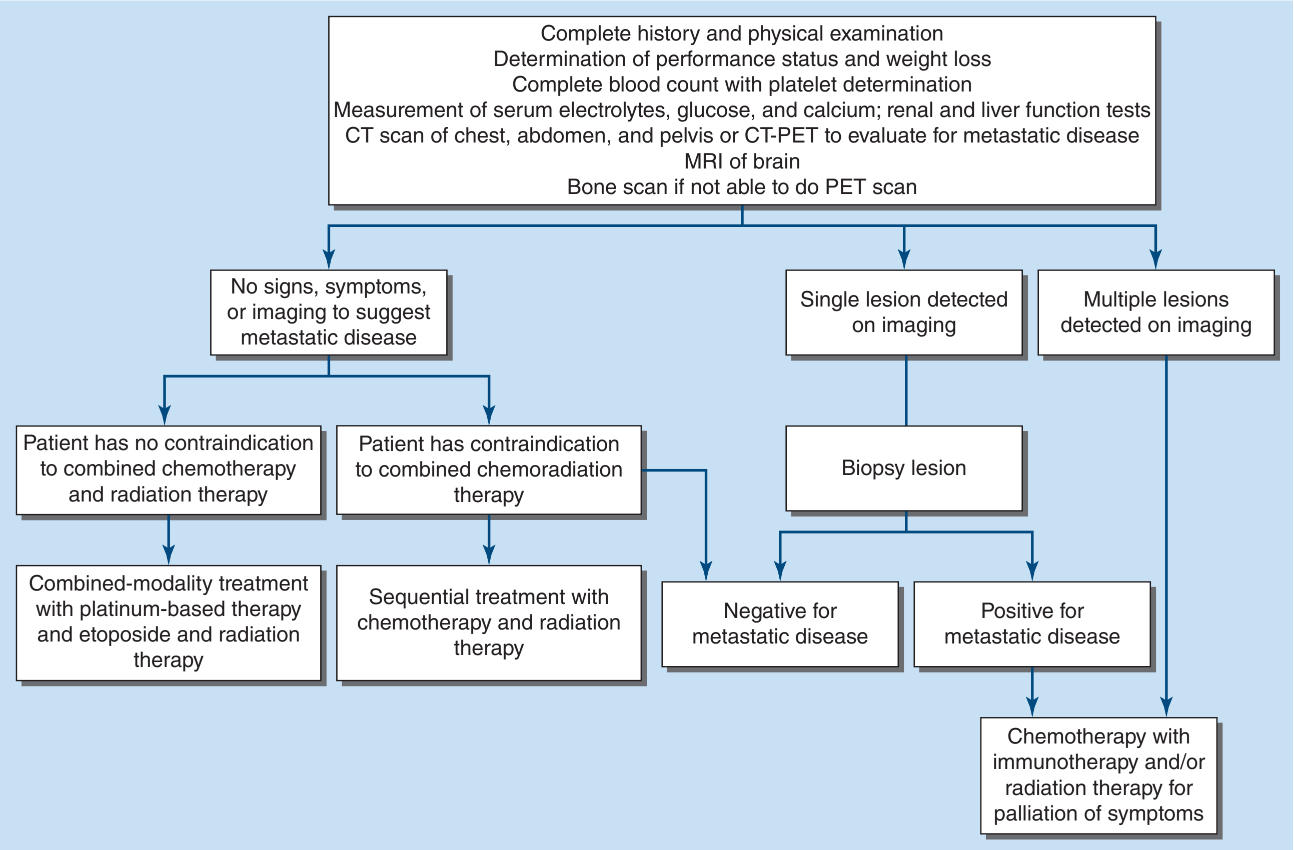

Complete history and physical examination
Determination of performance status and weight loss
Complete blood count with platelet determination
Measurement of serum electrolytes, glucose, and calcium; renal and liver function tests
CT scan of chest, abdomen, and pelvis or CT-PET to evaluate for metastatic disease
MRI of brain
Bone scan if not able to do PET scan
No signs, symptoms, or imaging to suggest metastatic disease
Single lesion detected on imaging
Multiple lesions detected on imaging
Patient has no contraindication to combined chemotherapy and radiation therapy
Patient has contraindication to combined chemoradiation therapy
Biopsy lesion
Combined-modality treatment with platinum-based therapy and etoposide and radiation therapy
Sequential treatment with chemotherapy and radiation therapy
Negative for metastatic disease
Positive for metastatic disease
Chemotherapy with immunotherapy and/or radiation therapy for palliation of symptoms

FIGURE 83-8 Algorithm for management of small-cell lung cancer. MRI, magnetic resonance imaging; PET, positron emission tomography.

all patients with limited-stage SCLC will progress on systemic therapy alone within 1 year. Meta-analyses indicate that chemotherapy combined with TRT improves 3-year survival by ~5% compared with chemotherapy alone. The 5-year survival rate, however, remains low at ~29–34%. Most commonly, TRT is combined with concurrent platinum and etoposide because of a better toxicity profile compared to anthracycline-based regimens. For limited-stage SCLC, concurrent chemoradiation is more effective than sequential chemoradiation but is associated with significantly more esophagitis and hematologic toxicity. Ideally, TRT should be administered by the second cycle of chemotherapy because later application appears slightly less effective. If for reasons of fitness or availability concurrent chemoradiation cannot be administered, TRT should follow induction chemotherapy in alignment with an extensive-stage SCLC paradigm. With respect to fractionation of TRT, twice-daily 1.5-Gy fractioned radiation therapy has been shown to improve survival in limited-stage SCLC patients but is associated with higher rates of grade 3 esophagitis and pulmonary toxicity. Patients should be carefully selected for concurrent chemoradiation therapy based on good performance status and adequate pulmonary reserve. The role of radiotherapy in ED-SCLC is mainly limited to palliation of tumor-related symptoms such as bone pain and bronchial obstruction. Notably, consolidation TRT has demonstrated a survival benefit in extensive-stage SCLC in prospective studies; however, translating these study findings into current clinical practice is challenging due to their completion before the introduction of immunotherapy in the first-line setting.

PROPHYLACTIC CRANIAL IRRADIATION

Prophylactic cranial irradiation (PCI) has historically been offered to all patients with SCLC without progression on initial treatment for limited-stage and extensive-stage disease. A meta-analysis including seven trials and 987 patients with limited-stage SCLC who had achieved a complete remission after upfront chemotherapy yielded a 5.4% improvement in overall survival for patients treated with PCI. However, the role of PCI has become controversial as these studies were conducted before brain MRIs for staging. In patients with ED-SCLC who have responded to first-line chemotherapy and had no CNS disease, patients randomized to observation had a higher incidence of brain metastases; however, use of PCI did not improve overall survival.

THYMIC TUMORS

Thymic tumors are rare malignancies, accounting for 0.5–1.5% of all malignancies in the United States with a higher incidence among Asian populations. They are particularly rare among children and young adults with incidence peaking in the fifth decade of life. There is no difference between sexes, and no clear risk factors have been identified.

CLINICAL MANIFESTATIONS

The majority of thymic tumors occur in the anterior mediastinum. Approximately 40% of patients with mediastinal masses will be asymptomatic with an incidental finding on chest imaging. In patients presenting with an anterior mediastinal mass, if appropriate, serum β-human chorionic gonadotropin (HCG) and α-fetoprotein (AFP) should be sent to rule out a germ cell tumor. A patient with a sign or symptom of thymoma or thymic carcinoma may present with chest pain, dyspnea, cough, or superior vena cava syndrome secondary to effects on adjacent organs or a paraneoplastic syndrome, most commonly myasthenia gravis, pure red cell aplasia, or hypogammaglobulinemia. More rare paraneoplastic syndromes include limbic encephalitis, aplastic anemia, hemolytic anemia, and autoimmune disease such as Sjögren's syndrome, polymyositis, rheumatoid arthritis, and ulcerative colitis, among others.

STAGING

Given the rarity of the tumor, patients with suspected thymoma should be evaluated by a multidisciplinary team including a surgeon, medical and radiation oncologist, and pathologist with experience in treating the disease. A CT scan of the chest with contrast is recommended to

TABLE 83-14 Staging Thymic Tumors

MASAOKA STAGE	DEFINITION
I	Grossly and microscopically encapsulated
IIA	Microscopic transcapsular invasion
IIB	Macroscopic invasion into surrounding tissue excluding pericardium, lung, and great vessels
III	Macroscopic invasion into neighboring organs of the lower neck or upper chest
IVA	Pleural or pericardial dissemination
IVB	Hematogenous or lymphatic dissemination to distal organs
WHO	
A	Tumor with few lymphocytes
AB	Tumor with features of type A and foci rich in lymphocytes
B1	Tumor with features of normal epithelial cells with vesicular nuclei and distinct nucleoli and an abundant population of lymphocytes. Also known as cortical thymoma, lymphocyte-rich thymoma
B2	Thymoma with no or mild atypia with round or polygonal-shaped cells with small component of lymphocytes
B3	Well-differentiated thymic carcinoma with mild atypia
C	Thymic carcinoma with high atypia

determine if the mass is resectable based on relationship to surrounding structures. An MRI with contrast may be performed if clinically indicated. A PET scan may be useful in the evaluation of a patient with thymic tumors, although it may be less useful in the staging of thymoma compared to thymic carcinoma. A core needle biopsy is considered standard of care for obtaining a histologic diagnosis of an anterior mediastinal tumor. This may be obtained via CT or ultrasound imaging. However, in some circumstances, a mediastinoscopy or open biopsy may be required.

Thymomas are commonly staged using the Masaoka system or the WHO staging system, as described in **Table 83-14**. WHO types A, AB, and B1 tend to be more well-differentiated, types B2 and B3 are moderately differentiated, and type C is poorly differentiated.

TREATMENT

Surgical resection is the mainstay of treatment for patients with Masaoka type I and II thymic tumors. In patients with type III and IV who have potentially resectable thymic tumors, neoadjuvant chemotherapy may be given to decrease the tumor size and allow for a resection with negative margins. Surgery remains controversial and provides a limited role in the treatment of stage III and IV disease. No additional therapy may be required in patients with type I who have a resection with negative margins. Postoperative radiation therapy may be recommended based on extracapsular extension and the presence of positive margins in patients with type II or III thymic tumors or histologic evaluation WHO B3 and C. Radiation therapy may be beneficial in patients with locally advanced disease (type III or IV) or in patients with symptoms secondary to compression of surrounding structures. Chemotherapy with cisplatin, doxorubicin, and cyclophosphamide (CAP) remains the mainstay of therapy in the neoadjuvant and adjuvant setting as well as first-line therapy in patients with metastatic thymoma, whereas carboplatin and paclitaxel are often employed in patients with thymic carcinoma. Limited additional agents are recommended based on small phase 2 trials as second-line therapy and beyond.

SUMMARY

The management of SCLC and NSCLC has undergone major change in the past decade, resulting in a reduction in lung cancer mortality. For patients with early-stage disease, advances in radiotherapy and surgical procedures as well as new systemic therapies in the neoadjuvant and perioperative settings have greatly improved prognosis in all diseases. For patients with advanced lung cancer, major progress in understanding tumor genetics and tumor immunology has led to the development of rational targets and specific inhibitors, which have documented efficacy in specific subsets of NSCLC. Furthermore, increased understanding of how to activate the immune system to drive antitumor immunity has proven to be a successful therapeutic strategy for a subset of patients with advanced lung cancer. However, only a small subset of patients responds to immune checkpoint inhibitors, and the majority of patients treated with targeted therapies or chemotherapy eventually develop resistance, which provides strong motivation for further research and enrollment of patients onto clinical trials in this rapidly evolving area.

ACKNOWLEDGMENT

David Johnson, Leora Horn, and Wade Iams contributed to this chapter in the prior edition and material from that chapter has been retained here.

FURTHER READING

Cascone T et al: Perioperative nivolumab in resectable lung cancer. N Engl J Med 390:1756, 2024.

Ettinger D et al: NCCN Guidelines® Insights: Non-Small Cell Lung Cancer, Version 2.2023. J Natl Compr Canc Netw 21:340, 2023.

Hsu M et al: Lung cancer survivorship: Physical, social, emotional, and medical needs of NSCLC survivors. J Natl Compr Canc Netw 22:e237072, 2024.

Owen D et al: Therapy for stage IV non-small-cell lung cancer with driver alterations: ASCO Living Guideline, Version 2023.2. J Clin Oncol 41:e63, 2023.

Rudin C et al: Small-cell lung cancer. Nature Rev Dis Primers 7:3, 2021.

Wakelee H: Chemotherapy and immunotherapy in early-stage NSCLC: Neoadjuvant vs adjuvant therapy. Clin Adv Hematol Oncol 21:648, 2023.

84

Breast Cancer

Nancy E. Davidson

Breast cancer is the most common nonskin cancer diagnosed in women in the world. In 2024, it is estimated that in the United States 310,000 women will be diagnosed with invasive breast cancer, >56,000 women will receive a diagnosis of ductal carcinoma in situ (DCIS), and about 42,250 women will die from breast cancer. Although largely a disease of women, about 2800 men will be diagnosed with and 530 men will die from breast cancer in 2024 in the United States. Thanks to advances in understanding of breast cancer biology, screening, diagnosis, treatment, and decreased use of hormone replacement therapy, 5-year relative survival in the United States is currently 91%. These advances have also made it possible to conceptualize the evolution of breast cancer and how available interventions can be applied across the continuum of changes to improve outcomes **(Fig. 84-1)**.

EPIDEMIOLOGY AND RISK FACTORS

NONGENETIC RISK FACTORS

Female gender and increasing age are the most common risk factors for breast cancer. About 70% of breast cancer in the United States is diagnosed in women 55 years and older, and median age of diagnosis is 63. Incidence is highest in non-Hispanic whites followed by non-Hispanic blacks and is lower in Hispanic, Asian/Pacific Islander, and American Indian/Alaska Native women; in contrast, mortality is highest in non-Hispanic blacks followed by non-Hispanic whites and individuals of other race/ethnicity. Both incidence and mortality vary considerable around the globe, but studies of immigrant populations show that populations who migrate from low-incidence regions to high-incidence regions will attain the breast cancer risk of the higher incidence region within one or two generations.

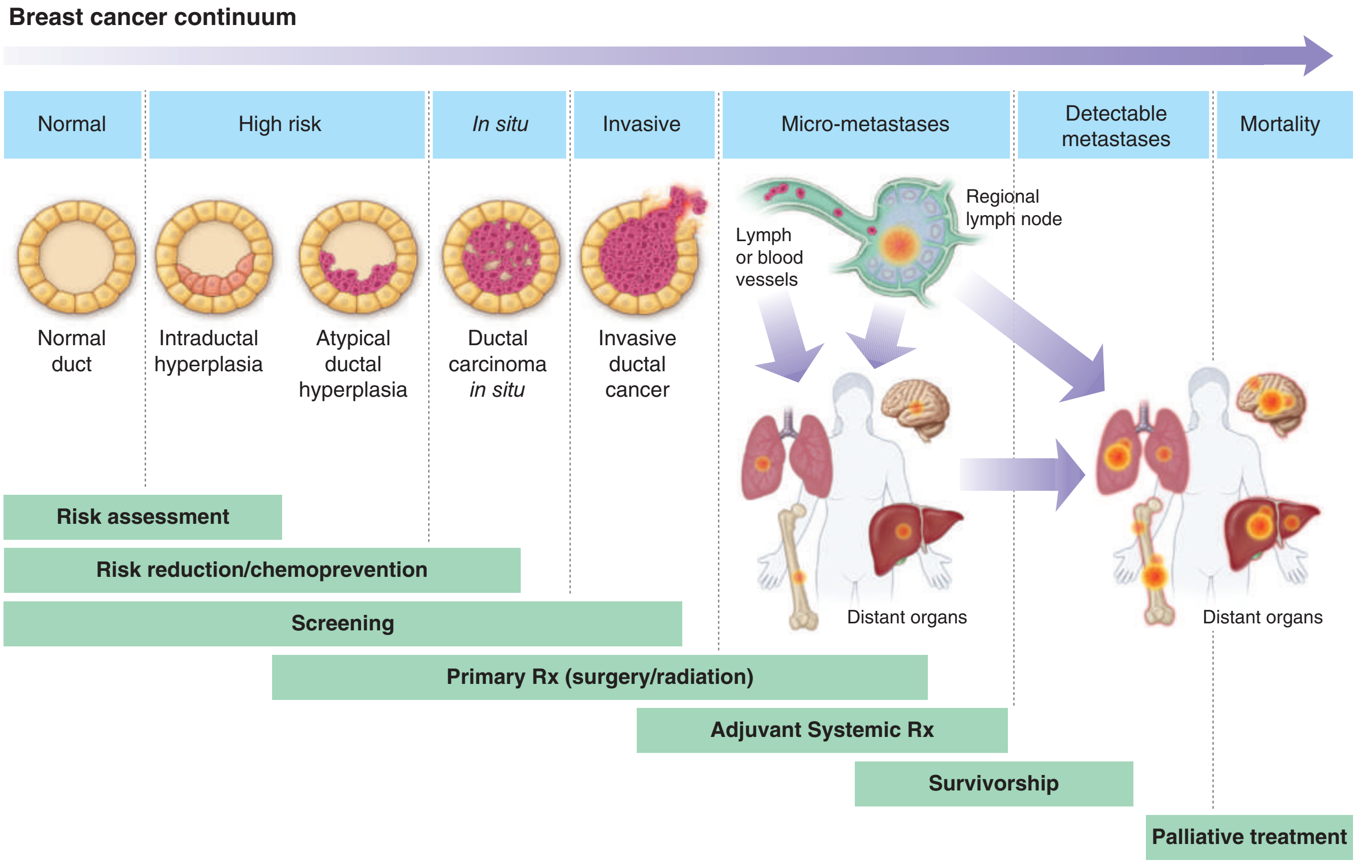

FIGURE 84-1 Breast cancer continuum conceptual model. Most breast cancers begin in epithelial cells within the lobules or ducts. They proceed through a continuum of atypia and hyperplasia to in situ malignancy to invasion into surrounding normal tissues followed by intravasation into lymph and blood channels to local lymph nodes and distant organs, culminating in distant metastases. This is a conceptual model. Not all metastatic breast cancers have progressed through these stages, and many lesions do not progress to the next.

Breast cancer is predominantly a disease resulting from prolonged exposure of the breast to estrogen. Thus, early menarche, late menopause, and late first pregnancy are known risk factors. Likewise, prolonged exposure to hormone replacement therapy (but paradoxically not estrogen replacement therapy) is associated with increased risk as is current use of oral contraceptives. Postmenopausal obesity (but not premenopausal obesity) is also a risk factor, likely because of increased estrogen exposure. Studies of diet and breast cancer risk have not been conclusive, but alcohol consumption is a risk factor.

The best documented exogenous risk factor for breast cancer is exposure to ionizing radiation during adolescence. Studies of other environmental factors such as pesticide or other chemical exposures have not been convincing.

Women diagnosed with some types of benign breast pathology also have a higher risk of subsequent invasive breast cancer diagnosis. In particular, the diagnosis of atypical ductal or lobular hyperplasia increases risk about four- to fivefold, whereas a diagnosis of lobular carcinoma in situ (LCIS) increases risk 7- to 12-fold. Recent studies also suggest that women with high breast density on mammography may be at increased risk of breast cancer.

GENETIC RISK FACTORS

Family history is a critical risk factor, although only 20% of women diagnosed with breast cancer have a family history. Diagnosis of breast cancer in a first-degree relative (parent, sibling, or daughter) doubles breast cancer risk. A personal diagnosis of previous invasive breast cancer also increases the risk of developing a new breast cancer in the ipsilateral or contralateral breast.

Prevention, diagnosis, and management of breast cancer have been revolutionized by the identification of a family of hereditary breast cancer susceptibility genes that account for 5–10% of breast cancers. These genes play a role in DNA damage repair, and inherited mutations, which are transmitted in an autosomal dominant fashion, generally lead to protein truncation and loss of function of DNA repair proteins. The most common mutations are in the *BRCA1* (located on chromosome 17q21) or *BRCA2* (located on chromosome 13q12) genes, and they impart a 50–80% risk of developing invasive breast cancer by age 80 years as well as a 30% risk of developing ovarian cancer. Germline mutations in other genes also lead to a higher risk of breast cancer including *TP53* (Li-Fraumeni syndrome), *PALB2*, *ATM*, *STK11* (Peutz-Jeghers syndrome), and *PTEN* (Cowden syndrome). Certain populations have a higher incidence of *BRCA* mutations, especially Ashkenazi Jews. Commercially available assays for germline, breast cancer susceptibility mutations have expanded from just the two originally discovered genes (*BRCA1/2*) to panels including these and the genes listed above to panels that include these and 20 or more additional genes. Two major concerns have arisen over time regarding the relative lack of clinical correlations inherent in these panels and the unknown association with risk, which have led to some confusion and discomfort among patients: (1) variants of unknown significance (VUS) in the genes known to be associated with increased risk; and (2) genes included in the panels about which little is known regarding clinical risk. Efforts to develop polygenic risk scores that evaluate single nucleotide polymorphisms (SNPs) in other genes are under study but are not yet ready for clinical application.

Testing for germline mutations is readily done using panel testing on DNA obtained from peripheral blood or saliva after appropriate counseling. It is not recommended for the general population. Evidence-based guidelines from the American Society of Clinical Oncology and Society of Surgical Oncology recommend consideration of testing for high-penetrance breast cancer susceptibility genes in individuals with a personal history of breast cancer with specific features including diagnosis ≤65 years. Older patients should be offered testing for clinical features such as Ashkenazi Jewish ancestry; multiple breast cancers; certain types of breast cancer, including breast cancer lacking expression of the estrogen and progesterone receptors and HER2 proteins (triple-negative breast cancer [TNBC]) and lobular breast cancer with personal or family history of diffuse gastric cancer; or close blood relative with early-onset or male breast cancer, ovarian cancer, pancreatic cancer, or metastatic prostate cancer. Testing for individuals without a cancer diagnosis should be considered for those with a family member who tests positive or those with a family history as outlined above or those with higher risk based on existing risk assessment tools such as Tyrer-Cuzick score, BRCAPro, or CanRisk.

PREVENTION

Current strategies to prevent breast cancer include surgery, chemoprevention, and lifestyle modification; their use and impact will vary depending on the underlying risk of developing breast cancer.

PROPHYLACTIC SURGERY

Prophylactic mastectomy reduces risk of developing breast cancer by about 90% including in individuals who carry a germline breast cancer susceptibility gene mutation. Women who opt for prophylactic mastectomy in this setting should be counseled that breast cancer may develop in residual breast tissue. However, preventive mastectomy is not likely to improve outcomes in women with low or only modest risk and should be discouraged. Though the primary role for prophylactic oophorectomy is for ovarian cancer prevention in germline mutation carriers, it can also reduce breast cancer incidence in premenopausal women by about 50% because of reduction in estrogen exposure.

CHEMOPREVENTION

Ample evidence exists for the use of chemoprevention approaches that target estrogen signaling pathways in high-risk women. Tamoxifen reduces risk of invasive breast cancer in women at higher risk (≥60 years, diagnosis of LCIS, or younger women with risk of developing invasive breast cancer ≥1.67% over 5 years based on current risk assessment tools). Newer data suggest that low-dose tamoxifen for 3 years may be effective and well tolerated. Side effects include postmenopausal symptoms and increased risk for endometrial cancer and thromboembolic events, especially in women over 50. Raloxifene may be an alternative for postmenopausal women. Though less effective in reducing breast cancer risk than tamoxifen, it is associated with fewer uterine cancers. The aromatase inhibitors, exemestane or anastrozole, also reduce breast cancer incidence by about 50% in postmenopausal women who are at moderate-high risk for breast cancer. The U.S. Preventive Services Task Force recommends that clinicians offer risk-reducing medications like tamoxifen, raloxifene, or aromatase inhibitors to women who are at increased risk for breast cancer and at low risk for medication side effects. Uptake is low despite the substantial evidence that demonstrates a huge benefit.

LIFESTYLE MODIFICATION

Potential lifestyle modifications to reduce breast cancer risk include maintenance of a normal body mass index, avoidance of alcohol, and minimizing use of hormone replacement therapy. Regular exercise, especially during adolescence, may be associated with reduced risk. Long-term follow-up from the dietary substudy of the Women's Health Initiative showed that a low-fat diet in postmenopausal women who were cancer-free at the time of study enrollment resulted in a nonsignificant reduction in breast cancer incidence but appeared to reduce risk of death from breast cancer. The role of newer GLP-1 agents has not been studied.

SCREENING

Breast cancer screening has been an area of active investigation and controversy for decades. Issues include lack of consensus around the goal of screening, the target population (e.g., age, risk), and the type and frequency of screening. Multiple guidelines exist and continue to evolve. Initial trials focused on three modalities: breast self-exam, clinical breast exam (CBE), and mammography. Though widely promoted in the past, emphasis on breast self-exam has waned, in part because of two randomized clinical trials in China and Russia that showed no benefit. In contrast, a randomized trial in India of CBE every 2 years in women aged 35–64 years with no history of breast cancer indicated that CBE led to a significant reduction in the proportion of women diagnosed with stage III or IV disease and a nonsignificant 15% reduction in breast cancer mortality, largely in women ≥50 years.

Nine randomized trials have studied screening mammography. In aggregate, they demonstrated that screening mammography reduced breast cancer mortality by about 20–25% in women ≥50 years without an impact on overall mortality. The UK Age trial suggested a similar benefit for women who began screening at 40 years. Screening frequency has varied from 1–2 years across trials. Mammography techniques have evolved over the years, and a current U.S. trial is evaluating the role of tomosynthesis as a means of improving benefit from mammography. Multiple guidelines exist and continue to evolve. Most recommend that women aged 50–70 years have mammography every 1–3 years, and many recommend screening for women aged 40–50 years as well. When to stop is not known, but it is generally accepted that benefits are limited to women with a predicted life expectancy of at least 10 years.

The role of magnetic resonance imaging (MRI) as a screening modality is less well studied. MRI is more sensitive, less specific, and more complex to perform. Some guidelines suggest that it be used for women who have a lifetime predicted risk of ≥20%, which includes those with germline pathogenic mutations in the *BRCA* genes. Its use has also been suggested for those with very dense breasts on mammography, but clear evidence of benefit is lacking. Screening ultrasound has also been studied, but evidence supporting its routine use is lacking.

There is considerable interest in the use of multicancer early detection (MCED) tests using circulating DNA as a screening test for diseases like breast cancer. Initial studies have not shown that these tests can substitute for or augment conventional mammographic screening, and women who do have MCED assays performed should still undergo mammographic screening, even if the MCED test is negative.

DIAGNOSIS AND STAGING

Clinical signs and symptoms suggestive of breast cancer may include a breast lump or skin or nipple changes or palpable regional nodes. Thanks to mammographic screening, many patients present with abnormal mammographic findings including a mass, distortion, and/or suspicious microcalcifications without any symptoms or physical exam findings. **Figure 84-2** shows an algorithm for diagnostic evaluation of breast abnormalities noted on physical exam or imaging. Final diagnosis rests on pathologic confirmation, which is generally carried out by image-guided core biopsy to confirm diagnosis, assess tumor grade and morphology, and carry out biomarker evaluation for expression of estrogen receptor (ER) and progesterone receptor (PR) and HER2 proteins and potentially HER2 gene amplification. Suspicious axillary lymph nodes should also be biopsied via image-guided techniques. It is important to image both breasts because up to 3% of women with newly diagnosed breast cancer have unsuspected contralateral disease. Evaluation of the breasts with MRI after a biopsy-proven diagnosis is controversial. On the one hand, MRI is more sensitive than radiologic mammography and will commonly detect previously unidentified disease. On the other hand, this observation may lead to additional surgery, including mastectomy, without obvious improvements in outcomes.

Staging is a cornerstone for breast cancer management because it provides information about natural history and informs decisions about therapy. The traditional TNM (tumor-node-metastasis) staging system has evolved to consider molecular testing including biomarkers as above and certain genomic tests such as the Oncotype 21 gene assay. Staging may take place at time of diagnosis (c or clinical staging) or after surgery (p or pathologic staging) or after preoperative systemic therapy followed by surgery (designated with a "y" prefix). For most asymptomatic individuals presenting with early breast cancer, a careful history and physical exam will be sufficient, and testing can be limited to breast imaging and any testing needed to ensure a safe surgical procedure. Individuals who present with symptoms suggestive of metastatic disease or physical findings of more advanced disease (e.g., large

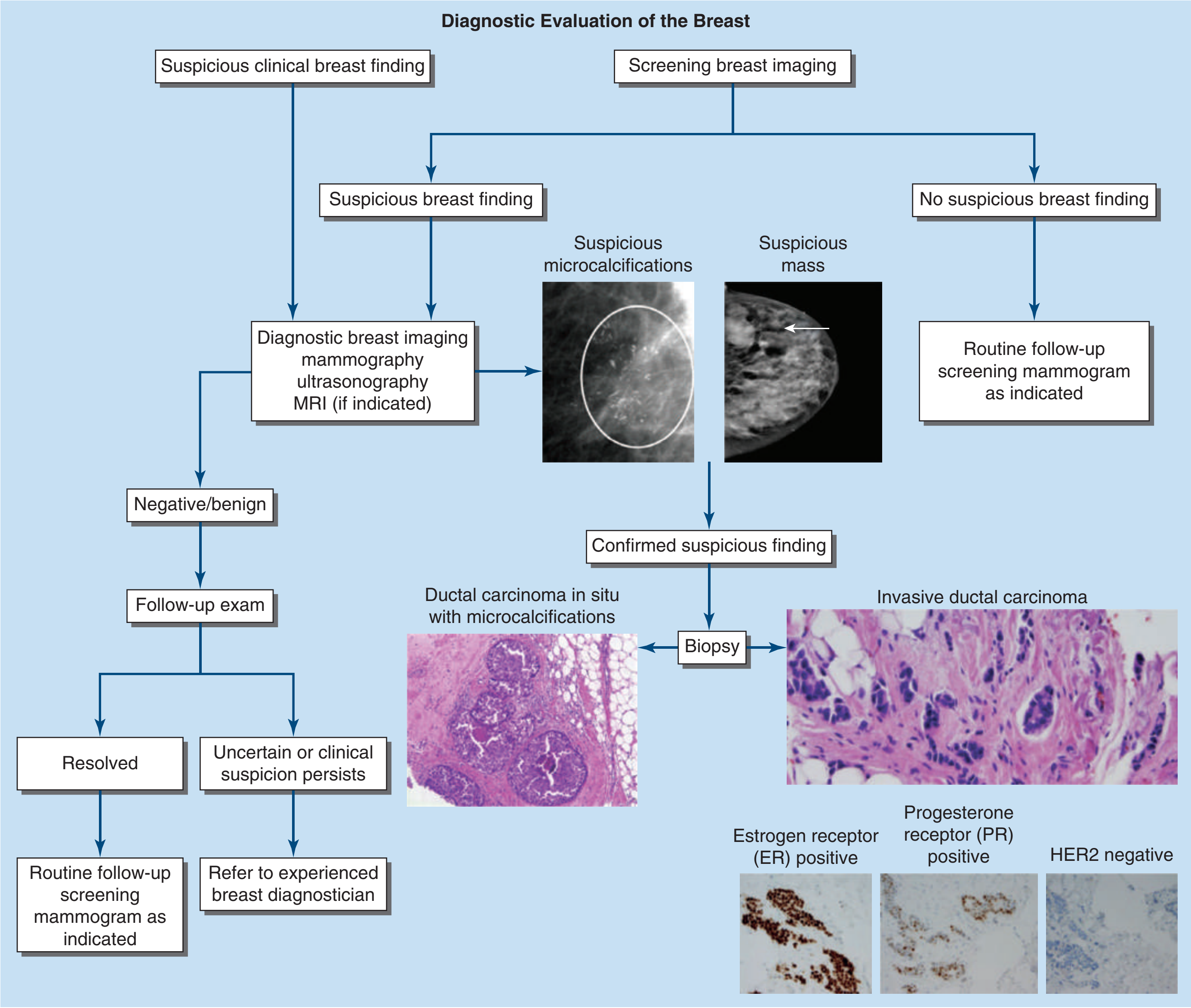
Diagnostic Evaluation of the Breast
Suspicious clinical breast finding
Screening breast imaging
Suspicious breast finding
No suspicious breast finding
Suspicious microcalcifications
Suspicious mass
Diagnostic breast imaging
mammography
ultrasonography
MRI (if indicated)
Routine follow-up screening mammogram as indicated
Negative/benign
Confirmed suspicious finding
Follow-up exam
Ductal carcinoma in situ with microcalcifications
Invasive ductal carcinoma
Biopsy
Resolved
Uncertain or clinical suspicion persists
Estrogen receptor (ER) positive
Progesterone receptor (PR) positive
HER2 negative
Routine follow-up screening mammogram as indicated
Refer to experienced breast diagnostician

FIGURE 84-2 Evaluation and workup of breast lesions. For more extensive details, see *https://www.nccn.org/professionals/physician_gls/pdf/breast-screening.pdf.* *(Mammographic images courtesy of Drs. Mark Helvie and Colleen Neal, Department of Radiology, Michigan Medicine. Photomicrographs courtesy of Dr. Celina Kleer, Department of Pathology, Michigan Medicine.)*

tumor, skin changes, extensive regional adenopathy) should undergo radiologic imaging with computed tomography and radionuclide scanning to look for overt metastatic disease. Suspicious lesions should be biopsied whenever possible to confirm a diagnosis of metastatic breast cancer because of the implications for prognosis and selection of therapy. Less than 10% of patients present with de novo metastatic breast cancer.

Breast cancer staging is currently carried out according to the guidelines of the American Joint Committee on Cancer (AJCC) eighth edition. Conceptually, stages I and II represent early-stage disease, which is confined to breast and ipsilateral nodes, whereas stage III comprises locally advanced breast cancer and stage IV denotes de novo metastatic breast cancer, designated as M1 in the current staging system. Common sites for breast cancer metastasis include soft tissues, lung, bone, liver, and brain. One study suggested that M1 disease can be further divided into at least four subgroups with substantially different prognoses, based on sites and burden of disease as well as biological features. However, these findings have not yet been incorporated into the formal AJCC staging system. Outcomes are directly related to stage at presentation and vary by race and ethnicity in the United States, as shown in **Table 84-1**.

CURRENT UNDERSTANDING OF BREAST CANCER BIOLOGY AND CLINICAL IMPLICATIONS

It is now accepted that invasive breast cancer is in fact a disease with diverse subtypes, both histologic and molecular. Two decades of molecular analyses have documented multiple transcriptional, epigenetic, and genetic changes that characterize invasive breast cancer.

TABLE 84-1 Five-Year Breast Cancer Relative Survival Rate (%) by Stage at Diagnosis and Race/Ethnicity in United States from 2012–2018[a]

STAGE	ALL	WHITE	BLACK	AIAN	HISPANIC	API
I	>99	>99	>99	>99	>99	>99
II	93	93	89	93	92	94
III	75	77	64	72	74	77
IV	29	31	20	35	29	34

[a] Race is exclusive of Hispanic origin.

Abbreviations: AIAN American Indian/Alaska Native; API Asian/Pacific Islander.

Source: Adapted from AN Giaquinto et al: Breast cancer statistics, 2022. CA Cancer J Clin 72:524, 2022.

This has led to the division of invasive breast cancer into at least four major molecular subtypes. Luminal A and B subtypes both express ER and are viewed as generally likely to be responsive to antiestrogen strategies (also called hormonal or endocrine therapies). However, luminal B tumors are characterized by other findings such as reduced PR expression or increased proliferation, usually measured by expression of Ki67, and are associated with a poorer outcome than luminal A tumors. The HER2 subtype expresses HER2 protein at high levels and is more likely to benefit from the use of HER2-targeted therapies. Basal tumors tend to lack expression of ER, PR, and HER2 proteins (TNBC). TNBCs carry a poorer prognosis and are generally managed with chemotherapy and immunotherapy. TNBC can be further divided into at least six subtypes, which may also have therapeutic implications, especially in the metastatic setting.

Pathologic evaluation of a breast cancer specimen begins with an assessment of morphology. Most invasive breast cancers are ductal, but about 10–15% are lobular histologies. Lobular cancers are usually ER positive and HER2 negative. They are distinguished by absence of E-cadherin staining or function, and they have a different natural history, with a propensity to spread to serosal surfaces, including omentum, pleura, and meninges or to various parenchyma, such as ovaries and upper and lower gastrointestinal tracts. This curious site for metastatic recurrence may result in unusual clinical presentations that may delay diagnosis if metastatic disease is not considered. Other special subtypes are also recognized such as mucinous, medullary, papillary, and metaplastic histologies. Mucinous and pure medullary cancers (distinguished by high levels of lymphocytic invasion surrounding the tumor) are associated with very favorable prognoses with local therapy only. Metaplastic cancers may present as very poorly differentiated tumors or even as squamous or sarcomatoid differentiations. They are generally approached as ductal cancers but may be treated in a manner consistent with their relative subtypes, such as squamous or sarcoma-like differentiation. Tumor grade is also assessed using standard approaches such as Nottingham grade, ranging from grade 1 to 3; higher grade is generally associated with poorer prognosis.

Further profiling of the primary tumor specimens may follow depending on the clinical circumstances. Early-stage ER-positive breast cancers are often evaluated by genomic transcriptional profiling assays to identify patients with such a favorable prognosis that adjuvant chemotherapy in addition to endocrine therapy is not needed; five of these are recommended by the American Society of Clinical Oncology: Oncotype Dx (21 genes), MammaPrint (70 genes), ProSigna (50 genes), EndoPredict (12 genes), and Breast Cancer Index (two genes plus a proliferation signature). Although any one of these appears to be relatively accurate in terms of prognosis, they do not measure the same thing and they have not been compared head-to-head. Therefore, ordering more than one assay for a single case is discouraged. At present, more detailed molecular analysis is usually restricted to metastatic tumor specimens to identify molecular alterations that are associated with response to specific targeted therapies, as discussed in the section on treatment of metastatic breast cancer.

TREATMENT

Breast Cancer

MANAGEMENT OF IN SITU BREAST CANCER

Ductal Carcinoma In Situ (DCIS) Routine use of screening mammography has led to a marked increase in the diagnosis of DCIS and LCIS. Untreated DCIS (which is confined to the duct without evidence of invasion through the basement membrane into surrounding interstitial tissue) is associated with a 30% risk of developing a subsequent invasive cancer in the same breast. Therapy is generally focused on excision of the lesion to negative margins and to confirm absence of invasion; evaluation of axillary lymph nodes is not generally undertaken. Excision may be followed by radiotherapy to the ipsilateral breast and consideration of endocrine therapy if the DCIS expresses ER protein to reduce risk of breast recurrence or contralateral primary. Patients with extensive DCIS may require mastectomy, whereas those with a small disease burden may opt for excision alone with regular mammographic surveillance. Molecular assays that can predict lack of benefit from radiotherapy after excision are sometimes used.

Lobular Carcinoma In Situ (LCIS) LCIS is often an incidental finding on a breast biopsy done for abnormalities on physical exam or mammography. It is viewed as a marker for increased risk of developing invasive breast cancer as about 25–30% of women with LCIS subsequently develop invasive disease in either breast. The primary management is continued breast cancer screening and consideration of chemoprevention with tamoxifen or aromatase inhibitor to decrease risk of invasive disease. Bilateral prophylactic mastectomy is sometimes considered but should be generally limited to those who have other risk factors or concerns.

TREATMENT OF EARLY-STAGE BREAST CANCER

Multidisciplinary care is a cornerstone for optimal treatment of early breast cancer. The current approach to treatment of early-stage breast cancer reflects evolution from the initial concept of breast cancer as an orderly disease that spreads from breast to axillary nodes to systemic disease to our current understanding of breast cancer as a potentially systemic disease almost from onset. Therefore, treatment approaches incorporate surgery and radiation to treat local disease and systemic therapy to eliminate or suppress any microscopic distant disease. The goal of these treatments in early-stage disease is reduction of subsequent distant recurrence and mortality. They are given in a manner to optimize efficacy while minimizing toxicity and avoiding overtreatment.

Surgery Six randomized trials have demonstrated equivalent survival with the use of breast-conservation therapy (lumpectomy and, usually, radiotherapy to the remaining breast) or modified radical mastectomy for individuals with early-stage breast cancer. Contraindications to breast conservation include patient preference, poor cosmesis, multifocal disease, previous chest radiation, and ongoing pregnancy that prevents timely administration of radiotherapy. Though not an absolute contraindication, germline mutation in a member of the *BRCA* gene family can drive a decision for bilateral mastectomy because of the high risk of subsequent disease in either breast. Patients who require or choose mastectomy should be informed about availability of immediate or delayed reconstruction options using autologous tissue or implants as clinically appropriate. It is clear, though, that most women with early breast cancer are well served by a lumpectomy to negative margins with axillary management, followed by radiation as appropriate.

Management of the ipsilateral axilla continues to evolve as well. For patients with small tumors and clinically negative axillary nodes, sentinel node localization is used to identify clinically relevant nodes for removal and pathologic assessment to finalize staging and inform decisions about extent of radiotherapy and selection of adjuvant systemic therapy. Use of axillary dissection (removal of level 1 and 2 lymph nodes) is restricted to individuals with palpable lymph nodes or those with substantial pathologic involvement of sentinel lymph nodes. Randomized trials have shown excellent outcomes with omission of any sentinel node evaluation for older women with small biologically favorable tumors and clinically negative axilla who will take adjuvant endocrine therapy as well as omission of axillary dissection for women with low tumor burden in sentinel nodes who will receive radiation.

Radiation As with surgery, the use of postoperative radiotherapy to minimize breast recurrence and nodal recurrence has also changed as a result of data accumulated from serial clinical trials. Current treatment strategies are tailored to the individual based on age, tumor size, and nodal status. Options for external-beam radiotherapy include a standard scheme of daily whole breast radiotherapy (WBR) for 4–6 weeks, hypofractionated schedules lasting 5 days to 3 weeks, or partial breast radiation, usually given in a hypofractionated regimen. Clinical trials of the latter have failed to demonstrate its benefit over WBR. Extension of the radiotherapy port to include regional and axillary nodes is also considered based on pathology and type of surgery. The safety of omitting radiotherapy for women older than 65 with low-risk ER-positive tumors who will take endocrine therapy has been demonstrated. Brachytherapy to apply a radiation source directly to the tumor bed is also employed in some cases as a way of shortening duration of therapy.

Patients who undergo mastectomy may also benefit from chest wall radiotherapy to reduce risk of both local and distant recurrence, thus improving survival. Potential candidates include those with large tumors or positive axillary lymph nodes, especially those with four or more involved nodes, while those with one to three positive nodes or close/minimally involved margins are also considered.

Systemic Therapy It has been recognized that breast cancer is often a systemic disease at time of diagnosis. Enhanced understanding of breast cancer biology, including the concept that resistant clones evolve as a function of random mutations over time, and the identification of multiple targets of therapy have greatly refined our understanding of the role of systemic therapy in early breast cancer. Such therapy may be given preoperatively (neoadjuvant) or postoperatively (adjuvant) and may take the form of cytotoxic chemotherapy, endocrine therapy directed toward ER, and other targeted therapies such anti-HER2, poly (ADP-ribose) polymerase (PARP) inhibitors in cancers with *BRCA1/2* mutations, and/or immunotherapy directed toward immune checkpoints. Treatment algorithms are constantly evolving based on results from clinical trials, and current evidence-based guidelines are available through groups like the National Comprehensive Cancer Network, American Society of Clinical Oncology, or European Society of Medical Oncology.

SELECTION OF NEOADJUVANT VERSUS ADJUVANT SYSTEMIC THERAPY

The primary goal of neoadjuvant therapy (also designated “preoperative” adjuvant therapy) compared to classic postoperative adjuvant therapy is to downsize tumor to make it more amenable to surgery. A second goal is to provide an in vivo assessment of tumor response to the selected systemic therapy and perhaps direct subsequent adjuvant therapy. Data suggest that individual patients who attain a complete pathologic response (pCR) to neoadjuvant therapy have a better outcome than those who do not, although randomized trials of adjuvant versus neoadjuvant therapy have not shown a difference in survival. At least one trial has demonstrated that patients who fail to experience a pCR to standard combination chemotherapy containing an anthracycline, an alkylating agent, and a taxane, especially those with TNBC, appear to benefit from subsequent capecitabine. Currently, neoadjuvant therapy with chemotherapy and anti-HER2 agents is considered for individuals with HER2-positive tumors that are ≥2 cm or node-positive disease, while neoadjuvant therapy with chemotherapy and possibly the checkpoint inhibitor pembrolizumab are considered for individuals with TNBC ≥2 cm or node-positive disease. Neoadjuvant endocrine therapy can be used for postmenopausal women with ER- or PR-positive breast cancer to reduce extent of surgery or perhaps avoid surgery for older women with multiple comorbidities. Women with small tumors of any subtype are often better served by initial surgery to establish pathologic stage followed by tailored selection of systemic therapy.

Types of Systemic Therapy for Early Breast Cancer Endocrine therapy is a mainstay for management of ER- and/or PR-positive invasive breast cancer; positive tumors are commonly defined as those with ≥1% staining by immunohistochemistry. Because of their mechanisms of action as blockers of interaction between estrogen ligand and ER, the selective ER modulators tamoxifen and toremifene can be used in women with any menopausal status. In contrast, estrogen deprivation approaches depend on menopausal status. The use of aromatase inhibitors, such as anastrozole, letrozole, and exemestane, as monotherapy is restricted to postmenopausal women. Ovarian suppression by gonadotropin hormone–releasing hormone (GnRH), such as the luteinizing hormone–releasing hormone (LHRH) agonists goserelin or leuprolide, or alternatively surgical oophorectomy, can be used to reduce circulating estradiol levels in premenopausal women to postmenopausal levels. These women can then also be treated either with tamoxifen or an AI. The CDK4/6 inhibitors abemaciclib and ribociclib may be used in combination with endocrine therapy for high-risk, early-stage, ER-positive breast cancer.

Multiple cytotoxic agents have shown efficacy for treatment of metastatic breast cancer and are now used in treatment of early-stage disease. The most common regimens include cyclophosphamide, doxorubicin, paclitaxel, docetaxel, 5-fluorouracil, methotrexate, and/or one of the platin salt compounds (cisplatin or carboplatin).

Several anti-HER2 therapies are also available. Two monoclonal antibodies, trastuzumab and pertuzumab, and two tyrosine kinase inhibitors, neratinib and lapatinib, are approved for use in early-stage breast cancer; multiple other agents are available for metastatic disease, as discussed below, and are undergoing testing for use in earlier stages of disease.

The role of immunotherapy with the checkpoint inhibitors is also under evaluation. At present, pembrolizumab for patients with TNBC is the only U.S. Food and Drug Administration (FDA)-approved agent for early-stage breast cancer, but others are under investigation.

A targeted agent, olaparib, an oral PARP inhibitor, is also used as an adjunct to chemotherapy for women with high-risk, germline *BRCA*-mutated, HER2-negative breast cancer who have completed (neo)adjuvant chemotherapy. It is particularly active in cancers with DNA repair pathway defects.

GENERAL GUIDELINES FOR (NEO)ADJUVANT SYSTEMIC THERAPY FOR EARLY BREAST CANCER

Approach to ER- and/or PR-Positive Breast Cancer Endocrine therapy is the foundation for adjuvant therapy for women with ER- and/or PR-positive breast cancer. Selection is based on menopausal status and consideration of side effect profiles. For premenopausal women with low-risk breast cancer, tamoxifen for 5–10 years is the standard. Premenopausal women with higher risk tumors are candidates for combined endocrine therapy with an LHRH agonist plus tamoxifen or aromatase inhibitor for 5 years. Multiple randomized trials have shown that aromatase inhibitors for 5 years provide modest but statistically significant superior outcome compared with tamoxifen for postmenopausal women. Women with ER-positive breast cancers are at risk for distant recurrence long after their original diagnosis, from 10–20 or more years. Some trials suggest that up to 10 years of endocrine therapy may be indicated for healthy postmenopausal women with higher risk tumors and good tolerance.

Given the chronic nature of adjuvant endocrine therapy, careful attention to side effects is warranted. Tamoxifen is associated with postmenopausal symptoms and a small chance of thromboembolic events or uterine cancer, especially in women over 50 years. Aromatase inhibitors are associated with postmenopausal symptoms, arthralgias, and bone loss, while LHRH agonists are associated with postmenopausal symptoms and bone loss. Although the three aromatase inhibitors appear to be equally effective, tolerance may vary; patients who are intolerant of one aromatase inhibitor may benefit from change to another member of the family or tamoxifen.

A critical decision in this setting is the value of chemotherapy in addition to endocrine therapy. Its absolute (as opposed to relative or proportional) benefit varies by menopausal status and nodal status. Genomic multigene expression testing of the tumor with assays like Oncotype, MammaPrint, ProSigna, EndoPredict, or Breast Cancer Index will identify a substantial portion of patients with zero to three positive nodes who would not benefit from adjuvant chemotherapy.

Neoadjuvant chemotherapy should be administered to healthy individuals with a large tumor burden or more than three positive nodes, especially premenopausal patients with a primary goal of reducing tumor size. However, in ER-positive disease, the likelihood of a pCR is low. Regardless, adjuvant chemotherapy (neoadjuvant or standard) is considered for healthy individuals with four or more positive lymph nodes as well as for those with lesser nodal burden who have high genomic signature score. Established chemotherapy regimens are outpatient based and use multiple agents for 3–6 months with appropriate supportive care including antiemetics and colony-stimulating factors as appropriate to reduce risk of febrile neutropenia. Long-term side effects can include premature menopause in premenopausal women and a small risk of doxorubicin-related cardiomyopathy, taxane-related peripheral neuropathy, or secondary leukemia, and possible cognitive dysfunction.

Women with node-positive or high-risk node-negative, ER- and/or PR-positive, HER2-negative breast cancer who have completed adjuvant chemotherapy should consider use of a CDK4/6 inhibitor in conjunction with endocrine therapy. Current information supports the use of abemaciclib for 2 years or ribociclib for 3 years; addition of palbociclib for 2 years did not improve outcomes. Those with HER2-negative germline *BRCA*-mutated breast cancer who have received (neo)adjuvant chemotherapy should consider a year of oral olaparib. When used in conjunction with chemotherapy or anti-HER2 therapy, endocrine therapy is generally delayed until completion of adjuvant chemotherapy to minimize toxicity.

Approach to HER2-Positive Breast Cancer Healthy individuals with tumor ≥3 cm or positive lymph nodes are candidates for neoadjuvant therapy with multiagent chemotherapy and trastuzumab with or without pertuzumab for 4–5 months. Those who achieve a pCR should receive further trastuzumab with or without pertuzumab therapy to complete a year. By contrast, those with residual invasive disease at surgery should switch to complete a year of the antibody-drug conjugate (ADC) trastuzumab emtansine (TDM-1). High-risk individuals who have completed a year of anti-HER2 therapy can also consider addition of oral neratinib for another year, although the benefits are incremental.

Toxicity of trastuzumab is relatively uncommon. Few patients have any side effects during treatment. However, the greatest risk is reduction of cardiac ejection fraction. Therefore, baseline cardiac evaluation with an echocardiogram should be obtained, and trastuzumab should not be given to those with low ejection fraction, or they should be evaluated by an experienced cardiologist if therapy is felt to be critical. Serial echocardiograms should be obtained during adjuvant trastuzumab therapy but are not indicated once therapy is discontinued if the patient has not had any evidence of cardiac dysfunction.

Pertuzumab is commonly associated with loose stools or diarrhea. This can be managed with conservative loperamide treatment. However, the added benefit of adjuvant pertuzumab to trastuzumab is modest, and if diarrhea is not easily controlled, trastuzumab should be continued alone.

Individuals who present with smaller tumors with clinically negative nodes should be considered for upfront surgery to establish pathologic stage. This is because women with pathologically node-negative tumors <3 cm have excellent outcomes with a regimen of single-agent weekly paclitaxel for 12 weeks in conjunction with a year of trastuzumab with or without pertuzumab.

Patients with HER2-positive tumors that are ER and/or PR positive should receive endocrine therapy as described above. Endocrine therapy is generally initiated after completion of any chemotherapy and can be given concurrent with anti-HER2 therapy. Neither CDK4/6 inhibitors nor PARP inhibitors have been tested as an adjuvant therapy in early-stage HER2-positive breast cancer.

Triple-Negative Breast Cancer (TNBC) Healthy patients with early-stage TNBC are frequently candidates for (neo)adjuvant chemotherapy. Those with tumors >2 cm or tumors >1 cm and positive axillary nodes should be considered for neoadjuvant multiagent chemotherapy, perhaps in conjunction with the checkpoint inhibitor pembrolizumab. If chemoimmunotherapy is utilized, then pembrolizumab is continued postoperatively to complete a year. Those who achieve a pCR receive pembrolizumab alone. Those patients who do not have a pCR may also consider further chemotherapy with capecitabine. In contrast, those with smaller tumors that are clinically node negative are best served by initial surgery to establish pathologic stage, which might permit a refinement of choice of chemotherapy regimen.

Follow-Up of Survivors of Early-Stage Breast Cancer Asymptomatic survivors of early-stage breast cancer should be followed regularly with history and physical examination to look for any evidence of recurrent disease and to assess for toxicities of prior or ongoing treatments. These exams are conducted every 3–6 months for the first 3 years and diminish in frequency over time. As noted, serial echocardiograms beyond 12 months of trastuzumab therapy are not needed. Annual breast imaging to look for ipsilateral or contralateral disease is the only routine testing needed. In the absence of symptoms or physical exam findings, routine imaging of other types or blood studies have not been shown to enhance well-being or outcome from breast cancer. Current research is focused on use of blood-based assays for circulating tumor cells (CTCs) or circulating free tumor DNA (ctDNA) as markers for early detection of recurrent disease, but these are investigational approaches at present. As many patients have long survival, routine follow-up by a primary care provider and adherence to age-appropriate general health guidelines are key.

Symptom management is a key aspect of follow-up, especially for those on adjuvant endocrine therapy as adherence is a crucial determinant of outcome. Hot flashes may be ameliorated by use of antidepressants like venlafaxine or gabapentin, while vaginal dryness should be addressed through topical agents or, if unsuccessful, vaginal estrogen. Avoidance of hormone replacement therapy is preferred. Aromatase inhibitor–associated musculoskeletal symptoms may be treated by switching from one aromatase inhibitor to another after a 4- to 6-week washout period, or they can be addressed by exercise, nonsteroidal anti-inflammatory agents, acupuncture, or the antidepressant duloxetine. Persistent taxane-related peripheral neuropathy is sometimes responsive to gabapentin.

Special Considerations in Management of Individuals with Early Breast Cancer

Use of Bisphosphonates Bone health can be compromised by breast cancer therapy, especially the estrogen-deprivation agents, aromatase inhibitors and LHRH agonists. Patients who receive these therapies may benefit from use of a bisphosphonate, zoledronate every 6 months for 3 years, or oral clodronate or ibandronate as a bone-strengthening agent. In addition, a meta-analysis suggests that this therapy is associated with decreased breast cancer recurrence, especially in bone. Data on denosumab are less compelling at present.

Pregnancy, Fertility, and Childbearing in Premenopausal Women The diagnosis of breast cancer during pregnancy can be difficult because of the evolving changes in the pregnant breast. Suspicious lumps or skin or nipple changes should be evaluated as above though imaging modalities of choice including ultrasound or MRI. If breast cancer is diagnosed and pregnancy is continued, the goal is to administer appropriate multimodality therapy to optimize maternal outcome from breast cancer and minimize toxicity to the fetus. Breast surgery can be safely undertaken during the second and third trimesters, but any therapeutic radiation must be delayed until after delivery. If indicated, certain chemotherapy agents such as doxorubicin and cyclophosphamide can be administered during the second and third trimesters, whereas others such as methotrexate and 5-fluorouracil should be avoided. In general, endocrine and targeted therapies should not be used until after delivery. Co-management by experienced multidisciplinary breast cancer and high-risk obstetrical teams is preferred.

Diagnosis of breast cancer in all premenopausal women necessitates discussion about fertility preservation and pregnancy with those women who desire later pregnancy. Premature menopause is a known consequence of adjuvant chemotherapy, and likelihood is related to type and duration of chemotherapy and age of patient. It is seen in <50% of women <40 years old but is common in those who are over 40. Patients are asked to avoid pregnancy while on adjuvant endocrine therapy. Thus, counseling about and implementation of fertility-preservation techniques are a priority in newly diagnosed premenopausal women if further childbearing is planned. A meta-analysis suggests that concurrent administration of an LHRH agonist to suppress ovarian function during adjuvant chemotherapy may help to preserve subsequent ovarian function without adverse breast cancer outcomes. Avoidance of oral contraceptives or hormone-based intrauterine devices is recommended for premenopausal breast cancer survivors who require contraception.

Though data are not extensive, pregnancy after early stage breast cancer diagnosis appears to be safe. There is no evidence of increased rate of recurrence compared to matched patients who did not become pregnant, nor is there evidence of increased fetal anomalies. A recent study has documented the safety and pregnancy success rate for women with treated early ER-positive breast cancer who stop adjuvant tamoxifen after 18–24 months of therapy and then resume after pregnancy to complete their prescribed course.

Male Breast Cancer Men with breast cancer usually present with a breast lump or other physical abnormality. Principles about diagnosis, staging, and local therapy are like those for women, although most men undergo mastectomy for primary management for anatomic reasons. Men with breast cancer should undergo genetic testing as germline *BRCA* mutation is seen in up to 14% of men with breast cancer. About 90% of male breast cancers express ER, and adjuvant endocrine therapy is standard. Tamoxifen is the drug of choice; if an aromatase inhibitor is indicated (as for example in an individual who has a history of thromboembolic disease), concomitant administration of an LHRH agonist is required. Guidelines for use of adjuvant chemotherapy are like those for postmenopausal women.

LOCALLY ADVANCED BREAST CANCER

In the United States, about 10% of patients present with locally advanced or stage III breast cancers, although this presentation is more common in less-resourced countries. These tumors are characterized by a large primary tumor, involvement of skin or chest wall, or fixed tumors or axillary lymph nodes, findings that make primary surgical resection difficult or impossible. Inflammatory tumors that present with rapid onset of erythema, swelling, and tenderness of the breast are a subset of locally advanced breast cancer and are sometimes confused with mastitis or other infection. Breast or skin biopsy is critical in the setting of breast inflammation that fails to resolve after antibiotics. Skin biopsy showing dermal lymphatic invasion with tumor cells is often associated with the diagnosis of inflammatory breast cancer. Because up to one-third of these patients have detectable metastasis at time of diagnosis, an evaluation for metastatic disease is recommended even in asymptomatic patients.

Combined-modality therapy begins with neoadjuvant systemic therapy, whose selection is guided by biomarker status as outlined above, to downstage tumor to permit resection. Mastectomy with axillary dissection is often required in the case of inflammatory or other T4 lesions because breast-conserving therapy has been associated with an unacceptably high incidence of locoregional recurrence. Postoperative radiotherapy is the norm in conjunction with systemic therapy with chemotherapy and/or anti-HER2 therapy and/or endocrine therapy tailored to the biological qualities of the tumor.

METASTATIC BREAST CANCER

Presentation and Evaluation About 20–25% of patients who are treated for early breast cancer subsequently develop metastatic disease, presumably because of micrometastatic disease at time of diagnosis that was or becomes resistant to adjuvant systemic therapy. Metastatic disease can be detected years to decades after primary diagnosis, especially in the setting of ER-positive disease, whereas it is most likely to be diagnosed within 3–7 years after treatment for TNBC or ER-negative/HER2-positive breast cancer. Only about 5% of patients present with de novo metastatic or stage IV breast cancer in the United States.

Patients may present with abnormal physical exam or symptoms suggestive of metastases. As shown in **Fig. 84-3**, careful evaluation of extent of disease with computed tomography and radionuclide imaging, routine blood studies, and measurement of tumor markers such as carcinoembryonic antigen (CEA) and either CA27/29 or CA15-3 should be undertaken. Wherever possible, it is critical to biopsy a suspicious lesion to confirm the diagnosis of metastatic breast cancer and assay tumor markers including ER, PR, and HER2 because they may have changed from the initial biopsy under the pressure of time or therapy. In addition, PD-L1 staining and tumor mutational burden should be assessed in tumor, and assays for *PIKC3A* and *ESR1* mutations should be performed in blood or tumor to delineate options for treatment. Next-generation sequencing of tumor may also be considered to simultaneously survey for any other targetable changes to guide selection of therapy or eligibility for a clinical trial.

Goals of Care and General Management Metastatic breast cancer is rarely curable, and the goals of therapy, which is chronic—to palliate or prevent symptoms without undue toxicity—should be explicitly discussed with the patient at the time of diagnosis of metastatic disease. Median survival is <3 years, although the range is wide. Some patients with favorable characteristics such as ER-positive disease, nonvisceral disease, long disease-free interval, and good performance status may survive using serial therapies for many years, whereas those with TNBC are more likely to progress sooner and succumb to their disease. Advances in our understanding of biology of ER- and HER2-positive and genetically mutated (e.g., *BRCA1/2*, *PIK3CA*) breast cancer and development of a myriad of targeted treatments have led to improved outcome for these subtypes in recent years.

Unlike in early breast cancer, the primary intervention in metastatic breast cancer is systemic therapy. Surgery is generally limited to excision of isolated local recurrence or a solitary brain metastasis or stabilization of a bone metastasis. Randomized trials suggest that patients who present with de novo metastatic breast cancer do not have improved outcomes with surgical treatment for the primary breast cancer in addition to systemic therapy and the focus should be on systemic therapy. Radiotherapy may be used at any time to palliate symptomatic localized disease such as bony or brain metastases.

All patients with metastatic breast cancer should have access to palliative therapy approaches in addition to antineoplastic therapy to maximize symptom control and quality of life. In addition, those with metastatic bone disease should be considered for regular administration of bisphosphonate or denosumab in addition to antineoplastic therapy to reduce the chance of skeletal morbidity including pain, fracture, and need for radiotherapy.

Patients with metastatic breast cancer require regular follow-up with history and physical exam to gauge response to therapy. Patient well-being and relief of symptoms are paramount, and use of imaging and blood studies should be personalized to the patient and therapy to adjust dose or schedule and to assist in decisions about efficacy and toxicity of therapy. Current algorithms suggest changes in therapy only if the patient has clear signs of disease progression or unacceptable toxicity. Studies to evaluate the role of serial liquid biopsies to track CTC or ctDNA as an indicator to switch therapy, and even to switch to a targeted therapy suggested by the mutational profile of the ctDNA, in the absence of clinical or radiologic evidence of disease progression are in progress.

Management of ER-Positive Metastatic Breast Cancer Whenever possible, serial use of endocrine therapy is the preferred approach

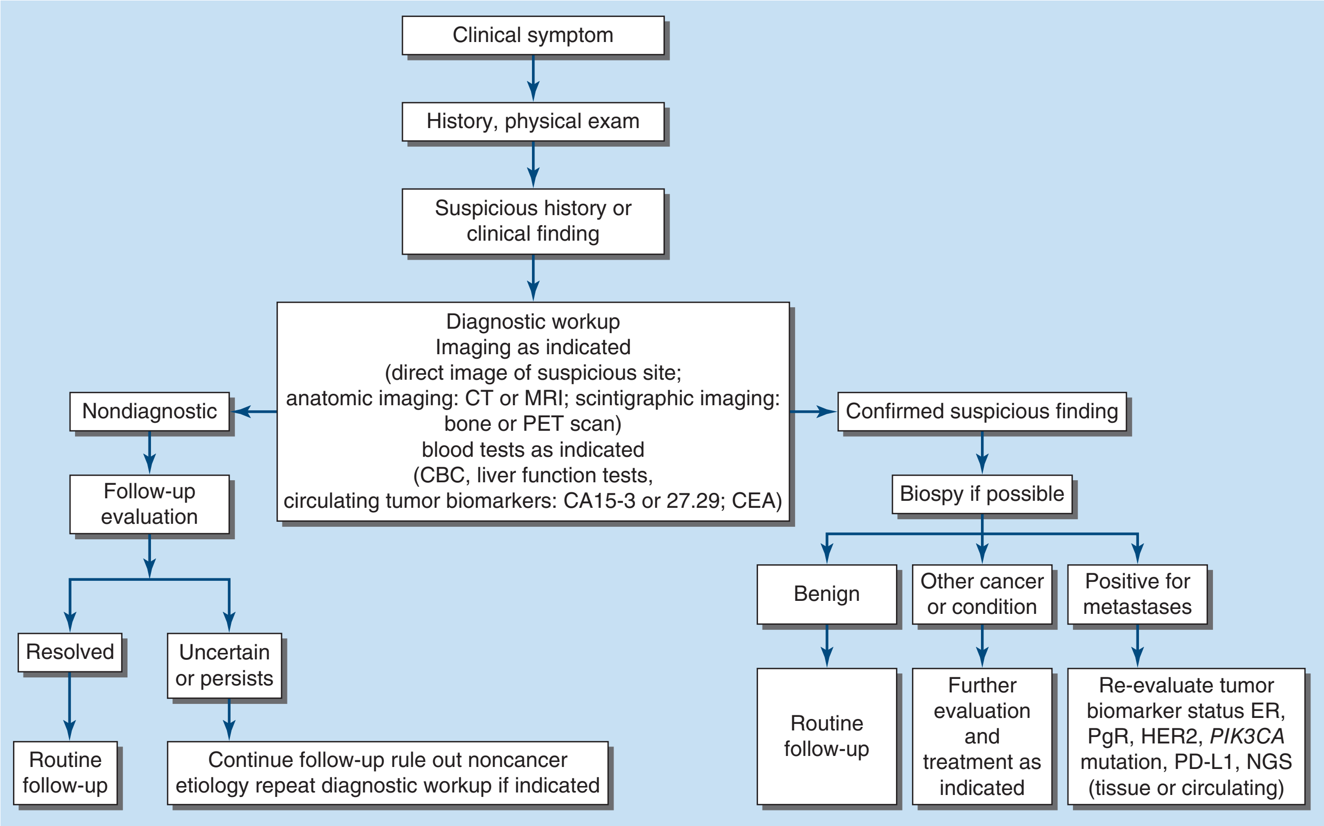
Clinical symptom
History, physical exam
Suspicious history or clinical finding
Diagnostic workup
Imaging as indicated
(direct image of suspicious site;
anatomic imaging: CT or MRI; scintigraphic imaging:
bone or PET scan)
blood tests as indicated
(CBC, liver function tests,
circulating tumor biomarkers: CA15-3 or 27.29; CEA)
Nondiagnostic
Follow-up evaluation
Resolved
Uncertain or persists
Routine follow-up
Continue follow-up rule out noncancer etiology repeat diagnostic workup if indicated
Confirmed suspicious finding
Biospy if possible
Benign
Other cancer or condition
Positive for metastases
Routine follow-up
Further evaluation and treatment as indicated
Re-evaluate tumor biomarker status ER, PgR, HER2, *PIK3CA* mutation, PD-L1, NGS (tissue or circulating)

FIGURE 84-3 Evaluation of new signs or symptoms in a patient with prior history of early-stage breast cancer. See text for details. CBC, complete blood count; CEA, carcinoembryonic antigen; ER, estrogen receptor; NGS, next-generation sequencing; PET, positron emission tomography; PgR, progesterone receptor.

to patients whose recurrent tumor is ER and/or PR positive and whose clinical presentation is not dire. Use of chemotherapy can be reserved for those with life-threatening visceral disease like lymphangitic lung metastases or impending liver failure. However, a prospective trial has suggested that overall survival was the same in patients with apparent rapidly growing visceral disease randomly assigned to either chemotherapy or endocrine therapy and a CDK4/6 inhibitor. There is no value for concurrent endocrine therapy and chemotherapy for management of metastatic breast cancer. Selection of type of endocrine therapy will depend on menopausal status and previous adjuvant endocrine therapy. It may be paired with targeted therapy depending on the molecular profile of the cancer and clinical scenario.

For the uncommon patient who presents with untreated ER-positive metastatic breast cancer, therapy generally begins with aromatase inhibitor plus CDK4/6 inhibitor for postmenopausal women and LHRH agonist plus aromatase inhibitor plus CDK4/6 inhibitor for premenopausal women; median progression-free survival is about 2 years, and median overall survival is >3 years with this approach.

More commonly, patients present with recurrent disease diagnosed while on or after completing adjuvant endocrine therapy. For those who have received aromatase inhibitor, up to one-half will have evidence of *ESR1* mutation and may be candidates for the oral selective estrogen receptor degrading (SERD) agent elacestrant if they have also received a CDK4/6 inhibitor. Those without evidence of *ESR1* mutation are often treated with an alternate SERD, fulvestrant, which is administered monthly by intramuscular injection; this may be coupled with a CDK4/6 inhibitor if the patient has not previously received such an agent. The value of continuing or switching to another CDK4/6 inhibitor for those who have previously received adjuvant CDK4/6 inhibitor is not clearly defined in this setting.

For postmenopausal patients previously treated with endocrine therapy who are found to have a *PIKC3A* mutation by liquid or tissue biopsy, the use of alpelisib (a selective inhibitor of PI3Kα) with fulvestrant can be considered. Postmenopausal patients with *PIK3CA/AKT1/PTEN* alterations who have received one endocrine therapy for metastatic disease or recur within 12 months of completing adjuvant therapy can receive capivasertib (a pan-AKT inhibitor) with fulvestrant. The use of aromatase inhibitor plus everolimus, an mTOR inhibitor, can be considered for patients without targetable mutations as it, too, has been shown to increase progression-free survival over endocrine therapy alone. Each of these targeted agents carries a unique set of toxicities, and careful monitoring of clinical status and blood studies is needed for safe administration.

Other approaches, especially for patients who have demonstrated repeated and lengthy responses to serial hormone therapies, can include switch to another aromatase inhibitor or tamoxifen or administration of additive hormone therapies such as progestins, androgens, and estrogens. The latter category of agents is seldom used at present because of concerns about toxicity and the advent of newer more targeted therapies.

At some point, it will become apparent that a patient has tumor that is unresponsive to endocrine therapy, which necessitates a transition to other types of systemic therapy. The timing of this decision may be informed in part by the observation that likelihood of disease control and the duration of benefit are decreased by about one-half with each successive switch in endocrine therapy.

Management of Endocrine-Unresponsive ER-Positive, HER2-Negative Breast Cancer or TNBC Chemotherapy is the backbone of systemic therapy for individuals with metastatic HER2-negative breast cancer that is not responsive to endocrine therapy. Unlike in the setting of early breast cancer where multiagent chemotherapy is the norm, use of combination chemotherapy in metastatic breast cancer should be reserved for the uncommon situation of visceral crisis where a rapid reduction in tumor burden is desirable. For most patients, serial use of single-agent chemotherapy is associated with an acceptable likelihood of disease palliation without excess toxicity; combination chemotherapy does not improve overall survival over sequential monotherapy.

Many chemotherapeutic agents are active in metastatic breast cancer, and selection is based on previous adjuvant chemotherapy and patient preference about route and schedule of administration and side effects. The oral agent capecitabine is often used initially because of ease of administration and acceptable side effect profile. However, several studies have demonstrated that a small percentage of the population harbors deleterious mutations in the *DPYD* gene, which is responsible for converting capecitabine to its inactive metabolites for excretion. Such patients are prone to suffering extreme and, in some cases, lethal toxicities, and some guidelines call for routine testing for this gene before initiating capecitabine or other fluoropyrimidines, such as 5-fluorouracil. Other options include taxanes, anthracycline (depending on previous adjuvant exposure), vinorelbine, gemcitabine, cyclophosphamide, methotrexate, eribulin, ixabepilone, and platinum-based agents; the last is often considered in TNBC because of evidence of activity of platinums in that subtype. For patients with TNBC that expresses PD-L1, administration of paclitaxel with the checkpoint inhibitor pembrolizumab can be considered because the combination results in better outcomes than taxane alone; activity of other checkpoint inhibitors is under investigation.

ADCs are increasingly finding a role in treatment of metastatic breast cancer. Sacituzumab govitecan, an antibody targeted against TROP-2SN-38 coupled to an irinotecan metabolite, has shown activity against TNBC and endocrine-unresponsive ER-positive, HER2-negative breast cancer. A second ADC, trastuzumab deruxtecan, which couples trastuzumab with the cytotoxic deruxtecan (discussed in greater detail below in the HER2 section) is also active against tumors that express low levels of HER2 (1+ or 2+/fluorescence in situ hybridization [FISH] negative) and were initially labeled HER2 negative and thus ineligible for adjuvant anti-HER2 therapy using other anti-HER2 agents. Development and testing of novel ADCs are very vibrant areas in treatment of metastatic breast cancer, and findings are beginning to challenge some of our concepts about targeted therapy.

For patients whose tumors have evidence of germline or somatic *BRCA* mutation, use of single-agent PARP inhibitor with olaparib or taloparib should also be considered, perhaps before the transition to infusional therapies because of convenience.

Management of HER2-Positive Breast Cancer The landscape of treatment for metastatic breast cancer has changed substantially in the 25 years since the importance of HER2 protein expression as a predictor for response to anti-HER2 therapy was demonstrated. FDA-approved HER2-directed therapies now include monoclonal antibodies (trastuzumab, pertuzumab, margetuximab), small-molecule tyrosine kinase inhibitors (lapatinib, neratinib, tucatinib), and ADCs (ado-trastuzumab emtansine or TDM-1, trastuzumab-deruxtecan), and others are under development. Apart from trastuzumab-deruxtecan, their clinical benefit is seen only in individuals whose tumors express 3+ HER2 staining by immunohistochemistry or are gene-amplified by FISH.

The preferred first-line treatment for metastatic HER2-positive breast cancer is a combination of taxane, pertuzumab, and trastuzumab for several months; taxane chemotherapy may be stopped when disease stability is reached, and the anti-HER2 agents are continued as maintenance therapy. In some cases, this approach can lead to long-term remission, and an active area of research is how long to continue anti-HER2 therapy for those with complete response. A randomized clinical trial has established the superiority of trastuzumab-deruxtecan for second-line therapy. Thereafter, serial administration of anti-HER2 agents, in some cases in combination with single-agent chemotherapy, is pursued. Patients who have HER2-positive, ER- and/or PR-positive breast cancer may benefit from concomitant endocrine therapy as well.

HER2-positive breast cancers have a propensity for involvement of the central nervous system. Because control of systemic disease with anti-HER2 therapies can be excellent and penetration of the antibody-based agents across the blood-brain barrier is poor, patients with this disease subtype are often found to have brain metastasis and may require radiotherapy for local disease control. The small-molecule inhibitors may also offer disease control as they cross into the central nervous system.

■ FURTHER READING

Andre F et al: Biomarkers for adjuvant endocrine and chemotherapy in early-stage breast cancer: ASCO Guideline Update. J Clin Oncol 40:1816, 2022.

Bedrosian I et al: Germline testing in patients with breast cancer: ASCO-Society of Surgical Oncology guideline. J Clin Oncol 42:584, 2024.

Benitez Fuentes JD et al: Global stage distribution of breast cancer at diagnosis: A systemic review and meta-analysis. JAMA Oncol 10:71, 2024.

Burstein HJ: Systemic therapy for estrogen receptor-positive, HER-2 negative breast cancer. N Engl J Med 383:2557, 2020.

Curigliano G et al: Understanding breast cancer complexity to improve patient outcomes: The St Gallen International Consensus Conference for the Primary Therapy of Individuals with Early Breast Cancer 2023. Ann Oncol 34:970, 2023.

Giordano S et al: Systemic therapy for advanced human epidermal growth factor receptor 2-positive breast cancer: ASCO guideline update. J Clin Oncol 40:2636, 2022.

Henry NL et al: Biomarkers for systemic therapy in metastatic breast cancer: ASCO guideline update. J Clin Oncol 40:3205, 2022.

Jackson EB, Chia SKL: Sequencing of endocrine therapy targeted therapies in hormone-sensitive, human epidermal growth factor receptor 2-negative advanced breast cancer. J Clin Oncol 41:3976, 2023.

Loibl S et al: Breast cancer. Lancet 397:1750, 2021.

Nielsen S, Narayan AK: Breast cancer screening modalities, recommendations, and novel imaging techniques. Surg Clin North Am 103:63, 2023.

PDQ Screening and Prevention Editorial Board: Breast Cancer Prevention (PDQR). *https://www.ncbi.nlm.nih.gov/books/NBK65884/*. Accessed January 7, 2024.

85

Upper Gastrointestinal Tract Cancers

David Kelsen

Cancers of the upper gastrointestinal tract include malignancies of the esophagus, stomach, and small bowel. Esophageal, gastroesophageal junction (GEJ), and gastric cancers are among the most common of human malignancies, with 1.693 million global new cases diagnosed in 2020; as a measure of the gravity of such a diagnosis, there were an estimated 1.312 million deaths. In the United States, a lower-risk area, it is estimated that in 2023, esophageal cancer will be diagnosed in 21,560 people and cause 16,120 deaths; for gastric cancer, 26,500 new cases will be diagnosed and 11,130 deaths will occur. Small intestine cancers are rare.

ESOPHAGEAL CANCER

■ GLOBAL DIFFERENCES IN INCIDENCE AND CAUSATIVE FACTORS

Two distinct forms of cancer with different epidemiologies, causative factors, and genomic profiles arise within the esophagus: squamous cell cancers (SCCs), which occur more frequent in the upper and mid esophagus; and adenocarcinomas, which are almost always located in the lower esophagus and at the GEJ. The incidence of esophageal cancer varies up to 20-fold in global geographic distribution: it is relatively uncommon in North America but has a high incidence in Asia (especially China), the Normandy coast of France, and Middle Eastern countries such as Iran. This marked global variation is likely due to different causative factors in the development of the malignancy, leading to two different cancer types within the same organ: globally, SCCs make up the majority of cases, as they are more common in high-incidence areas, usually with lower of the four-tier Human Development Index (HDI) scores (a measure of economic development that includes standard of living, health, and education). Overall, approximately 604,100 new cases of esophageal cancer were diagnosed globally in 2020; esophageal cancer was the tenth most common cause of malignancy and the sixth most common cause of cancer-related mortality, with an estimated 544,076 deaths.

The most established high-risk factors for the SCC subtype in Western countries are alcohol or tobacco abuse; concurrent alcohol and tobacco abuse further increases the risk. Ingestion of extremely hot substances (such as tea in Iran and mate [maté] in South America) has been proposed as a risk factor; in India, chewing the areca (betel) nut increases the risk of esophageal SCC. Less common risk factors include chronic achalasia, radiation therapy (such as is delivered for treatment of Hodgkin's lymphoma or breast cancer), lye ingestion, and Plummer-Vinson (Patterson-Kelly) syndrome (iron deficiency anemia, glossitis, cheilosis, and the development of esophageal webs). Adenocarcinoma of the lower esophagus and GEJ has been the predominant histologic subtype in the United States and Western Europe for several decades, now making up >75% of all incident cases. Risk factors for adenocarcinoma include chronic reflux esophagitis leading to inflammation and the development of Barrett's esophagus (the finding of glandular gastric type mucosa extending into the esophagus). Although obesity increases the risk of reflux esophagitis, a substantial number of patients with newly diagnosed adenocarcinoma of the esophagus and GEJ are younger and fit; Barrett's esophagus may still be found in these patients. In patients with adenocarcinoma of the lower esophagus in which Barrett's esophagus is not present, the disease may arise without Barrett's esophagus, or an extensive tumor found at diagnosis may obliterate previous areas of Barrett's. Genomic alterations may be identified even before the development of frank adenocarcinoma in patients with dysplasia associated with Barrett's esophagus. These include mutations of *TP53*, a gene critical in regulating uncontrolled cell division, and aneuploidy in dysplastic regions. Risk of progression of Barrett's esophagus to cancer is about 0.4–0.5% per year. **Management of Barrett's esophagus is discussed in Chap. 334.**

As opposed to other gastrointestinal malignancies, such as colorectal cancer, inherited cancer susceptibility genes are rarely associated with esophagus and GEJ cancers. An exception is the rare inherited cancer susceptibility gene driving tylosis palmaris and plantaris; a mutation in the *RHBDF2* gene is associated with an increased risk for SCC of the esophagus. Both Lynch syndrome and germline *BRCA* mutation carriers have a modestly increased risk of gastric and potentially GEJ adenocarcinomas. Identification of these germline mutations is important because it guides choices of systemic therapy, as well as counseling of family relatives (see below).

SCREENING AND SURVEILLANCE OF HIGHER RISK GROUPS

Because of its low incidence in North America and the absence of proven blood-based biomarkers for esophageal cancer assays, screening of the asymptomatic general population using, e.g., upper endoscopy is not currently recommended in the United States. Periodic endoscopy is used for surveillance of higher risk patients, such as those with Barrett's esophagus, especially with dysplasia, based on expert opinion guidelines.

Early-Onset GEJ and Gastric Cancers A marked increase in the incidence of early-onset gastrointestinal malignancies (EOGI; cancer occurring in people <50 years of age) has been noted over the past several decades. First noted were early-onset colorectal cancers, occurring particularly in the left-sided colon or rectum; the start of the steep increase was traced back to the mid-1990s. A similar trend has been noted in upper gastrointestinal tract cancers, particularly in women. The large majority of EOGI patients do not have inherited germline cancer susceptibility gene mutations or a family history of malignancy for colon cancer or small-bowel cancer or inflammatory bowel disease. Driving factors suggested for the steep increase include lifestyle changes such as being more sedentary (less exercise), dietary changes leading to obesity, use of alcohol and tobacco, changes in the microbiome, and exposure to an as not yet identified environmental agent, not previously used, which was introduced to the general population in the 1980s or 1990s. Because the absolute number of young people developing GEJ or gastric cancers is still small, surveillance upper endoscopy or imaging is not recommended (for colon cancer screening, current U.S. recommendations are to start at age 45 years). Efforts are underway to develop sensitive and specific early detection biomarkers such as blood-based assays. Because these cancers are uncommon in young patients, symptoms may be attributed to benign illnesses so that advanced stages are common at the time of diagnosis. Physicians should be aware that symptoms in a young person found more typically in older patients, such as dysphagia, odynophagia, or upper abdominal discomfort, may be related to malignancy, leading to the same diagnostic studies as performed in older patients.

GENOMIC ALTERATIONS

Genomic alteration analyses have revealed substantial genomic differences between adenocarcinomas and SCCs of the esophagus. The well-known integrated analysis involving several different genomic platforms, performed by The Cancer Genome Atlas (TCGA) Research Network investigators, demonstrated that esophageal SCC more closely resembled SCCs of other primary sites, such as the head and neck, than adenocarcinomas arising in the esophagus. Three molecular subclasses of SCC were identified (of note, as opposed to SCC of the head and neck, human papillomavirus was not identified in any of the three subgroups). Similarly, a study of 528 Chinese SCC patients, using whole genome sequencing, evaluated structural variations resulting in rearrangements; five types of structural variations were identified, and some of these may be of clinical significance. In adenocarcinomas, *HER2/ERBB2* is frequently amplified, and microsatellite instability can be found in both esophageal and GEJ adenocarcinomas. Epstein-Barr virus (EBV) infection is a driver in adenocarcinoma. The similar genomic profile for adenocarcinomas of the distal esophagus, GEJ, and cardia of the stomach suggests that proximal gastric and GEJ tumors may have a similar driving factor (see below). A genome-wide association study identified 27 risk loci for the development of adenocarcinomas of the esophagus or GEJ, with a difference between adenocarcinoma arising in Barrett's esophagus versus not. Other studies comparing transcriptomes of adenocarcinomas and SCC across organs (i.e., the same tumor histology arising in different organs, such as SCCs and adenocarcinomas from the esophagus, lung, and uterine cervix) found that the same histologies among the different organs showed more similarity than between the different histologies within the same organ. In addition to implications regarding driving factors in the initiation and progression of cancer, these genomic alterations are important for therapeutic decisions involving systemic agents given in the neoadjuvant or postoperative adjuvant setting or for advanced metastatic disease. For esophageal adenocarcinoma, genomic abnormalities that should be considered in prescribing drug-based therapy include analysis for *HER2/ERBB2* amplification, programed death ligand 1 (PD-L1) expression, and hypermutated tumors/microsatellite instability (see **Table 85-1**, and below). For SCC of the esophagus, the degree of PD-L1 expression may be of importance.

CLINICAL FEATURES

PRESENTING SYMPTOMS

The most common symptoms leading to suspicion of esophageal cancer are dysphagia or odynophagia and, less frequently, hematemesis or melena. More subtle symptoms include anorexia and weight loss, and

TABLE 85-1 AJCC Prognostic Stage Groups for Esophageal Cancer Using cTNM (Pretreatment)[a]

			5-YEAR SURVIVAL RATE	
TNM	CLINICAL STAGE	PRESENTING AT THIS STAGE[b,c]	SQUAMOUS	ADENOCARCINOMA
cTis, N0, M0	0	1.2%	75%	82%
cT1-2, N0, M0	I	17%	75%	78%
cT1-2, N1-3, M0	IIA	7%	53%	50%
cT3-4a, N0, M0	IIB	13%	40%	40%
cT3-4a, N1-3, M0[c]	III	31%	25%	25%
cT4b, any N, M0	IVA		17%	21%
cAny T, any N, M1	IVB	5%	10%	18%

Survival by ypTNM Staging After Neoadjuvant Chemotherapy

		ESTIMATED 5-YEAR SURVIVAL RATE	
TNM	yp STAGE	SQUAMOUS	ADENOCARCINOMA
T1-2, N0, M0 T1, N1, M0	I	46%	52%
T3, N0-1, M0 T2, N1-2 M0 T1, N2-3, M0 T4a, N0, M0	II	34%	38%
T4a, N1-3 M0 T4b, any N, M0 T3, N2-3, M0 T2, N3, M0	III	22%	27%
Any T, any N, M1	IV	10%	12%

[a]AJCC Cancer Staging Manual Version 9 expected in 2024. [b]Squamous cell and adenocarcinoma histologies combined. [c]Surgical series; underestimates incidence of M1 disease at presentation. [d]Incidence includes cT4b and cNanyM0.

Sources: Adapted from TW Rice et al: CA Cancer J Clin 67:304, 2017; TW Rice et al: Dis Esophagus 29:707, 2016; and TW Rice et al: personal communication.

fatigue and shortness of breath if anemia from gastrointestinal bleeding is present. Because the symptoms of dysphagia or odynophagia are usually not perceived by the patient until substantial obstruction of the esophageal lumen has occurred, the large majority of patients with esophageal cancer are found with locally advanced if not metastatic disease. Patients with symptoms of dysphagia and/or odynophagia should undergo upper endoscopy to determine the presence or absence of malignancy; biopsy should be performed at the same setting to determine histology. Depending on the tumor stage, molecular diagnostic or next-generation sequencing (NGS) analysis to assist in determining potential therapies should be performed. PD-LI, which may guide use of immune modulation therapy, is assessed by immunohistochemistry (IHC). NGS requires adequate tumor cellularity, which may be difficult to achieve from endoscopic biopsy. NGS is currently probably the most efficient methodology for genomic analysis (other than PD-L1), and genomic analysis should be done on all patients with metastatic disease because it will guide therapy. Some high-volume U.S. centers routinely perform NGS on all specimens, including from all but the most early-stage primary tumors of patients without metastatic disease.

STAGING

Therapeutic strategy is currently based on the stage of the disease using a system such as the eighth version of the American Joint Committee on Cancer (AJCC) Staging Manual (for esophageal and gastric cancers, the ninth version is expected in 2024) tumor-node-metastasis (TNM) staging system. The T stage is based on the size of the tumor and depth of penetration through the esophageal wall (which for most of its course is not covered by serosa so that invasion through the muscle layer leads directly into periesophageal tissues) **(Fig. 85-1)**. Patients with regional lymph node metastases are still potentially curable. Metastatic disease is generally treated with palliative intent with rare exceptions. Because neoadjuvant (preoperative) therapy is widely employed for esophageal cancer to improve subsequent surgical outcomes, the AJCC TNM staging system includes clinical, pathologic (for patients undergoing initial surgery as first treatment), and ypTNM staging assessment for those treated with preoperative therapy. See **Table 85-2** for the TNM staging classification for gastric cancer, which is similar to esophageal cancer.

Determining tumor extent includes careful physical examination, which may reveal palpable lymphadenopathy or hepatomegaly; imaging studies including computed tomography (CT) and fluorodeoxyglucose (FDG) positron emission tomography (PET)-CT scan are used to assess for metastatic disease. If no metastatic disease is identified, endoscopic ultrasonography (EUS) is commonly performed to more definitively determine depth of penetration of the primary tumor (T) and regional lymph node involvement. Deep learning artificial intelligence methodology may assist the imaging analysis to improve the diagnostic sensitivity and specificity of EUS for both esophageal and gastric cancers. For tumors of the mid and upper esophagus (5% of esophageal cancers are in the upper third of the esophagus, 20% in the middle third, and 75% in the lower third), bronchoscopy may be performed to rule out invasion of the tracheobronchial tree. The finding of invasion of the trachea or bronchus rules out surgical intervention with curative intent. Regional lymph nodes may be biopsied under EUS guidance. For GEJ tumors, as for gastric cancers, laparoscopy is performed at many centers to assess for peritoneal involvement if the T stage is more advanced. If metastatic disease is suspected, biopsy to confirm tumor staging and to obtain adequate tissue for molecular and genomic alterations analysis should be performed. Assay of plasma ctDNA (circulating tumor DNA) can be used to detect genomic alterations before treatment if biopsy tissue is too limited to allow NGS or molecular diagnostics; ctDNA is being studied as a measure of minimal residual disease after definitive regional therapy (surgery or chemoradiation therapy). If systemic therapy is indicated as a portion of the treatment (for metastatic disease or for preoperative therapy for locally advanced cancers), serial FDG-PET/CT scans, using decrease in FDG avidity as a surrogate measure of effectiveness, are being used to guide whether the initial therapy should be continued or changed.

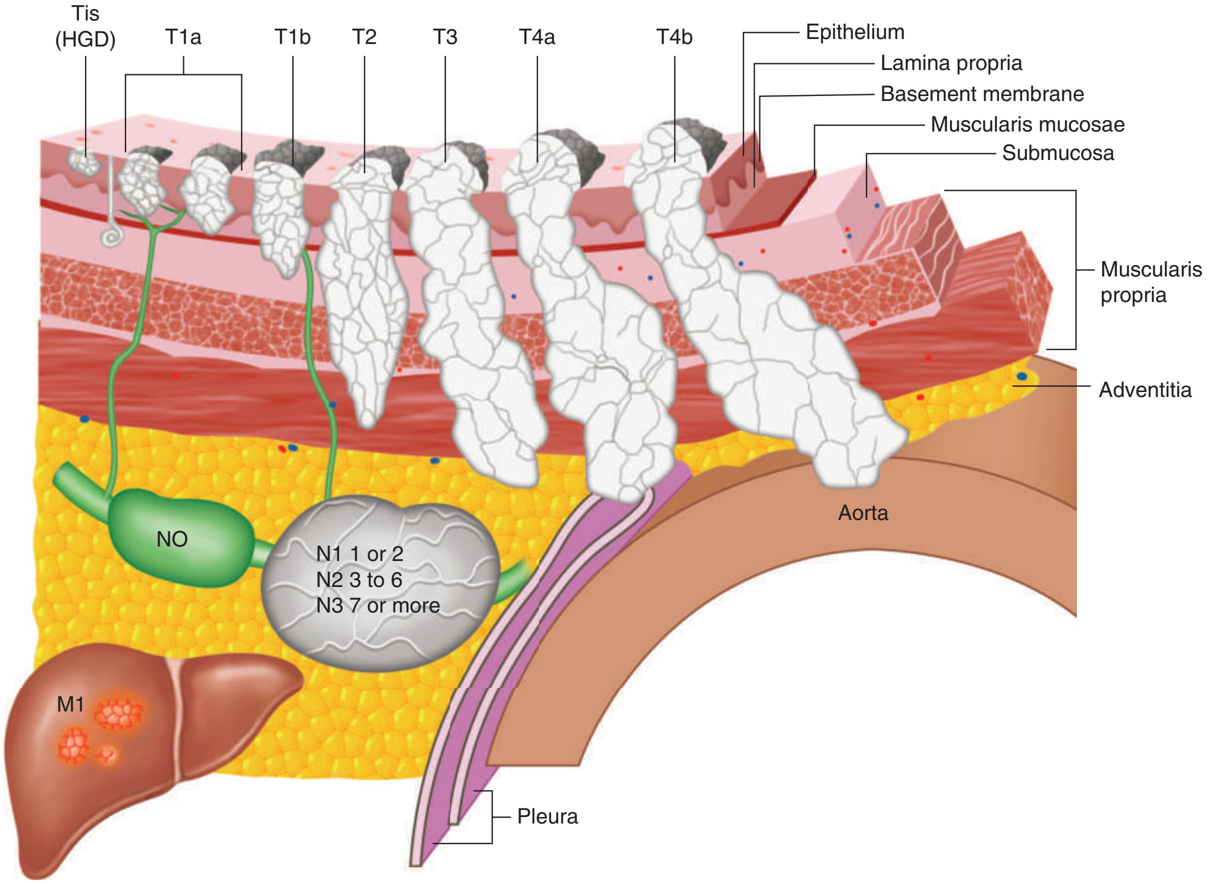

FIGURE 85-1 Patterns of spread of esophageal cancer and the basis for anatomic staging. HGD, high-grade dysplasia. *(Reproduced with permission from TW Rice et al: Cancer of the esophagus and esophagogastric junction: An eighth edition staging primer. J Thorac Oncol 12:36, 2017.)*

TREATMENT

Esophageal Cancer

Although the prognosis for patients with esophageal cancer (all stages) is still poor, a slow but steady improvement in 5-year survival has been noted. Because no cost-effective early detection methods exist in low-incidence countries, the number of U.S. patients found to have very-early-stage cancers at the time of diagnosis has not markedly increased; the modest improvement in survival is probably a combination of somewhat improved systemic therapy as well as decreased operative morbidity and mortality when surgery is performed by high-volume surgeons at high-volume centers, as well as improvements in the delivery of external-beam radiation therapy.

For patients without evidence of metastatic disease, the goal of therapy is cure, usually by employing combined-modality therapies. Except for patients with early-stage esophageal cancer, who might be treated by surgery alone (or for very-early-stage lesions [Tis or T1a], by endoscopic mucosal resection [EMR] or submucosal dissection [ESD] for smaller [<2 cm] tumors); EMR and ESD are also used as part of initial staging to determine more definitively T stage), preoperative treatment including systemic drug therapy is a standard of care option for patients with esophageal and GEJ cancers. For locally advanced adenocarcinomas of the esophagus and GEJ, systemic therapy alone may be indicated: in the ESOPEC trial, preoperative systemic FLOT chemotherapy resulted in improved survival compared to preoperative chemoradiation (using the CROSS regimen). For selected patients with adenocarcinomas (e.g., not able to tolerate regimens such as FLOT), and for patients with SCC of the esophagus, chemoradiation remains a standard of care option; for SCC, chemoradiation may be definitive therapy. If surgery is planned after preoperative chemoradiation, a prolonged delay before surgery has been associated with a poorer outcome. For selected patients with GEJ adenocarcinomas, systemic therapy alone may be given before definitive surgical resection. For patients with SCC of the upper and mid esophagus, combined chemotherapy plus concurrent radiation therapy is a standard option, with surgery reserved for patients not achieving a complete radiographic and endoscopic response. Radiation therapy alone is rarely given with curative intent, as chemotherapy plus concurrent radiation was superior to radiation therapy alone in several clinical trials. Increasingly, all systemic therapy with curative intent is given before operation, although if surgery is the initial therapy and the patient is found to have more locally advanced cancer at pathology (e.g., regional lymph node metastasis), postoperative systemic therapy is used in the adjuvant setting. Adjuvant chemotherapy is more frequently indicated in patients with adenocarcinoma than SCCs. The CheckMate 577 trial demonstrated that for esophageal or GEJ cancer patients who received preoperative chemoradiation therapy and underwent a R0 (negative margins) resection but who had residual tumor in the resection specimen (i.e., not ypT0N0M0), postoperative adjuvant therapy with nivolumab, an immune-modulating agent targeting PD-1, significantly improved survival. For patients with metastatic disease, the goal of therapy is symptom palliation and life extension. No randomized trials of supportive care only versus systemic therapy plus best supportive care have been reported in patients with esophageal cancers. For gastric cancer (an adenocarcinoma histology, as are most distal esophageal and GEJ tumors), clinical trials performed in the 1980s and 1990s indicated a modest improvement in 1- and 2-year survival when systemic therapy plus best supportive care was used versus best supportive care only. While the cytotoxic chemotherapy regimens used for palliation have not changed dramatically over the past 10 years (currently including, for first-line treatment, a platinum compound such as oxaliplatin and a fluorinated pyrimidine such as fluorouracil or capecitabine; the FLOT [fluorouracil, leucovorin, oxaliplatin, and docetaxel] regimen includes docetaxel and is used

TABLE 85-2 Staging System for Gastric Carcinoma[a]

			DATA FROM ACS IN THE UNITED STATES	
STAGE	TNM	FEATURES	NO. OF CASES, %	5-YEAR SURVIVAL, %
0	TisN0M0	Node negative; limited to mucosa	1	90
IA	T1N0M0	Node negative; invasion of lamina propria or submucosa	7	59
IB	T2N0M0 T1N1M0	Node negative; invasion of muscularis propria	10	44
II	T1N2M0 T2N1M0 T3N0M0	Node positive; invasion beyond mucosa but within wall *or* Node negative; extension through wall	17	29
IIIA	T2N2M0 T3N1-2M0	Node positive; invasion of muscularis propria or through wall	21	15
IIIB	T4N0-1M0	Node negative; adherence to surrounding tissue	14	9
IIIC	T4N2-3M0	>3 nodes positive; invasion of serosa or adjacent structures		
	T3N3M0	7 or more positive nodes; penetrates wall without invading serosa or adjacent structures		
IV	T4N2M0 T1-4N0-2M1	Node positive; adherence to surrounding tissue *or* Distant metastases	30	3

[a]AJCC Cancer Staging Manual, Eighth Edition.

Abbreviation: TNM, tumor-node-metastasis.

in very fit patients), subgroups of patients have been identified who benefit from therapies targeting specific genomic alterations (Table 85-3). Approximately 20–25% of patients with adenocarcinoma of the esophagus or GEJ are found to have amplified or overexpressed *HER2/ERBB2*; trastuzumab plus chemotherapy, or trastuzumab, chemotherapy, plus immune therapy using pembrolizumab (for progression-free survival), results in higher response rates and longer progression-free and overall survival compared to chemotherapy alone. For patients whose tumors do not express *HER2/ERBB2*, immune-modulating therapy using PD-1 inhibitors is a standard option as part of first-line therapy (in combination with chemotherapy) and as second-line palliative therapy for patients who have esophageal cancers expressing PD-L1. Patients whose tumors have high PD-L1 combined positive score (CPS) may have greater benefit. For patients with hypermutated or microsatellite-unstable tumors, immune therapy alone is considered, using agents such as pembrolizumab, or nivolumab plus ipilimumab. As noted above, molecular diagnostic or NGS genomic alteration analysis assays to identify these biomarkers should be performed routinely in patients with metastatic esophageal, GEJ, and gastric cancer to help guide therapy.

Supportive measures to improve nutrition and quality of life include placement of an endoluminal stent in the setting of high-grade obstruction; use of enteral nutrition can also be performed using a percutaneous gastrostomy. Photodynamic therapy and endoscopic laser therapy have been used to treat symptoms of endoluminal obstruction.

TABLE 85-3 Molecular Diagnostics and Genomic Alteration Analyses Help Guide Systemic Therapy

SITE HISTOLOGY	FACTOR	ASSAY TYPE	THERAPY OPTIONS
Esophageal SCC	PD-L1	IHC	Immunomodulation
Esophageal Adeno/ GEJ	PD-L1	IHC	Immunomodulation
Esophageal Adeno GEJ	ERRB2/HER2	IHC/FISH/NGS	Trastuzumab; trastuzumab-deruxtecan
Esophageal Adeno GEJ/gastric	dMMRP/MSI	IHC/NGS	Immunomodulation
Gastric Adeno	Claudin 18.2	IHC	Zolbetuximab[a]

[a]Zolbetuximab is currently investigational.

Abbreviations: Adeno, adenocarcinoma; dMMRP, deficient mismatch repair protein; FISH, fluorescent in situ hybridization; IHC, immunohistochemistry; MSI, microsatellite instability; NGS, next-generation sequencing; PD-LI, programmed death ligand 1; SCC, squamous cell carcinoma.

TUMORS OF THE STOMACH

ADENOCARCINOMA OF THE STOMACH

Incidence and Causative Factors Since the 1920s, the incidence of gastric cancer has steadily decreased; while the reason for this has not been definitively identified, it coincided with widespread use of refrigeration and a decreased need for food preservatives. In 2023, it is estimated that there will be 27,500 new cases of gastric cancer diagnosed in the United States; while now seen much less frequently, it remains a lethal disease, with 11,130 deaths. As noted above, globally, gastric cancer is the third most common cause of cancer mortality. High-incidence areas, as is the case for esophageal cancers, include large Asian countries such as China, Korea, and Japan; South American countries such as Chile; and Eastern European countries.

While the number of new cases of body and distal gastric cancers has decreased in Western, high-tier HDI countries, the incidence of adenocarcinomas of the GEJ has markedly increased in the same areas over the past several decades, including as described earlier in younger patients. For gastric cancer overall, a higher incidence in some countries may be due to the ingestion of high concentrations of nitrates found in dried, smoked, and salted foods. Chronic inflammation, caused by an infection, is a major cause of gastric cancer. *Helicobacter pylori* infection is a known driver in many cases of gastric cancer. While *H. pylori* is extremely common, occurring in approximately half of all humans, gastric cancer occurs in only a small subset of those infected. Higher cancer risk has been associated with certain strains of *H. pylori*; while analyses of *H. pylori* are identifying high-risk genomic factors, a specific human genomic profile increasing risk of gastric cancer in the setting of *H. pylori* has not yet been identified. *H. pylori* and ingestion of partially decayed bacterially contaminated food may lead to the generation of carcinogenic nitrites from nitrates. Supportive evidence that *H. pylori* infection is a causative factor in the development of gastric cancer includes prospective studies demonstrating that treatment of *H. pylori* infection decreases the overall risk of gastric cancer. For example, patients with *H. pylori* infection who had at least one first-degree relative with a history of gastric cancer (increasing their own risk of stomach cancer) were randomly assigned to placebo or treatment for *H. pylori*. The group receiving *H. pylori* eradication showed a significant decrease in the incidence of gastric cancer (especially for those in whom *H. pylori* was successfully eradicated) compared to the control group. Earlier studies had demonstrated that treatment of *H. pylori* in Korean patients who had a prior very-early-stage gastric cancer decreased the incidence of a second gastric cancer. These data suggest that treatment of asymptomatic *H. pylori* gastric infection should be considered for patients who have a first-degree relative who has had gastric cancer or who themselves have a prior history of an early-stage gastric cancer. The spectrum of the gastric microbiome beyond *H. pylori* may also be contributory to an increased risk of gastric cancer, with gastric cancer patients having a higher bacterial load than controls.

EBV has also been associated with an increased risk for gastric cancer, and gastric cancer EBV phenotype and genotype have been well described.

Another inflammation-driven cause is suspected to be reflux of gastric contents into the esophagus, particularly in obese people, increasing the risk of GEJ and esophageal adenocarcinomas. Obesity alone is not the cause, as a substantial number of GEJ patients are fit and not overweight.

In addition to chronic inflammatory conditions, inherited cancer susceptibility genes increase the risk of gastric cancer. These include mutations of *CDH1*, which encodes for the cell cohesion gene e-cadherin; germline *CDH1* mutations markedly increase the risk for the diffuse cell (signet cell) gastric cancer subtype (see below for discussion of histologic subtypes). Prophylactic gastrectomy as a risk-reducing surgical procedure is an option for patients with an inherited deleterious *CDH1* mutation. *CDH1* mutations also increase the risk for lobular breast cancer. Germline mutations in the mismatch repair pathway (Lynch syndrome) and in the homologous repair pathway (*BRCA* mutations) slightly increase the risk for gastric cancer. Other inherited cancer susceptibility genetic syndromes that increase the risk of gastric cancer include familial adenomatous polyposis, juvenile polyposis, and Peutz-Jeghers syndrome. Surveillance programs for the early detection of gastric cancer for the higher risk germline cancer susceptibility genes (e.g., *CDH1*) should be employed. In addition, systemic therapy options for patients with certain germline mutations may be different from those with sporadic gastric cancer (e.g., for Lynch syndrome, immune modulation therapy; and for *BRCA*, use of DNA-damaging agents such as platinum compounds).

Gastric cancer stem cells, mostly located in the isthmus of individual gastric glands and possibly originating in the bone marrow, may play an important role in the development of gastric cancer. *H. pylori* may be an inciting factor for recruitment of such bone marrow gastric stem cells. If this hypothesis is confirmed, it may have important implications for therapy of gastric cancers.

Clinical Features

SURVEILLANCE STRATEGIES As is the case for esophageal cancer, the incidence of gastric cancer in Western patients is relatively low, and early detection methodologies such as endoscopy are not routinely employed. The development of a "pan-cancer" blood-based biomarker or a biomarker specific for gastric cancer would allow broader screening. At this time, the overwhelming majority of Western patients with gastric cancer are symptomatic at the time of diagnosis. Early detection programs in Japan and Korea, where gastric cancer has been among the most common of malignancies (although its incidence has been decreasing), include upper endoscopy; these programs have increased the number of patients found with early gastric cancer and decreased mortality rates. This strategy has not been cost effective in populations in which the incidence of gastric cancer is much lower, such as in the United States. As noted above, in high-incidence areas, treatment of symptomatic *H. pylori* is a preventive measure.

PRESENTING SYMPTOMS Presenting symptoms include vague upper abdominal discomfort, hematemesis or melena, anorexia and early satiety, and unexplained weight loss. For patients with esophagogastric junction cancers, dysphagia or odynophagia may be the presenting symptom. Anemia may be found due to occult bleeding. These symptoms and signs lead to upper (and if site of bleeding is uncertain, lower) endoscopy and biopsy. Occasionally, imaging using CT performed to evaluate abdominal symptoms may identify gastric thickening or a gastric mass leading to upper endoscopy. Physical examination can reveal left supraclavicular adenopathy (Virchow's node), a periumbilical mass (Sister Mary Joseph nodule), a pelvic mass on rectal exam (Blumer's shelf), ascites, or an ovarian mass (Krukenberg tumor). More commonly, physical examination is unrevealing.

Upper endoscopy may reveal an ulcer or ulcerated mass, biopsy of which shows adenocarcinoma. For the diffuse subtype of gastric cancer, a mass or ulceration may not be seen, but rather, thickened gastric rugae may be noted. Initial biopsy may not reveal diffuse gastric cancer, which may track below the mucosal surface. In these patients, EUS may guide biopsy.

Histopathology Classification of Primary Gastric Adenocarcinomas

The large majority (~85%) of gastric malignancies are adenocarcinomas or subtypes of adenocarcinoma. Other malignancies, discussed below, include neuroendocrine tumors (carcinoid tumors), primary gastric lymphomas, gastrointestinal stromal tumors (GISTs), and other rare malignancies. Using the Lauren classification, pathologists classify adenocarcinomas on the basis of histopathology as intestinal (more common) or diffuse subtype (~20%). As noted above, the diffuse subtype is associated with inherited *CDH1* mutations; in addition, in the TCGA genomic analysis of gastric cancer and in other genomic analysis studies, approximately a third of diffuse subtype cases had somatic *CDH1* mutations. The intestinal subtype is associated with *H. pylori* infection and atopic gastritis. Histologic grade also influences the clinical course.

Genomic analysis performed by several groups has resulted in molecular classifications of gastric cancer that may, in the future,

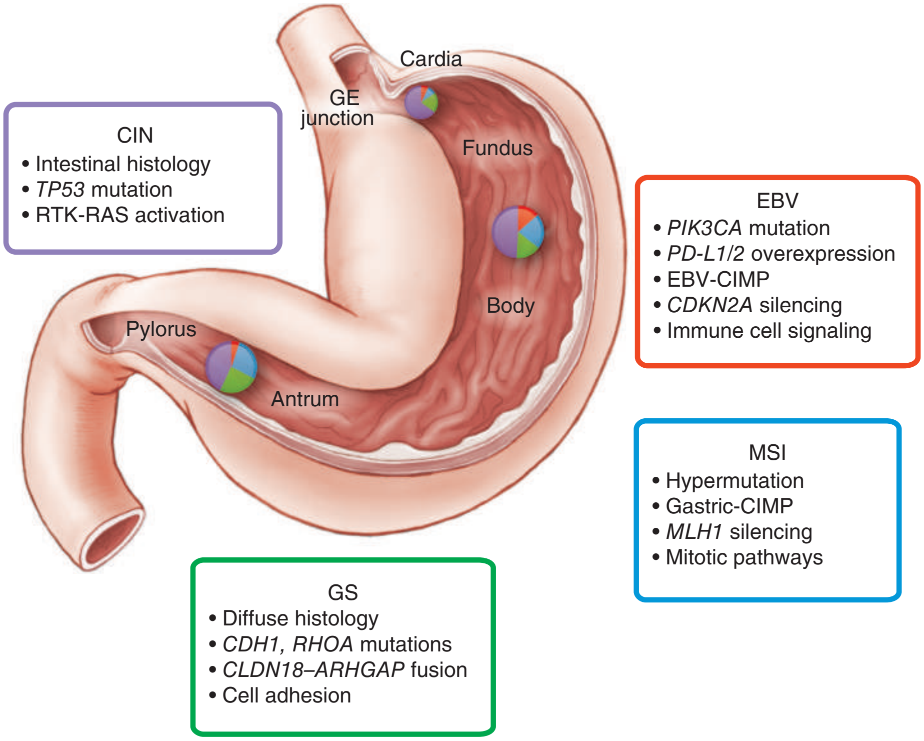

FIGURE 85-2 Molecular/genomic characterization of subtypes of gastric carcinomas. CIMP, CpG-island methylator phenotype; CIN, chromosomally unstable; EBV, Epstein-Barr virus-associated; GE, gastroesophageal; GS, genomically stable; MSI, microsatellite instability-associated.

inform staging systems; they already provide a better understanding of the driving factors in the development of gastric cancer and important information on treatment options as outlined above **(Fig. 85-2)**. For example, the TCGA group reported the results of a multiplatform analysis of 295 patients with previously untreated gastric cancer; both Western and Asian patients were included in the analysis. Four subtypes of gastric cancer were identified: high EBV burden, microsatellite unstable with hypermutation, genomically stable (associated with the diffuse subtype), and chromosomal unstable. The Asian Cancer Research Group (ACRG), studying primary tumors from 300 Korean patients, analyzed gene expression profiles and found four subtypes: mesenchymal, microsatellite unstable, microsatellite stable with *TP53* expressed, and microsatellite unstable with *TP53* mutated. Clinical outcome was correlated with genomic subtype in both studies, with microsatellite unstable tumors having the best outcome (in an era before use of immune modulation therapy) and genomically stable (TCGA) and mesenchymal (ACRG) types having the worst outcome.

In addition to histopathology, as is the case for esophageal cancer, molecular diagnostics and genomic alteration analysis using NGS are an important part of the pathology workup. The molecular subtypes have therapeutic implications; for example, as discussed above, ~20% of gastric cancer or GEJ cancer patients' tumors have overexpression or amplification of *HER2/ERBB2*, which would lead to the addition of agents such as trastuzumab-based therapy as part of systemic treatment for metastatic disease. Immune modulation therapy would be used in patients with hypermutated tumors, found by NGS (tumor mutation burden) or by polymerase chain reaction (PCR) for microsatellite instability (MSI). An evaluation for overexpression or amplification of *HER2/ERBB2*, quantification of PD-L1 by IHC, and assessment of MSI by PCR or deficient mismatch repair protein (dMMRP) expression should be a routine part of the pathology workup of patients with metastatic gastric cancer. Claudin 18.2 has recently been identified as a target for therapy; if agents targeting Claudin 18.2 are approved by the U.S. Food and Drug Administration (FDA), an assay for this factor should be included in the pathologic evaluation of metastatic gastric cancer. More controversial is whether these assays should also be routinely performed in patients with potentially operable gastric cancer because, for example, the addition of trastuzumab to neoadjuvant chemotherapy has not yet been shown to change outcome. In large-volume centers, NGS is routinely performed on pretreatment biopsies. Currently, the finding on pathologic assays of positive tumor EBV (identified in 8–10% of gastric cancer patients) does not change therapeutic options.

Staging Once a diagnosis of a primary gastric adenocarcinoma is made, algorithms for clinical evaluation of stage include physical examination and imaging studies (**Fig. 85-3**; Table 85-2). Tumor-related biomarkers such as carcinoembryonic antigen (CEA) or CA19-9 may be elevated but are nonspecific (may be elevated in a number of other gastrointestinal and other site cancers). Diagnostic CT scan of the chest, abdomen, and pelvis should be performed. If metastatic disease is suspected on imaging, a biopsy of a metastatic site should be strongly considered to confirm stage IV disease, which changes the goals of care from potentially curative to palliative treatment and provides tissue for molecular and genomic analyses. As is the case for esophageal cancer, FDG-PET/CT, which is more sensitive than diagnostic CT scan in identifying sites of metastatic disease, should be performed if the anatomic CT is negative for metastatic disease. Note, however, that FDG-PET may be noninformative (the primary tumor may not be FDG-avid, particularly in diffuse-type gastric cancer). If imaging does not reveal metastatic disease, EUS should be considered to determine depth of penetration of the primary tumor and the presence or absence of regional lymphadenopathy suspicious for metastasis. Lymph node biopsy and, on occasion, biopsy of left hepatic parenchymal lesions found on EUS can be performed at the same setting. Endoscopic biopsies usually provide enough tumor tissue for molecular diagnostic pathology testing for *HER2/ERBB2*, MSI/MMRP, and PD-L1 assessment; it may not provide enough tissue for NGS genomic alteration analysis. If neoadjuvant therapy is planned, laparoscopy should be considered to allow evaluation of the peritoneal cavity, with peritoneal washing for cytology if no peritoneal metastases are visible. The peritoneal cavity is a common site of metastases, especially from diffuse-type gastric cancer. The finding of peritoneal involvement either visibly or by positive cytology is staged as metastatic disease and generally rules out the chance for curative resection.

The AJCC staging classification for gastric cancer is summarized in Table 85-2. Three staging classifications are provided: cTNM clinical staging (before any therapy has been given), pTNM pathologic staging (for patients not undergoing preoperative therapy), and a post–neoadjuvant therapy classification staging (ypTNM). The three components take into account current standard-of-care options for therapy in which the AJCC prognostic stage groups from clinical staging guide therapeutic decisions. For example, after clinical evaluation, a large percentage of newly diagnosed patients will be found to have higher-stage primary cancers (penetrating through the gastric wall [T3 or T4] or lymph node–positive tumors), in which case perioperative (neoadjuvant) systemic therapy may be chosen. Pathologic examination of the resected specimen for prognostic stage classification must take into account exposure to preoperative therapies that may lead to downstaging (thus, ypTNM staging). Nomograms have been developed for predicting outcome in patients undergoing surgery as initial treatment.

TREATMENT

Gastric Cancer

POTENTIALLY CURABLE GASTRIC CANCER: SURGERY

Surgical removal of the primary tumor with negative microscopic margins (an R0 resection) and with resection of regional lymph nodes is currently the only curative therapy; with surgery alone, overall 5-year survival rates are approximately 25% (higher for very-early-stage tumors). If tumor cells are found at the margin of resection (R1) or if visible cancer is left at the time of surgical removal of the primary tumor (R2), surgery is palliative rather than curative. For patients with early-stage tumors (mostly clinical stage I), surgery without perioperative systemic therapy may be performed. For patients with more locally advanced tumors (clinical stages IIA, IIB, III), who compose approximately 70% of newly diagnosed operable patients, multimodality therapy (surgery and systemic chemotherapy) improves overall survival. Both neoadjuvant (preoperative) and postoperative systemic therapy are accepted approaches. If staging studies demonstrate a locally advanced cancer (or node positive), preoperative treatment is recommended in a medically fit patient. If surgery is performed first and a locally advanced cancer is found, postoperative chemotherapy or chemotherapy plus chemoradiation is recommended. For selected very-early-stage gastric cancers (primary tumors that are ≤2 cm in diameter, are well to moderately differentiated, do not invade the deep submucosa [T1], and do not show lymphovascular invasion or lymph node metastasis), which are not commonly found in the United States, EMR or ESD may be performed by experienced gastroenterologists in place of surgical resection, with favorable results in studies in high-incidence areas such as Japan.

For patients in whom the primary tumor is in the distal stomach, a subtotal gastrectomy is the preferred surgical procedure. For tumors of the proximal stomach, the options for resection include total gastrectomy or, alternatively, proximal gastrectomy. Esophagogastrectomy is performed for tumors involving the GEJ. In selected patients, a jejunostomy feeding tube may be placed if postoperative radiation therapy is part of the treatment plan.

As noted above, laparoscopy is commonly performed at high-volume centers before a final decision regarding the role of surgery. If staging has already demonstrated clinically suspicious lymph nodes or an advanced T stage tumor, but laparoscopy does not demonstrate peritoneal metastasis, perioperative chemotherapy is given before surgical resection.

Palliative resection of the primary tumor is usually performed only if symptoms such as uncontrollable bleeding or obstruction are present that cannot be relieved by other means.

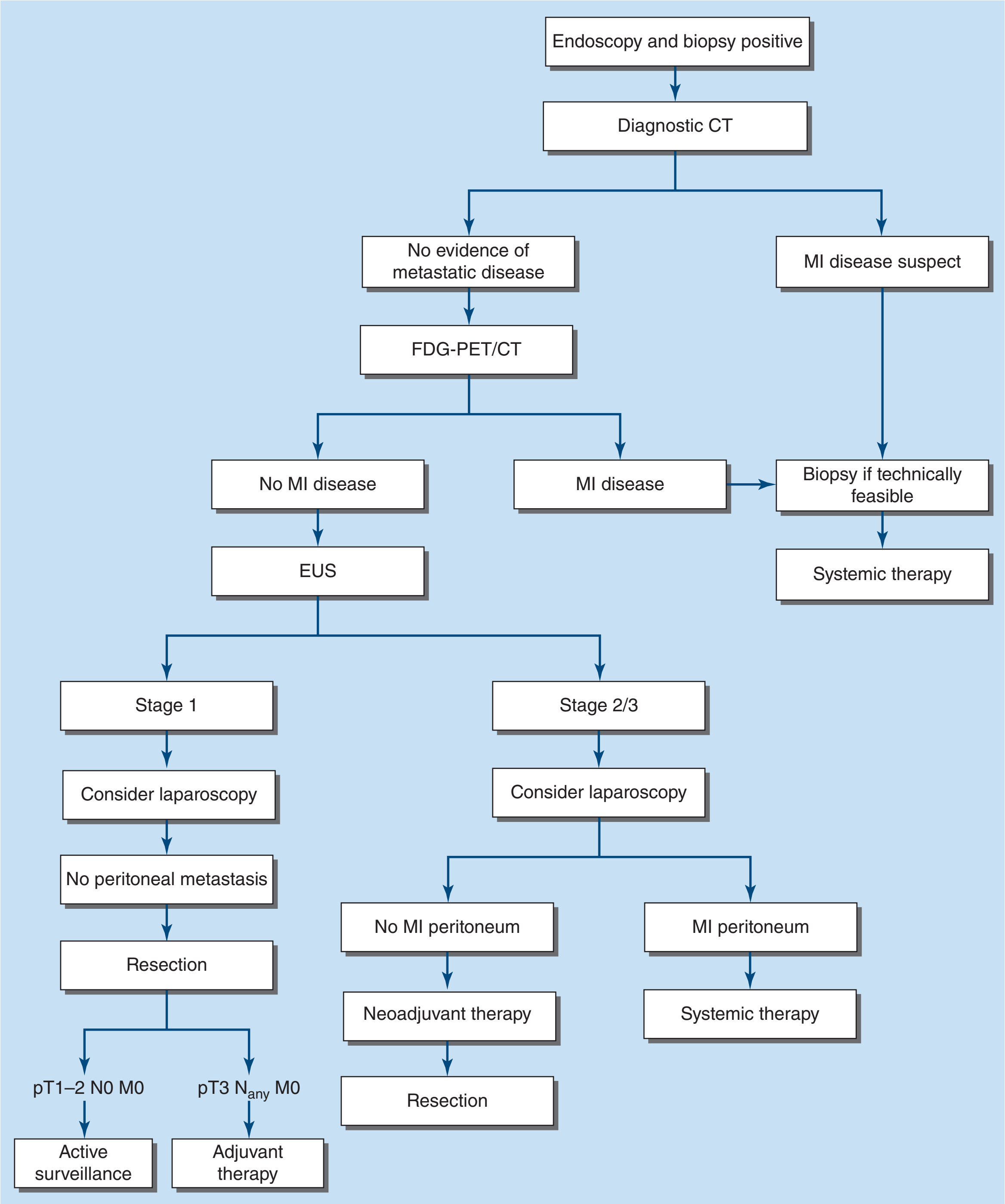
Endoscopy and biopsy positive
Diagnostic CT
No evidence of metastatic disease
MI disease suspect
FDG-PET/CT
No MI disease
MI disease
Biopsy if technically feasible
EUS
Systemic therapy
Stage 1
Stage 2/3
Consider laparoscopy
Consider laparoscopy
No peritoneal metastasis
No MI peritoneum
MI peritoneum
Resection
Neoadjuvant therapy
Systemic therapy
pT1–2 N0 M0
pT3 N_{any} M0
Resection
Active surveillance
Adjuvant therapy

FIGURE 85-3 Staging for gastric adenocarcinoma. CT, computed tomography; EUS, endoscopic ultrasound; FDG-PET, fluorodeoxyglucose positron emission tomography.

As is the case for colorectal cancer, a correlation exists between the number of lymph nodes removed and sampled and outcome. Sentinel lymph node biopsy is not performed in gastric cancer outside of a research study setting. The goal is to examine at least 15 lymph nodes from the resected specimen; it is more controversial whether more extensive lymph node resection itself affects outcome; the extent of lymphadenopathy can be classified using a D0–D3 system with a higher number meaning more extensive lymphadenopathy. In the United States, a modified D2 (D1+) resection preserving the spleen and avoiding pancreatectomy is recommended but should be performed by experienced surgeons at high-volume centers. Japanese investigators and others have used very extensive lymph node dissections, but studies have not demonstrated an advantage for a D3 resection. Both resection of the primary tumor and its regional lymph nodes can be performed laparoscopically in appropriate patients.

In the hands of experienced surgeons, operative mortality would be anticipated to be ≤2%.

NEOADJUVANT AND POSTOPERATIVE ADJUVANT THERAPY FOR RESECTABLE GASTRIC CANCER

The large majority of potentially resectable Western gastric cancer patients have locally advanced tumors (cTNM stage IIA/B or III). Multimodality therapy using systemic chemotherapy plus surgery improves 5-year survival rates by 10–15% compared to surgery alone. An older but widely cited study, the MAGIC clinical trial, randomly assigned patients with potentially resectable disease to receive perioperative chemotherapy or to proceed directly to surgery. Five-year overall survival for patients undergoing surgery alone was 23%; for those receiving pre- and postoperative chemotherapy, it was 36%. On the basis of this and other clinical trials, for most medically fit patients with stage cTNM II and III resectable gastric cancers, preoperative systemic chemotherapy followed by resection and, if tolerable, postoperative chemotherapy is a standard approach. Because it is challenging to give postgastrectomy systemic therapy using currently employed regimens (<50% of patients receive or complete postoperative treatment), total neoadjuvant therapy is favored at some centers. For patients with MSI/hypermutated tumors, immune modulation therapy (PD-1/PD-L1 or PD-1 plus CTLA-4 agents) can be considered in place of chemotherapy, although substantial pathologic response rates were seen in a small number of patients in the DANTE trial using FLOT alone. Preoperative chemoradiation as given for esophageal cancers is usually used for GEJ tumors. If an R1 or R2 resection is performed, postoperative chemoradiation can be considered. Use of postoperative chemoradiation therapy, with exceptions outlined above, did not improve outcome when compared to systemic chemotherapy alone. Preoperative chemotherapy followed by chemoradiation is under study. For patients being treated with multimodality therapy, close interactions among the surgeon, medical oncologist, and radiation oncologist are essential.

Clinical trials have compared different preoperative cytotoxic chemotherapy regimens, most of which include a platinum compound—either cisplatin or oxaliplatin. Currently, a platinum compound plus a fluorinated pyrimidine, such as fluorouracil or capecitabine, given for three to four cycles before surgery is a standard-of-care option. Drug combinations are favored; for very fit patients, a combination of FLOT may be chosen. An alternative is, for example, the FOLFOX regimen, which includes fluorouracil, leucovorin, and oxaliplatin. Addition of trastuzumab to preoperative chemotherapy has not improved outcomes for patients with *HER2*-positive cancers. Careful monitoring of chemotherapy-related toxicities with appropriate dose modifications is important. For patients receiving preoperative systemic chemotherapy and undergoing an R0 D2/D1+ dissection, postoperative chemoradiation therapy has not improved outcome.

For patients in whom the primary tumor has been resected and who did not receive preoperative chemotherapy, who are found to have stage II or III cancers, or who have <15 lymph nodes found in the resected specimen (less than D2 resection), postoperative chemotherapy or chemoradiation is a treatment option. Chemotherapy and chemoradiation therapy may also be given for unresectable cancers in selected patients. As noted above, plasma ctDNA as a measure of minimal residual disease (regional or metastatic cancer not identifiable by current imaging modalities) is being studied. Data from other gastrointestinal tumors suggest that conversion of a positive ctDNA assay to negative after systemic postoperative therapy is associated with a better outcome.

PALLIATIVE THERAPY FOR INCURABLE GASTRIC CANCER

Patients with clinical stage IV gastric cancers with an adequate performance status should be offered systemic drug therapies. Small clinical trials performed in the 1980s and 1990s showed a survival benefit for systemic therapy compared to best supportive care only. As discussed above, genomic and molecular diagnostic analysis has identified several distinct cohorts of gastric cancer patients. These analyses guide therapy. While the cytotoxic chemotherapy regimens most commonly employed are still based on a platinum compound and a fluorinated pyrimidine (e.g., FOLFOX, as is used in the perioperative setting), a major advance has been the demonstration that immune modulation therapy plus chemotherapy improves outcome (including overall survival) compared to chemotherapy alone. In the CheckMate 649 trial, 1581 patients with *HER2/ERBB2*-negative advanced gastric or GEJ cancers were randomly assigned to receive nivolumab plus chemotherapy (oxaliplatin plus capecitabine or FOLFOX) versus chemotherapy alone. Overall survival, progression-free survival, and response rates were significantly better for patients receiving nivolumab plus chemotherapy; this effect was most marked for patients with PD-L1 CPS of >5. CheckMate 649 and other data have led to guidelines recommending use of immune modulation therapy plus chemotherapy as an option for first-line treatment for patients with *HER2/EBB2*-nonexpressing, PD-L1 CPS >5 metastatic gastric and gastroesophageal cancers; chemotherapy plus immune modulation therapy may also have benefit for patients with lower PD-L1 CPS scores. Those whose tumors have overexpressed or amplified *HER2* should receive HER2-targeted agents such as trastuzumab or trastuzumab, pembrolizumab, plus cytotoxic chemotherapy. Beyond first-line therapy, additional HER2-targeted therapy using agents such as trastuzumab-deruxtecan, a monoclonal antibody-drug conjugate, is recommended for patients whose tumors have amplification of *HER2/ERBB2*. For patients with metastatic MSI/dMMRP gastric cancers, immune modulation therapy using either PD-1 inhibitors such as pembrolizumab or combinations such as nivolumab plus ipilimumab may be employed as first-line therapy; guidelines include the option of immune therapy plus chemotherapy.

When disease progresses after first-line treatment, other therapies include the combination of a VEGF receptor–targeted agent, ramucirumab, either alone or in combination with paclitaxel. Immune modulation inhibitors are an option for second-line therapy for patients whose tumors are microsatellite unstable (e.g., nivolumab plus ipilimumab if single-agent pembrolizumab was used as first-line therapy). For patients with microsatellite stable gastric cancers (the large majority), immune modulators are an option if not used in the first-line setting, if PD-L1 is positive. As above, for patients with *HER2/ERBB2*-amplified tumors, trastuzumab-deruxtecan can be used. Several other cytotoxic agents have activity in the third-line setting including trifluridine-tipiracil and irinotecan. Even with improvements in therapy including immune modulation therapy, the best results from clinical trials indicate overall survival for treated patients with stage IV disease is still only 12–15 months. Investigational studies include targeting Claudin 18.2, a cell surface type junction molecule that is overexpressed in ~50% of gastric cancers. In the SPOTLIGHT trial, previously untreated, *HER2/ERBB2*-negative patients assigned to zolbetuximab (targeting Claudin 18.2) plus FOLFOX chemotherapy had an improvement in survival compared to those receiving FOLFOX chemotherapy alone. Zolbetuximab plus chemotherapy may become another first-line option for advanced gastric cancer. Additional studies targeting Claudin 18.2 including those using chimeric antigen receptor (CAR) T cells are underway.

Radiation therapy using shorter regimens may be employed to palliate bleeding. For patients with advanced incurable disease, other supportive measures include placement of a duodenal stent to relieve gastric outlet obstruction; in selected patients, surgical procedures for gastric outlet obstruction may be performed. Radiation therapy might be used if not previously given. Enteral feeding using a jejunostomy tube may support nutritional needs.

GASTRIC LYMPHOMAS

Lymphomas of the stomach are an uncommon (~3%) but important subgroup of gastric malignancies. They are extranodal non-Hodgkin's lymphomas (NHL). The gastrointestinal tract is the most common site for extranodal NHL, and the stomach is the most common site within the gastrointestinal tract. The presenting symptoms can be similar to those of the much more common adenocarcinoma of the stomach described above, including pain, anorexia, and bleeding. Symptoms of fever and night sweats occur in 10–15% of patients with gastric NHL. Because the treatment options are so different, obtaining adequate tissue for definitive pathologic examination is crucial in diagnosing both gastric lymphomas (as opposed to adenocarcinoma) as well as defining the subtype of lymphoma. On occasion, this may be challenging because, similar to diffuse subtype adenocarcinoma, lymphomas may track below the mucosal surface. A fine-needle aspirate may not be sufficient; multiple deep biopsies or mucosal resection may be needed to provide enough tissue for definitive pathologic assessment.

Potential driving forces in the development of gastric lymphomas include active or prior *H. pylori* infection, which is associated with mucosa-associated lymphoid tissue (MALT) subtype gastric lymphomas and has been described as a potential driver in some patients with diffuse large B-cell lymphomas (DLBCLs). MALT lymphomas may develop in nearly any organ, but the stomach is the most frequent primary site, accounting for ~35% of all MALT lymphomas. Identifying a MALT lymphoma is important because antibacterial therapy directed against *H. pylori* alone can be a highly effective treatment. Other infections should be considered either as driving factors or as risks from treatment, such as reactivation of hepatitis B virus (HBV); testing for HBV is recommended before initiation of lymphoma therapy. Hepatitis C infection has been associated with DLBCL and with marginal zone lymphomas. Other forms of NHL may involve the stomach either as primary gastric lymphoma or as a secondary site of disease, including both B-cell (e.g., mantle cell lymphoma, Burkitt's lymphoma, and follicular lymphoma) and T-cell lymphomas (e.g., enteropathy-associated T-cell lymphoma, anaplastic large-cell lymphoma, and peripheral T-cell lymphoma).

Staging is performed using similar imaging modalities as for gastric adenocarcinoma, but the staging systems are different (see below, and presented in more detail in the chapter on lymphomas **[see Chap. 113]**). As a staging study, FDG-PET/CT is performed, as well as a contrast-enhanced, diagnostic-quality CT. FDG-PET/CT is used both as an initial staging study and to assess response. EUS may be used to determine depth of invasion in patients in whom no evidence of metastatic disease is noted. Examination of the peripheral blood and bone marrow aspirate should be considered as part of the workup. In all patients with gastric lymphoma, *H. pylori* infection status should be evaluated. If *H. pylori* testing is negative by histopathology, noninvasive testing by either stool antigen test or urea breath test should be used.

■ STAGING

The TNM staging system is not employed for gastric lymphomas. Several classification systems are used for lymphomas, including the Lugano staging system (a modification of Ann Arbor staging), the revised World Health Organization (WHO) Fifth Edition of the WHO Classification of Haematolymphoid Tumours (WHOHAEM5), and the International Consensus Classification (ICC). These are discussed in **Chap. 113** on lymphomas. The Lugano staging system for gastrointestinal lymphomas is still widely used; it divides patient groups into stages I, II, III, and IV. Stage I tumors are limited to the gastric wall; stage II tumors have regional lymph node involvement or invasion of local structures; stage III tumors have lymph node involvement above and below the diaphragm. Stage IV tumors have either more extensive lymph node involvement or have distant metastasis, including to the bone marrow or other extranodal sites.

■ PATHOLOGIC CLASSIFICATION

The two most common histologic subtypes of gastric lymphoma are marginal zone B-cell lymphomas (gastric marginal zone B-cell lymphomas or MALT; ~40% of newly diagnosed patients) and DLBCLs (~55%). The distinction is critical because therapeutic options are different. The large majority of MALT cases are associated with *H. pylori* infection. The finding of a t(11;18) translocation identifies a subgroup less likely to respond to *H. pylori* eradication (or in non-*H. pylori*–associated MALT). This translocation may be detected using PCR or fluorescent in situ hybridization (FISH); it creates a chimeric protein composed of the amino terminal of *API1* (apoptosis inhibitor) and the carboxy terminus of *MALT1*, leading to activation of nuclear factor-κB signaling. Note that even *H. pylori*–negative MALT tumors may respond to anti–*H. pylori* therapy. Other, rarer subtypes of primary gastric lymphomas include those of T-cell lineage.

TREATMENT

Gastric Lymphoma

Unlike adenocarcinoma of the stomach, surgical resection has no role in the treatment of primary gastric lymphoma in the absence of complications of therapy such as perforation or uncontrollable bleeding. Resection of gastric lymphoma does not improve clinical outcomes.

For patients with MALT lymphoma, eradication of *H. pylori* with antibiotics is highly effective therapy. If tests for *H. pylori* are positive and t(11;18) translocation assay is negative, one of the currently accepted antibiotic regimens for treating *H. pylori* should be the initial therapy. *H. pylori* eradication is associated with high response rates including complete remissions in the majority of patients. Treatment with an anti–*H. pylori* regimen may induce remission even in patients in whom standard testing for *H. pylori*, described above, is negative. The time to remission may be prolonged (in some studies averaging 15–16 months); therefore, careful monitoring is important before determining that a MALT tumor is not responding to anti–*H. pylori* therapy. For patients in whom the t(11;18) translocation assay is positive, options for therapy include anti–*H. pylori* antibiotic therapy plus involved-field radiation therapy or, if radiation is contraindicated, the use of single-agent rituximab, a monoclonal antibody targeting CD20. For patients who are *H. pylori* negative, in whom anti–*H. pylori* therapy may have been used without response, moderate-dose (24–30 Gy) involved-site radiation therapy or single-agent rituximab is a treatment option. For selected, more advanced, stage IV MALT patients who have not responded to or who have progressed after receiving anti–*H. pylori* antibiotic therapy and/or rituximab, cytotoxic chemotherapy regimens such as R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisolone), dose-adjusted R-EPOCH (an infusional regimen of rituximab, etoposide, prednisolone, vincristine, cyclophosphamide, and doxorubicin), or rituximab and lenalidomide may be considered.

DLBCL may be a result of transformation from more indolent MALT lymphoma or may arise de novo. De novo tumors are more likely to be BCL2 and CD10 positive. MALT lymphomas that have transformed to DLBCLs are more frequently BCL2 and CD10 negative.

For patients with DLBCL, earlier-stage tumors may be treated by combination chemotherapy alone or chemotherapy plus involved-field radiation therapy. For more advanced gastric DLBCL tumors, chemotherapy using the R-CHOP or R-EPOCH regimen is standard therapy. Some reports have suggested that eradication of *H. pylori* is effective treatment for early-stage DLBCL when the patient also has *H. pylori*.

UNCOMMON TUMORS OF THE ESOPHAGUS AND STOMACH

■ GASTROINTESTINAL STROMAL TUMORS

GISTs are rare tumors of the gastrointestinal tract associated with activating somatic mutations in the *cKIT* (the majority) or *PDGFRA* genes; in a minority of cases, neither gene is mutated. GISTs arise from Cajal cells, which bridge between the autonomic nerves to the muscle layer of the bowel. The stomach is the most frequent primary site (~60%), followed by the small bowel in about 25%. As endoscopy for other indications has become more widely used, otherwise asymptomatic and probably clinically insignificant small GISTs have been identified more frequently; it is not clear that the actual incidence has substantially increased. Symptoms associated with GISTs include acute gastrointestinal bleeding leading to melena and/or hematemesis. Anemia may be reflected in generalized weakness. With larger tumors, abdominal distention and pain may be presenting symptoms. At endoscopy, a nonspecific smooth bulging mass covered by normal mucosa is the most frequent finding. Initial biopsy may not reveal a mesenchymal neoplasm. The radiologic features found on diagnostic CT scan (which may identify an asymptomatic tumor) may raise the question of a GIST tumor. EUS can be performed, but if biopsy is done, it may not provide an adequately cellular specimen to allow genomic sequencing, which is required to help guide systemic therapy (if needed).

Histologically, a spindle cell neoplasm is the most common subtype (~70%), with epithelioid cells making up 20%; 10% of cases are mixed histology. KIT expression is found in ~95% and *PDGFRA* in ~80% of cases. NGS is usually used to assess for gain-of-function *KIT, PDGFRA*, or other, rarer mutations.

For nonmetastatic GIST tumors, risk of recurrence assessment (guiding use of postoperative therapy) is based on the size of the primary tumor, mitotic index, and location in the gastrointestinal tract. For gastric GISTs, resected tumors ≤2 cm with a mitotic index of ≤5% have a near 0 risk of progression. For larger tumors, especially those ≥10 cm, with a mitotic index >5%, the risk is moderate to high. Postoperative adjuvant imatinib for 3 years is a standard-of-care option. Avapritinib is used for tumors with certain *PDGFRA* mutations. For patients with metastatic disease, imatinib (for *KIT* mutation–bearing tumors) or avapritinib (targeting *PDGFRA*) is used; the location of the *KIT* mutation (exon 11 vs 9) is associated with degree of effectiveness. However, resistance almost invariably develops, and the development of newer agents effective in tumors with secondary mutations is a high priority; ripretinib, which targets both *KIT* and *PDGFRA*, is now also FDA approved.

SMALL-BOWEL NEOPLASMS

Small-bowel neoplasms make up ~3% of gastrointestinal tumors; in 2023, 12,000 new cases are expected in the United States. The spectrum of malignant small-bowel neoplasms includes neuroendocrine tumors (NETs; carcinoid), adenocarcinomas, lymphomas, and GISTs. In the United States, NETs, which have increased in incidence, are slightly more frequent (40–45%) than adenocarcinomas (30–40%), with the remainder mostly lymphomas and ~8% GISTs. The duodenum is the most common portion of the small bowel in which malignancies develop (~50%), with ~30% occurring in the jejunum and 20% in the ileum. NETs are the most common benign and malignant tumors of the ileum. Risk factors for the development of small-bowel adenocarcinoma include inflammatory bowel disease (Crohn's disease); inherited germline mutation syndromes such as Lynch syndrome, familial adenomatous polyposis (FAP), and Peutz-Jeghers syndrome; and celiac disease (which is also associated with an increased risk for lymphomas).

While an asymptomatic small-bowel primary adenocarcinoma might be found during surveillance in patients at high risk (e.g., FAP), for many small-bowel tumors, the finding of positive stool occult blood, unexplained anemia, or small-bowel obstruction leads to the diagnosis. Both adenocarcinoma and lymphomas might present with perforation. Evaluation by diagnostic CT imaging may reveal a small-bowel lesion. CT and/or magnetic resonance enterography (CTE or MRE) have a high positive predictive value (for adenocarcinoma, MRE appears to be more accurate). A variety of endoscopic techniques are used for diagnostic evaluation of the small bowel, including device-assisted enteroscopy (DAE), e.g., double-balloon enteroscopy; small-bowel capsule endoscopy (SBCE) may also be employed. DAE allows tissue diagnosis; a tattoo helps localize the small bowel neoplasm (SBN) for resection; DAE may also allow stent placement for palliation of obstruction. SBCE is contraindicated in the setting of obstruction. As is the case for gastric imaging studies, deep learning artificial intelligence is being studied to improve the analysis of small-bowel imaging. For NETs, a gallium-68 or copper-64 DOTATATE scan may identify both the primary site as well as metastatic disease. Blood tumor biomarkers are nonspecific for the primary site (e.g., CEA or CA19-9 for adenocarcinoma); these assays are better used to monitor response or progression of disease rather than for diagnosis.

The median age at diagnosis for sporadic small-bowel adenocarcinoma is in the seventh or eighth decade of life, but genetically predisposed patients and those with inflammatory bowel disease may be diagnosed at a much earlier age. African Americans have a higher incidence of small-bowel cancer than whites. While systemic therapies are usually modeled on agents used to treat colorectal cancer, genomic analyses have indicated that small-bowel adenocarcinoma has distinct genomic alterations compared to either colorectal or gastric cancers. Genomic alterations less frequent in small-bowel than in colorectal cancers include *TP53*, *BRAF* V600E, and *APC* mutations, whereas the rate of *KRAS* mutations is similar to colorectal cancer (a genomic analysis of Chinese small-bowel cancer patients, mostly duodenal cancers, noted a somewhat different profile than U.S. patients). Within small-bowel sites, the most striking difference is the higher rate of *ERBB2* alterations (of which a minority are amplifications) in duodenal cancers. Not surprisingly, because Lynch syndrome increases the risk of small-bowel adenocarcinoma, 15–20% of these tumors are MSI high or mismatch repair deficient; small-bowel adenocarcinoma associated with celiac disease also may have an increased rate of MSI-high tumors. MSI/MMRP status should be assessed in all patients with small-bowel adenocarcinoma. Somatic tumor genomic analysis may suggest a germline mutation, but appropriate genetic testing for a germline driver mutation should be performed in all patients with small-bowel adenocarcinoma.

Small-bowel adenocarcinoma has its own staging classification in the eighth edition of the AJCC Cancer Staging Manual.

TREATMENT

Small-Bowel Adenocarcinomas

Surgical resection with negative microscopic margins (R0), as is the case for other gastrointestinal tumors, is the best chance for cure. For duodenal adenocarcinomas, a Whipple procedure may be needed; for more distal duodenal cancers and jejunal adenocarcinomas, a segmental resection with adequate margins should be performed. Distal ileal tumors may require right hemicolectomy.

Small-bowel cancers are frequently found with locally advanced disease at the time of diagnosis. If the tumor is resectable, postoperative adjuvant systemic therapy is currently recommended for lymph node–positive patients, using regimens such as those employed for colorectal cancer (capecitabine and oxaliplatin or FOLFOX). Benefit from adjuvant therapy has not yet been proven. Small-bowel cancers developing in patients with Lynch syndrome probably have a better prognosis; if colorectal cancer is a model, postoperative adjuvant therapy for patients with Lynch syndrome may include consideration of immune modulation therapy. For duodenal cancers, chemoradiation is considered if the resection margins are still positive.

For patients with advanced metastatic disease, in the absence of Lynch syndrome or a hypermutated tumor, similar cytotoxic regimens as deployed for gastric (duodenal) or colon cancers (jejunum or ileum) have been widely used. For tumors that are MSI high or dMMRP positive, immunotherapy is indicated; for tumors that are *HER2* amplified or *BRAF* mutated, targeted therapy may be useful.

SMALL-BOWEL GASTROINTESTINAL STROMAL TUMORS

Like small-bowel adenocarcinomas, small-bowel GISTs may present with obstruction or bleeding. Diagnostic techniques are those employed for other small-bowel neoplasms. While the pathological criteria for malignant potential are somewhat different than those used for gastric GISTs, postoperative management and treatment of metastatic disease are the same as those described above for gastric GIST.

BENIGN NEOPLASMS OF THE SMALL BOWEL

As is the case for malignant small-bowel tumors, benign neoplasms of the small bowel are rare. In addition to cancers, the precursor lesion adenomas or hamartomas (from which cancers develop) may be driven by inherited cancer susceptibility genes (e.g., FAP, Lynch syndrome, Peutz-Jeghers syndrome). Other benign neoplasms include lipomas, leiomyomas, neurofibromas, and benign lymphoid nodular hyperplasia.

Patients with benign small-bowel neoplasms not associated with an inherited cancer susceptibility gene (in which case, a benign tumor may be found during surveillance) are usually asymptomatic. A mass may be noted on an imaging study (usually a CT) ordered for another reason. Workup for occult or overt bleeding or intussusception of the bowel may lead to the discovery of a benign small-bowel neoplasm.

Diagnostic evaluation is similar to that described above for malignant tumors. In general, benign neoplasms, if found during surveillance, are removed endoscopically, if technically feasible, to decrease the risk of intussusception in Peutz-Jeghers syndrome. Mucosectomy may be used to treat bleeding hemangiomas.

NEUROENDOCRINE TUMORS OF THE ESOPHAGUS, STOMACH, AND SMALL BOWEL

Neuroendocrine neoplasms include NETs (more well differentiated, classified as grades 1–3) and neuroendocrine carcinomas (NECs). Neuroendocrine neoplasms including pancreas NETs are described in detail in **Chap. 89**. The following describes NET and NEC of the esophagus, stomach, and small bowel.

■ ESOPHAGEAL AND GASTRIC NEUROENDOCRINE TUMORS

NETs of the esophagus are rare, accounting for <1% of gastrointestinal NETs. Presenting symptoms are similar to those of SCC or adenocarcinoma of the esophagus, with dysphagia or odynophagia or with more nonspecific symptoms such as substernal discomfort or burning consistent with reflux esophagitis. A potential driving factor of NEC is smoking. The initial diagnostic evaluation includes upper endoscopy and biopsy. Pathology may reveal a well-differentiated NET (grades 1–3; grade guided by percent Ki-67 positivity) with a low to relatively low metastatic potential; at the other end of the spectrum are small-cell or large-cell NECs, which are fully malignant and frequently metastasize. In the absence of metastatic disease, EUS to assess depth of penetration and presence or absence of regional lymph node metastasis is frequently performed. Imaging studies include CT; for high-grade NEC, an FDG-PET/CT scan to assess for metastatic disease is done. For NET, somatostatin receptor imaging studies such as gallium-68 or copper-64 DOTATATE scans may be performed if metastatic disease is suspected.

Gastric NETs (also called gastric carcinoid tumors) represent 7–9% of gastrointestinal NETs but <1% of gastric neoplasms; several types have been described, based in part on whether there is associated achlorhydria and hypergastrinemia. For all gastric NETs, initial evaluation includes upper endoscopy and biopsy. EUS may be helpful in assessing depth of invasion for larger tumors and for assessing regional lymph node metastases in higher grade tumors. Somatostatin analogue imaging using DOTATATE scanning may be performed if metastatic disease is suspected. As is the case for other NETs, the finding of unresectable metastatic disease that is DOTATATE avid not only provides staging information but also guides potential therapy using somatostatin receptor–targeted therapy. See **Chap. 89** for a more detailed discussion.

TREATMENT

Esophageal Neuroendocrine Tumors

For early-stage, lower-grade NETs, endoscopic resection including EMR or ESD may be performed. Surgery may be employed for more advanced-stage locoregional but not metastatic NET. Small-cell or large-cell NECs that are not metastatic are usually treated with chemotherapy plus external-beam radiation therapy using chemotherapy regimens similar to those employed for small- and large-cell neuroendocrine cancers of the lung **(Chap. 83)**. For metastatic NETs, somatostatin receptor DOTATATE scans guide therapy. See below discussion for small-bowel NETs for discussion of use of somatostatin receptor blockade with agents such as octreotide or lanreotide and the use of peptide-related radiation therapy (PRRT). Systemic therapy for metastatic small- and large-cell esophagogastric NECs is also modeled on therapy for small- and large-cell thoracic NECs.

TREATMENT

Gastric Neuroendocrine Tumors

Type 1 tumors can be treated endoscopically with polypectomy or endomucosal resection. For larger tumors (>2 cm) or tumors invading through the muscularis or to regional lymph nodes, surgical resection is recommended. Type 2 tumors have a higher risk for regional lymph node metastasis and are usually treated surgically, although selected patients may have a combination of endoscopic resection and limited surgical resection. Type 3 tumors are not associated with elevated gastrin levels and have a higher propensity for regional lymph node metastasis and distant metastasis. Surgery is the treatment of choice for localized type 3 tumors, although EMR has been used in selected patients. Adenocarcinoma of the stomach may be found in 5–10% of type 3 tumors. The role of surgery in mixed NEC-adenocarcinoma neoplasms is not well established.

■ NEUROENDOCRINE TUMORS (CARCINOID) OF THE SMALL BOWEL

In the United States, NET are the most common small-bowel neoplasms. For not yet identified reasons, the incidence of small-bowel carcinoid tumors has markedly increased over the past several decades. A recent study found a 9% incidence of inherited germline cancer susceptibility genes in patients with small-bowel NET; however, it is unlikely that this is the cause of the increased incidence **(Chap. 89)**. Even with an increase in incidence, small-bowel NETs are still uncommon (~12 cases per million in the United States); the disease is more common in African Americans than whites.

Anatomically, the ileum is the most common part of the small bowel affected (~49%), followed by the duodenum and the jejunum. The same grading system, based on histology and the Ki-67 proliferative index or mitotic count, as for other gastrointestinal NETs is used. In the absence of metastatic disease to the liver in the subgroup of patients whose tumors are functional (i.e., produce a hormone, usually serotonin; a duodenal NET may produce gastrin), presenting symptoms may be vague abdominal discomfort until or unless small-bowel obstruction occurs. Carcinoid syndrome (including diarrhea and/or flushing) may occur in patients whose tumors are diagnosed with already established hepatic metastasis. Because the liver is very efficient at clearing serotonin on first pass, carcinoid syndrome as a result of small-bowel carcinoid tumors usually does not occur in the absence of hepatic metastasis.

Clinical evaluation includes a diagnostic-quality CT or MRI of the abdomen and pelvis. For duodenal NET, upper endoscopy with EUS is also performed. For ileal NET, colonoscopy with evaluation of the terminal ileum is performed. Imaging studies are similar to those used for diagnosis and localization of small-bowel adenocarcinomas. Somatostatin analogue imaging using gallium-68 or copper-64 DOTATATE PET/CT is helpful in assessing extent of disease in patients whose tumors are somatostatin analogue avid, as well as in identifying patients who may benefit from therapy targeting the somatostatin receptor. Plasma or urine assay for 5-hydroxyindoleacetic acid (5-HIAA) is performed to assess for a functional small-bowel NET. The AJCC Ninth Edition Cancer Staging Manual has a specific small-bowel NET TNM stage classification (for the duodenum and ampulla and for the jejunum and ilium).

The majority of patients present with locoregional disease, with approximately 40% having identified lymph node metastasis. Ten to 15% of patients have metastatic disease (usually to the liver) at the time of initial diagnosis.

Initial management should be surgical resection with curative intent; for patients with extensive adenopathy involving the root of the mesentery, vascular reconstruction may be required. Since small-bowel NETs may involve multiple tumors (15–30% of patients), the entire small bowel should be carefully examined at surgery. For patients with functional carcinoid tumors, somatostatin analogue therapy using agents such as octreotide should be given before the induction of anesthesia to avoid a carcinoid crisis. For patients with hepatic metastasis, resection or regional therapy including ablation or hepatic artery embolization for functional tumors may provide effective palliation. Carcinoid syndrome may also be palliated by somatostatin receptor–targeted therapy in the patients in whom DOTATATE scanning is positive (the majority of patients with carcinoid syndrome), including agents such as octreotide or lanreotide, or by peptide-directed radiation therapy using agents such as lutetium-177. Everolimus, an mTOR kinase inhibitor, has modest activity in metastatic small-bowel carcinoid tumors.

FURTHER READING

Ben-Aharon I et al: Early-onset cancer in the gastrointestinal tract is on the rise-evidence and implications. Cancer Discov 13:538, 2023.

Cancer Genome Atlas Research Network: Comprehensive molecular characterization of gastric adenocarcinoma. Nature 513:202, 2014.

Cancer Genome Atlas Research Network et al: Integrated genomic characterization of oesophageal carcinoma. Nature 541:169, 2017.

Choi IJ et al: *Helicobacter pylori* and prevention of gastric cancer. N Engl J Med 378:2244, 2018.

Choi IJ et al: Family history of gastric cancer and *Helicobacter pylori* treatment. N Engl J Med 382:427, 2020.

Dermawan JK et al: Novel genomic risk stratification model for primary gastrointestinal stromal tumors (GIST) in the adjuvant therapy era. Clin Cancer Res. 29:3974, 2023.

Hoeppner J et al: Perioperative chemotherapy or preoperative chemoradiotherapy in esophageal cancer. N Engl J Med 392:323, 2025.

Janjigian YY et al: Pembrolizumab plus trastuzumab and chemotherapy for HER2-positive gastric or gastro-oesophageal junction adenocarcinoma: Interim analyses from the phase 3 KEYNOTE-811 randomised placebo-controlled trial. Lancet 402:2197, 2023.

Pourmand K, Itzkowitz SH: Small bowel neoplasms and polyps. Curr Gastroenterol Rep 18:23, 2016.

Watanabe M et al: Recent progress in multidisciplinary treatment for patients with esophageal cancer. Surg Today 50:12, 2020.

86

Colorectal Cancer

David P. Ryan

INCIDENCE

Colorectal cancer is the most common cancer of the gastrointestinal system in the United States. In 2024, 153,000 cases are expected, with ~106,000 cases in the colon and 46,000 cases in the rectum. It is the second most common cause of death from cancer (lung cancer is first) with 53,000 expected deaths. There is a slight male predominance in the incidence of colorectal cancer, and it is more common in black Americans than white Americans. Since 1985, there has been a steady decline in the incidence of colorectal cancer in both men and women likely due to the adoption of widespread screening guidelines for colonoscopy. This decline is seen in adults older than 50. However, a steady increase in the incidence of colorectal cancer has been seen in patients under the age of 50. While in older adults a gradual shift into predominantly right-sided cancer has been noted, a gradual increase in left-sided colon and rectal cancers has been seen in patients under the age of 50. This increase in incidence has resulted in an increase in mortality rates from colorectal cancer among young adults. Colon cancer is now the leading cause of cancer death for young men age 20–49 and the second leading cause of cancer death after breast cancer for women age 40–49.

ADENOMATOUS POLYPS

Most colorectal cancers arise from adenomatous polyps as opposed to a hyperplastic polyp, hamartomatous polyp, or a serrated polyp. While adenomatous polyps are clearly premalignant, only a minority of adenomatous polyps will ever develop into an adenocarcinoma. Adenomatous polyps are common in the United States and western societies, with approximately 25% of people having adenomatous polyps by age 50. The rate steadily rises as people age, and as many as 50% of 70-year-old people will have adenomatous polyps. Less than 5% of adenomatous polyps are expected to progress to an adenocarcinoma. The vast majority of adenomatous polyps are asymptomatic and do not bleed. The risk of progression to cancer is associated with size of the polyp and the histology. Villous adenomas develop into cancer threefold more often than tubular adenomas.

MOLECULAR PATHOGENESIS

The majority of colon cancers arise from a series of genetic and epigenetic events involving tumor suppressor genes and oncogenes. Three major types of molecular pathways are involved in sporadic colorectal cancer: the chromosomal instability (CIN) pathway, the mismatch repair (MMR) pathway, and the CpG island methylation pathway (CIMP). The CIN pathway is the most common molecular pathway associated with colon cancer and is characterized by early loss of the adenomatous polyposis coli (*APC*) gene, which is a tumor suppressor gene. Subsequent mutations in important oncogenes involving the MAP kinase pathway, such as *KRAS* and *BRAF*, are acquired sequentially. Finally, loss of *p53* tumor suppressor gene is a common final step.

The defects in the MMR pathway in sporadic colon cancer are often caused by hypermethylation of the promoters for mismatch repair genes. This results in a phenotype known as microsatellite instability high (MSI-high), which is characterized by altered sizes of various mono- and di-nucleotide repeat sequences. A defect in the MMR pathway is the etiology of Lynch syndrome, which involves a germline mutation in one of the genes encoding for the proteins involved in the MMR pathway. It is critical to test for MSI-high because these tumors are dramatically sensitive to immune checkpoint inhibition and less sensitive to chemotherapy. The third major molecular pathway is the CIMP phenotype. These tumors are often right sided and associated with methylation of MLH1. They can be either MSI-high or microsatellite stable (MSS). The three major pathways are not mutually exclusive, and often overlapping changes can be seen in individual cancers.

ETIOLOGY AND RISK FACTORS

Risk factors for the development of colorectal cancer are listed in Table 86-1.

HEREDITARY FACTORS AND SYNDROMES

Advances in germline genotyping have made this technology more widely available to the population of patients getting colon cancer. Approximately 10% of all patients with colorectal cancer will have an inherited, germline predisposition to colorectal cancer, and this rises to ~15% of young adults with colorectal cancer **(Table 86-2)**. The germline mutations can be divided into high penetrance and moderate penetrance based on their risk of colon cancer development.

Familial Adenomatous Polyposis (FAP) FAP and its variants are due to a mutation in the *APC* gene located on chromosome 5q21-22. It is inherited in an autosomal dominant pattern, although a significant minority of cases are due to de novo mutations. FAP accounts for <1% of all colon cancer and is characterized by the development of hundreds to thousands of colonic polyps. It is diagnosed in childhood in the classical form of FAP, and colorectal cancer develops in virtually 100% of individuals if left untreated. An attenuated form of FAP is characterized by tens of polyps and cancers developing at a later age. FAP

TABLE 86-1 Risk Factors for the Development of Colorectal Cancer

Diet: Animal fat, obesity
Hereditary syndromes
Polyposis coli
MYH-associated polyposis
Nonpolyposis syndrome (Lynch's syndrome)
Inflammatory bowel disease
Streptococcus bovis bacteremia
Tobacco use

TABLE 86-2 Heritable (Autosomal Dominant) Gastrointestinal Neoplasia Syndromes

SYNDROME	DISTRIBUTION OF POLYPS	HISTOLOGIC TYPE	MALIGNANT POTENTIAL	ASSOCIATED LESIONS
Familial adenomatous polyposis	Large intestine	Adenoma	Common	—
Gardner's syndrome	Large and small intestines	Adenoma	Common	Osteomas, fibromas, lipomas, epidermoid cysts, ampullary cancers, congenital hypertrophy of retinal pigment epithelium
Turcot's syndrome	Large intestine	Adenoma	Common	Brain tumors
MYH-associated polyposis	Large intestine	Adenoma	Common	None
Lynch syndrome (nonpolyposis syndrome)	Large intestine (often proximal)	Adenoma	Common	Endometrial and ovarian tumors (most frequently), gastric, genitourinary, pancreatic, biliary cancers (less frequently)
Peutz-Jeghers syndrome	Small and large intestines, stomach	Hamartoma	Rare	Mucocutaneous pigmentation; tumors of the ovary, breast, pancreas, endometrium
Juvenile polyposis	Large and small intestines, stomach	Hamartoma, rarely progressing to adenoma	Rare	Various congenital abnormalities

can be associated with extracolonic manifestations such as desmoid tumors (Gardner's syndrome) or brain tumors (Turcot's syndrome).

Patients with classic FAP should undergo colectomy, as well as those patients with attenuated FAP that is too difficult to screen endoscopically. Colectomy patients with FAP should be followed for extracolonic tumors and manifestations. Chemoprevention with nonsteroidal anti-inflammatory drugs (NSAIDs) can be tried because it has been associated with polyp regression and delayed progression, but its overall effect on cancer prevention has not been established.

MUTYH-Associated Polyposis (MAP) MAP is an autosomal recessive polyposis syndrome caused by a biallelic mutation in the *MUTYH* gene. The *MUTYH* gene is a base excision repair gene, and failure of base excision repair leads to somatic CG-AT transversions in multiple genes. MAP accounts for <1% of colorectal cancers. The clinical presentation of MAP overlaps with attenuated FAP in that individuals may have tens of polyps that develop by the fifth or sixth decade of life. Individuals with MAP are at increased risk of duodenal cancers and thyroid cancer. Screening and colectomy guidelines for this syndrome are less clear than for polyposis coli, but annual to biennial colonoscopic surveillance is generally recommended starting at age 25–30 years.

Lynch Syndrome Lynch syndrome, previously known as hereditary nonpolyposis colon cancer, is the most common inherited predisposition to colorectal cancer, accounting for ~3% of all colorectal cancers, and confers an up to 17-fold risk of colorectal cancer. Lynch syndrome is due to a mutation in one of the MMR genes. The MMR genes are *MLH1* (Chr 3p22), *MSH2* (2p21-16), *MSH6* (2p16), and *PMS2* (7p22). Lynch syndrome can also occur due to deletion of the *EPCAM* gene, which causes loss of expression of *MSH2*. These mutations are inherited in an autosomal dominant fashion. The MMR system maintains genomic integrity by correcting base substitution mismatches, and failure of the MMR system results in the accumulation of a thousand-fold more mutations in genes that drive carcinogenesis compared with MMR-proficient tumors. Lynch syndrome has multiple extracolonic manifestations including increased risk of cancers of the ovary, endometrium, stomach, small bowel, pancreaticobiliary system, genitourinary system, brain, and skin.

Lynch syndrome may not be associated with multiple polyps and therefore is often unrecognized. Universal testing of colorectal tumors for defects in the MMR system is recommended and accomplished by either immunohistochemistry for MMR proteins in the tumor sample or evaluation for microsatellite instability on the tumor sample. For patients without a Lynch syndrome–associated cancer, family history is the most reliable way to recognize Lynch syndrome, and several family history criteria (Amsterdam Criteria and Revised Bethesda Criteria) have been developed. Germline testing is recommended for all patients with defects in the MMR system noted on the tumor sample or for those non-cancer-affected patients with an appropriate family history. Germline testing should be done in the context of appropriate pre- and posttest genetic counseling.

***BRCA1/2* Mutation** The hereditary breast and ovarian cancer syndromes due to mutations in the tumor suppressor genes *BRCA1* and *BRCA2* may carry an increased risk of colorectal cancer. In particular, *BRCA1* mutation may confer a 1.5-fold increased risk of colorectal cancer. Currently, the evidence is inconsistent, and therefore, guidelines do not recommend increased screening. Nevertheless, *BRCA* carriers may have an increased risk of young adult–onset colorectal cancer, and clinicians should be aware of these data.

Diet The etiology for most cases of large-bowel cancer appears to be related to environmental factors. The disease occurs more often in upper socioeconomic populations who live in urban areas. Mortality from colorectal cancer is directly correlated with per capita consumption of calories, meat protein, and dietary fat and oil as well as elevations in the serum cholesterol concentration and mortality from coronary artery disease. Geographic variations in incidence largely are unrelated to genetic differences because migrant groups tend to assume the large-bowel cancer incidence rates of their adopted countries. Furthermore, population groups such as Mormons and Seventh Day Adventists, whose lifestyle and dietary habits differ somewhat from those of their neighbors, have significantly lower-than-expected incidence and mortality rates for colorectal cancer. The incidence of colorectal cancer has increased in Japan since that nation has adopted a more "Western" diet. At least three hypotheses have been proposed to explain the relationship to diet, none of which is fully satisfactory.

ANIMAL FATS One hypothesis is that the ingestion of animal fats found in red meats and processed meat leads to an increased proportion of anaerobes in the gut microflora (the "microbiome"), resulting in the conversion of normal bile acids into carcinogens. This provocative hypothesis is supported by several reports of increased amounts of fecal anaerobes (*Fusobacterium nucleatum*, *Bacteroides fragilis*) in the stools of patients with colorectal cancer. Diets high in animal (but not vegetable) fats are also associated with high serum cholesterol, which is also associated with enhanced risk for the development of colorectal adenomas and carcinomas.

INSULIN RESISTANCE The large number of calories in Western diets coupled with physical inactivity has been associated with a higher prevalence of obesity. Obese persons develop insulin resistance with increased circulating levels of insulin, leading to higher circulating concentrations of insulin-like growth factor type I (IGF-I). This growth factor appears to stimulate proliferation of the intestinal mucosa.

FIBER Contrary to prior beliefs, the results of randomized trials and case-controlled studies have *failed* to show any value for dietary fiber or diets high in fruits and vegetables in preventing the recurrence of colorectal adenomas or the development of colorectal cancer.

The weight of epidemiologic evidence, however, implicates diet as being the major etiologic factor for colorectal cancer, particularly diets high in animal fat and in calories.

INFLAMMATORY BOWEL DISEASE (CHAP. 337)

Colon and rectal cancers (but not anal cancers) are more common in patients with inflammatory bowel disease (IBD). Most of the data supporting this come from patients with ulcerative colitis. The risk of colorectal cancer in a patient with IBD is relatively small during the first 10 years of the disease but then appears to increase at a rate of ~0.5–1% per year. Cancer may develop in 8–30% of patients. The risk is higher in patients with more extensive, more severe, and longer-lasting colitis. The data are less consistent with Crohn's disease, but this may be due to the difficulty in distinguishing the two entities and also the different amount of pancolitis in the two diseases. The molecular pathogenesis of colorectal cancer arising from IBD appears to be distinctly different from sporadic cancers. For instance, *K-ras* and *APC* mutations appear to be less common in colon cancers arising in IBD patients.

Cancer surveillance strategies in patients with IBD are unsatisfactory due to the lack of the classic polyp-to-cancer sequence. Symptoms such as bloody diarrhea, abdominal cramping, and obstruction, which may signal the appearance of a tumor, are similar to the complaints caused by a flare-up of the underlying inflammatory disease. Surveillance for colorectal cancer in patients with IBD is highly technical and requires management by physicians specializing in this area. In patients with longstanding ulcerative colitis, surgical removal of the colon (subtotal or total) eliminates or significantly reduces the risk of colorectal cancer depending on the extent of the operation. In patients undergoing subtotal colectomy with preservation of the rectum, continued surveillance of the remaining rectum is required.

OTHER HIGH-RISK CONDITIONS

Tobacco Use Cigarette smoking is linked to the development of colorectal adenomas, particularly after >35 years of tobacco use. No biologic explanation for this association has yet been proposed.

Alcohol Use Moderate to excessive alcohol consumption has been associated with an increased risk of colorectal cancer. It is difficult to tease out the effect compared with other lifestyle risk factors.

Obesity, Insulin Resistance, and Diabetes Mellitus Multiple studies demonstrate an increased risk of colorectal cancer in obese individuals. Additionally, diabetes is associated with an increased risk of colorectal cancer. This has led to a theory that metabolic syndrome and hyperinsulinemia are risk factors for the development of colorectal cancer.

Vitamin D Deficiency Multiple studies have demonstrated an association between low levels of vitamin D and colon cancer. Nevertheless, no evidence suggests that vitamin D supplementation results in fewer colon cancers.

PRIMARY PREVENTION

Several orally administered compounds have been assessed as possible inhibitors of colon cancer. The most effective class of chemopreventive agents is aspirin and other NSAIDs, which are thought to suppress cell proliferation by inhibiting prostaglandin synthesis. Regular aspirin use as demonstrated in randomized studies reduces the risk of colon adenomas. Regular aspirin use in cohort studies has demonstrated reduced incidence of colon cancer. However, prospective trials using aspirin to prevent colon cancer have had conflicting results in part due to the delayed effects of aspirin on colon carcinogenesis; prevention may increase with the duration and dosage of aspirin, and the effects may take years to manifest. Meta-analyses have suggested that regular aspirin use prevents colon cancer. However, given the potential side effects of aspirin, regular use of aspirin is an individualized decision.

Antioxidant vitamins such as ascorbic acid, tocopherols, and β-carotene are ineffective at reducing the incidence of subsequent adenomas in patients who have undergone the removal of a colon adenoma. Estrogen replacement therapy has been associated with a reduction in the incidence of colorectal cancer in women but not mortality from colorectal cancer.

SCREENING

The rationale for colorectal cancer screening programs is that the removal of adenomatous polyps will prevent colorectal cancer and that earlier detection of localized, superficial cancers in asymptomatic individuals will increase the surgical cure rate. It is important to note that screening studies did not take into account the presence of inherited predisposition and germline genetics. Due to the rise in incidence of colorectal cancer in younger adults, the U.S. Preventative Services Task Force updated their screening recommendations in 2021 and lowered the age for recommended screening to 45.

TABLE 86-3 Screening Strategies for Colorectal Cancer

Digital rectal examination
Stool testing
- Occult blood
- Fecal DNA

Imaging
- Contrast barium enema
- Virtual (i.e., computed tomography colonography)

Endoscopy
- Flexible sigmoidoscopy
- Colonoscopy

Screening strategies for colorectal cancer that have been examined during the past several decades are listed in Table 86-3.

The rationale for screening asymptomatic individuals has undergone change over many decades, and it is important to understand how that change influences our current approach. Initially, patients were screened for the presence of occult blood in the stool, which would then lead to colonoscopy. Unfortunately, even when performed optimally, the fecal occult blood test has major limitations as a screening technique. About 50% of patients with documented colorectal cancers have a negative fecal occult blood test, consistent with the intermittent bleeding pattern of these tumors. When random cohorts of asymptomatic persons have been tested, 2–4% have fecal occult blood–positive stools. Colorectal cancers have been found in <10% of these "test-positive" cases, with benign polyps being detected in an additional 20–30%. Thus, a colorectal neoplasm will not be found in most asymptomatic individuals with occult blood in their stool. Nevertheless, prospectively controlled trials have shown a statistically significant reduction in mortality rate from colorectal cancer for individuals undergoing annual stool guaiac screening. However, this benefit only emerged after many years of follow-up and was due to colonoscopic intervention, which likely provided the opportunity for cancer prevention through the removal of potentially premalignant adenomatous polyps because the eventual development of cancer was reduced by 20% in the cohort undergoing annual screening.

Due to the importance of endoscopic screening in preventing death from colon cancer, many societies shifted from testing stool for the presence of occult blood to offering endoscopic screening for the asymptomatic population. Colonoscopy evaluates the entire colon but is associated with more complications, the need for cathartics to remove stool from the colon, the need for sedation in most cases, and more overall expense. Sigmoidoscopy does not require an enema, is associated with less perforations and morbidity, does not require sedation, and can identify patients at high risk of needing a full colonoscopy if polyps or cancer is found in the rectum or sigmoid colon. The recommendation for the inclusion of flexible sigmoidoscopy is strongly supported by randomized studies evaluating a one-time sigmoidoscopy as opposed to usual standard of care. A reduction in both colorectal cancer incidence and colorectal cancer mortality is obtained for patients undergoing screening, and this effect persists for >15 years. One of the downsides of using screening sigmoidoscopy alone is that in the presence of a normal sigmoidoscopy, ~1.5% of individuals will have either a high-risk polyp or adenocarcinoma in the proximal colon. The issue of missing advanced neoplasms in the proximal colon when using only sigmoidoscopy was addressed by increased frequency of sigmoidoscopy (every 5 years instead of every 10 years for colonoscopy) and the use of annual fecal immunohistochemical test (FIT) for occult blood. While randomized studies evaluating the combination are lacking, most professional societies offer this approach as an alternative to colonoscopy screening.

With the appreciation that the carcinogenic process leading to the progression of the normal bowel mucosa to an adenomatous polyp and then to a cancer is the result of a series of molecular changes, investigators have examined fecal DNA for evidence of mutations associated with such molecular changes as evidence of the occult presence of precancerous lesions or actual malignancies. A multitarget stool DNA test evaluates stool that is collected at home for DNA changes including methylation combined with a FIT test. This multitarget stool DNA test, when done in patients undergoing colonoscopy, has a >90% specificity and sensitivity for detecting a colorectal adenocarcinoma but a <50% sensitivity for detecting an advanced precancerous lesion.

The use of imaging studies to screen for colorectal cancers has also been explored. Air contrast barium enemas had been used to identify sources of occult blood in the stool prior to the advent of fiberoptic endoscopy; the cumbersome nature of the procedure and inconvenience to patients limited its widespread adoption. The introduction of computed tomography (CT) scanning led to the development of virtual (i.e., CT) colonography as an alternative to the growing use of endoscopic screening techniques. Virtual colonography requires use of colon cathartic bowel preparations and appears to have good accuracy at detecting large lesions (>10 mm) but has not been compared in a randomized study with colonoscopy. Two other disadvantages of virtual colonoscopy are that it may have interuser variability and any findings need to be followed up with a colonoscopy requiring a second bowel cathartic preparation.

Most professional societies agree that screening for colon cancer should begin at age 45 for patients at average risk. In the United States, there is ~60% compliance with screening and the emphasis is on getting patients screened initially with either direct observation using sigmoidoscopy or colonoscopy or indirect tests such as FIT or multitarget stool DNA. The best combination approach and frequency of screening likely will involve risk stratification factors in the future.

CLINICAL FEATURES

Presenting Symptoms Symptoms vary with the anatomic location of the tumor. Because stool is relatively liquid as it passes through the ileocecal valve into the right colon, cancers arising in the cecum and ascending colon may become quite large without resulting in any obstructive symptoms or noticeable alterations in bowel habits. Lesions of the right colon commonly ulcerate, leading to chronic, insidious blood loss without a change in the appearance of the stool. Consequently, patients with tumors of the ascending colon often present with symptoms associated with hypochromic, microcytic anemia, indicative of iron deficiency. As a result, the unexplained presence of iron-deficiency anemia in any adult (with the possible exception of a premenopausal, multiparous woman) mandates a thorough endoscopic and/or radiographic visualization of the entire large bowel (Fig. 86-1).

Because stool becomes more formed as it passes into the transverse and descending colon, tumors arising there tend to impede the passage of stool, resulting in the development of abdominal cramping, occasional obstruction, and even perforation. Radiographs and CTs of the abdomen often reveal characteristic annular, constricting lesions ("apple-core") (Fig. 86-2).

Cancers arising in the rectosigmoid are often associated with hematochezia, tenesmus, and narrowing of the caliber of stool. While these symptoms may lead patients and their physicians to suspect the presence of hemorrhoids, the development of rectal bleeding and/or altered bowel habits demands a prompt digital rectal examination and sigmoidoscopy.

Staging, Prognostic Factors, and Patterns of Spread The prognosis for individuals having colorectal cancer is related to the depth of tumor penetration into the bowel wall and the presence of both regional lymph node involvement and distant metastases. These variables are incorporated into a TNM classification method, in which T represents the depth of tumor penetration, N the presence of lymph node involvement, and M the presence or absence of distant metastases (Fig. 86-3). Superficial lesions that do not involve regional lymph nodes and do not penetrate through the submucosa (T1) or the muscularis (T2) are designated as *stage I* (T1–2N0M0) disease; tumors that penetrate through the muscularis but have not spread to lymph nodes are *stage II* disease (T3–4N0M0); regional lymph node involvement defines *stage III* (TXN1–2M0) disease; and metastatic spread to sites such as liver, lung, or bone indicates *stage IV* (TXNXM1) disease. Unless gross evidence of metastatic disease is present, disease stage cannot be determined accurately before surgical resection and pathologic analysis of the operative specimens.

The majority of recurrences occur within the initial 5 years after resection. Occasionally tumors can recur between 5 and 10 years after resection, and this appears to be more common with rectal cancer

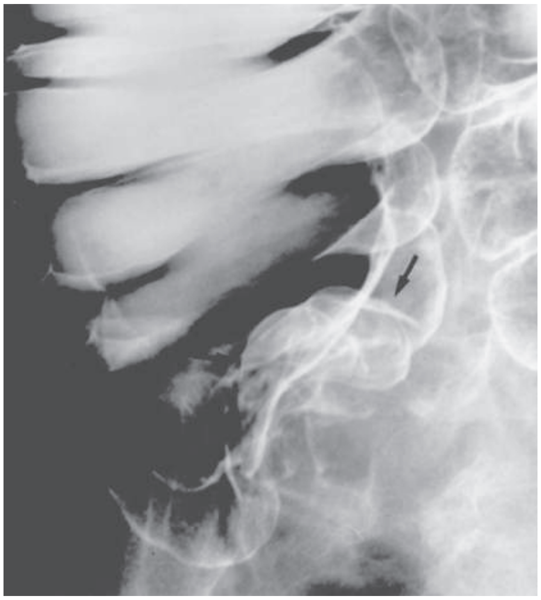

FIGURE 86-1 Double-contrast air-barium enema revealing a sessile tumor of the cecum in a patient with iron-deficiency anemia and guaiac-positive stool. The lesion at surgery was a stage II adenocarcinoma.

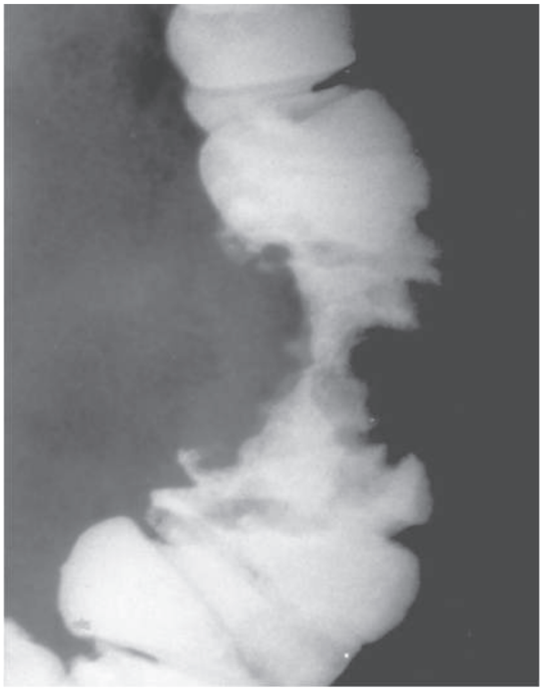

FIGURE 86-2 Annular, constricting adenocarcinoma of the descending colon. This radiographic appearance is referred to as an "apple-core" lesion and is always highly suggestive of malignancy.

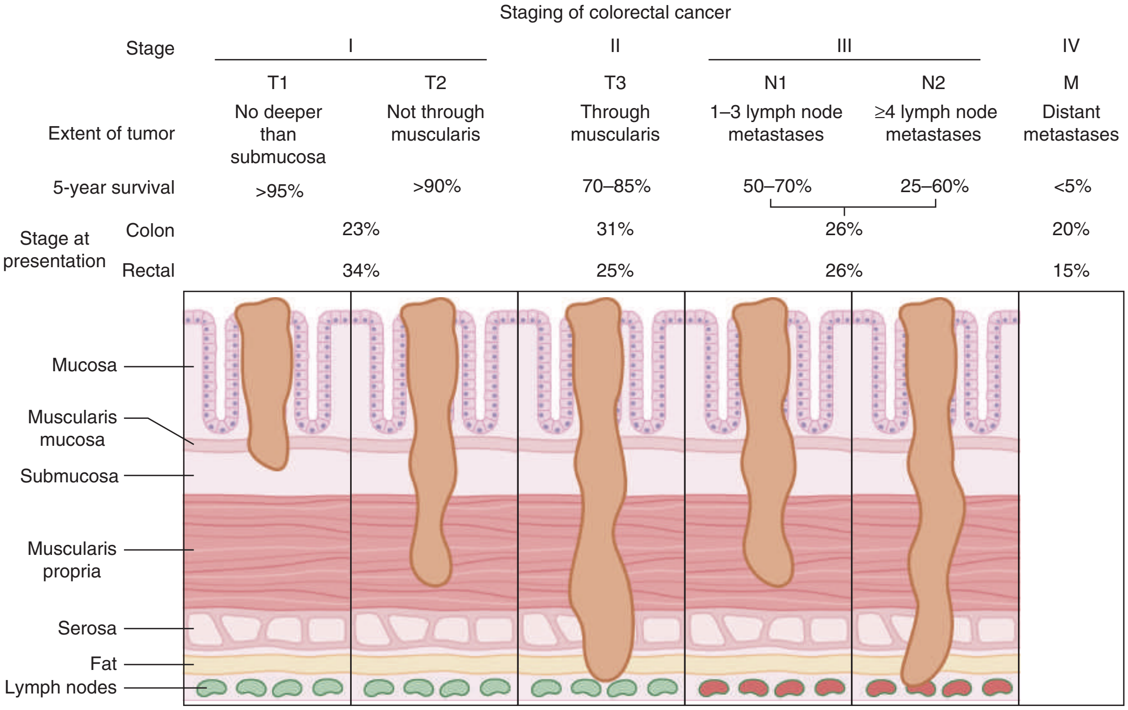

Staging of colorectal cancer

Stage	I		II	III		IV
	T1	T2	T3	N1	N2	M
Extent of tumor	No deeper than submucosa	Not through muscularis	Through muscularis	1–3 lymph node metastases	≥4 lymph node metastases	Distant metastases
5-year survival	>95%	>90%	70–85%	50–70%	25–60%	<5%
Stage at presentation: Colon	23%		31%	26%		20%
Stage at presentation: Rectal	34%		25%	26%		15%

FIGURE 86-3 Staging and prognosis for patients with colorectal cancer.

than colon cancer. The likelihood for 5-year survival in patients with colorectal cancer is stage-related (Fig. 86-3), and the survival rates per stage have been improving in the past several decades. The most plausible explanation for improved survival is the more thorough evaluation for metastatic disease and the positive effects of chemotherapy for treatment of adjuvant and systemic disease. A minimum of 12 sampled lymph nodes is thought necessary to accurately define tumor stage, and the more nodes examined, the better reliability that a patient is either node negative or node positive. Other predictors of a poor prognosis after a total surgical resection include tumor penetration through the bowel wall into pericolic fat, poorly differentiated histology, perforation and/or tumor adherence to adjacent organs (increasing the risk for an anatomically adjacent recurrence), and venous invasion by tumor **(Table 86-4)**. Tumors with microsatellite instability or defects in MMR have improved prognosis presumably due to an enhanced immune response against the tumor. Tumors arising in the left colon are associated with a better prognosis than those appearing in the right colon, likely due to differences in molecular patterns. In contrast to most other cancers, the prognosis in colorectal cancer is not influenced by the size of the primary lesion when adjusted for nodal involvement and histologic differentiation.

Colorectal cancer generally spreads to the liver, lungs, and peritoneal cavity. The liver represents the most frequent visceral site of metastasis; it is the initial site of distant spread in one-third of recurring colorectal cancers and is involved in more than two-thirds of such patients at the time of death. It can spread to the brain and bones but not typically as an initial site of metastatic disease, but rather after years of having systemic metastases. Nevertheless, prompt evaluation of signs and symptoms of bone and brain involvement should be efficiently done because lack of prompt treatment can have devastating consequences (e.g., cord compression, brain hemorrhage).

TABLE 86-4 Predictors of Poorer Outcomes Following Total Surgical Resection of Colorectal Cancer

Tumor spread to regional lymph nodes
Number of regional lymph nodes involved
Tumor penetration through the bowel wall
Poorly differentiated histology
Perforation
Tumor adherence to adjacent organs
Venous invasion
Preoperative elevation of CEA titer (>5 ng/mL)
Specific chromosomal deletion (e.g., mutation in the *b-raf* gene)
Right-sided location of primary tumor

Abbreviation: CEA, carcinoembryonic antigen.

TREATMENT

Colorectal Cancer

Staging should consist of a complete blood count, comprehensive metabolic panel, serum carcinoembryonic antigen (CEA), and a chest, abdomen, and pelvic CT scan. Additional imaging such as magnetic resonance imaging (MRI) and positron emission tomography (PET) scan can be dependent on the findings of the CT scan. When possible, a colonoscopy of the entire large bowel should be performed to identify synchronous neoplasms and/or polyps. While surgical resection is the standard of care for cure in patients with localized disease, initial treatment of colorectal cancer when possible should be delayed until MMR status or microsatellite stability status is known. MSI-high rectal cancers can be treated for cure with checkpoint inhibitors (PD-1 antibodies) without the need for surgery, radiation, and chemotherapy. Studies are underway involving management of MSI-high colon cancers with checkpoint inhibition.

LOCALIZED COLON CANCER

For the vast majority of patients with MSS colon adenocarcinoma, initial surgical resection is the standard of care. Surgery with minimally invasive techniques such as laparoscopy and robotic approaches has largely replaced open approaches. Following complete surgical resection, patients with stage I disease are considered cured. Patients with high-risk stage II and stage III colon cancer are considered for the appropriateness of adjuvant chemotherapy. The standard adjuvant approaches generally improve overall survival and cure rates by ~30%. Adjuvant chemotherapy for 3–6 months either with single-agent fluoropyrimidine (IV or oral) or in combination with oxaliplatin is considered the standard adjuvant treatment for MSS colon cancer.

Following recovery from a complete resection, patients should be observed carefully for 5 years using physical examinations and blood chemistry measurements at regular intervals. Some authorities favor measuring plasma CEA levels at 3-month intervals because of the sensitivity of this test as a marker for otherwise undetectable tumor recurrence. The value of periodically assessing plasma for the presence of circulating tumor DNA as a biomarker for residual or recurrent disease is under study. Subsequent endoscopic surveillance of the large bowel 1 year after resection and then usually every 3 years is indicated because patients who have been cured of one colorectal cancer have a 3–5% probability of developing an additional bowel cancer during their lifetime and a >15% risk for the development of adenomatous polyps. Anastomotic ("suture-line") recurrences are infrequent in colorectal cancer patients, provided the surgical resection margins were adequate and free of tumor. The value of periodic CT scans of the abdomen, assessing for an early, asymptomatic indication of tumor recurrence, is uncertain; however, CT has been recommended semi-annually to annually for the first 3 postoperative years.

LOCALIZED RECTAL CANCER

For the vast majority of patients with MSS rectal adenocarcinoma, surgical resection is considered the standard curative therapy. In addition to the staging above, pelvic MRI should be standard in staging for all patents with rectal cancer to assess mesorectal margin and nodal status. Surgical resection can either be a low anterior resection (LAR) or an abdominal-perineal resection (APR). An APR requires a permanent colostomy due to resection of the anal sphincter. Unlike colon cancer, rectal cancers (defined as tumor at or below the peritoneal reflection) have an increased risk of local recurrence after surgery. The risk of local recurrence increases with stage and closeness to the mesorectal border and decreases with distance from the anus. Thus, tumors requiring an APR have a high risk of local recurrence. The risk of local recurrence can be reduced with preoperative chemotherapy and radiation. Patients with stage II and III rectal cancer are deemed to have a risk high enough to warrant consideration of preoperative therapy.

Two new treatment paradigms have emerged for patients with rectal cancer in specific circumstances. The first situation involves patients in whom the consequences of radiation therapy are deemed unacceptable. Radiation is associated with the potential long-term side effects of radiation proctitis (increased risk in patients with IBD), small-bowel obstruction to adhesions, and malignancy (~1% risk after 10 years). Notably, radiation therapy in women causes infertility due to the effects of radiation on the uterus even when the ovaries are not included in the radiation field. Preoperative therapy with a fluoropyrimidine and oxaliplatin followed by surgical resection produces similar rates of local control to preoperative chemoradiation and can be considered for average-risk stage II and III rectal cancer patients.

The second circumstance involves patients with stage II and III rectal cancer treated with total preoperative therapy who experience a complete clinical response. Nonoperative management can be considered in these patients because 60–70% will be cured without an LAR or APR. It is important to note that 30–40% of patients will experience a recurrence so patients need to be followed intensively in the years following chemotherapy and radiation. The majority of patients who experience a local recurrence during nonoperative management can be cured with surgery.

METASTATIC COLORECTAL CANCER

The first principle to note is that complete resection of metastatic disease is considered curative for patients with colorectal cancer, particularly in the setting of metastases isolated to one site (liver, lungs, peritoneum). Consultation with the appropriate surgical specialists is preferred before starting systemic chemotherapy.

The detection of metastases should not preclude surgery to remove the primary tumor in patients with tumor-related symptoms such as gastrointestinal bleeding or obstruction but usually prompts the initiation of chemotherapy unless those symptoms are viewed to be life-threatening. Studies have demonstrated that resection of the primary tumor in the setting of unresectable metastatic disease does not improve overall survival and should be done only on an ad hoc basis.

Patients with MSI-high colon cancer should be treated with a checkpoint inhibitor as initial treatment. Long-term survival is possible in these patients, and some will have a complete response and may not ever have recurrence of their cancer. For the vast majority of patients with MSS colorectal adenocarcinoma who are not candidates for complete resection of metastatic disease, treatment with systemic chemotherapy is associated with marked improvement in overall survival compared with palliative care alone. A patient with metastatic disease lives on average for 2–3 years with systemic chemotherapy, and as many as 25% of patients will live for at least 5 years on and off chemotherapy.

The mainstays of therapy for patients with metastatic colon cancer are the following medications: 5-fluorouracil (5-FU; or capecitabine), irinotecan, oxaliplatin, and bevacizumab. These medications are used in combination all together (e.g., FOLFIRINOX-bevacizumab or FOLFOXIRI-bevacizumab) or in partial combination (e.g., FOLFOX-bevacizumab, FOFIRI-bevacizumab). 5-FU is an inhibitor of thymidylate synthase and is used in conjunction with folinic acid (leucovorin). It is metabolized by dihydropyrimidine dehydrogenase (DPD), which is deficient in ~1% of adults. Individuals with DPD deficiency can experience increased side effects from 5-FU, especially myelosuppression and mucositis. Testing for DPD gene mutations may identify patients at high risk of DPD deficiency and toxicity from 5-FU. Irinotecan is a topoisomerase I inhibitor and is metabolized by glucuronidation in the liver. Thus, patients with *UGT* mutations can experience decreased clearance and enhanced toxicity, principally myelosuppression. Oxaliplatin is a platinum analogue that is metabolized generally and can cause permanent peripheral neuropathy. Bevacizumab is a monoclonal antibody against vascular endothelial growth factor (VEGF). Side effects of bevacizumab include hypertension, nephrotic syndrome, and impaired wound healing. Untreated hypertension from bevacizumab has been associated with posterior reversible leukoencephalopathy syndrome.

Other agents with proven disease-altering effects in the metastatic setting include trifluridine-tipiracil, regorafenib, and fruquintinib. Trifluridine is a nucleoside analogue that is used with tipiracil, a thymidine phosphorylase inhibitor. Regorafenib is an oral multitargeted tyrosine kinase inhibitor, including VEGF. Fruquintinib is an oral VEFG tyrosine kinase inhibitor.

Molecular genotyping of the primary tumor should be done in all patients with metastatic colon cancer if possible. Patients without mutations of the RAS/RAF pathway can have excellent responses and prolonged survival with the addition of antibodies against epidermal growth factor (EGF), such as cetuximab and panitumumab. These agents are often given in combination with either FOLFOX or FOLFIRI. The effect of cetuximab and panitumumab appears to be limited to left-sided colon cancers because right-sided colon cancers have upregulation of receptor tyrosine kinases that create resistance to these antibodies. The main side effects of these two antibodies are the class effects of diarrhea and acneiform rash. Cetuximab, which is partially murine, is also associated with allergic reactions.

Patients who harbor *BRAF* V600E mutations may have a particularly virulent form of colon cancer that does not respond as well to standard chemotherapy. These mutations tend to occur more commonly in right-sided tumors and women. They often involve methylation of *MLH1* and can be MMR deficient. The BRAF inhibitor encorafenib is approved for use with cetuximab in patients with *BRAF*-mutant colorectal cancer.

HER2 overexpression can be seen in colorectal cancer, and consideration of anti-HER2–targeted therapy is warranted. Tucatinib (HER2 tyrosine kinase inhibitor) and trastuzumab (anti-HER2 monoclonal antibody) are approved in the United States for patients with HER2-positive colorectal cancers that lack a *RAS* mutation.

Rarely, *RET* fusions and *TRK* fusions can be seen in patients with metastatic colorectal cancer. These patients can have tremendous responses to the appropriate RET and TRK inhibitors with marked prolongation of survival.

CANCERS OF THE ANUS

Cancers of the anus account for 1–2% of the malignant tumors of the large bowel. Anal cancer by convention refers to squamous cell carcinomas arising in the anorectal region. Other types of cancers including melanoma, neuroendocrine cancer, lymphoma, and mesenchymal tumors can arise in these regions but are referred to by their histologic subtype and are not the subject of this chapter. True adenocarcinomas arising from the glands in the anal canal that are distinct from rectal adenocarcinoma can also arise but, for all intents and purposes, are indistinguishable from rectal adenocarcinoma and treated accordingly.

Most anal cancers arise in the anal canal, the anatomic area extending from the anorectal ring to a zone approximately halfway between the pectinate (or dentate) line and the anal verge. The dentate line can be quite variable in adults but, on average, is 4 cm from the anal verge. Squamous cell carcinomas are characterized as either nonkeratinizing or keratinizing. Older terms such as cloacogenic and basaloid are not used anymore, and these tumors are instead characterized as nonkeratinizing. Outcomes for keratinizing and nonkeratinizing squamous cell carcinoma of the anus are not different.

The development of anal cancer is associated with infection by human papillomavirus (HPV), the same virus etiologically linked to cervical and oropharyngeal cancers. The same HPV subtypes associated with cervical cancer are seen in anal cancer. The infection may lead to squamous intraepithelial lesions (SILs), which are classified as either low grade (LSIL) or high grade (HSIL). LSILs are associated with non–cancer-causing HPV subtypes, are not considered precancerous, and can spontaneously regress. HSILs are associated with HPV16 and are considered precancerous. Anal cancer risk is increased in both men and women with immunocompromised states including solid organ transplant patients, patients with chronic immunosuppression with glucocorticoids, and patients living with HIV, particularly with low CD4 counts. Anal cancers occur most commonly in middle-aged persons and are more frequent in women than men. At diagnosis, patients may experience bleeding, pain, sensation of a perianal mass, and pruritus. Examination of and attention to the inguinal lymph nodes on imaging is important because anal cancers often spread initially to the inguinal region rather than the iliac nodes like rectal cancer.

The standard treatment for anal cancers consists of 5-FU, mitomycin, and concurrent external beam radiation therapy. The majority of patients will experience a complete response to therapy. Some patients will have ongoing resolution of their cancer over the 6 months after completion of therapy. Therefore, the decision about whether a patient has had a complete response to therapy is not made until after 6 months following chemoradiation.

Patients who are considered primarily refractory to chemoradiation or those with locally recurrent disease can be cured with radical resection (i.e., APR). In addition to those experiencing a local recurrence, ~10% of patients may experience continued rectal incontinence due to the effects of radiation therapy and undergo a subsequent colostomy.

Metastatic anal cancer is considered incurable. It may respond well to carboplatin and paclitaxel. PD-1 antibodies have limited activity in patients with metastatic disease.

Acknowledgment

Robert J. Mayer contributed to this chapter in the prior edition and material from that chapter has been retained here.

FURTHER READING

André T et al: Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. N Engl J Med 383:2207, 2020.

Cercek A et al: PD-1 blockade in mismatch repair-deficient locally advanced rectal cancer. N Engl J Med 386:2363, 2022.

Colón-López V et al: Anal cancer risk among people with HIV infection in the United States. J Clin Oncol 36:68, 2018.

Dekker E et al: Colorectal cancer. Lancet 394:1467, 2019.

Eng C et al: Anal cancer: emerging standards in a rare disease. J Clin Oncol 40:2022.

Grothey A et al: Duration of adjuvant chemotherapy for stage III colon cancer. N Engl J Med 378:1177, 2019.

Inadomi JM: Screening for colorectal neoplasia. N Engl J Med 376:149, 2017.

Long H et al: Pathways of colorectal carcinogenesis. Gastroenterology 158:291, 2020.

Petrelli F et al: Prognostic survival associated with left-sided vs right-sided colon cancer. A systemic review and meta-analysis. JAMA Oncol 3:211, 2017.

Schrag D et al: Preoperative treatment of locally advanced rectal cancer. N Engl J Med 389:1631, 2023.

Siegel RL et al: Colorectal cancer statistics 2024. CA Cancer J Clin 12:49, 2024.

87

Tumors of the Liver and Biliary Tree

Josep M. Llovet

Liver cancer is the sixth most common cancer worldwide, the third leading cause of cancer-related deaths and the leading cause of death among cirrhotic patients. Liver cancer comprises a heterogeneous group of malignant tumors that range from hepatocellular carcinoma (HCC; ~85 cases), intrahepatic cholangiocarcinoma (iCCA; ~10%), and other malignancies, such as fibrolamellar HCC, mixed HCC-iCCA, epithelioid hemangiothelioma, and the pediatric cancer hepatoblastoma. The burden of liver cancer is increasing globally.

HEPATOCELLULAR CARCINOMA

EPIDEMIOLOGY AND RISK FACTORS

Overall, liver cancer accounts for 7% of all cancers (~900,000 new cases each year), and HCC represents 85% of primary liver cancers. The highest incidence rates of HCC occur in Asia and sub-Saharan Africa due to the high prevalence of hepatitis B virus (HBV) infection, with 20–35 cases per 100,000 inhabitants. Southern Europe and now North America have intermediate incidence rates (10 cases per 100,000), whereas Northern and Western Europe have low incidence rates of less than 5 cases per 100,000 inhabitants. In the United States, the incidence of liver cancer is 40,000 cases per year **(Fig. 87-1)**. HCC has a strong male preponderance, with a male-to-female ratio estimated to be 2.5:1. The incidence increases with age, reaching a peak at 65–70 years old. In Chinese and in black African populations (where vertical transmission of HBV occurs), the mean age is 40–50 years. By contrast, in Japan mean age in men is now around 75 years.

The main risk factors for HCC development are cirrhosis—an associated chronic liver damage caused by inflammation and fibrosis—of any etiology, chronic infection by HBV or hepatitis C virus (HCV) infection, alcohol abuse, metabolic syndrome, and hemochromatosis (associated to *HFE1* gene germline mutations) (Fig. 87-1). Cirrhotic patients represent 1% of the human population, and one-third of them will develop HCC during their lifetime. Long-term follow-up studies have established an annual risk of HCC development of 3–8% in HBV- or HCV-infected cirrhotic patients. HCC is less common (1–2% per year) in cirrhosis associated with alcohol, metabolic dysfunction–associated steatohepatitis (MASH, formerly known as nonalcoholic steatohepatitis or NASH), α_1-antitrypsin deficiency, autoimmune hepatitis, Wilson's disease, and cholestatic liver disorders. Predictors of liver cancer development among cirrhotic patients have been associated

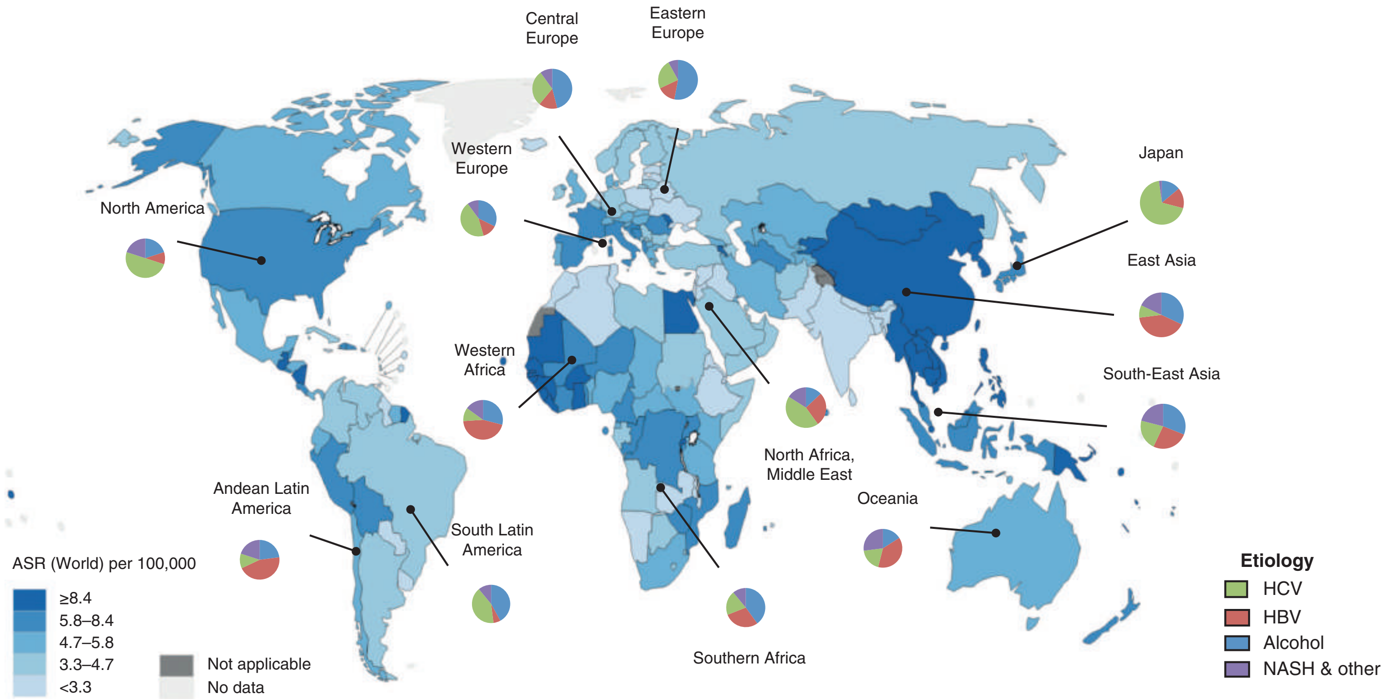

FIGURE 87-1 Distribution of hepatocellular carcinoma (HCC) incidence according to geographical area and etiology. HBV, hepatitis B virus; HCV, hepatitis C virus; NASH, nonalcoholic steatohepatitis. *(Reproduced with permission from JM Llovet et al: Hepatocellular carcinoma. Nat Rev Disease Primers 21:76, 2021.)*

with liver disease severity (platelet count of <100,000/μL, presence of portal hypertension), the degree of liver stiffness as measured by transient elastography, and liver gene signatures capturing the *cancer field effect*.

In terms of attributable risk fraction, HBV infection—a DNA virus that can cause insertional mutagenesis and affects ~300 million people globally—accounts for ~50% of HCC cases globally (60% in Asia and Africa and 20% in the Western world). Among patients with HBV infection, a family history of HCC, HBeAg seropositivity, high viral load, and genotype C are independent predictors of HCC development. HCV infection—an RNA virus that affects ~70 million people—is responsible for ~30% of cases and is the main cause of HCC in Europe and North America. Among patients with HCV infection, HCC occurs almost exclusively when relevant advanced liver damage and fibrosis are present, particularly if associated with HCV genotype 1b.

Alcohol consumption and metabolic syndrome due to diabetes and obesity are responsible for ~30% of cases. MASH is the fastest growing cause of cirrhosis in developed countries and currently represents ~15–20% of HCC cases in the West. The annual incidence of HCC in MASH-related cirrhosis (1–2%/year) justifies including cirrhotic patients in surveillance programs. Nonetheless, it has to be taken into account that 25–30% of MASH-associated HCCs occur in the absence of cirrhosis. A *PNPLA3* polymorphism is strongly associated with fatty and alcoholic chronic liver diseases and HCC occurrence. Finally, other cofactors contributing to HCC development in all etiologies are tobacco, aflatoxin B1 (a fungal carcinogen present in food supplies that induces *TP53* mutations), and aristolochic acid contained in Chinese medicine herbs.

MOLECULAR PATHOGENESIS

HCC development is a complex multistep process that starts with precancerous cirrhotic nodules, so-called low-grade dysplastic nodules (LGDN) that evolve to high-grade dysplastic nodules (HGDN) that can transform into early-stage HCC. Molecular studies support the pivotal role of adult hepatocytes as the cell of origin, either by directly transforming to HCC or by de-differentiating into hepatocyte precursor cells. Alternatively, progenitor cells also give rise to HCC with progenitor markers.

Genomic analysis has provided a clear picture of the main drivers responsible for HCC initiation and progression. This tumor results from the accumulation of around 40–60 somatic genomic alterations per tumor, among which 4–8 are considered driver cancer genes. HCC is a prototypical inflammation-associated cancer, where immune microenvironment and oxidative stress present in chronically damaged livers play pivotal roles in inducing mutations. In preneoplastic HGDN, mutations in telomere reverse transcriptase (*TERT*) gene (20% of cases) and gains in 8q have been described. Oncogenic transformation occurs upon additional genomic hits. The main molecular drivers of HCC are in the telomerase reverse transcriptase (*TERT*) promoter (56%), *TP53* (27%), and *CTNNB1* (26%), all of which are unactionable with molecular therapies. Genes commonly mutated in other solid tumors such as *EGFR*, *HER2*, *PIK3CA*, *BRAF*, or *KRAS* are rarely mutated in HCC (<5%) **(Table 87-1)**. Studies assessing copy-number alterations in HCCs have consistently identified: (1) high-level amplifications at 5–10% prevalence containing oncogenes in 11q13 (*CCND1* and *FGF19*) and 6p21 (*VEGFA*), *TERT* focal amplification, and homozygous deletion of *CDKN2A*; and (2) common amplifications containing *MYC* (8q gain). Overall, only ~20–25% of HCCs have at least one actionable mutation. Some risk factors have been associated with specific molecular aberrations. HBV integrates into the genome of driver genes, such as the *TERT* promoter, *MLL4*, and cyclin E1 (*CCNE1*). Alcohol abuse and HCV infection have been associated with *CTNNB1* mutations. *TP53* mutations are the most frequent alterations with a specific hotspot of mutation (R249S) in patients with aflatoxin B1 exposure.

Molecular and Immune Classes Genomic studies have revealed two molecular subclasses of HCC, each representing ~50% of patients. The proliferative subclass associated with poor outcomes, HBV-related etiologies, and overexpression of α-fetoprotein is enriched by activation of Ras, mammalian target of rapamycin (mTOR), and insulin-like growth factor (IGF) signaling and *FGF19* amplification. By contrast, the so-called nonproliferative subclass contains a subtype characterized by *CTNNB1* mutations and better outcome. Another classification based upon immune status has been proposed. It defines an inflamed HCC class in ~35% of cases (i.e., *hot* tumors), characterized by immune infiltrate with expression of PD-1/PD-L1, enrichment of T-cell activation, and better response to immunotherapies, and a noninflamed class (*cold* tumors), which includes the excluded subclass associated with activation of pathways related with immune escape (i.e., Wnt signaling) or absence of T-cell infiltrate. None of this molecular knowledge

TABLE 87-1 Molecular Aberrations Common in Hepatocellular Carcinoma (HCC)[a]

PATHWAY	TARGET	PREVALENCE (%)
Mutations		
Telomere stability	*TERT* promoter	56
p53/cell cycle control	*TP53*	27
	ATM	3
	RB1	3
Wnt/β-catenin signaling	*CTNNB1*	26
	AXIN1	5
Chromatin remodeling	*ARID1A*	6
	ARID2	7
	KMT2A	3
	KMT2C	3
Ras/PI3K/mTOR pathway	*RPS6KA3*	3
	TSC1/TSC2	3
Oxidative stress	*NFE2L2*	3
	KEAP1	3
High-level focal amplifications		
VEGF signaling	*VEGFA*	3
FGF signaling	*FGF19*	6
Cell-cycle control	CCND1 protein	7
Target with homozygous deletion		
TP53/cell-cycle control	*CDKN2A*	5
	TP53	4
	Retinoblastoma 1	4
Wnt/β-catenin signalling	*AXIN1*	3

[a]Recurrent mutations, focal amplifications, or homozygous deletions in HCC based on next-generation sequencing analyses.

has yet been translated into actual clinical benefits for any specific molecularly based subgroups, and thus, precision oncology is still an unmet goal of therapy.

PREVENTION AND EARLY DETECTION

Prevention Primary prevention of HCC can be achieved by vaccination against HBV and effective treatment of HBV and HCV infection. Universal vaccination against HBV infection is associated with a significant decrease of the incidence of HCC. Nowadays, HBV vaccination is recommended to all newborns and high-risk groups, following World Health Organization guidelines, and people with risk factors for acquiring HBV infection, such as health workers, travelers to areas where HBV infection is prevalent, injecting drug users, and people with multiple sex partners.

Effective antiviral treatments for patients with chronic HBV infection—achieving undetectable viral titers (circulating HBV-DNA)—result in 50–80% risk reduction of HCC development. Treatment of HCV with direct-acting antiviral agents (DAAs) yields >90% sustained virological response (SVR) rates after 12 weeks of treatment, thus significantly reducing HCC occurrence. Once cirrhosis is established, the incidence of HCC is lower for patients with SVR than for those with active viral disease, although they continue to have persistent HCC risk. Clinical practice guidelines recommend coffee consumption as a preventive strategy in patients with chronic liver disease. Aspirin, statins, and metformin have shown preventive effects but are not yet recommended as formal chemopreventive strategies.

Surveillance Surveillance programs aim to reduce cancer-related mortality. This is usually achieved through early detection that enhances the applicability and cost-effectiveness of curative therapies. U.S. and European guidelines recommend surveillance for patients at high risk for HCC on the basis of cost-effectiveness analyses.

Surveillance is recommended for cirrhotic patients owing to any cause, those with HCV-related advanced fibrosis, and patients with chronic HBV infection if Asian aged >40 years, African aged >20 years, family history of HCC, or patients with sufficient risk by risk scores such as PAGE-B. In terms of liver dysfunction, the presence of advanced cirrhosis (Child-Pugh class C) prevents potentially curative therapies from being employed, and thus surveillance is not recommended. As an exception, patients on the waiting list for liver transplantation, regardless of liver functional status, should be screened for HCC in order to detect tumors exceeding conventional criteria and to define priority policies for transplantation.

Ultrasonography every 6 months with serum α-fetoprotein (AFP) levels is the recommended method of surveillance. It has a sensitivity of ~65% and a specificity of >90% for early detection. A shorter follow-up interval (every 3–6 months) is recommended when a nodule of <1 cm has been detected. Computed tomography (CT) and magnetic resonance imaging (MRI) are not recommended as screening tools due to lack of data on accuracy, high cost, and possible harm (i.e., radiation with CT). Contrast-enhanced MRI can be considered in patients with obesity and fatty liver, where visualization with ultrasound is suboptimal. The accuracy of other serum biomarkers proposed, such as des-γ carboxyprothrombin (DCP) and the L3 fraction of AFP (AFP-L3), in early detection is not known.

Despite the fact that surveillance is cost-effective in HCC, the global implementation of such programs is estimated to engage ~50% of the target population in Europe and ~30% in the United States. Public health policies encouraging the implementation of such programs could lead to an increase in early tumor detection.

Diagnosis HCC is generally diagnosed at early or intermediate stages in Western countries but at advanced stages in most Asian (except Japan) and African countries. A surveillance program yields detection of early HCC in 70–80% of cases. At these stages, the tumor is asymptomatic, and diagnosis can be made by noninvasive (radiological) or invasive (biopsy) approaches. Without surveillance, HCC is discovered either as a radiological finding or due to cancer-related symptoms. If symptoms are present, the disease is already at an advanced stage, with a median life expectancy of <1 year. Symptoms include malaise, weight loss, anorexia, abdominal discomfort, or signs related to advanced liver dysfunction.

NONINVASIVE (RADIOLOGICAL) DIAGNOSIS Patients enrolled in a surveillance program are diagnosed by identification of a new liver nodule on abdominal ultrasound. Noninvasive criteria can only be applied to cirrhotic patients and are based on imaging techniques obtained by four-phase multidetector CT scan (four phases are unenhanced, arterial, venous, and delayed) or dynamic contrast-enhanced MRI. A flowchart of diagnosis and recall policy recommended by U.S. and European guidelines is summarized in Fig. 87-2. In nodules >1 cm or with AFP ≥20 ng/mL or rising AFP, multiphasic contrast-enhanced CT or MRI is recommended. Using these techniques, the typical hallmark of HCC consists of vascular uptake of the nodule in the arterial phase with washout in the portal venous or delayed phases. This radiological pattern captures the hypervascular nature characteristic of HCC and has a diagnostic specificity of ~95–100%, making a biopsy unnecessary. The Liver Imaging Reporting and Data System (LI-RADS) has been proposed as a way of classifying radiological findings. Essentially, nodules >10 mm visible on multiphase exams are assigned category codes reflecting their relative probability of being benign, HCC, or other hepatic malignant neoplasm. LI-RADS-1 lesions have a 0% probability of HCC, whereas lesions assigned to the LI-RADS-4 category are likely to be an HCC in 60–70% of cases and repeated imaging within 3–6 months or biopsy is recommended. Finally, LI-RADS-5 lesions have a ~95% probability of being HCCs. LI-RADS-M category comprises lesions that have malignant radiological features but are only HCC in ~35% of cases.

Nodules <1 cm in size are unlikely to be HCC and would be very difficult to diagnose, and thus ultrasound follow-up at 3–4 months is recommended. Contrast-enhanced ultrasound (CEUS) and angiography

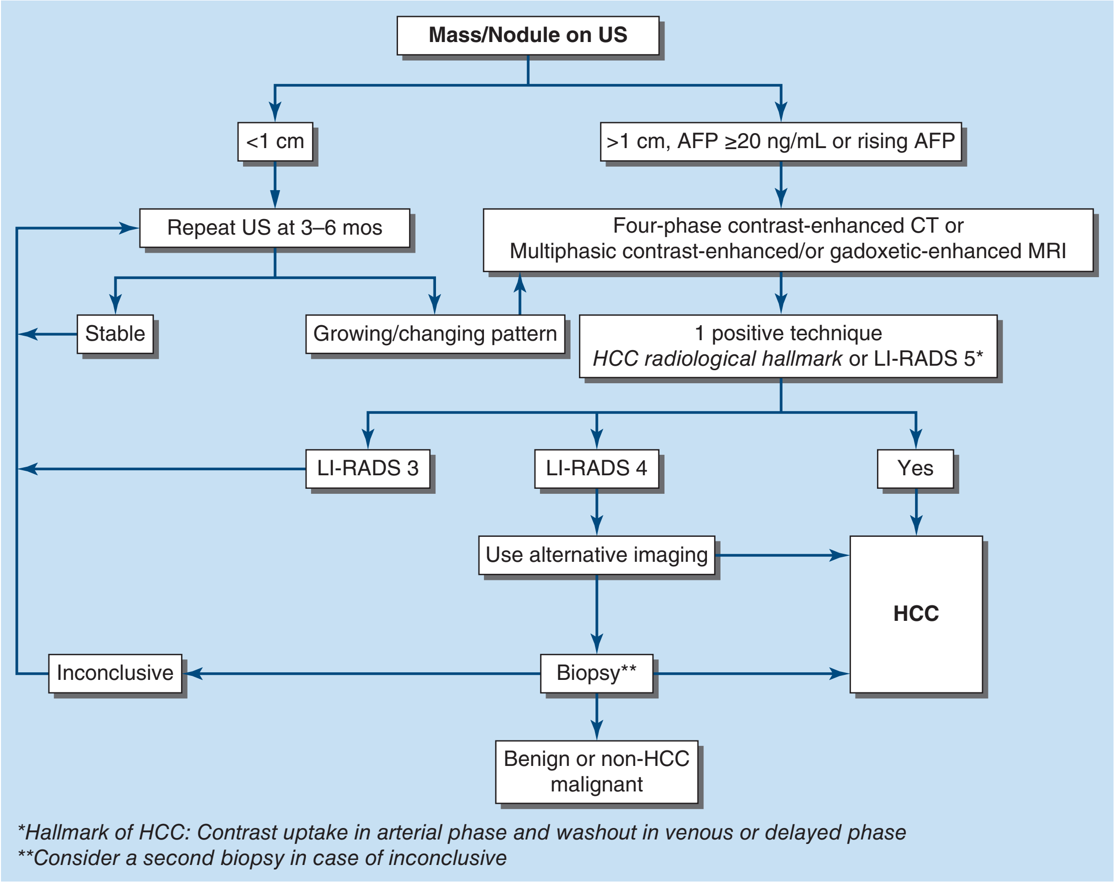
Mass/Nodule on US
<1 cm
>1 cm, AFP ≥20 ng/mL or rising AFP
Repeat US at 3–6 mos
Four-phase contrast-enhanced CT or
Multiphasic contrast-enhanced/or gadoxetic-enhanced MRI
Stable
Growing/changing pattern
1 positive technique
HCC radiological hallmark or LI-RADS 5*
LI-RADS 3
LI-RADS 4
Yes
Use alternative imaging
HCC
Inconclusive
Biopsy**
Benign or non-HCC malignant

**Hallmark of HCC: Contrast uptake in arterial phase and washout in venous or delayed phase*
***Consider a second biopsy in case of inconclusive*

FIGURE 87-2 Recall diagnosis schedule for HCC (EASL). EASL, European Association for the Study of Liver Disease; HCC, hepatocellular carcinoma. *(Modified with permission from EASL. J Hepatology 69:182, 2018.)*

are less accurate for HCC diagnosis. Positron emission tomography (PET) scan performs poorly for early diagnosis. AFP levels ≥400 ng/dL are highly suspicious but not diagnostic of HCC according to guidelines.

PATHOLOGICAL DIAGNOSIS Pathological diagnosis is required: (1) in patients without cirrhosis, (2) if radiology is not typical in at least one of two imaging techniques (CT and MRI), and (3) for a LI-RADS-4 lesion. Biopsy has not been used as the gold standard in clinical practice, although with the advent of molecular therapies, some guidelines advocate obtaining tissue samples in the setting of all research studies in HCC, even if radiological criteria are met. Sensitivity of liver biopsies ranges between 70 and 90% for all tumor sizes but decreases to <50% in tumors 1–2 cm in size. The risk of complications, such as tumor seeding and bleeding, after liver biopsy is ~2–3%. Biopsies should be assessed by an expert hepatopathologist. The use of special stains may help to resolve diagnostic uncertainties. Positive staining in two of four markers (glypican 3 [GPC3], glutamine synthetase, heat shock protein 70 [HSP70], and clathrin heavy chain) is highly specific for HCC. Additional staining can be considered to detect progenitor cell features (K19 and epithelial cell adhesion molecule [EpCAM]) or assess neovascularization (CD34). A negative biopsy does not eliminate the diagnosis of HCC. A second biopsy is recommended in case of inconclusive findings or growth or change in enhancement pattern identified during follow-up (Fig. 87-2).

■ TREATMENT

Overview The landscape of management of HCC has substantially changed during the last decade. For early stages, resection, liver transplantation, and local ablation have substantially improved life expectancy, with median overall survival (OS) times beyond 5 years **(Fig. 87-3)**. Adjuvant therapy with atezolizumab plus bevacizumab improves recurrence-free survival in patients at high risk of recurrence undergoing resection or ablation. For intermediate stages, transarterial chemoembolization (TACE) has improved the natural history of 16 months to ~30 months, and when combined with durvalumab plus bevacizumab, it improves progression-free survival. Systemic drugs for advanced tumors (atezolizumab plus bevacizumab, durvalumab plus tremelimumab, sorafenib, lenvatinib, regorafenib, cabozantinib, and ramucirumab) have improved median survivals from 8 months to even beyond 19 months in front-line and to 10 months in second-line treatment (Fig. 87-3).

Staging Systems and Treatment Allocation Staging systems are aimed at stratifying patients according to prognostic factors and outcome and allocating the best available therapies according to evidence. The most accepted staging system is the Barcelona Clinic Liver Cancer (BCLC) Classification, which is endorsed by U.S. and European clinical practice guidelines (Fig. 87-3). This staging system defines five prognostic subclasses and allocates specific treatments for each stage. The BCLC staging system has been externally validated by numerous studies. It is an evolving system that allows incorporation of new therapies and treatment-dependent variables as new evidence emerges. Several treatments improve survival in HCC, and thus have been incorporated in the therapeutic algorithm: surgical resection, liver transplantation, radiofrequency (RF) ablation, microwave, chemoembolization, and systemic therapies (atezolizumab-bevacizumab, durvalumab plus tremelimumab, sorafenib, lenvatinib, regorafenib, cabozantinib, and ramucirumab). The BCLC assigns each patient to a given treatment allocation. Treatment stage migration is also applied by this scheme, meaning that if patients are not candidates for the selected therapy, the next effective therapy at more advanced stages can be given.

In HCC, three parameters are relevant for defining treatment strategy: tumor status, cancer-related symptoms, and liver dysfunction. The BCLC staging captures all three variables and allocates patients to treatments according to evidence. Since >80% of patients have two diseases, HCC and cirrhosis, a clear measurement of liver dysfunction should be in place. The prognosis of chronic liver disease is commonly assessed using the Child-Pugh score, which uses five clinical

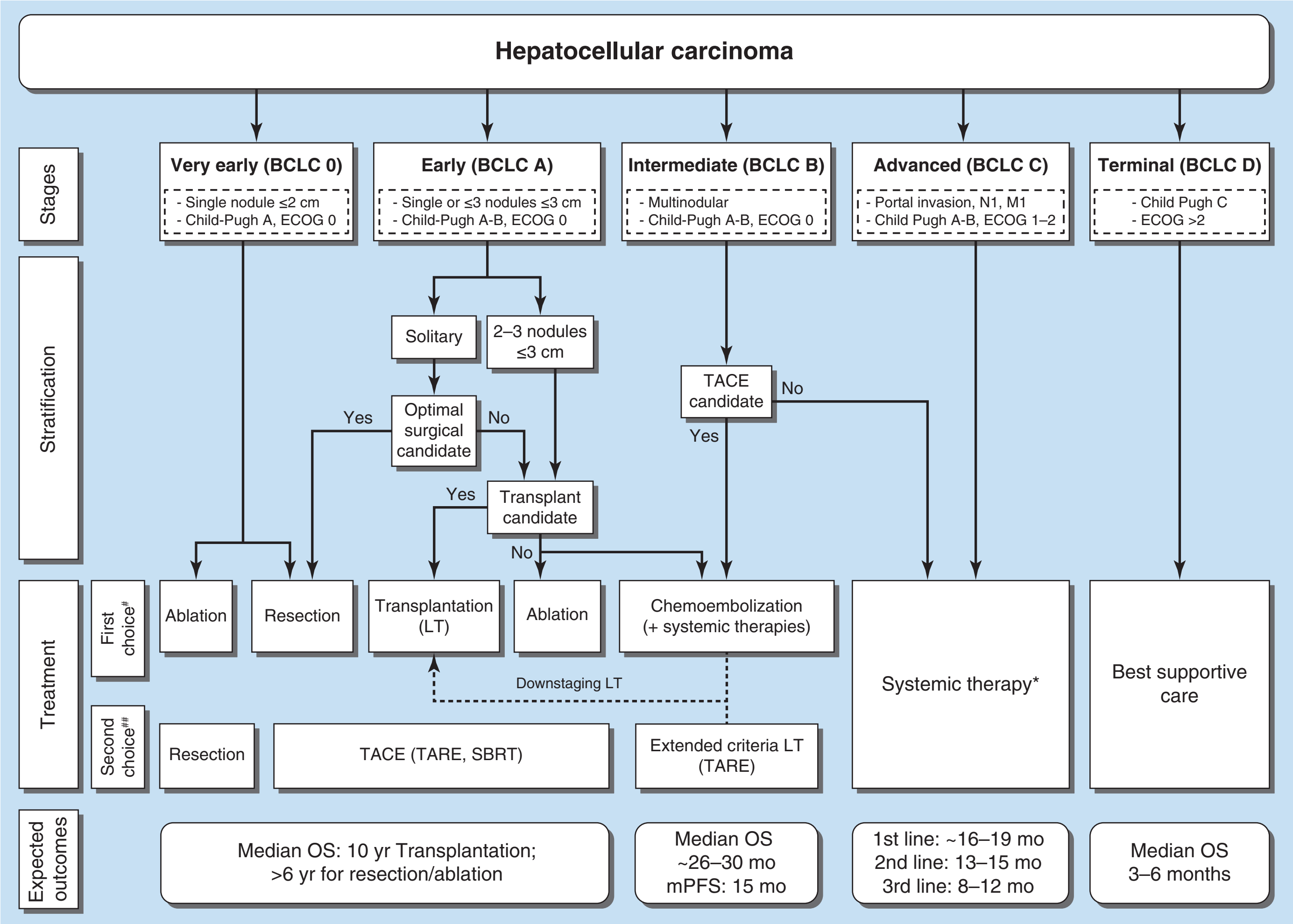
Hepatocellular carcinoma
Stages
Stratification
Treatment
Expected outcomes
Very early (BCLC 0)
- Single nodule ≤2 cm
- Child-Pugh A, ECOG 0
Early (BCLC A)
- Single or ≤3 nodules ≤3 cm
- Child-Pugh A-B, ECOG 0
Intermediate (BCLC B)
- Multinodular
- Child-Pugh A-B, ECOG 0
Advanced (BCLC C)
- Portal invasion, N1, M1
- Child Pugh A-B, ECOG 1–2
Terminal (BCLC D)
- Child Pugh C
- ECOG >2
Solitary
2–3 nodules ≤3 cm
Optimal surgical candidate
Yes
No
Transplant candidate
Yes
No
TACE candidate
No
Yes
First choice#
Second choice##
Ablation
Resection
Transplantation (LT)
Ablation
Chemoembolization (+ systemic therapies)
Downstaging LT
Systemic therapy*
Best supportive care
Resection
TACE (TARE, SBRT)
Extended criteria LT (TARE)
Median OS: 10 yr Transplantation; >6 yr for resection/ablation
Median OS ~26–30 mo mPFS: 15 mo
1st line: ~16–19 mo 2nd line: 13–15 mo 3rd line: 8–12 mo
Median OS 3–6 months

FIGURE 87-3 Staging system and therapeutic strategy. BCLC classification comprises five stages that select the best candidates for therapies according to evidence-based data. Patients with asymptomatic early tumors (stages 0 -A) are candidates for radical therapies (resection, transplantation, or local ablation). Asymptomatic patients with multinodular HCC (stage B) are suitable for transcatheter arterial chemoembolization (TACE), whereas patients with advanced symptomatic tumors and/or an invasive tumoral pattern (stage C) are candidates to receive systemic therapies. End-stage disease (stage D) includes patients with poor prognosis that should be treated by best supportive care. BCLC, Barcelona Clinic Liver Cancer; DDLT, deceased donor liver transplantation; EASL, European Association for the Study of Liver Disease; ECOG, Eastern Cooperative Oncology Group Performance Status; EORTC, European Organisation for Research and Treatment of Cancer; GRADE, grading of recommendations assessment, development, and evaluation; HCC, hepatocellular carcinoma; LDLT, living donor liver transplantation; OS, overall survival; PEI, percutaneous ethanol injection; RF, radiofrequency ablation; TACE, transcatheter arterial chemoembolization. *(Modified with permission from JM Llovet et al: Molecular pathogenesis and systemic therapies for hepatocellular carcinoma. Nature Cancer 3:386, 2022.)*

*Around 70–80% of patients are expected to receive this regimen

#Based on high level evidence studies.

##Based on low or moderate level of evidence studies.

measures—total bilirubin, serum albumin, prothrombin time, ascites severity, and hepatic encephalopathy grade—to classify patients into one of three groups (A–C) of predicted survival rates. In brief, Child-Pugh class A reflects well-preserved liver function, Child-Pugh class B moderate liver dysfunction, and Child-Pugh class C severe liver dysfunction. Other measurements of liver dysfunction, such as the Model for End-Stage Liver Disease (MELD) score or the albumin-bilirubin score, are not integrated in this staging system. Performance status is assessed by Eastern Cooperative Oncology Group (ECOG) scale (a 5-point scale where higher numbers indicate greater disability), and presence of cancer-related symptoms (ECOG 1–2) is considered a sign of advanced stage.

Considering all these prognostic/predictive variables and evidence-based treatment efficacy, five BCLC stages have been defined (Fig. 87-3). Patients with liver-only neoplastic disease, no symptoms (ECOG 0), and mild to moderate liver dysfunction (Child-Pugh A-B) can be classified as very early (stage 0), early (stage A), or intermediate (stage B) stage depending on tumor size and number. Very early HCC (BCLC 0) is defined by single tumors ≤2 cm (if pathology is available, the tumors should be well-differentiated with absence of microvascular invasion or satellites). Early HCC (BCLC A) includes either single tumors or a maximum of three nodules of ≤3 cm in diameter. Intermediate stage (BCLC B) is defined by all other liver-only tumors. Conversely, HCC is considered at advanced stages (BCLC C) when patients present with cancer-related symptoms (ECOG 1–2) or tumors with macrovascular invasion (of any type, including branch, hepatic, or portal vein), lymph node involvement, or extrahepatic spread. Finally, end-stage disease (BCLC D) is considered in cases of severe impairment of quality of life or severe cancer-related symptoms (ECOG 3–4) or severe liver dysfunction (Child-Pugh C).

Around 40% of patients are diagnosed at stages 0 and A and hence are eligible for potentially curative therapies, resection, transplantation, or local ablation. These treatments provide median survival rates of 60 months and beyond, which are in sharp contrast with outcomes of 36 months reported in historical controls (Fig. 87-3, Table 87-2). Adjuvant therapy with atezolizumab plus bevacizumab is recommended in patients after resection or local ablation who are at high risk of recurrence. Patients at intermediate stage (stage B) with preserved liver function have a documented natural history of around 16 months. These patients benefit from TACE, as reported in two randomized studies and one meta-analysis and achieve an estimated survival of 25–30 months. Combination of durvalumab plus bevacizumab with TACE

TABLE 87-2 Summary of Key Results of Randomized and Cohort Studies in the Management of Hepatocellular Carcinoma (HCC)

TREATMENT OF EARLY AND INTERMEDIATE STAGE HCC

TREATMENTS	HCC STAGE	TREATMENT ARMS	OUTCOMES (OS)
Treatments for early HCC			
Resection	Early	Optimal (single nodule; no portal hypertension)	5-year: 50–70%
		Suboptimal (multinodular or portal hypertension)	5-year: 35–55%
Resection + adjuvant	Early	Adjuvant atezolizumab + bevacizumab vs surveillance	12-month RFS: 78 vs 65%
Liver transplantation	Early	Milan (1 nodule <5 cm, 2–3 nodules ≤3 cm, no MVI, no EHS)	5-year: 70–80%
	Early/intermediate	Downstaged (1 nodule ≤6.5 cm, ≤3 nodules ≤4.5 cm and total diameter ≤8 cm, no MVI, no EHS)	5-year: 60–70%
Ablation	Early	RFA	Median: 50–60 months
Treatments for intermediate HCC			
Transarterial therapies	Intermediate	TACE	Median: 20–32 months Median PFS: 8 months
Locoregional + systemic	Intermediate	TACE + durvalumab + bevacizumab	Median PFS: 15.2 months

TREATMENT OF ADVANCED STAGE HCC

STUDY NAME	TREATMENT	MEDIAN OS, MONTHS (HR 95% CI)	MEDIAN PFS, MONTHS (HR 95% CI)	ORR mRECIST/RECIST
First-line therapies				
IMbrave150	Atezolizumab + bevacizumab	19.2 vs 13.5 (HR 0.66, 0.452–0.85)	6.9 (HR 0.65, 0.53–0.81)	35.4%/29.8%
HIMALAYA	Durvalumab + tremelimumab	16.4 (HR 0.78, 0.65–0.93)	3.7 (HR 0.90, 0.77–1.05)	NA/20%
SHARP	Sorafenib	10.7 (HR 0.69, 0.55–0.87)	10.7 (HR 0.69, 0.55–0.87)	NA/2%
REFLECT	Lenvatinib	13.6 (HR 0.92, 0.79–1.06)	7.4 (HR 0.66, 0.57–0.77)	24.1%/18.8%
Second-line therapies				
RESORCE	Regorafenib	10.6 (HR 0.63, 0.5–0.79)	3.1 (HR 0.46, 0.37–0.56)	11%/7%
CELESTIAL	Cabozantinib	10.2 (HR 0.76, 0.63–0.92)	5.2 (HR 0.44, 0.36–0.52)	NA/4%
REACH-2	Ramucirumab	8.5 (HR 0.71, 0.53–0.95)	2.8 (HR 0.45, 0.34–0.6)	NA/5%

Abbreviations: CI, confidence interval; EHS, extrahepatic spread; HCC, hepatocellular carcinoma; HR, hazard ratio; mRECIST, modified Response Evaluation Criteria in Solid Tumors; MVI, microvascular invasion; NA, not available; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; RFA, radiofrequency ablation; TACE, transarterial chemoembolization.

has shown benefits in progression-free survival compared with TACE alone. Patients progressing on TACE or at advanced stage (stage C) benefit from systemic treatments. First-line therapy for advanced HCC includes atezolizumab plus bevacizumab or durvalumab plus tremelimumab. Both combinations were superior to sorafenib in phase 3 trials. In patients with contraindications for immunotherapies, both sorafenib or lenvatinib are recommended Three additional targeted therapies have shown to improve survival compared to placebo in patients with HCC progressing on sorafenib: regorafenib, cabozantinib, and ramucirumab (only in patients with AFP >400 ng/mL). Therefore, these treatments have been adopted by guidelines and incorporated into the treatment algorithm. Patients with end-stage disease (BCLC D) should be considered for nutritional and psychological support and proper management of pain.

Although the BCLC establishes validated stages and treatment assignment according to evidence, clinical practice is not always aligned with this classification. In large cohort studies and surveys, only half of patients, or even less in Asia, are treated accordingly. Alternative staging or scoring systems have been proposed, such as the Hong Kong classification or the Japan Integrated Staging score. These systems capture extended indications for resection and TACE applied in clinical practice in Asia. Finally, the tumor-node-metastasis (TNM) staging system is not used in HCC since it does not incorporate the main prognostic variables related to liver function and performance status.

Due to the complexities of HCC diagnosis and management, it is recommended to refer patients to centers with multidisciplinary liver cancer programs that include a hepatologist, oncologist, hepatobiliary and transplant surgeons, interventional and body imaging radiologist, hepatopathologist, and specialized nurses.

SURGICAL THERAPIES

Resection Surgical resection is the first-line option for noncirrhotic patients with early-stage HCC (BCLC 0 or A) with solitary tumors (Fig. 87-3). In cirrhotic patients, ablation competes with resection for BCLC 0 tumors (<2 cm in diameter). Which is better is not defined. Cost-effectiveness approaches report a benefit for local ablation with RF. For single tumors >2 cm (BCLC A), resection remains the mainstay of treatment in patients with Child-Pugh A with normal bilirubin and absence of portal hypertension (no esophageal/gastric varices or platelet count >100,000/μL associated with splenomegaly). Anatomic resections following the functional segments of the liver are recommended to spare uninvolved liver parenchyma and to remove satellite tumors. Applying these criteria, resection is associated with perioperative decompensation rate of 5%, perioperative mortality of <1%, and 5-year survival of 60–70%, as opposed to ~35–55% for suboptimal candidates (Table 87-2). Macrovascular invasion, extrahepatic involvement, and liver dysfunction (Child-Pugh B-C) are major contraindications for resection.

Adjuvant Treatments Tumor recurrence represents a major complication of resection and local ablation (with 5-year rates of 50–70%). Predictors of recurrence are tumor size, tumor number, presence of microsatellites, or microvascular invasion at the specimen analysis. Most recurrences are intrahepatic metastases, but at least one-third are considered de novo tumors, new clones developing in the cirrhotic carcinogenic field. The type of recurrence can only be defined by molecular studies. An adjuvant regimen after resection or local ablation with atezolizumab plus bevacizumab for 12 months significantly improved recurrence-free survival in patients at high risk of recurrence and has been incorporated into the guidelines of practice.

Liver Transplantation Liver transplantation is the first treatment choice for cirrhotic patients with single tumors ≤5 cm and portal hypertension (including Child-Pugh B and C) or with small multinodular tumors (three or fewer nodules, each ≤3 cm) (Fig. 87-3). These so-called Milan criteria have been validated over the years and lead to median survival times of 10 years. Perioperative mortality rates have been reduced to <3%. Transplantation simultaneously cures the tumor and the underlying cirrhosis, and it is associated with a low risk of recurrence, around 10–15% at 5 years. No immunosuppressive regimens or antitumor therapies after transplantation have demonstrated any preventive effect on recurrence. Milan criteria are integrated in the treatment strategy (BCLC 0 and A) and have also been adopted by the United Network for Organ Sharing (UNOS) pretransplant staging for organ allocation in the United States (stage T2). Aside from size and number, conventional contraindications for organ transplantation procedures (e.g., ABO incompatibility, comorbidities) are applied in this setting.

Liver transplantation has a couple of important limitations, such as cost and donor availability, that limit this procedure to <5% of HCC cases worldwide. The scarcity of donors represents a major drawback of liver transplantation. Donor scarcity varies geographically, and deceased liver donation is almost zero in some Asian countries. Due to the shortage of donors, median waiting times in Western programs is ~6–12 months, leading to 20% of candidates dropping off the list due to tumor progression before receiving the procedure. Neoadjuvant treatments with locoregional therapies are recommended when the waiting time exceeds 6 months.

Expansion of Milan criteria by using locoregional therapies to effectively downstage the tumor (i.e., UNOS-downstaging criteria) have reported good results. Since policies for enhancing organ donation have reached a ceiling during the past several years, alternatives to donation have emerged. Living donor liver transplantation represents a plausible alternative that accounts for ~5% of total transplantations performed globally. Outcomes reported are similar to those with deceased liver donors, and it is recommended as an alternative option in patients on a waiting list exceeding 6 months. The risks and benefits of this procedure should take into account both donor (death is estimated in 0.3%) and recipient, a concept known as *double equipoise*. Due to the complexity of this treatment, it must be restricted to centers of excellence in hepatobiliary surgery and transplantation.

■ LOCOREGIONAL THERAPIES

Local Ablation Thermal ablation with RF or microwave (MWA) is recommended as the primary ablative technique **(Fig. 87-4)**. The energy generated by RF ablation (heating of tissue at 80°–100°C) induces coagulative necrosis of the tumor, producing a *safety ring* in the peritumoral tissue, which might eliminate small undetected satellites. Treatment consists of one or two sessions performed using a percutaneous approach, although in some instances, ablation with laparoscopy is needed. HCC patients treated by RF ablation have 5-year survival rates of ~60% (Table 87-2). In tumors <2 cm, both RF and MWA ablation achieve complete responses in 90–100% of cases with good long-term outcome and compete with resection in terms of cost-effectiveness as a first-line option. For BCLC A cases, local ablation techniques are considered in patients with three tumors <3 cm in diameter unsuitable for surgical therapies.

For patients with unresectable HCC >3 cm in diameter who are not candidates for liver transplantation, transarterial radioembolization (TARE) with yttrium-90 and stereotactic body radiation (SBRT) are considered alternative options based on propensity-matched score studies and phase 2 investigations.

Chemoembolization TACE is the most widely used primary treatment for unresectable HCC worldwide and the first-line indication for patients with intermediate BCLC B stage (Fig. 87-3). Conventional chemoembolization (c-TACE) consists of local hepatic artery administration of chemotherapy (either doxorubicin 50 mg/m^2 or cisplatin) mixed with an emulsion of lipiodol followed by obstruction of the feeding artery with sponge particles. The best randomized phase 3 investigations have provided median survivals for TACE of 20–30 months in properly selected populations (compared to 16 months for pooled control arms). Median objective response rates are 50–70%. In randomized studies, the treatment is either performed at a regular schedule of 0, 2, and 6 months (median number of sessions is three) or on demand according to tumor response. TACE procedures should be stopped upon tumor progression or any other contraindication. Around 50% of patients present with a limited postembolization syndrome of fever and abdominal pain related to ischemic injury and release of cytokines. Less than 5% of patients have major complications (liver abscess, ischemic cholecystitis, or liver failure), and

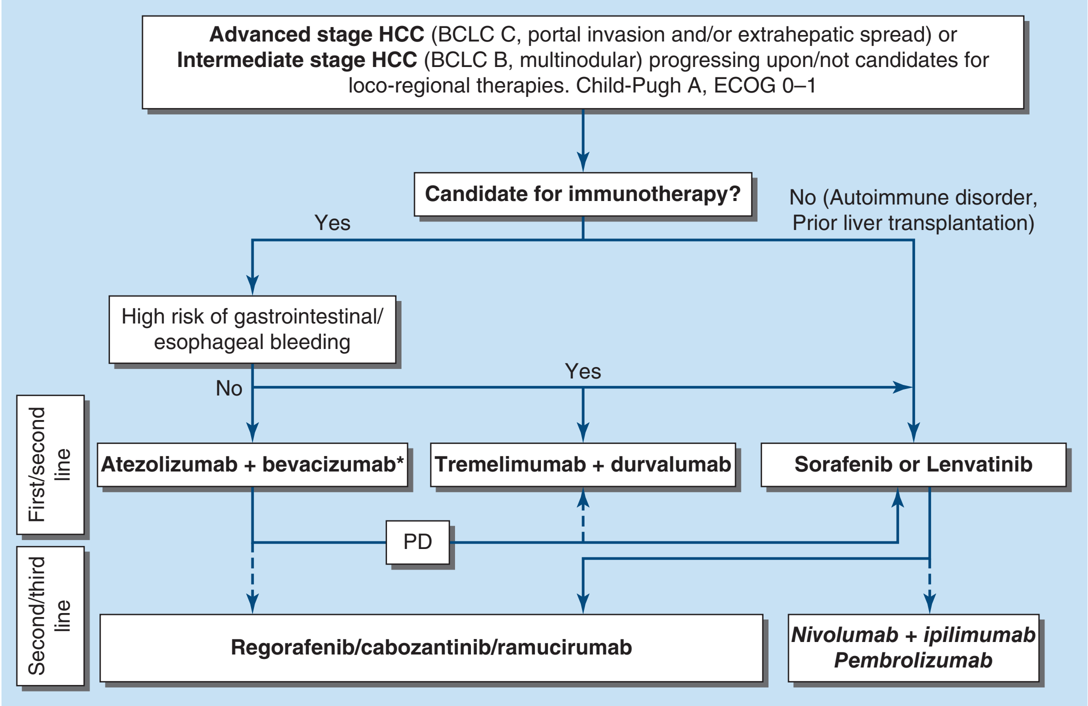
Advanced stage HCC (BCLC C, portal invasion and/or extrahepatic spread) or
Intermediate stage HCC (BCLC B, multinodular) progressing upon/not candidates for loco-regional therapies. Child-Pugh A, ECOG 0–1
Candidate for immunotherapy?
Yes
No (Autoimmune disorder, Prior liver transplantation)
High risk of gastrointestinal/ esophageal bleeding
Yes
No
First/second line
Atezolizumab + bevacizumab*
Tremelimumab + durvalumab
Sorafenib or Lenvatinib
PD
Second/third line
Regorafenib/cabozantinib/ramucirumab
Nivolumab + ipilimumab
Pembrolizumab

FIGURE 87-4 Treatment strategy for advanced hepatocellular carcinoma with systemic therapies. Drugs in bold have positive results from phase 3 trials with regulatory approval (atezolizumab plus bevacizumab, durvalumab plus tremelimumab, sorafenib, lenvatinib, regorafenib, cabozantinib, and ramucirumab). Drugs in bold italic have received accelerated approval from the Food and Drug Administration on the basis of promising efficacy results in phase 2 trials in second line (pembrolizumab and nivolumab ipilimumab). Key details of the patient populations are provided. BCLC, Barcelona Clinic Liver Cancer (classification); ECOG, Eastern Cooperative Oncology Group; PD, progressive disease. *(Modified from JM Llovet et al: Nat Rev Clin Oncol 15:599, 2018.)*

in <2% of cases, treatment-related death occurs. Drug-eluting bead chemoembolization (DEB-TACE) differs from c-TACE in the use of more standardized embolic spheres of regular size embedded with chemotherapy. This strategy achieves similar antitumor activity (objective responses of ~60%) as c-TACE and is associated with significantly fewer systemic toxic effects and better patient tolerance, but there are no clear differences in clinical outcomes.

Overall, TACE can only be applied to 50% of patients at intermediate stage, mostly as a result of the presence of liver failure (Child B or ascites or encephalopathy), technical contraindications to the procedure (i.e., impaired portal vein blood flow), or infiltrative/massive tumor burden (i.e., generally main tumor size >10 cm). Super-selective TACE minimizes the ischemic insult to nontumor tissue. According to guidelines, treatment stage migration allows performing TACE on patients at early stages not suitable for surgical or ablative therapies. TACE plus durvalumab plus bevacizumab significantly increases progression-free survival compared to TACE alone (median progression-free survival, 15.2 vs 8 months) with manageable treatment-related adverse events.

Radioembolization and Other Intra-arterial Therapies Transarterial radioembolization (TARE) using beads coated with yttrium-90 (Y-90)—an isotope that emits short-range β radiation—is the most promising alternative to TACE. Several phase II studies reported objective responses and overall outcome with a safe profile similar to TACE, and thus TARE has been endorsed by clinical practice guidelines. Radioembolization requires prevention of severe lung shunting and intestinal radiation before the procedure. Around 20% of patients experience liver-related toxicity, and <2% experience treatment-related death. Due to the minimally embolic effect of Y-90 microspheres, treatment can be safely used in patients with portal vein thrombosis, a setting where survival results in phase II studies were encouraging.

SYSTEMIC THERAPIES

Approximately 50–60% of patients with HCC are currently exposed to systemic therapies during their life span, either because they have been diagnosed at advanced stages or because they have progressed after locoregional therapies. In 2007, a phase III trial demonstrated survival benefits for patients with advanced-stage disease treated with sorafenib, thus becoming the first systemic therapy for HCC. Subsequently atezolizumab plus bevacizumab and durvalumab plus tremelimumab have demonstrated survival superiority compared to sorafenib, whereas lenvatinib showed noninferior effects (Fig. 87-4). In Asia, other combinations (i.e., carmelizumab plus rivoceranib) also were superior versus sorafenib in terms of survival. Three additional therapies, regorafenib, cabozantinib, and ramucirumab (only in patients with AFP >400 ng/mL), have been shown to benefit patients progressing on sorafenib. Of note, classical chemotherapy and radiotherapy have not resulted in benefits in survival. Similarly, due to the low prevalence of actionable molecular aberrations, precision oncology regimens are not yet available in HCC.

First-Line Therapies Atezolizumab (anti–PD-1 checkpoint inhibitor) plus bevacizumab (monoclonal antibody against VEGF-A) demonstrated survival differences compared to sorafenib (median OS, 19.2 vs 13.4 months) and has become the standard of care in first-line treatment (Fig. 87-4, Table 87-2). This combination treatment resulted in improved progression-free survival, patient-reported outcomes reflecting quality of life, and objective response rates (30 vs 12% for sorafenib). The combination had fewer adverse events compared to sorafenib (grade 3-4 adverse events, 36 vs 50%, respectively). The most common side effects associated with the combination are hypertension, proteinuria, and low-grade diarrhea, whereas autoimmune events are infrequent and manageable. Treatment-related adverse event rate leading to discontinuation of any drugs is 15%. Screening for varices with upper gastrointestinal endoscopies has become standard before first-line therapy in advanced HCC to mitigate the risk of bleeding associated with bevacizumab. In cases with varices, the use of one session of banding or carvedilol is recommended.

The combination of tremelimumab (a CTLA-4 inhibitor) and durvalumab (a PD-L1 inhibitor) (STRIDE regimen) has demonstrated a survival advantage versus sorafenib (median OS, 16.4 vs 13.7 months for sorafenib). No differences were identified regarding progression-free survival, and the response rate was 20.1% with the combination versus 5.1% with sorafenib. This regimen can be administered even in patients with portal hypertension or varices. STRIDE was associated with more immune-related adverse events, and 20% of patients required glucocorticoid treatments (Fig. 87-4).

Alternatively, sorafenib or lenvatinib is indicated for patients with advanced HCC with contraindications for immunotherapies (i.e., due to autoimmune disease or liver transplantation) (Fig. 87-4, Table 87-2). A phase 3 study comparing sorafenib (a multikinase inhibitor) versus placebo showed increased survival from 7.9 months to 10.7 months (hazard ratio [HR], 0.69; 31% reduction in risk of death). Patients with HCV-related HCC achieve significantly better outcomes with sorafenib, with a median survival of 14 months. Median treatment duration is about 6 months. Treatment is associated with manageable adverse events, such as diarrhea, hand-foot skin reactions, fatigue, and hypertension. These toxicities lead to treatment discontinuation in 15% of patients and dose reduction in up to half. It has been estimated that this therapy cannot be administered to approximately one-third of the targeted patients due to toxicity, advanced age, or liver failure (ascites or encephalopathy). Active vascular disease, either coronary or peripheral, is considered a formal contraindication. Median time to progression on sorafenib is 4–5 months in phase 3 trials.

Another alternative to sorafenib is the multikinase inhibitor lenvatinib, which was shown to be noninferior in a phase 3 investigation (Fig. 87-4). A phase 3 study comparing lenvatinib (an inhibitor of vascular endothelial growth factor receptor [VEGFR], fibroblast growth factor recepctor [FGFR], platelet-derived growth factor receptor [PDGFR], RET, and c-Kit) with sorafenib showed noninferiority of results in terms of OS (13.6 vs 12.3 months; HR, 0.92). Lenvatinib induces objective responses in 24% of cases. The main side effects are hypertension, proteinuria, asthenia, diarrhea, and weight loss. This treatment is associated with a 55% rate of grade 3-4 drug-related adverse events, resulting in a ~15% withdrawal rate.

Second-Line Therapies Three drugs (regorafenib, cabozantinib, and ramucirumab) have shown survival benefits versus placebo in patients progressing to sorafenib, and two additional treatments have been approved by the U.S. Food and Drug Administration (FDA) based on promising phase 2 data (pembrolizumab, and nivolumab plus ipilimumab) (Fig. 87-4). It is estimated that only half of patients progressing on sorafenib can be considered for second-line therapies, and their median survival with no treatment is 7–8 months (obtained from patients allocated to the placebo arm).

A phase 3 study comparing regorafenib (a more potent multikinase inhibitor than sorafenib, but targeting similar kinases) versus placebo in patients progressing to sorafenib reported an increase in survival from 7.8 to 10.6 months (HR, 0.62; 38% reduction in risk of death) **(Fig. 87-5)**. Response rate was 10% based on modified Response Evaluation Criteria in Solid Tumors. Median time on treatment was 3.5 months. Prevalence of toxicity (hand-foot reaction, fatigue, and hypertension) was higher compared with reported toxicity from sorafenib, but adverse events only led to treatment discontinuation in 10% of cases. Cabozantinib, a multikinase VEGFR inhibitor with activity against both AXL and cMET, improves survival compared to placebo after progression to sorafenib (10.2 months for cabozantinib vs 8.0 months in the placebo arm; HR, 0.76). Toxicity was manageable, with the most common grade 3–4 events being palmar-plantar erythrodysesthesia, hypertension, increased aspartate aminotransferase level, fatigue, and diarrhea. Ramucirumab, an anti–VEGFR-2 monoclonal antibody, is the only biomarker-guided therapy in HCC based on AFP levels. The randomized, placebo-controlled, phase 3 REACH-2 study studied patients with advanced HCC in second-line treatment with baseline AFP ≥400 ng/dL. This trial demonstrated positive survival results, and a further meta-analysis established a median survival for ramucirumab of 8.1 months compared to 5 months for patients receiving placebo. The most common grade 3–4 treatment-related adverse events were hypertension, hyponatremia, and increased aspartate

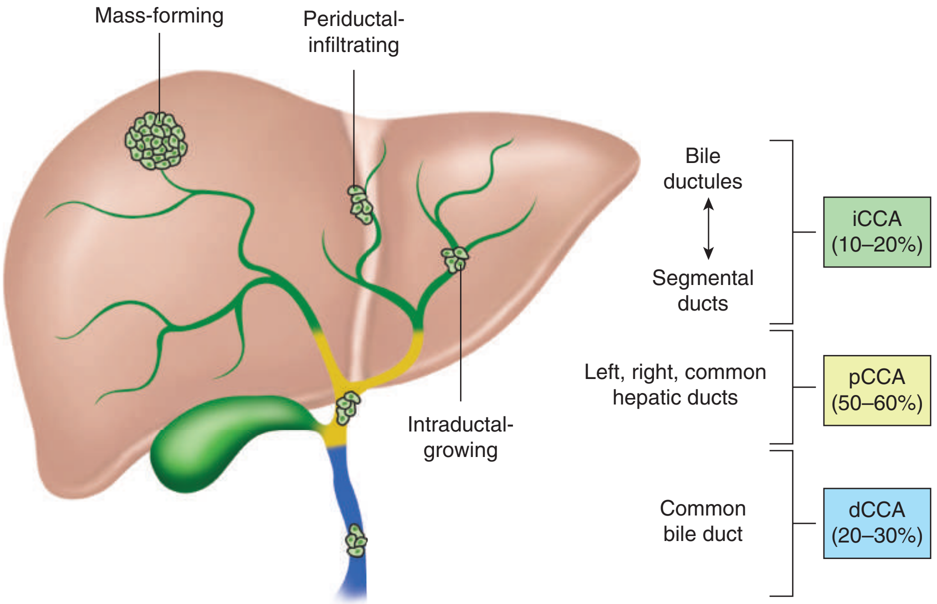

FIGURE 87-5 Anatomical classification of cholangiocarcinoma. Cholangiocarcinoma is classified as intrahepatic (iCCA) and extrahepatic (eCCA). eCCA can be subclassified as perihilar (pCCA) and distal (dCCA). *(Reproduced from JM Banales et al: Cholangiocarcinoma 2020: The next horizon in mechanisms and management. Nat Rev Gastroenterol Hepatol 17:557, 2020.)*

aminotransferase. Patients progressing after second-line therapy might be considered for third-line approaches. Patients with tumors at a BCLC D stage should receive best supportive palliative care, including management of pain, nutrition, and psychological support.

CHOLANGIOCARCINOMA

Cholangiocarcinoma (CCA) is classified according to its anatomic location as intrahepatic (iCCA; ~20–30%), perihilar (pCCA; ~50–60%), and distal (dCCA: ~20–30%). The latter two are also known as extrahepatic cholangiocarcinomas (eCCA), with the second-order bile ducts acting as the separation point (Fig. 87-5). The three subtypes of CCA differ in their anatomic location, epidemiology and risk factors, cell of origin, pathogenesis, and treatment. iCCA originates from adult cholangiocytes, transdifferentiation of adult hepatocytes, and hepatic progenitor cell–cholangiocyte precursors **(Fig. 87-6)**, as opposed to HCC, which originates only from hepatic progenitor cells or adult hepatocytes. Mixed HCC-iCCA originates from hepatic progenitor cells, whereas eCCA arises from the biliary epithelium and peribiliary glands. Moreover, their mutational profile also differs. *FGFR2* fusions and *IDH1/2* mutations mostly occur in iCCA, whereas *ERBB2/3* amplifications and *SMAD4* aberrations are characteristic of eCCA. iCCA has been recognized as a distinct entity with specific clinical practice guidelines, which should be tailored according to each biological/anatomical subtype of CCA.

EPIDEMIOLOGY, RISK FACTORS, AND MOLECULAR TRAITS

CCA is the second most common liver cancer following HCC, with a 5-year survival of 10%. iCCA has globally increasing incidence and mortality rates. The incidence of iCCA varies according to exposure to risk factors, ranging from 1–2 cases per 100,000 inhabitants in Europe and North America to the highest incidence in some areas of Southeast Asia, particularly in Thailand (>80 cases/100,000 inhabitants). The male-to-female ratio is 1.2:1. Only 30% of patients with CCA have a known risk factor. The classical risk factors for CCA development include primary sclerosing cholangitis (PSC), biliary duct cysts, hepatolithiasis, and Caroli's disease. Parasitic biliary infestation with flukes (most common are *Opisthorchis viverrini* and *Clonorchis sinensis*) is a prevalent etiology in Asia that can be prevented with an antihelminth therapy, praziquantel. PSC is a clear risk factor for iCCA and pCCA development, with a lifetime incidence ranging from 5 to 10%. Surveillance in PSC patients is recommended with annual imaging techniques and CA 19-9 serum determination. Common risk factors for HCC, such as HBV and HCV infection and cirrhosis, have been associated with iCCA development. More recently, sweetened beverages were reported to constitute a risk factor in the development of eCCA and gallbladder carcinoma (GBC) in a population cohort study.

Molecular Classification and Actionable Drivers A morphological and molecular classification subclassifies iCCA into the large duct type, which is associated with *IDH* (15–20%) and *FGFR2* (15%) mutations and better prognosis, and the small duct type, which is associated with *KRAS* (15%) and *SMARD4* (<5%) mutations and poor prognosis. Overall, up to 40% of iCCAs have a targetable mutation. Similarly, a molecular classification of eCCA has been proposed, dividing tumors into four categories (metabolic, proliferation, mesenchymal, and immune) based on molecular traits. It has been suggested that the proliferation class with enrichment of *ERBB2/3* mutations might respond to monoclonal antibodies against this receptor, while the immune class might respond to checkpoint inhibitors, a fact that needs clinical confirmation. The most common mutations of pCCA/dCCA are *P53* (~30%) and *KRAS* (~25%), whereas *ERBB2* amplifications (~20%) are common in gallbladder cancer.

INTRAHEPATIC CHOLANGIOCARCINOMA

Surveillance, Diagnosis, and Staging Guidelines currently only recommend surveillance for early diagnosis in the following at-risk subpopulations: (1) patients with primary biliary sclerosis (PBS; surveillance is recommended with CA 19-9 and magnetic resonance cholangiopancreatography [MRCP] every 12 months) and (2) patients with cirrhosis or those infected with liver flukes (surveillance is recommended with abdominal ultrasound every 6 months). Otherwise, incidental diagnosis occurs due to cross-sectional imaging performed for other reasons. In most cases, iCCA is diagnosed at advanced stages when symptoms such as weight loss, malaise, abdominal discomfort, or jaundice are present. Diagnosis of iCCA requires pathological confirmation. Differential diagnosis should be established with metastatic adenocarcinoma (i.e., colorectal, breast, and lung cancer) and mixed iCCA-HCC tumors. Immunohistochemistry using K7, K19, and K20 is useful to confirm iCCA and to distinguish it from metastatic liver cancer. Hepatocytic markers such as Hep-Par-1, GPC3, and HSP70 may aid in pointing to a mixed HCC-iCCA tumor. Current guidelines recommend the use of abdominal MRI, chest and abdomen CT scan, and PET scan for establishing the disease extension once the pathological diagnosis has been confirmed. Meta-analysis has defined a role for PET scanning in identifying lymph node metastasis in patients with no apparent lymph node invasion with MRI and/or CT scan. Lymph node sampling by endoscopic ultrasound with fine-needle aspiration would be considered before resection in selected unclear cases. Tumor biomarker CA 19-9 at a cutoff level of 100 U/mL has prognostic significance but lacks accuracy (sensitivity and specificity of ~60%) for early diagnosis.

TREATMENT

The European Association for the Study of Liver Disease–International Liver Cancer Association (ILCA) guidelines for management of iCCA proposed an updated treatment algorithm, which has been adapted to the current accepted treatment modalities **(Fig. 87-7)**. iCCA can be classified as early, intermediate, or advanced cases according to size of the nodules and invasion of lymph nodes (N1), metastasis, and ECOG

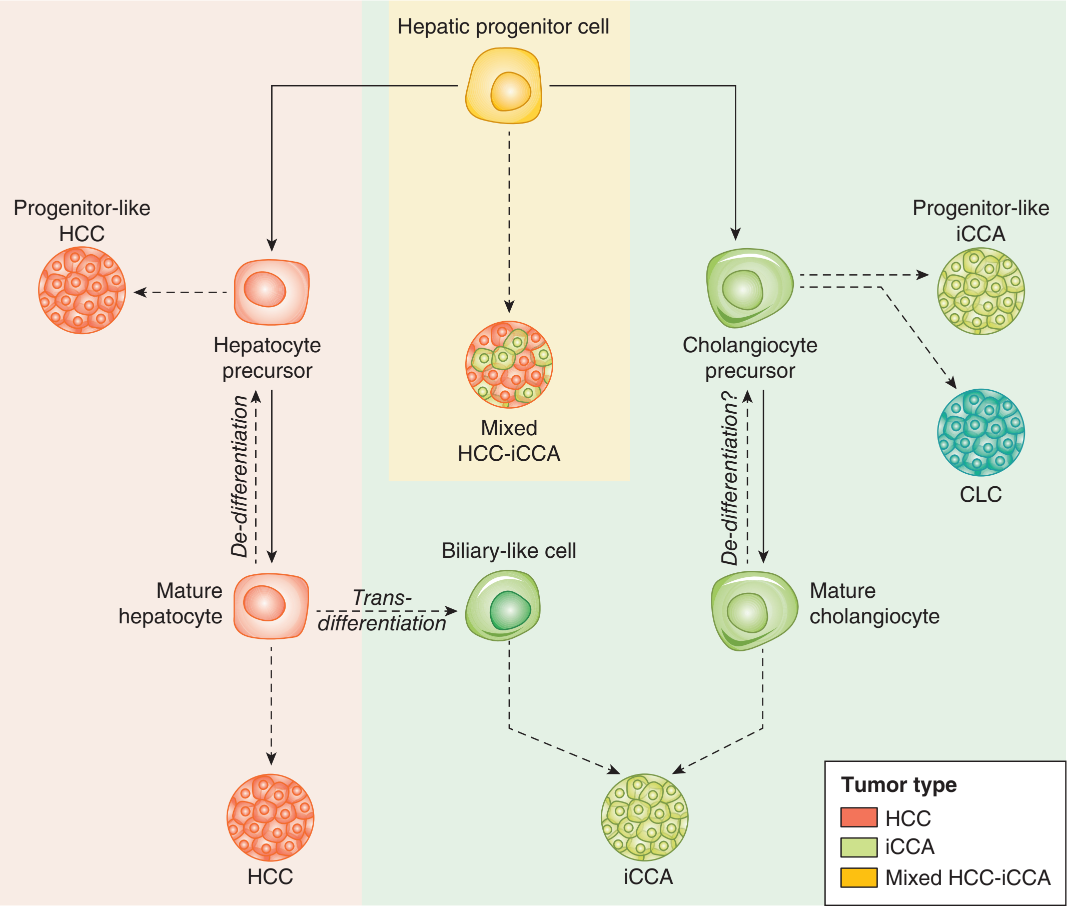

FIGURE 87-6 Cell of origin of liver cancer. Hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (iCCA) can develop from the neoplastic transformation of mature hepatocytes and cholangiocytes, respectively. There is evidence showing that hepatic progenitor cells (HPCs), their intermediate states, or dedifferentiated hepatocytes can originate liver cancers with progenitor-like features, including mixed HCC-CCA (e.g., cholangiolocellular carcinoma [CLC]). Mature hepatocytes can be also reprogrammed into cells that closely resemble biliary epithelial cells and induce the onset of iCCA. *(Printed with permission from © Mount Sinai Health System.)*

performance status and liver dysfunction. Approximately 30–40% of iCCA cases are deemed resectable, and the median reported survival for single tumors is ~40–50 months, whereas survival decreases to ~20 months in intermediate/multinodular resectable tumors. The main predictors of recurrence (~50–60% at 3 years) and survival are identified at the pathological examination, including presence of vascular invasion, lymph node metastases, and poor differentiation degree. In terms of adjuvant therapy, a phase 3 trial (BILCAP trial) including all types of CCA in a prespecified per-protocol analysis reported improved survival (53 vs 36 months; adjusted HR, 0.75). Based on this trial, American Society of Clinical Oncology guidelines recommend adjuvant capecitabine for a period of 6 months. A small proportion of patients with early/intermediate tumors can have contraindications for resection, particularly in cirrhotic patients, and should be first considered for local ablation or even liver transplantation if the diameter of the main tumor is <2 cm.

Nonsurgical candidates have a dismal life expectancy. Overall, patients with multinodular unresectable tumors might be considered for locoregional therapies, such as chemoembolization or radioembolization, but the level of evidence is low and mostly based on cohort studies. A meta-analysis of 14 trials testing locoregional therapies reported median survival times of 15 months. External-beam radiation therapy is not recommended as standard therapy. At more advanced stages in patients with ECOG of 0–1, systemic chemotherapy with the combination of gemcitabine, cisplatin, and durvalumab showed significantly better survival and progression-free survival compared with gemcitabine plus cisplatin alone (12.8 vs 11.5 months) in the setting of the TOPAZ1 phase 3 trial including 685 patients. Objective response was 27%. Similar results were obtained with the combination of chemotherapy with pembrolizumab versus chemotherapy alone in a phase 3 trial including 1069 patients. The median OS was 12.7 months in the pembrolizumab group and 10.9 months in the placebo group (HR, 0.83; 95% CI, 0.72–0.95; one-sided $p = .0034$). Therefore, triplet therapy is now considered the standard for the management of CCA. In the second-line setting, a phase 3 study that randomized patients who had progressed on cisplatin and gemcitabine to mFOLFOX (leucovorin, fluorouracil, and oxaliplatin) versus best supportive care showed improvement in median OS of 5.3 to 6.2 months (adjusted HR, 0.69). In addition, three molecular and immune therapies have been approved in the second-line setting in iCCA patients with *IDH1/2* mutations, *FGFR2* aberrations, or deficient mismatch repair (dMMR) or microsatellite instability high (MSI-H). A phase 3 trial compared ivosidenib, an IDH-1 inhibitor, versus placebo in the second-line setting and demonstrated an improved primary end point of progression-free survival (2.7 vs 1.4 months; HR, 0.37) and improved OS in an adjusted analysis. A single-arm phase 2 study assessing pemigatinib (FGFR2 inhibitor) in iCCA patients with *FGFR2* fusions showed a median survival of 21 months with an objective response of 35%, leading to FDA accelerated approval. Similar results have been observed in phase 3 trials testing other FGFR2 inhibitors, such as infigratinib and futibatinib. Finally, regulatory agencies granted approval of pembrolizumab for MSI-H or dMMR solid tumors that progressed following prior treatment based on a basket tissue-agnostic trial. This recommendation excludes patients receiving durvalumab in first line.

Mixed HCC-iCCA is a rare neoplasm accounting for <0.5% of all primary liver cancers. Diagnosis is based on pathology. The 2010 World Health Organization classification defined two subtypes: the classical type and the stem cell feature type. Molecular data have also

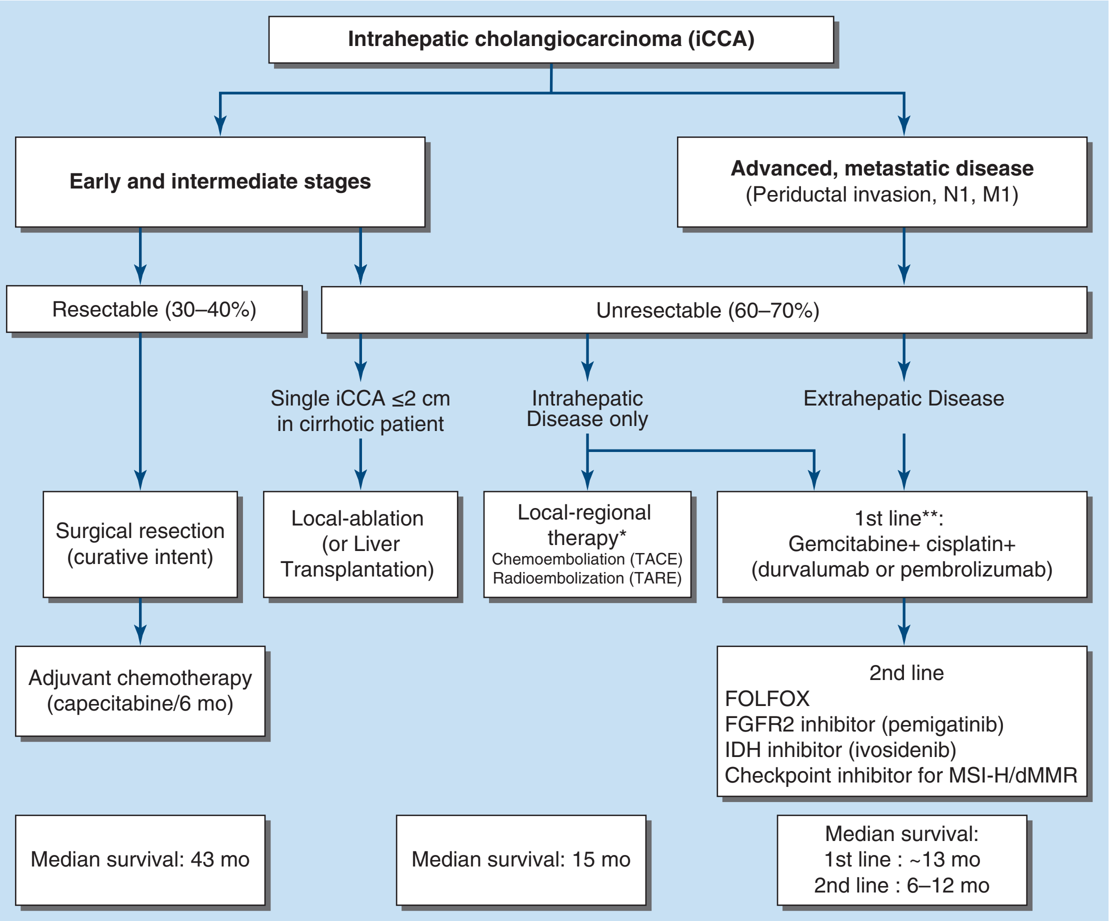
Intrahepatic cholangiocarcinoma (iCCA)
Early and intermediate stages
Advanced, metastatic disease
(Periductal invasion, N1, M1)
Resectable (30–40%)
Unresectable (60–70%)
Single iCCA ≤2 cm in cirrhotic patient
Intrahepatic Disease only
Extrahepatic Disease
Surgical resection (curative intent)
Local-ablation (or Liver Transplantation)
Local-regional therapy*
Chemoembolization (TACE)
Radioembolization (TARE)
1st line**:
Gemcitabine+ cisplatin+ (durvalumab or pembrolizumab)
Adjuvant chemotherapy (capecitabine/6 mo)
2nd line
FOLFOX
FGFR2 inhibitor (pemigatinib)
IDH inhibitor (ivosidenib)
Checkpoint inhibitor for MSI-H/dMMR
Median survival: 43 mo
Median survival: 15 mo
Median survival:
1st line : ~13 mo
2nd line : 6–12 mo

Treatments used as **standard of practice. No enough evidence for standard of care.*
***Patients with ECOG > 2 or liver dysfunction are only suitable for best supportive care.*

FIGURE 87-7 Staging and treatment schedule for intrahepatic cholangiocarcinoma (iCCA). *(Modified from EASL-ILCA Guidelines. J Hepatol 79:181, 2023.)*

characterized a third unique entity, cholangiolocellular carcinoma, with distinct molecular traits and better outcome. Due to their low incidence, the demographic features and clinical behavior of these tumors remain ill-defined. Survival and management are similar to iCCA.

EXTRAHEPATIC CHOLANGIOCARCINOMA

Perihilar and Distal Cholangiocarcinoma pCCA tumors arise between the second-order bile ducts up to the insertion of the cystic duct, whereas dCCAs arise from this point to the ampulla of Vater (Fig. 87-5). Thus, dCCA can be difficult to distinguish from early pancreatic cancer. Both entities have a similar diagnostic approach. Acute onset of painless jaundice occurs in 90% of patients with pCCA, and 10% of patients present with cholangitis. Primary biliary cholangitis with a cutoff for CA 19-9 >129 U/mL is suspicious for CCA. Imaging assessment starts with CT and MRI; they have a good sensitivity and specificity (>85%) for detecting the degree of bile duct involvement and hepatic and portal vein invasion. MRI cholangiography is optimal for defining the extent of the bile duct lesion. Ruling out IgG4 cholangiopathy by assessing serum IgG4 is mandatory. As a second step, endoscopic retrograde cholangiography with brushing to explore cytology and fluorescence in situ hybridization (FISH)—for exploring polysomy—are recommended. FISH enhances the sensitivity of cytology from 20 to ~40%.

Diagnosis is based on pathology. The treatment algorithm for pCCA indicates that in cases of a dominant stricture with positive cytology/biopsy or polysomy, a lymph node biopsy through endoscopic ultrasound should be obtained. pCCA with negative lymph node involvement is best treated by surgery, resection, or transplantation, the sole curative options. Staging laparoscopy is recommended to exclude metastatic disease prior to surgery, which occurs in 15% of cases. Resection entails hepatic and bile duct removal and Roux-en-Y hepaticojejunostomy with regional lymphadenectomy. Bilobular involvement is considered a surgical contraindication. Perioperative mortality rate is up to 10%, mostly as a result of liver failure. In a few referral centers, unresectable single pCCAs <3 cm without dissemination can be considered for liver transplantation with neoadjuvant chemoradiation. This procedure is associated with 5-year survival rates of ~70%. If lymph node involvement is present, systemic chemotherapy can be considered along with biliary tract stenting. Surgical resection (Whipple procedure) is the primary option for management of dCCA, a procedure that achieves a median survival of 24 months and 5-year survival rates of ~25%. Main contraindications for resection are presence of distant lymph node involvement, metastases, or major vascular invasion. At the pathological examination, perineural invasion, lymph node metastasis, R0 resection (absence of residual tumor at pathological examination), and tumor differentiation are predictors of survival. Adjuvant therapy with capecitabine for 6 months is accepted based on the BILCAP study, which has been previously discussed. For advanced cases, consensus statements endorse first-line (gemcitabine and cisplatin plus durvalumab or pembrolizumab) and second-line therapies (FOLFOX) similar to those for iCCA. No molecular targeted therapies are available for these entities.

GALLBLADDER CANCER

Gallbladder cancer is the most common cancer of the biliary tract worldwide. The estimated number of cases of gallbladder cancer in the United States in 2016 was 11,400, more than CCA. The female-to-male ratio is 3:1. Cholelithiasis is the major risk factor, but <1% of patients with cholelithiasis develop this cancer. Gallbladder polyps at risk of transformation are those ≥10 mm in diameter. Early cases are discovered incidentally at routine cholecystectomy. Clinical symptoms, such as jaundice, pain, and weight loss, are associated with advanced stages. Staging of gallbladder cancer involves local disease (tumor confined to the gallbladder) and advanced disease (tumor outside gallbladder with lymph node or distant metastases). The most accurate technique to define staging and vascular and biliary tract invasion is the magnetic resonance cholangiopancreatography. CT and PET scan can also be useful for preoperative staging.

The mainstay of treatment is surgical resection, either simple or radical cholecystectomy (partial hepatectomy and regional lymph node dissection) for local disease. Only ~20% of patients are candidates for surgery with curative intent, and 5-year survival rates range from 60 to 90% depending on prognostic factors such as lymph node or tumor invasion beyond muscular layer. Adjuvant therapy with capecitabine is recommended in R0 cases. Gallbladder cancers at advanced stage are considered unresectable. For patients with ECOG of 0–1, chemotherapy with gemcitabine, cisplatin, and durvalumab is the standard of care based on the TOPAZ 1 phase 3 trial that included 171 patients with gallbladder cancer (25%). Overall, median survival is 10–12 months in advanced cases. Second-line therapy includes FOLFOX chemotherapy. Percutaneous transhepatic drainage is indicated in case of biliary obstruction. Radiotherapy is not effective.

OTHER MALIGNANT LIVER TUMORS

Fibrolamellar Hepatocellular Carcinoma Fibrolamellar hepatocellular carcinoma (FLC) is a rare form of primary liver cancer that typically affects children and young adults (10–30 years of age) without background liver disease. FLC accounts for 0.85% of all primary hepatic malignancies in the United States, and its incidence rate is 0.02 cases per 100,000 inhabitants. FLC is considered a unique entity with a specific fusion oncogene *PRKACA-DNAJB1* present in 80–100% of cases. Few additional mutations have been described in <10% of cases. FLC has a better prognosis than HCC, probably due to the absence of cirrhosis and the earlier age of presentation. Surgical resection is the mainstay of treatment, and indications are less restrictive than for HCC. A retrospective series of 575 FLC cases reported a median survival of 70 months after resection. At advanced stages, the expected outcome is <20 months. There is no standard systemic therapy, and clinical trials are focused on targeting the fusion protein with kinase inhibitors or immunomodulatory agents.

Hepatoblastoma Hepatoblastoma (HB) is the most frequent primary liver tumor in children. The incidence of the disease is 1.5 cases per 1,000,000, and onset of the disease occurs before the age of 3 years. Background liver disease is rare in these patients. WNT signaling plays a major role, with *CTNNB1* mutations (70%) as the most frequently reported molecular event. Overexpression of IGF2 and genes in the 14q32 *DLK1/DIO3* locus is also very prevalent. At diagnosis, ~30% of patients are amenable to surgery. Resection followed by chemotherapy with doxorubicin/cisplatin is the mainstay treatment strategy. Among the remaining 70% of patients, neoadjuvant chemotherapy achieved response in >90% of cases and enabled resection with good clinical outcomes. A study including 1605 patients randomized in eight clinical trials reported better outcome for patients with stage I–II of the PRETEXT (Pretreatment Extent of Tumor) classification (out of four stages), age <3 years, AFP >1000 ng/mL, and absence of metastases. As opposed to HCC, low AFP <100 ng/mL indicates poor prognosis. Overall, 5-year survival is 70% (ranging from 50 to 90% depending on PRETEXT stage).

BENIGN LIVER TUMORS

The most common benign liver tumors are hemangiomas, focal nodular hyperplasia (FNH), and hepatocellular adenomas (HCA). Most benign tumors are identified incidentally by abdominal ultrasound or other imaging techniques. *Hemangiomas* are present in ~5% of the general population and are diagnosed by ultrasound except in cirrhotic patients or oncology patients, in whom contrast-enhanced imaging (contrast-enhanced ultrasound, CT, or MRI) is required. Conservative management is appropriate, and follow-up is not recommended. Exceptionally, growing lesions causing symptoms by compression can be considered for resection. FNH is a benign tumor present in <2% of the population and occurring mostly in females aged 40–50 years. FNH is a polyclonal hepatocellular proliferation due to an arterial malformation. MRI has the highest diagnostic accuracy with a specificity of 100%, when typical imaging features are present (homogeneous enhancement in the arterial phase with a central scar). Atypical FNH requires biopsy for diagnosis. Treatment is not recommended because these tumors do not degenerate or cause complications. In exceptional cases of expanding symptomatic lesions, surgery is the treatment of choice.

Hepatic adenomas are clonal benign proliferations resulting from single-gene driver mutations. HCAs have a low prevalence of 0.001% of the population and are frequently diagnosed in women aged 35–40 years. The female-to-male ratio is 10:1, and the main risk factors are oral contraceptives in females and use of anabolic androgenic steroids in male body builders. HCAs have the potential for hemorrhage and HCC development, particularly when >5 cm. Nowadays, there is a clear understanding of the molecular classification of HCA in subtypes defined by *CTNNB1* mutations (10–20%), *HNF1A* inactivation, and activation of inflammatory pathways (50–60%) or Hedgehog signaling pathway. Diagnosis is based on MRI, which is able to correlate with molecular subtypes in 80% of cases (inflammatory and HNF-1A type). For defining HCA with *CTNNB1* mutations, biopsy is required. Upon diagnosis, discontinuation of oral contraceptives and weight loss are recommended. Resection is indicated in all cases >5 cm, in men, or with *CTNNB1* mutation. For HCA <5 cm, 1-year follow-up is recommended. In case of active HCA bleeding, embolization followed by resection is the treatment of choice. The presence of multiple HCAs is common, and guidelines endorse treating them based on the size of the main nodule.

FURTHER READING

EASL-EORTC Clinical Practice Guidelines: Management of hepatocellular carcinoma. J Hepatol 69:182, 2018.

EASL-ILCA Clinical Practice Guidelines on the management of intrahepatic cholangiocarcinoma. J Hepatol 79:181, 2023.

Finn RS et al: Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. N Engl J Med 382:1894, 2020.

Haber PK et al: Evidence-based management of hepatocellular carcinoma: Systematic review and meta-analysis of randomized controlled trials (2002-2020). Gastroenterology 161:879, 2021.

Ilyas SI: Cholangiocarcinoma: novel biological insights and therapeutic strategies. Nat Rev Clin Oncol 20:470, 2023.

Llovet JM et al: Hepatocellular carcinoma. Nat Rev Dis Primers 21;7:6, 2021.

Llovet JM et al: Locoregional therapies in the era of molecular and immune treatments for hepatocellular carcinoma Nat Rev Gastroenterol Hepatol 18:293, 2021.

Llovet JM et al: Immunotherapies for hepatocellular carcinoma. Nat Rev Clin Oncol 19:151, 2022.

Llovet JM et al: Molecular pathogenesis and systemic therapies for hepatocellular carcinoma. Nat Cancer 3:386, 2022.

Oh DY et al: Durvalumab plus gemcitabine and cisplatin in advanced biliary tract cancer. NEJM Evid 1:2022.

Singal AG et al: AASLD Practice Guidance on prevention, diagnosis, and treatment of hepatocellular carcinoma. Hepatology 78:1922, 2023.

Villanueva A: Hepatocellular carcinoma. N Engl J Med 380:1450, 2019.

88

Pancreatic Cancer

Eileen M. O'Reilly, Andrew M. Lowy, Daniel D. Von Hoff

Pancreatic cancer is the third leading cause of cancer-related mortality in the United States, with >66,000 Americans diagnosed and >51,000 dying from the disease each year. Pancreatic cancer is projected to be the second leading cause of death from cancer in the United States by 2030. Worldwide, pancreatic cancer is the eleventh most common cancer, with about half a million diagnoses per annum. Pancreatic cancer is the most lethal human cancer, with an overall 5-year survival of 13%. However, that situation is changing. In particular, (1) an enhanced understanding of the pathobiology and genomic landscape of pancreatic cancer is steadily giving rise to new therapeutic opportunities, (2) the integration of multimodality therapy for patients with localized disease has improved outcomes, and (3) increasingly more molecularly defined groups of patients are eligible for targeted therapy beyond cytotoxic therapy.

EPIDEMIOLOGY

Pancreatic cancer accounts for 3% of all new cancer cases in the United States and ~8.3% of all deaths from cancer in the United States. The lifetime risk of developing pancreatic cancer is ~1.7%; however, the incidence has been increasing about 1.1% per year overall, but notably in people <55 years of age, it has been increasing 2.36% per year in women and 0.62% per year in men. Pancreatic cancer is more common with increasing age and slightly more common in men than in women. The 5-year survival rate for all stages has increased from 3% in 1975 to 13% in 2024. The latest information from the U.S. Surveillance, Epidemiology, and End Results (SEER) database predicts that the 5-year survival is about 44% for patients with localized pancreatic cancer, 16% for those with regional disease, and 3.2% for patients with metastatic disease. Pancreatic cancer is more common in developed countries (although it tracks with the prevalence of smoking). The incidence is highest in Eastern (e.g., Hungary and Slovakia) and Western (e.g., Germany) Europe and North America followed by other areas in Europe, Australia, New Zealand, and Southcentral Asia.

RISK FACTORS

Age is one of the strongest risk factors for pancreatic cancer with median age at diagnosis of 71 years (the disease is most frequently diagnosed in the 65–79 age group; for men, 65–69; for women, 75–79). The number of new cases per 100,000 persons and the number of deaths per 100,000 persons are higher for males and for black people of both sexes. People who have a non-O blood type are at higher risk of developing pancreatic cancer.

Environment The greatest risk factor for pancreatic cancer is cigarette smoking. The risk correlates with the increased number of cigarettes smoked and persists for at least 10 years following smoking cessation. Twenty-five to 30% of pancreatic cancers are believed to be caused by smoking. Exposure to cadmium as part of cigarette smoking or via exposure to welding, soldering, or dietary exposure has been weakly associated with an increased risk of pancreatic cancer.

Dietary factors may contribute to risk, acknowledging confounding issues; however, high intake of fat or meat (particularly well-done barbequed meat) are risk factors. High intake of citrus fruits and vegetables are associated with a decreased risk of pancreatic cancer. Coffee and low-to-moderate alcohol consumption are not associated with an increased risk for pancreatic cancer, although consumption of sugary carbonated drinks has been associated with elevated risk.

Hereditary Factors, Genetics, and Screening Hereditary factors account for 10–16% of all pancreatic cancers. Family members of patients with pancreatic cancer and selected individuals with certain pathogenic germline variants should seek participation in an early detection program with genetic counseling, definition of risk, and if appropriate, periodic magnetic resonance imaging (MRI) screening and endoscopic ultrasound of the pancreas, ideally enrolled on a prospective registry. Table 88-1 summarizes the various germline variants along with familial cancer syndromes where there is known to be an increased risk for pancreatic cancer. For an average-risk individual (without any known predisposing factors), there is currently no recommended screening.

Knowing that an individual carries a *BRCA1/2* or *PALB2* germline variant or any of the above mutations requires referral of that person to an early detection or high-risk screening clinic. For patients with a *BRCA1/2* or *PALB2* pathogenic germline variant, a poly(ADP-ribose) polymerase (PARP) inhibitor should be considered (see below) as part of treatment for metastatic pancreatic cancer. Other pathogenic germline variants are under study to determine their increased risk of pancreatic cancer, including *CFTR, PRSS2, CDK4, FANCC, APC, ATM, BRIP1, BRCA1, EPCAM, MEN1, MLH1, MSH2, MSH6, NF1, PMS2, SMAD4, TP53, TSC1, TSC2,* and *VHL*. Some of these variants are also associated with pancreatic neuroendocrine tumors **(Chap. 89)**.

In addition to recognized genetic syndromes from single gene variants, other individuals without any of these identifiable germline variants may be at higher risk for this cancer. For example, a family history of pancreatic cancer is associated with increased risk based on the number of affected relatives. Having one first-degree relative with pancreatic cancer, one's risk for developing the disease is increased 4.6-fold, having two first-degree relatives increases the risk 6.4-fold, and three or more first-degree relatives confers a 32-fold increased risk. The risk is also increased if a relative developed pancreatic cancer at <55 years old and is further compounded by smoking.

Medical Conditions Chronic pancreatitis that is nonfamilial is associated with an increased risk of pancreatic cancer (2.3–16.5-fold increase). Risk is also increased in people with chronic pancreatitis associated with cystic fibrosis or tropical pancreatitis.

A clear association exists between diabetes mellitus (both type 1 and type 2) and pancreatic cancer. Whether this is a causal association or whether the diabetes is the result of the cancer is not exactly clear. New-onset (particularly with concomitant weight loss) or unexpected worsening of type 2 (or type 3c) diabetes may be associated with pancreatic cancer, and research programs are evaluating screening in otherwise healthy individuals in this setting.

High body mass index (BMI) is considered a risk factor for pancreatic cancer. A high BMI of ≥30 is associated with a doubling of pancreatic cancer risk. As obesity is a risk factor for diabetes, the contribution of obesity alone is unclear. Interestingly, patients with severe obesity

TABLE 88-1 Germline Mutations, Familial Cancer Syndromes, and Fold Risk of Pancreatic Cancer

GERMLINE MUTATION	FAMILIAL CANCER SYNDROME	ESTIMATED INCREASED RISK (FOLD) OF PANCREATIC CANCER
BRCA2[a]	Familial breast/ovarian cancer and others	3.5–10
PALB2 (partner and localizer of *BRCA2*)	Familial breast cancer and others	~Sixfold
p16/CDKN2A	Familial atypical multiple mole melanoma (FAMMM)	15–22
STKII (*LKB1*)	Peutz-Jeghers syndrome	76–140
PRSS1 or *SPIN11*[b]	Hereditary (familial) pancreatitis	53
ATM	Ataxia-telangiectasia	6[c]
MLH1, MSH2, MSH6, PMS2	Heredity nonpolyposis colorectal syndrome or Lynch syndrome[d]	9–36

[a] Particularly common in individuals with Ashkenazi Jewish heritage. [b] Forty percent chance of pancreatic cancer by the age of 70. [c] Calculated at age 70. [d] Very important because this is associated with microsatellite instability, which is a marker for response to immune checkpoint blockade.

who undergo a bariatric intervention experience a reduction in the incidence of gastrointestinal cancers, including pancreatic cancer, by >30% in the first 3 years (along with a decrease in their hemoglobin A_{1c} and blood glucose level). Physical inactivity also has been associated with an increased risk of pancreatic cancer.

Other Considerations Most patients with pancreatic cancer relate that they have had developing symptoms over months to years before diagnosis. Efforts at early detection of the disease have not yet been fruitful, but this is an area of very active investigation.

PATHOLOGY AND MOLECULAR CONSIDERATION

Location The posterior location of the pancreas in the abdomen is one of the issues that makes diagnosis more challenging **(Fig. 88-1*A*)**.

Pathology Cancers of the pancreas can be divided into neoplasms of the endocrine pancreas **(Chap. 89)** and those of the exocrine pancreas. The most common neoplasm of the exocrine pancreas and most lethal is ductal adenocarcinoma. These cancers arise in the head, body, or tail of the pancreas and are characterized by infiltrating desmoplastic stromal reactions **(Fig. 88-1*B*)**.

Other subtypes of nonneuroendocrine pancreatic cancers include acinar cell carcinoma (tumors of the exocrine enzyme producing cell), colloid carcinoma, medullary carcinoma, adenosquamous carcinoma, and other rare subtypes. Each of these is different in behavior and in their molecular characteristics and often requires other specific types of treatment. Occasionally, metastases to the pancreas occur secondary to renal, breast, lung, and urothelial carcinomas, melanoma, lymphoma, and other malignancies.

Molecular Characteristics The mutational landscape of pancreatic ductal adenocarcinoma is characterized by alterations in four genes that are commonly mutated or inactivated. The most common of these is the *KRAS* oncogene (mutations primarily in codon 12). It is increasingly important to identify the specific mutant *KRAS* allele as novel drugs targeting specific alleles now exist and others are in development. *KRAS* mutations are seen in 90–95% of pancreatic adenocarcinomas. A subset of pancreatic cancers is termed "*KRAS* wildtype," where no alteration in *KRAS* is observed. This form of pancreatic cancer is more frequently observed in patients <55 years of age, and often other therapeutically actionable findings are identified in these cancers, including mutations in *BRAF*, rare oncogenic targetable fusions (e.g., *NRG-1, NTRK1-3*), or microsatellite instability. *p16/CDKN2A* mutation or epigenetic silencing is present in >90% of invasive pancreatic adenocarcinomas. *TP53* is mutated in ~75% and *DPC4/SMAD4* is mutated in about half of these tumors. As a reference point, the *BRCA2* gene noted in Table 88-1 is mutated in 3.5–10% of pancreatic adenocarcinomas.

Precursor Lesions Most pancreatic cancers are believed to arise from pancreatic intraepithelial neoplasia (PanINs), which have varying degrees of dysplasia designated as PanINs 1–3 (and constitute a progression model for pancreatic cancer). Genetic alterations become more frequent as the PanIN grade increases (e.g., grade 3). Most PanIN lesions will never progress to invasive malignancy. An alternate precursor lesion is an intraductal papillary neoplasm (IPMN), which is usually noninvasive. Unfortunately, even high-grade PanIN lesions are not detectable by currently available imaging modalities. In contrast, some pancreatic adenocarcinomas arise from noninvasive cystic epithelial precursor lesions. These lesions can be seen on magnetic resonance imaging (MRI) and computed tomography (CT) imaging, and thus detection may permit early diagnosis of pancreatic cancer. As compared to main pancreatic duct IPMN, a side branch duct IPMN is more likely to be noninvasive. MRI and endoscopic ultrasound are commonly used to assess malignant potential based on risk factors such as size, growth rate, main pancreatic duct size, and the presence of mural nodularity.

One other rare pancreatic tumor is the mucinous cystic neoplasm; they may be seen as incidental findings on scans and are more common in the body and tail of the pancreas and tend to occur in women. These lesions are less likely to progress to malignancy. Clinical signs of concern include large size (>3 cm) and the presence of mural nodules and/or thickening. Even after rare progression to mucinous adenocarcinoma, they typically have a more favorable prognosis relative to typical ductal adenocarcinoma.

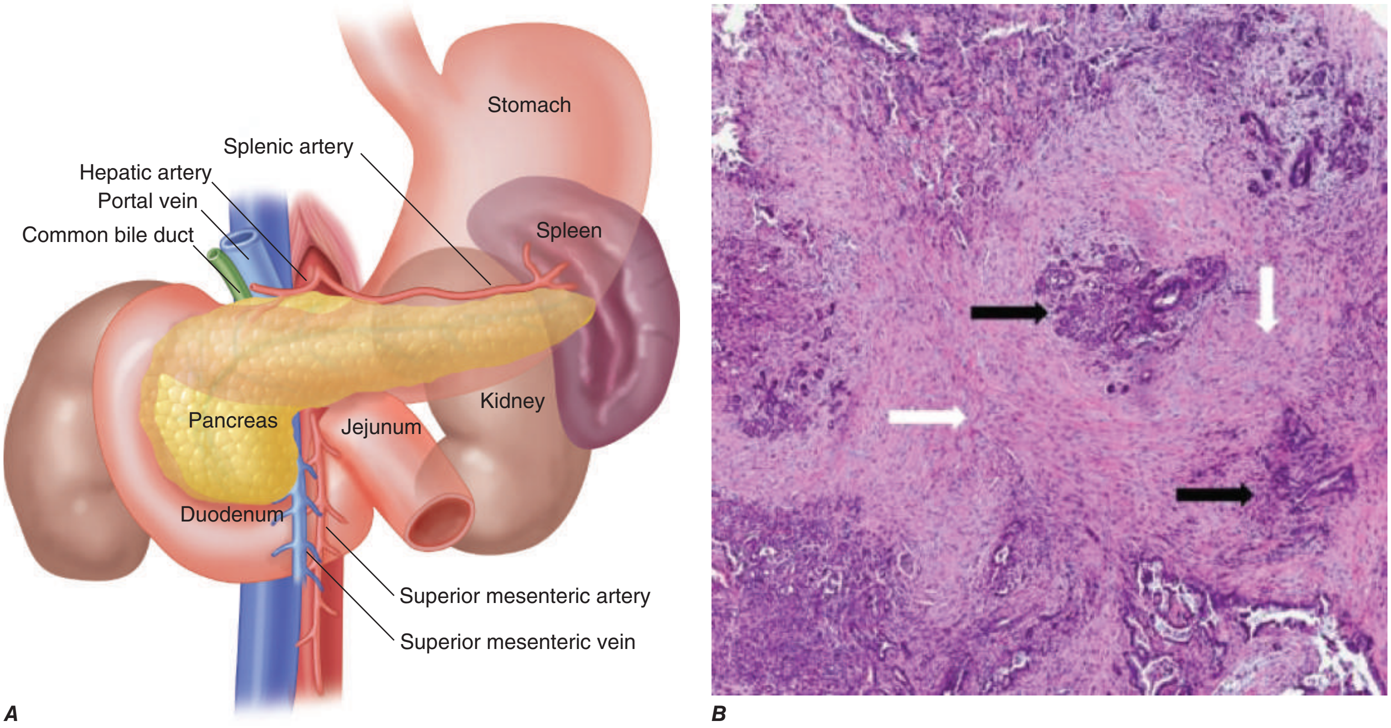

FIGURE 88-1 ***A.*** Note the relationship of the pancreas to the major vessels of the retroperitoneum. ***B.*** Ductal adenocarcinoma of the pancreas (*black arrows*), with intense stromal component (*white arrows*). *(Part A is courtesy of Mary Kay Washington, MD, PhD, Vanderbilt University. Part B is courtesy of Haiyong Han, PhD, Translational Genomics Research Institute [TGen].)*

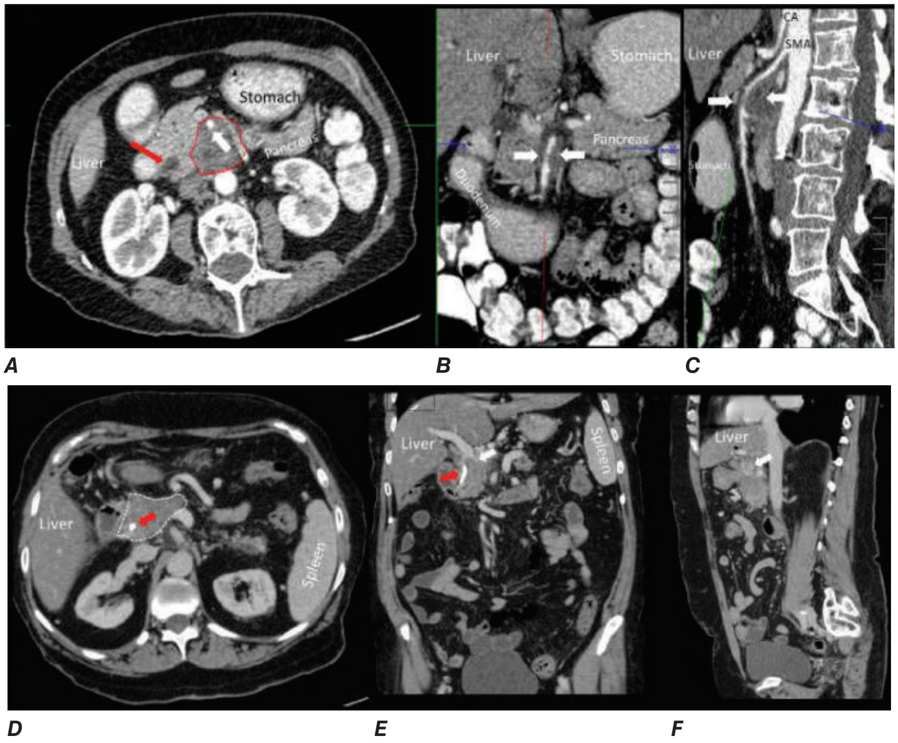

FIGURE 88-2 Selected images from contrast-enhanced computed tomography (CT) in patients with locally advanced adenocarcinoma of the pancreas. A high-quality contrast-enhanced CT scan (arterial phase in panels ***A–C*** and portal venous phase in panels ***D–F***) is required for optimal staging of pancreas cancer. Panel ***A*** demonstrates the typical features of adenocarcinoma of the pancreas on arterial phase axial CT scans (*dotted outline*) with tumor encasement of the superior mesenteric artery (*white arrow*). Note the dilatation of the common bile duct (*red arrow*). Panels ***B*** (magnified coronal) and ***C*** (sagittal) show reconstruction of CT images into additional orthogonal planes with exquisite details to confirm the unresectable nature of the tumor due to vascular encasement. Panel ***D*** demonstrates the typical features of adenocarcinoma of the pancreas on portal venous phase axial CT scans in a different subject. The dotted line outlines a pancreas cancer lesion in the pancreatic head, which is encasing the portal splenic confluence (*dotted outline*). Panels ***E*** (*white arrow*) and ***F*** show the pinched appearance of the portal splenic confluence by tumor abutment and invasion of the superior mesenteric vein (*white arrow*) on coronal and sagittal views. Note the presence of a stent in the common bile duct (*red arrow*) to help relieve biliary obstruction caused by the tumor. CA, celiac axis; SMA, superior mesenteric artery.

CLINICAL FEATURES

History and Physical A classic presentation for a patient with pancreatic cancer arising in the head of the gland is "painless" jaundice. Jaundice is visually detectable with a bilirubin of >2 mg/dL, and pruritus may also occur due to bile salt deposition in the skin. However, pain is also common in newly diagnosed patients and is typically mid epigastric (sometimes described as a "boring-like" pain) with radiation to the back (due to retroperitoneal invasion of the splanchnic nerve plexus). The pain may be exacerbated by eating or lying flat. In the presence of jaundice, light stool color from the absence of bile occurs (steatorrhea also causes malodorous stools). Other signs include the onset of diabetes (particularly with concomitant weight loss) or hyperglycemia/elevated hemoglobin A_{1c} in the preceding 1–2 years. The association of pancreatic cancer with depression remains controversial.

In addition to jaundice, physical signs of pancreatic cancer include evidence of weight loss, including loss of muscle mass, and a palpable gallbladder in the setting of biliary obstruction (Courvoisier's sign). Migratory superficial thrombophlebitis and presentation with both deep venous and arterial thromboses can occur (Trousseau's syndrome). Signs of late-stage disease include a lymph node palpable in the supraclavicular fossa (usually on the left where the thoracic duct enters the subclavian vein). This is referred to as Virchow's node. Occasionally, one can palpate subcutaneous metastases in the periumbilical area referred to as a Sister Mary Joseph's node—named after one of the operating room nurses from the Mayo Clinic who noted that when she prepped that area and felt those nodules, the patient often had peritoneal metastases.

The history and symptoms noted above lead to imaging as the next step, including ultrasonography (jaundice) and CT or MRI.

DIAGNOSTIC WORKUP

Imaging A key diagnostic tool is the use of a dual-phase contrast-enhanced CT using a pancreas cancer protocol, which allows arterial phase enhancement and portal venous phase enhancement and detailed visualization of tumor–blood vessel relationships to inform both resectability and staging. Figure 88-2 demonstrates such a CT scan (with vascular involvement). Figure 88-3 demonstrates the use of an ^{18}F-glucose positron emission tomography (PET) scan that can aid in the detection of otherwise occult metastatic disease.

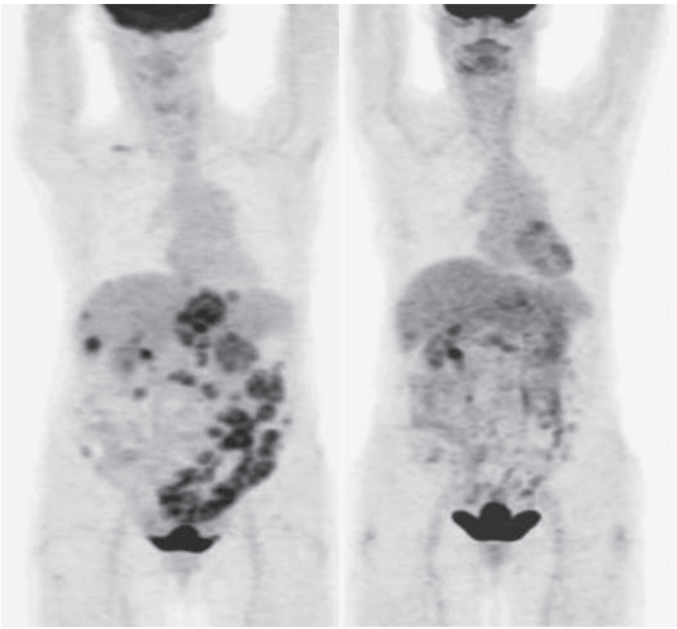

FIGURE 88-3 Positron emission tomography scan demonstrating metastatic disease—baseline and after 6 weeks of chemotherapy with partial resolution of liver metastases.

Histologic Diagnosis A histologic (tissue) diagnosis is essential and should be obtained with a Tru-Cut biopsy needle (fine-needle aspiration cytology is an acceptable but less preferred alternative). Obtaining adequate tissue not only secures a histologic/cytologic diagnosis but also facilitates molecular testing (next-generation sequencing for genomic alterations including *KRAS*) and microsatellite status. Increasingly subsets of patients based on genomic characterization are being identified and for whom therapies are available, making the need for adequate tissue sampling increasingly relevant to patient management.

A core needle (18- to 22-gauge) biopsy can be obtained via endoscopic ultrasound-guided technique for a tumor localized to the pancreas. To biopsy the liver or a lymph node, a larger (16- to 18-gauge) percutaneous needle biopsy by an interventional radiologist is typically undertaken.

Serum Markers Before treatment and, ideally, once the bilirubin level has normalized (biliary obstruction is associated with elevated CA19-9), a serum sample should be obtained for level of CA19-9, carcinoembryonic antigen (CEA), and CA125 (can be helpful for the 8–15% of patients with nondetectable CA19-9–Lewis antigen nonsecretors). Trends of these tumor markers over time can inform therapeutic decision-making.

IMPORTANT IMMEDIATE CONSIDERATIONS IN PATIENT CARE

During the life cycle of this cancer, biliary tract obstruction is a frequent occurrence for tumors arising in the head of the gland (and the attendant risk for sepsis from the biliary tree). A metallic biliary wall stent is typically placed endoscopically to alleviate jaundice and pruritus. If surgery is under consideration as the initial therapeutic step, biliary stent placement is not necessarily warranted; however, if neoadjuvant (preoperative) systemic therapy is planned, both a histologic/cytologic diagnosis and placement of a biliary wall stent are required (where the latter is medically indicated). Of important note, all patients with localized pancreatic cancer require multidisciplinary evaluation by surgical, medical, and radiation oncology to optimize therapeutic decision-making.

Patients with metastatic pancreatic cancer may have a hypercoagulable state and frequently have thrombophlebitis (Trousseau's sign) and deep vein thrombosis with pulmonary emboli and/or more rarely arterial thrombotic events. Anticoagulation using either a direct oral anticoagulant or low-molecular-weight heparin is required in these setting in the absence of a contraindication to anticoagulation.

Control of pain or of active cancer-related symptomatology including anorexia, weight loss, and malabsorption should be proactively treated. Pancreatic enzyme replacement therapy (PERT) is a critical measure to address symptoms of malabsorption, including abdominal cramping, bloating, steatorrhea, excess flatus, and weight loss. Early involvement of a supportive care team can help maximize symptom control, facilitate treatment readiness, and extend a better quality of life.

CLINICAL STAGING

The clinical staging of pancreatic cancer according to the American Joint Commission on Cancer staging is presented in **Table 88-2**.

Table 88-3 presents another clinical way to express extent of disease as well as therapeutic approaches (to be discussed later).

For optimal staging, a laparoscopy is selectively indicated either before or at the time of definitive surgery. If metastatic disease is identified at laparoscopy, curative intent surgery is not undertaken.

TREATMENT

Pancreatic Cancer

RESECTABLE DISEASE (10–20%)

For all patients with localized pancreatic cancer, multidisciplinary evaluation should be undertaken. A standard approach for resectable disease (10–15% of patients; as defined in Table 88-3) is upfront surgery. Neoadjuvant therapy (systemic therapy prior to surgery) is an option for patients with resectable disease. The rationales for preoperative chemotherapy include the following: (1) to control micrometastatic disease; (2) to assess tumor biology and response to the selected chemotherapy regimen (this allows patients with progressive disease to avoid nontherapeutic, morbid surgery); (3) to assure delivery of systemic therapy because therapy is better tolerated in the preoperative setting and is not subject to omission of drugs or altering doses due to delayed postoperative recovery; and (4) to achieve tumor downstaging to enhance the potential for a margin-negative surgical resection. The type of surgery for patients with tumors in the pancreatic head, neck, or uncinate process is typically a pancreaticoduodenectomy with or without pylorus preservation. For tumors in the body or tail, a distal subtotal pancreatectomy and splenectomy are performed. These operations may be performed using a traditional "open" surgical technique or using minimally invasive techniques. Clinical and pathologic findings of the resection are defined as either an R0 resection (no macroscopic or microscopic disease left after surgery) or an R1 resection, which refers to microscopic residual disease at the surgical margin. An R2 resection refers to gross residual disease remaining after surgery and is a highly undesirable oncologic outcome. The best outcomes are achieved in patients with a small tumor (<2 cm), no lymph node involvement (N0), and R0 (stage I), in whom 5-year survival ranges from 50 to 90% depending on actual tumor size and grade. Overall, only a small fraction of tumors are diagnosed when stage I; however, a key goal of screening is to "stage migrate" and increase the proportion of tumors that are diagnosed as early stage I lesions.

Several approaches are designed to maximize outcome for localized pancreatic cancer.

Postoperative Adjuvant Therapy A standard of care is 24 weeks of adjuvant treatment with the modified FOLFIRINOX regimen (folinic acid, 5-fluorouracil, irinotecan, oxaliplatin). The median survival was 54 months for modified FOLFIRINOX compared to 35 months for gemcitabine alone (hazard ratio [HR] 0.64; 95% confidence interval [CI] 0.48–0.86; $p = .003$) in the definitive trial supporting this therapy. Main side effects included fatigue, gastrointestinal toxicity, myelosuppression, and neuropathy.

Neoadjuvant Therapy A newer approach is the use of neoadjuvant chemotherapy (systemic therapy given before surgery) to shrink the tumor, eradicate micrometastatic disease, and normalize serum CA19-9 level. Neoadjuvant therapy is the standard approach for patients with borderline resectable or locally advanced disease. The role of radiation in the treatment of localized pancreatic cancer is controversial.

LOCALLY ADVANCED DISEASE (25–30% OF PATIENTS)

For patients with locally advanced disease (defined by tumor contact with arterial structures >180° and/or venous contact that does not permit resection/reconstruction), the median survival is poor (6–18 months). Complications from local disease progression can be significant and include pain, biliary and duodenal obstruction, vascular thromboses, varices, bleeding, and infection/sepsis. Typical treatment paradigms for locally advanced disease include systemic chemotherapy with or without radiation therapy. A proportion of patients with locally advanced disease (~20%) may be rendered operable following neoadjuvant therapy.

METASTATIC DISEASE (50–60% OF PATIENTS)

The mainstay of treatment for metastatic pancreatic cancer is systemic chemotherapy. Several treatment regimens including FOLFIRINOX, gemcitabine, albumin-bound paclitaxel, and, more recently, NALIFIRINOX (nanoliposomal irinotecan, 5-fluorouracil, leucovorin, oxaliplatin; NAPOLI-3) have all demonstrated a survival advantage compared to prior standards in untreated metastatic pancreatic cancer. In a previously treated setting, liposomal irinotecan and infusional 5-fluorouracil and leucovorin (NAPOLI-1) regimen has been shown to improve survival. Predictors of outcome in the

TABLE 88-2 Definition of Primary Tumor (T)

T CATEGORY	T CRITERIA
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma in situ This includes high-grade pancreatic intraepithelial neoplasia (PanIn-3), intraductal papillary mucinous neoplasm with high-grade dysplasia, intraductal tubulopapillary neoplasm with high-grade dysplasia, and mucinous cystic neoplasm with high-grade dysplasia
T1	Tumor ≤2 cm in greatest dimension
T1a	Tumor ≤0.5 cm in greatest dimension
T1b	Tumor >0.5 cm and <1 cm in greatest dimension
T1c	Tumor 1–2 cm in greatest dimension
T2	Tumor >2 cm and ≤4 cm in greatest dimension
T3	Tumor >4 cm in greatest dimension
T4	Tumor involves celiac axis, superior mesenteric artery, and/or common hepatic artery, regardless of size
N CATEGORY	**N CRITERIA**
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastases
N1	Metastasis in one to three regional lymph nodes
N2	Metastasis in four or more regional lymph nodes
M CATEGORY	**M CRITERIA**
M0	No distant metastasis
M1	Distant metastasis

AJCC Prognostic Stage Groups

WHEN T IS...	AND N IS...	AND M IS...	THEN THE STAGE GROUP IS....
Tis	N0	M0	0
T1	N0	M0	IA
T1	N1	M0	IIB
T1	N2	M0	III
T2	N0	M0	IB
T2	N1	M0	IIB
T2	N2	M0	III
T3	N0	M0	IIA
T3	N1	M0	IIB
T3	N2	M0	III
T4	Any N	M0	III
Any T	Any N	M1	IV

Source: Used with permission of the American College of Surgeons, Chicago, Illinois. The original source for this information is the *AJCC Cancer Staging System* (2023).

TABLE 88-3 Extent of Disease and Therapeutic Approach

DESIGNATION (MEDIAN SURVIVAL)	THERAPEUTIC APPROACHES
1. Resectable (localized): (18–23 mo) • No solid tumor contact with celiac axis, hepatic artery, or superior mesenteric artery (SMA), contact with superior mesenteric–portal veins of <180° • Patent superior mesenteric–portal veins • No extrapancreatic disease	Surgery followed by adjuvant therapy • mFOLFIRINOX or gemcitabine +/– capecitabine or nab-paclitaxel Neoadjuvant chemotherapy followed by surgery
2. Locally advanced: (6–18 mo) • Arterial involvement of >180° (superior mesenteric artery, others) • Venous occlusion (superior mesenteric vein [SMV] or portal) • No extrapancreatic disease	Chemotherapy Chemotherapy +/– radiation Evaluate for surgery following systemic therapy
3. Metastatic: (6–12 mo)	Chemotherapy with special consideration for tumors with specific targets (e.g., mismatch repair high, *BRCA1/2*, *BRAF*, etc.)

Abbreviation: mFOLFIRINOX, modified FOLFIRINOX (folinic acid, 5-fluorouracil, irinotecan, and oxaliplatin (T Conroy et al: N Engl J Med 379:2395, 2018).

metastatic setting include good physical functioning (performance status), favorable nutritional status, the absence of liver metastases, lung-only metastases, and the presence of select pathogenic germline variants. **Table 88-4** details combination regimens that have further improved survival modestly. Median overall survival ranges from 6 to 12 months. However, 1-year survival is improving, with subsets of patients whose survival with metastatic disease can reach a couple of years or longer.

PATIENTS WITH A SPECIFIC MOLECULAR PROFILE IN THEIR TUMOR/GERMLINE

PARP inhibitors have activity in patients with pathogenic germline or somatic (tumor-based) *BRCA2*, *BRCA1*, or *PALB2* (i.e., defective DNA repair proteins) variants. In addition, these tumors might be more sensitive to specific combinations of chemotherapy that include platinum agents (gemcitabine and cisplatin or FOLFIRINOX). About 1% of pancreatic cancers have microsatellite instability, typically with high numbers of mutations in their tumors, due to deficient mismatch repair, and these tumors are likely to benefit from immune checkpoint blockade with anti-PD-1 (pembrolizumab, nivolumab) and anti-CTLA-4 inhibitors. *KRAS* mutations occur ubiquitously in pancreatic cancer. Current treatment guidelines include drugs targeting *KRAS* G12C (~1% of pancreatic

TABLE 88-4 Combination Chemotherapy Regimens That Have an Impact on Survival in Stage IV Disease

STUDY DESIGN (AUTHOR/REF)	NO. OF PATIENTS	MEDIAN OVERALL SURVIVAL (MONTHS)
Gemcitabine + erlotinib vs gemcitabine (Moore et al: J Clin Oncol 26:1960, 2007)	569	6.24 vs 5.91 (HR 0.82; 95% CI 0.69–0.99; *p* = .038)
FOLFIRINOX (folinic acid + 5-fluorouracil + irinotecan + oxaliplatin) vs gemcitabine (Conroy et al: N Engl J Med 364:1817, 2011)	342	11.1 vs 6.8 (HR 0.57; 95% CI 0.45–0.70; *p* <.001)
Nab-paclitaxel + gemcitabine vs gemcitabine (Von Hoff et al: N Eng J Med 369:1691, 2013)	861	8.5 vs 6.7 (HR 0.72; 95% CI 0.62–0.83; *p* <.001[a]
Nanoliposomal irinotecan, 5-fluorouracil + folinic acid vs nanoliposomal irinotecan monotherapy vs 5-fluorouracil + folinic acid (Wang-Gillam et al: Lancet 387:545, 2015)	417	6.1 vs 4.2 (HR 0.67; 95% CI 0.49–0.92; *p* = .012[b])
NALIRIFOX (nanoliposomal irinotecan, 5-fluorouracil, folinic acid, oxaliplatin) vs nab-paclitaxel and gemcitabine (Wainberg et al: Lancet 402:1272, 2023)	770	11.1 vs 9.2 (HR 0.83; 95% CI 0.70–0.99; *p* = .036)

[a] The 2-year survival rate with this regimen is 9%, and the 3+ year rate is 4%. Other studies have not reported on these parameters. [b] HR is for nanoliposomal irinotecan + 5-fluorouracil + folinic acid vs 5-fluorouracil + folinic acid.

Abbreviations: CI, confidence interval; HR, hazard ratio.

cancers); however, agents that target more common *KRAS* mutations (G12D, G12V, G12R) are being evaluated in clinical trials and have high potential to become part of standard treatment algorithms in the future. Other rare actionable alterations include oncogenic fusions in *RET, ALK, MET, NRG-1, ROS*, and *BRAF* V600E mutations, for which therapeutic agents are available and are in clinical trials.

MAINTENANCE THERAPY FOR PATIENTS RESPONDING TO TREATMENT

For patients with a germline *BRCA1* or *BRCA2* mutation whose metastatic pancreatic cancer has not grown during an initial platinum-based regimen, the PARP inhibitor olaparib has been shown to improve progression-free survival (7.4 vs 3.8 months; HR 0.53; 95% CI 0.35–0.82; *p* = .004) and maintenance of quality of life, both relative to placebo.

FUTURE DIRECTIONS

Multiple novel therapies are under development in pancreatic cancer. Immunotherapy using personalized neoantigen vaccines or using an antigenic target such as a "public" or shared neoantigen such as *KRAS* has demonstrated early promise following resection of pancreatic cancer as an adjunct to standard chemotherapy. Mid-phase trials are planned/underway to further explore these signals. *KRAS* has heretofore been considered nondruggable; however, developments in organic chemistry, biosynthesis, and other innovations have led to a series of therapeutics directly targeting *KRAS* with promise that these agents in the proximate future will be integrated as part of standard therapy for pancreatic cancer. Multiple therapeutic approaches targeting the tumor immune microenvironment are being explored, capitalizing on an increased understanding of the pathobiology of this cancer. Other important developments include innovation in clinical trial design, novel approaches to screening and surveillance utilizing "liquid" biomarkers (incorporating DNA fragments, methylation, and proteomic profiles), and other technologic developments.

Acknowledgment

Thank you to the American Joint Committee on Cancer for providing the tables.

FURTHER READING

Conroy T et al: FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. N Engl J Med 364:1817, 2011.

Conroy T et al: FOLFIRINOX or gemcitabine as adjuvant therapy for pancreatic cancer. N Engl J Med 379:2395, 2018.

Golan T et al: Maintenance olaparib for germline BRCA-mutated metastatic pancreatic cancer. N Engl J Med 381:317, 2019.

Hu ZI, O'Reilly EM: Therapeutic developments in pancreatic cancer. Nat Rev Gastroenterol Hepatol 21:7, 2024.

Hruban RJ et al: Genetic progression in the pancreatic ducts. Am J Pathol 156:1821, 2000.

Park W et al: Pancreatic cancer: A review. JAMA 326:851, 2021.

Rahib L et al: Evaluation of pancreatic cancer clinical trials and benchmarks for clinically meaningful future trials: A systemic review. JAMA Oncol 2:1209, 2016.

Rawla P et al: Epidemiology of pancreatic cancer: Global trends, etiology, and risk factors. World J Oncol 10:10, 2019.

Solomon S et al: Inherited pancreatic cancer syndromes. Cancer J 18:485, 2012.

Von Hoff D et al: Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. N Engl J Med 369:1691, 2013.

Wainberg Z et al: NALIRIFOX versus nab-paclitaxel and gemcitabine in treatment-naive patients with metastatic pancreatic ductal adenocarcinoma (NAPOLI 3): A randomised, open-label, phase 3 trial. Lancet 402:1272, 2023.

Yuan C et al: Diabetes, weight change, and pancreatic cancer risk. JAMA Oncol 6:e202948, 2020.

89

Gastrointestinal Neuroendocrine Tumors

Matthew H. Kulke

Gastrointestinal (GI) neuroendocrine tumors (NETs) can be broadly grouped according to their site of origin as either extrapancreatic NETs, historically called carcinoid tumors, or pancreatic NETs. While NETs can pursue a broad range of clinical behaviors, they classically follow a course that is more indolent than many other malignancies. NETs also have the ability to synthesize peptides, growth factors, and bioactive amines that may be ectopically secreted, giving rise to a range of unique clinical syndromes.

INCIDENCE AND PREVALENCE

The diagnosed incidence of NETs has been steadily increasing over the past several decades. An analysis of data from the Surveillance, Epidemiology, and End Results (SEER) program, comprising population-based data in the United States from 1973 to 2012, showed that the overall incidence had increased 6.4-fold over this time period and that the estimated prevalence of patients who had been diagnosed with a NET was >170,000. Gastroenteropancreatic neuroendocrine tumors comprise the majority of neuroendocrine tumors, and recent analyses suggest the incidence in this subgroup has risen proportionately and continues to rise **(Fig. 89-1)**. These same studies have also found that overall survival durations for patients with NETs have improved significantly over time. The increasing incidence and improved survival durations for patients with NETs likely reflect, at least in part, advances in both diagnosis and treatment.

While environmental or other factors leading to an increased incidence of NETs cannot be excluded, common cancer risk factors such as tobacco or alcohol use and dietary patterns have not been clearly linked to NET development.

A minority of NETs develop in the context of autosomal inherited genetic syndromes associated with mutations in specific tumor-suppressor genes. The most common of these is multiple endocrine neoplasia type 1 (MEN 1), due to mutation and loss of function of the *menin* gene, located on chromosome 11q13 **(Chap. 400)**. Patients

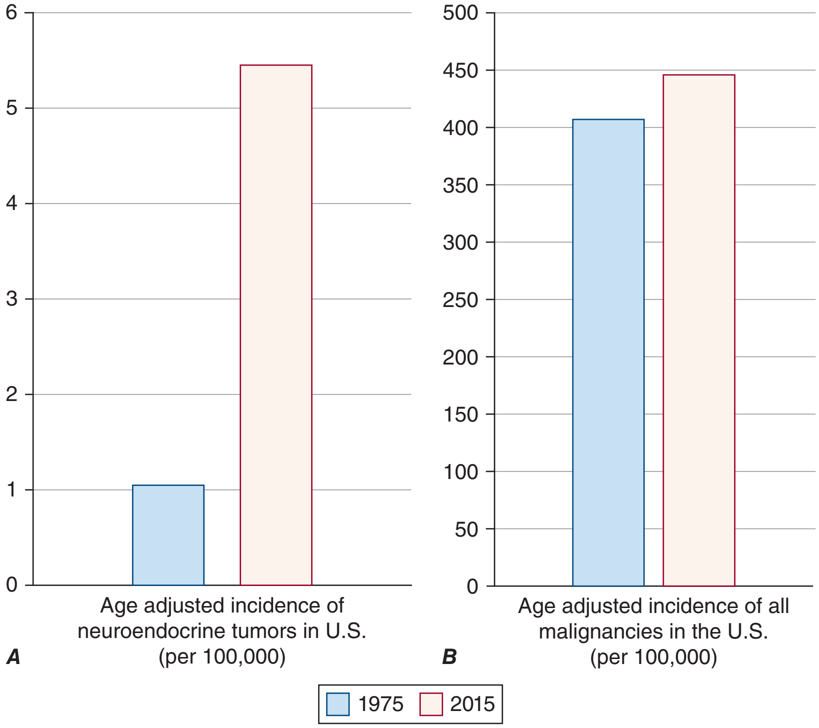

FIGURE 89-1 Incidence of gastroenteropancreatic neuroendocrine tumors (NETs). The incidence of gastroenteropancreatic neuroendocrine tumors has been increasing over the past several decades, an observation that has been attributed in part to improved diagnosis and classification. *(Adapted from Z Xu et al: Epidemiologic trends of and factors associated with overall survival for patients with gastroenteropancreatic neuroendocrine tumors in the United States. JAMA Network Open 4:e2124750, 2021, Figure 1A.)*

with MEN 1 are at risk for developing pancreatic NETs as well as hyperparathyroidism and pituitary adenomas; less commonly, they may develop bronchial and thymic NETs. Other inherited syndromes associated with NETs include von Hippel–Lindau disease (VHL), von Recklinghausen's disease (neurofibromatosis type 1), and tuberous sclerosis (Bourneville's disease). Inherited mutations in the VHL gene, located on chromosome 3p25, are associated with the development of cerebellar hemangioblastomas, renal cancer, and pheochromocytomas and, less commonly, pancreatic NETs. Mutations in neurofibromin (*NF1*) are associated with neurofibromatosis (von Recklinghausen's disease); patients with neurofibromatosis are at risk of developing both pancreatic and extrapancreatic NETs. Tuberous sclerosis is caused by mutations that alter either hamartin (*TSC1*) or tuberin (*TSC2*). Both hamartin and tuberin function as inhibitors of the phosphatidylinositol 3-kinase and the mechanistic target of rapamycin (mTOR) signaling cascades, and pancreatic NETs have been reported in these patients. Rare cases of familial small intestine NETs have also been reported; in these cases, multiple synchronous tumors generally arise within the small intestine. A characteristic inherited mutation has not been identified to date in the majority of these cases.

HISTOLOGIC CLASSIFICATION AND MOLECULAR FEATURES

The histologic features of NETs vary widely and are one of the most important determinants of both clinical behavior and treatment. NETs are classified based on the degree tumor differentiation (well or poorly differentiated), as assessed by a pathologist, and tumor grade (grades 1–3) **(Table 89-1)**. Tumor grade closely correlates with mitotic count and Ki-67 proliferative index. Classic, well-differentiated NETs are composed of monotonous sheets of small round cells with uniform nuclei and only rare mitoses. Immunocytochemical staining for chromogranins and synaptophysin is typical. Ultrastructurally, these tumors contain electron-dense neurosecretory granules containing peptides and bioactive amines that may be ectopically secreted, giving rise to a range of clinical syndromes. These classic well-differentiated NETs have low-grade features and generally have a mitotic index of <2 mitoses per 10 high-power fields (HPFs) and a Ki-67 proliferative index of <3%. Less commonly, well-differentiated NETs have an intermediate histologic grade and pursue a somewhat more aggressive clinical course. Intermediate-grade tumors typically have a mitotic count of 2–20 per 10 HPF and a mitotic index of 3–20%. Well-differentiated high-grade tumors are rare and have mitotic counts that exceed 20 per 10 HPF and a Ki-67 proliferative index of >20%. Poorly differentiated high-grade tumors form the most clinically aggressive category; prognosis and treatment for these tumors differ markedly from their well-differentiated counterparts.

Whole exome sequencing of sporadic pancreatic NETs has shown that the most frequently altered gene was *MEN1*, occurring in 44% of tumors. In addition, 43% of tumors had mutations in genes encoding two subunits of a transcription/chromatin remodeling complex consisting of DAXX (death-domain-associated protein) and ATRX (α-thalassemia/mental retardation syndrome X-linked). Mutations in genes associated with the mTOR pathway were identified in 15% of tumors. In contrast, recurrent mutations in extrapancreatic NETs appear to be rare. In one study that evaluated 180 small intestinal NETs using a combination of whole exome and more targeted genome-sequencing analysis, recurrent mutations were only observed in the *CDKN1B* gene (cyclin-dependent kinase inhibitor 1B [p27^{Kip1}]) in 8% of cases. Loss of chromosome 18 is a common finding in small-bowel NETs. Small-intestinal GI carcinoids commonly have epigenetic changes; however, the clinical significance of these alterations remains uncertain.

CLINICAL PRESENTATION AND MANAGEMENT OF LOCALIZED PANCREATIC NEUROENDOCRINE TUMORS

Pancreatic NETs have been subcategorized as either "functional," meaning associated with symptoms of hormone secretion, or nonfunctional, in which case they may be clinically silent until they cause anatomic symptoms. The clinical presentation of functional pancreatic NETs depends on the type of hormone secreted and can sometimes lead to dramatic clinical presentations **(Table 89-2)**. The most common functional pancreatic NETs are insulinomas, followed in incidence by glucagonomas and gastrinomas. Pancreatic NETs secreting other hormones, including somatostatin, vasoactive intestinal peptide (VIP), adrenocorticotropic hormone (ACTH), and parathyroid hormone (PTH), have also been described but are uncommon. Only ~20% of pancreatic NETs are associated with symptoms of hormone

TABLE 89-1 Histologic Classification of Neuroendocrine Tumors

CLASSIFICATION	DIFFERENTIATION	GRADE	MITOTIC COUNT	KI-67
Neuroendocrine tumor	Well differentiated	Low grade (grade 1)	<2 per 10 HPF	<3%
Neuroendocrine tumor	Well differentiated	Intermediate grade (grade 2)	2–20 per 10 HPF	3–20%
Neuroendocrine tumor	Well differentiated	High grade (grade 3)	>20 per 10 HPF	>20%
Neuroendocrine carcinoma	Poorly differentiated	High grade (grade 3)	>20 per 10 HPF	>20%

Abbreviation: HPF, high-power field.

TABLE 89-2 Clinical Presentation and Management of Secretory Syndromes Associated with Neuroendocrine Tumors

	CLINICAL SYMPTOMS AND MANIFESTATIONS	TREATMENT OPTIONS TO CONTROL SECRETORY SYMPTOMS
Pancreatic Neuroendocrine Tumors		
Gastrinoma (generally located in "gastrinoma triangle")	Zollinger-Ellison syndrome: gastroesophageal reflux, peptic ulcer disease, diarrhea	Proton pump inhibitors, somatostatin analogues
Insulinoma	Hypoglycemia leading to confusion, lethargy, coma; weight gain	Diazoxide, everolimus[a]
Glucagonoma	Skin rash (necrolytic migratory erythema), glucose intolerance, weight loss	Somatostatin analogues
VIPoma	Verner-Morrison syndrome: watery diarrhea, hypokalemia, achlorhydria	Somatostatin analogues
ACTHoma	Cushing's syndrome: hyperglycemia, weight gain, hypokalemia	Ketoconazole, metyrapone, consider adrenalectomy
Extrapancreatic Gastrointestinal Neuroendocrine Tumors		
Typically in setting of advanced disease from small intestine or appendiceal primary tumors	Carcinoid syndrome: flushing, diarrhea, right-sided valvular heart disease, mesenteric fibrosis	Somatostatin analogues, telotristat ethyl

[a] Successful use of monoclonal anti-insulin receptor antibodies to treat insulin-induced hypoglycemia has been reported but remains investigational.

Abbreviations: ACTH, adrenocorticotropic hormone; VIP, vasoactive intestinal peptide.

hypersecretion; the majority of pancreatic NETs are "nonfunctional" and are diagnosed either incidentally or after patients present with abdominal pain, weight loss, or other anatomic symptoms related to tumor bulk.

GASTRINOMA

Patients with gastrinoma typically present with Zollinger-Ellison syndrome (ZES) **(Chap. 335)**. The most common symptoms associated with this syndrome are abdominal pain, diarrhea, gastroesophageal reflux disease (GERD), and peptic ulcer disease. Peptic ulcer disease manifesting as multiple ulcers with associated diarrhea is a classic presentation. Up to 25% of patients with ZES have MEN 1, and a diagnosis of gastrinoma should prompt a family history as well as an assessment for concurrent hyperparathyroidism. Fasting hypergastrinemia is a nearly universal finding in patients with gastrinoma. Importantly, however, proton pump inhibitors (PPIs) can suppress acid secretion sufficiently to cause hypergastrinemia and can confound the diagnosis. Achlorhydria, usually in the context of chronic atrophic gastritis, will also elevate serum gastrin levels but can usually be easily distinguished from gastrinoma given the absence of other evidence of acid hypersecretion.

While often classified as pancreatic NETs, the majority of gastrinomas in fact arise in the "gastrinoma triangle," an anatomic region bounded by the duodenum, pancreas, and confluence of the cystic and common bile ducts. Most gastrinomas (50–90%) in sporadic ZES arise in the duodenum. They are frequently small and may be difficult to localize. Imaging studies generally include either computed tomography (CT) or magnetic resonance imaging (MRI); endoscopic ultrasound or somatostatin scintigraphy may also be helpful.

PPIs are generally highly effective in the treatment of symptoms related to gastrinoma and are considered the initial treatment of choice. Rapid resolution of both abdominal pain and diarrhea related to acid hypersecretion is common. Somatostatin analogues may also be helpful in controlling symptoms in refractory cases. Once symptoms are controlled, surgical resection is generally recommended for patients with sporadic gastrinomas, both to eliminate the cause of gastrin secretion and to decrease the risk of developing metastatic disease. The technique used for resection depends in large part on the precise location of the tumor. In some cases where preoperative imaging is not successful but a diagnosis is strongly suspected, exploratory laparotomy with intraoperative ultrasound may be undertaken. In gastrinoma patients who have underlying MEN 1, tumors are generally small and multiple; the role of routine surgery in this setting remains more controversial but generally is still recommended in patients with larger tumors measuring ≥1.5–2 cm in diameter.

INSULINOMA

Patients with insulinoma generally present with symptoms of hypoglycemia, which may include confusion, headache, disorientation, visual difficulties, irrational behavior, and even coma. In some cases, the diagnosis of insulinoma may not be immediately evident, and patients with insulinoma may initially be diagnosed with psychiatric illnesses that in retrospect were hypoglycemic symptoms. The diagnosis of insulinoma is generally confirmed with elevated fasting insulin levels in conjunction with elevated proinsulin and C-peptide. Fasting hypoglycemia can also be caused by severe liver disease, alcoholism, and poor nutrition. Postprandial hypoglycemia may also occur after gastric bypass surgery. Surreptitious use of insulin or hypoglycemic agents may be difficult to distinguish from an insulinoma. Evaluation of proinsulin and C-peptide levels, both of which should be normal in patients using exogenous insulin, and measurement of sulfonylurea levels in serum or plasma are helpful in such cases.

The hypoglycemia associated with insulinomas can be severe and challenging to manage. While somatostatin analogues are usually effective in treating symptoms of hormone hypersecretion associated with other types of NETs, they should be used with caution in patients with insulinoma. Somatostatin analogues may suppress counterregulatory hormones, such as growth hormone (GH), glucagon, and catecholamines, and precipitously worsen hypoglycemia. Diazoxide has historically been used in the initial management of patients with insulinoma and results in inhibition of insulin release, though it can also be associated with side effects including sodium retention and nausea. Everolimus, in addition to its antitumor effect (see below), is effective in improving glycemic control in patients with insulinoma. The benefits of everolimus in this setting may be related both to induction of insulin resistance and a direct antitumor effect. The use of anti-insulin receptor monoclonal antibodies may be highly effective in treating the hypoglycemia associated with insulinoma, although currently their use for this indication remains investigational Insulinomas may be difficult to localize, as they are less consistently avid on somatostatin scintigraphy than other pancreatic NETs. Insulinomas are also generally small, with the majority measuring <2 cm in diameter. Because of their generally small size, insulinomas are best localized with endoscopic ultrasound (EUS). In the absence of metastatic disease, surgical resection is usually recommended. The primary treatment for exophytic or peripheral insulinomas is enucleation. If enucleation is not possible because of invasion or the location of the tumor within the pancreas, then pancreatoduodenectomy for tumors in the head of the pancreas or distal pancreatectomy with preservation of the spleen for smaller tumors not involving splenic vessels may be considered.

GLUCAGONOMA

Patients with glucagonoma most commonly present with a characteristic dermatitis, called necrolytic migratory erythema **(Fig. 89-2)**. The rash usually involves intertriginous sites, especially in the groin or buttock, and can wax and wane. Other common presenting symptoms of glucagonoma include glucose intolerance and weight loss. The diagnosis of glucagonoma can be confirmed by demonstrating an increased plasma glucagon level, generally in excess of 1000 pg/mL. Somatostatin analogues are usually highly effective as an initial treatment to alleviate the symptoms and rash associated with glucagon hypersecretion. The majority of glucagonomas are large in size at presentation and arise in

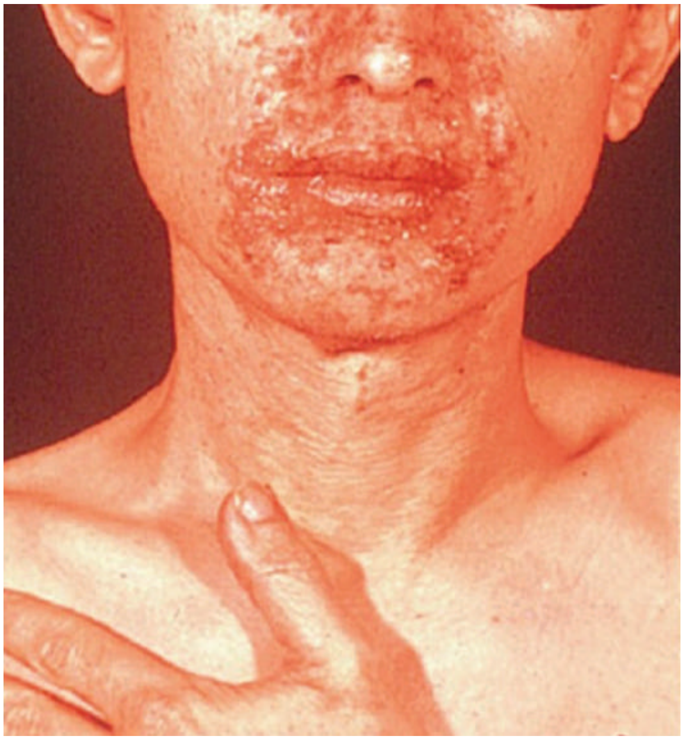

FIGURE 89-2 Glucagonoma syndrome. Patients with glucagonoma may present with a classic skin rash, necrolytic migratory erythema (shown). Other presenting symptoms include glucose intolerance and weight loss.

the tail of the pancreas. For patients with localized disease, distal pancreatectomy and splenectomy are recommended. A hypercoagulable state has been reported in up to 33% of patients with glucagonoma, and perioperative anticoagulation should generally be employed.

SOMATOSTATINOMA

Patients with somatostatinoma typically present with diabetes mellitus, gallbladder disease, diarrhea, and steatorrhea. Somatostatinomas occur primarily in the pancreas or duodenum, are usually large, and are commonly metastatic at presentation. They are only rarely associated with MEN 1. The diagnosis of somatostatinoma is based on the demonstration of elevated plasma somatostatin levels, and as such, the potential benefits of using somatostatin analogues as a treatment for patients with somatostatinoma are questionable. Surgery is recommended for patients with localized disease.

VIPOMA

VIPomas are associated with a distinct syndrome that has been variously called Verner-Morrison syndrome, pancreatic cholera, and WDHA syndrome (*w*atery *d*iarrhea, *h*ypokalemia, and *a*chlorhydria). VIP is a 28-amino-acid peptide that mimics the effects of the cholera toxin by stimulating chloride secretion in the small intestine and increasing smooth-muscle contractility, resulting in profound diarrhea. Treatment of dehydration, hypokalemia, and electrolyte losses with fluid and electrolyte replacement is the most critical initial treatment for patients with VIPoma. VIPomas are usually solitary and arise in the pancreatic tail. Elevated plasma levels of VIP are typical but should not be the only basis of the diagnosis of VIPomas because they can occur with some diarrheal states including inflammatory bowel disease, in the setting of small-bowel resection, and radiation enteritis. Chronic surreptitious use of laxatives/diuretics can be particularly difficult to detect clinically. Somatostatin analogues are usually highly effective in controlling the diarrhea; surgical resection is recommended for patients with localized disease.

OTHER SECRETORY PANCREATIC NETS

Pancreatic NETs secreting GH-releasing factor (GRF), calcitonin, ACTH, and PTH-related protein have also been described; it is also possible for pancreatic NETs to secrete more than one hormone or for the secretory profiles to evolve over time. Gastrinomas, in particular, may evolve and may be associated with secretion of ACTH, resulting in ectopic Cushing's syndrome. Tumors secreting these hormones may not be as responsive to treatment with somatostatin analogues as the more common pancreatic NETs and the associated hormonal symptoms may cause significant morbidity. As with other pancreatic NETs, patients with localized disease are generally treated with surgical resection. In patients with ACTH-secreting tumors, the associated symptoms of Cushing's syndrome can be alleviated with the use of metyrapone, an agent that directly inhibits cortisol synthesis. Adrenalectomy may also be considered if resection of the primary tumor is not possible or in the setting of metastatic disease.

PANCREATIC NETS ARISING IN THE SETTING OF MEN 1

Pancreatic NETs occurring in patients with MEN 1 are typically multiple and often pursue a relatively indolent course. Because of the high probability of multiple tumors, surgical resection of confirmed pancreatic NETs in patients with MEN 1 is usually undertaken with caution given the likelihood of tumors arising in the remaining pancreas if partial pancreatectomy is undertaken as well as the significant morbidities associated with total pancreatectomy. However, for symptomatic tumors or for growing tumors >2 cm in size, surgical resection may still be considered.

NONFUNCTIONING PANCREATIC NETS

As noted above, the majority of pancreatic NETs are not associated with symptoms of hormone hypersecretion and are considered "nonfunctional." As a result, they often remain clinically silent and either are diagnosed incidentally or are not diagnosed until widespread, metastatic disease is present resulting in anatomic symptoms. If they are localized at diagnosis, the general treatment recommendation is surgical resection; however, the management of small, asymptomatic pancreatic NETs is debated. Assuming tumors are low grade, patients with incidentally discovered, low-grade tumors measuring <1 cm in size can generally be safely followed. However, some studies have suggested that at least some tumors measuring <2 cm in size can pursue a more aggressive course and that surgical resection may be warranted in some cases. Management of small, incidentally discovered, asymptomatic, low-grade pancreatic NETs is therefore based on clinical judgement, taking into account surgical risk and patient comorbidities. For tumors measuring >2 cm in diameter, metastases pose a significant risk and surgical resection is generally recommended in patients for whom surgery is not contraindicated.

CLINICAL PRESENTATION AND MANAGEMENT OF LOCALIZED EXTRAPANCREATIC GASTROINTESTINAL NEUROENDOCRINE TUMORS

Extrapancreatic GI NETs, historically called carcinoid tumors, may arise virtually anywhere in the GI tract and differ significantly in their clinical characteristics depending on their location. The most common locations for extrapancreatic NETs are the stomach, distal small intestine, appendix, and rectum.

GASTRIC NETS

Gastric NETs can be categorized into three groups: type 1 (associated with chronic atrophic gastritis); type 2 (associated with gastrinomas and ZES), and type 3 (sporadic, gastric NETs). Type 1 gastric NETs are the most common of the three types. In type 1 gastric NETs, chronic atrophic gastritis results in loss of acid secretion with consequent loss of the negative feedback loop on gastrin-producing cells in the antrum of the stomach. Pernicious anemia is also commonly associated with this condition; classic laboratory findings are a markedly elevated gastrin level and low levels of vitamin B_{12}. Unchecked gastrin secretion in these patients results in hyperplasia of the endocrine cells in the gastric fundus. A typical finding on endoscopy is diffuse endocrine cell hyperplasia with multiple gastric carcinoid tumors (Fig. 89-3). These

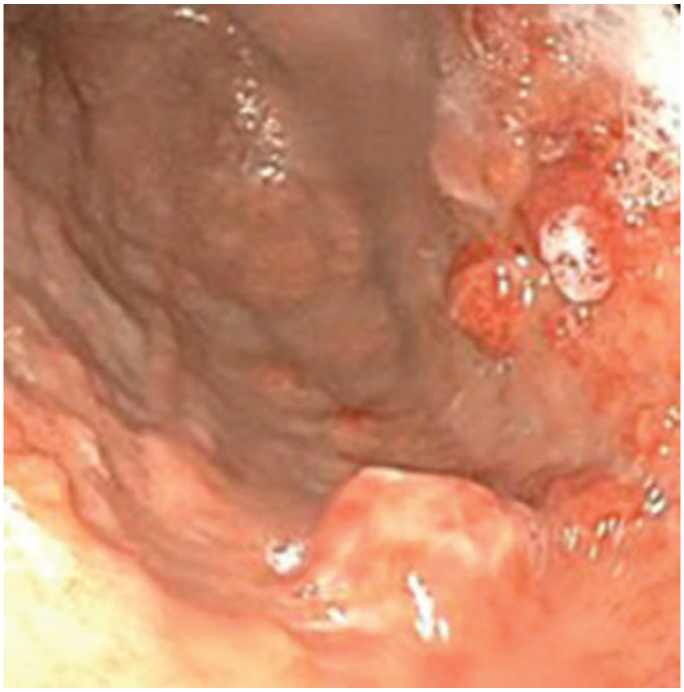

FIGURE 89-3 Multifocal gastric neuroendocrine tumor. *(Courtesy of Christopher Huang, MD, Boston Medical Center.)*

tumors generally pursue a benign course and can be monitored with serial endoscopy. In cases where tumors continue to grow or become symptomatic, antrectomy to remove the source of gastrin production can result in tumor regression. Type 2 tumors are rare and usually occur in the setting of gastrinoma; as with type 1 gastric NETs, elevated gastrin levels result in diffuse gastric neuroendocrine hyperplasia and multifocal gastric NETs. Resection of the gastrinoma, removing the source of gastrin production, is the treatment of choice.

In contrast to type 1 and type 2 gastric NETs, type 3 gastric NETs are generally solitary, arise in the setting of normal gastrin levels, and may pursue a far more aggressive course. For early-stage, smaller tumors, endoscopic or wedge resection may be performed. For larger tumors, partial gastrectomy with lymphadenectomy is recommended.

■ NETS OF THE SMALL INTESTINE

Small-bowel NETs occur most commonly in the terminal ileum and are notoriously difficult to diagnose at an early stage. One reason for this is that they arise within the muscularis, and their submucosal location makes them difficult to see during routine colonoscopy **(Fig. 89-4*A*)**. Small-bowel NETs are also often multifocal; multifocal tumors appear to arise independently throughout the small intestine, although the mechanisms underlying this phenomenon remain uncertain.

Small-bowel NETs are often associated with desmoplasia and mesenteric fibrosis, likely as a result of fibroblast proliferation stimulated by tumor serotonin secretion. Mesenteric fibrosis frequently results in intermittent small-bowel obstruction and, in some cases, bowel ischemia due to involvement of the mesenteric vessels. Patients may experience symptoms of intermittent abdominal pain and associated diarrhea, sometimes for months or years before diagnosis, that because of the difficulty in diagnosis are often attributed to irritable bowel syndrome. One classic finding that can aid in diagnosis is that the lymph node metastases associated with small intestine NETs are usually larger than the primary tumor and may be calcified, which, together with the tethering of the small intestine caused by the associated fibrosis, results in a classic "spoke and wheel" appearance on CT **(Fig. 89-4*B*)**.

Surgical resection of the primary tumor and associated metastases is recommended when feasible and is performed with curative intent when distant metastatic disease is not already present. Resection should also be considered in patients with metastatic disease experiencing intermittent obstruction or abdominal discomfort thought to be related to the primary tumor or associated mesenteric disease. Some have also advocated the routine resection of asymptomatic small-bowel primary tumors in patients with distant metastatic disease, with the rationale that this may be a way to prevent the future development of fibrosis and obstruction and preempt the development of unresectable disease due to tumor involvement of the mesenteric vessels. However, the available data on the benefits of resecting an asymptomatic primary tumor in this context are conflicting. Some studies have suggested that this practice results in an overall survival benefit, but the retrospective nature of these studies makes the data difficult to interpret given the high potential for selection bias in patients taken to surgery compared with those who were not. Other studies have suggested that prophylactic primary tumor resection confers no survival benefit and that surgery can be safely delayed until it is indicated based on the development of symptoms.

■ NETS OF THE APPENDIX

NETs are one of the most common tumors arising in the appendix. They are typically discovered incidentally in younger individuals undergoing appendectomy for acute appendicitis and not uncommonly are identified only at the time of pathology review. While the unexpected diagnosis of an appendiceal NET in such situations can cause considerable anxiety, in the majority of cases, the prognosis is excellent. Indeed, the clinical behavior of appendiceal NETs has been inferred from multiple large retrospective surgical series that suggest that the risk of lymph node or distant metastases from appendiceal NETs with well-differentiated histology and a tumor diameter measuring <2 cm is extremely low. In such cases, appendectomy alone is felt to be a sufficient surgical procedure.

In contrast, the risk of metastases for tumors measuring 2–3 cm is ~20–30% and is even greater for tumors measuring >3 cm. For patients with larger tumors, more formal staging studies with either

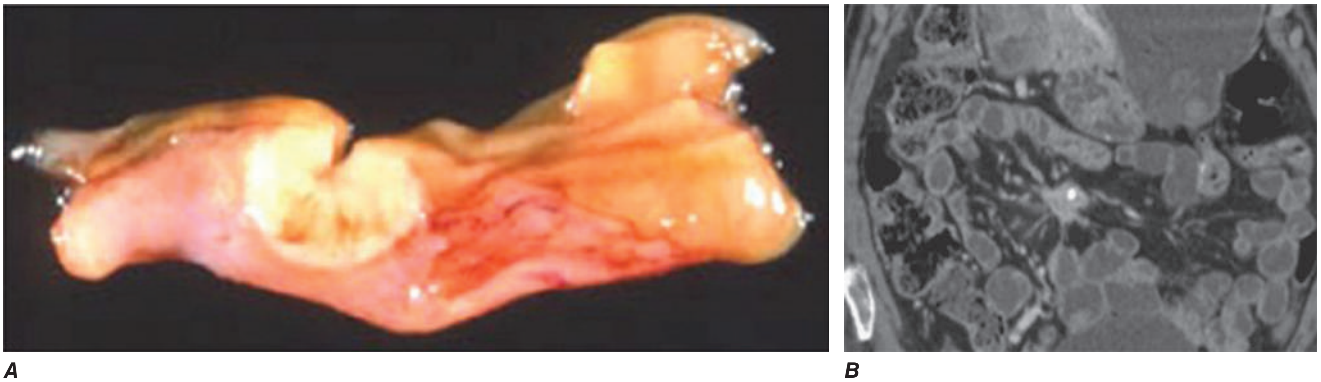

FIGURE 89-4 Small intestine neuroendocrine tumor. *A.* Small intestine neuroendocrine tumors arising in submucosal location. The submucosal location of small intestine neuroendocrine tumors, together with their location beyond the ileocecal valve in the terminal ileum, can make endoscopic detection challenging. ***B.*** Classic "spoke and wheel" appearance of calcified mesenteric mass associated with small intestine primary neuroendocrine tumor. Mesenteric fibrosis commonly leads to intermittent obstructive symptoms and can also lead to ischemia when the mesenteric vasculature is involved. *(Fig. B: Courtesy of Christina LeBedis, MD, Boston Medical Center.)*

cross-sectional imaging or somatostatin scintigraphy are generally recommended to assess for distant metastases, and a subsequent right colectomy to remove regional lymph nodes is performed if no distant metastases are observed. Whether right colectomy should be performed for tumors measuring <2 cm with features such as mesoappendiceal invasion or tumor origin at the appendiceal base, which in some series have suggested a poorer prognosis, remains uncertain. Additionally, tumors may arise in which neuroendocrine cells are admixed with mucin-producing cells or cells exhibiting features of frank adenocarcinoma. In such mixed neuroendocrine-adenocarcinoma tumors, sometimes termed "adenocarcinoids," treatment recommendations are generally dictated by the more aggressive component of the tumor and align with typical recommendations for colorectal adenocarcinoma.

■ RECTAL NETS

With the increased use of screening colonoscopy, the diagnosis of rectal NET has also become more common. For unclear reasons, the incidence of rectal carcinoid tumors shows geographic variation. In European studies, they compose up to 14% of all NETs, while in some Asian series (Japan, China, Korea), they compose up to 90% of all NETs. The majority of rectal NETs are small, measuring <1 cm in diameter, and have well-differentiated histology. These tumors rarely metastasize and can usually be safely removed endoscopically with subsequent endoscopic monitoring. In contrast, up to one-third of rectal NETs between 1 and 2 cm are associated with metastases, and those >2 cm, though uncommon, metastasize in >70% of patients. When identified early, these tumors generally require a surgical resection. In contrast to NETs of the appendix and small intestine, hormone secretion from rectal NETs, even when metastatic, is exceedingly rare.

CLINICAL PRESENTATION, DIAGNOSIS, AND EVALUATION OF PATIENTS WITH METASTATIC NEUROENDOCRINE TUMORS

While patients who undergo resection of localized NETs may be at risk of developing tumor recurrence or metastatic disease, postoperative treatment has not yet been shown to alter the risk of recurrence, and systemic adjuvant therapy is not recommended following resection of well-differentiated NETs, as it is for some other cancers. Whether adjuvant systemic therapy may be of benefit following resection of high-grade NETs is uncertain, and an approach utilizing platinum-based chemotherapy with or without external-beam radiation, analogous to that used in small-cell carcinoma, is sometimes considered.

The evaluation of patients with known or suspected metastatic disease generally includes both standard cross-sectional imaging such as CT or MRI and somatostatin scintigraphy. Somatostatin scintigraphy in this setting is based on the fact that >90% of NETs express somatostatin receptors. Gallium-68 (^{68}GA) dotatate, as well as ^{68}GA DOTATOC and copper-64 dotatate, are all radioligands bound to a somatostatin analogue and can be used as a nuclear medicine tracer to perform positron emission tomography (PET) scanning that is highly sensitive in detecting both primary NETs and metastases **(Fig. 89-5)**. Because of the sensitivity of these approaches, false-positive results can occur due to somatostatin receptor expression in other tissues. Physiologic uptake in the pancreatic uncinated process is common; uptake can also occur in the setting of sarcoidosis, in meningiomas, and in thyroid goiter or thyroiditis. Standard fluorodeoxyglucose (FDG) PET scans are often negative in well-differentiated NET due to their low metabolic activity but can show uptake in higher-grade tumors; conversely, rates of somatostatin expression tend to be lower in higher-grade tumors, and ^{68}GA dotatate scans may be negative in this setting.

The utility of blood-based tumor markers in NETs is controversial. The circulating tumor marker chromogranin A is commonly used as a screen for the presence of NETs and also to monitor for both recurrence and progression of disease in patients with known metastases. While chromogranin A is elevated in patients with metastatic NETs, it is neither particularly sensitive nor specific. A broad range of different assays for chromogranin A have also posed challenges in interpreting results in a standardized fashion. Chromogranin A is often elevated in a number of nonmalignant conditions, including in patients with impaired renal function and in patients who are taking PPIs. Elevated values of chromogranin A should be interpreted with caution in patients in whom a NET is being considered but in whom a diagnosis has not been established.

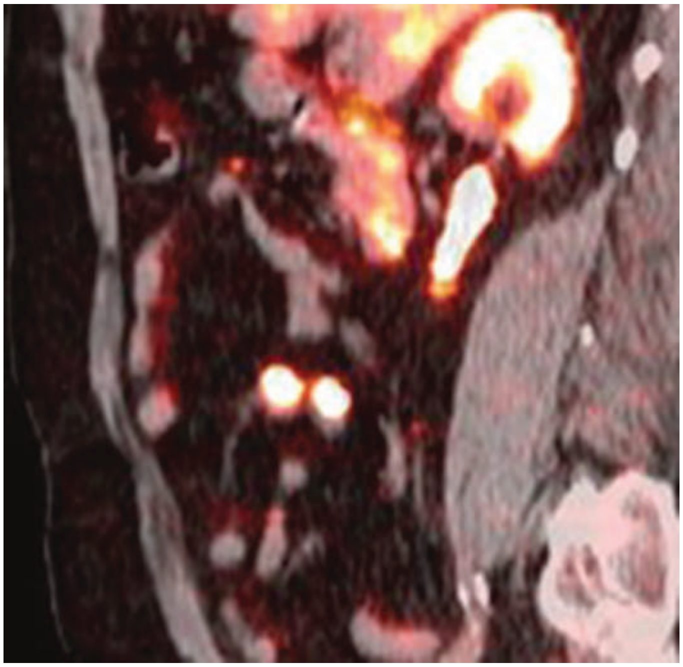

FIGURE 89-5 Gallium-68 dotatate positron emission tomography (PET) scan demonstrating a small-bowel neuroendocrine tumor and associated mesenteric mass. *(Courtesy of Sara Meibom, MD, Boston Medical Center.)*

The overall survival durations for patients with metastatic NETs vary significantly, depending on both the primary location of the tumor and the histologic grade. Median survival durations for patients with well-differentiated NETs have markedly increased in recent years, likely reflecting both earlier diagnoses and improved treatments. For example, in early analyses of the SEER database, the median survival for patients with advanced pancreatic NETs was ~2 years; this had increased to 4 years in a more recent analysis. Similar increases were observed in patients with advanced small intestine NETs, where the median survival for patients with well-differentiated small intestine NETs exceeds 5 years. The sometimes prolonged survival of patients with NETs can make it challenging to determine at what point to initiate treatment. In patients with symptoms of hormone secretion, decisions to initiate therapy are straightforward. In asymptomatic patients, on the other hand, observation off treatment can sometimes be appropriate. Nevertheless, the natural course of NETs is ultimately to progress, and if treatment is not initiated, close monitoring is essential to ensure patients maximize access to available treatment options over the course of their disease.

MANAGEMENT OF SYMPTOMS OF HORMONE HYPERSECRETION AND THE CARCINOID SYNDROME

Patients with advanced NETs may in some cases experience more symptoms from hormone hypersecretion than from tumor bulk. The management of hormonal symptoms associated with pancreatic NETs depends on the hormone being secreted (see above). Patients with GI NETs, particularly those with small intestine or appendiceal primaries, may develop the carcinoid syndrome. Flushing and diarrhea are the two most common symptoms associated with carcinoid syndrome. The characteristic flush is of sudden onset; it is a deep red or violaceous erythema of the upper body, especially the neck and face, often associated with a feeling of warmth. Flushes may be precipitated by stress, alcohol, exercise, and certain foods such as cheese. Flushing episodes initially are brief, lasting 2–5 min, though later in the disease course, they may last hours. The diarrhea associated with carcinoid syndrome

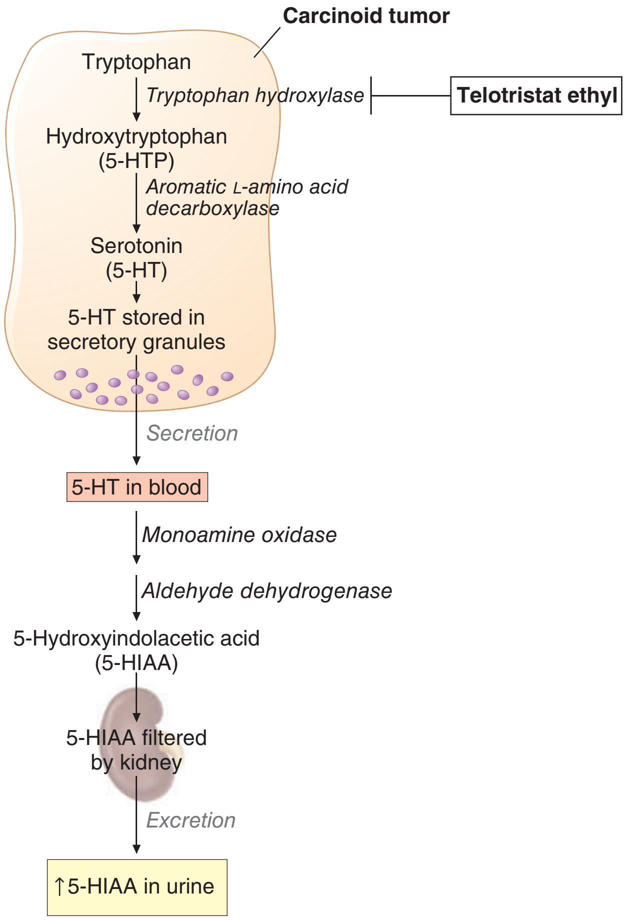

FIGURE 89-6 Serotonin synthesis and secretion in neuroendocrine tumors. Tryptophan is converted to hydroxytryptophan by tryptophan hydroxylase within the tumor cell and, subsequently, to serotonin (5-HT). Serotonin is subsequently converted to 5-hydroxyindole acetic acid (5-HIAA), which can be measured in a 24-h urine collection and can facilitate the diagnosis of carcinoid syndrome. Telotristat ethyl inhibits tryptophan hydroxylase and can be used as a treatment for carcinoid syndrome.

may or may not be associated with flushing and is described as watery in nature. Diarrhea can be profound, sometimes occurring in excess of 10 times daily and is one of the symptoms that most significantly interferes with activities of daily living. Less common manifestations of the carcinoid syndrome include wheezing or asthma-like symptoms. Impaired cognitive function has also been described in particularly advanced cases.

The main secretory product implicated in the carcinoid syndrome is serotonin (5-HT). Serotonin is synthesized from tryptophan by the enzyme tryptophan hydroxylase **(Fig. 89-6)**. Up to 50% of dietary tryptophan can be used in this synthetic pathway by tumor cells, resulting in inadequate supplies for conversion to niacin; hence, some patients develop symptoms of niacin deficiency and pellagra-like lesions. Serotonin has numerous biologic effects, including the stimulation of intestinal secretion, increasing intestinal motility, and the stimulation of fibroblast growth. Other secreted products contributing to carcinoid syndrome symptoms are thought to include histamines and tachykinins, including substance P.

■ DIAGNOSIS AND TREATMENT OF THE CARCINOID SYNDROME

While the carcinoid syndrome can develop in patients with NETs from almost any site, it is most commonly associated with appendiceal or small intestine NETs. In these patients, the syndrome usually develops only after the development of hepatic metastases or retroperitoneal lesions, allowing entry of serotonin and other vasoactive substances into the systemic circulation. While serotonin levels can be measured in plasma, such measurements are frequently highly variable. Evidence of excess serotonin secretion can be more reliably confirmed by measuring levels of the serotonin metabolite 5-hydroxyindole acetic acid (5-HIAA), either in plasma or using a 24-h urine collection. Urine collections can be challenging, and false-positive elevations may occur if the patient is eating serotonin-rich foods (e.g., salmon, eggs). As a result, elevated levels of 5-HIAA are suggestive but not diagnostic of the carcinoid syndrome. Patients with NETs may also experience symptoms of carcinoid syndrome related to other secreted products, including histamine, absent evidence of serotonin secretion. Conversely, patients without NETs may also describe symptoms analogous to carcinoid syndrome but due to other causes. The symptoms associated with systemic mastocytosis, in particular, can be easily confused with carcinoid syndrome.

The symptoms of carcinoid syndrome, including diarrhea, are generally refractory to standard antidiarrheals or other traditional medications but can often be well controlled with somatostatin analogues **(Fig. 89-7)**. Somatostatin is a 14-amino-acid peptide that inhibits the secretion of a broad range of hormones. Due to its short half-life, administration of somatostatin itself is not therapeutically practical. Longer-acting somatostatin analogues, including octreotide and lanreotide, share an 8-amino-acid binding domain with naturally occurring somatostatin and bind primarily to somatostatin receptor subtypes 2 and 5. Both have been shown to be effective in the treatment of carcinoid syndrome.

The presence of somatostatin receptors on NETs is predictive of response to somatostatin analogues and can be easily confirmed with uptake on somatostatin scintigraphy such as ^{68}GA dotatate PET scan.

Somatostatin analogue side effects are generally mild. Mild nausea, abdominal discomfort, bloating, and loose stools occur in up to one-third of patients during the first month or two of treatment but usually subsequently subside. Patients with persistent symptoms of bloating or loose stools may be experiencing pancreatic insufficiency associated with use of somatostatin analogues; use of pancreatic enzyme supplements can ameliorate these symptoms. Mild glucose intolerance may also occur due to inhibition of insulin secretion. One of the more significant side effects associated with somatostatin analogues is impaired gallbladder contractility, resulting in delayed gallbladder emptying, and long-term administration of somatostatin analogues has been associated with an increased risk of cholelithiasis. For this reason, patients with advanced NETs in whom surgery is planned and for whom somatostatin analogue therapy is being considered should generally also undergo prophylactic cholecystectomy.

Over time, for reasons that remain uncertain, patients receiving somatostatin analogues for symptoms of hormone secretion may become refractory to treatment. Not uncommonly, such patients experience symptom exacerbation toward the final week of each treatment cycle. Such patients may benefit from an increased frequency of administration (i.e., every 3 weeks) or use of additional short-acting octreotide for breakthrough symptoms.

The association between high levels of circulating serotonin and symptoms of the carcinoid syndrome has also led to efforts aiming to directly inhibit serotonin synthesis (Fig. 89-6). This approach was first undertaken in the late 1960s with the drug para-chlorophenylalanine, which was reported to reduce symptoms of carcinoid syndrome but also caused significant central nervous system (CNS) side effects. Telotristat ethyl, a tryptophan hydroxylase inhibitor with minimal CNS penetration, was evaluated in a randomized trial that enrolled 135 patients with persistent carcinoid syndrome–related diarrhea while receiving somatostatin analogues. Treatment with telotristat ethyl was associated with a reduction in bowel movement frequency as well as significant decreases in urinary 5-HIAA compared to placebo, consistent with its mechanism of directly inhibiting serotonin synthesis. Thus, telotristat is a treatment option for patients with carcinoid syndrome who have persistent diarrhea despite treatment with somatostatin analogues.

■ CARCINOID CRISIS

Carcinoid crisis has been described in the setting of tumor manipulation during surgery and, less commonly, after other interventions such as hepatic artery embolization or radionuclide therapy. It may also occur as a result of exogenous administration of epinephrine or during

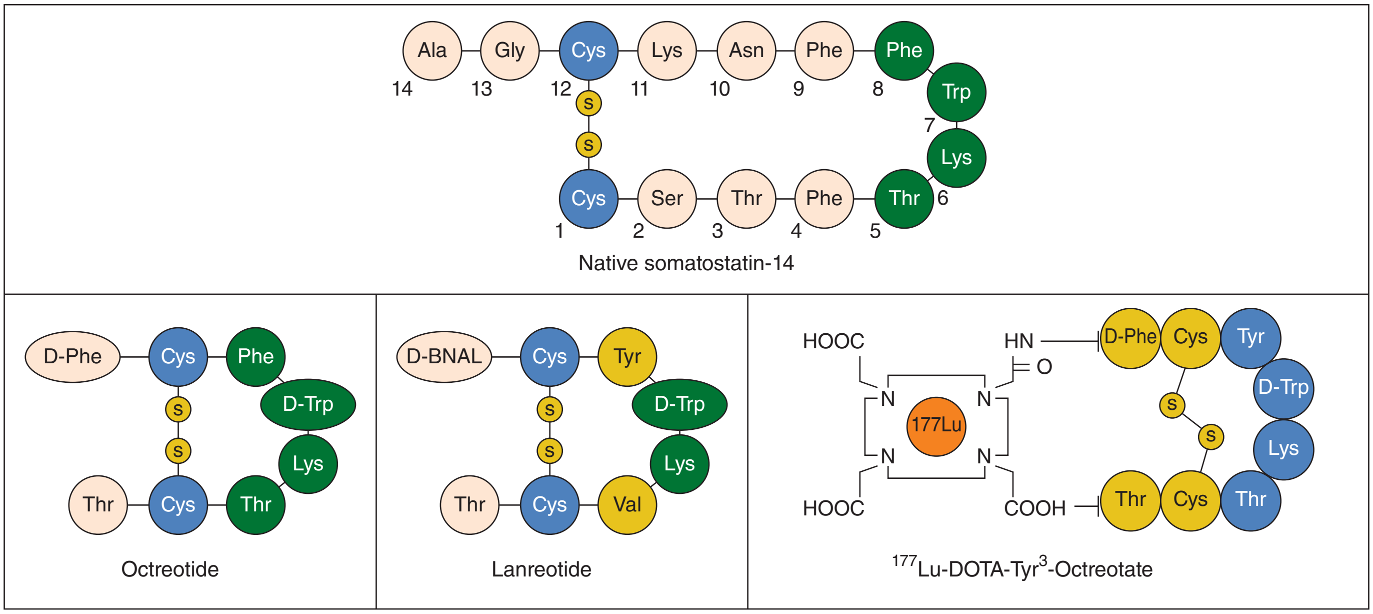

FIGURE 89-7 Somatostatin analogues. Commonly used somatostatin analogues include octreotide and lanreotide, which mirror the molecular structure of human somatostatin and bind to somatostatin receptors on neuroendocrine tumors. Somatostatin analogues inhibit tumoral hormone secretion and also have an antiproliferative effect. Radiolabeled somatostatin analogues such as ^{177}Lu-DOTA-octreotate, shown in the figure, share a similar molecular structure and are used therapeutically.

induction of anesthesia. It is most common in patients who already have significant symptoms of carcinoid syndrome and is thought to be caused by a sudden release of biologically active compounds from the tumor. Carcinoid crisis can be life-threatening and can manifest as either profound hypotension or hypertension. Prospective studies on the prevention and management of carcinoid crisis are limited; however, somatostatin analogues should be readily available during surgical procedures, and in some cases, continuous prophylactic intravenous administration of somatostatin analogues has been utilized as a way to mitigate risk.

CARCINOID HEART DISEASE

Carcinoid heart disease occurs in approximately two-thirds of patients with the carcinoid syndrome. Carcinoid heart lesions are characterized by plaque-like, fibrous endocardial thickening that classically involves the right side of the heart and often causes retraction and fixation of the leaflets of the tricuspid and pulmonary valves **(Fig. 89-8)**. The fibrosis in carcinoid heart disease is thought to be directly related to exposure of heart valve fibroblasts to high circulating levels of serotonin. Lesions similar to those observed in carcinoid heart disease were observed in patients receiving fenfluramine, a drug also known to increase serotonin signaling, as well as in patients receiving ergot-containing dopamine receptor agonists for Parkinson's disease. Metabolites of fenfluramine, as well as the dopamine receptor agonists, have high affinity for serotonin receptor subtype 5-HT$_{2B}$ receptors, whose activation is known to cause fibroblast mitogenesis and which are normally expressed in heart valve fibroblasts. These observations support the hypothesis that serotonin overproduction in patients with carcinoid syndrome mediates the valvular changes by activating 5-HT$_{2B}$ receptors in the endocardium.

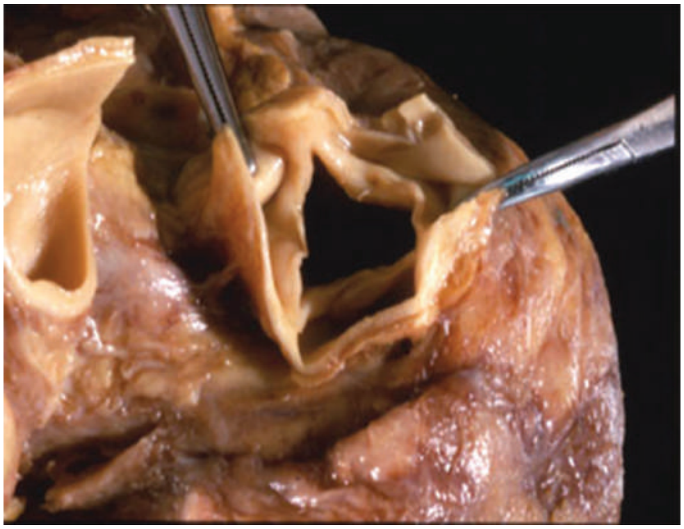

FIGURE 89-8 Carcinoid heart disease. Fibrosis secondary to elevated levels of circulating serotonin classically involves the tricuspid valve, resulting in valve retraction and tricuspid regurgitation.

Tricuspid regurgitation is a nearly universal feature of carcinoid heart disease; tricuspid stenosis, pulmonary regurgitation, and pulmonary stenosis may also occur. Left-sided heart disease occurs in <10% of patients and has been associated with the presence of a patent foramen ovale. The preponderance of lesions in the right heart is related directly to the fact that serotonin is secreted by liver metastases or retroperitoneal tumor deposits into the venous circulation and subsequently into the right atrium and right ventricle. The lower incidence of heart disease in the left heart is postulated to be due to the fact that serotonin is metabolized in the pulmonary vasculature before entering the left atrium and ventricle. Among patients with carcinoid syndrome, patients with heart disease exhibit higher levels of serum serotonin and urinary 5-HIAA excretion than patients without heart disease. Treatment with somatostatin analogues resulting in decreased serotonin secretion does not result in regression of cardiac lesions. Reduction of serotonin levels as a result of treatment with somatostatin analogues or with the tryptophan hydroxylase inhibitor telotristat ethyl seems likely to slow progression of carcinoid heart disease but has not been formally evaluated in clinical trials.

Right-sided heart failure in patients with carcinoid heart disease may lead to significant morbidity and mortality. The development of multiple new treatments to improve overall disease control in patients with advanced NETs has led to increased interest in valvular replacement, which may result in significant clinical benefit in appropriately selected patients with carcinoid heart disease. The appropriate timing of valve replacement in such patients can be challenging given the competing desires to perform surgery before the onset of severe right-sided heart failure, which can increase surgical morbidity, and the need to achieve adequate overall tumor control. However, advanced and less invasive techniques, including catheter-based valve replacement, have made valve replacement an increasingly attractive option for patients with this condition.

HEPATIC-DIRECTED THERAPY FOR METASTATIC NETS

The liver is one of the most common sites for metastases in patients with NETs and, in some cases, is the only site of metastatic disease. Hepatic-directed therapies can often be effective as a means of controlling, if not eliminating, metastases, particularly in patients who have more indolent tumors with well-differentiated histology. Common approaches for such patients include surgical resection, ablation or embolization, and orthotopic liver transplantation.

For patients with limited hepatic disease whose tumors have well-differentiated histology, surgical resection is generally considered the preferable option. While data are limited to retrospective series with the consequent risk of selection bias, long-term survival durations and symptomatic improvements reported in select populations of patients undergoing hepatic resection of neuroendocrine liver metastases compare favorably with outcomes associated with other management approaches, and 5-year survival rates approach 90% in some series. In patients in whom anatomy precludes resection or in whom a greater number of lesions are present, radiofrequency ablation or cryoablation can also be used, either as a primary treatment modality or as an adjunct to surgical resection. While ablation is considered to be less morbid than hepatic resection, it is generally utilized only in smaller tumors so that zones of ablation are limited.

In most cases, however, liver metastases are large, multiple, and involve both lobes of the liver. In such cases, the benefit of surgical resection and ablation is limited. Hepatic arterial embolization can be considered in these cases, assuming that extrahepatic disease remains relatively limited and that clinical benefit can be achieved by reducing hepatic tumor bulk. Hepatic artery embolization is based on the principle that tumors in the liver derive most of their blood supply from the hepatic artery, whereas healthy hepatocytes derive most of their blood supply from the portal vein. Multiple different embolization techniques have been explored, ranging from the simple infusion of gel foam powder into the hepatic artery (bland embolization) to the administration of chemotherapy or chemotherapy-eluting beads into the hepatic artery (chemoembolization) or the intra-arterial administration of radioisotope-tagged microspheres (radioembolization). Limited data suggest an optimal approach to embolization, and few studies have compared these approaches directly. Tumor response rates with all of these approaches generally exceed 50%. Specific approaches are therefore often tailored to the patient, taking into account tumor location, overall tumor burden, and comorbidities. Bland embolization, for example, may be associated with less morbidity, whereas chemoembolization or radioembolization may result in longer durations of response.

The role of orthotopic liver transplantation for the treatment of NETs remains uncertain. Data from available institutional series suggest that a small number of highly selected patients may achieve long-term survival. However, 5-year overall median survival durations in most series are ~50%, and the majority of patients undergoing hepatic transplantation develop tumor recurrence. Additionally, the widespread utility of hepatic transplantation is limited by organ availability. Decisions regarding proceeding with transplantation in patients with advanced NETs are therefore highly individualized.

SYSTEMIC TREATMENT TO CONTROL TUMOR GROWTH

While hepatic-directed therapies can be effective in the management of patients with liver-predominant disease, a majority of patients will either present with or ultimately develop more widespread metastases. A number of systemic treatment options have been developed and can be effective in treating such patients. These options include treatment with traditional somatostatin analogues, peptide receptor radioligand therapy, traditional cytotoxic chemotherapy, and an increasing array of molecularly targeted therapies targeting the mTOR or vascular endothelial growth factor (VEGF) pathways (Table 89-3). The choice and sequence of therapy depend in part on the type of tumor, the extent of disease, and patient symptoms and comorbidities.

TABLE 89-3 Selected Randomized Trials of Therapeutic Agents for the Treatment of Advanced Neuroendocrine Tumors (NETs)

TUMOR TYPE	NUMBER OF PATIENTS	PROGRESSION-FREE SURVIVAL
Pancreatic and Extrapancreatic NETs		
Lanreotide vs placebo (CLARINET)	204	65% vs 33% at 2 years ($p <.001$)
177-Lutetium dotatate vs octreotide (NETTER 2)	226 (limited to histologic grade 2 and 3 tumors)	22.8 vs 8.5 months ($p <.0001$)
Cabozantinib vs placebo (CABINET)	298 (203 extrapancreatic NETs [epNET] and 95 pancreatic NETs [pNET])	8.5 vs 4 months (epNET); 13.8 vs 4.5 months (pNET) ($p <.0001$ for both cohorts)
Pancreatic NET		
Everolimus vs placebo (RADIANT 3)	410	11 vs 4.6 months ($p <.001$)
Sunitinib vs placebo	171	11.4 vs 5.5 months ($p <.001$)
Surufatinib vs placebo	264	10.9 vs 3.7 months ($p = .001$)
Temozolomide/ capecitabine vs temozolomide	144	22.7 vs 14.4 months ($p = .021$)
Extrapancreatic NET		
Octreotide vs placebo (PROMID)	85	14.3 vs 6 months[a]
Everolimus + octreotide vs octreotide (RADIANT 2)	429	16.4 vs 11.3 months
Everolimus vs placebo (RADIANT 4)	302	11 vs 3.9 months
Surufatinib vs placebo	198	9.2 vs 3.8 months ($p <.0001$)
Pazopanib vs placebo	171	11.6 vs 8.5 months ($p <.0005$)
177-Lutetium dotatate vs octreotide (NETTER 1)	230	65.2 vs 10.8% at 20 months ($p <.001$)

[a] Time to tumor progression.

SOMATOSTATIN ANALOGUES

Somatostatin analogues were originally developed as a treatment to reduce hormone secretion in NETs but are also effective in slowing tumor growth. The biologic mechanisms underlying this effect remain uncertain, but clinical studies have been definitive. The first of these studies, the PROMID study, randomized patients with metastatic small-intestinal NET to receive either octreotide LAR at a dose of 30 mg monthly or placebo. The median time to tumor progression in patients receiving octreotide was 14 months compared to only 6 months for patients receiving placebo. Because the study was limited to patients with small-intestinal NET, the generalizability of these results to patients with NETs of other origins, including pancreatic NET, was initially uncertain. This question was ultimately addressed by the phase 3 CLARINET trial, which compared lanreotide, a somatostatin analogue that is similar to octreotide in its somatostatin receptor–binding affinities, to placebo in 204 patients with a range of advanced well- or moderately differentiated gastroenteropancreatic NETs. Progression-free survival duration at 2 years was 65% in patients receiving lanreotide and 33% in patients receiving placebo, a difference that was statistically significant. One unusual aspect of the PROMID and CLARINET studies is the difference in progression-free survival durations in the placebo arms of the studies, which has been attributed to differences in patient selection. Either octreotide or lanreotide is currently considered an acceptable option for control of tumor growth in patients with advanced NETs.

The timing of initiation of somatostatin analogues in patients with advanced NETs remains uncertain. The variable clinical course of NETs means that tumors can remain indolent for years even without treatment. For patients with asymptomatic, small-volume disease, observation alone may be an appropriate initial option. However, for patients with a larger disease burden, evidence of disease progression, or symptomatic disease, somatostatin analogues are generally used as an initial systemic treatment due to their ease of use and tolerability.

PEPTIDE RECEPTOR RADIOLIGAND THERAPY

Peptide receptor radioligand therapy employs the systemic administration of radiolabeled somatostatin analogues and is a treatment option for patients who require more aggressive treatment due to progression on traditional somatostatin analogues or other therapies (Fig. 89-7). Peptide receptor radioligand therapy may also be considered as an initial treatment in patients with significant symptoms or tumor burden. With this approach, a radioligand is coupled to a somatostatin analogue, using the somatostatin analogue to target the tumor. When bound to the tumor cell, the radioligand is then internalized, resulting in cell death. Due to its mechanism of action, peptide receptor radioligand therapy is only considered in patients whose tumors demonstrate uptake on somatostatin scintigraphy.

Several different radioligands have been evaluated, the most successful of which have been yttrium (^{90}Y) and lutetium (^{177}Lu). These two ligands differ from one another in terms of their particle energy and tissue penetration; of the two, ^{90}Y-DOTA-TOC emits higher-energy β particles and has deeper tissue penetration. ^{90}Y-DOTA-TOC has been evaluated in numerous series with overall tumor responses reported in approximately one-third of patients. Enthusiasm for this approach, however, has been tempered due to concerns about side effects including both renal and hematologic toxicity.

^{177}Lu-DOTA-octreotate emits both β particles and lower-energy γ particles and, in most studies, has been associated with less toxicity than ^{90}Y-DOTA-TOC. Initial single-center studies with ^{177}Lu-DOTA-octreotate showed promising antitumor activity, and based on these studies, a randomized trial of ^{177}Lu-dotatate in midgut GI NETs was undertaken. In this study (NETTER-1), 229 patients with inoperable, somatostatin receptor–positive midgut NETs were randomly assigned to receive either four doses of ^{177}Lu-dotatate administered intravenously every 8 weeks or treatment with high-dose octreotide LAR (60 mg) every 4 weeks. Treatment with ^{177}Lu-dotatate was associated with objective tumor responses in 18% of patients and also was associated with a significant improvement in progression-free survival: progression-free survival at month 20 was 10.8% for octreotide LAR alone and 65.2% in the ^{177}Lu-dotatate group. Subsequent analyses have also suggested improved overall survival associated with ^{177}Lu-dotatate treatment, as well as improvements in quality of life across a number of parameters, including global health status, overall physical functioning, fatigue, pain, and diarrhea.

One limitation of the NETTER-1 study was its restriction to patients with advanced small intestine NETs with low-grade histology. A subsequent study, NETTER-2, randomized 226 patients with a broader range of gastroenteropancreatic neuroendocrine tumors that were higher grade (grade 2 or 3) to receive treatment with ^{177}Lu-dotatate or octreotide alone. This study confirmed the activity of ^{177}Lu-dotatate in this patient population; median progression-free survival was 22.8 months for patients receiving ^{177}Lu-dotatate and 8.5 months for patients receiving octreotide alone.

The renal clearance of radiopeptides, including ^{177}Lu-DOTA-octreotate, poses a risk of renal toxicity; this risk can be mitigated with the coadministration of intravenous amino acids during treatment. Longer-term safety data are available from large institutional series that include >1000 patients. Reported toxicities from these series have included rare cases of acute leukemia and myelodysplastic syndrome, presumably associated with radiation exposure. Nevertheless, these studies generally support both the efficacy and safety of ^{177}Lu-dotatate as a treatment for patients with a range of somatostatin receptor–positive NETs.

ALKYLATING AGENTS

While the efficacy of traditional cytotoxic chemotherapy appears to be minimal in most extrapancreatic GI NETs, alkylating agents have a clear role in the treatment of advanced pancreatic NETs. Streptozocin-based combination therapy was historically used as treatment standard in such patients but has largely fallen out of favor due to both toxicity concerns and a cumbersome administration schedule. Temozolomide is an orally administered alkylating agent that has largely replaced streptozocin as a backbone in combination regimens used for the treatment of pancreatic NETs.

Initial studies evaluating temozolomide in combination with a range of different agents showed that temozolomide-based combination therapy was associated with tumor responses in 24–70% of patients. One of the most active combination regimens appeared to be temozolomide and capecitabine. This combination was subsequently compared to temozolomide alone in a prospective randomized study undertaken by the Eastern Cooperative Oncology Group (ECOG) that enrolled 144 patients with advanced pancreatic NETs. The overall response rates in the two arms were relatively similar; 33% of patients who received the combination of temozolomide and capecitabine experienced objective tumor responses as compared to 28% of the patients who received temozolomide as a single agent. However, progression-free survival was significantly longer in the combination arm (22.7 vs 14.4 months). Based on these results, the combination of temozolomide and capecitabine is now the preferred chemotherapy combination for advanced pancreatic NETs.

The reason that some pancreatic NETs respond to alkylating agents while others do not is uncertain. In patients with glioblastoma, methylation of the promoter region for methylguanine DNA methyltransferase (MGMT) is associated with decreased MGMT protein expression and is highly associated with temozolomide responsiveness. MGMT is an enzyme that is responsible for repairing DNA damage induced by alkylating agents. Reduced levels of MGMT presumably impair the ability of tumor cells to repair their DNA in response to treatment and enhance the cytotoxicity of temozolomide. Several retrospective studies, as well as data from the prospective ECOG study, have suggested that lack of MGMT expression in pancreatic NET may be associated with responsiveness to temozolomide-based therapy, although findings have not been definitive.

SMALL-MOLECULE TYROSINE KINASE INHIBITORS

The highly vascular nature of NETs combined with observations in preclinical models that disruption of signaling pathways associated with VEGF inhibits neuroendocrine cell growth prompted a number of clinical trials evaluating therapeutic agents that inhibit the VEGF pathway in both pancreatic and extrapancreatic NETs. The VEGF pathway is activated through the binding of VEGF to its cell surface receptor, which initiates an intracellular signaling cascade that promotes angiogenesis as well as cell growth, proliferation, and survival. Clinical trials of VEGF pathway inhibitors in NETs have included a number of small-molecule tyrosine kinase inhibitors that, while they differ to some extent in specificity, all have in common the property targeting VEGFR2, the receptor isoform most strongly implicated in promoting angiogenesis.

Sunitinib, a multitargeted tyrosine kinase inhibitor that inhibits a range of growth factor receptors including VEGFR2, was one of the first agents in this class found to have activity in pancreatic NETs. In an initial phase 2 trial, sunitinib was administered to 109 patients with either pancreatic or extrapancreatic NET. Of 61 patients with pancreatic NET enrolled in the study, 11 had evidence of an objective tumor response. Based on these observations, sunitinib was evaluated in an international, randomized trial in which continuous administration of sunitinib (37.5 mg daily) was compared with placebo in 171 patients with advanced, progressive pancreatic NET. The median progression-free survival was significantly longer in patients treated with sunitinib compared with patients treated with placebo (11.4 vs 5.5 months). Common side effects associated with sunitinib included hypertension, proteinuria, and fatigue.

A second VEGFR-targeted tyrosine kinase inhibitor, surufatinib, was evaluated in a randomized trial in which 264 patients with advanced pancreatic NETs from 21 centers in China were randomized to receive either surufatinib, administered at a dose of 300 mg daily, or placebo. Patients receiving surufatinib experienced a median progression-free survival duration of 10.9 months, as compared to 3.7 months in patients receiving placebo, closely mirroring the results of the earlier sunitinib study. Cabozantinib, a tyrosine kinase inhibitor with activity against not only VEGFR but also c-MET and related growth factor receptors, was evaluated against placebo in a randomized trial led by the ALLIANCE cooperative group that included both pancreatic and extrapancreatic NETs. Among the 95 pancreatic NET patients in the study, median progression-free survival was 13.8 months for those receiving cabozantinib and 4.5 months for those receiving placebo. Other small-molecule tyrosine kinase inhibitors that have been evaluated in smaller, single-arm studies and have shown activity in pancreatic NETs, include sorafenib, pazopanib, and axitinib.

Small-molecule tyrosine kinase inhibitors targeting the VEGF pathway have also been evaluated in patients with advanced nonpancreatic GI NET. In most of these studies, objective tumor response rates are lower than those seen in pancreatic NET, though many of these initial studies also revealed low rates of tumor progression and encouraging progression-free survival durations, suggesting that these agents had antitumor activity. Pazopanib was compared to placebo in a randomized study undertaken by the ALLIANCE cooperative group, which enrolled 171 patients with nonpancreatic NETs. Patients treated with pazopanib in this study had a superior progression-free survival compared to those who received placebo (11.6 vs 8.5 months), a difference that was statistically significant. Surufatinib was used in a randomized study of 198 patients with extrapancreatic NETs; the median progression-free survival was 9.2 months in patients receiving surufatinib and 3.8 months in those receiving placebo. The strongest results to date have come from a randomized study of cabozantinib (see above) that included 205 patients with extrapancreatic GI NET. In this study, median progression-free survival was 8.5 months in the cabozantinib arm as compared with 4 months in the placebo arm. Taken together with the results in the pancreatic NET cohort (above), these results provide a strong rationale for considering cabozantinib as a treatment option for patients with both advanced pancreatic and nonpancreatic GI NET.

mTOR INHIBITORS

mTOR is an intracellular protein kinase that has been implicated in the regulation of a number of processes regulating cell growth in both normal and malignant cells. It functions as a downstream component of the PI3-AKT-mTOR pathway. This pathway is negatively regulated by the tuberous sclerosis complex, comprising TSC1 and TSC2. An association between the development of pancreatic NETs and inherited mutations in *TSC2* in patients with tuberous sclerosis complex was a contributing factor to initial interest in exploring mTOR inhibition as a therapeutic approach in this setting.

Following initial evidence of antitumor activity associated with everolimus (10 mg daily) in an international, multicenter, phase 2 trial of 160 patients, everolimus monotherapy (10 mg daily) was compared with best supportive care alone in the RADIANT-3 trial that enrolled 410 patients with advanced progressing pancreatic NET. While overall objective responses were uncommon, treatment with everolimus was associated with a significant prolongation in median progression-free survival (11.0 vs 4.6 months) compared to placebo, supporting its use as a standard treatment to control tumor growth in patients with advanced pancreatic NET. Common toxicities associated with everolimus are generally mild and can include stomatitis and rash; a more severe but less common side effect is pneumonitis.

Everolimus was also associated with promising activity in early phase 2 studies enrolling patients with extrapancreatic NET. The first large, randomized study evaluating everolimus was the RADIANT-2 trial; 429 patients with advanced GI NETs were randomly assigned to receive octreotide LAR (30 mg intramuscularly every 28 days) with or without everolimus (10 mg daily). Treatment with everolimus in this study was associated with an improvement in median progression-free survival (16.4 vs 11.3 months), but the difference in this study was of only borderline statistical significance. A second study, the RADIANT-4 study, enrolled 302 patients with advanced NETs of either GI (excluding pancreatic) or lung origin, randomizing them to receive either everolimus or placebo. In this study, treatment with octreotide was not required. As in the RADIANT-3 study, objective tumor responses were uncommon; however, median progression-free survival in patients who received everolimus was significantly longer than in those who received placebo (11 vs 3.9 months). Based on the results of this study, everolimus is considered a standard treatment for control of tumor growth in extrapancreatic NETs.

OTHER SYSTEMIC TREATMENTS FOR CONTROL OF TUMOR GROWTH

Interferon α has been used as a treatment for advanced NETs for several decades. With the development of newer approaches, its routine use has diminished. The use of interferon α was based primarily on observations in large, retrospective series where low-dose interferon α was reported to both reduce symptoms of hormonal hypersecretion and slow tumor progression. Interferon can be myelosuppressive, requiring dose titration, and in some patients can induce both fatigue and depression. Antitumor activity has also been reported with oxaliplatin-based chemotherapy regimens. A combined analysis of two phase 2 trials examining oxaliplatin-fluoropyrimidine chemotherapy plus bevacizumab in advanced NET suggested antitumor activity for these regimens; the benefit appeared to be greatest in patients with intermediate-grade rather than low-grade tumors.

Treatment with immune checkpoint inhibitors has been found to be effective across multiple cancer types. The role of immune checkpoint inhibitors for the treatment of neuroendocrine tumors has not yet been clearly established and appears to depend at least in part on tumor grade. Immunohistochemical and transcriptomic profiling in well-differentiated neuroendocrine tumors has revealed only low levels of PD-1 and PD-L1 expression together with high levels of immunosuppressive gene expression in tumor-associated myeloid cells. The KEYNOTE-158 study investigated the efficacy of pembrolizumab, a monoclonal antibody targeting the checkpoint marker PD-1, in multiple cancers, including 107 patients with neuroendocrine tumors of various sites. The overall tumor response rate in the neuroendocrine tumor cohort was only 3.7%. Other immunotherapeutic approaches, including use of chimeric antigen receptor T cells for the treatment of neuroendocrine tumors, remain investigational.

SYSTEMIC THERAPY FOR HIGH-GRADE NEUROENDOCRINE CARCINOMA

High-grade NETs are relatively uncommon and tend to pursue an aggressive clinical course. In contrast to well-differentiated neuroendocrine tumors, high-grade neuroendocrine carcinomas may, at least in some cases, be quite responsive to immunotherapeutic approaches. A basket trial evaluating a combination of the anti-CTLA-4 monoclonal antibody ipilimumab together with the anti-PD-1 monoclonal antibody nivolumab enrolled 32 patients with poorly differentiated nonpancreatic neuroendocrine tumors; the overall tumor response rate was 44% in this cohort. Apart from these more recent immune-based approaches, chemotherapy for advanced high-grade neuroendocrine carcinoma has historically followed a paradigm analogous to that used for small-cell carcinoma of the lung, with combinations of either cisplatin or carboplatin administered together with etoposide generally considered the preferred first-line approach. One of the most important elements in determining the optimal chemotherapeutic approach is assessing the Ki-67 proliferative index. A large retrospective series that evaluated 252 patients with high-grade neuroendocrine carcinoma found that the activity of platinum-based therapy was greatest in patients who had a Ki-67 proliferative index of 55% or higher; in these patients, the overall tumor response rate was 42%. In contrast, the overall response rate in patients in whom the Ki-67 proliferative index was <55% was only 15%.

Acknowledgment

Dr. Robert Jensen contributed this chapter in previous editions and some material from his chapter is retained here.

FURTHER READING

Caplin ME et al: Lanreotide in metastatic enteropancreatic neuroendocrine tumors. N Engl J Med 371:224, 2014.

Dasari A et al: Trends in the incidence, prevalence, and survival outcomes for patients with neuroendocrine tumors in the United States. JAMA Oncol 3:1336, 2017.

Kulke MH et al: Telotristat ethyl, a tryptophan hydroxylase inhibitor for the treatment of carcinoid syndrome. J Clin Oncol 35:14, 2017.

Kunz PL et al: Randomized study of temozolomide or temozolomide and capecitabine in patients with advanced pancreatic neuroendocrine tumors (ECOG-ACRIN E2211). J Clin Oncol 41:1359, 2023.

Patel SP et al: A phase II basket trial of dual anti-CTLA-4 and anti-PD-1 blockade in rare tumors (DART SWOG 1609) in patients with nonpancreatic neuroendocrine tumors. Clin Cancer Res 26:2290, 2020.

Raymond E et al: Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. N Engl J Med 364:501, 2011.

Rindi G et al: A common classification framework for neuroendocrine neoplasms: An International Agency for Research on Cancer (IARC) and World Health Organization (WHO) expert consensus proposal. Mod Pathol 31:1770, 2018.

Scarpa A et al: Whole genome landscape of pancreatic neuroendocrine tumors. Nature 543:65, 2017.

Strosberg J et al: Phase 3 trial of 177 Lu-dotatate for midgut neuroendocrine tumors. N Engl J Med 376:125, 2017.

Xu Z et al: Epidemiologic trends of and factors associated with overall survival for patients with gastroenteropancreatic neuroendocrine tumors in the United States. JAMA Network Open 4:e2124750, 2021.

Yao JC et al: Everolimus for advanced pancreatic neuroendocrine tumors. N Engl J Med 364:514, 2011.

90

Renal Cell Carcinoma

Robert J. Motzer, Martin H. Voss

Renal cell carcinomas account for 90–95% of malignant neoplasms arising from the kidney. Notable features include frequent diagnosis without symptoms, resistance to cytotoxic agents, robust activity of angiogenesis-targeted agents, immune infiltration commonly rendering tumors susceptible to checkpoint-directed immunotherapy, and a variable clinical course for patients with metastatic disease, including anecdotal reports of spontaneous regression. Most of the remaining 5–10% of malignant neoplasms arising from the kidney are transitional cell carcinomas (urothelial carcinomas) originating in the lining of the renal pelvis. See **Chap. 91** for transitional cell carcinomas.

EPIDEMIOLOGY

The incidence of cancers of the kidney and renal pelvis rose for three decades, reached a plateau of approximately 64,000 cases annually in the United States between 2012 and 2018, but has since increased to approximately 81,000 cases annually, resulting in close to 15,000 deaths per year. It is the seventh most common cancer overall in the United States, the sixth most common in males, and the ninth most common in females; the male-to-female ratio is 2:1. Although this malignancy may be diagnosed at any age, it is uncommon in those under 45 years, and incidence peaks between the ages of 55 and 75 years. Many factors have been investigated as possible contributing causes; associations include cigarette smoking, obesity, and hypertension. Risk is also increased for patients with polycystic kidney disease that has been complicated by chronic renal failure.

Most cases of renal cell carcinoma (RCC) are sporadic, although familial forms have been reported **(Table 90-1)**. One well-established example includes clear cell RCC arising in the context of von Hippel-Lindau (VHL) syndrome, an autosomal dominant disorder. Genetic studies identified the *VHL* gene on the short arm of chromosome 3. Individuals with VHL syndrome have a 70% estimated lifetime risk of developing clear cell RCC. Other *VHL*-associated neoplasms include retinal hemangioma, hemangioblastoma of the spinal cord and cerebellum, pheochromocytoma, and neuroendocrine tumors. Belzutifan, an oral inhibitor of hypoxia-inducible factor (HIF-2α), is approved for treatment of *VHL*-associated cancers. Birt-Hogg-Dubé syndrome is a rare human autosomal dominant genetic disorder characterized by fibrofolliculomas (benign tumors arising in hair follicles), pulmonary cysts, and RCCs of varying histologies, most commonly the chromophobe type, occurring in about a third of patients. This disorder is associated with mutations in the *FLCN* gene, which codes for folliculin. Other hereditary syndromes are summarized in Table 90-1.

PATHOLOGY AND GENETICS

Renal cell malignancies represent a heterogeneous group of tumors with distinct histopathologic, genetic, and clinical features **(Table 90-2)**. Categories include clear cell carcinoma (70% of cases), papillary RCC (10–15%), chromophobe RCC (≤5%), molecularly defined entities such as *TFE3*-rearranged RCC (<5%), and other less common variants. Papillary tumors can be bilateral and multifocal. Chromophobe tumors tend to have a more indolent clinical course. *TFE3*-rearranged RCC, rare in adult patients, is the predominant histology in children. *SMARCB1*-deficient RCC, previously called renal medullary

TABLE 90-1 Hereditary Renal Cell Tumors

SYNDROME	CHROMOSOME(S)	GENE	PROTEIN	KIDNEY TUMOR TYPE	ADDITIONAL CLINICAL FINDINGS
von Hippel-Lindau syndrome	3p25	*VHL*	von Hippel-Lindau protein	Clear cell	Hemangioblastoma of the retina and central nervous system; pheochromocytoma; pancreatic and renal cysts; neuroendocrine tumors
Hereditary papillary RCC	7p31	*MET*	MET	Papillary	Bilateral and multifocal kidney tumors
Hereditary leiomyomatosis and RCC (HLRCC syndrome)	1q42	*FH*	Fumarate hydratase	FH-deficient/HLRCC syndrome–associated RCC	Leiomyoma; uterine leiomyoma/leiomyosarcoma
Birt-Hogg-Dubé syndrome	17p11	*FLCN*	Folliculin	Chromophobe; clear cell; oncocytoma	Facial fibrofolliculoma; pulmonary cysts
Tuberous sclerosis	9q34 16p13	*TSC1* *TSC2*	Hamartin Tuberin	Angiomyolipomas; TSC-associated lymphangioleiomyomatosis; rare RCC with variety of histologic appearances including eosinophilic solid and cystic (ESC) RCC	Angiofibroma, subungual fibroma; cardiac rhabdomyoma; adenomatous small intestine polyps; pulmonary and renal cysts; cortical tuber; subependymal giant cell astrocytomas
BAP1 tumor predisposition syndrome	3p21	*BAP1*	BAP1	Mostly clear cell, but chromophobe and papillary have also been reported	Atypical Spitz tumors; uveal melanoma; cutaneous melanoma; basal cell carcinoma; malignant mesothelioma

Abbreviations: ESC, eosinophilic solid and cystic; FH, fumarate hydratase; HLRCC; hereditary leiomyomatosis and RCC; RCC, renal cell carcinoma; TSC, tuberous sclerosis.

TABLE 90-2 Classification of Malignant Epithelial Neoplasms Arising from the Kidney

CARCINOMA TYPE	CHARACTERISTIC GROWTH PATTERN	CHROMOSOMAL EVENTS	GENES WITH RECURRENT SOMATIC ALTERATIONS
Clear cell	Varying growth patterns, including acinar, solid and sarcomatoid	3p–, 5q+, 14q–, 9p–	*VHL, PBRM1, BAP1, SETD2*
Papillary	Papillary or sarcomatoid	+7, +17, 9p–	*MET, CDKN2A* (focal deletions)
Chromophobe	Solid, tubular, or sarcomatoid	Whole arm losses (1, 2, 6, 10, 13, 17, and 21)	*TP53, PTEN, TERT* promotor
TFE3-rarranged and *TFEB*-altered renal cell carcinomas	Mimicking clear cell and papillary variants	Xp11.2 translocations; t(6;11) translocations	*TFE3* gene fusions, *TFEB* gene fusions
SMARCB1-deficient renal medullary carcinoma	Varying growth patterns, including cribriform, reticular, sarcomatoid, adenoid, and microcystic	+8q, 22q–, 22q translocations	*SMARCB1* (focal deletions, mutations, gene fusions), *SETD2*

carcinoma, is rare, very aggressive, and associated with sickle cell trait. Tumors that do not meet criteria for defined variants are generally referred to as "unclassified" with variable clinical courses.

Clear cell tumors, the predominant histology, are found in >80% of patients who develop metastases and arise from the epithelial cells of the proximal tubules. Loss of chromosome 3p is uniformly seen as the earliest event in the development of these cancers. This leads to loss of heterozygosity for a number of relevant 3p genes, including *VHL*, *PBRM1*, *BAP1*, and *SETD2*, which can be functionally silenced through secondary events in the remaining allele. *VHL* encodes a tumor-suppressor protein that is involved in regulating the transcription of vascular endothelial growth factor (VEGF) and a number of other effectors through ubiquitination of hypoxia-inducible factors (HIF). Inactivation of *VHL*, through upregulation of VEGF signaling, promotes tumor angiogenesis and growth, ultimately rendering clear cell RCC cells susceptible to antiangiogenesis therapy.

Large-scale sequencing efforts have helped elucidate recurrent patterns of genomic evolution that correlate with distinct clinical phenotypes, e.g., varying levels of aggressiveness or specific patterns of metastatic spread. For example, early loss of chromosome 9p appears to confer a high risk for early metastatic dissemination and correlates with poor cancer-specific survival.

A growing number of other RCC variants are well defined (see Table 90-2 for examples) yet can further vary by molecular features. For instance, up to 15% of RCCs are of the papillary subtype, and variant features can be distinguished by light microscopy but also molecular assays. Activating mutations in the *MET* oncogene or gain of chromosome 7 (where *MET* is located) are hallmark events of certain papillary variants and considered actionable via targeted MET inhibitors. Tumors of the less common chromophobe subtype originate from the distal nephron. They are typically characterized by aneuploidy with common loss of an entire chromosome copy for chromosomes 1, 2, 6, 10, 13, and 17.

CLINICAL PRESENTATION

Presenting signs and symptoms may include hematuria, flank or abdominal pain, and a palpable mass. Other symptoms are fever, weight loss, anemia, and a varicocele. Tumors are, however, commonly detected as an incidental finding on a radiograph. Widespread use of radiologic cross-sectional imaging (computed tomography [CT], magnetic resonance imaging [MRI]) contributes to earlier detection of renal masses during evaluation for other medical conditions. The increasing number of incidentally discovered low-stage tumors has contributed to an improved 5-year survival for patients with RCC and increased use of nephron-sparing surgery (partial nephrectomy). A spectrum of paraneoplastic syndromes has been associated with these malignancies, including erythrocytosis, hypercalcemia, nonmetastatic hepatic dysfunction (Stauffer's syndrome), and acquired dysfibrinogenemia. Erythrocytosis is noted at presentation in only about 3% of patients. Anemia, commonly a sign of more advanced disease, is more common. Kidney cancer was called the "internist's tumor" since it was often discovered from the initial presentation of a paraneoplastic syndrome. This was more common before the era of modern imaging, as was initial presentation by the classic triad of hematuria, flank pain, and a palpable abdominal mass.

The standard evaluation of patients with suspected renal tumors includes a CT scan of the abdomen and pelvis, chest radiograph, and urine analysis. If metastatic disease is suspected from the chest radiograph, a CT of the chest is warranted. MRI is useful in evaluating the inferior vena cava in cases of suspected tumor involvement or invasion by thrombus, or when intravenous contrast administration given with CT is prohibited by impaired renal function. In clinical practice, any solid renal masses should be considered malignant until proven otherwise; a definitive diagnosis is required. If no metastases are demonstrated, surgery is indicated, even if the renal vein or inferior vena cava is invaded. In small tumors (particularly those of clear cell variant), the risk of impending metastatic spread is lower and surgery can potentially be delayed. In that setting, a needle biopsy should be performed to confirm the underlying histology, and radiographic surveillance is indicated until the time of surgery. The differential diagnosis of a renal mass includes cysts, benign neoplasms (adenoma, angiomyolipoma, oncocytoma), inflammatory lesions (pyelonephritis or abscesses), and other malignancies originating in the kidney such as transitional cell carcinoma of the renal pelvis, sarcoma, lymphoma, and Wilms' tumor or metastases from cancers originating in other organs. All of these are less common causes of renal masses than is RCC. The most common sites of distant metastases are the lungs, lymph nodes, liver, bone, and brain. These tumors may follow an unpredictable and protracted clinical course.

STAGING AND PROGNOSIS

Staging is based on the American Joint Committee on Cancer (AJCC) staging system **(Fig. 90-1)**. Stage I tumors are ≤7 cm in greatest diameter and confined to the kidney; stage II tumors are >7 cm and confined to the kidney; stage III tumors extend through the renal capsule but are confined to Gerota's fascia, grossly infiltrate the renal vein, or involve regional lymph nodes (N1); and stage IV disease includes tumors that have invaded adjacent organs or involve nonregional lymph nodes or distant metastases. Sixty-five percent of patients present with stage I or II disease, 15–20% with stage III, and 15–20% with stage IV. The 5-year survival rate is currently 77% across all RCCs but varies greatly by stage.

Prognostic risk models are helpful for counseling patients diagnosed with metastatic disease and for anticipating survival rates when designing a clinical trial. A widely used prognostic model for advanced disease, the International Metastatic RCC Database Consortium (IMDC) risk model, incorporates six factors shown to correlate with worse survival: poor performance status, low hemoglobin concentration, high serum calcium, high neutrophil levels, high platelet levels, and <1-year interval from diagnosis to systemic treatment. Patients with zero risk factors achieve significantly longer median survival (≥5 years) than patients with one or two risk factors (~4 years) and those with three to six risk factors (~3 years) when treated with first-line checkpoint inhibitor–containing combination regimens (see below).

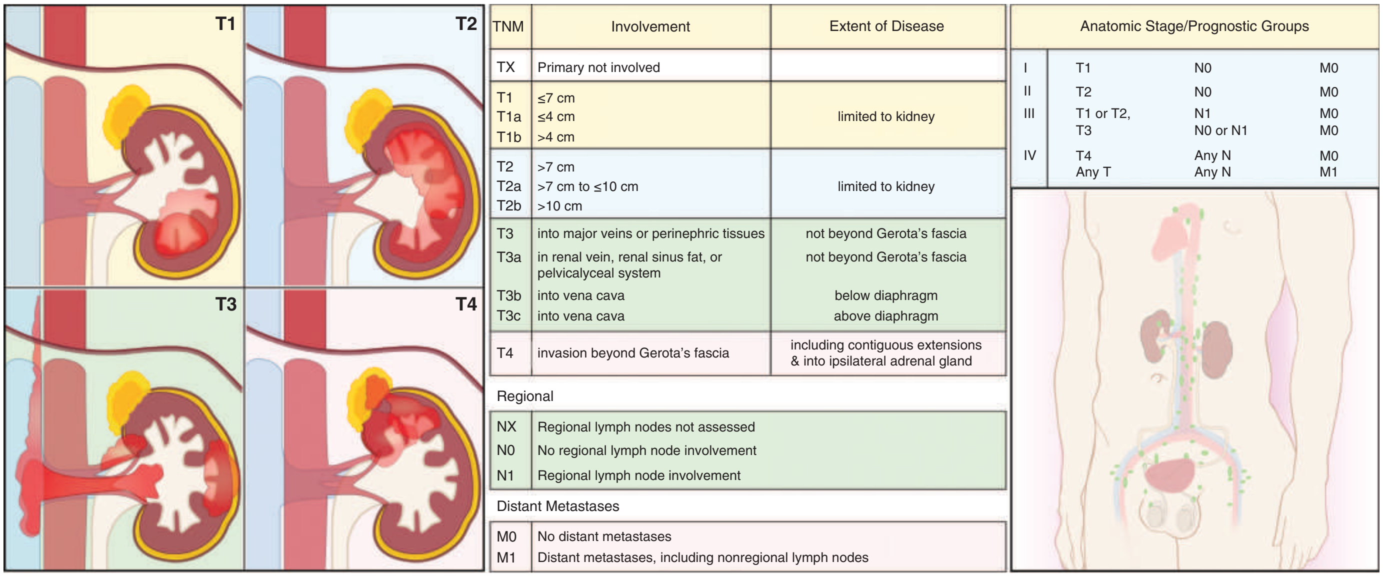

TNM	Involvement	Extent of Disease
TX	Primary not involved	
T1 T1a T1b	≤7 cm ≤4 cm >4 cm	limited to kidney
T2 T2a T2b	>7 cm >7 cm to ≤10 cm >10 cm	limited to kidney
T3	into major veins or perinephric tissues	not beyond Gerota's fascia
T3a	in renal vein, renal sinus fat, or pelvicalyceal system	not beyond Gerota's fascia
T3b	into vena cava	below diaphragm
T3c	into vena cava	above diaphragm
T4	invasion beyond Gerota's fascia	including contiguous extensions & into ipsilateral adrenal gland
Regional		
NX	Regional lymph nodes not assessed	
N0	No regional lymph node involvement	
N1	Regional lymph node involvement	
Distant Metastases		
M0	No distant metastases	
M1	Distant metastases, including nonregional lymph nodes	

Anatomic Stage/Prognostic Groups			
I	T1	N0	M0
II	T2	N0	M0
III	T1 or T2, T3	N1 N0 or N1	M0 M0
IV	T4 Any T	Any N Any N	M0 M1

FIGURE 90-1 Renal cell carcinoma staging. TNM, tumor-node-metastasis.

TREATMENT

Renal Cell Carcinoma

LOCALIZED TUMOR

The standard management for stage I or II tumors and selected cases of stage III disease is radical or partial nephrectomy. A radical nephrectomy involves en bloc removal of Gerota's fascia and its contents, including the kidney, and commonly the ipsilateral adrenal gland and regional lymph nodes that appear abnormal on imaging or intraoperatively. Open, laparoscopic, or robotic surgical techniques may be used. The role of a template lymphadenectomy in patients without apparent lymphadenopathy is controversial. Extension into the renal vein or inferior vena cava (stage III disease) does not preclude resection, which would then include thrombectomy.

Nephron-sparing approaches, i.e., open or laparoscopic partial nephrectomy, may be appropriate depending on the size and location of the tumor. This approach is particularly relevant for patients with solitary kidneys, bilateral tumors, or chronic renal insufficiency but can also be applied electively to resect small masses for patients with normal kidney function. Radical nephrectomy carries a greater risk for chronic kidney disease and cardiovascular morbidity and mortality.

Adjuvant systemic therapies, including cytokines and targeted agents, have been studied in randomized clinical trials, largely with negative results. The checkpoint inhibitor pembrolizumab is approved in patients with increased risk of recurrence following nephrectomy, where 12 months of therapy was shown to improve disease-free survival compared to placebo. For those with low risk of recurrence, the standard of care remains active surveillance after nephrectomy.

METASTATIC DISEASE

Surgery has a limited role for patients with metastatic disease. Long-term survival may occur in patients who relapse with a solitary site that is removed (metastasectomy). Nephrectomy despite presence of metastases (cytoreductive nephrectomy) is considered for carefully selected patients with stage IV disease. One indication for this approach can be to alleviate pain or hemorrhage of a primary tumor.

Radiation therapy is used for palliation of bone or brain metastases. The type of radiotherapy most commonly used is external-beam therapy, including stereotactic radiosurgery and other forms of image-guided radiotherapy.

Systemic therapy is the mainstay of care for metastatic disease. The timing of initiating such treatment should be carefully considered; some patients are asymptomatic at diagnosis, and with indolent behavior, it may be best to document progression before initiating treatment.

Metastatic RCC is refractory to cytotoxic chemotherapy. Patients are treated with molecularly targeted agents, including targeted immunotherapies, i.e., checkpoint inhibitors. Treatments are continued with noncurative intent while tolerated and until disease progression is evident on cross-sectional imaging. Outcomes for patients with metastatic disease improved when increased understanding of underlying biology led to the successful development of several tyrosine kinase inhibitors (TKIs) targeting proangiogenic signaling through the VEGF receptors as well as allosteric inhibitors of mammalian target of rapamycin (mTOR) signaling. Serial large-scale randomized trials demonstrated that such agents, typically orally available, could be administered sequentially and in combination. Pivotal studies, by design, defined a dedicated space for each regimen in treatment-naïve or pretreated patients **(Table 90-3)**.

Targeted immunotherapies were introduced after VEGF- and mTOR-directed agents had established standards of care in the first- and second-line setting. Nivolumab, a checkpoint inhibitor targeting PD-1, demonstrated superior overall survival compared to the mTOR inhibitor everolimus in a randomized trial in patients who had progressed on prior TKI therapy, challenging the standard approach in pretreated patients and positioning nivolumab as the new second-line agent of choice. Subsequently, immunotherapy combination regimens demonstrated efficacy in randomized trials conducted in treatment-naïve patients. In separate studies, four doublets demonstrated survival benefit over standard sunitinib therapy and changed the standard of care for untreated metastatic clear cell RCC toward two-drug, immunotherapy-containing regimens: nivolumab in combination with the CTLA-4–directed checkpoint inhibitor ipilimumab; the TKI axitinib together with the PD-1 inhibitor pembrolizumab; the TKI cabozantinib plus nivolumab; and the TKI lenvatinib paired with pembrolizumab. These combinations demonstrated objective radiographic responses in 40–70% of patients, and complete radiographic disappearance of cancer is achieved in about 10% of patients. The majority of such anticancer effects were reported to be long-lasting.

With an ever-growing number of approved options directed toward different molecular targets, biomarkers are urgently needed to help individualize therapeutic choices and to gain insight as to

TABLE 90-3 Commonly Used Systemic Regimens for Metastatic Renal Cell Carcinoma

CLASS	DRUG	FIRST FDA APPROVAL FOR RCC	CURRENTLY USED FOR
Antiangiogenic: TKIs	Sunitinib	2006	Advanced RCC, first line
	Pazopanib	2009	Advanced RCC, first line
	Axitinib	2012	Advanced RCC, pretreated
	Cabozantinib	2016	Advanced RCC, pretreated with antiangiogenic therapy
		2017	Advanced RCC, first line
	Tivozanib	2021	Advanced RCC, pretreated with two or more prior systemic therapies
Immunotherapy: checkpoint inhibitor	Nivolumab	2015	Advanced RCC, pretreated with antiangiogenic therapy
Combination therapies			
TKI + mTOR inhibitor	Lenvatinib + everolimus	2016	Advanced RCC, pretreated with one antiangiogenic therapy
PD-1 inhibitor + CTLA-4 inhibitor	Nivolumab + ipilimumab	2018	Advanced intermediate-risk or poor-risk RCC, first line
PD-1 inhibitor + TKI	Pembrolizumab + axitinib	2019	Advanced RCC, first line
	Nivolumab + cabozantinib	2021	Advanced RCC, first line
	Pembrolizumab + lenvatinib	2021	Advanced RCC, first line

Abbreviations: CTLA-4, cytotoxic T lymphocyte-associated protein; FDA, U.S. Food and Drug Administration; mTOR, mammalian target of rapamycin; PD-1, programmed cell death-1; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor.

whether and why treatments are working. Although a multitude of candidate biomarkers have been investigated for their predictive value in metastatic RCC, none have been validated for clinical use to date.

Projected overall survival in patients starting systemic therapies for newly diagnosed metastatic disease has tripled over the past 15–20 years; this can largely be attributed to the successful drug developments discussed here.

GLOBAL CONSIDERATIONS

Worldwide, >400,000 patients are diagnosed each year with malignant tumors arising from the kidney, resulting in >175,000 deaths annually. Kidney cancer is the sixth most common cancer in men and the 10th most common cancer in women. Higher incidence is observed in developed countries, including the United States, Canada, Europe, Australia, New Zealand, and Uruguay. Lower rates are reported in Southeast Asia and Africa, though it is speculated that these rates may be underreported due to lack of disease-reporting structure and reduced access to diagnostic facilities. The incidence of kidney cancer has been steadily increasing over the past four decades. Mortality trends have stabilized in Europe and the United States, but not in less developed countries. This is likely related to differences in access to optimal therapies. Treatment guidelines for both localized and metastatic renal cancer are similar between U.S. and European documents and contingent on the access to adequate health care and availability of targeted drugs to treat metastases.

FURTHER READING

Choueiri TK et al: Adjuvant pembrolizumab after nephrectomy in renal-cell carcinoma. N Engl J Med 385:683, 2021.

Choueiri TK et al: Nivolumab plus cabozantinib versus sunitinib for advanced renal-cell carcinoma. N Engl J Med 384:829, 2021.

Huang J et al: A global trend analysis of kidney cancer incidence and mortality and their associations with smoking, alcohol consumption, and metabolic syndrome. Eur Urol Focus 8:200, 2022.

Jonasch E et al: Belzutifan for renal cell carcinoma in von Hippel–Lindau Disease. N Engl J Med 385:2036, 2021.

Moch H et al: The 2022 WHO classification of tumours of the urinary system and male genital organs—Part A: renal, penile, and testicular tumours. Eur Urol 82:458, 2022.

Motzer RJ et al: Molecular subsets in renal cancer determine outcome to checkpoint and angiogenesis blockade. Cancer Cell 38:803, 2020.

Motzer RJ et al: Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. N Engl J Med 378:1277, 2018.

91

Cancer of the Bladder and Urinary Tract

Noah M. Hahn

GLOBAL CONSIDERATIONS

Within the United States, urothelial carcinomas of the bladder and urinary tract are most closely related to tobacco smoking history. However, within developing countries, water supplies contaminated with arsenic or schistosomiasis parasites also are major carcinogenic contributors.

INTRODUCTION

Cancers of the urinary tract including the bladder, renal pelvis, ureter, and urethra occur frequently, and they represent the second most common class of genitourinary cancers. Bladder cancer alone represents the sixth most common cancer diagnosis annually in the United States with 82,290 new cases and 16,710 deaths every year. Because cancers of the renal pelvis are often lumped in with all kidney cancers, the true incidence and mortality from nonbladder urinary tract cancers are less precise. While less frequent than bladder cancer, an additional 20,000 new cases and 5000 deaths are estimated every year. An accelerated understanding of the molecular underpinnings of bladder and urinary tract cancer biology has led to a significant increase in urothelial cancer clinical trials resulting in U.S. Food and Drug Administration (FDA) approval of multiple new therapeutic agents since 2016 with more expected to follow. This chapter reviews the established, current, and emerging evidence that serves as the basis for the rapidly evolving standards of care for patients with bladder and urinary tract cancers.

CLINICAL EPIDEMIOLOGY AND RISK FACTORS

Bladder cancer typically affects older patients with a median age at diagnosis of 73 years. Males are four times more frequently affected than females. Similarly, bladder cancer is more common in Caucasians than in Asian patients. Inheritable germline genetic risk factors have been observed in up to one-seventh of patients with bladder or urinary tract cancers. However, a predominant, singular germline genetic alteration has not been identified. Patients with defects in mismatch repair genes leading to microsatellite instability (*MLH1, MSH2, MSH6*, etc.) as part of the familial cancer Lynch syndrome are at particular risk of upper urinary tract cancers of the renal pelvis and ureter. Additionally, patients with Cowden disease (*PTEN* mutations) or retinoblastoma (*RB1* mutations) are at increased risk for developing bladder cancer.

Historically, associations have existed between environmental toxic exposures and higher rates of developing bladder cancer. The aromatic amines benzidine and β-naphthylamine that can be present in industrial dyes as well as arsenic that can be found in drinking water supplies in underdeveloped countries have been associated with increased bladder cancer risk. Other chemicals in the leather, paint, rubber, textiles, and printing industries have been associated with bladder cancer. Associations with exposures to hair dyes and hair sprays in workers in the hairstyling field have been suggested. Additionally, concern has been raised regarding use of the antidiabetic medication, pioglitazone, and bladder cancer risk. Extensive reviews and meta-analyses have produced differing conclusions. The data suggest a small risk of bladder cancer from long-term pioglitazone use, which has led to inclusion of bladder cancer risk within the pioglitazone prescribing information. An association between chronic inflammatory states and the development of squamous bladder cancer clearly exists in underdeveloped countries in patients chronically infected with the parasitic disease schistosomiasis and in paraplegic patients with chronic indwelling catheters. Above and beyond each of these associations, however, smoking of tobacco products (cigarettes, cigars, pipes, etc.) remains the overwhelming leading risk factor for development of bladder cancer. Among new bladder cancer diagnoses, 90% of cases occur in current or former smokers. Toxicologists have estimated that >70 confirmed carcinogenic toxins are present within tobacco smoke. It is estimated that one-third of bladder cancer cases could be prevented through simple modification of lifestyle choices, in particular cessation of smoking.

CLINICAL PRESENTATION AND DIAGNOSTIC WORKUP

Occasionally, patients will present with flank pain in association with an upper tract renal pelvis or ureter cancer or due to hydronephrosis in association with a bladder tumor obstructing the orifice of the ureter within the bladder. Only in rare cases do patients present with significant cachexia and widespread metastatic disease. For most patients, painless hematuria (either gross or microscopic) represents the initial manifestation of an underlying urinary tract cancer. In females, hematuria due to malignancy can often be mistaken for a urinary tract infection or menstrual bleeding. While treatment with antibiotics is warranted if a concurrent urinary tract infection is noted on initial urinalysis, persistent hematuria requires further workup. Painless hematuria in males is almost always abnormal and should be worked up. Initial investigations in patients of either sex should include urine cytology and visual examination of the bladder by cystoscopy. Cytology is successful in identifying cancer in only 50% of individuals with high-grade bladder cancers. In addition to urine cytology, radiographic evaluation of the kidneys and upper urinary tract by CT urogram should be performed. A magnetic resonance (MR) urogram may be substituted in patients with poor renal function. Additional diagnostic testing of the urine to assess for cancer-associated chromosomal changes by fluorescent in situ hybridization, increased levels of nuclear mitotic proteins, increased bladder tumor–associated antigens, or higher levels of staining on cells shed by the bladder may identify some cancers missed by traditional cytology testing. However, they may also produce abnormal results in patients who do not have cancer. For now, these adjunct molecular tests are primarily utilized in detecting recurrent cancer in patients with a prior diagnosis of urinary tract cancer. Small tumors, particularly flat noninvasive tumors of the bladder, may be detected at higher rates with the use of blue light cystoscopy or narrow-band imaging cystoscopy. Both blue light and narrow-band imaging cystoscopies are now used routinely in the monitoring of patients with bladder cancer. For patients with no bladder abnormalities in whom upper tract tumors are suspected, visualization of the upper urinary tracts and renal pelvises should be performed by ureteroscopy or retrograde pyelography.

In all patients with abnormalities noted in the bladder or upper urinary tracts, complete endoscopic resection for histologic diagnosis and staging should be performed when possible via either transurethral resection of bladder tumor (TURBT) or endoscopic resection of upper tract tumors. All metastatic patients should have genomic sequencing performed on their tumor tissue, circulating tumor DNA (ctDNA), or both. The role of genomic sequencing in earlier stages of disease is evolving.

HISTOLOGY

Urothelial carcinoma, often called transitional cell carcinoma in the past, is the most common urinary tract cancer histology and is observed in ~90% of cases. Squamous, glandular, micropapillary, plasmacytoid, sarcomatoid, and other variant features can often be found in portions of urothelial carcinoma tumors; however, pure variant histologies are rare. The presence of some variant histologies including micropapillary and plasmacytoid has been associated with worse surgical outcomes compared to urothelial carcinoma. Nonurothelial variant histologies including squamous cell carcinoma, adenocarcinoma, small-cell carcinoma, and carcinosarcoma collectively account for ≤10% of urinary tract tumors. Examples of traditional urothelial carcinoma and some of the variant histologies are shown in **Fig. 91-1**.

MOLECULAR BIOLOGY

Clinically, urothelial carcinoma of the bladder displays a biphasic phenotype characterized by (1) low-grade papillary tumors that frequently recur but rarely invade or metastasize and (2) high-grade sometimes flat tumors that invade early leading to lethal metastatic disease. In both of these phenotypes, loss of portions of chromosomes 9q and 9p by loss of heterozygosity is an early molecular event, whose exact significance is not clear. Potential candidate regulatory genes in these genomic regions include *CDNK2A,* a cyclin-dependent kinase inhibitor, and *TSC1,* a gene encoding hamartin mutated in tuberous sclerosis. Genomic investigations have demonstrated that low-grade tumors are characterized by alterations in the *RAS/RAF* signaling pathway with activating *FGFR3* mutations or gene fusions present in 60–80% of patients. In contrast, the high-grade invasive phenotype is notable for early deleterious mutations in *TP53* and *RB1*, alterations in *CDH1* (E-cadherin), and increased expression of *VEGFR2*. In urothelial carcinoma of the renal pelvis and ureter, 10–20% of cases may be associated with Lynch syndrome hereditary defects in the *MLH1*, *MSH2*, or *MSH6* mismatch repair genes, leading to microsatellite instability and frequent DNA mutations. Testing for germline mutations in these genes is recommended in patients with upper urinary tract urothelial carcinoma under the age of 60 at diagnosis, with a first-degree relative with a Lynch syndrome–associated cancer diagnosed under the age of 50, or with two first-degree relatives with a Lynch syndrome–associated cancer regardless of the age at diagnosis.

As genomic analysis technologies have improved, so has our understanding of the molecular biology unique to urothelial carcinoma. In 2017, the full bladder cancer results of The Cancer Genome Atlas (TCGA) project were published. This effort comprehensively analyzed gene mutations, fusions, expression, copy number variations, methylation, and microRNA across the genome of patients with bladder urothelial carcinoma treated with surgery. Key findings from this effort include (1) genomic alterations in genes (e.g., *FGFR3*, *EGFR*, *ERBB2*, *ERBB3*, *PIK3CA*, *TSC1*, etc.) targetable by currently approved drugs or drugs in development in 71% of patients; (2) genomic alterations in chromatin-modifying genes (*KMT2D, KDM6A, KMT2C, EP300, CREBBP*, etc.) in the majority of patients; (3) hypermethylation with epigenetic silencing of gene expression in one-fourth of patients; and (4) the identification by RNA sequencing of five distinct intrinsic molecular subtypes (luminal papillary, luminal infiltrated, luminal, basal squamous, and neuronal) closely resembling luminal and basal subclassifications of breast cancers. These bladder TCGA findings have led to clinical trial designs enriching for patients with specific gene mutation profiles as well as interrogation of candidate biomarkers according to intrinsic molecular subtypes.

STAGING AND OUTCOMES BY STAGE

The staging of bladder cancer is dependent on the depth of invasion within the bladder wall, involvement of lymph nodes, and spread to

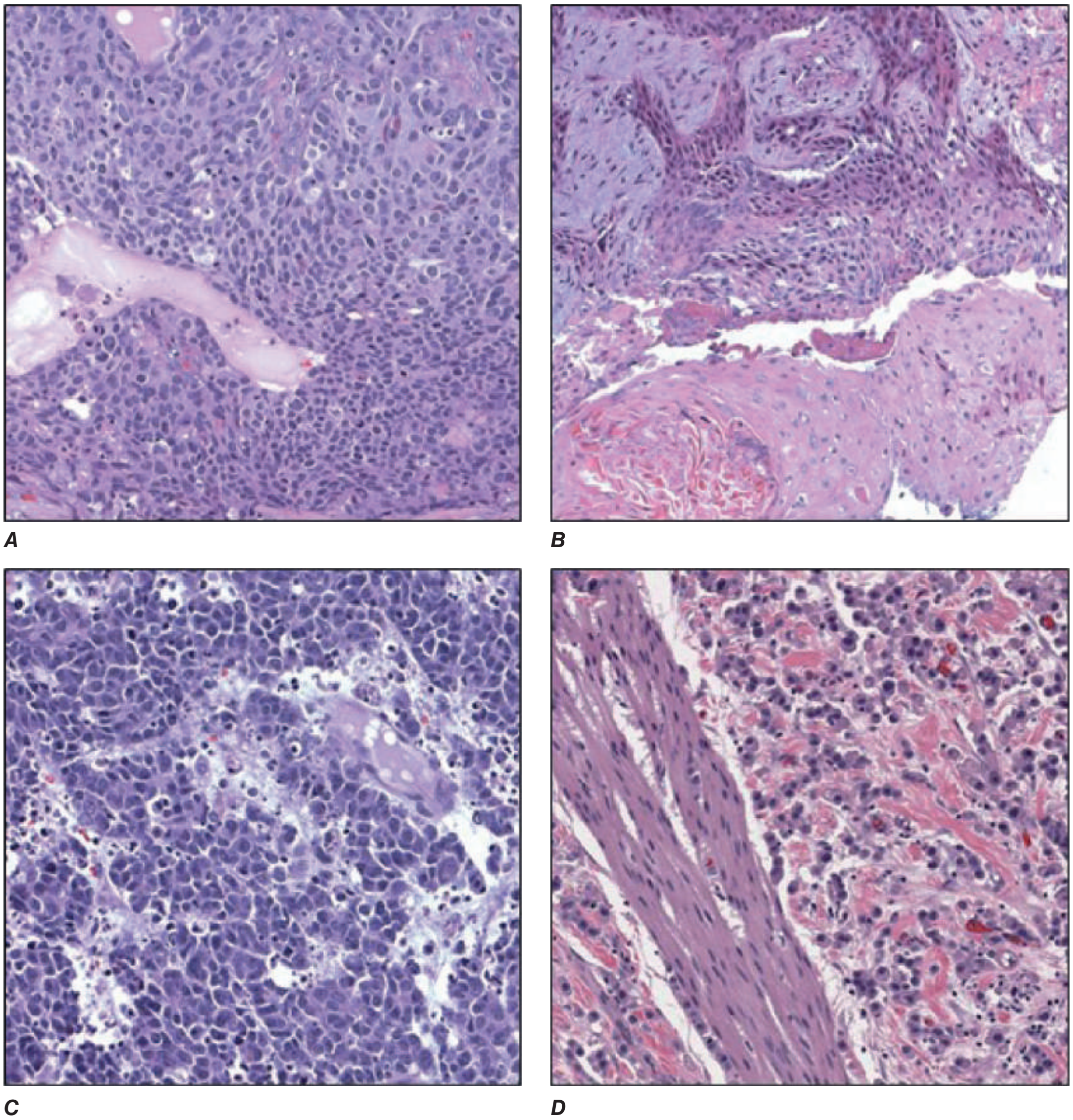

FIGURE 91-1 Bladder and urinary tract cancer histologies. ***A***. Urothelial carcinoma. ***B.*** Squamous cell carcinoma. ***C.*** Small-cell carcinoma. ***D.*** Plasmacytoid variant. *(Courtesy of Alex Baras, MD, PhD, Johns Hopkins University Department of Pathology.)*

surrounding and distant organs as depicted in **Fig. 91-2**. Approximately 75% of bladder cancer presents with non–muscle-invasive bladder cancer (NMIBC), 18% with disease invading into or through the muscular wall of the bladder, and only 3% with metastatic spread to distant organs. NMIBC is defined by tumors that involve only the immediate epithelial layer of cells (carcinoma in situ [CIS] and Ta) or that only penetrate into the connective tissue below the urothelium (T1) but not into the muscular layer known as the *muscularis propria*. Muscle-invasive bladder cancer (MIBC) is defined by tumors that invade into the muscularis propria (T2), through the muscularis propria to involve the surrounding serosa (T3), or into immediately adjacent pelvic organs such as the rectum, prostate, vagina, or cervix (T4). Lymph node staging is classified according to involvement of a solitary node within the true pelvis (N1), two nodes involved in the true pelvis (N2), or involvement of the common iliac nodes (N3). Any disease that has spread beyond the common iliac nodes is considered metastatic (M1). The staging of bladder cancer is driven primarily by the T stage of the tumor, with stages 0a–II defined entirely by the T stage in the absence of nodal or metastatic disease. Involvement of regional lymph nodes in the true pelvis or along the common iliac artery qualifies as stage III disease, whereas involvement of any distant metastases qualifies as stage IV disease. Clinical outcomes of patients with bladder cancer correlate closely with staging at diagnosis with 5-year overall survival rates of 70–90% for disease confined to the bladder (stage I–II), 39–50% for disease that penetrates through the bladder or has spread to regional lymph nodes (stage III), and only 8% for disease extending to metastatic sites (stage IV).

■ TREATMENT APPROACHES

Early-Stage Disease For NMIBC, removal of all visible tumors by TURBT in the operating room is considered the mainstay of surgical treatment. Risk of recurrence can be classified as low, intermediate, or high depending on the presence of features summarized in **Table 91-1**. For patients with low-risk disease, meta-analyses have demonstrated a 12% reduction in early relapses when a single chemotherapy treatment of mitomycin C, epirubicin, or gemcitabine was instilled directly into the bladder (intravesical therapy) within 24 hours of the TURBT. For patients with intermediate- or high-risk tumors, weekly intravesical instillations for 6 consecutive weeks of the attenuated mycobacterium strain known as *Bacille Calmette-Guérin* (BCG) reduce the risk of recurrence from 50 to 29%. In addition, BCG treatment has been shown to decrease the rate of progression to MIBC by 27%. Intravesical BCG is generally well tolerated. Side effects can include dysuria, urinary frequency, bladder spasms, hematuria, and, in rare cases (<5%), a systemic inflammatory response that can mimic disseminated BCG infection. Following a 6-week induction BCG schedule, additional maintenance BCG treatments given according to the Southwest

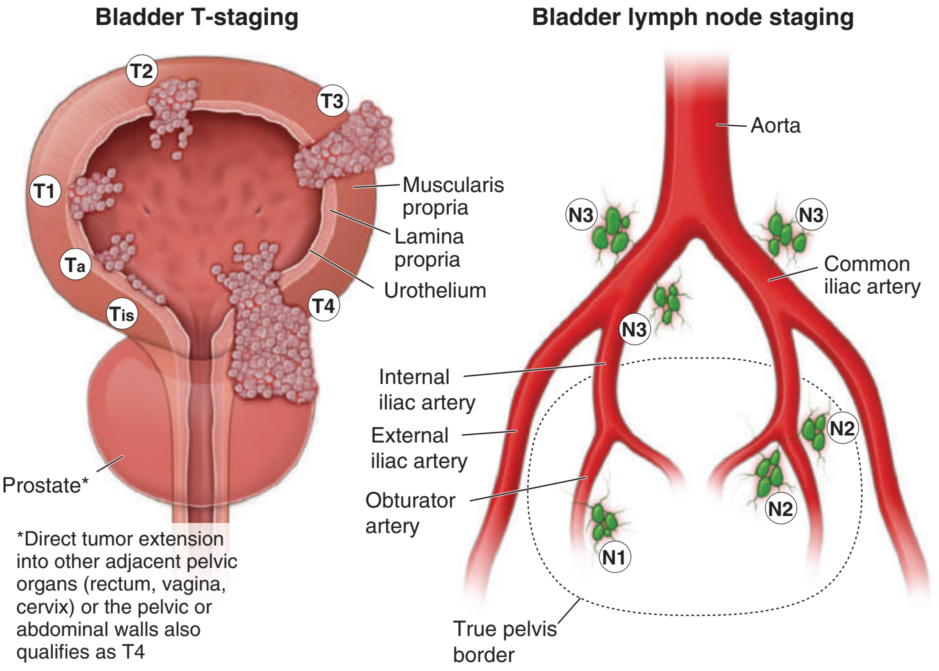

Bladder cancer prognosis according to stage

T	N	M	Stage	5-yr survival
Tis/Ta	N0	M0	0is/0a	96%
T1	N0	M0	1	90%
T2	N0	M0	2	70%
T3	N0	M0	3	50%
T1-T4	N1-N3	M0	3	36%
Any T	Any N	M1	4	5%

N1 - Cancer spread to 1 lymph node in the true pelvis

N2 - Cancer spread to 2 lymph nodes in the true pelvis

N3 - Cancer spread to lymph nodes along the common iliac artery

FIGURE 91-2 Bladder cancer staging and prognosis. TNM, tumor-node-metastasis.

Oncology Group schedule further reduce the risk of recurrent NMIBC compared to induction BCG alone. In patients with NMIBC that recurs long after initial BCG treatment, a repeat course of BCG can be considered. For patients with recurrence after a second adequate course of BCG or with relapsed NMIBC within 6 months of initial BCG exposure, surgical removal of the entire bladder by cystectomy is recommended due to the high risk of progression to MIBC and potentially metastatic disease. For patients who are not fit enough for or who refuse cystectomy, non-BCG alternative intravesical agents (nadofaragene firadenovec, mitomycin C, gemcitabine, docetaxel, valrubicin) or systemically administered agents that inhibit the PD-1/PD-L1 immune checkpoint pathway (pembrolizumab) can achieve durable tumor responses in a small fraction of patients.

Upper Tract Disease In patients with urothelial carcinoma of the renal pelvis or ureter, endoscopic tissue acquisition and staging are more challenging than primary tumors located in the bladder. Tumors possessing all of the following are considered low risk: solitary tumor, low grade, size <1 cm, and no invasive component on imaging. Low-risk tumors can successfully be treated by laser ureteroscopic ablation or surgical resection and reanastomosis of the remaining ureter ends in tumors that cannot be successfully eradicated endoscopically.

Muscle-Invasive Disease In patients with urothelial carcinoma of the bladder that invades into or through the muscularis propria but with no evidence of metastatic spread, more aggressive therapy options summarized in Table 91-2 are required to achieve cure. In carefully selected patients with no evidence of CIS or hydronephrosis, bladder-sparing combined-modality therapy with concurrent chemotherapy and radiation can achieve cure in ~65% of patients. Various chemotherapy regimens have been utilized in combination with radiation including cisplatin, carboplatin, 5-fluorouracil, mitomycin C, paclitaxel, and gemcitabine. It is important to note that a maximal debulking of all visible tumor by TURBT is required prior to initiation of combined-modality therapy. In patients who achieve a complete response to combined-modality therapy, regular cystoscopic monitoring of the bladder is required with salvage cystectomy offered to patients who develop MIBC in follow-up.

In a similar fashion, bladder-sparing partial cystectomy can be performed in a very small subset of MIBC patients. The ideal patient for partial cystectomy is the patient with a solitary, clinical T2 urothelial carcinoma in the dome of the bladder. In such patients, the tumor and immediate surrounding urothelium can

TABLE 91-1 Non–Muscle-Invasive Bladder Cancer Recurrence Risk Groups

RISK GROUP	CHARACTERISTICS
Low risk	Initial tumor, solitary tumor, low grade, <3 cm, no CIS
Intermediate risk	All tumors not defined in the two adjacent categories (between the category of low and high risk)
High risk	Any of the following: • T1 tumor • High-grade • CIS • Multiple and large (>3 cm) Ta low-grade tumors in patients over age 70 (all conditions must be met for this point on Ta low-grade tumors)

Abbreviation: CIS, carcinoma in situ.

TABLE 91-2 Treatment Approaches to MIBC Patients

TREATMENT	PATIENT SELECTION	CLINICAL OUTCOMES
Bladder-sparing chemoradiation	No CIS, no hydronephrosis, maximal TURBT required	65% cure, 55% bladder intact, highly dependent on patient selection
Bladder-sparing partial cystectomy	Solitary tumors in dome of bladder are ideal	Variable, highly dependent on patient selection
Cystectomy	Any MIBC patient	50% cure with surgery alone, highly dependent on pathologic stage
Neoadjuvant cisplatin-based chemotherapy	Cisplatin-eligible MIBC patients	5–10% improvement in overall survival compared to cystectomy alone
Adjuvant cisplatin-based chemotherapy	Cisplatin-eligible, high-risk, postcystectomy MIBC patients (pT3-4, N+)	Similar improvement as neoadjuvant treatment, data less robust, many patients not suitable for adjuvant treatment
Adjuvant nivolumab anti–PD-1 immunotherapy	Postsurgery, high-risk MIBC and UTUC patients (pT3-4, N+, cisplatin-eligible after neoadjuvant therapy OR pT2-4, N+, cisplatin-ineligible who did not receive neoadjuvant therapy)	30% improvement in disease-free survival compared to surgery alone

Abbreviations: CIS, carcinoma in situ; MIBC, muscle-invasive bladder cancer; TURBT, transurethral resection of bladder tumor; UTUC, upper tract urothelial carcinoma.

be resected with reconstruction of the remaining bladder to maintain near physiologic urinary function.

In the majority of patients, however, resection of the entire bladder is required. In males, a cystoprostatectomy with removal of the bladder, prostate, and pelvic lymph nodes is performed, whereas in females, an anterior exenteration with removal of the bladder, uterus, ovaries, cervix, and pelvic lymph nodes is performed. With the bladder removed, three options exist to reroute the urine outflow. In an ileostomy, the bilateral ureters are connected to a portion of ileum that is brought through an incision in the abdominal wall to create a stoma that drains urine into an affixed bag outside of the body. In a continent urinary reservoir or "Indiana pouch," the ureters are connected to a portion of ileum that has been separated on both ends from the rest of the small-bowel transit to form a urinary reservoir. The remaining small bowel is reanastomosed, and the urinary reservoir is brought up just beneath the abdominal wall muscles with patients catheterizing the urinary reservoir several times per day via a small stoma tract. Last, in a neobladder, the same urinary reservoir described previously is brought down into the pelvis and is anastomosed to the remaining urethra to provide the opportunity to the patient to void urine through the urethra. The choice of which urinary reconstruction to perform is affected not only by patient choice but also by anatomic tumor considerations and urologist experience with each procedure. Regardless of the type of surgery performed, all patients undergo a significant catabolic change in their metabolism following removal of the bladder. While many MIBC patients are affected by weight loss preoperatively, it is not uncommon for postcystectomy patients to lose an additional 10–15 pounds in the first month postoperatively. In addition, patients can experience long-term nutritional changes such as low B_{12} levels due to alterations in small-bowel physiology caused by all of the urinary diversion options.

Despite aggressive surgery, only half of patients undergoing cystectomy are cured by surgery alone. Therefore, many clinical trials have investigated the role of systemic chemotherapy and immunotherapy before (neoadjuvant) or after (adjuvant) surgery. Meta-analyses have shown a 5–10% absolute overall survival advantage when combination chemotherapy regimens utilizing cisplatin have been used before surgery. Importantly, non–cisplatin-containing chemotherapy regimens have proven inferior to cisplatin-containing regimens. Therefore, if patients are not suitable candidates for cisplatin administration due to poor functional status or comorbidities (e.g., poor renal function), patients should proceed directly to surgery and forego neoadjuvant therapy. A similar benefit exists with cisplatin-based combination chemotherapy given after surgery. However, in the postoperative setting, some patients may not recover sufficiently from their surgery within a time frame optimal for chemotherapy administration. In bladder and upper tract urothelial cancer patients with high-risk postsurgery pathology features, adjuvant treatment with the PD-1–targeting immune checkpoint inhibitor, nivolumab, has demonstrated a 30% improvement in disease-free survival compared to observation and is considered a standard treatment option.

For patients with high-risk urothelial carcinoma of the upper urinary tract, resection of the kidney and ureter (including the ureter bladder cuff) by nephroureterectomy is preferred. Segmental ureterectomy may be appropriate in patients with decreased renal function in which nephron-sparing outcomes are critical to prevent the need for dialysis. Similarly, in CIS patients, administration of BCG therapy via a nephrostomy tube can be considered to preserve intact renal function. The use of cisplatin-based neoadjuvant chemotherapy has been associated with a pathologic complete response at surgery of 14% in upper tract urothelial carcinoma patients. Similarly, in the post-nephroureterectomy setting, adjuvant platinum-based chemotherapy (carboplatin or cisplatin) reduced recurrence rates by 55% compared to surgery alone. The use of perioperative chemotherapy (before or after surgery) or PD-1–targeting immunotherapy (after surgery) is now recommended for upper tract urothelial carcinoma patients in national guidelines.

Metastatic Disease For patients with metastatic urothelial carcinoma regardless of primary tumor origin, systemic platinum-based chemotherapy remained the established initial treatment for decades. However, the regimen of a nectin-4–targeting antibody-drug conjugate, enfortumab vedotin, carrying the monomethyl auristatin E (MMAE) cytotoxic payload combined with the PD-1 immune checkpoint inhibitor pembrolizumab (EV+P) supplanted platinum-based chemotherapy in 2023 as the preferred first-line therapy for metastatic urothelial carcinoma patients. In a randomized, phase 3 clinical trial, EV+P treatment resulted in a highly significant improved median overall survival of 31.5 months compared to 16.2 months with platinum-based chemotherapy. These results led to the FDA approval of the EV+P regimen in first-line metastatic urothelial carcinoma patients. The optimal second-line therapy for patients who experience disease progression on front-line EV+P is not defined. Second-line options include platinum-based chemotherapy or fibroblast growth factor receptor (FGFR) inhibitor therapy (erdafitinib) for patients with activating *FGFR* genomic alterations. For patients with progression after receiving both platinum-based chemotherapy and a PD-1/PD-L1 immune checkpoint inhibitor, the Trop-2–targeting antibody-drug conjugate sacituzumab govitecan, containing a topoisomerase-inhibiting cytotoxic payload, SN-38, remains a third-line option. Prior to the approval of EV+P in the metastatic front-line setting, maintenance treatment with the PD-L1 immune checkpoint inhibitor avelumab had demonstrated an improvement in overall survival compared to platinum-based chemotherapy alone in patients demonstrating stable disease or objective response to their front-line treatment. The role of maintenance PD-1/PD-L1 immune checkpoint inhibitor therapy in the second-line and beyond settings is undefined and is being evaluated in ongoing clinical trials. Additional novel urothelial carcinoma therapeutics are under ongoing investigation.

With the increased number of therapeutic agents for metastatic urothelial cancer patients now available, differing toxicity profiles between treatment regimens commonly impact treatment decisions for individual patients. With platinum-based chemotherapy, renal insufficiency, myelosuppression, nausea, and neuropathy are common. With enfortumab vedotin, myelosuppression can occur; however, neuropathy is dose-limiting with rare but serious skin toxicity and new-onset diabetes events. Myelosuppression and significant diarrhea requiring intravenous fluid support are common with sacituzumab govitecan. In contrast, myelosuppression is uncommon with PD-1/PD-L1 immune checkpoint inhibitors. Rare immune-related toxicities can be severe and may include colitis, pneumonitis, hepatitis, nephritis, myocarditis, rash, hypothyroidism, Guillain-Barré syndrome, idiopathic thrombocytopenia purpura, and adrenal insufficiency. Lastly, erdafitinib side effects are notable for hyperphosphatemia, dystrophic nail changes, and rare central serous retinopathy. Thus, ophthalmologic evaluations are recommended for patients receiving FGFR inhibitor therapies.

FURTHER READING

American Cancer Society: *Cancer Facts & Figures 2023*. Atlanta, GA: Available from *https://www.cancer.org/cancer/bladder-cancer/detection-diagnosis-staging/survival-rates.html.*

Carlo MI et al: Cancer susceptibility mutations in patients with urothelial malignancies. J Clin Oncol 38:5, 2020.

Chou R et al: Intravesical therapy for the treatment of nonmuscle invasive bladder cancer: A systemic review and meta-analysis. J Urol 197:5, 2017.

Coleman JA et al: Diagnosis and management of non-metastatic upper tract urothelial carcinoma: AUA/SUO guideline. J Urol 209:6, 2023.

Dyrskjøt L et al: Bladder cancer. Nat Rev Dis Primers 9:1, 2023.

Howlader N et al: SEER Cancer Statistics Review, 1975-2017. Available from *https://seer.cancer.gov/csr/1975_2017/*. Based on November 2019 SEER data submission, posted to the SEER website, April 2020.

Powles T et al: Enfortumab vedotin and pembrolizumab in untreated advanced urothelial cancer. N Engl J Med 390:10, 2024.

Robertson AG et al: Comprehensive molecular characterization of muscle-invasive bladder cancer. Cell 171:3, 2017.

Siegel RL et al: Cancer statistics, 2023. CA Cancer J Clin 73:1, 2023.

Sylvester RJ et al: European Association of Urology (EAU) prognostic factor risk groups for non–muscle-invasive bladder cancer (NMIBC) incorporating the WHO 2004/2016 and WHO 1973 classification systems for grade: An update from the EAU NMIBC guidelines panel. Eur Urol 79:4, 2021.

US Food and Drug Administration: FDA approves enfortumab vedotin-ejfv with pembrolizumab for locally advanced or metastatic urothelial cancer. Available from: *https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-enfortumab-vedotin-ejfv-pembrolizumab-locally-advanced-or-metastatic-urothelial-cancer*.

92

Benign and Malignant Diseases of the Prostate

Oliver Sartor, James A. Eastham

Benign and malignant changes in the prostate increase with age. Autopsies of men in the eighth decade of life show hyperplastic changes in >90% and malignant changes in >70% of individuals. The high prevalence of these diseases among the elderly, who often have competing causes of morbidity and mortality, mandates a risk-adapted approach to diagnosis and treatment. This can be achieved by considering these diseases as a series of states. Each state represents a distinct clinical milestone for which therapy(ies) may be recommended based on disease extent, current symptoms, risk of developing symptoms, and/or the risk of death from prostate cancer in relation to death from other causes. Given the complexities and, at times, the uncertainties of optimal decision-making, patient preferences are an important consideration. For benign proliferative disorders, symptoms of bladder outlet obstruction and potential complications including urinary retention and urinary tract infection are weighed against the side effects and complications of medical or surgical intervention. For prostate malignancies, the likelihood that a clinically significant cancer is present in the gland and the concomitant risk of symptoms or death from cancer are necessarily balanced against the risks and morbidities of the recommended treatments.

ANATOMY AND PATHOLOGY

The prostate is in the pelvis and is adjacent to the rectum, the bladder, the periprostatic and dorsal vein complexes, the neurovascular bundles that are responsible for erectile function, and the urinary sphincter that is responsible for passive urinary control. The prostate is composed of branching tubuloalveolar glands arranged in lobules surrounded by fibromuscular stroma. The acinar unit includes an epithelial compartment made up of epithelial, basal, and neuroendocrine cells separated by a basement membrane and a stromal compartment that includes fibroblasts and smooth-muscle cells. Prostate-specific antigen (PSA) is produced in the epithelial cells. Both prostate epithelial cells and stromal cells express androgen receptors (ARs) and depend on androgens for growth. Testosterone, the major circulating androgen, is converted by prostatic 5α-reductase to dihydrotestosterone, a more potent androgen.

The periurethral portion of the gland increases in size during puberty and after the age of 55 years due to the growth of nonmalignant cells in the transition zone of the prostate that surrounds the urethra. Most cancers develop in the peripheral zone, and cancers in this location may or may not be palpated during a digital rectal examination (DRE).

PROSTATE CANCER

The American Cancer Society's estimates for prostate cancer in the United States for 2024 are 299,010 new prostate cancer cases and 35,250 deaths from prostate cancer. In the United States, prostate cancer is the most common nonskin malignancy in men and the second leading cause of cancer death. The absolute number of prostate cancer deaths has decreased in the past 30 years, attributed by some to the widespread use of PSA-based detection strategies. The incidence/mortality ratio for prostate cancer is extremely high relative to most other malignancies, and most men diagnosed with prostate cancer will not die from their disease.

■ EPIDEMIOLOGY

Epidemiologic studies show that the risk of being diagnosed with prostate cancer increases 2.5-fold if one first-degree relative is affected and even more if two or more are affected. Risk is greatest for those whose relatives are diagnosed before the age of 65. Current estimates are that 40% of early-onset and 5–10% of all prostate cancers are hereditary. Prostate cancer affects ethnic groups differently. Matched for age, African-American patients have a higher incidence and present at a more advanced stage with higher-grade, more aggressive cancers. The causes for this disparity are debated, but genetic, environmental, and societal factors are implicated. Genome-wide association studies (GWAS) have identified >100 prostate cancer susceptibility loci that are estimated to explain up to 25% of prostate cancer risk. Each prostatic risk loci typically confers only a small incremental risk of cancer, but combinations are more impactful. Some mutated genes such as *BRCA2, HOXB13, ATM*, and *PALB2* clearly confer an increased risk of prostate cancer; however, that risk can be substantially modulated by presence or absence of additional genomic prostatic risk loci. Polygenic risk scores to better ascertain prostate cancer risk are under development but not routinely available for clinical care today.

The prevalence of autopsy-detected cancers is relatively similar around the world, while the incidence of clinical disease varies. In part, this is due to the substantial variations in the use of PSA screening tests. Environmental and dietary factors may play a role in prostate cancer growth and progression. High consumption of dietary fats, such as α-linoleic acid or polycyclic aromatic hydrocarbons that form when red meats are cooked, is believed to increase risk. Like breast cancer in Asians, the risk of prostate cancer in Asians increases when they move to Western environments. Reasons for this are poorly understood but in part likely reflect a greater probability of PSA screening tests in the Western world. Protective factors may include consumption of lycopene found in tomatoes and inhibitors of cholesterol biosynthesis (e.g., statin drugs). Not smoking and avoiding obesity reduce the risk of disease progression.

■ DIAGNOSIS AND TREATMENT BY CLINICAL STATE

The prostate cancer continuum—from the appearance of a preneoplastic and invasive lesion that is localized to the gland to a metastatic lesion causing symptoms and, ultimately, mortality—can span years to decades. To limit overdiagnosis of clinically insignificant cancers and for disease management in general, competing risks are considered in the context of a series of clinical states **(Fig. 92-1)**. The states are defined operationally based on whether or not a cancer diagnosis has been established and, for those with a diagnosis, the state of the primary tumor (treated vs untreated), the presence of absence of metastases on imaging studies, and the measured level of testosterone in the blood. With this approach, an individual resides in only one state and remains in that state until progression. At each assessment, the decision to offer treatment and the specific form of treatment are based on the presence or absence of cancer-related symptoms and, if absent, the risk posed by the cancer relative to competing causes of morbidity and mortality that may be present in that individual. It follows that the more advanced the disease, the greater is the need for treatment. Individual decision-making and patient preferences are key for prostate cancer throughout the continuum. Considerable heterogeneity can be encountered in patients seeking treatment recommendations given widely varying perceptions of risks and benefits, especially for those with early-stage disease.

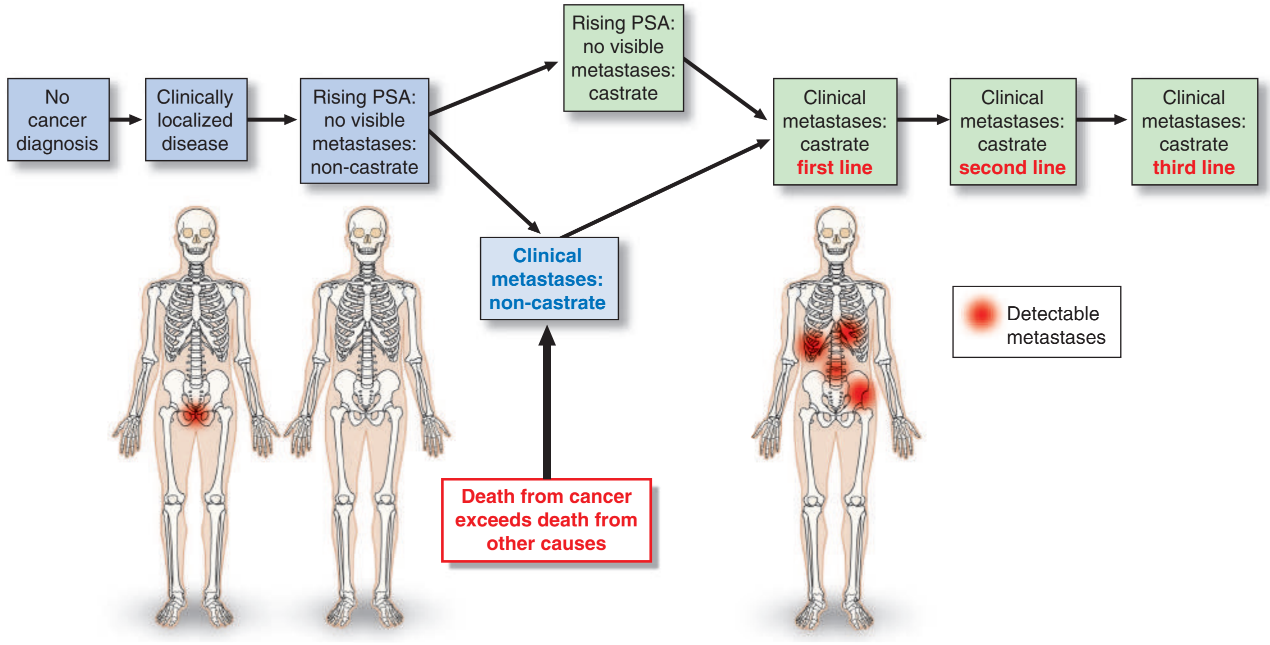
No cancer diagnosis
Clinically localized disease
Rising PSA: no visible metastases: non-castrate
Rising PSA: no visible metastases: castrate
Clinical metastases: non-castrate
Clinical metastases: castrate first line
Clinical metastases: castrate second line
Clinical metastases: castrate third line
Death from cancer exceeds death from other causes
Detectable metastases

FIGURE 92-1 Clinical states of prostate cancer. PSA, prostate-specific antigen.

For those without a cancer diagnosis, the decision to undergo testing to detect a cancer is based on the individual's estimated life expectancy and, separately, the probability that a clinically significant cancer may be present. For those with a prostate cancer diagnosis, the clinical states model considers the probability of developing symptoms or dying from the disease. Thus, a patient with a localized tumor that has been surgically removed remains in the state of localized disease if the PSA remains undetectable. The time within a state then becomes a measure of the efficacy of an intervention. Because many patients with active cancer are not at risk for developing metastases, symptoms, or death, the clinical states model allows a distinction between cure—the elimination of all cancer cells, the primary therapeutic objective of treatment for most cancers—and cancer control, by which the tempo of the illness is determined to be so slow or has been altered by treatment to the point where it is unlikely to cause symptoms, to metastasize, or to shorten a patient's life expectancy. Importantly, from a patient standpoint, both outcomes can be considered equivalent therapeutically assuming the patient has not experienced symptoms of the disease or the treatment needed to control it. Even when a recurrence is documented, immediate therapy is not always necessary. Rather, as at the time of diagnosis, the need for intervention is based on the tempo of the illness as it unfolds in the individual, relative to the risk-to-benefit ratio of the intervention being considered. Assessment of the future risk and tempo of the cancer, balanced with an understanding of the patient's morbidities and life expectancy, is key to management of prostate cancer.

■ NO CANCER DIAGNOSIS

Prevention No agent is currently approved for the prevention of prostate cancer. The results from several large double-blind, randomized chemoprevention trials have established 5α-reductase inhibitors (5ARIs) as the predominant therapy to reduce the future risk of a prostate cancer diagnosis; however, most of the prevented cancers were low risk, and those cancers typically require no treatment as their natural history is indolent. No drug is approved for prostate cancer prevention. Various vitamins and supplements have been studied in large-scale prostate cancer prevention trials, but no intervention has been shown to be effective.

Early Detection and Diagnosis The decision to pursue a diagnosis of prostate cancer must balance the benefit from detecting and treating clinically significant cancers that, left untreated, would adversely affect a patient's quality and duration of life, against the morbidity associated with the overtreatment of clinically insignificant cancers highly prevalent in the general population. The balance is best approached through shared decision-making between the patient and physician; however, the complexities of the subject and the limited time available for most primary care physicians often preclude truly informed decisions. Considerations regarding whether to pursue a diagnosis include symptoms, an abnormal DRE, or more typically, a change in or elevated serum PSA. Genetic risk and family history are also relevant in the decision-making process, and those with a family history of lethal cancer represent patients of particular concern.

PHYSICAL EXAMINATION The digital rectal exam (DRE) focuses on prostate size, symmetry, consistency, and abnormalities within and/or beyond the gland. Most cancers occur in the peripheral zone and may be palpated on DRE. Carcinomas are characteristically hard, nodular, irregular, and often result in an asymmetric DRE finding. Induration may also be due to benign prostatic hyperplasia (BPH) or calculi. Overall, 20–25% of patients with an abnormal DRE have prostate cancer. The DRE is a not accepted as a validated prostate cancer screening test given the frequent false positives and false negatives associated with this exam.

PROSTATE-SPECIFIC ANTIGEN PSA (kallikrein-related peptidase 3; KLK3) is a kallikrein-related serine protease that causes liquefaction of seminal coagulum. It is produced by both nonmalignant and malignant epithelial cells and, as such, is prostate specific, not prostate cancer specific. Serum levels of PSA may increase from prostatitis, BPH, or prostate cancer. PSA circulating in the blood is inactive and mainly occurs as a complex with the protease inhibitor α_1-antichymotrypsin and as free (unbound) PSA forms. A lower free PSA is more likely to be associated with cancer at the time of biopsy. PSA levels should be undetectable after about 6 weeks if the prostate has been completely removed (radical prostatectomy). If there is no prostate, there should be no detectable PSA. Persistence of measurable serum PSA after prostate removal is typical of recurrent cancer.

PSA testing was approved by the U.S. Food and Drug Administration (FDA) in 1994 for early detection of prostate cancer, and the widespread use of the test has played a significant role in the proportion of patients diagnosed with early-stage cancers: today, >70–80% of newly diagnosed cancers are clinically organ confined. The level of PSA in blood is strongly associated with the risk and outcome of prostate cancer. A single PSA measured at age 60 is associated (area under the curve [AUC] of 0.90) with lifetime risk of death from prostate cancer. Most (90%) prostate cancer deaths occur among patients with PSA levels in the top quartile (>2 ng/mL), although only a minority of patients with PSA >2 ng/mL will develop lethal prostate cancer. Despite this and mortality rate reductions reported from large, randomized prostate cancer screening trials, routine use of the test remains controversial given both false positives and false negatives and the fact that PSA testing may lead to the diagnosis and treatment of indolent cancers that pose little risk to the patient's overall health.

The U.S. Preventive Services Task Force (USPSTF) current guideline states, "For men aged 55 to 69 years, the decision to undergo periodic PSA-based screening for prostate cancer should be an individual one. Before deciding whether to be screened, men should have an opportunity to discuss the potential benefits and harms of screening with their clinician and to incorporate their values and preferences in the decision." The USPSTF recommends against screening men aged 70 or older. The USPSTF guideline can be accessed at *https://www.uspreventiveservicestaskforce.org/uspstf/draft-update-summary/prostate-cancer-screening-adults*. The American Urologic Association (AUA) guidelines indicate that "clinicians should engage in shared decision-making with people for whom prostate cancer screening would be appropriate and proceed based on a person's values and preferences." In contrast to the USPSTF, the AUA guidelines state that "clinicians should offer prostate cancer screening beginning at age 40 to 45 years for people at increased risk of developing prostate cancer based on the following factors: Black ancestry, germline mutations, strong family history of prostate cancer." Further, "clinicians should offer regular prostate cancer screening every 2 to 4 years to people aged 50 to 69 years." The current AUA guidelines are available at *https://www.auanet.org/guidelines-and-quality/guidelines/early-detection-of-prostate-cancer-guidelines*.

There is no rationale for recommending PSA screening in asymptomatic patients with a short life expectancy. Hence, patients over the age of 70 should only be tested in selected circumstances, such as a higher than median PSA measured before age 70 or excellent overall health. In addition, because a baseline PSA is a strong predictor of the future risk of lethal prostate cancer, patients with low PSAs, for example <1 ng/mL, can undergo testing less frequently, perhaps every 4 years, with screening possibly ending at age 60 if the PSA remains at ≤1 ng/mL

Though early detection of prostate cancer can potentially reduce the risk of prostate cancer death, high proportions of patients with screen-detected prostate cancer may not need immediate treatment and can be managed by active surveillance. The goal of prostate cancer screening should be to maximize the benefits of early cancer detection among patients who can benefit from treatment and minimize its harms.

The PSA criteria used to recommend additional testing have evolved over time. However, based on the commonly used cut point (a total PSA ≥4 ng/mL), most patients with a PSA elevation do not have histologic evidence of prostate cancer at biopsy. In addition, many patients with PSA levels below this cut point harbor cancer cells in their prostate. Information from the large Prostate Cancer Prevention Trial demonstrates that there is no PSA below which the risk of prostate cancer is zero. Thus, the PSA level helps establish the likelihood that a person will harbor cancer if he undergoes a prostate biopsy. The goal is to increase the sensitivity of the test for younger patients harboring clinically significant cancers that may shorten survival and to reduce the frequency of detecting cancers of low malignant potential in patients more likely to die of other causes. Patients with symptomatic bacterial prostatitis should have a course of antibiotics before biopsy. However, the routine use of antibiotics in an asymptomatic person with an elevated PSA level is discouraged.

SECOND-LINE SCREENING TESTS Several tests have been developed to better stratify patients with an elevated PSA test into those more or less likely to have clinically significant prostate cancer. The 4K Score Test (OPKO Lab, Nashville, TN) measures four prostate-specific kallikreins (total PSA, free PSA, intact PSA, and human kallikrein 2). The results are combined with clinical information in an algorithm that estimates an individual's percent risk of having an aggressive prostate cancer should that individual opt for a prostate biopsy.

The Prostate Health Index (PHI; Innovative Diagnostic Laboratory, Richmond, VA) is a blood test that estimates the risk of having prostate cancer. The PHI test is a combination of free PSA, total PSA, and the [−2]proPSA isoform of free PSA. These three tests are combined in a formula that calculates the PHI score. The PHI score is a better predictor of prostate cancer than the total PSA test alone or the free PSA test alone. Urine-based testing measuring exosomes (ExoDx Prostate Test) or mRNA levels of prostate cancer–related genes (Select-MDx) are also available.

The most important current test to stratify risk after detection of an elevated PSA is multiparametric prostate magnetic resonance imaging (mpMRI). The mpMRI varies from a traditional magnetic resonance imaging (MRI) by including not only the T1/T2-weighted images but also measurement of diffusion-weighted imaging with or without dynamic contrast enhancement. Using a structured reading system termed the Prostate Imaging Reporting and Data System (PI-RADS), patients can be risk stratified to determine the risk of clinically significant cancer. Data indicate that men with a PI-RADS 1–2 lesion can safely avoid biopsy, whereas patients with a PI-RADS 3–5 lesion can have targeted biopsies that are more likely to yield a true sense of the findings within the gland. The addition of an mpMRI in the prebiopsy decision-making unequivocally reduces the number of patients undergoing biopsy and reduces the diagnosis of Gleason 6 (grade group 1) cancers, without compromising the diagnosis of clinically relevant cancers. Today, an mpMRI of the prostate should be obtained in most cases prior to performing a prostate biopsy. Doing so will reduce the number of patients who are recommended to have a biopsy and improve the accuracy of the biopsies that are performed.

PROSTATE BIOPSY A diagnosis of cancer is established by an image-guided needle biopsy. Direct visualization of the gland by transrectal ultrasound (TRUS), MRI, or fusion of the ultrasound and MRI images ensures that all areas of the gland, including suspicious areas, are sampled. Contemporary schemas advise an extended-pattern 12-core biopsy that includes sampling from the peripheral zone as well as a lesion-directed palpable nodule or suspicious image-guided sampling. MRI-guided biopsy is superior to that of ultrasound-guided biopsy in terms of ascertaining an accurate depiction of the prostate pathology. Patients with an elevated PSA and a negative biopsy should continue to be followed as a negative biopsy does not rule out a future diagnosis of cancer.

PATHOLOGY Each core of the biopsy is examined for the presence of cancer, and the amount of cancer should be quantified based on the length of the cancer within the core and the percentage of the core involved. Of the cancers identified, >95% are adenocarcinomas; the rest are squamous or transitional cell tumors or, rarely, carcinosarcomas or small-cell histologies. Metastases to the prostate are rare, but colon cancers or transitional cell tumors of the bladder can invade the gland by direct extension.

When prostate cancer is diagnosed, a measure of histologic aggressiveness is assigned using the Gleason grading system, in which the dominant and secondary glandular histologic patterns are scored from 1 (well differentiated) to 5 (undifferentiated) and summed to give a total score of 6–10 for each tumor. The most poorly differentiated area of tumor (i.e., the area with the highest histologic grade) often determines biologic behavior. The presence or absence of perineural invasion and extracapsular spread should also be recorded.

Over the years, the Gleason grading system has undergone several changes. Currently, Gleason total scores of 2–5 are no longer assigned, and in practice, the lowest total score now assigned is 6. This leads to a logical yet incorrect assumption on the part of patients that a Gleason 6 cancer is in the middle of the scale, triggering the fear that their cancer is serious and the assumption that treatment is necessary despite

TABLE 92-1 TNM Classification

TNM (tumor, node, metastasis) Staging System for Prostate Cancer[a]	
Tx	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Localized Disease	
T1	Clinically inapparent tumor, neither palpable nor visible by imaging
T1a	Tumor incidental histologic finding in ≤5% of resected tissue; not palpable
T1b	Tumor incidental histologic finding in >5% of resected tissue
T1c	Tumor identified by needle biopsy (e.g., because of elevated PSA)
T2	Tumor confined within prostate[b]
T2a	Tumor involves half of one lobe or less
T2b	Tumor involves more than one half of one lobe, not both lobes
T2c	Tumor involves both lobes
Local Extension	
T3	Tumor extends through the prostate capsule[c]
T3a	Extracapsular extension (unilateral or bilateral)
T3b	Tumor invades seminal vesicles
T4	Tumor is fixed or invades adjacent structures other than seminal vesicles such as external sphincter, rectum, bladder, levator muscles, and/or pelvic wall
Metastatic Disease	
N1	Positive regional lymph nodes
M1	Distant metastases

[a]Revised from SB Edge et al (eds): *AJCC Cancer Staging Manual*, 7th ed. New York, Springer, 2010. [b]Tumor found in one or both lobes by needle biopsy, but not palpable or reliably visible by imaging, is classified as T1c. [c]Invasion into the prostatic apex or into (but not beyond) the prostatic capsule is classified not as T3 but as T2.

Abbreviation: PSA, prostate-specific antigen.

Gleason score 6 being favorable risk. To address these issues, a new five-grade group system has been developed:

Grade group 1 (Gleason score ≤6)
Grade group 2 (Gleason score 3 + 4 = 7)
Grade group 3 (Gleason score 4 + 3 = 7)
Grade group 4 (Gleason score 4 + 4 = 8)
Grade group 5 (Gleason scores 9 and 10)

The new system simplifies the grading of prostate cancer, appropriately classifying the lowest risk as grade group 1 (rather than Gleason score 6), and accurately predicts prognosis.

PROSTATE CANCER STAGING The TNM (tumor, node, metastasis) staging system includes categories for cancers that are identified solely based on an abnormal PSA (T1c), those that are palpable but clinically confined to the gland (T2), and those that have extended outside the gland (T3 and T4) (Table 92-1 and Fig. 92-2). DRE alone is inaccurate in determining the extent of disease within the gland, the presence or absence of capsular invasion, involvement of seminal vesicles, and extension of disease to lymph nodes. Unfortunately, no single test has been proven to perfectly indicate the stage or the presence of organ-confined disease, seminal vesicle involvement, or lymph node spread in part due to fact that cancer cells can microscopically spread below the limits of sensitivity for any imaging test. That said, mpMRI imaging can often reveal evidence of extracapsular spread or seminal vesicle invasion with a higher degree of accuracy than DRE, and taking MRI findings into account is typically done in staging today.

Prognosis and classification for nonmetastatic prostate cancer are typically done in three broad categories (low, intermediate, and high risk). Low-risk tumors are confined to the prostate with a PSA of <10 ng/mL and grade group 1. Intermediate-risk tumors are confined to the prostate, have a PSA between 10 and 20 ng/mL, or are grade group 2 or 3 (Gleason 7). This category is typically divided into a "favorable" and "unfavorable" intermediate risk depending on grade group 2 (favorable) or 3 (unfavorable) and number of positive biopsies (>50% of cores positive for cancer is unfavorable). High-risk tumors extend outside the prostate, have a PSA >20 ng/mL, or are grade group 4 or 5 (Gleason 8–10). There is also a subset of high-risk tumors termed "very high risk" in which the cancer extends into the seminal vesicles (T3b) or adjacent organs, i.e., rectum or bladder (T4), or there are multiple biopsy samples with high-grade cancer.

Computed tomography (CT) lacks sensitivity and specificity to detect extraprostatic extension and is inferior to MRI in visualization of lymph nodes. In general, mpMRI is superior to CT to detect cancers in the prostate, to assess local disease extent, and fused with ultrasound imaging, to guide sites to biopsy within the gland. Thus, mpMRI can be quite useful for the planning of surgery and radiation therapy.

The use of prostate-specific membrane antigen (PSMA) positron emission tomography (PET) scans has radically altered the assessment of prostate cancer spread beyond the prostate. This scan is now FDA approved for patients with intermediate- and high-risk prostate cancer and is substantially more sensitive and accurate in the detection of metastases as compared to conventional imaging (including MRI). Currently, state-of-the-art staging for Gleason grade group 3 and higher patients involves PSMA PET imaging before treatment. PSMA PET scans significantly improve specificity over conventional imaging while maintaining comparable sensitivity. PSMA PET scans are not capable of detecting low-volume microscopic disease. In studies where surgically removed lymph nodes were assessed for metastatic prostate cancer, sensitivity was consistently <50%. Simply stated, PET scans cannot detect microscopic disease (<2 mm focus), and a negative scan cannot rule out the possibility of cancer spread. We emphasize that no test is capable of microscopic cancer detection and note that a "negative" scan is not equated with the definitive absence of cancer.

Radionuclide bone scans (bone scintigraphy) have been used in the past to evaluate spread to osseous sites. This test is insensitive relative to PSMA PET and nonspecific because it does not detect the cancer itself, only reaction of the bone in the presence of the cancer. Staging

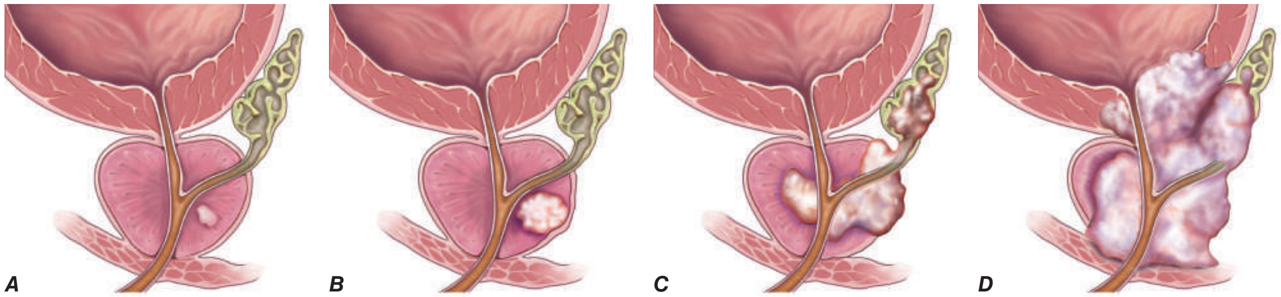

FIGURE 92-2 T stages of prostate cancer. ***A.*** T1—Clinically inapparent tumor, neither palpable nor visible by imaging. ***B.*** T2—Tumor confined within prostate. ***C.*** T3—Tumor extends through prostate capsule and may invade the seminal vesicles. ***D.*** T4—Tumor is fixed or invades adjacent structures. Eighty percent of patients present with local disease (T1 and T2), which is associated with a 5-year survival rate of 100%. An additional 12% of patients present with regional disease (T3 and T4 without metastases), which is also associated with a 100% survival rate after 5 years. Four percent of patients present with distant disease (T4 with metastases), which is associated with a 30% 5-year survival rate. (Three percent of patients are ungraded.) *(© 2010 Memorial Sloan-Kettering Cancer Center, New York, NY. All rights reserved. Republished with permission.)*

early-stage cancers with a bone scan is much less effective than using PSMA PET, and most centers no longer routinely use bone scans for initial staging in patients with suspected localized cancer given its known insensitivity and propensity for false-positive findings.

TREATMENT

Prostate Cancer

CLINICALLY LOCALIZED PROSTATE CANCER

Patients with clinically localized disease are typically managed by radical prostatectomy, radiation therapy, or active surveillance. Choice of therapy requires the consideration of several factors: the presence of symptoms, the probability that the untreated tumor will adversely affect the quality or duration of survival, and the probability that the tumor can be cured by single-modality therapy directed to the prostate versus requiring combinations of both local and systemic therapy.

There is no clear evidence for the superiority of any one form of local therapy relative to another. This is due to the lack of prospective randomized trials, referral bias and physician bias, variation in the experience of the treating teams, and differences in trial end points and the definitions of cancer control. Often, PSA relapse-free survival is used because an effect on metastatic progression or survival may not be apparent for years (if ever). A PSA recurrence does not necessarily mean that the disease will cause symptoms or shorten survival. The rate at which PSA doubles (PSA doubling time [PSADT]) is an important predictor of future clinical impact. Those with a PSADT of 14 months or greater are at low risk of clinical consequences, while those with a more rapid PSADT, especially those with a doubling time of 3 months or less, are at much higher risk for future clinically relevant events. Understanding the PSADT and the age/comorbidities of the patient is essential for decision-making.

After radical surgery to remove all prostate tissue, PSA should become undetectable in the blood within 6 weeks. If PSA remains or becomes detectable after radical prostatectomy, the patient is considered to have persistent or recurrent disease. A patient with no prostate should have no detectable PSA. After radiation therapy, in contrast, PSA does not become undetectable because the remaining nonmalignant elements of the gland continue to produce PSA even if the cancer cells have been eliminated.

Similarly, cancer control is not well defined by PSA for a patient managed by active surveillance. For patients with an intact prostate and a PSA of <10 ng/mL, the fluctuations in PSA are more reflective of the benign rather than malignant cells in the prostate. Other outcomes are time to objective progression (local or systemic), cancer-specific survival, and overall survival; however, these outcomes may take years to ascertain; thus, studies of localized disease are necessarily large, long, and often quite expensive.

The more extensive the local disease, the higher the probability of regional lymph node involvement (even when imaging studies are normal), the lower the probability of local control, and the higher the probability of relapse and the development of metastases. More important is that within the categories of clinical stage T1, T2, and T3 disease are cancers with a range of prognoses. Some T3 tumors are curable with therapy directed solely at the prostate, and some T1 lesions have a high probability of systemic relapse that requires the integration of local and systemic therapy to achieve cure. For T1c cancers, stage alone is inadequate to predict outcome and select treatment; other factors (especially Gleason score) must be considered. Genomic stratification of risk is increasingly viewed as important, and tools such as transcriptomic profiling may one day be routinely used in addition to traditional pathology.

To better assess risk and guide treatment selection, many groups have developed prognostic models or nomograms that use a combination of the initial clinical T stage, biopsy Gleason score, number of biopsy cores in which cancer is detected, and baseline PSA. Not all variables are equally important, and Gleason score is required to make accurate prognostication. Some prognostic models use discrete cut points, as mentioned above (PSA <10 or ≥10 ng/mL; Gleason score of ≤6, 7, or ≥8), while others employ nomograms that use PSA and Gleason score as continuous variables. More than 100 nomograms have been reported to predict (1) the probability that a clinically significant cancer is present, (2) disease extent (organ-confined vs non–organ-confined, node-negative or -positive), or (3) the relative probability of treatment success for specific local therapies using pretreatment variables.

Exactly what probability of success or failure would lead a physician to recommend and a patient to seek alternative approaches is controversial and highly dependent on the age and comorbidities of the patient. As an example, it may be appropriate to recommend radical surgery for a younger patient with intermediate-risk disease, while recommending surveillance for an older patient with a tumor having identical characteristics. Nomograms and various biomarkers are being refined continually to incorporate additional clinical parameters and biologic determinants that can affect outcomes, making treatment decisions a dynamic process. Patient preferences are also critically important in shared decision-making. Side effects of therapy are not the same for various treatment options, and patient preferences for one approach or another may require careful discussions that extend beyond the simple calculations of risk and benefit.

Radical Prostatectomy The goal of radical prostatectomy is to excise the cancer completely with a clear margin, to maintain continence by preserving the external sphincter, and to preserve potency by sparing the autonomic nerves in the neurovascular bundle. The procedure is advised for patients with a life expectancy of 10 years or more with a Gleason score 7 or higher disease. Radical prostatectomy can be performed via a retropubic or perineal approach or via a minimally invasive robotic-assisted or handheld laparoscopic approach. Outcomes can be predicted using postoperative nomograms that consider pretreatment factors and the pathologic findings at surgery. PSA failure is usually defined as a value >0.1 or 0.2 ng/mL, but as stated above, the patient without a prostate should typically have no detectable PSA. Specific criteria to guide the choice of one surgical approach over another are lacking. Minimally invasive approaches offer the advantage of a shorter hospital stay and reduced blood loss. Rates of cancer control, recovery of continence, and recovery of erectile function are comparable for robotic and open surgery. The individual surgeon, rather than the surgical approach used, is most important in determining outcomes after surgery.

Neoadjuvant hormonal treatment with gonadotropin-releasing hormone (GnRH) agonists/antagonists alone has also been explored to improve the outcomes of surgery for high-risk patients using a variety of definitions. The results of several large trials evaluating 3 or 8 months of androgen depletion before surgery have not been shown to improve clinically relevant outcomes. Currently, neoadjuvant hormonal therapies are not considered to be standard of care for surgically treated patients.

Factors associated with incontinence following radical prostatectomy include older age and membranous urethral length, which impacts the ability to preserve the urethra beyond the apex and the distal sphincter. The skill and experience of the surgeon are also critical factors.

The likelihood of recovery of erectile function is associated with younger age, quality of erections before surgery, and the absence of damage to the neurovascular bundles. In general, erectile function begins to return ~6 months after surgery if neurovascular tissue has been preserved. Potency is reduced by half if at least one neurovascular bundle is sacrificed. Overall, with the availability of drugs such as sildenafil, intraurethral inserts of alprostadil, and intracavernosal injections of vasodilators, many patients recover satisfactory sexual function.

Radiation Therapy Radiation therapy is given by external beam, by radioactive sources implanted into the gland, or by a combination of the two techniques.

Contemporary external beam intensity-modulated radiation therapy (IMRT) permits shaping of the dose and allows the delivery of higher doses to the prostate and a dramatic reduction in normal tissue exposure compared to three-dimensional conformal treatment alone. These advances have enabled the safe administration of doses >78 Gy and resulted in higher local control rates and fewer side effects. Proton beam radiation has not been demonstrated to have superior outcomes as compared to conventional radiation with IMRT. Lack of randomized studies prevents the use of routinely adopting proton beam as a standard of care.

Cancer control after radiation therapy has been defined by various criteria, including a decline in PSA to <0.5 or 1 ng/mL, "nonrising" PSA values, and (rarely) a negative biopsy of the prostate 2 years after completion of treatment. The standard definition of biochemical failure is a rise in PSA by ≥2 ng/mL higher than the lowest PSA achieved; however, newer data question whether the cutoff should be lower, particularly in those with higher grade cancers.

Radiation dose is critical to the eradication of prostate cancer. In a representative study, a PSA nadir of <1.0 ng/mL was achieved in 90% of patients receiving 81.0 Gy versus 76% and 56% of those receiving 70.2 and 64.8 Gy, respectively. Positive biopsy rates at 2.5 years were 4% for those treated with 81 Gy versus 27% and 36% for those receiving 75.6 and 70.2 Gy, respectively. Higher doses of radiation, however, may be associated with a higher risk of adverse events.

Hypofractionation schedules, utilizing fewer treatments of higher radiation doses, have been evaluated and shown to provide good cancer control rates based on posttreatment biopsies showing no evidence of cancer, with no apparent increase in treatment-related morbidity. Hypofractionated treatments can range from as few as 5 treatments to upward of 26 treatments, with both regimens representing substantial reductions in treatment length.

Multiple clinical trials have evaluated the use of androgen deprivation therapy (ADT) in combination with radiation. In patients with unfavorable intermediate-risk prostate cancer, short-course ADT (6 months), when combined with external beam radiotherapy, has demonstrated significant improvements in overall survival. Higher doses of radiation cannot substitute for ADT. In patients with high-risk localized disease, longer courses of ADT (18–36 months) have proven superior to shorter courses and represent the current standard of care when combined with radiotherapy. Novel genomic and transcriptomic biomarkers are being explored to assess the responsiveness of tumors to ADT, but as of yet, there are no accepted biomarkers that determine who should or should not receive ADT. Artificial intelligence assessments are promising, and new assays suggest the possibility that prostate cancers treated with radiation can be stratified into those that do and do not benefit from androgen deprivation.

For patients with nonmetastatic cancer by conventional imaging (bone scan, CT, or MRI) and two or more high risk factors (Gleason 8–10, stage T3/T4, or PSA ≥40 ng/mL) or positive pelvic lymph nodes, a large randomized trial demonstrated that the addition of abiraterone/prednisone to 2 years of ADT and radiotherapy is superior to 3 years of ADT and radiotherapy. This study defined a new standard of care for these particularly high-risk patients.

The Prostate Testing for Cancer and Treatment (ProtecT) trial investigated the effects of active monitoring, radical prostatectomy, and radical radiotherapy with hormones on patient-reported outcomes in patients diagnosed with predominantly low-risk prostate cancer (~75% with Gleason score 6 or grade group 1 cancer). Cancer outcomes were essentially the same for these three approaches. Patient-reported outcomes among 1643 patients who completed questionnaires before diagnosis, at 6 and 12 months, and annually thereafter were compared. Of the three treatments, prostatectomy had the greatest negative effect on sexual function and urinary continence, and although there was some recovery, these outcomes remained worse in the prostatectomy group than in the other groups throughout the trial. The negative effect of radiotherapy on sexual function was greatest at 6 months, but sexual function then recovered somewhat and was stable thereafter; radiotherapy had little effect on urinary continence. Sexual and urinary function declined gradually in the active-monitoring group. Bowel function was worse in the radiotherapy group at 6 months than in the other groups but then recovered somewhat, except for the increasing frequency of bloody stools; bowel function was unchanged in the other groups. Urinary voiding and nocturia were worse in the radiotherapy group at 6 months but then mostly recovered and were like the other groups after 12 months. Effects on quality of life mirrored the reported changes in function. No significant differences were observed among the groups in measures of anxiety, depression, or general health-related or cancer-related quality of life. Taken together, both surgery and radiation had clear side effects, but these side effects were distinct.

Brachytherapy Brachytherapy is the direct implantation of radioactive sources into the prostate. It is based on the principle that the deposition of radiation energy in tissues decreases as a function of the square of the distance from the source **(Chap. 78)**. The goal is to deliver intensive irradiation to the prostate, minimizing the exposure of the surrounding tissues. The current standard technique achieves a more homogeneous dose distribution by delivering radiation according to a customized template based on imaging assessment of the cancer and computer-optimized dosimetry. The implantation is typically performed transperineally as an outpatient procedure with real-time imaging.

Improvements in brachytherapy techniques have resulted in fewer complications and a marked reduction in local failure rates. In a series of 197 patients followed for a median of 3 years, 5-year actuarial PSA relapse–free survival for patients with pretherapy PSA levels of 0–4, 4–10, and >10 ng/mL were 98, 90, and 89%, respectively. In a separate report of 201 patients who underwent posttreatment biopsies, 80% were negative, 17% were indeterminate, and 3% were positive. The results did not change with longer follow-up. Brachytherapy is well tolerated, although most patients experience urinary frequency and urgency that can persist for several months. Higher complication rates are observed in patients who have undergone a prior transurethral resection of the prostate (TURP), while those with obstructive symptoms at baseline are at a higher risk for retention and persistent voiding symptoms. Proctitis has been reported in <2% of patients. Like surgery, brachytherapy is best performed by those well-practiced in the technique.

Active Surveillance With the advent of PSA testing, many patients are diagnosed with low-risk prostate cancers that may not pose a threat to either the quantity or quality of a person's life. Active surveillance, described previously as watchful waiting or deferred therapy, initially evolved from (1) studies that evaluated predominantly elderly patients with well-differentiated tumors (Gleason 6 or grade group 1) who remained untreated and demonstrated no clinically significant progression for protracted periods, (2) recognition of the contrast between incidence and disease-specific mortality, (3) the high prevalence of autopsy cancers, and (4) an effort to reduce overtreatment and treatment-related side effects. In practice, active surveillance is the treatment recommended to patients with cancers of low aggressiveness that can be safely monitored at fixed intervals with DREs, PSA measurements, imaging (best done with prostate mpMRI), and repeat prostate biopsies as indicated until histopathologic or serologic changes correlative of progression warrant treatment with curative intent. It is important to recognize that no treatment has proven superior to active surveillance for those with PSA-detected grade group 1 (Gleason score 6) disease. Surveillance remains underutilized today, and needless morbidity can result from overtreatment of those who have little chance to benefit from therapy given the indolent nature of the vast majority of patients with Gleason 6 localized disease.

Case selection is critical, and determining the clinical parameters predictive of cancer aggressiveness that can be used to reliably select patients most likely to benefit from active surveillance is an area of intense study. One set of criteria includes patients with clinical T1c tumors that are biopsy Gleason grade 6. Nomograms to help predict which patients can safely be managed by active surveillance continue to be refined, and as their predictive accuracy improves, it can be anticipated that more patients will be candidates. Older patients and those with substantial comorbidities can have more permissive criteria for surveillance including those with Gleason 3 + 4 = 7 (grade group 2) localized disease. Treatment decision-making for all localized prostate cancers depends on both the characteristics of the cancer and the age and comorbidities of the patient. Anxiety remains a problem for some patients with untreated cancers, and management of that anxiety can at times be problematic.

RISING PSA AFTER DEFINITIVE LOCAL THERAPY

Patients in this state include those in whom the sole manifestation of disease is a rising PSA after surgery and/or radiation therapy. For most of these patients in the initial phases, there is no evidence of disease on imaging studies. For these patients, the central issue is whether the rise in PSA results from persistent disease in the primary site, systemic disease, or both. Disease in the primary site may still be curable by additional local treatment.

The decision to recommend radiation therapy after prostatectomy is guided by the age and comorbidities of the patient, pathologic findings at surgery, the timing of PSA failure, the PSA doubling time, and the PSA level at the time of failure. Traditional imaging (MRI, CT, and radionucleotide bone scans), especially at low levels of PSA, is typically uninformative, but newer imaging methods may or may not be useful.

New PET tracers such as ^{18}F-fluciclovine and ^{18}F- or ^{68}Ga-PSMA that directly image the cancer are more sensitive and can detect low-volume disease in the prostate bed or other sites to better inform the decisions regarding treatment. Detection rates, both in and outside the prostate bed, correlate with the absolute level of PSA and the rate of PSA rise. Factors that predict for response to salvage radiation therapy after radical prostatectomy are a positive surgical margin, lower Gleason score in the surgical specimen, a long interval from surgery to PSA failure, slower PSA doubling time, a low (<0.5 ng/mL) PSA value at the time of radiation treatment, and the absence of distant disease when using a PSMA or fluciclovine PET scan.

For patients with a rising PSA after radiation therapy, salvage local therapy can be considered if persistent disease has been documented by a biopsy of the prostate and if no disease is detectable outside of the prostate bed or regional lymph nodes by imaging. Unfortunately, case selection is poorly defined in most series. Local treatment options for patients with recurrence after radiation include salvage radical prostatectomy, salvage cryotherapy, salvage radiation therapy, and salvage high-intensity focused ultrasound. Unfortunately, morbidities can be significant in the postradiated patient.

A rise in PSA after surgery or radiation therapy may indicate subclinical or micrometastatic disease with or without local recurrence. The need for treatment depends, in part, on the estimated probability that the patient will develop clinical symptoms and in what time frame. That immediate therapy is not always required was shown in a well-annotated series where patients who developed a rising PSA after radical prostatectomy received no systemic therapy until metastatic disease was documented. Overall, the median time to metastatic progression was 8 years, and 63% of the patients with rising PSA values remained free of metastases at 5 years. Factors associated with progression included the Gleason score of the radical prostatectomy specimen, time to recurrence after surgery, and PSADT. For those with Gleason score ≥8, the probability of metastatic progression was 37, 51, and 71% at 3, 5, and 7 years, respectively. If the time to recurrence was <2 years and PSADT was long (>10 months), the proportions with metastatic disease at the same time intervals were 23, 32, and 53%, versus 47, 69, and 79% if the doubling time was short (<10 months).

PSADTs are also prognostic for survival. Most physicians advise treatment when PSA doubling times are ≤10 months. A difficulty with predicting the risk of metastatic spread, symptoms, or death from disease in the rising PSA state is that many patients receive some form of therapy before the development of metastases as detected by conventional imaging. Nevertheless, predictive models continue to be refined with a consensus that lower risk patients can be watched, whereas higher risk patients may require some intervention. Given that new molecular imaging techniques are now FDA approved, the interplay between imaging and risk continues to be refined.

METASTATIC DISEASE: NONCASTRATE

The state of noncastrate metastatic disease includes patients with metastases visible on an imaging study at the time of diagnosis or after local therapy(ies) who have noncastrate levels of testosterone (>50 ng/dL). The prognosis is distinct for those with metastases detected by conventional imaging versus those detected only with molecular imaging. Symptoms of metastatic disease may include pain from osseous spread, although many patients are asymptomatic despite extensive spread. Less common are symptoms related to weight loss, fatigue, malaise, lymphedema, marrow infiltration by tumor (myelophthisis), coagulopathy, blood clots, or spinal cord compression. Standard treatment is to deplete or lower androgens via ADT by medical or surgical means, the latter being the least acceptable to patients. More than 90% of male hormones originate in the testes; <10% are synthesized in the adrenal gland **(Fig. 92-3)**.

Testosterone-Lowering Agents Medical therapies that lower testosterone levels include the GnRH agonists/antagonists, pure GnRH antagonists, 17,20-lyase inhibitors, CYP17 inhibitors, and estrogens such as oral diethylstilbestrol (DES) (Fig. 92-3). DES is not utilized today due to the excessive risk of vascular complications that include fluid retention, phlebitis, emboli, and stroke. GnRH agonists, such as leuprolide acetate and goserelin acetate, initially produce a very brief rise in luteinizing hormone and follicle-stimulating hormone followed by a downregulation of receptors in the pituitary gland, which causes a chemical castration. Regulatory approval was based on randomized trials showing reduced cardiovascular toxicities relative to DES, with equivalent potency. The initial rise in testosterone may occasionally result in a clinical flare of the disease, and as such, these agents are relatively contraindicated in patients with significant urinary obstructive symptoms, cancer-related pain, or spinal cord compromise. AR antagonists that block testosterone binding to the receptor are commonly used to reduce the risk of flare. Increases in testosterone do not occur with GnRH antagonists such as degarelix, given by injection, or relugolix, given orally, and these agents as well as abiraterone acetate rapidly achieve castrate levels of testosterone.

Agents that lower testosterone are associated with an androgen-deprivation syndrome that includes hot flushes, weakness, fatigue, loss of muscle mass, risk of osteoporosis and fracture, anemia, changes in cognition and personality, and depression. Changes in lipids, obesity, and insulin resistance and an increased risk of diabetes and cardiovascular disease may also be seen, along with a decrease in bone density that worsens over time and may result in an increased risk of clinical fractures. This is a particular concern in patients with preexisting osteopenia that results from hypogonadism that may be worsened with steroid or alcohol use and significantly underappreciated. Baseline fracture risk can be assessed using the FRAX scale, and to minimize fracture risk, patients can be advised to take a bisphosphonate or RANK-ligand inhibitor (denosumab) in combination with calcium and vitamin D supplementation. Hot flushes may be ameliorated by oxybutynin or certain antidepressants such as venlafaxine.

Antiandrogens Nonsteroidal first-generation antiandrogens such bicalutamide and nilutamide have largely been replaced by the more potent next-generation agents (enzalutamide, apalutamide, and darolutamide) that do not lower serum androgen levels and result in fewer hot flushes, less of an effect on libido, less muscle wasting, fewer personality changes, and less bone loss relative to

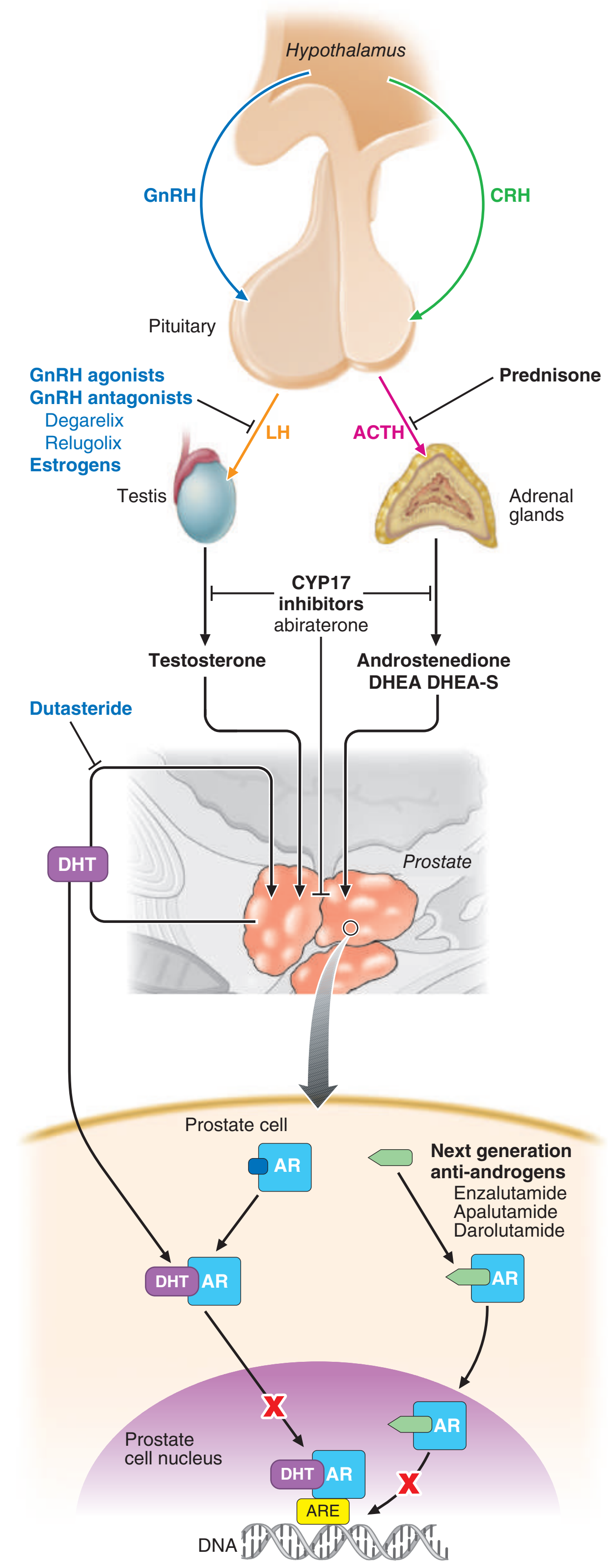

FIGURE 92-3 Sites of action of different hormone therapies.

testosterone-lowering therapies. However, over time, testosterone levels increase and are converted to estrogen, which can result in mastalgia and gynecomastia that limits long-term use. The side effect of antiandrogens as monotherapy can be prevented in part by tamoxifen, aromatase inhibitors, or prophylactic breast irradiation.

Most reported randomized trials in metastatic patients suggest that the cancer-specific outcomes are inferior when antiandrogens are used alone. Bicalutamide, even at a dose of 150 mg (three times the approved dose for use in combination in GnRH agonists), resulted in a shorter time to progression and inferior survival compared to surgical castration for patients with established metastatic disease. Newer studies for those without metastatic disease suggest that enzalutamide is appropriate in selected patients with a PSADT of 10 months or less.

Improving on the outcomes with ADT alone was a focus of the field for decades. Older antiandrogens such as bicalutamide or flutamide did not convincingly achieve these goals. Practice standards changed when an improvement in time to progression and overall survival was shown when ADT was combined with docetaxel relative to ADT alone. The greatest benefit was seen for patients with "high-volume" disease defined as the presence of ≥4 lesions on radionuclide bone scan or visceral disease. Longer progression-free and overall survival times have been noted in separate phase 3 trials comparing ADT with abiraterone, a CYP17 inhibitor that blocks androgen synthesis, and ADT with the AR antagonists such as enzalutamide and apalutamide versus the ADT standard, further changing the standard of care.

Intermittent Androgen Deprivation Therapy One way to reduce the side effects of androgen depletion is to administer hormonal therapy on an intermittent basis; however, the use of this approach in metastatic disease has yet to be proven useful for those being treated with ADT plus a novel hormone such as abiraterone, apalutamide, or enzalutamide. Additional studies in the metastatic setting with intermittent therapy are under consideration.

Chemotherapy Studies with ADT plus docetaxel indicate that patients with high-volume metastatic disease benefit from the addition of docetaxel. However, more recent studies indicate that ADT and docetaxel are not appropriate choices today given that the addition of abiraterone or darolutamide is superior to that of ADT plus docetaxel. Consultation with a clinician practiced in the art of advanced prostate cancer is appropriate given the rapidly changing standards of care.

Outcomes of Androgen Deprivation The anti–prostate cancer effects of the various newer androgen depletion strategies (abiraterone, apalutamide, or enzalutamide in combination with medical or surgical castration) are similar, and the clinical course is predictable: an initial response, a period of stability in which tumor cells are dormant and nonproliferative, followed after a variable period of time by a rise in PSA and regrowth that is visible on a scan as a castration-resistant lesion. Androgen depletion is not curative because cells that survive castration are present when the disease is first diagnosed. Considered by disease manifestation, PSA levels return to normal in 60–70% of patients, and measurable disease regression occurs in >50%. Duration of survival is inversely proportional to disease extent at the time androgen depletion is first started and the nadir level of PSA at 6–8 months. Patients with nadir PSA values >4 ng/mL have markedly inferior survival times and should be considered for alternative approaches.

METASTATIC DISEASE: CASTRATE

Castration-resistant prostate cancer (CRPC), disease that progresses while the measured levels of testosterone in the blood are 50 ng/mL or lower, can produce some of the most feared complications of the disease and is lethal for most patients. The most common manifestation is a rising PSA, frequently co-occurring with progression in bone. Nodal and/or visceral progression is less frequent. Symptoms may or may not be present. Critical for management is that therapeutic objectives be based on the manifestations of the disease in the individual at the time a change in therapy is being considered. As such, for the patient with symptomatic bone disease, relief of pain can be more clinically relevant than lowering the PSA. Naturally, for all patients, the central focus is delaying or preventing disease progression, symptoms, and death from disease.

Through 2010, docetaxel was the only FDA-approved life-prolonging therapy for CRPC. Since then, our understanding of the biology of the disease has increased significantly, which in turn has led to improved therapies. It is now recognized that the majority of metastatic CRPCs continue to express the AR and remain AR signaling dependent. For those with progression after ADT alone, the next-generation antiandrogen enzalutamide and the CYP17 inhibitor abiraterone acetate (given in combination with prednisone) have been proven to prolong life and are FDA approved for use in CRPC in both the pre- and postchemotherapy setting.

Large-scale molecular profiling efforts have led to a biologically based disease taxonomy that continues to evolve and showed a markedly higher than expected frequency of germline (~6% of patients) and somatic *BRCA* (also ~6% of patients) alterations, along with other genes in the DNA damage repair pathway that have been targeted successfully with poly-ADP ribose polymerase (PARP) inhibitors of which two, olaparib and rucaparib, are FDA approved as monotherapies. Combinations of abiraterone and olaparib are FDA approved for *BRCA1*/*BRCA2*-mutated CRPC patients, and enzalutamide/talozaparib combinations are approved for *BRCA1*/*BRCA2*-mutated patients and other DNA repair pathway mutations. Also approved are the checkpoint inhibitors pembrolizumab and dostarlimab for tumors with high microsatellite instability (MSI) scores, mismatch repair gene deficiency, or patients with a high tumor mutational burden. These alterations are found in ~3% of prostate cancers.

Other classes of therapy approved for selected CRPC patients based on demonstrated survival benefits include the biologic agent sipuleucel-T, the second-generation taxane cabazitaxel, the α-emitting bone-targeting radiopharmaceutical radium-223, and the PSMA-directed radionuclide therapy 177-lutetium PSMA-617. The use of 177-lutetium PSMA-617 has garnered intense interest given that this targeted form of radiation was able to extend survival in patients who had exhausted most conventional forms of therapy. Overall, an intense focus of current CRPC research is to understand the optimal sequence in which to utilize these agents to maximize benefit for the individual patient. Most of these agents are also being tested earlier in the course of the disease when tumor burdens are lower and the disease less heterogeneous. The result has been an increase in the frequency of late-state tumors that have undergone a lineage transformation from epithelial to alternative phenotypes (neuroendocrine and more) and are highly resistant to available therapies.

Pain Management Pain secondary to osseous metastases is one of the most feared complications of the disease and a major cause of morbidity, worsened by the narcotics needed to control symptoms.

Management requires accurate diagnoses because noncancer etiologies including degenerative disease, spinal stenosis, and vertebral collapse secondary to bone loss are common. Neurologic symptoms, including those suggestive of base of skull disease or spinal cord compromise, require emergency evaluation because loss of function may be permanent if not addressed quickly. Neurologic symptoms and loss of function are best treated with external beam radiation, as are single sites of pain. Diffuse symptoms in the absence of neurologic deficits can be treated with bone-seeking radioisotopes, such as radium-223 or the β emitter ^{153}Sm–ethylene-diamine-tetramethylene-phosphonic acid (EDTMP); mitoxantrone; or other systemic therapies. Radium-223 is indicated for patients with symptoms and has been shown to prolong survival, whereas ^{153}Sm-EDTMP and mitoxantrone are approved for the palliation of pain but have not been shown to prolong life. Abiraterone, enzalutamide, docetaxel, cabazitaxel, and Lu-177 PSMA-617 do not have a formal indication for pain but were shown to palliate pain in the registration trials that led to their approval by showing a survival benefit.

Other bone-targeting agents, including bisphosphonates such as zoledronic acid and the RANK-ligand inhibitor denosumab, have been shown to reduce the frequency and development of skeletal complications including pain requiring analgesia, neurologic compromise from epidural extension of tumor, and/or the need for surgery or radiation therapy to treat symptomatic osseous disease. It is important to note that, for all of these agents, the direct effects on the tumor are limited and benefits are seen without declines in PSA or improvements on imaging.

BENIGN DISEASE

BENIGN PROSTATIC HYPERPLASIA

BPH is a pathologic process that contributes to the development of lower urinary tract symptoms (LUTS) in patients. LUTS, arising from lower urinary tract dysfunction, are further subdivided into obstructive symptoms (urinary hesitancy, straining, weak stream, terminal dribbling, prolonged voiding, incomplete emptying) and irritative symptoms (urinary frequency, urgency, urge incontinence, small voided volumes). LUTS and other sequelae of BPH are not just due to a mass effect but are also likely due to a combination of the prostatic enlargement and age-related detrusor dysfunction.

Diagnostic Procedures and Treatment LUTS are generally measured using a validated, reproducible index that is designed to determine disease severity and response to therapy—the AUA's Symptom Index (AUASI), also adopted as the International Prostate Symptom Score (IPSS) (Table 92-2). Serial AUASI is particularly useful in

TABLE 92-2 AUA Symptom Index

	AUA SYMPTOM SCORE (CIRCLE 1 NUMBER ON EACH LINE)					
QUESTIONS TO BE ANSWERED	NOT AT ALL	LESS THAN 1 TIME IN 5	LESS THAN HALF THE TIME	ABOUT HALF THE TIME	MORE THAN HALF THE TIME	ALMOST ALWAYS
Over the past month, how often have you had a sensation of not emptying your bladder completely after you finished urinating?	0+	1	2	3	4	5
Over the past month, how often have you had to urinate again less than 2 h after you finished urinating?	0	1	2	3	4	5
Over the past month, how often have you found you stopped and started again several times when you urinated?	0	1	2	3	4	5
Over the past month, how often have you found it difficult to postpone urination?	0	1	2	3	4	5
Over the past month, how often have you had a weak urinary stream?	0	1	2	3	4	5
Over the past month, how often have you had to push or strain to begin urination?	0	1	2	3	4	5
Over the past month, how many times did you most typically get up to urinate from the time you went to bed at night until the time you got up in the morning?	(None)	(1 time)	(2 times)	(3 times)	(4 times)	(5 times)
Sum of 7 circled numbers (AUA Symptom Score): ____						

Abbreviation: AUA, American Urological Association.

Source: Modified with permission from MJ Barry et al: The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. J Urol 148:1549, 1992.

following patients as they are treated with various forms of therapy. Asymptomatic patients do not require treatment regardless of the size of the gland, while those with an inability to urinate, gross hematuria, recurrent infection, or bladder stones require evaluation and treatment. In patients with symptoms, uroflowmetry can identify those with normal flow rates who are unlikely to benefit from treatment, and bladder ultrasound can identify those with high postvoid residuals who may need intervention. Pressure-flow (urodynamic) studies detect primary bladder dysfunction. Cystoscopy is recommended if hematuria is documented and to assess the urinary outflow tract before surgery. Imaging of the upper tracts is advised for patients with hematuria, a history of calculi, or prior urinary tract problems.

Symptomatic relief is the most common reason patients seek treatment for BPH, and therefore, symptomatic relief is usually the goal of therapy for BPH. α-Adrenergic receptor antagonists are thought to treat the dynamic aspect of BPH by reducing sympathetic tone of the bladder outlet, thereby decreasing resistance and improving urinary flow. 5ARIs are thought to treat the static aspect of BPH by reducing prostate volume and having a similar, albeit delayed effect. 5ARIs have also proven beneficial in the prevention of BPH progression, as measured by prostate volume, the risk of developing acute urinary retention, and the risk of having BPH-related surgery. The use of an alpha-adrenergic receptor antagonist and a 5ARI as combination therapy seeks to provide symptomatic relief while preventing progression of BPH.

Another class of medications that has shown improvement in LUTS secondary to BPH is phosphodiesterase-5 (PDE5) inhibitors, used currently in the treatment of erectile dysfunction. All four of the PDE5 inhibitors available in the United States—sildenafil, vardenafil, tadalafil, and avanafil—appear to be effective in the treatment of LUTS secondary to BPH. The use of PDE5 inhibitors is not without controversy, however, given the fact that short-acting phosphodiesterase inhibitors such as sildenafil need to be dosed separately from alpha blockers such as tamsulosin because of potential hypotensive effects.

Symptoms due to BPH often coexist with symptoms due to overactive bladder, and the most common pharmacologic agents for the treatment of overactive bladder symptoms are anticholinergics. This has led to multiple studies evaluating the efficacy of anticholinergics for the treatment of LUTS secondary to BPH.

Surgical therapy is now considered second-line therapy and is usually reserved for patients after a trial of medical therapy. The goal of surgical therapy is to reduce the size of the prostate, effectively reducing resistance to urine flow. Surgical approaches include TURP, transurethral incision, or removal of the gland via a retropubic, suprapubic, or perineal approach. Also used are transurethral ultrasound–guided laser-induced prostatectomy (TULIP), stents, and hyperthermia.

FURTHER READING

Bergengren O et al: 2022 Update on prostate cancer epidemiology and risk factors: A systematic review. Eur Urol 84:191, 2023.

Deek MP et al: Multi-institutional analysis of metastasis-directed therapy with or without androgen deprivation therapy in oligometastatic castration-sensitive prostate cancer. Eur Urol Oncol 7:1403, 2024.

Donovan JL et al: Patient-reported outcomes 12 years after localized prostate cancer treatment. NEJM Evid 2:EVIDoa2300018, 2023. Erratum in: NEJM Evid 2:EVIDx2300122, 2023.

Eastham JA et al: Clinically localized prostate cancer: AUA/ASTRO Guideline. J Urol 208:505, 2022.

Fendler WP et al: Assessment of 68Ga-PSMA-11 PET accuracy in localizing recurrent prostate cancer: A prospective single-arm clinical trial. JAMA Oncol 5:856, 2019.

Haile ES et al: Medical management of benign prostatic hyperplasia. Cleve Clin J Med 91:163, 2024.

Hamdy FC et al: Fifteen-year outcomes after monitoring, surgery, or radiotherapy for prostate cancer. N Engl J Med 388:1547, 2023.

Hugosson J et al: Prostate cancer screening with PSA and MRI followed by targeted biopsy only. N Engl J Med 387:2126, 2022.

Hussain M et al: Survival with olaparib in metastatic castration-resistant prostate cancer. N Engl J Med 383:2345, 2020.

Kasivisvanathan V et al: MRI-targeted or standard biopsy for prostate-cancer diagnosis. N Engl J Med 378:1767, 2018.

Merseburger AS et al: Genomic testing in patients with metastatic castration-resistant prostate vancer: A pragmatic guide for clinicians. Eur Urol 79:519, 2021.

Morgan TM et al: Salvage therapy for prostate cancer: AUA/ASTRO/SUO guideline part I: Introduction and treatment decision-making at the time of suspected biochemical recurrence after radical prostatectomy. J Urol 211:509, 2024.

Pinsky PF, Parnes H: Screening for prostate cancer. N Engl J Med 388:1405, 2023.

Roach M 3rd et al: Prostate cancer, race, and health disparity: What we know. Cancer J 29:328, 2023.

Sartor O et al: Lutetium-177-PSMA-617 for metastatic castration-resistant prostate cancer. N Engl J Med 385:1091, 2021.

Virgo KS et al: Initial management of noncastrate advanced, recurrent, or metastatic prostate cancer: ASCO guideline update. J Clin Oncol 39:1274, 2021.

93

Testicular Cancer

David J. Vaughn

Testicular germ cell tumors (GCTs) represent 95% of all testicular neoplasms. Non-GCTs of the testis are much less common. Approximately 5% of GCTs arise in extragonadal locations including the mediastinum, retroperitoneum, and pineal gland. Treatment for testicular GCTs is determined by pathology and stage. The development of effective chemotherapy for this disease represents a landmark achievement in oncology. About 95% of newly diagnosed patients with testicular GCTs will be cured. For this reason, testicular cancer has been called “a model for a curable neoplasm.”

INCIDENCE AND GLOBAL CONSIDERATIONS

In 2023, ~9200 cases of testicular GCTs will be diagnosed in the United States, with 470 deaths. The incidence of testicular GCTs appears to be increasing worldwide. The disease has the highest incidence in Scandinavia, Western Europe, and Australia/New Zealand. Africa and Asia have the lowest incidence. The incidence in the United States and the United Kingdom is intermediate. While a distinct biology related to geography is not apparent, several countries have reported a migration to earlier stage disease in part related to public awareness and earlier diagnosis.

EPIDEMIOLOGY

Testicular GCT is the most common malignancy diagnosed in adolescent and young adult males (defined as age 15–39 years). The incidence in patients over 50 is increasing. Testicular GCT is most commonly diagnosed in Caucasians. The disease is much less commonly seen in African Americans. Testicular GCTs have an estimated heritability of almost 50%. Interestingly, the risk of GCT is higher in male siblings than in offspring of the patient. Although epidemiologic studies have been performed attempting to identify a relationship with environmental exposures, no conclusive causal links have been established.

Risk Factors The strongest risk factors for testicular GCT include a prior history of the disease, cryptorchidism, and a history of testicular germ cell neoplasia in situ. Patients with a prior history of testicular GCT have a 1–2% risk of developing a contralateral GCT. These are more commonly metachronous than synchronous. Men with cryptorchidism have approximately a four- to sixfold increased risk of developing testicular GCT. Orchidopexy before puberty decreases but does not eliminate this risk. Interestingly, the contralateral descended testis is also at risk for this disease. Men undergoing infertility evaluation in which a testicular biopsy demonstrates germ cell neoplasia in situ have a significant risk of developing GCT. Although scrotal ultrasound of patients with testicular GCT may demonstrate testicular microcalcifications that may be related to germ cell neoplasia in situ, the significance of testicular microcalcifications in the general population is unclear.

■ BIOLOGY

The primordial germ cell is the cell of origin for GCTs. Most malignant GCTs arise from in situ neoplasia. The molecular events that result in the development of germ cell neoplasia in situ and subsequent malignant GCT have not been fully determined. However, genetic analysis of GCTs has demonstrated an excess copy number of isochromosome 12p (i[12p]) in most cases. Several genome-wide association studies have identified multiple independent loci associated with testicular GCT risk. The strongest of these is the *KITLG* (KIT ligand) locus on chromosome 12. These loci contribute significantly to the heritable risk of this disease.

■ PATHOLOGY

GCTs are either seminomas or nonseminomas. For a tumor to be considered a seminoma, it must be 100% seminoma. Any mixed GCT should be approached as a nonseminomatous GCT. Seminomas represent ~50% of cases. Seminomas arise most commonly in patients in the fourth decade of life. Seminomas may contain syncytiotrophoblastic cells, which may secrete β-human chorionic gonadotropin (hCG). Seminomas do not secrete α-fetoprotein (AFP). Seminomas are exquisitely sensitive to both chemotherapy and radiation therapy. Nonseminomatous GCTs are most commonly diagnosed in the third decade of life. The histologic subtypes include embryonal carcinoma, yolk sac tumor, choriocarcinoma, and teratoma. Embryonal carcinoma is the most undifferentiated nonseminomatous GCT subtype with the potential to differentiate into the other subtypes. Embryonal carcinoma may secrete AFP, hCG, both, or neither. Yolk sac tumor often secretes AFP. Choriocarcinoma is an aggressive subtype, often secreting hCG at very high levels. These nonseminomatous GCT subtypes are all considered chemotherapy sensitive. Teratoma is composed of somatic cell types that are derived from two or more germinal layers (endoderm, mesoderm, and ectoderm). Teratomas are classified as mature, in which cell types resemble normal adult somatic tissue; immature, in which cell types resemble fetal somatic tissue; and malignant, in which the cell types have undergone malignant transformation into the malignant counterpart of the somatic tissue. Teratomas are chemotherapy resistant and must be treated surgically.

■ INITIAL PRESENTATION

Signs and Symptoms Although a painless testicular mass is pathognomonic of a GCT, most patients present with testicular swelling, firmness, discomfort, or a combination of these. The differential diagnosis may include epididymitis or orchitis, and a trial of antibacterials may be considered. Patients with retroperitoneal metastases may complain of back or flank pain. Patients may have cough, shortness of breath, or hemoptysis because of lung metastases. In patients with elevation of serum hCG, gynecomastia may be present. Diagnostic delay is not uncommon and may be associated with a more advanced stage at diagnosis.

Physical Examination Careful examination of the affected testis and the contralateral normal testis should be performed. Many tumors will have a hard consistency to palpation. Some patients may show testicular atrophy. Evaluation for supraclavicular lymphadenopathy, gynecomastia, and abdominal mass should be performed. Inguinal lymphadenopathy is rare. Most patients with lung metastases will have normal auscultation of the lungs.

Diagnostic Testing If a firm testicular mass is identified, a scrotal ultrasound should be performed. Patients with suspected epididymitis or orchitis who do not respond to antibiotics should also undergo scrotal ultrasound. Scrotal ultrasound should include both testicles. On ultrasound, a testicular GCT is hypoechoic and may be multifocal. A solid mass identified on ultrasound should be considered malignant until otherwise proven. Transscrotal aspiration or biopsy of a testicular mass should never be performed. Such scrotal violation may result in tumor seeding of the scrotum or inguinal lymph nodes.

Serum Tumor Markers Serum AFP, hCG, and lactate dehydrogenase (LDH) should be measured in patients suspected of testicular GCT. AFP is elevated in ~60–70% of patients who present with nonseminomatous GCTs. Seminomas never secrete AFP. A patient with a seminoma with elevation of AFP should be approached as having a nonseminomatous GCT. The half-life of AFP is 5–7 days. A falsely elevated AFP may be seen in patients with hepatic disease or a condition called hereditary persistence of AFP, in which patients may have baseline AFP levels that are mildly elevated. hCG may be elevated in both nonseminomatous GCTs as well as seminomas. Patients with choriocarcinoma may have markedly elevated levels of hCG. The half-life for hCG is 24–36 h. False-positive elevation of hCG may be seen secondary to hypogonadism, marijuana use, or because of interfering substances measured by the assay. LDH is a nonspecific marker for GCT. Its principal use is to help in the assessment of the risk classification of a patient with metastatic disease. Although elevation of serum tumor markers supports the diagnosis of a testicular GCT, most patients with seminoma and up to a third of patients with nonseminomatous GCTs do not have elevated levels. Serum microRNA (miR)-371a-3 has been identified as a promising biomarker for GCT, and validation studies are ongoing.

■ INITIAL MANAGEMENT

Inguinal Orchiectomy Prompt referral to urology should be performed if a testicular GCT is suspected. The initial treatment for most patients suspected of having a testicular GCT is radical inguinal orchiectomy with removal of the testicle and spermatic cord to the level of the internal inguinal ring. In patients who present with metastatic disease and the diagnosis of GCT is certain, orchiectomy may be deferred until completion of chemotherapy. Although some institutions perform testis-sparing surgery in select patients, the gold standard remains radical inguinal orchiectomy. Pathologic examination of the entire testicle is important, since testicular GCTs may be multifocal. Given the rarity of this cancer, review by an experienced pathologist is essential for accurate tumor classification. Serum tumor markers should be obtained before and after orchiectomy.

Staging The staging of testicular GCT is based on an understanding of the pattern of spread. The initial spread is by the lymphatic route to the retroperitoneal lymph nodes. A left-sided testicular GCT spreads first to the primary landing zone of left paraaortic lymph nodes inferior to the left renal vessels. A right-sided testicular GCT spreads first to the primary landing zone of the aortocaval nodes inferior to the right renal vessels. Nodal metastases may extend into the iliac regions. If scrotal violation occurred, inguinal lymph node metastases may be seen. Subsequent lymphatic spread is to the retrocrural, mediastinal, and supraclavicular lymph nodes. Hematogenous spread to the lung is the next most common site of metastasis. Metastases to the liver, bone, and brain are less common. Patients with newly diagnosed testicular GCTs should undergo computed tomography (CT) scan of the abdomen and pelvis. Chest x-ray should be performed. CT scan of the chest is performed if retroperitoneal metastases are present or if lung nodules are identified on chest x-ray. Bone scan and magnetic resonance imaging (MRI) of the brain are not routinely performed unless clinically indicated. Positron emission tomography (PET) has little role in the initial staging of testicular GCTs.

The American Joint Committee on Cancer tumor-node-metastasis (TNM) staging classification is used. There are three main stages of testicular GCT. Stage I is limited to the testis; stage II involves the retroperitoneal lymph nodes; and stage III includes lymph node involvement beyond the retroperitoneum and/or distant metastatic disease.

STAGE-BASED MANAGEMENT

Treatment of testicular GCT is based on two factors: (1) whether the tumor is seminoma or nonseminomatous GCT and (2) the stage of the patient. This is summarized in Fig. 93-1.

Stage I • **SEMINOMA** About 70% of newly diagnosed patients with seminoma present with stage I disease. This is defined as no evidence of metastatic disease on imaging of the chest, abdomen, and pelvis. Approximately 15% of patients with stage I seminoma have metastatic disease at the microscopic level, usually in the retroperitoneum. Historically, patients with stage I seminoma were treated with a course of adjuvant radiation therapy to the paraaortic lymph nodes. While still an option, this is not usually performed because of concerns for late radiation-induced secondary malignancies. Active surveillance is the most common approach elected by these patients following orchiectomy. With active surveillance, interval physical examination and CT scan of the abdomen are performed. For the 15% of patients who develop metastatic disease during active surveillance, treatment is curative in nearly all. A third option for clinical stage I seminoma is adjuvant chemotherapy with carboplatin monotherapy for one or two cycles. While effective in decreasing the risk of recurrence, most patients are cured by orchiectomy alone, and therefore, the additional treatment is unnecessary. In addition, long-term data on toxicity are not available.

NONSEMINOMATOUS GCTS About 40% of newly diagnosed patients with nonseminomatous GCTs present with stage I disease. Because nonseminomatous GCTs have an increased potential for invasion and metastasis, spread to the retroperitoneum and beyond is more common than with seminoma. If pre-orchiectomy serum tumor markers are elevated, these must normalize after orchiectomy to be considered stage I. Patients with persistently elevated or rising serum tumor markers after orchiectomy have stage IS disease and should be treated with cisplatin-based chemotherapy. If the tumor is limited to testis without lymphovascular invasion, the risk of recurrence is approximately 20%. However, if the tumor has high-risk features including lymphovascular invasion, invasion of the spermatic cord, or invasion of the scrotum, the risk of recurrence is ~50% or higher. Historically, a prophylactic retroperitoneal lymph node dissection (RPLND) was performed. This surgery is not only diagnostic but also therapeutic. In fact, most patients who undergo prophylactic RPLND will never require chemotherapy. While still an option, this approach subjects many patients to unnecessary major abdominal surgery. RPLND is also associated with a small risk of retrograde ejaculation due to nerve injury, and nerve-sparing techniques have been developed. Active surveillance is frequently performed especially for patients without lymphovascular invasion. Most patients who relapse will be treated with cisplatin-based chemotherapy and achieve cure rates approaching 100%. Active surveillance can also be employed for patients with higher risk features, although the risk of progression is significantly higher. For this reason, some advocate adjuvant cisplatin-based chemotherapy with BEP (bleomycin, etoposide, cisplatin) for one cycle for these patients. Other centers favor a prophylactic RPLND. Almost all patients who present with stage I nonseminomatous GCTs will achieve cure.

Stage II • **SEMINOMA** Approximately 15–20% of newly diagnosed patients with seminoma present with stage II disease. Patients are subgrouped into IIA, IIB, or IIC based on the size of the retroperitoneal nodes (≤2 cm, >2 to 5 cm, or >5 cm, respectively). Patients with stage IIA disease are usually treated with "dogleg" radiation therapy (referring to the shape of the radiation field), which includes the para-aortic and ipsilateral iliac nodes. Cisplatin-based chemotherapy may also be considered. Stage IIB disease is treated with cisplatin-based chemotherapy or, in select patients, radiation therapy. Most patients treated with radiation therapy who relapse will subsequently be cured with cisplatin-based chemotherapy. RPLND has been considered in select patients with stage IIA and nonbulky stage IIB seminoma in an effort to avoid chemotherapy and radiation therapy. For patients with stage IIC disease, cisplatin-based chemotherapy should be used.

NSGCTS Approximately 15% of newly diagnosed patients with nonseminomatous GCTs present with clinical stage II disease. Patients with stage IIA disease may be treated with primary RPLND. Alternatively, these patients may be treated with cisplatin-based chemotherapy. Patients with stage IIB and IIC disease are best initially managed with cisplatin-based chemotherapy.

Stage III Patients who present with stage III GCT (seminoma or nonseminomatous GCT) are treated with cisplatin-based chemotherapy. These patients are classified into good-, intermediate-, or poor-risk categories using the International Germ Cell Consensus Classification system, which is based on clinical factors including histology, site of primary, the presence of nonpulmonary visceral metastatic disease, and the level of postorchiectomy serum tumor markers (Table 93-1). Most patients with stage III GCT present with good-risk disease and >90% will be cured. The remainder present with intermediate-risk or poor-risk disease associated with 5-year survival rates of ~80% and 50%, respectively. Select patients with rapidly progressive metastatic disease and life-threatening symptoms such as hemoptysis in whom there is a high clinical suspicion of GCT should emergently initiate cisplatin-based chemotherapy, even without a tissue diagnosis.

Chemotherapy The development of cisplatin-based chemotherapy represents an important advance in cancer medicine. Through a series of carefully performed clinical trials with the aim of maximizing cure while minimizing the extent of treatment, the chemotherapy approach to the treatment of these patients has been standardized. Patients with good-risk metastatic GCT are treated with either three cycles of BEP or four cycles of etoposide and cisplatin (EP). Patients with intermediate- and poor-risk metastatic disease are treated with either four cycles of BEP or four cycles of etoposide, ifosfamide, and cisplatin (VIP). Maintaining dose and schedule is important, as dose modifications and delays have been associated with inferior outcomes. Serum tumor markers should be monitored throughout treatment and should normalize during or after treatment. Cisplatin-based chemotherapy is associated with myelosuppression, nausea and vomiting, and alopecia. Cisplatin may result in nephrotoxicity, ototoxicity, and peripheral neuropathy. Bleomycin may result in pulmonary toxicity, and risk factors for this include age >40, renal failure, tobacco use, and the cumulative dose of bleomycin received. For patients at increased risk of bleomycin-induced pneumonitis, non–bleomycin-containing regimens as noted above may be given. Cisplatin-based chemotherapy is also associated with sterility. Approximately 30% of newly diagnosed testicular GCT patients have severe oligospermia or azoospermia. For the remainder with normal baseline spermatogenesis who receive cisplatin-based chemotherapy, all will be azoospermic at the completion of therapy. Approximately 80% of these patients will recover spermatogenesis over a period of several years. For this reason, prechemotherapy sperm banking should be offered to all patients treated with chemotherapy.

Postchemotherapy Surgery Upon completion of cisplatin-based chemotherapy, many patients with normalized serum tumor markers will have radiographic evidence of residual masses. In approximately half of patients with nonseminomatous GCT, the residual mass is composed of necrosis and/or fibrosis. About 40% will have residual teratoma, and only 10% will have residual viable nonteratomatous GCT. Unfortunately, radiographic imaging cannot accurately differentiate between these entities. For this reason, nonseminomatous GCT patients with residual masses after chemotherapy undergo resection of all sites of disease. This most commonly includes a postchemotherapy RPLND. However, thoracotomy and neck dissection are required in some patients. Given the complexity of this surgery, patients should be referred to highly experienced centers. If the patients are found to have residual necrosis or teratoma, no additional therapy is required.

Stage 1
Testis

	Seminoma	NSGCT
Stage IA Testis only, no lymphovascular invasion	Active surveillance; or Adjuvant carboplatin × 1 or 2 cycles; or Adjuvant para-aortic RT	Active surveillance; or Nerve-sparing RPLND; or Adjuvant BEP × 1 cycle
Stage IB Testis only, with lymphovascular invasion or invasion of spermatic cord or scrotum	Active surveillance; or Adjuvant carboplatin × 1 or 2 cycles; or Adjuvant para-aortic RT	Active surveillance; or Adjuvant BEP × 1 cycle; or Nerve-sparing RPLND
Stage IS Elevated serum tumor markers post-orchiectomy	BEP × 3 cycles; or EP × 4 cycles	BEP × 3 cycles; or EP × 4 cycles

A

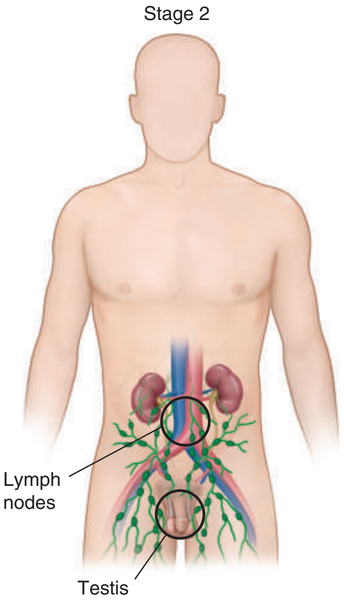

	Seminoma	NSGCT
Stage IIA N1: nodes ≤ 2 cm	Para-aortic and ipsilateral iliac RT; or BEP × 3 cycles or EP × 4 cycles; Nerve-sparing RPLND in select patients	Nerve-sparing RPLND; or BEP × 3 cycles or EP × 4 cycles
Stage IIB N2: nodes > 2 to 5 cm	BEP × 3 cycles or EP × 4 cycles; or Para-aortic and ipsilateral iliac RT	BEP × 3 cycles or EP × 4 cycles +/– postchemotherapy RPLND
Stage IIC N3: nodes > 5 cm	BEP × 3 cycles or EP × 4 cycles	BEP × 3 cycles or EP × 4 cycles +/– postchemotherapy RPLND

B

FIGURE 93-1 Stage-based management of testicular germ cell tumor.

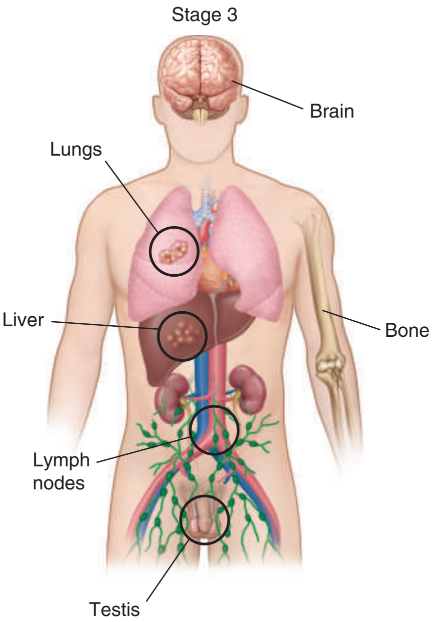

	Seminoma	NSGCT
Stage IIIA (good-risk)	BEP × 3 cycles; or EP × 4 cycles	BEP × 3 cycles; or EP × 4 cycles; +/– Postchemotherapy surgery
Stage IIIB (intermediate-risk)	BEP × 4 cycles; or VIP × 4 cycles	BEP × 4 cycles; or VIP × 4 cycles +/– Postchemotherapy surgery
Stage IIIC (poor-risk)	N/A	BEP × 4 cycles; or VIP × 4 cycles +/– Postchemotherapy surgery

Abbreviations: BEP, bleomycin, etoposide, cisplatin; EP, etoposide, cisplatin; N/A, not applicable; NSGCT, nonseminomatous germ cell tumor; RPLND, retroperitoneal lymph node dissection; RT, radiation therapy; VIP, etoposide, ifosfamide, cisplatin.

C

FIGURE 93-1 (Continued)

TABLE 93-1 International Germ Cell Consensus Classification System

RISK GROUP	SEMINOMA	NSGCT
Good	Any primary site; and normal AFP, any hCG, any LDH; and nonpulmonary visceral metastases absent	Gonadal or retroperitoneal primary; and nonpulmonary visceral metastases absent; and AFP <1000 ng/mL; and hCG <5000 mIU/mL; and LDH <1.5 × ULN
Intermediate	Any primary site; and normal AFP, any hCG, any LDH; and nonpulmonary visceral metastases present	Gonadal or retroperitoneal primary; and nonpulmonary visceral metastases absent; and one of the following: AFP 1000–10,000 ng/mL HCG 5000–50,000 mIU/mL LDH 1.5–10 × ULN
Poor	N/A	Mediastinal primary; or nonpulmonary visceral metastases present; or one of the following: AFP >10,000 ng/mL HCG >50,000 mIU/mL LDH >10 × ULN

Abbreviations: AFP, α-fetoprotein; hCG, human chorionic gonadotropin; LDH, lactate dehydrogenase; N/A, not applicable; NSGCT, nonseminomatous germ cell tumor; ULN, upper limit normal. Nonpulmonary visceral metastases include liver, bone, and brain.

Source: Reproduced with permission from International Germ Cell Cancer Collaborative Group: International Germ-Cell Consensus Classification: A prognostic factor based staging system for metastatic germ cell tumors. J Clin Oncol 15:594, 1997.

However, for patients with residual viable nonteratomatous GCT, two additional cycles of chemotherapy may be considered. It should be noted that in most centers, patients with minimal residual tumors defined as retroperitoneal lymph nodes of ≤1 cm forego postchemotherapy RPLND. Patients who experience normalization of serum tumor markers with first-line chemotherapy but have enlarging tumors, most often cystic masses in the retroperitoneum, may have "growing teratoma syndrome." These patients are best approached with surgery.

For patients with metastatic seminoma, most residual masses are necrotic and do not harbor viable tumor. Patients with residual masses of 3 cm or less may be observed without surgery. For patients with residual masses >3 cm, fluorodeoxyglucose (FDG)-PET may be used to distinguish necrosis from viable seminoma and identify patients who should be considered for postchemotherapy surgery or short interval imaging.

RELAPSED DISEASE

Approximately 20–30% of patients with metastatic GCTs treated with cisplatin-based chemotherapy will not achieve durable disease control. Most of these patients will experience disease progression within 2 years following completion of chemotherapy. The International Prognostic Factors Study Group developed a risk stratification classification system for patients in first relapse. Contributors to a worsened prognosis include NSGCT histology, extragonadal primary, incomplete response to first-line chemotherapy, time to relapse of 3 months or less, level of serum tumor markers at relapse, and the presence of nonpulmonary visceral metastatic disease.

Patients in first relapse may be treated with either conventional-dose salvage chemotherapy or high-dose salvage chemotherapy with autologous stem cell rescue. There is controversy concerning which approach is optimal. Some institutions advocate for risk stratification, with more favorable prognosis patients receiving conventional-dose chemotherapy and worse prognosis patients receiving high-dose chemotherapy. The most commonly utilized conventional-dose regimen includes paclitaxel, ifosfamide, and cisplatin (TIP). High-dose chemotherapy consists of initial salvage therapy followed by stem cell harvest and then two or three cycles of high-dose carboplatin and etoposide (CE) with stem cell rescue. A large retrospective analysis has compared conventional-dose salvage chemotherapy to high-dose salvage chemotherapy in patients in first relapse. This study reports a more favorable outcome with high-dose salvage chemotherapy across nearly all risk groups. However, given the retrospective nature of this study and the controversy concerning optimal approaches, an international randomized trial comparing conventional-dose chemotherapy (TIP) to high-dose chemotherapy with autologous stem cell rescue (TI-CE) has completed accrual and results are forthcoming.

Some patients who experience disease progression after conventional-dose salvage chemotherapy may successfully be treated with high-dose salvage chemotherapy with autologous stem cell rescue. Patients with disease progression after high-dose salvage chemotherapy may be treated with subsequent chemotherapy regimens that include gemcitabine/oxaliplatin, gemcitabine/paclitaxel, epirubicin/cisplatin, and oral etoposide. While these patients may benefit from third-line chemotherapy, few will achieve durable disease control. Select patients with relapsed but resectable disease may be candidates for salvage or so-called "desperation" surgery. Studies of molecularly targeted agents and immune checkpoint inhibitors in this population have to date been generally disappointing.

Patients who experience disease progression >2 years after chemotherapy are considered to have "late relapse." Late relapse appears to have a different biology than early relapse. These patients tend to have more chemotherapy-resistant disease. Patients with late relapse usually have nonseminomatous GCT with elevation of serum AFP. Many of these patients experience recurrence in the retroperitoneum many years after first-line chemotherapy, and this likely represents residual retroperitoneal disease that was not controlled after first-line therapy. These patients are best approached with salvage surgery.

EXTRAGONADAL GERM CELL TUMORS

Approximately 5% of patients who present with GCTs have extragonadal primaries. These mainly originate in the mediastinum or retroperitoneum. Patients suspected of extragonadal GCT should undergo scrotal ultrasound to exclude a gonadal primary. Extragonadal seminomas have a similar excellent prognosis as their gonadal counterparts and are approached the same. Mediastinal nonseminomatous GCTs are classified as poor risk and are treated with either four cycles of BEP or four cycles of VIP. These patients frequently require postchemotherapy thoracic surgery for residual disease. For this reason, some advocate avoiding bleomycin in this patient population. Klinefelter's syndrome is associated with an increased risk of mediastinal nonseminomatous GCTs. Rarely, mediastinal nonseminomatous GCTs are associated with hematologic disorders including acute myeloid leukemia. Nonseminomatous GCTs arising in the retroperitoneum do not have a worse prognosis than their gonadal counterparts. Many patients who present with extragonadal GCTs will undergo core needle biopsy for diagnosis. However, select patients with extragonadal tumors and definitive elevation of serum tumor markers may initiate chemotherapy without a tissue diagnosis.

Cancers of unknown primary are defined as histologically proven metastatic malignancy in which the primary site is not obvious. A subgroup of patients with cancer of unknown primary have occult GCTs. Male gender, age <65 years, midline tumors, and nonsmoking status increase the likelihood of this presentation. Pathology may demonstrate a poorly differentiated malignant neoplasm. Immunohistochemical staining is used to exclude lymphoma. Tumor may be analyzed by fluorescence in situ hybridization for i(12p), which confirms the diagnosis. Even if the diagnosis is not certain, patients should be treated with cisplatin-based chemotherapy, which will cure up to 20% of this patient group.

TESTICULAR NON–GERM CELL TUMORS

Rarely, patients may develop testicular non-GCTs. These include non-Hodgkin's lymphoma, most commonly occurring in men over the age of 50; sex cord stromal tumors including Leydig cell tumors and Sertoli cell tumors; mesothelioma of the tunica vaginalis; and paratesticular sarcoma. Metastasis to the testis is rare, most commonly occurring in patients with advanced prostate cancer and melanoma.

SURVIVORSHIP AND LATE EFFECTS

Because most patients with testicular GCT will experience long-term survival, survivorship care is important. Since primary care physicians will follow many of these patients, an understanding of the physical, psychological, and social late effects is important. Late effects are defined as health problems that occur months or years after a disease is diagnosed or after treatment has ended. Late effects may be related to the underlying cancer or to the treatment the patient received. In long-term survivors of testicular GCT, increased cardiovascular risk and increased secondary malignancies have been reported. Patients treated with cisplatin-based chemotherapy have an increased risk of hypertension, hyperlipidemia, metabolic syndrome, and cardiovascular events. Patients treated with high cumulative doses of etoposide (e.g., patients who receive standard chemotherapy, relapse, and then receive salvage high-dose chemotherapy) may experience up to a 1–2% risk of developing acute myeloid leukemia, typically 2–3 years after completing therapy and associated with an 11q23 translocation. Patients treated with radiation therapy, cisplatin-based chemotherapy, or both have an increased risk of developing secondary solid malignancies.

FURTHER READING

King J et al: Testicular cancer: Biology to bedside. Cancer Res 81:5369, 2021.

Lobo J et al: Molecular biomarkers with potential clinical application in testicular cancer. Mod Pathol 36:100307, 2023.

Pluta J et al: Identification of 22 novel susceptibility loci associated with testicular germ cell tumors. Nat Commun 12:4487, 2021.

Travis LB et al: Adolescent and young adult germ cell tumors: Epidemiology, genomics, treatment, and survivorship. J Clin Oncol 42:696-706, 2024.

94

Gynecologic Malignancies

David Spriggs

OVARIAN CANCER

INCIDENCE AND PATHOLOGY

Ovarian cancer remains a leading cause of cancer deaths in American women, ranking behind lung, breast, colon, and pancreatic cancers. The ovary is responsible for hormone production and egg production, including maturation and ovulation with the supporting cyclical production of sex steroid hormones. These complex biologic functions are linked to populations of stromal cells, ovarian germ cells, and the enveloping epithelial cells. Malignancies arising from each group include multiple histologic variants, each with unique neoplastic behaviors. Epithelial tumors are, by far, the most common histologic variant of ovarian neoplasms; they may be benign (50%), frankly malignant (33%), or of borderline malignancy (low malignant potential) (16%). In adnexal masses detected by imaging or physical exam, age influences risk of malignancy; tumors in younger women are more likely benign. In the malignant group, the most common tumors are epithelial. In the group of the ovarian epithelial malignancies are the serous tumors (60–70%), mucinous tumors (10%), endometrioid tumors (10–15%), and clear cell tumors (10–15%) tumors. The distribution of histologic types varies in different parts of the world. The less common stromal tumors arise from the ancillary, supportive cells such as steroid hormone–producing cells and likewise have different phenotypes and clinical presentations. Most stromal tumors do not produce estrogen, but ectopic hormone production can be seen in certain subtypes. Tumors arising in the ovarian germ cell lineage are generally similar in biology and behavior to testicular tumors in males, although their intraperitoneal location alters some metastatic behaviors **(Chap. 93)**. Ovarian tissue may also host metastatic epithelial tumors arising from breast, colon, gastric, and pancreatic primaries. Bilateral ovarian masses from metastatic mucin-secreting gastrointestinal cancers are termed *Krukenberg tumors*. Consideration of other potential malignancies is part of the diagnostic workup of ovarian masses.

OVARIAN CANCER OF EPITHELIAL ORIGIN

Epidemiology An American woman has approximately a 1 in 72 lifetime risk (1.6%) of developing ovarian cancer, with the majority of affected women developing epithelial tumors. In 2024 in the United States, ~19,710 cases of ovarian cancer are expected to be diagnosed, with >13,270 deaths. Sporadic (not familial) epithelial tumors of the ovary have a peak incidence in women in their fifties and sixties, although age at presentation ranges from the third decade to the eighties and nineties. Ovarian cancer risk has been linked to an interactive mixture of epidemiologic, environmental, and genetic factors. Nulliparity, obesity, diet, infertility treatments, talc exposure, and possibly hormone replacement therapy have all been linked to an increase in risk. Protective factors include the use of oral contraceptives, multiparity, tubal ligation, aspirin use, and breast-feeding. Other epidemiologic factors such as the historical use of perineal talc agents remain controversial. The mechanisms underlying the various protective factors are largely unknown, but it is increasingly clear that dysplasia and in situ cancers are seen within the fallopian tube and probably are the original site for a large percentage of cancers.

Genetics and Pathogenesis Ovarian cancers are divided into type 1 cancers and the more aggressive type 2 variant. The type 1 cancers are characterized by low-grade histology and generally indolent behavior. These tumors include the low malignant potential tumors, low-grade endometrial and mucinous histologies, and clear cell cancers (which are more aggressive). Genetic alterations in type 1 cancers include mutations in *KRAS*, *BRAF*, *PTEN*, and *PIK3CA*. In contrast, type 2, high-grade serous epithelial ovarian cancers show serial genetic changes in the fallopian tube with loss of *BRCA1/2* function and *TP53* mutation leading to intraepithelial cancer in the luminal epithelium. Following these early genetic events, additional mutations in these transformed cells lead to tumor cell shedding, metastasis, and invasion. These type 2, poorly differentiated, serous cancer cells can then spread to the ovaries and the peritoneal cavity, aided by the ovarian cancer cell's expression of MUC16 and binding to mesothelin-expressing cells.

Genetically, type 2 serous ovarian cancer is classically a disease characterized by loss of *TP53* (95%) and *BRCA1/2* function in nearly all cases. Widespread amplifications and deletions rather than single-gene point mutations or common gene fusions are also present. Low prevalence but statistically recurrent somatic mutations in seven other genes including *NF1*, *RB1*, and *CDK12* were also seen. The most common heritable abnormality linked to ovarian cancer is a germline mutation in either *BRCA1* (chromosome 17q12–21) or *BRCA2* (chromosome 13q12–13). These genes are essential parts of the homologous DNA repair machinery for double-stranded DNA break repair. Individuals inheriting a single copy of a mutant allele have an increased lifetime risk of breast (46–87% for *BRCA1*; 38–84% for *BRCA2*) and ovarian cancer (39–63% for *BRCA1*; 16.5–27% for *BRCA2*). Many of these women have a family history that includes multiple cases of breast and/or ovarian cancer at an early age. Male breast cancer, pancreatic cancer, and prostate cancer are also linked to familial *BRCA2* mutations. The most common malignancy in women carrying germline *BRCA1/2* mutations is breast carcinoma, although women harboring germline *BRCA1* mutations also have a marked increased risk of developing ovarian malignancies in their forties and fifties. Women harboring a mutation in *BRCA2* have a lower penetrance of ovarian cancer with onset typically in their fifties or sixties. Other uncommon germline mutations of other genes encoding proteins linked to homologous DNA repair (e.g., *PALB2*) can also contribute to cancer risk, although the frequency of mutation and magnitude of risk increment are much lower. Germline *BRCA1/2* testing is recommended for *all* incident epithelial ovarian cancers to detect probands to identify relatives for early therapeutic intervention. Women with these high-risk germline mutations are advised to undergo prophylactic removal of fallopian tubes and ovaries after completing childbearing, ideally before age 40. Early prophylactic salpingo-oophorectomy is highly protective. Salpingo-oophorectomy also appears to protect these women from subsequent breast cancer (risk reduction 50%). Prophylactic salpingectomy is almost certainly the key part of any surgical prophylaxis strategy for ovarian cancer prevention. Although less common, women with type II Lynch syndrome caused by mutations in one of the DNA mismatch repair genes (*MSH2, MLH1, MLH6, PMS1, PMS2*) are at risk for ovarian and endometrial cancer. Like *BRCA1/2*-related cancers, these cancers develop earlier than sporadic ovarian cancer.

Neoplasms of the ovary tend to be painless unless they undergo torsion. Nonspecific gastrointestinal symptoms like bloating and early satiety are common at presentation, probably related to compression of local organs or due to symptoms from metastatic disease. Women with ovarian tumors also may have an increased incidence of symptoms including pelvic discomfort, bloating, and perhaps changes in urinary or bowel pattern. Unfortunately, these same symptoms are common in primary care and are frequently dismissed by either the woman or her health care team until later stages of disease. The pathogenic factors and timing of spread beyond the ovary are still not well understood. The most common symptoms at presentation of advanced disease include a period of progressive complaints of nausea, early satiety, bloating, indigestion, constipation, and abdominal pain. Signs include the rapid increase in abdominal girth due to the accumulation of ascites that typically alerts the patient and her physician that the concurrent gastrointestinal symptoms are likely associated with malignant pathology. Radiologic evaluation typically demonstrates a complex adnexal mass with ascites, carcinomatosis, and pelvic, para-aortic, and mesenteric adenopathy in advanced disease. Positron emission tomography (PET) scans are generally not required. Laboratory evaluation often demonstrates a markedly elevated serum CA-125, the shed mucin component (MUC16) associated with, but not specific for, ovarian cancer. Ovarian cancers are divided into four stages, with stage I tumors confined to the ovary, stage II malignancies confined to the pelvis, and stage III confined to the peritoneal cavity and retroperitoneal nodes **(Table 94-1)**. These three stages are subdivided, with the most common presentation, stage IIIC, defined as tumors with bulky intraperitoneal disease or positive lymph node involvement. About 70% of women present with stage III disease. Stage IV disease includes women with parenchymal metastases (liver, lung, spleen) or, alternatively, abdominal wall or pleural disease. The 30% of patients not presenting with stage III disease are roughly evenly distributed among the other stages.

Screening Advanced ovarian cancer is a highly lethal condition. It is curable in early stages but seldom curable in advanced stages; hence, screening continues to be of considerable interest. Early-stage tumors often secrete excessive amounts of normal proteins that can be measured in the serum such as CA-125, mesothelin, and HE-4. Nevertheless, the incidence of ovarian cancer in the middle-aged female population is very low, with only ~1 in 2000 women between the ages of 50 and 60 carrying an asymptomatic and undetected tumor. Large, well-designed screening studies, even in the mutated *BRCA1/2* families, have thus far failed to decrease ovarian cancer mortality in prospective testing. Circulating DNA approaches have also been unsuccessful so far. Screening for ovarian cancer is currently not recommended outside

TABLE 94-1 Staging and Survival in Gynecologic Malignancies

STAGE	OVARIAN	5-YEAR SURVIVAL, %	ENDOMETRIAL	5-YEAR SURVIVAL, %	CERVIX	5-YEAR SURVIVAL, %
0	—		—		Carcinoma in situ	100
I	Confined to ovary	88–95	Confined to corpus	>90	Confined to uterus	85
II	Confined to pelvic organs	70–80	Involves corpus and cervix	~75	Invades beyond uterus but not to pelvic wall	65
III	Intra-abdominal spread to omentum, diaphragm, or lymph nodes	20–40	Extends outside the uterus but not outside the true pelvis	45–60	Extends to pelvic wall and/or lower third of vagina, or hydronephrosis	35
IV	Spread outside abdominal cavity, parenchymal spread, and pleural effusion cytology	10–20	Extends outside the true pelvis or involves the bladder or rectum	~20	Invades mucosa of bladder or rectum or extends beyond the true pelvis	7

of a clinical trial, but a careful history for familial cancers and directed genetic testing for susceptibility genes are definitely appropriate.

TREATMENT

Ovarian Cancer

Epithelial ovarian cancer can be divided into distinct "disease states" for the purpose of treatment selection, as shown in **Fig. 94-1**. Detection by ultrasonography generally can identify a complex ovarian mass as "suspicious," but surgery by a skilled gynecologic oncologist remains the preferred initial diagnostic and therapeutic option for an isolated adnexal mass or a more involved picture of peritoneal involvement. The amount of residual visible cancer at the end of a primary operation is strongly predictive of outcome and is paired with histology, grade, and stage to determine prognosis and treatment. Metastatic disease to the ovary can be seen from primary tumors of the colon, appendix, stomach (Krukenberg tumors), and breast. Needle biopsy of adnexal masses is contraindicated to avoid malignant contamination of the peritoneal cavity with malignant cells. Typically, women undergo laparoscopic evaluation and unilateral salpingo-oophorectomy for diagnostic purposes. If pathology reveals a primary ovarian malignancy or the laparoscopy proves disseminated disease is present, then the procedure should be followed by a total hysterectomy, removal of the remaining tube and ovary, omentectomy, and pelvic node sampling along with biopsies of the peritoneal cavity and diaphragms. This extensive surgical procedure is performed because ~30% of tumors that, by visual inspection, appear to be confined to the ovary have already disseminated to the peritoneal cavity and/or surrounding lymph nodes. As with axillary dissections in breast cancer, node sampling is diagnostic, but full lymphadenectomy appears to provide little or no additional therapeutic advantage over nodal sampling. The target outcome of an ovarian cancer surgery is always an R0 resection (no visible residual cancer). The less favorable "optimal resection" (no residual disease >1 cm in size) is still clinically useful, and the prognosis of those patients is much better than that of patients who are left with >1 cm of disease at the end of surgery. These "suboptimally debulked" patients derive very little benefit from their surgery. If large deposits of unresectable residual tumor are anticipated, the surgery should be delayed until after several cycles of neoadjuvant chemotherapy. Such "interval debulking" surgery

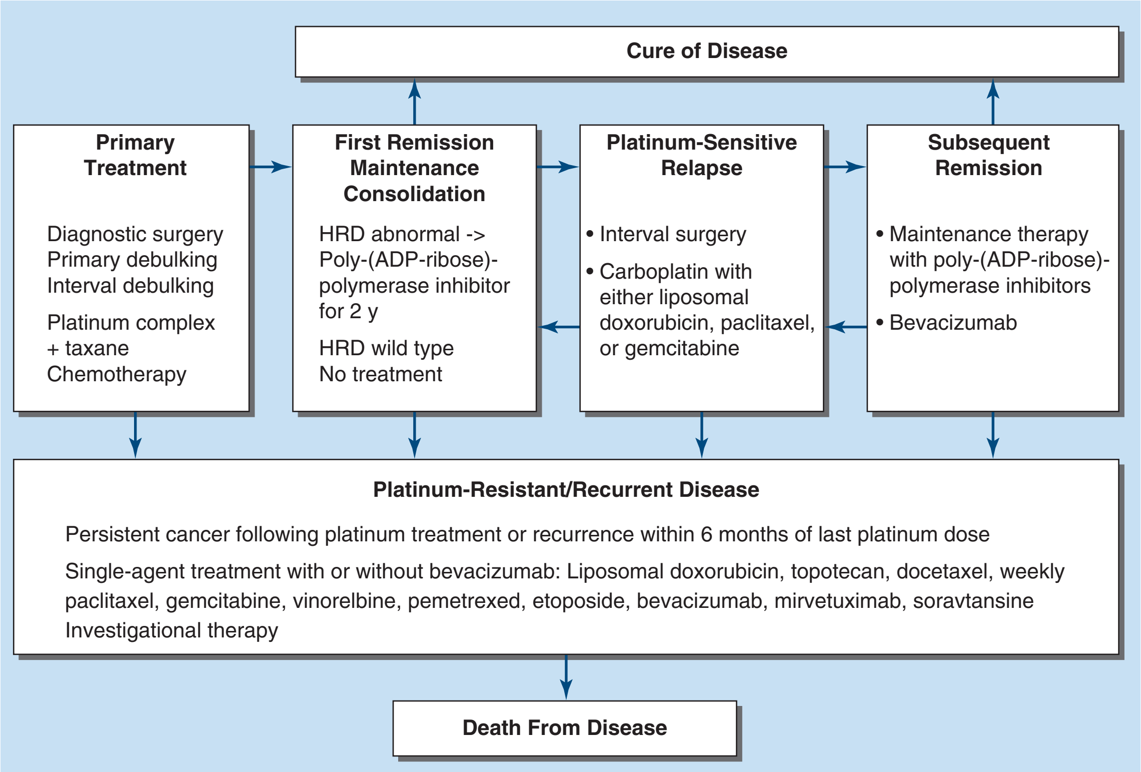
Cure of Disease
Primary Treatment
Diagnostic surgery
Primary debulking
Interval debulking
Platinum complex + taxane Chemotherapy
First Remission Maintenance Consolidation
HRD abnormal -> Poly-(ADP-ribose)-polymerase inhibitor for 2 y
HRD wild type No treatment
Platinum-Sensitive Relapse
• Interval surgery
• Carboplatin with either liposomal doxorubicin, paclitaxel, or gemcitabine
Subsequent Remission
• Maintenance therapy with poly-(ADP-ribose)-polymerase inhibitors
• Bevacizumab
Platinum-Resistant/Recurrent Disease
Persistent cancer following platinum treatment or recurrence within 6 months of last platinum dose
Single-agent treatment with or without bevacizumab: Liposomal doxorubicin, topotecan, docetaxel, weekly paclitaxel, gemcitabine, vinorelbine, pemetrexed, etoposide, bevacizumab, mirvetuximab, soravtansine
Investigational therapy
Death From Disease

FIGURE 94-1 Disease states model of epithelial ovarian cancer and its treatment. Each box represents a relatively homogenous group of patients who share a palette of potential treatment choices and have a similar prognosis. The *arrows* indicate that a single patient may move from one state to another during the course of her illness, and the choice of treatments will become different in her new disease state. HRD, homologous recombination deficiency.

achieves similar results to primary surgery with diminished surgical morbidity and more timely chemotherapy. Patients without gross residual disease (R0 resection) after resection have a median survival in excess of 60 months, compared to 28–42 months for those left with macroscopic tumor or those undergoing interval debulking, regardless of treatment strategy.

After appropriate surgical treatment, primary chemotherapy will consist of combination treatment with paclitaxel and carboplatin. Primary chemotherapy can be delivered intravenously, or alternatively, some therapy can be directly administered into the peritoneal cavity via an indwelling catheter. The intraperitoneal approach is technically more difficult and is increasingly replaced by carboplatin and paclitaxel, which appears to offer similar results. Although interest in immunotherapy or chemoimmunotherapy has been high, immunotherapeutics have not yet improved primary chemotherapy.

With optimal debulking surgery and platinum-based chemotherapy, 70% of women who present with advanced-stage tumors show tumor reduction, and 40–50% experience a complete remission with normalization of their CA-125, computed tomography (CT) scans, and physical examination. For women with evidence of functional homologous DNA repair defects, administration of oral poly-ADP ribose polymerase inhibitors (PARPi) such as niraparib, olaparib, or rucaparib will improve survival outcomes when administered at the completion of intravenous chemotherapy as consolidation. These drugs substantially delay recurrence and provide survival advantages as well. In the majority of patients, disease still recurs within 1–4 years from the completion of their primary therapy. CA-125 levels often increase as a first sign of relapse, and CT scan findings are eventually confirmatory. Recurrent disease is often successfully managed for years but rarely cured, despite a growing panel of chemotherapeutic agents and antibody-drug conjugates. Additional surgical therapy does not appear to extend survival in randomized trials. Patients with a treatment-free interval are often best treated with additional platinum doublets, combining carboplatin with liposomal doxorubicin, gemcitabine, or a taxane. Eventually all women who experience relapse develop chemotherapy-refractory disease. Refractory ascites, poor bowel motility, and obstruction or tumor-infiltrated aperistaltic bowel are all common premorbid events. Limited surgery to relieve intestinal obstruction, localized radiation therapy to relieve pressure or pain from masses, or palliative chemotherapy may be helpful. Agents with >15% response rates include gemcitabine, topotecan, liposomal doxorubicin, and bevacizumab. Five-year survival correlates with the stage of disease: stage I, 90–95%; stage II, 70–80%; stage III, 25–40%; stage IV, 10–15% (Table 94-1). Prognosis is improved by lower histologic grade and presence of *BRCA1* or *BRCA2* germline mutation.

■ UNCOMMON OVARIAN TUMORS

Low Malignant Potential Tumors (Borderline Tumors) These type 1 tumors are found in younger women (age 30–50 years) and are indolent in behavior, and few of these patients will succumb to their tumors (10-year survival may approach 98%), although recurrence is not uncommon. Certain features, such as micropapillary histology and microinvasion, are linked to more aggressive behavior. Tumors of low malignant potential have different mutations including mutations *BRAF* or *KRAS* and cyclin-dependent kinase inhibitor (*CDKN*) 2A/2B deletion. Borderline tumor patients are managed primarily by surgery, but targeted therapy for the RAS/RAF pathway and hormonal treatments sometimes have benefit.

Stromal Tumors Approximately 7% of ovarian neoplasms are stromal tumors, with ~1800 cases expected each year in the United States. Ovarian stromal tumors or sex cord tumors are most common in women in their fifties or sixties, but tumors can present at any age. These tumors arise from the mesenchymal components of the ovary, including both steroid-producing cells and fibroblasts. Most of these tumors are indolent tumors with limited metastatic potential and present as unilateral solid masses. These tumors primarily are discovered by the detection of an abdominal mass, sometimes with abdominal pain due to ovarian torsion, intratumoral hemorrhage, or rupture. Rarely, stromal tumors can produce estrogen and present with breast tenderness as well as precocious puberty in children, menstrual disturbances in reproductively active women, or postmenopausal bleeding. In some women, estrogen-associated secondary malignancies, such as endometrial or breast cancer, may present as synchronous malignancies. Sertoli-Leydig tumors often present with hirsutism and virilization due to increased production of androgens. Hormonally inert tumors include fibromas, which present as solitary masses often in association with ascites and occasionally hydrothorax, also known as Meigs's syndrome. A subset of these tumors presents in individuals with a variety of inherited disorders that predispose them to mesenchymal neoplasia including Ollier's disease (juvenile granulosa cell tumors) and Peutz-Jeghers syndrome (ovarian sex cord tumors). The treatment of these tumors is primarily complete surgical resection, without adjuvant chemotherapy. Chemotherapy with carboplatin and paclitaxel is generally reserved for either unresectable or multiply recurrent tumors.

Germ Cell Tumors of the Ovary Germ cell tumors, like their counterparts in the testis, are cancers of germ cells. These totipotent cells contain the programming for differentiation to essentially all tissue types, and hence, the germ cell tumors include a histologic menagerie of bizarre tumors, including benign teratomas (dermoid cysts) and a variety of malignant tumors, such as dysgerminoma, immature teratomas, yolk sac malignancies, and choriocarcinomas. Benign teratoma (or dermoid cyst) is the most common germ cell neoplasm of the ovary and often presents in young women. These tumors include a complex mixture of differentiated tissue including tissues from all three germ layers. In older women, these differentiated tumors can develop malignant transformation, most commonly squamous cell carcinomas. Malignant germ cell tumors include dysgerminomas, yolk sac tumors, immature teratomas, and embryonal and choriocarcinomas. Germ cell tumors can present at all ages, but the peak age of presentation tends to be in adolescents. Typically, these tumors will become large ovarian masses, which eventually present as palpable low abdominal or pelvic masses. Like sex cord tumors, torsion or hemorrhage may present urgently or emergently as acute abdominal pain. Some germ cell tumors produce elevated levels of human chorionic gonadotropin (hCG) or α-fetoprotein (AFP). Unlike epithelial ovarian cancer, these tumors have a higher proclivity for nodal or hematogenous metastases. Germ cell tumors typically present in women who are of childbearing age, and because bilateral tumors are uncommon (except in dysgerminoma, 10–15%), the typical treatment is unilateral oophorectomy or salpingo-oophorectomy with lymph node sampling. Most commonly, women with advanced malignant germ cell tumors typically receive bleomycin, etoposide, and cisplatin (BEP) chemotherapy, in an analogous fashion to the treatment of testicular cancers. In the majority of these women, even those with advanced-stage disease, cure is expected. Dysgerminoma is the ovarian counterpart of testicular seminoma and is highly curable. Although the tumor is highly radiation-sensitive, radiation produces infertility in many patients. BEP chemotherapy is as effective or more so without causing infertility.

FALLOPIAN TUBE CANCER

Transport of the egg to the uterus occurs through the fallopian tube, with the distal ends of these tubes composed of fimbriae that drape about the ovarian surface and capture the egg as it erupts from the ovarian cortex. As described above, the majority of type 2 ovarian cancers are now thought to arise from the tubal epithelium. Fallopian tube malignancies are typically of serous histology and share the same biology and recommended treatment approaches as serous ovarian cancer. These tumors often present as clinically isolated adnexal masses, but like ovarian cancer, these tumors spread relatively early throughout the peritoneal cavity. Fallopian tubal cancers have a natural history and treatment that are essentially identical to ovarian cancer (Table 94-1).

CERVICAL CANCER

ETIOLOGY AND GENETICS

Cervical cancer is the second most common and the most lethal malignancy in women worldwide. Infection with high-risk strains of human papillomavirus (HPV) is the primary neoplastic-initiating event in the vast majority of women with invasive cervical cancer. This double-stranded DNA virus infects epithelium near the transformation zone of the cervix where underlying columnar epithelium becomes squamous epithelium. More than 60 types of HPV are known, with ~20 types having the ability to generate high-grade dysplasia and malignancy. HPV16 and 18 are the types most frequently associated with high-grade dysplasia, but types 31, 33, 35, 52, and 58 are also considered to be high-risk variants. The large majority of sexually active adults are exposed to HPV, and most women clear the infection without specific intervention. The 8-kb HPV genome encodes seven early genes, most notably *E6* and *E7*, which can bind to *RB* and *p53*, respectively. High-risk types of HPV encode *E6* and *E7* molecules that are particularly effective at inhibiting the normal cell cycle checkpoint functions of these regulatory proteins, leading to immortalization but not full transformation of cervical epithelium. A minority of women will fail to clear the infection, with subsequent HPV integration into the host genome. Over as little as a few months to several years, some of these persistently infected women develop worsening dysplasia, a premalignant condition that, untreated, can progress to cervical carcinoma. Complete transformation to cancer occurs over a period of years and almost certainly requires the acquisition of other genetic mutations within the infected and immortalized epithelium.

In 2024, more than half a million new cases of cervical cancer will occur worldwide, with an estimated >300,000 deaths. Cancer incidence is particularly high in women residing in Central and South America, the Caribbean, and southern and eastern Africa. The mortality rate is disproportionately high in Africa. In the United States, an estimated 13,960 women will be diagnosed with cervical cancer in 2024 and ~4300 women will die of the disease.

In the integrated genomic characterization of cervical cancer by The Cancer Genome Atlas (TCGA), integration of HPV sequences was found in all of the HPV18-linked cancers and over three-quarters of the HPV16 cancers. The cervical tumors also showed a characteristic APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like; a family of cytidine deaminases that edit DNA and are endogenous mutagenic enzymes) pattern of mutagenesis, with *ERBB3, CASP8*, and *TGFRB2* identified as significantly mutated genes presumably linked to progression from dysplasia to carcinoma. In the much smaller number of HPV-negative cancers, which are more common in older women, mutations in oncogenes *KRAS, ARID1A,* and *PTEN* were frequently seen. The clinical behavior of these cancers is likely to be different.

HPV INFECTION AND PREVENTION

The Pap smear is the primary detection method for asymptomatic preinvasive cervical dysplasia of squamous epithelial lining during a gynecologic exam. Because the progression from dysplasia to cervical cancer takes several years, annual (or longer interval) screening and prevention strategies that detect precancerous dysplasia and carcinoma in situ can be implemented successfully. Annual or biannual cervical scraping for cytology (Pap smear) is highly effective in reducing the incidence of cervical cancer by early detection and subsequent surgical treatment of premalignant disease. The incorporation of HPV testing by polymerase chain reaction (PCR) or other molecular techniques increases the sensitivity of detecting cervical pathology but at the cost of lower sensitivity in that it identifies many women with transient infections who require no specific medical intervention. Unfortunately, both the collection of a Pap smear and its cytologic evaluation require infrastructure beyond the means of many middle- and low-income countries. High-throughput, low-technology prevention strategies and point-of-care testing are needed to identify and treat women bearing high-risk cervical dysplasia to prevent cancer development.

A primary prevention strategy relies on HPV vaccines. In the United States, the Gardasil-9 vaccine protects against HPV types 6, 11, 16, 18, 31, 33, 45, 52, and 58. Vaccination of girls and women between ages 9 and 45 years is recommended with three injections (0, 2, and 6 months). Vaccination before the initiation of sexual activity dramatically reduces the rate of high-risk HPV infection and subsequent dysplasia. Vaccination of both boys and girls is increasingly considered to reduce the risk of HPV-induced cancers of the pharynx. Vaccinated women are still at risk for HPV infection and still benefit from standard Pap smear screening.

CLINICAL PRESENTATIONS

Risk Factors Clinical risk factors include the prevalence of high-risk HPV subtypes in the population and HPV infection–linked features such as a high number of sexual partners, early age of first intercourse, and history of venereal disease. Smoking is a cofactor; heavy smokers have a higher risk of dysplasia with HPV infection. HIV infection, especially when associated with low CD4+ T-cell counts, is associated with a higher rate of high-grade dysplasia and likely a shorter latency period between infection and invasive disease. Histologically, the majority (80%) of cervical malignancies are squamous cell carcinomas associated with HPV, but adenocarcinomas are also HPV related, and both arise in the transitional zone of the endocervical canal; the lesions in the canal or cervical glands may not be seen by visual inspection of the cervix and can be missed by Pap smear screening. Other malignancies, such as vulvar cancer, anal cancer, and, increasingly, pharyngeal cancer, are also linked to HPV infection.

Diagnosis of Cervical Cancer Early cancer of the cervix is asymptomatic, and this biology underlies the recommendations for routine gynecologic care. Larger, invasive carcinomas often have symptoms or signs including postcoital spotting or intermenstrual cycle bleeding or menometrorrhagia. Foul-smelling or persistent yellow discharge may also be present. Symptoms such as pelvic or sacral pain suggest lateral extension into the pelvic nerve plexus by either the primary tumor or a pelvic node metastasis and indicate advanced-stage disease. Likewise, flank pain from hydronephrosis from ureteral compression or deep-venous thrombosis from iliac vessel compression suggests either extensive nodal disease or direct extension of the primary tumor to the pelvic sidewall. The most common finding upon physical exam is a visible tumor on the cervix, but deeper tumors in the cervical os and glands should be considered. Larger tumors may be identified by inspection and biopsied directly. Staging of cervical cancer is performed by expert clinical exam. Stage I cervical tumors are confined to the cervix, whereas stage II tumors extend into the upper vagina or paracervical soft tissue **(Fig. 94-2)**. Stage III tumors extend to the lower vagina or the pelvic sidewalls, whereas stage IV tumors invade the bladder or rectum or have spread to distant sites. While radiographic studies are not part of the formal clinical staging of cervical cancer, treatment planning requires them for appropriate therapy. CT can detect hydronephrosis indicative of pelvic sidewall disease but is not accurate at evaluating other pelvic structures. Magnetic resonance imaging (MRI) is more accurate at estimating uterine extension and paracervical extension of disease into soft tissues typically bordered by broad and cardinal ligaments that support the uterus in the central pelvis. Very small stage I cervical tumors can be treated with a variety of surgical procedures, but minimally invasive surgery has inferior outcome compared to standard open hysterectomy. In young women desiring to maintain fertility, radical trachelectomy removes the cervix with subsequent anastomosis of the upper vagina to the uterine corpus; however, subsequent pregnancies may be more problematic. Patients with large stage I cervical tumors (4 cm) confined to the cervix and all stage II to IV patients are treated with radiation therapy in combination with cisplatin-based immunochemotherapy with concurrent PD-1 blockers. This multimodality treatment can offer the patient with advanced-stage disease a 40–80% chance of cure depending on the clinical circumstances. Immunotherapy with PD-1 blockade, cisplatin, paclitaxel, bevacizumab, and tisotumab vedotin are generally considered as appropriate palliative choices for metastatic

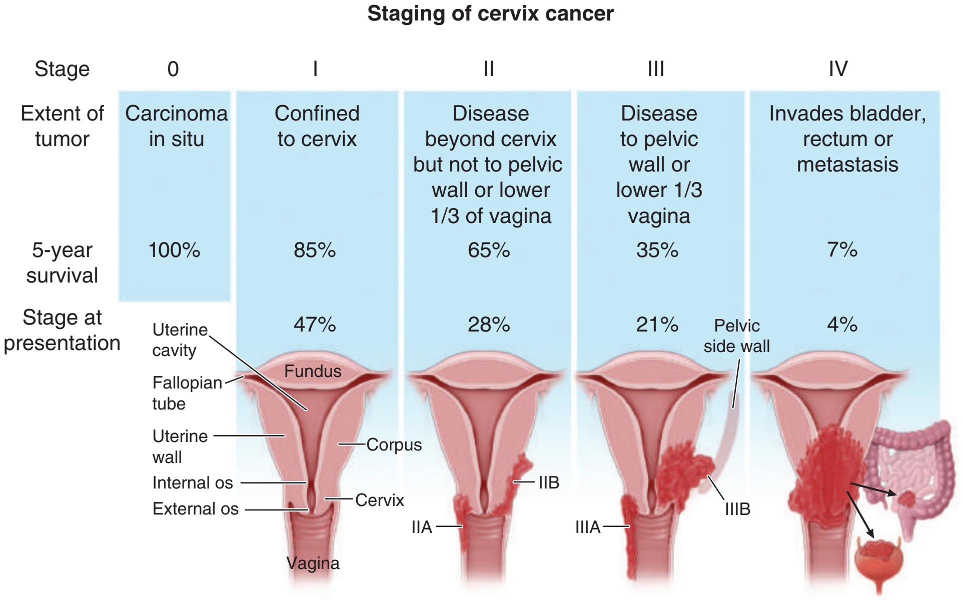

FIGURE 94-2 Anatomic display of the stages of cervix cancer defined by location, extent of tumor, frequency of presentation, and 5-year survival.

cervical cancer patients. Secondary chemotherapy confers minimal improvement in most patients. Additional immunotherapies targeting HPV antigens are potential avenues for improved outcomes in recurrent, unresectable cancers of the cervix.

UTERINE CANCER

EPIDEMIOLOGY

Several different tumor types arise in the uterine corpus. Most tumors arise in the glandular lining and are endometrial adenocarcinomas. Benign (leiomyomas) and malignant smooth muscle tumors (leiomyosarcomas) can also arise in the uterus and have very different clinical features. The endometrioid histologic subtype is the most common gynecologic malignancy in the United States. In 2024, the American Cancer Society predicted that 66,200 new cancers of the uterine corpus are expected in 2024 with 13,030 resulting deaths. Development of these tumors is a multistep process, with estrogen playing an important early role in driving endometrial gland proliferation. Relative overexposure to this class of hormones is the principal risk factor for the subsequent development of endometrioid tumors. In contrast, progestins drive glandular maturation and are protective. Hence, women with high endogenous or pharmacologic exposure to estrogens, especially if unopposed by progesterone, are at higher risk for endometrial cancer. Obese women, women treated with postmenopausal estrogens, or women with estrogen-producing tumors are at higher risk for endometrial cancer. In addition, long-term treatment with tamoxifen, which has antiestrogenic effects in breast tissue but can show weak estrogenic effects in uterine epithelium, is associated with an increased risk of endometrial cancer.

Genetics Women with a germline mutation in one of a series of DNA mismatch repair genes associated with the Lynch syndrome, also known as hereditary nonpolyposis colon cancer (HNPCC) syndrome, are at increased risk for endometrioid endometrial carcinoma. These individuals have germline mutations in *MSH2*, *MLH1*, and, in rare cases, *PMS1* and *PMS2*. Individuals who carry these mutations typically have a family history of cancer and are at markedly increased risk for colon cancer and modestly increased risk for ovarian cancer and a variety of other tumors. Middle-aged women with HNPCC carry a 4% annual risk of endometrial cancer and a relative overall risk of ~200-fold as compared to age-matched women without HNPCC. In sporadic cancers, secondary events such as mutation of the *PI3K* gene or the loss of the *PTEN* tumor-suppressor gene likely serve as secondary genetic "hits" in the carcinogenesis related to estrogenic excess. The molecular events that underlie less common endometrial cancers such as clear cell and papillary serous tumors of the uterine corpus are not well understood.

PATHOLOGY

Approximately 75–80% of endometrial cancers are adenocarcinomas and have been characterized as type 1 (estrogen-linked) endometrial cancers and type 2 cancers that have less clear associations with estrogens (clear cell cancers, serous cancers, and mucinous cancers). Endometrial serous cancers show *TP53* loss of function and behave clinically more like ovarian cancers with high risk for systemic recurrence. Prognosis for endometrial cancer depends on stage, histologic grade, and depth of myometrial invasion.

CLINICAL PRESENTATION

The majority of women with tumors of the uterine corpus present with postmenopausal vaginal bleeding due to shedding of the malignant endometrial lining. Premenopausal women often will present with atypical bleeding between typical menstrual cycles. These signs typically bring a woman to the attention of health care providers, and the majority of women have early-stage disease in which the tumor is confined to the uterine corpus and, consequently, have a high cure rate. Diagnosis is typically established by endometrial biopsy. Type 1 tumors may spread to pelvic or para-aortic lymph nodes and are generally subjected to sentinel lymph node biopsy at the time of primary surgery. Serous tumors tend to have patterns of spread similar to high-grade serous ovarian cancer, and patients may present with omental/peritoneal disease and sometimes ascites. Some women presenting with uterine sarcomas will present with pelvic pain. Uterine sarcomas (carcinosarcomas and leiomyosarcomas) commonly are found by detection of symptomatic large pelvic masses that may not be associated with dysfunctional vaginal bleeding.

TREATMENT

Uterine Cancer

Most women with endometrial cancer have disease that is localized to the uterus (75% are stage I, Table 94-1), and definitive treatment typically involves a hysterectomy with removal of the ovaries and fallopian tubes. The resection of lymph nodes does not improve outcome, but sentinel node resection provides important staging and prognostic information. Node involvement defines stage IIIC disease, which is treated with immunochemotherapy. Tumor grade and depth of invasion are two key prognostic variables in early-stage tumors, and women with low-grade and/or minimally invasive tumors (<50% myometrial penetration) are typically observed after definitive surgical therapy. Patients with high-grade tumors or tumors that are deeply invasive (stage IB) are at higher risk for pelvic recurrence or recurrence at the vaginal cuff, which is typically prevented by intravaginal brachytherapy. It is now routine to test all endometrial cancers for microsatellite instability (MSI) with a larger number of mutations in the tumor. MSI cancers, when recurrent or present at an advanced stage, are likely to respond to immune checkpoint therapy and PD-1–targeted therapy should be part of treatment for those patients. Women with regional metastases or metastatic disease (3% of patients) with low-grade tumors can be treated with progesterone or tamoxifen. In contrast, poorly differentiated tumors lack hormone receptors and are typically unresponsive to hormonal manipulation. The role of adjuvant chemotherapy in stage I–II disease is generally restricted to serous endometrial cancers. For more advanced-stage cancers (stage III–IV), chemotherapy and/or immune checkpoint blockade are administered because of the higher rates of recurrent systemic disease. Carboplatin and paclitaxel combinations with immune targeting agents are the current standard of care. Chemotherapy for metastatic disease is delivered with palliative intent. Even patients with advanced cancer and known mismatch repair deficits may respond well to immunotherapy with antagonists of the PD-1/PD-L1 axis. Lenvatinib and pembrolizumab (even for microsatellite-stable tumors) have become the most common second-line treatments. Other potentially active treatments include bevacizumab and mammalian target of rapamycin (mTOR) inhibitors (e.g., temsirolimus). Newer antibody-drug conjugates may have good responses in patients with expression of the target antigens. Carcinosarcomas of the uterus (also called Müllerian tumors) contain both mesenchymal and epithelial components but will often respond to paclitaxel and platinum complex therapy. Other uterine sarcomas require an entirely different approach and need histology-specific consideration. The most common are the leiomyosarcomas of the uterus, which are treated with docetaxel/gemcitabine at recurrence but do not appear to benefit from adjuvant therapy. Ifosfamide/doxorubicin and trabectedin can have some benefit in refractory disease.

GESTATIONAL TROPHOBLASTIC TUMORS

Gestational trophoblastic diseases represent a spectrum of neoplasia from benign hydatidiform mole to choriocarcinoma due to persistent trophoblastic disease associated most commonly with molar pregnancy but occasionally seen after normal gestation. The most common presentations of trophoblastic tumors are partial and complete molar pregnancies. These represent approximately 1 in 1500 conceptions in developed Western countries. The incidence widely varies globally, with areas in Southeast Asia having a much higher incidence of molar pregnancy. Regions with high molar pregnancy rates are often associated with diets low in carotene and animal fats.

■ RISK FACTORS

Trophoblastic tumors result from the outgrowth or persistence of placental tissue. They arise most commonly in the uterus but can also arise in other sites such as the fallopian tubes due to ectopic pregnancy. Risk factors include poorly defined dietary and environmental factors as well as conceptions at the extremes of reproductive age, with the incidence particularly high in females conceiving at younger than age 16 or older than age 50. In older women, the incidence of molar pregnancy might be as high as one in three, likely due to increased risk of abnormal fertilization of the aged ova. Most trophoblastic neoplasms are associated with complete moles, diploid tumors with all genetic material from the paternal donor (known as uniparental disomy). This is thought to occur when a single sperm fertilizes an enucleate egg that subsequently duplicates the paternal DNA. Trophoblastic proliferation occurs with exuberant villous stroma. If pseudopregnancy extends out past the 12th week, fluid progressively accumulates within the stroma, leading to "hydropic changes." Fetal development does not occur in complete moles.

Partial moles arise from the fertilization of an egg with two sperm cells; hence, two-thirds of genetic material is paternal in these triploid tumors. Hydropic changes are less dramatic, and fetal development can often occur through late first trimester or early second trimester, at which point spontaneous abortion is common. Laboratory findings will include excessively high hCG and high AFP. The risk of persistent gestational trophoblastic disease after partial mole is ~5%. Complete and partial moles can be noninvasive or invasive. Myometrial invasion occurs in no more than one in six complete moles and a lower portion of partial moles.

■ PRESENTATION OF INVASIVE TROPHOBLASTIC DISEASE

The clinical presentation of molar pregnancy is changing in developed countries due to the early detection of pregnancy with home pregnancy kits and the very early use of Doppler and ultrasound to evaluate the early fetus and uterine cavity for evidence of a viable fetus. Thus, in these countries, the majority of women presenting with trophoblastic disease have their moles detected early and have typical symptoms of early pregnancy including nausea, amenorrhea, and breast tenderness. With uterine evacuation of early complete and partial moles, most women experience spontaneous remission of their disease as monitored by serial serum β-hCG levels. These women require no chemotherapy. Patients with persistent elevation of β-hCG or rising β-hCG after uterine evacuation have persistent or actively growing gestational trophoblastic disease and require therapy. Most series suggest that between 15 and 25% of women will have evidence of persistent gestational trophoblastic disease after molar evacuation.

In women who lack access to prenatal care, presenting symptoms can be life-threatening, including the development of preeclampsia or even eclampsia. Hyperthyroidism can also be seen with very high β-hCG values. Evacuation of large moles can be associated with life-threatening complications including uterine perforation, volume loss, high-output cardiac failure, and adult respiratory distress syndrome (ARDS).

For women with evidence of rising β-hCG or radiologic confirmation of metastatic or persistent regional disease, prognosis can be estimated through a variety of scoring algorithms that identify women at low, intermediate, and high risk for requiring multiagent chemotherapy. In general, women with widely metastatic nonpulmonary disease, very elevated β-hCG, and prior normal antecedent term pregnancy are considered at high risk and typically require multiagent chemotherapy at an expert center for cure. Even very advanced gestational trophoblastic disease is almost uniformly curable when managed by an expert in this rare malignancy.

TREATMENT

Invasive Trophoblastic Disease

Management of invasive trophoblastic disease should be 100% curative, and complex patients should only be managed by clinicians experienced in this disease. The management for a persistent and rising β-hCG after evacuation of a molar conception is typically chemotherapy, although surgery can play an important role for chemotherapy-resistant disease that is isolated in the uterus (especially if childbearing is complete) or to control hemorrhage. For women wishing to maintain fertility or with metastatic disease, the preferred treatment is chemotherapy or immunotherapy targeting the PD-1 axis. Trophoblastic disease is exquisitely sensitive to chemotherapy, and guided by serial serum β-hCG testing, successful, curative treatment is the rule. Single-agent treatment with dactinomycin or methotrexate cures 90% of women with low-risk disease. Patients with high-risk disease (very high β-hCG levels, presentation ≥4 months after pregnancy, brain or liver metastases, failure of methotrexate therapy) are typically treated with multiagent chemotherapy (etoposide, methotrexate, and dactinomycin, alternating with cyclophosphamide and vincristine [EMA-CO]), which is typically curative even in women with extensive metastatic disease. A regimen of cisplatin and etoposide alternating with etoposide/methotrexate/dactinomycin is used for the highest-risk patients. In the highest-risk patients with liver, lung, and brain metastases, hemorrhage from the rich tumor vasculature is a major risk during chemotherapy initiation. Cured women may become pregnant again without evidence of increased fetal or maternal complications.

■ FURTHER READING

Longo DL: Personalized medicine for primary treatment of serous ovarian cancer. N Engl J Med 381:2471, 2019.

Lu KH, Broaddus RR: Endometrial cancer. N Engl J Med 383:2053, 2020.

Moore KN et al: Mirvetuximab soravtansine in FRα-positive, platinum-resistant ovarian cancer. N Engl J Med 389:2162, 2023.

95

Primary and Metastatic Tumors of the Nervous System

Mary Jane Lim-Fat, Patrick Y. Wen

An estimated 95,000 people will be diagnosed with a primary brain tumor annually in the United States. At least 27,000 of these tumors are malignant, and most of these are gliomas. Meningiomas account for 41% of all central nervous system (CNS) tumors, vestibular schwannomas 10%, and CNS lymphomas ~2%. Brain metastases are three times more common than all primary brain tumors combined and are diagnosed in ~150,000 people each year. Metastases to the leptomeninges and epidural space of the spinal cord each occur in ~2–12% of patients with systemic cancer and are also a major cause of neurologic disability.

APPROACH TO THE PATIENT

Primary and Metastatic Tumors of the Nervous System

CLINICAL FEATURES

Brain tumors of any type can present with a variety of symptoms and signs that fall into two categories: general and focal; patients often have a combination of the two **(Table 95-1)**. General symptoms include headache, with or without nausea or vomiting, cognitive difficulties, personality change, and gait disorder. These symptoms arise when the enlarging tumor and its surrounding edema cause an increase in intracranial pressure or compression of cerebrospinal fluid (CSF) circulation, leading to hydrocephalus. The classic brain tumor headache predominates in the morning and improves during the day, but this pattern is seen in a minority of patients. Headaches are often holocephalic but can be ipsilateral to the side of a tumor. Occasionally, headaches have features of a typical migraine with unilateral throbbing pain associated with visual scotoma. Personality changes may include apathy and withdrawal from social situations, mimicking depression. Focal or lateralizing findings include hemiparesis, aphasia, or visual field defect. Lateralizing symptoms are typically subacute and progressive; language difficulties may be mistaken for confusion. Seizures are common, occurring in ~25% of patients with brain metastases or malignant gliomas, and are the presenting symptom in up to 90% of patients with a low-grade glioma. All seizures arising from a brain tumor will have a focal onset whether or not it is apparent clinically.

NEUROIMAGING

Cranial magnetic resonance imaging (MRI) is the preferred diagnostic test for any patient suspected of having a brain tumor and should be performed with gadolinium contrast administration. Computed tomography (CT) scan should be reserved for those patients unable to undergo MRI. Malignant brain tumors—whether primary or metastatic—typically enhance with gadolinium, have central areas of necrosis, and are surrounded by edema of the neighboring white matter. Low-grade gliomas usually do not enhance with gadolinium and are best appreciated on fluid-attenuated inversion recovery (FLAIR) MRI sequences. Meningiomas have a typical appearance on MRI because they are dural-based enhancing tumors with a dural tail and compress but do not invade the brain. Dural metastases or a dural lymphoma can have a similar appearance. Imaging is characteristic for many primary and metastatic tumors and sometimes will suffice to establish a diagnosis when the location precludes surgical intervention (e.g., brainstem glioma). Functional MRI is useful in presurgical planning to define eloquent sensory, motor, or language cortex. Positron emission tomography (PET) is useful in determining the metabolic activity of the lesions seen on MRI; MR perfusion and spectroscopy can provide information on blood flow or tissue composition. These techniques may help distinguish tumor progression from tissue necrosis due to treatment with radiation and chemotherapy. Neuroimaging is the only test necessary to diagnose a brain tumor. Laboratory tests are rarely useful, although patients with metastatic disease may have elevation of a serum tumor marker (e.g., β human chorionic gonadotropin [β-hCG] from testicular cancer). Additional testing such as cerebral angiogram, electroencephalogram (EEG), or lumbar puncture is rarely indicated or helpful.

TREATMENT

Brain Tumors

Therapy of any intracranial malignancy requires both symptomatic and definitive treatments. Definitive treatment is based on the specific tumor type and includes surgery, radiotherapy, and chemotherapy. However, symptomatic treatments apply to brain tumors of any type. Most high-grade malignancies are accompanied by substantial surrounding edema, which contributes to neurologic disability and raised intracranial pressure. Glucocorticoids are highly effective at reducing perilesional edema and improving neurologic function, often within hours of administration. Dexamethasone has been the glucocorticoid of choice because of its relatively low mineralocorticoid activity; initial doses are 4–12 mg/d in one to two daily doses. Glucocorticoids rapidly ameliorate symptoms and signs, but their long-term use causes substantial toxicity including insomnia, weight gain, diabetes mellitus, steroid myopathy, and personality changes. Consequently, a taper is indicated as definitive treatment is administered and the patient improves.

Patients with brain tumors who present with seizures require antiepileptic drug therapy. Prophylactic antiepileptic drugs are occasionally used in the perioperative setting, but there is no role for extended use in patients who have not had a seizure. The agents of choice are drugs that do not induce the hepatic microsomal enzyme system. These include levetiracetam, topiramate,

TABLE 95-1 Symptoms and Signs at Presentation of Brain Tumors

	HIGH-GRADE GLIOMA (%)	LOW-GRADE GLIOMA (%)	MENINGIOMA (%)	METASTASES (%)
Generalized				
Impaired cognitive function	50	10	30	60
Hemiparesis	40	10	36	60
Headache	50	40	37	50
Lateralizing				
Seizures	20	70+	17	12
Aphasia	20	<5	—	18
Visual field deficit	—	—	—	7

lamotrigine, valproic acid, and lacosamide **(Chap. 447)**. Other drugs, such as phenytoin and carbamazepine, are used less frequently because they are potent enzyme inducers that can interfere with both glucocorticoid and chemotherapy metabolism. Venous thromboembolic disease occurs in 20–30% of patients with HGGs or brain metastases. Prophylactic anticoagulants should be used during hospitalization and in nonambulatory patients. Those who have had either a deep vein thrombosis or a pulmonary embolus can receive therapeutic doses of anticoagulation safely and without increasing the risk of hemorrhage into the tumor. Inferior vena cava filters are reserved for patients with absolute contraindications to anticoagulation such as recent craniotomy.

PRIMARY BRAIN TUMORS

EPIDEMIOLOGY

No etiology has been identified for most primary brain tumors. The only established risk factors are exposure to ionizing radiation (meningiomas, gliomas, and schwannomas) and immunosuppression (primary CNS lymphoma). There is no proven evidence for any association with exposure to electromagnetic fields including cellular telephones, head injury, foods containing *N*-nitroso compounds, or occupational risk factors. A small minority of patients have a family history of brain tumors. Some of these familial cases are associated with genetic syndromes **(Table 95-2)**.

MOLECULAR PATHOGENESIS

As with other neoplasms, brain tumors arise as a result of a multistep process driven by the sequential acquisition of genetic alterations. These include loss of tumor-suppressor genes (e.g., *p53*, cyclin-dependent kinase inhibitor 2A and 2B [*CDKN2A/B*], and phosphatase and tensin homolog on chromosome 10 [*PTEN*]) and amplification and overexpression of protooncogenes such as the epidermal growth factor receptor (*EGFR*) and platelet-derived growth factor receptors (*PDGFR*). The accumulation of these genetic abnormalities results in uncontrolled cell growth and tumor formation. Many brain tumors, including glioblastomas, are characterized by significant molecular heterogeneity, which contributes to the difficulty in developing effective therapies.

Important progress has been made in understanding the molecular pathogenesis of several types of brain tumors, including glioblastoma and medulloblastoma, allowing them to be separated into different subtypes with different prognoses. This has led the World Health Organization (WHO) to issue an update on the classification of CNS tumors in 2016 that for the first time incorporated molecular parameters in addition to traditional histology into the diagnosis of brain tumors. The most recent 2021 WHO Classification of Tumors of the CNS further stressed the use of integrated diagnoses based on both molecular diagnostics and histology. This has improved the classification of brain tumors, allowing for better understanding of the prognosis and optimal therapy for patients.

INTRINSIC “MALIGNANT” TUMORS

DIFFUSE GLIOMA

Gliomas are the most common type of malignant primary brain tumor. The 2021 WHO Classification now differentiates gliomas as “adult type” or “pediatric type” based on molecular genetic differences. Both adult-type and pediatric-type diffuse gliomas can be subclassified into high- or low-grade glioma based on histology and are further characterized by key molecular alterations **(Table 95-3)**. Although uncommon, pediatric-type diffuse gliomas also occur in adults, hence their inclusion here.

TABLE 95-2 Genetic Syndromes Associated with Primary Brain Tumors

SYNDROME	INHERITANCE	GENE/PROTEIN	ASSOCIATED TUMORS
Cowden’s syndrome	AD	Mutations of *PTEN* (ch10p23)	Dysplastic cerebellar gangliocytoma (Lhermitte-Duclos disease), meningioma, astrocytoma Breast, endometrial, thyroid cancer, trichilemmomas
Familial schwannomatosis	Sporadic Hereditary	Mutations in *INI1/SNF5* (ch22q11)	Schwannomas, gliomas
Gardner’s syndrome	AD	Mutations in *APC* (ch5q21)	Medulloblastoma, glioblastoma, craniopharyngioma Familial polyposis, multiple osteomas, skin and soft tissue tumors
Gorlin syndrome (basal cell nevus syndrome)	AD	Mutations in *Patched 1* gene (ch9q22.3)	Medulloblastomas Basal cell carcinoma
Li-Fraumeni syndrome	AD	Mutations in *p53* (ch17p13.1)	Gliomas, medulloblastomas Sarcomas, breast cancer, leukemias, others
Lynch syndrome	AD	Mutations in *MSH2, MSH1, MSH6, PMS2*	Glioblastoma and other gliomas Gastrointestinal, endometrial, and other cancers
Multiple endocrine neoplasia 1 (Wermer’s syndrome)	AD	Mutations in *Menin* (ch11q13)	Pituitary adenoma, malignant schwannomas Parathyroid and pancreatic islet cell tumors
NF1	AD	Mutations in *NF1*/neurofibromin (ch17q12-22)	Schwannomas, astrocytomas, optic nerve gliomas, meningiomas Neurofibromas, neurofibrosarcomas, others
NF2	AD	Mutations in *NF2*/merlin (ch22q12)	Bilateral vestibular schwannomas, astrocytomas, multiple meningiomas, ependymomas
TSC (Bourneville disease)	AD	Mutations in *TSC1/TSC2* (ch9q34/16)	Subependymal giant cell astrocytoma, ependymomas, glioma, ganglioneuroma, hamartoma
Turcot syndrome	AD	Mutations in *APC*[a] (ch5)	Gliomas, medulloblastomas
	AR	hMLH1 (ch3p21)	Adenomatous colon polyps, adenocarcinoma
VHL	AD	Mutations in *VHL* gene (ch3p25)	Hemangioblastomas Retinal angiomas, renal cell carcinoma, pheochromocytoma, pancreatic tumors and cysts, endolymphatic sac tumors of the middle ear

[a] Various DNA mismatch repair gene mutations may cause a similar clinical phenotype, also referred to as Turcot syndrome, in which there is a predisposition to nonpolyposis colon cancer and brain tumors.

Abbreviations: AD, autosomal dominant; APC, adenomatous polyposis coli; AR, autosomal recessive; ch, chromosome; NF, neurofibromatosis; PTEN, phosphatase and tensin homologue; TSC, tuberous sclerosis complex; VHL, von Hippel-Lindau.

TABLE 95-3 Summary of Gliomas and Relevant Molecular Alterations

TUMOR TYPE	CHARACTERISTIC MOLECULAR ALTERATIONS
Adult-Type Diffuse Gliomas	
Astrocytoma, IDH-mutant	*IDH1, IDH2*
Oligodendroglioma, IDH-mutant, 1p/19q-codeleted	*IDH1, IDH2,* 1p/19q
Glioblastoma, IDH wild type	Chromosome 7 gain and 10 loss, *TERT, EGFR*
Pediatric-Type Diffuse High-Grade Gliomas	
Diffuse midline glioma, H3 K27-altered	H3 K27M, H3K27me3, *EGFR, EZHIP*
Diffuse hemispheric glioma, H3 G34-mutant	H3.3 G34R/V
Diffuse pediatric-type high-grade glioma, H3 wild type and IDH wild type	*EGFR, PDGFRA, MYCN*
Pediatric-Type Diffuse Low-Grade Gliomas	
Diffuse low-grade glioma, MAPK pathway-altered	MAPK pathway genes (*BRAF V600E* mutation, *BRAF* fusion, *FGFR* mutation)

Abbreviations: IDH, isocitrate dehydrogenase; *TERT,* telomerase reverse transcriptase; *EGFR,* epidermal growth factor receptor; *FGFR,* fibroblast growth factor receptor; *MAPK,* mitogen-activated protein kinase; *EZHIP,* EZH inhibitor protein; *PDGFRA,* platelet-derived growth factor receptor alpha; *MYCN,* MYCN proto-oncogene.

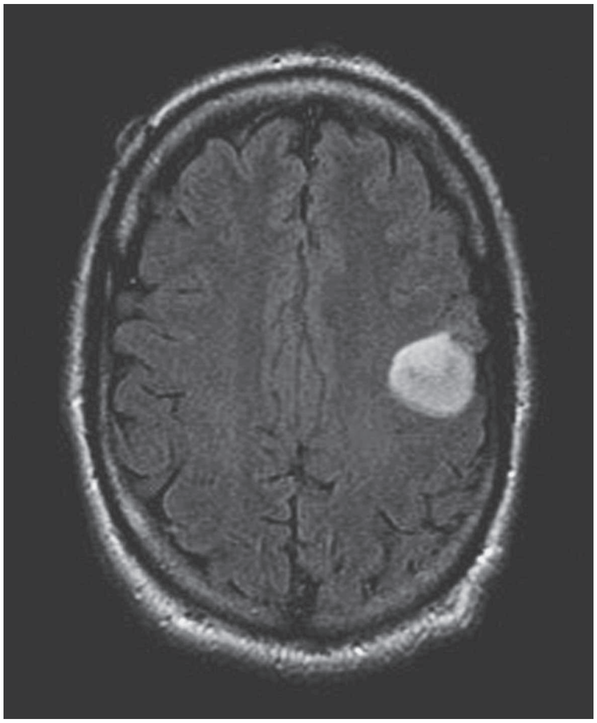

FIGURE 95-1 Fluid-attenuated inversion recovery (FLAIR) MRI of a left frontal low-grade astrocytoma. This lesion did not enhance.

ADULT-TYPE DIFFUSE GLIOMA

Adult-type diffuse gliomas are classified based on the presence of a mutation in a key driver, the isocitrate dehydrogenase (*IDH*) gene, followed by the presence of 1p/19q co-deletion. Both of these molecular alterations have significant prognostic impact and lead to the emergence of three distinct groups: *IDH*-mutant astrocytoma (1p/19q intact), *IDH*-mutant oligodendroglioma (1p/19q co-deleted), and *IDH* wild-type glioblastoma.

Diffuse gliomas can present rarely as widespread infiltration of the brain tissue without a focal mass. Such tumors usually present with cognitive problems, and the MRI demonstrates confluent, typically nonenhancing areas of increased signal on FLAIR sequences without significant mass effect. Formerly known as gliomatosis cerebri, these lesions are now categorized by the pathology identified on biopsy, but they can be diagnostically challenging when the nature of the imaging abnormalities is unclear. Often diagnosis is delayed until the patient develops worsening deficits or there is clear progression on imaging. Treatment is then determined by the pathology.

Astrocytoma, IDH-Mutant *IDH*-mutant astrocytoma can be further classified into grade 2, 3, or 4 based on histologic features, with higher grade tumors incorporating additional features of malignancy. *CDKN2A/B* homozygous deletion is associated with poor prognosis, and its presence confers a grade 4 to *IDH*-mutant astrocytomas with lower-grade histologic features. Low-grade (grade 2) *IDH*-mutant astrocytomas are infiltrative tumors that usually present with seizures in young adults or can be detected incidentally. They appear as nonenhancing tumors with increased T2/FLAIR signal **(Fig. 95-1)**. If feasible, patients should undergo maximal surgical resection, although complete resection is rarely possible because of the invasive nature of the tumor. In patients at higher risk for recurrence (subtotal resection or above the age of 40 years), there is evidence that radiation therapy (RT) followed by PCV (procarbazine, lomustine, and vincristine) or temozolomide chemotherapy may be of benefit. The tumor transforms to a higher-grade astrocytoma in most patients, leading to variable survival with a median of ~10 years. The *IDH*-inhibitor vorasidenib has been showed to be effective in delaying the progression-free survival and time to next intervention, and *IDH* inhibitors may become an integral part of the management for both *IDH*-mutant low-grade glioma astrocytoma and oligodendroglioma. High-grade *IDH*-mutant astrocytoma includes grade 3 astrocytoma (formerly termed anaplastic astrocytoma) and grade 4 astrocytoma (previously *IDH*-mutant glioblastoma) and generally presents in the fourth and fifth decades of life as variably enhancing tumors. Treatment is the same as for glioblastoma, consisting of maximal safe surgical resection followed by RT and adjuvant temozolomide alone or RT with concurrent and adjuvant temozolomide. The median survival for high-grade anaplastic astrocytoma can range from 3 to 9 years.

Oligodendroglioma, IDH-Mutant and 1p/19q Co-deleted Oligodendrogliomas account for ~15–20% of gliomas. They are characterized by co-deletion of 1p/19q and have *IDH* mutations. Oligodendrogliomas are classified by the WHO into grade 2 or grade 3 oligodendrogliomas (formerly anaplastic oligodendrogliomas). Oligodendrogliomas have distinctive pathologic features such as perinuclear clearing—giving rise to a “fried egg” appearance—and a reticular pattern of blood vessel growth. Some tumors have both an oligodendroglial as well as an astrocytic component. With molecular testing, it is now clear that almost all of these mixed tumors (oligoastrocytomas) are genetically either astrocytomas or oligodendrogliomas. As a result, the diagnosis of oligoastrocytoma is now rarely made unless molecular testing is not available.

Grade 2 oligodendrogliomas are generally more responsive to therapy and have a better prognosis than pure astrocytic tumors. These tumors present similarly to grade 2 astrocytomas in young adults. The tumors are nonenhancing and often partially calcified. They should be treated with surgery and, in patients with residual disease or aged >40 years, RT and chemotherapy. Targeting mutant *IDH* with inhibitors such as vorasidenib to delay progression and transformation to higher grade may become an effective strategy in low-grade oligodendroglioma. Patients with oligodendrogliomas have a median survival in excess of 10 years.

Grade 3 oligodendrogliomas present in the fourth and fifth decades as variably enhancing tumors. They are more responsive to therapy than grade 3 astrocytomas. Treatment involves maximal safe resection followed by RT and PCV or temozolomide chemotherapy. Median survival of patients is in excess of 10 years.

Glioblastoma Glioblastomas account for the majority of high-grade astrocytomas and are now defined by the absence of *IDH* mutations. With the new WHO classification, grade 2 and 3 astrocytomas without the classic histologic features of glioblastoma (necrosis and endothelial proliferation) but harboring molecular features of glioblastoma

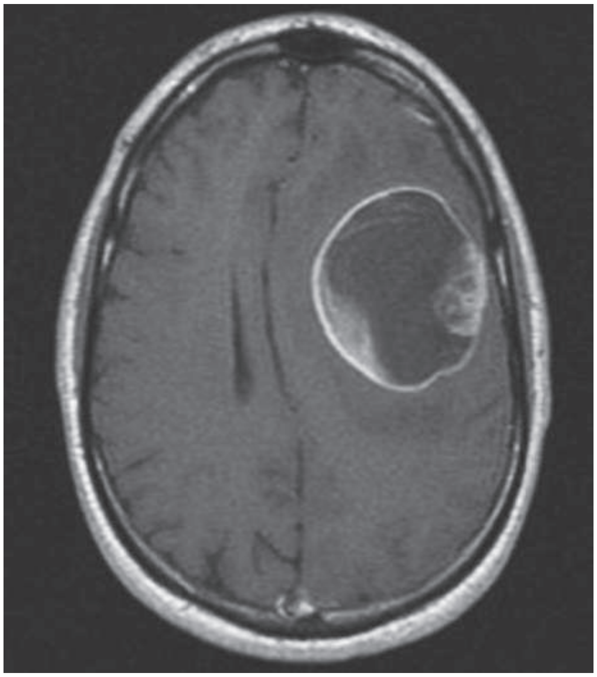

FIGURE 95-2 Postgadolinium T1 MRI of a large cystic left frontal glioblastoma.

(epidermal growth factor amplification, combined with whole chromosome 7 gain and 10 loss, or telomerase reverse transcriptase [*TERT*] promoter mutations) are considered molecular glioblastomas.

Glioblastomas are the most common malignant primary brain tumor, with >12,000 cases diagnosed each year in the United States. Patients usually present in the sixth and seventh decades of life with headache, seizures, or focal neurologic deficits. The tumors appear as ring-enhancing masses with central necrosis and surrounding edema **(Fig. 95-2)**. These are highly infiltrative tumors, and the areas of increased T2/FLAIR signal surrounding the main tumor mass contain invading tumor cells. Treatment involves maximal surgical resection followed by involved-field external-beam RT (6000 cGy in thirty 200-cGy fractions) with concomitant temozolomide, followed by 6 months of adjuvant temozolomide. With this regimen, median survival is increased to 15–18 months compared to only 12 months with RT alone, and 5-year survival is ~10%. Efforts to increase the dose of RT locally using brachytherapy or stereotactic radiosurgery (SRS) have failed to improve the outcome, and these treatments are not recommended. Patients whose tumor contains the DNA repair enzyme O^6-methylguanine-DNA methyltransferase (MGMT) are relatively resistant to temozolomide and have a worse prognosis compared to those whose tumors contain low levels of MGMT as a result of silencing of the *MGMT* gene by promoter hypermethylation. Implantation of biodegradable polymers containing carmustine chemotherapy into the tumor bed after resection of the tumor or addition of tumor treating fields (scalp electrodes delivering low-intensity electric currents) produces a modest improvement in survival.

For elderly patients aged >65–70 years, a hypofractionated RT regimen of 40 Gy over 3 weeks with temozolomide is well tolerated and likely leads to similar outcomes as the 6-week standard RT regimen.

Despite optimal therapy, glioblastomas invariably recur. Treatment options for recurrent disease may include reoperation, reirradiation, and treatment with bevacizumab and standard chemotherapeutic regimens. Bevacizumab, a humanized vascular endothelial growth factor (VEGF) monoclonal antibody, has activity in recurrent glioblastoma, increasing progression-free survival but not overall survival and reducing peritumoral edema and glucocorticoid use **(Fig. 95-3)**. Immune checkpoint inhibitors have been successful in a variety of solid tumors but have failed to demonstrate substantial activity in glioblastoma. A recent phase 3 trial comparing bevacizumab with nivolumab in recurrent glioblastoma demonstrated an identical median overall survival of approximately 10 months in the two arms, with similar toxicities. Treatment decisions for patients with recurrent glioblastoma must be made on an individual basis, taking into consideration such factors as previous therapy, time to relapse, performance status, and quality of life. Whenever feasible, patients should be enrolled in clinical trials. Novel therapies undergoing evaluation in patients with glioblastoma include targeted molecular agents directed at receptor tyrosine kinases and signal transduction pathways; immunotherapy using vaccines, novel checkpoint inhibitors, or chimeric antigen receptor (CAR) T cells; oncolytic viruses; antiangiogenic agents; chemotherapeutic agents that cross the blood-brain barrier more effectively than currently available drugs; and infusion of radiolabeled drugs and targeted toxins into the tumor and surrounding brain by means of convection-enhanced delivery.

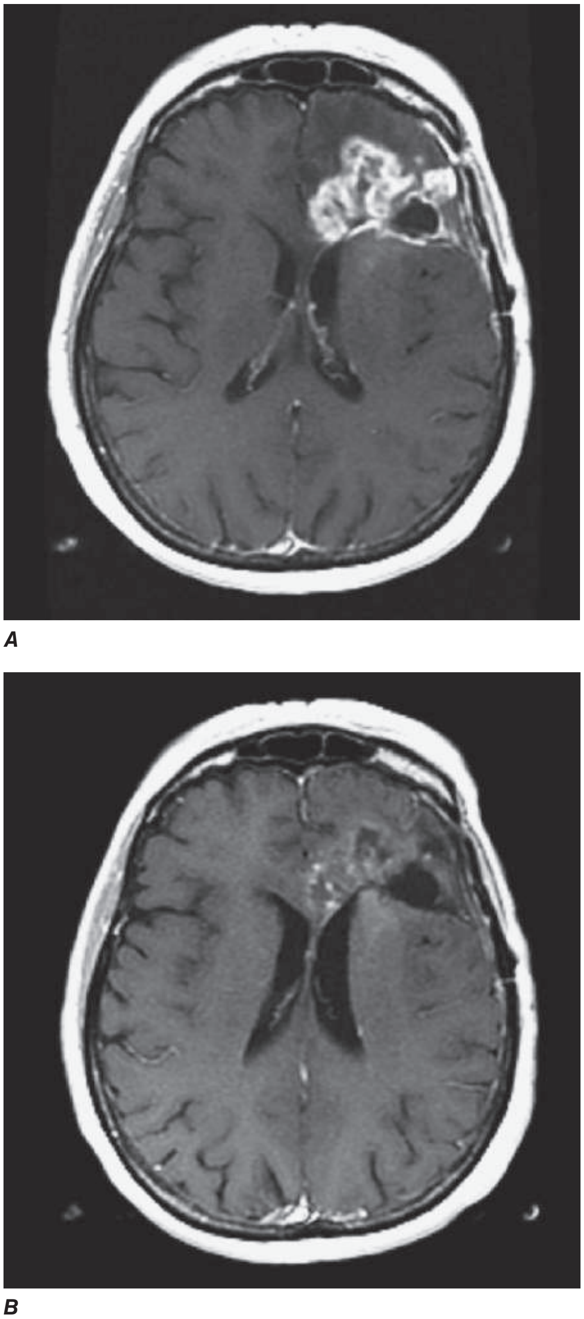

FIGURE 95-3 Postgadolinium T1 MRI of a recurrent glioblastoma before (***A***) and after (***B***) administration of bevacizumab. Note the decreased enhancement and mass effect.

The most important adverse prognostic factors in patients with glioblastomas are older age, unmethylated MGMT promoter, poor Karnofsky performance status, and unresectable tumor.

Gliosarcomas are a variant of glioblastoma containing both an astrocytic and a sarcomatous component and are treated in the same way as glioblastomas.

PEDIATRIC-TYPE DIFFUSE HIGH-GRADE GLIOMA

Pediatric-type diffuse high-grade gliomas (HGGs) are clinically and biologically distinct and are further classified based on their location, which also corresponds to specific molecular alterations. These tumors are typical in younger patients, including young adults, and prior to the recognition of these specific molecular drivers, many pediatric-type HGGs, which do not harbor *IDH* mutations, were often diagnosed and treated as glioblastoma.

Diffuse Midline Glioma, H3K27-Altered These tumors arise from midline structures including the pons, thalamus, or spinal cord and have an infiltrative appearance, often without contrast enhancement on MRI. They harbor mutations in the *H3F3A* gene, resulting in lysine-to-methionine substitution in amino acid residue 27 (K27M). H3K27-altered glioma can occur both in children and adults and carries a poor prognosis regardless of grade. As gross total resection is not feasible in these tumors, treatment mostly consists of RT, while systemic therapy options remain limited in efficacy. The median overall survival is about 1 year.

Diffuse Hemispheric Glioma, H3 G34-Mutant Another HGG that preferentially occurs in young adults is the diffuse hemispheric glioma with the histone variant H3.3 glycine to arginine or valine (H3.3-G34R/V) mutation. These tend to occur in a hemispheric location, and the progression remains poor, with median overall survival of 18–22 months. No standard-of-care treatment exists for these tumors, which are often treated similar to GBM with RT and chemotherapy.

PEDIATRIC-TYPE LOW-GRADE GLIOMA

While most common in pediatric patients, pediatric-type low-grade gliomas can also arise in young and older adults. They include avian myeloblastosis viral oncogene (*MYB*)-altered or MYB proto-oncogene like 1 (*MYBL1*)-altered diffuse gliomas and mitogen-activated protein kinase (*MAPK*)-altered diffuse gliomas (commonly v-Raf murine sarcoma viral oncogene homolog B [*BRAF*] or fibroblast growth factor receptor [*FGFR*] alterations). Gross total resection can be curative in many cases, and systemic therapy can provide durable response with oral BRAF or MAPK inhibitors when a targetable mutation or fusion is found.

CIRCUMSCRIBED ASTROCYTOMA AND OTHER GLIONEURONAL TUMORS

These tumors regroup several histologies of more circumscribed tumors, with a typically more indolent natural history. Pilocytic astrocytomas (WHO grade 1) are the most common tumor of childhood. They occur typically in the cerebellum but may also be found elsewhere in the neuraxis, including the optic nerves and brainstem. Frequently they appear as cystic lesions with an enhancing mural nodule. Often, they have *BRAF* fusions or mutations. These are well-demarcated lesions that are potentially curable if they can be resected completely. Subependymal giant cell astrocytomas (SEGAs) are usually found in the ventricular wall of patients with tuberous sclerosis, discussed later.

Gangliogliomas and pleomorphic xanthoastrocytomas occur in young adults. They behave as more indolent forms of grade 1 gliomas and are usually treated with surgery. Frequently they will have *BRAF* V600E mutations, which can be targeted with BRAF inhibitors.

EPENDYMOMAS

Ependymomas are tumors derived from ependymal cells that line the ventricular surface. They arise in three different compartments and are classified according to location as supratentorial, posterior fossa, or spinal ependymomas. Further molecular stratification within each location (supratentorial ependymoma with *ZFTA* or *YAP1* oncogenic fusion; posterior fossa A or posterior fossa B; and spinal with or without *MYCN* amplification) and grading can help guide management. Children typically present with posterior fossa ependymoma followed by supratentorial ependymoma. Although adults can have intracranial ependymomas, they occur more commonly in the spine, especially in the filum terminale of the spinal cord where they have a myxopapillary histology. Ependymomas that can be completely resected are potentially curable. Partially resected ependymomas will recur and require irradiation. The less common anaplastic ependymoma is more aggressive and is treated with resection and RT; chemotherapy has limited efficacy. Subependymomas are slow-growing benign lesions arising in the wall of ventricles that often do not require treatment.

PRIMARY CENTRAL NERVOUS SYSTEM LYMPHOMA

Primary central nervous system lymphoma (PCNSL) is a rare non-Hodgkin's lymphoma accounting for <3% of primary brain tumors. For unclear reasons, its incidence is increasing, particularly in immunocompetent, older individuals.

PCNSL in immunocompetent patients is usually a diffuse large B-cell lymphoma. Immunocompromised patients, especially those infected with the human immunodeficiency virus (HIV) or organ transplant recipients, are at risk for PCNSL that is typically large cell with immunoblastic and more aggressive features. Epstein-Barr virus (EBV) plays an important role in the pathogenesis of PCNSL in this population. These patients are usually severely immunocompromised, with CD4 counts of <50/mL.

Immunocompetent patients with PCNSL are older (median age, 60 years) than those with HIV-related PCNSL (median age, 31 years). PCNSL usually presents as a mass lesion, with neuropsychiatric symptoms, lateralizing signs, or seizures. Ocular and leptomeningeal involvement each occur in 15–20% of patients, and involvement of these compartments may be asymptomatic. Rarely, it may present as isolated ocular lymphoma or as primary leptomeningeal lymphoma. When restricted to the leptomeninges, it may present as a subacute or chronic meningitis that causes progressive cranial and spinal nerve dysfunction. CSF cytologic examination or flow cytometry is required to establish the diagnosis.

On contrast-enhanced MRI, PCNSL usually appears as a densely enhancing tumor (Fig. 95-4). Immunocompetent patients have solitary lesions more often than immunosuppressed patients. Frequently, there is involvement of the basal ganglia, corpus callosum, or periventricular region. Stereotactic biopsy is necessary to obtain a histologic diagnosis. Whenever possible, glucocorticoids should be withheld until after the biopsy has been obtained because they have a cytolytic effect on

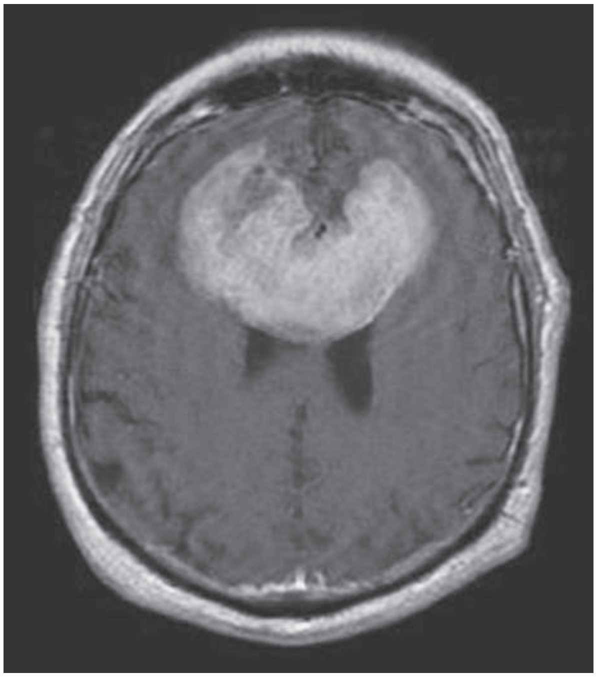

FIGURE 95-4 Postgadolinium T1 MRI demonstrating a large bifrontal primary central nervous system lymphoma (PCNSL). The periventricular location and diffuse enhancement pattern are characteristic of lymphoma.

lymphoma cells and may lead to nondiagnostic tissue. In addition, patients should be tested for HIV, and the extent of disease should be assessed by performing PET or CT of the body, MRI of the spine, CSF analysis, and slit-lamp examination of the eye. Bone marrow biopsy and testicular ultrasound are occasionally performed.

TREATMENT

Primary Central Nervous System Lymphoma

PCNSL is more sensitive to glucocorticoids, chemotherapy, and RT than other primary brain tumors. Durable complete responses and long-term survival are possible with these treatments. High-dose methotrexate, a folate antagonist that interrupts DNA synthesis, produces response rates ranging from 35 to 80% and median survival of up to 50 months. The combination of methotrexate with other chemotherapeutic agents such as cytarabine increases the response rate to 70–100%. The addition of whole-brain RT (WBRT) to methotrexate-based chemotherapy prolongs progression-free survival but not overall survival, but it is associated with delayed neurotoxicity, especially in patients aged >60 years. As a result, full-dose RT is frequently omitted, but there may be a role for reduced-dose RT. The anti-CD20 monoclonal antibody rituximab is often incorporated into the chemotherapy regimen, although there are studies questioning its benefit. For some patients, high-dose chemotherapy with autologous stem cell rescue may offer the best chance of preventing relapse. At least 50% of patients will eventually develop recurrent disease. Treatment options include RT for patients who have not had prior irradiation, retreatment with methotrexate, and other chemotherapeutic agents such as temozolomide and pemetrexed. High-dose chemotherapy with autologous stem cell rescue may be appropriate in selected patients with relapsed disease. Bruton's tyrosine kinase (BTK) inhibitors such as ibrutinib, immunomodulatory drugs such as pomalidomide and lenalidomide, and immune checkpoint inhibitors have shown promising preliminary activity and are being evaluated in clinical trials, as are CAR-T cells.

PCNSL IN IMMUNOCOMPROMISED PATIENTS

PCNSL in immunocompromised patients often produces multiple ring-enhancing lesions that can be difficult to differentiate from metastases or infections such as toxoplasmosis. The diagnosis is usually established by examination of the CSF for cytology and EBV DNA; toxoplasmosis serologic testing; brain PET imaging for hypermetabolism of the lesions, which, although nonspecific, can be consistent with tumor; and, if necessary, brain biopsy. Since the advent of highly active antiretroviral drugs, the incidence of HIV-related PCNSL has declined. These patients are preferably treated with high-dose methotrexate-based regimens and initiation of highly active antiretroviral therapy; WBRT is reserved for those who cannot tolerate systemic chemotherapy. In organ transplant recipients, reduction of immunosuppression may improve outcome.

MEDULLOBLASTOMA

Medulloblastomas are the most common malignant brain tumor of childhood, accounting for ~20% of all primary CNS tumors among children. They arise from granule cell progenitors or from multipotent progenitors from the ventricular zone. Approximately 5% of children with medulloblastoma have an inherited syndrome, such as Gorlin, Turcot, or Li-Fraumeni, which predisposes to the development of medulloblastoma. Histologically, medulloblastomas are highly cellular tumors with abundant dark staining, round nuclei, and rosette formation (Homer-Wright rosettes). In the 2016 WHO pathologic classification, they have been divided into four molecular subgroups: (1) WNT-activated (primarily affects children and has the best outcome); (2) SHH-activated (affects adults, infants, and children, with the younger patients having the better outcome and adults doing poorly); (3) non-WNT/non-SHH, group 3 (frequently has disseminated CNS disease at diagnosis and has the worst outcome); and (4) non-WNT/non-SHH, group 4 (30% have metastases at diagnosis, but 5-year progression-free survival is 95%). Regardless of subtype, patients present with headache, ataxia, and signs of brainstem involvement. On MRI, they appear as densely enhancing tumors in the posterior fossa, sometimes associated with hydrocephalus. Treatment involves maximal surgical resection, craniospinal irradiation, and chemotherapy with agents such as cisplatin, lomustine, cyclophosphamide, and vincristine. Approximately 70% of patients overall have long-term survival but usually at the cost of significant neurocognitive impairment. A major goal of current research is to improve survival while minimizing long-term complications, and clinical trials are now being designed for specific molecular subgroups.

PINEAL REGION TUMORS

A large variety of tumors can arise in the region of the pineal gland. These typically present with headache, visual symptoms, and hydrocephalus. Patients may have Parinaud's syndrome characterized by impaired upgaze and accommodation. Some pineal tumors such as pineocytomas and benign teratomas can be treated by surgical resection. Germinomas respond to irradiation, whereas pineoblastomas and nongerminomatous germ cell tumors require craniospinal RT and chemotherapy.

EXTRINSIC "BENIGN" TUMORS

MENINGIOMA

Meningiomas are diagnosed with increasing frequency as more people undergo neuroimaging for various indications. They are now the most common primary brain tumor, accounting for ~35% of the total. Their incidence increases with age. They tend to be more common in women and in patients with neurofibromatosis type 2 (NF2). They also occur more commonly in patients with a history of cranial irradiation.

Meningiomas arise from the dura mater and are composed of neoplastic meningothelial (arachnoidal cap) cells. They are most commonly located over the cerebral convexities, especially adjacent to the sagittal sinus, but they can also occur in the skull base and along the dorsum of the spinal cord. Meningiomas are classified by the WHO into three histologic grades of increasing aggressiveness: grade I (benign), grade II (atypical), and grade III (malignant).

Many meningiomas are found incidentally following neuroimaging for unrelated reasons. They can also present with headaches, seizures, or focal neurologic deficits. On imaging studies, they have a characteristic appearance usually of a densely enhancing extra-axial tumor arising from the dura **(Fig. 95-5)**. Typically they have a dural tail, consisting of thickened, enhanced dura extending like a tail from the mass. The main differential diagnosis of meningioma is a dural metastasis.

If the meningioma is small and asymptomatic, no intervention is necessary and the lesion can be observed with serial MRI studies. Larger, symptomatic lesions should be resected. If complete resection is achieved, the patient is cured. Incompletely resected tumors tend to recur, although the rate of recurrence can be very slow with grade I tumors. Tumors that cannot be resected or can only be partially removed may benefit from external-beam RT or SRS. These treatments may also be helpful in patients whose tumor has recurred after surgery. Hormonal therapy and chemotherapy are currently unproven.

Rarer tumors that resemble meningiomas include hemangiopericytomas and solitary fibrous tumors. Since they share similar molecular alterations (*NAB2-STAT6* fusion), the 2016 WHO classification introduced the combined term *solitary fibrous tumor/hemangiopericytoma* for this entity. These tumors are treated with surgery and RT but have a higher propensity to recur locally or metastasize systemically.

SCHWANNOMA

These are generally benign tumors arising from the Schwann cells of cranial and spinal nerve roots. The most common schwannomas, termed *vestibular schwannomas* or *acoustic neuromas*, arise from the vestibular portion of the eighth cranial nerve and account for ~9% of primary brain tumors. Patients with NF2 have a high incidence of vestibular schwannomas that are frequently bilateral. Schwannomas arising from other cranial nerves, such as the trigeminal nerve (cranial nerve V), occur with much lower frequency. Neurofibromatosis type 1

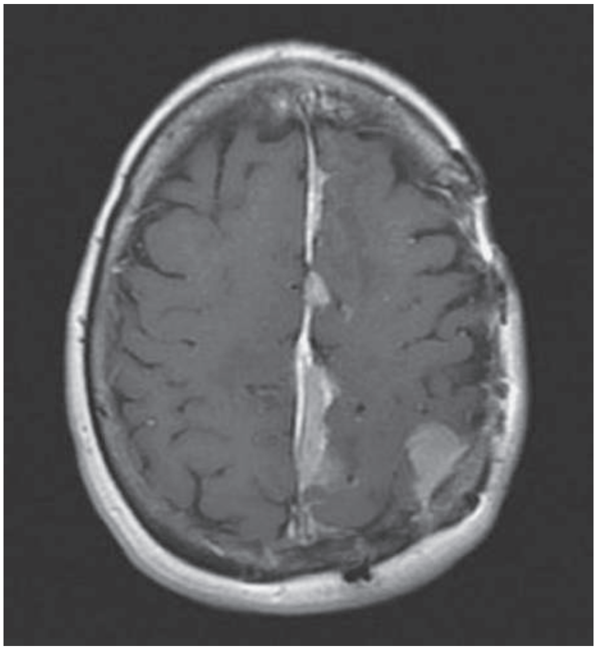

FIGURE 95-5 Postgadolinium T1 MRI demonstrating multiple meningiomas along the falx and left parietal cortex.

(NF1) is associated with an increased incidence of schwannomas of the spinal nerve roots.

Vestibular schwannomas may be found incidentally on neuroimaging or present with progressive unilateral hearing loss, dizziness, tinnitus, or, less commonly, symptoms resulting from compression of the brainstem and cerebellum. On MRI, they appear as densely enhancing lesions, enlarging the internal auditory canal and often extending into the cerebellopontine angle (Fig. 95-6). The differential diagnosis includes meningioma. Very small, asymptomatic lesions can be observed with serial MRIs. Larger lesions should be treated with surgery or SRS. The optimal treatment will depend on the size of the tumor, symptoms, and the patient's preference. In patients with small vestibular schwannomas and relatively intact hearing, early surgical intervention increases the chance of preserving hearing.

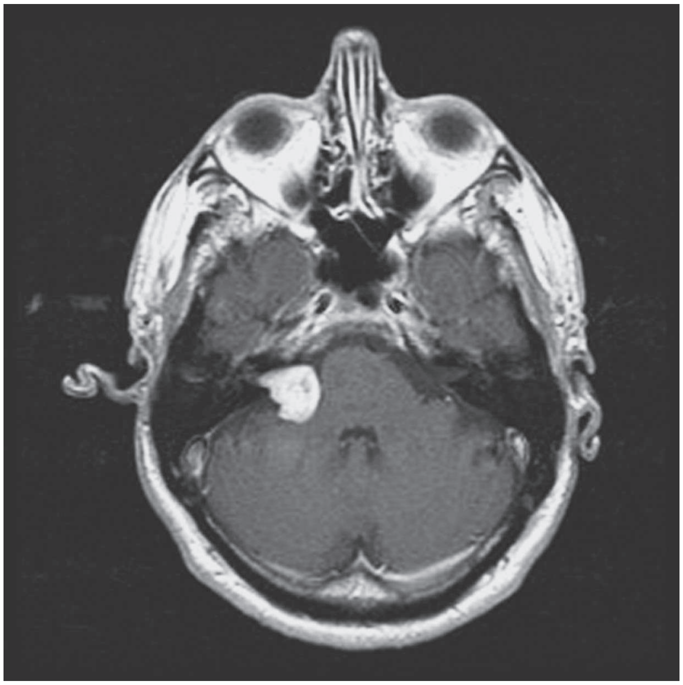

FIGURE 95-6 Postgadolinium MRI of a right vestibular schwannoma. The tumor can be seen to involve the internal auditory canal.

PITUITARY TUMORS

These are discussed in detail in Chap. 392.

CRANIOPHARYNGIOMAS

Craniopharyngiomas are rare, usually suprasellar, partially calcified, solid, or mixed solid-cystic benign tumors that arise from remnants of Rathke's pouch. They have a bimodal distribution, occurring predominantly in children but also between the ages of 55 and 65 years. They present with headaches, visual impairment, and impaired growth in children and hypopituitarism in adults. Treatment involves surgery, RT, or a combination of the two. The papillary subtype of craniopharyngiomas often has *BRAF* V600E mutations and can be treated with RAF/MEK inhibitors.

OTHER BENIGN TUMORS

Epidermoid Cysts These consist of squamous epithelium surrounding a keratin-filled cyst. They are usually found in the cerebellopontine angle and the intrasellar and suprasellar regions. They may present with headaches, cranial nerve abnormalities, seizures, or hydrocephalus. MRI demonstrates an extra-axial lesion with characteristics that are similar to CSF but have restricted diffusion. Treatment involves surgical resection.

Dermoid Cysts Like epidermoid cysts, dermoid cysts arise from epithelial cells that are retained during closure of the neural tube. They contain both epidermal and dermal structures such as hair follicles, sweat glands, and sebaceous glands. Unlike epidermoid cysts, these tumors usually have a midline location. They occur most frequently in the posterior fossa, especially the vermis, fourth ventricle, and suprasellar cistern. On MRI, dermoid cysts resemble lipomas, demonstrating T1 hyperintensity and variable signal on T2. Symptomatic dermoid cysts can be treated with surgery.

Colloid Cysts These usually arise in the anterior third ventricle and may present with headaches, hydrocephalus, and, very rarely, sudden death. Surgical resection is curative, or a third ventriculostomy may relieve the obstructive hydrocephalus and be sufficient therapy.

NEUROCUTANEOUS SYNDROMES (PHAKOMATOSES)

A number of genetic disorders are characterized by cutaneous lesions and an increased risk of brain tumors. Most of these disorders have an autosomal dominant inheritance with variable penetrance.

NEUROFIBROMATOSIS TYPE 1 (von RECKLINGHAUSEN'S DISEASE)

NF1 is an autosomal dominant disorder with variable penetrance and an incidence of ~1 in 2600–3000. Approximately one-half of cases are familial; the remainder are caused by new mutations arising in patients with unaffected parents. The *NF1* gene is located on chromosome 17q11.2 and encodes neurofibromin, a guanosine triphosphatase (GTPase) activating protein (GAP) that is a negative regulator of the RAS–mitogen-activated protein (MAP) kinase signaling pathway, which includes the downstream kinase MEK. It is a classic tumor suppressor, and biallelic loss can result in a variety of nervous system tumors including neurofibromas, plexiform neurofibromas, optic nerve gliomas, astrocytomas, and meningiomas. In addition to neurofibromas, which appear as multiple, soft, rubbery cutaneous tumors, other cutaneous manifestations of NF1 include café-au-lait spots and axillary freckling. NF1 is also associated with hamartomas of the iris termed Lisch nodules, pheochromocytomas, pseudoarthrosis of the tibia, scoliosis, epilepsy, and intellectual disability. The MEK inhibitor selumetinib has activity against inoperable plexiform neurofibromas and is the only treatment that targets the dysregulated signaling pathway.

NEUROFIBROMATOSIS TYPE 2

NF2 is less common than NF1, with an incidence of 1 in 25,000–40,000. It is an autosomal dominant disorder with full penetrance. As with NF1, approximately one-half of cases arise from new mutations.

The *NF2* gene on 22q encodes a cytoskeletal protein, merlin (moesin, ezrin, radixin-like protein), that functions as a tumor suppressor. NF2 is characterized by bilateral vestibular schwannomas in >90% of patients, multiple meningiomas, and spinal ependymomas and astrocytomas. Treatment of bilateral vestibular schwannomas can be challenging because the goal is to preserve hearing for as long as possible. These patients may also have diffuse schwannomatosis that may affect the cranial, spinal, or peripheral nerves; posterior subcapsular lens opacities; and retinal hamartomas.

TUBEROUS SCLEROSIS (BOURNEVILLE DISEASE)

This is an autosomal dominant disorder with an incidence of ~1 in 5000–10,000 live births. It is caused by mutations in either the *TSC1* gene, which encodes a protein termed hamartin, or the *TSC2* gene, which encodes the protein tuberin. Hamartin forms a complex with tuberin, which inhibits cellular signaling through the mammalian target of rapamycin (mTOR), and acts as a negative regulator of the cell cycle. Patients with tuberous sclerosis may have seizures, intellectual disability, adenoma sebaceum (facial angiofibromas), shagreen patch, hypomelanotic macules, periungual fibromas, renal angiomyolipomas, and cardiac rhabdomyomas. These patients have an increased incidence of subependymal nodules, cortical tubers, and SEGAs. Patients frequently require anticonvulsants for seizures. SEGAs do not always require therapeutic intervention, but the most effective therapy is with the mTOR inhibitors sirolimus or everolimus, which often decrease seizures as well as SEGA size.

TUMORS METASTATIC TO THE BRAIN

Brain metastases arise from hematogenous spread and frequently originate from a lung primary or are associated with pulmonary metastases. Most metastases develop at the gray matter–white matter junction in the watershed distribution of the brain where intravascular tumor cells lodge in terminal arterioles. The distribution of metastases in the brain approximates the proportion of blood flow such that ~85% of all metastases are supratentorial and 15% occur in the posterior fossa. The most common sources of brain metastases are lung and breast carcinomas; melanoma has the greatest propensity to metastasize to the brain, being found in 80% of patients at autopsy **(Table 95-4)**. Other tumor types such as ovarian and esophageal carcinoma rarely metastasize to the brain. Prostate and breast cancers also have a propensity to metastasize to the dura and can mimic meningioma. Leptomeningeal metastases are common from hematologic malignancies and also breast and lung cancers. Spinal cord compression primarily arises in patients with prostate and breast cancer, tumors with a strong propensity to metastasize to the axial skeleton.

DIAGNOSIS OF METASTASES

Brain metastases are best visualized on MRI, where they usually appear as well-circumscribed lesions **(Fig. 95-7)**. The amount of perilesional edema can be highly variable, with large lesions causing minimal edema and sometimes very small lesions causing extensive edema. Enhancement may be in a ring pattern or diffuse. Occasionally, intracranial metastases will hemorrhage; melanoma, thyroid, and kidney cancer have the greatest propensity to hemorrhage, but the most common cause of a hemorrhagic metastasis is lung cancer because it accounts for the majority of brain metastases. The radiographic appearance of brain metastasis is nonspecific, and similar-appearing lesions can occur with infection including brain abscesses, demyelinating lesions, sarcoidosis, radiation necrosis in a previously treated patient, or a primary brain tumor that may be a second malignancy in a patient with systemic cancer. Biopsy is rarely necessary for diagnosis because imaging alone in the appropriate clinical situation usually suffices. However, in ~10% of patients, a systemic cancer may present with a brain metastasis, and if there is not an easily accessible systemic site to biopsy, a brain lesion must be removed for diagnostic purposes.

TABLE 95-4 Frequency of Nervous System Metastases for Common Primary Tumors

	BRAIN (%)	LM (%)	ESCC (%)
Lung	41	17	15
Breast	19	57	22
Melanoma	10	12	4
Prostate	1	1	10
GIT	7	—	5
Renal	3	2	7
Lymphoma	<1	10	10
Sarcoma	7	1	9
Other	11	—	18

Abbreviations: ESCC, epidural spinal cord compression; GIT, gastrointestinal tract; LM, leptomeningeal metastases.

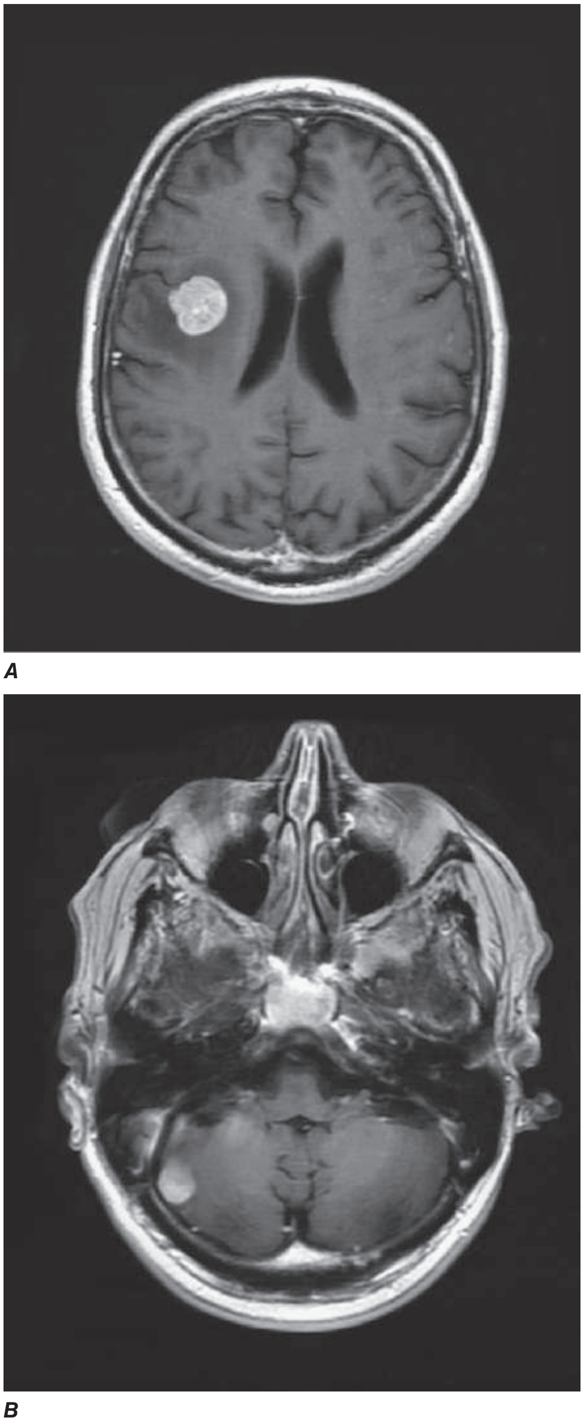

FIGURE 95-7 Postgadolinium T1 MRI of multiple brain metastases from non-small-cell lung cancer involving the right frontal (***A***) and right cerebellar (***B***) hemispheres. Note the diffuse enhancement pattern and absence of central necrosis.

TREATMENT

Tumors Metastatic to the Brain

DEFINITIVE TREATMENT

The number and location of brain metastases often determine the therapeutic options. The patient's overall condition and current or potential control of systemic disease are also major determinants. Brain metastases are single in approximately one-half of patients and multiple in the other half.

RADIATION THERAPY

The standard treatment for brain metastases has previously been WBRT usually administered to a total dose of 3000 cGy in 10 fractions. This affords rapid palliation, and ~80% of patients improve with glucocorticoids and RT. However, it is not curative, is associated with neurocognitive toxicity, and produces median survival of only 4–6 months. Hippocampal avoidance during WBRT can preserve cognitive function without increasing the risk of an intracranial relapse. The use of WBRT has declined with the development of more effective systemic options and access to SRS. If feasible, SRS has become the primary radiation oncology approach to brain metastases. It can be delivered through a variety of equally effective techniques including the gamma knife, linear accelerator, proton beam, or CyberKnife, all of which can deliver highly focused doses of RT, usually in a single fraction. SRS can effectively sterilize the visible lesions and afford local disease control in 80–90% of patients. Some patients have been cured of their brain metastases using SRS, whereas this is distinctly rare with WBRT. Traditionally SRS was used only for patients with 1–3 metastases, but recent data suggest that SRS can effectively treat 10 or more lesions. It is, however, confined to lesions of ≤3 cm and is most effective in metastases of ≤1 cm. The addition of WBRT to SRS improves disease control in the nervous system but does not prolong survival and thus is rarely employed.

SURGERY

Randomized controlled trials have demonstrated that surgical extirpation of a single brain metastasis followed by WBRT is superior to WBRT alone. Removal of two lesions or a single symptomatic mass, particularly if compressing the ventricular system, can also be useful. This is particularly important in patients who have highly radioresistant lesions such as renal carcinoma. Surgical resection can produce rapid amelioration of symptoms, improve control of edema, and result in prolonged survival. WBRT administered after complete resection of a brain metastasis improves disease control but does not prolong survival. There is increasing use of focal RT or even SRS to a resected cavity, especially if there is concern that tumor has been left behind, but most avoid postoperative WBRT because of its cognitive effects.

CHEMOTHERAPY

Chemotherapy and targeted therapy are becoming increasingly useful for brain metastases. Metastases from tumor types that are highly chemosensitive, such as germ cell tumors or small-cell lung cancer, may respond to chemotherapeutic regimens chosen according to the underlying malignancy. Increasingly, data demonstrate responsiveness of brain metastases to targeted therapeutics, such as for patients with lung cancer harboring *EGFR* mutations that sensitize them to EGFR inhibitors. Immunotherapy is also effective against those primary tumors that are sensitive to this approach, such as melanoma. Antiangiogenic agents such as bevacizumab are effective in the treatment of CNS metastases in those primary tumors for which it is approved.

LEPTOMENINGEAL METASTASES

Leptomeningeal metastases are also described as carcinomatous meningitis, meningeal carcinomatosis, or, in the case of specific tumors, leukemic or lymphomatous meningitis. Among the hematologic malignancies, acute leukemias most commonly metastasize to the subarachnoid space, followed in frequency by aggressive diffuse lymphomas. Among solid tumors, breast and lung carcinomas and melanoma most frequently spread in this fashion. Tumor cells reach the subarachnoid space via the arterial circulation or occasionally through retrograde flow in venous systems that drain metastases along the bony spine or cranium. In addition, leptomeningeal metastases may develop as a direct consequence of prior brain metastases and occur in almost 40% of patients who have a metastasis resected from the cerebellum.

CLINICAL FEATURES

Leptomeningeal metastases are characterized by multilevel symptoms and signs along the neuraxis. Combinations of lumbar and cervical radiculopathies, cranial neuropathies, seizures, confusion, and encephalopathy from hydrocephalus or raised intracranial pressure can be present. Focal deficits such as hemiparesis or aphasia are rarely due to leptomeningeal metastases unless there is direct brain infiltration. New-onset limb pain in patients with breast cancer, lung cancer, or melanoma should prompt consideration of leptomeningeal spread.

LABORATORY AND IMAGING DIAGNOSIS

Leptomeningeal metastases are particularly challenging to diagnose because identification of tumor cells in the subarachnoid compartment may be elusive. MRI can be definitive when there are clear tumor nodules adherent to the cauda equina or spinal cord, enhancing cranial nerves, or subarachnoid enhancement on brain imaging **(Fig. 95-8)**. Imaging is diagnostic in ~75% of patients and is more often positive in patients with solid tumors. Demonstration of tumor cells in the CSF is definitive and often considered the gold standard. However, CSF cytologic examination is positive in only 50% of patients on the first lumbar puncture and still misses 10% after three CSF samples. New technologies, such as rare cell capture, enhance identification of tumor cells in the CSF; molecular profiling of the CSF can also identify tumor-specific mutations, indicating malignancy in the leptomeninges. CSF cytologic examination is most useful in hematologic malignancies, especially when combined with flow cytometry to identify a clonal population. Accompanying CSF abnormalities include an elevated protein concentration and an elevated white blood cell count; hypoglycorrhachia is noted in <25% of patients but is useful when present. Identification of tumor markers may be helpful in some solid tumors.

TREATMENT

Leptomeningeal Metastases

The treatment of leptomeningeal metastasis is palliative because there is no curative therapy. RT to the symptomatically involved areas, such as skull base for cranial neuropathy, can relieve pain and sometimes improve function. Craniospinal irradiation (CSI) was previously avoided because it has significant toxicity with myelosuppression and gastrointestinal irritation as well as limited effectiveness. However, recent data on proton beam CSI suggest better disease control with fewer systemic toxicities. Systemic chemotherapy, targeted therapeutics, and immunotherapy have all demonstrated limited efficacy in the appropriate setting. Alternatively, intrathecal chemotherapy can be effective, particularly in hematologic malignancies. This is optimally delivered through an intraventricular cannula (Ommaya reservoir) rather than by lumbar puncture. Few drugs can be delivered safely into the subarachnoid space, and they have a limited spectrum of antitumor activity, perhaps accounting for the relatively poor response to this approach, particularly in solid tumors. In addition, impaired CSF flow dynamics can compromise intrathecal drug delivery. Surgery has a limited role in leptomeningeal metastasis. A ventriculoperitoneal shunt can relieve raised intracranial pressure but complicates the use of intrathecal drugs.

EPIDURAL METASTASIS

Epidural metastasis occurs in 3–5% of patients with a systemic malignancy and causes neurologic compromise by compressing the spinal cord or cauda equina. The most common cancers that metastasize to

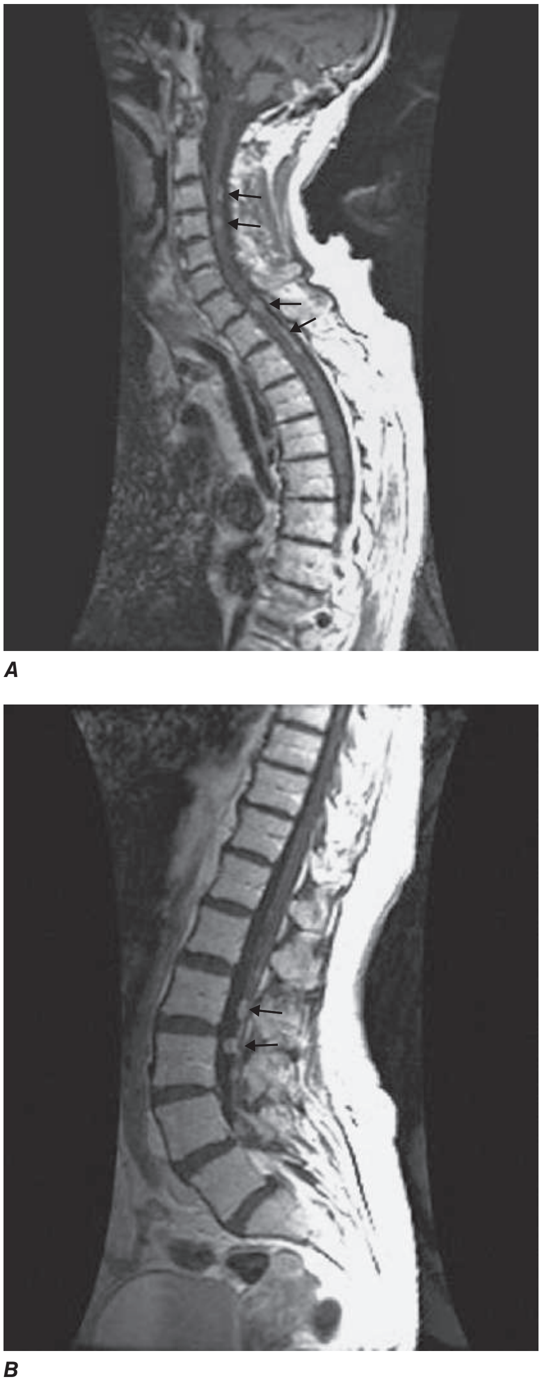

FIGURE 95-8 Postgadolinium MRI images of extensive leptomeningeal metastases from breast cancer. Nodules along the dorsal surface of the spinal cord (***A***) and cauda equina (***B***) are seen.

the epidural space are those malignancies that spread to bone, such as breast and prostate. Lymphoma can cause bone involvement and compression, but it can also invade an intervertebral foramen and cause spinal cord compression without bone destruction. The thoracic spine is affected most commonly, followed by the lumbar and then cervical spine.

■ CLINICAL FEATURES

Back pain is the presenting symptom of epidural metastasis in virtually all patients; the pain may precede neurologic findings by weeks or months. The pain is usually exacerbated by lying down; by contrast, arthritic pain is often relieved by recumbency. Leg weakness is seen in ~50% of patients, as is sensory dysfunction. Sphincter problems are present in ~25% of patients at diagnosis.

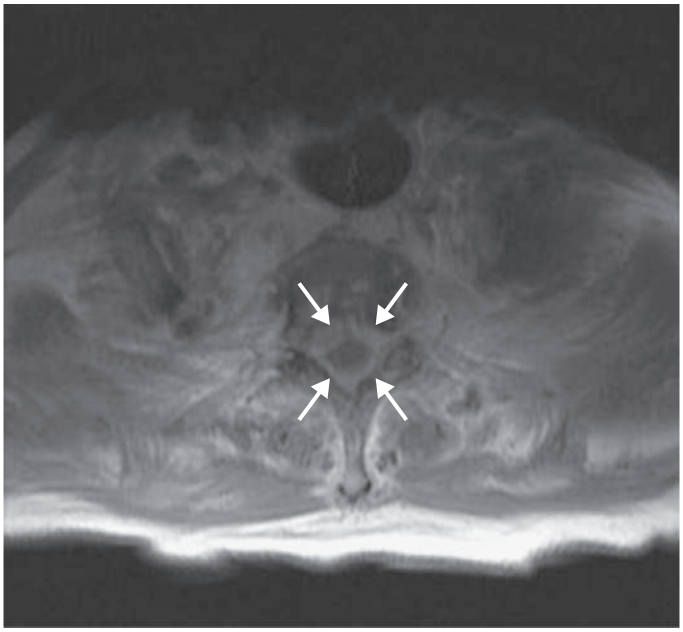

FIGURE 95-9 Postgadolinium T1 MRI showing circumferential epidural tumor around the thoracic spinal cord from esophageal cancer.

■ DIAGNOSIS

Diagnosis is established by imaging, preferably with an MRI of the entire spine (Fig. 95-9). Any patient with cancer who has severe back pain should undergo an MRI. Plain films, bone scans, or even CT scans may show bone metastases, but only MRI can reliably delineate epidural tumor. For patients unable to have an MRI, CT myelography should be performed to outline the epidural space. The differential diagnosis of epidural tumor includes epidural abscess, acute or chronic hematomas, epidural lipomatosis, and, rarely, extramedullary hematopoiesis.

TREATMENT

Epidural Metastasis

Epidural metastasis requires immediate treatment. A randomized controlled trial demonstrated the superiority of surgical resection followed by RT compared to RT alone. However, patients must be able to tolerate surgery, and the surgical procedure of choice is a complete removal of the mass, which is typically anterior to the spinal canal, necessitating an extensive approach and resection. Otherwise, RT is the mainstay of treatment and can be used for patients with radiosensitive tumors, such as lymphoma, or for those unable to undergo surgery. SRS or stereotactic body radiotherapy is increasingly being used, especially for radioresistant tumor types or for reirradiation. Chemotherapy is rarely used for epidural metastasis unless the patient has minimal to no neurologic deficit and a highly chemosensitive tumor such as lymphoma or germinoma. Patients generally fare well if treated before there is a severe neurologic deficit. Recovery from paraparesis is better after surgery than with RT alone, but survival is often short due to widespread metastatic tumor.

NEUROLOGIC TOXICITY OF THERAPY

■ TOXICITY FROM RADIOTHERAPY

RT can cause a variety of toxicities in the CNS. These are usually described based on their relationship in time to the administration of RT: acute (occurring within days of RT), early delayed (months), or late delayed (years). In general, the acute and early delayed syndromes resolve and do not result in persistent deficits, whereas the late delayed toxicities are usually permanent and sometimes progressive.

Acute Toxicity Acute cerebral toxicity may occur during RT to the brain. RT can cause a transient disruption of the blood-brain barrier, resulting in edema and elevated intracranial pressure. This is usually manifest as headache, lethargy, nausea, and vomiting and can be both prevented and treated with the administration of glucocorticoids.

Early Delayed Toxicity Early delayed toxicity is usually apparent weeks to months after completion of cranial irradiation and is likely due to focal demyelination. Clinically it may be asymptomatic or take the form of worsening or reappearance of a preexisting neurologic deficit. At times, a contrast-enhancing lesion can be seen on MRI/CT that can mimic the tumor for which the patient received the RT. For patients with a malignant glioma, this has been described as "pseudoprogression" because it mimics tumor recurrence on MRI, but it represents inflammation and necrotic debris engendered by effective therapy. This is seen with increased frequency when chemotherapy, particularly temozolomide, is given concurrently with RT. Pseudoprogression can resolve on its own or, if very symptomatic, may require glucocorticoids, resection, or bevacizumab.

In the spinal cord, early delayed RT toxicity is manifest as a Lhermitte symptom with paresthesias of the limbs or along the spine when the patient flexes the neck. Although frightening, it is benign, resolves on its own, and does not portend more serious problems.

Late Delayed Toxicity Late delayed toxicities are the most serious because they are often irreversible and cause severe neurologic deficits. In the brain, late toxicities can take several forms, the most common of which include radiation necrosis and leukoencephalopathy. Radiation necrosis is a focal mass of necrotic tissue that is contrast enhancing on CT/MRI and may be associated with significant edema. This may appear identical to pseudoprogression but is seen months to years after RT and is always symptomatic. Clinical symptoms and signs include seizures and findings referable to the location of the necrotic mass. The necrosis is caused by the effect of RT on cerebral vasculature with fibrinoid necrosis and occlusion of blood vessels. It can mimic tumor radiographically, but unlike tumor, it is typically hypometabolic on a PET scan and has reduced cerebral blood volume on perfusion MR sequences. It may require resection for diagnosis and treatment unless it can be managed with glucocorticoids. There are reports of improvement with bevacizumab and laser interstitial thermal therapy.

Leukoencephalopathy is seen most commonly after WBRT as opposed to focal RT. On T2 or FLAIR MR sequences, there is diffusely increased signal seen throughout the hemispheric white matter, often bilaterally and symmetrically. There tends to be a periventricular predominance that may be associated with atrophy and ventricular enlargement. Clinically, patients develop cognitive impairment, a gait disorder, and later urinary incontinence, all of which can progress over time. These symptoms mimic those of normal pressure hydrocephalus, and placement of a ventriculoperitoneal shunt can improve function in some patients but does not reverse the deficits completely. Increased age is a risk factor for leukoencephalopathy but not for radiation necrosis. Necrosis appears to depend on an unidentified predisposition.

Other late neurologic toxicities include endocrine dysfunction if the pituitary or hypothalamus was included in the RT port. An RT-induced neoplasm can occur many years after therapeutic RT for either a prior CNS or a head and neck tumor; accurate diagnosis requires surgical resection or biopsy. In addition, RT causes accelerated atherosclerosis, which can cause stroke either from intracranial vascular disease or carotid plaque from neck irradiation.

The peripheral nervous system is relatively resistant to RT toxicities. Peripheral nerves are rarely affected by RT, but the plexus is more vulnerable. Plexopathy develops more commonly in the brachial than in the lumbosacral distribution. It must be differentiated from tumor progression in the plexus, which is usually visualized by CT/MRI or PET scan demonstrating tumor infiltrating the region. Clinically, tumor progression is usually painful, whereas RT-induced plexopathy is painless. Radiation plexopathy is also more commonly associated with lymphedema and myokymia of the affected limb. Sensory loss and weakness are seen in both.

TABLE 95-5 Neurologic Toxicities Caused by Agents Commonly Used in Patients with Cancer

- Acute encephalopathy (delirium)
  - Methotrexate (high-dose IV, IT)
  - Cisplatin
  - Vincristine
  - Asparaginase
  - Procarbazine
  - 5-Fluorouracil (± levamisole)
  - Cytarabine (high-dose)
  - Nitrosoureas (high-dose or arterial)
  - Ifosfamide
  - Etoposide (high-dose)
  - Bevacizumab (PRES)
  - CAR-T cells
- Chronic encephalopathy (dementia)
  - Methotrexate
  - Carmustine
  - Cytarabine
  - Fludarabine
- Visual loss
  - Tamoxifen
  - Gallium nitrate
  - Cisplatin
  - Fludarabine
- Cerebellar dysfunction/ataxia
  - 5-Fluorouracil (± levamisole)
  - Cytarabine
  - Procarbazine
- Seizures
  - Methotrexate
  - Etoposide (high-dose)
  - Cisplatin
  - Vincristine
  - Asparaginase
  - Nitrogen mustard
  - Carmustine
  - Dacarbazine (intraarterial or high-dose)
  - Busulfan (high-dose)
- Myelopathy (IT drugs)
  - Methotrexate
  - Cytarabine
  - Thiotepa
- Peripheral neuropathy
  - Vinca alkaloids
  - Cisplatin
  - Procarbazine
  - Etoposide
  - Teniposide
  - Cytarabine
  - Taxanes
  - Suramin
  - Bortezomib

Abbreviations: CAR, chimeric antigen receptor; IT, intrathecal; IV, intravenous; PRES, posterior reversible encephalopathy syndrome.

TOXICITY FROM CHEMOTHERAPY

Neurotoxicity is second to myelosuppression as the dose-limiting toxicity of chemotherapeutic agents (Table 95-5). Chemotherapy causes peripheral neuropathy from many commonly used agents, and the type of neuropathy can vary depending on the drug. Vincristine causes paresthesias but little sensory loss and is associated with motor dysfunction, autonomic impairment (frequently ileus), and, rarely, cranial nerve compromise. Cisplatin causes large-fiber sensory loss resulting in sensory ataxia but little cutaneous sensory loss and no weakness. The taxanes also cause a predominately sensory neuropathy. Agents such as bortezomib and thalidomide also cause neuropathy. Sometimes a severe neuropathy emerges after multiple neurotoxic agents have been used together or in sequence.

Encephalopathy and seizures are common toxicities from chemotherapeutic drugs. Ifosfamide can cause a severe encephalopathy, which is reversible with discontinuation of the drug. Fludarabine also causes a severe global encephalopathy that may be permanent. Bevacizumab and other anti-VEGF agents can cause posterior reversible encephalopathy syndrome. Cisplatin can cause hearing loss and less frequently vestibular dysfunction. Immunotherapy with monoclonal antibodies such as ipilimumab or nivolumab can cause an autoimmune hypophysitis, Guillain-Barré syndrome, or an autoimmune encephalitis. CAR-T cells frequently cause a reversible encephalopathy due to an immune effector cell–associated neurotoxicity syndrome (Chap. 318).

FURTHER READING

Aizer AA et al: Brain metastases: A Society for Neuro-Oncology (SNO) consensus review on current management and future directions. Neuro Oncol 24:1613, 2022.

Grommes C et al: Comprehensive approach to diagnosis and treatment of newly diagnosed primary CNS lymphoma. Neuro Oncol 21:296, 2019.

Lamba N et al: Epidemiology of brain metastases and leptomeningeal disease. Neuro Oncol 23:1447, 2021.

Louis DN et al: The 2021 WHO classification of tumors of the central nervous system: A summary. Neuro-Oncology 23:1231, 2021.

Mellinghoff IK et al: Vorasidenib in IDH1- or IDH2-mutant low-grade glioma. N Engl J Med 389:589, 2023.

Ostrom QT et al: CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2016-2020. Neuro Oncol 25:iv1-iv99, 2023.

Schiff D et al: Recent developments and future directions in adult lower-grade gliomas: Society for Neuro-Oncology (SNO) and European Association of Neuro-Oncology (EANO) consensus. Neuro Oncol 21:837, 2019.

Wen PY et al: Glioblastoma in adults: A Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. Neuro Oncol 22:1073, 2020.

96

Soft Tissue and Bone Sarcomas and Bone Metastases

Shreyaskumar R. Patel

Sarcomas are rare (<1% of all malignancies) mesenchymal neoplasms that arise in bone and soft tissues. These tumors are usually of mesodermal origin, although a few are derived from neuroectoderm, and they are biologically distinct from the more common epithelial malignancies. Sarcomas affect all age groups; 15% are found in children <15 years of age, and 40% occur after age 55 years. Sarcomas are one of the most common solid tumors of childhood and are the fifth most common cause of cancer deaths in children. Sarcomas may be divided into two groups, those derived from bone and those derived from soft tissues.

SOFT TISSUE SARCOMAS

Soft tissues include muscles, tendons, fat, fibrous tissue, synovial tissue, vessels, and nerves. Approximately 60% of soft tissue sarcomas arise in the extremities, with the lower extremities involved three times as often as the upper extremities. Thirty percent arise in the trunk, with the retroperitoneum accounting for 40% of all trunk lesions. The remaining 10% arise in the head and neck.

INCIDENCE

Approximately 13,400 new cases of soft tissue sarcomas occurred in the United States in 2023. The annual age-adjusted incidence is 3 per 100,000 population, but the incidence varies with age. Soft tissue sarcomas constitute 0.7% of all cancers in the general population and 6.5% of all cancers in children.

EPIDEMIOLOGY

Malignant transformation of a benign soft tissue tumor is extremely rare, with the exception that malignant peripheral nerve sheath tumors (neurofibrosarcoma, malignant schwannoma) can arise from neurofibromas in patients with neurofibromatosis. Several etiologic factors have been implicated in soft tissue sarcomas.

Environmental Factors Trauma or previous injury is rarely involved, but sarcomas can arise in scar tissue resulting from a prior operation, burn, fracture, or foreign body implantation. Chemical carcinogens such as polycyclic hydrocarbons, asbestos, and dioxin may be involved in the pathogenesis.

Iatrogenic Factors Sarcomas in bone or soft tissues occur in patients who are treated with radiation therapy. The tumor nearly always arises in the irradiated field. The risk increases with time.

Viruses Kaposi's sarcoma (KS) in patients with HIV type 1, classic KS, and KS in HIV-negative homosexual men is caused by human herpesvirus (HHV) 8 **(Chap. 200)**. No other sarcomas are associated with viruses.

Immunologic Factors Congenital or acquired immunodeficiency, including therapeutic immunosuppression, increases the risk of sarcoma.

GENETIC CONSIDERATIONS

Li-Fraumeni syndrome is a familial cancer syndrome in which affected individuals have germline abnormalities of the tumor-suppressor gene *p53* and an increased incidence of soft tissue sarcomas and other malignancies, including breast cancer, osteosarcoma, brain tumor, leukemia, and adrenal carcinoma **(Chap. 76)**. Neurofibromatosis 1 (NF-1; peripheral form, von Recklinghausen's disease) is characterized by multiple neurofibromas and café-au-lait spots. Neurofibromas occasionally undergo malignant degeneration to become malignant peripheral nerve sheath tumors. The gene for *NF1* is located in the pericentromeric region of chromosome 17 and encodes neurofibromin, a tumor-suppressor protein with guanosine 5′-triphosphate (GTP)ase-activating activity that inhibits ras function **(Chap. 95)**. Germline mutation of the *RB1* locus (chromosome 13q14) in patients with inherited retinoblastoma is associated with the development of osteosarcoma in those who survive the retinoblastoma and of soft tissue sarcomas unrelated to radiation therapy. Other soft tissue tumors, including desmoid tumors, lipomas, leiomyomas, neuroblastomas, and paragangliomas, occasionally show a familial predisposition.

Insulin-like growth factor (IGF) type II is produced by some sarcomas and may act as an autocrine growth factor and as a motility factor that promotes metastatic spread. IGF-II stimulates growth through IGF-I receptors, but its effects on motility are through different receptors. If secreted in large amounts, IGF-II may produce hypoglycemia **(Chaps. 98 and 418)**. A large international sarcoma kindred study including 1162 patients and 6545 Caucasian controls revealed that about half the patients with sarcoma have putatively pathogenic monogenic and polygenic variation in previously reported and new cancer genes, some of them representing therapeutically actionable targets. These patients were diagnosed with sarcoma at an earlier age compared to controls.

CLASSIFICATION

Approximately 20 different groups of sarcomas are recognized on the basis of the pattern of differentiation toward normal tissue. For example, rhabdomyosarcoma shows evidence of skeletal muscle fibers with cross-striations; leiomyosarcomas contain interlacing fascicles of spindle cells resembling smooth muscle; and liposarcomas contain adipocytes. When precise characterization of the group is not possible, the tumors are called *unclassified sarcomas*. All of the primary bone sarcomas can also arise from soft tissues (e.g., extraskeletal osteosarcoma). The entity *malignant fibrous histiocytoma* (MFH) includes many tumors previously classified as fibrosarcomas or as pleomorphic variants of other sarcomas and is characterized by a mixture of spindle (fibrous) cells and round (histiocytic) cells arranged in a storiform pattern with frequent giant cells and areas of pleomorphism. As immunohistochemical suggestion of differentiation, particularly myogenic differentiation, may be found in a significant fraction of these patients, many are now characterized as poorly differentiated leiomyosarcomas, and the terms *undifferentiated pleomorphic sarcoma* (UPS) and *myxofibrosarcoma* are replacing MFH and myxoid MFH.

For purposes of treatment, most soft tissue sarcomas can be considered together. However, some specific tumors have distinct features. For example, *liposarcoma* can have a spectrum of behaviors. Pleomorphic liposarcomas and dedifferentiated liposarcomas behave like other high-grade sarcomas; in contrast, well-differentiated liposarcomas (better termed *atypical lipomatous tumors*) lack metastatic potential, and myxoid liposarcomas metastasize infrequently, but, when they do, they have a predilection for unusual metastatic sites containing fat, such as the retroperitoneum, mediastinum, and subcutaneous tissue. Rhabdomyosarcomas, Ewing's sarcoma, and other small-cell sarcomas tend to be more aggressive and are more responsive to chemotherapy than other soft tissue sarcomas.

Approximately a third of all soft tissue sarcomas have a translocation that may have diagnostic and prognostic relevance; for example, 90% of synovial sarcomas contain a characteristic chromosomal translocation t(X;18)(p11;q11) involving a nuclear transcription factor on chromosome 18 called *SYT* and two breakpoints on X. Patients with translocations to the second X breakpoint (*SSX2*) may have longer survival than those with translocations involving *SSX1*. Targeting these translocations for therapy is an area of ongoing investigation.

Gastrointestinal stromal tumors (GISTs), previously classified as gastrointestinal leiomyosarcomas, are now recognized as a distinct entity within soft tissue sarcomas. Its cell of origin resembles the interstitial cell of Cajal, which controls peristalsis. The majority of malignant GISTs have activating mutations of the c-*kit* gene that result in ligand-independent phosphorylation and activation of the KIT receptor tyrosine kinase, leading to tumorigenesis. Approximately 5–10% of tumors will have a mutation in the platelet-derived growth factor receptor α (*PDGFRA*). GISTs that are wild type for both *KIT* and *PDGFRA* mutations may show mutations in *SDH* B, C, or D and may be driven by the IGF-I pathway.

■ DIAGNOSIS

The most common presentation is an asymptomatic mass. Mechanical symptoms referable to pressure, traction, or entrapment of nerves or muscles may be present. All new and persistent or growing masses should be biopsied, either by a small incision or by a cutting needle (core-needle biopsy) placed so that it can be encompassed in the subsequent excision without compromising a definitive resection. Lymph node metastases occur in 5%, except in synovial and epithelioid sarcomas, clear cell sarcoma (melanoma of the soft parts), angiosarcoma, and rhabdomyosarcoma, where nodal spread may be seen in 17%. The pulmonary parenchyma is the most common site of metastases. Exceptions are GISTs, which metastasize to the liver; myxoid liposarcomas, which seek fatty tissue; and clear cell sarcomas, which may metastasize to bones. Central nervous system metastases are rare, except in alveolar soft part sarcoma.

Radiographic Evaluation Imaging of the primary tumor is best with plain radiographs and magnetic resonance imaging (MRI) for tumors of the extremities or head and neck and by computed tomography (CT) for tumors of the chest, abdomen, or retroperitoneal cavity. A radiograph and CT scan of the chest are important for the detection of lung metastases. Other imaging studies may be indicated, depending on the symptoms, signs, or histology.

■ STAGING AND PROGNOSIS

The histologic grade and size of the primary tumor are the most important prognostic factors. The current American Joint Committee on Cancer (AJCC) staging system is shown in **Table 96-1**. Prognosis is related to the stage. Cure is common in the absence of metastatic disease, but a small number of patients with metastases can also be cured. Historically, most patients with stage IV disease used to die within 12 months, but with availability of multiple lines of treatments, median survival in second-line and beyond ranges from 13 to 18 months, and some patients may live with stable or slowly progressive disease for many years.

TREATMENT

Soft Tissue Sarcomas

AJCC stage I patients are adequately treated with surgery alone. Stage II patients are considered for adjuvant radiation therapy. Stage III patients may benefit from neoadjuvant or adjuvant chemotherapy. Stage IV patients are managed primarily with systemic therapy, with or without other modalities.

TABLE 96-1 American Joint Commission on Cancer Staging System for Sarcomas, Eighth Edition

T1	Tumor ≤5 cm in greatest dimension
T2	Tumor >5 cm and ≤10 cm in greatest dimension
T3	Tumor >10 cm and ≤15 cm in greatest dimension
T4	Tumor >15 cm in greatest dimension
N0	No regional lymph node metastasis or unknown lymph node status
N1	Regional lymph node metastasis
M0	No distant metastasis
M1	Distant metastasis
Stage Groups	
Stage IA	T1; N0; M0; G1
Stage IB	T2, T3, T4; N0; M0; G1
Stage II	T1; N0; M0; G2/3
Stage IIIA	T1A, T2; N0; M0; G2/3
Stage IIIB	T3, T4; N0; M0; G2/3
Stage IV	Any T; N1; M0; any G Any T; any N; M1; any G

SURGERY

Soft tissue sarcomas tend to grow along fascial planes, with the surrounding soft tissues compressed to form a pseudocapsule that gives the sarcoma the appearance of a well-encapsulated lesion. This is invariably deceptive because "shelling out," or marginal excision, of such lesions results in a 50–90% probability of local recurrence. Wide excision with a negative margin, incorporating the biopsy site, is the standard surgical procedure for local disease. The adjuvant use of radiation therapy and/or chemotherapy improves the local control rate and permits the use of limb-sparing surgery with a local control rate (85–90%) comparable to that achieved by radical excisions and amputations. Limb-sparing approaches are indicated except when negative margins are not obtainable, when the risks of radiation are prohibitive, or when neurovascular structures are involved so that resection will result in serious functional consequences to the limb.

RADIATION THERAPY

External-beam radiation therapy is an adjuvant to limb-sparing surgery for improved local control. Preoperative radiation therapy allows the use of smaller fields and smaller doses but results in a higher rate of wound complications. Postoperative radiation therapy must be given to larger fields, because the entire surgical bed must be encompassed, and in higher doses to compensate for hypoxia in the operated field. This results in a higher rate of late complications. Brachytherapy or interstitial therapy, in which the radiation source is inserted into the tumor bed, is comparable in efficacy (except in low-grade lesions), less time consuming, and less expensive.

With the advent of stereotactic body radiotherapy (SBRT), the role of radiation therapy in oligometastatic disease in various visceral sites is evolving.

ADJUVANT CHEMOTHERAPY

Chemotherapy is the mainstay of treatment for Ewing's sarcomas/primitive neuroectodermal tumors (PNETs) and rhabdomyosarcomas. Meta-analysis of 14 randomized trials in non-small-cell sarcomas revealed a significant improvement in local control and disease-free survival in favor of doxorubicin-based chemotherapy. Overall survival improvement was 4% for all sites and 7% for the extremity site. An updated meta-analysis including four additional trials with doxorubicin and ifosfamide combination reported a statistically significant 6% survival advantage in favor of chemotherapy. A chemotherapy regimen including an anthracycline and ifosfamide with growth factor support improved overall survival by 19% for high-risk (high-grade, ≥5 cm primary, or locally recurrent) extremity soft tissue sarcomas. Long-term follow-up of a trial evaluating neoadjuvant use of the same combination confirms survival advantage and reports a 10-year survival of 61%. A more contemporary randomized trial compared the standard anthracycline and ifosfamide combination to specific histology-tailored chemotherapy as an active control and confirmed superiority of the standard regimen.

ADVANCED DISEASE

Metastatic soft tissue sarcomas are largely incurable, but up to 20% of patients who achieve a complete response become long-term survivors. The therapeutic intent, therefore, is to produce a complete remission with chemotherapy (<10%) and/or surgery (30–40%). Surgical resection of metastases, whenever possible, is an integral part of the management. Some patients benefit from repeated surgical excision of metastases. The two most active chemotherapeutic agents are doxorubicin and ifosfamide. These drugs show a steep dose-response relationship in sarcomas. Gemcitabine with or without docetaxel has become an established second-line regimen and is particularly active in patients with UPS and leiomyosarcomas. Dacarbazine also has some modest activity. Taxanes have selective activity in angiosarcomas, and vincristine, etoposide, and irinotecan are effective in rhabdomyosarcomas and Ewing's sarcomas. Pazopanib, an inhibitor of the vascular endothelial growth factor, platelet-derived growth factor (PDGF), and c-kit, is approved for patients with advanced soft tissue sarcomas excluding liposarcomas after failure of chemotherapy. Two additional chemotherapy drugs have gained approval from the U.S. Food and Drug Administration (FDA). Trabectedin was compared to dacarbazine in a large phase 3 randomized study in advanced leiomyosarcomas and liposarcomas after failure of an anthracycline and resulted in significant improvement in progression-free survival. Eribulin was also tested in a similar trial and showed improvement in survival, predominantly in the liposarcoma subgroup, and is therefore now approved for that subset. Tazemetostat, an EZH2 inhibitor, is now approved for use in metastatic epithelioid sarcomas characterized by loss of tumor-suppressor gene *INI1*, resulting in activation of the EZH2 pathway. Nab-sirolimus, an inhibitor of the mammalian target of rapamycin (mTOR) pathway, is now approved for treatment of metastatic perivascular epithelioid cell tumors (PEComas). The FDA granted approval for nirogacestat, a γ-secretase inhibitor, for the treatment of desmoid tumors. Imatinib targets KIT and PDGF tyrosine kinase activity and is standard therapy for advanced/metastatic GISTs and dermatofibrosarcoma protuberans. Imatinib is also indicated as adjuvant therapy for completely resected primary GISTs. Three years of adjuvant imatinib appear to be superior to 1 year of therapy for high-risk GISTs, although the optimal treatment duration remains unknown. Sunitinib and regorafenib are approved for second- and third-line use, respectively, in metastatic GIST after failure of or intolerance to imatinib. Ripretinib, an inhibitor of c-kit and PDGFRA, was approved for fourth-line use in metastatic GIST based on a placebo-controlled randomized trial reporting an improved median progression-free and overall survival. Avapritinib also received approval for use in the specific molecular subset of *PDGFRA* D842V–mutant metastatic GIST.

Immune checkpoint inhibitors have generally been ineffective in most forms of sarcoma, but atezolizumab, an anti-PD-L1 antibody, has produced responses in about one-third of patients with alveolar soft part sarcoma.

BONE SARCOMAS

INCIDENCE AND EPIDEMIOLOGY

Bone sarcomas are rarer than soft tissue sarcomas; they accounted for only 0.2% of all new malignancies and 3970 new cases in the United States in 2023. Several benign bone lesions have the potential for malignant transformation. Enchondromas and osteochondromas can transform into chondrosarcoma; fibrous dysplasia, bone infarcts, and Paget's disease of bone can transform into either UPS or osteosarcoma.

CLASSIFICATION

Benign Tumors The common benign bone tumors include enchondroma, osteochondroma, chondroblastoma, and chondromyxoid fibroma, of cartilage origin; osteoid osteoma and osteoblastoma, of bone origin; fibroma and desmoplastic fibroma, of fibrous tissue origin; hemangioma, of vascular origin; and giant cell tumor, of unknown origin.

Malignant Tumors The most common malignant tumors of bone are plasma cell tumors (Chap. 116). The four most common malignant nonhematopoietic bone tumors are osteosarcoma, chondrosarcoma, Ewing's sarcoma, and UPS. Rare malignant tumors include chordoma (of notochordal origin), malignant giant cell tumor, adamantinoma (of unknown origin), and hemangioendothelioma (of vascular origin).

Musculoskeletal Tumor Society Staging System Sarcomas of bone are staged according to the Musculoskeletal Tumor Society staging system based on grade and compartmental localization. A Roman numeral reflects the tumor grade: stage I is low grade, stage II is high grade, and stage III includes tumors of any grade that have lymph node or distant metastases. In addition, the tumor is given a letter reflecting its compartmental localization. Tumors designated A are intracompartmental (i.e., confined to the same soft tissue compartment as the initial tumor), and tumors designated B are extracompartmental (i.e., extending into the adjacent soft tissue compartment or into bone). The tumor-node-metastasis (TNM) staging system is shown in Table 96-2.

OSTEOSARCOMA

Osteosarcoma, accounting for almost 45% of all bone sarcomas, is a spindle cell neoplasm that produces osteoid (unmineralized bone) or bone. Approximately 60% of all osteosarcomas occur in children and

TABLE 96-2 Staging System for Bone Sarcomas

Primary tumor (T)	TX	Primary tumor cannot be assessed
	T0	No evidence of primary tumor
	T1	Tumor ≤8 cm in greatest dimension
	T2	Tumor >8 cm in greatest dimension
	T3	Discontinuous tumors in the primary bone site
Regional lymph nodes (N)	NX	Regional lymph nodes cannot be assessed
	N0	No regional lymph node metastasis
	N1	Regional lymph node metastasis
Distant metastasis (M)	MX	Distant metastasis cannot be assessed
	M0	No distant metastasis
	M1	Distant metastasis
	M1a	Lung
	M1b	Other distant sites
Histologic grade (G)	GX	Grade cannot be assessed
	G1	Well differentiated—low grade
	G2	Moderately differentiated—low grade
	G3	Poorly differentiated—high grade
	G4	Undifferentiated—high grade (Ewing's is always classed G4)

Stage Grouping				
Stage IA	T1	N0	M0	G1,2 low grade
Stage IB	T2	N0	M0	G1,2 low grade
Stage IIA	T1	N0	M0	G3,4 high grade
Stage IIB	T2	N0	M0	G3,4 high grade
Stage III	T3	N0	M0	Any G
Stage IVA	Any T	N0	M1a	Any G
Stage IVB	Any T	N1	Any M	Any G
	Any T	Any N	M1b	Any G

adolescents in the second decade of life, and ~10% occur in the third decade of life. Osteosarcomas in the fifth and sixth decades of life are frequently secondary to either radiation therapy or transformation in a preexisting benign condition, such as Paget's disease. Males are affected 1.5–2 times as often as females. Osteosarcoma has a predilection for metaphyses of long bones; the most common sites of involvement are the distal femur, proximal tibia, and proximal humerus. The classification of osteosarcoma is complex, but 75% of osteosarcomas fall into the "classic" category, which includes osteoblastic, chondroblastic, and fibroblastic osteosarcomas. The remaining 25% are classified as "variants" on the basis of (1) clinical characteristics, as in the case of osteosarcoma of the jaw, postradiation osteosarcoma, or Paget's osteosarcoma; (2) morphologic characteristics, as in the case of telangiectatic osteosarcoma, small-cell osteosarcoma, or epithelioid osteosarcoma; or (3) location, as in parosteal or periosteal osteosarcoma. Diagnosis usually requires a synthesis of clinical, radiologic, and pathologic features. Patients typically present with pain and swelling of the affected area. A plain radiograph reveals a destructive lesion with a moth-eaten appearance, a spiculated periosteal reaction (sunburst appearance), and a cuff of periosteal new bone formation at the margin of the soft tissue mass (Codman's triangle). A CT scan of the primary tumor is best for defining bone destruction and the pattern of calcification, whereas MRI is better for defining intramedullary and soft tissue extension. A chest radiograph and CT scan are used to detect lung metastases. Metastases to the bony skeleton should be imaged by a bone scan or by fluorodeoxyglucose positron emission tomography (FDG-PET). Almost all osteosarcomas are hypervascular and PET-avid. Pathologic diagnosis is established either with a core-needle biopsy, where feasible, or with an open biopsy with an appropriately placed incision that does not compromise future limb-sparing resection. Most osteosarcomas are high grade. The most important predictive factor for long-term survival is response to chemotherapy. Preoperative chemotherapy followed by limb-sparing surgery (which can be accomplished in >80% of patients) followed by postoperative chemotherapy is standard management. The effective drugs are doxorubicin, ifosfamide, cisplatin, and high-dose methotrexate with leucovorin rescue. The various combinations of these agents that have been used have all been about equally successful. Long-term survival rates in extremity osteosarcoma range from 60 to 80%. Osteosarcoma is radioresistant; radiation therapy has no role in the routine management. UPS is considered a part of the spectrum of osteosarcoma and is managed similarly. A randomized trial has shown improved progression-free survival with regorafenib compared to placebo.

CHONDROSARCOMA

Chondrosarcoma, which constitutes ~20–25% of all bone sarcomas, is a tumor of adulthood and old age, with a peak incidence in the fourth to sixth decades of life. It has a predilection for the flat bones, especially the shoulder and pelvic girdles, but can also affect the diaphyseal portions of long bones. Chondrosarcomas can arise de novo or as a malignant transformation of an enchondroma or, rarely, of the cartilaginous cap of an osteochondroma. Chondrosarcomas have an indolent natural history and typically present as pain and swelling. Radiographically, the lesion may have a lobular appearance with mottled or punctate or annular calcification of the cartilaginous matrix. It is difficult to distinguish low-grade chondrosarcoma from benign lesions by x-ray or histologic examination. The diagnosis is therefore influenced by clinical history and physical examination. A new onset of pain, signs of inflammation, and progressive increase in the size of the mass suggest malignancy. The histologic classification is complex, but most tumors fall within the classic category. Like other bone sarcomas, high-grade chondrosarcomas spread to the lungs. Most chondrosarcomas are resistant to chemotherapy, and surgical resection of primary or recurrent tumors, including pulmonary metastases, is the mainstay of therapy with the exception of two histologic variants. Dedifferentiated chondrosarcoma has a high-grade osteosarcoma or a UPS component that responds to chemotherapy. Mesenchymal chondrosarcoma, a rare variant composed of a small-cell element, also is responsive to systemic chemotherapy and is treated like Ewing's sarcoma.

EWING'S SARCOMA

Ewing's sarcoma, which constitutes ~10–15% of all bone sarcomas, is common in adolescence and has a peak incidence in the second decade of life. It typically involves the diaphyseal region of long bones and also has an affinity for flat bones. The plain radiograph may show a characteristic "onion peel" periosteal reaction with a generous soft tissue mass, which is better demonstrated by CT or MRI. This mass is composed of sheets of monotonous, small, round, blue cells and can be confused with lymphoma, embryonal rhabdomyosarcoma, and small-cell carcinoma. The presence of p30/32, the product of the *mic-2* gene (which maps to the pseudoautosomal region of the X and Y chromosomes), is a cell-surface marker for Ewing's sarcoma (and other members of the Ewing family of tumors, previously also called PNETs). Most PNETs arise in soft tissues; they include peripheral neuroepithelioma, Askin's tumor (chest wall), and esthesioneuroblastoma. The classic cytogenetic abnormality associated with this disease is a reciprocal translocation of the long arms of chromosomes 11 and 22, t(11;22), which creates a chimeric gene product of unknown function with components from the *fli-1* gene on chromosome 11 and *ews* on chromosome 22. This disease is very aggressive, and it is therefore considered a systemic disease. Common sites of metastases are lung, bones, and bone marrow. Systemic chemotherapy is the mainstay of therapy, often being used before surgery. Doxorubicin, cyclophosphamide or ifosfamide, etoposide, and vincristine are active drugs. Topotecan or irinotecan in combination with an alkylating agent is often used in relapsed patients. Local treatment for the primary tumor includes surgical resection, usually with limb salvage or radiation therapy. Overall cure rates approach 60–70% for localized tumors.

TUMORS METASTATIC TO BONE

Bone is a common site of metastasis for carcinomas of the prostate, breast, lung, kidney, bladder, and thyroid and for lymphomas and sarcomas. Prostate, breast, and lung primaries account for 80% of all bone metastases. Metastatic tumors of bone are more common than primary bone tumors. Tumors usually spread to bone hematogenously, but local invasion from soft tissue masses also occurs. In descending order of frequency, the sites most often involved are the vertebrae, proximal femur, pelvis, ribs, sternum, proximal humerus, and skull. Bone metastases may be asymptomatic or may produce pain, swelling, nerve root or spinal cord compression, pathologic fracture, or myelophthisis (replacement of the marrow). Symptoms of hypercalcemia may be noted in cases of bony destruction.

Pain is the most frequent symptom. It usually develops gradually over weeks, is usually localized, and often is more severe at night. When patients with back pain develop neurologic signs or symptoms, emergency evaluation for spinal cord compression is indicated **(Chap. 80)**. Bone metastases exert a major adverse effect on quality of life in cancer patients.

Cancer in the bone may produce osteolysis, osteogenesis, or both. Osteolytic lesions result when the tumor produces substances that can directly elicit bone resorption (vitamin D–like steroids, prostaglandins, or parathyroid hormone–related peptide) or cytokines that can induce the formation of osteoclasts (interleukin 1 and tumor necrosis factor). Osteoblastic lesions result when the tumor produces cytokines that activate osteoblasts. In general, purely osteolytic lesions are best detected by plain radiography, but they may not be apparent until they are >1 cm. These lesions are more commonly associated with hypercalcemia and with the excretion of hydroxyproline-containing peptides indicative of matrix destruction. When osteoblastic activity is prominent, the lesions may be readily detected using radionuclide bone scanning (which is sensitive to new bone formation), and the radiographic appearance may show increased bone density or sclerosis. Osteoblastic lesions are associated with higher serum levels of alkaline phosphatase and, if extensive, may produce hypocalcemia. Although some tumors may produce mainly osteolytic lesions (e.g., kidney cancer) and others mainly osteoblastic lesions (e.g., prostate cancer), most metastatic lesions produce both types of lesion and may go through stages where one or the other predominates.

In older patients, particularly women, it may be necessary to distinguish metastatic disease of the spine from osteoporosis. In osteoporosis, the cortical bone may be preserved, whereas cortical bone destruction is usually noted with metastatic cancer.

TREATMENT

Metastatic Bone Disease

Treatment of metastatic bone disease depends on the underlying malignancy and the symptoms. Some metastatic bone tumors are curable (lymphoma, Hodgkin's lymphoma), and others are treated with palliative intent. Pain may be relieved by local radiation therapy. Hormonally responsive tumors are responsive to hormone inhibition (antiandrogens for prostate cancer, antiestrogens for breast cancer). Strontium-89, samarium-153, and radium-223 are bone-seeking radionuclides that can exert antitumor effects and relieve symptoms. Denosumab, a monoclonal antibody that binds to RANK ligand, inhibits osteoclastic activity and increases bone mineral density. Bisphosphonates such as pamidronate may relieve pain and inhibit bone resorption, thereby maintaining bone mineral density and reducing risk of fractures in patients with osteolytic metastases from breast cancer and multiple myeloma. Careful monitoring of serum electrolytes and creatinine is recommended. Monthly administration prevents bone-related clinical events and may reduce the incidence of bone metastases in women with breast cancer. When the integrity of a weight-bearing bone is threatened by an expanding metastatic lesion that is refractory to radiation therapy, prophylactic internal fixation is indicated. Overall survival is related to the prognosis of the underlying tumor. Bone pain at the end of life is particularly common; an adequate pain relief regimen including sufficient amounts of narcotic analgesics is required. **The management of hypercalcemia is discussed in Chap. 422.**

■ FURTHER READING

Alvarez RA et al: Optimization of the therapeutic approach to patients with sarcoma: Delphi consensus. Sarcoma 2019:4351308, 2019.

Ballinger ML et al: Monogenic and polygenic determinants of sarcoma risk: An international genetic study. Lancet Oncol 17:1261, 2016.

Beird HC et al: Osteosarcoma. Nat Rev Dis Primers 8:77, 2022.

Gounder M et al: Nirogacestat, a γ-secretase inhibitor for desmoid tumors. N Engl J Med 388:898, 2023.

Meyer M, Seetharam M: First-line therapy for metastatic soft tissue sarcoma. Curr Treat Options Oncol 20:6, 2019.

Pasquali S, Gronchi A: Neoadjuvant chemotherapy in soft tissue sarcomas: Latest evidence and clinical implications. Ther Adv Med Oncol 9:415, 2017.

Patel S, Reichardt P: An updated review of the treatment landscape for advanced gastrointestinal stromal tumors. Cancer 127:2185, 2021.

Ratan R, Patel SR: Chemotherapy for soft tissue sarcoma. Cancer 122:2952, 2016.

Wagner MJ et al: Chemotherapy for bone sarcoma in adults. J Oncol Pract 12:208, 2016.

97

Carcinoma of Unknown Primary

Kanwal Raghav, James L. Abbruzzese

Carcinoma (or cancer) of unknown primary (CUP) is a biopsy-proven malignancy for which the anatomic site of origin remains unidentified after a standardized detailed diagnostic evaluation. CUP is one of the 10 most frequently diagnosed cancers globally, accounting for 3–5% of all malignancies. Most investigators limit CUP to epithelial or undifferentiated cancers and do not include lymphomas, metastatic melanomas, and metastatic sarcomas because these cancers have specific histology and stage-based management guidelines, even in the absence of a primary site. CUP can occur in patients of all age groups including adolescents and young adults.

The emergence of sophisticated imaging, robust immunohistochemistry (IHC), and genomic and proteomic tools has challenged the "unknown" designation. Additionally, effective targeted therapies in several cancers and tissue-agnostic biomarker-driven therapies have endorsed a change in paradigm from empiricism to a personalized approach to CUP management. The reasons cancers present as CUP remain unclear. One hypothesis is that the primary tumor either regresses after seeding the metastasis or remains so small that it is not detected. It is possible that CUP falls on the continuum of cancer presentation where the primary has been contained or eliminated by the natural body defenses, including the immune system. Alternatively, CUP may represent a specific malignant event that results in an increase in metastatic spread or survival relative to the primary. Whether the CUP metastases truly define a clone that is genetically and phenotypically unique to this diagnosis remains to be determined.

Since liver is a common site of CUP presentation, intrahepatic cholangiocarcinoma (ICC) can be often misdiagnosed as CUP. Of note, the incidence of ICC is increasing, whereas at the same time, that of CUP is declining. Improvements in diagnostic technologies including next-generation sequencing and other molecular techniques and awareness among clinicians to differentiate the two are possibly contributing to an increased recognition and incidence of ICC.

CUP BIOLOGY

Studies looking for unique signature abnormalities in CUP tumors have not been positive. Abnormalities in chromosomes 1 and 12 and other complex cytogenetic abnormalities have been reported. Aneuploidy has been described in 70% of CUP patients with metastatic adenocarcinoma or undifferentiated carcinoma. The overexpression of various genes, including *RAS*, *BCL2* (40%), *HER2* (11%), and *P53* (26–53%), has been identified in CUP samples, but they are found in many other malignancies. The extent of angiogenesis in CUP relative to that in metastases from known primaries has also been evaluated, but no consistent findings have emerged. Although current comprehensive genomic profiling efforts may help identify targeted therapeutic approaches to improve outcomes for this disease as discussed below, they have failed thus far to reveal a distinct molecular signature. More comprehensive and integrated multiomic efforts are needed to provide insights into CUP biology through recognition of molecular aberrations that specifically drive metastatic growth.

APPROACH TO THE PATIENT

Carcinoma (or Cancer) of Unknown Primary

Initial CUP evaluation has two goals: search for the primary or "putative primary" tumor based on clinicopathologic evaluation of the metastases and determine the extent of disease. Focused evaluation directed by clinicopathologic cues allows for judicious and efficient use of diagnostic tests. Obtaining a thorough medical history from CUP patients is essential, including paying particular attention to risk factors, prior surgeries, and family medical history to assess potential hereditary cancers. Adequate physical examination, including a digital rectal examination in men and breast and pelvic examinations in women, should be performed based on clinical presentation. Finally, all patients with CUP, in the absence of contraindication, must undergo a computed tomography (CT) scan of chest, abdomen, and pelvis as a part of their standard workup.

■ ROLE OF SERUM TUMOR MARKERS AND CYTOGENETICS

Most tumor markers, including carcinoembryonic antigen (CEA), CA-125, CA 19-9, and CA 15-3, are nonspecific and not helpful in determining the primary site. Men who present with adenocarcinoma and predominant osteoblastic metastasis should undergo a prostate-specific antigen (PSA) test. In patients with undifferentiated or poorly differentiated carcinoma (especially with a midline tumor), elevated β-human chorionic gonadotropin (β-hCG) and α fetoprotein (AFP) levels suggest the possibility of an extragonadal germ cell (testicular) tumor. With the availability of advanced IHC, cytogenetic studies are rarely needed.

ROLE OF IMAGING STUDIES

In the absence of contraindications, a baseline intravenous (IV) contrast CT scan of the chest, abdomen, and pelvis is the standard of care. This helps to search for the primary tumor, evaluate the extent of disease, and select the most accessible biopsy site. With precise imaging and reporting, latent primary cancers, defined as appearance of a new primary cancer after a latent period of several months to years, is uncommon and seen in ≤5% of CUP patients, usually in patients with very indolent presentations and/or highly responsive metastatic cancers that allows a latent primary to emerge (grow) over time.

Mammography should be performed in women who present with metastatic adenocarcinoma, especially in those with isolated axillary lymphadenopathy. Magnetic resonance imaging (MRI) of the breast can be considered in patients with axillary adenopathy and suspected occult primary breast carcinoma following a negative mammography and ultrasound. The results of these imaging modalities can influence surgical management; a negative MRI of the breast predicts a low tumor yield at mastectomy.

A conventional workup for a squamous cell carcinoma and cervical CUP (neck lymphadenopathy with no known primary tumor) includes a CT scan or MRI and invasive studies, including indirect and direct laryngoscopy, bronchoscopy, and upper endoscopy. Ipsilateral (or bilateral) staging tonsillectomy has been recommended for these patients. 18-Fluorodeoxyglucose positron emission tomography (18-FDG-PET) scans are useful in this patient population and may help guide the biopsy; determine the extent of disease; facilitate the appropriate treatment, including planning radiation fields; and help with disease surveillance. Several studies have evaluated the utility of PET in patients with squamous cervical CUP, and head and neck primary tumors were identified in ~21–30%.

The diagnostic contribution of PET to the evaluation of other CUP presentations (outside of the neck adenopathy indication) remains controversial and is not routinely recommended. PET-CT can be helpful for patients with bone metastases and those deemed candidates for aggressive multimodality therapy (surgical intervention/radiation) such as patients with solitary metastatic disease because the identification of disease in addition to the solitary metastatic site may affect treatment planning.

Invasive studies, including upper endoscopy, colonoscopy, and bronchoscopy, should be limited to symptomatic patients or those with laboratory, imaging, or pathologic abnormalities that suggest that these techniques will result in a high yield in finding a primary cancer.

TABLE 97-1 Major Histologies in Carcinoma (Cancer) of Unknown Primary

HISTOLOGY	PROPORTION, %
Well to moderately differentiated adenocarcinoma	60
Squamous cell cancer	5
Poorly differentiated adenocarcinoma, poorly differentiated carcinoma	30
Neuroendocrine	2
Undifferentiated malignancy	3

ROLE OF PATHOLOGIC STUDIES

A detailed pathologic examination of the most accessible biopsied tissue specimen is mandatory in CUP patients. Pathologic evaluation typically consists of hematoxylin and eosin stains and IHC tests. The importance of adequate tissue acquisition cannot be overemphasized in CUP. In addition to pathologic evaluation, tissue is also needed for molecular profiling, which can aid in identifying biomarkers suggesting the primary site for effective therapeutics including targeted agents, immunotherapy, and clinical trials.

Light Microscopy Evaluation Adequate tissue obtained preferably by excisional biopsy or core needle biopsy (instead of only a fine-needle aspiration) is stained with hematoxylin and eosin and subjected to light microscopic examination. On light microscopy, 60–65% of CUP is adenocarcinoma, and 5% is squamous cell carcinoma. The remaining 30–35% is poorly differentiated adenocarcinoma, poorly differentiated carcinoma, or poorly differentiated neoplasm. A small percentage of lesions are diagnosed as neuroendocrine cancers (2%), mixed tumors (adenosquamous or sarcomatoid carcinomas), or undifferentiated neoplasms (Table 97-1).

Role of IHC Analysis IHC stains are peroxidase-labeled antibodies against specific tumor antigens that are used to define tumor lineage. The number of available IHC stains is ever-increasing. However, a tiered and uniform approach to tissue evaluation in the CUP setting is lacking. For CUP cases, more is not necessarily better, and IHC stains should be used in conjunction with the patient's clinical presentation and imaging studies to select the best therapy. Communication between the clinician and pathologist is essential. No stain is 100% sensitive or specific, and under-/overinterpretation should be avoided. Poor differentiation, even in known primary tumors, decreases sensitivity of hallmark IHC markers. PSA and thyroglobulin tissue markers, which are positive in prostate and thyroid cancer, respectively, are two of the most specific markers. However, these cancers rarely present as CUP, so the yield of these tests may be low. Figure 97-1 delineates a simple algorithm for immunohistochemical staining in CUP cases. Table 97-2 lists additional tests that may be useful to further define the tumor lineage. A more comprehensive algorithm may improve the diagnostic accuracy but can make the process complex and increase

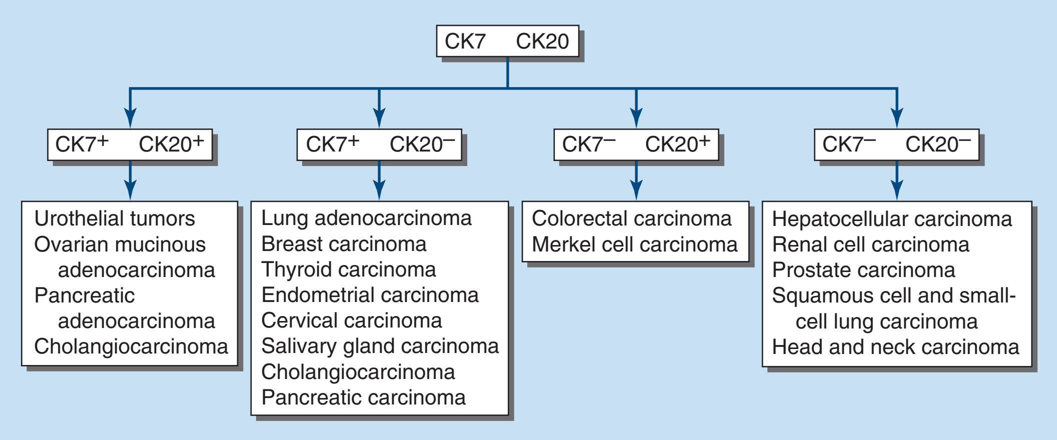

FIGURE 97-1 Approach to cytokeratin (CK7 and CK20) markers used in adenocarcinoma of unknown primary.

TABLE 97-2 Select Immunohistochemical Stains Useful in the Diagnosis of CUP

LIKELY PRIMARY PROFILE	COMMONLY CONSIDERED IHC TO ASSIST IN DIFFERENTIAL DIAGNOSIS OF CUP[a]
Breast	ER, GCDFP-15, mammaglobin, HER2/neu, GATA3
Ovarian/müllerian	ER, WT1, CK7, PAX8, PAX2
Lung adenocarcinoma	TTF-1; nuclear staining, napsin A, SP-A1
Germ cell	β-hCG, AFP, OCT3/4, CKIT, CD30 (embryonal), SALL4
Prostate	PSA, α-methylacyl CoA racemase/P504S (AMACR/P504S), P501S (prostein), PSMA, NKX3-1
Intestinal	CK7, CK20, CDX-2, CEA
Neuroendocrine	Chromogranin, synaptophysin, CD56
Sarcoma	Desmin (desmoid tumors), factor VIII (angiosarcomas), CD31, smooth muscle actin (leiomyosarcoma), MyoD1 (rhabdomyosarcoma)
Renal	RCC, CD10, PAX8, CD10
Hepatocellular carcinoma	Hep Par-1, Arg-1, glypican-3
Melanoma	S100, SOX-10, vimentin, HMB-45, tyrosinase, melan-A
Urothelial	CK7, CK20, thrombomodulin, uroplakin III
Mesothelioma	Calretinin, WT1, D2-40, mesothelin
Lymphoma	LCA, CD3, CD4, CD5, CD20, CD45
SCC	p63, p40 (lung SCC), CK5/6

[a] Patterns emerging from coexpression of stains are better than individual stains to suggest putative primary site. Even with optimization, no IHC panel is 100% sensitive or specific (e.g., ovarian mucinous carcinoma can exhibit positivity with intestinal markers).

Abbreviations: AFP, α fetoprotein; Arg-1, arginase-1; β-hCG, β-human chorionic gonadotropin; CEA, carcinoembryonic antigen; CUP, carcinoma of unknown primary; ER, estrogen receptor; GCDFP-15, gross cystic disease fibrous protein-15; IHC, immunohistochemistry; LCA, leukocyte common antigen; PSA, prostate-specific antigen; PSMA, prostate-specific membrane antigen; SCC, squamous cell carcinoma; SP-A1, surfactant protein A precursor; TTF, thyroid transcription factor; WT, Wilms' tumor.

cost. With the use of IHC markers, electron microscopic analysis, which is time-consuming and expensive, is rarely performed today.

There are >20 subtypes of cytokeratin (CK) intermediate filaments with different molecular weights and differential expression in various cell types and cancers. Monoclonal antibodies to specific CK subtypes have been used to help classify tumors according to their site of origin; commonly used CK stains in adenocarcinoma CUP are CK7 and CK20. CK7 is found in tumors of the lung, ovary, endometrium, breast, and upper gastrointestinal tract including pancreaticobiliary cancers, whereas CK20 is normally expressed in the gastrointestinal epithelium, urothelium, and Merkel cells. The nuclear CDX-2 transcription factor, which is the product of a homeobox gene necessary for intestinal organogenesis, is often used to aid in the diagnosis of gastrointestinal adenocarcinomas. However, CDX-2 positivity can be seen with enteric or mucinous differentiation in tumors from diverse primary sites (e.g., mucinous ovarian cancers).

Thyroid transcription factor 1 (TTF-1) nuclear staining is frequently positive in lung and thyroid cancers. Approximately 68% of adenocarcinomas and 25% of squamous cell lung cancers stain positive for TTF-1, which helps differentiate a lung primary tumor from metastatic adenocarcinoma in a pleural effusion, the mediastinum, or the lung parenchyma.

Gross cystic disease fibrous protein-15 (GCDFP-15), a 15-kDa monomer protein, is a marker of apocrine differentiation that is detected in 62–72% of breast carcinomas. GATA3 is being increasingly used in the CUP setting when there is concern for a breast primary and can be particularly useful as a marker for metastatic breast carcinoma, especially triple-negative and metaplastic carcinomas, which lack specific endocrine markers of mammary origin. UROIII, high-molecular-weight cytokeratin, thrombomodulin, and CK20 are the markers used to diagnose lesions of urothelial origin.

IHC performs the best when used in groups that give rise to patterns that are strongly indicative of certain profiles. For example, the TTF-1/CK7+ and CK20+/CDX-2+/CK7− phenotypes have been reported as very suggestive of lung and lower gastrointestinal cancer profiles, respectively. Despite their practical utility, these patterns have not been validated prospectively in CUP patients. IHC is not without its limitations; several factors affect tissue antigenicity (antigen retrieval, specimen processing, and fixation), interpretation of stains in tumor (nuclear, cytoplasmic, membrane) versus normal tissue, inter- and intraobserver variability, variable performance of different antibodies said to recognize the same antigen, and tissue heterogeneity and inadequacy (given small biopsy sizes). Communication with the pathologist is critical to determine if acquisition of additional tissue will be beneficial in the pathologic evaluation. Pathologic features should not always supersede clinical or radiologic findings when considering testing for biomarkers of therapeutic response (e.g., epidermal growth factor receptor [EGFR], *ALK* mutations, human epidermal growth factor receptor 2 [HER2]).

Role of Cancer Classifier Molecular Profiling In the absence of a known primary, developing therapeutic strategies for CUP is challenging. The current diagnostic yield with imaging and immunochemistry is ~20–30% for CUP patients. To reduce diagnostic uncertainty, sophisticated molecular analytics have been applied to CUP samples. These include gene expression profiling, messenger RNA (mRNA), microRNA, and genetic and epigenetic profiling to classify CUP.

Gene expression profiles are most commonly generated using quantitative reverse transcriptase polymerase chain reaction (RT-PCR) or DNA microarray. Neural network programs are then used to develop predictive algorithms from the gene expression profiles. Typically, a training set of gene profiles from known cancers (preferably from metastatic sites) is used to train the software. Comprehensive gene expression databases that have become available for common malignancies are then applied to CUP samples, and the program can then be used to predict the putative origin of a CUP sample.

mRNA- or microRNA-based tissue of origin cancer classifier assays have also been studied in prospective and retrospective CUP trials. Other assays such as DNA methylation profiling predicted a primary cancer in 87% of 216 CUP patients. Incorporation of machine learning and next-generation sequencing has furthered prediction of these classifiers.

Despite the sophistication of the cancer classifier molecular assays, most of the CUP studies have evaluated assay *performance*, although the challenge with validating the accuracy of an assay for CUP is that, by definition, the primary cancer diagnosis cannot be verified. Thus, current estimates of tissue of origin test accuracy have relied on indirect metrics, including comparison with pathology/IHC, clinical presentation, appearance of latent primaries, and autopsies. Using these measures, the assays suggest a plausible primary in ~70–80% of patients studied. Three outcomes-based studies have been performed. First, a single-arm study reported a median survival of 12.5 months for patients who received assay-directed site-specific therapy. Second, a phase 2 trial of site-specific therapy, including molecularly targeted therapy, based on predicted tumor site from an algorithm using gene expression and alteration profile showed a 1-year survival of 53.1%. However, two randomized clinical trials evaluating site-specific therapy directed by gene expression profiling versus empirical chemotherapy with either paclitaxel plus carboplatin or gemcitabine plus cisplatin failed to show a significant improvement in survival with this approach. Firm conclusions of therapeutic impact cannot be drawn from these studies given the sample size, design, statistical biases, confounding variables including use of subsequent lines of (empiric) therapy, and heterogeneity of the CUP cancers. Additional studies are needed to better understand the clinical impact of tissue of origin profiling tools and how these assays complement IHC and help guide therapy.

Role of Next-Generation Sequencing A significant push is being made toward personalized medicine across all cancer types with the goal of identifying driver mutation(s) in a patient who can be treated with targeted agents independent of the site of origin. A retrospective study of 200 CUP tumor specimens reported on genomic alterations using the hybrid capture–based FoundationOne assay. The authors reported that a large number of CUP samples (85%) harbored at least one clinically relevant genomic alteration with the potential to influence and personalize therapy. The mean number of genomic alterations was 4.2 per tumor, and the most common genetic alterations included *TP53* (55%), *KRAS* (20%), *CDKN2A* (19%), and *ARID1A* (11%). The adenocarcinoma CUP tumors were more frequently driven by genetic alterations in the receptor tyrosine kinase (RTK)/Ras/mitogen-activated protein kinase (MAPK) signaling pathway than nonadenocarcinoma CUP tumors. Although druggable genetic lesions seen in CUP are comparable to those in known primary cancer databases, whether molecularly stratified approaches for CUP will successfully improve outcomes remains to be seen and clinical trials are needed. In a single-arm phase 2 study of 97 patients with molecularly targeted therapy, five patients were found to have targetable *EGFR* mutations. Of these, four patients were treated with afatinib, an anti-EGFR drug, and two patients achieved a progression-free survival of >6 months. It is anticipated that second- and third-generation drugs targeting *EGFR* mutations are likely to be even more effective. The emerging use of assays looking for circulating tumor DNA (ctDNA), so-called liquid biopsies, has been increasingly useful within known tumor types and has stirred interest in their potential utility in CUP **(Chap. 503)**.

Ongoing histology and cellular-context agnostic prospective clinical trials are studying the presence of actionable mutations and matching patients to the right targeted drug. As this strategy gains traction, CUP would be a natural fit for genomic alteration–based targeted therapy independent of tumor site. Established tumor-agnostic therapies such as immune checkpoint inhibitors (pembrolizumab) for microsatellite instability high (MSI-H) or deficient mismatch repair (dMMR) tumors or tumors with high tumor mutation burden (TMB) and NTRK inhibitors for *NTRK* fusion–positive tumors can help a small minority of CUP patients.

TREATMENT

Carcinoma (or Cancer) of Unknown Primary

GENERAL CONSIDERATIONS

The treatment of CUP continues to evolve, albeit slowly. The median survival of most patients with disseminated CUP is ~6–10 months.

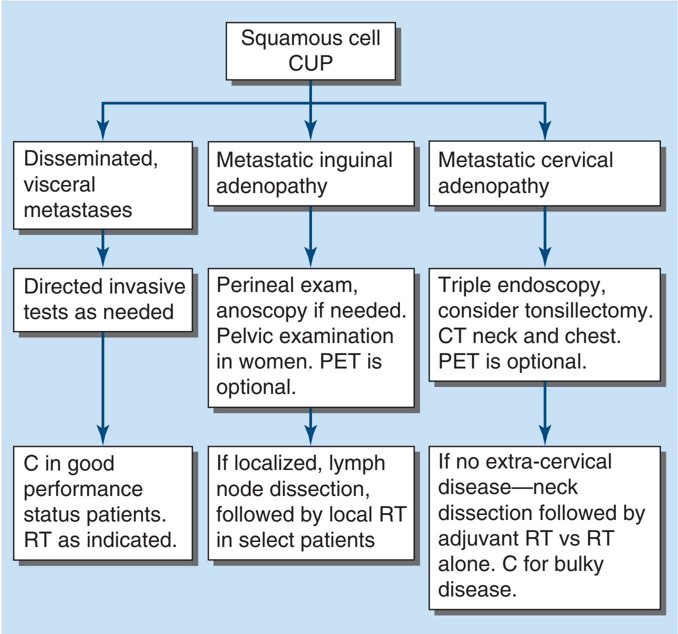
Squamous cell CUP
Disseminated, visceral metastases
Metastatic inguinal adenopathy
Metastatic cervical adenopathy
Directed invasive tests as needed
Perineal exam, anoscopy if needed. Pelvic examination in women. PET is optional.
Triple endoscopy, consider tonsillectomy. CT neck and chest. PET is optional.
C in good performance status patients. RT as indicated.
If localized, lymph node dissection, followed by local RT in select patients
If no extra-cervical disease—neck dissection followed by adjuvant RT vs RT alone. C for bulky disease.

FIGURE 97-3 Treatment algorithm for squamous cell carcinoma of unknown primary (CUP). C, chemotherapy; CT, computed tomography; PET, positron emission tomography; RT, radiation.

Systemic chemotherapy is the primary treatment modality in most patients with disseminated disease, but the careful integration of surgery, radiation therapy, and even periods of observation is important in the overall management of this condition **(Figs. 97-2 and 97-3)**. Prognostic factors include performance status, site and number of metastases, response to chemotherapy, and serum lactate dehydrogenase (LDH) levels. Culine and colleagues developed a prognostic model using performance status and serum LDH levels, which allowed the assignment of patients into two subgroups with divergent outcomes. Raghav and colleagues developed a prognostic nomogram to provide individualized survival estimates for patients with CUP based on baseline gender, Eastern Cooperative Oncology

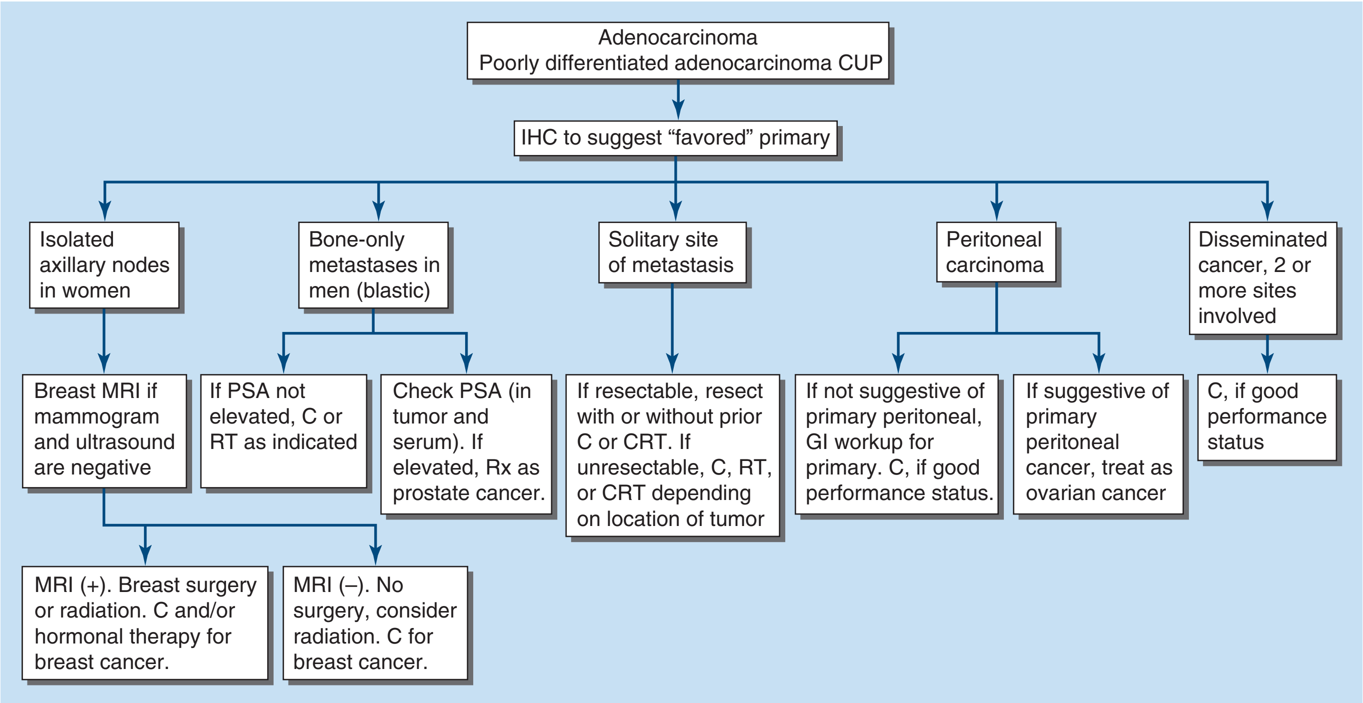
Adenocarcinoma
Poorly differentiated adenocarcinoma CUP
IHC to suggest “favored” primary
Isolated axillary nodes in women
Bone-only metastases in men (blastic)
Solitary site of metastasis
Peritoneal carcinoma
Disseminated cancer, 2 or more sites involved
Breast MRI if mammogram and ultrasound are negative
If PSA not elevated, C or RT as indicated
Check PSA (in tumor and serum). If elevated, Rx as prostate cancer.
If resectable, resect with or without prior C or CRT. If unresectable, C, RT, or CRT depending on location of tumor
If not suggestive of primary peritoneal, GI workup for primary. C, if good performance status.
If suggestive of primary peritoneal cancer, treat as ovarian cancer
C, if good performance status
MRI (+). Breast surgery or radiation. C and/or hormonal therapy for breast cancer.
MRI (–). No surgery, consider radiation. C for breast cancer.

FIGURE 97-2 Treatment algorithm for adenocarcinoma and poorly differentiated adenocarcinoma of unknown primary (CUP). C, chemotherapy; CRT, chemoradiation; GI, gastrointestinal; IHC, immunohistochemistry; MRI, magnetic resonance imaging; PSA, prostate-specific antigen; RT, radiation.

Group performance status, histology, number of metastatic sites, and neutrophil-lymphocyte ratio. Future prospective trials using this prognostic model are warranted. Clinically, some CUP diagnoses fall into a favorable prognostic subset. Others, including those with disseminated CUP, have a more unfavorable prognosis.

TREATMENT OF FAVORABLE CUP SUBSETS

Women with Isolated Axillary Adenopathy Women with isolated axillary adenopathy with adenocarcinoma or carcinoma are usually treated for stage II or III breast cancer based on pathologic findings. These patients should undergo a breast MRI if mammogram and ultrasound are negative. Radiation therapy to the ipsilateral breast is indicated if the MRI of the breast is positive. Chemotherapy and/or hormonal therapy are indicated based on patient's age (premenopausal or postmenopausal), nodal disease bulk, and hormone receptor and HER2 status **(Chap. 84)**. It is important to verify that the pathology suggests a breast cancer profile (morphology, IHC breast markers including estrogen receptor, mammaglobin, GCDFP-15, GATA3, HER2 gene expression) before embarking on a breast cancer therapeutic program.

Women with Peritoneal Carcinomatosis The term *primary peritoneal papillary serous carcinoma* (PPSC) has been used to describe CUP with carcinomatosis with the pathologic and laboratory (elevated CA-125 antigen) characteristics of ovarian cancer but no ovarian primary tumor identified on transvaginal sonography or laparotomy. Studies suggest that ovarian cancer and PPSC, which are both of müllerian origin, have similar gene expression profiles. Like patients with ovarian cancer, patients with PPSC are candidates for cytoreductive surgery, followed by adjuvant taxane- and platinum-based chemotherapy. In one retrospective study of 258 women with peritoneal carcinomatosis who had undergone cytoreductive surgery and chemotherapy, 22% of patients had a complete response to chemotherapy; the median survival duration was 18 months (range 11–24 months). However, not all peritoneal carcinomatosis in women is PPSC. Careful pathologic evaluation can help diagnose a colon cancer profile (CDX-2+, CK20+, CK7−) or a pancreaticobiliary cancer or even a mislabeled peritoneal mesothelioma (calretinin, D2-40 positive; BerEp4, MOC-31 negative).

Poorly Differentiated Carcinoma with Midline Adenopathy (Chap. 93) Men with poorly differentiated or undifferentiated carcinoma who present with midline adenopathy should be evaluated for extragonadal germ cell malignancy. If diagnosed and treated as such, they often experience a good response to treatment with platinum-based combination chemotherapy. Response rates of >50% have been noted, and long-term survival rates of 10–15% have been reported. Older patients, especially smokers, who present with mediastinal adenopathy are more likely to have a lung or head and neck cancer profile.

Neuroendocrine Cancer (Chap. 89) Low-grade neuroendocrine tumor (NET) often has an indolent course, and treatment decisions are based on symptoms and tumor bulk. Urine 5-HIAA and serum chromogranin may be elevated and can be followed as markers. Often the patient is treated with somatostatin analogues alone for hormone-related symptoms (diarrhea, flushing, nausea). Specific local therapies or systemic therapy would only be indicated if the patient is symptomatic with local pain secondary to significant growth of the metastasis or the hormone-related symptoms are not controlled with endocrine therapy. Novel therapy options have demonstrated benefit in patients with low-grade NET, including sunitinib (which targets the vascular endothelial growth factor pathway), everolimus (which inhibits the mammalian target of rapamycin), and lutetium-177 dotatate (a somatostatin peptide receptor radioligand). Patients with high-grade NET are treated with platinum-based doublet therapy; 20–25% show a complete response, and up to 10% patients with limited/oligo presentations survive for >5 years. Some degree of neuroendocrine differentiation can be seen in diverse poorly differentiated carcinomas.

Squamous Cell Carcinoma Presenting as Neck Adenopathy Patients with early-stage squamous cell carcinoma involving the cervical lymph nodes are candidates for node dissection and radiation therapy, which can result in long-term survival. The role of chemotherapy in these patients is undefined, although chemoradiation therapy or induction chemotherapy is often used and is beneficial in bulky N2/N3 lymph node disease.

Solitary Metastatic Sites Patients with solitary metastases can also experience good treatment outcomes. Some patients who present with locoregional disease are candidates for aggressive trimodality (chemotherapy, radiation, and surgery) management—both prolonged disease-free survival and, occasionally, cure are possible.

Men with Blastic Skeletal Metastases and Elevated PSA (Chap. 92) Blastic bone-only metastasis is a rare presentation, and elevated serum PSA or tumor staining with PSA may provide confirmatory evidence of prostate cancer in these patients. Those with elevated levels are candidates for hormonal or other therapy for prostate cancer, although it is important to rule out other primary tumors (lung most common).

MANAGEMENT OF DISSEMINATED CUP

Patients who present with liver, brain, and adrenal metastatic disease usually have a poor prognosis. Patients with peritoneal carcinomatosis secondary to metastatic adenocarcinoma have a broad differential diagnosis, which includes mainly gastrointestinal cancers including gastric, appendiceal, colon, and pancreaticobiliary cancers.

Traditionally, platinum-based combination chemotherapy regimens have been used to treat CUP. Several broadly used regimens have been studied in the past two decades; these include paclitaxel-carboplatin, gemcitabine-cisplatin, gemcitabine-oxaliplatin, and irinotecan and fluoropyrimidine-based therapies. These chemotherapeutic agents used as empiric regimens have shown response rates of 25–40%, and their use obtains median survival times of 6–13 months.

Outside of favorable subsets, there is a small group of patients with a "definitive" IHC profile. These patients usually have a single diagnosis based on their clinicopathologic presentation and are often treated for the putative primary tumor. This does not guarantee a response, although it increases the probability of response when select drugs are chosen from a class of drugs known to be effective in that cancer type. Efforts should be made to search for biomarkers of response to tumor-agnostic effective therapies such as immunotherapy for MSI-H/dMMR tumors. Patients who do not fall into those categories are candidates for broad-spectrum platinum-based regimens, clinical trials, and additional trial-based genomic and proteomic tests. Today, we do not have many effective drugs for several CUP cancer profiles, and treatments overlap for some cancers. Immunotherapy has been an area of active interest due to robust and durable responses in cancers with known primaries and has shown some activity in CUP. However, biomarkers of response and immune-sensitive subsets need to be defined within CUP.

SUMMARY

Patients with CUP should undergo a directed diagnostic search for the primary tumor based on clinical and pathologic data. Subsets of patients have prognostically favorable disease, as defined by clinical or histologic criteria, and may substantially benefit from aggressive treatment; in these patients, prolonged survival can be expected. However, for most patients who present with advanced CUP, the prognosis remains poor with early resistance to available cytotoxic therapy. The current focus has shifted away from empirical chemotherapeutic trials to understanding the metastatic phenotype, tissue of origin profiling in select patients, and next-generation sequencing to identify actionable mutations in CUP patients. As novel therapies evolve in cancers with known primaries, investigations to assess their value in CUP will likely have a positive impact on management of CUP patients.

FURTHER READING

Gatalica Z et al: Comprehensive analysis of cancers of unknown primary for the biomarkers of response to immune checkpoint blockade therapy. Eur J Cancer 94:179, 2018.

Hayashi H et al: Randomized phase II trial comparing site-specific treatment based on gene expression profiling with carboplatin and paclitaxel for patients with cancer of unknown primary site. J Clin Oncol 37:570, 2019.

Hayashi H et al: Site-specific and targeted therapy based on molecular profiling by next-generation sequencing for cancer of unknown primary site: A nonrandomized phase 2 clinical trial. JAMA Oncol 6:1, 2020.

Huey R et al: Feasibility and value of genomic profiling in cancer of unknown primary: Real-world evidence from prospective profiling study. J Natl Cancer Inst 115:994, 2023.

Kato S et al: Utility of genomic analysis in circulating tumor DNA from patients with carcinoma of unknown primary. Cancer Res 77: 4238, 2017.

National Comprehensive Cancer Network: Occult Primary (Cancer of Unknown Primary) version 2.2017, October 2016. *https://www.nccn.org/guidelines/guidelines-detail?category=1&id=1451*.

Raghav K et al: Cancer of unknown primary in adolescents and young adults: Clinicopathological features, prognostic factors and survival outcomes. PLoS One 11:e0154985, 2016.

Raghav K et al: Development and validation of a novel nomogram for individualized prediction of survival in cancer of unknown primary. Clin Cancer Res 27:3414, 2021.

Raghav K et al: Efficacy of pembrolizumab in patients with advanced cancer of unknown primary (CUP): A phase 2 non-randomized clinical trial. J Immunother Cancer 10:e004822, 2022.

Ross JS et al: Comprehensive genomic profiling of carcinoma of unknown primary site: New routes to targeted therapies. JAMA Oncol 1:40, 2015.

Varadhachary GR, Raber MN: Cancer of unknown primary site. N Engl J Med 371:757, 2014.

98

Paraneoplastic Syndromes: Endocrinologic/Hematologic

J. Larry Jameson, Dan L. Longo

Neoplastic cells can produce a variety of substances that can alter the physiology of hormonal, hematologic, dermatologic, rheumatologic, renal, and neurologic systems. *Paraneoplastic syndromes* refer to the disorders that accompany benign or malignant tumors but are not directly related to mass effects or invasion. Tumors of neuroendocrine origin, such as small-cell lung carcinoma (SCLC) and carcinoids, are common causes of paraneoplastic syndromes, but these syndromes are associated with many types of tumors that produce peptide hormones, cytokines, and growth factors and induce the production of antibodies. Studies of the prevalence of paraneoplastic syndromes indicate that they are more common than is generally appreciated. The signs, symptoms, and metabolic alterations associated with paraneoplastic disorders are easily overlooked in the context of a malignancy and its treatment. Consequently, atypical clinical manifestations in a patient with cancer should prompt consideration of a paraneoplastic syndrome. The most common hormonal and hematologic syndromes associated with underlying neoplasia will be discussed here.

ENDOCRINE PARANEOPLASTIC SYNDROMES

Etiology Hormones can be produced from eutopic or ectopic sources. *Eutopic* refers to the expression of a hormone from its normal tissue of origin, whereas *ectopic* refers to hormone production from an atypical tissue source. For example, adrenocorticotropic hormone (ACTH) is expressed eutopically by the corticotrope cells of the anterior pituitary, but it can be expressed ectopically in SCLC. Many hormones are produced at low levels from tissues other than the classic endocrine source. Thus, ectopic expression is often a quantitative change rather than an absolute change in tissue expression. Nevertheless, the term *ectopic expression* is firmly entrenched and conveys the abnormal physiology associated with hormone production by neoplastic cells. In addition to high levels of hormones, ectopic expression is often characterized by abnormal regulation of hormone production (e.g., defective feedback control in ectopic ACTH) and peptide processing (resulting in large, unprocessed precursor peptide such as proopiomelanocortin [POMC]).

Many different molecular mechanisms can cause ectopic hormone production. In rare instances, genetic rearrangements account for aberrant hormone expression. For example, translocation of the parathyroid hormone (*PTH*) gene can result in high levels of PTH expression in tissues other than the parathyroid gland because the genetic rearrangement brings the *PTH* gene under the control of atypical regulatory elements. A related phenomenon is well documented in many forms of leukemia and lymphoma, in which somatic genetic rearrangements confer a growth advantage and alter cellular differentiation and function. Although genetic rearrangements cause selected cases of ectopic hormone production, this mechanism is rare, as many tumors are associated with excessive production of numerous peptides. Cellular dedifferentiation probably underlies most cases of ectopic hormone production. Many cancers are poorly differentiated, and certain tumor products, such as human chorionic gonadotropin (hCG), PTH–related protein (PTHrP), and α-fetoprotein, are characteristic of gene expression at earlier developmental stages. In contrast, the propensity of certain cancers to produce particular hormones (e.g., squamous cell carcinomas produce PTHrP) suggests that dedifferentiation is partial or that selective pathways are derepressed. These expression profiles probably reflect epigenetic modifications that alter transcriptional repression, microRNA expression, and other pathways that govern cell differentiation.

Ectopic hormone production might be considered merely epiphenomenon associated with cancer if it did not cause clinical manifestations. Excessive and unregulated production of hormones such as ACTH, PTHrP, and vasopressin can lead to substantial morbidity and complicate the cancer treatment plan. Moreover, the paraneoplastic endocrinopathies may be a presenting clinical feature of underlying malignancy and prompt the search for an unrecognized tumor.

General features that confirm cancer-associated ectopic hormone syndromes include: (1) excess hormone production from an atypical tissue source; (2) documentation of tumor hormone production based on immunostaining, mRNA production, or hormone secretion in vitro; (3) hormone gradient across the tumor vascular supply; and (4) resolution or decline of hormone levels after reduction of tumor mass. Imaging studies, including computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and octreotide scintigraphy also play an important role in the detection and characterization of tumors associated with paraneoplastic syndromes, particularly when endocrine clinical manifestations precede a cancer diagnosis (i.e., ectopic ACTH in lung cancer). Treatment of the underlying tumor is the mainstay of all paraneoplastic endocrine syndromes. Depending on the hormone produced, specific therapies can be used (see below) to ameliorate symptoms but rarely influence overall survival or cancer progression.

A large number of paraneoplastic endocrine syndromes have been described, linking overproduction of particular hormones with specific types of tumors. However, certain recurring syndromes emerge from this group **(Table 98-1)**. The most common paraneoplastic endocrine

TABLE 98-1 Paraneoplastic Syndromes Caused by Ectopic Hormone Production

PARANEOPLASTIC SYNDROME	ECTOPIC HORMONE	TYPICAL TUMOR TYPES[a]
Common		
Hypercalcemia of malignancy	Parathyroid hormone–related protein (PTHrP)	Squamous cell (head and neck, lung, skin), breast, genitourinary, gastrointestinal; osteolytic metastases
	1,25-Dihydroxyvitamin D	Lymphomas
	Parathyroid hormone (PTH) (rare)	Lung, ovary
	Prostaglandin E_2 (PGE_2) (rare)	Renal, lung
Syndrome of inappropriate antidiuretic hormone secretion (SIADH)	Vasopressin	Lung (squamous, small cell), gastrointestinal, genitourinary, breast, ovary
Cushing's syndrome	Adrenocorticotropic hormone (ACTH)	Lung (small cell, bronchial carcinoid, adenocarcinoma, squamous), thymus, pancreatic islet, medullary thyroid carcinoma, pheochromocytoma
	Corticotropin-releasing hormone (CRH) (rare)	Pancreatic islet, carcinoid, lung, prostate
	Ectopic expression of gastric inhibitory peptide (GIP), luteinizing hormone (LH)/human chorionic gonadotropin (hCG), other G protein–coupled receptors (rare)	Macronodular adrenal hyperplasia
Less Common		
Non–islet cell hypoglycemia	Insulin-like growth factor type II (IGF-II)	Mesenchymal tumors, sarcomas, adrenal, hepatic, gastrointestinal, kidney, prostate
	Insulin (rare)	Cervix (small-cell carcinoma)
Male feminization	hCG[b]	Testis (embryonal, seminomas), germinomas, choriocarcinoma, lung, hepatic, pancreatic islet
Diarrhea or intestinal hypermotility	Calcitonin[c]	Lung, colon, breast, medullary thyroid carcinoma
	Vasoactive intestinal peptide (VIP)	Pancreas, pheochromocytoma, esophagus
Rare		
Oncogenic osteomalacia	Fibroblast growth factor 23 (FGF23) or phosphatonin	Hemangiopericytomas, osteoblastomas, fibromas, sarcomas, giant cell tumors, prostate, lung
Acromegaly	Growth hormone–releasing hormone (GHRH)	Pancreatic islet, bronchial, and other carcinoids
	Growth hormone (GH)	Lung, pancreatic islet
Hyperthyroidism	Thyroid-stimulating hormone (TSH)	Hydatidiform mole, embryonal tumors, struma ovarii
Hypertension	Renin	Juxtaglomerular tumors, kidney, lung, pancreas, ovary
Consumptive hypothyroidism	Type 3 deiodinase	Hepatic and other hemangiomas
Cancer immunotherapy associated autoimmune diseases	Autoimmune hormone deficiencies Thyroiditis, Graves' disease	Cancers treated with immunotherapy, particularly anti–CTLA-4, PD-1, PD-L1

[a] Only the most common tumor types are listed. For most ectopic hormone syndromes, an extensive list of tumors has been reported to produce one or more hormones. [b] hCG is produced eutopically by trophoblastic tumors. Certain tumors produce disproportionate amounts of the hCG α or hCG β subunit. High levels of hCG rarely cause hyperthyroidism because of weak binding to the TSH receptor. [c] Calcitonin is produced eutopically by medullary thyroid carcinoma and is used as a tumor marker.
Abbreviations: CTLA-4, cytotoxic T lymphocyte-associated protein-4; PD-1, programmed cell death protein 1; PD-L1, programmed death ligand 1.

syndromes include hypercalcemia from overproduction of PTHrP and other factors, hyponatremia from excess vasopressin, and Cushing's syndrome from ectopic ACTH.

HYPERCALCEMIA CAUSED BY ECTOPIC PRODUCTION OF PTHRP

(See also Chap. 422).

Etiology Humoral hypercalcemia of malignancy (HHM) occurs in up to 20% of patients with cancer. HHM is most common in cancers of the lung, head and neck, skin, esophagus, breast, and genitourinary tract and in multiple myeloma and lymphomas, as well as metastases associated with these, and other cancers. There are several distinct humoral causes of HHM, but it is caused most commonly by overproduction of PTHrP. In addition to acting as a circulating humoral factor, bone metastases (e.g., breast, multiple myeloma) may produce PTHrP and other chemokines, leading to local osteolysis and hypercalcemia. PTHrP may also affect the initiation and progression of tumors by acting through pro-survival and chemokine pathways.

PTHrP is structurally related to PTH and binds to the PTH receptor, explaining the similar biochemical features of HHM and hyperparathyroidism. PTHrP plays a key physiologic role in skeletal development and regulates cellular proliferation and differentiation in other tissues, including skin, bone marrow, breast, and hair follicles. The mechanism of PTHrP induction in malignancy is incompletely understood; however, tumor-bearing tissues commonly associated with HHM normally produce PTHrP during development or cell renewal. Hypomethylation of the *PTHLH* locus, which encodes PTHrP, suggests a role for epigenetic factors in upregulating PTHrP production. PTHrP expression is stimulated by hedgehog pathways and Gli transcription factors that are active in many malignancies. Transforming growth factor β (TGF-β), which is produced by many tumors, also stimulates PTHrP. Mutations in certain oncogenes, such as *Ras*, also can activate PTHrP expression, as does loss of the tumor suppressor, p53. In addition to its role in HHM, the PTHrP pathway may also provide a potential target for therapeutic intervention to impede cancer growth.

Another relatively common cause of HHM is excess production of 1,25-dihydroxyvitamin D. Like granulomatous disorders associated with hypercalcemia, lymphomas can produce an enzyme that converts 25-hydroxyvitamin D to the more active 1,25-dihydroxyvitamin D, leading to enhanced gastrointestinal calcium absorption. Other causes of HHM include tumor-mediated production of osteolytic cytokines and inflammatory mediators.

Clinical Manifestations The typical presentation of HHM is a patient with a known malignancy who is found to be hypercalcemic on routine laboratory tests. Less often, hypercalcemia is the initial presenting feature of malignancy. Particularly when calcium levels are markedly increased (>3.5 mmol/L [>14 mg/dL]), patients may experience fatigue, mental status changes, polyuria, dehydration, or symptoms of nephrolithiasis. Hypercalcemia can shorten ST segments and QT intervals, as well as bundle branch blocks and bradyarrhythmias.

Diagnosis Features that favor HHM, as opposed to primary hyperparathyroidism, include known malignancy, recent onset of hypercalcemia, and very high serum calcium levels. Like hyperparathyroidism, hypercalcemia caused by PTHrP is accompanied by hypercalciuria and hypophosphatemia. Patients with HHM typically have metabolic alkalosis rather than hyperchloremic acidosis, as is seen in hyperparathyroidism. In contrast to PTH, PTHrP does not appear to stimulate 1-α-hydroxylase and 1,25-dihydroxyvitamin D levels. Measurement of PTH is useful to exclude primary hyperparathyroidism; the PTH level should be suppressed in HHM. An elevated PTHrP level confirms the diagnosis, and it is increased in ~80% of hypercalcemic patients with cancer. 1,25-Dihydroxyvitamin D levels may be increased in patients with lymphoma.

TREATMENT

Humoral Hypercalcemia of Malignancy

The management of HHM begins with removal of excess calcium in the diet, medications, or intravenous (IV) solutions. Saline rehydration (typically 200–500 mL/h) is used to dilute serum calcium and promote calciuresis; exercise caution in patients with cardiac, hepatic, or renal insufficiency. Forced diuresis with furosemide (20–80 mg IV in escalating doses) or other loop diuretics can enhance calcium excretion but provides relatively little value except in life-threatening hypercalcemia. When used, loop diuretics should be administered only after complete rehydration and with careful monitoring of fluid balance. Oral phosphorus (e.g., 250 mg Neutra-Phos 3–4 times daily) should be given until serum phosphorus is >1 mmol/L (>3 mg/dL). Bisphosphonates such as zoledronate (4–8 mg IV), pamidronate (60–90 mg IV), and etidronate (7.5 mg/kg per day orally [PO] for 3–7 consecutive days) can reduce serum calcium within 1–2 days and suppress calcium release for several weeks. Bisphosphonate infusions can be repeated, or oral bisphosphonates can be used for chronic treatment. Denosumab (120 mg subcutaneously [SC] weekly for 4 weeks and then monthly) can be used in patients who do not respond adequately to bisphosphonates. It acts as a decoy receptor for RANK ligand to mitigate stimulation of osteoclasts. Cinacalcet (30 mg PO bid to 90 mg PO qid) stimulates calcium-sensing receptors to suppress PTH secretion and is therefore applicable in parathyroid carcinoma and rare cases of ectopic PTH-producing tumors. Hypercalcemia associated with lymphomas, multiple myeloma, or leukemia may respond to glucocorticoid treatment (e.g., prednisone 40–100 mg PO in four divided doses). Dialysis should be considered in severe hypercalcemia when saline hydration and bisphosphonate treatments are not possible or are too slow in onset. Previously used agents such as calcitonin and mithramycin have little utility now that bisphosphonates and other agents are available.

■ ECTOPIC VASOPRESSIN: TUMOR-ASSOCIATED SYNDROME OF INAPPROPRIATE ANTIDIURETIC HORMONE

(See also Chap. 56).

Etiology Vasopressin is an antidiuretic hormone normally produced by the posterior pituitary gland. Ectopic vasopressin production by tumors is a common cause of the syndrome of inappropriate antidiuretic hormone (SIADH), occurring in at least half of patients with SCLC. SIADH also can be caused by a number of nonneoplastic conditions, including central nervous system (CNS) trauma, infections, and medications **(Chap. 398)**. Compensatory responses to SIADH, such as decreased thirst, may mitigate the development of hyponatremia. However, with prolonged production of excessive vasopressin, the osmostat controlling thirst and hypothalamic vasopressin secretion may become reset. In addition, intake of free water, orally or intravenously, can quickly worsen hyponatremia because of reduced renal diuresis.

Tumors with neuroendocrine features, such as SCLC and carcinoids, are the most common sources of ectopic vasopressin production, but it also occurs in other forms of lung cancer and with CNS lesions, head and neck cancer, and genitourinary, gastrointestinal, and ovarian cancers. The mechanism of activation of the vasopressin gene in these tumors is unknown, but the frequent concomitant expression of the adjacent oxytocin gene suggests derepression of this locus.

Clinical Manifestations Most patients with ectopic vasopressin secretion are asymptomatic and are identified because of the presence of hyponatremia on routine chemistry testing. Symptoms may include weakness, lethargy, nausea, confusion, depressed mental status, and seizures. The severity of symptoms reflects the rapidity of onset as well as the severity of hyponatremia. Hyponatremia usually develops slowly but may be exacerbated by the administration of IV fluids or the institution of new medications.

Diagnosis The diagnostic features of ectopic vasopressin production are the same as those of other causes of SIADH **(Chaps. 56 and 398)**. Hyponatremia and reduced serum osmolality occur in the setting of an inappropriately normal or increased urine osmolality. Urine sodium excretion is normal or increased unless volume depletion is present. Other causes of hyponatremia should be excluded, including renal, adrenal, or thyroid insufficiency. Physiologic sources of vasopressin stimulation (CNS lesions, pulmonary disease, nausea), adaptive circulatory mechanisms (hypotension, heart failure, hepatic cirrhosis), and medications, including many chemotherapeutic agents, also should be considered as possible causes of hyponatremia. Vasopressin measurements are not usually necessary to make the diagnosis.

TREATMENT

Ectopic Vasopressin: Tumor-Associated SIADH

Most patients with ectopic vasopressin production develop hyponatremia over several weeks or months. The disorder should be corrected gradually unless mental status is altered or there is risk of seizures. Rapid correction can cause brain dehydration and central pontine myelinolysis. Treatment of the underlying malignancy may reduce ectopic vasopressin production, but this response is slow if it occurs at all. Fluid restriction to less than urine output, plus insensible losses, is often sufficient to correct hyponatremia partially. However, strict monitoring of the amount and types of liquids consumed or administered intravenously is required for fluid restriction to be effective. Salt tablets and saline are not helpful unless volume depletion is also present. Demeclocycline (150–300 mg orally 3–4 times daily) can be used to inhibit vasopressin action on the renal distal tubule, but its onset of action is relatively slow (1–2 weeks), and it has largely been supplanted by newer vasopressin receptor antagonists. The vaptan class of drugs acts by inhibiting vasopressin receptors (V_{1A}, V_{1B}, V_2) in the renal collecting ducts. Nonpeptide V_2-receptor antagonists, tolvaptan (15 mg PO daily) or conivaptan (20–120 mg PO bid or 10–40 mg IV), are particularly effective when used in combination with fluid restriction in euvolemic hyponatremia. Severe hyponatremia (Na <115 meq/L) or mental status changes may require treatment with hypertonic (3%) or normal saline infusion together with furosemide to enhance free water clearance. The rate of sodium correction should be slow (0.5–1 meq/L per hour) to prevent rapid fluid shifts and the possible development of central pontine myelinolysis.

■ CUSHING'S SYNDROME CAUSED BY ECTOPIC ACTH PRODUCTION

(See also Chap. 398).

Etiology Ectopic ACTH production accounts for 10–20% of cases of Cushing's syndrome. The syndrome is particularly common in neuroendocrine tumors. SCLC is the most common cause of ectopic ACTH, followed by bronchial and thymic carcinoids, islet cell tumors, other carcinoids, and pheochromocytomas. Ectopic ACTH production is caused by increased expression of the proopiomelanocortin (*POMC*) gene, which encodes ACTH, along with melanocyte-stimulating hormone (MSH), β-lipotropin, and several other peptides. In many tumors, there is abundant but aberrant expression of the *POMC* gene from an internal promoter, proximal to the third exon, which encodes ACTH. However, because this product lacks the signal sequence necessary for protein processing, it is not secreted. Increased production of ACTH arises instead from less abundant, but unregulated, *POMC* expression from the same promoter site used in the pituitary. Because tumors lack many of the enzymes needed to process the POMC polypeptide, it is typically released as multiple large, biologically inactive fragments along with relatively small amounts of fully processed, active ACTH.

Rarely, corticotropin-releasing hormone (CRH) is produced by pancreatic islet cell tumors, SCLC, medullary thyroid cancer, carcinoids, or prostate cancer. When levels are high enough, CRH can cause pituitary corticotrope hyperplasia and Cushing's syndrome. Tumors that produce CRH sometimes also produce ACTH, raising the possibility of a paracrine mechanism for ACTH production.

A distinct mechanism for ACTH-independent Cushing's syndrome involves ectopic expression of various G protein–coupled receptors in adrenal nodules. Ectopic expression of the gastric inhibitory peptide (GIP) receptor is the best-characterized example of this mechanism. In this case, meals induce GIP secretion, which inappropriately stimulates adrenal growth and glucocorticoid production.

Clinical Manifestations The clinical features of hypercortisolemia are detected in only a fraction of patients with documented ectopic ACTH production. Patients with ectopic ACTH syndrome generally exhibit less marked weight gain and centripetal fat redistribution, probably because the exposure to excess glucocorticoids is relatively brief and because cachexia reduces the propensity for weight gain and fat deposition. The ectopic ACTH syndrome is associated with several clinical features that distinguish it from other causes of Cushing's syndrome (e.g., pituitary adenomas, adrenal adenomas, iatrogenic glucocorticoid excess). The metabolic manifestations of ectopic ACTH syndrome are dominated by fluid retention and hypertension, hypokalemia, metabolic alkalosis, glucose intolerance, and occasionally steroid psychosis. The very high ACTH levels often cause increased pigmentation, reflecting increased activity of MSH derived from the POMC precursor peptide. The extraordinarily high glucocorticoid levels in patients with ectopic sources of ACTH can lead to marked skin fragility and easy bruising. In addition, the high cortisol levels often overwhelm the renal 11β-hydroxysteroid dehydrogenase type II enzyme, which normally inactivates cortisol and prevents it from binding to renal mineralocorticoid receptors. Consequently, in addition to the excess mineralocorticoids produced by ACTH stimulation of the adrenal gland, high levels of cortisol exert activity through the mineralocorticoid receptor, leading to severe hypokalemia.

Diagnosis The diagnosis of ectopic ACTH syndrome is usually not difficult in the setting of a known malignancy. Urine-free cortisol levels fluctuate but are typically greater than two to four times normal, and the plasma ACTH level is usually >22 pmol/L (>100 pg/mL). A suppressed ACTH level excludes this diagnosis and indicates an ACTH-independent cause of Cushing's syndrome (e.g., adrenal or exogenous glucocorticoid). In contrast to pituitary sources of ACTH, most ectopic sources of ACTH do not respond to glucocorticoid suppression. Therefore, high-dose dexamethasone (8 mg PO) suppresses 8:00 A.M. serum cortisol (50% decrease from baseline) in ~80% of pituitary ACTH-producing adenomas but fails to suppress ectopic ACTH in ~90% of cases. Bronchial and other carcinoids are well-documented exceptions to these general guidelines, as these ectopic sources of ACTH may exhibit feedback regulation indistinguishable from pituitary adenomas, including suppression by high-dose dexamethasone, and ACTH responsiveness to adrenal blockade with metyrapone. If necessary, petrosal sinus catheterization can be used to evaluate a patient with ACTH-dependent Cushing's syndrome when the source of ACTH is unclear. After CRH stimulation, a 3:1 petrosal sinus:peripheral ACTH ratio strongly suggests a pituitary ACTH source. Imaging studies (CT or MRI) are also useful in the evaluation of suspected carcinoid lesions, allowing biopsy and characterization of hormone production using special stains. If available, PET scans or octreotide scintigraphy may identify some sources of ACTH production.

TREATMENT

Cushing's Syndrome Caused by Ectopic ACTH Production

The morbidity associated with the ectopic ACTH syndrome can be substantial. Patients may experience depression or personality changes because of extreme cortisol excess. Metabolic derangements, including diabetes mellitus and hypokalemia, can worsen fatigue. Poor wound healing and predisposition to infections can complicate the surgical management of tumors, and opportunistic infections caused by organisms such as *Pneumocystis carinii* and mycoses are often the cause of death in patients with ectopic ACTH production. These patients have increased risk of venous thromboembolism, reflecting the combination of malignancy and altered coagulation factor profiles. Depending on prognosis and treatment plans for the underlying malignancy, measures to reduce cortisol levels are often indicated. Treatment of the underlying malignancy may reduce ACTH levels but is rarely sufficient to reduce cortisol levels to normal. Adrenalectomy is not practical for most of these patients but should be considered during surgery for the malignancy or if the underlying tumor is not resectable and the prognosis is otherwise favorable (e.g., carcinoid). Medical therapy with ketoconazole (300–600 mg PO bid), metyrapone (250–500 mg PO every 6 h), mitotane (3–6 g PO in four divided doses, tapered to maintain low cortisol production), etomidate (0.1–0.3 mg/kg/h IV), or other agents that block steroid synthesis or action is often the most practical strategy for managing the hypercortisolism associated with ectopic ACTH production. Glucocorticoid replacement should be provided to prevent adrenal insufficiency **(Chap. 398)**. Unfortunately, many patients eventually progress despite medical blockade. Mifepristone (200–1000 mg PO qd) inhibits both glucocorticoid and progesterone receptors, has rapid onset of action, and improves glucose intolerance and hypertension in a subset of patients. ACTH-neutralizing antibodies and ACTH receptor blockers are under investigation, as are selective inhibitors of the glucocorticoid receptor.

■ TUMOR-INDUCED HYPOGLYCEMIA CAUSED BY EXCESS PRODUCTION OF INSULIN-LIKE GROWTH FACTOR TYPE II

(See also **Chap. 418**) Mesenchymal tumors, hemangiopericytomas, hepatocellular tumors, adrenal carcinomas, and a variety of other large tumors have been reported to produce excessive amounts of insulin-like growth factor type II (IGF-II) precursor, which binds weakly to insulin receptors and more strongly to IGF-I receptors, leading to insulin-like actions. The gene encoding IGF-II resides on chromosome 11p15, a locus that is normally imprinted (that is, expression is exclusively from a single parental allele). Biallelic expression of the IGF-II gene occurs in a subset of tumors, suggesting loss of methylation and loss of imprinting as a mechanism for gene induction. In addition to increased IGF-II production, IGF-II bioavailability is increased due to complex alterations in circulating binding proteins. Increased IGF-II suppresses growth hormone (GH) and insulin, resulting in reduced IGF binding protein 3 (IGFBP-3), IGF-I, and acid-labile subunit (ALS). The reduction in ALS and IGFBP-3, which normally sequester IGF-II, causes it to be displaced to a small circulating complex that has greater access to insulin target tissues. For this reason, circulating IGF-II levels may not be markedly increased despite causing hypoglycemia. In addition to IGF-II–mediated hypoglycemia, tumors may occupy enough of the liver to impair gluconeogenesis.

In most cases, a tumor causing hypoglycemia is clinically apparent (usually >10 cm in size), and hypoglycemia develops in association with fasting. As with other causes of hypoglycemia, patients may present with sweating, tremors, palpitations, confusion, seizures, or coma. The diagnosis is made by documenting low serum glucose and suppressed insulin levels in association with symptoms of hypoglycemia. Serum IGF-II levels may not be increased (IGF-II assays may not detect IGF-II precursors), but an elevated IGF-II/IGF-I ratio greater than 10:1 is suggestive. Increased IGF-II mRNA expression is found in most of these tumors. Any medications associated with hypoglycemia should be eliminated. Treatment of the underlying malignancy, if possible, may reduce the predisposition to hypoglycemia. Frequent meals and IV glucose, especially during sleep or fasting, are often necessary to prevent hypoglycemia. Glucagon, recombinant GH, and glucocorticoids have also been used to enhance glucose production. Antibodies that inhibit IGF-II action are under development.

HUMAN CHORIONIC GONADOTROPIN

hCG is composed of α and β subunits and can be produced as intact hormone, which is biologically active, or as uncombined biologically inert subunits. Ectopic production of intact hCG occurs most often in association with testicular embryonal tumors, germ cell tumors, extragonadal germinomas, lung cancer, hepatoma, and pancreatic islet tumors. Eutopic production of hCG occurs with trophoblastic malignancies. hCG α subunit production is particularly common in lung cancer and pancreatic islet cancer. In men, high hCG levels stimulate steroidogenesis and aromatase activity in testicular Leydig cells, resulting in increased estrogen production and the development of gynecomastia. Precocious puberty in boys or gynecomastia in men should prompt measurement of hCG and consideration of a testicular tumor or another source of ectopic hCG production. Most women are asymptomatic. hCG is easily measured. Treatment should be directed at the underlying malignancy.

ONCOGENIC OSTEOMALACIA

Hypophosphatemic oncogenic osteomalacia, also called tumor-induced osteomalacia (TIO), is caused by excessive production of fibroblast growth factor 23 (FGF23), previously referred to as phosphotonin. Oncogenic osteomalacia is characterized by markedly reduced serum phosphorus and renal phosphate wasting, leading to muscle weakness, bone pain, and osteomalacia. Serum calcium and PTH levels are normal. FGF23 inhibits the renal conversion of 25-hydroxyvitamin D to 1,25-dihydroxyvitamin D, resulting in low levels of 1,25-dihydroxyvitamin D. Oncogenic osteomalacia is usually caused by benign mesenchymal tumors, such as hemangiopericytomas, fibromas, and giant cell tumors, often of the skeletal extremities or head. It has also been described in sarcomas and in patients with prostate or lung cancer. Resection of the tumor reverses the disorder, confirming its humoral basis. FGF23 levels are increased in some, but not all, patients with osteogenic osteomalacia. FGF23 forms a ternary complex with the klotho protein and renal FGF receptors to reduce renal phosphate reabsorption. Treatment involves removal of the tumor, if possible, and supplementation with phosphate and vitamin D. Octreotide treatment reduces phosphate wasting in some patients with tumors that express somatostatin receptor subtype 2. Octreotide scans may also be useful in detecting these tumors. A human monoclonal antibody against FGF23, burosumab (0.5 mg/kg every 4 weeks) has been approved for the treatment of osteogenic osteomalacia. If needed, it can be increased to 2 mg/kg every 2 weeks. Burosumab improves metabolic features of the disease and may improve bone structure and fracture risk, but these outcomes are still being evaluated. The calcium-sensing receptor agonist cinacalcet has been effective in some patients, apparently by reducing PTH-mediated phosphaturia. FGF receptor inhibitors hold promise as future therapies targeted either to pathways that stimulate FGR23 production (e.g., FGFR1) or inhibit its action (e.g., FGF23 receptor).

CONSUMPTIVE HYPOTHYROIDISM

Newborns with hepatic hemangiomas can develop a rare form of hypothyroidism caused by overexpression of type 3 deiodinase (D3), an enzyme that degrades and inactivates thyroxine (T_4) and triiodothyronine (T_3). The very high expression of D3 and consumption of thyroid hormones apparently outstrip the thyroid gland's rate of hormone production. The disorder is characterized by low T_4, low T_3, high TSH, and markedly elevated reverse T_3 (rT_3), reflecting the degradation of T_4 to rT_3. In addition to treating the underlying hemangioma (rarely other tumor types), patients are treated with L-thyroxine replacement, titrated to normalize TSH. Steroids and propranolol may provide benefit, perhaps by inhibiting growth factor pathways thought to stimulate D3 production.

CANCER IMMUNOTHERAPY-ASSOCIATED ENDOCRINOPATHIES (SEE ALSO CHAPS. 78, 401)

Although not strictly a paraneoplastic endocrine syndrome, the introduction of cancer immunotherapies, particularly immune checkpoint inhibitors (i.e., anti–CTLA-4, PD-1, PD-L1), is associated with a high incidence (~10%) of autoimmune endocrine disease. Hypophysitis and autoimmune thyroid diseases are most common, but autoimmune diabetes mellitus, adrenal insufficiency, hypoparathyroidism, and diabetes insipidus also occur. The mechanism of these autoimmune side effects remains unclear. CTLA-4 is expressed in the pituitary gland, likely explaining predisposition to anti–CTLA-4–associated hypophysitis. Genetic predisposition and underlying endocrine autoimmunity likely play a role in thyroid and other endocrinopathies, becoming exacerbated following checkpoint inhibitor–induced immune activation. Autoimmune endocrine disease can emerge in the early weeks of immunotherapy but can also present after several months. At present, screening and prophylactic treatment are not routinely recommended. However, clinicians should be attuned to clinical or laboratory features of these disorders as they can be challenging to identify during cancer management. Treatment is similar to that of individual hormone deficiencies **(Chaps. 391, 395, and 398)**. These endocrinopathies are generally irreversible and require lifelong hormone replacement.

HEMATOLOGIC SYNDROMES

The elevation of granulocyte, platelet, and eosinophil counts in most patients with myeloproliferative disorders is caused by the proliferation of the myeloid elements due to the underlying disease rather than to a paraneoplastic syndrome. The paraneoplastic hematologic syndromes in patients with solid tumors are less well characterized than are the endocrine syndromes because the ectopic hormone(s) or cytokines responsible have not been identified in most of these tumors **(Table 98-2)**. The extent of the paraneoplastic syndromes parallels the course of the cancer. With very rare exception, red cell, white cell, or platelet numbers are self-limited and not associated with symptomatic abnormalities. In some circumstances, elevations in platelet counts can be a marker that influences prognosis. By far, the most consequential hematologic abnormality in cancer patients is hypercoagulability.

ERYTHROCYTOSIS

Ectopic production of erythropoietin by cancer cells causes most paraneoplastic erythrocytosis. The ectopically produced erythropoietin

TABLE 98-2 Paraneoplastic Hematologic Syndromes

SYNDROME	PROTEINS	CANCERS TYPICALLY ASSOCIATED WITH SYNDROME
Erythrocytosis	Erythropoietin	Renal cancers, hepatocarcinoma, cerebellar hemangioblastomas
Granulocytosis	G-CSF, GM-CSF, IL-6	Lung cancer, gastrointestinal cancer, ovarian cancer, genitourinary cancer, Hodgkin's disease
Thrombocytosis	IL-6	Lung cancer, gastrointestinal cancer, breast cancer, ovarian cancer, lymphoma
Eosinophilia	IL-5	Lymphoma, leukemia, lung cancer
Thrombophlebitis	Unknown	Lung cancer, pancreatic cancer, gastrointestinal cancer, breast cancer, genitourinary cancer, ovarian cancer, prostate cancer, lymphoma

Abbreviations: G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte-macrophage colony-stimulating factor; IL, interleukin.

stimulates the production of red blood cells (RBCs) in the bone marrow and raises the hematocrit. Other lymphokines and hormones produced by cancer cells may stimulate erythropoietin release but have not been proved to cause erythrocytosis.

Most patients with erythrocytosis have an elevated hematocrit (>52% in men, >48% in women) that is detected on a routine blood count. Approximately 3% of patients with renal cell cancer, 10% of patients with hepatoma, and 15% of patients with cerebellar hemangioblastomas have erythrocytosis. In most cases, the erythrocytosis is asymptomatic.

Patients with erythrocytosis due to a renal cell cancer, hepatoma, or CNS cancer should have measurement of red cell mass. If the red cell mass is elevated, the serum erythropoietin level should be measured. Patients with a cancer that has been associated with erythrocytosis, elevated erythropoietin levels, and no other explanation for erythrocytosis (e.g., hemoglobinopathy that causes increased O_2 affinity; **Chaps. 66 and 103**) have the paraneoplastic syndrome.

TREATMENT

Erythrocytosis

Successful resection of the cancer usually resolves the erythrocytosis. If the tumor cannot be resected or treated effectively with radiation therapy or chemotherapy, phlebotomy may control any symptoms or risk related to erythrocytosis.

GRANULOCYTOSIS

Approximately 30% of patients with solid tumors have granulocytosis (granulocyte count >8000/μL). In about half of patients with granulocytosis and cancer, the granulocytosis has an identifiable nonparaneoplastic etiology (e.g., infection, tumor necrosis, glucocorticoid administration). The other patients have proteins in urine and serum that stimulate the growth of bone marrow cells. Tumors and tumor cell lines from patients with lung, ovarian, and bladder cancers have been documented to produce granulocyte colony-stimulating factor (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), and/or interleukin 6 (IL-6). However, the etiology of granulocytosis has not been characterized in most patients.

Patients with granulocytosis are nearly all asymptomatic, and the differential white blood cell count does not have a shift to immature forms of neutrophils. Granulocytosis occurs in 40% of patients with lung and gastrointestinal cancers, 20% of patients with breast cancer, 30% of patients with brain tumors and ovarian cancers, 20% of patients with Hodgkin's disease, and 10% of patients with renal cell carcinoma. Patients with advanced-stage disease are more likely to have granulocytosis than are those with early-stage disease.

Paraneoplastic granulocytosis does not require treatment. The granulocytosis resolves when the underlying cancer is treated.

THROMBOCYTOSIS

Some 35% of patients with thrombocytosis (platelet count >400,000/μL) have an underlying diagnosis of cancer. IL-6, a candidate molecule for the etiology of paraneoplastic thrombocytosis, stimulates the production of platelets in vitro and in vivo. Some patients with cancer and thrombocytosis have elevated levels of IL-6 in plasma. Another candidate molecule is thrombopoietin, a peptide hormone that stimulates megakaryocyte proliferation and platelet production. The etiology of thrombocytosis has not been established in most cases.

Patients with thrombocytosis are nearly all asymptomatic. Thrombocytosis is not clearly linked to thrombosis in patients with cancer. Thrombocytosis is present in 40% of patients with lung and gastrointestinal cancers; 20% of patients with breast, endometrial, and ovarian cancers; and 10% of patients with lymphoma. Patients with thrombocytosis are more likely to have advanced-stage disease and have a poorer prognosis than do patients without thrombocytosis. In ovarian cancer, IL-6 has been shown to directly promote tumor growth. Paraneoplastic thrombocytosis does not require treatment other than treatment of the underlying tumor.

EOSINOPHILIA

Eosinophilia is present in ~1% of patients with cancer. Tumors and tumor cell lines from patients with lymphomas or leukemia may produce IL-5, which stimulates eosinophil growth. Activation of IL-5 transcription in lymphomas and leukemias may involve translocation of the long arm of chromosome 5, to which the genes for IL-5 and other cytokines map.

Patients with eosinophilia are typically asymptomatic. Eosinophilia is present in 10% of patients with lymphoma, 3% of patients with lung cancer, and occasional patients with cervical, gastrointestinal, renal, and breast cancer. Patients with markedly elevated eosinophil counts (>5000/μL) can develop shortness of breath and wheezing. A chest radiograph may reveal diffuse pulmonary infiltrates from eosinophil infiltration and activation in the lungs.

TREATMENT

Eosinophilia

Definitive treatment is directed at the underlying malignancy. Tumors should be resected or treated with radiation or chemotherapy. In most patients who develop shortness of breath related to eosinophilia, symptoms resolve with the use of oral or inhaled glucocorticoids. IL-5 antagonists exist but have not been evaluated in this clinical setting.

THROMBOPHLEBITIS AND DEEP VENOUS THROMBOSIS

Deep venous thrombosis and pulmonary embolism are the most common thrombotic conditions in patients with cancer. Migratory or recurrent thrombophlebitis may be the initial manifestation of cancer. Nearly 15% of patients who develop deep venous thrombosis or pulmonary embolism have a diagnosis of cancer **(Chap. 122)**. The coexistence of peripheral venous thrombosis with visceral carcinoma, particularly pancreatic cancer, is called *Trousseau's syndrome*.

Pathogenesis Patients with cancer are predisposed to thromboembolism because they are often at bed rest or immobilized, and tumors may obstruct or slow blood flow. Postoperative deep venous thrombosis is twice as common in cancer patients who undergo surgery. Chronic IV catheters also predispose to clotting. In addition, clotting may be promoted by release of procoagulants or cytokines from tumor cells or associated inflammatory cells or by platelet adhesion or aggregation. The specific molecules that promote thromboembolism have not been identified.

Chemotherapeutic agents, particularly those associated with endothelial damage, can induce venous thrombosis. The annual risk of venous thrombosis in patients with cancer receiving chemotherapy is about 11%, sixfold higher than the risk in the general population. Bleomycin, L-asparaginase, nitrogen mustard, thalidomide analogues, cisplatin-based regimens, and high doses of busulfan and carmustine are all associated with an increased risk.

In addition to cancer and its treatment causing secondary thrombosis, primary thrombophilic diseases may be associated with cancer. For example, the antiphospholipid antibody syndrome is associated with a wide range of pathologic manifestations **(Chap. 369)**. About 20% of patients with this syndrome have cancers. Among patients with cancer and antiphospholipid antibodies, 35–45% develop thrombosis.

Clinical Manifestations Patients with cancer who develop deep venous thrombosis usually develop swelling or pain in the leg, and physical examination reveals tenderness, warmth, and redness. Patients who present with pulmonary embolism develop dyspnea, chest pain, and syncope, and physical examination shows tachycardia, cyanosis, and hypotension. Some 5% of patients with no history of cancer who have a diagnosis of deep venous thrombosis or pulmonary embolism will have a diagnosis of cancer within 1 year. The most common cancers associated with thromboembolic episodes include lung, pancreatic, gastrointestinal, breast, ovarian, and genitourinary cancers; lymphomas; and brain tumors. Patients with cancer who undergo surgical procedures requiring general anesthesia have a 20–30% risk of deep venous thrombosis.

Diagnosis The diagnosis of deep venous thrombosis in patients with cancer is made by impedance plethysmography or bilateral compression ultrasonography of the leg veins. Patients with a noncompressible venous segment have deep venous thrombosis. If compression ultrasonography is normal and there is a high clinical suspicion for deep venous thrombosis, venography should be done to look for a luminal filling defect. Elevation of D-dimer is not as predictive of deep venous thrombosis in patients with cancer as it is in patients without cancer; elevations are seen in people over age 65 years without concomitant evidence of thrombosis, probably as a consequence of increased thrombin deposition and turnover in aging.

Patients with symptoms and signs suggesting a pulmonary embolism should be evaluated with a chest radiograph, electrocardiogram, arterial blood gas analysis, and ventilation-perfusion scan. Patients with mismatched segmental perfusion defects have a pulmonary embolus. Patients with equivocal ventilation-perfusion findings should be evaluated as described above for deep venous thrombosis in their legs. If deep venous thrombosis is detected, they should be anticoagulated. If deep venous thrombosis is not detected, they should be considered for a pulmonary angiogram.

Patients without a diagnosis of cancer who present with an initial episode of thrombophlebitis or pulmonary embolus need no additional tests for cancer other than a careful history and physical examination. In light of the many possible primary sites, diagnostic testing in asymptomatic patients is wasteful. However, if the clot is refractory to standard treatment or is in an unusual site, or if the thrombophlebitis is migratory or recurrent, efforts to find an underlying cancer are indicated.

TREATMENT

Thrombophlebitis and Deep Venous Thrombosis

Patients with cancer and a diagnosis of deep venous thrombosis or pulmonary embolism should be treated initially with IV unfractionated heparin or low-molecular-weight heparin for at least 5 days, and warfarin should be started within 1 or 2 days. The warfarin dose should be adjusted so that the international normalized ratio (INR) is 2–3. Patients with proximal deep venous thrombosis and a relative contraindication to heparin anticoagulation (hemorrhagic brain metastases or pericardial effusion) should be considered for placement of a filter in the inferior vena cava (Greenfield filter) to prevent pulmonary embolism. Warfarin should be administered for 3–6 months. An alternative approach is to use low-molecular-weight heparin for 6 months. The new oral anticoagulants (factor Xa and thrombin inhibitors) are attractive because they do not require close monitoring of the prothrombin time and are not affected by dietary factors. Oral apixaban (10 mg bid for 7 days followed by 5 mg bid for 6 months) is noninferior to dalteparin in the treatment of cancer patients who develop deep vein thrombosis or pulmonary embolism. Patients with cancer who undergo a major surgical procedure should be considered for heparin prophylaxis or pneumatic boots. Breast cancer patients undergoing chemotherapy and patients with implanted catheters should be considered for prophylaxis. Guidelines recommend that hospitalized patients with cancer and patients receiving a thalidomide analogue receive prophylaxis with low-molecular-weight heparin or low-dose aspirin. Use of prophylaxis routinely during chemotherapy is controversial. Risk is affected by type of cancer, type of therapy, blood counts, and body mass index (all taken into account in the Khorana risk score; Table 98-3). Studies of Khorana high-risk patients with cancer using rivaroxaban and apixaban as clot prophylaxis have resulted in a 50% reduction in risk with a level of bleeding of about 5%. However, prophylaxis is not routinely recommended by the American Society of Clinical Oncology.

TABLE 98-3 Khorana Risk Score for Venous Thromboembolism in Cancer Patients

PATIENT CHARACTERISTICS	RISK SCORE POINTS
Site of cancer	
Very high risk (stomach, pancreas)	2
High risk (lung, lymphoma, gynecologic, genitourinary excluding prostate)	1
Prechemotherapy platelet count ≥350,000/μL	1
Hemoglobin level <10 g/dL or use of red cell growth factors	1
Prechemotherapy leukocyte count >11,000/μL	1
BMI ≥35 kg/m^2	1

RISK SCORE (POINTS)	RISK CATEGORY	RATES OF sVTE ACCORDING TO SCORES (%)
0	Low	0.3–0.8
1–2	Intermediate	1.8–2.0
≥3	High	6.7–7.1

Abbreviations: BMI, body mass index; sVTE, symptomatic venous thromboembolism.

Source: Reproduced from AJM Muñoz et al: Clinical guide SEOM on venous thromboembolism in cancer patients. Clin Transl Oncol 16:1079-1090, 2014.

MISCELLANEOUS REMOTE EFFECTS OF CANCER

Patients with cancer can develop paraneoplastic autoimmune disorders (e.g., thrombocytopenia) and dysfunction of organs not directly invaded or involved with the cancer (rheumatologic and renal abnormalities are among the most frequent). The pathogenesis of these disorders is undefined, but often, the conditions reverse if the tumor is removed or successfully treated.

Cutaneous paraneoplastic syndromes are discussed in Chap. 61. Neurologic paraneoplastic syndromes are discussed in Chap. 99.

■ FURTHER READING

Agnelli G et al: Apixaban for the treatment of venous thromboembolism associated with cancer. N Engl J Med 382:1599, 2020.

Asonitis N et al: Diagnosis, pathophysiology and management of hypercalcemia in malignancy: A review of the literature. Horm Metab Res 51:770, 2019.

Catani MV et al: The "Janus face" of platelets in cancer. Int J Mol Sci 21:788, 2020.

Dynkevich Y et al: Tumors, IGF-2, and hypoglycemia: Insights from the clinic, the laboratory, and the historical archive. Endocr Rev 34:798, 2013.

Feelders RA et al: Advances in the medical treatment of Cushing's syndrome. Lancet Diabetes Endocrinol 7:300, 2019.

Farge D et al: 2022 international clinical practice guidelines for the treatment and prophylaxis of thromboembolism in patients with cancer, including COVID19. Lancet Oncol 23:e334, 2022.

Hattersley R et al: Endocrine complications of immunotherapies: A review. Clin Med 21:e212, 2021.

Jan de Beur SM et al: Burosumab for the treatment of tumor-induced osteomalacia. J Bone Miner Res 36:627, 2021.

Lin RJ et al: Paraneoplastic thrombocytosis: The secrets of tumor self-promotion. Blood 124:184, 2014.

Onyema MC et al: Endocrine abnormality in paraneoplastic syndrome. Best Pract Res Clin Endocrinol Metab 36:101621, 2022.

Pelosof LC, Gerber DE: Paraneoplastic syndromes: An approach to diagnosis and treatment. Mayo Clin Proc 85:838, 2010.

Workeneh BT et al: Hyponatremia in the cancer patient. Kidney Int 98:870, 2020.

99

Paraneoplastic Neurologic Syndromes and Autoimmune Encephalitis

Josep Dalmau, Francesc Graus

Paraneoplastic neurologic disorders (PNDs) are cancer-related syndromes that can affect any part of the nervous system (**Table 99-1**). Initially defined as syndromes of unknown cause, currently most PNDs are considered to have an autoimmune pathogenesis triggered by the underlying cancer. In 60% of patients, the neurologic symptoms precede the cancer diagnosis. Clinically disabling PNDs occur in 0.5–1% of all cancer patients, but they affect 2–3% of patients with neuroblastoma or small-cell lung cancer (SCLC) and 30–50% of patients with thymoma.

PATHOGENESIS

Most PNDs are mediated by immune responses triggered by neuronal proteins ectopically expressed by tumors (e.g., SCLC and other cancers) or as a result of altered immunologic responses caused by some types of tumors such as thymomas or lymphomas. In PNDs of the central nervous system (CNS), many antibody-associated immune responses have been identified. These antibodies react with neurons and the patient's tumor, and their detection in serum or cerebrospinal fluid (CSF) variably predicts the presence of cancer. According to the frequency of an underlying tumor, these antibodies are classified as high risk (>70% probability of an underlying tumor; **Table 99-2**); intermediate risk (30–70% probability of an underlying tumor; **Table 99-3**), and low risk (<30% probability of an underlying tumor; **Table 99-4**). All the target antigens of high-risk antibodies are intracellular proteins except for Tr (DNER [delta/notch-like epidermal growth factor-related receptor]), which is expressed on the cell surface. By contrast, all target antigens of intermediate- and low-risk antibodies are cell-surface proteins or receptors except for GAD65 and glial fibrillary acidic protein (GFAP), which are intracellular. When the antigens are intracellular, most neurologic syndromes are associated with extensive infiltrates of CD4+ and CD8+ T cells, microglial activation, gliosis, and variable neuronal loss. The infiltrating T cells are often in close contact with neurons undergoing degeneration, suggesting a primary pathogenic role. T-cell–mediated cytotoxicity may contribute directly to cell death in these PNDs and probably underlies the resistance of many of these conditions to therapy.

TABLE 99-1 Paraneoplastic Syndromes of the Nervous System

CLASSIC SYNDROMES: HIGH RISK OF ASSOCIATED CANCER[a]	NONCLASSIC SYNDROMES: MODERATE OR LOW RISK OF ASSOCIATED CANCER
Encephalomyelitis	Brainstem encephalitis
Limbic encephalitis	Stiff-person syndrome
Cerebellar degeneration (adults)	Progressive encephalomyelitis with rigidity and myoclonus
Opsoclonus-myoclonus	Necrotizing myelopathy
Sensory neuronopathy	Motor neuron disease
Gastrointestinal pseudo-obstruction (enteric neuropathy)	Subacute axonal sensory-motor neuropathies
Dermatomyositis (adults)	Paraproteinemic neuropathies
Lambert-Eaton myasthenic syndrome	Pure autonomic neuropathy
Cancer- or melanoma-associated retinopathy	Acute necrotizing myopathy
	Polymyositis
	Optic neuropathy
	BDUMP
	Peripheral nerve hyperexcitability (neuromyotonia)
	Myasthenia gravis

[a] These syndromes frequently associate with cancer.

Abbreviation: BDUMP, bilateral diffuse uveal melanocytic proliferation.

TABLE 99-2 High-Risk Antibodies (>70% Probability of an Underlying Cancer), Syndromes, and Associated Tumors

ANTIBODY[a]	ASSOCIATED NEUROLOGIC SYNDROME(S)	TUMORS
Anti-Hu (ANNA1)	Encephalomyelitis, sensory neuronopathy	SCLC
Anti-Yo (PCA1)	Rapidly progressive cerebellar syndrome	Ovary, breast
Anti-Ri (ANNA2)	Cerebellar degeneration, opsoclonus, brainstem encephalitis	Breast, gynecologic, SCLC
Anti-CRMP5 (CV2)	Encephalomyelitis, chorea, optic neuritis, uveitis, peripheral neuropathy	SCLC, thymoma, other
Anti-Tr (DNER)	Rapidly progressive cerebellar syndrome	Hodgkin's lymphoma
Anti-Ma proteins	Limbic, hypothalamic, brainstem encephalitis	Testicular (Ma2), other (Ma)
Anti-PCA2 (MAP1B)	Sensorimotor neuropathy, rapidly progressive cerebellar syndrome, and encephalomyelitis	SCLC, non-SCLC, breast cancer
Anti-Kelch-like protein 11	Brainstem encephalitis, ataxia, sensorineural hearing loss	Seminoma, germ cell tumor, teratoma,
Anti-amphiphysin[b]	Stiff-person syndrome, encephalomyelitis	Breast, SCLC
Anti-SOX1	LEMS, rapidly progressive cerebellar syndrome with and without LEMS	SCLC

[a] All the antibodies of this table are against intracellular antigens, except for Tr (DNER), which is a cell-surface protein. [b] Amphiphysin is likely exposed to the cell surface during synaptic vesicle endocytosis.

Abbreviations: CRMP, collapsin response-mediator protein; DNER, delta/notch-like epidermal growth factor-related receptor; LEMS, Lambert-Eaton myasthenic syndrome; MAP1B, microtubule associated protein 1B; PCA, Purkinje cell antigen; SCLC, small-cell lung cancer.

In contrast to the predominant role of cytotoxic T-cell mechanisms in PNDs associated with antibodies against intracellular antigens, those associated with antibodies to antigens expressed on the neuronal cell surface of the CNS or at the neuromuscular junction are mediated by direct antibody effects on the target antigens and are more responsive to immunotherapy (Tables 99-3 and 99-4, **Fig. 99-1**). These disorders occur with and without a cancer association and may affect children and young adults. Some disorders are triggered by viral encephalitis such as herpes simplex virus encephalitis or Japanese encephalitis leading to autoimmune encephalitis.

TABLE 99-3 Intermediate-Risk Antibodies (30–70% Probability of an Underlying Cancer), Syndromes, and Associated Tumors

ANTIBODY[a]	NEUROLOGIC SYNDROME	TUMOR TYPE WHEN ASSOCIATED
Anti-NMDAR[b]	Anti-NMDAR encephalitis	Teratoma in young women (children and men rarely have tumors)
Anti-AMPAR[b]	Limbic encephalitis with relapses	SCLC, thymoma, breast
Anti-GABA$_B$R[c]	Limbic encephalitis with early and prominent seizures	SCLC
Anti-Caspr2[b]	Morvan syndrome	Thymoma
Anti-mGluR5[b]	Autoimmune encephalitis without distinctive features	Hodgkin lymphoma
Anti-VGCC[b]	LEMS, cerebellar degeneration	SCLC

[a] All the antibodies of this table are against neuronal cell-surface proteins. [b] A direct pathogenic role of these antibodies has been demonstrated in cultured neurons or animal models. [c] These antibodies are strongly suspected to be pathogenic.

Abbreviations: AMPAR, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor; Caspr2, contactin-associated protein-like 2; GABA$_B$R, γ-aminobutyric acid B receptor; LEMS, Lambert-Eaton myasthenic syndrome; mGluR, metabotropic glutamate receptor; NMDAR, *N*-methyl-D-aspartate receptor; SCLC, small-cell lung cancer; VGCC, voltage-gated calcium channel.

TABLE 99-4 Low-Risk Antibodies (<30% Probability of an Underlying Cancer), Syndromes, and Associated Tumors

ANTIBODY[a]	NEUROLOGIC SYNDROME	TUMOR TYPE WHEN ASSOCIATED
Anti-LGI1[b]	Limbic encephalitis, hyponatremia, faciobrachial dystonic seizures	Rarely thymoma
Anti-GABA$_A$R[b]	Encephalitis with prominent seizures and status epilepticus	Thymoma in ~20% of the patients
Anti-DPPX[b]	Agitation, myoclonus, tremor, seizures, hyperekplexia, encephalomyelitis with rigidity	No cancer, but frequent diarrhea or cachexia suggesting paraneoplasia
Anti-glycine receptor[b]	PERM, stiff-person syndrome	Rarely, thymoma, lung, Hodgkin's lymphoma
Anti-IgLON5[b]	NREM and REM sleep disorder, brainstem dysfunction, movement disorder, obstructive sleep apnea, stridor	No tumor association
Anti-GAD65	Stiff-person, cerebellar syndrome, encephalitis with seizures, limbic encephalitis	Infrequent tumor association (thymoma, lung, and breast cancer)
Anti-mGluR1[b]	Cerebellar syndrome	Hodgkin's lymphoma in <10% of patients
Anti-GFAP	Meningoencephalitis and myelitis	No cancer association
Anti-aquaporin 4[b]	Neuromyelitis optica spectrum disorders	Lung and breast adenocarcinomas in a few patients
Anti-MOG	ADEM, optic neuritis, myelitis, cortical encephalitis	No cancer association
Anti-AChR (muscle)[b]	Myasthenia gravis	Thymoma
Anti-AChR (neuronal)[b]	Autonomic ganglionopathy	SCLC

[a]All the antibodies of this table are against cell-surface proteins except for GAD65 and GFAP which are intracellular proteins. Rare antibodies reported in <30 patients are not included. [b]A direct pathogenic role of these antibodies has been demonstrated in cultured neurons or animal models.

Abbreviations: AChR, acetylcholine receptor; ADEM, acute disseminated encephalomyelitis; DPPX, dipeptidyl-peptidase-like protein-6; GABA$_A$R, γ-aminobutyric acid A receptor; GAD, glutamic acid decarboxylase; GFAP, glial fibrillary acidic protein; LGI1, leucine-rich glioma-inactivated 1; mGluR, metabotropic glutamate receptor; MOG, myelin oligodendrocyte glycoprotein; NREM, non–rapid eye movement; PERM, progressive encephalomyelitis with rigidity and myoclonus; REM, rapid eye movement; SCLC, small-cell lung cancer.

In patients with cancer, the use of immune checkpoint inhibitors is associated in <3% of cases with neurologic immune-related adverse events that may be accompanied by neuronal antibodies, in which case the associated syndromes are indistinguishable from PNDs.

Other PNDs are likely immune-mediated, although their antigens are unknown. The best example is opsoclonus-myoclonus syndrome associated with neuroblastoma or SCLC. For still other PNDs, the cause remains quite obscure. These include, among others, several neuropathies that occur in the terminal stages of cancer and a number of neuropathies associated with plasma cell dyscrasias or lymphoma without evidence of tumor infiltration or deposits of immunoglobulin, cryoglobulin, or amyloid.

APPROACH TO THE PATIENT

Paraneoplastic Neurologic Disorders

Three key concepts are important for the diagnosis and management of PNDs. First, it is common for symptoms to appear before the presence of a tumor is known; second, the neurologic syndrome usually develops rapidly, producing severe deficits in a short period of time; and third, there is evidence that prompt tumor control improves the neurologic outcome. Therefore, the major concern of the physician is to recognize a disorder as paraneoplastic as early as possible and to identify and treat the tumor.

PND OF THE CENTRAL NERVOUS SYSTEM AND DORSAL ROOT GANGLIA

When symptoms involve brain, spinal cord, or dorsal root ganglia, the suspicion of PND is usually based on a combination of clinical, radiologic, and CSF findings. Presence of antineuronal antibodies (Tables 99-2, 99-3, and 99-4) may help in the diagnosis, but only 60–70% of PNDs of the CNS and <20% of those involving the peripheral nervous system have neuronal or neuromuscular junction antibodies that can be used as diagnostic tests.

Magnetic resonance imaging (MRI) and CSF studies are important to rule out neurologic complications due to the direct spread of cancer, particularly metastatic and leptomeningeal disease. In most PNDs, the MRI findings are nonspecific. Paraneoplastic limbic encephalitis is usually associated with characteristic MRI abnormalities in the mesial temporal lobes (see below), but similar findings can occur with other disorders (e.g., nonparaneoplastic autoimmune limbic encephalitis and human herpesvirus type 6 [HHV-6] encephalitis) **(Fig. 99-2*A*)**. The CSF profile of patients with PND of the CNS or dorsal root ganglia typically consists of mild to moderate pleocytosis (<200 mononuclear cells, predominantly lymphocytes), an increase in the protein concentration, and a variable presence of oligoclonal bands. There are no specific electrophysiologic tests that are diagnostic of PND of the CNS. Moreover, a biopsy of the affected tissue is often difficult to obtain, and although useful to rule out other disorders (e.g., metastasis), the pathologic findings are not specific for PND.

PND OF NERVE AND MUSCLE

If symptoms involve peripheral nerve, neuromuscular junction, or muscle, the diagnosis of a specific PND is usually established on clinical, electrophysiologic, and pathologic grounds. The clinical history, accompanying symptoms (e.g., anorexia, weight loss), and type of syndrome dictate the studies and degree of effort needed to demonstrate a neoplasm. For example, the frequent association of Lambert-Eaton myasthenic syndrome (LEMS) with SCLC should lead to a chest and abdomen computed tomography (CT) or body positron emission tomography (PET) scan and, if negative, periodic tumor screening for at least 3 years after the neurologic diagnosis. In contrast, the weak association of polymyositis with cancer calls into question the need for repeated cancer screenings in this situation. Serum and urine immunofixation studies should be considered in patients with peripheral neuropathy of unknown cause; detection of a monoclonal gammopathy suggests the need for additional studies to uncover a B-cell or plasma-cell malignancy. In paraneoplastic neuropathies, diagnostically useful neuronal antibodies are limited to CRMP5 (CV2) and Hu (ANNA1).

For any type of PND, if neuronal antibodies are negative, the diagnosis relies on the demonstration of cancer and the exclusion of other cancer-related or independent neurologic disorders. Combined CT and PET scans often uncover tumors undetected by other tests. For germ cell tumors of the testis and teratomas of the ovary, ultrasound (testicular, transvaginal, or pelvic) and MRI or CT of the abdomen and pelvis may reveal tumors undetectable by PET.

SPECIFIC PARANEOPLASTIC NEUROLOGIC SYNDROMES

PARANEOPLASTIC ENCEPHALOMYELITIS AND FOCAL ENCEPHALITIS WITH HIGH-RISK ANTIBODIES FOR AN UNDERLYING CANCER

The term *encephalomyelitis* describes an inflammatory process with multifocal involvement of the nervous system, including brain, brainstem, cerebellum, and spinal cord. It is often associated with dorsal root ganglia and autonomic dysfunction. For any given patient, the clinical

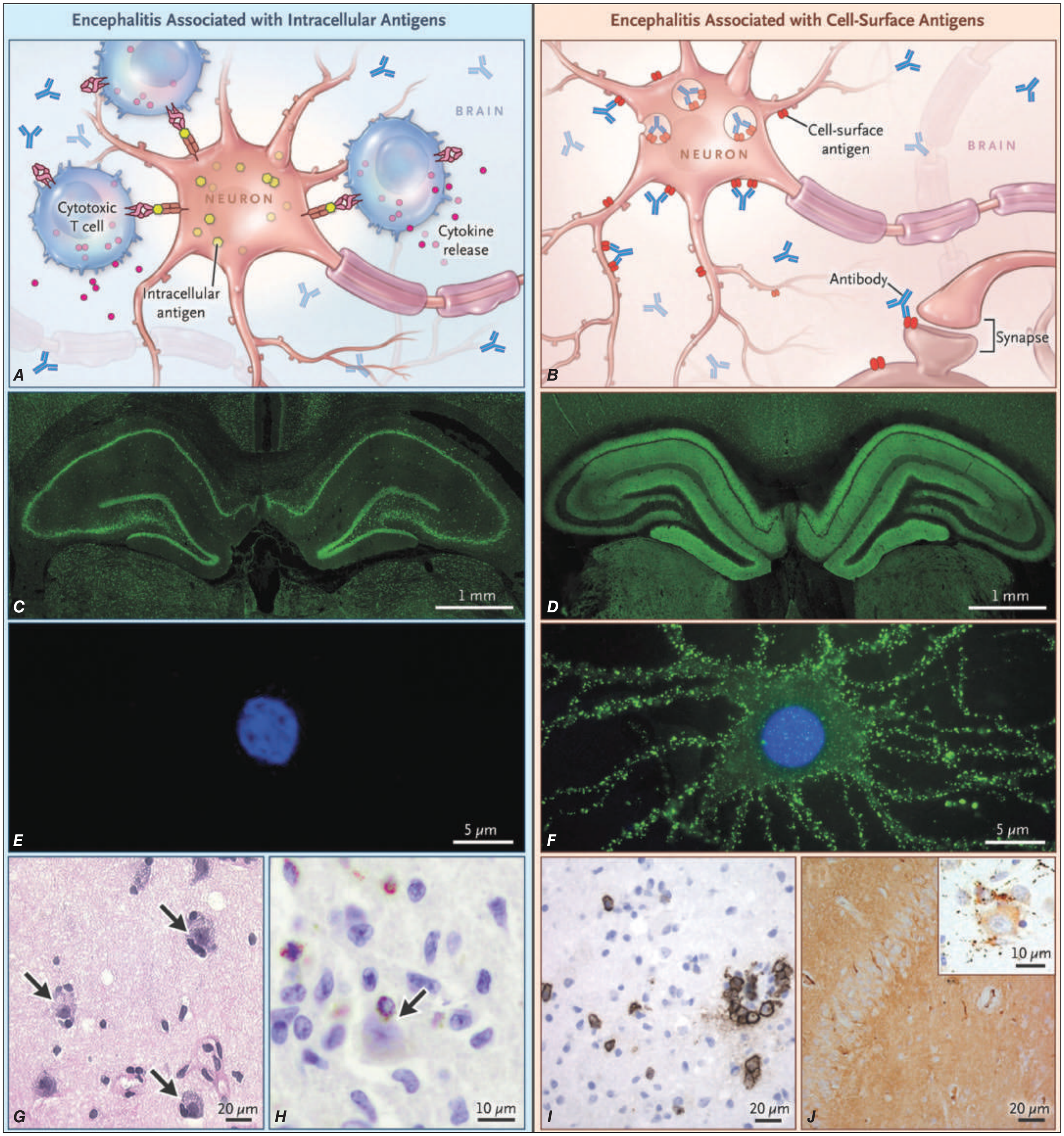
Encephalitis Associated with Intracellular Antigens
Encephalitis Associated with Cell-Surface Antigens
BRAIN
NEURON
Cytotoxic T cell
Cytokine release
Intracellular antigen
A
Cell-surface antigen
NEURON
BRAIN
Antibody
Synapse
B
1 mm
C
1 mm
D
5 μm
E
5 μm
F
20 μm
G
10 μm
H
20 μm
I
10 μm
20 μm
J

FIGURE 99-1 Antibody reactivity and pathologic findings in patients with antibodies against intracellular antigens compared with those of patients with antibodies against neuronal surface antigens. In encephalitis associated with antibodies against intracellular antigens, the antibodies cannot reach the intracellular epitopes and cytotoxic T-cell mechanisms are predominantly involved (***A***), whereas in encephalitis with antibodies against surface antigens, the antibodies have access to the epitopes and can potentially alter the structure and function of the antigen (***B***). The Hu antibodies (***C, E***) are shown here to exemplify the group of antibodies against intracellular antigens, and the NMDAR antibodies (***D, F***) are shown to exemplify the group of antibodies against cell-surface antigens. In rodent brain immunofluorescence with tissue permeabilized to allow entry of antibodies, the Hu antibodies produce a discrete pattern of cellular immunolabeling (***C***), whereas the NMDAR antibodies produce a pattern of neuropil-like immunolabeling (***D***). In contrast, with live cultured neurons, only the NMDAR antibodies have access to the target antigen showing intense immunolabeling (***F***), whereas the Hu antibodies cannot reach the intracellular antigen showing no immunolabeling (***E***). In autopsy studies, patients with encephalitis associated with antibodies to intracellular antigens (Hu or other) have extensive neuronal loss and inflammatory infiltrates (not shown); the T cells show direct contact with neurons (*arrows* in ***G***) likely contributing to neuronal degeneration via perforin and granzyme mechanisms (*arrow* in ***H***). In contrast, patients with antibodies against cell-surface antigens (NMDAR shown here, and probably applicable to other antigens) have moderate brain inflammatory infiltrates along with plasma cells (brown cells in ***I***), deposits of IgG (diffuse brown staining in ***J***), and microglial proliferation (inset in ***J***), without evidence of predominant T-cell–mediated neuronal loss (not shown). All human tissue sections (***G-J***) were obtained from hippocampus. *(From J Dalmau: Antibody mediated encephalitis. N Engl J Med 378:840, 2018. Copyright © 2018 Massachusetts Medical Society. Reprinted with permission.)*

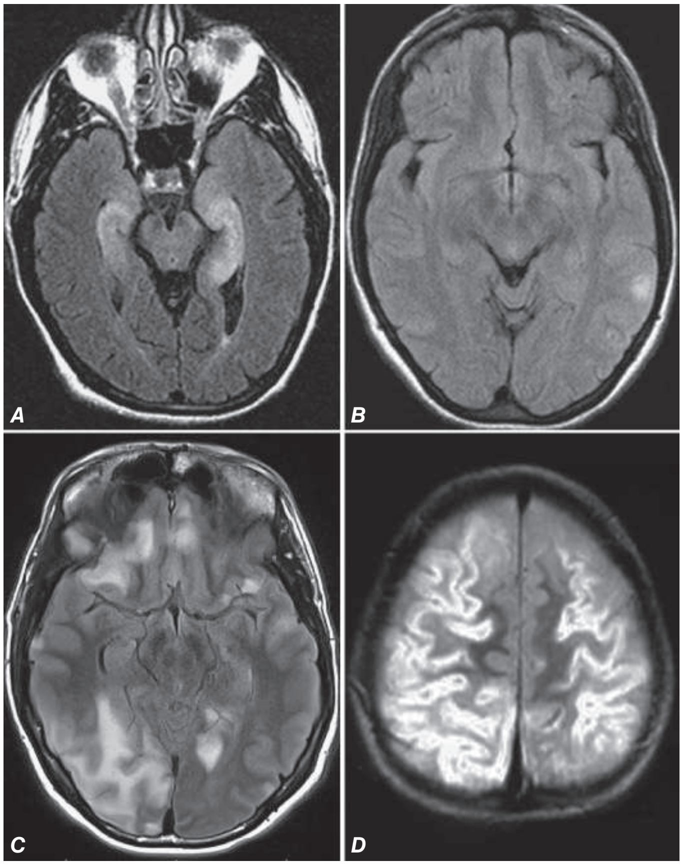

FIGURE 99-2 Brain MRI findings in paraneoplastic and autoimmune encephalitis. Representative MRI studies of patients with several types of autoimmune encephalitides. ***A.*** Limbic encephalitis (LE) may result from several different immune responses (Hu, Ma2, AMPAR, GABA$_B$R, LGI1, Caspr2) and typically manifests with unilateral or bilateral medial temporal lobe increased FLAIR signal. ***B.*** Anti-NMDAR encephalitis often occurs with normal MRI findings or mild FLAIR signal abnormalities. ***C.*** In contrast, anti-GABA$_A$R encephalitis usually occurs with multiple cortical-subcortical increased FLAIR signal changes. ***D.*** Cortical encephalitis can occur in patients with myelin oligodendrocyte glycoprotein (MOG) antibodies, as shown in this T2-weighted MRI image from a 3-year-old boy who presented with extensive cortical abnormalities with mild enhancement (not shown here) suggesting cortical necrosis. *(Panels A-C from J Dalmau: Antibody mediated encephalitis. N Engl J Med 378:840, 2018. Copyright © 2018 Massachusetts Medical Society. Reprinted with permission. Panel D from T Armangue: Associations of paediatric demyelinating and encephalitic syndromes with myelin oligodendrocyte glycoprotein antibodies: A multicentre observational study. Lancet Neurol 19:234, 2020.)*

manifestations are determined by the areas predominantly involved, but pathologic studies almost always reveal abnormalities beyond the symptomatic regions. Several clinicopathologic syndromes may occur alone or in combination: (1) *cortical encephalitis*, which may present as "epilepsia partialis continua"; (2) *limbic encephalitis*, characterized by confusion, depression, agitation, anxiety, severe deficit forming new memories ("short-term memory deficit"), and temporal lobe or generalized seizures (the MRI usually shows unilateral or bilateral medial temporal lobe abnormalities, best seen with T2 and fluid-attenuated inversion recovery [FLAIR] sequences); (3) *brainstem encephalitis,* resulting in eye movement disorders (nystagmus, opsoclonus, supranuclear or nuclear paresis), cranial nerve paresis, dysarthria, dysphagia, unsteady gait, and central autonomic dysfunction; (4) *cerebellar gait and limb ataxia*; (5) *myelitis*, which may cause lower or upper motor neuron symptoms, myoclonus, muscle rigidity, spasms, fasciculations, sensory deficits, and sphincter dysfunction; and (6) *autonomic dysfunction* as a result of involvement of the neuraxis at multiple levels, including hypothalamus, brainstem, and autonomic nerves (see "Paraneoplastic Peripheral Neuropathies," below). Cardiac arrhythmias, postural hypotension, and central hypoventilation can be the cause of death in patients with encephalomyelitis.

Paraneoplastic encephalomyelitis and focal encephalitis are usually associated with SCLC, but many other cancers have been implicated. Patients with SCLC and these syndromes usually have Hu antibodies in serum and CSF. CRMP5 antibodies occur less frequently; some of these patients may develop chorea, uveitis, or optic neuritis. Antibodies to Ma proteins are associated with limbic, hypothalamic, and brainstem encephalitis and occasionally with cerebellar symptoms; some patients develop hypersomnia, cataplexy, and severe hypokinesia. MRI abnormalities are frequent, including those described with limbic encephalitis and variable involvement of the hypothalamus, basal ganglia, or upper brainstem. Kelch-like protein 11 antibodies are predominantly associated with brainstem encephalitis, vertigo, sensorineural hearing loss, and seminomas, germ cell tumors, and teratomas. Amphiphysin antibodies usually are associated with paraneoplastic stiff-person syndrome, but in some patients, they can occur with paraneoplastic encephalomyelitis or isolated myelitis. The oncologic associations of these antibodies are shown in Table 99-2.

Most types of paraneoplastic encephalitis and encephalomyelitis in which the antigens are intracellular respond poorly to treatment. Stabilization of symptoms or partial neurologic improvement may occur, particularly if there is a satisfactory response of the tumor to treatment. Controlled trials of therapy are lacking, but many reports and the opinion of experts suggest that therapies aimed to remove the antibodies against intracellular antigens, such as intravenous immunoglobulin (IVIg) or plasma exchange, usually fail. The main concern should be to treat the tumor and consider immunotherapies aimed at cytotoxic T-cell responses. Approximately 30% of patients with anti-Ma2-associated encephalitis respond to treatment of the tumor (usually a germ cell neoplasm of the testis) and immunotherapy.

Cortical encephalitis can occur with antibodies against myelin oligodendrocyte glycoprotein (MOG) **(Chap. 456)**, and encephalomyelitis can occur with antibodies against GFAP **(Chap. 456)**. These two disorders rarely associate with cancer and respond better to immunotherapy than the indicated paraneoplastic syndromes. Encephalitis with seizures, limbic encephalitis, stiff-person syndrome, or cerebellar ataxia can occur with GAD65 antibodies; this type of autoimmunity rarely associates with cancer but responds less frequently to immunotherapy than similar syndromes associated with antibodies against neuronal surface proteins.

ENCEPHALITIDES WITH INTERMEDIATE- OR LOW-RISK ANTIBODIES FOR AN UNDERLYING CANCER

These disorders are important for four reasons: (1) they can occur with and without tumor association; less frequently, they develop after a viral encephalitis (herpes simplex or Japanese encephalitis); (2) some syndromes predominate in young individuals and children; (3) despite the severity of the symptoms, patients usually respond to treatment of the tumor, if found, and immunotherapy (e.g., glucocorticoids, IVIg, plasma exchange, rituximab, or cyclophosphamide); and (4) for many of these disorders, the antibody pathogenicity has been demonstrated in models using cultures of neurons or passive transfer of patients' antibodies to animals **(Fig. 99-3)**.

Encephalitis with N*-methyl-*D*-aspartate receptor (NMDAR) antibodies* usually occurs in young women and children, but men and older patients of both sexes can be affected. The disorder has a characteristic pattern of symptom progression that often includes a prodrome resembling a viral process, followed in a few days by the onset of severe psychiatric symptoms, sleep dysfunction (usually insomnia), reduced verbal output, memory loss, seizures, decreased level of consciousness, abnormal movements (orofacial, limb, and trunk dyskinesias, dystonic postures), autonomic instability, and frequent hypoventilation. Monosymptomatic episodes, such as pure psychosis, occur in about 5% of patients. Clinical relapses occur in 12–24% of patients (12% during the first 2 years after initial presentation). Most patients have intrathecal synthesis of antibodies, likely by infiltrating plasma cells in brain and meninges **(Fig. 99-1*I*)**. In about 65% of patients, the brain MRI is normal; in the other 35%, it shows FLAIR abnormalities that can affect cortical and subcortical regions, usually mild and transient, and rarely

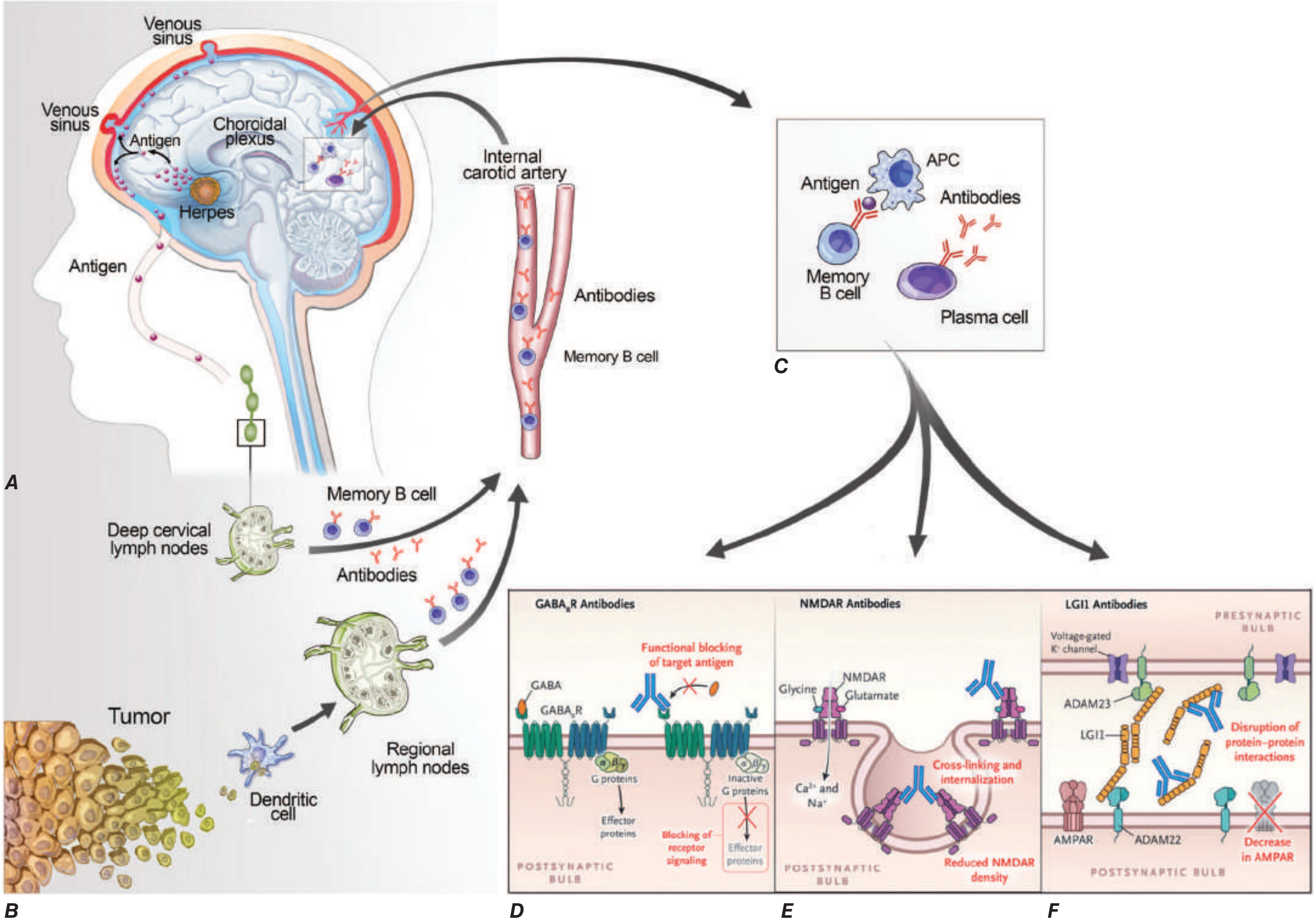
Venous sinus
Venous sinus
Antigen
Choroidal plexus
Herpes
Antigen
Internal carotid artery
Antibodies
Memory B cell
A
Deep cervical lymph nodes
Memory B cell
Antibodies
Tumor
Regional lymph nodes
Dendritic cell
B
Antigen
APC
Antibodies
Memory B cell
Plasma cell
C
GABA$_B$R Antibodies
Functional blocking of target antigen
GABA
GABA$_B$R
G proteins
Effector proteins
Inactive G proteins
Blocking of receptor signaling
Effector proteins
POSTSYNAPTIC BULB
D
NMDAR Antibodies
Glycine
NMDAR
Glutamate
Ca^{2+} and Na$^+$
Cross-linking and internalization
Reduced NMDAR density
POSTSYNAPTIC BULB
E
LGI1 Antibodies
PRESYNAPTIC BULB
Voltage-gated K$^+$ channel
ADAM23
LGI1
Disruption of protein–protein interactions
AMPAR
ADAM22
Decrease in AMPAR
POSTSYNAPTIC BULB
F

FIGURE 99-3 Proposed mechanisms of disease and functional interactions of autoantibodies with neuronal surface proteins. The graph shows a multistep process that results in antibody-mediated neuronal dysfunction; some of the steps have been demonstrated in reported studies, whereas others are based on proposed hypotheses. Two well-known triggers of autoimmune encephalitides are represented: herpes simplex encephalitis (***A***) and systemic tumors (***B***); the genetic susceptibility of some autoimmune encephalitides and unknown immunologic triggers are not depicted. It is postulated that antigens released by viral-induced neuronal destruction or apoptotic tumor cells are loaded into antigen-presenting cells (APCs; dendritic cells) and transported to regional lymph nodes. In the lymph nodes, naïve B cells exposed to the processed antigens, with cooperation of CD4+ T cells, become antigen-experienced and differentiate into antibody-producing plasma cells. After entering the brain, memory B cells undergo restimulation, antigen-driven affinity maturation, clonal expansion, and differentiation into antibody-producing plasma cells (***C***). The contribution of systemically produced antibodies to the pool of antibodies present in the brain is unclear and may depend on systemic antibody titers and integrity of the blood-brain barrier. Based on experimental models with cultured neurons, the presence of antibodies in the brain may lead to neuronal dysfunction by different mechanisms, including functional blocking of the target antigen (GABA$_B$R antibodies; ***D***), receptor crosslinking and internalization (NMDAR antibodies; ***E***), and disruption of protein-protein interaction, leading to downstream effects on receptors (LGI1 leading to a decrease of Kv1 potassium channels and AMPAR; ***F***). These mechanisms are influenced by the type of antibodies; for example, whereas IgG1 antibodies frequently crosslink and internalize the target antigen, IgG4 antibodies are less effective at crosslinking the target and more often alter protein-protein interactions. *(Panels D-F from J Dalmau: Antibody mediated encephalitis. N Engl J Med 378:840, 2018. Copyright © 2018 Massachusetts Medical Society. Reprinted with permission.)*

the presence of contrast enhancement **(Fig. 99-2*B*)**. The syndrome may be misdiagnosed as a viral or idiopathic encephalitis, neuroleptic malignant syndrome, or encephalitis lethargica, and some patients are initially evaluated by psychiatrists with the suspicion of acute psychosis as the presentation of a primary psychiatric disease. The detection of an associated teratoma is dependent on age and gender: 46% of female patients 12 years or older have uni- or bilateral ovarian teratomas, whereas <7% of girls younger than 12 have a teratoma **(Fig. 99-4*A*)**. In young male patients, the detection of a tumor is rare. Patients older than 45 years are more frequently male; about 20% of these patients have tumors (e.g., cancer of the breast, ovary, or lung). Prompt diagnosis and treatment with immunotherapy (and tumor removal when it applies) improve outcome. Treatment usually includes first-line immunotherapy (steroids, IVIg, or plasma exchange), and if there is no response, second-line immunotherapy (rituximab and cyclophosphamide). Rituximab is highly effective and is increasingly used as part of first-line treatments. The response to treatment can take several weeks, and the improvement is usually slow. Overall, about 85–90% of patients have substantial neurologic improvement or full recovery. Deficits of attention, memory, and executive functions may recover slowly over many months. During the process of recovery, patients often develop transient hypersomnia and, less frequently, hyperphagia, hypersexuality, apathy, and irritability, resembling a Kleine-Levin syndrome.

Approximately 25% of patients with herpes simplex encephalitis develop a form of autoimmune encephalitis that usually is associated with abnormal movements (choreoathetosis after herpes simplex encephalitis) in children and with cognitive and psychiatric symptoms in adults. This disorder develops a few weeks after the viral infection has resolved, is associated with new synthesis of antibodies against the NMDAR and other neuronal cell surface proteins, and is usually less responsive to immunotherapy than anti-NMDAR encephalitis (idiopathic or teratoma-associated).

Encephalitis with α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor antibodies affects middle-aged women, who develop acute limbic dysfunction or, less frequently, prominent psychiatric symptoms; 70% of patients have an underlying tumor in the lung, breast, or thymus **(Fig. 99-4*B*)**. In about 50% of cases, the brain MRI shows typical features of limbic encephalitis (similar to Fig. 99-2*A*).

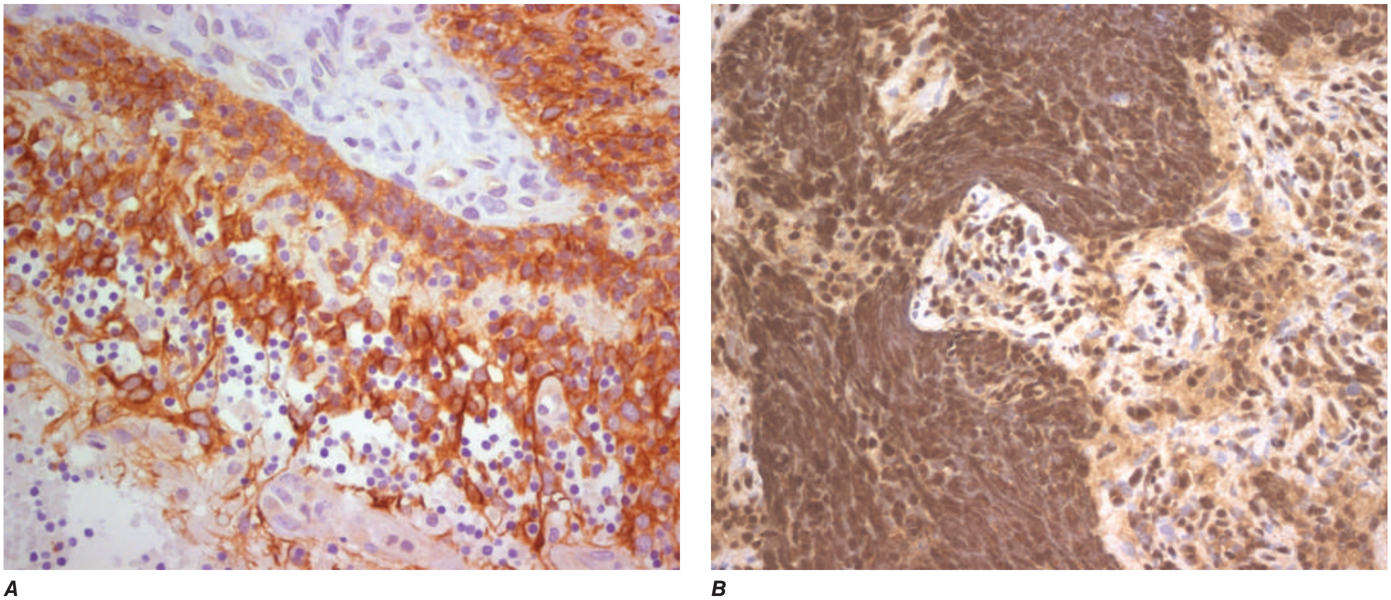

FIGURE 99-4 Immunopathological studies in tumors of patients with autoimmune encephalitis. ***A.*** Neurons and neuronal processes (brown cells; stained with MAP2) in the teratoma of a patient with anti-NMDA receptor encephalitis; these neurons express NMDAR (not shown). ***B.*** Lung cancer from a patient with anti–α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor encephalitis showing expression of AMPA receptors by the neoplastic cells (brown cells). *(Panel B from M Lai et al: AMPA receptor antibodies in limbic encephalitis alter synaptic receptor location. Ann Neurol 65:424, 2009.)*

Neurologic relapses may occur; these also respond to immunotherapy and are not necessarily associated with tumor recurrence.

Encephalitis with GluK2 antibodies can affect children and adults and is associated with rapidly progressive encephalopathy with cerebellar ataxia or cerebellitis. Symptoms of encephalopathy may include impairment of memory and level of consciousness and motor alterations such as dyskinesias, choreoathetosis, bradykinesia, and spastic paraparesis. Some patients develop intracranial hypertension. In one patient, the symptoms were associated with teratoma.

Encephalitis with γ-aminobutyric acid type A (GABA_A) receptor antibodies may affect children and adults and is associated with prominent seizures and status epilepticus often requiring a pharmacologically induced coma. In approximately 80% of patients, the brain MRI shows multifocal, asynchronous, cortical-subcortical T2/FLAIR abnormalities predominantly involving temporal and frontal lobes, but also basal ganglia and other regions **(Fig. 99-2C)**. Most patients do not have an underlying tumor, but some may have thymoma.

Encephalitis with GABA_B receptor antibodies is usually associated with limbic encephalitis and seizures. In >50% of cases, the MRI shows increased medial temporal lobe FLAIR changes characteristic of limbic encephalitis (similar to Fig. 99-2*A*). In rare instances, patients develop cerebellar symptoms and opsoclonus. Fifty percent of patients have SCLC or a neuroendocrine tumor of the lung. The outcome is substantially better in patients without cancer.

Encephalitis with glycine receptor (GlyR) antibodies usually manifests with a syndrome characterized by progressive encephalomyelitis with rigidity and myoclonus (PERM) or stiff-person spectrum of symptoms. The disease usually occurs in adults and rarely in children. About 20% of adult patients have a concurrent underlying tumor (thymoma, B-cell lymphoma, or breast or lung cancer) or past history of cancer (thymoma, breast cancer, Hodgkin's lymphoma, or melanoma).

Encephalitis with metabotropic glutamate receptor 5 (mGluR5) antibodies is characterized by nonspecific clinical features of encephalitis (confusion, agitation, memory loss, delusions, paranoid ideation, hallucinations, psychosis, or seizures) without distinctive MRI changes and frequent association with Hodgkin's lymphoma. The encephalitis is highly responsive to immunotherapy and treatment of the tumor.

Encephalitis with antibodies against dopamine-2 receptor has been reported in children with basal ganglia encephalitis manifesting with abnormal movements (coarse tremor, parkinsonism, chorea, oculogyric crises) along with psychiatric features, lethargy, drowsiness, brainstem dysfunction, or ataxia. The disorder is extremely rare and is not associated with cancer.

Encephalitis with leucine-rich glioma-inactivated 1 (LGI1) antibodies predominates in patients older than 50 years (65% male) and frequently presents with short-term memory loss and seizures (limbic encephalopathy), along with hyponatremia and sleep dysfunction. The MRI often shows increased FLAIR signal in one or both medial temporal lobes. In about 40% of patients, these symptoms are preceded by faciobrachial dystonic seizures, which consist of sudden, short-lasting, mainly distal muscle contractions involving the arm, face, or leg. These are unilateral but can independently affect both sides and occur multiple times during the day or night. About 15% of patients present with rapidly progressive cognitive decline, resembling a rapidly progressive dementia. Less than 5% of patients have thymoma. An association with the human leukocyte antigen (HLA) haplotypes DRB1*07:01, DQB1*02:02, DQA1*02:01, and DRB4 has been identified. All symptoms, including faciobrachial dystonic seizures, respond to immunotherapy, although about two-thirds of patients are left with memory or cognitive deficits. Despite the improvement with initial immunotherapy, video-polysomnography or prolonged electroencephalogram studies frequently uncover persistent alterations (faciobrachial dystonic seizures, other seizures, rapid eye movement [REM] sleep behavior disorder) only detectable during sleep. These alterations, which might contribute to residual memory or cognitive deficits, respond to additional or prolonged immunotherapy.

Encephalitis with contactin-associated protein-like 2 (Caspr2) antibodies predominates in patients older than 50 years and is associated with a form of encephalitis with three or more of the following core symptoms: encephalopathy (which can be similar to limbic encephalitis), cerebellar symptoms, peripheral nervous system hyperexcitability, dysautonomia, insomnia, neuropathic pain, and weight loss. Patients with Morvan syndrome, which includes clinical features of encephalitis (confusion, hallucinations, prominent sleep dysfunction, or "agrypnia excitata"), autonomic alterations, and peripheral nerve hyperexcitability or neuromyotonia, usually have Caspr2 antibodies. About 20% of patients with Caspr2 antibody–associated syndromes have thymoma; this percentage is higher (~50%) in patients with Morvan syndrome. An association of Caspr2 antibody–associated syndromes with HLA DRB1*11.01 has been reported.

Encephalitis with dipeptidyl-peptidase-like protein-6 (DPPX) antibodies is usually preceded or develops concurrently with diarrhea, other gastrointestinal symptoms, and substantial loss of weight that often suggest the presence of a gastrointestinal disease. Neurologic symptoms include agitation, hallucinations, paranoid delusions, and features of CNS hyperexcitability such as hyperekplexia, tremor, myoclonus, nystagmus, or seizures. Some patients develop a clinical picture similar to progressive encephalomyelitis with rigidity and myoclonus. The few patients reported with an associated tumor usually had B-cell neoplasms.

Encephalitis with antibodies against neurexin 3 alpha does not have distinctive clinical features; the experience is limited, and the disorder does not appear to be associated with cancer.

Anti-IgLON5 disease is a chronic or subacute encephalopathy that characteristically is associated with REM and non-REM (NREM) parasomnia along with obstructive sleep apnea that may be preceded or accompanied by bulbar symptoms (mainly dysphagia and stridor due to vocal cord palsy), unsteadiness, movement disorders (chorea, craniofacial dyskinesias), oculomotor dysfunction, and, in less than half of cases, cognitive decline. The median age of the patients is in the early 60s, and men and women are equally affected. The sleep disorder is characterized by abnormal sleep initiation with undifferentiated NREM sleep associated with frequent vocalizations and quasi-purposeful movements. Brain MRI is unrevealing or demonstrates minor changes of unclear clinical relevance; the CSF is usually normal or may show transient lymphocytic pleocytosis. It is not associated with cancer but shows a strong association with the HLA-DRB1*10:01 and HLA-DQB1*05 haplotypes, which are present in 60% of patients. The response to immunotherapy is poor. In some patients, neuropathologic studies show a neuronal tauopathy predominantly involving the hypothalamus and tegmentum of the brainstem; however, not all autopsy studies show deposits of hyperphosphorylated tau, and in some patients, mild inflammatory infiltrates have been identified.

With the exception of patients with anti-IgLON5 disease, who rarely respond to treatment, most patients with autoimmune or paraneoplastic encephalopathies associated with antibodies against cell-surface or synaptic proteins respond to immunotherapy and treatment of the tumor (if appropriate). Although there are no specific standardized treatment protocols, the most frequent approach is similar to that indicated for anti-NMDAR encephalitis and consists of progressive escalation of immunotherapy using first a combination of glucocorticoids, IVIg, and plasma exchange, and then, if there is no response, rituximab or cyclophosphamide.

Encephalitis with MOG antibodies can present with a clinical picture suggestive of autoimmune encephalitis related to neuronal antibodies. Most patients with MOG antibody–associated syndromes are children and young adults who present with optic neuritis, myelitis, brainstem (mainly pons) and cerebellar symptoms, or acute disseminated encephalomyelitis (ADEM). About 85% of patients with these syndromes respond to immunotherapy, although relapses occur in about 30% of cases. Besides these syndromes, there is a small group of adults and children who present with unilateral or bilateral cortical encephalitis, and their response to treatment is variable. In children, two phenotypes of poor prognosis include ADEM-like relapses progressing to leukodystrophy-like features and extensive cortical encephalitis evolving to atrophy **(Fig. 99-2*D*)**. In general, MOG antibody syndromes are not associated with tumors.

PARANEOPLASTIC CEREBELLAR DEGENERATION

This disorder is often preceded by a prodrome that may include dizziness, oscillopsia, blurry or double vision, nausea, and vomiting. A few days or weeks later, patients develop dysarthria, gait and limb ataxia, and variable dysphagia. The examination usually shows downbeating nystagmus and, rarely, opsoclonus. Brainstem dysfunction or upgoing toes may occur. Early in the course, MRI studies are usually normal; later, the MRI reveals cerebellar atrophy. The disorder results from extensive degeneration of Purkinje cells, with variable involvement of other cerebellar cortical neurons, deep cerebellar nuclei, and spinocerebellar tracts. The tumors more frequently involved are SCLC, cancer of the breast and ovary, and Hodgkin's lymphoma.

Anti-Yo (PCA1) antibodies in patients with breast or gynecologic cancers typically are associated with prominent or pure cerebellar degeneration. A variable degree of cerebellar dysfunction can be associated with virtually any of the antibodies and PND of the CNS shown in Table 99-2. A number of single case reports have described neurologic improvement after tumor removal, plasma exchange, IVIg, cyclophosphamide, rituximab, or glucocorticoids. However, most patients with paraneoplastic cerebellar degeneration and any of the antibodies shown in Table 99-2 do not improve with treatment; an exception are the cerebellar syndromes associated with antibodies against surface antigens (e.g., Tr antibodies and Hodgkin's lymphoma, VGCC antibodies and SCLC), which are more responsive to immunotherapy.

Antibody-associated cerebellar symptoms can also occur without cancer association. The most common is cerebellar ataxia with antibodies against GAD65 (an intracellular synaptic protein). The median age at diagnosis is 60 years and 75% of cases are women. Concurrent organ-specific autoimmune diseases, mainly type 1 diabetes, are present in 80% of patients. The clinical course is chronic, but in 40% of patients, the presentation is subacute (<6 months). Response to immunotherapy is poor. Subacute onset of symptoms and prompt initiation of immunotherapy are associated with a better chance for improvement. By contrast, the cerebellar syndrome associated with mGluR1 antibodies is more responsive to immunotherapy, and a substantial improvement occurs in almost half of the patients. Immunotherapy is also effective in patients with GluK2 antibodies, who can present with cerebellitis and posterior fossa edema with compression of the fourth ventricle, and in those with septin-5 antibodies. Although septin-5 is an intracellular antigen, it is likely exposed to the cell surface during the process of endo- and exocytosis at synaptic terminals.

PARANEOPLASTIC OPSOCLONUS-MYOCLONUS SYNDROME

Opsoclonus is a disorder of eye movement characterized by involuntary, chaotic saccades that occur in all directions of gaze; it is frequently associated with myoclonus and ataxia. Opsoclonus-myoclonus may be cancer-related or idiopathic. When the cause is paraneoplastic, the tumors involved are usually cancer of the lung and breast in adults and neuroblastoma in children. The pathologic substrate of opsoclonus-myoclonus is unclear, but studies suggest that disinhibition of the fastigial nucleus of the cerebellum is involved. Most patients do not have neuronal antibodies. A small subset of patients with ataxia, opsoclonus, and other eye-movement disorders have Ri antibodies; these patients may also develop muscle rigidity, laryngeal spasms, and autonomic dysfunction. The tumors most frequently involved in anti-Ri-associated syndromes are breast, ovarian, and lung cancers. If the tumor is not successfully treated, the syndrome in adults often progresses to encephalopathy, coma, and death. In addition to treating the tumor, symptoms may respond to immunotherapy (glucocorticoids, plasma exchange, and/or IVIg).

At least 50% of children with opsoclonus-myoclonus have an underlying neuroblastoma. Hypotonia, ataxia, behavioral changes, and irritability are frequent accompanying symptoms. Neurologic symptoms often improve with treatment of the tumor and glucocorticoids, adrenocorticotropic hormone (ACTH), plasma exchange, IVIg, rituximab, or cyclophosphamide. Many patients are left with psychomotor retardation and behavioral and sleep problems.

PARANEOPLASTIC SYNDROMES OF THE SPINAL CORD

The number of reports of paraneoplastic spinal cord syndromes, such as *subacute motor neuronopathy* and *acute necrotizing myelopathy*, has decreased over the years. This may represent a true decrease in incidence, due to improved and prompt oncologic interventions, or the identification of nonparaneoplastic etiologies. Some patients with cancer or lymphoma develop *upper* or *lower motor neuron dysfunction* or both, resembling amyotrophic lateral sclerosis. It is unclear whether these disorders have a paraneoplastic etiology or simply coincide with the presence of cancer.

Paraneoplastic myelitis may present with upper or lower motor neuron symptoms, segmental myoclonus, sensory deficits, sphincter dysfunction, and neurogenic pruritus and can be the first manifestation of encephalomyelitis. The spine MRI usually shows longitudinally extensive, symmetric tract or gray matter abnormalities in the spinal cord. It is mainly associated with breast and lung carcinomas and with CRMP5 or amphiphysin antibodies. The prognosis is poor. *Neuromyelitis optica (NMO) with aquaporin 4 antibodies* may occur in rare instances as a paraneoplastic manifestation of a cancer. NMO is discussed in detail in **Chap. 456**.

■ PARANEOPLASTIC STIFF-PERSON SYNDROME

This disorder is characterized by progressive muscle rigidity, stiffness, and painful spasms triggered by auditory, sensory, or emotional stimuli. Rigidity mainly involves the lower trunk and legs, but it can affect the upper extremities and neck. Sometimes, only one extremity is affected (*stiff-limb syndrome*). Symptoms improve with sleep and general anesthetics. Electrophysiologic studies demonstrate continuous motor unit activity. The associated antibodies target proteins (GAD65, amphiphysin) involved in the function of inhibitory synapses that use γ-aminobutyric acid (GABA) or glycine as neurotransmitters. The presence of amphiphysin antibodies usually indicates a paraneoplastic etiology related to SCLC and breast cancer. By contrast, GAD antibodies may occur in some cancer patients but are much more frequently present in the nonparaneoplastic disorder. GlyR antibodies may occur in some patients with stiff-person syndrome; these antibodies are more frequently detectable in patients with PERM **(Fig. 99-5)**.

Optimal treatment of stiff-person syndrome requires therapy of the underlying tumor, glucocorticoids, and symptomatic use of drugs that enhance GABA-ergic transmission (diazepam, baclofen, sodium valproate, tiagabine, vigabatrin). IVIg and plasma exchange are transiently effective in some patients, and there are reports of responses to rituximab and CD19 CAR T-cell therapy in patients who did not respond to other treatments.

■ PARANEOPLASTIC SENSORY NEURONOPATHY OR DORSAL ROOT GANGLIONOPATHY

This syndrome is characterized by sensory deficits that may be symmetric or asymmetric, painful dysesthesias, radicular pain, and decreased or absent reflexes. All modalities of sensation and any part of the body including face and trunk can be involved. Special senses such as taste and hearing can also be affected. Electrophysiologic studies show decreased or absent sensory nerve potentials with normal or near-normal motor conduction velocities. Symptoms result from an immune-mediated process that targets the dorsal root ganglia, causing neuronal loss and secondary degeneration of the posterior columns of the spinal cord. The dorsal and, less frequently, the anterior nerve roots and peripheral nerves may also be involved. This disorder often precedes or is associated with encephalomyelitis and autonomic dysfunction and has the same immunologic and oncologic associations (Hu antibodies, SCLC).

As with anti-Hu-associated encephalomyelitis, the therapeutic approach focuses on prompt treatment of the tumor and cytotoxic T-cell–mediated mechanisms. Glucocorticoids occasionally produce clinical stabilization or improvement. The benefit of IVIg and plasma exchange is not proven.

■ PARANEOPLASTIC PERIPHERAL NEUROPATHIES

These disorders may develop any time during the course of the neoplastic disease. Neuropathies occurring at late stages of cancer or lymphoma usually cause mild to moderate sensorimotor deficits due

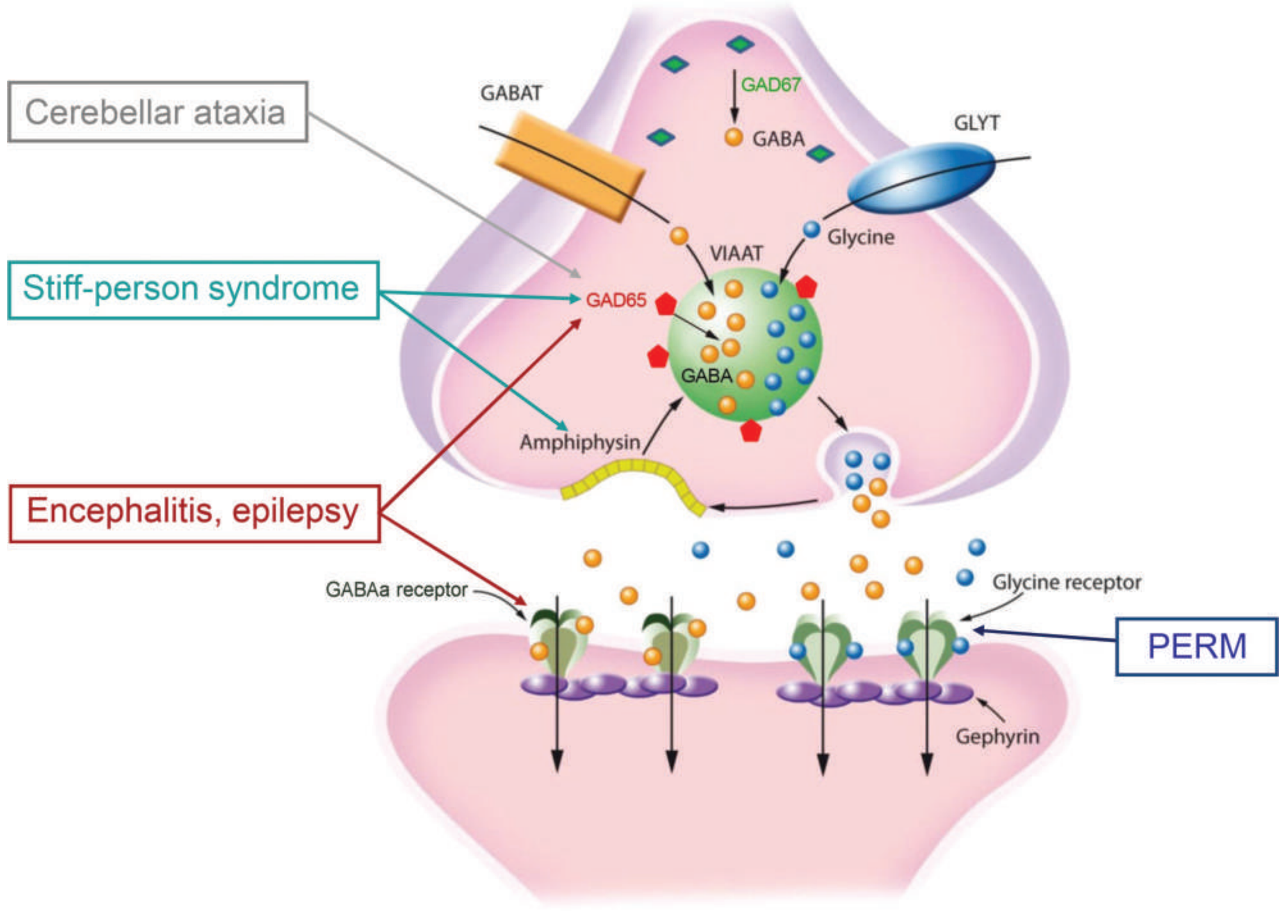

FIGURE 99-5 Schematic representation of an inhibitory synapse showing the main autoimmune targets (GAD, amphiphysin, GABA receptor, and glycine receptor) and the corresponding neurologic disorders. GAD antibodies predominantly occur in stiff-person syndrome (SPS), cerebellar ataxia, and epilepsy, sometimes in the setting of encephalitis. Amphiphysin antibodies are markers of paraneoplastic SPS and breast cancer, GlyR antibodies often associate with progressive encephalomyelitis with rigidity and myoclonus (PERM), and $GABA_A$ receptor antibodies occur in a form of autoimmune encephalitis that is frequently associated with refractory seizures and status epilepticus. *(Modified from F Graus et al: Nat Rev Neurol 16:353, 2020.)*

to axonal degeneration of unclear etiology. These neuropathies are often masked by concurrent neurotoxicity from chemotherapy and other cancer therapies. In contrast, the neuropathies that develop in the early stages of cancer frequently show a rapid progression, sometimes with a relapsing and remitting course, and evidence of inflammatory infiltrates and axonal loss or demyelination. Besides the subacute onset, an asymmetric distribution of symptoms, presence of neuropathic pain and dysautonomia, clinical or neuroimaging evidence of involvement of the spinal cord or nerve roots, and detection of an axonal pattern in the electrophysiological studies are features that suggest the possibility of a paraneoplastic etiology. The response to treatment is usually poor, but if demyelinating features predominate **(Chaps. 457 and 458)**, IVIg, plasma exchange, or glucocorticoids may improve symptoms. Occasionally, CRMP5 antibodies are present; detection of Hu antibodies suggests concurrent dorsal root ganglionitis.

Guillain-Barré syndrome **(Chap. 458)** and *brachial plexitis* **(Chap. 457)** have occasionally been reported in patients with Hodgkin's lymphoma, but there is no clear evidence of a paraneoplastic association.

Diseases associated with paraproteinemia such as multiple myeloma, osteosclerotic myeloma, cryoglobulinemia, amyloidosis, Waldenström's macroglobulinemia, or POEMS (polyneuropathy, organomegaly, endocrinopathy, M-protein spike, and skin manifestations) syndrome, among others, may cause neuropathy by a variety of mechanisms, including compression of roots and plexuses by metastasis to vertebral bodies and pelvis, by deposits of amyloid in peripheral nerves, or through a direct interaction of the abnormal immunoglobulin with peripheral nerve antigens. In other patients, the mechanisms underlying the neuropathy remain unknown and paraneoplastic immune-mediated mechanisms have not been ruled out. Neuropathies more often occur with IgM paraproteinemia followed by IgG and IgA. The phenotype of the neuropathy and likelihood of improvement with successful treatment of the paraproteinemia are dependent on the underlying hematologic disorder **(Chap. 458)**.

Vasculitis of the nerve and muscle causes a painful symmetric or asymmetric distal axonal sensorimotor neuropathy with variable proximal weakness. It predominantly affects elderly men and is associated with an elevated erythrocyte sedimentation rate and increased CSF protein concentration. SCLC and lymphoma are the primary tumors involved. Glucocorticoids and cyclophosphamide often result in neurologic improvement.

Peripheral nerve hyperexcitability (*neuromyotonia*, or *Isaacs' syndrome*) is characterized by spontaneous and continuous muscle fiber activity of peripheral nerve origin. Clinical features include cramps, muscle twitching (fasciculations or myokymia), stiffness, delayed muscle relaxation (pseudomyotonia), and spontaneous or evoked carpal or pedal spasms. The involved muscles may be hypertrophic, and some patients develop paresthesias and hyperhidrosis. The electromyogram (EMG) shows fibrillations; fasciculations; and doublet, triplet, or multiplet single-unit (myokymic) discharges that have a high intraburst frequency. Some patients have Caspr2 antibodies usually in the context of Morvan syndrome, but most patients with isolated neuromyotonia are antibody negative. The disorder often occurs without cancer; if paraneoplastic, benign and malignant thymomas and SCLC are the usual tumors. Phenytoin, carbamazepine, and plasma exchange improve symptoms.

Paraneoplastic autonomic neuropathy usually develops as a component of other disorders, such as LEMS and encephalomyelitis. It may rarely occur as a pure or predominantly autonomic neuropathy with cholinergic or adrenergic dysfunction at the pre- or postganglionic levels. Patients can develop several life-threatening complications, such as gastrointestinal paresis with pseudo-obstruction, cardiac dysrhythmias, and postural hypotension. Other clinical features include abnormal pupillary responses, dry mouth, anhidrosis, erectile dysfunction, and problems with sphincter control. The disorder occurs in association with several tumors, including SCLC, cancer of the pancreas or testis, carcinoid tumors, and lymphoma. Because autonomic symptoms can be the presenting feature of encephalomyelitis, serum Hu and CRMP5 antibodies should be sought. Antibodies to ganglionic (α3-type) neuronal acetylcholine receptors are the cause of autoimmune autonomic ganglionopathy, a disorder that frequently occurs without cancer association **(Chap. 451)**.

LAMBERT-EATON MYASTHENIC SYNDROME

LEMS is discussed in **Chap. 459**.

MYASTHENIA GRAVIS

Myasthenia gravis is discussed in **Chap. 459**.

POLYMYOSITIS-DERMATOMYOSITIS

Polymyositis and dermatomyositis are discussed in detail in **Chap. 377**.

IMMUNE-MEDIATED NECROTIZING MYOPATHY

Patients with this syndrome develop myalgias and rapid progression of weakness involving the extremities, neck, pharyngeal, respiratory, and sometimes cardiac muscles. Serum muscle enzymes are elevated, and muscle biopsy shows extensive necrosis with minimal or absent inflammation and sometimes deposits of complement. The disorder may occur without cancer association (sometimes as a result of statin exposure, connective tissue disease, or HIV) or with cancer association. Patients with antibodies against 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGCR) and seronegative patients are more likely to have an underlying cancer than those with antibodies against signal recognition particle. No specific type of cancer has been found to be predominant. Successful treatment of the tumor and aggressive immunotherapy (steroids, IVIg, and steroid-sparing immunosuppressants) may lead to complete or substantial recovery. Immune-mediated necrotizing myopathy is discussed in **Chap. 377**.

PARANEOPLASTIC VISUAL SYNDROMES

This group of disorders involves the retina and, less frequently, the uvea and optic nerves. The term *cancer-associated retinopathy* is used to describe paraneoplastic cone and rod dysfunction characterized by photosensitivity, progressive loss of vision and color perception, central or ring scotomas, night blindness, and attenuation of photopic and scotopic responses in the electroretinogram (ERG). The most commonly associated tumor is SCLC. Melanoma-associated retinopathy affects patients with metastatic cutaneous melanoma. Patients develop acute onset of night blindness and shimmering, flickering, or pulsating photopsias that often progress to visual loss. The ERG shows reduced b-waves with normal dark adapted a-waves. Paraneoplastic optic neuritis and uveitis can develop in association with encephalomyelitis. Patients with paraneoplastic uveitis and optic neuritis may harbor CRMP5 antibodies.

Some paraneoplastic retinopathies are associated with serum antibodies that specifically react with the subset of retinal cells undergoing degeneration, supporting an immune-mediated pathogenesis. The best-characterized retinal antibodies are against recoverin and have been described in patients with SCLC and cancer-associated retinopathy. However, the specificity of this and other retinal antibodies is low as they have also been reported in nonparaneoplastic autoimmune and nonautoimmune retinopathies. Paraneoplastic retinopathies rarely show substantial improvement after treatment of the tumor and immunotherapy; however, stabilization of symptoms and partial responses to a variety of immunotherapies (glucocorticoids, plasma exchange, IVIg, rituximab, or alemtuzumab) have been reported.

FURTHER READING

Armangue T et al: Associations of paediatric demyelinating and encephalitic syndromes with myelin oligodendrocyte glycoprotein antibodies: A multicentre observational study. Lancet Neurol 19:234, 2020.

Armangue T et al: Neurological complications in herpes simplex encephalitis: Clinical, immunological and genetic studies. Brain 146:4306, 2023.

Cellucci T et al: Clinical approach to the diagnosis of autoimmune encephalitis in the pediatric patient. Neurol Neuroimmunol Neuroinflamm 7:e663, 2020.

Dalmau J, Graus F: *Autoimmune Encephalitis and Related Disorders of the Nervous System*. Cambridge, Cambridge University Press, 2022.

Dalmau J, Graus F: Diagnostic criteria for autoimmune encephalitis: Utility and pitfalls for antibody-negative disease. Lancet Neurol 22:529, 2023.

Farina A et al: Neurological adverse events of immune checkpoint inhibitors and the development of paraneoplastic neurologic syndromes. Lancet Neurol 23:81, 2024.

Flanagan EP et al: Autoimmune myelopathies. Continuum 17:776, 2011.

Gaig C: Frequency and characterization of movement disorders in anti-IgLON5 disease. Neurology 97:e1367, 2021.

Graus F et al: Clinical approach to diagnosis of autoimmune encephalitis. Lancet Neurol 15:391, 2016.

Graus F: GAD antibodies in neurological disorders: Insights and challenges. Nat Rev Neurol 16:353, 2020.

Graus F et al: Updated diagnostic criteria for paraneoplastic neurological syndromes. Neurol Neuroimmunol Neuroinflamm 8:e1014, 2021.

Muñoz-Lopetegi A: Neurological, psychiatric, and sleep investigations after treatment of anti-leucine-rich glioma-inactivated protein 1 (LGI1) encephalitis in Spain: A prospective cohort study. Lancet Neurol 23:256, 2024.

100 Cancer Survivorship and the Long-Term Impact of Cancer and Its Treatment

Mark Roschewski, Dan L. Longo

A cancer diagnosis affects patients well beyond the initial treatment period with health implications that last throughout life. Long-term effects are toxicities that emerge during therapy and may persist after treatment is complete, while late effects are sequelae that may not emerge for months or years after treatment. Both long-term and late effects impact the health-related quality of life of cancer survivors and can affect the safe delivery of future therapy in the event of relapse. The risk of long-term effects varies across patients and depends on both patient-related and treatment-related factors. Older patients are more susceptible to acute treatment-related toxicities that emerge during therapy but can result in permanent disability. Adult survivors of childhood, adolescent, and young adult cancer face perpetual health consequences of cancer treatment related to premature physiologic aging and frailty and remain vulnerable to late effects, including premature death, for decades. (Table 100-1 lists potential long-term and late effects of cancer therapy by organ system). The emergence of novel targeted therapies and immunotherapy with unique mechanisms of action has introduced new considerations regarding the long-term impact of cancer therapy. Long-term risks are not immediately evident, and lifelong pharmacovigilance remains a critical cornerstone of caring for the cancer patient.

Cancer therapy has substantially improved, and the price of that success is that a greater number of cancer survivors are now at risk for the long-term impact of treatment. It is estimated that the United States currently has >18 million cancer survivors, representing approximately 5% of the total population. The prevalence is sharply increasing and is projected to reach 22 million by 2032 and 26 million by 2040. Almost 70% of patients have lived for at least 5 years since their diagnosis, and nearly half of survivors have lived longer than 10 years. Despite the magnitude of the growing problem, the core issues related to cancer survivorship remain understudied and continue to evolve with emerging therapies. Most studies are descriptive with fewer studies focused on the prevention and treatment of complications. Cancer survivorship remains an area that is ripe for further discovery, including improved risk-assessment tools, a deeper understanding of the pathophysiology and influence of genetics on host susceptibility, emerging biomarkers, and evidence-based clinical practice management guidelines.

TABLE 100-1 Toxicities by Organ System

ORGAN SYSTEM	LONG-TERM EFFECTS	LATE EFFECTS
Cardiovascular	Congestive heart failure Arrhythmias Myocarditis Pericardial disease Hypertension	Coronary artery disease Peripheral vascular disease Cardiac valvular disease
Pulmonary	Interstitial pneumonitis	Pulmonary fibrosis Radiation recall pneumonitis
Immunologic	Recurrent infections Autoimmune disease	Second malignancies Myelodysplasia Opportunistic infections
Endocrine	Gonadal dysfunction Premature ovarian failure Hypothyroidism Diabetes insipidus Hypopituitarism	Infertility Fractures Diabetes mellitus Osteoporosis
Neurologic	Peripheral neuropathy Autonomic dysfunction Hearing loss	Cognitive impairment
Gastrointestinal	Chronic diarrhea Malabsorption	Small-bowel obstruction Intestinal stricture
Genitourinary	Chronic renal failure Bladder dysfunction	Ureteral stricture Hemorrhagic cystitis Hydronephrosis Nephrotic syndrome
Psychological	Anxiety Depression Substance abuse disorders Cancer-related fatigue	Fear of cancer recurrence Posttraumatic stress disorder Sexual dysfunction Psychosocial dysfunction Financial hardship

Our primary understanding of the long-term impact of cancer treatment originated from survivors of childhood malignancies who are often cured after treatment with a combination of chemotherapy, radiation, and surgery. Modern 5-year survival rates for children and adolescents are >80%, with a goal of maintaining high cure rates while minimizing late effects. The cause-specific mortality following initial treatment of Hodgkin's lymphoma includes nearly 40% of deaths attributed to nonlymphoma causes. Indeed, more patients diagnosed with Hodgkin's lymphoma die from treatment-related late toxicity than from Hodgkin's lymphoma. More recently, improvements in the tolerability of cancer therapy and in supportive care have allowed a greater number of patients with comorbid conditions and advanced age to receive treatment, which has increased the rate of treatment-related morbidity. Approximately 67% of current cancer survivors are older than age 65, which carries a higher risk of treatment-related deaths from heart or lung complications, immunologic dysfunction, or adverse effects of drugs. Treatment paradigms for cancer continuously evolve, however, and newer treatments including targeted agents and immunotherapy have introduced unique long-term effects, particularly when these agents are administered indefinitely (Table 100-2 lists the potential long-term effects of specific cancer treatments). The complexities of cancer survivorship and the importance of longitudinal care of cancer patients are recognized as vital components to comprehensive cancer care. Still, the primary goal of cancer therapy remains long-term disease control, and the treating physician must maintain

TABLE 100-2 Toxicities by Type of Therapy

THERAPY TYPE	LONG-TERM EFFECT
Radiation	Second malignancies, coronary artery disease, pericardial disease, cardiac valvular disease, peripheral vascular disease, hypopituitarism, infertility, hypothyroidism, reduced bone density, intestinal or ureteral strictures
Chemotherapy and Hormonal Agents	
Anthracyclines (doxorubicin, daunorubicin)	Congestive heart failure, acute leukemia, cognitive impairment
Alkylating agents (cyclophosphamide, ifosfamide, melphalan)	Myelodysplasia, acute leukemia, impaired T-cell function, infertility, congestive heart failure, pulmonary fibrosis
Antimetabolites (5-flurouracil, fludarabine, gemcitabine)	Coronary artery disease, myelodysplasia, interstitial pneumonitis, cognitive impairment
Glycopeptide antibiotics (bleomycin)	Pulmonary fibrosis
Platinum agents (cisplatin, carboplatin, oxaliplatin)	Peripheral neuropathy, hearing loss, renal tubular dysfunction, interstitial pneumonitis, second malignancies
Vinca alkaloids (vincristine, vinblastine)	Peripheral neuropathy
Taxanes (paclitaxel, docetaxel)	Peripheral neuropathy, interstitial pneumonitis, cognitive impairment
Aromatase inhibitors (anastrozole, letrozole, fulvestrant)	Vasomotor symptoms, fractures
Antiandrogens (apalutamide, bicalutamide, darolutamide, enzalutamide, flutamide, nilutamide)	Sexual dysfunction
Immunotherapy Agents	
Immune checkpoint inhibitors (nivolumab, pembrolizumab, cemiplimab, avelumab, durvalumab, atezolizumab, dostarlimab, ipilimumab)	Autoimmune conditions, myocarditis, pericarditis, peripheral neuropathy
Bispecific T-cell engagers (mosunetuzumab, glofitamab, epcoritamab)	Recurrent infections
CAR-T therapy	Acute leukemia, myelodysplasia, T-cell lymphoma, peripheral neuropathy, recurrent infections
Targeted Agents	
Antibody-drug conjugates (brentuximab, polatuzumab, gemtuzumab)	Interstitial pneumonitis, peripheral neuropathy, sinusoidal obstruction syndrome
Immunomodulatory agents (thalidomide, lenalidomide, pomalidomide)	Second malignancies, prolonged neutropenia, peripheral neuropathy
Proteasome inhibitors (bortezomib, carfilzomib, ixazomib)	Peripheral neuropathy, congestive heart failure
HER2 inhibitors (trastuzumab, neratinib, capivasertib)	Congestive heart failure, interstitial pneumonitis
BTK inhibitors (ibrutinib, acalabrutinib, zanubrutinib)	Opportunistic infections, atrial and ventricular arrhythmias, hypertension
ALK inhibitors (alectinib, brigatinib, ceritinib, lorlatinib, crizotinib)	Atrial and ventricular arrhythmias, interstitial pneumonitis
CDK inhibitors (palbociclib, ribociclib, abemaciclib)	Hepatitis, colitis
Bcr-Abl inhibitors (imatinib, dasatinib, nilotinib, bosutinib)	Coronary artery disease, pleural effusions, pulmonary hypertension, peripheral vascular disease, nephrotic syndrome
FLT3 inhibitors (quizartinib)	Hypertension
Anti-VEGF agents (bevacizumab, pazopanib, sunitinib, sorafenib, axitinib)	Hypertension, myocardial ischemia, QTc prolongation, pulmonary hypertension, thrombosis
MEK inhibitors (selumetinib, trametinib, cobimetinib)	Retinopathy, hypertension, congestive heart failure
BRAF inhibitors (dabrafenib)	Hyperglycemia, secondary skin cancers, hypertension, congestive heart failure

proper perspective when considering these relative risks. The fear of long-term complications should not prevent the application of effective cancer treatment, particularly when delivered with curative intent. In a sense, managing long-term effects is a privilege only afforded those fortunate enough to overcome the initial threat to life represented by the cancer diagnosis.

CARDIOVASCULAR DYSFUNCTION

The long-term risks of cardiovascular disease after anthracycline chemotherapy and radiation that involves the mediastinum are well characterized and include arrhythmias, myocarditis, hypertension, congestive heart failure (CHF), and pericardial disease along with late effects of premature coronary artery disease, peripheral vascular disease, and cardiac valvular complications. It is estimated that one in eight childhood cancer survivors will experience a life-threatening cardiovascular event within 30 years after initial exposure, and cardiovascular disease is the most common cause of noncancer death in this population. Newer targeted agents and immunotherapy have introduced additional cardiovascular risks. A new discipline of cardio-oncology has emerged to address the increased risk of cardiovascular complications of cancer treatment in a multidisciplinary fashion and to better characterize individuals at high risk for treatment-related cardiac toxicities, develop surveillance strategies, and improve the management of long-term effects.

Radiation therapy, particularly mediastinal radiation that encompasses the heart, causes interstitial myocardial fibrosis, pericarditis, valvular disease, and premature coronary artery disease. Repeated or high (>6000 cGy) radiation doses are associated with greater risk, as is concurrent cardiotoxic chemotherapy exposure. Symptoms of acute pericarditis peak about 9 months after treatment and include dyspnea, chest pain, and fever. Chronic constrictive pericarditis and cardiac valvular diseases including aortic insufficiency and mitral regurgitation may develop 5–10 years following radiation therapy. Radiation is associated with premature coronary artery disease and peripheral vascular disease and a markedly increased risk of late effects such as myocardial infarction and thromboembolic stroke. Three-dimensional conformal techniques and newer particles including proton beams may more precisely spare normal tissue, but these are not widely available and the degree to which they will decrease long-term cardiovascular effects is unknown.

The myocardial toxicity of anthracyclines is dose-dependent and is associated with the pathognomonic finding of myofibrillar dropout on endomyocardial biopsy. Anthracycline cardiotoxicity occurs through a root mechanism of chemical free radical damage. Fe^{3+}-doxorubicin complexes damage DNA, nuclear and cytoplasmic membranes, and mitochondria. Approximately 5% of patients receiving >450–550 mg/m^2 total dose of doxorubicin will develop CHF, but it can also develop at substantially lower doses in some patients. Regimens that use continuous infusion of doxorubicin or liposomally encapsulated doxorubicin are associated with less cardiotoxicity. Genome-wide association studies have identified multiple genetic polymorphisms associated with a higher risk of cardiotoxicity, but the ability to risk-stratify is limited. Additional risk factors include age >60 years, smoking, hypertension, diabetes, chronic renal failure, and previous cardiovascular disease. Anthracycline-related CHF is often irreversible and carries a high rate of morbidity and mortality, making prevention crucial. Baseline screening with straining imaging echocardiography in patients who are at high risk of cardiac complications is recommended. A patient who develops symptoms suggestive of CHF should have therapy held and undergo additional testing. High-sensitivity troponins are emerging biomarkers that may predict cardiotoxicity, but further validation is needed. Prophylactic use of cardioprotective medications such as beta blockers, angiotensin-converting enzyme inhibitors, or statins is not yet routine.

Inhibitors of the human epidermal growth factor receptor 2 (HER2) (e.g., trastuzumab, lapatinib, neratinib, capivasertib) are also associated with CHF. These targeted agents are used in combination with chemotherapy as both adjuvant therapy and as treatment of metastatic breast cancer. In contrast to anthracyclines, cardiotoxicity is not dose related, is usually reversible, is not associated with pathologic changes on cardiac myofibrils, and has a different biochemical mechanism inhibiting intrinsic cardiac repair mechanisms. Monitoring for cardiac toxicity is typically performed every three or four doses using echocardiograms, and treatment is interrupted when ejection fractions significantly decline from baseline. Other potentially cardiotoxic chemotherapy agents include alkylating agents (cyclophosphamide, ifosfamide) at high doses and 5-flurouracil.

Tyrosine kinase inhibitors (TKIs) are novel classes of molecularly targeted anticancer agents that have become routinely applicable across a variety of malignancies. Although the overall tolerability of these drugs is often better than chemotherapy, they are frequently administered indefinitely and to older patients, which introduces new notions of cumulative long-term effects. These agents also carry risks of cardiovascular toxicities including CHF, atrial and ventricular arrhythmias, prolongation of the QT interval, and pulmonary and systemic hypertension. Other small-molecule inhibitors that have been associated with CHF include Bcr-Abl inhibitors (bosutinib, dasatinib, ponatinib, nilotinib), anti–vascular endothelial growth factor (VEGF) agents (pazopanib, sunitinib, sorafenib, axitinib, lenvatinib, nintedanib), MEK inhibitors (trametinib), and proteasome inhibitors (carfilzomib). Systemic hypertension is commonly associated with anti-VEGF agents, ponatinib, and trametinib, whereas dasatinib has a well-documented association with pulmonary hypertension. Bruton tyrosine kinase (BTK) inhibitors, including ibrutinib, acalabrutinib, and zanubrutinib, have all been associated with atrial fibrillation as well as ventricular arrythmias. As more of these small-molecule inhibitors become widely used, additional monitoring for long-term and late effects should be incorporated into the routine surveillance of oncologists, cardiologists, cardio-oncologists, and primary care providers.

Immunotherapy agents have emerged as anticancer therapies used in a variety of cancers. Immune checkpoint inhibitors (ICI) have been associated with a number of important cardiovascular toxicities including pericardial disease, vasculitis, and fulminant myocarditis. The mechanism of these toxicities is T cell mediated, and the toxicities often respond to rapid initiation of high-dose glucocorticoids, but they can be severe or fatal if not recognized promptly. Patients with definite myocarditis on endomyocardial biopsy may have normal findings on cardiac magnetic resonance imaging (MRI) and only a rise in troponin levels, highlighting the need for a high index of suspicion for this potential complication. Patients who develop ICI-mediated myocarditis should have the drug halted immediately, and it should not be restarted after recovery. Combinations of multiple ICIs increase the risk of these immune-related toxicities, and no clear pattern of which patients are most susceptible has emerged.

Chimeric antigen receptor T-cell (CAR-T) therapies are associated with cytokine release syndromes (CRS), which can be severe and associated with arrhythmias or CHF. Multiple cytokines including interleukin (IL) 1, IL-2, tumor necrosis factor (TNF) α, and IL-6 activate prostaglandins, and levels of IL-6 are associated with more severe CRS, possibly due to the recruitment of additional T cells. Inhibitors of IL-6 such as tocilizumab are commonly used to treat CRS and prevent complications.

PULMONARY DYSFUNCTION

Radiation-induced lung injury presents in early phases as acute pneumonitis at 4 weeks following treatment, but it can evolve into pulmonary fibrosis. Risk factors include advanced age, smoking, poor performance status, preexisting compromised pulmonary function, and radiation volume and dose. It occurs most commonly in patients with lung cancer, mediastinal lymphoma, and breast cancer, and the incidence is decreasing due to advances in radiation delivery techniques. Fever, cough, pleuritic chest pain, and dyspnea on exertion are characteristic symptoms, and the severity may be out of proportion to the lung volume irradiated. Fine, high-pitched "Velcro rales" may be an accompanying physical finding. The diffusion capacity of the lungs for carbon dioxide (DL_{CO}) is the most sensitive measure of pulmonary functional impairment, and ground-glass infiltrates often correspond with relatively sharp edges to the irradiated volume, although the pneumonitis may progress beyond the field. The mechanism of lung injury is a direct effect of radiation that leads to increased capillary permeability and pulmonary edema. Damage to type I and II pneumocytes leads to surfactant loss and the transudation of serum proteins into the alveoli. Cytokines including TNF-α are released from the damaged lung cells and attract inflammatory cells to the alveoli and interstitial space. The late phases of injury are caused by reactive oxygen species that stimulate collagen production and may lead to fibrosis. Transforming growth factor β (TGF-β) is particularly important in stimulating collagen synthesis and may represent a therapeutic target.

Up to 10% of patients will develop chemotherapy-induced pulmonary toxicity that can manifest as both long-term and delayed effects in cancer survivors. Bleomycin generates activated free radical oxygen species and causes pneumonitis associated with a radiographic or interstitial ground-glass appearance diffusely throughout both lungs, often worse in the lower lobes. A nonproductive cough may be an early sign. This toxicity is dose-related and dose-limiting. The DL_{CO} is a sensitive measure of toxicity and recovery, and a baseline value is generally obtained before administering bleomycin therapy. Doses are stopped if the baseline DL_{CO} falls 25% or more. Additive or synergistic risk factors include age, prior lung disease, and concomitant use of other chemotherapy, lung irradiation, and high concentrations of inspired oxygen. Other commonly used chemotherapeutic agents notable for pulmonary toxicity include gemcitabine, irinotecan, methotrexate, fludarabine, and the antibody-drug conjugates brentuximab and fam-trastuzumab deruxtecan, which can cause interstitial lung disease. Most of these toxicities are idiosyncratic and not dose-related. The pathogenesis for pulmonary toxicity is related to direct injury to pneumocytes, oxidative stress, and the release of cytokines that recruit additional inflammatory cells. High-dose alkylating agents, cyclophosphamide, ifosfamide, and melphalan are frequently used in the hematopoietic stem cell transplant setting, which may result in severe pulmonary fibrosis and/or pulmonary venoocclusive disease.

Radiation-induced lung injury and chemotherapy-induced pneumonitis are generally glucocorticoid responsive. Prednisone 1 mg/kg is often used to control acute symptoms and prevent pulmonary dysfunction with a slow taper over months. Antibiotics, bronchodilators, oxygen, and diuretics may all play an important role in management of pneumonitis, and consultation with a pulmonologist should be undertaken. Relapses can occur after an initial response to glucocorticoids, and refractory cases may respond to alternative immunosuppressive agents such as azathioprine or cyclosporine. No effective therapy exists for pulmonary fibrosis, and the treatment is primarily supportive with supplemental oxygen.

Pulmonary toxicity resulting from targeted anticancer agents including small-molecule inhibitors and immunotherapy agents can be potentially life-threatening and often leads to drug discontinuation. Noninfectious pneumonitis is associated with cough, dyspnea, and infiltrates on chest imaging and has been reported to be associated with anti-VEGF agents (sunitinib, sorafenib), epidermal growth factor receptor (EGFR) inhibitors (cetuximab, afatinib), and the ALK inhibitor crizotinib. The antibody-drug conjugate brentuximab vedotin may also cause severe pulmonary toxicity. The onset of drug-induced pneumonitis can be rapid, and prompt use of glucocorticoids is important once infectious causes are excluded. Severe pneumonitis is a reason to discontinue the offending drug permanently.

IMMUNE SYSTEM DYSFUNCTION

A significant risk of most anticancer treatment is cumulative effects on the host immune system leading to a higher risk of second malignancies and impaired long-term immune health. Second malignancies in cancer survivors are a major cause of death, and up to 20% of all cancer diagnoses now occur in cancer survivors. The risk of second malignancies is related to genetic susceptibility, environmental exposures, and specific cancer treatments. Often, the risk factors that led to the primary cancer remain, and a risk of a second malignancy persists. Genetic cancer syndromes (e.g., multiple endocrine neoplasia or Li-Fraumeni, Lynch's, Cowden's, and Gardner's syndromes) are examples of genetically based second malignancies of specific types, and specific cancer treatments do not increase the risk of secondary malignancies. Genetic disorders that result in DNA repair deficiencies including ataxia-telangiectasia, Bloom's syndrome, and Fanconi's anemia greatly increase the lifetime risk of cancers as well as the risks associated with DNA-damaging agents. Importantly, the risk of treatment-related second malignancies is greatly increased when radiation is combined with chemotherapy, and for such combined-therapy treatment approaches, it is important to establish the necessity of each component in the treatment program.

Patients receiving radiation have a lifelong risk of second malignancies that is 1–2% in the second decade following treatment but increases to >25% after 25 years and does not appear to plateau. The risk of second malignancies from radiation is dose-dependent and often occurs within the treatment field. Common radiation-related solid tumors include central nervous system (CNS), breast, bladder, colon, stomach, liver, lung, and thyroid cancers and sarcomas, which are often aggressive and have a poor prognosis. The age of the patient when they receive radiation affects their risk, and younger patients are more vulnerable to second malignancies. An example of an organ-, age-, and sex-dependent radiation-induced secondary malignancy is breast cancer, in which the risk is small with radiation in women aged older than 30 years but increases about 20-fold over baseline in women aged younger than 30 years. Females with Hodgkin's lymphoma who receive radiation have a cumulative risk of 25% for developing breast cancer at 30 years after treatment.

Chemotherapy is significantly associated with two second malignancies: acute leukemia and myelodysplastic syndromes. Two types of secondary leukemia have been described; in patients treated with alkylating agents, acute myeloid leukemia is associated with deletions in chromosome 5 or 7 and complex karyotypes and often is preceded by myelodysplasia. The lifetime risk is about 1–5%, is increased by radiation therapy, and increases with age. The incidence of these leukemias peaks at 5–8 years after therapy, with risk returning close to baseline at 10 years. The other type of acute myeloid leukemia is related to therapy with topoisomerase inhibitors, is associated with chromosome 11q23 translocations, has an incidence of <1%, generally occurs 2–3 years after treatment, and is rarely preceded by myelodysplasia. Both of these acute myeloid leukemias have a high mortality. A fraction of the population has underlying clonal hematopoiesis not related to prior cancer treatment, and the percentage increases with age. In such patients, the hematopoietic stem cells carry mutations that are associated with myeloid malignancy despite normal blood counts. The presence of these genetic lesions may predispose patients to develop myeloid malignancies, but currently, these are not assessed prior to starting therapy.

Other cytotoxic agents are associated with long-term alterations in immunity beyond the initial treatment period. Bendamustine has significant effects on both B-cell and T-cell subpopulations that can persist for years, and long-term studies of lymphoma patients treated with bendamustine-based regimens show higher rates of death from second malignancies compared to other chemotherapy regimens. Purine analogues like fludarabine also produce long-term T-cell suppression. The risk of solid tumors is also increased after chemotherapy, and alkylating agents such as cyclophosphamide increase the risk of sarcoma, gastrointestinal cancers, and breast cancer in a dose-dependent manner. Other chemotherapy agents, including platinum agents, which are commonly used to treat germ cell tumors, have been associated with second malignancies. Treatment of multiple myeloma with the immunomodulatory agent lenalidomide is associated with a significantly increased risk of second hematologic malignancies including lymphomas and leukemias.

Given the indefinite risk of second malignancies in cancer survivors, patients need lifelong surveillance. Guidelines for breast cancer surveillance in survivors exposed to chest radiation recommend that patients be screened with mammograms and/or breast MRI beginning at age 25 years or 8 years after treatment. Any organ in the treatment field is susceptible to developing a cancer (e.g., radiation to the chest may increase the risk of gastric or esophageal cancer). Patients exposed to abdominal or pelvic radiation should have annual colonoscopies starting at age 35 years or 10 years after exposure. For patients treated with neck radiation, fine-needle aspiration should be performed on any palpable thyroid nodules, and thyroid-stimulating hormone (TSH) levels should be monitored periodically.

Combination chemotherapy is also associated with impaired immune health and increases the long-term risk of recurrent sinopulmonary infections, autoimmune complications, and late-onset opportunistic infections. Survivors of lymphoma have elevated risks of developing autoimmune hemolytic anemia, viral or fungal pneumonias, meningitis, or other infections, and these risks remain decades after treatment. Agents or treatment regimens that result in significant T-cell depletion increase the risk of Epstein-Barr virus–associated B-cell lymphoproliferative disorders. Discontinuing immunosuppressive therapy, if possible, is often associated with complete disease regression. Anti-CD20 monoclonal antibodies, CAR-T therapy, bispecific T-cell engagers (monoclonal antibodies targeting both B cells and T cells), and ICIs have all been associated with long-term B-cell aplasia that often requires intravenous immunoglobulin replacement and persistent vigilance for recurrent sinopulmonary infections. The genetically modified CAR-T cells may also carry a small risk of insertional mutagenesis as T-cell lymphomas have been rarely reported. Rituximab and immunomodulatory agents have both been associated with late-onset neutropenia that can occur months after exposure to the drug and often requires growth factor support. Given these risks of impaired immune system function, all cancer survivors should undergo annual influenza and other seasonal vaccinations and should be considered for pneumococcal vaccination depending on age and immune health status.

REPRODUCTIVE AND ENDOCRINE DYSFUNCTION

Endocrine complications are prevalent in childhood cancer survivors. Nearly half of all survivors will have at least one hormonal disorder in their lifetime, and these most commonly present as late effects. Radiation to the head, neck, or pelvis is associated with the greatest risk of endocrine dysfunction. Testicles and ovaries in prepubertal patients are sensitive to radiation damage in a dose-related fashion; spermatogenesis is affected by low doses of radiation, and complete azoospermia occurs at 600–700 cGy. Leydig cell dysfunction, in contrast, occurs at <2000 cGy, and hence, endocrine function is lost at much higher radiation doses than spermatogenesis. Erectile dysfunction occurs in up to 80% of men treated with external-beam radiation therapy for prostate cancer. Ovarian function damage with radiation is age-related and occurs at doses of 150–500 cGy. Hormone replacement therapy is often contraindicated (as in estrogen receptor–positive breast cancer). Attention must be paid to maintenance of bone mass with calcium and vitamin D supplements and oral bisphosphonates, and bone mass should be monitored using bone density determinations. Long-term survivors of childhood cancer who have received cranial radiation may have altered leptin biology and growth hormone deficiency, leading to obesity and reduced strength, exercise tolerance, and bone density. Radiation therapy to the neck may lead to hypothyroidism, Graves' disease, and thyroiditis. Cranial radiation may also be associated with an array of endocrine abnormalities with disruption of normal pituitary-hypothalamic axis function, and a high index of suspicion needs to be maintained to identify and treat this toxicity. Patients who have received abdominal radiation should receive annual screens for obesity and diabetes mellitus with height and weight measurements along with a hemoglobin A_{1C} at least every 2 years.

Alkylating agents are associated with the highest rates of male and female infertility, which is directly dependent on dose and duration of treatment. The age at treatment is an important determinant of fertility outcome, with prepubertal patients having the highest tolerance. Chemotherapy reduces the number of primordial follicles in women, and the risk of infertility is age related. Females who resume menses after treatment are still at increased risk for premature ovarian failure. Males generally have reversible reductions in sperm counts and motility during alkylator chemotherapy, and long-term infertility is associated with doses of cyclophosphamide >9 g/m^2 and with high-intensity therapy, such as that used in hematopoietic stem cell transplantation. All patients should be counseled on the potential impact on fertility, and sperm, oocyte, and embryo cryopreservation are now routinely performed prior to therapy in patients of child-bearing age. Some evidence suggests that gonadotropin-releasing hormone agonists given concurrently with chemotherapy may improve the resumption of menses, but long-term data on fertility preservation are lacking and this should not replace oocyte harvesting. In vitro fertilization can be helpful to couples with chemotherapy-induced infertility.

Combination chemotherapy and aromatase inhibitors can impair bone health, and older patients may be more susceptible to these effects. Due to the combined risk of age-related osteoporosis and the effect of therapy, the risk of fractures in patients over the age of 70 years may be as high as 5–10% within a few years of finishing therapy. The risk of low bone mineral density is highest in certain high-risk groups including childhood cancer survivors, breast cancer patients, and those who have undergone hematopoietic stem cell transplant.

ICIs that target CTLA-4 and PD-1 have led to serious chronic toxicities including the breaking of self-tolerance and the autoimmune destruction of certain endocrine organs, particularly the thyroid and adenohypophysis (anterior pituitary). Hypophysitis is more commonly reported in association with CTLA-4 inhibitors, whereas thyroid dysfunction is more common with PD-1 inhibitors. Most immune-related toxicities occur within 8–12 weeks of starting treatment, but they can occur at any time during treatment or even after therapy has stopped. Patients with autoimmune thyroiditis or hypophysitis require lifelong hormone replacement, and early recognition is important.

TKIs have been associated with various endocrine-related toxicities, with hypothyroidism being the most common. Virtually all glands can be affected by TKIs, however, and adrenal insufficiency, growth retardation, increased bone turnover, and hypogonadism have all been reported. Growth deceleration is problematic in children, particularly when treatment is initiated before puberty. The mechanism is postulated to be related to disruptions in growth hormone signaling or inhibition of the insulin-like growth factor 1 (IGFR-1) receptor. Endocrine effects that have been reported to be associated with other small-molecule inhibitors include alterations in bone remodeling, reduced calcium and vitamin D levels, thyroid dysfunction, gonadal dysfunction, adrenal dysfunction, and altered glucose metabolism. Thyroid function tests should be monitored periodically while patients are on these targeted agents, and replacement hormones and/or vitamins should be prescribed as necessary. Careful consideration should be made to keep patients on therapy with limited dose reduction or interruption while managing the symptoms of glandular dysfunction.

NEUROLOGIC DYSFUNCTION

Delayed radiation CNS toxicity occurs weeks to months following therapy; is associated with nausea, drowsiness, hypersomnia, and ataxia; and is pathologically associated with reversible demyelination. Chronic, late radiation injury occurs 9 months to up to 10 years following therapy, and dysfunction increases over time. Diffuse radiation injury is associated with global CNS neurologic dysfunction and diffuse white matter changes on computed tomography (CT) or MRI. Pathologically, small vessel changes are prominent and focal necrosis is common. Prophylactic cranial irradiation in both childhood and adult leukemias/lymphomas has largely been abandoned due to the acute and long-term effects of therapy. Glucocorticoids may be symptomatically useful for acute toxicities but do not alter the course. Psychostimulants such as methylphenidate may improve attention and executive functioning in childhood survivors.

In children and adolescent cancers, younger age, higher cranial irradiation dose, larger brain volumes irradiated, and longer treatment times are associated with worse neurocognitive outcomes. In adult cancers, patients over the age of 60 who receive whole-brain radiation therapy are at high risk for neurocognitive impairment after therapy. Genetic polymorphisms may be associated with an increased risk of neurocognitive problems, and emerging evidence suggests polymorphisms in the folate pathway, oxidative stress genes, and enzymes that regulate both catecholamines and deamination of amines are associated with individual risk.

The most common chemotherapy-related neurologic toxicities that manifest in cancer survivors are peripheral neuropathy and cognitive impairment. Vinca alkaloids produce a characteristic "stocking-glove" neuropathy with pain, numbness, and tingling and increased sensitivity to temperature advancing to loss of motor function, which is highly dose related. Distal sensorimotor polyneuropathy prominently involves loss of deep tendon reflexes with initially loss of pain and temperature sensation, followed by proprioceptive and vibratory loss. The mechanism is axonal degeneration leading to myelin loss with some contribution of alterations in neurotransmitter signaling. Risk factors for peripheral neuropathy include preexisting neuropathies, diabetes mellitus, and older age. Genetic predictors and biomarkers to identify patients at the highest risk are the subject of ongoing research. The most effective management is prevention of acute toxicities resulting in permanent damage. This requires careful patient history and physical examination by experienced oncologists to decide when the drug must be stopped or reduced to prevent permanent damage, which can have a profound effect on the ability to deliver further therapy. Milder toxicity often slowly completely resolves after treatment discontinuation. Cisplatin is associated with sensorimotor neuropathy and hearing loss, especially at doses >400 mg/m^2, requiring audiometry in patients with preexisting hearing compromise. Carboplatin is often substituted in such cases given its lesser effect on hearing.

Chemotherapy-related cognitive impairment, or "chemo brain," occurs in nearly 35% of all cancer survivors and has a profound impact on functional ability and the health-related quality of life after therapy. It is clinically associated with impaired memory, learning, attention, executive functioning, and speed of information processing. The mechanisms are postulated to involve alterations in neurodegeneration, immune dysregulation, and cytokine production. No effective therapy has been identified, although regular exercise and cognitive rehabilitation are associated with improved symptoms. Most symptoms improve within a year of therapy, but symptoms can persist in 10–20% of patients for extended periods of time.

Antibody-drug conjugates (ADCs) are novel therapies in which a monoclonal antibody targeted to a tumor antigen is attached to a potent anticancer agent via a chemical linker. Often, these agents are associated with significant central and peripheral neurotoxicity, as has been described with agents such as brentuximab vedotin and pertuzumab.

These neuropathies often emerge during treatment similar to those seen with chemotherapy and are dose dependent. Early recognition of treatment-emergent neuropathy induced by ADCs mandates dose interruption or modification to less frequent dosing schedules in order for patients to remain on therapy. ICIs have been associated with autoimmune complications and unique neurologic manifestations such as optic neuritis that may be reversible after glucocorticoids.

Newer molecularly targeted agents and immunotherapy have also been associated with neurologic dysfunction and may exacerbate persistent neuropathy from prior therapy. Proteasome inhibitors are associated with neuropathic pain and motor neuropathy that can occur immediately or be delayed in onset. The proposed mechanism is through enhanced oxidative stress on neural cells. Subcutaneous administration of these agents is associated with less peripheral neuropathy than intravenous infusions. Immunotherapy agents such as CAR-T therapies and bispecific monoclonal antibodies targeting both T cells and B cells are associated with significant acute neurotoxicity known as immune effector cell–associated neurotoxicity syndrome (ICANS). Symptoms of ICANS include confusion, encephalopathy, seizures, and cerebellar symptoms. These symptoms are hypothesized to be related to the CRS associated with these therapies and often are time limited. However, a minority of patients treated with these agents will have persistent central and neurologic dysfunction for which management is mainly supportive. Management guidelines to prevent and treat ICANS have been developed and include rapid delivery of glucocorticoids during the acute onset of symptoms along with supportive care. Notably, treatment with IL-6 receptor antagonists such as tocilizumab, which are effective for decreasing the severity of CRS, is largely ineffective at preventing neurologic toxicities. Progressive multifocal leukoencephalopathy (PML) is a rare but serious brain infection that is caused by the JC virus and has been reported as a rare complication of treatment with rituximab, ibrutinib, and PI3K inhibitors. The inflammatory response to the virus presents as unifocal or multifocal hyperintense lesions involving the subcortical white matter that are best seen on T2/fluid-attenuated inversion recovery (FLAIR) images on MRI. Treatment is supportive and includes removal of the offending agent.

HEPATIC AND GASTROINTESTINAL DYSFUNCTION

Long-term hepatic damage from standard chemotherapy regimens is rare. High-dose chemotherapy alone or with radiation therapy in preparative regimens for bone marrow transplantation may result in venoocclusive disease of the liver. This potentially lethal complication classically presents with anicteric ascites, elevated alkaline phosphatase, and hepatosplenomegaly. Pathologically, there is venous congestion, epithelial cell proliferation, and hepatocyte atrophy progressing to frank fibrosis. Frequent monitoring of liver function tests during any chemotherapy is necessary to avoid both idiosyncratic and expected toxicities. Hepatic radiation damage depends on dose, volume, fractionation, preexisting liver disease, and synchronous or metachronous chemotherapy. In general, radiation doses to the liver >1500 cGy can produce hepatic dysfunction with a steep dose-injury curve. Radiation-induced liver disease closely mimics hepatic venoocclusive disease.

Novel targeted agents including immunotherapy agents have introduced a number of hepatic and gastrointestinal toxicities that occur during the course of treatment and persist as long-term toxicities including hepatitis, colitis, malabsorption, and chronic diarrhea. Early signs of serious liver injury should lead to discontinuation of the offending agent as dose reductions do not reliably reduce further liver injury. The mechanisms for colitis or hepatitis associated with these targeted agents are not completely understood but are hypothesized to be T cell mediated, and prompt initiation of glucocorticoids is often effective. Diarrhea with or without severe colitis can be associated with virtually all targeted agents through a variety of mechanisms including BCR-ABL inhibitors, BTK inhibitors, EGFR inhibitors, MEK inhibitors, CDK inhibitors, and immunomodulatory agents. Even if the diarrhea is not severe, the impact associated with indefinite treatment greatly interferes with the quality of life and often leads to discontinuation of targeted therapy. Immunomodulatory agents including lenalidomide are associated with late onset of diarrhea that is caused by bile acid malabsorption and often responds to bile acid sequestrants. ICIs are associated with colitis and hepatitis that may be responsive to prompt initiation of glucocorticoids that may require a long taper over many months.

RENAL AND BLADDER DYSFUNCTION

Cisplatin produces reversible decrements in renal function but may also produce severe irreversible toxicity in the presence of preexisting renal disease and may predispose to additional damage with subsequent therapy. Cyclophosphamide and ifosfamide are prodrugs primarily activated in the liver with cleavage products (acrolein) that can produce hemorrhagic cystitis. This can be prevented with the free radical scavenger MESNA (mercaptoethane sulfonate), which is required for ifosfamide administration. Hemorrhagic cystitis is most commonly an acute toxicity but may become a chronic and recurrent problem as long as 20 years after therapy.

A number of targeted agents, including anti-VEGF agents, FLT3 inhibitors, and BTK inhibitors, are associated with systemic hypertension, which can lead to late effects or a progressive decline in renal function. Pharmacovigilance studies have recently identified a link between the chronic use of BCR-ABL inhibitors and the late effects of chronic renal failure and the nephrotic syndrome. Renal dysfunction following ICIs is uncommon, but acute interstitial nephritis can occur. Similar to other immune-related toxicities, this acute toxicity requires prompt use of glucocorticoids to avoid long-term effects on renal function.

PSYCHOLOGICAL DYSFUNCTION AND SOCIOECONOMIC IMPACT OF SURVIVORSHIP

The diagnosis and treatment of cancer can introduce psychological effects that continue throughout life. Cancer survivors are at increased risk for anxiety, depression, attention problems, and posttraumatic stress syndromes. Many cancer patients experience intrusive or debilitating concerns about cancer recurrence following therapy. In addition, these patients may experience socioeconomic stressors that affect employment, insurance, and relationships and lead to financial and/or sexual difficulties. Survivors of childhood cancers are less likely to graduate from college or gain full-time employment than their peers and are more likely to engage in risky health behaviors such as substance abuse and excessive alcohol use. The long-term psychosocial effects of treatment are greatest in patients who undergo CNS-directed therapies and intensive combination chemotherapy regimens. One of the most challenging long-term effects of cancer therapy is cancer-related fatigue (CRF) that occurs in up to 60% of patients during treatment and may persist as physical, cognitive, or emotional exhaustion for years after therapy. The precise etiology of CRF is unclear, but it has a strong association with impaired sleep and lower physical activity. Patients with CRF have higher levels of proinflammatory cytokines such as IL-1, IL-6, and C-reactive protein, which may be therapeutic targets. Oncologists should ask about physical, emotional, and cognitive issues explicitly with patients and provide appropriate counseling or support systems. The overall risk of suicidal ideation and suicide is low but is greater in cancer patients and survivors than age-matched controls. Tailored cognitive-behavioral therapy may improve the anxiety and posttraumatic stress associated with cancer survivorship. Psychostimulants including methylphenidate and modafinil are modestly effective at treating the symptoms associated with CRF, but further studies are needed.

CANCER SURVIVORSHIP CARE PLANS

Survivorship starts at the time of diagnosis and continues indefinitely. Prevention of late effects may be improved with the implementation of risk-stratification tools prior to therapy. After therapy is initiated, focused surveillance plans for late effects are critical for early detection and implementation of interventions but also must include risk stratification to avoid unnecessary surveillance testing that leads to overdiagnosis and/or psychological distress. Survivorship care has traditionally been performed by oncologists, but the magnitude of the problem mandates that primary care providers and preventive medicine specialists be trained in the follow-up of treated cancer patients in remission or undergoing chronic therapy. All former cancer patients should undergo surveillance for recurrence and second malignancies and be monitored for long-term effects of treatment; however, nearly all recurrences are detected because of symptoms. Health promotion and disease prevention with age- and sex-specific routine screening tests (e.g., colonoscopy, Pap smears, mammography, human papillomavirus vaccination, dual-energy x-ray absorptiometry scans) should be a focus of survivorship care along with psychosocial well-being. Annual mammography should start no later than 10 years after breast radiation. Patients receiving radiation fields encompassing thyroid tissue should have regular thyroid examinations and TSH testing. Localized pain or palpable abnormality in a previously radiated field should prompt radiographic evaluation. Patients treated with alkylating agents or topoisomerase inhibitors should have a complete blood count every 6–12 months, and cytopenias, abnormal cells on peripheral smear, or macrocytosis should be evaluated with bone marrow biopsy and aspirate and include cytogenetics, flow cytometry, or fluorescence in situ hybridization (FISH) studies as appropriate.

As the population of cancer survivors increases and patients live longer, cancer survivorship has become increasingly important, and the Institute of Medicine and National Research Council have published a monograph entitled *From Cancer Patient to Cancer Survivor: Lost in Transition*. The monograph proposes a plan that would inform clinicians caring for cancer survivors of the complete details of patients' previous treatments, complications thereof, signs and symptoms of late effects, and recommended screening and follow-up procedures.

OUTLOOK

Survivorship care is a burgeoning problem facing oncologists today. The challenge is to develop cancer treatments that maximize clinical benefit including cure of disease while also mitigating the risks of long-term toxicity. As cancer treatments continue to improve, the prevalence of cancer survivors increases along with an increase in life expectancy. Further, since emerging therapies often have improved tolerability profiles, a greater number of patients with advanced age or comorbid medical conditions will become cancer survivors with persistent treatment-related toxicities. As treatment paradigms continue to evolve, the nature and biologic basis for toxicities will change and pharmacovigilance of new therapies is critical. Advances in genomic medicine may allow for more risk-stratified personalized care. The choice of therapy needs to be tailored to the type of cancer, expected outcomes, and patient-related risk factors for both acute and long-term toxicities. After therapy is complete, longitudinal monitoring of the health and health-related quality of life of cancer survivors is critical since the incidence of late effects of treatment does not appear to plateau over time.

Acknowledgment

We would like to acknowledge the contribution of Carl E. Freter who coauthored a previous version of this chapter; material from his chapter was retained in this version.

■ FURTHER READING

Armenian SH et al: Cardiovascular disease in survivors of childhood cancer: Insights into epidemiology, pathophysiology, and prevention. J Clin Oncol 36:2135, 2018.

Brinkman TM et al: Psychological symptoms, social outcomes, socioeconomic attainment, and health behaviors among survivors of childhood cancer: Current state of the literature. J Clin Oncol 36:2190, 2018.

Chow EJ et al: New agents, emerging late effects, and the development of precision survivorship. J Clin Oncol 36:2231, 2018.

Ehrhardt MJ et al: Health care transitions among adolescents and young adults with cancer. J Clin Oncol 42:743, 2024.

Lustberg MB et al: Mitigating long-term and delayed adverse events associated with cancer treatment: Implications for survivorship. Nat Rev Clin Oncol 20:527, 2023.

Rowland JH et al: Survivorship science at the NIH: Lessons learned from grants funded in fiscal year 2016. J Natl Cancer Inst 111:109, 2019.

Shapiro CL: Cancer survivorship. N Engl J Med 379:2438, 2018.

Shapiro CL et al: ReCAP: ASCO core curriculum for cancer survivorship education. J Oncol Pract 12:e08, 2016.

Shree T et al: Impaired immune health in survivors of diffuse large B-cell lymphoma. J Clin Oncol 38:1664, 2020.

Turcotte LM et al: Risk, risk factors, and surveillance of subsequent malignant neoplasms in survivors of childhood cancer: A review. J Clin Oncol 36:2145, 2018.

Section 2 Hematopoietic Disorders

101

Hematopoietic Stem Cells

David T. Scadden, Dan L. Longo

All of the cell types in the blood and some cells in every tissue of the body are derived from hematopoietic (*hemo*: blood; *poiesis*: creation) stem cells. If the hematopoietic stem cell is damaged and can no longer function (e.g., due to a nuclear accident), a person would survive 2–4 weeks in the absence of extraordinary support measures. With the clinical use of hematopoietic stem cells, tens of thousands of lives are saved each year **(Chap. 119)**. Stem cells produce hundreds of billions of blood cells daily from a stem cell pool that is estimated to be only 20,000–200,000. How stem cells do this, how they persist for many decades despite the production demands, and how they may be better used in clinical care are important issues in medicine.

The study of blood cell production has become a paradigm for how other tissues may be organized and regulated. Basic research in hematopoiesis includes defining stepwise molecular changes accompanying functional changes in maturing cells, aggregating cells into functional subgroups, and demonstrating hematopoietic stem cell regulation by a specialized microenvironment; these concepts are worked out in hematology and offer models for other tissues. Moreover, these concepts may not be restricted to normal tissue function but extend to malignancy.

CARDINAL FUNCTIONS OF HEMATOPOIETIC STEM CELLS

All stem cell types have two cardinal functions: self-renewal and differentiation **(Fig. 101-1)**. Stem cells exist to generate, maintain, and repair tissues. They function successfully if they can replace a wide variety of shorter-lived mature cells over prolonged periods. The process of self-renewal (see below) assures that a stem cell population can be sustained over time. Without self-renewal, the stem cell pool would become exhausted and tissue maintenance would not be possible. The process of differentiation leads to production of the effectors of tissue function: mature cells. Without proper differentiation, the integrity of tissue function would be compromised and organ failure or neoplasia would ensue.

In the blood, mature cells have variable average life spans, ranging from hours for mature neutrophils to a few months for red blood cells to many years for memory lymphocytes. However, the stem cell pool is the central, durable source of all blood and immune cells, maintaining a capacity to produce a broad range of cells from a single cell source, yet keeping itself vigorous over decades of life. As an individual stem cell divides, it has the capacity to accomplish one of three division outcomes: two stem cells, two cells destined for differentiation, or one stem cell and one differentiating cell. The former two outcomes are the

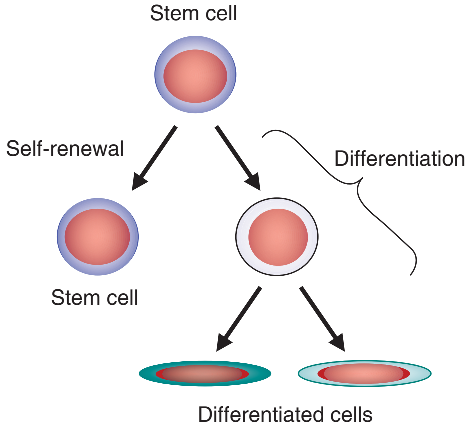

FIGURE 101-1 Signature characteristics of the stem cell. Stem cells have two essential features: the capacity to differentiate into a variety of mature cell types and the capacity for self-renewal. Intrinsic factors associated with self-renewal include expression of *Bmi-1, Gfi-1, PTEN, STAT5, Tel/Atv6, p21, p18, MCL-1, Mel-18, RAE28,* and *HoxB4*. Extrinsic signals for self-renewal include Notch, Wnt, SHH, angiogenin, and Tie2/Ang-1. Based mainly on murine studies, hematopoietic stem cells express the following cell surface molecules: CD34, Thy-1 (CD90), c-Kit receptor (CD117), CD133, CD164, and c-Mpl (CD110, also known as the thrombopoietin receptor).

result of symmetric cell division, whereas the latter indicates a different outcome for the two daughter cells—an event termed asymmetric cell division. The relative balance for these types of outcomes may change during development and under particular kinds of demands on the stem cell pool.

DEVELOPMENTAL BIOLOGY OF HEMATOPOIETIC STEM CELLS

During development, blood cells are produced at different sites. Initially, the yolk sac provides oxygen-carrying red blood cells and many of the macrophage-like cells that are resident in tissues: cells like microglia in the brain. The placenta and several sites of intraembryonic blood cell production then become involved in sequential order. These move from the genital ridge at a site where the aorta, gonadal tissue, and mesonephros are emerging to the fetal liver and then, in the second trimester, to the bone marrow and spleen. As the location of stem cells changes, the cells they produce also change. The yolk sac provides red cells expressing embryonic hemoglobins and tissue-resident macrophages. Intraembryonic sites of hematopoiesis generate stem cells, red cells, platelets, and the circulating cells of innate immunity. The production of the cells of adaptive immunity occurs then as well but becomes robust as the thymus forms and the bone marrow is colonized in the second trimester. Stem cell proliferation remains high, even in the bone marrow, until shortly after birth, when it appears to dramatically decline. The cells in the bone marrow are thought to arrive by the bloodborne transit of cells from the fetal liver after calcification of the long bones has begun. The presence of stem cells in the circulation is not unique to a time window in development, however, as hematopoietic stem cells circulate throughout life. The time that stem cells spend freely circulating appears to be brief (measured in minutes in the mouse), but the stem cells that do circulate are functional and can be used for transplantation. The number of stem cells that circulate can be increased in a number of ways to facilitate their harvest and transfer to the same or a different host.

MOBILITY OF HEMATOPOIETIC STEM CELLS

Cells entering and exiting the bone marrow do so through a series of molecular interactions. Circulating stem cells (through CD162 and CD44) engage the lectins (carbohydrate binding proteins) P- and E-selectin on the endothelial surface to slow the movement of the cells to a rolling phenotype. Stem cell integrins are then activated and accomplish firm adhesion between the stem cell and vessel wall, with a particularly important role for stem cell VCAM-1 engaging endothelial VLA-4. The chemokine CXCL12 (SDF1) interacting with stem cell CXCR4 receptors and ionic calcium interacting with the calcium-sensing receptor are important in the process of stem cells getting from the circulation to where they engraft in the bone marrow. This is particularly true in the developmental move from fetal liver to bone marrow.

In the adult, the role for CXCR4 is in retention of stem cells in the bone marrow as well as getting them there. Interrupting that retention process through specific molecular blockers of the CXCR4/CXCL12 interaction, cleavage of CXCL12, or downregulation of the CXCR4 receptor can result in the release of stem cells into the circulation. This process is an increasingly important aspect of recovering stem cells for therapeutic use as it has permitted the harvesting process to be done by leukapheresis rather than bone marrow punctures in the operating room. Granulocyte colony-stimulating factor and plerixafor, a macrocyclic compound that can block CXCR4, are both used clinically to mobilize marrow hematopoietic stem cells for transplant.

HEMATOPOIETIC STEM CELL MICROENVIRONMENT

The concept of a specialized microenvironment, or stem cell niche, was first proposed to explain why cells derived from the bone marrow of one animal could be used in transplantation and again be found in the bone marrow of the recipient. This niche is more than just a housing site for stem cells, however. It is an anatomic location where regulatory signals are provided that allow the stem cells to thrive, to expand if needed, and to provide varying amounts of descendant daughter cells. In addition, unregulated growth of stem cells may be problematic based on their undifferentiated state and self-renewal capacity. Thus, the niche also regulates the number of stem cells produced. In this manner, the niche has the dual function of serving as a site of nurture but imposing limits for stem cells: in effect, acting as both a nutritive and constraining home.

The niche for blood stem cells changes with each of the sites of blood production during development, but for most of human life, it is located in the bone marrow. Within the bone marrow, the perivascular space particularly in regions of trabecular bone serves as a niche. The mesenchymal and endothelial cells of the marrow microvessels produce kit ligand and CXCL12, both known to be important for hematopoietic stem cells. Other cell types, such as sympathetic neurons, nonmyelinating Schwann cells, macrophages, megakaryocytes, osteoclasts, and osteoblasts, have been shown to regulate stem cells, some by direct and others by indirect effects. Extracellular matrix proteins like osteopontin and heparan sulfates also affect stem cell function. The endosteal region appears to be particularly important for transplanted cells, in part because many of the mesenchymal cells and sinusoidal blood vessels of the central marrow are disrupted by the conditioning regimens used to prepare a patient for transplantation. The functioning of the niche as a supportive context for stem cells is of obvious importance for maintaining hematopoiesis and in transplantation. An active area of study involves determining whether the niche is altered in disease as experimental models have shown that mutations in niche cells can lead to myeloid malignancies.

EXCESS CAPACITY OF HEMATOPOIETIC STEM CELLS

In the absence of disease, one never runs out of hematopoietic stem cells. Indeed, serial transplantation studies in mice suggest that sufficient stem cells are present to reconstitute several animals in succession, with each animal having normal blood cell production. The fact that allogeneic stem cell transplant recipients also never run out of blood cells over decades argues that even the limiting numbers of stem cells provided to them are sufficient. How stem cells respond to different conditions to increase or decrease their mature cell production remains poorly understood. Clearly, negative feedback mechanisms affect the level of production of most of the cells, leading to the normal tightly regulated blood cell counts. However, many of the regulatory mechanisms that govern production of more mature progenitor cells do not apply or apply differently to stem cells. Similarly, most of the molecules shown to be able to change the size of the stem cell pool have little effect on more mature blood cells. For example, the growth factor erythropoietin, which stimulates red blood cell production from precursor cells, has no effect on stem cells. Similarly, granulocyte colony-stimulating factor drives the rapid proliferation of granulocyte precursors but has little or no effect on the cell cycling of stem cells. Rather, it changes the location of stem cells by indirect means, altering molecules such as CXCL12 that tether stem cells to their niche. Molecules shown to be important for altering the proliferation, self-renewal, or survival of stem cells, such as cyclin-dependent kinase inhibitors, transcription factors like Bmi-1, microRNA-processing enzymes like Dicer, or even metabolic regulators like pyruvate kinase isoforms, have little or different effects on progenitor cells. Hematopoietic stem cells have governing mechanisms that are distinct from the cells they generate.

HEMATOPOIETIC STEM CELL DIFFERENTIATION

Hematopoietic stem cells sit at the base of a branching hierarchy of cells culminating in the many mature cell types that compose the blood and immune system (Fig. 101-2). The maturation steps leading to terminally differentiated and functional blood cells take place both as a consequence of intrinsic changes in gene expression and external, niche-directed or cytokine-directed changes in the cells. Our knowledge of the details remains incomplete. As stem cells mature to progenitors, precursors, and, finally, mature effector cells, they undergo a series of functional changes. These include the acquisition of functions defining mature blood cells, such as phagocytic capacity or hemoglobin synthesis. They also include the progressive loss of plasticity (i.e., the ability to become other cell types). For example, some myeloid progenitors can make all cells in the myeloid series but none in the lymphoid series. As common myeloid progenitors mature, they become precursors for either monocytes and granulocytes or erythrocytes and megakaryocytes, but not both. Some amount of reversibility of this process may exist early in the differentiation cascade, but that is lost beyond a distinct stage in normal physiologic conditions.

As cells differentiate, they may also lose proliferative capacity (Fig. 101-3). Mature granulocytes are incapable of proliferation and only increase in number by increased production from precursors. The exceptions to the rule are some tissue-resident macrophages, which appear capable of proliferation, and lymphoid cells. Lymphoid cells retain the capacity to proliferate but have linked their proliferation to the recognition of particular proteins or peptides by specific antigen

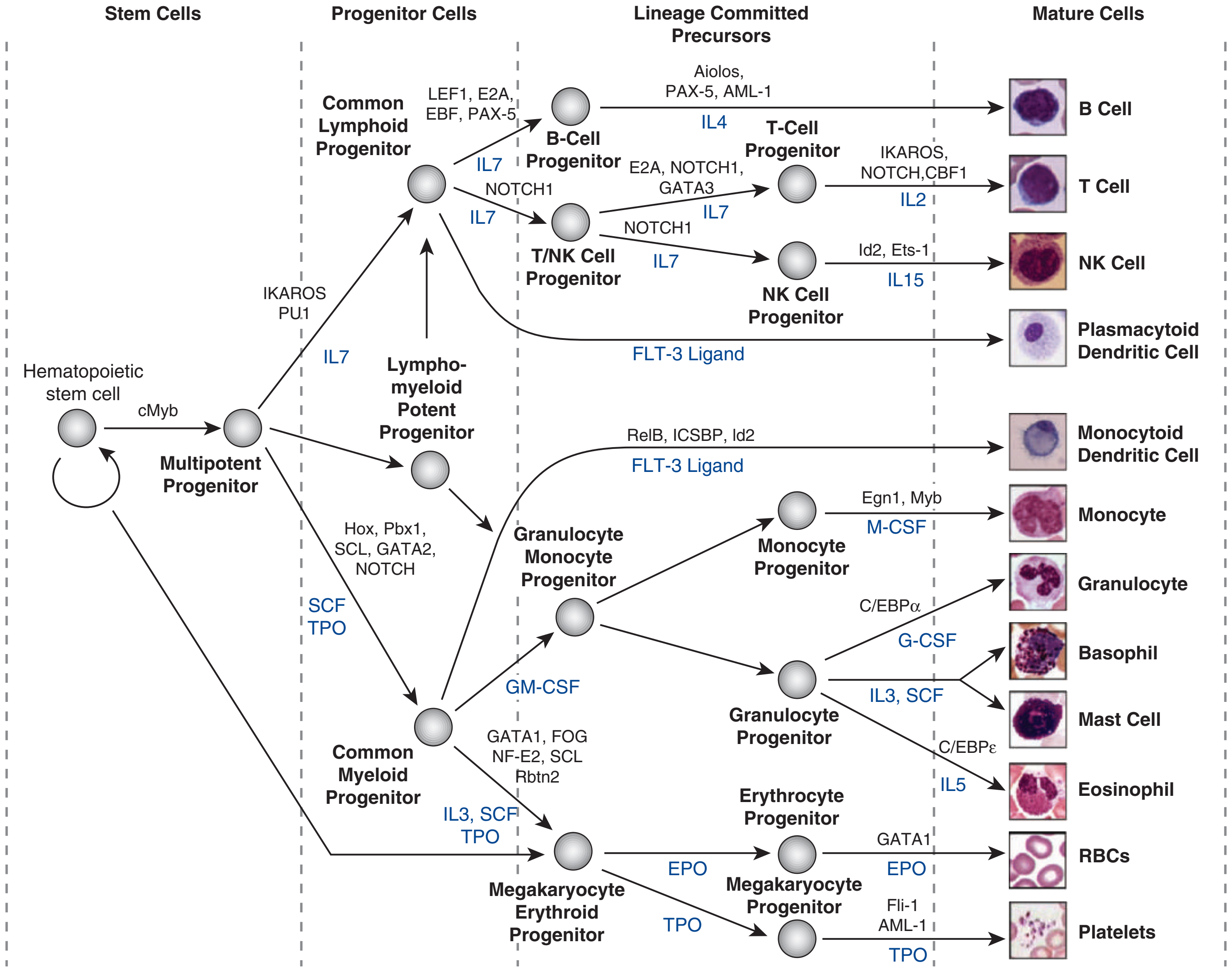
Stem Cells
Progenitor Cells
Lineage Committed Precursors
Mature Cells
Hematopoietic stem cell
cMyb
Multipotent Progenitor
IKAROS PU1
IL7
Common Lymphoid Progenitor
LEF1, E2A, EBF, PAX-5
IL7
B-Cell Progenitor
Aiolos, PAX-5, AML-1
IL4
B Cell
NOTCH1
IL7
T/NK Cell Progenitor
E2A, NOTCH1, GATA3
IL7
T-Cell Progenitor
IKAROS, NOTCH,CBF1
IL2
T Cell
NOTCH1
IL7
NK Cell Progenitor
Id2, Ets-1
IL15
NK Cell
FLT-3 Ligand
Plasmacytoid Dendritic Cell
Lympho-myeloid Potent Progenitor
RelB, ICSBP, Id2
FLT-3 Ligand
Monocytoid Dendritic Cell
Hox, Pbx1, SCL, GATA2, NOTCH
SCF TPO
Granulocyte Monocyte Progenitor
Monocyte Progenitor
Egn1, Myb
M-CSF
Monocyte
C/EBPα
G-CSF
Granulocyte
GM-CSF
Common Myeloid Progenitor
Granulocyte Progenitor
IL3, SCF
Basophil
Mast Cell
C/EBPε
IL5
Eosinophil
GATA1, FOG NF-E2, SCL Rbtn2
IL3, SCF TPO
Megakaryocyte Erythroid Progenitor
EPO
Erythrocyte Progenitor
GATA1
EPO
RBCs
TPO
Megakaryocyte Progenitor
Fli-1 AML-1
TPO
Platelets

FIGURE 101-2 Hierarchy of hematopoietic differentiation. *Stem cells* are multipotent cells that are the source of all descendant cells and have the capacity to provide either long-term (measured in years) or short-term (measured in months) cell production. *Progenitor cells* have a more limited spectrum of cells they can produce and are generally a shorter-lived, highly proliferative population also known as transient amplifying cells. *Precursor cells* are cells committed to a single blood cell lineage but with a continued ability to proliferate; they do not have all the features of a fully mature cell. *Mature cells* are the terminally differentiated product of the differentiation process and are the effector cells of specific activities of the blood and immune system. Progress through the pathways is mediated by alterations in gene expression. The regulation of the differentiation by soluble factors and cell-cell communications within the bone marrow niche are still being defined. The transcription factors that characterize particular cell transitions are illustrated on the *arrows*; the soluble factors that contribute to the differentiation process are in *blue*. This picture is a simplification of the process. Active research is revealing multiple discrete cell types in the maturation of B cells and T cells and has identified cells that are biased toward one lineage or another (rather than uncommitted) in their differentiation. EPO, erythropoietin; RBC, red blood cell; SCF, stem cell factor; TPO, thrombopoietin.

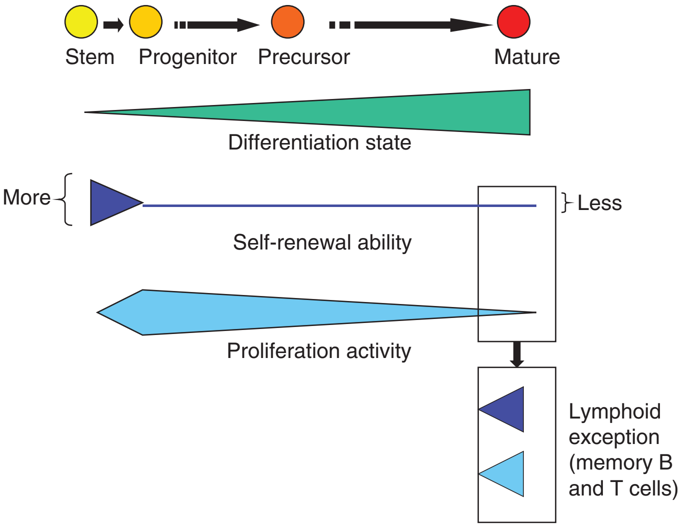

FIGURE 101-3 Relative function of cells in the hematopoietic hierarchy. The boxes represent distinct functional features of cells in the myeloid (*upper box*) versus lymphoid (*lower box*) lineages.

receptors on their surface. Like many tissues with short-lived mature cells such as the skin and intestine, blood cell proliferation is largely accomplished by a more immature progenitor population. In general, cells within the highly proliferative progenitor cell compartment are also relatively short-lived, making their way through the differentiation process in a defined molecular program involving the sequential activation of particular sets of genes. For any particular cell type, the differentiation program is difficult to speed up. The time it takes for hematopoietic progenitors to become mature cells is ~10–14 days in humans, evident clinically by the interval between cytotoxic chemotherapy and blood count recovery in patients.

Although hematopoietic stem cells are generally thought to have the capacity to form all cells of the blood, individual stem cells are heterogeneous in their differentiation potential. That is, some stem cells are "biased" to become mature cells of a particular type. In addition, individual stem cells may respond differently to proliferation or cell death signals. Therefore, the stem cell population is an aggregate of cells with somewhat distinctive properties that have the collective characteristics ascribed to hematopoietic stem cells.

■ SELF-RENEWAL AND CLONAL DYNAMICS

Self-renewal is the ability to divide while preserving an undifferentiated state. This stem cell characteristic is generally not seen in progenitor or mature cells where proliferation is coupled with progressive differentiation. Stem cells being able to asymmetrically divide, such that one daughter cell is the product of self-renewal and the other enters into the differentiating progenitor pool, enables the hematopoietic system to have a modestly sized pool of stem cells yet produce seven orders of magnitude greater numbers of mature blood cells each day.

Self-renewal has its risks, however. Genetic mutations that occur in a stem cell will durably persist in stem cells because of self-renewal. In contrast, mutations in progenitors will largely be lost as the cells terminally differentiate and die. It is the stem cell then that has greater potential to accumulate genetic mutations, a setting that can lead to cancer. Countering this risk is the small number of stem cells, the radioprotective environment of their bone localized niche, and the relative quiescence of stem cells.

Stem cells have distinctive cell cycle regulation. Some are deeply quiescent, serving as a deep reserve, whereas others are more proliferative and replenish the short-lived progenitor population. Hematopoietic stem cells are generally cytokine-resistant, remaining dormant even when cytokines drive bone marrow progenitors to proliferation rates measured in hours. Stem cells, in contrast, are thought to divide at far longer intervals, measured in months to years, for the most quiescent cells. This quiescence is difficult to overcome in vitro, limiting the ability to effectively expand human hematopoietic stem cells. The process may be controlled by particularly high levels of cyclin-dependent kinase inhibitors like p57 or CDKN1c that restrict entry of stem cells into the cell cycle by blocking the G_1-S transition. Exogenous signals from the niche also appear to enforce quiescence, including angiogenin, interleukin 18, and perhaps angiopoietin 1.

Individual stem cells may vary in their proliferation and self-renewal features. This can lead to distortions in the representation of any given clone of stem cells, a feature commonly seen with aging and often associated with acquired somatic mutations. Hematopoietic stem cells are estimated to acquire 17 somatic mutations with each year of life based on deep sequencing studies. Some of these appear to provide a fitness advantage to the stem cell as the presence of mutated or "variant" alleles contributing to >1% of blood cells is virtually uniform by age 70. Furthermore, it is estimated that 10–20 such clones exist in individuals by 70, and in aggregate, those clones provide between 30 and 60% of the blood cells. Thus, the diversity of active stem cell clones declines with age with a likely accompanying reduced diversity of functions that each clone provides. Whether this contributes to immune alterations or other aspects of aging is to be determined. However, some expanded clones are associated with "driver" mutations observed in myelodysplasia and myeloid leukemias. These clones do have a low frequency of progression to overt neoplastic disease, and at least one of them, an inactivating mutation of *TET2*, can also increase adverse outcomes from a number of chronic inflammatory conditions.

The most common mutations associated with clonal hematopoiesis are of epigenetic regulatory genes. For example, inactivating mutations of the DNA methyl transferase *DNMT3a* or the dioxygenase involved in DNA demethylation, *TET2*, are commonly found, as are mutations in *ASXL1*, a member of the polycomb family of genes whose products alter chromatin structure, a high-order DNA organization that affects transcription. Therefore, epigenetic control appears to be critical for homeostasis of the hematopoietic stem cell pool and constraining the outgrowth of potentially pathogenic stem cell clones.

THE RELATIONSHIP OF STEM CELLS TO CANCER

Some cancers have been shown to have a cellular hierarchy similar to normal tissues, with stem-like cells having the capacity to self-renew and differentiate. These stem-like cells can be experimentally transplanted into immunodeficient animals and initiate a new cancer. It is hypothesized that these stem-like cells may be the basis for disease relapse after therapy as they have distinctive molecular features from other cells in the cancer that may render them less vulnerable to therapies. Myeloid leukemias have been experimentally shown to be consistent with this model. Focusing therapies on the stem-like cells as opposed to the bulk population of the cancer cells as a means to improve cure rates is an active area of investigation.

Given that hematopoietic stem cells can accumulate genetic mutations by virtue of their self-renewal, it is logically appealing to regard them as the likely cell source of leukemias. Experimental testing of this hypothesis has shown that stem cells are more likely to result in malignancy following an oncogenic mutation. However, some more mature populations with less well-defined self-renewal capability can also be transformed to malignancy. Therefore, self-renewal may also be acquired by mutation, and cancer stem-like cells need not have originated in normal stem cells.

STEM CELLS AS TARGETS OF GENE THERAPY OR GENE EDITING

The hematopoietic stem cell is the ideal cell target for genetic therapies intended to durably change the genome of blood cells. Because stem cells can persist for the lifetime of an individual, genetically modifying them can provide curative therapies for genetic disorders such as hemoglobinopathies or congenital immunodeficiencies. The extensive cell proliferation with limited or no self-renewal among progenitor populations makes them less able to provide durable benefit if they are genetically modified. Therefore, modification of stem cells is critical and requires integration of the genetic therapy into the host cell's genome to persist. This is challenging because the relative quiescence of stem cells makes them less susceptible to the molecular events of integration of a transgene or editing of a target gene. Yet, many of these technical challenges have been overcome sufficiently that there are medicines composed of hematopoietic stem cells bearing a gene to offset a missing or defective gene for active cerebral adrenoleukodystrophy, β thalassemia, and sickle cell anemia. These are termed gene therapies as a full coding sequence of the abnormal gene is provided. There is also a gene-edited therapy for β thalassemia and sickle cell anemia based on CRISPR-based technology. CRISPR-Cas allows for selected deletion of regions of genes and is thus called gene editing. In the case of the hemoglobinopathies, a gene that suppresses fetal hemoglobin, *BLC11A*, is the target. Reducing *BLC11A* expression by editing leads to fetal hemoglobin expression that overcomes the abnormal adult β hemoglobins causing β thalassemia and sickle cell anemia.

These therapies are powerful new ways in which hematopoietic stem cells can be medically beneficial. By modifying the patient's own cells, this type of hematopoietic stem cell therapy does not require immunologic matching with a donor and avoids graft-versus-host disease. At present, it still involves myeloablative conditioning of the patient prior to receiving the cells as they are modified outside the body and must compete with endogenous stem cells in the bone marrow. Efforts to enable gene editing in vivo and thereby avoid transplantation are being actively investigated.

FURTHER READING

Jacobsen SEW, Nerlov C: Haematopoiesis in the era of advanced single-cell technologies. Nat Cell Biol 1:2, 2019.

Mitchell E et al: Clonal dynamics of haematopoiesis across the human lifespan. Nature 606:343, 2022.

Weeks LD, Ebert BL: Causes and consequences of clonal hematopoiesis. Blood 26:2235, 2023.

102

Iron Deficiency and Other Acquired Anemias Due to Decreased Erythrocyte Production

Clara Camaschella, Domenico Girelli

This chapter provides basic information on iron metabolism as an introduction to common anemias due to iron deficiency, inflammation, and aging, conditions in which alterations of iron balance play a prominent role. The pathogenetic mechanism of the anemia is hypoproliferation of the marrow precursors.

IRON METABOLISM

Iron is indispensable for multiple cell functions, including oxygen delivery, DNA replication and repair, metabolism, and energy production: hence, it is conserved and reutilized. However, to prevent potential toxicity due to the propensity of free iron to generate reactive oxygen species, iron is always bound to proteins, and its homeostasis is tightly regulated. Body iron cycle and distribution are illustrated in **Fig. 102-1**. Iron is provided to all cells by transferrin, a bilobed glycoprotein of hepatic origin abundant in plasma (204–360 mg/dL), that binds iron within a single (monoferric) or both lobes (diferric transferrin). Transferrin carries only 3–4 mg of iron but has a fast turnover. To produce about 200 billion erythrocytes daily, erythroid precursors need about 20–25 mg of iron, mainly supplied by macrophage recycling since dietary iron absorption is limited to 1–2 mg daily. Diferric transferrin binds with high affinity the iron importer transferrin receptor in all cells. The ligand-receptor complex is internalized via clathrin-coated pits to form the endosome. Released by low pH, iron is exported to cytosol by divalent metal transporter 1 (DMT1) and mainly used for heme and iron-sulfur cluster biosynthesis in mitochondria. Apotransferrin and transferrin receptor are recycled to cell surface and reutilized. Excess iron is safely sequestered by cytosolic ferritin, a large hollow shell formed by 24 (heavy and light) subunits in variable combination that stores iron into its central cavity, a process especially efficient in spleen macrophages. Ferritin is also secreted in blood where its concentration may be a marker of stored iron. In iron deficiency, cells can recover iron from ferritin by an autophagic process called ferritinophagy. Diferric transferrin also binds with lower affinity the iron sensor transferrin receptor 2 on hepatocyte **(Fig. 102-2*A*)** and erythroid cell surface. Erythroblasts express the highest number of transferrin receptors (~300,000–400,000/cell) and take up iron only from transferrin. Receptors not bound to transferrin are protease cleaved and released into blood where their concentration as soluble receptor is an estimate of total or iron-deficient erythropoiesis. Transferrin receptor is indispensable to all hematopoietic cells from stem cells to neutrophils. Lymphocyte function is strongly iron-dependent, as shown by the immunodeficiency resulting from rare mutations in transferrin receptor that hamper iron uptake. In the liver, transferrin receptor has the extra function of hepcidin regulator, binding and sequestering the hemochromatosis protein HFE (Fig. 102-2*A*). Liver, pancreas, and heart, but not erythroblasts, can take up non-transferrin-bound-iron (NTBI), which appears in the circulation when transferrin saturation exceeds 60%. No regulated mechanism for iron excretion exists. Physiologic iron loss occurs only through cell desquamation and menses in females and roughly balances the amount of daily absorbed iron (Fig. 102-1).

REGULATION OF IRON HOMEOSTASIS

Systemic iron homeostasis is regulated by *hepcidin*, a 25-amino-acid peptide mainly produced by hepatocytes that, blocking iron export and degrading the iron exporter ferroportin, inhibits plasma iron flux from duodenal enterocytes, iron-recycling macrophages, and iron-storing hepatocytes (Fig. 102-1). High hepcidin levels decrease, while low hepcidin levels increase, iron delivery to plasma. Hepcidin expression is increased by high plasma and liver iron that activate the BMP (bone morphogenetic protein)-SMAD (suppressors of mothers against decapentaplegic) signaling pathway. The proteins HFE, transferrin receptor 2, and hemojuvelin, whose mutations lead to different forms of hemochromatosis **(Chap. 426)**, contribute to hepcidin upregulation. Hepcidin is increased also by inflammatory cytokines, through interleukin (IL) 6 signal transducer and activator of transcription 3 (STAT3) pathway **(Fig. 102-2*A*)**.

Hepcidin expression is suppressed in iron deficiency by the hepatic transmembrane serine protease TMPRSS6 that antagonizes the BMP pathway. Another powerful inhibitor is erythroferrone **(Fig. 102-2*B*)**, the hormone secreted by erythroblasts upon erythropoietin stimulation, such as after a hemorrhage or in hypoxia, which sequesters the BMPs. Other hepcidin inhibitors include testosterone, which adapts iron supply to male erythropoiesis; fibrinogen-like protein 1 (FGL1), released by hepatocytes; and platelet-derived growth factor BB produced by many cell types in hypoxia.

At the cellular level, the iron regulatory proteins (IRPs) control the synthesis of transferrin receptor, ferritin, and ferroportin by interacting with the iron-responsive elements in their mRNA, thus coordinating cell iron uptake, with storage and efflux, according to the cell iron need. IRPs also tune erythropoiesis by regulating hypoxia-inducible factor (HIF) 2α and thus erythropoietin and the first enzyme of heme biosynthesis 5′-aminolevulinate synthase 2.

Intestinal Iron Absorption Iron needs change during life according to age and gender **(Table 102-1)**. In adults, only about 1–2 mg of iron are absorbed from the diet. Dietary iron content is closely related to caloric intake, with 6–7 mg elemental iron every 1000 calories: hence Western diet contains about 15 mg iron daily. The food composition

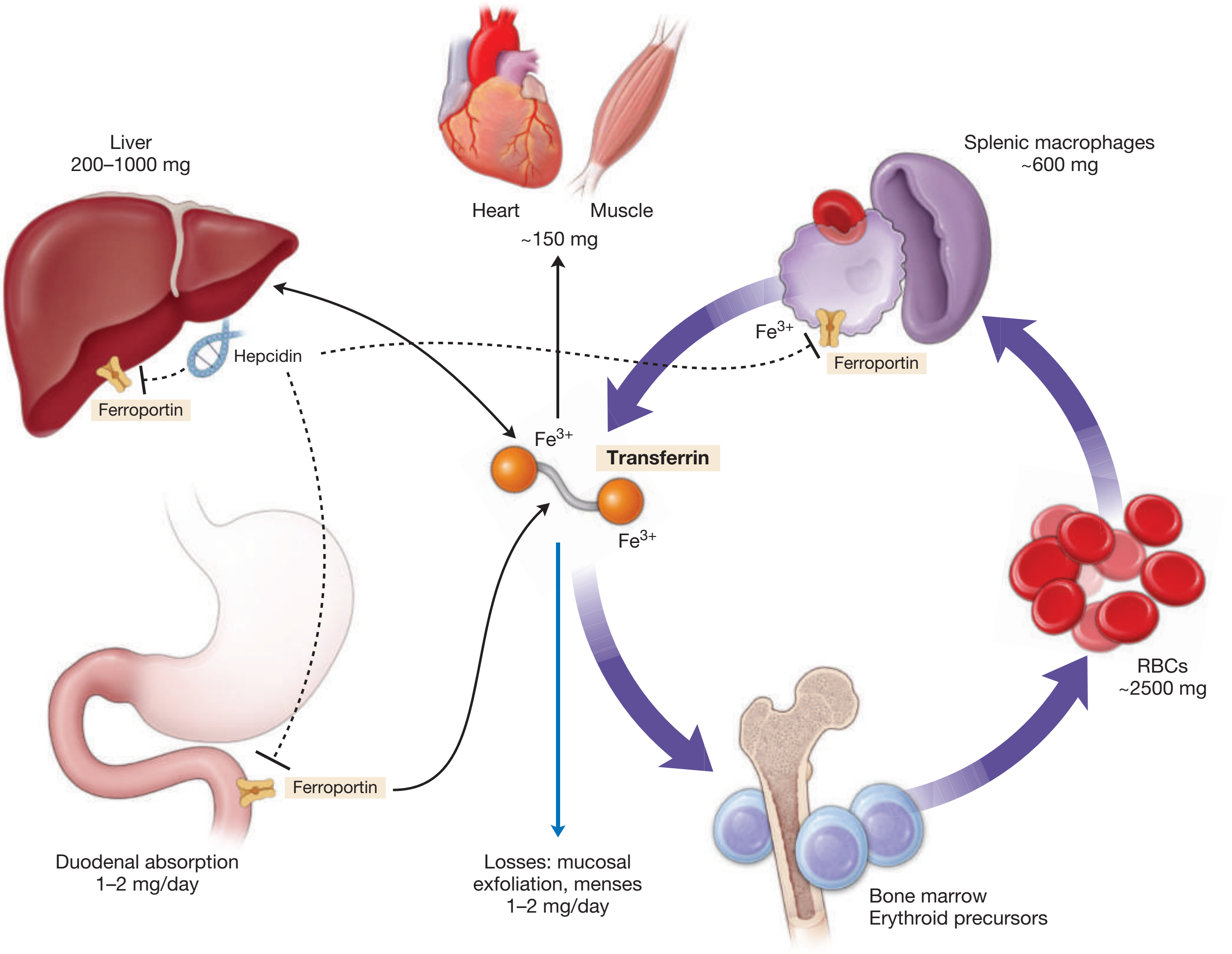

FIGURE 102-1 The iron cycle. Transferrin is central to iron trafficking. Most iron is recycled from macrophages, while only 1–2 mg are derived from intestinal absorption; mg indicates iron content of organs/tissues. Iron flux to plasma is controlled by hepcidin, a peptide produced by the liver that inhibits the activity of the iron exporter ferroportin in enterocytes, macrophages, and hepatocytes. Iron loss balances iron absorption.

affects iron bioavailability. Heme contained in red meat is easily taken up by enterocytes through a still undefined mechanism. Nonheme ferric (Fe^{3+}) iron is converted to ferrous (Fe^{2+}) iron by the low local pH and by the ferrireductase duodenal cytochrome B (DCYTB) before being imported from the lumen by DMT1. Phytates and phosphates contained in vegetables and cereals and tannin contained in tea and coffee reduce, while ascorbic acid enhances, inorganic iron absorption. Absorption takes place prevalently in duodenal enterocytes where iron can be utilized or released to plasma transferrin by ferroportin upon oxidation by hephaestin. Surplus iron is stored in cytosolic ferritin and lost by enterocytes sloughed into the lumen. Duodenal iron absorption depends both on the local levels of HIF-2α that enhance the expression of the absorption machinery components (DMT1, DCYTB, ferroportin) and on the hepcidin-mediated suppression of iron export. In iron deficiency, high levels of HIF-2α that favor luminal iron uptake and low hepcidin that facilitates export to plasma cooperate to maximize the absorption process. Iron in the gut may reshape the local microbiome, whose influence is emerging as modulator of body iron.

Macrophage Iron Recycling Specialized liver and spleen macrophages remove senescent erythrocytes (0.8–1% daily) and recycle to plasma the iron recovered from heme (Fig. 102-1). The high (20–25 mg daily) macrophage flux maintains steady-state iron supply to erythropoiesis and may further increase to compensate for hypoferremia or accelerated erythropoiesis. In contrast, decreased iron recycling, as caused by high hepcidin levels in inflammation, leads to anemia, with iron stored in macrophages. Because of its essential role in iron recycling, ferroportin is regulated at multiple levels, the most important being hepcidin-dependent. In intravascular hemolysis, hepatocytes and macrophages recover iron from hemoglobin-haptoglobin and heme-hemopexin complexes.

Iron Utilization Erythropoiesis utilizes a large amount of iron for heme/hemoglobin production and facilitates its own iron supply by releasing erythroferrone (Fig. 102-2*B*) upon erythropoietin stimulation. The release is transient; chronic erythroferrone production in ineffective erythropoiesis, as in non-transfusion-dependent β thalassemia **(Chap. 103)**, leads to iron overload. In every cell, iron is mainly utilized in mitochondria to produce heme and iron/sulfur clusters, components of enzymes and complexes involved in the cell cycle, oxidative phosphorylation and energy production, fundamental processes in heart, skeletal muscle, and brain. Although bone marrow consumes most available iron, excessive erythroid use at the expense of other organs is prevented; one impeding mechanism is the iron export by erythroid ferroportin that contributes to serum iron levels in hypoferremia.

IRON DEFICIENCY AND IRON-DEFICIENCY ANEMIA

Iron deficiency and iron-deficiency anemia are global health problems affecting billions of people and a main cause of years lived with disability, especially among women. In isolated iron deficiency, iron stores are depleted while hemoglobin levels are still normal. In iron-deficiency anemia, stores are depleted and iron is insufficient for erythropoiesis. Iron deficiency is *absolute* when the stores are exhausted and *functional* when erythropoiesis is iron-restricted but the stores are adequate or even increased.

PATHOPHYSIOLOGY

Isolated iron deficiency with low/absent iron stores is a common condition induced by rapid growth, inadequate nutrition, abundant menses. When iron is still sufficient for erythropoiesis, erythrocytes are normochromic and normocytic while ferritin is decreased (<15–30 μg/L). Isolated iron deficiency may progress to anemia of variable degrees. With transferrin saturation <15–20% iron supply to erythropoiesis is insufficient and erythrocytes become hypochromic and microcytic, the classical feature of iron-deficiency anemia. Though anemia is the most evident outcome, other organs, such as heart, muscles, and brain, are affected by iron deficiency. In chronic heart failure (CHF), iron deficiency causes symptoms even independently from anemia. Immune response may be blunted because lymphocytes require iron during the metabolic burst that mounts adaptative defense against infectious agents.

The relevance of low hepcidin in iron deficiency is clearly documented. Mutations of the negative regulator of hepcidin *TMPRSS6* cause a rare form of hereditary iron-refractory iron-deficiency anemia (IRIDA), characterized by reduced iron absorption and refractoriness to oral iron supplementation. Similarly, in inflammation, high hepcidin levels sequester iron into stores, leading to functional iron deficiency and iron-restricted erythropoiesis with low serum iron and transferrin saturation in the presence of elevated ferritin levels. Not infrequently, absolute and functional iron deficiency coexist, as in inflammatory bowel diseases, chronic kidney disease (CKD), and aging. Functional iron deficiency may be caused also by erythrocyte-stimulating agents (ESA) that, expanding erythropoiesis, unbalance iron demand and supply.

CAUSES OF IRON DEFICIENCY

Causes of absolute iron deficiency are detailed in **Table 102-2**. In general, iron supply is inadequate for iron needs because of excessive demand, insufficient intake, defective absorption, or blood loss. In clinical practice, multiple causes frequently co-occur, especially in the elderly and in patients with multimorbidity. Several genome-wide association studies showed that variants of iron genes, such as *TMPRSS6* and *TF* (encoding transferrin), may confer a genetic susceptibility to iron deficiency.

EPIDEMIOLOGY

Iron-deficiency anemia is estimated to affect ~1.2 billion individuals, roughly half of the world anemic population. The high iron requirements explain its high prevalence in preschool children (estimated ~25%), young women (37–40%), and pregnant women (>40% presenting with severe forms in low-income countries). Isolated iron deficiency is even more common, though precise figures are lacking. Both conditions are especially frequent in low-income countries, although high-income countries are not spared. In the United States, the prevalence of iron deficiency and iron-deficiency anemia in females aged 12–21 years based on ferritin <30 μg/L has been recently reported to be 38.6 and 6.3%, respectively. Anemia of inflammation,

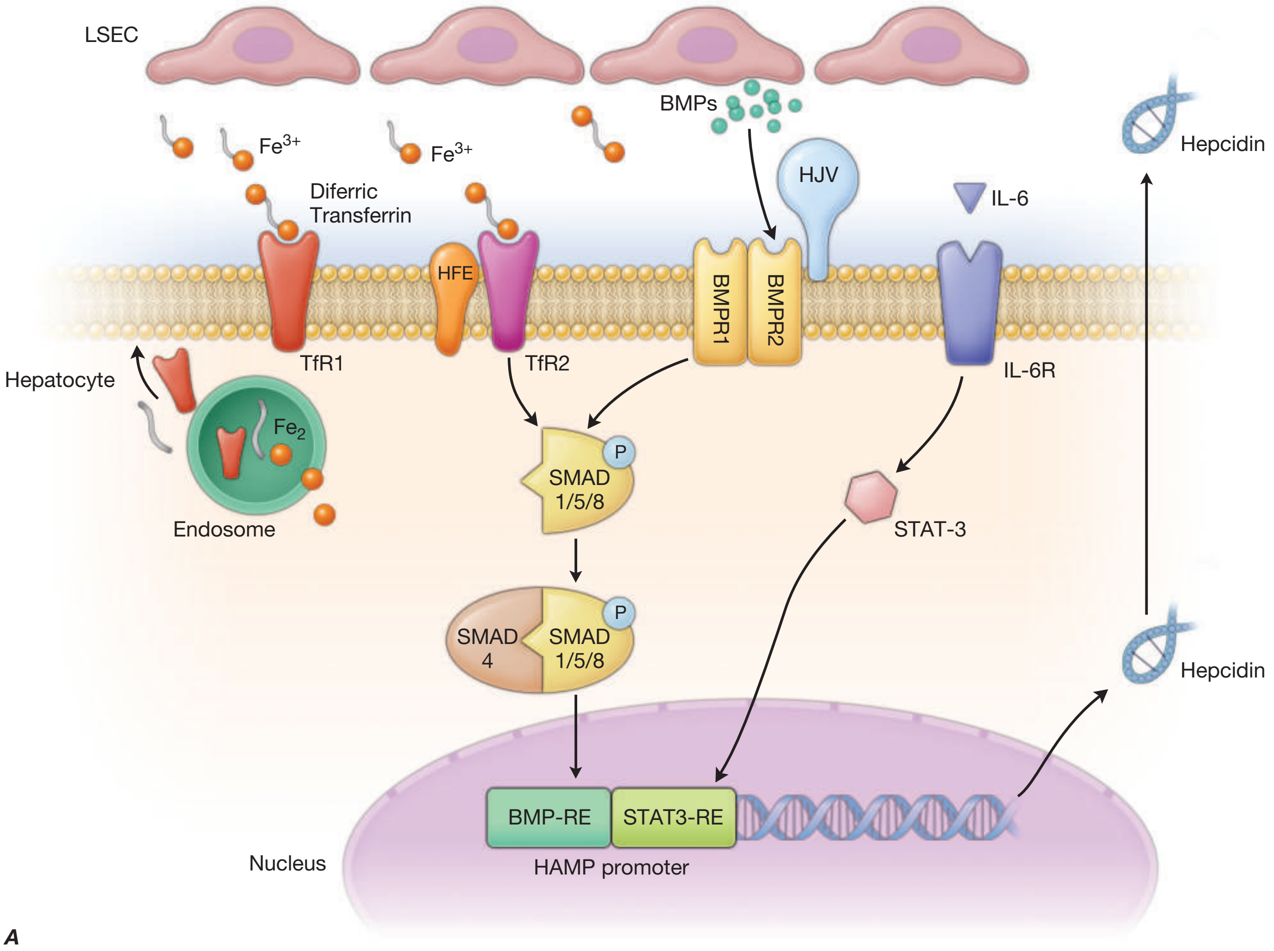

FIGURE 102-2 A. Molecular regulation of hepcidin in iron deficiency. In steady state, diferric transferrin binds TFR1 to release iron to hepatocytes and interacts with TFR2 to signal iron concentration. Liver sinusoidal endothelial cells (LSECs) produce bone morphogenic proteins (BMPs) that interact with BMP receptors type 1 and 2 (BMPR1 and BMPR2, respectively) to activate cytosolic suppressors of mothers against decapentaplegic (SMAD) 1/5/8, with the cooperation of the coreceptor hemojuvelin (HJV) and HFE bound to TFR2. SMAD 1/5/8 complex with SMAD 4 and translocate to the nucleus, where they interact with BMP responsive elements (BMPRE) in hepcidin (*HAMP*) promoter to upregulate its expression. In inflammation, hepcidin is upregulated also through the interleukin 6 (IL-6)–IL-6 receptor–STAT3 pathway.

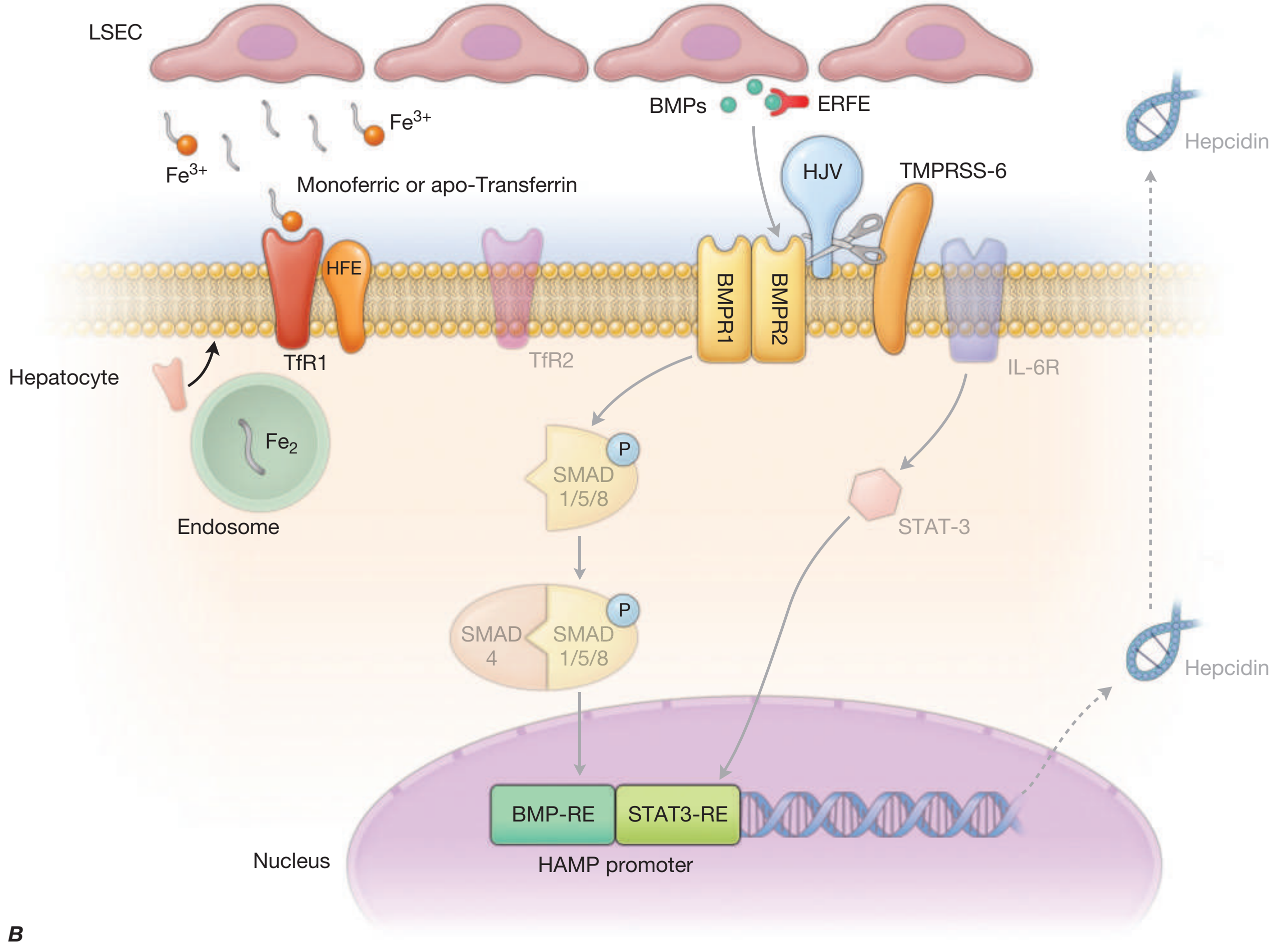

FIGURE 102-2 B. Molecular regulation of hepcidin in iron deficiency. In iron deficiency, transferrin is prevalently mono- or apotransferrin, the production of BMP by LSEC is scarce, the BMP coreceptor HJV is cleaved from the cell surface by the serine protease TMPRSS6, and TFR1 cannot transport iron inside the cell since it is bound to HFE. The SMAD and STAT3 pathways are inactive (indicated by *faint arrows*), and hepcidin is low/ undetectable.

due to functional iron deficiency, is the second most common anemia worldwide, especially in hospitalized patients, although precise figures are not available.

CLINICAL PRESENTATION

Anemia develops only in the late stage of iron deficiency, since the erythron is relatively spared compared to heart and skeletal muscle that may become iron depleted earlier. Because of slow progression and efficient compensatory mechanisms, anemia often remains undetected, especially in young people, until hemoglobin level falls below 8 g/dL. Isolated iron deficiency should not be underestimated since the resulting suboptimal metabolic function of myocytes, cardiomyocytes, and neurons may lead to fatigue, irritability, poor concentration, restless legs syndrome, and reduced exercise capability. As iron deficiency progresses, classical symptoms of anemia may appear (Chap. 66). Clinical clues pointing to iron-deficiency anemia include pica (craving and compulsive ingestion of nonnutritive substances such as ice, soil, or clay), nail fragility and koilonychia, hair loss, cheilosis, atrophic glossitis, and dysphagia caused by formation of esophageal webs (Plummer-Vinson syndrome). Clinical history is important to identify such symptoms as well as symptoms of any underlying disorder (Table 102-2).

TABLE 102-1 Daily Iron Requirements (mg) According to Age and Gender

Children	0.5–1.0
Adolescents	1.2–1.5
Adult males	1.0
Females (fertile age)	1.20–1.50
Pregnancy second trimester	2.8–3.0
Pregnancy third trimester	4.0–5.0
Breastfeeding	1.0-1.15
Females (postmenopausal)	1.0

Laboratory Iron Studies The diagnosis of iron deficiency with or without anemia is a two-step process that first requires laboratory tests, and then the identification of the cause(s) (Table 102-2). The laboratory diagnosis is usually simple, while tracing the cause may require imaging and endoscopy, procedures essential in males and postmenopausal females to rule out malignancy or other severe causes of bleeding. The absence of stainable iron in the bone marrow remains the gold standard for iron deficiency, although it is rarely needed and unacceptably invasive. The mainstay of absolute iron deficiency is low serum ferritin, which reflects reduced/exhausted iron stores. Although the precise cutoff is debated and varies in different settings, ferritin levels <30 μg/L have sufficiently high sensitivity and specificity (92 and 98%, respectively) to be used as a single test for diagnosing iron deficiency. Serum iron is reduced, but diurnal fluctuations prevent its use as a single test. Transferrin saturation (serum iron/total iron-binding capacity ratio) measures iron available for erythropoiesis, which becomes iron restricted for values <15–20%. Mean corpuscular volume (MCV) is <80 fL in established iron deficiency, but can be normal in early stages. Red cell distribution width (RDW) is often elevated, a measure of anisocytosis. Laboratory diagnosis may be demanding in patients with comorbidities. Classical microcytosis and hypochromia can be absent in elderly patients or masked by concomitant folate

TABLE 102-2 Causes of Absolute Iron Deficiency With or Without Anemia

MECHANISM	CONDITIONS
Increased iron needs	• Children, adolescents (growth spurt) • Pregnancy and breastfeeding • Treatment with erythropoiesis-stimulating agents
Inadequate iron intake	• Vegan or vegetarian diet • Malnutrition, poverty • Prolonged breastfeeding
Decreased iron absorption	• Gastrectomy, bariatric surgery • Celiac disease • *Helicobacter pylori* gastritis • Atrophic/autoimmune gastritis • Proton pump inhibitors[a] • Excess consumption of iron absorption inhibitors (e.g., tea, calcium) • IRIDA (*TMPRSS6* mutations)
Chronic blood losses	*Gastrointestinal* • Erosive esophagitis/gastritis • Peptic ulcer • Large hiatal hernia • Antral gastric ectasia (watermelon stomach) • Benign lesions (e.g., large polyps) • Meckel's diverticulum • Angiodysplasias (including Heyde's syndrome with concurrent aortic stenosis and acquired von Willebrand disease) • Malignancy (especially right colon cancer) • Large hemorrhoids • Inflammatory bowel diseases • Parasites (hookworms, others) *Gynecologic* • Heavy menstrual bleeding[b] • Fibroids • Uterine cancers *Urinary* • Intravascular hemolysis (e.g., PNH, prosthetic heart valves, march hemoglobinuria) • Urinary schistosomiasis (*Schistosoma haematobium*) • Bladder or renal cancer *Pulmonary* • Pulmonary hemosiderosis • Severe hemoptysis *Iatrogenic* • Hemodialysis • Surgical blood losses • Excess blood drawn for diagnostic testing • Drugs inducing or facilitating GI bleeding (aspirin, nonsteroidal anti-inflammatory drugs, corticosteroids, antithrombotic drugs) *Bleeding disorders* • von Willebrand disease, other coagulation defects, platelet disorders • Hereditary hemorrhagic telangiectasia (Rendu-Osler-Weber disease) *Frequent blood donors* *Self-inflicted in psychiatric disorders* (Munchausen's syndrome)

[a] Unlikely the sole cause. [b] May be exacerbated by undiagnosed von Willebrand disease or platelet disorders.

Abbreviations: IRIDA, iron-refractory iron-deficiency anemia; PNH, paroxysmal nocturnal hemoglobinuria.

and/or vitamin B_{12} deficiency in malabsorption disorders. **Table 102-3** lists all potentially useful laboratory parameters. Levels of soluble transferrin receptor increase in iron deficiency, since the receptor is highly expressed and released by erythroid precursors. Reduced reticulocyte hemoglobin content and increased percentage of hypochromic red cells reflect insufficient iron availability for erythropoiesis a few days or 2–3 months before testing, respectively. Serum levels of hepcidin are low/undetectable in absolute iron deficiency, a mechanism that maximizes dietary iron absorption. They are elevated in inflammation and in the rare IRIDA due to mutations of *TMPRSS6*. Overall, the clinical use of laboratory biomarkers other than ferritin and transferrin saturation is currently hampered by limited standardization and/or accessibility. Table 102-3 provides the differential diagnosis of iron deficiency based on iron parameters.

Severe iron-deficiency anemia may be accompanied by thrombocytosis due to prevalent megakaryocyte commitment of the common erythroid-platelet progenitor in the presence of insufficient iron.

TREATMENT

Both anemic patients and those with symptomatic isolated iron deficiency should receive iron replacement therapy. This includes oral and intravenous preparations **(Table 102-4)**, with different pharmacokinetic properties. The choice of the administration route depends mainly on the severity of anemia and the individual probability of iron absorption. To be efficacious, oral compounds require the integrity of the first digestive tract and full suppression of hepcidin. Iron salts, as iron sulfate, containing 65–100 mg of elemental iron per tablet, are the first choice in mild to moderate anemia, being widely available and cheap. Traditionally they were given once or twice daily, preferably before meals (Table 102-4). Gastrointestinal side effects, such as abdominal discomfort, nausea, epigastric burning, metallic taste, diarrhea, or stypsis, are common (30–60% of cases) and due to the toxicity of unabsorbed iron on the intestinal mucosa as a result of its limited absorption rate. The unabsorbed fraction can induce changes in the gut microbiome from harmless lactobacilli to pathogenic enterobacteria, promoting infectious diarrhea in children from low- to middle-income countries.

Another potential drawback that may reduce patient adherence to oral therapy is the need for prolonged administration, usually at least 3 months, to ensure recovery of hemoglobin levels and replenishment of iron stores to prevent early recurrence. Studies tracing iron absorption with stable isotopes are changing the paradigm of oral iron administration. An alternate-day schedule maximizes iron absorption as compared to daily administration, which induces a transient peak of hepcidin; moreover, it is better tolerated. Addition of 80 mL of orange juice (but not of pharmacologic doses of ascorbic acid) maximizes the fraction of absorbed iron, indicating that the best regimen is alternate-day iron salts plus "natural" vitamin C. Lack of reticulocytosis after 1 week or a <1 g/dL hemoglobin increase after 2 weeks of treatment defines the response as inadequate and should prompt a search for the cause, such as lack of compliance, ongoing blood loss, or malabsorption.

Intravenous iron preparations (Table 102-4) are best used in case of intolerance to oral iron or malabsorption and for rapid correction of severe anemia (hemoglobin ≤ 8 g/dL). These nanodrugs share a common structure with a shell of carbohydrate moieties surrounding an iron hydroxide core. Their use has long been limited by safety concerns, especially anaphylaxis reactions that occurred with early preparations, such as the high-molecular-weight dextran, that are no longer on the market. The latest generation of compounds show a good safety profile, do not elicit IgE-mediated response, and are highly stable, allowing the administration of the therapeutic dose in a single infusion or two infusions. A dose test is no longer required. The nanoparticles are taken up by macrophages, and iron is slowly released to transferrin through ferroportin. Infusion reactions may occur if a small fraction of iron is rapidly released from the nanoparticle core, depending on the stability of the shell. Reactions occur in $<0.1\%$ of cases and are usually mild and self-limited. More severe reactions, possibly complement mediated (complement activation–related pseudo-allergy), are extremely rare. Other side effects are flu-like symptoms 2–5 days after the infusion and hypophosphatemia, more commonly reported with ferric carboxymaltose and mediated by the increase of the phosphaturic hormone fibroblast growth factor 23. Hypophosphatemia is

TABLE 102-3 Laboratory Biomarkers for Differential Diagnosis of Iron Deficiency

	IRON DEFICIENCY	ANEMIA OF INFLAMMATION	MIXED FORMS	THALASSEMIA TRAIT
Ferritin (μg/L)	Low (≤15–30)	Normal to high (>100)	Low to normal (<70–100)	Normal to high
Serum iron (μmol/L)	Low	Low	Low	Normal to high
Transferrin saturation (%)	Low (<10)	Low (<20)	Low (<20)	Normal to high
MCV (fL)	Low (<80)	Low to normal	Low to normal	Low
RDW	High	Normal	Normal to high	Low to normal
% HYPO	High (>6%)	High	High	High
CHr	Low	Low	Low	Low
sTfR	High	Low to normal	Normal to high	Normal to high
sTfR/log ferritin	>2	<1	>2	NA
Hepcidin[a]	Very low	Normal to high	Low	Low to normal
Zn-PP[b] (μM/mol heme)	High (>80)	Normal to high	High (>80)	High

[a]Serum hepcidin levels are not routinely measured; reference values are not standardized. [b]Zn-PP indicate zinc protoporphyrins, which reflect iron/zinc competition in the last step of heme synthesis; used in screening programs.

Note: % HYPO (percentage of hypochromic red blood cells) and CHr (hemoglobin content of reticulocytes) are measured by automated analyzers and available upon request.

Abbreviations: MCV, mean corpuscular volume; NA, not assessed; RDW, red cell distribution width; sTfR, soluble transferrin receptor.

transient and asymptomatic in most cases, whereas patients undergoing multiple high-dose infusions or with micronutrient deficiencies are at risk of developing osteomalacia and bone fractures.

Overall, the latest generation of compounds have revolutionized iron supplementation because of easy administration, reduction of hospital visits, and reassuring safety profile. Consequently, indications have been expanded to treat severe anemia in the second and third trimesters of pregnancy and in complex conditions, such as CHF, perioperative anemia, and anemia in the emergency department (see below).

Iron-deficiency anemia should not be treated with blood transfusions unless patients have hemodynamic instability or are rapidly deteriorating.

TABLE 102-4 Iron Preparations for Replacement Therapy

Oral Compounds		
	ELEMENTAL IRON CONTENT,[a] mg	DOSAGE
Iron sulfate	65–105	65–100 mg 1–2 times a day or 65–100 mg on alternate day
Iron gluconate	27–38	1–2 doses per day
Iron fumarate	29–105	1–2 doses per day
Sucrosomial iron[b]	30	30 mg/d
Polysaccharide iron complexes (tablet or elixir)	50–200 15/mL 125/mL	50–200 mg/d
Ferric maltol	30	30 mg 2 times a day
Ferric citrate	45–210	3 doses per day in chronic kidney disease patients with hyperphosphatemia
Intravenous Compounds		
	IRON CONTENT, mg/mL	MAXIMUM SINGLE DOSE
Iron sucrose	20	200–300
Iron gluconate	12.5	125, slow infusion
Low-molecular-weight dextran	50	1000, over 1 h
Iron carboxymaltose	50	750–1000, over 15 min infusion
Iron derisomaltose	100	Up to 20/kg, rapid infusion
Ferumoxytol	30	510, rapid infusion

[a]The content of *elemental* iron in sulfates, gluconate, and fumarate varies from 10 to 33% (e.g., an iron sulfate tablet contains 100 mg of *elemental* iron in 325 mg of the total salt).

[b]Sucrosomial iron consists of ferric pyrophosphate conveyed by a phospholipid and sucrester matrix.

PREVENTION

Prevention is fundamental in low- to middle-income countries where iron-deficiency anemia is common due to the combination of malnutrition and increased blood loss by helminth infections. In this setting, prevention of iron deficiency at the population level employing food fortification or oral iron administration must take into account the infectious risk. Iron deficiency confers some protection from malaria in endemic areas, whereas iron supplementation may favor malaria progression by enhancing parasite trophism for reticulocytes that are increased during anemia recovery. In view of such bidirectional association, in malaria-endemic regions, the World Health Organization recommends iron supplementation only when concomitant strategies to prevent and treat malaria are implemented. Proposed preventive strategies for pregnancy to improve maternal and infant outcome are either routine prenatal supplementation or regular checking of hemoglobin/iron status and on-demand treatment. Frequent blood donors are another population at risk, considering that each unit of blood contains ~200 mg of iron. Prolonging the interdonation interval and monitoring ferritin levels are prevention measures, especially in female donors, who may eventually require iron supplementation. Iron deficiency in patients undergoing major surgery predisposes to anemia due to blood loss during the procedure and the frequent diagnostic testing. Preoperative screening and treatment of iron deficiency is one pillar of patient blood management strategies that aim to reduce perioperative complications and length of hospital stay and save cost. Intravenous route is preferred if surgery is within 6 weeks from the diagnosis of iron deficiency.

ANEMIA OF INFLAMMATION

Anemia of chronic diseases indicates the multifactorial anemia observed in a heterogeneous group of disorders, including infections, autoimmunity, malignancies, and inflammatory bowel diseases, that share mechanisms of immune activation and cytokine production. Similar mechanisms account for the anemia that develops in acute conditions such as sepsis, severe trauma, and after major surgery so that the broader term *anemia of inflammation* is now preferred. This type of anemia spans the whole spectrum of clinical inflammatory disorders, from hyperacute, exemplified by macrophage activation syndrome, to subclinical, as in CHF, diabetes, and obesity.

PATHOPHYSIOLOGY

At least three cytokine-driven mechanisms account for the diminished red cell production in inflammation **(Fig. 102-3)**. Increased hepcidin levels, mainly induced by IL-6, cause macrophage iron sequestration and restrict iron reutilization. Erythropoietin (EPO) production is suppressed by IL-1β and tumor necrosis factor α (TNF-α). Response of erythroid progenitors to EPO is blunted by interferon-γ (IFN-γ)

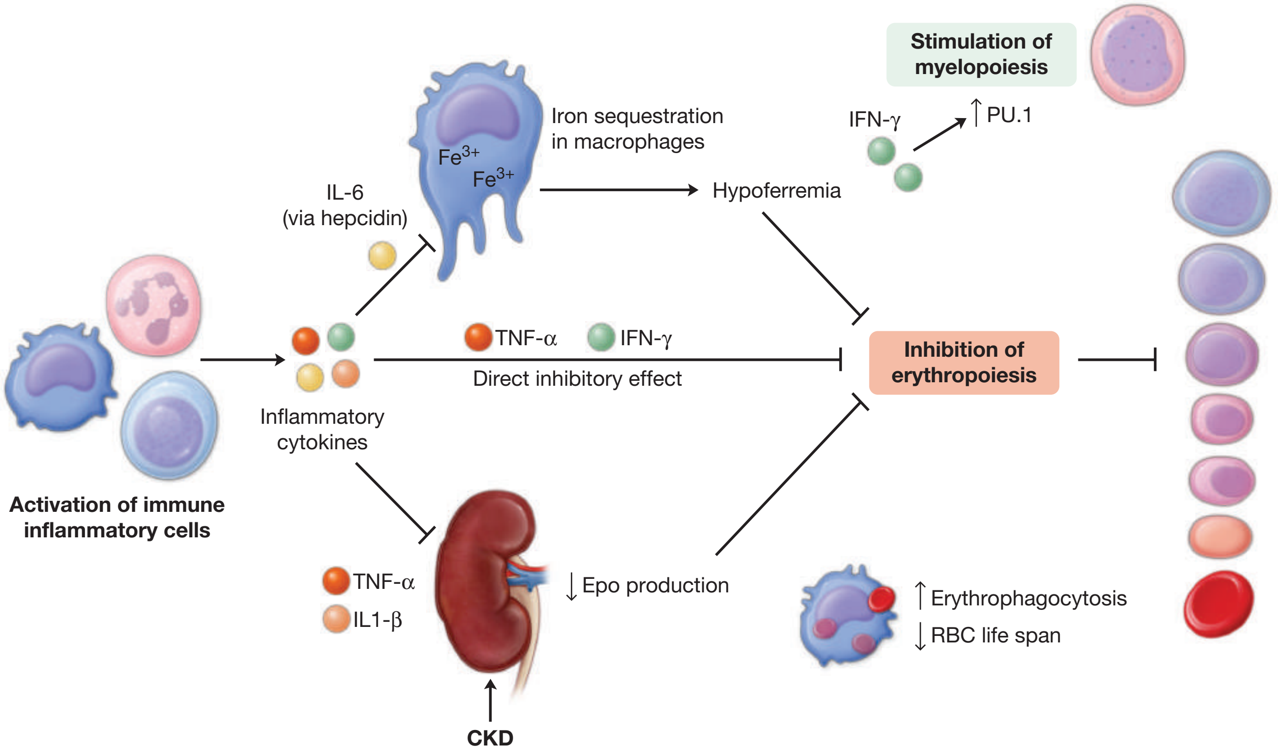

FIGURE 102-3 Pathophysiology of anemia of inflammation. Activation of immune cells leads to the production of inflammatory cytokines. IL-6 induces macrophage iron sequestration via hepcidin, tumor necrosis factor α (TNF-α) and interferon γ (IFN-γ) reduce erythropoietin (EPO) production and erythropoietic response; activation of PU.1 transcript induces myelopoiesis. Hemolysis of red cells phagocytosed by macrophages variably contribute to the pathophysiology of anemia, especially in acute inflammation.

and TNF-α. Moreover, bone marrow progenitor commitment is reprogrammed toward the myeloid lineage through increased activity of PU.1 and other transcription factors. In acute inflammation, immature white blood cells may be released to the circulation, the so-called "emergency myelopoiesis," as observed in severe COVID-19. A mild decrease of erythrocyte life span contributes to anemia. Iron sequestration in stores and hypoferremia are central to anemia of inflammation, emphasizing that it is a disorder of iron distribution. The contribution of the other mechanisms varies in different conditions. EPO deficiency is key in CKD. Massive red cell destruction is the main driver of anemia only in hyperinflammation, such as Still's disease and hemophagocytic lymphohistiocytosis. Mechanistically, hepcidin hyperexpression and hypoferremia are part of the innate immune response, aimed at depriving iron from and blocking the growth of the invading pathogens. However, in the long term and in noninfectious disorders, hypoferremia becomes a maladaptive response that leads to anemia.

CLINICAL FINDINGS AND DIAGNOSIS

Anemia of inflammation is mild to moderate, normocytic, and normochromic with low reticulocyte count. It can become microcytic when longstanding, due to iron-restricted erythropoiesis. It can be asymptomatic or contribute to disease burden, worsening the prognosis, as in CHF and CKD. A biochemical hallmark of inflammation is hyperferritinemia (Table 102-3). Ferritin, especially its heavy chain, is an acute phase reactant; thus, it cannot be used as an iron marker in inflammation. The combination of hypoferremia and hyperferritinemia in inflammatory disorders defines "functional" iron deficiency, indicating insufficient erythroid iron notwithstanding the presence of adequate stores. The boundaries between absolute and functional iron deficiency are not unequivocal, as shown by inflammatory bowel diseases, colon cancer, CKD, and CHF. Detecting true iron deficiency in these settings is challenging as both a universal definition and biomarkers are lacking. Ferritin <100 μg/L is the most accepted diagnostic threshold. In CHF and CKD, absolute iron deficiency is also defined by transferrin saturation <20%, irrespective of ferritin levels. A proposed diagnostic tool is the soluble transferrin receptor (sTfR)/log ferritin ratio (Table 102-3). The clinical usefulness of other laboratory parameters needs confirmation.

In CKD, anemia commonly develops when >50% of kidney function is lost and progresses in parallel with the reduction of glomerular filtration rate. It is a negative prognostic factor, predisposing to cardiovascular complications, hospitalization, and mortality. EPO deficiency, systemic inflammation, and negative iron balance are key contributors to the complex pathophysiology of anemia in CKD. Impaired iron absorption, chronic bleeding due to platelet dysfunction, and blood loss in the dialysis apparatus are causes of absolute iron deficiency, whereas increased production and decreased clearance of hepcidin result in functional iron deficiency. Deficiency of folate and vitamin B_{12} and a hemolytic component in uremic patients may contribute to anemia.

In CHF patients, the prevalence of anemia and iron deficiency is reported to be as high as 50–60% with negative prognostic impact. Besides mild inflammation, mechanisms include malnutrition, reduced iron absorption, and blood loss caused by the use of antithrombotic drugs. Current guidelines recommend periodic screening for iron deficiency in CHF patients. Supplementing intravenous iron may improve symptoms and quality of life, although influence on mortality is not proven.

Anemia is common in cancer, especially in hematologic malignancies such as myelodysplasia and multiple myeloma. It is multifactorial as it may be associated with iron and vitamin deficiency, blood loss, and chemotherapy side effects. Moderate to severe anemia is frequent in critically ill patients and is associated with unfavorable outcomes such as prolonged hospitalization and even mortality. Nearly 50% of patients admitted to an intensive care unit (ICU) have a hemoglobin level <9 g/dL, and most discharged survivors have persistent anemia. Anemia in ICU reflects anemia of inflammation, with the addition of blood losses due to blood sampling (up to 40 mL/d) for laboratory monitoring.

A mild anemia is common in elite endurance athletes, especially females. It is multifactorial and due to reduced iron absorption, iron losses, and increased hepcidin caused by both mild inflammation and indirect activation of the hepatic SMAD pathway by lactic acid.

TREATMENT

Treatment of the underlying disorder controls anemia of inflammation, as shown by the effect of anti-IL-6 agents in Castleman's disease and rheumatoid arthritis. Standard treatment of anemia in CKD is based on ESAs **(Chap. 322)** with a target hemoglobin established at <11.5 g/dL to avoid cardiovascular complications. Since iron deficiency leads to ESA hyporesponsiveness, iron supplementation is usually required, mostly parenterally, because in advanced CKD high hepcidin levels prevent iron absorption. Intravenous iron is administered at high or refracted doses, ideally to maintain transferrin saturation at 30–40% and ferritin between 200 and 700 μg/L. Iron supplementation may lower the dose of ESA required to maintain the hemoglobin target. An oral alternative to intravenous iron is ferric citrate, which, binding intestinal phosphate, provides iron while reducing phosphate levels. Prolyl hydroxylase inhibitors that stabilize HIF-2α have been approved in some countries with the aim of replacing with a single oral drug both ESA and iron, since HIF upregulates EPO production as well as duodenal iron absorption. Although these novel antianemic drugs are efficacious, safety concerns still exist for their long-term use.

Intravenous iron can be used in cancer-related anemia to reduce the need of transfusions, especially with ferritin <100 μg/L and before elective surgery. ESAs are used in low-risk myelodysplastic syndromes and in patients receiving chemotherapy notably when cancer is deemed uncurable. Intravenous iron supplementation is currently recommended in CHF patients with either ferritin <100 μg/L or transferrin saturation <20% with ferritin of 100–300 μg/L. Nonetheless, there is still some uncertainty as to whether this represents the best definition of iron deficiency in CHF. Novel drugs targeting the hepcidin-ferroportin axis to counteract iron maldistribution are under investigation.

Blood transfusions remain the therapeutic option for severe anemia in the ICU, with guidelines increasingly recommending restrictive hemoglobin thresholds (7–8 g/dL) for hemodynamically stable patients. Intravenous iron is proposed to minimize the use of transfusions, with some studies reporting improved clinical outcomes, although the risk of infections and hypophosphatemia are of concern in these fragile patients.

ANEMIA OF AGING

Prevalence of anemia in the elderly is ~10% at the age of 65 years, 25% in people >85 years, and 50% in those affected by multimorbidity, hospitalized, or institutionalized. Anemia is typically mild (hemoglobin 11–12 g/dL) and multifactorial, and its impact is frequently underestimated. Mounting evidence strengthens its negative influence on quality of life, muscle weakness, and the risk of falls and fractures, as well as its independent association with mortality. The adverse effects of anemia are seen in black patients at hemoglobin levels ~1 g/dL lower than in white patients.

Approximately one-third of the cases are estimated to be due to nutritional causes, especially iron deficiency; one-third are due to inflammation, including CKD; and one-third remain without a detectable origin and are defined "unexplained anemia of aging." Age-related decline in testosterone levels in males, stem cell exhaustion, and EPO reduction might contribute. Absolute iron deficiency is common, due to poor dietary iron intake, impaired absorption by chronic gastritis, or the use of proton pump inhibitors and blood losses. A gastrointestinal malignancy should always be ruled out in anemic elderly patients, but chronic bleeding may also result from nonmalignant intestinal lesions such as angiodysplasias and the use of antithrombotic drugs for common disorders such as atrial fibrillation. A recent study has shown that 100 mg of daily aspirin given for primary cardiovascular prevention in elderly subjects was associated with a 20% higher risk of anemia and iron deficiency as compared to placebo. The proposed ferritin cutoff for iron deficiency (45–70 μg/L) is higher than in younger patients. Treatment of iron deficiency may mitigate the negative prognostic implications of anemia.

Functional iron deficiency due to overt or subclinical inflammation often contributes, complicating the laboratory diagnosis. Traditionally "bone marrow senescence," now known as clonal hematopoiesis of indeterminate potential (CHIP), is present in >10% of otherwise-healthy individuals aged 70 years and increases sharply to >60% in the eighties. CHIP is associated with a chronic low-grade inflammatory status that worsens anemia and cardiovascular risk. Patients should be followed up for the risk of developing myelodysplastic syndromes **(Chap. 107)**. The unexplained anemia of aging is usually characterized by an EPO level that is lower than expected for the degree of anemia. EPO can increase the hemoglobin level, but it is unclear whether that change has an influence on the hospitalization, frailty, and mortality risks associated with the anemia.

FURTHER READING

Camaschella C: Iron deficiency. Blood 133:30, 2019.

Cleland JFG et al: Redefining both iron deficiency and anaemia in cardiovascular disease. Eur Heart J 44:1992, 2023.

Galy B et al: Mechanisms controlling cellular and systemic iron homeostasis. Nat Rev Mol Cell Biol 25:133, 2024.

Ganz T: Anemia of Inflammation. N Engl J Med 381:1148, 2019.

GBD 2021 Anaemia Collaborators: Prevalence, years lived with disability, and trends in anaemia burden by severity and cause, 1990-2021: Findings from the Global Burden of Disease Study 2021. Lancet Haematol 10:e713, 2023.

Macdougall I: Anaemia in CKD: Treatment standard. Nephrol Dial Transplant 26:770, 2024.

Oyedeji CI et al: How I treat anemia in older adults. Blood 143:205, 2024.

Pasricha SR et al: Iron deficiency. Lancet 397:233, 2021.

Patel KV et al: Haemoglobin concentration and the risk of death in older adults: Differences by race/ethnicity in the NHANES III follow-up. Br J Haematol 145:514, 2009.

Van Doren L, Auerbach M: IV iron formulations and use in adults. Hematol Am Soc Hematol Educ Program 1:622, 2023.

Weyand AC et al: Prevalence of iron deficiency and iron-deficiency anemia in US females aged 12-21 years, 2003-2020. JAMA 329:2191, 2023.

103

Disorders of Hemoglobin

Vijay G. Sankaran, Martin H. Steinberg

Hemoglobinopathies result from changes in the amino acid sequence of globin; in thalassemia, synthesis of normal globin is insufficient. Together the disorders of hemoglobin compose the most common Mendelian genetic diseases. In addition, they are responsible for most instances of hemolytic anemia. Sickle cell disease and the hemoglobin E (HbE)–associated syndromes are the most prevalent hemoglobinopathies; β and α thalassemia are the most prevalent thalassemias. In addition to these common disorders of hemoglobin, rare globin mutations cause hemoglobin instability, cause increased or decreased affinity of hemoglobin for oxygen (O_2), and allow spontaneous oxidation of hemoglobin, reducing its O_2 transport. O_2 transport can also be reduced by exposure to carbon monoxide (CO) and extrinsic oxidizing agents **(Table 103-1)**.

HEMOGLOBIN

Easy access to erythrocytes to study hemoglobin structure and function, reticulocytes to examine hemoglobin biosynthesis, and leukocyte DNA to define the mutations of hemoglobin, with the availability of hematopoietic stem and progenitor cells from blood and bone marrow, has placed hemoglobin disorders in the forefront of molecular medicine. The biology of hemoglobin provides the background for understanding the pathophysiology of its many genetic and acquired disorders and approaches to their treatment.

TABLE 103-1 Disorders of Hemoglobin

- I. **Hemoglobinopathies**—hemoglobin variants with amino acid sequence variants that alter the physical, chemical, or functional properties of hemoglobin
  - A. Common variants with unusual properties
    - 1. HbS—polymerization
    - 2. HbE—reduced biosynthesis
    - 3. HbC—hemoglobin-membrane interaction
  - B. Altered oxygen affinity
    - 1. High affinity—erythrocytosis
    - 2. Low affinity—cyanosis, anemia
  - C. Hemoglobins that oxidize readily
    - 1. Unstable hemoglobins—hemolytic anemia, jaundice
    - 2. M hemoglobins—methemoglobinemia, cyanosis
- II. **Thalassemias**—defective biosynthesis of globin chains
  - A. α Thalassemias
  - B. β Thalassemias
  - C. Complex thalassemias
- III. **Hereditary persistence of fetal hemoglobin**—persistence of higher than normal levels of HbF into adult life
  - A. Deletions within the *HBB* cluster—15–30% HbF in heterozygotes, pancellular HbF
  - B. Point mutations in *HBG2/1* promoters—5–30% HbF in heterozygotes; pancellular or heterocellular HbF
- IV. **Acquired hemoglobinopathies**
  - A. Methemoglobin due to toxic exposures
  - B. Sulfhemoglobin due to toxic exposures
  - C. Carboxyhemoglobin
  - D. HbH in erythroleukemia
  - E. Elevated HbF in myelodysplasia

DEVELOPMENTAL BIOLOGY

Successive waves of erythropoiesis direct the synthesis of different hemoglobin molecules that result from sequential activation and silencing of globin genes (Fig. 103-1). Hemoglobin is a tetramer of two pairs of unlike globin polypeptide chains, each containing a tetrapyrrole heme group. O_2 binds to heme as erythrocytes traverse the lungs and is released in the tissues. Heme is nestled within a protective pocket of each globin subunit.

GLOBIN GENE CLUSTERS

Globin is encoded in two nonallelic gene clusters. The β-globin gene cluster is on chromosome 11; the α-globin gene cluster is on chromosome 16 (Fig. 103-1). Tetramers of α-like and β-like globins form embryonic, fetal, and adult hemoglobins. Fetal hemoglobin (HbF, $\alpha_2\gamma_2$) production begins at 6–8 weeks' gestation, peaks during mid-gestation, then falls to <1% of total hemoglobin during the first 6 months of extrauterine life. Adult hemoglobin A (HbA, $\alpha_2\beta_2$) production follows a pattern reciprocal to that of HbF. The hemoglobin composition of normal adults is >95% HbA, ~1% HbF, and 2–3% HbA_2 (α2δ2). HbF and HbA_2 are not functionally important in normal adults because of their low concentrations. Measuring their levels can provide helpful diagnostic clues for thalassemia and some hemoglobinopathies. Hemoglobin is subject to posttranslational modifications. Most important clinically is the nonenzymatic glycosylation of HbA forming the adduct HbA_{1c}, which is useful in the management of diabetes mellitus.

HEMOGLOBIN STRUCTURE

All globin polypeptides have similar but not identical primary structures. α-Globins contain 141 amino acids, and β-like globins have 146 amino acids. This primary structure dictates, according to the constraints of protein folding, the secondary structure of globin into α-helical sections joined by small nonhelical stretches. Each globin chain folds into a tertiary conformation known as the globin fold, whereby charged amino acid residues face the exterior of the molecules while uncharged residues face the hydrophobic interior. The iron-containing tetrapyrrole heme moiety is protected from oxidation and located between two of the helical segments; O_2 loading and unloading occur when heme iron is in its reduced ferrous form. Globin gene mutations affecting critical heme-binding amino acid residues allow iron to be oxidized, forming methemoglobin, which has high O_2 affinity and does not release O_2 in tissues. Dimers of α- and non-α-globin chains reversibly assemble into the tetrameric quaternary structure.

HEMOGLOBIN FUNCTION

Hemoglobin transports O_2 from lungs to tissues and carbon dioxide (CO_2) from tissues to lungs. As a nitrate reductase, it releases nitric oxide (NO) from nitrite to promote vasodilation. Oxygen binding is defined by the hemoglobin-O_2 dissociation curve. P_{50} is a point on this sigmoidal curve that indicates the partial pressure of O_2 where hemoglobin is half saturated (Fig. 103-2). Normal P_{50} is ~26 mmHg; low P_{50} indicates that hemoglobin has high O_2 affinity, decreasing O_2 delivery

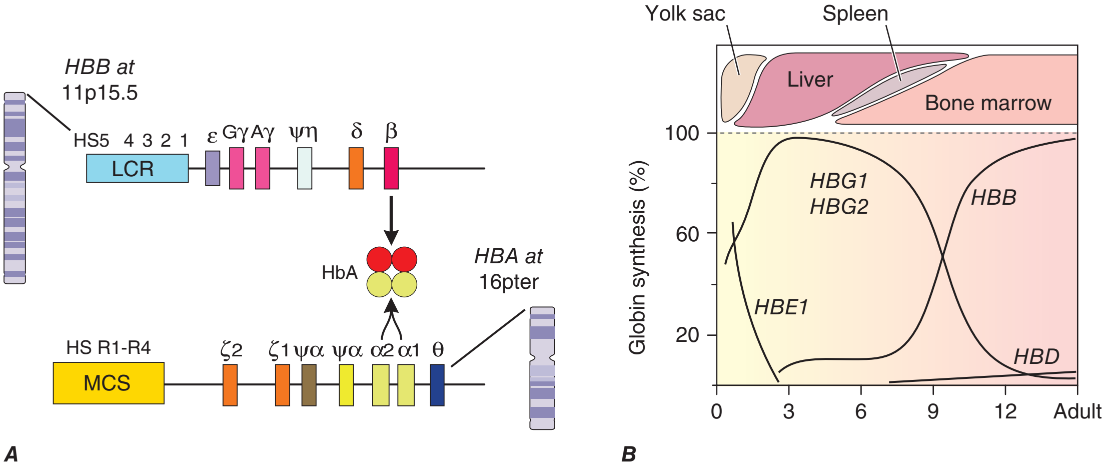

FIGURE 103-1 Globin gene clusters and their hemoglobin products during gestation. ***A.*** The order of globin genes in the β- and α-globin gene clusters along with their upstream enhancers, the β-globin locus control region (LCR) and α-globin multispecies conserved sequences (MCS), which contain critical regulatory elements. Normal hemoglobin tetramers contain two α-globin chains and two non-α-globin chains. In the example shown, this is adult HbA. The β-globin gene cluster contains an embryonic ε-globin gene (*HBE*), two nearly identical fetal γ-globin genes (*HBG2, HBG1*), a major adult β-globin gene (*HBB*), and a minor adult δ-globin gene (*HBD*). The α-globin gene cluster contains an embryonic ζ-globin gene (*HBZ*) and duplicated α-globin genes (*HBA2, HBA1*) coding for identical proteins. Embryonic hemoglobins include Gower I ($\zeta_2\varepsilon_2$), Gower II ($\alpha_2\varepsilon_2$), Portland I ($\zeta_2\gamma_2$), and Portland II ($\zeta_2\beta_2$). Fetal hemoglobin (HbF, $\alpha_2\gamma_2$) production begins at 6–8 weeks' gestation, peaks during mid-gestation, then falls to <1% of total hemoglobin during the first 6 months of extrauterine life. ***B.*** Sites of erythropoiesis and globin synthesized from the yolk sac and the early embryo (months 1–3), the fetus (months 3–9), after delivery (months 9–12), and afterward (adult).

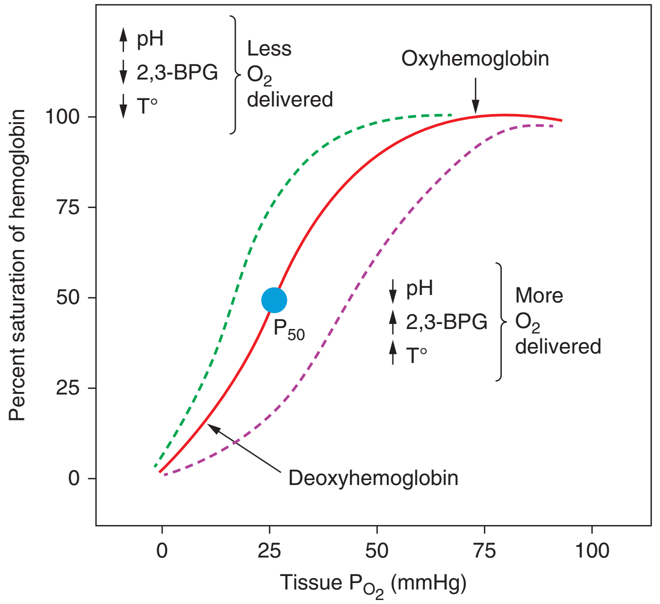

FIGURE 103-2 Hemoglobin-oxygen dissociation curve. The P_{50} is influenced by the binding of 2,3-bisphosphoglycerate (2,3-BPG), a product of glycolysis, in the central cavity of hemoglobin, pH, and temperature. The hemoglobin tetramer can bind up to four molecules of oxygen (O_2) in the iron-containing sites of the heme molecules. As O_2 is bound, 2,3-BPG and carbon dioxide (CO_2) are expelled. Salt bridges are broken, and each of the globin molecules changes its conformation to facilitate O_2 binding. O_2 release to the tissues is the reverse process, with salt bridges being formed and 2,3-BPG and CO_2 bound. Deoxyhemoglobin does not bind O_2 efficiently until the cell returns to conditions of higher pH, the most important modulator of O_2 affinity (Bohr effect). When acid is produced in the tissues, the dissociation curve shifts to the right, facilitating O_2 release and CO_2 binding. Alkalosis has the opposite effect, reducing O_2 delivery.

to tissues; high P_{50} indicates that hemoglobin has low O_2 affinity, releasing more O_2 to tissues. The conformation of hemoglobin fully saturated with O_2 is known as the R or relaxed state; desaturated hemoglobin is in the T or tense state. The transition between T and R states occurs when two or three O_2 molecules are bound. Cooperativity describes the progressively more rapid binding of O_2 once the first molecule is bound. Hemoglobin variants that decrease P_{50} are characterized by isolated erythrocytosis as compensation for hypoxia; variants with increased P_{50} sometimes are accompanied by cyanosis and anemia as hemoglobin becomes unsaturated and O_2 delivery is enhanced. Mutations of residues critical for heme binding, R-T transitions, or tetramer stability cause hemoglobinopathies characterized by hemolytic anemia, methemoglobinemia, erythrocytosis and cyanosis.

GLOBIN GENE SWITCHING

The sequential activation and inactivation of globin genes during development is called “hemoglobin switching.” Transcription factors along with epigenetic elements, such as DNA and histone methyltransferases and demethylases, interact with enhancers “upstream” of the gene clusters that contact globin gene promoters, silencing the embryonic and fetal genes. Developmental factors such as RNA-binding factors and microRNAs also impact hemoglobin switching.

β-Globin Gene Switching HbF reactivation by drugs and gene therapy is a prime therapeutic goal for treating sickle cell disease and β thalassemia, meriting a discussion of the controls of HbF gene silencing. An upstream enhancer called the β-globin locus control region (LCR) binds erythroid-specific and ubiquitous transcription factors. The LCR interacts directly with globin gene promoters; transcription factors that silence and activate genes also interact with elements of the globin genes. Competition among the β-like genes for the LCR and autonomous silencing of the embryonic and fetal globin genes depends on transcription factors. Silencing, first of the embryonic gene *HBE* and then of the two fetal genes, *HBG2* and *HBG1*, favors the interaction of the LCR with *HBB* allowing its expression (Fig. 103-1). The transcription factors BCL11A (2p16) and ZBTB7A (19p13) are the major repressors of HbF gene expression. BCL11A, a zinc finger protein that represses HbF genes, binds TGACCA motifs, the most important at position −115 in the promoter of each γ-globin gene. ZBTB7A binds 85 nucleotides upstream of these BCL11A binding sites. Mutations in these binding sites abolish the normal silencing of the HbF genes, leading to one type of the benign condition called hereditary persistence of fetal hemoglobin (HPFH). When binding of either BCL11A or ZBTB7A is disrupted, *HBG2* and *HBG1* are derepressed. *BCL11A* single nucleotide variants (SNVs) are common and are thought to underlie a large portion of the interindividual variation in HbF levels. Disruption of the *BCL11A* regulatory elements or the promotor binding sites for BCL11A and other repressive factors by gene editing in patient hematopoietic stem cells using CRISPR/Cas leads to 30–50% HbF with possible “cure” of sickle cell disease and β thalassemia, with the former strategy receiving U.S. Food and Drug Administration (FDA) approval for both conditions.

α-Globin Gene Switching A less complex switch takes place in the α-globin gene cluster. A regulatory locus of four elements termed R1–R4 is present within introns of the gene *NPRL3* that is upstream of *HBA2*. R1–R4 are critical for α-globin gene expression, as demonstrated by natural deletions causing thalassemia. A developmental switch from embryonic ζ- to adult α-globin gene expression occurs at about 6 weeks’ gestation.

Modulation of HbF Level and Haplotypes of the β-Globin Gene Cluster Variations in three quantitative trait loci (QTL), *BCL11A*, *MYB* (6q23), and a locus linked to the *HBB* cluster (11p15), are associated with HbF variation among normal individuals and patients with sickle cell anemia and β thalassemia. The *MYB* gene is essential for hematopoiesis and erythroid differentiation. *MYB* inhibits HbF expression directly by activation of *KLF1* and other repressors and indirectly through alteration of the kinetics of erythroid differentiation. The third QTL is marked by a common variant 158 nucleotides upstream of the transcription start site of *HBG2*. Five common haplotypes associated with the *HBB* cluster have been defined by its SNVs. Sickle cell anemia patients with the Senegal and Arab-Indian HbS gene-associated haplotypes have the common −158 C-T variant in the *HBG2* promoter. They have higher HbF levels than patients with Benin, Bantu, and Cameroon haplotypes. When young, they might have a milder clinical course.

GENERAL ASPECTS OF HEMOGLOBIN DIAGNOSIS

α-Globin gene mutations are expressed in the embryo and fetus and persist throughout life; HbF mutations are expressed in the fetus and in the first months of life, vanishing from notice afterward; δ-globin gene mutations are innocuous and usually not detected; β-globin gene mutations can become clinically apparent after the synthesis of HbF dwindles to stable adult levels.

With rare exceptions, all disorders of hemoglobin are autosomal recessive or co-dominant disorders; a family history, usually of anemia, a common feature of most symptomatic hemoglobinopathies and thalassemias, is often present. In addition to pallor and jaundice, splenomegaly is often present. A small number of laboratory tests can confirm the diagnosis starting with a complete blood count that includes a reticulocyte count with a careful review of a peripheral blood film. A sustained increase in reticulocyte count indicates the presence of hemolytic anemia. Hemoglobin fractionation by high-performance liquid chromatography (HPLC) or capillary electrophoresis, especially when, in addition to the index case, family members are available for study, is often sufficient to confirm a diagnosis at the level of hemoglobin phenotype. DNA sequencing of the globin genes allows definitive diagnosis; available from excellent reference laboratories, it is a prerequisite for genetic counseling.

Sickle cell disease and β thalassemia are chronic hemolytic anemias sharing hemolysis-related complications like venous thrombosis, leg ulcers, and pulmonary hypertension. Differences are that only deoxyHbS polymerizes, while ineffective erythropoiesis is the key pathophysiologic feature of β thalassemia. Both diseases are "cured" by successful allogeneic hematopoietic stem cell transplantation and gene therapy as discussed below.

SICKLE CELL DISEASE

Sickle cell disease is a clinical and hematologic phenotype caused by an assortment of genotypes **(Table 103-2)**. Sickle cell anemia, defined as homozygosity for the sickle hemoglobin mutation ($\alpha_2\beta^S{}_2$; glutamic acid [E] 7 valine [V] GAG-G*T*G), is the most common of these genotypes, followed by HbSC disease or compound heterozygosity for HbS and HbC ($\alpha_2\beta^C{}_2$; E 7 lysine [K] GAG-*A*AG) genes. Many different thalassemia mutations contribute to the HbS-β thalassemias. Compound heterozygous genotypes are less common than HbS homozygotes. HbS has been described in compound heterozygotes with many other variant hemoglobins. Few of these genotypes, other than HbSO^{Arab}, HbSE, and HbSD^{Punjab} are symptomatic.

ORIGIN, SPREAD, AND EPIDEMIOLOGY

HbS originated in Africa between 7000 and 22,000 years ago, reaching high frequencies because of the increased genetic fitness of heterozygotes under selective pressure from *Plasmodium falciparum*. HbS gene haplotypes have a loose association with the severity of disease because each haplotype has a different average level of HbF. In some regions of Africa, India, and the Middle East, nearly half the population have sickle cell trait. Nigeria alone has ~150,000 newborns each year with sickle cell anemia, about one-third of the world's total newborns; most die before age 5. Coerced and free population movement has spread the HbS gene throughout the world. The HbS carrier, or sickle cell trait, prevalence is 2–15% in emigrant populations; ~100,000 patients in the United States have sickle cell disease. In the United States, death in childhood is rare; the median age of death in patients with sickle cell anemia is in the fifth or sixth decade.

PATHOPHYSIOLOGY

Pathophysiologic features of sickle cell disease are summarized in **Fig. 103-3**. HbS is physiologically like HbA in most respects except it polymerizes when deoxygenated. Contacts between one of the β^7 valine residues of deoxyHbS and specific amino acid residues of β- and α-globin culminate in fascicles of hemoglobin that injure the sickle erythrocyte. A delay occurs between the initiation of polymerization and the accumulation of sufficient polymer to damage the cell. It is unclear how much polymer is needed for cell injury, but polymer leads directly and indirectly to the multiple abnormalities of the sickle erythrocyte that generate the pathophysiology of disease. Prominent among these abnormalities are HbS polymer penetration of the membrane causing vesiculation with membrane microparticle release; increased activity of the Gardos, K/CL cotransport, and P_{sickle} channels that dehydrate the cell, increasing mean corpuscular sickle hemoglobin concentration (MC[HbS]C), reducing cellular deformability, and increasing the polymerization potential of HbS; translocation of amino phospholipids such as phosphatidylserine to the outer leaflet of the membrane; and oxidation of erythrocyte contents. These and other abnormalities lead to the formation of irreversibly sickled cells (ISCs), which are sickle erythrocytes that are forever deformed because of permanent membrane damage regardless of whether HbS remains polymerized. Damaged sickle erythrocytes are responsible for initiating the vasoocclusive, hemolytic, and inflammatory features of the disease shown in Fig. 103-3.

DIAGNOSIS

Although sickle cell disease can appear in any ethnic group, most often it is present in people of African, Middle Eastern, Mediterranean, and Indian descent. The chief presenting symptom is pain. This might be an arthritis-like hand-foot syndrome in young children or the typical acute painful episode in older children and adults. In HbSC disease and HbS-β^+ thalassemia, acute vasoocclusive episodes occur at about half the rate as in sickle cell anemia while complications develop later; rarely, patients with these genotypes are asymptomatic. The key elements of laboratory diagnosis are outlined in Table 103-2 showing typical hematologic findings and hemoglobin fractions. **Figure 103-4** displays HPLC profiles and blood films in typical patients with sickle cell trait, sickle cell anemia, and HbSC disease. Clinical and basic laboratory diagnosis is sufficient for general management and counseling; genetic counseling and family planning usually require DNA-based diagnosis.

TABLE 103-2 Common Sickle Hemoglobinopathies

GENOTYPE	CLINICAL ABNORMALITIES	HEMOGLOBIN LEVEL, g/L (g/dL)/MCV, fL	HEMOGLOBIN FRACTIONS (%)
Sickle cell trait (HbAS)	8% of African Americans; hematuria, papillary necrosis, hyposthenuria, increased incidence of chronic kidney disease; 2–4 times increased VTE risk;? stroke; splenic infarction at altitude; rhabdomyolysis	Normal	HbA: 60–70 HbS: 30–40 Percent HbS dependent on presence or absence of α thalassemia
Sickle cell anemia (HbSS)	Vasoocclusion related: pain, acute chest syndrome, osteonecrosis, splenic infarction Hemolysis related: stroke, pulmonary and systemic vasculopathy, nephropathy, leg ulceration gallstones, priapism	70–100 (7–10)/80–100	HbS: >75 HbF: 2–25 HbA_2: 3–4
HbS-β^0 thalassemia	Rate of complications similar to HbSS	80–100 (8–11)/60–85	HbS: >75 HbF: 2–15 HbA_2: 5–6
HbS-β^+ thalassemia	Rate of complications about half the rate of HbSS depending on percent HbA	100–140 (10–14)/70–80	HbS: 60–90 HbA: 5–40 HbF: 1–10 HbA_2: 5–6
Hemoglobin SC disease (HbSC)	Nearly asymptomatic to disease as severe as HbSS; about half the rate of complications as HbSS. Increased risk of retinopathy	100–140 (10–14)/70–100	HbS: 50 HbC: 50
HbSE	Resembles clinically HbS-β^+ thalassemia; symptoms delayed; often Asian/Indian ancestry	90–130 (9–13)/65–75	HbS: 65 HbE: 35 HbF: 1–5
HbSS-α thalassemia	Present in 30% of HbSS; phenocopies HbS-β^0 thalassemia because of microcytosis and high HbA_2; like HbSS but with fewer strokes and leg ulcers and less pulmonary vascular and renal disease	80–100 (8–11)/60–85	HbS: >75 HbF: 2–15 HbA_2: 4–5

Note: Laboratory values are averages in untreated adults.

Abbreviation: VTE, venous thromboembolism.

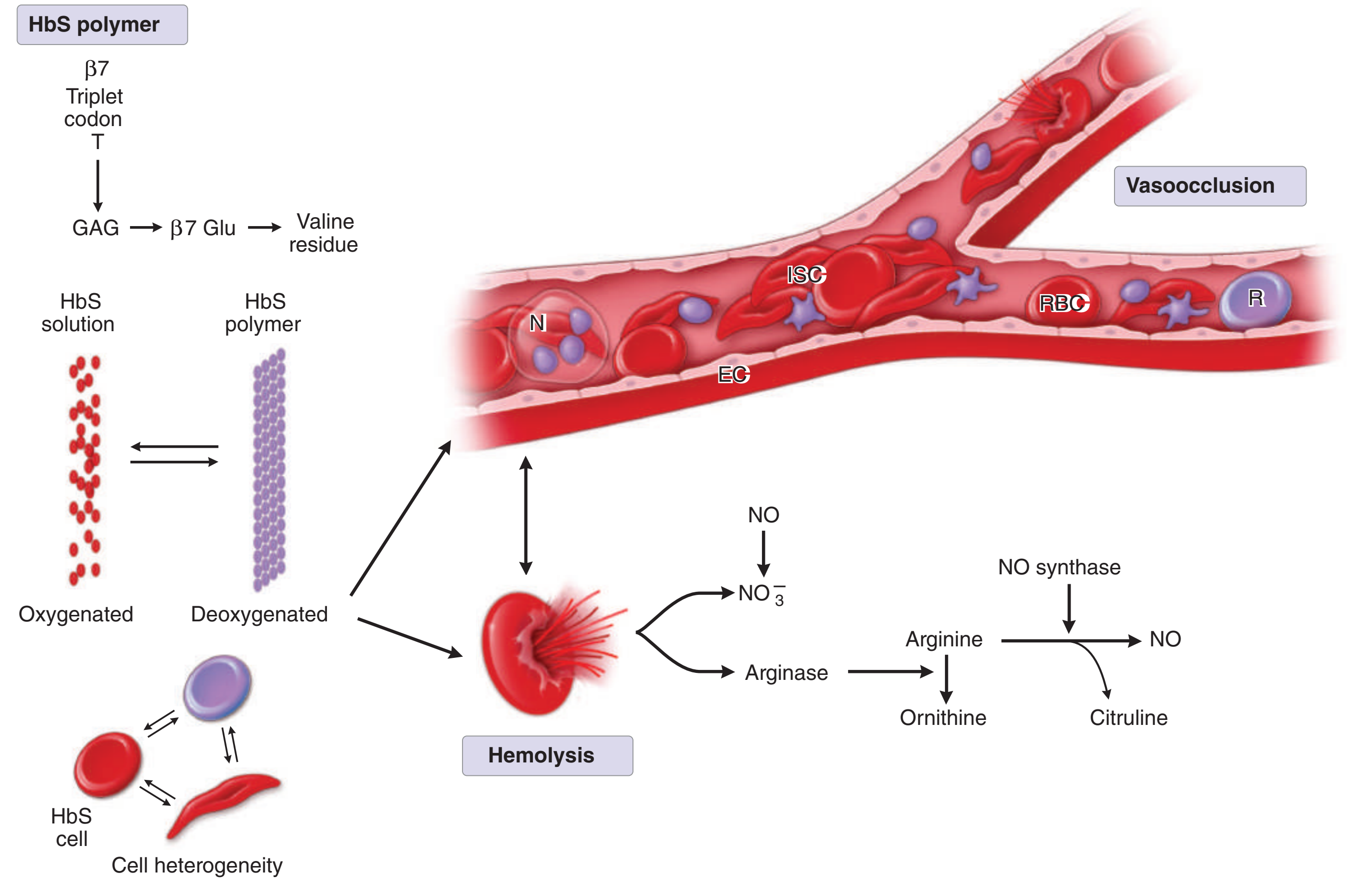

FIGURE 103-3 Pathophysiology of sickle cell disease. HbS is in solution when oxygenated but reversibly polymerizes when deoxygenated. Polymerization is dependent on the 30th power of hemoglobin concentration. In the sickle cell, this means that small changes in hemoglobin concentration or cell hydration can have large effects on polymerization. Polymerization begins seconds to minutes following deoxygenation. Erythrocyte deformation, or sickling, is initially reversible, but after an undetermined number of cell sickling events, the cell becomes irreversibly deformed. These are known as irreversibly sickled cells (ISCs). Their membrane is permanently damaged, although depending on their oxygen (O_2) content, HbS could be in solution. Sickle erythrocytes lead to the clinical and laboratory phenotypes of disease. Sickle cells interact with endothelial cells and other blood cells, occluding flow in small and sometimes large vessels and causing the many complications thought to be a result of vasoocclusion. Sickle cells also live <20 days (normal ~120 days) hemolyzing intra- and extravascularly. Intravascular hemolysis depletes haptoglobin and hemopexin while liberating heme, arginase, and other danger-associated molecular patterns (DAMPs) into the blood. This scavenges nitric oxide (NO), activates platelets and endothelium, reduces antioxidant activity, causes vasoconstriction, and is proinflammatory.

COMPLICATIONS

Complications of sickle cell disease can be grouped into those that likely are a consequence of the related entities of sickle vasoocclusion and those due to intravascular hemolysis (Fig. 103-3). Complications associated with vasoocclusion seem to respond best to induction of HbF. Some complications of disease are presented in Table 103-3.

Acute Painful Episodes Characterized by unprovoked severe pain in extremities or the torso that is often symmetrical and stereotypical for each patient and that usually requires treatment with strong opioids in the emergency department, acute painful episodes are the most common acute events in sickle cell disease. They are the chief cause of concern for patients, most of whom have them at some time in their life. Their frequency varies; most patients have one to two episodes a year; some rarely have them; others are hardly ever without them. Acute painful episodes last days to weeks. Pain in sickle cell disease can also be chronic from osteonecrosis, osteoporosis, or leg ulcers. Chronic and acute pain can overlap. Pain can also be induced by the opioids. Most of the time, patients have some degree of pain that does not reach the intensity of the acute episode. This can be treated with oral opioids dispensed monthly.

No diagnostic test can confirm or refute the presence of an acute pain episode whose cause is uncommonly identified. Physical examination is not often useful diagnostically. Some patients will have pain on pressure over an affected area, perhaps accompanied by swelling; mild fever is common. Often a 1–2 g/dL decrease in hemoglobin level and a modest increase in the leukocyte count are noted. The presence of ISCs and the reticulocyte count are of no diagnostic value. Drastic decreases in hemoglobin and platelet levels with more extreme leukocytosis can portend development of severe acute chest syndrome or multiorgan failure.

Some patients die suddenly shortly after admission for an acute painful episode. The cause of this sudden unexpected death is usually unknown; among the possibilities are arrhythmias and pulmonary embolism. Admitting patients to monitored beds or using continuous pulse oximetry for the first 48–72 h of hospitalization might prevent some of these deaths and help identify early acute chest syndrome that follows within 72 h in about a quarter of admissions for acute pain. After searching for possible precipitants such as infection or dehydration and treating these appropriately, the foundation of treatment is the proper dosing of opioid analgesics. By the time a patient presents at the emergency department or clinic requesting treatment, they have usually tried nonsteroidal anti-inflammatory drugs (NSAIDs) and oral opioids. In most patients, relief of pain requires intravenous opioids. Many patients are opioid tolerant, requiring higher than usual doses for satisfactory relief. Dosing should not be on an “as-needed” schedule; patient-controlled analgesia or frequent fixed doses of opioids with rescue doses for breakthrough pain are the preferred means of treatment, with frequent assessments to ensure pain relief without excessive sedation. Adjunctive treatment includes incentive spirometry to forestall pulmonary complications, maintaining hydration with half-normal saline with care not to overhydrate, prophylaxis for thromboembolism, and antihistamines and laxatives to counter expected side effects of opioids. Unless hypoxia is present, supplemental O_2 is unnecessary. NSAIDs have little value in patients receiving intravenous opioids.

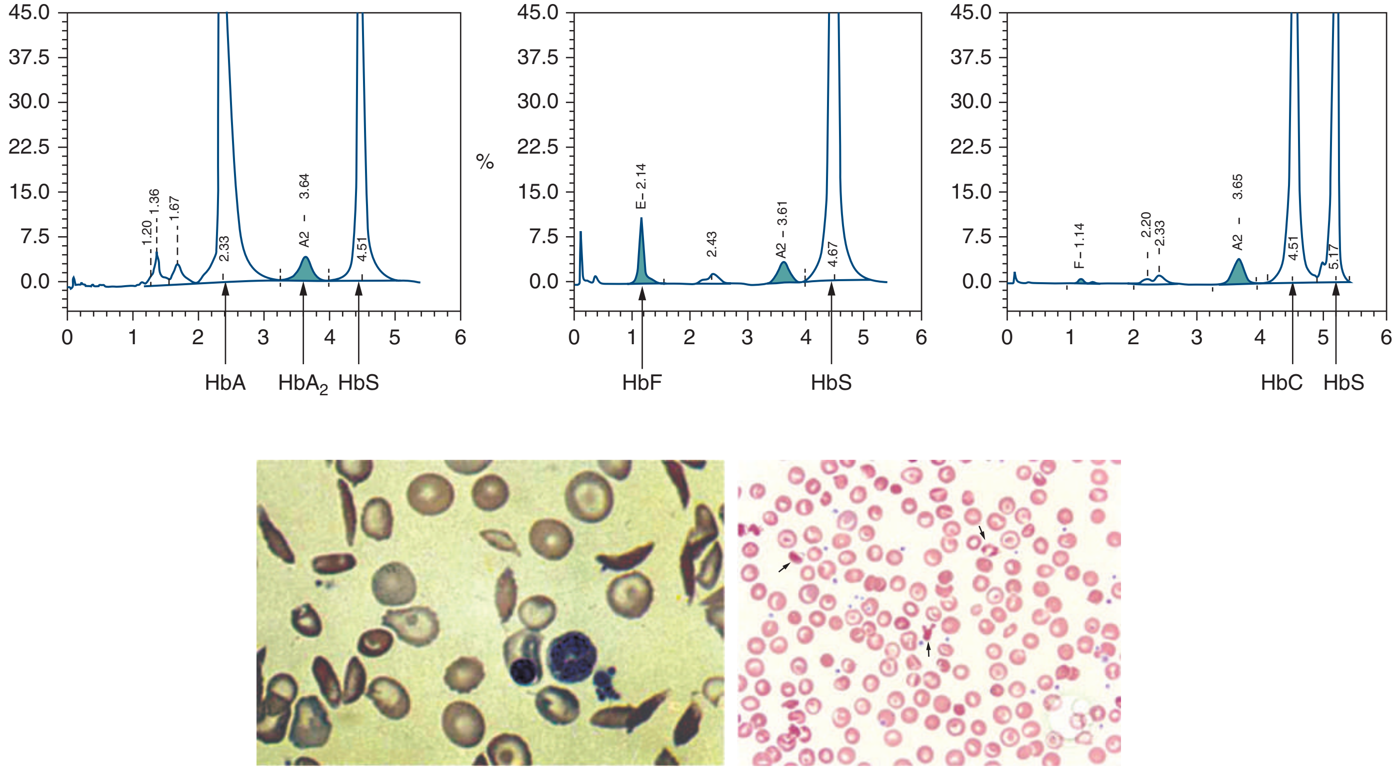

FIGURE 103-4 Diagnosis of sickle cell disease. ***A.*** From *left to right*, high-performance liquid chromatography separation in sickle cell trait, sickle cell anemia, and HbSC disease. Beneath each chromatogram, the individual protein peaks are identified. ***B.*** *Left:* Dense, elongated, and pointed cells are the irreversibly sickled cells characteristic of the sickle cell anemia and sickle cell-β^0 thalassemia. Target cells and nucleated red cells are also present. *Right:* Target cells, cells with squared ends of HbC crystals, cells folded like tacos, and contracted microspherocytes are typical of HbSC disease. *(Source: B [right]: Reproduced with permission from American Society of Hematology.)*

TABLE 103-3 Complications of Sickle Cell Disease

COMPLICATION	INCIDENCE, DIAGNOSIS, AND FEATURES	TREATMENT
Priapism	~30% of males; can be episodic and short duration (stuttering); severe episodes can cause impotence; associated with markers of hemolysis	Many therapies including α-adrenergic agonists, stilbesterol; consult urology for treatment, which is time-critical
Stroke and silent infarction	10–15% of all cases; infarction in early childhood into adulthood; hemorrhagic in adults; neurocognitive abnormalities in adults even without apparent stroke; associated with markers of hemolysis	Transcranial Doppler screening in children ages 2–16; transfusion for at-risk patients; hydroxyurea
Gallstones/surgery	~40% of patients; bilirubin levels and stones related to polymorphisms of *UGT1A*; in surgery requiring general anesthesia, simple preoperative transfusion to a hemoglobin of 10 g/dL is recommended	If asymptomatic, usually let be; otherwise, laparoscopic cholecystectomy
Hepatic disease	>80% of patients have hepatomegaly; intrahepatic cholestasis can have bilirubin ~100 mg/dL; viral hepatitis, iron overload, RBC sequestration, extrahepatic cholestasis also contribute	Exchange transfusion for intrahepatic cholestasis; transplant for end-stage liver failure
Nephropathy	~30% of adults age >30 years; hyperfiltration in children, renal failure in adults; early albuminuria, later nephrotic-range proteinuria; associated with markers of hemolysis	Screen for microalbuminuria by age 10 years; avoid NSAIDs; use ACE inhibitors or receptor antagonists for albuminuria; erythropoietin for symptomatic anemia; dialysis or transplant for renal failure
Lung/pulmonary hypertension	Restrictive disease; asthma common; 5–10% have pulmonary hypertension by right heart catheterization; 30% have increased TRV that portends poor prognosis; associated with markers of hemolysis	Consult expert pulmonologist; screen yearly by echocardiography measurement of TRV
Retinopathy	30% in HbSC disease, 3% in HbSS[a]; develops in peripheral retina; vitreous hemorrhage and retinal detachment can cause blindness	Screen annually starting at age 10 tears with fluorescein angiography; laser photocoagulation for proliferative disease
Acute anemic episodes	B19 parvovirus infection, folic acid deficiency, splenic sequestration, delayed hemolytic transfusion reaction with destruction of transfused and sometimes autologous red cells	RBC transfusion if symptomatic; splenectomy if more than 1 or 2 episodes of sequestration; anti-parvovirus IgM positive in acute infection, IgG in past infection
Multiorgan failure	Can accompany severe acute chest syndrome; often confused with sepsis and can coexist with sepsis; CNS liver, muscle, lung, kidney affected	Exchange transfusion, ICU support
Pregnancy	Screening both partners for hemoglobin disorders with risk counseling is critical component of family planning	All pregnancies are "high risk"; transfuse if sickle cell events increase, if previous miscarriage, multiple fetuses

[a]Sickle cell anemia (HbSS).

Abbreviations: ACE, angiotensin-converting enzyme; CNS, central nervous system; ICU, intensive care unit; NSAIDs, nonsteroidal anti-inflammatory drugs; RBC, red blood cell; TRV, tricuspid regurgitant jet velocity.

Acute Chest Syndrome This pneumonia-like illness is the second most frequent acute sickle cell–related event. It occurs in >50% of patients, often more than once. Acute chest syndrome can be mild, especially in children, in whom it can result from viral infection, or devastating, where multiple lobes of the lung are affected with severe hypoxia, multiorgan failure, and death. Chest pain, cough, fever, and hypoxia and a pulmonary infiltrate on chest x-ray are the major diagnostic criteria. The etiology includes in situ thrombosis, emboli, any type of infection, and postoperative hypoventilation. Management in adults is dictated by the severity of the episode and the need for supplemental oxygen. Patients who are hypoxic and febrile can be admitted directly to the intensive care unit. Antibiotics are almost always used in these patients even though a causative bacterium is not often cultured. Supplemental O_2 is given for an O_2 saturation <95%. Overhydration and excessive opioids can compound dyspnea and hypoxia. Hypoxic patients who are febrile with leukocytosis with more than a trivial infiltrate on x-ray are transfused. In the severely ill patient, exchange transfusion, if possible, is the preferred modality. However, if transfusion of the severely ill patient is indicated and hours are needed to arrange red cell exchange, simple or top-up transfusion should be started first. Simple transfusions also suffice for less severely affected patients. Most patients survive acute chest syndrome, but in the most severe cases, often caused by embolization of necrotic bone marrow, death can be rapid even with prompt and proper treatment. Thrombocytopenia, leukocyte counts >20,000/dL, and rapidly developing acute anemia often portend severe acute chest syndrome with its possibility of acute respiratory distress syndrome and multiorgan failure. Asthma is very common in patients with sickle cell disease. Some adults have chronic lung disease with reduced diffusing capacity for CO that could be a sequela of acute chest syndrome.

Osteonecrosis This painful and sometimes crippling complication that most often affects hips bilaterally occurs in about half of all patients with sickle cell anemia and is also common in HbSC disease; shoulders are less often affected. Beginning with chronic pain that can become severe, loss of function, especially in the hips, is often the final stage. Magnetic resonance imaging (MRI) can detect the earliest stages, whereas x-ray is less sensitive. Physical therapy and NSAIDs provide some initial relief; oral opioids are sometimes required. Joint replacement can restore lost mobility while relieving pain. Life span of prosthetic joints is finite, so surgery should be delayed as long as mobility is satisfactory and pain tolerable.

Leg Ulcers The incidence of leg ulcers is highly dependent on geography and hemoglobin genotype. They are far less common in HbSC disease and HbS-β^+ thalassemia than in sickle cell anemia and HbS-β^0 thalassemia. In temperate climates, 10–20% of patients are affected; tropical and subtropical areas have an incidence rate up to 75%; ulcers rarely occur in patients from the Middle East. Leg ulcers can be small and superficial or deep and encompass most of the lower leg. They can be extraordinarily painful. Long-standing, recurrent large ulcers are difficult to treat. Wet-to-dry dressings and Unna boots are reasonable choices for initial treatment.

SICKLE CELL TRAIT (CARRIERS, OR SIMPLE HETEROZYGOSITY FOR THE HBS GENE)

Carriers of sickle cell trait outnumber patients with the disease by 25 to 1. Although testing for sickle cell disease is part of most perinatal cord blood screening programs, counseling and follow-up of detected carriers are imperfect, so adolescents and adults can be unaware they carry sickle cell trait. Counseling carriers about the complications of sickle cell trait and their likelihood of having offspring with sickle cell disease is essential. Carriers should be counseled prior to participation in sports because of the risk, albeit small, of sudden death from heat-related exertional rhabdomyolysis. Optimal hydration before and during exercise can prevent most episodes of heat-related illness. Usually a benign condition with a normal life expectancy, some complications are shown in Table 103-2.

TREATMENT, SCREENING, COUNSELING, AND ANTENATAL DIAGNOSIS

Patients should, if possible, be referred to a sickle cell center for initial consultation, institution of therapy, and follow-up. Cooperation among primary care providers, hematologists, and other specialists provides the best preventive care and management of complications. The frequency at which a patient is seen depends on their therapeutic regimen and complications.

Remarkable changes in the treatment landscape have occurred with the promise of even greater benefits from new curative approaches. The following discussion focuses on treatment to prevent the complications of disease.

Hydroxyurea Hydroxyurea is the standard of care for all patients with sickle cell anemia and HbS-β^0 thalassemia regardless of symptoms. Although in some symptomatic patients with HbSC disease, its benefits in this genotype are understudied. The major mechanism of action of hydroxyurea is to induce high levels of HbF. Hydroxyurea increases HbF unevenly among red cells (heterocellularly), so some cells have greater protection from HbS polymerization than others. When started in adults, where the average baseline HbF is ~5%, HbF increases to ~10%. Nevertheless, pain and acute chest syndrome are reduced by about half, hemoglobin concentration increases by ~1 g/dL, and after 17.5 years of follow-up, mortality was reduced by 49%. In contrast, all young children respond robustly to hydroxyurea. When started at <1 year of age at a dose of ~27 mg/kg, HbF levels were 33.3 ± 9.1% and hemoglobin concentration was 10.1 ± 1.3 g/dL. Acute events were markedly reduced with little toxicity. Based on these and other studies in high- and low-resource countries, unless there is a contraindication, hydroxyurea should be given to all infants with sickle cell anemia and HbS-β^0 thalassemia starting at 9 months of age at a dose of ~20 mg/kg and titrated to the maximal tolerated dose based on neutrophil and platelet counts.

Voxelotor Voxelotor increases the affinity of the hemoglobin molecule for O_2. At a dose of 1500 mg daily, hemoglobin concentration increased ~1 g/dL in 59% of patients with a reduction in the biomarkers of hemolysis. Its effects on acute vasoocclusive events are unclear. Many questions remain about the long-term effects of voxelotor. Less hemolysis reduces the propensity for stroke, nephropathy, pulmonary hypertension, leg ulcers, and priapism. Will voxelotor be accompanied by these long-term benefits? Could the high O_2 affinity of a modified hemoglobin be harmful for some patients? The answers to these important questions require further study.

Crizanlizumab Downstream effects of HbS polymerization include adhesive interactions among endothelial cells, leukocytes, platelets, and erythrocytes. P-selectin is one molecule involved in these interactions; blocking selectins prevents sickle cell–endothelial cell adhesion. Crizanlizumab, a P-selectin-blocking monoclonal antibody given intravenously every month, reduced acute painful episodes by ~45%. Hemolysis was unaffected. A follow-up trial failed to replicate the results of the rigorous study that led to FDA approval.

L-Glutamine The mechanism of action of this agent, presumed to be the reduction of oxidative stress in sickle erythrocytes, is unsettled. A phase 3 placebo-controlled trial showed that L-glutamine was associated with a 25% reduction in painful episodes and 33% reduction in hospitalization.

There is little consensus regarding how these recently approved drugs should be integrated into treatment with hydroxyurea. The effects of voxelotor and crizanlizumab appear to be additive to those of hydroxyurea. Voxelotor can be added to hydroxyurea if the benefits of hydroxyurea alone are insufficient, as they are in most adults. If both hydroxyurea and voxelotor are taken at effective doses and acute vasoocclusive complications continue, crizanlizumab might then be added. The dropout rates in the crizanlizumab and L-glutamine trials were ~35%, so adherence to these therapeutics could be problematic.

Transfusion Transfusions are overutilized and underutilized. Major indications for transfusion include severe symptomatic anemia; treatment and prevention of stroke; increasing hemoglobin level to ~10 g/dL before surgery requiring general anesthesia; and severe acute chest syndrome. Sometimes transfusions are given during pregnancy when there is a history of complications or fetal loss. Transfusions should usually be avoided in acute pain episodes and for repair of stable chronic anemia. Automated red cell exchange transfusion is preferred in acute stroke, severe acute chest syndrome, or multiorgan failure or when chronic transfusions are planned. Expert guidelines recommended extended red cell antigen profiling, if possible, before the first transfusion and antigen matching for Rh (C, E or C/c, E/e) and K antigens in addition to ABO/RhD. Complications of transfusion include hyperviscosity, alloimmunization (which occurred in 18.6% of patients transfused between 1979 and 1984 and 27.3% of patients transfused between 2001 and 2011), iron overload, and delayed hemolytic transfusion reactions with hyperhemolysis.

Stem Cell Transplantation Given the excellent results of human leukocyte antigen (HLA)–identical related donor transplants, which have an event-free survival of >95%, this option might be extended to all patients with a suitable donor. Unfortunately, only 15% of patients have a fully matched donor. New approaches to haploidentical transplants are improving event-free survival in these patients.

Gene Therapy Two forms of ex vivo gene therapy are approved for sickle cell disease. Both use mobilized autologous CD34+ stem cells and increase levels of a hemoglobin that inhibits HbS polymerization. In one approach, an HbA gene containing the β^{T87Q} mutation responsible for the antipolymerization effects of HbF is introduced into stem cells via a lentiviral vector. In the second, the major enhancer of the HbF repressor, *BCL11A*, is disrupted using CRISPR/Cas9 gene editing. In both approaches, following myeloablative conditioning, engineered cells are reinfused and engraft. Both treatments have resulted in near pancellular distribution of 30–50% HbF or HbA^{T87Q}, reduced hemolysis, and total hemoglobin concentrations of >12 g/dL, with nearly total prevention of acute vasoocclusive events. Long-term safety and cure rate will take many more years of follow-up to establish.

Preventive Measures and Screening Cord blood screening for sickle cell disease is done in many countries and all 50 states. Affected patients are then directed to clinics that can initiate early preventive care. In childhood, transcranial Doppler screening beginning at age 2 years and repeated annually until age 16 years, prophylactic penicillin (125 mg for children younger than 3 years; 250 mg for children 3 years and older) twice daily until age 5 years, and vaccination with pneumococcal vaccines are the main measures to prevent stroke and invasive pneumococcal infection. Folic acid, 1 mg daily, is given to prevent megaloblastic erythropoiesis; it is probably unnecessary in people with nutritious diets.

All women planning pregnancy should be screened for disorders of hemoglobin by blood counts, erythrocyte indices, and HPLC analysis of hemoglobin. Individuals with HbS or β thalassemia trait should have their partners tested. Only then is it possible to know the risks of a fetus having sickle cell disease (Table 103-2). Antenatal diagnosis using chorionic villus sampling or cell free DNA testing is widely available.

THALASSEMIA

Thalassemia is caused by reduced accumulation of either α- or β-globin chains causing a relative excess of the unaffected chain. Unbalanced globin synthesis is the hallmark of thalassemia and the proximate cause of its pathophysiology; unpaired globin chains damage the developing erythroblast. Like the HbS mutation and many other red cell traits, thalassemia reached polymorphic levels in tropical and subtropical populations because heterozygotes are protected from severe forms of *P. falciparum* infection. Estimates are that 1–5% of the world's population carries a thalassemia mutation; in some locales, most people have a thalassemia mutation. These mutations can affect any globin gene, but clinically, β and α thalassemia are the most important. With nearly 500 unique thalassemia-causing mutations (*www.globin.bx.psu.edu*) that can interact with each other and with hemoglobinopathies, thalassemia syndromes are remarkably diverse. Where resources permit and the mutation is known, genetic counseling can be provided and antenatal diagnosis is possible.

HbE (β^{27} glu-lys) is a common variant whose biosynthesis is reduced because the site of the mutation alters its mRNA processing. Its reduced biosynthesis leads to a deficit of β^E-globin chains and features of β thalassemia. Hemoglobin Constant Spring is caused by a mutation of the termination codon of *HBA2* that leads to the synthesis of an elongated α-globin chain that is unstable and suboptimally synthesized. This variant therefore behaves as an α thalassemia variant.

β THALASSEMIA

EPIDEMIOLOGY

Once known as Mediterranean anemia, because of the concentration of cases in Italy, Greece, and other countries bordering the Mediterranean Sea, or as Cooley's anemia after the physician first describing cases, β thalassemia is common in most areas of the world where malaria was endemic, including the Mediterranean region, Asia, and the Middle East. Effective programs of screening, counseling, and antenatal diagnosis have reduced the birth of new cases in a number of regions. About 40,000 β thalassemia patients are born yearly. In the United States, there are ~1000 cases of severe β thalassemia.

CLASSIFICATION

β^0 Thalassemia mutations totally prevent the accumulation of any globin from the affected gene; β^+ thalassemia mutations cause minor or moderate reductions in β-globin synthesis. β Thalassemia major and β thalassemia intermedia are now categorized as transfusion-dependent and non-transfusion-dependent based on the number and frequency of transfusions required to sustain a good quality of life.

Pathophysiology Single nucleotide changes are the most common β thalassemia mutations, but gene deletions also occur. A partial listing of the classes of mutations causing β thalassemia include mutations in the promoter elements affecting gene transcription causing mild and sometimes silent β^+ thalassemia; mutations in the junctions between exons and introns that affect mRNA processing causing β^0 and β^+ thalassemia; introduction of alternative splice sites into introns or exons usually causing β^+ thalassemia; 3′ end-processing sequence mutations preventing RNA polyadenylation leading to mild or silent β^+ thalassemia; mutations preventing initiation of translation causing β^0 thalassemia; and introduction of stop codons that prematurely terminate translation (nonsense mutations) producing reading frameshifts and resulting in truncated globin mRNA and β^0 thalassemia. In addition, rare causes of β thalassemia have been identified that are unlinked from the β globin locus and caused by mutations in general transcription regulators, such as SUPT5H and TFIIH, or erythroid transcription factors like GATA1.

In β thalassemia, the deficit in β-globin chain synthesis allows α-globin chains to accumulate in excess. Without a non-α-globin chain partner in dimer and tetramer formation, unpaired α-globin chains are unstable, cannot form a tetramer, and precipitate within the developing erythroblast, causing membrane lipid oxidation and damage. The predominant cause of anemia is intramedullary destruction of erythroid precursors, known as ineffective erythropoiesis. Reduced deformability and phosphatidyl serine exposure also cause extra- and intravascular hemolysis of those erythrocytes that gain entrance into the circulation. In poorly treated β thalassemia, severe anemia leads to bone marrow expansion; hepatosplenomegaly; iron accumulation in liver, heart, and endocrine organs; pulmonary hypertension; and thromboembolic disease.

Frightening pictures of children with severe β thalassemia permeate the literature. These examples of near-terminal disease should be relegated to history because treatment with transfusion and iron chelation can prevent their occurrence, hematopoietic stem cell transplantation can "cure" patients who have suitable donors, and efficacious gene therapies are now approved.

DIAGNOSIS

Heterozygous β thalassemia, also known as β thalassemia trait and β thalassemia minor, has mild or no anemia but microcytic/hypochromic erythrocytes with minimal or no increase in reticulocyte count. After recognizing these hematologic abnormalities and excluding iron deficiency, finding an elevated level of HbA_2 and perhaps HbF by HPLC is sufficient to establish this diagnosis. The hematologic characteristics of this heterozygous carrier state are listed in **Table 103-4**. Sometimes, the spleen is enlarged. Before genetic counseling and antenatal diagnosis are considered after carrier identification by red cell indices and quantitation of HbA_2, the thalassemia-causing mutation should be identified. Sequencing is the key to preventing homozygotes or compound heterozygotes with transfusion-dependent thalassemia.

The more severe forms of β thalassemia are hemolytic anemias with hypochromia, microcytosis, reticulocytosis, marked anisocytosis, and poikilocytosis with variable numbers of circulating nucleated red cells **(Fig. 103-5)**.

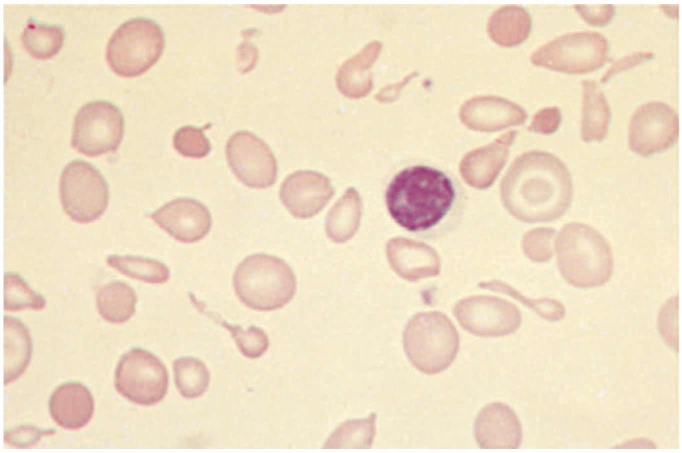

FIGURE 103-5 α **Thalassemia intermedia.** Target cells and marked variation in cell size and shape but with general hypochromia and microcytosis characterize the blood film. A lymphocyte is shown for size comparison.

COMPLICATIONS

Complications of severe β thalassemia are many. They are a consequence of chronic hemolytic anemia, chronic transfusion, and iron loading. Increased iron absorption is especially common in non-transfusion-dependent thalassemia. Most complications, listed in **Table 103-5**, develop because of either inadequate blood transfusion and/or poor iron chelation and iron loading. Even when chelation is optimized, some complications attributable to iron toxicity will develop. Many complications have complex and multifactorial etiologies. Iron stores are estimated by serum ferritin levels; MRI is the most widespread means of noninvasively measuring iron accumulation in liver and heart.

MANAGEMENT, SCREENING, COUNSELING, AND ANTENATAL DIAGNOSIS

Heterozygote screening and counseling couples at risk for affected fetuses, with antenatal diagnosis, if needed, is an effective preventive approach. Severe thalassemia should be dealt with in specialized centers where these and other services are available and managed by a team led by a hematologist experienced with this disease with help from endocrinologists, cardiologists, transfusion medicine specialists, and social services.

Transfusion and Iron Chelation Transfusion every 2–4 weeks with a goal pretransfusion hemoglobin concentration of 9–10.5 g/dL to suppress ineffective erythropoiesis, coupled with iron chelation to prevent the accumulation of excess toxic iron that accompanies transfusion, has prevented the development of cardiomyopathy and endocrinopathies while extending life to at least 50 years. When to begin transfusions, whether partial exchange transfusion is preferable to simple transfusion, and the choice of blood product require consultation with experts. To be effective, transfusions and iron chelation must be started early, be uninterrupted, and continue lifelong. Older patients who did not have the advantage of effective chelation are more likely to develop multiple disease-related morbidities such as osteoporosis, endocrinopathies, liver disease, and renal failure. Two orally effective chelating agents, deferasirox and deferiprone, and one intravenous chelator, deferoxamine, are available.

Improving Ineffective Erythropoiesis Luspatercept is a fusion protein containing the extracellular domain of human activin type IIB receptor and the Fc domain of human IgG. By binding transforming growth factor β superfamily ligands and reducing Smad2/3 signaling, luspatercept enhances late-stage erythropoiesis. Given subcutaneously, 1 mg/kg every 3 weeks, it was associated with a 33% reduction in transfusion requirements.

TABLE 103-4 β Thalassemias

CLASSIFICATION	HEMOGLOBIN (g/dL)/MCV (fL)	HEMOGLOBIN FRACTIONS (%)	CLINICAL FEATURES
β-Thalassemia trait	100–140 (10–14)/60–80	HbA: 94 HbF:1–2 HbA_2: 4–6	Heterozygosity for β^+ or β^0 thalassemia mutations; "silent" carriers can have normal HbA_2 and red cell indices.
Non-transfusion-dependent β thalassemia (thalassemia intermedia)	70–120 (7–12)/65–80	HbA: 60–90 HbF: 10–40 HbA_2: 4–6	Defined by infrequent or no transfusion requirement; caused by many different genotypes including homozygosity for "mild" β^+ mutations, combinations of β and α thalassemia, homozygous β thalassemia with high HbF-producing capacity, and many others. Iron loading, thromboembolic disease, and pulmonary hypertension are major clinical events.
Transfusion-dependent β thalassemia (Thalassemia major)	20–40 (2–4)/50–80	HbA: 0–5 HbF: 90–100 HbA_2: 2–5	Caused by many different genotypes including homozygosity and compound heterozygosity for β^0 and β^+ mutations, combinations of β and α thalassemia; transplantation curative; iron chelation required.
HbE-β thalassemia	50–80 (5–8)/60–70	HbE: 50–70 HbF: 30–50	Common in Southeast Asian populations; in some parts of the world, the most prevalent severe thalassemia; in HbE-β^0 thalassemia, only HbE and HbF are found; in HbE-β^+ thalassemia, HbA is present. Transfusion dependence depends in part on the thalassemia mutation.
δβ Thalassemia and Hb Lepore	110–120 (11–12)/65–75	HbA: 70 HbF: 7–13 HbA_2: 2	Rare; deletions removing the δ- and β-globin genes cause δβ thalassemia; Lepore hemoglobins are fusion globin chains; values are for heterozygotes; homozygotes have 100% HbF with hemoglobin 10–11 g/dL.
Gene deletion hereditary persistence of fetal hemoglobin (HPFH)	120–140 (12–14)/75–85	HbA: 70 HbF: 15–30 HbA_2: 2	Rare; large deletions removing the δ- and β-globin genes; values are for heterozygotes; homozygotes, who are asymptomatic, have 100% HbF without anemia.

Note: Laboratory results are averages in adults.

TABLE 103-5 Complications of β Thalassemia

COMPLICATION	INCIDENCE, DIAGNOSIS, AND FEATURES
Growth retardation	Most often a feature of delayed or inadequate transfusions but can occur in well-transfused children.
Delayed puberty; secondary amenorrhea	50% and 25%, respectively.
Splenomegaly	Can trap 1–40% of red blood cell volume; increases plasma volume, worsening heart failure. Splenectomy indicated when transfusion requirement to maintain ideal hemoglobin increases. Prophylactic penicillin after splenectomy.
Heart	Due to chronic anemia, heightened sensitivity to iron toxicity, thromboembolic pulmonary hypertension, other causes. Progresses through stages to congestive failure and arrhythmias. Assessed by T2* on magnetic resonance imaging (MRI). The available chelating agents might have differential effects on different measure of cardiac function and can be used in combination.
Leg ulcers	Common in thalassemia intermedia.
Hepatic disease	Fibrosis progressing to cirrhosis is related to hepatic iron concentration that can be monitored by MRI. Hepatitis also plays a role.
Lung disease/ pulmonary hypertension	Fibrosis, chronic thromboembolic disease, restrictive pathophysiology, intravascular hemolysis, and reduced nitric oxide bioavailability.
Thromboembolism	Multifactorial etiology, including platelet activation, red cell–endothelial interactions, thrombocytosis, endothelial activation, splenectomy.
Endocrinopathies	Diabetes, hypothyroidism, hypoparathyroidism, adrenal insufficiency; hypogonadism; hypothalamic-pituitary axis might be especially sensitive to iron.
Bone disease	Caused by bone marrow expansion, severe iron loading, hypogonadism; osteoporosis in ~50% of patients, even those well treated. Extramedullary hematopoietic masses are a feature of thalassemia intermedia.
Infections	Transfusion associated; linked to iron overload (*Yersinia*); malaria.

Hematopoietic Stem Cell Transplantation There is consensus that patients with available donors should be offered transplantation because of the difficulty of lifelong transfusion and chelation and its imperfect efficacy. Quality of life in successfully transplanted patients exceeds that in patients treated with transfusion and chelation. Transplantation from matched sibling donors is curative in >80% of all cases. Unfortunately, only a third of patients have matched donors. The best results are in the youngest patients who have been effectively chelated and received fewer transfusions. Graft failure, graft rejection, graft-versus-host disease, and a mortality of 5–20% depending on risk factors are the major drawbacks of this procedure. Results of haploidentical and unrelated donor transplants are improving but lag those of matched sibling donors.

Gene Therapy The same gene therapy approaches approved for sickle cell disease are approved for transfusion-dependent β thalassemia (see Sickle Cell Disease). CRISPR/Cas editing to downregulate *BCL11A* has resulted in increases in total hemoglobin ≥12 g/dL and HbF ≥10 g/dL, leading to transfusion independence in >90% of 52 patients aged between 12 and 35 years with transfusion-dependent β thalassemia. Results of lentiviral-mediated HbA^{T87Q} additive gene therapy were best with non-β^0/β^0 genotypes, although some individuals with β^0/β^0 genotypes could be effectively treated as higher viral titers were used in subsequent gene therapy protocols.

α THALASSEMIA

In some respects, the obverse of β thalassemia, clinically consequential α thalassemia is less common than severe β thalassemia. α Thalassemia is most often found in Asian populations and is usually caused by deletion of α-globin genes rather than point mutations.

EPIDEMIOLOGY

Carriers of the most common α thalassemia chromosomes (Table 103-6) are found in 5–80% of people from tropical and subtropical regions of Africa, the Middle East, Asia, and Melanesia. About 30% of African Americans carry the common $-\alpha^{3.7}$ chromosome that contains a single functional α-globin gene. HbH disease, the chief clinically important α thalassemia, is most prevalent in southern China and Southeast Asia. Estimates are that in Thailand ~3500 patients with severe α thalassemia are born yearly. Pregnancies affected by hemoglobin (Hb) Bart's hydrops fetalis occur mainly in southern China and southeastern Asia.

CLASSIFICATION

Each normal chromosome 16 contains two α-globin genes; normal diploid individuals have four α-globin genes. A classification of inherited α thalassemia, as summarized in Table 103-6, is based on the number of functional α-globin genes. If one or two α-globin genes are missing or poorly expressed, these people have α thalassemia

TABLE 103-6 α Thalassemias

CLASSIFICATION	α-GLOBIN GENE ARRANGEMENT	HEMOGLOBIN LEVEL, g/L (g/dL)/MCV (fL)	CLINICAL FEATURES
α-Thalassemia trait	–α/αα –α/–α – –/αα $\alpha^{T}\alpha/\alpha\alpha$	120–150 (12–15)/65–80	The chromosome with one deleted α gene (—α/) is called α^{+} thalassemia (α thalassemia-2); the chromosome with both deleted α genes is α^{0} thalassemia (α thalassemia-1); non–gene deletion α thalassemias (α^{T}) often have a more severe phenotype.
Hemoglobin H disease	– –/–α $\alpha^{T}\alpha/--$ $\alpha^{T}\alpha/\alpha^{T}\alpha$	50–120 (5–12)/60–70	Mild to moderate anemia depending on genotype; non–gene deletion forms of α thalassemia can produce severe HbH disease.
Hb Bart's hydrops fetalis	––/––		Fatal in utero or at birth with rare survivors. Hydrops can also result from combinations of gene deletion and non–gene deletion α thalassemia.
α Thalassemia/intellectual disability syndromes (ATR-16) (ATR-X)	– –/αα or – –/–α in ATR-16 αα/αα in ATR-X		ATR-16: Large deletions and rearrangements in chr16p. ATR-X: No α-globin gene deletion or mutation, *ATRX* mutations, X-linked.
α Thalassemia with myelodysplasia (ATMDS)	αα/αα		Mutations in *ATRX*; striking male predominance. Hematologic findings of HbH disease.

Note: Laboratory values are averages in adults. αα/denotes the chromosome with two intact α-globin genes; –α/chromosome with one α-globin gene deleted; – –/chromosome with both α-globin genes deleted; α^{T} represents non–gene deletion α thalassemia caused by point mutations. The –α/chromosome, referred to as α^{+} or α thalassemia-2, most often has a deletion of 3.7 kb of DNA ($-\alpha^{3.7}$) or 4.2 kb of DNA ($-\alpha^{4.2}$) that leaves a single α-globin gene intact. The chromosome where both α-globin genes are deleted (– –/) is called α^{0} thalassemia or α thalassemia-1. These chromosomes are caused by different-sized deletions that are usually named after their regions of highest frequency such as -SEA, -MED, -FI, and -THAI.

trait. Their hematologic abnormalities are almost always trivial. HbH disease is usually caused by deletion or malfunction of three α-globin genes. Hb Bart's hydrops fetalis fetuses have no normally functioning α-globin genes. Hundreds of different-sized deletions and rarer point mutations affect the production of α-globin and the magnitude of imbalanced globin synthesis. Because of this mutational complexity, many different variations of the common α thalassemia syndromes are found.

PATHOPHYSIOLOGY

Reduced accumulation of α-globin leaves non-α-globins unpaired and unable to participate in the formation of functional hemoglobin tetramers. In the fetus, absent or reduced synthesis of α-globin allows unpaired γ-globin chains, part of the HbF tetramer, to form γ_4 or Hb Bart's; in adults, when γ-globin synthesis is mostly silenced, unpaired β-globin chains, lacking a suitable partner to form HbA, tetramerize as β_4 or HbH. Both Hb Bart's and HbH have very high O_2 affinity and do not unload O_2 in tissues; HbH is also unstable. Severe anemia in Hb Bart's hydrops fetalis is a result of absent normal hemoglobin and ineffective erythropoiesis; in HbH disease, unstable HbH leads to oxidative membrane damage with extravascular hemolysis in the spleen and ineffective erythropoiesis.

DIAGNOSIS

Microcytosis/hypochromia with nearly normal hemoglobin concentrations, in the absence of iron deficiency and the increased level of HbA_2 that is diagnostic of β thalassemia, is sufficient for a presumptive diagnosis of α thalassemia trait. When genetic counseling is needed and antenatal diagnosis contemplated, the molecular basis of the presumed α thalassemia is required. HbH disease, which is usually due to compound heterozygosity for one chromosome with both α-globin genes deleted and one chromosome with only a single α-globin gene, is defined by the hematologic findings shown in Table 103-6 along with varying levels of reticulocytosis. At birth, when hemoglobin is separated by HPLC, 20–30% Hb Bart's is present; in adults, traces to 40% HbH are present along with residual Hb Bart's in some cases. HbH inclusions can be induced in some red cells after incubation and staining with brilliant cresyl blue. Hemoglobin composition in Hb Bart's hydrops fetalis is predominantly Hb Bart's with some Hb Portland if the deletion removing α-globin genes preserves the ζ-globin gene.

COMPLICATIONS

HbH disease is very heterogeneous because of the different combinations of genotypes that can cause this phenotype. Generally, when non–gene deletion mutants, such as Hb Constant Spring, contribute to the genotype, the disease is more severe, and intermittent or regular transfusions are necessary. In the most common – –/–α genotype, mean hemoglobin in adults is ~11 g/dL. Hepatosplenomegaly, jaundice, thalassemic bone changes in the face, and growth impairment are seen in 20–50% of cases, depending on the underlying genotype. Iron loading occurs but is not the severe problem it is in β thalassemia. Pregnancy in these patients should be considered high risk and managed accordingly. Mothers of infants with Hb Bart's hydrops fetalis have a history of stillbirth and develop preeclampsia, polyhydramnios, and antepartum hemorrhage and have difficult labor and delivery. Intrauterine transfusion of the fetus is possible.

MANAGEMENT, SCREENING, COUNSELING, AND ANTENATAL DIAGNOSIS

When planning families, couples from regions where α thalassemia is common who have red cell indices that suggest the possibility of carrying an α thalassemia gene should have genetic counseling based on DNA analysis of their globin genes. Iron should be avoided in non-iron-deficient individuals with α thalassemia trait and microcytosis. Transfusions are not usually needed in HbH disease. Nevertheless, depending on the genotype of disease, transfusions might be necessary especially when anemia becomes more severe, for example, with acute anemic episodes or pregnancy. Iron stores should be checked periodically by measuring serum ferritin or MRI; chelation does not appear to be needed.

Hb Bart's hydrops fetalis is best prevented by screening couples at risk and antenatal diagnosis. Intrauterine therapy and perinatal intensive care have permitted survival of some infants with Hb Bart's hydrops fetalis. As growth retardation affects ~40% and neurodevelopmental delay is present in 20% of survivors, prevention is the best approach.

OTHER HEMOGLOBINOPATHIES OF CLINICAL IMPORTANCE (TABLE 103-7)

More than 1500 mutations affecting hemoglobin structure have been described (*www.globin.bx.psu.edu*). Most are clinically silent. HbC and HbE are common. HbC is found in people of African descent and HbE in South China and Southeast Asia. Heterozygotes for HbC and HbE are unaffected clinically. Even individuals homozygous for these mutations, where the variant hemoglobin comprises >90% of the hemolysate, are clinically well with very mild anemia and microcytosis. The major importance of these variants is the interaction of HbC with HbS and HbE with β thalassemia, as outlined in Tables 103-2 and 103-4. A definitive diagnosis for all rare variants depends on DNA analysis.

TABLE 103-7 HbC, HbE, and Rare Hemoglobinopathies

CLASSIFICATION	CLINICAL ABNORMALITIES	HEMOGLOBIN LEVEL, g/L (g/dL)/MCV, fL	HEMOGLOBIN FRACTIONS (%)
HbC trait	2% of African Americans; target cells; no disease	Normal	HbC: 30–40 HbA_2: 2–3
HbC disease	Target cells; HbC crystals; mild reticulocytosis; splenomegaly	100–130 (10–13)/60–70	HbC: >95 HbF: 2–4 HbA_2: 2–3
HbE trait	50% incidence in some Asian populations; a few target cells; clinically normal	120–140 (12–14)/80–90	HbE: 27–31 HbF: 1 HbA_2: 3
HbE disease	No hemolysis; 20–80% target cells; no splenomegaly	100–120 (10–12)/65–75	HbE: 85–95 HbF: 3–7 HbA_2: 3
High O_2 affinity hemoglobins	Isolated erythrocytosis; often familial; no splenomegaly; no $JAK2^{V617F}$ mutation	150–200 (15–20)	Variants in α- and β-globin genes; patients are heterozygotes; ~25–50% variant
Low O_2 affinity hemoglobins	Asymptomatic mild anemia; cyanosis	100–140 (10–14)	~50% variant
Unstable hemoglobins	Pigmenturia; hemolysis; reticulocytosis; splenomegaly	90–140 (9–14)/70–90	20–35% variant; rare hyperunstable variants can be undetectable and have the phenotype of thalassemia
M hemoglobins	Some have mild hemolysis; few symptoms	100–140 (10–14)/80–90	20–50% variant depending on gene affected

Note: Laboratory values are averages in adults. As noted for HbAS, the amount of HbC and HbE in heterozygotes depends on the number of α-globin genes.

Unexpected low O_2 saturation by pulse oximetry (SpO_2) with normal O_2 saturation of arterial blood is occasionally seen in rare hemoglobin variants with clinical phenotypes. Asymptomatic patients with unexpectedly low SpO_2 should not be subjected to unneeded cardiopulmonary investigations in search of the cause of their "hypoxemia" until the existence of a hemoglobin variant is excluded.

M HEMOGLOBINS

M (met) hemoglobins are characterized by oxidation of the heme-iron from its ferrous (Fe^{++}) to ferric (Fe^{+++}) form. The major clinical feature of these disorders is asymptomatic cyanosis. Thirteen M hemoglobin variants have been described. In nine, the mutation involves histidine residues that interact with heme. Asymptomatic slate gray/brownish pseudocyanosis is the main clinical finding. Spectrophotometric recording of the visible spectrum of the hemolysate is diagnostic. To distinguish M hemoglobins from methemoglobinemia due to drugs or cytochrome b5 reductase (*CYB5R3*) deficiency, potassium cyanide (KCN) can be added to the hemolysate; methemoglobin-containing blood will turn red, but KCN has no effect on M hemoglobin. Treatment is not needed.

UNSTABLE HEMOGLOBINS

Sometimes referred to as congenital Heinz body hemolytic anemias, some mutations result in a hemoglobin tetramer that is unstable and precipitates intracellularly. One-hundred-fifty-six such variants have been described and are often a result of a new mutation that affects the tertiary or quaternary structure of the molecule. The most common class of mutations introduce a proline residue in the α helix or a polar amino acid into the interion of the molecule. Heinz bodies are intraerythrocytic precipitates that are detectable as dark globular aggregates after staining with a dye such as brilliant cresyl blue. Three unstable hemoglobins are the most common. Hemoglobin Köln (β^{99} val-met) has been found in multiple families, Hb Hasharon (α^{47} asp-his) is found in Ashkenazi Jews, and Hb Zurich (β^{63} his-arg) is susceptible to oxidant drug-induced hemolysis. Unstable variants present with nonspherocytic hemolytic anemia, but presentation is highly variable. The associated disease is usually mild and does not require transfusion. Heating blood to 50°C or incubation with isopropanol precipitates unstable hemoglobins but must be done with careful controls. Some variants can be detected by HPLC.

HEMOGLOBINS WITH HIGH OXYGEN AFFINITY AND LOW OXYGEN AFFINITY

Rare mutations in areas involved in the R-T transition, at critical interfaces between globin chains of the tetramer that reduce the affinity for 2,3-bisphosphoglycerate, or present in the heme pocket account for most of these variants. High O_2 affinity hemoglobins (103) outnumber low O_2 affinity variants (48). Isolated erythrocytosis in the absence of splenomegaly suggests the presence of a high O_2 affinity hemoglobin. High O_2 affinity hemoglobin variants shift the hemoglobin-O_2 dissociation curve leftward, causing a low P_{50} and thereby stimulating erythropoiesis. Many of these variants are due to new mutations. The clinical course is benign, and phlebotomy because of erythrocytosis is usually not required. Early diagnosis is important to forestall unnecessary diagnostic procedures and therapeutics such as cardiac catheterization to exclude congenital heart disease or treatment for polycythemia vera. Low O_2 affinity variants often present with cyanosis. Their hemoglobin-O_2 dissociation curve is right-shifted with high P_{50}. HPLC might reveal the presence of a hemoglobin variant. Treatment is often not necessary.

ACQUIRED DISORDERS OF HEMOGLOBIN

CO binds hemoglobin with high affinity forming carboxyhemoglobin. Carboxyhemoglobin levels can be accurately measured by co-oximetry of arterial blood. Standard pulse oximeters cannot accurately make this measurement. Some newly developed pulse oximeters can measure both carboxyhemoglobin and methemoglobin. Bound CO inhibits the transport of O_2; the hemoglobin-O_2 binding curve is left-shifted. Acute and chronic CO intoxication, caused by occupational exposure and other sources of incomplete combustion of hydrocarbons, presents with headache, altered mental status, and other constitutional symptoms. High-flow O_2 via facemask is the preferred treatment; criteria have been developed to guide the use of hyperbaric O_2.

Acquired methemoglobinemia and methemoglobinemia due to deficiency of *CYB5R3* are more common than the M hemoglobins. *CYB5R3* is required for the reduction of methemoglobin by NADH. Affected individuals with "toxic" methemoglobinemia can be cyanotic and symptomatic. As in carboxyhemoglobinemia, O_2 transport is reduced and reflected by the left shift in the hemoglobin-O_2 binding curve. *CYB5R3* deficiency usually affects only erythrocytes (type I), causing a mild disorder; when all cells are affected (type II), a severe disease results. Intravenous methylene blue is the preferred treatment in symptomatic patients with acquired methemoglobinemia and 40–60% methemoglobin. The usual dose is 1–2 mg/kg. Alternative treatment with ascorbic acid is preferable in people who are glucose-6-phosphate dehydrogenase deficient. Methylene blue interferes with co-oximetry, reducing the value of co-oximetry for monitoring treatment.

Many drugs and chemicals can induce methemoglobin in the absence of *CYB5R3* deficiency. Dapsone and topical anesthetics such as benzocaine are the most common offending agents.

FURTHER READING

Frangoul H et al: Exagamglogene autotemcel for severe sickle cell disease. N Engl J Med 290:1649, 2024.

Hardouon G et al: Sickle cell disease: From genetics to curative approaches. Ann Rev Genomics Hum Genet 24:255, 2023.

Leonard A et al: Gene therapy for hemoglobinopathies: Beta-thalassemia, sickle cell disease. Hematol Oncol Clin North Am 36:769, 2022.

Locatelli F et al: Defining curative endpoints for sickle cell disease in the era of gene therapy and gene editing. Am J Hematol 99:430, 2024.

Piel FB et al: Defining global strategies to improve outcomes in sickle cell disease: A Lancet Haematology Commission. Lancet Haematol. 10:e633, 2023.

Pinto VM et al: Management of the aging beta-thalassemia transfusion-dependent population: The Italian experience. Blood Rev 38:100594, 2019.

Pinto VM et al: Management of the sickle cell trait: An opinion by expert panel members. J Clin Med 12:3441, 2023.

Ribeil J-A et al: An integrated therapeutic approach to sickle cell disease management beyond infancy. Am J Hematol 98:1087, 2023.

Sheth S et al: Management of luspatercept therapy in patients with transfusion-dependent β-thalassaemia. Br J Haematol 201:824, 2023.

Taher AT et al: Beta-thalassemia. N Engl J Med 384:727, 2021.

104

Megaloblastic Anemias

A. Victor Hoffbrand

The megaloblastic anemias are a group of disorders characterized by the presence of distinctive morphologic appearances of the developing red cells in the bone marrow. The marrow is usually hypercellular, and the anemia is based on ineffective erythropoiesis. The cause is usually a deficiency of either cobalamin (vitamin B_{12}) or folate, but megaloblastic anemia may occur because of genetic or acquired abnormalities that affect the metabolism of these vitamins or because of defects in DNA synthesis not related to cobalamin or folate **(Table 104-1)**.

COBALAMIN

Cobalamin (vitamin B_{12}) exists in a number of different chemical forms. All have a cobalt atom at the center of a corrin ring. In nature, the vitamin is mainly in the 2-deoxyadenosyl (ado) form, which is located in mitochondria. It is the cofactor for the enzyme L-methylmalonyl

TABLE 104-1 Causes of Megaloblastic Anemia

- Cobalamin deficiency or abnormalities of cobalamin metabolism **(see Tables 104-3, 104-4)**
- Folate deficiency or abnormalities of folate metabolism **(see Table 104-5)**
- Therapy with antifolate drugs (e.g., methotrexate)
- Independent of either cobalamin or folate deficiency and refractory to cobalamin and folate therapy:
  - Some cases of acute myeloid leukemia, myelodysplasia
  - Therapy with drugs interfering with synthesis of DNA (e.g., cytosine arabinoside, hydroxyurea, 6-mercaptopurine, azidothymidine [AZT])
  - Orotic aciduria (responds to uridine)
  - Thiamine-responsive

coenzyme A (CoA) mutase. The other major natural cobalamin is methylcobalamin, the form in human plasma and in cell cytoplasm. It is the cofactor for methionine synthase. Minor amounts of hydroxocobalamin are also present to which methyl- and ado-cobalamin are converted rapidly by exposure to light.

DIETARY SOURCES AND REQUIREMENTS

Cobalamin is synthesized solely by microorganisms. Ruminants obtain cobalamin from the foregut, but the only source for humans is food of animal origin, for example, meat, fish, and dairy products. Vegetables, fruits, and other foods of nonanimal origin are free from cobalamin unless they are contaminated by bacteria. A normal Western diet contains 5–30 μg of cobalamin daily. Adult daily losses (mainly in the urine and feces) are 1–3 μg (~0.1% of body stores), and because the body does not have the ability to degrade cobalamin, daily requirements are also about 1–3 μg. Body stores are of the order of 2–3 mg, sufficient for 3–4 years if supplies are completely cut off.

ABSORPTION

Two mechanisms exist for cobalamin absorption. One is passive, occurring equally through buccal, duodenal, and ileal mucosa; it is rapid but extremely inefficient, with <1% of an oral dose being absorbed by this process. The normal physiologic mechanism is active; it occurs through the ileum and is efficient for small (a few micrograms) oral doses of cobalamin, and it is mediated by gastric intrinsic factor (IF). Dietary cobalamin is released from protein complexes by enzymes in the stomach, duodenum, and jejunum; it combines rapidly with a salivary glycoprotein that belongs to the family of cobalamin-binding proteins known as haptocorrins (HCs). In the intestine, the HC is digested by pancreatic trypsin and the cobalamin is transferred to IF.

IF (gene at chromosome 11q13) is produced in the gastric parietal cells of the fundus and body of the stomach, and its secretion parallels that of hydrochloric acid. Normally, a vast excess of IF is available. The IF-cobalamin complex passes to the ileum, where IF attaches to a specific receptor (cubilin) on the microvillus membrane of the enterocytes. Cubilin also is present in yolk sac and renal proximal tubular epithelium. Cubilin appears to traffic by means of amnionless (AMN), an endocytic receptor protein that directs sublocalization and endocytosis of cubilin with its ligand IF-cobalamin complex. The cobalamin-IF complex enters the ileal cell, where IF is destroyed. After a delay of about 6 h, the cobalamin appears in portal blood attached to transcobalamin (TC) II.

Between 0.5 and 5 μg of cobalamin enter the bile each day. This binds to IF, and a major portion of biliary cobalamin normally is reabsorbed together with cobalamin derived from sloughed intestinal cells. Because of the appreciable amount of cobalamin undergoing enterohepatic circulation, cobalamin deficiency develops more rapidly in individuals who malabsorb cobalamin than it does in vegans, in whom reabsorption of biliary cobalamin is intact.

TRANSPORT

Two main cobalamin transport proteins exist in human plasma; they both bind cobalamin—one molecule for one molecule. One an HC, also known as transcobalamin (TC) I, is closely related to other cobalamin-binding HCs in milk, gastric juice, bile, saliva, and other fluids. The gene *TCNL* is at chromosome 11q11-q12.3. These HCs differ from each other only in the carbohydrate moiety of the molecule. TC I is derived primarily from the specific granules in neutrophils. Normally, it is about two-thirds saturated with cobalamin, which it binds tightly. TC I does not enhance cobalamin entry into tissues. Glycoprotein receptors on liver cells are involved in the removal of TC I from plasma, and TC I may play a role in the transport of cobalamin analogues (which it binds more effectively than does IF) to the liver for excretion in bile.

The other major cobalamin transport protein in plasma is transcobalamin, also known as TC II. The gene is on chromosome 22q11-q13.1. As for IF and HCs, there are nine exons. The three proteins are likely to have a common ancestral origin. TC II is synthesized by liver and by other tissues, including macrophages, ileum, and vascular endothelium. It normally carries only 20–60 ng of cobalamin per liter of plasma and readily gives up cobalamin to marrow, placenta, and other tissues, which it enters by receptor-mediated endocytosis involving the TC II receptor and megalin (encoded by the *LRP-2* gene). The TC II cobalamin is internalized by endocytosis via clathrin-coated pits; the complex is degraded, but the receptor probably is recycled to the cell membrane as is the case for transferrin. Export of "free" cobalamin is via the ATP-binding cassette drug transporter alias multidrug resistance protein 1.

FOLATE

DIETARY FOLATE

Folic (pteroylglutamic) acid is a yellow, crystalline, water-soluble substance. It is the parent compound of a large family of natural folate compounds, which differ from it in three respects: (1) they are partly or completely reduced to dihydrofolate (DHF) or tetrahydrofolate (THF) derivatives, (2) they usually contain a single carbon unit **(Table 104-2)**, and (3) 70–90% of natural folates are folate-polyglutamates. These usually have a chain of four to six glutamate moieties rather than one, as in the monoglutamate folic acid. The whole family is known as folate or vitamin B_9.

Most foods contain some folate. The highest concentrations are found in liver, yeast, spinach, other greens, and nuts (>100 μg/100 g). The total folate content of an average Western diet is 400–500 μg daily, but the amount varies widely according to the type of food eaten and the method of cooking. Folate is easily destroyed by heating, particularly in large volumes of water. Total-body folate in the adult is ~10 mg, with the liver containing the largest store. Daily adult requirements are 100–200 μg, and so stores are sufficient for only 3–4 months in normal adults, and severe folate deficiency may develop rapidly.

ABSORPTION

Folates are absorbed rapidly from the upper small intestine. The absorption of folate polyglutamates is less efficient than that of monoglutamates; on average, ~50% of food folate is absorbed. Polyglutamate forms are hydrolyzed to the monoglutamate derivatives either in the lumen of the intestine or within the mucosa. All dietary folates are converted to 5-methyl-THF (5-MTHF) within the small intestinal mucosa before entering portal plasma. Monoglutamates are actively transported across the enterocyte by a proton-coupled folate transporter (PCFT, SCL46A1). This is situated at the apical brush border and is most active at pH 5.5, which is about the pH of the duodenal and jejunal surface. Genetic mutations of this protein underlie hereditary malabsorption of folate (see below). Pteroylglutamic acid at doses >400 μg is absorbed largely unchanged and converted to natural folates in the liver. Lower doses are converted to 5-MTHF during absorption through the intestine.

About 60–90 μg of folate enter the bile each day and are excreted into the small intestine. Loss of this folate, together with the folate of sloughed intestinal cells, accelerates the speed with which folate deficiency develops in malabsorption conditions.

TRANSPORT

Folate is transported in plasma; about one-third is loosely bound to albumin, and two-thirds are unbound. In all body fluids (plasma, cerebrospinal fluid, milk, bile), folate is largely, if not entirely, 5-MTHF

TABLE 104-2 Biochemical Reactions of Folate Coenzymes

REACTION	COENZYME FORM OF FOLATE INVOLVED	SINGLE CARBON UNIT TRANSFERRED	IMPORTANCE
Formate activation	THF	$-CHO$	Generation of 10-formyl-THF
Purine synthesis			
Formation of glycinamide ribonucleotide	5,10-Methylene-THF	$-CHO$	Formation of purines needed for DNA, RNA synthesis, but reactions probably not rate-limiting
Formylation of aminoimidazole carboxamide ribonucleotide (AICAR)	10-Formyl (CHO)THF		
Pyrimidine synthesis			
Methylation of deoxyuridine monophosphate (dUMP) to thymidine monophosphate (dTMP)	5,10-Methylene-THF	$-CH_3$	Rate limiting in DNA synthesis Oxidizes THF to DHF Some breakdown of folate at the C-9–N-10 bond
Amino acid interconversion			
Serine-glycine interconversion	THF	$=CH_2$	Entry of single carbon units into active pool
Homocysteine to methionine	5-Methyl(M)THF	$-CH_3$	Demethylation of 5-MTHF to THF; also requires cobalamin, flavine adenine dinucleotide, ATP, and adenosylmethionine
Forminoglutamic acid to glutamic acid in histidine catabolism	THF	$-HN-CH=$	

Abbreviations: DHF, dihydrofolate; THF, tetrahydrofolate.

(the monoglutamate form). Three types of folate-binding protein are involved. A reduced folate transporter (RFC, SLC19A1) is the major route of delivery of plasma folate (5-MTHF) to cells. Two folate receptors, FR2 and FR3 embedded in the cell membrane by a glycosyl phosphatidylinositol anchor, transport folate into the cell via receptor-mediated endocytosis. The third protein, proton-coupled folate transporter (PCFT), transports folate at low pH from the vesicle to the cell cytoplasm. The reduced folate transporter also mediates uptake of methotrexate by cells.

BIOCHEMICAL FUNCTIONS

Folates (as the intracellular polyglutamate derivatives) act as coenzymes in the transfer of single-carbon units (Fig. 104-1 and Table 104-2). Two of these reactions are involved in purine synthesis and one in pyrimidine synthesis necessary for DNA and RNA replication. Folate is also a coenzyme for methionine synthesis, in which methylcobalamin is also involved and in which THF is regenerated. THF is the acceptor of single carbon units newly entering the active pool via conversion of serine to glycine. Methionine, the other product of the methionine synthase reaction, is the precursor for *S*-adenosylmethionine (SAM), the universal methyl donor involved in >100 methyltransferase reactions (Fig. 104-1).

During thymidylate synthesis, 5,10-methylene-THF is oxidized to DHF. The enzyme DHF reductase converts this to THF. The drugs methotrexate, pyrimethamine, and (mainly in bacteria) trimethoprim inhibit DHF reductase and so prevent formation of active THF coenzymes from DHF. A small fraction of the folate coenzyme is not recycled during thymidylate synthesis but is degraded at the C9-N10 bond.

BIOCHEMICAL BASIS OF MEGALOBLASTIC ANEMIA

The common feature of all megaloblastic anemias is a defect in DNA synthesis that affects rapidly dividing cells in the bone marrow. Conditions that give rise to megaloblastic changes have in common a disparity in the availability of the four immediate precursors of DNA or a block in their condensation to form DNA. The four precursors are the deoxyribonucleoside triphosphates (dNTPs)—dA(adenine)TP and dG(guanine)TP (purines), dT(thymine)TP, and dC(cytosine)TP (pyrimidines). In deficiencies of either folate or cobalamin, conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP), the precursor of deoxythymidine triphosphate (dTTP) (Fig. 104-1) fails. This occurs because folate is needed as the coenzyme 5,10-methylene-THF polyglutamate for conversion of dUMP to dTMP. The availability of 5,10-methylene-THF is reduced in either cobalamin or folate deficiency. Because of the shortage of one or more precursor, DNA replication from multiple origins along the chromosome is slower than normal during mitosis, and the incomplete replicons fail to join up with resulting single-stranded DNA breaks. An alternative and less likely theory for megaloblastic anemia in cobalamin or folate deficiency is misincorporation of uracil into DNA because of the accumulation of deoxyuridine triphosphate (dUTP) at the DNA replication fork as a consequence of the block in conversion of dUMP to dTMP.

COBALAMIN-FOLATE RELATIONS

Folate is required for many reactions in mammalian tissues (Table 104-2). Only two reactions in the body are known to require cobalamin. Methylmalonyl-CoA isomerization requires adocobalamin, and the methylation of homocysteine to methionine requires both methylcobalamin and 5-MTHF (Fig. 104-1). This reaction is the first step in the pathway by which 5-MTHF, which enters bone marrow and other cells from plasma, is converted into all the intracellular folate coenzymes. The coenzymes are all polyglutamated (the larger size aiding retention in the cell), but the enzyme folate polyglutamate synthase can use only THF, not 5-MTHF, as substrate. In cobalamin deficiency, 5-MTHF accumulates in plasma, and intracellular folate concentrations fall due to failure of formation of THF, the substrate on which folate polyglutamates are built. This has been termed *THF starvation*, or the *methylfolate trap*. This trap also occurs at the polyglutamate level with accumulation of the methyl form at the expense of the other one-carbon forms.

This theory explains the abnormalities of folate metabolism that occur in cobalamin deficiency (high serum folate, low cell folate, positive purine precursor aminoimidazole carboxamide ribonucleotide [AICAR] excretion) (Table 104-2) and also why the anemia of cobalamin deficiency responds to folic acid in large doses, which overcome the methylfolate trap (Fig 104-1).

CLINICAL FEATURES

Many symptomless patients are detected through the finding of a raised mean corpuscular volume (MCV) on a routine blood count. The main clinical features in more severe cases are those of anemia. Anorexia is usually marked, and weight loss, diarrhea, or constipation may be present. Glossitis, angular cheilosis, a mild fever in more severely anemic patients, jaundice (unconjugated), and reversible melanin skin hyperpigmentation also may occur with a deficiency of either folate or cobalamin. Thrombocytopenia sometimes leads to bruising, and this may be aggravated by vitamin C deficiency or alcohol in malnourished patients. The anemia and low leukocyte count may predispose to infections, particularly of the respiratory and urinary tracts. Cobalamin

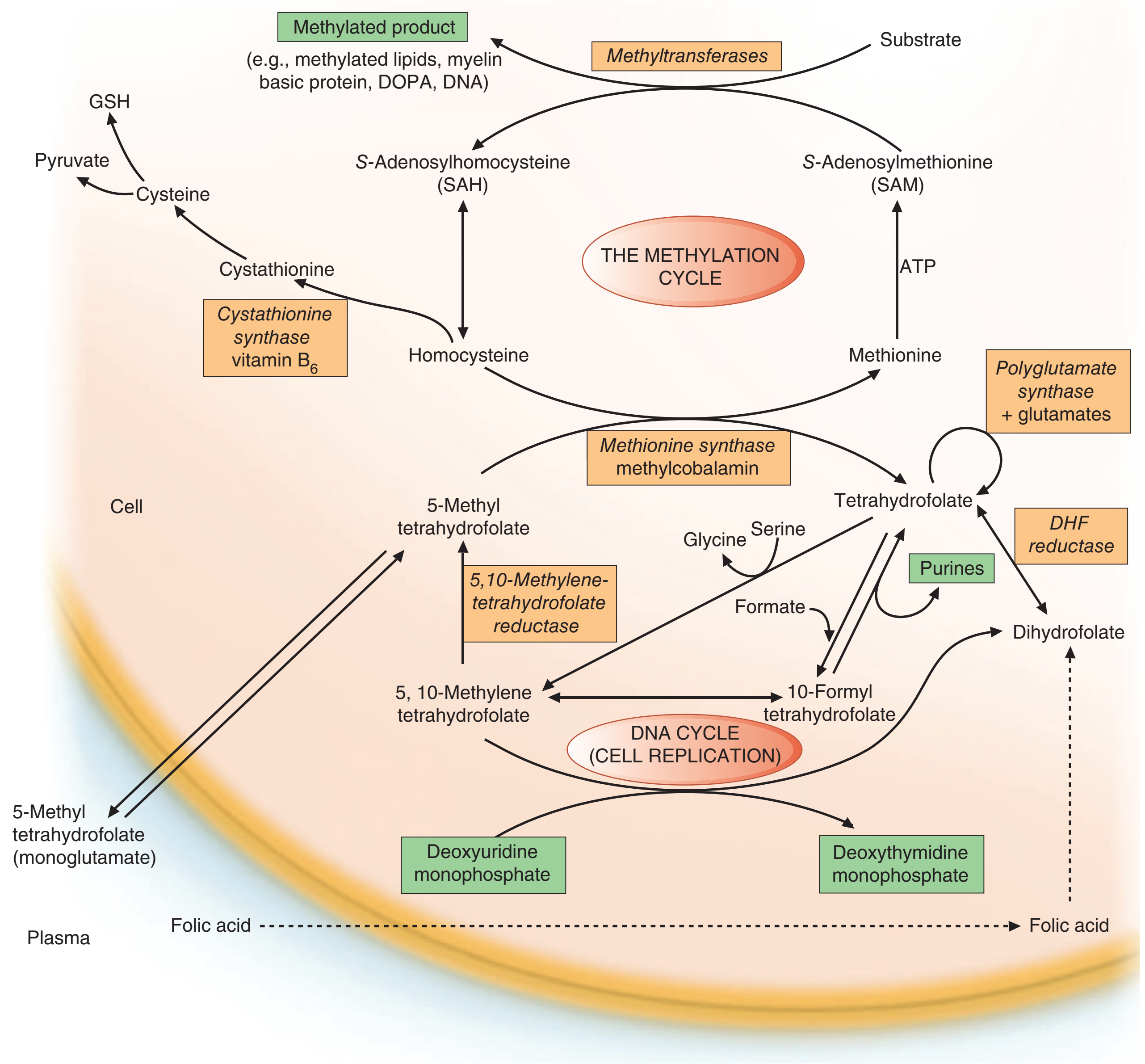
Methylated product
(e.g., methylated lipids, myelin basic protein, DOPA, DNA)
Methyltransferases
Substrate
GSH
Pyruvate
Cysteine
S-Adenosylhomocysteine (SAH)
S-Adenosylmethionine (SAM)
THE METHYLATION CYCLE
ATP
Cystathionine
Cystathionine synthase vitamin B_6
Homocysteine
Methionine
Polyglutamate synthase + glutamates
Methionine synthase methylcobalamin
Tetrahydrofolate
DHF reductase
Cell
5-Methyl tetrahydrofolate
Glycine
Serine
Purines
5,10-Methylene-tetrahydrofolate reductase
Formate
Dihydrofolate
5, 10-Methylene tetrahydrofolate
10-Formyl tetrahydrofolate
DNA CYCLE (CELL REPLICATION)
5-Methyl tetrahydrofolate (monoglutamate)
Deoxyuridine monophosphate
Deoxythymidine monophosphate
Plasma
Folic acid
Folic acid

FIGURE 104-1 The role of folates in DNA synthesis and in formation of *S*-adenosylmethionine (SAM), which is involved in numerous methylation reactions. DHF, dihydrofolate; GSH, glutathione. *(Reproduced with permission from AV Hoffbrand et al [eds]: Postgraduate Haematology, 5th ed. Oxford, UK, Blackwell Publishing, 2005.)*

deficiency has also been associated in a few studies with impaired bactericidal function of phagocytes and with osteoporosis.

Neurologic Manifestations Cobalamin is needed for the myelination of the central nervous system. Its deficiency may cause a bilateral peripheral neuropathy or degeneration (demyelination) of the cervical and thoracic posterior and lateral (pyramidal) tracts of the spinal cord and, less frequently, of the cranial nerves and of the white matter of the brain. Optic atrophy and cerebral symptoms including dementia, depression, psychotic symptoms, and cognitive impairment may be prominent. Anosmia and loss of taste may occur. Magnetic resonance imaging (MRI) may show the "spongy" degeneration of the cord.

The patient, more frequently male, typically presents with paresthesias, muscle weakness, or difficulty in walking but sometimes may present with dementia, psychotic disturbances, or visual impairment. Loss of proprioception and vibration sensation is usually present with positive Romberg and Lhermitte signs. Gait may be ataxic with spasticity (hyperreflexia). Autonomic nervous dysfunction can result in postural hypotension, impotence, and incontinence.

Long-term nutritional cobalamin deficiency in infancy leads to poor brain development and impaired intellectual development. In infancy, feeding difficulties, lethargy, and coma may be noted. Convulsions and myoclonus have been described. An important clinical problem is the nonanemic patient with neurologic or psychiatric abnormalities and a low or borderline serum cobalamin level. In such patients, it is necessary to try to establish whether significant cobalamin deficiency is present, for example, by careful examination of the blood film for macrocytosis or hypersegmented neutrophils (see below), tests for pernicious anemia (PA) by serum gastrin level and antibodies to IF or parietal cells, and serum methylmalonic acid (MMA) measurement. A trial of cobalamin therapy for at least 3 months will usually also be needed to determine whether the symptoms improve.

The biochemical basis for cobalamin neuropathy remains obscure. Its occurrence in the absence of methylmalonic aciduria in TC II deficiency suggests that the neuropathy is related to the defect in homocysteine-methionine conversion. Accumulation of *S*-adenosylhomocysteine in the brain, resulting in inhibition of transmethylation reactions, has been suggested. Folate deficiency has been suggested to cause organic neurologic disease, but this is uncertain, although methotrexate injected into the cerebrospinal fluid may cause brain or spinal cord damage.

Psychiatric disturbance, as discussed above, is common in both folate and cobalamin deficiencies. This, like the neuropathy, has been attributed to a failure of the synthesis of SAM, which is needed in methylation of biogenic amines (e.g., dopamine) as well as that of proteins, phospholipids, and neurotransmitters in the brain (Fig. 104-1).

GENERAL TISSUE EFFECTS OF COBALAMIN AND FOLATE DEFICIENCIES

Epithelial Surfaces After the marrow, the next most frequently affected tissues are the epithelial cell surfaces of the mouth (with glossitis), stomach, small intestine, and respiratory, urinary, and female genital tracts. The cells show macrocytosis with increased numbers of multinucleate and dying cells. The deficiencies may cause cervical smear abnormalities.

Complications of Pregnancy The gonads are also affected, and infertility is common in both men and women with severe deficiency of either vitamin. Maternal folate deficiency has been implicated as a cause of prematurity, and both folate and cobalamin deficiencies have been implicated in recurrent fetal loss and neural tube defects.

Neural Tube Defects Folic acid supplements at the time of conception and in the first 12 weeks of pregnancy can reduce by ~80% the incidence of neural tube defects (NTDs) (anencephaly, meningomyelocele, encephalocele, and spina bifida) in the fetus. Most of this protective effect can be achieved by taking folic acid, 0.4 mg daily, before and at the time of conception.

The incidence of cleft palate and harelip also can be reduced by prophylactic folic acid. No clear simple relationship exists between maternal folate status and these fetal abnormalities, although for NTDs, it has been established that the lower the maternal folate, the greater is the risk to the fetus. NTDs also can be caused by antifolate and antiepileptic drugs.

An underlying maternal folate metabolic abnormality has also been postulated. One abnormality has been identified: reduced activity of the enzyme 5,10-methylene-THF reductase (MTHFR) (Fig. 104-1) caused by a common C677T polymorphism in the *MTHFR* gene. In one study, the prevalence of this polymorphism was found to be higher than in controls in the parents of NTD fetuses and in the fetuses themselves: homozygosity for the TT mutation was found in 13% of cases compared with 5% of control subjects. The polymorphism codes for a thermolabile form of MTHFR. The homozygous state results in a lower mean serum and red cell folate level compared with control subjects, as well as significantly higher serum homocysteine levels. Tests for mutations in other enzymes possibly associated with NTDs, for example, methionine synthase and serine–glycine hydroxymethylase, have been negative. Serum cobalamin levels are also lower in the sera of mothers of NTD infants than in controls. In addition, maternal TC II receptor polymorphisms are associated with increased risk of NTD births. However, no studies have been undertaken that show that dietary fortification with cobalamin reduces the incidence of NTDs.

Cardiovascular Disease Children with severe homocystinuria (blood levels ≥100 μmol/L) due to deficiency of one of three enzymes (methionine synthase, MTHFR, or cystathionine synthase; Fig. 104-1) have vascular disease, for example, ischemic heart disease, cerebrovascular disease, or pulmonary embolus, as teenagers or in young adulthood. Lesser degrees of raised serum homocysteine and low levels of serum folate and homozygous inherited mutations of *MTHFR* have been found to be associated with cerebrovascular, peripheral vascular, and coronary heart disease and with deep vein thrombosis. Prospective randomized trials of lowering homocysteine levels with supplements of folic acid, vitamin B_{12}, and vitamin B_6 against placebo over a 5-year period in patients with vascular disease or diabetes have not, however, shown a reduction of first event fatal or nonfatal myocardial infarction, nor have these supplements reduced the risk of recurrent cardiovascular disease after an acute myocardial infarct. Meta-analysis showed an 18% reduction in strokes. The benefit for stroke prevention has been confirmed by a large (>20,000 subjects) randomized prospective study in hypertensive subjects in China. This showed a significant reduction in the first incidence of stroke in subjects receiving enalapril and folic acid compared to enalapril alone. The effect was especially marked in the subjects commencing the prospective trial with the lowest serum folate levels. Venous thrombosis has been reported to be more frequent in folate-deficient or cobalamin-deficient subjects than in controls and to occur at unusual sites such as cerebral venous sinuses. This tendency was ascribed to raised plasma homocysteine levels in folate or cobalamin deficiency, but no evidence exists that folic acid or cobalamin supplements reduce the prevalence of venous thrombosis.

Cognitive Decline Association between low serum folate or cobalamin levels and higher homocysteine levels with the development of decreased cognitive function and of dementia in Alzheimer's disease has been reported. A meta-analysis of randomized, placebo-controlled trials of homocysteine-lowering B-vitamin supplementation of individuals with and without cognitive impairment, however, showed that supplementation with vitamin B_{12}, vitamin B_6, and folic acid alone or in combination did not improve cognitive function or slow cognitive decline. It is unknown whether prolonged treatment with these B vitamins can reduce the risk of dementia in later life.

Malignancy Prophylactic folic acid in pregnancy has been found in some but not all studies to reduce the subsequent incidence of acute lymphoblastic leukemia (ALL) in childhood. A significant negative association has also been found with the *MTHFR* C677T polymorphism and leukemias with mixed lineage leukemia (MLL) translocations, but a positive association was found with hyperdiploidy in infants with ALL or acute myeloid leukemia or with childhood ALL. A second polymorphism in the *MTHFR* gene, A1298C, is also strongly associated with hyperdiploid leukemia. Various positive and negative associations are noted between polymorphisms in folate-dependent enzymes and the incidence of adult ALL. The C677T polymorphism is thought to lead to increased thymidine pools and "better quality" of DNA synthesis by shunting one-carbon groups toward thymidine and purine synthesis. This may explain its reported association with a lower risk for colorectal cancer. Most but not all studies suggest that prophylactic folic acid also protects against colon adenomas. Other tumors that have been associated with folate polymorphisms or status include follicular lymphoma, breast cancer, and gastric cancer.

A meta-analysis of 50,000 individuals given folic acid (0.5–40 mg daily) or placebo in cardiovascular (n = 10) or colon adenoma prevention (n = 3) trials found that folic acid supplementation did not significantly increase or decrease the overall incidence of cancer or of any site-specific cancer during a weighted average scheduled treatment duration of 5.7 years. Because folic acid may "feed" tumors, it probably should be avoided in those with established tumors unless severe megaloblastic anemia due to folate deficiency is present.

HEMATOLOGIC FINDINGS

PERIPHERAL BLOOD

Oval macrocytes, usually with considerable anisocytosis and poikilocytosis, are the main feature **(Fig. 104-2*A*)**. The MCV is usually >100 fL unless a cause of microcytosis (e.g., iron deficiency or thalassemia trait) is present. Some of the neutrophils are hypersegmented (more than five nuclear lobes). There may be leukopenia due to a reduction in granulocytes and lymphocytes, but this is usually >1.5 × 10^9/L; the platelet count may be moderately reduced, rarely to <40 × 10^9/L. The severity of all these changes parallels the degree of anemia. In a nonanemic patient, the presence of a few macrocytes and hypersegmented neutrophils in the peripheral blood may be the only indication of the underlying disorder.

BONE MARROW

In a severely anemic patient, the marrow is hypercellular with an accumulation of primitive cells due to selective death by apoptosis of more mature forms. The erythroblast nucleus maintains a primitive, fine chromatin appearance despite maturation and hemoglobinization of the cytoplasm. The cells are larger than normoblasts, and an increased number of cells with eccentric lobulated nuclei or nuclear fragments may be present **(Fig. 104-2*B*)**. Giant and abnormally shaped metamyelocytes and enlarged hyperpolyploid megakaryocytes are characteristic. In severe cases, the accumulation of primitive cells ("blasts")

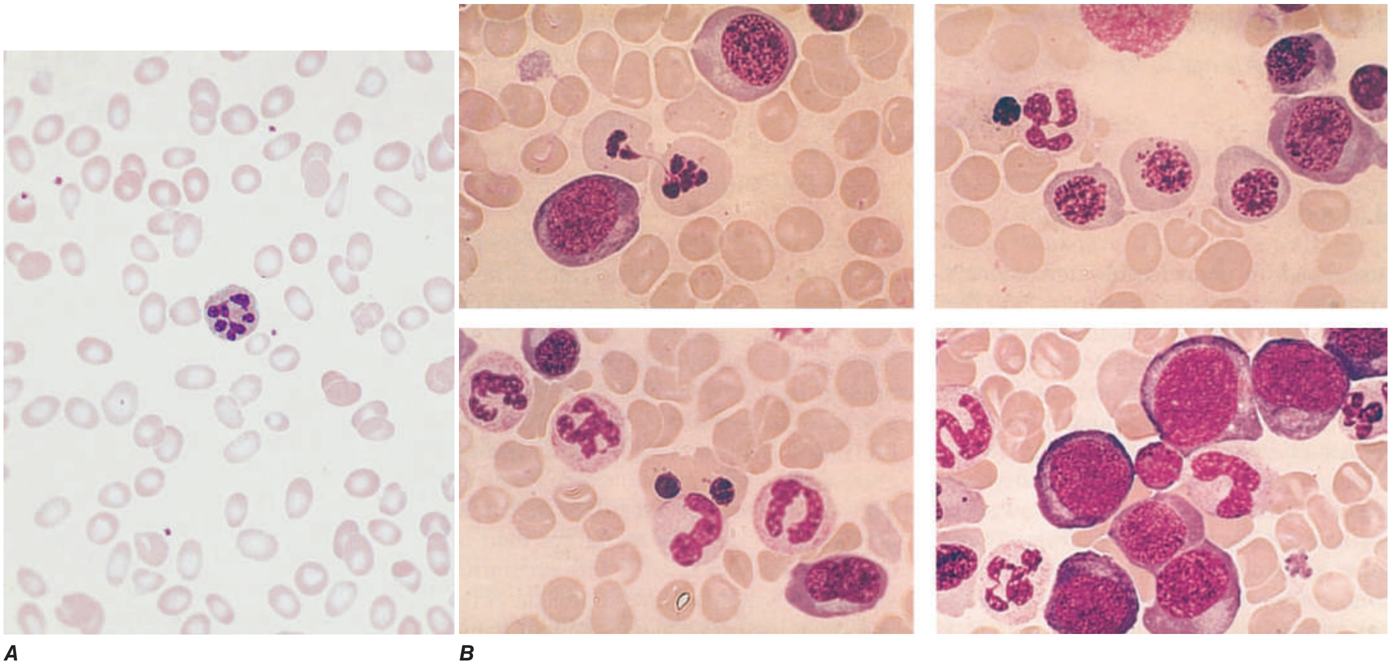

FIGURE 104-2 ***A.*** The peripheral blood in severe megaloblastic anemia. ***B.*** The bone marrow in severe megaloblastic anemia. *(Reprinted from AV Hoffbrand et al [eds]: Postgraduate Haematology, 5th ed. Oxford, UK, Blackwell Publishing, 2005; with permission.)*

may mimic acute myeloid leukemia, whereas in less anemic patients, the changes in the marrow may be difficult to recognize. The terms *intermediate*, *mild*, and *early* have been used. The term *megaloblastoid* is best avoided. It has been used to describe cells with both immature-appearing nuclei and defective hemoglobinization refractory to folic acid or cobalamin therapy, especially seen in myelodysplasia.

CHROMOSOMES

Bone marrow cells, transformed lymphocytes, and other proliferating cells in the body show a variety of chromosomal changes, including random breaks, reduced contraction, spreading of the centromere, and exaggeration of secondary chromosomal constrictions and over-prominent satellites. Similar abnormalities may be produced by antimetabolite drugs (e.g., cytarabine, hydroxyurea, thioguanine, and methotrexate) that interfere with either DNA replication or folate metabolism and that also cause megaloblastic appearances.

INEFFECTIVE HEMATOPOIESIS

Unconjugated bilirubin accumulates in plasma due to the death of nucleated red cells in the marrow (ineffective erythropoiesis). Other evidence for this includes raised urine urobilinogen, reduced haptoglobins, positive urine hemosiderin, and raised serum lactate dehydrogenase. A weakly positive direct antiglobulin test due to complement only can lead to a false diagnosis of autoimmune hemolytic anemia.

CAUSES OF COBALAMIN DEFICIENCY

Cobalamin deficiency is usually due to malabsorption. The only other cause is inadequate dietary intake.

INADEQUATE DIETARY INTAKE

Adults Dietary cobalamin deficiency arises in vegans who omit meat, fish, eggs, cheese, and other animal products from their diet. The largest group in the world consists of Hindus, and it is likely that many millions of Indians are at risk of deficiency of cobalamin on a nutritional basis. Subnormal serum cobalamin levels are found in up to 50% of randomly selected, young, adult Indian vegans, but the deficiency usually does not progress to megaloblastic anemia since the diet of most vegans is not totally lacking in cobalamin and the enterohepatic circulation of cobalamin is intact. Dietary cobalamin deficiency may also arise rarely in nonvegetarian individuals who exist on grossly inadequate diets because of poverty or psychiatric disturbance.

Infants Cobalamin deficiency has been described in infants born to severely cobalamin-deficient mothers. These infants develop megaloblastic anemia at about 3–6 months of age, presumably because they are born with low stores of cobalamin and because they are fed breast milk with low cobalamin content. The babies have also shown growth retardation, impaired psychomotor development, and other neurologic sequelae. MRI shows delayed myelination and brain atrophy.

GASTRIC CAUSES OF COBALAMIN MALABSORPTION

See Tables 104-3 and 104-4.

Formerly, the pathogenesis of cobalamin malabsorption was distinguishable based on the results of a Schilling test in which a radioactive form of cobalamin was administered orally and its appearance in the urine was a sign of absorption. Radioactive cobalamin is no longer available, and Schilling tests are no longer performed. Other approaches to the differential diagnosis of cobalamin malabsorption are now employed.

Pernicious Anemia PA, the dominant cause of severe cobalamin deficiency in Western countries, may be defined as a severe lack of IF due to gastric atrophy. It is a common disease in northern Europeans but occurs in all countries and ethnic groups. It is more frequent in people of African than Asian ancestry. The overall incidence is about

TABLE 104-3 Causes of Cobalamin Deficiency Sufficiently Severe to Cause Megaloblastic Anemia

NUTRITIONAL	VEGANS
Malabsorption	Pernicious anemia
Gastric causes	Congenital absence of intrinsic factor or functional abnormality
	Total or partial gastrectomy
Intestinal causes	Intestinal stagnant loop syndrome: jejunal diverticulosis, ileocolic fistula, anatomic blind loop, intestinal stricture, etc.
	Ileal resection and Crohn's disease
	Selective malabsorption with proteinuria
	Tropical sprue
	Transcobalamin II deficiency
	Fish tapeworm

TABLE 104-4 Malabsorption of Cobalamin May Occur in the Following Conditions but Is Not Usually Sufficiently Severe and Prolonged to Cause Megaloblastic Anemia

Gastric causes
Simple atrophic gastritis (food cobalamin malabsorption)
Zollinger-Ellison syndrome
Gastric bypass or bariatric surgery
Use of proton pump inhibitors
Intestinal causes
Gluten-induced enteropathy
Severe pancreatitis
HIV infection
Radiotherapy
Graft-versus-host disease
Deficiencies of cobalamin, folate, protein,? riboflavin,? nicotinic acid
Therapy with colchicine, para-aminosalicylate, neomycin, slow-release potassium chloride, anticonvulsant drugs, metformin,[a] cytotoxic drugs
Alcohol

[a]It is now thought that metformin lowers serum vitamin B_{12} level by lowering the level of transcobalamin I.

120 per 100,000 population in the United Kingdom (UK). The ratio of incidence in men and women among whites is ~1:1.6, and the median age of onset is 70–80 years, with only 10% of patients being <40 years of age. However, in some ethnic groups, notably blacks and Latin Americans, the age at onset of PA is generally lower. The disease occurs more commonly than by chance in close relatives and in persons with other organ-specific autoimmune diseases, for example, thyroid diseases, vitiligo, hypoparathyroidism, type 1 diabetes, and Addison's disease. It is also associated with hypogammaglobulinemia, premature graying or blue eyes, and persons of blood group A. An association with human leukocyte antigen (HLA) 3 has been reported in some but not all series and, in those with endocrine disease, with HLA-B8, -B_{12}, and -BW15. Life expectancy is normal in women once regular treatment has begun. Men had in earlier decades a slightly subnormal life expectancy as a result of a higher incidence of carcinoma of the stomach than in control subjects, but current data on their life expectancy are unavailable. Gastric output of hydrochloric acid, pepsin, and IF is severely reduced. The serum gastrin level is raised, and serum pepsinogen I levels are low.

Gastric Biopsy A single endoscopic examination is recommended if PA is diagnosed. Gastric biopsy usually shows atrophy of all layers of the body and fundus, with loss of glandular elements, an absence of parietal and chief cells and replacement by mucous cells, a mixed inflammatory cell infiltrate, and perhaps intestinal metaplasia. The infiltrate of plasma cells and lymphocytes contains an excess of CD4 cells. These are directed against gastric H/K-ATPase. The antral mucosa is usually well preserved. *Helicobacter pylori* infection occurs infrequently in PA, but it has been suggested that *H. pylori* gastritis occurs at an early phase of atrophic gastritis and presents in younger patients as iron-deficiency anemia and in older patients as PA. *H. pylori* is suggested to stimulate an autoimmune process directed against parietal cells. It has been suggested that *H. pylori* infection is replaced, in some individuals, by the autoimmune process.

Serum Antibodies Two types of IF immunoglobulin G antibody may be found in the sera of patients with PA. The "blocking," or type I, antibody prevents the combination of IF and cobalamin, whereas the "binding," or type II, antibody prevents attachment of IF to ileal mucosa. Type I occurs in the sera of ~55% of patients, and type II in 35%. IF antibodies cross the placenta and may cause temporary IF deficiency in a newborn infant. Type I antibody has been detected rarely in the sera of patients without PA but with thyrotoxicosis, myxedema, Hashimoto's disease, or diabetes mellitus and in relatives of PA patients. IF antibodies have also been detected in gastric juice in ~80% of PA patients. These gastric antibodies may reduce absorption of dietary cobalamin by combining with small amounts of remaining IF. Patients with PA also show cell-mediated immunity to IF.

Parietal cell antibody is present in the sera of almost 90% of adult patients with PA but is frequently present in other subjects. Thus, it occurs in as many as 16% of randomly selected female subjects age >60 years. The parietal cell antibody is directed against the α and β subunits of the gastric proton pump (H^+, K^+-ATPase).

JUVENILE PERNICIOUS ANEMIA

This usually occurs in older children and resembles PA of adults. Gastric atrophy, achlorhydria, and serum IF antibodies are all present, although parietal cell antibodies are usually absent. About one-half of these patients show an associated endocrinopathy such as autoimmune thyroiditis, Addison's disease, or hypoparathyroidism; in some, mucocutaneous candidiasis occurs.

CONGENITAL INTRINSIC FACTOR DEFICIENCY OR FUNCTIONAL ABNORMALITY

An affected child usually presents with megaloblastic anemia in the first to third year of life; a few have presented as late as the second decade. The child usually has no demonstrable IF but has a normal gastric mucosa and normal secretion of acid. The inheritance is autosomal recessive. Parietal cell and IF antibodies are absent. Variants have been described in which the child is born with IF that can be detected immunologically but is unstable or functionally inactive, unable to bind cobalamin or to facilitate its uptake by ileal receptors.

GASTRECTOMY

After total gastrectomy, cobalamin deficiency is inevitable, and prophylactic cobalamin therapy should be commenced immediately after the operation. After partial gastrectomy, 10–15% of patients also develop this deficiency. The exact incidence and time of onset are most influenced by the size of the resection and the preexisting size of cobalamin body stores.

FOOD COBALAMIN MALABSORPTION

Failure of release of cobalamin from binding proteins in food is responsible for this condition, which is more common in the elderly. It is associated with low serum cobalamin levels, with or without raised serum levels of MMA and homocysteine. Typically, these patients have normal cobalamin absorption, as measured with crystalline cobalamin, but show malabsorption when a modified test using food-bound cobalamin is used. It is usually due to mild forms of atrophic gastritis or therapy with proton pump inhibitors. Bariatric surgery is likely to be an increasing cause of this form of cobalamin malabsorption and deficiency. The frequency of progression to severe cobalamin deficiency and the reasons for this progression are not clear.

INTESTINAL CAUSES OF COBALAMIN MALABSORPTION

Intestinal Stagnant Loop Syndrome Malabsorption of cobalamin occurs in a variety of intestinal lesions in which there is colonization of the upper small intestine by fecal organisms. This may occur in patients with jejunal diverticulosis, entero-anastomosis, or an intestinal stricture or fistula or with an anatomic blind loop due to Crohn's disease, tuberculosis, or an operative procedure.

Ileal Resection Removal of ≥1.2 m of terminal ileum causes malabsorption of cobalamin. In some patients after ileal resection, particularly if the ileocecal valve is incompetent, colonic bacteria may contribute further to the onset of cobalamin deficiency.

Selective Malabsorption of Cobalamin with Proteinuria (Imerslund's Syndrome; Imerslund-Gräsbeck Syndrome; Congenital Cobalamin Malabsorption; Autosomal Recessive Megaloblastic Anemia; MGA1) This autosomal recessive disease is the most common cause of megaloblastic anemia due to cobalamin deficiency in infancy in Western countries. More than 200 cases have been reported with familial clusters in Finland, Norway, the Middle East, and North Africa. The patients secrete normal amounts of IF and gastric acid but are unable to absorb cobalamin. In Finland, impaired synthesis, processing, or ligand binding of cubilin due to inherited mutations is found. In Norway, mutation of the gene for *AMN* has been reported. Other tests of intestinal absorption are normal. Over 90% of these patients show nonspecific proteinuria, but renal function is otherwise normal, and renal biopsy has not shown any consistent renal defect. A few have shown aminoaciduria and congenital renal abnormalities, such as duplication of the renal pelvis.

Tropical Sprue Nearly all patients with acute and subacute tropical sprue show malabsorption of cobalamin. This may persist as the principal abnormality in the chronic form of the disease, when the patient may present with megaloblastic anemia or neuropathy due to cobalamin deficiency. Absorption of cobalamin usually improves after antibiotic and, in the early stages, folic acid therapy.

Fish Tapeworm Infestation The fish tapeworm (*Diphyllobothrium latum*) lives in the small intestine of humans and accumulates cobalamin from food, rendering the cobalamin unavailable for absorption. Individuals acquire the worm by eating raw or partly cooked fish. Infestation was common around the lakes of Scandinavia, Germany, Japan, North America, and Russia. Megaloblastic anemia or cobalamin neuropathy occurs only in those with a heavy infestation.

Gluten-Induced Enteropathy Malabsorption of cobalamin occurs in ~30% of untreated patients (presumably those in whom the disease extends to the ileum). Cobalamin deficiency is not severe in these patients and is corrected with a gluten-free diet.

Severe Chronic Pancreatitis In this condition, lack of trypsin is thought to cause dietary cobalamin attached to gastric non-IF (HC) binder to be unavailable for absorption. It also has been proposed that, in pancreatitis, the concentration of calcium ions in the ileum falls below the level needed to maintain normal cobalamin absorption.

HIV Infection Serum cobalamin levels tend to fall in patients with HIV infection and are subnormal in 10–35% of those with AIDS. Malabsorption of cobalamin not corrected by IF has been shown in some, but not all, patients with subnormal serum cobalamin levels. Cobalamin deficiency sufficiently severe to cause megaloblastic anemia or neuropathy is rare.

Zollinger-Ellison Syndrome Malabsorption of cobalamin has been reported in the Zollinger-Ellison syndrome. It is thought that there is a failure to release cobalamin from HC binding protein due to inactivation of pancreatic trypsin by high acidity, as well as interference with IF binding of cobalamin.

Radiotherapy Both total-body irradiation and local radiotherapy to the ileum (e.g., as a complication of radiotherapy for carcinoma of the cervix) may cause malabsorption of cobalamin.

Graft-Versus-Host Disease This commonly affects the small intestine. Malabsorption of cobalamin due to abnormal gut flora, as well as damage to ileal mucosa, is common.

Drugs The drugs that have been reported to cause malabsorption of cobalamin are listed in Table 104-4. However, megaloblastic anemia due to these drugs is rare. It has been suggested that metformin lowers serum cobalamin by lowering TC I level rather than causing malabsorption of cobalamin.

■ ABNORMALITIES OF COBALAMIN METABOLISM

Congenital Transcobalamin II Deficiency or Abnormality Infants with TC II deficiency usually present with megaloblastic anemia within a few weeks of birth. Serum cobalamin and folate levels are normal, but the anemia responds to massive (e.g., 1 mg three times weekly) injections of cobalamin. Some cases show neurologic complications. The protein may be present but functionally inert. Genetic abnormalities found include mutations of an intraexonic cryptic splice site, extensive deletion, single nucleotide deletion, nonsense mutation, and an RNA editing defect. Malabsorption of cobalamin occurs in all cases, and serum immunoglobulins are usually reduced. Failure to institute adequate cobalamin therapy or treatment with folic acid may lead to irreversible neurologic damage.

Congenital Methylmalonic Acidemia and Aciduria Infants with this abnormality are ill from birth with vomiting, failure to thrive, severe metabolic acidosis, ketosis, and intellectual disability. Anemia, if present, is normocytic and normoblastic. The condition may be due to a functional defect in either mitochondrial methylmalonyl-CoA mutase or its cofactor adocobalamin. Mutations in the methylmalonyl-CoA mutase are not responsive or are only poorly responsive to treatment with cobalamin. A proportion of infants with failure of adocobalamin synthesis respond to cobalamin in large doses. Some children have combined methylmalonic aciduria and homocystinuria due to defective formation of both cobalamin coenzymes. This usually presents in the first year of life with feeding difficulties, developmental delay, microcephaly, seizures, hypotonia, and megaloblastic anemia.

Acquired Abnormality of Cobalamin Metabolism: Nitrous Oxide Inhalation Nitrous oxide (N_2O) irreversibly oxidizes methylcobalamin to an inactive precursor; this inactivates methionine synthase. Megaloblastic anemia has occurred in patients undergoing prolonged N_2O anesthesia (e.g., in intensive care units). A neuropathy resembling cobalamin neuropathy has been described in dentists and anesthetists who are exposed repeatedly to N_2O. Methylmalonic aciduria does not occur as adocobalamin is not inactivated by N_2O.

CAUSES OF FOLATE DEFICIENCY

(Table 104-5)

■ NUTRITIONAL

Dietary folate deficiency is common except in countries that fortify their diet with folic acid. In most patients with folate deficiency, a nutritional element is present. Certain individuals are particularly prone to have diets containing inadequate amounts of folate (Table 104-5). In the United States and other countries where fortification

TABLE 104-5 Causes of Folate Deficiency

- Dietary[a]
  - Particularly in: old age, infancy, poverty, alcoholism, chronic invalids, and the psychiatrically disturbed; may be associated with scurvy or kwashiorkor
- Malabsorption
  - Major causes of deficiency
    - Tropical sprue, gluten-induced enteropathy in children and adults, and in association with dermatitis herpetiformis, specific malabsorption of folate, intestinal megaloblastosis caused by severe cobalamin or folate deficiency
  - Minor causes of deficiency
    - Extensive jejunal resection, Crohn's disease, partial gastrectomy, congestive heart failure, Whipple's disease, scleroderma, amyloid, diabetic enteropathy, systemic bacterial infection, lymphoma, sulfasalazine (Salazopyrin)
- Excess utilization or loss
  - Physiologic
    - Pregnancy and lactation, prematurity
  - Pathologic
    - Hematologic diseases: chronic hemolytic anemias, sickle cell anemia, thalassemia major, myelofibrosis
    - Malignant diseases: carcinoma, lymphoma, leukemia, myeloma
    - Inflammatory diseases: tuberculosis, Crohn's disease, psoriasis, exfoliative dermatitis, malaria
    - Metabolic disease: homocystinuria
    - Excess urinary loss: congestive heart failure, active liver disease
    - Hemodialysis, peritoneal dialysis
- Antifolate drugs[b]
  - Anticonvulsant drugs (phenytoin, primidone, barbiturates), sulfasalazine
  - Nitrofurantoin, tetracycline, antituberculosis (less well documented)
- Mixed causes
  - Liver diseases, alcoholism, intensive care units

[a] In severely folate-deficient patients with causes other than those listed under Dietary, poor dietary intake is often present. [b] Drugs inhibiting dihydrofolate reductase are discussed in the text.

of the diet with folic acid has been adopted, the prevalence of folate-deficient megaloblastic anemia has dropped dramatically and is now restricted to high-risk groups with increased folate needs. Nutritional folate deficiency occurs in kwashiorkor and scurvy and in infants with repeated infections or those who are fed solely on goats' milk, which has a low folate content.

■ MALABSORPTION

Malabsorption of dietary folate occurs in tropical sprue and in gluten-induced enteropathy. In the rare congenital recessive syndrome of selective malabsorption of folate due to mutation of the PCFT, there is an associated defect of folate transport into the cerebrospinal fluid, and these patients show megaloblastic anemia, which responds to physiologic doses of folic acid given parenterally but not orally. They also show intellectual disability, convulsions, and other central nervous system abnormalities. Minor degrees of malabsorption may also occur after jejunal resection or partial gastrectomy, in Crohn's disease, and in systemic infections, but in these conditions, if severe deficiency occurs, it is usually largely due to poor nutrition. Malabsorption of folate has been described in patients receiving sulfasalazine (Salazopyrin), cholestyramine, and triamterene.

■ EXCESS UTILIZATION OR LOSS

Pregnancy Folate requirements are increased by 50% daily in a normal pregnancy, partly because of transfer of the vitamin to the fetus but mainly because of increased folate catabolism due to cleavage of folate coenzymes in rapidly proliferating tissues. A dietary folate intake of 600 µg daily is recommended. Megaloblastic anemia due to folate deficiency is prevented by prophylactic folic acid therapy. It occurred in 0.5% of pregnancies in the UK and other Western countries before prophylaxis with folic acid, but the incidence is much higher in countries where the general nutritional status is poor. During lactation, folate requirements are increased about 25 and a dietary intake of 500 µg of folate daily is advised.

Prematurity A newborn infant, whether full term or premature, has higher serum and red cell folate concentrations than an adult. However, a newborn infant's demand for folate has been estimated to be up to 10 times that of adults on a weight basis, and the neonatal folate level falls rapidly to the lowest values at about 6 weeks of age. The falls are steepest and are liable to reach subnormal levels in premature babies, a number of whom develop megaloblastic anemia responsive to folic acid at about 4–6 weeks of age. This occurs particularly in the smallest babies (<1500 g birth weight) and those who have feeding difficulties or infections or have undergone multiple exchange transfusions. In these babies, prophylactic folic acid should be given.

Hematologic Disorders Folate deficiency frequently occurs in chronic hemolytic anemias, particularly in sickle cell disease, autoimmune hemolytic anemia, and congenital spherocytosis. In these and in other conditions of increased cell turnover (e.g., myelofibrosis, malignancies), folate deficiency arises because it is not completely reutilized after performing coenzyme functions.

Inflammatory Conditions Chronic inflammatory diseases such as tuberculosis, rheumatoid arthritis, Crohn's disease, psoriasis, exfoliative dermatitis, bacterial endocarditis, and chronic bacterial infections cause deficiency by reducing the appetite and increasing the demand for folate. Systemic infections also may cause malabsorption of folate. Severe deficiency is virtually confined to the patients with the most active disease and the poorest diet.

Homocystinuria This is a rare metabolic defect in the conversion of homocysteine to cystathionine. Folate deficiency occurring in most of these patients may be due to excessive utilization because of compensatory increased conversion of homocysteine to methionine.

Long-Term Dialysis Because folate is only loosely bound to plasma proteins, it is easily removed from plasma by dialysis. In patients with anorexia, vomiting, infections, and hemolysis, folate stores are particularly likely to become depleted. Routine folate prophylaxis is now given.

Congestive Heart Failure and Liver Disease Excess urinary folate losses of >100 µg per day may occur in some of these patients. It appears to be due to release of folate from damaged liver cells.

■ ANTIFOLATE DRUGS

A large number of people with epilepsy receiving long-term therapy with phenytoin or primidone, with or without barbiturates, develop low serum and red cell folate levels. The exact mechanism is unclear. Alcohol may also be a folate antagonist, as patients who are drinking spirits may develop megaloblastic anemia that will respond to normal quantities of dietary folate or to physiologic doses of folic acid only if alcohol is withdrawn. Macrocytosis of red cells is associated with chronic alcohol intake even when folate levels are normal. Inadequate folate intake is the major factor in the development of deficiency in spirit-drinking alcoholics. Beer is relatively folate-rich in some countries, depending on the technique used for brewing.

The drugs that inhibit DHF reductase include methotrexate, pyrimethamine, and trimethoprim. Methotrexate has the most powerful action against the human enzyme, whereas trimethoprim is most active against the bacterial enzyme and is likely to cause megaloblastic anemia only when used in conjunction with sulfamethoxazole in patients with preexisting folate or cobalamin deficiency. The activity of pyrimethamine is intermediate. The antidote to these drugs is a reduced form of folate, folinic acid (5-formyl-THF).

■ CONGENITAL ABNORMALITIES OF FOLATE METABOLISM

Some infants with congenital defects of folate enzymes (e.g., cyclohydrolase or methionine synthase) have had megaloblastic anemia.

DIAGNOSIS OF COBALAMIN AND FOLATE DEFICIENCIES

The diagnosis of cobalamin or folate deficiency has traditionally depended on the recognition of the relevant abnormalities in the peripheral blood and analysis of the blood levels of the vitamins.

■ COBALAMIN DEFICIENCY

Serum Cobalamin This is measured by an automated enzyme-linked immunosorbent assay (ELISA) or competitive-binding luminescence assay (CBLA). Normal serum levels range from 118–148 pmol/L (160–200 ng/L) to ~738 pmol/L (1000 ng/L). In patients with megaloblastic anemia due to cobalamin deficiency, the level is usually <74 pmol/L (100 ng/L). In general, the more severe the deficiency, the lower is the serum cobalamin level. In patients with spinal cord damage due to the deficiency, levels are very low even in the absence of anemia. Values between 74 and 148 pmol/L (100 and 200 ng/L) are regarded as borderline. They may occur, for instance, in pregnancy, in patients with megaloblastic anemia due to folate deficiency. They may also be due to heterozygous, homozygous, or compound heterozygous mutations of the gene *TCN1* that codes for HC (TC I). There is then no clinical or hematologic abnormality. The serum cobalamin level is sufficiently robust, cost-effective, and most convenient to rule out cobalamin deficiency in the vast majority of patients suspected of having this problem. However, problems have arisen with commercial CBLA assays involving IF in PA patients with intrinsic antibodies in serum. These antibodies may cause false normal serum cobalamin levels in up to 50% of cases tested. Where clinical indications of PA are strong, a normal serum cobalamin does not rule out the diagnosis. Serum MMA levels will be elevated in untreated PA (see below).

Folate deficiency, TC I (HC) deficiency, oral contraceptives, and multiple myeloma have all been associated with low serum cobalamin levels that do not indicate cobalamin deficiency. On the other hand, high serum cobalamin levels are usually due to raised serum TC I levels and can be due to the presence of liver, renal, or myeloproliferative diseases or to cancer of the breast, colon, or liver.

Serum Methylmalonate and Homocysteine In patients with cobalamin deficiency sufficient to cause anemia or neuropathy, the serum MMA level is raised. Sensitive methods for measuring MMA and homocysteine in serum have been introduced and recommended for the early diagnosis of cobalamin deficiency, even in the absence of hematologic abnormalities or subnormal levels of serum cobalamin. Serum MMA levels fluctuate, however, in patients with renal failure. Mildly elevated serum MMA and/or homocysteine levels occur in up to 30% of apparently healthy volunteers, with serum cobalamin levels up to 258 pmol/L (350 ng/L) and normal serum folate levels; 15% of elderly subjects, even with cobalamin levels >258 pmol/L (>350 ng/L), have this pattern of raised metabolite levels. These findings bring into question the exact cutoff points for normal MMA and homocysteine levels. It is also unclear at present whether these mildly raised metabolite levels have clinical consequences.

Serum homocysteine is raised in both early cobalamin and folate deficiency but may be raised in other conditions, for example, chronic renal disease, alcoholism, smoking, pyridoxine deficiency, hypothyroidism, and therapy with steroids, cyclosporin, and other drugs. Levels are also higher in serum than in plasma, in men than in premenopausal women, in women taking hormone replacement therapy or in oral contraceptive users, and in elderly persons and patients with several inborn errors of metabolism affecting enzymes in trans-sulfuration pathways of homocysteine metabolism. Thus, homocysteine levels must be carefully interpreted for diagnosis of cobalamin or folate deficiency.

Tests for the Cause of Cobalamin Deficiency Only vegans, strict vegetarians, or people living on a totally inadequate diet will become cobalamin deficient because of inadequate intake. Studies of cobalamin absorption once were widely used, but difficulty in obtaining radioactive cobalamin and ensuring that IF preparations are free of viruses has made these tests obsolete. Tests to diagnose PA include serum gastrin, which is raised; serum pepsinogen I, which is low in PA (90–92%) but also in other conditions; and gastric endoscopy. Tests for IF and parietal cell antibodies are also used, as well as tests for individual intestinal diseases.

Patients with atrophic gastritis may also have sufficient occult gastrointestinal blood loss to have iron deficiency as well as cobalamin deficiency. Iron deficiency may blunt the development of macrocytosis. Iron deficiency is much more common than cobalamin deficiency. In people older than age 60 years, cobalamin deficiency may accompany iron deficiency in 15–20% of cases. Thus, older patients diagnosed with iron-deficiency anemia should have cobalamin levels assessed, and those diagnosed with cobalamin deficiency should have their iron status assessed.

FOLATE DEFICIENCY

Serum Folate This is also measured by a chemiluminescence immunoassay or ELISA technique. In most laboratories, the normal range is from 11 nmol/L (2 μg/L) to ~82 nmol/L (15 μg/L). The serum folate level is low in all folate-deficient patients. It also reflects recent diet. Because of this, serum folate may be low before there is hematologic or biochemical evidence of deficiency. Serum folate rises in severe cobalamin deficiency because of the block in conversion of 5-MTHF to THF inside cells; raised levels have also been reported in the intestinal stagnant-loop syndrome due to absorption of bacterially synthesized folate.

Red Cell Folate The red cell folate assay is a valuable test of body folate stores. It is less affected than the serum assay by recent diet and traces of hemolysis. In normal adults, concentrations range from 880 to 3520 μmol/L (160–640 μg/L) of packed red cells. Subnormal levels occur in patients with megaloblastic anemia due to folate deficiency but also in nearly two-thirds of patients with severe cobalamin deficiency. False-normal results may occur if a folate-deficient patient has received a recent blood transfusion or if a patient has a raised reticulocyte count. Serum homocysteine assay is discussed earlier.

Tests for the Cause of Folate Deficiency The diet history is important. Tests for transglutaminase antibodies are performed to confirm or exclude gluten–induced enteropathy. If positive, duodenal biopsy is needed. An underlying disease causing increased folate breakdown should also be excluded.

TREATMENT

Cobalamin and Folate Deficiency

It is usually possible to establish which of the two deficiencies, folate or cobalamin, is the cause of the anemia and to treat only with the appropriate vitamin. In patients who enter the hospital severely ill, however, it may be necessary to treat with both vitamins in large doses once blood samples have been taken for cobalamin and folate assays and a bone marrow biopsy has been performed (if deemed necessary). Transfusion is usually unnecessary and inadvisable. If it is essential, packed red cells should be given slowly, one or two units only, with the usual treatment for heart failure if present. Occasionally, an excessive rise in platelets occurs after 1–2 weeks of therapy. Antiplatelet therapy, for example, aspirin, should be considered if the platelet count rises to >800 × 10^9/L.

COBALAMIN DEFICIENCY

It is usually necessary to treat patients who have developed severe cobalamin deficiency, as from PA, with lifelong regular cobalamin injections. In the UK, hydroxocobalamin is used; in the United States, cyanocobalamin. In a few instances, the underlying cause of cobalamin deficiency can be permanently corrected, for example, fish tapeworm, tropical sprue, or an intestinal stagnant loop that is amenable to surgery. The indications for starting cobalamin therapy are a well-documented megaloblastic anemia or other hematologic abnormalities and/or neuropathy due to the deficiency. Cobalamin should be given routinely to all patients who have had a total gastrectomy or ileal resection. Patients who have undergone gastric reduction for control of obesity or who are receiving long-term treatment with proton pump inhibitors should be screened and, if necessary, given cobalamin replacement.

Replenishment of body stores should be complete with six 1000-μg IM injections of hydroxocobalamin given at 3- to 7-day intervals. More frequent doses are usually used in patients with cobalamin neuropathy, but there is no evidence that they produce a better response. Allergic reactions are rare and may require desensitization or antihistamine or glucocorticoid cover. For maintenance therapy, 1000 μg hydroxocobalamin IM once every 3 months is satisfactory. Because of the poorer retention of cyanocobalamin, protocols generally use higher and more frequent doses, for example, 1000 μg cyanocobalamin IM, monthly, for maintenance treatment.

Because a small fraction of cobalamin can be absorbed passively through mucous membranes even when there is complete failure of physiologic IF-dependent absorption, large daily oral doses (1000–2000 μg) of cyanocobalamin may be used in PA for replacement (especially in Canada and Sweden) and maintenance of normal cobalamin status. Sublingual therapy has also been proposed for those in whom injections are difficult because of a bleeding tendency and who may not tolerate parenteral therapy. If oral therapy is used, it is important to monitor compliance, particularly with elderly, forgetful patients. This author prefers parenteral therapy for initial treatment for PA, particularly in severe anemia or if a neuropathy is present, and for maintenance.

Treatment of patients with subnormal serum cobalamin levels with a normal MCV and no hypersegmentation of neutrophils and a negative IF antibody is, however, problematic. Some (perhaps 15%) cases may be due to TC I (HC) deficiency. Homocysteine and/or MMA measurements may help, but in the absence of these tests and with otherwise normal gastrointestinal function, repeat serum cobalamin assay after 6–12 months may help one decide whether to start cobalamin therapy. Oral cyanocobalamin therapy with low doses (e.g., 50 μg daily) has a large role in treating patients thought to have food malabsorption of cobalamin.

Cobalamin injections are used in a wide variety of diseases, often neurologic, despite normal serum cobalamin and folate levels and a normal blood count and in the absence of randomized, double-blind, controlled trials. These conditions include multiple sclerosis and chronic fatigue syndrome/myalgic encephalomyelitis (ME). It seems probable that any benefit is due to the placebo effect of a usually painless, pink injection. In ME, oral cobalamin therapy, despite providing equally large amounts of cobalamin, has not been beneficial, supporting the view of the effect of the injections being placebo only.

FOLATE DEFICIENCY

Oral doses of 5–15 mg of folic acid daily are satisfactory, as sufficient folate is absorbed from these extremely large doses even in patients with severe malabsorption. The length of time therapy must be continued depends on the underlying disease. It is customary to continue therapy for about 4 months, when all folate-deficient red cells will have been eliminated and replaced by new folate-replete populations.

Before large doses of folic acid are given, cobalamin deficiency must be excluded and, if present, corrected; otherwise, cobalamin neuropathy may develop as the deficiency progresses despite a response of the anemia of cobalamin deficiency to folate therapy.

Long-term folic acid therapy is required when the underlying cause of the deficiency cannot be corrected and the deficiency is likely to recur, for example, in chronic dialysis or chronic hemolytic anemias. It may also be necessary in gluten-induced enteropathy that does not respond to a gluten-free diet. Where mild but chronic folate deficiency occurs, it is preferable to encourage improvement in the diet after correcting the deficiency with a short course of folic acid. In any patient receiving long-term folic acid therapy, it is important to measure the serum cobalamin level at regular (e.g., once-yearly) intervals to exclude the coincidental development of cobalamin deficiency.

Folinic Acid (5-Formyl-THF) This is a stable form of fully reduced folate. It is given orally or parenterally to overcome the toxic effects of methotrexate or other DHF reductase inhibitors, for example, trimethoprim or cotrimoxazole.

PROPHYLACTIC FOLIC ACID

Prophylactic folic acid is used in chronic dialysis patients and in parenteral feeds. Prophylactic folic acid has been used to reduce homocysteine levels to prevent cardiovascular disease and for cognitive function in the elderly, but there are no firm data to show any benefit.

Pregnancy In over 80 countries (but none in Europe), food is fortified at a level of 120–250 μg/100 g with folic acid (in grain, flour, or rice) to reduce the risk of NTDs. In all countries that have studied this, fortification has led to a lower prevalence of NTD pregnancies and births. Nevertheless, folic acid, 400 μg daily, should also be given as a supplement before and throughout pregnancy to prevent megaloblastic anemia and reduce the incidence of NTDs, even in countries with fortification of the diet. Most if not all the folic acid used in fortification and eaten over three meals a day will be converted during absorption to 5-MTHF. This compound at the levels achieved by fortification will not correct the anemia in cobalamin deficiency. Studies in the United States suggest that there is no increase in the proportion of individuals with low serum cobalamin levels and no anemia since food fortification with folic acid. It is unknown if there has been a change in incidence of cobalamin neuropathy, but no country has reported this since mandating fortification.

Data in early pregnancy show significant lack of compliance with taking folic acid supplements, emphasizing the need for food fortification. In women who have had a previous fetus with an NTD and others at high risk (e.g., diabetes, sickle cell anemia), a dose of 5 mg daily is recommended when pregnancy is contemplated and throughout the subsequent pregnancy.

Infancy and Childhood The incidence of folate deficiency is so high in the smallest premature babies during the first 6 weeks of life that folic acid (e.g., 1 mg daily) should be given routinely to those weighing <1500 g at birth and to larger premature babies who require exchange transfusions or develop feeding difficulties, infections, or vomiting and diarrhea.

The World Health Organization currently recommends routine supplementation with iron and folic acid in children in countries where iron deficiency is common and child mortality, largely due to infectious diseases, is high. However, some studies suggest that in areas where malaria rates are high, this approach may increase the incidence of severe illness and death. Even where malaria is rare, there appears to be no survival benefit.

MEGALOBLASTIC ANEMIA NOT DUE TO COBALAMIN OR FOLATE DEFICIENCY OR ALTERED METABOLISM

This may occur with many antimetabolite drugs (e.g., hydroxyurea, cytarabine, mercaptopurine, thioguanine) that inhibit DNA replication. Antiviral nucleoside analogues used in treatment of HIV infection may also cause macrocytosis and megaloblastic marrow changes. In the rare disease orotic aciduria, two consecutive enzymes in purine synthesis are defective. The condition responds to therapy with uridine, which bypasses the block. In thiamine-responsive megaloblastic anemia, there is a genetic defect in the high-affinity thiamine transporter (*SLC19A2*) gene. This causes defective RNA ribose synthesis through impaired activity of transketolase, a thiamine-dependent enzyme in the pentose cycle. This defect leads to reduced nucleic acid production. It may be associated with diabetes mellitus and deafness and the presence of many ringed sideroblasts in the marrow. The explanation is unclear for megaloblastic changes in the marrow in some patients with acute myeloid leukemia and myelodysplasia.

The most frequent causes of macrocytosis without megaloblastic changes are alcohol, liver disease, hypothyroidism, and pregnancy. Myelodysplasia, myeloma and other paraproteinemias, aplastic anemia, and smoking are other causes.

■ FURTHER READING

Berry RJ: Lack of historical evidence to support folic acid exacerbation of the neuropathy caused by vitamin B12 deficiency. Am J Clin Nutr 110:554, 2019.

Bunn HF: Vitamin B12 and pernicious anemia: The dawn of molecular medicine. N Engl J Med 370:773, 2014.

Del Bo C et al: Effect of two different sublingual dosages of vitamin B12 on cobalamin nutritional status in vegans and vegetarians with a marginal deficiency: A randomized controlled trial. Clin Nutr 38:575, 2019.

Green R: Vitamin B12 deficiency from the perspective of a practicing hematologist. Blood 129:2603, 2017.

Green R et al: Vitamin B12 deficiency. Nat Rev Dis Primers 3:17040, 2017.

Hesdorffer CS, Longo DL: Drug-induced megaloblastic anemia. N Engl J Med 373:1649, 2015.

Hoffbrand V: *The Folate Story: A Vitamin Under the Microscope*. Leicestershire, UK, Troubador Publishing, 2023.

Kancherla V et al: Preventing birth defects, saving lives, and promoting health equity: An urgent call to action for universal mandatory food fortification with folic acid. Lancet Global Health 10:e1053, 2022.

Ma F et al: Effects of folic acid and vitamin B12 alone and in combination on cognitive function and inflammatory factors in the elderly with mild cognitive impairment: A single blind experimental design. Curr Alzheimer Res 16:622, 2019.

Miller JW: Proton pump inhibitors, H2-receptor antagonists, metformin, and vitamin B-12 deficiency: Clinical implications. Adv Nutr 9:511S, 2018.

O'connor DMA et al: Low folate predicts cognitive decline: 8-year follow–up of 3140 older adults in Ireland. Eur J Clin Nutr 76:950, 2022.

Rogers LM et al: Global folate status in women of reproductive age: A systematic review with emphasis on methodological issues. Ann N Y Acad Sci 1431:35, 2018.

Salinas M et al: High frequency of anti-parietal cell antibody (APCA) and intrinsic factor blocking antibody (IFBA) in individuals with severe vitamin B12 deficiency: An observational study in primary care patients. Clin Chem Lab Med 58:424, 2020.

Wald NJ: Folic acid and neural tube defects: Discovery, debate and the need for policy change. J Med Screening 29:138, 2022.

Zaric BL et al: Homocysteine and hyperhomocysteinaemia. Curr Med Chem 26: 2948, 2019.

105

Hemolytic Anemias

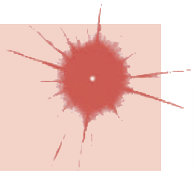

Lucio Luzzatto, Lucia De Franceschi

DEFINITIONS

Turnover is typical of all blood cells, including erythrocytes, that have a finite life span. Hence, a logical, time-honored classification of anemias is in three groups: (1) decreased production of red cells, (2) increased destruction of red cells, and (3) acute blood loss. Decreased production is covered in **Chaps. 102, 103, and 107**; acute blood loss in **Chap. 106**; increased destruction is covered in this chapter.

All patients who are anemic as a result of either increased destruction of red cells or acute blood loss have one important element in common: the anemia results from overconsumption of red cells from the peripheral blood, whereas the supply of cells from the bone marrow is normal (indeed, it is usually increased). However, with blood loss, as in acute hemorrhage, the red cells are physically lost *from* the body itself; this is fundamentally different from destruction of red cells *within* the body, as in hemolytic anemias (HAs).

With respect to primary etiology, HAs may be *inherited* or *acquired*; from a clinical point of view, they may be more *acute* or more *chronic*, and they may vary from mild to very severe; the site of hemolysis may be predominantly *intravascular* or *extravascular*. With respect to mechanisms, HAs may be due to *intracorpuscular* causes or to *extracorpuscular* causes **(Table 105-1)**. But before reviewing the individual types of HA, it is appropriate to consider what general features they have in common, in terms of clinical aspects and pathophysiology.

GENERAL CLINICAL AND LABORATORY FEATURES

The clinical presentation of a patient with anemia is greatly influenced in the first place by whether the onset is abrupt or gradual, and HAs are no exception. A patient with autoimmune HA or with favism may be a medical emergency, whereas a patient with mild hereditary spherocytosis (HS) or with cold agglutinin disease (CAD) may be diagnosed after years. This is due in large measure to the remarkable ability of the body to adapt to anemia when it is slowly progressing **(Chap. 66)**.

What differentiates HAs from other anemias is that the patient has signs and symptoms arising directly from hemolysis **(Table 105-2)**. At the clinical level, the main sign is *jaundice*; in addition, the patient may report discoloration of the urine. In many cases of HA, the spleen is enlarged because it is a preferential site of hemolysis; in some cases, the liver may be enlarged as well; and gallstones are common. In all severe congenital forms of HA, there may also be skeletal changes due to overactivity of the bone marrow: they are never as severe as in thalassemia major because ineffective erythropoiesis is less, or even absent. Since several forms of HA are inherited, it is important to include family history in the initial appraisal.

TABLE 105-1 Classification of Hemolytic Anemias[a]

	INTRACORPUSCULAR DEFECTS	EXTRACORPUSCULAR FACTORS
Inherited	Hemoglobinopathies Enzymopathies Membrane-cytoskeletal defects	Familial (atypical) hemolytic-uremic syndrome
Acquired	Paroxysmal nocturnal hemoglobinuria (PNH)	Mechanical destruction (microangiopathic) Toxic agents Drugs Infectious Autoimmune

[a] Hereditary causes correlate with intracorpuscular defects because these defects are due to inherited mutations; the one exception is PNH because the defect is due to an acquired somatic mutation. Conversely, acquired causes correlate with extracorpuscular factors because mostly these factors are exogenous; the one exception is familial hemolytic-uremic syndrome (HUS; often referred to as atypical HUS [aHUS]) because here an inherited abnormality permits complement activation triggered by exogenous factors to become excessive, with bouts of production of membrane attack complex capable of destroying normal red cells. Interestingly, in both PNH and aHUS, hemolysis is complement-mediated.

TABLE 105-2 Features Common to Most Patients with a Hemolytic Disorder

GENERAL EXAMINATION	JAUNDICE, PALLOR
Other physical findings	Spleen may be enlarged; bossing of skull in severe congenital cases
Hemoglobin level	From normal to severely reduced
MCV, MCH	Usually increased
Reticulocytes	Usually increased
Bilirubin	Almost always increased (mostly unconjugated)
LDH	Increased (up to 10× normal with intravascular hemolysis)
Haptoglobin	Reduced to absent if hemolysis is at least in part intravascular

Abbreviations: LDH, lactate dehydrogenase; MCH, mean corpuscular hemoglobin; MCV, mean corpuscular volume.

The laboratory features of HA are related to (1) hemolysis per se and (2) the erythropoietic response of the bone marrow. In most cases, hemolysis is largely extravascular, and it produces an increase in unconjugated bilirubin and aspartate aminotransferase (AST) in the serum; urobilinogen will be increased in both urine and stool. If hemolysis is mainly intravascular, the telltale sign is hemoglobinuria (often associated with hemosiderinuria); in the serum, there is free hemoglobin, lactate dehydrogenase (LDH) is increased, and haptoglobin is reduced. In contrast, the serum bilirubin level may be normal or only mildly elevated. The main sign of the erythropoietic response by the bone marrow is an increase in reticulocytes (a test all too often neglected in the initial workup of a patient with anemia). Usually the increase will be reflected in both the percentage of reticulocytes (the more commonly quoted figure) and in the absolute reticulocyte count (the more definitive parameter). The increased number of reticulocytes is associated with an increased mean corpuscular volume (MCV) in the blood count. On the blood smear, this is reflected in the presence of macrocytes; polychromasia is also present, and sometimes one sees nucleated red cells. In most cases, a bone marrow aspirate is not necessary in the diagnostic workup; if it is done, it will show erythroid hyperplasia. In practice, once an HA is suspected, specific tests will usually be required for a definitive diagnosis of a specific type of HA.

GENERAL PATHOPHYSIOLOGY

The mature red cell is the product of a developmental pathway that brings the phenomenon of differentiation to an extreme. An orderly sequence of events produces synchronous changes, whereby the gradual accumulation of a huge amount of hemoglobin in the cytoplasm (to a final level of 340 g/L, i.e., about 5 mM) goes hand in hand with the gradual loss of cellular organelles and of biosynthetic abilities. In the end, the erythroid cell undergoes a process that has features of apoptosis, including nuclear pyknosis and eventually extrusion of the nucleus. However, the final result is more altruistic than suicidal; the cytoplasmic body, instead of disintegrating, is now able to provide oxygen to all cells in the human organism for some remaining 120 days of the red cell life span.

As a result of this unique process of differentiation and maturation, intermediary metabolism is drastically curtailed in mature red cells **(Fig. 105-1)**; for instance, cytochrome-mediated oxidative phosphorylation has been lost with the loss of mitochondria (through a process of physiologic autophagy); therefore, there is no backup to anaerobic glycolysis, which in the red cell is the only provider of adenosine triphosphate (ATP). Also, the capacity of making protein has been lost with the loss of ribosomes. This places the cell's limited metabolic apparatus at risk, because if any protein component deteriorates, it cannot be replaced, as it would be in most other cells; and in fact, the activity of most enzymes gradually decreases as red cells age. At the same time, during their long time in circulation, various red cell

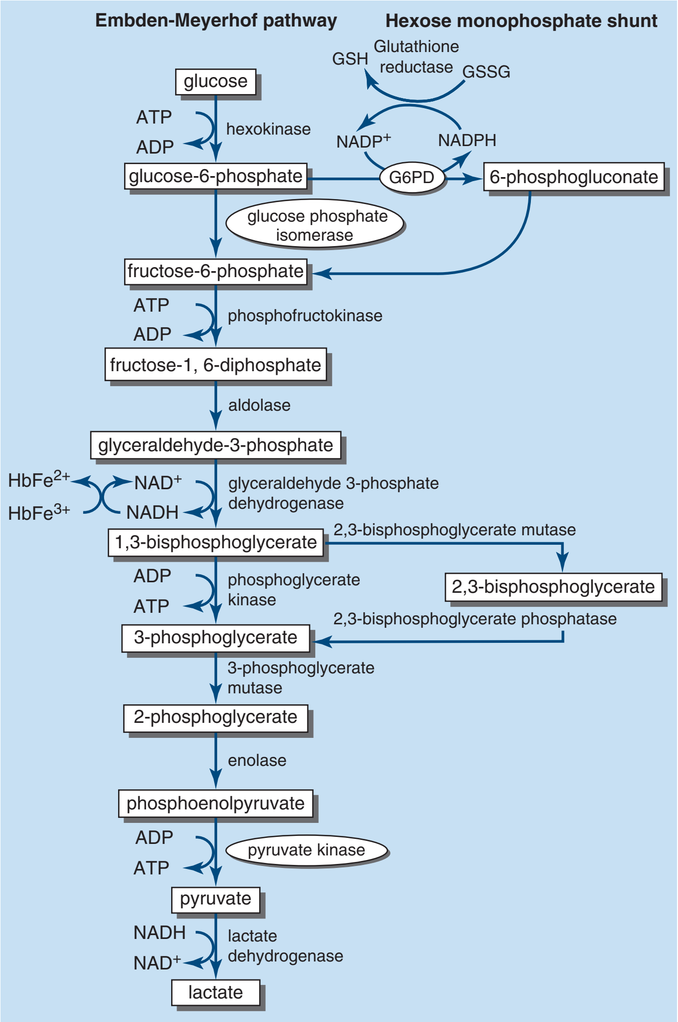

FIGURE 105-1 Red blood cell (RBC) metabolism. The Embden-Meyerhof pathway (glycolysis) generates ATP required for cation transport and for membrane maintenance. The generation of NADH maintains hemoglobin iron in a reduced state. The hexose monophosphate shunt generates NADPH that is used to reduce glutathione, which protects the red cell against oxidant stress; the 6-phosphogluconate, after decarboxylation, can be recycled via pentose sugars to glycolysis. Regulation of the 2,3-bisphosphoglycerate level is a critical determinant of oxygen affinity of hemoglobin. Enzyme deficiency states in order of prevalence: glucose-6-phosphate dehydrogenase (G6PD) > pyruvate kinase > glucose-6-phosphate isomerase > rare deficiencies of other enzymes in the pathway. The more common enzyme deficiencies are encircled.

components inevitably accumulate damage and become physically denser. The anion exchanger known as band 3 is the most abundant protein in the red cell membrane **(Fig. 105-2 and Table 105-3)**, with about 1.2 million molecules per red cell. As red cells age and become denser, probability is increased that a region of the band 3 molecule becomes exposed on the cell surface and contributes to creating an antigenic site recognizable by low-avidity naturally occurring anti–band 3 IgG antibodies. This process might be enhanced by the clustering of band 3 molecules favored by the antibody itself and by the binding of hemichromes arising from hemoglobin degradation. Senescent red cells thus become opsonized, and this is the signal for phagocytosis by macrophages in the spleen, in the liver, and elsewhere. This process may become accelerated in various ways in HA.

Another consequence of the relative simplicity of red cells is that they have a limited range of ways to manifest distress under hardship; in essence, any sort of metabolic failure will eventually lead either to structural damage to the membrane or to failure of the cation pump. In either case, the life span of the red cell is reduced, which is the definition of a *hemolytic disorder*. If the rate of red cell destruction exceeds the capacity of the bone marrow to produce more red cells, the hemolytic disorder will manifest as HA.

Thus, the essential pathophysiologic process common to all HAs is an increased red cell turnover; in many HAs, this is due at least in part to an acceleration of the senescence process described above. The gold standard for proving that the life span of red cells is reduced (compared to the normal value of ~120 days) is a *red cell survival* study, which can be carried out by labeling the red cells with ^{51}Cr and measuring the fall in radioactivity over several days or weeks (this classic test can now be replaced by a methodology using the nonradioactive isotope ^{15}N). If the hemolytic event is transient, it does not usually cause any long-term consequences, except for an increased requirement for erythropoietic factors, particularly folic acid. However, if hemolysis is recurrent or persistent, the increased bilirubin production favors the formation of gallstones. If a considerable proportion of hemolysis takes place in the spleen, as is often the case, splenomegaly may become increasingly a feature, and hypersplenism may develop, with consequent neutropenia and/or thrombocytopenia.

The increased red cell turnover has important consequences. In normal subjects, the iron from effete red cells is very efficiently recycled by the body; however, with chronic intravascular hemolysis, the persistent hemoglobinuria will cause considerable iron loss, needing replacement. With chronic extravascular hemolysis, the opposite problem, iron overload, is more common, especially if the patient needs frequent blood transfusions. Even without blood transfusion, when erythropoiesis is massively increased, the release of erythroferrone from erythroid cells suppresses hepcidin, causing increased iron absorption. In the long run, in the absence of iron-chelation therapy, iron overload will cause secondary hemochromatosis; this will cause damage particularly to the liver, eventually leading to cirrhosis, and to the heart muscle, eventually causing heart failure.

Compensated Hemolysis versus Hemolytic Anemia Red cell destruction is a potent stimulus for erythropoiesis, which is mediated by erythropoietin (EPO) produced by the kidney. This mechanism is so effective that in many cases the increased output of red cells from the bone marrow can fully balance an increased destruction of red cells. In such cases, we say that hemolysis is *compensated*. The pathophysiology of compensated hemolysis is similar to what we have just described, except there is no anemia. This notion is important from the diagnostic point of view, because a patient with a hemolytic condition, even an inherited one, may present without anemia; and it is also important from the point of view of management because compensated hemolysis may become "decompensated," i.e., anemia may suddenly appear in certain circumstances, for instance in pregnancy, folate deficiency, or renal failure interfering with adequate EPO production. Another general feature of chronic HAs is seen when any intercurrent condition, such as an acute infection, depresses erythropoiesis. When this happens, in view of the increased rate of red cell turnover, the effect will be predictably much more marked than in a person who does not have hemolysis. The most dramatic example is infection by parvovirus B19, which may cause a rather precipitous fall in hemoglobin—an occurrence sometimes referred to as *aplastic crisis*.

INHERITED HEMOLYTIC ANEMIAS

The red cell has three essential components: (1) hemoglobin, (2) the membrane-cytoskeleton complex, and (3) the metabolic machinery necessary to keep hemoglobin and the membrane-cytoskeleton complex in working order. Diseases caused by inherited abnormalities of hemoglobin, or hemoglobinopathies, are covered in **Chap. 103**. Here we will deal with diseases of the other two components.

Hemolytic Anemias due to Abnormalities of the Membrane-Cytoskeleton Complex The detailed architecture of the red cell membrane is complex, but its basic design is relatively simple (Fig. 105-2). The lipid bilayer incorporates phospholipids and cholesterol, and it is spanned by a number of proteins that have their hydrophobic transmembrane domain(s) embedded in the membrane; most of these proteins also extend to both the outside (extracellular domains) and the inside of the cell (cytoplasmic domains). Other proteins are tethered to the membrane through a glycosylphosphatidylinositol (GPI) anchor; these have only an extracellular domain. Membrane proteins include energy-dependent ion transporters, ion channels, receptors

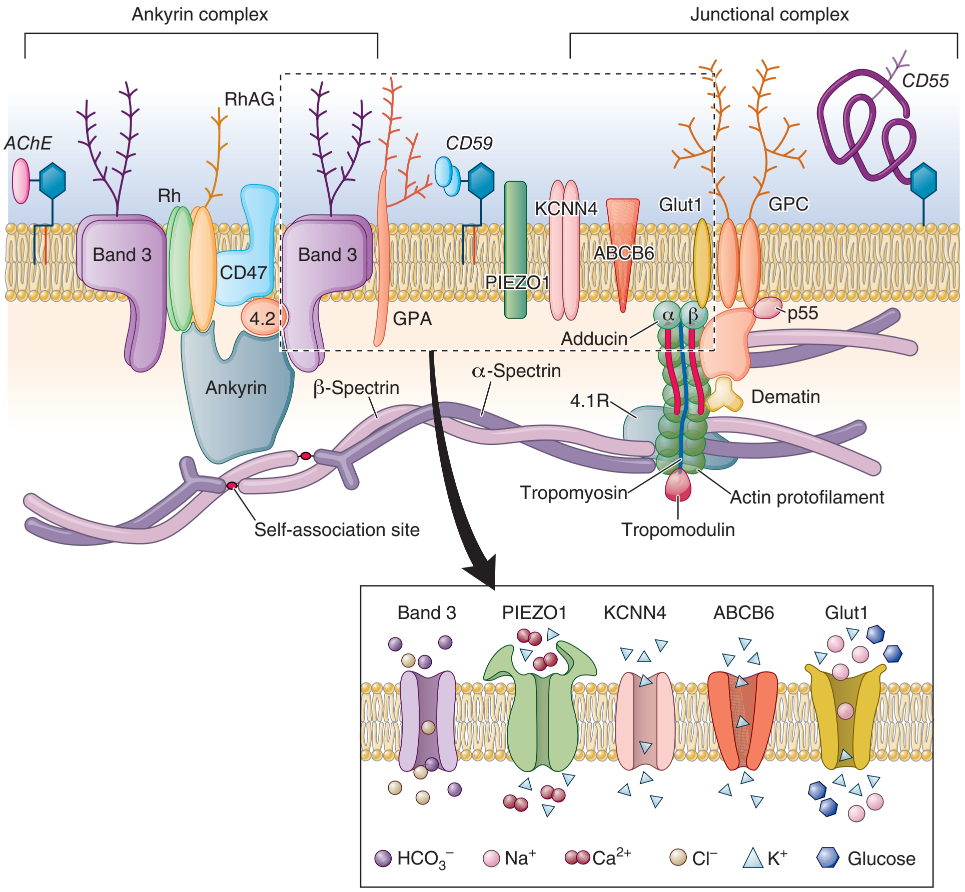

FIGURE 105-2 The red cell membrane and cytoskeleton schematic diagram. Within the membrane lipid bilayer, several integral membrane proteins are shown (see inset). Other proteins, e.g., acetylcholinesterase (AChE) and the two complement-regulatory proteins CD59 and CD55, are tethered to the membrane through the glycosylphosphatidylinositol (GPI) anchor: in these cases, the entire polypeptide chain is extracellular. Many of the membrane proteins bear polypeptide and/or carbohydrate red cell antigens. Underneath the membrane, the α-β spectrin dimers, which associate head-to-head into tetramers, together with actin and other proteins, form most of the cytoskeleton. The *ankyrin complex*, which also involves the band 4.2 protein, and the *junctional complex*, which involves the band 4.1 protein and dematin, connect the membrane to the cytoskeleton. The ankyrin complex provides mainly radial (also called vertical) connections; the junctional complex provides mainly tangential (also called horizontal) connections. Pathogenic changes in the former can cause spherocytosis, whereas pathogenic changes in the latter can cause elliptocytosis; pathogenic changes in spectrin can cause either. Branched lines symbolize carbohydrate moiety of proteins. The various molecules are obviously not drawn to the same scale. ***Inset.*** Schematic diagram of membrane transporters, abnormalities of which underlie channelopathies. Band 3 (anion exchanger 1 [AE1]) is the most abundant. Mutations of *SLC4A1* (encoding band 3) can inactivate anion exchange, causing cation leak or negatively affecting interactions with neighboring membrane proteins. PIEZO1 is a huge protein that is embedded in the membrane as a homo-trimer with a three-bladed, propeller-shaped structure; it is a mechano-sensitive cation channel. *KCNN4* encodes a Ca^{2+} activated K^+ channel, also known as the Gardos channel. *PIEZO1* mutations can be associated with abnormal Ca^{2+} entry, resulting in overactivation of the Gardos channel; this promotes K^+ efflux and red cell dehydration Mutations on KCC4 induce alterations of channel properties (kinetic or ion trafficking), resulting again in abnormal K^+ efflux and red cell dehydration. *ABCB6* encodes a mitochondrial porphyrin transporter, but some mutations can cause increased K^+ efflux from red cells and stomatocytosis. Glut1 facilitates glucose transport; it is encoded by *SLC2A1*, and mutations can block the glucose transport and induce Na^+, K^+ leakage. *(Modified from N Young et al: Clinical Hematology. Philadelphia, Elsevier, 2006; and from A Iolascon et al: Br J Haematol 187:13, 2019.)*

for complement components, and receptors for other ligands. The most abundant red cell membrane proteins are glycophorins and the so-called band 3, an anion transporter that is an integral membrane protein. The extracellular domains of many of these proteins are heavily glycosylated, and they carry antigenic determinants that correspond to blood groups. Underneath the membrane, and tangential to it, is a network of other proteins that make up the cytoskeleton. The main cytoskeletal protein is the spectrin tetramer, consisting of a head-to-head association of two α-spectrin-β-spectrin heterodimers. The cytoskeleton is linked to the membrane through the *ankyrin complex* (that includes also band 4.2) and the *junctional complex* (that includes adducin and band 4.1) (Fig. 105-2). These multiprotein complexes make membrane and cytoskeleton intimately connected to each other, thus supporting membrane stability and at the same time providing the erythrocyte with the important property of deformability.

The membrane-cytoskeleton complex has essentially three functions: it is an envelope for the red cell cytoplasm; it maintains the normal red cell shape; and it provides cross-membrane transport of electrolytes and of metabolites such as glucose and amino acids. In the membrane-cytoskeleton complex, the individual components are so intimately associated with each other that an abnormality of almost any of them will be disturbing or disruptive, causing mechanical instability of the membrane and/or reduced red cell deformability, ultimately causing hemolysis. These abnormalities are almost invariably inherited mutations; thus, diseases of the membrane-cytoskeleton complex belong to the category of inherited HAs. Before the red cells lyse, they often exhibit more or less specific changes that alter the normal biconcave disk shape. Thus, the majority of the diseases in this group have been known for over a century as hereditary spherocytosis (HS) and hereditary elliptocytosis (HE). More recently, a third morphologic entity, whereby on a blood smear the round-shaped central pallor of a red cell is replaced by a linear-shaped central pale area, has earned the name *stomatocytosis*; because this abnormal shape is related to abnormalities of channel molecules, the underlying disorders are also referred to as channelopathies. From an understanding of the molecular basis of these disorders, it has emerged (Table 105-3) that, although these disorders are predominantly monogenic, no one-to-one correlation exists between a certain gene and a certain disorder. Rather, what has been regarded as a single disorder (e.g., HS) can arise through mutation of one of several genes; conversely, what have been regarded as different disorders can arise through different mutations of the very same gene **(Fig. 105-3)**.

HEREDITARY SPHEROCYTOSIS This is most common among this group of HAs, with an estimated prevalence of 1:2000–1:5000 in populations of European ancestry. Its identification is credited to Minkowsky and Chauffard, who, at the end of the nineteenth century, reported families who had spherocytes in their peripheral blood **(Fig. 105-4A)**. In vitro studies revealed that the red cells were abnormally susceptible to lysis in hypotonic media; indeed, the increase in *osmotic fragility* became the main diagnostic test for HS. Today we know that HS, thus defined, is genetically heterogeneous; i.e., it can arise from a variety of mutations in one of several genes (Table 105-3). It has been also recognized that the inheritance of HS is not always autosomal dominant (with the patient being heterozygous); indeed, some of the most severe forms are instead autosomal recessive (with the patient being homozygous).

Clinical Presentation and Diagnosis The spectrum of clinical severity of HS is broad. Severe cases may present in infancy with severe anemia, whereas mild cases may present in young adults or even later in life. The main clinical findings are jaundice, an enlarged spleen, and often gallstones; indeed, it may be the finding of gallstones in a young person that triggers diagnostic investigations.

TABLE 105-3 Inherited Diseases of the Red Cell Membrane-Cytoskeleton Complex

GENE	CHROMOSOMAL LOCATION	PROTEIN PRODUCED	DISEASE(S) WITH CERTAIN MUTATIONS (INHERITANCE)	COMMENTS
SPTA1	1q22-q23	α-Spectrin	HS (recessive) HE (dominant)	Rare Mutations of this gene account for about 65% of HE. More severe forms may be due to coexistence of an otherwise silent mutant allele.
SPTB	14q23-q24.1	β-Spectrin	HS (dominant) HE (dominant)	Rare Mutations of this gene account for about 30% of HE, including some severe forms.
ANK1	8p11.2	Ankyrin	HS (dominant)	May account for majority of HS.
SLC4A1	17q21	Band 3; also known as AE (anion exchanger) or AE1	HS (dominant) Southeast Asia ovalocytosis (dominant) Stomatocytosis (cryohydrocytosis)	Mutations of this gene may account for about 25% of HS. Polymorphic mutation (deletion of nine amino acids); in heterozygotes, clinically asymptomatic and protective against *Plasmodium falciparum.* Certain specific missense mutations shift protein function from anion exchanger to cation conductance.
EPB41	1p33-p34.2	Band 4.1	HE (dominant)	Mutations of this gene account for about 5% of HE, mostly with prominent morphology but little/no hemolysis in heterozygotes; severe hemolysis in homozygotes.
EPB42	15q15-q21	Band 4.2	HS (recessive)	Mutations of this gene account for about 3% of HS.
RHAG	6p21.1-p11	Rhesus-associated glycoprotein	Chronic nonspherocytic hemolytic anemia (recessive)	Very rare; associated with total loss of all Rh antigens. One specific mutation in this gene entails loss of stomatin from the cell membrane, causing overhydrated stomatocytosis.
PIEZO1	16q23-q24	PIEZO1 (mechanosensitive ion channel component 1 channel)	Dehydrated hereditary stomatocytosis (dominant)	Also known as xerocytosis with pseudohyperkalemia. Patients may present with perinatal edema.
KCNN4	19q13.31	KCNN4 Intermediate conductance calcium-activated potassium channel protein 4 (Gardos channel)	Dehydrated hereditary stomatocytosis (dominant)	Clinical presentation similar to that of *PIEZO1* mutants.
ABCB6	2q35-q36	ATP-binding cassette subfamily B member 6	Familial pseudohyperkalemia (dominant)	Increased potassium leakage upon storage in blood bank condition: this can cause hyperkalemia in the recipient. *ABCB6* mutation is present in 0.3% of blood donors.
SLC2A1	1p34.2	GLUT1 glucose transporter	Overhydrated hereditary stomatocytosis	Associated with serious neurologic manifestations.

Note: PIEZO1, KCNN4, ABCB6, and *GLUT1* are channel molecules; conditions associated with mutations in the respective genes are appropriately named channelopathies.

Abbreviations: HE, hereditary elliptocytosis; HS, hereditary spherocytosis.

The variability in clinical manifestations that is observed among patients with HS is largely due to the different underlying molecular lesions (Table 105-3). Not only are mutations of several genes involved, but also different mutations of the same gene can give very different clinical manifestations. In milder cases, hemolysis is often compensated (see above), but changes in clinical expression may be seen even in the same patient because intercurrent conditions (e.g., pregnancy, infection) may cause decompensation. The anemia is usually normocytic with the characteristic morphology that gives the disease its name. An increased mean corpuscular hemoglobin concentration (MCHC >34 g/dL) and increased red cell distribution width (RDW >14%) associated with normal or slightly decreased MCV on an ordinary blood count report should raise the suspicion of HS. The spleen plays a key role in HS through a dual mechanism. On one hand, because HS red cells are less deformable, transit through the splenic circulation makes them more prone to vesiculate; on the other hand, like in many other HAs, the spleen itself is a major site of destruction through phagocytosis by macrophages.

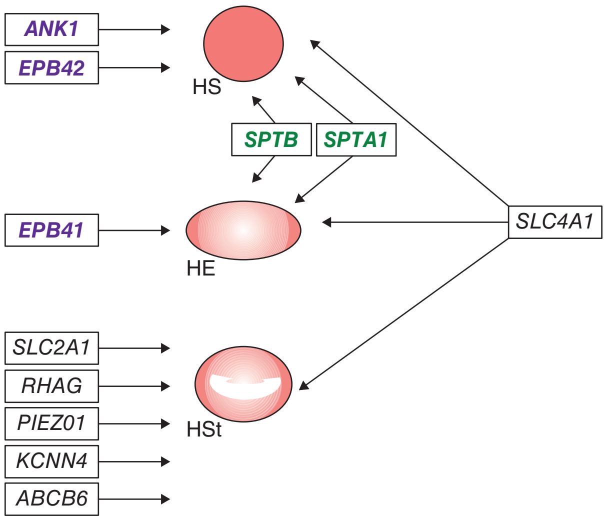

FIGURE 105-3 Hereditary spherocytosis (HS), hereditary elliptocytosis (HE), and hereditary stomatocytosis (HSt) are three morphologically distinct forms of congenital hemolytic anemia. It has emerged that each one can arise from mutation of one of several genes and that different mutations of the same gene can give one or another form. (See also Table 105-3.) Genes encoding membrane proteins are in *black*; genes encoding cytoskeleton proteins are in *green*; genes encoding proteins in the junctional and ankyrin complexes are in *purple*.

When there is a family history, it is usually easy to make a diagnosis based on features of HA and typical red cell morphology. However, family history may be negative for at least two reasons. First, the patient may have a de novo mutation, i.e., a mutation that has taken place in a germ cell of one of the patient's parents or early after zygote formation. Second, the patient may have a recessive form of HS (Table 105-3). In such cases, more extensive laboratory investigations are required, including osmotic fragility, the acid glycerol lysis test, the eosin-5′-maleimide (EMA)–binding test, sodium dodecyl sulfate (SDS)-gel electrophoresis of membrane proteins, and ektacytometry (testing red cell deformability as a function of shear stress at different osmolality); these tests are usually carried out in laboratories with special expertise in this area. Sometimes a definitive diagnosis can be obtained only by molecular studies demonstrating a mutation in one of the genes underlying HS (Table 105-3).

TREATMENT

Hereditary Spherocytosis

We do not have a causal treatment for HS; i.e., no way has yet been found to correct the basic defect in the membrane-cytoskeleton structure. Given the special role of the spleen in HS (see above), splenectomy is often beneficial. Current recommendations are to proceed with splenectomy at the age of 4–6 years in severe cases, to delay splenectomy until puberty in moderate cases, and to avoid splenectomy in mild cases. Partial splenectomy can be considered in certain cases, and it is helpful to know about the outcome of splenectomy in the patient's affected relatives. Before splenectomy, vaccination against encapsulated bacteria (*Neisseria meningitidis* and *Streptococcus pneumoniae*) is imperative; penicillin prophylaxis after splenectomy is controversial. Along with splenectomy, cholecystectomy should not be carried out automatically, but it should be carried out, usually by the laparoscopic approach, whenever it is clinically indicated, mainly when gallstones are symptomatic. The most severe cases of HS (estimated at <10%) are transfusion dependent, and in infants with severe HS, erythropoietin may prove to be transfusion sparing.

HEREDITARY ELLIPTOCYTOSIS HE is at least as heterogeneous as HS, both from the genetic point of view (Table 105-3, Fig. 105-3) and from the clinical point of view. The global incidence of HE is 1:2000–4000 subjects. Again, it is the shape of the red cells **(Fig. 105-4*B*)** that gives the name to the condition, but there is no direct correlation between the elliptocytic morphology and clinical severity. In fact, some mild or even asymptomatic cases may have nearly 100% elliptocytes (or ovalocytes). Indeed, the diagnosis of HE is generally incidental, because hemolysis may be compensated and there may be no anemia, although this may become evident in the course of infection. One particular in-frame deletion of nine amino acids in the *SLC4A1* gene encoding band 3 underlies the so-called Southeast Asia ovalocytosis (SAO): it is not a disease, but rather a polymorphism with a frequency of up to 5–7% in certain populations (e.g., Papua New Guinea, Indonesia, Malaysia, Philippines), presumably as a result of malaria selection. It is asymptomatic in heterozygotes and probably lethal in homozygotes. The cases of HE with the most severe HA are those with biallelic mutations of one of the genes involved (see Fig. 105-3), and these are said to have hereditary pyropoikilocytosis (HPP): here the instability of the cytoskeleton protein network may result from decreased tetramerization of spectrin dimers. The red cell volume is decreased (MCV: 50–60 fL), and all kinds of bizarre poikilocytes are seen on the blood smear **(Fig. 105-4*C*)**. HPP patients have splenomegaly and often benefit from splenectomy.

Channelopathies These rare conditions (see Fig. 105-3) are characterized by abnormalities in red cell ion content and alteration of erythrocyte volume. Cation leak can cause hyperkalemia; in some cases, this leak is accelerated in the cold (the resulting spuriously high serum K^+ is then referred to as pseudo-hyperkalemia). The less rare form, dehydrated stomatocytosis (DHS; also referred to as xerocytosis), is a (usually compensated) macrocytic hemolytic disorder, with increased MCHC (generally >36 g/dL) associated with mild jaundice. Mutations in either *PIEZO1*, encoding an ion channel activated by pressure (mechanoreceptor), or in *KCCN4*, encoding the Ca^{2+} activated K^+ channel (Gardos channel) have been recognized to cause DHS (see Table 105-3).

Another form is overhydrated stomatocytosis (OHS). OHS is also macrocytic (MCV >110 fL), but the MCHC is low (<30 g/dL). The underlying mutation is in the Rhesus gene *RHAG*, which encodes an ammonia channel. Yet other patients with stomatocytosis (Table 105-3) have mutations in *SLC4A1* (encoding band 3) and *SLC2A1* (encoding the glucose transporter GLUT1). Mutations of the latter are responsible for *cryohydrocytosis*, a channelopathy in which the red cells swell and burst when they are cooled. In vivo hemolysis can vary from relatively mild to quite severe. Familial hyperkalemia has been recently linked to mutations in *ABCB6*, resulting in abnormal cation leak with extracellular release of a large amount of K^+ (hyperkalemia). Mutations in *ABCB6* have been identified in almost 0.3% of blood donors. However, splenectomy is contraindicated in stomatocytosis due to the significant proportion of severe thromboembolic complications observed in splenectomized DHS patients.

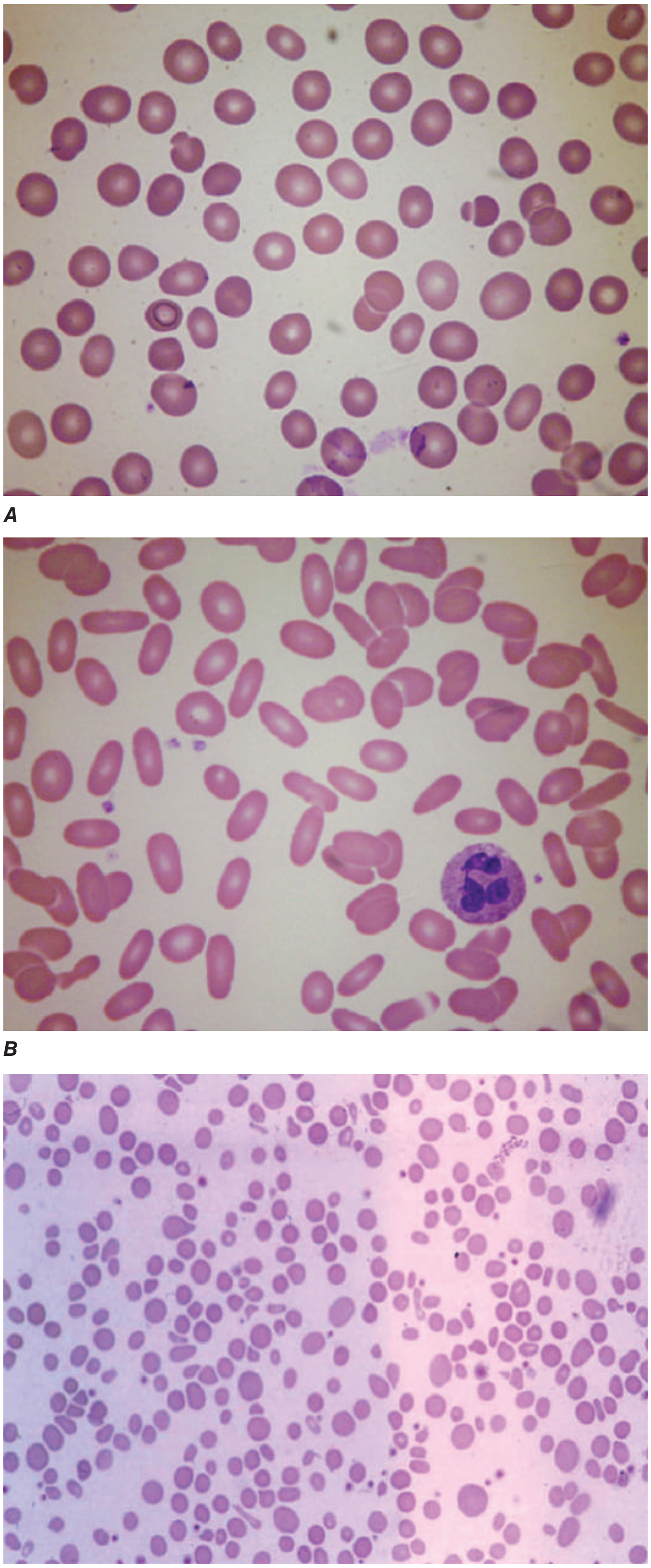

FIGURE 105-4 Peripheral blood smear from patients with membrane-cytoskeleton abnormalities. ***A.*** Hereditary spherocytosis. ***B.*** Hereditary elliptocytosis, heterozygote. ***C.*** Pyropoikilocytosis, with both alleles of the α-spectrin gene mutated.

Laser diffraction analysis, or *ektacytometry*, can measure the deformability of red blood cells subjected to either increasing shear stress or to an osmotic stress. This technique has been used extensively to investigate membrane-cytoskeleton abnormalities, and it can differentiate stomatocytosis from spherocytosis.

Enzyme Abnormalities When an important defect in a component of the membrane-cytoskeleton complex is present, hemolysis is a direct consequence of the fact that the very structure of the red cell is compromised. Instead, when one of the enzymes is defective, the consequences will depend on the precise role of that enzyme in the metabolic machinery of the red cell. This machinery has two main functions: (1) to provide energy in the form of ATP, and (2) to prevent oxidative damage to hemoglobin and to other proteins by providing sufficient reductive potential; the key molecule for this is NADPH, required for regeneration of glutathione (GSH) and for degradation of H_2O_2.

ABNORMALITIES OF THE GLYCOLYTIC PATHWAY Because red cells, in the course of their differentiation, have sacrificed not only their nucleus and their ribosomes but also their mitochondria, they rely exclusively on the anaerobic portion of the glycolytic pathway for producing ATP, most of which is required by the red cell for cation transport against a concentration gradient across the membrane. If this fails due to a defect of any of the enzymes of the glycolytic pathway (Table 105-4), the result will be hemolytic disease.

Pyruvate Kinase Deficiency Abnormalities of the glycolytic pathway are all inherited and all rare. Among them, deficiency of pyruvate kinase (PK) is the least rare, with an estimated prevalence in most populations of 1:10,000. However, recently, a polymorphic PK mutation (E277K) was found in some African populations with heterozygote frequencies of 1–7%, suggesting that this may be another malaria-related polymorphism. HA secondary to PK deficiency is an autosomal recessive disease (Fig. 105-5).

The clinical picture of homozygous (or biallelic) PK deficiency is that of an HA that often presents in the newborn with severe neonatal jaundice, requiring nearly always phototherapy and frequently exchange transfusion; the jaundice often persists, and it is often associated with reticulocytosis. The anemia is of variable severity; sometimes it is so severe as to require regular blood transfusion treatment, whereas sometimes it is mild, bordering on a nearly compensated hemolytic disorder. As a result, the diagnosis may be delayed: in some cases, it is made, for instance, in a young woman during her first pregnancy, when the anemia may get worse. The delay in diagnosis may be caused in part by the fact that the anemia is often remarkably well tolerated because the metabolic block at the last step in glycolysis causes an increase in 2,3-bisphosphoglycerate (or DPG; Fig. 105-1), a major effector of the hemoglobin-oxygen dissociation curve; thus, for a certain level of hemoglobin, the oxygen delivery to the tissues is enhanced, a remarkable compensatory feat.

TREATMENT

Pyruvate Kinase Deficiency

Until recently, the management of PK deficiency has been supportive. Folic acid supplements should be given constantly. Blood transfusion should be used as necessary, and iron chelation may be required even in some patients who, though not receiving blood transfusion, may be developing iron overload (see "General Pathophysiology" above). About one-half of patients sooner or later undergo splenectomy, which usually provides a modest but significant increase in hemoglobin (paradoxically, reticulocytes also often increase, because they were previously trapped in the

TABLE 105-4 Red Cell Enzyme Abnormalities Causing Hemolysis

ENZYME (ACRONYM)	GENE SYMBOL; CHROMOSOMAL LOCATION	PREVALENCE OF ENZYME DEFICIENCY (RANK)	CLINICAL MANIFESTATIONS EXTRA-RED CELL	COMMENTS
Glycolytic Pathway				
Hexokinase (HK)	*HK1*; 10q22	Very rare		May benefit from splenectomy; BMT[c]
Glucose-6-phosphate isomerase (G6PI)	*GPI*; 19q31.1	Rare (4); at least 60 cases reported[a]	NM, CNS	May benefit from splenectomy
Phosphofructokinase (PFK)[b]	*PFKM*; 12q13	Very rare	Myopathy; myoglobinuria	
Aldolase	*ALDOA*; 16q22-24	Very rare	Myopathy	
Triose phosphate isomerase (TPI)	*TPI1*; 12p13.31	Very rare	CNS (severe), NM	
Glyceraldehyde 3-phosphate dehydrogenase (GAPD)	*GAPDH*; 12p13.31	Very rare	Myopathy	
Bisphosphoglycerate mutase (BPGM)	*BPGM*; 7q33	Very rare		Erythrocytosis rather than hemolysis; some of the rare mutations are in the enzyme active site
Phosphoglycerate kinase (PGK)	*PGK1*; Xq21.1	Very rare	CNS, NM	May benefit from splenectomy; BMT[c]
Pyruvate kinase (PK)	*PKLR*; 1q22	Rare (2)[a]		May benefit from splenectomy; BMT[c]
Redox				
Glucose-6-phosphate dehydrogenase (G6PD)	*G6PD*; Xq28	Common (1)[a]	Very rarely granulocytes	In almost all cases, only AHA from exogenous trigger
Glutathione synthase	*GSS*; 20q11.22	Very rare	CNS	
Glutathione reductase	*GSR*; 8p12	Very rare	Cataracts	AHA from exogenous trigger (favism)
γ-Glutamylcysteine synthase	*GCLC*; 6p12.1	Very rare	CNS	Mutations affect catalytic subunit
Cytochrome b5 reductase	*CYB5R3*; 22q13.2	Rare	CNS	Methemoglobinemia rather than hemolysis
Nucleotide Metabolism				
Adenylate kinase (AK)	*AK1*; 9q34.11	Very rare	CNS	May benefit from splenectomy
Pyrimidine 5′ nucleotidase (P5N)	*NTSC3A*; 7p14.3	Rare (3)[a]		May benefit from splenectomy

[a] The numbers from (1) to (4) indicate the ranking order of these enzymopathies in terms of frequency. [b] PFK deficiency is associated with increased glycogen in muscle, and it is also known as glycogen storage disease type VII or Tarui's disease. [c] Occasional report of successful treatment of the hematologic manifestations by BMT.

Abbreviations: AHA, acquired hemolytic anemia; BMT, bone marrow transplantation; CNS, central nervous system; NM, neuromuscular.

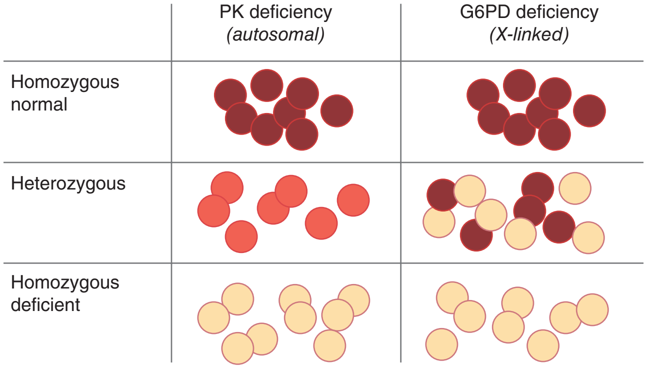

FIGURE 105-5 Different phenotypes of heterozygotes for red cell enzymopathies. In a heterozygote for deficiency of pyruvate kinase (PK), encoded by an autosomal gene (see Table 105-4), the level of enzyme is about one-half of normal in all red cells. Because this level of enzyme is sufficient, there are no clinical consequences, i.e., PK deficiency is recessive. In a heterozygote for deficiency of glucose-6-phosphate dehydrogenase (G6PD), encoded by an X-linked gene, the situation is quite different: X-chromosome inactivation generates red cell mosaicism, whereby some red cells are entirely normal and others are G6PD deficient. Therefore, G6PD deficiency is expressed in heterozygotes; it is not recessive.

spleen). Cholecystectomy may also be required. A major advance has been the introduction of *mitapivat,* an allosteric activator of PK, and the first drug for a red cell enzymopathy, approved on grounds of a significant increase in hemoglobin in one-half of PK-deficient patients—those in whom either one or both *PKLR* mutations are of the missense type. Some patients with severe disease have received bone marrow transplantation (BMT) from human leukocyte antigen (HLA)-identical PK-normal sibling. Prenatal diagnosis has been carried out in a mother who had already had an affected child. Rescue of inherited PK deficiency through lentiviral-mediated human PK gene transfer has been successful in mice and is currently undergoing a clinical trial in patients.

Other Glycolytic Enzyme Abnormalities All of these defects are rare to very rare (Table 105-4), and most of them cause HA with varying degrees of severity. It is not unusual for the presentation to be in the guise of severe neonatal jaundice, which may require exchange transfusion; if the anemia is less severe, it may present later in life, or it may even remain asymptomatic and be detected incidentally when a blood count is done for unrelated reasons. The spleen is often enlarged. When other systemic manifestations occur, they can involve the central nervous system (sometimes entailing severe intellectual disability, particularly in the case of triose phosphate isomerase deficiency), the neuromuscular system, or both (see Table 105-4). This is not altogether surprising if we consider that these are housekeeping genes, i.e., expressed in all tissues. The diagnosis of HA is usually not difficult, thanks to the triad of normo-macrocytic anemia, reticulocytosis, and hyperbilirubinemia. Enzymopathies should be considered in the differential diagnosis of any chronic Coombs-negative HA. Unlike with membrane disorders, in most cases of glycolytic enzymopathies, morphologic abnormalities are conspicuous by their absence. A definitive diagnosis can be made only by demonstrating the deficiency of an individual enzyme by quantitative assays; these are carried out in only a few specialized laboratories. If a particular molecular abnormality is already known in the family, then one could test directly for that defect at the DNA level, thus bypassing the need for enzyme assays. Of course, the time may be getting nearer when a patient will present with their exome already sequenced, and we will need to concentrate on which genes to look up within the file. The principles for the management of these conditions are similar as for PK deficiency. In isolated clinically severe cases of glycolytic enzyme abnormalities, BMT has been carried out successfully, although unfortunately, nonhematologic manifestations, if any, are not reversed.

ABNORMALITIES OF REDOX METABOLISM • Glucose-6-Phosphate Dehydrogenase (G6PD) Deficiency G6PD is a housekeeping enzyme critical in the redox metabolism of all aerobic cells (Fig. 105-1). In red cells, its role is even more critical because it is the only source of NADPH, which directly and via GSH defends these cells against oxidative stress **(Fig. 105-6)**. G6PD deficiency–related HA is a prime example of an HA due to interaction between an intracorpuscular cause and an extracorpuscular cause; indeed, in the vast majority of cases, hemolysis is triggered by an exogenous agent. Although the G6PD activity is decreased in most tissues of G6PD-deficient subjects, in other cells, the decrease is much less pronounced than in red cells, and it does not seem to impact on clinical expression.

GENETIC CONSIDERATIONS

The G6PD gene is X-linked, and this has important implications. First, because males have only one G6PD gene (i.e., they are hemizygous for this gene), they must be either normal or G6PD deficient. By contrast, females, who have two G6PD genes, can be either normal or deficient (homozygous) or intermediate (heterozygous). Second, as a result of the phenomenon of X chromosome inactivation, heterozygous females are genetic mosaics (see Fig. 105-5), with a highly variable ratio of G6PD-normal to G6PD-deficient cells and an equally variable degree of clinical expression; some heterozygotes can be just as affected as hemizygous males. The enzymatically active form of G6PD is either a dimer or a tetramer of a single protein subunit of 514 amino acids. G6PD-deficient subjects have been found invariably to have mutations in the coding region of the G6PD gene. Almost all of the over 230 different mutations known are single missense point mutations, entailing single amino acid replacements in the G6PD protein. In most cases, these mutations cause G6PD deficiency by decreasing the in vivo stability of the protein; thus, the physiologic decrease in G6PD activity that takes place with red cell aging is greatly accelerated. In some cases, an amino acid replacement can also affect the catalytic function of the enzyme. The genetic heterogeneity of G6PD deficiency is clinically important, and for the variants that are widespread, it also has public health implications. Therefore, the World Health Organization (WHO) has worked out a classification **(Table 105-5)**.

Among these mutations, those underlying chronic nonspherocytic hemolytic anemia (CNSHA; see below) are a discrete subset. This much more severe clinical phenotype can be ascribed in some cases to adverse qualitative changes (e.g., a decreased affinity for the substrate glucose-6-phosphate) or simply to the fact that the enzyme deficit is more extreme because of a more severe instability of the enzyme. For instance, a cluster of mutations map at or near the dimer interface, and clearly, they compromise severely the formation of the dimer.

Epidemiology G6PD deficiency is widely distributed in tropical and subtropical parts of the world (Africa, southern Europe, the Middle East, Southeast Asia, and Oceania) **(Fig. 105-7)** and wherever people from those areas have migrated. A conservative estimate is that at least 500 million people have a G6PD deficiency gene. In several of these areas, the frequency of a G6PD deficiency gene may be as high as 20% or more. It would be quite extraordinary for a trait that causes significant pathology to spread widely and reach high frequencies in many populations without conferring some biologic advantage. Indeed, G6PD is one of the best-characterized examples of genetic polymorphisms in the human species. Clinical field studies and in vitro experiments strongly support the view that G6PD deficiency has been selected by *Plasmodium falciparum* malaria because it confers a relative resistance against this highly lethal infection. As in other cases of balanced polymorphism, it is heterozygotes, therefore females, who are protected. Different G6PD variants (class B) underlie most of the prevalence of G6PD deficiency in different parts of the world. Examples of widespread variants are G6PD Mediterranean on the shores of that sea, in the Middle East, and elsewhere; G6PD A– in Africa, in the Middle East, and in southern Europe; G6PD Orissa in India; G6PD Viangchan and G6PD Mahidol in Southeast Asia; G6PD Kaiping and G6PD Canton in China; and G6PD Union worldwide. The heterogeneity of polymorphic G6PD variants is proof of their independent origin, further supporting the notion of selection by a common environmental

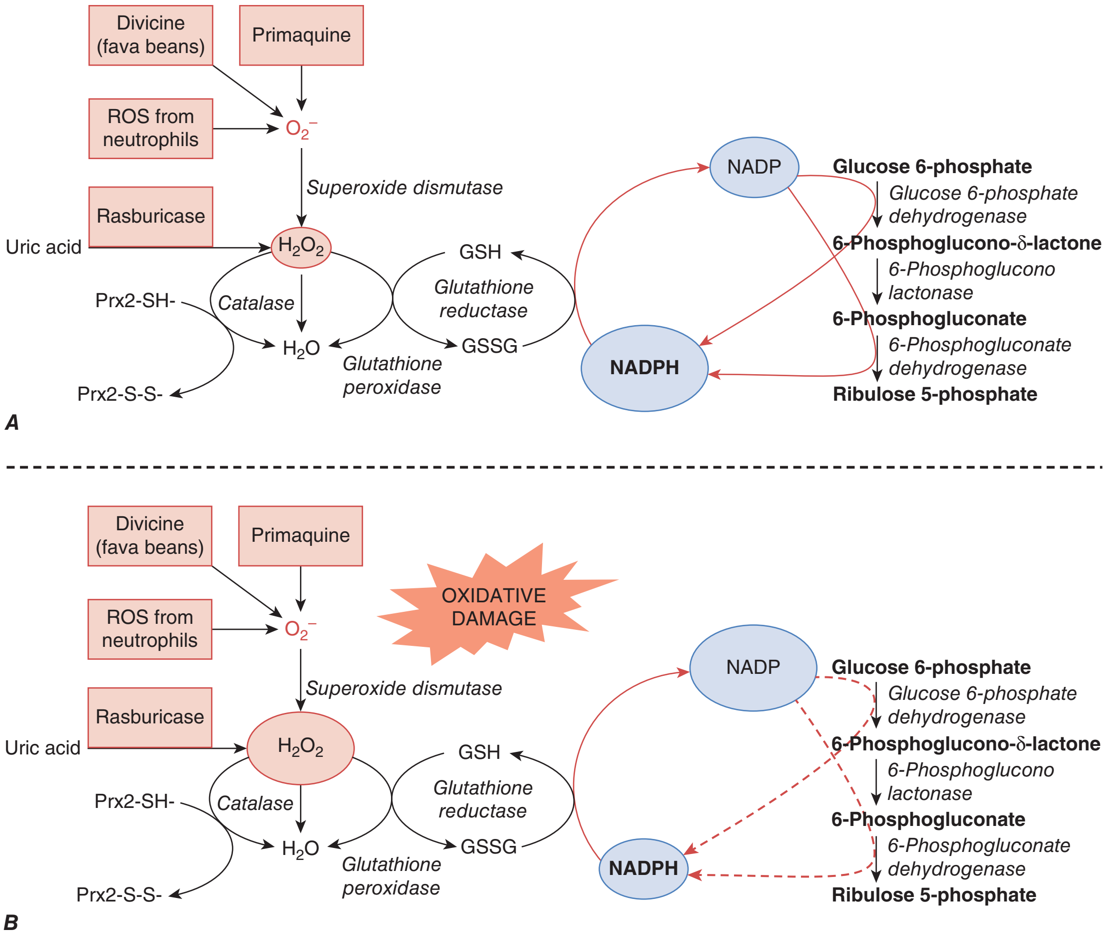
Divicine (fava beans)
Primaquine
ROS from neutrophils
O_2^-
Superoxide dismutase
Rasburicase
Uric acid
H_2O_2
Prx2-SH-
Catalase
H_2O
Prx2-S-S-
GSH
Glutathione reductase
GSSG
Glutathione peroxidase
NADP
NADPH
Glucose 6-phosphate
Glucose 6-phosphate dehydrogenase
6-Phosphoglucono-δ-lactone
6-Phosphoglucono lactonase
6-Phosphogluconate
6-Phosphogluconate dehydrogenase
Ribulose 5-phosphate
A
Divicine (fava beans)
Primaquine
ROS from neutrophils
O_2^-
OXIDATIVE DAMAGE
Superoxide dismutase
Rasburicase
Uric acid
H_2O_2
Prx2-SH-
Catalase
H_2O
Prx2-S-S-
GSH
Glutathione reductase
GSSG
Glutathione peroxidase
NADP
NADPH
Glucose 6-phosphate
Glucose 6-phosphate dehydrogenase
6-Phosphoglucono-δ-lactone
6-Phosphoglucono lactonase
6-Phosphogluconate
6-Phosphogluconate dehydrogenase
Ribulose 5-phosphate
B

FIGURE 105-6 The role of glucose-6-phosphate dehydrogenase (G6PD) in protecting red cells from oxidative damage. ***A.*** In G6PD-normal red cells, G6PD and 6-phosphogluconate dehydrogenase—two of the enzymes of the pentose phosphate pathway—provide ample supply of NADPH, which in turn regenerates glutathione (GSH) when this is oxidized by reactive oxygen species (e.g., O_2^- and H_2O_2). Thus when O_2^- (meant here to represent itself and other reactive oxygen species [ROS]) is produced by pro-oxidant compounds such as primaquine, or the glucosides in fava beans (divicine), or the oxidative burst of neutrophils, these ROS are rapidly neutralized; similarly, when rasburicase administered to degrade uric acid produces an equimolar amount of hydrogen peroxide, this is rapidly degraded by the combined action of glutathione peroxidase, catalase, and Prx2 (peroxiredoxin-2; all three mechanisms are NADPH dependent). ***B.*** In G6PD-deficient red cells, where the enzyme activity is reduced, NADPH production is limited, and it may not be sufficient to cope with the excess ROS generated by pro-oxidant compounds and the consequent excess hydrogen peroxide. This diagram also explains why a defect in GSH reductase has very similar consequences to G6PD deficiency.

agent, namely malaria, in keeping with the concept of convergent evolution (Fig. 105-7).

Clinical Manifestations The vast majority of people with G6PD deficiency remain clinically asymptomatic throughout their lifetime; however, all of them have an increased risk of developing neonatal jaundice (NNJ) and a risk of developing acute HA (AHA) when challenged by a number of oxidative agents. NNJ related to G6PD deficiency is rarely present at birth; the peak incidence of clinical onset is between day 2 and day 3, and in most cases, the anemia is not severe. However, NNJ can be very severe in some G6PD-deficient babies, especially in association with prematurity, infection, and/or environmental factors (e.g., naphthalene-camphor balls, which may be used in babies' bedding and clothing); and the risk of severe NNJ is also increased by the coexistence of a monoallelic or biallelic mutation in the uridyl transferase gene (*UGT1A1*; the same mutations are associated with Gilbert's syndrome). It is imperative to manage promptly NNJ associated with G6PD deficiency because it can produce kernicterus and permanent neurologic damage.

TABLE 105-5 Current World Health Organization Classification of Glucose 6-Phosphate Dehydrogenase (G6PD) Variants

G6PD VARIANT CLASS	MEDIAN OF G6PD ACTIVITY (% OF NORMAL)	ASSOCIATED CLINICAL MANIFESTATIONS
A[a]	<20%[b]	Chronic hemolytic anemia
B[c]	<45%	Neonatal jaundice; acute hemolytic anemia triggered by certain medicines, fava beans, and infections
C[a]	>60%	None reported
U[d]	Any	Uncertain clinical significance

[a] Class A corresponds to former class I; class C corresponds to former class IV. [b] A variant with <20% activity will be in class A only if it is associated with chronic hemolytic anemia. [c] In view of the extensive overlap in enzyme activity and in clinical expression of variants in former class II and class III, these have been merged into class B. [d] A temporary assignment for variants for which there is currently insufficient information regarding clinical manifestations.

AHA can develop as a result of three types of triggers: (1) fava beans, (2) infections, and (3) drugs (Table 105-6). Typically, a hemolytic attack starts with malaise, weakness, and abdominal or lumbar pain. Within a time frame of several hours to 2–3 days, the patient develops jaundice and often dark urine. The onset can be extremely abrupt, especially with favism in children. The anemia is moderate to extremely severe, usually normocytic and normochromic, and due partly to intravascular hemolysis; hence, it is associated with hemoglobinemia, hemoglobinuria, high LDH, and low or absent plasma haptoglobin. The blood film shows anisocytosis, polychromasia, and spherocytes; in addition, the most typical feature of G6PD deficiency is the presence of bizarre poikilocytes, with red cells that appear to have unevenly distributed hemoglobin ("hemighosts") and red cells that appear to have had parts of them bitten away ("bite cells" or "blister cells") (Fig. 105-8). A classical test, now rarely carried out, is supravital staining with methyl violet, which, if done promptly, reveals the presence of Heinz bodies (consisting of precipitates of denatured hemoglobin and hemichromes), which are regarded as a signature of oxidative damage to red cells (they are

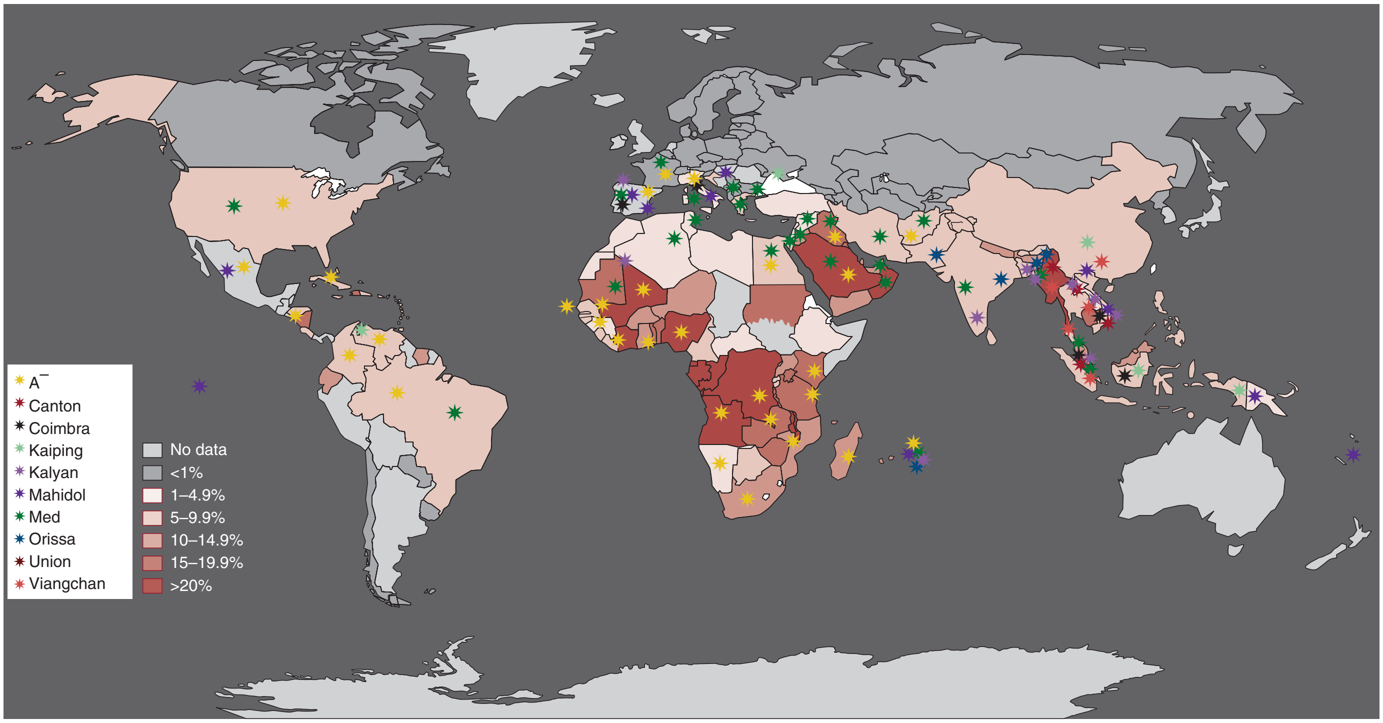

FIGURE 105-7 Epidemiology of glucose-6-phosphate dehydrogenase (G6PD) deficiency throughout the world. Each country on the map is shaded in a color based on the best estimate of the mean frequency of G6PD deficiency allele(s) in that country (this is the same as the frequency of G6PD-deficient males). The small panel on the left gives the key to color shadings corresponding to each country. The larger panel gives a color-coded list of 10 common G6PD variants associated with G6PD deficiency: *asterisk-shaped symbols* in the corresponding colors are shown in the countries where these variants have been observed (for graphic reasons, symbols could not be inserted in all countries). *(Reproduced with permission from L Luzzatto, M Ally, R Notaro. Glucose-6-phosphate dehydrogenase deficiency. 136:1225, 2020.)*

also seen with unstable hemoglobins). Since there is also a substantial component of extravascular hemolysis, unconjugated bilirubin is high and there is often clinical icterus. The most serious threat from AHA in adults is the development of acute renal failure (this is exceedingly rare in children). Once the threat of acute anemia is over and in the absence of comorbidity, full recovery from AHA associated with G6PD deficiency is the rule.

TABLE 105-6 Drugs That Carry Risk of Clinical Hemolysis in Persons with Glucose 6-Phosphate Dehydrogenase Deficiency

	RISK OF ACUTE HEMOLYTIC ANEMIA		
DRUG CLASS	**HIGH**	**MEDIUM TO LOW**	
Antimalarials	Primaquine Tafenoquine	Chloroquine Hydroxychloroquine Quinine	
Sulphonamides/ sulphones	Dapsone	Sulfadimidine Sulfamethoxazole Sulfasalazine	
Antibacterial/ antibiotics		Chloramphenicol Ciprofloxacin Cotrimoxazole Nalidixic acid Nitrofurantoin Norfloxacin *p*-Aminosalicylic acid	
Anti-helmint		Niridazole	
Antipyretic/ analgesics	Acetylsalicylic acid high dose (>3 g/d)	Acetaminophen Acetanilide Phenacetin Phenazopyridine	
Other	Rasburicase Pegloticase Methylene blue Toluidine blue	Ascorbic acid (>1 g) Doxorubicin Probenecid Vitamin K analogues	

It was primaquine (PQ)-induced AHA that led to the discovery of G6PD deficiency, but this drug has not been very prominent subsequently because it is not necessary for the treatment of life-threatening *P. falciparum* malaria. Today there is a revival in the use of PQ for two reasons. First, it is the only effective agent for eliminating the gametocytes of *P. falciparum* (thus preventing further transmission): a small single dose (0.25 mg/kg) is required, and it is safe for G6PD-deficient persons. Second, a 14-day course of PQ is the standard treatment for eliminating the hypnozoites of *Plasmodium vivax* (thus preventing endogenous relapse) **(Chap. 231)**. In countries aiming to eliminate malaria, there may be a call for mass administration of PQ; this ought

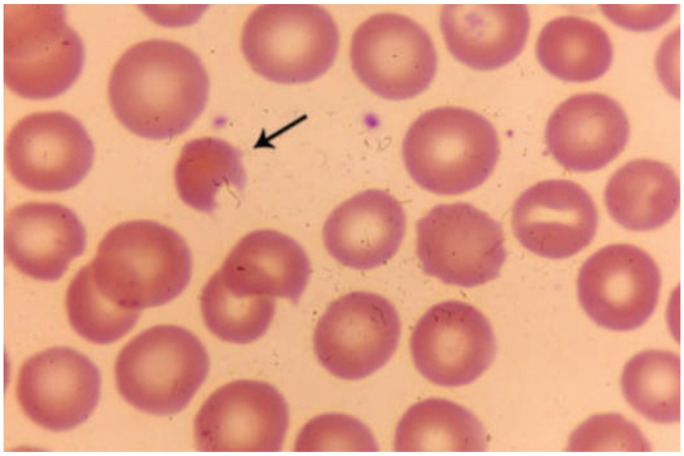

FIGURE 105-8 Peripheral blood smear from a glucose-6-phosphate dehydrogenase (G6PD)-deficient boy experiencing hemolysis. Note the red cells that are misshapen and called "bite" cells. *(From MA Lichtman et al: Lichtman's Atlas of Hematology: http://www.accessmedicine.com. Copyright © The McGraw-Hill Companies, Inc. All rights reserved.)*

to be associated with G6PD testing. At the other end of the historic spectrum, the latest additions to the list of potentially hemolytic drugs (Table 105-6) are rasburicase and pegloticase; again, G6PD testing ought to be made mandatory before giving either of these drugs because fatal cases have been reported upon using one of these drugs, which generate hydrogen peroxide, in newborns with kidney injury and in children and adults with tumor lysis syndrome.

Although drug-induced AHA has been prominent in the study of G6PD deficiency, the most common clinical manifestations are in fact NNJ and favism, both of which are of public health importance in many populations. Contrary to beliefs that are still widespread, fava bean pollen inhalation does not cause favism, and other beans are safe.

A very small minority of subjects with G6PD deficiency, those who have a class A variant, suffer from CNSHA of variable severity. The patient is nearly always a male, usually with a history of NNJ, who may present with anemia, unexplained jaundice, or gallstones later in life. The spleen may be enlarged. The severity of anemia ranges in different patients from borderline to transfusion dependent. The anemia is usually normo-macrocytic, with reticulocytosis. Bilirubin and LDH are increased. Although hemolysis is, by definition, chronic in these patients, they are also vulnerable to acute oxidative damage, and therefore, the same agents that can cause AHA in people with the ordinary type of G6PD deficiency will cause severe exacerbations in people with CNSHA associated with G6PD deficiency. In some cases of CNSHA, the deficiency of G6PD is so severe in granulocytes that it limits their capacity to produce an oxidative burst, with consequent increased susceptibility to some bacterial infections.

Laboratory Diagnosis The suspicion of G6PD deficiency can be confirmed by semiquantitative methods often referred to as screening tests, which are suitable for population studies and can correctly classify male subjects, in the steady state, as G6PD normal or G6PD deficient. However, in clinical practice, a diagnostic test is usually needed when the patient has had a hemolytic attack, whereby the oldest, most G6PD-deficient red cells have been selectively destroyed, and young red cells, having higher G6PD activity, are being released into the circulation. Under these conditions, only a quantitative test can give a definitive result. In males, this test will identify normal hemizygotes and G6PD-deficient hemizygotes; among females, some heterozygotes will be missed, but those who are at most risk of hemolysis will be identified. Of course, G6PD deficiency also can be diagnosed by DNA testing. Currently easy-to-use point-of-care tests for G6PD deficiency are becoming available, geared especially to the prospect of mass administration of PQ or of the newly introduced derivative tafenoquine.

TREATMENT

G6PD Deficiency

The AHA of G6PD deficiency is largely preventable by avoiding exposure to triggering factors of previously screened subjects. Of course, the practicability and cost-effectiveness of screening depend on the prevalence of G6PD deficiency in each individual community. Favism is entirely preventable in G6PD-deficient subjects by not eating fava beans. Drug-induced hemolysis can be prevented by testing for G6PD deficiency before prescribing; in many cases, one can use alternative drugs. When AHA develops and once its cause is recognized, no specific treatment is needed in most cases. However, if the anemia is severe, it may be a medical emergency, especially in children, requiring immediate action, including blood transfusion. This has been the case with an antimalarial drug combination containing dapsone (called Lapdap, introduced in 2003) that has caused severe acute hemolytic episodes in children with malaria in several African countries; after a few years, the drug was taken off the market. If there is acute renal failure, hemodialysis may be necessary, but if there is no previous kidney disease, recovery is the rule. The management of NNJ associated with G6PD deficiency is no different from that of NNJ due to other causes.

In cases with CNSHA, if the anemia is not severe, regular folic acid supplements and regular hematologic surveillance will suffice. It will be important to avoid exposure to potentially hemolytic drugs, and blood transfusion may be indicated when exacerbations occur, mostly in concomitance with intercurrent infection. In rare patients, regular blood transfusions may be required, in which case appropriate iron chelation should be instituted. Unlike in HS, there is no evidence of selective red cell destruction in the spleen; however, in practice, splenectomy has proven beneficial in severe cases.

Other Abnormalities of the Redox System As mentioned previously, GSH is a key player in the defense against oxidative stress. Inherited defects of GSH metabolism are exceedingly rare, but each one can give rise to chronic HA (Table 105-4). A rare, peculiar, and severe but usually self-limited HA occurring in the first month of life, called *infantile poikilocytosis*, may be associated with deficiency of glutathione peroxidase (GSHPX) due not to an inherited abnormality, but to transient nutritional deficiency of selenium, an element essential for the activity of GSHPX.

PYRIMIDINE 5′-NUCLEOTIDASE (P5N) DEFICIENCY P5N is a key enzyme in the catabolism of nucleotides arising from the degradation of nucleic acids that takes place in the final stages of erythroid cell maturation. How exactly its deficiency causes HA is not well understood, but a highly distinctive feature of this condition is a morphologic abnormality of the red cells known as *basophilic stippling*. The condition is rare, but it probably ranks third in frequency among red cell enzyme defects (after G6PD deficiency and PK deficiency). The anemia is lifelong, is of variable severity, and may benefit from splenectomy.

Familial (Atypical) Hemolytic-Uremic Syndrome (aHUS)

This term is used to designate a group of rare disorders, mostly affecting children, characterized by microangiopathic HA with presence of fragmented erythrocytes in the peripheral blood smear, thrombocytopenia (usually mild), and acute renal failure. (The word *atypical* in this phrase should be consigned to history; it was introduced originally to distinguish this condition from the hemolytic-uremic syndrome [HUS] caused by infection with *Escherichia coli* producing the Shiga toxin, regarded as *typical.*) The disease is caused by dysregulation of the complement alternative pathway: in some cases, this results from anti–factor H autoantibodies, but in the majority of cases, it is a consequence of mutations in genes encoding proteins that are components or regulators of the complement system, i.e., *C3* (encoding complement component C3), *CFB* (encoding complement factor B), *CFH* (encoding complement factor H), *CD46* (encoding the membrane cofactor protein), *CFI* (encoding complement factor I), *THBD* (encoding thrombomodulin), and rarely others. Patients may have abnormalities in one, two, or even three of the above genes. Pathogenic variants of these genes predispose to HUS with autosomal dominant inheritance; because clinical aHUS requires a triggering factor, most commonly infection, the penetrance is incomplete. Thus, whereas all other inherited HAs are due to intrinsic red cell abnormalities, this group is unique in that hemolysis results from an inherited defect external to red cells (Table 105-1). Because the regulation of the complement cascade has considerable redundancy, in the steady state, any of the above abnormalities can be tolerated. However, when an intercurrent infection or some other trigger briskly activates complement, the deficiency of one of the complement regulators becomes critical. Endothelial cells get damaged, especially in the kidney. At the same time, and partly as a result of this, there will be brisk hemolysis; thus, the more common Shiga toxin–related HUS **(Chap. 172)** can be regarded as a phenocopy of aHUS. aHUS is a severe disease; before anticomplement therapy was available, mortality was up to 15% in the acute phase, and up to 50% of cases progressed to end-stage renal disease (ESRD). Not infrequently, aHUS undergoes spontaneous remission. Because it is an inherited abnormality, it is not surprising that, given renewed exposure to a trigger, the syndrome will tend to recur; when it does, the prognosis is always serious. The traditional treatment has been plasma exchange, which will supply the deficient complement regulator and clear complement activation products. This has changed since the introduction of the anti-C5 complement inhibitor eculizumab (see "Paroxysmal Nocturnal Hemoglobinuria"), which was found to greatly ameliorate the microangiopathic picture, with improvement in platelet counts and in renal function, thus abrogating the need for plasma exchange, which is not always effective and not free of complications. Once a full remission is obtained, with urine dipstick negative for hemoglobinuria, within 3–6 months, eculizumab can be discontinued. Given the genetic background of aHUS, relapses are possible. Fortunately, patients who relapsed after discontinuing eculizumab have responded again. No current evidence supports continuing eculizumab indefinitely.

For the diagnosis of inherited red cell abnormalities, including those of the membrane-cytoskeleton, channelopathies, and enzymopathies, DNA sequencing of a gene panel (e.g., next-generation sequencing [NGS]) has become increasingly popular. This approach (which may or may not be more expensive) has the advantage of being potentially comprehensive and of providing definitive data. On the other hand, the mutations identified may be of uncertain significance, in which case, the diagnosis must be still confirmed by conventional methodology. Thus, the use of NGS is specially useful with unsolved cases of hereditary HAs.

ACQUIRED HEMOLYTIC ANEMIA

Mechanical Destruction of Red Cells Although red cells are characterized by the remarkable deformability that enables them to squeeze through capillaries narrower than themselves for thousands of times in their lifetime, there are at least two situations in which they succumb to shear, if not to wear and tear; the result is intravascular hemolysis, resulting in hemoglobinuria (Table 105-7). One situation is acute and self-inflicted, *march hemoglobinuria*. Why sometimes a marathon runner may develop this complication, whereas on another occasion, this does not happen, we do not know (perhaps their footwear needs attention). A similar syndrome may develop after prolonged barefoot ritual dancing or intense playing of bongo drums. The other situation is chronic and iatrogenic (it has been called *microangiopathic hemolytic anemia*). It takes place in patients with prosthetic heart valves, especially when paraprosthetic regurgitation is present. If the hemolysis consequent on mechanical trauma to the red cells is mild, and if the supply of iron is adequate, the loss may be largely compensated; if more than mild anemia develops, reintervention to correct regurgitation may be required.

Infection By far, the most frequent infectious cause of HA in endemic areas is malaria (Chap. 231). In other parts of the world, the most frequent direct cause is probably Shiga toxin–producing *E. coli* O157:H7, now recognized as the main etiologic agent of HUS, which is more common in children than in adults (Chap. 166). Life-threatening intravascular hemolysis, due to a toxin with lecithinase activity, occurs with *Clostridium perfringens* sepsis, particularly following open wounds, septic abortion, or as a disastrous accident due to a contaminated blood unit. Rarely, and if at all in children, HA is seen with sepsis or endocarditis from a variety of organisms. In addition, bacterial and viral infections can cause HA by indirect mechanisms (see Table 105-7).

Immune Hemolytic Anemias These can arise through at least two distinct mechanisms. First, when an antibody directed against a certain molecule (e.g., a drug) reacts with that molecule, red cells may get caught in the reaction (the so-called innocent bystander mechanism; see later section "Hemolytic Anemia from Toxic Agents and Drugs"), whereby they are damaged or destroyed. Second, and more frequently, a true autoantibody is directed against a red cell antigen, i.e., a molecule present on the surface of red cells. Autoimmune HAs have been originally classified into two types, depending on the thermal amplitude of the autoantibodies involved; this classification is valid because the two types have different pathophysiologic and clinical features.

AUTOIMMUNE HEMOLYTIC ANEMIA, WARM TYPE (WAIHA; FOR SIMPLICITY, WE WILL USE THE ACRONYM AIHA) AIHA has an estimated incidence in the United States of about 1–3:100,000 per year, and a prevalence of 17:100,000. AIHA can be serious because even with appropriate management the mortality is ~5–10%.

Clinical Features and Diagnosis The onset is often abrupt and can be dramatic. The hemoglobin level may drop, within days, to as low as 4 g/dL; the massive red cell removal will produce jaundice, and sometimes the spleen is enlarged. When this triad is present, the suspicion of AIHA must be high. The reticulocyte count is typically elevated, except when erythroid precursors are also targeted by the autoantibody attack. LDH may also be elevated. In some cases, AIHA can be associated, on first presentation or subsequently, with autoimmune thrombocytopenia. This double autoimmune condition, referred to as Evans syndrome, may be a manifestation of common variable immune deficiency, and in children, it may suggest one of several primary immune deficiency syndromes. Evans syndrome signals high-risk disease. Other predictors of the outcome and of the probability of relapse of AIHA are severe anemia (hemoglobin <6 g/dL), certain characteristics of the antibody, acute renal failure, and infection.

There are few situations in hematology where one laboratory test is as informative as the direct antiglobulin test developed in 1945 by R. R. A. Coombs and known since then by this name. The currently recommended version of this test uses in the first instance a "broad-spectrum" reagent, i.e., one that will detect not only immunoglobulins (Ig) but also complement (C) components (usually C3 fragments) bound to the surface of the patient's red cells. If the test is positive, it is, almost on its own, diagnostic of AIHA. False positives may occur as a result of previous blood transfusion or, much more rarely, following organ transplantation procedures or administration of antilymphocyte or immunoglobulin products. One can then determine, by using specific reagents, whether Ig or C or both are implicated. The sensitivity of the Coombs test varies depending on the techniques that are used. In general, the test is positive if there is an average of at least 400 molecules of Ig and/or C on each red cell; but with more advanced techniques involving flow cytometry analysis or enzyme-linked radiolabeled tests, the sensitivity can be pushed to as low as 30–40 molecules

TABLE 105-7 Diseases and Clinical Situations in Which Hemolysis Is Largely Intravascular

	ONSET/TIME COURSE	MAIN MECHANISM	APPROPRIATE DIAGNOSTIC PROCEDURE	COMMENTS
Mismatched blood transfusion	Abrupt	Nearly always ABO incompatibility	Repeat cross-match	
Paroxysmal nocturnal hemoglobinuria (PNH)	Chronic with acute exacerbations	Complement (C)-mediated destruction of CD59(–) red cells	Flow cytometry to display a CD59(–) red cell population	Exacerbations due to C activation through any pathway
Paroxysmal cold hemoglobinuria (PCH)	Acute	Immune lysis of normal red cells	Test for Donath-Landsteiner antibody	Often triggered by viral infection
Septicemia	Very acute	Exotoxins produced by *Clostridium perfringens*	Blood cultures	Other organisms may be responsible
Microangiopathic	Acute or chronic	Red cell fragmentation	Red cell morphology on blood smear	Different causes ranging from endothelial damage to hemangioma to leaky prosthetic heart valve
March hemoglobinuria	Abrupt	Mechanical destruction	Targeted history taking	Has been reported after extreme ritual dancing
Favism	Acute	Destruction of older fraction of G6PD-deficient red cells	G6PD assay	Triggered by ingestion of large dish of fava beans[a]

[a] The trigger of acute hemolytic anemia, often with hemoglobinuria, can be infection or a drug (see Table 105-5) rather than fava beans. Hemoglobinuria may or may not be reported by patient, but it is often macroscopic, i.e., recognizable by simple inspection of urine.

Abbreviation: G6PD, glucose 6-phosphate dehydrogenase.

per red cell. Therefore, liaison with a specialized laboratory is desirable; a dual direct antiglobulin test has also been developed. In the past, the diagnosis of "Coombs-negative AIHA" was regarded as a last resort, but it is important to know that a patient with this label may have severe AIHA, because if the antibody is powerful (high affinity/avidity), few molecules may be sufficient to opsonize red cells. Based on the Coombs test findings as well as on the thermal characteristics and the antigenic specificities of the autoantibodies **(Table 105-8)**, AIHA has been classified into subtypes.

In AIHA, the autoantibody reacts best at 37°C, and it is usually Rhesus specific (sometimes specifically anti-e). The main mechanism of hemolysis in AIHA is that the Fc portion of the IgG antibody bound to red cells is recognized by the Fc receptor of macrophages. This will trigger erythrophagocytosis wherever macrophages are abundant, i.e., in the liver, in the bone marrow, but especially in red pulp of the spleen **(Fig. 105-9)**, which, also because of its special anatomy, is often the predominant site of red cell destruction.

AIHA may be seen in isolation (and it is then called *idiopathic*) or as secondary to other disorders such as systemic autoimmune disorders (systemic lupus erythematosus [SLE]; sometimes, AIHA may be the first manifestation that leads to a diagnosis of SLE) or lymphoproliferative disorders (Table 105-8). Like all autoimmune diseases, AIHA must arise from a dysregulation of immunity. It is therefore not surprising that it is increasingly being recognized in chronic lymphocytic leukemia (CLL), whether treated or untreated; after BMT; and after solid organ transplantation entailing immunosuppressive treatment. Recently, warm antibody AIHA has also occurred as a side effect of the use of immune checkpoint inhibitors, such as nivolumab, in patients with various types of cancer.

TABLE 105-8 Classification of Acquired Immune Hemolytic Anemias

	TYPE OF ANTIBODY	
CLINICAL SETTING	COLD, MOSTLY IgM, OPTIMAL TEMPERATURE 4°C–30°C	WARM, MOSTLY IgG, OPTIMAL TEMPERATURE 37°C; OR MIXED
Primary	CAD	AIHA (idiopathic)
Secondary to viral infection	EBV CMV Other	Parvovirus B19 HIV HCV EBV Viral vaccines
Secondary to other infection	Mycoplasma infection: paroxysmal cold hemoglobinuria	Babesia
Secondary to/associated with other disease	CAD in: Waldenström's disease Lymphoma	AIHA in: SLE, scleroderma, RA CLL Lymphoproliferative disorders Multiple myeloma Other malignancy Chronic inflammatory disorders (e.g., IBD) Thyroiditis (including Hashimoto) After allogeneic HSCT Common variable immunodeficiency After immune checkpoint modulating drugs
Secondary to drugs: drug-induced immune hemolytic anemia	Small minority (e.g., with lenalidomide)	Majority: currently most common culprit drugs are cefotetan, ceftriaxone, piperacillin, methyldopa, fludarabine
	Drug-dependent: antibody destroys red cells only when drug present (e.g., rarely penicillin) Drug-independent: antibody can destroy red cells even when drug no longer present (e.g., methyldopa)	
Associated with	Pregnancy	

Abbreviations: AIHA, autoimmune hemolytic anemia; CAD, cold agglutinin disease; CLL, chronic lymphocytic leukemia; CMV, cytomegalovirus; EBV, Epstein-Barr virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; HSCT, hematopoietic stem cell transplantation; IBD, inflammatory bowel disease; SLE, systemic lupus erythematosus; RA, rheumatoid arthritis.

TREATMENT

Warm Antibody Autoimmune Hemolytic Anemia

Severe acute AIHA can be a medical emergency. The immediate treatment almost invariably includes transfusion of red cells. This may pose a special problem because many or all of the blood units cross-matched may be incompatible. In these cases, it is often correct, if paradoxical, to transfuse ABO-matched but incompatible blood, with the rationale being that the transfused red cells will be destroyed no less—but no more—than the patient's own red cells, and in the meantime, the patient stays alive. A situation like this requires close liaison and understanding between the clinical unit treating the patient and the blood transfusion/serology lab. Whenever the anemia is not immediately life-threatening, blood transfusion should be withheld (because compatibility problems may increase with each unit of blood transfused) and medical treatment started immediately with prednisone (1 mg/kg per day), which will produce a remission promptly in at least one-half of patients. Rituximab (anti-CD20), previously regarded as second-line treatment, is increasingly being used at a relatively low dose (100 mg/week × 4), together with prednisone as part of first-line treatment. It is especially encouraging that this approach seems to reduce the rate of relapse, a common occurrence in AIHA.

For patients who do relapse or are refractory to medical treatment, additional therapeutic strategies are now available. Splenectomy does not cure the disease, but it can produce significant benefit by removing a major site of hemolysis, thus improving the anemia and/or reducing the need for other therapies (e.g., the dose of prednisone); of course, splenectomy is not free of risk, as it entails increased risk of sepsis and of thrombosis. The response rates to splenectomy and rituximab are similar. Since the introduction of rituximab, azathioprine, cyclophosphamide, cyclosporine, mycophenolate, and intravenous immunoglobulin have become second- or third-line agents. In very rare severe refractory cases, one may have to consider a high dose of cyclophosphamide (50 mg/kg per day for 4 days), followed by a myelo-stimulating agent, or the anti-CD52 agent alemtuzumab. Fostamatinib, an inhibitor of intracellular cell signaling (Syk kinase) mediating erythrophagocytosis, has shown beneficial effects in an open-label clinical study, becoming an interesting third-line therapeutic option (Fig. 105-9). When severe anemia is associated with inadequate reticulocyte output, the use of EPO may help to reduce or avoid the requirement for transfusion of red cells. Several new agents, including inhibitors of the neonatal Fc receptor and others, are currently under investigation.

PAROXYSMAL COLD HEMOGLOBINURIA (PCH) PCH is a rare form of AIHA occurring mostly in children, usually triggered by a viral infection, usually self-limited, and characterized by the so-called Donath-Landsteiner antibody. In vitro, this antibody has unique serologic features; it has usually anti-P specificity, and it binds to red cells only at a low temperature (optimally at 4°C), but when the temperature is shifted to 37°C, lysis of red cells takes place in the presence of complement. Consequently, in vivo, there is intravascular hemolysis, resulting in hemoglobinuria. Clinically the differential diagnosis must include other causes of hemoglobinuria (Table 105-7), but the presence of the Donath-Landsteiner antibody will prove PCH. Active supportive treatment, including blood transfusion, may be needed to control the anemia; subsequently, recovery is the rule.

COLD AGGLUTININ DISEASE This designation indicates the other main type of AIHA, which has quite different features when compared with warm antibody AIHA. First, cold agglutinin disease (CAD) is a

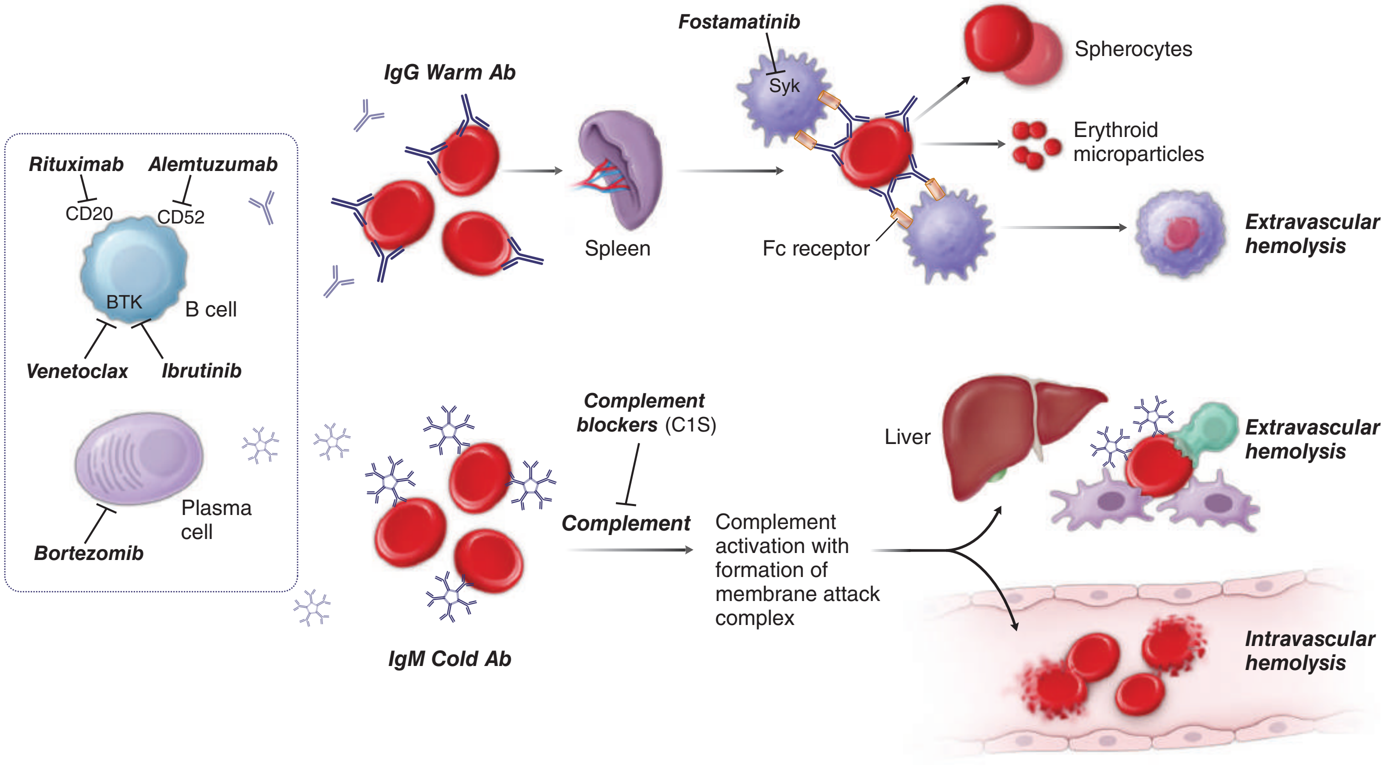

FIGURE 105-9 Basic mechanisms involved in warm antibody– and cold antibody-mediated autoimmune hemolytic anemias. ***Top.*** With warm antibodies (usually IgG), opsonized red cells are removed by Fc receptor–bearing macrophages, largely in the spleen (extravascular hemolysis); fragmentation and spherocyte formation also play a role, with the spleen again being the main site. ***Bottom.*** In cold agglutinin disease (CAD), the antibody is IgM, which, once bound to red cells, causes complement activation through the classic pathway, with consequent intravascular hemolysis. In addition, C3b-opsonized red cells will undergo erythrophagocytosis by Kupffer cells in the liver, with consequent extravascular hemolysis. The inset on the left illustrates the B cells that make these autoantibodies: polyclonal in the case of warm antibody–mediated autoimmune hemolytic anemia and monoclonal in the case of CAD. The new therapeutic approaches in addition to glucocorticoids for autoimmune hemolytic anemias are shown in red. The inset highlights drugs targeting immune cells involved in generation of antibodies. Ab, antibody; BTK, Bruton's tyrosine kinase; Syk, spleen tyrosine kinase.

chronic and more frequently indolent condition—in contrast to the abrupt onset of warm antibody AIHA. Second, the term *cold* refers to the fact that the autoantibody involved reacts with red cells poorly or not at all at 37°C, whereas it reacts strongly at lower temperatures. As a result, hemolysis is more prominent the more the body is exposed to the cold. Third, the antibody is produced by a clone of autoreactive B lymphocytes. Sometimes the antibody concentration in the serum is high enough to show up as a spike in plasma protein electrophoresis, thus qualifying CAD as an IgM monoclonal gammopathy; however, it differs from Waldenström macroglobulinemia by not having the characteristic *MYD88* mutation **(see Chap. 116)**; instead, a somatic mutation in the *KMT2D* gene, encoding a lysine histone methylase, is present in the B-cell clone of a majority of CAD patients: this seems to favor proliferation. The antibody produced by the B-cell clone is IgM; usually it has an anti-I specificity (the I antigen is present on the red cells of almost everybody), and it may have a very high titer (1:100,000 or more has been observed). IgM, when bound to red cells, is a powerful activator of the complement cascade, with ultimate formation of the membrane attack complex (see Fig. 105-9); this will directly cause destruction of red cells (*intravascular hemolysis*; indeed, CAD patients may present with hemoglobinuria). In addition, once complement is activated, C3b will bind to red cells that, thus opsonized, will be destroyed by macrophages (*extravascular hemolysis*); unlike in AIHA, the spleen is not predominant in this process.

In mild forms of CAD, avoidance of exposure to cold may be all that is needed to enable the patient to have a reasonably comfortable quality of life; but in more severe forms, the management of CAD is not easy. Plasma exchange will remove antibodies and is, therefore, in theory, a rational approach in severe cases. However, the management of CAD has changed significantly with the advent of the anti-CD20 antibody rituximab; up to 60% of patients respond. If remission is followed by relapse, a new course of rituximab may be again effective, and remissions may be more durable with a combination of rituximab with either fludarabine or bendamustine, in particular in CAD associated with a clinically manifested lymphoproliferative disorder. Therefore, even in the absence of a formal trial, rituximab has become de facto first-line treatment, especially since previously used immunosuppressive/cytotoxic agents, although they can reduce the antibody titer, have limited clinical efficacy and, in view of the chronic nature of CAD, their side effects may prove unacceptable. Unlike in AIHA, prednisone and splenectomy are ineffective. In the management of CAD in relapse, the B-cell receptor inhibitors venetoclax and ibrutinib, as well as for the proteasome inhibitor bortezomib, are emerging as effective agents. A different approach targeting complement inhibitors has been also explored by using sutimlimab (anti-C1s); a limitation of this approach is that hemolysis will be curbed only for as long as this agent is administered (Fig. 105-9).

In terms of supportive treatment, blood transfusion may be helpful, despite the fact that red cells from the donor, being I positive, will survive no longer than those of the patient. Both the blood bag and the patient's extremities must be kept warm during transfusion.

Hemolytic Anemia from Toxic Agents and Drugs A number of chemicals with oxidative potential, whether medicinal or not, can cause hemolysis even in people who are not G6PD deficient (for which, see above). Examples are hyperbaric oxygen (or 100% oxygen), nitrates, chlorates, methylene blue, dapsone, cisplatin, and numerous aromatic (cyclic) compounds. Other chemicals may be hemolytic through nonoxidative, largely unknown mechanisms; examples include arsine, stibine, copper, and lead. The HA caused by lead poisoning is characterized by basophilic stippling; it is in fact a phenocopy of that seen in P5N deficiency (see above), suggesting it is mediated at least in part by lead inhibiting this enzyme.

In these cases, hemolysis appears to be mediated by a direct chemical action on red cells. But drugs can cause hemolysis through at least two other mechanisms. (1) A drug can behave as a hapten and induce antibody production; in rare subjects, this happens, for instance, with penicillin. Upon a subsequent exposure, red cells are caught, as innocent bystanders, in the reaction between penicillin and antipenicillin antibodies. Hemolysis will subside as soon as penicillin administration is stopped. (2) A drug can trigger, perhaps through mimicry, the production of an antibody against a red cell antigen. The best-known example is methyldopa, an antihypertensive agent no longer in use, which in a small fraction of patients stimulated the production of the Rhesus antibody anti-e. In patients who have this antigen, the anti-e is a true autoantibody, which then causes true AIHA (see above). Usually this will gradually subside once methyldopa is discontinued.

Severe intravascular hemolysis can be caused by the venom of certain snakes (cobras and vipers), and HA can also follow spider bites.

Paroxysmal Nocturnal Hemoglobinuria (PNH) PNH is an acquired chronic HA characterized by persistent intravascular hemolysis with occasional or frequent recurrent exacerbations. In addition to (1) hemolysis, there may be (2) pancytopenia and (3) a distinct tendency to venous thrombosis. This triad makes PNH a truly unique clinical condition; however, when not all of these three features are manifest on presentation, the diagnosis is often delayed, although it can always be made by appropriate laboratory investigations (see below).

PNH is encountered in all populations throughout the world, but it is a rare disease, with an estimated prevalence of 5–10 per million (it may be somewhat less rare in Southeast Asia and in the Far East). PNH has about the same frequency in men and women. PNH is not inherited, and it has never been reported as a congenital disease, but it can present in small children or as late as in the seventies, although most patients are young adults.

CLINICAL FEATURES When seeking medical attention, the patient may report that one morning, they "passed blood instead of urine." This distressing or frightening event may be regarded as the classic presentation; however, more frequently, this symptom is not noticed or not reported. Indeed, the patient often presents simply as a problem in the differential diagnosis of anemia, whether symptomatic or discovered incidentally. Sometimes the anemia is associated from the outset with neutropenia, thrombocytopenia, or both, thus signaling an element of bone marrow failure (see below). Some patients may present with recurrent attacks of severe abdominal pain eventually found to be related to thrombosis in abdominal veins, or attributable to nitric oxide depletion associated with intravascular hemolysis. When thrombosis affects the hepatic vein, it may produce acute hepatomegaly and ascites, i.e., a full-fledged Budd-Chiari syndrome, which, in the absence of liver disease, ought to raise the suspicion of PNH.

The natural history of PNH can extend over decades. In the past, with supportive treatment only, the median survival was estimated to be about 10–20 years, with the most common cause of death being venous thrombosis, followed by infection secondary to severe neutropenia and hemorrhage secondary to severe thrombocytopenia. Rarely (estimated 1–2% of all cases), PNH may terminate in acute myeloid leukemia. On the other hand, full spontaneous recovery from PNH has been documented, albeit rarely.

LABORATORY INVESTIGATIONS AND DIAGNOSIS The most consistent blood finding is anemia, which may range from mild to moderate to very severe. The anemia is usually normo-macrocytic, with unremarkable red cell morphology. If the MCV is high, it is usually largely accounted for by reticulocytosis, which may be quite marked (up to 20%, or up to 400,000/μL). The anemia may become microcytic if the patient is allowed to become iron deficient as a result of chronic iron loss through hemoglobinuria. Unconjugated bilirubin is mildly or moderately elevated; LDH is typically markedly elevated (values in the thousands are common); and haptoglobin is usually undetectable. All of these findings make the diagnosis of HA compelling. Hemoglobinuria may be overt in a random urine sample; if it is not, it may be helpful to obtain serial urine samples because hemoglobinuria can vary dramatically from day to day and even from hour to hour. The bone marrow is usually cellular, with marked to massive erythroid hyperplasia, often with mild to moderate dyserythropoietic features (these overlap with those seen in myelodysplastic syndromes, but PNH remains a separate entity). At some stage of the disease, the marrow may become hypocellular or even frankly aplastic (see below).

The definitive diagnosis of PNH must be based on the demonstration that a substantial proportion of the patient's red cells are deficient in proteins (notably CD59 and CD55) that normally protect the red cells from activated complement (C). The sucrose hemolysis test is unreliable; in contrast, the acidified serum (Ham) test is highly reliable. The gold standard today is flow cytometry, which can be carried out on granulocytes as well as on red cells and has a very high sensitivity. In PNH, characteristically, one sees a bimodal distribution of cells, with a discrete population that is CD59 and CD55 negative. Although very small populations of CD59(–) cells are of interest in terms of pathophysiology (particularly of aplastic anemia [AA]), no patient should be diagnosed with PNH unless the proportion is substantial: in first approximation, at least 5% of the total red cells and at least 20% of the total granulocytes.

PATHOPHYSIOLOGY Hemolysis in PNH is mainly intravascular and is due to an intrinsic abnormality of the red cell, which makes it exquisitely sensitive to activated C, whether C is activated through the alternative pathway or through an antigen-antibody reaction (classic pathway). The former mechanism is mainly responsible for chronic hemolysis in PNH; the latter explains why the hemolysis can be dramatically exacerbated in the course of a viral or bacterial infection. Hypersusceptibility to C is due to deficiency in the red cell membrane of several protective proteins **(Fig. 105-10)**, among which CD59 is the most important because it is able to hinder the insertion into the membrane of C9 polymers (the so-called membrane attack complex [MAC]). The molecular basis for the deficiency of these proteins has been pinpointed not to a defect in any of the respective genes, but rather to the shortage of a unique glycolipid molecule, GPI (Fig. 105-2), which, through a peptide bond, anchors these proteins to the surface membrane of cells. The shortage of GPI is due in turn to a somatic mutation in an X-linked gene, called *PIGA*, required for an early step in GPI biosynthesis. As a result, the patient's marrow is a mosaic of mutant and nonmutant cells, and the peripheral blood always contains both GPI-negative (PNH) cells and GPI-positive (non-PNH) cells; in most cases, the former prevail. Thrombosis is one of the most immediately life-threatening complications of PNH and yet one of the least understood in its pathogenesis. It could be that deficiency of CD59 on the PNH platelet causes inappropriate platelet activation; however, other mechanisms are possible. In very rare cases, PNH can be caused by biallelic mutations of the autosomal *PIGT* gene, which maps to chromosome 20q, in the absence of a *PIGA* mutation. In these cases, because GPI is produced but cannot bind to proteins, the clinical picture is further complicated by the coexistence of a chronic inflammatory state.

BONE MARROW FAILURE (BMF) AND RELATIONSHIP BETWEEN PNH AND AA It is not unusual that patients with firmly established PNH have a previous history of AA, sometimes well documented; in many cases, BMF preceding overt PNH may have passed unnoticed, but it is probably the rule rather than the exception. On the other hand, sometimes a patient with PNH becomes less hemolytic and more pancytopenic and ultimately has the clinical picture of AA. The relationship between PNH and AA manifested in the clinical course of patients reflects a close link in pathogenesis. AA is thought to be an organ-specific autoimmune disease, in which T cells cause damage to hematopoietic stem cells via an as yet unidentified molecular target. The same may be true of PNH, and in this condition, the target might be the GPI molecule itself. This would explain why GPI-negative (PNH) stem cells are spared; *PIGA* mutations can be demonstrated in normal people. Thus, PNH results from the combined action of two factors: failure of normal hematopoiesis and massive expansion of a PNH clone. There is evidence from mouse models that PNH stem cells do not expand on their own, and there is evidence from human patients that expansion is associated with negative selection against GPI-positive cells by GPI-specific T cells. Thus, PNH is a prime example of a clonal disease that is not malignant.

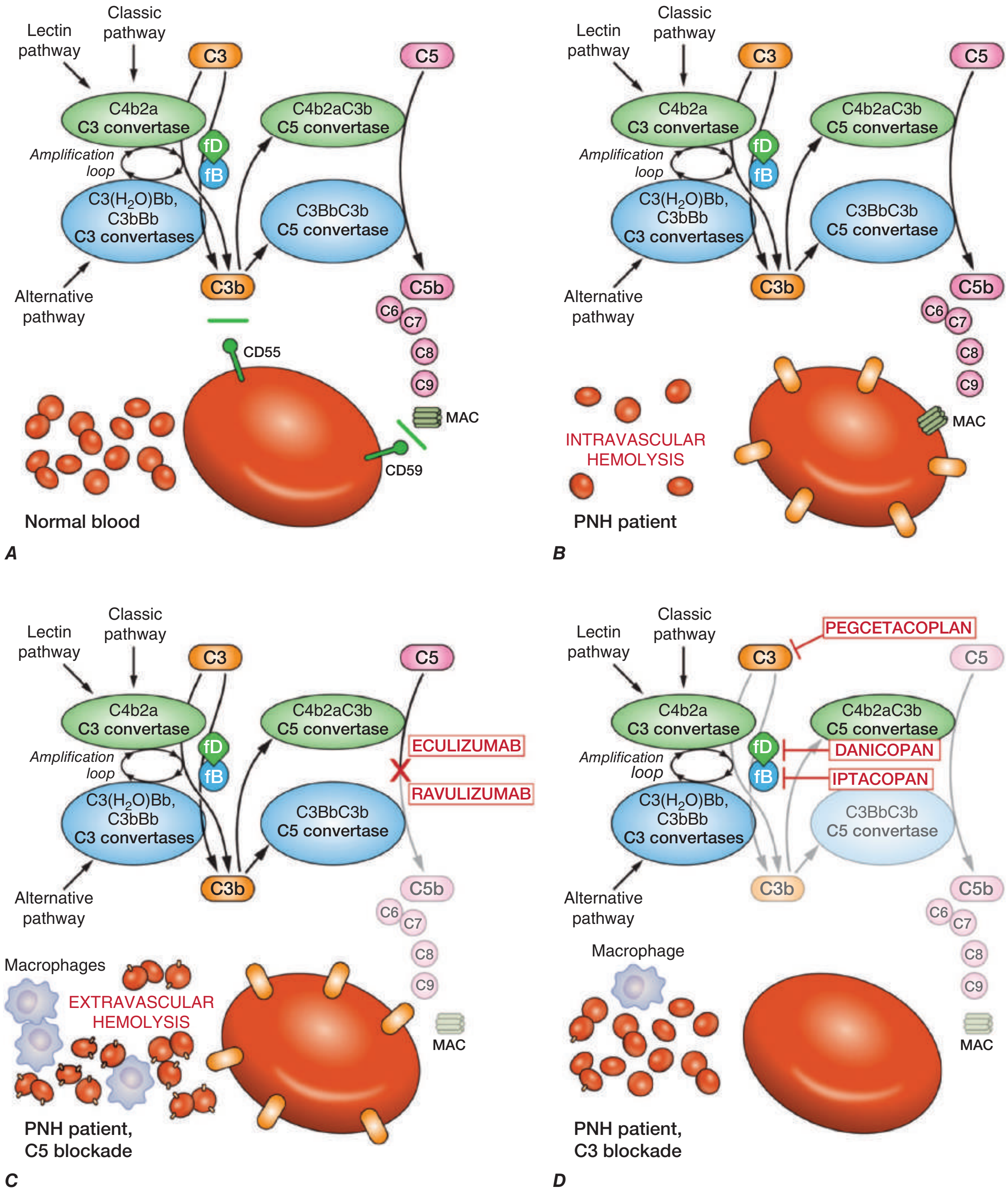
Lectin pathway
Classic pathway
C3
C5
C4b2a C3 convertase
C4b2aC3b C5 convertase
fD
fB
Amplification loop
C3(H2O)Bb, C3bBb C3 convertases
C3BbC3b C5 convertase
Alternative pathway
C3b
C5b
C6
C7
C8
C9
CD55
MAC
CD59
Normal blood
A
INTRAVASCULAR HEMOLYSIS
PNH patient
B
ECULIZUMAB
RAVULIZUMAB
Macrophages
EXTRAVASCULAR HEMOLYSIS
PNH patient, C5 blockade
C
PEGCETACOPLAN
DANICOPAN
IPTACOPAN
Macrophage
PNH patient, C3 blockade
D

FIGURE 105-10 The complement cascade and the fate of red cells. ***A.*** In normal blood, when complement is activated, red cells are protected from lysis in several ways: primarily by the two glycosylphosphatidylinositol (GPI)-linked surface proteins CD55 (prevents binding of C3 fragments) and CD59 (prevents the membrane attack complex [MAC] from inserting into the membrane). ***B.*** Paroxysmal nocturnal hemoglobinuria (PNH) red cells are deficient in CD55 and CD59 because the glycosylphosphatidylinositol (GPI) biosynthetic pathway is blocked as a result of a PIGA mutation; therefore, C3 fragments, particularly C3d, bind to their surface, and the red cells are rapidly lysed by the action of the MAC. ***C.*** With drugs (monoclonal antibodies) that bind to C5 and prevent it from splitting into C5a and C5b, the entire distal pathway from C5 onward is blocked, MAC is not formed, and intravascular hemolysis (IVH) is abrogated. However, red cells opsonized by C3d will be destroyed in the spleen and elsewhere; this drug-induced extravascular hemolysis (EVH) varies in severity between patients. The Coombs test, which is characteristically negative in PNH, becomes positive (provided that a "broad-spectrum" or an anticomplement reagent is used). ***D.*** With a drug that targets C3, C3b formation is inhibited, and the distal pathway is not triggered by C3b. Therefore, again, no MAC is formed (abrogating IVH), and at the same time, opsonization of red cells by C3d is prevented, so that EVH is also curbed. The same is largely true for drugs that target factor B or factor D, although C3b can still be formed through the classical pathway. *(Reproduced with permission from L Luzzatto: Control of hemolysis in patients with PNH. Blood 138:1908, 2021.)*

TREATMENT

Paroxysmal Nocturnal Hemoglobinuria

Until around 20 years ago, there were essentially two treatment options for PNH: either allogeneic BMT, providing a definitive cure at the cost of nonnegligible risks; or continued supportive treatment for what, unlike other acquired HAs, may be a lifelong condition. A major advance has been the introduction in 2007 of a humanized monoclonal antibody, eculizumab, which binds to the complement component C5 near the site that, when cleaved, will trigger the distal part of the complement cascade leading to formation of the MAC. With C5 blocked by anti-C5, the patient is relieved of intravascular hemolysis and of its attendant consequences, including hemoglobinuria, with a reduced risk of thrombosis. In the majority of patients who needed regular blood transfusion, the transfusion requirement is either abolished or significantly reduced. For many

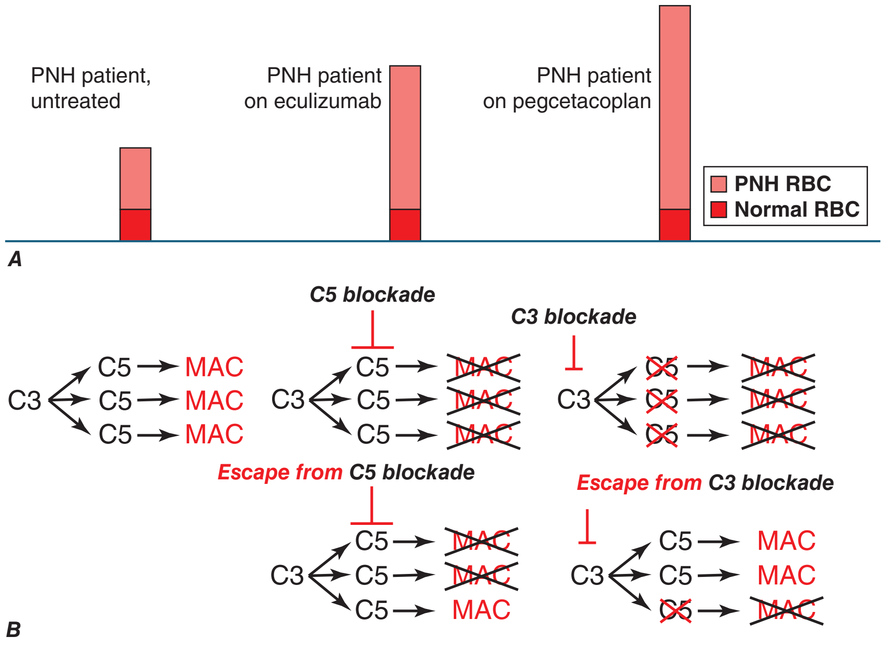

FIGURE 105-11 Impact and implications of anticomplement therapy in paroxysmal nocturnal hemoglobinuria (PNH). ***A.*** The *three vertical bars* symbolize red cell composition in examples of PNH patients. In the untreated patient, severe anemia is present with ~30% PNH red cells. In the patient on eculizumab, PNH red cells are protected from intravascular hemolysis, although at the expense of iatrogenic extravascular hemolysis (see text), and therefore, mild anemia is still present and can become severe. In the patient on pegcetacoplan, both intravascular and extravascular hemolysis are prevented, and the anemia is completely corrected. Note that the increase in red cells in the treated patients consists entirely of PNH red cells. ***B.*** Schematics of complement action and inhibition. In the untreated patient (*left*), the membrane attack complex (MAC) produces intravascular hemolysis. On eculizumab (*middle*), C5 blockade prevents MAC formation; if blockade is incomplete, some MAC is formed, with consequent *breakthrough hemolysis.* On pegcetacoplan (*right*), C5 convertase cannot be formed, and therefore, MAC formation is equally prevented; however, since there is an enzymatic cascade upstream of C5 cleavage, if blockade is incomplete, MAC formation will be more abundant with consequently more severe *breakthrough hemolysis,* potentially more massive as the PNH red cells are greatly increased in both relative and absolute terms. RBC, red blood cell.

PNH patients, eculizumab has meant a real transformation in the quality of life, as well as a decrease in complications, particularly thrombosis; thrombosis may still occur, but much more rarely. At the same time, it is important to know that in patients on eculizumab, the PNH red cells, now protected from being lysed through the MAC, do still bind C3 fragments and thus become opsonized; thus, the Coombs test becomes positive and hemolysis persists, which is now iatrogenic and mainly extravascular. The extent to which this happens depends in part on a genetic polymorphism of the complement receptor CR1. Based on its half-life, eculizumab must be administered intravenously every 14 days. Ravulizumab, a derivative of eculizumab with longer life in circulation, is administered at 8-week intervals, with obvious practical advantage. Patients who, upon C5 blockade, are still receiving blood transfusions are at risk of iron overload.

Persistent blood transfusion requirement and extravascular hemolysis in PNH patients on C5 blockade therapy have been a stimulus to developing agents that may inhibit complement activation more upstream, at the level of C3. This has been found not only to prevent intravascular hemolysis but also to avoid C3 opsonization of red cells, thus not causing extravascular hemolysis. Pegcetacoplan, an inhibitor of C3, and iptacopan, an inhibitor of factor B (which, once activated, becomes part of the alternative pathway C3 convertase) are already approved drugs; and danicopan, an inhibitor of factor D (required for the activation of factor B), has given promising results in clinical trials. Thus, several drug therapy options are available for an ultra-rare disease **(Fig. 105-11)**. Compared to C5 blockade, C3 blockade has a greater chance to completely correct anemia. We must consider that correction of anemia takes place by the buildup of a much larger population of PNH red cells than would ever be possible in an untreated PNH patient, and this population may undergo sudden lysis if complement blockade is abrogated, whether by omission or otherwise.

Eculizumab and the other complement inhibitors are very expensive and, for this reason, not accessible to patients in many parts of the world. Therefore, the management of PNH by supportive treatment is still very important. Folic acid supplements (at least 3 mg/d) are mandatory; the serum iron should be checked periodically, and iron supplements should be administered as appropriate. Transfusion of white cell-free red cells should be used whenever necessary, which, for some patients, means quite frequently. Long-term glucocorticoids are not indicated because there is no evidence that they have any effect on chronic hemolysis; in fact, they are contraindicated because their side effects are considerable. A short course of prednisone may be useful when an inflammatory process exacerbates hemolysis. Any PNH patient who has had venous thrombosis or who has been found after a thrombophilia screen to have a genetic risk factor should be on regular anticoagulant prophylaxis. With thrombotic complications that do not resolve otherwise, thrombolytic treatment with tissue plasminogen activator may be indicated.

Where anti-C5 therapy is available, the proportion of PNH patients receiving BMT has decreased significantly. However, when an HLA-identical sibling is available, BMT should be taken into consideration for any young patient with severe PNH, and for patients with the so-called PNH-AA syndrome, since C inhibitors have no effect on BMF. For these patients, immunosuppressive treatment with antithymocyte globulin and cyclosporine A may be an alternative, and it may be compatible with concurrent administration of eculizumab.

■ FURTHER READING

Berentsen S, Barcellini W: Autoimmune hemolytic anemias. N Engl J Med 385:1407, 2021.

Brodsky RA: Warm autoimmune hemolytic anemia. N Engl J Med 381:64, 2019.

Dacie J: *The Haemolytic Anaemias*. London, Churchill Livingstone, volumes 1-5, 1985–1999.

De Franceschi L et al: Acute hemolysis by hydroxycloroquine was observed in G6PD-deficient patient with severe COVD-19 related lung injury. Eur J Intern Med 77:136, 2020.

Grace RF et al: Safety and efficacy of mitapivat in pyruvate kinase deficiency. N Engl J Med 381:933, 2019.

Loirat C et al: An international consensus approach to the management of atypical hemolytic uremic syndrome in children. Pediatr Nephrol 31:15, 2016.

Luzzatto L et al: Glucose-6-phosphate dehydrogenase deficiency. Blood 136:1225, 2020.

Notaro R, Luzzatto L: Breakthrough hemolysis in PNH with proximal or terminal complement inhibition. N Engl J Med 387:160, 2022.

Roy NBA et al: The use of next-generation sequencing in the diagnosis of rare inherited anaemias: A joint BSH/EHA Good Practice Paper. Br J Haematol 198:459, 2022.

Uyoga S et al: Glucose-6-phosphate dehydrogenase deficiency and susceptibility to childhood diseases in Kilifi, Kenya. Blood Adv 4:5942, 2020.

106

Anemia Due to Acute Blood Loss

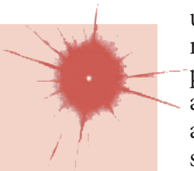

Dan L. Longo

Blood loss causes anemia by two main mechanisms: (1) by the direct loss of red cells; and (2) if the loss of blood is protracted, it will gradually deplete iron stores, eventually resulting in iron deficiency. The latter type of anemia is covered in **Chap. 102**. Here, we are concerned with the former type, that is, *posthemorrhagic anemia*, which follows *acute* blood loss. This can be *external* (e.g., after trauma or obstetric hemorrhage) or *internal* (e.g., from bleeding in the gastrointestinal tract, rupture of the spleen, rupture of an ectopic pregnancy, subarachnoid hemorrhage, leaking aneurysm). In any of these cases, after the sudden loss of a large amount of blood, there are three clinical/pathophysiologic stages. (1) At first, the dominant feature is hypovolemia, which poses a threat particularly to organs that normally have a high blood supply, like the brain and the kidneys; therefore, loss of consciousness and acute renal failure are major threats. It is important to note that at this stage an ordinary blood count will not show anemia because the hemoglobin concentration is not affected. As hypovolemia is corrected with intravenous fluids acutely, the hemoglobin will gradually fall over several hours. On physical exam, tachycardia, tachypnea, decreased pulse pressure, cold skin that appears pale and mottled, and decreased urine output may be noted reflecting the efforts of the sympathetic nervous system to compensate. (2) Next, as an emergency response, baroreceptors and stretch receptors will cause release of vasopressin and other peptides, and the body will shift fluid from the extravascular to the intravascular compartment, producing hemodilution; thus, the hypovolemia gradually converts to anemia. The degree of anemia will reflect the amount of blood lost. If after 3 days the hemoglobin is, for example, 7 g/dL, it means that about half of the entire blood has been lost. (3) Provided bleeding does not continue, the bone marrow response will gradually ameliorate the anemia. In this phase of the process, the reticulocyte count and erythropoietin levels will be elevated. The physiologic increase in marrow red cell production reflected by the increase in reticulocytes is similar to the marrow response to hemolysis. Hemolysis and compensatory marrow increase in red blood cell (RBC) production are the two major mechanisms associated with anemia that is accompanied by an increase in reticulocyte count.

The diagnosis of acute posthemorrhagic anemia (APHA) is usually straightforward, although sometimes internal bleeding episodes (e.g., after a traumatic injury), even when large, may not be immediately obvious. When the scene is bloody, often the estimate of the volume of blood loss is overestimated. Always check the patient carefully. Look for physical findings that may help localize the bleeding if the site of bleeding is not obvious. Grey Turner sign (flank ecchymosis) may reflect retroperitoneal bleeding. Cullen sign (umbilical ecchymosis) may suggest intraperitoneal or retroperitoneal bleeding. Dullness to chest percussion may suggest intrapleural bleeding. Whenever an abrupt fall in hemoglobin has taken place, whatever history is given by the patient, APHA should be suspected. Supplementary history may have to be obtained by asking the appropriate questions, and appropriate investigations (e.g., a sonogram or an endoscopy) may have to be carried out.

If blood loss is mild, enhanced O_2 delivery is achieved through changes in the O_2–hemoglobin dissociation curve mediated by a decreased pH or increased CO_2 (*Bohr effect*). With acute blood loss, hypovolemia dominates the clinical picture, and the hematocrit and hemoglobin levels do not reflect the volume of blood lost. Signs of vascular instability appear with acute losses of 25% or more of the total blood volume. The donation of a unit of blood (~20% of the blood volume) is often minimally symptomatic. In patients who lose a larger percentage of the blood volume, the issue is not anemia but hypotension and decreased organ perfusion. When >30% of the blood volume is lost suddenly, patients are unable to compensate with the usual mechanisms of sympathetic nervous system increases in heart rate, vascular contraction, and changes in regional blood flow. The patient prefers to remain supine and will show postural hypotension and tachycardia. If the volume of blood lost is >40% (i.e., >2 L in the average-sized adult), signs of hypovolemic shock including confusion, dyspnea, diaphoresis, hypotension, and tachycardia appear. Such patients have significant deficits in vital organ perfusion and require immediate volume replacement.

Symptoms associated with more chronic or progressive anemia depend on the age of the patient and the adequacy of blood supply to critical organs. Symptoms associated with moderate anemia include fatigue, loss of stamina, breathlessness, and tachycardia (particularly with physical exertion). However, because of the intrinsic compensatory mechanisms that govern the O_2–hemoglobin dissociation curve, the gradual onset of anemia—particularly in young patients—may not be associated with signs or symptoms until the anemia is severe (hemoglobin <70–80 g/L [7–8 g/dL]). When anemia develops over a period of days or weeks, the total blood volume is normal to slightly increased, and changes in cardiac output and regional blood flow help compensate for the overall loss in O_2-carrying capacity. Changes in the position of the O_2–hemoglobin dissociation curve account for some of the compensatory response to anemia. With chronic anemia, intracellular levels of 2,3-bisphosphoglycerate rise, shifting the dissociation curve to the right and facilitating O_2 unloading. This compensatory mechanism can only maintain normal tissue O_2 delivery in the face of a 20–30 g/L (2–3 g/dL) deficit in hemoglobin concentration. Finally, further protection of O_2 delivery to vital organs is achieved by the shunting of blood away from organs that are relatively rich in blood supply, particularly the kidney, gut, and skin.

TREATMENT

Anemia Due to Acute Blood Loss

In patients who are hemodynamically unstable, the usual airway, breathing, and circulation assessments take priority. In the face of bleeding associated with hypotension, pharmacologic support with vasopressors is critical. With respect to anemia treatment, a two-pronged approach is imperative. (1) In many cases, the blood lost needs to be replaced promptly. Unlike with many chronic anemias, when finding and correcting the cause of the anemia is the first priority and blood transfusion may not be even necessary because the body is adapted to the anemia, with acute blood loss, the reverse is true. Because the body is not adapted to the anemia, blood transfusion takes priority. (2) While the emergency is being confronted, it is imperative to stop the hemorrhage and to eliminate its source.

In an acute hemorrhage situation, plasma may be preferred to saline for volume expansion since dilution of clotting factors with crystalloid may interfere with hemostasis. Furthermore, trauma can lead to vascular and platelet abnormalities that enhance the bleeding risk.

A special type of APHA is blood loss during and immediately after surgery, which can be substantial (e.g., up to 2 L in the case of a radical prostatectomy). Of course, with elective surgical procedures, the patient's own stored blood may be available (through preoperative autologous blood donation), and in any case, blood loss ought to have been carefully monitored/measured. The fact that this blood loss is iatrogenic dictates that ever more effort should be invested in optimizing its management. The special features of transfusion medicine are discussed in **Chap. 118**.

A Holy Grail of emergency medicine for a long time has been the idea of a blood substitute that would be universally available, suitable for all recipients, easy to store and to transport, safe, and as effective as blood itself. Two main paths have been pursued: (1) fluorocarbon synthetic chemicals that bind oxygen reversibly, and (2) artificially modified hemoglobins, known as hemoglobin-based oxygen carriers (HBOCs). Although there are numerous anecdotal reports of the use of both approaches in humans, and although HBOCs have reached the stage of phase 2–3 clinical trials, no "blood substitute" has yet become standard treatment.

FURTHER READING

Gianola S et al: Volume replacement in the resuscitation of trauma patients with acute hemorrhage: An umbrella review. Int J Emerg Med 16:87, 2023.

Khromov AV et al: Blood substitutes with gas transfer function. Bull Exp Biol Med 176:709, 2024.

Kornblith LZ et al: Trauma-induced coagulopathy: the past, present and future. J Thromb Haemostat 17:652, 2019.

Schiller AM et al: The physiology of blood loss and shock: New insights from a laboratory model of hemorrhage. Exp Biol Med 242:874, 2017.

107

Bone Marrow Failure Syndromes Including Aplastic Anemia and Myelodysplasia

Neal S. Young, Emma M. Groarke

Bone marrow failure diseases include aplastic anemia, myelodysplastic syndrome (MDS), pure red cell aplasia (PRCA), and other single-lineage cytopenias. Low blood counts in marrow failure result from deficient hematopoiesis, as distinguished from blood count depression due to peripheral destruction of red cells (hemolytic anemias), platelets (idiopathic thrombocytopenic purpura [ITP] or due to splenomegaly), and granulocytes (as in the immune leukopenias). Marrow damage and dysfunction also may be secondary to infection, inflammation, or cancer.

Hematopoietic failure syndromes are classified by dominant morphologic features of the bone marrow **(Table 107-1)**. Although practical distinction among these syndromes usually is clear from the marrow pathology, some processes are so closely related that the diagnosis may be complex. Separation between aplastic anemia and hypoplastic MDS can be particularly difficult. Patients may seem to suffer from two or three related diseases simultaneously, or one diagnosis may appear to evolve into another. Many of these syndromes share an immune-mediated mechanism of marrow destruction and some element of genomic instability resulting in a higher rate of malignant transformation.

It is important that the internist and general practitioner recognize the marrow failure syndromes because quality of life and prognosis is poor if the patient is untreated; effective therapies are often available but sufficiently complicated in their choice and delivery so as to warrant the care of a hematologist or oncologist. While the identification of pathogenic mutations on genomic screen, often on testing ordered by the internist and pediatrician, has revolutionized the diagnosis of the marrow failure syndromes, these results often require the interpretation of the hematologist and oncologist.

TABLE 107-1 Differential Diagnosis of Pancytopenia

Pancytopenia with Hypocellular Bone Marrow	
Acquired (immune) aplastic anemia	
Constitutional aplastic anemia (Fanconi anemia, telomere biology disorders, and others)	
Hypoplastic myelodysplastic syndrome	
Rare aleukemic leukemia	
Some acute lymphoid leukemia	
Rare lymphomas of bone marrow	
Copper deficiency	
Pancytopenia with Cellular Bone Marrow	
Primary bone marrow diseases	Secondary to systemic diseases
Myelodysplastic syndromes	Systemic lupus erythematosus
Paroxysmal nocturnal hemoglobinuria (PNH)	Hypersplenism
Myelofibrosis	B_{12}, folate deficiency
Aleukemic leukemia	Copper deficiency
Myelophthisis	Alcohol
Bone marrow lymphoma	HIV infection
Hairy cell leukemia	Brucellosis
	Sarcoidosis
	Tuberculosis
	Leishmaniasis
	Sepsis
Hypocellular Bone Marrow ± Pancytopenia	
Q fever	
Legionnaires' disease	
Anorexia nervosa, starvation	
Mycobacterium	

APLASTIC ANEMIA

DEFINITION

Aplastic anemia is pancytopenia with bone marrow hypocellularity. Based on etiology, pathophysiology, and clinical features, an empty bone marrow may be the result of physical and chemical damage, of immune-mediated destruction of hematopoietic cells, or a component of an inherited syndrome **(Fig. 107-1)**. Acquired aplastic anemia, usually immune mediated, is distinguished from constitutional aplastic anemia, and within the acquired category, marrow damage as occurs consistently after intensive cytotoxic chemotherapy for cancer, conditioning for transplant, and occasionally as in accidental radiation poisoning, from sporadic immune aplastic anemia. Aplastic anemia can also be constitutional, due to germline genetic defects, termed the inherited bone marrow failure syndromes (IBMFS). Classical inherited syndromes such as Fanconi anemia and dyskeratosis congenita typically are diagnosed in early childhood in affected pedigrees and have characteristic physical anomalies. However, telomere biology disorders **(see Chap. 482)** and germline predisposition disorders, such as *GATA2* deficiency and *RUNX1* familial platelet disorder, can present as marrow failure in normal-appearing adults. Immune aplastic anemia is often stereotypical in its manifestations, with the abrupt onset of low blood counts in a previously well young adult; seronegative hepatitis or a course of an incriminated medical drug may precede the onset. The diagnosis in these instances is uncomplicated. Sometimes blood count depression is moderate or incomplete, resulting in anemia, leukopenia, and thrombocytopenia in some combination. Aplastic anemia is related to both paroxysmal nocturnal hemoglobinuria (PNH; **Chap. 105**) and to MDS, and a clear distinction among these disorders may not be possible.

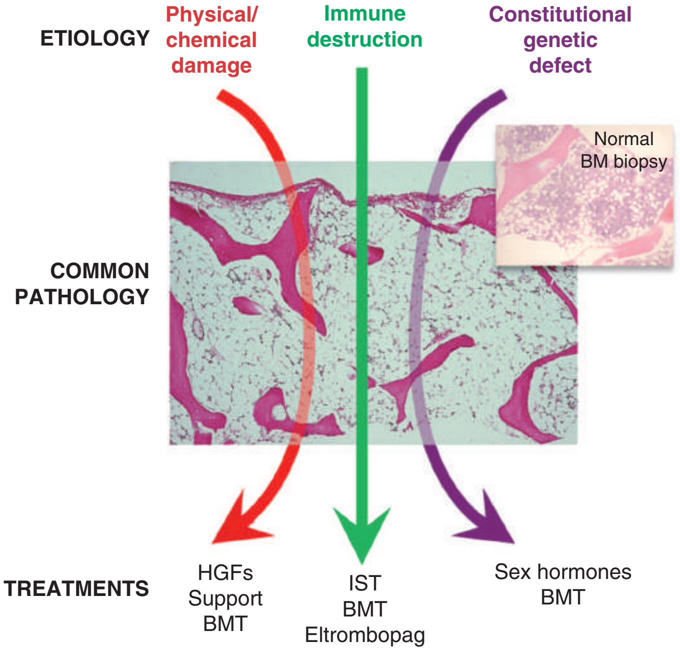

FIGURE 107-1 Pathophysiologies of aplastic anemia. The common pathology of the bone marrow replaced by fat can result from chemical or physical damage (iatrogenic; benzene); immune destruction (primarily T cell driven); and as a constitutional defect in genes important for cell maintenance and immune regulation. BM, bone marrow; BMT, bone marrow transplantation; HGF, hematopoietic growth factors; IST, immunosuppressive therapy.

TABLE 107-2 Classification of Aplastic Anemia and Single Cytopenias

ACQUIRED	INHERITED/CONSTITUTIONAL
Aplastic Anemia	
Secondary	Fanconi anemia
Radiation	Dyskeratosis congenita/telomere disease
Drugs and chemicals	Shwachman-Diamond syndrome
Regular effects	Familial aplastic anemia/leukemia predisposition syndromes: *GATA2, RUNX1, CTLA4*, and others
Idiosyncratic reactions	
Viruses	
Epstein-Barr virus (infectious mononucleosis)	Nonhematologic syndromes (Down, Dubowitz, Seckel)
Hepatitis (non-A, non-B, non-C hepatitis)	
Parvovirus B19 (transient aplastic crisis, pure red cell aplasia [PRCA])	
HIV-1 (AIDS)	
Immune diseases	
Eosinophilic fasciitis	
Hypoimmunoglobulinemia	
Large granular lymphocytosis (LGL)	
Thymoma/thymic carcinoma	
Graft-versus-host disease in immunodeficiency	
Paroxysmal nocturnal hemoglobinuria (PNH)	
Pregnancy	
Idiopathic (immune)	
Cytopenias	
PRCA (see Table 107-4)	Congenital PRCA (Diamond-Blackfan anemia)
Neutropenia/agranulocytosis	
Idiopathic	Kostmann syndrome
Drugs, toxins	Shwachman-Diamond syndrome
LGL	Reticular dysgenesis
Pure white cell aplasia (+/– thymoma)	
Thrombocytopenia	
Drugs, toxins	Amegakaryocytic thrombocytopenia
Acquired amegakaryocytic thrombocytopenia	Thrombocytopenia with absent radii
	Other rare germline mutations

EPIDEMIOLOGY

The incidence of acquired aplastic anemia in Europe and Israel is two cases per million persons annually. In Thailand and China, rates of five to seven per million have been established. Men and women are affected with equal frequency, but the age distribution is biphasic, with the major peak in the teens and twenties and a second in older adults.

ETIOLOGY

Historically, the origins of aplastic anemia were inferred from recurring clinical associations (Table 107-2); unfortunately, these relationships, as with a medical drug, may not be reliable in an individual patient, and associations are not equivalent to causation. In addition, although most cases of aplastic anemia are immune, little other than history separates these cases from those with a presumed etiology such as a drug exposure.

Radiation Marrow aplasia is a major acute sequela of radiation, anticipated from medical radiation therapy and rarely after accidental and/or occupational exposure Radiation damages DNA; tissues dependent on active mitosis are particularly susceptible. Whereas the radiation dose can be approximated from the rate and degree of decline in blood counts, dosimetry by reconstruction of the exposure can help to estimate the patient's prognosis and also to protect medical personnel from contact with radioactive tissue and excreta. MDS and leukemia, but probably not aplastic anemia, are late effects of radiation.

Chemicals Benzene is a notorious cause of bone marrow failure: epidemiologic, clinical, and laboratory data link benzene to aplastic anemia, acute leukemia, and blood and marrow abnormalities. For leukemia, incidence is correlated with cumulative exposure, but susceptibility must also be important because only a minority of even heavily exposed workers develop myelotoxicity. Employment history is important, especially in industries where benzene is used for a secondary purpose, usually as a solvent. Benzene-related blood diseases have declined with regulation of industrial exposure. The association between marrow failure and other chemicals is much less well substantiated. There is scant direct evidence of marrow failure as a late effect of exposure, even to benzene.

Drugs (Table 107-3) Many chemotherapeutic drugs have marrow suppression as a major toxicity; effects are dose-dependent and will occur in all recipients. In contrast, idiosyncratic reactions to a large and diverse group of drugs may lead to aplastic anemia without a clear dose-response relationship. A large international study in Europe in the 1980s quantitated drug relationships, especially for nonsteroidal analgesics, sulfonamides, thyrostatic drugs, some psychotropics, penicillamine, allopurinol, and gold. Association does not equal causation: a drug may have been used to treat the first symptoms of bone marrow failure (antibiotics for fever or a preceding viral illness) or provoked the first symptom of a preexisting disease (petechiae by nonsteroidal anti-inflammatory agents administered to the thrombocytopenic patient). In the context of total drug use, idiosyncratic reactions, although individually devastating, are rare events. Risk estimates are usually lower when determined in population-based studies. Furthermore, the low absolute risk is also made more obvious: even a 10- or 20-fold increase in risk translates, in a rare disease, to just a handful of drug-induced aplastic anemia cases among hundreds of thousands of exposed persons.

Infections Transient, mild blood count depression is frequent in the course of many viral and bacterial infections. Aplastic anemia can

TABLE 107-3 Some Drugs and Chemicals Associated with Aplastic Anemia

- Agents that regularly produce marrow depression as major toxicity in commonly used doses or normal exposures:
  - Cytotoxic drugs used in cancer chemotherapy: *alkylating agents*, *antimetabolites*, *antimitotics*, some antibiotics
- Agents that frequently but not inevitably produce marrow aplasia:
  - *Benzene*
- Agents associated with aplastic anemia but with a relatively low probability:
  - *Chloramphenicol*
  - Insecticides
  - Antiprotozoals: *quinacrine* and chloroquine, mepacrine
  - Nonsteroidal anti-inflammatory drugs (including *phenylbutazone*, indomethacin, ibuprofen, sulindac, aspirin)
  - Anticonvulsants (*hydantoins*, *carbamazepine*, phenacemide, felbamate)
  - Heavy metals (*gold*, arsenic, bismuth, mercury)
  - Sulfonamides: some antibiotics, antithyroid drugs (methimazole, methylthiouracil, propylthiouracil), antidiabetes drugs (tolbutamide, chlorpropamide), carbonic anhydrase inhibitors (acetazolamide and methazolamide)
  - Antihistamines (*cimetidine*, chlorpheniramine)
  - D-Penicillamine
  - Estrogens (in pregnancy and in high doses in animals)
- Agents whose association with aplastic anemia is more tenuous:
  - Other antibiotics (streptomycin, tetracycline, methicillin, mebendazole, trimethoprim/sulfamethoxazole, flucytosine)
  - Sedatives and tranquilizers (chlorpromazine, prochlorperazine, piperacetazine, chlordiazepoxide, meprobamate, methyprylon)
  - Allopurinol
  - Methyldopa
  - Quinidine
  - Lithium
  - Guanidine
  - Potassium perchlorate
  - Thiocyanate
  - Carbimazole

Note: Terms set in italics show the most consistent association with aplastic anemia.

rarely follow infectious mononucleosis. Parvovirus B19 is a cause of pure red cell aplasia and rarely implicated in generalized bone marrow failure.

Immunologic Diseases Aplasia is a major consequence and the inevitable cause of death in transfusion-associated graft-versus-host disease (GVHD) that can occur after infusion of nonirradiated blood products to an immunodeficient recipient. Aplastic anemia occurs with the rare collagen vascular syndrome eosinophilic fasciitis that is characterized by painful induration of subcutaneous tissues **(Chap. 372)**. Thymoma and hypoimmunoglobulinemia can co-occur with aplastic anemia. Pancytopenia with marrow hypoplasia can be seen in systemic lupus erythematosus (SLE).

Hepatitis Posthepatitis marrow failure accounts for ~5% of etiologies in most series. Patients are usually young men who have recovered from a bout of liver inflammation 1–2 months earlier; the subsequent pancytopenia is very severe. The hepatitis is seronegative (non-A, non-B, non-C); intensive laboratory efforts including deep sequencing have not disclosed an infectious agent, and the hepatitis is presumed to be immune-mediated. Fulminant liver failure in childhood can follow seronegative hepatitis, and marrow failure occurs at a high rate in these patients.

Pregnancy Aplastic anemia very rarely may complicate pregnancy and resolve with delivery or with spontaneous or induced abortion.

Paroxysmal Nocturnal Hemoglobinuria An acquired mutation in the *PIG-A* gene in a hematopoietic stem cell is required for the development of PNH, but *PIG-A* mutations are also present in the marrow cells of normal individuals. If the *PIG-A* mutant stem cell proliferates, the result is a clone of progeny deficient in glycosylphosphatidylinositol-linked cell surface membrane proteins **(Chap. 105)**. Clones of deficient cells can be detected by sensitive flow cytometry tests in one-half or more of patients with aplastic anemia at the time of presentation. Functional studies of bone marrow from PNH patients, even those with mainly hemolytic manifestations, show evidence of defective hematopoiesis. Patients with an initial clinical diagnosis of PNH, especially younger individuals, may later develop frank marrow aplasia and pancytopenia; patients with an initial diagnosis of aplastic anemia may suffer later from hemolytic PNH years after recovery of blood counts.

Constitutional Syndromes Fanconi anemia, an autosomal recessive disorder, manifests as congenital developmental anomalies, progressive pancytopenia, and an increased risk of malignancy. Chromosomes in Fanconi anemia are susceptible to DNA cross-linking agents, the basis for a diagnostic assay. Patients with Fanconi anemia typically have short stature, café au lait spots, and anomalies involving the thumb, radius, and genitourinary tract. At least 23 different genetic defects have been defined; the most common, type A Fanconi anemia, is due to a mutation in *FANCA*. Most of the Fanconi anemia gene products form a protein complex that activates FANCD2 by monoubiquitination to play a role in the cellular response to DNA damage and especially interstrand cross-linking.

Diamond-Blackfan anemia (see below) and Shwachman-Diamond syndrome are ribosomopathies, genetic defects in ribosome assembly that are tissue specific. In Shwachman-Diamond syndrome, presentation is early in life with neutropenia, pancreatic insufficiency, and malabsorption; most patients have compound heterozygous mutations in *SBDS*.

In the telomere biology disorders, inherited genetic defects alter telomere repair or one of the shelterin protein components of the telomere. The pediatric syndrome dyskeratosis congenita is characterized by the triad of mucous membrane leukoplasia, dystrophic nails, reticular hyperpigmentation, and early development of aplastic anemia **(Chap. 482)**. Mutations in genes of the telomere repair complex, which acts to maintain telomere length in replicating cells, cause telomere biology disorders: the X-linked variety is due to mutations in the *DKC1* (*dyskerin*) gene; the autosomal dominant types are due to mutations in *TERC* and TERT, more rarely, mutations can also occur in genes such as *TINF2* that encode shelterin proteins, which bind telomere DNA. Mutations in *TERC* and *TERT*, which encode; the catalytic reverse transcriptase telomerase, can have subtle and milder effects on hematopoietic function; presentation in adults is not unusual, as moderate aplastic anemia, which can be chronic and not progressive, and isolated macrocytic anemia or thrombocytopenia. Physical anomalies are usually not present, but early hair graying is a clue to the diagnosis. A detailed personal and family history may disclose pulmonary fibrosis and hepatic cirrhosis. Variable penetrance means that *TERT* and *TERC* mutations represent risk factors for marrow failure, as family members with the same mutations may have normal or only slight hematologic abnormalities but more subtle evidence of (compensated) hematopoietic insufficiency. Patients can suffer pulmonary or liver fibrosis in the absence of hematologic abnormalities.

■ PATHOPHYSIOLOGY

Bone marrow failure results from severe damage to the hematopoietic cell compartment. In aplastic anemia, replacement of the bone marrow by fat is apparent in the morphology of the biopsy specimen (Fig. 107-1) and magnetic resonance imaging (MRI) of the spine. Cells bearing the CD34 antigen, a marker of early hematopoietic cells, are greatly diminished, and in functional studies, committed and primitive progenitor cells are virtually absent; in vitro assays have suggested that the stem cell pool is reduced to ≤1% of normal in severe disease at the time of presentation.

Constitutional Genetic Syndromes An intrinsic stem cell defect exists for the constitutional aplastic anemias: in a critical DNA repair pathway in Fanconi anemia, manifested in the laboratory as chromosome damage and cell death on exposure to certain chemical agents. In the telomere biology disorders, inability to repair telomeres or to protect chromosome ends is the result of mutations in genes of the telomerase complex or the shelterin proteins; telomere defects limit the cell's capacity to proliferate. Mutations in the *GATA2* and *RUNX1* genes affect signal transduction and transcriptional regulation in hematopoietic gene networks.

Chemical and Drug Injury Extrinsic damage to the marrow follows massive physical or chemical insults such as high doses of radiation and toxic chemicals. For the more common idiosyncratic reaction to modest doses of medical drugs, altered drug metabolism has been invoked as a mechanism. The metabolic pathways of many drugs and chemicals, especially if they are polar and have limited water solubility, involve enzymatic degradation to highly reactive electrophilic compounds; these intermediates are toxic because of their propensity to bind to cellular macromolecules. For example, derivative hydroquinones and quinolones are responsible for benzene-induced tissue injury. Excessive generation of toxic intermediates or failure to detoxify the intermediates may be genetically determined and apparent only on specific drug challenge; the complexity and specificity of the pathways imply multiple susceptibility loci and would provide an explanation for the rarity of idiosyncratic drug reactions.

Immune-Mediated Stem Cell Destruction The recovery of marrow function in some patients prepared for bone marrow transplantation with antilymphocyte globulin first suggested that aplastic anemia might be immune mediated. Blood and bone marrow cells of patients can suppress normal hematopoietic progenitor cell growth, and removal of T cells from aplastic anemia bone marrow improves hematopoiesis in vitro. Increased numbers of activated cytotoxic T-cell clones, which can be visualized in sophisticated single-cell experiments (single-cell RNA sequencing, time-of-flight flow cytometry) usually decline with successful immunosuppressive therapy. Type 1 cytokines are implicated, and interferon γ (IFN-γ) induces Fas expression on CD34 cells, leading to apoptotic cell death. Immune selection can result in loss of human leukocyte antigen (HLA) expression on hematopoietic stem cells, and PNH clones are observed almost exclusively in immune marrow failure. The rarity of aplastic anemia despite common exposures (medicines, seronegative hepatitis) suggests that genetically determined features of the immune response can convert a normal physiologic response into a sustained abnormal autoimmune process, including polymorphisms in histocompatibility antigens, cytokine genes, and genes that regulate T-cell polarization (maturation toward helper or cytotoxic phenotypes) and effector function.

CLINICAL FEATURES

History Aplastic anemia can appear abruptly or insidiously. Bleeding is the most common early symptom; a complaint of days to weeks of easy bruising, oozing from the gums, nose bleeds, heavy menstrual flow, and sometimes petechiae will have been noticed. With thrombocytopenia, massive hemorrhage is unusual, but small amounts of bleeding in the central nervous system can result in catastrophic intracranial or retinal hemorrhage. Symptoms of anemia are also frequent, including lassitude, weakness, shortness of breath, and a pounding sensation in the ears. Infection is an unusual first symptom in aplastic anemia (unlike in agranulocytosis, where pharyngitis, anorectal infection, or frank sepsis occurs early). Patients often feel and look remarkably well despite drastically reduced blood counts. Systemic complaints and weight loss should point to other etiologies of pancytopenia. Prior medical drug use, chemical exposure, and preceding viral illnesses must often be elicited with directed questioning. A family history of hematologic diseases or blood abnormalities is suspicious for IBMFS. Specific features of the history can help: pulmonary or liver fibrosis or early hair graying points to a telomere biology disorder, and a family history of unusual infections and warts to *GATA2* deficiency.

Physical Examination Petechiae and ecchymoses are typical, and retinal hemorrhages may be present. Pallor of the skin and mucous membranes is common. Infection on presentation is unusual but may occur if the patient has been symptomatic for a few weeks. Lymphadenopathy and splenomegaly are highly atypical of aplastic anemia. Suggestive findings of IBMFS include café au lait spots and short stature for Fanconi anemia; peculiar nails and leukoplakia for dyskeratosis congenita; early graying (and use of hair dyes to mask it!) for a telomere biology disorder.

LABORATORY STUDIES

Blood The smear shows large erythrocytes and a paucity of platelets and granulocytes. Mean corpuscular volume (MCV) is commonly increased. Reticulocytes are absent or few, and lymphocyte numbers may be normal or reduced. The presence of immature myeloid forms suggests leukemia or MDS; nucleated red blood cells (RBCs) suggest marrow fibrosis or tumor invasion; abnormal platelets suggest either peripheral destruction or MDS.

Bone Marrow The bone marrow is usually readily aspirated but dilute on smear, and the fatty biopsy specimen may be grossly pale on withdrawal; a "dry tap" instead suggests fibrosis or myelophthisis. In severe aplasia, the smear of the aspirated specimen shows only red cells, residual lymphocytes, and stromal cells; the biopsy (which should be >1 cm in length) is superior for determination of cellularity and shows mainly fat under the microscope, with hematopoietic cells occupying <25% of the marrow space; sometimes, the biopsy is virtually all fat. The correlation between marrow cellularity and disease severity is imperfect; patients with moderate disease by blood counts can have "empty" iliac crest biopsies, whereas "hot spots" of hematopoiesis may be seen in severe cases. Residual hematopoietic cells should have normal morphology, except for mildly megaloblastic erythropoiesis; megakaryocytes are greatly reduced and usually absent. Granulomas may indicate an infectious etiology of the marrow failure.

Ancillary Studies Chromosome breakage studies of peripheral blood using diepoxybutane or mitomycin C should be performed on children and younger adults to diagnose Fanconi anemia. Very short telomere length of peripheral blood leukocytes is seen in telomere biology disorders, which can be pursued by family studies and nucleotide sequencing. Chromosome studies of bone marrow cells are often abnormal in MDS but should be negative in typical aplastic anemia. Flow cytometry offers a sensitive diagnostic test for PNH; a PNH clone is strongly suggestive of immune-mediated disease. Serologic studies may show evidence of recent viral infection, such as Epstein-Barr virus and HIV. Posthepatitis aplastic anemia is seronegative.

Genomics Next-generation sequencing allows for a large number of genes to be tested for the presence of pathogenic mutations. Panels are available commercially and in academic laboratories. While expensive, they are very useful and sometimes critical in establishing the correct diagnosis. Germline gene panels examine 50 or more genes etiologic in constitutional bone marrow failure, including many for which functional assays (described above) are not available. A germline panel should be considered for all children and those adults with suspicious clinical features or family histories. Somatic mutations are sought when MDS is suspected using myeloid neoplasm gene panels that can query about 100 genes that are recurrently mutated in MDS and acute myeloid leukemia (AML). Pathogenic mutations in spliceosome genes and genes in the cohesion gene family are frequent in MDS and unexpected in aplastic anemia.

DIAGNOSIS

The diagnosis of aplastic anemia is usually straightforward, based on the combination of pancytopenia with a fatty bone marrow. Immune aplastic anemia is a disease of the young and should be a leading diagnosis in the pancytopenic adolescent or young adult. When pancytopenia is secondary, the primary diagnosis is usually obvious from either history or physical examination: the massive spleen of alcoholic cirrhosis, the history of metastatic cancer or SLE, or miliary tuberculosis on chest radiograph (Table 107-1).

Diagnostic problems can occur with atypical presentations and among related hematologic diseases. Patients with bone marrow hypocellularity may have depression of only one or two of three blood lines, later progressing to pancytopenia. The most important differential diagnoses are between acquired and constitutional aplastic anemia, and between aplastic anemia and MDS. The bone marrow in constitutional aplastic anemia is usually morphologically indistinguishable from the aspirate in acquired disease (an exception is *GATA2* deficiency with its characteristic megakaryocyte atypia). The diagnosis can be suggested by family history, abnormal blood counts since childhood, or the presence of associated, sometimes subtle physical anomalies. Flow cytometry may show distinctive patterns to guide diagnosis, for sample, the absence of B and natural killer (NK) cells seen in *GATA2* deficiency. Genomic testing for pathogenic mutations in genes etiologic in constitutional marrow failure syndromes can discriminate acquired from inherited aplastic anemia (but results may not return for several weeks, a problem in the severely pancytopenic patient). AML and MDS in a pedigree should prompt screening for an inherited predisposition syndrome, such as familial *RUNX1* disorder. Aplastic anemia may be difficult to distinguish from hypocellular MDS: MDS is favored by finding morphologic abnormalities, particularly of megakaryocytes and myeloid precursor cells, and typical cytogenetic abnormalities and somatic mutations on genomic screening of myeloid neoplasm genes (see above). There remains an unclear boundary between immune aplastic anemia and low-risk MDS: patients with deletion of 13q and 20q may respond well to immunosuppression, and mutations in genes such as *DNMT3A* and *ASXL1* occur in both diseases.

PROGNOSIS

The natural history of severe aplastic anemia is rapid deterioration and death. Historically, provision first of RBCs and later of platelet transfusions and effective antibiotics were of some benefit, but few patients showed spontaneous recovery. The major prognostic determinant is the blood count. Severe disease historically has been defined by the presence of two of three parameters: absolute neutrophil count <500/μL, platelet count <20,000/μL, and corrected reticulocyte count <1% (or absolute reticulocyte count <60,000/μL). In the era of effective immunosuppressive therapies, absolute numbers of reticulocytes (>25,000/μL) and lymphocytes (>1000/μL) are better predictors of response to treatment and long-term outcome.

Other prognostic factors include the presence of a PNH clone, short telomeres on presentation, and somatically mutated white cells. Even small PNH clones may indicate an immune pathophysiology and responsiveness to immunosuppressive therapies. Telomere shortening in the granulocytes of many patients likely reflects stem cell reserve, regenerative stress, and susceptibility to chromosomal instability; shorter telomeres in lymphocytes are more consistent with a telomere biology disorder. Collectively, the presence of mutations in the same myeloid neoplasia genes that are mutated in some clonal hematopoiesis of indeterminate potential (*ASXL1*) is associated with worse prognosis and clonal evolution.

TREATMENT

Aplastic Anemia

Severe acquired aplastic anemia can be cured by replacement of the absent hematopoietic cells (and the immune system) by stem cell transplant, or it can be ameliorated by suppression of the immune system to allow recovery of the patient's residual bone marrow function. Glucocorticoids are not of value as primary therapy, and among their multiple toxicities is increased susceptibility to fungal infection. Suspicious drugs or chemicals should be discontinued; however, spontaneous recovery of severe blood count depression is rare, and a waiting period before beginning treatment may not be advisable unless the blood counts are only modestly depressed.

HEMATOPOIETIC STEM CELL TRANSPLANTATION

This is the first choice for the younger patient with a fully histocompatible sibling donor **(Chap. 119)**. HLA typing should be ordered as soon as the diagnosis of aplastic anemia is established in transplant-eligible patients. In transplant candidates, transfusion of blood from family members should be avoided so as to prevent sensitization to histocompatibility antigens. In general, limited numbers of blood products probably do not greatly affect outcome, especially when blood products are depleted of leukocytes. For allogeneic transplant from fully matched siblings, long-term survival rates for children are ~90% or better. Transplant morbidity and mortality are increased among adults, due to the higher risk of chronic GVHD and infections. Nevertheless, transplant should be considered early in all but the most elderly, including from alternative donors.

Most patients do not have a suitable sibling donor. Occasionally, a full phenotypic match can be found within the family and serve as well. More broadly, matched unrelated donors in large registries are available for the majority of Caucasian patients. With high-resolution matching at HLA, outcomes are similar to those with sibling donors, although complications (GVHD and infection) are more frequent. Cord blood also can be a source of stem cells, especially for children. Matched unrelated donor transplants are often considered as initial treatment in children and as salvage therapy for adults after failed immunosuppression. Transplantation from an HLA haploidentical family donor is increasingly popular, as a donor is almost always readily available. Posttransplant cyclophosphamide appears effective in preventing GVHD and is increasingly popular; lymphocyte depletion before donor cell infusions has been developed and employed. Transplant protocols for marrow failure, except in the case of alternative donor transplant, now usually do not include radiation in order to avoid late occurrence of cancer.

IMMUNOSUPPRESSION WITH ELTROMBOPAG

Early, serendipitous observation of spontaneous improvement of blood counts despite lack of engraftment of donor marrow led to speculation of an immune pathophysiology and also purposeful administration of antilymphocyte globulin as sole treatment. Addition of cyclosporine to antithymocyte globulin (ATG) increased the response rate, and cyclosporine helped maintain remissions. Many other attempts to improve on the regimen of ATG plus cyclosporine failed. Although hematopoietic growth factors generally are of little utility in immune aplastic anemia, thrombopoietic mimetics, surprisingly, rescued refractory aplastic anemia as monotherapy and increased both overall and complete response rates in combination with immunosuppression in treatment-naïve disease.

The current standard regimen of ATG, cyclosporine, and eltrombopag ("triple therapy") induces hematologic recovery (independence from transfusion and a leukocyte count adequate to prevent infection) in 70–80% of patients. Children do especially well with immunosuppressive therapy, whereas older adult patients can suffer complications due to the presence of comorbidities. Children have less improvement in response with the addition of eltrombopag to immunosuppressive therapy than do adults. An early robust hematologic response correlates with long-term survival. Improvement in granulocyte number is generally apparent within 2 months of treatment. Most recovered patients continue to have some degree of blood count depression, the MCV remains elevated, and bone marrow cellularity returns toward normal very slowly if at all. Relapse (some degree of recurrent pancytopenia) is frequent, often occurring as cyclosporine or eltrombopag is tapered or discontinued; most, but not all, patients respond to reinstitution of therapy. Some patients become dependent on continued cyclosporine or eltrombopag administration. "Clonal evolution," isolated chromosomal abnormalities or the development of MDS, with typical cytogenetic aberrations and abnormal marrow morphology, occurs in ~10–15% of treated patients over a decade following initiation of ATG, usually but not invariably associated with a return of pancytopenia, and some patients develop leukemia. A laboratory diagnosis of PNH can generally be made at the time of presentation of aplastic anemia by flow cytometry; recovered patients may have frank hemolysis if the PNH clone expands. Bone marrow examinations should be performed if there is an unfavorable change in blood counts.

Horse ATG is administered as intravenous infusions and requires hospitalization. Rabbit ATG is much less effective, perhaps because it reduces T-regulatory cell numbers in patients. Serum sickness, a flulike illness with a characteristic cutaneous eruption and arthralgia, may develop ~10 days after initiating treatment. Methylprednisolone is administered with ATG to ameliorate the immune consequences of heterologous protein infusion. (Excessive or extended glucocorticoid therapy is associated with avascular joint necrosis.)

Cyclosporine is administered orally at an initial high dose, with subsequent adjustment according to blood levels. Its most important side effects are nephrotoxicity, hypertension, and seizures. Some centers replace cyclosporine with tacrolimus.

Eltrombopag, a thrombopoietin (TPO) mimetic, is administered once daily for 6 months; its most important side effect is hepatotoxicity. The likely mechanism of action of TPO mimetics is stimulation of the hematopoietic stem cell, but iron chelation and increased regulatory T cells are also possibly beneficial effects. Eltrombopag added to first-line immunosuppression with horse ATG markedly increased overall and complete response rates, to about 80 and 50%, respectively, and triple therapy became the U.S. Food and Drug Administration (FDA)-approved standard-of-care therapy in 2018. Eltrombopag was first licensed in the refractory setting where it showed response rates of ~40%. Romiplostim, another TPO mimetic, also has been reported as effective.

Most patients with aplastic anemia lack a fully matched marrow donor, and immunosuppression is the treatment of choice. Overall survival is equivalent with transplantation and immunosuppression. However, successful transplant cures marrow failure, whereas patients who recover adequate blood counts after immunosuppression remain at risk of relapse and malignant evolution. Increasing age and the severity of neutropenia are the most important factors weighing in the decision between transplant and immunosuppression in adults who have a matched family donor: older patients do better with ATG and cyclosporine, whereas transplant is preferred if neutropenia is profound. Patients in whom immunosuppression has failed can be salvaged by stem cell transplant later.

ANDROGENS

The effectiveness of androgens has not been verified in controlled trials, but occasional patients will respond or even demonstrate blood count dependence on continued therapy. Sex hormones upregulate telomerase gene activity in vitro, which is possibly also their mechanism of action in improving marrow function. For patients with moderate disease, especially if a telomere gene defect is present, a 3- to 4-month trial may improve all blood counts **(Chap. 483)**.

SUPPORTIVE CARE

Meticulous medical attention is required so that the patient may survive to benefit from definitive therapy or, having failed treatment, to maintain a reasonable existence in the face of pancytopenia. First and most important, infection in the presence of severe neutropenia must be aggressively treated by prompt institution of parenteral, broad-spectrum antibiotics. Therapy is empirical and must not await results of culture, although specific foci of infection such as oropharyngeal or anorectal abscesses, pneumonia, sinusitis, and typhlitis (necrotizing colitis) and infected lines should be sought on clinical assessment and with radiographic studies. Persistent or recrudescent fever implies fungal disease: *Candida* and *Aspergillus* are common, especially after several courses of antibacterial antibiotics. A major reason for the improved prognosis in aplastic anemia has been the development of better antifungal drugs and the timely institution of such therapy when infection is suspected. Granulocyte transfusions can be effective when bacterial or fungal infection is progressive or refractory to antibiotics. Nonabsorbed antibiotics for gut decontamination are poorly tolerated and unproven, nor does reverse isolation reduce mortality from infections.

Both platelet and erythrocyte numbers can be maintained by transfusion. Alloimmunization historically limited the usefulness of platelet transfusions and is now minimized by several strategies, including use of single donors to reduce exposure and physical or chemical methods to diminish leukocytes in the product; HLA-matched platelets are usually effective in alloimmunized patients refractory to random donor products. The use of low-dose glucocorticoids to induce "vascular stability" is unproven and not recommended. With prophylactic platelet transfusions, the goal is to maintain the platelet count >10,000/μL (oozing from the gut increases sharply at counts <5000/μL). Menstruation should be suppressed either by oral estrogens or follicle-stimulating hormone/luteinizing hormone antagonists. Aspirin and other nonsteroidal anti-inflammatory agents must be avoided in the presence of thrombocytopenia.

RBCs should be transfused so as to allow patients a normal level of activity, usually at a hemoglobin value of 70 g/L (90 g/L if there is underlying cardiac or pulmonary disease); a regimen of 2 units every 2 weeks will replace normal losses in a patient without a functioning bone marrow. In chronic anemia, the iron chelators deferoxamine and deferasirox should be added at approximately the fiftieth transfusion to avoid secondary hemochromatosis.

Hematopoietic growth factors (HGFs) such as erythropoietin (EPO) and granulocyte colony-stimulating factor (G-CSF) are not effective in aplastic anemia.

PURE RED CELL APLASIA

Other more restricted forms of marrow failure occur, in which only a single cell type is affected and the marrow shows corresponding absence or decreased numbers of specific precursor cells: aregenerative anemia as in PRCA (see below), thrombocytopenia with amegakaryocytosis **(Chap. 120)**, and neutropenia without marrow myeloid cells in agranulocytosis **(Chap. 67)**. In general, and in contrast to aplastic anemia and MDS, the unaffected lineages appear quantitatively and qualitatively normal. Agranulocytosis, the most frequent of these syndromes, is usually a complication of medical drug use, either by a mechanism of direct chemical toxicity or by immune destruction. Agranulocytosis has an incidence similar to aplastic anemia (but geographically more frequent in Europe than in Asia); in contrast to aplastic anemia, agranulocytosis is more prevalent among older adults and in women. Agranulocytosis should resolve with discontinuation of exposure, but significant mortality is attached to neutropenia in the older and often previously unwell patient. Both pure white cell aplasia (agranulocytosis without incriminating drug exposure) and amegakaryocytic thrombocytopenia are exceedingly rare and, like PRCA, appear to be due to a destructive immune response. In all of the single-lineage failure syndromes, progression to pancytopenia or leukemia is unusual.

DEFINITION AND DIFFERENTIAL DIAGNOSIS

PRCA is characterized by anemia, reticulocytopenia, and absent or rare erythroid precursor cells in the bone marrow. The classification of PRCA is shown in **Table 107-4**. In adults, PRCA is usually acquired. An identical syndrome can occur constitutionally: Diamond-Blackfan anemia, or congenital PRCA, is diagnosed at birth or in early childhood and often responds to glucocorticoid treatment; mutations in ribosome protein genes are etiologic. Temporary red cell failure occurs in transient aplastic crisis of hemolytic anemias due to acute parvovirus infection **(Chap. 202)** and in transient erythroblastopenia of childhood, which occurs in normal children.

CLINICAL ASSOCIATIONS AND ETIOLOGY

PRCA has important associations with immune system diseases. A minority of cases occur with a thymoma. More frequently, red cell aplasia can be the major manifestation of large granular lymphocytosis or complicates chronic lymphocytic leukemia. Some patients may be hypogammaglobulinemic. A ribosomal protein gene is deleted in the 5q− syndrome, such that the MDS may manifest as an acquired red cell aplasia. Occasionally (as compared to agranulocytosis), PRCA can

TABLE 107-4 Classification of Pure Red Cell Aplasia

- Self-limited
  - Transient erythroblastopenia of childhood
  - Transient aplastic crisis of hemolysis (acute B19 parvovirus infection)
- Fetal red blood cell aplasia
  - Nonimmune hydrops fetalis (in utero B19 parvovirus infection)
- Constitutional pure red cell aplasia
  - Congenital pure red cell aplasia (Diamond-Blackfan anemia)
- Acquired pure red cell aplasia
  - Primary immune
  - MDS (5q- syndrome)
  - Cancer
    - Thymoma
    - Lymphoid malignancies (and more rarely other hematologic diseases)
    - Paraneoplastic to solid tumors
- Connective tissue disorders with immunologic abnormalities
  - Systemic lupus erythematosus, juvenile rheumatoid arthritis, rheumatoid arthritis
  - Multiple endocrine gland insufficiency
- Viruses
  - Persistent B19 parvovirus, hepatitis, adult T-cell leukemia virus, Epstein-Barr virus
- Pregnancy
- Drugs
  - Especially phenytoin, azathioprine, chloramphenicol, procainamide, isoniazid
- Antibodies to erythropoietin
- Idiopathic (immune)

be due to an idiosyncratic drug reaction. Subcutaneous administration of EPO has provoked PRCA mediated by neutralizing antibodies to the hormone. PRCA due to antibodies to blood group antigens (isoagglutins) is a complication of allogeneic stem cell transplant. For most PRCAs, T-cell inhibition is probably the prevalent immune mechanism.

PERSISTENT PARVOVIRUS B19 INFECTION

Chronic parvovirus infection is a treatable cause of red cell aplasia. This common virus causes a benign exanthem of childhood (fifth disease) and a polyarthralgia/arthritis syndrome in adults. In patients with underlying hemolysis (or any condition that increases demand for RBC production), parvovirus infection can cause a transient aplastic crisis and an abrupt but temporary worsening of the anemia due to failed erythropoiesis. In healthy individuals, acute infection is resolved by production of neutralizing antibodies to the virus, but in the setting of congenital, acquired, or iatrogenic immunodeficiency, persistent viral infection may occur. The bone marrow shows red cell aplasia and the presence of giant pronormoblasts **(Fig. 107-2)**, which is the cytopathic sign of B19 parvovirus infection. Viral tropism for human erythroid progenitor cells is due to its use of erythrocyte P antigen as a cellular receptor for entry. Direct cytotoxicity of virus causes anemia if demands on erythrocyte production are high; in normal individuals, the temporary cessation of red cell production is not clinically apparent, and skin and joint symptoms are mediated by immune complex deposition.

TREATMENT

Pure Red Cell Aplasia

History, physical examination, and routine laboratory studies may disclose an underlying disease or a drug exposure. Thymoma should be sought by radiographic procedures; tumor excision is indicated, but anemia does not necessarily improve with surgery. The diagnosis of parvovirus infection requires detection of viral DNA sequences in the blood (IgG and IgM antibodies are

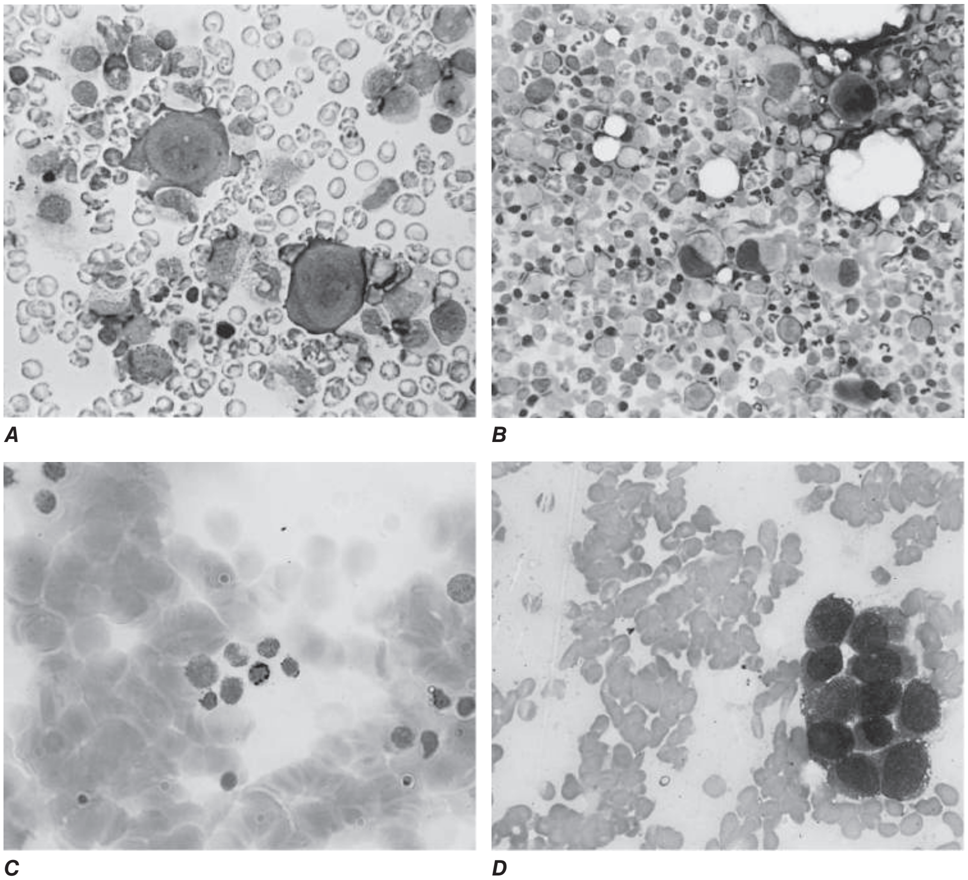

FIGURE 107-2 Pathognomonic cells in marrow failure syndromes. ***A.*** Giant pronormoblast, the cytopathic effect of B19 parvovirus infection of the erythroid progenitor cell. ***B.*** Uninuclear megakaryocyte and microblastic erythroid precursors typical of the 5q– myelodysplastic syndrome (MDS). ***C.*** Ringed sideroblast showing perinuclear iron granules, typical of *SF3B1* mutated MDS. ***D.*** Tumor cells present on a touch preparation made from the marrow biopsy of a patient with metastatic carcinoma.

TABLE 107-5 World Health Organization (WHO) 2022 Classification of Myelodysplastic Syndromes (MDS)/Neoplasms

NAME	BLASTS	CYTOGENETICS	MUTATIONS
MDS with Defining Genetic Abnormalities			
MDS with low blasts and isolated 5q deletion (MDS-5q)	<5% BM and <2% PB	5q deletion alone, or with 1 other abnormality other than –7/7q del	
MDS with low blasts and *SF3B1* mutation (MDS-*SF3B1*)[a]	<5% BM and <2% PB	Absence of 5q–, –7, or complex karyotype	*SF3B1*
MDs with biallelic TP53 inactivation	<20% BM and PB	Usually complex	Two or more *TP53* mutations or evidence of copy number loss
MDS, Morphologically Defined			
MDS with low blasts	<5% BM and <2% PB		
MDS, hypoplastic[b]	<5% BM and <2% PB		
MDS with increased blasts-1	5–9% BM or 2–4% PB		
MDS with increased blasts-2	10–19% BM or 5–19% PB or Auer rods		
MDS with fibrosis	5–19% BM; 2–19% PB		

[a]Detection of ≥15% ring sideroblasts may substitute *SF3B1*; [b]≤25% BM cellularity.

Abbreviations: BM, bone marrow; PB, peripheral blood.

commonly absent). The presence of erythroid colonies has been considered predictive of response to immunosuppressive therapy in idiopathic PRCA.

Red cell aplasia is compatible with long-term survival with supportive care alone: a combination of erythrocyte transfusions and iron chelation. For persistent B19 parvovirus infection, almost all patients respond to intravenous immunoglobulin therapy. The majority of patients with acquired PRCA respond favorably to immunosuppression: glucocorticoids, cyclosporine, ATG, azathioprine, and cyclophosphamide are effective.

MYELODYSPLASTIC SYNDROMES

DEFINITION

The MDS are a heterogeneous group of hematologic disorders characterized by both (1) cytopenias due to bone marrow failure and (2) a high risk of development of AML. Anemia due to ineffective erythropoiesis, often with thrombocytopenia and neutropenia, occurs with dysmorphic (abnormal appearing) and usually a cellular bone marrow, or with specific chromosome abnormalities or acquired mutations. In patients with "low-risk" MDS, marrow failure dominates the clinical course. In other patients, myeloblasts are present at diagnosis, chromosomes are abnormal, and the "high risk" is due to leukemic progression. MDS may be fatal due, most often, to complications of pancytopenia or to progression to leukemia, but a large proportion of patients will die of concurrent disease, the comorbidities typical in an elderly population. An initial nosology of these often-confusing entities was first developed by the French-American-British Cooperative Group in 1983. Five subtypes were defined then: refractory anemia (RA), refractory anemia with ringed sideroblasts (RARS), refractory anemia with excess blasts (RAEB), refractory anemia with excess blasts in transformation (RAEB-t), and chronic myelomonocytic leukemia (CMML). The World Health Organization (WHO) classification (2002) recognized that the distinction between RAEB-t and AML was arbitrary, grouped them together as acute leukemia, and clarified that CMML behaves as a myeloproliferative disease. Since then, classification has become more refined and now incorporates molecularly defined subtypes. Two classifications currently exist: the fifth edition of the WHO and the International Consensus Classification (ICC) (**Tables 107-5 and 107-6**). Both classification systems incorporate morphology, cytopenias, bone marrow blasts, cytogenetics, and molecular findings. In the ICC classification, a new category has been introduced called MDS/AML, which emphasizes the disease continuum and similar outcomes between high blast count MDS and AML. In both classifications, a new molecularly defined subtype of MDS incorporating biallelic or multi-hit *TP53* has been added. In the future, MDS will become increasingly molecularly defined, as more is learned about correlation with clinical outcomes.

The diagnosis of MDS can be a challenge, even for the expert, because sometimes subtle clinical and pathologic features must be distinguished, and precise diagnostic categorization requires a hematopathologist knowledgeable in the latest classification scheme. Unfortunately, agreement among pathologists on morphologic features and classification is imperfect: changes in the appearance of megakaryocytes are more reliable than loss of granules in neutrophil precursors or dyserythropoiesis. Dysplastic changes can be observed in normal individuals, and they also can occur with vitamin deficiencies and as drug effects. Genomic testing is increasingly routine but can be difficult to interpret, as in differences between somatic and germline mutations, pathogenic mutations versus those of unknown significance (clonal hematopoiesis increases in frequency with normal aging), and clone size and changes over time. It is important that the internist and primary

TABLE 107-6 International Consensus Classification (ICC) 2022 Classification of Myelodysplastic Syndromes (MDS)/Neoplasms

NAME	BLASTS	CYTOGENETICS	MUTATIONS
MDS with 5q deletion (MDS-5q)	<5% BM and <2% PB	5q deletion alone, or with 1 other abnormality other than –7/7q del	Any, except multi-hit *TP53*
MDS with *SF3B1* mutation (MDS-*SF3B1*)	<5% BM and <2% PB	Absence of 5q–, –7, abn3q26.2, or complex karyotype	*SF3B1* (≥10% VAF), without multi-hit *TP53*, or *RUNX1*
MDS, NOS without dysplasia (no dysplastic lineages)	<5% BM and <2% PB	–7/7q del or complex	Any except multi-hit *TP53* or *SF3B1* (VAF ≥10%)
MDS, NOS with single-lineage dysplasia (1 dysplastic lineage)	<5% BM and <2% PB	Any except 5q–	Any except multi-hit *TP53* or *SF3B1* (VAF ≥10%)
MDS, NOS with multi-lineage dysplasia (≥2 dysplastic lineages)	<5% BM and <2% PB	Any except 5q–	Any except multi-hit *TP53* or *SF3B1* (VAF ≥10%)
MDS with increased blasts-1	5–9% BM or 2–9% PB	Any	Any except multi-hit *TP53*
MDS with increased blasts-2	10–19% BM or PB	Any, except AML defining	Any except *TP53, NPM1, bZIP, CEBPA*

Abbreviations: AML, acute myeloid leukemia; BM, bone marrow; NOS, not otherwise specified; PB, peripheral blood; VAF, variant allele frequency.

care physician be sufficiently familiar with MDS to expedite referral to a hematologist because many new therapies are now available to improve hematopoietic function and the judicious use of supportive care can improve the patient's quality of life.

EPIDEMIOLOGY

MDS is a disease of the elderly; the mean age at onset is older than 70 years. There is a slight male predominance. MDS is a relatively common form of bone marrow failure, with reported incidence rates of 35 to >100 per million persons in the general population and 120 to >500 per million in older adults. Estimates of incidence in the United States range from 30,000 to 40,000 new cases annually and a prevalence of 60,000–120,000 in the population. Rates of MDS have increased over time due to better recognition of the syndrome by physicians and an aging population.

MDS is rare in children, in whom it often has a constitutional genetic basis that can be identified on genomic screens of myeloid cancer predisposition panels.

Secondary or therapy-related MDS, usually related to previous iatrogenic exposure to alkylating agents and other chemotherapy or radiation, is not age related.

ETIOLOGY AND PATHOPHYSIOLOGY

MDS is associated with environmental exposures such as radiation, benzene, and smoking; other risk factors have been reported inconsistently. Secondary, therapy-related MDS occurs as a late toxicity of cancer treatment; radiation and the radiomimetic alkylating agents such as busulfan, nitrosourea, or procarbazine (with a latent period of 5–7 years); or the DNA topoisomerase inhibitors (2-year latency). Acquired aplastic anemia, Fanconi anemia, and other constitutional marrow failure diseases can evolve into MDS; occasionally, MDS in adults is recognized as due to germline *GATA2*, *RUNX1,* or telomere gene mutations. MDS in patients <40 years of age has a high probability of being constitutional, and germline genetic screening is indicated. The typical MDS patient does not have a suggestive environmental exposure history or a preceding hematologic disease.

MDS is a clonal hematopoietic stem cell disorder characterized by disordered cell proliferation, impaired differentiation, and aberrant hematopoiesis, resulting in cytopenias and risk of progression to leukemia. MDS is a disease of older persons consistent with accumulation of mutations within a hematopoietic stem cell in an aging marrow environment. Both chromosomal and genetic instability lead to MDS; both are aging-related. Cytogenetic abnormalities are found in approximately one-half of patients, some of the same specific lesions are also seen in leukemia; aneuploidy (chromosome loss or gain) is more frequent than translocations. Accelerated telomere attrition may destabilize the genome in marrow failure and predispose to acquisition of chromosomal lesions. Cytogenetic abnormalities are not random (loss of all or part of 5, 7, and 20, trisomy of 8) and may be related to etiology (11q23 following topoisomerase II inhibitors). The type and number of cytogenetic abnormalities strongly correlate with the probability of leukemic transformation and survival.

Genomics has illuminated the role of specific mutations and distinct molecular pathways in the pathophysiology of MDS. Somatic mutations in about 100 genes, which are recurrently present in myeloid neoplasms, arise in the abnormal marrow cells (and are absent in the germline). Many of the same genes are mutated in AML and in MDS, whereas others are distinctive in subtypes of MDS. A prominent example is *SF3B1*, in which mutations strongly associate with sideroblastic anemia. Some mutations correlate with prognosis: spliceosome defects (like *SF3B1*) correlate with favorable outcome, and mutations in *EZH2, TP53, RUNX1*, and *ASXL1* correlate with poor outcome. Correlation and exclusion in the pattern of mutations indicate a functional genomic architecture. Driver genes mutated early are consistent with normal blood counts and marrow morphology, but expanded clones of mutated cells are susceptible to malignant transformation with the acquisition of additional mutations. Deep sequencing in patients whose MDS evolved to AML shows clonal succession, with founder clones acquiring additional mutations to become dominant clones. Mutations and cytogenetic abnormalities are not independent: *TP53* mutations associate with complex cytogenetic abnormalities and *TET2* mutations with normal cytogenetics. The prevalence of abnormal cells by morphology underestimates bone marrow involvement by MDS clones, as cells normal in appearance are derived from the abnormal clones. Presenting and evolving hematologic manifestations result from the accumulation of multiple genetic lesions: loss of tumor-suppressor genes, activating oncogene, epigenetic pathways that affect mRNA processing and methylation status, and other harmful alterations. Pathophysiology has been linked to mutations and chromosome abnormalities in some specific MDS syndromes. The 5q− deletion leads to heterozygous loss of a ribosomal protein gene, which mimics constitutional red cell aplasia. An immune pathophysiology may be important in lower-risk MDS, as cytopenias can respond to immunosuppressive therapy as administered for aplastic anemia. In general for MDS, our understanding remains incomplete regarding the immune system and its cells and cytokines; the role of the hematopoietic stem cell niche, the microenvironment, and cell–cell interactions; the fate of normal cells in the Darwinian competitive environment of the dysplastic marrow; and how mutant cells produce marrow failure.

CLINICAL FEATURES

Anemia dominates the early course. Most symptomatic patients complain of the gradual onset of fatigue and weakness, dyspnea, and pallor, but at least one-half of patients are asymptomatic, and their MDS is discovered only incidentally on routine blood counts. Previous chemotherapy or radiation exposure is an important historic fact. Fever and weight loss are more often features of a myeloproliferative rather than myelodysplastic process. MDS in childhood is rare and, when diagnosed, implies an underlying genetic disease. Children with Down syndrome are susceptible to MDS as well as leukemia. A family history may indicate a hereditary form of sideroblastic anemia, Fanconi anemia, or a telomeropathy. Inherited *GATA2* mutations, as in the MonoMAC syndrome (with increased susceptibility to viral, mycobacterial, and fungal infections, as well as deficient numbers of monocytes, natural killer cells, and B lymphocytes), predispose to MDS. Germline *RUNX1* mutations also confer a high risk of MDS and leukemia, often preceded by years of modest thrombocytopenia. A family history is important in all MDS patients, as constitutional mutations may not result in manifest disease until adulthood.

The physical examination in MDS is remarkable for signs of anemia; approximately 20% of patients have splenomegaly. Accompanying autoimmune syndromes are not infrequent. MDS with coexisting inflammatory disease (including Sweet's syndrome) should prompt consideration of the newly described vacuoles, E1 enzyme, X-linked, autoinflammatory, somatic (VEXAS) syndrome, a hematopoietic cell disorder caused by somatic mutations in the *UBA1* gene. In the younger patient, stereotypical anomalies point to a constitutional syndrome (short stature, abnormal thumbs in Fanconi anemia; early graying in the telomeropathies; cutaneous warts in *GATA2* deficiency).

LABORATORY STUDIES

Blood Anemia is present in most cases, either alone or as part of bi- or pancytopenia; isolated neutropenia or thrombocytopenia is more unusual. Macrocytosis is common, as in most marrow failure disease. Platelets also are large and lack granules. In functional studies, they may show marked abnormalities, and patients may have bleeding symptoms despite seemingly adequate numbers. Neutrophils are hypogranulated; have hyposegmented, ringed, or abnormally segmented nuclei; contain Döhle bodies; and may be functionally deficient. Circulating myeloblasts usually correlate with marrow blast numbers, and their quantity is important for classification and prognosis. The total white blood cell (WBC) count is usually normal or low, except in CMML. As in aplastic anemia, MDS can be associated with a clonal population of PNH cells. Genetic testing is commercially available for constitutional syndromes.

Bone Marrow The bone marrow is usually normal or hypercellular, but in about 20% of cases, it is sufficiently hypocellular to be confused with aplastic anemia. No single characteristic feature of marrow morphology distinguishes MDS, but the commonly observed are dyserythropoietic changes (especially nuclear abnormalities) and ringed sideroblasts in the erythroid lineage; hypogranulation and hyposegmentation in granulocytic precursors, with an increase in myeloblasts; and megakaryocytes showing reduced numbers of or disorganized nuclei. Megaloblastic nuclei and defective hemoglobinization in the erythroid lineage are common. Prognosis strongly correlates with the proportion of marrow blasts, which should be enumerated manually on the marrow smear and by flow cytometry of the aspirate. Flow cytometry can also reveal characteristically aberrant hematopoietic differentiation. Cytogenetics and fluorescent in situ hybridization can identify chromosomal abnormalities.

■ DIFFERENTIAL DIAGNOSIS

Deficiencies of vitamin B_{12} or folate should be excluded by appropriate blood tests; vitamin B_6 deficiency can be assessed by a therapeutic trial of pyridoxine if the bone marrow shows ringed sideroblasts. Copper deficiency can lead to cytopenias and dysplastic marrows of varying cellularity. Transient marrow dysplasia can be observed in acute viral infections, drug reactions, or chemical toxicity. Myeloid and erythroid precursor cell vacuolation is characteristic of VEXAS, though vacuolization can also occur in copper deficiency and zinc or alcohol toxicity. More difficult are the distinctions between hypocellular MDS and aplasia or between MDS with excess blasts and acute leukemia: the WHO considers 20% blasts in the marrow as the criterion that separates AML from MDS. In young patients, underlying, predisposing genetic diseases should be considered and appropriate genomic testing performed (see above).

■ PROGNOSIS

The median survival varies greatly from years for patients with 5q− or sideroblastic anemia to a few months in those with high-risk MDS with excess blasts. The International Prognostic Scoring System (IPSS), revised in 2012 to the IPSS-R **(Table 107-7)**, assists in making predictions. In 2022, the IPSS-M, incorporating molecular data, is now emerging as the current standard tool for risk assessment. Even "lower-risk" MDS has significant morbidity and mortality. Prognostic scoring systems can separate those with intermediate-1 risk or higher who have relatively poor prognoses. Prognostic systems have been developed based on survival from diagnosis, but prognosis changes over time, and hazard ratios for survival and leukemic transformation converge over time among risk categories, consistent with dynamic changes in clonal architecture.

Most patients die as a result of complications of pancytopenia and not due to leukemic transformation; perhaps one-third succumb to diseases unrelated to their MDS. Precipitous worsening of pancytopenia, acquisition of new chromosomal abnormalities on serial cytogenetic determination, increased blasts, and marrow fibrosis are all poor prognostic indicators. The outlook in therapy-related MDS, regardless of type, is extremely poor, and most patients progress within a few months to refractory AML.

TABLE 107-7 Revised International Prognostic Scoring System (IPSS-R)

1. Marrow blast categories
   ≤2%, >2%–<5%, 5–10%, >10–30%
2. Refined cytogenetic abnormalities and risk groups
   16 (vs 6) specific abnormalities, 5 (vs 3) subgroups[a]
3. Evaluation of depth of cytopenias[b]
   Clinically and statistically relevant cut points used
4. Inclusion of differentiating features
   Age, performance status, serum ferritin, LDH, β_2-microglobulin
5. Prognostic model with 5 (vs 4) risk categories
   Improved predictive power

[a]Good, normal, −Y, del(5q), del (20q); poor, complex (≥3 abnormalities) or chromosome 7 abnormalities; intermediate, all other abnormalities. [b]Cytopenias at baseline, cut points: hemoglobin <80, 80–<100, or ≥100 g/L; platelet count <50, 50–100, or ≥100,000/μL, and absolute neutrophil count <800 versus ≥800/μL.

Abbreviation: LDH, lactate dehydrogenase.

TREATMENT

Myelodysplasia

Historically, therapy of MDS has been unsatisfactory, but several drugs may not only improve blood counts but also delay onset of leukemia and improve survival. The choice of therapy for an individual patient, administration of treatment, and management of toxicities are complicated and require hematologic expertise.

Only hematopoietic stem cell transplantation offers cure of MDS. The survival rate in selected patient cohorts is ~50% at 3 years but improving. Results using unrelated matched donors are similar to those with siblings, and data are emerging on the use of alternative donor transplant in MDS. Older patients can now be successfully transplanted using reduced intensive conditioning (RIC) regimens, but these regimens have a higher risk of relapse. Treatment-related mortality and morbidity increase with recipient age. The transplant conundrum is that the high-risk patient by prognostic score, for whom the procedure is most obviously indicated, has a high probability of a poor outcome from transplant-related mortality or disease relapse, whereas the low-risk patient, who is more likely to tolerate transplant, also may do well for years with less aggressive therapies. In reality, only a small proportion of MDS patients undergo transplantation.

MDS has been regarded as particularly refractory to cytotoxic chemotherapy regimens, and as in AML in the older adult, drug toxicity is frequent and often fatal, and remissions, if achieved, are brief. Low doses of cytotoxic drugs have been administered for their "differentiation" potential, and from this experience, drug therapies have emerged based on pyrimidine analogues. These drugs are classified as epigenetic modulators, believed to act through a demethylating mechanism to alter gene regulation and allow differentiation to mature blood cells from the abnormal MDS stem cell. The hypomethylating agents (HMA) azacitidine and decitabine are frequently used in MDS clinics. Azacitidine improves blood counts and survival in MDS, compared to best supportive care. Azacitidine is usually administered subcutaneously, daily for 7 days, at 4-week intervals, for at least four cycles before assessing for response. Overall, improved blood counts with a decrease in transfusion requirements occurred in ~50% of patients in published trials. Response is dependent on continued drug administration, and most patients eventually become refractory and experience recurrent cytopenias or progression to AML. Decitabine is closely related to azacitidine; 30–50% of patients show responses in blood counts, with a duration of response of almost a year. Decitabine is usually administered by continuous intravenous infusion in regimens of varying doses and durations of 3–10 days in repeating cycles. The major toxicity of azacitidine and decitabine is myelosuppression, leading to worsening blood counts. Hypomethylating agents are frequently used in the high-risk patient who is not a candidate for stem cell transplant. In the lower risk patient, they are also effective, but alternative therapies should first be considered. Decitabine-cedazuridine, an oral HMA, is FDA approved in patients with high-risk MDS.

Lenalidomide, a thalidomide derivative with a favorable toxicity profile, is particularly effective in reversing anemia in MDS patients with 5q− syndrome; not only do a high proportion of these patients become transfusion independent with normal or near-normal hemoglobin levels, but their cytogenetics also become normal. The drug has many biologic activities, and it is unclear which is critical for clinical efficacy. Lenalidomide is administered orally. Most patients will improve within 3 months of initiating therapy. Toxicities include myelosuppression (worsening thrombocytopenia and neutropenia, necessitating blood count monitoring) and an increased risk of deep vein thrombosis and pulmonary embolism.

Immunosuppression also may produce sustained independence from transfusion and improve survival. ATG, cyclosporine, and the anti-CD52 monoclonal antibody alemtuzumab are especially effective in younger MDS patients (<60 years old) with more favorable IPSS. In a consortium retrospective review, about 50% of patients with mainly refractory anemia responded to ATG, usually combined with cyclosporine, particularly patients with hypocellular marrow.

HGFs can improve blood counts but, as in most other marrow failure states, have been most beneficial to patients with the least severe pancytopenia. EPO alone or in combination with G-CSF can improve hemoglobin levels, particularly in those with low serum EPO levels who have no or a modest need for transfusions. Survival is improved by EPO and its amelioration of anemia. G-CSF treatment alone failed to improve survival in a controlled trial. Thrombopoietin mimetics appear to improve platelet counts in some patients, but TPO agonists may increase the rate of leukemic progression in high-risk MDS. No clear evidence suggests that they increase leukemic transformation in low-risk MDS.

Luspatercept, which affects transforming growth factor β–mediated suppression of erythropoiesis, has been approved by the FDA for anemia in low-risk MDS, particularly those with *SF3B1* mutations. The FDA has approved ivosidenib, an *IDH1* inhibitor, for MDS. Venetoclax, a BCL2 inhibitor, is FDA approved for patients with AML and has been used in combination with HMA in high-risk MDS; however, it remains investigational.

Promising novel agents in trials include targeted therapies (i.e., TP53, splicing factor mutations), inflammation pathway inhibitors (including inflammasome and interleukin 1 receptor–associated kinase), and imetelstat, a telomerase inhibitor. Likely, HMA alone will be replaced with HMA combination therapies.

The same principles of supportive care described for aplastic anemia apply to MDS. Many patients will be anemic for years. RBC transfusion support should be accompanied by iron chelation to prevent secondary hemochromatosis.

MYELOPHTHISIC ANEMIAS

Fibrosis of the bone marrow (see **Fig. 65-19**), usually accompanied by a characteristic blood smear picture called *leukoerythroblastosis*, can occur as a primary hematologic disease, called *myelofibrosis* or *myeloid metaplasia* **(Chap. 108)**, and as a secondary process, called *myelophthisis*. Myelophthisis, or secondary myelofibrosis, is reactive. Fibrosis can be a response to invading tumor cells, usually an epithelial cancer of breast, lung, or prostate origin or neuroblastoma. Marrow fibrosis may occur with infection of mycobacteria (both *Mycobacterium tuberculosis* and *Mycobacterium avium*), fungi, or HIV and in sarcoidosis. Intracellular lipid deposition in Gaucher disease and obliteration of the marrow space related to absence of osteoclast remodeling in congenital osteopetrosis also can produce fibrosis. Secondary myelofibrosis is a late consequence of radiation therapy or treatment with radiomimetic drugs. Usually the infectious or malignant underlying processes are obvious. Marrow fibrosis can also be a feature of a variety of hematologic syndromes, especially chronic myeloid leukemia, multiple myeloma, lymphomas, myeloma, and hairy cell leukemia.

The pathophysiology has three distinct features: proliferation of fibroblasts in the marrow space (myelofibrosis); the extension of hematopoiesis into the long bones and into extramedullary sites, usually the spleen, liver, and lymph nodes (myeloid metaplasia); and ineffective erythropoiesis. The etiology of the fibrosis is unknown but most likely involves dysregulated production of growth factors: platelet-derived growth factor and transforming growth factor β have been implicated. Abnormal regulation of other hematopoietins would lead to localization of blood-producing cells in nonhematopoietic tissues and uncoupling of the usually balanced processes of stem cell proliferation and differentiation. Myelofibrosis is remarkable for pancytopenia despite very large numbers of circulating hematopoietic progenitor cells.

Anemia is dominant in secondary myelofibrosis, usually normocytic and normochromic. The diagnosis is suggested by the characteristic leukoerythroblastic smear. Erythrocyte morphology is highly abnormal, with circulating nucleated RBCs, teardrops, and shape distortions. WBC numbers are often elevated, sometimes mimicking a leukemoid reaction, with circulating myelocytes, promyelocytes, and myeloblasts. Platelets may be abundant and are often of giant size. Inability to aspirate the bone marrow, the characteristic "dry tap," can allow a presumptive diagnosis in the appropriate setting before the biopsy is decalcified.

The course of secondary myelofibrosis is determined by its etiology, usually a metastatic tumor or an advanced hematologic malignancy. Treatable causes must be excluded, especially tuberculosis and fungus. Transfusion support can relieve symptoms.

■ FURTHER READING

Arber DA et al: International Consensus Classification of myeloid neoplasms and acute leukemias: Integrating morphologic, clinical, and genomic data. Blood 140:1200, 2022.

Bernard E et al: Molecular International Prognostic Scoring System for myelodysplastic syndromes. NEJM Evid 1:EVIDoa2200008, 2022.

Cazzola M: Myelodysplastic syndromes. N Engl J Med 383:1358, 2020.

DeFilipp Z et al: Hematopoietic cell transplantation in the management of myelodysplastic syndrome: An evidence-based review from the American Society for Transplantation and Cellular Therapy Committee on Practice Guidelines. Transplant Cell Ther 29:71, 2022.

Gurnari C, Maciejewski JP: How I manage acquired pure red cell aplasia in adults. Blood 15;137:2001, 2021.

Hellstrom-Lindberg ES et al: Clinical decision-making and treatment of myelodysplastic syndromes. Blood 142:2268, 2023.

Khoury JD et al: The 5th edition of the World Health Organization classification of haematolymphoid tumours: Myeloid and histiocytic/dendritic neoplasms. Leukemia 36:1703, 2022.

Mustjoki S, Young NS: Somatic mutations in "benign" disease. N Engl J Med 384:2039, 2021.

Townsley DM et al: Eltrombopag added to standard immunosuppression for aplastic anemia. N Engl J Med 376:1540, 2017.

Young NS: Aplastic anemia. N Engl J Med 379:1643, 2018.

108

Polycythemia Vera and Other Myeloproliferative Neoplasms

Jerry L. Spivak

The World Health Organization (WHO) classification of the chronic myeloproliferative neoplasms (MPNs) includes eight disorders, some of which are rare or poorly characterized **(Table 108-1)** but all of which share an origin in a hematopoietic stem cell; overproduction of one or more of the formed elements of the blood without significant dysplasia; and a predilection to extramedullary hematopoiesis, myelofibrosis, and transformation at varying rates to acute leukemia. Within this broad

TABLE 108-1 World Health Organization Classification of Chronic Myeloproliferative Neoplasms

Chronic myeloid leukemia, *BCR-ABL*–positive
Chronic neutrophilic leukemia
Chronic eosinophilic leukemia, not otherwise specified
Polycythemia vera
Primary myelofibrosis
Essential thrombocytosis
Mastocytosis
Myeloproliferative neoplasms, unclassifiable

classification, however, significant phenotypic heterogeneity exists. Some diseases such as chronic myelogenous leukemia (CML), chronic neutrophilic leukemia (CNL), and chronic eosinophilic leukemia (CEL) express primarily a myeloid phenotype, whereas in other diseases, such as polycythemia vera (PV), primary myelofibrosis (PMF), and essential thrombocytosis (ET), erythroid or megakaryocytic hyperplasia predominates. The latter three disorders, in contrast to the former three, also appear capable of transforming into each other.

Such phenotypic heterogeneity has a genetic basis; CML is the consequence of the balanced translocation between chromosomes 9 and 22 (t[9;22][q34;11]); CNL has been associated with a mutation of *CSF3R* and a t(15;19) translocation; and CEL occurs with a deletion or balanced translocations involving the *PDGFRα* usually with the *FIP1L1*, *PDGFRβ*, *FGFR1*, and *PCM1-JAK2* genes. By contrast, PV, PMF, and ET are characterized by driver mutations that directly or indirectly constitutively activate JAK2, a tyrosine kinase essential for the function of the erythropoietin and thrombopoietin receptors and also utilized by the granulocyte colony-stimulating factor receptor. This important distinction is reflected in the natural histories of CML, CNL, and CEL, which are usually measured in years, with a high rate of leukemic transformation. The natural histories of PV, PMF, and ET, by contrast, are usually measured in decades, and transformation to acute leukemia is uncommon in the absence of chemotherapy. This chapter focuses only on PV, PMF, and ET because their clinical features and driver mutation overlap are substantial, although their disease duration and clinical manifestations vary.

The other chronic MPNs will be discussed in **Chaps. 110** and **115**.

POLYCYTHEMIA VERA

PV is a clonal hematopoietic stem cell disorder in which phenotypically normal red cells, granulocytes, and platelets accumulate in the absence of a recognizable physiologic stimulus. The most common of the MPNs, PV occurs in 2.5 per 100,000 persons, sparing no adult age group and increasing with age to rates >10/100,000. Familial transmission is infrequent, and women under age 50 predominate among sporadic cases.

ETIOLOGY

Nonrandom chromosome abnormalities such as deletion 20q and deletion 13q or trisomy 9 occur in up to 30% of untreated PV patients, but unlike CML, no consistent cytogenetic abnormality has been associated with the disorder. However, a mutation in the autoinhibitory pseudokinase domain of the tyrosine kinase JAK2 that replaces valine with phenylalanine (V617F), causing constitutive kinase activation, has a central role in PV pathogenesis.

JAK2 is a member of an evolutionarily well-conserved, nonreceptor tyrosine kinase family and serves as the cognate tyrosine kinase for the erythropoietin and thrombopoietin receptors. It also functions as an obligate chaperone for these receptors in the Golgi apparatus and is responsible for their cell-surface expression. The conformational change induced in the erythropoietin and thrombopoietin receptors following binding to their respective cognate ligands, erythropoietin or thrombopoietin, leads to JAK2 autophosphorylation, receptor phosphorylation, and phosphorylation of proteins involved in cell proliferation, differentiation, and resistance to apoptosis. Transgenic animals lacking *JAK2* die as embryos from severe anemia. Constitutive activation of JAK2, on the other hand, explains the erythropoietin hypersensitivity, erythropoietin-independent erythroid colony formation, rapid terminal differentiation, increased Bcl-X_L expression, and apoptosis resistance in the absence of erythropoietin that characterize the in vitro behavior of PV erythroid progenitor cells.

More than 95% of PV patients express this mutation, as do ~50% of PMF and ET patients. Importantly, the *JAK2* gene is located on the short arm of chromosome 9, and loss of heterozygosity on chromosome 9p involving the segment containing the *JAK2* locus over time due to mitotic recombination (uniparental disomy) is the most common cytogenetic abnormality in PV. Loss of heterozygosity in this region leads to homozygosity for *JAK2* V617F and occurs in ~60% of PV patients and to a lesser extent in PMF but is rare in ET. Most PV patients who do not express *JAK2* V617F express a mutation in exon 12 of the gene and are not clinically different from those who do, with the exception of a higher frequency of isolated erythrocytosis, nor do *JAK2* V617F heterozygotes differ clinically from homozygotes. Importantly, the predisposition to acquire *JAK2* mutations appears to be associated with a specific *JAK2* gene haplotype, GGCC. *JAK2* V617F is the basis for many of the phenotypic and biochemical characteristics of PV such as increased blood cell production and increased inflammatory cytokine production; however, it cannot solely account for the entire PV phenotype and is probably not the initiating lesion in any of the MPNs. First, PV patients with the same phenotype and documented clonal disease can have mutations in *LNK*, a JAK2 inhibitor, or rarely, calreticulin (*CALR*), an ER chaperone, since MPN driver mutations are not mutually exclusive. Second, ET and PMF patients have the same mutations but different clinical phenotypes. Third, familial PV can occur without the mutation, even when other members of the same family express it. Finally, in some *JAK2* V617F–positive PV or ET patients, acute leukemia can occur in a *JAK2* V617F–negative progenitor cell, suggesting the presence of an ancestral precursor cell.

CLINICAL FEATURES

Although PV is a panmyelopathy, isolated thrombocytosis, leukocytosis, or splenomegaly can be its presenting manifestation, but most often, the disorder is first recognized by the incidental discovery of a high hemoglobin, hematocrit, or red cell count. With the exception of aquagenic pruritus or erythromelalgia, no symptoms distinguish PV from other causes of erythrocytosis.

Uncontrolled erythrocytosis causes hyperviscosity, leading to neurologic symptoms such as vertigo, tinnitus, headache, visual disturbances, and transient ischemic attacks (TIAs). Systolic hypertension is also a feature of the red cell mass elevation. In some patients, venous or arterial thrombosis may be the presenting manifestation of PV. Any vessel can be affected, but cerebral, cardiac, and mesenteric vessels are most commonly involved. Hepatic venous thrombosis (Budd-Chiari syndrome) is particularly common in young women and may be catastrophic if sudden and complete obstruction of the hepatic vein occurs; portal vein thrombosis is more common in male PV patients. Indeed, PV should be suspected in any woman who develops hepatic vein thrombosis, since this is the only type of thrombosis associated with *JAK2* V617F expression. Digital ischemia, easy bruising, epistaxis, acid-peptic disease, or gastrointestinal hemorrhage may occur due to vascular stasis or extreme thrombocytosis (>900,000/mL). In the latter instance, absorption and proteolysis of high-molecular-weight von Willebrand multimers by the large platelet mass cause acquired von Willebrand's disease. Erythema, burning, and pain in the extremities, a symptom complex known as erythromelalgia, is another complication of thrombocytosis in PV due to increased platelet stickiness. Given the large turnover of hematopoietic cells, hyperuricemia with secondary gout, uric acid stones, and symptoms due to hypermetabolism can also complicate the disorder.

DIAGNOSIS

When PV presents with erythrocytosis in combination with leukocytosis, thrombocytosis, or splenomegaly or any combination of these, the diagnosis is apparent. However, when patients present with an elevated hemoglobin, hematocrit, and red cell count alone, the diagnostic evaluation is more complex because of the many diagnostic possibilities **(Table 108-2)**. Furthermore, unless the hemoglobin level is ≥20 g/dL (hematocrit ≥60%), it is not possible to distinguish true erythrocytosis from disorders causing plasma volume contraction. This is because uniquely in PV, in contrast to other causes of true erythrocytosis, there is expansion of the plasma volume, which can mask the elevated red cell mass, particularly in women; thus, red cell mass and plasma volume determinations are necessary to establish the presence of an absolute erythrocytosis and distinguish this from relative erythrocytosis due to a reduction in plasma volume alone (also known as *stress* or *spurious erythrocytosis* or *Gaisböck's syndrome*). **Figure 66-18** illustrates a diagnostic algorithm for the evaluation of suspected erythrocytosis. While an assay for *JAK2* or rarely *LNK* mutations in the presence of a normal arterial oxygen saturation appears to provide an alternative

TABLE 108-2 Causes of Erythrocytosis

Relative Erythrocytosis	
Hemoconcentration secondary to dehydration, diuretics, ethanol abuse, androgens, or tobacco abuse	
Absolute Erythrocytosis	
Hypoxia	**Tumors**
Carbon monoxide intoxication	Hypernephroma
High-oxygen-affinity hemoglobin	Hepatoma
High altitude	Cerebellar hemangioblastoma
Pulmonary disease	Uterine myoma
Right-to-left cardiac or vascular shunts	Adrenal tumors
Sleep apnea syndrome	Meningioma
Hepatopulmonary syndrome	Pheochromocytoma
Renal Disease	**Drugs**
Renal artery stenosis	Androgens
	SGLT2 inhibitors
Focal sclerosing or membranous glomerulonephritis	Recombinant erythropoietin
Postrenal transplantation	**Familial (with normal hemoglobin function)**
Renal cysts	Erythropoietin receptor mutations
Bartter's syndrome	*VHL* mutations (Chuvash polycythemia)
	2,3-*BPG* mutation
	PHD2 (*EGLN1*) and *HIF2α* (*EPAS1*) mutations
	LNK mutations
	Polycythemia vera

Abbreviations: 2,3-BPG, 2,3-bisphosphoglycerate; VHL, von Hippel-Lindau.

diagnostic approach to isolated erythrocytosis since red cell mass and plasma volume determinations are not usually available, isolated erythrocytosis is uncommon as an initial manifestation of PV, and not everyone expressing a low *JAK*2 V617F quantitative mutation allele burden (variant allele frequency [VAF] ≤5%) actually has a blood disease. In addition, a normal serum erythropoietin level does not exclude the presence of PV, but an elevated erythropoietin level is most consistent with a secondary cause for the erythrocytosis.

Other laboratory studies that may aid in diagnosis include the red cell count, mean corpuscular volume, and red cell distribution width (RDW), particularly when the hematocrit or hemoglobin levels are <60% or 20 g/dL, respectively. Only three situations cause microcytic erythrocytosis: β-thalassemia trait, hypoxic erythrocytosis, and PV. With β-thalassemia trait, the RDW is usually normal, whereas with hypoxic erythrocytosis or PV, the RDW may be elevated due to associated iron deficiency. Today, however, the quantitative assay for *JAK2* V617F or *JAK*2 exon 12 mutations using next-generation sequencing technology has superseded the other surrogate tests for establishing the diagnosis of PV.

A bone marrow aspirate and biopsy provide no specific diagnostic information because these may be normal or indistinguishable from ET or PMF. Similarly, no specific cytogenetic abnormality is associated with the disease, and the absence of a cytogenetic marker does not exclude the diagnosis.

COMPLICATIONS

Many of the clinical complications of PV relate directly to the increase in blood viscosity associated with red cell mass elevation and indirectly to the increased turnover of red cells, leukocytes, and platelets with the attendant increase in uric acid and inflammatory cytokine production. The latter also appears to be responsible for some of the constitutional symptoms in PV. Peptic ulcer disease can also be due to *Helicobacter pylori* infection, the incidence of which is increased in PV, while the pruritus associated with this disorder may be a consequence of mast cell activation by *JAK2* V617F. A sudden increase in spleen size can be associated with painful splenic infarction. Myelofibrosis appears to be part of the natural history of the disease but is a reactive, reversible process that does not itself impede hematopoiesis and by itself has no prognostic significance in PV. In ~15% of patients, however, myelofibrosis is associated with hematopoietic stem cell failure, manifested by substantial extramedullary hematopoiesis in the liver and spleen and transfusion-dependent anemia. Organomegaly can cause significant mechanical discomfort, portal hypertension, and progressive cachexia.

Although the incidence of acute myeloid leukemia is increased in PV, the incidence of acute leukemia in patients not exposed to chemotherapy or radiation therapy is very low. Interestingly, chemotherapy, including hydroxyurea, has been associated with acute leukemia in *JAK2* V617F–negative hematopoietic stem cells (HSCs) in some PV patients. *Erythromelalgia* is a curious syndrome of unknown etiology associated with thrombocytosis, primarily involving the lower extremities and usually manifested by erythema, warmth, and pain of the affected appendage and occasionally digital infarction. It occurs with a variable frequency and is usually responsive to salicylates. Some of the central nervous system symptoms observed in patients with PV, such as ocular migraine, appear to represent a variant of erythromelalgia.

Left uncontrolled, erythrocytosis can lead to thrombosis involving vital organs such as the liver, heart, brain, or lungs. Patients with massive splenomegaly are particularly prone to thrombotic events because the associated increase in plasma volume masks the true extent of the red cell mass elevation measured by the hematocrit or hemoglobin level. A "normal" hematocrit or hemoglobin level in a PV patient with massive splenomegaly should be considered indicative of an elevated red cell mass until proven otherwise.

TREATMENT

Polycythemia Vera

PV is generally an indolent disorder, the clinical course of which is measured in decades, and its management should reflect its tempo. Thrombosis due to erythrocytosis is the most significant complication and often the presenting manifestation; maintenance of the hemoglobin level at ≤140 g/L (14 g/dL; hematocrit <45%) in men and ≤120 g/L (12 g/dL; hematocrit <42%) in women is mandatory to avoid thrombotic complications. Phlebotomy serves initially to reduce hyperviscosity by reducing the red cell mass to normal while further expanding the plasma volume. Periodic phlebotomies thereafter serve to maintain the red cell mass within the normal range and induce a state of iron deficiency that prevents accelerated reexpansion of the red cell mass. In most PV patients, once an iron-deficient state is achieved, phlebotomy is usually only required at 3-month intervals. Neither phlebotomy nor iron deficiency increases the platelet count relative to the effect of the disease itself, and neither thrombocytosis nor leukocytosis is correlated with thrombosis in PV, in contrast to the strong correlation between erythrocytosis and thrombosis. The use of salicylates to prevent thrombosis in PV patients is potentially harmful not only if the red cell mass is not controlled by phlebotomy, but also due to an increased incidence of bleeding, particularly in patients over age 60.

Anticoagulation is indicated when a thrombosis has occurred, and the newer oral anticoagulants may be preferable to a vitamin K antagonist since they do not require monitoring. Asymptomatic hyperuricemia (<10 mg/dL) requires no therapy, but allopurinol should be administered to avoid further elevation of the uric acid when chemotherapy is used to reduce splenomegaly or leukocytosis or to treat pruritus. Generalized pruritus intractable to antihistamines or antidepressants such as doxepin can be a major problem in PV; the JAK1/2 inhibitor ruxolitinib, pegylated interferon α (IFN-α), psoralens with ultraviolet light in the A range (PUVA) therapy, and hydroxyurea are other methods of palliation. Asymptomatic thrombocytosis requires no therapy unless the platelet count is sufficiently high to cause bleeding due to acquired von Willebrand's disease, but bleeding in this situation is not usually spontaneous and is responsive to tranexamic acid or ε-aminocaproic acid. Symptomatic splenomegaly can be treated with either ruxolitinib or pegylated

IFN-α. Both ruxolitinib and pegylated IFN-α target the involved HSCs in PV and are not mutagenic; hydroxyurea does not target the involved HSCs in PV and is mutagenic. Furthermore, pegylated IFN-α allows only biweekly administration and produced complete hematologic and molecular remissions in ~20% of PV patients. Anagrelide, a phosphodiesterase inhibitor, can reduce the platelet count and, if tolerated, is preferable to hydroxyurea because it lacks marrow toxicity and is also protective against venous thrombosis, whereas hydroxyurea is not. However, chronic anagrelide therapy is cardiotoxic and nephrotoxic, particularly in older PV patients.

A reduction in platelet number may be necessary for the treatment of erythromelalgia or ocular migraine if salicylates are not effective or if the platelet count is sufficiently high to increase the risk of hemorrhage but only to the degree that symptoms are alleviated. Alkylating agents and radioactive sodium phosphate (^{32}P) are leukemogenic in PV, and their use should be avoided. If a cytotoxic agent must be used, hydroxyurea is preferred, but this drug does not prevent either thrombosis or myelofibrosis in PV, is itself leukemogenic, and should be considered as a short-duration therapy. Previously, PV patients with massive splenomegaly unresponsive to reduction by chemotherapy or IFN required splenectomy. However, with the introduction of the nonspecific JAK1/2 inhibitor ruxolitinib, it has been possible in the majority of patients with PV complicated by myelofibrosis and myeloid metaplasia to reduce spleen size while at the same time alleviating constitutional symptoms and pruritus due to cytokine release while reducing the phlebotomy requirement. However, in contrast to PMF, PV patients have a more chronic course; in contrast to other malignancies, PV patients have a low rate of mutation accumulation, and the acquisition of deleterious mutations such as *TP*53 mutations as detected by next-generation sequencing is usually associated with leukemic transformation. Since hydroxyurea antagonizes *TP53* and also causes del17p, leading to *TP53* haploinsufficiency, its chronic use should be constrained in PV.

Ruxolitinib has also been demonstrated in a phase 3 clinical trial to be effective in PV patients without myelofibrosis who are intolerant or refractory to hydroxyurea or best available supportive therapy. Three other JAK2 inhibitors, fedratinib, pacritinib, and momelotinib, have been approved for treatment of PV patients with myelofibrosis in whom ruxolitinib treatment failed or who were intolerant to the drug. In some PV patients with end-stage disease, pulmonary hypertension may develop due to fibrosis or extramedullary hematopoiesis. A role for bone marrow transplantation, either allogeneic or haploidentical, in PV has not been defined.

Most patients with PV can live long lives without functional impairment when their red cell mass is effectively managed with phlebotomy alone. Chemotherapy is never indicated to control the red cell mass in PV, but when venous access is an issue, ruxolitinib or pegylated IFN is the preferred therapy. Interestingly, hepcidin production is suppressed, but not absent in PV, and a hepcidin agonist has been shown to reduce phlebotomy requirements in PV.

PRIMARY MYELOFIBROSIS

Chronic PMF (other designations include *idiopathic myelofibrosis*, *agnogenic myeloid metaplasia*, or *myelofibrosis with myeloid metaplasia*) is a clonal HSC disorder associated with mutations in *JAK2, MPL*, or *CALR,* and characterized by marrow fibrosis, extramedullary hematopoiesis, variable suppression of hematopoiesis, and splenomegaly. PMF is the least common MPN, and establishing its diagnosis in the absence of a specific clonal marker is difficult because myelofibrosis and splenomegaly are also features of both PV and CML. Furthermore, myelofibrosis and splenomegaly also occur in a variety of benign and malignant disorders **(Table 108-3)**, many of which are amenable to specific therapies not effective in PMF. In contrast to the other MPNs and so-called acute or malignant myelofibrosis, which can occur at any age, PMF primarily afflicts men in their sixth decade or later.

TABLE 108-3 Disorders Causing Myelofibrosis

MALIGNANT	NONMALIGNANT
Acute leukemia (lymphocytic, myelogenous, megakaryocytic)	HIV infection
Chronic myeloid leukemia	Hyperparathyroidism
Hairy cell leukemia	Renal osteodystrophy
Hodgkin's disease	Systemic lupus erythematosus
Primary myelofibrosis	Tuberculosis
Lymphoma	Vitamin D deficiency
Multiple myeloma	Thorium dioxide exposure
Myelodysplasia	Gray platelet syndrome
Metastatic carcinoma	
Polycythemia vera	
Systemic mastocytosis	

ETIOLOGY

Nonrandom chromosome abnormalities such as 9p, 20q−, 13q−, trisomy 8 or 9, or partial trisomy 1q are common in PMF, but no cytogenetic abnormality specific to the disease has been identified. *JAK2* V617F is present in ~55% of PMF patients, and mutations in the thrombopoietin receptor, *MPL*, occur in ~4%. Most of the rest have mutations in the calreticulin gene (*CALR*) that alter the carboxy-terminal portion of the protein, permitting it to bind and activate MPL while presenting it at the cell surface. The degree of myelofibrosis and the extent of extramedullary hematopoiesis are not related. Fibrosis in this disorder is associated with overproduction of transforming growth factor β and tissue inhibitors of metalloproteinases and thrombopoietin, while osteosclerosis is associated with overproduction of osteoprotegerin, an osteoclast inhibitor. Marrow angiogenesis occurs due to increased production of vascular endothelial growth factor. Importantly, fibroblasts in PMF are polyclonal and not part of the neoplastic clone but can be induced by it to produce inflammatory cytokines.

CLINICAL FEATURES

No signs or symptoms are specific for PMF. Many patients are asymptomatic at presentation, and the disease is often detected by the discovery of splenic enlargement and/or abnormal blood counts during a routine examination. In contrast to its companion MPN, night sweats, fatigue, and weight loss are common presenting complaints. A blood smear will show the characteristic features of extramedullary hematopoiesis: teardrop-shaped red cells, nucleated red cells, myelocytes, and promyelocytes; myeloblasts may also be present **(Fig. 108-1)**. Anemia, usually mild initially, is common, whereas the leukocyte and platelet counts are either normal or increased, but either can be depressed. Mild hepatomegaly may accompany splenomegaly but is unusual in its

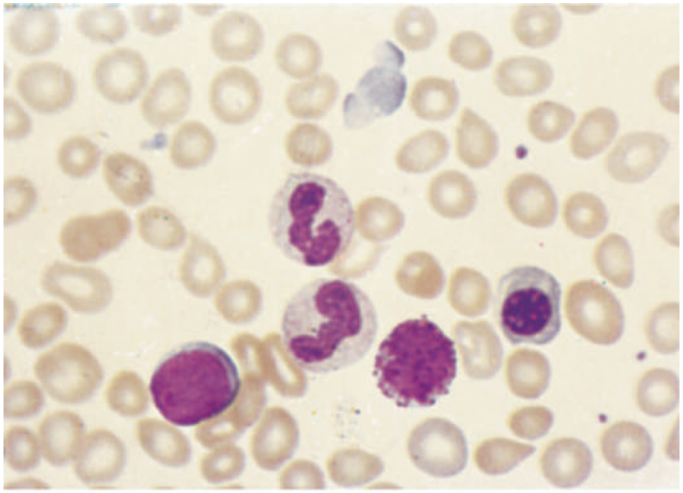

FIGURE 108-1 Teardrop-shaped red blood cells indicative of membrane damage from passage through the spleen, a nucleated red blood cell, and immature myeloid cells indicative of extramedullary hematopoiesis are noted. This peripheral blood smear is related to any cause of extramedullary hematopoiesis.

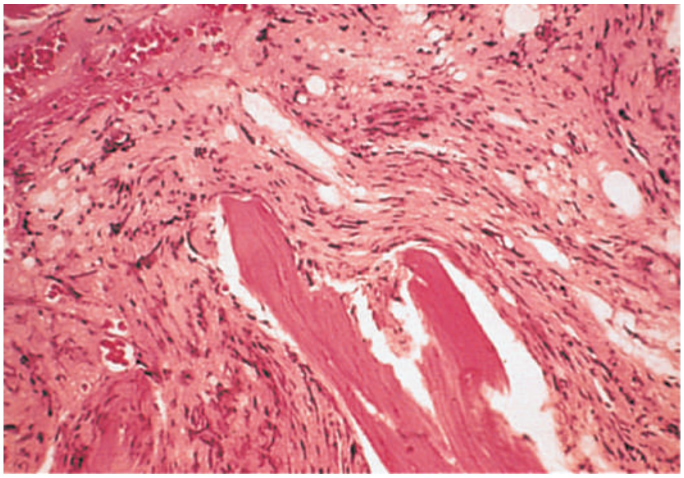

FIGURE 108-2 This marrow section shows the marrow cavity replaced by fibrous tissue composed of reticulin fibers and collagen. When this fibrosis is due to a primary hematologic process, it is called *myelofibrosis*. When fibrosis is secondary to a tumor or a granulomatous process, it is called *myelophthisis*.

absence; isolated lymphadenopathy should suggest another diagnosis. Both serum lactate dehydrogenase and alkaline phosphatase levels can be elevated. Marrow is usually inaspirable due to the myelofibrosis **(Fig. 108-2)**, and bone x-rays may reveal osteosclerosis. Exuberant extramedullary hematopoiesis can cause ascites; portal, pulmonary, or intracranial hypertension; intestinal or ureteral obstruction; pericardial tamponade; spinal cord compression; or skin nodules. Splenic enlargement can be sufficiently rapid to cause splenic infarction with fever and pleuritic chest pain. Hyperuricemia and secondary gout may ensue.

DIAGNOSIS

While the clinical picture described above is characteristic of PMF, all of these clinical features can be observed in PV or CML. Massive splenomegaly commonly masks erythrocytosis in PV, and reports of intraabdominal thrombosis in PMF most likely represent instances of unrecognized PV. In some PMF patients, erythrocytosis has developed during the course of the disease. Importantly, because many other disorders have features that overlap with PMF but respond to distinctly different therapies, the diagnosis of PMF is one of exclusion, which requires that the disorders listed in Table 108-3 be ruled out.

The presence of teardrop-shaped red cells, nucleated red cells, myelocytes, and promyelocytes establishes the presence of extramedullary hematopoiesis, while the presence of leukocytosis, thrombocytosis with large and bizarre platelets, and circulating myelocytes suggests the presence of an MPN as opposed to a secondary form of myelofibrosis (Table 108-3). Marrow is usually inaspirable due to increased marrow reticulin, but marrow biopsy will usually reveal a hypercellular marrow with trilineage hyperplasia and, in particular, increased numbers of megakaryocytes in clusters and with large, dysplastic nuclei. A small number of PMF patients with a low *JAK2* V617F mutation allele burden (≤25%) have a faster time to anemia and leukopenia and a shortened survival. However, there are no specific bone marrow morphologic abnormalities that distinguish PMF from the other MPNs. Splenomegaly due to extramedullary hematopoiesis may be sufficiently massive to cause portal hypertension and variceal formation. In some patients, exuberant extramedullary hematopoiesis dominates the clinical picture.

An intriguing feature of PMF is the occurrence of autoimmune abnormalities such as immune complexes, antinuclear antibodies, rheumatoid factor, or a positive Coombs' test. Whether these represent a host reaction to the disorder or are involved in its pathogenesis is unknown. Cytogenetic analysis of the blood is useful both to exclude CML and for prognostic purposes, because the development of complex karyotype abnormalities portends a poor prognosis in PMF. It is thought that impaired expression of the cytokine CXCL4 is responsible for the markedly increased number of circulating CD34+ cells in PMF (>15,000/μL) compared to PV patients, unless they too develop extramedullary hematopoiesis.

Importantly, ~55% of PMF patients, like patients with its companion MPNs, express the *JAK2* V617F mutation, often as homozygotes. Such patients are usually older and have higher hematocrits than patients with *MPL* (4%) or *CALR* (36%) mutations; PMF patients expressing an *MPL* mutation tend to be more anemic and have lower leukocyte counts than *JAK2* V617F–positive patients. Somatic mutations (due to deletions [type 1] or insertions [type 2]) in exon 9 of *CALR* have been found in a majority of patients with PMF who lack mutations in either *JAK2* or *MPL*. In some studies, type 1 mutations, the most common *CALR* mutation in PMF, had a survival advantage compared to *JAK2* or *MPL* mutations but not with respect to leukemic transformation. PMF patients who lack a known MPN driver mutation (triple-negative) appear to have the worst prognosis.

TABLE 108-4 Three Current Scoring Systems for Estimating Prognosis in PMF Patients

RISK FACTOR	IPSS (2009)[a]	DIPSS (2010)[b]	DIPSS PLUS (2011)[c]
Anemia (<10 g/dL)	X	X	X
Leukocytosis (>25,000/μL)	X	X	X
Peripheral blood blasts (≥1%)	X	X	X
Constitutional symptoms	X	X	X
Age (>65 years)	X	X	X
Unfavorable karyotype			X
Platelet count (<100,000/μL)			X
Transfusion dependence			X

[a]Blood 113:2895, 2009. [b]Blood 115:1703, 2010. [c]J Clin Oncol 29:392, 2011.

Note: The Dynamic International Prognostic Scoring System (DIPSS) was developed to determine if the International Prognostic Scoring System (IPSS) risk factors identified as important for survival at the time of primary myelofibrosis (PMF) diagnosis could also be used for risk stratification following their acquisition during the course of the disease. One point is assigned to each risk factor for IPSS scoring. For DIPSS, the same is true, but anemia is assigned 2 points. The DIPSS Plus scoring system represents recognition that the addition of unfavorable karyotype, thrombocytopenia, and transfusion dependence improved the DIPSS risk stratification system for which additional points are assigned (Table 108-5). More recent studies suggest that mutational analysis of the *ASXL1, EZH2, SRSF2,* and *IDH1/2* genes further improves risk stratification for survival and leukemic transformation (Leukemia 27:1861, 2013), as can cytogenetic abnormalities (Leukemia 32:1631, 2018). These prognostic scoring systems are not accurate for risk assessment in polycythemia vera or essential thrombocytosis patients who have developed myelofibrosis (Haematologica 99:e55, 2014).

COMPLICATIONS

Survival in PMF varies according to specific risk factors at diagnosis **(Tables 108-4 and 108-5)** but is much shorter than in PV and ET patients. The natural history of PMF is one of increasing marrow failure with transfusion-dependent anemia and increasing organomegaly due to extramedullary hematopoiesis. As with CML, PMF can evolve from a chronic to an accelerated phase with constitutional symptoms and increasing marrow failure. About 10% of patients spontaneously transform to an aggressive form of acute leukemia for which therapy is usually ineffective. Additional important prognostic factors for disease acceleration during the course of PMF include the presence of complex

TABLE 108-5 IPSS and DIPSS Risk Stratification Systems

	NUMBER OF RISK FACTORS		
RISK CATEGORIES[a]	IPSS	DIPSS	DIPSS PLUS
Low	0	0	0
Intermediate-1	1	1–2	1
Intermediate-2	2	3–4	2–3
High	≥3	>4	4–6

[a]The corresponding survival curves for each risk category can be found in the references cited in the footnotes of Table 108-4.

Abbreviations: DIPSS, Dynamic International Prognostic Scoring System; IPSS, International Prognostic Scoring System.

cytogenetic abnormalities, thrombocytopenia, and transfusion-dependent anemia. Mutations in the *ASXL1*, *EZH2*, *SRSF2*, and *IDH1/2* genes have been identified as risk factors for early death or transformation to acute leukemia, as have complex cytogenetic abnormalities, and have proved to be more useful for PMF risk assessment than the clinical scoring systems.

TREATMENT

Primary Myelofibrosis

No specific therapy exists for PMF. The causes for anemia are multifarious and include ineffective erythropoiesis uncompensated by splenic extramedullary hematopoiesis, hemodilution due to splenomegaly, splenic sequestration, blood loss secondary to thrombocytopenia or portal hypertension, folic acid deficiency, systemic inflammation, and autoimmune hemolysis. Neither recombinant erythropoietin nor androgens such as danazol have proven to be consistently effective as therapy for anemia. Erythropoietin may worsen splenomegaly and will be ineffective if the serum erythropoietin level is >125 mU/L. Given the inflammatory milieu that characterizes PMF, glucocorticoids can ameliorate anemia as well as constitutional symptoms such as fever, chills, night sweats, anorexia, and weight loss, and combining these with low-dose thalidomide has proved effective as well. Thrombocytopenia can be due to impaired marrow function, splenic sequestration, or autoimmune destruction and may also respond to low-dose thalidomide and prednisone.

Splenomegaly is by far the most distressing and intractable problem for PMF patients, causing abdominal pain, portal hypertension, easy satiety, and cachexia, whereas surgical removal of a massive spleen is associated with significant postoperative complications including mesenteric venous thrombosis, hemorrhage, rebound leukocytosis and thrombocytosis, and hepatic extramedullary hematopoiesis with no amelioration of either anemia or thrombocytopenia when present. For unexplained reasons, splenectomy also increases the risk of blastic transformation.

Splenic irradiation is, at best, temporarily palliative and associated with a significant risk of neutropenia, infection, and subsequent operative hemorrhage if splenectomy is attempted. Allopurinol can control significant hyperuricemia, and bone pain can be alleviated by local irradiation. Pegylated IFN-α can ameliorate fibrosis in early PMF, but in advanced disease, it may exacerbate bone marrow failure. The JAK1/2 inhibitor ruxolitinib has proved effective in reducing splenomegaly and alleviating constitutional symptoms in a majority of advanced PMF patients while possibly prolonging survival, but in some patients, ruxolitinib is associated with *RAS* mutations. Although anemia and thrombocytopenia are its major side effects, these are dose-dependent, and with time, anemia stabilizes, and thrombocytopenia may improve. Fedratinib, a new tyrosine kinase inhibitor with anti-FLT3 activity, has proved useful in patients with disease refractory to ruxolitinib. Two other JAK2 inhibitors have been approved for PMF therapy, pacritinib, which is useful when thrombocytopenia is present, and momelotinib, which may be useful in improving red cell production.

In some patients, hypomethylating agents such as azacytidine or decitabine in combination with high-dose ruxolitinib have been used to control the disease or prepare patients for bone marrow transplantation. Transformation to acute leukemia in PMF, like PV or ET, is usually refractory to treatment.

Allogeneic bone marrow transplantation is the only curative treatment for PMF and should be considered in younger patients and older patients with high-risk disease; nonmyeloablative conditioning regimens permit hematopoietic cell transplantation to be extended to older individuals.

ESSENTIAL THROMBOCYTOSIS

ET (other designations include *essential thrombocythemia*, *idiopathic thrombocytosis*, *primary thrombocytosis*, and *hemorrhagic thrombocythemia*) is a clonal hematopoietic stem cell disorder associated with mutations in *JAK2* (V617F), *MPL*, or *CALR* and manifested clinically by overproduction of platelets without a definable cause. ET has an incidence of 1–2/100,000 and a distinct female predominance. Canonical MPN driver mutations distinguish 90% of ET patients from the more common nonclonal, reactive forms of thrombocytosis **(Table 108-6)**; mutation-negative ET patients may have either uncommon *MPL* mutations, *JAK2* V617F expression limited to the platelets, or a hereditary form of thrombocytosis. Once considered a disease of the elderly and responsible for significant morbidity due to hemorrhage or thrombosis, it is now clear that ET can occur at any age in adults and often without symptoms or disturbances of hemostasis. There is an unexplained female predominance in contrast to PMF or the reactive forms of thrombocytosis where no sex difference exists. Because no specific clonal marker is available, clinical and laboratory criteria have been proposed to distinguish ET from other MPNs, which may also present initially with isolated thrombocytosis but have differing prognoses and therapies (Table 108-6). These criteria are useful in identifying disorders such as CML, PV, PMF, or myelodysplasia, which can masquerade as ET. Furthermore, as with "idiopathic" erythrocytosis, nonclonal benign forms of thrombocytosis exist (e.g., hereditary overproduction of thrombopoietin and those with noncanonical *JAK2* driver mutations) that are not widely recognized because we currently lack diagnostic assays. Approximately 50% of ET patients express *JAK2* V617F, 30% *CALR* (both type 1 and type 2), and 8% *MPL* mutations. ET patients lacking a canonical MPN driver mutation usually have a benign prognosis.

TABLE 108-6 Causes of Thrombocytosis

Tissue inflammation: collagen vascular disease, inflammatory bowel disease	Hemorrhage
Malignancy	Iron-deficiency anemia
Infection	Surgery
Myeloproliferative disorders: polycythemia vera, primary myelofibrosis, essential thrombocytosis, chronic myelogenous leukemia	Rebound: Correction of vitamin B_{12} or folate deficiency, post-ethanol abuse, postsplenectomy
Myelodysplastic disorders: 5q– syndrome, idiopathic refractory sideroblastic anemia	Hemolysis
Postsplenectomy or hyposplenism	Familial: Thrombopoietin overproduction, *JAK2, CALR*, or *MPL* mutations

ETIOLOGY

Megakaryocytopoiesis and platelet production depend on thrombopoietin and its receptor MPL. As in the case of early erythroid and myeloid progenitor cells, early megakaryocytic progenitors require the presence of interleukin 3 (IL-3) and stem cell factor for optimal proliferation in addition to thrombopoietin. Their subsequent terminal development is also enhanced by the chemokine stromal cell-derived factor 1 (SDF-1). Interestingly, terminal megakaryocyte maturation and platelet production do not require thrombopoietin.

Megakaryocytes are unique among hematopoietic progenitor cells because reduplication of their genome is endomitotic rather than mitotic and promoted by thrombopoietin. Unlike erythropoietin, thrombopoietin is produced only in the liver but has important functions in the bone marrow where it functions to maintain hematopoietic stem cells quiescent in their endosteal niches; once released from their niches, thrombopoietin promotes the proliferation of these cells in the sinusoidal niches. Like plasma erythropoietin and its target erythroblasts, an inverse correlation exists between the platelet count and plasma thrombopoietin. However, unlike erythropoietin, thrombopoietin is only constitutively produced and the plasma thrombopoietin level is controlled by the size of the platelet and megakaryocyte progenitor cell pools. Also, in contrast to erythropoietin, but like its myeloid counterparts, granulocyte and granulocyte-macrophage colony-stimulating factors, thrombopoietin not only enhances the proliferation of its target cells but also enhances the reactivity of their end-stage product, the platelet. Paradoxically, in the three MPNs, expression of the thrombopoietin receptor, MPL, is impaired and plasma thrombopoietin is increased despite the increased number of megakaryocytes and platelets.

The clonal nature of ET was established by analysis of glucose-6-phosphate dehydrogenase isoenzyme expression in patients hemizygous for this gene. Although thrombocytosis is its principal manifestation, like the other MPNs, the hematopoietic stem cell is involved in ET. Furthermore, a number of families have been described in which ET was inherited, in one instance as an autosomal dominant trait. In addition to ET, PMF and PV have also been observed in such kindreds.

CLINICAL FEATURES

Clinically, ET is most often identified incidentally when a platelet count is obtained during the course of a routine medical evaluation. Occasionally, review of previous blood counts will reveal that an elevated platelet count was present but overlooked for many years. No symptoms or signs are specific for ET, but these patients can have hemorrhagic and thrombotic tendencies expressed as easy bruising for the former and microvascular occlusive events for the latter such as erythromelalgia, ocular migraine, or a TIA. Physical examination is generally unremarkable. Splenomegaly is indicative of another MPN, in particular PV, PMF, or CML.

Anemia is unusual, but a mild neutrophilic leukocytosis is not. The blood smear is most remarkable for the number of platelets present, some of which may be very large. The large mass of circulating platelets may prevent the accurate measurement of serum potassium due to release of platelet potassium upon blood clotting. This type of hyperkalemia is a test tube artifact and not associated with electrocardiographic abnormalities. Similarly, arterial oxygen measurements can be inaccurate unless thrombocythemic blood is collected on ice. The prothrombin and partial thromboplastin times are normal, whereas abnormalities of platelet function such as a prolonged bleeding time and impaired platelet aggregation can be present. However, despite much study, no platelet function abnormality is characteristic of ET, and no platelet function test predicts the risk of clinically significant bleeding or thrombosis.

The elevated platelet count may hinder marrow aspiration, but marrow biopsy usually reveals megakaryocyte hypertrophy and hyperplasia, as well as an overall increase in marrow cellularity. If marrow reticulin is increased, another diagnosis should be considered. The absence of stainable iron demands an explanation because iron deficiency alone can cause thrombocytosis, and absent marrow iron in the presence of marrow hypercellularity is a feature of PV.

Nonrandom cytogenetic abnormalities occur in ET but are uncommon, and no specific or consistent abnormality is notable, even those involving chromosomes 3 and 1, where the genes for thrombopoietin and its receptor, MPL, respectively, are located.

DIAGNOSIS

Thrombocytosis is encountered in a broad variety of clinical disorders (Table 108-6), in many of which inflammatory cytokine production is increased. The absolute level of the platelet count is not a useful diagnostic aid for distinguishing between benign and clonal causes of thrombocytosis. About 50% of ET patients express the *JAK2* V617F mutation. When *JAK2* V617F is absent, cytogenetic evaluation is mandatory to determine if the thrombocytosis is due to CML or a myelodysplastic disorder such as the 5q– syndrome or sideroblastic anemia. Because the *BCR-ABL* translocation can be present in the absence of the Ph chromosome, and because the *BCR-ABL* reverse transcriptase polymerase chain reaction is associated with false-positive results, fluorescence in situ hybridization (FISH) analysis for *BCR-ABL* is the preferred assay in patients with thrombocytosis in whom a cytogenetic study for the Ph chromosome is negative. *CALR* mutations (type 1 or type 2) are present in 30% and *MPL* mutations are present in 8% of ET patients who do not have a *JAK2* mutation. Anemia and ringed sideroblasts are not features of ET, but they are features of idiopathic refractory sideroblastic anemia with the *SF3B1* mutation, and in some of these patients, thrombocytosis occurs in association with expression of *JAK2* V617F, *CALR*, or an *MPL* mutation. Significant splenomegaly should suggest the presence of another MPN such as PV or PMF, because splenomegaly can mask the presence of erythrocytosis. Importantly, what appears to be ET can evolve into PV (usually in women with *JAK2* V617F) or PMF (usually in men with type 1 *CALR* mutations) after a period of many years due to clonal evolution or succession. There is sufficient overlap of the *JAK2* V617F neutrophil allele burden between ET and PV that this test cannot be used as a distinguishing diagnostic feature with the exception that, in ET, the quantitative *JAK2* V617F neutrophil allele is never greater than 50%, and importantly in this regard, 64% of *JAK2* V617F–positive ET patients in one study actually were found to have PV when red cell mass and plasma volume determinations were performed. Claims that ET and PV form a biological continuum are unfounded as these disorders have different gene expression profiles and different natural histories.

COMPLICATIONS

Perhaps no other condition in clinical medicine has caused otherwise astute physicians to intervene inappropriately more often than thrombocytosis, particularly if the platelet count is $>1 \times 10^6/\mu L$. It is commonly believed that a high platelet count causes thrombosis; however, no controlled clinical study has ever established this association, and in patients younger than age 60 years, the incidence of thrombosis was not greater in patients with thrombocytosis than in age-matched controls, and tobacco use appears to be the most important risk factor for thrombosis in ET patients.

To the contrary, very high platelet counts are associated primarily with hemorrhage due to acquired von Willebrand's disease. This is not meant to imply that an elevated platelet count cannot cause symptoms in an ET patient, but rather that the focus should be on the patient, not the platelet count. For example, some of the most dramatic neurologic problems in ET are migraine-related and respond only to lowering of the platelet count, whereas other symptoms such as erythromelalgia respond simply to platelet cyclooxygenase-1 inhibitors such as aspirin or ibuprofen, without a reduction in platelet number. Still others may represent an interaction between an atherosclerotic vascular system and a high platelet count, and others may have no relationship to the platelet count whatsoever. Recognition that PV can present with thrombocytosis alone as well as the discovery of previously unrecognized causes of hypercoagulability **(Chaps. 121 and 122)** make the older literature on the complications of thrombocytosis unreliable.

TREATMENT

Essential Thrombocytosis

Survival of ET patients is not different than the general population regardless of their driver mutation. An elevated platelet count in an asymptomatic patient without cardiovascular risk factors or tobacco use requires no therapy. Indeed, before any therapy is initiated in a patient with thrombocytosis, the cause of symptoms must be clearly identified as due to the elevated platelet count. When the platelet count rises above $1 \times 10^6/\mu L$, a substantial quantity of high-molecular-weight von Willebrand multimers are removed from the circulation and destroyed by the enlarged platelet mass, resulting in an acquired form of von Willebrand's disease. This can be identified by a reduction in ristocetin cofactor activity. In this situation, aspirin could promote hemorrhage. Bleeding in this situation is rarely spontaneous and usually responds to tranexamic acid or ε-aminocaproic acid, which can be given prophylactically before and after elective minor surgery.

Plateletpheresis is at best a temporary and inefficient remedy that is rarely required. Importantly, ET patients treated with ^{32}P or alkylating agents are at risk of developing acute leukemia without any proof of benefit; combining either therapy with hydroxyurea increases this risk. If platelet reduction is deemed necessary on the basis of symptoms refractory to salicylates alone, pegylated IFN-α, the quinazoline derivative anagrelide, or hydroxyurea can be used to reduce the platelet count, but none of these is uniformly effective or without significant side effects. Hydroxyurea and aspirin were more effective than anagrelide and aspirin for prevention of TIA because hydroxyurea is a nitric oxide donor, but they were not more effective for the prevention of other types of arterial thrombosis and actually less effective for venous thrombosis. The risk of gastrointestinal bleeding is also higher when aspirin is combined with anagrelide. Normalizing the platelet count does not prevent either arterial or venous thrombosis. Pegylated IFN can produce a complete molecular remission in some ET patients, but a role for it or ruxolitinib in ET management has not yet been established.

As more clinical experience is acquired, ET appears more benign than previously thought. Evolution to acute leukemia is more likely to be a consequence of therapy than of the disease itself. In managing patients with thrombocytosis, the physician's first obligation is to do no harm.

FURTHER READING

Alvarez-Larran A et al: Antiplatelet therapy versus observation in low-risk essential thrombocythemia with CALR mutation. Haematologica 101:926, 2016.

Guglielmelli P et al: Clinical impact of mutated JAK2 allele burden reduction in polycythemia vera and essential thrombocythemia. Am J Hematol 99:1550, 2024.

Passamonti F et al: Myelofibrosis. Blood 141:1954, 2023.

Spivak JL: How I treat polycythemia vera. Blood 134:341, 2019.

Wouters HJCM et al: Erythrocytosis in the general population: Clinical characteristics and association with clonal hematopoiesis. Blood Adv 4:6353, 2020.

109

Acute Myeloid Leukemia

William Blum

INCIDENCE

Acute myeloid leukemia (AML) is a neoplasm characterized by infiltration of the blood, bone marrow, and other tissues by proliferative, clonal, poorly differentiated cells of the hematopoietic system. These leukemias comprise a spectrum of malignancies that untreated are uniformly fatal. In 2023, the estimated number of new AML cases in the United States was 20,380. AML is the diagnosis in 1% of all cancer cases and 31% of all new acute leukemias but causes 62% of leukemic deaths. AML is the most common acute leukemia in older patients, with a median age at diagnosis of 69 years. U.S. registry data report that only 32% of patients survive 5 years.

ETIOLOGY

Most cases of AML are idiopathic. Genetic predisposition, radiation, chemical/other occupational exposures, and drugs have been implicated in the development of AML, but AML with established etiology is relatively uncommon. No direct evidence suggests a viral etiology. Genome sequencing studies suggest that most cases of AML arise from a limited number of mutations that accumulate with advancing age. Indeed, genome sequencing provides paradigm-shifting advances in our understanding of leukemogenesis. The Cancer Genome Atlas (TCGA) and other databases demonstrate that blood cells from up to 5–6% of normal individuals aged >70 years contain potentially "premalignant" mutations that are associated with clonal expansion.

Use of the term *premalignant* to describe these lesions is not precisely accurate; rather, these mutations represent clonal hematopoiesis of *indeterminate* potential (CHIP; sometimes called age-related clonal hematopoiesis). The genes most commonly mutated in CHIP are the epigenetic regulators *DNMT3A, TET2,* and *ASXL1.* Study of CHIP is important because it has relevance not just to blood cancer evolution but also other medical conditions. Clonal expansion driven by the acquisition of new mutations is associated with a 10-fold increase in risk for developing a hematologic malignancy (compared to matched patients without CHIP), but it is clear that additional "hits" must occur to drive toward leukemia. We do not yet fully understand why or how these secondary lesions occur.

Patients with CHIP also have increased risk of cardiovascular mortality that is not fully explained. The link between these two seemingly unrelated issues (cardiovascular and hematologic malignancy) may lie in the interactions between circulating clonally expanded blood cells and vascular endothelium. A "proinflammatory" state caused by clonal, infiltrating monocytes leads to accelerated atherosclerotic plaque development and altered cardiac remodeling. Similar phenomena may occur in the marrow/blood. An altered relationship between hematopoietic stem cells and the marrow microenvironment (along with altered immune surveillance) contributes to clonal survival and expansion. These perturbations increase the likelihood that a clone with somatic mutations may survive, acquire additional mutations, and then further expand eventually to leukemia. Whether early identification of CHIP in patients will provide therapeutic opportunities for patients remains to be seen. Certainly, modifying cardiovascular risk in patients with CHIP seems prudent, but development of mutation-directed therapy designed to eliminate a problematic clone and prevent future leukemia is likely to be more elusive.

Genetic Predisposition Myeloid neoplasms occur sporadically in adults; inherited predisposition is uncommon. Yet, it is clear that myeloid neoplasms with germline predisposition represent an important and growing subset of disease. Germline mutations associated with increased risk of developing a myeloid neoplasm include *CEBPA*, *DDX41*, *TP53*, *RUNX1*, *ANKRD26*, *ETV6*, and *GATA2*, and others (Table 109-1). Myeloid neoplasms with germline predisposition are

TABLE 109-1 World Health Organization 2022, Subtypes of Myeloid Neoplasms Associated with Germline Predisposition

Myeloid neoplasms with germline predisposition without a preexisting platelet disorder or organ dysfunction
• Germline *CEBPA* P/LP variant (CEBPA-associated familial AML)
• Germline *DDX41* P/LP variant[a]
• Germline *TP53* P/LP variant[a] (Li-Fraumeni syndrome)
Myeloid neoplasms with germline predisposition and preexisting platelet disorder
• Germline *RUNX1* P/LP variant[a] (familial platelet disorder with associated myeloid malignancy [FPD-MM])
• Germline *ANKRD26* P/LP variant[a] (thrombocytopenia 2)
• Germline *ETV6* P/LP variant[a] (thrombocytopenia 5)
Myeloid neoplasms with germline predisposition and potential organ dysfunction
• Germline *GATA2* P/LP variant (GATA2 deficiency)
• Bone marrow failure syndromes
• Severe congenital neutropenia (SCN)
• Shwachman-Diamond syndrome (SDS)
• Fanconi anemia (FA)
• Telomere biology disorders
• RASopathies (neurofibromatosis type 1, CBL syndrome, Noonan syndrome or Noonan syndrome-like disorders[a])
• Down syndrome[a]
• Germline *SAMD9* P/LP variant (MIRAGE syndrome)
• Germline *SAMD9L* P/LP variant (SAMD9L-related ataxia pancytopenia syndrome)[b]
• Biallelic germline *BLM* P/LP variant (Bloom syndrome)

[a]Lymphoid neoplasms can also occur. [b]Ataxia is not always present.

Abbreviations: LP, likely pathogenic; P, pathogenic.

Source: Modified from JD Khoury et al: The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and histiocytic/dendritic neoplasms. Leukemia 36:1703, 2022.

a feature of several well-described clinical syndromes, including bone marrow failure disorders (e.g., Fanconi anemia, Shwachman-Diamond syndrome, Diamond-Blackfan anemia) and telomere biology disorders (e.g., dyskeratosis congenita). As new mutations and associations are added to a rapidly growing list, it is clear that genetic predisposition plays a larger role than has been previously understood.

Several genetic syndromes with somatic cell chromosome aneuploidy, such as Down syndrome with trisomy 21, are associated with an increased incidence of AML. Down syndrome–associated AML in young children (<4 years) is typically of megakaryocytic differentiation and is associated with mutation in the *GATA1* gene. Such patients have excellent clinical outcomes but require dose modification of chemotherapy due to high treatment-related toxicities. Inherited diseases with defective DNA repair (e.g., Fanconi anemia, Bloom syndrome, and ataxia-telangiectasia) are also associated with AML. Each syndrome is associated with unique clinical features and atypical toxicities with chemotherapy, requiring expert care. Congenital neutropenia (Kostmann syndrome), due to mutations in the genes encoding the granulocyte colony-stimulating factor receptor and neutrophil elastase, is another disorder that may evolve into AML.

Chemical, Radiation, and Other Exposures Anticancer drugs are the leading cause of therapy-associated AML. AML post chemotherapy (AML-pCT) with alkylating agents occurs 4–6 years after exposure; affected individuals often have multilineage dysplasia, monosomy/aberrations in chromosomes 5 and 7, mutations of *TP53*, and poor prognosis. AML-pCT with topoisomerase II inhibitors occurs 1–3 years after exposure; affected individuals often have AML with monocytic features and aberrations involving chromosome 11q23 (involved gene previously called *MLL*, now *KMT2A*). The risk of acute leukemia is much higher after combined-modality therapy with alkylating agent–based chemotherapy plus external beam radiation therapy. Exposure to ionizing radiation, benzene, chloramphenicol, phenylbutazone, and other drugs can result in bone marrow failure that may evolve into AML.

■ CLASSIFICATION

Historically, marrow (or blood) myeloid blast count of ≥20% established the diagnosis of AML. However, biologically distinct groups are now primarily classified based on genetic aberrations, in addition to clinical features and light microscopy. It is increasingly understood that genetic aberrations drive clinical presentation and clinical course, and thus, in 2022, the World Health Organization (WHO) eliminated the blast percent requirement among cases with specific, defined genetic aberrations **(Table 109-2)**. In the WHO system referenced here, genetic aberrations are notated by the gene mutation or fusion involved; learners may find clarity by also reviewing the European LeukemiaNet (ELN) risk classification **(Table 109-3)**, which lists associated cytogenetic abnormalities alongside the gene fusions. Dueling classification systems of the WHO and the International Consensus Classification (ICC) assign AML diagnosis inconsistently with regard to blast percent, but the differences are semantic. Cases with <20% blasts that are still classified as AML have recurrent genetic abnormalities including t(15;17), t(8;21), inv(16), t(16;16), rearrangements involving *KMT2A* (with many different fusion partners), mutations in *NPM1* (nucleophosmin), and other defined genetic aberrations. The emergence of novel targeted treatment options for specific aberrations, such as menin inhibitors for patients with *KMT2A* rearrangement, *NPM1* mutation, or (possibly) *NUP98* rearrangement, suggests that a practical approach to classification based on the presence of a specific aberration regardless of blast count may indeed be most appropriate. All AML cells contain genetic mutations, most of which are recurring; in 2024, the majority of AML patients will have a genetic lesion that can be specifically targeted with a novel drug.

TABLE 109-2 World Health Organization 2022 Classification of Acute Myeloid Leukemia

Acute myeloid leukemia with defining genetic abnormalities
Acute promyelocytic leukemia with *PML*::*RARA* fusion
Acute myeloid leukemia with *RUNX1*::*RUNX1T1* fusion
Acute myeloid leukemia with *CBFB*::*MYH11* fusion
Acute myeloid leukemia with *DEK*::*NUP214* fusion
Acute myeloid leukemia with *RBM15*::*MRTFA* fusion
Acute myeloid leukemia with *BCR*::*ABL1* fusion[a]
Acute myeloid leukemia with *KMT2A* rearrangement
Acute myeloid leukemia with *MECOM* rearrangement
Acute myeloid leukemia with *NUP98* rearrangement[b]
Acute myeloid leukemia with *NPM1* mutation
Acute myeloid leukemia with *CEBPA* mutation[a,c]
Acute myeloid leukemia, myelodysplasia-related[a,d]
Acute myeloid leukemia with other defined genetic alterations[a]
Acute myeloid leukemia, defined by differentiation
Acute myeloid leukemia with minimal differentiation
Acute myeloid leukemia without maturation
Acute myeloid leukemia with maturation
Acute basophilic leukemia
Acute myelomonocytic leukemia
Acute monocytic leukemia
Acute erythroid leukemia
Acute megakaryoblastic leukemia

[a] Requires ≥20% blasts. [b] NUP98 rearrangements involve 11p15, with many fusion partners, usually cryptic on cytogenetic analysis. [c] Includes biallelic (biCEBPA) as well as single mutations located in the basic leucine zipper (bZIP) region. [d] Defining characteristics of Acute myeloid leukemia, myelodysplasia-related (AML-MR) include history of myelodysplastic syndrome (MDS) or MDS/myeloproliferative neoplasm (MPN) and/or one of the following: complex karyotype, several other specific chromosome aberrations typical of MDS, or mutation of *ASXL1, BCOR, EZH2, SF3B1, SRSF2, STAG2, U2AF1,* or *ZRSR2.*

Source: Modified from JD Khoury et al: The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and histiocytic/dendritic neoplasms. Leukemia 36:1703, 2022.

Genetic Findings Subtypes of AML are recognized due to the presence or absence of specific, recurrent cytogenetic, and/or genetic abnormalities. For example, the diagnosis of *acute promyelocytic leukemia* (APL) is based on the presence of the t(15;17)(q22;q12) cytogenetic rearrangement or the *PML-RARA* fusion. Similarly, core binding factor (CBF) AML is designated based on the presence of t(8;21)(q22;q22), inv(16)(p13.1q22), or t(16;16)(p13.1;q22) or the respective fusion products *RUNX1-RUNX1T1* and *CBFB-MYH11*. Each of these three groups identifies patients with favorable clinical outcomes when appropriately treated.

Many genetic AML subtypes are associated with a specific morphologic appearance such as a complex karyotype (and/or mutation of *TP53*) and dysplastic morphology in AML, myelodysplasia-related (AML-MR). One abnormality is invariably associated with a specific morphologic feature: t(15;17)(q22;q12) or the molecular fusion *PML-RARA* with APL. Further examples include inv(16)(p13.1q22) with AML and abnormal bone marrow eosinophils; t(8;21)(q22;q22) and slender Auer rods, expression of CD19, and increased normal eosinophils; and rearrangements involving *KMT2A* with monocytic features. AML with mutation of *NPM1*, especially when co-occurring with mutation of *FLT3* (fms-related tyrosine kinase 3), often presents with blasts having "cup-shaped" nuclear morphology. Recurring genetic aberrations in AML may also be loosely associated with specific clinical characteristics. More commonly associated with younger age are t(8;21) and t(15;17), and with older age are del(5q), del(7q), and mutated *TP53*. Myeloid sarcomas are associated with t(8;21); disseminated intravascular coagulation (DIC) is associated with t(15;17). *KMT2A* aberrations and monocytic leukemia are associated with extramedullary sites of involvement at presentation, especially gingival hypertrophy. High leukocyte count is commonly observed with *NPM1* and/or *FLT3* mutation. Many other cytogenetic and genetic findings commonly, but not always, are associated with a morphologic description, highlighting the necessity of genetic and cytogenetic testing for precise diagnosis.

TABLE 109-3 2022 European LeukemiaNet Risk Classification of Acute Myeloid Leukemia (AML) by Genetics at Initial Diagnosis[a]

RISK CATEGORY	GENETIC ABNORMALITY
Favorable	• t(8;21)(q22;q22.1)/*RUNX1*::*RUNX1T1*[b,c] • inv(16)(p13.1q22) or t(16;16)(p13.1;q22)/*CBFB*::*MYH11*[b,c] • Mutated *NPM1*[b,d] without *FLT3*-ITD • bZIP in-frame mutated *CEBPA*[e]
Intermediate	• Mutated *NPM1*[b,d] with *FLT3*-ITD • Wild-type *NPM1* with *FLT3*-ITD (without adverse-risk genetic lesions) • t(9;11)(p21.3;q23.3)/*MLLT3*::*KMT2A*[b,f] • Cytogenetic and/or molecular abnormalities not classified as favorable or adverse
Adverse	• t(6;9)(p23.3;q34.1)/*DEK*::*NUP214* • t(v;11q23.3)/*KMT2A*-rearranged[g] • t(9;22)(q34.1;q11.2)/*BCR*::*ABL1* • t(8;16)(p11.2;p13.3)/*KAT6A*::*CREBBP* • inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2)/*GATA2, MECOM(EVI1)* • t(3q26.2;v)/*MECOM*(*EVI1*)-rearranged • –5 or del(5q); –7; –17/abn(17p) • Complex karyotype,[h] monosomal karyotype[i] • Mutated *ASXL1, BCOR, EZH2, RUNX1, SF3B1, SRSF2, STAG2, U2AF1,* and/or *ZRSR2*[j] • Mutated *TP53*[k]

[a]Frequencies, response rates, and outcome measures should be reported by risk category and, if sufficient numbers are available, by specific genetic lesions indicated. Acute promyelocytic leukemia is excluded from this table. [b]Mainly based on results observed in intensively treated patients. Initial risk assignment may change during the treatment course based on the results from analyses of measurable residual disease. [c]Concurrent *KIT* and/or *FLT3* gene mutation does not alter risk categorization. [d]AMLs with *NPM1* mutation and adverse-risk cytogenetic abnormalities are categorized as adverse risk. [e]Only in-frame mutations affecting the basic leucine zipper (bZIP) region of *CEBPA*, irrespective of whether they occur as monoallelic or biallelic mutations, have been associated with favorable outcome. [f]The presence of t(9;11)(p21.3;q23.3) takes precedence over rare, concurrent adverse-risk gene mutations. [g]Excluding *KMT2A* partial tandem duplication (PTD). [h]Complex karyotype: three or more unrelated chromosome abnormalities in the absence of other class-defining recurring genetic abnormalities; excludes hyperdiploid karyotypes with three or more trisomies (or polysomies) without structural abnormalities. [i]Monosomal karyotype: presence of two or more distinct monosomies (excluding loss of X or Y), or one single autosomal monosomy in combination with at least one structural chromosome abnormality (excluding core-binding factor AML). [j]For the time being, these markers should not be used as an adverse prognostic marker if they co-occur with favorable-risk AML subtypes. [k]*TP53* mutation at a variant allele fraction of at least 10%, irrespective of the *TP53* allelic status (mono- or biallelic mutation); *TP53* mutations are significantly associated with AML with complex and monosomal karyotype.

Source: Reproduced with permission from H Döhner et al: Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood 140:1345, 2022.

WHO classification incorporates molecular abnormalities by recognizing fusion genes or specific genetic mutations with a role in leukemogenesis. As a classic example, t(15;17) results in the fusion gene *PML-RARA* that encodes a chimeric protein, promyelocytic leukemia (Pml)–retinoic acid receptor α (Rarα), which is formed by the fusion of the retinoic acid receptor α (*RARA*) gene from chromosome 17 and the promyelocytic leukemia (*PML*) gene from chromosome 15. Unique clinical therapy with retinoic acid and arsenic trioxide has revolutionized the care of APL patients (see "Treatment of Acute Promyelocytic Leukemia" section). Similar examples of molecular subtypes included in the category of AML with recurrent genetic abnormalities are those characterized by the leukemogenic fusion genes *RUNX1-RUNX1T1* and *CBFB-MYH11* and the so-called CBF AML subtypes noted cytogenetically as t(8;21), inv(16), or t(16;16). Additional examples of fusions are *MLLT3-KMT2A* and *DEK-NUP214*, resulting from t(9;11) and t(6;9)(p23;q34). Mutated genes are also critical elements of AML classification. The most common is AML with mutated *NPM1*, a mutation seen in 30% of AML patients and in 60% of those with cytogenetically normal AML (CN-AML). Another subtype is AML with mutated *CEBPA* (specifically in-frame bZIP mutation). Both are associated with more favorable clinical outcomes, though the presence of coexisting mutation in *FLT3* negatively affects *NPM1* prognostic impact. Activating mutations of *FLT3* are present in ~30% of adult AML patients, primarily due to internal tandem duplications (ITDs) in the juxtamembrane domain that have negative prognostic impact. In contrast, point mutations of the activating loop of the kinase, called tyrosine kinase domain (TKD) mutations, have uncertain prognostic impact. Aberrant activation of the *FLT3*-encoded protein provides increased proliferation and antiapoptotic signals to the myeloid progenitor cell. *FLT3*-ITD, the most common of the *FLT3* mutations, also occurs preferentially in patients with CN-AML. The importance of identifying *FLT3*-ITD at diagnosis relates to the fact that it is not only useful as a negative prognosticator but also predicts response to specific treatment such as a tyrosine kinase inhibitor (TKI). Several TKIs targeting FLT3 are either approved for AML (e.g., midostaurin or quizartinib, only in first-line therapy in combination with chemotherapy; gilteritinib, in relapse as monotherapy) or currently in clinical investigation.

Immunophenotypic Findings The immunophenotype of human leukemia cells can be studied by multiparameter flow cytometry after the cells are labeled with monoclonal antibodies to cell-surface antigens. This can be important in quickly distinguishing AML from acute lymphoblastic leukemia and for identifying some subtypes of AML. For example, AML with minimal differentiation, characterized by immature morphology and no lineage-specific cytochemical reactions, may be diagnosed by flow-cytometric demonstration of the myeloid-specific antigens cluster designation (CD) 13 and/or 117. Similarly, acute megakaryoblastic leukemia can often be diagnosed only by expression of the platelet-specific antigens CD41 and/or CD61. Although flow cytometry is widely used, and in some cases essential for the diagnosis of AML, it has only a supportive role in establishing the different subtypes of AML through the WHO classification given the paramount importance of genetics. Increasingly, multiparameter flow cytometry is used for the measurement of measurable residual disease (MRD) after remission is achieved.

PROGNOSTIC FACTORS

Several factors predict outcome of AML patients treated with chemotherapy; they should be used for risk stratification and treatment guidance.

Chromosome and molecular investigations performed at diagnosis provide the most important prognostic information. Patients with t(15;17) have a very good prognosis (~85% cured), and those with t(8;21) and inv(16) have a good prognosis (~55% cured), whereas those with no cytogenetic abnormality have an intermediate outcome risk (~40% cured). Patients with *NPM1* mutation without a *FLT3*-ITD also have favorable risk and high cure rate; conversely, those with *TP53* mutation, complex karyotype, t(6;9), inv(3), or –7 have adverse risk and very poor clinical outcomes with virtually no chance for cure without transplantation.

For patients lacking prognostic cytogenetic abnormalities, i.e., those with CN-AML, testing for several mutated genes can help to risk-stratify. In addition to *NPM1* mutation and *FLT3*-ITD as described above, in-frame bZIP *CEBPA* mutation has favorable prognosis. Given the proven prognostic importance of *NPM1*, *FLT3*, and *CEBPA*, molecular assessment of these genes and others at diagnosis has been incorporated into AML management guidelines by the National Comprehensive Cancer Network (NCCN) and the ELN. The same markers help to define genetic groups in the ELN standardized reporting system, which is based on both cytogenetic and molecular abnormalities and is used for comparing clinical features/treatment response among subsets of patients reported across different clinical studies (Table 109-3). These genetic groups should be used for risk stratification and treatment guidance (note that APL is excluded from the table).

In addition to *NPM1, FLT3, CEBPA*, and *TP53* mutations, molecular aberrations in other genes are routinely used for prognostication. Among these mutated genes are those encoding receptor tyrosine kinases, transcription factors (*RUNX1* and *WT1*), and epigenetic modifiers (*ASXL1*, *DNMT3A*, isocitrate dehydrogenase 1 [*IDH1*], *IDH2*, and *TET2*). Among an expanding panel of mutated genes associated with adverse risk (at least when not coexisting with favorable aberrations) are *ASXL1, BCOR, EZH2, RUNX1, SF3B1, SRSF2, STAG2, U2AF1, ZRSR2,* and *TP53*. Because prognostic molecular markers in AML are not mutually exclusive and often occur concurrently (>80% patients have at least two or more prognostic gene mutations), distinct marker combinations will certainly be part of continued evolution of AML classification, prognosis, and treatment.

Novel drugs that inhibit/modulate cellular pathways activated by genetic aberrations (especially *FLT3*, *IDH1*, *IDH2*, and *NPM1*/*KMT2A*/*NUP98*) have been remarkably effective in subsets of disease (see section on treatment of AML). Epigenetic changes (e.g., DNA methylation and/or posttranslational histone modification) and microRNAs are often involved in deregulation of genes involved in hematopoiesis, contribute to leukemogenesis, and may associate with the previously discussed prognostic gene mutations. These changes have been shown to provide biologic insights into leukemogenic mechanisms and provide independent prognostic information. Therapeutic progress based on advances in understanding the role of epigenetic changes in AML over the last decade has been tremendous. For example, in patients with mutations of *IDH1* or *IDH2*, novel enzymes produced from these respective mutations have aberrant activity and "hijack" the citric acid cycle. These mutations lead to production of a novel "oncometabolite," 2-hydroxyglutarate, which disrupts a myriad of epigenetic processes. Pharmacologic inhibition of these aberrant enzymes can reverse these leukemogenic activities and restore normal marrow function (though monotherapy is not typically curative).

In addition to cytogenetic and molecular aberrations, several clinical factors are associated with outcome in AML. Age at diagnosis is the most important. Advancing age is associated with a poor prognosis for two reasons: (1) its influence on the ability to survive induction therapy due to coexisting medical comorbidities, and (2) with each successive decade of age, a greater proportion of patients have intrinsically more resistant disease/adverse genetic risk. Next, a prolonged symptomatic interval with cytopenias preceding AML diagnosis, or a history of antecedent hematologic disorders including myelodysplastic syndrome (MDS) or myeloproliferative neoplasm, is often found in older patients. Preexisting cytopenia is a clinical feature associated with a lower complete remission (CR) rate and shorter survival time. The CR rate is lower in patients who have had anemia, leukopenia, and/or thrombocytopenia for >3 months before the diagnosis of AML when compared to those without such a history. Responsiveness to chemotherapy declines as the duration of the antecedent disorder increases. Likewise, AML-pCT, typically developing after treatment with cytotoxic agents for other malignancies, is often resistant to treatment given its association with adverse genetic features. In general, older patients less frequently harbor favorable genetic aberrations (e.g., t[8;21], inv[16], and t[16;16], *NPM1* mutation) and more frequently harbor adverse genetic aberrations (e.g., complex karyotypes, mutations in *ASXL1*, *TP53*).

Other factors independently associated with worse outcome are poor performance status, which influences ability to survive induction therapy, and a high presenting leukocyte count that in some series is an adverse prognostic factor for attaining a CR. Among patients with hyperleukocytosis (>100,000/μL), early central nervous system bleeding and pulmonary leukostasis contribute to poor outcomes.

Following administration of therapy, achievement of CR is associated with better outcome and longer survival, of course. CR is defined after examination of both blood and bone marrow and essentially represents eradication of detectable leukemia *and* restoration of normal hematopoiesis. The blood neutrophil count must be ≥1000/μL and the platelet count ≥100,000/μL for formal criteria; CR with incomplete recovery of counts is a lesser but still meaningful response. Hemoglobin concentration is not considered in determining CR. Circulating blasts should be absent. Although rare blasts may be detected in the blood during marrow regeneration, they should disappear on successive studies. At CR, the bone marrow should contain <5% blasts, and extramedullary leukemia should not be present.

CLINICAL PRESENTATION

Symptoms Patients with AML usually present with nonspecific symptoms that begin gradually, though sometimes abruptly, and are the consequence of anemia, leukocytosis, leukopenia/leukocyte dysfunction, or thrombocytopenia. Nearly half have symptoms for ≤3 months before the leukemia is diagnosed.

Fatigue is a frequent first symptom among AML patients. Anorexia and weight loss are common. Fever with or without an identifiable infection is the initial symptom in ~10% of patients. Signs of abnormal hemostasis (bleeding, easy bruising) are common. Bone pain, lymphadenopathy, nonspecific cough, headache, or diaphoresis may also occur.

Rarely, patients may present with symptoms from a myeloid sarcoma (a tumor mass consisting of myeloid blasts occurring at anatomic sites other than bone marrow). Sites involved are most commonly the skin, lymph node, gastrointestinal tract, soft tissue, and testis. This may precede or coincide with blood and/or marrow involvement by AML. Patients who present with isolated myeloid sarcoma typically develop blood and/or marrow involvement quickly thereafter and cannot be cured with local therapy (radiation or surgery) alone.

Physical Findings Fever, infection, and hemorrhage are often found at the time of diagnosis; splenomegaly, hepatomegaly, lymphadenopathy, and "bone pain" may also be present but less commonly. Hemorrhagic complications are most commonly and, classically, found in APL. APL patients often present with DIC-associated minor hemorrhage but may have significant gastrointestinal bleeding, intrapulmonary hemorrhage, or intracranial hemorrhage. Counterintuitively, thrombosis while less frequent is another well recognized clinical feature of DIC in APL. Complications associated with coagulopathy may also occur in monocytic AML and with extreme degrees of leukocytosis or thrombocytopenia in other morphologic subtypes. Retinal hemorrhages are detected in 15% of patients. Infiltration of the gingiva, skin, soft tissues, or meninges with leukemic blasts at diagnosis is characteristic of the monocytic subtypes and those with *KMT2A* chromosomal abnormalities.

Hematologic Findings Anemia is usually present at diagnosis, although it is not typically severe. The anemia is usually normocytic normochromic. Decreased erythropoiesis in the setting of AML often results in a reduced reticulocyte count, and red blood cell (RBC) survival is decreased by accelerated destruction. Active blood loss may rarely contribute to the anemia.

The median presenting leukocyte count for new AML cases is ~15,000/μL. Lower presenting leukocyte counts are more typical of older patients and those with antecedent hematologic disorders. Between 25 and 40% of patients have counts <5000/μL, and 20% have counts >100,000/μL. Fewer than 5% have no detectable leukemic cells in the blood. In AML blasts, the cytoplasm often contains primary (nonspecific) granules, and the nucleus shows fine, lacy chromatin with one or more nucleoli characteristic of immature cells. Abnormal rod-shaped granules called Auer rods are not uniformly present, but when they are, AML (and not acute lymphocytic leukemia) is certain **(Fig. 109-1)**.

Platelet counts <100,000/μL are found at diagnosis in ~75% of patients, and ~25% have counts <25,000/μL. Both morphologic and functional platelet abnormalities can be observed, including large and bizarre shapes with abnormal granulation and inability of platelets to aggregate or adhere normally to one another.

Pretreatment Evaluation Once the diagnosis of AML is suspected, thorough evaluation and initiation of appropriate therapy should follow. In addition to clarifying the subtype of leukemia, initial studies should evaluate the overall functional integrity of the major organ systems, including the cardiovascular, pulmonary, hepatic, and renal systems **(Table 109-4)**. Factors that have prognostic significance, either for achieving CR or for CR duration, should also be assessed before initiating treatment including cytogenetics and molecular markers. Leukemic

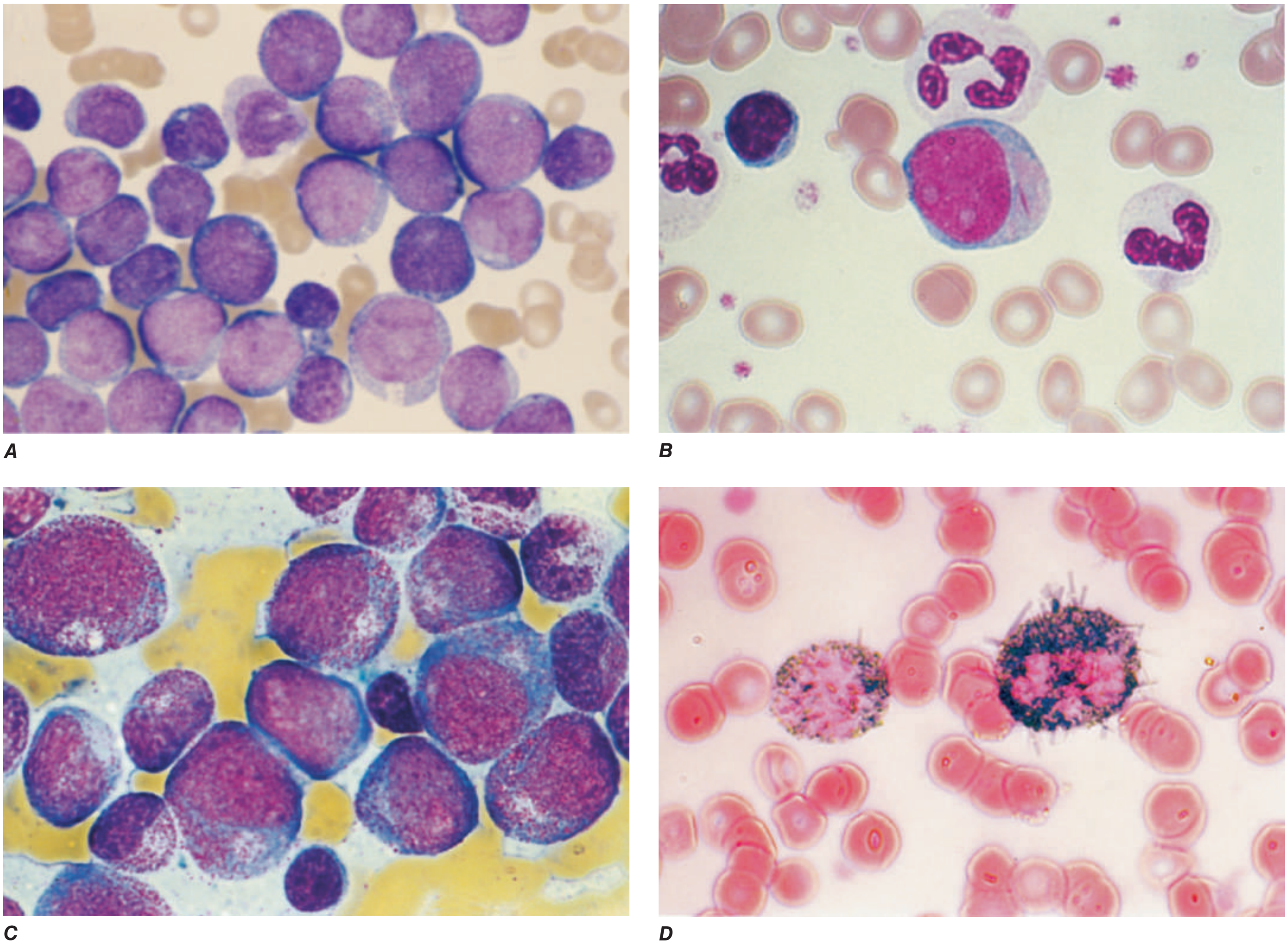

FIGURE 109-1 Morphology of acute myeloid leukemia (AML) cells. ***A.*** Uniform population of primitive myeloblasts with immature chromatin, nucleoli in some cells, and primary cytoplasmic granules. ***B.*** Leukemic myeloblast containing an Auer rod. ***C.*** Promyelocytic leukemia cells with prominent cytoplasmic primary granules. ***D.*** Peroxidase stain shows dark blue color characteristic of peroxidase in granules in AML.

cells should be obtained from all consenting patients and cryopreserved for future investigational testing as well as potential use as new diagnostics and therapeutics become available. All patients should be evaluated for infection. Patients with respiratory symptoms should undergo testing for the presence of the novel coronavirus, SARS-CoV-2, and other viruses before initiation of chemotherapy.

Most patients are anemic and thrombocytopenic at presentation. Replacement of the appropriate blood components, if necessary, should begin promptly. Because qualitative platelet dysfunction or the presence of an infection may increase the likelihood of bleeding, evidence of hemorrhage justifies the immediate use of platelet transfusion even if the platelet count is only moderately decreased.

About 50% of patients have a mild to moderate elevation of serum uric acid at presentation. Only 10% have marked elevations, but renal precipitation of uric acid and the nephropathy that may result is a serious but uncommon complication. The initiation of chemotherapy may aggravate hyperuricemia, and patients are usually started immediately on allopurinol and hydration at diagnosis. Rasburicase (recombinant uric oxidase) is also useful for treating uric acid nephropathy and often can normalize the serum uric acid level within hours with a single dose of treatment, although its expense suggests that limiting its use to patients with severe hyperuricemia and/or kidney injury may be prudent (and it should be avoided in setting of glucose-6-phosphate dehydrogenase deficiency). The presence of high concentrations of lysozyme, a marker for monocytic differentiation, may be etiologic in renal tubular dysfunction for a minority of patients.

TREATMENT

Acute Myeloid Leukemia

Treatment of the newly diagnosed patient with AML is usually divided into two phases, induction and postremission management (consolidation) **(Fig. 109-2)**. The initial goal is to induce CR. Once CR is obtained, further therapy must be given to prolong survival and achieve cure. The initial induction treatment and subsequent postremission therapy are chosen based on the patient's age, overall fitness, and cytogenetic/molecular risk. Intensive therapy with cytarabine and anthracycline in younger patients (<60 years) increases the cure rate of AML. In older patients, the benefit of intensive therapy is controversial in all but favorable-risk patients; novel approaches for selecting patients predicted to be responsive to treatment and new therapies are being pursued. Importantly, even infirm older patients should be considered for therapy; treatment is better than supportive care for all candidates. Improved options have emerged for older AML patients such as the addition of the BCL2 antagonist venetoclax to one of several low-intensity chemotherapies. Venetoclax is currently in testing in combination with intensive chemotherapies as well. Likewise, novel oral drugs targeting IDH1/IDH2, alone (IDH1) or in combination (IDH1 and IDH2) with low-intensity chemotherapy, may be considered as initial therapy for unfit older patients who have mutations in those respective pathways.

TABLE 109-4 Initial Diagnostic Evaluation and Management of Adult Patients with AML

History
Increasing fatigue or decreased exercise tolerance (anemia)
Excess bleeding or bleeding from unusual sites (DIC, thrombocytopenia)
Fevers or recurrent infections (neutropenia)
Headache, vision changes, nonfocal neurologic abnormalities (CNS leukemia or bleed)
Early satiety (splenomegaly)
Family history of AML (Fanconi, Bloom, or Kostmann syndromes or ataxia-telangiectasia)
History of cancer (exposure to alkylating agents, radiation, topoisomerase II inhibitors)
Occupational exposures (radiation, benzene, petroleum products, paint, smoking, pesticides)
Physical Examination
Performance status (prognostic factor)
Ecchymosis and oozing from IV sites (DIC, possible acute promyelocytic leukemia)
Fever and tachycardia (signs of infection)
Papilledema, retinal infiltrates, cranial nerve abnormalities (CNS leukemia)
Poor dentition, dental abscesses
Gum hypertrophy (leukemic infiltration, most common in monocytic leukemia)
Skin infiltration or nodules (leukemia infiltration, most common in monocytic leukemia)
Lymphadenopathy, splenomegaly, hepatomegaly
Back pain, lower extremity weakness (spinal granulocytic sarcoma, most likely in t[8;21] patients)
Laboratory and Radiologic Studies
CBC with manual differential cell count
Chemistry tests (electrolytes, creatinine, BUN, calcium, phosphorus, uric acid, hepatic enzymes, bilirubin, LDH, amylase, lipase)
Clotting studies (prothrombin time, partial thromboplastin time, fibrinogen, D-dimer)
Viral serologies (CMV, HSV-1, varicella-zoster)
RBC type and screen
HLA typing for potential allogeneic HCT
Bone marrow aspirate and biopsy (morphology, cytogenetics, flow cytometry, molecular studies)
Cryopreservation of viable leukemia cells
Myocardial function (echocardiogram or MUGA scan)
PA and lateral chest radiograph
Placement of central venous access device
Interventions for Specific Patients
Dental evaluation (for those with poor dentition)
Lumbar puncture (for those with symptoms of CNS involvement)
Screening spine MRI (for patients with back pain, lower extremity weakness, paresthesias)
Social work referral for patient and family psychosocial support
Counseling for All Patients
Provide patients with information regarding their disease and genetic risks, sperm banking or menstrual suppression, financial counseling, support group contact, and consent for tissue banking of leukemic cells

Abbreviations: AML, acute myeloid leukemia; BUN, blood urea nitrogen; CBC, complete blood count; CMV, cytomegalovirus; CNS, central nervous system; DIC, disseminated intravascular coagulation; HLA, human leukocyte antigen; HCT, hematopoietic stem cell transplantation; HSV, herpes simplex virus; IV, intravenous; LDH, lactate dehydrogenase; MRI, magnetic resonance imaging; MUGA, multigated acquisition; PA, posteroanterior; RBC, red blood (cell) count.

INDUCTION CHEMOTHERAPY

The most commonly used induction regimens (for patients other than those with APL) consist of combination chemotherapy with cytarabine and an anthracycline (e.g., daunorubicin, idarubicin).

Cytarabine is a cell cycle S-phase–specific antimetabolite that becomes phosphorylated intracellularly to an active triphosphate form that interferes with DNA synthesis. Anthracyclines are DNA intercalators. Their primary mode of action is thought to be inhibition of topoisomerase II, leading to DNA breaks.

In adults, cytarabine used at standard dose (100–200 mg/m^2) is administered as a continuous intravenous infusion for 7 days. With cytarabine, anthracycline therapy generally consists of daunorubicin (60–90 mg/m^2) or idarubicin (12 mg/m^2) intravenously on days 1, 2, and 3 (the 7+3 regimen). Other agents can be added (e.g., gemtuzumab ozogamicin) when 60 mg/m^2 of daunorubicin is used. Patients failing remission after one induction are offered reinduction with the same (or slightly modified) therapy. The CD33-targeting immunoconjugate gemtuzumab ozogamicin may be added to induction therapy for subsets of patients, especially those with CBF AML. Many alternative intensive approaches other than 7+3 chemotherapy exist and are commonly used.

In older patients (age ≥60–65 years), the outcome with conventional intensive therapy is generally poor due to a higher frequency of resistant disease and increased rate of treatment-related morbidity and mortality. Patients still fare far better with treatment than with supportive care only. Conventional therapy for fit older patients is similar to that for younger patients: the 7+3 regimen with standard-dose cytarabine and idarubicin (12 mg/m^2) or daunorubicin (60 mg/m^2). For patients aged >65 years, high-dose daunorubicin (90 mg/m^2) has increased toxicity and is not recommended. A liposomal preparation of cytarabine and daunorubicin in a fixed molar ratio may instead be administered, especially to fit patients with AML-MR. Patients over 75 years and those unable to receive intensive therapy due to medical comorbidity may receive repetitive cycles of lower intensity therapy with a hypomethylating agent (HMA; decitabine or azacitidine) or low-dose cytarabine in combination with daily venetoclax (BCL2 antagonist). As noted, targeted IDH1- or IDH2-directed therapy is another consideration for particularly infirm patients. All patients should be considered for clinical trials.

With the 7+3 regimen (or other similar approaches), 60–80% of younger and 33–60% of older patients (among those who are candidates for intensive therapy) with primary AML achieve CR. Response rates around 60% in first line have been similarly reported with the combination of HMA plus venetoclax in older or infirm patient groups. Induction death is more frequent with advancing age and medical comorbidity, but the most common reason for treatment failure is lack of remission. Patients with refractory disease after inductions should be considered for salvage treatments, preferably on clinical trials. Planning for the possibility of allogeneic hematopoietic stem cell transplantation (HCT) for all eligible patients under age 75 years is part of optimal initial AML care. Typically, allogeneic HCT is performed only for patients who are in CR but at risk for relapse, but fit younger patients with primary refractory disease (not in remission after initial induction) have ~15–20% cure rates with allogeneic HCT after myeloablative conditioning. For this reason, early planning for possible future allogeneic HCT (including human leukocyte antigen [HLA] typing, donor search, etc.) should be part of the initial approach for most AML patients.

POSTREMISSION THERAPY

Induction of a durable first CR (CR1) is critical to long-term survival in AML. However, without further therapy, virtually all CR patients will eventually relapse. Thus, postremission therapy is designed to eradicate residual (typically undetectable) leukemic cells to prevent relapse and prolong survival. As with induction, the type of postremission therapy in AML is selected for each individual patient based on age, fitness, and cytogenetic/molecular risk.

The choice between consolidation with chemotherapy or with transplantation is complex and based on age, risk, and practical considerations. In younger patients receiving chemotherapy, postremission therapy with intermediate- or high-dose cytarabine for two to four cycles is standard practice. Higher doses of cytarabine during postremission therapy appear more effective than

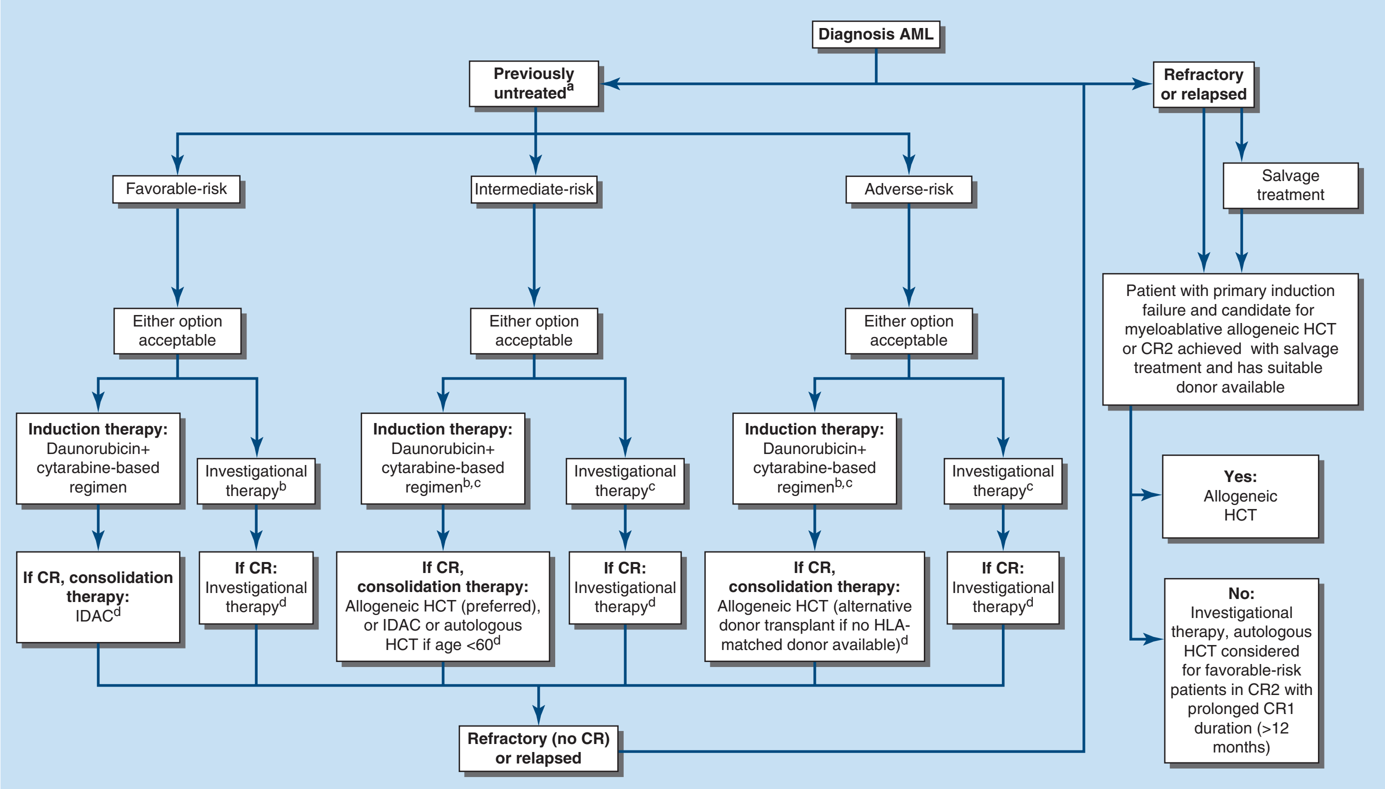
Diagnosis AML
Previously untreated[a]
Refractory or relapsed
Favorable-risk
Intermediate-risk
Adverse-risk
Salvage treatment
Either option acceptable
Either option acceptable
Either option acceptable
Patient with primary induction failure and candidate for myeloablative allogeneic HCT or CR2 achieved with salvage treatment and has suitable donor available
Induction therapy: Daunorubicin+ cytarabine-based regimen
Investigational therapy[b]
Induction therapy: Daunorubicin+ cytarabine-based regimen[b,c]
Investigational therapy[c]
Induction therapy: Daunorubicin+ cytarabine-based regimen[b,c]
Investigational therapy[c]
Yes: Allogeneic HCT
If CR, consolidation therapy: IDAC[d]
If CR: Investigational therapy[d]
If CR, consolidation therapy: Allogeneic HCT (preferred), or IDAC or autologous HCT if age <60[d]
If CR: Investigational therapy[d]
If CR, consolidation therapy: Allogeneic HCT (alternative donor transplant if no HLA-matched donor available)[d]
If CR: Investigational therapy[d]
No: Investigational therapy, autologous HCT considered for favorable-risk patients in CR2 with prolonged CR1 duration (>12 months)
Refractory (no CR) or relapsed

FIGURE 109-2 Algorithm for the therapy of newly diagnosed acute myeloid leukemia (AML). [a]Risk stratification according to the European LeukemiaNet (see Table 109-3). [b]Younger patients (<60–65 years) should routinely be offered investigational therapy on a backbone of standard intensive chemotherapy for induction and consolidation. [c]Older patients, especially those >65 years or with adverse risk disease, or those who are unfit for intensive anthracycline + cytarabine regimens, may be considered for investigational therapy alone or in combination or lower intensity chemotherapy plus venetoclax. Whether venetoclax-containing regimens will be effective in younger, fit patients when compared to intensive chemotherapy is an important question. [d]Novel allogeneic transplantation approaches are preferred for nonfavorable risk when available; other investigational therapy or oral azacitidine (approved in nonfavorable risk) as maintenance should be considered following consolidation.

Allogeneic hematopoietic cell transplantation (HCT) is a consideration for all eligible patients in first complete remission (CR) with non–favorable-risk disease and highly recommended for all older patients (60–75 years) and those with adverse risk.

For all forms of AML in fit patients, except acute promyelocytic leukemia (APL), standard induction therapy includes a regimen based on a 7-day continuous infusion of cytarabine (100–200 mg/m^2/d) and a 3-day course of daunorubicin (60–90 mg/m^2/d) with or without additional drugs. Idarubicin (12 mg/m^2/d) can be used in place of daunorubicin (not shown). The value of postremission/consolidation therapy for older patients (>60 years) who do not have favorable-risk disease is uncertain. Patients who achieve CR undergo postremission consolidation therapy, including sequential courses of intermediate-dose cytarabine, allogeneic HCT, autologous HCT, or novel therapies, based on their predicted risk of relapse (i.e., risk-stratified therapy). Patients receiving induction of lower intensity chemotherapy with venetoclax (or investigational therapy) typically receive repetitive cycles of same on an attenuated schedule, if necessary due to myelotoxicity, after achieving remission. Patients with APL (see text for treatment) usually receive tretinoin and arsenic trioxide–based regimens with or without anthracycline-based chemotherapy and possibly maintenance with tretinoin. HLA, human leukocyte antigen; IDAC, intermediate-dose cytarabine.

standard doses (such as are used in induction) for those who do not have adverse-risk genetics. Studies have shown that the long-standing practice of high-dose cytarabine (3 g/m^2, every 12 h on days 1, 3, and 5) may not improve survival over intermediate-dose cytarabine (IDAC; 1–1.5 g/m^2) for such patients. Thus, the ELN has recommended IDAC at 1–1.5 g/m^2, every 12 h, on days 1–3, as the optimal postremission chemotherapy approach for favorable- and intermediate-risk younger patients, for two to four cycles. While high-dose cytarabine may not be necessary, it is important to note that younger, favorable-risk patients have worse outcomes when doses <1 g/m^2 are used. In contrast to favorable-risk patients, intermediate- or adverse-risk patients should proceed with allogeneic HCT in CR1 when feasible (see transplant discussion below). Because older patients have increased toxicities with higher doses of cytarabine, ELN recommends relatively attenuated cytarabine doses (0.5–1 g/m^2, every 12 h, on days 1–3) in favorable-risk older patients. There is no clear value for intensive postremission therapy in non–favorable-risk older patients; allogeneic HCT in CR1 (up to age 75 years) or investigational postremission therapy is recommended. Indeed, postremission therapy is an appropriate setting for introduction of new agents in both older and younger patients. For older patients (with nonfavorable cytogenetic risk) in CR after intensive therapy who have no transplantation option, maintenance treatment with prolonged low-dose oral azacitidine improves survival.

For patients treated initially with lower intensity regimens that include venetoclax, the current practice is to continue repetitive cycles of the same combination of agents after remission until disease progression. Therapy often must be abbreviated over time due to cumulative myelotoxicity. In the largest published trial with azacitidine and venetoclax in older AML, patients had median survival of <15 months; survival depended on genetic risk. The median duration of remission (including those with incomplete count recovery) was 17.5 months.

Allogeneic HCT is the best relapse-prevention strategy currently available for AML. Allogeneic HCT is best understood as an opportunity for immunotherapy; residual leukemia cells potentially elicit an immunologic response from donor immune cells, the so-called graft-versus-leukemia (GVL) effect. The benefit of GVL in relapse risk reduction unfortunately is offset somewhat by increased morbidity and mortality from complications of HCT including graft-versus-host disease (GVHD). Given that relapsed AML is typically resistant to chemotherapy, allogeneic HCT in CR1 (i.e., before relapse ever occurs) is a favored strategy. We have often explained to patients that transplant can effectively "eliminate the needle in a haystack, but not a stack of needles." Transplant is recommended for patients age <75 years who do not have favorable-risk disease and who have an available HLA-compatible donor (related or unrelated). We recommend allogeneic HCT in CR1 for patients with intermediate-risk disease (Table 109-3). However, considerable debate exists regarding whether allogeneic HCT in CR1 is a requirement for younger patients with intermediate-risk AML, as one large series from the Medical Research Council reported that such patients have similar outcomes if transplanted only after relapse (and achievement of CR2), sparing some the long-term morbidity of transplantation. That said, allogeneic HCT is generally recommended as soon as possible after CR1 is achieved unless the patient is in a favorable-risk group. Patients without HLA-matched donors are considered for alternative donor transplants (e.g., HLA-mismatched unrelated, haploidentical related, and umbilical cord blood) even in CR1. More effective and safe methods of in vivo T-cell depletion (e.g., posttransplant cyclophosphamide following mismatched transplantation or use of abatacept) have broadened the availability of potential allogeneic HCT donors. Now, virtually any patient with a healthy parent or child has an available donor suitable for allogeneic HCT if desired. Long-term outcomes with conventional chemotherapy for older patients are dismal; transplantation for such patients is expanding and improving outcomes.

Trials comparing allogeneic HCT with intensive chemotherapy or autologous HCT have shown improved duration of remission with allogeneic HCT. The relapse risk reduction observed with allogeneic HCT, however, is partially offset by the increase in fatal treatment-related toxicity (GVHD, organ toxicity). Despite this, there is no debate that patients with adverse-risk AML have improved long-term survival with early allogeneic HCT. For rare patients with no allogeneic donor option, high-dose chemotherapy with autologous HCT rescue is another postremission approach in non–adverse-risk subsets. Autologous HCT patients receive their own stem cells (collected during remission and cryopreserved), following administration of myeloablative chemotherapy. The toxicity is relatively low with autologous HCT (5% mortality rate), but the relapse rate is higher than with allogeneic HCT due to the absence of the GVL effect. Favorable- and intermediate-risk patients may benefit from autologous HCT; it is not recommended in adverse-risk patients. Practically speaking, autologous HCT in AML patients is less frequently employed currently due to enhanced relapse risk reduction seen with allogeneic HCT and the growing availability of HLA-mismatched donors (and novel transplantation approaches).

Prognostic factors help to select the appropriate postremission therapy in patients in CR1. Our approach includes allogeneic HCT in CR1 for patients without favorable cytogenetics or genotype. Patients with adverse-risk disease should proceed to allogeneic HCT at CR1 if possible. The decision for allogeneic HCT for younger intermediate-risk patients is complex and individualized as described above; we recommend it. Subsets of patients may benefit from targeted therapy given during remission; emerging data demonstrate survival benefit from incorporation of the FLT3 inhibitors midostaurin or quizartinib, for example, into induction and postremission therapies for patients with *FLT3*-mutated AML. Allogeneic transplantation in CR1 is still recommended for most of these patients.

For patients in morphologic CR, measurement of MRD remains a very important and challenging research area. Cytogenetics are a mainstay of disease assessment, and persistence of abnormal karyotype (despite morphologic CR) is clearly associated with poor clinical outcomes. Immunophenotyping (flow cytometry) to detect minute populations of blasts and/or sensitive molecular assays such as quantitative reverse transcriptase polymerase chain reaction (RT-PCR) to detect AML-associated molecular abnormalities when present (e.g., *NPM1*, *RUNX1/RUNX1T1* and *CBFB/MYH11* transcripts, *PML/RARA*) can be performed to assess MRD at sequential time points during or after treatment. We know that continued detection of MRD after therapy is unfavorable. Whether emerging next-generation sequencing or serial quantitative assessment using flow or PCR, performed during remission, can effectively direct subsequent therapy and improve clinical outcome remains to be determined for most subtypes. Currently, no consensus exists for the optimal MRD measurement technique or its application, although testing is increasingly employed in clinical practice. Data suggest that MRD measurement can in some settings be a reliable discriminator between patients who will continue in CR or relapse, but whether subsequent therapy (i.e., allogeneic HCT or additional therapy) can effectively eradicate disease in such patients is not yet clear. For patients with *NPM1* mutation, MRD-negative status by *NPM1* PCR after two courses of chemotherapy predicts favorable outcome without transplantation (even for patients with coexisting *FLT3* mutation); emerging data demonstrate that transplantation of MRD-positive patients does indeed improve outcome (rather than just predict they will do poorly irrespective of subsequent therapy given). In the subset of patients with APL, serial PCR (for the *PML/RARA* transcript) is a very useful and reliable tool to detect early relapse and to direct initiation of reinduction therapy prior to onset of overt relapse. Critical in the general understanding of MRD in all disease subsets is the recognition that even patients with undetectable levels of MRD remain at risk for leukemic relapse.

SUPPORTIVE CARE

Measures geared to supporting patients through several weeks of neutropenia and thrombocytopenia are critical to successful AML therapy. Patients with AML should be treated in centers expert in providing supportive care. Multi-lumen central venous catheters should be inserted as soon as newly diagnosed AML patients have been stabilized. They should be used thereafter for administration of intravenous medications/chemotherapy and transfusions, as well as for blood drawing instead of venipuncture during prolonged periods of myelosuppression.

Adequate and prompt blood bank support is critical to therapy of AML. Platelet transfusions should be given as needed to maintain a platelet count ≥10,000/μL. The platelet count should be kept at higher levels in febrile patients and during episodes of active bleeding or DIC. Patients with poor posttransfusion platelet count increments may benefit from administration of ABO-matched platelets or platelets from HLA-matched donors. RBC transfusions should be considered to keep the hemoglobin level >70–80 g/L (7–8 g/dL) in the absence of active bleeding, DIC, or congestive heart failure, which may require higher hemoglobin levels. Blood products leukodepleted by filtration should be used to avert or delay alloimmunization as well as febrile reactions. Blood products may also be irradiated to prevent transfusion-associated GVHD.

Neutropenia (neutrophils <500/μL or <1000/μL and predicted to decline to <500/μL over the next 48 h) can be part of the initial presentation and/or a side effect of the chemotherapy treatment in AML patients. Thus, infectious complications remain the major cause of morbidity and death during induction and postremission chemotherapy for AML. Antibacterial (e.g., quinolones) and antifungal (e.g., posaconazole) prophylaxis, especially in conjunction with regimens that cause mucositis, is beneficial. For patients who are herpes simplex virus or varicella-zoster seropositive, antiviral prophylaxis should be initiated (e.g., acyclovir, valacyclovir).

Fever develops in most patients with AML, but infections are documented in only half of febrile patients. Empiric initiation of empirical broad-spectrum antibacterial and antifungal antibiotics has significantly reduced the number of patients dying of infectious complications **(Chap. 79)**. An antibiotic regimen adequate to treat gram-negative organisms should be instituted at the onset of fever in a neutropenic patient after clinical evaluation, including a detailed physical examination with inspection of the indwelling catheter exit site and a perirectal examination (for perirectal abscess), as well as procurement of cultures and radiographs aimed at documenting the source of fever. Specific antibiotic regimens should be based on institutional antibiotic sensitivity data obtained from where the patient is being treated. Acceptable regimens for empiric antibiotic therapy include monotherapy with imipenem-cilastatin, meropenem, piperacillin/tazobactam, or an extended-spectrum antipseudomonal cephalosporin (cefepime or ceftazidime). The combination of an aminoglycoside with an antipseudomonal penicillin (e.g., piperacillin) or an aminoglycoside in combination with an extended-spectrum antipseudomonal cephalosporin should be considered in complicated or resistant cases. Aminoglycosides should be avoided, if possible, in patients with renal insufficiency. Empirical vancomycin should be added in neutropenic patients with catheter-related infections, blood cultures positive for gram-positive bacteria before final identification and susceptibility testing, hypotension or shock, or known colonization with penicillin/cephalosporin-resistant pneumococci or methicillin-resistant *Staphylococcus aureus*. In special situations where decreased susceptibility to vancomycin, vancomycin-resistant organisms, or vancomycin toxicity is documented, other options including linezolid and daptomycin need to be considered.

Caspofungin (or a similar echinocandin), voriconazole, isavuconazonium, or liposomal amphotericin B should be considered for antifungal treatment if fever persists for 4–7 days following initiation of empiric antibiotic therapy. Although liposomal formulations of amphotericin B have improved the toxicity profile of this agent, use has been limited to situations with high risk of or documented mold infections, especially in those in whom an azole fails. Caspofungin has been approved for empiric antifungal treatment. Voriconazole has also been shown to be equivalent in efficacy and less toxic than amphotericin B; isavuconazonium may also be effective with fewer drug-drug interactions. Unfortunately, use of prophylactic or empiric antibiotics contributes to the development of resistance and increased incidence of nosocomial infections such as *Clostridium difficile* colitis, so hospital-wide antibiotic surveillance and isolation strategies should be employed to reduce these complications. Recombinant hematopoietic growth factors have a limited role in AML; myeloid growth factors may be useful in the postremission setting but are not recommended in induction or for "palliative" care for patients not in remission.

TREATMENT FOR REFRACTORY OR RELAPSED AML

In patients who relapse after achieving CR, the length of first CR is predictive of response to salvage chemotherapy treatment; patients with longer first CR (>12 months) generally relapse with drug-sensitive disease and have a higher chance of attaining a CR, even with the same chemotherapeutic agents used for first remission induction. Patients with short prior CR duration are at high risk for treatment failure. Similar to patients with refractory disease, patients with relapsed disease are rarely, if ever, cured by salvage chemotherapy treatments alone. Therefore, patients who eventually achieve a second CR and are eligible for allogeneic HCT should be transplanted. For patients who relapse after allogeneic HCT, no consensus for best therapy exists; outcomes in this setting are very poor.

Because achievement of a second CR with routine salvage therapies is relatively uncommon, especially in patients who relapse rapidly after achievement of first CR (<12 months), these patients and those lacking HLA-compatible donors or who are not candidates for allogeneic HCT should be considered for innovative approaches on clinical trials. Many new agents are in current testing **(Table 109-5)**. The discovery of novel gene mutations and mechanisms of leukemogenesis that might represent actionable therapeutic targets has prompted the development of many new targeting agents. In addition to kinase inhibitors for *FLT3*-mutated AML, other compounds targeting the aberrant activity of mutant proteins (e.g., IDH1/2 inhibitors) and numerous other biologic mechanisms are either approved by the Food and Drug Administration (FDA) or being tested in clinical trials. Inhibitors of FLT3 (gilteritinib, quizartinib), IDH1 (ivosidenib, olutasidenib), or IDH2 (enasidenib) are approved in AML. Exciting early clinical data with menin inhibitors in AML with *KMT2A* rearrangement, *NPM1* mutation, or other genetic aberrations with shared biology of *HOXA* cluster gene upregulation are likely to result in FDA approval of the first agent in this pathway during 2024. Furthermore, approaches with antibodies targeting markers commonly expressed on leukemia blasts (e.g., CD33) or leukemia-initiating cells (e.g., CD123) are also under investigation. Next-generation immune compounds such as bispecific or trispecific antibodies are promising and under study. Investigation of these in combination with other molecular targeting compounds and/or chemotherapy should be pursued.

TABLE 109-5 Novel Therapies in Clinical Development in Acute Myeloid Leukemia (AML)

	UNDER INVESTIGATION	APPROVED BY FOOD AND DRUG ADMINISTRATION SINCE 2017
Kinase inhibitors/ cell signaling	FLT3 inhibitors IRAK-4 inhibitors KIT inhibitors PI3K/AKT/mTOR inhibitors Aurora and polo-like kinase inhibitors, CDK4/6 inhibitors, CDK9 inhibitors, CHK1, WEE1, CSFR1, and MPS1 inhibitors SRC and HCK inhibitors Syk inhibitors	Midostaurin (FLT3) Gilteritinib (FLT3) Quizartinib (FLT3) Pemigatinib (FGFR1)
Epigenetic modulators	Menin inhibitors, other spliceosome modulators DNA methyltransferase inhibitors Histone methylation or acetylation modulators Other spliceosome modulators IDH1 and IDH2 inhibitors DOT1L inhibitors BET-bromodomain inhibitors	Revumenib (menin inhibitor; FDA review pending, 2024) Enasidenib (IDH2) Ivosidenib (IDH1) Olutasidenib (IDH1)
Chemotherapeutic agents	Liposomal preparations Nucleoside analogues	CPX-351 (liposomal cytarabine and daunorubicin) Oral azacitidine
Mitochondrial inhibitors	BH3 mimetics; Bcl-2, Bcl-xL, and Mcl-1 inhibitors Caseinolytic protease inhibitors	Venetoclax (BCL2)
Therapies targeting oncogenic proteins	Fusion transcript targeting EVI1 targeting NPM1 targeting Hedgehog inhibitors	Glasdegib (hedgehog)
Antibodies and immunotherapies	Monoclonal antibodies against CD33, CD44, CD47, CD123, CLEC12A Immunoconjugates Bispecific T-cell engagers (BiTEs) and dual-affinity retargeting molecules (DARTs) for CD33, CD123, others Trispecific T-cell or NK-cell engagers Chimeric antigen receptor (CAR) T cells, genetically engineered T-cell receptor (TCR) T cells, CAR-NK cells Immune checkpoint inhibitors (PD-1/PD-L1, CTLA-4, LAG-3, LILRB4) Vaccines	Gemtuzumab ozogamicin (CD33-toxin)
Therapies targeting AML environment	CXCR4 and CXCL12 antagonists Antiangiogenic therapies	

Source: Reproduced with permission from H Döhner et al: Diagnosis and management of acute myeloid leukemia in adults: 2017 recommendations from an international expert panel. Blood 129:424, 2017.

TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA

APL is a highly curable AML subtype, and ~85% of these patients achieve long-term survival with current approaches. APL has long been shown to be responsive to cytarabine and daunorubicin. However, in the past, patients who were treated with these drugs alone frequently died from DIC induced by the release of granule components by the chemotherapy-treated leukemia cells. The prognosis of APL patients changed dramatically first with the introduction of tretinoin (all-*trans*-retinoic acid [ATRA]) and then with combined ATRA and arsenic trioxide (ATO). ATRA is an oral drug that induces the differentiation of leukemic cells bearing the t(15;17), which disrupts the *RARA* gene encoding retinoid acid receptor. ATRA decreases the frequency of DIC but often produces another complication called the APL (differentiation) syndrome. Occurring within the first 3 weeks of treatment, it is characterized by fever, fluid retention, dyspnea, chest pain, pulmonary infiltrates, pleural and pericardial effusions, and hypoxemia. The syndrome is related to adhesion of differentiated neoplastic cells to the pulmonary vasculature endothelium. Glucocorticoids, chemotherapy for cytoreduction, and/or supportive measures can be effective for management of the APL syndrome. Temporary discontinuation of ATRA is necessary in cases of severe APL syndrome (i.e., patients developing renal failure or requiring admission to the intensive care unit due to respiratory distress). The mortality rate of this syndrome is ~10% if unrecognized. APL syndrome may also occur, less commonly, with ATO.

In adults with low-risk APL (low leukocyte count at presentation), ATRA (45 mg/m^2/d) plus ATO (0.15 mg/kg/d) was compared to ATRA plus concurrent idarubicin chemotherapy. ATRA/ATO was superior and is the new standard of care for such patients. CR rates in low-risk disease approach 100%, with excellent long-term survival. Notably, patients with high-risk APL (defined as leukocyte count >10,000/μL) must be uniquely treated, as they require immediate cytoreduction with chemotherapy due to life-threatening APL syndrome and rapidly rising leukocyte count after initiation of ATRA. High-risk patients are at increased risk for induction death due to this syndrome as well as increased frequency of hemorrhagic complications (related to DIC).

Assessment of residual disease by PCR amplification of the t(15;17) chimeric gene product *PML-RARA* following the final cycle of treatment is important. Disappearance of the signal is associated with long-term disease-free survival; its persistence or reemergence invariably predicts relapse. Sequential monitoring by PCR for *PML-RARA* is now considered standard for postremission monitoring of APL, at least in high-risk patients.

FURTHER READING

Dinardo CD et al: Venetoclax combined with decitabine or azacitidine in treatment-naive, elderly patients with acute myeloid leukemia. Blood 133:7, 2019.

Döhner H et al: Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood 140:1345, 2020.

Issa GC et al: The menin inhibitor revumenib in KMT2A-rearranged or NPM1-mutant leukaemia. Nature 615: 920, 2023.

Jaiswal S, Ebert BL: Clonal hematopoiesis in human aging and disease. Science 366:eaan4673, 2019.

Jongen-Lavrenic M et al: Molecular minimal residual disease in acute myeloid leukemia. N Engl J Med 378:1189, 2018.

Khoury JD et al: The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and histiocytic/dendritic neoplasms. Leukemia. 36:1703, 2022.

Perl AE et al: Gilteritinib or chemotherapy for relapsed or refractory FLT3-mutated AML. N Engl J Med 381:1728, 2019.

Pollyea DA et al: Enasidenib, an inhibitor of mutant IDH2 proteins, induces durable remissions in older patients with newly diagnosed acute myeloid leukemia. Leukemia 33:2575, 2019.

Roboz GJ et al: Ivosidenib induces deep durable remissions in patients with newly diagnosed IDH1-mutant acute myeloid leukemia. Blood 135:463, 2020.

Stone RM et al: Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 mutation. N Engl J Med 377:454, 2017.

110

Chronic Myeloid Leukemia

Hagop Kantarjian, Elias Jabbour

Chronic myeloid leukemia (CML) is a clonal hematopoietic myeloproliferative stem cell neoplasm. The disease is driven by the *BCR::ABL1* chimeric gene that codes for a constitutively active tyrosine kinase, resulting from a reciprocal balanced translocation between the long arms of chromosomes 9 and 22, t(9;22)(q34.1;q11.2), known as the Philadelphia chromosome (Ph) **(Fig. 110-1)**. Untreated, the course of CML is typically biphasic or triphasic, with an early indolent or chronic phase, followed often by an accelerated phase and a terminal blastic phase. Before the era of BCR::ABL1 tyrosine kinase inhibitors (TKIs), the median survival in CML was 3–7 years, and the 10-year survival rate was 30% or less. Introduced into standard CML therapy in 2000, TKIs have revolutionized the treatment, natural history, and prognosis of CML. Today, the estimated 10-year survival rate with imatinib mesylate, the first BCR::ABL1 TKI approved, is >85% and approaches that of a normal age-matched population. Allogeneic hematopoietic stem cell transplantation (HSCT), a curative approach but one that involves more risks, is now offered as later-line therapy after failure of TKIs.

INCIDENCE AND EPIDEMIOLOGY

CML accounts for ~15% of all cases of leukemia. There is a slight male predominance (male-to-female ratio 1.6:1). The median age at diagnosis is 55–65 years. It is uncommon in children; only 3% of patients with CML are younger than 20 years, although in recent years, a higher proportion of young patients have been diagnosed. The incidence of CML increases gradually with age, with a steeper increase after the age of 40–50 years. The annual incidence of CML is 2 cases per 100,000 individuals. In the United States, this translates into about 9000 new cases per year. The incidence of CML has not changed over several decades. By extrapolation, the worldwide annual incidence of CML is about 250,000 cases. With a median survival of 3–6 years before 2000, the disease prevalence in the United States was ~30,000 cases. With TKI therapy, the annual mortality has been reduced from 10–20% to about 1–2%. Therefore, the prevalence of CML is expected to continue to increase. Based on these estimates (incidence of 9000 cases, annual mortality of 2%), the plateau prevalence of CML is estimated to be reached at ~450,000 in the United Stated (9000 × 100/2) by about 2040, with full TKI optimal treatment penetration. The worldwide prevalence will depend on the treatment penetration of TKIs and their effect on reduction of worldwide annual mortality. Ideally, with full TKI treatment penetration, the worldwide prevalence should plateau at 35 times the incidence, or ~9–10 million patients. These estimates are based on extrapolations from the incidence and prevalence of CML in the United States, as well as an estimated annual mortality of 2% with modern TKI therapy; they could vary considerably if the estimates were to change.

ETIOLOGY

There are no familial associations in CML. The risk of developing CML is not increased in monozygotic twins or in relatives of patients with CML. No etiologic agents are incriminated, and no associations exist with exposures to benzene or other toxins, fertilizers, insecticides, or viruses. CML is not a frequent secondary leukemia following therapy of other cancers with alkylating agents and/or radiation. Exposure to ionizing radiation (e.g., nuclear accidents, higher doses of radiation treatment) has increased the risk of CML, which peaks at 5–10 years after exposure and is dose-related. The median time to development of CML among atomic bomb survivors was 6.3 years. Following the Chernobyl accident, no increase in the incidence of CML was reported, suggesting that larger dose exposures of radiation are required to cause CML. Because of adequate protection, the risk of CML has not

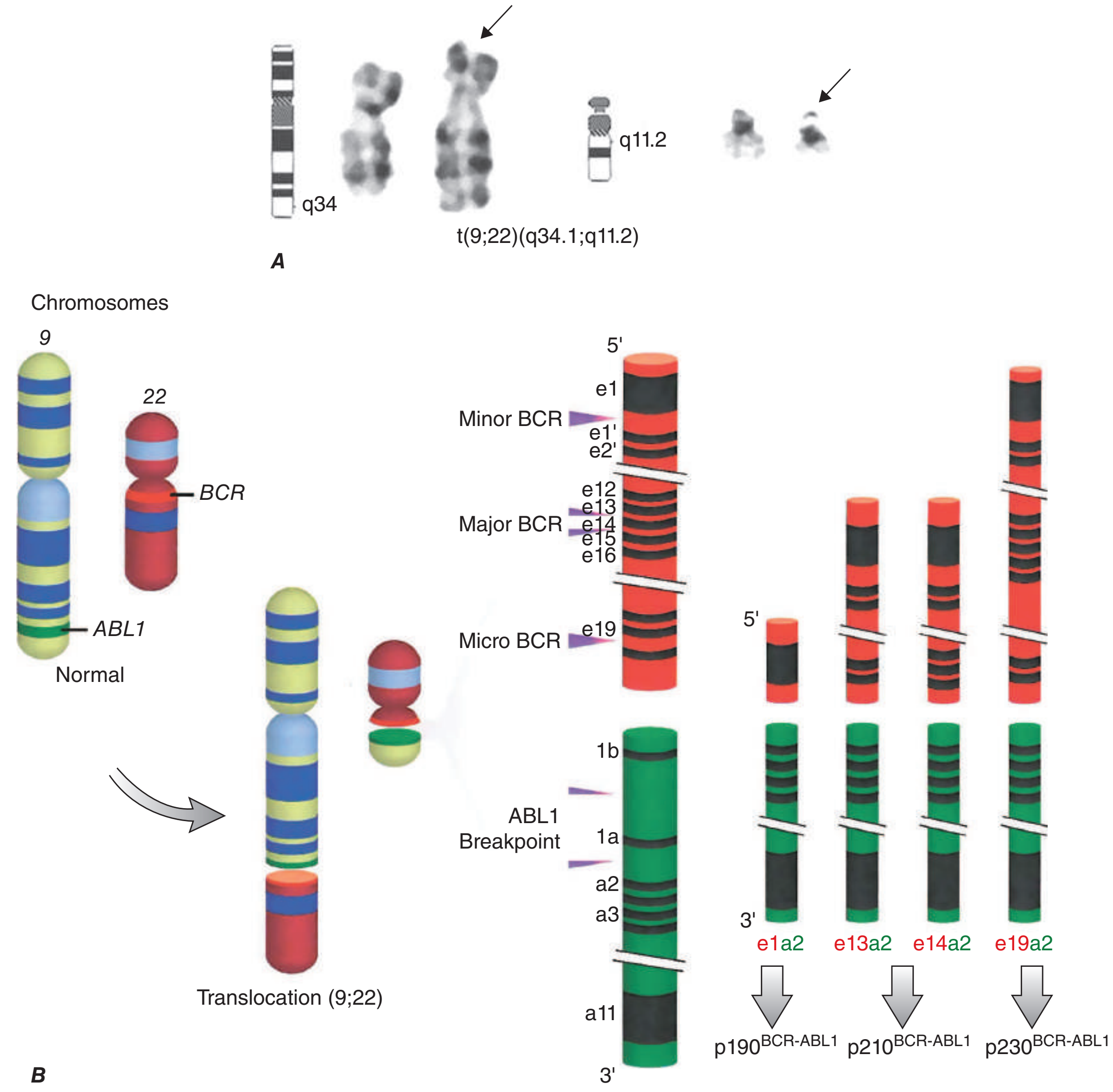

FIGURE 110-1 ***A.*** The Philadelphia (Ph) chromosome cytogenetic abnormality. ***B.*** Breakpoints in the long arms of chromosome 9 (*ABL1* locus) and chromosome 22 (*BCR* regions) result in at least three different BCR::ABL1 oncoprotein messages, p210$^{\text{BCR::ABL1}}$ (most common message in chronic myeloid leukemia [CML]), p190$^{\text{BCR::ABL1}}$ (present in two-thirds of patients with Ph-positive acute lymphoblastic leukemia; rare in CML), and p230$^{\text{BCR::ABL1}}$ (rare in CML and associated with an indolent course). Other rearrangements (e.g., e14a3, e14a3) are less common. *(© 2013 The University of Texas MD Anderson Cancer Center.)*

increased among individuals working in the nuclear industry or among radiologists.

PATHOPHYSIOLOGY

The t(9;22)(q34.1;q11.2) is present in >90% of classical CML cases. It results from a balanced reciprocal translocation between the long arms of chromosomes 9 and 22. It is present in hematopoietic cells (myeloid, erythroid, megakaryocytes, and monocytes; less often mature B lymphocytes; rarely mature T lymphocytes, but not stromal cells), but not in other cells in the human body. As a result of the genetic translocation, DNA sequences from the cellular *ABL* proto-oncogene 1 (*ABL1*) are juxtaposed to the major breakpoint cluster region (*BCR*) gene on chromosome 22, generating a hybrid *BCR::ABL1* oncogene. Depending on the breakpoint site in the major *BCR* region on chromosome 22 (e13 or e14), two main messenger RNA transcripts occur, e13a2 (previously b2a2) and e14a2 (previously b3a2). Both transcripts encode for a novel oncoprotein of molecular weight 210 kDa, referred to as p210$^{\text{BCR::ABL1}}$ (**Fig. 110-1*B***). This oncoprotein exhibits constitutive kinase activity that leads to increased proliferation and reduced apoptosis of CML cells, endowing them with a growth advantage over their normal counterparts. Over time, normal hematopoiesis is suppressed, but normal stem cells can persist and reemerge following effective anti-CML therapy, for example with TKIs. In two-thirds of patients with Ph-positive acute lymphoblastic leukemia (ALL) and in rare cases of CML, the breakpoint in *BCR* is more centromeric, in a region called the minor *BCR* region (*mBCR*). As a result, a shorter sequence of *BCR* is fused to *ABL1*, with a consequent e1a2 transcript and a smaller BCR::ABL1 oncoprotein, p190$^{\text{BCR::ABL1}}$. When occurring in Ph-positive CML, this translocation is associated with a worse outcome. A rarer breakpoint in *BCR* occurs telomeric to the major *BCR* region in the *micro-BCR* (μ-BCR) region. It juxtaposes a larger fragment of the *BCR* gene to *ABL1* and produces an e19a2 transcript and a larger p230$^{\text{BCR::ABL1}}$ oncoprotein (associated with a more indolent CML course). Other rearrangements (based on different breakpoints in the *ABL* region), such as e13a3 or e14a3 (also resulting in a p210$^{\text{BCR::ABL1}}$ oncoprotein), occur much less frequently. These are not readily identifiable nor quantifiable with the routine polymerase chain reaction (PCR) probes, thus producing falsely negative PCR levels on follow-up studies if not tested at diagnosis.

The constitutive activation of *BCR::ABL1* results in autophosphorylation and activation of multiple downstream pathways that affect gene transcription, apoptosis, stromal adherence, skeletal organization, and degradation of inhibitory proteins. These transduction pathways involve RAS, mitogen-activated protein (MAP) kinases, signal transducers and activators of transcription (STAT), phosphatidylinositol-3-kinase (PI3k), MYC, and others. These interactions are mostly mediated through tyrosine phosphorylation and require binding of BCR::ABL1 to adapter proteins such as GRB-2, CRK, CRK-like (CRK-L) protein, and Src homology containing proteins (SHC). Most BCR::ABL1 TKIs (imatinib, dasatinib, bosutinib, nilotinib, ponatinib) bind to the BCR::ABL1 ATP-binding domain, inhibiting its kinase activity, preventing the activation of transformation pathways, and inhibiting downstream signaling. As a result, proliferation of CML cells is inhibited and apoptosis induced, allowing the reemergence of normal hematopoiesis. An additional layer of complexity in CML is related to differences in signal transduction between CML-differentiated cells and early progenitors. Beta-catenin, Wnt1, Foxo3a, transforming growth factor β, interleukin 6, PP2A, SIRT1, and others have been implicated in CML stem cell survival. ABL1 also has a myristoyl site that functions as a negative regulator of its kinase activity. This site and its negative regulatory activity are lost upon fusion with BCR. Asciminib, a novel "third-generation" TKI (third generation refers to novel TKIs that inhibit the *ABL1*-T315I-mutated CML disease, such as ponatinib), is the first-in-class TKI that works through a novel mechanism that specifically targets the ABL1 myristoyl pocket (STAMP is an acronym for specifically targets the ABL1 myristoyl pocket). Asciminib binds this myristoyl site and restores the lost inhibitory activity. Mutations in other cancer-associated genes may also occur at diagnosis, most frequently in *ASXL1*, *IKZF1*, and *RUNX1*. Their significance is being clarified. Some reports suggest that their presence may be associated with worse response to TKI therapy and a higher risk of transformation to blastic phase. An *ASXL1* mutation is often associated with a higher incidence of recurrent cytopenias on TKI therapy, resulting in frequent treatment interruptions, TKI dose reductions, failure to achieve optimal therapeutic milestones, and the need to proceed to allogeneic HSCT.

Experimental models have established the causal relationship between the *BCR::ABL1* rearrangement and the development of CML. In animal models, expression of *BCR::ABL1* in normal hematopoietic cells produced CML-like disorders or lymphoid leukemia, demonstrating the leukemogenic potential of *BCR::ABL1* as a single oncogenic abnormality. Other models, however, suggest the need for a "second hit."

The cause of the *BCR::ABL1* rearrangement is unknown. Molecular techniques that detect *BCR::ABL1* at a level of 1 in 10^8 cells identify this molecular abnormality in the blood of up to 25% of normal adults and 5% of infants, but 0% of cord blood samples. This suggests that *BCR::ABL1* is not sufficient to cause overt CML in the overwhelming majority of individuals in whom it occurs. Because CML develops in only 2 of 100,000 individuals annually, additional molecular events or poor immune recognition of the rearranged cells may contribute to overt CML.

CML is defined by the presence of the *BCR::ABL1* fusion gene in a patient with a myeloproliferative neoplasm. In some patients with a typical morphologic picture of CML, the Ph chromosome is not detectable by standard G-banding karyotype, but fluorescence in situ hybridization (FISH) and/or molecular studies (PCR) detect *BCR::ABL1*. These patients have a course similar to patients with Ph-positive CML and respond to TKI therapy. Many of the remaining patients have atypical morphologic or clinical features and have other diseases, such as atypical CML, chronic myelomonocytic leukemia, and myelodysplastic syndrome/myeloproliferative neoplasms (MDS/MPN). These individuals do not respond to TKI therapy and usually have a poor prognosis with a median survival of ~2–3 years. Detection of mutations in the granulocyte colony-stimulating factor receptor (*CSF3R*) in chronic neutrophilic leukemia (80% of cases) and in some cases of atypical CML (5–10% of cases), mutations in *SETBP1* in atypical CML (25% of cases), and mutations in *SF3B1* in MDS/MPN with ringed sideroblasts and marked thrombocytosis (MDS/MPN-RS-T; 50–70% of cases, associated with longer median survival of 7 years vs 3.3 years with wild-type *SF3B1*) supports the notion that these are distinct molecular and biologic entities. Patients with chronic neutrophilic leukemia or atypical CML whose disease is associated with *CSF3R* mutation may respond to ruxolitinib (a JAK2 inhibitor) therapy (complete response rate 50–60%).

The events associated with the transition of CML from a chronic to the accelerated-blastic phase are poorly understood. Characteristic chromosomal abnormalities such as a double Ph, trisomy 8, isochromosome 17 or deletion of 17p (loss of *TP53*), 20q–, translocations involving 3q26, and others may be noted with disease acceleration. Molecular events associated with transformation include mutations in *TP53*, retinoblastoma 1 (*RB1*), myeloid transcriptions factors like *RUNX1*, and cell cycle regulators like *p16*. A plethora of other mutations or functional abnormalities have been implicated in blastic transformation, but no unifying theme has emerged other than the fact that *BCR::ABL1* itself induces genetic instability that favors the acquisition of additional molecular defects and eventually results in blastic transformation. One critical effect of TKIs is to stabilize the CML genome, leading to a reduced transformation rate. In particular, the previously observed sudden blastic transformations (i.e., abrupt transformation to blastic phase in a patient who had been in cytogenetic response) have become uncommon, occurring rarely in younger patients in the first 1–2 years of TKI therapy (usually sudden lymphoid blastic transformation). Sudden blastic transformations beyond the third year of TKI therapy are rare in patients who continue on TKI therapy. The incidence of any CML blastic transformation is now reduced to 5–6% at 10 years. Moreover, the course of CML is now frequently more indolent in patients treated with TKI, even without cytogenetic response, compared to previous experience with hydroxyurea/busulfan, suggesting a definite clinical benefit of continued inhibition of the kinase activity.

Among patients developing resistance to TKIs, several resistance mechanisms have been observed. The most clinically relevant one is the development of *ABL1* kinase domain mutations that may prevent the binding of TKIs to the catalytic site (ATP-binding site) of the kinase or maintain the kinase activity despite the presence of a TKI. More than 100 *ABL1* kinase domain mutations have now been described, many of which confer relative or absolute resistance to imatinib. Consequently, second-generation (i.e., dasatinib, nilotinib, bosutinib) and third-generation TKIs (ponatinib, asciminib, olverembatinib [approved in China; ongoing trials worldwide since 2023]) were developed. All three third-generation TKIs demonstrate significant efficacy against T315I, a "gatekeeper" mutation that prevents binding of, and causes resistance to, imatinib and the second-generation TKIs (dasatinib, bosutinib, nilotinib).

CLINICAL PRESENTATION

The presenting signs and symptoms in CML depend on the availability of and access to health care, including physical examinations and screening tests. In developed countries, because of the wider access to health care screening and physical examinations, 50–60% of patients are diagnosed on routine blood tests and have minimal symptoms at presentation, such as fatigue. In geographic locations where access to health care is more limited, patients often present with high CML disease burden including splenomegaly, anemia, and related symptoms (abdominal pain, weight loss, fatigue), associated with a higher frequency of high-risk CML. Presenting findings in patients diagnosed in the United States are shown in **Table 110-1**.

Symptoms Most patients with CML (90%) present in the indolent or chronic phase. Depending on the timing of diagnosis, patients are often asymptomatic (if the diagnosis is discovered during health care screening tests). Common symptoms, when present, are manifestations of anemia and splenomegaly. These include fatigue, malaise, weight loss (if high leukemia burden), or early satiety and left upper quadrant pain or masses (from splenomegaly). Less common presenting findings include thrombotic or hyperviscosity-related events from severe leukocytosis or thrombocytosis. These include priapism, cardiovascular complications, myocardial infarction, venous thrombosis, visual disturbances, dyspnea and pulmonary insufficiency, drowsiness, loss of coordination, confusion, or cerebrovascular accidents. Manifestations of bleeding diatheses include retinal hemorrhages, gastrointestinal bleeding, and others. Patients who present with, or progress to, the accelerated or blastic phases frequently have additional symptoms including unexplained fever, significant weight loss, severe fatigue, bone and joint pain, bleeding and thrombotic events, and infections.

TABLE 110-1 Presenting Signs and Symptoms of Newly Diagnosed Philadelphia Chromosome–Positive Chronic Myeloid Leukemia in Chronic Phase

PARAMETER	PERCENTAGE
Age ≥60 years (median)	40–50 (55–65)
Female gender	35–45
Splenomegaly	30
Hepatomegaly	5–10
Lymphadenopathy	5
Other extramedullary disease	2
Hemoglobin <10 g/dL	10–15
Platelets	
>450 × 10^9 cells/L	30–35
<100 × 10^9 cells/L	3–5
White blood cells ≥50 × 10^9 cells/L	35–40
Marrow	
≥5% blasts	5
≥5% basophils	10–15
Peripheral blood	
≥3% blasts	8–10
≥7% basophils	10
Additional chromosomal abnormalities (other than the Philadelphia chromosome)	4–5
Sokal risk	
Low	60–65
Intermediate	25–30
High	10

Physical Findings Splenomegaly is the most common physical finding, occurring in 20–70% of patients depending on health care screening frequency. Less common findings include hepatomegaly (5–10%), lymphadenopathy (5%), and extramedullary disease (skin or subcutaneous lesions). The latter indicates CML transformation if a biopsy confirms predominance of blasts. Other physical findings are manifestations of complications of high tumor burden described earlier (e.g., cardiovascular, cerebrovascular, bleeding). High basophil counts may be associated with histamine overproduction causing pruritus, diarrhea, flushing, and even gastrointestinal ulcers.

Hematologic and Marrow Findings In untreated CML, leukocytosis ranging from 10–500 × 10^9/L is common. The peripheral blood differential shows left-shifted hematopoiesis with predominance of neutrophils and the presence of bands, myelocytes, metamyelocytes, promyelocytes, and blasts (usually ≤5%). Basophils and/or eosinophils are frequently increased. Thrombocytosis is common, but thrombocytopenia is rare and, when present, suggests a worse prognosis, disease acceleration, or an unrelated etiology. Anemia is present in one-third of patients. Cyclic oscillations of counts are noted in 10–20% of patients without treatment. Biochemical abnormalities include a low leukocyte alkaline phosphatase score and high levels of vitamin B_{12}, uric acid, lactic dehydrogenase, and lysozyme. The presence of unexplained and sustained leukocytosis, with or without splenomegaly, should lead to a marrow examination and cytogenetic analysis.

The bone marrow is hypercellular with marked myeloid hyperplasia and a high myeloid-to-erythroid ratio of 15–20:1. Marrow blasts are typically 5% or less; when higher, they carry a worse prognosis or represent transformation to accelerated phase (if they are ≥15%). Increased reticulin fibrosis (detected with silver stain) is common, with 30–40% of patients demonstrating grade 3–4 reticulin fibrosis. This was considered adverse in the pre-TKI era. With TKI therapy, reticulin fibrosis resolves in most patients and is not an indicator of poor prognosis. Collagen fibrosis (Wright-Giemsa stain) is rare at diagnosis. Disease progression with a "spent phase" of myelofibrosis (myelophthisis, or burnt-out marrow) was a common end-stage CML condition with busulfan therapy (20–30%); it is extremely rare now with TKI therapy.

Cytogenetic and Molecular Findings The diagnosis of CML is straightforward and depends on documenting the translocation t(9;22)(q34.1;q11.2), which is identified by G-banding in 90% of cases. This is known as the Philadelphia chromosome (initially identified in Philadelphia as a minute chromosome; later identified to be chromosome 22) (Fig. 110-1). Some patients (~10%) may have complex translocations (complex variant Ph) involving three or more chromosomes including chromosomes 9 and 22 and one or more additional chromosomes. Others may have a "masked Ph," involving translocations between chromosome 9 and a chromosome other than 22 (but molecularly showing the *ABL1* rearrangement; known as simple variant Ph). The prognosis of these patients and their response to TKI therapy are like those in patients with Ph. Translocation (9;22)(q34;p13)/*ETV6::ABL1* is now classified as a myeloproliferative neoplasm-eosinophilia and may respond better to second-generation TKIs. About 5–10% of patients may have additional chromosomal abnormalities (ACAs) in the Ph-positive cells at diagnosis. These usually involve trisomy 8, a double Ph, isochromosome 17 or 17p deletion, 20q–, or others. This was historically a sign of adverse prognosis, particularly when trisomy 8, double Ph, or chromosome 17 abnormalities were noted. A less common abnormality involving chromosome 3q26.2 occurs with disease progression and carries a poor prognosis.

Techniques such as FISH and PCR are now used to aid in the diagnosis of CML. They are more sensitive to estimate the CML burden in patients on TKI therapy. They can be done on peripheral blood and thus are more convenient to patients. Patients with CML at diagnosis should have a FISH analysis to quantify the percentage of Ph-positive cells, if FISH is used to replace marrow cytogenetic analysis in monitoring response to therapy. FISH will not detect additional chromosomal abnormalities; thus, a cytogenetic analysis is recommended at the time of diagnosis. In addition, 10–15% of patients may develop chromosomal abnormalities in Ph-negative metaphases after responding to TKIs. These abnormalities may carry a worse prognosis but are not detected by FISH unless already identified and FISH is used to follow them. Molecular studies at diagnosis are important to document the type and presence of *BCR::ABL1* transcripts to avoid spurious "undetectable" *BCR::ABL1* transcripts on follow-up studies, with the false impression of a complete molecular response. The presence of the Philadelphia chromosome with "negative" PCR with standard methodology should prompt investigation of atypical transcripts (e13a3, e14a3, e19a2, others).

Both FISH and PCR studies can be falsely positive at low levels or falsely negative because of technical issues. Therefore, a diagnosis of CML must always rely on a marrow analysis with routine cytogenetics. The diagnostic bone marrow confirms the presence of the Ph chromosome, detects additional chromosomal abnormalities, and quantifies the percentage of marrow blasts and basophils. In 10% of patients, the percentage of marrow blasts and basophils can be significantly higher than in the peripheral blood, conferring poorer prognosis or even representing disease transformation.

Monitoring patients on TKI therapy by cytogenetics, FISH, and PCR has become an important standard practice to assess response to therapy, emphasize compliance, evaluate possible treatment resistance, identify the need to change TKI therapy, and determine the need to assess for kinase domain mutations. Because of the decreasing reliance of bone marrow aspirations to monitor response, equivalence has been established to correlate cytogenetic results with PCR values. These are not absolute correlations but provide adequate guidance. A partial cytogenetic response is defined as the presence of 35% or less Ph-positive metaphases by routine cytogenetic analysis. This is roughly equivalent to *BCR::ABL1* transcripts on the International Scale (IS) of 10% or less. A complete cytogenetic response refers to the absence of Ph-positive metaphases (0% Ph positivity). This is approximately equivalent to *BCR::ABL1* transcripts (IS) ≤1% (MR2). A major molecular response (MMR or MR3) refers to *BCR::ABL1* transcripts (IS) ≤0.1%, or roughly a 3-log or greater reduction of *BCR::ABL1* transcripts from a standardized baseline. MR4 (deep molecular response; DMR) refers to *BCR::ABL1* transcripts (IS) ≤0.01%, and MR4.5 refers to *BCR::ABL1* transcripts (IS) ≤0.0032%, roughly equivalent to a 4.5-log reduction or greater of transcripts.

Findings in CML Transformation Progression of CML is usually associated with leukocytosis resistant to therapy, increasing anemia, fever and constitutional symptoms, and increased blasts and basophils in the peripheral blood or marrow. Criteria of accelerated-phase CML, historically associated with median survival of <2 years, include the presence of 15% or more peripheral blasts, 30% or more peripheral blasts plus promyelocytes, 20% or more peripheral basophils, cytogenetic clonal evolution (presence of additional chromosomal abnormalities other than Ph), and thrombocytopenia $<100 \times 10^9$/L (unrelated to therapy). About 5–10% of patients present with de novo accelerated phase or blastic phase. The prognosis of de novo accelerated phase with TKI therapy has improved, with an estimated 8-year survival rate of 60–70%. The median survival of accelerated phase evolving from chronic phase has also improved from a historical median survival of 18 months to an estimated 3-year survival rate of 50% on TKI therapy. Therefore, the criteria for accelerated-phase CML should be revisited because most clinical criteria defining accelerated phase have lost much of their prognostic significance. The newest World Health Organization (WHO) classification suggested eliminating accelerated-phase CML as an entity and classifying such patients as high-risk CML. However, the survival of patients with both de novo and evolved accelerated-phase CML is significantly worse than in high-risk CML. The accelerated-phase definition should be maintained in order to recognize such patients and treat them differently than chronic-phase CML, either with existing standards of care (combinations of TKIs with other agents, allogeneic HSCT) or on investigational trials of novel TKIs or other modalities. Blastic-phase CML is defined by the presence of 30% or more peripheral or marrow blasts or the presence of sheets of blasts in extramedullary disease (usually skin, soft tissues, or lytic bone lesions). Blastic-phase CML is commonly myeloid (60%) but can present uncommonly as erythroid, promyelocytic, monocytic, or megakaryocytic. Lymphoid blastic phase occurs in about 25% of patients. Lymphoblasts are terminal deoxynucleotide transferase positive and peroxidase negative (although occasionally with low positivity up to 3–5%) and express lymphoid markers (CD10, CD19, CD20, CD22). They also often express myeloid markers (50–80%), resulting in diagnostic challenges. Proper immunophenotypic diagnosis is important because lymphoid blastic-phase CML is responsive to anti-ALL-type chemotherapy (e.g., hyper-CVAD [cyclophosphamide, vincristine, doxorubicin, and dexamethasone]) in combination with TKIs and immunotherapies (complete response rate 70%; median survival 3 years; high rates of bridging to allogeneic HSCT and possible cure).

■ PROGNOSIS AND CML COURSE

Before the TKI era, the annual mortality in CML was 10% in the first 2 years and 15–20% thereafter. The median survival in CML was 3–7 years (with hydroxyurea-busulfan and interferon α). Without a curative option of allogeneic HSCT, the course of CML was toward transformation to, and death from, accelerated or blastic phases for most patients. Even disease stability was unpredictable, with some patients demonstrating sudden transformation to a blastic phase. With imatinib therapy, the annual mortality rate in CML has decreased to 1–2% in the first 20 years of observation. The 10-year survival rate is about 85%, the 10-year CML-specific survival rate is 90%, and CML-specific mortality rate is 10%; the 10-year incidence of blastic transformation is 5–6%. More than half of the deaths are from conditions other than CML, such as old age, comorbidities, accidents, suicides, other cancers, and other medical conditions (e.g., infections, surgical procedures) **(Fig. 110-2)**. The course of CML has also become quite predictable. In the first 2 years of TKI therapy, rare sudden transformations are still reported (1–2%), usually lymphoid blastic transformations that respond to combinations of chemotherapy and TKIs followed by allogeneic HSCT. These may be explained by the intrinsic mechanisms of sudden transformation already existing in the CML clones before the start of therapy that were not amenable to TKI inhibition, in particular imatinib. Second-generation TKIs (nilotinib, dasatinib, bosutinib) used as frontline therapy have reduced the incidence of transformation in the first 2–3 years from 4–6% with imatinib to 1–2% with second-generation TKIs. Disease transformation to accelerated or blastic phase is rare with continued TKI therapy, estimated at <1% annually in years 4–10 of follow-up. Patients usually develop resistance in the form of cytogenetic relapse, followed by hematologic relapse and subsequent transformation, rather than the previously feared sudden transformation without the warning signals of cytogenetic-hematologic relapse.

Before the imatinib era, several pretreatment prognostic factors predicted for worse outcome in CML and have been incorporated into prognostic models and staging systems. These have included older age, significant splenomegaly, anemia, thrombocytopenia or thrombocytosis, high percentages of blasts and basophils (and/or eosinophils), marrow fibrosis, additional chromosomal abnormalities, and others. The introduction of TKIs into CML therapy has decreased or eliminated the prognostic impact of these prognostic factors and the significance of the CML models (e.g., Sokal, Hasford, European Treatment and Outcome Study [EUTOS]). Treatment-related prognostic factors have emerged as the most important prognostic factors in the era of imatinib therapy. Achievement of complete cytogenetic response (MR2) has become the major therapeutic endpoint and is the only endpoint associated with improvement in survival. Achievement of MMR (MR3) is associated with decreased risk of events (relapse) and CML transformation but not with survival prolongation among patients with complete cytogenetic response. This may be due to the efficacy of salvage TKI therapies, which are implemented at the first evidence of cytogenetic relapse. Achievement of durable DMR (MR4/MR4.5) may offer the possibility of treatment-free remission (TFR). A durable DMR for 2+ years is associated with a TFR rate of 50%; a durable DMR for 5+ years is associated with a TFR rate of 80%+. A TFR may allow a temporary TKI therapy interruption in women pursuing pregnancy. The lack of achievement of MMR or DMR should not be considered as "failure" of a particular TKI therapy and/or an indication to change the TKI or to consider allogeneic HSCT.

Long-term updates of randomized trials suggest that second-generation TKIs and imatinib are similarly effective in lower-risk CML; second-generation TKIs may offer a therapeutic advantage in high-risk CML and when TFR is an important treatment endpoint (younger patients).

TREATMENT

Chronic Phase Chronic Myeloid Leukemia

Since 2001, six oral BCR::ABL1 TKIs have been approved by the U.S. Food and Drug Administration (FDA) for the treatment of CML. These include imatinib (Gleevec, Glivec), nilotinib (Tasigna), dasatinib (Sprycel), bosutinib (Bosulif), ponatinib (Iclusig), and asciminib (Scemblix). Dasatinib, nilotinib, and bosutinib are referred to as second-generation TKIs; ponatinib and asciminib are referred to as third-generation TKIs, a term also used generally for the more recently developed TKIs that are active against T315I-mutated CML (asciminib is also referred to as a STAMP inhibitor because of its different mechanism of action). Nilotinib is similar in structure to imatinib but 30 times more potent. Dasatinib and bosutinib inhibit the *SRC* family of kinases in addition to *ABL1*, with dasatinib reported to be 300 times more potent and bosutinib 30–50 times more potent than imatinib **(Table 110-2)**. Ponatinib inhibits vascular endothelial growth factor receptor (VEGFR), which may be partly responsible for the high incidence of hypertension observed with this agent (Table 110-2).

ESTABLISHED AND EVOLVING THERAPEUTIC CONCEPTS IN CML

In the early days of TKI development, the primary aim in CML therapy was to improve survival, as the therapeutic miracle of

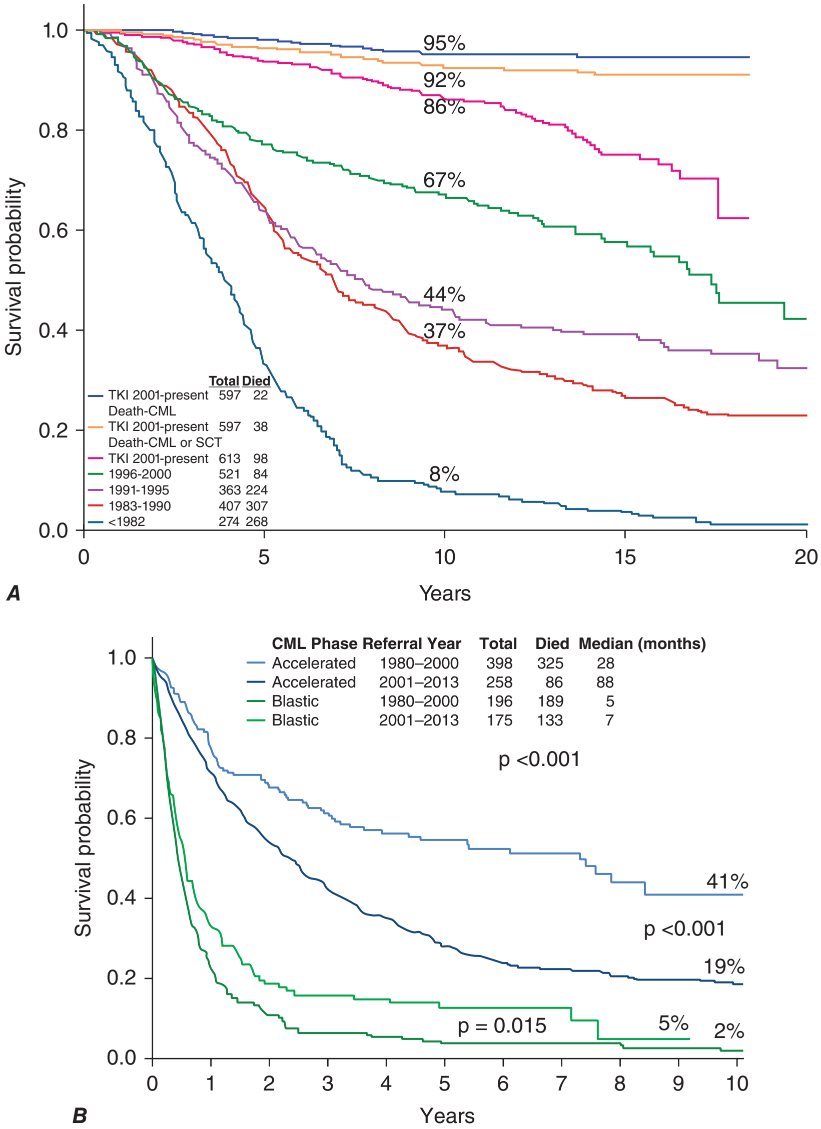

FIGURE 110-2 ***A.*** Survival in newly diagnosed chronic-phase chronic myeloid leukemia (CML) by era of therapy (MD Anderson Cancer Center experience from 1965 to present). Top blue curve is survival with tyrosine kinase inhibitors (TKIs), accounting for only CML-related deaths. The orange curve (second from top) accounts for deaths related to CML or CML treatment complications (e.g., deaths following allogeneic hematopoietic stem cell transplant [HSCT]). The red curve (third from top) is survival including all deaths regardless of causality (old age, car accidents, suicide, gun shots, second cancers, complications of unrelated surgeries, infections, others). The difference in the denominators, 613 minus 597 cases, is because 16 deaths were from unknown/undocumented causes (outside MD Anderson and no good tracking for cause of death). ***B.*** Survival in patients with accelerated- and blastic-phase CML referred to MD Anderson Cancer Center by era of therapy, demonstrating the significant survival benefit in the TKI era in accelerated-phase CML but the modest benefit in blastic-phase CML. Referred cases included de novo and post-chronic-phase transformations.

TABLE 110-2 Medical Therapeutic Options in Chronic Myeloid Leukemia

AGENT (BRAND NAME)	APPROVED INDICATIONS	DOSE SCHEDULE	NOTABLE TOXICITIES
Imatinib mesylate (Gleevec)	All phases	400 mg daily	See text
Dasatinib (Sprycel)	All phases	First-line: 100 mg daily Salvage: 100 mg daily in chronic phase; 140 mg daily in transformation	Myelosuppression; pleural and pericardial effusions; pulmonary hypertension
Nilotinib (Tasigna)	All phases except blastic phase	First-line: 300 mg twice daily Salvage: 400 mg twice daily	Diabetes; arterio-occlusive events; pancreatitis
Bosutinib (Bosulif)	All phases	First line: 400 mg daily Salvage: 500 mg daily	Diarrhea; liver toxicity; renal dysfunction
Ponatinib (Iclusig)	T315I mutation; failure of ≥2 tyrosine kinase inhibitors	45 mg daily (may consider lower starting doses, e.g., 30 mg daily; lower the dose to 15 mg daily once a complete cytogenetic response is achieved)	Skin rashes (10–20%); pancreatitis (5%); arterio-occlusive events (10–20%); systemic hypertension (10–15%)
Asciminib (Scemblix)	Third-line therapy; T315I mutation	40 mg twice daily or 80 mg daily; T315I: 200 mg twice daily	Arterial occlusive events; hypertension; (others?)
Omacetaxine mepesuccinate (Synribo)	Failure ≥2 tyrosine kinase inhibitors	1.25 mg/m^2 subcutaneously twice daily for 14 days of induction; 7 days of maintenance every month (consider shorter dose schedules, 7 days of induction, 2–5 days of maintenance)	Myelosuppression

near-normal survival with TKI therapy was never anticipated. It was thought that CML cells would develop resistance mechanisms after a period of TKI exposure. This happened, as CML was the first tumor demonstrating the development of mutations in the ABL1 kinase domain that prevented the binding of the TKIs. However, the true resistance rate was only 10% after 10 years of imatinib exposure. The TKIs were developed at a dose level below the maximum tolerated dose (MTD; traditional development of chemotherapy agents in cancer), based on the toxicities noted in the first one to two courses or few months of therapy. As the success of therapy required treatment continuation for years, or even a patient's lifetime, additional unanticipated toxicities were observed with long-term exposures. This shifted the concept to develop these targeted therapies in CML (and in other cancers) at an "optimal biologic dose" (OBD; a dose that presumably maintains the same efficacy but reduces toxicities). Such OBDs were derived initially from clinical experience in CML, although ongoing trials are basing them on translational studies of the drug exposure and target modulation, compared with preclinical research findings. This led, in CML, to study dasatinib 50 mg daily as frontline therapy, which proved to be as effective and less toxic than 100 mg daily. Similar results were reported with other TKIs where dose adjustments after achieving a good molecular response maintained efficacy. Other early concepts included the following: (1) changing a TKI if evidence of toxicity (rather than lowering the TKI dose, with the concern being that lowering the TKI dose will compromise efficacy and cause resistance); (2) changing a TKI if the National Comprehensive Cancer Network (NCCN)/European Leukemia Network (ELN) landmark-defined milestones of "failure" or "warning" were noted; and (3) changing TKI therapy in a responding patient (MMR or even MR4), in order to achieve a complete molecular response and aim for a TFR status. As experience matured, it was noted that such strategies may be harmful, as they increased the cost of therapy and increased the risk of potential additional side effects, without benefiting patients. At present, therapy of CML may consider lower-dose schedules of TKIs as frontline or later-line therapy; reducing the TKI dose once a good molecular response (≥MR2) is achieved; not changing TKI therapy in patients in good molecular response; and perhaps not changing TKI therapy based on the NCCN/ELN milestones (as the long-term follow-up data showed survival to be favorable even among patients who did not achieve these molecular milestones). These concepts will be discussed further.

At the beginning of the TKI experience, survival prolongation was the primary goal of therapy, but once near-normal survival was accomplished, investigators started addressing additional treatment endpoints. Thus, a second treatment goal was whether CML is curable (defined, as in other cancers, as the ability to stop TKI therapy without disease recurrence after several years of observation). This led to studies aimed at stopping TKI therapy after 2–5 years of durable DMR and achieving a TFR status. The third treatment goal is to make TKI therapies available and affordable to all patients (rather than the few who can afford them), a problem addressed with the availability of generic TKIs. The fourth treatment goal is to minimize the early and long-term toxicities (currently addressed with studies investigating the OBDs of TKIs).

Frontline Therapy of CML Imatinib, dasatinib, bosutinib, and nilotinib are all acceptable frontline therapies in CML. The long-term results of imatinib are very favorable. The 10-year cumulative rate of MMR is 90%, and of DMR 80%. The 10-year survival rate was 82%, and 10-year relative (compared with age-matched population) survival rate 92%. About 25% had to change to second-generation TKIs, 10% because of resistance to imatinib and 15% for other reasons (adverse events). The 10-year incidence of blastic-phase CML was 5.8%. In multiple randomized studies, (e.g., ENESTnd, DASISION, BFORE), the second-generation TKIs resulted in better outcomes in early surrogate endpoints (higher rates of MMR and MR4.5; lower rates of blastic transformation). However, none showed survival benefit. This may be because the rate of complete cytogenetic response (MR2) was ultimately similarly high with imatinib versus second-generation TKIs and because later-line salvage TKI therapy (following close observation and treatment change at progression) provided highly effective therapy; this ensured adequate long-term outcome despite resistance or intolerance to initial imatinib therapy.

The choice of TKI frontline CML therapy depends on several factors: (1) the treatment aims (survival, TFR), which are tightly linked to age; (2) the TKI cost and affordability (generic vs patented TKIs); and (3) the patient comorbidities (e.g., avoid dasatinib if history of lung injury or chronic lung disease; avoid nilotinib if history of pancreatitis, diabetes mellitus, or arterio-/veno-occlusive spastic events; avoid bosutinib if history of liver or renal dysfunction of gastrointestinal problems [colitis, diverticulitis]). In general, if survival is the treatment aim in CML, all four TKIs achieve similar outcomes. However, if TFR is the treatment aim, then second-generation TKIs may be preferred. Second-generation TKIs may also be preferred in high-risk chronic-phase CML. Two additional important aims include the comparative long-term toxicities and the TKI treatment value.

The frontline TKI dose schedules are as follows: imatinib 400 mg orally daily; dasatinib 100 mg orally daily; bosutinib 400 mg orally daily (use dose escalation of 100–200 mg daily for 1–2 weeks, 300 mg daily for 2 weeks, then 400 mg daily to avoid the early self-limited gastrointestinal toxicity [diarrhea, nausea, and vomiting]); and nilotinib 300 mg orally twice a day (on an empty stomach). Dasatinib 50 mg orally daily is as effective in frontline therapy as 100 mg daily and significantly less toxic.

A recent randomized trial compared asciminib to investigator's TKI of choice in frontline CML therapy. The 12-month MMR was significantly higher with asciminib versus other TKIs (68% vs 49%; p<.001) and with asciminib versus imatinib (69% vs 40%); p<.001), but not with asciminib versus second-generation TKIs (66% vs 58%). Whether the higher MMR rate will translate into better long-term survival or TFR rates, lower long-term side-effects, or better treatment value is an open question.

Management of TKI Toxicities In the first 10–15 years of the TKI experience, it was common practice to change TKI therapy when toxicities occurred because of the erroneous assumption that reducing the TKI dose may reduce efficacy. This was not borne out in the long-term experience, particularly in patients who are in good molecular response. Among such patients, the TKI dose can be reduced for mild-moderate or even severe reversible toxicities. Imatinib can be reduced to 100–300 mg daily, dasatinib to 20–50 mg daily, nilotinib to 200 mg daily or 150 mg twice daily, bosutinib to 100–300 mg daily, and ponatinib to 15–30 mg daily, depending on the toxicities and molecular response.

Side effects of TKIs are generally mild to moderate, although with long-term TKI therapy, they could affect the patient's quality of life. Serious side effects occur in <5–10% of patients. With imatinib therapy, common mild to moderate side effects include fluid retention, weight gain, nausea, diarrhea, skin rashes, periorbital edema, bone or muscle aches, fatigue, and others (rates of 10–20%). In general, second-generation TKIs are associated with lower rates of these bothersome adverse events. However, dasatinib 100 mg daily is associated with higher rates of myelosuppression (20–30%), particularly thrombocytopenia; pleural (10–25%) or pericardial effusions (≤5%); and pulmonary hypertension (<5%). A lower dose of dasatinib (50 mg daily instead of 100 mg daily) used in frontline CML therapy resulted in similar efficacy and a lower incidence of serious side effects (pleural effusions <5%, myelosuppression <10%). Nilotinib is associated with higher rates of hyperglycemia (10–20%), pruritus and skin rashes, hyperbilirubinemia (typically among patients with Gilbert's syndrome and mostly of no clinical consequences), and headaches. Nilotinib is also associated with occasional instances of pancreatitis (<5%). Nilotinib 300–400 mg twice daily is associated with a 10-year cumulative incidence of cardiovascular complications of 25–35%. Bosutinib is associated with higher rates of liver toxicity, renal dysfunction, and early and self-limited gastrointestinal adverse events, particularly diarrhea. Occasionally, the gastrointestinal symptoms mimic chronic severe enterocolitis, which reverses with treatment discontinuation. Ponatinib 45 mg daily is associated with higher rates of serious skin rashes (10–15%), pancreatitis (10%), elevations of amylase/lipase (10%), and systemic hypertension (50–60%; severe in 20%). Arterio-occlusive events (cardiovascular, cerebrovascular, and peripheral arterial) have been reported with most TKIs. The incidence appears to be highest with ponatinib, but both nilotinib and dasatinib are associated with these events at an incidence significantly higher than imatinib or bosutinib. Among the TKIs, imatinib and bosutinib are associated with the lowest incidence of cardiovascular events.

With long-term follow-up, rare but clinically relevant serious toxicities are emerging. Renal dysfunction and occasionally renal failure (creatinine elevations >2–3 mg/dL) are observed in 2–3% of patients, more frequently with imatinib and bosutinib than other TKIs, and usually reverse with TKI discontinuation and/or dose reduction. Rarely, patients may develop TKI-related peripheral neuropathy or even central neurotoxicities that are misdiagnosed as dementia or Alzheimer's disease; these may reverse slowly after TKI discontinuation.

Some toxicities are prohibitive and require a change of TKI therapy: recurrent pleural effusions (most commonly with dasatinib; least with imatinib and nilotinib; responsive to a short course of steroids); vasospastic or vaso-occlusive events (cerebrovascular accidents, myocardial infarction or unstable angina; more common with ponatinib and nilotinib; least with imatinib and bosutinib); pulmonary hypertension (1–2% with dasatinib, but can occur with other TKIs; slowly reversible with a short course of steroids and sildenafil citrate), pancreatitis (2% with nilotinib, 2–4% with ponatinib); neurologic problems (dementia-like, parkinsonism; rare and slowly reversible with TKI discontinuation); immune-mediated events (pneumonitis, myocarditis, pericarditis, hepatitis, nephritis; usually reversible with TKI discontinuation and a short course of high-dose steroids [e.g., methylprednisolone 50 mg twice daily for 3–5 days]); and severe colitis (bosutinib; reversible with discontinuation).

When switching TKI therapy for prohibitive toxicity, the dose of the new TKI does not have to be the dose recommended for "failure" (a term that historically encompassed resistance and intolerance). Many of these patients are already in good molecular response, and the dose of the new TKI used for previous TKI intolerance can be lower, particularly if the patient is already in ≥MR2: dasatinib 20–50 mg daily; bosutinib 100–300 mg daily; nilotinib 200 mg daily or 150 mg twice a day; ponatinib 15–30 mg daily.

TKI cross-intolerance may be more common than previously thought. Because the TKIs have different chemical structures, cross-intolerance was thought to be uncommon. With experience, it appears that patients who have intolerance to one TKI may more often have intolerance to others, with the intolerance/side effect manifesting as the same or as a different one.

Discontinuation of TKIs and TFR Several studies have confirmed that TKI discontinuation among patients who achieve DMR (MR4) for longer than 2–3 years can result in TFR rates of 40–60%. Discontinuation of TKI therapy after 5+ years of DMR is associated with TFR rates of 80%+. Since the incidence of durable MR4 is 60–80%, ~30–60% of all patients with CML on TKI therapy may potentially achieve TFR with optimization of the current TKI strategies. Suggested conditions to attempt TFR include low or intermediate Sokal risk CML in first chronic phase (no evidence or history of transformation); quantifiable *BCR::ABL1* transcripts (e13a2, e14a2); long-term TKI therapy (5+ years); and documented DMR for >2–5 years. Once TFR is attempted, patients should be monitored molecularly every 1.5–2 months in the first 6–12 months, every 2–3 months in the second year, and then every 3–6 months in subsequent years (perhaps for up to 6–8 years of TFR status). TKIs should not be restarted unless the *BCR::ABL1* transcripts (IS) increase to >0.1% (documented at least twice).

Management of TKI Resistance Resistance to a TKI refers today to *BCR::ABL1* transcripts (IS) >10% after 6 months of TKI frontline therapy or *BCR::ABL1* transcripts (IS) >1% after 12+ months of frontline TKI therapy. The NCCN and ELN updates of the criteria for resistance and suboptimal response have evolved over time, settling on more conservative criteria to change TKI therapy. The long-term follow-up studies have shown that patients who previously met suboptimal or resistance criteria still had excellent long-term survival without changing TKI therapy. For example, older patients with persistent *BCR::ABL1* transcripts (IS) 1–10% at 2 years of imatinib therapy still had 10-year survival rates similar to those with transcripts <1%.

Before switching TKI therapy for resistance, it is important to check for BCR::ABL1 kinase domain mutations. About 50% of patients with TKI resistance have BCR::ABL1 mutations that may selectively respond to particular TKIs. For example, mutations involving Y253H, E255K/V, and F359V/C/I respond better to dasatinib or bosutinib. Mutations involving V299L, T315A, and F317L/F/I/C respond better to nilotinib. T315I mutations require therapy with ponatinib or asciminib and serious consideration of allogeneic HSCT.

Dasatinib, bosutinib, nilotinib, ponatinib, and asciminib are approved for CML salvage therapy. The second-generation TKIs are excellent second-line therapies after imatinib frontline resistance. In patients with resistance to one second-generation TKI, switching to another second-generation TKI yields poor results (MR2 rate 10–20%; unless specific guiding mutations), and switching to ponatinib is indicated. Ponatinib is a very potent TKI with high activity in T315I-mutated CML and in CML after two TKI exposures (particularly if there is resistance to a second-generation TKI). In third-line therapy, ponatinib produces high molecular response rates and improves survival compared with second-generation TKIs. Asciminib is approved as third-line therapy and for T315I-mutated disease.

The TKI dose schedule in later-line therapy for CML resistance is as follows: dasatinib 100 mg daily; nilotinib 400 mg twice daily; bosutinib 500 mg daily; ponatinib 45 mg daily; and asciminib 40 mg twice daily or 80 mg once daily or 200 mg twice daily in T315I-mutated CML. Because ponatinib 45 mg daily may be associated with serious side effects, a response-directed dose-adjusted regimen (starting dose of 45 mg and reduction to 15 mg once MR2 is achieved) was investigated and resulted in a lower incidence of arterio-occlusive events.

TKI later-line therapy in chronic-phase CML with dasatinib, nilotinib, bosutinib, ponatinib, or asciminib is associated with MR2 rates of 30–80%, depending on the line of therapy (second vs later), CML status (cytogenetic/molecular relapse vs hematologic relapse), prior response to other TKIs, number of prior TKIs used, and the mutations at the time of relapse. The estimated 6- to 8-year survival rates with second-generation TKIs as second-line therapy are 65–75% (compared with <30–40% before their availability). Ponatinib third-line therapy resulted in an MR2 rate of 50–60%, MMR rate of 40%, and a 5-year overall survival rate of 70–75%. Its results appear even surprisingly better in real-word data: MR2 rate 80%, MMR rate 70–75%, and MR4 rate 40%.

ALLOGENEIC HEMATOPOIETIC STEM CELL TRANSPLANT

Allogeneic HSCT, a curative modality in CML, is associated with long-term survival rates of 50–70% when implemented in chronic phase. It carries the risk of complications related to myelosuppression, infections, and graft-versus-host disease (GVHD). Consequently, the early (1-year) mortality rate is 5–30%, and 10–15% of patients die in the subsequent 1–2 decades from long-term complications of the transplant (rather than from CML relapse). These are related to GVHD, organ dysfunction, development of second cancers, occasional late relapses, and hazard ratios for mortality higher than in the normal population. Other significant morbidities include infertility, chronic immune-mediated complications, cataracts, hip necrosis, and other morbidities affecting quality of life. The cure and early mortality rates in chronic-phase CML are also associated with several factors: patient age, duration of chronic phase, whether the donor is related or unrelated, degree of matching, preparative regimen, and others. In accelerated-phase CML, the cure rates with allogeneic HSCT are 30–50%, depending on the definition of accelerated disease. Patients with clonal evolution as the only criterion have cure rates of 40–50%. Patients undergoing allogeneic HSCT in second chronic phase have cure rates of 40–50%. The cure rates with allogeneic HSCT in blastic-phase CML are ≤20%. Post–allogeneic HSCT strategies in the setting of molecular or cytogenetic relapse include the use of TKIs for prevention or treatment of relapse, donor lymphocyte infusions, and second allogeneic HSCTs, among others. TKIs are successful at reinducing cytogenetic/molecular remissions in the setting of cytogenetic or molecular relapse after allogeneic HSCT.

Choice and Timing of Allogeneic HSCT Allogeneic HSCT was considered first-line CML therapy before 2000. The positive experience with TKIs has now relegated its use as a later-line approach. It should be considered in any patient in chronic-phase CML who develops resistance to a second-generation TKI or with a T315I-mutated disease. Relying on third-generation TKIs in these settings as long-term therapy can be very expensive ($250,000 to $1.5 million per year) and may be associated with serious toxicities. Among patients who present with or evolve to blastic phase, combinations of chemotherapy and TKIs should be used to induce remission, followed by allogeneic HSCT as soon as possible. The same applies to patients who evolve from chronic to accelerated phase. Patients with de novo accelerated-phase CML may do well with long-term TKI therapy (estimated 8-year survival rate 75%). Older patients with CML (age 65–70 years or older) have a reasonable CML-specific 10-year overall survival rate even with persistent molecular disease after 2 years of TKI therapy, with *BCR::ABL1* transcripts (IS) 1–10%, or even > 10%. Such patients may opt to continue on daily TKI therapy alone or with the addition of other agents (hydroxyurea, decitabine, low-dose cytarabine). They can then remain in chronic-phase CML disease without MR2 and avoid the adverse events and poorer quality of life associated with allogeneic HSCT.

MONITORING THERAPY IN CML

Achievement of complete cytogenetic response by 12 months of imatinib therapy and its persistence later, the only consistent prognostic factor associated with prolonged survival, is now the main therapeutic endpoint in CML. Failure to achieve a complete cytogenetic response by 12 months or occurrence of later cytogenetic or hematologic relapse is considered as treatment failure and an indication to change therapy. Because later-line TKI therapy may re-establish good outcome, it is important to ensure patient compliance to continued TKI therapy and change therapy when cytogenetic relapse is confirmed unless this is related to nonadherence. Patients on frontline imatinib therapy should be closely monitored until documentation of complete cytogenetic response, at which time they can be monitored every 6 months with peripheral blood PCR or more frequently (e.g., every 3 months) if there are concerns about changes in *BCR::ABL1* transcripts. Cytogenetic relapse on imatinib is an indication of treatment failure and need to change TKI therapy. Mutational analysis in this instance helps in the selection of the next TKI and identifies mutations in 30–50% of patients. Mutational studies by standard Sanger sequencing (which is the technique currently available in most clinical laboratories) in patients in complete cytogenetic response (in whom there may be concerns of increasing *BCR::ABL1* transcripts) identify mutations in ≤5% and are therefore not indicated.

Changes of TKI therapy for patients with "slow" molecular response have not been proven to be of long-term benefit compared to changes when more obvious signs of resistance appear.

TREATMENT OF ACCELERATED AND BLASTIC PHASES

Patients in accelerated or blastic phase may receive therapy with TKIs, preferably second- or third-generation TKIs (dasatinib, nilotinib, bosutinib, ponatinib), alone or in combination with chemotherapy, to reduce the CML burden, before undergoing allogeneic HSCT. Response rates (major hematologic) with single-agent TKIs range from 30 to 50% in accelerated phase and from 20 to 30% in blastic phase. Cytogenetic responses, particularly complete cytogenetic responses, are uncommon (10–30%) and transient in blastic phase. Studies of TKIs in combination with chemotherapy show that combined TKI-chemotherapy strategies increase the response rates and their durability and improve survival. This is particularly true in CML lymphoid blastic phase, where the combination of anti-ALL chemotherapy with TKIs results in complete response rates of 70% and median survival times of 3 years (compared with historical response rates of 40–50% and median survival times of 12–18 months). This allows many patients to undergo allogeneic HSCT in a state of minimal CML burden or second chronic phase, which are associated with higher probability of long-term survival. In CML nonlymphoid blastic phase, anti–acute myeloid leukemia chemotherapy combined with TKIs results in complete remission rates of 30–50% and median survival times of 12 months (compared with historical response rates of 20–30% and median survival times of 3–5 months). In accelerated phase, response to single TKIs is significant in conditions where "softer" accelerated phase criteria are considered (e.g., clonal evolution alone, thrombocytosis alone, significant splenomegaly or resistance to hydroxyurea, but without evidence of high blast and basophil percentages). In accelerated phase, combinations frequently include TKIs with low-intensity chemotherapy such as low-dose cytarabine, decitabine, interferon α, hydroxyurea, or others.

OTHER TREATMENTS AND SPECIAL THERAPEUTIC CONSIDERATIONS

Interferon α Interferon α is considered in combination with TKIs (an investigational approach) and occasionally in patients during pregnancy.

Chemotherapeutic Agents Hydroxyurea remains a safe and effective agent (at daily doses of 0.5–10 g) to reduce initial CML burden, as a temporary measure in between definitive therapies, or in combination with TKIs to sustain complete hematologic or cytogenetic responses. Busulfan is often used in allogeneic HSCT preparative regimens. Because of its side effects (delayed myelosuppression, Addison-like disease, pulmonary and cardiac fibrosis, myelofibrosis), it is now rarely used in the chronic management of CML. Omacetaxine, low-dose cytarabine, decitabine, 6-mercaptopurine, 6-thioguanine, thiotepa, anagrelide, and other agents are sometimes useful in different CML settings to control the disease burden, usually in combination with a daily TKI.

Others Leukapheresis is occasionally used in patients presenting with extreme leukocytosis and leukostatic complications. Single doses of high-dose cytarabine or high doses of hydroxyurea, with tumor lysis management, may be as effective and less cumbersome.

Pregnancy and CML TKI therapy in the first trimester of pregnancy is associated with fetal malformations (2–10%). Women with CML who become pregnant should discontinue TKI therapy immediately. Among 125 babies delivered to women with CML who discontinued imatinib therapy as soon as the pregnancy was known, three babies were born with neurologic, skeletal, and renal malformations, suggesting the teratogenicity of imatinib known from animal studies. A similar experience has been reported with dasatinib, where the incidence of malformations was reported to be higher, 10–12%. Data are scant with other TKIs. Control of CML during pregnancy can be managed with leukapheresis for severe symptomatic leukocytosis in the first trimester and with hydroxyurea subsequently until delivery. TKI therapy with imatinib (but not dasatinib; scant data with nilotinib or bosutinib), if indicated, may be safe after 20 weeks of gestation and in the third trimester. There are reports of successful pregnancies and deliveries of normal babies with interferon α therapy and registry studies in essential thrombocytosis, but interferon α has side effects that may be troublesome during pregnancy, can be antiangiogenic, and may increase the risk of spontaneous abortions.

Cytogenetic Abnormalities in Ph-Negative Cells; Mutations and CML Approximately 5% of patients on TKI therapy and in cytogenetic/molecular response may develop chromosomal abnormalities in the Ph-negative cells. These may involve loss of chromosome Y, trisomy 8, 20q−, chromosome 5 or 7 abnormalities, and others. Most chromosomal abnormalities disappear spontaneously and may be indicative of the genetic instability of the hematopoietic stem cells that predisposes the patient to develop CML in the first place. Rarely (in <1% of instances), abnormalities involving chromosomes 5 or 7 may be truly clonal and evolve into myelodysplastic syndrome, acute myeloid leukemia, or myeloproliferative neoplasms. This is thought to be part of the natural course of patients in whom CML was suppressed and who live long enough to develop other hematologic malignancies. The presence of mutations (*ASXL1, DNMT3A, RUNX1, TET2, EZH2, IDH1/2*) at diagnosis or later during CML was discussed earlier.

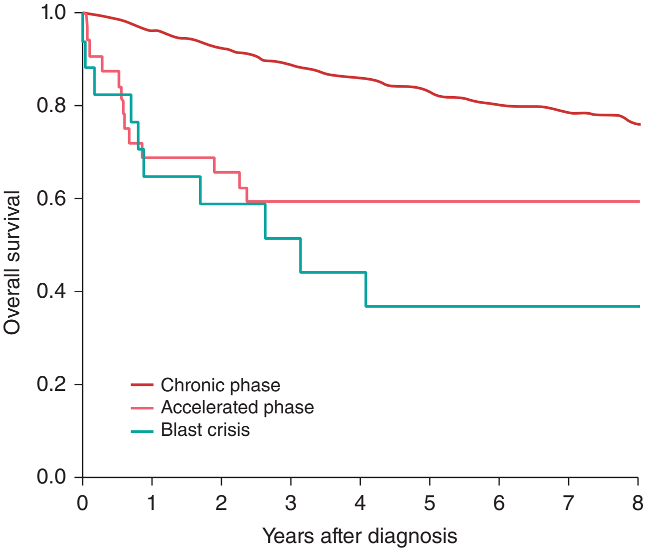

No. at risk

	0	1	2	3	4	5	6	7	8
CP	699	672	631	530	446	365	278	212	144
AP	32	22	21	18	14	10	5	4	2
BC	17	11	10	7	6	5	4	4	1

FIGURE 110-3 Survival in chronic (CP), accelerated (AP), and blastic phase (BP) phases of chronic myeloid leukemia (CML) in the population-based Swedish national registry study. The accelerated- and blastic-phase cases are de novo presentations. The favorable outcome with de novo blastic phase may be due to use of 20% blasts or more to define blastic phase. *(From Dr. Martin Hoglund, Swedish CML Registry, 2013.)*

GLOBAL ASPECTS OF CML

Routine physical examinations and blood tests in the United States and advanced countries result in early detection of CML in most patients. About 50–70% of patients with CML are diagnosed incidentally, and high-risk CML as defined by prognostic models (e.g., Sokal risk groups) is found in only 10% of patients. This is different in emerging nations where most patients are diagnosed following evaluation for symptoms and many present with high tumor burden, such as massive splenomegaly, and advanced phases of CML (high-risk CML documented in 20–30%). Therefore, the prognosis of such patients on TKI therapy may be worse than the published experience.

The high cost of TKI therapies (annual costs of $120,000–270,000 in the United States; lower but variable in the rest of the world) makes the general affordability of such treatments difficult. Fortunately, in 2024, some imatinib generic formulations cost about $500/year in the United States and worldwide. Generic dasatinib is also available in many geographies, and generic formulations of bosutinib, nilotinib, and ponatinib may become available by 2027. Therefore, in frontline CML therapy, if survival is the treatment endpoint, generic imatinib is a good choice. If TFR is the endpoint, generic dasatinib (50 or 100 mg daily) is a good choice. Although TKI treatment penetration is high in nations with universal health care and where cost of therapy is not an issue (e.g., Europe, Canada, Australia, United Kingdom), it may be less so in other nations, even in advanced ones like the United States, where out-of-pocket expenses may be prohibitive to a subset of patients. The estimated 10-year survival rate in CML is >85% in single-institution or national studies in countries with TKI affordability (Sweden) (**Figs.** 110-2 and **110-3**), whereas the estimated 10-year survival rate worldwide is likely to be <50%. The Surveillance, Epidemiology, and End Results (SEER) data from the United States report an estimated 5-year survival rate of 70% in the era of TKIs. It appears that the treatment penetration of imatinib and other TKIs into CML therapy worldwide is still not optimal.

The current high cost of TKI therapies, particularly in later-line therapy, encouraged considering allogeneic HSCT as third-line therapy (one-time cost of $20,000–500,000) despite the associated mortality and morbidities. Safe and effective generic TKIs should be preferred frontline and second-line therapies in CML, precluding the necessity of an allogeneic HSCT until evidence of resistance to generic second-generation TKIs.

FURTHER READING

Cortes JE et al: Bosutinib versus imatinib for newly diagnosed chronic myeloid leukemia: Results from the randomized BFORE Trial. J Clin Oncol 36:231, 2018.

Cortes JE et al: Ponatinib efficacy and safety in Philadelphia chromosome-positive leukemia: Final 5-year results of the phase 2 PACE trial. Blood 132:393, 2018.

Gener-Ricos G et al: Low-dose dasatinib (50 mg daily) frontline therapy in newly diagnosed chronic phase chronic myeloid leukemia: 5-year follow-up results. Clin Lymphoma Myeloma Leuk 23:742, 2023.

Haddad FG, Kantarjian H: Navigating the management of chronic phase CML in the era of generic BCR::ABL1 tyrosine kinase inhibitors. J Natl Compr Canc Netw 22:e237116, 2024.

Hochhaus A et al: Asciminib in newly diagnosed chronic myeloid leukemia. N Engl J Med 391:885, 2024.

Kantarjian H et al: Long-term outcomes with frontline nilotinib versus imatinib in newly diagnosed chronic myeloid leukemia in chronic phase: ENESTnd 10-year analysis. Leukemia 35:440, 2021.

Kantarjian HM et al: Revising six established practices in the treatment of chronic myeloid leukaemia. Lancet Haematol 10:e860, 2023.

Kantarjian HM: What is the impact of failing to achieve TKI therapy milestones in chronic myeloid leukemia. Leukemia 37:2324, 2023.

Mahon FX et al: European Stop Tyrosine Kinase Inhibitor Trial (EURO-SKI) in chronic myeloid leukemia: Final analysis and novel prognostic factors for treatment-free remission. J Clin Oncol 42:1875, 2024.

Senapati J et al: Management of chronic myeloid leukemia in 2023: Common ground and common sense. Blood Cancer J 13:58, 2023.

Shih YT et al: Treatment value of second-generation BCR-ABL1 tyrosine kinase inhibitors compared with imatinib to achieve treatment-free remission in patients with chronic myeloid leukaemia: A modelling study. Lancet Haematol 6:e398, 2019.

111 Acute Lymphoid Leukemia

Dieter Hoelzer

In acute lymphoblastic leukemia (ALL), the malignant clone arises from hematopoietic progenitors in the bone marrow or lymphatic system resulting in an increase of immature nonfunctioning leukemic cells. Infiltration of bone marrow leads to anemia, granulocytopenia, and thrombocytopenia with the clinical manifestations of fatigue, weakness, infection, and hemorrhage. These symptoms are more often the reason a patient first seeks medical advice rather than consequences of tumor bulk, such as lymph node enlargement, hepatosplenomegaly caused by leukemic infiltration, or symptoms of the central nervous system (meningeosis leukemica).

INCIDENCE AND AGE

ALL is the most frequent neoplastic disease in children with an early peak at the age of 3–4 years. The incidence in adults ranges from 0.7 to 1.8/100,000 per year, being somewhat higher in adolescents and young adults (AYAs), decreasing in adults but increasing again in elderly people. Thus, Philadelphia chromosome–positive ALL (Ph+ ALL; *BCR/ABL* translocation) is observed in half of the elderly B-lineage patients. The frequency of immunologic, cytogenetic, and genetic subtypes changes substantially with age.

ETIOLOGY

The etiology of acute leukemias is unknown. Internal and external factors influence the incidence of leukemia. Exposure to ionizing radiation or to chemicals, including prior chemotherapy, is associated with an increased risk of developing leukemia, more often observed in acute myeloid leukemia (AML). However, increasingly, secondary ALLs have been observed, particularly after cytostatic treatment with alkylating agents and topoisomerase inhibitors as treatment for primary tumors, most often for AML, myelodysplastic syndromes, or breast cancer.

CONGENITAL DISORDERS

Patients with some rare congenital chromosomal abnormalities have a higher risk of developing acute leukemia (e.g., Klinefelter's syndrome, Fanconi's anemia, Bloom's syndrome, ataxia-telangiectasia, and neurofibromatosis). Those with Down's syndrome have a 20-fold increased incidence of leukemia; ALL is increased in childhood and AML at an older age.

INFECTIOUS AGENTS

No direct evidence implicates viruses as a major cause of human acute leukemia. However, viruses are involved in the pathogenesis of two lymphoid neoplasias. In the endemic African type of Burkitt's lymphoma, the Epstein-Barr virus, a DNA virus of the herpes family, has been implicated as a potential causative agent (see Chap. 199). Endemic infection with human T-cell leukemia virus I in Japan and the Caribbean has been shown to be an etiologic agent for rare cases of adult T-cell leukemia/lymphoma (see Chap. 207).

DIAGNOSIS AND CLASSIFICATION

The diagnosis of acute leukemia is first made by examination of the peripheral blood and bone marrow. For further classification of the leukemic blast cells, cytochemical stains, immunologic markers, and cytogenetic and molecular analysis are required. The immunologic markers are still the major criteria to subdivide into B-cell lineage or T-cell lineage ALL leukemias.

PERIPHERAL BLOOD

Peripheral blood counts and a differential count from a Wright-Giemsa–stained blood smear are essential at the time of presentation.

TABLE 111-1 Laboratory Values at Diagnosis of Acute Lymphoblastic Leukemia (ALL)

		ALL
NO.		1273
Initial white blood cell count (× 10^9/L)	<10	41%
	10–50	31%
	>50–100	28%
	>100	16%
Neutrophils (× 10^9/L)	<50–100	12%
	<100,000	16%
Platelets (× 10^9/L)	<20	22%
	21–40	22%
	41–100	29%
	>100	27%
Hemoglobin (g/dL)	<7	20%
	7–9	33%
	>9	47%
Leukemic blasts in peripheral blood	0%	8%
	25–75%	34%
	>75%	36%
Leukemic blasts in bone marrow	<50%	4%
	51–90%	25%
	>90%	71%

Source: Data from three consecutive German Multicenter Trials for Adult ALL (GMALL).

The white blood cell (WBC) count in ~40% of ALL patients is reduced or normal (Table 111-1). Only 16% of patients have a WBC above 100 × 10^5/L. It is noteworthy that in 8% of ALL patients, no circulating leukemic blast cells were observed. Thus, in the frequently used automatic blood cell counting, the diagnosis may not be detected.

Peripheral blood characteristically shows anemia, thrombocytopenia, and neutropenia. Nearly one-third of patients have hemoglobin levels <7–8 g/dL. A platelet count below the critical number of 20 × 10^9/L, associated with the risk of bleeding, and neutropenia (neutrophils <0.5 × 10^9/L), associated with a higher risk of infection, are each noted in one-fifth of adults with ALL.

BONE MARROW EXAMINATION

Bone marrow aspirates/biopsies are important to assess immunologic, cytogenetic, and genetic markers. A biopsy of the bone marrow is essential to confirm the diagnosis of acute leukemia and to distinguish between AML and ALL. The bone marrow is usually heavily packed with leukemic blast cells with >90% in ~70% of patients, and thus, the normal hemopoietic elements are greatly reduced or absent.

LUMBAR PUNCTURE

The examination of the cerebrospinal fluid is an essential routine diagnostic measure for ALL. Central nervous system (CNS) leukemia is diagnosed if ≥5 cells/μL or leukemic blast cells were observed by morphology in cerebrospinal fluid. Opinions differ as to when the first lumbar puncture should be done—i.e., either delay lumbar puncture until remission is achieved to avoid seeding of the CNS with leukemic blast cells from the peripheral blood during the spinal tap or perform the lumbar puncture before treatment starts, since early recognition of CNS disease will lead to immediate CNS-specific therapy. Lumbar puncture is restricted to patients with an adequate platelet count (>20 × 10^9/L) and without manifest clinical hemorrhages. To eliminate potentially transferred blast cells, patients should receive intrathecal methotrexate at the first lumbar puncture.

MORPHOLOGIC SUBTYPES IN ALL

The French-American-British (FAB) classification distinguished three subgroups. L1 and L2 morphology has no clinical consequences. Only the L3 morphology, observed in up to 5% of adult patients, is indicative for mature B-cell lineage ALL (B-ALL) (see Chap. 65).

IMMUNOLOGIC SUBTYPES

A series of monoclonal antibodies is employed to identify antigens expressed on the surface of leukemic cells, corresponding to the pathways of normal B-cell differentiation (see Fig. 113-2). The immunologic classification aims to subdivide ALLs according to the presence or absence of B-cell or T-cell markers. A marker is considered positive if >20% of the cells are stained with the monoclonal antibody.

There are different immunologic classifications, such as that of the European Group for the Immunological Characterization of Leukemias (EGIL), with clear therapeutic implications. Table 111-2 gives a simplified correlation of immunologic subtypes, cytogenetics and molecular aberrations, and clinical characteristics.

B-Cell Lineage ALL (B-ALL) More than 70% of adult ALLs are of B-cell origin, and the most frequent immunologic subtype, common ALL, is characterized by the presence of the ALL antigen CD10 without markers of relatively mature B cells such as cytoplasmic or surface membrane immunoglobulins. Pre-B-ALL (early B-ALL) is characterized by the expression of cytoplasmic immunoglobulin, which is negative in common ALL, but otherwise is identical with respect to all other cell markers. Pro-B-ALL corresponds to early B-cell differentiation and was formerly termed non-T-, non-B-ALL or null ALL because neither T-cell nor B-cell features could be demonstrated. This subtype is HLA-DR, terminal deoxynucleotidyl transferase, and CD19 positive and composes ~12% of adults ALL.

Mature B-ALL is seen in 3–4% of adults and is also known as Burkitt's leukemia. In mature B-ALL, blast cells express surface antigens of mature B cells, including the sIgM.

T-Cell Lineage ALL (T-ALL) Approximately 25% of adult ALLs are of T-cell lineage. All cases express the T-cell antigen CD7 and cytoplasmatic CD3 (CyCD3) or surface CD3. According to their stage of T-cell differentiation, they may express other T-cell antigens (e.g., the E-rosette receptor CD2 and/or the cortical thymocyte antigen CD1a). Early pro/pre-T-ALL (also termed early T precursor ALL [ETP-ALL]), cortical or thymic T-ALL, and mature T-ALL can be distinguished with these markers. ETP-ALL is characterized by lack of CD1a and CD8, weak CD5 expression, and at least one myeloid/stem cell marker.

Biphenotypic or Mixed Leukemias Biphenotypic leukemias are defined as those expressing markers of both lymphoid and myeloid lineages on the same leukemic cells. Bilineage leukemias are those with two populations of blast cells with either lymphoid or myeloid antigens. It is not clear whether these patients should receive an ALL or AML treatment protocol. In pediatric studies, starting with a pediatric ALL protocol seemed preferable, which was then followed by AML consolidation elements.

CYTOGENETIC AND MOLECULAR ANALYSIS

Cytogenetic and molecular analyses should be performed in all cases. They are important to define ALL subtypes, can identify independent prognostic markers of disease-free survival, and may determine specific targeted therapies.

The diagnostic techniques for ALL are standard cytogenetics, fluorescence in situ hybridization, and reverse transcriptase polymerase chain reaction. These methods allow the detection of Ph+ ALL, with the chromosomal translocation t(9;22)(q34;q11) and the detection of the corresponding *BCR-ABL1* gene rearrangement. Further ALL entities that have been identified are t(4;11)(q21;q23)/*MLL-AFA4*, abn11q23/*MLL*, and t(1;19)(q23;p13)/*PBX-E2A*.

Gene expression profiling, single nucleotide polymorphism array analysis, array-comparative genomic hybridization, and next-generation sequencing recognize the newly defined ALL entities: ETP-ALL and Ph-like ALL.

TABLE 111-2 Immunologic, Cytogenetic, Molecular, and Clinical Characteristics of Adult Acute Lymphoblastic Leukemia (ALL)

SUBTYPES	MARKER	INCIDENCE	FREQUENT CYTOGENETIC ABERRATIONS	GENETIC ABERRATIONS AND FUSION TRANSCRIPTS	CLINICAL CHARACTERISTICS	RELAPSE KINETICS AND LOCALIZATION
B-lineage ALL (B-ALL)	**HLA-DR+, TdT+, CD19+, and/or CD79a+, and/or cyCD22+**	**76%**				
Pro B-ALL	No additional differentiation markers Frequent myeloid coexpression (>50%) CD10–	112%	t(4;11) (q21;q23)	70% *ALL1-AF4* (20% Flt3 in MLL+)	High WBC (>100,000/μL) (26%)	Mainly BM (>90%)
Common ALL	CD10+	49%	t(9;22)(q34;q11) del(6q)	33% *BCR::ABL1* with 54% *IKFZ1* del >25% *CDKN2A/B*	Higher age >50 years (24%)	Mainly BM (>90%) Prolonged relapse kinetics (up to 5–7 years)
Pre-B-ALL	CD10±, cyIg+	12%	t(9;22)(q34;q11) t(1;19)(q23;p13)	4% t(1;19)/*PBX-E2A*		
Mature B-ALL	TdT–, CD34–, sIg+	4%	t(8;14)(q24;q32) t(2;8)(p12;q24) t(8;22)(q24;q11)		Higher age >55 years (27%) Frequent organ involvement (32%) and CNS involvement (13%)	Frequent CNS (10%) Short relapse kinetics (up to 1–1.5 years)
T-lineage ALL (T-ALL)	**TdT±, cyCD3, CD7+**	**24%**	t(10;14)(q24;q11) t(11;14)(p13;q11)	50% *NOTCH1B* 33% *HOX11b* 5% *HOX11L2b* 4% *NUP213-ABL1*	Younger age (90% <50 years) Frequent mediastinal tumors (60%) Frequent CNS involvement (8%) High WBC (>50/μL) (46%)	Frequent CNS (up to 10%) Extramedullary (6%) Intermediate relapse kinetics (up to 3–4 years) When relapsed, fast progression
Early Pro/Pre T-ALL	No additional differentiation markers, mostly CD2–/(+), SCD3–, CD1a–	6%				
Cortical T-ALL	CD1a+, sCD3±	12%				
Mature T-ALL	sCD3+, CD1a–	6%				

Abbreviations: BM, bone marrow; CNS, central nervous system; WBC, white blood cells.

TABLE 111-3 Response Parameters According to Minimal Residual Disease (MRD)

TERMINOLOGY	DEFINITION
Complete hematologic remission (CHR)	Leukemic cells not detectable by light microscopy (<5% blast cells in bone marrow [BM])
Complete molecular remission/MRD negativity	Patient in complete remission, MRD not detectable, ≤0.01% = ≤1 leukemia cell in 10,000 BM cells
Molecular failure/MRD positivity	Patient in complete hematologic remission, but not in molecular complete remission >0.01%
Molecular relapse/MRD positivity	Patient still in complete remission, had prior molecular complete remission, leukemic blast cells in BM not detectable (<5%)
Hematologic relapse	>5% blast cells in BM/blood

Ph-like ALL, also known as *BCR-ABL1*-like ALL, is characterized by genetic lesions similar to Ph+ ALL, associated with *IKZF1* (Ikaros) gene deletion, *CLRF2* (gene for cytokine-like receptor-2) overexpression, and tyrosine kinase activating rearrangements involving *ABL1*, *JAK2*, *PDGFRB*, and several other genes; however, it is *BCR-ABL1* negative. The frequency is 10% in children and 25–30% in young adults but does not increase further with age like Ph+ ALL. Treatment based on the underlying genetic lesion with BCR-ABL inhibitors (e.g., dasatinib) or JAK2 inhibitors (e.g., ruxolitinib) has so far had limited success in adults.

MINIMAL RESIDUAL DISEASE

Minimal residual disease (MRD) is the detection of residual leukemic cells that are not recognizable by light microscopy. Methods for determining MRD are based on the detection of leukemia-specific aberrant immunophenotypes by flow cytometry, the evaluation of leukemia-specific rearranged immunoglobulin or T-cell receptor sequences by real-time quantitative polymerase chain reaction, or the detection of fusion genes associated with chromosomal abnormalities (e.g., *BCR-ABL*, *MLL-AF4*). The detection limit with these methods is 10^{-3}–10^{-5} (0.1–0.001%). With new techniques such as next-generation sequencing (NGS) or digital droplet polymerase chain reaction (ddPCR), the sensitivity may increase to 10^{-5}–10^{-6}. The phenotypic aberrations are unique to each patient with ALL and can be detected in up to 95% of individuals. Collection of bone marrow at diagnosis for identification of patients' individual markers is essential for follow-up of MRD.

MOLECULAR RESPONSE AFTER INDUCTION THERAPY AND IMPACT ON OUTCOME

Achievement of molecular complete response/molecular remission is the most relevant independent prognostic factor for disease-free and overall survival in pediatric and adult ALL **(Table 111-3)**. Patients with molecular complete remission after induction therapy have significantly superior outcomes in several studies, with a disease-free survival rate of ~70% compared to <40% for MRD-positive patients. Patients with molecular failure after induction should proceed to a targeted therapy to reduce the tumor load, followed by allogeneic stem cell transplantation (SCT), if possible.

PROGNOSTIC FACTORS, RISK STRATIFICATION, AND MRD

The aim of identification of prognostic parameters at diagnosis, which include age, WBC count, immunophenotype, and cytogenetic and genetic aberrations, is to stratify patients into risk groups: standard-risk patients are patients without any risk factors, and high-risk patients are those with one or more risk factors. High-risk patients are most often candidates for SCT in first complete remission (CR). MRD is thus the most important prognostic factor during therapy **(Fig. 111-1)**; 20–30% of adult ALL patients who are MRD negative after induction will relapse. Potential reasons include loss of sensitivity, evolution of leukemic subclones, and extramedullary origin of disease. If the MRD status of a patient is not available, risk stratification should rely on clinical and laboratory risk factors evaluated at diagnosis.

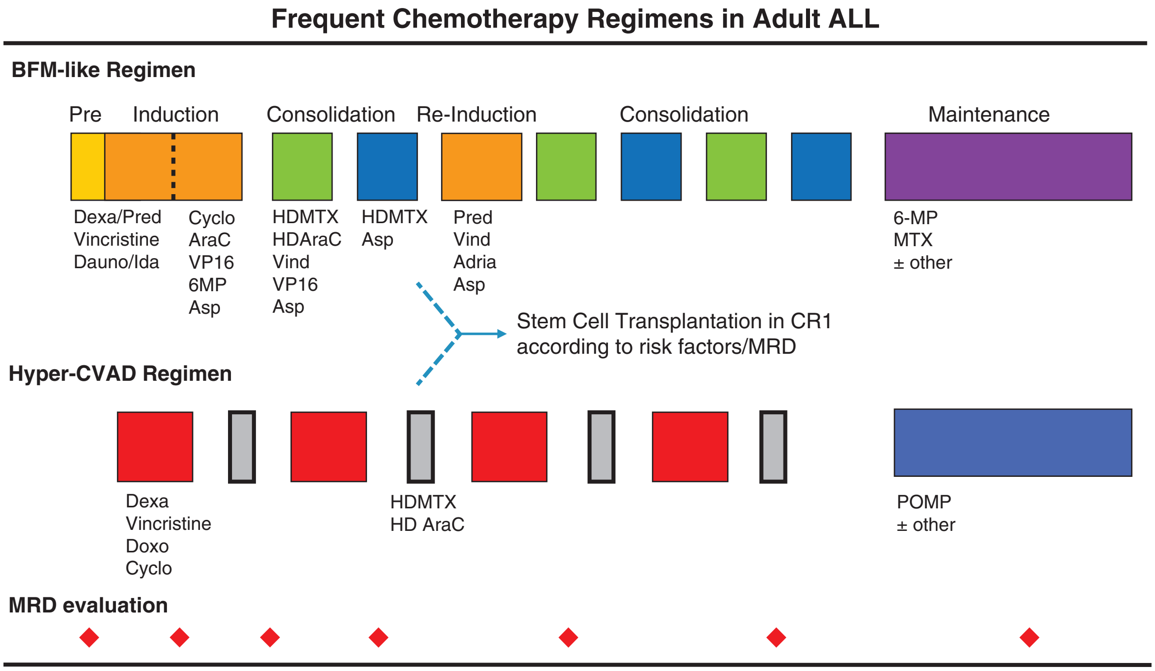

- Prophylactic CNS treatment; intrathecal monotherapy; MTX or intratecal triple MTX, AraC, Dexa/Pred, +/– cranial irradiation (24 Gy)
- MRD evaluation; material collection at diagnosis, evaluation after Induction I, Induction II, Consolidation I, then every 3 months.
- Rituximab in B-lineage, nelarabine in T-lineage
- Maintenance therapy, ~2 years in all subtypes (except Burkitt)

FIGURE 111-1 A schematic treatment algorithm in acute lymphoblastic leukemia (ALL). 6-MP, 6-mercaptopurine; Adria, Adriamycin (doxorubicin); AraC, cytarabine; Asp, asparaginase; BFM, Berlin-Frankfurt-Münster; CNS, central nervous system; CR1, first complete remission; Cyclo, cyclophosphamide; Dauno, daunorubicin; Dexa, dexamethasone; Doxo, doxorubicin; HD, high-dose; Hyper-CVAD, hyperfractionated cyclophosphamide, vincristine, doxorubicin, and dexamethasone; Ida, idarubicin; MRD, minimal residual disease; MTX, methotrexate; POMP, mercaptopurine, vincristine, methotrexate, and prednisolone; Pred, prednisolone; Vind, vindesine; VP16, etoposide.

TREATMENT PRINCIPLES

Treatment of ALL consists usually of pre-phase therapy, induction therapy, consolidation cycles, and maintenance treatment. Treatment should start immediately when the diagnosis of ALL is established and can later be specified or adapted when ALL subtype is known (e.g., Ph+).

Pre-Phase Therapy Pre-phase therapy consisting of glucocorticoids (prednisone 20–60 mg/d or dexamethasone 6–16 mg/d, both IV or PO) alone or in combination with another drug (e.g., vincristine, cyclophosphamide) is usually given for ~5–7 days. It allows safe tumor reduction to avoid tumor lysis syndrome, to initiate supportive therapy, such as substitution of platelets/erythrocytes, or to treat infections. The time required for pre-phase therapy will also allow time to obtain results of the diagnostic workup (e.g., cytogenetics, molecular genetics).

Induction Therapy The goal of induction therapy is the achievement of a CR or, even better, a molecular CR. With current regimens, the CR rate has increased to 80–90% and is higher for standard-risk patients (>90%) and lower for high-risk patients (~60%).

Induction regimens are centered around vincristine, glucocorticoids, and anthracyclines with or without cyclophosphamide or cytarabine. L-Asparaginase is the only ALL-specific drug and is now more intensively used in adults. Pegylated asparaginase has the advantage of a significantly longer period of asparagine depletion. Dexamethasone is often preferred to prednisone because it penetrates the blood-brain barrier and also acts on resting leukemic blast cells.

Two chemotherapy regimens are widespread (Fig. 111-1). One is patterned after the pediatric BFM (Berlin-Frankfurt-Münster) protocol, which is mostly used in European adult ALL trials. Another approach is to repeat two different alternating intensive chemotherapy cycles, identical for induction and consolidation, for eight cycles, such as Hyper-CVAD (hyperfractionated cyclophosphamide, vincristine, doxorubicin, and dexamethasone) protocol, which is preferentially used in the United States but also in many other parts of the world.

Postremission Consolidation Usual protocols use six to eight courses and often contain systemic high-dose (HD) therapy to reach sufficient drug levels in sanctuary sites such as the CNS. Most often HD methotrexate (1–1.5 g/m^2 and up to 3–5 g/m^2) and/or HD cytarabine (4–12 doses at 1–3 g/m^2) are administered.

Maintenance Therapy Maintenance therapy, a strategy transferred from childhood ALL, is mandatory. It consists of 6-mercaptopurine and methotrexate plus intrathecal therapy. The potential effect of further intensification cycles during maintenance remains unclear. The duration of maintenance therapy for T-ALL and B-ALL is 2–2.5 years, except for Burkitt's leukemia, for which a half year to 1 year is sufficient. In Ph+ ALL, patients also require maintenance therapy that should include a tyrosine kinase inhibitor (TKI), most likely the TKI that has been used during induction and consolidation therapy. It is also standard to give a TKI after allogeneic SCT. The duration of maintenance therapy with a TKI is also 2–2.5 years and should be guided by MRD evaluation. TKI use is often interrupted or switched to another TKI if toxicity occurs.

TREATMENT OF ALL PATIENTS ACCORDING TO AGE

The outcome of ALL is strictly related to the age of a patient, with cure rates of ~90% in children, decreasing to <10% in elderly or frail patients. Thus, age-adapted protocols have emerged, where the age limits are directed by the hematologic and nonhematologic toxicities. Table 111-4 provides a summary of the best results obtained in adult ALL according to ALL subtype, age, and treatment. The major risk of relapse is in the first 2 years and is less likely after 5 years.

TABLE 111-4 Best Results in Recent Studies for Adult Acute Lymphoblastic Leukemia (ALL)

SUBTYPE	TREATMENT	OVERALL SURVIVAL
Burkitt's leukemia	Short intensive chemotherapy + rituximab; no SCT; no maintenance	80–90%
B-lineage ALL, Ph–		
AYA 15–35/45 years	Pediatric inspired, few/no SCT	≥70–80%
Adults 45–55 years	Intensive chemotherapy +/– SCT	50–60%
Elderly 55–70 years	Less intensive chemotherapy + immunotherapy	~30%
Frail >70/75 years	Various	≤10%
B-lineage ALL, Ph+		
Ph BCR-ABL	Intensive chemotherapy + TKI +/– SCT	60–70%
Ph-like ALL	Chemotherapy + dasatinib/JAK inhibitors	≤50%
T-lineage ALL		
Early (ETP)	Intensive chemotherapy + nelarabine + SCT	40–50%
Cortical/thymic	Intensive chemotherapy + nelarabine, no SCT	70–84%
Mature	Intensive chemotherapy + nelarabine + SCT	30–50%

Abbreviations: AYA, adolescent and young adult; ETP, early T precursor; Ph, Philadelphia chromosome; SCT, stem cell transplantation; TKI, tyrosine kinase inhibitor.

PROPHYLAXIS AND TREATMENT OF CENTRAL NERVOUS SYSTEM LEUKEMIA

Prophylactic CNS therapy in ALL is essential in order to prevent CNS leukemia and to avoid spread of leukemic cells from the CNS back to the periphery. Treatment options include intrathecal therapy, systemic HD chemotherapy, and cranial radiation therapy (CRT). Wide variations exist in CNS prophylaxis regimens. Intrathecal therapy mostly consists of methotrexate as a single drug or in combination with cytosine arabinoside (AC) with or without glucocorticoids. The route of intrathecal therapy application is generally lumbar puncture. Systemic HD chemotherapy may comprise HDAC or HD methotrexate since both drugs reach cytotoxic drug levels in the cerebrospinal fluid and show effectiveness in overt CNS leukemia. CRT (18–24 Gy in 12 fractions over 16 days) is also effective as preventive treatment of CNS leukemia. Using combined modalities for CNS prophylaxis, the CNS relapse rate has decreased to 2–5%.

Particular attention to CNS prophylaxis is required for targeted therapies. In Ph+ ALL, not all TKIs cross the blood-brain barrier equally such as imatinib and nilotinib but dasatinib and probably ponatinib do. In immunotherapies, intrathecal therapy is required because most antibodies do not enter the CNS.

CNS involvement at diagnosis is observed in 5–10% of adult patients and is higher in mature B-ALL (up to 10–15%) and T-ALL (up to 10%). Treatment consists of the standard chemotherapy with additional intrathecal applications 3–5 times per week until blast cells are cleared in the spinal fluid. Patients with initial CNS involvement have a similar overall survival as CNS-negative patients.

Relapse in CNS is usually accompanied by bone marrow involvement, and if blast cells are not seen morphologically, MRD as a sign of discrete infiltration is positive in nearly all cases. CNS relapse requires local as well as systemic therapy. The outcome after CNS relapse is dismal, and salvage chemotherapy followed by allogeneic SCT is the most effective option. Chimeric antigen receptor (CAR) T cells (most often targeting CD19) can cross the blood-brain barrier and achieve CRs in patients with CNS relapse.

Extramedullary manifestation and relapses in ALL are often observed. Patients should have the general treatment and, if required, local intervention (e.g., local radiation if residual mediastinal mass).

STEM CELL TRANSPLANTATION

SCT is an essential part of the treatment strategy for adult ALL. Peripheral blood cells are more often being used as a stem cell source, instead of bone marrow. If no matched sibling stem cell donors are available, increasingly matched unrelated donors or haploidentical donors are used. Indications for SCT in first CR are controversial. However, in most studies, SCT is recommended for high-risk patients defined either by conventional prognostic factors or by MRD positivity. High-risk patients transplanted in first CR have a survival rate of 50% or greater. Decreasing transplant-related mortality from 20–30% to 10–15% has contributed substantially to better outcomes. For standard-risk patients with sustained molecular remission, allogeneic SCT in first CR is not recommended. Autologous SCT should be restricted to MRD-negative patients, *BCR-ABL*–negative patients and older patients because it is less toxic but associated with a substantially higher relapse rate. For all relapsed adult ALL patients, an allogeneic SCT is thus far the only curative option.

■ PEDIATRIC-INSPIRED THERAPIES FOR ADOLESCENTS AND YOUNG ADULTS

The principle of pediatric-inspired therapies is to have higher doses and more applications of ALL-specific drugs such as glucocorticoids, vincristine, and L-asparaginase and fewer myeloablative anthracyclines or alkylating agents, with strict adherence to time-dose intensity, thereby reducing the role of SCT. The overall survival rates for AYAs are 70–80%.

■ ADULT ALL

The treatment results for adult ALL patients have greatly improved with more intensive chemotherapy, optimized SCT, and better supportive care. In several recent multicenter prospective trials, the overall survival rate for standard-risk patients was >70% with chemotherapy alone, and for high-risk patients, the overall survival rate has increased from 20–30% to >50%.

■ ALL IN THE ELDERLY

Palliative treatment regimens for elderly patients have failed, with CR rates of ~40%, a high early death rate of 24%, and a poor overall survival of only a few months. Intensive chemotherapy has also failed, with a higher CR rate of 56%, but still an early death rate of 23%, and only moderate improvement of overall survival to 14 months. Specific elderly ALL protocols with less intensive therapy based on glucocorticoids, vincristine, and asparaginase, largely avoiding anthracyclines and alkylating agents, have improved outcomes. The early treatment-related death rate decreased to <10%, CR rates improved to ~90%, and overall survival of ~30 months was noted.

Frail patients above the age of 70–75 years have very poor survival of <10%. Hopefully, this will improve with ongoing targeted therapies with either TKIs in Ph+ ALL or immunotherapies.

■ TARGETED THERAPIES

Substantial progress in adult ALL has been made in the past decade by the introduction of new targeted therapies, including TKIs and immunotherapeutic approaches (Table 111-5).

■ TYROSINE KINASE INHIBITORS IN PHILADELPHIA-POSITIVE ALL

Patients with Ph+ ALL constitute ~25% of adult B-ALL patients, with the frequency increasing to ~50% among elderly patients. In the pre-imatinib era, CR rates were 60–70%; survival with chemotherapy was ~10%, and after allogeneic SCT, it was ~30%. With the first-generation TKI imatinib, CR rates increased to 80–90%, the rate of *BCR-ABL* negativity increased from 5 to 50%, and the 5- to 10-year overall survival improved to 50–70%.

Faster and deeper molecular responses are achieved with second-generation TKIs (dasatinib, nilotinib), and these responses apparently translate into a survival benefit. The third-generation TKI ponatinib is also effective in tumors bearing mutations (particularly T315I) that convey resistance to earlier-generation TKIs.

Treating adult Ph+ ALL with an allogeneic SCT in first CR is still a good treatment option for adult patients, with a 5-year overall survival of 60–70%. In elderly patients, when low-intensity chemotherapy was combined with dasatinib, the CR rate was >90%. In a next step, by combining mini-chemotherapy with a TKI and adding immunotherapy with inotuzumab (an anti-CD22 antibody), the CR rate was >90% and the overall survival improved further. A pilot experience with a chemotherapy-free regimen composed of dexamethasone, the TKI dasatinib, and the bispecific antibody blinatumomab (anti-CD19 and anti-CD3) demonstrated a CR rate of 98% and 2-year overall and disease-free survival rates of 95% and 88%, respectively. Blinatumomab eliminates Ph+ leukemic cells with resistant mutations.

TABLE 111-5 Targeted Therapies in Adult Acute Lymphoblastic Leukemia (ALL)

Tyrosine Kinase Inhibitors (TKIs)
Ph/BCR-ABL+ ALL
TKIs
Imatinib, dasatinib, nilotinib, bosutinib, ponatinib, asciminib
Ph/BCR-ABL-like ALL
ABL1, ABL2: dasatinib, ponatinib; JAK2: ruxolitinib
Immunologic Approaches
Antibodies directed against leukemia surface antigens
Monovalent antibodies
Bivalent antibodies against the tumor and CD3 (e.g., blinatumomab)
Adoptive cellular therapy
T cells engineered to kill leukemic cells
Checkpoint Inhibitors
PD-1 inhibitors: pembrolizumab, nivolumab
CTLA-4 inhibitors: ipilimumab
Targeted Agents
Proteasome inhibitors: bortezomib, ixazomib
BCL-2 inhibitors: venetoclax, navitoclax

■ IMMUNOTHERAPEUTIC APPROACHES

Treatments involving monoclonal antibodies or activated T cells are currently changing the treatment paradigm of ALL. The prerequisite is that B-lineage blast cells express a variety of specific antigens, such as CD19, CD20, and CD22 (Table 111-6) that are targetable with a wide variety of monoclonal antibodies. A new treatment principle is the activation of the patient's T cells to destroy their CD19+ leukemic blasts.

Anti-CD20 The anti-CD20 monoclonal antibody rituximab has improved the outcome of patients with de novo Burkitt's leukemia/lymphoma. With repeated short cycles of intensive chemotherapy combined with rituximab, the overall survival increased to >80%. Rituximab is now included in most B-ALL regimens and is given at the usual dose of 375 mg/m^2 on day −1 before chemotherapy for at

TABLE 111-6 Expression of Antigens in B-Cell Lineage Acute Lymphoblastic Leukemia (ALL) for Potential Antibody Therapy

SURFACE ANTIGEN	ALL SUBTYPES	EXPRESSION ON LBC[a]	MONOCLONAL ANTIBODY
CD20	Burkitt's lymphoma/ leukemia B-precursor	86–100% 30–40%	Rituximab Ofatumumab
CD22	B-precursor Mature B-ALL	93–98% ~100%	Inotuzumab Epratuzumab Moxetumomab pasudotox
CD19	B-precursor Mature B-ALL	95–<100% 94–<100%	T cell–activating therapies Blinatumomab Bispecific CD3/CD19 Chimeric antigen receptor modified T cells (CAR T cells)

[a] Defined as ≥20% positive blast cells.

Abbreviation: LBC, leukemic blast count.

Source: Reproduced with permission from D Hoelzer: Novel antibody-based therapies for acute lymphoblastic leukemia. *Hematology Am Soc Hematol Educ Program* 2011:243, 2011.

least eight or more cycles. This leads to a significant increase in MRD negativity and improved survival.

Anti-CD22 Monoclonal antibodies directed against CD22 are linked to cytotoxic agents, such as calicheamicin (inotuzumab ozogamicin), or to plant or bacterial toxins (epratuzumab). In a randomized trial of relapsed or refractory ALL patients, the CR rate was 66% and significantly superior to the CR rate with standard chemotherapy. Inotuzumab is now included in first-line therapy for Ph+ and Ph– patients.

Anti-CD19 Targeting CD19 is of great interest because this antigen is highly expressed in all B-lineage cells, most likely including early lymphoid precursor cells. A new promising approach is the bispecific antibody blinatumomab, which combines single-chain antibodies to CD19 and CD3, such that T cells lyse the CD19-bearing B cells.

Blinatumomab is particularly effective in MRD-positive patients, with a 70–80% conversion to MRD negativity, translating into improved overall survival; ~25% of MRD-negative patients survived without any further treatment. Blinatumomab has also moved to frontline therapy.

CAR-T Cells The adoptive transfer of CAR-modified T cells directed against CD19 is a promising approach for the treatment of CD19+ childhood or adult ALL. In the first three larger studies in adults with relapsed or refractory ALL, the CR rate ranged from 67 to 91% with MRD negativity in 60–81% of the patients who achieved CR. Overall survival is 50% or more at ≥2 years, which is remarkable for these heavily pretreated patients. CAR-T cells are also effective in CNS leukemia and in other extramedullary sites. CAR-T cell therapy in relapsed or refractory ALL was first considered as a bridge to allogeneic SCT, applied in 10–50% of patients, but the necessity for an allogeneic SCT after CAR-T cells is unclear. CAR-T cell therapies are also moving to the frontline. CD19-negative relapses after CAR-T cell therapy or blinatumomab due to downregulation of CD19 expression are a relevant obstacle.

Toxicities of Immunotherapies The anti-CD22 agent inotuzumab ozogamicin is associated with hepatotoxicity, including veno-occlusive disease, particularly after allogeneic SCT, but can be managed by reduced dosing and limitation of cycles. For anti-CD19 therapies, cytokine release syndrome and severe neurotoxicity are the most prominent toxicities and often require intensive care unit care (more so after CAR-T cells than blinatumomab). Management of these complications has improved with early recognition. Because toxic death after immunotherapies is very low compared to intensive chemotherapy or allogeneic SCT, immunotherapies are now increasingly included in frontline therapy.

TREATMENT OF T-ALL

Immunotherapy for T-ALL is still not available, and intensive chemotherapy is still the mainstay in combination with the T cell–specific drug nelarabine. Currently, γ-secretase targeting *NOTCH1*, checkpoint inhibitors such as bortezomib and venetoclax, and HDAC inhibitors are being explored.

CONCLUSION AND FUTURE DIRECTIONS

Cytogenetic and molecular analysis at diagnosis allows identification of ALL subentities requiring different treatment options. Evaluation of MRD is the most important parameter for treatment decisions. The greatest progress has been achieved by targeted therapies, such as TKIs for Ph+ ALL and new immunotherapeutic approaches. This will lead to further improved outcome of adult ALL patients, 50% of whom are already surviving 5–10 years and are most likely cured. New options and advances, such as low-intensity chemotherapy, reduction of SCT, incorporation of targeted therapies, and reduction of toxicities, will improve the quality of life of patients and lead to individualized approaches for each patient.

FURTHER READING

Caracciolo D et al: The emerging scenario of immunotherapy for T-cell acute lymphoblastic leukemia: Advances, challenges and future perspectives. Exp Hematol Oncol 12:5, 2023.

Gökbuget N et al: Diagnosis, prognostic factors and assessment of ALL in adults: 2023 ELN recommendations from a European expert panel. Blood 143:1891, 2024.

Gökbuget N et al: Management of ALL in adults: 2023 ELN recommendations from a European expert panel. Blood 143:1903, 2024.

Hoelzer D et al: ESMO Clinical Practice Guideline interim update on the use of targeted therapy in acute lymphoblastic leukaemia. Ann Oncol 35:15, 2024.

Prockop S, Wachter F: The current landscape: Allogeneic hematopoietic stem cell transplant for acute lymphoblastic leukemia. Best Pract Res Clin Haematol 36:101485, 2023.

Ribera J-M, Chiaretti S: Modern management options for Ph+ ALL. Cancers 14:4554, 2022.

Shimony S et al: Nelarabine: when and how to use in the treatment of T-cell acute lymphoblastic leukemia. Blood Adv 8:23, 2024.

Short N et al: Using immunotherapy and novel trial designs to optimise front-line therapy in adult acute lymphoblastic leukaemia: Breaking with the traditions of the past. Lancet Haematol 10:e382, 2023.

112

Chronic Lymphocytic Leukemia

Jennifer A. Woyach, John C. Byrd

Chronic lymphocytic leukemia (CLL) is a monoclonal proliferation of mature B lymphocytes defined by an absolute number of malignant cells in the blood (5×10^9/L). The presence of malignant B cells under this count in the blood without nodal, spleen, or liver involvement and absent cytopenias is a precursor of this disease called *monoclonal B-cell lymphocytosis* (MBL) with ~1–2% chance per year of progressing to overt CLL. CLL is a heterogeneous disease in terms of natural history, with some patients presenting asymptomatically and never requiring therapy, whereas most will need therapy and a small subset will present with symptomatic disease, require multiple lines of therapy, and eventually die of their disease. Over the past two decades, the understanding of CLL origin and biology has grown exponentially, leading first to more refined disease definition, prognostic markers, and, subsequently, introduction of novel therapies that have significantly changed the natural history of this disease to where only a small minority will die from CLL. In this way, CLL has served as a prototype of a cancer where understanding the biologic underpinnings in absence of unified mutation drivers has informed therapy development. In this chapter, we review the epidemiology, biology, and management of CLL, with a focus on new knowledge that has changed and continues to change standards of care.

EPIDEMIOLOGY

CLL is primarily a disease of older adults, with a median age at diagnosis of 71 and an age-adjusted incidence of 4.6/100,000 people in the United States; 18,740 peoples were diagnosed and 4490 people died of CLL in the United States in 2023. The prevalence of CLL has increased over the past decades due to improvements in therapy for this disease and also survival of older patients from other medical ailments. In 1980, the 5-year overall survival of patients was 70%, and this increased to 92% in 2015 and is likely even higher today. The male-to-female ratio is 2:1; however, as patients age, the ratio becomes more even, and over the age of 80, the incidence is equal between men and women. The disease is most common in Caucasians, less common in Hispanic and African Americans, and is rare in the Asian population.

Unlike many other malignancies, there have been no definitive links between CLL and exposures. Indeed, CLL is one of the only types of leukemia not linked to radiation exposure. Agent Orange exposure has been implicated, and CLL is thus a service-connected condition for those who were exposed to Agent Orange in the Vietnam conflict, burn pit exposure from the Middle East conflicts, and Camp Lejune water.

CLL is one of the most familial-associated malignancies, and the first-degree relative of a CLL patient has an 8.5-fold elevated risk of developing CLL than the general population. MBL is also more common in families with two first-degree relatives having CLL, further supporting a genetic predisposition of this disease. Despite this, specific genes conferring risk in the familial setting outside of specific families have been difficult to identify. In genome-wide association studies (GWAS), ~30 single nucleotide polymorphisms (SNPs) have been identified, which is estimated to account for 19% of the familial risk of CLL. Genes involved in apoptosis, telomere function, B-cell receptor (BCR) activation, and B-cell differentiation have all been implicated in GWAS. Variants in shelterin complex proteins involved in telomere maintenance such as POT1 have been identified in a small number of families.

BIOLOGY AND PATHOPHYSIOLOGY

CELL OF ORIGIN

The cell of origin in CLL has not definitively been established. The morphology, immunophenotype, and gene expression pattern of CLL cells are that of a mature B cell **(Fig. 112-1)**, and so it has long been presumed that the initiating cell is a mature lymphocyte, perhaps memory B cells. However, many facets of CLL biology do not support this idea, including antigen-binding characteristics of CLL cells and the presence of stereotyped BCRs. Other possibilities include a stepwise process including a series of transforming events at various stages of B-cell development, potentially including de-differentiation of more mature cells. The self-renewing, multipotent hematopoietic stem cell (HSC) might also be the originating cell of CLL, postulated based on transplant studies in mice showing clonal leukemic cell development with different characteristics from donor leukemia after transplantation of HSC from CLL patients. More work will be required to elucidate the origins of CLL.

B-CELL RECEPTOR SIGNALING IN CLL

Perhaps the most important advancement in CLL biology is the understanding of the role of BCR signaling in the disease. CLL has distinct BCR signaling as compared to normal B cells, which is characterized by low-level IgM expression, variable response to antigen stimulation, and tonic activation of antiapoptotic signaling pathways that promote tumor survival. CLL cells by gene expression profiling share many features with antigen-activated mature B cells, suggesting a role for activation of BCR signaling in the disease pathogenesis. Tissue-based microarrays have revealed upregulation of BCR pathway genes in the lymph nodes and bone marrow compared to the peripheral blood, suggesting a particular importance of this pathway in microenvironmental homing.

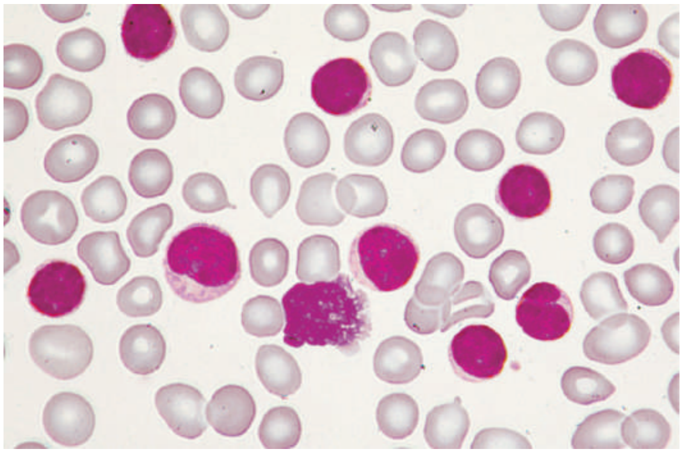

FIGURE 112-1 Chronic lymphoid leukemia in the peripheral blood. *(From Williams Hematology, 7th ed, in M Lichtman et al [eds]: New York, McGraw-Hill, 2005.)*

Fitting with the role of BCR signaling in CLL, one of the most influential prognostic factors identified in this disease is the mutational status of the immunoglobulin heavy chain variable (IGHV) region. During normal B-cell maturation, the variable regions of the immunoglobulin heavy chain undergo somatic hypermutation. In CLL, ~60% of patients have IGHV that is ≥2% mutated from germline. This may indicate a more mature, postgerminal center progenitor, and is typically associated with a more indolent disease course. Conversely, ~40% of patients will have IGHV <2% mutated from germline, which is associated with more rapid progression of disease and short survival before the era of therapeutics that target BCR. Unfavorable biologic properties including enhanced telomerase activity, overexpression of activation-induced cytidine deaminase, increased nuclear factor-κB (NF-κB) activity, high-risk genomic mutations (e.g., *NOTCH1, SF3B1, TP53, ATM*), and clonal evolution are also associated with IGHV unmutated disease. In addition to the mutational status of IGHV, about 30% of CLL patients express "stereotyped" BCRs, where the stereotype subset predicts clinical course, with subsets 1 and 2 predicting higher-risk disease.

CYTOGENETIC ABNORMALITIES

Besides IGHV mutational status, recurrent cytogenetic abnormalities are the most robust prognostic factor clinically available in CLL. These abnormalities are typically identified by fluorescent in situ hybridization (FISH) analysis; however, stimulated metaphase karyotype has a role as well. The most well-characterized abnormalities include del(13)(q14.3), trisomy 12, del(11)(q22.3), and del(17)(p13.1) **(Fig. 112-2)**. The presence of sole del(13)(q14.3) is associated with more indolent disease, prolonged survival, and good response to traditional therapies. Usually this abnormality is not seen on banded karyotype analysis, and when present on karyotype, it indicates a larger deletion involving the retinoblastoma gene, which negates the favorable prognosis associated with this marker. Trisomy 12 has a more intermediate prognosis. The del(11)(q23.3) results in deletion of the *ATM* gene and is associated with bulky lymphadenopathy and aggressive disease in young patients, with inferior prognosis, and more rapid progression to symptomatic disease. The del(17)(p13.1) results in loss of one allele of the tumor suppressor TP53 and is associated with the poorest prognosis in CLL with rapid disease progression, poor response to traditional therapies, and shorter survival. Other abnormalities have been shown to be important in smaller studies but are not routinely performed at all centers. Finally, complex karyotype (three or more abnormalities) on stimulated metaphase karyotype analysis has significant adverse impact on time to treatment and overall survival, with data indicating that increasing complexity is even more deleterious to response and survival.

Clonal evolution, or acquisition of cytogenetic or molecular abnormalities, is common in CLL, especially in patients with IGHV unmutated CLL. Because the cytogenetics of patients can change even in the absence of therapy, it is recommended that FISH, with or without cytogenetics, is checked before every line of therapy, mostly to evaluate acquisition of del(17)(p13.1).

GENE MUTATIONS AND MIR ALTERATIONS

Compared with many other malignancies, the genome in CLL is relatively simple, with an average CLL genome carrying ~20 nonsynonymous alterations and ~5 structural abnormalities. And, unlike many other hematologic malignancies, there is no unifying genetic lesion, and most recurrent genetic driving mutations exist at frequencies of <5%. Whole genome and whole exome sequencing have identified the most common mutations in CLL to be in *SF3B1*, *NOTCH1*, *MYD88*, *ATM*, and *TP53* **(Table 112-1)**. Most of the identified mutations in these genes are common among different malignancies, and with the exception of *MYD88*, they are generally identified with much higher frequency in IGHV unmutated disease.

NOTCH1 mutations are present in ~15% of CLL patients and are commonly associated with trisomy 12. Although multiple different

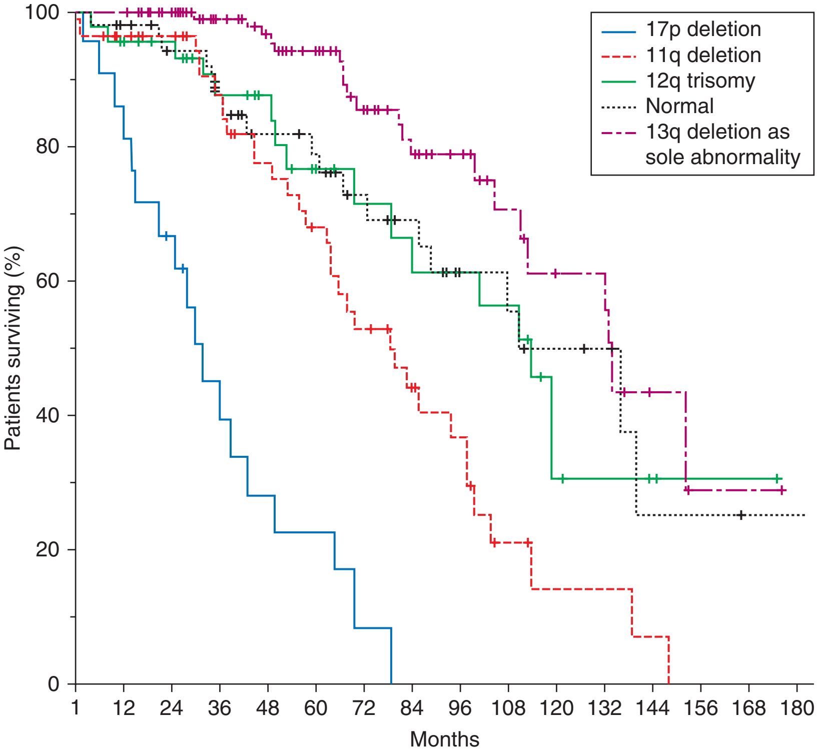

No. at Risk																
17p deletion	23	18	13	8	5	4	1	0	0	0	0	0	0	0	0	0
11q deletion	56	53	47	43	33	27	20	15	10	4	2	2	1	0	0	0
12q trisomy	47	44	41	29	24	17	14	13	12	11	4	3	2	1	1	0
Normal	57	51	45	37	30	27	20	17	12	11	6	5	2	2	1	1
13q deletion as sole abnormality	117	117	106	91	80	63	45	36	24	16	12	11	3	1	1	0

FIGURE 112-2 Outcomes among CLL patients with various cytogenetic abnormalities. *(From The New England Journal of Medicine, Genomic Aberrations and Survival in Chronic Lymphocytic Leukemia, H Dohner et al: 343: 1910. Copyright @2000 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

mutations are seen, most are located within the PEST (proline, glutamic acid, serine, and threonine) domain and result in constitutive NOTCH signaling. *NOTCH1* mutations have been associated with lower sensitivity to CD20 antibody therapy and increased risk of transformation to aggressive diffuse large B-cell lymphoma (DLBCL; Richter's transformation), although its relevance in the era of targeted therapies is less clear. *SF3B1* is a component of the RNA spliceosome and is mutated in 10–15% of CLL cases. Mutations appear to be associated with intermediate-risk disease, and functionally, *SF3B1* may be important in the response to DNA damage.

Mutations of the tumor suppressor *TP53* are found in ~5% of CLLs in previously untreated early-stage disease and up to 40% in later stages. Seventy percent of the time, these mutations coexist with del(17)(p13.1), effectively eliminating *TP53* function. As expected, and consistent with other malignancies, *TP53* mutations are associated with a poor prognosis and expected lack of response to DNA-damaging therapies.

TABLE 112-1 Recurrent Mutations in CLL

GENE	FREQUENCY OF MUTATIONS (%)
SF3B1	8–14
TP53	5–13
NOTCH1	10–13
MYD88	4–8
ATM	8–11
BIRC3	<5
XPO1	<5
FBXW7	<5
POT1	<5
BRAF	<5
EGR2	<5
IKZF3	<5

Abbreviation: CLL, chronic lymphocytic leukemia.

ATM mutations, which are heterogeneous and occur throughout the gene, occur in 10–15% of CLL patients. *ATM* mutations often coexist with del(11)(q22.3), eliminating *ATM* on the alternate allele. Similar to *TP53*, mutations in *ATM* tend to result in impaired response to DNA damage, which can reduce responsiveness to chemotherapy.

In contrast to the aforementioned mutations, those in *MYD88* tend to occur in IGHV-mutated CLL and be associated with a more indolent prognosis. This gene is involved in Toll-like receptor signaling, and the most common mutation, L265P, results in constitutive activation and NF-κB activity.

Along with abnormalities in coding genes, it has become apparent that noncoding genes such as microRNAs are recurrently altered in CLL. The most common cytogenetic abnormality, del(13)(q14.3), results in loss of the miR15/16 cluster, which is important in the pathogenesis of CLL. In normal cells, miR15A/miR16A inhibits antiapoptotic gene expression (including *BCL2*, *CCND1*, *CCND3*, and *CDK6*), and this specific deletion allows for overexpression of these genes and thus increased cell survival. Loss of other miR expression such as miR-181a leads to overexpression of proteins such as the antiapoptotic genes *MCL1* and *TCL1*. Overexpression of miR-155, an onco-miR associated with B-cell transformation, has also been documented in the majority of CLL patients.

■ IMMUNOLOGY

CLL is characterized by dysregulation of the normal immune system in addition to the malignant immune cells. Besides numerical abnormalities due to bone marrow dysfunction, even in the early stages of disease, there are skewed ratios of immune cells and functional abnormalities.

Innate immune system defects associated with CLL include reduced complement proteins and activity, qualitative neutrophil defects, and functional defects of natural killer cells.

More focus has been placed on the impairments in the adaptive immune system in this disease. Within the CD4+ T-cell compartment, a qualitative defect is noted similar to chronic antigen stimulation inducing a phenotype of T-cell exhaustion typical of what is seen in chronic viral infections such as hepatitis. This has been demonstrated to lead to impaired T-cell cytotoxic capacity and reduced proliferative ability. Additionally, there are physical changes in the T-cell cytoskeleton that cause impaired immune synapse formation with antigen presenting cells. In addition to a lack of capacity to respond to pathogens, the T-cell defect in CLL also likely leads to tumor cell tolerance. During the course of the disease, the polarization of the CD4+ T cells shifts from a Th1 (cytotoxic) phenotype to a Th2 phenotype, which leads to expansion of immunosuppressive cytokines such as interleukin (IL) 10. Additionally, in the later stage of disease, T regulatory cells are expanded, which contributes to an immunosuppressive phenotype.

Other components of the immune microenvironment are altered as well to form a more supportive environment for the malignant cells. M2 monocytes have been shown to differentiate into a type of tumor-associated macrophage known as a *nurse-like cell* in CLL. These cells promote survival by secreting chemokines and cytokines that increase migration and activation.

The humoral immune system in CLL is also dysregulated, as is expected for a malignancy that results in very few normal B cells. Hypogammaglobulinemia is very common and affects all subclasses of immunoglobulins, occurring in ~85% of patients at some time in their disease course, and is more common as disease progresses. A correlation between low IgG and IgA and infection risk has been established, but isolated IgM reduction does not seem to be associated with excess infection risk. Also, CLL cells can secrete monoclonal IgM or IgG in a small number of cases, and this can correlate with disease progression.

CLINICAL PRESENTATION AND DIAGNOSIS OF CLL

CLINICAL PRESENTATION AND DIAGNOSIS

The presentation of CLL most commonly occurs as an incidental diagnosis made at the time of medical evaluation for another cause. In this regard, CLL is most commonly diagnosed on routine blood work demonstrating an elevated lymphocyte count in asymptomatic individuals, although some patients present with symptoms and require early therapy. When noting either an elevated total white blood cell (WBC) count with lymphocytic predominance or a normal WBC with a differential showing a lymphocytosis, the next step is to perform flow cytometry on the peripheral blood. In CLL, this will reveal the typical immunophenotype that includes the typical B-cell markers CD19, CD20, CD22, CD23, CD200, the T-cell marker CD5 (CD5 is also expressed on the B1 subset of B cells that typically has unmutated immunoglobulin and responds to antigens independent of cognate T-cell help), and dim surface immunoglobulin of either kappa or lambda type **(Table 112-2)**. Atypical phenotypes can be seen as well and usually can be differentiated on the basis of morphology, cytogenetics, or clinical presentation. In cases in which the clonal B-cell count based on flow cytometry is $\geq 5 \times 10^9$/L, no further workup is needed to confirm the diagnosis of CLL.

Some patients will present with a small clonal proliferation of CLL cells in the peripheral blood but will also have lymphadenopathy or splenomegaly. In these cases, the likely diagnosis is small lymphocytic lymphoma (SLL), a semantic designation from CLL that denotes a primarily tissue-based disease rather than bone marrow/blood-based disease. The genetic and molecular features of SLL are identical to those of CLL. The retention of the cells in tissues may be related to the expression of a particular adhesion molecule. Thus, SLL patients are managed identically to CLL patients, and often in the later stages of disease, these patients will have blood and bone marrow involvement as well.

MONOCLONAL B-CELL LYMPHOCYTOSIS

Patients who do not meet the diagnostic criteria for CLL based on quantification of clonal B cells in the peripheral blood and who do not have associated signs of CLL including lymphadenopathy, organomegaly, or cytopenias have a disorder known as monoclonal B-cell lymphocytosis (MBL), which is now thought to precede every case of CLL. Analogous to monoclonal gammopathy of uncertain significance (MGUS) in myeloma, not all MBL progresses to CLL. MBL is initially characterized by a CLL-like immunophenotype in ~75% of cases but can also be atypical (CD23 negative or bright CD20) or CD5 negative. More relevant for prognosis is characterization by count, with low-count MBL defining patients with $<0.5 \times 10^9$ clonal B cells/L and high-count MBL defining patients with $>0.5 \times 10^9$ but $<5 \times 10^9$/L. Patients with low-count MBL have a negligible rate of progression to CLL, whereas those with high count progress to overt CLL at a rate of 1–2% per year, warranting continued monitoring. Population-based studies have estimated the prevalence of MBL up to ~12% in the general population, where it is most common in elderly men. It is especially common in first-degree relatives of CLL patients, where the frequency is ~18%.

Although the risk of MBL progression is relatively low, it has become apparent that patients still experience complications that suggest an immune dysfunction in MBL that is similar to that seen with CLL. Rates of serious infections requiring hospitalization appear to be significantly increased in MBL, similar to the rates seen in CLL. In a case-control study, patients with MBL had a 16% chance of hospitalization over a 4-year time period, compared with 18.4% in patients with newly diagnosed CLL. Secondary cancers also appear to be increased in MBL. These data suggest that monitoring for patients with MBL should focus on vaccinations and age-appropriate cancer screening, as the probability of complications appears to be higher than the risk of progression in most of these patients. Follow-up for patients with MBL can occur with the primary care physician as this does not represent a malignancy, whereas CLL is mostly co-managed with both a primary care physician and a hematologist.

COMPLICATIONS OF CLL

A significant amount of morbidity and mortality related to CLL is due to complications of the disease. In general, complications besides disease progression include infections, secondary cancers, autoimmune complications, and transformation to a more aggressive clonally related lymphoma.

INFECTIONS

Infections are a leading cause of both disease-related morbidity and death in patients with CLL, with ~30–50% of deaths in CLL patient attributed to infection. Owing to the immune dysfunction associated with the disease, patients are at risk for both typical and atypical infections. Besides this baseline risk of infections, most CLL therapies, even targeted therapies, can increase infection risk. Viral prophylaxis

TABLE 112-2 Typical Immunophenotype of CLL Compared with Other B-Cell Malignancies

DISEASE	CD5	CD10	CD19	CD20	CD23	CYCLIN D1	SURFACE IG
CLL	+	–	+	+ (dim)	+	–	+ (dim)
Mantle cell lymphoma	+	–	+	+ (mod/bright)	–	+	+ (mod/bright)
Marginal zone lymphoma	–/+	–	+	+ (mod/bright)	–/+	–	+ (mod/bright)
Follicular lymphoma	–	+	+	+	+	–	

Abbreviation: CLL, chronic lymphocytic leukemia.

is also indicated for many patients when therapy is initiated (even with targeted agents) and for patients with a history of varicella-zoster to diminish reactivation and morbidity from this virus.

Because of the abnormalities in cellular and humoral immunity, vaccine responses in CLL are limited in many patients, especially in the later stages of disease. In one study, one dose of 13-valent pneumococcal vaccine produced an adequate immune response in only 58% of patients compared with 100% in age-matched controls. Vaccine efficacy can be improved in CLL patients by repeated booster vaccinations, vaccine adjuvants, and protein conjugation. Despite the known limitations, vaccination against influenza, COVID-19, varicella-zoster, pneumococcal pneumonia (Prevnar 20), and respiratory syncytial virus is recommended in CLL. The recombinant zoster vaccine has approximately a 60% response in previously untreated CLL, is safe, and should be considered for this patient group. Efficacy of the newer 20-valent pneumococcal vaccine in patients with CLL has not yet been reported. In contrast, live vaccines should be avoided in the setting of CLL because of the small risk of viral reactivation with an immunocompromised host.

Vaccine effectiveness in terms of humoral response is also decreased by most CLL therapies, although this can be overcome in some cases by booster vaccinations. When possible, receiving vaccinations before the initiation of therapy is recommended.

As discussed earlier, hypogammaglobulinemia is common in CLL and can be associated with significant risk for infections, primarily of mucocutaneous etiology such as sinusitis and bronchitis. In addition, women can have frequent urinary tract infections. While administration of prophylactic intravenous immunoglobulin (IVIg) has not been shown to improve survival, it has been shown to reduce the number of minor or moderate bacterial infections, and thus is indicated in patients with hypogammaglobulinemia who suffer from recurrent infections or have pulmonary bronchiectasis. It is also our practice to administer at least one dose of immunoglobulin to CLL patients who develop influenza with coexisting hypogammaglobulinemia to diminish risk of postinfluenza pneumococcal pneumonia. IVIg is also indicated in patients who have been hospitalized for a serious infection and in those whose IgG level is <300–500 mg/dL.

SECONDARY MALIGNANCIES

Multiple population-based studies have shown that patients with CLL are at an elevated risk of developing other cancers, with a rate up to three times that of the general population, even in the absence of cytotoxic chemotherapy. The most common types of cancers seen in CLL are skin, prostate, and breast cancers, although other cancers are seen as well. Skin cancers are particularly common, with a rate of 8- to 15-fold higher than the general population, and may behave more aggressively. All CLL patients should be counseled on the use of sunscreen while outdoors and should undergo preventative skin examinations at least yearly.

In one single-center study, older age at CLL diagnosis, male sex, high β_2-microglobulin, high lactate dehydrogenase (LDH), and chronic kidney disease were associated with excess risk of other cancers; other CLL-specific risk factors have not shown association with other cancer risk.

While cancer risk is higher, there are no specific recommendations for increased cancer screening in CLL patients. Age- and sex-appropriate screenings should be recommended. In addition, we extend screening beyond age 70–75 years for CLL patients based on the higher frequency of cancers.

Conflicting data exist regarding the risk of cancers after CLL-specific therapy. Chemoimmunotherapy, in particular alkylator-containing regimens, seems to be associated with an increased risk for secondary cancers. Secondary cancers are also seen in the setting of targeted therapies. Bruton tyrosine kinase (BTK) inhibitors appear to have a secondary cancer risk similar to what is seen in the CLL population in general, but potentially a higher rate of nonmelanoma skin cancers. Compared with chemoimmunotherapy, venetoclax plus obinutuzumab has been associated with a numerically but not statistically higher rate of secondary cancers at follow-up of 6 years.

AUTOIMMUNE COMPLICATIONS

Autoimmune complications are frequent in CLL. Most commonly, these include autoimmune cytopenias, but autoimmune complications of other organs including glomerulonephritis, vasculitis, and neuropathies have also been reported. Of the autoimmune cytopenias, the most common is autoimmune hemolytic anemia (AIHA), which is an antibody-mediated destruction of autologous red blood cells (RBCs). Second most common is immune thrombocytopenia (ITP), which shares some features with AIHA and has a similar mechanism targeting platelets. These two syndromes may occur in isolation, sequentially in the same patient, or present in combination as Evan's syndrome. Pure red cell aplasia (PRCA) and autoimmune granulocytopenia (AIG) are comparatively rare and can occur alone or in combination with other autoimmune cytopenias. It is difficult to tease out whether autoimmune cytopenias lead to worse prognosis in CLL because of various complicating factors. However, it is clear that these can lead to significant morbidity, both due to the process itself and due to therapies required for management.

AIHA usually presents as an isolated anemia with an elevated reticulocyte count and features of hemolysis including elevated bilirubin and LDH and low haptoglobin. Detection of a warm IgG antibody on the surface of RBCs with a Coombs test can help solidify the diagnosis, although Coombs-negative cases can occur. Immediate therapy is almost always necessary and consists of transfusion and immunosuppression. Glucocorticoids are often used for initial therapy, although in most cases, additional treatment is needed due to either poor response or recurrence with taper of steroid dosing. Rituximab can be successful, and therapy directed toward the underlying CLL is often effective in more resistant cases. Transfusion of blood in cases of robust AIHA must be initiated with caution as transfusion reactions can be seen due to poorly matched blood, but it should be pursued in those with severe, symptomatic anemia. Death from uncontrolled AIHA can occur in the absence of appropriate supportive care.

ITP can be more difficult to diagnose, as it may be difficult to differentiate from progression of disease due to the lack of laboratory tests that identify platelet destruction from this mechanism. Signs that point toward ITP include isolated thrombocytopenia and rapid decline in platelet levels in the absence of an alternative etiology. A bone marrow biopsy showing normal or increased megakaryocytes can be used to confirm the diagnosis but is often not necessary. In CLL, treatment for ITP is usually instituted when platelet levels drop to 20,000–30,000 or if there is evidence of bleeding complications or need for invasive procedures. Like AIHA, initial therapy consists of glucocorticoids and IVIg, with rituximab also being an effective method to induce long-term remissions. Also, the thrombopoietin receptor agonists romiplostim and eltrombopag are effective in secondary ITP. In many cases, ITP can be successfully treated without treating the underlying CLL. In cases in which anemia or thrombocytopenia appear, it is important to investigate the mechanism because the approach to therapy of autoimmune cytopenias in CLL differs from that for cytopenias due to marrow replacement.

RICHTER'S TRANSFORMATION

One of the most devastating complications of CLL is Richter's transformation, transformation of CLL to an aggressive lymphoma, most commonly DLBCL. The World Health Organization also recognizes Hodgkin's lymphoma (HL) as a variant of Richter's transformation; other aggressive lymphomas are rarely identified. Some older series have included prolymphocytic transformation in this category, although this has much less prognostic impact on long-term outcome. The prevalence of Richter's transformation is difficult to estimate based on previous studies, but one prospective observational study estimated a rate of 0.5% per year for DLBCL and 0.05% per year for HL. Risk factors for development include bulky lymphadenopathy, *NOTCH1* mutations, del(17)(p13.1), and a specific stereotyped IGVH usage. Lymphomas arising in the setting of CLL can either be clonally related or unrelated to the initial CLL, with prognosis significantly better for clonally unrelated lymphomas. In addition, patients with Hodgkin's transformation have improved outcome when treated with standard

Hodgkin's disease treatment. B-cell prolymphocytic leukemia (PLL) arising from CLL is currently classified as Richter's transformation as well; however, clinical features and therapy are quite different, so these two should be differentiated for therapeutic purposes.

Clinical signs of Richter's transformation include rapid progression in adenopathy, often in a specific area, and constitutional symptoms including fatigue, night sweats, fever, and weight loss. LDH is usually high. In suspected cases, the first step is ^{18}FDG-PET/CT (fluorodeoxyglucose–positron emission tomography combined with computed tomography) scan to localize an area for biopsy. Standardized uptake values (SUVs) <5 are consistent with CLL and can rule out Richter's transformation in many cases. SUVs >5 are suspicious for Richter's transformation, with SUVs ≥10 being very concerning. Excisional biopsy is the preferred mode of diagnosis, and fine-needle aspiration should be discouraged.

Therapy for DLBCL Richter's transformation usually involves combination chemoimmunotherapy. Outcomes are poor, with median survivals of 6–16 months in most series for clonally related Richter's versus ~5 years for clonally unrelated. This highlights an area of unmet need in CLL therapy and an area of active investigation. Intensive chemotherapy is ineffective for most patients and results in significant toxicity. For fit patients who achieve a response with therapy, stem cell transplantation has the possibility to induce long-term remissions and should be explored. In addition, chimeric antigen receptor (CAR) T-cell (CAR-T) therapy has shown promising results in small groups of patients and remains an area of active clinical investigation. Limited data using bispecific antibodies are also encouraging. Patients with Hodgkin's disease can be treated according to the algorithm for this disease, with many individuals being cured.

WORKUP OF CLL AND APPROACH TO THERAPY

WORKUP AND STAGING

Workup of a patient with new diagnosis of CLL based on typical immunophenotyping includes a detailed history of infectious disease; family history of CLL; and careful physical examination with attention to the lymph nodes, spleen, and liver. In patients desiring to know the expected natural history of their CLL, prognostic testing using FISH and stimulated karyotype and sequencing for *TP53* and IGHV mutation status can be performed. Imaging with CT scan is usually not necessary unless there are symptoms and concern for intraabdominal nodes out of proportion to peripheral nodes. Bone marrow biopsy is not undertaken until therapy is initiated, except in cases of unexplained cytopenias.

STAGING

There are two widely used staging systems in CLL: The Rai staging system is used more commonly in the United States, whereas the Binet system is more commonly used in Europe. Both characterize CLL on the basis of disease bulk and marrow failure **(Table 112-3)**. Both rely on physical examination and laboratory studies and do not require imaging or bone marrow analysis. While the initial staging systems could reliably predict survival in CLL, with the changes in therapy since the original description of the stages, the impact of initial stage on survival is not as clear. Cytogenetic and genomic testing can help refine outcome of these staging tests. An international collaboration integrated both clinical and genomic staging to better predict outcome at diagnosis and time of initial treatment, which led to development of the CLL International Prognostic Index **(Table 112-4)**. This index has been shown to be useful in prediction of both time to first treatment and outcome with chemoimmunotherapy. Validation in the setting of novel targeted therapies has not occurred.

TABLE 112-3 Staging of CLL

Rai Staging System	
Low risk (stage 0)	Lymphocytosis only
Intermediate risk (stage I/II)	Lymphocytosis with lymphadenopathy, with or without splenomegaly or hepatomegaly
High risk (stage III/IV)	Lymphocytosis with anemia or thrombocytopenia due to bone marrow involvement
Binet Staging System	
A	<3 areas of lymphadenopathy
B	≥3 areas of lymphadenopathy
C	Hemoglobin ≤10 g/dL and/or platelets <100,000/μL

Abbreviation: CLL, chronic lymphocytic leukemia.

TABLE 112-4 CLL International Prognostic Index

Risk Score

VARIABLE	ADVERSE FACTOR	RISK SCORE
TP53 status	Deleted or mutated	4
IGHV mutational status	Unmutated	2
β_2-Microglobulin concentration	>3.5 mg/L	2
Clinical stage	Rai I–IV or Binet B–C	1
Age	>65 years	1

Implications of Risk Score

RISK SCORE	RISK CLASSIFICATION	5-YEAR SURVIVAL (TRAINING SET DATA)
0-1	Low	93.2%
2-3	Intermediate	79.3%
4-6	High	63.3%
7-10	Very high	23.3%

CRITERIA FOR THE INITIATION OF THERAPY

Currently, a watchful waiting strategy is used for most patients with CLL, with therapy reserved for patients with symptomatic disease. This recommendation is based on multiple trials showing no survival advantage with earlier therapy, although this question continues to be a focus of active investigation.

With the exception of patients participating in early intervention studies in CLL, disease-related symptoms that require the initiation of therapy are outlined in **Table 112-5**. Except for the rare patient who presents with disease requiring urgent therapy, in most cases, these symptoms can be monitored over short periods to determine relatedness to CLL and need for therapy.

INITIAL THERAPY FOR CLL

Over the past decade, the initial therapy of CLL has dramatically changed. Whereas chemoimmunotherapy was once standard for all patients, now most patients are treated with oral therapies targeted

TABLE 112-5 Criteria for the Initiation of Therapy

Symptoms Indicating Need for Therapy in CLL
Evidence of progressive marrow failure (worsening of anemia or thrombocytopenia not due to autoimmune destruction)
Massive (≥6 cm below costal margin), progressive, or symptomatic splenomegaly
Massive (≥10 cm), progressive, or symptomatic lymphadenopathy
Progressive lymphocytosis with an increase of ≥50% over a 2-month period or lymphocyte doubling time <6 months
Autoimmune anemia or thrombocytopenia not responsive to standard therapy
Symptomatic or functional extranodal involvement
Constitutional symptoms (one or more of the following: unintentional weight loss ≥10% over 6 months, significant fatigue, fevers ≥100.5°F for 2+ weeks without infection, night sweats for >1 month without infection)

Abbreviation: CLL, chronic lymphocytic leukemia.

against BTK or BCL2 with an anti-CD20 monoclonal antibody. This continues to be an area of active investigation, with standards of care shifting rapidly. The major classes of these therapies will be outlined here.

BTK Inhibitors BTK is an attractive target in CLL because, unlike other kinases in the BCR pathway, BTK does not have natural redundancy and is relatively selective for B cells, so inhibition leads to a predominant B-cell–specific phenotype. The first-in-class covalent BTK inhibitor is ibrutinib, which is relatively selective for BTK but also inhibits a number of structurally similar kinases. As initial therapy, ibrutinib was initially compared with chlorambucil (RESONATE study), and there was an 84% lower risk of progression or death with ibrutinib, with 59% of ibrutinib-treated patients alive and progression-free at 7 years. Subsequent studies have compared ibrutinib alone or with the anti-CD20 antibody rituximab to standard chemoimmunotherapy with fludarabine plus cyclophosphamide plus rituximab (FCR) or bendamustine plus rituximab (BR) and shown superiority of targeted therapy to chemoimmunotherapy. Ibrutinib also has immune modulatory roles to expand and also enhance T-cell function that may improve CAR-T efficacy. Side effects noted to occur with this class of agents include arthralgias/myalgias, rash, diarrhea, dyspepsia, bruising/bleeding (particularly when on antiplatelet/anticoagulation therapy or with surgery), hypertension, atrial fibrillation, and ventricular arrhythmias.

Two second-generation covalent BTK inhibitors, acalabrutinib and zanubrutinib, were developed to be more specific for BTK than ibrutinib and consequentially show better tolerability. Two trials performed in relapsed/refractory CLL have positioned acalabrutinib and zanubrutinib as preferred BTK inhibitors compared with ibrutinib. Acalabrutinib was compared head-to-head with ibrutinib in previously treated patients with high-risk relapsed CLL in the ELEVATE-RR trial. Acalabrutinib was shown to be noninferior to ibrutinib in terms of efficacy and to have lower rates of atrial fibrillation, hypertension, myalgias/arthralgias, bruising, and skin and nail changes than reported with ibrutinib. Zanubrutinib was compared head-to-head with ibrutinib in the ALPINE trial, which enrolled patients with relapsed CLL without regard to risk status. In this study, zanubrutinib was shown to be superior to ibrutinib in terms of overall response rate and progression-free survival (PFS). Zanubrutinib was also associated with lower rates of atrial fibrillation but similar rates of other adverse events.

Because acalabrutinib and zanubrutinib are newer, follow-up is shorter than with ibrutinib, but all three covalent BTK inhibitors appear to be similarly active in the frontline setting. Acalabrutinib was studied in the treatment-naïve setting in the ELEVATE-TN trial, where acalabrutinib resulted in a 6-year PFS of 62% and acalabrutinib given with obinutuzumab resulted in a 4-year PFS of 78%. Zanubrutinib was studied in treatment-naïve CLL in the SEQUOIA study, where 2-year PFS was 85.5% for patients treated with zanubrutinib.

Importantly, with the follow-up that is available from frontline studies of covalent BTK inhibitors in CLL, traditional prognostic factors, including IGHV mutational status and FISH, have less impact on outcome. Indeed, age and performance status were the only variables that predicted survival in the ELEVATE-TN trial.

BCL2 Inhibitor Venetoclax is an orally bioavailable, selective allosteric inhibitor of the antiapoptotic protein BCL2, which is upregulated in CLL. Similar to the BTK inhibitors, phase 3 trials support the frontline use of venetoclax in combination with obinutuzumab (VO) compared with chemoimmunotherapy. The CLL14 study compared VO versus chlorambucil plus obinutuzumab in previously untreated patients with coexisting medical conditions. Unlike BTK inhibitors, which are continuously administered until disease progression, VO treatment is administered for a fixed duration of 1 year. Median PFS for VO was 76.2 months, compared with 36.4 months for patients treated with chlorambucil plus obinutuzumab. The CLL13/GAIA trial also studied VO, as well as VO plus ibrutinib (IVO) and venetoclax plus rituximab (VR), compared with effective chemoimmunotherapy regimens (FCR for younger patients and BR for older patients) in fit patients. At 4 years, VO and IVO showed superior PFS compared with chemoimmunotherapy (81.8% and 85.5%, respectively, vs 77.2%), but VR was not superior to chemoimmunotherapy. Side effects associated with venetoclax include tumor lysis syndrome, neutropenia, and nausea/vomiting/diarrhea. Importantly, genomic risk features including IGHV mutational status and FISH/cytogenetics appear to be more relevant to PFS with fixed-duration VO therapy as compared with indefinite BTK inhibitor therapy.

Targeted Therapy Combinations Due to synergy between BTK and BCL2 inhibition, there has been considerable interest in studies combining agents with these two mechanisms to allow for fixed-duration therapies. As described earlier, 1-year fixed-duration IVO was shown to be superior to chemoimmunotherapy in previously untreated fit patients in the CLL13/GAIA study. The GLOW study is a registration trial comparing ibrutinib plus venetoclax (IV) to chemoimmunotherapy in previously untreated older or unfit patients. In this trial with 1 year of IV, 42-month PFS was 74.6% for IV compared with 24.8% for chlorambucil plus obinutuzumab. Although this study led to regulatory approval of IV in Europe, it is not approved by the U.S. Food and Drug Administration (FDA). It is not yet clear whether IV or IVO is a more effective therapy than the other fixed-duration standard, VO. Current studies are also using more selective inhibitors of BTK in an attempt to improve efficacy and safety of these combinations.

Chemoimmunotherapy For the most part, targeted therapy has supplanted chemoimmunotherapy in CLL. However, long-term follow-up of studies of FCR has demonstrated that a subset of patients treated with this regimen can have durable responses over 20 years, with a likely cure of CLL. This group is composed almost exclusively of patients with mutated *IGVH* and favorable cytogenetics. However, despite the efficacy of this regimen, short- and long-term toxicities limit its adaptability to many patients with IGHV-mutated disease. Short-term toxicities are mostly related to myelosuppression and include neutropenia and infection. Long-term cytopenias are less common, but they do occur. Also, there is about a 3–5% risk of therapy-related myeloid neoplasm with this regimen that is almost always fatal. In the E1912 study of FCR versus ibrutinib plus rituximab (IR), at follow-up, there was no difference in PFS or overall survival between FCR and IR for patients with mutated IGHV, suggesting that a place for this regimen may remain in clinical practice. In addition, current studies are focused on limiting chemotherapy and/or adding novel agents in efforts to achieve cure but limit toxicity.

THERAPY OF RELAPSED CLL

Currently, the mainstays of treatment for relapsed CLL are the same classes as initial therapy, and the choice of second-line therapy is heavily dependent on the agent that was used in the frontline setting. The optimal sequencing of targeted agents in CLL has not been established; however, the available data suggest that the sequence of either BTK inhibitor and then BCL2 inhibitor and the reverse are both acceptable. In addition, some patients treated initially with venetoclax-based regimens can likely be successfully retreated with venetoclax. In a trial of venetoclax for patients who had relapsed after ibrutinib therapy, overall response rate (ORR) was 65%, with a median PFS of ~2 years, in a very heavily pretreated patient population. Retrospective data of a BTK inhibitor given after venetoclax suggest that this sequence is also effective, with an ORR of 84% and median PFS of 32 months. PI3K inhibitors also have activity in relapsed CLL; however, activity following both BTK and BCL2 inhibitors is likely minimal. In addition, many new agents are in development in CLL, including novel oral targeted therapies, antibodies, and immune-based treatments.

Noncovalent Inhibitors of BTK Despite the activity of covalent BTK inhibitors in CLL, a subset of patients will eventually relapse, and the primary mechanism of acquired resistance to ibrutinib, acalabrutinib, and zanubrutinib is acquisition of a mutation at the binding site of the drug (predominantly *BTK* C481S). Noncovalent BTK inhibitors

TABLE 112-6 Response Criteria in CLL

	LYMPHOCYTE COUNT	LYMPH NODES[a]	SPLEEN/LIVER SIZE[b]	BONE MARROW[c]	PERIPHERAL BLOOD COUNTS
CR	<4000/μL	None >1.5 cm	Not palpable	Normocellular, <30% lymphocytes, no B lymphoid nodules	• Platelet count >100,000/μL • Hemoglobin >11 g/dL • Neutrophils >1500/μL
PR	Decrease ≥50% from baseline	Decrease ≥50% from baseline	Decrease ≥50% from baseline	Infiltrate ≤50% of baseline	One of the following: • Platelet count >100,000/μL or ≥50% from baseline • Hemoglobin >11 g/dL or ≥50% from baseline • Neutrophils >1500/μL or ≥50% from baseline
Stable disease	Not meeting CR/PR/PD criteria	Not meeting CR/PR/PD criteria	Not meeting CR/PR/PD criteria	Not meeting CR/PR/PD criteria	Not meeting CR/PR/PD criteria
PD	Increase ≥50%	Increase ≥50%	Increase ≥50%		• Platelet count ≤50% of baseline due to CLL • Hemoglobin decrease >2 g/dL due to CLL

[a]Refers to sum of the products of multiple lymph nodes evaluated by CT scan. [b]Based on physical examination. [c]Bone marrow only required to confirm CR.

Abbreviations: CLL, chronic lymphocytic leukemia; CR, complete response; PD, progressive disease; PR, partial response.

have been developed to overcome this resistance mechanism by binding both reversibly and to alternative sites on BTK. Pirtobrutinib is the first in class of these agents and is an extremely selective inhibitor of BTK that has achieved accelerated approval by the U.S. FDA for use in patients who have been previously treated with both BTK and BCL2 inhibitors. This is based on the BRUIN study, where an overall response rate of 82.2% and median PFS of 19.6 months were seen in a heavily pretreated high-risk group of patients, with equivalent efficacy noted for patients with C481S mutations in *BTK*.

Immune Therapies Immune therapies in CLL are currently focused in the relapsed setting and include allogenic stem cell transplantation, CAR-T therapy, and bispecific antibodies.

Stem cell transplantation is a curative approach to CLL. Because most CLL patients are older and many have significant comorbidities, myeloablative transplants incur extensive morbidity and mortality, making them prohibitive in many individuals. Reduced-intensity conditioning (RIC) allogeneic transplants have been successfully incorporated into the treatment of patients up to ~75 years in age but still have a ≥50% frequency of chronic graft-versus-host disease. This is still considered a standard treatment in CLL but has fallen out of favor with the introduction of well-tolerated novel agents, as well as clinical trials of CAR-T therapy. CD19 CAR-T trials have not been as successful in CLL as they have in other B-cell malignancies, due to the immunosuppression associated with the disease. The most robust data have come with lisocabtagene maraleucel (liso-cel), where in the TRANSCEND CLL 004 trial, the ORR was 43%, with a median PFS of 11.9 months in patients previously treated with BTK inhibitors and venetoclax. Many current trials are focused on optimizing CD19 CAR-T by adding agents such as BTK inhibitors or PI3K inhibitors or modifying the CAR-T structure, and other studies are testing different targets outside of CD19. This area remains a focus of intense investigation in CLL.

ASSESSING RESPONSE TO THERAPY AND MINIMAL RESIDUAL DISEASE IN CLL

Following the completion of therapy or during therapy for indefinite targeted agents, response is initially assessed using physical examination and laboratory studies (Table 112-6). If residual disease is not detected using these methodologies, CT scans are used to assess response. Bone marrow biopsies with flow cytometry are indicated if no disease is detected to confirm CR.

It has been established in various malignancies that complete tumor eradication is associated with longer survival. In CLL, if no malignant cells can be detected in the bone marrow down to a level of 1 CLL cell in 10^4 leukocytes (0.01%), the patient is said to be negative for minimal residual disease (MRD). Following combination chemoimmunotherapy, eradication of MRD correlates with long-term survival and potentially cure in a subset of patients receiving FCR chemoimmunotherapy. Undetectable MRD in blood or bone marrow is also associated with improvement in PFS in venetoclax-based regimens. However, eradication of MRD has not been shown to be a meaningful endpoint with BTK inhibitors as monotherapy. Higher sensitivity of 1 CLL in 10^6 leukocytes (0.0001%) can be obtained using next-generation sequencing methods such as ClonoSeq. Treatment of MRD-relapsing CLL is being explored as part of clinical trials.

CONCLUSION

CLL is treated only when it becomes symptomatic. At the time of therapy, FCR chemoimmunotherapy in a small subset of young patients with very good risk CLL is potentially curative. In the majority of patients with symptomatic CLL, targeted therapy directed at BTK and/or BCL2 can produce durable remissions and allow patients many years of disease-free survival.

FURTHER READING

Brown JR et al: Zanubrutinib or ibrutinib in relapsed or refractory chronic lymphocytic leukemia. N Engl J Med 388:319, 2023.

Byrd JC et al: Acalabrutinib versus ibrutinib in previously treated chronic lymphocytic leukemia: Results of the first randomized phase III trial. J Clin Oncol 39:3441, 2021.

Fischer K et al: Venetoclax and obinutuzumab in patients with CLL and coexisting conditions. N Engl J Med 380:2225, 2019.

Hallek M et al: iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. Blood 131:2745, 2018.

Landau DA et al: Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell 152:714, 2013.

Mato AR et al: Pirtobrutinib after a covalent BTK inhibitor in chronic lymphocytic leukemia. N Engl J Med 389:33, 2023.

Puente XS et al: Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature 475:101, 2011.

Sharman JP et al: Acalabrutinib with or without obinutuzumab versus chlorambucil and obinutuzumab for treatment-naïve chronic lymphocytic leukemia (ELEVATE TN): A randomized, controlled, phase 3 trial. Lancet 395:1278, 2020.

Siddiqi T et al: Lisocabtagene maraleucel in chronic lymphocytic leukaemia and small lymphocytic lymphoma (TRANSCEND CLL 004): A multicentre, open-label, single-arm, phase 1–2 study. Lancet 402:641, 2023.

Thompson PA et al: Fludarabine, cyclophosphamide, and rituximab treatment achieves long-term disease-free survival in IGHV-mutated chronic lymphocytic leukemia. Blood 127:303, 2016.

113

Non-Hodgkin's Lymphoma

Caron A. Jacobson, Dan L. Longo

Non-Hodgkin's lymphomas (NHLs) are cancers of mature B, T, and natural killer (NK) cells. They were distinguished from Hodgkin's lymphoma (HL) upon recognition of the Reed-Sternberg (RS) cell and differ from HL with respect to their biologic and clinical characteristics. Whereas ~80–85% of patients with HL will be cured of their lymphoma by chemotherapy with or without radiotherapy, the prognosis and natural history of NHL tend to be more variable. NHL can be classified as either a mature B-NHL or a mature T/NK-NHL depending on whether the cancerous lymphocyte is a B, T, or NK cell, respectively. Within each category are lymphomas that grow quickly and behave aggressively, as well as lymphomas that are more indolent, or slow growing, in nature. For a list of the World Health Organization (WHO) classification of lymphoid neoplasms, see Table 113-1.

EPIDEMIOLOGY AND ETIOLOGY

In 2023, it is estimated that there will be 80,550 new cases of NHL in the United States, ~4% of all new cancers in both males and females, making it the seventh most common cause of cancer-related death in both women and men. The incidence is nearly 10 times the incidence of HL. There is a slight male-to-female predominance and a higher incidence for Caucasians than for African Americans. The incidence rises steadily with age, especially after age 40, but lymphomas are also among the most common malignancies in adolescent and young adult patients. The incidence of NHL has nearly doubled over the past 20–40 years and continues to rise by 1.5–2% each year. Patients with both primary and secondary immunodeficiency states are predisposed to developing NHL. These include patients with HIV infection, patients who have undergone organ transplantation, and patients with inherited immune deficiencies and autoimmune conditions. The 5-year survival rate for NHL is 74% and is higher for Caucasians than it is for African Americans.

TABLE 113-1 WHO-HAEM5 Classification of Lymphoid Malignancies

B CELL	T CELL
Mature (peripheral) B-cell neoplasms	Mature (peripheral) T-cell neoplasms
Lymphoplasmacytic lymphoma (Waldenström's macroglobulinemia)	T-cell granular lymphocytic leukemia
Hairy cell leukemia	Adult T-cell leukemia/lymphoma (HTLV-1+)
Splenic marginal zone B-cell lymphoma	Extranodal NK/T-cell lymphoma, nasal type
Extranodal marginal zone B-cell lymphoma of MALT type	Enteropathy-associated T-cell lymphoma
Nodal marginal zone B-cell lymphoma	Hepatosplenic T-cell lymphoma
Follicular lymphoma	Subcutaneous panniculitis-like T-cell lymphoma
Mantle cell lymphoma	Mycosis fungoides
Diffuse large B-cell lymphoma (including subtypes)	Sezary syndrome
High-grade B-cell lymphoma with MYC and BCL2 rearrangements	Peripheral T-cell lymphoma NOS
High-grade B-cell lymphoma NOS	Angioimmunoblastic T-cell lymphoma
High-grade B-cell lymphoma with 11q aberrations	Anaplastic large-cell lymphoma, ALK+
Burkitt's lymphoma/Burkitt's cell leukemia	Anaplastic large-cell lymphoma, ALK–
Primary mediastinal large B-cell lymphoma	
Mediastinal grey zone lymphoma	
Primary large B-cell lymphoma of immune-privileged sites	
Plasmablastic lymphoma	
Primary effusion lymphoma	
HHV8+ DLBCL NOS	
Intravascular large B-cell lymphoma	
ALK+ large B-cell lymphoma	

Abbreviations: DLBCL, diffuse large B-cell lymphoma; HHV, human herpesvirus; HTLV, human T-cell lymphotropic virus; MALT, mucosa-associated lymphoid tissue; NK, natural killer; NOS, not otherwise specified; WHO, World Health Organization.

Source: Adapted from R Alaggio et al: The 5th edition of the World Health Organization classification of haematolymphoid tumours: Lymphoid neoplasms. Leukemia 36:1720, 2022.

The incidence of NHL and the patterns of expression of the various subtypes differ geographically and across age groups. T-cell lymphomas are more common in Asia than in Western countries, whereas certain subtypes of B-cell lymphomas such as follicular lymphoma (FL) are more common in Western countries. A specific subtype of NHL known as the angiocentric nasal T/NK-cell lymphoma has a striking geographic occurrence, being most frequent in southern Asia and parts of Latin America. Another subtype of NHL associated with infection by human T-cell lymphotropic virus (HTLV) 1 is seen particularly in southern Japan and the Caribbean (see Chap. 207). Likewise, there are differences in the age-dependent incidence of NHL by histologic subtype, with aggressive lymphomas like diffuse large B-cell lymphoma (DLBCL) and Burkitt's lymphoma (BL) being the most common entities in children, and DLBCL and indolent lymphomas including FL being the most common forms in adults. The relative frequencies of the various types of lymphoid malignancies, including HL, plasma cell disorders, and lymphoid leukemias, are shown in **Fig. 113-1.**

A number of environmental factors have been implicated in the occurrence of NHL, including infectious agents, chemical exposures, and medical treatments. Several studies have demonstrated an association between exposure to agricultural chemicals and an increased incidence of NHL. Patients treated for HL can develop NHL; it is unclear whether this is a consequence of the HL or its treatment, especially radiation.

Several NHLs are associated with infectious agents (Table 113-2). Epstein-Barr virus (EBV) is associated with the development of BL in Central Africa and the occurrence of aggressive NHL in immunosuppressed patients in Western countries. The majority of primary central nervous system (CNS) lymphomas are associated with EBV. EBV infection is strongly associated with the occurrence of extranodal nasal NK/T-cell lymphomas in Asia and South America. HTLV-1 infects T cells and leads directly to the development of adult T-cell lymphoma (ATL) in a small percentage of patients infected as babies through ingestion of breast milk of infected mothers. The median age of patients with ATL is ~56 years; thus, HTLV-1 demonstrates a long latency from infection to oncogenesis (Chap. 207). Infection with HIV predisposes to the development of aggressive B-cell NHL. This may be through overexpression of interleukin 6 by infected macrophages. Infection of the stomach by the bacterium *Helicobacter pylori* induces the development of gastric mucosa-associated lymphoid tissue (MALT) lymphomas. This association is supported by evidence that patients treated with antibiotics to eradicate *H. pylori* have regression of their MALT lymphoma. The bacterium does not transform lymphocytes to produce the lymphoma; instead, a vigorous immune response is made to the bacterium, and the chronic antigenic stimulation leads to the neoplasia. MALT lymphomas of the skin may be related to *Borrelia* species infections in Europe, those of the eyes to *Chlamydophila psittaci*, and those of the small intestine to *Campylobacter jejuni*. Chronic hepatitis C virus infection has been associated with the development of lymphoplasmacytic lymphoma and splenic marginal zone lymphoma (MZL). Human herpesvirus 8 is associated with primary effusion lymphoma in HIV-infected persons and multicentric Castleman's disease, a diffuse lymphadenopathy associated with systemic symptoms of fever, malaise, and weight loss.

In addition to infectious agents, a number of other diseases or exposures may predispose to developing lymphoma (Table 113-3). Diseases of inherited and acquired immunodeficiency as well as autoimmune

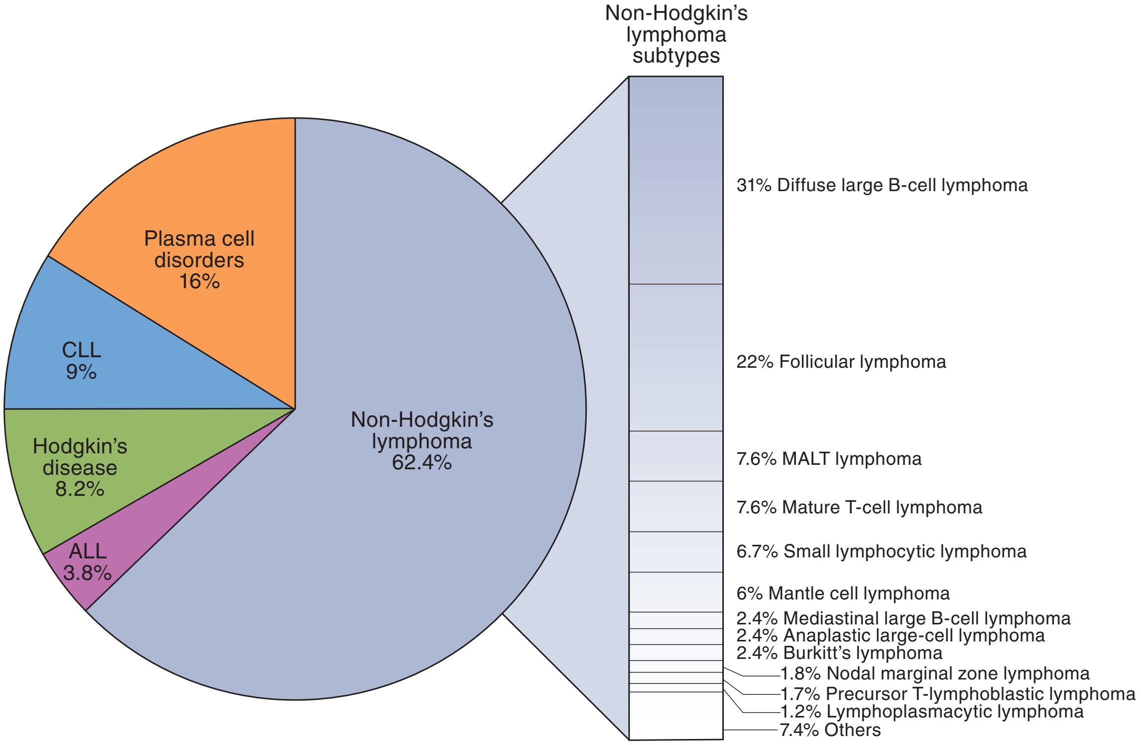

FIGURE 113-1 Relative frequency of lymphoid malignancies. ALL, acute lymphoid leukemia; CLL, chronic lymphocytic leukemia; MALT, mucosa-associated lymphoid tissue.

diseases are associated with an increased incidence of lymphoma. The association between immunosuppression and induction of NHLs is compelling because if the immunosuppression can be reversed, a percentage of these lymphomas regress spontaneously. The incidence of NHL is nearly a hundredfold increased for patients undergoing organ transplantation necessitating chronic immunosuppression and is greatest in the first year posttransplant. About 30% of these arise as a polyclonal B-cell proliferation that evolves into a clonal B-cell malignancy. The NHLs that occur in the context of immunosuppression or immunodeficiency, including HIV infection, are frequently associated with EBV. Histologically, DLBCLs are most frequently associated with immunosuppression and autoimmune diseases, although almost all histologies can be seen, especially MALT lymphomas in the context of autoimmune diseases such as Sjögren's syndrome and Hashimoto's thyroiditis. The rare inherited immunodeficiency diseases X-linked lymphoproliferative syndrome, Wiskott-Aldrich syndrome, Chédiak-Higashi syndrome, ataxia-telangiectasia, and common variable immunodeficiency syndrome are complicated by highly aggressive lymphomas. The elevated incidence of lymphoma in iatrogenic immunosuppression, AIDS, and autoimmune disease argues strongly for immune dysregulation contributing to the pathogenesis of some lymphomas. An increased risk of NHL has been observed in first-degree relatives with NHL, HL, or chronic lymphocytic leukemia (CLL). In large database studies, ~9% of patients with lymphoma or CLL have a first-degree relative with a lymphoproliferative disorder.

TABLE 113-2 Infectious Agents Associated with the Development of Lymphoid Malignancies

INFECTIOUS AGENT	LYMPHOID MALIGNANCY
Epstein-Barr virus	Burkitt's lymphoma Post–organ transplant lymphoma Primary CNS diffuse large B-cell lymphoma Hodgkin's lymphoma Extranodal NK/T-cell lymphoma, nasal type
HTLV-1	Adult T-cell leukemia/lymphoma
HIV	Diffuse large B-cell lymphoma Burkitt's lymphoma
Hepatitis C virus	Lymphoplasmacytic lymphoma
Helicobacter pylori	Gastric MALT lymphoma
Human herpesvirus 8	Primary effusion lymphoma Multicentric Castleman's disease

Abbreviations: CNS, central nervous system; HIV, human immunodeficiency virus; HTLV, human T-cell lymphotropic virus; MALT, mucosa-associated lymphoid tissue; NK, natural killer.

IMMUNOLOGY

All lymphoid cells are derived from a common hematopoietic progenitor that gives rise to lymphoid, myeloid, erythroid, monocyte, and megakaryocyte lineages. Through the ordered and sequential activation of a series of transcription factors, the cell first becomes committed to the lymphoid lineage and then gives rise to B and T cells.

About 90% of all lymphomas are of B-cell origin. A cell becomes committed to B-cell development when it expresses the master B lineage transcription factor PAX5, which ultimately results in a transcriptional program that leads to the rearrangement of its immunoglobulin genes, which involves chromosomal recombination as well as somatic hypermutation to create an immunoglobulin gene that is unique to that B cell. The sequence of cellular changes, including changes in

TABLE 113-3 Diseases or Exposures Associated with Increased Risk of Development of Malignant Lymphoma

Inherited immunodeficiency disease	Autoimmune disease
Klinefelter's syndrome	Sjögren's syndrome
Chédiak-Higashi syndrome	Celiac sprue
Ataxia-telangiectasia syndrome	Rheumatoid arthritis and systemic lupus erythematosus
Wiskott-Aldrich syndrome	Chemical or drug exposures
Common variable immunodeficiency disease	Phenytoin
Acquired immunodeficiency diseases	Dioxin, phenoxy herbicides
Iatrogenic immunosuppression	Radiation
HIV-1 infection	Prior chemotherapy and radiation therapy
Acquired hypogammaglobulinemia	Anti-TNF drugs

Abbreviations: HIV, human immunodeficiency virus; TNF, tumor necrosis factor.

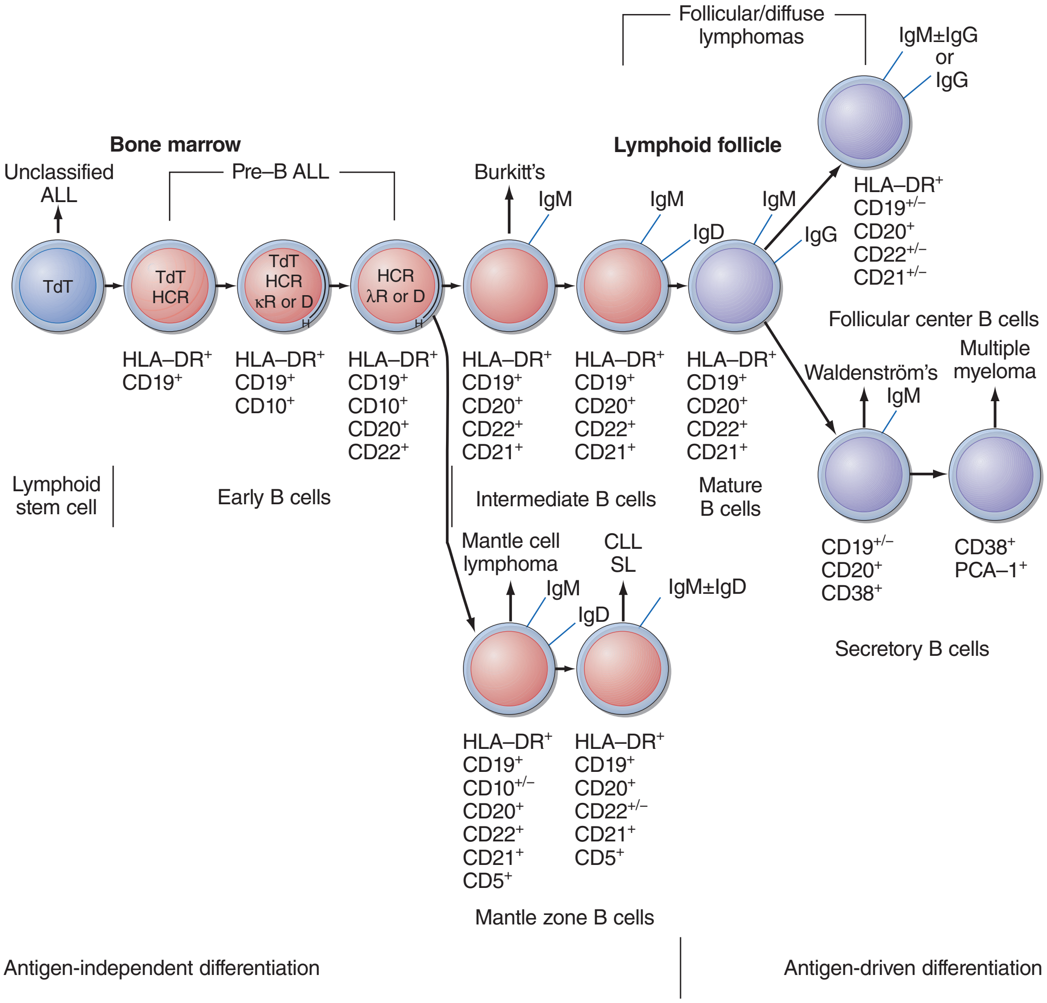
Follicular/diffuse lymphomas
IgM±IgG or IgG
Bone marrow
Lymphoid follicle
Unclassified ALL
Pre–B ALL
Burkitt's
IgM
IgM
IgD
IgM
IgG
TdT
TdT HCR
TdT HCR κR or D
HCR λR or D
H
H
HLA–DR⁺ CD19^+/– CD20⁺ CD22^+/– CD21^+/–
Follicular center B cells
HLA–DR⁺ CD19⁺
HLA–DR⁺ CD19⁺ CD10⁺
HLA–DR⁺ CD19⁺ CD10⁺ CD20⁺ CD22⁺
HLA–DR⁺ CD19⁺ CD20⁺ CD22⁺ CD21⁺
HLA–DR⁺ CD19⁺ CD20⁺ CD22⁺ CD21⁺
HLA–DR⁺ CD19⁺ CD20⁺ CD22⁺ CD21⁺
Waldenström's
IgM
Multiple myeloma
Lymphoid stem cell
Early B cells
Intermediate B cells
Mature B cells
Mantle cell lymphoma
CLL SL
IgM
IgD
IgM±IgD
CD19^+/– CD20⁺ CD38⁺
CD38⁺ PCA–1⁺
Secretory B cells
HLA–DR⁺ CD19⁺ CD10^+/– CD20⁺ CD22⁺ CD21⁺ CD5⁺
HLA–DR⁺ CD19⁺ CD20⁺ CD22^+/– CD21⁺ CD5⁺
Mantle zone B cells
Antigen-independent differentiation
Antigen-driven differentiation

FIGURE 113-2 Pathway of normal B-cell differentiation and relationship to B-cell lymphomas. HLA-DR, CD10, CD19, CD20, CD21, CD22, CD5, and CD38 are cell markers used to distinguish stages of development. Terminal transferase (TdT) is a cellular enzyme. Immunoglobulin heavy chain gene rearrangement (HCR) and light chain gene rearrangement or deletion (κR or D, λR or D) occur early in B-cell development. The approximate normal stage of differentiation associated with particular lymphomas is shown. ALL, acute lymphoid leukemia; CLL, chronic lymphocytic leukemia; SL, small lymphocytic lymphoma.

cell-surface phenotype that characterizes normal B-cell development, is shown in **Fig. 113-2.** Most B-cell lymphomas arise following the process of immunoglobulin gene recombination and somatic hypermutation, which leads to class switching and affinity maturation of the mature immunoglobulin, respectively, suggesting that it is the error-prone nature of these genetic events that contributes to oncogenesis. Certainly the frequency of chromosomal translocations that result in the activation of an oncogene or the inactivation of a tumor-suppressor gene in B-cell NHL may be the result of these normal cellular processes gone awry (see below). In addition, the key roles of the transcription factors MYC and BCL6 and the antiapoptotic protein BCL2 in the process of B-cell development explain why the genes encoding these proteins are commonly mutated in B-cell lymphomas.

A cell becomes committed to T-cell differentiation upon migration to the thymus and rearrangement of T-cell receptor (TCR) genes. This requires the expression of the T-cell master regulatory transcription factor, NOTCH-1. As in B cells, the development of the mature TCR involves the rearrangement and recombination of the TCR loci, which is error-prone and potentially oncogenic. The sequence of the events that characterize T-cell development is depicted in **Fig. 113-3.**

Although lymphoid malignancies often retain the cell-surface phenotype of lymphoid cells at particular stages of differentiation, this information is of little clinical or prognostic consequence. The so-called stage of differentiation of a malignant lymphoma does not predict its natural history. The antigen footprint, or immunophenotype, of the cell, however, is valuable diagnostically as it allows for the distinguishing of specific NHL subtypes. It can be detected by flow cytometry of single-cell suspension from blood, bone marrow, body fluid, or disaggregated tissue using fluorescently labeled antibodies against these antigens or by immunohistochemical staining of paraffin-embedded tissue sections with enzyme-linked antibodies against these antigens followed by a colorimetric reaction.

As already mentioned, malignancies of lymphoid cells are associated with recurring genetic abnormalities including chromosomal translocations and genetic mutations that may in part be the result of aberrant immunoglobulin or TCR development. While specific genetic abnormalities have not been identified for all subtypes of lymphoid malignancies, it is presumed that they exist. As previously discussed, B cells are even more susceptible to acquiring mutations during their maturation in germinal centers; the generation of antibody of higher affinity requires the introduction of mutations into the variable region genes in the germinal centers. Given this, other nonimmunoglobulin genes, e.g., *bcl-6*, may acquire mutations as well. Likewise, many lymphomas contain balanced chromosomal translocations involving the antigen receptor genes; immunoglobulin genes on chromosomes 2, 14, and 22 in B cells; and T-cell antigen receptor genes on chromosomes 7 and 14 in T cells. The rearrangement of chromosome segments to generate mature antigen receptors must create a site of vulnerability to aberrant recombination. Examples of this type of event include the (8;14)(q24;q32) translocation in BL, involving the *MYC* proto-oncogene and the IgH gene; the (14;18)(q32;q32) translocation in FL, involving the *BCL2* proto-oncogene and the IgH gene; and the (11;14) (q13;q32) translocation in mantle cell lymphoma (MCL), involving the gene encoding cyclin D1 (*CCDN1*) and the IgH gene. Less commonly, chromosomal

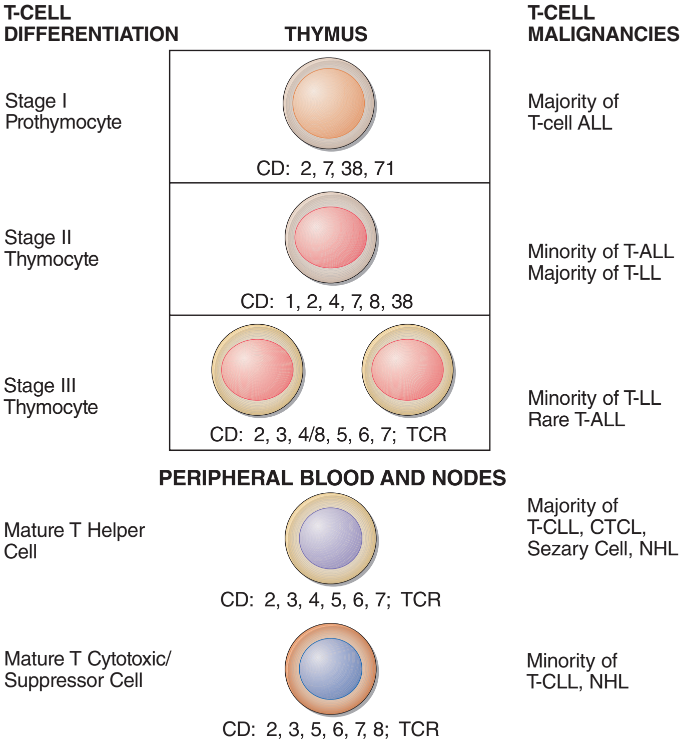
T-CELL DIFFERENTIATION
THYMUS
T-CELL MALIGNANCIES
Stage I Prothymocyte
CD: 2, 7, 38, 71
Majority of T-cell ALL
Stage II Thymocyte
CD: 1, 2, 4, 7, 8, 38
Minority of T-ALL Majority of T-LL
Stage III Thymocyte
CD: 2, 3, 4/8, 5, 6, 7; TCR
Minority of T-LL Rare T-ALL
PERIPHERAL BLOOD AND NODES
Mature T Helper Cell
CD: 2, 3, 4, 5, 6, 7; TCR
Majority of T-CLL, CTCL, Sezary Cell, NHL
Mature T Cytotoxic/ Suppressor Cell
CD: 2, 3, 5, 6, 7, 8; TCR
Minority of T-CLL, NHL

FIGURE 113-3 Pathway of normal T-cell differentiation and relationship to T-cell lymphomas. CD1, CD2, CD3, CD4, CD5, CD6, CD7, CD8, CD38, and CD71 are cell markers used to distinguish stages of development. T-cell antigen receptors (TCR) rearrange in the thymus, and mature T cells emigrate to nodes and peripheral blood. ALL, acute lymphoid leukemia; CTCL, cutaneous T-cell lymphoma; NHL, non-Hodgkin's lymphoma; T-ALL, T-cell ALL; T-CLL, T-cell chronic lymphocytic leukemia; T-LL, T-cell lymphoblastic lymphoma.

TABLE 113-4 Genetic Features of B- and T-Cell Lymphomas

GENETIC FEATURE	GENES	LYMPHOMA
t(8;14) t(2;8) t(8;22)	*MYC/IgH* *MYC/Igκ* *MYC/Ig λ*	Burkitt's lymphoma
t(11;14)	*BCL1 (CCND1)/IgH*	Mantle cell lymphoma; multiple myeloma
t(14;18) t(3;14)	*BCL2/IgH* *BCL6/IgH*	Follicular lymphoma, diffuse large B-cell lymphoma (DLBCL)
t(11;18) t(1;14) t(14;18) t(3;14)	*API2/MALT1* *BCL10/IgH* *MALT1/IgH* *FOXP1/IgH*	MALT lymphoma
Trisomy 3 7q21 deletion	Unknown *CDK6*	Splenic marginal zone lymphoma
t(9;14) 6q21 deletion	*PAX5/IgH* Unknown	Lymphoplasmacytic lymphoma
inv(14) t(14;14)	*TCRα/TCL1*	Peripheral T-cell lymphoma, NOS; T-PLL
t(2;5) t(1;2) t(2;3) t(2;17) inv(2)	*NPM1/ALK* *TPM3/ALK* *TFG/ALK* *CTLC/ALK* *ATIC/ALK*	Anaplastic large-cell lymphoma (ALCL)
Trisomy 3 Trisomy 5	Unknown Unknown	Angioimmunoblastic T-cell lymphoma
Isochromosome 7q	Unknown	Hepatosplenic T-cell lymphoma

Abbreviations: MALT, mucosa-associated lymphoid tissue; NOS, not otherwise specified; T-PLL, T-cell prolymphocytic leukemia.

translocations produce fusion genes that encode chimeric oncogenic proteins. Examples of this include the (2;5)(p23;q35) translocation involving the *ALK* and *NPM1* genes in anaplastic large-cell lymphoma (ALCL) and the t(11;18)(q21;q21) translocation involving the *API2* and *MLT* genes in MALT lymphoma. Table 113-4 presents the most common translocations and associated oncogenes for various subtypes of lymphoid malignancies.

Gene profiling using array technology allows the simultaneous assessment of the expression of thousands of genes. This technology provides the possibility to identify new genes with pathologic importance in lymphomas, the identification of patterns of gene expression with diagnostic and/or prognostic significance, and the identification of new therapeutic targets. Recognition of patterns of gene expression is complicated and requires sophisticated mathematical techniques. Early successes using this technology in lymphoma include the identification of previously unrecognized subtypes of DLBCL whose gene expression patterns resemble either those of follicular or germinal center B (GCB) cells or activated peripheral blood B cells (ABC). Patients whose lymphomas have a GCB-like pattern of gene expression have a considerably better prognosis than those whose lymphomas have a pattern resembling ABCs. This improved prognosis is independent of other known prognostic factors. These subcategories have been more specifically refined into five subcategories, using more advanced genetic sequencing techniques, that differ with respect to biology and driver genes, as well as prognosis, and may have important treatment implications in the future. Similar information is being generated in FL and MCL. The challenge remains to provide information from such techniques in a clinically useful time frame.

APPROACH TO THE PATIENT

Regardless of the type of lymphoid malignancy, the initial evaluation of the patient should include performance of a careful history and physical examination. These will help confirm the diagnosis, identify those manifestations of the disease that might require prompt attention, and aid in the selection of further studies to optimally characterize the patient's status to allow the best choice of therapy. It is difficult to overemphasize the importance of a carefully done history and physical examination. They might provide observations that lead to reconsidering the diagnosis, provide hints at etiology, clarify the stage, and allow the physician to establish rapport with the patient that will make it possible to develop and carry out a therapeutic plan.

The duration of symptoms and pace of symptomatic progression are important in distinguishing aggressive from more indolent lymphomas, as are the presence or absence of "B" symptoms, such as fevers, night sweats, or unexplained weight loss. Patients should be asked about localizing symptoms that may point toward lymphomatous involvement of specific sites, such as the chest, abdomen, or CNS. Comorbid diagnoses that may impact therapy or monitoring on therapy should be reviewed and acknowledged, including a history of diabetes or congestive heart failure. A physical examination should pay close attention to all the peripherally accessible sites of lymph nodes; the liver and spleen size; Waldeyer's ring; whether there is a pleural or pericardial effusion or abdominal ascites; whether there is an abdominal, testicular, or breast mass; and whether there is cutaneous involvement because all of these findings may influence further evaluation and disease management.

Laboratory studies should include a complete blood count, routine chemistries, liver function tests, and serum protein electrophoresis to document the presence of circulating monoclonal paraproteins. The serum β_2-microglobulin level and serum lactate dehydrogenase (LDH) are important independent prognostic factors in NHL. Staging of certain diseases may involve a bone marrow biopsy; results of other laboratory and staging studies may also warrant a marrow evaluation. A lumbar puncture for evaluation of lymphomatous involvement may be indicated in the setting of concerning neurologic signs or symptoms or diseases that are high risk for CNS involvement. The latter may include disease involving the paranasal sinuses, testes, breast, kidneys, adrenal glands, and epidural space, as well as highly aggressive histologies like BL. Since HIV and hepatitis B and C infection can be risk factors for developing NHL, and since treatment for some NHLs can result in the potentially life-threatening reactivation of hepatitis B, patients with a new diagnosis of NHL should be screened for these viruses as well.

Lymphoma histology and clinical presentation dictate which imaging studies should be ordered. Chest, abdominal, and pelvic computed tomography (CT) scans are essential for accurate staging to assess lymphadenopathy for indolent lymphomas, whereas positron emission tomography (PET) using ^{18}F-fluorodeoxyglucose (FDG-PET) is useful for aggressive lymphomas, including BL, DLBCL, plasmablastic lymphoma, and the aggressive T-cell NHLs. FDG-PET is highly sensitive for detecting both nodal and extranodal sites involved by NHL. The intensity of FDG avidity, or standardized uptake value (SUV), correlates with histologic aggressiveness, and may be useful in cases when disease transformation of an indolent lymphoma to a diffuse aggressive lymphoma is suspected. PET scanning can also differentiate between treated disease and active disease at the end of therapy in patients with residual masses on CT scans. Consensus recommendations regarding PET scanning were published as a result of an International Harmonization Project and state that PET should only be used for DLBCL and HL, that scanning during therapy should only be done as part of clinical trials, and that the end-of-treatment scan should not be done before 3 weeks but preferably 6–8 weeks after chemotherapy and 8–12 weeks after radiation or chemoradiotherapy. There is no evidence that long-term follow-up should include PET scanning. More recently, though, PET scan results at the end of therapy for FL have been associated with prognosis, with patients with residual PET-avid disease at the end of treatment having a poorer prognosis than those who are PET negative, and so it may be used for this prognostic purpose. Finally, magnetic resonance imaging (MRI) is useful in detecting bone, bone marrow, and CNS disease in the brain and spinal cord. The staging evaluation is outlined in **Table 113-5**.

The Ann Arbor staging system developed in 1971 for HL was adapted for staging NHLs **(Table 113-6)**. This staging system focuses on the number of tumor sites (nodal and extranodal), location, and the presence or absence of systemic, or B, symptoms. Table 113-6 summarizes the essential features of the Ann Arbor system.

This anatomic based system is less useful in NHL, which disseminates widely, not in an ordered stepwise fashion. A majority of patients with NHL have advanced-stage disease at diagnosis. Apart from early-stage disease limited to a radiation field where local therapy with radiation is an option, all other disease is treated the same regardless of stage. Histology and clinical parameters at presentation are more important than stage with respect to prognosis. The International Prognostic Index (IPI) is perhaps the best predictor of outcome **(Table 113-7)**. The IPI was developed based on the analysis of >2000 patients with aggressive NHLs treated with an anthracycline-containing regimen. Age (≤60 vs >60), serum LDH (≤ normal vs > normal), performance status (0 or 1 vs 2–4), stage (I or II vs III or IV), and extranodal involvement (<1 site vs >1 site) were identified as independently prognostic for overall survival (OS). A point is awarded for each risk factor and then summed, defining four risk groups: low (0 or 1); low-intermediate (2); high-intermediate (3); and high (4–5). The 5-year OS rates for patients with scores of 0–1, 2, 3, and 4–5 were 73, 51, 43, and 26%, respectively. The age-adjusted IPI separates patients ≤60 from patients >60. For the age-adjusted IPI, only stage, LDH, and performance status were important. Younger patients with 0, 1, 2, or 3 risk factors had 5-year survival rates of 83, 69, 46, and 32%, compared to 56, 44, 37, and 21% for older patients. When factoring in the introduction and clinical benefit of rituximab, the 4-year progression-free survival rates are 94, 80, and 53% for 0–1, 2, or 3 or more risk factors, respectively.

The Follicular Lymphoma International Prognostic Index (FLIPI) is a similar predictive model for FL, derived from the analysis of >4000 patients. Age >60, stage III/IV disease, the

TABLE 113-5 Staging Evaluation for Non-Hodgkin's Lymphoma

- Physical examination
- Documentation of B symptoms
- Laboratory evaluation
  - Complete blood counts
  - Liver function tests
  - Uric acid
  - Calcium
  - Serum protein electrophoresis
  - Serum β_2-microglobulin
- Chest radiograph
- CT scan of abdomen, pelvis, and usually chest
- Bone marrow biopsy
- Lumbar puncture in lymphoblastic, Burkitt's, and diffuse large B-cell lymphoma with positive marrow biopsy
- Gallium scan (SPECT) or PET scan in large-cell lymphoma

Abbreviations: CT, computed tomography; PET, positron emission tomography; SPECT, single-photon emission computed tomography.

TABLE 113-6 Ann Arbor Staging for Lymphoma[a]

STAGE	DESCRIPTION
I	Involvement of a single lymph node region (I) or single extranodal site (IE)
II	Involvement of two or more lymph node regions or lymphatic structures on the same side of the diaphragm alone (II) or with involvement of limited, contiguous, extralymphatic organ or tissue (IIE)
III	Involvement of lymph node regions on both sides of the diaphragm (III), which may include the spleen (IIIS), or limited, contiguous, extralymphatic organ or tissue (IIIE), or both (IIIES)
IV	Diffuse or disseminated foci of involvement of one or more extralymphatic organs or tissues, with or without associated lymphatic involvement

[a] All stages are further subdivided according to the absence (A) or presence (B) of systemic B symptoms including fevers, night sweats, and/or weight loss (>10% of body weight over 6 months prior to diagnosis).

TABLE 113-7 International Prognostic Index for NHL

Five Clinical Risk Factors	
Age ≥60 years	
Serum lactate dehydrogenase levels elevated	
Performance status ≥2 (ECOG) or ≤70 (Karnofsky)	
Ann Arbor stage III or IV	
>1 site of extranodal involvement	
For Diffuse Large B-Cell Lymphoma	
0, 1 factor = low risk	35% of cases; 5-year survival, 73%
2 factors = low-intermediate risk	27% of cases; 5-year survival, 51%
3 factors = high-intermediate risk	22% of cases; 5-year survival, 43%
4, 5 factors = high risk	16% of cases; 5-year survival, 26%
For Diffuse Large B-Cell Lymphoma Treated With R-CHOP	
0 factor = good	10% of cases; 4-year survival, 94%
1, 2 factors = intermediate	45% of cases; 4-year survival, 80%
3, 4, 5 factors = poor	45% of cases; 4-year survival, 53%

Abbreviations: ECOG, Eastern Cooperative Oncology Group; NHL, non-Hodgkin's lymphoma; R-CHOP, rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone.

presence of >4 nodal sites, an elevated serum LDH concentration, and a hemoglobin <12 were identified as independent prognostic variables, and summation of each variable identified three risk groups. The median 10-year survival rates for patients with zero to one (low risk), two (intermediate risk), or three or more (high risk) of these adverse factors were 71, 51, and 36%, respectively. Similar disease-specific IPIs have been developed for MCL and peripheral T-cell lymphoma (PTCL) as well. These prognostic indices take into account the proliferative index and cell-surface markers, respectively.

Finally, as mentioned previously, gene expression profiling has identified DLBCLs with differential prognoses: GCB and ABC, where GCB-like DLBCL is associated with a significantly better OS. A more readily accessible immunohistochemical algorithm has been developed, based on the presence or absence of CD10, BCL6, and MUM1, that correlates closely with gene expression profiles and can differentiate the majority of GCB from non–GCB-like DLBCL. These profiles have prognostic importance but, to date, do not alter treatment recommendations for the primary treatment of DLBCL. Current clinical trials do stratify by DLBCL subtype, and it appears that agents like the Bruton tyrosine kinase (BTK) inhibitor ibrutinib and lenalidomide are most active in non-GCB DLBCL in the relapsed setting. Treatment may then be differentiated by these subtypes in the future.

CLINICAL FEATURES, TREATMENT, AND PROGNOSIS OF SPECIFIC NHL

■ MATURE B-CELL NEOPLASMS

B-cell NHLs can be characterized into two broad groups—those that behave aggressively, require immediate or urgent treatment with combination chemotherapy regimens, and are potentially curable; and those that are more indolent in nature, can be observed and treated only when they cause symptoms or signs of organ function impairment, are very responsive to therapy, but are not ultimately curable in the vast majority of cases. Among the aggressive diseases, the most common is DLBCL, and the most rapidly growing is BL. FL is the second most common NHL and the most common indolent NHL. Other indolent NHLs include MZL, lymphoplasmacytic lymphoma (LPL), and hairy cell leukemia (HCL). MCL is an intermediate-grade lymphoma that shares some characteristics with the aggressive lymphomas (fairly urgent need for treatment and aggressive upfront combination chemotherapy regimens), but like the indolent lymphomas, it is not readily curable with conventional-dose therapies.

Burkitt's Lymphoma Burkitt's lymphoma/leukemia (BL) is a rare disease in adults in the United States, making up <1% of NHL, but it makes up ~30% of childhood NHL. It is one of the fastest growing neoplasms, with a doubling time of <24 h. In general, it is a pediatric tumor that has three major clinical presentations. The endemic (African) form presents as a jaw or facial bone tumor that spreads to extranodal sites including ovary, testis, kidney, breast, and especially the bone marrow and meninges. The nonendemic form has an abdominal presentation with massive disease, ascites, and renal, testis, and/or ovarian involvement and, like the endemic form, also spreads to the bone marrow and CNS. Immunodeficiency-related cases more often involve lymph nodes and may present as acute leukemia. BL has a male predominance and is typically seen in patients <35 years of age.

On biopsy, there is a monotonous infiltration of medium-sized cells with round nuclei, multiple nucleoli, and basophilic cytoplasm with vacuoles. The proliferation rate is ~100%, and tingible body macrophages give rise to the classic "starry sky" appearance of this tumor **(Fig. 113-4)**. Tumor cells are positive for B-cell antigens CD19 and CD20 and surface immunoglobulin. They are also uniformly positive for CD10 and BCL6 but negative for BCL2. Endemic BLs are EBV positive, whereas the majority of nonendemic BLs are EBV negative. BL is associated with a translocation involving MYC on chromosome 8q24 in >95% of the cases. The most common partners are chromosomes 14, 2, or 22, rearrangements that produce fusions of MYC with either the IgH (80%), kappa (15%), or lambda (5%) light chain genes, respectively.

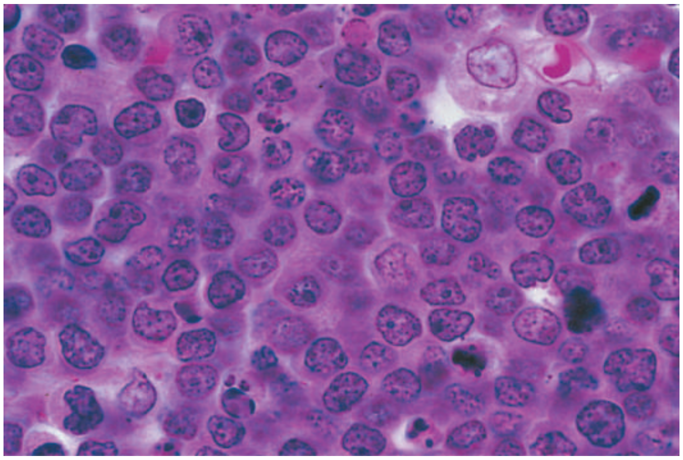

FIGURE 113-4 Burkitt's lymphoma. The neoplastic cells are homogeneous, medium-sized B cells with frequent mitotic figures, a morphologic correlate of high growth fraction. Reactive macrophages are scattered through the tumor, and their pale cytoplasm in a background of blue-staining tumor cells gives the tumor a so-called starry sky appearance.

While exquisitely chemosensitive, it is imperative that treatment for BL be initiated quickly given the rapid doubling time and high morbidity of this disease. There are several effective intensive combination chemotherapy regimens, all of which incorporate high doses of cyclophosphamide. Prophylactic therapy to the CNS is mandatory. Cure can be expected in 80–90% of patients when treated promptly and correctly. Modified Magrath and dose-adjusted EPOCH-R (rituximab, infusional etoposide/vincristine/doxorubicin, cyclophosphamide, prednisone) are highly effective regimens. Salvage therapy has been generally ineffective in patients whose disease progresses after upfront therapy, emphasizing the importance of the initial treatment approach and referral to a tertiary cancer center with experience treating this disease.

Diffuse Large B-Cell Lymphoma DLBCL is the most common histologic subtype of NHL diagnosed, representing about one-third of all cases. Previously felt to be "one disease," it is now recognized as a heterogeneous collection of multiple entities. It is slightly more common in Caucasians and men, and the median age at diagnosis is 64. The relative risk (RR) of DLBCL is higher among people with affected first-degree relatives (RR 3.5-fold), and patients with congenital or acquired immunodeficiency, patients on immunosuppression, and patients with autoimmune disorders also have a higher risk of developing DLBCL, often EBV-related. The majority of patients present with advanced-stage disease, with only 30–40% of patients having stage I or II disease; ~40% of patients will have "B" symptoms, and 50% of patients will have an elevated LDH. Up to 40% of patients will have involvement of non–lymph node sites including bone marrow, CNS, gastrointestinal tract, thyroid, liver, and skin. Patients with extensive bone marrow involvement or involvement of the testes, breast, kidney, adrenal gland, paranasal sinus, or epidural space are at increased risk of CNS dissemination.

The tumor consists of a diffuse proliferation of large, atypical lymphocytes with a high proliferative index **(Fig. 113-5)**. These cells typically express the B-cell antigens CD19, CD20, and CD79a. Expression of CD10 and BCL6 is consistent with the tumor cell being of germinal center origin (GCB), while the expression of MUM1 corresponds with the non–germinal center or ABC subtype. BCL2 is overexpressed in anywhere from 25 to 80% of DLBCLs, whereas BCL6 is positive in more than two-thirds of cases, as the result of translocations, gain of copy number, or promoter mutations. MYC is rearranged in 10% of DLBCLs, and ~20% of MYC-rearranged cases have a concurrent BCL2 rearrangement, a combination referred to as "double-hit lymphoma." These double-hit lymphomas constitute one subtype of high-grade B-cell lymphoma (HGBL) and are associated with an extremely poor

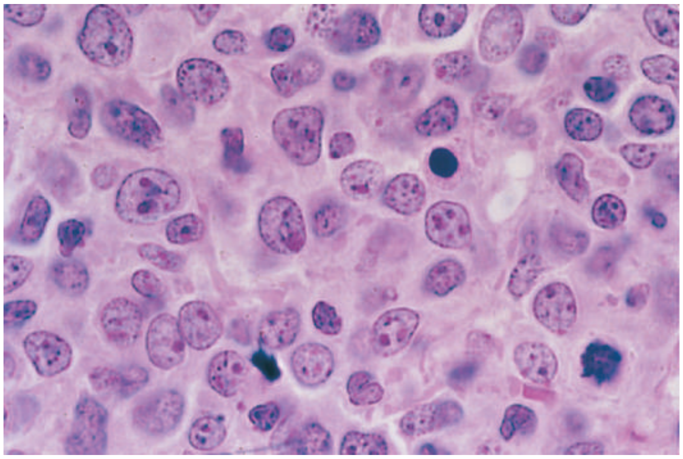

FIGURE 113-5 Diffuse large B-cell lymphoma. The neoplastic cells are heterogeneous but predominantly large cells with vesicular chromatin and prominent nucleoli.

prognosis with a median OS of only 12–18 months. The other subtype of HGBL is called HGBL not otherwise specified (NOS) and is defined based on blastoid or Burkitt-like morphologic features. Amplification and/or overexpression of MYC independent of rearrangements or amplification have also been described and are also associated with a poor, albeit better, prognosis.

Combination chemotherapy offers potentially curative therapy for DLBCL, regardless of the stage. The addition of the anti-CD20 antibody rituximab to cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) improved survival beyond CHOP alone and has been the standard first-line chemotherapy for this disease for decades. The combination of R-CHP (R-CHOP without vincristine) and the antibody-drug conjugate (ADC) polatuzumab was compared to R-CHOP in the randomized POLARIX study and demonstrated a progression-free survival benefit with polatuzumab–R-CHP, leading to the approval of this regimen as a new standard option for patients with IPI 2 or higher disease. For patients with early-stage disease localized to a radiation field, treatment options include full-course chemotherapy with R-CHOP every 3 weeks for six cycles or abbreviated chemotherapy for three to four cycles followed by involved field radiotherapy. The randomized FLYER studies compared four versus six cycles of R-CHOP chemotherapy for early favorable-risk DLBCL and showed no benefit to a more extended course of chemotherapy. For advanced-stage DLBCL, therapy is with a full course of chemotherapy. On average, ~65–70% of patients with DLBCL can be expected to be cured with this approach, and the likelihood of cure is predicted by the IPI, gene expression profile cell of origin, and/or MYC cytogenetics and expression. Several studies, other than the above mentioned POLARIX study, have investigated alternative anthracycline-containing chemotherapy regimens and/or consolidation autologous stem cell transplantation in first remission for higher-risk disease without improvement over R-CHOP alone. Dose-adjusted R-EPOCH is one such regimen. Although this regimen did not appear to be better than R-CHOP for DLBCL in one multicenter clinical trial, it is often used to treat primary mediastinal large B-cell lymphoma and double-hit DLBCL based on results from phase 2 and retrospective studies, respectively. CNS prophylaxis with either intrathecal chemotherapy or high-dose systemic methotrexate and leucovorin rescue offers unclear benefit to patients at high risk for CNS relapse and is increasingly not being used based on results of retrospective and observational studies.

Over one-third of patients will either have primary refractory disease or disease that relapses after first-line chemotherapy. These patients may still be cured with either salvage chemotherapy regimens followed by autologous stem cell transplantation (for patients relapsing >12 months from the completion of therapy) or CD19 chimeric antigen receptor (CAR) T cells (for patients with primary refractory or early relapsing disease or those who are not felt to be autologous stem cell transplantation candidates). However, patients with a poor performance status or certain comorbid conditions who are not candidates for such approaches are often managed with palliative intentions. Radiation to symptomatic areas of disease can be transiently helpful. Less intensive chemotherapy with drugs such as gemcitabine, cytarabine, or bendamustine can help control disease and symptoms for a limited period of time. Newer drugs including the CD79b ADC polatuzumab in combination with bendamustine and rituximab (BR), the high-affinity CD19 monoclonal antibody (mAb) tafasitamab in combination with lenalidomide, the CD19 ADC loncastuximab, and the CD20-CD3 bispecific antibodies epcoritamab and glofitamab have been approved for use in these palliative settings as well as for patients who relapse after CAR-T therapy or autologous stem cell transplantation. Some of these agents can be used as a bridge to a definitive allogeneic stem cell transplantation

For patients in whom more aggressive therapy is an option, treatment for late relapsing patients is with combination chemotherapy using various combinations of drugs primarily in order to identify patients with chemosensitive disease. Patients with chemosensitive disease have the greatest likelihood of benefiting from high-dose chemotherapy and autologous stem cell transplant, which improves response duration and survival over salvage chemotherapy alone and leads to long-term disease-free survival in ~40–50% of patients. The randomized ZUMA-7 and TRANSFORM studies of the CD19 CAR-T therapies axicabtagene ciloleucel (axi-cel) and lisocabtagene autoleucel (liso-cel) established both of these CAR-T therapies as the most effective option for patients with primary refractory or early (<12 months) relapsing disease over salvage chemotherapy and autologous stem cell transplantation.

For this therapy, T cells are collected from a patient and are then genetically modified to express a receptor that will bind to a surface antigen expressed on the patient's own tumor cells. In the case of B-cell malignancies, CD19 has been targeted most commonly. After infusion, autologous CAR-T cells home to sites of disease and persist over time. The CARs consist of an extracellular antigen recognition domain (typically a single-chain Fv variable fragment from a monoclonal antibody) linked via a transmembrane domain to an intracellular signaling domain (usually the CD3ζ endodomain), resulting in the redirection of T-cell specificity toward target antigen-positive cells, and one or more costimulatory domains including CD28, 4-1BB, or OX40 to enhance cytokine secretion and effector cell expansion and prevent activation-induced apoptosis and immune suppression by tumor-related metabolites. For late relapsing patients with chemorefractory disease, CAR-T therapies such as axi-cel, liso-cel, and tisagenlecleucel (tisa-cel) offer a potentially curative option in the third-line setting as well. In this setting, the response rate of CAR-T cells is >80%, with >50% of patients achieving a complete response. These responses appear to be durable, with 40% of patients in remission at long-term follow-up.

Other large B-cell lymphomas include intravascular large B-cell lymphoma, T-cell/histiocyte–rich large B-cell lymphoma, EBV-positive DLBCL of the elderly, and ALK-positive large B-cell lymphoma. Patients with the latter two diseases tend to have a poor prognosis, whereas the addition of rituximab to CHOP chemotherapy has dramatically improved outcomes with intravascular large B-cell lymphoma, and the outcomes in T-cell/histiocyte–rich large B-cell lymphoma are similar to DLBCL. R-CHOP remains the treatment of choice for each of these lymphomas.

Follicular Lymphoma FLs are the second leading NHL diagnosis in the United States and Europe and make up 22% of NHLs worldwide and at least 30% of NHLs diagnosed in the United States. This type of lymphoma can be diagnosed accurately on morphologic findings alone and has been the diagnosis in the majority of patients in therapeutic trials for "low-grade" lymphoma in the past.

Evaluation of an adequate biopsy by an expert hematopathologist is sufficient to make a diagnosis of FL. The tumor is composed of small cleaved and large cells in varying proportions organized in a follicular pattern of growth **(Fig. 113-6)**. Confirmation of B-cell immunophenotype (monoclonal immunoglobulin light chain, CD19, CD20, CD10, and BCL6 positive, and CD5 and CD23 negative) and the existence of t(14;18) and abnormal expression of BCL2 protein are confirmatory.

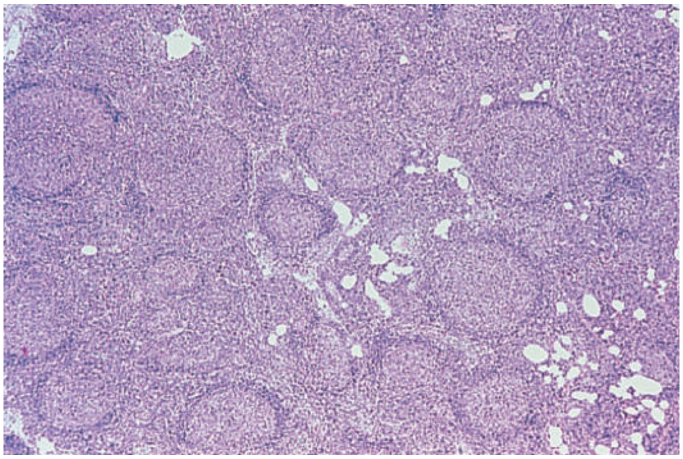

FIGURE 113-6 Follicular lymphoma. The normal nodal architecture is effaced by nodular expansions of tumor cells. Nodules vary in size and contain predominantly small lymphocytes with cleaved nuclei along with variable numbers of larger cells with vesicular chromatin and prominent nucleoli.

While >85% of FLs will harbor a t(14;18) and overexpress the antiapoptotic protein BCL2, this genetic event is necessary but not sufficient for malignant transformation of the B lymphocytes, and multiple genetic events are required for the development of FL. Studies have identified the most common recurrent genetic events in FL, and they included mutations in several epigenetic modifying genes, including *MLL2, EZH2, CREBBP*, and *EP300*. The major differential diagnosis is between lymphoma and reactive follicular hyperplasia. The coexistence of DLBCL must be considered. Patients with FL are often subclassified, or graded, into those with predominantly small cells, those with a mixture of small and large cells, and those with predominantly large cells. The WHO classification adopted grading from I to III based on the number of centroblasts, or large cells, counted per high-power field (hpf): grade I, from 0 to 5 centroblasts/hpf; grade II, from 6 to 15 centroblasts/hpf; and grade III, >15 centroblasts/hpf. Grade III has been subdivided into grade IIIa, in which centrocytes predominate, and follicular large B-cell lymphoma (FLBCL), in which there are sheets of centroblasts. While this distinction cannot be made simply or very reproducibly, these subdivisions do have prognostic significance. Patients with FL with predominantly large cells have a higher proliferative fraction, progress more rapidly, and have a shorter OS with simple chemotherapy regimens. FLBCL is an aggressive disease and considered most similar to DLBCL and treated as such with curative intent.

The most common presentation for FL is with new, painless lymphadenopathy. Multiple sites of lymphoid involvement are typical, and unusual sites such as epitrochlear nodes are sometimes seen. However, essentially any organ can be involved, and extranodal presentations do occur. Most patients do not have an elevated LDH or fevers, night sweats, or weight loss, although histologic transformation to DLBCL does occur at a rate of ~3% per year and can be associated with these signs or symptoms. As discussed previously, prognosis is best predicted by the FLIPI. Staging is typically done with CT scans of the chest, abdomen, and pelvis, as well as the neck if neck disease is suspected, although PET/CT scans can be helpful in cases where disease transformation is suspected, as transformed disease will be more FDG avid than indolent disease, or for confirmation of early-stage disease, where definitive local therapy with radiation may be considered.

Although FL is highly sensitive to chemotherapy and radiotherapy, these therapies are usually not ultimately curative, except in the setting of early-stage disease. If the disease can be encompassed in a radiation field, involved field radiotherapy at a dose of 24–30 Gy may be curative, with 5-, 10-, and 15-year freedom from treatment failure rates of 72, 46, and 39%, and overall 5-, 10-, and 15-year survival rates of 93, 75, and 62%, respectively. If radiation therapy would not be tolerated or if a patient prefers not to receive radiation, observation is a reasonable alternative with a median time to treatment not reached at 7 years of follow-up in one study. Many of these patients are diagnosed incidentally or at a time when their lymphoma is not causing symptoms or signs of organ function impairment. Numerous studies have shown that treating patients with asymptomatic disease does not improve survival compared with a program of close observation, with treatment reserved for symptomatic disease progression or organ dysfunction. Thus, asymptomatic patients should be observed.

When systemic treatment is indicated, a variety of treatment options are available, including the use of the monoclonal antibody against CD20, rituximab, alone or in combination with chemotherapy or with the oral drug lenalidomide. Treatment decisions are often determined by the indication for treatment and/or by the volume of disease being treated. For patients requiring therapy for inflammatory or autoimmune phenomenon thought to be driven by FL or for patients with low-volume disease, single-agent rituximab is associated with a response rate of ~70% and a median response duration of >2 years. This response duration is improved with the addition of maintenance rituximab following a favorable response to rituximab induction therapy. For patients with a larger volume of disease at the time of treatment initiation, the addition of rituximab (R) to chemotherapy regimens such as CHOP or cyclophosphamide, vincristine, and prednisone (CVP) has improved survival in this disease. The combination of bendamustine and rituximab (BR) has been compared to R-CHOP and results in longer response duration and less toxicity. Thus, BR has become the standard of care for the first-line therapy of medium- to high-volume FL. Similarly, the addition of maintenance rituximab following a good response to R-CHOP or R-CVP improves response duration when used in newly treated FL patients. A newer anti-CD20 antibody, obinutuzumab, has been tested in combination with chemotherapy in a randomized trial against rituximab plus chemotherapy in previously untreated FL. The obinutuzumab combinations resulted in improvements in minimal residual disease (MRD) negativity as well as progression-free survival at the expense of more infection and infusion reactions. Based on these results, both rituximab plus chemotherapy and obinutuzumab plus chemotherapy are options for untreated FL in need of treatment. The superiority of one over the other has not been established. Finally, a randomized study has compared rituximab plus chemotherapy with either BR, R-CHOP, or R-CVP to rituximab plus lenalidomide, and results were similar in both arms, thus making the chemotherapy-free rituximab-lenalidomide regimen an option for the frontline treatment of FL.

In patients with FL, the disease nearly always recurs following therapy, after which retreatment is again reserved for symptomatic disease or disease interfering with organ function. Single-agent rituximab or alternative chemotherapy regimens, with both rituximab and obinutuzumab, can again be employed. Both autologous and allogeneic hematopoietic stem cell transplantations yield high complete response rates in patients with relapsed FL, and long-term remissions can occur in 40 and 60% of patients, respectively. The latter is associated with considerable treatment-related morbidity and mortality and so is usually reserved for patients with multiply relapsed FL that is no longer responsive to chemotherapy. More targeted oral therapies like lenalidomide and the EZH2 inhibitor tazemetostat are active in relapsed FL. The PI3 kinase inhibitors idelalisib, duvelisib, and copanlisib are active in FL but are no longer available given the lack of randomized confirmatory evidence to support their efficacy. The anti-CD19–directed CAR-T therapies axi-cel and tisa-cel have been approved for relapsed FL in the third-line setting and beyond, with complete responses seen in >80% of patients with multiply relapsed disease, and with many of those responses proving durable, albeit with limited follow-up. Longer follow-up is needed to determine if this may be a definitive treatment strategy for a subset of relapsed FL patients. The CD20-CD3 bispecific antibody mosunetuzumab has been approved for FL in the third-line setting and beyond as well and offers a highly effective and potentially less toxic option for patients compared with CAR-T therapy. On average, most patients will live with FL for 15–20 years, a number that is increasing given our improved understanding of the genetics and microenvironment of FL and the increasing number of drugs and therapies being tested in this disease. However, in addition to a high-risk FLIPI, patients who do not have a complete metabolic response by PET/CT scanning to their primary therapy and patients who relapse within 2 years of the completion of their primary chemotherapy tend to do poorly with chemotherapy and should be considered for some of these non-chemotherapy options. Randomized trials comparing chemotherapy, lenalidomide-based therapy, and CAR-T therapy in the second-line setting for these refractory or early relapsing patients are underway.

Patients with FL have a high rate of histologic transformation to DLBCL (~3% per year). This is recognized ~40% of the time during the course of the illness by repeat biopsy and is present in almost all patients at autopsy. This transformation is usually heralded by rapid growth of lymph nodes—often localized—and the development of systemic symptoms such as fevers, sweats, and weight loss. When this happens in patients who have had previously untreated FL, treatment with R-CHOP chemotherapy, as for DLBCL, can be curative for the aggressive component while the FL may eventually recur. In patients with previously treated FL that transforms to DLBCL, prognosis is poor, and successful therapy with an aggressive combination chemotherapy regimen should be consolidated with an autologous stem cell transplant. CAR-T therapy and other therapies for relapsed/refractory DLBCL are options for chemorefractory transformations. Finally, as discussed previously, grade IIIb FL is more similar to DLBCL than it is to FL and should be treated as such.

Marginal Zone Lymphoma The second most common indolent B-cell NHL is MZL. There are three main types: splenic MZL, extranodal MZL of MALT, and nodal MZL.

Nodal MZL most closely resembles FL clinically, and much of the way we manage and treat it is based on studies done in FL. Tumor biopsies in this disease show parafollicular and perivascular infiltration by monocytoid-appearing atypical lymphocytes with folded nuclear contours that are positive for CD19, CD20, and CD79a but negative for CD10 and largely negative for CD5. Some cases can have plasmacytoid differentiation and can be associated with a monoclonal expression of kappa or lambda light chains and with small monoclonal immunoglobulin spikes. Treatment is often similar to that of FL, with the exception that the BTK inhibitors ibrutinib and zanubrutinib are highly active in this disease, while largely disappointing in FL, and are good treatment options for relapsed nodal MZL as well as other MZL subtypes. CAR-T therapy is not approved in MZL but has been tested in clinical trials with similar efficacy to what has been shown in FL.

Splenic MZL is largely a disease of older Caucasian patients; infection with hepatitis C is a risk factor for this disease, and treatment of hepatitis C can result in regression of the lymphoma. Patients present with a lymphocytosis with or without cytopenias and splenomegaly. Bone marrow involvement is common. Diagnosis can be made by flow cytometry of the peripheral blood; malignant lymphocytes will be positive for surface immunoglobulin, CD19, and CD20 and will generally lack CD5 and CD10. On peripheral smear, they have small nuclei and abundant cytoplasm with "shaggy" or villous projections. It can be differentiated from HCL by the absence of CD25, CD103, and annexin A1. Recurrent cytogenetic abnormalities include trisomy 3 and abnormalities of chromosome 7q. Therapy is indicated for symptomatic disease or significant cytopenias. Splenectomy is reasonable for selected patients with excellent relief of symptoms and cytopenias. Splenectomy is associated with an overall response rate of 85% and estimated progression-free survival and OS rates at 5 years of 58 and 77%, respectively. Single-agent rituximab can improve splenomegaly and cytopenias in >90% of patients. In a study of induction with weekly rituximab followed by maintenance, the response rate was 95%, with OS and progression-free survival rates at 5 years of 92 and 73%, respectively. Other options for therapy at relapse are similar to those used for FL and include retreatment with rituximab, alkylating agents, and purine analogues in combination with rituximab. The survival rate of patients is in excess of 70% at 10 years.

MALT lymphoma is an MZL lymphoma of extranodal tissue, most commonly the stomach, but other common sites include the skin, salivary glands, lung, small bowel, ocular adnexa, breasts, bladder, thyroid, dura, and synovium. It is associated with states of chronic inflammation due to either autoimmune diseases like Sjögren's syndrome or Hashimoto's thyroiditis or chronic infections with organisms such as *H. pylori* (gastric), *Borrelia burgdorferi* (skin), *C. psittaci* (conjunctiva), *C. jejuni* (intestines), and hepatitis C virus. The essential pathologic feature of MALT lymphoma is the presence of lymphoepithelial lesions, which result from invasion of mucosal glands and crypts by the neoplastic lymphocytes. These cells are positive for CD19, CD20, and CD79a and negative for CD5 and CD10. Recurrent cytogenetic abnormalities include t(11;18), t(14;18), t(1;14), t(3;14), and trisomy 8. The t(11;18) is most common, occurring in up to 50% of MALT lymphomas. It results in the fusion of the apoptosis inhibitor 2 (*API2*) gene and the *MALT1* gene, resulting in activation of nuclear factor-κB (NF-κB). Unlike other indolent B-cell lymphomas, MALT lymphomas present most commonly with stage I or II disease. In these cases, radiation therapy may be curative. Alternatively, patients may respond to antibiotics for the associated underlying infection. Treatment of symptomatic or organ-impairing relapsed, refractory, or advanced-stage disease is similar to approaches used in FL with chemotherapy, immunotherapy, or chemoimmunotherapy.

Lymphoplasmacytic Lymphoma About 1% of all NHLs will be LPLs, which are indolent B-cell NHLs with lymphoplasmacytic differentiation, most commonly associated with a monoclonal IgM paraprotein. Nearly all patients will have stage IV disease at diagnosis with bone marrow involvement. Patients with high levels of circulating IgM paraproteins constitute a specific entity known as Waldenström's macroglobulinemia and can have symptoms due to hyperviscosity as a result of the circulating IgM. Activating mutations in MYD88, an adaptor protein that is involved in signaling downstream of the Ig receptor leading to NF-κB activation, are present in >90% of cases. Tumor biopsies are notable for proliferation of small lymphocytes, lymphoplasmacytic cells, and plasma cells, and malignant lymphocytes are positive for CD19, CD20, and surface IgM but generally negative for CD5 and CD10. Like the other indolent NHLs, treatment is indicated for disease that causes symptoms or interferes with organ function; hyperviscosity related to elevated serum IgM and paraneoplastic neuropathy are additional indications for therapy. Single-agent rituximab may be useful for low-volume disease but can be associated with a transient rise in serum IgM concentrations that can cause or exacerbate hyperviscosity. Chemoimmunotherapy with regimens such as BR and rituximab, cyclophosphamide, and dexamethasone is active, as are myeloma therapies such as bortezomib. Ibrutinib and zanubrutinib in combination with rituximab are highly active in this disease and are options for both previously untreated and relapsed disease. Given that 85% of IgM remains intravascular, acute relief of hyperviscosity symptoms can be obtained by plasmapheresis. For recurrent disease, one can often use agents that were previously used. For patients with more refractory LPL, the mammalian target of rapamycin (mTOR) inhibitor everolimus and the bcl-2 inhibitor venetoclax are active. Selected patients with relapsed disease are considered for high-dose therapy with autologous or allogeneic stem cell transplantation. The results seen are similar to those of other indolent lymphomas.

Mantle Cell Lymphoma MCL composes ~6% of NHLs. It is an intermediate-grade lymphoma that, like the indolent B-cell NHLs, is not curable with conventional therapies but, like the aggressive lymphomas, often requires more aggressive chemoimmunotherapy regimens with or without an autologous stem cell transplant to achieve a reasonable response duration. This therapy is not curative, however, and median survival with this disease is on the order of 5–10 years. An exception to this is a more indolent SOX11 variant that often presents with circulating disease with splenomegaly but without significant lymphadenopathy and with a low Ki67 (<10%). This subset behaves more like the indolent B-cell NHLs and can be observed until treatment is indicated by symptoms or organ function impairment. Similarly, there is a blastic variant with a high Ki67 index that is associated with a poor prognosis and a median OS of only 18 months. For other patients, prognosis is best predicted by the biologic MCL International Prognostic Index (MIPI), which factors in age, performance status, LDH, white blood cell count, and Ki67 expression to determine a risk group. This disease is more common in men, and the average age of diagnosis is 63. MCLs with a mutation in *TP53* or a complex karyotype are particularly high risk as well. Over two-thirds of patients will have stage IV disease, mostly with bone marrow and peripheral blood involvement, at the time of diagnosis. Another common extranodal site of involvement is the gastrointestinal tract, where diffuse lymphomatous polyposis may be seen.

The pathognomonic cytogenetic finding in MCL is t(11;14), which brings the gene for the cell cycle control protein cyclin D1 under the control of the immunoglobulin heavy chain gene promoter on chromosome 14. This translocation is present in >90% of cases. The remaining cases usually overexpress cyclin D2, cyclin D3, or cyclin E. Tumor cells also are positive for B-cell markers CD19 and CD20, as well as CD5. They usually lack CD10 and CD23.

Therapies for MCL are evolving. Patients with localized disease might be treated with combination chemotherapy followed by radiotherapy; however, these patients are exceedingly rare. Similarly, patients with the indolent variant can be observed until disease progresses to cause symptoms or signs of organ function impairment. For the usual presentation with disseminated disease, standard lymphoma treatments like R-CHOP have been unsatisfactory, with the minority of patients achieving complete remission. The addition of high-dose cytarabine to an R-CHOP–like backbone with or without consolidation autologous stem cell transplantation in first remission has improved progression-free survival, but it has not elicited cures in this disease. These include the Nordic regimens and R-HyperCVAD (rituximab, cyclophosphamide, vincristine, doxorubicin, dexamethasone, cytarabine, and methotrexate). BR has activity in this disease and is more effective and better tolerated than R-CHOP. Newer studies with short follow-up suggest that strategies that combine BR with cytarabine with or without autologous stem cell transplant may be effective and well tolerated. The SHINE study randomized patients to BR alone versus BR plus ibrutinib and showed a modest benefit with the addition of ibrutinib. It is not clear, however, if this benefit is greater than the sequencing of upfront BR followed by BTK inhibition at relapse. Maintenance rituximab, following a good response to induction chemotherapy or after autologous stem cell transplant, also improves outcomes over observation alone. Increasing evidence suggests that with modern induction therapies, high-dose chemotherapy and autologous stem cell rescue may no longer provide benefit in the frontline setting. The randomized European TRIANGLE study showed no benefit of autologous transplant when added to induction chemoimmunotherapy with ibrutinib followed by rituximab and ibrutinib maintenance. An ongoing randomized study in the United States is asking a similar question for patients who are treated with chemoimmunotherapy without upfront BTK inhibition. For relapsed disease, the BTK inhibitors ibrutinib, acalabrutinib, and zanubrutinib have single-agent activity with a response rate of almost 70% but a response duration of only 18–24 months. These drugs are being explored in combination with chemotherapy as well as with the BCL2 antagonist venetoclax. Anti-CD19–directed CAR-T therapies are approved for the treatment of relapsed/refractory MCL; two-thirds of patients who had progressed after chemoimmunotherapy (with or without an autologous stem cell transplant) and BTK inhibition have achieved complete responses, many of which are durable through limited follow-up. As in FL, longer follow-up is needed to determine if some of these patients may be cured, which would make this the only curative therapy for this disease outside of an allogeneic stem cell transplantation. The noncovalent BTK inhibitor pirtobrutinib has activity in MCL that has progressed on prior covalent BTK inhibitors and is now approved in this setting. Drugs such as lenalidomide, venetoclax, bortezomib, and temsirolimus can similarly induce transient partial responses. Appropriate patients who respond to salvage therapy, with the exception of CAR-T therapy, should be considered for allogeneic stem cell transplant, which can lead to long-term disease-free survival in 30–50% of patients.

MATURE (PERIPHERAL) T-CELL DISORDERS

Mature T-cell disorders include cutaneous lymphomas, such as mycosis fungoides, and the PTCLs, some of which are distinguished based on specific clinical presentations or contexts or by molecular or biologic features, but many of which fall into the category of PTCL not otherwise specified (NOS). T-cell NHLs are significantly rarer than B-cell NHLs, and as such, our understanding of their biology is less advanced and our therapies are less well developed. While some T-cell lymphomas, like mycosis fungoides, can behave indolently and some, like ALK-positive ALCL, can be cured with chemotherapy, the majority are associated with a poor prognosis. The advent of genomic technologies is enhancing our ability to understand the genetic and biologic basis of these neoplasms.

Mycosis Fungoides Mycosis fungoides is also known as cutaneous T-cell lymphoma. This lymphoma is more often seen by dermatologists than internists. The median age of onset is in the mid-fifties, and the disease is more common in males and in blacks.

Mycosis fungoides is an indolent lymphoma, with patients often having several years of eczematous or dermatitic skin lesions before the diagnosis is finally established. The skin lesions progress from patch stage to plaque stage to cutaneous tumors. Early in the disease, biopsies are often difficult to interpret, and the diagnosis may only become apparent by observing the patient over time. Adenopathy may reflect involvement with mycosis fungoides or be read as dermatopathic change. In advanced stages, the lymphoma can spread to lymph nodes and visceral organs. Patients with this lymphoma may develop generalized erythroderma and circulating tumor cells, called *Sézary's syndrome*.

Rare patients with localized early-stage mycosis fungoides can be cured with radiotherapy, often total-skin electron beam irradiation. More advanced disease has been treated with topical glucocorticoids, topical nitrogen mustard, phototherapy, psoralen with ultraviolet A (PUVA), extracorporeal photopheresis, retinoids (bexarotene), electron beam radiation, interferon, antibodies, fusion toxins, histone deacetylase inhibitors, brentuximab (for CD30+ disease), and systemic cytotoxic therapy. Mogamulizumab, an anti-CCR4 antibody, has activity in this disease and has been approved by the U.S. Food and Drug Administration for this indication. Unfortunately, these treatments are palliative.

Peripheral T-Cell Lymphoma, Not Otherwise Specified PTCLs include a number of entities, which constitute 15% of all NHLs in adults. PTCL NOS, which composes 6% of all NHLs, is the term used for cases that are not other entities defined in the WHO classification. Named varieties include ALCL, angioimmunoblastic T-cell lymphoma (AITL), hepatosplenic T-cell lymphoma, enteropathy-associated T-cell lymphoma, and subcutaneous panniculitis T-cell lymphoma. PTCL NOS is a disease of older individuals, with a median age at presentation of 65, and the majority of patients will have advanced-stage disease at diagnosis, with involvement of the bone marrow, liver, spleen, and skin being common. Associated "B" symptoms and pruritis are also common. These lymphomas can be associated with a reactive eosinophilia as well as hemophagocytic syndrome. The IPI has been applied to PTCL NOS and provides some assessment of outcomes, but even the low-risk group has a median OS of just >2 years.

This diagnostic category is a collection of heterogeneous lymphomas that vary widely and lack typical findings of other specific PTCL subgroups. Because of this heterogeneity, histology, immunophenotype, and genetics are variable. Often lymph nodes are effaced by atypical lymphoid cells of various sizes, sometimes associated with vascular proliferation or an infiltrate of eosinophils and/or macrophages. As most of these lymphomas behave aggressively, note is often made of mitotic and apoptotic figures as well as geographic necrosis. The cells often are positive for CD3, and the majority of PTCL NOS is positive for CD4 rather than CD8, but some are negative for both markers. There can be loss of more mature T-cell markers like CD5 and CD7, and this is associated with a more aggressive course. There are some recurrent translocations, including t(7;14), t(11;14), inv(14), and t(14;14), all of which involve the TCR genes.

The most common primary therapy for PTCL NOS involves a CHOP-like chemotherapy backbone—either CHOP alone or CHOP in combination with etoposide (CHOEP). The latter may provide the most benefit to younger patients and patients with more favorable disease risk factors. Brentuximab in combination with cyclophosphamide, doxorubicin, and prednisone (CHP) has been tested in a randomized clinical trial against CHOP in CD30+ T-cell lymphomas; progression-free survival was improved with the brentuximab-containing arm, and this was most pronounced for patients with ALCL (see below). Autologous stem cell transplant has been investigated for patients in their first remission and does seem to improve progression-free survival in certain contexts. Drugs such as gemcitabine, bendamustine, and pralatrexate have activity in relapsed disease, as do the histone deacetylase inhibitors romidepsin and belinostat. The PI3 kinase inhibitor duvelisib is being investigated in these diseases with early signals of activity. All of these agents are associated with transient responses in a minority of patients. Patients should be considered for clinical trials. For patients who do achieve remission, reduced-intensity allogeneic stem cell transplantation can yield long-term nonrelapse survival rates of ~40–50%.

Angioimmunoblastic T-Cell Lymphoma AITL constitutes ~20% of T-cell NHLs and ~4% of all NHLs diagnosed. Patients present with a variety of signs and symptoms, most often including lymphadenopathy, hepatosplenomegaly, "B" symptoms, rash, polyarthritis, and hemolytic anemia. Over 80% of patients have advanced-stage disease at diagnosis, and bone marrow involvement is common. Polyclonal hypergammaglobulinemia is common, as are elevated LDH, eosinophilia, a positive Coombs test, and opportunistic infections.

On biopsy, lymph nodes are effaced by a polymorphous infiltrate of lymphocytes, ranging in size and shape, and of immunoblasts. The neoplastic lymphocytes are positive for CD3 as well as CXCL13, PD-1, CD10, and BCL6, most closely resembling CD4-positive follicular helper T cells. There is an expanded follicular dendritic cell network surrounding tumor cells. Scattered immunoblasts are often EBV positive and may give rise to secondary EBV-positive B-cell lymphomas at a later time. Genetic analysis of this disease has revealed recurrent mutations in *TET2* (76%), *DNMT3* (33%), and *IDH2* (20%).

A subset of AITL can remit with immunosuppression with agents like glucocorticoids or methotrexate. Most patients, however, will need combination chemotherapy with regimens like those used in PTCL NOS. Median response duration is short, and median OS is only 15–36 months. Treatment of relapsed disease is similar to that of relapsed PTCL NOS.

Anaplastic Large-Cell Lymphoma ALCL is the next most common T-cell lymphoma after AITL but is more common in children, accounting for up to 10% of pediatric lymphomas. Approximately 40–60% of cases harbor t(2;5), which fuses a portion of the nucleolar protein nucleophosmin-1 (*NPMI*) gene to a part of the anaplastic lymphoma kinase (*ALK*) gene, the product of which has constitutive tyrosine kinase activity. These patients have a much more favorable prognosis compared to ALK-negative ALCL, akin to that of DLBCL. There is an additional, more indolent and favorable subtype that occurs in the breast tissue of patients with breast implants, and there is a cutaneous variant. In general, this is a disease that is more common in men. ALK-positive disease is a disease of younger patients, with a median age at diagnosis of 34 years, whereas the median age at diagnosis of ALK-negative patients is 58. With the exception of the cutaneous variant and the variant associated with breast implants, most patients present with rapidly growing lymphadenopathy with or without extranodal involvement; "B" symptoms are common.

Most cases of ALCL involve large atypical lymphocytes with horseshoe-shaped nuclei with prominent nucleoli ("hallmark" cells). Tumor cells tend to be localized within the lymph node sinuses, and almost all are positive for CD30 but negative for CD15. A majority will also express CD3, CD25, CD43, and CD4. ALK-rearranged ALCL can be diagnosed by fluorescence in situ hybridization (FISH) cytogenetics for t(2;5) or by immunohistochemical staining for ALK.

ALCL is generally treated with CHOP, although like PTCL NOS, CHOEP may benefit younger patients, particularly with ALK-positive disease. Overall, ALCL has a better prognosis than PTCL, and this is particularly true for ALK-positive disease, which has an 8-year OS rate of 82%, versus 49% for ALK-negative disease. Relapsed ALK-positive ALCL is treated similarly to relapsed DLBCL, with salvage combination chemotherapy to identify chemotherapy sensitivity followed by autologous stem cell transplant. For patients with chemotherapy-insensitive disease or for ALK-negative disease, the conjugated anti-CD30 antibody to monomethyl auristatin E (MMAE)/brentuximab is highly active, with a response rate of 86% and a complete response rate of 57%. As mentioned earlier, brentuximab in combination with CHP chemotherapy is an approved frontline regimen for the treatment of CD30+ T-cell lymphomas, including ALCL. The ALK inhibitors, including crizotinib, are active in refractory ALK-positive ALCL with excellent outcomes.

Other PTCL Subtypes Enteropathy-associated T-cell lymphoma, hepatosplenic T-cell lymphoma, and subcutaneous panniculitis-like T-cell lymphoma are other less common PTCL subtypes. *Enteropathy-type intestinal T-cell lymphoma* is a rare disorder. Type I occurs in patients with a history of gluten-sensitive enteropathy and is associated with HLADQA1*0501, DQB1*0201; a gluten-free diet can prevent the development of this lymphoma. Type II is now referred to as monomorphic epitheliotropic intestinal T-cell lymphoma and is not associated with celiac disease. Patients are frequently cachectic and sometimes present with intestinal perforation. The prognosis is poor, with a median survival of 10 months. Therapy is often with combination chemotherapy, including high-dose methotrexate, and autologous stem cell transplant in first remission.

Hepatosplenic γδ T-cell lymphoma is a systemic illness that presents with sinusoidal infiltration of the liver, spleen, and bone marrow by malignant T cells. Tumor masses generally do not occur. The disease is associated with systemic symptoms and is often difficult to diagnose. Recurrent genetic events include isochromosome 7q and trisomy 8. Treatment outcome is poor, but regimens that include ifosfamide, such as ifosfamide, carboplatin, and etoposide (ICE) or ifosfamide, etoposide, and cytarabine (IVAC), are associated with better outcomes in small series of patients. Responding patients should be considered for allogeneic stem cell transplantation.

Subcutaneous panniculitis-like T-cell lymphoma is a rare disorder that is often confused with panniculitis. Patients present with multiple subcutaneous nodules, which progress and can ulcerate. There is a more indolent form that tends to express α/β TCRs and can be managed with immune suppression, whereas lymphomas that express γ/δ TCRs are more aggressive and are associated with a worse prognosis and coincident hemophagocytic syndrome. This is a disease of young men in their fifth and sixth decades of life. Patients with aggressive disease are managed with multiagent chemotherapy, and responding patients should be considered for allogeneic stem cell transplantation.

Adult T-Cell Leukemia/Lymphoma Adult T-cell leukemia/lymphoma (ATLL) is a disease that is most prevalent in Japan and the Caribbean basin **(Chap. 207)**. It is a neoplasm that is driven by HTLV-1, often contracted through the breast milk of infected mothers. The average age at diagnosis is 60, so there is a long latency between viral infection and viral transformation, and only 4% of infected patients will develop the disease. This suggests that HTLV-1 may not be sufficient to cause the malignant phenotype. There are four disease variants: acute (60% of patients), lymphomatous (20% of patients), chronic (15% of patients), and smoldering (5% of patients); prognosis varies across these groups, with median survival times of 6, 10, and 24 months, and not yet reached, respectively. Presentation depends on the subtype, but most commonly, patients present with circulating disease and bone marrow involvement, hypercalcemia, lytic bone lesions, lymphadenopathy, hepatosplenomegaly, skin lesions, and opportunistic infections.

The pathognomonic finding is the malignant "flower cell" that is positive for CD4 and CD25, as well as CD2, CD3, and CD5 but lacking CD7 **(Fig. 113-7)**. Combination chemotherapy is generally used, but for patients fortunate enough to respond, response durations are very short. Other active agents in this disease include the antiretroviral agent zidovudine, interferon α, arsenic and mogamulizumab, a humanized monoclonal antibody that blocks the CCR4 chemokine receptor. In any patients who do respond to therapy, allogeneic stem cell transplant should be considered.

Extranodal NK/T-Cell Lymphoma, Nasal Type Extranodal NK/T-cell lymphoma, nasal type, is a lymphoma that is associated with

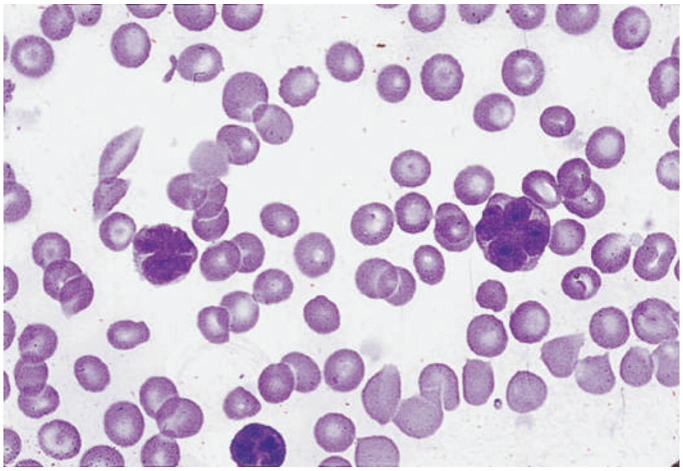

FIGURE 113-7 Adult T-cell leukemia/lymphoma. Peripheral blood smear showing leukemia cells with typical “flower-shaped” nucleus.

EBV infection in nearly all cases and more common in Asia and native populations in Peru. It usually presents with a mass and obstructive symptoms in the upper aerodigestive tract with occasional extranodal sites, but over two-thirds of patients will have localized disease. It is more common in men, and the median age at diagnosis is 60. This disease has its own prognostic score, which takes into account the presence or absence of “B” symptoms, disease stage, whether LDH is elevated, and whether there is lymph node involvement. EBV viral load at diagnosis and at the end of therapy is also predictive.

Treatment for early-stage disease is usually with combined-modality therapy of chemotherapy (commonly using etoposide, ifosfamide, cisplatin, and dexamethasone) and intensity-modulated radiation therapy (50–55 Gy), and patients with localized disease involving the nasal passages do quite well, with 3-year OS of ~85%. Patients with more advanced-stage disease do poorly, with disseminated extranodal relapse occurring frequently, and the median OS is only 4.3 months. The most commonly used treatment regimen is the SMILE regimen (dexamethasone, methotrexate, ifosfamide, L-asparaginase, and etoposide).

FURTHER READING

Horwitz S et al: The ECHELON-2 Trial: 5-year results of a randomized, phase III study of brentuximab vedotin with chemotherapy for CD30-positive peripheral T-cell lymphoma. Ann Oncol 33:288, 2022.

Morschhauser F et al: Rituximab plus lenalidomide in advanced untreated follicular lymphoma. N Engl J Med 379:934, 2018.

Tille H et al: Polatuzumab vedotin in previously untreated diffuse large B-cell lymphoma. N Engl J Med 386:351, 2022.

Westin JR et al: Survival with axi-cabtagene ciloleucel in large B-cell lymphoma. N Engl J Med 389:148, 2023.

114

Hodgkin’s Lymphoma

Caron A. Jacobson, Dan L. Longo

Hodgkin’s lymphoma (HL) is a malignancy of mature B lymphocytes. It represents ~10% of all lymphomas diagnosed each year. The majority of HL diagnoses are classical HL (cHL), but there is a second subtype of HL, nodular lymphocyte-predominant HL (NLPHL). While this diagnosis does resemble cHL morphologically in certain respects, there is some evidence that it is more related to the indolent B-cell non-Hodgkin’s lymphomas (NHLs) biologically than it is to cHL. The majority of this chapter will be specific to cHL, with a discussion of NLPHL at the end.

cHL is one of the success stories of modern oncology. Until the advent of extended-field radiotherapy in the mid-twentieth century, it was a highly fatal disease of young people. Radiation therapy cured some patients with early-stage disease, and the introduction of multiagent chemotherapy in the 1970s resulted in further improved cure rates, for both patients with early- and advanced-stage disease. Cure rates now are >85%. The new challenge in the treatment of HL is late therapy-related toxicity, including a high rate of secondary malignancies and cardiovascular disease. Current clinical trials are aimed at minimizing this risk while preserving efficacy.

EPIDEMIOLOGY AND ETIOLOGY

HL is of B-cell origin. The incidence of HL appears fairly stable, with an estimated 8830 new cases diagnosed in the United States in 2023. HL is more common in whites than in blacks and more common in males than in females. A bimodal distribution of age at diagnosis has been observed, with one peak incidence occurring in patients in their twenties and the other in those in their eighties. Some of the late age peak may be attributed to confusion among entities with similar appearance such as anaplastic large-cell lymphoma and T-cell/histiocyte–rich B-cell lymphoma. There are four distinct subtypes of cHL that are differentiated based on their histopathologic features **(Table 114-1)**: nodular sclerosis, mixed cellularity, lymphocyte-rich, and lymphocyte-depleted. Patients in the younger age groups diagnosed in the United States largely have the nodular sclerosing subtype of HL. Elderly patients, patients infected with HIV, and patients in developing countries more commonly have mixed-cellularity HL or lymphocyte-depleted HL. Together, nodular sclerosis and mixed-cellularity types account for nearly 95% of cases. Infection by HIV is a risk factor for developing HL. In addition, an association between infection by Epstein-Barr virus (EBV) and HL has been suggested. A monoclonal or oligoclonal proliferation of EBV-infected cells in 20–40% of the patients with HL has led to proposals for this virus having an etiologic role in HL. However, the matter is not settled definitively. Viral oncogenesis appears to play a greater role in HIV-related cHL: EBV can be detected in nearly all cases of HIV-associated cHL, compared to only one-third of cases of non–HIV-associated cHL. Reed-Sternberg (HRS) cells are the malignant cells in HL. HRS cells in HIV-associated cHL express the EBV-transforming protein latent membrane protein 1 (LMP-1), and the EBV genomes from multiple disease sites in the same HIV-associated cHL patient are episomal and clonal, suggesting that EBV is directly involved in early lymphomagenesis.

Histologically, the HRS cell is diagnostic of cHL **(Fig. 114-1)**. These cells are large cells with abundant cytoplasm with bilobed and/or multiple nuclei. By immunohistochemistry, they are often PAX-5 positive but have low to no expression of other B-cell antigens like CD19 and CD20. They express CD15 and CD30 in 85 and 100% of cases, respectively. These cells, though, comprise <1% of the tumor cellularity, with the majority of the tumor made up of a surrounding inflammatory infiltrate of polyclonal lymphocytes, eosinophils, neutrophils, macrophages, plasma cells, fibroblasts, and collagen. The HRS cell interacts with its microenvironment via cell-cell contact and elaboration of growth factors and cytokines, which results in a surrounding cellular milieu that protects it from host immune attack. The surrounding environmental cells likewise support the HRS cells via cell-cell signaling and cytokine production, which provides signals

TABLE 114-1 World Health Organization Classification of Hodgkin’s Lymphoma

Nodular lymphocyte-predominant Hodgkin’s lymphoma
Classical Hodgkin’s lymphoma
Nodular sclerosis
Lymphocyte-rich
Mixed cellularity
Lymphocyte-depleted

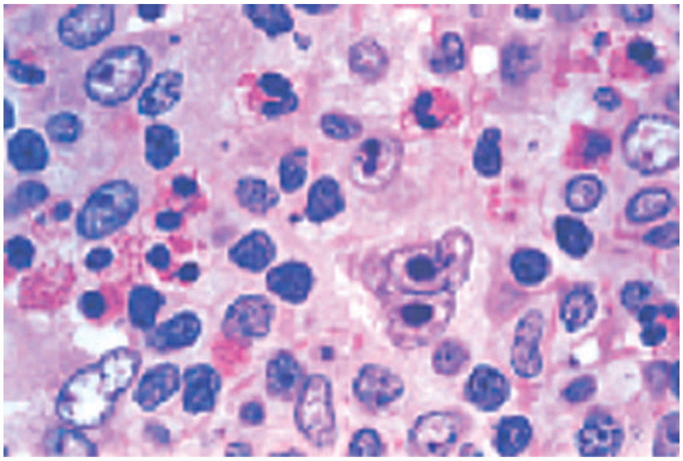

FIGURE 114-1 Hodgkin's disease: A classic Reed-Sternberg (RS) cell is present near the center of the field. RS cells are large cells with a bilobed nucleus and prominent nucleoli surrounded by a pleiomorphic cellular infiltrate. *(From DL Kasper: Harrison's Principles of Internal Medicine, 16th ed. New York, NY: McGraw-Hill; 2005, Fig. 97-11, p. 654.)*

that promote proliferation and survival of the HRS cell itself. Interestingly, 97% of HRS cells in cHL harbor genetic aberrations in the PD-L1 locus on chromosome 9p24.1, resulting in overexpression of PD-L1, the ligand for the inhibitory PD-1 receptor on immune cells. This is one mechanism whereby the HRS cell may be able to avoid immune destruction in its inflammatory microenvironment and may contribute to the generalized immune suppression in HL patients.

APPROACH TO THE PATIENT

Classical Hodgkin's Lymphoma

Most patients with cHL present with palpable lymphadenopathy that is nontender; in most patients, these lymph nodes are in the neck, supraclavicular area, and axilla. More than half of the patients will have mediastinal adenopathy at diagnosis, and this is sometimes the initial manifestation. Subdiaphragmatic presentation of cHL is unusual and more common in older males. One-third of patients present with fevers, night sweats, and/or weight loss, or "B" symptoms. Occasionally, HL can present as a fever of unknown origin. This is more common in older patients who are found to have mixed-cellularity HL in an abdominal site. Rarely, the fevers persist for days to weeks, followed by afebrile intervals and then recurrence of the fever. This pattern is known as *Pel-Ebstein* fever. HL can occasionally present with unusual manifestations. These include severe and unexplained itching, cutaneous disorders such as erythema nodosum and ichthyosiform atrophy, paraneoplastic cerebellar degeneration and other distant effects on the CNS, nephrotic syndrome, immune hemolytic anemia and thrombocytopenia, hypercalcemia, and pain in lymph nodes on alcohol ingestion.

Evaluation of patients with HL will typically begin with a careful history and physical examination. Patients should be asked about the presence or absence of "B" symptoms. Comorbid diagnoses that may impact therapy should be reviewed, including a history of pulmonary disease and congestive heart failure given the use of chemotherapy drugs that can cause both lung and heart toxicity. A physical examination should pay attention to the peripherally accessible sites of lymph nodes and to the liver and spleen size. Laboratory evaluation should include a complete blood count with differential; erythrocyte sedimentation rate (ESR); chemistry studies reflecting major organ function including serum albumin; and HIV and hepatitis virus testing. A positron emission tomography (PET)/computed tomography (CT) scan is used for staging and is more accurate than a bone marrow biopsy for evaluation of bone marrow involvement as the bone marrow involvement in cHL tends to be patchy and therefore potentially missed on a unilateral bone marrow biopsy. The initial evaluation of a patient with HL or NHL is similar. In both situations, the determination of an accurate anatomic stage is an important part of the evaluation. Staging is done using the Ann Arbor staging system **(Table 114-2)**.

TABLE 114-2 The Ann Arbor Staging System for Hodgkin's Lymphoma

STAGE	DEFINITION
I	Involvement of a single lymph node region or lymphoid structure (e.g., spleen, thymus, Waldeyer's ring)
II	Involvement of two or more lymph node regions on the same side of the diaphragm (the mediastinum is a single site; hilar lymph nodes should be considered "lateralized" and, when involved on both sides, constitute stage II disease)
III	Involvement of lymph node regions or lymphoid structures on both sides of the diaphragm
III$_1$	Subdiaphragmatic involvement limited to spleen, splenic hilar nodes, celiac nodes, or portal nodes
III$_2$	Subdiaphragmatic involvement includes paraaortic, iliac, or mesenteric nodes plus structures in III$_1$
IV	Involvement of extranodal site(s) beyond that designated as "E" More than one extranodal deposit at any location Any involvement of liver or bone marrow
A	No symptoms
B	Unexplained weight loss of >10% of the body weight during the 6 months before staging investigation Unexplained, persistent, or recurrent fever with temperatures >38°C during the previous month Recurrent drenching night sweats during the previous month
E	Localized, solitary involvement of extralymphatic tissue, excluding liver and bone marrow

The diagnosis of HL is established by review of an adequate biopsy specimen by an expert hematopathologist. HL is a tumor characterized by rare neoplastic cells of B-cell origin (immunoglobulin genes are rearranged but not expressed) in a tumor mass that is largely polyclonal inflammatory infiltrate, probably a reaction to cytokines produced by the tumor cells. The differential diagnosis of a lymph node biopsy suspicious for HL includes inflammatory processes, mononucleosis, NHL, phenytoin-induced adenopathy, and nonlymphomatous malignancies.

Staging for cHL is anatomically based given the propensity of the disease to march from one lymph node group to the next group, often contiguous to the first. Staging is important for selecting therapy of appropriate duration and intensity, but the outcome of optimal therapy for all the stages is excellent. Patients are stratified based on whether they have early-stage disease (stage I or II) or advanced-stage disease (stage III or IV). Patients with early-stage disease have a better prognosis overall but are further classified as favorable or unfavorable based on a variety of factors. These factors vary from study to study but include bulky disease, number of lymph node areas involved, an elevated ESR (>30 if "B" symptoms are present; >50 if "B" symptoms are absent), and age. Prognosis in advanced-stage disease is best predicted by the International Prognostic Score (IPS), which ascribes 1 point for male sex, older age (>45 years), stage IV disease, serum albumin <4 g/dL, hemoglobin <10.5 g/dL, white blood cell count ≥15,000/μL, and a lymphocyte count <600/μL and/or <8% of white blood cell count. Five-year progression-free survival ranges from 88% for patients with no risk factors to 62% for patients with four or more factors, but very few patients have multiple risk factors.

TREATMENT

Classical Hodgkin's Lymphoma

The overwhelming majority of patients with HL will be cured with either chemotherapy alone or a combination of chemotherapy and radiation therapy. It has long been appreciated that patients with advanced-stage disease do not benefit from the addition of radiation therapy to chemotherapy and are thus treated with chemotherapy alone. For early-stage disease, however, treatment with combined-modality therapy has been associated with a small decrease in risk of relapse but with an increased risk of late toxicity including secondary malignancies, thyroid disease, and premature cardiovascular disease and stroke resulting in minimal or no improvement in long-term survival. Much of this risk can be attributed to radiation therapy. Thus, investigation into the treatment of early-stage HL at present is aimed at trying to maximize treatment outcome without using radiotherapy. This is an area of controversy in the treatment of HL.

EARLY-STAGE DISEASE

The most common chemotherapy regimen used to treat early-stage HL in the United States is ABVD (doxorubicin, bleomycin, vinblastine, and dacarbazine). This regimen is given every other week, with each cycle including two treatments. In patients with low-risk, or favorable, disease, the use of four to six cycles of ABVD alone, without radiation therapy, results in progression-free and overall survival rates of 88–92% and 97–100%, respectively, at 5–7 years. This may be associated with a slightly increased risk of relapse when compared with abbreviated chemotherapy (ABVD for four cycles) followed by involved field radiation therapy (30 Gy), but with no difference in overall survival owing to the excellent salvage strategies used for relapsed HL and to the late toxicities seen following radiation therapy to the chest. German studies have examined a very abbreviated chemotherapy regimen (ABVD for two cycles) and low-dose radiation (20 Gy) for particularly good-risk disease with two or fewer lymph node areas involved and found that this was equally effective to standard combined-modality therapy of ABVD for four cycles and 30 Gy of radiation. However, long-term follow-up is not yet available to assess the impact of the lower radiotherapy dose on late toxicities. Finally, the use of an early interim PET/CT scan can aid decisions regarding the duration and extent of therapy. In one study, a negative PET/CT scan after three cycles of ABVD predicted for excellent outcomes with no additional therapy; in another, a negative PET/CT scan after two cycles of ABVD predicted for good outcomes with two additional cycles of ABVD alone, without radiation therapy.

For unfavorable-risk disease, the omission of radiation therapy following chemotherapy is associated with a more significant increased risk of relapse compared to favorable-risk disease, but again with no change in overall survival. For these patients, treatment options would include ABVD for four cycles followed by involved field radiation therapy or ABVD alone for six cycles. Treatment decisions are often based on the extent of the radiation field and the unfavorable risk factor, with patients with nonbulky disease being candidates for chemotherapy alone if radiation would be contraindicated for another reason. Combined modality therapy has typically been used for patients with bulky disease, although patients with bulky disease who have a negative PET/CT scan after chemotherapy may not benefit from additional radiation therapy.

Alternative chemotherapy regimens to ABVD have been developed and include the Stanford V regimen and escalated BEACOPP (bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, and prednisone). Neither of these regimens has resulted in improved outcomes in patients with early-stage disease.

ADVANCED-STAGE DISEASE

Patients with advanced-stage disease do not benefit from the addition of radiation therapy after a complete response to chemotherapy alone and should be treated with chemotherapy alone. The most common regimens used in the United States include ABVD or brentuximab vedotin plus doxorubicin, vinblastine, and dacarbazine (AVD) for six cycles. Brentuximab is an antibody-drug conjugate (ADC) that targets CD30 on the HRS cell and is conjugated to the microtubule inhibitor monomethyl auristatin E (MMAE). It was approved in the relapsed setting as a single agent and then was tested in phase 1 and 2 trials in combination with AVD chemotherapy for the upfront treatment of advanced-stage cHL. A phase 3 study, ESCHELON-1, randomized patients with advanced-stage cHL to either ABVD or brentuximab-AVD and demonstrated both a progression-free and overall survival benefit with brentuximab-AVD. Again, Stanford V and escalated BEACOPP have been evaluated in advanced-stage disease and are not associated with an improvement in overall survival but are associated with increased toxicity. The small fraction of patients who do not achieve complete remission with chemotherapy alone (partial responders with persistent PET scan positivity account for <10% of patients) may benefit from the addition of involved field radiotherapy.

Drugs that target the PD-1/PD-L1 axis have been developed for the treatment of relapsed HL based on the known genomic alterations leading to PD-L1 overexpression on the HRS cell (see "Relapsed Disease," below). In the setting of relapsed disease, these drugs, which include pembrolizumab and nivolumab, have very high response rates and are associated with durable responses. Phase 2 studies combining these drugs with either brentuximab (nivolumab) or ICE (nivolumab; ifosfamide, carboplatin, and etoposide) or GND (pembrolizumab; gemcitabine, vinorelbine, and doxorubicin) chemotherapy have demonstrated high complete response rates in order to get patients to autologous stem cell transplantation. A multicenter randomized trial comparing brentuximab-AVD to nivolumab-AVD in the upfront treatment of advanced-stage cHL has been presented, and there was a progression-free survival benefit with nivolumab-AVD and the regimen had an arguably more favorable toxicity profile. We await longer follow-up and U.S. Food and Drug Administration review of these results to determine if this study establishes a new standard of care for the upfront treatment of advanced-stage cHL.

RELAPSED DISEASE

Patients who relapse after primary therapy of HL can frequently still be cured. Patients who relapse after an effective chemotherapy regimen are usually not curable with subsequent chemotherapy administered at standard doses. Alternative salvage chemotherapy administered at standard doses, then, is given in order to document sensitivity to chemotherapy and to achieve maximum reduction of tumor mass. For patients who respond completely or nearly so, autologous stem cell transplantation can cure over half of patients. Standard salvage chemotherapy regimens have historically included ICE and GND. Newer combinations, including brentuximab with immune checkpoint inhibitors such as nivolumab, have also been tested with promising results for patients who have not seen brentuximab in the frontline setting. The combinations of ICE or GND with nivolumab or pembrolizumab, respectively, are similarly highly effective and lead to high complete remission rates and success in getting patients to definitive autologous stem cell transplantation. These novel combinations have largely replaced standard salvage chemotherapy approaches, as the chemotherapy/immunotherapy combinations appear to be more effective and are associated with higher rates of durable response after autologous stem cell transplantation, perhaps owing to the chemosensitizing effects that have been observed following immune checkpoint inhibition. Studies are ongoing to investigate whether autologous transplant is necessary for patients who have a favorable response to chemotherapy/immunotherapy combinations based on this observed phenomenon. For patients with early-stage disease who do not respond sufficiently to salvage chemotherapy, radiation therapy can be very effective to achieve a remission; whether to consolidate such a remission with an autologous stem cell transplant is debated. Brentuximab is also used as a maintenance therapy following successful autologous stem cell transplantation based on results of the AETHERA study, a randomized trial of brentuximab maintenance versus observation. Finally, anti-CD30 chimeric antigen receptor (CAR) T-cell therapy has been tested in multiply relapsed cHL with promising early results; these products are now being tested in multicenter phase 2 clinical trials.

SURVIVORSHIP

Because of the very high cure rate in patients with HL, long-term complications have become a major focus for clinical research. In fact, in some series of patients with early-stage disease, more patients died from late complications of therapy than from HL itself. This is particularly true in patients with localized disease. The most serious late side effects include second malignancies and cardiac injury. Patients are at risk for the development of acute leukemia in the first 10 years after treatment with combination chemotherapy regimens that contain alkylating agents plus radiation therapy. The risk for development of acute leukemia is greater after MOPP-like (mechlorethamine, vincristine, procarbazine, and prednisone) and BEACOPP-like regimens than with ABVD or brentuximab-AVD. The risk of development of acute leukemia after treatment for HL is also related to the number of exposures to potentially leukemogenic agents (i.e., multiple treatments after relapse) and the age of the patient being treated, with those aged >60 years at particularly high risk. The development of carcinomas as a complication of treatment for HL is a major problem. These tumors usually occur ≥10 years after treatment and are associated with use of radiotherapy. For this reason, young women treated with thoracic radiotherapy for HL should institute screening mammograms or breast MRI exams 5–10 years after treatment, and all patients who receive thoracic radiotherapy for HL should be discouraged from smoking. Mediastinal radiation also accelerates coronary artery disease, and patients should be encouraged to minimize risk factors for coronary artery disease such as smoking and elevated cholesterol levels. Cervical radiation therapy increases the risk of carotid atherosclerosis and stroke and thyroid disease, including cancer.

A number of other late side effects from the treatment of HL are well known. Patients who receive thoracic radiotherapy are at very high risk for the eventual development of hypothyroidism and should be observed for this complication; intermittent measurement of thyrotropin should be made to identify the condition before it becomes symptomatic. Lhermitte's syndrome occurs in ~15% of patients who receive thoracic radiotherapy. This syndrome is manifested by an "electric shock" sensation into the lower extremities on flexion of the neck. Because of the young age at which HL is often diagnosed, infertility is a concern for patients undergoing treatment for HL. Chemotherapy regimens containing alkylating agents induce permanent infertility in nearly all men. The risk of permanent infertility in women treated with alkylating agent–containing chemotherapy is age-related, with younger women more likely to recover fertility. Infertility is very rare after treatment with ABVD.

NODULAR LYMPHOCYTE-PREDOMINANT HODGKIN'S LYMPHOMA

NLPHL is now recognized as an entity distinct from cHL. Previous classification systems recognized that biopsies from a small subset of patients diagnosed as having HL contained a predominance of small lymphocytes and rare Reed-Sternberg–like cells; tumors had a nodular growth pattern and a clinical course that varied from that of patients with cHL. This is an unusual clinical entity and represents <5% of cases of HL and defines NLPHL.

NLPHL has a number of characteristics that suggest its relationship to NHL, rather than cHL, however. The HRS-like cell, or L&H (lymphocyte and histiocyte) or "popcorn" cell, is a clonal proliferation of B cells that are positive for B-cell markers CD45, CD79a, CD20, CD19, and BCL2. They do not express two markers normally found on HRS cells, CD30 and CD15. This lymphoma tends to have a chronic, relapsing course and sometimes transforms to diffuse large B-cell lymphoma, including a specific subtype of diffuse large B-cell lymphoma known as T-cell/histiocyte–rich B-cell lymphoma, which shares an immunophenotype with the L&H cell. This natural history most closely resembles that of the indolent B-cell NHLs outlined in **Chaps. 113** and **115**.

Patients with NLPHL are more commonly male (75%). Like cHL, the age distribution of patients with this disease has two peaks, but unlike cHL, these peaks include children and adults age 30–40 years, respectively. The majority of patients diagnosed have stage I or II disease (75%), with a minority having advanced-stage disease at diagnosis. "B" symptoms are uncommon.

Patients with early-stage disease at diagnosis should be treated with definitive radiotherapy. This is associated with a 15-year nonrelapse survival rate of 82%. The treatment of patients with advanced-stage NLPHL is controversial. Some clinicians favor no treatment of asymptomatic disease and merely close follow-up, akin to the indolent B-cell NHLs. For patients who need therapy due to symptoms or signs of organ function impairment, both cHL regimens and B-cell NHL regimens have been used, including ABVD and R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone). A single-institution experience with R-CHOP resulted in a 100% response rate in a small group of patients without a single relapse with 42 months of follow-up. Although this is short follow-up for an indolent disease, some believe R-CHOP may be curative in this disease and advocate treating patients with advanced-stage disease at diagnosis, regardless of symptoms or organ function.

■ FURTHER READING

Ansell SM et al: Overall survival with brentuximab vedotin in stage III or IV Hodgkin's lymphoma. N Engl J Med 387:310, 2022.

Chen R et al: Pembrolizumab in relapsed or refractory Hodgkin lymphoma: 2-year follow-up of KEYNOTE-087. Blood 134:1144, 2019.

Gillessen S et al: Intensified treatment of patients with early stage, unfavourable Hodgkin lymphoma: Long-term follow-up of a randomised, international phase 3 trial of the German Hodgkin Study Group (GHSG HD14). Lancet Haematol 8:e278, 2021.

Herrera AF et al: Nivolumab+AVD in advanced-stage classic Hodgkin's lymphoma. N Engl J Med 391:1379, 2024.

Moskowitz CH et al: Five-year PFS from the AETHERA trial of brentuximab vedotin for Hodgkin lymphoma at high risk of progression or relapse. Blood 132:2639, 2018.

Rashidi A et al: Allogeneic hematopoietic stem cell transplantation in Hodgkin lymphoma: A systemic review and meta-analysis. Bone Marrow Transplant 51:521, 2016.

Straus DJ et al: CALGB 50604: Risk-adapted treatment of nonbulky early-stage Hodgkin lymphoma based on interim PET. Blood 132:1013, 2018.

115

Less Common Lymphoid and Myeloid Malignancies

Ayalew Tefferi, Dan L. Longo

The most common lymphoid malignancies are discussed in **Chaps. 111, 112, 113, 114, and 116**, myeloid leukemias in **Chaps. 109 and 110**, myelodysplastic syndromes (MDS) in **Chap. 107**, and myeloproliferative syndromes in **Chap. 108**. This chapter will focus on the more unusual forms of hematologic malignancy. The diseases discussed here are listed in **Table 115-1**. Each of these entities accounts for <1% of hematologic neoplasms.

RARE LYMPHOID MALIGNANCIES

All the lymphoid tumors discussed here are mature B-cell or T-cell natural killer (NK) cell neoplasms.

TABLE 115-1 Unusual Lymphoid and Myeloid Malignancies[a]

Lymphoid
Mature B-cell neoplasms
B-cell prolymphocytic leukemia
Splenic marginal zone lymphoma
Hairy cell leukemia
Nodal marginal zone B-cell lymphoma
Mediastinal large B-cell lymphoma
Intravascular large B-cell lymphoma
Primary effusion lymphoma
Lymphomatoid granulomatosis
Mature T-cell and natural killer (NK) cell neoplasms
T-cell prolymphocytic leukemia
T-cell large granular lymphocytic leukemia
Aggressive NK cell leukemia
Extranodal NK/T-cell lymphoma, nasal type
Enteropathy-type T-cell lymphoma
Hepatosplenic T-cell lymphoma
Subcutaneous panniculitis-like T-cell lymphoma
Blastic NK cell lymphoma
Primary cutaneous CD30+ T-cell lymphoma
Angioimmunoblastic T-cell lymphoma
Myeloid
Chronic neutrophilic leukemia
Chronic eosinophilic leukemia/hypereosinophilic syndrome
Histiocytic and Dendritic Cell Neoplasms
Histiocytic sarcoma
Langerhans cell histiocytosis
Langerhans cell sarcoma
Interdigitating dendritic cell sarcoma
Follicular dendritic cell sarcoma
Mast cells
Mastocytosis
Cutaneous mastocytosis
Systemic mastocytosis
Mast cell sarcoma
Extracutaneous mastocytoma

[a] This list is not exhaustive. Many named entities are very rare and not discussed here. A complete listing is available in the online version of this chapter.

MATURE B-CELL NEOPLASMS

B-Cell Prolymphocytic Leukemia (B-PLL) This is a malignancy of medium-sized (about twice the size of a normal small lymphocyte), round lymphocytes with a prominent nucleolus and light blue cytoplasm on Wright's stain. It predominantly affects the blood, bone marrow (BM), and spleen and usually does not cause adenopathy. The median age of affected patients is 70 years, and men are more often affected than women (male-to-female ratio is 1.6). This entity is distinct from chronic lymphoid leukemia (CLL) and does not develop as a consequence of that disease.

Clinical presentation is generally from symptoms of splenomegaly or incidental detection of an elevated white blood cell (WBC) count. The clinical course can be rapid. The cells express surface IgM (with or without IgD) and typical B-cell markers (CD19, CD20, CD22). CD23 is absent, and about one-third of cases express CD5. The CD5 expression along with the presence of the t(11;14) translocation in 20% of cases leads to confusion in distinguishing B-PLL from the leukemic form of mantle cell lymphoma. No reliable criteria for the distinction have emerged, and gene expression studies suggest a close relationship between mantle cell lymphoma and B-PLL and significant differences with CLL. About half of patients have mutation or loss of p53, and deletions have been noted in 11q23 and 13q14. Nucleoside analogues like fludarabine and cladribine and combination chemotherapy (cyclophosphamide, doxorubicin, vincristine, and prednisone [CHOP]) have produced responses. CHOP plus rituximab may be more effective than CHOP alone, but the disease is sufficiently rare that large series have not been reported. Splenectomy can produce palliation of symptoms but appears to have little or no impact on the course of the disease. BM transplantation may be curative. Imatinib may also have activity.

Splenic Marginal Zone Lymphoma (SMZL) This tumor of mainly small lymphocytes originates in the marginal zone of the spleen white pulp, grows to efface the germinal centers and mantle, and invades the red pulp. Splenic hilar nodes, BM, and peripheral blood (PB) may be involved. The circulating tumor cells have short surface villi and are called villous lymphocytes. Table 115-2 shows differences in tumor cells of a number of neoplasms of small lymphocytes that aid in the differential diagnosis. SMZL cells express surface immunoglobulin and CD20 but are negative for CD5, CD10, CD43, and CD103. Lack of CD5 distinguishes SMZL from CLL, and lack of CD103 separates SMZL from hairy cell leukemia.

The median age of patients with SMZL is mid-fifties, and men and women are equally represented. Patients present with incidental or symptomatic splenomegaly or incidental detection of lymphocytosis in the PB with villous lymphocytes. Autoimmune anemia or thrombocytopenia may be present. The immunoglobulin produced by these cells contains somatic mutations that reflect transit through a germinal center, and ongoing mutations suggest that the mutation machinery has remained active. About 40% of patients have either deletions or translocations involving 7q21, the site of the *FLNC* gene (filamin Cγ, involved in cross-linking actin filaments in the cytoplasm). *NOTCH2* mutations are seen in 25% of patients. Chromosome 8p deletions may also be noted. The genetic lesions typically found in extranodal marginal zone lymphomas (e.g., trisomy 3 and t[11;18]) are uncommon in SMZL.

The clinical course of disease is generally indolent with median survivals exceeding 10 years. Patients with elevated lactate dehydrogenase (LDH) levels, anemia, and hypoalbuminemia generally have a poorer prognosis. Long remissions can be seen after splenectomy. Rituximab, ibrutinib, and PI3 kinase inhibitors are also active. A small fraction of patients undergo histologic progression to diffuse large B-cell lymphoma with a concomitant change to a more aggressive natural history. Experience with combination chemotherapy in SMZL is limited.

Hairy Cell Leukemia Hairy cell leukemia is a tumor of small lymphocytes with oval nuclei, abundant cytoplasm, and distinctive membrane projections (hairy cells). Patients have splenomegaly and diffuse BM involvement. While some circulating cells are noted, the clinical picture is dominated by symptoms from the enlarged spleen and pancytopenia. The mechanism of the pancytopenia is not completely clear and may be mediated by both inhibitory cytokines and

TABLE 115-2 Immunophenotype of Tumors of Small Lymphocytes

	CD5	CD20	CD43	CD10	CD103	sIG	CYCLIN D1
Follicular lymphoma	neg	pos	pos	pos	neg	pos	neg
Chronic lymphoid leukemia	pos	pos	pos	neg	neg	pos	neg
B-cell prolymphocytic leukemia	pos	pos	pos	neg	neg	pos	pos
Mantle cell lymphoma	pos	pos	pos	neg	neg	pos	pos
Splenic marginal zone lymphoma	neg	pos	neg	neg	neg	pos	neg
Hairy cell leukemia	neg	pos	?	neg	pos	pos	neg

Abbreviations: neg, negative; pos, positive.

TABLE 115-3 Differential Diagnosis of "Dry Tap"—Inability to Aspirate Bone Marrow

Dry taps occur in about 4% of attempts and are associated with:	
Metastatic carcinoma infiltration	17%
Chronic myeloid leukemia	15%
Myelofibrosis	14%
Hairy cell leukemia	10%
Acute leukemia	10%
Lymphomas, Hodgkin's disease	9%
Normal marrow	Rare

marrow replacement. The marrow has an increased level of reticulin fibers; indeed, hairy cell leukemia is a common cause of inability to aspirate BM or so-called "dry tap" (Table 115-3). Monocytopenia is profound and may explain a predisposition to atypical mycobacterial infection that is observed clinically. The tumor cells have strong expression of CD22, CD25, and CD103; soluble CD25 level in serum is an excellent tumor marker for disease activity. The cells also express tartrate-resistant acid phosphatase. The immunoglobulin genes are rearranged and mutated, indicating the influence of a germinal center. No specific cytogenetic abnormality has been found, but most cases contain the activating *BRAF* mutation V600E.

The median age of affected patients is mid-fifties, and the male-to-female ratio is 5:1. Treatment options are numerous. Splenectomy is often associated with prolonged remission. Nucleosides including cladribine and deoxycoformycin are highly active but are also associated with further immunosuppression and can increase the risk of certain opportunistic infections. However, after brief courses of these agents, patients usually obtain very durable remissions during which immune function spontaneously recovers. Interferon α is also an effective therapy but is not as effective as nucleosides. Chemotherapy-refractory patients have responded to vemurafenib, a BRAF inhibitor. Vemurafenib does not appear to be curative, but responses can be maintained with chronic treatment. More durable remissions occur when rituximab is added to vemurafenib.

Nodal Marginal Zone B-Cell Lymphoma This rare node-based disease bears an uncertain relationship with extranodal marginal zone lymphomas, which are often mucosa-associated and are called mucosa-associated lymphoid tissue (MALT) lymphomas, and SMZLs. Patients may have localized or generalized adenopathy. The neoplastic cell is a marginal zone B cell with monocytoid features and has been called monocytoid B-cell lymphoma in the past. Up to one-third of the patients may have extranodal involvement, and involvement of the lymph nodes can be secondary to the spread of a mucosal primary lesion. In authentic nodal primaries, the cytogenetic abnormalities associated with MALT lymphomas (trisomy 3 and t[11;18]) are very rare. The clinical course is indolent. Patients often respond to combination chemotherapy, although remissions have not been durable. Few patients have received CHOP plus rituximab, which is likely to be an effective approach to management.

Mediastinal (Thymic) Large B-Cell Lymphoma This entity was originally considered a subset of diffuse large B-cell lymphoma; however, additional study has identified it as a distinct entity with its own characteristic clinical, genetic, and immunophenotypic features. This is a disease that can be bulky in size but usually remains confined to the mediastinum. It can be locally aggressive, including progressing to produce a superior vena cava obstruction syndrome or pericardial effusion. About one-third of patients develop pleural effusions, and in 5–10% of cases, disease can disseminate widely to kidney, adrenal, liver, skin, and even brain. The disease affects women more often than men (male-to-female ratio is 1:2–3), and the median age is 35–40 years.

The tumor is composed of sheets of large cells with abundant cytoplasm accompanied by variable, but often abundant, fibrosis. It is distinguished from nodular sclerosing Hodgkin's disease by the paucity of normal lymphoid cells and the absence of lacunar variants of Reed-Sternberg cells. However, more than one-third of the genes that are expressed to a greater extent in primary mediastinal large B-cell lymphoma than in usual diffuse large B-cell lymphoma are also overexpressed in Hodgkin's disease, suggesting a possible pathogenetic relationship between the two entities that affect the same anatomic site. Tumor cells may overexpress *MAL*. The genome of tumor cells is characterized by frequent chromosomal gains and losses. The tumor cells in mediastinal large B-cell lymphoma express CD20, but surface immunoglobulin and human leukocyte antigen (HLA) class I and class II molecules may be absent or incompletely expressed. Expression of lower levels of class II HLA identifies a subset with poorer prognosis. The cells are CD5 and CD10 negative but may show light staining with anti-CD30. The cells are CD45 positive, unlike cells of classical Hodgkin's disease.

Methotrexate, leucovorin, doxorubicin, cyclophosphamide, vincristine, prednisone, and bleomycin (MACOP-B) and rituximab plus CHOP are effective treatments, achieving 5-year survival of 75–87%. Dose-adjusted therapy with prednisone, etoposide, vincristine, cyclophosphamide, and doxorubicin (EPOCH) plus rituximab has produced 5-year survival of 97%. A role for mediastinal radiation therapy has not been definitively demonstrated, but it is frequently used, especially in patients whose mediastinal area remains positron emission tomography–avid after 4–6 cycles of chemotherapy.

Intravascular Large B-Cell Lymphoma This is an extremely rare form of diffuse large B-cell lymphoma characterized by the presence of lymphoma in the lumen of small vessels, particularly capillaries. It is also known as malignant angioendotheliomatosis or angiotropic large-cell lymphoma. It is sufficiently rare that no consistent picture has emerged to define a clinical syndrome or its epidemiologic and genetic features. It is thought to remain inside vessels because of a defect in adhesion molecules and homing mechanisms, an idea supported by scant data suggesting absence of expression of β-1 integrin and ICAM-1. Patients commonly present with symptoms of small-vessel occlusion, skin lesions, or neurologic symptoms. The tumor cell clusters can promote thrombus formation. A subset of patients have tumors with *MYD88* or *CD79B* mutations. In general, the clinical course is aggressive and the disease is poorly responsive to therapy. Often a diagnosis is not made until very late in the course of the disease or at autopsy. Diagnosis may be revealed in random skin biopsies in settings where localized findings are limited.

Primary Effusion Lymphoma This entity is another variant of diffuse large B-cell lymphoma that presents with pleural effusions, usually without apparent tumor mass lesions. It is most common in the setting of immune deficiency disease, especially AIDS, and is caused by human herpes virus 8 (HHV-8)/Kaposi's sarcoma herpes virus (KSHV). It is also known as *body cavity–based lymphoma*. Some patients have been previously diagnosed with Kaposi's sarcoma. It can also occur in the absence of immunodeficiency in elderly men of Mediterranean heritage, similar to Kaposi's sarcoma but even less common.

The malignant effusions contain cells positive for HHV-8/KSHV, and many are also co-infected with Epstein-Barr virus. The cells are large with large nuclei and prominent nucleoli that can be confused with Reed-Sternberg cells. The cells express CD20 and CD79a (immunoglobulin-signaling molecule), although they often do not express immunoglobulin. Some cases aberrantly express T-cell markers such as CD3 or rearranged T-cell receptor genes. No characteristic genetic lesions have been reported, but gains in chromosome 12 and X material have been seen, similar to other HIV-associated lymphomas. The clinical course is generally characterized by rapid progression and death within 6 months. CHOP plus lenalidomide or bortezomib may produce responses. Highly active antiretroviral therapy for HIV should be maintained during treatment.

Lymphomatoid Granulomatosis This is an angiocentric, angiodestructive lymphoproliferative disease comprised by neoplastic Epstein-Barr virus–infected monoclonal B cells accompanied and outnumbered by a polyclonal reactive T-cell infiltrate. The disease is graded based on histologic features such as cell number and atypia in the B cells. It is most often confused with extranodal NK/T-cell lymphoma, nasal type, which can also be angiodestructive and is Epstein-Barr virus–related. The disease usually presents in adults (males > females) as a pulmonary infiltrate. Involvement is often entirely extranodal and can include kidney (32%), liver (29%), skin (25%), and brain (25%). The disease often but not always occurs in the setting of immune deficiency.

The disease can be remitting and relapsing in nature or can be rapidly progressive. The course is usually predicted by the histologic grade. The disease is highly responsive to combination chemotherapy and is curable in most cases. Some investigators have claimed that low-grade disease (grade I and II) can be treated with interferon α.

MATURE T-CELL AND NK CELL NEOPLASMS

T-Cell Prolymphocytic Leukemia This is an aggressive leukemia of medium-sized prolymphocytes involving the blood, marrow, nodes, liver, spleen, and skin. It accounts for 1–2% of all small lymphocytic leukemias. Most patients present with elevated WBC count (often >100,000/μL), hepatosplenomegaly, and adenopathy. Skin involvement occurs in 20%. The diagnosis is made from PB smear, which shows cells about 25% larger than those in small lymphocytes, with cytoplasmic blebs and nuclei that may be indented. The cells express T-cell markers like CD2, CD3, and CD7; two-thirds of patients have cells that are CD4+ and CD8–, and 25% have cells that are CD4+ and CD8+. T-cell receptor β chains are clonally rearranged. In 80% of patients, inversion of chromosome 14 occurs between q11 and q32. Ten percent have t(14;14) translocations that bring the T-cell receptor alpha/beta gene locus into juxtaposition with oncogenes *TCL1* and *TCL1b* at 14q32.1. Chromosome 8 abnormalities are also common. Deletions in the *ATM* gene are also noted. Activating *JAK3* mutations have also been reported.

The course of the disease is generally rapid, with median survival of about 12 months. Responses have been seen with the anti-CD52 antibody alemtuzumab, nucleoside analogues, and CHOP chemotherapy. Histone deacetylase inhibitors like vorinostat and romidepsin may also have activity. Small numbers of patients with T-cell prolymphocytic leukemia have also been treated with high-dose therapy, and allogeneic BM transplantation after remission has been achieved with alemtuzumab or conventional-dose therapy.

T-Cell Large Granular Lymphocytic Leukemia T-cell large granular lymphocytic (LGL) leukemia is characterized by increases in the number of LGLs in the PB (2000–20,000/μL) often accompanied by severe neutropenia, with or without concomitant anemia. Pure red cell aplasia may occur in 15–20% of patients. Splenomegaly is seen in 25% of patients; adenopathy is generally absent. B symptoms are rare, but 20–30% of patients may have infections related to the severe neutropenia. Patients may have splenomegaly and frequently have evidence of systemic autoimmune disease, including rheumatoid arthritis, hypergammaglobulinemia, autoantibodies, and circulating immune complexes. BM involvement is mainly interstitial in pattern, with <50% lymphocytes on differential count. Usually the cells express CD3, T-cell receptors, usually TCRα/β, and CD8; NK-like variants may be CD3–. Like other T-cell neoplasms, loss of expression of CD5 and/or CD7 is common. The leukemic cells often express Fas and Fas ligand. The JAK/STAT pathway is often activated.

The course of the disease is generally indolent and dominated by the neutropenia. Paradoxically, immunosuppressive therapy with cyclosporine, methotrexate, or cyclophosphamide plus glucocorticoids can produce an increase in granulocyte counts. Nucleosides have been used anecdotally. Occasionally the disease can accelerate to a more aggressive clinical course.

Aggressive NK Cell Leukemia NK neoplasms are very rare, and they may follow a range of clinical courses from very indolent to highly aggressive. They are more common in Asians than whites, and the cells frequently harbor a clonal Epstein-Barr virus episome. The PB white count is usually not greatly elevated, but abnormal large lymphoid cells with granular cytoplasm are noted. The aggressive form is characterized by symptoms of fever and laboratory abnormalities of pancytopenia. Hepatosplenomegaly is common; node involvement is less common. Patients may have hemophagocytosis, coagulopathy, or multiorgan failure. Serum levels of Fas ligand are elevated.

The cells express CD2 and CD56 and do not have rearranged T-cell receptor genes. Deletions involving chromosome 6 are common. The disease can be rapidly progressive. Some forms of NK neoplasms are more indolent. They tend to be discovered incidentally with LGL lymphocytosis and do not manifest the fever and hepatosplenomegaly characteristic of the aggressive leukemia. The cells are also CD2 and CD56 positive, but they do not contain clonal forms of Epstein-Barr virus and are not accompanied by pancytopenia or autoimmune disease.

Extranodal NK/T-Cell Lymphoma, Nasal Type Like lymphomatoid granulomatosis, extranodal NK/T-cell lymphoma tends to be an angiocentric and angiodestructive lesion, but the malignant cells are not B cells. In most cases, they are CD56+ Epstein-Barr virus–infected cells; occasionally, they are CD56–Epstein-Barr virus–infected cytotoxic T cells. They are most commonly found in the nasal cavity. Historically, this illness was called lethal midline granuloma, polymorphic reticulosis, and angiocentric immunoproliferative lesion. This form of lymphoma is prevalent in Asia, Mexico, and Central and South America; it affects males more commonly than females. When it spreads beyond the nasal cavity, it may affect soft tissue, the gastrointestinal tract, or the testis. In some cases, hemophagocytic syndrome (HPS) may influence the clinical picture. Patients may have B symptoms. Many of the systemic manifestations of disease are related to the production of cytokines by the tumor cells and the cells responding to their signals. Deletions and inversions of chromosome 6 are common.

Many patients with extranodal NK/T-cell lymphoma, nasal type, have excellent antitumor responses with combination chemotherapy regimens, particularly those with localized disease. Radiation therapy is often used after completion of chemotherapy. Four risk factors have been defined, including B symptoms, advanced stage, elevated LDH, and regional lymph node involvement. Patient survival is linked to the number of risk factors: 5-year survival is 81% for zero risk factors, 64% for one risk factor, 32% for two risk factors, and 7% for three or four risk factors. Combination regimens without anthracyclines have been touted as superior to CHOP, but data are sparse. High-dose therapy with stem cell transplantation has been used, but its role is unclear.

Enteropathy-Type T-Cell Lymphoma Enteropathy-type T-cell lymphoma is a rare complication of longstanding celiac disease. It most commonly occurs in the jejunum or the ileum. In adults, the lymphoma may be diagnosed at the same time as celiac disease, but the suspicion is that the celiac disease was a longstanding precursor to the development of lymphoma. The tumor usually presents as multiple ulcerating mucosal masses but may also produce a dominant exophytic mass or multiple ulcerations. The tumor expresses CD3 and CD7 nearly always and may or may not express CD8. The tumor may express CD30, but therapies directed at CD30 have not been adequately tested. The normal-appearing lymphocytes in the adjacent mucosa often have a similar phenotype to the tumor. Most patients have the HLA genotype associated with celiac disease, HLA DQA1*0501 or DQB1*0201.

The prognosis of this form of lymphoma is typically poor (median survival is 7–11 months), but some patients have a good response to CHOP chemotherapy. Patients who respond can develop bowel perforation from responding tumor. If the tumor responds to treatment, recurrence may develop elsewhere in the celiac disease–affected small bowel.

An indolent form of T-cell or NK cell lymphoma occurs rarely that affects mainly the small intestine and presents with dyspepsia, vomiting, and diarrhea. The cells often contain genetic changes that result in JAK-STAT activation. The disease is most often chronic with little or no propensity to spread and develop aggressive growth. A variety of approaches have been tested; none are reliably curative.

Hepatosplenic T-Cell Lymphoma Hepatosplenic T-cell lymphoma is a malignancy derived from T cells expressing the gamma/delta

T-cell antigen receptor that affects mainly the liver and fills the sinusoids with medium-size lymphoid cells. When the spleen is involved, dominantly the red pulp is infiltrated. It is a disease of young people, especially young people with an underlying immunodeficiency or with an autoimmune disease that demands immunosuppressive therapy. The use of thiopurine and infliximab is particularly common in the history of patients with this disease. The cells are CD3+ and usually CD4– and CD8–. The cells may contain isochromosome 7q, often together with trisomy 8. The lymphoma has an aggressive natural history. Combination chemotherapy may induce remissions, but most patients relapse. Cytarabine/etoposide/platinum-based regimens appear more effective than CHOP-based regimens. Median survival is about 2 years. The tumor does not appear to respond to reversal of immunosuppressive therapy.

Subcutaneous Panniculitis-Like T-Cell Lymphoma Subcutaneous panniculitis-like T-cell lymphoma involves multiple subcutaneous collections of neoplastic T cells that are usually cytotoxic cells in phenotype (i.e., contain perforin and granzyme B and express CD3 and CD8). The rearranged T-cell receptor is usually alpha/beta-derived, but occasionally, the gamma/delta receptors are involved, particularly in the setting of immunosuppression. The cells are negative for Epstein-Barr virus. A history of autoimmune disease, particularly lupus erythematosus, in the patient or the family is present in almost one-third of patients. Patients may have an HPS in addition to the skin infiltration; fever and hepatosplenomegaly may also be present. Nodes are generally not involved. Patients frequently respond to combination chemotherapy, including CHOP. When the disease is progressive, the HPS can be a component of a fulminant downhill course. Effective therapy can reverse the HPS.

Blastic NK Cell Lymphoma The neoplastic cells express NK cell markers, especially CD56, and are CD3 negative. They are large blastic-appearing cells and may produce a leukemia picture, but the dominant site of involvement is the skin. Morphologically, the cells are similar to the neoplastic cells in acute lymphoid and myeloid leukemia. No characteristic chromosomal abnormalities have been described. The clinical course is rapid, and the disease is largely unresponsive to typical lymphoma treatments.

Primary Cutaneous CD30+ T-Cell Lymphoma This tumor involves the skin and is composed of cells that appear similar to the cells of anaplastic T-cell lymphoma. Among cutaneous T-cell tumors, ~25% are CD30+ anaplastic lymphomas. If dissemination to lymph nodes occurs, it is difficult to distinguish between the cutaneous and systemic forms of the disease. The tumor cells are often CD4+, and the cells contain granules that are positive for granzyme B and perforin in 70% of cases. The typical t(2;5) of anaplastic T-cell lymphoma is absent; indeed, its presence should prompt a closer look for systemic involvement and a switch to a diagnosis of anaplastic T-cell lymphoma. This form of lymphoma has sporadically been noted as a rare complication of silicone or saline breast implants. The natural history of breast implant–associated lymphoma is generally indolent. Cutaneous CD30+ T-cell lymphoma often responds to therapy. The anti-CD30 immunotoxin conjugate brentuximab vedotin is active. Radiation therapy can be effective, and surgery can also produce long-term disease control. Five-year survival exceeds 90%.

Angioimmunoblastic T-Cell Lymphoma Angioimmunoblastic T-cell lymphoma is a systemic disease that accounts for ~15% of all T-cell lymphomas. Patients frequently have fever, advanced stage, diffuse adenopathy, hepatosplenomegaly, skin rash, polyclonal hypergammaglobulinemia, and a wide range of autoantibodies including cold agglutinins, rheumatoid factor, and circulating immune complexes. Patients may have edema, arthritis, pleural effusions, and ascites. The nodes contain a polymorphous infiltrate of neoplastic T cells and non-neoplastic inflammatory cells together with proliferation of high endothelial venules and follicular dendritic cells (FDCs). The most common chromosomal abnormalities are trisomy 3, trisomy 5, and an extra X chromosome. Aggressive combination chemotherapy can induce regressions. The underlying immune defects make conventional lymphoma treatments more likely to produce infectious complications.

RARE MYELOID MALIGNANCIES

The World Health Organization (WHO) and the International Consensus Classification (ICC) systems use PB counts, PB smear analysis, BM morphology, and cytogenetic and molecular genetic tests in order to classify myeloid malignancies into several major categories **(Table 115-4)**. Among them, acute myeloid leukemia (AML) and AML-related disorders are discussed in **Chap. 109**, MDS and MDS/AML in **Chap. 107**, chronic myeloid leukemia (CML) in **Chap. 110**, and *JAK2* mutation–prevalent myeloproliferative neoplasms (MPN), including essential thrombocythemia, polycythemia vera, and primary myelofibrosis, in **Chap. 108**. In this chapter, we focus on some of the remaining myeloid neoplasms listed in Table 115-4, which are less frequent: (1) other MPNs including chronic neutrophilic leukemia (CNL), chronic eosinophilic leukemia, not otherwise specified (CEL-NOS), and MPN, unclassifiable (MPN-U); (2) MDS/MPN overlap including chronic myelomonocytic leukemia (CMML), atypical CML (aCML), MDS/MPN with mutated *SF3B1* and thrombocytosis, MDS/MPN with ring sideroblasts and thrombocytosis, not otherwise specified (MDS/MPN-RS-T-NOS), and MDS/MPN, not otherwise specified (MDS/MPN-NOS); (3) juvenile myelomonocytic leukemia (JMML); (4) transient myeloproliferative disorder (TMD); (5) hypereosinophilia including those associated with tyrosine kinase gene fusions (TKGFs) and hypereosinophilic syndrome (HES); (6) mastocytosis; and (7) histiocytic and dendritic cell neoplasms (hemophagocytic lymphohistiocytosis [HLH] is discussed in **Chap. 68**).

■ CHRONIC NEUTROPHILIC LEUKEMIA

CNL is a clonal proliferation of mature neutrophils with few or no circulating immature granulocytes. Other clinical features include hepatosplenomegaly and constitutional symptoms. The disease is molecularly characterized by activating mutations of the gene (*CSF3R*) encoding for the receptor for granulocyte colony-stimulating factor (G-CSF), also known as colony-stimulating factor 3 (CSF3). Patients with CNL might be asymptomatic at presentation but can also display constitutional symptoms, splenomegaly, anemia, and thrombocytopenia. A population-based study suggested an overall incidence of 0.1 cases/million individuals, using combined Surveillance, Epidemiology, and End Results and National Cancer Database data. CNL typically presents in elderly patients with a median age at diagnosis of 66.5 years (range, 15–86 years) and slight male preponderance (56–58% of cases). Median survival is ~2 years, and causes of death include transformation to acute leukemia, progressive disease associated with severe cytopenias, and marked treatment-refractory leukocytosis.

CSF3 is the main growth factor for granulocyte proliferation and differentiation. Accordingly, recombinant CSF3 is used for the treatment of severe neutropenia, including severe congenital neutropenia (SCN). Some patients with SCN acquire *CSF3R* mutations, and the frequency of such mutations is significantly higher (~80%) in patients who experience leukemic transformation. SCN-associated *CSF3R* mutations occur in the region of the gene coding for the cytoplasmic domain of CSF3R and result in truncation of the C-terminal–negative regulatory domain. In 2013, Maxson et al described a different class of *CSF3R* mutations in ~90% of patients with CNL; these were mostly membrane proximal, the most frequent being a C-to-T substitution at nucleotide 1853 (T618I). In a subsequent confirmatory study, *CSF3R* mutations were found to be specific to WHO-defined CNL. About 40% of the T618I-mutated cases also harbored *SETBP1* mutations. *CSF3R* T618I has been shown to induce lethal myeloproliferative disorder in a mouse model and to have in vitro sensitivity to JAK inhibition.

Diagnosis of CNL requires exclusion of the more common causes of neutrophilia including infections and inflammatory processes **(Table 115-5)**. In addition, one should be mindful of the association between some forms of metastatic cancer or plasma cell neoplasms with secondary neutrophilia. Neoplastic neutrophilia also occurs in other myeloid malignancies, which should be excluded during the

TABLE 115-4 International Consensus Classification of Myeloid Neoplasms

6. Acute myeloid leukemia (AML)
    a. AML diagnosis requiring ≥10% bone marrow (BM) or peripheral blood (PB) blasts
        i. Acute promyelocytic leukemia
        ii. Core binding factor AML
        iii. AML with *KMT2A* rearrangement
        iv. AML with *DEK::NUP214*
        v. AML with *MECOM* rearrangements
        vi. AML with *NPM1* mutation
        vii. AML with in-frame *bZIP CEBPA* mutations
        viii. AML with other rare recurring translocations
        ix. Myelodysplastic syndrome (MDS)/AML with *TP53* mutations
        x. MDS/AML with myelodysplasia-related mutations
        xi. MDS/AML with myelodysplasia-related karyotype
        xii. MDS/AML not otherwise specified (NOS)
    b. AML diagnosis requiring ≥20% BM or PB blasts
        i. AML with t(9;22)-*BCR::ABL1*
        ii. AML with *TP53* mutations, other than pure erythroid leukemia
        iii. AML with myelodysplasia-related gene mutations
        iv. AML with myelodysplasia-related karyotype
        v. AML NOS
7. AML-related disorders
    a. Pure erythroid leukemia (PEL; *TP53* mutated)
    b. Myeloid sarcoma
    c. Blastic plasmacytoid dendritic cell neoplasm
    d. Acute leukemia of ambiguous lineage
    e. Acute undifferentiated leukemia
    f. Mixed phenotype acute leukemia
8. Myelodysplastic syndromes (MDS)
    a. MDS with mutated *TP53*
    b. MDS with excess blasts (5–9% BM or 2–9% PB)
    c. MDS without excess blasts (<5% BM and <2% PB)
        i. MDS with del(5q) [isolated or accompanied by only one other cytogenetic abnormality other than 7/del(7q); no multi-hit *TP53*]
        ii. MDS with *SF3B1* [variant allele frequency ≥10%/no *RUNX1* or multi-hit *TP53*; no del(5q), –7/del(7q), complex karyotype, or abnormal 3q26.2]
        iii. MDS, NOS–single-lineage dysplasia
        iv. MDS, NOS–multilineage dysplasia
        v. MDS, NOS without dysplasia
9. MDS/AML
    a. MDS/AML (BM/PB blasts 10–19%)
    b. MDS/AML with mutated *TP53*
10. Myeloproliferative neoplasms (MPN)
    a. Chronic myeloid leukemia
    b. Polycythemia vera
    c. Essential thrombocythemia
    d. Primary myelofibrosis (PMF)
        i. Early/prefibrotic PMF
        ii. Overt PMF
    e. MPN, unclassifiable (MPN-U)
    f. Chronic neutrophilic leukemia
    g. Chronic eosinophilic leukemia, NOS
11. MDS/MPN
    a. Chronic myelomonocytic leukemia (CMML) (≥0.5 × 10^9/L absolute and ≥10% PB monocytes)
        i. CMML-1 (<10% BM and <5% PB blasts)
        ii. CMML-2 (10–19% BM or 5–19% PB blasts)
    b. Atypical chronic myeloid leukemia
    c. MDS/MPN with mutated *SF3B1* and thrombocytosis
    d. MDS/MPN with ring sideroblasts and thrombocytosis, NOS
    e. MDS/MPN, NOS
        i. MDS/MPN with isolated isochromosome (17q)
12. Eosinophilic disorders
13. Mastocytosis
14. Hematologic neoplasms with germline predisposition
15. Pediatric myeloid malignancies
16. Premalignant clonal hematopoiesis

diagnostic workup (Table 115-5). Accordingly, the ICC diagnostic criteria for CNL are designed to exclude the possibilities of both secondary/reactive neutrophilia and leukocytosis associated with myeloid malignancies other than CNL (Table 115-5). The discovery of *CSF3R* mutations (see above) and their almost invariable association with ICC-defined CNL has allowed its incorporation in the ICC diagnostic criteria (Table 115-5). In general, the presence of a membrane proximal *CSF3R* mutation in a patient with predominantly neutrophilic granulocytosis should be sufficient for the diagnosis of CNL, regardless of the degree of leukocytosis. Unfortunately, several exclusionary criteria still need to be met for diagnosing CNL in the absence of *CSF3R* mutations (Table 115-5).

Current treatment in CNL is largely palliative and suboptimal in its efficacy. Several drugs alone or in combination have been tried, and none have shown remarkable efficacy. As such, allogeneic hematopoietic stem cell transplant (ASCT) is reasonable to consider in the presence of symptomatic disease, especially in younger patients. Otherwise, cytoreductive therapy with hydroxyurea is probably as good as anything, and a more intensive combination chemotherapy may not have additional value. However, response to hydroxyurea therapy is often transient, and some have successfully used interferon α as an alternative drug. JAK inhibitor therapy has emerged as an additional therapeutic option but is not necessarily superior to hydroxyurea (estimated response rate of 30%). It is thus recommended that CNL patients first and foremost be evaluated for eligibility and disposition for ASCT, with the remaining therapeutic agents being aimed at controlling myeloproliferation (targeting leukocytes <25–30 × 10^9/L) and alleviating symptoms.

CHRONIC EOSINOPHILIC LEUKEMIA, NOT OTHERWISE SPECIFIED

In a Mayo Clinic survey of 1416 patients with PB eosinophilia evaluated between 2008 and 2019, 17 patients (1.2%) fulfilled the ICC criteria for CEL-NOS (Table 115-5); median age was 63 years, with the vast majority of patients (88%) presenting with systemic symptoms. Organ involvement was a prominent feature including spleen, cardiac, pulmonary, and distal esophagus. Laboratory abnormalities included anemia, leukocytosis, and eosinophilia (median eosinophil count of 6.4 × 10^9/L; range, 2.0–53.1 × 10^9/L). The most common bone marrow abnormalities included abnormal eosinophils, abnormal and increased megakaryocytes, and fibrosis (18%). Cytogenetic abnormalities occurred in 88% of patients and included trisomy 8, complex karyotype, 13q–, 20q–, and chromosome 1 abnormalities. All seven patients with next-generation sequencing studies harbored one or more mutations including *ASXL1* (43%) and *IDH1* (29%). Half of patients treated with hydroxyurea-based regimens responded with a persistent decline in eosinophil count for a median duration of 18 months. One-third of patients treated with prednisone responded, with a median duration of response at 13 months. Three patients were treated with imatinib, of whom two had normalization of eosinophil count. At a median follow-up of 13 months, nine patients had died, including three who underwent leukemic transformation.

MYELOPROLIFERATIVE NEOPLASM, UNCLASSIFIABLE

The category of MPN-U includes MPN-like neoplasms that cannot be clearly classified as one of the other subcategories of MPN listed in Table 115-1. Examples include patients presenting with unusual thrombosis or unexplained organomegaly with normal blood counts but found to carry MPN-characteristic mutations such as *JAK2* and *CALR* or display bone marrow morphology that is consistent with MPN. It is possible that some cases of MPN-U represent earlier disease stages in polycythemia vera (PV) or essential thrombocythemia (ET), which, however, fail to meet the threshold hemoglobin levels or platelet counts that are required per WHO diagnostic criteria. Specific treatment interventions might not be necessary in asymptomatic patients with MPN-U, whereas patients with arterial thrombotic complications might require cytoreductive and aspirin therapy and those with venous thrombosis might require systemic anticoagulation.

TABLE 115-5 International Consensus Classification (ICC) Diagnostic Criteria for Chronic Neutrophilic Leukemia (CNL), Atypical Chronic Myeloid Leukemia (aCML), and Chronic Myelomonocytic Leukemia (CMML)

VARIABLES	CNL[a]	aCML	CMML
PB leukocyte count	≥13 × 10^9/L[d]	≥13 × 10^9/L	
PB segmented neutrophils/bands	≥80%		
PB neutrophil precursors[b]	<10%	≥10%	
PB blasts	Usually absent[e]	<20%	<20%
PB monocyte count	<10% of leukocytes	No or minimal monocytosis	≥0.5 × 10^9/L[g] Persistent and lasting for at least 3 months
Cytopenia[h]		Yes	Yes
Dysgranulopoiesis		Yes	
PB basophil/eosinophil percentage		<10%	
PB monocyte percentage		<10%	≥10%
BM	Hypercellular ↑ Neutrophils, number and % <5% blasts Normal neutrophilic maturation	Hypercellular ↑ Granulocyte proliferation Granulocytic dysplasia ± erythroid/megakaryocyte Dysplasia <20% blasts	Hypercellular due to myeloproliferation and increased monocytes and lacking diagnostic features of AML, MPN, or other conditions associated with monocytosis
BCR-ABL1	No	No	No
Tyrosine kinase gene fusions[f]	No	No	No
CSF3R T618I or other activating CSF3R mutation or persistent neutrophilia, splenomegaly, no identifiable cause of reactive neutrophilia,[c] if plasma cell neoplasm is present, need demonstration of clonality of myeloid cells by cytogenetic or molecular studies	Yes		
PB and BM blasts/promonocytes		<20%	<20%
Evidence for other MPN: CML, PV, ET, PMF	No	No	No
Evidence for reactive leukocytosis or monocytosis	No	No	No

[a]Diagnosis requires meeting all criteria. [b]Neutrophil precursors include myeloblasts, promyelocytes, myelocytes, and metamyelocytes. [c]Causes of reactive neutrophilia include plasma cell neoplasms, solid tumor, infections, and inflammatory processes. [d]≥25 × 10^9/L in cases lacking *CSF3R* T618I or another activating *CSF3R* mutation. [e]10–19% blasts constitute accelerated phase and ≥20% blast phase. [f]Tyrosine kinase gene fusions involve *PDGFRA*, *PDGFRB*, *FGFR1*, *ABL1*, *JAK2*, and *FLT3*. [g]PB monocytes ≥1 × 10^9/L in cases without evidence of clonality; the latter is signified by abnormal karyotype or a myeloid neoplasm associated mutation with ≥10% variant allele frequency. [h]Hemoglobin <12 g/dL in females and <13 g/dL in males, absolute neutrophil count <1.8 × 10^9/L, and/or platelets <150 × 10^9/L, that is not explained by another condition.

Abbreviations: AML, acute myeloid leukemia; BM, bone marrow; CML, chronic myeloid leukemia; ET, essential thrombocythemia; MDS, myelodysplastic syndromes; MPN, myeloproliferative neoplasms; PB, peripheral blood; PMF, primary myelofibrosis; PV, polycythemia vera.

CHRONIC MYELOMONOCYTIC LEUKEMIA

CMML is classified under the ICC category of MDS/MPN neoplasms and is defined by sustained (>3 months) PB monocytosis (≥0.5 × 10^9/L; monocytes ≥10% of leukocyte count), consistent BM morphology, <20% BM or PB blasts (including promonocytes), and cytogenetic or molecular evidence of clonality. The median age at CMML diagnosis is ~73–75 years, with a male preponderance (1.5–3:1). The exact incidence of CMML remains unknown but is estimated at 4 cases per 100,000 persons per year. Clinical presentation is variable and depends on whether the disease presents with MDS-like (MDS-CMML) or MPN-like (MP-CMML) phenotype, based on the presence or absence of leukocyte count of ≥13 × 10^9/L; the former is associated with cytopenias and the latter with splenomegaly and features of myeloproliferation such as fatigue, night sweats, weight loss, and cachexia. About 20% of patients with CMML experience unique symptoms including systemic inflammatory syndromes (e.g., arthritis, pericardial effusion, pleural effusion, ascites), autoimmune diseases, leukemia cutis, and lysozyme-induced nephropathy.

During the diagnostic workup of CMML, it is important to first exclude reactive causes of monocytosis, including tuberculosis, fungal infections, subacute bacterial endocarditis, viral and protozoal infections, connective tissue diseases, sarcoidosis, lipid storage disorders, postsplenectomy state, and the recovery phase of an acute infection or BM regeneration after chemotherapy. Other myeloid neoplasms in the differential diagnosis include CML (*BCR::ABL1*-defined) and other fusion gene-associated entities including those with rearrangements of *PDGFRA*, *PDGFRB*, *FGFR1*, *JAK2*, *FLT3*, and *ABL1*. Similarly, it should be noted that monocytosis can be associated with MPN such as primary myelofibrosis (PMF) and PV, where its presence adversely impacts survival. BM examination often shows morphologic dysplasia in at least one hematopoietic lineage and granulocytic and monocytic proliferation. On immunophenotyping, the abnormal cells often express myelomonocytic antigens such as CD13 and CD33, with variable expression of CD14, CD64, CD68, and CD163. Monocytic-derived cells are almost always positive for the cytochemical nonspecific esterases (e.g., butyrate esterase), while normal granulocytic precursors are positive for lysozyme and chloroacetate esterase. In CMML, it is common to have a hybrid cytochemical staining pattern with cells expressing both chloroacetate and butyrate esterases simultaneously (dual esterase staining).

Based on flow cytometric expression of CD14/CD16, monocytes can be classified into classical MO1 (CD14+/CD16–), intermediate MO2 (CD14+/CD16+), and nonclassical MO3 (CD14–/CD16+) fractions, with MO1 constituting the major monocyte population (85%) in healthy conditions. CMML patients have a characteristic increase in classical monocytes, distinguishing CMML from other causes of reactive and clonal monocytosis. Almost all patients with CMML harbor somatic mutations that are neither specific nor disease-defining, including (1) mutations in epigenetic control of transcription, such as histone modification (*EZH2*, *ASXL1*, *UTX*), and DNA methylation (*TET2*, *DNMT3A*, *IDH1*, *IDH2*); (2) mutations in the spliceosome machinery (*SF3B1*, *SRSF2*, *U2AF1*, *ZRSR2*, *PRPF8*); (3) mutations in genes that regulate cell signaling (*JAK2*, *KRAS*, *NRAS*, *CBL*, *PTPN11*, *NF1*, *FLT3*); (4) mutations in transcription factors and nucleosome assembly regulators (*RUNX1*, *GATA2, SETBP1*); and (5) mutations in DNA damage response genes such as *TP53* and *PHF6*. Of these, those involving *TET2* (~60%), *SRSF2* (~50%), *ASXL1* (~40%), and the oncogenic *RAS* pathway (~30%) are the most frequent, with only frameshift and nonsense *ASXL1* mutations consistently and independently adversely impacting survival. Clonal cytogenetic abnormalities are seen in about a third of patients with CMML and include trisomy 8 and abnormalities of chromosome 7.

Several risk models serve similar purposes in identifying high-risk patients who are considered for ASCT earlier than later. Risk factors in the Mayo Molecular Model (MMM) include presence of truncating *ASXL1* mutations, absolute monocyte count >10 × 10^9/L, hemoglobin <10 g/dL, platelet count <100 × 10^9/L, and the presence of circulating immature myeloid cells. The resulting four-tiered risk categorization includes high (three or more risk factors), intermediate-2 (two risk factors), intermediate-1 (one risk factor), and low (no risk factors); the corresponding median survivals are 16, 31, 59, and 97 months, respectively. ASCT is the only treatment modality that secures cure or long-term survival and is appropriate for MMM high/intermediate-2 risk disease. In one of the largest retrospective cohorts involving 513 CMML patients (median age, 53 years), the European Group for Blood and Marrow Transplantation reported a 4-year relapse-free survival rate of 27% and an overall survival rate of 33%. At present, for younger patients with higher risk disease and an acceptable comorbidity index, ASCT is the preferred treatment modality. Drug therapy is currently not disease-modifying and includes hydroxyurea and hypomethylating agents; a phase 3 study (DAKOTA) comparing hydroxyurea and decitabine in high-risk MP-CMML showed similar overall survival at 23.1 versus 18.4 months, respectively, despite response rates being higher for decitabine (56 vs 31%).

ATYPICAL CHRONIC MYELOID LEUKEMIA

aCML is formally classified under the MDS/MPN category of myeloid malignancies and is characterized by left-shifted granulocytosis and dysgranulopoiesis. Diagnostic criteria are listed in Table 115-2 and include leukocyte count of ≥13 × 10^9/L, dysgranulopoiesis, cytopenia, ≥10% immature granulocytes, <20% PB or BM myeloblasts, <10% PB monocytes, <10% PB eosinophilia, absence of otherwise specific mutations such as *BCR::ABL1* or TKGFs (involving *PDGFRA, PDGFRB, FGFR1, JAK2, ABL1*, or *FLT3*; **Table 115-6**), and not meeting WHO criteria for CML, PMF, PV, or ET. The BM in aCML is hypercellular with granulocyte proliferation and dysplasia with or without erythroid or megakaryocytic dysplasia. The differential diagnosis of aCML includes CML, which is distinguished by the presence of *BCR::ABL1*; CNL, which is distinguished by the absence of dysgranulopoiesis and presence of *CSF3R* mutations; and CMML, which is distinguished by the presence of monocytosis (absolute monocyte count ≥0.5 × 10^9/L).

The molecular pathogenesis of aCML is incompletely understood; about one-fourth of patients express *SETBP1* mutations, which are, however, also found in several other myeloid malignancies, including CNL and CMML. *SETBP1* mutations in aCML were prognostically detrimental and mostly located between codons 858 and 871; similar mutations are seen with Schinzel-Giedion syndrome (a congenital disease with severe developmental delay and various physical stigmata including midface retraction, large forehead, and macroglossia). A somatic missense mutation in ethanolamine kinase 1 (*ETNK1* N244S) was described in 9% of patients with aCML but was also seen in 14% of patients with CMML, 6% of patients with mastocytosis (especially in association with eosinophilia), and rarely in other MPNs.

In a series of 55 patients with WHO-defined aCML, median age at diagnosis was 62 years, with female preponderance (57%); splenomegaly was reported in 54% of the patients, red cell transfusion requirement in 65%, abnormal karyotype in 20% (20q– and trisomy 8 being the most frequent), and leukemic transformation in 40%. Median survival was 25 months. Outcome was worse in patients with marked leukocytosis, transfusion requirement, and increased immature cells in the PB. In a more recent Mayo Clinic study of 25 molecularly annotated and strictly WHO-defined aCML patients, median age was 70 years and 84% were male. Cytogenetic abnormalities were seen in 36% and gene mutations in 100%. Mutational frequencies were as follows: *ASXL1* 28%, *TET2* 16%, *NRAS* 16%, *SETBP1* 12%, *RUNX1* 12%, *ETNK1* 8%, and *PTPN11* 4%. Median survival was 10.8 months, and at last follow-up (median, 11 months), 17 (68%) deaths and 2 (8%) leukemic transformations were documented. In multivariable analysis, advanced age, low hemoglobin, and *TET2* mutations were shown to carry independent prognostic significance; other mutations, including *ASXL1* and *SETBP1* lacked prognostic significance. Conventional chemotherapy is largely ineffective in the treatment of aCML. Similarly, treatment response to the JAK1/2 inhibitor ruxolitinib has not been impressive. However, a favorable experience with ASCT was reported in nine patients; after a median follow-up of 55 months, the majority of the patients remained in complete remission.

MDS/MPN WITH MUTATED SF3B1 OR WITH RING SIDEROBLASTS, BOTH ASSOCIATED WITH THROMBOCYTOSIS OR NOT OTHERWISE SPECIFIED

The ICC classifies patients with morphologic and laboratory features that resemble both MDS and MPN as "MDS/MPN overlap." This category is broad and is distinguished from MPN by the presence of "cytopenia." Leukocytosis is also part of the definition for the subcategories of MDS/MPN, including MDS/MPN with mutated *SF3B1* (MDS/MPN-T-*SF3B1*), MDS/MPN with ring sideroblasts and thrombocytosis, NOS (MDS/MPN-RS-T-NOS), and MDS/MPN-NOS. Diagnostic criteria for MDS/MPN-T-*SF3B1* include thrombocytosis (≥450 × 10^9/L), anemia, blasts <1% in PB and <5% in BM, presence of *SF3B1*

TABLE 115-6 Primary Eosinophilia Classification

VARIABLES	EOSINOPHILIA ASSOCIATED WITH TKGF (M/LN-EO-TK)	CHRONIC EOSINOPHILIC LEUKEMIA, NOT OTHERWISE SPECIFIED (CEL-NOS)	LYMPHOCYTIC VARIANT HYPEREOSINOPHILIA	HYPEREOSINOPHILIC SYNDROME
Absolute eosinophil count	≥1500 × 10^9/L	≥1500 × 10^9/L	≥1500 × 10^9/L	≥1500 × 10^9/L
PB eosinophil %	≥10%	≥10%	≥10%	≥10%
Documentation of chronicity of eosinophilia	Advised	Advised	Advised	Advised
Comorbidity associated with secondary eosinophilia	Absent	Absent	Absent	Absent
PB blast ≥2% or BM blast ≥5%	Yes or no	Yes or no	No	No
Abnormal karyotype	Yes or no	Yes or no	No	No
TKGF	Yes	No	No	No
BCR-ABL1	No	No	No	No
Abnormal T lymphocyte phenotype or clonal T-cell clones	No	No	Yes	No
Eosinophil-mediated tissue/organ damage	Yes or no	Yes or no	Yes or no	Yes

Abbreviations: ABL1 (e.g., *ETV6::ABL1*); BM, bone marrow; *FGFR1* (e.g., *ZMYM2::FGFR1*); *FLT3* (*ETV6::FLT3*); *JAK2* (*PCM1::JAK2*); M/LN-eo-TK, myeloid/lymphoid neoplasms with eosinophilia and TKGF; PB, peripheral blood; TKGF, tyrosine kinase gene fusions, often involving *PDGFRA* (e.g. *FIP1L1::PDGFRA*) or *PDGFRB* (e.g., *ETV6::PDGFRB*).

(variant allele frequency [VAF] >10%), and not otherwise classified as another myeloid neoplasm; corresponding criteria for MDS/MPN-RS-T-NOS also include thrombocytosis and anemia and absence of excess blasts but, in addition, require presence of ≥15% BM ring sideroblasts and absence of *SF3B1* mutation. The term MDS/MPN-NOS is reserved for MDS/MPN that does not meet criteria for either of the aforementioned MDS/MPN entities despite displaying thrombocytosis (≥450 × 10^9/L) or leukocytosis (≥13 × 10^9/L).

JUVENILE MYELOMONOCYTIC LEUKEMIA

JMML is primarily a disease of early childhood and is now considered a unique clonal disorder of childhood, separated from MDS/MPN. Both CMML and JMML feature leukocytosis, monocytosis, and hepatosplenomegaly. Additional characteristic features in JMML include thrombocytopenia and elevated fetal hemoglobin. Myeloid progenitors in JMML display granulocyte-macrophage colony-stimulating factor (GM-CSF) hypersensitivity that has been attributed to dysregulated RAS/MAPK signaling. The latter is believed to result from mutually exclusive mutations involving *RAS*, *PTPN11*, and *NF1*. A third of patients with JMML that is not associated with Noonan syndrome carry *PTPN11* mutations, while the incidence of *NF1* in patients without neurofibromatosis type 1 and *RAS* mutations is ~15% each. In general, ~85% of JMML cases have one of the classical RAS pathway mutations (*PTPN11*, *NRAS*, *KRAS*, *NF1*, or *CBL*); in addition, a myriad of other mutations, such as *ASXL1*, *RUNX1*, *SETBP1*, *JAK3*, and *CUX1*, among others, have recently been reported. Taken together, it is currently believed that almost all patients with JMML harbor mutations in the RAS pathway; clonal disorders that mimic JMML but do not harbor a RAS pathway mutation are classified as JMML-like neoplasms.

The 2022 ICC diagnostic criteria for JMML require the presence of PB monocyte count ≥1 × 10^9/L, <20% blasts in PB or BM, splenomegaly, and absence of *BCR::ABL1*. Diagnosis also requires the presence of one of the following: (1) somatic mutations of *PTPN11*, *KRAS*, *NRAS*, or *RRAS*; (2) germline *NF1* mutation and loss of heterozygosity of *NF1* or clinical diagnosis of neurofibromatosis type 1; and (3) germline mutation and loss of heterozygosity of *CBL*. Drug therapy is relatively ineffective in JMML, and the treatment of choice is ASCT, which results in a 5-year survival of ~50%.

TRANSIENT MYELOPROLIFERATIVE DISORDER

TMD, also referred to as transient abnormal myelopoiesis (TAM), constitutes an often but not always transient phenomenon of abnormal megakaryoblast proliferation, which occurs in ~10% of infants with Down syndrome. TMD is usually recognized at birth and either undergoes spontaneous regression (75% of cases) or progresses to acute megakaryoblastic leukemia (AMKL; 25% of cases). Almost all patients with TMD and TMD-derived AMKL display somatic *GATA1* mutations. TMD-associated *GATA1* mutations constitute exon 2 insertions, deletions, or missense mutations, affecting the N-terminal transactivation domain of GATA-1 and resulting in loss of full-length (50 kD) GATA-1 and its replacement with a shorter isoform (40 kD) that retains friend of GATA-1 (FOG-1) binding. In contrast, inherited forms of exon 2 *GATA1* mutations produce a phenotype with anemia, whereas exon 4 mutations that affect the N-terminal, FOG-1-interactive domain produce familial dyserythropoietic anemia with thrombocytopenia or X-linked macrothrombocytopenia.

HYPEREOSINOPHILIA

Eosinophilia refers to a PB absolute eosinophil count (AEC) that is above the upper normal limit of the reference range. The term *hypereosinophilia* (HE) is used when the AEC is ≥1500 × 10^9/L. The ICC recommends both BM and PB examination for diagnostic evaluation of HE. The former should include cytogenetic and molecular analysis as well as immunohistochemistry for mast cells (CD117, tryptase, CD25), and the latter should include lymphocyte flow cytometry with T-cell panel, TCR gene rearrangement studies, and serum tryptase. The most frequent causes of HE include infections, especially those related to tissue-invasive helminths, allergic/vasculitic diseases, drugs, and metastatic cancer. Primary HE is the focus of this chapter and is considered when a cause for secondary eosinophilia is not readily apparent.

In the presence of normal BM morphology and absence of genetic abnormalities, the two major diagnostic possibilities are lymphocytic variant HE and idiopathic HE (Table 115-3); the former is characterized by the presence of an abnormal T-cell phenotype or clone. Both conditions might be associated with tissue/organ dysfunction due to eosinophilic infiltrates, in which case their nomenclature is modified into lymphocytic variant hypereosinophilic syndrome and idiopathic hypereosinophilic syndrome (iHES), respectively. In the presence of a TKGF, a diagnosis of myeloid/lymphoid neoplasm with eosinophilia and TKGF is considered (Table 115-3). The genes involved in TKGF-associated HE are listed in Table 115-6: *PDGFRA*, *PDGFRB*, *FGFR1*, *ABL*, *JAK2*, and *FLT3*. Once the latter possibility is excluded, other ICC-defined myeloid or lymphoid neoplasms (e.g., systemic mastocytosis, acute myeloid or lymphoblastic leukemia, CML, MPN or MDS/MPN, Hodgkin's and non-Hodgkin's lymphoma) must be considered and excluded. CEL-NOS is considered in the presence of cytogenetic abnormalities, excess blasts, or morphologic evidence of dysplasia, including that of megakaryocytes.

The diagnostic workup for HE that is not associated with morphologically overt myeloid malignancy should start with PB mutation screening for *PDGFRA* and *PDGFRB* mutations using fluorescence in situ hybridization (FISH) or reverse transcriptase polymerase chain reaction. If mutation screening is negative, a BM examination with cytogenetic and molecular studies is indicated. In this regard, one must first pay attention to the presence or absence of TKGF or associated cytogenetic abnormalities (Table 115-6). CEL-NOS is a subset of clonal eosinophilia that is neither molecularly defined nor classified as an alternative clinicopathologically assigned myeloid malignancy. We prefer to use the term strictly in patients with an HES phenotype who also display either a clonal cytogenetic/molecular abnormality, excess blasts in the BM or PB, or abnormal BM morphology (Table 115-6).

HYPEREOSINOPHILIA ASSOCIATED WITH TYROSINE KINASE GENE FUSIONS

Both platelet-derived growth factor receptors α (*PDGFRA*; located on chromosome 4q12) and β (*PDGFRB*; located on chromosome 5q31-q32) are involved in MPN-relevant activating mutations. Clinical phenotype in both instances includes prominent blood eosinophilia and excellent response to imatinib therapy. In regard to *PDGFRA* mutations, the most popular is *FIP1L1-PDGFRA*, a karyotypically occult del(4)(q12), which was described in 2003 as an imatinib-sensitive activating mutation. Functional studies have demonstrated transforming properties in cell lines and the induction of MPN in mice. Cloning of the *FIP1L1-PDGFRA* fusion gene identified a novel molecular mechanism for generating this constitutively active fusion tyrosine kinase, wherein an ~800 kb interstitial deletion within 4q12 fuses the 5′ portion of *FIP1L1* to the 3′ portion of *PDGFRA*. *FIP1L1-PDGFRA* occurs in a very small subset of patients who present with the phenotypic features of either systemic mastocytosis (SM) or HES, but the presence of the mutation reliably predicts complete hematologic and molecular response to imatinib therapy.

In a retrospective survey of 151 patients with *FIP1L1-PDGFRA*–associated eosinophilia (143 males; mean age at diagnosis, 49 years), organopathy involved the spleen (44%), skin (32%), lungs (30%), heart (19%), and CNS (9%); none of 31 patients initially treated with corticosteroids achieved complete hematologic remission, whereas all 148 patients treated with imatinib achieved complete hematologic responses and also molecular responses, when evaluated. Treatment discontinuation was documented in 46 patients followed by a 57% relapse rate; the 1-, 5-, and 10-year overall survival rates in imatinib-treated patients were 99%, 95%, and 84%, respectively. Other studies have confirmed the possibility of treatment-free remissions in some patients after imatinib discontinuation. Infrequent occurrence of *FIP1L1-PDGFRA*–mutated AML associated with eosinophilia has also been shown to respond to low-dose imatinib therapy (100 mg/d).

The association between eosinophilic myeloid malignancies and *PDGFRB* rearrangement was first characterized and published in 1994 when fusion of the tyrosine kinase encoding region of *PDGFRB* to the *ets*-like gene, *ETV6* (*ETV6-PDGFRB*, t(5;12)(q33;p13), was demonstrated. The fusion protein was transforming to cell lines and resulted in constitutive activation of PDGFRB signaling. Since then, several other *PDGFRB* fusion transcripts with similar disease phenotypes have been described, cell line transformation and MPD induction in mice have been demonstrated, and imatinib therapy was effective when employed.

The 8p11 myeloproliferative syndrome (EMS) (also known as human stem cell leukemic/lymphoma syndrome) constitutes a clinical phenotype with features of both lymphoma and eosinophilic MPN and is characterized by a fusion mutation that involves the gene for fibroblast growth factor receptor 1 (*FGFR1*), which is located on chromosome 8p11 [e.g., *ZMYM2::FGFR1*, t(8;13)(p11.1;q12.1)]; disease phenotypes include T-cell acute lymphoblastic leukemia, large B-cell lymphoma, and MPN-like disease with HE. In EMS, both myeloid and lymphoid lineage cells exhibit the 8p11 translocation, thus demonstrating the stem cell origin of the disease. The disease features several 8p11-linked chromosome translocations, and some of the corresponding fusion *FGFR1* mutants have been shown to transform cell lines and induce EMS- or CML-like disease in mice depending on the specific *FGFR1* partner gene (*ZNF198* or *BCR*, respectively). Consistent with this laboratory observation, some patients with *BCR-FGFR1* mutation manifest a more indolent CML-like disease. The mechanism of *FGFR1* activation in EMS is similar to that seen with *PDGFRB*-associated MPN; the tyrosine kinase domain of *FGFR1* is juxtaposed to a dimerization domain from the partner gene. EMS is an aggressive disease often requiring combination chemotherapy followed by ASCT. Pemigatinib, which targets FGFR1/2/3, has been introduced and shown to induce hematologic and cytogenetic response in >70% of patients with *FGFR1*-rearranged myeloid/lymphoid neoplasms.

The 2022 ICC includes a number of other subcategories of myeloid/lymphoid neoplasms with eosinophilia and TKGFs (M/LN-eo-TK): (1) *ETV6::ABL1*, t(9;12)(q34.1;p13.2), phenotypically similar to CML and treated the same way with imatinib or similar tyrosine kinase inhibitors (TKIs) with good response; (2) *PCM1::JAK2* or *BCR::JAK2* or other *JAK2* partners, t(8;9)(p22;p24.1), phenotypically similar to MPN or MDS/MPN with >90% response rate to ruxolitinib but not durable and requiring bridging to ASCT; and (3) *ETV6::FLT3*, t(12;13)(p13.2;q12.2), phenotypically similar to lymphoblastic leukemia or lymphoma, CEL, or MDS/MPN, with some responses seen with *FLT3* inhibitor therapy.

HYPEREOSINOPHILIC SYNDROME

Blood eosinophilia that is neither secondary nor clonal is operationally labeled as being "idiopathic." HES is a subcategory of idiopathic eosinophilia with persistent increase of the AEC to ≥1.5 × 10^9/L and presence of eosinophil-mediated organ damage, including cardiomyopathy, gastroenteritis, cutaneous lesions, sinusitis, pneumonitis, neuritis, and vasculitis. In addition, some patients manifest thromboembolic complications, hepatosplenomegaly, and either cytopenia or cytosis.

BM histologic and cytogenetic/molecular studies should be examined before a working diagnosis of HES is made. Additional blood studies that are currently recommended during the evaluation of HES include serum tryptase (an increased level suggests mastocytosis and warrants molecular studies to detect *FIP1L1-PDGFRA*), T-cell immunophenotyping, and T-cell receptor antigen gene rearrangement analysis (a positive test suggests an underlying clonal or phenotypically abnormal T-cell disorder). In addition, initial evaluation in HES should include echocardiogram and measurement of serum troponin levels to screen for myocardial involvement by the disease.

Initial evaluation of the patient with eosinophilia should include tests that facilitate assessment of target organ damage, including complete blood count, chest x-ray, echocardiogram, and serum troponin level. Increased level of serum cardiac troponin has been shown to correlate with the presence of cardiomyopathy in HES. Typical echocardiographic findings in HES include ventricular apical thrombus, posterior mitral leaflet or tricuspid valve abnormality, endocardial thickening, dilated left ventricle, and pericardial effusion.

In a Mayo Clinic study of 98 consecutive patients with idiopathic eosinophilia, including HES, median age was 53 years (55% male), and overt organ involvement was seen in >80% of the cases, including 54% involving organs other than the skin. The frequencies of cardiac involvement, hepatosplenomegaly, and increased serum tryptase and interleukin (IL) 5 levels were 8%, 4%, 24%, and 31%, respectively. The study also revealed that 11% of the affected patients harbored pathogenetic mutations including *TET2*, *ASXL1*, and *KIT*; the presence of such mutations did not appear to influence phenotype, and the number of informative cases was too small to assess prognostic relevance. Instead, the study identified anemia and presence of cardiac involvement or hepatosplenomegaly as risk factors for survival.

Glucocorticoids are the cornerstone of therapy in HES. Treatment with oral prednisone is usually started at 1 mg/kg per day and continued for 1–2 weeks before the dose is tapered slowly over the ensuing 2–3 months. If symptoms recur at a prednisone dose level of >10 mg/d, either hydroxyurea or interferon α is used as steroid-sparing agent. In patients in whom usual therapy fails as outlined above, mepolizumab or alemtuzumab might be considered. Mepolizumab is a monoclonal antibody that targets IL-5, which is a well-recognized growth factor for eosinophils. Alemtuzumab targets the CD52 antigen, which has been shown to be expressed by eosinophils but not by neutrophils. In a recently reported placebo-controlled phase 3 study, HES patients received subcutaneous mepolizumab (300 mg) every 4 weeks, in addition to their preprotocol therapy, and experienced significantly fewer disease flare-ups or treatment discontinuations (28 vs 56% for placebo), without excess adverse events. Mepolizumab was U.S. Food and Drug Administration approved for use in HES on September 25, 2020. In a smaller phase 2 study, benralizumab (monoclonal antibody targeting the receptor for IL-5; 30 mg given subcutaneously every 4 weeks) was also shown to reduce eosinophil count more efficiently compared to placebo (90 vs 30%).

MASTOCYTOSIS

SM is characterized by proliferation of neoplastic mast cells (MCs) in BM and/or other extracutaneous organs and is distinguished from cutaneous mastocytosis (CM; skin involvement only) and mast cell sarcoma (MCS; high-grade focal MC tumor). According to the ICC and WHO-proposed fifth edition (WHO5) classification systems (Table 115-4), SM is subclassified into indolent (ISM), smoldering (SSM), aggressive (ASM), SM with associated myeloid (SM-AMN, per ICC) or hematologic (SM-AHN, per WHO5) neoplasm, and mast cell leukemia (MCL). WHO5 also includes an additional "low-grade" SM subtype, namely BM mastocytosis (BMM); the latter is described as consisting of (1) absence of skin lesions, (2) absence of B findings, and (3) serum tryptase level <125 ng/mL. The ICC considers BMM as a clinicopathologic variant and not an SM subtype. Both ICC and WHO5 have also refined their diagnostic criteria for SM in general. Diagnosis per ICC requires the presence of a major criterion (multifocal aggregates of ≥15 MCs) or, in its absence, the presence of at least three minor criteria, including (1) BM biopsy or extracutaneous organ section with >25% MCs with atypical morphology; (2) MC expression of CD25, CD2, and/or CD30; (3) *KIT* D816V or other activating *KIT* mutation; and (4) increased serum tryptase >20 ng/mL (needs to be adjusted in the presence of hereditary α-tryptasemia); in addition, presence of myeloid/lymphoid neoplasm with eosinophilia with TKGRs must be excluded. SM diagnosis per WHO5 requires the presence of the major criterion as well as one other minor criterion or at least three minor criteria.

The term *advanced SM* (AdvSM) includes ASM, SM-AMN/AHN, and MCL; AdvSM is distinguished from non-AdvSM by the presence of either an associated myeloid/hematologic neoplasm (e.g., AML, CMML, MDS, MPN) or organopathy resulting from MC infiltration. MC-associated organopathy is defined by the presence of one or more C findings: (1) ≥1 cytopenia (hemoglobin <10 g/dL, absolute neutrophil count <1 × 10^9/L, or platelet count <100 × 10^9/L); (2) palpable hepatomegaly with abnormal liver function tests, ascites, or portal hypertension; (3) palpable splenomegaly with thrombocytopenia attributed to hypersplenism; (4) MC infiltration of the gastrointestinal system with resultant malabsorption with weight loss; and (5) large osteolytic lesions with or without pathologic fractures.

The ICC system requires the AHN component in SM-AHN to be of myeloid lineage, resulting in a revised nomenclature (i.e., SM-AMN). By contrast, WHO5-defined SM-AHN allows the AHN component to be of either myeloid or lymphoid lineage. Additional divergence between ICC and WHO5 concerns the definition of MCL; both systems require the presence of ≥20% MCs in BM aspirate, but ICC criteria require, in addition, immature cytomorphology (i.e., promastocytes, metachromatic blast-like cells, or multinucleated or highly pleomorphic MC) of the excess MCs; furthermore, ICC no longer differentiates between leukemic (≥10% circulating MC) versus aleukemic MCL variants. Both the ICC and WHO5 use similar B findings to distinguish indolent from smoldering SM, with the latter requiring the presence of two or more B findings, including (1) MCs >30% of BM cellularity on BM biopsy *and* serum tryptase >200 ng/mL; (2) cytopenia not meeting criteria for C findings or cytosis; and (3) palpable hepatomegaly without liver function impairment or splenomegaly without thrombocytopenia or >1 cm lymphadenopathy on palpation or imaging; in addition, *KIT* D816V VAF ≥10% qualifies as a B finding, per WHO5.

In a study of 329 patients with AdvSM, including WHO5 subcategories of SM-AHN (64%), ASM (30%), and MCL (6%) or ICC subcategories of SM-AMN (64%), ASM (33%), and MCL (3%), multivariable analysis that included the Mayo Alliance risk factors for survival in SM (age >60 years, anemia, thrombocytopenia, increased alkaline phosphatase) revealed more accurate survival prediction with the ICC versus WHO5 classification order: (1) survival was significantly worse with ICC-defined MCL versus WHO5-defined MCL with otherwise mature MC cytomorphology; (2) prognostic distinction was confirmed for ICC-defined MCL versus ICC-defined SM-AMN but not for WHO5-defined MCL versus WHO-defined SM-AHN; (3) survival was similar between WHO5-defined MCL with mature cytomorphology versus ICC-defined SM-AMN; and (4) ICC-defined SM-AMN but not WHO-defined SM-AHN with lymphoid lineage was prognostically distinct from ASM. Accordingly, our views on the classification of AdvSM are in line with those of the ICC system. We believe that these details are therapeutically relevant considering the emergent nature of ICC-defined MCL and the fact that SM-AMN, as opposed to SM-AHN, carries a prognostically worse designation that might require therapeutic intervention with ASCT sooner than later.

Currently available drugs for treatment include KIT inhibitors (KITi), which exhibit remarkable activity in reducing MC and mutant *KIT* burden but have not been shown to extend survival. On the other hand, we are impressed by the remarkable activity of currently available KITi on the MC component of SM, with the caveat that such drugs are expensive, have significant side effects, and need long-term use. The first step in treatment decision-making is to identify whether the AMN component takes priority over the SM component of the disease, for treatment purposes. High-grade AMNs, including AML, high-/very-high-risk MDS or CMML, or those with >10% BM blasts, likely take precedence in this regard. Accordingly, in a fit patient with SM-AML or SM-high/very-high-risk MDS, intensive remission induction therapy or hypomethylating agent (HMA) therapy can be considered. If the SM is incidentally diagnosed or is associated with minimal symptoms, we generally limit SM therapy to supportive/symptomatic care, including with antihistamine and/or antileukotriene agents or cromolyn, as well as taking precautions for anaphylaxis prevention/treatment. For SM with significantly high MC or symptom burden, concurrent MC cytoreduction can be considered, particularly for those with relapsed/refractory disease, although there are no clear protocols in this regard.

ASCT has an important role in the treatment of SM-high-grade AMN, although there is no consensus regarding optimal timing, debulking strategy, and so on; overall survival appears to be most favorable for SM-AMN patients compared to other subgroups, and survival was superior in those receiving myeloablative versus reduced-intensity conditioning, thereby indicating the need for effective cytoreduction prior to stem cell transplantation. For SM-AMN patients with a low-grade AMN, such as PV or ET, or low-risk MDS, conventional therapy for the AMN including therapeutic phlebotomies, low-dose aspirin, hydroxyurea or interferon α, or erythropoiesis-stimulating agents is pursued. If the SM disease component does not warrant cytoreduction (e.g., no C findings), supportive/symptomatic care, as described earlier, is pursued, along with intermittent monitoring of disease status. If MC cytoreduction is warranted, monotherapy with a TKI or cladribine can be pursued, depending on individual risk, benefit considerations, treatment availability, cost considerations, and so on.

We find cladribine, midostaurin, and avapritinib to be reasonable drug considerations for AdvSM as well as MCL. While avapritinib has theoretical advantages over midostaurin, including a more potent inhibitory effect on *KIT* D816V and proven efficacy in patients previously treated with midostaurin, we are agnostic regarding the choice of TKI given the lack of head-to-head comparison; instead, we recommend an individualized approach to TKI selection that weighs drug accessibility/affordability, comorbid conditions, risk tolerance for anticipated adverse events, and provider/center experience with the particular TKI. We continue to utilize cladribine as a reasonable alternative to TKI therapy, based on decades-long experience, long-term safety record, and qualitatively better toxicity profile that is mostly limited to cytopenias, as opposed to cognitive impairment and other side effects with avapritinib and intolerance due to diarrhea with midostaurin. For ASM patients who exhibit disease progression and/or leukemic transformation despite adequate TKI dosing or cladribine, we recommend ASCT as salvage therapy. We consider "true" MCL an oncologic emergency, and collaboration with an experienced hematopathologist to expeditiously confirm the diagnosis is critical, as is expedited molecular testing for *KIT* D816V and other AMN-relevant mutations.

■ DENDRITIC AND HISTIOCYTIC NEOPLASMS

Dendritic cell (DC) and histiocyte/macrophage neoplasms are extremely rare. DCs are antigen-presenting cells, whereas histiocyte/macrophages are antigen-processing cells. BM myeloid stem cells (CD34+) give rise to monocyte (CD14+, CD68+, CD11c+, CD1a−) and DC (CD14−, CD11c+/−, CD1a+/c) precursors. Monocyte precursors, in turn, give rise to macrophages (CD14+, CD68+, CD11c+, CD163+, lysozyme+) and interstitial DCs (CD68+, CD1a−). DC precursors give rise to Langerhans cell DCs (Birbeck granules, CD1a+, S100+, langerin+) and plasmacytoid DCs (CD68+, CD123+). Follicular DCs (CD21+, CD23+, CD35+) originate from mesenchymal stem cells. Dendritic and histiocytic neoplasms are operationally classified into macrophage/histiocyte-related and DC-related. The former includes histiocytic sarcoma/malignant histiocytosis and the latter Langerhans cell histiocytosis, Langerhans cell sarcoma, interdigitating DC sarcoma, and follicular DC sarcoma.

Histiocytic Sarcoma/Malignant Histiocytosis Histiocytic sarcoma represents malignant proliferation of mature tissue histiocytes and is often localized. Median age at diagnosis is estimated at 46 years with slight male predilection. Some patients might have a history of lymphoma, MDS, or germ cell tumors at time of disease presentation. The three typical disease sites are lymph nodes, skin, and the gastrointestinal system. Patients may or may not have systemic symptoms including fever and weight loss, and other symptoms include hepatosplenomegaly, lytic bone lesions, and pancytopenia. Immunophenotype includes presence of histiocytic markers (CD68, lysozyme, CD11c, CD14) and absence of myeloid or lymphoid markers. Prognosis is poor, and treatment is often ineffective. The term *malignant histiocytosis* refers to a disseminated disease and systemic symptoms. Lymphoma-like treatment induces complete remissions in some patients, and median survival is estimated at 2 years. In one of the largest series of histiocytic sarcoma, 330 cases were included with median age of 61 years (59% male). In the latter study, the most common sites of presentation were skin, connective tissue, lymph nodes, gastrointestinal tract, and hematopoietic system; median overall survival was 6 months, and treatment included systemic chemotherapy, radiotherapy, and surgery. Factors associated with poor outcome included older age, high comorbidity index, and disease involving the hematopoietic and reticuloendothelial system.

Langerhans Cell Histiocytosis Langerhans cells (LCs) are specialized DCs that reside in mucocutaneous tissue and upon activation become specialized for antigen presentation to T cells. LC histiocytosis (LCH; also known as histiocytosis X) represents neoplastic proliferation of LCs (S100+, CD1a+, and Birbeck granules on electron microscopy). LCH incidence is estimated at 5 per million, and the disease typically affects children, with a male predilection. Presentation can be either unifocal (eosinophilic granuloma) or multifocal. The former usually affects bones and less frequently lymph nodes, skin, and lung, while the latter is more disseminated. Unifocal disease often affects older children and adults, while multisystem disease affects infants. LCH of the lung in adults is characterized by bilateral nodules. Prognosis depends on organs involved. Only 10% of patients progress from unifocal to multiorgan disease. LCH of the lung might improve upon cessation of smoking. Approximately 55% of patients with LCH harbor *BRAF* V600E gain-of-function mutations, which indicates high-risk disease and resistance to first-line therapy, while responses to targeted therapy with vemurafenib have been reported. Other forms of treatment for LCH include combination chemotherapy and MEK inhibitors in *BRAF* wild-type but with other MAPK pathway mutations. Unfortunately, such targeted therapy has not secured long-lasting treatment-free remissions.

In one retrospective study, 33 adult patients with LCH were studied including 21 with single-system LCH, 10 with multisystem LCH, and 2 with pulmonary LCH. Patients with single-system unifocal involvement were successfully treated with local therapies such as surgery and radiotherapy. Most of the multisystem LCH patients and patients with single-system multifocal involvement were treated with systemic chemotherapy. Cladribine was the first choice in 10 of 11 patients who received chemotherapy. Among all patients, the overall response rate (ORR) was 97%. Among those who had cladribine in the first line, the ORR was 81%. All these patients achieved a complete remission and were alive at the last visit. The median follow-up was 38 months (range, 2–183 months). The median progression-free survival (PFS) has not yet been reached. Ten-year PFS was 90.9%. Expert consensus recommendations for treatment include local therapies for unifocal disease, smoking cessation as first-line therapy for pulmonary LCH, and systemic therapy for multifocal and multisystem disease; the latter might include cladribine, cytarabine, and targeted therapy with BRAF and MEK inhibitors.

In one study, 26 adult patients with non-LCH, including 17 with Erdheim-Chester disease (ECD), 3 with Rosai-Dorfman disease (RDD), 5 with ECD/RDD, and 1 with ECD/LCH, were treated with the MEK inhibitor trametinib; the most common treatment-related toxicity was rash (27%), whereas the response rate of the 17 evaluable patients was 71%, including 73% without a detectable *BRAF* V600E; median time-to-treatment failure was 37 months; most patients harbored mutations in *BRAF* (either classic *BRAF* V600E or other *BRAF* alterations) or alterations in other genes involved in the MAPK pathway (e.g., *MAP2K*, *NF1*, *GNAS*, or *RAS*).

Langerhans Cell Sarcoma Langerhans cell sarcoma (LCS) also represents neoplastic proliferation of LCs with overtly malignant morphology. The disease can present de novo or progress from antecedent LCH. There is a female predilection, and median age at diagnosis is estimated at 41 years. Immunophenotype is similar to that seen in LCH, and liver, spleen, lung, and bone are the usual sites of disease. Prognosis is poor, and treatment is generally ineffective.

Interdigitating Dendritic Cell Sarcoma Interdigitating DC sarcoma (IDCS), also known as reticulum cell sarcoma, represents neoplastic proliferation of interdigitating DCs. The disease is extremely rare and affects elderly adults with no sex predilection. Typical presentation is asymptomatic solitary lymphadenopathy. Immunophenotype includes S100+ and negative for vimentin and CD1a. Prognosis ranges from benign local disease to widespread lethal disease.

Follicular Dendritic Cell Sarcoma Follicular dendritic cells (FDCs) reside in B-cell follicles and present antigen to B cells. FDC neoplasms (FDCNs) are usually localized and often affect adults. FDCN might be associated with Castleman's disease in 10–20% of cases, and increased incidence in schizophrenia has been reported. Cervical lymph nodes are the most frequent site of involvement in FDCNs, and other sites include maxillary, mediastinal, and retroperitoneal lymph nodes; oral cavity; the gastrointestinal system; skin; and breast. Sites of metastasis include lung and liver. Immunophenotype includes CD21, CD35, and CD23. Clinical course is typically indolent, and treatment includes surgical excision followed by regional radiotherapy and sometimes systemic chemotherapy.

Hemophagocytic Lymphohistiocytosis (see Chap. 68) Hemophagocytic lymphohistiocytosis (HLH) represents nonneoplastic proliferation and activation of macrophages and cytotoxic lymphocytes that induce cytokine-mediated bone marrow suppression, features of intense phagocytosis in bone marrow and liver, and multiorgan dysfunction including cytopenias, coagulopathy, and fever. HLH may result from genetic (primary) or acquired (secondary) disorders of macrophages. The former entail genetically determined inability to regulate macrophage proliferation and activation and might include alterations in familial HLH genes, including those of perforin (*PRF1*, *UNC13D*, *STXBP2*, and *STX11*), granule/pigment abnormality genes (*RAB27A*, *LYST*, and *AP3B1*), or X-linked lymphoproliferative disease genes (*SH2D1A* and *XIAP*). Acquired HLH is often precipitated by viral infections, including Epstein-Barr virus. HLH might also accompany certain malignancies such as T-cell lymphoma and autoimmune diseases, ASCT, and chimeric antigen receptor (CAR) T-cell therapy. In a recent population-based study from Sweden, the annual incidence of malignancy-associated HLH had increased 10-fold and was at least 0.71 per 100,000 adults from 2012 to 2018, and early survival improved significantly, likely due to increased awareness and more HLH-directed therapy. Regardless of the cause, the common tissue/organ-damaging pathway involves excessive inflammatory cytokine release, including IL-6, IL-2, IL-1, interferon γ, and tumor necrosis factor (TNF).

Clinical presentation of HLH includes fever, severe constitutional symptoms, enlarged lymph nodes, hepatosplenomegaly, neurologic dysfunction, and abnormalities in multiple organ function tests. Diagnosis is accomplished by either detection of HLH-related mutations or meeting five of the following eight conventional criteria: (1) hemophagocytosis in the bone marrow/spleen/lymph nodes; (2) serum ferritin ≥500 μg/L; (3) hypofibrinogenemia (fibrinogen ≤1.5 g/L) or hypertriglyceridemia (triglycerides ≥3 mmol/L); (4) low NK cell activity; (5) elevated soluble IL-2 receptor (CD25) ≥2400 U/mL; (6) bi- or tri-cytopenia (platelets <100 × 10^9/L, hemoglobin <9 g/dL, absolute neutrophil count <1 × 10^9/L); (7) splenomegaly palpable >3 cm below left costal margin; and (8) fever. Clinical course is often fulminant and fatal with reported 1-year survival rates of <30% in patients with hematologic malignancy. Current therapeutic approaches for primary or secondary HLH include the so-called HLH-94 protocol, which consists of weekly treatments with etoposide and dexamethasone, stem cell transplant, emapalumab (a monoclonal antibody that binds and neutralizes interferon γ and is approved in primary HLH), and the JAK1/2 inhibitor ruxolitinib. The latter has recently been shown to increase survival rate in affected patients to >80%.

FURTHER READING

Alaggio R et al: The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid neoplasms. Leukemia 36:1720, 2022.

Arber DA et al: International Consensus Classification of myeloid neoplasms and acute leukemias: Integrating morphologic, clinical, and genomic data. Blood 140:1200, 2022.

de Leval L et al: A practical approach to the modern diagnosis and classification of T- and NK-cell lymphomas. Blood 144:1855, 2024.

Khoury JD et al: The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and histiocytic/dendritic neoplasms. Leukemia 36:1703, 2022.

Miranda RN et al: The 5th edition of the World Health Organization Classification of Hematopoietic and lymphoid tissues: Mature T-cell, NK-cell, and stroma-derived neoplasms of lymphoid tissues. Mod Pathol 37:100512, 2024.

Szuber N et al: Chronic neutrophilic leukemia: 2022 update on diagnosis, genomic landscape, prognosis, and management. Am J Hematol 97:491, 2022.

116 Plasma Cell Disorders

Nikhil C. Munshi, Dan L. Longo, Kenneth C. Anderson

The *plasma cell disorders* are monoclonal neoplasms related to each other by virtue of their development from common progenitors in the late B-lymphocyte lineage. Multiple myeloma (MM), Waldenström's macroglobulinemia, primary amyloidosis **(Chap. 117)**, and the heavy chain diseases comprise this group and may be designated by a variety of synonyms such as *monoclonal gammopathies, paraproteinemias, plasma cell dyscrasias*, and *dysproteinemias*. Mature B lymphocytes destined to produce IgG bear surface immunoglobulin molecules of both μ and γ heavy chain isotypes with both isotypes having identical idiotypes (variable regions). Under normal circumstances, maturation to antibody-secreting plasma cells and their proliferation is stimulated by exposure to the antigen for which the surface immunoglobulin is specific; however, in the plasma cell disorders, the control over this process is lost. The clinical manifestations of all the plasma cell disorders relate to the expansion of the neoplastic cells, to the secretion of cell products (immunoglobulin molecules or subunits, lymphokines), and to some extent to the host's response to the tumor. Normal development of B lymphocytes is discussed in **Chap. 360** and depicted in **Fig. 113-2**.

Three categories of structural variation are present among immunoglobulin molecules that form antigenic determinants, and these are used to classify immunoglobulins. *Isotypes* are those determinants that distinguish among the main classes of antibodies of a given species and are the same in all normal individuals of that species. Therefore, isotypic determinants are, by definition, recognized by antibodies from a distinct species (heterologous sera) but not by antibodies from the same species (homologous sera). There are five heavy chain isotypes (M, G, A, D, E) and two light chain isotypes (κ, λ). *Allotypes* are distinct determinants that reflect regular small differences between individuals of the same species in the amino acid sequences of otherwise similar immunoglobulins. These differences are determined by allelic genes; by definition, they are detected by antibodies made in the same species. *Idiotypes* are the third category of antigenic determinants. They are unique to the molecules produced by a given clone of antibody-producing cells. Idiotypes are formed by the unique structure of the antigen-binding portion of the molecule.

Antibody molecules **(Fig. 116-1)** are composed of two heavy chains (~50,000 molecular weight [mol wt]) and two light chains (~25,000 mol wt). Each chain has a constant portion (limited amino acid sequence variability) and a variable region (extensive sequence variability). The light and heavy chains are linked by disulfide bonds and are aligned so that their variable regions are adjacent to one another. This variable region forms the antigen recognition site of the antibody molecule; its unique structural features form idiotypes that are reliable markers for a particular clone of cells because each antibody is formed and secreted by a single clone. Because of the mechanics of the gene rearrangements necessary to specify the immunoglobulin variable regions (VDJ joining for the heavy chain, VJ joining for the light chain), a particular clone rearranges only one of the two chromosomes to produce an immunoglobulin molecule of only one light chain isotype and only one allotype (allelic exclusion) (Fig. 116-1). After exposure to antigen, the variable region may become associated with a new heavy chain isotype (class switch). Each clone of cells performs these sequential gene arrangements in a unique way. This results in each clone producing a unique immunoglobulin molecule. In most plasma cells, light chains are synthesized in slight excess, secreted as free light chains, and cleared by the kidney, but <10 mg of such light chains are excreted per day.

Electrophoretic analysis permits separation of components of the serum proteins **(Fig. 116-2)**. The immunoglobulins move heterogeneously in an electric field and form a broad peak in the gamma region. There is a sharp spike in this region called an *M component* (M for monoclonal) in the sera of patients with plasma cell tumors. Less commonly, the M component may appear in the β_2 or α_2 globulin region. The monoclonal antibody must be present at a concentration of at least 5 g/L (0.5 g/dL) to be accurately quantitated by this method. This corresponds to ~10^9 antibody producing cells. Confirmation of the type of immunoglobulin and that it is truly monoclonal is determined by immunoelectrophoresis that reveals a single heavy and/or light chain type. Hence, immunoelectrophoresis and electrophoresis provide qualitative and quantitative assessment of the M component, respectively. The amount of M component in the serum is a reliable measure of the tumor burden and an excellent tumor marker to manage therapy, yet it is not specific enough to be used to screen asymptomatic patients. In addition to the plasma cell disorders, M components may be detected in other lymphoid neoplasms such as chronic lymphocytic leukemia (CLL) and lymphomas of B- or T-cell origin; nonlymphoid neoplasms such as chronic myeloid leukemia, breast cancer, and colon cancer; a variety of nonneoplastic conditions such as cirrhosis, sarcoidosis, parasitic diseases, Gaucher's disease, and pyoderma gangrenosum; and a number of autoimmune conditions, including rheumatoid arthritis, myasthenia gravis, and cold agglutinin disease. Monoclonal proteins are also observed in immunosuppressed patients after organ transplant and, rarely, allogeneic transplant. At least two very rare skin diseases—lichen myxedematosus (also known as papular mucinosis) and necrobiotic xanthogranuloma—are associated with a monoclonal gammopathy. In papular mucinosis, highly cationic IgG is deposited in the dermis of patients. This organ specificity may reflect the specificity of the antibody for some antigenic component of the dermis. Necrobiotic xanthogranuloma is a histiocytic infiltration of the skin, usually of the face, that produces red or yellow nodules that can enlarge to plaques. Approximately 10% progress to myeloma. Five percent of patients with sensory motor neuropathy also have a monoclonal paraprotein.

The nature of the M component is variable in plasma cell disorders. It may be an intact antibody molecule of any heavy chain subclass, or it may be an altered antibody or fragment. Isolated light or heavy chains may be produced. In some plasma cell tumors such as extramedullary or solitary bone plasmacytomas, less than one-third of patients will have an M component. In ~20% of myelomas, only light chains are produced and, in most cases, are secreted in the urine as Bence Jones proteins. The frequency of myelomas of a particular heavy chain class is roughly proportional to the serum concentration, and therefore, IgG myelomas are more common than IgA and IgD myelomas. In ~1% of patients with myeloma, biclonal or triclonal gammopathy is observed.

MULTIPLE MYELOMA

DEFINITION

MM represents a malignant proliferation of plasma cells derived from a single clone. The tumor, its products, and the host response to it result in a number of organ dysfunctions and symptoms, including bone pain or fracture, renal failure, susceptibility to infection, anemia, hypercalcemia, and occasionally clotting abnormalities, neurologic symptoms, and manifestations of hyperviscosity.

ETIOLOGY

The cause of myeloma is not known. Myeloma occurred with increased frequency in those exposed to the radiation of nuclear warheads in World War II after a 20-year latency. Myeloma has been seen more commonly than expected among farmers, wood workers, leather workers, and those exposed to petroleum products. A variety of recurrent chromosomal alterations have been found in patients with myeloma: hyperdiploidy (trisomies involving one or more of chromosomes 3, 5, 7, 9, 11, 15, 19, or 21) is observed in half of the patients, while the other half have translocations involving the 14q32 chromosome with variable partners including t(11;14)(q13;q32), t(4;14)(p16;q32), and t(14;16). Other frequent abnormalities include 13q14 deletion, 1q amplification or 1p deletion, and 17p13 deletions. Evidence is strong that errors in switch recombination—the genetic mechanism to change antibody heavy chain isotype—participate in the early transformation process. However, no single common molecular pathogenetic pathway has yet

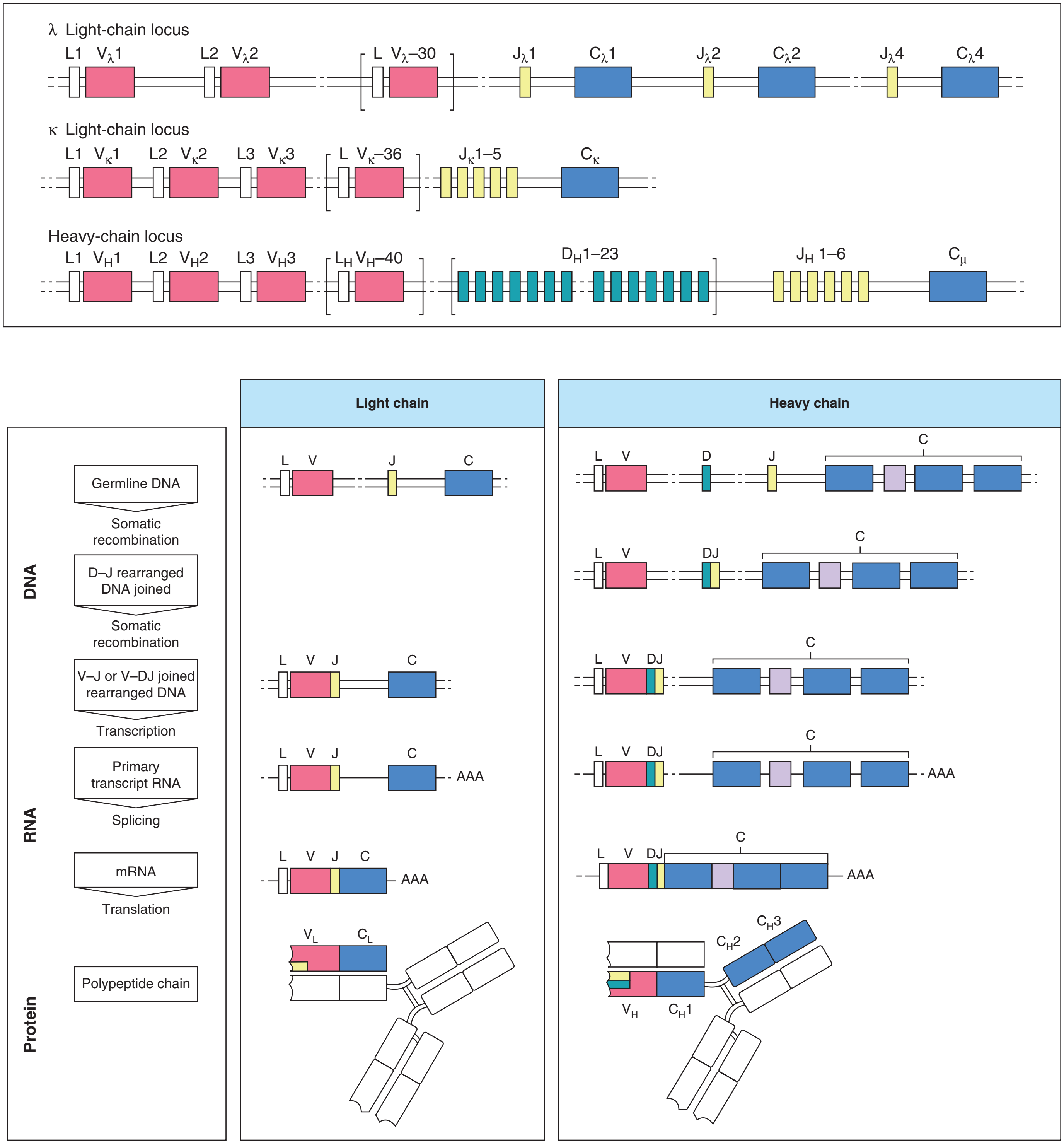
λ Light-chain locus
L1 Vλ1
L2 Vλ2
L Vλ–30
Jλ1
Cλ1
Jλ2
Cλ2
Jλ4
Cλ4
κ Light-chain locus
L1 Vκ1
L2 Vκ2
L3 Vκ3
L Vκ–36
Jκ1–5
Cκ
Heavy-chain locus
L1 VH1
L2 VH2
L3 VH3
LH VH–40
DH1–23
JH 1–6
Cμ
Light chain
Heavy chain
DNA
RNA
Protein
Germline DNA
Somatic recombination
D–J rearranged DNA joined
Somatic recombination
V–J or V–DJ joined rearranged DNA
Transcription
Primary transcript RNA
Splicing
mRNA
Translation
Polypeptide chain
L
V
D
J
DJ
C
AAA
VL
CL
VH
CH1
CH2
CH3

FIGURE 116-1 Immunoglobulin genetics and the relationship of gene segments to the antibody protein. The *top* portion of the figure is a schematic of the organization of the immunoglobulin genes, λ on chromosome 22, κ on chromosome 2, and the heavy chain locus on chromosome 14. The heavy chain locus is >2 megabases, and some of the D region gene segments are only a few bases long, so the figure depicts the schematic relationship among the segments, not their actual size. The *bottom* portion of the figure outlines the steps in going from the noncontiguous germline gene segments to an intact antibody molecule. Two recombination events juxtapose the V-D-J (or V-J for light chains) segments. The rearranged gene is transcribed, and RNA splicing cuts out intervening sequences to produce an mRNA, which is then translated into an antibody light or heavy chain. The sites on the antibody that bind to antigen (the so-called CDR3 regions) are encoded by D and J segments for heavy chains and the J segments for light chains. *(Adapted from Janeway's Immunobiology, 8th ed by Kenneth Murphy. Copyright © 2012 by Garland Science, Taylor & Francis Group, LLC. Used by permission of W. W. Norton & Company, Inc.)*

emerged. Genome sequencing efforts have allowed for characterization of critical genes, pathways, and clonal heterogeneity in myeloma. The median number of mutations per transcribed genome in myeloma is ~58, and within the whole genome, it is >7000. A very heterogeneous mutational landscape with no unifying mutation has been observed. The most frequently mutated genes are *KRAS* and *NRAS* (~20% each), followed by *TP53, DIS3, FAM46C*, and *BRAF*, all mutated in 5–10% of patients. All other mutations were observed in <5% of the patients. These results are now being applied to develop new targeted personalized therapies in myeloma. Evidence of complex clusters of subclonal variants is present at diagnosis, and additional mutations are acquired over time, indicative of genomic evolution that may drive disease progression. Interleukin (IL) 6 may play a role in driving myeloma cell proliferation. It remains difficult to distinguish benign from malignant

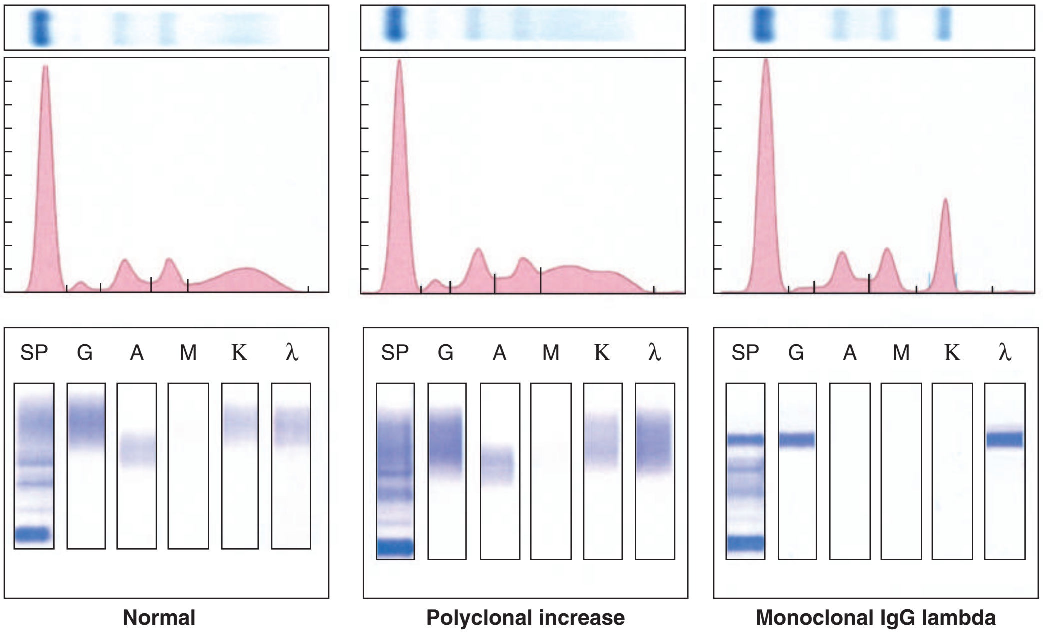

FIGURE 116-2 Representative patterns of serum electrophoresis and immunofixation. The *upper panels* represent agarose gel, middle panels are the densitometric tracing of the gel, and *lower panels* are immunofixation patterns. The panel on the *left* illustrates the normal pattern of serum protein on electrophoresis. Because there are many different immunoglobulins in the serum, their differing mobilities in an electric field produce a broad peak. In conditions associated with increases in polyclonal immunoglobulin, the broad peak is more prominent (*middle panel*). In monoclonal gammopathies, the predominance of a product of a single cell produces a "church spire" sharp peak, usually in the γ globulin region (*right panel*). The immunofixation (*lower panel*) identifies the type of immunoglobulin. For example, normal and polyclonal increases in immunoglobulins produce no distinct bands; however, the *right panel* shows distinct bands in IgG and lambda protein lanes, confirming the presence of IgG lambda monoclonal protein. *(Courtesy of Dr. Neal I. Lindeman.)*

plasma cells based on morphologic criteria in all but a few cases **(Fig. 116-3)**.

INCIDENCE AND PREVALENCE

In 2024 in the United States, 35,780 new cases of myeloma were estimated to be diagnosed, and 12,540 people were estimated to die from the disease. Myeloma increases in incidence with age. The median age at diagnosis is 69 years; it is uncommon under age 40. Males are more commonly affected than females, and blacks have nearly twice the incidence of whites. In 2021, myeloma accounted for 1.8% of all malignancies, with incidence rates per 100,000 of 8.1 and 5.1 in white and 17.1 and 13.0 in black men and women, respectively.

GLOBAL CONSIDERATIONS

The incidence of myeloma is highest in blacks and Pacific Islanders; intermediate in Europeans and North American whites; and lowest in people from developing countries including Asia. The higher incidence in more developed countries may result from the combination of a longer life expectancy and more frequent medical surveillance. Incidence of MM in other ethnic groups including native Hawaiians, female Hispanics, American Indians from New Mexico, and Alaskan natives is higher relative to U.S. whites in the same geographic area. Chinese and Japanese populations have a lower incidence than whites. Immunoproliferative small-intestinal disease (IPSID) with α heavy chain disease is most prevalent in the Mediterranean area. Despite these differences in prevalence, the characteristics, response to therapy, and prognosis of myeloma are similar worldwide.

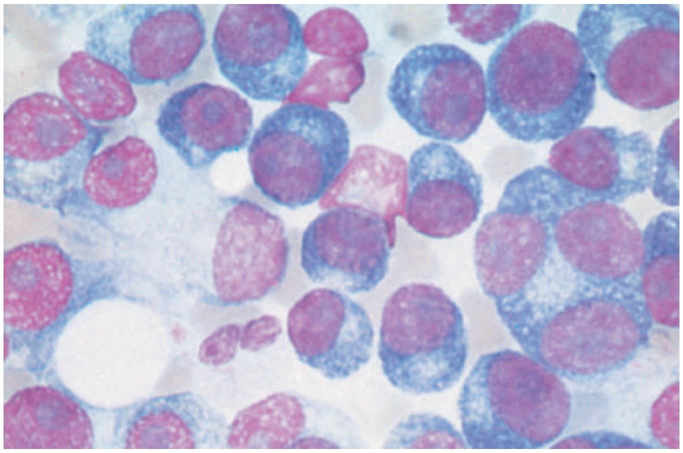

FIGURE 116-3 Multiple myeloma (marrow). The cells bear characteristic morphologic features of plasma cells: round or oval cells with an eccentric nucleus composed of coarsely clumped chromatin, a densely basophilic cytoplasm, and a perinuclear clear zone containing the Golgi apparatus. Binucleate and multinucleate malignant plasma cells can be seen.

PATHOGENESIS AND CLINICAL MANIFESTATIONS

MM cells bind via cell-surface adhesion molecules to bone marrow stromal cells (BMSCs) and extracellular matrix (ECM), which triggers MM cell growth, survival, drug resistance, and migration in the bone marrow milieu **(Fig. 116-4)**. These effects are due both to direct MM cell–BMSC binding via adhesion molecules and to induction of various cytokines, including IL-6, insulin-like growth factor type 1 (IGF-1), and vascular endothelial growth factor (VEGF). Growth, drug resistance, and migration are mediated via Ras/Raf/mitogen-activated protein kinase, PI3K/Akt, and protein kinase C signaling cascades, respectively. Other cellular elements in the bone marrow microenvironment also significantly impact MM cell growth and survival, especially interactions with endothelial cells and osteoclasts. Immune cells such as plasmacytoid dendritic cells (pDC), myeloid-derived suppressor cells (MDSC), and T helper 17 (T_H17) cells are increased in number and support myeloma growth, while antimyeloma immune responses, especially T helper and cytotoxic cells, B cells, and natural killer T cells, are suppressed.

Bone pain is the most common symptom in myeloma, affecting nearly 70% of patients. The bone lesions of myeloma are caused by the proliferation of tumor cells, activation of osteoclasts that destroy bone, and suppression of osteoblasts that form new bone. The increased osteoclast activity is mediated by osteoclast activating factors (OAFs) produced by the myeloma cells (mediated by several cytokines, including IL-1, lymphotoxin, VEGF, receptor activator of nuclear factor-κB [RANK] ligand, macrophage inhibitory factor [MIP]-1α, and tumor

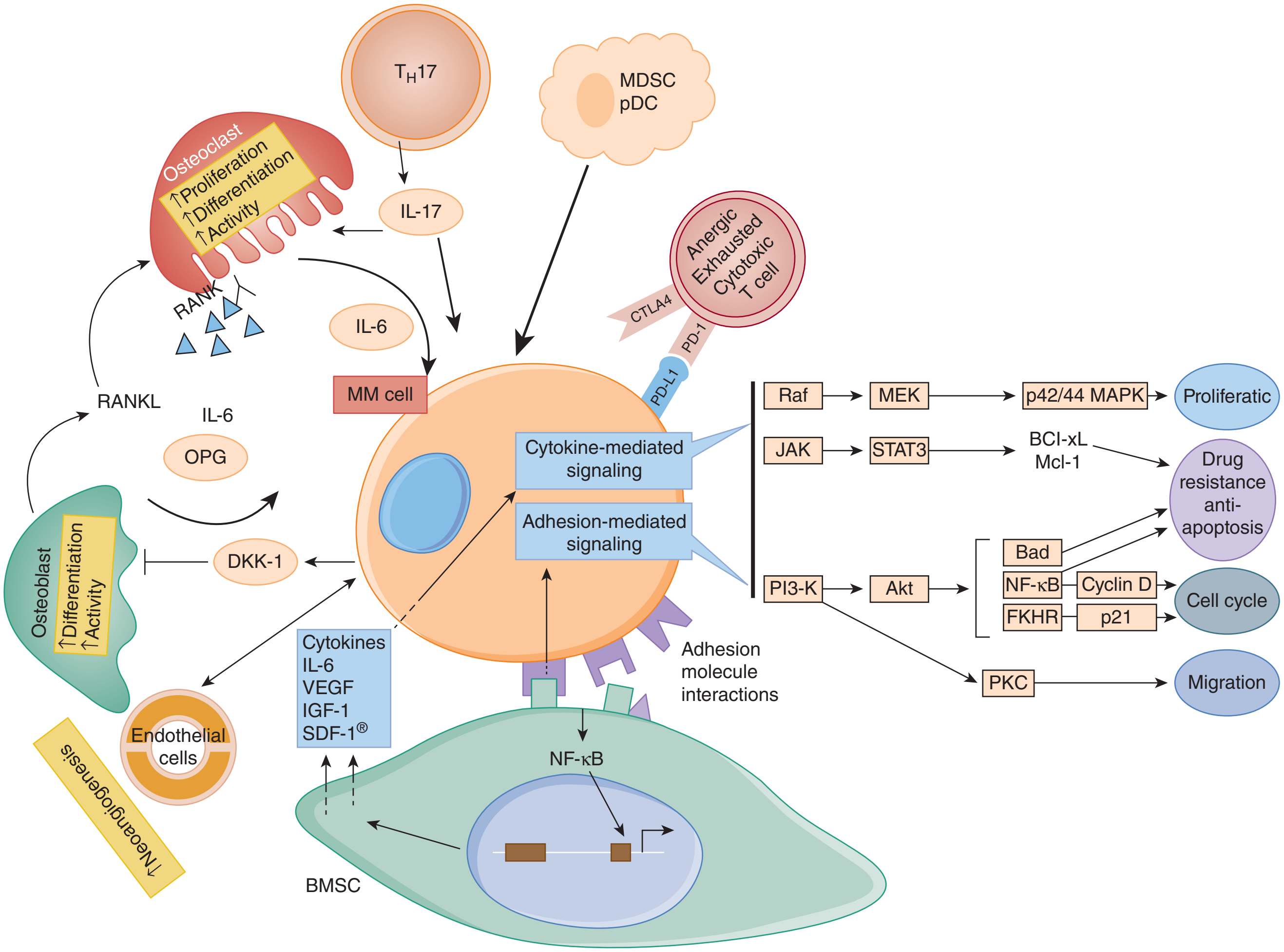

FIGURE 116-4 Pathogenesis of multiple myeloma. Multiple myeloma (MM) cells interact with bone marrow stromal cells (BMSCs) and extracellular matrix proteins via adhesion molecules, triggering adhesion-mediated signaling as well as cytokine production. This triggers cytokine-mediated signaling that provides growth, survival, and antiapoptotic effects as well as development of drug resistance. Additional bidirectional interactions lead to inhibition of osteoblast and increase in osteoclast activity, which leads to bone-related issues in myeloma. Similar interactions with immune microenvironment lead to augmentation of tumor-promoting immune responses and suppression of tumor protective immune responses, overall allowing myeloma cell growth. *(Adapted from G Bianchi, NC Munshi: Blood 125: 3049, 2015.)*

necrosis factor [TNF]). The bone lesions are lytic in nature **(Fig. 116-5)** and are rarely associated with osteoblastic new bone formation due to their suppression by dickkopf-1 (DKK-1) produced by myeloma cells. Therefore, radioisotopic bone scan is less useful in diagnosis than is plain radiography. The bony lysis results in substantial mobilization of calcium from bone, and serious acute and chronic complications of hypercalcemia may dominate the clinical picture (see below). Localized bone lesions may cause pathological fracture or the collapse of vertebrae, leading to spinal cord compression. The next most common clinical problem in patients with myeloma is susceptibility to bacterial infections. The most common infections are pneumonias and pyelonephritis, and the most frequent pathogens are *Streptococcus pneumoniae*, *Staphylococcus aureus*, and *Klebsiella pneumoniae* in the lungs and *Escherichia coli* and other gram-negative organisms in the urinary tract. In ~25% of patients, recurrent infections are the presenting features, and >75% of patients will have a serious infection at some time in their course. The susceptibility to infection has several contributing causes. First, patients with myeloma have diffuse hypogammaglobulinemia if the M component is excluded. The hypogammaglobulinemia is related to both decreased production and increased destruction of normal antibodies. The large M component results in fractional catabolic rates of 8–16% instead of the normal 2%. Moreover, some patients generate a population of circulating regulatory cells in response to their myeloma that can suppress normal antibody synthesis. These patients have very poor antibody responses, especially to polysaccharide antigens such as those on bacterial cell walls. Various abnormalities in T-cell function are also observed including decreased T_H1 response, increase in T_H17 cells producing proinflammatory cytokines, and aberrant T regulatory cell function. Granulocyte lysozyme content is low, and granulocyte migration is not as rapid as normal in patients with myeloma, probably the result of a tumor product. There are also a variety of abnormalities in complement functions in myeloma patients. All these factors contribute to the immune deficiency in these patients. Some commonly used therapeutic agents may significantly affect immune function; e.g., dexamethasone suppresses immune responses and increases susceptibility to bacterial and fungal infection, B-cell maturation antigen (BCMA)–targeting chimeric antigen receptor T (CAR-T) cells and bispecific antibodies can eliminate plasma cells inducing hypogammaglobulinemia, and bortezomib predisposes to herpesvirus reactivation.

Renal failure occurs in nearly 25% of myeloma patients, and some renal pathology is noted in >50%. Of many contributing factors, hypercalcemia is the most common cause of renal failure. Glomerular deposits of amyloid, hyperuricemia, recurrent infections, frequent use of nonsteroidal anti-inflammatory agents for pain control, use of iodinated contrast dye for imaging, bisphosphonate use, and occasional infiltration of the kidney by myeloma cells all may contribute to renal dysfunction. However, tubular damage associated with the excretion of light chains is common. Normally, light chains are filtered, reabsorbed in the tubules, and catabolized. With the increase in the amount of light chains presented to the tubule, the tubular cells become overloaded with these proteins, and tubular damage results either directly from light chain toxic effects or indirectly from the release of intracellular lysosomal enzymes. The earliest manifestation of this tubular damage is the adult Fanconi's syndrome (a type 2 proximal renal tubular acidosis), with loss of glucose and amino acids, as well as defects in the ability of the kidney to acidify and concentrate the urine.

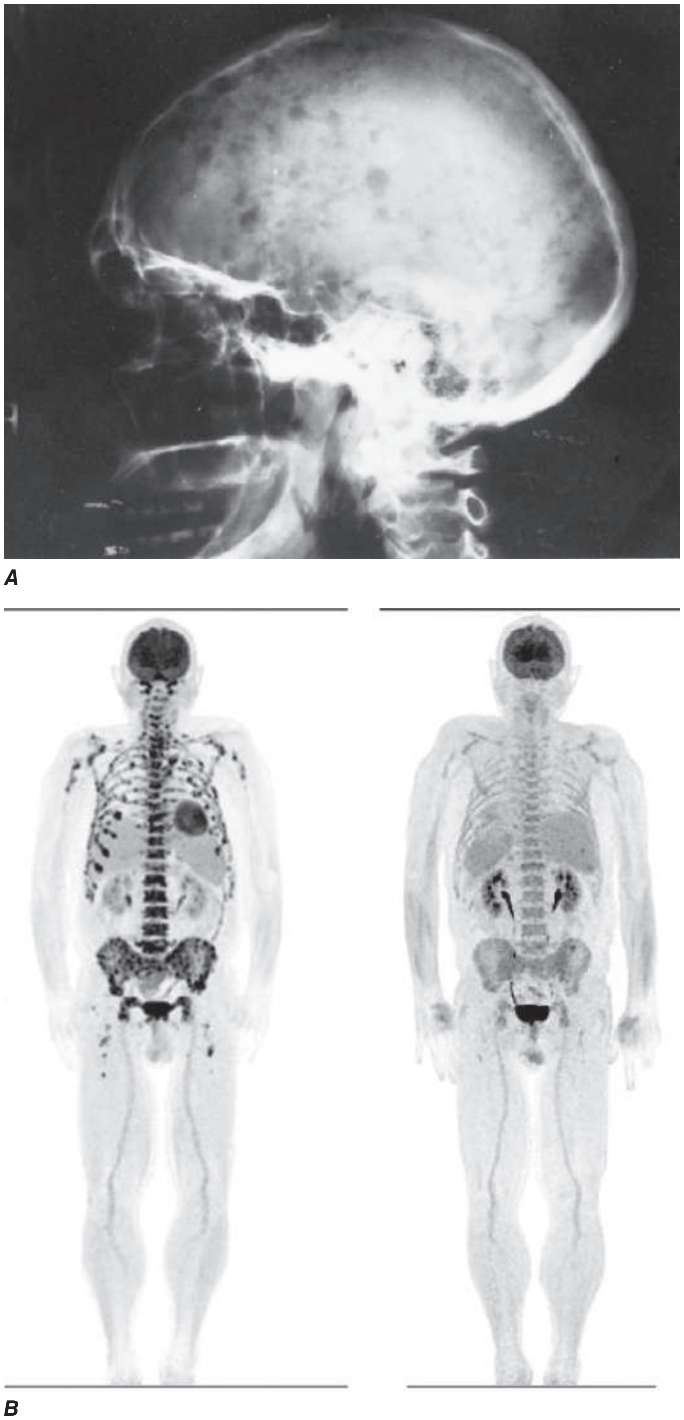

FIGURE 116-5 Bony lesions in multiple myeloma (MM). ***A.*** The skull demonstrates the typical "punched out" lesions characteristic of MM. The lesion represents a purely osteolytic lesion with little or no osteoblastic activity (*above*). ***B.*** Positron emission tomography/computed tomography showing multiple fluorodeoxyglucose (FDG)-avid lesions in skeleton (*left panel*) with their resolution on achieving complete response (CR) (*right panel*). *(Part A courtesy of Dr. Geraldine Schechter; with permission. Part B courtesy of Dr. Sundar Jagannath; with permission.)*

The proteinuria is not accompanied by hypertension, and the protein is nearly all light chains. Generally, very little albumin is in the urine because glomerular function is usually normal. When the glomeruli are involved, nonselective proteinuria is also observed. Patients with myeloma also have a decreased anion gap [i.e., $Na^+ - (Cl^- + HCO_3^-)$] because the M component is cationic, resulting in retention of chloride. This is often accompanied by hyponatremia that is felt to be artificial (pseudohyponatremia) because each volume of serum has less water as a result of the increased protein. Renal dysfunction due to light chain deposition disease, light chain cast nephropathy, and amyloidosis is partially reversible with effective therapy. Myeloma patients are susceptible to developing acute renal failure if they become dehydrated.

Normocytic and normochromic anemia occurs in ~80% of myeloma patients. It is usually related to the replacement of normal marrow by expanding tumor cells, to the inhibition of hematopoiesis by factors produced by the tumor, to reduced production of erythropoietin by the kidney, and to the effects of long-term therapy. In addition, mild hemolysis may contribute to the anemia. A larger than expected fraction of patients may have megaloblastic anemia due to either folate or vitamin B_{12} deficiency. Granulocytopenia and thrombocytopenia are rare except when therapy-induced. Clotting abnormalities may be seen due to the failure of antibody-coated platelets to function properly; the interaction of the M component with clotting factors I, II, V, VII, or VIII; antibody to clotting factors; or amyloid damage of endothelium. Deep venous thrombosis is also observed with the use of thalidomide, lenalidomide, or pomalidomide. Raynaud's phenomenon and impaired circulation may result if the M component forms cryoglobulins, and hyperviscosity syndromes may develop depending on the physical properties of the M component (most common with IgM, IgG3, and IgA paraproteins). Hyperviscosity is defined based on the relative viscosity of serum as compared with water. Normal relative serum viscosity is 1.8 (i.e., serum is normally almost twice as viscous as water). Symptoms of hyperviscosity occur at a level greater than 4 centipoises (cP), which is usually reached at paraprotein concentrations of ~40 g/L (4 g/dL) for IgM, 50 g/L (5 g/dL) for IgG3, and 70 g/L (7 g/dL) for IgA; however, chemico-physical properties of the paraproteins may cause it at lower levels.

Although neurologic symptoms occur in a minority of patients, they may have many causes. Hypercalcemia may produce lethargy, weakness, depression, and confusion. Hyperviscosity may lead to headache, fatigue, shortness of breath, exacerbation or precipitation of heart failure, visual disturbances, ataxia, vertigo, retinopathy, somnolence, and coma. Bony damage and collapse may lead to cord compression, radicular pain, and loss of bowel and bladder control. Infiltration of peripheral nerves by amyloid can be a cause of carpal tunnel syndrome and other sensorimotor mono- and polyneuropathies. Neuropathy associated with monoclonal gammopathy of undetermined significance (MGUS) and myeloma is more frequently sensory than motor neuropathy and is associated with IgM more than other isotypes. In >50% of patients with neuropathy, the IgM monoclonal protein is directed against myelin-associated globulin (MAG). Sensory neuropathy is also a side effect of therapy, specifically thalidomide and bortezomib.

Many of the clinical features of myeloma, e.g., cord compression, pathologic fractures, hyperviscosity, sepsis, and hypercalcemia, can present as medical emergencies. Despite the widespread distribution of plasma cells in the body, tumor expansion is dominantly within bone and bone marrow and, for reasons unknown, rarely causes enlargement of spleen, lymph nodes, or gut-associated lymphatic tissue.

DIAGNOSIS AND STAGING

The diagnosis of myeloma requires marrow plasmacytosis (>10%), a serum and/or urine M component, and at least one of the myeloma-defining events detailed in **Table 116-1**. Bone marrow plasma cells are CD138+ and either monoclonal kappa or lambda light chain positive. The most important differential diagnosis in patients with myeloma involves their separation from individuals with MGUS or smoldering multiple myeloma (SMM). MGUS is vastly more common than myeloma, occurring in 1% of the population aged >50 years and in up to 10% of individuals aged >75 years. The diagnostic criteria for MGUS, SMM, and myeloma are described in Table 116-1. Although ~1% of patients per year with MGUS go on to develop myeloma, all cases of myeloma are preceded by MGUS. Non-IgG subtype, abnormal kappa/lambda free light chain ratio, and serum M protein >15 g/L (1.5 g/dL) are associated with higher incidence of progression of MGUS to myeloma. Absence of all three features predicts a 5% chance of progression, whereas higher-risk MGUS with the presence of all three features predicts a 60% chance of progression over 20 years. The features

TABLE 116-1 Diagnostic Criteria for Multiple Myeloma, Myeloma Variants, and Monoclonal Gammopathy of Undetermined Significance

Monoclonal Gammopathy of Undetermined Significance (MGUS)

Serum monoclonal protein (non-IgM type) <30 g/L

Clonal bone marrow plasma cells <10%[a]

Absence of myeloma-defining events or amyloidosis that can be attributed to the plasma cell proliferative disorder

Smoldering Multiple Myeloma (Asymptomatic Myeloma)

Both criteria must be met:

- Serum monoclonal protein (IgG or IgA) ≥30 g/L or urinary monoclonal protein ≥500 mg per 24 h and/or clonal bone marrow plasma cells 10–60%
- Absence of myeloma-defining events or amyloidosis

Symptomatic Multiple Myeloma

Clonal bone marrow plasma cells or biopsy-proven bony or extramedullary plasmacytoma[a] and any one or more of the following myeloma-defining events:

- Evidence of one or more indicators of end-organ damage that can be attributed to the underlying plasma cell proliferative disorder, specifically:
  - Hypercalcemia: serum calcium >0.25 mmol/L (>1 mg/dL) higher than the upper limit of normal or >2.75 mmol/L (>11 mg/dL)
  - Renal insufficiency: creatinine clearance <40 mL/min[b] or serum creatinine >177 μmol/L (>2 mg/dL)
  - Anemia: hemoglobin value of >20 g/L below the lower limit of normal, or a hemoglobin value <100 g/L
  - Bone lesions: one or more osteolytic lesions on skeletal radiography, CT, or PET-CT[c]
  - Any one or more of the following biomarkers of malignancy:
    - Clonal bone marrow plasma cell percentage[a] ≥60%
    - Involved: uninvolved serum free light chain ratio[d] ≥100
    - >1 focal lesion on MRI studies[e]

Nonsecretory Myeloma

No M protein in serum and/or urine with immunofixation[f]

Bone marrow clonal plasmacytosis ≥10% or plasmacytoma[a]

Myeloma-related organ or tissue impairment (end-organ damage, as described above)

Solitary Plasmacytoma

Biopsy-proven solitary lesion of bone or soft tissue with evidence of clonal plasma cells

Normal bone marrow with no evidence of clonal plasma cells[a]

Normal skeletal survey and MRI (or CT) of spine and pelvis (except for the primary solitary lesion)

Absence of end-organ damage such as hypercalcemia, renal insufficiency, anemia, or bone lesions (CRAB) that can be attributed to a lymphoplasma cell proliferative disorder

POEMS Syndrome

All of the following four criteria must be met:

1. Polyneuropathy
2. Monoclonal plasma cell proliferative disorder
3. Any one of the following: (a) sclerotic bone lesions; (b) Castleman's disease; (c) elevated levels of vascular endothelial growth factor (VEGF)
4. Any one of the following: (a) organomegaly (splenomegaly, hepatomegaly, or lymphadenopathy); (b) extravascular volume overload (edema, pleural effusion, or ascites); (c) endocrinopathy (adrenal, thyroid, pituitary, gonadal, parathyroid, and pancreatic); (d) skin changes (hyperpigmentation, hypertrichosis, glomeruloid hemangiomata, plethora, acrocyanosis, flushing, and white nails); (e) papilledema; (f) thrombocytosis/polycythemia[g]

[a]Clonality should be established by showing κ/λ light chain restriction on flow cytometry, immunohistochemistry, or immunofluorescence. Bone marrow plasma cell percentage should preferably be estimated from a core biopsy specimen; in case of a disparity between the aspirate and core biopsy, the highest value should be used. [b]Measured or estimated by validated equations. [c]If bone marrow has <10% clonal plasma cells, more than one bone lesion is required to distinguish from solitary plasmacytoma with minimal marrow involvement. [d]These values are based on the serum Freelite assay (The Binding Site Group, Birmingham, United Kingdom). The involved free light chain must be ≥100 mg/L. [e]Each focal lesion must be ≥5 mm in size. [f]A small M component may sometimes be present. [g]These features should have no other attributable causes and have temporal relation with each other.

Abbreviations: CT, computed tomography; MRI, magnetic resonance imaging; PET-CT, ^{18}F-fluorodeoxyglucose positron emission tomography with computed tomography; POEMS, polyneuropathy, organomegaly, endocrinopathy, M-protein, and skin changes.

responsible for higher risk of progression from SMM to MM are bone marrow plasmacytosis >10%, abnormal kappa/lambda free light chain ratio, and serum M protein >30 g/L (3 g/dL). Patients with only one of these three features have a 25% chance of progression to MM in 5 years, whereas patients with high-risk SMM with all three features have a 76% chance of progression. Two important variants of myeloma are solitary bone plasmacytoma and solitary extramedullary plasmacytoma. These lesions are associated with an M component in <30% of the cases, they may affect younger individuals, and both are associated with median survivals of ≥10 years. Solitary bone plasmacytoma is a single lytic bone lesion without marrow plasmacytosis. Extramedullary plasmacytomas usually involve the submucosal lymphoid tissue of the nasopharynx or paranasal sinuses without marrow plasmacytosis. Both tumors are highly responsive to local radiation therapy. If an M component is present, it should disappear after treatment. Solitary bone plasmacytomas may recur in other bony sites or evolve into myeloma. Extramedullary plasmacytomas rarely recur or progress.

LABORATORY INVESTIGATION

Serum protein electrophoresis and measurement of serum immunoglobulins and free light chains are useful for detecting and characterizing M spikes, supplemented by immunoelectrophoresis, which is especially sensitive for identifying low concentrations of M components not detectable by protein electrophoresis. A 24-h urine specimen is necessary to quantitate Bence Jones protein (immunoglobulin light chain) excretion. The serum M component is IgG in 53% of patients, IgA in 25%, and IgD in 1%; 20% of patients will have only light chains in serum and urine. Dipsticks for detecting proteinuria are not reliable at identifying light chains, and the heat test for detecting Bence Jones protein is falsely negative in ~50% of patients with light chain myeloma. Fewer than 1% of patients have no identifiable M component; these patients usually have light chain myeloma in which renal catabolism has made the light chains undetectable in the urine. In most of these patients, light chains can now be detected by serum free light chain assay. Mass spectrometry has been investigated to accurately assess M protein. IgD myeloma may also present with light chain disease. About two-thirds of patients with serum M components also have urinary light chains. The light chain isotype may have an impact on disease behavior. Whether this is due to some genetically important determinant of cell proliferation or because lambda light chains are more likely to cause renal damage and form amyloid than are kappa light chains is unclear. About half of patients with IgM paraproteins develop hyperviscosity compared with only 2–4% of patients with IgA and IgG M components. Among IgG myelomas, it is the IgG3 subclass that has the highest tendency to form both concentration- and temperature-dependent aggregates, leading to hyperviscosity and cold agglutination at lower serum concentrations. A standard workup directed at detecting monoclonal plasma cells and myeloma-defining events as well as prognosis is detailed in **Table 116-2**.

A complete blood count with differential may reveal anemia. Erythrocyte sedimentation rate is elevated. Rare patients (~1%) may have plasma cell leukemia with >2000 plasma cells/μL. This may be seen in disproportionate frequency in IgD (12%) and IgE (25%) myelomas. Serum calcium, urea nitrogen, creatinine, and uric acid levels may be elevated. Serum alkaline phosphatase is usually normal even with extensive bone involvement because of the absence of osteoblastic activity. It is also important to quantitate serum β_2-microglobulin and albumin (see below).

Chest and bone radiographs may reveal lytic lesions or diffuse osteopenia. Magnetic resonance imaging (MRI) offers a sensitive means to document extent of bone marrow infiltration and cord or root compression in patients with pain syndromes. ^{18}F-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography (PET)/computed tomography (CT) is a valuable tool to assess bone damage and detect extramedullary sites of the disease (Fig. 116-5). The use of ^{18}F-FDG PET/CT is now considered standard to distinguish between smoldering and active MM and to confirm a suspected diagnosis of solitary plasmacytoma. It is also a valuable tool to evaluate response in patients with oligo- or nonsecretory myeloma.

TABLE 116-2 Standard Investigative Workup in Multiple Myeloma (MM)

Investigations to Evaluate for Clonal Plasma Cells
Bone marrow aspirate and biopsy (fine-needle aspiration of plasmacytoma if indicated) • Histology • Clonality by kappa/lambda immunostaining by flow cytometry or immunohistochemistry
Investigations to Evaluate Clonal Paraprotein
Serum protein electrophoresis and immunofixation
Quantitative serum immunoglobulin levels (IgG, IgA, and IgM)
24-h urine protein electrophoresis and immunofixation
Serum free light chain and ratio
Immunofixation for IgD or IgE in select cases
Investigation to Evaluate End-Organ Damage
Hemogram to assess for anemia.
Chemistry panel for renal function and calcium
Skeletal survey or PET/CT scan to evaluate bone lesions
PET/CT or MRI if smoldering MM or solitary plasmacytoma
Investigation for Risk Stratification
β_2-Microglobulin and serum albumin for ISS stage
DNA-sequencing or if not available Fluorescent in situ hybridization for hyperdipoidy, del17p, t(4;14); t(11;14), t(14;16), t(14;20), amp1q34, and del1p and p53 mutation on bone marrow sample
LDH
Specialized Investigation in Selected Cases
Abdominal fat pad for amyloid
Serum viscosity if IgM component or high IgA levels or serum M component >7 g/dL
Myd88 and *CXCR4* mutation analysis if IgM component

Abbreviations: CT, computed tomography; ISS, International Staging System; LDH, lactate dehydrogenase; MRI, magnetic resonance imaging; PET/CT, positron emission tomography/computed tomography.

PROGNOSIS

Serum β_2-microglobulin is the single most powerful predictor of survival and can substitute for staging. β_2-Microglobulin is the light chain of the class I major histocompatibility antigens (HLA-A, -B, -C) on the surface of every cell. Combination of serum β_2-microglobulin and albumin levels forms the basis for a three-stage International Staging System (ISS) **(Table 116-3)** that predicts survival. With the use of high-dose therapy and the newer agents, the Durie-Salmon staging system is unable to predict outcome and is no longer used. High labeling index, circulating plasma cells, performance status, and high levels of lactate dehydrogenase are also associated with poor prognosis.

With low proliferative activity of plasma cells, karyotype is now not recommended in myeloma. Fluorescent in situ hybridization (FISH) has become a standard investigative technique, and the current recommendation suggests switching to DNA sequencing-based methods. Chromosome 17p deletion with 20% clonality or *TP53* mutation, and translocations t(4;14), (14;16), or t(14;20) associated with del 1p or amp1q changes along with high β2-microglobulin with normal renal function are considered prognostic for high-risk disease, as described in Table 116-3. Chromosome 13q deletion and t(11;14) are not considered predictors of high-risk myeloma.

TREATMENT

Multiple Myeloma

MGUS, SMM, AND SOLITARY PLASMACYTOMA

No specific intervention is indicated for patients with MGUS. Follow-up once a year or less frequently is adequate except in higher-risk MGUS, where serum protein electrophoresis, complete blood count, creatinine, and calcium should be repeated every 6 months. A patient with MGUS and severe polyneuropathy is considered for therapeutic intervention if a causal relationship can be assumed, especially in the absence of any other potential causes for neuropathy. Therapy can include plasmapheresis and occasionally rituximab in patients with IgM MGUS or myeloma-like therapy

TABLE 116-3 Risk Stratification in Myeloma

METHOD	STANDARD RISK (75–80%) (EXPECTED SURVIVAL 8–10+ YEARS)	HIGH RISK (20–25%) (EXPECTED SURVIVAL 3–4 YEARS)
DNA sequencing or FISH if sequencing not available	t(11;14)	del(17p) (20% cutoff) and/or TP53 mut
	del(13)	t(4;14)/t(14;16)/t(14;20) with 1q gain and/or 1p del
	Hyperdiploidy	Del 1p + 1q gain or biallelic del 1p
Biochemical assessment		β_2M >5.5 with normal creatinine
INTERNATIONAL STAGING SYSTEM (ISS)		
INTERNATIONAL STAGING SYSTEM (ISS)	**STAGE**	**MEDIAN SURVIVAL, MONTHS**
β_2M <3.5, ALB ≥3.5	I (28%)	62
β_2M <3.5, ALB <3.5 ***or*** β_2M <3.5, ALB ≥3.5 *or* β_2M = 3.5–5.5	II (39%)	44
β_2M >5.5 *or* β_2M <3.5, ALB <3.5 *or* β_2M = 3.5–5.5	III (33%)	29

[a] Percentage of patients presenting at each stage.

Abbreviations: β_2M, serum β_2-microglobulin in mg/L; ALB, serum albumin in g/dL; FISH, fluorescent in situ hybridization; LDH, lactate dehydrogenase.

in those with IgG or IgA disease. A subset of patients with MGUS develop renal dysfunction usually based on renal damage from the monoclonal antibody. The damage may affect the glomeruli, tubules, or vessels. No consensus exists on management, but lowering the level of the monoclonal antibody with bortezomib has had some advocates.

About 10% of patients have SMM and will have an indolent course demonstrating only slow progression of disease over many years. For patients with SMM, no specific therapeutic intervention is indicated in general. But early intervention with lenalidomide alone or in combination with dexamethasone has been shown to prevent progression from high-risk SMM to active MM. At present, patients with SMM only require antitumor therapy when myeloma-defining events are identified.

Patients with solitary bone plasmacytomas and extramedullary plasmacytomas may be expected to enjoy prolonged disease-free survival after local radiation therapy at a dose of ~40 Gy. Occult marrow involvement may occur at low incidence in patients with solitary bone plasmacytoma. Such patients are usually identified because their serum M component falls slowly or disappears initially after local therapy, only to return after a few months. These patients respond well to systemic therapy.

SYMPTOMATIC MM

Patients with symptomatic myeloma require therapeutic intervention. In general, such therapy has two purposes: (1) systemic therapy to control myeloma; and (2) supportive care to control symptoms of the disease, its complications, and adverse effects of therapy. Therapy can significantly prolong survival and improve the quality of life for myeloma patients.

The therapy of myeloma includes an initial induction regimen followed by consolidation and/or maintenance therapy and, on subsequent progression, management of relapsed disease. All agents available for use at various stages of the therapy and their doses, schedules, and combinations are detailed in **Table 116-4**. Therapy is partly dictated by the patient's age and comorbidities, which may

TABLE 116-4 Standard Therapeutic Agents in Myeloma

CLASS	AGENT	STANDARD DOSAGE AND ADMINISTRATION	COMBINATION	MYELOMA INDICATION
Immunomodulatory drugs (IMiD)	Thalidomide (T)	Oral 50–200 mg qd	TD, VTD	Newly diagnosed and relapsed
	Lenalidomide (R)	Oral 5–25 mg daily × 21 days q 4 weeks	RD, RVD, DaRD, ERD, KRD, IRD, DaRVD	Newly diagnosed, maintenance, and relapsed
	Pomalidomide (P)	Oral 2–4 mg daily × 21 days q 4 weeks	PD, KPD, DaPD	Relapsed
Proteasome inhibitors (PI)	Bortezomib (V)	IV or SC 1.3 mg/m^2 days 1, 4, 8, 11 OR days 1, 8, 15	VD, VTD, VRD, DaVD, VCD DaRVD	Newly diagnosed and relapsed
	Carfilzomib (K)	IV 20–56 mg/m^2 days 1, 2, 8, 9, 15, 16 q 4 weeks	KD, KRD, KPD, Da KD, Da KRD, IsaKD	Newly diagnosed and relapsed
	Ixazomib (I)	Oral 4 mg days 1, 8, 15	IRD	Relapsed, maintenance
Antibodies	Daratumumab (Da)	IV or SC 16 mg/kg per week for 8 weeks then every 2 weeks for 16 weeks and then every 4 weeks thereafter	Dara, DaRD, DaVD, DaPD, DaKD	Newly diagnosed, maintenance, and relapsed
	Elotuzumab (E)	IV 10 mg/kg days 1, 8, 15, and 22 for first two cycles, then on days 1 and 15; along with RD	ERD, EPD	Relapsed
	Isatuximab (Isa)	IV 10 mg/kg weekly for 4 weeks and then every 2 weeks	IsaPD, IsaKD	Relapsed
	Belantamab mafodotin	IV 2.5 mg/kg once every 3 weeks		Relapsed or refractory—4 prior lines of therapy
Selective inhibitor of nuclear export (SINE)	Selinexor (S)	Oral 80 mg on days 1 and 3 of each week	SVD	Relapsed
Histone deacetylase inhibitor	Panobinostat (Pa)	Oral 20 mg once every other day for 3 doses/week for 2 weeks every 21 days	PaVD	Relapsed
Alkylating agents	Melphalan (M)	Oral 0.25 mg/kg per day for 4 days (with P) every 4–6 weeks	MP, MPT, MPR, MPV, DaMPV, high-dose M	Newly diagnosed and relapsed conditioning
	Cyclophosphamide	IV—300–500 mg/m^2 weekly × 2 q 4 weeks Oral—50 mg qd × 21 days	VCD	Newly diagnosed and relapsed
	Bendamustine (B)	IV 70–90 mg days 1, 2 OR days 1, 8 q 4 weeks	BD or BVD	Relapsed
	Melflufen (Me)	IV 40 mg day 1 (with D 40 mg on days 1, 8, 15, and 22) q 28 days	MeD	Relapsed or refractory—4 prior lines of therapy
Glucocorticoid	Dexamethasone (D) Prednisone (P)	Oral 10–40 mg q week Oral 1 mg/kg		All stages
Cellular therapy	Idecabtagene vicleucel (Ide-cel)	IV 450 × 10^6 cells	None	At least two prior lines of therapy that includes an PI, IMiD and anti-CD38 antibody
	Ciltacabtagene autoleucel (Cilta-cel)	IV	None	At least one prior line of therapy, including PI and IMiD and who are refractory to lenalidomide
Bispecific antibodies	Teclistamab Anti-BCMA–anti-CD3	Step-up doses of 0.06 mg and 0.3 mg per kilogram SC days 1, 4 and 1.5 mg per kilogram of body weight day 8 and q week	None	Relapsed or refractory—4 prior lines of therapy with prior exposure to PI, IMiD, and anti-CD38 antibody
	Elranatamab Anti-BCMA–anti-CD3	Step-up doses of 12 and 32 mg on days 1, 4 and 76 mg SC day 8 and then q week for 25 weeks and then q 2 weeks	None	
	Talquetamab Anti-GPRC5D–anti-CD3	Two step-up doses during first week and then every week or every other week regimen	None	

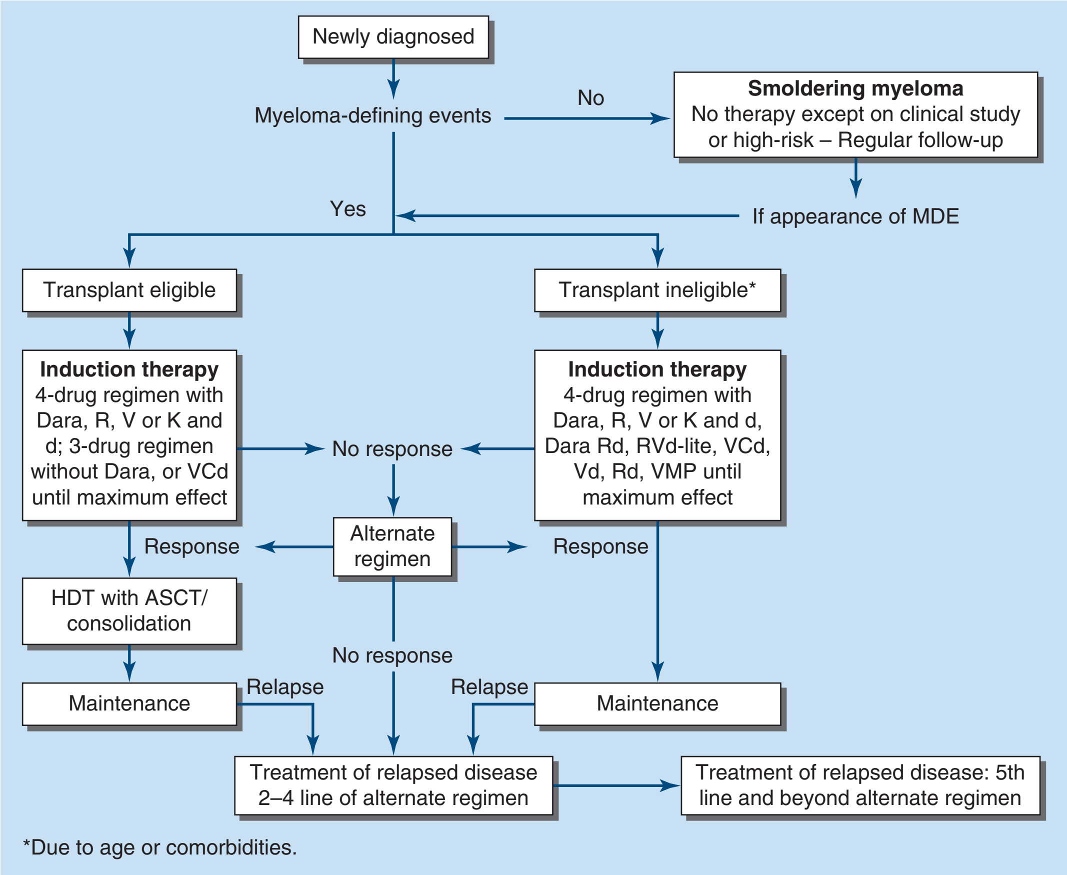
Newly diagnosed
Myeloma-defining events
No
Smoldering myeloma
No therapy except on clinical study or high-risk – Regular follow-up
If appearance of MDE
Yes
Transplant eligible
Transplant ineligible*
Induction therapy
4-drug regimen with Dara, R, V or K and d; 3-drug regimen without Dara, or VCd until maximum effect
Induction therapy
4-drug regimen with Dara, R, V or K and d, Dara Rd, RVd-lite, VCd, Vd, Rd, VMP until maximum effect
No response
Alternate regimen
Response
Response
HDT with ASCT/ consolidation
No response
Maintenance
Relapse
Relapse
Maintenance
Treatment of relapsed disease 2–4 line of alternate regimen
Treatment of relapsed disease: 5th line and beyond alternate regimen
*Due to age or comorbidities.

FIGURE 116-6 Treatment algorithm for multiple myeloma. *Alternate regimen* indicates combinations including daratumumab, elotuzumab, panobinostat, carfilzomib, ixazomib, pomalidomide, or other agents. ASCT, autologous stem cell transplantation; C, cyclophosphamide; D, dexamethasone; HDT, high-dose therapy; M, melphalan; MDE, myeloma-defining events; P, prednisone; R, lenalidomide; RVD-lite, weekly regimen; V, bortezomib.

affect a patient's ability to undergo high-dose therapy and transplantation **(Fig. 116-6)**.

Three important classes of agents approved for treatment of newly diagnosed MM are immunomodulatory agents, proteasome inhibitors, and antibodies targeting CD38. The combination of lenalidomide with a proteasome inhibitor (bortezomib or carfilzomib) and dexamethasone achieves close to a 100% response rate and a >50% complete response (CR) rate, making this combination one of the preferred induction regimens in transplant-eligible patients. Other similar three-drug combinations (bortezomib, thalidomide, and dexamethasone or bortezomib, cyclophosphamide, and dexamethasone) also achieve >90% response rate. Addition of a fourth agent, daratumumab or isatuximab, an anti-CD38 antibody, is providing even deeper responses and such four-drug regimens as induction therapy are becoming standard of care. Usually between four and six cycles of these combination regimens are utilized to achieve initial deep cytoreduction before consideration of high-dose therapy with autologous stem cell transplantation.

In patients who are not transplant candidates due to physiologic age >70 years, significant cardiopulmonary problems, or other comorbid illnesses, the same three-drug combinations described above are considered standard of care as induction therapy with age- and frailty-guided dose and schedule modifications. Combination of daratumumab with lenalidomide and dexamethasone (DRd) and modified lenalidomide-bortezomib-dexamethasone (RVD lite) combination achieve high overall response rates (93 and 86%, respectively) and CR rates (47 and 32%, respectively). The four-drug combinations used in younger patients are also utilized in healthy older patients with excellent response. In the past, the combination of melphalan with prednisone (MP) alone or with thalidomide or bortezomib has been utilized effectively, but with the availability of newer agents and combinations, MP-based combinations are now not utilized.

HIGH-DOSE THERAPY WITH AUTOLOGOUS STEM CELL TRANSPLANTATION

High-dose therapy (HDT) and consolidation/maintenance are standard practice in the majority of eligible patients. In patients who are transplant candidates and receiving lenalidomide, stem cells should be collected within 6 months because the continued use of lenalidomide may compromise the ability to collect adequate numbers of stem cells. Randomized studies comparing standard-dose therapy to high-dose melphalan therapy with hematopoietic stem cell support have shown that HDT can achieve higher overall response rates, with up to 25–40% additional CRs and prolonged progression-free and overall survival; however, few, if any, patients are cured. Although two successive HDTs (tandem transplantations) are more effective than single HDT, the benefit is only observed in the subset of patients who do not achieve a complete or very good partial response to the first transplantation, which is a rare subset. Moreover, a randomized study failed to show any significant difference in overall survival between early transplantation after induction therapy versus delayed transplantation at relapse. These data allow an option to delay transplantation, especially with the availability of newer agents and combinations. Allogeneic transplantations may also produce high response rates, but with significant toxicities. Nonmyeloablative allogeneic transplantation can reduce toxicity but is recommended only under the auspices of a clinical trial to exploit an immune graft-versus-myeloma effect while avoiding attendant toxicity. Due to effective cellular therapies, allogeneic transplantation is very rarely used.

Maintenance therapy prolongs remissions following standard-dose regimens as well as HDT. Several phase 3 studies have demonstrated improved progression-free survival, and one study showed prolonged overall survival in patients receiving lenalidomide compared to placebo as maintenance therapy after HDT. In nontransplant candidates, two phase 3 studies showed prolonged progression-free survival with lenalidomide maintenance after MP plus lenalidomide or lenalidomide plus dexamethasone induction therapy. Although concern arises regarding an increased incidence of second primary malignancies in patients receiving lenalidomide maintenance, its benefits in reducing the risk of progressive disease and death from myeloma far outweigh the small increased risk of second cancers. In patients with high-risk cytogenetics, lenalidomide and bortezomib or an oral proteasome inhibitor, ixazomib, show promise as maintenance combination therapy after transplantation. A phase 3 study has also demonstrated the benefit of maintenance therapy with daratumumab after HDT.

RELAPSED DISEASE

Relapsed myeloma can be treated with a number of agents including lenalidomide and/or bortezomib, if previously not used. The second-generation proteasome inhibitor carfilzomib and immunomodulatory agent pomalidomide have shown efficacy in relapsed and refractory MM, even MM refractory to lenalidomide and bortezomib. An oral proteasome inhibitor, ixazomib, has also been approved in combination with lenalidomide and dexamethasone as an all-oral regimen for relapsed MM. Four antibodies are approved for treatment of relapsed MM. Daratumumab targeting CD38 achieves high response rates and improved progression-free survival as a single agent with further improvement in response and survival when added to bortezomib and dexamethasone or lenalidomide and dexamethasone. A formulation of daratumumab for subcutaneous administration provides decreased toxicity and improved convenience. Isatuximab, another antibody targeting CD38, achieves high response rates and improved progression-free survival in combination with pomalidomide or carfilzomib and dexamethasone. Elotuzumab, which targets SLAMF7, has shown significant activity in combination with lenalidomide and dexamethasone in relapsed/refractory myeloma but not as a single agent. Finally, belantamab mafodotin, an antibody-drug conjugate, targets B-cell maturation antigen (BCMA), which is expressed mainly on normal plasma cells and myeloma cells and delivers auristatin, a microtubule inhibitor, to the tumor cells and achieves responses in relapsed/refractory myeloma. The drug has a unique ophthalmologic toxicity that requires close monitoring. Its efficacy increases significantly when combined with pomalidomide or proteasome inhibitor.

Panobinostat, a histone deacetylase inhibitor, in combination with bortezomib and dexamethasone has been approved for treatment of relapsed/refractory myeloma based on superior response and progression-free survival compared to bortezomib and dexamethasone alone. Two additional newer agents have unique mechanisms of action: selinexor is a first-in-class exportin inhibitor that blocks export of proteins from the cell nucleus, and melflufen is an alkylating agent conjugated to a peptide to improve specific delivery to myeloma cells that express aminopeptidase required for cleaving of the peptide to deliver the drug intracellularly in myeloma cells. Both agents have been approved based on their effectiveness in relapsed/refractory myeloma.

Immunotherapeutic Approaches Two cellular therapies, both targeting BCMA, are approved for relapsed myeloma. The anti-BCMA CAR transduced T cell, idecabtagene vicleucel (Ide-cel), is approved after at least two prior lines of therapy that includes an immunomodulatory agent, a proteasome inhibitor, and an anti-CD38 monoclonal antibody, whereas ciltacabtagene autoleucel (Cilta-cel) is approved for patients who have received at least one prior line of therapy, including a proteasome inhibitor and an immunomodulatory agent and who are refractory to lenalidomide. In patients with advanced myeloma with a median of six prior lines of treatment, 81% of patients receiving target dose of Ide-cel responded, and a CR rate of 33% was observed. Cilta-cel in a similar patient population achieved 98% overall response with 82% achieving CR. Cytokine release syndrome and neurotoxicity remain primary toxicities requiring close monitoring and aggressive management.

Three bispecific antibodies are also approved for treatment of relapsed and refractory myeloma after four prior lines of therapies. Bispecific antibodies can bind to two different cell surface targets through two distinct binding domains. One domain binds to CD3 on human lymphocytes, while the other domain binds to a cell surface target specific for myeloma cells, bringing the T cells closer and leading to myeloma cell death. Two bispecific antibodies, teclistamab and elranatamab, target BCMA, whereas talquetamab targets a novel cell surface target on myeloma cells, G protein–coupled receptor class C group 5 (GPRC5D). In patients with relapsed refractory myeloma, these antibodies result in a 60–70% response rate, with a quarter of the patients achieving CR. Bispecific antibodies are associated with toxicities including cytokine release syndrome, neurotoxicity, and significant susceptibility to infectious complications. Incorporation of the large number of active agents at various stages of treatment, including in newly diagnosed patients, is improving survival as well as quality of life.

THERAPY ENDPOINT

Improvement in the serum M component may lag behind the symptomatic improvement due to longer serum half-life (~3 weeks) of the immunoglobulin. The fall in M component depends on the rate of tumor kill and the fractional catabolic rate of immunoglobulin. Serum and urine light chains with a functional half-life of ~6 h may fall much quicker within the first week of treatment. Because urine light chain levels may relate to renal tubular function, they are not a reliable measure of tumor cell kill in patients with renal dysfunction. Achieving CR, defined as disappearance of serum and urine monoclonal protein with normal bone marrow by light microscopy, has been a standard goal of therapy. However, sequencing or multicolor flow cytometry–based assessment of minimal residual disease (MRD) in bone marrow to measure the presence of one myeloma cell in a million cells is being considered as an important new endpoint, both in newly diagnosed and relapsed patients. Absence of MRD at this sensitivity predicts for both longer progression-free survival and longer overall survival. Although patients may not achieve complete remission, clinical responses may last for long periods of time in small numbers of patients.

The median overall survival of patients with myeloma is 8+ years, with subsets of younger patients surviving >10 years. The major causes of death are progressive myeloma, renal failure, sepsis, or therapy-related myelodysplasia. Nearly a quarter of patients die of myocardial infarction, chronic lung disease, diabetes, or stroke, which are all intercurrent illnesses related more to the age of the patient group than to the tumor.

SUPPORTIVE THERAPY

Herpes zoster prophylaxis is indicated if bortezomib is used, and neuropathy attendant to bortezomib can be decreased both by its subcutaneous administration and by administration on a weekly schedule. Lenalidomide use requires prophylaxis for deep-vein thrombosis (DVT) with either aspirin or, if patients are at a greater risk of DVT, warfarin, low-molecular-weight heparin, or direct-acting anticoagulants DOACs. Patients receiving anti-BCMA CAR-T cell therapy or bi-specific antibodies may need supplementation with intravenous γ globulin due to induction of prolonged hypogammaglobulinemia.

Supportive care directed at the anticipated complications of the disease may be as important as primary antitumor therapy. Hypercalcemia generally responds well to bisphosphonates, glucocorticoid therapy, hydration, and natriuresis and rarely requires calcitonin as well. Bisphosphonates (e.g., pamidronate 90 mg or zoledronate 4 mg initially once a month for 12–24 months and later every 2–3 months) reduce osteoclastic bone resorption and preserve performance status and quality of life, decrease bone-related complications, and may also have antitumor effects. Osteonecrosis of the jaw and renal dysfunction can occur in a minority of patients receiving bisphosphonate therapy. Denosumab is an alternative agent administered intravenously at 120 mg monthly and achieves a similar level of effect as bisphosphonates to prevent bone-related complications in myeloma. Treatments aimed at strengthening the skeleton such as fluorides, calcium, and vitamin D, with or without androgens, have been suggested but are not of proven efficacy. Kyphoplasty or vertebroplasty should be considered in patients with painful collapsed vertebra. Iatrogenic worsening of renal function may be prevented by maintaining a high fluid intake to prevent dehydration and enhance excretion of light chains and calcium. In the event of acute renal failure, plasmapheresis is ~10 times more effective at clearing light chains than peritoneal dialysis; however, its role in reversing renal failure remains controversial. Importantly, reducing the protein load by effective antitumor therapy with agents such as bortezomib may result in improvement in renal function in over half of the patients. Use of lenalidomide in renal failure is possible but requires dose modification because it is renally excreted. Urinary tract infections should be watched for and treated early. Plasmapheresis may be the treatment of choice for hyperviscosity syndromes. Although the pneumococcus is a dreaded pathogen in myeloma patients, pneumococcal polysaccharide vaccines may not elicit an antibody response. The pneumococcal conjugate vaccines are more protective. Prophylactic administration of intravenous γ globulin preparations is used in the setting of recurrent serious infections and in patients receiving CAR-T therapy or bispecific antibodies. Routine chronic oral antibiotic prophylaxis is not warranted. Patients receiving bortezomib or CD38-directed therapies receive prophylaxis against herpes zoster. Patients with myeloma and even SMM and MGUS may be at a higher risk of developing COVID-19 and thus are encouraged to have vaccination for prevention and, if positive for infection, to receive antiviral therapy for COVID-19. Patients developing neurologic symptoms in the lower extremities, severe localized back pain, or problems with bowel and bladder control may need emergency MRI and local radiation therapy and glucocorticoids if cord compression is identified. In patients in whom neurologic deficit is increasing or substantial, emergent surgical decompression may be necessary. Most bone lesions respond to analgesics and systemic therapy, but certain painful lesions may respond more promptly to localized radiation. The anemia associated with myeloma may respond to erythropoietin along with hematinics (iron, folate, cobalamin). The pathogenesis of the anemia should be established and specific therapy instituted, whenever possible.

WALDENSTRÖM'S MACROGLOBULINEMIA

In 1948, Waldenström described a malignancy of lymphoplasmacytoid cells that secreted IgM. In contrast to myeloma, the disease was associated with lymphadenopathy and hepatosplenomegaly, but the major clinical manifestation was hyperviscosity syndrome. The disease resembles the related diseases CLL, myeloma, and lymphocytic lymphoma. It originates from a post–germinal center B cell that has undergone somatic mutations and antigenic selection in the lymphoid follicle and has the characteristics of an IgM-bearing memory B cell. Waldenström's macroglobulinemia (WM) and IgM myeloma follow a similar clinical course, but therapeutic options are different. The diagnosis of IgM myeloma is usually reserved for patients with lytic bone lesions and predominant infiltration with CD138+ plasma cells usually with t(11;14) translocation in the bone marrow. Such patients are at greater risk of pathologic fractures than patients with WM.

A familial occurrence is common in WM, but its molecular bases are yet unclear. A distinct *MYD88* L265P somatic mutation is present in >90% of patients with WM and the majority of IgM MGUS. Other commonly occurring mutations include *CXCR4* (30–40%), *ARID1A* (17%), and *CD79B* (8–15%). Presence of *MYD88* mutation status is now used as a diagnostic test to discriminate WM from marginal zone lymphomas (MZLs), IgM-secreting myeloma, and CLL with plasmacytic differentiation. This mutation also explains the molecular pathogenesis of the disease with involvement of Toll-like receptor (TLR) and interleukin 1 receptor (IL-1R) signaling leading to activation of IL-1R–associated kinase (IRAK) 4 and IRAK1 followed by nuclear factor-κB (NF-κB) activation. *MYD88* mutation also triggers Bruton's tyrosine kinase (BTK) and hemopoietic cell kinase (HCK)-mediated growth and survival signaling, which are now important therapeutic targets in WM. *CXCR4* mutations induce AKT and extracellular regulated kinase 1/2 (ERK1/2) signaling. This pathway can lead to development of drug resistance in the presence of its ligand CXCL12.

The disease is similar to myeloma in being slightly more common in men and occurring with increased incidence with increasing age (median age 64 years). The IgM in some patients with macroglobulinemia may have specificity for myelin-associated glycoprotein (MAG), a protein that has been associated with demyelinating disease of the peripheral nervous system and may be lost earlier and to a greater extent than the better-known myelin basic protein in patients with multiple sclerosis. Sometimes patients with macroglobulinemia develop a peripheral neuropathy, and half of these patients are positive for anti-MAG antibody. The neuropathy may precede the appearance of the neoplasm. The whole process may begin with a viral infection that may elicit an antibody response that cross-reacts with a normal tissue component.

Like myeloma, the disease involves the bone marrow, but unlike myeloma, it does not cause bone lesions or hypercalcemia. Bone marrow shows >10% infiltration with lymphoplasmacytic cells (surface IgM+, CD19+, CD20+, and CD22+, rarely CD5+, but CD10− and CD23−) with an increase in number of mast cells. Like myeloma, an M component is present in the serum in excess of 30 g/L (3 g/dL), but unlike myeloma, the size of the IgM paraprotein results in little renal excretion, and only ~20% of patients excrete light chains. Therefore, renal disease is not common. The light chain isotype is kappa in 80% of the cases. Patients present with weakness, fatigue, and recurrent infections similar to myeloma patients, but epistaxis, visual disturbances, and neurologic symptoms such as peripheral neuropathy, dizziness, headache, and transient paresis are much more common in macroglobulinemia. Presence of *MYD88* and *CXCR4* mutations also affects disease presentation. Presence of *CXCR4* mutations is associated with higher bone marrow disease burden and higher incidence of hyperviscosity. Patients with wild-type *MYD88* show lower bone marrow disease burden.

Physical examination reveals adenopathy and hepatosplenomegaly, and ophthalmoscopic examination may reveal vascular segmentation and dilation of the retinal veins characteristic of hyperviscosity states. Patients may have a normocytic, normochromic anemia, but rouleaux formation and a positive Coombs test are much more common than in myeloma. Malignant lymphocytes are usually present in the peripheral blood. About 10% of macroglobulins are cryoglobulins. These are pure M components and are not the mixed cryoglobulins seen in rheumatoid arthritis and other autoimmune diseases. Mixed cryoglobulins are composed of IgM or IgA complexed with IgG, for which they are specific. In both cases, Raynaud's phenomenon and serious vascular symptoms precipitated by the cold may occur, but mixed cryoglobulins are not commonly associated with malignancy. Patients suspected of having a cryoglobulin based on history and physical examination should have their blood drawn into a warm syringe and delivered to the laboratory in a container of warm water to avoid errors in quantitating the cryoglobulin.

TREATMENT

Waldenström's Macroglobulinemia

A diagnosis of WM requires lymphoplasmacytic infiltrate of any level in the bone marrow and an IgM monoclonal paraprotein of any size. Treatment is usually not initiated unless the disease is symptomatic or increasing anemia, hyperviscosity, lymphadenopathy, or hepatosplenomegaly is present. Control of serious hyperviscosity symptoms such as an altered state of consciousness or paresis can be achieved acutely by plasmapheresis because 80% of the IgM paraprotein is intravascular. The median survival of affected individuals is ~50 months. However, many patients with WM have indolent disease that does not require therapy. Pretreatment parameters including older age, male sex, general symptoms, and cytopenias define a high-risk population. BTK inhibitors (ibrutinib), alkylating drugs (bendamustine and cyclophosphamide), and proteasome inhibitors (bortezomib, carfilzomib, and ixazomib), alone or more frequently in combination with rituximab, are considered as first-line therapy for symptomatic patients with WM. Ibrutinib targets the constitutively activated BTK. In patients with one prior line of therapy, the overall response to ibrutinib was 91%. Best responses to ibrutinib are observed in patients with mutated *MYD88* and wild-type *CXCR4* status, while delayed and lower response rates to ibrutinib are observed in patients with mutated *CXCR4*. At first relapse, in patients with an initial durable response, either the previous regimen or another primary therapy regimen can be used. The therapeutic choice is dependent upon the genomic features, drug availability, and the patient's clinical profile.

Rituximab can produce an IgM flare, so either plasmapheresis should be used before rituximab or its use should be initially withheld in patients with high IgM levels. Fludarabine (25 mg/m^2 per d for 5 days every 4 weeks) is also an effective single agent. With identification of the *MYD88* mutation, novel BTK inhibitors (acalabrutinib, zanubrutinib, and tirabrutinib), inhibitors targeting IRAK1/4, and the BCL2 antagonist venetoclax are being explored for the treatment of WM. HDT plus autologous transplantation has been utilized in the past, but it is now not recommended due to the availability of other effective agents.

POEMS SYNDROME

The features of this syndrome are *p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathy, *M*-protein, and *s*kin changes (POEMS). Diagnostic criteria are described in Table 116-1. Patients usually have a severe, progressive sensorimotor polyneuropathy associated with sclerotic bone lesions from myeloma. Polyneuropathy occurs in ~1.4% of myelomas, but the POEMS syndrome is only a rare subset of that group. Unlike typical myeloma, hepatomegaly and lymphadenopathy occur in about two-thirds of patients, and splenomegaly is seen in one-third. The lymphadenopathy frequently resembles Castleman's disease histologically, a condition that has been linked to IL-6 overproduction. The endocrine manifestations include amenorrhea in women and impotence and gynecomastia in men. Hyperprolactinemia due to loss of normal inhibitory control by the hypothalamus may be associated with other central nervous system manifestations such as papilledema and elevated cerebrospinal fluid pressure and protein. Type 2 diabetes mellitus occurs in about one-third of patients. Hypothyroidism and adrenal insufficiency are occasionally noted. Skin changes are diverse: hyperpigmentation, hypertrichosis, skin thickening, and digital clubbing. Other manifestations include peripheral edema, ascites, pleural effusions, fever, and thrombocytosis. Not all the components of POEMS syndrome may be present initially.

The pathogenesis of the disease is unclear, but high circulating levels of the proinflammatory cytokines IL-1, IL-6, VEGF, and TNF have been documented, and levels of the inhibitory cytokine transforming growth factor β are lower than expected. Treatment of the myeloma may result in an improvement in the other disease manifestations.

Patients are often treated similarly to those with myeloma. Plasmapheresis does not appear to be of benefit in POEMS syndrome. Patients presenting with isolated sclerotic lesions may have resolution of neuropathic symptoms after local therapy for plasmacytoma with radiotherapy. Similar to MM, novel agents and HDT with autologous stem cell transplantation have been pursued in selected patients and have been associated with prolonged progression-free survival.

HEAVY CHAIN DISEASES

The heavy chain diseases are rare lymphoplasmacytic malignancies. Their clinical manifestations vary with the heavy chain isotype. Patients have absence of light chain and secrete a defective heavy chain that usually has an intact Fc fragment and a deletion in the Fd region. Gamma, alpha, and mu heavy chain diseases have been described, but no reports of delta or epsilon heavy chain diseases have appeared. Molecular biologic analysis of these tumors has revealed structural genetic defects that may account for the aberrant chain secreted.

GAMMA HEAVY CHAIN DISEASE (FRANKLIN'S DISEASE)

This disease affects individuals of widely different age groups and countries of origin. It is characterized by lymphadenopathy, fever, anemia, malaise, hepatosplenomegaly, and weakness. It is frequently associated with autoimmune diseases, especially rheumatoid arthritis. Its most distinctive symptom is palatal edema, resulting from involvement of nodes in Waldeyer's ring, and this may progress to produce respiratory compromise. The diagnosis depends on the demonstration of an anomalous serum M component (often <20 g/L [<2 g/dL]) that reacts with anti-IgG but not anti–light chain reagents. The M component is typically present in both serum and urine. Most of the paraproteins have been of the γ_1 subclass, but other subclasses have been seen. The patients may have thrombocytopenia, eosinophilia, and nondiagnostic bone marrow that may show increased numbers of lymphocytes or plasma cells that do not stain for light chain. Patients usually have a rapid downhill course and die of infection; however, some patients have survived 5 years with chemotherapy. Therapy is indicated when symptomatic and involves chemotherapeutic combinations used in low-grade lymphoma. Rituximab has also been reported to show efficacy.

ALPHA HEAVY CHAIN DISEASE (SELIGMANN'S DISEASE)

This is the most common of the heavy chain diseases. It is closely related to a malignancy known as Mediterranean lymphoma, a disease that affects young persons in parts of the world where intestinal parasites are common, such as the Mediterranean, Asia, and South America. The disease is characterized by an infiltration of the lamina propria of the small intestine with lymphoplasmacytoid cells that secrete truncated alpha chains. Demonstrating alpha heavy chains is difficult because the alpha chains tend to polymerize and appear as a smear instead of a sharp peak on electrophoretic profiles. Despite the polymerization, hyperviscosity is not a common problem in alpha heavy chain disease. Without J chain–facilitated dimerization, viscosity does not increase dramatically. Light chains are absent from serum and urine. The patients present with chronic diarrhea, weight loss, and malabsorption and have extensive mesenteric and paraaortic adenopathy. Respiratory tract involvement occurs rarely. Patients may vary widely in their clinical course. Some may develop diffuse aggressive histologies of malignant lymphoma. Chemotherapy may produce long-term remissions. Rare patients appear to have responded to antibiotic therapy, raising the question of the etiologic role of antigenic stimulation, perhaps by some chronic intestinal infection. Chemotherapy plus antibiotics may be more effective than chemotherapy alone. IPSID is recognized as an infectious pathogen–associated human lymphoma associated with *Campylobacter jejuni*. It involves mainly the proximal small intestine, resulting in malabsorption, diarrhea, and abdominal pain. IPSID is associated with excessive plasma cell differentiation and produces truncated alpha heavy chain proteins lacking the light chains as well as the first constant domain. Early-stage IPSID responds to antibiotics (30–70% complete remission). Most untreated IPSID patients progress to lymphoplasmacytic and immunoblastic lymphoma. Patients not responding to antibiotic therapy are considered for treatment with combination chemotherapy used to treat low-grade lymphoma.

MU HEAVY CHAIN DISEASE

The secretion of isolated mu heavy chains into the serum appears to occur in a very rare subset of patients with CLL. The only features that may distinguish patients with mu heavy chain disease are the presence of vacuoles in the malignant lymphocytes and the excretion of kappa light chains in the urine. The diagnosis requires ultracentrifugation or gel filtration to confirm the nonreactivity of the paraprotein with the light chain reagents because some intact macroglobulins fail to interact with these serums. The tumor cells seem to have a defect in the assembly of light and heavy chains because they appear to contain both in their cytoplasm. Such patients are not treated differently from other patients with CLL **(Chap. 107)**.

■ FURTHER READING

Corre J et al: Risk factors in multiple myeloma: is it time for a revision? Blood 137:16, 2021.

Hideshima T, Anderson KC: Signaling pathway mediating myeloma cell growth and survival. Cancers (Basel) 13:216, 2021.

Hillengass J et al: International myeloma working group consensus recommendations on imaging in monoclonal plasma cell disorders. Lancet Oncol 20:e302, 2019.

Kumar S et al: International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. Lancet Oncol 17:e328, 2016.

Moreau P et al: Treatment of relapsed and refractory multiple myeloma: Recommendations from the International Myeloma Working Group. Lancet Oncol 22:e105, 2021.

Munshi NC et al: A large meta-analysis establishes the role of MRD negativity in long-term survival outcomes in patients with multiple myeloma. Blood Adv 4:5988, 2020.

Raje NS et al: Consensus guidelines and recommendations for infection prevention in multiple myeloma: A report from the International Myeloma Working Group. Lancet Haematol 9:e143 2022.

Rajkumar SV et al: International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. Lancet Oncol 15:e538, 2014.

Richardson PG et al: Triplet therapy, transplantation, and maintenance until progression in myeloma. N Engl J Med 387:132, 2022.

Robiou du Pont S et al: Genomics of multiple myeloma. J Clin Oncol 35:963, 2017.

Terpos E et al: Treatment of multiple myeloma-related bone disease: Recommendations from the Bone Working Group of the International Myeloma Working Group. Lancet Oncol 22:e119, 2021.

Treon SP et al: How I use genomics and BTK inhibitors in the treatment of Waldenstrom macroglobulinemia. Blood 143:1702, 2024.

117

Amyloidosis

John L. Berk, Vaishali Sanchorawala

■ GENERAL PRINCIPLES

Amyloidosis is the term for a group of protein misfolding disorders characterized by the extracellular deposition of insoluble polymeric protein fibrils in tissues and organs. A robust cellular machinery exists to chaperone proteins during the process of synthesis and secretion, to ensure that they achieve correct tertiary conformation and function, and to eliminate proteins that misfold. However, genetic mutation, incorrect processing, and other factors may favor misfolding, with consequent loss of normal protein function and intracellular or extracellular aggregation. Many diseases, ranging from cystic fibrosis to Alzheimer's disease, are now known to involve protein misfolding. In the amyloidoses, the aggregates are typically extracellular, and the misfolded protein subunits assume a common antiparallel, β-pleated sheet–rich structural conformation that leads to the formation of higher-order oligomers and then fibrils with unique staining properties. The term *amyloid* was coined around 1854 by the pathologist Rudolf Virchow, who thought that these deposits resembled starch (Latin *amylum*) under the microscope.

Amyloid diseases, defined by the biochemical nature of the protein composing the fibril deposits, are classified according to whether they are systemic or localized, whether they are acquired or inherited, and their clinical patterns **(Table 117-1)**. The standard nomenclature is *AX*, where *A* indicates amyloidosis and *X* represents the protein present in the fibril. This chapter focuses primarily on the systemic forms. *AL* amyloidosis refers to amyloid composed of immunoglobulin light chains; this disorder, formerly termed *primary systemic amyloidosis*, arises from a clonal B-cell or plasma cell disorder and can be associated with myeloma or lymphoma. *ATTR* amyloidosis, the most prevalent of the *familial amyloidoses*, refers to amyloid derived from wild-type or mutated transthyretin (TTR), the transport protein for thyroid hormone and retinol-binding protein. *AA* amyloid is composed of the acute-phase reactant protein serum amyloid A (SAA) and occurs in the setting of chronic inflammatory or infectious diseases; for this reason, this type was formerly known as *secondary amyloidosis*. *A*β$_2$*M* amyloid results from misfolded β$_2$-microglobulin, occurring in individuals with long-standing renal disease who have undergone dialysis, typically for years. *A*β, the most common form of localized amyloidosis, is found in the brain of patients with Alzheimer's disease after abnormal proteolytic processing and aggregation of polypeptides derived from the amyloid precursor protein.

Diagnosis and treatment of the amyloidoses rest upon the histopathologic identification of amyloid deposits and immunohistochemical, biochemical, or genetic determination of amyloid type **(Fig. 117-1)**. In the systemic amyloidoses, the clinically involved organs can be biopsied, but amyloid deposits may be found in any tissue of the body. Historically, blood vessels of the gingiva or rectal mucosa were often examined, but the most easily accessible tissue—positive in more than 80% of patients with systemic amyloidosis—is abdominal fat. After local anesthesia, fat is aspirated with a 16-gauge needle from the subcutaneous layer of the abdominal wall. Fat globules expelled onto a glass slide can be stained for amyloid by Congo red dye, thus avoiding a surgical procedure. If this material is negative, more invasive biopsies of the involved organ like kidney, heart, liver, tongue, or gastrointestinal tract can be considered in patients in whom amyloidosis is suspected. The regular β-sheet structure of amyloid deposits exhibits a unique "green" birefringence by polarized light microscopy when stained with Congo red dye; other regular protein structures (e.g., collagen) appear white under these conditions. The 10-nm-diameter fibrils can also be visualized by electron microscopy of paraformaldehyde-fixed tissue. Once amyloid is found, the precursor protein type must be determined by immunohistochemistry, immunoelectron microscopy, or extraction and biochemical analysis employing mass spectrometry; gene sequencing is used to identify mutants causing hereditary amyloidosis. However, a mass spectrometry–based analysis of the amyloid-containing tissues is now considered the best approach, with a reported sensitivity of 88% and specificity of 96%, which are higher than immunochemical techniques, and this technique does not require a large panel of antisera to identify non-AL amyloidosis. The patient's history, physical findings, and clinical presentation, including age and ethnic origin, organ system involvement, underlying diseases, and family history, may provide helpful clues as to the type of amyloidosis. However, there can be considerable overlap in clinical presentations, and accurate typing of the amyloidogenic protein is essential to guide appropriate therapy and offer genetic counseling as appropriate.

The mechanisms of fibril formation and tissue toxicity remain controversial. The "amyloid hypothesis," as it is currently understood, proposes that precursor proteins undergo a process of reversible unfolding or misfolding; misfolded proteins form oligomeric aggregates, higher-order polymers, and then fibrils that deposit in tissues. Accumulating evidence suggests that the oligomeric intermediates may constitute the most toxic species. Oligomers are more capable than fibrils of interacting with cells and inducing formation of reactive oxygen species and stress signaling. Ultimately, the fibrillar tissue deposits are likely to interfere with normal organ function. However, direct proteotoxicity of the soluble oligomers also can lead to organ dysfunction. A more sophisticated understanding of the mechanisms leading to amyloid formation and cell and tissue dysfunction will continue to provide new targets for therapies.

The clinical syndromes of the amyloidoses are associated with relatively nonspecific alterations in routine laboratory tests. Blood counts are usually normal, although the erythrocyte sedimentation rate is

TABLE 117-1 Amyloid Precursor Proteins and Their Clinical Syndromes

DESIGNATION	PRECURSOR	CLINICAL SYNDROME	CLINICAL INVOLVEMENT
Systemic Amyloidoses			
AL	Immunoglobulin light chain	Primary or myeloma-associated[a]	Any
AH	Immunoglobulin heavy chain	Rare variant of primary or myeloma-associated	Any
AA	Serum amyloid A protein	Secondary; reactive[b]	Renal, heart, other
$A\beta_2M$	β_2-Microglobulin	Hemodialysis-associated	Synovial tissue, bone
ATTR	Transthyretin	Familial (mutant) Age-related (wild type)	Cardiac, peripheral and autonomic nerves, soft tissues, spine, bladder
AApoAI	Apolipoprotein AI	Familial	Hepatic, renal
AApoAII	Apolipoprotein AII	Familial	Renal
Agel	Gelsolin	Familial	Cornea, cranial nerves, skin, renal
AFib	Fibrinogen Aα	Familial	Renal, vascular
ALys	Lysozyme	Familial	Renal, hepatic
ALECT2	Leukocyte chemotactic factor 2	Undefined	Renal
Localized Amyloidoses			
Aβ	Amyloid β protein	Alzheimer's disease; Down's syndrome	Central nervous system
ACys	Cystatin C	Cerebral amyloid angiopathy	Central nervous system, vascular
APrP	Prion protein	Spongiform encephalopathies	Central nervous system
AIAPP	Islet amyloid polypeptide (amylin)	Diabetes-associated	Pancreas
Acal	Calcitonin	Medullary carcinoma of the thyroid	Thyroid
AANF	Atrial natriuretic factor	Atrial fibrillation	Cardiac atria
APro	Prolactin	Endocrinopathy	Pituitary
ASgl	Semenogelin I	Age-related; incidental autopsy or biopsy finding	Seminal vesicles

[a] Localized AL deposits can occur in skin, conjunctiva, urinary bladder, and the tracheobronchial tree. [b] Secondary to chronic inflammation or infection or to a hereditary periodic fever syndrome such as familial Mediterranean fever.

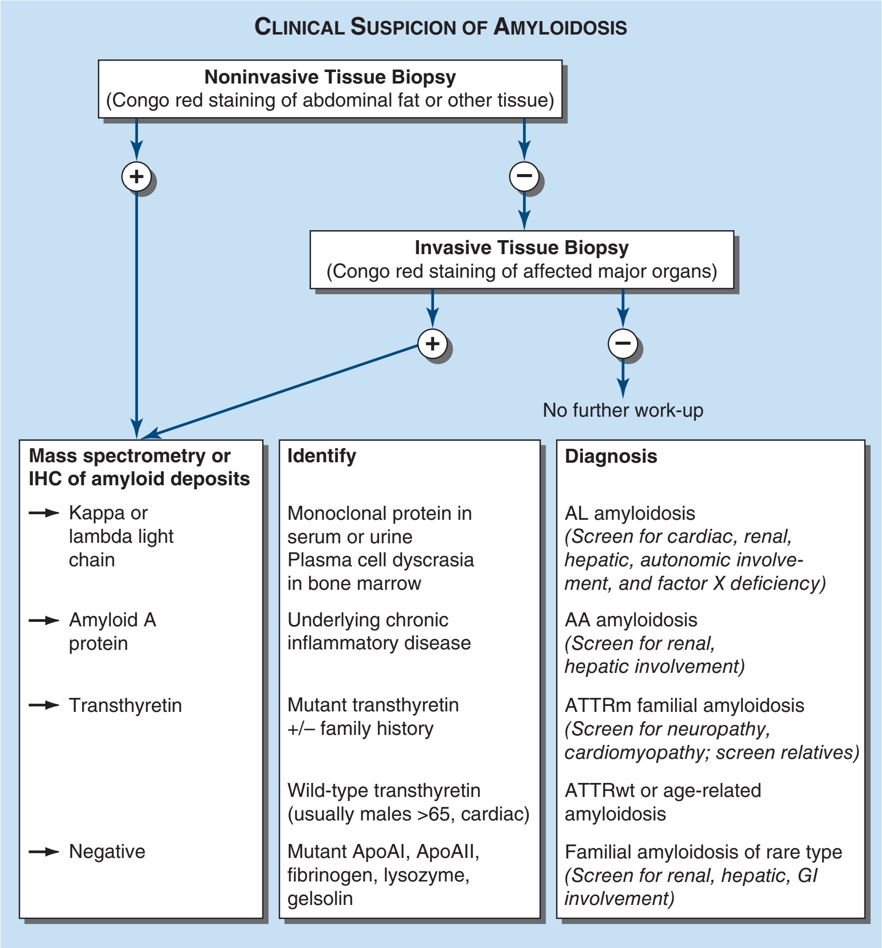
CLINICAL SUSPICION OF AMYLOIDOSIS
Noninvasive Tissue Biopsy
(Congo red staining of abdominal fat or other tissue)
+
−
Invasive Tissue Biopsy
(Congo red staining of affected major organs)
+
−
No further work-up
Mass spectrometry or IHC of amyloid deposits
Kappa or lambda light chain
Amyloid A protein
Transthyretin
Negative
Identify
Monoclonal protein in serum or urine
Plasma cell dyscrasia in bone marrow
Underlying chronic inflammatory disease
Mutant transthyretin +/– family history
Wild-type transthyretin (usually males >65, cardiac)
Mutant ApoAI, ApoAII, fibrinogen, lysozyme, gelsolin
Diagnosis
AL amyloidosis
(Screen for cardiac, renal, hepatic, autonomic involvement, and factor X deficiency)
AA amyloidosis
(Screen for renal, hepatic involvement)
ATTRm familial amyloidosis
(Screen for neuropathy, cardiomyopathy; screen relatives)
ATTRwt or age-related amyloidosis
Familial amyloidosis of rare type
(Screen for renal, hepatic, GI involvement)

FIGURE 117-1 Algorithm for the diagnosis of amyloidosis and determination of type. Clinical suspicion: unexplained nephropathy, cardiomyopathy, neuropathy, enteropathy, arthropathy, and macroglossia. ApoAI, apolipoprotein AI; ApoAII, apolipoprotein AII; GI, gastrointestinal; IHC, immunohistochemistry.

frequently elevated. Patients with glomerular kidney involvement generally have proteinuria, often in the nephrotic range, leading to hypoalbuminemia that may be severe; patients with serum albumin levels <2 g/dL generally have pedal edema or anasarca. Amyloid cardiomyopathy is characterized by concentric ventricular hypertrophy and diastolic dysfunction associated with elevation of brain natriuretic peptide (BNP) or *N*-terminal pro–brain natriuretic peptide (NT-proBNP) as well as troponin. These cardiac biomarkers can be used for disease staging, prognostication, and disease activity monitoring in patients with AL amyloidosis. Notably, renal insufficiency can falsely elevate levels of these biomarkers. Biomarkers of cardiac remodeling—that is, matrix metalloproteinases and tissue inhibitors of metalloproteinases—are altered in the serum of patients with amyloid cardiomyopathy. Electrocardiographic and echocardiographic features of amyloid cardiomyopathy are described below. Patients with liver involvement, even when advanced, usually develop cholestasis with an elevated alkaline phosphatase concentration with minimal alteration of the aminotransferases and preservation of synthetic function. In AL amyloidosis, endocrine organs may be involved, and hypothyroidism, hypoadrenalism, or even hypopituitarism can occur. Although none of these findings is specific for amyloidosis, the presence of abnormalities in multiple organ systems should raise suspicions of the diagnosis.

AL AMYLOIDOSIS

Etiology and Incidence AL amyloidosis is most frequently caused by a clonal expansion of bone marrow plasma cells that secrete a monoclonal immunoglobulin light chains forming amyloid fibrils and deposits in tissues. Whether the clonal plasma cells produce a light chain that misfolds and leads to AL amyloidosis or a light chain that folds properly, allowing the cells to inexorably expand over time and develop into multiple myeloma **(Chap. 116)**, may depend upon primary sequence of the clonal light chain or other genetic or epigenetic factors. AL amyloidosis can occur with multiple myeloma or other B lymphoproliferative diseases, including non-Hodgkin's lymphoma **(Chap. 113)** and Waldenström's macroglobulinemia **(Chap. 116)**. AL amyloidosis is the most common type of systemic amyloidosis diagnosed in North America. Its incidence has been estimated at 8–12 cases per 100,000 population; however, ascertainment continues to be inadequate, and the true incidence may be much higher. AL amyloidosis, like other plasma cell disorders, usually occurs after age 40 and is often progressive and fatal if untreated.

Pathology and Clinical Features Amyloid deposits are usually widespread in AL amyloidosis and can be present in the interstitium of any organ outside the central nervous system. The amyloid fibril deposits are composed of full-length 23-kDa monoclonal immunoglobulin light chains as well as fragments. Accessory molecules codeposited with light chain fibrils (as well as with other amyloid fibrils) include serum amyloid P component, apolipoproteins e and A-IV, glycosaminoglycans, and metal ions. Although all kappa and lambda light chain subtypes have been identified in AL amyloid fibrils, lambda subtypes predominate.

AL amyloidosis is often a rapidly progressive disease that presents as a pleiotropic set of clinical syndromes, recognition of which is key for initiation of the appropriate workup. Nonspecific symptoms of fatigue and weight loss are common; however, the diagnosis is rarely considered until symptoms referable to a specific organ develop. The kidneys are a frequently involved organ and are affected in 60–70% of patients. Renal amyloidosis usually manifests as proteinuria, often in the nephrotic range and associated with hypoalbuminemia, secondary hypercholesterolemia and hypertriglyceridemia, and edema or anasarca. In some patients, interstitial rather than glomerular amyloid deposition can produce azotemia without proteinuria. The heart is the other commonly affected organ (70–80% of patients), and cardiac involvement is the leading cause of death from AL amyloidosis. Early on, the electrocardiogram may show low voltage in the limb leads with a pseudo-infarct pattern. Echocardiographic features of disease include concentrically thickened ventricles and diastolic dysfunction with an abnormal global longitudinal strain pattern; a "sparkly" appearance has been described but is often not seen with modern high-resolution echocardiographic techniques. Poor atrial contractility occurs even in sinus rhythm, and patients with cardiac amyloidosis are at risk for development of atrial thrombi and thromboembolic complications. Cardiac magnetic resonance imaging (MRI) can show increased wall thickness and characteristic delayed gadolinium enhancement of the subendocardium. Nervous system symptoms include peripheral sensorimotor neuropathy and/or autonomic dysfunction manifesting as gastrointestinal motility disturbances (early satiety, diarrhea, constipation), dry eyes and mouth, impotence, orthostatic hypotension, and/or neurogenic bladder. Macroglossia **(Fig. 117-2*A*)**, a pathognomonic sign of AL amyloidosis, is seen in only ~10% of patients. Liver involvement causes cholestasis and hepatomegaly. The spleen is frequently involved, and there may be functional hyposplenism in the absence of significant splenomegaly. Many patients experience "easy bruising" due to amyloid deposits in capillaries or deficiency of clotting factor X due to binding to amyloid fibrils; cutaneous ecchymoses appear, particularly around the eyes, producing another uncommon but pathognomonic finding, the "raccoon-eye" sign **(Fig. 117-2*B*)**. Other findings include nail dystrophy **(Fig. 117-2*C*)**, alopecia, and amyloid arthropathy with thickening of synovial membranes in the wrists and shoulders. The presence of a multisystemic illness or general fatigue along with any of these clinical syndromes should prompt a workup for amyloidosis.

Diagnosis Identification of an underlying clonal plasma cell or B lymphoproliferative process and a clonal light chain are key to the diagnosis of AL amyloidosis. Serum protein electrophoresis and urine protein electrophoresis, although of value in multiple myeloma, are *not* useful screening tests if AL amyloidosis is suspected because the clonal light chain or whole immunoglobulin often is not present in sufficient amounts to produce a monoclonal "M-spike" in the serum or light chain (Bence Jones) protein in the urine. However, more than 90% of patients with AL amyloidosis have serum or urine monoclonal light chain or whole immunoglobulin detectable by immunofixation electrophoresis of serum (SIFE) or urine (UIFE) **(Fig. 117-3*A*)** or by nephelometric measurement of serum "free" light chains (i.e., light chains circulating in monomeric form rather than in an immunoglobulin tetramer with heavy chain). Examining the ratio as well as the absolute amount of serum-free light chains is essential, as renal insufficiency reduces light chain clearance, nonspecifically elevating both isotypes. In addition, an increased percentage of plasma cells in the bone marrow—typically 5–30% of nucleated cells—is found in ~90% of patients. Kappa or lambda clonality should be demonstrated by flow cytometry, immunohistochemistry, or in situ hybridization for light chain mRNA **(Fig. 117-3*B*)**. More sensitive mass spectrometry–based assays can have higher levels of detection for low concentration of monoclonal protein.

A monoclonal serum protein by itself is not diagnostic of amyloidosis, since monoclonal gammopathy of uncertain significance is common in older patients **(Chap. 116)**. However, when monoclonal gammopathy of uncertain significance is found in patients with biopsy-proven amyloidosis, the AL type should be ruled out. Similarly, patients thought to have "smoldering myeloma" because of a modest elevation of bone-marrow plasma cells should be screened for AL amyloidosis if

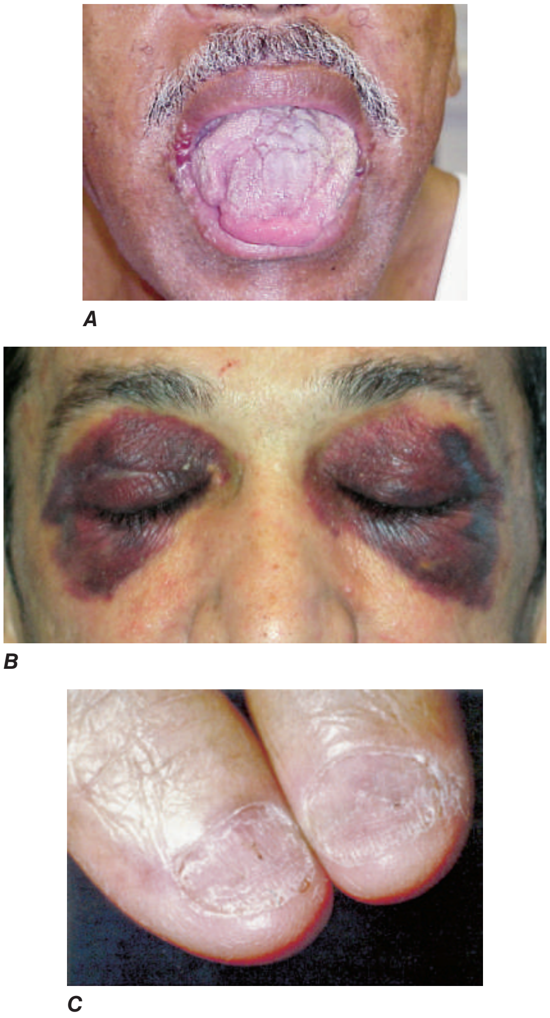

FIGURE 117-2 Clinical signs of AL amyloidosis. ***A.*** Macroglossia. ***B.*** Periorbital ecchymoses. ***C.*** Fingernail dystrophy.

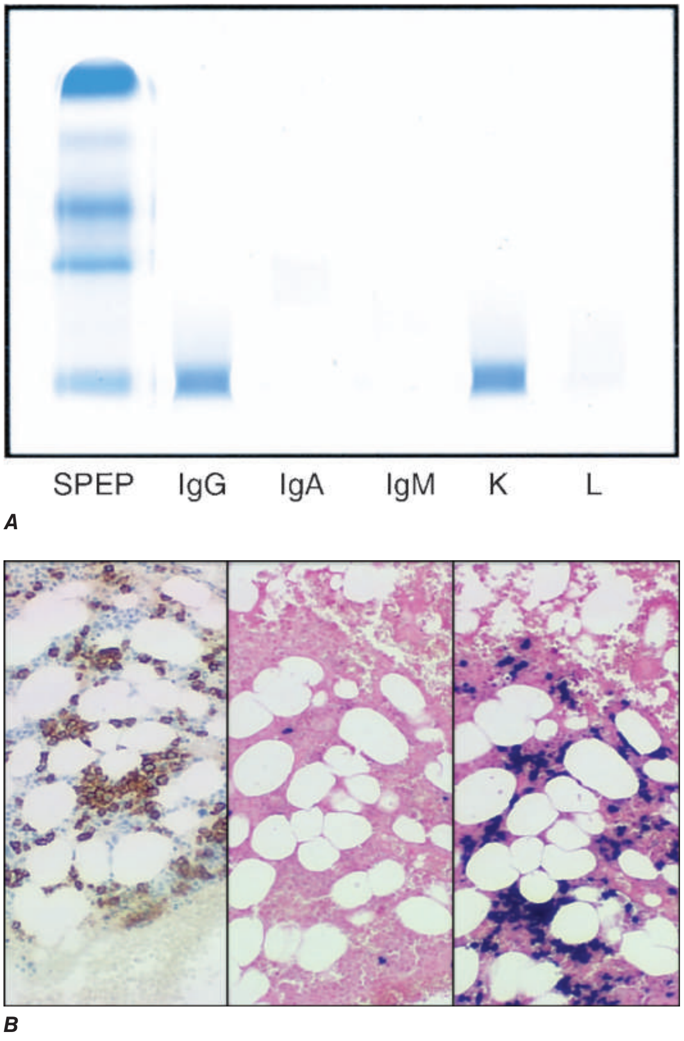

FIGURE 117-3 Laboratory features of AL amyloidosis. ***A.*** Serum immunofixation electrophoresis reveals an IgGκ monoclonal protein in this example; serum protein electrophoresis is often normal. ***B.*** Bone marrow biopsy sections stained by immunohistochemistry with antibody to CD138 (syndecan, highly expressed on plasma cells) (*left*) or by in situ hybridization with fluorescein-tagged probes (Ventana Medical Systems) binding to κ mRNA (*center*) and λ mRNA (*right*) in plasma cells. *(Photomicrograph courtesy of C. O'Hara; with permission.)*

they have signs or symptoms of renal, cardiac, or neurologic disease. Accurate tissue amyloid typing is essential for appropriate treatment. Immunohistochemical staining of the amyloid deposits is useful if they selectively bind one light chain antibody in preference to the other; some AL deposits bind antibodies nonspecifically. Commercial antibodies used for immunohistochemistry may not be accurate in amyloid typing. Immunoelectron microscopy is more reliable; laser capture microdissection and tandem mass spectrometry–based typing of the amyloid precursor protein have become the diagnostic standard. In ambiguous cases, other forms of amyloidosis should be thoroughly excluded with appropriate genetic and other testing.

Staging System and Risk Stratification The current staging systems for systemic AL amyloidosis are based on the biomarkers of plasma cell dyscrasia and cardiac and renal involvement. The Mayo 2004 staging system is based on the levels of NT-proBNP and cardiac troponins and was modified by European investigators to identify and classify very-high-risk patients. This cardiac staging system is the most widely used to determine patient management. This staging system was modified (Mayo 2012) to include clonal burden, assessed by dFLC (difference between involved and uninvolved circulating free light chain) concentration, which has independent ability to predict survival. Boston University investigators introduced a staging system incorporating BNP and troponin I that also is able to predict survival. Patients with AL amyloidosis with a very low (<50 mg/L) dFLC level have a significantly better outcome irrespective of cardiac stage. A renal staging system based on 24-h urine protein excretion and estimated glomerular filtration rate (eGFR) predicting the progression to dialysis at 2 years has also been developed and validated. Several other biomarkers have been shown to predict outcomes and survival but have not been incorporated in staging systems yet.

TREATMENT

AL Amyloidosis

Extensive multisystemic involvement typifies AL amyloidosis, and historically, the median survival without treatment was usually only ~1–2 years from the time of diagnosis. Marked progress in the outcome and survival has taken place over the past four decades with advent of new therapies, increased awareness, and accurate diagnosis. Current therapies target the clonal bone marrow plasma cells, using approaches employed for multiple myeloma. High-dose intravenous (IV) melphalan followed by autologous stem cell transplantation (HDM/SCT) produces complete hematologic responses in ~40% of treated patients, as determined by loss of clonal plasma cells in the bone marrow and disappearance of the amyloidogenic monoclonal light chain, as determined by SIFE/UIFE and free light chain quantitation. Six to 12 months after achieving a hematologic response, improvements in organ function and quality of life may occur. Hematologic responses appear to be more durable after HDM/SCT than in multiple myeloma, with remissions continuing in some patients beyond 15 years without additional treatment. Unfortunately, only ~20–30% of all AL amyloidosis patients are suitable for aggressive treatment, and even at specialized treatment centers, transplantation-related morbidity and mortality rates are higher than those for other hematologic diseases because of impaired organ function at initial presentation. Amyloid cardiomyopathy, poor nutritional and performance status, and multiorgan disease contribute to excess morbidity and mortality. A bleeding diathesis resulting from adsorption of clotting factor X to amyloid fibrils also increases mortality rates; however, this syndrome occurs in only 5–10% of patients. A randomized multicenter trial conducted in France compared oral melphalan and dexamethasone with HDM/SCT and failed to show a benefit of dose-intensive treatment, although the transplantation-related mortality rate in this study was very high. It has become clear that careful selection of patients and expert peritransplantation management are essential in reducing transplantation-related complications.

The best therapy for those who are not eligible to receive SCT is based on a U.S. Food and Drug Administration–approved therapy of CyBorD (cyclophosphamide, bortezomib [a proteasome inhibitor], and dexamethasone) with daratumumab. Patient characteristics should be considered when choosing a regimen; for example, treatment with bortezomib plus oral melphalan and dexamethasone (MDex) can overcome the effects of both gain of 1q21 (which confers a poorer outcome with oral melphalan) and t(11;14) (which confers a poorer outcome with bortezomib). Transplant-ineligible patients in whom bortezomib is contraindicated due to preexisting peripheral neuropathy can be treated with MDex or combinations based on immunomodulatory drugs (e.g., lenalidomide or pomalidomide). High-risk patients represent ~15–20% of all individuals with AL amyloidosis and are a challenge owing to advanced cardiac stage (IIIb) or severe heart failure (New York Heart Association class III or IV) as they are excluded from most of the clinical trials.

Novel antifibril monoclonal antibodies are currently undergoing clinical trials in combination with treatments directed against the plasma cell dyscrasia (CyBorD plus daratumumab [anti-CD38]) in patients with newly diagnosed AL amyloidosis. Clinical trials are essential in improving therapy for this rare disease.

Supportive care is important for patients with any type of amyloidosis. For nephrotic syndrome, diuretics and support stockings can ameliorate edema; angiotensin-converting enzyme inhibitors should be used with caution and have not been shown to slow renal disease progression. Effective diuresis can be facilitated with albumin infusions to raise intravascular oncotic pressure. Congestive heart failure due to amyloid cardiomyopathy is best treated with diuretics; it is important to note that digitalis, calcium channel blockers, and beta blockers are relatively contraindicated as they can interact with amyloid fibrils and produce heart block and worsening heart failure. Amiodarone has been used for atrial and ventricular arrhythmias. Automatic implantable defibrillators appear to have reduced effectiveness due to the thickened myocardium, but they may benefit some patients. Atrial ablation is an effective approach for atrial fibrillation. For conduction abnormalities, ventricular pacing may be indicated. Atrial contractile dysfunction is common in amyloid cardiomyopathy and associated with increased thromboembolic complications, prompting considerations of anticoagulation even in the absence of atrial fibrillation. Autonomic neuropathy can be treated with α agonists such as midodrine to support postural blood pressure; gastrointestinal dysfunction may respond to motility or bulk agents. Nutritional supplementation, either oral or parenteral, is also important.

In localized AL amyloidosis, amyloid deposits can be produced by clonal plasma cells infiltrating local sites in the airways, bladder, skin, or lymph nodes (Table 117-1). These deposits may respond to surgical intervention or elimination of the responsible plasma cell clone by low-dose radiation therapy (typically only 20 Gy); systemic treatment generally is not appropriate. Patients should be referred to a center familiar with management of these rare manifestations of amyloidosis.

AA AMYLOIDOSIS

Etiology and Incidence AA amyloidosis can occur in association with almost any chronic inflammatory state (e.g., rheumatoid arthritis, inflammatory bowel disease, ankylosing spondylitis, familial Mediterranean fever **[Chap. 381]**, or other periodic fever syndromes) or chronic infections such as tuberculosis, osteomyelitis, or subacute bacterial endocarditis. In the United States and Europe, AA amyloidosis has become less common, occurring in fewer than 2% of patients with these diseases, presumably because of advances in anti-inflammatory and antimicrobial therapies. It has also been described in association with Castleman's disease, lymphomas, and renal cell carcinoma, emphasizing the diagnostic importance of computed tomography (CT) scanning to look for such tumors as well as serologic and microbiologic studies. In up to 20% of patients, AA amyloidosis can also be seen without any identifiable underlying disease.

Pathology and Clinical Features Organ involvement in AA amyloidosis usually begins in the kidneys. Hepatomegaly, splenomegaly, and autonomic neuropathy can also occur as the disease progresses; cardiomyopathy is a late manifestation in ~10–25% of patients. The symptoms and signs of AA disease cannot be reliably distinguished from those of AL amyloidosis. AA amyloid fibrils are usually composed of an 8-kDa, 76-amino-acid *N*-terminal portion of the 12-kDa precursor protein SAA. This acute-phase protein is synthesized in the liver and transported by high-density lipoprotein (HDL3) in the plasma. Several years of an underlying inflammatory disease causing chronic elevation of SAA levels usually precede fibril formation, although infections can lead to AA amyloid deposition more rapidly.

TREATMENT

AA Amyloidosis

Primary therapy for AA amyloidosis consists of treatment of the underlying inflammatory or infectious disease. Treatment that suppresses or eliminates the inflammatory state or infection decreases the circulating levels of SAA, slowing the rate of amyloid fibril formation. For familial Mediterranean fever, colchicine at a dose of 1.2–1.8 mg/d is the standard treatment. However, colchicine has not been helpful for AA amyloidosis of other causes or for other amyloidoses. Tumor necrosis factor and interleukin 1 and interleukin 6 antagonists can effectively interrupt cytokine signaling that drives many inflammatory syndromes, inhibiting hepatic SAA production and limiting AA amyloid deposition. Development of a fibril-specific agent (eprodisate) that interferes with the interaction of serum amyloid A protein and glycosaminoglycans to prevent or disrupt fibril formation failed in phase 3 trials.

ATTR AND OTHER HEREDITARY AMYLOIDOSES

The familial amyloidoses are autosomal dominant diseases in which mutated or variant plasma proteins misfold or aggregate to form beta-sheet rich amyloid deposits. These diseases are rare, with an estimated case incidence of <1/100,000 population in the United States, although founder effects in remote areas of Portugal, Sweden, and Japan produce a higher local prevalence of disease. The most prevalent form of hereditary amyloidosis arises from mutation of the abundant liver-derived plasma protein transthyretin (TTR, also known as *prealbumin*) and is termed ATTR variant (ATTRv) amyloid. More than 130 TTR mutations typically conferring one-amino-acid substitutions have been described, with most inducing clinical ATTR amyloid disease. Toxic TTR oligomers and ATTR amyloid deposits target peripheral and autonomic nervous systems and the heart. One TTR variant, V122I, occurs in nearly 4% of the African-American and Afro-Caribbean populations and is associated with late-onset cardiac amyloidosis. The actual incidence and penetrance of disease in the African-American population are the subject of ongoing research, but consideration of V122I ATTR amyloidosis is warranted in African-American patients who present with concentric cardiac hypertrophy and evidence of diastolic heart failure, particularly in the absence of a history of hypertension or valvular disease. Other familial amyloidoses, caused by variant apolipoproteins AI or AII, gelsolin, fibrinogen Aα, or lysozyme, are reported with lower prevalence worldwide. New amyloidogenic serum proteins continue to be identified periodically, including leukocyte chemotactic factor LECT2, which is a cause of renal amyloidosis in Hispanic and Pakistani populations. Although the clustering of ALECT2 cases suggests heritability, no LECT2 gene-coding sequence variations have been identified.

Normal (wild-type) transthyretin can also misfold and aggregate to form ATTR amyloid, principally expressed in men beginning in the seventh decade with increasing prevalence with age. Formerly termed senile systemic amyloidosis, ATTRwt amyloid is reported at autopsy in 25% of hearts from male patients who are 80 years and older. Although it is unclear why a wild-type protein becomes amyloidogenic, aging inefficiencies of intracellular quality-assurance mechanisms (termed the unfolded protein response) likely predispose to secretion of proteins prone to misaggregation. Due to the numbers of aging men globally, ATTRwt is the most prevalent and rapidly growing form of amyloidosis in the world today. Data to date characterize ATTRwt amyloidosis as a disease of aging, not inheritance.

Clinical Features and Diagnosis ATTRv amyloidosis has varied presentations predicted by the specific TTR mutation. Consequently, kindreds typically express similar disease timing and clinical course. Apparent sporadic presentations (no recognized family history) often reflect incomplete penetrance of the TTR mutation and not a spontaneous event. ATTRv amyloidosis presents as familial amyloidotic polyneuropathy (nerve damage) or familial amyloidotic cardiomyopathy (heart damage), although the majority of cases exhibit multiorgan disease. Peripheral neuropathy begins as a length-dependent small-fiber sensorimotor neuropathy first exhibited in the feet with ascending progression to the upper extremities. Autonomic neuropathy manifests as smooth muscle dysmotility (dysphagia, diarrhea, urinary retention), vascular dysregulation (orthostatic hypotension, erectile dysfunction), and anhidrosis. Soft tissue disease (carpal tunnel syndrome, tendinopathy, and spinal stenosis) commonly precedes nerve or heart manifestations of disease by one to two decades, particularly in ATTRwt amyloid patients who frequently report bicipital, patellar, or Achilles tendon rupture. Less common expressions of ATTRv include vitreous opacities and leptomeningeal amyloid deposition from variant protein produced by the retinal epithelium and choroid plexus, respectively. ATTR amyloid involvement of the heart is clinically better tolerated than AL amyloid cardiomyopathy as reflected by both the time from heart failure presentation to death in untreated cases of ATTR (median 42–48 months) versus AL (median 6 months) amyloidosis, and the dramatically greater burden of disease by echocardiographic measures at symptomatic presentation.

Typical syndromes associated with non-ATTR forms of hereditary (AF) disease include renal amyloidosis with mutant fibrinogen, lysozyme, or apolipoproteins; hepatic amyloidosis with apolipoprotein AI; and amyloidosis of cranial neuropathy with corneal lattice dystrophy pathognomonic of gelsolin (Finnish) amyloidosis. Patients with AF amyloidosis can present with clinical syndromes that mimic those of patients with AL disease. Rarely, AF carriers can develop AL disease or AF patients may have monoclonal gammopathy without AL. Thus, it is important to screen for plasma cell disorders and for protein mutations in patients with amyloidosis. Although mass spectrometry often detects amino acid sequence variations, it is not designed to definitively identify specific protein variations; DNA sequencing is the diagnostic standard for AF mutations.

TREATMENT

ATTR Amyloidosis

Untreated, survival after onset of ATTR disease is 4–15 years depending on whether the disease affects primarily the heart or nervous system, respectively. To date, therapeutic strategies used to control ATTR amyloidosis include: (1) orthotopic liver transplantation (OLT) to replace the factory of the mutated protein (only applicable to ATTRv); (2) stabilization of circulating TTR tetramers, preventing TTR monomer release and amyloid fibril formation; and (3) TTR gene silencing (RNA interference or antisense oligonucleotide agents), suppressing hepatic TTR production and subsequent ATTR fibril formation. After nearly 30 years as the principal treatment, OLT is now rarely employed, limited to patients with ATTRv amyloid, early peripheral neuropathy (V30M ATTR), and minimal systemic amyloid burden. Patients with more extensive amyloid (late V30M and non-V30M TTR mutations) who undergo OLT often suffer posttransplant disease progression due to allograft wild-type ATTR complexing on preexisting amyloid deposits. The TTR small-molecule thyroxine mimetic agents, diflunisal and tafamidis, bind to the kinetically stable tetrameric TTR conformation, limiting release and misfolding of monomeric protein, which is the critical step in TTR amyloidogenesis. International phase 3 randomized controlled trials demonstrate that TTR stabilizers slow but infrequently stop progression of ATTR polyneuropathy (diflunisal) and cardiomyopathy (tafamidis). TTR gene silencers (patisiran, inotersen, vutrisiran, eplontersen) more reliably halt neurologic disease progression by minimizing production of the amyloidogenic protein by the liver. Indeed, 35–60% of treated patients with familial amyloid polyneuropathy exhibit improved sensory nerve deficits, a novel finding. Therapeutic drug trials are underway to examine the safety, tolerability, and effectiveness of TTR gene silencers for ATTR cardiomyopathy. Preliminary data suggest TTR gene silencers may promote heart remodeling and improve systolic function in patients with wild type and variant ATTR amyloid cardiomyopathy.

Future clinical trials are set to examine the applicability of (1) one-time CRISPR/cas9 gene editing or (2) ATTR amyloid-depleting antibodies in patients with either ATTR polyneuropathy or cardiomyopathy. These antibodies are designed to recognize and bind nonnative (misfolded) TTR epitopes, mobilizing macrophages and monocytes to disrupt existing amyloid deposits. Whether disrupting amyloid deposits renews heart and nerve function will be determined by the outcome of these pivotal trials.

The extraordinary pace of drug development harnessing cutting-edge science in this orphan disease has extended survival and improved quality of life. Ironically, these advances expose previously unrecognized leptomeningeal (brain) and vitreous (eye) ATTR disease due to their occurrence late in disease, highlighting the unmet need for effective amyloid treatments that penetrate the blood-brain barrier.

Aβ$_2$M AMYLOIDOSIS

Aβ$_2$M amyloid is composed of β$_2$-microglobulin, the invariant chain of class I human leukocyte antigens, and produces rheumatologic manifestations in patients undergoing long-term hemodialysis and, rarely, in patients with a hereditary form of disease. β$_2$-Microglobulin is excreted by the kidney, and levels become elevated in end-stage renal disease. The molecular mass of β$_2$M is 11.8 kDa—above the cutoff of some dialysis membranes. The incidence of this disease appears to be declining with the use of newer membranes in high-flow dialysis techniques. Aβ$_2$M amyloidosis usually presents as carpal tunnel syndrome, persistent joint effusions, spondyloarthropathy, or cystic bone lesions. Carpal tunnel syndrome is often the first symptom. In the past, persistent joint effusions accompanied by mild discomfort were found in up to 50% of patients who had undergone dialysis for >12 years. Involvement is bilateral, and large joints (shoulders, knees, wrists, and hips) are most frequently affected. The synovial fluid is noninflammatory, and β$_2$M amyloid can be found if the sediment is stained with Congo red. Although less common, visceral β$_2$M amyloid deposits do occasionally occur in the gastrointestinal tract, heart, tendons, and subcutaneous tissues of the buttocks. There are no proven specific therapies for Aβ$_2$M amyloidosis, but cessation of dialysis after renal allografting may lead to symptomatic improvement.

SUMMARY

A diagnosis of amyloidosis should be considered in patients with unexplained nephropathy, cardiomyopathy (particularly with diastolic dysfunction), neuropathy (either peripheral or autonomic), enteropathy, or the pathognomonic soft tissue findings of macroglossia or periorbital ecchymoses. Pathologic identification of amyloid fibrils can be made with Congo red staining of aspirated abdominal fat or of an involved-organ biopsy specimen. Accurate typing by a combination of immunologic, biochemical, and genetic testing is essential in selecting appropriate therapy (Fig. 117-1). Systemic amyloidosis should be considered a treatable condition, as anti–plasma cell chemotherapy is highly effective in AL disease and targeted therapies are being developed for AA and ATTR disease. The combination of precursor and end-organ amyloid therapeutics potentially provides not only disease control but also functional and quality-of-life improvements for patients with amyloidosis. Tertiary referral centers can provide specialized diagnostic techniques and access to clinical trials for patients with these rare diseases.

FURTHER READING

Adams D et al: Efficacy and safety of vutrisiran for patients with hereditary transthyretin-mediated amyloidosis with polyneuropathy: A randomized clinical trial. Amyloid 30:1, 2023.

Coelho T et al: Eplontersen for hereditary transthyretin amyloidosis with polyneuropathy. JAMA 330:1448, 2023.

Griffin JM et al: ATTR amyloidosis: Current and emerging management strategies: JACC: CardioOncology state-of-the-art review. JACC CardioOncol 3:488, 2021.

Gustine JN et al: Predictors of hematologic response and survival with stem cell transplantation in AL amyloidosis: A 25-year longitudinal study. Am J Hematol 97:1189, 2022.

Kastritis E et al: Daratumumab-based treatment for immunoglobulin light-chain amyloidosis. N Engl J Med 385:46, 2021.

Maurer MS et al: Patisiran treatment in patients with transthyretin cardiac amyloidosis. N Engl J Med 389:1553, 2023.

Merlini G et al: Systemic immunoglobulin light chain amyloidosis. Nat Rev Dis Primers 4:38, 2018.

Staron A et al: Marked progress in AL amyloidosis survival: A 40-year longitudinal natural history study. Blood Cancer J 11:139, 2021.

118 Transfusion Therapy and Biology

Pierre Tiberghien, Olivier Garraud, Jacques Chiaroni

Transfusion encompasses the use of blood components (BCs) to prevent or treat anemia, hemorrhage, and bleeding disorders. Occasionally, BCs may be used to treat infection or relapse of malignant blood diseases after allogeneic hematopoietic transplantation. BCs comprise mainly red blood cell concentrates (RBCCs), platelet concentrates (PCs), and plasma for transfusion use as opposed to plasma fractionated into medicinal products (such as immunoglobulin, albumin, and clotting factors). Alongside transfusion safety, ensuring BC quality, assessing in vivo efficacy, and promoting evidence-based transfusion practices are critical aspects of transfusion medicine.

Donor medicine does not fall within the scope of this chapter. While particularly safe, blood donations can cause adverse reactions, among which are fainting reactions and iron deficiency. These reactions require preventive approaches and appropriate treatment when needed.

BLOOD COMPONENTS

BC collection and manufacturing processes are described in Table 118-1. BCs are collected as whole blood or directly as components by apheresis. The vast majority of BCs are homologous. Autologous BCs, collected ahead of planned surgery, are now exceptional as they present little to no advantage over homologous BCs. Nevertheless, such donation may still be of benefit in the presence of a rare blood group phenotype.

All BCs comply with common quality and performance standards and guidelines. Quality assurance encompasses well-defined processing steps and stringent BC quality controls as defined by health authorities. Reporting of adverse reactions and events associated with blood collection, BC processing, and transfusion is highly recommended.

With the obvious exception of granulocyte concentrates and mononuclear cells, most BCs are now leukocyte-reduced, and universal prestorage leukocyte reduction has been recommended. These BCs contain <1–5.10^6 donor leukocytes and are associated with reduced incidence of febrile nonhemolytic transfusion reactions (FNHTRs), infections with intracellular pathogens such as cytomegalovirus (CMV), alloimmunization, and immunomodulation.

BCs may undergo additional processing steps. These include irradiation to prevent graft-versus-host disease (GVHD) in immunosuppressed patients, pathogen reduction to further reduce the risk of transfusion-transmitted infections, plasma reduction in patients with severe allergic reactions to BCs, or the manufacturing of specific units for young children, neonates, or intrauterine transfusion.

BC constituents undergo centrifugation and filtration and are placed in contact with needles, plastic tubing, and bags, as well as anticoagulants and various additive solutions. BCs are subjected to gas exchanges that are significantly different from aerobic breathing and are maintained at temperatures that are not physiologic, such as 22°C or 4°C. Any of these elements may contribute to so-called "storage lesions" and to the presence of bioactive molecules such extracellular vesicles and cell-free mitochondrial DNA in the BC. The clinical impact of such lesions is still under investigation with currently no consensus on this issue.

Furthermore, plasma present in BCs contains donor antibodies (Abs). When directed toward antigens (Ags) present in the recipient, such as blood group or tissue (human leukocyte antigen [HLA]) Ags, such Abs may result in adverse events. RBCCs bring only a limited amount of donor plasma (10–30 mL), unlike PCs and obviously plasma. The use of platelet additive solution can replace two-thirds of plasma in PCs, while still leaving the equivalent of one plasma unit of 200 mL per transfused PC.

BLOOD GROUP ANTIGENS AND ANTIBODIES

Red blood cells (RBCs), as well as other blood constituents such as platelets and neutrophils, express allogeneic determinants. Transfusion may therefore result in alloimmunization and the production of alloantibodies (alloAbs). These alloAbs comprise anti-RBC Abs, anti-HLA, anti-human platelet Ag (HPA) Abs, and anti-human neutrophil Ag (HNA) Abs. Anti-RBC immunization may result in hemolysis, whereas anti-HLA or anti-HPA Abs may result in complications such as fever and platelet transfusion refractoriness. Furthermore, anti-HLA and anti-HNA immunization in the donor may result in a severe lung disorder called transfusion-related acute lung injury (TRALI). The Abs against red cell Ags may be IgM or IgG immunoglobulin classes. Some IgG or IgM can activate complement, and some IgG, crossing the placental barrier, may induce hemolytic disease of the fetus and newborn.

Erythrocyte blood groups refer to antigenic molecules that are expressed on the surface of RBC and other cells, genetically transmitted, and recognized by specific Abs. The polymorphism of such molecules explains their immunizing potential in situations such as transfusion, pregnancy, and transplantation. Blood groups can also interact with the environment and with infectious pathogens, leading to individual susceptibilities. For example, malaria is less severe in type O than non-O patients. Currently, ~390 different blood group Ags have been described, classified within ~45 different systems. Blood group Ags belong to two broad categories based on their biochemical nature: carbohydrate blood groups and protein blood groups. RBC Ags may be the target of autoantibodies (autoAbs) generating autoimmune hemolytic anemia. Some of them, mostly IgG, are active at 37°C, called "warm autoAbs," and are most often directed against Rh Ags, whereas others, most often IgM, are active at 4°C, called "cold autoAbs," and may be directed against ABO, HI, I, i, P, and other Ags.

Carbohydrate blood groups are headed by the **ABO system**, which comprises two main Ags, A and B, encoded by two alleles, which are the *A* and *B* alleles, respectively. In addition to these active alleles, there is an inactive allele: *O*. Depending on the genotype, four different phenotypes are produced (Table 118-2). Other carbohydrate systems (H, P1PK, Lewis, I, GLOB and SID) share many characteristics with the ABO system. The *A* allele encodes the A enzyme, which binds the A-type sugar (GalNac) A to the H substrate (expressed by action of the H enzyme encoded by the *H* allele, which happens to be inactive in the Bombay phenotype); sugars are attached to protein substrates on the surface of the RBC and so forth.

Carbohydrate Ags are ubiquitously distributed in the body. The ABO Ags, expressed on endothelial cells, are genuine "tissue" groups and may be involved in graft rejection. These Ags are not specific to humans but are shared by many species including viruses and bacteria. The presence of A and B Ags in the environment and, in particular, on the bacteria of the microbiota explains the synthesis of so-called "natural" or "regular" Abs, aside from any transfusion or pregnancy. Such Abs have a major hemolytic capacity as they bind complement and activate its cascade up to the membrane attack complex. This imposes donor-recipient stringent compatibility rules for RBCCs and whole blood transfusion and, albeit less stringently, for plasma and PC transfusion.

Protein blood groups are headed by the **Rh system** for RBCs (Table 118-3). As these Ags are specific to humans, the occurrence of immunization can only occur upon allogeneic stimulation. Abs directed against Ags of RBC groups other than ABO must be detected before RBCC transfusion or transplantation and during pregnancy. Of the 45 RBC group systems described, five (Rh, Kell, Duffy, Kidd, and MNS) are routinely investigated due to the clinical significance of Abs and their frequency. Testing for all five types ensures routine transfusion compatibility of 95%.

The Rh system comprises nearly 56 Ags, the most immunogenic of which is the D Ag (RH1). The Rh system has two $RH{*}D$ and $RH^{*}CE$ genes located on chromosome 1. The $RH^{*}D$ gene codes for the RhD protein expressing the D Ag (RH1) present in 85%, 93%, and >99%

TABLE 118-1 Blood Components: Collection and Manufacturing Processes

BLOOD OR APHERESIS COLLECTION AND INITIAL PROCESSING	BLOOD COMPONENT	ADDITIONAL COMPONENT PROCESSING (OPTIONAL TO MANDATORY)	RATIONALE	VOLUME AND CONTENT	STORAGE CONDITIONS AND DURATION
Whole blood collection: Separation into RBCC and platelet-rich plasma (PRP) by slow centrifugation, followed by high-speed centrifugation of the PRP to yield one unit of platelets (most often subsequently pooled) and one unit of plasma. *or* Separation into a RBCC, a plasma, and a “buffy coat” containing leukocytes and platelets by high-speed centrifugation, followed by pooling and slow-speed centrifugation of the buffy coat to produce a pooled platelet unit. Alternatively, the buffy coat may undergo high-speed centrifugation to produce a granulocyte unit that will be subsequently pooled. Apheresis collection: Various apheresis devices allow for the collection of BCs either as individual BCs such as plasma or PC (possibly double, such as double RBCC) or combined BCs, such as PC and plasma, or RBCC, platelets, and plasma.	RBCC from whole blood or apheresis	Leukocyte reduction Deleukocytation to <1–5.10^6 leukocytes per unit (highly recommended): initial whole blood filtration or RBC elective filtration (highly recommended)	Reduction of fever and chills Reduction of intracellular pathogens (including CMV) Reduction of alloimmunization	250–300 mL (including additive solution, no more than 40–50 mL of plasma) Hemoglobin: 22–40 g/dL Hematocrit: 50–70% Hemolysis ≤0.8% at issuing	4 +/– 2°C Duration depends on the additive solution: 25–42 days After irradiation: 24 h After plasma reduction: 24 h to 10 days depending on reduction methodology
		Irradiation: X-ray or gamma, ~25–35 Gy; most often units not older than 28 days after collection	GVHD prevention in immunosuppressed patients or intrafamily transfusions		
		Plasma reduction	Prevention of allergic reactions in patients with prior severe transfusion reactions	Lesser volume, 10% reduction in RBC content	
		Pediatric preparation	Adjustment to low-weight recipients	Adjusted content	
		Cryopreservation (glycerol)	Most often to ensure availability of RBCCs with a rare blood group for immunized “public-negative” recipients or recipients with complex alloimmunizations[a]	Same Hb content Hematocrit: 40–80% Glycerol ≤1 g	N2 or –80°C electric freeze drying N2: unlimited; –80°C: 30 years 7 days after thawing in suitable additive solutions, 24 h without additive solution
	PC from whole blood (individual units or pools of 4–6 units of ABO identical units) or from apheresis	Suspension in a platelet additive solution (PAS). PAS contains ingredients such as acetate, potassium, phosphate, and magnesium to sustain platelet storage	Reduction of fever and chills Plasma orientation toward fractionation	From 100 to 700 mL ≥2.10^{11} platelets pH ≥6.4	At 20–24°C and under permanent motion: 3–7 days *or* At 4°C without motion: up to 14–21 days If irradiated: <24 h
		Leukocyte reduction (<1–5.10^6 leukocytes per unit) (highly recommended): initial whole blood filtration or PC elective filtration	Reduction of fever and chills Reduction of intracellular pathogens (including CMV infections) Reduction of alloimmunization		
		Pathogen reduction: Most often nucleic acid cross-linker and/or UV illumination	Reduction of transfusion-transmitted infections Prevention of GVHD		
		Volume reduction	Prevention of allergic reactions in patients with prior severe reactions		
		Irradiation: X-ray or gamma, ~25–35 Gy; in general, on bags no older than 3 days after collection	Prevention of GVHD		
		Pediatric	Volume and content adjustment		
		Cryopreservation (DMSO)	To ensure continuous availability in remote locations To ensure availability of platelets with rare HPA groups		6 h after thawing (depending on cryopreservation procedure, may be resuspended in plasma)
	Plasma from whole blood or from apheresis	Cryopreservation at –18°C (most often)	Shelf-life extension	200–300 mL Coagulation factors, including fibrinogen (≥2 g/L), factor VIII (≥0.5 IU/mL), protein C and S, antithrombin	1–2 years if cryopreserved Up to 28 days if kept unfrozen

(*Continued*)

TABLE 118-1 Blood Components: Collection and Manufacturing Processes (Continued)

BLOOD OR APHERESIS COLLECTION AND INITIAL PROCESSING	BLOOD COMPONENT	ADDITIONAL COMPONENT PROCESSING (OPTIONAL TO MANDATORY)	RATIONALE	VOLUME AND CONTENT	STORAGE CONDITIONS AND DURATION
		Leukocyte reduction (<1–5.10^{6} leukocytes per product): Initial whole blood filtration and/or plasma elective filtration	Reduction of fever and chills Reduction of intracellular pathogens (including CMV) Reduction of alloimmunization		
		Pathogen reduction: Nucleic acid cross-linker and/or UV illumination or solvent detergent treatment (most often on pooled products)	Reduction of transfusion-transmitted infections		
		Lyophilization	To facilitate transportation and storage, as well as immediate availability, in remote locations		
	Granulocyte concentrates from whole blood (pools of 10–20 ABO-identical units) or from apheresis[b]	Irradiation (mandatory)	Prevention of GVHD	≤650 mL ≤2.10^{10} granulocytes	Room temperature ≤24 h after the end of collection
	Whole blood	Leukocyte reduction with a platelet-sparing device	Reduction of posttransfusion fever and chills Reduction of intracellular pathogens (including CMV) Reduction of alloimmunization	~520 mL (including additive solution)	At 2–4°C 21–35 days
	Peripheral blood mononuclear cells (apheresis)	May undergo cryopreservation (N2)	Increased practicability Repeated administration	Number of cells adjusted for a predetermined number of T lymphocytes 10^{5}–10^{7} CD3+ cells/recipient kg	N2: unlimited Never frozen or thawed: <6 h
	Cryoprecipitate (collected after thawing and centrifugation of plasma)	Resuspension in plasma (10–15 mL) and cryopreservation	N/A	Cold-insoluble plasma proteins (fibrinogen, factor VIII, von Willebrand factor)	12 months After thawing, may be stored at 20–24°C for up to 6 h

[a] Antigen frequency below 1% to 1/1000 of the population and contraindication for using regular blood units, depending on country-specific regulations. [b] Granulocyte collection by apheresis requires donor preadministration of steroids and/or hematopoietic growth factor and exposure to heparin and hydroxyethyl starch during the apheresis procedure.

Abbreviations: BC, blood component; CMV, cytomegalovirus; DMSO, dimethyl sulfoxide; GVHD, graft-versus-host disease; Hb, hemoglobin; HPA, human platelet antigen; N2, nitrogen gas; N/A, not applicable; RBC, red blood cell; RBCC, red blood cell concentrate; PC, platelet concentrate; UV, ultraviolet.

of individuals of Caucasian, African, and Asian ancestry, respectively. The *RH**CE* gene codes for RhCE proteins expressing C (RH2) and/or c (RH4), and E (RH3) and/or e (RH4) Ags. The presence of the D Ag confers Rh "positivity," while its absence confers Rh "negativity." The *RH**D* and *RH**CE* genes determine eight main haplotypes (*DCe, DcE, Dce, DCE, dce, dCe, dcE,* and *dCE*) whose frequencies differ considerably among different geographical populations. The high diversity of the Rh Ags includes weak and/or partial expression. Identifying individuals (especially young females of childbearing potential and multitransfused patients) with a weak or partial D Ag is important to adequately select D-positive or -negative RBCs. Molecular biology is now routinely applied to resolve such situations.

TABLE 118-2 ABO Blood Groups and Antibodies: Transfusion Compatibility

GENOTYPE(S)	ENZYME(S)/IMMUNODOMINANT SUGAR(S)	PHENOTYPE	NATURAL ANTIBODIES	TRANSFUSION COMPATIBILITY REQUIREMENTS		
				RBCC	PC[a]	PLASMA
A/A or A/0	"A" transferase/N-acetylgalactosamine (GalNac)	A	Anti-B	A or 0	A, 0[b], B[b], or AB[b]	A or A,B
B/B or B/0	"B" transferase/galactose (Gal)	B	Anti-A	B or 0	B, 0, A[b], or AB[b]	B or A,B
A/B	"A" transferase and "B" transferase GalNac and Gal	A,B	None	A,B or A or B or 0	A,B, 0[b], or A[b] or B[b]	A,B
0/0	Inactive Unconverted H antigen	0	Anti-A and Anti-B	0	0, A, B, or A,B	A or B or A,B or 0

[a] Order of priority. [b] Without high-titer anti-A and/or anti-B antibody.

Abbreviations: PC, platelet concentrate; RBCC, red blood cell concentrate.

TABLE 118-3 Red Blood Cell Group Systems and Antibodies: Clinical Significance and Transfusion Recommendations

ISBT NO./ SYSTEM	SYMBOL/GENE(S)	ANTIGENS (NO.)	MAIN ANTIBODIES (ANTI-)	HEMOLYSIS CHARACTERISTICS		RBCC TRANSFUSION RECOMMENDATIONS
				TRANSFUSION	HDFN	
1/ABO	ABO/*ABO*	4	A, B	None to severe; immediate and/or delayed	None to moderate (rarely severe)	Ab-negative RBCC
2/MNS	MNS/*GYPA, GYPB, (GYPE)*	49	M	None (except in extremely rare cases if active at 37°C)	None (except in extremely rare cases if active at 37°C)	Compatible RBCC (negative IAT at 37°C) Ag-negative red cells in the case of sickle cell disease
			N	None (may be clinically significant in the case of the rare N–S–s–U– phenotype)	None	Compatible RBCC (negative IAT at 37°C) Ag-negative RBCC in the case of N–S–s–U– phenotype
			S, s	None to moderate (rare)	None to severe (rare)	Ag-negative RBCC
			U	Mild to severe	Mild to severe (one reported case requiring an intrauterine transfusion)	Ag-negative RBCC
3/P1PK	P1PK/*A4GALT*	3	P1	None to moderate; delayed (rare)	None	Compatible RBCC (negative IAT at 37°C)
			P1, Pk, P (Tj^a)	None to severe	None to severe	Ag-negative RBCC
4/Rh	RH/*RHD, RHCE*	55	D, C, E, c, e	Mild to severe; immediate or delayed	Mild to severe	Ag-negative RBCC
6 /Kell	KEL/*KEL*	36	K	Mild to severe; delayed	Mild to severe (rare)	Ag-negative RBCC
7/Lewis	LE/*FUT3*	6	Le^a, Le^b	None (rare cases of hemolytic reactions)	None	Compatible RBCC (negative IAT at 37°C)
8/Duffy	FY/*ACKR1*	5	Fy^a, Fy^b	Mild to severe (rare); immediate/delayed	Mild to severe (rare)	Ag-negative RBCC
			Fy3, Fy5	Mild to moderate; immediate (rare)/delayed	Mild (rare) (no data for anti-Fy5)	Ag-negative RBCC
9/Kidd	JK/*SLC14A1*	3	Jk^a, Jk^b	None to severe; immediate or delayed	Mild to moderate (rare)	Ag-negative RBCC
			Jk3	None to severe; immediate or delayed	None to mild	Ag-negative RBCC
18/H	H/*FUT1*	1	H (Bombay)	None to severe; immediate/ delayed	Not none	Ag-negative RBCC
20/Globoside	GLOB/*B3GALNT1*	2	P	None to severe	None to mild	Ag-negative RBCC

Abbreviations: Ab, antibody; Ag, antigen; HDFN, hemolytic disease of the fetus and newborn; IAT, indirect antiglobulin test (indirect Coombs test); ISBT, International Society of Blood Transfusion; RBCC, red blood cell concentrate.

The **Kell system** comprises 38 Ags, one of which is routinely determined: the K antigen (KEL1); 9% and 2% of individuals of Caucasian and African ancestry are K positive (KEL:1), respectively, whereas 91% and 98%, respectively, are K negative (KEL:–1). The immunogenicity of Kell is third behind the ABO and Rh systems. The Kell protein is linked to another blood group protein called Kx. The rare absence of this protein (controlled by a gene on X) is associated with a weak KEL Ag, acanthocytosis, shortened RBC survival, and a progressive form of muscular dystrophy that includes cardiac defects. This rare condition is called the McLeod phenotype.

The **Duffy system** (FY) comprises five Ags, two of which are routinely tested: the Fy^a Ag (FY1), coded by the *Fy*^{a} allele, and the Fy^b Ag (FY2), coded by the *Fy*^{b} allele. Depending on the combination of alleles, three common phenotypes are expected: Fy (a+b+), which has the two alleles *Fy*^{a} and *Fy*^{b}; Fy (a+b–), which has only the *Fy*^{a} allele in a double dose; and Fy (a–b+), which has only a double dose of the Fy*b* allele. A particular phenotype characterized by the absence of the Fy^a and Fy^b Ags, the Fy(a–b–) phenotype, is exclusive (with some exceptions) to individuals of African ancestry where it can reach frequencies of 70–100% depending on the population. It is linked to the presence of a double dose of a silent *FY***0* allele. This distribution may be related to the fact that the Fy Ags serve as receptors for *Plasmodium vivax* and therefore the Fy(a–b–) phenotype. However, these individuals may develop Abs against two high-frequency Ags (FY3 and FY5) after transfusion or pregnancy. They may also have low a granulocyte count but are not associated with any disease.

The **Kidd system** (JK) comprises three Ags, two of which are routinely tested: the Jk^a Ag (JK1), coded by the *Jk*^{a} allele, and the Jk^b Ag (JK2), coded by the *Jk*^{b} allele. Depending on the combinations of alleles, three common phenotypes are seen: Jk(a+b+) displaying the two alleles *Jk*^{a} and *Jk*^{b}, Jk(a+b–) displaying only the *Jk*^{a} allele in a double dose, and Jk(a–b+) displaying only a double dose of the *Jk*^{b} allele. A particular phenotype is characterized by the absence of the Jk^a and Jk^b Ags: the Jk(a–b–) phenotype found in Polynesian populations. It is linked to the presence of a double dose of a silent *JK***0* allele. These people may develop Abs against the high-frequency anti-JK3 Ag after transfusion or pregnancy.

The **MNS system** comprises 50 Ags, four of which are routinely tested. Two genes (*GYPA, GYPB*) encode two pairs of so-called "antithetical" Ags. The M (MNS1) and N (MNS2) pair Ags encoded by the *M* and *N* alleles, respectively, are branched on the glycophorin A molecule. Their combination will determine whether or not they are present. M+ and N+ subjects have both alleles; an M+, N– subject is homozygous for the *M* allele; and an M–, N+ subject is homozygous for the *N* allele. The same holds true for the other pair of Ags, S (MNS3) and s (MNS4) expressed on glycophorin B. Therefore, an M+, N–, S–, s+ subject (in international nomenclature, this is written as MNS:1,–2,–3,5) will be homozygous for the *M* and s alleles. A rare phenotype, S–s–, found exclusively in individuals of African ancestry, can develop an Ab against the high-frequency U Ag (MNS5) after transfusion or pregnancy.

■ RARE RBC PHENOTYPES

Some patients present with rare genotype/phenotype assortments, and their RBCs display so-called private Ags or, conversely, lack public Ags (i.e., widely shared Ags) toward which the patient may develop an immune response when exposed to these Ags. Public-negative immunized individuals are virtually impossible to transfuse using conventional blood bank resources and require access to designated blood banks that have access to rare blood programs. Their primary responsibility is to identify and collect blood from donors exhibiting particular Ag displays on their RBCs or platelets that are uncommon in the given jurisdiction. Specific ethnic populations may be targeted, as some may display genotype specificities, such as the Bombay group in southwestern Indians. Several hemoglobinopathies, such as sickle cell disease, are more common in individuals of African ancestry. Such patients may display RBC phenotypes that are uncommon in countries in the Northern Hemisphere, resulting in difficulties adequately identifying donors to match the need, as a last resort, for highly valued cryopreserved BCs.

CLINICAL INDICATIONS AND EFFICACY ASSESSMENT OF BLOOD COMPONENTS

BCs are life-saving therapies but also scarce resources. Furthermore, transfusion may result in well-identified adverse reactions as well as more ill-defined adverse reactions, including inflammation and therapeutic inefficacy. As highlighted in patient blood management programs, transfusion should be considered within a multidisciplinary approach that includes optimization of hematopoiesis and minimization of blood loss during surgical interventions. Clinical indications of BCs as well as means to assess therapeutic efficacy are detailed in **Table 118-4**.

ADVERSE REACTIONS TO BLOOD COMPONENTS

Adverse reactions to transfused BCs are most commonly non-life-threatening, although serious reactions can present with mild symptoms. Transfused patients should be closely monitored for warning signs suggestive of adverse reactions, as described in **Table 118-5**. When an adverse reaction is suspected, the transfusion must be stopped while the recipient's clinical status is assessed, and supportive care is initiated as needed. An average of 37 transfusion-associated fatalities with possible to definite imputability were reported yearly to the U.S. Food and Drug Administration (FDA) between 2017 and 2021 among ~15 million transfused BCs. Most frequent causes of death were transfusion-associated circulatory overload (TACO) (32%), followed by TRALI (21%), hemolysis (21%), and microbial contamination (13%).

Adverse reactions to BCs may result in immune and nonimmune mechanisms. Immune-mediated reactions are often due to recipient or donor alloimmunization and the presence of preformed recipient or donor Abs. Nonimmune causes of reactions are from the physical or chemical properties of BCs or from pathogens present in the BC.

■ IMMUNE-MEDIATED ADVERSE REACTIONS

Hemolytic Transfusion Adverse Reactions Immune-mediated acute hemolysis occurs when the recipient preformed Abs lyse transfused donor RBCs and may occur during or 24 h after transfusion. The anti-A or anti-B Abs are responsible for most of the most severe reactions, which can be fatal. However, alloAbs directed against other RBC Ags (i.e., Rh, Kell, and Duffy) are also responsible for severe hemolytic reactions. Such dramatic reactions are usually caused by a failure in product or patient identification, erroneous blood grouping, or unidentified anti-RBC alloimmunization in the recipient. Hemolysis, most often of lesser severity, may also occur upon transfusion of BCs containing incompatible plasma with a large amount of alloAbs directed against the recipient's RBCs. This may typically occur after transfusion of a PC containing ABO-incompatible plasma. Estimated frequencies of acute and chronic hemolytic adverse reactions are 1–10 and 5–40 per 10^5 transfused BCs, respectively.

Mechanisms of transfusion hemolytic reactions are described in **Fig. 118-1.**

Prevention of hemolytic reactions relies on **pretransfusion testing** of potential recipients. Testing will include determination of the ABO D phenotype (and anti-ABO Abs) as well as additional typing for the other main Rh Ags (CcEe), K Ag of the Kell system, and more rarely, Duffy (Fy^a and Fy^b), Kidd (Jk^a and Jk^b), and MNS (S and s) Ags, depending on the clinical setting. These determinations are most often performed by serology. However, molecular typing is increasingly being used to predict RBC phenotype and facilitate the selection of a compatible component. Special care must be taken to verify the patient's identity and apply adequate tube labeling. A double ABO determination performed separately may be considered, especially in the absence of a systematic crossmatch.

Testing will also include the screening and identification of alloAbs directed against RBC Ags other than ABO. This screen is performed by mixing patient serum with type O RBCs expressing Ags from most blood group systems and whose extended phenotype is known. The specificity of the alloAb is identified by correlating the presence or absence of Ag with the induced—or not—agglutination. Special attention should be paid to patients receiving monoclonal Ab treatment that may bind to erythrocytes in vivo (such as anti-CD38 IgG treatment for multiple myeloma) and therefore interfere with alloAb screening. Such interference may be offset by sample dithiothreitol pretreatment.

Crossmatching between the recipient plasma/serum and the selected RBCs may be performed, especially when the recipient is alloimmunized against RBC or is frequently transfused, as well as in specific clinical settings such as sickle cell disease, even if the Ab screening is negative.

The selection of a compatible BC should consider pretransfusion testing as well as the recipient's clinical status. In the case of D-negative patients, every attempt must be made to provide Rh-negative RBCC to prevent anti-D alloimmunization. In an emergency, D-positive RBCC can be safely transfused to a D-negative patient who lacks anti-D. However, an estimated 20–22% of RBCC recipients will become alloimmunized and produce anti-D Abs after transfusion with D-positive RBCs. Such alloimmunization can occur after PC transfusion, although at a much lower frequency (~1%). Whenever possible, females with childbearing potential (to include prepubertal girls) should be transfused with D- and K-compatible RBCCs and D-compatible PCs to prevent alloimmunization and protect a future fetus/newborn from an alloimmune-mediated hemolytic disease. D-negative females with childbearing potential who are transfused with BCs containing D-positive RBCs should receive anti-D Ab to prevent allosensitization.

Hemolysis, most often of lesser severity, may also occur after transfer of alloAbs directed against the recipient's RBC Ags. Such ABO "plasmatic" incompatibility, called "minor ABO incompatibility," will occur mainly with PC transfusions, where platelets are suspended in ~100–300 mL of plasma (depending on whether part of the plasma is substituted by additive solution). BCs containing plasma with high-titer anti-A/B Ab may induce a hemolytic reaction. When the transfusion of ABO-identical (vs ABO-compatible) PCs is not feasible, PCs provided by donors with low-titer anti-A/B only should be preferred. While there is no universal definition of high-titer Abs, a threshold titer of 1/64 (as assessed by hemagglutination) may be appropriate. The use of an additive solution in PCs substantially mitigates this risk. Lastly, ABO plasmatic incompatibility can lead to the formation of immune complexes with soluble A and/or B Ags and ensuing inflammation and platelet activation.

Acute hemolytic reactions may present with hypotension, tachypnea, tachycardia, fever (+1–2°C), chills, chest and back pain, hemoglobinuria, and hemoglobinemia. In the most severe cases, disseminated intravascular coagulation (DIC), acute renal failure, shock, and death may occur.

Delayed hemolytic reactions, with icterus and persisting or worsening anemia as the main clinical manifestations, result from an anamnestic response. Such reactions may occur in patients previously sensitized to RBC Ags who have a negative alloAb screen at the time of transfusion due to low Ab levels. The alloAb is detectable 1–2 weeks after the transfusion.

Diagnosis of transfusion-associated hemolysis relies on persistent and/or worsening anemia, depleted plasma haptoglobin levels,

TABLE 118-4 Blood Components: Clinical Use

COMPONENT	THERAPEUTIC INDICATION	GOAL	DONOR/RECIPIENT COMPATIBILITY	DOSAGE	EFFICACY EVALUATION
RBCC	Transfusion: Anemia and/or tissue ischemia (treatment or prevention) Hb below a given threshold (to be considered in relation with clinical symptoms): <7 g/dL for patients hemodynamically stable, except for patients undergoing orthopedic surgery, cardiac surgery, or with preexisting cardiovascular disease (<8 g/dL) as well as for patients with acute coronary disease (<9–10 g/dL). Such thresholds do not apply to neonates and patients with severe thrombocytopenia and chronic transfusion-dependent anemia. Not recommended: nutritional anemia (iron, vitamin B_{12}, or folate deficiency)	Improve systemic and tissue oxygenation	ABO compatible (cellular) and ABO identical when achievable. RhD compatibility in young and childbearing-age females, and if multitransfused RhC/c/E/e; Kell-compatible RBCCs if multitransfused whenever possible. Additional compatibility may be required depending on the clinical setting and screening results.	1 unit at a time (250–350 mL, including additive solution), repeated per clinical status and Hb level	Reduction of anemia-related symptoms, clinical improvement Increased Hb (+1 g/dL) and hematocrit (+3%)
	RBC exchange: Anemia/sickle cell crisis in hemoglobinopathies (sickle cell disease, thalassemia)	Replace altered RBCs with donor RBCs and compensate for hemolysis, prevention of sickle cell occlusive crisis		25–30 mL/kg	Sickle cell disease: reduced percentage of HbS
PC (from pooled whole blood–derived platelets or single donor apheresis), maintained at room temperature (most often) or at 4°C	Thrombocytopenia-related bleeding disorders: Treatment (cold or room temperature PC) or prevention (room temperature PC) Platelet level below a given threshold: ≤5000/μL in the absence of fever or infection, ≤10,000/μL to 20,000/μL if fever or infection; ≤50,000/μL if surgery, DIC, endoscopy, invasive procedures; ≤80,000/μL if neurosurgery, eye surgery, or ECMO Acute hypovolemic coagulopathy (see below) Immune thrombocytopenia, thrombotic microangiopathy, and heparin-induced thrombocytopenia: Not recommended	Correct impaired primary hemostasis, including vessel healing Cold stored platelets, despite lower in vivo survival, have maintained and possibly improved hemostatic capacity compared with room temperature stored platelets	ABO identical preferable; if not, ABO compatible (cellular) with low-titer anti-A/B Ab; RhD compatible preferred in premenopausal women HLA compatible (negative lymphocyte crossmatch) or HLA identical in case of refractoriness related to the presence of anti-HLA Ab HPA compatible in thrombocytopenic neonates to HPA immunized mother (fetal neonatal alloimmune thrombocytopenia)	0.5–0.7 × 10^{10} platelets/kg (apheresis or pooled whole blood–derived PCs)	Prevention and/or resolution of bleeding CCI[a] >7.5 to 10 within 1 h and > 4.5 to 5 within 24 h after transfusion (not applicable to cold/cryopreserved platelets)
Plasma (thawed frozen, never frozen and maintained at 4°C or at room temperature, freeze-dried)	Transfusion: Coagulation factor–related bleeding disorders Acute hypovolemic coagulopathy (see below) Infectious disease (convalescent plasma containing pathogen-specific Abs): Argentina hemorrhagic fever, viral respiratory infections (experimental)	Correct impaired hemostasis by providing missing elements of coagulation or fibrinolysis cascade, as well as elements to heal injured vessel endothelium Provide Abs against relevant pathogens	ABO compatible (plasma)	10–15 mL/kg Not determined	Reduced bleeding disorder Infection resolution
	Plasma exchange (plasma or combined plasma and albumin): Pathogenic Ab removal and supplementation of lacking enzyme (e.g., thrombotic thrombocytopenic microangiopathy or Guillain-Barré syndrome) Pathogenic Ab removal (e.g., anti-HLA Ab prior to kidney transplantation)	Deplete pathogenic elements in the blood (autoantibodies such as anti-ADAMTS-13 Ab in case of TTP, excess cholesterol, etc.); plasma may also bring anti-inflammatory and/or immunomodulatory factors such as immunoglobulin	ABO compatible (plasma)	45–60 mL/kg	Improved disease-specific symptomatology (i.e., apyrexia and platelet recovery in case of TTP) Reduced antibody levels (e.g., anti-HLA antibodies prior to organ transplantation)

(*Continued*)

Whole blood	Acute hypovolemic coagulopathy requiring massive transfusion	Balanced provision of blood components maintained at 4°C and without an additive solution and related dilution	ABO-identical or group O with low-titer anti-A/B Ab	Repeated per clinical status	Normovolemia; bleeding resolution
Multicomponent (RBCC, PC, and plasma)	Acute hypovolemic coagulopathy requiring massive transfusion	Appropriate ratio is under investigation; a ratio of 1 RBCC/1 plasma/0.25 PC (platelet content of a whole blood) is currently favored	Standard RBCC, PC, and plasma compatibility	1 RBCC/1 plasma/0.25 PC ratio, repeated per clinical status	Normovolemia; bleeding resolution
Granulocyte concentrates (apheresis or a pool of whole blood–derived granulocytes)	Severe refractory bacterial or fungal infection (mainly soft tissues and lung) in patients with neutropenia (<100/μL) or with dysfunctional granulocytes (CGD). Neutropenia can be acquired (chemotherapy) or congenital. Formal proof of efficacy is lacking.	Correct impaired granulocyte function in relation to granulocytopenia or granulocyte dysfunction	ABO compatible	1–2 × 10^{10}, repeated per clinical status	Infection resolution (or stabilization until recovery from neutropenia)
Donor mononuclear cells	Relapse of malignant hemopathy after allogeneic hematopoietic cell transplantation	Graft-versus-leukemia effect (and graft enhancement effect)	N/A	10^5–10^7 T lymphocytes/kg	Disease specific (remission)
Cryoprecipitate	Acute bleeding coagulopathy, type II (dysfunctional factor) or type III (absent factor) von Willebrand disease, hemophilia A in the absence of factor VIII concentrates	Provision of fibrinogen, factor VIII, von Willebrand factor, and factor XIII	ABO compatibility is not required	10–15 mL/unit, pool of 4–5 units	Increased plasma fibrinogen (0.3–1 g/L)

[a] CCI calculation:

$$\text{CCI} = \frac{\text{Postransfusion count } (/\mu\text{L}) - \text{pretransfusion count } (/\mu\text{L})}{\text{Number of platelets transfused x } 10^{11}} \text{ x Body surface area } (\text{m}^2)$$

Abbreviations: Ab, antibody; CCI, corrected count increment; CGD, chronic granulomatous disease; DIC, disseminated intravascular coagulation; ECMO, extracorporeal membrane oxygenation; Hb, hemoglobin; HLA, human leukocyte antigen; N/A, not applicable; RBC, red blood cell; RBCC, red blood cell concentrate, PC, platelet concentrate; TTP, thrombotic thrombocytopenic purpura.

TABLE 118-5 Transfusion Adverse Reactions: Main Warning Signs

Fever (≥38°C)	+1–2°C within 4 h	FNHTR Anti-HLA immunization and cognate Ag in the blood product TRALI (with dyspnea at the forefront)
	+1–2°C within 15 min +/–: • Chills • Dyspnea • Hypotension • Digestive disorders • Disseminated intravascular coagulation • Hemoglobinuria	Transfusion-transmitted bacterial infection Hemolysis
	>2°C	Transfusion-transmitted bacterial infection
Hypotension (≥30 mmHg decrease in systolic blood pressure)		Hemolytic shock Anaphylactic shock Septic shock TRALI (with dyspnea at the forefront)
Dyspnea		TRALI (within 6 h of transfusion) TACO (within 12 h of transfusion) Severe allergy (immediate; within 4 h)
Hemoglobinuria		Intravascular hemolysis • Immunologic • Mechanical • Toxic • Thermic
Rash	<2/3 of the body within 2–3 h	Minor allergy
	>2/3 of the body during or within 2–3 h	Severe allergy
	>2/3 of the body within 5 min Associated with dyspnea and shock	Anaphylaxis
Icterus		Delayed hemolysis
New alloantibody		Alloimmunization
Rash, diarrhea, and fever occurring 2 days to 6 weeks after transfusion		GVHD
Gum bleeding, purpura 5–12 days after transfusion		Posttransfusion purpura
Cardiac, hepatic, and/or renal insufficiency in frequently transfused patients		Posttransfusion iron overload
Top-down investigation after a blood donor is subsequently found to be infected Bottom-up investigation after another recipient of a same blood donation is found to be infected Infectious symptoms within 6 months		Transfusion-transmitted infection

Abbreviations: Ag, antigen; FNHTR, febrile nonhemolytic transfusion reaction; GVHD, graft-versus-host disease; HLA, human leukocyte antigen; TACO, transfusion-associated circulatory overload; TRALI, transfusion-related acute lung injury.

hemoglobinemia and hemoglobinuria, and elevated plasma lactate dehydrogenase and unconjugated bilirubin. The direct antiglobulin test (DAT, or direct Coombs test) that detects immunoglobulin, and possibly complement (C3d), on the surface of the recipient's RBC will most often be positive **(Fig. 118-2)**. Similarly, a positive indirect antiglobulin test (IAT, or indirect Coombs test) that detects anti-RBC alloAb in the serum will also be positive. An elution of the Ab on the surface of the RBC may allow for the identification of the culprit alloAb.

The management of an immune-mediated acute hemolytic transfusion reaction is mainly supportive. Prompt interruption of the transfusion, biological workup, and a thorough clerical check to prevent a possible second misidentified transfusion are crucial initial steps. Vigorous hydration with isotonic saline and diuretics to maintain urine output is recommended. Although often self-limiting, acute hemolysis may also require forced alkaline diuresis, correction of electrolyte abnormalities, and pressor support as needed. In patients with DIC and severe bleeding, PC, plasma, and cryoprecipitate or fibrinogen may be required. When transfusion of incompatible RBCCs is unavoidable, prophylaxis with steroids (100 mg of hydrocortisone) just before the transfusion and repeated 24 h later and polyvalent immunoglobulin (1.2–2.0 g/kg per day over 2–3 days, initiated just before the transfusion) have been successfully used to prevent or minimize acute and delayed hemolysis. Polyvalent immunoglobulin, anticomplement (C3) Ab, anti-B-cell Ab, or plasma exchange may be considered in case of severe posttransfusion hyperhemolysis in sickle cell patients.

Immune-mediated hemolysis may also occur after allogeneic hematopoietic transplantation (most often involving a peripheral blood stem cell graft) or, more seldomly, solid organ transplantation. Minor ABO incompatibility, with subsequent red cell destruction in the recipient, is the most common cause of clinically significant hemolysis in such cases. Viable donor B lymphocytes, called "passenger lymphocytes," transferred passively with the graft, may produce alloAbs (including anti-D or anti-A1 in an A2 donor) that target recipient red cells. Such hemolysis has been reported to develop 5–14 days after transplantation. Reduced-intensity conditioning regimens and cyclosporine as prophylaxis against GVHD or rejection have been associated with increased risk. Transfusing RBCs compatible with the graft donor and the use of GVHD prophylaxis able to target B cells (e.g., methotrexate) have significantly reduced the incidence of passenger lymphocyte syndrome. Allogeneic hematopoietic transplantation may also result in acute hemolysis due to incompatible donor-derived red cell (and precursor) destruction by the recipient alloAbs (i.e., major ABO incompatibility). Prolonged pure red cell aplasia may occur in such a situation. Graft deserythrocytation will reduce the risk of early acute hemolysis.

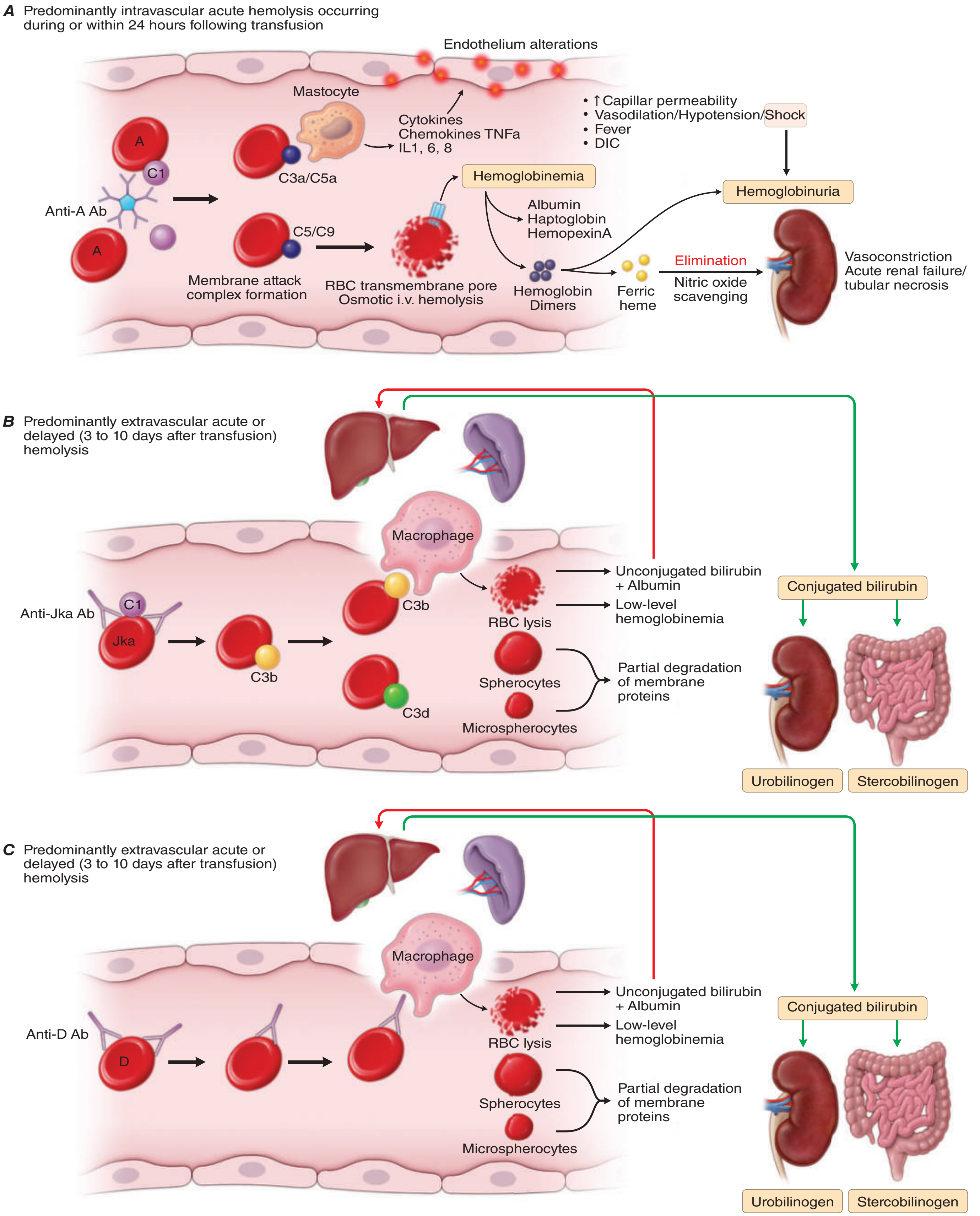
A Predominantly intravascular acute hemolysis occurring during or within 24 hours following transfusion
Endothelium alterations
Mastocyte
Cytokines
Chemokines TNFa
IL1, 6, 8
• ↑Capillar permeability
• Vasodilation/Hypotension/Shock
• Fever
• DIC
A
C1
Anti-A Ab
A
C3a/C5a
C5/C9
Membrane attack complex formation
Hemoglobinemia
RBC transmembrane pore
Osmotic i.v. hemolysis
Albumin
Haptoglobin
HemopexinA
Hemoglobin Dimers
Ferric heme
Elimination
Nitric oxide scavenging
Hemoglobinuria
Vasoconstriction
Acute renal failure/
tubular necrosis
B Predominantly extravascular acute or delayed (3 to 10 days after transfusion) hemolysis
Macrophage
Anti-Jka Ab
C1
Jka
C3b
C3b
C3d
RBC lysis
Spherocytes
Microspherocytes
Unconjugated bilirubin + Albumin
Low-level hemoglobinemia
Partial degradation of membrane proteins
Conjugated bilirubin
Urobilinogen
Stercobilinogen
C Predominantly extravascular acute or delayed (3 to 10 days after transfusion) hemolysis
Macrophage
Anti-D Ab
D
RBC lysis
Spherocytes
Microspherocytes
Unconjugated bilirubin + Albumin
Low-level hemoglobinemia
Partial degradation of membrane proteins
Conjugated bilirubin
Urobilinogen
Stercobilinogen

FIGURE 118-1 Mechanisms of transfusion hemolytic reactions. ***A.*** Acute responses will involve preexisting antibodies (Abs), naturally occurring anti-A/anti-B IgM or IgG directed against other RBC antigens (Ags) and resulting from prior sensitization. Upon interaction with cognate antigen on transfused red blood cells (RBCs), recipient allogeneic Ab (alloAb), mostly natural anti-A/anti-B IgM, may fix and activate complement up to C5/C9. Formation of membrane attack complex (MAC) will create pores in transfused RBCs with resulting intravascular hemolysis, release of toxic moieties including free hemoglobin responsible for end-organ damage including renal failure, and tissue factors contributing to occurrence of disseminated intravascular coagulation (DIC). ***B.*** Alternatively, complement activation may be incomplete, as typically observed in a delayed hemolytic transfusion reaction involving neoformed allogeneic IgG. In such cases, complement activation up to C3 results in C3b-mediated opsonization of RBCs, extravascular hemolysis, and clearance through immunophagocytosis. Anemia and jaundice will be the primary clinical manifestations. ***C.*** Lastly, alloAb may not fix complement while ensuring antibody-dependent cellular cytotoxicity (ADCC)–mediated phagocytosis of targeted RBC. *(Adapted from SR Panch et al: Hemolytic transfusion reactions. N Engl J Med 381:150, 2019.)*

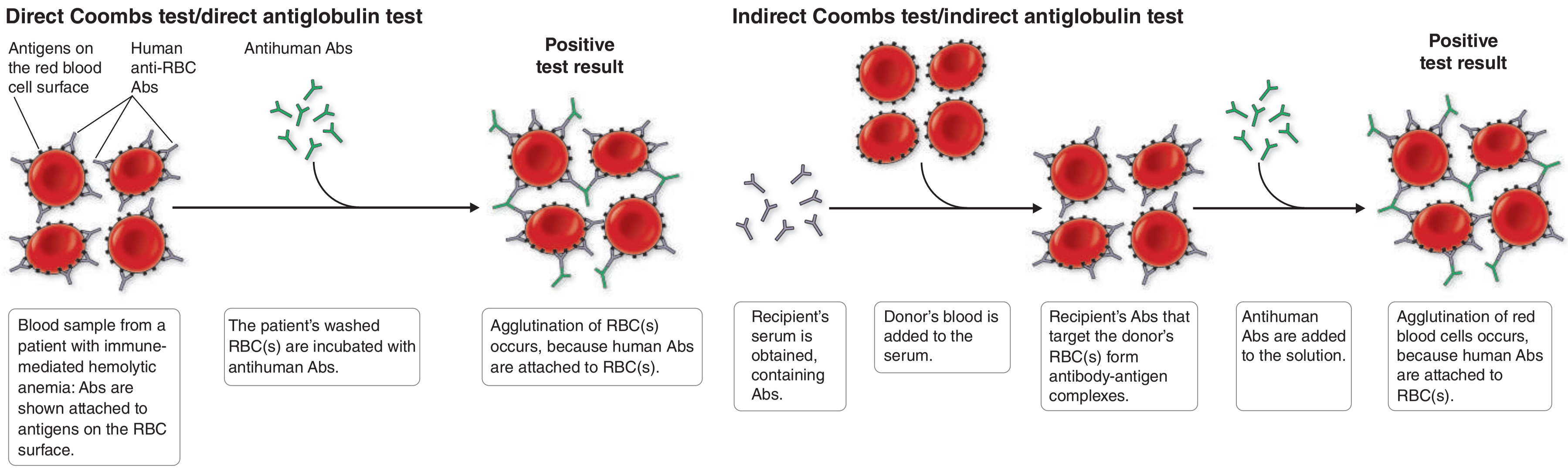
Direct Coombs test/direct antiglobulin test
Antigens on the red blood cell surface
Human anti-RBC Abs
Antihuman Abs
Positive test result
Blood sample from a patient with immune-mediated hemolytic anemia: Abs are shown attached to antigens on the RBC surface.
The patient's washed RBC(s) are incubated with antihuman Abs.
Agglutination of RBC(s) occurs, because human Abs are attached to RBC(s).
Indirect Coombs test/indirect antiglobulin test
Positive test result
Recipient's serum is obtained, containing Abs.
Donor's blood is added to the serum.
Recipient's Abs that target the donor's RBC(s) form antibody-antigen complexes.
Antihuman Abs are added to the solution.
Agglutination of red blood cells occurs, because human Abs are attached to RBC(s).

FIGURE 118-2 Direct and indirect Coombs test. The direct Coombs/antiglobulin test detects the presence of Abs (or complement) on the surface of erythrocytes. The indirect Coombs/antiglobulin test detects Abs in the serum that may bind to donor erythrocytes. Abs, antibodies; RBC, red blood cell. *(Adapted from http://upload.wikimedia.org/wikipedia/commons/1/1c/coombs_test_schematic.png.)*

Polyvalent immunoglobulin may contain high titers of anti-A (mostly) and/or anti-B Abs and induce acute hemolysis, most often of limited severity. Such hemolysis is particularly described in group A or AB children receiving high-dose immunoglobulin, notably for Kawasaki's disease, as well as in adults treated for thrombotic thrombocytopenic purpura. A similar mechanism may lead to hemolysis after anti-D immunoglobulin treatment for immune thrombocytopenia in RhD-positive patients.

Nonimmune mechanisms of transfusion-associated hemolysis include thermal (overheated or cold BCs), osmotic (concurrent hypo-osmotic perfusion), and mechanical (pressure related to high-flow transfusion filtering during cell saver processing) mechanisms.

Autoimmune and drug-induced hemolytic anemias may be exacerbated by transfusion and can therefore mimic hemolytic transfusion reactions. Transfusion of RBCs with enzymatic defects may mimic immune-mediated hemolysis as well. Notably, severe hemolytic reactions in patients receiving long-term transfusions for hemoglobinopathies (mainly sickle cell disease) can precipitate bystander hemolysis, in addition to clearing transfused red cells. The mechanisms of this hyperhemolytic transfusion reaction may be a mediated RBC hemolysis-related systemic inflammatory response and resulting lysis of red cell precursors by macrophages. This process may be immediate or delayed, with hemoglobin levels falling below the pretransfusion values, often to life-threatening levels. Further RBCC transfusion typically exacerbates ongoing hemolysis, with the exogenous (transfused) allogeneic Ags probably triggering further nonspecific hemolysis.

Febrile Nonhemolytic Transfusion Reaction The most frequent reaction associated with the transfusion of cellular BCs is FNHTR (up to 300 per 10^5 BCs). This reaction is characterized by chills and rigors and a ≥1°C rise in body temperature and is caused by proinflammatory cytokines in the BC or by recipient Abs directed against donor cell Ags present in the BC. FNHTR is diagnosed when other causes of fever, notably infection and hemolysis, have been excluded. Leukocyte reduction, especially prestorage, can prevent the occurrence of FNHTR. Moreover, the use of additive solutions decreases FNHTR frequency associated with PC transfusion. Premedication with antipyretics has generally proven ineffective at decreasing the rate of such reactions and may mask relevant clinical symptoms.

Allergic Reactions Most allergic transfusion reactions are mild and include rash, pruritus, urticaria, and localized edema. More rarely, allergic reactions may be severe to life-threatening with an anaphylactic reaction that can involve bronchospasm, respiratory distress, hypotension, nausea, vomiting, and shock. Frequencies of mild and severe allergic reactions are ~100 and ~5 per 10^5 BCs, respectively. Allergic reactions are related to plasma proteins found in transfused components. Mild reactions may be treated by temporarily stopping the transfusion and administering antihistamine drugs. Patients with a history of allergic transfusion reaction may be premedicated with an antihistamine, although there is no consensus on this issue. Cellular components can be washed to remove residual plasma for extremely sensitized patients. Most of the allergic presentation may not depend on preformed Abs and may be attributable to soluble mediators triggering histamine and serotonin release from platelets and leukocytes. An anaphylactic reaction may occur after the transfusion of only a few milliliters of the BC. Treatment includes stopping the transfusion, maintaining vascular access, and administering adrenaline (0.3–0.5 mg subcutaneously). Additional treatment with steroids, antihistamine drugs, and bronchodilators may also be required.

Patients who are IgA deficient (<1% of the population) may be sensitized to this immunoglobulin isotype and may be at risk of anaphylactic reactions associated with plasma transfusion. As a precaution, individuals with severe IgA deficiency should therefore receive, where available, IgA-deficient plasma and washed cellular BCs. Patients who have anaphylactic or repeated allergic reactions to BCs should be tested for IgA deficiency. It should be noted that the importance, or even the reality, of such a transfusion-related allergic risk is currently debated.

Graft-Versus-Host Disease GVHD is an extremely rare adverse reaction caused by transfusion, although it is a frequent complication of allogeneic hematopoietic transplantation. Transfusion-related GVHD is mediated by engrafted donor T lymphocytes in a recipient unable to reject such allogenic lymphocytes (as in severely immunosuppressed patients or patients homozygous for an HLA haplotype shared with the donor). Such donor T lymphocytes interact with host HLA Ags and mount an immune response, which is manifested clinically by the development, 5–10 days after transfusion, of cytopenia, fever, a characteristic skin rash, diarrhea, and liver function abnormalities. Transfusion-associated GVHD is highly resistant to treatment with immunosuppressive therapies as well as ablative therapy followed by allogeneic bone marrow transplantation and is fatal in >90% of cases. Prevention in at-risk patients relies on the irradiation of cellular BCs (minimum of 25 Gy) or treating BCs with pathogen reduction technology that will deplete all living cells in the component. At-risk patients include patients with inherited immune deficiency, patients undergoing autologous or allogeneic hematopoietic transplantation, patients treated with immunosuppressive drugs such as purine or pyrimidine analogues, anti-CD52 Ab, or antithymocyte globulin, fetuses receiving intrauterine transfusions, and recipients of BCs provided by a blood relative. Because granulocyte concentrates contain a large number of lymphocytes, they should always be irradiated.

Transfusion-Related Acute Lung Injury TRALI is characterized by the occurrence or worsening of hypoxia and noncardiogenic pulmonary edema with bilateral interstitial infiltrates on chest x-ray during or within 6 h after transfusion, although delayed cases may occur up to 72 h later. Frequency of TRALI is BC dependent and ranges, on average, from 0.5 to 10 per 10^5 BCs. TRALI may be difficult to distinguish from other causes of hypoxia, such as circulatory overload, and is among the most common causes of transfusion-related fatalities. Treatment is supportive only. TRALI usually results from the transfusion of donor plasma that contains high-titer anti-HLA class II Abs that bind recipient cognate Ag. Anti-HLA class I and anti-HNA Abs may also be involved. TRALI mediated by cytokines and chemokines in the absence of an HLA-mediated interaction may occur also. Leukocytes, especially when primed by either a bacterial moiety such as lipopolysaccharide or a cytokine/chemokine, aggregate in the pulmonary vasculature and release inflammatory mediators. The transfusion of plasma and PCs from male donors and nulliparous or parous female donors without anti-HLA Abs has significantly reduced the risk of TRALI where implemented. Recipient factors associated with an increased risk of TRALI include smoking, chronic alcohol use, shock, liver surgery (transplantation), cancer surgery, mechanical ventilation, and positive fluid balance.

Posttransfusion Purpura This rare reaction (~1/10^5 BCs) is defined as a thrombocytopenia-related bleeding disorder developing 5–12 days after PC (and more rarely RBCC) transfusion, predominantly in women. Platelet-specific alloAbs are found in the recipient, most frequently anti-HPA-1a in HPA-1a-negative alloimmunized individuals. The delayed thrombocytopenia is due to a secondary increased production of alloAbs. The mechanisms for the destruction of the patient's own platelets remain unclear. Management is mostly supportive but may require polyvalent immunoglobulin, glucocorticoids, or plasma exchange. Additional platelet transfusions may worsen the thrombocytopenia or be associated with poor increments. Prevention of recurrence includes use of washed BCs or BCs from HPA-compatible donors.

Alloimmunization/Platelet Refractoriness A recipient may become alloimmunized to a number of Ags on cellular blood elements and plasma proteins. AlloAbs to RBC Ags are detected during pretransfusion testing, and their presence may delay finding Ag-negative crossmatch-compatible products for transfusion. Women of childbearing age who are sensitized to RBC Ags (i.e., D, c, E, Kell, or Duffy) are at risk of bearing a fetus with hemolytic disease of the fetus or newborn. Ag matching is the only pretransfusion selection test to prevent RBC alloimmunization, which is found to occur with a frequency of

~100/10^5 RBCC transfusions. Alloimmunization to Ags on leukocytes and platelets, most often anti-HLA Abs, can result in refractoriness to PC transfusions (as defined by a low increase in platelet count after transfusion). Once alloimmunization has developed, HLA-compatible (crossmatched) PCs should be preferred if available. If not, repeated PCs at shortened intervals may be considered. Use of leukocyte-reduced cellular BCs will reduce the incidence of immunization. Transfusion refractoriness may also result from an anti-HPA alloimmunization, although less commonly. Recipient factors associated with platelet refractoriness include fever, splenomegaly, bleeding, DIC, and medications such as amphotericin B. Notably, cold-stored (and cryopreserved) PCs have been found to have preserved hemostatic function in acutely bleeding patients despite poor platelet increments.

Immunomodulation Transfusion of allogeneic blood may be associated with immunosuppression, as evidenced early on by the beneficial effect of pretransplant transfusion on kidney graft survival. The intensity of such an effect is debated and, if present, is most probably attenuated with leukoreduced BCs. Transfusion-related immunomodulation is indeed thought to be mainly mediated by donor leukocytes, whether transfused to the recipient or undergoing apoptosis during storage. However, leukoreduced RBCCs or PCs still release immunomodulatory mediators during storage. These mediators, along with the transfused RBCs or platelets, may exert various, possibly opposing, immune effects in vivo, including immunosuppression and inflammation.

NONIMMUNOLOGIC TRANSFUSION ADVERSE REACTIONS

Fluid Overload TACO is a common and underrecognized transfusion adverse reaction. Estimated frequencies vary from ~10 to 100 per 10^5 BCs. TACO is now the main cause of death from transfusion since the TRALI risk has been mitigated. Risk factors include older age, renal failure, preexisting fluid overload, cardiac dysfunction, administration of a large volume of BCs, and an excessive rate of transfusion in relation to the patient's hemodynamic tolerance. TACO results in dyspnea, hypoxia, bilateral and predominantly alveolar infiltrates on chest x-ray, frequent systolic hypertension, and elevated brain natriuretic peptide. Fever may also exist. Prevention involves identifying at-risk patients, close monitoring, a slow transfusion rate (1 RBCC over 3–4 h), and use of diuretics in hemodynamically stable patients with a history of TACO. Treatment requires stopping the transfusion and administrating oxygen and diuretics.

Massive Transfusion-Associated Reactions/Electrolyte and Cold Toxicity Reactions Reactions related to massive transfusion, i.e., transfusion of 50% of the patient's total blood volume over 3 h or >5–10 units of RBCCs (plus associated BCs), include citrate toxicity, hypothermia, hyperkalemia, and dilutional coagulopathy. Citrate, which is commonly used to anticoagulate BCs, chelates calcium. Hypocalcemia, manifested by circumoral paresthesia, and changes in cardiac function may result from multiple rapid transfusions. Although citrate is quickly metabolized to bicarbonate, calcium infusion (through a separate line) may be required. Rapid transfusion of BCs still at 4°C can result in hypothermia and cardiac dysrhythmias. Use of an inline warmer will prevent this complication. RBC leakage during storage, longer storage, and irradiation increase the concentration of potassium in the unit. Neonates and patients with renal failure or other comorbidities (e.g., hyperglycemia or hypocalcemia) are at risk of hyperkalemia and resulting acute cardiac toxicity. Treatment includes insulin, glucose, calcium gluconate, and furosemide, and prevention includes the use of washed or plasma-reduced RBCCs or a storage age of <7–10 days and the avoidance of RBCCs stored for >24 h after irradiation.

Iron Overload Each RBCC contains 200–250 mg of iron. In frequently transfused recipients, iron accumulation that is left untreated will affect endocrine, hepatic, and cardiac function. Death may occur from cardiac failure or arrhythmia. Iron overload can be assessed by means of serum ferritin measurements, magnetic resonance imaging, and liver biopsy. Prevention and treatment of this frequently underreported transfusion adverse event rely on careful monitoring and iron chelation.

Hypotensive Reactions Acute hypotensive transfusion reactions are defined as an abrupt drop in blood pressure of >30 mmHg early after the start of transfusion and resolving quickly once the transfusion is stopped, without further intervention. Respiratory, gastrointestinal, or mild allergic reactions may also be present. Estimated frequency is 1–10/10^5 BCs. These reactions may result from the generation of vasoactive kinins in the BCs and are more likely to occur in hypertensive patients taking angiotensin-converting enzyme (ACE) inhibitors who are therefore less able to metabolize bradykinin. Upon resolution, the same blood product should not be restarted. Switching from an ACE inhibitor to an alternative drug should be considered for patients requiring further transfusions.

Adverse Transfusion Reactions of Uncertain Imputability Necrotizing enterocolitis, which is common in preterm and very-low-birth-weight neonates, has been infrequently described with close temporal association with RBC transfusion. However, the causality of any association remains to be further ascertained, as does the efficacy of withholding feeds during transfusion to prevent such a complication. Posterior reversible encephalopathy syndrome is a rare syndrome characterized by acute reversible neurologic symptoms related to subcortical vasogenic brain edema. It has been described within 10 days after RBCC transfusion, mainly in women with severe (and long-standing) anemia. The prognosis is most often favorable, although irreversible neurologic disturbance has been described. Prevention may include avoiding rapid correction of chronic severe anemia. Again, causality remains to be established.

INFECTIOUS ADVERSE REACTIONS

Donor screening involves the selection of healthy donors without high-risk lifestyles, medical conditions, or exposure to transmissible pathogens. Tests are performed on donated blood to detect the presence of pathogens by testing for relevant Abs or Ags or by directly detecting infectious agents by nucleic acid amplification. Increasing sensitivity of testing methods has progressively narrowed the "window" period early on after infection during which a low-titer undetectable pathogen may be present in the blood and result in a transfusion-transmitted infection.

Transfusion-transmitted bacterial infection remains a significant concern, notably with PCs stored at room temperature, which allows for bacterial proliferation. However, some gram-negative bacteria such as *Yersinia* can grow at 4°C and therefore may be implicated in infections related to RBCC transfusion. Recipients of contaminated BCs may develop abrupt (during transfusion and up to several hours after) fever and chills, which can deteriorate to septic shock, DIC, and death. Endotoxin formed within the BC may be implicated. After sampling for bacterial culture, broad-spectrum antibiotics should be promptly initiated.

Pathogen reduction of PC and plasma, and perhaps soon of RBCCs as well, offers an additional means of reducing transfusion infection risks. Although effective for a wide range of pathogens, such processes are most often ineffective for bacterial spores and nonenveloped viruses such as hepatitis A virus, parvovirus B19, and hepatitis E virus. Postdonation information (i.e., fever occurring within 24 h after donation) allows the involved BCs to be quarantined and may provide an additional safety measure.

Transfusion-transmitted infections are increasingly rare. However, new or previously unidentified infectious risks may occur, as highlighted by the emergence of the transfusion-associated West Nile virus infection and babesiosis in early 2000 in the United States, as well as transfusion-associated hepatitis E in early 2010 in Europe. Such occurrences require active surveillance programs and the appropriate implementation of mitigation measures such as additional testing, pathogen reduction, and travel-related deferral criteria. Along with West Nile virus, a number of other arbovirus-related infections

TABLE 118-6 Infectious Transfusion Adverse Events

PATHOGEN		DONATION PREVALENCE (/10^4 BLOOD DONATIONS)	PREVENTION MEASURES (IN ADDITION TO DONOR DEFERRAL)	INFECTION PREVALENCE IN RECIPIENTS (/10^6 BLOOD PRODUCTS TRANSFUSED)
Bacteria	Pyogenic bacteria	PC: 10–20	Venipuncture sepsis, diversion of the initial 10–30 mL of blood, bacterial detection, pathogen reduction (for PC)	Sepsis: PC: 5–30; with bacterial detection: <1 to 10; with pathogen reduction: <1 RBCC: <0.2
	Treponema pallidum (syphilis)	~1[a]	Serology[b,c]	<0.01
Virus	HIV-1/2	~0.1	Serology, NAT (+/– p24 Ag)[b,c]	0.1–1[d]
	HBV	~0.5	Serology, NAT[b,c]	<0.5 (3 without NAT)[d]
	HCV	0.2–1.2	Serology, NAT[b,c]	<0.1–1[d]
	HTLV-1/2	0.05–0.1[a]	Serology, BC leukocyte reduction [b,c]	0.1–0.3[d]
	HEV	0–10 (in endemic regions)	NAT	Endemic regions: <0.1 with NAT; a transmission rate from infected donors of ~50% has been reported
	CMV	Undetermined	Serology, BC leukocyte reduction[b,c]	<0.1 in leukocyte-reduced BCs
	Parvovirus B19	~0.5 with viral DNA >10^6 IU/mL,[e] up to 100 overall	NAT	Most adults are immune to parvovirus B19; up to 0.12% in seronegative adults has been reported
	West Nile virus	Up to 3 in high season endemic regions[a]	NAT[b]	High season endemic regions: <1 with NAT
Parasite[f]	*Plasmodium* (malaria)	~4 (40–>50 in donors from endemic regions)[a]	Serology (NAT soon available)	<0.1 in nonendemic regions
	Babesia	~90 (in endemic regions)[a]	Serology (NAT implementation underway)	ND (0.04% donors may be within the serology window period)
	Trypanosoma cruzi (Chagas disease)	~0.14 in donors/mothers from endemic regions[a]	Serology, leukocyte reduction	ND

[a]As assessed based on seropositivity, i.e., including a varying percentage of individuals not harboring the pathogen in their blood. [b]Prevention measures may also include pathogen reduction (for PC and plasma), [c]Prevention measures may also include a quarantine of the (cryopreserved) BC pending a negative serology on a subsequent donation (for plasma), [d]Estimated residual risk. [e]Transfusion risk deemed as absent below this threshold, [f]Visceral leishmaniasis is a transfusion risk as well in endemic regions, a risk that may be mitigated by leukocyte reduction.

Note. Other pathogens associated with transfusion-transmitted infections at a very low frequency include arboviruses other than West Nile (dengue, Zika virus), hepatitis A, human herpesvirus-8, Japanese encephalitis virus, tick-borne encephalitis virus complex, and the prion responsible for variant Creutzfeldt-Jakob disease (4 cases in the United Kingdom, in the context of the bovine spongiform encephalopathy epidemic, before implementation of systematic leukocyte reduction).

Abbreviations: Ag, antigen; BC, blood component; CMV, cytomegalovirus; HBV, hepatitis B virus; HCV, hepatitis C virus; HEV, hepatitis E virus; HTLV, human T-cell leukemia virus; NAT, nucleic acid detection test; ND, not determined; PC platelet concentrate; RBCC, red blood cell concentrate.

possibly transmissible by blood transfusion are endemic or involved in epidemic outbreaks. Despite being possibly present in the blood at asymptomatic phases of the disease, documented cases of transfusion-transmitted infections involving these arboviruses have been very rare (Zika), often without a discernible clinical impact (Dengue), or absent (Chikungunya). Route of infection (i.e., intravenous vs mosquito bite), pathogen dose, storage conditions, recipient immune status, and ongoing treatments may all impact the ability of a pathogen in the donor to induce a disease in the recipient. Estimated frequencies of transfusion-relevant infections in donors and of transfusion-transmitted infections are reported in Table 118-6. Such frequencies depend heavily on variables such as local epidemiology, donor deferral rules, risk reduction measures, and data reporting, and may vary considerably.

ALTERNATIVES AND PERSPECTIVES

In addition to promoting appropriate transfusion indications, patient blood management programs have highlighted several transfusion-sparing strategies, such as the treatment of iron deficiency before surgery and minimization of blood loss. Antifibrinolytic agents such tranexamic acid (TXA) have been shown to be effective in preventing bleeding complications in various forms of hemorrhage. Erythropoietin stimulates erythrocyte production in patients with anemia from chronic renal failure and other conditions. Thrombopoietin receptor agonists may reduce platelet transfusion needs resulting from chemotherapy-induced thrombopenia. Bone marrow transplantation and gene therapy approaches, including CRISPR-Cas9-mediated gene editing, dramatically reduce the transfusion needs in patients with sickle cell or major thalassemia. Stem cell–derived blood cells such as RBCs or platelets may in the future become a suitable alternative to rare blood donors.

Importantly, issues surrounding transfusion safety now fully encompass transfusion efficacy. Linked large-scale databases pertaining to blood donors and transfused patients will also be instrumental in further assessing and understanding the basis of transfusion safety and efficacy. Optimal transfusion care will soon require consideration of new criteria in relation to donor, blood product, and/or recipient characteristics.

ACKNOWLEDGMENTS

The authors are indebted to Jeffery S. Dzieczkowski and Kenneth C. Anderson who co-authored the chapter in a previous edition and expertly paved the way for this chapter.

FURTHER READING

Carson JL et al: Red blood cell transfusion 2023 AABB International guidelines. JAMA 330:1892, 2023.

Delaney M et al: Transfusion reactions: Prevention, diagnosis, and treatment. Lancet 388:2825, 2016.

Panch SR et al: Hemolytic transfusion reactions. N Engl J Med 381:150, 2019.

Stanworth SJ, Shah A: How I use platelet transfusions. Blood 140:1925, 2022.

119 Hematopoietic Cell Transplantation

Frederick R. Appelbaum

Bone marrow transplantation was the original term used to describe the collection and transplantation of hematopoietic stem cells, but with the demonstration that peripheral blood and umbilical cord blood are also useful sources of stem cells, *hematopoietic cell transplantation* has become the preferred generic term for this process. Hematopoietic cell transplantation is used to treat patients with an abnormal but nonmalignant lymphohematopoietic system by replacing it with one from a normal donor. Hematopoietic cell transplantation is also used to treat malignancy by allowing the administration of higher doses of myelosuppressive therapy than would otherwise be possible and, in the setting of allogeneic hematopoietic cell transplantation, by conferring an immunologic graft-versus-tumor effect. The use of hematopoietic cell transplantation is increasing, as it becomes safer and applicable to more diseases and as donor availability expands.

The Worldwide Network for Blood and Marrow Transplantation (*http://www.wbmt.org*) estimates that worldwide more than 100,000 transplants were performed in 2022. The frequency of transplantation varied widely from country to country, with a close association of transplant rates with gross national income (GNI) per capita. However, even among countries with similar GNIs per capita, there are substantial differences between countries and regions regarding the frequency of transplantation, disease indications, and choice of donor type.

THE HEMATOPOIETIC STEM CELL

Several features of the hematopoietic stem cell (HSC) make transplantation clinically feasible, including its remarkable regenerative capacity, its ability to home to the marrow space following intravenous injection, and the ability of the stem cell to be cryopreserved **(Chap. 101)**. Transplantation of a single stem cell can replace the entire lymphohematopoietic system of an adult mouse. In humans, transplantation of a small percentage of a donor's bone marrow volume regularly results in complete and sustained replacement of the recipient's entire lymphohematopoietic system, including all red cells, granulocytes, B and T lymphocytes, and platelets, as well as cells comprising the fixed macrophage population, including Kupffer cells of the liver, pulmonary alveolar macrophages, osteoclasts, and Langerhans cells of the skin. Homing of HSCs to their marrow niche initially involves interactions between P- and E-selectins on marrow sinusoidal endothelium with integrins including VLA-4 on HSCs. Once tethered to the vascular endothelium, changes in integrin conformation result in tight adhesion following which stem cells migrate through the endothelium and extracellular matrix eventually reaching the stem cell niche. This last step is facilitated by CXCL12 produced by the niche stroma interacting with the chemokine CXCR4 on HSCs. Human hematopoietic stem cells can survive freezing and thawing with little, if any, damage, making it possible to remove and store a portion of the patient's own bone marrow for later reinfusion following treatment of the patient with high dose myelotoxic therapy.

CATEGORIES OF HEMATOPOIETIC CELL TRANSPLANTATION

Hematopoietic cell transplantation can be described according to the relationship between the patient and the donor and by the anatomic source of stem cells. In ~1% of cases, patients have identical twins who can serve as donors. With the use of syngeneic donors, there is no risk of graft-versus-host disease (GVHD), and unlike the use of autologous marrow, there is no risk that the stem cells are contaminated with tumor cells.

Allogeneic transplantation involves a donor and a recipient who are not genetically identical. Following allogeneic transplantation, immune cells transplanted with the stem cells or developing from them can react against the patient, causing GVHD. Alternatively, if the immunosuppressive preparative regimen used to treat the patient before transplant is inadequate, immunocompetent cells of the patient can cause graft rejection. The risks of these complications are influenced by the degree of matching between donor and recipient for human leukocyte antigen (HLA) molecules encoded by genes of the major histocompatibility complex.

HLA molecules are responsible for binding antigenic proteins and presenting them to T cells. The antigens presented by HLA molecules may derive from exogenous sources (e.g., during active infections) or may be endogenous proteins. If individuals are not HLA-matched, T cells from one individual will react strongly to the mismatched HLA, or "major antigens," of the second. Even if the individuals are HLA-matched, the T cells of the donor may react to differing endogenous or "minor antigens" presented by the HLA of the recipient. Reactions to minor antigens tend to be less vigorous. The genes of major relevance to transplantation include HLA-A, -B, -C, and -D; they are closely linked and therefore tend to be inherited as haplotypes, with only rare crossovers between them. Thus, the odds that any one full sibling will match a patient are one in four, and the probability that the patient has an HLA-identical sibling is $1 - (0.75)n$, where n equals the number of siblings.

With conventional techniques, the risk of graft rejection is 1–3%, and the risk of severe, life-threatening acute GVHD is ~15% following transplantation between HLA-identical siblings. The incidence of graft rejection and GVHD increases progressively with the use of family member donors mismatched for one, two, or three antigens. Newer approaches to GVHD prophylaxis, including the use of posttransplant high-dose cyclophosphamide, have diminished the impact of HLA mismatching, making transplantation between donor/recipient pairs who share only one HLA haplotype possible. Since the formation of the National Marrow Donor Program and other registries, HLA-matched unrelated donors can be identified for many patients. The genes encoding HLA antigens are highly polymorphic, and thus the odds of any two unrelated individuals being HLA identical are extremely low, somewhat less than 1 in 10,000. However, by recruiting >40 million volunteer donors, HLA-matched donors can be found for ~60% of patients for whom a search is initiated, with higher rates among whites and lower rates among minorities and patients of mixed race. It takes, on average, 3–4 months to complete a search and schedule and initiate an unrelated donor transplant. With improvements in HLA typing and supportive care measures, survival following matched unrelated donor transplantation is essentially the same as that seen with HLA-matched siblings.

Allogeneic hematopoietic cell transplantation can be carried out across ABO blood barriers by removing isoagglutinins and/or incompatible red blood cells from the donor graft. However, depending on the direction of the mismatch, hemolysis of donor cells by persistent isoagglutinins in the host, or hemolysis of recipient red cells by isoagglutinins in the graft or developing from it may occur despite appropriate manipulation of the donor cell product.

Autologous transplantation involves the removal and storage of the patient's own stem cells with subsequent reinfusion after the patient receives high dose myeloablative therapy. Unlike allogeneic transplantation, there is no risk of GVHD or graft rejection with autologous transplantation. On the other hand, autologous transplantation lacks a graft-versus-tumor (GVT) effect, and the autologous stem cell product can be contaminated with tumor cells, which could lead to relapse. A variety of techniques have been developed to "purge" autologous products of tumor cells, but no prospective randomized trials have shown that any approach decreases relapse rates or improves disease-free or overall survival.

Bone marrow aspirated from the posterior and anterior iliac crests initially was the source of hematopoietic stem cells for transplantation. Typically, anywhere from 1.5 to 5×10^8 nucleated marrow cells per kilogram are collected for allogeneic transplantation. Several studies have found improved survival following both matched sibling and unrelated transplantation by transplanting higher numbers of bone marrow cells.

Hematopoietic stem cells circulate in the peripheral blood but in very low concentrations. Following the administration of a myeloid growth factor such as granulocyte colony-stimulating factor (G-CSF) and during recovery from intensive chemotherapy, the concentration of hematopoietic progenitor cells in blood, as measured either by colony-forming units or expression of the CD34 antigen, increases markedly. This makes it possible to harvest adequate numbers of stem cells from the peripheral blood for transplantation. Donors are typically treated with 4 or 5 days of hematopoietic growth factor, following which stem cells are collected in one or two 4-h pheresis sessions. In the autologous setting, transplantation of $>2.5 \times 10^6$ CD34 cells per kilogram, a number that can be collected in most circumstances, leads to rapid and sustained engraftment in virtually all cases. In the 5–10% of patients who fail to mobilize enough CD34+ cells with growth factor alone, the addition of plerixafor, an antagonist of CXCR4, may be useful. Blocking CXCR4 allows more stem cells to escape the marrow. When compared to the use of autologous marrow, use of peripheral blood stem cells results in more rapid hematopoietic recovery. Although this more rapid recovery diminishes the morbidity rate of transplantation, no studies show improved survival.

In the setting of allogeneic transplantation, the use of growth factor–mobilized peripheral blood stem cells also results in faster engraftment than seen with marrow but at the cost of more chronic GVHD because of donor T-cell contamination. With matched sibling donors, the increased chronic GVHD is more than balanced by reductions in rates of relapse and nonrelapse mortality, resulting in improved overall survival. However, in the setting of matched unrelated donor transplantation, use of peripheral blood results in more chronic GVHD without a compensatory survival advantage. Nonetheless, because of ease of collection, peripheral blood continues to be the more commonly used source of stem cells.

Umbilical cord blood contains a high concentration of hematopoietic progenitor cells, allowing for its use as a source of stem cells for transplantation. Cord blood transplantation from family members has been used when the immediate need for transplantation precludes waiting the 9 or so months generally required for the baby to mature to the point of donating marrow. Use of cord blood results in slower peripheral count recovery than seen with marrow but a lower incidence of GVHD, perhaps reflecting the low number of T cells in cord blood. Multiple cord blood banks have been developed to harvest and store cord blood for possible transplantation to unrelated patients from material that would otherwise be discarded. Currently >800,000 units are cryopreserved and available for use. The advantages of unrelated cord blood are rapid availability and decreased immune reactivity allowing for the use of partially matched units, which is of particular importance for those without matched unrelated donors. The risks of graft failure and transplant-related mortality are related to the dose of cord blood cells per kilogram, which previously limited the application of single cord blood transplantation to pediatric and smaller adult patients. Subsequent trials have found that for patients without suitable single cord units, the use of double cord transplants diminishes the risk of graft failure and early mortality even though only one of the donors ultimately engrafts. Given the similar survival rates seen with cord blood, matched unrelated, and haploidentical family member donors, a source of allogeneic stem cells can now be found for almost every patient in need (**Table 119-1**).

TABLE 119-1 Probability of Identifying a Donor Based on Stem Cell Source and Patient Ethnicity

	UNRELATED ADULT %		UNRELATED CORD %	HAPLOIDENTICAL
Ethnicity	8/8[a]	7/8[a]	≥4/6[b]	
Caucasian	75	90	>95	95
Hispanic	35	75	95	95
Black	18	70	90	95

[a]Matching for HLA-A, -B, -C, and DRB1. [b]Matching for HLA-A, -B, and DRB1.

THE TRANSPLANT PREPARATIVE REGIMEN

The treatment regimen administered to patients immediately preceding transplantation is designed to eradicate the patient's underlying disease and, in the setting of allogeneic transplantation, immunosuppress the patient adequately to prevent rejection of the transplanted stem cells. The appropriate regimen therefore depends on the disease setting and graft source. For example, when transplantation is performed to treat severe combined immunodeficiency and the donor is a histocompatible sibling, no preparative regimen is needed because no host cells require eradication, and the patient is already too immune incompetent to reject the transplanted graft. For aplastic anemia, there is no large population of cells to eradicate, and high-dose cyclophosphamide plus antithymocyte globulin are sufficient to immunosuppress the patient adequately to accept the marrow graft. In the setting of thalassemia and sickle cell anemia, high-dose busulfan is frequently added to cyclophosphamide to eradicate hyperplastic host hematopoiesis. A variety of different regimens have been developed to treat malignant diseases. Most regimens include agents with high activity against the tumor in question at conventional doses and with myelosuppression as their predominant dose-limiting toxicity. Therefore, these regimens commonly include busulfan, cyclophosphamide, melphalan, thiotepa, carmustine, etoposide, and total-body irradiation in various combinations.

Although high-dose treatment regimens were the initial approach to transplantation for malignancies, the realization that much of the antitumor effect of transplantation derives from an immunologically mediated GVT response led investigators to ask if reduced-intensity conditioning regimens might be effective and more tolerable. Evidence for a GVT effect comes from studies showing that posttransplant relapse rates are lowest in patients who develop acute and chronic GVHD, higher in those without GVHD, and higher still in recipients of T cell–depleted allogeneic or syngeneic marrow. The demonstration that complete remissions can be obtained in many patients who have relapsed after transplant by simply administering viable lymphocytes from the original donor further strengthens the argument for a potent GVT effect. Accordingly, a variety of alternative regimens have been studied, ranging from nonmyeloablative, which are the very minimum required to achieve engraftment (e.g., fludarabine plus 200 cGy total-body irradiation) and would cause only transient myelosuppression if no transplant were performed, to so-called reduced-intensity regimens, which would cause significant but not necessarily fatal myelosuppression in the absence of transplantation (e.g., fludarabine plus melphalan). With the use of nonmyeloablative and reduced-intensity regimens, engraftment is readily achieved with less toxicity than seen with conventional transplantation. In general, relapse rates are higher following reduced-intensity conditioning, but transplant-related mortality is lower, favoring the use of reduced-intensity conditioning in older patients and those with significant comorbidities. High-dose regimens are favored in those felt able to tolerate the treatment, particularly if patients have any evidence of measurable disease at the time of transplantation.

THE TRANSPLANT PROCEDURE

Marrow is usually collected from the donor's posterior and sometimes anterior iliac crests, with the donor under general or spinal anesthesia. Typically, 10–15 mL/kg of marrow is aspirated, placed in heparinized media, and filtered through 0.3- and 0.2-mm screens to remove fat and bony spicules. The collected marrow may undergo further processing depending on the clinical situation, such as the removal of red cells to prevent hemolysis in ABO-incompatible transplants, the removal of donor T cells to prevent GVHD, or attempts to remove possible contaminating tumor cells in autologous transplantation. Marrow donation is safe, with only very rare complications reported.

Peripheral blood stem cells are collected by leukapheresis after the donor has been treated with hematopoietic growth factors or, in the setting of autologous transplantation, sometimes after treatment with a combination of chemotherapy and growth factors. Stem cells for transplantation are infused through a large-bore central venous catheter. Such infusions are usually well tolerated, although occasionally patients develop fever, cough, or shortness of breath. These symptoms typically resolve with slowing of the infusion. When the stem cell product has been cryopreserved using dimethyl sulfoxide, patients sometimes experience short-lived nausea or vomiting due to the taste (and smell) of the cryoprotectant.

ENGRAFTMENT AND IMMUNE RECONSTITUTION

Peripheral blood counts reach their nadir several days to a week after transplant as a consequence of the preparative regimen; then cells produced by the transplanted stem cells begin to appear in the peripheral blood. The rate of recovery depends on the source of stem cells and use of posttransplant growth factors. If marrow is the source, recovery to 100 granulocytes/μL occurs on average by day 16 and to 500/μL by day 22. Use of G-CSF–mobilized peripheral blood stem cells speed the rate of recovery by ~1 week compared to marrow, whereas engraftment following cord blood transplantation is typically delayed by ~1 week. Use of a myeloid growth factor after transplant accelerates recovery by 3–5 days. Platelet counts usually recover shortly after granulocytes.

While granulocytes and other components of innate immunity recover rapidly after hematopoietic cell transplantation, adaptive immunity, which consists of cellular (T cell) and humoral (B cell) immunity, may take 1–2 years to fully recover. Survival and peripheral expansion of infused donor T cells is the dominant mechanism for T-cell recovery in the first months after hematopoietic cell transplantation and results in mostly CD8+ T cells with a limited repertoire. After several months, de novo generation of donor-derived CD4+ and CD8+ T cells becomes dominant, providing a more diverse T-cell repertoire. B-cell counts recover by 6 months after autologous hematopoietic cell transplantation and 9 months after allogeneic hematopoietic cell transplantation. In general, immune recovery occurs more rapidly after autologous than allogeneic hematopoietic cell transplantation and after receipt of unmodified grafts compared to the setting of in vivo or ex vivo T-cell depletion.

Following allogeneic transplantation, engraftment can be documented using fluorescence in situ hybridization of sex chromosomes if donor and recipient are sex-mismatched or by analysis of short tandem repeat polymorphisms after DNA amplification.

COMPLICATIONS FOLLOWING HEMATOPOIETIC CELL TRANSPLANTATION

Early Direct Chemoradiotoxicities The transplant preparative regimen may cause a spectrum of acute toxicities that vary according to the intensity of the regimen and the specific agents used but frequently include nausea, vomiting, and mild skin erythema **(Fig. 119-1)**.

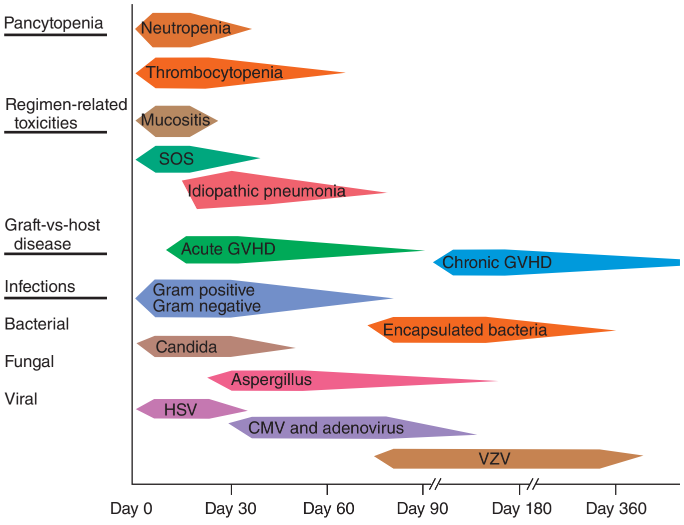

FIGURE 119-1 Major syndromes complicating marrow transplantation. CMV, cytomegalovirus; GVHD, graft-versus-host disease; HSV, herpes simplex virus; SOS, sinusoidal obstructive syndrome (formerly venoocclusive disease); VZV, varicella-zoster virus. The size of the shaded area roughly reflects the period of risk of the complication.

High-dose cyclophosphamide can result in hemorrhagic cystitis, which can usually be prevented by bladder irrigation or with the sulfhydryl compound mercaptoethanesulfonate (MESNA). Most high-dose preparative regimens will result in oral mucositis, which typically develops 5–7 days after transplant and often requires narcotic analgesia. Use of a patient-controlled analgesic pump provides the greatest patient satisfaction and results in a lower cumulative dose of narcotic. Keratinocyte growth factor (palifermin) can shorten the duration of mucositis by several days following autologous transplantation. Patients begin losing their hair 5–6 days after transplant and by 1 week are usually profoundly pancytopenic.

Depending on the intensity of the conditioning regimen, 3–10% of patients will develop sinusoidal obstruction syndrome (SOS) of the liver (formerly called venoocclusive disease), a syndrome that results from direct cytotoxic injury to hepatic-venular and sinusoidal endothelium, with subsequent deposition of fibrin and the development of a local hypercoagulable state. This chain of events leads to the clinical symptoms of tender hepatomegaly, ascites, jaundice, and fluid retention. These symptoms can develop any time during the first month after transplant, with the peak incidence at day 16. Predisposing factors include prior exposure to intensive chemotherapy, pretransplant hepatitis of any cause, and use of more intense conditioning regimens. The mortality rate of sinusoidal obstruction syndrome is ~30%, with progressive hepatic failure culminating in a terminal hepatorenal syndrome. Treatment of severe SOS with defibrotide, a polydeoxyribonucleotide, reduces mortality.

Although most pneumonias developing early after transplant are caused by infectious agents, in a small percentage of patients, a diffuse interstitial pneumonia will develop that is a result of direct toxicity of high-dose preparative regimens. Bronchoalveolar lavage usually shows alveolar hemorrhage, and biopsies are typically characterized by diffuse alveolar damage, although some cases may have a more clearly interstitial pattern. High-dose glucocorticoids or anti–tumor necrosis factor therapies are sometimes used as treatment, although randomized trials proving their utility have not been reported.

Transplant-associated thrombotic microangiopathy is seen in 5–10% of patients, appearing on average about 1 month after transplant. The syndrome is characterized by presence of schistocytes on peripheral smear, elevated lactate dehydrogenase, thrombocytopenia, and acute kidney injury and is the result of endothelial injury and complement activation. Since calcineurin and mTOR inhibitors are thought to contribute to the pathogenesis of the syndrome, changing immunosuppressive regimens is sometimes effective. Patients sometimes respond to eculizumab.

Late Direct Chemoradiotoxicities Two categories of chronic pulmonary disease occur in patients >3 months after hematopoietic cell transplantation. Cryptogenic organizing pneumonia is a restrictive lung disease characterized by dry cough, shortness of breath, and chest imaging showing a diffuse, fluffy infiltrate. Biopsy shows granulation tissue within alveolar spaces and small airways and no infectious agents. The disease responds well to corticosteroids and is entirely reversible. Bronchiolitis obliterans is an obstructive disease presenting with cough, progressive dyspnea, and radiologic evidence of air trapping. Pathology shows collagen and granulation tissue in and around bronchial structures and eventually obliteration of small airways. The disease is usually associated with chronic GVHD, and although it may respond to increasing immunosuppression, complete reversal is uncommon.

Other late complications of the preparative regimen include decreased growth velocity in children and delayed development of secondary sex characteristics. These complications can be partly ameliorated with the use of appropriate growth and sex hormone replacement. Most men become azoospermic, and most postpubertal women will develop ovarian failure, which should be treated. However, pregnancy is possible after transplantation, and patients should be counseled accordingly. Thyroid dysfunction, usually well compensated, is sometimes seen. Cataracts develop in 10–20% of patients and are most common in patients treated with total-body irradiation and those who receive glucocorticoid therapy after transplant for treatment of GVHD. Aseptic necrosis of the femoral head is seen in 10% of patients and is particularly frequent following chronic glucocorticoid therapy. Both acute and late chemoradiotoxicities (except those due to glucocorticoids and other agents used to treat GVHD) are less frequent in recipients of reduced-intensity compared to high-dose preparative regimens.

Graft Failure Although complete and sustained engraftment is usually seen after transplant, occasionally marrow function either does not return or, after a brief period of engraftment, is lost. Graft failure after autologous transplantation can be the result of inadequate numbers of stem cells being transplanted, damage during ex vivo treatment or storage, or exposure of the patient to myelotoxic agents after transplant. Infections with cytomegalovirus (CMV) or human herpesvirus type 6 have also been associated with loss of marrow function. Graft failure after allogeneic transplantation can also be due to immunologic rejection of the graft by immunocompetent host cells. Such rejection is generally thought to be mostly T-cell–mediated, but the presence before hematopoietic cell transplantation of donor-specific HLA antibodies in the patient is associated with poor engraftment, leading to the recommendation for screening for donor-directed anti-HLA antibodies in recipients prior to transplant. Immunologically based graft rejection is more common following use of less immunosuppressive preparative regimens, in recipients of T-cell–depleted stem cell products, and in patients receiving grafts from HLA-mismatched donors or cord blood.

Treatment of graft failure involves removing all potentially myelotoxic agents from the patient's regimen and attempting a short trial of a myeloid growth factor. Persistence of lymphocytes of host origin in allogeneic transplant recipients with graft failure indicates immunologic rejection. Reinfusion of donor stem cells in such patients is usually unsuccessful unless preceded by a second immunosuppressive preparative regimen. Standard high-dose preparative regimens are tolerated poorly if administered within 100 days of a first transplant because of cumulative toxicities. However, reduced-intensity conditioning regimens have been effective in some cases.

Graft-Versus-Host Disease Acute GVHD occurs within the first 3 months after allogeneic transplant with a peak onset around 4 weeks and is characterized by an erythematous maculopapular rash; by persistent anorexia or diarrhea, or both; and by liver disease with increased serum levels of bilirubin, alanine and aspartate aminotransferase, and alkaline phosphatase. Because many conditions can mimic acute GVHD, the diagnosis usually requires skin, liver, or endoscopic biopsy for confirmation. In all these organs, endothelial damage and lymphocytic infiltrates are seen. In skin, the epidermis and hair follicles are damaged; in liver, the small bile ducts show segmental disruption; and in intestines, destruction of the crypts and mucosal ulceration may be noted. A commonly used rating system for acute GVHD is shown in Table 119-2. Grade I acute GVHD is of little clinical significance, does not affect the likelihood of survival, and does not require treatment. In contrast, grades II to IV GVHD are associated with significant symptoms and a poorer probability of survival and require aggressive therapy. The incidence of acute GVHD is higher in recipients of stem cells from mismatched or unrelated donors, in older patients, and in patients unable to receive full doses of drugs used to prevent the disease.

Historically, the standard approach to GVHD prevention was the administration of a calcineurin inhibitor (cyclosporine or tacrolimus) paired with an antimetabolite (most commonly methotrexate) following transplantation. Prospective randomized trials have demonstrated the benefit of adding a third drug, either mycophenolate mofetil, abatacept, or posttransplant cyclophosphamide, to the two-drug regimen. Other approaches include the addition of anti–T-cell immune globulin (ATG) to the GVHD prophylactic regimen or the removal of subsets or all T cells from the stem cell inoculum.

Despite prophylaxis, significant acute GVHD will develop in ~30% of recipients of stem cells from matched siblings. Factors associated with a greater risk of acute GVHD include HLA-mismatching between recipient and donor, patient and donor age, use of more intense preparative regimens, and use of multiparous women as donors. Presumably, multiparous women have more alloreactivity based on carriage of genetically disparate fetuses. Disruption of the intestinal microbiota leading to loss of diversity and overgrowth by a single taxon is associated with a higher risk of GVHD and transplant-associated mortality. Biomarkers, including ST2, REG32, and TNF R1, have been identified that predict the severity of acute GVHD. The disease is usually treated with prednisone at a daily dose of 1–2 mg/kg. Patients in whom the acute GVHD fails to respond to prednisone sometimes respond to the oral JAK2 inhibitor ruxolitinib.

Chronic GVHD occurs most commonly between 3 months and 2 years after allogeneic transplant, developing in 20–50% of recipients. The disease is more common in older patients, with the use of peripheral blood rather than marrow as the stem cell source, in recipients of mismatched or unrelated stem cells, and in those with a preceding episode of acute GVHD. The disease resembles an autoimmune disorder with malar rash, sicca syndrome, arthritis, obliterative bronchiolitis, and bile duct degeneration with cholestasis. Mild chronic GVHD can sometimes be managed using local therapies (topical glucocorticoids to skin and cyclosporine eye drops). More severe disease requires systemic therapy usually with prednisone, which leads to responses in 40–60% of patients. Three drugs have received U.S. Food and Drug Administration approval for the treatment of steroid-resistant chronic GVHD: ibrutinib, ruxolitinib, and belumosudil. All three are kinase inhibitors, a class of compounds that reduces growth signals and activation of key cellular proteins involved with cell activation, migration, and proliferation. Mortality rates from chronic GVHD average around 15% but range from 5 to 50% depending on severity. In most patients, chronic GVHD resolves, but it may require 1–3 years of immunosuppressive treatment before these agents can be withdrawn without the disease recurring. Because patients with chronic GVHD are susceptible to significant infection, they should receive prophylactic trimethoprim-sulfamethoxazole, and all suspected infections should be investigated and treated aggressively.

Although onset before or after 3 months after transplant is often used to discriminate between acute and chronic GVHD, occasional

TABLE 119-2 Clinical Staging and Grading of Acute Graft-Versus-Host Disease

CLINICAL STAGE	SKIN	LIVER—BILIRUBIN, λmol/L (mg/dL)	GUT
1	Rash <25% body surface	34–51 (2–3)	Diarrhea 500–1000 mL/d
2	Rash 25–50% body surface	51–103 (3–6)	Diarrhea 1000–1500 mL/d
3	Generalized erythroderma	103–257 (6–15)	Diarrhea >1500 mL/d
4	Desquamation and bullae	>257 (>15)	Ileus
OVERALL CLINICAL GRADE	**SKIN STAGE**	**LIVER STAGE**	**GUT STAGE**
I	1–2	0	0
II	1–3	1	1
III	1–3	2–3	2–3
IV	2–4	2–4	2–4

patients will develop signs and symptoms of acute GVHD after 3 months (late-onset acute GVHD), whereas others will exhibit signs and symptoms of both acute and chronic GVHD (overlap syndrome). No data suggest that these patients should be treated differently than those with classic acute or chronic GVHD.

From 3 to 5% of patients will develop an autoimmune disorder following allogeneic hematopoietic cell transplantation, most commonly autoimmune hemolytic anemia or idiopathic thrombocytopenic purpura. Unrelated donor source and chronic GVHD are risk factors, but autoimmune disorders have been reported in patients with no obvious GVHD. Treatment is with prednisone, cyclosporine, or rituximab.

Infection Posttransplant patients, particularly recipients of allogeneic transplantation, require unique approaches to the problem of infection. Early after transplantation, patients are profoundly neutropenic, and because the risk of bacterial infection is so great, most centers place patients on broad-spectrum antibiotics once the granulocyte count falls to <500/μL. Prophylaxis against fungal infections reduces rates of infection and improves overall survival. Fluconazole is often used for patients with standard risk, while prophylaxis with mold active agents (voriconazole, posaconazole, or isavuconazonium) should be considered for patients at higher risk, such as those with a prior fungal infection. Patients seropositive for herpes simplex should receive acyclovir or valacyclovir prophylaxis. One approach to infection prophylaxis is shown in **Table 119-3**. Despite these prophylactic measures, most patients will develop fever and signs of infection after transplant. The management of patients who become febrile despite bacterial and fungal prophylaxis is a difficult challenge and is guided by individual aspects of the patient and by the institution's experience.

The general problem of infection in the immunocompromised host is discussed in **Chap. 148**.

Once patients engraft, the incidence of bacterial infection diminishes; however, patients, particularly allogeneic transplant recipients, remain at significant risk of infection. During the period from engraftment until about 3 months after transplant, the most common causes of infection are gram-positive bacteria, fungi (particularly *Aspergillus*), and viruses including CMV. CMV disease, which in the past was frequently seen and often fatal, can be prevented in seronegative patients transplanted from seronegative donors using either seronegative blood products or products from which the white blood cells have been removed. In seropositive patients or patients transplanted from seropositive donors, either prophylaxis or preemptive therapy is used. Letermovir administered over the first 3 months after transplant is effective as prophylaxis. An alternative approach is to monitor blood of patients after transplant using polymerase chain reaction assays for viral DNA and to treat reactivation preemptively with ganciclovir before clinical disease develops. Foscarnet is effective for some patients who develop CMV antigenemia or infection despite the use of ganciclovir or who cannot tolerate the drug, but it can be associated with severe electrolyte wasting.

TABLE 119-3 Approach to Infection Prophylaxis in Allogeneic Transplant Recipients

ORGANISM	AGENT	APPROACH
Bacterial	Levofloxacin	750 mg PO or IV daily
Fungal	Fluconazole	400 mg PO qd to day 75 posttransplant
Pneumocystis jirovecii	Trimethoprim-sulfamethoxazole	1 double-strength tablet PO bid 2 days/week until day 180 or off immunosuppression
Viral		
Herpes simplex	Acyclovir	800 mg PO bid to day 30
Varicella-zoster	Acyclovir	800 mg PO bid to day 365
Cytomegalovirus	Ganciclovir	5 mg/kg IV bid for 7 days, then 5 (mg/kg)/d 5 days/week to day 100

Pneumocystis jirovecii pneumonia, once seen in 5–10% of patients, can be prevented by treating patients with oral trimethoprim-sulfamethoxazole for 1 week before transplant and resuming the treatment once patients engraft.

Respiratory viruses that cause community-acquired infections, including respiratory syncytial virus (RSV), parainfluenza virus, influenza virus, and metapneumovirus, can be life threatening or fatal in the posttransplant patient. Protection of patients from infected visitors and staff by avoiding such contacts is critical. Neuraminidase inhibitors are effective for influenza infections. Oral or inhaled ribavirin is sometimes used for RSV.

The risk of infection diminishes considerably beyond 3 months after transplant unless chronic GVHD requiring continuous immunosuppression develops. Most transplant centers recommend continuing trimethoprim-sulfamethoxazole prophylaxis while patients are receiving any immunosuppressive drugs and also recommend careful monitoring for late CMV reactivation. In addition, many centers recommend prophylaxis against varicella-zoster, using acyclovir for 1 year after transplant. Antibody titers to vaccine-preventable diseases (e.g., tetanus, polio, mumps, rubella, and encapsulated organisms) decline after allogeneic or autologous transplantation if the recipient is not revaccinated. Vaccination begins at 3 months after transplantation for SAR-CoV-2 and 6 months for influenza (or 3–4 months when seasonal prevalence is high). Other nonlive routine childhood vaccinations should be repeated, usually starting at 12 months after transplantation. Live vaccines (measles, mumps, and rubella [MMR] or MMR plus varicella [MMR-V]) are generally not administered before 2 years after hematopoietic cell transplantation.

TREATMENT

Nonmalignant Diseases

Evidence-based indications for hematopoietic cell transplantation have been published by several organizations and are guided not only by disease-related factors but also by patient comorbidities, socioeconomic issues, caregiver and donor availability, and patient preference.

IMMUNODEFICIENCY DISORDERS

By replacing abnormal stem cells with cells from a normal donor, hematopoietic cell transplantation can cure patients of a variety of immunodeficiency disorders including severe combined immunodeficiency, Wiskott-Aldrich syndrome, and Chédiak-Higashi syndrome. The widest experience is with severe combined immunodeficiency disease, where cure rates of >90% can be expected with allogeneic transplantation from a suitable related or unrelated donor when carried out shortly after birth **(Table 119-4)**. Treatment of severe refractory autoimmune diseases with hematopoietic stem cell transplantation is also beginning to be explored (see below).

APLASTIC ANEMIA

Transplantation from matched siblings after a preparative regimen of high-dose cyclophosphamide and antithymocyte globulin cures >95% of patients age <40 years with severe aplastic anemia. Historically, results in older patients and in recipients of mismatched family member or unrelated marrow were less favorable, and therefore, a trial of immunosuppressive therapy was recommended for such patients before considering transplantation. However, results with transplantation have improved leading many to recommend transplantation as initial therapy. Transplantation is effective in all forms of aplastic anemia including, for example, the syndromes associated with paroxysmal nocturnal hemoglobinuria and Fanconi's anemia. Patients with Fanconi's anemia are abnormally sensitive to the toxic effects of alkylating agents, and so less intensive preparative regimens are used in their treatment **(Chap. 107)**.

HEMOGLOBINOPATHIES

Marrow transplantation from an HLA-identical sibling following a preparative regimen of busulfan and cyclophosphamide can cure

TABLE 119-4 Estimated 3-Year Survival Rates Following Transplantation[a]

DISEASE	ALLOGENEIC, %	AUTOLOGOUS, %
Severe combined immunodeficiency	95	NA
Aplastic anemia	98	NA
Thalassemia	95	NA
Acute myeloid leukemia		
First remission	57	ID
Second remission	53	ID
Acute lymphocytic leukemia		
First remission	64	ID
Second remission	45	ID
Chronic myeloid leukemia		
Chronic phase	71	NA
Accelerated phase	67	NA
Blast crisis	28	NA
Chronic lymphocytic leukemia	64	NA
Myelodysplasia	50	NA
Multiple myeloma—initial therapy	NA	80
Non-Hodgkin's lymphoma		
First relapse/second remission	45	55
Hodgkin's disease		
First relapse/second remission	68	89

[a]These estimates are mostly based on results of transplants performed between 2010 and 2020 reported by the Center for International Blood and Marrow Transplant Research (CIBMTR). The analysis has not been reviewed by their Advisory Committee.

Abbreviations: ID, insufficient data; NA, not applicable.

80–90% of patients with thalassemia major. The best outcomes can be expected if patients are transplanted before they develop hepatomegaly or portal fibrosis and if they have been given adequate iron chelation therapy. Among such patients, the probabilities of 5-year survival and disease-free survival are 95 and 90%, respectively. Although prolonged survival can be achieved with aggressive chelation therapy, transplantation is the only curative treatment for thalassemia. Transplantation is potentially curative for patients with sickle cell anemia. Two-year survival and disease-free survival rates of 95 and 85%, respectively, have been reported following matched sibling or cord blood transplantation. Decisions about patient selection and the timing of transplantation remain difficult, but transplantation is a reasonable option for children and young adults who have suffered complications of sickle cell anemia including stroke, recurrent vasoocclusive pain, sickle cell lung disease, or sickle nephropathy **(Chap. 103)**. As new gene therapies become available, the indications for allogeneic hematopoietic cell transplantation for thalassemia and sickle cell disease may change.

OTHER NONMALIGNANT DISEASES

Theoretically, hematopoietic cell transplantation should be able to cure any disease that results from an inborn error of the lymphohematopoietic system. Transplantation has been used successfully to treat congenital disorders of white blood cells such as Kostmann's syndrome, chronic granulomatous disease, and leukocyte adhesion deficiency. Congenital anemias such as Blackfan-Diamond anemia can also be cured with transplantation. Since the penetrance of some congenital marrow failure states is variable, potential family member donors should be carefully screened before use to assure they are not affected. Infantile malignant osteopetrosis is due to an inability of the osteoclast to resorb bone, and because osteoclasts derive from the marrow, transplantation can cure this rare inherited disorder.

Hematopoietic cell transplantation has been used as treatment for several storage diseases caused by enzymatic deficiencies, such as Gaucher's disease, Hurler's syndrome, Hunter's syndrome, and infantile metachromatic leukodystrophy. Transplantation for these diseases has not been uniformly successful, but treatment early in the course of these diseases, before irreversible damage to extramedullary organs has occurred, increases the chance for success.

Transplantation is being applied as a treatment for severe acquired autoimmune disorders. These approaches are based on studies demonstrating that transplantation can reverse autoimmune disorders in animal models and on the observation that occasional patients with coexisting autoimmune disorders and hematologic malignancies have been cured of both with transplantation. A prospective randomized trial found that patients with severe scleroderma have improved event-free and overall survival if treated with hematopoietic cell transplantation. Randomized studies are exploring a similar approach for patients with multiple sclerosis.

ACUTE LEUKEMIA

Allogeneic hematopoietic cell transplantation cures ~30% of patients who do not achieve a complete response after induction chemotherapy for acute myeloid leukemia (AML) and is the only form of therapy that can cure such patients. Thus, all patients with AML who are possible transplant candidates should have their HLA type determined soon after diagnosis to enable hematopoietic cell transplantation for those who fail to enter remission. Cure rates of 45–50% are seen when patients are transplanted in second remission or in first relapse. The best results with allogeneic transplantation are achieved when applied during first remission, with long-term disease-free survival rates averaging 55–60%. Meta-analyses of studies comparing matched related donor transplantation to chemotherapy for adult AML patients age <60 years show a survival advantage with transplantation. This advantage is greatest for those with adverse and intermediate-risk disease but is not seen in patients with favorable-risk AML. Some centers rely on measurements of minimal residual disease (MRD) as determined by either multidimensional flow cytometry or molecular methods to further define transplant candidacy, proceeding with transplantation in otherwise favorable risk patients if MRD positive and withholding transplantation in MRD-negative intermediate-risk patients. While hematopoietic cell transplantation can be performed in patients up to age 80, prospective trials comparing hematopoietic cell transplantation with chemotherapy are lacking for older patients. Autologous transplantation has no defined role in the treatment of AML.

Similar to patients with AML, adults with acute lymphocytic leukemia who do not achieve a complete response to induction chemotherapy can be cured in ~30% of cases with immediate transplantation. Cure rates improve to 40–50% in second remission, and therefore, transplantation can be recommended for adults who have persistent disease after induction chemotherapy or who subsequently relapse. Transplant outcomes in second remission are improved if carried out when MRD assessments are negative, and so use of agents such as blinatumomab to achieve an MRD-negative state before transplantation is recommended. Transplantation in first remission results in cure rates of about 65%. Transplantation appears to offer a survival advantage over chemotherapy for patients with high-risk disease as defined by molecular profiling. Debate continues about whether adults with standard-risk disease should be transplanted in first remission or whether transplantation should be reserved until relapse. Autologous transplantation is associated with a higher relapse rate but a somewhat lower risk of nonrelapse mortality when compared to allogeneic transplantation. Autologous transplantation has no obvious role in treatment for acute lymphocytic leukemia in first remission, and for second-remission patients, most experts recommend use of allogeneic stem cells if an appropriate donor is available.

CHRONIC LEUKEMIA

Allogeneic hematopoietic cell transplantation is indicated for patients with chronic myeloid leukemia (CML) who are in chronic phase but have failed therapy with two or more tyrosine kinase inhibitors. In such patients, cure rates of 70% can be expected.

Hematopoietic cell transplantation is also recommended for patients with CML who present or progress to accelerated phase or blast crisis, although lower cure rates are seen in such patients **(Chap. 110)**.

Although allogeneic transplantation can cure patients with chronic lymphocytic leukemia (CLL), it has not been extensively studied because of the chronic nature of the disease, the age profile of patients, and more recently, the availability of multiple effective therapies. In those cases where it was studied, complete remissions were achieved in the majority of patients, with disease-free survival rates of ~65% at 3 years, despite the advanced stage of the disease at the time of transplant.

MYELODYSPLASIA AND MYELOPROLIFERATIVE DISORDERS

Between 40 and 65% of patients with myelodysplasia can be cured with allogeneic transplantation. Results are better among younger patients and those with less advanced disease. However, patients with early-stage myelodysplasia can live for extended periods without intervention, and so transplantation is generally reserved for patients with an International Prognostic Scoring System (IPSS) score of Int-2 or higher, or for selected patients with an IPSS score of Int-1 who have other poor prognostic features **(Chap. 107)**. Allogeneic hematopoietic cell transplantation can cure patients with primary myelofibrosis or myelofibrosis secondary to polycythemia vera or essential thrombocythemia, with 5-year progression-free survival rates in excess of 65% being reported. It may require many months for the fibrosis to resolve.

LYMPHOMA

Patients with disseminated intermediate- or high-grade non-Hodgkin's lymphoma who have not been cured by first-line chemotherapy and are transplanted in first relapse or second remission can still be cured in 50–60% of cases. This represents a clear advantage over results obtained with conventional-dose salvage chemotherapy. It is unsettled whether patients with high-risk disease benefit from transplantation in first remission. Most experts favor the use of autologous rather than allogeneic transplantation for patients with intermediate- or high-grade non-Hodgkin's lymphoma, because fewer complications occur with this approach and survival appears equivalent. The use of chimeric antigen receptor T cells targeting CD19 has been reported to yield results similar to those achieved with autologous stem cell transplantation. As yet, no consensus has been reached about how to sequence these two therapies. Although autologous transplantation results in high response rates in patients with recurrent disseminated indolent non-Hodgkin's lymphoma, the availability of newer agents for this category of patient leaves the role of transplantation unsettled. Reduced-intensity conditioning regimens followed by allogeneic transplantation result in high rates of complete and enduring complete responses in patients with recurrent indolent lymphomas.

The role of transplantation in Hodgkin's lymphoma is similar to that in intermediate- and high-grade non-Hodgkin's lymphoma. With transplantation, 3-year disease-free survival is 40–50% in patients who never achieved a first remission with standard chemotherapy and up to 80% for those transplanted in second remission. Transplantation has no defined role in first remission in Hodgkin's lymphoma.

MYELOMA

Patients with myeloma whose disease progresses after first-line therapy can sometimes benefit from allogeneic or autologous transplantation. Prospective randomized studies demonstrate that the inclusion of autologous transplantation as part of initial therapy results in improved disease-free survival and overall survival. Further benefit is seen with the use of lenalidomide maintenance therapy following transplantation. The use of autologous transplantation followed by nonmyeloablative allogeneic transplantation has yielded mixed results.

SOLID TUMORS

Patients with testicular cancer in whom first-line platinum-containing chemotherapy has failed can still be cured in ~50% of cases if treated with high-dose chemotherapy with autologous stem cell support, an outcome better than that seen with low-dose salvage chemotherapy. The use of high-dose chemotherapy with autologous stem cell support is being studied for several other solid tumors, including neuroblastoma and pediatric sarcomas. As in most other settings, the best results were obtained in patients with limited amounts of disease and in whom the remaining tumor remains sensitive to conventional-dose chemotherapy. Few randomized trials of transplantation in these diseases have been completed.

POSTTRANSPLANT RELAPSE

Patients who relapse following autologous transplantation sometimes respond to further chemotherapy and may be candidates for possible allogeneic transplantation, particularly if the remission following the initial autologous transplant was long. Several options are available for patients who relapse following allogeneic transplantation. Treatment with infusions of unirradiated donor lymphocytes results in complete responses in as many as 75% of patients with chronic myeloid leukemia, 40% with myelodysplasia, 25% with AML, and 15% with myeloma. Major complications of donor lymphocyte infusions include transient myelosuppression and the development of GVHD. These complications depend on the number of donor lymphocytes given and the schedule of infusions, with less GVHD seen with lower dose, fractionated schedules.

■ FURTHER READING

Dadwal SS et al: How I prevent viral reactivation in high-risk patients. Blood 141:2062, 2023.

DeFilipp Z et al: Hematopoietic cell transplantation in the treatment of adult acute lymphoblastic leukemia: Updated 2019 evidence-based review from the American Society for Transplantation and Cellular Therapy. Biol Blood Marrow Transplant 25:2113, 2019.

Duarte RF et al: Indications for haematopoietic stem cell transplantation for haematological diseases, solid tumours and immune disorders: current practice in Europe, 2019. Bone Marrow Transplantation 54:1525, 2019.

Jamy O et al: Novel developments in the prophylaxis and treatment of acute GVHD. Blood 142:1037, 2023.

McDonald GB et al: Survival, nonrelapse mortality, and relapse-related mortality after allogeneic hematopoietic cell transplantation: Comparing 2003-2007 versus 2013-2017 cohorts. Ann Intern Med 172:229, 2020.

Miller PDE et al: Joint consensus statement on the vaccination of adult and paediatric haematopoietic stem cell transplant recipients: Prepared on behalf of the British Society of Blood and Marrow Transplantation and Cellular Therapy (BSBMTCT), the Children's Cancer and Leukaemia Group (CCLG), and British Infection Association (BIA). J Infect 86:1, 2023.

Niederwieser D et al: One and a half million hematopoietc stem cell transplants: Continuous and differential improvement in worldwide access with the use of non-identical family donors. Haematologica 107:1045, 2022.

Scott BL et al: Myeloablative versus reduced-intensity conditioning for hematopoietic cell transplantation in acute myelogenous leukemia and myelodysplastic syndromes: Long-term follow-up of the BMT CTN 0901 clinical trial. Transplant Cell Ther 27:483, 2021.

Westin J, Sehn LH: CAR T cells as a second-line therapy for large B-cell lymphoma: A paradigm shift? Blood 139:2737, 2022.

Zeiser R, Lee SJ: Three Food and Drug Administration-approved therapies for chronic GVHD. Blood 139:1642, 2022.

Section 3 Disorders of Hemostasis

120 Disorders of Platelets and Vessel Wall

Barbara A. Konkle

Hemostasis is a dynamic process in which the platelet and the blood vessel wall play key roles. Platelets are activated upon adhesion to von Willebrand factor (VWF) and collagen in the exposed subendothelium after injury. Platelet activation is also mediated through shear forces imposed by blood flow itself, particularly in areas where the vessel wall is diseased, and is also affected by the inflammatory state of the endothelium. The activated platelet surface provides the major physiologic site for coagulation factor activation, which results in further platelet activation and fibrin formation. Genetic and acquired influences on the platelet and vessel wall, as well as on the coagulation and fibrinolytic systems, determine whether normal hemostasis or bleeding or clotting symptoms will result.

THE PLATELET

Platelets are released from the megakaryocyte, likely under the influence of flow in the capillary sinuses. The normal blood platelet count is 150,000–450,000/μL. The major regulator of platelet production is the hormone thrombopoietin (TPO), which is synthesized in the liver and other organs. Synthesis is increased with inflammation and specifically by interleukin 6. TPO binds to its receptor on platelets and megakaryocytes, by which it is removed from the circulation. Thus, a reduction in platelet and megakaryocyte mass increases the level of TPO, which then stimulates platelet production. Platelets circulate with an average life span of 7–10 days. Approximately one-third of the platelets reside in the spleen, and this number increases in proportion to splenic size, although the platelet count rarely decreases to <40,000/μL as the spleen enlarges. Platelets are physiologically very active, but are anucleate, and thus have limited capacity to synthesize new proteins.

Normal vascular endothelium contributes to preventing thrombosis by inhibiting platelet function **(Chap. 69)**. When vascular endothelium is injured, these inhibitory effects are overcome, and platelets adhere to the exposed intimal surface primarily through VWF, a large multimeric protein present in both plasma and in the extracellular matrix of the subendothelial vessel wall. Platelet adhesion results in the generation of intracellular signals that lead to activation of the platelet glycoprotein (Gp) IIb/IIIa ($\alpha_{\text{IIb}}\beta_3$) receptor and resultant platelet aggregation.

Activated platelets undergo release of their granule contents, which include nucleotides, adhesive proteins, growth factors, and procoagulants that serve to promote platelet aggregation and blood clot formation and influence the environment of the forming clot. During platelet aggregation, additional platelets are recruited to the site of injury, leading to the formation of an occlusive platelet thrombus. The platelet plug is stabilized by the fibrin mesh that develops simultaneously as the product of the coagulation cascade.

THE VESSEL WALL

Endothelial cells line the surface of the entire circulatory tree, totaling 1–6 × 10^{13} cells, enough to cover a surface area equivalent to about six tennis courts. The endothelium is physiologically active, controlling vascular permeability, flow of biologically active molecules and nutrients, blood cell interactions with the vessel wall, the inflammatory response, and angiogenesis.

The endothelium normally presents an antithrombotic surface **(Chap. 69)** but rapidly becomes prothrombotic when stimulated, which promotes coagulation, inhibits fibrinolysis, and activates platelets. In many cases, endothelium-derived vasodilators are also platelet inhibitors (e.g., nitric oxide), and conversely, endothelium-derived vasoconstrictors (e.g., endothelin) can also be platelet activators. The net effect of vasodilation and inhibition of platelet function is to promote blood fluidity, whereas the net effect of vasoconstriction and platelet activation is to promote thrombosis. Thus, blood fluidity and hemostasis are regulated by the balance of antithrombotic/prothrombotic and vasodilatory/vasoconstrictor properties of endothelial cells.

DISORDERS OF PLATELETS

THROMBOCYTOPENIA

Thrombocytopenia results from one or more of three processes: (1) decreased bone marrow production; (2) sequestration, usually in an enlarged spleen; and/or (3) increased platelet destruction. Disorders of production may be either inherited or acquired. In evaluating a patient with thrombocytopenia, a key step is to review the peripheral blood smear and to first rule out "pseudothrombocytopenia," particularly in a patient without an apparent cause for the thrombocytopenia. Pseudothrombocytopenia **(Fig. 120-1*B*)** is an in vitro artifact resulting from platelet agglutination via antibodies (usually IgG, but also IgM and IgA) when the calcium content is decreased by blood collection in ethylenediamine tetraacetic (EDTA) (the anticoagulant present in tubes [purple top] used to collect blood for complete blood counts [CBCs]). If a low platelet count is obtained in EDTA-anticoagulated blood, a blood smear should be evaluated and a platelet count determined in blood collected into sodium citrate (blue top tube) or heparin (green top tube), or a smear of freshly obtained unanticoagulated blood, such as from a finger stick, can be examined.

APPROACH TO THE PATIENT

Thrombocytopenia

The history and physical examination, results of the CBC, and review of the peripheral blood smear are all critical components in the initial evaluation of thrombocytopenic patients **(Fig. 120-2)**. The overall health of the patient and whether they are receiving drug treatment will influence the differential diagnosis. A healthy young adult with thrombocytopenia will have a much more limited differential diagnosis than an ill hospitalized patient who is receiving multiple medications. Except in less common inherited disorders, decreased platelet production usually results from bone marrow disorders that also affect red blood cell (RBC) and/or white blood cell (WBC) production. Because myelodysplasia can present with isolated thrombocytopenia, the bone marrow should be examined in patients presenting with isolated thrombocytopenia who are older than 60 years of age or who do not respond to initial therapy. While inherited thrombocytopenia is uncommon, any prior platelet counts should be retrieved and a family history regarding thrombocytopenia obtained. A careful history of drug ingestion should be obtained, including nonprescription and herbal remedies, because drugs are the most common cause of thrombocytopenia.

The physical examination can document an enlarged spleen, evidence of chronic liver disease, and other underlying disorders. Mild to moderate splenomegaly may be difficult to appreciate in many individuals due to body habitus and/or obesity but can be easily assessed by abdominal ultrasound. A platelet count of approximately 5000–10,000 is required to maintain vascular integrity in the microcirculation. When the count is markedly decreased, petechiae first appear in areas of increased venous pressure, the ankles and feet in an ambulatory patient. Petechiae are pinpoint, nonblanching hemorrhages and are usually a sign of a decreased platelet number and not platelet dysfunction. Wet purpura, blood blisters that form on the oral mucosa, are thought to denote an increased risk of life-threatening hemorrhage in the thrombocytopenic patient. Excessive bruising is seen in disorders of both platelet number and function.

Infection-Induced Thrombocytopenia Many viral and bacterial infections result in thrombocytopenia and are the most common

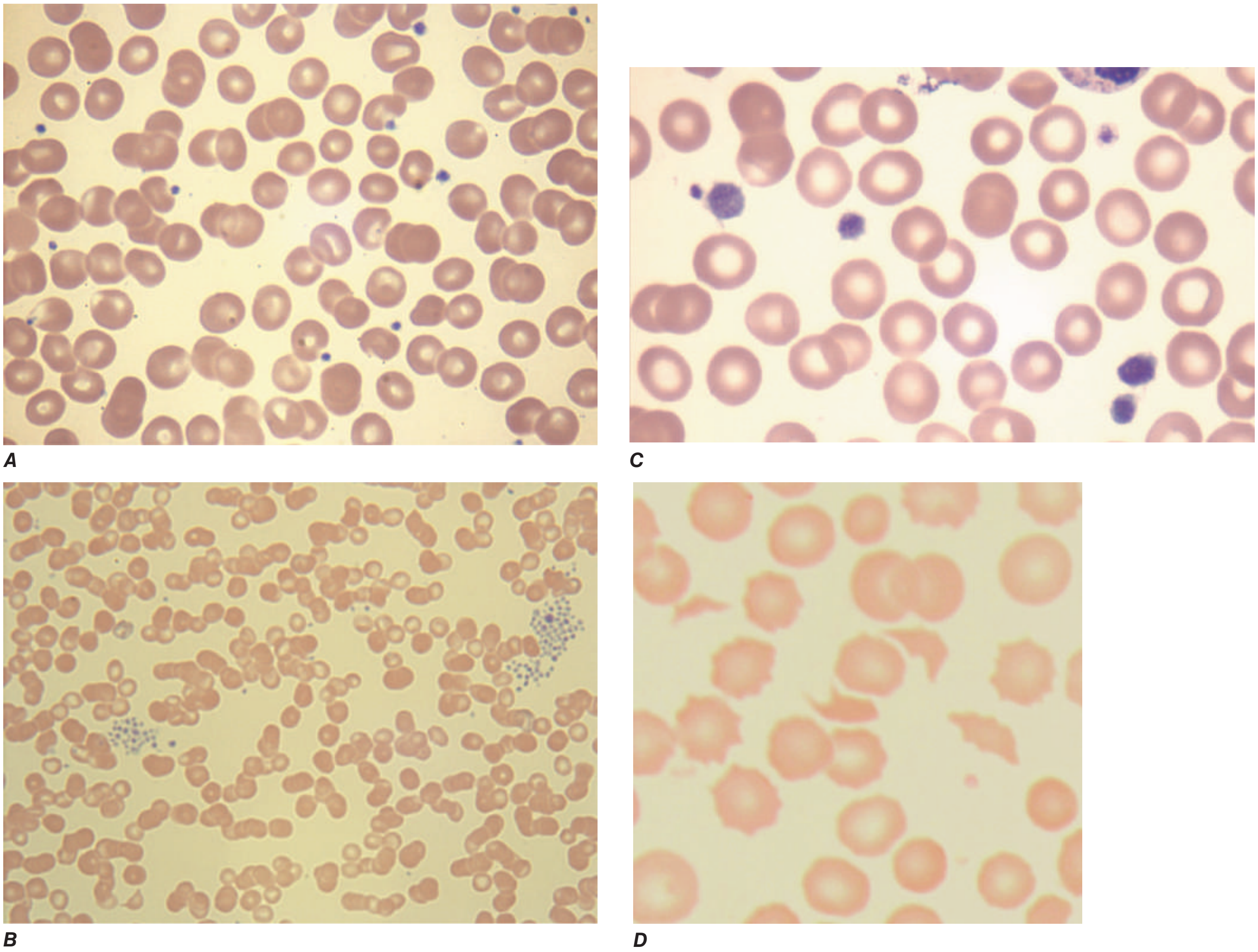

FIGURE 120-1 Photomicrographs of peripheral blood smears. ***A.*** Normal peripheral blood. ***B.*** Platelet clumping in pseudothrombocytopenia. ***C.*** Abnormal large platelet in autosomal dominant macrothrombocytopenia. ***D.*** Schistocytes and decreased platelets in microangiopathic hemolytic anemia.

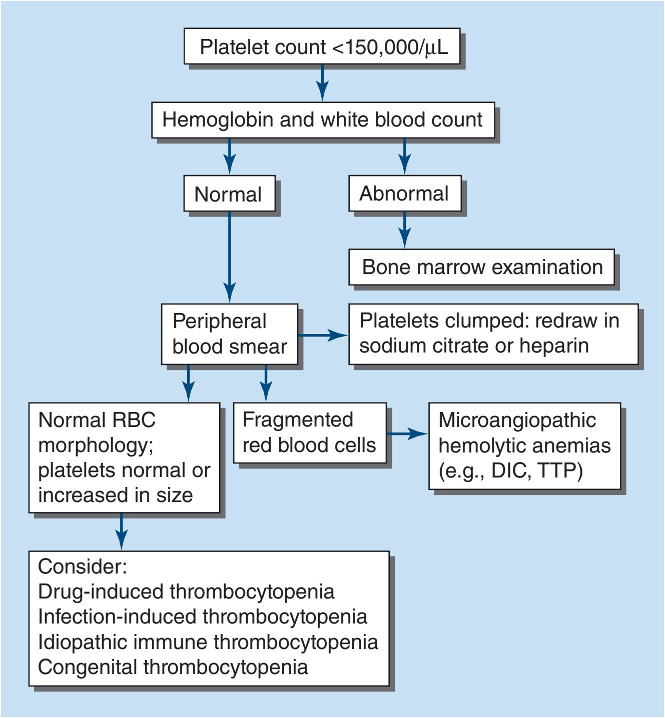
Platelet count <150,000/µL
Hemoglobin and white blood count
Normal
Abnormal
Bone marrow examination
Peripheral blood smear
Platelets clumped: redraw in sodium citrate or heparin
Normal RBC morphology; platelets normal or increased in size
Fragmented red blood cells
Microangiopathic hemolytic anemias (e.g., DIC, TTP)
Consider:
Drug-induced thrombocytopenia
Infection-induced thrombocytopenia
Idiopathic immune thrombocytopenia
Congenital thrombocytopenia

FIGURE 120-2 Algorithm for evaluating the thrombocytopenic patient. DIC, disseminated intravascular coagulation; RBC, red blood cell; TTP, thrombotic thrombocytopenic purpura.

noniatrogenic cause of thrombocytopenia. This may or may not be associated with laboratory evidence of disseminated intravascular coagulation (DIC), which is most commonly seen in patients with systemic infections with gram-negative bacteria and is seen in patients ill with COVID-19. Infections can affect both platelet production and platelet survival. In addition, immune mechanisms can be at work, as in infectious mononucleosis and early HIV infection. Late in HIV infection, pancytopenia and decreased and dysplastic platelet production are more common. Immune-mediated thrombocytopenia in children usually follows a viral infection and almost always resolves spontaneously. This association of infection with immune thrombocytopenic purpura is less clear in adults.

Drug-Induced Thrombocytopenia Many drugs have been associated with thrombocytopenia. A predictable decrease in platelet count occurs after treatment with many chemotherapeutic drugs due to bone marrow suppression (Chap. 78). Drugs that cause isolated thrombocytopenia and have been confirmed with positive laboratory testing are listed in Table 120-1, but all drugs should be suspect in a patient with thrombocytopenia without an apparent cause and should be stopped, or substituted, if possible. Although not as well studied, herbal and over-the-counter preparations may also result in thrombocytopenia and should be discontinued in patients who are thrombocytopenic.

Classic drug-dependent antibodies are antibodies that react with specific platelet surface antigens and result in thrombocytopenia only when the drug is present. Many drugs are capable of inducing these antibodies, but for some reason, they are more common with quinine and sulfonamides. Drug-dependent antibody binding can be demonstrated by laboratory assays, showing antibody binding in the presence

TABLE 120-1 Drugs Reported as Definitely or Probably Causing Isolated Thrombocytopenia[a]

- Abciximab
- Acetaminophen
- Amiodarone
- Amlodipine
- Ampicillin
- Carbamazepine
- Ceftriaxone
- Cephamandole
- Ciprofloxacin
- Diazepam
- Eptifibatide
- Furosemide
- Gold
- Haloperidol
- Heparin
- Ibuprofen
- Lorazepam
- Mirtazapine
- Naproxen
- Oxaliplatin
- Penicillin
- Phenytoin
- Piperacillin
- Quinidine
- Quinine
- Ranitidine
- Rosiglitazone
- Roxifiban
- Sulfisoxazole
- Suramin
- Tirofiban
- Tranilast
- Trimethoprim/sulfamethoxazole
- Vancomycin

[a] Based on scoring requiring a compatible clinical picture and positive laboratory testing. Note, this is not inclusive of newer medications.

Source: Adapted from DM Arnold et al: J Thromb Hemost 11:169, 2013.

of, but not without, the drug present in the assay. The thrombocytopenia typically occurs after a period of initial exposure (median length 21 days), or upon reexposure, and usually resolves in 7–10 days after drug withdrawal. The thrombocytopenia caused by the platelet Gp IIb/IIIa inhibitory drugs, such as abciximab, differs in that it may occur within 24 h of initial exposure. This appears to be due to the presence of naturally occurring antibodies that cross-react with the drug bound to the platelet.

Heparin-Induced Thrombocytopenia Drug-induced thrombocytopenia due to heparin differs from that seen with other drugs in two major ways. (1) The thrombocytopenia is not usually severe, with nadir counts rarely <20,000/μL. (2) Heparin-induced thrombocytopenia (HIT) is not associated with bleeding and, in fact, markedly increases the risk of thrombosis. The pathogenesis of HIT is complex. It results from antibody formation to a complex of the platelet-specific protein platelet factor 4 (PF4) and heparin or other glycosaminoglycans. The anti-heparin/PF4 antibody can activate platelets through the FcγRIIa receptor and also activate monocytes, endothelial cells, and coagulation proteins. Many patients exposed to heparin develop antibodies to heparin/PF4 but do not appear to have adverse consequences. A fraction of those who develop antibodies will develop HIT, and a portion of those (up to 50%) will develop thrombosis (HITT).

HIT can occur after exposure to low-molecular-weight heparin (LMWH) as well as unfractionated heparin (UFH), although it is more common with the latter. Most patients develop HIT after exposure to heparin for 5–14 days **(Fig. 120-3)**. It occurs before 5 days in those who were exposed to heparin in the prior few weeks or months (<~100 days) and have circulating anti-heparin/PF4 antibodies. Rarely, thrombocytopenia and thrombosis begin several days after all heparin has been stopped (termed *delayed-onset HIT*), and more rarely, spontaneous HIT, or autoimmune HIT syndrome, occurs where there is no history of heparin exposure. A syndrome similar to spontaneous HIT has been described rarely after COVID-19 vaccination with the adenovirus-based ChAdOx1-S/nCoV-19 vaccine and termed *vaccine-induced immune thrombocytopenia and thrombosis* (VITT). Even more rarely, a similar thrombotic syndrome can follow an adenovirus infection. then go to the 4Ts. The "4Ts" are recommended to be used in a diagnostic algorithm for HIT: *t*hrombocytopenia, *t*iming of platelet count drop, *t*hrombosis and other sequelae such as localized skin reactions, and o*t*her causes of thrombocytopenia not evident. Application of the 4T scoring system is very useful in excluding a diagnosis of HIT but will result in overdiagnosis of HIT in situations where thrombocytopenia and thrombosis due to other etiologies are common, such as in the intensive care unit.

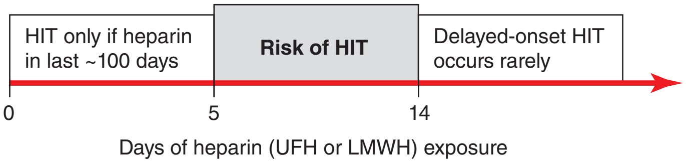
HIT only if heparin in last ~100 days
Risk of HIT
Delayed-onset HIT occurs rarely
0
5
14
Days of heparin (UFH or LMWH) exposure

FIGURE 120-3 Time course of heparin-induced thrombocytopenia (HIT) development after heparin exposure. The timing of development after heparin exposure is a critical factor in determining the likelihood of HIT in a patient. HIT occurs early after heparin exposure in the presence of preexisting heparin/platelet factor 4 (PF4) antibodies, which disappear from circulation by ~100 days following a prior exposure. Rarely, HIT may occur later after heparin exposure (termed delayed-onset HIT). In this setting, heparin/PF4 antibody testing is usually markedly positive. HIT can occur after exposure to either unfractionated (UFH) or low-molecular-weight heparin (LMWH).

LABORATORY TESTING FOR HIT Because of the prevalence of anti-heparin antibodies without clinical disease, testing should be done in individuals who are at intermediate or high risk based on clinical pretest assessment. HIT (anti-heparin/PF4) antibodies can be detected using two types of assays. The most widely available are immunoassays usually with PF4/polyanion complex as the antigen. Because many patients develop antibodies but do not develop clinical HIT, the test has a low specificity for the diagnosis of HIT. This is especially true in patients who have undergone surgery requiring cardiopulmonary bypass, where approximately 50% of patients develop these antibodies postoperatively. The other assay is a platelet activation assay, most commonly the serotonin release assay, which measures the ability of the patient's serum to activate platelets in the presence of heparin in a concentration-dependent manner. This test has lower sensitivity but higher specificity than the enzyme-linked immunosorbent assay (ELISA). However, HIT remains a clinical diagnosis.

TREATMENT

Heparin-Induced Thrombocytopenia

Early recognition is key in treatment of HIT, with prompt discontinuation of heparin and use of alternative anticoagulants if bleeding risk does not outweigh thrombotic risk. Thrombosis is a common complication of HIT, even after heparin discontinuation, and can occur in both the venous and arterial systems. In patients diagnosed with HIT, imaging studies to evaluate the patient for thrombosis (at least lower extremity duplex Doppler imaging) are recommended. Patients requiring anticoagulation should be switched from heparin to an alternative anticoagulant. The direct thrombin inhibitor (DTI) argatroban is effective in HITT. The DTI bivalirudin, the antithrombin-binding pentasaccharide fondaparinux, and direct oral anticoagulants (DOACs) appear to have efficacy, although none of these are approved by the U.S. Food and Drug Administration (FDA) for this indication. Of the DOACs, the most experience reported is with rivaroxaban. In general, an intravenous DTI should be used in acutely ill patients with transition to fondaparinux or a DOAC when they are more stable, which then allows outpatient treatment. HIT antibodies cross-react with LMWH, and these drugs should not be used in the treatment of HIT.

Because of the high rate of thrombosis in patients with HIT, anticoagulation should be considered, even in the absence of thrombosis. In patients with thrombosis, anticoagulation is continued for 3–6 months, but in patients without thrombosis, the duration of anticoagulation is less well defined, but should be continued at least until platelet recovery. An increased risk of thrombosis is present for at least 1 month after diagnosis; however, most thromboses occur early, and whether thrombosis occurs later if the patient is initially anticoagulated is unknown. Introduction of warfarin alone in the setting of HIT or HITT may precipitate thrombosis, particularly venous gangrene, presumably due to clotting activation and severely reduced levels of proteins C and S. Warfarin therapy, if started, should be overlapped with a DTI, fondaparinux, or a DOAC and started after resolution of the thrombocytopenia and lessening of the prothrombotic state.

The rare VITT syndrome is characterized by high D-dimer levels and thrombosis in unusual sites like the cerebral venous sinuses. Fatal in about 20%, treatment is usually intravenous gamma globulin (IVIgG) to block platelet activation through Fc receptors, the pathogenic effect of the anti–PF4-polyanion antibody, and a nonheparin anticoagulant.

Immune Thrombocytopenic Purpura Immune thrombocytopenic purpura (ITP; also termed *idiopathic thrombocytopenic purpura*) is an acquired disorder in which there is immune-mediated destruction of platelets and possibly inhibition of platelet release from the megakaryocyte. In children, it is usually an acute disease, most commonly following an infection, and with a self-limited course. In adults, it is a more chronic disease, although in some adults, spontaneous remission occurs, usually within months of diagnosis. ITP is termed *secondary* if it is associated with an underlying disorder; autoimmune disorders, particularly systemic lupus erythematosus (SLE), and infections, such as HIV and hepatitis C, are common causes. The association of ITP with *Helicobacter pylori* infection is unclear but appears to have a geographic distribution.

ITP is characterized by mucocutaneous bleeding and a low, often very low, platelet count, with an otherwise normal peripheral blood cells and smear. Patients usually present either with ecchymoses and petechiae or with thrombocytopenia incidentally found on a routine CBC. Mucocutaneous bleeding, such as oral mucosa, gastrointestinal, or heavy menstrual bleeding, may be present. Rarely, life-threatening, including central nervous system, bleeding can occur. Wet purpura (blood blisters in the mouth) and retinal hemorrhages may herald life-threatening bleeding.

LABORATORY TESTING IN ITP Laboratory testing for antibodies (serologic testing) is usually not helpful due to the low sensitivity and specificity of the current tests. Bone marrow examination can be reserved for those who have other signs or laboratory abnormalities not explained by ITP or in patients who do not respond to initial therapy. The peripheral blood smear may show large platelets, with otherwise normal morphology. Depending on the bleeding history, iron-deficiency anemia may be present.

Laboratory testing is performed to evaluate for secondary causes of ITP and should include testing for HIV infection and hepatitis C (and other infections if indicated). Serologic testing for SLE, serum protein electrophoresis, immunoglobulin levels to potentially detect hypogammaglobulinemia, selective testing for IgA deficiency or monoclonal gammopathies, and testing for *H. pylori* infection should be considered, depending on the clinical circumstance. If anemia is present, direct antiglobulin testing (Coombs' test) should be performed to rule out combined autoimmune hemolytic anemia with ITP (Evans' syndrome).

TREATMENT

Immune Thrombocytopenic Purpura

The treatment of ITP uses drugs that decrease reticuloendothelial uptake of the antibody-bound platelet, decrease antibody production, and/or increase platelet production. The diagnosis of ITP does not necessarily mean that treatment must be instituted. Patients with platelet counts >30,000/μL appear not to have increased mortality related to the thrombocytopenia.

Initial treatment in patients without significant bleeding symptoms, severe thrombocytopenia (<5000/μL), or signs of impending bleeding (e.g., retinal hemorrhage or large oral mucosal hemorrhages) can be instituted as an outpatient using single agents. Traditionally, this has been prednisone at 1 mg/kg or a 4-day course of dexamethasone, 40 mg/d, although Rh_0(D) immune globulin therapy, at 50–75 μg/kg, is also being used in this setting. Rh_0(D) immune globulin must be used only in Rh-positive patients because the mechanism of action is production of limited hemolysis, with antibody-coated cells "saturating" the Fc receptors, inhibiting Fc receptor function. Monitoring patients for 8 h after infusion is now advised by the FDA because of the rare complication of severe intravascular hemolysis. IVIgG, which is pooled, primarily IgG antibodies, also blocks the Fc receptor system, but appears to work primarily through different mechanism(s). IVIgG has more efficacy than anti-Rh_0(D) in postsplenectomized patients. IVIgG is dosed at 1–2 g/kg total, given over 1–5 days. Side effects are usually related to the volume of infusion and infrequently include aseptic meningitis and renal failure. All immunoglobulin preparations are derived from human plasma and undergo treatment for viral inactivation. Rituximab, an anti-CD20 (B-cell) antibody, has shown efficacy in the treatment of refractory ITP, although long-lasting remission only occurs in approximately 30% of patients. TPO receptor agonists (TPO-RA), administered subcutaneously (romiplostim) or orally (eltrombopag, avatrombopag), are effective in raising platelet counts in patients with chronic ITP.

For patients with severe ITP and/or symptoms of bleeding, hospital admission is required, and combined-modality therapy is given using high-dose glucocorticoids with IVIgG or anti-Rh_0(D) therapy and, as needed, additional immunosuppressive agents.

For chronic ITP, broad immune therapy and splenectomy are now generally used less frequently compared to TPO-RAs due to unfavorable side effects and a negative impact on quality of life. Newer drugs are under study. In individuals who achieve a complete response to a TPO-RA, many will maintain a sustained response off therapy.

In patients who are to undergo splenectomy, vaccination against encapsulated organisms (especially pneumococcus, but also meningococcus and *Haemophilus influenzae*, depending on patient age and potential exposure) is recommended. Accessory spleens are a very rare cause of relapse.

Inherited Thrombocytopenia Thrombocytopenia is rarely inherited, either as an isolated finding or as part of a syndrome, and may be inherited in an autosomal dominant, autosomal recessive, or X-linked pattern. Multiple genetic variants have now been identified in individuals with isolated thrombocytopenia and cytopenic syndromes, some of which carry an increased risk of hematologic malignancy. Many forms of autosomal dominant macrothrombocytopenia are now known to be associated with variants in the nonmuscle myosin heavy chain *MYH9* gene. These include the May-Hegglin anomaly, and Sebastian, Epstein's, and Fechtner syndromes, all of which have distinct distinguishing features. A common feature of these disorders is large platelets **(Fig. 120-1*C*)**. It is important that family history be explored in any individual with unexplained thrombocytopenia.

■ THROMBOTIC THROMBOCYTOPENIC PURPURA AND HEMOLYTIC-UREMIC SYNDROME

Thrombotic thrombocytopenic microangiopathies are a group of disorders characterized by microangiopathic hemolytic anemia (MAHA) defined by thrombocytopenia and fragmented RBCs **(Fig. 120-1*D*)** on peripheral blood smear, laboratory evidence of hemolysis (elevated lactate dehydrogenase [LDH] and unconjugated bilirubin and decreased haptoglobin), and microvascular thrombosis. They include thrombotic thrombocytopenic purpura (TTP) and hemolytic-uremic syndrome (HUS), as well as syndromes complicating bone marrow transplantation, certain medications and infections, pregnancy, and vasculitis. In DIC, although thrombocytopenia and microangiopathy are seen, a coagulopathy predominates, with consumption of clotting factors and fibrinogen resulting in an elevated prothrombin time (PT) and often activated partial thromboplastin time (aPTT). The PT and aPTT are characteristically normal in TTP or HUS.

Thrombotic Thrombocytopenic Purpura TTP was first described in 1924 by Eli Moschcowitz and characterized by a pentad of findings that include microangiopathic hemolytic anemia, thrombocytopenia, renal failure, neurologic findings, and fever. The full-blown

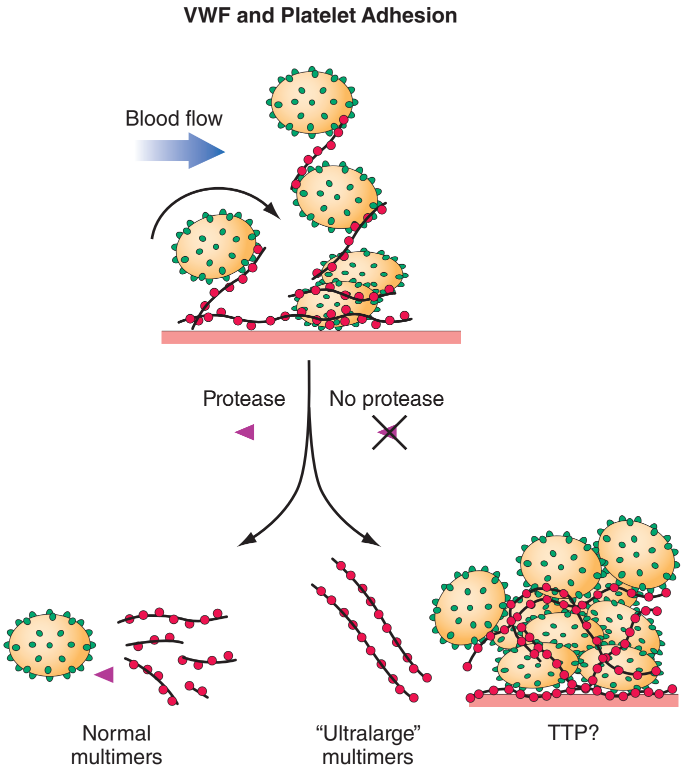

FIGURE 120-4 Pathogenesis of thrombotic thrombocytopenic purpura (TTP). Normally the ultra-high-molecular-weight multimers of von Willebrand factor (VWF) produced by the endothelial cells are processed into smaller multimers by a plasma metalloproteinase called ADAMTS13. In TTP, the activity of the protease is inhibited, and the ultra-high-molecular-weight multimers of VWF initiate platelet aggregation and thrombosis.

syndrome is less commonly seen now, probably due to earlier diagnosis. The introduction of treatment with plasma exchange markedly improved the prognosis in patients, with a decrease in mortality from 85–100% to 10–30%.

The pathogenesis of inherited (Upshaw-Schulman syndrome) and idiopathic TTP (ITTP) is related to a deficiency of, or antibodies to, the metalloprotease ADAMTS13, which cleaves VWF. VWF is normally secreted as ultra-large multimers, which are then cleaved by ADAMTS13. The persistence of ultra-large VWF molecules is thought to contribute to pathogenic platelet adhesion and aggregation **(Fig. 120-4)**. This defect alone, however, is not sufficient to result in TTP because individuals with a congenital absence of ADAMTS13 develop TTP only episodically, including during first pregnancy. The level of ADAMTS13 activity, as well as antibodies to ADAMTS13, can be detected by laboratory assays, which play a critical role in the differential diagnosis of MAHA. ADAMTS13 activity levels of <10% are diagnostic of TTP.

Idiopathic TTP appears to be more common in women than in men. No geographic or racial distribution has been defined. TTP is more common in patients with HIV infection and in pregnant women. Medication-related MAHA may be secondary to antibody formation (ticlopidine and possibly clopidogrel) or direct endothelial toxicity (cyclosporine, gemcitabine, mitomycin C, tacrolimus), although this is not always so clear, and fear of withholding treatment, as well as lack of other treatment alternatives, may result in initial application of plasma exchange. However, withdrawal, or reduction in dose, of endothelial toxic agents usually decreases the microangiopathy.

TREATMENT

Thrombotic Thrombocytopenic Purpura

TTP is a devastating disease if not diagnosed and treated promptly. In patients presenting with new thrombocytopenia, with or without evidence of renal insufficiency and other elements of classic TTP, laboratory data (PT, aPTT, CBC with platelet count and peripheral smear, ADAMTS13 activity, LDH, bilirubin, haptoglobin, direct antiglobulin assay) should be obtained to rule out DIC and to evaluate for evidence of MAHA.

Therapeutic plasma exchange (TPE) remains the mainstay of treatment of TTP. TPE is continued until the platelet count is normal and signs of hemolysis are resolved for at least 2 days. Although never evaluated in clinical trials, the use of glucocorticoids seems a reasonable approach but should only be used as an adjunct to plasma exchange. The addition of rituximab to initial therapy decreases the risk of relapse. Caplacizumab, an anti-VWF nanobody, decreases mortality and burden of care when used in patients with ADAMTS13 <10% or with high clinical probability of disease. Guidelines from the International Society of Thrombosis and Hemostasis recommend starting caplacizumab and rituximab only in individuals with diagnostic ADAMTS13 levels (usually <10%) and, additionally for rituximab, in patients with evidence of an inhibitor, given potential side effects and costs.

Patients with persistently low ADAMTS13 have a greater risk of ongoing sequelae including stroke. There is a significant relapse rate; in patients treated with TPE, 25–45% of patients relapse within 30 days of initial "remission," and 12–40% of patients have late relapses. Relapses are more frequent in patients with severe ADAMTS13 deficiency at presentation. Treatment of patients with TTP relapses should be initiated before confirmatory laboratory assays are available.

Hemolytic-Uremic Syndrome HUS is a syndrome characterized by acute renal failure, microangiopathic hemolytic anemia, and thrombocytopenia. It is seen preceded by an episode of diarrhea, often hemorrhagic in nature, predominantly in children. *Escherichia coli* O157:H7 is the most frequent, although not only, etiologic serotype. HUS not associated with diarrhea is more heterogeneous in presentation and course. Atypical HUS (aHUS) is usually due to genetic defects in complement genes or antibodies directed against complementary regulatory proteins that result in chronic complement activation. Laboratory testing for DNA variants in complement regulatory genes is available, although assigning pathogenicity to variants remains challenging. Currently, a commercially available functional assay is not available that is diagnostic of the disease.

TREATMENT

Hemolytic-Uremic Syndrome

Treatment of HUS is primarily supportive. In HUS associated with diarrhea, many (~40%) children require at least some period of support with dialysis; however, the overall mortality is <5%. In HUS not associated with diarrhea, the mortality is higher, approximately 26%. Plasma infusion or plasma exchange has not been shown to alter the overall course in HUS or aHUS, except in patients with antibodies to factor H. ADAMTS13 levels are generally reported to be normal in HUS, although occasionally they have been reported to be decreased. In patients with aHUS, anticomplement therapy has efficacy in resolution of aHUS and improving or preserving renal function. Patients with aHUS may initially be treated with plasma exchange, until the ADAMTS13 level is returned and the diagnosis is more clear, since aHUS remains a diagnosis of exclusion. However, plasma exchange has not been shown to affect clinical outcomes in aHUS.

THROMBOCYTOSIS

Thrombocytosis is almost always due to (1) iron deficiency; (2) inflammation, cancer, or infection (reactive thrombocytosis); or (3) an underlying myeloproliferative process (essential thrombocythemia or polycythemia vera) **(Chap. 108)** or, rarely, the 5q– myelodysplastic process **(Chap. 107)**. Patients presenting with an elevated platelet count should be evaluated for underlying inflammation and malignancy, and iron deficiency should be ruled out. Thrombocytosis in response to acute or chronic inflammation has not been clearly associated with an increased thrombotic risk. In fact, patients with markedly elevated platelet counts (>1.5 million), usually seen in the setting of a myeloproliferative disorder, have an increased risk of bleeding. This appears to be due, at least in part, to acquired von Willebrand disease (VWD) due to platelet-VWF binding and removal from the circulation.

■ QUALITATIVE DISORDERS OF PLATELET FUNCTION

Inherited Disorders of Platelet Function Inherited platelet function disorders are thought to be relatively rare, although the prevalence of mild disorders of platelet function is unclear, in part because our testing for such disorders is suboptimal. Rare qualitative disorders include the autosomal recessive disorders Glanzmann's thrombasthenia (absence of the platelet Gp IIb/IIIa receptor) and Bernard-Soulier syndrome (absence of the platelet Gp Ib-IX-V receptor). Both are inherited in an autosomal recessive fashion and present with bleeding symptoms in childhood.

Platelet storage pool disorder (SPD) is the classic autosomal dominant qualitative platelet disorder. This results from abnormalities of platelet granule formation. It is also seen as a part of inherited disorders of granule formation, such as Hermansky-Pudlak syndrome. Bleeding symptoms in SPD are variable but often are mild. The most common inherited disorders of platelet function prevent normal secretion of granule content and are termed *secretion defects*. An increasing number of genetic variants are being found in patients with these disorders, although assigning pathogenicity remains challenging.

TREATMENT

Inherited Disorders of Platelet Dysfunction

Bleeding symptoms or prevention of bleeding in patients with severe platelet dysfunction frequently requires platelet transfusion. Care must be taken to limit the risk of alloimmunization by limiting exposure and using human leukocyte antigen–matched single donor platelets for transfusion when needed. Recombinant factor VIIa (rFVIIa) is FDA approved in Glanzmann's thrombasthenia and Bernard Soulier syndrome where use can avoid platelet alloimmunization and antireceptor antibody formation. Platelet disorders associated with milder bleeding symptoms frequently respond to desmopressin (1-deamino-8-D-arginine vasopressin [DDAVP]). DDAVP increases plasma VWF and factor VIII levels; it may also have a direct effect on platelet function. Particularly for mucosal bleeding symptoms, antifibrinolytic therapy (tranexamic acid or ε-aminocaproic acid) is used alone or in conjunction with DDAVP or platelet therapy.

Acquired Disorders of Platelet Function Acquired platelet dysfunction is common, usually due to medications, either intentionally as with antiplatelet therapy or unintentionally as with high-dose penicillins. Acquired platelet dysfunction occurs in uremia. This is likely multifactorial, but the resultant effect is defective adhesion and activation. The platelet defect is improved most by dialysis but may also be improved by increasing the hematocrit to 27–32%, giving DDAVP (0.3 μg/kg), or use of conjugated estrogens. Platelet dysfunction also occurs with cardiopulmonary bypass due to the effect of the artificial circuit on platelets, and bleeding symptoms respond to platelet transfusion. Platelet dysfunction seen with underlying hematologic disorders can result from nonspecific interference by circulating paraproteins or intrinsic platelet defects in myeloproliferative and myelodysplastic syndromes.

■ VON WILLEBRAND DISEASE

VWD is the most common inherited bleeding disorder, with prevalence of symptomatic disease of 1 in 1000 to 1 in 10,000 individuals. VWF serves two roles: (1) as the major adhesion molecule that tethers the platelet to the exposed subendothelium; and (2) as the binding protein for factor VIII (FVIII), resulting in significant prolongation of the FVIII half-life in circulation. The platelet-adhesive function of VWF is critically dependent on the presence of large VWF multimers, whereas FVIII binding is not. Most of the symptoms of VWD are "platelet-like" except in more severe VWD when the FVIII is low enough to produce symptoms similar to those found in FVIII deficiency (hemophilia A).

TABLE 120-2 Laboratory Diagnosis of von Willebrand Disease (VWD)

TYPE	aPTT	VWF ANTIGEN	VWF ACTIVITY	FVIII ACTIVITY	MULTIMER
1	Nl or ↑	↓	↓	↓	Normal distribution, decreased in quantity
2A	Nl or ↑	↓	↓↓	↓	Loss of high- and intermediate-MW multimers
2B[a]	Nl or ↑	↓	↓↓	↓	Loss of high-MW multimers
2M	Nl or ↑	↓	↓↓	↓	Normal distribution, decreased in quantity
2N	↑↑	Nl or ↓[b]	Nl or ↓[b]	↓↓	Normal distribution
3	↑↑	↓↓	↓↓	↓↓	Absent

[a]Usually also decreased platelet count. [b]For type 2N, in the homozygous state, factor VIII is very low; in the heterozygous state, it is only seen in conjunction with type 1 VWD.

Abbreviations: aPTT, activated partial thromboplastin time; F, factor; MW, molecular weight; Nl, normal; VWF, von Willebrand factor.

VWD has been classified into three major types, with four subtypes of type 2 (**Table 120-2**). By far, the most common type of VWD is type 1 disease, with a parallel decrease in VWF protein, VWF function, and FVIII levels, accounting for at least 80% of cases. Type 1C is a subtype associated with increased VWF clearance. In type 1 VWD, patients have predominantly mucosal bleeding symptoms, although procedure-related and other bleeding is also seen. Bleeding symptoms are uncommon in infancy and usually manifest later in childhood with excessive bruising and epistaxis. Because these symptoms occur commonly in childhood, the clinician should particularly note bruising at sites unlikely to be traumatized and/or prolonged epistaxis requiring medical attention. Heavy menstrual bleeding is a common manifestation of VWD. Menstrual bleeding resulting in anemia should warrant an evaluation for VWD and, if negative, functional platelet disorders and other bleeding disorders. Type 1 VWD may first manifest with dental extractions, particularly wisdom tooth extraction, or tonsillectomy.

Not all patients with low VWF levels have bleeding symptoms. Whether patients bleed or not will depend on the overall hemostatic balance they have inherited, along with environmental influences and the type of hemostatic challenges they experience. Although the inheritance of VWD is autosomal, many factors modulate both VWF levels and bleeding symptoms. These have not all been defined, but include blood type, thyroid hormone status, race, stress, exercise, hormonal (both endogenous and exogenous) influences, and modulators of VWF clearance. Patients with type O blood have VWF protein levels of approximately one-half those of patients with AB blood type, and in fact, the normal range for patients with type O blood overlaps that which has been considered diagnostic for VWD. Patients with mildly decreased VWF levels should be diagnosed with VWD only in the setting of bleeding symptoms and/or a family history of VWD.

Patients with type 2 VWD have functional defects; thus, the VWF antigen measurement is significantly higher than the test of function. For types 2A, 2B, and 2M VWD, platelet-binding and/or collagen-binding VWF activity is decreased. In type 2A VWD, the impaired function is due either to increased susceptibility to cleavage by ADAMTS13, resulting in loss of intermediate- and high-molecular-weight multimers, or to decreased production of these multimers by the cell. Type 2B VWD results from gain-of-function DNA variants that result in increased binding of VWF to platelets in circulation with subsequent increased ADAMTS13 cleavage and clearance. The resulting VWF in the patients' plasma lacks the highest molecular-weight multimers, and the platelet count is usually modestly reduced, but not uniformly. Type 2M occurs as a consequence of DNA variants that result in a dysfunctional protein not affecting multimer structure.

Type 2N VWD is due to variants in the *VWF* gene that affect binding of FVIII. As FVIII is stabilized by binding to VWF, the FVIII in patients with type 2N VWD has a very short half-life, and the FVIII level is markedly decreased. This is sometimes termed *autosomal hemophilia*. Type 3 VWD, or severe VWD, describes patients with virtually no VWF protein and usually FVIII levels <10%. Patients experience mucosal and joint bleeding, surgery-related bleeding, and other bleeding symptoms. Some patients with type 3 VWD, particularly those with large *VWF* gene deletions, are at risk of developing antibodies to infused VWF.

Acquired VWD or von Willebrand syndrome can be seen in patients with underlying lymphoproliferative disorders, including monoclonal gammopathies of underdetermined significance (MGUS), multiple myeloma, and Waldenström's macroglobulinemia. It is seen most commonly in the setting of MGUS and should be suspected in patients, particularly elderly patients, with a new onset of severe mucosal bleeding symptoms. Laboratory evidence of acquired VWD is found in patients with cardiac valvular disease. Heyde's syndrome (aortic stenosis with gastrointestinal bleeding) is attributed to the presence of angiodysplasia of the gastrointestinal tract in patients with aortic stenosis. The shear stress on blood passing through the stenotic aortic valve appears to unfold VWF, making it susceptible to proteolysis. Consequently, large multimer forms are lost, leading to an acquired type 2 VWD, but return when the stenotic valve is replaced.

TREATMENT

Von Willebrand Disease

The mainstay of treatment for type 1 VWD is DDAVP (desmopressin), which results in release of VWF and FVIII from endothelial stores. DDAVP can be given intravenously, by high-concentration intranasal spray (1.5 mg/mL), or when a concentrated form is available, by subcutaneous injection. The peak activity when given intravenously is approximately 30 min, whereas it is 2 h when given intranasally. The usual dose is 0.3 µg/kg intravenously or two squirts (one in each nostril) for patients >50 kg (one squirt for those <50 kg). It is recommended that patients with VWD be tested with DDAVP to assess their response before using it. In patients who respond well (increase in laboratory values greater than twofold with levels >50% for at least 4 h), it can be used for procedures with minor to moderate risk of bleeding. Depending on the procedure, additional doses may be needed; it is usually given every 12–24 h. Less frequent dosing may result in less tachyphylaxis, which occurs when synthesis cannot compensate for the released stores. The major side effect of DDAVP is hyponatremia due to decreased free water clearance. This occurs most commonly in the very young and the very old, but fluid restriction should be advised for all patients for the 24 h following each dose.

Some patients with type 2A VWD respond to DDAVP such that it can be used for minor procedures. For the other subtypes, for type 3 disease, and for major procedures requiring longer periods of normal hemostasis, VWF replacement can be given. Virally inactivated VWF-plasma-derived and recombinant factor concentrates are safer than cryoprecipitate as the replacement product.

Antifibrinolytic therapy using either tranexamic acid (TXA) or ε-aminocaproic acid is an important therapy, either alone or in an adjunctive capacity, particularly for the prevention or treatment of mucosal bleeding. These agents are particularly useful in treatment of heavy menstrual bleeding (TXA 1300 mg every 8 h) and postpartum hemorrhage, as prophylaxis for dental procedures, and with DDAVP or factor concentrate for dental extractions, tonsillectomies, and prostate procedures. Antifibrinolytic agents are contraindicated in the setting of upper urinary tract bleeding due to the risk of ureteral obstruction.

DISORDERS OF THE VESSEL WALL

The vessel wall is an integral part of hemostasis, and separation of a fluid phase is artificial, particularly in disorders such as TTP or HIT that clearly involve the endothelium as well. Inflammation localized to the vessel wall, such as vasculitis, and inherited connective tissue disorders are abnormalities inherent to the vessel wall.

Metabolic and Inflammatory Disorders Acute febrile illnesses may result in vascular damage. This can result from immune complexes containing viral antigens or the viruses themselves. Certain pathogens, such as the rickettsiae causing Rocky Mountain spotted fever, replicate in endothelial cells and damage them. SARS-CoV-2 also infects endothelial cells, resulting in activation and damage contributing to COVID-19 pathogenicity. Vascular purpura may occur in patients with polyclonal gammopathies but more commonly occurs in those with monoclonal gammopathies, including Waldenström's macroglobulinemia, multiple myeloma, and cryoglobulinemia. Patients with mixed cryoglobulinemia develop a more extensive maculopapular rash due to immune complex–mediated damage to the vessel wall.

Patients with scurvy (vitamin C deficiency) develop painful episodes of perifollicular skin bleeding as well as more systemic bleeding symptoms. Vitamin C is needed to synthesize hydroxyproline, an essential constituent of collagen. Patients with Cushing's syndrome or on chronic glucocorticoid therapy develop skin bleeding and easy bruising due to atrophy of supporting connective tissue. A similar phenomenon is seen with aging, where following minor trauma, blood spreads superficially under the epidermis. This has been termed *senile purpura*. It is most common on skin that has been previously damaged by sun exposure.

Immunoglobulin A vasculitis, formerly called Henoch-Schönlein purpura, is a distinct, self-limited type of vasculitis that occurs in children and young adults. Patients have an acute inflammatory reaction with IgA and complement components in capillaries, mesangial tissues, and small arterioles leading to increased vascular permeability and localized hemorrhage. The syndrome is often preceded by an upper respiratory infection, commonly with streptococcal pharyngitis, or is triggered by drug or food allergies. Patients develop a purpuric rash on the extensor surfaces of the arms and legs, usually accompanied by polyarthralgias or arthritis, abdominal pain, and hematuria from focal glomerulonephritis. All coagulation tests are normal, but renal impairment may occur. Glucocorticoids can provide symptomatic relief but do not alter the course of the illness.

Inherited Disorders of the Vessel Wall Patients with inherited disorders of the connective tissue matrix, such as Marfan's syndrome, Ehlers-Danlos syndrome, and pseudoxanthoma elasticum, frequently report easy bruising. Inherited vascular abnormalities can result in increased bleeding. This is notably seen in hereditary hemorrhagic telangiectasia (HHT, or Osler-Weber-Rendu disease), a disorder where abnormal telangiectatic capillaries result in frequent bleeding episodes, primarily from the nose and gastrointestinal tract. Arteriovenous malformation (AVM) in the lung, brain, and liver may also occur in HHT. The telangiectasia can often be visualized on the oral and nasal mucosa. Signs and symptoms develop over time. Epistaxis begins, on average, at the age of 12 and occurs in >95% of affected individuals by middle age. Approximately 25% have gastrointestinal bleeding usually beginning after the age of 50. HHT is caused by pathogenic DNA variants in a number of genes involved in the TGFβ/BMP signaling cascade.

FURTHER READING

Boender J et al: A diagnostic approach to mild bleeding disorders. J Thromb Haemost 14:1507, 2016.

Cines DB, Greinacher A: Vaccine-induced immune thrombotic thrombocytopenia. Blood 141:1659, 2023.

Connell NT: ASH ISTH NHF WFH 2021 guidelines on the management of von Willebrand disease. Blood Adv 5:301, 2021.

Cuker A et al: American Society of Hematology 2018 guidelines for management of venous thromboembolism: Heparin-induced thrombocytopenia. Blood Adv 2:23360, 2018.

Gomez K: Advances in the diagnosis of heritable platelet disorders. Blood Rev 56:100972, 2022.

James PD et al: ASH ISTH NHF WFH 2021 guidelines on the diagnosis of von Willebrand disease. Blood Adv 5:280, 2021.

Jiang D et al: Changing paradigms in ITP management: Newer tools for an old disease. Transfus Med Rev 36:188, 2022.

Jokiranta TS: HUS and atypical HUS. Blood 129:2847, 2017.

May J et al: Heparin-induced thrombocytopenia: An illustrated review. Res Pract Thromb Haemost 7:100283, 2023.

Skeith L et al: A practical approach to evaluating postoperative thrombocytopenia. Blood Adv 4:776, 2020.

Vayne C et al: Pathophysiology and diagnosis of drug-induced immune thrombocytopenia. J Clin Med 9:2212, 2020.

Zheng XL et al: ISTH guidelines for the diagnosis of thrombotic thrombocytopenic purpura. J Thromb Hemost 18:2486, 2020.

121 Coagulation Disorders

Jean M. Connors

Deficiencies of coagulation factors have been recognized for centuries. Patients with genetic deficiencies of plasma coagulation factors exhibit lifelong recurrent bleeding episodes into joints, muscles, and closed spaces, either spontaneously or following an injury. The most common inherited factor deficiencies are the hemophilias, X-linked diseases caused by deficiency of factor (F) VIII (hemophilia A) or FIX (hemophilia B). Rare congenital bleeding disorders due to deficiencies of other factors, including FII (prothrombin), FV, FVII, FX, FXI, FXIII, and fibrinogen, are commonly inherited in an autosomal recessive manner **(Table 121-1)**. Disease phenotype often correlates with the level of factor activity. While patients can have a congenital deficiency of FXII accompanied by a significant prolongation in the activated partial thromboplastin time (aPTT), FXII deficiency is not accompanied by a bleeding phenotype, likely due to redundant paths to activation of the intrinsic pathway of the coagulation cascade, including direct activation of FXI by thrombin generated through the extrinsic pathway **(Fig. 121-1)**. Advances in characterization of the molecular basis of clotting factor deficiencies have contributed to better understanding of the disease phenotypes allowing the development of more targeted therapeutic approaches, including the use of small molecules, recombinant proteins, or cell- and gene-based therapies.

The two most commonly used tests of hemostasis, the prothrombin time (PT) and the aPTT, were designed to perform the first screen for clotting factor deficiency (Fig. 121-1). An isolated prolonged PT suggests FVII deficiency, whereas a prolonged aPTT indicates an intrinsic pathway factor deficiency, most commonly hemophilia A or B (FVIII or FIX, respectively) or FXI deficiency (Fig. 121-1). The prolongation of both PT and aPTT suggests a deficiency of FV, FX, FII, or fibrinogen abnormalities. A mixing study, in which the addition of normal pooled plasma to the patient's plasma, will correct a prolonged aPTT or PT due to a factor deficiency, and is the next step in determining if there is a coagulation factor deficiency. If the clotting time does not correct, it suggests the presence of an inhibitor, an antibody to a specific factor; however, a mixing study will also detect the presence of anticoagulants. Many labs have testing methods for detecting inhibitors that neutralize anticoagulants. If the mixing study corrects with normal plasma, individual factor activity assays are performed to determine which factor is deficient.

Acquired deficiencies of plasma coagulation factors are more frequent than congenital disorders; the most common disorders include hemorrhagic diathesis of liver disease, disseminated intravascular coagulation (DIC), and vitamin K deficiency. In these disorders, blood coagulation is hampered by the deficiency of more than one clotting factor, and the bleeding episodes are the result of perturbation of both primary (e.g., platelet and vessel wall interactions) and secondary (coagulation) hemostasis.

The development of alloantibodies to coagulation plasma proteins, clinically termed *inhibitors*, is a relatively rare disease that often affects hemophilia A or B and FXI-deficient patients on repetitive exposure to the missing protein to control bleeding episodes. Inhibitory autoantibodies also occur among subjects without genetic deficiency of clotting factors and, although rare, can be seen in the postpartum setting, as a manifestation of underlying autoimmune or neoplastic disease, or idiopathically. Rare cases of acquired inhibitors to thrombin or FV have been reported in patients receiving topical bovine thrombin preparation as a local hemostatic agent in complex surgeries. A mixing study that does not correct with the addition of normal plasma indicates the presence of an inhibitor, requiring additional tests to identify the specificity of the inhibitor and measure its titer. Inhibitor detection in patients with hemophilia is of particular importance, with yearly screening performed at most hemophilia treatment centers.

The treatment of coagulation factor deficiencies in the setting of bleeding requires replacement of the deficient protein(s) using recombinant or purified plasma-derived products or fresh-frozen plasma

TABLE 121-1 Genetic and Laboratory Characteristics of Inherited Coagulation Disorders

CLOTTING FACTOR DEFICIENCY	INHERITANCE	PREVALENCE IN GENERAL POPULATION	LABORATORY ABNORMALITY[a]			MINIMUM HEMOSTATIC LEVELS	TREATMENT	PLASMA HALF-LIFE
			aPTT	PT	TT			
Fibrinogen	AR	1 in 1,000,000	+	+	+	100 mg/dL	Cryoprecipitate	2–4 d
Prothrombin	AR	1 in 2,000,000	+	+	–	20–30%	FFP/PCC	3–4 d
Factor V	AR	1 in 1,000,000	+/–	+/–	–	15–20%	FFP[c]	36 h
Factor VII	AR	1 in 500,000	–	+	–	15–20%	FFP/PCC	4–6 h
Factor VIII	X-linked	1 in 5000	+	–	–	30%	FVIII concentrates	8–12 h
Factor IX	X-linked	1 in 30,000	+	–	–	30%	FIX concentrates	18–24 h
Factor X	AR	1 in 1,000,000	+/–	+/–	–	15–20%	FFP/PCC	40–60 h
Factor XI	AR	1 in 1,000,000	+	–	–	15–20%	FFP	40–70 h
Factor XII	AR	ND	+	–	–	[b]	[b]	60 h
HK	AR	ND	+	–	–	[b]	[b]	150 h
Prekallikrein	AR	ND	+	–	–	[b]	[b]	35 h
Factor XIII	AR	1 in 2,000,000	–	–	+/–	2%–5%	Cryoprecipitate/FXIII concentrates	11–14 d

[a] Values within normal range (–) or prolonged (+). [b] No risk for bleeding; treatment is not indicated. [c] Since platelets contain FV, platelet transfusion can be used as therapy.

Abbreviations: aPTT, activated partial thromboplastin time; AR, autosomal recessive; FFP, fresh-frozen plasma; HK, high-molecular-weight kininogen; ND, not determined; PCC, prothrombin complex concentrates; PT, prothrombin time; TT, thrombin time.

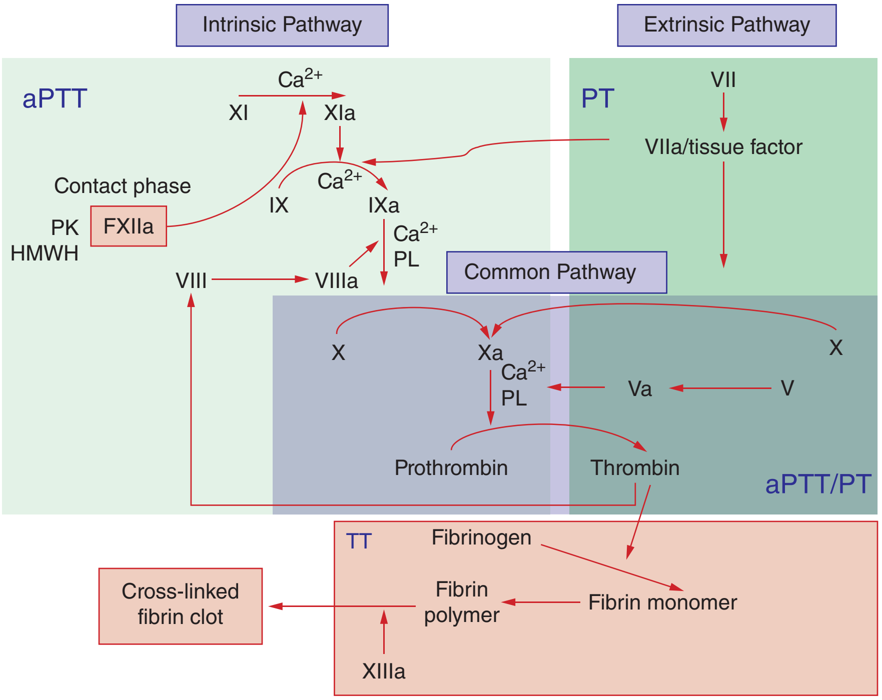

FIGURE 121-1 Coagulation cascade and laboratory assessment of clotting factor deficiency by activated partial thromboplastin time (aPTT), prothrombin time (PT), thrombin time (TT), and phospholipid (PL).

(FFP). Prothrombin complex concentrates (PCCs) are intermediate-purity plasma-derived factor concentrates initially used as sources of FVIII or FIX for hemophilia patients, but because they contain the vitamin K–dependent factors, they are also used for warfarin reversal. Three-factor PCC (3F-PCC) is less frequently used now for warfarin reversal because these preparations contain low levels of FVII, requiring FFP as a source of FVII. Four-factor PCC (4F-PCC), especially the one used in the United States, contains FII, FIX, FX, higher levels of FVII than 3F-PCC, and protein S and protein C.

HEMOPHILIA A AND B

PATHOGENESIS AND CLINICAL MANIFESTATIONS

Hemophilia is an X-linked recessive hemorrhagic disease due to mutations in the *F8* gene (hemophilia A or classic hemophilia) or *F9* gene (hemophilia B). The disease affects 1 in 10,000 males worldwide, in all ethnic groups; hemophilia A represents 80% of all cases. The large size of the *F8* gene makes it more susceptible to mutation events than the smaller *F9* gene. Male subjects are clinically affected; women, who carry a single mutated gene, are generally asymptomatic. However, increased bleeding tendencies with procedures are now more commonly appreciated based on F8 or F9 level. Family history of the disease is absent in ~30% of cases, and in these cases, 80% of the mothers are carriers of the de novo mutated allele. More than 500 different mutations have been identified in the *F8* or *F9* gene. One of the most common hemophilia A mutations results from an inversion of the intron 22 sequence, and it is present in 40% of cases of severe hemophilia A. Advances in molecular diagnosis now permit precise identification of mutations, allowing accurate diagnosis of women carriers of the hemophilia gene in affected families.

Clinically, hemophilia A and hemophilia B are indistinguishable. The disease phenotype correlates with the activity of FVIII or FIX and can be classified as severe (<1%), moderate (1–5%), or mild (6–30%). In the severe and moderate forms, the disease is characterized by bleeding into the joints (hemarthrosis), soft tissues, and muscles after minor trauma or even spontaneously. Patients with mild disease experience infrequent bleeding, usually secondary to trauma. Among those with baseline FVIII or FIX activity >25%, the disease is discovered only with bleeding after major trauma or during routine laboratory tests, usually an isolated prolongation of the aPTT that requires mixing study evaluation. FVIII has a short circulating half-life of 25–30 min that is extended to roughly 12 h when complexed with its carrier protein von Willebrand factor (VWF). In patients without a known history of hemophilia, a diagnosis of von Willebrand disease (VWD) needs to be excluded in patients with a prolonged aPTT and low FVIII activity. Early in life, bleeding may present after circumcision or rarely as intracranial hemorrhages. The disease is more evident when children begin to walk or crawl. In the severe form, the most common bleeding manifestations are recurrent hemarthroses, affecting primarily the knees, elbows, ankles, shoulders, and hips. Acute hemarthroses are painful, and clinical signs are local swelling and erythema. To avoid pain, the patient may adopt a fixed position, which leads eventually to muscle contractures. Very young children unable to communicate verbally show irritability and a lack of movement of the affected joint. Chronic hemarthroses are debilitating with synovial thickening and synovitis in response to the intraarticular blood. After a joint has been damaged, recurrent bleeding episodes result in the clinically recognized “target joint,” which then establishes a vicious cycle of bleeding, resulting in progressive joint deformity that in critical cases requires surgery as the only therapeutic option. Hematomas into the muscle of distal parts of the limbs may lead to external compression of arteries, veins, or nerves that can result in compartment syndrome.

Bleeding into the oropharyngeal spaces, central nervous system (CNS), or retroperitoneum is life-threatening and requires immediate therapy. Retroperitoneal hemorrhages can accumulate large quantities of blood with formation of masses with calcification and inflammatory tissue reaction (pseudotumor syndrome) and also result in damage to the femoral nerve. Pseudotumors can also form in bones, especially long bones of the lower limbs. Hematuria is frequent among hemophilia patients, even in the absence of genitourinary pathology. It is often self-limited and may not require specific therapy.

TREATMENT

Hemophilia

Without treatment, severe hemophilia may limit life expectancy. Advances in the blood fractionation industry during World War II resulted in the realization that plasma could be used to treat hemophilia, but the volumes required to achieve even modest elevation of circulating factor levels limit the utility of plasma infusion as an approach to disease management. The discovery in the 1960s that the cryoprecipitate fraction of plasma was enriched for FVIII, and the eventual purification of FVIII and FIX from plasma, led to the introduction of home infusion therapy with factor concentrates in the 1970s. The availability of factor concentrates resulted in a dramatic improvement in life expectancy and quality of life for people with severe hemophilia. However, the contamination of the blood supply with hepatitis viruses and HIV resulted in transmission of these bloodborne infections within the hemophilia population. The introduction of viral inactivation steps in the preparation of plasma-derived products in the mid-1980s greatly reduced the risk of HIV and hepatitis; the risks were further reduced by the production of recombinant FVIII and FIX proteins in the 1990s. It is uncommon for hemophilic patients born after 1985 to have contracted either hepatitis or HIV, and for these individuals, life expectancy is now ~65 years. In fact, since 1998, new infections with viral hepatitis or HIV have not been reported in hemophilia patients.

Factor replacement for hemophilia has been the mainstay of therapy for half a century; however, advances including uniquely functioning molecules and gene therapy have expanded treatment approaches. Factor replacement has been provided either in response to a bleeding episode or as prophylactic treatment. Primary prophylaxis is defined as maintaining the missing clotting factor at levels ~1% or higher on a regular basis to prevent bleeds, especially the onset of hemarthroses. Hemophilic boys receiving regular infusions of FVIII (3 days/week) or FIX (2 days/week) can reach puberty without detectable joint abnormalities. Therefore, prophylactic treatment has become more common. The Centers for Disease Control and Prevention reported that >51% of children with severe hemophilia who are aged <6 years receive prophylaxis, increasing considerably from 33% in 1995. Although prophylaxis with factor concentrates is the standard care for children and adults with severe hemophilia, teenagers and young adults do not always maintain treatment due to high cost and lifestyle factors including difficulties accessing peripheral veins for infusions that occur two to three times a week or potential infectious and thrombotic risks of long-term central vein catheters.

Treatment of hemophilia bleeds requires the following: (1) prompt initiation of factor replacement as symptoms often precede objective evidence of bleeding, especially for classic symptoms of bleeding into the joint in a reliable patient, headaches, or major trauma; and (2) avoidance of antiplatelet drugs.

FVIII and FIX are dosed in units. One unit is defined as the amount of FVIII (100 ng/mL) or FIX (5 μg/mL) in 1 mL of normal plasma. One unit of FVIII per kilogram of body weight increases the plasma FVIII level by 2%. One can calculate the dose needed to increase FVIII levels to 100% in a 70-kg severe hemophilia patient (<1%) using the simple formula below. Thus, 3500 units of FVIII will raise the circulating level to 100%.

$$\text{FVIII dose (IU)} = \text{Target FVIII levels} - \text{FVIII baseline levels} \\ \times \text{body weight (kg)} \times 0.5\ \text{unit/kg}$$

The doses for FIX replacement are different from those for FVIII, because FIX recovery after infusion is usually only 50% of the predicted value. Therefore, the formula for FIX replacement is as follows:

$$\text{FIX dose (IU)} = \text{Target FIX levels} - \text{FIX baseline levels} \\ \times \text{body weight (kg)} \times 1\ \text{unit/kg}$$

The FVIII half-life of 8–12 h requires injections twice a day to maintain therapeutic levels, whereas the FIX half-life is longer, ~24 h, so that once-a-day injection is sufficient. In specific situations such as after surgery, continuous infusion of factor may be desirable because of its safety in achieving sustained factor levels at a lower total cost.

Cryoprecipitate is enriched with FVIII protein bound to VWF (each bag contains ~80 IU of FVIII). This product should be used only in emergencies when factor concentrates are not available, although cryoprecipitate may be the only source of FVIII in developing countries.

Mild bleeds such as uncomplicated hemarthroses or superficial hematomas require achieving an initial factor level of 30–50%. Additional doses to maintain levels of 15–25% for 2 or 3 days are indicated for severe hemarthroses, especially when these episodes affect the "target joint." Large hematomas, or bleeds into deep muscles, require factor levels of 50% or even higher if the clinical symptoms do not improve, and factor replacement may be required for a period of 1 week or longer. The control of serious bleeds, including those that affect the oropharyngeal spaces, CNS, and the retroperitoneum, requires sustained protein levels of 50–100% for 7–10 days. Prophylactic replacement for surgery is aimed at achieving normal factor levels (100%) for a period of 7–10 days; replacement can then be tapered depending on the extent of the surgical wounds. Oral surgery is associated with extensive tissue damage that usually requires factor replacement for 1–3 days coupled with oral antifibrinolytic drugs.

NONTRANSFUSION THERAPY IN HEMOPHILIA

DDAVP (1-Amino-8-D-Arginine Vasopressin) DDAVP is a synthetic vasopressin analogue that causes a transient rise in FVIII and VWF, but not FIX, by release from stores in vascular endothelial cells. Patients with moderate or mild hemophilia A should be tested to determine if they respond to DDAVP before use. DDAVP at doses of 0.3 μg/kg body weight, over a 20-min period, is expected to raise FVIII levels by two- to threefold over baseline, peaking between 30 and 60 min after infusion. DDAVP does not improve FVIII levels in severe hemophilia A patients because no stores are available to release. Repeated dosing of DDAVP results in tachyphylaxis as storage pools are depleted. After three consecutive doses, if further therapy is indicated, exogenous FVIII is required.

Antifibrinolytic Drugs Bleeding in the gums, the gastrointestinal tract, and during oral surgery can be treated with oral antifibrinolytic drugs such as ε-aminocaproic acid (EACA) or tranexamic acid to prevent fibrin degradation by plasmin. The duration of the treatment depending on the clinical indication is 1 week or longer. Tranexamic acid is given at doses of 25 mg/kg three to four times a day. EACA treatment requires a loading dose of 200 mg/kg (maximum of 10 g) followed by 100 mg/kg per dose (maximum 30 g/d) every 6 h. These drugs are not indicated to control hematuria because of concern for forming an occlusive clot in the lumen of genitourinary tract structures.

COMPLICATIONS

Inhibitor Formation The formation of alloantibodies to FVIII or FIX is the major complication of hemophilia treatment. The prevalence of inhibitors to FVIII is estimated to be ~30% in severe hemophilia A patients and 10% among patients with nonsevere hemophilia A. Inhibitors to FIX are detected in only 3–5% of all hemophilia B patients. The high-risk group for inhibitor formation includes severe deficiency (>80% of all cases of inhibitors), familial history of inhibitor, African descent, mutations in the FVIII or FIX gene resulting in deletion of large coding regions, or gross gene rearrangements. Inhibitors usually appear early in life, at a median of 2 years of age, and after 10 cumulative days of exposure. However, intensive replacement therapy such as for major surgery, intracranial bleeding, or trauma increases the risk of inhibitor formation for patients of all ages and severity; all patients require close laboratory monitoring following these events.

The clinical diagnosis of an inhibitor is suspected when patients do not respond to factor replacement at therapeutic doses. Inhibitors increase both morbidity and mortality in hemophilia. Because early detection of an inhibitor is critical to a successful correction of the bleeding or to eradication of the antibody, most hemophilia centers perform annual screening with aPTT and mixing studies. The Bethesda assay uses a similar principle as a mixing study and defines the specificity of the inhibitor and its titer. The results are expressed in Bethesda units (BU), in which 1 BU is the amount of antibody that neutralizes 50% of the FVIII or FIX present in normal plasma after 2 h of incubation at 37°C. Clinically, inhibitor patients are classified as low responders or high responders, with response defined as increase in antibody titer; knowledge of responder type guides therapy. Therapy for inhibitor patients has two goals: the control of acute bleeding episodes and the eradication of the inhibitor. For the control of bleeding episodes, low responders, those with titer <5 BU, respond well to high doses of human FVIII (50–100 U/kg), with minimal or no increase in the inhibitor titers. However, high-responder patients, those with initial inhibitor titer >5 BU or an anamnestic response with increase in the antibody titer to >5 BU, even if low titer initially, do not respond to FVIII. The control of bleeding episodes in high-responder patients can be achieved by using concentrates enriched for prothrombin, FVII, FIX, FX (PCCs but usually activated PCCs [aPCCs]), and recombinant activated FVII (FVIIa), known as "bypass agents" because they activate coagulation downstream of the inhibited/absent factor or through a different pathway (Fig. 121-1). For FIX inhibitor patients, high doses of FIX can be used (<5 BU); however, allergic or anaphylactic reactions are common in FIX inhibitor patients; thus, bypass products should be used to treat or prevent bleeding as well as for those cases of high titer inhibitors. For eradication of the inhibitory antibody, immunosuppression alone is not effective. The most effective strategy is immune tolerance induction (ITI) based on daily infusion of the missing protein until the inhibitor disappears, typically requiring periods >1 year, with success rates of ~60%. The management of patients with severe hemophilia and inhibitors resistant to ITI is challenging. The use of anti-CD20 monoclonal antibody (rituximab) combined with ITI was thought to be effective, but although it reduces the inhibitor titers in some cases, sustained eradication is uncommon.

Other Therapeutic Approaches for Hemophilia A and B Engineered clotting factors, using fusion to polyethylene glycol (FVIII, FIX), IgG1-Fc (FVIII, FIX), or albumin (FIX) or other strategies, extend the plasma half-life of the coagulation factor. A number of products have been approved for use. These extended half-life products (for FVIII and FIX) facilitate prophylaxis with fewer weekly injections to maintain circulating levels >1%, decreasing injections from 3 to 2 days a week in hemophilia A and to once a week for hemophilia B. Novel approaches to manipulating the coagulation cascade components, including targeting the natural anticoagulants and inhibitors of activation of coagulation, have shown promising clinical trial results but do not yet have regulatory approval.

Emicizumab is an asymmetric bispecific antibody with one immunoglobulin variable chain region that binds FIXa and another that binds FX bringing them in close contact and resulting in activation of FX by FIXa. FXa subsequently cleaves prothrombin to thrombin—without the need for FVIII **(Fig. 121-2)**. It is effective in patients with severe hemophilia A with or without inhibitors. Similar molecules are in development. After initial once-a-week subcutaneous injections (an improvement over intravenous administration of factors) for 4 weeks, patients can usually be maintained with once-a-month dosing to prevent spontaneous bleeds, an overwhelmingly dramatic improvement in quality of life when compared to even the twice-weekly infusion schedule of "long-acting" FVIII compounds. Breakthrough bleeds can occur, however, and need to be carefully managed, as a small number of patients with inhibitors treated with aPCC or recombinant FVIIa developed thrombotic events or fatal thrombotic microangiopathy. Routine aPTT and FVIII activity measurements are inaccurate when emicizumab is present; to detect endogenous or infused FVIII activity, a chromogenic FVIII activity assay using bovine factor substrates is required.

These X-linked disorders are ideally suited for gene therapy as small increases in plasma factor level will result in significant clinical improvement. FIX has been the most studied as the gene is smaller and easier to package in the viral vectors used. In one approach, the sequence of a known spontaneous FIX gain-of-function mutation that has marked increase in specific activity, FIX Padua, is used so that small increments in plasma level of FIX are also accompanied by even greater increase in functional activity. The larger FVIII gene has also been successfully transferred through an adeno-associated viral vector to a few patients with hemophilia A. The early results appear promising. Complications include transaminitis and loss of gene expression for a variety of reasons. Gene therapy using adenoviral vector approaches have now been approved for patients with hemophilia A or B. Details on how to implement gene therapy, selection of appropriate patients, and follow-up monitoring are still being developed **(Chap. 483)**.

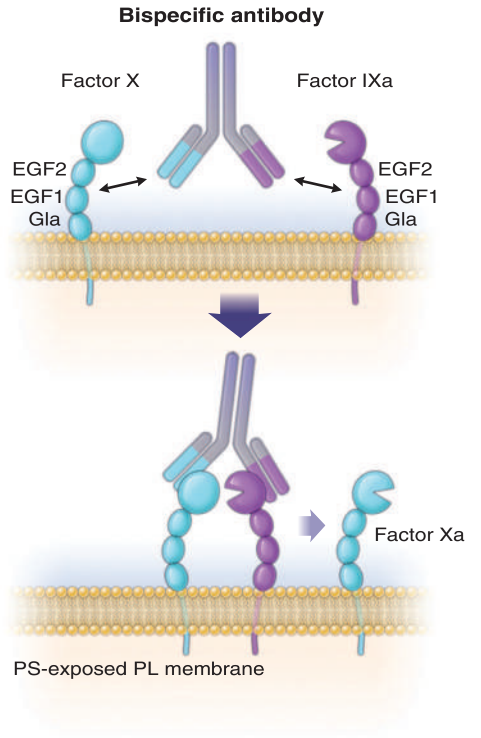

FIGURE 121-2 Mechanism of action of emicizumab. Emicizumab is a bifunctional antibody; the two binding sites recognize different protein sequences, unlike normal antibodies where both variable regions recognize the same antigen. One arm of emicizumab recognizes factor IXa and the other factor X. It functions to bring these two factors in proximity so that factor IXa can activate factor X to factor Xa, which then cleaves prothrombin to thrombin and activates the clotting cascade. *(From T Kitazawa, M Shima: Emicizumab, a humanized bispecific antibody to coagulation factors IXa and X with a factor VIIIa-cofactor activity. Int J Hematol 111:20, 2020.)*

INFECTIOUS DISEASES

Hemophilia patients treated with clotting factor concentrates before the development of recombinant factors in the 1990s were almost universally infected with hepatitis C virus (HCV) and HIV, which became the second leading cause of death. Co-infection of HCV and HIV, present in almost 50% of hemophilia patients, is an aggravating factor for the evolution of liver disease as correction of both genetic and acquired (secondary to liver disease) factor deficiencies may be needed. Effective treatments for both HIV and HCV have altered the devastating prognosis. In some select cases with cirrhosis, liver transplant has been performed, which also is curative for hemophilia.

EMERGING CLINICAL PROBLEMS IN AGING HEMOPHILIA PATIENTS

Patients with hemophilia now live well into adulthood, with life expectancy of patients with severe hemophilia now only ~10 years shorter than the general male population and near normal in patients with mild or moderate hemophilia. The older hemophilia population has distinct needs relating to more severe arthropathy, chronic pain, and high rates of HCV and/or HIV infections.

Although mortality from coronary artery disease is lower in hemophilia patients with hypocoagulability decreasing thrombus formation, atherogenesis is not prevented. Typical cardiovascular risk factors such as age, obesity, and smoking, along with physical inactivity, hypertension, and chronic renal disease, are seen in these patients as in the general population.

Management of an acute ischemic event and coronary revascularization should include collaboration among hematologists, cardiologists, and internists. Cancer due to HIV- and HCV-related malignancies is also a concern in this population, with hepatocellular carcinoma (HCC) the most common cause of death in HIV-negative patients. The recommendations for cancer screening for the general population should be the same for age-matched hemophilia patients, including routine screening for HCC. Hemophilia patients benefit from the same preventive and therapeutic approaches to minimize the risk of cardiovascular disease and malignancy as the general population.

MANAGEMENT OF CARRIERS OF HEMOPHILIA

Women carriers of hemophilia with factor levels ~50% of normal may not have an increased risk for bleeding. However, a wide range of factor activity (22–116%) due to random inactivation of the X chromosome (*lyonization*) can occur and lead to unexpected bleeding in women with low levels. The factor level of carriers should be measured to optimize perioperative management. During pregnancy, FVIII levels increase approximately two- to threefold in most carriers compared to nonpregnant women, whereas the FIX increase is less pronounced. After delivery, a rapid fall in the pregnancy-induced rise of maternal clotting factor levels occurs, resulting in increased risk for postpartum hemorrhage that can be prevented by infusion of factor concentrate to levels of 50–70% for 3 days for vaginal delivery and up to 5 days for cesarean delivery. In mild cases, the use of DDAVP and/or antifibrinolytic drugs is recommended.

FACTOR XI DEFICIENCY

FXI deficiency, also known as hemophilia C, is a rare autosomal bleeding disorder that occurs at a frequency of one in a million. However, it is highly prevalent among Ashkenazi and Iraqi Jewish populations, reaching a frequency of 6% heterozygotes and 0.1–0.3% homozygotes. More than 65 mutations in the FXI gene have been reported, whereas fewer mutations (two to three) are found among affected Jewish populations.

Normal FXI clotting activity levels range from 70–150 U/dL. Levels vary depending on the presence of heterozygous, homozygous, or double heterozygous mutations with levels <1 U/dL seen in the latter two. Patients with FXI levels <10% of normal have a high risk of bleeding, but the phenotype does not always correlate with FXI clotting activity. The family history is informative, with the bleeding risk based on bleeding in kindreds. Clinically, spontaneous bleeding is rare, but mucocutaneous bleeding such as bruises, gum bleeding, epistaxis, hematuria, and menorrhagia are common, especially following trauma. This hemorrhagic phenotype suggests that tissues rich in fibrinolytic activity are more susceptible to FXI deficiency. Postoperative bleeding is common but not always present, even among patients with very low FXI levels.

FXI replacement is indicated in patients with severe disease for major surgical procedures. A negative history of bleeding complications following invasive procedures does not exclude the possibility of an increased risk for hemorrhage.

TREATMENT

Factor XI Deficiency

Sources of FXI are limited to FFP in the United States, whereas a plasma-derived FXI concentrate is available in other countries. FFP at doses of 15–20 mL/kg to increase levels by 10–20% can be given every other day in the setting of bleeding or major surgery as FXI has a half-life of 40–70 h. Antifibrinolytic drugs can be used for minor bleeds and as adjunctive treatment with FXI replacement with the exception of genitourinary tract bleeding. The development of an FXI inhibitor can be seen in 10% of severely FXI-deficient patients. Although inhibitors are not associated with spontaneous bleeding, bleeding with surgery or trauma can be severe; treatment with PCC/aPCC or recombinant activated FVII is effective. Data for use of a single very low dose of recombinant activated FVII (10–15 μg/kg) and an antifibrinolytic agent before surgery in patients with severe FXI deficiency are good, avoiding the need for plasma products.

RARE BLEEDING DISORDERS

Inherited disorders resulting from deficiencies of clotting factors other than FVIII, FIX, and FXI (Table 121-1) occur infrequently. Bleeding manifestations vary from generally asymptomatic as with dysfibrinogenemia or FVII deficiency to life-threatening as with FX or FXIII deficiency. In contrast to hemophilia, hemarthroses are rare, but bleeding in the mucosal tract or after umbilical cord clamping is common. Individuals heterozygous for plasma coagulation deficiencies are often asymptomatic. The laboratory assessment for the specific deficient factor following screening with general coagulation tests (Table 121-1) identifies the diagnosis.

Replacement therapy using FFP or PCCs for deficiencies provides adequate hemostasis for bleeds or prophylactic treatment, although specific concentrates for FX and fibrinogen are available. Cryoprecipitate or FXIII concentrate is needed for FXIII deficiency. FVII deficiency, like FXI, has an increased prevalence in the Ashkenazi Jewish population and is best treated with recombinant FVIIa rather than FFP or PCCs depending on the severity of bleeding or type of surgery. It should be noted that deficiency of FXII is associated with significant prolongation of the aPTT but no bleeding phenotype.

FAMILIAL MULTIPLE COAGULATION DEFICIENCIES

Several bleeding disorders are characterized by the inherited deficiency of more than one plasma coagulation factor. To date, the genetic defects in two of these diseases have been characterized, and they provide new insights into the regulation of hemostasis by gene-encoding proteins outside blood coagulation.

Combined Deficiency of FV and FVIII Patients with combined FV and FVIII deficiency exhibit ~5% of residual clotting activity of each factor, yet have a mild bleeding tendency, often following trauma. A mutation in the lectin mannose binding 1 (*LMAN1*) gene, a mannose-binding protein localized in the Golgi apparatus that functions as a chaperone for both FV and FVIII, is responsible. In other families, mutations in the multiple coagulation factor deficiency 2 (*MCFD2*) gene have been defined; this gene product forms a Ca^{2+}-dependent complex with *LMAN1*, providing cofactor activity for intracellular mobilization of both FV and FVIII. Replacement therapy to control or prevent bleeding consists of FFP to maintain FV levels and DDAVP or FVIII concentrate to achieve FVIII levels of 20–40%. Alternatively, platelets, which contain FV, can also be used.

Multiple Deficiencies of Vitamin K–Dependent Coagulation Factors Two enzymes involved in vitamin K metabolism have been associated with combined deficiency of all vitamin K–dependent proteins, including the procoagulant proteins prothrombin (II), VII, IX, and X and the anticoagulant proteins C and S. Vitamin K, a fat-soluble vitamin, is a cofactor for carboxylation of the gamma carbon of the glutamic acid residues in the vitamin K–dependent factors, a critical step for calcium and phospholipid binding **(Fig. 121-3)**. The enzymes γ-glutamylcarboxylase and epoxide reductase are critical for the metabolism and regeneration of vitamin K. Mutations in the genes encoding the γ-carboxylase (GGCX) or vitamin K epoxide reductase complex 1 (VKORC1) result in defective enzymes and thus in vitamin K–dependent factors with reduced activity, varying from 1–30%

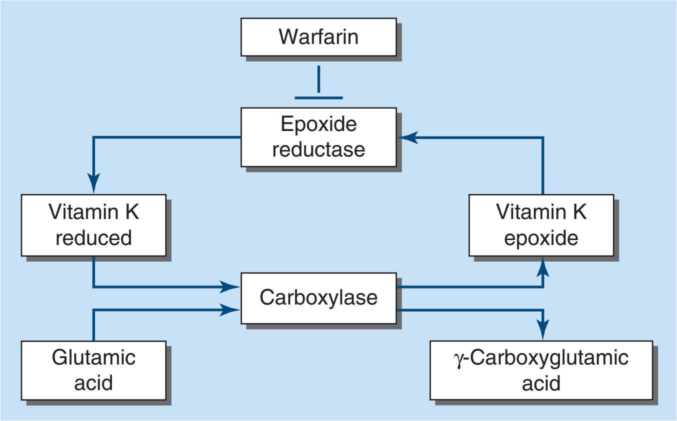

FIGURE 121-3 The vitamin K cycle. Vitamin K is a cofactor for the formation of γ-carboxyglutamic acid residues on coagulation proteins. Vitamin K–dependent γ-glutamylcarboxylase, the enzyme that catalyzes the vitamin K epoxide reductase, regenerates reduced vitamin K. Warfarin blocks the action of the reductase and competitively inhibits the effects of vitamin K.

of normal. Patients can have mild to severe bleeding episodes present from birth. Some patients respond to oral vitamin K_1 (5–20 mg/d) or parenteral vitamin K_1 at doses of 5–20 mg/week. For severe bleeding, replacement therapy with PCC may be necessary.

DISSEMINATED INTRAVASCULAR COAGULATION

In 2001, the International Society on Thrombosis and Haemostasis (ISTH) defined DIC as "an acquired syndrome characterized by the intravascular activation of coagulation with loss of localization arising from different causes that can originate from and cause damage to the microvasculature, which if sufficiently severe, can produce organ dysfunction." Many disparate processes are associated with DIC **(Table 121-2)**.

TABLE 121-2 Common Clinical Causes of Disseminated Intravascular Coagulation

SEPSIS	**IMMUNOLOGIC DISORDERS**
• Bacterial: Staphylococci, streptococci, pneumococci, meningococci, gram-negative bacilli • Viral • Mycotic • Parasitic • Rickettsial	• Acute hemolytic transfusion reaction • Organ or tissue transplant rejection • Immunotherapy • Graft-versus-host disease
TRAUMA AND TISSUE INJURY	**DRUGS**
• Brain injury (gunshot) • Extensive burns • Fat embolism • Rhabdomyolysis	• Fibrinolytic agents • Aprotinin • Warfarin (especially in neonates with protein C deficiency) • Prothrombin complex concentrates • Recreational drugs (amphetamines)
VASCULAR DISORDERS	**ENVENOMATION**
• Giant hemangiomas (Kasabach-Merritt syndrome) • Large vessel aneurysms (e.g., aorta)	• Snake • Insects
OBSTETRICAL COMPLICATIONS	**LIVER DISEASE**
• Abruptio placentae • Amniotic fluid embolism • Dead fetus syndrome • Septic abortion	• Fulminant hepatic failure • Cirrhosis • Fatty liver of pregnancy
CANCER	**MISCELLANEOUS**
• Adenocarcinoma (prostate, pancreas, etc.) • Hematologic malignancies (acute promyelocytic leukemia)	• Shock • Respiratory distress syndrome • Massive transfusion

The most common causes are bacterial sepsis, although viral and fungal sepsis can also cause DIC; trauma; obstetric causes such as abruptio placentae or amniotic fluid embolism; and malignant disorders, especially mucin-producing adenocarcinomas and acute promyelocytic leukemia. Activation of inflammatory pathways in response to infectious pathogens results in increased expression of tissue factor, activation of neutrophils and monocytes with release of cytokines and development of neutrophil extracellular traps, and release of polyphosphates that engage in cross-talk with the coagulation system to cause thrombin generation; this process is known as *thrombo-inflammation*. Damage to vascular endothelial cells results in the loss of their native antithrombotic properties; such damage especially occurs with sepsis and trauma. Systemic inflammatory response syndrome (SIRS) and cytokine storm are cytokine-mediated exuberant inflammatory responses often in the setting of infection that are associated with increased mortality and DIC. Purpura fulminans is a severe form of DIC resulting in thrombosis of extensive areas of the skin; it affects predominantly young children following viral or bacterial infection, particularly those with inherited or acquired hypercoagulability due to deficiencies of the components of the protein C pathway. Neonates homozygous for protein C deficiency can develop neonatal purpura fulminans with or without thrombosis of large vessels.

The central mechanism of DIC is the uncontrolled generation of thrombin by multiple mechanisms **(Fig. 121-4)**. Simultaneous disruption of the physiologic anticoagulant mechanisms and abnormal fibrinolysis further accelerate the process. These abnormalities contribute to systemic fibrin deposition in small and midsize vessels. The duration and intensity of the fibrin deposition can compromise the blood supply of many organs, especially the lung, kidney, liver, and brain, with consequent organ failure; for example, pulmonary microvascular thrombosis is a component of adult respiratory distress syndrome (ARDS). The sustained activation of coagulation and formation of fibrin can result in consumption of clotting factors and platelets, which in turn leads to systemic bleeding that can be aggravated by secondary hyperfibrinolysis that occurs in late stages of DIC.

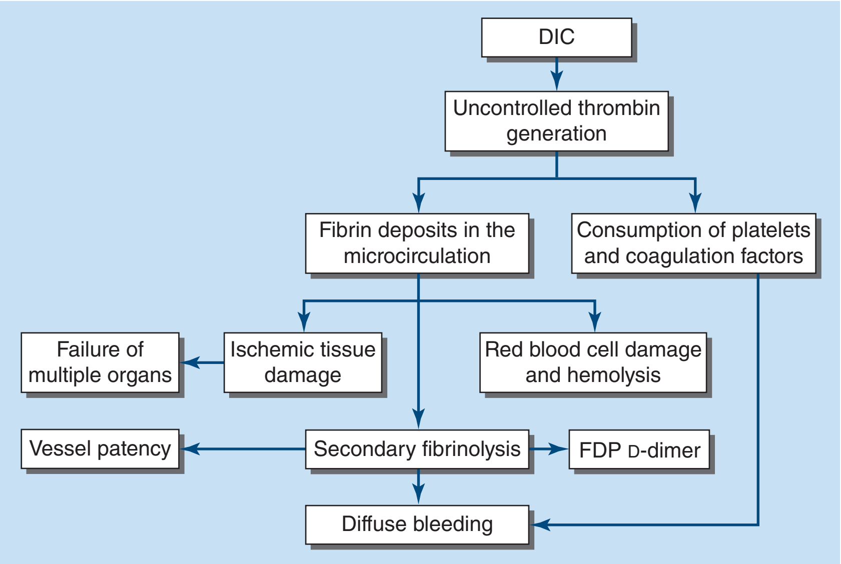
DIC
Uncontrolled thrombin generation
Fibrin deposits in the microcirculation
Consumption of platelets and coagulation factors
Failure of multiple organs
Ischemic tissue damage
Red blood cell damage and hemolysis
Vessel patency
Secondary fibrinolysis
FDP D-dimer
Diffuse bleeding

FIGURE 121-4 The pathophysiology of disseminated intravascular coagulation (DIC). Interactions between coagulation and fibrinolytic pathways result in bleeding and thrombosis in the microcirculation in patients with DIC. FDP, fibrin degradation product.

TABLE 121-3 International Society on Thrombosis and Haemostasis Criteria for Overt Disseminated Intravascular Congestion (DIC)

PARAMETER	VALUE	POINTS
Platelets	$>100{,}000 \times 10^9$/L	0
	>50–$<100 \times 10^9$/L	1
	$<50 \times 10^9$/L	2
D-Dimer[a]	Normal	0
	Moderate increase	2
	Severe increase	3
Prothrombin time (PT) prolonged	<3 s	0
	3–<6 s	1
	>6 s	2
Fibrinogen	>1 g/L	0
	<1 g/L	1
Total Score		**<5 Low-grade DIC** **>5 Overt DIC**

[a] D-Dimer assays are not standardized and have different ranges of normal. Check your institution range of normal to assess degree of increase.

Note: A score of <5 suggests nonovert DIC/low-grade DIC and should be repeated every 1–2 days. A score of >5 suggests overt DIC; lab values should be repeated daily to assess critical changes. Not to be used in pregnant patients.

Clinical manifestations of DIC are related to the magnitude of the imbalance of hemostasis, to the underlying disease, or to both. The most common clinical findings include petechiae, ecchymoses, and bleeding ranging from oozing from venipuncture sites to severe hemorrhage from the gastrointestinal tract, lung, or into the CNS. In chronic DIC, the bleeding symptoms are discrete and restricted to skin or mucosal surfaces. The hypercoagulability of DIC manifests as the occlusion of vessels in the microcirculation resulting in organ failure. Thrombosis of large vessels and cerebral embolism can also occur. Hemodynamic complications and shock are common among patients with acute DIC, due to the underlying disease, with mortality ranging from 30 to >80%.

Making the diagnosis of DIC can be difficult. The ISTH has developed a validated scoring tool to aid in the diagnosis of overt DIC with a separate tool for pregnant women. It incorporates platelet count, D-dimer level, PT, and fibrinogen level, and assigns points for different levels of each with the aggregate score helping to make the diagnosis of DIC **(Table 121-3)**. The peripheral smear should be assessed for schistocytes. The laboratory diagnosis of DIC should prompt a search for the underlying disease if not already apparent. In critically ill patients, these tests should be repeated over a period of 6–8 h as patients can rapidly deteriorate.

Chronic DIC Low-grade, compensated DIC can occur in clinical situations including giant hemangioma, metastatic carcinoma, or late gestation fetal demise. Plasma levels of fibrin degradation product or D-dimers are elevated. aPTT, PT, and fibrinogen values are within the normal range or high. Mild thrombocytopenia or normal platelet counts are also common findings. Red cell fragmentation is often detected but at a lower degree than in acute DIC.

Differential Diagnosis Distinguishing between DIC and severe liver disease is challenging and requires serial measurements of the laboratory parameters of DIC. Patients with severe liver disease manifest laboratory features including thrombocytopenia due to platelet sequestration, portal hypertension, or hypersplenism; decreased synthesis of coagulation factors and natural anticoagulants; and elevated levels of D-dimer. However, in contrast to DIC, these laboratory parameters in liver disease do not change rapidly.

Although microangiopathic disorders such as immune thrombotic thrombocytopenic purpura present with acute onset accompanied by thrombocytopenia, red cell fragmentation, and multiorgan failure, the clinical presentation and laboratory findings including presence of an inhibitor to ADAMTS13 assist in diagnosis **(Chap. 120)**.

TREATMENT

Disseminated Intravascular Coagulation

The morbidity and mortality associated with DIC are primarily related to the underlying disease. Management of the underlying disease is required to control and eliminate DIC; however, support with platelets and coagulation factors may be needed until the inciting cause is under control. Many patients with overt DIC are critically ill, usually requiring management in the intensive care unit to treat shock physiology and other manifestations of the underlying illness.

MANAGEMENT OF HEMORRHAGIC SYMPTOMS

Patients with active bleeding or at high risk of bleeding during invasive procedures or after chemotherapy require transfusion support; however, transfusion solely to correct mildly to moderately abnormal coagulation parameters is not indicated. Platelet transfusion for platelet counts <10,000–20,000/μL and replacement of fibrinogen and coagulation factors with FFP, cryoprecipitate, or fibrinogen concentrate as a source of fibrinogen are indicated with amounts determined by the degree of abnormal PT, aPTT, and fibrinogen levels, as well as severity of bleeding or bleeding risk with invasive procedures. For these situations, fibrinogen level should be maintained at >150 mg/dL and PT prolonged no more than 3 s above the upper limit of normal. Vitamin K should be given. Patients should be frequently monitored, and transfusion support adjusted as the patient's condition changes and dictates.

REPLACEMENT OF COAGULATION OR FIBRINOLYSIS INHIBITORS

Anticoagulants such as heparin, concentrates of antithrombin and thrombomodulin, and antifibrinolytic drugs have all been tried in the treatment of DIC. Low doses of continuous-infusion heparin (5–10 U/kg per h) may be effective in patients with low-grade DIC associated with solid tumors, acute promyelocytic leukemia, or in a setting with recognized thrombosis. Heparin is also indicated for the treatment of purpura fulminans. In acute hemorrhagic DIC, the use of heparin is likely to aggravate bleeding. The use of heparin in patients with severe DIC, although demonstrating improved coagulation parameters, is not associated with a survival benefit; professional society recommendations for use vary widely. Although the use of concentrates of the serine protease inhibitors, antithrombin and thrombomodulin, for sepsis demonstrated little efficacy in all treated patients, post hoc analyses of those with sepsis and confirmed DIC suggest a survival advantage and require further study. Activated protein C treatment for septic shock was withdrawn from the market years ago as findings in clinical practice did not replicate the mortality advantage seen in the clinical trial; impact on DIC was not evaluated.

In patients who have DIC characterized by a primary hyperfibrinolytic state with concomitant severe bleeding, the administration of antifibrinolytics may be considered. However, concern for increasing the risk of thrombosis has led to consideration of concomitant use of heparin. Patients with acute promyelocytic leukemia or those with chronic DIC associated with giant hemangiomas are among the few patients who may benefit from this therapy.

VITAMIN K DEFICIENCY

Vitamin K–dependent proteins are a heterogeneous group, including clotting factor proteins and proteins found in bone, lung, kidney, and placenta. Vitamin K mediates posttranslational modification of glutamate residues to γ-carboxylglutamate, which is necessary for calcium binding and proper assembly on phospholipid membranes (Fig. 121-3). Inherited mutations with decreased functional activity of the enzymes GGCX or VKORC1 (see above) result in bleeding disorders. Vitamin K in the diet is often limiting for the carboxylation reaction; thus, recycling of the vitamin K by these enzymes is essential to maintain normal levels of vitamin K–dependent proteins. In adults, severe vitamin K deficiency due to low dietary intake is rare but is common in association with the use of broad-spectrum antibiotics or with disease or surgical interventions that affect the ability of the intestinal tract to absorb vitamin K, through anatomic alterations or by changing the fat content of bile salts and pancreatic enzymes in the proximal small bowel. Chronic liver diseases such as primary biliary cirrhosis also deplete vitamin K stores. Neonatal vitamin K deficiency and the resulting hemorrhagic disease of the newborn have been almost entirely eliminated by routine administration of vitamin K to all neonates. Prolongation of PT values is the most common and earliest finding in vitamin K–deficient patients due to the short half-life of FVII and occurs before prolongation of the aPTT. Parenteral administration of 10 mg of vitamin K is sufficient to restore normal levels of clotting factor within 8–10 h. More rapid correction of the coagulopathy requires replacement with FFP or PCC, with the choice depending on patient intravascular volume status and need for rapidity of correction. The reversal of excessive anticoagulant therapy with vitamin K antagonists, such as warfarin, can be achieved by minimal doses of vitamin K (1 mg orally or by intravenous injection) for asymptomatic patients. This strategy can diminish the risk of bleeding while maintaining therapeutic anticoagulation for an underlying prothrombotic state. For emergent reversal of warfarin in the setting of life-threatening bleeding or need for emergency surgery, use of 4F-PPC is the standard of care.

In patients with underlying vascular disease, vascular trauma, atrial fibrillation, and other comorbidities, re-initiation of anticoagulation needs to be carefully considered to prevent subsequent thromboembolic complications.

COAGULATION DISORDERS ASSOCIATED WITH LIVER FAILURE

The liver is the site of synthesis and clearance of most procoagulant and natural anticoagulant proteins and of essential components of the fibrinolytic system. Acute liver failure is associated with a high risk of bleeding due to deficient synthesis of procoagulant factors and enhanced fibrinolysis; hepatologists refer to this as accelerated intravascular coagulation and fibrinolysis (AICF). Thrombocytopenia is common in patients with liver disease and may be due to decreased thrombopoietin that is synthesized in the liver, congestive splenomegaly (hypersplenism), or immune-mediated shortened platelet life span (primary biliary cirrhosis). In addition, several anatomic abnormalities secondary to underlying liver disease further increase the risk of bleeding (Table 121-4). Dysfibrinogenemia is a relatively common finding in patients with liver disease due to impaired fibrin polymerization. The development of DIC in patients with chronic liver disease is not uncommon and may enhance the risk for bleeding. Laboratory evaluation is mandatory for an optimal therapeutic strategy, either to control ongoing bleeding or before invasive procedures. Typically, these patients present with prolonged PT, aPTT, and thrombin time (TT) depending on the degree of liver damage, thrombocytopenia, and normal or slight increase in D-dimer. Fibrinogen levels are low only in fulminant hepatitis, decompensated cirrhosis, advanced liver disease, or in the presence of DIC. The presence of prolonged TT and normal fibrinogen and D-dimer levels suggests dysfibrinogenemia. FVIII levels are often normal or elevated in patients with liver failure, and decreased levels suggest superimposed DIC. FV is only synthesized in the hepatocyte and is not a vitamin K–dependent protein; therefore, reduced levels of FV may be an indicator of liver failure. Normal levels of FV and low levels of FVII suggest vitamin K deficiency. Vitamin K levels may be reduced in patients with liver failure due to compromised storage in hepatocellular disease, changes in bile acids, or cholestasis that can diminish the absorption of vitamin K. Replacement with intravenous vitamin K may improve hemostasis.

TABLE 121-4 Coagulation Disorders and Hemostasis in Liver Disease

Bleeding
Portal hypertension
Esophageal varices
Thrombocytopenia
Splenomegaly
Chronic or acute DIC
Decreased synthesis of clotting factors
Hepatocyte failure
Vitamin K deficiency
Systemic fibrinolysis
DIC
Dysfibrinogenemia
Thrombosis
Decreased synthesis of coagulation inhibitors: protein C, protein S, antithrombin
Hepatocyte failure
Vitamin K deficiency (protein C, protein S)
Failure to clear activated coagulation proteins (DIC)
Dysfibrinogenemia

Abbreviation: DIC, disseminated intravascular coagulation.

Patients with chronic stable liver disease are now recognized to have rebalanced hemostasis with simultaneous changes in both procoagulant and anticoagulant proteins. Although traditional clinical laboratory tests may suggest increased bleeding risk, these patients in fact do not have spontaneous bleeding and often do not need treatment for minor to moderate bleeding risk procedures. If the patient is bleeding, treatment with FFP was the standard approach to correcting hemostasis in patients with acute liver failure; however, the use of 4F-PCC is now favored due to lower volume, less increase in portal pressure, reduced risk of circulatory overload, and other complications associated with FFP transfusion. As in any clinical situation, treatment should not be given simply to correct laboratory abnormalities in a patient who is not bleeding or with no need for invasive procedures. Platelet concentrates are indicated when platelet counts are <10,000–20,000/μL to control bleeding or immediately before an invasive procedure if counts are <50,000/μL. Cryoprecipitate is indicated only when fibrinogen levels are <100–150 mg/mL unless the patient is bleeding, in which case a higher target is used. The use of antifibrinolytic drugs as adjuncts to control bleeding in patients with liver failure is not thought to result in an increased risk of thrombosis; however, their impact on acute thrombosis propagation is not well studied.

Liver Disease and Thromboembolism Bleeding in patients with stable liver disease is often mild or even asymptomatic. However, as the disease progresses, the hemostatic balance is precarious and easily disturbed; comorbid complications such as infections and renal failure can rapidly upset this balance (Fig. 121-5). Past assumptions based on abnormal coagulation tests have been that patients with liver disease have a decreased risk of thrombosis; however, multiple factors contribute to hypercoagulability, including decreased levels of the natural anticoagulant proteins S and C, as well as endothelial cell changes and hemodynamic changes that result in stasis such that portal vein thrombosis is common. Patients with liver disease can also develop deep-vein thrombosis and pulmonary embolism; those with cirrhosis appear to have a 1.5- to 2-fold increase in the rate of venous thromboembolism (VTE). Patients with compensated cirrhosis do not appear to have increased bleeding with the use of VTE prophylaxis or even therapeutic dose heparin to treat acute portal vein thrombosis when carefully managed. In the outpatient setting, warfarin is avoided, but low-molecular-weight heparin and direct oral anticoagulants have been safely used to treat thrombosis.

Acquired Inhibitors of Coagulation Factors An acquired inhibitor is an immune-mediated disease characterized by the presence of an autoantibody against a specific clotting factor. Almost half of patients with an acquired factor inhibitor will have an underlying autoimmune or immunoproliferative disorder, have a malignancy, or be peripartum. FVIII is the most common target of antibody formation and is often referred to as acquired hemophilia A, but inhibitors to prothrombin (FII), FV, FIX, FX, and FXI are also reported. Acquired inhibitor to FVIII occurs predominantly in older adults (median age of 60 years) but occasionally in pregnant or postpartum women with

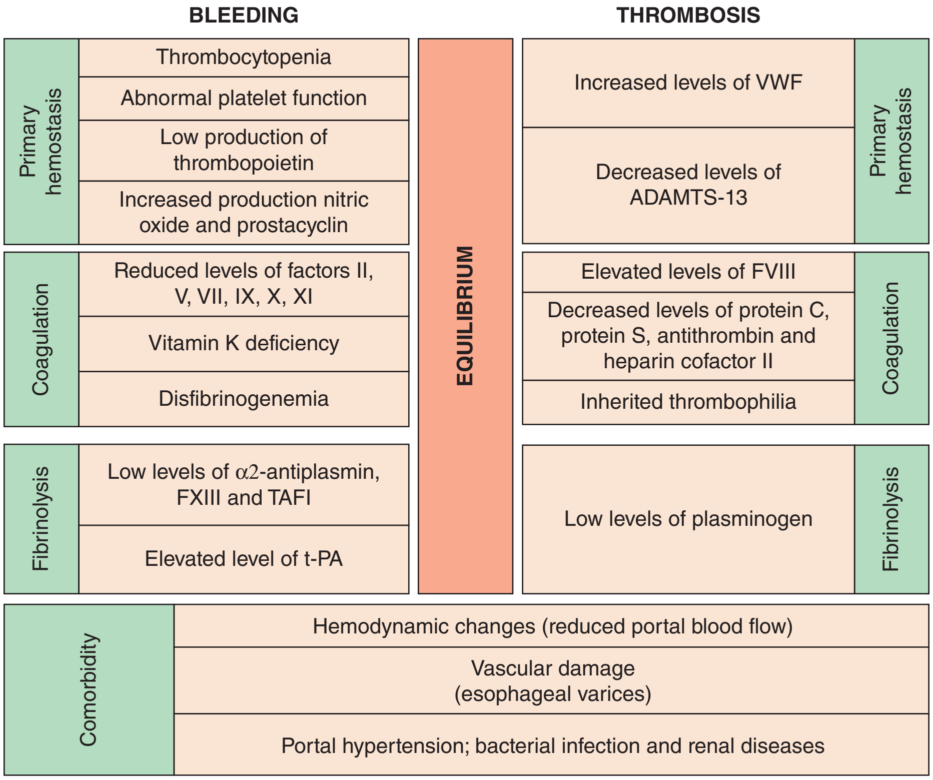
BLEEDING
THROMBOSIS
Primary hemostasis
Thrombocytopenia
Abnormal platelet function
Low production of thrombopoietin
Increased production nitric oxide and prostacyclin
Coagulation
Reduced levels of factors II, V, VII, IX, X, XI
Vitamin K deficiency
Disfibrinogenemia
Fibrinolysis
Low levels of α2-antiplasmin, FXIII and TAFI
Elevated level of t-PA
EQUILIBRIUM
Increased levels of VWF
Decreased levels of ADAMTS-13
Primary hemostasis
Elevated levels of FVIII
Decreased levels of protein C, protein S, antithrombin and heparin cofactor II
Inherited thrombophilia
Coagulation
Low levels of plasminogen
Fibrinolysis
Comorbidity
Hemodynamic changes (reduced portal blood flow)
Vascular damage (esophageal varices)
Portal hypertension; bacterial infection and renal diseases

FIGURE 121-5 Balance of hemostasis in liver disease. TAFI, thrombin-activated fibrinolytic inhibitor; t-PA, tissue plasminogen activator; VWF, von Willebrand factor.

no previous history of bleeding. Bleeding episodes occur commonly in soft tissues, the gastrointestinal or urinary tracts, and skin. In contrast to congenital hemophilia, hemarthrosis is rare in these patients. Retroperitoneal hemorrhages and other life-threatening bleeding may appear suddenly. The overall mortality in untreated patients ranges from 8–22%, and most deaths occur within the first few weeks after presentation. The diagnosis is based on the prolonged aPTT with normal PT and TT and a mixing study that does not correct with normal pooled plasma. The Bethesda assay using factor-specific deficient plasma as performed for inhibitor detection in hemophilia will confirm the diagnosis. Treatment of acquired inhibitors of coagulation factors requires control of bleeding and eradication of the inhibitor. Many patients can have life-threatening bleeding. The use of activated "bypass products" such as aPCC or recombinant FVIIa is required. The use of recombinant porcine FVIII can be effective for acquired inhibitors of FVIII. The use of emicizumab to treat acquired FVIII inhibitors has been reported, and trials in this population are underway in Europe.

In contrast to inhibitors in patients with congenital factor deficiencies, acquired inhibitors are typically responsive to immune suppression, and treatment should be initiated early for most cases. High-dose intravenous γ-globulin and anti-CD20 monoclonal antibody are reported to be effective in patients with autoantibodies to FVIII; however, no firm evidence confirms that these alternatives are superior to the first line of immunosuppressive drugs (glucocorticoids and cyclophosphamide), effective in 70% of patients. Relapse of an inhibitor to FVIII is relatively common (up to 20%) within the first 6 months following withdrawal of immunosuppression; patients should be followed up regularly for relapse.

Topical plasma-derived bovine and human thrombin are commonly used during major cardiovascular, thoracic, neurologic, and pelvic surgeries as well as in trauma patients with extensive burns. Antibody formation to the xenoantigen or its contaminant (bovine clotting protein) has the potential to cross-react with human clotting factors, particularly FV and thrombin, and can result in bleeding that can be life-threatening. The development of antibodies to FV with the use of topical preparations of recombinant human thrombin has also been reported. The clinical diagnosis of these acquired coagulopathies is rare but is often complicated by the fact that the bleeding episodes may be detectable during or immediately following major surgery and could be assumed to be due to the procedure itself.

The risk of developing a cross-reacting antibody is increased by repeated exposure to topical thrombin preparations. Thus, a careful medical history of previous surgical interventions that may have occurred even decades earlier is critical to assessing risk.

The laboratory abnormalities include a combined prolongation of the aPTT and PT that often fails to improve by transfusion of FFP and vitamin K, and a mixing study that does not correct with normal pooled plasma. The specificity of the antibody is determined by the measurement of the residual activity of human FV or other suspected human clotting factor. No assays specific for bovine thrombin coagulopathy are commercially available.

No treatment guidelines have been established. Platelet transfusions have been used as a source of FV replacement for patients with FV inhibitors. FFP and vitamin K supplementation may function as co-adjuvants rather than as effective treatments for the coagulopathy itself. Experience with recombinant FVIIa as a bypass agent is limited, and outcomes have been generally poor. Specific treatments to eradicate the antibodies based on immunosuppression with glucocorticoids, intravenous immunoglobulin, or serial plasmapheresis have been sporadically reported. Patients should be advised to avoid any topical thrombin sealant in the future.

The presence of lupus anticoagulant can be associated with venous or arterial thrombotic disease. However, bleeding has also been reported rarely with lupus anticoagulants due to antibodies to prothrombin, resulting in hypoprothrombinemia. Both disorders show a prolonged aPTT that does not correct on mixing. To distinguish acquired inhibitors from lupus anticoagulant, note that the dilute Russell viper venom time (dRVVT) and the hexagonal-phase phospholipids test will be negative in patients with an acquired inhibitor and positive in patients with lupus anticoagulants. Moreover, lupus anticoagulant interferes with the clotting activity of many factors (FVIII, FIX, FXI, FXII), which can be assessed in the clinical laboratory; acquired inhibitors are specific to a single factor.

ACKNOWLEDGMENT

Valder Arruda and Katherine High wrote this chapter in prior editions and some material from their chapter is included here.

FURTHER READING

KITAZAWA T, SHIMA M: Emicizumab, a humanized bispecific antibody to coagulation factors IXa and X with a factor VIIIa-cofactor activity. Int J Hematol 111:20, 2020.

LEVI M, SCULLY M: How I treat disseminated intravascular coagulation. Blood 131:845, 2018.

MENEGATTI M et al: Management of rare acquired bleeding disorders. Hematology Am Soc Hematol Educ Program 2019:80, 2019.

PIPE S et al: Delivery of gene therapy in haemophilia treatment centres in the United States: Practical aspects of preparedness and implementation. Haemophilia 29:1430, 2023.

ROBERTS LN: How to manage hemostasis in patients with liver disease during interventions. Hematology Am Soc Hematol Educ Program 2023:274, 2023.

SRIVASTAVA A et al: WFH guidelines for the management of hemophilia, 3rd edition. Haemophilia 26:1, 2020.

122

Arterial and Venous Thrombosis

Jane E. Freedman, Joseph Loscalzo

OVERVIEW OF THROMBOSIS

GENERAL OVERVIEW

Thrombosis is hemostasis "at the wrong place and at the wrong time" (MacFarlane). The obstruction of blood flow due to the formation of thrombus may result in tissue anoxia and damage, and it is a major cause of morbidity and mortality in a wide range of arterial and venous diseases and patient populations. As reported in 2020, 655,000 Americans die from heart disease each year or about 1 in 4 deaths. In 2020, coronary disease killed 382,820 people and stroke killed 160,747 in the United States with approximately 805,000 suffering from heart attacks and 795,000 having strokes.

It is estimated that as many as 600,000 people each year have a pulmonary embolism or deep-venous thrombotic event, and 60,000–80,000 Americans die of these conditions annually. In the nondiseased state, physiologic hemostasis reflects a delicate interplay between factors that promote and inhibit blood clotting, favoring the former. This response is crucial as it prevents uncontrolled hemorrhage and exsanguination following injury. In specific settings, the same processes that regulate normal hemostasis can cause pathologic thrombosis, leading to arterial or venous occlusion. Importantly, many commonly used therapeutic interventions may also alter the thrombotic–hemostatic balance adversely.

Hemostasis and thrombosis primarily involve the interplay among three factors: the vessel wall, coagulation and fibrinolytic proteins, and platelets. Many prevalent acute vascular diseases are due to thrombus formation within a vessel, including myocardial infarction, thrombotic cerebrovascular events, and venous thrombosis. Although the end result is vessel occlusion and tissue ischemia, the pathophysiologic processes governing these pathologies have similarities as well as distinct differences. While many of the pathways regulating thrombus formation are similar to those that regulate hemostasis, the processes triggering or perpetuating thrombosis may be distinct and can vary in different clinical and genetic settings. In venous thrombosis, primary hypercoagulable states reflecting defects in the proteins governing coagulation and/or fibrinolysis or secondary hypercoagulable states involving abnormalities of blood vessels and blood flow or stasis lead to thrombosis. By contrast, arterial thrombosis is highly dependent on the state of the vessel wall, the platelet, and factors related to blood flow.

ARTERIAL THROMBOSIS

OVERVIEW OF ARTERIAL THROMBOSIS

In arterial thrombosis, platelets and abnormalities of the vessel wall typically play a key role in vessel occlusion. Arterial thrombus forms via a series of sequential steps in which platelets adhere to the vessel wall, additional platelets are recruited, and thrombin is activated **(Fig. 122-1)**. The regulation of platelet adhesion, activation, aggregation, and recruitment will be described in detail below. In addition, while the primary function of platelets is regulation of hemostasis, our understanding of their role in other processes, such as immunity, metastasis, wound healing, and inflammation, continues to evolve.

ARTERIAL THROMBOSIS AND VASCULAR DISEASE

Arterial thrombosis is a major cause of morbidity and mortality both in the United States and, increasingly, worldwide. Although the rates have declined in the United States, the overall burden remains high. Overall, in 2020, heart disease was estimated to cause about 1 of every 4 deaths in the United States. In addition to the 605,000 Americans who will have a new coronary event annually, an additional 200,000 myocardial infarctions occur in those with previous heart attacks. Although the rate of strokes has fallen, each year about 795,000 people experience a new or recurrent ischemic stroke. In 2018, about 1 in 6 deaths from cardiovascular disease were due to stroke in the United States.

THE PLATELET

Many processes in platelets have parallels with other cell types, such as the presence of specific receptors and signaling pathways; however, unlike most cells, platelets lack a nucleus and are unable to adapt to changing biologic settings by altered gene transcription. Platelets sustain limited protein synthetic capacity from megakaryocyte-derived and intracellularly transported messenger RNA (mRNA) and microRNA (miRNA). Most of the molecules needed to respond to various stimuli, however, are maintained in storage granules and membrane compartments.

Platelets are disc-shaped, very small, anucleate cells (1–5 μm in diameter) that circulate in the blood at concentrations of 200–400,000/μL, with an average life span of 7–10 days. Platelets are derived from megakaryocytes, polyploidal hematopoietic cells found in the bone marrow. The primary regulator of platelet formation is thrombopoietin (TPO). The precise mechanism by which megakaryocytes produce and release fully formed platelets is unclear, but the process likely involves formation of proplatelets, pseudopod-like structures generated by the

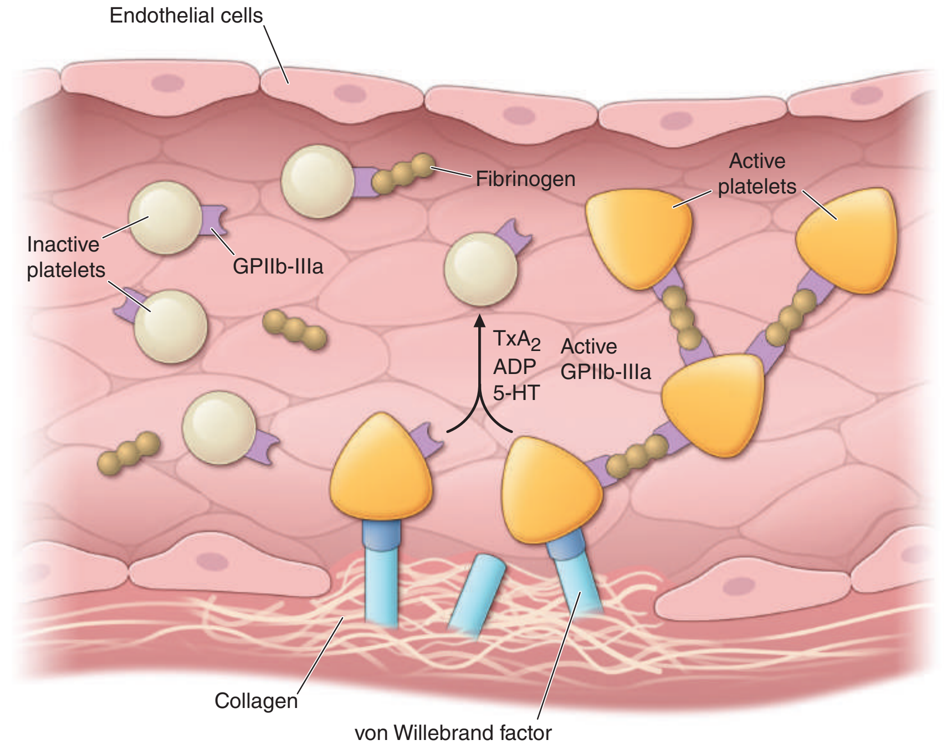

FIGURE 122-1 Platelet activation and thrombosis. Platelets circulate in an inactive form in the vasculature. Damage to the endothelium and/or external stimuli activate platelets that adhere to the exposed subendothelial von Willebrand factor and collagen. This adhesion leads to activation of the platelet, shape change, and the synthesis and release of thromboxane (TxA_2), serotonin (5-HT), and adenosine diphosphate (ADP). Platelet stimuli cause conformational change in the platelet integrin glycoprotein (GP) IIb/IIIa receptor, leading to the high-affinity binding of fibrinogen and the formation of a stable platelet thrombus.

evagination of the cytoplasm from which platelets bud. After release into the circulation, (young, large) platelets may continue to divide. Platelet granules are synthesized in megakaryocytes before thrombopoiesis and contain an array of prothrombotic, proinflammatory, and antimicrobial mediators. The two major types of platelet granules, alpha and dense, are distinguished by their size, abundance, and content. Alpha-granules contain soluble coagulation proteins, adhesion molecules, growth factors, integrins, cytokines, and inflammatory modulators. Platelet dense-granules are smaller than alpha-granules and less abundant. Whereas alpha-granules contain proteins that may be more important in the inflammatory response, dense-granules contain high concentrations of small molecules, including adenosine diphosphate (ADP) and serotonin, that influence platelet aggregation and other related vascular processes, such as vasomotor tone.

Platelet Adhesion (See Fig. 122-1) The formation of a thrombus is initiated by the adherence of platelets to the damaged vessel wall. Damage exposes subendothelial components responsible for triggering platelet reactivity, including collagen, von Willebrand factor, fibronectin, and other adhesive proteins, such as vitronectin and thrombospondin. The hemostatic response may vary, depending on the extent of damage, the specific proteins exposed, and flow conditions. Certain proteins are expressed on the platelet surface that subsequently regulate collagen-induced platelet adhesion, particularly under flow conditions, and include glycoprotein (GP) IV, GPVI, and the integrin $\alpha_2\beta_1$. The platelet GPIb-IX-V complex adhesive receptor is central both to platelet adhesion and to the initiation of platelet activation. Damage to the blood vessel wall exposes subendothelial von Willebrand factor and collagen to the circulating blood. The GPIb-IX-V complex binds to the exposed von Willebrand factor, causing platelets to adhere (Fig. 122-1). In addition, the engagement of the GPIb-IX-V complex with ligand induces signaling pathways that lead to platelet activation. von Willebrand factor–bound GPIb-IX-V promotes a calcium-dependent conformational change in the GPIIb/IIIa receptor, transforming it from an inactive low-affinity state to an active high-affinity receptor for fibrinogen.

Platelet Activation The activation of platelets is controlled by a variety of surface receptors that regulate various functions in the activation process. Platelet receptors control many distinct processes and are stimulated by a wide variety of agonists and adhesive proteins that result in variable degrees of activation. In general terms, the stimulation of platelet receptors triggers two specific processes: (1) activation of internal signaling pathways that lead to further platelet activation and granule release, and (2) the capacity of the platelet to bind to other adhesive proteins/platelets. Both of these processes contribute to the formation of a thrombus. Stimulation of nonthrombotic receptors results in platelet adhesion or interaction with other vascular cells, including endothelial cells, neutrophils, and mononuclear cells.

Many families and subfamilies of receptors are found on platelets that regulate a variety of platelet functions. These include the seven transmembrane receptor family, which is the main agonist-stimulated receptor family. Several seven transmembrane receptors are found on platelets, including the ADP receptors, prostaglandin receptors, lipid receptors, and chemokine receptors. Receptors for thrombin comprise the major seven transmembrane receptors found on platelets. Among this last group, the first identified was the protease activation receptor 1 (PAR1). The PAR class of receptors has a distinct mechanism of activation that involves specific cleavage of the N-terminus by thrombin, which, in turn, acts as a ligand for the receptor. Other PAR receptors are present on platelets, including PAR2 (not activated by thrombin) and PAR4. Adenosine receptors are responsible for transduction of ADP-induced signaling events, which are initiated by the binding of ADP to purinergic receptors on the platelet surface. There are several distinct ADP receptors, classified as $P2X_1$, $P2Y_1$, and $P2Y_{12}$. The activation of both the $P2Y_{12}$ and $P2Y_1$ receptors is essential for ADP-induced platelet aggregation. The thienopyridine derivatives, clopidogrel and prasugrel, are clinically used inhibitors of ADP-induced platelet aggregation.

Platelet Aggregation Activation of platelets results in a rapid series of signal transduction events, including tyrosine kinase, serine/threonine kinase, and lipid kinase activation. In unstimulated platelets, the major platelet integrin GPIIb/IIIa is maintained in an inactive conformation and functions as a low-affinity adhesion receptor for fibrinogen. This integrin is unique as it is only expressed on platelets. After stimulation, the interaction between fibrinogen and GPIIb/IIIa forms intercellular connections between platelets, leading to the formation of a platelet aggregate (Fig. 122-1). A calcium-sensitive conformational change in the extracellular domain of GPIIb/IIIa enables the high-affinity binding of soluble plasma fibrinogen as a result of a complex network of inside-out signaling events. The GPIIb/IIIa receptor serves as a bidirectional conduit with GPIIb/IIIa-mediated signaling (outside-in signaling) occurring immediately after the binding of fibrinogen. This binding interaction leads to additional intracellular signaling that further stabilizes the platelet aggregate and transforms platelet aggregation from a reversible to an irreversible process (inside-out signaling).

THE ROLE OF PLATELETS AND THROMBOSIS IN INFLAMMATION

Inflammation plays an important role during the acute thrombotic phase of acute coronary and other vascular occlusive syndromes. In the setting of acute upper respiratory infections, people are at higher risk of myocardial infarction and thrombotic stroke. Patients with acute coronary syndromes have not only increased interactions between platelets (homotypic aggregates) but also increased interactions between platelets and leukocytes (heterotypic aggregates) detectable in circulating blood. These latter aggregates form when platelets are activated, often directly by pathogens, and adhere to circulating leukocytes as part of their contribution to the immune process. Platelets bind via P-selectin (CD62P) expressed on the surface of activated platelets to the leukocyte receptor, P-selectin glycoprotein ligand 1 (PSGL-1). This association leads to increased expression of CD11b/CD18 (Mac-1) on leukocytes, which amplifies immunity but may also support further interactions with platelets partially via bivalent fibrinogen linking this integrin with its platelet surface counterpart, GPIIb/IIIa. Platelet surface P-selectin also induces the expression of tissue factor on monocytes, which promotes fibrin formation.

In addition to platelet–monocyte aggregates, the immunomodulator, soluble CD40 ligand (CD40L or CD154), also reflects a link between thrombosis and inflammation. The CD40 ligand is a trimeric transmembrane protein of the tumor necrosis factor family and, with its receptor CD40, is an important contributor to the inflammatory process, leading both to thrombosis and atherosclerosis. While many immunologic and vascular cells have been found to express CD40 and/or CD40 ligand, in platelets, CD40 ligand is rapidly translocated to the surface after stimulation and is upregulated in the newly formed thrombus. The surface-expressed CD40 ligand is cleaved from the platelet to generate a soluble fragment (soluble CD40 ligand).

Links have also been established among platelets, infection, immunity, and inflammation. Bacterial and viral infections are associated with a transient increase in the risk of acute thrombotic events, such as acute myocardial infarction and stroke. In addition, platelets contribute significantly to the pathophysiology and high mortality rates of sepsis. The expression, functionality, and signaling pathways of Toll-like receptors (TLRs) have been established in platelets. Stimulation of platelet TLR2, TLR3, and TLR4 directly and indirectly activates the platelet's thrombotic and inflammatory responses, and live bacteria induce a proinflammatory response in platelets in a TLR2-dependent manner, suggesting a mechanism by which specific bacteria and bacterial components can directly activate platelet-dependent thrombosis. Additionally, viruses, such as SARS-CoV-2, HIV, hepatitis C virus, and Dengue, are also known to elevate thrombosis; recently, platelets have been shown to regulate immune responses to viruses via receptors TLR7 and TLR8.

Risk Factors for Arterial Thrombosis In addition to immune burden, various factors increase the risk of developing arterial thrombosis. Classically, the cardiovascular-dependent risk factors implicated in thrombosis have been hypertension, high levels of low-density lipoprotein cholesterol, and smoking. However, diabetes, pregnancy, age, and chemotherapeutic agents may also contribute to arterial thrombosis.

Stillbirth and loss of multiple pregnancies may increase the risk of ischemic stroke and myocardial infarction, as does hormonal replacement therapy. Systemic lupus erythematosus and rheumatoid arthritis are now well-recognized risks for thrombosis, and the former, in particular, may contribute in the pediatric population. The antiphospholipid syndrome is also another widely recognized autoimmune prothrombotic risk for arterial (and venous) thrombosis.

GENETICS OF ARTERIAL THROMBOSIS

Some studies have associated arterial thrombosis with genetic variants **(Table 122-1*A*)**; however, the associations have been weak and not consistently confirmed in larger series. Platelet count and mean platelet volume have been studied by genome-wide association studies (GWAS), and this approach identified signals located to noncoding regions. Of 15 quantitative trait loci associated with mean platelet volume and platelet count, one located at 12q24 is also a risk locus for coronary artery disease.

In the area of genetic variability and platelet function, studies have primarily dealt with pharmacogenetics, the field of pharmacology dealing with the interindividual variability in drug response based on genetic determinants **(Table 122-2)**. This focus has been driven by the wide variability among individuals in terms of response to antithrombotic drugs and the lack of a common explanation for this variance. The best described is the issue of "aspirin resistance," although heterogeneity for other antithrombotics (e.g., clopidogrel) has also been extensively examined. Primarily, platelet-dependent genetic determinants have been defined at the level of (1) drug effect, (2) drug compliance, and (3) drug metabolism. Many candidate platelet genes have been studied for their interaction with antiplatelet and antithrombotic agents.

TABLE 122-1 Heritable Causes of Arterial and Venous Thrombosis

A. Arterial Thrombosis
Platelet Receptors
β3 and α2 integrins
P$_{I}$A2 polymorphism
Fc(gamma)RIIA
GPIV T13254C polymorphism
GPIb
Thrombin receptor PAR1-5061 → D
Redox Enzymes
Plasma glutathione peroxidase, GPx3, promoter haplotype H2
H2 promoter haplotype
Endothelial nitric oxide synthase
–786T/C, –922A/G, –1468T/A
Paraoxonase
–107T allele, 192R allele
Homocysteine
Cystathionine β-synthase 833T → C
5,10-Methylene tetrahydrofolate reductase (MTHFR) 677C → T
B. Venous Thrombosis
Procoagulant Proteins
Fibrinogen
–455G/A, –854G/A
Prothrombin (20210G → A)
Protein C Anticoagulant Pathway
Factor V Leiden: 1691G → A (Arg506Gln)
Thrombomodulin 1481C → T (Ala455Val)
Fibrinolytic Proteins with Known Polymorphisms
Tissue plasminogen activator (tPA)
7351C/T, 20 099T/C in exon 6, 27 445T/A in intron 10
Plasminogen activator inhibitor (PAI-1)
4G/5G insertion/deletion polymorphism at position –675
Homocysteine
Cystathionine β-synthase 833T → C
5,10-MTHFR 677C → T

TABLE 122-2 Genetic Variation and Pharmacogenetic Responses to Platelet Inhibitors

POTENTIAL GENE ALTERED	TARGET THERAPEUTIC CLASS	SPECIFIC DRUG
P2Y1 and *P2Y12 CYP2C19, CYP3A4, CYP3A5*	ADP receptor inhibitors	Clopidogrel, prasugrel
COX1, COX2	Cyclooxygenase inhibitors	Aspirin
PIA1/A2	Receptor inhibitors	Abciximab, eptifibatide, tirofiban
INTB3, GPIbA	Glycoprotein IIb-IIIa receptor inhibitors	

Many patients have an inadequate response to the inhibitory effects of aspirin. Heritable factors contribute to the variability; however, ex vivo tests of residual platelet responsiveness after aspirin administration have not provided firm evidence for a pharmacogenetic interaction between aspirin and *COX1* or other relevant platelet receptors. As such, currently there is no clinical indication for genotyping to optimize aspirin's antiplatelet efficiency. For the platelet P2Y12 receptor inhibitor clopidogrel, additional data suggest that genetics may affect the drug's responsiveness and utility. The responsible genetic variant appears not to be the expected P2Y12 receptor but an enzyme responsible for drug metabolism. Clopidogrel is a prodrug, and liver metabolism by specific cytochrome P450 enzymes is required for activation. The genes encoding the CYP-dependent oxidative steps are polymorphic, and carriers of specific alleles of the CYP2C19 and CYP3A4 loci have increased platelet aggregability. Increased platelet activity has also been specifically associated with the CYP2C19*2 allele, which causes loss of platelet function in select patients. Because these are common genetic variants, this observation has been shown to be clinically relevant in large studies. In summary, although the loss-of-function polymorphisms in CYP2C19 is the strongest individual variable affecting pharmacokinetics and antiplatelet response to clopidogrel, it only accounts for 5–12% of the variability in ADP-induced platelet aggregation by clopidogrel. In addition, genetic variables do not appear to contribute significantly to the clinical outcomes of patients treated with the P2Y12 receptor antagonists prasugrel or ticagrelor.

VENOUS THROMBOSIS

OVERVIEW OF VENOUS THROMBOSIS

Coagulation is the process by which thrombin is activated and soluble plasma fibrinogen is converted into insoluble fibrin. These steps account for both normal hemostasis and the pathophysiologic processes influencing the development of venous thrombosis. The primary forms of venous thrombosis are deep-vein thrombosis (DVT) in the extremities and the subsequent embolization to the lungs (pulmonary embolism), referred to together as venous thromboembolic disease (VTE). Although the majority of venous thromboembolic events occur as pulmonary embolism or DVT of the lower extremities, up to 10% of events may occur in other vascular locations. Venous thrombosis has both heritable causes **(Table 122-1*B*)** and acquired causes **(Table 122-3)**.

DEEP-VENOUS THROMBOSIS AND PULMONARY EMBOLISM

It is estimated that DVT or pulmonary embolism (PE) occurs in ~1–2 individuals per 1000 each year, resulting in 300,000–600,000 new cases of venous thromboembolism each year in the United States. Approximately, 60,000–80,000 deaths are attributed to DVT or PE annually. Of new cases, up to 30% of patients die within 30 days and one-fifth suffer

TABLE 122-3 Acquired Causes of Venous Thrombosis

Surgery
Neurosurgery
Major abdominal surgery
Other
Trauma
Antiphospholipid syndrome
Pregnancy
Long-distance travel
Obesity
Oral contraceptives/hormone replacement
Myeloproliferative disorders
Polycythemia vera

sudden death due to PE; 30% go on to develop recurrent VTE within 10 years. Data from the Atherosclerosis Risk in Communities (ARIC) study reported a 9% 28-day fatality rate from DVT and a 15% fatality rate from PE. PE in the setting of cancer has a 25% fatality rate. The mean incidence of first DVT in the general population is 5 per 10,000 person-years; the incidence is similar in males and females when adjusting for factors related to reproduction and birth control and increases dramatically with age from 2–3 per 10,000 person-years at 30–49 years of age to 20 per 10,000 person-years at 70–79 years of age.

OVERVIEW OF THE COAGULATION CASCADE AND ITS ROLE IN VENOUS THROMBOSIS

Coagulation is defined as the formation of fibrin by a series of linked enzymatic reactions in which each reaction product converts the subsequent inactive zymogen into an active serine protease **(Fig. 122-2)**. This coordinated sequence is called the coagulation cascade and is a key mechanism for regulating hemostasis. Central to the function of the coagulation cascade is the principle of amplification: owing to a series of linked enzymatic reactions, a small stimulus can lead to much greater quantities of fibrin, the end product that prevents hemorrhage at the site of vascular injury. In addition to the known risk factors relevant to hypercoagulopathy, stasis, and vascular dysfunction, newer areas of research have identified contributions from procoagulant microparticles, inflammatory cells, microvesicles, and fibrin structure.

The coagulation cascade is primarily initiated by vascular injury exposing tissue factor to blood components (Fig. 122-2). Tissue factor may also be found in bloodborne cell-derived microparticles and, under pathophysiologic conditions, in leukocytes or platelets. Plasma factor VII (FVII) is the ligand for and is activated (FVIIa) by binding to tissue factor exposed at the site of vessel damage. The binding of FVII/VIIa to tissue factor activates the downstream conversion of factor X (FX) to active FX (FXa). In an alternative reaction, the FVII/FVIIa–tissue factor complex initially converts FIX to FIXa, which then activates FX in conjunction with its cofactor factor VIII (FVIIIa). Thrombin can also activate factor XI (FXI) to active FXI (FXIa), which, in turn, can also activate FIX. FXa with its cofactor FVa converts prothrombin to thrombin, which then converts soluble plasma fibrinogen to insoluble fibrin, leading to clot or thrombus formation. Thrombin also activates FXIII to FXIIIa, a transglutaminase that covalently cross-links and stabilizes the fibrin clot. Formation of thrombi is affected by mechanisms governing fibrin structure and stability, including specific fibrinogen variants and how they alter fibrin formation, strength, and structure.

Several antithrombotic factors also regulate coagulation; these include antithrombin, tissue factor pathway inhibitor (TFPI), heparin cofactor II, and protein C/protein S. Under normal conditions, these factors limit the production of thrombin to prevent the perpetuation of coagulation and thrombus formation. Typically, after the clot has caused occlusion at the damage site and begins to expand toward adjacent uninjured vessel segments, the anticoagulant reactions governed by the normal endothelium become pivotal in limiting the extent of this hemostatically protective clot.

RISK FACTORS FOR VENOUS THROMBOSIS

An array of different factors contributes to the risk of VTE; it is notable that women and men of all ages, races, and ethnicities are at risk for VTE. The risk factors for venous thrombosis are primarily related to hypercoagulability, which can be genetic (Table 122-1) or acquired, or due to immobilization and venous stasis. Independent predictors for recurrence include increasing age, obesity, malignant neoplasm, and acute extremity paresis. It is estimated that 5–8% of the U.S. population has a genetic risk factor known to predispose to venous thrombosis. Often, multiple risk factors are present in a single individual. Significant risk is incurred by major orthopedic, abdominal, or neurologic surgeries. Cancer patients have an approximately fourfold increased risk of VTE as compared with the general population, and cancer patients with VTE have reduced survival. Hospitalized patients have a greatly increased risk of venous thrombosis with risk factors (increased age, male, ethnicity) and comorbid conditions, including infection, renal disease, and weight loss. Community- or hospital-acquired infection is also associated with increased risk of VTE. Supportive of this risk, nearly 20% of hospitalized COVID-19 patients were noted to have coagulation abnormalities as well as increased PE, DVT, and peripheral thrombotic risk. Moderate risk is promoted by prolonged bedrest; certain types of cancer; pregnancy; hormone replacement therapy or oral contraceptive use; and other sedentary conditions, such as long-distance plane travel. It has been reported that the risk of developing a VTE event doubles after air travel lasting 4 h, although the absolute

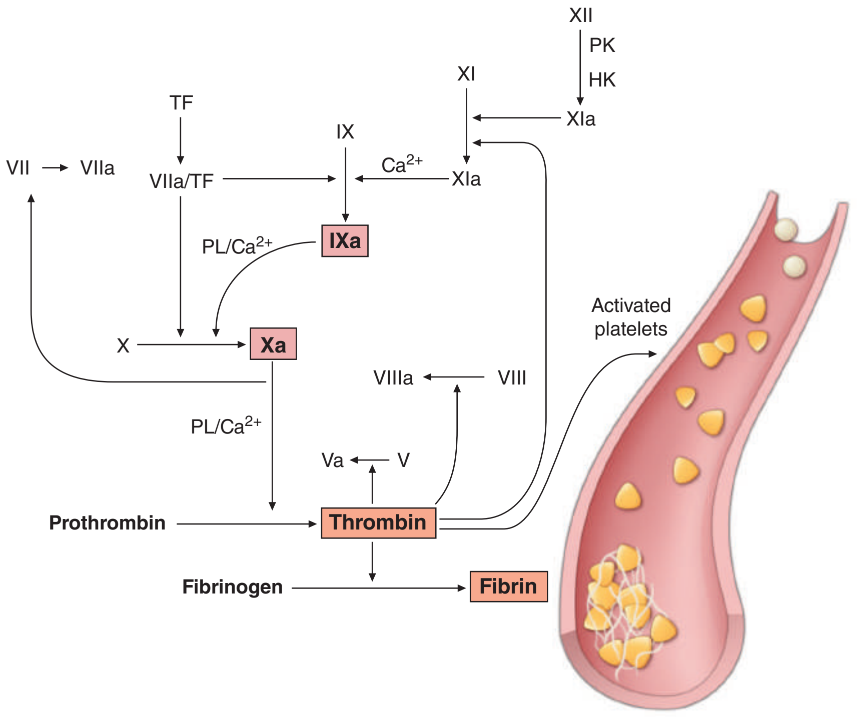

FIGURE 122-2 Summary of the coagulation pathways. Specific coagulation factors ("a" indicates activated form) are responsible for the conversion of soluble plasma fibrinogen into insoluble fibrin. This process occurs via a series of linked reactions in which the enzymatically active product subsequently converts the downstream inactive protein into an active serine protease. In addition, the activation of thrombin leads to stimulation of platelets. HK, high-molecular-weight kininogen; PK, prekallikrein; TF, tissue factor.

risk remains low (1 in 6000). The relative risk of VTE among pregnant or postpartum women is 4.3, and the overall incidence (absolute risk) is 199.7 per 100,000 woman-years.

■ GENETICS OF VENOUS THROMBOSIS

(See Table 122-2) Less common causes of venous thrombosis are those due to genetic variants. These abnormalities include loss-of-function mutations of endogenous anticoagulants as well as gain-of-function mutations of procoagulant proteins. Heterozygous antithrombin deficiency and homozygosity of the factor V Leiden mutation significantly increase the risk of venous thrombosis. While homozygous protein C or protein S deficiencies are rare and may lead to fatal purpura fulminans, heterozygous deficiencies are associated with a moderate risk of thrombosis. Activated protein C impairs coagulation by proteolytic degradation of FVa. Patients resistant to the activity of activated protein C may have a point mutation in the FV gene, a mutant denoted factor V Leiden. Mildly increased risk has been attributed to elevated levels of procoagulant factors, as well as low levels of tissue factor pathway inhibitor. Polymorphisms of methylene tetrahydrofolate reductase as well as hyperhomocysteinemia have been shown to be independent risk factors for venous thrombosis, as well as arterial vascular disease; however, many of the initial descriptions of genetic variants and their associations with thromboembolism are being questioned in larger, more contemporary studies.

■ FIBRINOLYSIS AND THROMBOSIS

Specific abnormalities in the fibrinolytic system have been associated with enhanced thrombosis. Factors such as elevated levels of tissue plasminogen activator (tPA) and plasminogen activator inhibitor type 1 (PAI-1) have been associated with decreased fibrinolytic activity and an increased risk of arterial thrombotic disease. Specific genetic variants have been associated with decreased fibrinolytic activity, including the 4G/5G insertion/deletion polymorphism in the (plasminogen activator type 1) *PAI-1* gene. Additionally, the 311-bp Alu insertion/deletion in tPA's intron 8 has been associated with enhanced thrombosis; however, genetic abnormalities have not been associated consistently with altered function or tPA levels, raising questions about the relevant pathophysiologic mechanism. Thrombin-activatable fibrinolysis inhibitor (TAFI) is a carboxypeptidase that regulates fibrinolysis; elevated plasma TAFI levels have been associated with an increased risk of both DVT and cardiovascular disease.

The metabolic syndrome also is accompanied by altered fibrinolytic activity. This syndrome, which comprises abdominal fat (central obesity), altered glucose and insulin metabolism, dyslipidemia, and hypertension, has been associated with atherothrombosis. The mechanism for enhanced thrombosis appears to be due both to altered platelet function and to a procoagulant and hypofibrinolytic state. One of the most frequently documented prothrombotic abnormalities reported in this syndrome is an increase in plasma levels of PAI-1.

In addition to contributing to platelet function, inflammation plays a role in both coagulation-dependent thrombus formation and thrombus resolution. Both polymorphonuclear neutrophils and monocytes/macrophages contribute to multiple overlapping thrombotic functions, including fibrinolysis, chemokine and cytokine production, and phagocytosis.

THE DISTINCTION BETWEEN ARTERIAL AND VENOUS THROMBOSIS

Although there is overlap, venous thrombosis and arterial thrombosis are initiated differently, and clot formation progresses by somewhat distinct pathways. In the setting of stasis or states of hypercoagulability, venous thrombosis is activated with the initiation of the coagulation cascade primarily due to exposure of tissue factor; this leads to the formation of thrombin and the subsequent conversion of fibrinogen to fibrin. In the artery, thrombin formation also occurs, but thrombosis is primarily promoted by the adhesion of platelets to an injured vessel and stimulated by exposed extracellular matrix (Figs. 122-1 and 122-2). There is wide variation in individual responses to vascular injury, an important determinant of which is the predisposition an individual has to arterial or venous thrombosis. This concept has been supported indirectly in prothrombotic animal models in which there is poor correlation between the propensity to develop venous versus arterial thrombosis.

Despite considerable progress in elucidating the role of hypercoagulable states in VTE, the contribution of hypercoagulability to arterial vascular disease is much less well understood. Although specific thrombophilic conditions, such as factor V Leiden and the prothrombin G20210A mutation, are risk factors for DVT, PE, and other VTE events, their contribution to arterial thrombosis is less well defined. In fact, to the contrary, many of these thrombophilic factors have not been found to be clinically important risk factors for arterial thrombotic events, such as acute coronary syndromes.

Clinically, although the pathophysiology is distinct, arterial and venous thrombosis do share common risk factors, including age, obesity, cigarette smoking, diabetes mellitus, arterial hypertension, hyperlipidemia, and metabolic syndrome. Select genetic variants, including those of the glutathione peroxidase-3 (*GPx3*) gene, have also been associated with arterial and venous thrombo-occlusive disease. Importantly, arterial and venous thrombosis may both be triggered by pathophysiologic stimuli responsible for activating inflammatory and oxidative pathways.

The diagnosis and treatment of ischemic heart disease are discussed in Chap. 284. Stroke diagnosis and management are discussed in Chap. 318. The diagnosis and management of DVT and PE are discussed in Chap. 290.

■ FURTHER READING

Ackermann M et al: Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in Covid-19. N Engl J Med 383:120, 2020.

Asmis L, Hellstern P: Thrombophilia testing: A systematic review. Clin Lab 69, 2023.

Becattini C, Aegnelli G: Acute treatment of venous thromboembolism. Blood 135:305, 2020.

Engelmann B, Massberg S: Thrombosis as an intravascular effector of innate immunity. Nat Rev Immunol 13:34, 2013.

Furie B, Furie BC: Mechanisms of thrombus formation. N Engl J Med 359:938, 2008.

Kaiser R et al: Hemostasis without clot formation: How platelets guard the vasculature in inflammation, infection, and malignancy. Blood 142;1413, 2023.

Koupenova M et al: Thrombosis and platelets: An update. Eur Heart J 38:785, 2017.

123

Antiplatelet, Anticoagulant, and Fibrinolytic Drugs

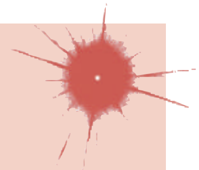

Jeffrey I. Weitz

Thromboembolic disorders are major causes of morbidity and mortality. Thrombosis can occur in arteries or veins. Arterial thrombosis is the most common cause of acute myocardial infarction (MI), ischemic stroke, and limb gangrene. VTE encompasses DVT, which can lead to postthrombotic syndrome, and PE, which can be fatal or can result in chronic thromboembolic pulmonary hypertension.

Most arterial thrombi are superimposed on disrupted atherosclerotic plaque because plaque rupture exposes thrombogenic material in the core to the blood. This material then triggers platelet aggregation and fibrin formation, which results in the generation of a platelet-rich thrombus that can temporarily or permanently occlude blood flow. In contrast, venous thrombi rarely form at sites of obvious vascular

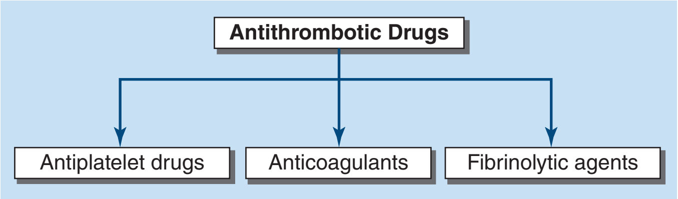

FIGURE 123-1 Classification of antithrombotic drugs.

disruption. Although they can develop after surgical trauma to veins or secondary to indwelling venous catheters, venous thrombi usually originate in the valve cusps of the deep veins of the calf or in the muscular sinuses. Sluggish blood flow reduces the oxygen supply to the avascular valve cusps. Endothelial cells lining these valve cusps become activated and express adhesion molecules on their surface. Tissue factor–bearing leukocytes and microvesicles adhere to these activated cells and induce coagulation. DNA extruded from neutrophils forms neutrophil extracellular traps (NETs) that provide a scaffold that binds platelets, promotes their activation and aggregation, and activates factor XII. Local thrombus formation is exacerbated by reduced clearance of activated clotting factors because of impaired blood flow. If the thrombi extend from the calf veins into the popliteal and more proximal veins of the leg, thrombus fragments can dislodge, travel to the lungs, and produce PE.

Arterial and venous thrombi are composed of platelets, fibrin, and trapped red blood cells, but the proportions differ. Arterial thrombi are rich in platelets because of the high shear in the injured arteries. In contrast, venous thrombi, which form under low shear conditions, contain relatively few platelets and are predominantly composed of fibrin and trapped red cells. Because of the predominance of platelets, arterial thrombi appear white, whereas venous thrombi are red in color, reflecting the trapped red cells.

Antithrombotic drugs are used for the prevention and treatment of thrombosis. Targeting the components of thrombi, these agents include (1) antiplatelet drugs, (2) anticoagulants, and (3) fibrinolytic agents **(Fig. 123-1)**. With the predominance of platelets in arterial thrombi, strategies to attenuate arterial thrombosis focus mainly on antiplatelet agents, although, in the acute setting, they may include anticoagulants and fibrinolytic agents. The addition of low-dose rivaroxaban, an oral factor Xa inhibitor, to dual-antiplatelet therapy reduces recurrent ischemic events and stent thrombosis in patients with acute coronary syndrome, whereas its addition to aspirin reduces the risk of major adverse coronary and limb events in patients with stable coronary or peripheral artery disease. These findings highlight the utility of combining low-dose anticoagulants with antiplatelet agents for secondary prevention in patients at high risk for recurrent atherothrombotic events.

Anticoagulants are the mainstay of prevention and treatment of VTE because fibrin is the predominant component of venous thrombi. Antiplatelet drugs are less effective than anticoagulants in this setting because of the limited platelet content of venous thrombi. Fibrinolytic therapy is used in selected patients with VTE. For example, patients with massive PE can benefit from systemic or catheter-directed fibrinolytic therapy. Pharmaco-mechanical therapy is also used to restore blood flow in patients with extensive DVT involving the iliac and/or femoral veins.

ANTIPLATELET DRUGS

■ ROLE OF PLATELETS IN ARTERIAL THROMBOSIS

In healthy vasculature, circulating platelets are maintained in an inactive state by nitric oxide (NO) and prostacyclin released by endothelial cells lining the blood vessels. In addition, endothelial cells also express CD39 on their surface, a membrane-associated ecto-adenosine diphosphatase (ADPase) that degrades ADP released from activated platelets. When the vessel wall is damaged, the release of these substances is impaired and the subendothelial matrix is exposed. Platelets adhere to exposed collagen via $\alpha_2\beta_1$ and glycoprotein (Gp) V1 and to von Willebrand factor (VWF) via Gp Ibα—receptors that are constitutively expressed on the platelet surface. Adherent platelets undergo a change in shape, secrete ADP from their dense granules, and synthesize and release thromboxane A2. Released ADP and thromboxane A_2, which are platelet agonists, activate ambient platelets and recruit them to the site of vascular injury **(Fig. 123-2)**.

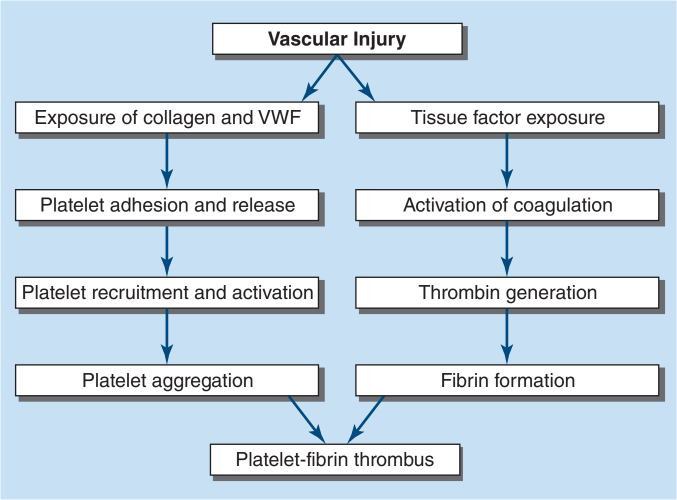
Vascular Injury
Exposure of collagen and VWF
Tissue factor exposure
Platelet adhesion and release
Activation of coagulation
Platelet recruitment and activation
Thrombin generation
Platelet aggregation
Fibrin formation
Platelet-fibrin thrombus

FIGURE 123-2 Coordinated role of platelets and the coagulation system in thrombogenesis. Vascular injury simultaneously triggers platelet activation and aggregation and activation of the coagulation system. Platelet activation is initiated by exposure of subendothelial collagen and von Willebrand factor (VWF), onto which platelets adhere. Adherent platelets become activated and release ADP and thromboxane A_2, platelet agonists that activate ambient platelets and recruit them to the site of injury. When platelets are activated, glycoprotein IIb/IIIa on their surface undergoes a conformational change that enables it to ligate fibrinogen and/or VWF and mediate platelet aggregation. Coagulation is triggered by tissue factor exposed at the site of injury. Tissue factor triggers thrombin generation. As a potent platelet agonist, thrombin amplifies platelet recruitment to the site of injury. Thrombin also converts fibrinogen to fibrin, and the fibrin strands then weave the platelet aggregates together to form a platelet/fibrin thrombus.

Disruption of the vessel wall also exposes tissue factor–expressing cells to the blood. Tissue factor binds to factor VII and induces its activation, and the tissue factor–factor VIIa complex then initiates coagulation. Activated platelets potentiate coagulation by providing a negatively charged surface that binds clotting factors and supports the assembly of activation complexes that enhance thrombin generation. In addition to converting fibrinogen to fibrin, thrombin serves as a potent platelet agonist and recruits more platelets to the site of vascular injury. Thrombin also amplifies its generation by feedback activation of factors V, VIII, and XI and solidifies the fibrin network by activating factor XIII, which then cross-links the fibrin strands and renders them more resistant to degradation.

When platelets are activated, Gp IIb/IIIa, the most abundant receptor on the platelet surface, undergoes a conformational change that enables it to bind fibrinogen and, under high shear conditions, VWF. Divalent fibrinogen or multivalent VWF molecules bridge adjacent platelets together to form platelet aggregates. Fibrin strands, generated through the action of thrombin, then weave these aggregates together to form a platelet-rich fibrin clot.

Antiplatelet drugs target various steps in this process. The commonly used antiplatelet drugs include aspirin, ADP receptor inhibitors, which include thienopyridines (clopidogrel and prasugrel) as well as ticagrelor and cangrelor, dipyridamole, Gp IIb/IIIa antagonists, and vorapaxar.

■ ASPIRIN

The most widely used antiplatelet agent worldwide is aspirin. As a cheap and effective antiplatelet drug, aspirin serves as the foundation of most antiplatelet strategies.

Mechanism of Action Aspirin produces its antithrombotic effect by irreversibly acetylating and inhibiting platelet cyclooxygenase

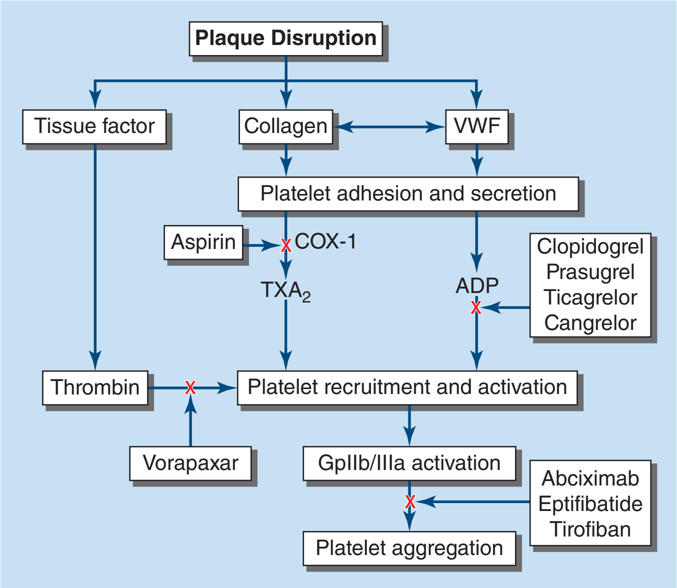

FIGURE 123-3 Site of action of antiplatelet drugs. Aspirin inhibits thromboxane A$_2$ (TXA$_2$) synthesis by irreversibly acetylating cyclooxygenase-1 (COX-1). Reduced TXA$_2$ release attenuates platelet activation and recruitment to the site of vascular injury. Clopidogrel and prasugrel irreversibly block P2Y$_{12}$, a key ADP receptor on the platelet surface; cangrelor and ticagrelor are reversible inhibitors of P2Y$_{12}$. Abciximab, eptifibatide, and tirofiban inhibit the final common pathway of platelet aggregation by blocking fibrinogen and von Willebrand factor (VWF) binding to activated glycoprotein (Gp) IIb/IIIa. Vorapaxar inhibits thrombin-mediated platelet activation by targeting protease-activated receptor-1 (PAR-1), the major thrombin receptor on human platelets.

(COX)-1 **(Fig. 123-3)**, a critical enzyme in the biosynthesis of thromboxane A$_2$. At high doses (~1 g/d), aspirin also inhibits COX-2, an inducible COX isoform found in endothelial cells and inflammatory cells. In endothelial cells, COX-2 initiates the synthesis of prostacyclin, a potent vasodilator and inhibitor of platelet aggregation.

Indications Aspirin is widely used for secondary prevention of cardiovascular events in patients with established coronary artery, cerebral artery, or peripheral artery disease. Compared with placebo in this setting, aspirin produces a 25% reduction in the risk of cardiovascular death, MI, or stroke. Aspirin is no longer used routinely for primary prevention because recent studies suggest that the risk of gastrointestinal and intracerebral hemorrhage outweigh the benefits. Consequently, aspirin is only recommended for primary cardiac prevention if the baseline cardiovascular risk is at least 1% per year over 10 years and patients are at low risk for bleeding.

Dosages Aspirin is usually administered at doses of 75–325 mg once daily. Higher doses of aspirin are not more effective than lower aspirin doses, and some analyses suggest reduced efficacy with higher doses. Because the side effects of aspirin are dose-related, daily aspirin doses of 75–100 mg are recommended for most indications. When rapid platelet inhibition is required, an initial aspirin dose of at least 160 mg should be given.

Side Effects The most common side effects are gastrointestinal and range from dyspepsia to erosive gastritis or peptic ulcers with bleeding and perforation. These side effects are dose related. Use of enteric-coated or buffered aspirin in place of plain aspirin does not eliminate gastrointestinal side effects. The overall risk of major bleeding with aspirin is 1–3% per year. The risk of bleeding is increased two- to threefold when aspirin is given in conjunction with other antiplatelet drugs, such as clopidogrel or ticagrelor, or with anticoagulants, such as warfarin. When dual or triple therapy is prescribed, low-dose aspirin should be given (75–100 mg daily). Eradication of *Helicobacter pylori* infection and administration of proton pump inhibitors may reduce the risk of aspirin-induced upper gastrointestinal bleeding in patients with peptic ulcer disease.

Aspirin should not be administered to patients with a history of aspirin allergy characterized by bronchospasm. This problem occurs in ~0.3% of the general population but is more common in those with chronic urticaria or asthma, particularly in individuals with nasal polyps or chronic rhinitis. Hepatic and renal toxicity are observed with aspirin overdose.

Aspirin Resistance Clinical aspirin resistance is defined as the failure of aspirin to protect patients from ischemic vascular events. This is not a helpful definition because it is made after the event occurs. Furthermore, it is not realistic to expect aspirin, which only blocks thromboxane A$_2$–induced platelet activation, to prevent all vascular events.

Aspirin resistance has also been described biochemically as a failure of the drug to produce its expected inhibitory effects on tests of platelet function, such as thromboxane A$_2$ synthesis or arachidonic acid–induced platelet aggregation. Potential causes of aspirin resistance include poor compliance, reduced absorption, drug-drug interaction with ibuprofen, and overexpression of COX-2. Unfortunately, the tests for aspirin resistance have not been well standardized, and there is little evidence that they identify patients at increased risk of recurrent vascular events, or that resistance can be reversed by giving higher doses of aspirin or by adding other antiplatelet drugs. Until such information is available, testing for aspirin resistance remains a research tool.

■ ADP RECEPTOR ANTAGONISTS

The ADP receptor antagonists include the thienopyridines (clopidogrel and prasugrel) as well as ticagrelor and cangrelor. All these drugs target P2Y$_{12}$, the key ADP receptor on platelets.

Thienopyridines • **MECHANISM OF ACTION** The thienopyridines are structurally related drugs that selectively inhibit ADP-induced platelet aggregation by irreversibly blocking P2Y$_{12}$ (Fig. 123-3). Clopidogrel and prasugrel are prodrugs that require metabolic activation by the hepatic cytochrome P450 (CYP) enzyme system. Prasugrel is about 10-fold more potent than clopidogrel and has a more rapid onset of action because of better absorption and more streamlined metabolic activation.

INDICATIONS When compared with aspirin in patients with recent ischemic stroke, recent MI, or a history of peripheral arterial disease, clopidogrel reduced the risk of cardiovascular death, MI, and stroke by 8.7%. Therefore, clopidogrel is more effective than aspirin but is also more expensive. Clopidogrel and aspirin are often combined to capitalize on their capacity to block complementary pathways of platelet activation. For example, the combination of aspirin plus clopidogrel is recommended for at least 4 weeks after implantation of a bare metal stent in a coronary artery and at least a year in those with a drug-eluting stent. Concerns about late in-stent thrombosis with drug-eluting stents have led some experts to recommend long-term use of clopidogrel plus aspirin for the latter indication. The course of dual-antiplatelet therapy is often shortened in patients at high risk of bleeding with the dropping of aspirin and continuation of the P2Y$_{12}$ inhibitor.

The combination of clopidogrel and aspirin is effective in patients with non-ST-segment elevation MI. Thus, in 12,562 such patients, the risk of cardiovascular death, MI, or stroke was 9.3% in those randomized to the combination of clopidogrel and aspirin and 11.4% in those given aspirin alone. This 20% relative risk reduction with combination therapy was highly statistically significant. However, combining clopidogrel with aspirin increases the risk of major bleeding to about 2% per year. This bleeding risk persists even if the daily dose of aspirin is ≤100 mg. Therefore, the combination of clopidogrel and aspirin should only be used when there is a clear benefit. For example, this combination was not proven to be superior to clopidogrel alone in patients with acute ischemic stroke or to aspirin alone for primary prevention in those at risk for cardiovascular events.

Prasugrel was compared with clopidogrel in 13,608 patients with acute coronary syndrome who were scheduled to undergo percutaneous coronary intervention. The incidence of the primary efficacy endpoint, a composite of cardiovascular death, MI, or stroke, was significantly lower with prasugrel than with clopidogrel (9.9% and 12.1%, respectively), mainly reflecting a reduction in nonfatal MI. The incidence of stent thrombosis also was significantly lower with prasugrel than with clopidogrel (1.1% and 2.4%, respectively). However, these advantages were at the expense of significantly higher rates of fatal bleeding (0.4% and 0.1%, respectively) and life-threatening bleeding (1.4% and 0.9%, respectively) with prasugrel. Because patients older than 75 years of age and those with a history of prior stroke or transient ischemic attack are at high risk of bleeding, prasugrel should generally be avoided in older patients, and the drug is contraindicated in those with a history of cerebrovascular disease. Caution is required if prasugrel is used in patients weighing <60 kg or in those with renal impairment.

When prasugrel was compared with clopidogrel in 7243 patients with unstable angina or MI without ST-segment elevation, prasugrel failed to reduce the rate of the primary efficacy endpoint, which was a composite of cardiovascular death, MI, and stroke. Because of the negative results of this study, prasugrel is reserved for patients undergoing percutaneous coronary intervention. In this setting, prasugrel is usually given in conjunction with aspirin. To reduce the risk of bleeding, the daily aspirin dose should be ≤100 mg.

For patients with noncardioembolic stroke or high-risk transient ischemic attack, the combination of clopidogrel or ticagrelor plus aspirin for 21–30 days followed by aspirin alone thereafter reduces the risk of stroke, MI, and vascular death by up to 30% compared with aspirin alone. Therefore, dual antiplatelet therapy is often administered for the first 3–4 weeks in such patients.

DOSING Clopidogrel is given once daily at a dose of 75 mg. Loading doses of clopidogrel are given when rapid ADP receptor blockade is desired. For example, patients undergoing coronary stenting are often given a loading dose of 300–600 mg, which produces inhibition of ADP-induced platelet aggregation in about 4–6 h. After a loading dose of 60 mg, prasugrel is given once daily at a dose of 10 mg. Patients older than age 75 years or weighing <60 kg should receive a lower daily prasugrel dose of 5 mg.

SIDE EFFECTS The most common side effect of clopidogrel and prasugrel is bleeding. Because of its greater potency, bleeding is more common with prasugrel than with clopidogrel. To reduce the risk of bleeding, clopidogrel and prasugrel should be stopped 5–7 days before major surgery. In patients taking clopidogrel or prasugrel who present with serious bleeding, platelet transfusion may be helpful.

Hematologic side effects, including neutropenia, thrombocytopenia, and thrombotic thrombocytopenic purpura, are rare.

THIENOPYRIDINE RESISTANCE The capacity of clopidogrel to inhibit ADP-induced platelet aggregation varies among subjects. This variability reflects, at least in part, genetic polymorphisms in the CYP isoenzymes involved in the metabolic activation of clopidogrel. The most important of these is CYP2C19. Clopidogrel-treated patients who are homozygous for the loss-of-function *CYP2C192* allele exhibit reduced platelet inhibition compared with those with the wild-type *CYP2C19* allele and experience a higher rate of cardiovascular events. This is important because estimates suggest that up to 25% of whites, 30% of African Americans, and 50% of Asians carry the loss-of-function allele, which would render them resistant to clopidogrel. Even patients with the reduced-function *CYP2C19* alleles may derive less benefit from clopidogrel than those with the full-function *CYP2C19* allele. Concomitant administration of clopidogrel with proton pump inhibitors, which are inhibitors of *CYP2C19*, produces a small reduction in the inhibitory effects of clopidogrel on ADP-induced platelet aggregation. The extent to which this interaction increases the risk of cardiovascular events remains controversial.

In contrast to their effect on the metabolic activation of clopidogrel, *CYP2C19* polymorphisms appear to be less important determinants of the activation of prasugrel. Thus, no association was detected between the loss-of-function allele and decreased platelet inhibition or increased rate of cardiovascular events with prasugrel. The observation that genetic polymorphisms affecting clopidogrel absorption or metabolism influence clinical outcomes raised the possibility that pharmacogenetic profiling might be useful for identifying clopidogrel-resistant patients and that point-of-care assessment of the extent of clopidogrel-induced platelet inhibition may help detect patients at higher risk for subsequent cardiovascular events. Clinical trials designed to evaluate these possibilities have thus far been negative. Although administration of higher doses of clopidogrel can overcome a reduced response to clopidogrel, the clinical benefit of this approach is uncertain. Instead, prasugrel or ticagrelor may be better choices for these patients.

Ticagrelor As an orally active inhibitor of $P2Y_{12}$, ticagrelor differs from the thienopyridines in that ticagrelor does not require metabolic activation and it produces reversible inhibition of the ADP receptor.

MECHANISM OF ACTION Like the thienopyridines, ticagrelor inhibits $P2Y_{12}$. Because it does not require metabolic activation, ticagrelor has a more rapid onset and offset of action than clopidogrel, and it produces greater and more predictable inhibition of ADP-induced platelet aggregation than clopidogrel.

INDICATIONS Ticagrelor is indicated for the secondary prevention of atherothrombotic events in patients with an acute coronary syndrome treated medically or with percutaneous coronary intervention (PCI) with or without stent implantation or with coronary artery bypass graft (CABG) surgery. Ticagrelor is also indicated for up to 3 years for secondary prevention in patients with a prior history of MI at least 1 year ago who are at high risk for atherothrombotic events. For patients with acute coronary syndrome undergoing PCI, guidelines give preference to ticagrelor over clopidogrel, particularly in higher risk patients.

DOSING Ticagrelor is initiated with an oral loading dose of 180 mg followed by 90 mg twice daily. The dose does not require adjustment in patients with renal impairment, but the drug should be used with caution in patients with hepatic disease and in those receiving potent inhibitors or inducers of *CYP3A4* because ticagrelor is metabolized in the liver via *CYP3A4*. Ticagrelor is usually administered in conjunction with aspirin; the daily aspirin dose should not exceed 100 mg.

SIDE EFFECTS In addition to bleeding, the most common side effects of ticagrelor are dyspnea, which can occur in up to 15% of patients, and asymptomatic ventricular pauses. The dyspnea, which tends to occur soon after initiating ticagrelor, is usually self-limiting and mild in intensity. The mechanism responsible for this side effect is unknown.

To reduce the risk of bleeding, ticagrelor should be stopped at least 5 days before major surgery. Platelet transfusion is unlikely to be of benefit in patients with ticagrelor-related bleeding or those requiring urgent surgery because the drug will bind to $P2Y_{12}$ on the transfused platelets. Bentracimab, an antibody fragment that binds ticagrelor and its metabolite with high affinity and rapidly reverses their platelet inhibitory effects, is under development for ticagrelor reversal before urgent surgery or intervention or for patients with serious bleeding.

Cangrelor Cangrelor is a rapidly acting reversible inhibitor of $P2Y_{12}$ that is administered intravenously. It has an immediate onset of action, a half-life of 3–5 min, and an offset of action within an hour. Cangrelor is licensed for use in patients undergoing PCI and produces rapid ADP receptor blockade in those who have not received pretreatment with clopidogrel, prasugrel, or ticagrelor.

Cangrelor is administered as a 30 μg/kg IV bolus before PCI followed by an infusion of 4 μg/kg per minute for at least 2 h or for the duration of the procedure, whichever is longer. When transitioning to oral $P2Y_{12}$ inhibitor therapy, ticagrelor can be given at a loading dose of 180 mg at any time during the cangrelor infusion or immediately after discontinuation. In contrast, loading doses of prasugrel or clopidogrel (60 and 600 mg, respectively) should only be given after cangrelor is stopped because cangrelor blocks the interaction of their active metabolites with $P2Y_{12}$.

DIPYRIDAMOLE

Dipyridamole is a relatively weak antiplatelet agent on its own, but an extended-release formulation of dipyridamole combined with low-dose aspirin, a preparation known as Aggrenox, is sometimes used for secondary prevention in patients with transient ischemic attacks or ischemic stroke.

MECHANISM OF ACTION By inhibiting phosphodiesterase, dipyridamole blocks the breakdown of cyclic adenosine monophosphate (AMP). Increased levels of cyclic AMP reduce intracellular calcium and inhibit platelet activation. Dipyridamole also blocks the uptake of adenosine by platelets and other cells. This produces a further increase in local cyclic AMP levels because the platelet adenosine A_2 receptor is coupled to adenylate cyclase (Fig. 123-4).

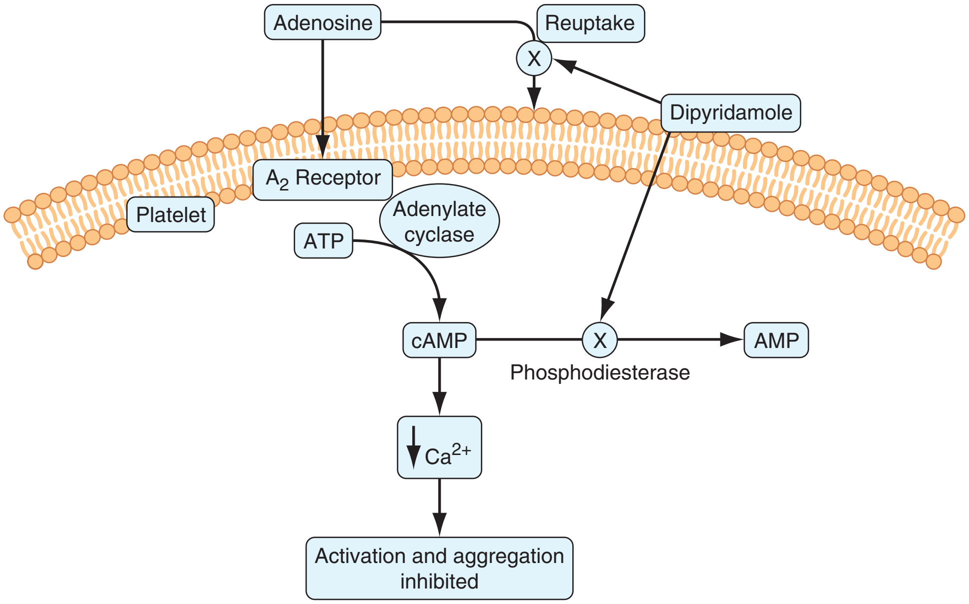

FIGURE 123-4 Mechanism of action of dipyridamole. Dipyridamole increases levels of cyclic AMP (cAMP) in platelets by (1) blocking the reuptake of adenosine and (2) inhibiting phosphodiesterase-mediated cyclic AMP degradation. By promoting calcium uptake, cyclic AMP reduces intracellular levels of calcium. This, in turn, inhibits platelet activation and aggregation.

INDICATIONS Dipyridamole plus aspirin was compared with aspirin or dipyridamole alone, or with placebo, in patients with an ischemic stroke or transient ischemic attack. The combination reduced the risk of stroke by 22.1% compared with aspirin and by 24.4% compared with dipyridamole. A second trial compared dipyridamole plus aspirin with aspirin alone for secondary prevention in patients with ischemic stroke. Vascular death, stroke, or MI occurred in 13% of patients given combination therapy and in 16% of those treated with aspirin alone. Another trial randomized 20,332 patients with noncardioembolic ischemic stroke to either Aggrenox or clopidogrel. The primary efficacy endpoint of recurrent stroke occurred in 9.0% of those given Aggrenox and in 8.8% of patients treated with clopidogrel. Although this difference was not statistically significant, the study failed to meet the prespecified margin to claim noninferiority of Aggrenox relative to clopidogrel. These results have dampened enthusiasm for the use of Aggrenox.

Because of its vasodilatory effects and the paucity of data supporting the use of dipyridamole in patients with symptomatic coronary artery disease, Aggrenox should not be used for stroke prevention in such patients. Clopidogrel is a better choice in this setting.

DOSING Aggrenox is given twice daily. Each capsule contains 200 mg of extended-release dipyridamole and 25 mg of aspirin.

SIDE EFFECTS Because dipyridamole has vasodilatory effects, it must be used with caution in patients with coronary artery disease. Gastrointestinal complaints, headache, facial flushing, dizziness, and hypotension can also occur. These symptoms often subside with continued use of the drug.

GP IIB/IIIA RECEPTOR ANTAGONISTS

As a class, parenteral Gp IIb/IIIa receptor antagonists have a niche in patients with acute coronary syndrome. The three agents in this class are abciximab, eptifibatide, and tirofiban. However, abciximab is no longer available.

Mechanism of Action A member of the integrin family of adhesion receptors, Gp IIb/IIIa is found on the surface of platelets and megakaryocytes. With about 80,000 copies per platelet, Gp IIb/IIIa is the most abundant receptor. Consisting of a noncovalently linked heterodimer, Gp IIb/IIIa is inactive on resting platelets. When platelets are activated, inside-outside signal transduction pathways trigger a conformational activation of the receptor. Once activated, Gp IIb/IIIa binds adhesive molecules, such as fibrinogen and, under high shear conditions, VWF. Binding is mediated by the Arg-Gly-Asp (RGD) sequence found on the α chains of fibrinogen and on VWF, and by the Lys-Gly-Asp (KGD) sequence located within a unique dodecapeptide domain on the γ chains of fibrinogen. Once bound, fibrinogen and/or VWF bridge adjacent platelets together to induce platelet aggregation. Although abciximab, eptifibatide, and tirofiban all target the Gp IIb/IIIa receptor, they are structurally and pharmacologically distinct (Table 123-1). Abciximab is a Fab fragment of a humanized murine monoclonal antibody directed against the activated form of Gp IIb/IIIa. Abciximab binds to the activated receptor with high affinity and blocks the binding of adhesive molecules. In contrast, eptifibatide and tirofiban are synthetic small molecules. Eptifibatide is a cyclic heptapeptide that binds Gp IIb/IIIa because it incorporates the KGD motif, whereas tirofiban is a nonpeptidic tyrosine derivative that acts as an RGD mimetic. Abciximab has a long half-life and can be detected on the surface of platelets for up to 2 weeks; eptifibatide and tirofiban have short half-lives.

Indications Abciximab and eptifibatide are used in patients undergoing PCIs, particularly those who have not been pretreated with an ADP receptor antagonist. Tirofiban and eptifibatide are used in high-risk patients with non-ST-segment elevation MI.

Dosing All the Gp IIb/IIIa antagonists are given as an IV bolus followed by an infusion. The recommended dose of abciximab is a bolus of 0.25 mg/kg followed by an infusion of 0.125 μg/kg per minute to a maximum of 10 μg/kg for 12 h. In patients undergoing PCI, eptifibatide is given as two 180 μg/kg boluses given 10 min apart, followed by an infusion of 2.0 μg/kg per minute for 18–24 h. For patients with acute coronary syndrome, the second eptifibatide bolus is withheld. Tirofiban is started at a rate of 0.4 μg/kg per minute for 30 min; the drug is then continued at a rate of 0.1 μg/kg per minute for up to 18 h. Because eptifibatide and tirofiban are cleared by the kidneys, the doses must be reduced in patients with renal insufficiency. Thus, the eptifibatide infusion is reduced to 1 μg/kg per minute in patients with a creatinine clearance below 50 mL/min, whereas the dose of tirofiban is cut in half for patients with a creatinine clearance below 30 mL/min.

TABLE 123-1 Features of Gp IIb/IIIa Antagonists

FEATURE	ABCIXIMAB	EPTIFIBATIDE	TIROFIBAN
Description	Fab fragment of humanized mouse monoclonal antibody	Cyclical KGD-containing heptapeptide	Nonpeptidic RGD mimetic
Specific for Gp IIb/IIIa	No	Yes	Yes
Plasma half-life	Short (min)	Long (2.5 h)	Long (2.0 h)
Platelet-bound half-life	Long (days)	Short (s)	Short (s)
Renal clearance	No	Yes	Yes

Abbreviation: Gp, glycoprotein.

Side Effects In addition to bleeding, thrombocytopenia is the most serious complication. Thrombocytopenia is immune-mediated and is caused by antibodies directed against neoantigens on Gp IIb/IIIa that are exposed upon antagonist binding. With abciximab, thrombocytopenia occurs in up to 5% of patients. Thrombocytopenia is severe in ~1% of these individuals. Thrombocytopenia is less common with the other two agents, occurring in ~1% of patients.

VORAPAXAR

An orally active PAR-1 antagonist, vorapaxar blocks thrombin-induced platelet activation. Vorapaxar has a half-life of about 200 h.

Indications When compared with a placebo in 12,944 patients with acute coronary syndrome without ST-segment elevation, vorapaxar failed to significantly reduce the primary efficacy endpoint, a composite of cardiovascular death, MI, stroke, recurrent ischemia requiring rehospitalization, and urgent coronary revascularization. Moreover, vorapaxar was associated with increased rates of bleeding, including intracranial bleeding.

In a second trial, vorapaxar was compared with placebo for secondary prevention in 26,449 patients with prior MI, ischemic stroke, or peripheral arterial disease. Overall, vorapaxar reduced the risk of cardiovascular death, MI, or stroke by 13%, but doubled the risk of intracranial bleeding. In the prespecified subgroup of 17,779 patients with prior MI, however, vorapaxar reduced the risk for cardiovascular death, MI, or stroke by 20% compared with placebo (from 9.7% to 8.1%, respectively). The rate of intracranial hemorrhage was higher with vorapaxar than with placebo (0.6% and 0.4%, respectively; $p = .076$) as was the rate of moderate or severe bleeding (3.4% and 2.1%, respectively; $p < .0001$). Based on these data, vorapaxar is licensed for patients younger than 75 years of age with MI or peripheral artery disease who have no history of stroke, transient ischemic attack, or intracranial bleeding and weigh >60 kg.

Dosing Vorapaxar is given at a dose of 2.08 mg once daily.

Side Effects The major side effect is bleeding. Platelet transfusion may be of benefit for vorapaxar reversal.

ANTICOAGULANTS

There are both parenteral and oral anticoagulants. The parenteral anticoagulants include heparin, low-molecular-weight heparin (LMWH), fondaparinux (a synthetic pentasaccharide), lepirudin, desirudin, bivalirudin, and argatroban. Currently available oral anticoagulants include warfarin; dabigatran etexilate, an oral thrombin inhibitor; and rivaroxaban, apixaban, and edoxaban, which are oral factor Xa inhibitors.

PARENTERAL ANTICOAGULANTS

Heparin A sulfated polysaccharide, heparin is isolated from mammalian tissues rich in mast cells. Most commercial heparin is derived from porcine intestinal mucosa and is a polymer of alternating D-glucuronic acid and *N*-acetyl-D-glucosamine residues.

MECHANISM OF ACTION Heparin acts as an anticoagulant by activating antithrombin (previously known as antithrombin III) and accelerating the rate at which antithrombin inhibits clotting enzymes, particularly thrombin and factor Xa. Antithrombin, the obligatory plasma cofactor for heparin, is a member of the serine protease inhibitor (serpin) superfamily. Synthesized in the liver and circulating in plasma at a concentration of 2.6 ± 0.4 μM, antithrombin acts as a suicide substrate for its target enzymes.

To activate antithrombin, heparin binds to the serpin via a unique pentasaccharide sequence that is found on one-third of the chains of commercial heparin (Fig. 123-5). Heparin chains without this pentasaccharide sequence have little or no anticoagulant activity. Once bound to antithrombin, heparin induces a conformational change in the reactive center loop of antithrombin that renders it more readily accessible to its target proteases. This conformational change enhances the rate at which antithrombin inhibits factor Xa by at least two orders of magnitude but has little effect on the rate of thrombin inhibition. To catalyze thrombin inhibition, heparin serves as a template that binds antithrombin and thrombin simultaneously. Formation of this ternary complex brings the enzyme in close apposition to the inhibitor, thereby promoting the formation of a stable covalent thrombin-antithrombin complex.

Only pentasaccharide-containing heparin chains composed of at least 18 saccharide units (which correspond to a molecular weight of 5400) are of sufficient length to bridge thrombin and antithrombin

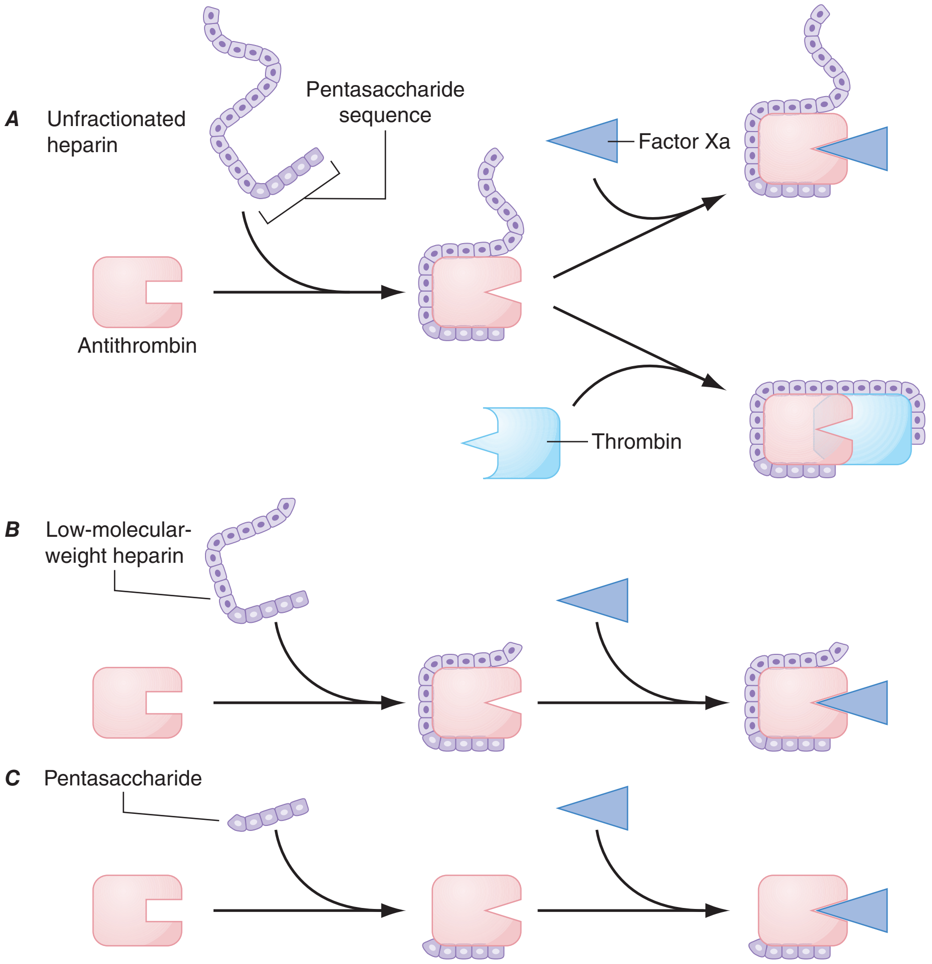

FIGURE 123-5 Mechanism of action of heparin, low-molecular-weight heparin (LMWH), and fondaparinux, a synthetic pentasaccharide. ***A.*** Heparin binds to antithrombin via its pentasaccharide sequence. This induces a conformational change in the reactive center loop of antithrombin that accelerates its interaction with factor Xa. To potentiate thrombin inhibition, heparin must simultaneously bind to antithrombin and thrombin. Only heparin chains composed of at least 18 saccharide units, which corresponds to a molecular weight of 5400, are of sufficient length to perform this bridging function. With a mean molecular weight of 15,000, all the heparin chains are long enough to do this. ***B.*** LMWH has greater capacity to potentiate factor Xa inhibition by antithrombin than thrombin because, with a mean molecular weight of 4500–5000, at least half of the LMWH chains are too short to bridge antithrombin to thrombin. ***C.*** Fondaparinux, a synthetic pentasaccharide, only accelerates factor Xa inhibition by antithrombin because it is too short to bridge antithrombin to thrombin.

together. With a mean molecular weight of 15,000, and a range of 5000–30,000, almost all the chains of unfractionated heparin are long enough to do so. Consequently, by definition, heparin has equal capacity to promote the inhibition of thrombin and factor Xa by antithrombin and is assigned an anti–factor Xa to anti–factor IIa (thrombin) ratio of 1:1.

Heparin causes the release of tissue factor pathway inhibitor (TFPI) from the endothelium. A factor Xa–dependent inhibitor of tissue factor–bound factor VIIa, TFPI may contribute to the antithrombotic activity of heparin. Longer heparin chains induce the release of more TFPI than shorter ones.

PHARMACOLOGY Heparin must be given parenterally. It is usually administered SC or by continuous IV infusion. When used for therapeutic purposes, the IV route is most often employed. If heparin is given SC for treatment of thrombosis, the dose of heparin must be high enough to overcome the limited bioavailability associated with this method of delivery.

In the circulation, heparin binds to the endothelium and to plasma proteins other than antithrombin. Heparin binding to endothelial cells explains its dose-dependent clearance. At low doses, the half-life of heparin is short because it binds rapidly to the endothelium. With higher doses of heparin, the half-life is longer because heparin is cleared more slowly once the endothelium is saturated. Clearance is mainly extrarenal; heparin binds to macrophages, which internalize and depolymerize the long heparin chains and secrete shorter chains back into the circulation. Because of its dose-dependent clearance mechanism, the plasma half-life of heparin ranges from 30 to 60 min with bolus IV doses of 25 and 100 units/kg, respectively.

Once heparin enters the circulation, it binds to plasma proteins other than antithrombin, a phenomenon that reduces its anticoagulant activity. Some of the heparin-binding proteins found in plasma are acute-phase reactants whose levels are elevated in ill patients. Others, such as high-molecular-weight multimers of VWF, are released from activated platelets or endothelial cells. Activated platelets also release platelet factor 4 (PF4), a highly cationic protein that binds heparin with high affinity. The large amounts of PF4 found in the vicinity of platelet-rich arterial thrombi can neutralize the anticoagulant activity of heparin. This phenomenon may attenuate heparin's capacity to suppress thrombus growth.

Because the levels of heparin-binding proteins in plasma vary from person to person, the anticoagulant response to fixed or weight-adjusted doses of heparin is unpredictable. Consequently, coagulation monitoring is essential to ensure that a therapeutic response is obtained. This is particularly important when heparin is administered for treatment of established thrombosis because a subtherapeutic anticoagulant response may render patients at risk for recurrent thrombosis, whereas excessive anticoagulation increases the risk of bleeding.

MONITORING THE ANTICOAGULANT EFFECT Heparin therapy can be monitored using the activated partial thromboplastin time (aPTT) or anti–factor Xa level. Although the aPTT is the test most often used for this purpose, there are problems with this assay. aPTT reagents vary in their sensitivity to heparin, and the type of coagulometer used for testing can influence the results. Consequently, laboratories must establish a therapeutic aPTT range with each reagent-coagulometer combination by measuring the aPTT and anti–factor Xa level in plasma samples collected from heparin-treated patients. For most of the aPTT reagents and coagulometers in current use, therapeutic heparin levels are achieved with a two- to threefold prolongation of the aPTT. Anti–factor Xa levels also can be used to monitor heparin therapy. With this test, therapeutic heparin levels range from 0.3 to 0.7 units/mL.

Up to 25% of heparin-treated patients with VTE require >35,000 units/d to achieve a therapeutic aPTT. These patients are considered heparin resistant. It is useful to measure anti–factor Xa levels in heparin-resistant patients because many will have a therapeutic anti–factor Xa level despite a subtherapeutic aPTT. This dissociation in test results occurs because elevated plasma levels of fibrinogen and factor VIII, both of which are acute-phase proteins, shorten the aPTT but have no effect on anti–factor Xa levels. Heparin therapy in patients who exhibit this phenomenon is best monitored using anti–factor Xa levels instead of the aPTT. Patients with congenital or acquired antithrombin deficiency and those with elevated levels of heparin-binding proteins may also need high doses of heparin to achieve a therapeutic aPTT or anti–factor Xa level. If there is good correlation between the aPTT and the anti–factor Xa level, either test can be used to monitor heparin therapy.

DOSING For prophylaxis, heparin is usually given in fixed doses of 5000 units SC two or three times daily. With these low doses, coagulation monitoring is unnecessary. In contrast, monitoring is essential when the drug is given in therapeutic doses. Fixed-dose or weight-based heparin nomograms are used to standardize heparin dosing and to shorten the time required to achieve a therapeutic anticoagulant response. At least two heparin nomograms have been validated in patients with VTE and reduce the time required to achieve a therapeutic aPTT. Weight-adjusted heparin nomograms have also been evaluated in patients with acute coronary syndromes. After an IV heparin bolus of 5000 units or 70 units/kg, a heparin infusion rate of 12–15 units/kg per hour is usually administered. In contrast, weight-adjusted heparin nomograms for patients with VTE use an initial bolus of 5000 units or 80 units/kg, followed by an infusion of 18 units/kg per hour. Thus, patients with VTE appear to require higher doses of heparin to achieve a therapeutic aPTT than do patients with acute coronary syndromes. This may reflect differences in the thrombus burden. Heparin binds to fibrin, and the amount of fibrin in patients with extensive DVT is greater than that in those with coronary thrombosis.

LIMITATIONS Heparin has pharmacokinetic and biophysical limitations (Table 123-2). The pharmacokinetic limitations reflect heparin's propensity to bind in a pentasaccharide-independent fashion to cells and plasma proteins. Heparin binding to endothelial cells explains its dose-dependent clearance, whereas binding to plasma proteins results in a variable anticoagulant response and can lead to heparin resistance.

The biophysical limitations of heparin reflect the inability of the heparin-antithrombin complex to inhibit factor Xa when it is incorporated into the prothrombinase complex, the complex that converts prothrombin to thrombin, and to inhibit thrombin bound to fibrin. Consequently, factor Xa bound to activated platelets within platelet-rich thrombi has the potential to generate thrombin, even in the face of heparin. Once this thrombin binds to fibrin, it too is protected from inhibition by the heparin-antithrombin complex. Clot-associated thrombin can then trigger thrombus growth by locally activating platelets and amplifying its own generation through feedback activation of factors V, VIII, and XI. Further compounding the problem is the potential for heparin neutralization by the high concentrations of PF4 released from activated platelets within the platelet-rich thrombus.

SIDE EFFECTS The most common side effect of heparin is bleeding. Other complications include thrombocytopenia, osteoporosis, and elevated levels of transaminases.

Bleeding The risk of bleeding rises as the dose of heparin is increased. Concomitant administration of drugs that affect hemostasis, such as antiplatelet or fibrinolytic agents, increases the risk of bleeding, as does

TABLE 123-2 Pharmacokinetic and Biophysical Limitations of Heparin

LIMITATIONS	MECHANISM
Poor bioavailability at low doses	Binds to endothelial cells and macrophages
Dose-dependent clearance	Binds to macrophages
Variable anticoagulant response	Binds to plasma proteins whose levels vary from patient to patient
Reduced activity in the vicinity of platelet-rich thrombi	Neutralized by platelet factor 4 released from activated platelets
Limited activity against factor Xa incorporated in the prothrombinase complex and thrombin bound to fibrin	Reduced capacity of heparin-antithrombin complex to inhibit factor Xa bound to activated platelets and thrombin bound to fibrin

TABLE 123-3 Features of Heparin-Induced Thrombocytopenia

FEATURES	DETAILS
Thrombocytopenia	Platelet count of ≤100,000/μL or a decrease in platelet count of ≥50%
Timing	Platelet count falls 5–14 days after starting heparin
Type of heparin	More common with unfractionated heparin than low-molecular-weight heparin
Type of patient	More common in surgical patients and patients with cancer than general medical patients; more common in women than in men
Thrombosis	Venous thrombosis more common than arterial thrombosis

recent surgery or trauma. Heparin-treated patients with serious bleeding can be given protamine sulfate to neutralize the heparin. Protamine sulfate, a mixture of basic polypeptides isolated from salmon sperm, binds heparin with high affinity, and the resultant protamine-heparin complexes are then cleared. Typically, 1 mg of protamine sulfate neutralizes 100 units of heparin. Protamine sulfate is given IV at a maximum amount of 50 mg per dose. Anaphylactoid reactions to protamine sulfate can occur, and drug administration by slow IV infusion is recommended to reduce the risk.

Thrombocytopenia Heparin can cause thrombocytopenia. Heparin-induced thrombocytopenia (HIT) is an antibody-mediated process that is triggered by antibodies directed against neoantigens on PF4 that are exposed when heparin binds to this protein. These antibodies, which are usually of the IgG isotype, bind simultaneously to the heparin-PF4 complex and to platelet Fc receptors. Such binding activates the platelets and generates platelet microparticles. Circulating microparticles are procoagulant because they express anionic phospholipids on their surface and can bind clotting factors and promote thrombin generation. The clinical features of HIT are illustrated in Table 123-3. Typically, HIT occurs 5–14 days after initiation of heparin therapy, but it can manifest earlier if the patient has received heparin within the past 3 months. A platelet count <100,000/μL or a 50% decrease in the platelet count from the pretreatment value should raise the suspicion of HIT. HIT is more common in surgical patients than in medical patients and, like many autoimmune disorders, occurs more frequently in females than in males.

HIT can be associated with thrombosis, either arterial or venous. Venous thrombosis, which manifests as DVT and/or PE, is more common than arterial thrombosis. Arterial thrombosis can manifest as ischemic stroke or acute MI. Rarely, platelet-rich thrombi in the distal aorta or iliac arteries can cause critical limb ischemia.

The diagnosis of HIT is established using enzyme-linked assays to detect antibodies against heparin-PF4 complexes or with platelet activation assays. Enzyme-linked assays are sensitive but can be positive in the absence of any clinical evidence of HIT. The most specific diagnostic test for HIT is the serotonin release assay. This test is performed by quantifying serotonin release when washed platelets loaded with labeled serotonin are exposed to patient serum in the absence or presence of varying concentrations of heparin. If the patient serum contains the HIT antibody, heparin addition induces platelet activation and serotonin release. Management of HIT is outlined in Table 123-4. Heparin should be stopped in patients with suspected or documented HIT, and an alternative anticoagulant should be administered to prevent or treat thrombosis. The agents most often used for this indication are parenteral direct thrombin inhibitors, such as argatroban or bivalirudin, or factor Xa inhibitors, such as fondaparinux or rivaroxaban. A HIT-like syndrome known as vaccine-induced thrombotic thrombocytopenia was reported as a rare complication after vaccination with adenovirus COVID-19 vaccines. Characterized by thrombosis and thrombocytopenia that occurred 4–28 days after vaccination, patients presented with cerebral or splanchnic vein thrombosis as well as DVT or PE. The diagnosis is established by the detection of antibodies against PF4 and a positive serotonin release assay with added PF4. Treatment can include intravenous immunoglobulin, steroids, and plasma exchange to offset the effects of the antibodies against PF4, as well as anticoagulants such as argatroban, fondaparinux, or rivaroxaban to treat the thrombosis.

TABLE 123-4 Management of Heparin-Induced Thrombocytopenia

- Stop all heparins.
- Give an alternative anticoagulant, such as argatroban, bivalirudin, fondaparinux, or rivaroxaban.
- Do not give platelet transfusions.
- Do not give warfarin until the platelet count returns to its baseline level. If warfarin was administered, give vitamin K to restore the INR to normal.
- Evaluate for thrombosis, particularly deep vein thrombosis.

Abbreviation: INR, international normalized ratio.

Patients with HIT, particularly those with associated thrombosis, often have evidence of increased thrombin generation that can lead to consumption of protein C. If these patients are given warfarin without a concomitant anticoagulant that inhibits thrombin or thrombin generation, the further decrease in protein C levels induced by the warfarin can trigger skin necrosis. To avoid this problem, patients with HIT should be treated with a direct thrombin inhibitor or with fondaparinux until the platelet count returns to normal levels. At this point, low-dose warfarin therapy can be introduced, and the parenteral anticoagulant can be discontinued when the international normalized ratio (INR) has been therapeutic for at least 2 days. Alternatively, a direct oral anticoagulant, such as apixaban or rivaroxaban, can be given.

Osteoporosis Treatment with therapeutic doses of heparin for >1 month can cause a reduction in bone density. This complication has been reported in up to 30% of patients given long-term heparin therapy, and symptomatic vertebral fractures occur in 2–3% of these individuals.

Heparin affects the activity of osteoblasts and osteoclasts and causes bone loss both by decreasing bone formation and by enhancing bone resorption.

Elevated Levels of Transaminases Therapeutic doses of heparin are frequently associated with modest elevations in the serum levels of hepatic transaminases without a concomitant increase in the level of bilirubin. The levels of transaminases rapidly return to normal when heparin is stopped. The mechanism responsible for this phenomenon is unknown.

Low-Molecular-Weight Heparin Consisting of smaller fragments of heparin, LMWH is prepared from unfractionated heparin by controlled enzymatic or chemical depolymerization. The mean molecular weight of LMWH is about 5000, one-third the mean molecular weight of unfractionated heparin. LMWH has advantages over heparin (Table 123-5) and has replaced heparin for most indications.

MECHANISM OF ACTION Like heparin, LMWH exerts its anticoagulant activity by activating antithrombin. With a mean molecular weight of 5000, which corresponds to about 17 saccharide units, at least half of the pentasaccharide-containing chains of LMWH are too short to bridge thrombin to antithrombin (Fig. 123-5). However, these chains retain the capacity to accelerate factor Xa inhibition by antithrombin

TABLE 123-5 Advantages of LMWH Over Heparin

ADVANTAGE	CONSEQUENCE
Better bioavailability and longer half-life after subcutaneous injection	Can be given subcutaneously once or twice daily for both prophylaxis and treatment
Dose-independent clearance	Simplified dosing
Predictable anticoagulant response	Coagulation monitoring is unnecessary in most patients
Lower risk of heparin-induced thrombocytopenia	Safer than heparin for short- or long-term administration
Lower risk of osteoporosis	Safer than heparin for extended administration

Abbreviation: LMWH, low-molecular-weight heparin.

because this activity is largely the result of the conformational changes in antithrombin evoked by pentasaccharide binding. Consequently, LMWH catalyzes factor Xa inhibition by antithrombin more than thrombin inhibition. Depending on their unique molecular weight distributions, LMWH preparations have anti–factor Xa to anti–factor IIa ratios ranging from 2:1 to 4:1.

PHARMACOLOGY Although usually given SC, LMWH also can be administered IV if a rapid anticoagulant response is needed. LMWH has pharmacokinetic advantages over heparin. These advantages reflect the fact that shorter heparin chains bind less avidly to endothelial cells, macrophages, and heparin-binding plasma proteins. Reduced binding to endothelial cells and macrophages eliminates the rapid, dose-dependent, and saturable mechanism of clearance that is a characteristic of unfractionated heparin. Instead, the clearance of LMWH is dose-independent and its plasma half-life is longer. Based on measurement of anti–factor Xa levels, LMWH has a plasma half-life of ~4–6 h. LMWH is cleared almost exclusively by the kidneys, and the drug can accumulate in patients with renal insufficiency.

LMWH exhibits about 90% bioavailability after SC injection. Because LMWH binds less avidly to heparin-binding proteins in plasma than heparin, LMWH produces a more predictable dose response, and resistance to LMWH is rare. With a longer half-life and more predictable anticoagulant response, LMWH can be given SC once or twice daily without coagulation monitoring, even when the drug is given in treatment doses. These properties render LMWH more convenient to administer than unfractionated heparin. Capitalizing on this feature, studies in patients with VTE have shown that home treatment with LMWH is as effective and safe as in-hospital treatment with continuous IV infusions of heparin. Outpatient treatment with LMWH streamlines care, reduces health care costs, and increases patient satisfaction.

MONITORING In most patients, LMWH does not require coagulation monitoring. If monitoring is necessary, anti–factor Xa levels must be measured because LMWH preparations have little effect on the aPTT. Therapeutic anti–factor Xa levels for once-daily and twice-daily dosing of LMWH range from 0.5 to 1.2 units/mL and 1.0 to 2.0 units/mL, respectively, when measured 3–4 h after drug administration. When LMWH is given in prophylactic doses, peak anti–factor Xa levels of 0.2–0.5 units/mL are desirable.

Indications for LMWH monitoring include renal impairment and obesity. LMWH monitoring in patients with a creatinine clearance ≤30 mL/min is advisable to ensure that there is no drug accumulation. Although weight-adjusted LMWH dosing appears to produce therapeutic anti–factor Xa levels in patients who are overweight, this approach has not been extensively evaluated in those with morbid obesity. It may also be advisable to monitor the anticoagulant activity of LMWH during pregnancy because dose requirements can change, particularly in the third trimester. Monitoring should also be considered in high-risk settings, such as in pregnant women with mechanical heart valves who are given LMWH to prevent valve thrombosis, and when LMWH is used in treatment doses in infants or children.

DOSING The doses of LMWH recommended for prophylaxis or treatment vary depending on the LMWH preparation. For prophylaxis, once-daily SC doses of 4000–5000 units are often used, whereas doses of 2500–3000 units are given when the drug is administered twice daily. For treatment of VTE, a dose of 150–200 units/kg is given if the drug is administered once daily. If a twice-daily regimen is used, a dose of 100 units/kg is given. In patients with unstable angina, LMWH is given SC on a twice-daily basis at a dose of 100–120 units/kg.

SIDE EFFECTS The major complication of LMWH is bleeding. Meta-analyses suggest that the risk of major bleeding is lower with LMWH than with unfractionated heparin. HIT and osteoporosis are less common with LMWH than with unfractionated heparin.

Bleeding Like the situation with heparin, bleeding with LMWH is more common in patients receiving concomitant therapy with antiplatelet or fibrinolytic drugs. Recent surgery, trauma, or underlying hemostatic defects also increase the risk of bleeding with LMWH.

Although protamine sulfate can be used as an antidote for LMWH, protamine sulfate incompletely neutralizes the anticoagulant activity of LMWH because it only binds the longer chains of LMWH. Because longer chains are responsible for catalysis of thrombin inhibition by antithrombin, protamine sulfate completely reverses the anti–factor IIa activity of LMWH. In contrast, protamine sulfate only partially reverses the anti–factor Xa activity of LMWH because the shorter pentasaccharide-containing chains of LMWH do not bind to protamine sulfate. Consequently, patients at high risk for bleeding may be more safely treated with continuous IV unfractionated heparin than with SC LMWH. Andexanet alfa, a recombinant factor Xa variant licensed for reversal of oral factor Xa inhibitors, reverses that anti–factor Xa activity of LMWH but not its anti–factor IIa activity. The utility of andexanet for LMWH reversal is uncertain, and it is not licensed for this indication.

Thrombocytopenia The risk of HIT is about fivefold lower with LMWH than with heparin. LMWH binds less avidly to platelets and causes less PF4 release. Furthermore, with lower affinity for PF4 than heparin, LMWH is less likely to induce the conformational changes in PF4 that trigger the formation of HIT antibodies.

LMWH should not be used to treat HIT patients because most HIT antibodies exhibit cross-reactivity with LMWH. This in vitro cross-reactivity is more than a laboratory phenomenon because there are case reports of thrombosis when HIT patients were switched from heparin to LMWH.

Osteoporosis Because the risk of osteoporosis is lower with LMWH than with heparin, LMWH is a better choice than heparin for extended treatment.

Fondaparinux A synthetic analogue of the antithrombin-binding pentasaccharide sequence, fondaparinux differs from LMWH in several ways **(Table 123-6)**. Fondaparinux is licensed for thromboprophylaxis in general medical or surgical patients and in high-risk orthopedic patients and as an alternative to heparin or LMWH for initial treatment of patients with established VTE. Although fondaparinux is used in Europe as an alternative to heparin or LMWH in patients with acute coronary syndrome, the drug is not licensed for this indication in the United States.

MECHANISM OF ACTION Fondaparinux has a molecular weight of 1728. Fondaparinux binds only to antithrombin (Fig. 123-5) and is too short to bridge thrombin to antithrombin. Consequently, fondaparinux catalyzes factor Xa inhibition by antithrombin and does not enhance the rate of thrombin inhibition.

PHARMACOLOGY Fondaparinux exhibits complete bioavailability after SC injection. With no binding to endothelial cells or plasma proteins, the clearance of fondaparinux is dose independent, and its plasma half-life is 17 h. The drug is given SC once daily. Because fondaparinux is cleared unchanged via the kidneys, it is contraindicated in patients with a creatinine clearance <30 mL/min and should be used with caution in those with a creatinine clearance <50 mL/min.

DOSING Fondaparinux produces a predictable anticoagulant response after administration in fixed doses because it does not bind to plasma proteins. The drug is given at a dose of 2.5 mg once daily for prevention of VTE. For initial treatment of established VTE, fondaparinux is

TABLE 123-6 Comparison of LMWH and Fondaparinux

FEATURES	LMWH	FONDAPARINUX
Number of saccharide units	15–17	5
Catalysis of factor Xa inhibition	Yes	Yes
Catalysis of thrombin inhibition	Yes	No
Bioavailability after subcutaneous administration (%)	90	100
Plasma half-life (h)	4	17
Renal excretion	Yes	Yes
Induces release of tissue factor pathway inhibitor	Yes	No
Neutralized by protamine sulfate	Partially	No

given at a dose of 7.5 mg once daily. The dose can be reduced to 5 mg once daily for those weighing <50 kg and increased to 10 mg for those >100 kg. When given in these doses, fondaparinux is as effective as heparin or LMWH for initial treatment of patients with DVT or PE and is associated with similar rates of bleeding.

Fondaparinux is used at a dose of 2.5 mg once daily in patients with acute coronary syndrome. When this prophylactic dose of fondaparinux was compared with treatment doses of enoxaparin in patients with non-ST-segment elevation MI, there was no difference in the rate of cardiovascular death, MI, or stroke at 9 days. However, the rate of major bleeding was 50% lower with fondaparinux than with enoxaparin, a difference that likely reflects the fact that the dose of fondaparinux was lower than that of enoxaparin. In acute coronary syndrome patients who require PCI, there is a risk of catheter thrombosis with fondaparinux unless adjunctive heparin is given at the time of the procedure.

SIDE EFFECTS Fondaparinux does not cause HIT because it does not bind to PF4. In contrast to LMWH, there is no cross-reactivity of fondaparinux with HIT antibodies. Consequently, fondaparinux appears to be effective for treatment of HIT patients, although large clinical trials supporting its use are lacking.

The major side effect of fondaparinux is bleeding. Fondaparinux has no antidote. Protamine sulfate has no effect on the anticoagulant activity of fondaparinux because it fails to bind to the drug. Andexanet alfa has been reported to reverse fondaparinux in vitro, but studies in patients are lacking. Recombinant activated factor VII reverses the anticoagulant effects of fondaparinux in volunteers, but it is unknown whether this agent controls fondaparinux-induced bleeding.

Parenteral Direct Thrombin Inhibitors Direct thrombin inhibitors bind directly to thrombin and block its interaction with its substrates. Approved parenteral direct thrombin inhibitors include recombinant hirudins (lepirudin and desirudin), argatroban, and bivalirudin (Table 123-7). Lepirudin and desirudin are no longer available.

ARGATROBAN A univalent inhibitor that targets the active site of thrombin, argatroban is metabolized in the liver. Consequently, this drug must be used with caution in patients with hepatic insufficiency. Argatroban is administered by continuous IV infusion and has a plasma half-life of ~45 min. The aPTT is used to monitor its anticoagulant effect, and the dose is adjusted to achieve an aPTT 1.5–3 times the baseline value, but not to exceed 100 s. Argatroban also prolongs the INR, a feature that can complicate the transitioning of patients from argatroban to warfarin. This problem can be circumvented by using the levels of factor X to monitor warfarin instead of the INR. Alternatively, argatroban can be stopped for 2–3 h before INR determination.

Argatroban is licensed for treatment of patients with HIT or a history of HIT, including those requiring PCI. In such patients, argatroban is most useful for those with severe kidney disease because unlike fondaparinux and bivalirudin, it is not cleared through the kidneys.

BIVALIRUDIN A synthetic 20-amino-acid analogue of hirudin, bivalirudin is a divalent thrombin inhibitor. Thus, the N-terminus of bivalirudin interacts with the active site of thrombin, whereas its C-terminus binds to exosite 1. Bivalirudin has a plasma half-life of 25 min, the shortest half-life of all the parenteral direct thrombin inhibitors. Bivalirudin is degraded by peptidases and is partially excreted via the kidneys. When given in high doses in the cardiac catheterization laboratory, the anticoagulant activity of bivalirudin is monitored using the activated clotting time. With lower doses, its activity can be assessed using the aPTT.

Bivalirudin is licensed as an alternative to heparin in patients undergoing PCI. Bivalirudin also has been used successfully in HIT patients who require PCI or cardiac bypass surgery.

TABLE 123-7 Comparison of the Properties of Desirudin, Bivalirudin, and Argatroban

	DESIRUDIN	BIVALIRUDIN	ARGATROBAN
Molecular mass	7000	1980	527
Site(s) of interaction with thrombin	Active site and exosite 1	Active site and exosite 1	Active site
Renal clearance	Yes	No	No
Hepatic metabolism	No	No	Yes
Plasma half-life (min)	60 (IV) 120–180 (SC)	25	45

ORAL ANTICOAGULANTS

For many years, vitamin K antagonists such as warfarin were the only available oral anticoagulants. This situation changed with the introduction of the direct oral anticoagulants, which include dabigatran, rivaroxaban, apixaban, and edoxaban.

Warfarin A water-soluble vitamin K antagonist initially developed as a rodenticide; warfarin is the coumarin derivative most often prescribed in North America. Like other vitamin K antagonists, warfarin interferes with the synthesis of the vitamin K–dependent clotting proteins, which include prothrombin (factor II) and factors VII, IX, and X. The synthesis of the vitamin K–dependent anticoagulant proteins, proteins C and S, is also reduced by vitamin K antagonists.

MECHANISM OF ACTION All of the vitamin K–dependent clotting factors possess glutamic acid residues at their N termini. A posttranslational modification adds a carboxyl group to the γ-carbon of these residues to generate γ-carboxyglutamic acid. This modification is essential for expression of the activity of these clotting factors because it permits their calcium-dependent binding to negatively charged phospholipid surfaces. The γ-carboxylation process is catalyzed by a vitamin K–dependent carboxylase. Thus, vitamin K from the diet is reduced to vitamin K hydroquinone by vitamin K reductase (Fig. 123-6). Vitamin K hydroquinone serves as a cofactor for the carboxylase enzyme, which in the presence of carbon dioxide replaces the hydrogen on the γ-carbon of glutamic acid residues with a carboxyl group. During this process, vitamin K hydroquinone is oxidized to vitamin K epoxide, which is then reduced to vitamin K by vitamin K epoxide reductase.

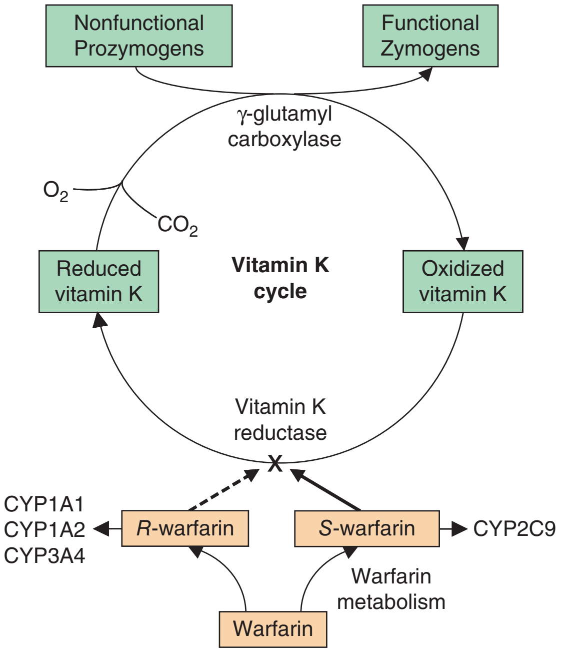

FIGURE 123-6 Mechanism of action of warfarin. A racemic mixture of *S*- and *R*-enantiomers, *S*-warfarin is most active. By blocking vitamin K epoxide reductase, warfarin inhibits the conversion of oxidized vitamin K into its reduced form. This inhibits vitamin K–dependent γ-carboxylation of factors II, VII, IX, and X because reduced vitamin K serves as a cofactor for a γ-glutamyl carboxylase that catalyzes the γ-carboxylation process, thereby converting pro-zymogens to zymogens capable of binding calcium and interacting with anionic phospholipid surfaces. *S*-warfarin is metabolized by *CYP2C9*. Common genetic polymorphisms in this enzyme can influence warfarin metabolism. Polymorphisms in the C1 subunit of vitamin K reductase (*VKORC1*) also can affect the susceptibility of the enzyme to warfarin-induced inhibition, thereby influencing warfarin dosage requirements.

Warfarin inhibits vitamin K epoxide reductase (VKOR), thereby blocking the γ-carboxylation process. This results in the synthesis of vitamin K–dependent clotting proteins that are only partially γ-carboxylated. Warfarin acts as an anticoagulant because these partially γ-carboxylated proteins have little or no biological activity. The onset of action of warfarin is delayed until the newly synthesized clotting factors with reduced activity gradually replace their fully active counterparts.

The antithrombotic effect of warfarin depends on a reduction in the functional levels of factor X and prothrombin, clotting factors that have half-lives of 24 and 72 h, respectively. Because the antithrombotic effect of warfarin is delayed, patients with established thrombosis or at high risk for thrombosis require concomitant treatment with a rapidly acting parenteral anticoagulant, such as heparin, LMWH, or fondaparinux, for at least 5 days.

PHARMACOLOGY Warfarin is a racemic mixture of R and S isomers. Warfarin is rapidly and almost completely absorbed from the gastrointestinal tract. Levels of warfarin in the blood peak about 90 min after drug administration. Racemic warfarin has a plasma half-life of 36–42 h, and >97% of circulating warfarin is bound to albumin. Only the small fraction of unbound warfarin is biologically active.

Warfarin accumulates in the liver where the two isomers are metabolized via distinct pathways. _CYP2C9_ mediates oxidative metabolism of the more active S isomer (Fig. 123-6). Two relatively common variants, _CYP2C9*2_ and _CYP2C9*3_, encode an enzyme with reduced activity. Patients with these variants require lower maintenance doses of warfarin. Approximately 25% of Caucasians have at least one variant allele of _CYP2C9*2_ or _CYP2C9*3_, whereas those variant alleles are less common in African Americans and Asians **(Table 123-8)**. Heterozygosity for _CYP2C9*2_ or _CYP2C9*3_ decreases the warfarin dose requirement by 20–30% relative to that required in subjects with the wild-type _CYP2C9*1/*1_ alleles, whereas homozygosity for the _CYP2C9*2_ or _CYP2C9*3_ alleles reduces the warfarin dose requirement by 50–70%.

Consistent with their decreased warfarin dose requirement, subjects with at least one _CYP2C9_ variant allele are at increased risk for bleeding. Compared with individuals with no variant alleles, the risk of warfarin-associated bleeding is almost twofold higher in _CYP2C9*2_ or _CYP2C9*3_ carriers.

Polymorphisms in _VKORC1_ also can influence the anticoagulant response to warfarin. Several genetic variations of _VKORC1_ are in strong linkage disequilibrium and have been designated as non-A haplotypes. _VKORC1_ variants are more prevalent than variants of _CYP2C9_. Asians have the highest prevalence of _VKORC1_ variants, followed by Caucasians and African Americans (Table 123-8). Polymorphisms in _VKORC1_ likely explain 30% of the variability in warfarin dose requirements. Compared with _VKORC1_ non-A/non-A homozygotes, the warfarin dose requirement decreases by 25 and 50% in A haplotype heterozygotes and homozygotes, respectively. These findings prompted the U.S. Food and Drug Administration (FDA) to amend the prescribing information for warfarin to indicate that lower initiation doses should be considered for patients with _CYP2C9_ and _VKORC1_ genetic variants. In addition to genotype data, other pertinent patient information has been incorporated into warfarin dosing algorithms. Although such algorithms help predict suitable warfarin doses, it remains unclear whether better dose identification improves patient outcome in terms of reducing hemorrhagic complications or recurrent thrombotic events.

TABLE 123-8 Frequencies of _CYP2C9_ Genotypes and _VKORC1_ Haplotypes in Different Populations and their Effect on Warfarin Dose Requirements

GENOTYPE/ HAPLOTYPE	FREQUENCY, %			PERCENTAGE DOSE REDUCTION COMPARED WITH WILD-TYPE
	CAUCASIANS	AFRICAN AMERICANS (A/A)	ASIANS (A)	
CYP2C9				
*1/*1	70	90	95	—
*1/*2	17	2	0	22
*1/*3	9	3	4	34
*2/*2	2	0	0	43
*2/*3	1	0	0	53
*3/*3	0	0	1	76
VKORC1				
Non-A/non-A	37	82	7	—
Non-A/A	45	12	30	26
A/A	18	6	63	50

In addition to genetic factors, the anticoagulant effect of warfarin is influenced by diet, drugs, and various disease states. Fluctuations in dietary vitamin K intake affect the activity of warfarin. A wide variety of drugs can alter absorption, clearance, or metabolism of warfarin. Because of the variability in the anticoagulant response to warfarin, coagulation monitoring is essential to ensure that a therapeutic response is obtained.

MONITORING Warfarin therapy is most often monitored using the prothrombin time, a test that is sensitive to reductions in the levels of prothrombin, factor VII, and factor X. The test is performed by adding thromboplastin, a reagent that contains tissue factor, phospholipid, and calcium, to citrated plasma and determining the time to clot formation. Thromboplastins vary in their sensitivity to reductions in the levels of the vitamin K–dependent clotting factors. Thus, less sensitive thromboplastins will trigger the administration of higher doses of warfarin to achieve a target prothrombin time. This is problematic because higher doses of warfarin increase the risk of bleeding.

The INR was developed to circumvent many of the problems associated with the prothrombin time. To calculate the INR, the patient's prothrombin time is divided by the mean normal prothrombin time, and this ratio is then multiplied by the international sensitivity index (ISI), which is an index of the sensitivity of the thromboplastin used for prothrombin time determination to reductions in the levels of the vitamin K–dependent clotting factors. Sensitive thromboplastins have an ISI near 1.0. Most current thromboplastins have ISI values that range from 0.9 to 1.4.

Although the INR has helped to standardize anticoagulant practice, problems persist. The precision of INR determination varies depending on reagent-coagulometer combinations. This leads to variability in the INR results. Also complicating INR determination is unreliable reporting of the ISI by thromboplastin manufacturers. Furthermore, every laboratory must establish the mean normal prothrombin time with each new batch of thromboplastin reagent. To accomplish this, the prothrombin time must be measured in fresh plasma samples from at least 20 healthy volunteers using the same coagulometer that is used for patient samples.

For most indications, warfarin is administered in doses that produce a target INR of 2.0–3.0. An exception is patients with mechanical heart valves, particularly those in the mitral position or older ball and cage valves in the aortic position, and valves in any position associated with atrial fibrillation, where a target INR of 2.5–3.5 is recommended. Studies in atrial fibrillation demonstrate an increased risk of cardioembolic stroke when the INR falls below 1.7 and an increase in bleeding with INR values >4.5. These findings highlight the fact that vitamin K antagonists have a narrow therapeutic window. In support of this concept, a study in patients receiving long-term warfarin therapy for unprovoked VTE demonstrated a higher rate of recurrent VTE with a target INR of 1.5–1.9 compared with a target INR of 2.0–3.0.

DOSING Warfarin is usually started at a dose of 5–10 mg. Lower doses are used for patients with _CYP2C9_ or _VKORC1_ polymorphisms, which affect the pharmacodynamics or pharmacokinetics of warfarin and render patients more sensitive to the drug. The dose is then titrated to achieve the desired target INR. Because of its delayed onset of action, patients with established thrombosis or those at high risk for thrombosis are given concomitant initial treatment with a rapidly acting parenteral anticoagulant, such as heparin, LMWH, or fondaparinux. Early prolongation of the INR reflects reduction in the functional levels of factor VII. Consequently, concomitant treatment with the parenteral anticoagulant should be continued until the INR has been therapeutic for at least 2 consecutive days. A minimum 5-day course of parenteral anticoagulation is recommended to ensure that the levels of factor Xa and prothrombin have been reduced into the therapeutic range with warfarin.

Because warfarin has a narrow therapeutic window, frequent coagulation monitoring is essential to ensure that a therapeutic anticoagulant response is maintained. Even patients with stable warfarin dose requirements should have their INR determined every 3–4 weeks although there are studies suggesting that less frequent monitoring is feasible. More frequent monitoring is necessary when new medications are introduced because so many drugs enhance or reduce the anticoagulant effects of warfarin and when the dosing regimen has been changed.

SIDE EFFECTS Like all anticoagulants, the major side effect of warfarin is bleeding. A rare complication is skin necrosis. Warfarin crosses the placenta and can cause fetal abnormalities. Consequently, warfarin should not be used during pregnancy.

Bleeding At least half of the bleeding complications with warfarin occur when the INR exceeds the therapeutic range. Bleeding complications may be mild, such as epistaxis or hematuria, or more severe, such as retroperitoneal or gastrointestinal bleeding. Life-threatening intracranial bleeding can also occur.

To minimize the risk of bleeding, the INR should be maintained in the therapeutic range. In asymptomatic patients whose INR is between 3.5 and 10, warfarin should be withheld until the INR returns to the therapeutic range. If the INR is over 10, oral vitamin K can be administered at a dose of 2.5–5 mg, although there is no evidence that doing so reduces the bleeding risk. Higher doses of oral vitamin K (5–10 mg) produce more rapid reversal of the INR but may render patients temporarily resistant to warfarin when the drug is restarted.

Patients with an increased INR associated with serious bleeding should be given 5–10 mg of vitamin K by slow IV infusion. Additional vitamin K should be given until the INR is in the normal range. Treatment with vitamin K should be supplemented with four-factor prothrombin complex concentrate, which contains all four vitamin K–dependent clotting proteins. Prothrombin complex concentrate normalizes the INR more rapidly than transfusion of fresh-frozen plasma.

Warfarin-treated patients who experience bleeding when their INR is in the therapeutic range require investigation into the cause of the bleeding. Those with gastrointestinal or genitourinary bleeding often have underlying disorders.

Skin Necrosis A rare complication of warfarin, skin necrosis usually is seen 2–5 days after initiation of therapy. Well-demarcated erythematous lesions form on the thighs, buttocks, breasts, or toes. Typically, the center of the lesion becomes progressively necrotic. Examination of skin biopsies taken from the border of these lesions reveals thrombi in the microvasculature.

Warfarin-induced skin necrosis is seen in patients with congenital or acquired deficiencies of protein C or protein S. Initiation of warfarin therapy in these patients produces a precipitous fall in plasma levels of proteins C or S, thereby eliminating this important anticoagulant pathway before warfarin exerts an antithrombotic effect through lowering of the functional levels of factor X and prothrombin. The resultant procoagulant state triggers thrombosis. Why the thrombosis is localized to the microvasculature of fatty tissues is unclear.

Treatment involves discontinuation of warfarin and reversal with vitamin K, if needed. An alternative anticoagulant, such as heparin or LMWH, should be given in patients with thrombosis. Protein C concentrate can be given to protein C–deficient patients to accelerate healing of the skin lesions; fresh-frozen plasma may be of value if protein C concentrate is unavailable and for those with protein S deficiency. Occasionally, skin grafting is necessary when there is extensive skin loss.

Because of the potential for skin necrosis, patients with known protein C or protein S deficiency require overlapping treatment with a parenteral anticoagulant when initiating warfarin therapy. Warfarin should be started in low doses in these patients, and the parenteral anticoagulant should be continued until the INR is therapeutic for at least 2–3 consecutive days. Alternatively, treatment with rivaroxaban or apixaban could be given, although there is limited information about their efficacy and safety in patients with severe protein C or S deficiency.

Pregnancy Warfarin crosses the placenta and can cause fetal abnormalities or bleeding. The fetal abnormalities include a characteristic embryopathy, which consists of nasal hypoplasia and stippled epiphyses. The risk of embryopathy is highest if warfarin is given in the first trimester of pregnancy. Central nervous system abnormalities can also occur with exposure to warfarin at any time during pregnancy. Finally, maternal administration of warfarin produces an anticoagulant effect in the fetus that can cause bleeding. This is of particular concern at delivery when trauma to the head during passage through the birth canal can lead to intracranial bleeding. Because of these potential problems, warfarin is contraindicated in pregnancy, particularly in the first and third trimesters. Instead, heparin, LMWH, or fondaparinux can be given during pregnancy for prevention or treatment of thrombosis.

Warfarin does not pass into the breast milk. Consequently, warfarin can safely be given to nursing mothers.

Special Problems Patients with antiphospholipid syndrome and those who need urgent or elective surgery present special challenges. Although observational studies suggested that patients with thrombosis complicating antiphospholipid syndrome required higher intensity warfarin regimens to prevent recurrent thromboembolic events, two randomized trials showed that targeting an INR of 2.0–3.0 is as effective as higher intensity treatment and produces less bleeding. Monitoring warfarin therapy can be problematic in patients with the lupus anticoagulant because it prolongs the baseline INR; factor X levels can be used instead of the INR in such patients.

There is no need to stop warfarin before procedures associated with a low risk of bleeding; these include dental cleaning, simple dental extraction, cataract surgery, or skin biopsy. For procedures associated with a moderate or high risk of bleeding, warfarin should be stopped 5 days before the procedure to allow the INR to return to normal levels. Patients at high risk for thrombosis, such as those with mechanical heart valves, can be bridged with once- or twice-daily SC injections of LMWH when the INR falls to <2.0. The last dose of LMWH should be given 12–24 h before the procedure, depending on whether LMWH is administered twice or once daily. After the procedure, treatment with warfarin can be restarted.

Direct Oral Anticoagulants The direct oral anticoagulants (DOACs) include dabigatran, which inhibits thrombin, and rivaroxaban, apixaban, and edoxaban, which inhibit factor Xa. These drugs have a rapid onset and offset of action and have half-lives that permit once- or twice-daily administration. Designed to produce a predictable level of anticoagulation, the DOACs are more convenient to administer than warfarin because they are given in fixed doses without routine coagulation monitoring.

MECHANISM OF ACTION The DOACs are small molecules that bind reversibly to the active site of their target enzyme. Table 123-9 summarizes the distinct pharmacologic properties of these agents.

INDICATIONS All four DOACs are licensed for stroke prevention in patients with atrial fibrillation except those with mechanical heart valves or severe rheumatic mitral valve disease, and for treatment of VTE. Dabigatran, rivaroxaban, and apixaban are licensed for thromboprophylaxis after elective hip or knee arthroplasty; edoxaban is only licensed for this indication in Japan. Finally, low-dose rivaroxaban is licensed for use with aspirin for secondary prevention in patients with coronary or peripheral artery disease.

DOSING For prevention of stroke in patients with atrial fibrillation, rivaroxaban is given at a dosage of 20 mg once daily, with a reduction to 15 mg once daily in patients with a creatinine clearance of 15–49 mL/min; dabigatran is given at a dosage of 150 mg twice daily, with a reduction to 75 mg twice daily in those with a creatinine clearance of 15–30 mL/min; apixaban is given at a dosage of 5 mg twice daily, with a reduction to 2.5 mg twice daily for patients with at least two of the

TABLE 123-9 Comparison of the Pharmacologic Properties of the Direct Oral Anticoagulants

CHARACTERISTIC	RIVAROXABAN	APIXABAN	EDOXABAN	DABIGATRAN
Target	Factor Xa	Factor Xa	Factor Xa	Thrombin
Prodrug	No	No	No	Yes
Bioavailability	80%	60%	50%	6%
Dosing	qd (bid)	bid	qd	bid (qd)
Half-life	7–11 h	12 h	9–11 h	12–17 h
Renal excretion	33% (66%)	25%	35%	80%
Interactions	3A4/P-gp	3A4/P-gp	P-gp	P-gp

Abbreviations: bid, twice a day; P-gp, P-glycoprotein; qd, once a day.

"ABC" criteria (i.e., age >80 years, body weight <60 kg, and creatinine >1.5 g/dL); and edoxaban is given at a dosage of 60 mg once daily for patients with a creatinine clearance of 50–95 mL/min and with a reduction to 30 mg once daily for patients with any one of the following criteria: creatinine clearance of 15–50 mL/min, body weight of 60 kg or less, or use of potent P-glycoprotein inhibitors, such as verapamil or quinidine. At doses of 15 or 20 mg once daily, rivaroxaban must be administered with food to enhance absorption. Apixaban and edoxaban can be given with or without food. Administration of dabigatran with food may reduce dyspepsia.

For treatment of VTE, dabigatran and edoxaban are started after patients have received at least a 5-day course of treatment with a parenteral anticoagulant such as LMWH. Dabigatran is given at a dose of 150 mg twice daily provided the creatinine clearance is >30 mL/min. The dosage regimen for edoxaban is identical to that used in patients with atrial fibrillation. Rivaroxaban and apixaban can be given in all-oral regimens; rivaroxaban is started at a dose of 15 mg twice daily for 21 days and is then reduced to 20 mg once daily thereafter, whereas apixaban is started at a dose of 10 mg twice daily for 7 days and is then reduced to 5 mg twice daily thereafter. For secondary VTE prevention after 6 months of full-dose treatment, the dosage of apixaban can be lowered to 2.5 mg twice daily while the dose of rivaroxaban can be lowered to 10 mg once daily, doses that have safety profiles like those of placebo and aspirin, respectively.

Thromboprophylaxis after elective hip or knee replacement surgery is started after surgery and is often continued for 30 days in patients undergoing hip replacement and for 10–14 days in patients undergoing knee replacement. Dabigatran is given at a dose of 220 mg once daily, whereas rivaroxaban and apixaban are given at doses of 10 mg once daily and 2.5 mg twice daily, respectively. In lower risk patients undergoing hip or knee replacement surgery, a 5-day course of rivaroxaban followed by a 30-day course of aspirin at a dose of 81 mg daily appears to be as effective and safe as extended thromboprophylaxis with rivaroxaban.

For secondary prevention of adverse cardiac or limb events in patients with coronary or peripheral artery disease, rivaroxaban is given at a dose of 2.5 mg twice daily on top of aspirin (81 or 100 mg once daily).

MONITORING Although designed to be administered without routine monitoring, there are situations where determination of the anticoagulant activity of the DOACs can be helpful. These include assessment of adherence, detection of accumulation or overdose, identification of bleeding mechanisms, and determination of activity before surgery, intervention, or reversal. For qualitative assessment of anticoagulant activity, the prothrombin time can be used for factor Xa inhibitors and the aPTT for dabigatran. Rivaroxaban and edoxaban prolong the prothrombin time more than apixaban. In fact, because apixaban has such a limited effect on the prothrombin time, anti–factor Xa assays are needed to assess its activity. The effect of the drugs on tests of coagulation varies depending on the time that the blood is drawn relative to the timing of the last dose of the drug and the reagents used to perform the tests. Chromogenic anti–factor Xa assays and the diluted thrombin clotting time or ecarin clot time with appropriate calibrators provide quantitative assays to measure the plasma levels of the factor Xa inhibitors and dabigatran, respectively.

SIDE EFFECTS Like all anticoagulants, bleeding is the most common side effect of the DOACs. The DOACS are associated with less intracranial bleeding than warfarin, but the higher dose regimens of dabigatran, rivaroxaban, and edoxaban are associated with more gastrointestinal bleeding.

Dyspepsia occurs in up to 10% of patients treated with dabigatran; this problem improves with time and can be minimized by administering the drug with food. Dyspepsia is rare with rivaroxaban, apixaban, and edoxaban.

PERIPROCEDURAL MANAGEMENT Like warfarin, the DOACs must be stopped before procedures associated with a moderate or high risk of bleeding. The drugs should be held for 1–2 days, or longer if renal function is impaired. Assessment of residual anticoagulant activity before procedures associated with a high bleeding risk is prudent.

MANAGEMENT OF BLEEDING With minor bleeding, withholding one or two doses of drug is usually sufficient. With more serious bleeding, the approach is like that with warfarin, except that vitamin K administration is of no benefit; the anticoagulant and any long-acting antiplatelet drugs should be withheld, the patient should be resuscitated with fluids and blood products as necessary, and the bleeding site should be identified and managed. Coagulation testing or measurement of the DOAC level will determine the extent of anticoagulation, and renal function should be assessed so that the half-life of the drug can be calculated. Timing of the last dose of anticoagulant is important; in cases of overdose, oral activated charcoal may help prevent absorption of drug administered in the past 4 h. If >24 h have elapsed since the last intake, the DOAC is unlikely to be responsible for the bleeding unless there is marked impairment of renal function.

Anticoagulant reversal should be considered if bleeding continues despite supportive measures or if the bleeding is life-threatening or occurs in a critical organ (e.g., intracranial) or in a closed space (e.g., the pericardium or retroperitoneum). Idarucizumab is licensed for dabigatran reversal in such patients or in those requiring urgent surgery or intervention. A humanized antibody fragment, idarucizumab, binds dabigatran with high affinity to form an essentially irreversible complex that is cleared by the kidneys. Idarucizumab is given intravenously as a 5-g bolus and is supplied in a box containing two 50-mL vials, each containing 2.5 g of idarucizumab. Idarucizumab rapidly reverses the anticoagulant effects of dabigatran and normalizes the aPTT, diluted thrombin time, or ecarin clot time.

Andexanet alfa is available for reversal of rivaroxaban and apixaban. A recombinant variant of factor Xa without catalytic activity, andexanet serves as a decoy to sequester oral factor Xa inhibitors until they are cleared from the circulation. Low- or high-dose IV andexanet regimens are used. The low-dose regimen starts with a bolus of 400 mg followed by an infusion of 4 mg/min for up to 120 min, whereas the high-dose regimen starts with a bolus of 800 mg followed by an infusion of 8 mg/min for up to 120 min. The low-dose regimen is used for reversal of doses of rivaroxaban or apixaban of 10 mg or 5 mg or less, respectively, or for any dose of rivaroxaban or apixaban if the last dose was taken >8 h before presentation. The high-dose regimen is used to reverse rivaroxaban or apixaban doses over 10 and 5 mg, respectively, if the last dose was taken <8 h since presentation, or if the timing of the last dose of rivaroxaban or apixaban is unknown.

Andexanet alfa is expensive and is not available in all hospitals. Because of its cost, andexanet alfa is often reserved for reversal in patients with life-threatening bleeds such intracranial hemorrhage or bleeds into a closed space such as retroperitoneal or pericardial bleeds. If andexanet is unavailable, the results of prospective cohort studies suggest that four-factor prothrombin complex concentrate (25–50 units/kg) also is effective at restoring hemostasis. If there is continued bleeding, activated prothrombin complex concentrate (50 units/kg) or recombinant factor VIIa (90 μg/kg) can be considered.

Neither andexanet alfa nor four-factor prothrombin complex concentrate has been evaluated for reversal in patients requiring urgent surgery or intervention. Furthermore, andexanet alfa not only reverses oral factor Xa inhibitors but also reverses heparin and LMWH. This could be problematic in patients who require cardiac surgery or vascular surgery, procedures where heparin is used routinely. To circumvent this problem, most surgical procedures and interventions can be undertaken without reversal, and four-factor prothrombin complex concentrate can be given if necessary. For patients requiring surgery to stop bleeding such as those with a ruptured aortic aneurysm or with bleeding secondary to polytrauma, upfront four-factor prothrombin concentrate administration can be considered.

PREGNANCY As small molecules, the DOACs pass through the placenta. Consequently, these agents are contraindicated in pregnancy, and when used by women of childbearing potential, appropriate contraception is important. DOACs should be avoided in nursing mothers because small amounts have been found in breast milk.

FIBRINOLYTIC DRUGS

ROLE OF FIBRINOLYTIC THERAPY

Fibrinolytic drugs are used to degrade thrombi and are administered systemically or can be delivered via catheters directly into the substance of the thrombus. Systemic delivery is used for treatment of acute MI, acute ischemic stroke, and most cases of massive PE. The goal of therapy is to produce rapid thrombus dissolution, thereby restoring blood flow. In the coronary circulation, restoration of blood flow reduces morbidity and mortality rates by limiting myocardial damage, whereas in the cerebral circulation, rapid thrombus dissolution decreases the neuronal death and brain infarction that produce irreversible brain injury. For patients with massive PE, the goal of thrombolytic therapy is to restore pulmonary artery perfusion.

Peripheral arterial thrombi and thrombi in the proximal deep veins of the leg are most often treated using catheter-directed thrombolytic therapy. Catheters with multiple side holes can be used to enhance drug delivery. In some cases, intravascular devices that fragment and extract the thrombus are used to hasten treatment. These devices can be used alone or in conjunction with fibrinolytic drugs.

MECHANISM OF ACTION

Currently approved fibrinolytic agents include streptokinase; acylated plasminogen streptokinase activator complex (anistreplase); urokinase; recombinant tissue-type plasminogen activator (rtPA), which is also known as alteplase or activase; and two recombinant derivatives of rtPA, tenecteplase and reteplase. Of these, streptokinase, anistreplase, and urokinase are no longer available in the United States. All these agents act by converting plasminogen, the zymogen, to plasmin, the active enzyme **(Fig. 123-7)**. Plasmin then degrades the fibrin matrix of thrombi and produces soluble fibrin degradation products.

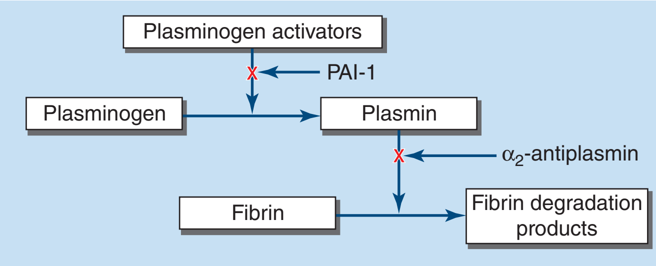

FIGURE 123-7 The fibrinolytic system and its regulation. Plasminogen activators convert plasminogen to plasmin. Plasmin then degrades fibrin into soluble fibrin degradation products. The system is regulated at two levels. Type 1 plasminogen activator inhibitor (PAI-1) inhibits the plasminogen activators, whereas α_2-antiplasmin serves as the major inhibitor of plasmin.

Endogenous fibrinolysis is regulated at two levels. Plasminogen activator inhibitors, particularly the type 1 form (PAI-1), prevent excessive plasminogen activation by regulating the activity of tPA and urokinase-type plasminogen activator (uPA). Once plasmin is generated, it is regulated by plasmin inhibitors, the most important of which is α_2-antiplasmin. The plasma concentration of plasminogen is twofold higher than that of α_2-antiplasmin. Consequently, with pharmacologic doses of plasminogen activators, the concentration of plasmin that is generated can exceed that of α_2-antiplasmin. In addition to degrading fibrin, unregulated plasmin can also degrade fibrinogen and other clotting factors. This process, which is known as the systemic lytic state, reduces the hemostatic potential of the blood and increases the risk of bleeding.

The endogenous fibrinolytic system is geared to localize plasmin generation to the fibrin surface. Both plasminogen and tPA bind to fibrin to form a ternary complex that promotes efficient plasminogen activation. In contrast to free plasmin, plasmin generated on the fibrin surface is relatively protected from inactivation by α_2-antiplasmin, a feature that promotes fibrin dissolution. Furthermore, C-terminal lysine residues, exposed as plasmin degrades fibrin, provide binding sites for additional plasminogen and tPA molecules. This creates positive feedback that enhances plasmin generation. When used pharmacologically, the various plasminogen activators capitalize on these mechanisms to a lesser or greater extent.

Plasminogen activators that preferentially activate fibrin-bound plasminogen are considered fibrin-specific. In contrast, nonspecific plasminogen activators do not discriminate between fibrin-bound and circulating plasminogen. Activation of circulating plasminogen results in the generation of unopposed plasmin that can trigger the systemic lytic state. Alteplase and its derivatives are fibrin-specific plasminogen activators, whereas streptokinase, anistreplase, and urokinase are nonspecific agents.

STREPTOKINASE

Unlike other plasminogen activators, streptokinase is not an enzyme and does not directly convert plasminogen to plasmin. Instead, streptokinase forms a 1:1 stoichiometric complex with plasminogen. Formation of this complex induces a conformational change in plasminogen that exposes its active site **(Fig. 123-8)**. The streptokinase-plasminogen complex then converts additional plasminogen to plasmin.

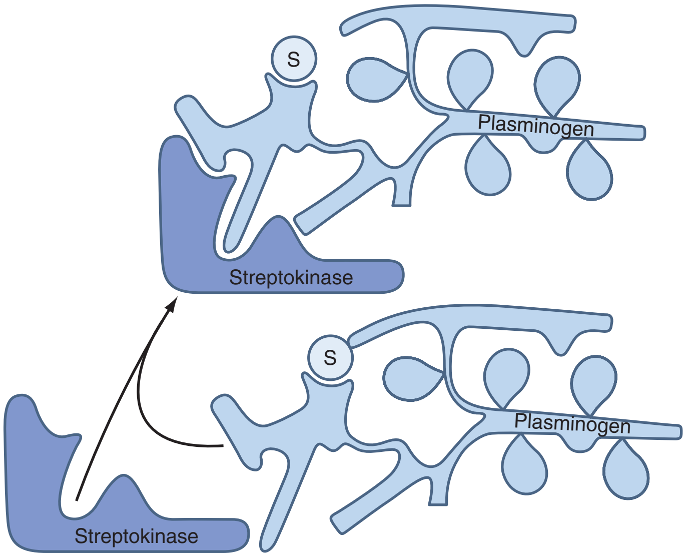

FIGURE 123-8 Mechanism of action of streptokinase. Streptokinase binds to plasminogen and induces a conformational change in plasminogen that exposes its active site. The streptokinase/plasmin(ogen) complex then serves as the activator of additional plasminogen.

Streptokinase has no affinity for fibrin, and the streptokinase-plasminogen complex activates both free and fibrin-bound plasminogen. Activation of circulating plasminogen generates enough plasmin to overwhelm α_2-antiplasmin. Unopposed plasmin not only degrades fibrin in the occlusive thrombus but also induces a systemic lytic state.

When given systemically to patients with acute MI, streptokinase reduces mortality. For this indication, the drug is usually given as an IV infusion of 1.5 million units over 30–60 min. Patients who receive streptokinase can develop antibodies against the drug, as can patients with prior streptococcal infection. These antibodies can reduce the effectiveness of streptokinase.

Allergic reactions occur in ~5% of patients treated with streptokinase. These may manifest as a rash, fever, chills, and rigors. Although anaphylactic reactions can occur, these are rare. Transient hypotension is common with streptokinase and has been attributed to plasmin-mediated release of bradykinin from kininogen. The hypotension usually responds to leg elevation and administration of IV fluids and low doses of vasopressors, such as dopamine or norepinephrine.

ANISTREPLASE

To generate this drug, streptokinase is combined with equimolar amounts of Lys-plasminogen, a plasmin-cleaved form of plasminogen with a Lys residue at its N terminal. The active site of Lys-plasminogen that is exposed upon combination with streptokinase is then masked with an anisoyl group. After IV infusion, the anisoyl group is slowly removed by deacylation, giving the complex a half-life of ~100 min. This allows drug administration via a single bolus infusion.

Although it is more convenient to administer, anistreplase offers few mechanistic advantages over streptokinase. Like streptokinase, anistreplase does not distinguish between fibrin-bound and circulating plasminogen. Consequently, it too produces a systemic lytic state. Likewise, allergic reactions and hypotension are just as frequent with anistreplase as they are with streptokinase.

When anistreplase was compared with alteplase in patients with acute MI, reperfusion was obtained more rapidly with alteplase than with anistreplase. Improved reperfusion was associated with a trend toward better clinical outcomes and reduced mortality rate with alteplase. These results and the high cost of anistreplase dampened the enthusiasm for its use.

UROKINASE

Urokinase is a two-chain serine protease derived from cultured fetal kidney cells with a molecular weight of 34,000. Urokinase converts plasminogen to plasmin directly by cleaving the Arg560-Val561 bond. Unlike streptokinase, urokinase is not immunogenic, and allergic reactions are rare. Urokinase produces a systemic lytic state because it does not discriminate between fibrin-bound and circulating plasminogen.

Despite many years of use, urokinase has never been systemically evaluated for coronary thrombolysis. Instead, urokinase is often employed for catheter-directed lysis of thrombi in the deep veins or the peripheral arteries. Because of production problems, urokinase is no longer available.

ALTEPLASE

A recombinant form of single-chain tPA, alteplase has a molecular weight of 68,000. Alteplase is rapidly converted into its two-chain form by plasmin. Although single- and two-chain forms of tPA have equivalent activity in the presence of fibrin, in its absence, single-chain tPA has 10-fold lower activity.

Alteplase consists of five discrete domains **(Fig. 123-9)**; the N-terminal A chain of two-chain alteplase contains four of these domains. Residues 4 through 50 make up the finger domain, a region that resembles the finger domain of fibronectin; residues 50 through 87 are homologous with epidermal growth factor, whereas residues 92 through 173 and 180 through 261, which have homology to the kringle domains of plasminogen, are designated as the first and second kringle, respectively. The fifth alteplase domain is the protease domain; it is located on the C-terminal B chain of two-chain alteplase.

The interaction of alteplase with fibrin is mediated by the finger domain and, to a lesser extent, by the second kringle domain. The affinity of alteplase for fibrin is considerably higher than that for fibrinogen. Consequently, the catalytic efficiency of plasminogen activation by alteplase is two to three orders of magnitude higher in the presence of fibrin than in the presence of fibrinogen. This phenomenon helps to localize plasmin generation to the fibrin surface.

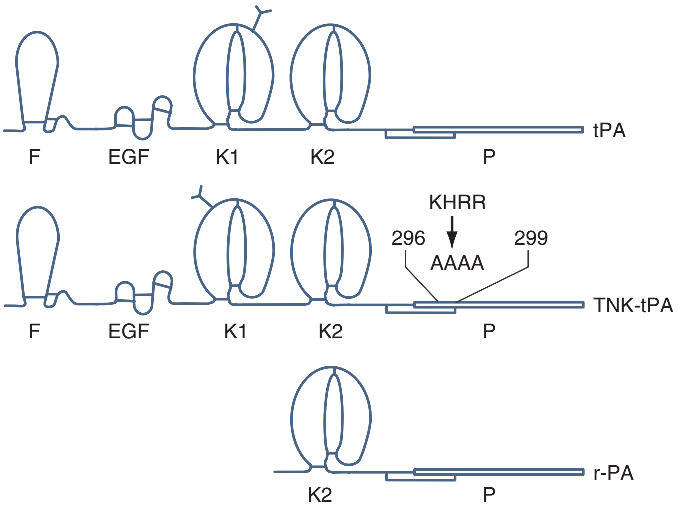

FIGURE 123-9 Domain structures of alteplase (tPA), tenecteplase (TNK-tPA), and reteplase (r-PA). The finger (F), epidermal growth factor (EGF), first and second kringles (K1 and K2, respectively), and protease (P) domains are illustrated. The glycosylation site (Y) on K1 has been repositioned in tenecteplase to endow it with a longer half-life. In addition, a tetra-alanine substitution in the protease domain renders tenecteplase resistant to type 1 plasminogen activator inhibitor (PAI-1) inhibition. Reteplase is a truncated variant that lacks the F, EGF, and K1 domains.

Although alteplase preferentially activates plasminogen in the presence of fibrin, alteplase is not as fibrin selective as was first predicted. Its fibrin specificity is limited because, like fibrin, (DD)E, the major soluble degradation product of cross-linked fibrin, binds alteplase and plasminogen with high affinity. Consequently, (DD)E is as potent as fibrin as a stimulator of plasminogen activation by alteplase. Whereas plasmin generated on the fibrin surface results in thrombolysis, plasmin generated on the surface of circulating (DD)E degrades fibrinogen. Fibrinogen degradation results in the accumulation of fragment X, a high-molecular-weight clottable fibrinogen degradation product. Incorporation of fragment X into hemostatic plugs formed at sites of vascular injury renders them susceptible to lysis. This phenomenon may contribute to alteplase-induced bleeding.

A trial comparing alteplase with streptokinase for treatment of patients with acute MI demonstrated significantly lower mortality with alteplase than with streptokinase, although the absolute difference was small. The greatest benefit was seen in patients age <75 years with anterior MI who presented <6 h after symptom onset.

For treatment of acute MI or acute ischemic stroke, alteplase is given as an IV infusion over 60–90 min. The total dose of alteplase usually ranges from 90 to 100 mg. Allergic reactions and hypotension are rare, and alteplase is not immunogenic.

TENECTEPLASE

Tenecteplase is a genetically engineered variant of tPA and was designed to have a longer half-life than tPA and to be resistant to inactivation by PAI-1. To prolong its half-life, a new glycosylation site was added to the first kringle domain (Fig. 123-9). Because addition of this extra carbohydrate side chain reduced fibrin affinity, the existing glycosylation site on the first kringle domain was removed. To render the molecule resistant to inhibition by PAI-1, a tetra-alanine substitution was introduced at residues 296–299 in the protease domain, the region responsible for the interaction of tPA with PAI-1.

Tenecteplase is more fibrin-specific than tPA. Although both agents bind to fibrin with similar affinity, the affinity of tenecteplase for (DD)E is significantly lower than that of tPA. Consequently, (DD)E does not stimulate systemic plasminogen activation by tenecteplase to the same extent as tPA. As a result, tenecteplase produces less fibrinogen degradation than tPA.

For coronary thrombolysis, tenecteplase is given as a single IV bolus. In a large phase III trial that enrolled >16,000 patients, the 30-day mortality rate with single-bolus tenecteplase was like that with accelerated-dose tPA. Although rates of intracranial hemorrhage were similar with both treatments, patients given tenecteplase had fewer noncerebral bleeds and a reduced need for blood transfusions than those treated with tPA. The improved safety profile of Tenecteplase likely reflects its enhanced fibrin specificity.

RETEPLASE

Reteplase is a single-chain, recombinant tPA derivative that lacks the finger, epidermal growth factor, and first kringle domains (Fig. 123-9). This truncated derivative has a molecular weight of 39,000. The affinity of reteplase for fibrin is lower than that of tPA likely because reteplase lacks the finger domain. Because it is produced in *Escherichia coli*, reteplase is not glycosylated. This endows it with a plasma half-life longer than that of tPA. Consequently, reteplase is given as two IV boluses, which are separated by 30 min. Clinical trials have demonstrated that reteplase is at least as effective as streptokinase for treatment of acute MI, but the agent is not superior to tPA.

CONCLUSIONS AND FUTURE DIRECTIONS

Thrombosis involves a complex interplay among the vessel wall, platelets, the coagulation system, and the fibrinolytic pathways. Activation of coagulation also triggers inflammatory pathways that may exacerbate thrombosis. A better understanding of the biochemistry of blood coagulation and advances in structure-based drug design have identified new targets and resulted in the development of novel antithrombotic drugs. Well-designed clinical trials have provided detailed information on which drugs to use and when to use them. Despite these advances, however, thromboembolic disorders remain a major cause of morbidity and mortality. Therefore, the search for better and safer targets continues.

FURTHER READING

Arepally GM, Padmanabhan A: Heparin-induced thrombocytopenia: A focus on thrombosis. Arterioscler Thromb Vasc Biol 41:141, 2021.

Berger JS: Aspirin for primary prevention—time to rethink our approach. JAMA Netw Open 5:e2210144, 2022.

Fei Y et al: Efficacy and safety of newer P2Y$_{12}$ inhibitors for acute coronary syndrome: A network meta-analysis. Sci Rep 10:16794, 2020.

Greinacher A et al: Thrombotic thrombocytopenia after ChAdOx1 nCov-19 vaccination. N Engl J Med 384:2092, 2021.

Hao C et al: Low molecular weight heparins and their clinical applications. Prog Mol Biol Transl Sci 163:21, 2019.

Phipps MS, Cronin CA: Management of acute ischemic stroke. BMJ 368:l6983, 2020.

Prince M, Wenham T: Heparin-induced thrombocytopenia. Postgrad Med J 94:453, 2018.

Rivera-Caravaca JM et al: Treatment strategies for patients with atrial fibrillation and anticoagulant-associated intracranial hemorrhage: An overview of the pharmacotherapy. Expert Opin Pharmacother 21:1867, 2020.

Satoh K et al: Recent advances in the understanding of thrombosis. Arterioscler Thromb Vasc Biol 39:e159, 2019.

Scully M et al: Pathologic antibodies to platelet factor 4 after ChAdOx1 nCoV-19 vaccination. N Engl J Med 384:2202, 2021.

Steffel J et al: The COMPASS Trial: Net clinical benefit of low-dose rivaroxaban plus aspirin as compared with aspirin in patients with chronic vascular disease. Circulation 142:40, 2020.

White K et al: New agents for DOAC reversal: a practical management review. Br J Cardiol 29:1, 2022.

Section 1 Basic Considerations in Infectious Diseases

124 Approach to the Patient with an Infectious Disease

Neeraj K. Surana, Dennis L. Kasper

HISTORIC PERSPECTIVE

The origins of the field of infectious diseases are humble. The notion that communicable diseases were due to a *miasma* ("bad air") can be traced back to at least the mid-sixteenth century. Not until the work of Louis Pasteur and Robert Koch in the late nineteenth century was there credible evidence supporting the germ theory of disease—i.e., that microorganisms are the direct cause of infections. In contrast to this relatively slow start, the twentieth century saw remarkable advances in the field of infectious diseases, and the etiologic agents of numerous infectious diseases were soon identified. Furthermore, the discovery of antibiotics and the advent of vaccines against some of the most deadly and debilitating infections greatly altered the landscape of human health. Indeed, the twentieth century saw the elimination of smallpox, one of the great scourges in the history of humanity. These remarkable successes prompted Sir Frank MacFarlane Burnet, a noted immunologist and Nobel laureate, to write in a 1962 publication entitled *Natural History of Infectious Diseases*: "In many ways one can think of the middle of the twentieth century as the end of one of the most important social revolutions in history, the virtual elimination of infectious disease." Professor Burnet was not alone in this view. Robert Petersdorf, a renowned infectious disease expert and former editor of this textbook, wrote in 1978 that "even with my great personal loyalties to infectious diseases, I cannot conceive a need for 309 more [graduating trainees in infectious diseases] unless they spend their time culturing each other." Given the enormous growth of interest in the microbiome in the past 20 years, Dr. Petersdorf's statement might have been ironically clairvoyant, although he could have had no idea what was in store for humanity, with an onslaught of new, emerging, and reemerging infectious diseases.

Clearly, even with all the advances of the twentieth century, infectious diseases continue to represent a formidable challenge for patients and physicians alike. Furthermore, during the latter half of the century, several chronic diseases were demonstrated to be directly or indirectly caused by infectious microbes; perhaps the most notable examples are the associations of *Helicobacter pylori* with peptic ulcer disease and gastric carcinoma, human papillomavirus with cervical cancer, and hepatitis B and C viruses with liver cancer. In fact, ~16% of all malignancies are now known to be associated with an infectious cause. In addition, numerous emerging and reemerging infectious diseases continue to have a dire impact on global health: HIV/AIDS, SARS-CoV-2, Ebola, and mpox are but a few examples. The fear of weaponizing pathogens for bioterrorism is ever present and poses a potentially enormous threat to public health. Moreover, escalating antimicrobial resistance in clinically relevant microbes (e.g., carbapenem-resistant Enterobacteriaceae and *Acinetobacter* spp., *Candida auris*, drug-resistant *Mycobacterium tuberculosis*, and vancomycin-resistant enterococci) signifies that the administration of antimicrobial agents—once thought to be a panacea—requires appropriate stewardship. For all these reasons, infectious diseases continue to exert grim effects on individual patients as well as on international public health. Even with all the successes of the past century, physicians must be as thoughtful about infectious diseases now as they were at the beginning of the twentieth century.

GLOBAL CONSIDERATIONS

Infectious diseases remain the second leading cause of death worldwide. Although the rate of infectious disease–related deaths has decreased dramatically over the past 25 years, there were still 9.6 million such deaths in 2019 **(Fig. 124-1*A*)**. These deaths disproportionately affect children <1 year of age, adults older than 70 years, and persons living in low- and middle-income countries **(Fig. 124-1*B* and 124-1*C*; Chap. 487)**; in 2019, ~17% of all deaths worldwide were related to infectious diseases, with a rate as high as ~69% in sub-Saharan Africa.

Given that infectious diseases are still a major cause of global mortality, understanding the local epidemiology of disease is critically important in evaluating patients. Diseases such as HIV/AIDS have decimated southern Africa, with HIV-infected adults representing 16–20% of the total population in countries like South Africa, Botswana, and Lesotho, and more than 25% in Eswatini. Moreover, drug-resistant tuberculosis is rampant throughout the former Soviet-bloc countries, India, China, and South Africa. The ready availability of this type of information allows physicians to develop appropriate differential diagnoses and treatment plans for individual patients. Programs such as the Global Burden of Disease seek to quantify human losses (e.g., deaths, disability-adjusted life-years) due to diseases by age, sex, and country over time; these data not only help inform local, national, and international health policy but can also help guide local medical decision-making.

Even though some diseases (e.g., pandemic influenza, mpox) are seemingly geographically restricted, the increasing ease of rapid worldwide travel has raised concern about their swift spread around the globe. Indeed, human migration has historically been the source of epidemics: *Yersinia pestis* spread along trade routes in the fourteenth century, Native American populations were devastated by diseases such as smallpox and *Salmonella* that were imported by European explorers in the fifteenth and sixteenth centuries, military maneuvers helped facilitate the spread of the 1918 influenza pandemic, and religious pilgrimages (e.g., the Hajj) provide the means for worldwide dissemination of diseases. The continued effects of global travel on the spread of infectious diseases are perhaps best highlighted by the SARS-CoV-2 pandemic **(Chap. 204)**. Although this virus was first identified in Wuhan, China, it quickly spread across the globe and brought an abrupt end to virtually all travel and commerce throughout the world, plunging economies into a deep recession, resulting at one point in more than half the world's population living under stay-at-home orders, and causing the death of ~7 million people worldwide. Not only can travelers carry person-to-person transmitted infections (e.g., SARS-CoV-2, HIV) anywhere in the world, but they can also introduce vector-borne infections to new geographic areas (e.g., chikungunya and Zika viruses) and contribute to the worldwide spread of multidrug-resistant organisms. The world's increasing interconnectedness has profound implications not only for the global economy but also for medicine and the spread of infectious diseases.

UNDERSTANDING THE MICROBIOTA

Normal, healthy humans are colonized with ~40 trillion bacteria as well as countless viruses, fungi, and archaea; taken together, these microorganisms outnumber human cells by ~10 times in the human body **(Chap. 484)**. The major reservoir of these microbes is the gastrointestinal tract, but substantial numbers of microbes live in the female genital tract, the oral cavity, and the nasopharynx. There is increasing interest in the skin and lungs as sites where microbial colonization might be highly relevant to the biology and disease susceptibility of the host. These commensal organisms provide the host with myriad benefits, from aiding in metabolism to shaping the immune system. With regard to infectious diseases, the vast majority of infections are caused by organisms that are part of the normal microbiota (e.g., *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Pseudomonas aeruginosa*), with relatively few infections due to organisms that are strictly pathogens (e.g., *Neisseria gonorrhoeae*, rabies virus). Perhaps it is not surprising that a general understanding

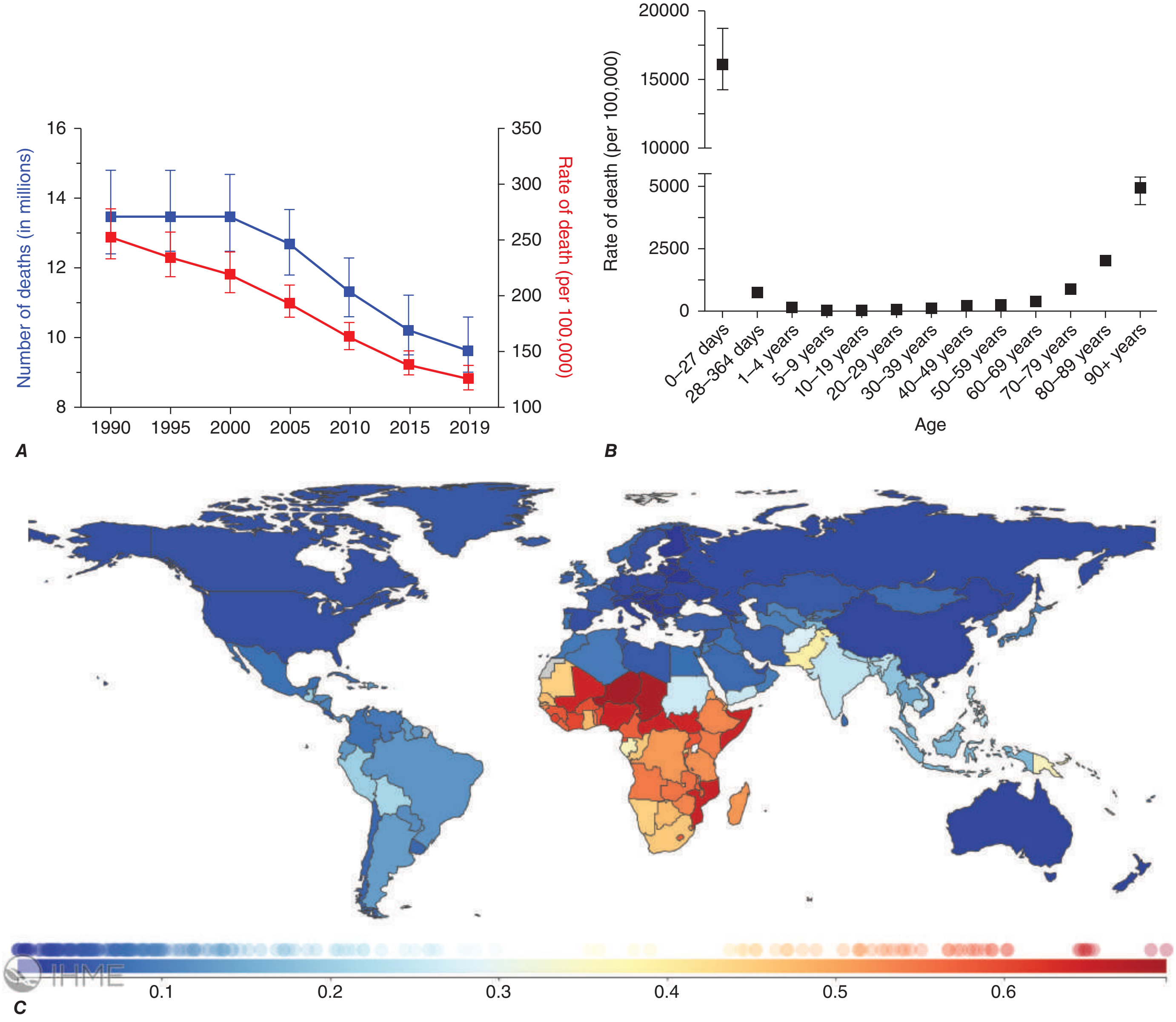

FIGURE 124-1 Magnitude of infectious disease–related deaths globally. ***A.*** The absolute number (*blue line; left axis*) and rate (*red line; right axis*) of infectious disease–related deaths throughout the world since 1990. ***B.*** Age-specific rates of infectious disease–related deaths in 2019. In both ***A*** and ***B***, the charts depict the mean estimate and 95% uncertainty intervals. ***C.*** A map depicting country-specific data for the percent of total deaths that were attributable to communicable, maternal, neonatal, and nutritional disorders in 2019. *(Source: Institute for Health Metrics and Evaluation. Used with permission. All rights reserved.)*

of the microbiota is essential in the evaluation of infectious diseases. Individuals' microbiotas have a major impact on their susceptibility to infectious diseases and even their responses to vaccines. Site-specific knowledge of the indigenous microbiota may facilitate appropriate interpretation of culture results, aid in selection of empirical antimicrobial therapy based on the likely causative agents, and provide additional impetus for rational antibiotic use to minimize the untoward effects of these drugs on the "beneficial" microbes that inhabit the body.

■ WHEN TO CONSIDER AN INFECTIOUS ETIOLOGY

The title of this chapter may appear to presuppose that the physician knows when a patient has an infectious disease. In reality, this chapter can serve only as a guide to the evaluation of a patient in whom an infectious disease is a possibility. Once a specific diagnosis is made, the reader should consult the subsequent chapters that deal with specific microorganisms in detail. The challenge for the physician is to recognize which patients may have an infectious disease as opposed to some other underlying disorder. This task is greatly complicated by the fact that infections have an infinite range of presentations, from acute life-threatening conditions (e.g., meningococcemia) to chronic diseases of varying severity (e.g., *H. pylori*–associated peptic ulcer disease) to no symptoms at all (e.g., latent *M. tuberculosis* infection). While it is impossible to generalize about a presentation that encompasses all infections, common findings in the history, physical examination, and basic laboratory testing often suggest that the patient either has an infectious disease or should be more closely evaluated for one. This chapter focuses on these common findings and how they may direct the ongoing evaluation of the patient.

APPROACH TO THE PATIENT

Infectious Disease

See also Chap. 127.

HISTORY

As in all of medicine, a complete and thorough history is paramount in the evaluation of a patient with a possible infectious disease. The history is critical for developing a focused differential diagnosis and for guiding the physical exam and initial diagnostic testing. Although a detailing of all the elements of a history is beyond the scope of this chapter, specific components relevant to infectious diseases require particular attention. In general, these aspects focus on two areas: (1) an exposure history that may identify microorganisms with which the patient may have come into contact and (2) host-specific factors that may predispose to the development of an infection.

Exposure History • **History of infections or exposure to drug-resistant microbes** Information about a patient's previous infections, with the associated microbial susceptibility profiles, is very helpful in determining possible etiologic agents. Specifically, knowing whether a patient has a history of infection with drug-resistant organisms (e.g., methicillin-resistant *S. aureus,* vancomycin-resistant *Enterococcus* species, enteric organisms that produce an extended-spectrum β-lactamase or carbapenemase) or may have been exposed to drug-resistant microbes (e.g., during a recent stay in a hospital, nursing home, or long-term acute-care facility) may alter the choice of empirical antibiotics. For example, a patient presenting with sepsis who is known to have a history of invasive infection with a multidrug-resistant isolate of *P. aeruginosa* should be treated empirically with an antimicrobial regimen that will cover this strain.

Social history Although the social history taken by physicians is often limited to inquiries about a patient's alcohol and tobacco use, a complete social history can offer a number of clues to the underlying diagnosis. Knowing whether the patient has any high-risk behaviors (e.g., unsafe sexual behaviors, intravenous [IV] drug use), potential hobby-associated exposures (e.g., avid gardening, with possible *Sporothrix schenckii* exposure), or occupational exposures (e.g., increased risk for *M. tuberculosis* exposure in funeral service workers) can facilitate diagnosis. The importance of the social history is exemplified by a case in 2009 in which a laboratory researcher died of a *Y. pestis* infection acquired during his work; although this patient had visited both an outpatient clinic and an emergency department, his records at both sites failed to include his occupation—information that potentially could have led quickly to appropriate treatment and infection control measures.

Dietary habits Because certain pathogens are associated with specific dietary habits, inquiring about a patient's diet can provide insight into possible exposures. For example, Shiga toxin–producing strains of *Escherichia coli*, and *Toxoplasma gondii* are associated with the consumption of raw or undercooked meat; *Salmonella typhimurium*, *Listeria monocytogenes*, and *Mycobacterium bovis* with unpasteurized milk; *Leptospira* species, parasites, and enteric bacteria with unpurified water; and *Vibrio* species, norovirus, helminths, and protozoa with raw seafood.

Animal exposures Because animals are often important vectors of infectious diseases, patients should be asked about exposures to any animals, including contact with their own pets, visits to petting zoos, or random encounters (e.g., home rodent infestation). For example, dogs can carry ticks that serve as agents for the transmission of several infectious diseases, including Lyme disease, Rocky Mountain spotted fever, and ehrlichiosis. Cats are associated with *Bartonella henselae* infection, reptiles with *Salmonella* infection, rodents with leptospirosis, and rabbits with tularemia **(Chap. 146)**.

Travel history Attention should be paid to both international and domestic travel. Fever in a patient who has recently returned from abroad significantly broadens the differential diagnosis **(Chap. 130)** and, as exemplified by the COVID-19 pandemic, can help identify the beginnings of international outbreaks. Even a remote history of international travel may reflect patients' exposure to infections with pathogens such as *M. tuberculosis* or *Strongyloides stercoralis*. Similarly, domestic travel may have exposed patients to pathogens that are not normally found in their local environment and therefore may not routinely be considered in the differential diagnosis. For example, a patient who has recently visited California or Martha's Vineyard may have been exposed to *Coccidioides immitis* or *Francisella tularensis*, respectively. Beyond simply identifying locations that a patient may have visited, the physician needs to delve deeper to learn what kinds of activities and behaviors the patient engaged in during travel (e.g., the types of food and sources of water consumed, freshwater swimming, animal exposures) and whether the patient had the necessary immunizations and/or took the necessary prophylactic medications prior to travel; these additional exposures, which the patient may not think to report without specific prompting, are as important as exposures during a patient's routine daily living.

Host-Specific Factors Because many opportunistic infections (e.g., with *Pneumocystis jirovecii*, *Aspergillus* species, or JC virus) affect primarily immunocompromised patients, it is of vital importance to determine the immune status of the patient. Defects in the immune system may be due to an underlying disease (e.g., malignancy, HIV infection, malnutrition), a medication (e.g., chemotherapy, glucocorticoids, monoclonal antibodies to components of the immune system), a treatment modality (e.g., total body irradiation, splenectomy), or a primary immunodeficiency. The type of infection for which the patient is at increased risk varies with the specific type of immune defect. In concert with determining whether a patient is immunocompromised for any reason, the physician should review the immunization record to ensure that the patient is adequately protected against vaccine-preventable diseases **(Chap. 129)**.

PHYSICAL EXAMINATION

Like the history, a thorough physical examination is crucial in evaluating patients with an infectious disease. Some elements of the physical exam (e.g., skin, lymphatics) that are often performed in a cursory manner as a result of the ever-increasing pace of medical practice may help identify the underlying diagnosis. Moreover, serial exams are critical since new findings may appear as the illness progresses. A description of all the elements of a physical exam is beyond the scope of this chapter, but the following components have particular relevance to infectious diseases.

Vital Signs Given that elevations in temperature are often a hallmark of infection, paying close attention to the temperature may be of value in diagnosing an infectious disease **(Chap. 20)**. The idea that 37°C (98.6°F) is the normal human body temperature dates to the nineteenth century and was initially based on axillary measurements. Rectal temperatures more accurately reflect the core body temperature and are 0.4°C (0.7°F) and 0.8°C (1.4°F) higher than oral and axillary temperatures, respectively. This idea of a "normal" body temperature does not consider the fact that temperatures tend to be higher later in the day, in women, and in younger people. Moreover, the average body temperature seems to have dropped ~0.03°C every decade since the early 1800s to a new normal of ~36.7°C. Although the definition of fever varies greatly throughout the medical literature, the most common definition, which is based on studies defining fever of unknown origin **(Chap. 22)**, uses a core temperature ≥38.3°C (≥101°F). Although fever is very commonly associated with infection, it is also documented in many other diseases **(Chap. 20)**. For every 1°C (1.8°F) increase in core temperature, the heart rate typically rises by ~10 beats/min. **Table 124-1** lists infections that are associated with relative bradycardia (*Faget's sign*), where patients have a lower heart rate than might be expected for a given body temperature. Although this pulse–temperature dissociation is not highly sensitive or specific for establishing a diagnosis, it is potentially useful in low-resource settings given its ready availability and simplicity.

Lymphatics There are ~600 lymph nodes throughout the body, and infections are an important cause of lymphadenopathy. A physical examination should include evaluation of lymph nodes in multiple regions (e.g., popliteal, inguinal, epitrochlear, axillary, multiple cervical regions), with notation of the location, size (normal, <1 cm), presence or absence of tenderness, and consistency (soft, firm, or rubbery) and of whether the nodes are matted (i.e., connected and moving together). Nodes that are small and firm can also be described as "shotty," referring to the size and consistency of buckshot pellets. Of note, palpable epitrochlear nodes are always

TABLE 124-1 Causes of Relative Bradycardia

Infectious Causes	
Intracellular organisms	
Gram-negative bacteria	*Salmonella typhi* *Francisella tularensis* *Brucella* spp. *Coxiella burnetii* (Q fever) *Leptospira interrogans* *Legionella pneumophila* *Mycoplasma pneumoniae*
Tick-borne organisms	*Rickettsia* spp. *Orientia tsutsugamushi* (scrub typhus) *Babesia* spp.
Other	*Corynebacterium diphtheriae* *Plasmodium* spp. (malaria)
Viruses/viral infections	Yellow fever virus Dengue virus Viral hemorrhagic fevers[a] Viral myocarditis
Noninfectious Causes	
	Drug fever Beta blocker use Central nervous system lesions Malignant lymphoma Factitious fever

[a] Primarily early in the course of infection with Marburg or Ebola virus.

pathologic. Of patients presenting with lymphadenopathy, 75% have localized findings, and the remaining 25% have generalized lymphadenopathy (i.e., that involving more than one anatomic region). Localized lymphadenopathy in the head and neck region is found in 55% of patients, inguinal lymphadenopathy in 14%, and axillary lymphadenopathy in 5%. Determining whether the patient has generalized versus localized lymphadenopathy can help narrow the differential diagnosis, as various infections present differently.

Skin The fact that many infections have cutaneous manifestations gives the skin examination particular importance in the evaluation of patients **(Chaps. 21, 61, 134, and A1)**. It is important to perform a complete skin exam, with attention to both front and back. Specific rashes are often extremely helpful in narrowing the differential diagnosis of an infection **(Chaps. 21 and A1)**. In numerous anecdotal instances, patients in the intensive care unit have had "fever of unknown origin" that was actually due to unrecognized pressure ulcers. Moreover, close examination of the distal extremities for splinter hemorrhages, Janeway lesions, or Osler's nodes may yield evidence of endocarditis or other causes of septic emboli.

Foreign Bodies As previously mentioned, many infections are caused by members of the indigenous microbiota. These infections typically occur when these microbes escape their normal habitat and enter a new one. Thus, maintenance of epithelial barriers is one of the most important mechanisms in protection against infection. However, hospitalization of patients is often associated with breaches of these barriers—e.g., due to placement of IV lines, surgical drains, or tubes (e.g., endotracheal tubes and Foley catheters) that allow microorganisms to localize in sites to which they normally would not have access **(Chap. 147)**. Accordingly, knowing what lines, tubes, and drains are in place is helpful in ascertaining what body sites might be infected.

DIAGNOSTIC TESTING

Laboratory and radiologic testing has advanced greatly over the past few decades and has become an important component in the evaluation of patients. The dramatic increase in the number of serologic diagnostics, antigen tests, and molecular diagnostics available to the physician has, in fact, revolutionized medical care. However, all of these tests should be viewed as adjuncts to the history and physical examination—not a replacement for them. The selection of initial tests should be based directly on the patient's history and physical exam findings. Moreover, diagnostic testing should generally be limited to those conditions that are reasonably likely and treatable, important in terms of public health considerations, and/or capable of providing a definitive diagnosis that will consequently limit other testing.

White Blood Cell (WBC) Count Elevations in the WBC count are often associated with infection, although many viral infections are associated with leukopenia. It is important to assess the WBC differential, given that different classes of microbes are associated with various leukocyte types. For example, bacteria are associated with an increase in polymorphonuclear neutrophils, often with elevated levels of earlier developmental forms such as bands; viruses are associated with an increase in lymphocytes; and certain parasites are associated with an increase in eosinophils. **Table 124-2** lists the major infectious causes of eosinophilia.

Inflammatory Markers The erythrocyte sedimentation rate (ESR) and the C-reactive protein (CRP) level are indirect and direct measures of the acute-phase response, respectively, that can be used to assess a patient's general level of inflammation. Moreover, these markers can be followed serially over time to monitor disease progress/resolution. It is noteworthy that the ESR changes relatively slowly, and its measurement more often than weekly usually is not useful; in contrast, CRP concentrations change rapidly, and daily measurements can be useful in the appropriate context. Although these markers are sensitive indicators of inflammation, neither is very specific. An extremely elevated ESR (>100 mm/h) has a 90% predictive value for a serious underlying disease **(Table 124-3)**. Work is ongoing to identify other potentially useful inflammatory markers (e.g., procalcitonin, serum amyloid A protein); their clinical utility requires further validation.

Analysis of Cerebrospinal Fluid (CSF) Assessment of CSF is critical for patients with suspected meningitis or encephalitis. An opening pressure should always be recorded, and fluid should routinely be sent for cell counts, Gram's stain and culture, and determination of glucose and protein levels. A CSF Gram's stain typically requires $>10^5$ bacteria/mL for reliable positivity; its specificity approaches 100%. **Table 124-4** lists the typical CSF profiles for various infections. In general, CSF with lymphocytic pleocytosis and a low glucose concentration suggests either infection (e.g., with *Listeria*, *M. tuberculosis*, or a fungus) or a noninfectious disorder (e.g., neoplastic meningitis, sarcoidosis). Bacterial antigen tests of CSF (e.g., latex agglutination tests for *Haemophilus influenzae* type b, group B *Streptococcus*, *S. pneumoniae*, and *Neisseria meningitidis*) are not recommended for screening, given that these tests are no more sensitive than Gram's stain; however, these assays can be helpful in presumptively identifying organisms seen on Gram's stain. In contrast, other antigen tests (e.g., for *Cryptococcus*) and some CSF serologic testing (e.g., for *Treponema pallidum*, *Coccidioides*) are highly sensitive and are useful for select patients. In addition, polymerase chain reaction (PCR) analysis of CSF is increasingly being used for the diagnosis of bacterial (e.g., *N. meningitidis*, *S. pneumoniae*, mycobacteria) and viral (e.g., herpes simplex virus, enterovirus) infections; while these molecular tests permit rapid diagnosis with a high degree of sensitivity and specificity, they often do not allow determination of antimicrobial resistance profiles.

Cultures The mainstays of infectious disease diagnosis include the culture of infected tissue (e.g., surgical specimens) or fluid (e.g., blood, urine, sputum, pus from a wound). Samples can be sent for culture of bacteria (aerobic or anaerobic), fungi, or viruses. Ideally, specimens are collected before the administration of antimicrobial therapy; in instances where this order of events is not clinically

TABLE 124-2 Major Infectious Causes of Eosinophilia[a]

ORGAN INVOLVED	ORGANISM	EXPOSURE	GEOGRAPHIC DISTRIBUTION	DEGREE OF EOSINOPHILIA[b]
Central nervous system	*Angiostrongylus*	Raw seafood	Asia	Mild
	Gnathostoma	Raw poultry and seafood	Asia	Moderate to extreme
Eye	*Loa loa*	Insect bite	Africa	Moderate (expatriates), mild (patients living in endemic areas)
	Onchocerca	Insect bite	Africa	Mild (expatriates), moderate (patients living in endemic areas)
Lung	*Chlamydia trachomatis*	Sexual transmission	Worldwide	Mild
	Strongyloides	Soil	Tropical	Moderate (acute), mild (chronic)
	Toxocara canis/Toxocara cati[c]	Dogs, soil	Worldwide	Moderate to extreme
	Paragonimus	Crabs and crayfish	Asia	Moderate (acute), mild (chronic)
	Coccidioides immitis	Soil	Southwestern United States	Mild (acute), extreme (disseminated)
	Brugia malayi	Insect bite	Asia	Mild to moderate
	Pneumocystis jirovecii	Air	Worldwide	Mild
Liver	*Schistosoma japonicum*	Freshwater swimming	Asia	Moderate (acute), mild (chronic)
	Schistosoma mansoni	Freshwater swimming	Africa, Middle East, Latin America	Moderate (acute), mild (chronic)
	Fasciola	Watercress	Worldwide	Moderate
	Clonorchis	Raw seafood	Asia	Mild to moderate
	Opisthorchis	Raw seafood	Asia	Mild to moderate
Intestines	*Ascaris*[d]	Raw fruits and vegetables, contaminated water	Worldwide	Mild to extreme
	Hookworm	Soil	Worldwide	Mild to moderate
	Trichuris	Raw fruits and vegetables, contaminated water	Tropical	Mild
	Cystoisospora belli	Contaminated water and food	Worldwide	Mild
	Dientamoeba fragilis	Unclear; spread via fecal–oral route	Worldwide	Mild
	Capillaria	Raw seafood	Asia	Extreme
	Heterophyes	Raw seafood	Asia, Middle East	Mild
	Anisakis	Raw seafood	Worldwide	Mild
	Baylisascaris procyonis[e]	Soil	North America	Moderate to extreme
	Hymenolepis nana	Contaminated water, soil	Worldwide	Mild
Bladder	*Schistosoma haematobium*	Freshwater swimming	Africa, Middle East	Moderate (acute), mild (chronic)
Muscle	*Trichinella*	Pork	Worldwide	Moderate to extreme
Lymphatics	*Wuchereria bancrofti*[d]	Insect bite	Tropical	Moderate to extreme[f]
	Bartonella henselae	Cats	Worldwide	Mild
Other	Recovery from bacterial or viral infections	—	—	Mild
	HIV	Contaminated bodily fluid	Worldwide	Mild
	Cryptococcus neoformans	Soil	Worldwide	Moderate to extreme (disseminated)

[a] There are numerous noninfectious causes of eosinophilia, such as atopic disease, DRESS (drug reaction with eosinophilia and systemic symptoms) syndrome, and pernicious anemia, which can cause mild eosinophilia; drug hypersensitivity and serum sickness, which can cause mild to moderate eosinophilia; collagen vascular disease, which can cause moderate eosinophilia; and malignancy, Churg-Strauss syndrome, and hyper-IgE syndromes, which can cause moderate to extreme eosinophilia. [b] Mild: 500–1500 cells/μL; moderate: 1500–5000 cells/μL; extreme: >5000 cells/μL. [c] Can also affect the liver and the eyes. [d] Can also affect the lungs. [e] Can also affect the eyes and the central nervous system. [f] Levels are typically higher with pulmonary infections.

feasible, microscopic examination of the specimen (e.g., Gram-stained or potassium hydroxide [KOH]–treated preparations) is particularly important. Culture of the organism(s) allows identification of the etiologic agent(s), determination of the antimicrobial susceptibility profile, and—when there is concern about an outbreak—isolate typing. While cultures are extremely useful in the evaluation of patients, determining whether culture results are clinically meaningful or represent contamination (e.g., a non-*aureus*, non-*lugdunensis* staphylococcal species growing in a blood culture) can sometimes be challenging and requires an understanding of the patient's immune status, exposure history, and microbiota. In some cases, serial cultures to demonstrate clearance of the organism may be helpful.

Pathogen-Specific Testing Numerous pathogen-specific tests (e.g., serology, antigen testing, PCR testing) are commercially available, and many hospitals now offer some of these tests in-house to facilitate rapid turnaround that ultimately enhances patient care. The reader is directed to relevant chapters on the pathogens of interest for specific details. Some of these tests (e.g., universal PCRs, shotgun metagenomic sequencing) identify organisms that currently are not easily cultivable and have unclear relationships to disease, thereby complicating diagnosis. As these tests become more commonplace, the relevance of some of these previously unrecognized bacteria to human health will likely become more apparent.

Radiology Imaging provides an important adjunct to the physical examination, allowing evaluation for lymphadenopathy in regions that are not externally accessible (e.g., mediastinum, intraabdominal sites), assessment of internal organs for evidence of infection, and facilitation of image-guided percutaneous sampling of deep spaces. The choice of imaging modality (e.g., CT, MRI, ultrasound,

TABLE 124-3 Causes of an Extremely Elevated Erythrocyte Sedimentation Rate (>100 mm/h)

ETIOLOGIC CATEGORY (% OF CASES)	SPECIFIC CAUSES
Infectious diseases (35–40)	Subacute bacterial endocarditis Abscesses Osteomyelitis Tuberculosis Urinary tract infection
Inflammatory diseases (15–20)	Giant cell arteritis Rheumatoid arthritis Systemic lupus erythematosus
Malignancies (15–20)	Multiple myeloma Leukemias Lymphomas Carcinomas
Other (20–35)	Drug hypersensitivity reactions (drug fever) Ischemic tissue injury/trauma Renal diseases

nuclear medicine, use of contrast) is best made in consultation with a radiologist to ensure that the results will address the physician's specific concerns.

TREATMENT

Physicians often must balance the need for empirical antibiotic treatment with the patient's clinical condition. When clinically feasible, it is best to obtain relevant samples (e.g., blood, CSF, tissue, purulent exudate) for culture prior to the administration of antibiotics, as antibiotic treatment often makes subsequent diagnosis more difficult. Although a general maxim for antibiotic treatment is to use a regimen with as narrow a spectrum as possible **(Chap. 149)**, empirical regimens are necessarily somewhat broad, given that a specific diagnosis has not yet been made. **Table 124-5** lists empirical antibiotic treatment regimens for commonly encountered infectious presentations. These regimens should be narrowed as appropriate once a specific diagnosis is made. In addition to antibiotics, there is sometimes a role for adjunctive therapies, such as intravenous immunoglobulin G (IVIG) pooled from healthy adults or hyperimmune globulin prepared from the blood of individuals with high titers of specific antibodies to select pathogens (e.g., cytomegalovirus, hepatitis B virus, rabies virus, vaccinia virus, *Clostridium tetani*, varicella-zoster virus, *Clostridium botulinum* toxin). Although the data suggesting efficacy are limited, IVIG is sometimes used for patients with suspected staphylococcal or streptococcal toxic shock syndrome.

INFECTION CONTROL

When evaluating a patient with a suspected infectious disease, the physician must consider what infection control methods are necessary to prevent transmission of any possible infection to other people. In 2007, the U.S. Centers for Disease Control and Prevention published guidelines for isolation precautions that are available for download at *www.cdc.gov/infectioncontrol/guidelines/isolation/*. Persons exposed to certain pathogens (e.g., *N. meningitidis*, HIV, *Bacillus anthracis*) should receive postexposure prophylaxis to prevent disease acquisition. (See relevant chapters for details on specific pathogens.)

WHEN TO OBTAIN AN INFECTIOUS DISEASE CONSULT

At times, primary physicians need assistance with patient management from a diagnostic and/or therapeutic perspective. Multiple studies have demonstrated that an infectious disease consult is associated with improved outcomes, shorter length of hospital stay, and decreased costs for patients with various diseases. For example, in a prospective cohort study of patients with *S. aureus* bacteremia, infectious disease consultation was independently associated with a 56% reduction in 28-day mortality. While artificial intelligence–based chatbots are beginning to be utilized in healthcare settings, they are not yet sophisticated enough to supplant an actual infectious disease consultation. In addition, infectious disease specialists provide other services (e.g., infection control, antimicrobial stewardship, management of outpatient antibiotic therapy, occupational exposure programs) that have been shown to benefit patients. Whenever such assistance would be advantageous to a patient with a possible infection, the primary physician should opt for an infectious disease consult. Specific situations that might prompt a consult include (1) difficult-to-diagnose patients with presumed infections, (2) patients who are not responding to treatment as expected, (3) patients with a complicated medical history (e.g., organ transplant recipients, patients immunosuppressed due to autoimmune or inflammatory conditions), and (4) patients with "exotic" diseases (i.e., diseases that are not typically seen within the region).

TABLE 124-4 Typical Cerebrospinal Fluid Profiles for Meningitis and Encephalitis[a]

	NORMAL	BACTERIAL MENINGITIS	VIRAL MENINGITIS	FUNGAL MENINGITIS[b]	PARASITIC MENINGITIS	TUBERCULOUS MENINGITIS	ENCEPHALITIS
WBC count (per μL)	<5	>1000	25–500	40–600	150–2000	25–100	50–500
Differential of WBC	60–70% lymphocytes, ≤30% monocytes/macrophages	↑↑PMNs (≥80%)	Predominantly lymphocytes[c]	Lymphocytes or PMNs, depending on specific organism	↑↑ Eosinophils (≥50%)[d]	Predominantly lymphocytes[c]	Predominantly lymphocytes[c]
Gram's stain	Negative	Positive (in >60% of cases)	Negative	Rarely positive	Negative	Occasionally positive[e]	Negative
Glucose (mg/dL)	40–85	<40	Normal	↓ to normal	Normal	<50 in 75% of cases	Normal
Protein (mg/dL)	15–45	>100	20–80	150–300	50–200	100–200	50–100
Opening pressure (mmH_2O)	50–180	>300	100–350	160–340	Normal	150–280	Normal to ↑
Common causes	—	*Streptococcus pneumoniae*, *Neisseria meningitidis*	Enteroviruses	*Candida*, *Cryptococcus*, and *Aspergillus* spp.	*Angiostrongylus cantonensis*, *Gnathostoma spinigerum*, *Baylisascaris procyonis*	*Mycobacterium tuberculosis*	Herpesviruses, enteroviruses, influenza virus, rabies virus

[a]Numbers indicate typical results, but actual results may vary. [b]Cerebrospinal fluid characteristics depend greatly on the specific organism. [c]Neutrophils may predominate early in the disease course. [d]Patients typically have striking eosinophilia as well. [e]Sensitivity can be increased by examination of a smear of protein coagulum (pellicle) and the use of acid-fast stains.

Abbreviations: PMNs, polymorphonuclear neutrophils; WBC, white blood cell.

TABLE 124-5 Initial Empirical Antibiotic Therapy for Common Infectious Disease Presentations[a]

CLINICAL SYNDROME	COMMON ETIOLOGIES	ANTIBIOTIC(S)	COMMENTS	SEE CHAPTER(S)
Septic shock	*Staphylococcus aureus*, *Streptococcus pneumoniae*, enteric gram-negative bacilli	Vancomycin, 15 mg/kg q12h[b] ***plus*** A broad-spectrum antipseudomonal β-lactam (piperacillin-tazobactam, 4.5 g q6h; imipenem, 1 g q8h; meropenem, 1 g q8h; or cefepime, 1–2 g q8–12h)	If a pseudomonal species is likely, a second antipseudomonal agent should be added.	**315**
Meningitis	*S. pneumoniae*, *Neisseria meningitidis*	Vancomycin, 15 mg/kg q12h[b] ***plus*** Ceftriaxone, 2 g q12h	Dexamethasone (0.15 mg/kg IV q6h for 2–4 d) should be added for patients with suspected or proven pneumococcal meningitis, with the first dose administered 10–20 min before the first dose of antibiotics.	**143** and pathogen-specific chapters
CNS abscess	*Streptococcus* spp., *Staphylococcus* spp., anaerobes, gram-negative bacilli	Vancomycin, 15 mg/kg q12h[b] ***plus*** Ceftriaxone, 2 g q12h ***plus*** Metronidazole, 500 mg q8h	—	**143**
Acute endocarditis (native valve)	*S. aureus*, *Streptococcus* spp., coagulase-negative staphylococci	Vancomycin, 15 mg/kg q12h[b] ***plus*** Cefepime, 2 g q8h	—	**133**
Pneumonia Community-acquired, outpatient	*S. pneumoniae*, *Mycoplasma pneumoniae*, *Haemophilus influenzae*, *Chlamydia pneumoniae*	No comorbidities[h]: Azithromycin, 500 mg PO? 1, then 250 mg PO qd 4 days With comorbidities[h]: Levofloxacin, 750 mg PO qd	If MRSA is a consideration, add vancomycin (15 mg/kg q8–12h[b]) or linezolid (600 mg q12h); daptomycin should not be used in patients with pneumonia.	**131** and pathogen-specific chapters
Inpatient, non-ICU	Above plus *Legionella* spp.	A respiratory fluoroquinolone (moxifloxacin, 400 mg IV/PO qd; gemifloxacin, 320 mg PO qd; or levofloxacin, 750 mg IV/PO qd) ***or*** A β-lactam (cefotaxime, ceftriaxone, or ampicillin-sulbactam) *plus* azithromycin		
Inpatient, ICU	Above plus *S. aureus*	A β-lactam ***plus*** Azithromycin *or* a respiratory fluoroquinolone		
Hospital-acquired pneumonia[d]	*S. pneumoniae*, *H. influenzae*, *S. aureus*, gram-negative bacilli (e.g., *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Acinetobacter* spp.)	An antipseudomonal β-lactam (cefepime, 2 g q8h; ceftazidime, 2 g q8h; imipenem, 500 mg q6h; meropenem, 1 g q8h; or piperacillin-tazobactam, 4.5 g q6h) ***plus*** An antipseudomonal fluoroquinolone (levofloxacin, 700 mg qd, or ciprofloxacin, 400 mg q8h) *or* an aminoglycoside (amikacin, 15–20 mg/kg q24h[c]; gentamicin, 5–7 mg/kg q24h[e]; or tobramycin, 5–7 mg/kg q24h[e])	If MRSA is a consideration, add vancomycin (15 mg/kg q8–12h[b]) or linezolid (600 mg q12h); daptomycin should not be used in patients with pneumonia.	
Complicated intraabdominal infection Mild to moderate severity	Anaerobes (*Bacteroides* spp., *Clostridium* spp.), gram-negative bacilli (*Escherichia coli*), *Streptococcus* spp.	Cefoxitin, 2 g q6h ***or*** A combination of metronidazole (500 mg q8–12h) *plus one of the following*: cefazolin (1–2 g q8h), cefuroxime (1.5 g q8h), ceftriaxone (1–2 g q12–24h), cefotaxime (1–2 g q6–8h), ciprofloxacin (400 mg q12h), levofloxacin (750 mg qd)	If MRSA is a consideration, add vancomycin (15 mg/kg q12h[b])	**137**, **182**, and pathogen-specific chapters
High-risk patient or high degree of severity	Same as above	A carbapenem (imipenem, 500 mg q6h; meropenem, 1 g q8h; doripenem, 500 mg q8h) ***or*** Piperacillin-tazobactam, 3.375 g q6h[f] ***or*** A combination of metronidazole (500 mg q8h) *plus* an antipseudomonal cephalosporin (cefepime, 2 g q8h; ceftazidime, 2 g q8h)		

(*Continued*)

TABLE 124-5 Initial Empirical Antibiotic Therapy for Common Infectious Disease Presentations[a] (Continued)

CLINICAL SYNDROME	COMMON ETIOLOGIES	ANTIBIOTIC(S)	COMMENTS	SEE CHAPTER(S)
Skin and soft tissue infection	*S. aureus*, *Streptococcus pyogenes*	Dicloxacillin, 250–500 mg PO qid ***or*** Cephalexin, 250–500 mg PO qid ***or*** Clindamycin, 300–450 mg PO tid ***or*** Nafcillin/oxacillin, 1–2 g q4h	If MRSA is a consideration, clindamycin, vancomycin (15 mg/kg q12h[b]), linezolid (600 mg IV/PO q12h), or TMP-SMX (1–2 double-strength tablets PO bid[g]) can be used.	**134** and pathogen-specific chapters

[a]This table refers to immunocompetent adults with normal renal and hepatic function. All doses listed are for parenteral administration unless indicated otherwise. Local antimicrobial susceptibility profiles may influence the choice of antibiotic. Therapy should be tailored once a specific etiologic agent and its susceptibilities are identified. [b]Trough levels for vancomycin should be 15–20 μg/mL. [c]Trough levels for amikacin should be <4 μg/mL. [d]In patients with late onset (i.e., after ≥5 days of hospitalization) or risk factors for multidrug-resistant organisms. [e]Trough levels for gentamicin and tobramycin should be <1 μg/mL. [f]If *P. aeruginosa* is a concern, the dosage may be increased to 3.375 g IV q4h or 4.5 g IV q6h. [g]Data on the efficacy of TMP-SMX in skin and soft tissue infections are limited. [h]Comorbidities include chronic heart, lung, liver, or renal disease; diabetes mellitus; alcoholism; malignancy; or asplenia.

Abbreviations: CNS, central nervous system; ICU, intensive care unit; MRSA, methicillin-resistant *S. aureus*; TMP-SMX, trimethoprim-sulfamethoxazole.

PERSPECTIVE

The study of infectious diseases is really a study of host–microbial interactions and represents evolution by both the host and the microbe—an endless struggle in which microbes have generally been more creative and adaptive. Given that nearly one-sixth of deaths worldwide are still related to infectious diseases, it is clear that the war against infectious diseases has not been won. For example, a cure for HIV infection is still lacking, there have been only marginal improvements in the methods for detection and treatment of tuberculosis after more than a half century of research, new infectious disease outbreaks (e.g., viral hemorrhagic fevers, SARS-CoV-2, mpox) continue to emerge, and the threat of microbial bioterrorism remains high. The subsequent chapters in Part 5 detail—on both a syndrome and a microbe-by-microbe basis—the current state of medical knowledge about infectious diseases. At their core, all these chapters carry a similar message: Despite numerous advances in the diagnosis, treatment, and prevention of infectious diseases, much work and research are required before anyone can confidently claim we have achieved "the virtual elimination of infectious disease." In reality, this goal will never be attained, given the rapid adaptability of microbes.

FURTHER READING

Baker RE et al: Infectious disease in an era of global change. Nat Rev Microbiol 20:193, 2022.

Bartlett JG: Why infectious diseases. Clin Infect Dis 59:S85, 2014.

Khabbaz RF et al: Challenges of infectious diseases in the USA. Lancet 384:53, 2014.

Maillard A et al: Can chatbot artificial intelligence replace infectious diseases physicians in the management of bloodstream infections? A prospective cohort study. Clin Infect Dis 78:825, 2024.

McQuillen DP, MacIntyre AT: The value that infectious disease physicians bring to the healthcare system. J Infect Dis 216:S588, 2017.

Rubin EJ et al: Audio interview: Dr. Fauci on infectious disease challenges. N Engl J Med 388:e82, 2023.

Verghese A et al: Inadequacies of physical examination as a cause of medical errors and adverse events: A collection of vignettes. Am J Med 128:1322, 2015.

125

Molecular Mechanisms of Microbial Pathogenesis

Jordan B. Jastrab, Marcia B. Goldberg

Infectious diseases involve intricate interactions among the infecting microbe, human tissue and immune system, and the host microbiome. The co-evolution of humans and microbes has led to the emergence of microbial factors that promote infection and corresponding human cellular responses to microbes. Among the microbial factors that promote disease are those that alter human cells, inhibit host immune responses, and respond to the microbiota. The process of infection can be divided into several stages: microbial entry into the human body (*colonization*), microbial attachment in its favored niche and avoidance of host defenses (*infection*), deployment of microbial factors and host inflammatory responses that damage human tissue (*disease*), and microbial release into the environment, where the microbe can infect others (*transmission*). It is notable that for most pathogens, the host inflammatory response contributes substantially to symptoms and tissue damage. Moreover, the microbiota (the collection of microbes that reside in and on the human body) modulates, directly or indirectly, every stage of infection **(Chap. 484)**. This chapter describes the best-understood molecular and cellular mechanisms that contribute to human disease caused by bacterial pathogens.

ENTRY INTO THE HUMAN HOST

Infectious diseases occur when a live pathogen enters or a toxic pathogen product is ingested by the host. Entry of a live microbe into a host can occur by crossing a tissue barrier, through entry into or penetration of the skin or a mucous membrane of the respiratory, gastrointestinal, or genitourinary tract. If live bacteria spread to the connective tissue surrounding a barrier, an infectious collection may form (e.g., skin or intraabdominal abscess, empyema). Microbial penetration of barrier surfaces may also result in bloodstream infection, which can in turn lead to infection of other organ systems.

Infection of the skin and underlying soft tissue (*cellulitis*) typically occurs when bacteria reach the dermis via a physical disruption of the epidermal barrier. This disruption can result from trauma, fungal infection, pressure ulcers, or venous stasis ulcers. Bacteria that are already associated with the skin are most likely to traverse this barrier defect; therefore, *Staphylococcus aureus* and group A streptococci are the most common causes of cellulitis. In patients with diabetic foot ulcers, infections with *Pseudomonas aeruginosa* and anaerobic bacteria are also common.

Entry into the respiratory tract occurs via respiratory droplet nuclei (airborne particles 1–5 μm in diameter) or via fomites introduced on a contaminated hand. Infectious droplet nuclei are generated when an individual with a communicable respiratory infection (e.g., tuberculosis, influenza, COVID-19 infection) sneezes, coughs, or talks, or when environmental fluid contaminated with a noncommunicable infectious agent is aerosolized (e.g., Legionnaires' disease, aspergillosis). A cough may generate 3000 particles, whereas a sneeze may generate up to 40,000 particles; large particles may evaporate down to 0.5 μm, and hygroscopic particles can increase in size as they pass through the moist respiratory tree.

Entry into the gastrointestinal tract occurs via ingestion of contaminated food or water or via person-to-person contact. Pathogens for which the minimum infectious inoculum is large (e.g., 10^8 organisms for epidemic spread of *Vibrio cholerae*, 10^5 organisms for *Salmonella enterica* serovar Typhimurium) are generally acquired via contaminated food or water, whereas pathogens for which the minimum infectious inoculum is small (e.g., 10^1 to 10^2 organisms for *Shigella* spp.) are typically acquired by person-to-person spread.

Entry into the genitourinary tract occurs via colonization of the urethral meatus or vaginal introitus with fecal organisms followed by ascension of the organisms into the bladder or kidneys or via instrumentation. Pyelonephritis can also result from hematogenous seeding of the kidney.

ESTABLISHMENT OF INFECTION

NICHE

Live Pathogens Many bacterial pathogens display tissue tropism; sites of infection are often pathogen-specific and restricted, such that adjacent tissues may be uninvolved. For example, group A streptococci cause pharyngitis and soft tissue infection, but rarely pneumonia. Cholera is an infection of the small intestine, whereas *Shigella* spp. cause disease only in the rectosigmoid colon. To establish infection, the pathogen must access and then survive within its niche (Table 125-1). In the respiratory tree, the site at which pathogens settle is determined by mode of spread: droplet nuclei reach the bronchial tree or alveoli, whereas fomites reach the pharynx or nasal passages. Pathogens move through the gastrointestinal tract via normal intestinal motility.

Tissue association and invasion are dictated by bacterial interaction with host factors, such as glycan-decorated receptors and the extracellular matrix. The environmental conditions of the niche trigger expression of virulence factors required for establishing infection; for example, bile salts in the gut stimulate the expression of *V. cholerae* adhesins and the germination of *Clostridioides difficile* spores. Bacteria commonly manipulate their niche environment in ways that facilitate infection. For example, the gastroduodenal pathogen *Helicobacter pylori* converts urea into ammonia, thereby increasing the pH of the acidic stomach environment, which creates a more hospitable environment.

TABLE 125-1 Bacterial Pathogens, Diseases, and Niches

MOST COMMON TROPISM	BACTERIUM	DISEASE
Skin, respiratory tract, small intestine	*Bacillus anthracis*	Anthrax
Respiratory tract	*Bordetella pertussis*	Pertussis
Systemic	*Borrelia burgdorferi*	Lyme disease
Systemic	*Brucella abortus*	Brucellosis
Systemic	*Burkholderia pseudomallei*	Melioidosis
Eyes, venereal	*Chlamydia trachomatis*	Various chlamydioses, including trachoma
Colon	*Clostridioides difficile*	Colitis
Pharynx	*Corynebacterium diphtheriae*	Diphtheria
Systemic	*Coxiella burnetii*	Q fever
Colon	Enterohemorrhagic *Escherichia coli*	
Stomach	*Helicobacter pylori*	Gastritis, gastric ulcers
Respiratory tract	*Legionella pneumophila*	Legionnaires' disease
Systemic, central nervous system	*Listeria monocytogenes*	Listeriosis
Respiratory tract	*Mycobacterium tuberculosis*	Tuberculosis
Urogenital tract	*Neisseria gonorrhoeae*	Gonorrhea
Respiratory tract	*Pseudomonas aeruginosa*	
Systemic	*Salmonella enterica* serovar Typhi	Typhoid fever
Gastrointestinal tract	*Salmonella enterica* serovar Typhimurium	
Colon, rectum	*Shigella* spp.	Dysentery, shigellosis
Multiple sites	*Staphylococcus aureus*	
Soft tissue	Group A *Streptococcus*	
Small intestine	*Vibrio cholerae*	Cholera
Systemic	*Yersinia pestis*	Plague

Preformed Toxins A small number of diseases are caused by ingestion of preformed bacterial toxins. The most common among these are *S. aureus* enterotoxins and the *Bacillus cereus* emetic enterotoxin. *S. aureus* enteroxins are strongly emetic and can be present in prepared foods such as dairy, meat, eggs, salads, and produce. *B. cereus* produces the peptide toxin cereulide, which acts as a potassium ionophore and induces emesis; this toxin is most often found in rice or other starchy food that has been improperly refrigerated. Although these pathogens may be killed when food is cooked, their toxins are heat stable. Less common but important is botulinum toxin, produced by *Clostridium botulinum* and the cause of botulism. Botulinum toxin blocks neurotransmitter release in motor neurons, inhibiting the central nervous system and resulting in a flaccid paralysis that may be complicated by fatal respiratory failure. Although botulinum toxin is heat-labile, *C. botulinum* spores are heat-stable.

ATTACHMENT

Attachment of bacteria to host tissue surfaces is a prerequisite for the pathogen's establishment of an infection and is mediated by specific receptor–ligand interactions. The tissue specificity of the host cell surface receptor repertoire is a critical factor in delimiting a pathogen's niche(s). These physical associations additionally facilitate the pathogen's avoidance of host clearance mechanisms (see "Avoidance of Innate Immune Responses," below) and may contribute to formation of biofilms by the pathogen (see "Biofilms," below). Because adhesion to cellular receptors often triggers cellular signal transduction and innate immune signaling, therapeutic blockade of this interaction may in some circumstances exacerbate infection.

Adhesins Bacterial pathogens have evolved a wide range of strategies by which to attach to the diverse host cell structures they encounter. For many bacterial pathogens, ligands or adhesins for specific host receptors are known. Adhesins comprise a wide variety of surface structures, including single proteins, carbohydrates, glycoproteins, lipids, lipoproteins, and multiprotein filamentous complexes that extend several micrometers from the bacterial surface, each anchored in the outer-surface cell envelope. Most bacteria produce multiple adhesins with varying specificity, enabling the pathogen to interact with multiple receptors, including those on several distinct cell and tissue types encountered during the process of infection. These interactions are often partially redundant, are serologically variable, and contribute additively or synergistically with other binding interactions.

Common classes of adhesins are pili (also known as fimbriae), flagella, and autotransporter proteins (Table 125-2). Pili are hairlike extensions consisting of a polymer of the major pilin subunit capped with minor pilins that provide the adherence function of the structure. Pili are classified by type and are produced by many gram-negative bacteria and a smaller number of gram-positive bacteria. To date, efforts to prevent infection with pilus-based vaccines have been unsuccessful.

Types of pili include type I, type P, and type IV. Type I pili frequently function at mucosal surfaces. For example, they mediate the close association of uropathogenic *Escherichia coli* (UPEC) with bladder epithelial cells and the ability of this pathogen to persist, causing relapsing urinary tract infections. UPEC also produces type P pili and afimbrial adhesins. These adhesins bind sugar moieties on host surface glycoproteins, with varied specificity depending on the adhesin. The minor pilin lectins at the tip of the pili generally bind D-mannose glycans, albeit with strain specificity. For example, the type I pilus adhesin FimH of intestinal *E. coli* strains often preferentially binds oligomannose, whereas the FimH of UPEC strains commonly binds monomannose. Thus, the same pilus structures in different bacterial strains can dictate adherence to distinct tissues. Type IV pili (Tfp) are widespread among gram-negative bacteria, and similar structures exist among gram-positive bacteria. These are evolutionarily related to the type II secretion system (T2SS) and, in addition to mediating adherence, allow the uptake of

TABLE 125-2 Classes of Bacterial Adhesion Proteins and Their Host Receptors

ADHESIN	EXAMPLE	RECEPTOR
Type I pili	Fim protein, uropathogenic *Escherichia coli*	Terminal mannose of uroplakin *N*-glycan in urinary epithelial cells
Type P pili	Pap protein, uropathogenic *E. coli*	Galactose disaccharides
Type IV pili	Tfp protein, *Neisseria gonorrhoeae*	CD64, CR3, I domain–containing integrins
MSCRAMM	SdrC protein, *Staphylococcus aureus*	β-Neurexin
Opa	Opa protein, *Neisseria meningitidis*	CEACAMs[a]
Flagellum	FliC protein, *Pseudomonas aeruginosa*	Asialo-GM1 ganglioside
Autotransporter	Invasin, *Yersinia pseudotuberculosis*	β1-Integrins
Autotransporter	Ag85, *Mycobacterium tuberculosis*	Fibronectin

[a] Carcinoembryonic antigen–related cell adhesion molecules.

DNA into bacteria and the motility of bacteria on surfaces. The Tfp of *Neisseria* spp. and *V. cholerae* mediate aggregation of individual bacteria into microcolonies, which promotes colonization.

Flagella are polymeric helical filaments that propel bacteria through liquid environments by rotating about their long axis. Because flagella confer the ability to swim toward a target surface, often following a chemotactic gradient where chemical sensing influences the direction of motility, they are vital virulence factors of many pathogenic bacteria. Flagella can be localized to one end of the bacterial cell (*polar*) or distributed around the bacterial surface (*peritrichous*). Flagella are evolutionarily related to type III secretion systems (T3SSs) (see "Replicative Niche" and "Survival in the Vacuole," below) and have been shown in some instances to be responsible for the secretion of bacterial toxins. Flagella may also act as adhesins, binding to mucins of mucosal surfaces in the case of the gastrointestinal pathogens enteropathogenic *E. coli* (EPEC) and enterohemorrhagic *E. coli* (EHEC).

Autotransporter proteins comprise a subdivision of type V secretion systems (T5SSs), which are prevalent among gram-negative bacteria (Fig. 125-1). Extended adhesive projections anchored in the bacterial outer membrane, autotransporter proteins are pivotal virulence determinants for several human pathogens, including the filamentous hemagglutinin of *Bordetella pertussis* (the etiologic agent of whooping cough) and the IcsA adhesin and intracellular motility determinant of *Shigella* spp. (responsible for dysenteric diseases). The autotransporter proteins intimin and invasin are required for the intimate association of adhering and effacing pathogens, such as EPEC, and invasion by *Yersinia* spp., respectively. Besides these main classes of adhesins, other bacterial surface proteins involved in adhesion include the Opa family of membrane proteins of *Neisseria* spp. and the gram-positive cell wall–anchored microbial surface component recognizing adhesive matrix molecules (MSCRAMMs) of *S. aureus* and various enterococci.

Receptors Carbohydrates (glycans) on the surface of and secreted by human cells play major roles in the adherence of bacterial pathogens. The surfaces of human cells are coated with glycoproteins and glycolipids, whereas the extracellular matrix (ECM) scaffold of tissues is mainly composed of secreted proteoglycans. Most mucosal surfaces are covered with a layer of mucus, which consists primarily of mucins, a family of proteins that are heavily glycosylated. Among the human enzymes involved in glycan decoration of secreted proteins is *FUT2*, for which ~20% of the human population harbors two nonfunctional alleles; the importance of human glycans in infection is highlighted by the observation that individuals who lack functional *FUT2* display increased susceptibility to certain bacterial infections, decreased susceptibility to certain viral infections, and altered susceptibility to chronic noninfectious inflammatory diseases. These data are confounded by a relative increase in secretion of sialylated glycans in individuals lacking functional *FUT2*. Glycans participate in many adhesive interactions and serve as receptors for certain bacterial toxins. Certain bacteria enzymatically alter host glycans in a manner that enables improved access to the epithelial surface, as in the case of *V. cholerae* and the oral pathogen *Tannerella forsythia*.

The GM1 ganglioside, a glycolipid, is the receptor for cholera toxin (CTX), *Pseudomonas aeruginosa* flagella, and the *Clostridium perfringens* α toxin, and contributes to tropism of the gastric pathogen *H. pylori* through interactions with the bacterial membrane proteins BabA and SabA. Heparan sulfate proteoglycans and other glycosaminoglycan-conjugated proteins are commonly associated with the basolateral membranes of epithelial layers and act as ligands for the

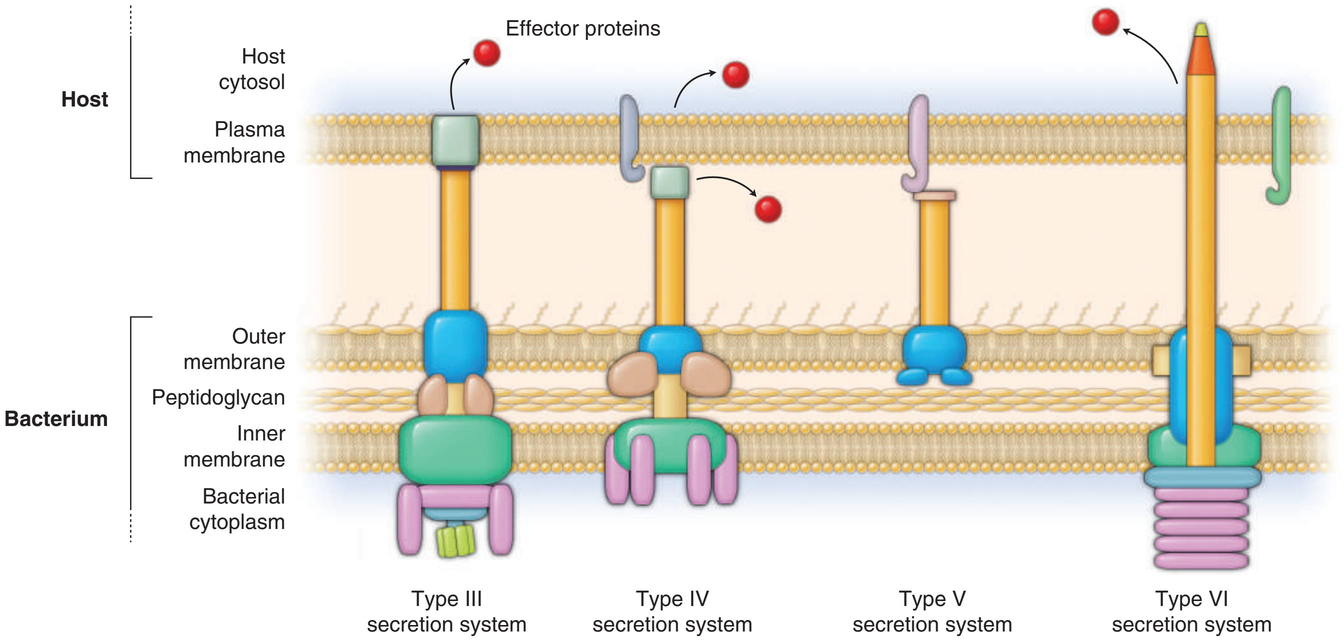

FIGURE 125-1 Major bacterial secretion systems involved in pathogenesis. Schematic of the types III, IV, V, and VI secretion systems (T3SS, T4SS, T5SS, and T6SS, respectively) of gram-negative bacteria. The T3SS, T4SS, and T6SS deliver bacterial effector proteins into host cells, whereas the T5SS participates in adhesion to the surface of cells. The architecture of the extracellular portion of the T4SS and how it translocates effector proteins remain poorly defined. Colored shapes with hooks in the host plasma membrane represent host membrane proteins that, it is thought, may participate in these processes.

chlamydial OmcB protein and the *E. coli* cytotoxic necrotizing factor toxin; these interactions promote initial bacterial adherence in proximity to the plasma membrane and lead to engagement of additional surface molecules.

In mammals, the ECM and integrin proteins are ubiquitous. The ECM consists of laminin, vimentin, and type IV collagen, which interact via fibronectin with integrin receptors in the plasma membrane. Because of direct interaction with plasma membrane–embedded integrins, ECM alterations can result in signal transduction that directly influences immune cell behavior. Many bacterial pathogens engage integrins and the functionally related carcinoembryonic antigen–related cell adhesion molecules (CEACAMs) as host cell receptors and as triggers for their internalization into human cells (see "Mechanisms of Microbial Entry into Cells," below). The receptor for adherent-invasive *E. coli*, a pathogenic type of *E. coli*, is CEACAM6, whose levels are increased on epithelial cells in inflammatory bowel disease; adherent-invasive *E. coli* adhere at increased levels in Crohn disease, and sites of excessive bacterial adherence display increased inflammation.

Among the many other host cell surface proteins that serve as receptors for bacterial adherence is the cystic fibrosis transmembrane conductance regulator (CFTR), a chloride channel involved in the maintenance of adequate hydration of mucosal surfaces. Mutation of the *CFTR* gene gives rise to the hereditary disease cystic fibrosis. These patients are hypersusceptible to respiratory infection because the cilia are unable to clear viscous mucus from the bronchial epithelial surface. The remarkably high frequency of cystic fibrosis (>2.5%) in Caucasian populations has been attributed to relative resistance to *Salmonella enterica* serovar Typhi infection, since *S.* Typhi adheres to CFTR via type IV pili. Epithelial cells utilize CFTR during internalization-mediated clearance of the extracellular pathogen *P. aeruginosa*, potentially contributing to the decreased ability of cystic fibrosis patients to clear infections by this pathogen.

■ REPLICATIVE NICHE

Once bacterial pathogens find their niche and associate with target host cells, they must replicate to persist. Some pathogens remain predominantly associated with the surface of epithelial cells (e.g., *Bordetella*, *Pseudomonas*, *Vibrio*, and *Clostridium* spp.); others predominantly enter into cells, where they may survive within the cytosol (e.g., *Shigella*, *Francisella*, and *Listeria* spp.) or within a membrane-bound compartment (vacuole) comprised of host-derived lipids (e.g., *Salmonella*, *Legionella pneumophila*). Phagocytic cells actively engulf pathogens for destruction (*phagocytosis*); to remain associated with the cell surface, extracellular bacteria must avoid or inhibit phagocytosis. Many pathogens deliver cytoskeleton-disrupting proteins that perturb the phagocytic process. Others create structures, such as *P. aeruginosa* biofilms and streptococcal cell chains, that impede phagocytosis. A subpopulation of any infecting pathogen will nevertheless be phagocytosed by macrophages, neutrophils, and/or dendritic cells (phagocytes). Many pathogens have evolved mechanisms to survive within phagocytes; for example, during colonization of the nasopharynx, *Streptococcus pneumoniae* survives within vacuoles in dendritic cells and macrophages. Nonphagocytic cells, such as epithelial cells, may also internalize bacteria in processes triggered by either the host cell (as a clearance mechanism) or the pathogen (for tissue invasion).

Extracellular Pathogens Extracellular pathogens replicate at the surface of host cells. Many secrete enzymes that liberate nutrients and other factors that fashion a hospitable niche. For example, *V. cholerae* secretes the mucinase vibriolysin that degrades the mucus barrier overlying intestinal epithelial cells, providing bacterial access to the cell surface. When phagocytosed, *V. cholerae* inhibits further phagocytosis of extracellular bacteria. *Yersinia* spp., which associates with the surface of leukocytes, delivers into these cells bacterial proteins that inhibit phagocytosis.

EPEC and EHEC efface the microvilli of the brush border of intestinal epithelial cells and adhere to the surface of these cells. These bacteria induce formation of actin-rich pedestals at the plasma membrane that inhibit bacterial internalization. Tight junctions mediate intimate associations between neighboring epithelial cells, thereby maintaining tissue barrier function; the dissolution of tight junctions by pathogens facilitates bacterial penetration into the tissue.

BIOFILMS Some extracellular pathogens, including *S. aureus* and *P. aeruginosa*, establish chronic infections by producing extracellular polymeric matrices called *biofilms*, which encase the bacteria. Biofilms commonly develop where tissue integrity has been compromised, such as in burn wounds and on prosthetic material. The biofilm matrix is composed of extracellular polysaccharides and DNA, to which the bacteria adhere. Biofilms protect bacteria from phagocytosis while impairing the diffusion of antibodies and antibiotics, enabling the bacteria to avoid elimination and persist.

Mechanisms of Microbial Entry into Cells Bacterial pathogens that are predominantly intracellular during infection can induce internalization into host cells. Even pathogens that replicate within normally phagocytic cells such as macrophages (e.g., *Salmonella* spp., *L. monocytogenes*) possess mechanisms for inducing internalization into these cells.

TRIGGER MECHANISM Bacterial systems that deliver proteins into human cells are critical to many aspects of pathogenesis (Fig. 125-1). An important example is the type III secretion system (T3SS), which delivers proteins from the bacterial cytoplasm directly into the cytosol of eukaryotic cells. Evolutionarily conserved T3SSs are found in many gram-negative bacterial pathogens. Delivery of bacterial proteins (*effector* proteins) into the host cell can have dramatic effects on host cell function; for many pathogens, these effects include induction of bacterial uptake into the host cell and alteration of cellular processes in ways that promote infection.

The T3SS forms an apparatus on the bacterial surface that resembles a needle and syringe, with a channel down its long axis (Fig. 125-1). The base of the organelle spans the two bacterial membranes. Anchored to the organelle base and protruding from the bacterial surface is a long needle-like structure. Upon contact of the needle tip with the eukaryotic plasma membrane, two proteins secreted through the channel form a pore in the cell's plasma membrane. The tip of the needle docks onto the extracellular face of the pore, forming a continuous conduit between the bacterial cytoplasm, where proteins are synthesized, and the host cell cytosol. Bacterial effector proteins are then delivered into the cell through this conduit.

For those pathogens that use a T3SS to induce entry into human cells (*Salmonella*, *Shigella*, and *Chlamydia* spp.), several of the first effector proteins to be translocated activate actin polymerization immediately beneath the point where the bacterium is docked on the plasma membrane. Polymerized actin pushes outwardly against the plasma membrane, generating large membrane ruffles that engulf the bacterium into a membrane-bound vacuole **(Fig. 125-2)**. *Shigella* spp. escape the vacuole and reside in the host cell cytosol, whereas *Salmonella* and *Chlamydia* spp. predominantly reside within the vacuole.

ZIPPER MECHANISM Several invasive pathogens enter host cells using a zipper-like mechanism that does not require a T3SS. In these instances, bacterial surface molecules engage with and stimulate the clustering of host cell receptors, which form close associations with the bacterium in a processive fashion until the plasma membrane is tightly opposed to and surrounding the bacterium (Fig. 125-2). For example, *Y. pseudotuberculosis* binds β1-integrins; this binding stimulates phosphoinositide 3-kinase (PI3K) and protein kinase B (Akt) signaling pathways, leading to zipper-mediated entry. *L. monocytogenes* engages E-cadherin and the Met receptor, activating β-catenin signaling and zipper-mediated entry. Subpopulations of bacteria classically considered extracellular, such as *P. aeruginosa* and *S. aureus*, may also promote bacterial uptake in a similar fashion.

ENTRY OF CYTOSOLIC BACTERIA INTO NEIGHBORING CELLS Bacteria are initially taken up into a vacuole (nonphagocytic cells) or a phagosome (phagocytic cells). Yet for some pathogens, residence in the vacuole or phagosome is transient; the lifestyles of these pathogens include residence in the cytosol, actin-based motility, and cell-to-cell spread. *L. monocytogenes*, *Shigella* spp., spotted fever group *Rickettsia* spp. (the

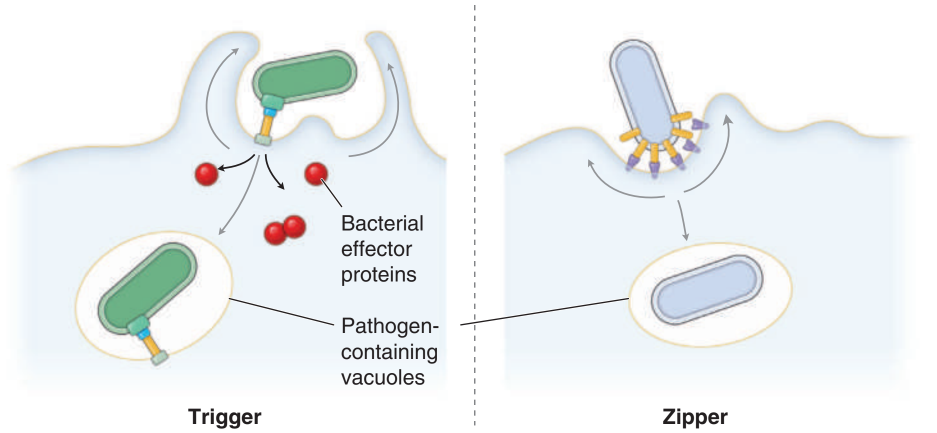

FIGURE 125-2 Common mechanisms of bacterial invasion. The mechanisms for internalization of bacteria into nonphagocytic cells are typically classified as *trigger* or *zipper* mechanisms. As examples of the trigger mechanism, *Shigella* and *Salmonella* spp. use their T3SSs to deliver into host cells effector proteins that manipulate the cytoskeleton in ways leading to the formation of cytoskeleton-supported membrane ruffles. These membrane ruffles extend and surround the pathogen, with consequent endocytosis of the bacterium. As examples of the zipper mechanism, bacterial membrane proteins of *Yersinia* and *Listeria* spp. induce clustering of host receptors. Clustering and subsequent intracellular signaling result in the uptake of the bacterium in a tightly opposed vacuole.

etiologic agents of Rocky Mountain spotted fever, rickettsialpox, and other spotted fever diseases), and *Burkholderia pseudomallei* (the etiologic agent of melioidosis) rapidly lyse the vacuole. *L. monocytogenes* produces a pore-forming toxin, listeriolysin O, and phospholipases that mediate escape into the cytosol, whereas the mechanism of vacuolar escape of *Shigella* spp. is still uncertain except for a requirement of its T3SS.

Once they have escaped into the cytosol, *L. monocytogenes*, *Shigella* spp., *B. pseudomallei*, and spotted fever group *Rickettsia* spp. spread into adjacent cells via actin-based motility. These organisms recruit actin-polymerizing host factors that drive actin polymerization at one end of the bacterium, propelling the microbe through the cell cytosol. Actin-based motility pushes bacteria into host cell membranes to form protrusions that are taken up by the neighboring cell, resolving into a double-membrane vacuole containing the bacterium. This vacuole is lysed once more, releasing bacteria into the cytosol to begin another round of replication. Subversion of the host cytoskeleton aids bacterial cytosolic survival and promotes dissemination through epithelial tissues. *Burkholderia* spp. instead trigger the infected host cell to fuse its membranes with adjacent cells in a T6SS-dependent manner, forming a multinucleated giant cell with contiguous cytoplasm.

SURVIVAL IN THE VACUOLE

After uptake into cells, the majority of invasive bacterial pathogens remain in a vacuole. Uptake vacuoles normally enter the endosomal pathway, acquiring host proteins that promote vesicle maturation, acidification, and degradation of contents. Many bacterial pathogens contained in these endosomes subvert this intracellular trafficking in ways that block endosomal maturation and lysosomal degradation of the pathogen. A distinct subset of bacteria actively damage vacuolar membrane integrity, thereby escaping into the cytosol, where they replicate. For bacteria that remain in the vacuole, various bacterial effector proteins are delivered by secretion systems either into the vacuolar space or across the vacuolar membrane into the cytosol (Fig. 125-1), where these proteins manipulate host processes to the benefit of the pathogen.

■ TYPE III SECRETION SYSTEMS

T3SSs (Fig. 125-1) are versatile virulence systems, helping bacteria such as *P. aeruginosa* remain extracellular, but promoting both the uptake and the intracellular survival of the intracellular pathogens *S. enterica* serovar Typhimurium, *Chlamydia* spp., and *Shigella* spp. *S. enterica* serovar Typhimurium uses distinct T3SSs for the invasion and subsequent maintenance of the *Salmonella*-containing vacuole (SCV), whereas *Chlamydia* and *Shigella* each encode only one such system. The effector proteins delivered are the major determinants of the lifestyles of these pathogens; *Shigella* quickly escapes the vacuole in a T3SS-dependent manner, whereas *Chlamydia* remains in the vacuolar compartment (the *inclusion*) and uses effector proteins to hijack cellular trafficking and perturb innate immune responses. The SCV and the *Chlamydia* inclusion exhibit distinct traits: the SCV is associated with thin membranous tubules that aid nutrient acquisition, whereas the inclusion is localized at the microtubule-organizing center, which is thought to facilitate recruitment of vesicles.

■ TYPE IV SECRETION SYSTEMS

For other pathogens, type IV secretion systems (T4SSs; Fig. 125-1)—conceptually similar, evolutionarily distinct effector protein delivery systems—are key to the ability to survive in cellular phagocytic vacuoles. These T4SSs are similar to the T3SSs described above in that they form a multiprotein apparatus that contains a continuous channel between the bacterial cytoplasm and the eukaryotic cell cytosol. However, T4SSs are functionally more diverse than T3SSs, in that (1) they are present among both gram-negative and gram-positive bacteria as well as some archaea, (2) a large subset of T4SSs transport DNA in a process called conjugation, and (3) some T4SSs deliver proteins, typically toxins, into bacterial cells rather than eukaryotic cells. Not surprisingly, T4SSs also display great structural diversity.

For the human pathogens *L. pneumophila* and *Coxiella burnetii*, bacterial effector proteins delivered across the vacuolar membrane into the cell cytosol by a T4SS alter maturation of the vacuole in a manner that makes it hospitable for bacterial survival; the resulting vacuole is known as a *Legionella*-containing or *Coxiella*-containing vacuole (LCV or CCV, respectively). The delivered effector proteins block fusion of lysosomes with the vacuole, thereby preventing bacterial degradation by lysosomal enzymes; manipulate host vesicular trafficking pathways; and remodel intracellular membranes to alter the lipid and protein content of the vacuolar membrane. For example, the *L. pneumophila* phospholipase VipD, a T4SS substrate, reduces the levels of phosphatidylinositol 3-phosphate on the LCV, thereby preventing recruitment of the host protein Rab5 GTPase, which is involved in endosomal maturation. The mature LCV membrane contains many features that resemble cellular endoplasmic reticulum. In the case of *C. burnetii*, the vacuole displays more lysosome-like characteristics, including a relatively low pH, which induces the activation of T4SS effector delivery. The delivered bacterial effectors stimulate efficient vesicle recruitment from endosomal and autophagosomal (see "Autophagy," below) networks, massively increasing the membrane of the CCV until it occupies much of the cytosol. For both *L. pneumophila* and *C. burnetii*, the mature bacterium-containing vacuoles are hospitable to bacterial replication and long-term survival.

■ OTHER SECRETION SYSTEMS

Francisella tularensis, the etiologic agent of the zoonotic infection tularemia, is a gram-negative facultative intracellular bacterium that displays a tropism for macrophages. *F. tularensis* resides in a phagosome, the maturation of which it delays by remodeling the lipid content of the vacuolar membrane through the action of its type VI secretion system (T6SS; Fig. 125-1). *F. tularensis* then avoids destruction by lysosomal enzymes by escaping into the cytosol, where it replicates. Another pathogen that escapes into the cytosol of macrophages is the respiratory pathogen *Mycobacterium tuberculosis*. Here, the ESX-1 type VII secretion system (T7SS) is required for lysis of the phagosomal membrane. For both of these pathogens, few effectors of their cognizant secretion system have been characterized, and the process of phagosomal membrane lysis remains poorly understood.

NUTRIENT ACQUISITION

Regardless of the site of bacterial infection, microbes must acquire nutrients to replicate. Both extracellular and vacuolar pathogens must circumvent the nutrient-poor conditions of these sites. Also impacting extracellular bacteria is the microbiota, which contributes to nutritional immunity through competition for resources **(Chap. 484)**. Bacterial pathogens possess an arsenal of nutrient acquisition mechanisms to overcome these challenges. For example, upon oxidative stress, to maximize nutrient uptake, several pathogens secrete metal ion-binding proteins through T4SSs (Fig. 125-1) and upregulate expression of cell surface receptors. Many bacteria manipulate host restriction of nutrients, with intracellular bacteria commonly delivering factors that inhibit host mRNA translation to increase the available pool of amino acids. Pathogens within vacuoles have evolved methods of redirecting host nutrient transport to these vacuoles (see "Cellular Trafficking," below).

CELLULAR TRAFFICKING

Because the vacuolar environment is poor in nutrients, bacteria that survive in vacuoles have evolved intricate strategies for the manipulation of host processes to meet their metabolic needs. Vacuolar pathogens modulate endosomal trafficking to prevent degradation by lysosomal compartments. Whereas cytoskeletal rearrangements are regulated by small GTPases of the Rho family, the transport of vesicles through the cell cytosol is controlled by Rab and Arf GTPases. The co-evolution of vacuole-residing bacteria with their eukaryotic hosts has promoted the acquisition of eukaryotic-like domains in many secreted bacterial effector proteins. Many of these effector proteins mimic host cellular trafficking proteins such as guanine nucleotide exchange factors (GEFs), which regulate GTPase activation, and soluble *N*-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs), which control vesicle fusion with vacuolar and other membranes. These bacterial effectors cause the accumulation of host protein and lipid markers on vacuolar membranes, thereby causing the vacuoles to resemble other host compartments. Consequently, host vesicles are redirected from the endoplasmic reticulum, Golgi apparatus, and secretory pathways to the pathogen-containing vacuole, thus delivering nutrients to the pathogen. *Chlamydia* spp., which are obligate intracellular bacteria, acquire host-derived lipids in the bacterial membranes. Dissection of the modes of action of bacterial effectors has revealed novel biochemical modifications of host trafficking regulators. For example, the *L. pneumophila* T4SS secretes two effector proteins that inactivate the cellular GTPase Rab1: AnkX inactivates Rab1 by transferring to it the phospholipid phosphatidylcholine, and SidM modifies Rab1 by adding an adenosine monophosphate molecule to it. This bacterium delivers >300 effector proteins through its T4SS; some of these proteins (*meta-effectors*) modulate the activity of other effectors rather than host factors. The coordinated activity of the numerous effector proteins provides precise spatiotemporal control of host processes.

AVOIDANCE OF INNATE IMMUNE RESPONSES

For the host to respond to and clear an infection, it must first be capable of sensing it. The route of infection and properties of the pathogen contribute to determining the nature of the immune response. The human immune system is generally considered to consist of two arms: the innate immune system, initiated by germline-encoded sensors, and the adaptive immune system, where lymphocytes are clonally selected in response to specific antigens. Initial interactions of pathogens are with specific host cell types, influenced by bacterial tropism; the resulting release of danger signals and cytokines from the innate immune response shapes the later adaptive immune response.

Epithelia (e.g., in the intestinal and respiratory tracts) are commonly the first tissues that a pathogen encounters. The mucosal surfaces of these epithelia are covered in a gel-like mucus layer that contains secretory immunoglobulins, antimicrobial peptides, and commensal bacteria, which offer the first line of defense that infecting pathogens must overcome **(Chap. 484)**. In the gastrointestinal tract, this barrier can be overcome through hijacking of microfold cells (M cells) in the follicle-associated epithelia. M cells sample antigens, including intact pathogenic bacteria, from the gut lumen and deliver them to the underlying gut-associated lymphoid tissue in a process known as *transcytosis*. Within the gut-associated lymphoid tissue, dendritic cells, macrophages, and neutrophils engulf transcytosed material for destruction. Lysosomal degradation of pathogenic bacteria in macrophages and dendritic cells can then lead to antigen presentation, where the hydrolytic products are processed for display by the major histocompatibility complexes (MHCs). These antigen-presenting cells migrate through the lymphatic system to lymph nodes, where the adaptive immune response is stimulated through activation of B-cell and T-cell clones.

COMPLEMENT

Once bacteria have crossed epithelial barriers into the lamina propria, they encounter components of humoral immunity. The complement system is a complex of plasma and tissue-resident proteins that can undergo activation from their pre-protein or zymogen forms to either label pathogens for phagocytosis (*opsonization*) or lyse them directly **(Fig. 125-3)**. Complement activation commences with detection of the pathogen through the binding of circulating antibodies, which are recognized by the C1 complex (*classical pathway*), or of lectins (*lectin pathway*) to surface carbohydrates. These pathways converge at the deposition of the C3 convertase complex on the bacterium. The *alternative pathway* is the third route of complement activation, whereby an alternative C3 convertase complex is formed either spontaneously or by the action of plasma factors, promoting amplification of the cascade. Activation of further complement components results in chemokine production, bacterial opsonization, or assembly of the membrane attack complex pore on the bacterial surface, which lyses the pathogen. Bacterial virulence factors have been described that inhibit each step of the complement pathway, highlighting the importance of this antimicrobial system to host immunity. The various strategies of bacterial interference include prevention of initial recognition by assembly of a polysaccharide capsule, modification of bacterial lipopolysaccharide (LPS), degradation of complement components, and masking of surfaces with host proteins. Streptococci employ numerous mechanisms for complement evasion. The streptococcal proteases ScpA and SpeB degrade many complement proteins, whereas the bacterial surface-exposed M proteins bind the complement inhibitor C4BP, thereby preventing C3b deposition. The thick peptidoglycan layer of gram-positive bacteria provides modest resistance to membrane attack complex–mediated lysis. Furthermore, streptococci produce a hyaluronic capsule that shields the bacterial surface from complement recognition. In comparison, *Borrelia burgdorferi*, the etiologic agent of Lyme disease, encodes a CD59-like protein that inhibits the completion of membrane attack complex assembly. Whereas many pathogens inhibit opsonization, thereby avoiding phagocytosis, other bacteria, including *Francisella* and *Yersinia* spp., promote this process, leading to uptake into the phagosomal compartment, wherein they replicate and prevent lysosomal fusion.

LYSOSOMES

Lysosomes are vesicular organelles of the endosomal system that contain hydrolytic and antimicrobial enzymes within an acidic environment. These enzymes are crucial for recycling of cellular material but also participate in direct killing of microbes and initiation of antibacterial immunity. Lysozyme is a lysosomal enzyme that hydrolyzes the polymeric glycan chains in peptidoglycan, resulting in bacterial lysis. Liberated peptidoglycan fragments may then stimulate innate immune receptors to drive antibacterial immunity. Lysozyme is highly effective against most gram-positive bacteria; however, the peptidoglycan layer of gram-negative bacteria lies between the bacterial outer and inner membranes, making it inaccessible to lysozyme. Gram-positive bacteria can decrease their susceptibility to lysozyme with peptidoglycan modifications that block access of lysozyme to the glycan chain, such as acetylation of sugars in the glycan strand. Metabolites present within the lysosome may also contribute to immunity; for example, the mitochondria-derived metabolite itaconate accumulates within lysosomes during infection of macrophages with *S. enterica* serovar Tymphimurium, where it limits bacterial growth.

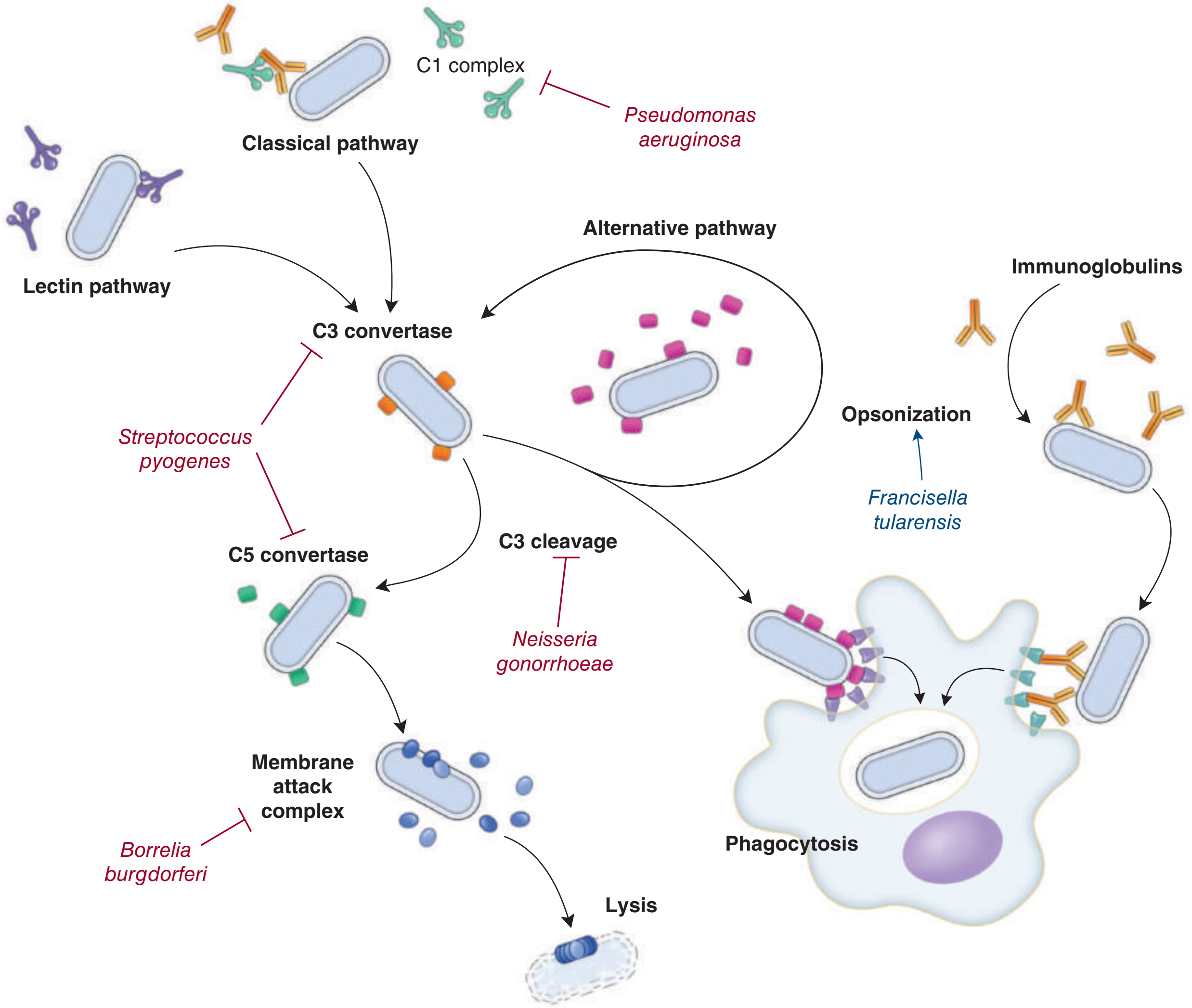

FIGURE 125-3 Overview of the complement system in bacterial infection. The *classical* and *lectin* pathways of complement activation are initiated by the binding of the C1 complex to antibodies bound to bacteria or to lectins binding carbohydrate moieties on the bacterial surface, respectively. A cascade of proteolytic cleavage generates the C3 convertase complex on the bacterial surface, which, upon activation, leads to C5 convertase formation. In the *alternative* pathway, spontaneous C3 cleavage forms alternative C3 products that lead to bacterial opsonization and to amplification of C3 convertase activation. The three complement pathways converge at the point of C3 convertase activation. In a manner similar to that seen for bacteria opsonized with circulating immunoglobulin, opsonized bacteria bind to surface receptors of phagocytes, triggering their engulfment for destruction and stimulation of the adaptive immune system. The C5 convertase, generated by conventional C3 convertase activation, recruits further complement proteins to form the membrane attack complex that directly lyses bacteria. Many bacterial pathogens inhibit specific points in the complement cascade, a selection of which are shown with blunt-ended arrows; in contrast, others, such as *Francisella* spp., promote their own opsonization and subsequent phagocytosis, enabling them to reach their intracellular replicative niche.

ANTIMICROBIAL PEPTIDES

Antimicrobial peptides, such as the positively charged defensins and cathelicidins, are concentrated in degradative compartments. These molecules bind the anionic surface of bacteria and perturb the integrity of LPS of the gram-negative outer membrane and wall teichoic acids in the gram-positive cell wall. Many bacterial pathogens modify their surfaces to reduce the net negative charge, thereby decreasing the binding of positively charged antimicrobial peptides. *S. aureus* and the respiratory pathogen *Burkholderia cenocepacia* secrete proteases that cleave antimicrobial peptides, facilitating survival at mucosal surfaces and within phagosomes.

OXIDATIVE BURST

Pathogens are exposed to reactive oxygen species and reactive nitrogen species produced by cells responding to infection. A well-characterized example is the neutrophil oxidative burst, in which the membrane complex nicotinamide adenine dinucleotide phosphate oxidase (NADPH oxidase) produces superoxide radicals through the transfer of electrons to molecular oxygen. NADPH oxidase activation at the site of infection generates reactive oxygen species in the phagosome of neutrophils and/or in the extracellular space. Bacterial detoxification enzymes, including superoxide dismutases, peroxiredoxins, and catalases, enable bacteria to survive this oxidative burst. Many bacteria can also block the activity of NADPH oxidase; several pathogens prevent assembly of the NADPH oxidase complex by T3SS effector protein–mediated posttranslational modification of small GTPases. The importance of NADPH oxidase in immune defense is highlighted clinically by the marked susceptibility to certain bacterial and fungal infections in patients with chronic granulomatous disease, which results from nonfunctional alleles for essential subunits of this complex.

NEUTROPHIL EXTRACELLULAR TRAPS

The oxidative burst also stimulates the release of the proinflammatory tumor necrosis factor α, azurophilic granules, and neutrophil extracellular traps (NETs). NETs are an extruded mesh of decondensed chromatin that are thought to ensnare pathogens, thereby restricting their spread and destroying them. Streptococci encode deoxyribonucleases (DNases) and *S. aureus* nuclease degrades NETs. *S. aureus* then converts the nucleotide products of this degradation into deoxyadenosine, which functions as a potent macrophage toxin. This process showcases how bacterial pathogens may utilize antimicrobial responses to their own benefit.

PATTERN RECOGNITION RECEPTORS, GUARD PROTEINS, AND THEIR EVASION

The first line of detection of pathogens by the innate immune system is via recognition of molecular patterns that are indicative of the presence of pathogens, termed pathogen-associated molecular patterns (PAMPs). Host proteins referred to as pattern-recognition receptors (PRRs) recognize PAMPs, enabling cells to establish that an infection is occurring. Membrane-associated PRRs surveil the extracellular or endosomal milieu, whereas cytosolic PRRs monitor for threats within the cytosol. Cytosolic ligands tend to stimulate stronger proinflammatory responses than extracellular ligands, and detection of bacteria-derived PAMPs tends to elicit a more powerful inflammatory response than detection of virus-derived PAMPs. Thus, both the subcellular localization and ligand specificity of PRRs play important roles in dictating the character and amplitude of an immune response. Another class of innate immune proteins, termed *guards*, can also initiate innate immune signaling in a process referred to as effector-triggered immunity. In contrast to the direct binding of PRRs to PAMPs, guards become activated upon sensing activities or cellular changes associated with infection.

TOLL-LIKE RECEPTOR DETECTION OF EXTRACELLULAR AND PHAGOSOMAL LIGANDS

The most extensively studied PRRs for bacterial PAMPs are Toll-like receptors (TLRs), which are membrane-associated. On phagocytic cells, TLRs may employ scavenger receptors as co-stimulatory surface receptors. Different TLRs recognize distinct conserved molecules that include LPS, peptidoglycan, flagellin, and nucleic acids **(Table 125-3)**. Upon binding their cognate ligands, TLRs trigger assembly of supramolecular organizing centers, large protein complexes that are signaling platforms for signal transduction. Ultimately, TLR signaling results in nuclear translocation of transcription factors, including nuclear factor-κB (NF-κB), AP-1, and interferon-regulatory factors (IRFs), which induce expression of cytokines, chemokines, and other immunity-related genes **(Fig. 125-4)**.

ALTERATION OF PAMPS TO PREVENT IMMUNE RECOGNITION Pathogens have evolved a plethora of strategies to avoid immunity; one common strategy is to mask PAMPs. For example, TLR4 binds efficiently to LPS that contains hexa-acylated lipid A. *Y. pestis*, the etiologic agent of plague, produces hexa-acylated LPS in the flea vector, but upon transmission into a human host switches to production of tetra-acylated LPS, which does not activate TLR4. Tetra-acylated LPS is also synthesized by *F. tularensis*, *H. pylori,* and the periodontal pathogen *Porphyromonas gingivalis*.

Bacteria can also mask immune recognition of flagella. The flagellum of *H. pylori* facilitates bacterial access to the gastric epithelium and is essential for virulence. Flagellin, the monomeric protein subunit of the flagellum, is a ligand for TLR5, but *H. pylori* bypasses immune recognition by producing a flagellin epitope that does not bind TLR5. CpG motifs in mammalian DNA are usually methylated, but CpG motifs in bacterial DNA are not; thus, methylation state provides a method of discriminating self from nonself. TLR9 recognizes CpG-containing unmethylated DNA in intracellular membrane compartments of the endosomal system. Group A streptococci secrete the Sda1 DNase that degrades CpG DNA, thereby reducing the immunogenicity of lysed bacteria.

TABLE 125-3 Pattern Recognition Receptors of the Innate Immune System and Their Ligands

PATTERN RECOGNITION RECEPTOR	LIGAND OR MODE OF ACTIVATION
TLR2 (with TLR1 or TLR6)	Lipoproteins
TLR4	LPS
TLR5	Flagellin
TLR9	CpG DNA
NLRP1	Enzymatic cleavage
NLRP3	Ionic flux, mitochondrial damage
NLRP6	Lipoteichoic acid, RNA
NAIPs/NLRC4	Flagellin/T3SS
STING	Cyclic dinucleotides
NOD1 and NOD2	Peptidoglycan

INHIBITION OF NF-κB AND MAPK SIGNALING NF-κB signaling plays a central role in controlling both antibacterial and steady-state immunity. A plethora of bacterial effector proteins target components of the NF-κB pathway in ways that maintain repression of NF-κB-dependent proinflammatory genes, thereby counteracting activation of innate immune responses triggered by their conserved PAMPs.

Many bacteria produce virulence factors with eukaryotic-like domains that enable pathogens to hijack host signal transduction, including the NF-κB pathway. *Ubiquitination*, a posttranslational modification in which a chain of the small protein ubiquitin is covalently linked to a target protein, is a common bacterial mechanism of host signaling inhibition. Ubiquitin is added to a target protein by an E3 ubiquitin ligase, which transfers ubiquitin from an E2 ubiquitin-conjugating enzyme to the target protein. Depending on the precise structure of the ubiquitin chain, the modified protein is subsequently degraded by the proteasome, targeted to particular vesicular compartments, or recruits adaptor proteins that initiate signal transduction.

The translocation of NF-κB into the nucleus requires the successive phosphorylation, ubiquitination, and degradation of IκBα by the IκB kinase (IKK) complex, which in resting cells inhibits NF-κB; the dissociation of the IKK complex from NF-κB enables the latter to translocate into the nucleus, where it functions as a transcription factor **(Fig. 125-4C)**. The T3SS of *Shigella* spp. delivers into cells several E3 ubiquitin ligases that target components of the NF-κB pathway for degradation. The intracellular pathogens *C. trachomatis* and *B. pseudomallei* secrete bacterial deubiquitinases through T3SS and T2SS, respectively, which cleave the ubiquitin moiety from IκBα, thereby preventing its degradation, maintaining inhibited NF-κB in the cytosol, and repressing proinflammatory gene expression.

Other enzymatic activities of bacterial effectors can also inhibit NF-κB activation by ubiquitin-dependent or -independent processes. *Shigella* spp. and *L. pneumophila* deliver effectors via their T3SS and T4SS, respectively, that inactivate E2 enzymes to prevent modification and degradation of the IKK complex. EPEC and *Shigella* spp. release cysteine methyltransferases that inhibit IKK modification, whereas *S. enterica* serovar Typhimurium targets NF-κB directly for degradation via T3SS effectors with metalloprotease activity.

Another cellular pathway of pro-inflammatory signaling stimulated by PRRs is the mitogen-activated protein kinase (MAPK) phosphorylation cascade, which activates transcription factor AP-1. The *B. anthracis* metalloprotease exotoxin lethal factor ablates this kinase cascade by cleaving the N-terminal region of MAPKs. Homologous T3SS effectors of *Yersinia* spp., *S. enterica* serovar Typhimurium, and *V. parahaemolyticus* acetylate MAPKs, preventing their phosphorylation. Finally, *Shigella* spp. and *S. enterica* serovar Typhimurium produce phosphothreonine lyases that cleave phosphorylated threonine residues from MAPKs, preventing their activities.

The broad range of effector protein mechanisms that bacteria employ to subvert TLR and MAPK signaling underscores the critical roles of these pathways in preventing pathogens from establishing infection (Fig. 125-4).

PRR SENSING OF CYTOSOLIC INFECTION Whereas the extracellular environment is surveilled for signs of infection by TLRs, the cytosol is monitored by other PRRs. Among the signaling pathways activated upon recognition of intracellular bacterial pathogens are the cytosolic PRRs NOD1 and NOD2, which activate NF-κB signaling upon binding modified peptidoglycan fragments. These PRRs are critical for defense against numerous bacterial pathogens, yet polymorphisms in the *NOD1* and *NOD2* genes confer susceptibility to inflammatory bowel disease. Another important cytosolic (and occasionally nuclear) innate immune surveillance system is the cyclic guanosine monophosphate–adenosine monophosphate (cGAMP) synthase-stimulator of interferon genes

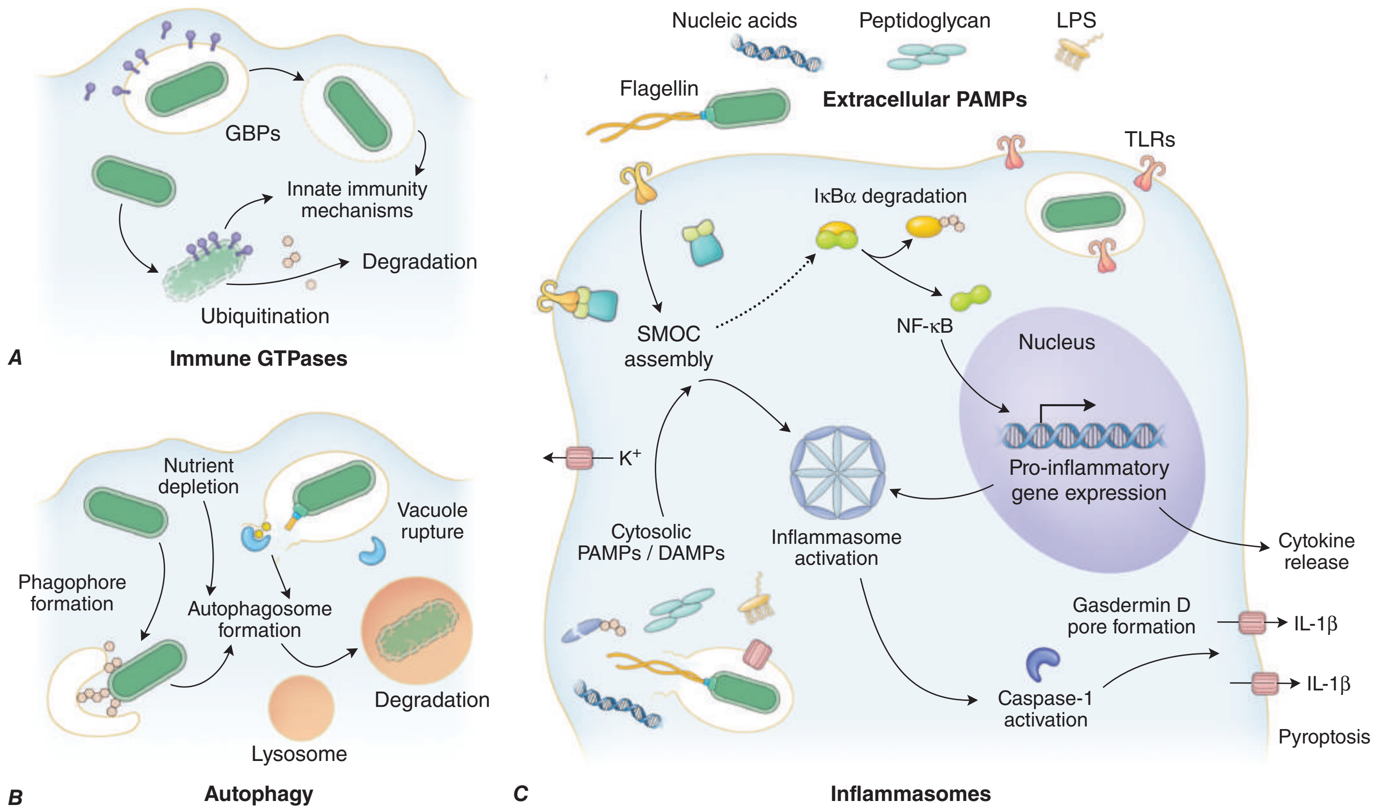

FIGURE 125-4 Overview of innate immune recognition and response to bacterial pathogens. ***A.*** Immune GTPases. Proteins of the GBP family associate with the bacterial surface, LPS shed from the bacterial surface, or pathogen-containing vacuoles, disrupting the integrity of the pathogen and leading to bacterial lysis. Alternatively, labeling of intracellular pathogens with immune GTPases may stimulate ubiquitination of the bacterium and its destruction by cell-autonomous immune mechanisms such as autophagy. ***B.*** Autophagy. Recognition of bacterial invasion leads to host-mediated sequestration of the pathogen in an autophagosome and destruction of the pathogen upon fusion of the autophagosome with lysosomes. Bacteria may be directly labeled by ubiquitination, damaged vacuoles are recognized by the binding of galectins to exposed glycans, and/or the cellular autophagy may be induced in response to sensing starvation. ***C.*** Pattern recognition receptors and innate immune signaling. Conserved PAMPs, DAMPs, and danger signals are recognized by all cells, triggering innate immune signaling pathways and proinflammatory gene expression. TLRs surveil the cell surface and endosomal interior for PAMPs that include LPS, nucleic acids, and flagellin. Cytosolic PRRs and guards, including NLRs and ALRs, recognize similar ligands as well as distinct stimuli, including dysregulation of ion homeostasis. TLR activation promotes supramolecular organizing center assembly and NF-κB signaling. Under resting conditions, NF-κB is sequestered in the cytosol by IκBα, but TLR activation leads to IκBα phosphorylation, ubiquitination, and degradation, with the consequent release of NF-κB. Resultant translocation of NF-κB into the nucleus leads to pro-inflammatory gene expression and the production and release of cytokines. Stimulation of cytosolic PRRs and guards may lead to assembly of supramolecular organizing centers known as inflammasomes, resulting in activation of inflammatory caspases. A well-characterized example is the activation of caspase-1 by canonical and noncanonical inflammasomes, subsequent activation of the pore-forming protein GSDMD, maturation and release of proinflammatory cytokines IL-1β and IL-18, and pyroptotic death of the cell.

(cGAS-STING) pathway. In response to recognition of cytosolic DNA, cGAS synthesizes the second messenger cGAMP. Upon binding cGAMP, STING activates interferon regulatory factor 3 (IRF3) and NF-κB, which subsequently promote transcription of interferon-regulated proinflammatory genes. The products of these genes include cytokines that alert bystander cells and recruit immune cells from the circulation, creating a heightened immune state in the local area.

INFLAMMASOMES Some cytosolic PRRs, including members of the NOD- and AIM2-like receptor (NLR and ALR, respectively) families, can seed formation of supramolecular organizing centers known as *inflammasomes*. Upon binding to a PAMP, these PRRs oligomerize and recruit the zymogen procaspase-1; some PRRs recruit procaspase-1 directly, whereas others bind the adaptor protein ASC, which then recruits procaspase-1. Upon its recruitment, procaspase-1 undergoes autocatalytic cleavage, thereby converting into active caspase-1 and completing formation of an active inflammasome. PRRs capable of triggering inflammasome assembly include NAIP-NLRC4, AIM2, and NLRP6. Select guard proteins can also trigger inflammasome assembly, including NLRP3, Pyrin, and NLRP1. In addition, caspase-4 and -5 (and their murine homolog caspase-11) serve as PRRs that drive assembly of the "non-canonical" NLRP3 inflammasome (see below). The importance of inflammasome signaling is underscored by a plethora of polymorphisms within inflammasome-associated genes that predispose to autoinflammatory disorders or susceptibility to infection.

Upon their activation within an inflammasome, caspases drive immunity by cleaving inflammasome effector proteins. The best-described inflammasome effectors are gasdermin D (GSDMD) and IL-1 family cytokines, including interleukin-1β (IL-1β) and IL-18. Cleavage of GSDMD liberates its N-terminal domain, which forms pores in the plasma membrane; cleavage of the zymogen forms of IL-1 family cytokines converts them into mature, active cytokines, which are released through GSDMD pores and bind the IL-1 receptor, in turn driving robust inflammation (Fig. 125-4). The formation of GSDMD pores in the plasma membrane is often associated with cell death, termed *pyroptosis* (inflammatory cell death).

Inflammasome-seeding PRRs are activated by binding to PAMPs. NAIPs sense the needle protein of the T3SS or the related flagellin protein and subsequently recruit NLRC4 to seed inflammasome formation. Many pathogens have evolved mechanisms to overcome activation of the NAIP-NLRC4 inflammasome. In *L. pneumophila*, *L. monocytogenes*, *P. aeruginosa*, and *S. enterica* serovar Typhimurium, flagellin expression is repressed during infection; similarly, in *S. enterica* serovar Typhimurium, expression of an immunogenic T3SS is repressed and that of a less immunogenic T3SS is upregulated. *S. enterica* serovar Typhimurium also produces a T3SS effector that blunts *NLRC4* expression. AIM2 is activated upon binding cytosolic double-stranded DNA, which is released during lysis of cytosolic bacteria. Bacterial cell wall and membrane modifications that reduce bacteriolysis dampen AIM2 activation during cytosolic infection. Overall, the inhibition of inflammasomes by bacterial pathogens enables pathogens to avoid elimination by the innate immune system.

The NLRP3 inflammasome appears to be activated upon sensing cell stress and homeostatic dysregulation, including via ion flux, mitochondrial damage, and release of oxidized mitochondrial DNA. Because these processes are triggered by membrane damage, NLRP3 can be activated by many bacterial pore-forming toxins; bacteria may avoid stimulating NLRP3 by downregulating the production of toxins. The pyrin inflammasome becomes activated upon the inactivation of Rho GTPases, thereby serving as a sensor of cytoskeletal perturbations. *C. difficile, C. botulinum*, and *B. cenocepacia* all produce effectors that inactivate RhoA via distinct covalent modifications, which induces cell death by alteration of the actin cytoskeleton yet ultimately drives pyrin activation. Meanwhile, NLRP1 acts as bait for cleavage by microbial proteases or E3 ubiquitin ligases; partial degradation of NLRP1 stimulates its activity and subsequent pyroptosis.

Caspases-4 and -5 trigger formation of the noncanonical NLRP3 inflammasome in response to LPS. In human cells, LPS is bound by the PRR NLRP11, which facilitates activation of caspase-4; alternatively, caspases-4 and -5 can themselves serve as PRRs for LPS. Upon activation, these caspases cleave GSDMD but not IL-1β; the formation of GSDMD membrane pores indirectly drives activation of the NLRP3 inflammasome, which subsequently cleaves IL-1β. *S. flexneri* produces an effector that inactivates caspase-4 via ADP-riboxanation, which blocks host cell mechanisms of cytosolic bacteriolysis, thereby preventing release of LPS into the cytosol. Although caspases-4 and -5 play important roles in host defense, they also contribute to pathologic inflammation during gram-negative sepsis, as the murine homolog caspase-11 drives mortality in murine models of sepsis.

Immune GTPases During many infections, the initial pathogen interaction triggers expression of host immunity-related genes that prime host cells for a more robust response. For example, detection of bacteria can drive interferon signaling, which induces production of the guanylate binding protein (GBP) family of immune GTPases. GBPs are recruited to pathogen-containing vacuoles or bacterial membranes and promote the release of PAMPs, making them available as ligands for PRRs. Some GBPs extract LPS from the surface of cytosolic gram-negative bacterial pathogens, facilitating noncanonical inflammasome activation. GBPs contribute to lysis of *F. novicida* (a subspecies of *F. tularensis*) and *Neisseria meningitidis*, causing release of bacterial DNA into the cytosol. Recruitment of GBPs to *Shigella* spp. prevents bacterial actin-based motility and promotes bacterial ubiquitination, which targets the bacteria for proteasomal degradation; simultaneously, *Shigella* spp. secrete an effector that ubiquitinates GBPs, targeting them for degradation.

Inhibition of Nonpyroptotic Cell Death Several types of regulated (*programmed*) cell death besides pyroptosis can occur, of which inflammatory necroptosis and noninflammatory apoptosis are best characterized. In *necroptosis*, activation of any of a variety of innate immune signaling pathways (interferon-γ, TLRs) leads to kinase-dependent assembly of a plasma membrane pore complex that induces cell lysis and release of DAMPs into the extracellular space, events that induce an intense inflammatory response. Necroptosis can drive both pathologic inflammation and antibacterial immunity.

In *apoptosis*, cells condense and form membranous blebs without releasing cytosolic contents and thus without inducing inflammation in adjacent tissues. Apoptotic blebs and dead cells are disposed of by macrophages in a process known as *efferocytosis*. In addition to its role during infection, apoptosis plays a key role in development and is required for normal cell turnover in many epithelial tissues, including the intestine. Apoptosis can be initiated by intrinsic or extrinsic cellular stimuli, resulting in activation of apoptosis-specific caspases that demolish cellular contents. The intrinsic pathway occurs upon dissolution of the mitochondrial membrane potential, which commonly occurs during infection due to cellular stress or the action of bacterial toxins. The extrinsic pathway can be stimulated by interactions of cell-surface death receptors with proinflammatory ligands like TNF-α.

The gastrointestinal pathogens *S. enterica* serovar Typhimurium, *Yersinia enterocolitica*, and pathogenic *E. coli* inhibit signal transduction from cell-surface death receptors by modifying or cleaving the components of death receptor–associated supramolecular organizing centers (Fig. 125-4), thereby blocking cell death. Inhibition of extrinsic apoptotic signaling can stimulate alternative inflammatory cell death pathways, including necroptosis, highlighting the functional redundancy of cell death pathways in some contexts.

Vacuolar *Chlamydia* spp. activate survival pathways through phosphoinositide 3-kinase (PI3K) and Wnt/β-catenin signaling, while simultaneously inhibiting apoptosis. T4SS effectors of vacuolar *L. pneumophila* and *C. burnetii* inhibit apoptosis by sequestering proapoptotic host factors. Cyclomodulins, a group of bacterial effector proteins that arrest the cell cycle, are delivered into cells by EPEC and EHEC. The cyclomodulin Cif deamidates ubiquitin and related proteins, inactivating host E3 ligases involved in cell cycle progression and thereby reducing epithelial turnover; this disruption can be sensed by host cells and triggers other cell death pathways. The diversity of mechanisms that bacteria have accumulated to modulate host cell death pathways highlights the importance of these pathways during infection.

Autophagy Host cells deploy mechanisms of cell-autonomous immunity to neutralize microbial threats. One such mechanism is autophagy, which recycles organelles and protein complexes by generating double-membrane compartments (*autophagosomes*) around the cargo and delivering their contents to the lysosome; the cargo is subsequently degraded to amino acids and other molecules. Autophagy can isolate and dispose of cytosolic pathogens; host polysaccharide-binding molecules termed galectins can bind microbes and induce ubiquitination of the bacterial surface, which drives recruitment of LC3, a host protein that triggers engulfment into autophagosomes.

Intracellular bacterial pathogens must inhibit autophagy to survive. Common mechanisms used by bacteria to interfere with autophagy include masking bacterial ubiquitination sites, inhibiting autophagosome maturation, and altering host pathways that regulate autophagy. *L. monocytogenes* recruits the host factor Arp2/3 to mask the bacterial surface, *F. tularensis* produces an LPS O-antigen that prevents surface access of E3 ligases, and *S. flexneri* produces an effector protein that blocks ubiquitination of the bacterial surface protein IcsA. To inhibit autophagosomal flux, *L. pneumophila* secretes an effector protein that cleaves LC3, *S. enterica* serovar Typhimurium secretes effector proteins that block activation of a kinase required for autophagosome generation, and *M. tuberculosis* and *C. trachomatis* prevent autophagosome-lysosome fusion. An important metabolic trigger of autophagy is intracellular depletion of amino acids, which occurs upon replication of metabolically active intracellular pathogens and triggers inactivation of a master regulator of cell metabolism, mammalian target of rapamycin (mTOR). mTOR inactivation derepresses autophagy; invasive pathogens including *M. tuberculosis*, *S. enterica* serovar Typhimurium, and *Shigella* spp. can reactivate mTOR signaling during infection, thereby blunting autophagy.

Epigenetic Control of Innate Immune Responses Dysregulated immune responses (e.g., sepsis, severe COVID-19 infection) can cause morbidity or death, whereas weak immune responses may lead to pathogen-induced morbidity and mortality. A cellular mechanism that promotes appropriate amplitude of the immune response is innate immune training, in which cells are epigenetically primed for immunity via alterations in chromatin structure. Posttranslational modification of histones influences chromatin structure and access of transcription factors to regulatory elements in the DNA. Mounting evidence indicates that the microbiota modifies the methylation state of promoters of genes involved in immunity, facilitating appropriately tuned responses to infection. Pathogens also regulate host responses at the epigenetic level. *B. anthracis* and *L. pneumophila* deliver into cells histone methyltransferases that control inflammation and ribosome activity. *L. monocytogenes* listeriolysin O, the *P. aeruginosa* T3SS pore, and other pore-forming toxins exert epigenetic modulation through alterations in ion homeostasis, thereby manipulating the gene expression profiles of infected host cells.

INHIBITION OF ADAPTIVE IMMUNE RESPONSES

Early interactions of pathogens with host cells trigger the innate immune response, releasing cytokines that recruit additional antigen-presenting immune cells to the site of infection. To establish chronic infection, bacterial pathogens need not only to suppress the innate immune system but also to avoid elimination by the adaptive immune response. The adaptive immune system comprises clonally expanded lineages of B and T lymphocytes that have been activated by antigen-presenting cells, the best characterized of which are dendritic cells and macrophages. B cells are required for humoral immunity; genetic diseases that affect B-cell function often manifest as the inability to produce adequate antibody titers to clear extracellular bacterial infections. T cells generally mediate cell-mediated immunity, helping the host clear infected cells. The activation of T cells occurs by the presentation of processed antigen on MHC molecules of antigen-presenting cells. The activation of T cells is controlled by the specificity of MHC molecules and of their receptors. The large variety of receptors possessed by T cells enables collective recognition of a wide variety of antigens. Mutations that ablate the development of the adaptive immune system result in severe combined immunodeficiency, rendering the individual extremely vulnerable to infection.

M. tuberculosis is notorious for its ability to establish latent infections, avoiding elimination by the immune system for decades. Once phagocytosed, the pathogen induces secretion of immunosuppressive cytokines, including IL-6, IL-10, and transforming growth factor β. The result is inhibition of interferon-γ-dependent gene expression, with downregulation of MHC class II and other immune-stimulatory molecules, which inhibits induction of helper $CD4^+$ T lymphocytes. Similarly, *B. pertussis* induces dendritic cells to produce IL-10, which skews T-cell maturation into regulatory T cells, dampening the immune response. The *S. enterica* serovar Typhimurium T3SS effector protein SteD depletes MHC class II molecules from the surface of dendritic cells by the activation of the E3 ubiquitin ligase MARCH8. MARCH8 ubiquitinates MHC class II, interfering with its trafficking to the cell surface and thereby decreasing interaction with and subsequent activation of T lymphocytes.

In addition to preventing antigen-presenting cells from stimulating lymphocytes, pathogens may also directly alter B-cell and T-cell activity. The *Yersinia* T3SS effector YopH, a protein tyrosine phosphatase, dephosphorylates B-cell and T-cell receptors and thus prevents both signal transduction upon stimulation by antigen-presenting cells and lymphocyte activation. Staphylococcal toxic shock syndrome is induced by the production of superantigens by *S. aureus*. These extremely inflammatory exotoxins are potent activators of T cells, stimulating exuberant and at times fatal cytokine production. Rather than binding the specificity-determining variable regions of MHC molecules and the T-cell receptor, superantigens bind invariable regions and are therefore able to nonspecifically activate vast numbers of T cells, with a consequent cytokine storm.

BACTERIAL CYTOTOXINS

Many bacteria produce *cytotoxins*—toxins that trigger host cell death. The best-characterized group of bacterial cytotoxins are the AB toxins, which are composed of an enzymatically active (A) subunit and a binding (B) subunit, which interacts with the cellular receptor. This family of toxins includes Shiga toxins of *Shigella* spp. and some strains of pathogenic *E. coli*, diphtheria toxin of *Corynebacterium diphtheriae*, pertussis toxin of *B. pertussis*, and CTX of *V. cholerae*. In general, upon binding to cell surface receptors, the toxin is endocytosed, whereupon the A-subunit translocates across the endosome membrane into the cytosol where its toxic enzymatic activity is stimulated. CTX and pertussis toxin ADP-ribosylate G-protein regulators of adenylate cyclases, thereby increasing cellular cyclic AMP (cAMP) concentrations. Within the cell, increased cAMP perturbs ion homeostasis and apoptosis; on a tissue level, increased intracellular cAMP induces chloride secretion via CFTR and inhibits sodium chloride absorption, driving massive fluid secretion into the lumen of the small intestine and the diarrheal symptoms of cholera. *C. perfringens* and *P. aeruginosa* produce non-AB family toxins that ADP-ribosylate elongation factor-2, thereby inhibiting host translation. *S. enterica* serovar Typhimurium and *C. difficile* produce toxins that target actin, inhibiting normal cellular cytoskeletal rearrangements. As dictated by the host targets modified, the effects of bacterial toxins on the host cell generally fall into the few broad categories of cytoskeletal manipulation, inhibition of innate immune signaling, and hijacking of cellular trafficking.

Pore-forming toxins are secreted virulence factors produced by many extracellular pathogens, including *S. pneumoniae*, *C. perfringens*, *B. anthracis*, and *L. monocytogenes*, among many others. *S. aureus* produces several types of pore-forming toxins, including leukocidins, hemolysins, and phenol-soluble modulins. Leukocidins and α- and γ-hemolysins oligomerize within host plasma membranes, leading to host cell lysis. The cell-type specificities of leukocidins and hemolysins are driven by toxin binding to specific cell surface receptors. Conversely, phenol-soluble modulins are small amphipathic peptides that directly insert into cell membranes. Lytic toxins protect *S. aureus* from phagocytosis and prevent an infected host from developing a protective adaptive immune response. Host cells can counter cytolysins by releasing *exosomes*, vesicles that act as decoys by removing cytolysins from the local environment.

TISSUE DAMAGE AND PATHOGEN DISSEMINATION

Much of the pathology associated with bacterial infection results from proinflammatory immune responses. Infected cells may continually signal in ways that alert the immune system, even though, as described above, many bacteria avoid elimination by cell-autonomous and immune cell–mediated mechanisms. In the intestine, tight junctions mediate intimate associations between epithelial cells that maintain tissue barrier function, linking the cytoskeletal networks of adjacent cells through intimate association of protein complexes across the cell membranes. Many intestinal pathogens perturb the integrity of the gut epithelium either by manipulating cell polarity or by disrupting intercellular junctions. *C. perfringens*, *V. cholerae*, pathogenic *E. coli*, *Shigella* spp., and *S. enterica* serovar Typhi all produce toxins in the gastrointestinal tract that disrupt tight junctions, consequently disrupting the barrier function of the tissue and facilitating access of the pathogen to deeper tissue. The RtxA multifunctional repeats-in-toxin (MARTX) toxin of *V. cholerae* causes cell rounding and barrier failure through actin-crosslinking activity, yet avoids eliciting substantial immune responses by simultaneously inactivating phospholipases and Rho GTPases. The serine protease autotransporters (SPATEs) of *Shigella* spp. and some pathogenic *E. coli* are typically secreted into the gut lumen or mucus layer, whereupon they cleave components of epithelial junctions and mucins in ways that facilitate tissue penetration. Tissue damage permits access to the underlying mucosal layers, the lymphatics, and the bloodstream; for some pathogens, this access enables seeding of other organs.

Transmission to New Hosts The host represents a replicative niche for bacterial pathogens, in which they multiply and are transmitted to new hosts. The mode of transmission is typically aligned with the mode of entry. For example, for respiratory pathogens, coughing induced by tissue damage in the lung aerosolizes the pathogen, enabling inhalation by and colonization of a new host. Similarly, gastrointestinal pathogens elicit diarrhea and are transmitted via the direct fecal–oral route or via contamination of crops or food with waste from an infected individual. The understanding of the spread of infectious diseases permits the institution of basic hygiene procedures that greatly diminish transmission rates—for example, hand washing, decontamination of communal surfaces, and adoption of social distancing measures.

Bacteriophages and Pathogen Lifestyle The reservoir of *V. cholerae* is aquatic environments. Disease is acquired through the ingestion of contaminated seawater or seafood. In regions where cholera is endemic, the disease displays seasonal peaks. This seasonality is associated with blooms of bacteria-targeting viruses (*bacteriophages*), which infect *V. cholerae* organisms, replicate, and lyse and kill the bacterial host. Bacterial lysis releases viral particles into the aquatic environment, whereupon they can infect other *V. cholerae*. Consequently, the bacteriophages regulate the abundance of pathogens, with ramifications for the epidemiology of the disease. Whether bacteriophages contribute to cyclical control of other pathogens is currently unknown.

■ SUMMARY

Bacterial pathogens display myriad mechanisms of colonization, adhesion, invasion, dissemination, and manipulation of host pathways. Infectious diseases result when pathogens successfully establish themselves within the host. Symptoms are usually the result of ensuing fights between the pathogen and the immune system. The incredible diversity of virulence determinants highlights the success of the host in combating infection. Further elucidating how bacteria cause infection will better our understanding of both human and microbial biology and will provide new opportunities for successful therapeutic intervention against infectious and inflammatory diseases.

■ FURTHER READING

Costa TRD et al: Secretion systems in gram-negative bacteria: Structural and mechanistic insights. Nat Rev Microbiol 13:343, 2015.

Fitzgerald KA, Kagan JC: Toll-like receptors and the control of immunity. Cell 180:1044, 2020.

Galluzzi L et al: Molecular mechanisms of cell death: Recommendations of the Nomenclature Committee on Cell Death 2018. Cell Death Differ 25:486, 2018.

Jastrab JB, Kagan JC: Strategies of bacterial detection by inflammasomes. Cell Chem Biol 31:835, 2024.

Lamason RL, Welch MD: Actin-based motility and cell-to-cell spread of bacterial pathogens. Curr Opin Microbiol 35:48, 2017.

Remick BC et al: Effector-triggered immunity. Annu Rev Immunol 41:453, 2023.

Ribet D, Cossart P: How bacterial pathogens colonize their hosts and invade deeper tissues. Microbes Infect 17:173, 2015.

Rivera-Cuevas R et al: Human guanylate-binding proteins in intracellular pathogen detection, destruction, and host cell death induction. Curr Opin Immunol 84:102373, 2023.

Stones DH, Krachler AM: Against the tide: The role of bacterial adhesion in host colonization. Biochem Soc Trans 44:1571, 2016.

Tsolis RM, Bäumler AJ: Gastrointestinal host–pathogen interaction in the age of microbiome research. Curr Opin Microbiol 53:78, 2020.

Wu YW, Li F: Bacterial interaction with host autophagy. Virulence 10:352, 2019.

126

Microbial Genomics and Infectious Disease

Roby P. Bhattacharyya, Yonatan H. Grad, Deborah T. Hung

Just as microscopy opened the worlds of microbiology by providing a tool with which to visualize microorganisms, technological advances in genomics provide microbiologists with powerful methods to characterize the genetic map that underlies all microbes with unprecedented resolution, thereby illuminating their complex and dynamic interactions with each other, the environment, and human health. The field of infectious disease genomics encompasses a vast frontier of active research that is transforming public health and the clinical practice of infectious diseases. While genetics has long played a key role in elucidating the process of infection and impacting clinical infectious diseases, the ability to extend our thinking and approaches beyond the study of single genes to an examination of the sequence, structure, and function of entire genomes allows us to identify new possibilities for research and opportunities to change clinical practice and disease surveillance. From the development of diagnostics with unprecedented sensitivity, specificity, and speed to the design of novel public health surveillance tools and interventions, technical and statistical genomic innovations are reshaping our understanding of the influence of the microbial world on human health and providing us with new tools to diagnose, track, and combat infection. In this chapter, we explore the application of genomics methods to microbial pathogens and the infections they cause. We discuss innovations that are driving the development of diagnostic approaches as well as the discovery of new pathogens, providing insight into novel therapeutic approaches and paradigms, and advancing methods in infectious disease epidemiology and the study of pathogen evolution that can inform infection control measures, public health responses to outbreaks, and vaccine development. We draw on examples in current practice and from the recent scientific literature as signposts that point toward ways in which pathogen genomics may influence infectious diseases in the short and long terms, and we highlight applications to SARS-CoV-2 and the COVID-19 pandemic. **Table 126-1** provides definitions for a selection of important terms used in genomics.

MICROBIAL DIAGNOSTICS

The basic goals of a clinical microbiology laboratory are to establish the presence of a pathogen in a clinical sample, to identify the pathogen, and, when possible, to provide other information that can help guide clinical management and affect prognosis, such as antibiotic susceptibility profiles or the presence of virulence factors. To date, clinical microbiology laboratories have largely approached these goals phenotypically by growth-based assays and biochemical testing. Bacteria, for instance, were historically grouped algorithmically into species by their characteristic microscopic appearance, nutrient requirements for growth, and ability to catalyze certain reactions. Antibiotic susceptibility is still determined in most cases by assessing bacterial growth in the presence of antibiotic.

With the sequencing revolution paving the way to easy access of complete pathogen genomes, we can more systematically define the genetic basis for these observable phenotypes. Compared with traditional growth-based methods for bacterial diagnostics that dominate the clinical microbiology laboratory, nucleic acid–based diagnostics that build on this genomic information promise improved speed, sensitivity, specificity, and breadth of information. Bridging clinical and research laboratories, adaptations of genomic technologies have begun to deliver on this promise **(Table 126-2)**.

■ HISTORICAL LIMITATIONS AND PROGRESS THROUGH GENETIC APPROACHES

The molecular diagnostics revolution in the clinical microbiology laboratory is well under way, born of necessity in the effort to identify and characterize microbes that are refractory to traditional culture methods. Historically, diagnosis of many so-called unculturable pathogens has relied largely on serology and antigen detection. However, these methods provide only limited clinical information because of their suboptimal sensitivity and specificity, and, for serology, the long delays that diminish their utility for real-time patient management and the inability to characterize pathogens beyond identifying past exposure. Newer tests to detect pathogens based on nucleic acid content have already offered improvements in the select cases in which they have been applied.

Unlike direct pathogen detection, serologic diagnosis—measurement of the host's response to pathogen exposure—can typically be made only in retrospect, requiring both acute- and convalescent-phase serum samples. For chronic infections, distinguishing active from latent infection or identifying repeat exposure from serology alone can be difficult or impossible, depending on the syndrome. In addition, serologic diagnosis is variably sensitive, depending on the organism and the patient's immune status. For instance, tuberculosis is notoriously difficult to identify by serologic methods; tuberculin skin testing using purified protein derivative (PPD) is especially insensitive in

TABLE 126-1 Glossary of Selected Terms in Genomics

TERM	DEFINITION
Contig	A DNA sequence representing a continuous fragment of a genome, assembled from overlapping sequences; relevant for de novo assembly of sequence data that do not align to previously sequenced genomes
Genome	The entire set of heritable genetic material within an organism
Horizontal gene transfer	The transfer of genes between organisms through mechanisms other than by clonal descent, such as through transformation, conjugation, or transduction
Genomic epidemiology	An approach to inferring and reconstructing transmission events, population dynamics, and epidemiological patterns using comparative analysis of microbial genome sequences
Metagenomics	Analysis of genetic material from multiple species directly from primary samples without requiring prior culture steps
Microarray	A collection of DNA oligonucleotides ("oligos") spatially arranged on a solid surface and used to detect or quantify sequences in a sample of interest that are complementary (and therefore bind) to one or more of the arrayed oligos
Microbial genome-wide association study (GWAS)	An analytic framework to test statistical associations between microbial genotypes and phenotypes of interest, such as antibiotic resistance and virulence
Mobile genetic elements	DNA elements that can move within a genome and can be transferred between genomes through horizontal gene transfer (e.g., plasmids, bacteriophages, and transposons)
Multilocus sequence typing	A method for typing organisms based on DNA sequence fragments from a prespecified set of genes
Next-generation sequencing	High-throughput sequencing using a parallelized sequencing process that produces millions of sequences concurrently, far beyond the capacity of prior dye-terminator methods
Nucleic acid amplification test (NAAT)	A biochemical assay that evaluates for the presence of a particular string of nucleic acids through amplification by one of several methods, including polymerase and ligase chain reactions
Polymerase chain reaction (PCR)	A type of NAAT used to amplify a specific region of DNA by means of specific oligonucleotide primers and a DNA polymerase
Single-nucleotide polymorphism (SNP)	Point mutations in DNA, the number of which in different microbial isolates is a measure of their genetic distance from one another
Transcriptome	The catalog of the full set of messenger RNA (mRNA) transcripts from a cell or organism, which are typically measured by microarray or by next-generation sequencing of complementary DNA (cDNA) via a process called RNA-Seq
Whole genome sequencing (WGS)	A process that determines the full DNA sequence of an organism's genome; has been greatly facilitated by next-generation sequencing technology

active disease and possibly cross-reactive with vaccines or other mycobacteria. Even the newer interferon γ release assays (IGRAs), which measure cytokine release from T lymphocytes in response to *Mycobacterium tuberculosis*–specific antigens in vitro, have limited sensitivity in immunodeficient hosts. Neither PPD testing nor IGRAs can distinguish latent from active infection. Serologic Lyme disease diagnostics suffer similar limitations: in patients from endemic regions, the presence of IgG antibodies to *Borrelia burgdorferi* may reflect prior exposure rather than active disease, while IgM antibodies are imperfectly sensitive and specific (50% and 80%, respectively, in early disease). The complicated nature of these tests, particularly in view of the nonspecific symptoms that may accompany Lyme disease, has had substantial implications for public perception of Lyme disease and antibiotic misuse in endemic areas. Similarly, syphilis, a chronic infection caused by *Treponema pallidum*, is notoriously difficult to stage by serology alone, requiring multiple different nontreponemal and treponemal tests (e.g., rapid protein reagin and fluorescent treponemal antibody, respectively) in conjunction with clinical suspicion. Complementing serology, antigen detection can improve sensitivity and specificity in select cases but has been validated only for a limited set of infections. Typically, structural elements of pathogens are detected, including components of viral envelopes (e.g., hepatitis B surface antigen, HIV p24 antigen, SARS-CoV-2 spike protein), cell surface markers in certain bacteria (e.g., *Streptococcus pneumoniae*, *Legionella pneumophila* serotype 1) or fungi (e.g., *Cryptococcus*, *Histoplasma*), and less specific fungal cell-wall components such as galactomannan and β-glucan (e.g., *Aspergillus* and other fungi).

Given the impracticality of culture and the lack of sensitivity or sufficient clinical information afforded by serologic and antigenic methods, the push toward nucleic acid–based diagnostics originated in pursuit of viruses and fastidious bacteria, becoming part of the standard of care for select organisms in U.S. hospitals. Such tests, including polymerase chain reaction (PCR) and other nucleic acid amplification tests (NAATs), are now widely used for many viral infections, both chronic (e.g., HIV infection, hepatitis C) and acute (e.g., influenza, SARS-CoV-2, respiratory syncytial virus). NAATs provide essential information about both the initial diagnosis and the response to therapy and, in some cases, genotypically predict drug resistance. Indeed, progression from antigen detection to PCR transformed our understanding of the natural course of HIV infection, with profound implications for treatment **(Fig. 126-1*A*)**. In the early years of the AIDS pandemic, p24 antigenemia was detected in acute HIV infection but then disappeared for years before emerging again with progression to AIDS **(Fig. 126-1*B*)**. Without a marker demonstrating viremia, the role of treatment during HIV infection prior to the development of clinical AIDS was uncertain, and assessing treatment efficacy was challenging. With the emergence of PCR as a progressively more sensitive test (now able to detect as few as 20 copies of virus per milliliter of blood), viremia was recognized as a near-universal feature of HIV infection. Given the challenges of phenotypic assays, genotypic antiviral resistance testing was also adopted early for HIV and is now the standard of care before the initiation of therapy in developed countries. These developments have been transformative in guiding therapy in early disease and, together with the development of less toxic therapies, have helped to shape policy that is moving toward ever-earlier introduction of antiretroviral therapy in HIV infection. Reverse transcriptase PCR (RT-PCR) assays are the core method for detecting SARS-CoV-2 virus in the acute phase, forming a critical component of the clinical and public health response to COVID-19, just as they were on a smaller scale for the related coronaviruses SARS-CoV and Middle East respiratory syndrome (MERS)-CoV. Tests for SARS-CoV-2 represent the largest implementation of a molecular infectious disease assay to date and play a critical role in both clinical diagnostics and public health measures to contain the COVID-19 pandemic.

As they are for viral testing, nucleic acid–based tests have become the diagnostic tests of choice for fastidious bacteria, including the common sexually transmitted bacterial pathogens *Neisseria gonorrhoeae* and *Chlamydia trachomatis* as well as the tick-borne *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum*. More recently, nucleic acid amplification–based detection has offered improved sensitivity for diagnosis of the important nosocomial pathogen *Clostridioides difficile*, and NAATs have provided clinically relevant information on the presence of cytotoxins A and B as well as molecular markers of hypervirulence, such as the North American pulsotype 1 (NAP1) strain that is enriched in severe illness. The importance of genomics in selecting loci for diagnostic assays and in monitoring test sensitivity was highlighted by the emergence in Sweden of a newly recognized variant of *C. trachomatis* with a deletion that includes the gene targeted by a set of commercial NAATs. By evading detection through this deletion (and thus avoiding treatment), this strain came to be highly prevalent in some areas of Sweden. While nucleic acid–based tests remain the diagnostic approach of choice for fastidious bacteria, this example of "diagnostic escape" serves as a reminder of the need for careful development and ongoing monitoring of molecular diagnostics.

In contrast, for typical bacterial pathogens for which culture methods are well established, growth-based assays still dominate in the

TABLE 126-2 Selected Clinical Applications of Infectious Disease Genomics

APPLICATION	TECHNOLOGY	NOTES/EXAMPLES
Organism Identification		
Viral detection	PCR, RT-PCR	Identification of HIV, HBV, HCV, respiratory viruses including SARS-CoV-2 and influenza, and others for diagnosis and response to therapy
TB detection	PCR	Amplification of the *rpoB* gene for species-specific identification of *Mycobacterium tuberculosis*
Pathogen detection	PCR, RT-PCR, NAAT	Multiplexed identification of dozens of viruses, bacteria, yeasts, and parasites from a variety of clinical specimens
Bacterial detection	16S ribosomal gene sequencing	Targeted amplification and sequencing of regions of the 16S rRNA gene for identification of suspected bacterial infections undiagnosed by conventional methods
Pathogen detection	Cell-free DNA sequencing	Unbiased amplification and sequencing of cell-free nucleic acid from blood, with analytical comparison of resulting nonhuman sequences to genomes of known pathogens and contaminants, in order to identify circulating pathogen DNA; anecdotal clinical use to establish etiology of systemic or focal infection, though clinical utility and optimal use cases still evolving
Pathogen Discovery		
Bacterial pathogens	Sequencing, metagenomic assembly	Unbiased "shotgun" sequencing of isolated nucleic acid from patient samples to identify associated pathogens; proofs-of-concept: new *Bradyrhizobium* species associated with cord colitis; *Escherichia coli* 0104:H4 from 2011 diarrheal outbreak in Germany; *Leptospira* species from one patient's cerebrospinal fluid; research use only at this time
Viral pathogens	Microarray, sequencing	Hybridization of clinical samples to microarrays from phylogenetically diverse known viruses identified the first SARS coronavirus and others. Direct sequencing has identified SARS-CoV-2, West Nile virus, and MERS-CoV, among others. Use is primarily in research.
Antibiotic Resistance		
MRSA detection	PCR	Detection of the *mecA* gene, the genotypic cause of methicillin resistance in *Staphylococcus aureus*
VRE detection	PCR	Detection of the *vanA* or *vanB* gene, the main genotypic causes of vancomycin resistance in *Enterococcus*
MDR-TB detection	PCR, NAAT	Detection of polymorphisms in the *rpoB* gene from *M. tuberculosis*, which account for 95% of rifampin resistance. Other probes available for *inhA* and *katG* genes can detect up to 85% of isoniazid resistance.
Carbapenemase detection	PCR	Detection of genes encoding one of several types of enzymes (KPC, NDM, OXA-48, IMP, VIM) that hydrolyze carbapenems, accounting for much but not all carbapenemase resistance in Enterobacteriaceae
HIV resistance detection	Targeted sequencing	Targeted sequencing of specific genes with known resistance-conferring mutations; now the standard of care prior to initial therapy in the United States and Europe
Epidemiology		
Outbreak and epidemic tracking	Sequencing	Application to tracking outbreaks and epidemics on local and international scales, including spread of carbapenemase-producing *Klebsiella*, *S. aureus*, *M. tuberculosis*, *E. coli*, *Vibrio cholerae*, Ebola virus, Zika virus, and influenza virus
Evolution and spread of pathogens	Sequencing	Sequencing collections of pathogens from individual patients or environmental reservoirs such as wastewater to shed light on pathogen dissemination, virulence factors, and antibiotic resistance determinants; innumerable examples, including *V. cholerae*, influenza virus, Ebola virus, Zika virus, and SARS-CoV-2

Abbreviations: HBV, hepatitis B virus; HCV, hepatitis C virus; MDR, multidrug-resistant; MERS, Middle East respiratory syndrome; MRSA, methicillin-resistant *S. aureus*; NAAT, nucleic acid amplification test; PCR, polymerase chain reaction; RT, reverse transcriptase; SARS, severe acute respiratory syndrome; TB, tuberculosis; VRE, vancomycin-resistant enterococci.

clinical laboratory. Informed by decades of clinical microbiology, these tests have served clinicians well, yet the limitations of growth-based tests—in particular, the delays associated with waiting for growth—have left opportunities for improvements. Driven by this need, mass spectrometry–based assays that offer highly accurate organism identification within a few hours of a positive blood culture are widely adopted in clinical microbiology laboratories, largely supplanting biochemical tests for organism identification in well-resourced areas. Looking ahead, molecular diagnostics, greatly informed by the vast quantity of microbial genome sequences generated in recent years, offers a way forward. First, sequencing studies can readily identify key genes (or noncoding nucleic acids) that can be developed into targets for clinical assays using PCR or hybridization assay platforms. Second, sequencing itself is becoming cheap and rapid enough to be performed on clinical specimens in certain cases, with consequent unbiased detection or characterization of pathogens.

One of the biggest drivers for the implementation of novel molecular technologies in the diagnosis of infectious diseases is the desire for more rapid—or even real-time—pathogen identification, ideally with antibiotic susceptibility information on those microbes for which resistance to the current anti-infective armamentarium is of concern. Such real-time tests have the potential to transform infectious disease management, impacting antibiotic stewardship in the outpatient setting, mortality risk in the critically ill (i.e., patients in whom early administration of effective antibiotics is the most significant factor in decreasing mortality risk), hospital admission, and length of hospital stay; the extent of this impact will depend on the economic forces that will help define the breadth of their deployment. On the public health level, such tests will likely play a role in improving antibiotic stewardship, thereby influencing the rise of antibiotic resistance and enabling surveillance of outbreaks by local, national, and international networks. In the United States and the United Kingdom, for example, public health agencies use genome sequencing to track food-borne pathogens and identify outbreaks and have rapidly expanded the routine use of genomics in identifying and characterizing other pathogens, from mycobacteria (both *M. tuberculosis* and nontuberculous mycobacteria) to *N. gonorrhoeae*. Further, international efforts to track the spread of viral diseases, particularly the vast efforts to sequence SARS-CoV-2 and monitor the emergence and spread of its variants, as well as recent work on mpox, Ebola, and Zika outbreaks and ongoing work on seasonal influenza, offer opportunities for improving interventions, surveillance, and prevention efforts, ranging from more accurate selection of the strains to include in vaccine development to improved design of trials to evaluate novel vaccines and therapies.

Technological innovations are lowering several critical barriers to the widespread adoption of genomics and other molecular methods. Specifically, for NAAT, the need for rapid thermal cycling and cold-chain storage for reagents has significantly impeded implementation

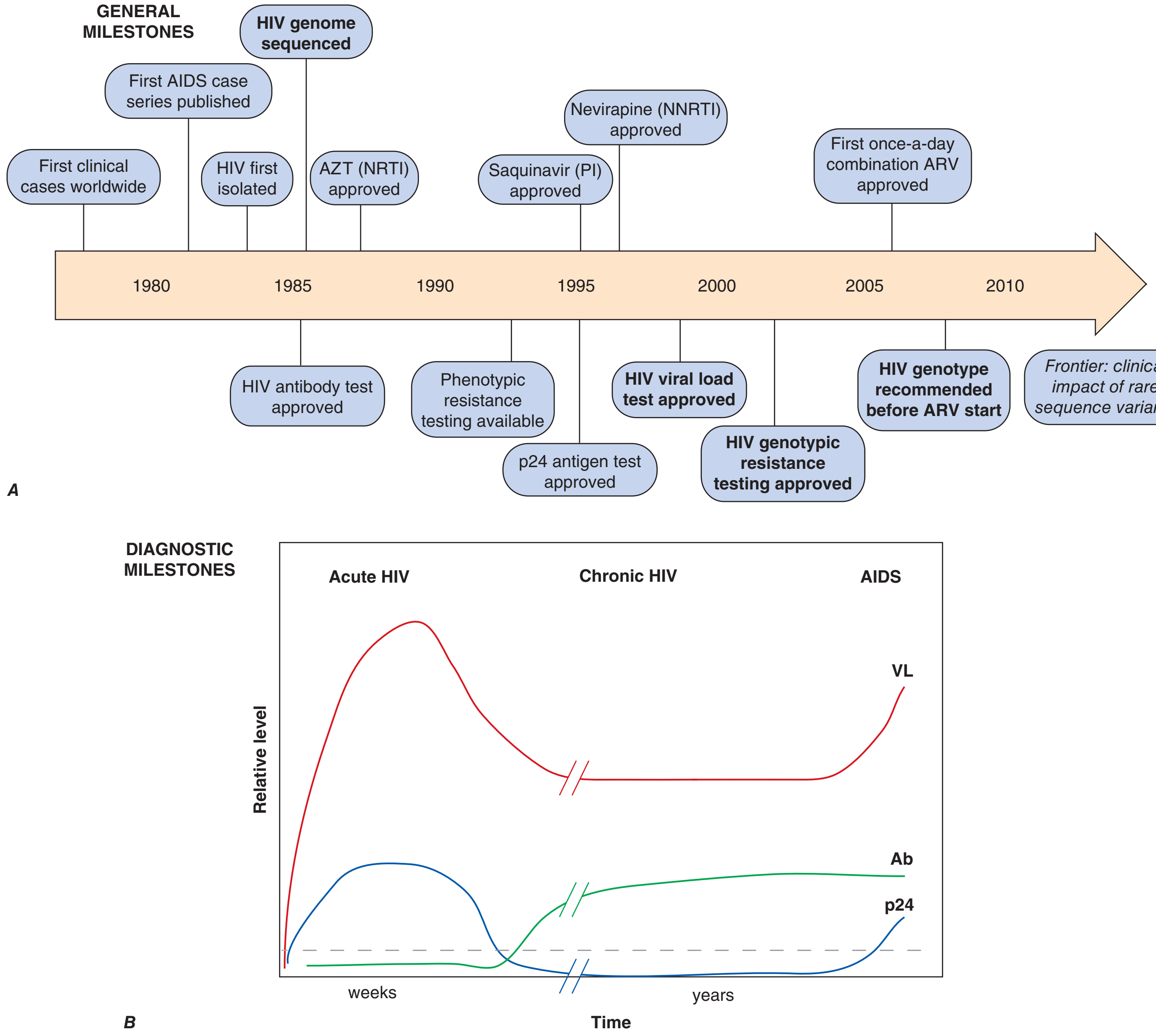

FIGURE 126-1 ***A.*** Timeline of select milestones in HIV management. Genomic advances are shown in bold type. The approvals and recommendations indicated apply to the United States. ARV, antiretroviral; AZT, zidovudine; NRTI, nucleoside reverse transcriptase (RT) inhibitor; NNRTI, nonnucleoside RT inhibitor; PI, protease inhibitor. ***B.*** Viral dynamics in the natural history of HIV infection. Three diagnostic markers are shown: HIV antibody (Ab), p24 antigen (p24), and viral load (VL). *Dashed gray line* represents limit of detection. *(Adapted from data in HH Fiebig et al: Dynamics of HIV viremia and antibody seroconversion in plasma donors: Implications for diagnosis and staging of primary HIV infection. AIDS 17:1871, 2003.)*

in resource-limited settings. Recent efforts aim to overcome these challenges by developing isothermal amplification protocols and lyophilized reagents that do not require refrigeration or sophisticated instrumentation. For clinical sequencing, (1) the cost and speed of sequencing and analysis methods continue to fall precipitously; (2) automation and miniaturization of the preparation of a sample for sequencing promise to reduce cost and minimize the expertise needed; and (3) direct sequencing technologies that eliminate the complex molecular biology required to prepare clinical samples for sequencing are improving in accuracy and robustness. Further barriers exist, including the need for careful processing of clinical samples to minimize contamination, and standardized pipelines to process data and present clinicians with easily interpretable and readily actionable results. However, as these advances give rise to rapid, accurate diagnostic tests, the ultimate goal is to inform a clinician in real time whether antibiotics are indicated and, if so, which will be effective. Real-time diagnostics will allow more efficient deployment of our precious antibiotic arsenal, thus improving both societal and patient-specific outcomes in much the same way that a rapid, sensitive troponin assay has transformed bedside management of chest pain.

■ ORGANISM IDENTIFICATION

To adapt nucleic acid detection to diagnostic tests and thus to identify pathogens on a wide scale, sequences must be found that are conserved enough within a species to identify the diversity of strains that may be encountered in various clinical settings, but divergent enough to distinguish one species from another. Until recently, this problem has been solved for bacteria by targeting the element of a bacterial genome that is most highly conserved within a species, the 16S ribosomal RNA (rRNA) subunit. Among many examples, this method has now been used to confirm *Mycobacterium chimaera* infections in several patients after cardiothoracic surgery, leading ultimately to recognition of a widespread outbreak. At present, 16S PCR amplification from tissue specimens can be performed by specialty laboratories, though its sensitivity and clinical utility to date have remained somewhat limited, in part because of the scarcity and relative fragility of pathogen nucleic acid in the sampled tissue, which necessitates reliable, sensitive nucleic acid amplification. As such barriers are reduced through technological advances and as the causes of culture-negative infection are clarified (perhaps in part through sequencing efforts), these tests may become both more accessible and more helpful.

With the wealth of sequencing data now available, other regions beyond 16S rRNA can be targeted for bacterial species identification. These other genomic loci can provide additional information about a clinical isolate that is relevant to patient management. For instance, detection of the presence, or potentially even the expression, of toxin genes such as *C. difficile* toxins A and B or Shiga toxin can provide clinicians with additional information that will help distinguish commensals or colonizing bacteria from pathogens and thus aid in prognostication and management as well as in diagnosis.

Beyond bacteria, one commonly used approach to PCR-based pathogen detection is so-called "syndromic panels" of multiplexed PCR to identify common causes of clinical infection syndromes, including upper respiratory infection, gastroenteritis, and meningoencephalitis. The most frequently deployed syndromic panel is the respiratory viral panel, which typically includes primer sets targeting a combination of influenza, parainfluenza, respiratory syncytial virus, adenovirus, rhinovirus, enterovirus, metapneumovirus, and common-cold coronaviruses, sometimes in conjunction with unculturable bacteria such as *Mycoplasma*, *Chlamydophila*, and *Bordetella* species. With the advent of the COVID-19 pandemic, SARS-CoV-2 was quickly added to this multiplex PCR panel within months of its development as a stand-alone PCR test. The goal of such panels is to capture common infectious causes of these syndromes in a single, standardized diagnostic test, ideally streamlining the diagnostic evaluation. The ready identification of a plausible etiologic agent may offer diagnostic clarity if judiciously used and carefully considered in the clinical context of each patient. The widespread adoption of SARS-CoV-2 PCR testing in the earliest days of the COVID-19 pandemic underscored the crucial role of precise etiologic pathogen diagnosis in patient care, triage, infection control, and epidemiology.

One challenge with PCR-based assays is the relative complexity of the molecular biology and consequent need for advanced technology for implementation, including instruments and reagents. Several recent approaches have advanced the molecular biology of nucleic acid detection with the aim of increasing deployability of NAATs for use in resource-limited or even field settings. These methods couple nucleic acid detection to an enzymatic readout, enabling catalytic signal amplification. Several such approaches build on the intrinsic sensitivity, specificity, and amplification of the CRISPR (*c*lustered, *r*egularly *i*nterspaced *s*hort *p*alindromic *r*epeats) effectors Cas12a and Cas13a as nucleic acid sensors. Distinct from the famous gene editing CRISPR effector Cas9, these robust and versatile enzymes recognize short nucleic acid targets with high specificity and transduce this binding event into "collateral cleavage" of nearby nucleic acids that can be engineered to create a signal using fluorescent reporter constructs. Crucially, this biotechnology can be made to work in conjunction with isothermal enzymatic preamplification steps to achieve remarkable sensitivity, all robustly enough to withstand lyophilization that enables long-term storage before being reconstituted in the field. Such assays are still in the early stages of development, but they have shown promise and could play a critical role in global diagnostics and surveillance.

While amplification tests such as PCR and other NAATs exemplify one approach to nucleic acid detection, other approaches exist, including detection by hybridization. Although not currently used in the clinical realm, techniques for multiplexed detection and identification of pathogens by hybridization to microarrays or in solution are being developed for other purposes. Of note, these different detection techniques require different degrees of conservation. Highly sensitive amplification methods require a high degree of sequence identity between PCR primer pairs and their short, specific target sequences; even a single base-pair mismatch (particularly near the 3′ end of the primer) may interfere with detection. In contrast, hybridization-based tests are more tolerant of mismatch and thus can be used to detect important regions that may be less precisely conserved within a species, thus potentially allowing detection of clinical isolates from a given species with greater diversity between isolates. Such assays take advantage of the predictable binding interactions of nucleic acids and do not require enzymology, broadening the range of conditions under which such assays are feasible, including directly on primary clinical specimens. The applicability of hybridization-based methods toward either DNA or RNA opens the possibility of expression profiling, which can uncover phenotypic information from nucleic acid content.

Both PCR and hybridization methods target specific, known organisms. At the other extreme, as sequencing costs decline, metagenomic sequencing from patient samples is increasingly feasible. This shotgun sequencing approach is unbiased—i.e., can detect any microbial sequence, however divergent or unexpected. In one example, a clinical sample of cerebrospinal fluid from an immunocompromised patient with signs and symptoms of chronic meningitis was found through metagenomic sequencing and analysis to contain small amounts of *Leptospira* DNA. In light of this information, retrospective PCR testing confirmed the diagnosis of neuroleptospirosis, which had been missed prior to the sequencing result. The patient was treated with penicillin G and clinically recovered. Increasingly, efforts are under way to bring whole genome sequencing to other clinical samples, including sputum and blood, to more readily identify pathogens. One such assay certified for clinical use in the United States—a shotgun metagenomic sequencing approach applied to cell-free DNA circulating in the bloodstream that aims to identify pathogens both in blood and other body sites—is gaining adoption in cases where traditional culture-based methods fail to identify a pathogen in patients with symptoms consistent with infection. This new approach brings its own set of challenges, however, including the need to recognize pathogenic sequences against a background of expected host and commensal sequences and to distinguish true pathogens from either colonizers or laboratory contaminants. The burgeoning field of microbiome research is driving technology development for sequencing and analyzing complex microbial communities. Lessons from this field will inform diagnostic efforts.

PATHOGEN DISCOVERY

In addition to clinical diagnostic applications, novel genomic technologies, including whole genome sequencing, are being applied to clinical research specimens with a goal of identifying new pathogens in a variety of circumstances. The tremendous sensitivity and unbiased nature of sequencing is also ideal in searching clinical samples for unknown or unsuspected pathogens.

Causal inference in infectious diseases has progressed since the time of Koch, whose historical postulates provided a rigorous framework for attributing a disease to a microorganism. To modernize Koch's postulates, an organism, whether it can be cultured or not, should induce disease upon introduction into a healthy host if it is to be implicated as a causative pathogen. Current sequencing technologies are ideal for advancing this modern version of Koch's postulates because they can identify candidate causal pathogens with unprecedented sensitivity and in an unbiased way, unencumbered by limitations such as culturability. Yet, as direct sequencing on primary patient samples greatly expands our ability to recognize associations between microbes and disease states, critical thinking and experimentation will remain vital in establishing causality.

Virus discovery in particular has been greatly facilitated by new nucleic acid technology. These frontiers were first notably explored with high-density microarrays containing spatially arrayed sequences from a phylogenetically diverse collection of viruses. Despite bias toward those with homology to known viruses, novel viruses in clinical samples were successfully identified on the basis of their ability to hybridize to these prespecified sequences. This methodology famously contributed to identification of the coronavirus causing severe acute respiratory syndrome (SARS). Once discovered, the SARS coronavirus was rapidly sequenced: the full genome was assembled in April 2003, <6 months after recognition of the first case.

With the advent of next-generation sequencing, unbiased pathogen discovery is now addressed through a process known as *metagenomic assembly* **(Fig. 126-2)**, largely supplanting other methods. Sequences of random nucleotide fragments can be generated from clinical specimens with no a priori knowledge of pathogen identity through a process called *shotgun sequencing*. This collection of sequences can then be computationally aligned to host (i.e., human) sequences, with aligned sequences removed and remaining sequences compared with other

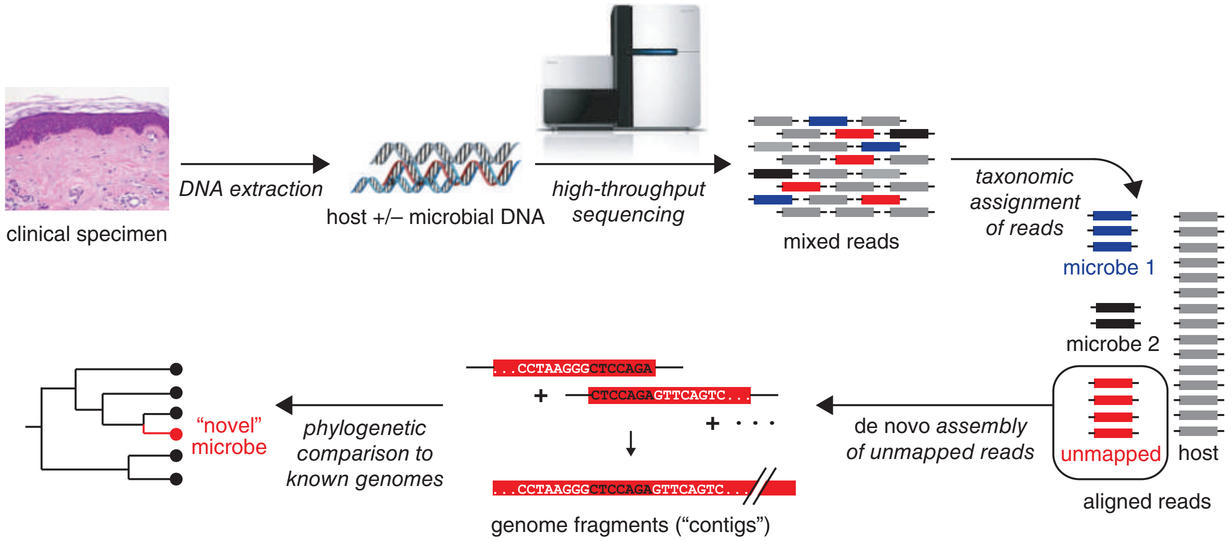
clinical specimen
DNA extraction
host +/– microbial DNA
high-throughput sequencing
mixed reads
taxonomic assignment of reads
microbe 1
microbe 2
unmapped
host
aligned reads
de novo assembly of unmapped reads
...CCTAAGGGCTCCAGA
+
CTCCAGAGTTCAGTC...
+ . . .
...CCTAAGGGCTCCAGAGTTCAGTC...
genome fragments (“contigs”)
phylogenetic comparison to known genomes
“novel” microbe

FIGURE 126-2 Workflow of metagenomic assembly for pathogen discovery. DNA is isolated from a specimen of interest (e.g., tissue, body fluid) containing a mixture of host DNA and nucleic acids from coexisting microbes, either commensal or pathogenic. All DNA (and RNA, if a reverse transcription step is added) is then sequenced, yielding a mixture of DNA sequence fragments (“reads”) from the organisms present. Except for reads that do not align (“map”) to any known sequence, these reads are aligned to existing reference genomes for the host or any known microbes. The unmapped reads are computationally assembled de novo into the largest contiguous stretches of DNA possible (“contigs”), representing fragments of previously unsequenced genomes. These genome fragments (contigs) are then mapped onto a phylogenetic tree based on their sequence. Some may represent known but as-yet-unsequenced organisms, while others will represent novel species. *(Figure prepared with valuable input from Dr. Ami S. Bhatt, personal communication.)*

known genomes to detect the presence of known microorganisms. Sequence fragments that remain unaligned suggest the presence of an additional organism that cannot be matched to a known, characterized genome; these reads can be assembled into contiguous nucleic acid stretches that can be compared with known sequences to construct the genome of a potentially novel organism. Assembled genomes (or parts of genomes) can then be compared with known genomes to infer the phylogeny of new organisms and identify related classes or traits. Thus, not only can this process identify unanticipated pathogens, but it can even identify undiscovered organisms.

The emergence of COVID-19 provides a dramatic example that illustrates advances in pathogen discovery technology in the intervening 16 years since SARS-CoV was discovered: the causal coronavirus, SARS-CoV-2, was identified through metagenomic sequencing within about 1 month of the first known case and just weeks after the outbreak was first recognized. Sequencing and assembly were completed within 5 days of the discovery of the new virus, and a NAAT was released 1 day later. Given the ensuing ravages of COVID-19 and the cost of delays of even a few weeks in implementing this new diagnostic test in some locations, it is sobering to contemplate the added harm had this outbreak occurred even a decade earlier. This timeline illustrates the advancing power and speed of new diagnostic technologies but also underscores the pressing need for continued progress.

As metagenomic sequencing and assembly techniques become more robust, this technology holds great promise for identifying microorganisms that are associated with clinical conditions of unknown etiology. Conventional methods already have unexpectedly linked numerous conditions with specific agents of infection—e.g., cervical and oropharyngeal cancers with human papillomavirus (HPV), Kaposi’s sarcoma with human herpesvirus 8, and certain lymphomas and, more recently, multiple sclerosis with Epstein-Barr virus. Recently, Zika virus, first described in the 1940s, was found to be increasing in incidence as a cause of febrile syndromes, particularly in Central and South America. A concurrent increase in the incidence of microcephaly was noted that temporally and geographically matched the Zika epidemics. Zika was suspected to be neurotropic because of a previously recognized association with Guillain-Barré syndrome, but the strongest link between Zika virus and microcephaly came when the virus itself was detected by both quantitative reverse transcription PCR (RT-qPCR) and whole genome sequencing in postmortem fetal brain tissue from microcephalic infants. An argument for causality was built on the foundation of epidemiologic evidence and direct viral detection, both of which were built on nucleic acid detection and genome sequencing. Sequencing techniques offer unprecedented sensitivity and specificity for identifying foreign nucleic acid sequences that may suggest other such pathogen-associated conditions—from malignancies to inflammatory conditions to unexplained fevers or other clinical syndromes—associated with organisms from viruses to bacteria to parasites. Caution is needed, though: in the absence of the ability to fulfill Koch’s postulates, sequence-based identification of a microbe from patient specimens is not, on its own, sufficient to ascribe pathogenicity. The increasing sensitivity of these methods warrants greater rigor and care in defining what is “noise” and what represents a pathogen.

As sequencing-based discovery expands, microbes may be found to be associated with conditions not classically thought of as infectious, such as the link between maternal Zika virus infection and fetal microcephaly. Studies of bowel flora in laboratory animals and even humans already suggest correlations between microbe composition and various aspects of metabolic and cardiovascular health. Improved methods for pathogen detection will continue to uncover unexpected correlations between microbes and disease states, but the mere presence of a microbe does not establish causality. Fortunately, once the relatively laborious and computationally intensive metagenomic sequencing and assembly efforts have identified a pathogen, further detection can more easily be undertaken with targeted methods such as PCR or hybridization, which may be more scalable and amenable to in situ confirmation. This capacity should facilitate the additional careful investigation that will be required to progress beyond correlation and to draw causal inference.

ANTIBIOTIC RESISTANCE

At present, antibiotic resistance in bacteria and fungi is conventionally determined by isolating a single colony from a cultured clinical specimen and testing its growth in the presence of drug. The requirement for multiple growth steps in these conventional assays has several consequences. First, only culturable pathogens can be readily processed. Second, this process requires considerable infrastructure to support the sterile environment needed for culture-based testing of diverse organisms. Finally, and perhaps most significantly, even the fastest-growing organisms require 1–2 days of processing for identification and 2–3 days for determination of susceptibilities. Some slow-growing organisms take even longer; for instance, weeks must pass before drug-resistant *M. tuberculosis* can be identified by growth phenotype. Given the clinical imperative in serious illness to begin effective therapy early, this inherent delay in susceptibility determination has obvious implications for empirical antibiotic use: broad-spectrum antibiotics often must be chosen up front in situations where it is later shown that preferred narrower-spectrum drugs would have been effective or even that no antibiotics were appropriate (i.e., in viral infections). Even with this strategy, as resistant organisms become more common, the empirical choice can be incorrect, often with devastating consequences. Real-time identification of the infecting organism and information on its susceptibility profile would guide initial therapy and support judicious antibiotic use, ideally improving patient outcomes while aiding in the ever-escalating fight against antibiotic resistance by reserving the use of broad-spectrum agents for cases in which they are truly needed.

Molecular diagnostics and sequencing offer a way to accelerate detection of a pathogen's antibiotic susceptibility profile. If a genotype that confers resistance can be identified, this genotype can be targeted for molecular detection. In infectious disease, this approach has most comprehensively come to fruition for HIV (Fig. 126-1*A*). (In a conceptually parallel application of genomic analysis, molecular detection of certain resistance determinants in cancers informs selection of targeted chemotherapy.) Extensive sequencing of HIV strains and correlations drawn between viral genotypes and phenotypic resistance have delineated the majority of mutations in key HIV genes, such as reverse transcriptase, protease, and integrase, that confer resistance to the antiretroviral agents that target these proteins. For instance, the single amino acid substitution K103N in the HIV reverse transcriptase gene predicts resistance to the first-line nonnucleoside reverse transcriptase inhibitor efavirenz, and its detection informs a clinician to choose a different agent. The effects of these common mutations on HIV susceptibility to various drugs—as well as on viral fitness—are curated in publicly available databases. Thus, genotypes are now routinely used to predict drug resistance in HIV, as phenotypic resistance assays are far more cumbersome than targeted sequencing. Though it was not implemented at the level of individual patients, sequencing-based detection of circulating SARS-CoV-2 variants predicted susceptibility to monoclonal antibodies due to changes in the spike protein and, thus, informed decisions made by the U.S. Food and Drug Administration (FDA) on granting and rescinding approval for such COVID-19 therapies as the pandemic progressed. As new therapies are introduced, this targeted sequencing–based approach to drug resistance will likely prove important in other viral infections.

The challenge of predicting drug susceptibility from genotype is more daunting for bacteria than for HIV, yet considerable progress has been made toward sequencing-based determination of bacterial antibiotic susceptibility. Bacteria have far more complex genomes than viruses, with thousands of genes on their chromosomes (many of which can functionally interact in ways that escape a priori prediction) and the capacity to acquire many more through horizontal gene transfer of plasmids and mobile genetic elements within and between species. Thus, the task of comprehensively defining all possible genetic resistance mechanisms is orders of magnitude more complex in bacteria than in viruses, which typically have far more limited genomes. Despite these challenges, considerable progress has been made in recent years. In select cases where biological factors appear to have constrained the genotypic basis for resistance to a small, well-defined set of mutations, genotypic assays for antibiotic resistance are already being integrated into clinical practice. One important example is the detection of methicillin-resistant *Staphylococcus aureus* (MRSA). *S. aureus* is one of the most common and serious bacterial pathogens of humans, particularly in health care settings. Resistance to methicillin—the most effective class of antistaphylococcal antibiotics—is common, even in community-acquired strains. Vancomycin—an alternative drug to methicillin—is effective against MRSA but is measurably inferior to methicillin against methicillin-susceptible *S. aureus* (MSSA). Analysis of clinical MRSA isolates has demonstrated that the molecular basis for resistance to methicillin in essentially all cases stems from the expression of an alternative penicillin-binding protein (PBP2A) encoded by the gene *mecA*, which is found within a transferable genetic element called *mec*. This mobile cassette has spread rapidly through the *S. aureus* population via horizontal gene transfer and selection from widespread antibiotic use. Because methicillin resistance is essentially always due to the presence of the *mecA* gene, MRSA is particularly amenable to molecular detection. A PCR test for the *mecA* gene, which saves hours to days compared with standard culture-based methods, has been approved by the FDA to augment (but not replace) traditional culture-based susceptibility testing. Similar to MRSA, vancomycin-resistant enterococci (VRE) harbor one of a limited number of *van* genes found to be responsible for resistance to this important antibiotic, which occurs through alteration of the mechanism for cell wall cross-linking that vancomycin inhibits. Detection of one of these genes by PCR indicates resistance. More recently, multiplexed PCR assays targeting carbapenemase genes (those encoding the KPC, NDM, OXA-48, IMP-1, and VIM carbapenemases), which are responsible for a significant fraction of carbapenem resistance (though not all instances), can predict some resistance to this crucial antibiotic class, though they are not comprehensive enough to confirm susceptibility if absent. Despite these caveats, such assays to detect even a limited set of high-value resistance genes are gaining use in high-resourced settings. Finally, a PCR assay targeting the highly conserved RNA polymerase gene serves not only to detect *M. tuberculosis* directly in sputum samples but also to detect resistance to rifampin, since the determinants of resistance to this RNA polymerase inhibitor map almost exclusively to a short region of this gene. Since rifampin resistance is epidemiologically associated with, though not causal for, multidrug resistance, this assay identifies strains at high risk for multidrug resistance, enhancing its value. This test has transformed tuberculosis testing where available, improving sensitivity and providing a limited measure of susceptibility testing to guide therapy.

Although identification and rapid detection of monogenic resistance determinants have improved, bacteria typically evolve multiple, diverse resistance mechanisms to most antibiotics; thus, outside of these edge cases, resistance prediction often requires probing for and integration of multiple genetic lesions, targets, or mechanisms. For instance, at least five distinct modes of resistance to fluoroquinolones are known: reduced import, increased efflux, target site mutation, drug modification, and shielding of the target sites by expression of another protein. These mechanisms are typically present in combination in clinically resistant isolates; thus, the problem of detecting genetic resistance is often a combinatorial one. In another clinically important example, while carbapenem resistance in Enterobacteriaceae is often explained by the presence of carbapenemases, resistance may also develop when other, less broad-spectrum β-lactamases are found in combination with porin mutations or upregulated efflux pumps. These more complex mechanisms prevent PCR-based carbapenemase detection assays from identifying other mechanisms of carbapenem resistance. Additionally, plasmids and transposable elements, which often are enriched for antibiotic resistance determinants, may be more technically and analytically challenging to sequence, although newer long-read sequencing technologies are beginning to address these challenges. To further complicate genetic prediction, changes in gene expression (which may be detectable through mutations in promoter regions or regulatory genes without coding mutations in known resistance determinants) and even gene copy number (which may occur without changes in primary sequence) of resistance determinants play critical roles in some cases of genetic resistance. Thus, while predicting resistance when determinants are found is rapidly becoming feasible, the more clinically relevant task of predicting *susceptibility* when no known resistance determinants are found remains more difficult.

To build on early successes with the goal of advancing beyond binary detection of monogenic resistance determinants, the ultimate frontier for genetic prediction of bacterial antibiotic resistance lies in more comprehensive prediction of a resistance phenotype from sequence information—a task similar to HIV resistance prediction. Yet there is no comprehensive compendium of genetic elements conferring resistance and their pairwise and higher-order interactions with each other and with the genetic background of bacterial pathogens. Nonviral genomes are much larger than viral ones, and their abundance and diversity are such that thousands of genetic differences often exist between clinical isolates of the same species, of which perhaps only one or a few may contribute to resistance. In addition, new mechanisms may emerge in the face of antibiotic deployment or with the release of new drugs, and genetic prediction of resistance will inevitably lag behind the emergence of unforeseen mechanisms. While confident prediction of bacterial antibiotic resistance from sequencing determinants may therefore seem daunting, the vast expansion of microbial sequencing capacity, combined with analytic methods such as microbial genome-wide association studies and machine learning algorithms, offers powerful analytical approaches to this "needle in a haystack" problem and has permitted remarkable advances in the predictive power of sequence determinants to date. Particularly in *M. tuberculosis*, where horizontal gene transfer is minimal and the pathogen is essentially restricted to human hosts to facilitate more representative sampling, a remarkable fraction of phenotypic resistance can be explained by known genetic determinants. Because of these biologic advantages, as well as the slow and laborious growth process that impedes traditional phenotypic assessment, whole genome sequencing has proven quite effective at predicting susceptibility profiles in this organism, to the point that the United Kingdom now routinely performs whole genome sequencing in parallel with phenotypic antibiotic susceptibility testing for *M. tuberculosis* in what some hope will be a precursor to fully whole genome sequencing–based antibiotic susceptibility testing. Even in more highly variable pathogens, with sequencing of sufficient numbers of susceptible and resistant pathogens, and more sophisticated analytical algorithms, sequence-based prediction methods are improving in predictive accuracy, at least within the geographic region from which the test samples have been sequenced.

It is important to note that genotype-based analytical methods largely identify correlates, not necessarily surrogates or determinants, of resistance. In HIV diagnostics, surrogates (i.e., causal determinants of resistance) were found to be more reliable predictors than mere correlates in expanding sequencing-based resistance prediction to the general population. Without a mechanistic understanding of genetic resistance, a correlative relationship may be lineage-specific and less generalizable. Especially with multiple possible mechanisms of resistance to a given antibiotic and ongoing evolutionary pressure resulting in the development and acquisition of new modes of resistance, a genotypic approach to diagnosing antibiotic resistance is likely to remain challenging and to require ongoing vigilance in constantly correlating genotypic with more traditional phenotypic methods. An important corollary benefit of a genomic approach to resistance prediction, anchored in phenotypic validation, could be the systematic identification of outliers with unexplained resistance. These strains can form the basis for understanding newly emerging resistance mechanisms, which can in turn inform new drug development endeavors. Understanding resistance mechanisms may also help direct infection control efforts. For instance, the first identification of the *mcr-1* (mobilized colistin resistance) gene on a plasmid, together with other antibiotic resistance determinants, heightened concern about colistin-resistant Enterobacterales identified first in China and later elsewhere because it implied transferrable multidrug resistance. Early recognition of these potentially dangerous strains elucidated the immediate need for strict containment protocols. Furthermore, metagenomic studies have highlighted the silent carriage of resistance genes in the microbiomes of returning travelers and their subsequent spread to household contacts.

In parallel with advancing sequencing technologies, progress in computational techniques, bioinformatics and statistics, and data storage, as well as experimental confirmatory testing of hypotheses, will be needed to advance toward the ambitious goal of a comprehensive compendium of global antibiotic resistance determinants. Open sharing and careful curation of new sequence information will be of paramount importance, as will iterative or even continuous comparison of predictions with ongoing phenotypic testing to assess performance and allow prediction algorithms to keep up with newly evolving or emerging resistance mechanisms.

We continuously observe the accumulation of new or unanticipated modes of resistance from ongoing evolutionary pressure caused by the widespread clinical use of antibiotics. Even with MRSA, perhaps the best-studied case of antibiotic resistance and a model of relative simplicity with a single known monogenic resistance determinant (*mecA*), a genotype-based approach to resistance detection proved imperfect. One limitation was a recall of the initial commercial genotypic resistance assay that was deployed for the identification of MRSA. A clinical isolate of *S. aureus* that emerged in Belgium expressed a variant of the *mec* cassette not detected by the assay's PCR primers. New primers were added to detect this new variant, and the assay was reapproved for use. This example illustrates the need for ongoing monitoring of any genotypic resistance assay. A second limitation is that a contradiction can occur between genotypic and phenotypic evidence for resistance. Up to 5% of MSSA strains have been reported to carry a copy of the *mecA* gene that is either nonfunctional or not expressed. Thus, the erroneous identification of these strains as MRSA by genotypic detection would lead to administration of the inferior antibiotic vancomycin rather than the preferred β-lactam therapy.

These examples illustrate one of the prime challenges of moving beyond growth-based assays: genotype is merely a proxy for the resistance phenotype that directly informs patient care. Alternative approaches currently under development attempt to circumvent the limitations of genotypic resistance testing by returning to phenotypic assays, albeit more rapid ones. One such approach is informed by genomic methods: transcriptional profiles serve as a rapid phenotypic signature for antibiotic response. Conceptually, since dying cells are transcriptionally distinct from cells fated to survive, susceptible bacteria enact different transcriptional profiles after antibiotic exposure than resistant ones, independent of the mechanism of resistance. These differences can be measured and, since transcription is one of the most rapid responses to cell stress (minutes to hours), can be used to determine whether cells are resistant or susceptible much more rapidly than is possible if growth in the presence of antibiotics is awaited (days). Like DNA, RNA can be readily detected through predictable rules governing base pairing via either amplification or hybridization-based methods. Changes in a carefully selected set of transcripts form an expression signature that can represent the total cellular response to antibiotic without requiring full characterization of the entire transcriptome. Preliminary proof-of-concept studies suggest that this approach may identify antibiotic susceptibility based on transcriptional phenotype much more quickly than is possible with growth-based assays. Other rapid phenotype-based approaches to antibiotic susceptibility testing, including automated microscopy, ultrafine measurements of mass fluctuations, and others, are under development as well, with the former approved for clinical use.

Because of its sensitivity in detecting even very rare nucleic acid fragments, sequencing provides an unprecedented depth of study into complex populations of cells and tissues. The strength of this depth and sensitivity applies not only to the detection of rare, novel pathogens in a sea of host signal but also to the identification of heterogeneous pathogen subpopulations in a single host that may differ, for example, in drug resistance profiles or pathogenesis determinants. For instance, recent studies have highlighted the diversification of pathogens in chronic bacterial infections, such as *Pseudomonas* in the lungs of patients with cystic fibrosis or *M. tuberculosis* in disseminated infection, perhaps allowing for niche specialization within the host. Such diversification has long been recognized in chronic viral populations, as exemplified by HIV. Future studies will be needed to elucidate the clinical significance of these variable subpopulations, even as deep sequencing is now providing unprecedented levels of detail about majority and minority members of this population.

HOST-BASED DIAGNOSTICS

While pathogen-based diagnostics continue to be the mainstay for confirming infection, serologic testing and nonspecific biomarkers—such as erythrocyte sedimentation rate, C-reactive protein level, and even total white blood cell and neutrophil counts—have long been the basis of a strategy for measuring host responses to aid in the diagnosis of infection. Even recently identified host biomarkers of bacterial infection, such as procalcitonin, have fallen short in their versatility, with positive and negative predictive values that are thus far adequate for only a few narrow applications but inadequate for generalized clinical use. Here, too, the application of genomics is now being explored to improve upon this approach, given the previously described limitations of serologic testing and the lack of specificity of protein biomarkers identified to date. Rather than using antibody responses as a retrospective biomarker for infection, recent efforts have focused on transcriptomic analysis of the host response as a new direction with diagnostic implications for human disease.

For instance, while pathogen-based diagnostic tests to distinguish active from latent tuberculosis infection have proven elusive, the transcriptional profile of circulating white blood cells exhibits a differential pattern of expression of nearly 400 transcripts that distinguish active from latent tuberculosis; this expression pattern is driven in part by changes in interferon-inducible genes in the myeloid lineage. In a validation cohort, this transcriptional signature was able to distinguish patients with active versus latent disease, to distinguish tuberculosis infection from other pulmonary inflammatory states or infections, and to track responses to treatment in as little as 2 weeks, with normalization of expression toward that of patients without active disease over 6 months of effective therapy. Such a test could play an important role not only in the management of patients but also as a marker of efficacy in clinical trials of new therapeutic agents. More recently, a distilled three-transcript signature has shown promise for distinguishing active from latent tuberculosis, raising hopes of a deployable assay in the near term.

Similarly, considerable progress has been made toward identifying host transcriptional signatures in circulating blood cells that distinguish viral from bacterial causes of upper respiratory infection, with better performance characteristics than current clinical parameters or available protein biomarkers. Additional host signatures have been reported that distinguish among bacterial infection, viral infection, and inflammatory states; identify Lyme disease; identify influenza; and even distinguish between gram-positive and gram-negative bacterial infections. In some cases, results have been extended to different host populations—including adults and children, and those with varying immune function—which obviously will be critical for generalizing such an approach. Thus, profiling of host transcriptional dynamics could augment the information obtained from studies of pathogens, both enhancing diagnosis and monitoring the progression of illness and the response to therapy. The frontier of genomic applications to understand host response to infection, with the potential of identifying biomarkers or even underlying disease biology, continues to rapidly advance, incorporating novel technological and computational approaches such as single-cell host transcriptional profiling of infected patients, to understand complex processes such as sepsis.

In this era of genome-wide association studies and attempts to move toward personalized medicine, genomic approaches are also being applied to the identification of host genetic loci and factors that contribute to infection susceptibility. Such loci will have undergone strong selection among populations in which the disease is endemic. Through identification of the beneficial genetic alleles among individuals who survive in such settings, markers for susceptibility or resistance are being discovered; these markers can be translated to diagnostic tests to identify susceptible individuals to implement preventive or prophylactic interventions. Further, such studies may offer mechanistic insight into the pathogenesis of infection and inform new methods of therapeutic intervention. Such beneficial genetic associations were recognized long before the advent of genomics, as in the protective effects of the negative Duffy blood group or heterozygous hemoglobin abnormalities against *Plasmodium* infection. Genomic approaches allow more systematic and widespread application of this principle to identify not only people with increased susceptibility to prevalent diseases (e.g., HIV infection, tuberculosis, and cholera) but also host factors that contribute to and thus might predict the severity of disease. In one recent example, polymorphisms in certain genetic loci, or specific circulating autoantibodies, were found to be associated with severe COVID-19.

THERAPEUTICS

Genomics has the potential to impact infectious disease therapeutics in two ways. By transforming the speed or type of diagnostic information that can be attained, it can influence therapeutic decision-making. Alternatively, by opening new avenues to a better understanding of pathogenesis, providing new ways to disrupt infection, and delineating new approaches to antibiotic discovery or characterization of vaccine targets, it has the potential to facilitate the development of new therapeutic agents.

GENOMIC DIAGNOSTICS INFORMING THERAPEUTICS

Efforts at antibiotic discovery are declining, with few new agents in the pipeline and even fewer new drugs (in particular, few agents with new mechanisms of action) entering the market. This phenomenon is due in part to the lack of economic incentives for the private sector; however, it is also attributable in part to the enormous challenges involved in the discovery and development of antibiotics. Most recent efforts have focused on broad-spectrum antibiotics; the development of a chemical entity that works across an extremely diverse set of organisms (i.e., species more divergent from each other than a human is from an amoeba) is far more challenging than the development of an agent that is designed to target a single bacterial species. Nevertheless, the concept of narrow-spectrum antibiotics has heretofore been rejected because of the lack of early diagnostic information that would guide the selection of such agents. Thus, rapid diagnostics providing antibiotic susceptibility information that can guide antibiotic selection in real time has the potential to alter and simplify antibiotic strategies by allowing a paradigm shift away from broad-spectrum drugs and toward narrow-spectrum agents. Such a paradigm shift clearly would have additional implications for antibiotic resistance, helping to limit selective pressure applied to pathogens and commensal bacteria during therapy.

In yet another diagnostic paradigm with the potential to impact therapeutic interventions, genomics is opening new avenues to a better understanding not only of different host susceptibilities to infection but also of different host responses to therapy. For example, the role of glucocorticoids in tuberculous meningitis has long been debated. Recently, polymorphisms in the human genetic locus *LTA4H*, which encodes a leukotriene-modifying enzyme, were found to modulate the inflammatory response to tuberculosis. Patients with tuberculous meningitis who were homozygous for the proinflammatory *LTA4H* allele were most helped by adjunctive glucocorticoid treatment, while those who were homozygous for the anti-inflammatory allele were negatively affected by steroid treatment. Steroids have become part of the standard of care in tuberculous meningitis, but this study suggests that perhaps only a subset of patients benefit from this anti-inflammatory adjunct (while others may be harmed) and further suggests a genetic means of prospectively identifying this subset. Thus, genomic diagnostic tests may eventually approach the goal of personalized medicine, informing diagnosis, prognosis, and treatment decisions by revealing the pathogenic potential of the microbe and by detecting individualized host responses to both infection and therapy.

GENOMICS IN DRUG AND VACCINE DEVELOPMENT

Genomic technologies are dramatically changing research on host–pathogen interactions, with a goal of increasingly influencing the process of therapeutic discovery and development. Sequencing offers several possible avenues into antimicrobial therapeutic discovery. First, genome-scale molecular methods have paved the way for comprehensive identification of all essential genes encoded by a pathogen, thereby systematically identifying critical vulnerabilities within a pathogen that could be targeted therapeutically. Second, genome-scale methodologies offer rapid ways to address the mechanism of action of newly identified hits from compound screens. Whole genome sequencing offers a rapid, unbiased way to detect mutations arising in resistant mutants during selection. Similarly, transcriptional profiling can provide insights into mechanisms of action of new candidate drugs. For instance, the transcriptional signature of cell wall disruptors (e.g., β-lactams) is distinct from that of DNA-damaging agents (e.g., fluoroquinolones) or protein synthesis inhibitors (e.g., aminoglycosides). Either approach can thus suggest a mechanism of action or flag compounds for prioritization because of a potentially novel activity. In an alternative strategy for determining mechanisms of action, genome-wide RNA interference or CRISPR screens can be used to identify genes required for antimicrobial efficacy. This approach provided new insights into the mechanism of action of drugs that have been in use for decades for human African trypanosomiasis. Third, sequencing can readily identify the most conserved regions of a pathogen's genomes and corresponding gene products; this information is invaluable in narrowing antigen candidates in vaccine development. These surface proteins can be expressed recombinantly and tested for the ability to elicit a serologic response and protective immunity. This process, termed *reverse vaccinology*, has proved particularly useful for pathogens that are difficult to culture or poorly immunogenic. After decades of development, the utility of this approach became dramatically apparent with the rapid development of mRNA vaccines targeting conserved regions of the SARS-CoV-2 genome, fueling the most rapid development of a vaccine in history. Comparative genomics informed by prior coronavirus sequences enabled the first mRNA vaccine design to begin within days of the first SARS-CoV-2 sequence being made publicly available, and now annual updates are guided by ongoing sequencing of circulating variants, as discussed in detail below.

Genomics has been employed in both developing vaccines and defining their impact on microbial epidemiology and ecology. Examples include recent studies of influenza, malaria, *S. pneumoniae*, and HPV following vaccine introduction. Extensive sequencing of influenza viruses has been valuable in understanding the modest efficacy of seasonal influenza vaccination, and the combination of genomics and antigenic cartography is used to select strains to include in subsequent influenza vaccines. Beyond this, sequence conservation informs efforts to design more robust pan-influenza or pan-coronavirus vaccines. The RTS,S/AS01 malaria vaccine was analyzed by targeted sequencing of parasites from vaccinated and control populations during a phase 3 trial conducted at 11 sites in Africa; these analyses revealed reduced vaccine efficacy against parasites with amino acid mutations in the circumsporozoite protein targeted by the vaccine. Similarly, studies of the more established pneumococcal vaccines (the 7- and 13-valent polysaccharide conjugate vaccines, PCV-7 and PCV-13) documented serotype replacement: strains targeted by the vaccine have dramatically decreased in prevalence following widespread vaccination campaigns. Given that specific serotypes of HPV (e.g., types 16 and 18) clearly are more strongly associated than others with carcinogenesis, HPV vaccines have capitalized on serotype replacement, targeting vaccine strains to specifically prevent infection with the more dangerous serotypes. Such a strategy, informed by pathogen genomics, aims to protect individuals and ideally to decrease the circulating burden of more virulent strains within society.

Large-scale gene content analysis from sequencing or expression profiling enables new research directions that provide novel insights into the interplay of pathogen and host during infection or colonization. One important goal of such research is to suggest new therapeutic approaches to disrupt this interaction in favor of the host. Indeed, one of the most immediate applications of next-generation sequencing technology has come from simply characterizing human pathogens and related commensal or environmental strains and then finding genomic correlates for pathogenicity. For instance, as *Escherichia coli* varies from a simple nonpathogenic, lab-adapted strain (K-12) to a Shiga toxin–producing enterohemorrhagic gastrointestinal pathogen (O157:H7), it displays up to a 25% difference in gene content, though it is classified as the same species. Similarly, some isolates of *Enterococcus*—a genus notorious for its increasing incidence of resistance to common antibiotics such as ampicillin, vancomycin, and aminoglycosides—also contain recently acquired genetic material comprising up to 25% of the genome on mobile genetic elements. This fact suggests that horizontal gene transfer plays an important role in the organisms' adaptation as nosocomial pathogens. On closer study, this genome expansion is associated with loss of CRISPR elements, which protect the bacterial genome from invasion by certain foreign genetic material, and may thus facilitate the acquisition of antibiotic resistance–conferring genetic elements. While loss of this regulation appears to impose a competitive disadvantage in antibiotic-free environments, these drug-resistant strains thrive in the presence of even some of the best antienterococcal therapies. In addition to insights gained from genome sequencing, transcriptomic and proteomic profiling of pathogens under various conditions that mimic colonization or infection, including existence as biofilms or in polymicrobial communities, intracellular infection models, antibiotic exposure, and nutrient starvation, has begun to reveal novel biologic features that may be targeted by the next generation of therapies. At the cutting edge of the host–pathogen interface, single-cell transcriptomic methodologies are rapidly increasing in feasibility and extent, revealing previously unknown heterogeneity in the potential outcomes of intracellular infection.

Thus, genomic studies are transforming our understanding of infection, offering evidence of virulence factors or toxins and providing insight into ongoing evolution of pathogenicity and drug resistance. One goal of such studies is to identify therapeutic agents that can disrupt the pathogenic process. There is currently much interest in the theoretical concept of antivirulence drugs that inhibit virulence factors rather than killing the pathogen outright as a means to intervene in infection. Further, with sequencing ever more accessible and efficient, ongoing large-scale studies have unprecedented statistical power to associate clinical outcomes with pathogen and host genotypes and thus to further reveal vulnerabilities in the infection process that can be targeted for disruption. Although this is just the beginning, such studies point to a tantalizing future in which the clinician is armed with genomic predictors of infection outcome and therapeutic response to guide clinical decision-making.

EPIDEMIOLOGY OF INFECTIOUS DISEASES

Epidemiologic studies of infectious diseases have several main goals: to identify and characterize outbreaks, to describe the pattern and dynamics of an infectious disease as it spreads through populations, and to identify interventions that can limit or reduce the burden of disease. One classic, paradigmatic example is John Snow's elucidation of the origin of the 1854 London cholera outbreak. Snow used careful geographic mapping of cases to determine that the likely source of the outbreak was contaminated water from the Broad Street pump, and by removing the pump handle, he aborted the outbreak. Whereas that effort was undertaken without knowledge of the causative agent of cholera, advances in microbiology and genomics have expanded the purview of epidemiology to consider not just the disease but also the pathogen and its genetic variants, its virulence factors, and the complex relationships between microbial and host populations.

Through use of genomic tools such as high-throughput sequencing, the diversity of a microbial population can be rapidly described with unprecedented resolution, with discrimination between isolates that have single-nucleotide differences across the entire genome and advancement beyond prior approaches that relied on phenotypes (such as antibiotic susceptibility profiles) or genetic markers (such as multilocus sequence typing). The development of statistical methods grounded in molecular genetics and evolutionary theory has established analytical approaches that translate descriptions of microbial population diversity and structure into descriptions of the origin and history of pathogen spread. By linking phylogenetic reconstruction with epidemiologic and demographic data, genomic epidemiology presents the opportunity to track transmission from person to person and across demographic and geographic boundaries, to infer transmission patterns of both pathogens and sequence elements that confer phenotypes of interest, and to estimate the transmission dynamics of outbreaks.

■ TRANSMISSION NETWORKS

Whole genome sequencing of pathogen genomes can be used to infer transmission and identify point-source outbreaks. As reported in a seminal paper in 2010, a study of MRSA in a Thai hospital demonstrated the use of whole genome sequencing in reconstructing the transmission of a pathogen from patient to patient by integrating the analysis of accumulation of mutations over time with the dates and hospital locations of the infected individuals. Since then, multiple instances of the use of whole genome sequencing to define and motivate interventions aimed at interrupting transmission chains have been reported. In another MRSA outbreak in a special-care baby unit in Cambridge, United Kingdom, whole genome sequencing extended the traditional infection control analysis, which relies on typing organisms by their antibiotic susceptibilities, to sequencing of isolates from clinical samples. This approach identified an otherwise unrecognized outbreak of a specific MRSA strain that was occurring against a background of the usual pattern of infection caused by a diverse circulating population of MRSA strains. The analysis showed evidence of transmission among mothers within the special-care baby unit and in the community and demonstrated the key role of MRSA carriage in a single health care provider in the persistence of the outbreak. In yet another example, in response to the observation of 18 cases of infection by carbapenemase-producing *Klebsiella pneumoniae* over 6 months at the National Institutes of Health Clinical Research Center, genome sequencing of the isolates was used to discriminate between the possibilities that these cases represented multiple, independent introductions into the health care system or a single introduction with subsequent transmission. On the basis of network and phylogenetic analysis of genomic and epidemiologic data, the authors reconstructed the likely relationships among the isolates from patient to patient, demonstrating the nosocomial spread of a single resistant *Klebsiella* strain. Similar approaches have elucidated the extent to which presumed nosocomial *C. difficile*, VRE, and carbapenem-resistant Enterobacterales represent within-hospital transmission—of bacterial strains or of plasmids and mobile genetic elements—versus independent acquisitions of unrelated pathogens. With these demonstrations of the potential contribution of genomics to hospital infection-control efforts, an important avenue of research seeks to develop statistical methods with which to ascertain when such tools are useful and their cost-effectiveness when compared with that of current nongenomic approaches.

Genome sequencing of clinical specimens of viruses has been used to understand their patterns of spread and the clinical and epidemiologic implications of genetic variants. As RNA viruses use an error-prone RNA-dependent RNA polymerase, they accumulate mutations at a rapid rate, facilitating inferences about the dynamics and patterns of spread. These tools have been applied to the study of outbreaks of well-known viruses, such as recent outbreaks of yellow fever in South America and mumps in the United States, as well as recent zoonotic pathogens, such as the coronaviruses MERS-CoV and SARS-CoV. The sequencing of SARS-CoV-2 in the context of the pandemic has offered a powerful example of the contributions that genomic epidemiology can make to, and its increasingly central role in, tracking the spread of a pathogen both locally and globally and informing policy and public health decision-making. Moreover, tracking variants and combining the genome sequences with epidemiologic line-list data enable investigation of the extent to which new variants cause differing symptom profiles and levels of severe disease.

The uncovering of unexpected transmission events by genomic epidemiology studies is motivating investigations into pathogen ecology and modes of transmission. Whole genome sequencing established the clonality of several high-profile outbreaks, enabling the discovery of dangerous contaminants such as *Burkholderia pseudomallei* in aromatherapy spray, *Pseudomonas* in eyedrops, *Exserohilum* in injectable corticosteroids, *Fusarium* in epidural anesthetic preparations, *E. coli* O157:H7 in beef, and *Mycobacterium chimaera* in the temperature-control systems used during cardiac bypass. Each of these outbreaks caused considerable morbidity and in some cases mortality before being localized and prevented on the basis of investigations informed and accelerated by genomic epidemiology.

As more studies aim to carefully define the origins and spread of infectious agents using the high-resolution lens of whole genome sequencing, fundamental questions arise about the diversity of infecting and colonizing microbial populations. Traditional microbiologic methods include taking a single colony from a growth plate as representative of the population. However, the more diverse the colonizing or infecting pathogen population, the less representative these individual isolates are and the greater the possibility for introducing error into whole genome sequencing–based methods while reconstructing transmission. Sequencing studies of multiple colonies of an *S. aureus* strain colonizing a single individual showed a "cloud" of diversity. What is the clinical significance of this diversity? What are the processes that generate and limit it? What amount of diversity is transmitted under different conditions and routes of transmission? How do the answers to these questions vary by infectious organism, type of infection, host, and response to treatment? More comprehensive descriptions of diversity, population dynamics, transmission bottlenecks, and the forces that shape and influence the growth and spread of microbial populations will be a critically important focus of future investigations.

ORIGINS AND DYNAMICS OF PATHOGEN SPREAD

In addition to reconstructing the transmission chains of local outbreaks, genomics-based epidemiologic methods reveal broad-scale geographic and temporal spread of pathogens. Four examples include the origins of cholera in Haiti, the history of HIV-1 group M, the spread of Ebola in West Africa, and the timing and nature of spread of the zoonotic COVID-19 pandemic, which brought genomic tracking of pathogen variants to the forefront of public consciousness for a time. Cholera, a dehydrating diarrheal illness caused by infection with *Vibrio cholerae*, first spread worldwide from the Indian subcontinent in the 1800s and has since caused seven pandemics; the seventh pandemic has been ongoing since the 1960s. An investigation into the geographic patterns of cholera spread in the seventh pandemic used genome sequences from a global collection of 154 *V. cholerae* strains representing isolates from 1957 to 2010. This investigation revealed that the seventh pandemic has comprised at least three overlapping waves spreading out from the Indian subcontinent **(Fig. 126-3*A*)**. Further, analysis of the genome of an isolate of *V. cholerae* from the 2010 outbreak of cholera in Haiti showed it to be more closely related to isolates from South Asia than to isolates from neighboring Latin America, supporting the hypothesis that the outbreak was derived from *V. cholerae* introduced into Haiti by human travel (likely from Nepal) rather than by environmental or more geographically proximal sources. A subsequent study that dated the time to the most recent common ancestor of a population of *V. cholerae* isolates from Haiti provided further support for a single point-source introduction from Nepal. Application of similar methods that integrate pathogen genome sequences, mutation rates, geographic locations, and phylogenetic inference to HIV-1 group M dated the origin of the virus to the 1920s and the city of Kinshasa (then called Leopoldville), the capital of the Democratic Republic of the Congo (then called the Belgian Congo). This work established an understanding of how a boom in industry and a city with extensive railroad connections provide a scaffolding along which a virus can rapidly spread geographically.

Genome sequencing has proven invaluable in understanding the geographic, demographic, climatic, and administrative factors that drove, sustained, and limited the 2013–2016 Ebola outbreak that ravaged West Africa **(Fig. 126-3*B*)** as well as the factors and patterns of transmission of Zika virus in the Americas and most recently the timing and origins of SARS-CoV-2 transmission in human populations. With the rapid availability of the SARS-CoV-2 viral genome sequence, data from a set of cases from early in the pandemic enabled inference of the time to the most recent common ancestor, supporting that SARS-CoV-2 entered circulation in human populations in Wuhan, China, sometime in late November to early December of 2019. Subsequently, large, coordinated sequencing networks have been able to recreate its patterns of global spread, discover new variants, and monitor as these variants disseminate. For the first time, scientists and even the public were able to watch viral evolution through a population in almost real time. First, a single-nucleotide polymorphism (D614G in the viral spike protein) displaced the original strain in early 2020, followed in subsequent years by other variants such as Alpha, Delta, Omicron, and then various Omicron sublineages becoming dominant either

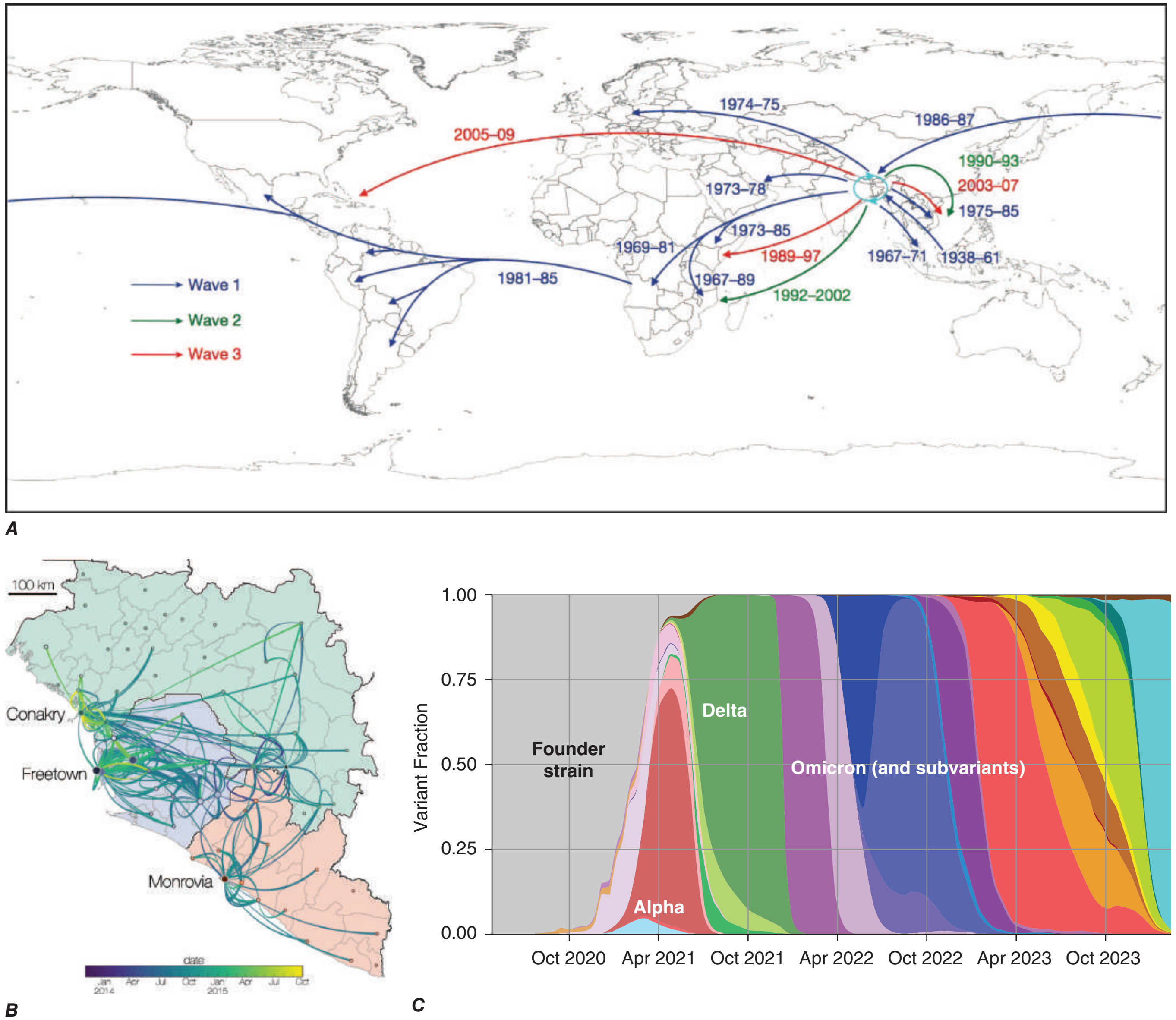

FIGURE 126-3 ***A.*** Transmission events inferred from phylogenetic reconstruction of 154 *Vibrio cholerae* isolates from the seventh cholera pandemic. Date ranges represent estimated time to the most recent common ancestor for strains transmitted from source to destination locations, based on a Bayesian model of the phylogeny. *(Reprinted by permission from the Nature Publishing Group, Nature 477:462. Evidence for several waves of global transmission in the seventh cholera pandemic, A Mutreja et al © 2011.)* ***B.*** Inferred Ebola virus spread in West Africa (Liberia, *red*; Guinea, *green*; and Sierra Leone, *blue*) by phylogeographic methods using virus genome sequences, dates, and an evolutionary model. The lines reflect spread between population centroids of each administrative region, going from the thin end to the thick end and colored by a time scale. *(Reprinted by permission from Nature Publishing Group, Nature 544:309. Virus genomes reveal factors that spread and sustained the Ebola epidemic, G Dudas et al © 2017.)* ***C.*** SARS-CoV-2 variant proportions in the United States from the start of the COVID-19 pandemic through March 6, 2024. Each major variant family that rose to dominance in the United States is labeled. *(Variant proportions are taken from https://covariants.org, accessed April 21, 2024, which derives underlying variant data from the Global Initiative on Sharing All Influenza Data [GISAID].)*

regionally or globally **(Fig. 126-3*C*)** as their mutation profile conferred advantages in increasingly immune populations.

One compelling area of innovation in molecular epidemiology is the development of wastewater-based surveillance. Wastewater surveillance has been used to monitor for the appearance of poliovirus, among others, but its widespread and widely publicized use for SARS-CoV-2 in "nowcasting" quantitative estimates of epidemic trends has galvanized interest in using wastewater to monitor disease trends and pathogens and their characteristics. The U.S. Centers for Disease Control and Prevention has sought to establish a national wastewater surveillance system, building on the SARS-CoV-2 work. These tools can help provide a population-level view of the first appearance of a pathogen, variant, or antibiotic resistance determinant and can also help monitor the population burden and hence the epidemic curve. Work remains to develop robust, repeatable, and accurate methods across many pathogens, particularly those that can replicate or exchange DNA in wastewater environments. Nonetheless, this data source has great promise for providing early signals to decision-makers to help inform public health actions.

These efforts illustrate the remarkable promise of genome sequencing in improving outbreak response strategies by elucidating previously hidden origins and paths of disease spread and details of the forces that shape epidemics. The combination of in-the-field sequencing with portable sequencing platforms, rapid data sharing, and rapid open analysis through sites such as *nextstrain.org* offers a paradigm by which real-time genomic epidemiology, including wastewater-based epidemiology, may contribute to "weather maps," enabling prediction of epidemic patterns in space and time and thus providing guidance for public health interventions to slow or control their spread.

Increasing numbers of investigations into the spread of many pathogens are contributing to a growing atlas of maps describing routes, patterns, and tempos of microbial diversification and dissemination, not just for agents of emerging infectious diseases but for common pathogens as well. Such studies will create a vast amount of data that can be used to investigate the diversity and microbiologic links within distinct niches and the patterns of spread from one niche to another. The increasingly broad adoption of genome sequencing by health care and public health institutions ensures that the available catalog of genome sequences and associated epidemiologic data will grow very rapidly. For example, updating from the pulsed-field gel electrophoresis techniques that have been used to define strains of food-borne pathogens since the late 1980s, PulseNet—the U.S. Centers for Disease Control and Prevention network for monitoring these pathogens—has instituted routine genome sequencing. The COVID-19 pandemic further underscores the importance of building a new global public health infrastructure in which sequencing plays a central role to facilitate early disease discovery, rapid and close tracking of spread, and development of diagnostics and targeted effective interventions. With higher-resolution description of microbial diversity and of the dynamics of that diversity over time and across epidemiologic and demographic boundaries and evolutionary niches, we will gain even greater insights into the relationships of transmission routes and patterns of historical spread.

EPIDEMIC POTENTIAL

Defining pathogen transmissibility is a critical step in the development of public health surveillance and intervention strategies because this information can help to predict the epidemic potential of an outbreak. Transmissibility can be estimated by a variety of methods, including inference from the growth rate of an epidemic and the generation time of an infection (the mean interval between infection of an index case and infection of the people infected by that index case). Genome sequencing and analysis of a well-sampled population provide another method by which to derive similar fundamental epidemiologic parameters. One key measure of transmissibility is the basic reproduction number, defined as the number of secondary infections generated from a single primary infectious case. When the basic reproduction number is >1, an outbreak has epidemic potential; when it is <1, the outbreak will become extinct. On the basis of sequences from influenza virus samples obtained from infected patients very early in the 2009 H1N1 influenza pandemic, the basic reproduction number was estimated through a population genomic analysis at 1.2; this result provided greater confidence to estimates derived by traditional epidemiologic data, which ranged from 1.4 to 1.6. In addition, with the assumption of a molecular clock model, sequences of H1N1 samples together with information about when and where the samples were obtained have been used to estimate the date and location of the pandemic's origin, providing insight into disease origins and dynamics. Integrating viral genomics with other types of data—such as the timing and nature of mitigation efforts and the impact of those efforts on mobility—will expand the toolkit with which to assess the impact of public health interventions on slowing and controlling disease spread. These tools may also be applied to institutional infection control: with the development of return-to-work protocols, sequencing offers one option to help learn the extent to which infections arose from within-institution spread. Because the magnitude and intensity of the public health response are guided by the predicted size of an outbreak, the ability of genomic methods to cast light on a pathogen's origin and epidemic potential adds an important dimension to the contributions of these methods to infectious disease epidemiology.

PATHOGEN EVOLUTION

Beyond describing transmission and dynamics, pathogen genomics can provide insight into the evolution of pathogens and the interactions of selective pressures, the host, and pathogen populations, which can have implications for clinical decision-making and the development of vaccines and therapeutics. From a clinical perspective, this process is central to the acquisition of antibiotic resistance, the generation of increasing pathogenicity or new virulence traits, the evasion of host immunity and clearance (leading to chronic infection), and vaccine efficacy.

Microbial genomes evolve through a variety of mechanisms, including mutation, duplication, insertion, deletion, recombination, and horizontal gene transfer. Segmented viruses (e.g., influenza virus) can reassort gene segments within multiply infected cells. The pandemic 2009 H1N1 influenza A virus, for example, appears to have been generated through reassortment of several avian, swine, and human influenza strains. Such potential for the evolution of novel pandemic strains has precipitated concern about the possible evolution to transmissibility of virulent strains that have been associated with high mortality rates but have not yet exhibited efficient human infectivity. Experiments with H5N1 avian influenza, for example, have defined five mutations that render it transmissible, at least in ferrets—the animal model system for human influenza. Studies that examine the genomes of pathogens collected longitudinally from individual infections have similarly demonstrated their evolution as they adapt to host environments and new immune and therapeutic pressures.

The continuous antigenic evolution of seasonal influenza offers an example of how studies of pathogen evolution can impact surveillance and vaccine development. Frequent updates to the annual influenza vaccine are needed to ensure protection against the dominant strains. These updates are based on anticipating which viral populations from a pool of substantial locally and globally diverse circulating viruses will predominate in the upcoming season. Toward that end, sequencing-based studies of influenza virus dynamics have shed light on the global spread of influenza, providing concrete data on patterns of spread and helping to elucidate the origins, emergence, and circulation of novel strains. Through analysis of >1000 influenza A H3N2 virus isolates over the 2002–2007 influenza seasons, Southeast Asia was identified as the usual site from which diversity originates and spreads worldwide. Further studies of global isolate collections have shed further light on the diversity of circulating virus, showing that some strains persist and circulate outside of Asia for multiple seasons. Similar studies with SARS-CoV-2, including genome sequences from orders of magnitude more clinical specimens than the aforementioned influenza study, have helped identify where new SARS-CoV-2 variants have emerged and informed vaccine composition.

Not only do genomic epidemiology studies have the potential to help guide vaccine selection and development, but they also help to track what happens to pathogens circulating in the population in response to vaccination. By describing pathogen evolution under the selective pressure of a vaccinated population, such studies can play a key role in surveillance and identification of virulence determinants and perhaps may even help to predict the future evolution of escape from vaccine protection. The seven-valent pneumococcal conjugate vaccine (PCV-7) targeted the seven serotypes of *S. pneumoniae* responsible for the majority of invasive disease at the time of its introduction in 2000; since then, PCV-7 has dramatically reduced the incidence of pneumococcal disease and mortality. However, sequencing of >600 Massachusetts pneumococcal isolates from 2001 to 2007 has shown that, in the pneumococcal population, previously rare nonvaccine serotypes are replacing vaccine serotypes and that some vaccine strains have persisted despite vaccination by recombining the vaccine-targeted capsule locus with a cassette of capsule genes from non-vaccine-targeted serotypes. Studying the virulence of these persistently circulating strains can help to rationally update vaccine composition.

The large collections of pathogen genome sequences are driving development of tools to decipher the genetic basis for antibiotic resistance, virulence, and infection risk. Some pathogens have distinct types of clinical manifestations, the basis for which we are just beginning to unravel with the aid of genomics. For example, *Listeria* is a food-borne pathogen that can cause both central nervous system infections and maternal/neonatal infections. Although all *Listeria* isolates are treated the same from a public health perspective, variation in outcomes exists and appears to be linked to the strains' genomic background. Molecular analysis of a national reference laboratory's collections of well-characterized specimens, based on the fraction of immunocompetent people in which they caused disease, revealed that some clonal complexes of *Listeria* appear to be more virulent than others. Linking epidemiology and comparative genomics then enabled enumeration of putative virulence factors that contribute to the clinical phenotypes as well as identification and confirmation of a novel gene cluster that mediates central nervous system tropism. This approach illustrates progress toward a future in which we can link pathogen identification with risk, thereby informing resource use and allocation.

GLOBAL CONSIDERATIONS

While cutting-edge genomic technologies are largely implemented in the developed world, their application to infectious diseases perhaps offers the biggest potential impact in less developed regions where the burden of these infections is greatest. This globalization of genomic technology and its extensions has already begun in each of the areas of focus highlighted in this chapter; it has occurred both through the application of advanced technologies to samples collected in the developing world and through the adaptation and importation of technologies directly to the developing world for on-site implementation as they become more globally accessible.

Genomic characterization of the pathogens responsible for such important global illnesses, such as tuberculosis, malaria, trypanosomiasis, cholera, and COVID-19, has led to insights in diagnosis, treatment, and infection control. For instance, with the increasing burden of drug-resistant tuberculosis in the developing world, a molecular diagnostic test has been developed to detect rifampin-resistant tuberculosis. The genetic basis for rifampin resistance has been well defined by targeted sequencing: characteristic mutations in the molecular target of rifampin, RNA polymerase, account for the vast majority of instances of rifampin resistance. At least in areas that can afford to implement it, a rapid, automated PCR assay that can detect both *M. tuberculosis* and a rifampin-resistant allele of RNA polymerase directly in clinical samples has been implemented in parts of Africa and Asia, transforming the recognition and management of incident tuberculosis and multidrug resistance where they are most prevalent. Since rifampin resistance frequently accompanies resistance to other antibiotics, this test can suggest the presence of multidrug-resistant *M. tuberculosis* within hours instead of weeks, without the infrastructure required for culture.

High-resolution genomic tracking of the spread of epidemics—from cholera to Ebola to Zika to COVID-19—has yielded insights into which public health measures may prove most effective in controlling local epidemics. Many genomic tracking efforts have involved close collaborations with local scientists and public health officials, and considerable investment in sequencing infrastructure in sub-Saharan Africa has made on-location epidemic tracking in the event of another such outbreak feasible. Such investment can not only enable real-time outbreak recognition and tracking but also provide the infrastructure needed to capitalize on the many other benefits of high-throughput sequencing as they are developed. The early returns of such investments are exemplified by the rapid reporting of genome sequences for SARS-CoV-2, with substantial insights from sequencing efforts across the world; perhaps most notably, the Omicron variant was first identified in South Africa, enabling preparations for what became a global sweep in the subsequent weeks. Overall, sequencing efforts have become cheaper and have moved closer to point-of-care with each passing year. As these technologies synergize with efforts to globalize information-technology resources, global implementation of genomic methods promises to spread state-of-the-art methods for diagnosis, treatment, and epidemic tracking of infections to areas that need these capabilities the most.

GENOMICS AND THE COVID-19 PANDEMIC

The COVID-19 pandemic, which began in 2019 and spread worldwide in 2020, resulted in hundreds of millions of documented infections and millions of deaths and serves as a prime example of the pandemic potential of infectious pathogens. It also demonstrated the central role that genomic tools now play in response to infectious outbreaks, ranging from enabling diagnostics and vaccines to tracking evolution, virulence, and transmissibility of the pathogen. The rapid discovery of SARS-CoV-2 and sequencing of its genome was complete within weeks of the recognition of the clinical syndrome. The rapid public sharing of this genome sequence led directly to two key interventions: diagnostic assay development via RT-qPCR and vaccine design. Crucially, vaccine development was informed by homology of the SARS-CoV-2 sequence to SARS and MERS coronaviruses. The dominant antigen of those viruses, the surface protein Spike, was well characterized, enabling the design of the first SARS-CoV-2 vaccines to begin the day after the genome sequence was shared. The progress of the most rapidly developed and validated vaccine in human history was unquestionably accelerated by genomic technology. Whole genome sequencing has also played a large role in outbreak tracking and confirmation of case clusters in institutional settings such as hospitals or congregate living facilities, in helping to distinguish reinfections from recrudescence or prolonged viral shedding, in monitoring spread through societies, and in tracking pathogen evolution, including the emergence of new variants of concern with altered transmissibility, severity, and/or partial evasion of the immune response generated to prior versions of the virus, vaccines, or monoclonal antibody therapeutics. Finally, cutting edge genomic methods including single-cell transcriptional profiling and genome-wide association studies are contributing to our understanding of the wide variability in outcomes of SARS-CoV-2 infection, ranging from asymptomatic carriage to death. Overall, just as the global response to the COVID-19 pandemic underscores the indispensable role that genomics methods have come to play in the clinical and public health management of infectious diseases, the devastating impact of this pandemic reveals the urgent need for further development and implementation of tools for disease surveillance and response.

SUMMARY

By illuminating the genetic information that encodes the most fundamental processes of life, genomic technologies are transforming many aspects of medicine. In infectious diseases, methods such as next-generation sequencing and genome-scale expression analysis offer information of unprecedented depth about individual microbes as well as microbial communities. This information is expanding our understanding of the interactions of microorganisms with each other, their human hosts, and the environment. Despite technological and financial barriers that slow the widespread adoption of large-scale pathogen sequencing in clinical and public health settings, genomic methodologies have utterly transformed the research landscape in infectious disease and are beginning to make meaningful inroads into clinical settings. As even vaster amounts of data are generated, innovations in data storage, development of bioinformatics tools to manipulate the data, standardization of methods, and training of end-users in both the research and clinical realms will be required. The cost-effectiveness and applicability of whole genome sequencing, particularly in the clinic, remain to be studied, and studies of the impact of genome sequencing on patient outcomes will be needed to clarify the contexts in which these new methodologies can make the greatest contributions to patient well-being. The ongoing efforts to overcome limitations through collaboration, teaching, and reduction of financial obstacles should be applauded and expanded. With advances in genomic technologies and computational analysis, our ability to detect, characterize, treat, monitor, prevent, and control infections has advanced rapidly in recent years and will continue to do so, with the hope of heralding a new era where the clinician is better armed to combat infection and promote human health.

■ FURTHER READING

Bullman S et al: Emerging concepts and technologies for the discovery of microorganisms involved in human disease. Annu Rev Pathol 12:217, 2017.

Burnham CD et al: Diagnosing antimicrobial resistance. Nat Rev Microbiol 15:697, 2017.

Levy JI et al: Wastewater surveillance for public health. Science 379:26, 2023.

CRyPTIC Consortium et al: Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. N Engl J Med 379:1403, 2018.

Dudas G et al: Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature 544:309, 2017.

Gardy JL, Loman NJ: Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet 19:9, 2018.
Grubaugh ND et al: Tracking virus outbreaks in the twenty-first century. Nat Microbiol 4:10, 2019.
Loman NJ, Pallen MJ: Twenty years of bacterial genome sequencing. Nat Rev Microbiol 13:787, 2015.
Mutreja A et al: Evidence for several waves of global transmission in the seventh cholera pandemic. Nature 477:462, 2011.
Viana R et al: Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. Nature 603:679, 2022.

127

Approach to the Acutely Ill Infected Febrile Patient

Tamar F. Barlam

The physician treating the acutely ill febrile patient must be able to recognize infections that require emergent attention. If such infections are not adequately evaluated and treated at initial presentation, the opportunity to alter an adverse outcome may be lost. In this chapter, the clinical presentations of and approach to patients with infectious disease emergencies are discussed. These infectious processes and their treatments are discussed in detail in other chapters.

APPROACH TO THE PATIENT

Acute Febrile Illness

Before the history is elicited and a physical examination is performed, an immediate assessment of the patient's general appearance can yield valuable information. The perceptive physician's subjective sense that a patient is septic or toxic often proves accurate. Visible agitation or anxiety in a febrile patient can be a harbinger of critical illness.

HISTORY

Presenting symptoms are frequently nonspecific. Detailed questions should be asked about the onset and duration of symptoms and about changes in severity or rate of progression over time. Host factors, such as extremes of age and comorbid conditions, may increase the risk of infection with certain organisms or of a more fulminant course than is usually seen. Lack of splenic function, alcoholism with significant liver disease, IV drug use, HIV infection, diabetes, malignancy, morbid obesity, organ transplantation, and chemotherapy all predispose to specific infections and frequently to increased severity. The patient should be questioned about factors that might help identify a nidus for invasive infection, such as recent upper respiratory tract infections, influenza, or varicella; prior trauma; disruption of cutaneous barriers due to lacerations, burns, surgery, body piercing, or decubiti; and the presence of foreign bodies or prosthetic devices. Travel, presence during a natural disaster such as a hurricane or tsunami, contact with pets or other animals, or activities that might result in tick or mosquito exposure can lead to diagnoses that would not otherwise be considered. Recent dietary intake, medication use, social or occupational contact with ill individuals, vaccination history, recent sexual contacts, and menstrual history may be relevant. Pregnancy might increase the risk and severity of some illnesses, such as influenza or COVID-19, or increase the risk of significant morbidity for the fetus, as in *Listeria* or Zika virus infection. A detailed review of systems should include any neurologic signs or sensorium alterations, rashes or skin lesions, and focal pain or tenderness.

PHYSICAL EXAMINATION

A complete physical examination should be performed, with special attention to several areas that are sometimes given short shrift in routine examinations such as assessment of the patient's general appearance and a detailed skin, soft tissue, and neurologic evaluation.

The patient may appear either anxious and agitated or lethargic and apathetic. Fever is usually present, although elderly patients and compromised hosts (e.g., patients who are uremic or cirrhotic and those who are taking glucocorticoids or nonsteroidal anti-inflammatory drugs) may be afebrile despite serious underlying infection. Critically ill patients may be hypothermic, with a high risk of organ failure and mortality. Mortality at 30 days decreases with increasing temperature at presentation. Measurement of blood pressure, heart rate, and respiratory rate and oxygen saturation helps determine the degree of hemodynamic and metabolic compromise.

The etiologic diagnosis may become evident in the context of a thorough skin examination **(Chap. 21)**. Petechial rashes are typically seen with meningococcemia or Rocky Mountain spotted fever (RMSF; **see Fig. A1-16**); erythroderma is associated with toxic shock syndrome (TSS). On soft tissue and muscle examination, areas of erythema or duskiness, edema, and tenderness may indicate underlying necrotizing fasciitis, myositis, or myonecrosis. The neurologic examination must include a careful assessment of mental status for signs of early encephalopathy. Evidence of nuchal rigidity or focal neurologic findings should be sought.

DIAGNOSTIC WORKUP

After a quick clinical assessment, diagnostic material should be obtained rapidly and antibiotic and supportive treatment begun. Blood (for cultures; baseline complete blood count with differential; measurement of serum electrolytes, blood urea nitrogen, serum creatinine, and serum glucose; liver function tests and serum lactate; and D-dimer) can be obtained at the time an IV line is placed and before antibiotics are administered. Three sets of blood cultures should be performed for patients with possible acute endocarditis. Blood smears from patients at risk for severe parasitic disease, such as malaria or babesiosis **(Chaps. 231, 232, and A2)**, must be examined for the diagnosis and quantitation of parasitemia. Blood smears may also be diagnostic in ehrlichiosis and anaplasmosis. Testing of a nasopharyngeal sample for COVID-19 or influenza may be indicated.

Patients with possible meningitis should have cerebrospinal fluid (CSF) obtained before the initiation of antibiotic therapy. Focal findings, depressed mental status, or papilledema should be evaluated by brain imaging prior to lumbar puncture, which, in this setting, could initiate herniation. *Antibiotics should be administered before imaging but after blood for cultures has been drawn.* If CSF cultures are negative, blood cultures will provide the diagnosis in 50–70% of cases. Molecular diagnostic techniques (e.g., broad-range 16S rRNA gene polymerase chain reaction testing for bacterial meningitis pathogens) are of increasing importance in the rapid diagnosis of life-threatening infections.

CT or MRI may be necessary to evaluate focal processes that require urgent surgical intervention. Other diagnostic procedures, such as wound cultures, should not delay the initiation of treatment for more than minutes. Once emergent evaluation, diagnostic procedures, and (if appropriate) surgical consultation (see below) have been completed, other laboratory tests can be conducted. Appropriate radiography; CT and/or MRI imaging; urinalysis; measurement of the erythrocyte sedimentation rate, C-reactive protein, and/or procalcitonin; and transthoracic or transesophageal echocardiography all may prove important.

TREATMENT

The Acutely Ill Patient

In the acutely ill patient, empirical antibiotic therapy for presumed bacterial or fungal infection is critical and should be administered without undue delay in addition to fluid resuscitation and vasopressor support as needed. Increased prevalence of antibiotic resistance in community-acquired bacteria must be considered when antibiotics are selected. Table 127-1 lists first-line empirical regimens for infections considered in this chapter. In addition to the rapid initiation of antibiotic therapy, several of these infections require urgent surgical attention. Neurosurgical evaluation for subdural empyema, otolaryngologic surgery for possible mucormycosis, and cardiothoracic surgery for critically ill patients with acute endocarditis are as important as antibiotic therapy. For infections such as necrotizing fasciitis and clostridial myonecrosis, rapid surgical intervention supersedes other diagnostic or therapeutic maneuvers.

Adjunctive treatments may reduce morbidity and mortality rates and include dexamethasone for bacterial meningitis or IV immunoglobulin for TSS. Adjunctive therapies should usually be initiated within the first hours of treatment; however, dexamethasone for bacterial meningitis must be given before or at the time of the first dose of antibiotic. Glucocorticoids may also be harmful—e.g., when given in the setting of cerebral malaria or viral hepatitis.

SPECIFIC PRESENTATIONS

The infections considered below according to common clinical presentation can have rapidly catastrophic outcomes, and their immediate recognition and treatment can be life-saving.

■ SEPSIS WITHOUT AN OBVIOUS FOCUS OF PRIMARY INFECTION

Patients initially have a brief prodrome of nonspecific symptoms and signs that progresses quickly to hemodynamic instability with hypotension, tachycardia, tachypnea, respiratory distress, and altered mental status. Disseminated intravascular coagulation (DIC) with clinical evidence of a hemorrhagic diathesis is a poor prognostic sign.

Septic Shock (See also Chap. 315) Patients with bacteremia leading to septic shock may have a primary site of infection (e.g., pneumonia, pyelonephritis, or cholangitis) that is not evident initially. Elderly patients who may have atypical presentations and often have comorbid conditions, hosts compromised by malignancy and neutropenia, and patients who have recently undergone a surgical procedure or hospitalization are at increased risk for an adverse outcome. Gram-negative bacteremia with organisms such as *Pseudomonas aeruginosa* or *Escherichia coli* and gram-positive infection with organisms such as *Staphylococcus aureus* (including methicillin-resistant *S. aureus* [MRSA]) or group A streptococci can present as intractable hypotension and multiorgan failure. Treatment can usually be initiated empirically on the basis of the presentation, host factors (Chap. 315), and local patterns of bacterial resistance. Outcomes are worse when antimicrobial treatment is delayed or when the responsible pathogen ultimately proves not to be susceptible to the initial regimen. The increasing prevalence of multidrug-resistant organisms makes this especially relevant. Broad-spectrum antimicrobial agents are therefore recommended and should be instituted rapidly, preferably within the first hours after presentation with septic shock. Pharmacodynamics are altered in sepsis due to increased volume of distribution and renal clearance, so it is important to adequately dose antimicrobials. Risk factors for fungal infection should be assessed, as the incidence of fungal septic shock is increasing. Nonbacterial causes of shock, such as dengue virus infection, should be considered in endemic areas. Glucocorticoids are often considered for patients with severe sepsis who do not respond to fluid resuscitation and vasopressor therapy, but conclusive evidence for efficacy in this setting is lacking.

Overwhelming Infection in Asplenic Patients (See also Chap. 315) Patients without splenic function are at risk for overwhelming bacterial sepsis compared with the general population. The median interval between splenectomy and sepsis is 4–6 years, with a range of 1–19 years. Almost 50% or more of these infections occur within the first 1 or 2 years, but the increased risk persists throughout life. Encapsulated bacteria cause the majority of infections. *Streptococcus pneumoniae* is the most common isolate, causing 40–70% of cases. Children less than 5 years of age are at 15 times higher risk of invasive pneumococcal disease than adults, who are more likely to have antibody to these organisms. The risk of infection with *Haemophilus influenzae* or *Neisseria meningitidis* also is greater in patients without splenic function, but reported cases are declining. Severe clinical manifestations of infections due to other organisms, such as *E. coli*, *S. aureus*, *Bordetella holmesii*, *Capnocytophaga*, *Babesia*, and *Plasmodium* species, have been described.

Babesiosis (See also Chap. 232) A history of recent travel to endemic areas raises the possibility of infection with *Babesia*. Cases are increasing in the United States, particularly in the Northeast. Between 1 and 4 weeks after a tick bite, the patient experiences chills, fatigue, anorexia, myalgia, arthralgia, shortness of breath, nausea, and neurologic symptoms such as headache, confusion, delirium or impaired consciousness; ecchymosis and/or petechiae are occasionally seen. The tick that most commonly transmits *Babesia*, *Ixodes scapularis*, also transmits *Borrelia burgdorferi* (the agent of Lyme disease) and *Anaplasma*; co-infection can occur and may result in more severe disease. Infection with the European species *Babesia divergens* is more frequently fulminant than that due to the U.S. species *Babesia microti. B. divergens* causes a febrile syndrome with hemolysis, jaundice, hemoglobinemia, and renal failure and is associated with a mortality rate as high as 40%. Severe babesiosis is especially common in asplenic hosts but does occur in hosts with normal splenic function, particularly those >60 years of age and those with underlying immunosuppressive conditions such as HIV infection or malignancy. Complications include renal failure, acute respiratory failure, heart failure, DIC, and splenic rupture.

Other Sepsis Syndromes Tularemia (Chap. 175) has been reported in every U.S. state except Hawaii. This disease is associated with wild rabbit, tick, horse-fly, and tabanid fly contact. It can be transmitted by arthropod bite, handling of infected animal carcasses, consumption of contaminated food and water, or inhalation. The typhoidal form can be associated with gram-negative septic shock and a mortality rate of >30%, especially in patients with underlying comorbid or immunosuppressive conditions. Plague occurs infrequently in the United States (Chap. 176), primarily after contact with ground squirrels, prairie dogs, or chipmunks, but is endemic in other parts of the world; >90% of all cases occur in Africa with Madagascar especially affected. The septic form is particularly rare and is associated with shock, multiorgan failure, and a 30% mortality rate. Pneumonic plague is rapidly progressive and fatal without treatment. These infections should be considered in the appropriate epidemiologic setting. The Centers for Disease Control and Prevention (CDC) lists *Francisella tularensis* and *Yersinia pestis* (the agents of tularemia and plague, respectively) along with *Bacillus anthracis* (the agent of anthrax) as important organisms that might be used for bioterrorism (Chap. S4).

■ SEPSIS WITH SKIN MANIFESTATIONS (SEE ALSO CHAP. 21)

Sepsis can be associated with diverse skin findings. In one study, almost 18% of patients with severe sepsis had secondary skin findings, such as purpura, petechiae, or ecchymoses. Maculopapular rashes may reflect early meningococcal or rickettsial disease but are usually associated with nonemergent infections. Exanthems are usually viral. Primary HIV infection commonly presents with a rash that is typically maculopapular and involves the upper part of the body but can spread to the palms and soles. The patient is usually febrile and can have lymphadenopathy, severe headache, dysphagia, diarrhea, myalgias, and arthralgias. Recognition of this syndrome provides an opportunity to prevent transmission and to institute early treatment.

TABLE 127-1 Empirical Treatment for Common Infectious Disease Emergencies[a]

CLINICAL SYNDROME	POSSIBLE ETIOLOGIES	TREATMENT	COMMENTS	SEE CHAP(S)
Sepsis without a Clear Focus				
Septic shock	*Pseudomonas* spp., gram-negative enteric bacilli, *Staphylococcus* spp., *Streptococcus* spp.	Vancomycin (15 mg/kg q12h)[b] ***plus either*** Piperacillin/tazobactam (4.5 g) q8h via extended infusion (EI)[c] *or* cefepime (2 g) q8h via EI	Empirical therapy should be tailored to local resistance patterns. Carbapenem or aminoglycoside antibiotics should be considered for empirical therapy when rates of multidrug-resistant gram-negative organisms are high or for patients with risk factors for resistant organisms. Adjust treatment when culture data become available.	152, 153, 166, 170, 315
Overwhelming post-splenectomy sepsis	*Streptococcus pneumoniae, Haemophilus influenzae, Neisseria meningitidis*	Ceftriaxone (2 g q12h) *plus* vancomycin (15 mg/kg q12h)[b]	If a β-lactam–sensitive strain is identified, vancomycin can be discontinued and a narrower-spectrum agent, such as penicillin, considered based on susceptibility testing.	315
Babesiosis	*Babesia microti* (U.S.), *B. divergens* (Europe)	Atovaquone (750 mg q12h) *plus* azithromycin (500 mg q24h)	Clindamycin (600 mg q8h) *plus* quinine (650 mg q8h) can be used in severe disease not responding to atovaquone and azithromycin. Treatment with doxycycline (100 mg bid) for potential co-infection with *Borrelia burgdorferi* or *Anaplasma* spp. may be prudent.	229, 232
Sepsis with Skin Findings				
Meningococcemia	*N. meningitidis*	Ceftriaxone (2 g q12h) or penicillin (4 mU q4h)	Ceftriaxone eradicates nasopharyngeal carriage of the organism. Close contacts require chemoprophylaxis with rifampin (600 mg q12h for 2 days) or ciprofloxacin (a single dose, 500 mg).	160
Rocky Mountain spotted fever (RMSF)	*Rickettsia rickettsii*	Doxycycline (100 mg bid)	If both meningococcemia and RMSF are being considered, use ceftriaxone (2 g q12h) *plus* doxycycline (100 mg bid). If RMSF is diagnosed, ceftriaxone can be discontinued.	192
Purpura fulminans	*S. pneumoniae, H. influenzae, N. meningitidis*	Ceftriaxone (2 g q12h) *plus* vancomycin (15 mg/kg q12h)[b]	If a β-lactam-sensitive strain is identified, vancomycin can be discontinued.	151, 160, 162, 315
Erythroderma: toxic shock syndrome	Group A *Streptococcus, Staphylococcus aureus*	Vancomycin (15 mg/kg q12h)[b] *plus* clindamycin (600 mg q8h)	If a penicillin- or oxacillin-sensitive strain is isolated, these agents are superior to vancomycin (penicillin, 2 mU q4h; or oxacillin, 2 g IV q4h). The site of toxigenic bacteria should be debrided; IV immunoglobulin can be used in severe cases.[d]	152, 153
Sepsis with Soft Tissue Findings				
Necrotizing fasciitis	Group A *Streptococcus,* mixed aerobic/anaerobic flora	Vancomycin (15 mg/kg q12h)[b] *plus* piperacillin/tazobactam (4.5 q q8h via EI)[c] *plus* clindamycin (600 mg q8h)	Urgent surgical evaluation is critical. Empirical therapy should be tailored to local resistance patterns. For mixed aerobic/anaerobic infections, clindamycin can be discontinued. Adjust treatment when culture data become available.	134, 152, 153
Clostridial myonecrosis	*Clostridium perfringens*	Penicillin (2 mU q4h) *plus* clindamycin (600 mg q8h)	Urgent surgical evaluation is critical.	159
Neurologic Infections				
Bacterial meningitis	*S. pneumoniae, N. meningitidis*	Ceftriaxone (2 g q12h) *plus* vancomycin (15 mg/kg q12h)[b]	If a β-lactam–sensitive strain is identified, vancomycin can be discontinued. If the patient is >50 years old or has comorbid disease, add ampicillin (2 g q4h) for *Listeria* coverage. Dexamethasone (10 mg q6h for 4 days) started before, or at the time of, the first dose of antibiotic improves outcome in adults with meningitis (especially pneumococcal).	143
Brain abscess, suppurative intracranial infections	*Streptococcus* spp., *Staphylococcus* spp., anaerobes, gram-negative bacilli	Vancomycin (15 mg/kg q12h)[b] *plus* metronidazole (500 mg q8h) *plus* ceftriaxone (2 g q12h)	Urgent surgical evaluation is critical. If a penicillin- or oxacillin-sensitive strain is isolated, these agents are superior to vancomycin (penicillin, 4 mU q4h; *or* oxacillin, 2 g q4h).	143
Cerebral malaria	*Plasmodium falciparum*	Artesunate (2.4 mg/kg IV at 0, 12, and 24 h; then once daily)	Avoid glucocorticoids. Until IV artesunate is available, treatment can be initiated with oral artemether-lumefantrine. Atovaquone-proguanil, quinine, and mefloquine are other options.	229, 231
Spinal epidural abscess	*Staphylococcus* spp., gram-negative bacilli	Vancomycin (15 mg/kg q12h)[b] ***plus either*** Piperacillin/tazobactam (4.5 g q8h via EI) *or* cefepime (2 g q8h via EI)[c]	Surgical evaluation is essential. If a β-lactam–sensitive strain is isolated, oxacillin, 2 g q4h or cefazolin 2 gm IV q8h is superior to vancomycin.	453
Focal Infections				
Acute bacterial endocarditis	*S. aureus,* β-hemolytic streptococci, HACEK group,[e] *Neisseria* spp., *S. pneumoniae*	Ceftriaxone (2 g q12h) or cefepime (2 g q8h via EI)[c] *plus* vancomycin (15 mg/kg q12h)[b]	Adjust treatment when culture data become available. Evaluation by cardiology and cardiothoracic surgery is essential.	133

[a]These empirical regimens include coverage for gram-positive pathogens that are resistant to β-lactam antibiotics. Local resistance patterns should be considered and may alter the need for empirical vancomycin or for expanded coverage for antibiotic-resistant gram-negative pathogens. [b]A vancomycin loading dose of 20–25 mg/kg can be considered in critically ill patients. Dosing must be adjusted based on pharmacokinetic/pharmacodynamic monitoring. Daptomycin (10 mg/kg once daily) can be considered in place of vancomycin as alternate coverage for β-lactam resistant gram-positive organisms. Data for use in central nervous system infections are limited but emerging. [c]EI, extended infusion. β-Lactam antibiotics may exhibit unpredictable pharmacodynamics in sepsis. Prolonged or continuous infusions are often used. [d]The optimal dose of IV immunoglobulin has not been determined, but the median dose in observational studies is 2 g/kg (total dose administered for 1–5 days). [e]*Haemophilus* spp., *Aggregatibacter* spp., *Cardiobacterium hominis*, *Eikenella corrodens*, and *Kingella kingae*.

Petechial rashes caused by viruses are seldom associated with hypotension or a toxic appearance, although there can be exceptions (e.g., severe measles or arboviral infection). Petechial rashes limited to the distribution of the superior vena cava are rarely associated with severe disease. In other settings, petechial rashes require more urgent attention.

Meningococcemia (See also Chap. 160) Almost three-quarters of patients with *N. meningitidis* bacteremia have a rash. Meningococcemia most often affects young children (i.e., those 6 months to 5 years old). In sub-Saharan Africa, the high prevalence of serogroup A meningococcal disease has been a threat to public health for more than a century. Thousands of deaths occur annually in this area, which is known as the "meningitis belt," and large epidemic waves occur approximately every 8–12 years. In the setting of aggressive vaccination programs against serogroup A, disease due to other serogroups, such as serogroups C, W135, and X, are increasing. Outside Africa, outbreaks for the past 50 years reported in the United States and Europe are caused primarily by serogroup C (approximately 60%) followed by serogroup B (29%). In the United States, sporadic cases and outbreaks occur in day-care centers, schools (grade school through college, particularly among college freshmen living in residential halls), and army barracks. Outbreaks have also been described in people experiencing homelessness. People with underlying comorbidities, such as cancer, renal or liver disease, or solid organ transplant recipients, are at increased risk. Household contacts of index cases are at 400–800 times greater risk of disease than the general population. Patients may have fever, headache, nausea, vomiting, myalgias, changes in mental status, and meningismus. However, the rapidly progressive form of disease is not usually associated with meningitis. The rash is initially pink, blanching, and maculopapular, appearing on the trunk and extremities, but then becomes hemorrhagic, forming petechiae. Petechiae are first seen at the ankles, wrists, axillae, mucosal surfaces, and palpebral and bulbar conjunctiva, with subsequent spread on the lower extremities and to the trunk. A cluster of petechiae may be seen at pressure points—e.g., where a blood pressure cuff has been inflated. In rapidly progressive meningococcemia (10–20% of cases), the petechial rash quickly becomes purpuric (see Fig. A1-41), and patients develop DIC, multiorgan failure, and shock; 50–60% of these patients die, and survivors often require extensive debridement or amputation of gangrenous extremities. Hypotension with petechiae for <12 h is associated with significant mortality. Cyanosis, coma, oliguria, metabolic acidosis, and elevated partial thromboplastin time also are associated with a fatal outcome. Antibiotics given in the office by the primary care provider before hospital evaluation and admission may improve prognosis; this observation suggests that early initiation of treatment may be lifesaving. Members of the patient's household and other persons with close contact should receive antibiotic prophylaxis with ciprofloxacin, rifampin, or ceftriaxone. Meningococcal conjugate vaccines are protective against serogroups A, C, Y, and W135 and are recommended for children 11–12 years of age with a booster dose at 16 years of age, and for other high-risk patients. Vaccines active against serogroup B are recommended for high-risk individuals ≥10 years of age and may be appropriate for teens and young adults (16 through 23 years of age).

Rocky Mountain Spotted Fever and Other Rickettsial Diseases (See also Chap. 192) RMSF is a tickborne disease that occurs throughout North and South America. It is caused primarily by *Rickettsia rickettsii* but can be caused by other rickettsiae (e.g., *R. parkeri*, *R. akari*). Up to 40% of patients do not report a history of a tick bite, but a history of travel or outdoor activity (e.g., camping in tick-infested areas) can often be ascertained. For the first 3 days, headache, fever, malaise, myalgias, nausea, vomiting, and anorexia are documented. By day 3, half of patients have skin findings. Blanching macules develop initially on the wrists and ankles and then spread over the legs and trunk. The lesions become hemorrhagic and are frequently petechial. The rash spreads to palms and soles later in the course. The centripetal spread is a classic feature of RMSF but occurs in a minority of patients. Moreover, 10–15% of patients with RMSF never develop a rash. The patient can be hypotensive and develop noncardiogenic pulmonary edema, confusion, lethargy, and encephalitis progressing to coma. The CSF contains 10–100 cells/μL, usually with a predominance of mononuclear cells. The CSF glucose level is often normal; the protein concentration may be slightly elevated. Renal and hepatic injury as well as bleeding secondary to vascular damage are noted. Delayed recognition and treatment are associated with a greater risk of death; mortality rates are 20–35% if treatment is delayed or not prescribed compared with ~4% when treated with doxycycline within 5 days of onset. Native Americans, Alaskan natives, Pacific Islanders, children 5–9 years of age, adults >70 years old, and persons with underlying immunosuppression are at increased risk of death as well.

Other rickettsial diseases cause significant morbidity and mortality worldwide. *Mediterranean spotted fever* caused by *Rickettsia conorii* is found in Africa, southwestern and south-central Asia, and southern Europe. Patients have fever, flu-like symptoms, and an inoculation eschar at the site of the tick bite. A maculopapular rash develops within 1–7 days, involving the palms and soles but sparing the face. Elderly patients or those with diabetes, alcoholism, uremia, or congestive heart failure are at risk for severe disease characterized by neurologic involvement, respiratory distress, and gangrene of the digits or purpura fulminans. Mortality rates associated with this severe form of disease approach 50% without treatment, but a single day of doxycycline is associated with much improved outcomes. *Epidemic typhus*, caused by *Rickettsia prowazekii*, is transmitted in louse-infested environments and emerges in conditions of extreme poverty, war, refugee camps, and natural disaster. Patients experience a sudden onset of high fevers, severe headache, cough, myalgias, and abdominal pain. A maculopapular rash develops (primarily on the trunk) in more than half of patients and can progress to petechiae and purpura. Serious signs include delirium, coma, seizures, noncardiogenic pulmonary edema, skin necrosis, and peripheral gangrene. Mortality rates approached 60% in the preantibiotic era and continue to exceed 10–15% in contemporary outbreaks. *Scrub typhus*, caused by *Orientia tsutsugamushi* (a separate genus in the family Rickettsiaceae), is transmitted by larval mites or chiggers and is one of the most common infections in southeastern Asia and the western Pacific. The organism is found in areas of heavy scrub vegetation (e.g., along riverbanks) and most common in rural areas among farming communities. Patients may have an inoculation eschar and may develop a maculopapular rash, lymphadenopathy, and dyspnea. Severe cases progress to pneumonia, meningoencephalitis, myocarditis, DIC, and renal failure. Mortality rates range from 1% to 70% and vary by location, increasing age, myocarditis, delirium, pneumonitis, or signs of hemorrhage.

If recognized in a timely fashion, rickettsial disease is very responsive to tetracycline-based treatment. Doxycycline (100 mg twice daily for 1–14 days) is the treatment of choice for both adults and children. Combination therapy with doxycycline and azithromycin reduces complications and death in patients with severe scrub typhus.

Purpura Fulminans (See also Chaps. 160 and 315) Purpura fulminans is the cutaneous manifestation of DIC and presents as large ecchymotic areas and hemorrhagic bullae. Progression of petechiae to purpura, ecchymoses, and gangrene is associated with congestive heart failure, septic shock, acute renal failure, acidosis, hypoxia, hypotension, and death. Purpura fulminans has been associated primarily with *N. meningitidis* but, in splenectomized patients, may be associated with *S. pneumoniae*, *H. influenzae*, and *S. aureus*.

Ecthyma Gangrenosum Septic shock caused by *P. aeruginosa* or, less often, *Aeromonas hydrophila* or other gram-negative organisms, can be associated with ecthyma gangrenosum (see Figs. 170-1 and A1-34): hemorrhagic vesicles surrounded by a rim of erythema with central necrosis and ulceration, most frequently located on the legs and trunk. Echthyma gangrenosum is most common among patients with neutropenia, extensive burns, and hypogammaglobulinemia.

Other Infections Associated with Rash *Vibrio vulnificus* and other noncholera *Vibrio* bacteremic infections (Chap. 173) can cause focal skin lesions and overwhelming sepsis in hosts with chronic liver disease, heavy alcohol consumption, iron storage disorders, diabetes, renal insufficiency, hematologic disease, or malignancy or other immunocompromising conditions. More than 95% of the cases are in the subtropical Pacific and Atlantic Oceans coastal regions in the Northern Hemisphere. After ingestion of contaminated raw shellfish (typically oysters from the Gulf Coast in U.S. cases), there is a sudden onset of malaise, chills, fever, and hypotension. The patient develops bullous or hemorrhagic skin lesions, usually on the lower extremities, and 75% of patients have leg pain. The mortality rate can be as high as 35–60%, particularly when the patient presents with hypotension and septicemia. Outcomes are improved when patients are treated with fluoroquinolones with or without cephalosporins or with tetracycline-containing regimens. Other infections, caused by agents such as *Aeromonas*, *Klebsiella*, and *E. coli*, can cause hemorrhagic bullae and death due to overwhelming sepsis in cirrhotic patients. *Capnocytophaga canimorsus* can cause septic shock in asplenic or cirrhotic patients. Infection typically follows a dog bite. Serovars A–C appear more virulent, constituting 92% of human infections but only 7.6% of canine isolates. Patients present with fever, chills, myalgia, vomiting, diarrhea, dyspnea, confusion, and headache. Findings can include an exanthem or erythema multiforme **(see Figs. 59-9 and A1-24A)**, cyanotic mottling or peripheral cyanosis, petechiae, and ecchymosis. About one-third of patients with sepsis develop septic shock, and 11–30% of patients with this fulminant form die of overwhelming sepsis and DIC. Survivors may require amputation because of gangrene.

Erythroderma TSS **(Chaps. 152 and 153)** is usually associated with erythroderma. The patient presents with fever, malaise, myalgias, nausea, vomiting, diarrhea, and confusion. There is a sunburn-type rash that may be subtle and patchy but is usually diffuse and is found on the face, trunk, and extremities. Erythroderma, which desquamates after 1–2 weeks, is more common in *Staphylococcus*-associated than in *Streptococcus*-associated TSS. Hypotension develops rapidly—often within hours—after the onset of symptoms. Early renal failure may precede hypotension and distinguishes this syndrome from other septic shock syndromes. There may be no indication of a primary focal infection, although possible cutaneous or mucosal portals of entry for the organism can be ascertained through a careful history. Colonization rather than overt infection of the vagina or a postoperative wound, for example, is typical with staphylococcal TSS, and the mucosal areas appear hyperemic but not infected. Streptococcal TSS is more often associated with skin or soft tissue infection (including necrotizing fasciitis), and patients are more likely to be bacteremic. TSS caused by *Clostridium sordellii* is associated with childbirth or with skin injection of black-tar heroin. The diagnosis of TSS is defined by the clinical criteria of fever, rash, hypotension, and multiorgan involvement, although fever is typically absent when TSS is caused by *C. sordellii*. The mortality rate is 5% for menstruation-associated TSS, 10–15% for nonmenstrual TSS, 30–70% for streptococcal TSS, and up to 90% for obstetric *C. sordellii* TSS. Clindamycin improves outcomes when included in the treatment regimen. The use of IV immunoglobulin is associated with improved survival in some studies, in conjunction with clindamycin therapy.

Viral Hemorrhagic Fevers Viral hemorrhagic fevers **(Chaps. 215 and 216)** are zoonotic illnesses caused by viruses that reside in either animal reservoirs or arthropod vectors. These diseases occur worldwide and are restricted to areas where the host species live. They are caused by four major groups of viruses: Arenaviridae (e.g., Lassa fever in Africa), Bunyaviridae (e.g., Rift Valley fever in Africa; hantavirus hemorrhagic fever with renal syndrome in Asia; and Crimean-Congo hemorrhagic fever, which has an extensive geographic distribution), Filoviridae (e.g., Ebola and Marburg virus infections in Africa), and Flaviviridae (e.g., yellow fever in Africa and South America and dengue in Asia, Africa, and the Americas). Lassa fever and Ebola and Marburg virus infections are also transmitted from person to person. The vectors for most viral fevers are found in rural areas; dengue and yellow fever are important exceptions. After a prodrome of fever, myalgias, and malaise, patients develop evidence of vascular damage, petechiae, and local hemorrhage. Shock, multifocal hemorrhaging, and neurologic signs (e.g., seizures or coma) predict a poor prognosis. Dengue **(Chap. 215)** is the most common arboviral disease worldwide. More than 100–400 million cases occur each year, ~25% of which are symptomatic with at least 12,000–20,000 deaths. Patients have a triad of symptoms: hemorrhagic manifestations, evidence of plasma leakage, and platelet counts of <100,000/μL. Mortality rates are 10–20%. If dengue shock syndrome develops, which is associated with severe hepatitis and neurologic involvement, mortality rates can reach 40%. Ebola infection has been associated with outbreaks with high mortality rates. From 1976 to 2022, 35 outbreaks were reported with an average case fatality rate of approximately 50%. Symptoms can appear 2–21 days after exposure, but most patients become ill within 9 days. The patient first presents with fatigue, fever, headache, and muscle pains, and the illness can progress to multiorgan failure and hemorrhaging. Careful volume-replacement therapy to maintain blood pressure and intravascular volume is key to survival in these infections.

Other viral illnesses with rash, such as measles, can be associated with significant mortality rates, especially in patients who develop organ complications such as pneumonia, pancreatitis or encephalitis. Measles continues to be responsible for more than 100,000 deaths per year worldwide, and to cause outbreaks in populations with low vaccination rates.

■ SEPSIS WITH A SOFT TISSUE/MUSCLE PRIMARY FOCUS

See also Chap. 134.

Necrotizing Fasciitis This infection is characterized by extensive necrosis of the subcutaneous tissue and fascia. It may arise at a site of minimal trauma or surgical incision and may also be associated with recent varicella, childbirth, or muscle strain. Diabetes mellitus, IV drug use, chronic liver or renal disease, and malignancy are associated risk factors. The most common causes of necrotizing fasciitis are group A streptococci alone **(Chap. 153)** and a mixed facultative and anaerobic flora **(Chap. 134)**. The incidence of group A streptococcal necrotizing fasciitis has been increasing for the past quarter-century; surveillance in the United States links 13% of the cases to active IV drug use or homelessness. Physical findings are initially minimal except for soft tissue edema and mild erythema compared with the severity of pain and the degree of fever. The infected area is red, hot, shiny, swollen, and exquisitely tender. In untreated infection, the overlying skin develops blue-gray patches after 36 h, and cutaneous bullae and necrosis develop after 3–5 days. Necrotizing fasciitis due to a mixed flora, but not that due to group A streptococci, can be associated with gas production. Without treatment, pain decreases because of thrombosis of the small blood vessels and destruction of the peripheral nerves—an ominous sign. The mortality rate is 11–34% overall and increases when in association with TSS. With surgery, outcomes are significantly better. Necrotizing fasciitis may also be due to *Clostridium perfringens* **(Chap. 159)**; in this condition, the patient is extremely toxic and the mortality rate is high. Within 48 h, rapid tissue invasion and systemic toxicity associated with hemolysis and death ensue. The distinction between this entity and clostridial myonecrosis is made by muscle biopsy.

Clostridial Myonecrosis (See also Chap. 159) Myonecrosis is often associated with trauma or surgery but can develop spontaneously. The incubation period is usually 12–24 h long, and massive necrotizing gangrene develops within hours of onset. Systemic toxicity, shock, and death can occur within 12 h. The patient's pain and toxic appearance are out of proportion to physical findings. On examination, the patient is febrile, apathetic, tachycardic, and tachypneic and may express a feeling of impending doom. Hypotension and renal failure develop later, and hyperalertness is evident preterminally. The skin over the affected area is bronze-brown, mottled, and edematous. Bullous lesions with serosanguineous drainage and a mousy or sweet odor can develop. Crepitus can occur secondary to gas production in muscle tissue. The mortality rate is >65% for spontaneous myonecrosis, which is often associated with *Clostridium septicum* or *C. tertium* and underlying malignancy. The mortality rates associated with trunk and limb infection are 63% and 12%, respectively, and any delay in surgical treatment increases the risk of death.

NEUROLOGIC INFECTIONS WITH OR WITHOUT SEPTIC SHOCK

Bacterial Meningitis (See also Chap. 143) Bacterial meningitis is one of the most common infectious disease emergencies involving the central nervous system. Although hosts with cell-mediated immune deficiency (including transplant recipients, diabetic patients, elderly patients, and cancer patients receiving certain chemotherapeutic agents) are at particular risk for *Listeria monocytogenes* meningitis, most cases in adults are due to *S. pneumoniae* (30–60%) and *N. meningitidis* (10–35%). The classic presentation of fever, meningismus, and altered mental status is seen in only one-third to one-half of patients and is more common in adults over 60 years of age, although the elderly can also present without fever or meningeal signs. Cerebral dysfunction is evidenced by confusion, delirium, and lethargy that can progress to coma. In some cases, the presentation is fulminant, with sepsis and brain edema; papilledema at presentation is unusual and suggests another diagnosis (e.g., an intracranial lesion). Focal signs, including cranial nerve palsies (IV, VI, VII), can be seen in 10–20% of cases; 50–70% of patients have bacteremia. A poor outcome is associated with coma, seizures, hypotension, a purpuric rash, a pneumococcal etiology, respiratory distress, a CSF glucose level of <0.6 mmol/L (<10 mg/dL), a CSF protein level of >2.5 g/L, and peripheral leukopenia or thrombocytopenia. Rapid initiation of treatment is essential; the odds of an unfavorable outcome may increase by 30% for each hour that treatment is delayed. Dexamethasone is an adjunctive treatment for meningitis in adults, especially for infections caused by *S. pneumoniae,* but is not recommended for *Listeria* meningitis. It must be given before or with the first dose of antibiotics; otherwise, it is unlikely to improve outcomes.

Suppurative Intracranial Infections (See also Chap. 145) Suppurative intracranial infections present along with sepsis and hemodynamic instability. Rapid recognition of the toxic patient with central neurologic signs is crucial to improvement of the prognosis of these entities. Patients with diabetes or hematologic disease may be at increased risk for these infections. *Subdural empyema* arises from the paranasal sinus in 60–70% of cases. Microaerophilic streptococci and staphylococci are the predominant etiologic organisms. The patient is toxic, with fever, headache, and nuchal rigidity. Of all patients, 75% have focal signs and 6–20% die. Despite improved survival rates, 15–44% of patients are left with permanent neurologic deficits. *Septic cavernous sinus thrombosis* follows a facial or sphenoid sinus infection; 70% of cases are due to staphylococci (including MRSA), and the remainder are due primarily to aerobic or anaerobic streptococci. Fungi have been common in some series. A unilateral or retro-orbital headache progresses to a toxic appearance and fever within days. Three-quarters of patients have unilateral periorbital edema that becomes bilateral and then progresses to ptosis, proptosis, ophthalmoplegia, and papilledema. The mortality rate is as high as 30% in older studies with frequent neurologic sequelae but recent reports indicate improved survival as high as 90%. *Septic thrombosis of the superior sagittal sinus* spreads from the ethmoid or maxillary sinuses and is caused by *S. pneumoniae*, other streptococci, and staphylococci. The fulminant course is characterized by headache, nausea, vomiting, rapid progression to confusion and coma, nuchal rigidity, and brainstem signs. Broad-spectrum antibiotics and early surgical intervention at the primary site of infection may improve outcomes. Anticoagulation or steroids are of uncertain benefit.

Brain Abscess (See also Chap. 145) Brain abscess often occurs without systemic signs. Almost half of patients are afebrile, and presentations are more consistent with a space-occupying lesion in the brain; 70% of patients have headache and/or altered mental status, 50% have focal neurologic signs, and 25% have papilledema. Abscesses can present as single or multiple lesions resulting from contiguous foci or hematogenous infection, such as endocarditis, or after surgery or trauma. The infection progresses over several days from cerebritis to an abscess with a mature capsule. More than one-fourth to one-half of infections are polymicrobial, with an etiology consisting of aerobic bacteria (primarily streptococcal species) and anaerobes. Abscesses arising hematogenously are especially apt to rupture into the ventricular space, causing a sudden and severe deterioration in clinical status and a high mortality rate. Otherwise, mortality is low (<20%) but morbidity is high (30–55%). Patients presenting with stroke and a parameningeal infectious focus, such as sinusitis or otitis, may have a brain abscess; physicians must maintain a high level of suspicion. Prognosis worsens in patients with a fulminant course, delayed diagnosis, abscess rupture into the ventricles, multiple abscesses, or abnormal neurologic status at presentation. Patients who receive medical therapy alone, compared with those receiving combined medical and surgical interventions, have higher mortality and a greater chance of neurologic sequelae.

Cerebral Malaria (See also Chap. 231) This entity should be urgently considered for inhabitants of, or recent travelers to, areas endemic for malaria. Any patient with a change in mental status or repeated seizure in the setting of fulminant malaria has cerebral malaria. Fulminant malaria is caused by *Plasmodium falciparum* and is associated with temperatures of >40°C (>104°F), hypotension, jaundice, acute respiratory distress syndrome, and bleeding. Patients present with a febrile illness and lethargy or other neurologic signs, In adults, this nonspecific febrile illness progresses to coma over several days; occasionally, coma occurs within hours and death within 24 h. Nuchal rigidity and photophobia are rare. On physical examination, symmetric encephalopathy is typical, and upper motor neuron dysfunction with decorticate and decerebrate posturing can be seen in advanced disease. Unrecognized infection results in a 20–30% mortality rate. Intravenous artesunate is associated with better outcomes. Children with neurologic deficit at hospital discharge, seizure recurrence during treatment, and/or ischemic neural injury on MRI are at particular risk for neurologic and mental health sequelae.

Intracranial and Spinal Epidural Abscesses (See also Chap. 453) Spinal and intracranial epidural abscesses (SEAs and ICEAs) can result in permanent neurologic deficits, sepsis, and death. At-risk patients include those with diabetes mellitus; IV drug use; chronic alcohol abuse; recent spinal trauma, surgery, or epidural anesthesia; and other comorbid conditions, such as HIV infection. Fungal epidural abscess and meningitis have been linked to epidural or paraspinal glucocorticoid injections. In the United States and Canada, where early treatment of otitis and sinusitis is typical, ICEA is rare but the number of cases of SEA is on the rise. ICEAs typically present as fever, mental status changes, and neck pain, while SEAs often present as fever, localized spinal tenderness, and back pain. ICEAs are typically polymicrobial, whereas SEAs are most often due to hematogenous seeding, with staphylococci the most common etiologic agent. Early diagnosis and treatment, which may include surgical drainage, minimize rates of mortality and permanent neurologic sequelae. Outcomes are worse for SEA due to MRSA, for infection at a higher vertebral-body level, for impaired neurologic status on presentation, and for dorsal rather than ventral abscess location. Elderly patients and persons with renal failure, malignancy, and other comorbidities also have less favorable outcomes.

OTHER FOCAL SYNDROMES WITH A FULMINANT COURSE

Infection at virtually any primary focus (e.g., osteomyelitis, pneumonia, pyelonephritis, or cholangitis) can result in bacteremia and sepsis. Rapid clinical deterioration and death can be associated with destruction of the primary site of infection, as is seen in endocarditis and in infections of the oropharynx (e.g., Ludwig's angina or epiglottitis, in which edema suddenly compromises the airway). TSS has been associated with focal infections such as septic arthritis, peritonitis, sinusitis, and wound infection. Lemierre's syndrome is jugular septic thrombophlebitis primarily caused by *Fusobacterium necrophorum,* but it can be caused by other *Fusobacterium* spp., anaerobic streptococci and other bacteria. Lemierre's syndrome is associated with metastatic infectious emboli (primarily to the lung but sometimes to the liver or other organs) and sepsis, with mortality rates of >15%. *Fusobacterium* bloodstream infections have been associated with occult gastrointestinal or genitourinary malignancy.

Rhinocerebral Mucormycosis (See also Chap. 224) Individuals with diabetes or immunocompromising conditions such as solid organ transplants or hematologic malignancies are at risk for invasive rhinocerebral mucormycosis. Patients present with low-grade fever, dull sinus pain, diplopia, decreased mental status, decreased ocular motion, chemosis, proptosis, dusky or necrotic nasal turbinates, and necrotic hard-palate lesions that respect the midline. Without rapid recognition, surgical intervention, and antifungal therapy, the process continues on an inexorable invasive course, with mortality rates of 50–85% or greater. Uncontrolled diabetes and increasing age are negative prognostic factors.

Acute Bacterial Endocarditis (See also Chap. 133) This entity presents with a much more aggressive course than subacute endocarditis. Bacteria such as *S. aureus*, *S. pneumoniae*, *L. monocytogenes*, *Haemophilus* species, and streptococci of groups A, B, and G attack native valves. Native-valve endocarditis caused by *S. aureus* (including MRSA strains) is increasing. Mortality rates range from 10% to 40%. The host may have comorbid conditions such as underlying malignancy, diabetes mellitus, IV drug use, or alcoholism. Rheumatic valvular disease remains the most prevalent risk factor in Africa but is much less common in the United States. The patient presents with fever, fatigue, and malaise <2 weeks after onset of infection. On physical examination, a changing murmur and congestive heart failure may be noted. Hemorrhagic macules on palms or soles (*Janeway lesions*) sometimes develop. Petechiae, Roth's spots, splinter hemorrhages, and splenomegaly are unusual. Rapid valvular destruction, particularly of the aortic valve, results in pulmonary edema and hypotension. Myocardial abscesses can form, eroding through the septum or into the conduction system and causing life-threatening arrhythmias or high-degree conduction block. Large friable vegetations can result in major arterial emboli, metastatic infection, or tissue infarction including stroke. Older patients with *S. aureus* endocarditis are especially likely to present with nonspecific symptoms—a circumstance that delays diagnosis and worsens prognosis. Rapid intervention is crucial for a successful outcome.

Inhalational Anthrax (See also Chap. S4) Inhalational anthrax, the most severe form of disease caused by *B. anthracis*, had not been reported in the United States for more than 25 years until the use of this organism as an agent of bioterrorism in 2001. Patients presented with malaise, fever, cough, nausea, drenching sweats, shortness of breath, and headache. Rhinorrhea was unusual. All patients had abnormal chest roentgenograms at presentation. Pulmonary infiltrates, mediastinal widening, and pleural effusions were the most common findings. Hemorrhagic meningitis was documented in 38% of these patients. Survival was more likely when antibiotics were given during the prodromal period and when multidrug regimens with at least two bacteriocidal antibiotic agents were used. In the absence of urgent intervention with antimicrobial agents and supportive care, inhalational anthrax progresses rapidly to hypotension, cyanosis, and death.

Viral Respiratory Tract Illness Viral respiratory tract illnesses can cause severe disease; several new syndromes have been described in the past decade. For patients who present with a respiratory illness and a relevant exposure and travel history, these viral illnesses must be considered and appropriate infection control measures instituted in addition to supportive care.

Avian and Swine Influenza (See also Chap. 206) Human cases of avian influenza have occurred primarily in Southeast Asia, particularly Vietnam (H5N1) and China (H7N9). Avian influenza should be considered in patients with severe respiratory tract illness, particularly if they have been exposed to poultry. Patients present with high fever, an influenza-like illness, and lower respiratory tract symptoms; this illness can progress rapidly to bilateral pneumonia, acute respiratory distress syndrome, multiorgan failure, and death. Younger age appears to be associated with a lower risk of complications. Early antiviral treatment with neuraminidase inhibitors should be initiated along with aggressive supportive measures. Unlike avian influenza, whose human-to-human transmission has so far been rare and has not been sustained, influenza caused by a novel swine-associated A(H1N1) virus emerged in 2009 and caused more than 60 million cases and 12,000 deaths in 1 year. Patients most at risk of severe disease were children <5 years of age, elderly persons, patients with underlying chronic conditions, and pregnant women. Obesity also has been identified as a risk factor for severe illness. The virus continues to circulate as a seasonal influenza virus, along with A(H1N1) strains. Although influenza cases dropped during the peak of the COVID-19 pandemic, they are again increasing and cause 3–5 million cases of severe illness and 300,000–650,000 deaths each year.

SARS, COVID-19, and MERS (See Chaps. 204 and 205) Novel betacoronaviruses have now been identified as the cause of severe acute respiratory syndrome (SARS, caused by SARS-CoV-1), Middle East respiratory syndrome (MERS, caused by MERS-CoV), and COVID-19 (caused by SARS-CoV-2). Although the three viruses have many similarities, they differ in degree of contagion and case fatality rates. SARS and MERS "super-spreaders" are very symptomatic and typically critically ill, hospitalized patients. This is in contrast to COVID-19 where patients with highly transmissible virus are often asymptomatic or have mild disease. SARS emerged in 2002 in China and was diagnosed in several countries, primarily in Asia. Possible animal reservoirs include bats and civets. SARS-CoV-1 is characterized by efficient human transmission but relatively low mortality. The potential SARS-CoV-1 pandemic was controlled through identification and isolation of infected patients. There have been no reported cases since 2004.

MERS likely has a bat reservoir and was first recognized in 2012 in Saudi Arabia. Human cases have been associated with direct and indirect contact with dromedary camels. Unlike SARS and COVID-19, MERS exhibits inefficient human transmission but carries a high mortality rate. As of 2023, >2600 cases had been confirmed, most in the Arabian Peninsula, with 35% mortality. MERS ranges from asymptomatic infection to acute respiratory distress syndrome, multiorgan failure, and death. Elderly men with comorbidities appear to be at highest risk for poor outcomes. Despite little documented human-to-human transmission in the community, nosocomial infection must be prevented by adherence to strict infection control practices. MERS is currently a low-level public health threat and is likely to remain so unless the virus mutates and its transmissibility increases.

COVID-19 caused by SARS-CoV-2 resulted in a pandemic of historic proportion. Originally linked to cases in China, SARS-CoV-2 spread internationally at a rapid pace. As of December 2023, cases approached 773 million worldwide with almost 7 million confirmed deaths; the United States has reported more than 103 million cases and over 1.1 million deaths. Actual infection may exceed reported cases by 8–10 fold. The primary mode of transmission of COVID-19 is through direct person-to-person contact via respiratory droplets; transmission by the airborne route or by contact with contaminated surfaces is much less common. Greatest risk of infection is associated with close and prolonged contact such as in household or congregate settings. Rates of secondary infection among household contacts are high: >50% in studies early in the pandemic.

Although ~80% of patients with symptomatic COVID-19 infection have mild disease such as cough, sore throat, fever, gastrointestinal symptoms, and taste and smell alterations, ~5% of patients develop respiratory failure, shock, and multiorgan system failure. Patients typically have been symptomatic for 5–7 days before progression to severe pneumonia and hypoxemia. Shock, often in the setting of cardiac injury/myocarditis, arrhythmias or cardiomyopathy, and thromboembolic complications such as pulmonary emboli or stroke, can occur. Other manifestations include acute kidney or liver injury. Laboratory findings include lymphopenia, elevated lactate dehydrogenase, and often evidence of a cytokine-release type syndrome with elevated C-reactive protein, D-dimer, ferritin, and proinflammatory cytokines such as interleukin 6 (IL-6). Chest radiographs and CT imaging reveal diffuse ground-glass opacities. Consolidation, predominantly in the lower lobes and peripherally, is a frequent finding. Mortality rates for patients requiring intensive care and mechanical ventilation have improved significantly over the course of the pandemic but remain ≥20% for patients who progress to that disease severity. Risk factors for severe disease and poor outcomes include age ≥65 years, morbid obesity, diabetes, cardiovascular or cerebrovascular disease, hypertension, chronic obstructive lung disease, and chronic kidney disease. Severe disease and mortality is also increased in males and pregnant women.

Treatment currently includes supportive care, antiviral therapy, anti-inflammatory therapy, and anticoagulation prophylaxis or treatment. Patients with respiratory distress are hypoxemic but maintain pulmonary mechanical functioning; thus, invasive mechanical ventilation should be delayed until other interventions have been exhausted. In a meta-analysis, dexamethasone or other glucocorticoids reduced 28-day mortality in patients with severe COVID-19 disease compared with standard care: 32% and 40%, respectively (odds ratio, 0.66; 95% confidence interval, 0.53–0.82). Thus, dexamethasone (6 mg IV/PO daily) is indicated for patients with hypoxemia. IL-6 inhibitors such as tocilizumab may be indicated for patients with rapidly escalating oxygen requirements. Secondary bacterial pneumonia in COVID-19 is not common, and antibacterial agents should not be routinely prescribed. With the availability of highly effective vaccines, prevention is the most important strategy to control COVID-19. Vaccination has reduced progression to severe disease and hospitalization. In addition, treatment with antiviral agents such as nirmatrelvir/ritonavir or remdesivir can reduce risk of progression to severe disease and hospitalization for patients with mild symptoms.

Hantavirus Pulmonary Syndrome (See also Chap. 215) Hantavirus pulmonary syndrome has been documented in the United States since 1993 (primarily the southwestern states, west of the Mississippi River), Canada, and South America. Most cases occur in rural areas and are associated with exposure to rodents. Patients present with a nonspecific viral prodrome of fever, malaise, myalgias, nausea, vomiting, and dizziness that may progress to pulmonary edema, respiratory failure, and death. Hantavirus pulmonary syndrome causes myocardial depression and increased pulmonary vascular permeability; therefore, careful fluid resuscitation and use of pressor agents are crucial. Use of venoarterial extracorporeal membrane oxygenation (ECMO) for severe Hantavirus pulmonary syndrome has been successful and should be considered. Aggressive cardiopulmonary support during the first few hours of illness can be life-saving in this high-mortality syndrome. The early onset of thrombocytopenia may help distinguish this syndrome from other febrile illnesses in an appropriate epidemiologic setting.

SUMMARY

Acutely ill febrile patients with the syndromes discussed in this chapter require close observation, aggressive supportive measures, and—in most cases—admission to intensive care units. The most important task of the physician is to distinguish these patients from other infected febrile patients whose illness will not progress to fulminant disease. The alert physician must recognize the acute infectious disease emergency and proceed with appropriate urgency.

FURTHER READING

Hasburn R: Progress and challenges in bacterial meningitis: A review. JAMA 328:2147, 2022.

Hua C et al: Necrotising soft-tissue infections. Lancet Infect Dis 23:e81, 2023.

Kollef MH et al: Timing of antibiotic therapy in the ICU. Crit Care 25:360, 2021.

Theilacker C et al: Overwhelming postsplenectomy infection: A prospective multicenter cohort study. Clin Infect Dis 62:871, 2016.

Wilder-Smith A et al: Dengue. Lancet 393:350, 2019.

128 Principles of Immunization

Kathleen M. Neuzil, Kathryn M. Edwards

The birth of vaccinology is often ascribed to Edward Jenner's observation in the 18th century that infection of milkmaids with cowpox, termed *vaccinia* and hence the term *vaccination*, conferred protection against smallpox. This discovery catalyzed a sequence of events that ultimately led to the eradication of smallpox, one of the most disfiguring and lethal infections of humans. Vaccination programs are among the most impactful and cost-effective health interventions. Considered to be one of the greatest achievements in public health, vaccines are estimated to have averted more than 37 million deaths globally between 2000 and 2019. This represents a 45% reduction in deaths compared with a scenario of no vaccination.

Vaccines are defined as inactivated or attenuated pathogens or components of a pathogen (nucleic acid, protein) that, when administered to the host, stimulate a protective response by the cells of the immune system **(Table 128-1)**. Adjuvants may be added to vaccines to nonspecifically boost the immune response. Vaccination is the act of introducing a vaccine into the body to induce protection from a specific disease. Immunization is a process by which a person becomes resistant to a particular infectious disease or pathogen, usually by vaccination. However, immunity may also be conferred passively.

IMMUNE RESPONSES TO VACCINES

While **Chapter 360** provides an overview of immune responses, a few additional concepts are relevant to vaccines. Vaccines targeted for new infectious pathogens, such as SARS-CoV-2 at the onset of the pandemic, or vaccines given to infants and young children without prior exposure to a pathogen, will trigger a primary immune response with peak antibody titers achieved several weeks after administration. It may take several doses of a vaccine to induce a priming response, particularly if the vaccine is inactivated or a toxoid. Live, attenuated vaccines generate more robust responses and generally only require one or two doses. Subsequent booster doses of vaccines trigger a secondary response with rapid induction of antibody within several days to a week. Vaccines administered against common pathogens, such as influenza, will trigger secondary booster responses since all older children and adults would have encountered influenza virus and/or will have received past influenza vaccines. Immune responses to vaccines are generally measured in terms of serum antibody levels, but mucosal antibodies, cellular responses, and memory cells also are stimulated after vaccination and function to prevent disease and sustain immunity.

Passive Immunity Vaccines stimulate active immunity—inducing an immune response by the host to the vaccine antigen—and will be the focus of most of this section. Passive immunity also serves an important role in prophylaxis. Passive immunity results from the exogenous introduction of antibodies. Originally, such antibodies came from another person or animal and were relatively short-lived. Examples included hepatitis A, tetanus, and rabies immunoglobulins. The more recent capability to generate monoclonal antibodies to infectious pathogens, along with methods to extend the half-life of antibodies, has expanded their prophylactic use. Monoclonal antibodies targeted to SARS-CoV-2 strains were beneficial during the pandemic, particularly in populations who did not mount robust immune responses to vaccines. Currently, monoclonal antibodies are routinely used in infants and young children for the prevention of respiratory syncytial virus (RSV) infections, and their use in adults will likely continue or increase in the years ahead. These agents are evaluated for their safety and efficacy using similar methodology as in vaccine studies but will not be discussed extensively in this chapter.

Another type of passive antibody protection is maternal immunization, whereby vaccines are given to the pregnant person and the

TABLE 128-1 Categories and Characteristics of Approved Vaccines and Adjuvants

VACCINES

TECHNOLOGY	DESCRIPTION	EXAMPLES OF APPROVED VACCINES	ADVANTAGES	DISADVANTAGES
Live attenuated	Weakened or attenuated form of the pathogen that causes disease	Measles, mumps, rubella, varicella, oral poliomyelitis, nasal influenza vaccine, oral rotavirus	Mimics natural infection Effective priming with durable immunity Large-scale manufacturing capabilities Single or two doses often sufficient	Difficult to reach desired level of attenuation Safety concerns for certain populations (e.g., pregnant women, immunocompromised patients) Stability
Inactivated	Pathogens or toxins rendered nonreplicating through heat or chemical processes (may be entire pathogen or parts of it)	Inactivated poliomyelitis, hepatitis A, whole-cell pertussis, tetanus and diphtheria toxoids	Induces broad immune response to multiple antigens Large-scale manufacturing capabilities	Reactogenic Multiple doses needed May require adjuvant
Purified protein-based (split or subunit)	Specific parts of the pathogen are produced in culture	Influenza, acellular pertussis, recombinant shingles	Highly specific immune response Noninfectious Low reactogenicity Ease of production	Multiple doses may be needed May require adjuvant Limited cross-protective immunity
Polysaccharide–protein conjugates	Type of subunit vaccine that combines a polysaccharide antigen with a protein carrier to improve immune responses	Pneumococcal, meningococcal, typhoid, Hib	Conjugate protein may also provide immunity (e.g., tetanus) Strong immune responses in all ages, including infants	Technically difficult and costly to produce Individual conjugate required for each vaccine serotype
Virus-like particles	One or more proteins arranged to closely resemble viruses	Hepatitis B, HPV	Broad and robust immunity Noninfectious	Technically difficult to produce Lower stability
Replicating viral vector	Replicating, nonpathogenic viruses deliver nucleic acid to host for in vivo production of antigen	Vesicular stomatitis virus–based Ebola vaccine	Induces broad immune response High manufacturing scalability	Immunity to vector may dampen immune response Safety of vector—may not be suitable for certain immunocompromised populations
Nonreplicating viral vector	Replication-deficient nonpathogenic viruses deliver nucleic acid to host for in vivo production of antigen	Chimp adenovirus–based COVID-19 vaccine	Induces broad immune response High manufacturing scalability	Immunity to vector may dampen immune response
Nucleic acid (DNA, RNA)	Lipid nanoparticles deliver nucleic acid to host for in vivo production of antigen	COVID-19 mRNA vaccines	Rapid development and manufacturing timelines Effective and safe for majority of population	May require several doses/boosters to maintain immunity Reactogenic

ADJUVANTS

TECHNOLOGY	EXAMPLES	CURRENT USE	MECHANISM OF ACTION
Alum	Aluminum hydroxide or aluminum phosphate	Multiple vaccines—e.g., HPV, hepatitis A, DTaP	Possible depot effect; increases antibody production
Oil-in-water emulsion	AS03, MF59	Influenza	Increases recruitment of innate immune cells and enhanced antigen uptake
Toll-like receptor (TLR) agonists	Monophosphoryl lipid A (MPL) Cytosine-phosphate-guanine (CpG)	HPV (AS04—MPL and alum) Shingles (MPL and QS21) Hepatitis B (CPG)	Activates TLR to enhance antigen presentation and enhance adaptive immune responses
Saponins	QS-21, Matrix-M	Malaria vaccines (MPL, QS21, and Matrix-M) COVID-19 (Matrix M)	Prolongs antigen bioavailability and enhances antigen signals
Delivery platforms	Lipid nanoparticle (LNP)	COVID-19 mRNA vaccines	Improved lymphatic transport enhances antigen uptake and presentation

antibody generated by their immune system is transferred to the fetus through the placenta. Examples of vaccines given to the pregnant person, largely for the protection of the baby, include tetanus, acellular pertussis, and RSV vaccines. Other vaccines given to pregnant persons for protection of both the mother and the baby include influenza and COVID-19.

VACCINE FORMULATIONS

Vaccine formulations have evolved over time (Fig. 128-1). In the late 19th to mid-20th century, methods to inactivate whole bacteria and detoxify bacterial toxins led to early whole-cell typhoid and cholera vaccines, as well as diphtheria and tetanus toxoids. The discovery that repeated passage of organisms in unnatural hosts could select for avirulent strains provided the foundation for Bacille Calmette-Guérin (BCG) vaccine for tuberculosis and yellow fever vaccines. The ability to grow viruses in cell culture informed the development of live attenuated vaccines against poliomyelitis, measles, mumps, rubella, and varicella viruses between 1950 and 1980. During the next three decades, the fields of biotechnology and molecular genetics burgeoned. New vaccines that circumvented poor immune responses to bacterial polysaccharides by their conjugation to proteins included pneumococcal, *Haemophilus influenzae* type b, and meningococcal vaccines. The development of recombinant DNA technology allowed antigens to be expressed in cell lines, leading to hepatitis B and human papillomavirus (HPV) vaccines.

With the COVID-19 pandemic, a new generation of vaccines and platform technologies emerged with unprecedented speed. While many of these new technologies had been in development for years, the public health need and the influx of government resources propelled these technologies through the approval and deployment phases at an accelerated pace. The gene-based (e.g., mRNA) and vector-based technological advancements have been most visible. For years,

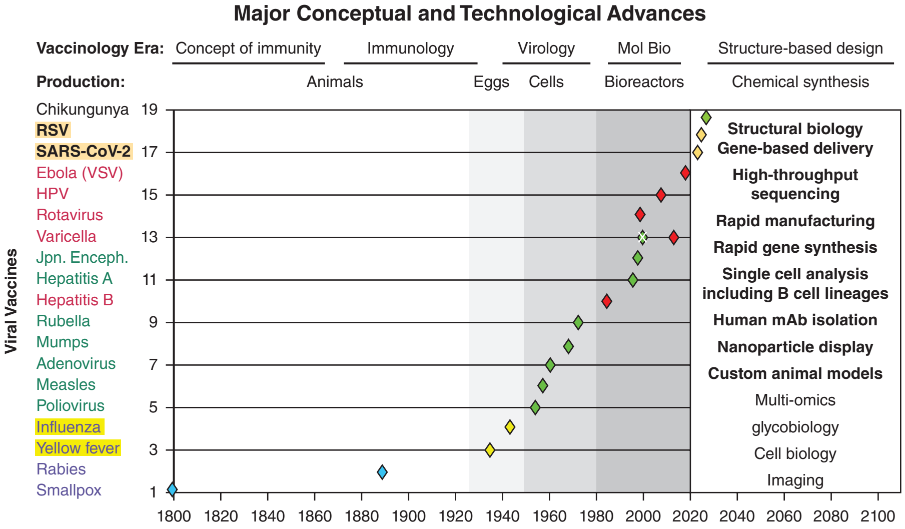

FIGURE 128-1 Major conceptual and technological advances for viral vaccines in the vaccinology era. *(Reproduced with permission from Barney Graham.)*

mRNA technology was limited by its instability and inefficiency of in vivo transfection, until the lipid nanoparticle (LNP) delivery system improved the RNA-based vaccine delivery. Further, new adjuvants and structure-guided antigen design of the SARS-CoV-2 spike protein transformed the field. These technological advances were accompanied by innovations in trial design, real-time safety surveillance and effectiveness evaluations at a massive scale, transmission modeling, and sophisticated analysis of immunologic correlates of protection.

Adjuvants Adjuvants are substances that are added to vaccines to enhance the immune responses to the antigen. Replicating viral vaccines generally do not need adjuvants, but they are often needed in inactivated or protein-based vaccines. Early emulsions, such as Freund's adjuvant, stimulated humoral and cellular responses but were very reactogenic. Alum, the oldest adjuvant formulated as aluminum hydroxide or phosphate, has been in use for more than 90 years. It promotes humoral immune responses, has an excellent safety record, and is included in many of the vaccines administered to infants and young children. As the understanding of molecular mechanisms involved in immune responses progressed, new adjuvants were developed. Oil-in-water emulsions, including MF59 and AS03, contain squalene as a vehicle and enhance innate and adaptive immune responses. They are approved in combination with certain influenza vaccines. As Toll-like receptors were discovered, adjuvants were developed to combine with the newer, more purified protein and subunit vaccines. The Toll-like receptor agonists TLR4 and TLR9 have been included in licensed vaccines. These include monophosphoryl lipid A (MPL) with alum (AS04) for HPV and cytosine-phosphate-guanine (CpG) for hepatitis B. More recently, MPL has been combined with a natural saponin component, QS-21, for use in the recombinant shingles vaccine and the RTS,S malaria vaccine. Likewise, a saponin-based adjuvant, Matrix-M, is approved for use in the recombinant COVID-19 vaccine and the R21 malaria vaccine. In addition to enhancing delivery of mRNA vaccines, LNPs induce inflammation and have an adjuvanting effect. Additional adjuvants are being studied in experimental vaccines, and others will likely be included in licensed products in the future.

VACCINE DEVELOPMENT

The pathway from vaccine discovery to population impact encompasses a broad set of components summarized in Fig. 128-2. These steps include understanding of the burden of the pathogen to be prevented and the immune factors involved in protection; antigen discovery; preclinical animal studies advancing to human trials assessing immune responses, safety, and efficacy; vaccine licensure and policymaking; implementation; and finally monitoring of vaccine uptake, impact, and safety after implementation. Well-established systems and processes, both pre- and postlicensure, ensure that vaccines are safe and effective.

Establishing the Need for an Immunization Generally, before extensive resources are expended to formulate and develop vaccines, it is important to establish the burden of disease, the population at risk for the disease, and the mechanisms involved in protection. Much of this can be done through existing surveillance systems, such as those at the Centers for Disease Control and Prevention (CDC), that measure the burden of disease and the population at risk. In certain situations, vaccines are developed before their potential benefit is fully known, based on the theoretical risk to the population. Examples of these vaccines are those directed against potentially pandemic influenza strains or other pandemic threats.

Phases of Vaccine Development Like the development of pharmaceutical agents, vaccine development progresses through preclinical and three distinct clinical stages: phases 1, 2, and 3. Under specific circumstances where phase 3 trials may be delayed or challenging, controlled human infection models (CHIM) may be part of the development pathway.

Initially, research focuses on the identification of an antigen or antigens with the potential to stimulate a protective immune response that will prevent disease upon encounter with the pathogen. For most vaccines the generation of an antibody directed to an important component of the pathogen is associated with protection. Examples are vaccines directed against the attachment proteins such as the spike in SARS-CoV-2 vaccines or against toxins such as the tetanus or diphtheria vaccines. These proposed vaccines are administered in *preclinical* studies to small animals, often mice; their immune responses are measured, and safety established. Generation of immunity against the pathogen must be demonstrated, or the vaccine will not undergo further testing. Toxicity studies are conducted in animals to detect safety signals, and some vaccinated animals are challenged with wild-type pathogens after vaccination to establish protection.

Phase 1, 2, and 3 Clinical Trials When preclinical studies in animals demonstrate that the vaccine stimulates an immune response and there are no toxicity concerns, vaccines then undergo *phase 1 clinical trials*. In the United States, prior to the onset of phase 1 studies, the vaccine manufacturer must submit an Investigational New Drug (IND) application to the U.S Food and Drug Administration (FDA) that outlines the vaccine approach prior to the start of the study. When the IND is approved, phase 1 trials can begin. These trials enroll limited numbers of healthy participants, usually fewer than 100 individuals and generally between the ages of 18 to approximately 55 years, primarily to test the safety of the new experimental vaccine, although immunogenicity also is measured. Subjects enrolled in phase 1 studies are well informed about the risks and the potential benefits of the vaccines and are screened for their ability to be monitored closely and to adhere to rigorous safety assessments. These assessments include daily

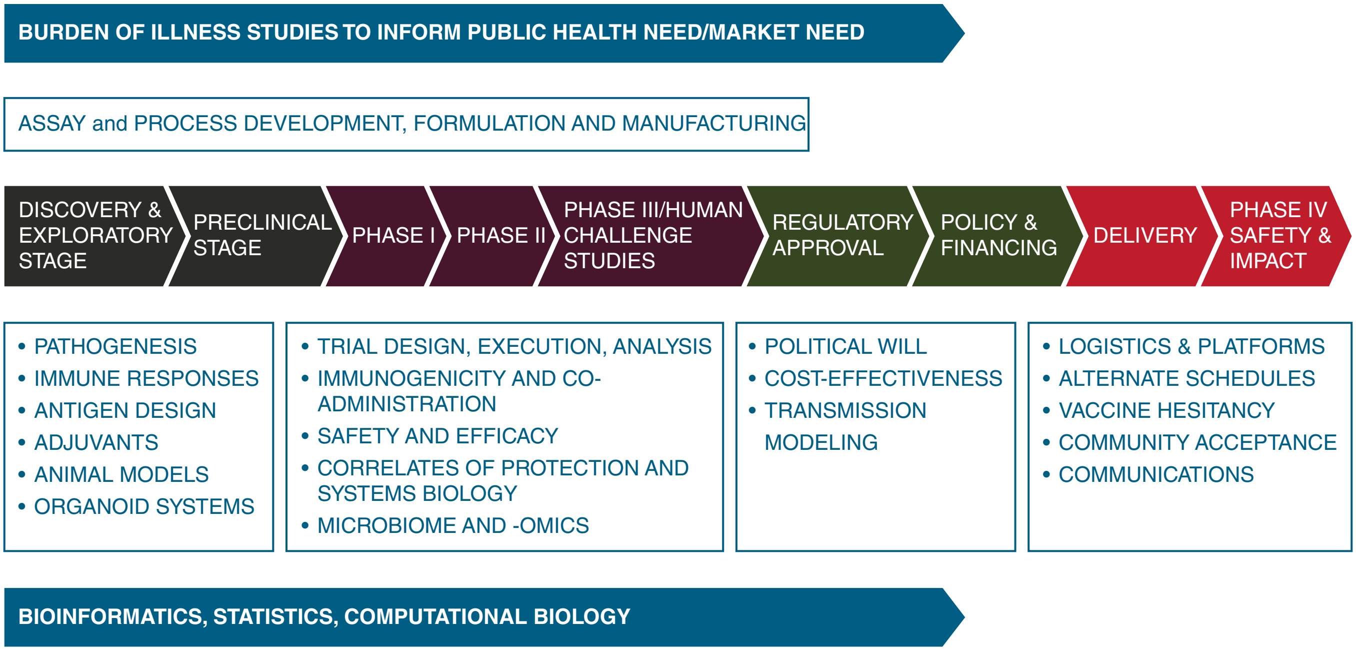
BURDEN OF ILLNESS STUDIES TO INFORM PUBLIC HEALTH NEED/MARKET NEED
ASSAY and PROCESS DEVELOPMENT, FORMULATION AND MANUFACTURING
DISCOVERY & EXPLORATORY STAGE
PRECLINICAL STAGE
PHASE I
PHASE II
PHASE III/HUMAN CHALLENGE STUDIES
REGULATORY APPROVAL
POLICY & FINANCING
DELIVERY
PHASE IV SAFETY & IMPACT
• PATHOGENESIS
• IMMUNE RESPONSES
• ANTIGEN DESIGN
• ADJUVANTS
• ANIMAL MODELS
• ORGANOID SYSTEMS
• TRIAL DESIGN, EXECUTION, ANALYSIS
• IMMUNOGENICITY AND CO-ADMINISTRATION
• SAFETY AND EFFICACY
• CORRELATES OF PROTECTION AND SYSTEMS BIOLOGY
• MICROBIOME AND -OMICS
• POLITICAL WILL
• COST-EFFECTIVENESS
• TRANSMISSION MODELING
• LOGISTICS & PLATFORMS
• ALTERNATE SCHEDULES
• VACCINE HESITANCY
• COMMUNITY ACCEPTANCE
• COMMUNICATIONS
BIOINFORMATICS, STATISTICS, COMPUTATIONAL BIOLOGY

FIGURE 128-2 Vaccine development: a continuum from antigen discovery to population impact. *(Figure created by Kathleen M. Neuzil, MD, MPH.)*

monitoring of local and systemic adverse events with measurement of temperature and size of redness and swelling at the injection site, and with detailed assessments of systemic reactions that may result in limitations of normal activities after vaccine administration. Phase 1 studies often involve a dose-ranging component such that the first enrolled subjects are administered the lowest doses of vaccine, and, if tolerated, the doses are increased. Many phase 1 studies have independent Data Safety and Monitoring Committees (DSMCs) composed of vaccine- and pathogen-specific experts, independent of the study site investigators and sponsors of the study, who assess the reactions and approve the advancing of the dose level based on the safety profile of the administered vaccines. Phase 1 studies have halting rules such that if severe reactions are seen, the study will be stopped. These halting rules might include hospitalization for a vaccine-related adverse event, or a severe reaction attributed to the vaccine in several participants. All participants in the phase 1 trials will have immunologic studies performed as well to determine the magnitude of the antibody responses and, in some situations, the function of the antibody generated and/or T lymphocyte responses to the vaccines. At the conclusion of the phase 1 trial, the safety and immunogenicity data are reviewed and presented to the FDA for approval to advance to phase 2 studies.

Phase 2 studies involve several hundred subjects and often have a larger age range than that evaluated in phase 1. They expand the safety profile and assess immune responses in larger numbers of subjects. Again, meticulous attention to safety assessment is included in these studies, often with independent DSMCs to assess the reaction profile. Finally, if phase 2 studies confirm safety and immunogenicity in the expanded populations and with approval of the FDA, phase 3 trials can begin.

Phase 3 trials are designed to determine whether the vaccines will prevent a predefined endpoint, generally the prevention of laboratory-confirmed disease with the vaccine-directed pathogen. Subjects enrolled in phase 3 studies are randomized and blind to the receipt of either another irrelevant vaccine or a control agent, usually consisting of a placebo. Subjects and investigators are blind to group assignment and remain so throughout the phase 3 trial. After administration of the vaccine, study participants are followed closely, and if they develop predefined symptoms of disease they are tested for the presence of the pathogen. The determination of the primary endpoint of the study and the level of vaccine efficacy is the study goal. When feasible, one of the additional outcomes of the phase 3 trial is the determination of a serologic correlate of protection. For example, a correlate could be an antibody level associated with protection from disease. This would facilitate evaluation of additional vaccines in population groups not involved in the original trial and would not require assessment of actual protection from disease. While this is a highly sought-after endpoint, it is often not possible to determine. Prior to submission of the vaccine for licensure there will need to be lot-consistency assessments that determine that several lots of the vaccines can be made that are consistent in their safety and immunogenicity endpoints.

Controlled Human Infection Models (CHIMs) CHIMs may be part of the development pathway of a vaccine. Such models involve intentionally infecting healthy participants with a pathogen and following them carefully for development of disease. The evaluation of vaccines in these models may provide unique immunologic insights into correlates of protection, may establish a rapid pathway for downselection of candidates to move to later-phase trials or to endemic areas, and may inform vaccine approval. For example, an oral cholera vaccine for travelers in the United States was approved by the FDA based on a controlled human challenge study that established efficacy against clinical cholera. The determination to use CHIM is made in consultation with the FDA and based on the premise that the size, complexity, and expense of a study—for example, a study of a cholera vaccine with a disease endpoint in travelers—would render a phase 3 trial infeasible and significantly delay the approval of the vaccine. Other well-established CHIMs that are used to assess vaccines include those for influenza, typhoid fever, shigellosis, and malaria.

Vaccine Licensure At the conclusion of the phase 3 efficacy trials and consistency lot testing, data are presented to the national regulatory authorities who ultimately will decide whether the vaccine will be licensed. In the United States, the FDA will require that the product be labeled for how it will be used, based on the data generated in the phase 3 clinical trial. They will seek the guidance of a standing advisory group of experienced clinicians, vaccine experts, epidemiologists, and other subject matter experts, called the Vaccine and Related Biological Products Advisory Committee (VRBPAC), to advise whether the vaccine should be licensed. The FDA poses specific questions to the VRBPAC that relate to both vaccine efficacy and safety. All the adverse events reported in clinical trials are comprehensively presented and discussed with the committee. Any significant safety concerns raised may delay licensure, and additional safety studies may be proposed. Ultimately, although the FDA will seek the guidance of the VRBPAC, it will decide independently whether to license the vaccine and what requirements will be placed on that license. The inclusion of children and pregnant women in vaccine licensure will generally not occur unless the vaccines

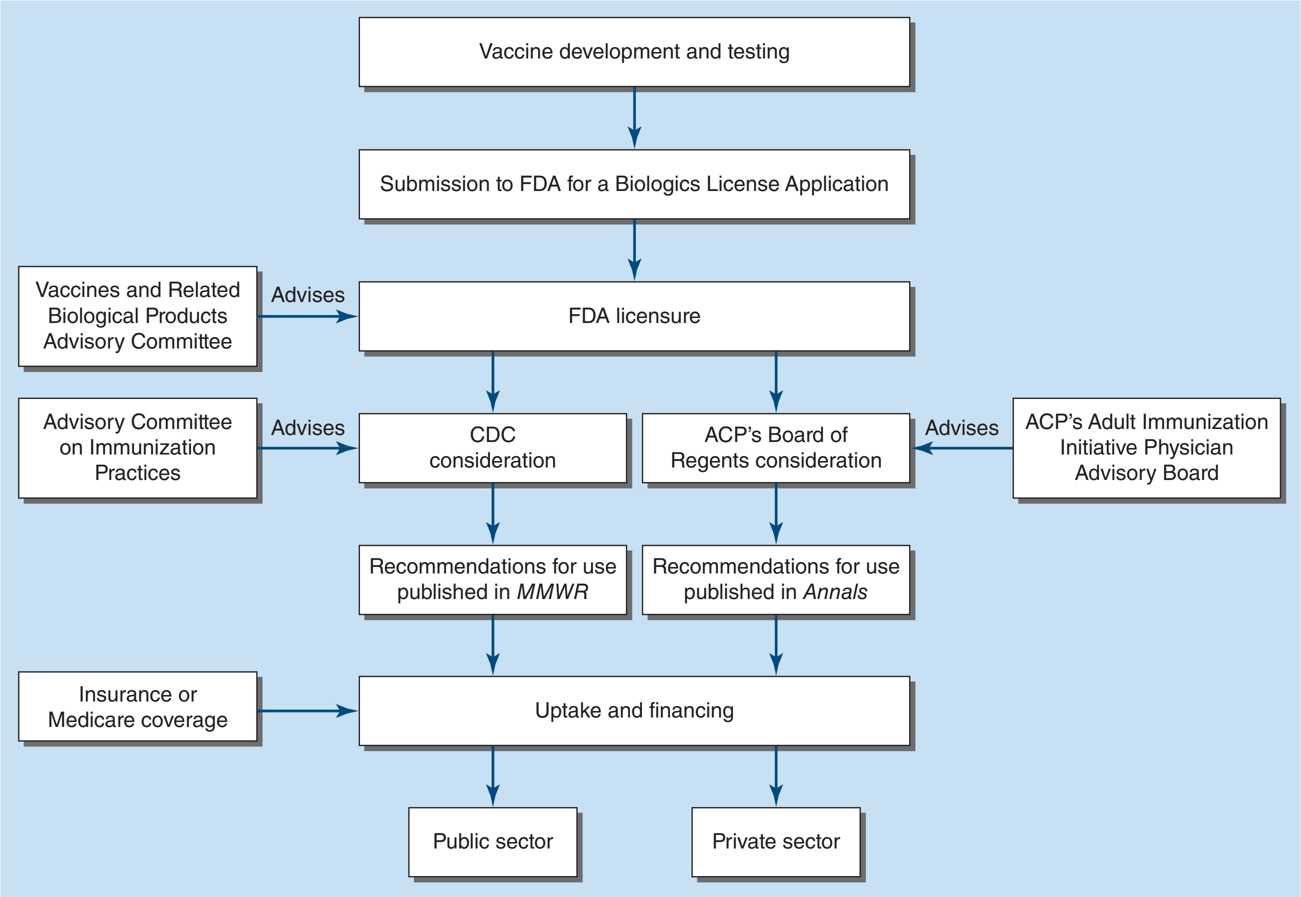
Vaccine development and testing
Submission to FDA for a Biologics License Application
Vaccines and Related Biological Products Advisory Committee
Advises
FDA licensure
Advisory Committee on Immunization Practices
Advises
CDC consideration
ACP's Board of Regents consideration
Advises
ACP's Adult Immunization Initiative Physician Advisory Board
Recommendations for use published in *MMWR*
Recommendations for use published in *Annals*
Insurance or Medicare coverage
Uptake and financing
Public sector
Private sector

FIGURE 128-3 Development and dissemination of vaccine recommendations and policies for adults. ACP, American College of Physicians. *(From Annals of Internal Medicine, JC Smith et al: Immunization Policy Development in the United States: The Role of the Advisory Committee on Immunization Practices, Vol. 150, No. 1. Copyright © 2009 American College of Physicians. All Rights Reserved. Reprinted with the permission of American College of Physicians, Inc.)*

have been specifically studied in those populations for their safety and immunogenicity. This process is outlined in Fig. 128-3.

The Advisory Committee on Immunization Practices (ACIP) Process Following licensure of vaccines by the FDA, the group that has played the major role in determining recommendations for vaccination in the United States is the Advisory Committee on Immunization Practices (ACIP) of the CDC. The ACIP provides recommendations for the use of vaccines in children, adolescents, and adults and issues a yearly vaccination schedule for all age groups and clinical conditions. Since 1995, the ACIP, American Academy of Pediatrics, American Academy of Family Physicians, American College of Physicians, American College of Obstetricians and Gynecologists, and other professional organizations have worked together to coordinate vaccine schedules and policies.

Members of the ACIP have expertise in a variety of disciplines related to the charge to the committee, including clinical medicine, epidemiology, vaccinology, public health, implementation science, and economic evaluations. The ACIP charter states, "Committee deliberations on use of vaccines to control disease in the U.S. shall include consideration of disease epidemiology and burden of disease, vaccine efficacy and effectiveness, vaccine safety, the quality of evidence reviewed, analyses and implementation issues." The committee may revise or withdraw their recommendation(s) regarding a particular vaccine as new information on disease epidemiology, effectiveness or safety, economic considerations, or other issues becomes available. Once approved by the CDC Director, the ACIP recommendations are published in the *Morbidity and Mortality Weekly Report (MMWR)* and become official CDC immunization use policy. Among the responsibilities of the ACIP is prioritizing use in settings where there are shortages of vaccines so that the highest-priority groups for whom vaccine might be recommended can receive them.

The immunization delivery system in the United States consists of both public and private providers. Since 1993 the Vaccines for Children (VFC) Program provides vaccines to children whose parents or guardians may not be able to afford them. Serving as one of the nation's most important contributors to health equity, the program helps ensure that all children have a better chance of getting their recommended vaccinations on schedule and staying healthy. In contrast, most vaccines for adults are delivered by private healthcare providers, although many of the vaccines are purchased using federal or other government funds. This may contribute to the relatively low coverage of adult vaccines compared with childhood vaccines. The Affordable Care Act requires private insurers to cover ACIP-recommended vaccines at in-network providers.

Vaccine Policy and Postlicensure Surveillance Once a vaccine is licensed for use in the United States and recommendations are made by the ACIP, systems need to be put in place to provide access to vaccines and to remove barriers to access, such as cost. Ongoing monitoring of vaccine effectiveness and safety are critical to evaluate issues such as waning immunity, to identify groups at heightened risk for vaccine failure, and to determine adverse events following vaccination that are causally related, especially rare events that could not be detected in prelicensure trials. Estimations of the burden of causally vaccine-related adverse events can then be weighed against the benefits of the vaccine to determine if any changes in recommendations are warranted. Disease surveillance is crucial to determine who continues to acquire disease, risk factors for disease, and the role of vaccine failure versus failure to vaccinate in disease prevention.

After Vaccine Implementation Prelicensure clinical trials usually enroll thousands of participants. However, there are questions that cannot be answered in these studies. These questions include (1) What is the duration of protection from disease, and is there a need for boosters? (2) What is the effectiveness of the vaccine in population groups not evaluated in the clinical trials? Effectiveness can be measured in observational studies post-licensure by comparing disease incidence in vaccinees versus nonvaccinees; (3) Are there rare causally related adverse events? If so, what is the incidence rate, and are there risk factors for developing the adverse event that could lead to vaccine contraindications? and (4) What is the impact of vaccination in protecting the community (i.e., herd immunity) by preventing or reducing transmission? Answering these questions requires a comprehensive surveillance system to detect and determine characteristics of disease in the postvaccine era and whether such disease is the result of failure to vaccinate or vaccine failure. If the former, what measures can be taken to enhance vaccine uptake, or should recommendations for vaccination be broadened if substantial numbers of cases are occurring in groups for whom vaccine is not recommended? If there is evidence of vaccine failure, is it the result of vaccine mishandling (e.g., improper storage) or is the rate of failure within expected levels (e.g., the measured vaccine effectiveness is within levels expected based on the prelicensure trials)? If effectiveness is low, are there groups at higher risk for vaccine failure, and if so, would additional doses of vaccines or alternative schedules reduce that risk?

Assessing Vaccine Safety Adequately assessing vaccine safety is critical to the success of immunization programs and requires an existing comprehensive system to monitor safety. In the United States, there are several systems in place to assess safety in the postlicensure setting. The Vaccine Adverse Event Reporting System (VAERS) allows providers, parents, and patients to report adverse events. The VAERS program functions more to raise hypotheses about whether receipt of a vaccine or vaccines causes an adverse event than it does to evaluate causation. The Vaccine Safety Datalink (VSD) is a collaborative project between CDC's Immunization Safety Office (ISO) and eight health care organizations. The VSD was initiated in 1990 and continues to monitor safety of vaccines over large populations and to conduct studies to assess rare and serious adverse events following immunization. The Clinical Immunization Safety Assessment (CISA) Project is a national network of vaccine safety experts from the CDC's ISO, seven medical research centers, and other partners, which provides a comprehensive vaccine safety public health service to the nation.

Vaccine Injury Compensation Most vaccine-preventable diseases are transmitted person-to-person. Thus, when individuals are vaccinated, they are protected from disease and indirectly are protecting others in society who either cannot be vaccinated (e.g., have medical contraindications to vaccine) or fail to make an adequate immune response. Therefore, if someone is injured by vaccine, society should provide that person compensation. This is the basis of the National Vaccine Injury Compensation Program (NVICP). This program offers compensation for the injured vaccine recipient and reduces the risk of liability for the vaccine provider and the manufacturer, as persons who receive vaccines covered by the NVICP must first go through this compensation process before suing the provider or manufacturer.

Vaccine Hesitancy Over the past decade there has been an increase in vaccine hesitancy. Reasons for vaccine hesitancy are multifactorial and include concerns over vaccine safety, questioning whether specific vaccines are needed to prevent disease, and mistrust of the public health sector and medical professionals recommending the vaccines. This hesitancy has increased since the COVID-19 pandemic coincident with the proliferation of antivaccine tropes and with concern over loss of autonomy of individuals to make decision for themselves. These concerns can often be alleviated or lessened by addressing the specific questions raised by the hesitant individual and by a presumptive approach where vaccines are presented as the standard of care. A strong recommendation for a vaccine by a trusted health care provider is one proven mechanism to reduce vaccine hesitancy. Websites maintained by the CDC, NIH, and various specialty and subspeciality organizations have extensive information about vaccines, and providers should refer their patients to these credible sources.

SUMMARY

Few programs have had the impact of vaccines in reducing health burdens. This is the result of a rigorous system to ensure that recommended vaccines are safe and effective, and of an equally rigorous system to ensure that persons for whom vaccines are recommended receive them. With the development and introduction of new vaccines, public health gains are expected to increase in coming years as long as high vaccine coverage levels can be maintained.

FURTHER READING

Cunningham AL et al: Vaccine development: From concept to early clinical testing. Vaccine 34:6655, 2016.

Li X et al: Estimating the health impact of vaccination against ten pathogens in 98 low-income and middle-income countries from 2000 to 2030: A modelling study. Lancet 397:398, 2021.

McCarthy NL et al: Monitoring vaccine safety using the Vaccine Safety Datalink: Utilizing immunization registries for pandemic influenza. Vaccine 29:4891, 2011.

National Academies of Sciences, Engineering and Medicine: The critical public health value of vaccines: Tackling issues of access and hesitancy: Proceedings of a workshop. Washington DC, National Academies Press, 2021.

National Vaccine Advisory Committee: Protecting the public's health: Critical functions of the Section 317 Immunization Program—a report of the National Vaccine Advisory Committee. Public Health Rep 128:78, 2013.

O'Leary ST et al: Strategies for improving vaccine communication and uptake. Pediatrics 153:e2023065483, 2024.

Pickering LK et al: Principles of vaccine licensure, approval and recommendations for use. Mayo Clin Proc 95:600, 2020.

Plotkin SA: Correlates of protection induced by vaccination. Clin Vaccine Immunol 17:1055, 2010.

Shattock AJ et al: Contribution of vaccination to improved survival and health: Modelling 50 years of the Expanded Programme on Immunization. Lancet 403:2307, 2024.

Walton LR et al: The history of the United States Advisory Committee on Immunization Practices (ACIP). Vaccine 33:405, 2015.

129

Immunization Principles and Vaccine Use

Sarah Meyer, Amanda Cohn, Georgina Peacock

Few medical interventions of the past century can rival the effect that immunization has had on longevity, economic savings, and quality of life. Twenty-one diseases are now preventable through vaccines routinely administered to children and adults in the United States (**Table 129-1**), and most vaccine-preventable diseases of childhood are at historically low levels (**Table 129-2**). In the past few years, the adult immunization landscape has changed substantially with an increased number of recommended routine vaccines.

The COVID-19 pandemic identified a need for improved infrastructure to ensure adult vaccination, especially for those who are uninsured and who have limited access to health systems. Health care providers in a variety of settings deliver the vast majority of vaccines in the United States and therefore play an integral role in the nation's public health system.

VACCINE IMPACT

Direct and Indirect Effects Many immunizations against specific infectious diseases protect individuals against infection and thereby prevent symptomatic illnesses. In addition, specific vaccines may blunt the severity of clinical illness (e.g., rotavirus vaccines and severe gastroenteritis) or reduce complications (e.g., zoster vaccines and

TABLE 129-1 Diseases Preventable with Vaccines Routinely Administered in the United States to Children and/or Adults

CONDITION	TARGET POPULATION(S) FOR ROUTINE USE
Pertussis	Children, adolescents, adults
Diphtheria	Children, adolescents, adults
Tetanus	Children, adolescents, adults
Poliomyelitis	Children
Measles	Children
Mumps	Children
Rubella, congenital rubella syndrome	Children
Hepatitis B	Children and high-risk adults
Haemophilus influenzae type b infection	Children and high-risk adults
Hepatitis A	Children and high-risk adults
Influenza	Children, adolescents, adults
Varicella	Children
Pneumococcal disease	Children, older adults, and high-risk adults[a]
Serogroups A, C, W, Y meningococcal disease	Adolescents and high-risk children and adults
Serogroup B meningococcal disease	High-risk children and adults[a]
Rotavirus infection	Infants
Human papillomavirus infection, cervical and anogenital cancers	Adolescents and young adults[a]
Zoster	Older adults and high-risk adults
Respiratory syncytial virus	Infants, high-risk children, pregnant persons[a]
COVID-19	Children, adolescents, adults
Dengue	High-risk children (i.e., those living in endemic areas and with laboratory confirmation of prior infection)
Mpox	High-risk adults

[a]Others in certain age groups may be vaccinated based on shared clinical decision-making.

TABLE 129-2 Decline in Vaccine-Preventable Diseases in the United States following Widespread Implementation of National Vaccine Recommendations

CONDITION	ANNUAL NO. OF PREVACCINE CASES (AVERAGE)	NO. OF CASES REPORTED IN 2023[a]	REDUCTION (%) IN CASES AFTER WIDESPREAD VACCINATION
Smallpox	29,005	0	100
Diphtheria	21,053	2	>99
Measles	530,217	47	>99
Mumps	162,344	429	>99
Pertussis	200,752	5611	97
Polio (paralytic)	16,316	0	100
Rubella	47,745	3	>99
Congenital rubella syndrome	152	0	100
Tetanus	580	15	97
Haemophilus influenzae type b infection age <5 years	20,000	27[b]	>99
Hepatitis A	117,333	11,500[c]	90
Hepatitis B (acute)	66,232	13,300[c]	80
Invasive pneumococcal infection: all ages	63,067	17,700[d]	72
Rotavirus hospitalizations (<3 years old)	62,500	16,250[e]	74
Varicella	4,085,120	26,919[f]	99

[a]2023 reported cases unless otherwise specified (provisional as of February 2024). [b]An additional 12 type b infections are estimated to have occurred among 257 reports of *H. influenzae* infection caused by unknown types among children <5 years of age. [c]Data from the CDC's Viral Hepatitis Surveillance, 2021. [d]Unpublished data from the CDC's Active Bacterial Core Surveillance, 2020. [e]Data from the CDC's New Vaccine Surveillance Network, 2021; U.S. rotavirus disease now has a biennial pattern. [f]Data from CDC's varicella program, 2021.

Source: Adapted from SW Roush et al: JAMA 298:2155, 2007 and Morb Mortal Wkly Rep 65:924, 2017.

postherpetic neuralgia). Some immunizations also reduce transmission of infectious disease agents from immunized persons to others, thereby reducing the impact of infection spread. This indirect impact is known as *herd immunity*. The level of immunization in a population that is required to achieve indirect protection of unimmunized persons varies substantially with the specific vaccine and disease. For example, because of how transmissible measles is, an estimated 95% of the population needs to be vaccinated to achieve herd immunity, whereas with polio, an estimated 80% coverage is needed.

Over the past 30 years, due to the implementation of the Vaccines for Children Program, children have had broad access to routine childhood vaccines, regardless of insurance status. For routinely recommended vaccines in the United States, major declines in rates of vaccine-preventable diseases among both children and adults have become evident (Table 129-2). For example, vaccination of children <5 years of age against *Streptococcus pneumoniae* has led to not only a 96% reduction in invasive pneumococcal disease, but also substantial reductions in incidence among adults through herd immunity. Among children born during 1994–2023, the childhood vaccination series will prevent 508 million illnesses and 1,129,000 deaths over the course of their lifetime and save nearly $2.7 trillion in societal costs (U.S.).

Control, Elimination, and Eradication of Vaccine-Preventable Diseases Immunization programs are associated with the goals of controlling, eliminating, or eradicating a disease. *Control* of a vaccine-preventable disease reduces poor illness outcomes and often limits the disruptive impacts associated with outbreaks of disease in communities, schools, and institutions. Control programs can also reduce absences from work for ill persons and for parents caring for sick children, decrease absences from school, and limit health care utilization associated with treatment visits.

Elimination of a disease is a more demanding goal than control, usually requiring the reduction to zero of cases in a defined geographic area but sometimes defined as reduction in the indigenous sustained transmission of an infection in a geographic area. As of 2024, the United States had eliminated indigenous transmission of measles, rubella, poliomyelitis, and diphtheria. Importation of pathogens from other parts of the world continues to be important, and public health efforts are intended to respond promptly to such cases in order to limit forward spread of the infectious agent.

Eradication of a disease is achieved when its elimination can be sustained without the need to continue interventions. The only vaccine-preventable disease of humans that has been globally eradicated thus far is smallpox. Although smallpox vaccine is no longer given routinely, the disease has not reemerged naturally because all chains of human transmission were interrupted through earlier vaccination efforts and humans were the only natural reservoir of the virus. Currently, a major health initiative is targeting the global eradication of polio. Two of the three wild poliovirus types (types 2 and 3) have been eradicated globally. However, endemic transmission of wild poliovirus type 1 continues in Afghanistan and Pakistan, and circulating vaccine-derived polioviruses have been detected in areas of the world where poliovirus had been eliminated, including a case of vaccine-derived poliovirus type 2 in New York in 2022. Detection of a case of disease that has been targeted for eradication or elimination is considered a sentinel event that could permit the infectious agent to become reestablished in the community or region. Therefore, such episodes must be promptly reported to public health authorities.

Epidemic and Pandemic Preparedness and Response Clusters of cases of a vaccine-preventable disease detected in an institution, a medical practice, or a community may signal important changes in the pathogen, vaccine, or environment. Several factors can give rise to increases in vaccine-preventable disease, including (1) low rates of immunization that result in an accumulation of susceptible persons (e.g., measles resurgence among vaccination abstainers); (2) changes in the infectious agent that permit it to escape vaccine-induced protection (e.g., non-vaccine-type pneumococci); (3) waning of vaccine-induced immunity (e.g., pertussis among adolescents and adults vaccinated in early childhood); and (4) point-source introductions of large inocula (e.g., food-borne exposure to hepatitis A virus). Reporting episodes of outbreak-prone diseases to public health authorities can facilitate recognition of clusters that require further interventions.

The COVID-19 pandemic highlighted the importance of a strong immunization program: robust surveillance systems to detect emerging infectious disease threats; public–private partnerships to accelerate the development of novel vaccines; and systems in place to rapidly implement a vaccination program and monitor vaccine safety and effectiveness. On a global scale, the "100 Days Mission" is being explored by the Coalition for Epidemic Preparedness Innovations (CEPI) and partners as a response to the next "Disease X" to make safe, effective vaccines, therapeutics, and diagnostics within 100 days of identification.

PUBLIC HEALTH REPORTING Recognition of suspected cases of diseases targeted for elimination or eradication—along with other diseases that require urgent public health interventions, such as contact tracing, administration of chemo- or immunoprophylaxis, or epidemiologic investigation for common-source exposure—is typically associated with special reporting requirements. Many diseases against which vaccines are routinely used, including measles, pertussis, and *Haemophilus influenzae* type b invasive disease, are nationally notifiable. Clinicians and laboratory staff have a responsibility to report some vaccine-preventable disease occurrences to local or state public health authorities according to specific case-definition criteria. All providers should be aware of state or city disease-reporting requirements and the best ways to contact public health authorities. A prompt response to vaccine-preventable disease outbreaks can greatly enhance the effectiveness of control measures.

GLOBAL CONSIDERATIONS Vaccinations are estimated to prevent 3.5–5 million deaths every year on a global scale. The COVID-19 pandemic led to disruptions of routine vaccination programs, resulting in declines in vaccine coverage in more than 100 countries. Although there have been promising signs of recovery, with diphtheria-pertussis-tetanus (DPT) coverage in 2022 nearly recovered to 2019 levels, gaps remain for other vaccines. For example, 22 million children missed their first dose of measles vaccine in 2022 compared with 19 million in 2019. As a result, the World Health Organization, the United Nations Children's Fund, Gavi, the Vaccine Alliance, and the Bill & Melinda Gates Foundation launched "The Big Catch-Up" to support recovery of childhood vaccination rates to at least prepandemic levels.

Enhancing Immunization in Adults Although immunization has become a centerpiece of routine pediatric medical visits, it has not been as well integrated into routine health care for adults. This chapter focuses on immunization principles and vaccine use in adults. Accumulating evidence suggests that immunization coverage can be increased through efforts directed at consumer-, provider-, institution-, and system-level factors. The literature suggests that the application of multiple strategies is more effective at raising coverage rates than is the use of any single strategy.

RECOMMENDATIONS FOR ADULT IMMUNIZATIONS The CDC's Advisory Committee on Immunization Practices (ACIP) makes recommendations for administration of vaccines approved or authorized by the U.S. Food and Drug Administration (FDA) for use in children and adults in the U.S. civilian population. The ACIP is a federal advisory committee that consists of up to 20 voting members (experts in fields associated with immunization) appointed by the Secretary of the U.S. Department of Health and Human Services, as well as ex officio members representing federal agencies and nonvoting representatives of various liaison organizations, including major medical societies and managed-care organizations. ACIP recommendations are available at *www.cdc.gov/acip-recs/hcp/vaccine-specific/*.

ACIP makes several types of recommendations. Routine, catch-up, and risk-based recommendations are those in which everyone in a particular age or risk group is recommended to receive vaccination. Examples include recombinant zoster vaccination for adults age ≥50 years and hepatitis B vaccination for adults age 19–59 years. Shared clinical decision-making recommendations are individually based and informed by a decision process between the health care provider and patient. With shared clinical decision-making recommendations, the decision to vaccinate is informed by the best available evidence on who may benefit from vaccination; the individual's characteristics, values, and preferences; the health care provider's clinical discretion; and the characteristics of the vaccine being considered. Examples of shared clinical decision-making recommendations include human papillomavirus vaccination of adults age 27-45 years and serogroup B meningococcal vaccination of adolescents and young adults age 16-23 years.

ADULT IMMUNIZATION SCHEDULES Immunization schedules for adults in the United States are updated regularly, through an addendum after ACIP votes on a new recommendation as well as an annual update, and can be found online (*www.cdc.gov/vaccines/hcp/imz-schedules/adult-age.html*). The adult immunization schedule is also approved by seven provider organizations and published annually in *Annals of Internal Medicine* and *Morbidity and Mortality Weekly Report* (*www.cdc.gov/mmwr*). The adult immunization schedules for 2024 are summarized in **Fig. 129-1**. Additional information and specifications are contained in the footnotes to these schedules.

IMMUNIZATION PRACTICE STANDARDS

Administering immunizations to adults involves a number of processes, such as deciding whom to vaccinate, assessing vaccine contraindications and precautions, providing vaccine information statements (VISs), ensuring appropriate storage and handling of vaccines, administering vaccines, and maintaining vaccine records. In addition, provider reporting of adverse events that follow vaccination is an essential component of the vaccine safety monitoring system. In 2014, the Standards for Adult Immunization Practice were revised to help providers take steps to ensure that their patients are fully immunized, including assessing the immunization status of patients at every clinical encounter, strongly recommending vaccines that patients need, administering vaccines or referring the patient to a vaccination provider, and documenting vaccines received by the patient.

Deciding Whom to Vaccinate Every effort should be made to ensure that adults receive all indicated vaccines as expeditiously as possible. When adults present for care, their immunization history should be assessed and recorded, and this information should be used to identify needed vaccinations according to the most current version of the adult immunization schedule. Decision-support tools incorporated into electronic health records can provide prompts for needed vaccinations. Standing orders, which are often used for routinely indicated vaccines (e.g., influenza and zoster vaccines), permit a nurse or another approved licensed practitioner to administer vaccines without a specific physician order, thus lowering barriers to adult immunization.

Assessing Contraindications and Precautions Before vaccination, all patients should be screened for contraindications and precautions. A *contraindication* is a condition that increases the risk of a serious adverse reaction to vaccination. A vaccine should not be administered when a contraindication is documented. For example, a history of an anaphylactic reaction to a dose of vaccine or to a vaccine component is a contraindication for further doses. A *precaution* is a condition that may increase the risk of an adverse event or that may compromise the ability of the vaccine to evoke immunity (e.g., administering measles vaccine to a person who has recently received immune globulins or other blood products and may consequently have transient passive immunity to measles virus). Normally, a vaccine is

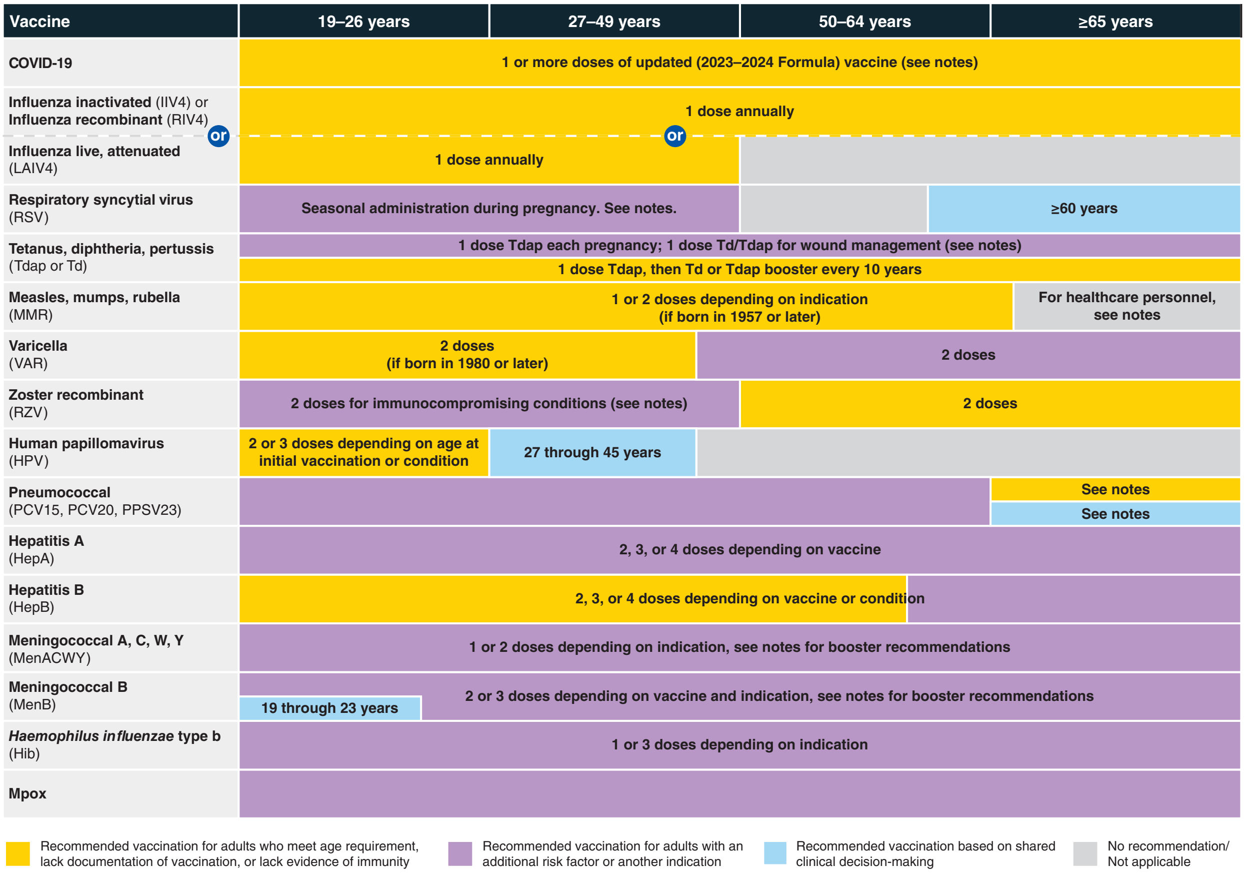

Vaccine	19–26 years	27–49 years	50–64 years	≥65 years
COVID-19	🟨 1 or more doses of updated (2023–2024 Formula) vaccine (see notes)	🟨	🟨	🟨
Influenza inactivated (IIV4) or **Influenza recombinant** (RIV4)	🟨 1 dose annually	🟨	🟨	🟨
or		or		
Influenza live, attenuated (LAIV4)	🟨 1 dose annually	🟨	⬜	⬜
Respiratory syncytial virus (RSV)	🟪 Seasonal administration during pregnancy. See notes.	🟪	⬜ 🟦 ≥60 years	🟦
Tetanus, diphtheria, pertussis (Tdap or Td)	🟪 1 dose Tdap each pregnancy; 1 dose Td/Tdap for wound management (see notes) 🟨 1 dose Tdap, then Td or Tdap booster every 10 years	🟪 🟨	🟪 🟨	🟪 🟨
Measles, mumps, rubella (MMR)	🟨 1 or 2 doses depending on indication (if born in 1957 or later)	🟨	🟨	🟨 ⬜ For healthcare personnel, see notes
Varicella (VAR)	🟨 2 doses (if born in 1980 or later)	🟨 🟪	🟪 2 doses	🟪
Zoster recombinant (RZV)	🟪 2 doses for immunocompromising conditions (see notes)	🟪	🟨 2 doses	🟨
Human papillomavirus (HPV)	🟨 2 or 3 doses depending on age at initial vaccination or condition	🟦 27 through 45 years ⬜	⬜	⬜
Pneumococcal (PCV15, PCV20, PPSV23)	🟪	🟪	🟪	🟨 See notes 🟦 See notes
Hepatitis A (HepA)	🟪 2, 3, or 4 doses depending on vaccine	🟪	🟪	🟪
Hepatitis B (HepB)	🟨 2, 3, or 4 doses depending on vaccine or condition	🟨	🟨 🟪	🟪
Meningococcal A, C, W, Y (MenACWY)	🟪 1 or 2 doses depending on indication, see notes for booster recommendations	🟪	🟪	🟪
Meningococcal B (MenB)	🟪 2 or 3 doses depending on vaccine and indication, see notes for booster recommendations 🟦 19 through 23 years	🟪	🟪	🟪
***Haemophilus influenzae* type b** (Hib)	🟪 1 or 3 doses depending on indication	🟪	🟪	🟪
Mpox	🟪	🟪	🟪	🟪

🟨 Recommended vaccination for adults who meet age requirement, lack documentation of vaccination, or lack evidence of immunity

🟪 Recommended vaccination for adults with an additional risk factor or another indication

🟦 Recommended vaccination based on shared clinical decision-making

⬜ No recommendation/ Not applicable

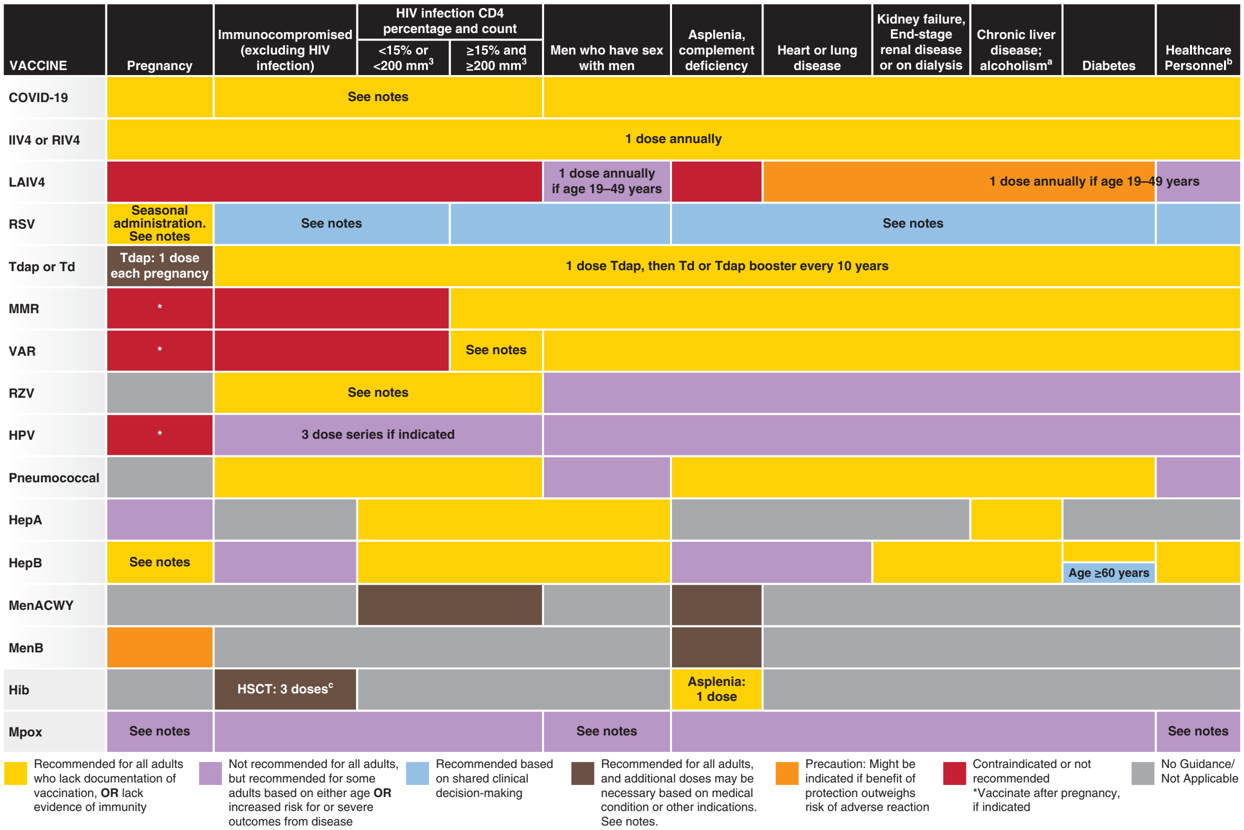

VACCINE	Pregnancy	Immunocompromised (excluding HIV infection)	HIV infection CD4 percentage and count		Men who have sex with men	Asplenia, complement deficiency	Heart or lung disease	Kidney failure, End-stage renal disease or on dialysis	Chronic liver disease; alcoholism[a]	Diabetes	Healthcare Personnel[b]
			<15% or <200 mm^3	≥15% and ≥200 mm^3							
COVID-19	🟨	🟨 See notes	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
IIV4 or RIV4	🟨 1 dose annually	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
LAIV4	🟥	🟥	🟥	🟥	🟪 1 dose annually if age 19–49 years	🟥	🟧 1 dose annually if age 19–49 years	🟧	🟧	🟧	🟪
RSV	🟨 Seasonal administration. See notes	🟦 See notes	🟦	🟦	🟦	🟦 See notes	🟦	🟦	🟦	🟦	🟦
Tdap or Td	🟫 Tdap: 1 dose each pregnancy	🟨 1 dose Tdap, then Td or Tdap booster every 10 years	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
MMR	🟥 *	🟥	🟥	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
VAR	🟥 *	🟥	🟥	🟨 See notes	🟨	🟨	🟨	🟨	🟨	🟨	🟨
RZV	⬜	🟨 See notes	🟨	🟨	🟪	🟪	🟪	🟪	🟪	🟪	🟪
HPV	🟥 *	🟪 3 dose series if indicated	🟪	🟪	🟪	🟪	🟪	🟪	🟪	🟪	🟪
Pneumococcal	⬜	🟨	🟨	🟨	🟪	🟨	🟨	🟨	🟨	🟨	🟪
HepA	🟪	⬜	🟨	🟨	🟨	⬜	⬜	⬜	🟨	⬜	⬜
HepB	🟨 See notes	🟪	🟨	🟨	🟨	🟪	🟪	🟨	🟨	🟨 🟦 Age ≥60 years	🟨
MenACWY	⬜	⬜	🟫	🟫	⬜	🟫	⬜	⬜	⬜	⬜	⬜
MenB	🟧	⬜	⬜	⬜	⬜	🟫	⬜	⬜	⬜	⬜	⬜
Hib	⬜	🟫 HSCT: 3 doses[c]	⬜	⬜	⬜	🟨 Asplenia: 1 dose	⬜	⬜	⬜	⬜	⬜
Mpox	🟪 See notes	🟪	🟪	🟪	🟪 See notes	🟪	🟪	🟪	🟪	🟪	🟪 See notes

🟨 Recommended for all adults who lack documentation of vaccination, **OR** lack evidence of immunity

🟪 Not recommended for all adults, but recommended for some adults based on either age **OR** increased risk for or severe outcomes from disease

🟦 Recommended based on shared clinical decision-making

🟫 Recommended for all adults, and additional doses may be necessary based on medical condition or other indications. See notes.

🟧 Precaution: Might be indicated if benefit of protection outweighs risk of adverse reaction

🟥 Contraindicated or not recommended *Vaccinate after pregnancy, if indicated

⬜ No Guidance/ Not Applicable

a. Precaution for LAIV4 does not apply to alcoholism.

b. See notes for influenza; hepatitis B; measles, mumps, and rubella; and varicella vaccinations.

c. Hematopoietic stem cell transplant.

FIGURE 129-1 Recommended adult immunization schedules, United States, 2024. Additional information, including footnotes for each vaccine, contraindications, and precautions, can be found at *www.cdc.gov/vaccines/hcp/imz-schedules/adult-age.html*. The recommendations in this schedule were approved by the Centers for Disease Control and Prevention (CDC) Advisory Committee on Immunization Practices (ACIP), the American Academy of Family Physicians (AAFP), the American College of Physicians (ACP), the American College of Obstetricians and Gynecologists (ACOG), and the American College of Nurse-Midwives (ACNM). For complete statements by the ACIP, visit *www.cdc.gov/acip-recs/hcp/vaccine-specific/*.

not administered when a precaution is noted. However, situations may arise when the benefits of vaccination outweigh the estimated risk of an adverse event, and the provider may decide to vaccinate the patient despite the precaution.

In some cases, contraindications and precautions are temporary and may lead to mere deferral of vaccination until a later time. For example, moderate or severe acute illness with or without fever is generally considered a transient precaution to vaccination and results in postponement of vaccine administration until the acute phase has resolved; thus, the superimposition of adverse effects of vaccination on the underlying illness and the mistaken attribution of a manifestation of the underlying illness to the vaccine are avoided. Contraindications and precautions to vaccines licensed in the United States for use in adults are summarized in **Table 129-3**. It is important to recognize conditions that are *not* contraindications in order not to miss opportunities for vaccination. For example, in most cases, mild acute illness (with or without fever) or a history of a mild to moderate local reaction to a previous dose of the vaccine are not contraindications to vaccination.

History of Immediate Hypersensitivity to a Vaccine Component A severe allergic reaction (e.g., anaphylaxis) to a previous dose of a vaccine or to one of its components is a contraindication to vaccination. While most vaccines have many components, substances to which individuals are most likely to have had a severe allergic reaction include egg protein, gelatin, and yeast. In addition, although natural rubber (latex) is not a vaccine component, some vaccines are supplied in vials or syringes that contain natural rubber latex. These vaccines can be identified by the product insert and should not be administered to persons who report a severe (anaphylactic) allergy to latex unless the benefit of vaccination clearly outweighs the risk for a potential allergic reaction. The much more common local or contact hypersensitivity to latex, such as to medical gloves (which contain synthetic latex that is not linked to allergic reactions), is *not* a contraindication to administration of a vaccine supplied in a vial or syringe that contains natural rubber latex.

SPECIFIC ADULT POPULATIONS

Pregnant Persons Several vaccines are recommended for pregnant persons in the United States: tetanus and diphtheria toxoids and acellular pertussis (Tdap), inactivated influenza, COVID-19, and respiratory syncytial virus (RSV) vaccine, as well as hepatitis B vaccine (if not already vaccinated). Tdap and RSV vaccine are administered to pregnant persons specifically to prevent severe pertussis and RSV disease, respectively, in their infants, while influenza, COVID-19, and hepatitis B (HepB) vaccines are given to protect both the pregnant person and infant.

Tdap vaccine is recommended during each pregnancy at the 27th through 36th week of gestation (preferably during the earlier part of this time period), regardless of prior vaccination status, in order to prevent pertussis in young infants. Annual influenza vaccination is recommended for everyone age ≥6 months, including pregnant persons in any trimester, ideally during September or October. Vaccination during July or August can be considered for pregnant persons in their third trimester if vaccine is available. Staying up to date with COVID-19 vaccination is recommended for everyone age ≥6 months, including pregnant persons in any trimester. It is recommended that every infant receive protection against severe RSV disease, either through maternal RSV vaccination during the 32nd through 36th week of gestation (during September to January) or through infant administration of nirsevimab, a long-acting monoclonal antibody.

Pregnant persons may need additional vaccines as well, such as HepB vaccine if not already vaccinated, since all adults age 19 through 59 years are recommended to receive HepB vaccination. Pregnancy is not a contraindication to administration of most other inactivated vaccines when otherwise indicated or when the benefits of vaccination are judged to outweigh potential risks (e.g., serogroup B meningococcal vaccination); it is recommended that human papillomavirus (HPV) and recombinant zoster vaccination are delayed until after pregnancy. Live-virus vaccines (e.g., measles, mumps, and rubella [MMR], varicella) are contraindicated during pregnancy because of the hypothetical risk that vaccine virus replication will cause congenital infection or have other adverse effects on the fetus. Patients who are breast-feeding or trying to get pregnant should stay up to date on all recommended vaccines, especially as certain congenital infections (e.g., rubella) are preventable through vaccination.

Immunocompromised Persons Immunocompromised persons are at increased risk for severe outcomes from infectious diseases. Therefore, staying up to date on recommended vaccines is important for this population. Immunocompromised patients may need to receive vaccines at an earlier age than the general population (e.g., pneumococcal and zoster at age ≥19 years instead of age ≥50 years), receive additional doses of a recommended vaccine (e.g., three-dose primary series and additional doses of COVID-19), receive vaccines based on a particular type of immunocompromising condition (e.g., meningococcal vaccination for those with complement deficiency), or have different vaccine recommendations based on degree of immune suppression, particularly for live-virus vaccines. Live-virus vaccines elicit an immune response due to replication of the attenuated (weakened) vaccine virus that is contained by the recipient's immune system. In persons with compromised immune function, enhanced replication of vaccine viruses is possible and could lead to disseminated infection with the vaccine virus. For this reason, live-virus vaccines are contraindicated for persons with severe immunosuppression, the definition of which may vary with the vaccine. Severe immunosuppression may be caused by many disease conditions, including hematologic or other malignancy. In some of these conditions, all affected persons are severely immunocompromised. In others (e.g., HIV infection), the degree to which the immune system is compromised depends on the severity of the condition, which in turn depends on the stage of disease or treatment. For example, MMR vaccine may be given to HIV-infected persons with CD4 percentage of ≥15 and count of ≥200/μL.

Older Adults As age increases, the ability to mount adaptive and innate immune responses declines, resulting in increased susceptibility to infectious diseases and reduced responses after vaccination. This immunosenescence, along with increased rates of underlying conditions, increases the risk of severe outcomes of infectious diseases in older adults. Long-term care facility residents among this age group are potentially at even greater risk due to increased frailty as well as risk of transmission associated with congregate settings. In addition to staying up to date on all recommended vaccines (e.g., COVID-19, Td/Tdap), there are specific vaccine recommendations for older adults: zoster vaccine in adults age ≥50 years, pneumococcal conjugate vaccine for adults age ≥50 years, and RSV vaccine in all adults aged ≥75 years and in those age 60–74 years at increased risk for severe RSV. Furthermore, adults age ≥65 years are recommended to receive a higher-dose influenza vaccine (i.e., high-dose, adjuvanted, or recombinant vaccine) rather than the standard-dose vaccine.

Health Care Workers Health care workers are recommended to stay up to date with all vaccines recommended for them based on age or underlying condition (including influenza and COVID-19 vaccines), and they may be recommended to receive additional vaccines due to their occupational exposure (e.g., meningococcal vaccine for laboratory personnel who handle *Neisseria meningitidis* isolates). As part of their participation in the Centers for Medicare and Medicaid Services' Hospital Inpatient Quality Reporting program, acute-care hospitals and select other facilities are required to report the proportion of their health care personnel who have received seasonal influenza and COVID-19 vaccination. Some institutions and jurisdictions have added mandates on influenza vaccination of health care workers and have expanded on earlier requirements related to vaccination or proof of immunity for hepatitis B, measles, mumps, rubella, and varicella.

VACCINE INFORMATION STATEMENTS

A Vaccine Information Statement (VIS) is a one-page (two-sided) information sheet produced by the CDC that informs vaccine

TABLE 129-3 Contraindications and Precautions for Commonly Used Vaccines in Adults

VACCINE FORMULATION	CONTRAINDICATIONS AND PRECAUTIONS
All vaccines	**Contraindication** Severe allergic reaction (e.g., anaphylaxis) after a previous vaccine dose or to a vaccine component **Precaution** Moderate or severe acute illness with or without fever. Defer vaccination until illness resolves.
Td	**Precautions** GBS within 6 weeks after a previous dose of TT-containing vaccine History of Arthus-type hypersensitivity reactions after a previous dose of TD- or DT-containing vaccines (including MenACWY). Defer vaccination until at least 10 years have elapsed since the last dose.
Tdap	**Contraindication** History of encephalopathy (e.g., coma or prolonged seizures) not attributable to another identifiable cause within 7 days of administration of a vaccine with pertussis components, such as DTaP or Tdap **Precautions** GBS within 6 weeks after a previous dose of TT-containing vaccine Progressive or unstable neurologic disorder, uncontrolled seizures, or progressive encephalopathy. Defer vaccination until a treatment regimen has been established and the condition has stabilized. History of Arthus-type hypersensitivity reactions after a previous dose of TT- or DT-containing vaccines (including MenACWY). Defer vaccination until at least 10 years have elapsed since the last dose.
HPV	**Contraindications** History of immediate hypersensitivity to yeast Pregnancy (vaccinate after pregnancy if indicated)
MMR	**Contraindications** History of immediate hypersensitivity reaction to gelatin[a] or neomycin Pregnancy (vaccinate after pregnancy if indicated) Known severe immunodeficiency (e.g., hematologic and solid tumors; chemotherapy; congenital immunodeficiency; long-term immunosuppressive therapy; severe immunocompromise due to HIV infection) Family history of altered immunocompetence (i.e., congenital or hereditary immunodeficiency in a first-degree relative), unless the immune competence of the potential vaccine recipient has been substantiated clinically or verified by a laboratory **Precautions** Recent receipt (within 11 months) of antibody-containing blood product History of thrombocytopenia or thrombocytopenic purpura Need for tuberculin skin testing or interferon γ release assay (IGRA) testing
Varicella	**Contraindications** Pregnancy (vaccinate after pregnancy if indicated) Known severe immunodeficiency History of immediate hypersensitivity reaction to gelatin[a] or neomycin Family history of altered immunocompetence (i.e., congenital or hereditary immunodeficiency in a first-degree relative), unless the immune competence of the potential vaccine recipient has been substantiated clinically or verified by a laboratory **Precaution** Recent receipt (within 11 months) of antibody-containing blood product Receipt of specific antiviral drugs (acyclovir, famciclovir, or valacyclovir) 24 h before vaccination Use of aspirin or aspirin-containing products
Influenza, inactivated, injectable	**Precaution**[b] History of GBS within 6 weeks after a previous influenza vaccine dose
Influenza, live attenuated nasal spray	**Contraindications**[b] Pregnancy Immunosuppression, including that caused by medications or by HIV infection; known severe immunodeficiency (e.g., hematologic and solid tumors; chemotherapy; congenital immunodeficiency; long-term immunosuppressive therapy; severe immunocompromise due to HIV infection) Close contact with severely immunosuppressed persons who require a protected environment, such as isolation in a bone marrow transplantation unit Receipt of oseltamivir or zanamivir within 48 h before vaccination, peramivir within the previous 5 days, or baloxavir within the previous 17 days. Active cerebrospinal fluid/oropharyngeal communications/leaks Cochlear implants Anatomic or functional asplenia (e.g., sickle cell disease) **Precautions** History of GBS within 6 weeks of a previous influenza vaccine dose Medical conditions that might predispose to higher risks of complications from influenza, such as diabetes mellitus; chronic pulmonary disease (including asthma); chronic cardiovascular disease (except hypertension); renal, hepatic, neurologic/neuromuscular, hematologic, or metabolic disorders

(*Continued*)

TABLE 129-3 Contraindications and Precautions for Commonly Used Vaccines in Adults (Continued)

VACCINE FORMULATION	CONTRAINDICATIONS AND PRECAUTIONS
Pneumococcal polysaccharide	None, other than those listed for all vaccines
Pneumococcal conjugate	None, other than those listed for all vaccines
Hepatitis A	**Contraindication** History of immediate hypersensitivity to neomycin
Hepatitis B	**Contraindications** History of immediate hypersensitivity to yeast (for Engerix-B and Recombivax-HB) Pregnancy: Heplisav-B and PreHevbrio are not recommended; use other hepatitis B vaccines if indicated
Meningococcal conjugate	None, other than those listed for all vaccines
Serogroup B meningococcal	**Precaution** Pregnancy (vaccination may be indicated if benefits of protection outweigh risks of adverse reaction)
Zoster	**Precaution** Current herpes zoster infection
COVID-19	**Precaution** Diagnosed nonsevere allergy (e.g., urticaria beyond the injection site) or nonsevere, immediate (onset <4 h) allergic reaction after administration of a previous dose Myocarditis or pericarditis within 3 weeks after a dose Multisystem inflammatory syndrome in children (MIS-C) or multisystem inflammatory syndrome in adults (MIS-A)
Respiratory syncytial virus	None, other than those listed for all vaccines
Mpox	None, other than those listed for all vaccines

[a]Extreme caution must be exercised in administering MMR, varicella, or live zoster vaccine to persons with a history of anaphylactic reaction to gelatin or gelatin-containing products. Before administration, skin testing for sensitivity to gelatin can be considered. However, no specific protocols for this purpose have been published. [b]History of severe allergic reaction (e.g., anaphylaxis) to egg is a labeled contraindication to the use of inactivated influenza vaccine and live attenuated influenza vaccine. However, CDC's Advisory Committee on Immunization Practices recommends that any licensed, recommended, and appropriate inactivated influenza vaccine or recombinant influenza vaccine may be administered to persons with egg allergy of any severity (*www.cdc.gov/acip-recs/hcp/vaccine-specific/flu.html*).

Abbreviations: DT, diphtheria toxoid; DTaP, diphtheria, tetanus, and pertussis; GBS, Guillain-Barré syndrome; HPV, human papillomavirus; MenACWY, quadrivalent meningococcal conjugate vaccine; MMR, measles, mumps, and rubella; Td, tetanus and diphtheria toxoids; Tdap, tetanus and diphtheria toxoids and acellular pertussis; TT, tetanus toxoid.

recipients (or their parents or legal representatives) about the benefits and risks of a vaccine. VISs are mandated by the National Childhood Vaccine Injury Act (NCVIA) of 1986 and—whether the vaccine recipient is a child or an adult—must be provided for any vaccine covered by the Vaccine Injury Compensation Program (VICP). As of February 2024, vaccines that are covered by the NCVIA and that are licensed for use in adults include tetanus, diphtheria, hepatitis A, hepatitis B, HPV, influenza, MMR, pneumococcal conjugate, meningococcal, polio, and varicella vaccines. When combination vaccines for which no separate VIS exists are administered (e.g., hepatitis A and B combination vaccine), all relevant VISs should be provided. In addition, although provision of a VIS is not specifically mandated under the Public Readiness and Emergency Preparedness (PREP) Act, which authorizes the Countermeasures Injury Compensation Program (CICP), CDC has published VISs for the approved or authorized vaccines currently covered under the CICP, including COVID-19 and mpox vaccines. The CICP provides compensation for injuries that occur after the administration of certain countermeasures.

VISs also exist for some vaccines not covered by the VICP or CICP, such as pneumococcal polysaccharide, Japanese encephalitis, rabies, herpes zoster, typhoid, and yellow fever vaccines. The use of these VISs is encouraged but is not mandated.

All current VISs are available at two websites: the CDC's Vaccine Information Statements site (*www.cdc.gov/vaccines/hcp/vis/*) and the Immunization Action Coalition's site (*www.immunize.org/vaccines/vis/about-vis/*). (The latter site also includes translations of the VISs.) VISs from these sites can be downloaded and printed.

■ STORAGE AND HANDLING

Injectable vaccines are packaged in multidose vials, single-dose vials, or manufacturer-filled single-dose syringes. The live attenuated nasal-spray influenza vaccine is packaged in single-dose sprayers. Oral typhoid vaccine is packaged in capsules. Some vaccines, such as MMR and varicella, come as lyophilized (freeze-dried) powders that must be reconstituted (i.e., mixed with a liquid diluent) before use. The lyophilized powder and the diluent come in separate vials. Diluents are not interchangeable but rather are specifically formulated for each type of vaccine; only the specific diluent provided by the manufacturer for each type of vaccine should be used. Once lyophilized vaccines have been reconstituted, their shelf-life is limited and they must be stored under appropriate temperature and light conditions. For example, varicella must be protected from light and administered within 30 min of reconstitution; recombinant zoster and MMR vaccines likewise must be protected from light but can be used up to 6 and 8 h after reconstitution, respectively.

Vaccines are stored either at refrigerator temperature (2–8°C) or at freezer temperature (–15°C or colder). In general, inactivated vaccines (e.g., inactivated influenza, pneumococcal polysaccharide, and meningococcal conjugate vaccines) are stored at refrigerator temperature, while vials of lyophilized-powder live-virus vaccines (e.g., varicella, MMR vaccines) are stored at freezer temperature. Diluents for lyophilized vaccines may be stored at refrigerator or room temperature. Live attenuated influenza vaccine—a live-virus liquid formulation administered by nasal spray—is stored at refrigerator temperature.

Vaccine storage and handling errors can result in the loss of vaccines worth millions of dollars, and administration of improperly stored vaccines may elicit inadequate immune responses or adverse events in patients. To improve the standard of vaccine storage and handling practices, the CDC has published detailed guidance (available at *www.cdc.gov/vaccines/hcp/admin/storage/toolkit/storage-handling-toolkit.pdf*). For vaccine storage, the CDC recommends stand-alone units—i.e., self-contained units that either refrigerate or freeze but do not do both—as these units maintain the required temperatures better than combination refrigerator/freezer units. Dormitory-style combined refrigerator/freezer units should never be used for vaccine storage.

The temperature of refrigerators and freezers used for vaccine storage must be monitored and recorded at least twice each workday. Ideally, continuous thermometers that measure and record temperature all day and all night are used, and minimal and maximal temperatures are read and documented each workday. The CDC recommends the use of calibrated digital thermometers with a probe in thermal-buffered material; more detailed information on specifications of storage units and temperature-monitoring devices is provided at the link given above.

■ ADMINISTRATION OF VACCINES

Most parenteral vaccines recommended for routine administration to adults in the United States are given by either the intramuscular (IM) or the subcutaneous (SC) route; one influenza vaccine formulation approved for use in persons 2–49 years of age is given intranasally. Some vaccines can be given by multiple routes; for example, MMRV vaccine, one of the MMR vaccines (M-M-R II), and 23-valent pneumococcal polysaccharide vaccine can be given by either IM or SC route, and the mpox vaccine can be given intradermally or subcutaneously.

Vaccines given to adults by the SC route are administered with a 5/8-inch needle into the upper outer-triceps area. Vaccines administered to adults by the IM route are injected into the deltoid muscle **(Fig. 129-2)** with a needle whose length should be selected on the basis of the recipient's sex and weight to ensure adequate penetration into the muscle. Current guidelines indicate that, for men and women weighing <152 lb (<70 kg), a 1-inch needle is sufficient; for women weighing 152–200 lb (70–90 kg) and men weighing 152–260 lb (70–118 kg), a 1- to 1.5-inch needle is needed; and for women weighing >200 lb (>90 kg) and men weighing >260 lb (>118 kg), a 1.5-inch needle is required. Additional illustrations of vaccine injection locations and techniques may be found at *www.immunize.org/wp-content/uploads/catg.d/p2020.pdf*.

Aspiration, the process of pulling back on the plunger of the syringe after skin penetration but prior to injection, is not necessary because no large blood vessels are present at the recommended vaccine injection sites.

Multiple vaccines can be administered at the same visit; indeed, administration of all needed vaccines at one visit is encouraged. Studies have shown that in general, vaccines are as effective when administered simultaneously as they are individually, and simultaneous administration of multiple vaccines is not associated with an increased risk of adverse effects and may increase uptake due to not needing to return for subsequent vaccinations. If more than one vaccine must be administered in the same limb, the injection sites should be separated by 1–2 inches so that any local reactions can be differentiated. If a vaccine and an immune globulin preparation are administered simultaneously (e.g., Td vaccine and tetanus immune globulin), a separate anatomic site should be used for each injection.

For certain vaccines (e.g., hepatitis B vaccine), multiple doses are required for an adequate and persistent antibody response. The recommended vaccination schedule specifies the interval between doses. Many adults who receive the first dose in a multiple-dose vaccine series do not complete the series or do not receive subsequent doses within the recommended interval; this lack of adherence to protocol compromises vaccine efficacy and/or the duration of protection. Providers should implement recall systems that will prompt patients to return for subsequent doses in a vaccination series at the appropriate intervals. With some exceptions (e.g., oral typhoid vaccine), an interruption in the schedule does not require restarting of the entire series or the addition of extra doses.

Syncope may follow vaccination, especially in adolescents and young adults. Serious injuries, including skull fracture and cerebral hemorrhage, have occurred. Adolescents and adults should be seated or lying down during vaccination. The majority of reported syncope episodes after vaccination occur within 15 min. The CDC recommends that vaccine providers consider observing vaccine recipients, particularly adolescents, with patients seated or lying down for 15 min after vaccination to decrease the risk of injury should they develop syncope. If syncope develops, patients should be observed until the symptoms resolve.

Anaphylaxis is a rare complication of vaccination. All facilities providing immunizations should have an emergency kit containing aqueous epinephrine for administration in the event of a systemic anaphylactic reaction.

■ MAINTENANCE OF VACCINE RECORDS

All vaccines administered should be fully documented in the patient's permanent medical record. Documentation should include the date of administration, the name or common abbreviation of the vaccine, the vaccine lot number and manufacturer, the administration site, the VIS edition, the date the VIS was provided, and the name, address, and title of the person who administered the vaccine. Increasing use

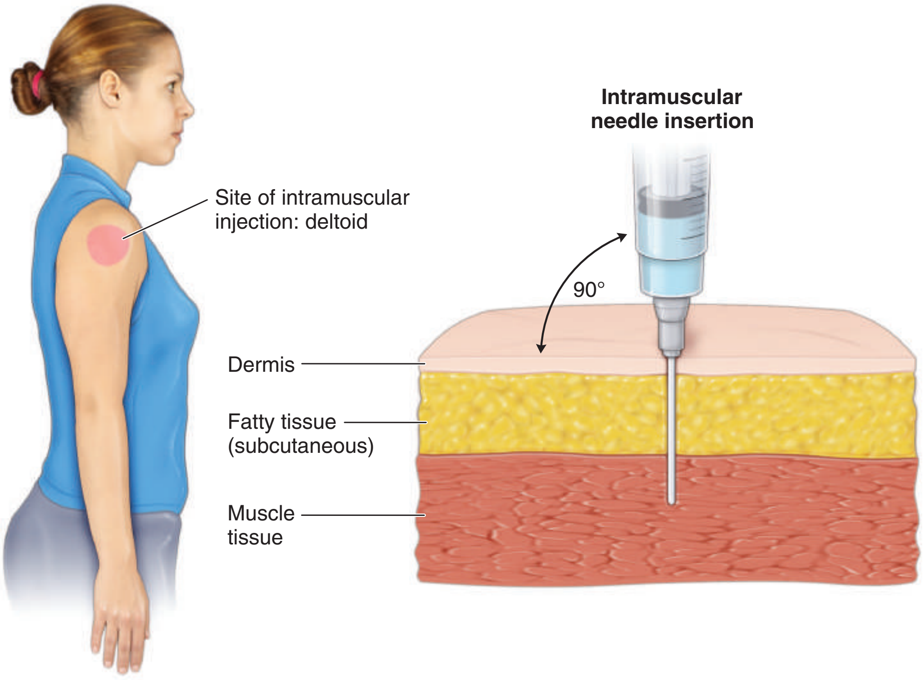
Intramuscular needle insertion
Site of intramuscular injection: deltoid
90°
Dermis
Fatty tissue (subcutaneous)
Muscle tissue

FIGURE 129-2 Technique for IM administration of vaccine.

of two-dimensional bar codes on vaccine vials and syringes that can be scanned for data entry into compatible electronic medical records and immunization information systems may facilitate more complete and accurate recording of required information.

■ VACCINE SAFETY MONITORING AND ADVERSE EVENT REPORTING

Prelicensure Evaluations of Vaccine Safety Before vaccines are licensed by the FDA, they are evaluated in clinical trials with volunteers. These trials are conducted in three progressive phases. Phase 1 trials are small, usually involving <100 volunteers. Their purposes are to provide a basic evaluation of safety and to identify common adverse events. Phase 2 trials, which are larger and may involve several hundred participants, collect additional information on safety and are usually designed to evaluate immunogenicity as well. Data gained from phase 2 trials can be used to determine the composition of the vaccine, the number of doses required, and a profile of common adverse events. Vaccines that appear promising are evaluated in phase 3 trials, which typically involve several hundred to several thousand volunteers and are generally designed to demonstrate vaccine efficacy and provide additional information on vaccine safety.

Postlicensure Monitoring of Vaccine Safety After licensure, a vaccine's safety is assessed by several mechanisms. The NCVIA of 1986 requires health care providers to report certain adverse events that follow vaccination. As a mechanism for that reporting, the Vaccine Adverse Event Reporting System (VAERS) was established in 1990 and is jointly managed by the CDC and the FDA. This safety surveillance system collects reports of adverse events associated with vaccines currently approved or authorized in the United States. *Adverse events* are defined as untoward events that occur after immunization and that might be caused by the vaccine product or vaccination process. While the VAERS was established in response to the NCVIA, reporting of any adverse event following vaccination—whether in a child or an adult, and whether or not it is believed to have actually been caused by vaccination—is encouraged through the VAERS. The adverse events that health care providers are required to report are listed in the reportable-events table at *vaers.hhs.gov/docs/VAERS_Table_of_Reportable_Events_Following_Vaccination.pdf*. The number of VAERS reports submitted varies each year. In 2019, VAERS received more than 48,000 reports. Approximately 85–90% of the reports described mild side effects such as fever, arm soreness, or mild irritability; the remaining reports are classified as serious. However, a report to VAERS does not mean that a vaccine caused an adverse event.

Anyone can file a VAERS report, including health care providers, manufacturers, and vaccine recipients or their parents or guardians. VAERS reports may be submitted online or in paper form (*vaers.hhs.gov/reportevent.html*); additional information can be obtained by email (*info@vaers.org*) or phone (800-822-7967). The VAERS form asks for the following information: the type of vaccine received; the timing of vaccination; the time of onset of the adverse event; and the recipient's current illnesses or medications, history of adverse events following vaccination, and demographic characteristics (e.g., age and sex). This information is entered into a database. The individual who reported the adverse event then receives a confirmation letter by mail with a VAERS identification number that can be used if additional information is submitted later. In selected cases of serious adverse reaction, the patient's recovery status may be followed up at 60 days and 1 year after vaccination. The FDA and the CDC have access to VAERS data and use this information to monitor vaccine safety and conduct research studies. VAERS data (minus personal information) are also available to the public.

While the VAERS provides useful information on vaccine safety, this passive reporting system has important limitations. One is that events following vaccination are merely reported; the system cannot assess whether a given type of event occurs more often than expected after vaccination. A second is that event reporting is incomplete and is biased toward events that are believed to be more likely to be due to vaccination and that occur relatively soon after vaccination. To obtain more systematic information on adverse events occurring in both vaccinated and unvaccinated persons, the Vaccine Safety Datalink project was initiated in 1991. Directed by the CDC, this project includes 11 managed-care organizations in the United States; member databases include information on immunizations, medical conditions, demographics, laboratory results, and medication prescriptions. In addition to these systems, CDC, FDA, and other federal partners use multiple other systems and data sources to conduct comprehensive vaccine safety monitoring, including CDC's V-safe system, FDA's Biologics Effectiveness and Safety (BEST) system, CMS's Medicare claims database, and the Department of Defense's Vaccine Adverse Event Clinical System. In addition, postlicensure evaluations of vaccine safety may be conducted by the vaccine manufacturer. In fact, such evaluations are often required by the FDA as a condition of vaccine licensure.

■ CONSUMER ACCESS TO AND DEMAND FOR IMMUNIZATION

By removing barriers to the consumer or patient, providers and health care institutions can improve vaccine use. Financial barriers have traditionally been important constraints. Fortunately, some progress has been made to mitigate out-of-pocket costs for vaccination of adults. The Affordable Care Act enacted in 2010 requires that most private insurance plans provide coverage for immunizations that appear on the approved immunization schedules without deductibles or copays when administered by an in-network provider. In addition, as of 2023, the Inflation Reduction Act eliminated out-of-pocket costs for ACIP-recommended vaccines for patients with Medicare prescription drug coverage and requires state Medicaid programs to cover and pay for ACIP-recommended adult vaccines without cost-sharing. However, barriers to vaccination remain for uninsured adults. A comprehensive Vaccines for Adults program to provide vaccines to uninsured adults has been proposed but has not yet been authorized for funding. To help ensure access to COVID-19 vaccines for uninsured and underinsured adults, CDC launched the Bridge Access Program for COVID-19 vaccines in 2023; this program ended in 2024.

In addition to removing financial barriers, other strategies that enhance patients' access to vaccination include extended office hours (e.g., evening and weekend hours) and scheduled vaccination-only clinics where waiting times are reduced. In recent years, pharmacies have become an increasingly important venue for adult vaccination and have helped improve equitable access to vaccines, given that the majority of the adult population lives within 5 miles of a pharmacy. Other locations outside the "medical home" (e.g., through occupational clinics, universities, and retail settings) also can expand access for adults who do not make medical visits frequently.

Health promotion efforts aimed at increasing the demand for immunization are common. Direct-to-consumer advertising by pharmaceutical companies has been used for some newer adolescent and adult vaccines. Efforts to raise consumer demand for vaccines have not increased immunization rates unless implemented in conjunction with other strategies that target strengthening of provider practices or reduction of consumer barriers. Attitudes and beliefs related to vaccination can be considerable impediments to consumer demand. Many adults view vaccines as important for children but are less familiar with vaccinations targeting disease prevention in adults. Several vaccines are recommended for adults with certain medical risk factors, but self-identification as a high-risk individual is relatively rare. Communication research suggests that adults are motivated to get vaccines to protect their own health and many would get vaccinated to protect loved ones. Adults with chronic conditions are more likely to be aware that they need to protect their own health.

■ STRATEGIES FOR PROVIDERS AND HEALTH CARE FACILITIES

Recommendation from the Provider Health care providers can have great influence on patients with regard to immunization. Studies repeatedly show that a health care provider is the most trusted source of vaccine information, and patients who receive a strong vaccine recommendation from a provider are more likely to get vaccinated than those who do not. Strong provider recommendations using a presumptive approach (e.g., “You’re due for the flu shot today” vs “What do you want to do about the flu shot?”) have been shown to improve vaccine acceptance **(Chap. 3)**. Providers should be well informed about vaccine risks and benefits so that they can address patients’ common concerns. The CDC and seven provider organizations review and update the schedule for adult immunization on an annual basis and have developed educational materials to facilitate provider–patient discussions about vaccination (*www.cdc.gov/vaccines/hcp/*).

System Supports Medical offices can incorporate a variety of methods to ensure that providers consistently offer specific immunizations to patients with indications for specific vaccines. Decision-support tools have been incorporated into some electronic health records to alert the provider when specific vaccines are indicated. Manual or automated reminders and standing orders have been discussed (see “Deciding Whom to Vaccinate,” above) and have consistently improved vaccination coverage in both office and hospital settings. Most clinicians’ estimates of their own performance diverge from objective measurements of their patients’ immunization coverage; quantitative assessment and feedback have been shown in pediatric and adolescent practices to increase immunization performance significantly. Some health plans have instituted incentives for providers with high rates of immunization coverage. Specialty providers, including obstetrician–gynecologists, may be the only providers serving some high-risk patients with indications for selected vaccines (e.g., pneumococcal polysaccharide vaccine, zoster vaccine for immunocompromised adults) as well as for other routine vaccinations (e.g., influenza, COVID-19).

Immunization Requirements Vaccination against selected communicable diseases is required for attendance at many universities and colleges as well as for service in the U.S. military and in some occupational settings (e.g., child care, laboratory, veterinary, and health care). Immunizations are recommended and sometimes required for travel to certain countries **(Chap. 130)**.

VACCINATION SETTINGS

Vaccination of adults occurs in a variety of settings. While the “medical home” remains a critical setting for vaccination of adults, other settings such as pharmacies have become an increasingly important venue for adult vaccination and help to expand equitable access and convenience to vaccination. Other vaccination settings outside the medical home include health department venues, workplaces, and schools or colleges.

Regardless of setting, it remains important for standards of immunization practice to be followed. Consumers should be provided with information on the vaccine and how to report adverse events (e.g., via provision of a VIS), and procedures should ensure that documentation of vaccine administration is forwarded to the primary care provider and the state or city public health immunization registry. Detailed documentation may be required for employment, school attendance, and travel. Personalized health records can help consumers keep track of their immunizations, and some occupational health clinics have incorporated automated immunization reports that help employees stay up to date with recommended vaccinations. Some pharmacy chain establishments are using automated systems to report immunization information to the state or local immunization information system.

PERFORMANCE MONITORING

Tracking of immunization coverage at national, state, institution, and practice levels can yield feedback to practitioners and programs and facilitate quality improvement. Healthcare Effectiveness Data and Information Set (HEDIS) measures related to adult immunization facilitate comparison of health plans. CDC utilizes a number of surveys to monitor vaccination coverage among adults and track progress toward achievement of Healthy People 2030 targets for immunization coverage. These sources include the National Health Interview Survey, the Behavioral Risk Factor Surveillance System, the National Immunization Surveys, Internet panel surveys, and other sources. Vaccination coverage among adults remains suboptimal, and state-specific immunization coverage for certain vaccines reveals substantial geographic variation in coverage. There are persistent disparities in adult immunization coverage rates by race and ethnicity, as well as other sociodemographic factors. In contrast, racial and economic disparities in immunization of young children have been dramatically reduced during the past 30 years. Much of this progress is attributed to the Vaccines for Children Program, which since 1994 has entitled eligible children, including those who are uninsured or underinsured, to receive free vaccines.

FUTURE TRENDS

Although most vaccines developed in the twentieth century targeted common acute infectious diseases of childhood, more recently developed vaccines prevent chronic conditions prevalent among adults. Hepatitis B vaccine prevents hepatitis B–related cirrhosis and hepatocellular carcinoma, HPV vaccine prevents some types of cervical cancer, genital warts, and anogenital cancers and may also prevent some oropharyngeal cancers, and the herpes zoster subunit vaccine protects against zoster and postherpetic neuralgia. New targets of vaccine development and research may further broaden the definition of vaccine-preventable disease. Research is ongoing on vaccines to prevent insulin-dependent diabetes mellitus, nicotine addiction, and Alzheimer’s disease. Expanding strategies for vaccine development are incorporating molecular approaches such as RNA, DNA, vector, and peptide vaccines. New technologies, such as the use of transdermal and other needle-less routes of administration, are being applied to vaccine delivery.

Acknowledgment

The authors thank Anne Schuchat, MD, Lisa A. Jackson, MD, and Nancy Messonnier, MD, for their significant contributions to this chapter in the previous editions.

FURTHER READING

Centers for Disease Control and Prevention: *Epidemiology and Prevention of Vaccine-Preventable Diseases*, 13th ed, Hamborsky J et al (eds). Washington DC, Public Health Foundation, 2015.

Kroger A et al: General best practice guidelines for immunization. Best practices guidance of the Advisory Committee on Immunization Practices (ACIP). Available at *www.cdc.gov/vaccines/hcp/acip-recs/general-recs/downloads/general-recs.pdf*. Accessed May 11, 2024.

McNeil MM et al: The vaccine safety datalink: Successes and challenges monitoring vaccine safety. Vaccine 32:5390, 2014.

National Vaccine Advisory Committee: Recommendations from the National Vaccine Advisory Committee: Standards for adult immunization practice. Public Health Rep 129:115, 2014.

Plotkin SA et al (eds): *Plotkin’s Vaccines*, 7th ed. Philadelphia, Elsevier, 2017.

Whitney CW et al: Benefits from immunization during the Vaccines for Children Program era—United States, 1994–2013. MMWR Morb Mortal Wkly Rep 63:352, 2014.

130

Health Recommendations for International Travel

Jesse Waggoner, Henry M. Wu

In recent decades, international travel has increased dramatically with globalization and greater access to international flights. According to the United Nations World Tourism Organization, international tourist arrivals increased 50.3% from 2010 to 2019; arrivals exceeded 1.4 billion in 2019, with the highest rate of growth in arrivals to destinations in Asia and the Pacific. In 2018, according to the United Nations Conference on Trade and Development, total global merchandise exports reached a record 19.5 trillion U.S. dollars, a nearly threefold increase over the previous two decades. Although travel in 2020 dropped drastically during the COVID-19 pandemic, there has been recovery since 2021, resuming prepandemic growth trends.

International travel has brought social, economic, and cultural benefits to the world; however, travel also widens the range of infections to which an individual may be exposed. The speed of air travel has been a major factor in the ease with which emerging infectious diseases have quickly spread worldwide in recent years. In the nineteenth century, intercontinental travel took long enough that travelers often recovered or perished from acute infections before arrival at their destinations. However, in the jet age, the time required to circumnavigate the globe has decreased to <24 h. This duration is shorter than the incubation periods for almost all infections, increasing the likelihood that infected travelers can arrive at their destinations prior to symptom onset. Epidemics can result; examples include severe acute respiratory syndrome (SARS) in 2003, Ebola virus disease in 2014, and the COVID-19 pandemic in 2020. Furthermore, introduction of pathogens into vulnerable regions can subsequently lead to infections becoming endemic, as was observed with the reintroduction of dengue throughout much of the Americas beginning in the 1970s and the global spread of HIV infections in the 1980s.

Additional challenges include the increasing diversity of travelers. While tourism, business travel, and mission work continue to be popular, recent decades have seen increasing numbers of other types of travelers, including students, migrants, medical tourists, and persons visiting their countries of origin ("visiting friends and relatives" [VFR] travelers). Furthermore, an increasing range of individuals with risk factors for illness or injury are traveling internationally, including elderly persons, infants, pregnant women, and persons with chronic medical conditions (e.g., immunocompromising conditions). Whether practicing travel medicine, primary care, or other specialties, providers will encounter patients who travel internationally. This chapter outlines key considerations and preventive measures for international travelers, particularly those traveling to low- and middle-income countries.

EPIDEMIOLOGY OF TRAVEL-RELATED CONDITIONS

Unanticipated medical problems during travel are common. Although reported rates of travel-related morbidity and mortality vary widely by destination, traveler type, and study methodology, as many as 43–79% of travelers report developing a travel-related illness. Most illnesses are minor, with diarrhea often the most commonly reported and fewer than 1–3% of travelers reporting hospitalization. Among vaccine-preventable infections in travelers to lower-income countries, influenza historically has been the most commonly reported, although COVID-19 has also become common among travelers since 2020. Typhoid and hepatitis A are reported much less often, but typically are still reported more frequently than other infections commonly discussed in travel medicine but not frequently diagnosed in travelers, including cholera, Japanese encephalitis, meningococcal disease, rabies, poliomyelitis, and yellow fever. However, outbreaks such as the emergence of yellow fever in coastal southeastern Brazil in 2017–2018 or cholera associated with the 2010 earthquake in Haiti can result in an increased incidence of travel-associated cases. Among causes of death in travelers, studies suggest that cardiovascular events (likely associated with preexisting cardiac conditions) and injury are much more common than infections. Among U.S. citizens, the injuries most commonly causing deaths during international travel in 2016–2017 were motor vehicle accidents, homicide, drowning and maritime injuries, and suicide. Stressors encountered during travel can also exacerbate or uncover psychiatric disorders, and psychological conditions such as depression and anxiety are a common reason for medical evacuation.

GENERAL APPROACH TO ADVISING INTERNATIONAL TRAVELERS

Whether advising travelers in a travel clinic or in a primary care setting, providers must cover a few key elements in a pretravel consultation **(Table 130-1)**. These include (1) a trip risk assessment based on detailed review of the itinerary and the traveler's medical profile; (2) immunizations; (3) prevention of arthropod-borne infections, including malaria chemoprophylaxis (when indicated); (4) food and water precautions and travelers' diarrhea management; and (5) prevention of injuries and other conditions associated with travel.

A detailed itinerary, including cities and areas in a country to be visited, activities, and type of accommodations, is critical for assessment of the risks of the trip and determination of the indications for specific vaccinations, malaria prophylaxis, and other preventive measures. Trip duration, sequence of countries visited, and transit stops are important considerations, especially in the assessment of immunization requirements such as those for yellow fever. Numerous online resources offer recommendations for immunizations and malaria prophylaxis, which vary significantly among countries or even within certain countries **(Table 130-2)**. The U.S. Centers for Disease Control and Prevention (CDC) travelers' health website (*www.cdc.gov/travel*) is a source of comprehensive, up-to-date, country-specific recommendations on numerous topics, including immunizations, malaria chemoprophylaxis, and travel health notices for outbreaks and emerging infections. Because recommendations and requirements can change unexpectedly, providers are advised to routinely review guidance for each country prior to making recommendations.

Consideration of the traveler's medical profile is critical to recommendations for appropriate preventive measures. Key considerations include the patient's age, medical and vaccination histories, current medications, allergies (drug, food, and environmental), and pregnancy status. Although any chronic medical problem can be relevant to travel, common issues that can require particular attention include immunocompromising conditions (including HIV infection and treatment with immunosuppressive and immunomodulatory medications), cardiac and pulmonary conditions, pregnancy status, and severe allergies. Significant hepatic and renal impairment due to any etiology can affect the choice of malaria prophylaxis.

Although travel clinics with providers specializing in travel medicine—e.g., those with a Certificate in Travelers' Health issued by the International Society of Travel Medicine—are now common in many cities worldwide, many travelers do not seek pretravel consultations, often because of an underappreciation of travel risks or a lack of awareness of the resource. Primary care providers are encouraged to routinely ask their patients about upcoming travel. Although some travel-specific vaccinations are often available only at specialized clinics, many recommended vaccines are available from general practitioners or pharmacies. Reasons to refer a patient to a travel specialist include the need for vaccines used exclusively for travel, complex itineraries or traveler medical histories, or unfamiliarity with recommended immunizations or malaria chemoprophylaxis. Because several vaccines are given as a series and all vaccines theoretically take a week or more to induce protective immunity, referral to a travel clinic at least 4–6 weeks before travel is ideal. However, when this time frame is not possible, consultations can still provide much benefit.

While the decision of whether to travel is ultimately that of the traveler, travel medicine providers play a key role in helping travelers identify the risks of a trip so that they can make an informed decision based on their personal risk tolerance. Occasionally, some situations can warrant advice against travel, including trips to areas with dangerous outbreaks or security situations or trips by a traveler who is unable to undertake critical preventive measures (e.g., travel to highly malarious areas without chemoprophylaxis). Unfortunately, travel-related vaccinations and chemoprophylactic medications can be expensive and often are not covered by health insurance. To assist travelers on a limited budget, providers can prioritize preventive measures according to degree of risk so that decisions to decline a recommendation are not based on cost alone.

IMMUNIZATIONS FOR TRAVELERS

Historically, the field of travel medicine considered immunizations as routine, recommended, or required. Because infections prevented by *routine* immunizations for children and adults are encountered worldwide, travelers should be up to date with these immunizations.

TABLE 130-1 Overview of the Pretravel Consultation

CONSULTATION ELEMENT	ITEMS TO BE COVERED	INTERVENTIONS, ADVICE
Risk Assessment		
Itinerary	• Destination countries and regions • Timing • Duration of trip • Mode of travel • Accommodations • Reason for travel and anticipated activities • Altitude	• Risk assessment that considers the itinerary, traveler, and ability to implement recommended preventive measures • Shared decision-making regarding whether to travel
Traveler	• Medical history, medications • Allergies • Pregnancy status and planning • General risk tolerance	
Immunizations		
Itinerary	Recommended and required vaccinations for itinerary	• Administration of vaccines to meet recommendations and requirements • Provision of official documentation (ICVP[a]) for required immunizations
Traveler	• Immunization history • Precautions and contraindications for specific vaccines	
Malaria and Arthropod-borne Infection Prevention		
Itinerary	• Malaria and other arthropod-borne infection risk at destination • Accommodations and activities • Local malaria resistance to chemoprophylaxis drugs	• Prescription of malaria chemoprophylaxis when indicated • Administration of vaccines for arthropod-borne infections when indicated and available • Arthropod-bite avoidance advice • Advice on early recognition of malaria symptoms
Traveler	• Precautions and contraindications to specific malaria chemoprophylaxis agents • Drug–drug interactions of regular medications with malaria chemoprophylaxis	
Gastrointestinal Illness		
Itinerary	• Destination hygienic standards and water quality • Source of meals (e.g., restaurants, street vendors, home-cooking)	See Fig. 130-2
Traveler	• Travel style • Adventurous eating habits • Drug–drug interactions of regular medications with self-treatment antibiotics	
Other Possible Topics to Address		
	• Waterborne infection (schistosomiasis, leptospirosis) prevention • Injury and crime avoidance • Animal bite and rabies prevention • Sexually transmitted infections • Altitude illness • Venous thromboembolism • Jet lag • Motion sickness • Severe food, insect sting, and environmental allergies • Travel health and medical evacuation insurance • Traveler health kits and travel with medications • Mental health and cultural adaptation	

[a]ICVP, International Certificate of Vaccine Prophylaxis ("yellow card").

Recommended travel vaccines in adults are those that are not included in routine schedules but that should be considered because of anticipated risks during travel. *Required* immunizations are those that are mandated by international regulations or specific countries for entry or exit. These three categories are not mutually exclusive or fixed, as many vaccinations that originally were used exclusively for travel (e.g., hepatitis A vaccine) are now given routinely in the United States.

For required travel immunizations, proof of vaccination is provided on the International Certificate of Vaccination Prophylaxis (ICVP, commonly called the "yellow card") issued by travel medicine providers. When appropriate, medical waivers for required vaccinations can be granted and documented on the ICVP. While enforcement of vaccine requirements can be unpredictable, travelers without proof or medical waiver for a required immunization can be subject to entry barriers, vaccination upon arrival, quarantine, or other penalties. In some situations, a vaccine may not be routinely recommended for a specific country but is still required for entry. The most common of these situations involves yellow fever vaccination, when a traveler who has recently been in an endemic country enters certain nonendemic countries (see "Yellow Fever," below).

TABLE 130-2 Online Resources for Travelers and Travel Medicine Providers

SUBJECT	RESOURCES[a]
General and country-specific recommendations, clinic directories	**CDC Travelers' Health,** *www.cdc.gov/travel* Country-specific immunization and malaria prevention advice, travel health notices, travel and yellow fever vaccine clinic listings **CDC Health Information for International Travel (*Yellow Book*)**, available at *www.cdc.gov/travel* Comprehensive travel medicine reference covering general topics and specific infections, immunizations, special traveler populations, and common itineraries **Heading Home Healthy,** *www.headinghomehealthy.org* Traveler and provider tools for trip-specific CDC recommendations **U.S. State Department,** *www.travel.state.gov* Country profiles, travel advisories, Smart Traveler Enrollment Program (STEP), traveler advice **Government of Canada Travel and Tourism,** *www.travel.gc.ca* Canadian guidelines and advice for international travel **National Travel Health Network and Centre (NaTHNaC),** *www.nathnac.net* British resource for international travel and travel medicine providers **International Society of Travel Medicine,** *www.istm.org* Global travel clinic directory, resources for travel medicine providers **International Association for Medical Assistance to Travellers,** *www.iamat.org* International clinic directory, advice on travel health insurance **World Health Organization,** *www.who.int/travel-advice* Travel health updates, traveler advice, technical guidance **American Society of Tropical Medicine and Hygiene,** *www.astmh.org* Directory for clinical specialists in tropical medicine, travel medicine, and medical parasitology
Jet lag prevention	**Jet Lag Rooster Jet Lag Calculator,** *https://sleepopolis.com/calculators/jet-lag/* Online tool to create jet lag prevention plan for a specific itinerary
For travelers with specific conditions	**American College of Obstetrics and Gynecology,** *acog.org/search#q=travel&sort=relevancy* Advice for pregnant travelers **The Global Database on HIV-Specific Travel & Residence Restrictions,** *www.hivtravel.org* General information for HIV-infected travelers and preexposure prophylaxis users and database on HIV-related entry restrictions **Asthma and Allergy Foundation of America,** *www.aafa.org/traveling-with-asthma-allergies* Advice for traveling with asthma and allergies **FARE,** *www.foodallergy.org* Resources for persons with severe food allergies, including chef card templates in several languages

[a]All websites last accessed March 21, 2024.

Abbreviation: CDC, Centers for Disease Control and Prevention.

All vaccines that are commonly administered for travel can generally be given on the same day; however, oral typhoid vaccine should be administered at least 8 h after oral cholera vaccine. Limited evidence on immunogenicity suggests that the response to live virus vaccines may be impaired if they are given on different days <28 days apart. For this reason, live virus vaccines (i.e., yellow fever, measles–mumps–rubella [MMR], live attenuated influenza, and varicella) should be given on the same day or spaced at least 28 days apart. If neither of these schedules is possible, the recommendation is to repeat the second vaccination after at least 28 days. Table 130-3 outlines common immunizations for travel in adults.

IMMUNIZATIONS FOR TRAVELERS TO MOST DESTINATIONS

Hepatitis A Hepatitis A is one of the more commonly reported vaccine-preventable infections in travelers. Transmission occurs primarily through direct person-to-person contact (fecal–oral transmission) or contaminated food and water, and travel is among the most common risk factors for infection among cases reported in the United States. Travelers are at highest risk in countries with inadequate sanitation and hygienic practices; levels of hepatitis A endemicity are highest in South Asia and sub-Saharan Africa. Although hepatitis A immunization is now routinely recommended for persons with certain medical conditions and for all children in the United States, many adults have not been vaccinated. A single dose of monovalent hepatitis A vaccine is considered protective for younger, healthy adults when given prior to travel, and a booster vaccine dose given 6–18 months after the primary dose confers lifelong immunity. For persons >40 years old, immunocompromised persons, and other individuals with chronic medical conditions that might impair immune response, administration of hepatitis A immune globulin (0.1 mL/kg) at a separate site at the time of primary vaccination can be considered. No efficacy data are available to support single-dose use of hepatitis A/B combined vaccine (Twinrix) before travel.

Hepatitis B Hepatitis B is transmitted through contact with contaminated blood, blood products, or other bodily fluids. Travelers are strongly advised against high-risk activities, including tattooing, body piercing, and unprotected sexual intercourse. Even when avoiding activities that pose a high risk of exposure to hepatitis B, travelers seeking health care can be exposed through inadequate infection-control measures or blood-product screening. While all travelers may benefit from hepatitis B vaccination, ensuring immunity to hepatitis B is particularly important for long-term travelers, health care workers, and persons who have sexual encounters. Nonimmune travelers who are departing too soon to complete the standard schedule for recombinant hepatitis B vaccine can consider the rapid hepatitis A/B combined vaccine (Twinrix) series or the recombinant hepatitis B vaccine with novel adjuvant (Heplisav-B).

Influenza and COVID-19 Since respiratory infections such as seasonal influenza and COVID-19 are the most common vaccine-preventable infections acquired during travel, travelers should be advised to be up to date with routine seasonal influenza and COVID-19 vaccinations. Because of variations in influenza seasons worldwide and

TABLE 130-3 Common Travel Immunizations

VACCINE	PRIMARY SERIES IN UNVACCINATED ADULTS	BOOSTER INTERVAL	PREGNANCY CONSIDERATIONS
Consider for Most Destinations			
Hepatitis A, inactivated (Havrix, Vaqta)	2 doses 6–12 months apart (Havrix); 2 doses 6–18 months apart (Vaqta)	None recommended	Limited data, generally considered safe, use if protection recommended
Hepatitis A/B combined (Twinrix)	3 doses at 0, 1, and 6 months; accelerated series: 3 doses on days 0, 7, and 21–30	Not recommended except booster at 12 months after accelerated primary series	Limited data, generally considered safe, use if protection recommended
Hepatitis B, recombinant and recombinant with novel adjuvant (Heplisav-B)	Recombinant, 3 doses at 0, 1, and 6 months; recombinant with novel adjuvant, 2 doses at 0 and 1 month	Not recommended after routine vaccination schedule	Recombinant vaccine, 3-dose schedule not contraindicated; Heplisav-B, no data
Measles, mumps, and rubella (MMR)	2 doses (≥28 days apart)	None recommended	Contraindicated
Specific Destinations or Activities			
Chikungunya vaccine, live	1 dose	Undetermined	Inadequate data, precaution advised
Cholera, live attenuated (Vaxchora), inactivated oral (Dukoral)	1 dose (Vaxchora); 2 doses 1–6 weeks apart (Dukoral)	Undetermined for Vaxchora; 2 years for Dukoral	Inadequate data, not recommended
Japanese encephalitis, inactivated Vero cell culture–derived (IXIARO)	2 doses on days 0 and 7–28	≥1 year after primary series	Inadequate data, precaution advised against use unless risk of infection outweighs theoretical vaccine risk
Meningococcal meningitis, quadrivalent conjugate	1 dose	5 years	May be used if indicated
Poliomyelitis, inactivated	3 doses if previously unvaccinated	Single lifetime adult booster for persons who received primary series as children	May be used if indicated
Rabies, human diploid cell (HDCV), purified chick embryo cell (PCECV)	2 doses on days 0 and 7	A third dose (21 days to 3 years after initial series) or titer check (1–3 years after primary series) recommended for travelers with ongoing (>3 year) exposure risk[a]	May be used if indicated
Tick-borne encephalitis vaccine, inactivated	3 doses; second dose 14 days to 3 months after the first, and third dose 5–12 months after the second[b]	≥3 years after primary series	Inadequate data; consider if risk of infection outweigh potential risks
Typhoid, Vi capsular polysaccharide and oral live attenuated	1 dose	2 years for Vi capsular polysaccharide; 5 years for oral live attenuated	Inadequate data; use Vi polysaccharide if protection needed
Yellow fever	1 dose	None routine; 10-year boosters recommended for certain groups	Precaution

[a]Readers are advised to refer to the Centers for Disease Control and Prevention *Yellow Book* or Advisory Committee on Immunization Practices (ACIP) 2022 guidance on rabies vaccination for detailed rabies preexposure prophylaxis recommendations based on exposure risk categories. [b]Travelers unable to receive the third dose before travel should receive two doses before travel.

year-round risk in tropical regions, influenza vaccine should be offered to unvaccinated travelers even when the influenza season has already peaked locally. Travel medicine providers can reinforce the importance of these immunizations for all travelers by emphasizing that even uncomplicated infection can disrupt travel and put their contacts at risk. Preventing infections among those at higher risk for severe illness (e.g., the elderly and persons with immunocompromising conditions) is particularly important, since the availability of appropriate treatments and supportive care can be limited in many destinations. For this reason, other vaccines for respiratory infections not typically associated with travel, including pneumococcus and respiratory syncytial virus, are potentially important for certain travelers at higher risk of severe illness. Other preventative advice for respiratory infections (wearing masks in high-risk situations or avoiding crowded, poorly ventilated areas) can be important for persons at high risk for infection.

For travelers at high risk for complicated influenza or COVID-19 infections, providing a course of antivirals for self-treatment might be considered for travel to areas with limited access to treatment. However, travelers must be counseled on the proper use of antivirals, i.e., reserving their use for confirmed or strongly suspected infections. In addition to available SARS-CoV-2 home test kits, rapid multiplex influenza and SARS-CoV-2 tests are now available for home use in the United States. These kits are useful for travelers; however, at-home tests are not sufficiently sensitive to rule out an infection and travelers should seek formal medical evaluation whenever possible.

Measles, Mumps, and Rubella According to the World Health Organization (WHO), numbers of measles cases have increased globally in recent years, with more cases worldwide in 2019 than in any other year since 1996 (Chap. 211). In the past decade, countries at all income levels have had major outbreaks. Although endemic measles transmission was eliminated in the United States in 2000, cases have increased in recent years because of outbreaks initiated by infected travelers, mostly those who were unvaccinated or undervaccinated. Unfounded vaccine hesitancy (Chap. 3) directed against measles vaccines is not uncommon in the United States and Europe, contributing to a recent worldwide resurgence. Disruption of vaccination programs during the COVID-19 pandemic has also contributed to recent outbreaks. Transmission events in airports and aboard aircraft have been reported, so all nonimmune travelers are potentially at risk. Travelers should have evidence of immunity to measles. For U.S. adults, acceptable criteria include (1) documented vaccination with two doses of live measles virus–containing vaccine given at least 28 days apart; (2) serologic evidence of immunity; (3) laboratory-confirmed illness; or (4) birth before 1957. Although two doses of measles virus–containing vaccine have been recommended for children in the United States since 1989, a significant number of older U.S. travelers born after 1956 may have received only one dose. In this country, the only measles vaccine available for use in adults is the combined MMR vaccine. Travelers should also have evidence of immunity to mumps and rubella; the criteria for immunity to these diseases are similar to those for measles except that a single dose of rubella virus–containing vaccine is considered adequate.

Tetanus, Diphtheria, and Pertussis Travelers should be up to date with age-appropriate tetanus, diphtheria, and pertussis vaccinations **(Chap. 129)**. While the risk of tetanus and pertussis exists worldwide, diphtheria is endemic or epidemic primarily in countries without adequate levels of vaccination. Because a tetanus booster is recommended for tetanus-prone injuries if the preceding booster dose was received >5 years earlier, some experts consider an early booster (before 10 years) for travelers engaging in activities with a high risk of injury in destinations with limited access to health care. A diphtheria booster within 5 years of travel is also recommended for travelers going to countries experiencing a diphtheria outbreak.

Typhoid Fever Typhoid fever, caused by *Salmonella enterica* serotype Typhi, is transmitted through ingestion of contaminated food and water. Most cases of typhoid fever reported in the United States are diagnosed in travelers after being acquired in South Asia. Africa and Southeast Asia also are considered high risk. East Asia, South America, and the Caribbean are considered lower risk. As with other food- and waterborne infections, travelers to endemic areas are at increased risk when consuming food or drink under unhygienic conditions. The risk of typhoid fever is usually lower than the risk of travelers' diarrhea and hepatitis A; however, rising levels of antimicrobial resistance in endemic regions (particularly in South Asia) have increased the importance of prevention. Two vaccines are approved for travelers: injectable Vi capsular polysaccharide vaccine and oral live attenuated vaccine. Oral vaccine is contraindicated for immunocompromised persons, and completion of the vaccination course requires 1 week (four doses separated by 48 h). Neither vaccine provides protection against paratyphoid fever (caused by *S. enterica* subtype Paratyphi A, B, or C), which is less commonly reported in U.S. travelers.

Varicella Travelers should have evidence of varicella immunity. For most U.S. travelers, this immunity can consist of documented receipt of two doses of varicella vaccine, laboratory evidence of immunity, confirmation of prior varicella or herpes zoster by a health care provider, or birth in the United States before 1980.

IMMUNIZATIONS FOR CERTAIN REGIONS OR SITUATIONS

Yellow Fever Yellow fever is endemic in much of sub-Saharan Africa and South America **(Fig. 130-1)**. Requirements for yellow fever immunization are among the most common entry rules encountered by travelers. Some endemic countries require proof of immunization for all international arrivals. Other countries, including many nonendemic countries that are prone to epidemics, require proof of immunization for arriving travelers who have recently (i.e., within 10 days prior to arrival) traveled to endemic countries. Transit stops for ≥12 h in an endemic country also can result in a requirement for proof of immunization. The United States has no requirement regarding yellow fever immunization for travelers entering the country. Country-specific recommendations and requirements for yellow fever immunization are available from the CDC Travelers' Health website (Table 130-2).

Yellow fever vaccine is available only through state-authorized official yellow fever vaccination clinics, and its administration is recorded on the ICVP. Yellow fever immunization is considered valid for entry purposes beginning 10 days after administration and extending for the lifetime of the vaccinee. Evidence indicates that a single dose of yellow fever vaccine provides most recipients with long-term protection; therefore, the previous requirement for boosting every 10 years was removed by the WHO from the International Health Regulations in 2016. Booster doses of yellow fever vaccine are still recommended after 10 years for certain individuals, including women who were pregnant during primary immunization, persons who were infected with HIV at the time of vaccination, and persons who received a hematopoietic stem cell transplant after immunization (provided they are sufficiently immunocompetent). Booster doses are also recommended for travelers who will be at particularly high risk of yellow fever during travel, including travel to areas experiencing epidemics, prolonged stays in highly endemic areas, or travel during peak transmission seasons.

All licensed yellow fever vaccines are live attenuated products. Contraindications include severe immunosuppression (e.g., during immunosuppressive or immunomodulatory therapy, in primary immunodeficiencies, or with symptomatic HIV infection or a CD4+ T lymphocyte count of <200/μL), malignant neoplasms, thymus gland disorders, and severe egg allergies. Precautions in adults include an age of ≥60 years, pregnancy, breastfeeding, and asymptomatic HIV infection with a CD4+ T lymphocyte count of 200–499/μL. Although medical waivers can be issued by yellow fever clinics, travelers must also consider the risks of traveling to endemic areas without vaccination. Common mild adverse reactions to vaccination include fevers, body aches, lymphadenopathy, localized swelling, and rash. Rare severe adverse events include anaphylaxis, neurologic complications (e.g., meningitis, encephalitis, Guillain-Barré syndrome), and yellow fever vaccine–associated viscerotropic disease (YEL-AVD). YEL-AVD is similar to yellow fever and can result in death. The risk of YEL-AVD is estimated to be ~0.3 case per 100,000 doses administered, with increased risk among immunosuppressed and elderly persons.

Poliomyelitis Although wild-type poliovirus has been eradicated from most of the world, poliomyelitis caused by circulating vaccine-derived poliovirus has been sporadically reported in numerous countries in Africa, the Middle East, and Asia where immunization rates are inadequate. For adults who have had the primary childhood polio vaccination series and are traveling to countries with reported wild-type or circulating vaccine-derived poliovirus transmission in the previous 12 months, a single booster of inactivated poliovirus vaccine before travel is recommended. This recommendation is sometimes extended to countries bordering those with poliovirus transmission when the risk of imported cases is high, especially for health care or humanitarian aid workers. Travelers who stay >4 weeks in certain countries considered high risk for exporting polio can also be subject to exit requirements for proof of recent vaccination (4 weeks to 12 months before departure). Because the list of countries with recommendations for polio booster doses is continually updated, providers should routinely review current polio booster guidance in online resources (Table 130-2).

Cholera Most countries at risk for cholera are in Africa and Asia, with the exception of the island of Hispaniola in the Americas. In 2022, an upsurge in cases was observed worldwide, and this trend continued into 2023. Transmission occurs mostly through consumption of contaminated water, although contaminated food or person-to-person contact also can be responsible. The risk to travelers is extremely low when safe food and water precautions are followed. Cholera vaccination can be considered for travelers to endemic regions, particularly those visiting areas experiencing outbreaks, health care workers, or travelers who cannot adhere to strict hygienic practices. Individuals at higher risk for severe disease (e.g., those with type O blood or comorbid conditions) and those who will be in situations where access to health care will be difficult also might consider vaccination. A single-dose live attenuated oral cholera vaccine (Vaxchora) is approved for travelers in the United States and the European Union. This vaccine had an efficacy of 90% at 10 days and 80% at 3 months after administration; the duration of protection and the need for booster doses remain to be determined. Oral killed cholera vaccines are available outside the United States.

Meningococcal Disease Endemic and epidemic meningococcal disease can occur worldwide; however, immunization is primarily recommended for high-risk travelers, including those going to countries in the African "meningitis belt" during the dry season (December to June), when large-scale epidemics can take place. Historically, meningococcal disease in the meningitis belt was caused by serogroup A; however, more recently, serogroups C, W-135, and X have caused outbreaks following the successful serogroup A vaccination program (Meningitis Vaccine Project, *https://www.path.org/our-impact/articles/about-meningitis-vaccine-project/*). Travelers should receive quadrivalent meningococcal vaccine, which protects against serogroups A, C, Y, and W-135. Conjugated meningococcal vaccines are generally preferred over polysaccharide vaccines because of their increased immunogenicity and the reduced carriage of meningococci by vaccinees; in fact, the polysaccharide vaccines are no longer available in the

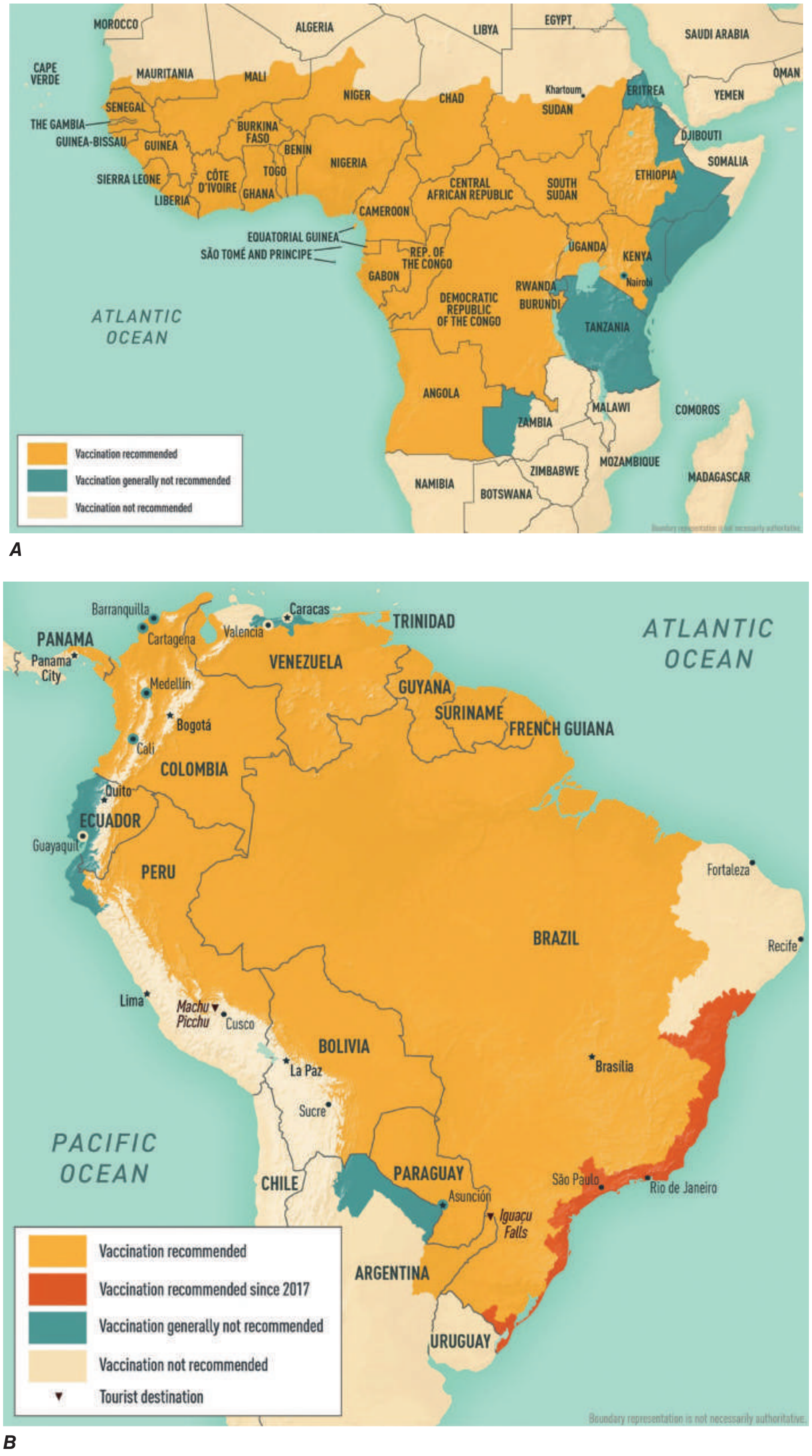
MOROCCO
ALGERIA
LIBYA
EGYPT
SAUDI ARABIA
CAPE VERDE
MAURITANIA
MALI
OMAN
Khartoum
ERITREA
YEMEN
NIGER
CHAD
SENEGAL
SUDAN
THE GAMBIA
BURKINA FASO
DJIBOUTI
GUINEA-BISSAU
GUINEA
BENIN
SOMALIA
NIGERIA
SIERRA LEONE
CÔTE D'IVOIRE
TOGO
CENTRAL AFRICAN REPUBLIC
SOUTH SUDAN
ETHIOPIA
GHANA
LIBERIA
CAMEROON
EQUATORIAL GUINEA
SÃO TOMÉ AND PRINCIPE
REP. OF THE CONGO
UGANDA
KENYA
GABON
Nairobi
RWANDA
BURUNDI
DEMOCRATIC REPUBLIC OF THE CONGO
ATLANTIC OCEAN
TANZANIA
ANGOLA
MALAWI
COMOROS
ZAMBIA
Vaccination recommended
Vaccination generally not recommended
Vaccination not recommended
MOZAMBIQUE
ZIMBABWE
MADAGASCAR
NAMIBIA
BOTSWANA
Boundary representation is not necessarily authoritative.
A
Barranquilla
Caracas
TRINIDAD
ATLANTIC OCEAN
Cartagena
Valencia
PANAMA
Panama City
VENEZUELA
Medellín
GUYANA
SURINAME
FRENCH GUIANA
Bogotá
Cali
COLOMBIA
Quito
ECUADOR
Guayaquil
PERU
Fortaleza
BRAZIL
Recife
Lima
Machu Picchu
Cusco
BOLIVIA
La Paz
Brasília
Sucre
PACIFIC OCEAN
PARAGUAY
São Paulo
Rio de Janeiro
CHILE
Asunción
Iguaçu Falls
Vaccination recommended
Vaccination recommended since 2017
ARGENTINA
Vaccination generally not recommended
URUGUAY
Vaccination not recommended
Tourist destination
Boundary representation is not necessarily authoritative.
B

FIGURE 130-1 Yellow fever recommendations in (*A*) Africa and (*B*) the Americas. Vaccination of travelers to areas with low exposure risk (designated in green) is not routinely recommended but can be considered for travelers at increased risk due to high exposure to mosquitoes or prolonged travel. Recommendations current as of November 2022. See Centers for Disease Control and Prevention Travelers' Health website for current recommendations. *(Reproduced with permission from J Nemhauser et al [eds]: Yellow Book 2024 Health Information for International Travel. New York, Oxford University Press; 2023.)*

United States. Pilgrims traveling to the Kingdom of Saudi Arabia for the Hajj and Umrah pilgrimages are required to demonstrate proof of vaccination with quadrivalent meningococcal vaccine within the preceding 3 years. Vaccination against serogroup B meningococcal disease is not recommended for travelers except in specific outbreak situations.

Japanese Encephalitis Japanese encephalitis is a potentially severe viral infection passed to humans by evening-biting mosquitoes in much of Asia and parts of the western Pacific **(Chap. 215)**. Although the WHO estimates that as many as 68,000 cases occur each year in Asia, the risk to travelers from nonendemic areas is estimated to be <1 case per 1 million travelers. However, certain travelers are at increased risk, such as long-term expatriates, persons traveling in rural areas during peak transmission seasons, and those with increased outdoor exposure (particularly during the evening). Short-term travelers to urban areas appear to be at lowest risk. In addition to mosquito avoidance measures, vaccination can be considered for travelers at risk for infection. Multiple Japanese encephalitis vaccines are available worldwide. An inactivated Vero cell culture–derived vaccine (IXIARO) is available in the United States and Europe and can be given to adults as a two-dose primary series, with doses administered 7–28 days apart.

Rabies Rabies **(Chap. 214)** is endemic to all continents except Antarctica and to numerous islands worldwide. Although many mammalian species can be infected with rabies virus, terrestrial carnivores and bats are the main reservoirs. In countries without animal control or routine pet vaccination, bites from infected dogs are often the most common source of rabies infection. Management of animal bites carrying a rabies risk is a common reason travelers seek urgent health care during or after travel. Individuals at higher risk for exposure may include children, long-term travelers, travelers whose activities will put them at higher risk of animal contact (e.g., field biologists, cavers), and travelers to remote areas. All travelers should be advised about animal bite avoidance and management of bite and scratch injuries (even minor), including thorough washing of the wound with soap and water (or povidone iodine) and immediate medical evaluation to determine whether rabies postexposure prophylaxis is indicated. The CDC-recommended postexposure prophylaxis regimen requires administration of rabies immune globulin and a rabies vaccination series. Rabies preexposure immunization series can be considered for travelers at higher risk of exposure, particularly those to destinations where access to rabies immune globulin can be difficult. Individuals receiving the rabies preexposure vaccination series should be advised that they will require urgent postexposure booster doses of vaccine following potential exposures. In the United States, the rabies preexposure vaccination series was revised in 2022, and a two-dose series (7 days apart) is now approved for travelers with risk of exposure. A third dose (between 21 days and 3 years after the initial series) or confirmation of protective titer (1–3 years after the initial series) is recommended for travelers who will be at risk for >3 years (see detailed discussion in the CDC *Yellow Book*, Table 130-2).

Tick-Borne Encephalitis Tick-borne encephalitis **(Chap. 215)** is endemic to parts of Europe and Asia, ranging from France to northern Japan. The infection is transmitted primarily through the bite of an infected *Ixodes* tick. Transmission is also possible through ingestion of unpasteurized dairy products from infected goats, sheep, or cows. Travelers to endemic regions who engage in outdoor activities in forested areas should consider vaccination prior to travel. An inactivated tick-borne encephalitis vaccine is approved in the United States, given with a three-dose primary series (second dose 14 days to 3 months after the first, and the third dose 5–12 months after the second). For short-term travelers unable to receive the third dose before travel, evidence suggests that vaccine efficacy following the initial two doses is sufficient enough to warrant vaccination with two doses. All travelers should use personal protective measures against tick bites (see "Prevention of Arthropod-Borne Infections," below).

Dengue Dengue **(Chap. 215)** is the most common arthropod-borne virus (arbovirus) worldwide, transmitted by *Aedes* mosquitoes. It is the leading cause of fever in travelers returning from most tropical and subtropical destinations outside of sub-Saharan Africa, where malaria is most common. While most travel-related cases are uncomplicated, severe dengue can occur in travelers, including dengue hemorrhagic fever and dengue shock syndrome. There is currently no U.S. Food and Drug Administration (FDA)-approved dengue vaccine in the United States for travelers, since the only FDA-approved vaccine (Dengvaxia; tetravalent live-attenuated dengue vaccine) has indications limited to individuals residing in regions of endemicity with evidence of a past dengue virus infection. In the European Union (and some other non-U.S. countries), a newer tetravalent live attenuated dengue vaccine (Qdenga) is approved for persons aged 4 and older. It can be considered for use in travelers, although questions remain regarding its ability to elicit protection against all dengue virus serotypes, specifically in seronegative recipients. All travelers should use personal protective measures against mosquito bites (see "Prevention of Arthropod-Borne Infections," below).

Chikungunya Chikungunya virus **(Chap. 215)** is an arbovirus transmitted by *Aedes* mosquitoes that has spread worldwide in recent decades. Approximately 100–200 chikungunya cases are reported annually in U.S. travelers. The highest risk of infection is among travelers to countries experiencing outbreaks. Chikungunya causes an acute febrile illness, and although life-threatening illness is rare, many patients develop severe arthralgias that might become chronic. A live chikungunya vaccine was approved in the United States in 2023 with an indication for individuals 18 years of age and older at increased risk of exposure to chikungunya virus. The vaccine is given as a single dose and is contraindicated in immunocompromised persons. In these trials, ~12% of individuals developed chikungunya-like symptoms. While most reactions were mild or moderate in severity and short-lived, 1.6% of vaccine recipients developed more significant symptoms or fever ≥39°C. New guidance on its use in travelers has been recently released (*https://www.cdc.gov/chikungunya/hcp/vaccine/index.html*).

PREVENTION OF ARTHROPOD-BORNE INFECTIONS

■ MALARIA PREVENTION

Malaria **(Chap. 231)** results from infection with one or more of five species of protozoan parasites: *Plasmodium falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*. The disease remains a significant cause of morbidity and mortality worldwide, and it is the leading cause of life-threatening infections in travelers. The majority of the disease burden, both in endemic populations and in travelers, occurs in regions of sub-Saharan Africa where *P. falciparum* is endemic. In 2018, nearly all of the 1823 malaria cases reported in the United States were acquired during travel to endemic countries. In most cases, the patient had not taken chemoprophylaxis, had adhered to the regimen inconsistently, or had taken an incorrect regimen. Among cases in U.S. residents, visiting friends or relatives was the most common reason for travel.

Although there have been significant recent developments in malaria vaccines for use in endemic countries, these vaccines are not approved for travelers. Key elements of malaria prevention in travelers include mosquito avoidance, chemoprophylaxis, and early recognition of infection to prevent the development of severe disease. *Plasmodium* species are transmitted by *Anopheles* mosquitoes, which typically bite from dusk until dawn. Limiting activities to enclosed or well-screened areas during these hours is recommended. Travelers should wear long-sleeved shirts and long pants to limit the amount of exposed skin and should apply a recommended insect repellent to any skin that remains exposed. Both the CDC and the WHO provide lists of reliable insect repellents, with products that contain DEET (20–50%) as the active ingredient being preferred by most experts. Travelers may also wear permethrin-treated clothing to increase protection. If sleeping areas are not well screened, it is recommended that individuals sleep under insecticide-treated bed nets. Indoor insecticide sprays or spatial repellents should be used with caution; the efficacy of these measures in malaria prevention has not been proven, and there may be risks from direct inhalation.

In addition to mosquito avoidance, chemoprophylaxis to prevent symptomatic disease is recommended for travelers to higher-risk regions. Current drugs used for malaria chemoprophylaxis include atovaquone-proguanil (Malarone), chloroquine (and hydroxychloroquine), doxycycline, mefloquine, primaquine, and tafenoquine (see **Chap. 231** for detailed regimens). Chemoprophylaxis recommendations should be based on careful review of the traveler's itinerary to determine malaria risk, predominant malaria species, and drug resistance patterns. The CDC and other travel medicine resources (Table 130-2) provide current country-specific risk assessments and chemoprophylaxis recommendations. Other important considerations include the traveler's medical history and routine medications; this information is essential in assessing for contraindications to specific drugs or drug–drug interactions. Primaquine and tafenoquine are active against the dormant liver stages of *P. vivax* and *P. ovale,* and both drugs are useful as prophylaxis in persons without glucose-6-phosphate dehydrogenase (G6PD) deficiency traveling to areas where these species are more common. Primaquine can also be used for presumptive antirelapse therapy in returned travelers without G6PD deficiency who had a prolonged risk of exposure to *P. vivax* or *P. ovale* and did not use primaquine or tafenoquine for chemoprophylaxis. Currently approved regimens are well tolerated, and concerns that severe side effects are common are not supported by clinical evidence (e.g., hepatitis with atovaquone-proguanil or psychosis with mefloquine). Providers should emphasize that malaria chemoprophylaxis can be life-saving and that, when prophylaxis is indicated, the benefits outweigh the low risk of serious adverse reactions.

After the onset of malaria, progression to severe disease can be rapid; therefore, early recognition of symptoms is crucial. Febrile travelers should seek evaluation as soon as possible during or after travel to malarious areas. Although the CDC does not recommend that travelers carry malaria treatment medications for standby emergency therapy as an alternative to chemoprophylaxis, this approach is accepted by some authorities for some travelers to lower-risk areas instead of chemoprophylaxis. However, because chemoprophylaxis failures can occur and locally acquired antimalarial agents can be substandard, prescription of a "reliable-supply" treatment course can be considered (in addition to chemoprophylaxis), to be used as needed under the advice of a medical professional. In the United States, atovaquone-proguanil and artemether-lumefantrine (Coartem) are available oral treatments for malaria, although atovaquone-proguanil should not be prescribed as a reliable-supply treatment course when it is being used for chemoprophylaxis.

■ PREVENTION OF OTHER ARTHROPOD-BORNE INFECTIONS

Arthropod-borne infections are present worldwide, and arboviral infections are particularly numerous. Risk of arbovirus acquisition may vary from year to year and by season (rainy vs dry). In addition, risk can increase dramatically during outbreak periods, as witnessed during recent outbreaks of infection with chikungunya virus and Zika virus in the Americas. While dengue, chikungunya, and Zika infections result from the bite of infected *Aedes* mosquitoes, arboviruses may also be transmitted by non-*Aedes* mosquitoes (Japanese encephalitis virus, Mayaro virus, o'nyong-nyong virus), ticks (tick-borne encephalitis virus, Powassan virus), and biting midges (Oropouche virus). In addition to arboviral illness, travelers are at risk for numerous other arthropod-borne infections, including—but not limited to—leishmaniasis (sand flies), filariasis (mosquitoes), African trypanosomiasis (tsetse flies), Chagas disease (triatomine bugs), and many tick-borne infections such as rickettsial diseases (e.g., African tick bite fever, scrub typhus) and Lyme disease.

Travel immunizations for arthropod-borne infections are limited to yellow fever, Japanese encephalitis, tick-borne encephalitis, dengue, and chikungunya vaccines, although availability of these vaccines can vary worldwide. Many arboviral infections have no approved vaccines, and furthermore, some travelers might be unable to receive certain vaccines due to cost or medical barriers (precautions or contraindications). For these reasons, arthropod bite avoidance remains a cornerstone of arthropod-borne infection prevention. Recommendations for avoiding arthropod bites in general are similar to those provided for *Anopheles* mosquito avoidance to prevent malaria. However, in contrast to malaria, many disease-transmitting arthropods bite during the day and may be encountered across a wider range of environments. Travelers at risk for tick bites can tuck pants into socks and perform daily self-inspections for attached ticks. Avoidance of sleeping in mud or thatch housing in areas endemic for Chagas disease is recommended.

GASTROINTESTINAL ILLNESS

Depending on the itinerary and season, as many as 30–70% of travelers report travelers' diarrhea (TD). Symptoms can include urgently passed loose stools, abdominal pain, fever, and vomiting, and more severe cases can result in volume depletion or bloody diarrhea (dysentery). The majority (~80%) of TD cases result from bacterial infections, with the most common pathogens being *Escherichia coli* and *Campylobacter*, *Shigella*, and *Salmonella* species. A minority of cases are caused by viruses, preformed bacterial toxins, and protozoa (most commonly *Giardia*). Although travelers may develop gastrointestinal illness during travel to any destination, TD is most likely to occur in low- and middle-income countries, where unhygienic food preparation practices constitute the greatest risk for development of disease.

Precautions Recommendations for the prevention of TD center on appropriate food and beverage selection as well as hand hygiene. Food is safest when cooked and served hot. Uncooked fruits and vegetables, unless they can be washed and peeled by the traveler, are considered risky. Dairy products should be pasteurized. Travelers are advised to drink bottled or purified water during travel and to avoid drinking beverages with ice, which may be made with water from unsafe sources. Before drinking any bottled or canned beverage, travelers should confirm that the factory seal is intact, as refilling of bottles with untreated water or questionable beverages is common. Street-side vendors can be particularly risky. Individuals traveling to more remote areas without access to bottled water can use one of a number of methods to purify water, including boiling, chemical treatments, filtration, or ultraviolet irradiation devices.

Prophylaxis For prophylaxis of TD, bismuth subsalicylate can be considered for short-term use. This medication, taken as two tablets (or 2 oz of liquid) four times daily, has been shown to decrease TD incidence by 50% in Mexico. The safety of bismuth subsalicylate prophylaxis when used for >21 days has not been established. Furthermore, the high dosing frequency of this regimen, common side effects (constipation, black tongue), and potential drug interactions (e.g., with acetazolamide or warfarin) limit its utility. No recommendation can be made for probiotics for TD prevention given the limited number of studies and varied results, which may vary by specific strain of probiotic used. Likewise, data supporting prebiotics for TD prevention are limited and inconclusive. Prophylactic antibiotics generally are not recommended for TD prevention given increasing concerns about adverse reactions, colonization or infection with multidrug-resistant pathogens, and development of *Clostridioides difficile* infection. However, short-term antibiotic prophylaxis (with rifaximin favored over fluoroquinolones) can be considered in rare situations for travelers at high risk for complications from TD. Oral killed cholera vaccine (Dukoral, available in Europe and Canada) shows some cross-protection against enterotoxigenic *E. coli*; however, given the wide range of TD pathogens, the protection conferred by this vaccine against TD is likely to be minimal.

Self-Treatment In general, TD is a self-limited illness, with symptoms resolving in 3–7 days for bacterial infections. Recovery times are typically shorter for infections with viral pathogens and may be prolonged for parasitic infections. For travelers who develop TD, initiation of self-treatment should be based on the patient-assessed functional impact of illness **(Fig. 130-2)**. TD can be considered *mild* (not distressing and has no impact on activities), *moderate* (distressing and may interfere with planned activities), or *severe* (incapacitating

<table>
<tr><td rowspan="2">Pre-travel</td><td colspan="4">Counsel travelers on</td></tr>
<tr><td colspan="4">• Food and water precautions
• Definition of travelers' diarrhea and severity classification
• Importance of oral rehydration and salt intake for all types of travelers' diarrhea
• Different travelers' diarrhea treatments and possible provision of antibiotics for self-treatment
• Antibiotic prophylaxis (considered only for travelers at high risk for complications from travelers' diarrhea)</td></tr>
<tr><td rowspan="4">During travel</td><td colspan="4">Self-determination of travelers' diarrhea severity</td></tr>
<tr><td rowspan="2">Mild
Tolerable, not distressing, and does not interfere with planned activities</td><td rowspan="2">Moderate
Distressing or interferes with activities</td><td colspan="2">Severe
Incapacitating or prevents planned activities</td></tr>
<tr><td>Non-dysentery</td><td>Dysentery</td></tr>
<tr><td>May use:
Loperamide
or
bismuth subsalicylates</td><td>May use:
Loperamide
or
antibiotics*
or
loperamide and antibiotics*</td><td>Should use:
Antibiotics*
with or without
loperamide</td><td>Should use:
Antibiotics*</td></tr>
</table>

*Azithromycin is preferred as first line for severe diarrhea or diarrhea acquired in areas with widespread quinolone resistance including Southeast and South Asia. Quinolones and rifaximin may be considered second-line treatment in moderate diarrhea or severe diarrhea without dysentery or high fever.

FIGURE 130-2 Management of travelers' diarrhea. *(Adapted from MS Riddle et al: Guidelines for the prevention and treatment of travelers' diarrhea: A graded expert panel report. J Travel Med 24:S63, 2017.)*

or prevents participation in planned activities). All dysenteric TD is considered severe. For all levels of severity, replacement of fluid and electrolyte losses resulting from diarrhea and/or vomiting is a mainstay of treatment. In severe cases, replacement with oral rehydration solution (available over the counter) is ideal; however, milder cases can be managed with any potable liquid. In addition, for patients with mild or moderate TD, self-treatment with antimotility agents alone (e.g., loperamide) can be considered.

Antibiotic treatment can decrease the duration of TD to 1–2 days, with potential further benefits from adjunctive loperamide. However, the risks of adverse effects, drug–drug interactions, and alterations in the traveler's microbiota are increasingly recognized in patients treated with antibiotics. Consequences of an altered microbiota can include *C. difficile* colitis and acquisition of multidrug-resistant organisms. Studies have shown that international travelers, particularly those who take antibiotics during travel, are at risk for becoming colonized with multidrug-resistant organisms, including extended-spectrum β-lactamase–producing Enterobacteriaceae. Travelers colonized with multidrug-resistant organisms may be at elevated risk for drug-resistant infections (e.g., urinary tract infection). The role of travelers in the global spread of multidrug-resistant organisms is uncertain. Given these concerns, routine self-treatment with antibiotics is recommended only in severe TD. Self-treatment with antibiotics, with or without antimotility agents, can be considered for moderate TD cases. Increasing quinolone resistance, most clearly documented in *Campylobacter* in Southeast and South Asia, has limited the utility of this antibiotic class for TD. The authors' preferred antibiotic regimen for TD self-treatment is azithromycin at a dose of 500 mg daily for 3 days, although a single 1000-mg dose also is effective. Rifaximin or rifamycin SV can also be considered, especially for persons unable to take other antibiotics because of drug–drug interactions. However, *Campylobacter* species are resistant to rifamycins, and their efficacy against dysentery has not been established. Treatment regimens for TD are covered in detail in **Chap. 138**. Empirical treatment of acute TD with antiprotozoal agents such as metronidazole is not recommended.

A small proportion of individuals develop prolonged symptoms (≥14 days), which may result from persistent infection (most often secondary to protozoa), secondary infection (*C. difficile*), or postinfectious irritable bowel syndrome. Antibiotic treatment for acute TD has not been proven to reduce the incidence of postinfectious irritable bowel syndrome, and patients with protracted symptoms should undergo a thorough evaluation.

PREVENTION OF OTHER TRAVEL-RELATED PROBLEMS

ACTIVITY-SPECIFIC INFECTION RISKS

Travelers should avoid direct contact with freshwater bodies (lakes, ponds, rivers) because of possible risks of leptospirosis and schistosomiasis. Schistosomiasis is endemic in Africa, Asia, and South America. Diving in African Rift Valley lakes (especially Lake Malawi) and rafting on the Nile River are popular activities that put travelers at risk for schistosomiasis. Appropriate footwear is important in tropical countries to prevent infection with *Strongyloides stercoralis* and hookworm as well as snakebites. Animals of all types (wild, stray, or even pets) are best avoided to minimize bite risk.

Travelers who engage in casual sex, including that with commercial sex workers, should be aware that the risk of sexually transmitted infections **(Chap. 141)** can be high, especially when barrier protections are not used. The multinational outbreak of mpox in 2022 was spread through close (skin-to-skin) contact, largely among networks of men who have sex with men (MSM). Given the continued transmission of mpox worldwide, vaccination against mpox should be considered for travelers who may engage in high-risk activity. Injection drug use, tattooing, and even acupuncture in unhygienic settings can pose a high risk for blood-borne infections such as HIV infection and hepatitis B and C.

VENOUS THROMBOEMBOLISM

Travelers are at risk for venous thromboembolism **(Chap. 122)**, particularly after long-haul flights or other extended periods of limited mobility. General precautions for prevention include ambulation during travel, calf exercises, and aisle seating. Travelers at increased risk for venous thromboembolism may benefit from graduated compression stockings. Anticoagulation may be considered for high-risk individuals.

ALTITUDE ILLNESS

Travelers to high-altitude destinations (>2500 m) should be counseled on altitude illness, and the prescription of medications for prophylaxis, such as acetazolamide, may be indicated **(Chap. 475)**. Popular high-altitude destinations include Cusco, Peru (the usual gateway to Machu

Picchu), mountains that attract climbers (e.g., Mt. Kilimanjaro), and Nepal (trekking).

MOTION SICKNESS

Motion sickness triggered by various modes of transportation can be significant for many travelers. Nonpharmacologic interventions can be helpful, such as avoiding known triggers, staying hydrated, getting adequate rest, and optimizing positions in vehicles near windows. For travelers with significant motion sickness history, over-the-counter and prescription antihistamines (e.g., dimenhydrinate, cyclizine, promethazine) can be effective in preventing and treating motion sickness; however, sedation and anticholinergic effects can limit their use. Nonsedating antihistamines appear to be less effective. A prescription for transdermal scopolamine patches can be helpful for susceptible travelers who anticipate prolonged risk periods (e.g., on cruises or extended bus trips), although like the antihistamines, anticholinergic and central nervous system effects are a concern with scopolamine, particularly in elderly travelers. While the evidence supporting complementary and integrative health approaches to prevent and treat motion sickness is limited or anecdotal, approaches such as acupressure bands or ginger are relatively simple to use and might have some benefit for individual travelers. Travelers should be counseled about potential drug–drug interactions between supplements and their routine medications.

JET LAG

Jet lag **(Chap. 33)** occurs when travel across time zones causes the traveler's circadian rhythm to become asynchronous with the local time zone. Symptoms are most significant with travel across more than three time zones and can result in poor sleep, daytime sleepiness (with poor physical and mental performance), gastrointestinal symptoms, and altered mood. Strategies to help circadian rhythms adjust to new time zones include shifting of sleep schedules prior to travel, timed light exposures after arrival, and melatonin use. Online resources to assist travelers in timing interventions to minimize jet lag are available (Table 130-2), although none has been validated in clinical trials. While caffeine use can reduce daytime drowsiness, it can also disrupt sleep. Prescription of hypnotics (e.g., zolpidem) for travel-related insomnia should generally be avoided, since adverse effects, including excessive fatigue and impaired cognition upon awakening, can be problematic during travel. When used, the lowest effective dose of a hypnotic medication should be used, and travelers should be cautioned about use during flights (when extended immobilization is problematic) or any situation when a full course of sleep is not possible. Sedative-hypnotic and anxiolytic medications are among the classes of drugs that are potentially restricted by certain countries (see "Traveling with Prescription Medications," below).

INJURIES

International travel presents numerous factors that contribute a higher risk of injuries and death. Travelers may face unfamiliar environments and language barriers and rely heavily on other people (e.g., drivers and tour operators) for protection. Furthermore, in low- and middle-income countries, safety protections that are typical in high-income countries are often less stringent, unenforced, or nonexistent. Travelers often exhibit increased risk-taking behaviors during travel, frequently in association with the use of alcohol. When injuries do occur, access to adequate trauma care can be limited. Motor vehicle accidents are a common cause of injury deaths in travelers. In addition to poor road conditions, traffic rules are often less strictly enforced overseas. Riding on motorbikes (especially without a helmet), on overcrowded public transit, in improperly maintained vehicles, and in vehicles without seatbelts should be avoided. Drowning prevention and crime avoidance are important safety topics. The U.S. State Department provides country-specific safety and security advice for U.S. travelers (Table 130-2).

TRAVELERS WITH PREEXISTING MEDICAL CONDITIONS

Travel is increasingly common for persons with chronic medical conditions. Risks vary depending on the condition, destination, and activities. Travelers with chronic medical conditions are encouraged to plan their trips carefully and to consult with their physicians to assess fitness for travel. Notably, cardiovascular events are a frequent cause of in-flight emergencies and death during travel. Travel across time zones and changes in diet can create challenges in conditions—e.g., diabetes mellitus—that require regulation of diet and consistent medication timing. Providers can assist travelers by providing copies of prescriptions (or medication lists), a medical problems list, and a baseline electrocardiogram.

Adverse events caused by drug–drug interactions can be difficult to manage during travel, especially in destinations with limited emergency care. Therefore, providers should review the traveler's medication list for potential drug interactions when prescribing prophylactic or self-treatment medications. Azithromycin and quinolones prescribed for travelers' diarrhea self-treatment can cause additive QT interval prolongation when used with some antidepressant and antiarrhythmic medications. Malaria prophylaxis medications can affect the international normalized ratio in patients who are taking warfarin.

IMMUNOCOMPROMISED TRAVELERS

An increasing number of immunocompromised persons are traveling, including organ transplant recipients, HIV-infected persons, cancer patients, persons with asplenia, and persons receiving immunosuppressive therapies (e.g., biologic agents, antimetabolites, or chronic high-dose glucocorticoids). Although each of these conditions has unique risks, general concerns include increased susceptibility to infection, decreased vaccine efficacy, and—for patients with severe immunosuppression—contraindications to live virus vaccines. Routinely recommended precautions (e.g., food and water hygiene, insect bite avoidance) are particularly important for travelers with immunocompromising conditions.

Conditions associated with severe immunocompromise that preclude use of live virus vaccines include active leukemia or lymphoma, generalized malignancy, graft-versus-host disease, HIV/AIDS (with a CD4+ count of <200 cells/μL), and congenital immunodeficiencies. Immunosuppressive therapies that preclude live virus vaccines include high-dose glucocorticoid treatment (defined as ≥20 mg of prednisone or the equivalent daily for ≥2 weeks), alkylating agent administration, antimetabolite therapy (e.g., azathioprine, methotrexate), transplant-related immunosuppression, cancer chemotherapy, radiation therapy, and treatment with biologic agents, including tumor necrosis factor blockers, checkpoint inhibitors, and lymphocyte-depleting agents. If possible, travel immunizations should be administered prior to iatrogenic immunosuppression (≥2 weeks for inactivated vaccines and ≥4 weeks for live vaccines). The duration of immunosuppression after discontinuation of immunosuppressive therapies can be prolonged, particularly for biologic agents. Travelers unable to receive a required yellow fever vaccine because of immunosuppression should be given a medical waiver if travel cannot be avoided. Providers are advised to review the detailed guidance for immunizing immunocompromised travelers provided in the CDC *Yellow Book* (Table 130-2) and other resources.

HIV-INFECTED TRAVELERS

Infection risk in HIV-infected individuals generally correlates with the level of immunosuppression (i.e., the CD4+ T-cell count). Adults with HIV infection and CD4+ counts of >500 cells/μL are generally considered to have levels of risk similar to those faced by travelers without immunocompromising conditions. Live MMR and varicella vaccines can generally be administered to HIV-infected travelers with a CD4+ count of >200 cells/μL for ≥6 months. Guidance for yellow fever immunization of HIV-infected persons is reviewed above. Oral live attenuated typhoid and live attenuated influenza vaccines should not be administered to HIV-infected persons, given the availability of polysaccharide and inactivated versions of these vaccines, respectively.

The number of countries that restrict the entry of HIV-infected persons has decreased in recent years, particularly for short-term travelers and tourists. However, HIV-infected travelers should review the policies of their destination, especially when they plan to work abroad or stay for longer terms. Unnecessary disclosures of HIV status to customs or immigration officials should be avoided. Resources include embassies in destination countries, the U.S. State Department, and online resources for travelers with HIV infection (Table 130-2).

PREGNANT TRAVELERS

Travel medicine providers should assess pregnancy status and the possibility of conception during travel. The pregnant traveler faces numerous unique risks. These include limited availability of emergency care for pregnancy complications, increased risk of certain infections, and exposure to specific infections that can result in pregnancy complications. Although most airlines will allow pregnant women to travel up to 36 weeks of pregnancy, the American College of Obstetrics and Gynecology recommends travel during the middle period of pregnancy (weeks 14–28), when morning sickness has improved, before mobility becomes impaired, and when the risk of spontaneous abortion or premature labor is minimal. Pregnancy-related contraindications and relative contraindications to travel are numerous and are reviewed in the CDC *Yellow Book* (Table 130-2).

Pregnant travelers are at increased risk of various infections (e.g., malaria, influenza, hepatitis E, listeriosis) and/or severity of illness when traveling. Some infections, notably Zika virus disease, toxoplasmosis, and rubella, can result in birth defects or fetal death. Pregnant travelers should contemplate the infectious risks of their destination and consider delaying travel to areas where particularly dangerous infections, such as malaria or Zika, are present. Currently, only mefloquine and chloroquine are approved for malaria chemoprophylaxis in pregnancy, and plasmodial resistance to these drugs can further limit options. Travel immunizations considered safe during pregnancy also are limited (Table 130-3). When used as directed, insect repellents registered by the Environmental Protection Agency, such as DEET, are considered safe during pregnancy.

TRAVELERS WITH SEVERE ALLERGIES

Travelers with severe allergies to food, insect stings, and environmental allergens can be at increased risk during travel, particularly in destinations without adequate emergency care. Avoidance of food allergens can be challenging, particularly when eating in restaurants or catered settings. Language barriers can present difficulties in avoiding exposures to allergens in food or medications. Regional variations in culinary practices and ingredients can lead to unexpected food allergen exposure. Outdoor activities can increase the risk of stings by hymenopterous insects (bees, wasps, ants). Solo travelers can face particular challenges when experiencing severe allergies.

Providers should ensure that travelers have an emergency care plan for severe allergies and an adequate supply of emergency self-treatment, including epinephrine auto-injectors and antihistamines. For travelers who will have no immediate access to health care, it can be prudent to bring other medications that might be indicated, including rescue bronchodilator inhalers (for individuals with asthma) and short courses of glucocorticoids. Written documentation of the allergic disorder and self-treatment medications should be carried, especially when injectable medications are involved. Travelers with severe food allergies should alert restaurants and hosts. Printable food allergy alert cards in various languages are available online (Table 130-2).

OTHER PRETRAVEL PREPARATIONS

TRAVEL HEALTH KITS

A carefully planned travel health kit can minimize the need to seek care for self-treatable conditions. The ideal contents depend on the destination, duration, and activities during travel as well as on individual health issues. Routine and trip-specific prescription medications (e.g., malaria prophylaxis, travelers' diarrhea self-treatment antibiotics) should be carried in original labeled prescription bottles to aid in identification. A digital thermometer and typically used over-the-counter medications, such as analgesics and antipyretics, antidiarrheal medications, medications for motion sickness, antacids, laxatives, oral rehydration salts, antihistamines, and topical steroid creams, can be important. Basic first-aid items, such as gloves, bandages, tape, antibiotic ointment, and tweezers, are helpful. Critical medications should always be carried and not packed in checked luggage; however, travelers must consider any restrictions about flying with sharp objects or liquids, particularly in their carry-on baggage.

TRAVELING WITH PRESCRIPTION MEDICATIONS

Carrying copies of prescriptions (or a signed medication list from a physician's office) is recommended. Many countries, particularly in Asia, the Middle East, and Africa, have stringent restrictions on certain drugs that are less restricted in the United States. These regulations can include controlled substances such as opioid analgesics, anxiolytics and sedatives, and medications for attention deficit hyperactivity disorder. Even some over-the-counter medications such as pseudoephedrine and diphenhydramine are restricted in certain countries. Requirements for traveling with restricted medications can include carrying copies of prescriptions or even obtaining advance approval from the destination country's health ministry. Levels of enforcement and penalties for violations vary widely. Travelers who plan to carry potentially restricted medications should contact the embassy of their destination to review policies. The International Narcotics Control Board also maintains a listing of regulations for travelers carrying controlled medications to specific countries (Table 130-2). Travelers taking antiviral medications for HIV preexposure prophylaxis going to countries that restrict entry to persons living with HIV/AIDS might take specific precautions to avoid challenges when entering (see online resource for HIV-infected travelers, Table 130-2). Travelers should carry amounts appropriate for the duration of their itinerary, including a limited number of extra doses for use in cases of a delay in return.

Although many medications that require prescription in the United States are available over the counter overseas, the quality of locally acquired pharmaceuticals can vary. Counterfeit medications, particularly antimalarials, are common in much of the world. Whenever possible, travelers are cautioned to avoid obtaining critical medications such as malaria chemoprophylaxis during travel.

HEALTH CARE OVERSEAS AND TRAVEL HEALTH INSURANCE

Travelers should consider where they would seek urgent or emergency health care, particularly if they have chronic health conditions, are pregnant, or will participate in activities with a high risk of injury or illness (e.g., altitude sickness). International travelers should be aware that most countries do not accept routine health insurance from other countries and that such insurance is unlikely to cover out-of-pocket health care costs or to provide assistance in identifying providers overseas. Travelers are advised to review their health insurance policies before travel to assess the scope of international coverage (including emergency care, hospitalization, psychiatric care, and obstetric care, if applicable) and the availability of 24-h physician-backed support.

For many travelers, supplemental insurance coverage of some type is prudent. *Travel insurance* usually consists of coverage for financial losses due to trip cancellation (e.g., due to unexpected illness) or lost baggage. *Supplemental travel health insurance* policies cover health care costs overseas and typically provide 24-h support centers. *Medical evacuation (medevac) insurance* can be a part of a travel health insurance policy or a stand-alone policy and covers medical evacuation when it is determined that the local level of care is inadequate. Further information on travel health insurance is available from online resources listed in Table 130-2.

SPECIAL TRAVEL POPULATIONS

Travelers are increasingly diverse in their reasons for and types of travel, each of which poses unique risks and challenges **(Table 130-4)**. A major challenge in travel medicine is presented by VFR travelers who are visiting their countries of origin. VFR travelers face increased risks for travel-related infections, as they often travel to areas not frequented by tourists, stay in local homes, and adopt local food and transportation habits. Immunity resulting from malaria infection is not long-term, but immigrants from malaria-endemic countries often incorrectly assume that they are immune. Barriers to appropriate

TABLE 130-4 Risks and Prevention Strategies in Special Travel Populations

GROUP	RISKS AND CHALLENGES	PREVENTION STRATEGIES
Travelers visiting friends and relatives (VFR)	• Greater likelihood of visiting areas outside usual travel destinations • Frequent adoption of food, accommodation, and transportation habits similar to those of locals • May underestimate importance of travel immunizations or malaria prophylaxis • Financial or cultural barriers to seeking pretravel advice or immunization	• Questions about planned travel during routine care visits • Prioritization of vaccines and prophylaxis for highest-risk infections when resources are limited
Budget travelers	• Financial barriers to seeking pretravel advice or immunization • Lower quality of accommodations, transportation, and food establishments	• Prioritization of vaccines and prophylaxis for highest-risk infections when resources are limited • Education about high-risk activities (e.g., motorbike taxis)
Last-minute travelers	• Minimal advance notice for pretravel consultations or immunizations	• Several vaccines are effective with a single dose. • Some vaccination series can be accelerated. • Some malaria chemoprophylaxis can be started 1 day prior to entering risk areas. • Consider broad immunization coverage and standing malaria chemoprophylaxis supply for aircrews or other travelers with unpredictable, last-minute trips.
Long-term travelers	• Increased risk of infection and injury due to longer duration • Increased likelihood of adopting local food, accommodation, and transportation standards • Potential need for extended supply of malaria chemoprophylaxis	• Increased emphasis on importance of certain vaccines, such as hepatitis B, rabies, typhoid, or Japanese encephalitis (in endemic areas) • Long-term malaria chemoprophylaxis
Health care workers on medical missions	• Risk of infections acquired through patient care because of inadequate infection-control standards • Potentially high prevalence of untreated transmissible infections in patients • Limited or no access to urgent postexposure prophylaxis for HIV infection and hepatitis B • Exposure to emerging infections and outbreaks	• Ensure that traveler has received recommended immunizations for health care workers. • Advise traveler to assess availability of adequate personal protective equipment and medications for HIV postexposure prophylaxis and to consider potential need to bring own supplies. • Advise traveler against working with organizations inexperienced in delivering care in the destination area.
Medical tourists	• Nosocomial infections and other complications of medical procedures overseas • Substandard accreditation, infection control, safety guidelines, drugs, and blood-product screening • Increased risk of thromboembolism following surgery	• Advise traveler on potential risks. • Direct tourist to internationally accredited facilities and providers. • Tourist should acquire a copy of medical records for providers who will provide follow-up care. • See CDC *Yellow Book* (Table 130-2) for specific resources.

Abbreviation: CDC, Centers for Disease Control and Prevention.

pretravel advice can include financial and language issues or a lack of trust in the medical system.

POSTTRAVEL MEDICAL CARE

Acute febrile illness in returning travelers can represent a potentially life-threatening illness, such as malaria, typhoid fever, or leptospirosis. Early diagnosis and treatment can be critical and potentially lifesaving for many travel-related infections. Although most acute febrile illnesses have incubation periods of <14 days, infections including typhoid, malaria, leptospirosis, and acute schistosomiasis can have prolonged incubation periods. Travelers should be advised to always inform their health care providers of their travel history, even when their travel does not immediately precede illness onset. Exposure risk occurring as much as a year prior to illness onset should be considered for malaria. Providers unfamiliar with infections common to a region recently visited by an acutely ill traveler should consult with infectious disease or travel medicine specialists and/or their local public health departments.

OUTBREAKS AND EMERGING INFECTIOUS DISEASES

Emerging and reemerging infectious diseases create challenges for international travel. During outbreaks of novel or emerging infections (e.g., the 2020 COVID-19 pandemic, recent Ebola virus disease epidemics in western and central Africa, or the emergence of Zika in the Americas in 2015–2016), information can be limited or can rapidly change. Significant travel disruption can also occur when outbreaks of well-known infections take place in previously unaffected areas, such as the 2017–2018 yellow fever outbreak in southeastern Brazil, which included the metropolitan areas of Rio de Janeiro and São Paolo, two major travel destinations for which vaccination had not previously been recommended. Providers counseling travelers can help them make informed decisions by carefully reviewing available travel health notices, surveillance reports, and potential risks to travelers. Individuals may be advised against travel to certain areas when risks are significant.

FURTHER READING

Angelo KM et al: What proportion of international travellers acquire a travel-related illness? A review of the literature. J Travel Med 24:1, 2017.

Keystone JS et al (eds): *Travel Medicine*, 4th ed. Edinburgh, Elsevier, 2019.

Mace KE: Malaria Surveillance—United States, 2018. MMWR Surveill Summ 71:1, 2022.

Nemhauser J et al (eds): *Yellow Book 2024 Health Information for International Travel*. New York, Oxford University Press, 2023.

Rao AK et al: Use of a modified preexposure prophylaxis vaccination schedule to prevent human rabies: Recommendations of the Advisory Committee on Immunization Practices–United States, 2022. MMWR Morb Mortal Wkly Rep 71:619, 2022.

Riddle MS et al: Guidelines for the prevention and treatment of travelers' diarrhea: A graded expert panel report. J Travel Med 24:S63, 2017.

Section 2 Clinical Syndromes: Community-Acquired Infections

131 Pneumonia

Lionel A. Mandell,
Michael S. Niederman

DEFINITION

Pneumonia is an infection of the pulmonary parenchyma. Despite significant morbidity and mortality, it is often misdiagnosed, mistreated, and underestimated. Pneumonia has usually been classified as community-acquired (CAP), hospital-acquired (HAP), or ventilator-associated (VAP). A fourth category, healthcare-associated pneumonia, should be discontinued because it did not reliably predict infection with resistant pathogens and was associated with increased use of broad-spectrum antibiotics. Rather than relying on a predefined subset of pneumonia cases, it is better to assess patients individually based on risk factors for infection with a resistant organism, such as certain comorbid illnesses, recent hospitalization, or recent antibiotic therapy.

Pneumonia caused by macroaspiration of oropharyngeal or gastric contents, usually referred to as aspiration pneumonia, is best thought of as a point on the continuum that includes CAP and HAP. Estimates suggest that aspiration pneumonia accounts for 5–15% of CAP cases, but reliable figures for HAP are unavailable. The airways or pulmonary parenchyma may be involved, and patients usually represent a clinical phenotype with risk factors for macroaspiration and involvement of characteristic anatomic pulmonary locations.

In this chapter, we will not be dealing with pneumonia in immunocompromised hosts.

PATHOPHYSIOLOGY

Pneumonia is the result of the proliferation of microbial pathogens in the alveoli and the host's response to them. Until recently, it was thought that the lungs were sterile and that pneumonia resulted from the introduction of potential pathogens into this sterile environment. Typically, this introduction occurred through microaspiration of oropharyngeal organisms into the lower respiratory tract. The overcoming of innate and adaptive immunity by such microorganisms could result in the clinical syndrome of pneumonia.

A complex and diverse community of bacteria in the lungs constitutes the lung microbiota. Awareness of this microbiota has prompted a rethinking of how pneumonia develops. Mechanical factors, such as the hairs and turbinates of the nares, the branching tracheobronchial tree, mucociliary clearance, and gag and cough reflexes, play roles in host defense but are insufficient to effectively block bacterial access to the lower airways. In the absence of a sufficient barrier, microorganisms may reach the lower respiratory tract by a variety of pathways, including inhalation, microaspiration, and direct mucosal dispersion.

The constitution of the lung microbiota is determined by three factors: microbial entry into the lungs, microbial elimination, and regional growth conditions for bacteria, such as pH, oxygen tension, and temperature. The key question, however, is how a dynamic homeostasis among bacterial communities results in acute infection. Pneumonia therefore does not appear to be the result of the invasion of a sterile space by a particular microorganism but is more likely an emergent phenomenon dependent upon a number of mechanisms, including self-accelerating positive feedback loops.

A possible model for pneumonia is as follows. An inflammatory event resulting in epithelial and/or endothelial injury results in the release of cytokines, chemokines, and catecholamines, some of which may selectively promote the growth of certain bacteria, such as *Streptococcus pneumoniae* and *Pseudomonas aeruginosa*. This cycle of inflammation, enhanced nutrient availability, and release of potential bacterial growth factors may result in a positive feedback loop that further accelerates inflammation and the growth of particular bacteria, which may then become dominant. In cases of CAP and HAP, the trigger may be a viral infection compounded by microaspiration of oropharyngeal organisms. In cases of true aspiration pneumonia, the trigger may possibly be the macroaspiration event itself.

Once triggered, innate and adaptive immune responses can help contain potential pathogens and prevent the development of pneumonia. However, in the face of continuing inflammation (and especially if a positive feedback loop becomes sustainable), the process may proceed to a full-fledged pneumonia syndrome. Inflammatory mediators such as interleukin 6 and tumor necrosis factor result in fever, and chemokines such as interleukin 8 and granulocyte colony-stimulating factor increase local neutrophil numbers. Mediators released by macrophages and neutrophils may create an alveolar capillary leak, resulting in impaired oxygenation, hypoxemia, and radiographic infiltrates. Bacteria themselves may produce toxins that further amplify the inflammatory response. Moreover, some bacterial pathogens appear to interfere with the hypoxic vasoconstriction that would normally occur with fluid-filled alveoli, possibly resulting in severe hypoxemia. Decreased compliance due to capillary leak, hypoxemia, increased respiratory drive, increased secretions, and occasionally infection-related bronchospasm all lead to worsening dyspnea. If severe enough, changes in lung mechanics secondary to reductions in lung volume, compliance, and intrapulmonary shunting of blood may cause respiratory failure.

PATHOLOGY

Classic pneumonia evolves through a series of stages. The initial stage is edema with a proteinaceous exudate and often bacteria in the alveoli. Next is a rapid transition to the red hepatization phase. Erythrocytes in this intraalveolar exudate give this stage its name. In the third phase, gray hepatization, erythrocytes have been lysed and degraded. The neutrophil is the predominant cell, fibrin deposition is abundant, and bacteria have disappeared. This phase corresponds with the successful containment of the infection and improvement in gas exchange. In the final phase, resolution, the macrophage reappears as the dominant cell in the alveolar space and the debris of neutrophils, bacteria and fibrin has been cleared, as has the inflammatory response.

This pattern has been described best for lobar pneumococcal pneumonia but may not apply to pneumonia of all etiologies. In VAP, respiratory bronchiolitis may precede the development of a radiologically apparent infiltrate. A bronchopneumonia pattern is most common in nosocomial pneumonias, whereas a lobar pattern is more common in bacterial CAP. Despite the radiographic appearance, viral and *Pneumocystis* pneumonias represent alveolar rather than interstitial processes.

COMMUNITY-ACQUIRED PNEUMONIA

ETIOLOGY

Numerous microbes may cause CAP, including a variety of bacteria, viruses, fungi, and protozoa.

Newer viral pathogens include metapneumoviruses, the coronaviruses responsible for severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), and the SARS-CoV-2 coronavirus. First described in December 2019, SARS-CoV-2 and its associated clinical disease, COVID-19, reached pandemic proportions and are a cause of significant morbidity and mortality. The COVID-19 pandemic has changed the etiologic profile of CAP, and the ultimate role that the SARS-CoV-2 virus will play as a cause of CAP remains to be seen. The virus and the disease are discussed in detail in **Chap. 205**.

Although most CAP cases are caused by relatively few pathogens, an accurate determination of their prevalence is difficult because laboratory testing methods are often insensitive and indirect **(Table 131-1)**. Separation of potential agents into "typical" bacterial pathogens and "atypical" organisms may be helpful, although both types of pathogens can lead to similar clinical syndromes. The former group includes *S. pneumoniae*, *Haemophilus influenzae*, and, in selected patients, *Staphylococcus aureus* and gram-negative bacilli such as *Klebsiella pneumoniae* and *P. aeruginosa*. The "atypical" organisms include *Mycoplasma pneumoniae*,

TABLE 131-1 Microbial Causes of Community-Acquired Pneumonia, by Site of Care

	HOSPITALIZED PATIENTS	
OUTPATIENTS	**NON-ICU**	**ICU**
Streptococcus pneumoniae *Mycoplasma pneumoniae* *Haemophilus influenzae* *Chlamydia pneumoniae* Respiratory viruses[a]	*S. pneumoniae* *M. pneumoniae* *C. pneumoniae* *H. influenzae* *Legionella* spp. Respiratory viruses[a]	*S. pneumoniae* *Staphylococcus aureus* *Legionella* spp. Gram-negative bacilli *H. influenzae* Respiratory viruses[a]

[a] Influenza A and B viruses, SARS-CoV-2 and other coronaviruses, human metapneumovirus, adenoviruses, respiratory syncytial viruses, parainfluenza viruses.

Abbreviation: ICU, intensive care unit.

TABLE 131-2 Epidemiologic Factors Suggesting Possible Causes of Community-Acquired Pneumonia

FACTOR	POSSIBLE PATHOGEN(S)
Alcoholism	*Streptococcus pneumoniae*, oral anaerobes, *Klebsiella pneumoniae*, *Acinetobacter* spp., *Mycobacterium tuberculosis*
COPD and/or smoking	*Haemophilus influenzae*, *Pseudomonas aeruginosa*, *Legionella* spp., *S. pneumoniae*, *Moraxella catarrhalis*, *Chlamydia pneumoniae*
Structural lung disease (e.g., bronchiectasis)	*P. aeruginosa*, *Burkholderia cepacia*, *Staphylococcus aureus*
Dementia, stroke, decreased level of consciousness	Oral anaerobes, gram-negative enteric bacteria
Lung abscess	CA-MRSA, oral anaerobes, endemic fungi, *M. tuberculosis*, nontuberculous mycobacteria
Travel to Ohio or St. Lawrence river valley	*Histoplasma capsulatum*
Travel to southwestern United States	Hantavirus, *Coccidioides* spp.
Travel to Southeast Asia	*Burkholderia pseudomallei*, avian influenza virus
Stay in hotel or on cruise ship in previous 2 weeks	*Legionella* spp.
Local influenza activity	Influenza virus, *S. pneumoniae*, *S. aureus*
Exposure to infected humans	SARS-CoV-2
Exposure to birds	*H. capsulatum*, *Chlamydia psittaci*
Exposure to rabbits	*Francisella tularensis*
Exposure to sheep, goats, parturient cats	*Coxiella burnetii*

Abbreviations: CA-MRSA, community-acquired methicillin-resistant *Staphylococcus aureus*; COPD, chronic obstructive pulmonary disease; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Chlamydia pneumoniae, and *Legionella* species as well as respiratory viruses such as influenza, adenoviruses, human metapneumoviruses, respiratory syncytial virus, and coronaviruses. With the increasing use of pneumococcal vaccine, the incidence of pneumococcal pneumonia is decreasing. *M. pneumoniae* plays more of a role in ambulatory cases, whereas *Legionella* tends to be associated with more serious cases and can be found in outbreaks, as well. *C. pneumoniae* now appears to account for <1% of CAP cases. Viruses are recognized as increasingly important in pneumonia, and polymerase chain reaction (PCR)–based testing indicates their presence in the respiratory tract of 20–30% of healthy adults and in the same percentage of pneumonia patients, including those who are severely ill. The most common are influenza, parainfluenza, and respiratory syncytial viruses. Whether they are true etiologic pathogens, co-pathogens, or simply colonizers cannot always be determined. Atypical organisms cannot be cultured on standard media or seen on Gram stain, but their frequency and importance have significant implications for therapy. They are intrinsically resistant to all β-lactam antibiotics and require treatment with a macrolide, fluoroquinolone, or a tetracycline. In the 10–15% of CAP cases that are polymicrobial, the etiology usually includes a combination of typical and atypical pathogens.

Earlier literature suggested that aspiration pneumonia was caused primarily by anaerobes, with or without aerobic pathogens. A shift, however, has been noted recently: if aspiration pneumonia is acquired in a community or hospital setting, the likely pathogens are those usually associated with CAP or HAP. Anaerobes may still play a role, especially in patients with poor dentition, lung abscess, necrotizing pneumonia, or empyema.

S. aureus pneumonia is known to complicate influenza virus infection. However, methicillin-resistant *S. aureus* (MRSA) has been reported as a primary etiologic agent of CAP. Although cases caused by MRSA are relatively uncommon, clinicians must be aware of its potentially serious consequences, such as necrotizing pneumonia. Two factors have led to this problem: the spread of MRSA from the hospital setting to the community and the emergence of genetically distinct MRSA strains in the community associated with bacterial toxin production. Community-associated MRSA (CA-MRSA) strains may infect healthy individuals who have had no association with health care.

Despite a careful history, physical examination, and radiographic studies, the causative pathogen is often difficult to predict with certainty, and in more than half of cases, a specific etiology is not determined. Nevertheless, epidemiologic and risk factors may suggest certain pathogens **(Table 131-2)**.

EPIDEMIOLOGY

It is estimated that 7 million or more CAP cases occur annually in the United States. The annual incidence in adults ranges from 16 to 23 per 1000 population. The incidence of hospitalization is 650/100,000 but rises dramatically to 2000/100,000 yearly in the elderly. Overall, approximately 30% of patients are hospitalized, resulting in 1.5 million admissions. Along with influenza, CAP is the eighth leading cause of death in the United States, resulting in >60,000 deaths annually. The mortality rate among outpatients is usually <5% but ranges from ~12 to 40% among hospitalized patients. The exact rate depends on whether treatment takes place in or outside the intensive care unit (ICU). In the United States, CAP is the leading cause of death from infection among patients >65 years of age. Moreover, 18% of hospitalized CAP patients are readmitted within 1 month of discharge. The overall yearly CAP cost is estimated at $17 billion.

Risk factors for CAP in general and pneumococcal pneumonia in particular have implications for treatment. They include alcoholism, asthma, immunosuppression, institutionalization, and age >70 years. In the elderly, decreased cough and gag reflexes and reduced antibody and Toll-like receptor responses increase the likelihood of pneumonia. Risk factors for pneumococcal pneumonia include dementia, seizure disorders, heart failure, cerebrovascular disease, alcoholism, tobacco smoking, chronic obstructive pulmonary disease (COPD), and HIV infection. CA-MRSA pneumonia is more likely in patients with skin colonization or infection with CA-MRSA at other sites, and after viral infection. Enterobacteriaceae tend to infect patients recently hospitalized, given antibiotics, or who have comorbidities such as alcoholism, heart failure, or renal failure. *P. aeruginosa* is a particular problem in patients with severe structural lung disease (e.g., bronchiectasis, cystic fibrosis, or severe COPD). Risk factors for *Legionella* infection include diabetes, hematologic malignancy, cancer, severe renal disease, HIV infection, smoking, male gender, and a recent hotel stay or cruise ship trip.

CLINICAL MANIFESTATIONS

The clinical presentation of pneumonia can vary from indolent to fulminant and from mild to fatal. Manifestations of worsening severity include both constitutional findings and those limited to the lung and associated structures. The patient is frequently febrile and/or tachycardic and may experience chills and/or sweats. Cough may be nonproductive or productive of mucoid, purulent, or blood-tinged sputum. Gross hemoptysis is suggestive of necrotizing pneumonia (e.g., that due to CA-MRSA). Depending on severity, shortness of breath may be present, and pleural involvement may result in chest pain. Up to 20% of patients may have gastrointestinal symptoms such as nausea, vomiting, or diarrhea. Other symptoms may include fatigue, headache, myalgias, and arthralgias.

Findings on physical examination vary with the degree of pulmonary consolidation and the presence or absence of a significant pleural effusion. An increased respiratory rate and use of accessory muscles of respiration are common. Palpation may reveal increased or decreased tactile fremitus, and the percussion note can vary from dull to flat, reflecting underlying consolidated lung and pleural fluid, respectively. Crackles, bronchial breath sounds, and possibly a pleural friction rub may be heard. The clinical presentation may be less obvious in the elderly, who may initially display new-onset or worsening confusion, or worsening of a chronic illness, but few other manifestations. Severely ill patients may have septic shock and organ failure. In cases of CAP, symptoms can range from almost nonexistent to severe, and in those with aspiration pneumonia, chest radiographic findings are often in gravity-dependent parts of the lung.

■ DIAGNOSIS

When confronted with possible CAP, the physician must ask two questions: Is this pneumonia, and, if so, what is the likely pathogen? The former question is answered by clinical and radiographic methods, whereas the latter requires laboratory techniques.

Clinical Diagnosis The differential diagnosis includes infectious and noninfectious entities, including acute bronchitis, exacerbations of chronic bronchitis, heart failure, and pulmonary embolism. The importance of a careful history cannot be overemphasized. The diagnosis of CAP requires a compatible history, such as cough, sputum production, fever and dyspnea, and a new infiltrate on chest radiography.

Unfortunately, the sensitivity and specificity of physical examination findings are only 58% and 67%, respectively. Chest radiography is often unable to differentiate CAP from other conditions. Radiographic findings may suggest increased severity (e.g., cavitation or multilobar involvement) and occasionally suggest an etiologic diagnosis, such as pneumatoceles in *S. aureus* infection or an upper-lobe cavitating lesion in tuberculosis. Computed tomography (CT) may be of value in suspected loculated effusion or cavitary cases or in postobstructive pneumonia caused by a tumor or foreign body. For outpatients, clinical and radiologic assessments are usually all that is required before treatment is started since most laboratory results are not available soon enough to influence initial management. In certain cases, the availability of rapid point-of-care outpatient tests can be important, such as for rapid diagnosis of influenza infection, and can prompt specific anti-influenza treatment and secondary prevention measures.

Etiologic Diagnosis The etiology of pneumonia usually cannot be determined solely on the basis of clinical or radiographic presentation. Data from >17,000 emergency department CAP cases showed an etiologic determination in only 7.6%. Except for CAP patients admitted to the ICU, no data exist showing that treatment directed at a specific pathogen is statistically superior to empirical therapy. The benefit of establishing a microbial etiology may be questioned, particularly in light of the cost of diagnostic testing. However, a number of reasons exist for attempting an etiologic diagnosis. Identification of a specific or unexpected pathogen allows focusing of the initial empirical regimen, with a consequent decrease in antibiotic selection pressure and the risk of resistance. Pathogens with important public safety implications, such as *Mycobacterium tuberculosis,* influenza, and SARS-CoV-2 viruses, may be found. Finally, without susceptibility data, trends in resistance cannot be followed accurately, and appropriate empirical therapeutic regimens are harder to devise.

GRAM STAIN AND CULTURE OF SPUTUM The main purpose of the sputum Gram stain is to ensure suitability of a specimen for culture. To be suitable, a sputum sample must have >25 neutrophils and <10 squamous epithelial cells per low-power field. However, staining may also identify certain pathogens (e.g., *S. pneumoniae*, *S. aureus*, and gram-negative bacteria). The sensitivity and specificity of the sputum Gram stain and culture are highly variable. Even in cases of proven bacteremic pneumococcal pneumonia, the positive sputum cultures are ≤50%.

Many patients, particularly elderly individuals, may be unable to produce an appropriate sputum sample. Others may be taking antibiotics that interfere with culture results. Inability to produce sputum can be caused by dehydration, whose correction may result in increased sputum production and a more obvious infiltrate on chest radiography. For patients admitted to the ICU and intubated, a deep-suction aspirate or bronchoalveolar lavage sample has a high yield on culture when sent to the laboratory as soon as possible. Since pathogens in severe and mild CAP may differ (Table 131-1), the greatest benefit of staining and culturing respiratory secretions is to alert the physician to unexpected and/or resistant pathogens and to permit appropriate modification of therapy. Other stains and cultures (e.g., for *M. tuberculosis* or fungi) may be useful as well. The sputum Gram stain and culture are recommended only for hospitalized CAP patients, particularly those with severe cases or those with risks of MRSA or *P. aeruginosa* infection.

BLOOD CULTURES The yield from blood cultures, even when samples are collected before antibiotic therapy, is disappointingly low. Only 5–14% of cultures from hospitalized CAP patients are positive, and the most common pathogen is *S. pneumoniae* followed by *S. aureus* and *P. aeruginosa.* Since recommended empirical regimens all provide pneumococcal coverage, a blood culture positive for this pathogen has little, if any, effect on clinical outcome. However, susceptibility data may allow narrowing of antibiotic therapy in appropriate cases, and such data help to track microbial resistance patterns on a national basis. Because of the low yield and lack of significant impact on outcome, blood cultures are not considered *de rigueur* for all hospitalized CAP patients. Certain high-risk patients should have blood cultured, including those with neutropenia secondary to pneumonia, asplenia, complement deficiencies, chronic liver disease, or severe CAP and those at risk of MRSA or *P. aeruginosa* infection.

URINARY ANTIGEN TESTS Two commercially available tests detect pneumococcal and *Legionella* urinary antigens. The *Legionella pneumophila* test detects only serogroup 1, which accounts for most community-acquired cases of Legionnaires' disease in the United States. Its sensitivity and specificity are 70% and 99%, respectively. The pneumococcal urine antigen test also is quite sensitive and specific (70% and >90%, respectively). Although false-positive results can be obtained for pneumococcus-colonized children, the test is generally reliable. Both tests can detect antigen even after initiation of appropriate antibiotic therapy. Testing for urinary pneumococcal antigen can be reserved for severe cases; *Legionella* antigen can be sought in severe cases and when relevant epidemiologic factors are present.

POLYMERASE CHAIN REACTION PCR tests amplify a microorganism's DNA or RNA, and multiplex PCR panels test for a number of viral and bacterial pathogens. These tests dramatically improve response times, compared to standard cultures, but the contamination of respiratory specimens by upper-airway flora may make semiquantitative or quantitative assays necessary for best results. PCR of nasopharyngeal swabs has become the standard for diagnosis of respiratory viral infection, including influenza and coronaviruses. PCR can also detect the nucleic acid of *Legionella* species, *M. pneumoniae*, *C. pneumoniae*, and mycobacteria. The cost-effectiveness of PCR testing, however, has not been definitively established.

SEROLOGY A fourfold rise in specific IgM antibody titer between acute- and convalescent-phase serum samples is generally considered diagnostic of infection with a particular pathogen. Until recently, serologic tests were used to help identify atypical pathogens as well as selected unusual organisms such as *Coxiella burnetii*. However, these tests have fallen out of favor because of the delays in obtaining convalescent phase results and difficulties with interpretation.

BIOMARKERS Two of the most commonly used markers are C-reactive protein (CRP) and procalcitonin (PCT). Levels of these acute-phase reactants increase in the presence of an inflammatory response, particularly to bacterial pathogens. Nevertheless, PCT is insufficiently accurate for use in the diagnosis of bacterial CAP, and initial serum

PCT levels should not be used as a basis for withholding initial antibiotic treatment. CRP is considered even less sensitive than PCT for detecting bacterial pathogens. These tests should not be used alone but, in conjunction with findings from the history, physical examination, radiography, and laboratory tests, may facilitate antibiotic stewardship and appropriate management of seriously ill CAP patients.

TREATMENT

Community-Acquired Pneumonia

SITE OF CARE

The decision to hospitalize a patient with CAP has considerable implications. The cost of inpatient management exceeds outpatient treatment by a factor of 20, and hospitalization accounts for most CAP-related expenditures. However, late admission to the ICU is associated with increased mortality rates. The choice can be difficult: some patients can be managed at home, while others require hospitalization. Tools that objectively assess the risk of adverse outcomes, including severe illness and death, can help to minimize unnecessary hospitalizations. The two most frequently used are the Pneumonia Severity Index (PSI), a prognostic model that identifies patients at low risk of dying, and the CURB-65 criteria, which yield a severity-of-illness score.

To determine the PSI, points are given for 20 variables, including age, coexisting illness, and abnormal physical and laboratory findings. Based on the score, patients are assigned to one of five classes with these mortality rates: class 1, 0.1%; class 2, 0.6%; class 3, 2.8%; class 4, 8.2%; and class 5, 29.2%. Using the PSI results in lower admission rates for class 1 and 2 patients. Class 3 patients could ideally be admitted to an observation unit pending further decisions.

The CURB-65 criteria include five variables: confusion (C); urea >7 mmol/L (U); respiratory rate ≥30/min (R); blood pressure—systolic ≤90 mmHg or diastolic ≤60 mmHg (B); and age ≥65 years. Patients with a score of 0 (a 30-day mortality rate of 1.5%) can be treated as outpatients. With a score of 1 or 2, the patient should be hospitalized unless the score is entirely or in part attributable to an age of ≥65 years; in such cases, hospitalization may not be necessary. Among patients with scores of ≥3, mortality rates are 22% overall; these patients may require ICU admission. The PSI score has greater efficacy and has been more robustly validated than the CURB-65 criteria but is more difficult to calculate.

In general, if a patient is unable to maintain oral intake, if compliance may be an issue when assessed on the basis of mental condition or living situation (e.g., cognitive impairment or homelessness), or if the patient's O_2 saturation on room air is <92%, hospitalization is necessary. If these considerations do not apply, clinical judgment in conjunction with a prediction rule should be used to determine the site of care.

Neither PSI nor CURB-65 is accurate in determining the need for ICU admission. Patients with septic shock requiring vasopressors or with acute respiratory failure requiring intubation and mechanical ventilation should be admitted directly to an ICU **(Table 131-3)**, and those with three of the nine minor criteria listed in the latter table should be admitted to an ICU or a high-level monitoring unit. Mortality rates were higher among less ill patients who were admitted to a medical floor but then deteriorated than among equally ill patients initially monitored in the ICU.

TABLE 131-3 Criteria for Severe Community-Acquired Pneumonia

Minor criteria
Respiratory rate ≥30 breaths/min
PaO_2/FiO_2 ratio ≤250
Multilobar infiltrates
Confusion/disorientation
Uremia (BUN level ≥20 mg/dL)
Leukopenia (WBC count <4000 cells/μL)
Thrombocytopenia (platelet count <100,000 cells/μL)
Hypothermia (core temperature <36°C)
Hypotension requiring aggressive fluid resuscitation
Major criteria
Respiratory failure requiring invasive mechanical ventilation
Septic shock requiring vasopressors

Abbreviations: BUN, blood urea nitrogen; PaO_2/FiO_2, arterial oxygen pressure/fraction of inspired oxygen; WBC, white blood cell.

ANTIBIOTIC RESISTANCE

Antimicrobial resistance is a significant problem that threatens to diminish our therapeutic armamentarium. Antibiotic misuse results in increased antibiotic selection pressure that can affect resistance locally and globally by clonal dissemination. For CAP, the main resistance issues currently involve *S. pneumoniae* and CA-MRSA.

S. pneumoniae In general, pneumococcal resistance to β-lactams is acquired by (1) direct DNA incorporation and remodeling of penicillin-binding proteins through contact with closely related oral commensal bacteria (e.g., viridans group streptococci), (2) the process of natural transformation, or (3) mutation of certain genes.

Susceptibility to penicillins depends upon treatment with intravenous (IV) or oral agents.

For IV: Susceptible: minimum inhibitory concentration (MIC) ≤2 μg/mL

Intermediate: MIC >2 and ≤4 μg/mL
Resistant: MIC ≥4 μg/mL

For oral: Susceptible: MIC ≤0.06 μg/mL

Intermediate: MIC >0.06 and ≤1 μg/mL
Resistant: MIC >1 μg/mL

For non–central nervous system pneumococcal infections, decreased susceptibility to penicillin is mitigated by usual doses. In the United States, only 0.6% of pneumococci are resistant to ceftriaxone and cefotaxime. Risk factors for penicillin-resistant pneumococcal infection include recent antimicrobial therapy, age of <2 or >65 years, attendance at a day-care center, recent hospitalization, and HIV infection.

In contrast to penicillin resistance, macrolide resistance is increasing in *S. pneumoniae* through several mechanisms. *Target-site modification* caused by ribosomal methylation in 23S rRNA encoded by the *ermB* gene results in high-level resistance (MIC, ≥64 μg/mL) to macrolides, lincosamides, and streptogramin B–type antibiotics. The *efflux mechanism* encoded by the *mef* gene (*M phenotype*) is usually associated with low-level resistance (MIC, usually <16 μg/mL). These two mechanisms account for ~40% and ~60%, respectively, of resistant pneumococcal isolates in the United States. High-level resistance to macrolides is more common in Europe, whereas lower-level resistance predominates in North America. The prevalence of macrolide-resistant *S. pneumoniae* exceeds 25% in some countries; in Canada, it is ~22%, and in the United States approximately 40%. Much of this resistance is high-level, and treatment failures may result. In these situations, a macrolide should not be used as empirical monotherapy. In the United States and Canada, 87.5% of pneumococci are susceptible to doxycycline.

The rate of pneumococcal resistance to fluoroquinolones (e.g., levofloxacin, moxifloxacin, and gemifloxacin) is usually <2%. Changes can occur in one or both target sites (topoisomerases II and IV) and are attributable to mutations in the *gyrA* and *parC* genes, respectively. An efflux pump may also play a role in pneumococcal resistance to fluoroquinolones.

Isolates resistant to drugs from three or more antimicrobial classes with different mechanisms of action are considered multidrug-resistant (MDR) strains. The propensity for an association of pneumococcal resistance to penicillin with reduced susceptibility to other drugs, e.g. macrolides, tetracyclines, and trimethoprim-sulfamethoxazole, is of concern. In the United States, 58.9% of penicillin-resistant pneumococcal blood isolates are also resistant to macrolides.

The most important risk factor for antibiotic-resistant pneumococcal infection is use of a specific antibiotic within the previous 3 months. A history of prior antibiotic treatment is a critical factor in avoiding the use of an inappropriate antibiotic.

CA-MRSA CAP due to MRSA may be caused by the classic hospital-acquired strains or by genotypically and phenotypically distinct community-acquired strains. Most infections with the former are acquired either directly or indirectly during contact with the health care environment. However, in some hospitals, CA-MRSA strains are displacing the classic hospital-acquired strains, suggesting that the newer community-acquired strains may be more robust.

Methicillin resistance in *S. aureus* is determined by the *mecA* gene, which encodes for resistance to all β-lactam drugs At least five staphylococcal chromosomal cassette *mec* (SCC*mec*) types have been described. The typical hospital-acquired strain usually has a type II or III SCC*mec* element, whereas CA-MRSA has type IV. CA-MRSA isolates tend to be less resistant than the older hospital-acquired strains and are often susceptible to trimethoprim-sulfamethoxazole, clindamycin, and tetracycline in addition to vancomycin and linezolid. CA-MRSA strains also carry genes for superantigens such as enterotoxins B and C and Panton-Valentine leukocidin; the latter is a membrane-tropic toxin that can create cytolytic pores in neutrophils, monocytes, and macrophages. Risk factors for MRSA include colonization or prior infection and MRSA as suggested by gram-positive cocci in clusters on sputum Gram stain. Other factors that may raise suspicion of MRSA infection include recent antibiotics, hospitalization, influenza, cavitary or necrotizing pneumonia, or empyema.

M. pneumoniae Macrolide-resistant *M. pneumoniae* has been reported in a number of countries, including Germany (3%), Japan (30%), China (95%), and France and the United States (5–13%). *Mycoplasma* resistance to macrolides is increasing as a result of binding-site mutation in domain V of 23S rRNA.

Gram-Negative Bacilli A detailed discussion of resistance among gram-negative bacilli is beyond the scope of this chapter **(see Chap. 166)**. Fluoroquinolone resistance among community isolates of *Escherichia coli* is increasing. *Enterobacter* species are typically resistant to cephalosporins, and the drugs of choice to treat these organisms are usually fluoroquinolones or carbapenems. Similarly, when infections due to bacteria producing extended-spectrum β-lactamases (ESBLs) are documented or suspected, a carbapenem should be considered.

INITIAL ANTIBIOTIC MANAGEMENT

Since the etiology of CAP is rarely known at the outset of treatment, initial therapy is usually empirical and designed to cover the likeliest pathogens. In all cases, treatment should be initiated as expeditiously as possible. CAP treatment guidelines in the United States from the American Thoracic Society (ATS) and the Infectious Diseases Society of America (IDSA) consider the likely pathogens, risk of antimicrobial resistance, severity of illness, site of care, and risk of infection with specific bacteria such as MRSA and *P. aeruginosa* **(Fig. 131-1, Tables 131-4 and 131-5)**. In the figure and the tables, the antibiotics are not listed in order of preference.

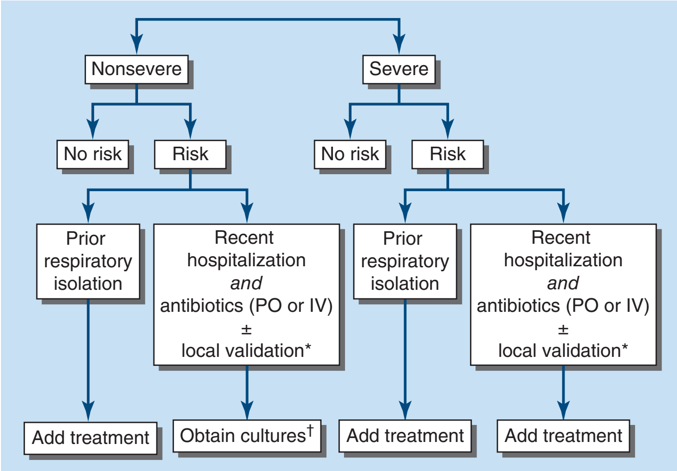
Nonsevere
Severe
No risk
Risk
No risk
Risk
Prior respiratory isolation
Recent hospitalization *and* antibiotics (PO or IV) ± local validation*
Prior respiratory isolation
Recent hospitalization *and* antibiotics (PO or IV) ± local validation*
Add treatment
Obtain cultures†
Add treatment
Add treatment

FIGURE 131-1 Algorithm for assessment of inpatient risk of infection with methicillin-resistant *Staphylococcus aureus* (MRSA) or *Pseudomonas aeruginosa*. Underlying lung disease (e.g., bronchiectasis or very severe chronic obstructive pulmonary disease) are also risks for *P. aeruginosa* infection. *Local validation consists of information on local prevalence, resistance, and risk factors. †Can also use MRSA rapid nasal polymerase chain reaction (PCR) if available. IV, intravenous; PO, oral.

The approach to treatment of aspiration pneumonia is based on a number of factors, including site of acquisition (community vs hospital), normal or abnormal chest radiograph, and additional variables such as illness severity, state of dentition, and risk of infection with an MDR pathogen. Routine coverage of anaerobes is unnecessary unless dentition is poor or there is a lung abscess or necrotizing pneumonia.

Our approach to CAP treatment (Tables 131-4 and 131-5) is very similar to the CAP guidelines with the exceptions listed below.

Outpatients The exceptions to the CAP guidelines that we follow in treating patients are as follows:

- We usually initiate coverage that includes atypical organisms as well as *S. pneumoniae.*
- Generally, we do not consider the risk of infection with *P. aeruginosa* or MRSA particularly significant in outpatients.
- Prior antibiotic use should include both oral and parenteral agents.

Patients are stratified into two groups: those without comorbidity or risk factors for antibiotic resistance and those with comorbidities (e.g., chronic heart, lung, liver, or kidney disease; diabetes; alcoholism; malignancy; or asplenia) with or without risk factors for resistance (Table 131-4). As a general rule, if patients have been treated with a drug from a particular class of antibiotics within the previous 3 months, drugs from a different class should be used to minimize resistance issues.

For those without comorbidity or resistance risk factors, amoxicillin alone or doxycycline is recommended in the guidelines. Monotherapy with amoxicillin is based on evidence of its efficacy in the treatment of hospitalized CAP patients. This recommendation is a change from the 2007 IDSA/ATS CAP guidelines. As a rule, however, we usually tend to initiate treatment that includes coverage for S. *pneumoniae* as well as the atypical pathogens (Table 131-4).

TABLE 131-4 Initial Treatment Strategies for Outpatients with Community-Acquired Pneumonia

STATUS	STANDARD REGIMEN
No comorbidities or risk factors for antibiotic resistance[a]	Combination therapy with amoxicillin (1 g tid) + either a macrolide[b] or doxycycline (100 mg bid) *or*
	Monotherapy with doxycycline (100 mg bid) *or*
	Monotherapy with a macrolide[b,c]
With comorbidities[d] ± risk factors for antibiotic resistance[a]	Combination therapy with
	amoxicillin/clavulanate[e] or a cephalosporin[f] + either a macrolide[b] or doxycycline (100 mg bid)
	or
	Monotherapy with a respiratory fluoroquinolone[g]

[a] Antibiotic treatment within the past 3 months or contact with the health care system. [b] Azithromycin (500 mg on day 1, then 250 mg/d for 4 days), clarithromycin (500 mg bid), or clarithromycin ER (1000 mg/d). [c] If local prevalence of pneumococcal resistance is <25%. [d] Including chronic heart, lung, liver, or kidney disease; diabetes mellitus; alcoholism; malignancy; or asplenia. [e] 500/125 mg tid or 875/125 mg bid. [f] Cefpodoxime (200 mg bid) or cefuroxime (500 mg bid). [g] Levofloxacin (750 mg/d), moxifloxacin (400 mg/d), or gemifloxacin (320 mg/d).

TABLE 131-5 Initial Treatment for Inpatients with or without Risk Factors for Infection with MRSA or *Pseudomonas aeruginosa*

DISEASE SEVERITY, RISK STATUS	REGIMEN
Nonsevere	
No risk factors	A β-lactam[a] + a macrolide[b] *or* A respiratory fluoroquinolone[c]
Prior respiratory isolation	Add coverage for MRSA[d] or *Pseudomonas aeruginosa*[e]
Recent hospitalization, antibiotic treatment, ± LV[f]	Add coverage for MRSA[d] or *P. aeruginosa*[e] only if cultures are positive
Severe	
No risk factors	A β-lactam[a] + a macrolide[b] *or* A β-lactam[a] + respiratory fluoroquinolone[c]
Prior respiratory isolation	Add coverage for MRSA[d] or *P. aeruginosa*[e]
Recent hospitalization, antibiotic treatment ± LV[f]	Add coverage for MRSA[d] or *P. aeruginosa*[e]

[a]Ampicillin-sulbactam (1.5–3 g q6h), ceftriaxone (1–2 g/d), cefotaxime (1–2 g q8h), ceftaroline (600 mg q12h), or ertapenem (1 g/d). [b]Azithromycin (500 mg/d) or clarithromycin (500 mg bid). [c]Levofloxacin (750 mg/d), moxifloxacin (400 mg/d), or gemifloxacin (320 mg/d). [d]Vancomycin (15 mg/kg q12h, with adjustment based on serum levels) or linezolid (600 mg q12h). [e]Piperacillin-tazobactam (4.5 g q6h), cefepime (2 g q8h), ceftazidime (2 g q8h), imipenem (500 mg q6h), meropenem (1 g q8h), or aztreonam (2 g q8h). [f]Obtain cultures. MRSA rapid nasal polymerase chain reaction can also be used if available.

Abbreviations: LV, local validation (local prevalence, resistance, risk factors); MRSA, methicillin-resistant *Staphylococcus aureus*.

Monotherapy with a macrolide is recommended in the guidelines only if there are contraindications to amoxicillin or doxycycline and there is documented low risk of macrolide resistance (<25%). Otherwise, outpatient treatment is quite similar to the regimens recommended in the 2007 IDSA/ATS guidelines. Two relatively newer agents, lefamulin (a pleuromutilin) and omadacycline (a tetracycline) are possible options for CAP patients unable to take β-lactams and/or wanting to avoid the fluoroquinolones. They are available in the United States but not in Canada. Treatment of influenza **(Chap. 206)** and COVID-19 **(Chap. 205)** is discussed in their own chapters.

Inpatients Our exceptions to the recommendations in the CAP guidelines are as follows:

- As a general rule, when initiating treatment for infection with *P. aeruginosa*, we use double coverage.
- The presence of all three risk factors is not required for drug resistance (recent hospitalization, recent oral or IV antibiotic treatment, ± local validation) (Fig. 131-1, Table 131-5).

The main considerations for determining initial empirical treatment of hospitalized CAP patients are clinical severity and risk of infection with drug-resistant pathogens such as MRSA or *P. aeruginosa*. Hospitalization alone is not now considered a significant risk factor for these pathogens. Hospitals should collect local data on MRSA and *P. aeruginosa* regarding prevalence, risk factors for infection, and antibiotic susceptibilities. Patients can be categorized as having nonsevere or severe CAP (Table 131-3), and those in each category may or may not have risk factors for MRSA or *P. aeruginosa* (Fig. 131-1). In scenarios involving these variables in hospitalized CAP patients, empirical treatment for either of these pathogens should be added to standard therapy in those previously colonized or infected with these pathogens, but not in the patient who is considered nonsevere and whose only risk factors are recent hospitalization and antibiotic treatment ± local validation data (Fig. 131-1). In this setting, if we begin treatment, we try to de-escalate if appropriate. In most patients, cultures should be performed but treatment usually withheld unless culture results or rapid nasal PCR results for MRSA are positive.

Nonsevere, No Risk Factors For patients with nonsevere infection and no risk factors, treatment should consist of either a combination of a β-lactam and a macrolide or monotherapy with a respiratory fluoroquinolone (Table 131-5). In the event of contraindications to macrolides and fluoroquinolones, a β-lactam plus doxycycline may be used. Treatment with a β-lactam plus macrolide combination or a fluoroquinolone alone results in lower mortality than monotherapy with a β-lactam.

Severe, No Risk Factors Patients with severe infection but no risk factors should receive combination therapy with either a β-lactam plus macrolide or a β-lactam and a respiratory fluoroquinolone (Table 131-5). Observational studies suggest that combination therapy with a β-lactam plus macrolide may be preferable to a β-lactam plus fluoroquinolone.

Nonsevere and Severe, with Risk Factors To date, there are no prediction rules reliably identifying patients who should be started empirically on treatment for MRSA or *P. aeruginosa*. Current risk factors for infection with these pathogens are hierarchical. Prior isolation of these organisms, especially from the respiratory tract within the previous year, is a more robust risk factor than recent hospitalization and exposure to parenteral antibiotics. For *P. aeruginosa*, underlying lung disease (e.g., bronchiectasis or very severe COPD) also is an important risk factor. If MRSA or *P. aeruginosa* has been isolated previously, appropriate empirical therapy should be started in both severe and nonsevere cases (Table 131-5). We prefer linezolid over vancomycin as first-line treatment for MRSA because of its inhibition of bacterial exotoxin and better lung penetration. If the organism is not isolated from respiratory secretions or blood and/or the nasal or bronchoalveolar lavage PCR test for MRSA is negative and the patient is improving at 48 h, treatment may be de-escalated to a standard regimen.

If, however, the risk factors are recent hospitalization and antibiotic use within the previous 3 months, appropriate samples should be obtained for culture, and, in severe cases only, extended-spectrum treatment for MRSA or *P. aeruginosa* should be initiated. Depending on the severity of infection, local data on *P. aeruginosa* resistance, and antibiotic use within the previous 90 days, single- or double-drug coverage should be used such as antipseudomonal β-lactam plus ciprofloxacin, levofloxacin, or aminoglycoside.

If two antipseudomonal agents are started, the drugs should be from different classes. Whenever possible, assessment for possible de-escalation of therapy is urged. If the patient's illness is not severe, empirical extended treatment should be withheld until culture results are available.

Regardless of the site of care, CAP patients with proven influenza should be given anti-influenza treatment (e.g., oseltamivir) as well as appropriate antibacterial therapy. Physicians should be vigilant about possible superinfection with MRSA. If a viral pathogen such as influenza or SARS-CoV-2 is found and no bacterial pathogen is obvious, antibacterial treatment can be discontinued. However, in those with severe illness, the possibility of bacterial-viral coinfection should be considered.

Although hospitalized patients have traditionally received initial therapy by the IV route, some drugs, particularly the fluoroquinolones, are very well absorbed and may be given orally from the outset to select patients. For those initially treated with IV agents, a switch to oral treatment is appropriate when the patient can ingest and absorb the drugs, is hemodynamically stable, and is showing clinical improvement. A 5-day course of treatment is usually sufficient for uncomplicated CAP, but longer treatment may be required for patients who have not stabilized clinically and for those with bacteremia, metastatic infection, or infection with a more virulent pathogen such as *P. aeruginosa* or MRSA. There are some data to suggest that in select patients who are doing well and are clinically stable, treatment may be discontinued after 3 days.

ADJUNCTIVE MEASURES

In addition to appropriate antimicrobial therapy, certain adjunctive measures should be used. Adequate hydration, oxygen therapy for hypoxemia, vasopressor treatment, and assisted ventilation when necessary are critical to successful treatment. Glucocorticoids may be beneficial in cases of severe CAP requiring invasive or noninvasive mechanical ventilation or with shock. Recent data show a mortality benefit for corticosteroid therapy in those with severe CAP (ventilated and nonventilated), especially if there is a high level of systemic inflammation (CRP >15 mg/dL); therapy is usually continued for 8–14 days and given either by intermittent or continuous infusion. Data support the use of dexamethasone plus a Janus kinase inhibitor or an interleukin 6 inhibitor in COVID-19 patients with rapidly increasing oxygen needs and systemic inflammation.

FAILURE TO IMPROVE

Patients slow to respond to therapy should be reevaluated at about day 3 (or sooner if their condition is worsening), with several scenarios considered. A number of noninfectious conditions mimic pneumonia, including pulmonary edema, pulmonary embolism, lung carcinoma, radiation and hypersensitivity pneumonitis, and connective tissue disease involving the lungs. If the patient truly has CAP and empirical treatment is aimed at the likely expected pathogens, lack of response may be explained in a number of ways. The pathogen may be resistant to the drug, or a sequestered focus (e.g., lung abscess or empyema) may prevent antibiotic access to the pathogen. The patient may be getting the wrong drug or the correct drug at the wrong dose or frequency of administration. Another possibility is that CAP has been diagnosed correctly but an unexpected pathogen (e.g., CA-MRSA, *M. tuberculosis*, or a fungus) is the cause. Nosocomial superinfections—both pulmonary and extrapulmonary—are other possible explanations for a hospitalized patient's failure to improve. In all cases of delayed response or worsening condition, the patient must be carefully reassessed and appropriate studies initiated, possibly including CT or bronchoscopy.

COMPLICATIONS

Complications of severe CAP include respiratory failure, shock and multiorgan failure, and exacerbation of comorbid illnesses. Three particularly noteworthy conditions are metastatic infection, lung abscess, and complicated pleural effusion. Metastatic infection (e.g., brain abscess or endocarditis) is unusual and requires a high degree of suspicion and a detailed workup for proper treatment. Lung abscess may occur in association with aspiration pneumonia or with infection caused by pathogens such as CA-MRSA, *P. aeruginosa*, or (rarely) *S. pneumoniae.* A significant pleural effusion should be tapped for diagnostic and therapeutic purposes. If the fluid has a pH <7.2, a glucose level of <2.2 mmol/L, and a lactate dehydrogenase concentration of >1000 U/L or if bacteria are seen or cultured, drainage is needed.

Cardiovascular events with pneumonia, particularly in the elderly and usually in association with pneumococcal pneumonia and influenza, are increasingly recognized. These events, which may be acute or whose occurrence may extend to at least 1 year, include congestive heart failure, arrhythmia, myocardial infarction, or stroke and may be caused by a variety of mechanisms, including increased myocardial load and/or destabilization of atherosclerotic plaques by inflammation.

FOLLOW-UP

Fever and leukocytosis usually resolve within 2–4 days in otherwise healthy CAP patients, but physical findings may persist longer. Chest radiographic abnormalities are slowest to resolve (4–12 weeks), with the speed of clearance depending on the patient's age and underlying lung disease, and the etiologic pathogen. Patients may be discharged from hospital once their clinical condition, including any comorbidity, is stable. The site of residence after discharge (nursing home, home with family, home alone) is an important consideration, particularly for elderly patients. For a hospitalized patient, we generally recommend a follow-up radiograph ~4–6 weeks later. If relapse or recurrence occurs, particularly in the same lung segment, the possibility of an underlying neoplasm or other local abnormalities (e.g., focal bronchiectasis) must be considered. For individuals managed as outpatients, routine follow-up chest radiography is not necessary if they are nonsmokers, if they are otherwise well, and if symptoms resolved within 5–7 days.

PROGNOSIS

The prognosis depends on the patient's age, comorbidities, and site of treatment (inpatient or outpatient). Young patients without comorbidity do well and usually recover fully after ~2 weeks. Older patients and those with comorbid conditions may take several weeks longer to recover fully. The overall mortality rate for the outpatient group is <5%. For patients requiring hospitalization, overall mortality ranges from 12% to 40%, depending on the patient category and the processes of care, particularly the timely administration of appropriate antibiotics. Recent data, especially in older patients, show that the 1-year mortality following CAP exceeds the 30-day mortality.

PREVENTION

The main preventive measure is vaccination. Recommendations of National Advisory Committees on Immunization Practices should be followed **(Chap. 129)**.

VENTILATOR-ASSOCIATED PNEUMONIA

Research on hospital-acquired pneumonia has focused on VAP (onset ≥48 h after mechanical ventilation). However, the same information and principles can also be applied to ventilated HAP and to non-ICU HAP. Approximately 70% of HAP cases are acquired outside the ICU and 30% in the ICU; the fact that 35% of all HAP patients need mechanical ventilation defines ventilated HAP as a distinct entity. In nonintubated patients with HAP, an expectorated sputum sample is used for microbiologic diagnosis, but results are confounded by frequent colonization by oral pathogens. Microbiologic information in VAP and ventilated HAP is obtained from direct access to deep lower respiratory tract samples, which provide reliable microbiologic data; however, these samples can also contain colonizing pathogens.

ETIOLOGY

Potential etiologic agents of VAP include both MDR and non-MDR bacterial pathogens **(Table 131-6)**. The non-MDR group of "core pathogens" is nearly identical to the pathogens found in severe CAP (Table 131-1); it is not surprising that such pathogens predominate if VAP develops in the first 5–7 days of the hospital stay. However, if patients have other risk factors (particularly prior antibiotic treatment), MDR pathogens are a consideration, even early in the hospital course. The relative frequency of individual MDR pathogens can vary significantly from hospital to hospital and even between different critical care units within the same institution. Most hospitals have problems with *P. aeruginosa* and MRSA, but other MDR pathogens are often institution-specific. Less commonly, fungal and viral pathogens cause VAP, usually affecting severely immunocompromised patients. Rarely,

TABLE 131-6 Microbiologic Causes of Ventilator-Associated Pneumonia

NON-MDR PATHOGENS (CORE PATHOGENS)	MDR PATHOGENS
Streptococcus pneumoniae	*Pseudomonas aeruginosa*
Other *Streptococcus* spp.	Methicillin-resistant *S. aureus*
Haemophilus influenzae	*Acinetobacter* spp.
Methicillin-sensitive *Staphylococcus aureus*	Antibiotic-resistant Enterobacteriaceae
Antibiotic-sensitive Enterobacteriaceae	ESBL-positive strains
Escherichia coli	Carbapenem-resistant strains
Klebsiella pneumoniae	*Legionella pneumophila*
Proteus spp.	*Burkholderia cepacia*
Enterobacter spp.	*Aspergillus* spp.
Serratia marcescens	

Abbreviations: ESBL, extended-spectrum β-lactamase; MDR, multidrug-resistant.

community-associated viruses cause mini-epidemics, usually when introduced by ill health care workers.

■ EPIDEMIOLOGY

Pneumonia is a common complication among patients requiring mechanical ventilation. Prevalence estimates vary between 6 and 52 cases per 100 patients, depending on the population studied. On any given day in the ICU, an average of 10% of patients will have pneumonia—VAP in the overwhelming majority of cases. Although in recent years the frequency of this infection was declining as a result of effective prevention strategies, with the advent of COVID-19, there has been an increase in its frequency. The frequency of VAP changes with the duration of mechanical ventilation, with the highest hazard ratio in the first 5 days and a plateau in additional cases (1% per day) after ~2 weeks. However, the cumulative rate among patients who remain ventilated for as long as 30 days is as high as 70%. These rates often do not reflect the recurrence of VAP in the same patient. Once a ventilated patient is transferred to a chronic-care facility or to home, the incidence of pneumonia drops significantly, especially in the absence of other risk factors for pneumonia. However, in chronic ventilator units, purulent tracheobronchitis becomes a significant issue, often interfering with efforts to wean patients off mechanical ventilation **(Chap. 313)**.

Three factors are critical in the pathogenesis of VAP: colonization of the oropharynx with pathogenic microorganisms, aspiration of these organisms from the oropharynx into the lower respiratory tract, and compromise of normal host defense mechanisms. Most risk factors and their corresponding prevention strategies pertain to one of these three factors **(Table 131-7)**.

The most important risk factor is the endotracheal tube, which bypasses the normal mechanical factors preventing aspiration. While the presence of an endotracheal tube may prevent large-volume aspiration, microaspiration is actually exacerbated by secretions pooling above the cuff. The endotracheal tube and the concomitant need for suctioning can damage the tracheal mucosa, thereby facilitating tracheal colonization. In addition, pathogenic bacteria can form a glycocalyx biofilm on the tube's surface that protects them from both antibiotics and host defenses. The bacteria can also be dislodged during suctioning (done preferably with a closed catheter system) and can reinoculate the trachea, or tiny fragments of a glycocalyx can embolize to distal airways, carrying bacteria with them. The ventilator circuit tubing can harbor pathogenic organisms that can wash back to the patient if manipulated too often; thus, circuits are changed only when soiled and with each new patient. Heat moisture exchangers are changed every 5–7 days or if visibly soiled or malfunctioning.

In a high percentage of critically ill patients, the normal oropharyngeal flora is replaced by pathogenic microorganisms. The most important risk factors are antibiotic selection pressure, cross-infection from other infected/colonized patients or contaminated equipment, severe systemic illness, and malnutrition. Of these factors, antibiotic exposure poses the greatest risk by far. Pathogens such as *P. aeruginosa* almost never cause infection in patients without prior exposure to antibiotics. The recent emphasis on hand hygiene has lowered the cross-infection rate.

Almost all intubated patients experience microaspiration and are at least transiently colonized with pathogenic bacteria. However, only around one-third of colonized patients develop VAP. Colony counts increase to high levels, sometimes days before the development of clinical pneumonia; these increases suggest that the final step in VAP development, independent of aspiration and oropharyngeal colonization, is the overwhelming of host defenses by a large bacterial inoculum. Severely ill patients with sepsis and trauma appear to enter a state of immunoparalysis several days after admission to the ICU—a time that corresponds to the greatest risk of developing VAP. The mechanism of this immunosuppression is not clear, although hyperglycemia and frequent transfusions adversely affect the immune response.

TABLE 131-7 Pathogenic Mechanisms and Corresponding Prevention Strategies for Ventilator-Associated Pneumonia

PATHOGENIC MECHANISM	PREVENTION STRATEGY
Oropharyngeal colonization with pathogenic bacteria	
Elimination of normal flora, overgrowth by pathogenic bacteria	Avoidance of prolonged antibiotic courses; consider oral chlorhexidine[a]
Large-volume oropharyngeal aspiration around time of intubation	Short course of prophylactic antibiotics for comatose patients; short course of prophylactic inhaled aminoglycoside antibiotics[b]
Gastroesophageal reflux	Postpyloric enteral feeding with orally placed feeding tube[a]; avoidance of high gastric residuals, prokinetic agents
Bacterial overgrowth of stomach	Avoidance of prophylactic agents that raise gastric pH[a]; selective decontamination of digestive tract with nonabsorbable antibiotics[a]
Cross-infection from other colonized patients	Hand washing, especially with alcohol-based hand rub; intensive infection control education[b]; isolation; proper cleaning of reusable equipment
Large-volume aspiration	Endotracheal intubation; rapid-sequence intubation technique; avoidance of sedation; decompression of small-bowel obstruction
Ventilator circuit humidification	Change ventilator circuits only when soiled and with new patient; drain ventilator circuit condensate away from patient; replace heat moisture exchanger every 5–7 days or if soiled or malfunctioning[a]
Microaspiration around endotracheal tube	
Endotracheal intubation	Noninvasive ventilation[b]
Prolonged duration of ventilation	Daily awakening from sedation,[b] weaning protocols[b]
Abnormal swallowing function	Early percutaneous tracheostomy[b]
Secretions pooled above endotracheal tube	Head of bed elevated[b]; continuous aspiration of subglottic secretions with specialized endotracheal tube[b]; avoidance of reintubation; minimization of sedation and patient transport; prophylactic PEEP[c] of 5–8 cm
Altered lower respiratory host defenses	Tight glycemic control[a]; lowering of hemoglobin transfusion threshold

[a]Strategies with negative randomized trials or conflicting results. [b]Strategies demonstrated to be effective in at least one randomized controlled trial. [c]Positive end-expiratory pressure.

■ CLINICAL MANIFESTATIONS

The clinical manifestations of HAP and VAP are nonspecific: fever, leukocytosis, increased respiratory secretions, and pulmonary consolidation on physical examination, along with a new or changing radiographic infiltrate. The frequency of abnormal chest radiographs before the onset of pneumonia in intubated patients and the limitations of portable radiographic technique make interpretation of radiographs more difficult than in patients who are not intubated. Other clinical features may include tachypnea, tachycardia, worsening oxygenation, and increased minute ventilation. Serial changes in oxygenation may identify pneumonia earlier than other findings and may also be a means to monitor improvement with therapy.

■ DIAGNOSIS

No single set of criteria is reliably diagnostic of pneumonia in a ventilated patient. The inability to accurately identify such patients compromises efforts to prevent and treat VAP and even calls into question estimates of the impact of VAP on mortality rates.

Application of clinical criteria typical for CAP consistently results in overdiagnosis of VAP, largely because of (1) frequent tracheal colonization with pathogenic bacteria in patients with endotracheal tubes, (2) multiple alternative causes of radiographic infiltrates in mechanically ventilated patients, and (3) the high frequency of other sources of fever in critically ill patients. The differential diagnosis of VAP includes atypical pulmonary edema, pulmonary contusion, alveolar hemorrhage, hypersensitivity pneumonitis, acute respiratory distress syndrome, and pulmonary infarction. Findings of fever and/or leukocytosis may have alternative causes, including antibiotic-associated diarrhea, central line–associated infection, sinusitis, urinary tract infection, pancreatitis, and drug fever. Conditions mimicking pneumonia are often documented in patients in whom VAP has been ruled out by accurate diagnostic techniques. Most of these alternative diagnoses do not require antibiotic treatment; require antibiotics different from those used to treat VAP (fungal or viral pneumonia); or require some additional intervention, such as surgical drainage or catheter removal, for optimal management.

This diagnostic dilemma has led to debate and controversy about whether a quantitative-culture approach as a means of eliminating false-positive clinical diagnoses is superior to a clinical approach enhanced by principles learned from quantitative-culture studies. The most recent IDSA/ATS guidelines for HAP/VAP give a weak recommendation for a clinical approach based on semiquantitative cultures, with consideration of the availability of resources, cost, and the availability of expertise. The guidelines acknowledge that the use of a quantitative approach may result in less antibiotic use, which may be critical for antibiotic stewardship in the ICU. Therefore, the approach at each institution—or potentially for each patient—should be individualized and based on local colonization rates, local diagnostic expertise, and recent history of antibiotic therapy.

Quantitative-Culture Approach This method uses quantitative cultures of deep respiratory tract samples to distinguish colonization from true infection. The more distal in the respiratory tree the diagnostic sampling, the more specific are the results and therefore the lower the threshold of growth necessary to diagnose pneumonia and exclude colonization. For example, an endotracheal aspirate yields proximal samples, and the diagnostic threshold is 10^6 cfu/mL. The protected specimen brush method, in contrast, collects distal samples and has a threshold of 10^3 cfu/mL. Conversely, sensitivity declines as more distal secretions are obtained, especially when they are collected blindly (i.e., by a technique other than bronchoscopy). Additional tests that may increase the diagnostic yield include Gram staining, differential cell counts, staining for intracellular organisms, and detection of local protein levels elevated in response to infection.

If the quantitative approach is used, therapy decisions should be linked to culture results (no antibiotics if below the diagnostic threshold), with antibiotics withheld until results are available unless the patient is critically ill. Studies have documented less antibiotic use with this approach than with the clinical approach, but the results are less clear if antibiotic decisions are not directly linked to culture data. One common limitation of the quantitative approach is that the use of a new and effective antibiotic agent in the 24–48 h prior to sampling can lead to false-negative results. With antimicrobial-sensitive microorganisms, a single antibiotic dose can reduce colony counts below the diagnostic threshold. After 3 days, the operating characteristics of the tests improve to the point at which they are equivalent to results obtained when no prior antibiotic therapy has been given. Conversely, colony counts above the diagnostic threshold during antibiotic therapy suggest that the current antibiotics are ineffective. In addition, quantitative cultures may give results below the diagnostic threshold if samples are collected early in the course of infection or if sampling is delayed until after an effective host response has reduced bacterial counts. Ideally, a specimen should be obtained as soon as pneumonia is suspected and before antibiotic therapy is initiated or changed.

Clinical Approach The lack of specificity of a clinical diagnosis of VAP has hampered its utility, but this approach has been improved by the addition of microbiologic and other laboratory data. Tracheal aspirates generally yield at least twice as many potential pathogens as quantitative cultures, but the causative pathogen is almost always present. The absence of bacteria in Gram-stained endotracheal aspirates makes pneumonia an unlikely cause of fever or pulmonary infiltrates. These findings, coupled with a heightened awareness of the alternative diagnoses possible in patients with suspected VAP, can prevent inappropriate antibiotic overtreatment. Furthermore, the absence of an MDR pathogen in tracheal aspirate cultures eliminates the need for MDR coverage, allowing de-escalation of empirical antibiotic therapy. Similarly, with newer and more sensitive molecular diagnostic methods, a suspected MDR pathogen can be eliminated as a therapy target if test results are negative. A clinical approach that focuses on careful antimicrobial use and de-escalation of therapy after culture results become available may have an impact on the avoidance of antimicrobial overuse and the consideration of alternative sites of infection similar to that of a quantitative-culture approach.

TREATMENT

Ventilator-Associated Pneumonia

Many studies have demonstrated higher mortality rates with the delay of initially appropriate empirical antibiotic therapy. The key to appropriate antibiotic management of VAP is an appreciation of the resistance patterns of the most likely pathogens in a given patient and consideration of local microbiology.

ANTIBIOTIC RESISTANCE

Because of a higher risk of infection with MDR pathogens (Table 131-6), VAP is treated with antibiotics different from those used for severe CAP. Antibiotic selection pressure leads to the frequent involvement of MDR pathogens by selecting either for drug-resistant isolates of common pathogens (e.g., MRSA and Enterobacteriaceae producing ESBLs or carbapenemases) or for intrinsically resistant pathogens (e.g., *P. aeruginosa* and *Acinetobacter* species). Frequent use of β-lactam drugs, especially cephalosporins, appears to be the major risk factor for infection with MRSA and ESBL-positive strains.

P. aeruginosa can develop resistance to all routinely used antibiotics, and, even if initially sensitive, *P. aeruginosa* isolates may develop resistance during treatment. Either derepression of resistance genes or selection of resistant clones within the large bacterial inoculum associated with most pneumonias may be the cause. *Acinetobacter* species, *Stenotrophomonas maltophilia*, and *Burkholderia cepacia* are intrinsically resistant to many of the empirical antibiotic regimens employed (see below). VAP caused by these pathogens typically emerges during treatment of other infections, and resistance is always evident at initial diagnosis.

EMPIRICAL THERAPY

Recommended options for empirical therapy are listed in **Table 131-8**. Treatment should be started once diagnostic specimens have been obtained. The major factors in the selection of agents are the presence of risk factors for MDR pathogens and the predicted risk of death (≤15% is considered low risk). Choices among the various options listed depend on local patterns of resistance and—a very important factor—the patient's prior antibiotic exposure. Knowledge of the local hospital's—and even the specific ICU's—antibiogram and the local incidence of specific MDR pathogens (e.g., MRSA) is critical in selecting appropriate empirical therapy.

The majority of patients *without* risk factors for MDR infection can be treated with a single agent. In fact, mortality is lower with a single agent than with combination therapy for those with a low mortality risk. For these patients, monotherapy options listed in Table 131-8 are active against the core pathogens, as well as nonresistant *P. aeruginosa*. However, in selected settings, it may be possible to use a nonpseudomonal agent such as cefotaxime or moxifloxacin. Unfortunately, the proportion of patients with no MDR risk factors is <10% in some ICUs and is unknown for HAP patients. The major difference from CAP is the markedly lower incidence of atypical pathogens in VAP; the exception is *Legionella*, which can be a nosocomial pathogen, especially with local epidemics due to breakdowns in the treatment of potable water in the hospital. The standard recommendation for patients *with* risk factors for MDR infection and a high mortality risk is for three antibiotics:

TABLE 131-8 Empirical Antibiotic Treatment of Hospital-Acquired and Ventilator-Associated Pneumonia

NO RISK FACTORS FOR RESISTANT GRAM-NEGATIVE PATHOGEN	RISK FACTORS FOR RESISTANT GRAM-NEGATIVE PATHOGEN[a] *(CHOOSE ONE FROM EACH COLUMN)*	
Piperacillin-tazobactam (4.5 g IV q6h) Cefepime (2 g IV q8h) Levofloxacin (750 mg IV q24h)	Piperacillin-tazobactam (4.5 g IV q6h) Cefepime (2 g IV q8h) Ceftazidime (2 g IV q8h) Imipenem (500 mg IV q6h) Meropenem (1 g IV q8h) Consider newer agents[c]	Amikacin (15–20 mg/kg IV q24h) Gentamicin (5–7 mg/kg IV q24h) Tobramycin (5–7 mg/kg IV q24h) Ciprofloxacin (400 mg IV q8h) Levofloxacin (750 mg IV q24h) Colistin (loading dose of 5 mg/kg IV followed by maintenance doses of 2.5 mg × [1.5 × CrCl + 30] IV q12h) Polymyxin B (2.5–3.0 mg/kg per day IV in 2 divided doses)
Risk Factors for MRSA[b] *(Add to above)*		
Linezolid (600 mg IV q12h) *or* Adjusted-dose vancomycin (trough level, 15–20 mg/dL)		

[a] Prior antibiotic therapy, prior hospitalization, local antibiogram.

[b] Prior antibiotic therapy, prior hospitalization, known MRSA colonization, chronic hemodialysis, local documented MRSA pneumonia rate >10% (or local rate unknown).

[c] Newer agents can have activity against resistant *P. aeruginosa* (ceftazidime-avibactam, ceftolozane-tazobactam, imipenem-relebactam, plazomicin), carbapenem-resistant *Enterobacteriaceae* (ceftazidime-avibactam, imipenem-relebactam, meropenem-vaborbactam), metallo-β-lactamase–producing *Enterobacteriaceae* (ceftazidime-avibactam, cefiderocol), *Stenotrophomona*s (cefiderocol), and *Acinetobacter* spp. (cefiderocol, sulbactam-durlobactam).

Abbreviations: CrCl, creatinine clearance rate; MRSA, methicillin-resistant *Staphylococcus aureus.*

two directed at *P. aeruginosa* and other resistant gram-negative organisms and one directed at MRSA. However, in the absence of septic shock, a single agent may be effective for these patients, provided there is a single agent that is likely to be effective against at least 90% of the gram-negative pathogens in that ICU. Empirical combination therapy enhances the likelihood of initially appropriate therapy over that with monotherapy. A β-lactam agent provides the greatest coverage, yet even the broadest-spectrum agent—a carbapenem—still constitutes inappropriate initial therapy in up to 10–15% of cases at some centers. The emergence of carbapenem resistance at some institutions requires the addition of polymyxins to the combination-therapy options. A number of emerging agents may modify our approach to therapy. New antipseudomonal agents include ceftazidime–avibactam, ceftolozane–tazobactam, imipenem–relebactam, and plazomicin. Therapy for carbapenem-resistant *Enterobacteriaceae* can consist of ceftazidime–avibactam, imipenem–relebactam, or meropenem–vaborbactam, while *Enterobacteriaceae* that produce metallo-β-lactamases can be treated with ceftazidime–avibactam or cefiderocol. *Acinetobacter* spp. can be treated with cefiderocol (as part of a combination regimen) or with sulbactam-durlobactam, and *Stenotrophomonas* can be treated with cefiderocol.

SPECIFIC TREATMENT

Once an etiologic diagnosis is made, broad-spectrum empirical therapy can be modified (de-escalated) to specifically address the known pathogen. For patients with MDR risk factors, antibiotic regimens can be reduced to a single agent in most cases. Only a minority of cases require a complete course with two or three drugs. A negative tracheal-aspirate culture or growth below the threshold for quantitative cultures of samples obtained before any antibiotic change strongly suggests that antibiotics should be discontinued or that an alternative diagnosis should be pursued. Identification of other confirmed or suspected sites of infection may require ongoing antibiotic therapy, but the spectrum of pathogens (and the corresponding antibiotic choices) may be different from those for VAP. A 7- or 8-day course of therapy is just as effective as a 2-week course and is associated with less frequent emergence of antibiotic-resistant strains. Exceptions include cases in which initial therapy is inappropriate or consists of second-line antibiotics and cases caused by some more resistant organisms, such as carbapenemase-producing *Acinetobacter* species. In these situations, serial measurements of procalcitonin may help guide duration of therapy.

A major controversy regarding specific therapy for VAP concerns the need for ongoing combination treatment of *Pseudomonas* pneumonia. No randomized controlled trials have demonstrated a benefit of combination therapy with a β-lactam and an aminoglycoside, nor have subgroup analyses in other trials found a survival benefit with such a regimen. Combination therapy may increase the likelihood of initially appropriate therapy and may have value in bacteremic infection with septic shock, but the benefit may last for only a few days. The unacceptably high rates of clinical failure and death despite combination therapy among patients with VAP caused by *P. aeruginosa* (see "Failure to Improve," below) indicate that better regimens are needed, perhaps including aerosolized antibiotics. In most cases of *Pseudomonas* pneumonia, current guidelines recommend against continuing combination therapy after the isolate's antimicrobial susceptibility is known.

FAILURE TO IMPROVE

Treatment failure is not uncommon in VAP, especially that caused by MDR pathogens. VAP caused by MRSA is associated with a 40% clinical failure rate when treated with standard-dose vancomycin. One proposed but unproven solution is the use of high-dose individualized treatment, although the risk of renal toxicity increases with this strategy. In addition, the MIC of MRSA to vancomycin has been increasing, and a high percentage of clinical failures occur when the MIC is in the upper range of sensitivity (i.e., 1.5–2 μg/mL). Linezolid appears to be 15% more efficacious than even adjusted-dose vancomycin and is preferred in patients with renal insufficiency and those infected with high-MIC isolates of MRSA. VAP due to *Pseudomonas* has a 40–50% failure rate, no matter what the regimen. Therapy-related causes of clinical failure include not using the recommended combination regimen (Table 131-8) and inadequate antibiotic dosing. However, the emergence of β-lactam resistance during therapy is an important problem, especially in infection with *Pseudomonas* and *Enterobacter* species. Recurrent VAP caused by the same pathogen is possible because the biofilm on endotracheal tubes allows persistence and reintroduction of the microorganism. Studies of VAP caused by *Pseudomonas* show that approximately half of recurrent cases are caused by a new strain. Some studies have suggested that treatment failure may be less common with optimized β-lactam dosing and use of either prolonged or continuous infusion therapy.

Possible causes of treatment failure can be difficult to determine early in the therapeutic course and can include superinfection, the presence of extrapulmonary infection, as well as patient factors such as severe comorbid illness and immunosuppression. Serial quantitative cultures may clarify the microbiologic response, and recent data in ICU patients have shown a role for biomarkers, such as procalcitonin, to guide duration of therapy in conjunction with the patient's initial response to treatment.

COMPLICATIONS

Apart from death, the major complication of VAP is prolongation of mechanical ventilation, with corresponding increases in the duration of ICU and hospital stay. In most studies, the need for additional mechanical ventilation resulting from VAP justifies aggressive efforts at prevention.

In rare cases, necrotizing pneumonia (e.g., due to *P. aeruginosa* or *S. aureus*) can cause significant pulmonary hemorrhage or empyema. More commonly, necrotizing infections result in the long-term complications of bronchiectasis and parenchymal scarring leading to recurrent pneumonia. Other long-term complications of pneumonia can include need for prolonged oxygen therapy, a catabolic state in a patient already nutritionally at risk, the necessity for ongoing rehabilitation, and—in the elderly—an inability to return to independent function and the need for nursing home placement.

FOLLOW-UP

Clinical improvement, if it occurs, is usually evident within 48–72 h of the initiation of antimicrobial treatment, usually with an improvement in oxygenation. Because findings on chest radiography often worsen initially during treatment, they are less helpful than clinical criteria as an indicator of response to therapy.

PROGNOSIS

VAP is associated with crude mortality rates as high as 50–70%, but the real issue is attributable mortality. Many patients with VAP have underlying diseases that would result in death even if VAP did not occur. Attributable mortality exceeded 25% in one matched-cohort study, while more recent studies have suggested much lower rates (5–10%), although patients with VAP complicating COVID-19 have a higher attributable mortality than those with other forms of VAP. Some variability in VAP mortality rates is clearly related to the type of patient and ICU studied. VAP in trauma patients is not associated with attributable mortality, possibly because many of the patients were otherwise healthy before being injured. The causative pathogen also plays a major role. Generally, MDR pathogens are associated with significantly greater attributable mortality than non-MDR pathogens. Pneumonia caused by some pathogens (e.g., *S. maltophilia*) is simply a marker for a patient whose immune system is highly compromised and is therefore at high risk.

PREVENTION (TABLE 131-7)

Because endotracheal intubation is a risk factor for VAP, the most important preventive intervention is to avoid intubation or minimize its duration. Successful noninvasive ventilation avoids many of the problems associated with endotracheal tubes. Strategies that minimize the duration of ventilation through daily holding of sedation and formal weaning protocols have also been highly effective in preventing VAP.

Unfortunately, a tradeoff in risks is sometimes necessary. Aggressive attempts to extubate early may result in reintubation(s) and increase aspiration, posing a risk of VAP. Heavy continuous sedation increases VAP risk, but self-extubation because of insufficient sedation is also a risk. The tradeoffs also apply to antibiotic therapy. Short-course antibiotic prophylaxis can decrease the risk of early-onset VAP in comatose patients requiring intubation, and data suggest that antibiotics decrease VAP rates overall. Conversely, prolonged courses of antibiotics consistently increase the risk of MDR VAP; pseudomonal VAP is rare among patients who have not recently received antibiotics. In one recent randomized trial, 3 days of daily inhaled aminoglycoside prophylaxis reduced the occurrence of VAP for the next 28 days, with no impact on mortality or antibiotic use.

Minimizing microaspiration around the endotracheal tube cuff also can prevent VAP. Simply elevating the head of the bed (at least 30° above horizontal, but preferably 45°) and using specially modified endotracheal tubes that allow removal of the secretions pooled above the cuff can prevent microaspiration. The risk-to-benefit ratio of transporting the patient outside the ICU for diagnostic tests or procedures should be carefully considered since VAP rates are increased among transported patients.

The role played by overgrowth of the normal bowel flora in the stomach—in the presence of elevated gastric pH—in the pathogenesis of VAP is questionable. Therefore, avoidance of agents that raise gastric pH may be relevant only in certain populations, such as liver transplant recipients and patients who have undergone other major intraabdominal procedures or who have bowel obstruction. MRSA and nonfermenters such as *P. aeruginosa* and *Acinetobacter* species are not normally part of the bowel flora but reside primarily in the nose and on the skin, respectively.

In outbreaks of VAP due to specific pathogens, the possibility of a breakdown in infection control measures (particularly contamination of reusable equipment) should be investigated. Even high rates of pathogens that are already common in a particular ICU may result from cross-infection. Education and reminders of the need for consistent hand washing and other infection-control practices can minimize this risk.

HOSPITAL-ACQUIRED PNEUMONIA

While less well studied than VAP, HAP in nonintubated patients—both inside and outside the ICU—is similar to VAP. The main differences are the higher frequency of non-MDR pathogens and the generally better underlying host immunity in nonintubated patients. The lower frequency of MDR pathogens allows monotherapy in a larger proportion of cases of HAP than of VAP. However, the bacteriology and outcome of ventilated HAP patients may be very similar to those of patients with VAP.

The only pathogens that may be more common in the non-VAP population are anaerobes because of a greater risk of macroaspiration and the lower oxygen tensions in the lower respiratory tract of these patients. Anaerobes usually contribute only to polymicrobial pneumonias, and specific therapy targeting anaerobes probably is not needed since many of the recommended antibiotics are active against anaerobes.

Diagnosis is even more difficult for HAP in the nonintubated patient than for VAP. Lower respiratory tract samples appropriate for culture are considerably more difficult to obtain from nonintubated patients. Many of the underlying diseases that predispose a patient to HAP are also associated with an inability to cough adequately. Since blood cultures are infrequently positive (<15% of cases), the majority of patients with HAP do not have culture data on which antibiotic modifications can be based, and de-escalation is less likely. Despite these difficulties, the better host defenses in non-ICU patients result in lower mortality rates than are documented for VAP and for ventilated HAP. In addition, the risk of antibiotic failure is lower in HAP.

GLOBAL IMPACT

From the available data, it is virtually impossible to accurately assess the impact of pneumonia from a global perspective. Any differences in incidence, disease burden, and costs across different age, ethnic, and racial groups are compounded by differences among countries in terms of etiologic pathogens, resistance rates, access to health-care and diagnostic facilities, and vaccine availability and use.

A standard approach with clearly defined outcome measures is needed before the impact of pneumonia can be accurately evaluated. However, simple extrapolation from U.S. data for CAP and HAP/VAP shows that pneumonia has a significant impact on quality of life, morbidity, health costs, and mortality rates and that this impact has implications for patients and for society as a whole.

Acknowledgment

The authors gratefully acknowledge the contributions of Richard Wunderink, MD, to this chapter in a prior edition.

FURTHER READING

Dequin PF et al: Hydrocortisone in severe community–acquired pneumonia. N Engl J Med 388:1931, 2023.

Dickson RP et al: Towards an ecology of the lung: New conceptual models of pulmonary microbiology and pneumonia pathogenesis. Lancet Respir Med 2:238, 2014.

File TM Jr: Community-acquired pneumonia. N Engl J Med 389:632, 2023.

Jain S et al: Community-acquired pneumonia requiring hospitalization among U.S. adults. N Engl J Med 373:415, 2015.

Kalil AC et al: Management of adults with hospital-acquired and ventilator-associated pneumonia: 2016 clinical practice guidelines by the Infectious Diseases Society of America and the American Thoracic Society. Clin Infect Dis 63:e61, 2016.

Mandell LA, Niederman MS: Aspiration pneumonia. N Engl J Med 380:651, 2019.

Mandell LA et al: Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis 44:S27, 2007.

Martin-Loeches I et al: ERS/ESICM/ESCMID/ALAT guidelines for the management of severe community-acquired pneumonia. Eur Respir J 61:2200735, 2023.

Metlay JP et al: Diagnosis and treatment of adults with community-acquired pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious Diseases Society of America. Am J Respir Crit Care Med 200:e45, 2019.

Vacheron CH et al: Attributable mortality of ventilator-associated pneumonia among patients with COVID-19. Am J Respir Crit Care Med 206:161, 2022.

132

Lung Abscess

Rebecca M. Baron, Beverly W. Baron, Miriam Baron Barshak

Lung abscess represents necrosis and cavitation of the lung following microbial infection. Lung abscesses can be single or multiple but usually are marked by a single dominant cavity >2 cm in diameter.

ETIOLOGY

The low prevalence of lung abscesses makes them difficult to study in randomized controlled trials. Although the incidence of lung abscesses has decreased in the antibiotic era, they are still a source of significant morbidity and mortality.

Lung abscesses usually are characterized as either primary (~80% of cases) or secondary. *Primary* lung abscesses generally arise from aspiration, often are caused principally by anaerobic bacteria, and occur in the absence of an underlying pulmonary or systemic condition. *Secondary* lung abscesses arise in the setting of an underlying condition, such as a postobstructive process (e.g., bronchial foreign body, tumor) or a systemic process (e.g., HIV infection, another immunocompromising condition). Lung abscesses can also be characterized as acute (<4–6 weeks in duration) or chronic (~40% of cases).

EPIDEMIOLOGY

The majority of the existing epidemiologic information involves primary lung abscesses. In general, middle-aged men are more commonly affected than middle-aged women. The major risk factor for primary lung abscesses is aspiration. Patients at particular risk for aspiration, such as those with altered mental status, alcoholism, drug overdose, seizures, bulbar dysfunction, prior cerebrovascular or cardiovascular events, or neuromuscular disease, are most commonly affected. At additional risk are patients with esophageal dysmotility or esophageal lesions (strictures or tumors) and those with gastric distention and/or gastroesophageal reflux, especially those who spend substantial time in the recumbent position.

It is widely thought that colonization of the gingival crevices by anaerobic bacteria or microaerophilic streptococci (especially in patients with gingivitis and periodontal disease), combined with a risk of aspiration, is important in the development of lung abscesses. In fact, many physicians consider it extremely rare for lung abscesses to develop in the absence of teeth as a nidus for bacterial colonization.

The importance of these risk factors in the development of lung abscesses is highlighted by a significant reduction in abscess incidence in the late 1940s that coincided with a change in oral surgical technique: beginning at that time, these operations were no longer performed with the patient in the seated position without a cuffed endotracheal tube, and the frequency of perioperative aspiration events was thus decreased. In addition, the introduction of penicillin around the same time significantly reduced the incidence of and mortality rate from lung abscess.

TABLE 132-1 Examples of Microbial Pathogens That Can Cause Lung Abscesses

CLINICAL CONDITION	PATHOGENS
Primary lung abscess (often with risk factors for aspiration)	Anaerobes (e.g., *Peptostreptococcus* spp., *Prevotella* spp., *Bacteroides* spp., milleri group streptococci), microaerophilic streptococci
Secondary lung abscess (often with underlying immunocompromise)	*Staphylococcus aureus*, gram-negative rods (e.g., *Pseudomonas aeruginosa*, Enterobacteriaceae), *Nocardia* spp., *Aspergillus* spp., Mucorales, *Cryptococcus* spp., *Legionella* spp., *Rhodococcus equi*, *Pneumocystis jirovecii*
Embolic lesions	*Staphylococcus aureus* (often from endocarditis), *Fusobacterium necrophorum* (Lemierre's syndrome; see text for details)
Endemic infections (with or without underlying immunocompromise)	*Mycobacterium tuberculosis* (as well as *Mycobacterium avium* and *Mycobacterium kansasii*), *Coccidioides* spp., *Histoplasma capsulatum*, *Blastomyces* spp., parasites (e.g., *Entamoeba histolytica*, *Paragonimus westermani*, *Strongyloides stercoralis*)
Miscellaneous conditions	Bacterial pathogen (often *S. aureus*) after influenza or another viral infection, *Actinomyces* spp.

PATHOGENESIS

Primary Lung Abscesses The development of primary lung abscesses is thought to originate when chiefly anaerobic bacteria (as well as microaerophilic streptococci) in the gingival crevices are aspirated into the lung parenchyma in a susceptible host (Table 132-1). Patients who develop primary lung abscesses usually carry an overwhelming burden of aspirated material or are unable to clear the bacterial load. Pneumonitis develops initially (exacerbated in part by tissue damage caused by gastric acid); then, over a period of 7–14 days, the bacteria produce parenchymal necrosis and cavitation whose extent depends on host–pathogen interaction (Fig. 132-1). Anaerobes are thought to produce more extensive tissue necrosis in polymicrobial infections in which virulence factors of the various bacteria can act synergistically to cause more significant tissue destruction.

Secondary Lung Abscesses The pathogenesis of secondary abscesses depends on the predisposing factor. For example, in cases of bronchial obstruction from malignancy or a foreign body, the

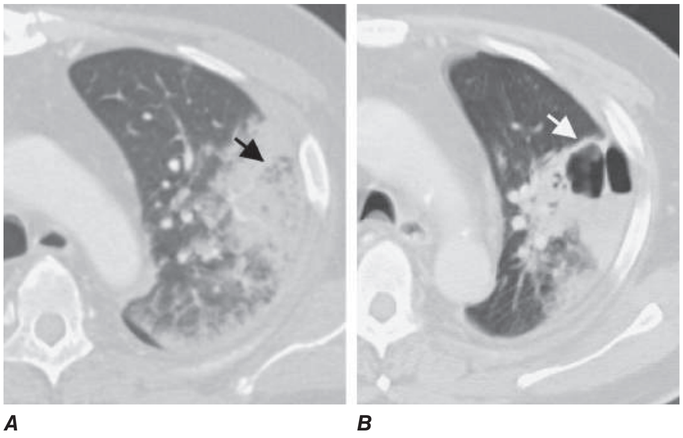

FIGURE 132-1 Representative chest CT scans demonstrating development of lung abscesses. This patient was immunocompromised from underlying lymphoma and developed severe *Pseudomonas aeruginosa* pneumonia, as represented by a left lung infiltrate with concern for central regions of necrosis (*panel **A**, black arrow*). Two weeks later, areas of cavitation with air-fluid levels were visible in this region and were consistent with the development of lung abscesses (*panel **B**, white arrow*). *(Images provided by Dr. Ritu Gill, formerly of the Division of Chest Radiology, Brigham and Women's Hospital, Boston; with permission.)*

obstructing lesion prevents clearance of oropharyngeal secretions, leading to abscess development. With underlying systemic conditions (e.g., immunosuppression after bone marrow or solid organ transplantation), impaired host defense mechanisms lead to increased susceptibility to the development of lung abscesses caused by a broad range of pathogens, including opportunistic organisms (Table 132-1).

Lung abscesses also arise from septic emboli, either in tricuspid valve endocarditis (often involving *Staphylococcus aureus*) or in Lemierre's syndrome, in which an infection begins in the pharynx (classically involving *Fusobacterium necrophorum*) and then spreads to the neck and the carotid sheath (which contains the jugular vein) to cause septic thrombophlebitis.

PATHOLOGY AND MICROBIOLOGY

Primary Lung Abscesses The dependent segments (posterior upper lobes and superior lower lobes) are the most common locations of primary lung abscesses, given the predisposition of aspirated materials to be deposited in these areas. Generally, the right lung is affected more commonly than the left because the right mainstem bronchus is less angulated.

Primary lung abscesses often are polymicrobial, primarily including anaerobic organisms as well as microaerophilic streptococci (Table 132-1). The retrieval and culture of anaerobes can be complicated by the contamination of samples with microbes from the oral cavity, the need for expeditious transport of the cultures to the laboratory, the need for early plating with special culture techniques, the prolonged time required for culture growth, and the need for collection of specimens prior to administration of antibiotics. When attention is paid to these factors, rates of recovery of specific isolates are reportedly as high as 78%.

Because it is not clear that knowing the identity of the causative anaerobic isolate alters the approach to treatment of a primary lung abscess, practice has shifted away from the use of specialized techniques to obtain material for culture, such as transtracheal aspiration and bronchoalveolar lavage with protected brush specimens that allow recovery of culture material while avoiding contamination from the oral cavity. When no pathogen is isolated from a primary lung abscess (which occurs as often as 40% of the time), the abscess is termed a *nonspecific lung abscess*, and the presence of anaerobes is often presumed. A *putrid lung abscess* refers to cases with foul-smelling breath, sputum, or empyema; these manifestations are essentially diagnostic of an anaerobic lung abscess.

Secondary Lung Abscesses The location of secondary abscesses may vary with the underlying cause. The microbiology of secondary lung abscesses can encompass a broad bacterial spectrum, with infection by *Pseudomonas aeruginosa* and other gram-negative rods the most common. In addition, a broad array of pathogens can be identified in patients from certain endemic areas and in specific clinical scenarios (e.g., a significant incidence of fungal infections among immunosuppressed patients following bone marrow or solid organ transplantation). Because immunocompromised hosts and patients without the classic presentation of a primary lung abscess can be infected with a wide array of unusual organisms (Table 132-1), it is of special importance to obtain culture material to target therapy.

CLINICAL MANIFESTATIONS

Clinical manifestations initially may be similar to those of pneumonia, with fevers, cough, sputum production, and chest pain; a more chronic and indolent presentation that includes night sweats, fatigue, and anemia is often observed with anaerobic lung abscesses. A subset of patients with putrid lung abscesses may report discolored phlegm and foul-tasting or foul-smelling sputum. Patients with lung abscesses due to non-anaerobic organisms, such as *S. aureus*, may present with a more fulminant course characterized by high fevers and rapid progression.

Findings on physical examination may include fevers, poor dentition, and/or gingival disease as well as amphoric and/or cavernous breath sounds on lung auscultation. Additional findings may include digital clubbing and the absence of a gag reflex.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of lung abscesses is broad and includes other noninfectious processes that result in cavitary lung lesions, including lung infarction, malignancy, sequestration, cryptogenic organizing pneumonia, sarcoidosis, vasculitides and autoimmune diseases (e.g., granulomatosis with polyangiitis), lung cysts or bullae containing fluid, and septic emboli (e.g., from tricuspid valve endocarditis). Other less common entities can include pulmonary manifestations of diseases that usually present at locations other than the chest (e.g., inflammatory bowel disease, pyoderma gangrenosum).

DIAGNOSIS

Lung abscesses are documented by chest imaging. Although a chest radiograph usually detects a thick-walled cavity with an air-fluid level, CT permits better definition and may provide earlier evidence of cavitation. CT may also yield additional information regarding a possible underlying cause of lung abscess, such as malignancy, and may help distinguish a peripheral lung abscess from a pleural infection. This distinction has important implications for treatment because a pleural space infection, such as an empyema, may require urgent drainage.

As described earlier (see "Pathology and Microbiology," above), more invasive diagnostics (such as transtracheal aspiration) were traditionally undertaken for primary lung abscesses, whereas empirical therapy that includes drugs targeting anaerobic organisms currently is used more often. While sputum can be collected noninvasively for Gram's stain and culture, which may yield a pathogen, the infection is likely to be polymicrobial, and culture results may not reflect the presence of anaerobic organisms. Increasing use of molecular techniques for bacterial detection (e.g., 16S RNA gene amplification) may eventually yield more specific pathogen identification. As stated above, many physicians consider putrid-smelling sputum to be virtually diagnostic of an anaerobic infection.

When a secondary lung abscess is present or empirical therapy fails to elicit a response, sputum and blood cultures are advised in addition to serologic studies for opportunistic pathogens (e.g., viruses and fungi causing infections in immunocompromised hosts). Additional diagnostics, such as bronchoscopy with bronchoalveolar lavage or protected brush specimen collection and CT-guided percutaneous needle aspiration, can be undertaken. Risks posed by these more invasive diagnostics include spillage of abscess contents into the other lung (with bronchoscopy) and pneumothorax and bronchopleural fistula development (especially with CT-guided needle aspiration). However, early diagnostics in secondary abscesses, especially in immunocompromised hosts, are particularly important because the patients involved may be especially fragile, at risk for infection with a broad array of pathogens, and therefore less likely than other patients to respond to empirical therapy.

TREATMENT

Lung Abscess

The availability of antibiotics in the 1940s and 1950s established therapy with this drug class as the primary approach to the treatment of lung abscess. Previously, surgery had been relied upon much more frequently. For many decades, penicillin was the antibiotic of choice for primary lung abscesses in light of its anaerobic coverage; however, because oral anaerobes can produce β-lactamases, clindamycin has proved superior to penicillin in clinical trials. For primary lung abscesses, the recommended regimens are (1) clindamycin (600 mg IV three times daily; then, with the disappearance of fever and clinical improvement, 300 mg PO four times daily) or (2) an IV-administered β-lactam/β-lactamase combination, followed—once the patient's condition is stable—by orally administered amoxicillin-clavulanate. This therapy should be continued until imaging demonstrates that the lung abscess has cleared or regressed to a small scar. Treatment duration may range from 3–4 weeks to as long as 14 weeks, with some literature suggesting that a course of at least 6 weeks may be associated with better outcomes. One small study suggested that moxifloxacin (400 mg/d PO) is as effective and well tolerated as ampicillin-sulbactam. Notably, metronidazole is not effective as a single agent: it covers anaerobic organisms but not the microaerophilic streptococci that are often components of the mixed flora of primary lung abscesses.

In secondary lung abscesses, antibiotic coverage should be directed at the identified pathogen, and a prolonged course (until resolution of the abscess is documented) is often required. Treatment regimens and courses vary widely, depending on the immune state of the host and the identified pathogen. Other interventions may be necessary as well, such as relief of an obstructing lesion or treatment directed at the underlying condition predisposing the patient to lung abscess. Similarly, if the condition of patients with presumed primary lung abscess fails to improve, additional studies to rule out an underlying predisposing cause for a secondary lung abscess are indicated.

Although it can take as long as 7 days for patients receiving appropriate therapy to defervesce, as many as 10–20% of patients may not respond at all, with continued fevers and progression of the abscess cavity on imaging. An abscess >6–8 cm in diameter is less likely to respond to antibiotic therapy without additional interventions. Options for patients who do not respond to antibiotics and whose additional diagnostic studies fail to identify a pathogen that can be treated include surgical resection and percutaneous drainage of the abscess, especially when the patient is a poor surgical candidate. Timing of surgical intervention can be challenging; the goal is to balance the morbidity/mortality risk of a procedure with the need for definitively clearing the abscess in the setting of persistent infection that is not responsive to nonsurgical approaches. Possible complications of percutaneous drainage include bacterial contamination of the pleural space as well as pneumothorax and hemothorax. Traversing normal lung parenchyma might represent a risk factor for major complications from percutaneous abscess drainage.

COMPLICATIONS

Larger cavity size on presentation may correlate with the development of persistent cystic changes (pneumatoceles) or bronchiectasis. Additional possible complications include recurrence of abscesses despite appropriate therapy, extension to the pleural space with development of empyema, life-threatening hemoptysis, and massive aspiration of lung abscess contents.

PROGNOSIS AND PREVENTION

Reported mortality rates for primary abscesses have been as low as 2%, while rates for secondary abscesses are generally higher—as high as 75% in some case series. Other poor prognostic factors include age >60, malignancy-related abscesses, the presence of aerobic bacteria, sepsis at presentation, symptom duration of >8 weeks, and abscess size >6 cm.

Mitigation of underlying risk factors may be the best approach to prevention of lung abscesses, with attention directed toward airway protection, oral hygiene, and minimized sedation with elevation of the head of the bed for patients at risk for aspiration. Prophylaxis against certain pathogens in at-risk patients (e.g., recipients of bone marrow or solid organ transplants or patients whose immune systems are significantly compromised by HIV infection) may be undertaken.

APPROACH TO THE PATIENT

Lung Abscess

For patients with a lung abscess and a low likelihood of malignancy (e.g., smokers <45 years old) and with risk factors for aspiration, it is reasonable to administer empirical treatment and then to pursue further evaluation if therapy does not elicit a response. However, some clinicians may opt for up-front cultures, even in primary lung abscesses. In patients with risk factors for malignancy or other underlying conditions (especially immunocompromised hosts) or with an atypical presentation, earlier diagnostics should be considered, such as bronchoscopy with biopsy or CT-guided needle aspiration. Bronchoscopy should be performed early in patients whose history, symptoms, or imaging findings are consistent with possible bronchial obstruction. In patients from areas endemic for tuberculosis or patients with other risk factors for tuberculosis (e.g., underlying HIV infection), induced sputum samples should be examined early in the workup to rule out this disease.

FURTHER READING

Bartlett JG: How important are anaerobic bacteria in aspiration pneumonia: When should they be treated and what is optimal therapy. Infect Dis Clin North Am 27:149, 2013.

Desai H, Agrawal A: Pulmonary emergencies: Pneumonia, acute respiratory distress syndrome, lung abscess, and empyema. Med Clin North Am 96:1127, 2012.

Lee JH et al: Percutaneous transthoracic catheter drainage for lung abscess: A systematic review and meta-analysis. Eur Radiol 32:1184, 2022.

Maitre T et al: Pyogenic lung abscess in an infectious disease unit: A 20-year retrospective study. Ther Adv Respir Dis 17534666211003012, 2021.

Ott SR et al: Moxifloxacin vs ampicillin/sulbactam in aspiration pneumonia and primary lung abscess. Infection 36:23, 2008.

Raymond D: Surgical intervention for thoracic infections. Surg Clin North Am 94:1283, 2014.

Vaarst JK et al: Lung abscess: Clinical characteristics of 222 Danish patients diagnosed from 2016 to 2021. Respir Med 216: 107305, 2023.

133

Infective Endocarditis

Sara E. Cosgrove, Michael T. Melia

The prototypic lesion of infective endocarditis (IE), the *vegetation* **(Fig. 133-1)**, is a mass of platelets, fibrin, microorganisms, and scant inflammatory cells. Infection most commonly involves heart valves but may also occur on the low-pressure side of a ventricular septal defect, on mural endocardium damaged by aberrant jets of blood or foreign bodies, or on intracardiac devices. The analogous process involving arteriovenous shunts, arterio-arterial shunts (patent ductus arteriosus), or a coarctation of the aorta is called *infective endarteritis*.

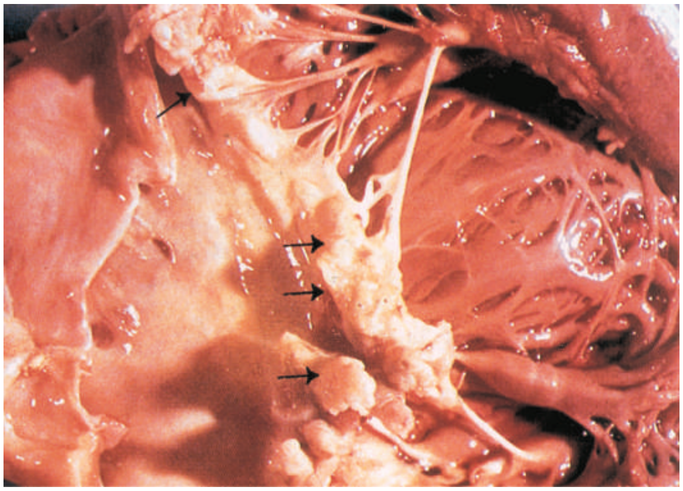

FIGURE 133-1 Vegetations (*arrows*) due to viridans streptococci endocarditis involving the mitral valve.

IE can be classified according to the temporal evolution of disease, the site of infection, the cause of infection, or the predisposing risk factor (e.g., injection drug use, health care–associated). *Acute IE* is a hectically febrile illness that rapidly damages cardiac structures, seeds extracardiac sites, and, if untreated, progresses to death within weeks. *Subacute IE* follows an indolent course; causes structural cardiac damage only slowly, if at all; rarely metastasizes; and is gradually progressive unless complicated by a major embolic event or a ruptured mycotic aneurysm.

In the United States and likely in other developed countries, the incidence of IE is estimated to be 15 cases per 100,000 population per year, with progressive increases during recent decades. While congenital heart diseases remain a constant predisposition, predisposing conditions in developed countries have shifted from chronic rheumatic heart disease (still common in developing countries) to injection drug use, degenerative valve disease, and intracardiac devices. Although the incidence of IE is increased among the elderly, recent data indicate age-adjusted mortality rates in people ≥55 years old have declined in the United States. Recently, however, there has been acceleration in mortality in people aged 25–44 years, likely associated with an increase in opioid use disorder (OUD) and injection drug use in this age group. In developed countries, 25–35% of cases of native-valve endocarditis (NVE) are health care–associated, and 16–30% of all cases are prosthetic-valve infections (PVE). The risk of PVE is greatest during the initial year after valve replacement; gradually declines to a low, stable rate thereafter; and is greater for bioprosthetic valves than mechanical valves. The incidence and rate of decline of transcatheter aortic valve replacement (TAVR)-PVE are similar to those for surgically implanted bioprosthetic aortic valves. IE involving cardiovascular implantable electronic devices (CIED-IE)—greater on implanted defibrillators and resynchronization devices than on permanent pacemakers—occurs in 0.5–1.14 cases per 1000 recipients.

■ ETIOLOGY

Although many species of bacteria and fungi cause sporadic episodes of IE, a few bacterial species cause the majority of cases **(Table 133-1)**. Recent large studies from developed areas identify *Staphylococcus aureus* as the most common bacterial species causing IE. The oral cavity, skin, and upper respiratory tract are the respective primary portals for viridans streptococci, staphylococci, and HACEK organisms (*Haemophilus* species, *Aggregatibacter* species, *Cardiobacterium hominis*, *Eikenella corrodens*, and *Kingella kingae*). *Streptococcus gallolyticus* subspecies *gallolyticus* (formerly *S. bovis* biotype 1) originates from the gastrointestinal tract and is associated with colonic polyps and tumors. Enterococci enter the bloodstream primarily from the genitourinary tract. Health care–associated IE, most commonly caused by *S. aureus*, coagulase-negative staphylococci (CoNS), and enterococci, may have either a nosocomial onset (55%) or a community onset (45%). IE complicates 8–25% of episodes of catheter-associated *S. aureus* bacteremia; the higher rates are detected in high-risk patients studied by transesophageal echocardiography (TEE) (see "Cardiac Imaging," below).

PVE arising within 2 months of valve surgery—i.e., early PVE—is generally nosocomial and is the result of intraoperative contamination of the prosthesis or a postoperative infection. This nosocomial origin is reflected in the microbial causes: *S. aureus*, CoNS, facultative gram-negative bacilli, diphtheroids, and fungi. The portals of entry and organisms causing PVE beginning >12 months after surgery—i.e., late PVE—are similar to those in community-acquired NVE. Regardless of the time of onset after surgery, the majority of CoNS strains that cause PVE are resistant to methicillin. The microbiology of TAVR-PVE, while generally similar to that of PVE, is notable for an increased frequency of enterococci. Risk factors associated with TAVR-PVE include male sex, diabetes, renal failure, and moderate postimplantation aortic valve regurgitation.

CIED-IE involves the device or the endothelium at points of device contact. Occasionally, there is concurrent valvular infection. One-third of cases of CIED-IE present within 3 months after device implantation or manipulation, one-third between 3 and 12 months, and one-third >1 year. *S. aureus* and CoNS cause the majority of cases.

IE in people who inject drugs (PWID), especially that involving the tricuspid valve, is commonly caused by *S. aureus*, which is often resistant to methicillin. Left-sided valve infections in PWID have a more varied etiology. In addition to the usual causes of IE, infection due to Enterobacterales, *Pseudomonas aeruginosa*, *Candida* species, and sporadically by unusual organisms (*Bacillus*, *Lactobacillus*, *Corynebacterium* species) is encountered.

About 5–15% of patients with IE have negative blood cultures; in one-third to one-half of these cases, cultures are negative because of prior antibiotic exposure. The remainder are infected by fastidious organisms, such as some streptococci; nutritionally variant bacteria now designated *Granulicatella*, *Gemella*, and *Abiotrophia* species; *Coxiella burnetii*; and *Bartonella* species. Some fastidious organisms

TABLE 133-1 Organisms Causing Major Clinical Forms of Infective Endocarditis (IE)

	PROPORTION OF CASES						
	NATIVE-VALVE IE		PROSTHETIC-VALVE IE AT INDICATED TIME OF ONSET (MONTHS) AFTER VALVE SURGERY			TAVR PVE	CIED-IE
ORGANISM(S)	COMMUNITY-ACQUIRED (*N* = 1718)	HEALTH CARE–ASSOCIATED (*N* = 1110)	<2 (*N* = 144)	>2-12 (*N* = 31)	>12 (N = 194)	(*N* = 295)	(*N* = 337)
Streptococci[b]	40	13	1	10	31	18	2
Pneumococci	2	—	—	—	—	-	—
Enterococci[c]	9	16	8	13	11	24	4
Staphylococcus aureus	28[a]	52[d]	22	13	18	23	36
Coagulase-negative staphylococci	5	11	33	35	11	20	41
Fastidious gram-negative coccobacilli (HACEK group)[e]	3	—	—	—	6	—	—
Gram-negative bacilli	1	1	13	3	6	1	6
Candida spp.	<1	1	8	13	1	1	2
Polymicrobial/miscellaneous	3	3	3	6	5	8	2
Diphtheroids	—	<1	6	—	3		1
Culture-negative	9	3	5	7	8	5	6

[a] Includes methicillin-susceptible and -resistant isolates. [b] Includes viridans streptococci; *Streptococcus gallolyticus*; other non–group A, groupable streptococci; and *Abiotrophia* and *Granulicatella* spp. (nutritionally variant, pyridoxal-requiring streptococci). [c] Primarily *E. faecalis* or nonspeciated isolates; occasionally *E. faecium* or other less likely species. [d] Methicillin resistance is common among these *S. aureus* strains. [e] Includes *Haemophilus* spp., *Aggregatibacter* spp., *Cardiobacterium hominis*, *Eikenella corrodens*, and *Kingella kingae*.

Abbreviations: CIED, cardiac implantable electronic device; TAVR, transcatheter aortic valve replacement.

Note: Data are compiled from multiple studies.

occur in characteristic geographic settings (e.g., *C. burnetii* and *Bartonella* species in Europe, *Brucella* species in the Middle East). *Tropheryma whipplei* causes an indolent, culture-negative form of IE. *C. burnetii* has a predilection for prosthetic valves. *Corynebacterium* species and *Cutibacterium acnes* may involve intracardiac devices and be slow to grow in blood cultures. *Mycobacterium chimaera,* which may be difficult to recover from blood cultures unless special media is used, has caused a global outbreak of PVE and disseminated infection as a result of aerosols from contaminated heater-cooler machines used during cardiopulmonary bypass. Lastly, atrial myxoma, marantic endocarditis, and the antiphospholipid antibody syndrome may mimic culture-negative IE.

■ PATHOGENESIS

The undamaged endothelium is resistant to infection by most bacteria. Endothelial injury (e.g., at the site of impact of high-velocity blood jets or on the low-pressure side of a cardiac structural lesion) allows either direct infection by virulent organisms or the development of a platelet–fibrin thrombus—a condition called *nonbacterial thrombotic endocarditis* (NBTE). This thrombus serves as a site of bacterial attachment during transient bacteremia. The cardiac conditions most commonly resulting in NBTE are mitral regurgitation, aortic stenosis, aortic regurgitation, ventricular septal defects, and complex congenital heart disease. NBTE also arises as a result of a hypercoagulable state; this phenomenon gives rise to *marantic endocarditis* (uninfected vegetations seen in patients with malignancy and chronic diseases) and to bland vegetations complicating systemic lupus erythematosus and antiphospholipid antibody syndrome.

Organisms that cause IE enter the bloodstream from colonized body surfaces or sites of infection. *S. aureus* adherence to intact endothelium may be mediated by local inflammation inducing von Willebrand factor on endothelial cell surfaces with resulting adherence of both platelets and *S. aureus.* Alternatively, *S. aureus* adherence to injured endothelium may be mediated by local deposition of fibrin and circulating von Willebrand factor on exposed subendothelial tissue to which in turn *S. aureus* adhere directly. Other microorganisms in the blood adhere to NBTE. The organisms that commonly cause IE have surface adhesin molecules, collectively called microbial surface components recognizing adhesin matrix molecules (MSCRAMMs) that mediate adherence to NBTE sites or injured endothelium. Adherence is facilitated by fibronectin-binding proteins present on many gram-positive bacteria; by clumping factor (a fibrinogen- and fibrin-binding surface protein) on *S. aureus*; by fibrinogen-binding surface proteins (Fss2), collagen-binding surface protein (Ace), and Ebp pili (the latter mediating platelet adherence) on *Enterococcus faecalis*; and by glucans or FimA (a member of the family of oral mucosal adhesins) on streptococci. Fibronectin-binding proteins are required for *S. aureus* invasion of intact endothelium; thus, these surface proteins may facilitate infection of previously normal valves. If resistant to the bactericidal activity of serum and the microbicidal peptides released locally by platelets, adherent organisms proliferate to form dense microcolonies. Microorganisms also induce platelet deposition and a localized procoagulant state by eliciting tissue factor from the endothelium and, in the case of *S. aureus,* from monocytes as well. Fibrin deposition combines with platelet aggregation and microorganism proliferation to generate an infected vegetation. Organisms deep in vegetations are metabolically inactive (nongrowing) and relatively resistant to killing by antimicrobial agents. Proliferating surface organisms are shed into the bloodstream continuously.

The clinical manifestations of IE—other than constitutional symptoms, which probably result from cytokine production—arise from damage to intracardiac structures; embolization of vegetation fragments leading to infection or infarction of remote tissues; hematogenous infection of sites during bacteremia; and tissue injury due to the deposition of circulating immune complexes or immune responses to deposited bacterial antigens.

■ CLINICAL MANIFESTATIONS

The highly variable clinical IE syndrome spans a continuum between acute and subacute presentations. Most forms of IE share clinical and laboratory manifestations **(Table 133-2)**. The causative microorganism is primarily responsible for the temporal course of IE. β-Hemolytic streptococci, *S. aureus*, and pneumococci typically result in an acute course, and IE caused by *Staphylococcus lugdunensis* (a coagulase-negative species) or by enterococci may present acutely. Subacute IE is typically caused by viridans streptococci, enterococci, CoNS, and the HACEK group. IE caused by *Bartonella* species, *T. whipplei*, *C. burnetii*, or *M. chimaera* is exceptionally indolent.

TABLE 133-2 Clinical and Laboratory Features of Infective Endocarditis

FEATURE	FREQUENCY, %
Fever	80–90
Chills and sweats	40–75
Anorexia, weight loss, malaise	25–50
Myalgias, arthralgias	15–30
Back pain	7–15
Heart murmur	80–85
New/worsened regurgitant murmur	20–50
Arterial emboli	20–50
Splenomegaly	15–50
Clubbing	10–20
Neurologic manifestations	20–40
Peripheral manifestations (Osler's nodes, subungual hemorrhages, Janeway lesions, Roth's spots)	2–15
Petechiae	10–40
Laboratory manifestations	
Anemia	70–90
Leukocytosis	20–30
Microscopic hematuria	30–50
Elevated erythrocyte sedimentation rate	60–90
Elevated C-reactive protein level	>90
Rheumatoid factor	50
Circulating immune complexes	65–100
Decreased serum complement	5–40

In patients with subacute presentations, fever is typically low-grade, rarely exceeding 39.4°C (103°F); in contrast, temperatures of 39.4°–40°C (103°–104°F) are often noted in acute IE. Fever may be blunted in patients who are elderly, are severely debilitated, or have renal failure.

Cardiac Manifestations Although heart murmurs are usually indicative of the predisposing cardiac pathology rather than of IE, valvular damage and ruptured chordae may result in new regurgitant murmurs. In acute IE involving a normal valve, murmurs may be absent initially but ultimately are detected in 85% of cases. Congestive heart failure (CHF) resulting from valve dysfunction or, occasionally, intracardiac fistulae develop in 30–40% of patients. Extension of leaflet infection into adjacent annular or myocardial tissue results in paravalvular abscesses, which in turn may cause intracardiac fistulae with new murmurs. Aortic paravalvular infection may burrow into the upper ventricular septum and interrupt the conduction system, leading to varying degrees of heart block. Mitral paravalvular abscesses are more distant from the conduction system and rarely cause conduction abnormalities. Coronary artery emboli occur in 2% of patients and may result in myocardial infarction.

Noncardiac Manifestations The classic nonsuppurative peripheral manifestations of subacute IE (e.g., Janeway lesions; **Fig. 133-2*A*)** are related to prolonged infection; with early diagnosis and treatment, these have become infrequent. In contrast, septic embolization mimicking some of these lesions (subungual hemorrhage, Osler's nodes) is common in patients with acute *S. aureus* IE **(Fig. 133-2*B*)**. Musculoskeletal pain usually remits promptly with treatment but

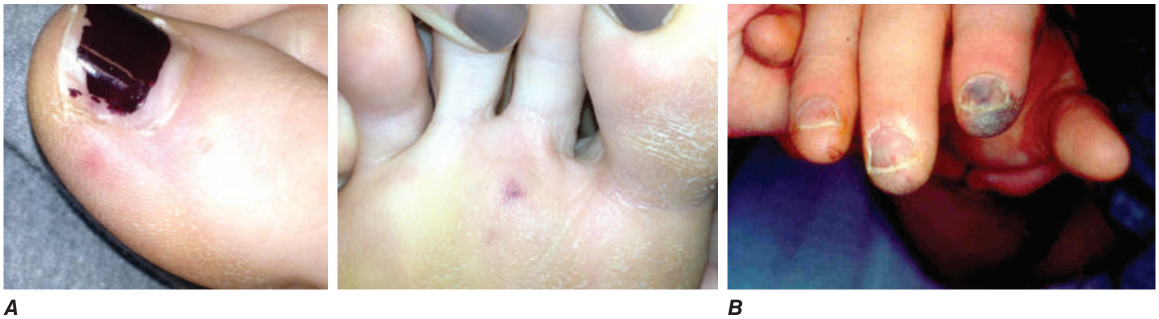

FIGURE 133-2 ***A.*** Janeway lesions on the toe (*left*) and plantar surface (*right*) of the foot in subacute *Neisseria mucosa* infective endocarditis (IE). *(Images courtesy of Rachel Baden, MD.)* ***B.*** Septic emboli with hemorrhage and infarction due to acute *Staphylococcus aureus* IE.

must be distinguished from focal metastatic infections (e.g., spondylodiscitis), which may complicate 10–15% of cases. Hematogenously seeded focal infection occurs most often in the skin, spleen, kidneys, skeletal system, and meninges. Arterial emboli, one-half of which precede the diagnosis of IE, are clinically apparent in up to 50% of patients. *S. aureus* IE, mobile vegetations >10 mm in diameter, and infection involving the mitral valve anterior leaflet are independently associated with an increased risk of embolization. Embolic arterial occlusion causes regional pain or ischemia-induced organ dysfunction (e.g., of the kidney, spleen, bowel, extremity). Cerebrovascular emboli presenting as strokes or occasionally as encephalopathy complicate 15–35% of cases; however, evidence of clinically asymptomatic emboli is found on magnetic resonance imaging (MRI) in 30–65% of patients with left-sided IE. The frequency of stroke is 8 per 1000 patient-days during the week prior to diagnosis and decreases to 4.8 and 1.7 per 1000 patient-days during the first and second weeks of effective antimicrobial therapy, respectively. Only 3% of strokes occur after 1 week of effective therapy. Emboli occurring late during or after effective therapy do not in themselves constitute evidence of failed antimicrobial treatment.

Other neurologic complications include aseptic or purulent meningitis, intracranial hemorrhage due to hemorrhagic infarcts or ruptured mycotic aneurysms, and seizures. *Mycotic aneurysms* are focal dilations of arteries occurring at points in the artery wall that have been weakened by infection in the vasa vasorum or where septic emboli have lodged. Microabscesses in the brain and meninges occur commonly in *S. aureus* IE; intracerebral abscesses requiring surgical drainage are infrequent.

Immune complex deposition on the glomerular basement membrane causes diffuse hypocomplementemic glomerulonephritis and renal dysfunction, which typically improve with effective antimicrobial therapy. Embolic renal infarcts cause flank pain and hematuria but rarely renal dysfunction. Splenic infarcts or abscess can manifest as left upper abdominal, pleuritic chest, or left shoulder pain.

Manifestations with Specific Predisposing Conditions Among PWID, 35–60% of IE is limited to the tricuspid valve and presents with fever with faint or no murmur and without peripheral manifestations. Septic pulmonary emboli, which are common with tricuspid IE, cause cough, pleuritic chest pain, nodular pulmonary infiltrates, and occasionally empyema or pyopneumothorax. Infection of the aortic or mitral valve presents with the typical clinical features of IE, including peripheral manifestations.

Health care–associated IE has typical manifestations unless associated with an intracardiac device or masked by the symptoms of concurrent illness. CIED-IE may be associated with obvious (especially within 6 months of device manipulation) or cryptic generator pocket infection or arise through bacteremic seeding without pocket infection. Fever, sepsis, minimal murmur, and occasionally pulmonary symptoms due to septic emboli are seen. Late-onset PVE and TAVR-PVE present with typical clinical features. In early PVE, symptoms may be masked by recent surgery. In both early and late PVE, paravalvular infection is common and often results in partial valve dehiscence, regurgitant murmurs, CHF, or disruption of the conduction system.

DIAGNOSIS

Careful clinical, microbiologic, and echocardiographic evaluations should be pursued when febrile patients have IE predispositions, cardiac or noncardiac (e.g., stroke or splenic infarct) features of IE, or blood cultures yielding an IE-associated organism.

Duke Criteria The diagnosis of IE is established with certainty only when vegetations are examined histologically and microbiologically. Nevertheless, a common clinical approach utilizes a diagnostic schema based on clinical, laboratory, and echocardiographic findings commonly encountered in patients with IE **(Table 133-3)**. Now known as the Duke–International Society for Cardiovascular Infectious Diseases (Duke-ISCVID) Criteria for IE, the criteria were updated in 2023 to encompass new epidemiologic data and microbiologic testing and imaging methodologies; these modifications have been validated in at least four bacteremia cohorts and show sensitivity in the 80–93% range. Clinical judgment must be exercised to use the criteria effectively. A clinical diagnosis of definite IE requires two major criteria, one major and three minor criteria, or five minor criteria. IE is rejected if an alternative diagnosis is established, if there is no recurrence despite therapy for <4 days, or if surgery or autopsy after <4 days of antimicrobial therapy yields no histologic evidence of IE. Cases not classified as definite or rejected are considered possible IE when either one major and one minor criterion or three minor criteria are fulfilled. Absent extenuating circumstances, patients with definite and possible IE are treated as having IE.

Blood Cultures The Duke-ISCVID Criteria reiterate that multiple blood cultures are the gold standard for diagnosing IE. Collection of three cultures from separate venipuncture sites is still recommended but is no longer required. Further, the bacterial species considered "typical" for causing IE are defined as those whose recovery from blood has been strongly associated with IE; the pathogen list has been expanded to include all streptococcal species except *S. pneumoniae* and *S. pyogenes*, *S. lugdunensis*, *E. faecalis* regardless of the primary source, and "streptococcus-like bacteria" (e.g., *Granulicatella* spp., *Abiotrophia* spp., and *Gemella* spp.). To fulfill a major criterion, a typical organism that causes IE (e.g., those listed previously plus *S. aureus* and HACEK organisms) must be recovered in two or more blood culture sets; other organisms, considered "nontypical" must grow in three or more blood culture sets and the clinical presentation must be unexplained by an extracardiac focus of infection. In patients with intracardiac prosthetic material, CoNS, *Corynebacterium striatum*, *Corynebacterium jeikeium*, *C. acnes*, *Serratia marcescens*, *Pseudomonas aeruginosa*, nontuberculous mycobacteria (e.g., *M. chimaerae*), and *Candida* species should be considered "typical." Otherwise these organisms must be found in three or more blood culture sets to satisfy a major criterion.

In patients with suspected NVE, PVE, TAVR-PVE, or CIED-IE who have not received antibiotics during the prior 2 weeks, three two-bottle blood culture sets containing the appropriate volume of blood (10 mL per bottle) should be obtained, ideally from different venipuncture sites. If the cultures remain negative after 48–72 h, two or three additional blood culture sets should be obtained, and the laboratory should be consulted for advice regarding optimal culture techniques.

TABLE 133-3 The Modified Duke Criteria for the Clinical Diagnosis of Infective Endocarditis (IE)[a]

Major Criteria

A. Microbiologic Criteria

1. Positive blood culture

   Microorganism that commonly cause IE in 2 or more separate blood culture sets (Typical—i.e., *S. aureus*, *S. lugdunensis*, *E. faecalis*, all streptococcal species except *S. pneumoniae* and *S. pyogenes*, *Granulicatella* spp., *Abiotrophia* spp., *Gemella* spp., HACEK group organisms A*nd* in the setting of intracardiac prosthetic material, CoNS, *C. striatum*, *C. jeikeium*, *S. marcescens*, *P. aeruginosa*, *C. acnes*, nontuberculosis mycobacteria, *Candida* spp.)

   or

   Microorganisms that occasionally or rarely cause IE isolated from 3 or more separate blood culture sets (Nontypical)

2. Positive laboratory tests

   Positive polymerase chain reaction (PCR) for *Coxiella burnetti*, *Bartonella* spp., or *Tropheryma whipplei* from blood

   or

   Single blood culture growing *C. burnetti* or phase I IgG Ab titer ≥ 1:800

   or

   Indirect immunofluorescence assays (IFA) for IgM Ab and IgG Ab to *B. henselae* or *B. quintana* with IgG Ab titer ≥ 1:800

B. Imaging Criteria

1. Echocardiography or cardiac CT showing vegetation, valvular/leaflet perforation/aneurysm, abscess, pseudoaneurysm or intracardiac fistula

   or

   New valvular regurgitation (new/worsening murmur is NOT sufficient; requires echocardiographic evidence)

   or

   New prosthetic valve dehiscence/insufficiency

2. FDG-PET/CT with abnormal metabolic activity involving a native or prosthetic valve, ascending aortic graft, intracardiac device leads, or other prosthetic material

C. Surgical Criteria

Evidence of IE documented by direct inspection during heart surgery

Minor Criteria

A. Predisposition: previous history of IE, injection drug use, prosthetic valve, previous valve repair, congenital heart disease (e.g., bicuspid AV), CIED, more than mild regurgitation or stenosis, hypertrophic cardiomyopathy

B. Fever: T > 38.0°C (100.4°F)

C. Vascular phenomenon: arterial emboli, septic pulmonary infarcts, mycotic aneurysm, intracranial hemorrhage, conjunctival hemorrhage, Janeway lesions, cerebral/splenic abscesses, purulent purpura

D. Immunologic phenomena: glomerulonephritis, Osler nodes, Roth spots, positive rheumatoid factor

E. Microbiologic evidence: positive blood cultures (not meeting above criteria)

or

Positive culture, PCR, or other nucleic acid-based test for an organism consistent with IE from a non-endovascular site or single finding of a skin bacterium by PCR on a valve or wire without additional clinical or microbiological supporting evidence

F. Imaging criteria: abnormal metabolic activity on FDG-PET/CT within 3 months of implantation of a prosthetic valve, ascending aortic graft, intracardiac device lead

G. Physical exam criteria: New valvular regurgitation on auscultation

DIAGNOSIS

Definite IE: 1. Pathologic criteria (microorganisms or active endocarditis identified in a vegetation/intra-cardiac abscess from cardiac tissue, prosthetic material, arterial embolus) 2. Clinical criteria (2 major or 1 major + 3 minor or 5 minor)

Possible IE: 1 major + 1 minor or 3 minor

Rejected IE: does not meet above criteria or firm alternative dx or lack of recurrence with <4 days of antibiotics or no evidence on autopsy

Abbreviations: Ab, antibody; AV, aortic valve; CIED, cardiac implantable electronic device; CoNS, coagulase-negative staphylococci; CT, computed tomography; FDG, ^{18}F-fluorodeoxyglucose; HACEK, *Haemophilus* species, *Aggregatibacter* species, *Cardiobacterium hominis*, *Eikenella corrodens*, and *Kingella kingae*; PET, positron emission tomography.

Source: Reproduced with permission from VG Fowler Jr et al: The 2023 Duke-International Society for Cardiovascular Infectious Diseases criteria for infective endocarditis: Updating the modified Duke criteria. Clin Infect Dis. 77:518, 2023, Table 2.

Pending culture results, empirical antimicrobial therapy should be withheld initially from hemodynamically and clinically stable patients with suspected subacute IE, especially those who have received antibiotics within the preceding 2 weeks. The delay allows blood for additional cultures to be obtained without the confounding effect of empirical treatment. Patients with sepsis or deteriorating hemodynamics who may require urgent surgery should receive empirical treatment immediately after the initial three sets of blood cultures are obtained.

Non–Blood Culture Tests Non–blood culture laboratory criteria have been added to microbiologic major criteria in the Duke-ISCVID Criteria to implicate organisms that are difficult to recover by blood culture. These include polymerase chain reaction (PCR) or other nucleic acid–based techniques identifying *C. burnetii*, *Bartonella* spp., or *T. whipplei* from the blood and indirect immunofluorescence assays for IgM and IgG antibodies to *Bartonella henselae* or *B. quintana* with IgG titer ≥1:800. Next-generation (shotgun metagenomic) sequencing of pathogen DNA from serum has emerged as a novel nonculture technology capable of identifying a wide array of organisms in blood culture–negative IE. However, Duke-ISCVID Criteria recommend results from such testing that yield organisms other than *C. burnetii*, *Bartonella* spp., or *T. whipplei* be considered minor criteria at this time.

In vegetations recovered at surgery or by embolectomy, pathogens can be identified by culture and histopathologic examination with special stains. A sample of the vegetation should be collected using sterile technique and saved for molecular testing using PCR with organism-specific primers (e.g., *C. burnetii*, *Bartonella*, *T. whipplei*, *C. acnes*, *Mycoplasma hominis*) or broad-range PCR targeting 16S ribosomal RNA (or 28S rRNA, if fungi are suspected) followed by sequencing for organism identification. Histopathology may inform the selection of specific molecular tests. Molecular testing is a useful diagnostic technology when the histopathology of a vegetation is consistent with IE; however, it cannot be used to establish the viability of residual bacteria in vegetations. Additionally, molecular testing is only moderately sensitive, and thus, a negative test cannot exclude IE. When tissue is limited, molecular testing should be prioritized over culture.

Cardiac Imaging Echocardiography anatomically confirms and measures vegetations, detects intracardiac complications, and assesses cardiac function. Transthoracic echocardiography (TTE) is exceptionally specific; however, in 20% of patients, the images are inadequate. TTE fails to detect vegetations in 20–35% of patients with definite clinical IE, missing vegetations <2 mm in diameter. It is not optimal for evaluating prosthetic valves, especially TAVR with large stents, or detecting intracardiac complications. TEE detects vegetations in >90% of patients with definite IE; nevertheless, initial studies may yield false-negative results in 6–18% of IE patients, especially in TAVR-PVE. A negative TEE, when IE is likely, does not exclude the diagnosis but rather warrants repeating the study in 7–10 days. TEE is sometimes augmented by three-dimensional TEE, which can better visualize vegetations and perivalvular extension of infection.

Other imaging should be pursued when anatomic confirmation of IE is unclear, when TEE is not confirmatory or is contraindicated, and in suspected PVE. Electrocardiographic-gated multislice cardiac CT angiogram (CTA), which is less sensitive than TEE in detection of vegetations, valvular perforation, and paravalvular leakage but superior in defining pseudoaneurysm or abscess, may be definitive. Further, it can be used in lieu of preoperative cardiac catheterization to assess coronary artery patency in patients at low to intermediate risk of coronary disease. ^{18}F-Fluorodeoxyglucose positron emission tomography (FDG-PET)/CT is less sensitive than TEE or CTA in detecting intracardiac pathology in NVE or CIED-IE but provides increased sensitivity in assessing suspected PVE, including TAVR-PVE, infection of ascending aorta grafts, extracardiac complications, left ventricular assist device (LVAD) infection, and CIED pocket and lead infection. As a whole-body image, findings may modify therapy in 25% of NVE and PVE patients. However, FDG-PET/CT is costly, requires preprocedure patient preparation, can have false-positive results in patients with recent valve surgery (<3 months), and requires experienced radiographers for interpretation. Of note, findings indicative of IE on CTA and

FDG-PET/CT are major criteria and considered equivalent to echocardiography in the Duke-ISCVID schema.

In population-based studies and large series (using various diagnostic criteria), IE occurs frequently among patients who have monomicrobial bacteremia due to those gram-positive organisms that are commonly associated with IE. For example, 12–17% of patients with blood cultures growing *E. faecalis* have IE; 7% of patients with blood cultures growing non-β-hemolytic streptococci have IE; and 8–14% of patients with blood cultures growing *S. aureus* have IE. Among patients with one or more positive monomicrobial blood culture, IE risk-prediction scoring systems have been developed to identify patients who are at sufficient risk of IE to justify echocardiographic assessment **(Table 133-4)**. Because *S. aureus* bacteremia is associated with a high prevalence of IE and a resultant high risk for mortality, echocardiographic evaluation (high-quality TTE or preferably TEE) is recommended routinely. Prediction scores suggest that with *S. aureus* bacteremia, a patient with any of the features listed in Table 133-4 incurs at least a 6% risk of IE, with risk increasing when multiple features are present. Thus, when present, these findings are a strong indication for early TEE. In their absence, TTE should suffice unless other findings suggest IE. Among patients with either monomicrobial *E. faecalis* or non-β-hemolytic streptococcal bacteremia, any three of the respective listed features (Table 133-4) are associated with a significant frequency of IE. For these patients, the estimated number needed to test with TEE to detect IE is 2.4 and 3.6, respectively. While these predictive scoring systems need further evaluation and should be used with clinical judgment, they appear to have a high sensitivity and therefore a high negative predictive value, which allows identification of patients at low risk of IE where echocardiography, particularly TEE, can be omitted. An approach to echocardiographic evaluation of patients with suspected IE is illustrated in **Fig. 133-3**.

Other Studies Many studies that are not diagnostic—i.e., complete blood count, creatinine determination, liver function tests, chest radiography, and electrocardiography—are important in the management of patients with IE. The erythrocyte sedimentation rate, C-reactive protein level, rheumatoid factor, and circulating immune complex titer are commonly increased in IE (Table 133-2). Brain MRI/magnetic resonance angiography (MRA) should be obtained in patients with neurologic signs or symptoms, including unusual headache, to assess for emboli, hemorrhage, or mycotic aneurysms. The findings can support the IE diagnosis as well as provide evidence requiring changes in planned surgical treatment. Patients with recalcitrant back pain or focal spine tenderness should undergo spine MRI targeted to the appropriate level based on symptoms to assess for osteomyelitis and epidural abscess. Contrast-enhanced whole-body tomography to detect silent emboli in patients without localizing symptoms is not likely to enhance diagnostic accuracy and is associated with significant risk of kidney injury due to contrast media exposure; thus, it should not be performed routinely.

TABLE 133-4 Features Guiding the Need for Echocardiographic Assessment in Patients with Selected Monomicrobial Bacteremia

BLOOD CULTURE ISOLATE		
***S. AUREUS*[a]**	***E. FAECALIS*[b]**	**NON-β-HEMOLYTIC STREPTOCOCCI[c]**
Intracardiac device	Symptoms ≥7 days	Symptoms ≥7 days
Preexisting valve disease (including prior endocarditis or valve prosthesis)	Emboli	>2 positive cultures
Injection drug use	≥2 positive cultures	One species: *S. gallolyticus*, *S. sanguinis*, *S. mutans* (not *S. anginosus*)
Cerebral/peripheral emboli	Unknown origin (no focus)	Preexisting valve disease (including prior endocarditis or valve prosthesis)
Meningitis	Heart murmur	Heart murmur
Persistent bacteremia (≥72 h)	Preexisting valve disease (including prior endocarditis or valve prosthesis)	Community acquisition
Vertebral osteomyelitis		
Community acquisition		
Nonnosocomial health care associated (including hemodialysis)		

Source: [a]S Tubiana et al: J Infect 72:544, 2016 and A Showler et al: JACC Cardiovasc Imaging 8:924, 2015. [b]A Berge et al: Infection 47:45, 2019. [c]T Sunnerhagen et al: Clin Infect Dis 66:693, 2018.

TREATMENT

Infective Endocarditis

ANTIMICROBIAL THERAPY

To cure IE, all bacteria in the vegetation must be killed. This is difficult because local host defenses are deficient and because the bacteria are largely nongrowing and metabolically inactive and thus are less easily killed by antibiotics. Consequently, therapy must be prolonged. Antibiotics are generally given parenterally to achieve serum concentrations that, through passive diffusion, result in effective concentrations in the depths of the vegetation. The decision to initiate treatment empirically must balance the need to establish a microbiologic diagnosis against the potential disease progression or the need to control infection prior to urgent surgery (see “Blood Cultures,” above). Infection at other sites (such as the meninges), allergies, end-organ dysfunction, interactions with concomitantly administered medications, and risks of adverse events must be considered in the selection of therapy.

The regimens recommended for the treatment of PVE (except that caused by staphylococci), although given for several weeks longer, are similar to those used to treat NVE **(Table 133-5)**. Recommended antibiotic dosing and duration of therapy, which is measured from the time blood cultures become negative, should be followed unless alterations are required by end-organ dysfunction or adverse events.

Organism-Specific Therapies • **Streptococci** The recommended therapies for streptococcal IE are based on the minimal inhibitory concentration (MIC) of penicillin for the causative isolate (Table 133-5). The 2-week penicillin/gentamicin and ceftriaxone/gentamicin regimens should not be used to treat NVE complicated by cardiac or extracardiac abscess or PVE. Caution should be exercised in considering aminoglycoside-containing regimens in patients at increased risk for aminoglycoside toxicity (renal or eighth cranial nerve). The regimens recommended for relatively penicillin-resistant streptococci are advocated for treatment of group B, C, or G streptococcal IE. *Granulicatella*, *Abiotrophia*, and *Gemella* species are treated with the regimens for moderately penicillin-resistant streptococci, as is PVE caused by these organisms or by streptococci with a penicillin MIC of >0.1 μg/mL (Table 133-5).

Enterococci Enterococci are resistant to oxacillin, nafcillin, and the cephalosporins and are only inhibited—not killed—by the cell wall–active agents penicillin, ampicillin, teicoplanin (not available in the United States), and vancomycin, with ampicillin and penicillin being preferred when susceptible. Enterococci are killed by the synergistic interaction of these cell wall–active antibiotics combined with gentamicin, unless the isolate exhibits high-level resistance to gentamicin, defined as growth of the isolate in the presence of gentamicin at ≥500 μg/mL. Bactericidal synergy with other aminoglycosides—tobramycin, netilmicin, kanamycin, and amikacin—is unpredictable even in the absence of high-level resistance; thus, they are not used to treat enterococcal IE.

Although the dose of gentamicin used to achieve bactericidal synergy in treating enterococcal IE is smaller than that used in standard therapy, nephrotoxicity is not uncommon during treatment

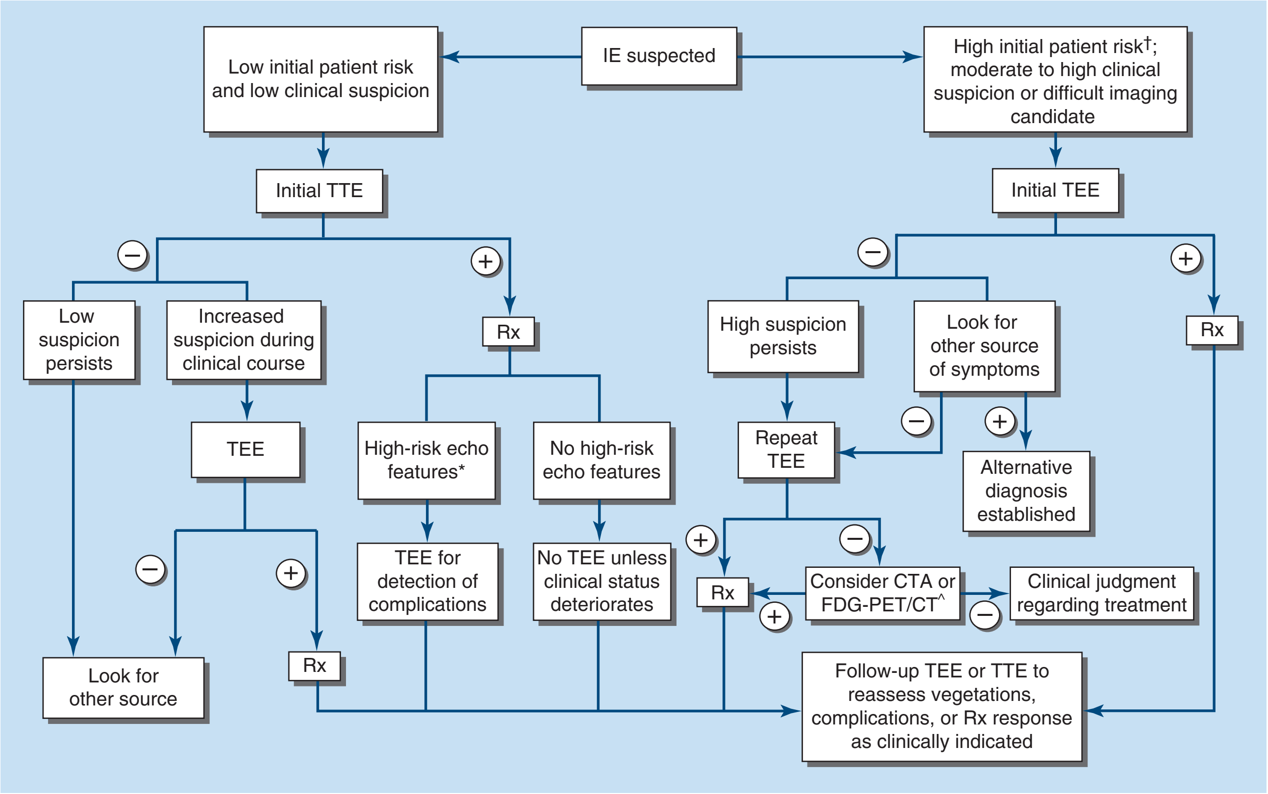
IE suspected
Low initial patient risk and low clinical suspicion
High initial patient risk†; moderate to high clinical suspicion or difficult imaging candidate
Initial TTE
Initial TEE
−
+
Low suspicion persists
Increased suspicion during clinical course
Rx
High suspicion persists
Look for other source of symptoms
Rx
TEE
High-risk echo features*
No high-risk echo features
Repeat TEE
Alternative diagnosis established
TEE for detection of complications
No TEE unless clinical status deteriorates
Rx
Consider CTA or FDG-PET/CT^
Clinical judgment regarding treatment
Look for other source
Rx
Follow-up TEE or TTE to reassess vegetations, complications, or Rx response as clinically indicated

FIGURE 133-3 The diagnostic use of transesophageal and transthoracic echocardiography (TEE and TTE, respectively). †High initial patient risk for infective endocarditis (IE) or evidence of intracardiac complications (new regurgitant murmur, new electrocardiographic conduction changes, or congestive heart failure). *High-risk echocardiographic features include large vegetations, valve insufficiency, paravalvular infection, or ventricular dysfunction. Rx indicates initiation of antibiotic therapy. CTA, electrocardiogram-gated cardiac computed tomography (CT) angiogram; FDG-PET/CT, fluorodeoxyglucose-positron emission tomography CT. ^See text for discussion of these modalities. *(Reproduced with permission from AS Bayer: Diagnosis and management of infective endocarditis and its complications. Circulation 98:2936, 1998, Figure 1.)*

lasting 4–6 weeks. High concentrations of ampicillin plus ceftriaxone or cefotaxime, by expanded binding of penicillin-binding proteins, also kill *E. faecalis* in vitro and in animal models of IE. Nonrandomized comparative studies suggest that high-dose regimens using ampicillin-ceftriaxone appear comparable and less nephrotoxic than penicillin or ampicillin plus gentamicin for treatment of *E. faecalis* (but not *E. faecium*) IE and may also provide effective treatment when strains possess high-level gentamicin resistance. This regimen has been used increasingly to address not only high-level gentamicin-resistant strains but also to minimize nephrotoxicity. Alternatively, if there is a contraindication to an ampicillin-ceftriaxone regimen, shorter 2- to 3-week courses of synergistic gentamicin can be considered. The combinations of vancomycin (or teicoplanin) or gentamicin with ceftriaxone are not bactericidal for *E. faecalis* and are not recommended for treatment of enterococcal IE. If a combination regimen cannot be used due to resistance or toxicity, an 8- to 12-week course of a single cell wall–active agent can be considered, although the patient should be followed carefully for evidence of failure.

Treatment of IE caused by *E. faecium*, which is generally more antibiotic resistant than *E. faecalis* and may be vancomycin resistant, is not well established. Successful treatment of IE caused by vancomycin-resistant enterococci with high-dose daptomycin (10–12 mg/kg IV once daily), often in combination with ampicillin or other β-lactams, has been reported. If the isolate susceptibility allows treatment with penicillin or ampicillin plus gentamicin, this is preferred. These cases should be managed in conjunction with an infectious disease consultant.

Staphylococci Management of *S. aureus* bacteremia and IE in conjunction with infectious disease consultants has been associated with improved outcomes and is recommended. Treatment of staphylococcal IE (Table 133-5) is based on the presence of a prosthetic valve or foreign device, the native valve(s) involved (right vs left side), and the antibiotic susceptibility of the isolate. Penicillin resistance and, except in specific countries, methicillin resistance are widespread among staphylococci. Thus, empirical therapy for possible staphylococcal IE should use a regimen effective against methicillin-resistant organisms. Therapy should be revised to an antistaphylococcal penicillin if the isolate is susceptible to methicillin. Cefazolin is generally considered an alternative β-lactam agent for the treatment of methicillin-susceptible *S. aureus* (MSSA) IE. Ease of administration and reduced adverse events compared to treatment with an antistaphylococcal penicillin have prompted use of cefazolin as a primary agent in this setting. Concerns, however, have been raised about inactivation of cefazolin by type A and C staphylococcal β-lactamases (these do not hydrolyze antistaphylococcal penicillins), resulting in treatment failure in high-inoculum infections. Initiating treatment with an antistaphylococcal penicillin until there is source control and a reduced inoculum and then transitioning to cefazolin should be considered. The addition of gentamicin to a β-lactam antibiotic or vancomycin to enhance therapy for left-sided NVE has not improved survival rates and is associated with nephrotoxicity. Guidelines do not recommend the routine addition of gentamicin, fusidic acid, rifampin, or daptomycin to regimens for MSSA NVE.

For treatment of NVE due to methicillin-resistant *S. aureus* (MRSA), vancomycin, dosed to achieve trough concentrations of 15 μg/mL (or an area under the time-concentration curve/broth microdilution MIC ratio [AUC:MIC] >400 achieved with the assistance of a pharmacist), is recommended, with the caveat that high vancomycin trough concentrations may be associated with nephrotoxicity. Although resistance to vancomycin among staphylococci is rare, reduced vancomycin susceptibility among MRSA strains is increasingly encountered. Isolates with a vancomycin MIC of 4–8 μg/mL have intermediate susceptibility and are referred to as

TABLE 133-5 Antibiotic Treatment for Infective Endocarditis Caused by Common Organisms[a]

ORGANISM(S)	DRUG (DOSE, DURATION)	COMMENTS
Streptococci		**For PVE 6-week regimens are preferred.**
Penicillin-susceptible streptococci, *S. gallolyticus* (MIC ≤0.12 μg/mL)	• Penicillin G (2–3 mU IV q4h for 4 weeks)	Can use ampicillin or amoxicillin (2 g IV q4h) if penicillin is unavailable.
	• Ceftriaxone (2 g once daily for 4 weeks)	Can use ceftriaxone in patients with nonimmediate penicillin allergy.
	• Vancomycin[b] (15 mg/kg IV q12h for 4 weeks)	Use vancomycin for patients with immediate (urticarial) or severe penicillin allergy. Obtain allergy consultation for further evaluation including role of β-lactam desensitization.
	• Penicillin G (2–3 mU IV q4h) *or* ceftriaxone (2 g IV once daily) for 2 weeks *plus* Gentamicin[c] (3 mg/kg daily IV or IM, as a single dose[d] for 2 weeks)	Can use ampicillin or amoxicillin (2 g IV q4h) if penicillin is unavailable. Avoid 2-week regimen when risk of aminoglycoside toxicity is increased and in prosthetic-valve or complicated endocarditis. Penicillin G at a dose of 4 mU IV q4h or ceftriaxone 2 g once daily for 6 weeks both with or without gentamicin during the initial 2 weeks preferred for PVE caused by streptococci with penicillin MICs ≤0.12 μg/mL.
Relatively penicillin-resistant streptococci, *S. gallolyticus* (MIC >0.12 μg/mL and <0.5 μg/mL[e])	• Penicillin G (4 mU IV q4h) *or* ceftriaxone (2 g IV daily) for 4 weeks *plus* Gentamicin[c] (3 mg/kg daily IV or IM, as a single dose[d] for 2 weeks)	Can use ampicillin or amoxicillin (2 g IV q4h) if penicillin is unavailable. Penicillin G at a dose of 4 mU IV q4h or ceftriaxone 2 g once daily for 6 weeks both with gentamicin during the initial 2 weeks preferred for PVE caused by streptococci with penicillin MICs >0.12 μg/mL.
	• Vancomycin[b] as noted above for 4 weeks	Use vancomycin for patients with immediate (urticarial) or severe penicillin allergy. Obtain allergy consultation for further evaluation including role of β-lactam desensitization. Ceftriaxone alone or with gentamicin can be used in patients with nonimmediate β-lactam allergy.
Moderately penicillin-resistant streptococci (MIC, ≥0.5 μg/mL and <8 μg/mL); *Granulicatella*, *Abiotrophia*, or *Gemella* spp.	• Penicillin G (4–5 mU IV q4h) *or* ceftriaxone (2 g IV daily) for 6 weeks *plus* Gentamicin[c] (3 mg/kg daily IV or IM in 2–3 equally divided doses for 6 weeks)	Can use ampicillin or amoxicillin (2 g IV q4h) if penicillin is unavailable.
	• Vancomycin[b] as noted above for 6 weeks	Regimen is preferred by some.
Enterococci[e]		**For PVE, 6-week regimens are preferred.**
	• Ampicillin (2 g IV q4h) *plus* ceftriaxone (2 g IV q12h), both for 6 weeks	Use for *E. faecalis* isolates with or without high-level resistance to gentamicin or for patients at risk for aminoglycoside nephrotoxicity.
	• Penicillin G (4–5 mU IV q4h) for 4–6 weeks *plus* gentamicin[c] (3 mg/kg daily IV or 1 mg/kg IV q8h) for 2–6 weeks	Can treat NVE for 4 weeks if symptoms last <3 months. Treat NVE for 6 weeks if >3 months of symptoms. Can abbreviate gentamicin course in some patients (see text).
	• Ampicillin (2 g IV q4h) for 4–6 weeks *plus* gentamicin[c] (3 mg/kg daily IV or 1 mg/kg IV q8h) for 2–6 weeks	Can use IV amoxicillin in lieu of ampicillin (same dose). Can abbreviate gentamicin course in some patients (see text).
	• Vancomycin[b] (15 mg/kg IV q12h) for 6 weeks *plus* gentamicin[c] (3 mg/kg daily IV or 1 mg/kg IV q8h) for 2–6 weeks	Use vancomycin plus gentamicin only for penicillin-allergic patients (preferable to desensitize to penicillin if immediate [urticarial] allergy; consult allergy) and for isolates resistant to penicillin/ampicillin.
Staphylococci (*S. aureus* and coagulase-negative)		
MSSA infecting native valves (no foreign devices) including complicated right-sided and left-sided endocarditis.	• Nafcillin, oxacillin, *or* flucloxacillin (2 g IV q4h for 4–6 weeks)	Addition of gentamicin is not recommended. For uncomplicated right-sided endocarditis, a 2-week course may be effective (see text).
	• Cefazolin (2 g IV q8h for 4–6 weeks)	Can use cefazolin regimen for patients with nonimmediate penicillin allergy; see text regarding cefazolin vs antistaphylococcal penicillin as primary therapy. Addition of gentamicin not recommended.
	• Vancomycin[b] (15 mg/kg IV q12h for 4–6 weeks)	Only use vancomycin for patients with immediate (urticarial) or severe penicillin allergy until allergy consultation can be obtained for β-lactam desensitization evaluation; addition of gentamicin not recommended.
MRSA infecting native valves (no foreign devices)	• Vancomycin[b] (15 mg/kg IV q8–12h) or daptomycin (8–10 mg/kg daily) for 4–6 weeks	No role for routine use of rifampin (see text). For daptomycin treatment, see text.
MSSA infecting prosthetic valves	• Nafcillin, oxacillin, *or* flucloxacillin (2 g IV q4h for 6–8 weeks) *plus* Gentamicin[c] (1 mg/kg IM or IV q8h for 2 weeks) *plus* • Rifampin[f] (300 mg PO q8h for 6–8 weeks)	Use gentamicin during initial 2 weeks; determine gentamicin susceptibility and await blood culture clearance before initiating rifampin (see text); if patient is highly allergic to penicillin, use regimen for MRSA and obtain allergy consultation; if β-lactam allergy is of the minor nonimmediate type, cefazolin can be substituted for oxacillin, nafcillin, or flucloxacillin.
MRSA infecting prosthetic valves	• Vancomycin[b] (15 mg/kg IV q12h for 6–8 weeks) *plus* Gentamicin[c] (1 mg/kg IM or IV q8h for 2 weeks) *plus* • Rifampin[f] (300 mg PO q8h for 6–8 weeks)	Use gentamicin during initial 2 weeks; determine gentamicin susceptibility and await blood culture clearance before initiating rifampin (see text). Daptomycin (8–10 mg/kg daily) is an alternative to vancomycin.

(*Continued*)

TABLE 133-5 Antibiotic Treatment for Infective Endocarditis Caused by Common Organisms[a] (Continued)

ORGANISM(S)	DRUG (DOSE, DURATION)	COMMENTS
HACEK organisms		**For PVE, 6-week regimens are preferred.**
	• Ceftriaxone (2 g once daily IV for 4 weeks)	
	• Ampicillin/sulbactam (3 g IV q6h for 4 weeks)	Use ampicillin alone (2 g IV q4h) only if β-lactamase production can be excluded. If the isolate is susceptible, ciprofloxacin (500 mg by mouth q12h or 400 mg IV q12h) can be used.
Coxiella burnetii		
	• Doxycycline (100 mg PO q12h) *plus* hydroxychloroquine (200 mg PO q8h), both for at least 18 (native valve) or 24 (prosthetic valve) months	Follow serology to monitor response during treatment (antiphase I IgG and IgA decreased fourfold and IgM antiphase II negative) and thereafter for relapse.
***Bartonella* spp.**		
	• Doxycycline (100 mg q12h PO) for 6 weeks *plus* Gentamicin (1 mg/kg IV q8h for 2 weeks)	If doxycycline is not tolerated, use azithromycin (500 mg PO daily). Some experts recommend that doxycycline be continued for 3–6 months unless all infection is resected surgically.

[a]Regimens adapted from the guidelines of the American Heart Association and the European Society of Cardiology (ESC). Doses of gentamicin, vancomycin, and daptomycin must be adjusted for reduced renal function. Ideal body weight is used to calculate doses of gentamicin and daptomycin per kilogram (men = 50 kg + 2.3 kg per inch over 5 feet; women = 45.5 kg + 2.3 kg per inch over 5 feet). [b]Vancomycin dose is based on actual body weight. Adjust for trough level of 10–15 μg/mL for streptococcal and enterococcal infections and 15–20 μg/mL for staphylococcal infections (see text). [c]Target peak and trough serum concentrations of divided-dose gentamicin 1 h after a 20- to 30-min infusion or IM injection are ~3.5 μg/mL and ≤1 μg/mL, respectively. [d]Netilmicin (4 mg/kg qd, as a single dose) can be used in lieu of gentamicin for streptococcal infection only. [e]Antimicrobial susceptibility must be evaluated; see text. [f]Rifampin increases warfarin and dicumarol requirements for anticoagulation.

Abbreviations: HACEK, *H*aemophilus species, *A*ggregatibacter species, *C*ardiobacterium hominis, *E*ikenella corrodens, and *K*ingella kingae; MIC, minimal inhibitory concentration; MRSA, methicillin-resistant *S. aureus*; MSSA, methicillin-susceptible *S. aureus*; NVE, native-valve endocarditis; PVE, prosthetic-valve endocarditis.

vancomycin-intermediate S. aureus (VISA). Isolates with a vancomycin MIC of 2 μg/mL may harbor subpopulations with higher MICs. These isolates, called *heteroresistant VISA* (hVISA), are not detectable by routine susceptibility testing and yet may impair vancomycin effectiveness.

Daptomycin has provided effective alternative treatment for left-sided NVE caused by documented daptomycin-susceptible MRSA and is associated with less nephrotoxicity than vancomycin (although it is associated with myositis and, rarely, eosinophilic pneumonitis). Although it is U.S. Food and Drug Administration (FDA) approved only for right-sided IE at a dose of 6 mg/kg daily, most recommend doses of 8–10 mg/kg daily for treatment of left-sided IE. Because receipt of vancomycin has been associated with emergence of daptomycin nonsusceptibility, caution should be exercised when switching from vancomycin to daptomycin monotherapy in a patient who has received vancomycin without clearing blood cultures. Daptomycin activity against MRSA—even against some daptomycin-nonsusceptible isolates—is enhanced when given in combination with nafcillin or ceftaroline. Case series suggest that high-dose daptomycin combined with nafcillin or ceftaroline, ceftaroline combined with trimethoprim/sulfamethoxazole, or ceftaroline alone (600 mg IV q8h) may be effective salvage treatment for vancomycin-unresponsive MRSA IE and should be considered in patients with sepsis, in those with multifocal sites of infection, or when bacteremia persists for 4 or more days. Eradicable sources of bacteremia should always be addressed; failure of source control is a very common reason for persistent MRSA bacteremia.

MSSA IE that is uncomplicated and limited to the tricuspid or pulmonic valve can often be treated with 2 weeks of oxacillin or nafcillin (but not vancomycin). Prolonged fever (≥5 days) during therapy or multiple septic pulmonary emboli mandate standard-duration therapy. Right-sided MRSA IE is treated for at least 4 weeks with vancomycin or daptomycin (6 mg/kg daily); 2-week courses of therapy are suboptimal. Left-sided staphylococcal IE is treated with a minimum of 6 weeks of therapy.

Staphylococcal PVE is treated with a multidrug regimen for 6–8 weeks (Table 133-5). To achieve long-term bacterial eradication, rifampin, which kills staphylococci embedded in biofilm adherent to foreign material, is considered an essential component of this regimen. Rifampin resistance can emerge during therapy. To prevent emergence of resistance, administration of rifampin should be delayed until initial therapy with two agents (gentamicin plus an antistaphylococcal penicillin or vancomycin selected on the basis of susceptibility testing) has eradicated bacteremia (reduced the inoculum). The isolate's susceptibility to gentamicin or an alternative agent as well as to rifampin should be established before rifampin treatment is begun. The impact of adding rifampin and gentamicin on survival of patients with staphylococcal PVE has not been demonstrated convincingly; thus, it is reasonable to discontinue these agents if patients are experiencing toxicity or drug interactions.

Other Organisms In the absence of meningitis, IE caused by *Streptococcus pneumoniae* isolates with a penicillin MIC of ≤4 μg/mL can be treated with IV penicillin (4 million units IV every 4 h), ceftriaxone (2 g once daily), or vancomycin. Ceftriaxone or vancomycin is preferred for pneumococcal strains with a penicillin MIC of ≥2 μg/mL. If meningitis is suspected, treatment with vancomycin plus ceftriaxone—at the doses advised for meningitis—should be initiated until susceptibility results are known. Definitive therapy should then be selected on the basis of meningitis breakpoints (penicillin MIC, 0.06 μg/mL; ceftriaxone MIC, 0.5 μg/mL). Pneumococcal NVE is treated for 4 weeks and pneumococcal PVE for 6 weeks. *P. aeruginosa* IE is treated with an antipseudomonal cephalosporin and traditionally high doses of tobramycin, and IE caused by Enterobacterales is treated with a β-lactam antibiotic, also traditionally, plus an aminoglycoside, although data substantiating a need for combination therapy for these organisms are limited. Therapy for *Candida* IE consists of a lipid formulation of amphotericin B (5 mg/kg IV daily) with or without flucytosine (25 mg/kg PO q6h) (if using flucytosine, monitor renal function, flucytosine levels, and bone marrow function). Alternatively, a high-dose echinocandin regimen can be used. If there is valve dysfunction or PVE, early surgery is advised, as is long-term (if not indefinite) oral azole suppression. Absent valve dysfunction, medical treatment with long-term oral azole suppression may achieve results comparable to surgical treatment.

Empirical Therapy and Treatment for Culture-Negative IE In designing therapy to be administered before culture results are known or when cultures are truly negative, clinical clues to etiology (e.g., acute vs subacute presentation, NVE, early or late PVE, the patient's predispositions) as well as epidemiologic clues (e.g., region of residence, animal exposure) must be considered. Thus, empirical therapy for acute IE should cover MRSA and in a PWID or for health care–associated NVE potentially antibiotic-resistant gram-negative bacilli. Treatment with vancomycin plus gentamicin or cefepime, initiated immediately after blood cultures are obtained, covers these organisms as well as many others. For empirical treatment of NVE with a subacute presentation, vancomycin plus ceftriaxone is reasonable. For blood culture–pending PVE, vancomycin, gentamicin, and cefepime should be used if the prosthetic valve has been in place for ≤1 year. Empirical therapy for late PVE (valve in place for >1 year) is similar to that for culture-negative NVE. Therapy is revised once a pathogen has been identified.

In the treatment of blood culture–negative episodes, marantic endocarditis and the antiphospholipid antibody syndrome must be considered. In the absence of prior antibiotic therapy, it is unlikely that infection due to *S. aureus*, CoNS, enterococci, or Enterobacterales will present with negative blood cultures; thus, recommended empirical therapy targets fastidious streptococci, *Abiotrophia*, *Gemella*, *Granulicatella*, the HACEK group, and *Bartonella* species. Pending the availability of diagnostic data, blood culture–negative subacute NVE is treated with vancomycin plus ampicillin-sulbactam (12 g every 24 h) or ceftriaxone; doxycycline (100 mg twice daily) is added for enhanced *Bartonella* coverage. If cultures are negative because of prior antibiotic administration, pathogens likely to be inhibited by the specific prior therapy should be considered.

TAVR-PVE The vast majority of these patients are treated medically with classic PVE antibiotic regimens for the given pathogen. Selection of empirical therapy pending blood culture results similarly parallels that for classic PVE but with recognition that enterococcal infection occurs with increased frequency.

CIED-IE Antimicrobial therapy for CIED-IE (as well as for generator pocket and lead infection) is adjunctive to complete removal of the device. The antimicrobial selected is based on the causative organism and should be used as recommended for NVE (Table 133-5). Bacteremic CIED infection may be complicated by coincident left-sided NVE, PVE, or remote-site infection (e.g., osteomyelitis), which may require modification of antimicrobial therapy. A 4- to 6-week course of IE-targeted therapy is recommended for patients with CIED-IE and for those with bacteremia that continues after device removal. Generator pocket infection without bacteremia is treated with a 10- to 14-day course, some of which can be given orally. In the absence of another source, *S. aureus* bacteremia (and persistent CoNS bacteremia) is likely indicative of CIED-IE or valvular IE and should be investigated and managed accordingly. However, not all bloodstream infections in these patients indicate IE. If evidence suggesting CIED-IE is lacking, bloodstream infection due to gram-negative bacilli, streptococci, and enterococci species may not indicate CIED-IE and can be treated with antimicrobial therapy for the alternative diagnosis. However, in these patients, relapse after antimicrobial therapy increases the likelihood of CIED-IE and warrants treatment as such. Attempted salvage of an infected CIED with antibiotics alone and long-term suppressive therapy is usually unsuccessful and should be reserved for patients whose devices cannot be removed or who decline removal. Careful follow-up is required.

Partial Oral Antibiotic Treatment of IE Recent studies have examined the use of oral antibiotics to complete therapy in patients who have received an initial course of intravenous treatment (with or without cardiac surgery). A noninferiority, multicenter, randomized study found mortality among patients with left-sided IE caused by streptococci, enterococci, and staphylococci who received partial oral treatment comparable to that of patients who were treated intravenously for the full course of therapy (6.5% and 7.5%, respectively). Four hundred clinically stable patients (20% of the population screened) who had received at least 10 days of parenteral therapy (or at least 7 days after surgery) were enrolled. IE was caused by streptococci (49%), enterococci (24%), and MSSA (22%); no patients had MRSA. Of note, the median duration of intravenous treatment before the switch to oral agents was 16 days (interquartile range, 13–23 days); thus, some patients may have been effectively cured by intravenous therapy with or without surgery prior to transition to oral therapy. The results in this highly selected and monitored cohort with relatively small numbers of patients with enterococcal and *S. aureus* IE may not be generalizable to most patients with IE. If oral therapy is considered in patients who meet the criteria in the studied population, the decision on agents and duration should be made in consultation with infectious diseases.

Use of Long-Acting Lipoglycopeptide Agents Dalbavancin, a lipoglycopeptide with a long half-life allowing for infrequent dosing, is FDA approved for treatment of skin and soft tissue infections. Only limited data are available supporting its use in the treatment of IE, particularly due to *S. aureus*, with most being case series of patients who received dalbavancin upon discharge from the hospital after having received several weeks of active intravenous therapy. Consideration can be given to using this agent in patients who are clinically stable with negative blood cultures and in patients with complicated IE who are not candidates to continue hospitalization or to transition to outpatient parenteral therapy, in consultation with infectious diseases.

Outpatient Parenteral Antimicrobial Therapy Fully compliant, clinically stable patients who are no longer bacteremic, are not febrile, and have no clinical or echocardiographic findings that suggest an impending complication may complete IV therapy as outpatients. Careful follow-up and a stable residential setting are necessary, as are predictable IV access and use of antimicrobial agents that are stable in solution and less frequently associated with severe adverse effects. Recommended regimens should not be compromised to accommodate outpatient therapy.

Monitoring Antimicrobial Therapy Antibiotic-related adverse events occur in 25–40% of IE patients and commonly arise after several weeks of therapy. Blood tests to detect renal, hepatic, and hematologic toxicity should be performed periodically. Serum concentrations of aminoglycosides and vancomycin should be monitored and doses adjusted to optimize treatment and minimize toxicity.

Control of peripheral sites of infection—source control—should be addressed promptly. Blood cultures should be repeated daily until sterile in patients with IE due to *S. aureus* or difficult-to-treat organisms and rechecked if there is recrudescent fever. Blood cultures become sterile after 2 days of appropriate therapy when infection is caused by viridans streptococci, *E. faecalis*, or HACEK organisms. In MSSA IE, β-lactam therapy results in sterile cultures in 3–5 days, whereas in MRSA IE, the duration of bacteremia is often longer with vancomycin or daptomycin treatment. MRSA bacteremia persisting despite an appropriate dosage of vancomycin or daptomycin may indicate emergence of reduced susceptibility in the infecting strain and point to a need for alternative therapy. When fever persists for 7 days despite appropriate antibiotic therapy, patients should be evaluated further for paravalvular abscess, extracardiac abscesses (spleen, kidney), or complications (embolic events). Recrudescent fever raises the possibility of these complications but also of drug reactions or complications of hospitalization. It is advisable to obtain follow-up echocardiography after the completion of therapy to assess valvular function. Of note, vegetations become smaller with effective therapy; however, 3 months after cure, 50% are unchanged, and 25% each are slightly larger or smaller.

Antithrombotic Therapy Because patients with IE are at risk for hemorrhagic transformation of embolic strokes and for intracerebral hemorrhage from septic arteritis or ruptured mycotic aneurysms, initiation of antithrombotic (anticoagulant or antiplatelet) therapy requires careful consideration of the risks and benefits. Antithrombotic therapy can render such bleeding catastrophic. Neither anticoagulant nor antiplatelet therapy reduces the risk of emboli in patients with NVE, and thus, such treatment is not

TABLE 133-6 Indications for Cardiac Surgical Treatment in Patients with Endocarditis

Surgery Required for Optimal Outcome
Native-valve or prosthetic-valve endocarditis
Moderate or severe congestive heart failure or shock due to valve dysfunction
Paravalvular extension of infection with abscess, fistula, or heart block
Persistent bacteremia without an extracardiac cause despite 7–10 days of optimal antimicrobial therapy
Lack of effective antimicrobial therapy (e.g., fungal [see text regarding *Candida* spp.], *Brucella*, multidrug-resistant gram-negative bacillary endocarditis)
Prosthetic-valve endocarditis
Partially dehisced unstable prosthetic valve
Surgery to Be Strongly Considered for Improved Outcome[a]
Prosthetic-valve endocarditis
S. aureus infection with intracardiac complications
Relapse after optimal antimicrobial therapy
Native-valve endocarditis
Large (>10-mm) hypermobile vegetation, particularly with prior systemic embolus and significant valve dysfunction[b]
Very large (>30-mm) vegetation
Persistent unexplained fever (≥10 days) in blood culture–negative endocarditis
Poorly responsive or relapsed endocarditis due to highly antibiotic-resistant enterococci or gram-negative bacilli

[a] Carefully consider surgery. Multiple findings are often combined to justify surgery.
[b] In the group with an estimated low cardiac-surgery mortality risk (see text).

indicated for that purpose. However, patients with IE may have coexisting conditions wherein anticoagulation is indicated. Thus, in the absence of a contraindication (i.e., no clinical or imaging evidence of a recent large embolic stroke, intracerebral hemorrhage, or mycotic aneurysm), anticoagulant therapy is given to patients who have a mechanical prosthetic valve, atrial fibrillation with either mitral stenosis or a CHA_2DS_2-VASc score ≥2, or deep-vein thrombophlebitis. Most experts use unfractionated or low-molecular-weight heparin for ease of reversal. Anticoagulant therapy should be reversed, at least temporarily, in most patients who have had an acute ischemic stroke or an intracerebral hemorrhage.

SURGICAL TREATMENT

The indications for cardiac surgical treatment of IE **(Table 133-6)** have been derived from observational studies and expert opinion. The strength of specific indications varies; thus, the risks and benefits as well as the timing of surgery must be individualized **(Table 133-7)**. These are best weighed by a team that includes cardiologists, cardiac surgeons, infectious disease physicians, and neurologists if there have been neurologic complications. Between 25% and 40% of patients with left-sided IE undergo cardiac surgery during active infection, with slightly higher surgery rates for PVE than NVE. The benefit of surgery has been assessed primarily in retrospective studies comparing populations of medically and surgically treated patients matched for the necessity of surgery, with adjustments for predictors of death (comorbidities) and the timing of surgical intervention (a correction for survival bias). Although study results vary, surgery for NVE based on current indications appears to convey a significant survival benefit (27–55%), which is greatest among those with the most pressing indications. The survival benefit becomes more apparent after ≥6 months. The effect of surgery for PVE is more nuanced, with survival benefits accruing largely to those with intracardiac complications. Of note, surgery itself carries mortality risks that may offset survival benefits in patients with lesser indications. Among patients with TAVR-PVE, 50–80% are reported to have an indication for surgical intervention—yet because of high pre-TAVR estimated operative mortality, <15% undergo surgery. Some patients with significant aortic regurgitation after medical cure of infection have undergone valve-in-valve redo-TAVR.

Indications • **Congestive Heart Failure** Moderate to severe refractory CHF caused by new or worsening valve dysfunction or intracardiac fistulae is the major indication for cardiac surgery. Surgery can relieve functional stenosis due to large vegetations or restore competence to damaged regurgitant valves by repair or replacement. At 6–12 months of follow-up of patients with left-sided NVE or PVE and moderate to severe CHF due to valve

TABLE 133-7 Timing of Cardiac Surgical Intervention in Patients with Endocarditis

	INDICATION FOR SURGICAL INTERVENTION	
TIMING	STRONG SUPPORTING EVIDENCE	CONFLICTING EVIDENCE, BUT MAJORITY OF OPINIONS FAVOR SURGERY
Emergent (same day)	Valve dysfunction with pulmonary edema or cardiogenic shock Acute aortic regurgitation plus preclosure of mitral valve Sinus of Valsalva abscess ruptured into right heart Rupture into pericardial sac	
Urgent (within 1–2 days)	Valve obstruction by vegetation Unstable (dehisced) prosthesis Acute aortic or mitral regurgitation with heart failure (New York Heart Association class III or IV) Septal perforation Paravalvular extension of infection with or without new electrocardiographic conduction system changes Lack of effective antibiotic therapy	Vegetation diameter >10 mm plus severe but not urgent aortic or mitral valve dysfunction[a] Major embolus plus persisting large vegetation (>10 mm) Mobile vegetation >30 mm
Elective (earlier usually preferred)	Progressive paravalvular prosthetic regurgitation Valve dysfunction plus persisting infection after ≥7–10 days of antimicrobial therapy Fungal (mold) endocarditis	Staphylococcal prosthetic-valve endocarditis with intracardiac complications Early prosthetic-valve endocarditis (≤2 months after valve surgery) *Candida* spp. endocarditis (see text) Antibiotic-resistant organisms

[a] Supported by a single-institution randomized trial showing benefit from early surgery. Implementation requires clinical judgment. If surgery is elected, it must be done early (see text).

Source: Reproduced with permission from L Olaison, G Pettersson: Current best practices and guidelines: Indications for surgical intervention in infective endocarditis. Infect Dis Clin North Am 16:453, 2002.

dysfunction, survival is significantly improved among those treated surgically compared with those treated medically. The survival benefit with surgery is inversely related to the severity of preoperative CHF; thus, surgery should not be delayed in the face of deteriorating hemodynamics.

Paravalvular Infection This complication, which is most common with aortic valve infection, occurs in 10–15% of patients with NVE and in 45–60% of those with PVE. It is suggested clinically by persistent unexplained fever during appropriate therapy, new electrocardiographic conduction disturbances, or pericarditis. TEE with color Doppler is the test of choice to detect paravalvular abscesses (sensitivity, ≥85%). Occasionally, three-dimensional TEE, ECG-gated CTA, or FDG-PET/CT demonstrates paravalvular infection not detected by TEE. For optimal outcome, paravalvular infection requires surgery, especially when fever persists, fistulae develop, prostheses are dehisced and unstable, or infection relapses after appropriate treatment. Cardiac rhythm must be monitored since high-grade heart block may require insertion of a pacemaker.

Uncontrolled Infection Continued positive blood cultures or otherwise unexplained persistent fevers despite optimal antibiotic therapy may reflect uncontrolled infection that warrants surgery. Surgical treatment is also advised for IE caused by organisms against which effective antimicrobial therapy is lacking (e.g., yeasts, molds, *P. aeruginosa*, other highly antibiotic-resistant bacteria, *Brucella* species).

S. *aureus* IE The mortality rate for *S. aureus* PVE exceeds 50% with medical treatment and may be reduced with surgical treatment. Nevertheless, surgery is not routinely advised for uncomplicated *S. aureus* PVE. Rather, survival benefits are most likely in those with paravalvular infection, dysfunctional valves, and CHF. Surgical treatment of *S. aureus* NVE should be guided by the standard indications. Isolated tricuspid-valve *S. aureus* IE, even with persistent fever, rarely requires surgery.

Prevention of Systemic Emboli Persisting morbidity and/or death may result from cerebral or coronary artery emboli. Antithrombotic therapy does not prevent systemic emboli in NVE. The frequency of embolization decreases rapidly with effective antimicrobial therapy. Thus, if emboli are to be prevented through surgical intervention, surgery must occur very early. Vegetation characteristics defined echocardiographically can identify patients at high risk of embolization but do not identify those patients in whom surgery to prevent emboli will increase survival. In a small randomized trial in patients who were at low risk of surgery-related mortality and had large vegetations (>10 mm) and significant valve dysfunction, emboli were prevented by early surgery (≤48 h after diagnosis), but there was no survival benefit. Rarely is prevention of emboli the sole indication for surgery; more often, this may be an additional benefit of early surgery for other indications. Valve repair, with the consequent avoidance of a prosthesis, improves the benefit-to-risk ratio of surgery performed to eliminate vegetations.

CIED-IE Removal of all hardware is recommended for patients with established CIED-IE as well as for pocket or intracardiac lead infection. Percutaneous lead extraction is preferred; if hardware remains after attempted percutaneous extraction, surgical removal should be considered. With lead vegetations >2 cm, there is a risk of a pulmonary embolism; nevertheless, the need for CIED removal surgically is unclear. Removal of the infected CIED during the initial hospitalization is associated with increased 30-day and 1-year survival rates over those attained with antibiotic therapy and device retention. The CIED, if needed, can be reimplanted at a new site after at least 10–14 days of effective antimicrobial therapy. CIEDs should be replaced when patients undergo surgery for IE.

Timing of Cardiac Surgery With life-threatening indications for surgery (valve dysfunction and severe CHF, perivalvular abscess, major prosthesis dehiscence), surgery during the initial days of therapy is associated with greater survival than later surgery. With less compelling indications, surgery may reasonably be delayed to allow further treatment as well as improvement in overall health (Table 133-7). Recrudescent IE on a newly implanted prosthetic valve follows surgery for active NVE and PVE in 2% and 6–15% of patients, respectively. These frequencies do not justify the increased mortality risk associated with delaying surgery in patients with severe heart failure, valve dysfunction, and uncontrolled infections. Delay is justified when infection and CHF are controlled with medical therapy.

Neurologic complications of IE may be exacerbated during cardiac surgery. The risk of neurologic deterioration is related to the type and severity of the preoperative neurologic complication and the interval between the complication and surgery. In a nonobtunded patient with an ischemic stroke and hemorrhage excluded by imaging, cardiac surgery, if urgent, should be performed early because risk of hemorrhagic conversion or other neurologic side effects is low. In patients with intracranial hemorrhage and need for urgent surgery, early surgery should be pursued as long as the patient is expected to have meaningful recovery. Similar rates of postoperative intracranial hemorrhage and no difference in mortality have been observed in patients with and without preoperative intracranial hemorrhage. Nonurgent cardiac surgery should be delayed for 2–3 weeks after a large nonhemorrhagic embolic infarction and for 4 weeks after a significant cerebral hemorrhage. A ruptured mycotic aneurysm should be treated before cardiac surgery.

Antibiotic Therapy after Cardiac Surgery Organisms have been detected on Gram stain—or their DNA has been detected by PCR—in excised valves from 45% of patients who have completed the recommended therapy for IE. However, organisms, most of which are unusual or antibiotic resistant, are rarely cultured from these valves. Detection of organisms or their DNA does not necessarily indicate antibiotic failure; in fact, relapse after surgery for active IE is uncommon. Thus, in uncomplicated NVE caused by susceptible organisms, the duration of preoperative plus postoperative treatment should equal the total duration of recommended therapy. For IE complicated by perivalvular abscess, partially treated PVE, or valves culture-positive for the original organism, a full course of therapy should be given postoperatively.

Treatment of IE in PWID PWID should be treated according to the standard guidelines for antibiotic selection and surgical intervention. Additionally, OUD must be recognized as an ongoing predisposition for IE and treated; this includes medication-assisted therapy that is initiated during hospitalization and continued without delay upon discharge. Addressing OUD significantly increases completion of antibiotic therapy, decreases resumption of injection drug use, and decreases recurrent IE and requirement for cardiac surgery.

Extracardiac Complications Splenic abscess develops in 3–5% of patients with IE. Effective therapy requires either image-guided percutaneous drainage or splenectomy. Mycotic aneurysms occur in 2–15% of IE patients; one-half of these cases involve the cerebral arteries and present as headaches, focal neurologic symptoms, or hemorrhage. Cerebral aneurysms should be monitored by angiography. Some will resolve with effective antimicrobial therapy, but those that have leaked or persist or enlarge should be treated surgically, if possible. Extracerebral aneurysms present as local pain, a mass, local ischemia, or bleeding; these are treated surgically.

OUTCOME

IE is a heterogeneous disease that occurs in extremely heterogeneous patient populations. Adverse outcomes are associated with older age, severe comorbid conditions and diabetes, delayed diagnosis, involvement of prosthetic valves or the aortic valve, an invasive (*S. aureus*) or antibiotic-resistant (*P. aeruginosa*, yeast) pathogen, intracardiac and major neurologic complications, and health care–associated infection.

TABLE 133-8 Antibiotic Regimens for Prophylaxis of Endocarditis in Adults with High-Risk Cardiac Lesions[a,b]

A. Standard oral regimen
Amoxicillin: 2 g PO 1 h before procedure
B. Inability to take oral medication
Ampicillin: 2 g IV or IM within 1 h before procedure
C. Penicillin allergy
1. Cephalexin[c]: 2 g PO 1 h before procedure
2. Clarithromycin or azithromycin: 500 mg PO 1 h before procedure
3. Doxycycline: 100 mg PO 1 h before procedure
D. Penicillin allergy, inability to take oral medication
Cefazolin[c] or ceftriaxone[c]: 1 g IV or IM 30 min before procedure

[a]Dosing for children: for amoxicillin, ampicillin, cephalexin, or cefadroxil, use 50 mg/kg PO; cefazolin, 25 mg/kg IV; clindamycin, 20 mg/kg PO or 25 mg/kg IV; clarithromycin, 15 mg/kg PO; and vancomycin, 20 mg/kg IV. [b]For high-risk lesions, see Table 133-9. Prophylaxis is not advised for other lesions. [c]Do not use cephalosporins in patients with immediate hypersensitivity (urticaria, angioedema, anaphylaxis) to penicillin.

Source: Table created using the guidelines published by the American Heart Association and the European Society of Cardiology (W Wilson et al: Circulation 116:1736, 2007; W Wilson et al: Circulation 143:e963, 2021; and G Habib et al: Eur Heart J 30:2369, 2009).

Death or poor outcome often is related not to failure of antibiotic therapy but rather to the interactions of comorbidities and IE-related end-organ complications. In developed countries, overall survival rates are 80–85%; however, rates vary considerably among subpopulations of IE patients. The outcome for a given patient depends on that individual's infection, the complexity of required therapy, and preexisting comorbidities. About 85–90% of patients with NVE caused by viridans streptococci, HACEK organisms, or enterococci (susceptible to synergistic therapy) survive. For *S. aureus* NVE in patients who do not inject drugs, survival rates are 55–70%; rates are 85–90% among PWID. However, 1-year mortality rises to 20–30% among PWID if substance use disorder is not successfully addressed. PVE beginning within 2 months after valve replacement results in mortality rates of 40–50%, whereas rates are only 10–20% in late-onset cases. In the elderly population with TAVR-PVE the in-hospital mortality is 35–50% and increases to 60–75% at 1 year. Crude survival rates after successful treatment of IE generally are 80–90% and 70–80% at 1 and 2 years, respectively.

PREVENTION

Prevention of IE has been a goal of clinical practice; however, the evidence establishing benefit from *antibiotic prophylaxis for IE* is insufficient to recommend it as a widespread standard of care. The American Heart Association and the European Society of Cardiology recommend limiting prophylactic antibiotics **(Table 133-8)** to only patients at highest risk for severe morbidity or death from IE **(Table 133-9)**.

TABLE 133-9 High-Risk Cardiac Lesions for Which Endocarditis Prophylaxis Is Advised Before Dental Procedures

Prosthetic heart valves or material
Left ventricular assist devices or implantable heart
Prior endocarditis
Unrepaired cyanotic congenital heart disease, including palliative shunts or conduits
Completely repaired congenital heart defects during the 6 months after repair
Repaired congenital heart disease with residual defects adjacent to prosthetic material
Surgical or transcatheter pulmonary artery valve or conduit placement
Valvulopathy developing after cardiac transplantation

Source: Table created using the guidelines published by the American Heart Association and the European Society of Cardiology (W Wilson et al: Circulation 116:1736, 2007; W Wilson et al: Circulation 143:e963, 2021; and G Habib et al: Eur Heart J 30:2369, 2009).

In at-risk patients, maintaining good dental hygiene is recommended, and antibiotic prophylaxis is recommended only when there is manipulation of gingival tissue or the periapical region of the teeth or perforation of the oral mucosa (including with respiratory tract surgery). Recent studies suggest that severe adverse events related to amoxicillin prophylaxis are exceedingly rare; however, clindamycin prophylaxis has been associated with low but significant rates of fatal and nonfatal adverse reactions with *Clostridioides difficile* infection. Consequently, the American Heart Association now recommends against the use of clindamycin for prophylaxis. Although prophylaxis is not advised for patients undergoing gastrointestinal or genitourinary tract procedures, genitourinary tract infections (or skin infection) should be treated before or when these sites undergo procedures.

In patients with aortic or mitral valve regurgitation or a prosthetic valve, treatment of acute Q fever for 12 months with doxycycline plus hydroxychloroquine (see Table 133-4) is highly effective in preventing *C. burnetii* IE.

Acknowledgment

The authors would like to acknowledge the important contributions of Adolf W. Karchmer, MD, to this chapter in prior editions.

FURTHER READING

Baddour LM et al: Infective endocarditis in adults: Diagnosis, antimicrobial therapy, and management of complications: A scientific statement for healthcare professionals from the American Heart Association. Circulation 132:1435, 2015.

Baddour LM et al: Update on cardiovascular implantable electronic device infections and their prevention, diagnosis, and management: A scientific statement from the American Heart Association. Circulation 149:e201, 2024.

Baddour LM et al: Management of infective endocarditis in people who inject drugs: A scientific statement from the American Heart Association. Circulation 146:e187, 2022.

Bourque JM et al: 18F-FDG PET/CT and radiolabeled leukocyte SPECT/CT imaging for the evaluation of cardiovascular infection in the multimodality context: ASNC Imaging Indications (ASNC I2) Series Expert Consensus Recommendations from ASNC, AATS, ACC, AHA, ASE, EANM, HRS, IDSA, SCCT, SNMMI, and STS. Clin Infect Dis 2024. Corrected and republished in Heart Rhythm 21:e1, 2024.

Chirouze C et al: Impact of early valve surgery on outcome of *Staphylococcus aureus* prosthetic valve infective endocarditis: Analysis in the International Collaboration of Endocarditis–Prospective Cohort Study. Clin Infect Dis 60:741, 2015.

Chobufo MD et al: Trends in infective endocarditis mortality in the United States: 1999 to 2020: A cause for alarm. J Am Heart Assoc 12:e031589, 2023.

Delgado V et al: 2023 ESC guidelines for the management of endocarditis. Eur Heart J 44:3948, 2023.

Duval X et al: Impact of systematic whole-body 18F-fluorodeoxyglucose PET/CT on the management of patients suspected of infective endocarditis: The prospective multicenter TEPvENDO study. Clin Infect Dis 73:393, 2021.

Fowler VG Jr et al: The 2023 Duke-International Society for Cardiovascular Infectious Diseases criteria for infective endocarditis: Updating the modified Duke criteria. Clin Infect Dis 77:518, 2023.

Liesman RM et al: Laboratory diagnosis of infective endocarditis. J Clin Microbiol 55:2599, 2017.

Regueiro A et al: Association between transcatheter aortic valve replacement and subsequent infective endocarditis and in-hospital death. JAMA 316:1083, 2016.

Wilson W et al: Prevention of viridans group streptococcal infective endocarditis: A scientific statement from the American Heart Association. Circulation 143:e963, 2021.

134 Infections of the Skin, Muscles, and Soft Tissues

Dennis L. Stevens, Amy E. Bryant

Skin and soft tissue infections occur in all races, all ethnic groups, and all geographic locations, although some have unique geographic niches. In modern times, the frequency and severity of some skin and soft tissue infections have increased for several reasons. First, microbes are rapidly disseminated throughout the world via efficient air travel, acquiring genes for virulence factors and antibiotic resistance. Second, natural disasters, such as earthquakes, tsunamis, tornadoes, and hurricanes, appear to be increasing in frequency, and the injuries sustained during these events commonly cause major skin and soft tissue damage that predisposes to infection. Third, trauma and casualties resulting from combat and terrorist activities can markedly damage or destroy tissues and provide both endogenous and exogenous pathogens with ready access to deeper structures. Unfortunately, because the marvels of modern medicine may not be available during human-instigated and natural disasters, primary treatment may be delayed and the likelihood of severe infection and death increased.

ANATOMIC RELATIONSHIPS: CLUES TO THE DIAGNOSIS OF SOFT TISSUE INFECTIONS

Skin and soft tissue infections have been common human afflictions for centuries. However, between 2000 and 2004, hospital admissions for these infections rose by 27%, a remarkable increase that was attributable largely to the emergence of the USA300 clone of methicillin-resistant *Staphylococcus aureus* (MRSA). This chapter provides an anatomic approach to understanding the types of soft tissue infections and the diverse microbes responsible.

Protection against infection of the epidermis depends on the mechanical barrier afforded by the stratum corneum since the epidermis itself is devoid of blood vessels **(Fig. 134-1)**. Disruption of this layer by burns or bites, abrasions, foreign bodies, primary dermatologic disorders (e.g., herpes simplex, varicella, ecthyma gangrenosum), surgery, or vascular or pressure ulcers allows penetration of bacteria to the deeper structures. Similarly, the hair follicle can serve as a portal either for components of the normal flora (e.g., *Staphylococcus*) or for extrinsic bacteria (e.g., *Pseudomonas* in hot-tub folliculitis). Intracellular infection of the squamous epithelium with vesicle formation may arise from cutaneous inoculation, as in infection with herpes simplex virus (HSV) type 1; from the dermal capillary plexus, as in varicella and infections due to other viruses associated with viremia; or from cutaneous nerve roots, as in herpes zoster. Bacteria infecting the epidermis, such as *Streptococcus pyogenes*, may be translocated laterally to deeper structures via lymphatics, an event that results in the rapid superficial spread of erysipelas. Later, engorgement or obstruction of lymphatics causes flaccid edema of the epidermis, another characteristic of erysipelas.

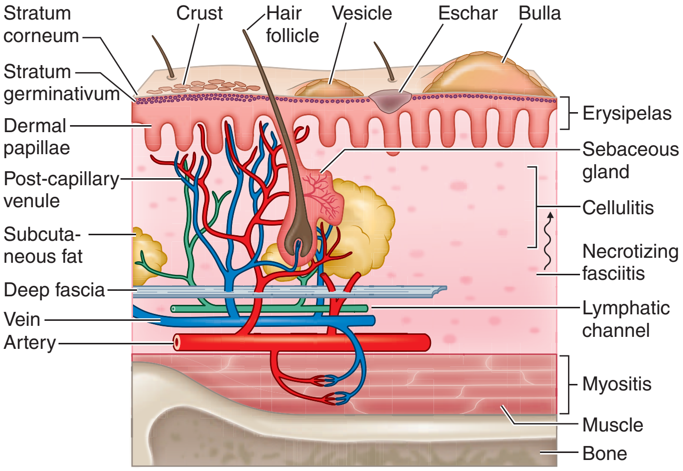

FIGURE 134-1 Structural components of the skin and soft tissues, superficial infections, and infections of the deeper structures. The rich capillary network beneath the dermal papillae plays a key role in the localization of infection and in the development of the acute inflammatory reaction.

The rich plexus of capillaries beneath the dermal papillae provides nutrition to the stratum germinativum, and physiologic responses of this plexus produce important clinical signs and symptoms. For example, infective vasculitis of the plexus results in petechiae, Osler nodes, Janeway lesions, and palpable purpura, which, if present, are important clues to the existence of endocarditis **(Chap. 133)**. In addition, metastatic infection within this plexus can result in cutaneous manifestations of disseminated fungal infection **(Chap. 222)**, gonococcal infection **(Chap. 161)**, *Salmonella* infection **(Chap. 171)**, *Pseudomonas* infection (i.e., ecthyma gangrenosum; **Chap. 170**), meningococcemia **(Chap. 160)**, and staphylococcal infection **(Chap. 152)**. The plexus also provides bacteria with access to the circulation, thereby facilitating local spread or bacteremia. The postcapillary venules of this plexus are a prominent site of polymorphonuclear leukocyte sequestration, diapedesis, and chemotaxis to the site of cutaneous infection.

Amplification of these physiologic mechanisms by excessive levels of cytokines or bacterial toxins causes leukostasis, venous occlusion, and pitting edema. Edema with purple bullae, ecchymosis, and cutaneous anesthesia suggests loss of vascular integrity and necessitates exploration of the deeper structures for evidence of necrotizing fasciitis or myonecrosis. An early diagnosis requires a high level of suspicion in instances of unexplained fever and of pain and tenderness in the soft tissue, even in the absence of acute cutaneous inflammation.

Table 134-1 indicates the chapters in which the infections described below are discussed in greater detail. Many of these infections are illustrated in the chapters cited or in **Chap. A1**.

INFECTIONS ASSOCIATED WITH VESICLES

(Table 134-1) Vesicle formation due to infection is caused by viral proliferation within the epidermis. In varicella and variola, viremia precedes the onset of a diffuse centripetal rash that progresses from macules to vesicles, then to pustules, and finally to scabs over the course of 1–2 weeks. Vesicles of varicella have a "dewdrop" appearance and develop in crops randomly about the trunk, extremities, and face over 3–4 days. Herpes zoster occurs in a single dermatome; the appearance of vesicles is preceded by pain for several days. Zoster may occur in persons of any age but is most common among immunosuppressed individuals and elderly patients, whereas most cases of varicella occur in young children. Vesicles due to HSV are found on or around the lips (HSV-1) or genitals (HSV-2) but also may appear on the head and neck of young wrestlers (herpes gladiatorum) or on the digits of health care workers (herpetic whitlow). Recurrent herpes labialis (HSV-1) and herpes genitalis (HSV-2) commonly follow primary infection. Coxsackievirus A16 characteristically causes vesicles on the hands, feet, and mouth of children. Orf is caused by a DNA virus related to smallpox virus and infects the fingers of individuals who work around goats and sheep. Molluscum contagiosum virus induces flaccid vesicles on the skin of healthy and immunocompromised individuals. Although variola (smallpox) in nature was eradicated as of 1977, postmillennial terrorist events have renewed interest in this devastating infection **(Chap. S4)**. Viremia beginning after an incubation period of 12 days is followed by a diffuse maculopapular rash, with rapid evolution to vesicles, pustules, and then scabs. Secondary cases can occur among close contacts.

Rickettsialpox begins after mite-bite inoculation of *Rickettsia akari* into the skin. A papule with a central vesicle evolves to form a 1- to 2.5-cm painless crusted black eschar with an erythematous halo and proximal adenopathy. While more common in the northeastern United States and Ukraine in 1940–1950, rickettsialpox has recently been described in Ohio, Arizona, and Utah. Blistering dactylitis is a painful, vesicular, localized *S. aureus* or group A streptococcal infection of the pulps of the distal digits of the hands.

TABLE 134-1 Skin and Soft Tissue Infections

LESION, CLINICAL SYNDROME	INFECTIOUS AGENT(S)	SEE ALSO CHAP(S).
Vesicles		
Smallpox	Variola virus	**S4**
Chickenpox	Varicella-zoster virus	**198**
Shingles (herpes zoster)	Varicella-zoster virus	**198**
Cold sores, herpetic whitlow, herpes gladiatorum	Herpes simplex virus	**197**
Hand-foot-and-mouth disease	Coxsackievirus A16	**210**
Orf	Parapoxvirus	**201**
Molluscum contagiosum	Molluscum contagiosum poxvirus	**201**
Rickettsialpox	*Rickettsia akari*	**192**
Blistering distal dactylitis	*Staphylococcus aureus* or *Streptococcus pyogenes*	**152, 153**
Bullae		
Staphylococcal scalded-skin syndrome	*S. aureus*	**152**
Necrotizing fasciitis	*S. pyogenes*, *Clostridium* spp., mixed aerobes and anaerobes	**153, 159, 182**
Gas gangrene	*Clostridium* spp.	**159**
Halophilic *Vibrio*	*Vibrio vulnificus*	**173**
Crusted lesions		
Bullous impetigo/ecthyma	*S. aureus*	**152**
Impetigo contagiosa	*S. pyogenes*	**153**
Ringworm	Superficial dermatophyte fungi	**225**
Sporotrichosis	*Sporothrix schenckii*	**225**
Histoplasmosis	*Histoplasma capsulatum*	**218**
Coccidioidomycosis	*Coccidioides immitis*	**219**
Blastomycosis	*Blastomyces dermatitidis*	**220**
Cutaneous leishmaniasis	*Leishmania* spp.	**233**
Cutaneous tuberculosis	*Mycobacterium tuberculosis*	**183**
Nocardiosis	*Nocardia asteroides*	**179**
Folliculitis		
Furunculosis	*S. aureus*	**152**
Hot-tub folliculitis	*Pseudomonas aeruginosa*	**170**
Swimmer's itch	*Schistosoma* spp.	**241**
Acne vulgaris	*Propionibacterium acnes*	**60**
Papular and nodular lesions		
Fish-tank or swimming-pool granuloma	*Mycobacterium marinum*	**185**
Creeping eruption (cutaneous larva migrans)	*Ancylostoma braziliense*	**238**
Dracunculiasis	*Dracunculus medinensis*	**240**
Cercarial dermatitis	*Schistosoma mansoni*	**241**
Verruca vulgaris	Human papillomaviruses 1, 2, 4	**203**
Condylomata acuminata (anogenital warts)	Human papillomaviruses 6, 11, 16, 18	**203**
Onchocerciasis nodule	*Onchocerca volvulus*	**240**
Cutaneous myiasis	*Dermatobia hominis*	**472**
Verruca peruana	*Bartonella bacilliformis*	**177**
Cat-scratch disease	*Bartonella henselae*	**177**
Lepromatous leprosy	*Mycobacterium leprae*	**184**
Secondary syphilis (papulosquamous and nodular lesions, condylomata lata)	*Treponema pallidum*	**187**
Tertiary syphilis (nodular gummatous lesions)	*T. pallidum*	**187**
Ulcers with or without eschars		
Anthrax	*Bacillus anthracis*	**S4**
Ulceroglandular tularemia	*Francisella tularensis*	**175, S4**
Bubonic plague	*Yersinia pestis*	**176, S4**
Buruli ulcer	*Mycobacterium ulcerans*	**185**
Leprosy	*M. leprae*	**184**
Cutaneous tuberculosis	*M. tuberculosis*	**183**
Chancroid	*Haemophilus ducreyi*	**162**
Primary syphilis	*T. pallidum*	**187**
Erysipelas	*S. pyogenes*	**153**
Cellulitis	*Staphylococcus* spp., *Streptococcus* spp., various other bacteria	Various
Necrotizing fasciitis		
Streptococcal gangrene	*S. pyogenes*	**153**
Fournier gangrene	Mixed aerobic and anaerobic bacteria	**182**
Staphylococcal necrotizing fasciitis	Methicillin-resistant *S. aureus*	**152**
Myositis and myonecrosis		
Pyomyositis	*S. aureus*	**152**
Streptococcal necrotizing myositis	*S. pyogenes*	**153**
Gas gangrene	*Clostridium* spp.	**159**
Nonclostridial (crepitant) myositis	Mixed aerobic and anaerobic bacteria	**182**
Synergistic nonclostridial anaerobic myonecrosis	Mixed aerobic and anaerobic bacteria	**182**

The recent spike in mpox (formerly monkeypox virus; MPXV) **(Chap. 201)** cases has raised concerns due to its clinical resemblance to smallpox in terms of symptom onset, timing of rash occurrence, and rash distribution. MPXV infection is generally less severe than smallpox in terms of complication rate and levels of scarification. The case fatality rate (1–10%) is also less than that of smallpox. Though largely endemic in West and Central Africa, cases have recently emerged in at least 10 other African countries. In addition, cases have been reported in Israel, the United Kingdom, and in six midwestern states of the United States. Most non-African cases are associated with tourism to endemic African countries or with exposure to small mammals imported from such areas. However, human-to-human disease transmission was clearly documented in a British health care worker who attended a patient that had acquired the infection during a trip to Nigeria. Most patients are in lower age groups (<40 years) due to a lack cross-protective immunity, having been born after discontinuation of the smallpox eradication campaign in the 1980s. This finding heightens the concern for potential global disease spread.

INFECTIONS ASSOCIATED WITH BULLAE

(Table 134-1) Staphylococcal scalded-skin syndrome (SSSS) in neonates is caused by a toxin (exfoliatin) from phage group II *S. aureus*. SSSS must be distinguished from toxic epidermal necrolysis (TEN), which occurs primarily in adults, is drug-induced, and is associated with a higher mortality rate. Punch biopsy with frozen section is useful in making this distinction since the cleavage plane is the stratum corneum in SSSS and the stratum germinativum in TEN (Fig. 134-1). Treatment with intravenous immune globulin plus corticosteroids may reduce recovery time and improve prognosis in patients with TEN. Necrotizing fasciitis and gas gangrene also induce bulla formation (see "Necrotizing Fasciitis," below). Halophilic *Vibrio* infection can be as aggressive and fulminant as necrotizing fasciitis; a helpful clue in its diagnosis is a history of exposure to waters of the Gulf of Mexico or the Atlantic seaboard or (in a patient with cirrhosis) the ingestion of raw seafood. The etiologic organism (*Vibrio vulnificus*) is highly susceptible to tetracycline.

INFECTIONS ASSOCIATED WITH CRUSTED LESIONS

(Table 134-1) Impetigo contagiosa is caused by *S. pyogenes*, and bullous impetigo is due to *S. aureus*. Both skin lesions may have an early bullous stage but then appear as thick crusts with a golden-brown color. Epidemics of impetigo caused by MRSA have been reported. Streptococcal lesions are most common among children 2–5 years of age, and epidemics may occur in settings of poor hygiene, particularly among children in lower socioeconomic settings in tropical climates. It is important to recognize impetigo contagiosa because of its relationship to poststreptococcal glomerulonephritis. Rheumatic fever is not a complication of skin infection caused by *S. pyogenes*. Superficial dermatophyte infection (ringworm) can occur on any skin surface, and skin scrapings with KOH staining are diagnostic. Primary infections with dimorphic fungi such as *Blastomyces dermatitidis* and *Sporothrix schenckii* can initially present as crusted skin lesions resembling ringworm. Disseminated infection with *Coccidioides immitis* can also involve the skin, and biopsy and culture should be performed on crusted lesions when the patient is from an endemic area. Crusted nodular lesions caused by *Mycobacterium chelonae* have been described in HIV-seropositive patients.

FOLLICULITIS

(Table 134-1) Hair follicles serve as portals for a number of bacteria, although *S. aureus* is the most common cause of localized folliculitis. Sebaceous glands empty into hair follicles and ducts and, if these portals are blocked, form sebaceous cysts that may resemble staphylococcal abscesses or may become secondarily infected. Inflammation of sweat glands (hidradenitis suppurativa) also can mimic infection of hair follicles, particularly in the axillae, but new treatments with potent anti-inflammatory agents hold promise. Chronic folliculitis is uncommon except in acne vulgaris, where constituents of the normal flora (e.g., *Propionibacterium acnes*) may play a role.

Diffuse folliculitis occurs in two settings. *Hot-tub folliculitis* is caused by *Pseudomonas aeruginosa* in waters that are insufficiently chlorinated and maintained at temperatures of 37–40°C. Infection is usually self-limited, although bacteremia and shock have been reported. *Swimmer's itch* occurs when a skin surface is exposed to water infested with freshwater avian schistosomes. Warm water temperatures and alkaline pH are suitable for mollusks that serve as intermediate hosts between birds and humans. Free-swimming schistosomal cercariae readily penetrate human hair follicles or pores but quickly die and elicit a brisk allergic reaction, causing intense itching and erythema.

PAPULAR AND NODULAR LESIONS

(Table 134-1) Raised lesions of the skin occur in many different forms. *Mycobacterium marinum* infections of the skin may present as cellulitis or as raised erythematous nodules. Similar lesions caused by *Mycobacterium abscessus* and *M. chelonae* have been described among patients undergoing cosmetic laser surgery and tattooing, respectively. Erythematous papules are early manifestations of cat-scratch disease (with lesions developing at the primary site of inoculation of *Bartonella henselae*) and bacillary angiomatosis (also caused by *B. henselae*). Raised serpiginous or linear eruptions are characteristic of cutaneous larva migrans, which is caused by burrowing larvae of dog or cat hookworms (*Ancylostoma braziliense*) and which humans acquire through contact with soil that has been contaminated with dog or cat feces. Similar burrowing raised lesions are present in dracunculiasis caused by migration of the adult female nematode *Dracunculus medinensis*. Nodules caused by *Onchocerca volvulus* measure 1–10 cm in diameter and occur mostly in persons bitten by *Simulium* flies in Africa. The nodules contain the adult worm encased in fibrous tissue. Migration of microfilariae into the eyes may result in blindness. Verruga peruana is caused by *Bartonella bacilliformis*, which is transmitted to humans by the sandfly *Phlebotomus*. This condition can take the form of single gigantic lesions (several centimeters in diameter) or multiple small lesions (several millimeters in diameter). Numerous subcutaneous nodules may also be present in cysticercosis caused by larvae of *Taenia solium*. Multiple erythematous papules develop in schistosomiasis; each represents a cercarial invasion site. Skin nodules as well as thickened subcutaneous tissue are prominent features of lepromatous leprosy. Large nodules or gummas are features of tertiary syphilis, whereas flat papulosquamous lesions are characteristic of secondary syphilis. Human papillomavirus may cause singular warts (verruca vulgaris) or multiple warts in the anogenital area (condylomata acuminata). The latter are major problems in HIV-infected individuals and others with defects in cell-mediated immunity.

ULCERS WITH OR WITHOUT ESCHARS

(Table 134-1) Cutaneous anthrax begins as a pruritic papule, which develops within days into an ulcer with surrounding vesicles and edema and then into an enlarging ulcer with a black eschar. Cutaneous anthrax may cause chronic nonhealing ulcers with an overlying dirty-gray membrane, although lesions may also mimic psoriasis, eczema, or impetigo. Ulceroglandular tularemia may have associated ulcerated skin lesions with painful regional adenopathy. Although buboes are the major cutaneous manifestation of plague, ulcers with eschars, papules, or pustules also are present in 25% of cases.

Mycobacterium ulcerans typically causes chronic skin ulcers on the extremities of individuals living in the tropics. *Mycobacterium leprae* may be associated with cutaneous ulcerations in patients with lepromatous leprosy related to Lucio's phenomenon, in which immune-mediated destruction of tissue bearing high concentrations of *M. leprae* bacilli occurs, usually several months after initiation of effective therapy. *Mycobacterium tuberculosis* also may cause ulcerations, papules, or erythematous macular lesions of the skin in both immunocompetent and immunocompromised patients.

Decubitus ulcers are due to tissue hypoxemia secondary to pressure-induced vascular insufficiency and may become secondarily infected with components of the skin and gastrointestinal flora, including anaerobes. Ulcerative lesions on the anterior shins may be due to pyoderma gangrenosum, which must be distinguished from similar lesions of infectious etiology by histologic evaluation of biopsy sites. Ulcerated lesions on the genitals may be either painful (chancroid) or painless (primary syphilis).

ERYSIPELAS

(Table 134-1) Erysipelas is due to *S. pyogenes* and is characterized by an abrupt onset of fiery-red swelling of the face or extremities. The distinctive features of erysipelas are well-defined indurated margins, particularly along the nasolabial fold; rapid progression; and intense pain. Flaccid bullae may develop during the second or third day of illness, but extension to deeper soft tissues is rare. Treatment with penicillin is effective; swelling may progress despite appropriate treatment, although fever, pain, and the intense red color diminish. Desquamation of the involved skin occurs 5–10 days into the illness. Infants and elderly adults are most commonly afflicted, and the severity of systemic toxicity varies.

CELLULITIS

(Table 134-1) Cellulitis is an acute inflammatory condition of the skin that is characterized by localized pain, erythema, swelling, and heat. It may be caused by indigenous flora colonizing the skin and appendages (e.g., *S. aureus* and *S. pyogenes*) or by a wide variety of exogenous bacteria. Because the exogenous bacteria involved in cellulitis occupy unique niches in nature, a thorough history (including epidemiologic data) offers important clues to etiology. When there is drainage, an open wound, or an obvious portal of entry, Gram stain and culture provide a definitive diagnosis. In the absence of these findings, the bacterial etiology of cellulitis is difficult to establish, and in some cases, staphylococcal and streptococcal cellulitis may have similar features. Even with needle aspiration of the leading edge or a punch biopsy of the cellulitis tissue itself, cultures are positive in only 20% of cases. This observation suggests that relatively low numbers of bacteria may cause cellulitis and that the expanding area of erythema within the skin may be a direct effect of extracellular toxins or of the soluble mediators of inflammation elicited by the host.

Bacteria may gain access to the epidermis through cracks in the skin, abrasions, cuts, burns, insect bites, surgical incisions, and IV catheters. Cellulitis caused by *S. aureus* spreads from a central localized infection, such as an abscess, folliculitis, or an infected foreign body (e.g., a splinter, a prosthetic device, an IV catheter). MRSA is rapidly replacing methicillin-sensitive *S. aureus* (MSSA) as a cause of cellulitis in both inpatient and outpatient settings. Cellulitis caused by MSSA or MRSA is usually associated with a focal infection, such as a furuncle, a carbuncle, a surgical wound, or an abscess; the U.S. Food and Drug Administration preferentially refers to these types of infection as *purulent cellulitis*. In contrast, cellulitis due to *S. pyogenes* is a more rapidly spreading, diffuse process that is frequently associated with lymphangitis and fever and should be referred to as *nonpurulent cellulitis*. Recurrent streptococcal cellulitis of the lower extremities may be caused by organisms of group A, C, or G in association with chronic venous stasis or with saphenous venectomy for coronary artery bypass surgery. Streptococci also cause recurrent cellulitis among patients with chronic lymphedema resulting from elephantiasis, lymph node dissection, or congenital disorders. Recurrent staphylococcal cutaneous infections are more common among individuals who have Job syndrome, the autosomal dominant hyper-IgE recurrent infection syndrome, and among nasal carriers of staphylococci. Cellulitis caused by *Streptococcus agalactiae* (group B *Streptococcus*) occurs primarily in elderly patients and those with diabetes mellitus or peripheral vascular disease. *Haemophilus influenzae* typically causes periorbital cellulitis in children in association with sinusitis, otitis media, or epiglottitis. It is unclear whether this form of cellulitis will (like meningitis) become less common as a result of the impressive efficacy of the *H. influenzae* type b vaccine.

Many other bacteria also cause cellulitis. It is fortunate that these organisms occur in such characteristic settings that a good history provides useful clues to the diagnosis. Cellulitis associated with cat bites and, to a lesser degree, with dog bites is commonly caused by *Pasteurella multocida*, although in the latter case *Staphylococcus intermedius* and *Capnocytophaga canimorsus* also must be considered. Sites of cellulitis and abscesses associated with dog bites and human bites also contain a variety of anaerobic organisms, including *Fusobacterium*, *Bacteroides*, aerobic and anaerobic streptococci, and *Eikenella corrodens*. *Pasteurella* is notoriously resistant to dicloxacillin and nafcillin but is sensitive to other β-lactam antimicrobial agents as well as to quinolones, tetracycline, and erythromycin. Amoxicillin-clavulanate, ampicillin-sulbactam, and cefoxitin are good choices for the treatment of animal or human bite infections. *Aeromonas hydrophila* causes aggressive cellulitis and occasionally necrotizing fasciitis in tissues surrounding lacerations sustained in freshwater (lakes, rivers, and streams). This organism remains sensitive to aminoglycosides, fluoroquinolones, chloramphenicol, trimethoprim-sulfamethoxazole, and third-generation cephalosporins; it is resistant to ampicillin, however. *P. aeruginosa* causes three types of soft tissue infection: ecthyma gangrenosum in neutropenic patients, hot-tub folliculitis, and cellulitis following penetrating injury. Most commonly, *P. aeruginosa* is introduced into the deep tissues following a significant penetrating injury, such as occurs when a person steps on a nail. Treatment includes surgical inspection and drainage, particularly if the injury also involves bone or joint capsule. Choices for empirical treatment (while antimicrobial susceptibility data are awaited) include an aminoglycoside, a third-generation cephalosporin (ceftazidime, cefoperazone), a semisynthetic penicillin (ticarcillin or piperacillin), or a fluoroquinolone **(Chap. 170)**.

Gram-negative bacillary cellulitis, including that due to *P. aeruginosa*, is most common among hospitalized, immunocompromised hosts. Cultures and sensitivity tests are critically important in this setting because of multidrug resistance **(Chap. 170)**.

The gram-positive aerobic rod *Erysipelothrix rhusiopathiae* is most often associated with fish and domestic swine and causes cellulitis primarily in bone renderers and fishmongers. *E. rhusiopathiae* remains susceptible to most β-lactam antibiotics (including penicillin), erythromycin, clindamycin, tetracycline, and cephalosporins but is resistant to sulfonamides, chloramphenicol, and vancomycin. Its resistance to vancomycin, which is unusual among gram-positive bacteria, is of potential clinical significance since this agent is sometimes used in empirical therapy for skin infection. Fish food containing the water flea *Daphnia* is sometimes contaminated with *M. marinum*, which can cause cellulitis or granulomas on skin surfaces exposed to the water in aquariums or injured in swimming pools **(Chap. 185)**. Ethambutol and either clarithromycin or azithromycin are the best therapies.

NECROTIZING FASCIITIS

(Table 134-1) Necrotizing fasciitis, formerly called *streptococcal gangrene*, may be associated with group A *Streptococcus* or mixed aerobic–anaerobic bacteria or may occur as a component of gas gangrene caused by *Clostridium perfringens*. Strains of MRSA that produce the Panton-Valentine leukocidin (PVL) toxin have been reported to cause necrotizing fasciitis. Early diagnosis may be difficult when pain or unexplained fever is the only presenting manifestation. Swelling then develops and is followed by brawny edema and tenderness. With progression, dark-red induration of the epidermis appears, along with bullae filled with blue or purple fluid. Later the skin becomes friable and takes on a bluish, maroon, or black color. By this stage, thrombosis of blood vessels in the dermal papillae (Fig. 134-1) is extensive. Extension of infection to the level of the deep fascia causes this tissue to take on a brownish-gray appearance. Rapid spread occurs along fascial planes, through venous channels and lymphatics. Patients in the later stages are toxic and frequently manifest shock and multiorgan failure.

Necrotizing fasciitis caused by mixed aerobic–anaerobic bacteria begins with a breach in the integrity of a mucous membrane barrier, such as the mucosa of the gastrointestinal or genitourinary tract. The portal can be a malignancy, a diverticulum, a hemorrhoid, an anal fissure, or a urethral tear. Other predisposing factors include peripheral vascular disease, diabetes mellitus, surgery, and penetrating injury to the abdomen. Leakage into the perineal area results in a syndrome called *Fournier gangrene*, characterized by massive swelling of the scrotum and penis with extension into the perineum or the abdominal wall and the legs.

Necrotizing fasciitis caused by *S. pyogenes* has increased in frequency and severity since 1985. There are two distinct clinical presentations: patients without, versus those with, a defined portal of bacterial entry. Infections in the first category often begin deep at the site of a nonpenetrating, relatively minor trauma, such as a bruise or a muscle strain. Seeding of the site via transient bacteremia is likely, although most patients deny antecedent streptococcal infection. Affected patients present with only severe pain and fever and are frequently misdiagnosed (e.g., thrombophlebitis), given pain-relieving drugs, and sent home. Later in the course, the classic signs of necrotizing fasciitis, such as purple (violaceous) bullae, skin sloughing, and progressive toxicity, develop. Mortality in this setting is high, and survivors often undergo repeated surgeries including amputations. In infections of the second type, *S. pyogenes* may reach the deep fascia from a site of cutaneous infection or penetrating trauma. These patients have early signs of superficial skin infection with progression to necrotizing fasciitis. In either setting, toxicity is severe, and renal impairment may precede the development of shock. In 20–40% of cases, myositis occurs concomitantly, and, as in gas gangrene (see below), serum creatine phosphokinase levels may be markedly elevated. Necrotizing fasciitis due to mixed aerobic–anaerobic bacteria may be associated with gas in deep tissue, but gas usually is not present when the cause is *S. pyogenes* or MRSA. Prompt surgical exploration down to the deep fascia and muscle is essential. Necrotic tissue must be surgically removed, and Gram staining and culture of excised tissue are useful in establishing whether group A streptococci, mixed aerobic–anaerobic bacteria, MRSA, or *Clostridium* species are present (see “Treatment,” below).

MYOSITIS AND MYONECROSIS

(Table 134-1) Muscle involvement can occur with viral infection (e.g., influenza, dengue, or coxsackievirus B infection) or parasitic invasion (e.g., trichinellosis, cysticercosis, or toxoplasmosis). Although myalgia develops in most of these infections, severe muscle pain is the hallmark of pleurodynia (coxsackievirus B), trichinellosis, and bacterial infection. Acute rhabdomyolysis predictably occurs with clostridial and streptococcal myositis but may also be associated with influenza virus, echovirus, coxsackievirus, Epstein-Barr virus, and *Legionella* infections.

Pyomyositis is usually due to *S. aureus*, is common in tropical areas, and generally has no known portal of entry. Cases of pyomyositis caused by MRSA producing the PVL toxin have been described among children in the United States. Muscle infection begins at the exact site of blunt trauma or muscle strain. Infection remains localized, and shock does not develop unless organisms produce toxic shock syndrome toxin 1 or certain enterotoxins and the patient lacks antibodies to the toxin produced by the infecting organisms. In contrast, *S. pyogenes* may induce primary myositis (referred to as *streptococcal necrotizing myositis*) in association with severe systemic toxicity. Myonecrosis occurs concomitantly with necrotizing fasciitis in ~50% of cases. Both are part of the streptococcal toxic shock syndrome.

Gas gangrene usually follows severe penetrating injuries that result in interruption of the blood supply and introduction of soil into wounds. Such cases of traumatic gangrene are usually caused by the clostridial species *C. perfringens*, *C. septicum*, and *C. histolyticum*. Rarely, latent or recurrent gangrene can occur years after penetrating trauma; dormant spores that reside at the site of previous injury are most likely responsible. Spontaneous nontraumatic gangrene among patients with neutropenia, gastrointestinal malignancy, diverticulosis, or recent radiation therapy to the abdomen is caused by several clostridial species, of which *C. septicum* is the most commonly involved. The tolerance of this anaerobe to oxygen probably explains why it can initiate infection spontaneously in normal tissue anywhere in the body.

Gas gangrene of the uterus, especially that due to *Clostridium sordellii,* historically occurred as a consequence of illegal or self-induced abortion and nowadays also follows spontaneous abortion, vaginal delivery, and Caesarean section. *C. sordellii* has also been implicated in medically induced abortion. Postpartum *C. sordellii* infections in young, previously healthy women present with little or no fever, lack of a purulent discharge, refractory hypotension, extensive peripheral edema and effusions, hemoconcentration, and a markedly elevated white blood cell count. The infection is often fatal, with death ensuing rapidly. *C. sordellii* and *C. novyi* have also been associated with cutaneous injection of black tar heroin; mortality rates are lower among these individuals, probably because their injection-site infections are readily apparent and diagnosis is therefore prompt.

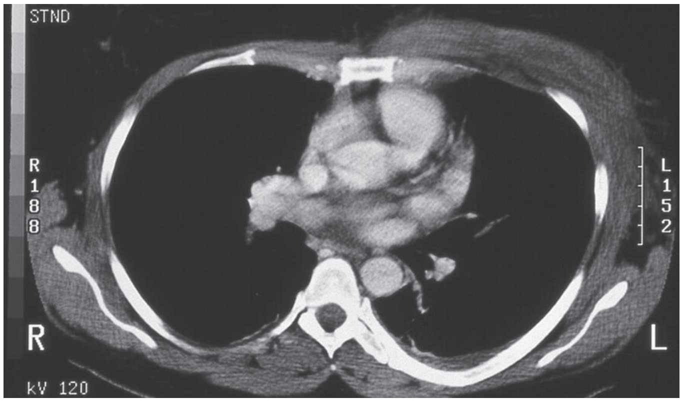

FIGURE 134-2 CT showing edema and inflammation of the left chest wall in a patient with necrotizing fasciitis and myonecrosis caused by group A *Streptococcus.*

Synergistic nonclostridial anaerobic myonecrosis, also known as *necrotizing cutaneous myositis* and *synergistic necrotizing cellulitis*, is a variant of necrotizing fasciitis caused by mixed aerobic and anaerobic bacteria with the exclusion of clostridial organisms (see “Necrotizing Fasciitis,” above).

DIAGNOSIS

This chapter emphasizes the physical appearance and location of lesions within the soft tissues as important diagnostic clues. Other crucial considerations in narrowing the differential diagnosis are the temporal progression of the lesions as well as the patient’s travel history, animal exposure or bite history, age, underlying disease status, and lifestyle. However, even the astute clinician may find it challenging to diagnose all infections of the soft tissues by history and inspection alone. Soft tissue radiography, computed tomography (CT) **(Fig. 134-2)**, and magnetic resonance imaging (MRI) may be useful in determining the depth of infection and should be performed when the patient has rapidly progressing lesions or evidence of a systemic inflammatory response syndrome. These tests are particularly valuable for defining a localized abscess or detecting gas in tissue. Unfortunately, they may reveal only soft tissue swelling and thus are not specific for fulminant infections such as necrotizing fasciitis or myonecrosis caused by group A *Streptococcus* (Fig. 134-2), where gas is not found in lesions.

Aspiration of the leading edge or punch biopsy with frozen section may be helpful if the results of imaging tests are positive, but negative results occur in ~80% of cases. There is some evidence that aspiration alone may be superior to injection and aspiration with normal saline. Frozen sections are especially useful in distinguishing SSSS from TEN and are quite valuable in cases of necrotizing fasciitis. Open surgical inspection, with debridement as indicated, is the best way to determine the extent and severity of infection and to obtain material for Gram staining and culture. Such an aggressive approach is important and may be lifesaving if undertaken early in the course of fulminant infections when there is evidence of systemic toxicity.

TREATMENT

Infections of the Skin, Muscles, and Soft Tissues

A full description of the treatment of all the clinical entities described herein is beyond the scope of this chapter. As a guide to the clinician in selecting appropriate treatment, the antimicrobial agents useful in the most common and the most fulminant cutaneous infections are listed in **Table 134-2**. Newer antibiotics approved

TABLE 134-2 Treatment of Common Infections of the Skin

DIAGNOSIS/CONDITION	PRIMARY TREATMENT	ALTERNATIVE TREATMENT	SEE ALSO CHAP(S).
Animal bite (prophylaxis or early infection)[a]	Amoxicillin–clavulanate (875/125 mg PO bid)	Doxycycline (100 mg PO bid)	146
Animal bite[a] (established infection)	Ampicillin–sulbactam (1.5–3 g IV q6h)	Clindamycin (600–900 mg IV q8h) *plus* Ciprofloxacin (400 mg IV q12h) *or* cefoxitin (2 g IV q6h)	146
Bacillary angiomatosis	Erythromycin (500 mg PO qid)	Doxycycline (100 mg PO bid)	177
Herpes simplex (primary genital)	Acyclovir (400 mg PO tid for 10 days)	Famciclovir (250 mg PO tid for 5–10 days) *or* valacyclovir (1000 mg PO bid for 10 days)	197
Herpes zoster (immunocompetent host >50 years of age)	Acyclovir (800 mg PO 5 times daily for 7–10 days)	Famciclovir (500 mg PO tid for 7–10 days) *or* valacyclovir (1000 mg PO tid for 7 days)	198
Cellulitis (staphylococcal or streptococcal[b,c])	Nafcillin or oxacillin (2 g IV q4–6h)	Cefazolin (1–2 g q8h) *or* ampicillin/sulbactam (1.5–3 g IV q6h) *or* erythromycin (0.5–1 g IV q6h) *or* clindamycin (600–900 mg IV q8h)	152, 153
MRSA skin infection[d]	Vancomycin (1 g IV q12h)	Linezolid (600 mg IV q12h)	152
Necrotizing fasciitis (group A streptococcal[b])	Clindamycin (600–900 mg IV q6–8h) *plus* penicillin G (4 million units IV q4h)	Clindamycin (600–900 mg IV q6–8h) *plus* a cephalosporin (first- or second-generation)	153
Necrotizing fasciitis (mixed aerobes and anaerobes)	Ampicillin (2 g IV q4h) *plus* clindamycin (600–900 mg IV q6–8h) *plus* ciprofloxacin (400 mg IV q6–8h)	Vancomycin (1 g IV q6h) *plus* metronidazole (500 mg IV q6h) *plus* ciprofloxacin (400 mg IV q6–8h)	182
Gas gangrene	Clindamycin (600–900 mg IV q6–8h) *plus* penicillin G (4 million units IV q4–6h)	Clindamycin (600–900 mg IV q6–8h) *plus* cefoxitin (2 g IV q6h)	159

[a] *Pasteurella multocida,* a species commonly associated with both dog and cat bites, is resistant to cephalexin, dicloxacillin, clindamycin, and erythromycin. *Eikenella corrodens*, a bacterium commonly associated with human bites, is resistant to clindamycin, penicillinase-resistant penicillins, and metronidazole but is sensitive to trimethoprim-sulfamethoxazole and fluoroquinolones.

[b] The frequency of erythromycin resistance in group A *Streptococcus* is currently ~5% in the United States but has reached 70–100% in some other countries. Most, but not all, erythromycin-resistant group A streptococci are susceptible to clindamycin; clindamycin-resistant group A streptococci are susceptible to linezolid and to tedizolid. Approximately 90% of *Staphylococcus aureus* strains are sensitive to clindamycin, but resistance—both intrinsic and inducible—is increasing.

[c] Severe hospital-acquired *S. aureus* infections or community-acquired *S. aureus* infections that are not responding to the β-lactam antibiotics recommended in this table may be caused by methicillin-resistant strains, requiring vancomycin, daptomycin, or linezolid or tedizolid.

[d] Some strains of methicillin-resistant *S. aureus* (MRSA) remain sensitive to tetracycline and trimethoprim-sulfamethoxazole. Daptomycin (4 mg/kg IV q24h) or tigecycline (100-mg loading dose followed by 50 mg IV q12h) is an alternative treatment for MRSA.

by the U.S. Food and Drug Administration for uncomplicated skin and soft tissue infections including ceftaroline, dalbavancin, oritavancin, tedizolid, delafloxacin, and omadacycline.

Furuncles, carbuncles, and abscesses caused by MRSA and MSSA are common, and their treatment depends upon the size of the lesion. Furuncles <2.5 cm in diameter are usually treated with moist heat. Those that are larger (4.5 cm of erythema and induration) require surgical drainage, and the occurrence of these larger lesions when associated with fever, chills, or leukocytosis requires both drainage and antibiotic treatment. Previous studies in children demonstrated that surgical drainage of abscesses (mean diameter, 3.8 cm) was as effective when used alone as when combined with trimethoprim-sulfamethoxazole treatment. However, the rate of recurrence of new lesions was lower in the group undergoing both drainage and antibiotic treatment. A more recent study in patients with predominantly MRSA localized abscesses suggested that a 7- to 10-day course of treatment with trimethoprim-sulfamethoxazole or clindamycin was associated with higher cure rates and fewer recurrences.

Early and aggressive surgical exploration is essential in cases of suspected necrotizing fasciitis, myositis, or gangrene to (1) visualize the deep structures, (2) remove necrotic tissue, (3) reduce compartment pressure, and (4) obtain suitable material for Gram staining and for aerobic and anaerobic cultures and antimicrobial sensitivity testing. Appropriate empirical antibiotic treatment for mixed aerobic–anaerobic infections is reviewed in **Chap. 159**. Group A streptococcal and clostridial infection of the fascia and/or muscle carries a mortality rate of 20–50% with penicillin treatment. In experimental models of streptococcal and clostridial necrotizing fasciitis/myositis, clindamycin has exhibited markedly superior efficacy, but no comparative clinical trials have been performed. A retrospective study of children with invasive group A streptococcal infection demonstrated higher survival rates with clindamycin treatment than with β-lactam antibiotic therapy. Current guidelines recommend treatment with clindamycin (or linezolid) plus a β-lactam antibiotic. The emergence of macrolide resistance among group A streptococci has, however, created therapeutic dilemmas for patients allergic to penicillin and for those with life-threatening infections, such as necrotizing fasciitis/myonecrosis and streptococcal toxic shock syndrome, where clindamycin has been the treatment of choice. In such instances, linezolid or tedizolid could be used.

The use of intravenous immune globulin (IVIg) in patients with invasive group A streptococcal infections has been controversial, and early reports failed to demonstrate efficacy or were terminated early due to low enrollment. More recent prospective studies and one five-study meta-analysis demonstrated that administration of IVIg (often in combination with clindamycin) significantly reduced 30- and 90-day mortality. The combination of clindamycin and IVIg may help reduce production and activities of circulating toxins produced by group A *Streptococcus*. IVIg could be given early, and more than one dose should be used because different batches of IVIg have variable neutralizing activity against streptococcal exotoxins.

Hyperbaric oxygen (HBO) treatment has been suggested to be useful in gas gangrene due to clostridial species. However, no definitive comparative trials of HBO have been done for the treatment for invasive group A streptococcal infections, although some reports state that such treatment reduces mortality and the need for further debridements. The use of HBO should not delay, or be used in preference to, surgical debridement when the latter is indicated; such delays significantly contribute to mortality. Antibiotic treatment should be continued until all signs of systemic toxicity have resolved, all devitalized tissue has been removed, and granulation tissue has developed **(Chaps. 153, 159, and 182)**.

In summary, infections of the skin and soft tissues are diverse in presentation and severity and offer a great challenge to the clinician. This chapter provides an approach to diagnosis and understanding of the pathophysiologic mechanisms involved in these infections. More in-depth information is found in chapters on specific infections.

FURTHER READING

Aronoff DM, Marrazzo JM: Infections caused by *Clostridium perfringens* and *Paeniclostridium sordellii* after unsafe abortion. Lancet Infect Dis 23:e48, 2023.

Bruun T et al: Risk factors and predictors of mortality in streptococcal necrotizing soft-tissue infections: A multicenter prospective study. Clin Infect Dis 72:293, 2021.

Bryant AE et al: Emerging erythromycin and clindamycin resistance in group A *Streptococcus*: Efficacy of linezolid and tedizolid in experimental necrotizing infection. J Glob Antimicrob Resist 22:601, 2020.

Daum RS et al: A placebo-controlled trial of antibiotics for smaller skin abscesses. N Engl J Med 376:2545, 2017.

Davies MR et al: Emergence of scarlet fever *Streptococcus pyogenes emm*12 clones in Hong Kong is associated with toxin acquisition and multidrug resistance. Nat Genet 47:84, 2015.

Gessain A et al: Monkeypox. N Engl J Med 387:1783, 2022.

Linner A et al: Clinical efficacy of polyspecific intravenous immunoglobulin therapy in patients with streptococcal toxic shock syndrome: A comparative observational study. Clin Infect Dis 59:851, 2014.

Rafei R et al: A global snapshot on the prevalent macrolide-resistant emm types of group A *Streptococcus* worldwide, their phenotypes and their resistance marker genotypes during the last two decades: A systematic review. Infect Genet Evol 99:105258, 2022.

Stevens DL et al: Necrotizing soft tissue infections. Infect Dis Clin North Am 35:135, 2021.

Talan DA et al: Bacteriologic analysis of infected dog and cat bites. Emergency Medicine Animal Bite Infection Study Group. N Engl J Med 340:85, 1999.

135

Infectious Arthritis

Nongnooch Poowanawittayakom,
Lawrence C. Madoff

Although *Staphylococcus aureus*, streptococci, and *Neisseria gonorrhoeae* are the most common causes of infectious arthritis, various mycobacteria, spirochetes, fungi, and viruses also infect joints (Table 135-1). Since acute bacterial infection can destroy articular cartilage rapidly, all inflamed joints must be evaluated without delay to exclude noninfectious processes and determine appropriate antimicrobial therapy and drainage procedures. For more detailed information on infectious arthritis caused by specific organisms, the reader is referred to the chapters on those organisms.

Acute bacterial infection typically involves a single joint or a few joints. Subacute or chronic monoarthritis or oligoarthritis suggests mycobacterial or fungal infection; episodic inflammation is seen in syphilis, Lyme disease, and the reactive arthritis that follows enteric infections and chlamydial urethritis. Acute polyarticular inflammation occurs as an immunologic reaction during the course of endocarditis, rheumatic fever, disseminated neisserial infection, and acute viral hepatitis. Viruses often infect multiple joints; however, bacterial infections generally cause mono- or oligoarthritis except in persons with underlying diseases such as rheumatoid arthritis.

TABLE 135-1 Differential Diagnosis of Arthritis Syndromes

ACUTE MONARTICULAR ARTHRITIS	CHRONIC MONARTICULAR ARTHRITIS	POLYARTICULAR ARTHRITIS
Staphylococcus aureus *Streptococcus pneumoniae* β-Hemolytic streptococci Gram-negative bacilli *Neisseria gonorrhoeae* *Candida* spp. Crystal-induced arthritis Fracture Hemarthrosis Foreign body Osteoarthritis Ischemic necrosis Monoarticular rheumatoid arthritis	*Mycobacterium tuberculosis* Nontuberculous mycobacteria *Borrelia burgdorferi* *Treponema pallidum* *Candida* spp. *Sporothrix schenckii* *Coccidioides immitis* *Blastomyces dermatitidis* *Aspergillus* spp. *Cryptococcus neoformans* *Nocardia* spp. *Brucella* spp. Legg-Calvé-Perthes disease Osteoarthritis	*Neisseria meningitidis* *N. gonorrhoeae* Nongonococcal bacterial arthritis Bacterial endocarditis *Candida* spp. Poncet's disease (tuberculous rheumatism) Hepatitis B virus Parvovirus B19 HIV Human T-lymphotropic virus type 1 Rubella virus Arthropod-borne viruses Sickle cell disease flare Reactive arthritis Serum sickness Acute rheumatic fever Inflammatory bowel disease Systemic lupus erythematosus Rheumatoid arthritis/ Still's disease Other vasculitides Sarcoidosis

APPROACH TO THE PATIENT

Infectious Arthritis

Aspiration of synovial fluid or arthrocentesis—an essential element in the evaluation of potentially infected joints—can be performed without difficulty in most cases by the insertion of a large-bore needle into the site of maximal fluctuance or tenderness or by the route of easiest access. Ultrasonography or computed tomography (CT) may be used to guide aspiration of difficult-to-localize effusions of the hip and, occasionally, the shoulder and other joints. Normal synovial fluid contains <180 cells (predominantly mononuclear cells) per microliter. Synovial cell counts averaging 100,000/μL (range, 25,000–250,000/μL), with >90% neutrophils, are characteristic of acute bacterial infections. Crystal-induced, rheumatoid, and other noninfectious inflammatory arthritides usually are associated with <30,000–50,000 cells/μL; cell counts of 10,000–30,000/μL, with 50–70% neutrophils and the remainder lymphocytes, are common in mycobacterial and fungal infections. Definitive diagnosis of an infectious process relies on identification of the pathogen in stained smears of synovial fluid, isolation of the pathogen from cultures of synovial fluid and blood, or detection of microbial nucleic acids and proteins by nucleic acid amplification tests (NAATs) and immunologic techniques. Gram stain is positive in about 30–50% of cases, and synovial fluid culture is positive in >60% of nongonococcal bacterial arthritis cases. Matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) mass spectrometry is helpful in patients who have negative culture and high suspicion of infectious arthritis. Sonication of explanted prosthetic joints (placement of the material into liquid and then immersion in an ultrasound bath) increases the yield of organism detection, especially in the case of prior antibiotic use within 14 days.

ACUTE BACTERIAL ARTHRITIS

PATHOGENESIS

Bacteria enter the joint from the bloodstream; from a contiguous site of infection in bone or soft tissue; or by direct inoculation during surgery, injection, animal or human bite, or trauma. In hematogenous infection, bacteria escape from synovial capillaries, which have no limiting basement membrane, and within hours provoke neutrophilic infiltration of the synovium. Neutrophils and bacteria enter the joint space; later, bacteria adhere to articular cartilage. Degradation of cartilage begins within 48 h as a result of increased intraarticular pressure, release of proteases and cytokines from chondrocytes and synovial macrophages, and invasion of the cartilage by bacteria and inflammatory cells. Histologic studies reveal bacteria lining the synovium and cartilage as well as abscesses extending into the synovium, cartilage, and—in severe cases—subchondral bone. Synovial proliferation results in the formation of a pannus over the cartilage, and thrombosis of inflamed synovial vessels develops. Bacterial factors that appear important in the pathogenesis of infective arthritis include various surface-associated adhesins in *S. aureus* that permit adherence to cartilage and endotoxins that promote chondrocyte-mediated breakdown of cartilage.

MICROBIOLOGY

The hematogenous route of infection is the most common route in all age groups, and nearly every bacterial pathogen is capable of causing septic arthritis. In infants, group B streptococci, gram-negative enteric bacilli, and *S. aureus* are the most common pathogens. Since the advent of the *Haemophilus influenzae* vaccine, the predominant causes among children <5 years of age have been *S. aureus*, *Streptococcus pyogenes* (group A *Streptococcus*), and (in some centers) *Kingella kingae*. Among young adults and adolescents, *N. gonorrhoeae* is the most commonly implicated organism. *S. aureus* (including methicillin-resistant *S. aureus* [MRSA]) accounts for most nongonococcal isolates in adults of all ages; gram-negative bacilli, pneumococci, and β-hemolytic streptococci—particularly groups A and B but also groups C, G, and F—are involved in up to one-third of cases in older adults, especially those with underlying comorbid illnesses. Gram-negative bacilli such as *Pseudomonas* may occur in immunocompromised patients or intravenous drug users.

Infections after surgical procedures or penetrating injuries are due most often to *S. aureus* and occasionally to other gram-positive bacteria or gram-negative bacilli. Infections with coagulase-negative staphylococci are unusual except after the implantation of prosthetic joints or arthroscopy. Anaerobic organisms, often in association with aerobic or facultative bacteria, are found after human bites and when decubitus ulcers or intraabdominal abscesses spread into adjacent joints. Polymicrobial infections complicate traumatic injuries with extensive contamination. Bites and scratches from cats and other animals may introduce *Pasteurella multocida* or *Bartonella henselae* into joints either directly or hematogenously, and bites from humans may introduce *Eikenella corrodens* or other components of the oral flora. Penetration of a sharp object through a shoe is associated with *Pseudomonas aeruginosa* arthritis in the foot.

NONGONOCOCCAL BACTERIAL ARTHRITIS

Epidemiology Although hematogenous infections with virulent organisms such as *S. aureus*, *H. influenzae*, and pyogenic streptococci occur in healthy persons, there is an underlying host predisposition in many cases of septic arthritis. Patients with rheumatoid arthritis have the highest incidence of infective arthritis (most often secondary to *S. aureus*) because of chronically inflamed joints; glucocorticoid therapy; and frequent breakdown of rheumatoid nodules, vasculitic ulcers, and skin overlying deformed joints. Diabetes mellitus, glucocorticoid therapy, hemodialysis, intravenous drug use, and malignancy all carry an increased risk of infection with *S. aureus* and gram-negative bacilli. Tumor necrosis factor inhibitors (e.g., etanercept, infliximab), which increasingly are used for the treatment of rheumatoid arthritis, predispose to mycobacterial infections and possibly to other pyogenic bacterial infections and could be associated with septic arthritis in this population. Pneumococcal infections complicate alcoholism, deficiencies of humoral immunity, and hemoglobinopathies. Pneumococci, *Salmonella* species, and *H. influenzae* cause septic arthritis in persons infected with HIV. Persons with primary immunoglobulin deficiency are at risk for mycoplasmal arthritis, which, while rare, results in permanent joint damage if tetracycline and replacement therapy with IV immunoglobulin are not administered promptly. IV drug users acquire staphylococcal and streptococcal infections from their own flora and acquire pseudomonal and other gram-negative infections from drugs and injection paraphernalia.

Clinical Manifestations Patients with acute septic arthritis usually present with joint pain often with limitation of passive and active joint movement, joint swelling, and/or erythema. Approximately 90% of patients present with involvement of a single joint—most commonly the knee; less frequently the hip; and still less often the shoulder, wrist, or elbow. Small joints of the hands and feet are more likely to be affected after direct inoculation or a bite. Among IV drug users, infections of the spine, sacroiliac joints, and sternoclavicular joints **(Fig. 135-1)** are more common than infections of the appendicular skeleton. Polyarticular infection is most common among patients with rheumatoid arthritis and may resemble a flare of the underlying disease.

The usual presentation consists of moderate to severe pain that is uniform around the joint, effusion, muscle spasm, and decreased range of motion. Fever in the range of 38.3–38.9°C (101–102°F) and sometimes higher is common but may not be present, especially in persons with rheumatoid arthritis, renal or hepatic insufficiency, or conditions requiring immunosuppressive therapy. The inflamed, swollen joint is usually evident on examination except in the case of a deeply situated joint such as the hip, shoulder, or sacroiliac joint. Cellulitis, bursitis, and acute osteomyelitis, which may produce a similar clinical picture, should be distinguished from septic arthritis by preservation of passive range of motion and less-than-circumferential swelling. A focus of extraarticular infection, such as a boil, pneumonia, or endocarditis, should be sought. Peripheral-blood leukocytosis with a left shift and elevation of the erythrocyte sedimentation rate or C-reactive protein level are common.

Plain radiographs show evidence of soft tissue swelling, joint space widening, and displacement of tissue planes by the distended capsule. Narrowing of the joint space and bony erosions indicate advanced infection and a poor prognosis. Ultrasound is useful for detecting effusions in the hip, and CT or MRI can demonstrate infections of the sacroiliac joint, the sternoclavicular joint, and the spine very well.

Laboratory Findings Specimens of peripheral blood and synovial fluid should be obtained before antibiotics are administered. Blood cultures are positive in up to 50–70% of *S. aureus* infections but are less frequently positive in infections due to other organisms. The synovial fluid is turbid, serosanguineous, or frankly purulent. Gram-stained smears confirm the presence of large numbers of neutrophils. Levels of total protein and lactate dehydrogenase in synovial fluid are elevated,

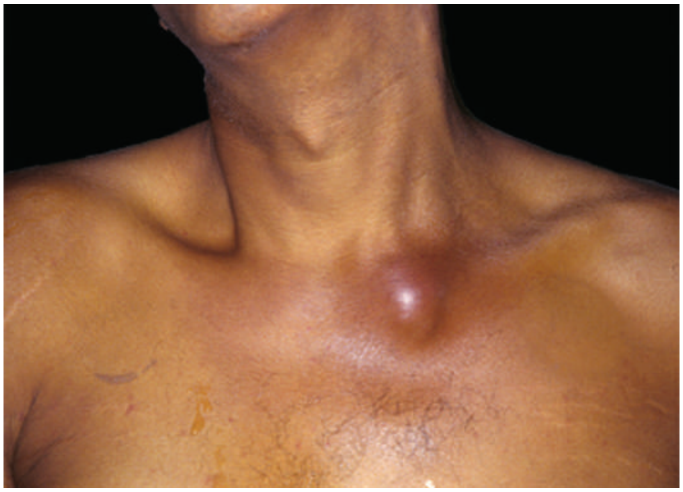

FIGURE 135-1 Acute septic arthritis of the sternoclavicular joint. A man in his forties with a history of cirrhosis presented with a new onset of fever and lower neck pain. He had no history of IV drug use or previous catheter placement. Jaundice and a painful swollen area over his left sternoclavicular joint were evident on physical examination. Cultures of blood drawn at admission grew group B *Streptococcus*. The patient recovered after treatment with IV penicillin. *(Courtesy of the late Francisco M. Marty, MD, Brigham and Women's Hospital, Boston; with permission.)*

and the glucose level is depressed; however, these findings are not specific for infection, and measurement of these levels is not necessary for diagnosis. The synovial fluid should be examined for crystals because gout and pseudogout can resemble septic arthritis clinically, and infection and crystal-induced disease occasionally occur together. Organisms are seen on synovial fluid smears in nearly three-quarters of infections with *S. aureus* and streptococci and in 30–50% of infections due to gram-negative and other bacteria. Cultures of synovial fluid are positive in >90% of cases. Inoculation of synovial fluid into bottles containing liquid media for blood cultures increases the yield of a culture, especially if the pathogen is a fastidious organism or the patient is taking an antibiotic, but should be interpreted in the context of the Gram's stain result. Pathogen nucleic acid amplification based assays (NAAT) or MALDI-TOF mass spectrometry, when available, can be useful for the diagnosis of partially treated or culture-negative arthritis. Inflammatory markers such as erythrocyte sedimentation rate and C-reactive protein tend to be elevated in septic arthritis but are nonspecific. Serum procalcitonin elevation is only ~50% sensitive and should not be used to rule out infectious arthritis. Synovial fluid procalcitonin might be useful, although data are limited.

TREATMENT

Nongonococcal Bacterial Arthritis

Prompt administration of systemic antibiotics and drainage of the involved joint can prevent destruction of cartilage, postinfectious degenerative arthritis, joint instability, or deformity. Once samples of blood and synovial fluid have been obtained for culture, empirical antibiotics should be directed against the bacteria visualized on smears or the pathogens that are likely in light of the patient's age and risk factors. Initial therapy should consist of IV-administered bactericidal agents; direct instillation of antibiotics into the joint is not necessary to achieve adequate levels in synovial fluid and tissue. If there are gram-positive cocci on the smear, IV vancomycin (15–20 mg/kg/dose) every 8–12 h should be started empirically. If methicillin-resistant *S. aureus* is an unlikely pathogen (e.g., when it is not widespread in the community), cefazolin (2 g every 8 h), oxacillin (2 g every 4 h), or nafcillin (2 g every 4 h) should be given.

If initial Gram's stain shows gram-negative bacilli, the risk for *P. aeruginosa* infection should be evaluated. If the patient does not have risk factors for *P. aeruginosa* infection, an IV third-generation cephalosporin such as cefotaxime (2 g every 8 h) or ceftriaxone (2 g every 24 h) provides adequate empirical coverage for most community-acquired infections. In addition, if the patient has a higher risk for *P. aeruginosa* infection, then anti-pseudomonal coverage such as cefepime (2 g every 8–12 h) or ceftazidime (2 g every 8 h) should be given. Double coverage of *Pseudomonas* with cephalosporin and ciprofloxacin or aminoglycoside can be considered empirically in severely ill patients or in the setting of highly resistant *Pseudomonas* infection incidence.

Definitive therapy is based on the identity and antibiotic susceptibility of the bacteria isolated in culture. Infections due to staphylococci are treated with cefazolin, oxacillin, nafcillin, or vancomycin for 4 weeks. In patients without evidence of endocarditis, IV antibiotics can be used for 7–14 days of treatment followed by oral antibiotics to complete the treatment course. Pneumococcal and streptococcal infections due to penicillin-susceptible organisms respond to 2 weeks of therapy with penicillin G (2 million units IV every 4 h); infections caused by *H. influenzae* and by strains of *Streptococcus pneumoniae* that are resistant to penicillin are treated with cefotaxime or ceftriaxone for 2 weeks. Most enteric gram-negative infections can be cured in 3–4 weeks by a second- or third-generation cephalosporin given IV or by a fluoroquinolone such as levofloxacin (500 mg IV or PO every 24 h). *P. aeruginosa* infection should be treated for at least 2 weeks with antipseudomonal cephalosporin such as ceftazidime (2 g IV every 8 h), cefepime (2 g every 8–12 h). If tolerated, this regimen is continued for an additional 2 weeks; alternatively, a fluoroquinolone such as ciprofloxacin (750 mg PO twice daily) may be given.

Timely drainage of pus and necrotic debris from the infected joint is required for a favorable outcome. Needle aspiration of readily accessible joints such as the knee may be adequate if loculations or particulate matter in the joint does not prevent its thorough decompression. Arthroscopic drainage and lavage may be employed initially or within several days if repeated needle aspiration fails to relieve symptoms, decrease the volume of the effusion and the synovial white cell count, and clear bacteria from smears and cultures. In some cases, arthrotomy is necessary to remove loculations and debride infected synovium, cartilage, or bone. Septic arthritis of the hip is best managed with arthrotomy, particularly in young children, in whom infection threatens the viability of the femoral head. Septic joints do not require immobilization except for pain control before symptoms are alleviated by treatment. Weight bearing should be avoided until signs of inflammation have subsided, but frequent passive motion of the joint is indicated to maintain full mobility. Although some clinical studies suggest that adjunctive use of glucocorticoids showed some benefit in children with septic arthritis, they are not widely used in clinical practice.

GONOCOCCAL ARTHRITIS

Epidemiology In the past, gonococcal arthritis **(Chap. 161)** accounted for up to 70% of episodes of infectious arthritis in persons <40 years of age in the United States. As the rates of mucosal gonorrhea have fallen in the United States, it is likely that the proportion of septic arthritis caused by *N. gonorrhoeae* also has fallen considerably. Arthritis due to *N. gonorrhoeae* is a consequence of bacteremia arising from gonococcal infection or, more frequently, from asymptomatic gonococcal mucosal colonization of the urethra, cervix, or pharynx. Women are at greatest risk during menses and during pregnancy and overall are two to three times more likely than men to develop disseminated gonococcal infection (DGI) and arthritis. Persons with complement deficiencies, especially of the terminal components, are prone to recurrent episodes of gonococcemia. Eculizumab, which is a long-acting monoclonal antibody targeting the C5 complement component and used mainly for the treatment of paroxysmal nocturnal hemoglobinuria, also has been reported to be associated with disseminated gonococcal infection.

Clinical Manifestations and Laboratory Findings The most common manifestation of DGI is a syndrome of arthritis-dermatitis. Patients present with fever, chills, rash, tenosynovitis, and articular symptoms. Small numbers of papules that progress to hemorrhagic pustules develop on the trunk and the extensor surfaces of the distal extremities. Migratory arthritis and tenosynovitis of the knees, hands, wrists, feet, and ankles are prominent. The cutaneous lesions and articular findings are believed to be the consequence of an immune reaction to circulating gonococci and immune-complex deposition in tissues. Thus, cultures of synovial fluid are consistently negative, and blood cultures are positive in <45% of patients. Synovial fluid may be difficult to obtain from inflamed joints and usually contains only 10,000–20,000 leukocytes/μL.

True gonococcal septic arthritis is less common than the DGI syndrome and always follows DGI, which is unrecognized in one-third of patients. A single joint such as the hip, knee, ankle, or wrist is usually involved. Synovial fluid, which contains >50,000 leukocytes/μL, can be obtained with ease; the gonococcus is evident only occasionally in Gram-stained smears, and cultures of synovial fluid are positive in <40% of cases. Blood cultures are almost always negative.

Because it is difficult to isolate gonococci from synovial fluid and blood, specimens for culture should be obtained from potentially infected mucosal sites. NAAT-based urine tests also may be positive. Culture requires endocervical (in female patients) or urethral (in male patients) swab specimens. Culture is available for detection of rectal, oropharyngeal, and conjunctival gonococcal infection. Cultures and Gram-stained smears of skin lesions are occasionally positive.

All specimens for culture should be plated onto Thayer-Martin agar directly or in special transport media at the bedside and transferred promptly to the microbiology laboratory in an atmosphere of 5% CO_2. NAAT assays are extremely sensitive in detecting gonococcal DNA in synovial fluid, but they are not FDA approved for this purpose. A dramatic alleviation of symptoms within 12–24 h after the initiation of appropriate antibiotic therapy supports a clinical diagnosis of the DGI syndrome if cultures are negative.

TREATMENT

Gonococcal Arthritis

Initial treatment consists of ceftriaxone (1 g IV or IM every 24 h) to cover possible penicillin-resistant organisms. Once local and systemic signs are clearly resolving, a 7-day course of antibiotics may be completed with daily IM ceftriaxone given at 500 mg daily (or 1g for those weighing over 150 kg). An oral fluoroquinolone such as ciprofloxacin (500 mg twice daily) may be used if the organism is known to be susceptible. If penicillin-susceptible organisms are isolated, amoxicillin (500 mg four times daily) may be used. Suppurative arthritis usually responds to needle aspiration of involved joints and 7–14 days of antibiotic treatment. Arthroscopic lavage or arthrotomy is rarely required. The Centers for Disease Control recommends that when chlamydial infection has not been excluded, treatment with PO doxycycline 100 mg twice daily for 7 days should be given. Sexual partners should be offered testing and presumptive treatment for gonorrhea and chlamydial infection. It is noteworthy that arthritis symptoms similar to those seen in DGI occur in meningococcemia. A dermatitis–arthritis syndrome, purulent monoarthritis, and reactive polyarthritis have been described. All respond to treatment with appropriate antibiotics.

SPIROCHETAL ARTHRITIS

LYME DISEASE

Lyme disease **(Chap. 191)** due to infection with the spirochete *Borrelia burgdorferi* causes arthritis in up to 60% of persons who are not treated. Intermittent arthralgias and myalgias—but not arthritis—occur within days or weeks of inoculation of the spirochete by the *Ixodes* tick. Later, there are three patterns of joint disease: (1) Fifty percent of untreated persons experience intermittent episodes of monoarthritis or oligoarthritis involving the knee and/or other large joints. The symptoms wax and wane without treatment over months, and each year, 10–20% of patients report loss of joint symptoms. (2) Twenty percent of untreated persons develop a pattern of waxing and waning arthralgias. (3) Ten percent of untreated patients develop chronic inflammatory synovitis that results in erosive lesions and destruction of the joint. Serologic tests for IgG antibodies to *B. burgdorferi* are positive in >90% of persons with Lyme arthritis, and a NAAT detects *Borrelia* DNA in synovial fluid in 85% of patients.

TREATMENT

Lyme Arthritis

Lyme arthritis generally responds well to therapy. A regimen of oral doxycycline (100 mg twice daily for 28 days), oral amoxicillin (500 mg three times daily for 28 days), or parenteral ceftriaxone (2 g/d for 2–4 weeks) is recommended. Patients who do not respond to a total of 2 months of oral therapy or 1 month of parenteral therapy are unlikely to benefit from additional antibiotic therapy and are treated with anti-inflammatory agents or synovectomy. Failure of therapy is associated with host features such as the human leukocyte antigen DR4 (HLA-DR4) genotype, persistent reactivity to OspA (outer-surface protein A), and the presence of hLFA-1 (human leukocyte function–associated antigen 1), which cross-reacts with OspA.

SYPHILITIC ARTHRITIS

Articular manifestations occur in different stages of syphilis **(Chap. 187)**. In early congenital syphilis, periarticular swelling and immobilization of the involved limbs (*Parrot's pseudoparalysis*) complicate osteochondritis of long bones. *Clutton's joint*, a late manifestation of congenital syphilis that typically develops between ages 8 and 15 years, is caused by chronic painless synovitis with effusions of large joints, particularly the knees and elbows. Secondary syphilis may be associated with arthralgias, with symmetric arthritis of the knees and ankles and occasionally of the shoulders and wrists, and with sacroiliitis. The arthritis follows a subacute to chronic course with a mixed mononuclear and neutrophilic synovial-fluid pleocytosis (typical cell counts, 5000–15,000/μL). Immunologic mechanisms may contribute to the arthritis, and symptoms usually improve rapidly with penicillin therapy. In tertiary syphilis, Charcot joint results from sensory loss due to tabes dorsalis. Penicillin is not helpful in this setting.

MYCOBACTERIAL ARTHRITIS

Tuberculous arthritis **(Chap. 183)** accounts for ~1% of all cases of tuberculosis and 10% of extrapulmonary cases. The most common presentation is chronic granulomatous monoarthritis. An unusual syndrome, *Poncet's disease*, is a reactive symmetric form of polyarthritis that affects persons with visceral or disseminated tuberculosis. No mycobacteria are found in the joints, and symptoms resolve with antituberculous therapy.

Unlike tuberculous osteomyelitis (*Pott's disease*) **(Chap. 136)**, which typically involves the thoracic and lumbar spine (50% of cases), tuberculous arthritis primarily involves the large weight-bearing joints, in particular the hips, knees, and ankles, and only occasionally involves smaller non-weight-bearing joints. Progressive monoarticular swelling and pain develop over months or years, and systemic symptoms are seen in only half of all cases. Tuberculous arthritis occurs as part of a disseminated primary infection or through late reactivation, often in persons with HIV infection or other immunocompromised hosts. Coexistent active pulmonary tuberculosis is unusual.

Aspiration of the involved joint yields fluid with an average cell count of 20,000/μL, with ~50% neutrophils. Acid-fast staining of the fluid yields positive results in fewer than one-third of cases, and cultures are positive in 80%. Culture of synovial tissue taken at biopsy is positive in ~90% of cases and shows granulomatous inflammation in most. NAAT can shorten the time to diagnosis to 1 or 2 days. Radiographs reveal peripheral erosions at the points of synovial attachment, periarticular osteopenia, and eventually joint-space narrowing. Therapy for tuberculous arthritis is the same as that for tuberculous pulmonary disease, requiring the administration of multiple agents for 6–9 months. Therapy is more prolonged in immunosuppressed individuals, such as those infected with HIV.

Various atypical mycobacteria **(Chap. 185)** found in water and soil may cause chronic indolent arthritis. Such disease results from trauma and direct inoculation associated with farming, gardening, or aquatic activities. Smaller joints, such as the digits, wrists, and knees, are usually involved. Involvement of tendon sheaths and bursae is typical. The mycobacterial species involved include *Mycobacterium marinum*, *M. avium complex*, *M. terrae*, *M. kansasii*, *M. fortuitum*, and *M. chelonae*. In persons who have HIV infection or are receiving immunosuppressive therapy, hematogenous spread to the joints has been reported for *M. kansasii*, *M. avium* complex, and *M. haemophilum*. Diagnosis usually requires biopsy and culture, and therapy is based on antimicrobial susceptibility patterns.

FUNGAL ARTHRITIS

Fungi are an unusual cause of chronic monoarticular arthritis. Granulomatous articular infection with the endemic dimorphic fungi *Coccidioides immitis*, *Blastomyces dermatitidis*, and (less commonly) *Histoplasma capsulatum* **(Fig. 135-2)** results from hematogenous seeding or direct extension from bony lesions in persons with disseminated disease. Joint involvement is an unusual complication of sporotrichosis (infection with *Sporothrix schenckii*) among gardeners and other persons who work with soil or sphagnum moss. Articular sporotrichosis

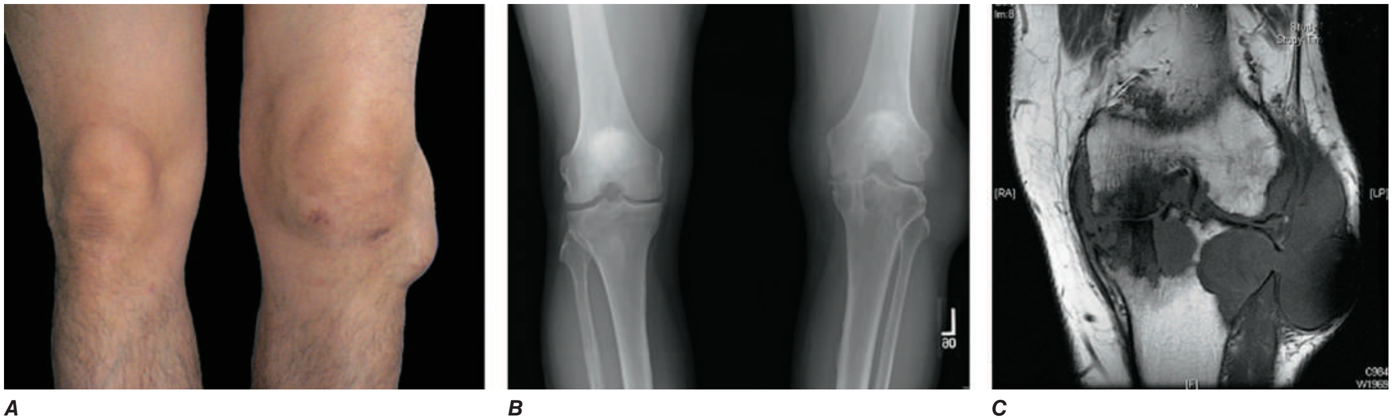

FIGURE 135-2 Chronic arthritis caused by *Histoplasma capsulatum* in the left knee. ***A.*** A man in his sixties from El Salvador presented with a history of progressive knee pain and difficulty walking for several years. He had undergone arthroscopy for a meniscal tear 7 years before presentation (without relief) and had received several intraarticular glucocorticoid injections. The patient developed significant deformity of the knee over time, including a large effusion in the lateral aspect. ***B.*** An x-ray of the knee showed multiple abnormalities, including severe medial femorotibial joint-space narrowing, several large subchondral cysts within the tibia and the patellofemoral compartment, a large suprapatellar joint effusion, and a large soft tissue mass projecting laterally over the knee. ***C.*** MRI further defined these abnormalities and demonstrated the cystic nature of the lateral knee abnormality. Synovial biopsies demonstrated chronic inflammation with giant cells, and cultures grew *H. capsulatum* after 3 weeks of incubation. All clinical cystic lesions and the effusion resolved after 1 year of treatment with itraconazole. The patient underwent a left total-knee replacement for definitive treatment. *(Courtesy of the late Francisco M. Marty, MD, Brigham and Women's Hospital, Boston; with permission.)*

is six times more common among men than among women, and alcoholics and other debilitated hosts are at risk for polyarticular infection.

Candida infection involving a single joint—usually the knee, hip, or shoulder—results from surgical procedures, intraarticular injections, or (among critically ill patients with debilitating illnesses such as diabetes mellitus or hepatic or renal insufficiency and patients receiving immunosuppressive therapy) hematogenous spread. *Candida* infections in IV drug users typically involve the spine, sacroiliac joints, or other fibrocartilaginous joints. Unusual cases of arthritis due to *Aspergillus* species, *Cryptococcus neoformans*, *Pseudallescheria boydii*, and the dematiaceous fungi also have resulted from direct inoculation or disseminated hematogenous infection in immunocompromised persons. In the United States, a 2012 national outbreak of fungal arthritis (and meningitis) caused by *Exserohilum rostratum* was linked to intraspinal and intraarticular injection of a contaminated preparation of methylprednisolone acetate.

The synovial fluid in fungal arthritis usually contains 10,000–40,000 cells/μL, with ~70% neutrophils. Stained specimens and cultures of synovial tissue often confirm the diagnosis of fungal arthritis when studies of synovial fluid give negative results. Treatment consists of drainage and lavage of the joint and systemic administration of an antifungal agent directed at a specific pathogen. The doses and duration of therapy are the same as for disseminated disease **(see Part 5, Section 16)**. In fungal prosthetic joint infection, the removal of all prosthetic joint material is highly recommended.

VIRAL ARTHRITIS

Viruses produce arthritis by infecting synovial tissue during systemic infection or by provoking an immunologic reaction that involves joints. As many as 50% of women report persistent arthralgias and 10% report frank arthritis within 3 days of the rash that follows natural infection with rubella virus and within 2–6 weeks after receipt of live-virus vaccine. Episodes of symmetric inflammation of fingers, wrists, and knees uncommonly recur for >1 year, but a syndrome of chronic fatigue, low-grade fever, headaches, and myalgias can persist for months or years. IV immunoglobulin has been helpful in selected cases. Self-limited monoarticular or migratory polyarthritis may develop within 2 weeks of the parotitis of mumps; this sequela is more common among men than women. Approximately 10% of children and 60% of women develop polyarthritis or polyarthralgia in small joints after infection with parvovirus B19. In adults, arthropathy sometimes occurs without fever or rash. Pain and stiffness, with less prominent swelling (primarily of the hands but also of the knees, wrists, and ankles), usually resolve within weeks, although a small proportion of patients develop chronic arthropathy.

About 2 weeks before the onset of jaundice, up to 10% of persons with acute hepatitis B develop an immune complex–mediated, serum sickness–like reaction with maculopapular rash, urticaria, fever, and arthralgias. Less common developments include symmetric arthritis involving the hands, wrists, elbows, or ankles and morning stiffness that resembles a flare of rheumatoid arthritis. Symptoms resolve at the time jaundice develops. Many persons with chronic hepatitis C infection report persistent arthralgia or arthritis, both in the presence and in the absence of cryoglobulinemia.

Painful arthritis often accompanies the fever and rash of several arthropod-borne viral infections, including those caused by Zika, chikungunya, O'nyong-nyong, Ross River, Mayaro, and Barmah Forest viruses **(Chap. 215)**. Symmetric arthritis involving the hands and wrists may occur during the convalescent phase of infection with lymphocytic choriomeningitis virus. Patients infected with an enterovirus frequently report arthralgias, and echovirus has been isolated from patients with acute polyarthritis.

Several arthritis syndromes are associated with HIV infection. An incomplete form of reactive arthritis with painful lower-extremity oligoarthritis may follow an episode of urethritis in HIV-infected persons. HIV-associated reactive arthritis appears to be extremely common among persons with the HLA-B27 haplotype, but sacroiliac joint disease is unusual and is seen mostly in the absence of HLA-B27. Up to one-third of HIV-infected persons with psoriasis develop psoriatic arthritis. Painless monoarthropathy and persistent symmetric polyarthropathy occasionally complicate HIV infection. Chronic persistent oligoarthritis of the shoulders, wrists, hands, and knees occurs in women infected with human T-lymphotropic virus type 1. Synovial thickening, destruction of articular cartilage, and leukemic-appearing atypical lymphocytes in synovial fluid are characteristic, but progression to T cell leukemia is unusual.

PARASITIC ARTHRITIS

Arthritis due to parasitic infection is rare. The guinea worm *Dracunculus medinensis* may cause destructive joint lesions in the lower extremities as migrating gravid female worms invade joints or cause ulcers in adjacent soft tissues that become secondarily infected. Hydatid cysts infect bones in 1–2% of cases of infection with *Echinococcus granulosus*. The expanding destructive cystic lesions may spread to and destroy adjacent joints, particularly the hip and pelvis. In rare cases, chronic synovitis has been associated with the presence of schistosomal eggs in synovial biopsies. Monoarticular arthritis in children with lymphatic filariasis appears to respond to therapy with diethylcarbamazine even in the absence of microfilariae in synovial fluid. Reactive arthritis has been attributed to hookworm, *Strongyloides*, *Cryptosporidium*, and *Giardia* infection in case reports, but confirmation is required.

POSTINFECTIOUS OR REACTIVE ARTHRITIS

Reactive polyarthritis develops several weeks after ~1% of cases of nongonococcal urethritis and 2% of enteric infections, particularly those due to *Yersinia enterocolitica*, *Shigella flexneri*, *Campylobacter jejuni*, *Clostridioides difficile*, and *Salmonella* species. Only a minority of these patients have the other findings of classic reactive arthritis, including urethritis, conjunctivitis, uveitis, oral ulcers, and rash. Studies have identified microbial DNA or antigen in synovial fluid or blood, but the pathogenesis of this condition is poorly understood. The arthritis may occur several days or weeks after the infection and can be associated with dactylitis, enthesitis, or extraarticular involvement such as conjunctivitis.

Reactive arthritis is most common among young men (except after *Yersinia* infection) and has been linked to the HLA-B27 locus as a potential genetic predisposing factor. Patients report painful, asymmetric oligoarthritis that affects mainly the knees, ankles, and feet. Low back pain is common, and radiographic evidence of sacroiliitis is found in patients with long-standing disease. Most patients recover within 6 months, but prolonged recurrent disease is more common in cases that follow chlamydial urethritis. Anti-inflammatory agents help relieve symptoms, but the role of prolonged antibiotic therapy in eliminating microbial antigen from the synovium is controversial.

Migratory polyarthritis and fever constitute the usual presentation of acute rheumatic fever in adults **(Chap. 371)**. This presentation is distinct from that of poststreptococcal reactive arthritis, which also follows infections with group A *Streptococcus* but is not migratory, lasts beyond the typical 3-week maximum of acute rheumatic fever, and responds poorly to aspirin.

INFECTIONS IN PROSTHETIC JOINTS

Infection complicates 0.5–2% of total joint replacements. Prosthetic joint infection occurs more often in knee arthroplasty compared with hip arthroplasty. The majority of infections are acquired intraoperatively or immediately postoperatively as a result of wound breakdown or infection; less commonly, these joint infections develop later after joint replacement and are the result of hematogenous spread or direct inoculation. The presentation may be acute, with fever, pain, and local signs of inflammation, especially in infections due to *S. aureus*, pyogenic streptococci, and gram-negative bacilli. Alternatively, infection may persist for months or years without causing constitutional symptoms when less virulent organisms—such as coagulase-negative staphylococci, *Cutibacterium* (formerly *Propionibacterium*) species, enterococci, or diphtheroids—are involved. Such indolent infections usually are acquired during joint implantation and are discovered during evaluation of chronic unexplained pain or after a radiograph shows loosening of the prosthesis; the erythrocyte sedimentation rate and C-reactive protein level are usually elevated in such cases.

The diagnosis is best made by needle aspiration of the joint; accidental introduction of organisms during aspiration must be avoided meticulously. Synovial fluid pleocytosis with a predominance of polymorphonuclear leukocytes is highly suggestive of infection, since other inflammatory processes uncommonly affect prosthetic joints. Culture and Gram's stain usually yield the responsible pathogen. Sonication of explanted prosthetic material can improve the yield of culture, presumably by breaking up bacterial biofilms on the surfaces of prostheses. Semiquantitative culture should be obtained since a low number of organisms can represent contaminants. Molecular diagnostic testing of synovial fluid, sonicate fluid, or periprosthetic tissue also may improve the yield when routine culture is negative; however, it is not always available in microbiology laboratories. Use of special media for unusual pathogens such as fungi, atypical mycobacteria, and *Mycoplasma* may be necessary if routine and anaerobic cultures are negative. Molecular diagnostic results should be interpreted cautiously since they may represent colonization rather than a true pathogen. Testing of synovial fluid for alpha-defensin, C-reactive protein, leukocyte esterase, and calprotectin can be considered in challenging cases. These tests must be interpreted in the context of the synovial fluid leukocyte count and neutrophil percentage.

TREATMENT

Prosthetic Joint Infections

Treatment includes surgery and high doses of parenteral antibiotics, which are given for 4–6 weeks because bone is usually involved. In most cases, the prosthesis must be removed and replaced to cure the infection. Implantation of a new prosthesis is best delayed for several weeks or months because relapses of infection occur most commonly within this time frame. In some cases, reimplantation is not possible, and the patient must manage without a joint, with a fused joint, or even with amputation. Cure of infection without removal of the prosthesis is occasionally possible in cases that are due to streptococci or pneumococci and that lack radiologic evidence of loosening of the prosthesis. In these cases, antibiotic therapy must be initiated within several days of the onset of infection, and the joint should be drained vigorously by open arthrotomy or arthroscopically, preferably with polyethylene liner exchange, to have a more successful outcome. In selected patients who prefer to avoid the high morbidity rate associated with joint removal and reimplantation, lifelong suppression of the infection with antibiotics may be a reasonable goal. A high cure rate with retention of the prosthesis has been reported when the combination of oral rifampin and another antibiotic (e.g., a quinolone, an antistaphylococcal penicillin, or vancomycin) is given for 3–6 months to persons with staphylococcal prosthetic joint infection of short duration. This approach is based on the ability of rifampin to kill organisms adherent to foreign material and in the stationary growth phase.

■ PREVENTION

To avoid the disastrous consequences of infection, candidates for joint replacement should be selected with care. All modifiable risk factors should be identified and minimized preoperatively to improve the surgical outcome and prevent surgical site infection. Preoperative screening for *S. aureus* with decolonization should be considered. Rates of infection are particularly high among patients with rheumatoid arthritis, persons who have undergone previous surgery on the joint, and persons with medical conditions requiring immunosuppressive therapy. Perioperative antibiotic prophylaxis, usually with cefazolin, and measures to decrease intraoperative contamination, such as laminar flow, have lowered the rates of perioperative infection to <1% in many centers. After implantation, measures should be taken to prevent or rapidly treat extraarticular infections that might give rise to hematogenous spread to the prosthesis. The effectiveness of prophylactic antibiotics for the prevention of hematogenous infection after dental procedures has not been demonstrated; in fact, viridans streptococci and other components of the oral flora are extremely unusual causes of prosthetic joint infection. Accordingly, the American Dental Association and the American Academy of Orthopaedic Surgeons do not recommend antibiotic prophylaxis for most dental patients with total joint replacements and have stated that there is no convincing evidence to support its use. Similarly, guidelines issued by the American Urological Association and the American Academy of Orthopaedic Surgeons do not recommend the use of prophylactic antibiotics for most patients with prosthetic joints who are undergoing urologic procedures but state that prophylaxis should be considered in certain situations—e.g., for patients (especially immunocompromised patients) who are undergoing a procedure posing a relatively high risk of bacteremia (e.g., lithotripsy or surgery involving bowel segments).

■ FURTHER READING

Bardin T: Gonococcal arthritis. Best Pract Res Clin Rheumatol 17:201, 2003.

Beam E, Osmon D: Prosthetic joint infection update. Infect Dis Clin North Am 32:843, 2018.

Borzio R et al: Predictors of septic arthritis in the adult population. Orthopedics 39:e657, 2016.

Franssila R, Hedman K: Infection and musculoskeletal conditions: Viral causes of arthritis. Best Pract Res Clin Rheumatol 20:1139, 2006.

Harrington JT: Mycobacterial and fungal arthritis. Curr Opin Rheumatol 10:335, 1998.

Huotari K et al: High cure rate for acute streptococcal prosthetic joint infections treated with debridement, antimicrobials, and implant retention in a specialized tertiary care center. Clin Infect Dis 67:1288, 2018.

Meehan AM et al: Outcome of penicillin-susceptible streptococcal prosthetic joint infection treated with debridement and retention of the prosthesis. Clin Infect Dis 36:845, 2003.

Mohammad M et al: The role of *Staphylococcus aureus* lipoproteins in hematogenous septic arthritis. Sci Rep 10:7936, 2020.

Osmon DR et al: Diagnosis and management of prosthetic joint infection: Clinical practice guidelines by the Infectious Diseases Society of America. Clin Infect Dis 56:e1, 2013.

Patel R: Periprosthetic joint infection. N Engl J Med 388:251, 2023.

Ross J et al: Septic arthritis and the opioid epidemic: 1465 cases of culture-positive native joint septic arthritis from 1990–2018. Open Forum Infect Dis 7:ofaa089, 2020.

Zeller V et al: One-stage exchange arthroplasty for chronic periprosthetic hip infection: Results of a large prospective cohort study. J Bone Joint Surg Am 96:e1, 2014.

Zimmerli W et al: Prosthetic-joint infections. N Engl J Med 351:145, 2004.

136 Osteomyelitis

Werner Zimmerli

Osteomyelitis, an infection of bone, can be caused by various microorganisms that arrive at bone through different routes. Spontaneous hematogenous osteomyelitis may occur in otherwise healthy individuals, whereas local microbial spread mainly affects either individuals who have underlying disease (e.g., vascular insufficiency) or patients who have compromised skin or other tissue barriers, with consequent exposure of bone. The latter situation typically follows surgery involving bone, such as sternotomy or orthopedic repair.

The manifestations of osteomyelitis are different in children and adults. In children, circulating microorganisms seed mainly long bones, whereas in adults, the vertebral column is the most commonly affected site.

Management of osteomyelitis differs greatly depending on whether an implant is involved. The most important aim of the management of either type of osteomyelitis is to prevent progression to chronic osteomyelitis by rapid diagnosis and prompt treatment. Therefore, unambiguous case definitions are required. Device-related bone and joint infection necessitates a multidisciplinary approach requiring antibiotic therapy and, in many cases, surgical removal of the device. For most types of osteomyelitis, the optimal duration and route of antibiotic treatment have not been established in clinical trials. Therefore, the recommendations for therapy in this chapter reflect mainly expert opinions.

CLASSIFICATION

There is no generally accepted, comprehensive system for classification of osteomyelitis, primarily because of the multifaceted presentation of this infection. Different specialists are confronted with different facets of bone disease. Most often, however, general practitioners or internists are the first to encounter patients with the initial signs and symptoms of osteomyelitis. These primary care physicians should be able to recognize this disease in any of its forms. Osteomyelitis cases can be classified by various criteria, including pathogenesis, duration of infection, location of infection, and presence or absence of foreign material. The widely used Cierny-Mader staging system is useful mainly for trauma surgeons. It classifies osteomyelitis according to anatomic site, comorbidity, and radiographic findings, with stratification of long-bone osteomyelitis to optimize surgical management; this system encompasses both systemic and local factors affecting immune status, metabolism, and local vascularity.

Any of three mechanisms can underlie osteomyelitis: (1) hematogenous spread; (2) spread from a contiguous site following surgery; and (3) secondary infection in the setting of vascular insufficiency or concomitant neuropathy. Hematogenous osteomyelitis in adults typically involves the vertebral column. In only about half of patients a primary focus can be detected. The most common primary foci of infection are the urinary tract, skin/soft tissue, intravascular catheterization sites, and the endocardium. Spread from a contiguous source follows either bone trauma or surgical intervention. Wound infection leading to osteomyelitis typically occurs after cardiovascular intervention involving the sternum, orthopedic repair after open fracture, or prosthetic joint insertion. Osteomyelitis secondary to vascular insufficiency or peripheral neuropathy most often follows chronic, progressively deep skin and soft tissue infection of the foot. The most common underlying condition is diabetes. In diabetes that is poorly controlled, the *diabetic foot syndrome* is caused by skin, soft tissue, and bone ischemia combined with motor, sensory, and autonomic neuropathy.

Classification of osteomyelitis according to the duration of infection, although ill defined, is useful because the management of acute and chronic osteomyelitis differs. However, not a defined duration of infection, but the presence or absence of bone necrosis (sequesters) is crucial. Acute osteomyelitis without bone necrosis can generally be treated with antibiotics alone. In contrast, for chronic osteomyelitis antibiotic treatment should be combined with debridement surgery. Acute hematogenous or contiguous osteomyelitis evolves over a short period—i.e., a few days or weeks. In contrast, subacute or chronic osteomyelitis lasts for weeks or months before treatment is started. Typical examples of a subacute course are vertebral osteomyelitis due to tuberculosis or brucellosis and delayed implant-associated infections caused mainly by low-virulence microorganisms (coagulase-negative staphylococci, *Cutibacterium acnes*). A special form of osteomyelitis is Brodie's abscess, mainly occurring in young males. Chronic osteomyelitis develops when insufficient therapy leads to persistence or recurrence, most often after sternal, mandibular, or foot infection.

Classification by location distinguishes among cases in the long bones, the vertebral column, and the periarticular bones. Long bones are generally involved after hematogenous seeding in children or contiguous spread following trauma or surgery. The risk of vertebral osteomyelitis in adults increases with age. Periarticular osteomyelitis, which complicates septic arthritis that has not been adequately treated, is especially common in periprosthetic joint infection.

Osteomyelitis involving a foreign device requires surgical management for cure. Even acute implant-associated infection calls for prolonged antimicrobial therapy. Therefore, identification of this type of disease is of practical importance.

VERTEBRAL OSTEOMYELITIS

■ PATHOGENESIS

Vertebral osteomyelitis, also referred to as *disk-space infection*, *septic diskitis*, *spondylodiskitis*, or *spinal osteomyelitis*, is the most common manifestation of hematogenous bone infection in adults. This designation reflects a pathogenic process leading to involvement of the adjacent vertebrae and the corresponding intervertebral disk. In adults, the disk is avascular. Microorganisms invade via the segmental arterial circulation in adjacent endplates and then spread into the disk. Alternative routes of infection are retrograde seeding through the prevertebral venous plexus and direct inoculation during spinal surgery, epidural infiltration, or trauma. In the setting of implant surgery, microorganisms are inoculated either during the procedure or, if wound healing is impaired, in the early postoperative period.

EPIDEMIOLOGY

Vertebral osteomyelitis occurs more often in male than in female patients (ratio, 1.5:1). Between 1995 and 2008, the incidence rate increased from 2.2 to 5.8 cases/100,000 person-years. There is a clear age-dependent increase. Men age ≥70 years have a sixfold higher incidence rate than those <70 years. The observed increase in reported cases over time may reflect improvements in diagnosis resulting from the broad availability of MRI technology. In addition, the fraction of cases of vertebral osteomyelitis acquired in association with health care is increasing as a consequence of comorbidity and the rising number of invasive interventions.

MICROBIOLOGY

Vertebral osteomyelitis is typically classified as pyogenic or nonpyogenic. However, this distinction is arbitrary: in "nonpyogenic" cases (tuberculous, brucellar), macroscopic pus formation (caseous necrosis, abscess) is quite common. A more accurate scheme is to classify cases as acute or subacute/chronic. Whereas the microbiologic spectrum of acute cases is similar in different parts of the world, the spectrum of subacute/chronic cases varies according to the geographic region. The great majority of cases are monomicrobial in etiology. Of episodes of acute vertebral osteomyelitis, 40–50% are caused by *Staphylococcus aureus*, 12% by streptococci, and 20% by gram-negative bacilli—mainly *Escherichia coli* (9%) and *Pseudomonas aeruginosa* (6%). Subacute vertebral osteomyelitis is typically caused by *Mycobacterium tuberculosis* or *Brucella* species in regions where these microorganisms are endemic. Osteomyelitis due to viridans streptococci also has a subacute presentation; these infections most often occur as secondary foci in patients with endocarditis. In vertebral osteomyelitis due to *Candida* species, the diagnosis is often delayed by several weeks; this etiology should be suspected in IV drug users who do not use sterile paraphernalia. In implant-associated spinal osteomyelitis, coagulase-negative staphylococci and *C. acnes*—which, in the absence of an implant, are generally considered contaminants—typically cause low-grade (chronic) infections. As an exception, coagulase-negative staphylococci can cause native spinal osteomyelitis in cases of prolonged bacteremia (e.g., in patients with infected pacemaker electrodes or implanted vascular catheters that are not promptly removed).

CLINICAL MANIFESTATIONS

The signs and symptoms of vertebral osteomyelitis are nonspecific. Only about half of patients develop fever >38°C (>100.4°F), perhaps because patients frequently use analgesic drugs. Back pain is the leading initial symptom (>85% of cases). The location of the pain corresponds to the site of infection: the cervical spine in ~10% of cases, the thoracic spine in 30%, and the lumbar spine in 60%. One exception is involvement at the thoracic level in two-thirds of cases of tuberculous osteomyelitis and at the lumbar level in only one-third. This difference is due to direct mycobacterial spread via pleural or mediastinal lymph nodes in pulmonary tuberculosis.

Neurologic deficits, such as radiculopathy, weakness, or sensory loss, are observed in about one-third of cases of vertebral osteomyelitis. Neurologic signs and symptoms are caused mostly by spinal epidural abscess. This complication starts with severe localized back pain and progresses to radicular pain, reflex changes, sensory abnormalities, motor weakness, bowel and bladder dysfunction, and paralysis.

A primary focus should always be sought but is found in only half of cases. Overall, endocarditis is identified in ~10% of patients. In osteomyelitis caused by viridans streptococci, endocarditis is the source in about half of patients.

Implant-associated spinal osteomyelitis can present as either early- or late-onset infection. Early-onset infection is diagnosed within 30 days after implant placement. *S. aureus* is the most common pathogen. Wound healing impairment and fever are the leading findings. Late-onset infection is diagnosed beyond 30 days after surgery, with low-virulence organisms such as coagulase-negative staphylococci or *C. acnes* as typical infecting agents. Fever is rare. One-quarter of patients have a sinus tract. Because of the delayed course and the lack of classic signs of infection, rapid diagnosis requires a high degree of suspicion.

DIAGNOSIS

Leukocytosis and neutrophilia have low levels of diagnostic sensitivity (only 65% and 40%, respectively). In contrast, an increased erythrocyte sedimentation rate or C-reactive protein (CRP) level has been reported in 98% and 100% of cases, respectively; thus, these tests are helpful in excluding vertebral osteomyelitis. The fraction of blood cultures that yield positive results depends heavily on whether the patient has been pretreated with antibiotics; across studies, the range is 30–78%. In view of this low rate of positive blood culture after antibiotic treatment, such therapy should be withheld until microbial growth is proven unless the patient has sepsis syndrome. In patients with negative blood cultures, CT-guided or open biopsy is needed. Whether a CT-guided biopsy with a negative result is repeated or followed by open biopsy depends on the experience of personnel at the specific center. Bone samples should be cultured for aerobic, anaerobic, and fungal agents, with a portion of the sample sent for histopathologic study. In cases with a subacute/chronic presentation, a suggestive history, or a granuloma detected during histopathologic analysis, mycobacteria and brucellae also should be sought. When blood and tissue cultures are negative despite suggestive histopathology, nonculture techniques (eubacterial or multiplex polymerase chain reaction analysis, metagenomics) of biopsy specimens or aspirated pus should be considered. These techniques allow detection of unusual pathogens such as *Helicobacter* spp. or *Tropheryma whipplei*.

Given that signs and symptoms of osteomyelitis are nonspecific, the clinical differential diagnosis of febrile back pain is broad, including pyelonephritis, pancreatitis, and viral syndromes. In addition, multiple noninfectious pathologies of the vertebral column, such as osteoporotic fracture, seronegative spondylitis (ankylosing spondylitis, psoriasis, reactive arthritis, enteropathic arthritis), and spinal stenosis must be considered.

Imaging procedures are the most important tools not only for the diagnosis of vertebral osteomyelitis but also for the detection of pyogenic complications and alternative conditions (e.g., bone metastases or osteoporotic fractures). Plain radiography is a reasonable first step in evaluating patients without neurologic symptoms and may reveal an alternative diagnosis. Because of its low sensitivity, plain radiography generally is not helpful in acute osteomyelitis, but it can be useful in subacute or chronic cases. The gold standard is MRI, which should be performed expeditiously in patients with neurologic impairment in order to rule out a herniated disk or to detect pyogenic complications in a timely manner **(Fig. 136-1, *left*)**. Even if the pathologic findings on MRI suggest vertebral osteomyelitis, alternative diagnoses should be considered, especially when blood cultures are negative. The most common alternative diagnosis is erosive osteochondrosis. Septic bone necrosis, gouty spondylodiskitis, and erosive diskovertebral lesions (Andersson lesions) in ankylosing spondylitis may likewise mimic vertebral osteomyelitis. CT is less sensitive than MRI but may be helpful in guiding a percutaneous biopsy. Positron emission tomography (PET) with ^{18}F-fluorodeoxyglucose, which has a high degree of diagnostic accuracy, is an alternative imaging procedure when MRI is contraindicated **(Fig. 136-1, *right*)**. ^{18}F-fluorodeoxyglucose PET should be considered for patients with implants and patients in whom several foci are suspected.

TREATMENT

Vertebral Osteomyelitis

The aims of therapy for vertebral osteomyelitis are (1) elimination of the pathogen(s), (2) protection from further bone loss, (3) relief of back pain, (4) prevention of complications, and (5) stabilization, if needed.

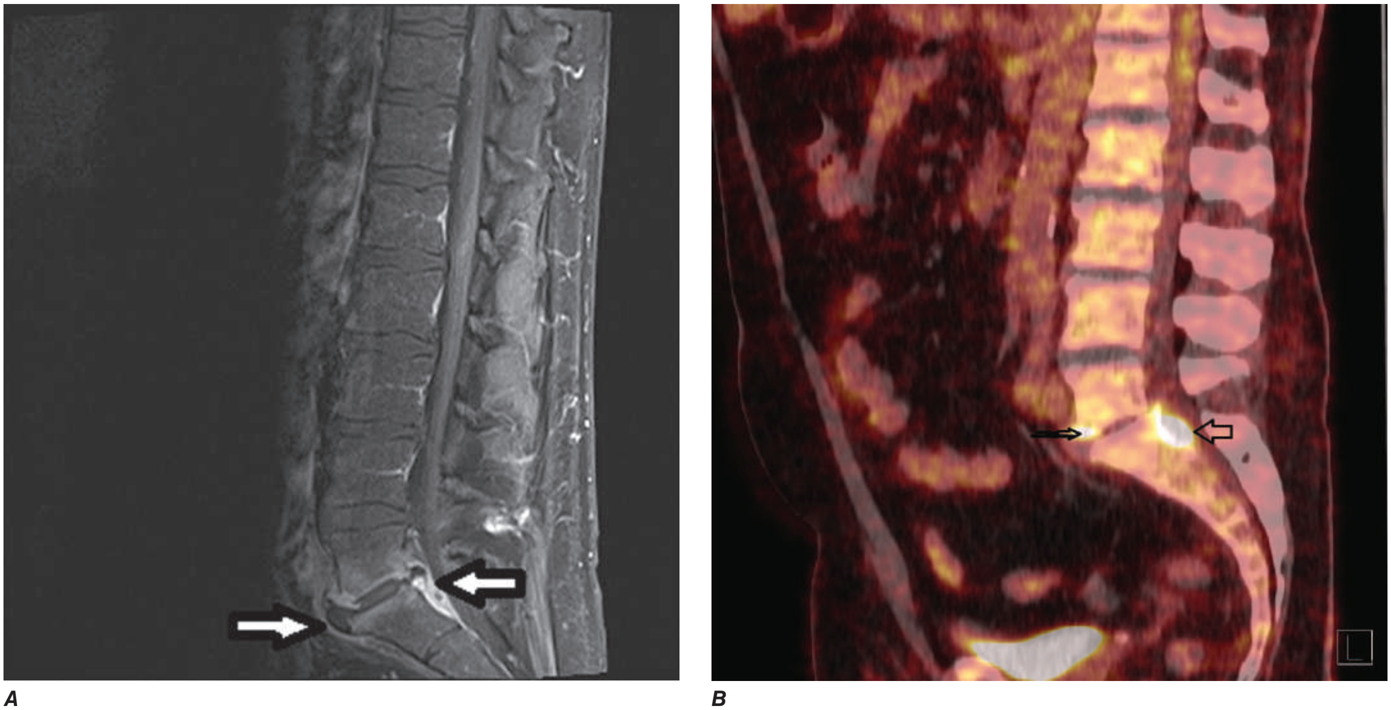

FIGURE 136-1 **Left:** MRI from a 53-year-old man suffering from prosthetic aortic valve endocarditis (*Aggregatibacter actinomycetemcomitans*). In addition, he experienced lumbar pain for 7 weeks. MRI sagittal sequence shows on T1 fat-saturated post-gadolinium image enhancement in the intervertebral disk space (*ventral arrow*) and a small epidural abscess (*dorsal arrow*). **Right:** PET/CT from the same patient 4 weeks earlier. PET/CT fusion shows fluorodeoxyglucose uptake at L5 ventral (*small arrow*) and dorsal of S1 (*large arrow*: epidural abscess). *(Figures courtesy of Damien Toia, MD, Kantonsspital Baselland; with permission.)*

Table 136-1 summarizes suggested antimicrobial regimens for infections attributable to the most common etiologic agents. For optimal antimicrobial therapy, identification of the infecting agent is required. Therefore, in patients without sepsis syndrome, antibiotics should not be administered until the pathogen is identified in a blood culture, a bone biopsy, or an aspirated pus collection. Traditionally, bone infections are at least initially treated by the IV route. However, the preference for the IV route is not evidence based. There are no good arguments for the assumption that IV therapy is superior to oral administration if the following requirements are met: (1) optimal antibiotic spectrum, (2) excellent bioavailability of the oral drug, (3) clinical studies confirming efficacy of the oral drug, (4) normal intestinal function, and (5) no vomiting. Indeed, in a controlled trial in patients with bone and joint infections, including vertebral osteomyelitis, oral antibiotic therapy was noninferior to intravenous therapy when used during the first 6 weeks. Nevertheless, a short initial course of parenteral therapy with a β-lactam antibiotic may lower the risk of emergence of fluoroquinolone resistance, especially if *P. aeruginosa* infection is treated with ciprofloxacin or staphylococcal infection with the combination of a fluoroquinolone plus rifampin. These suggestions are based on observational studies and expert opinion. A randomized, controlled trial showed that 6 weeks of antibiotic treatment is not inferior to a 12-week course in patients with pyogenic vertebral osteomyelitis. The cure rate was 90.9% in both groups 1 year after therapy. Thus, prolonged antibiotic therapy is required only for patients with undrained abscesses and for patients with spinal implants. Treatment efficacy should be regularly monitored through inquiries about signs and symptoms (fever, pain) and assessment for signs of inflammation (elevated CRP concentrations). Follow-up MRI is appropriate only for patients with pyogenic complications since the correlation between clinical healing and improvement on MRI is very poor.

Surgical treatment generally is not needed in acute hematogenous vertebral osteomyelitis. However, it is always necessary in implant-associated spinal infection. Early infections (those occurring up to 30 days after internal stabilization) can be cured with debridement, implant retention, and a 3-month course of antibiotics **(Table 136-2)**. In contrast, in late infection with a duration of >30 days, implant removal and a 6-week course of antibiotics (Table 136-1) are required for complete elimination of the infection. If implants cannot be removed, oral suppressive long-term treatment should follow the initial course of IV antibiotics. The optimal duration of suppressive therapy is unknown. However, if antibiotic therapy is discontinued after, for example, 1 year, close clinical and laboratory (CRP) follow-up is needed.

COMPLICATIONS

Complications should be suspected when there is persistent pain, a persistently increased CRP level, and new-onset or persistent neurologic impairment. In cases of persistent pain with or without signs of inflammation, paravertebral, epidural, or psoas abscesses must be sought. Epidural abscesses occur in 15–20% of cases. This complication is more common in the cervical column (30%) than in the lumbar spine (12%). Risk factors for severe neurologic deficit were epidural abscess, cervical and/or thoracic involvement, and *S. aureus* vertebral osteomyelitis. Persistent pain despite normalization of CRP values indicates mechanical complications such as severe osteonecrosis or spinal instability. These patients require a consult with an experienced orthopedic surgeon.

GLOBAL CONSIDERATIONS

The incidence rate of acute vertebral osteomyelitis is similar in different regions of the world. In contrast, subacute/chronic vertebral osteomyelitis predominates in defined regions. Cases attributable to brucellosis predominate in endemic areas such as the Middle East, Africa, Central and South America, and the Indian subcontinent. Tuberculosis is an especially frequent cause in Africa and Asia (India, Indonesia, China), where more than two-thirds of the global tuberculosis burden is reported. Thus, specific diagnostic tests are needed in patients either living in or having traveled to these regions.

OSTEOMYELITIS IN LONG BONES

PATHOGENESIS

Osteomyelitis in long bones is a consequence of hematogenous seeding, exogenous contamination during trauma (open fracture), or

TABLE 136-1 Antibiotic Therapy for Osteomyelitis in Adults without Implants[a]

MICROORGANISM	ANTIMICROBIAL AGENT (DOSE,[b] ROUTE)
Staphylococcus spp.	
Methicillin-susceptible	Nafcillin or oxacillin[c] (2 g IV q6h) ***followed by*** Rifampin (300–450 mg PO q12h) plus levofloxacin (750 mg PO q24h or 500 mg PO q12h)
Methicillin-resistant	Vancomycin[d] (15 mg/kg IV q12h) or daptomycin (8–10 mg/kg IV q24h) ***followed by*** Rifampin (300–450 mg PO q12h) ***plus*** Levofloxacin (750 mg PO q24h or 500 mg PO q12h) or TMP-SMX[e] (1 double-strength tablet PO q8h) or fusidic acid (500 mg PO q8h)
Streptococcus spp.	Penicillin G[c] (5 million units IV q6h) or ceftriaxone (2 g IV q24h)
Enterobacteriaceae	
Quinolone-susceptible	Ciprofloxacin (750 mg PO q24h)
Quinolone-resistant[f]	Imipenem (500 mg IV q6h) or meropenem (1–2g IV q8h)
Pseudomonas aeruginosa	Cefepime or ceftazidime (2 g IV q8h) *plus* an aminoglycoside[g] ***or*** Piperacillin-tazobactam (4.5 g IV q8h) plus an aminoglycoside[g] for 2–4 weeks ***followed by*** Ciprofloxacin[h] (750 mg PO q12h)
Anaerobes	Clindamycin (600 mg IV q6–8h) for 2–4 weeks ***followed by*** Clindamycin[i] (300 mg PO q6h)

[a]Unless otherwise indicated, the total duration of antimicrobial treatment is generally 6 weeks. [b]All dosages are for adults with normal renal function. [c]When the patient has delayed-type penicillin hypersensitivity, cefuroxime (1.5 g IV q6–8h) can be administered. When the patient has immediate-type penicillin hypersensitivity, the penicillin should be replaced by vancomycin (1 g IV q12h). [d]Target vancomycin trough level: 15–20 µg/mL. [e]Trimethoprim-sulfamethoxazole. A double-strength tablet contains 160 mg of trimethoprim and 800 mg of sulfamethoxazole. [f]Including isolates producing extended-spectrum β-lactamase. [g]The need for addition of an aminoglycoside has not yet been proven. However, this addition may decrease the risk of emergence of resistance to the β-lactam. [h]The rationale for starting ciprofloxacin treatment only after pretreatment with a β-lactam is the increased risk of emergence of quinolone resistance in the presence of a heavy bacterial load. [i]Alternatively, penicillin G (5 million units IV q6h) or ceftriaxone (2 g IV q24h) can be used against gram-positive anaerobes (e.g., *Cutibacterium acnes*), and metronidazole (500 mg IV/PO q8h) can be used against gram-negative anaerobes (e.g., *Bacteroides* spp.).

Source: From W Zimmerli: Vertebral osteomyelitis. N Engl J Med 362:1022, 2010. Copyright © 2010 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

perioperative contamination during surgery involving bone. Hematogenous infection in long bones typically occurs in children. In adults, the leading pathogenic source is exogenous fracture-related infection, mainly associated with internal fixation devices. For classification, the presence of a sequestrum and the status of the surrounding soft tissue are crucial for the decision as to whether a surgical intervention is required. Chronic osteomyelitis can be reactivated after a symptom-free interval of >70 years. Such recurrences are most common among elderly patients who developed *S. aureus* osteomyelitis in the preantibiotic era.

EPIDEMIOLOGY

In adults, most cases of long-bone osteomyelitis are posttraumatic or postsurgical; less frequently, late recurrence arises from hematogenous infections during childhood. For postoperative or postsurgical osteomyelitis, the term "facture-related infection" was generated. The risk of infection depends on the type of fracture. After closed fracture, implant-associated infection occurs in fewer than 1% of patients. In contrast, after open fracture, the risk of osteomyelitis ranges from ~2% up to 30%, with the precise figure depending on the degree of tissue damage during trauma and the time between injury and admission to a specialized center.

MICROBIOLOGY

The spectrum of microorganisms causing hematogenous long-bone osteomyelitis does not differ from that in vertebral osteomyelitis. *S. aureus* is most commonly isolated in each type of osteomyelitis. In rare cases, mycobacteria or fungal agents such as *Cryptococcus* species, *Sporothrix schenckii*, *Blastomyces dermatitidis*, or *Coccidioides* species are found in patients who live or have traveled in endemic regions. Impaired cellular immunity (e.g., in HIV infection or after transplantation) predisposes to these etiologies. Coagulase-negative staphylococci are the second most common etiologic agents (after *S. aureus*) in implant-associated osteomyelitis. After open fracture, contiguous long-bone osteomyelitis is typically caused by gram-negative bacilli or a polymicrobial mixture of organisms.

CLINICAL MANIFESTATIONS

The leading symptoms in adults with primary or recurrent hematogenous long-bone osteomyelitis are pain and low-grade fever. Infection occasionally manifests as clinical sepsis and local signs of inflammation (erythema and swelling). After internal fixation, osteomyelitis can be classified as early (acute; <3 weeks), delayed (3–10 weeks), or late (chronic) infection. Early/acute long-bone osteomyelitis manifests as signs of surgical site infection, such as erythema and impaired wound healing. Acute implant-associated infection may also follow hematogenous seeding at any time after implantation of a device. Typical symptoms are new-onset pain and signs of sepsis. Delayed or late (chronic) infections are usually caused by low-virulence microorganisms or occur after ineffective treatment of early-onset infection. Patients may present with persisting pain, subtle local signs of inflammation, intermittent discharge of pus, or fluctuating erythema over the scar **(Fig. 136-2)**. A special form of subacute osteomyelitis is Brodie's abscess, which is characterized by pain (98%) and swelling (53%), mainly in the tibia or the femur. Fever and inflammatory markers are typical. The median delay from symptoms to diagnosis is 3 months. Thus, a young patient with unclear localized pain in the tibia or femur should be worked-up with an imaging modality (plain x-ray, MRI, or CT).

DIAGNOSIS

The diagnostic workup for acute hematogenous long-bone osteomyelitis is similar to that for vertebral osteomyelitis. Bone remodeling and thus marker uptake are increased for at least 1 year after surgery. Therefore, the three-phase bone scan is not useful during this interval. However, in late recurrences it allows rapid diagnosis at low cost. If the results are positive, CT is required to estimate the extent of inflamed tissue and detect bone necrosis (sequesters). Implant-associated infection should be suspected if CRP values do not return to the normal range or rise after an initial decrease. Clinical and laboratory suspicion should prompt surgical exploration and sampling.

In osteomyelitis of >1 year's duration, single-photon emission CT plus conventional CT (SPECT/CT) is a good option, either with ^{99m}Tc methylene diphosphonate (^{99m}Tc-MDP)–labeled leukocytes or with labeled monoclonal antibodies to granulocytes. Surgical debridement is needed for diagnostic (biopsy culture, histology) and therapeutic reasons.

TREATMENT

Osteomyelitis in Long Bones

Treatment for acute hematogenous infection in long bones is identical to that for acute vertebral osteomyelitis (Table 136-1). The suggested duration of antibiotic therapy is 4–6 weeks. In patients with good soft tissue condition and no sequestra or implants, generally no surgical intervention is required. According to a controlled trial, oral treatment can be given, provided that a regimen with excellent oral biocompatibility is available. An initial IV course can be as short as a few days, if the microorganism and

TABLE 136-2 Antibiotic Therapy for Osteomyelitis Associated with Orthopedic Devices

MICROORGANISM	ANTIMICROBIAL AGENT[a] (DOSE, ROUTE)
Staphylococcus spp.	*Recommendation for initial treatment phase (2 weeks with implant)*
Methicillin-susceptible	Rifampin (450 mg PO/IV q12h[b]) ***plus*** Nafcillin or oxacillin[c] (2 g IV q6h)
Methicillin-resistant	Rifampin (450 mg PO/IV q12h[b]) ***plus*** Vancomycin (15 mg/kg IV q12h) or daptomycin (8–10 mg/kg IV q24h)
Staphylococcus spp.	*Recommendation after completion of initial treatment phase* Rifampin (450 mg PO q12h[b]) ***plus*** Levofloxacin (750 mg PO q24h or 500 mg PO q12h) or ciprofloxacin (750 mg PO q12h) or fusidic acid (500 mg PO q8h) or TMP-SMX[d] (1 double-strength tablet PO q8h) or minocycline (100 mg PO q12h) or linezolid (600 mg PO q12h) or clindamycin (1200–1350 mg/d PO in 3 or 4 divided doses)
Streptococcus spp.[e]	Penicillin G[c] (18–24 million units/d IV in 6 divided doses) or ceftriaxone (2 g IV q24h) for 4 weeks ***followed by*** Amoxicillin (750–1000 mg PO q6–8h) or clindamycin (1200–1350 mg/d PO in 3 or 4 divided doses)
Enterococcus spp.[f]	
Penicillin-susceptible	Penicillin G[c] (24 million units/d IV in 6 divided doses) *or* ampicillin or amoxicillin[g] (2 g IV q4–6h)
Penicillin-resistant	Vancomycin (15 mg/kg IV q12h) or daptomycin (6–10 mg/kg IV q24h) or linezolid (600 mg IV/PO q12h)
Enterobacteriaceae	A β-lactam selected in light of in vitro susceptibility profile for 2 weeks[h] ***followed by*** Ciprofloxacin (750 mg PO q12h)
Enterobacter spp.[i] and nonfermenters[j] (e.g., *Pseudomonas aeruginosa*)	Cefepime or ceftazidime (2 g IV q8h) *or* meropenem (1–2 g IV q8h[k]) for 2–4 weeks ***followed by*** Ciprofloxacin (750 mg PO q12h)
Cutibacterium spp.	Penicillin G[c] (18–24 million units/d IV in 6 divided doses) or clindamycin (600–900 mg IV q8h) for 2–4 weeks ***followed by*** Amoxicillin (750–1000 mg PO q6–8h) or clindamycin (1200–1350 mg/d PO in 3 or 4 divided doses)
Gram-negative anaerobes (e.g., *Bacteroides* spp.)	Metronidazole (500 mg IV/PO q8h)
Mixed bacteria (without methicillin-resistant staphylococci)	Ampicillin-sulbactam (3 g IV q6h) or amoxicillin-clavulanate[l] (2.2 g IV q6h) or piperacillin-tazobactam (4.5 g IV q8h) or imipenem (500 mg IV q6h) or meropenem (1–2 g IV q8h[k]) for 2–4 weeks ***followed by*** Individualized oral regimens chosen in light of antimicrobial susceptibility

[a] Antimicrobial agents should be chosen in light of the isolate's in vitro susceptibility, the patient's drug allergies and intolerances, potential drug interactions, and contraindications to specific drugs. All dosages recommended are for adults with normal renal and hepatic function. See text for total durations of antibiotic treatment. [b] Other dosages and intervals of administration with equivalent success rates have been reported. [c] When the patient has delayed-type penicillin hypersensitivity, cefazolin (2 g IV q8h) can be administered. When the patient has immediate-type penicillin hypersensitivity, the penicillin should be replaced by vancomycin (1 g IV q12h). [d] Trimethoprim-sulfamethoxazole. A double-strength tablet contains 160 mg of trimethoprim and 800 mg of sulfamethoxazole. [e] Determination of the minimal inhibitory concentration (MIC) of penicillin is advisable. [f] Combination therapy with an aminoglycoside is optional since its superiority to monotherapy for prosthetic joint infection is unproved. When using combination therapy, monitor for signs of aminoglycoside ototoxicity and nephrotoxicity; the latter is potentiated by other nephrotoxic agents (e.g., vancomycin). [g] For patients with hypersensitivity to penicillin, see treatment options for penicillin-resistant enterococci. [h] Ciprofloxacin (PO or IV) can be administered to patients with hypersensitivity to β-lactams. [i] Ceftriaxone and ceftazidime should not be administered for treatment targeting *Enterobacter* species, even strains that test susceptible in the laboratory, but can be used against nonfermenters. Strains producing extended-spectrum β-lactamases should not be treated with any cephalosporin, including cefepime. *Enterobacter* infections can also be treated with ertapenem (1 g IV q24h); however, ertapenem is not effective against *Pseudomonas* spp. and other nonfermenters. [j] Addition of an aminoglycoside is optional. Use of two active drugs can be considered in light of the patient's clinical condition. [k] The recommended dosage is in line with the guidelines of the Infectious Diseases Society of America. In Europe, 2 g IV q8h is suggested for *P. aeruginosa* infections. [l] Not available as an IV formulation in the United States.

Source: Modified from W Zimmerli et al: N Engl J Med 351:1645, 2004. Massachusetts Medical Society.

its antibiotic susceptibility is known. In recurrences of chronic osteomyelitis as well as in each type of exogenous osteomyelitis (acute, chronic, with or without an implant), a combination of surgical debridement, obliteration of dead space, and long-term antibiotic therapy is required. The length of therapy depends on the completeness of the surgical intervention (removal of sequestra, implants, and necrotic tissue).

The therapeutic aims in patients whose infections are associated with internal fixation devices are consolidation of the fracture and prevention of chronic osteomyelitis. Stable implants can be maintained except in patients with uncontrolled sepsis. In a systematic review reporting the outcome of 276 patients, the success rate with retention of the implant was 86–100% in early, 82–89% in delayed, and only 67% in late fracture-related infection. Appropriate antimicrobial therapies are listed in Table 136-2. The cure rate for early staphylococcal implant-associated infections treated with a fluoroquinolone plus rifampin is >90%. Rifampin is efficacious against staphylococcal biofilms of ≤3 weeks' duration. Similarly, fluoroquinolones are active against biofilms formed by gram-negative bacilli. In these cases, a short initial course of IV therapy with a β-lactam antibiotic is suggested to minimize the risk of emergence of resistance to the oral drugs. The total duration of treatment is 3 months, and the device can be retained even after antibiotics have been discontinued. In contrast, in cases caused by rifampin-resistant staphylococci or fluoroquinolone-resistant gram-negative bacilli, all hardware should be removed after consolidation of the fracture and before discontinuation of antibiotics. These patients are treated with an oral antibiotic (suppressive therapy) as long as the hardware is retained.

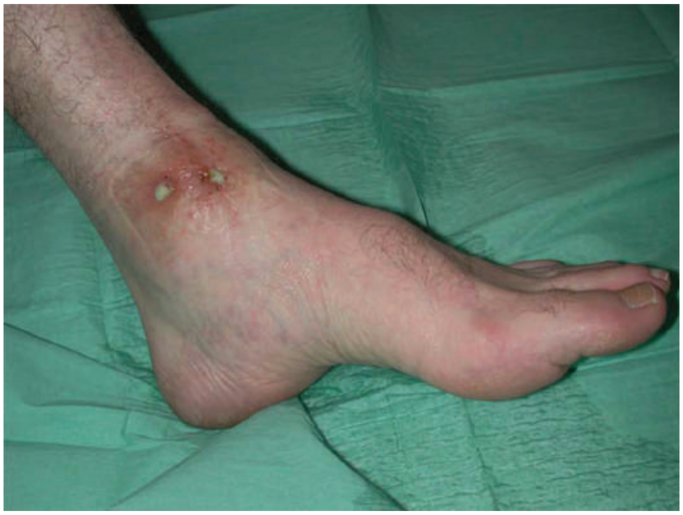

FIGURE 136-2 A 42-year-old man who had sustained a malleolar fracture 6 weeks previously had persistent pain and slight inflammation after orthopedic repair. His infection was treated with oral antibiotics without debridement surgery. This insufficient management of an implant-associated *Staphylococcus aureus* infection was complicated by a sinus tract.

COMPLICATIONS

The main complication of long-bone osteomyelitis is the persistence of infection with progression to chronic osteomyelitis. This risk is especially high after internal fixation of an open fracture and among patients with implant-associated osteomyelitis that is treated without surgical debridement. In longstanding osteomyelitis, recurrent sinus tracts result in severe damage to skin and soft tissue (Fig. 136-2). Patients who have chronic open wounds need a therapeutic approach combining orthopedic repair and plastic reconstructive surgery.

GLOBAL CONSIDERATIONS

In North American and Western European countries, tuberculous osteomyelitis is extremely rare, occurring mainly in very old people, HIV-infected patients, and immigrants from endemic countries. In contrast, in countries where the prevalence of tuberculosis is high (India, Indonesia, China), tuberculous osteomyelitis must routinely be considered.

PERIPROSTHETIC JOINT INFECTION

PATHOGENESIS

Implanted foreign material is highly susceptible to local infection due to local immunodeficiency around the device. Infection occurs by either the exogenous or the hematogenous route. More rarely, contiguous spread from adjacent sites of osteomyelitis or deep soft-tissue infection may cause periprosthetic joint infection (PJI). The fact that foreign devices are covered with host proteins such as fibronectin favors the adherence of staphylococci and the formation of a biofilm that resists phagocytosis.

EPIDEMIOLOGY

The risk of infection manifesting during the first 2 postoperative years varies according to the joint. It is lowest after hip and knee arthroplasty (0.3–1.5%) and highest after ankle and elbow replacement (4–10%). The risk of hematogenous PJI is highest in the early postoperative period. However, hematogenous seeding occurs throughout life, and most cases therefore develop >2 years after implantation. The rate of risk for secondary PJI during *S. aureus* bacteremia is 30–40%.

MICROBIOLOGY

About 50–70% of cases of PJI are caused by staphylococci (*S. aureus* and coagulase-negative staphylococci), 6–10% by streptococci, 4–10% by gram-negative bacilli, and the rest by various other microorganisms. In patients with hematogenous PJI, the fraction of streptococci and gram-negative bacilli is higher and the fraction of coagulase-negative staphylococci is much lower. All microorganisms can cause PJI, including fungi and mycobacteria. *C. acnes* causes up to one-third of episodes of periprosthetic shoulder infection.

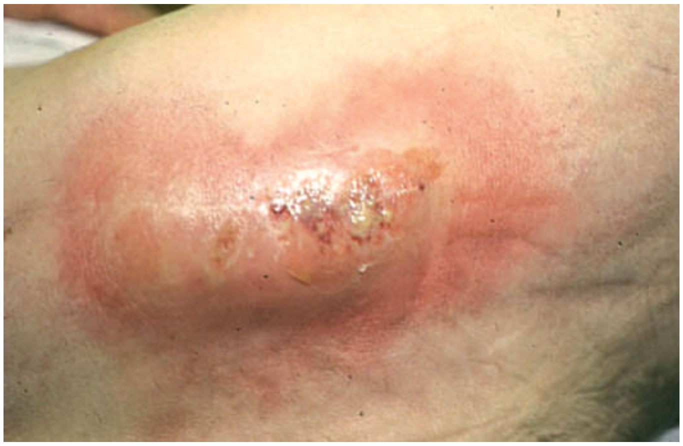

FIGURE 136-3 Acute postoperative periprosthetic joint infection of the left hip caused by group B streptococci in a 68-year-old woman.

CLASSIFICATION AND CLINICAL MANIFESTATIONS

PJI is traditionally classified as early (<3 months after implantation), delayed (3–24 months after surgery), or late (>2 years after implantation). For therapeutic decision-making (see below), it is more useful to classify PJI as (1) acute hematogenous PJI with <3 weeks of symptoms, (2) early postinterventional PJI manifesting within 1 month after surgery, or (3) chronic PJI with symptom duration of >3 weeks.

Acute exogenous PJI typically presents with local signs of infection **(Fig. 136-3)**. In contrast, acute hematogenous PJI is most often caused by *S. aureus* and is characterized by new-onset pain. Local inflammatory signs are rare in hip PJI but frequent in knee PJI. Fever is rare after the initial phase of bacteremia. Key findings in chronic PJI are joint effusion, local pain, implant loosening, and occasionally a sinus tract. Chronic PJI is most commonly caused by low-virulence microorganisms such as coagulase-negative staphylococci or *C. acnes*. These infections are characterized by nonspecific symptoms, such as chronic pain caused by low-grade inflammation or early loosening.

DIAGNOSIS

Blood tests such as the measurement of CRP (elevated levels, ≥10 mg/L) and erythrocyte sedimentation rate (elevated rates, ≥30 mm/h) are sensitive (91–97%) but not specific (70–78%). Synovial fluid cell counts are ~90% sensitive and specific, with threshold values of 1700 leukocytes/μL in periprosthetic knee infection and 4200 leukocytes/μL in periprosthetic hip infection. A biomarker, α-defensin, can be tested in synovial fluid; this biomarker is highly specific and therefore useful in confirming PJI. However, this test is expensive and its sensitivity is limited; therefore, it should not be used for screening. During debridement surgery, at least three but optimally six tissue samples should be obtained for culture and histopathology. If implant material (modular parts, screws, or the prosthesis) is removed, sonication of this material followed by culture and/or use of molecular methods to examine the sonicate fluid allows the detection of microorganisms in biofilms.

The three-phase bone scan is very sensitive for detecting PJI but is not specific. As mentioned above, this test does not differentiate bone remodeling from infection and therefore is not useful during at least the first year after implantation. CT and MRI detect soft tissue infection, prosthetic loosening, and bone erosion, but imaging artifacts caused by metal implants limit their use. ^{18}F-fluorodeoxyglucose PET (^{18F}-FD-PET) is an alternative method with good sensitivity but low specificity for the detection of PJI. Therefore, ^{18}F-FDG-PET/CT is useful only in excluding but not confirming PJI.

TREATMENT

Periprosthetic Joint Infection

The outcome following treatment of PJI is better when managed using a multidisciplinary approach involving an experienced orthopedic surgeon, an infectious disease specialist, a plastic reconstructive surgeon, and a microbiologist. Therefore, most patients are referred to a specialized center. In general, the goal of treatment is cure—i.e., a pain-free functional joint with complete eradication of the infecting pathogen(s). However, for patients with severe comorbidity, lifelong suppressive antimicrobial therapy may be preferred. As a rule, antimicrobial therapy without surgical intervention is not curative but merely suppressive. There are four curative surgical options: debridement and implant retention, one-stage implant exchange, two-stage implant exchange, and implant removal without replacement. The least invasive treatment option should be selected for each patient without compromising the cure rate. The choice can be guided following a treatment algorithm. However, the different surgical options have not been tested in a randomized controlled trial. Implant retention offers a good chance of infection-free survival (>80%) only if the following conditions are fulfilled: (1) acute infection, (2) stable implant, (3) pathogen susceptible to a biofilm-active antimicrobial agent (see below), and (4) skin and soft tissue in good condition.

Table 136-2 summarizes pathogen-specific antimicrobial therapy for PJI. Initial IV therapy is followed by long-term oral antibiotics. Efficacious treatment is best defined in staphylococcal implant-associated infections. Rifampin exhibits excellent activity against biofilms composed of susceptible staphylococci. Because of the risk of rapid emergence of resistance, rifampin must always be combined with another effective antibiotic. If gram-negative infections are treated with implant retention, fluoroquinolones should be used because of their activity against gram-negative biofilms.

PREVENTION OF HEMATOGENOUS INFECTION

As mentioned above, hematogenous seeding may occur throughout life. This risk is highest during *S. aureus* bacteremia from a distant focus. Therefore, documented bacterial infections should be promptly treated in patients with prosthetic joints. However, according to a prospective case-control study, the risk of prosthetic hip or knee infection is not increased following dental procedures. Therefore, antibiotic prophylaxis is not needed during dental work.

GLOBAL CONSIDERATIONS

Rifampin and fluoroquinolones are still the only antimicrobial agents with good activity against staphylococcal and gram-negative biofilms, respectively. Thus, in countries with high rates of rifampin resistance in staphylococci and/or high rates of fluoroquinolone resistance in gram-negative bacilli, debridement with implant retention generally does not yield a good cure rate.

STERNAL OSTEOMYELITIS

PATHOGENESIS

Sternal osteomyelitis occurs primarily after sternal surgery (with the entry of exogenous organisms) and more rarely by hematogenous seeding or contiguous extension from adjacent sites of sternocostal arthritis. Exogenous sternal osteomyelitis after open sternal surgery is also called *deep sternal-wound infection.* Exogenous infection may follow minor sternal trauma, sternal fracture, and manubriosternal septic arthritis. Tuberculous sternal osteomyelitis typically manifests during hematogenous seeding in children or as reactivated infection in adults. Reactivation is sometimes preceded by blunt trauma. In rare cases, tuberculous sternal osteomyelitis is caused by continuous infection from an infected internal mammary lymph node.

EPIDEMIOLOGY

The incidence of poststernotomy wound infection varies from 0.5% to 2%, but figures are even higher among patients with risk factors such as diabetes, obesity, chronic renal failure, emergency surgery, use of bilateral internal mammary artery grafts, and re-exploration for bleeding. Rapid diagnosis and correct management of superficial sternal wound infection prevent its progression to sternal osteomyelitis. Primary (hematogenous) sternal osteomyelitis accounts for only 0.3% of all cases of osteomyelitis. Risk factors are IV drug use, HIV infection, radiotherapy, blunt trauma, cardiopulmonary resuscitation, alcohol abuse, liver cirrhosis, and hemoglobinopathy.

MICROBIOLOGY

Poststernotomy osteomyelitis is generally caused by *S. aureus* (10–20% of cases), coagulase-negative staphylococci (40–60%), gram-negative bacilli (5-15%), or *C. acnes* (2–10%). The spectrum of microorganisms greatly varies in different populations (as represented in different publications); this is perhaps due to local epidemiologic conditions, such as antimicrobial stewardship and the incidence of chronic infections. Polymicrobial cases are rare in acute infection and indicate exogenous superinfection during therapy. Hematogenous sternal osteomyelitis is caused most commonly by *S. aureus*. Other microorganisms play a role in special populations—e.g., *P. aeruginosa* in IV drug users, *Salmonella* species in individuals with sickle cell anemia, and *M. tuberculosis* in patients from endemic areas who have previously had tuberculosis.

CLINICAL MANIFESTATIONS

Exogenous sternal osteomyelitis manifests as fever, increased local pain, erythema, wound discharge, and sternal instability **(Fig. 136-4)**. Contiguous mediastinitis is a feared complication, occurring in ~10–30% of patients with sternal osteomyelitis. Hematogenous sternal osteomyelitis is characterized by sternal pain, swelling, and erythema. In addition, most patients have systemic signs and symptoms of sepsis.

The differential diagnosis of hematogenous sternal osteomyelitis includes immunologic processes typically presenting as systemic or multifocal inflammation of the sternum or of the sternoclavicular or sternocostal joints (e.g., SAPHO [synovitis, acne, pustulosis, hyperostosis, osteitis], vasculitis, chronic multifocal relapsing osteomyelitis).

DIAGNOSIS

In primary sternal osteomyelitis, the diagnostic workup does not differ from that in other types of hematogenous osteomyelitis (see above). When a patient has grown up in regions where tuberculosis is endemic, a specific workup for mycobacterial infection should be performed, especially if osteomyelitis had its onset after a blunt sternal trauma. In secondary sternal osteomyelitis, leukocyte counts may be normal, but the CRP level is >100 mg/L in most cases. Tissue sampling for microbiologic studies is crucial. In osteomyelitis associated with sternal wires, low-virulence microorganisms, such as coagulase-negative staphylococci, play an important role. In order to differentiate between colonization and infection, samples from at least three deep biopsies should be subjected to microbiologic examination. Superficial swab cultures are not diagnostic and may be misleading. No studies have

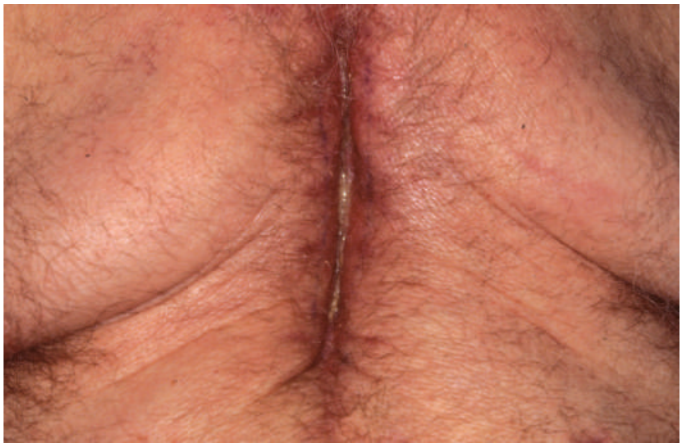

FIGURE 136-4 Sternal osteomyelitis caused by *Staphylococcus epidermidis* 5 weeks after sternotomy for aortocoronary bypass in a 72-year-old man.

compared the value of the various imaging modalities in suspected primary sternal osteomyelitis. However, MRI is the current gold standard for detection of each type of osteomyelitis.

TREATMENT

Sternal Osteomyelitis

In cases of deep sternal-wound infection, a combined approach using both surgery and antibiotic treatment is required. Antimicrobial therapy should be started immediately after samples have been obtained for microbiologic analyses in order to control clinical sepsis. To protect a newly inserted heart valve, initial treatment should be directed against staphylococci, with consideration of the local susceptibility pattern. In centers with a high prevalence of methicillin-resistant *S. aureus*, vancomycin or daptomycin should be added to a broad-spectrum β-lactam drug. As soon as cultures of blood and/or deep wound biopsies have confirmed the pathogen's identity and susceptibility pattern, treatment should be optimized and narrowed accordingly. Tables 136-1 and 136-2 show appropriate therapeutic choices for the most frequently identified microorganisms causing sternal osteomyelitis in the absence and presence, respectively, of an implanted device. In a recent observational study of patients with staphylococcal deep sternal-wound infection, the use of a rifampin-containing regimen was predictive of success. The optimal duration of antibiotic therapy has not been established. In acute sternal osteomyelitis without hardware, a 6-week course is the rule. In patients with remaining sternal wires, treatment duration is generally prolonged to 3 months (Table 136-2). Like other types of tuberculous bone infection, tuberculous sternal osteomyelitis is treated for 6–12 months.

Primary sternal osteomyelitis can generally be treated without surgery. In contrast, in secondary sternal osteomyelitis, debridement is always required. This procedure should be performed by a team of experienced surgeons, since mediastinitis, bone infection, and skin and soft tissue damage may need to be treated during the same intervention.

PROGNOSIS

Primary sternal osteomyelitis poses a minimal mortality risk. In contrast, the mortality rates from secondary sternal osteomyelitis during the first year after diagnosis are 6–20%.

GLOBAL CONSIDERATIONS

In endemic areas, microorganisms such as *M. tuberculosis*, *Salmonella* species, and *Brucella* species should be considered during sampling for microbiologic diagnosis.

FOOT OSTEOMYELITIS

PATHOGENESIS

Osteomyelitis of the foot is a feared complication in patients with diabetic foot ulcers. It also usually occurs in patients with peripheral arterial insufficiency, or peripheral neuropathy and after foot surgery. These entities are often linked to each other, especially in diabetic patients with late complications. However, foot osteomyelitis is also seen in patients with isolated peripheral neuropathy and can manifest as implant-associated osteomyelitis in patients without comorbidity due to a deep wound infection after foot surgery (hallux valgus surgery, arthrodesis, total ankle arthroplasty). Foot osteomyelitis is acquired almost exclusively by the exogenous route. It is a complication of deep pressure ulcers and of impaired wound healing after surgery.

EPIDEMIOLOGY

About 34% of people with diabetes develop a foot ulcer during their lifetime. Among patients with foot ulcer, the annual incidence of progression to foot osteomyelitis is 5%. The condition starts with skin and soft tissue lesions and progresses to osteomyelitis, especially in patients with risk factors. Diabetic foot osteomyelitis increases the risk of amputation. With adequate management of the early stage of diabetic foot infections, the rate of amputation can be lowered.

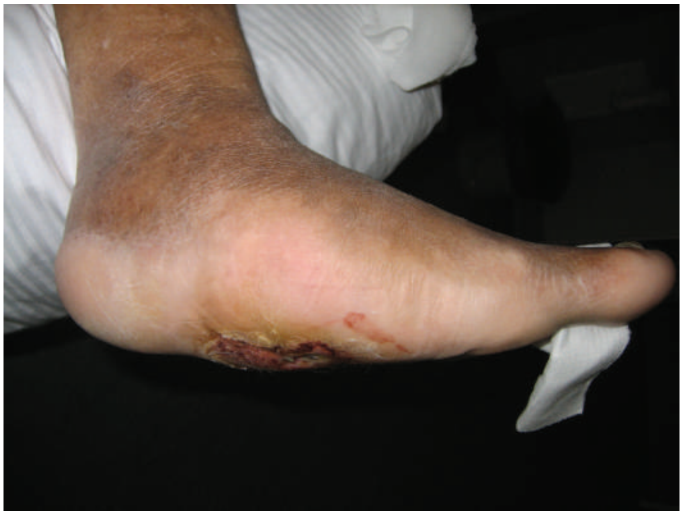

FIGURE 136-5 Neuropathic joint disease (Charcot foot) complicated by chronic foot osteomyelitis in a 78-year-old woman with diabetes mellitus complicated by severe neuropathy.

RISK FACTORS

Risk factors for diabetic foot infection are (1) peripheral motor, sensory, and autonomic neuropathy; (2) neuro-osteoarthropathic deformities (Charcot foot; **Fig. 136-5**); (3) arterial insufficiency; (4) uncontrolled hyperglycemia; (5) disabilities such as reduced vision; and (6) maladaptive behavior.

MICROBIOLOGY

The correlation between cultures from bone biopsy and those from wound swabs or even deep soft-tissue punctures is poor. In a study of 31 patients with simultaneous sampling, the correlation between needle biopsy and bone biopsy cultures was only 24%. The correlation is better when *S. aureus* is isolated (40–50%) than when anaerobes (20–35%), gram-negative bacilli (20–30%), or coagulase-negative staphylococci (0–20%) are identified. When only bone-biopsy samples are considered, the leading pathogens are *S. aureus* (25–40%), anaerobes (5–20%), and various gram-negative bacilli (18–40%). The precise distribution depends on whether the patient has already been treated with antibiotics. Anaerobes are especially prevalent in chronic wounds. Pretreatment typically selects for *P. aeruginosa*, methicillin-resistant *S. aureus*, or enterococci.

DIAGNOSIS

In many cases, foot osteomyelitis can be diagnosed clinically, without imaging procedures. Most clinicians rely on the "probe-to-bone" test, which has a positive predictive value of ~90% in populations with a high pretest probability. Thus, in a patient with diabetes who is hospitalized for a chronic deep foot ulcer, the diagnosis of foot osteomyelitis is highly probable if bone can be directly touched with a metal instrument. In a patient with a lower pretest probability, MRI should be performed because of its high sensitivity (80–100%) and specificity (80–90%). Plain x-rays have typical features characterizing diabetic foot osteomyelitis. However, their sensitivity is only 30–90% and the specificity varies between 50% and 90%; thus, it is mainly useful for follow-up of patients with confirmed diabetic foot osteomyelitis.

TREATMENT

Foot Osteomyelitis

As mentioned above, correlation between cultures of bone and those of wound swabs or wound punctures is poor. Antibiotic treatment should be based on bone culture. If no bone biopsy is performed, empirical therapy chosen in light of the most common infecting agents and the type of clinical syndrome should be given. In a controlled therapeutic trial of diabetic patients in whom no infected bone needed to be resected, the outcome of a 6-week course of antibiotics was not different from a 12-week course. In patients undergoing surgical debridement, a 3-week course of systemic antibiotics was noninferior to a 6-week course in a prospective, randomized pilot trial. Wound debridement combined with antibiotics renders amputation unnecessary in about two-thirds of patients. For the diagnosis and treatment of diabetic foot infections, the following management strategies should be considered. If a foot ulcer is clinically infected, prompt empirical antimicrobial therapy may prevent progression to osteomyelitis. When the risk of methicillin-resistant *S. aureus* is considered high, an agent active against these strains (e.g., vancomycin) should be chosen. If the patient has not recently received antibiotics, the spectrum of the selected antibiotic must include gram-positive cocci (e.g., clindamycin, ampicillin-sulbactam). If the patient has received antibiotics within the past month, the spectrum of empirical antibiotics should include gram-negative bacilli (e.g., clindamycin plus a fluoroquinolone). If the patient has risk factors for *Pseudomonas* infection (previous colonization, residence in a warm climate, frequent exposure of the foot to water), an empirical antipseudomonal agent (e.g., piperacillin-tazobactam, cefepime) is indicated. If osteomyelitis is suspected either on clinical grounds (probe to bone) or on the basis of imaging procedures (MRI), bone biopsy should be performed. If infected bone is not entirely removed by surgery, the patient should be treated for 4–6 weeks in line with the identified pathogen(s) and their susceptibility. Treatment should initially be given by the IV route. Whether therapy can later be administered by the oral route depends on the bioavailability of oral drugs that cover the infecting agents. In the future, phage therapy may be a valuable additive therapeutic option.

GLOBAL CONSIDERATIONS

The number of multiresistant microorganisms causing diabetic foot infection is increasing. The prevalence of methicillin-resistant *S. aureus* is 5–43% in various countries. Gram-negative pathogens are more common than gram-positive bacteria in Asia. In a study of 102 patients with diabetic foot infection from India, 69% of aerobic gram-negative bacilli produced extended-spectrum β-lactamase and 43% of *S. aureus* isolates were methicillin resistant. Risk factors for multidrug-resistant microorganisms are poor glycemic control, prolonged duration of infection, and large ulcer size.

FURTHER READING

Depypere M et al: Pathogenesis and management of fracture-related infection. Clin Microbiol Infect 26:572, 2020.

Li H-K et al: Oral versus intravenous antibiotics for bone and joint infection. N Engl J Med 380:425, 2019.

Osmon Dr et al: Diagnosis and management of prosthetic joint infection: Clinical practice guidelines by the Infectious Diseases Society of America. Clin Infect Dis 56:e1, 2013.

Senneville E et al: IWGDF/IDSA guidelines on the diagnosis and treatment of diabetes-related foot infections (IWGDF/IDSA 2023). 40:e3687, 2024.

Zimmerli W (ed): *Bone and Joint Infections. From Microbiology to Diagnostics and Treatment*, 2nd ed. West Sussex, UK, John Wiley & Sons, Ltd, 2021.

Zimmerli W: Clinical practice. Vertebral osteomyelitis. N Engl J Med 362:1022, 2010.

137

Intraabdominal Infections and Abscesses

Miriam Baron Barshak

Intraperitoneal infections generally arise because a normal anatomic barrier is disrupted. This disruption may result from a variety of causes—e.g., when the appendix, a diverticulum, or an ulcer ruptures; when the bowel wall is weakened by ischemia, tumor, or inflammation (e.g., in inflammatory bowel disease); or with adjacent inflammatory processes, such as pancreatitis or pelvic inflammatory disease, in which enzymes (in the former case) or organisms (in the latter) may leak into the peritoneal cavity. Whatever the inciting event, once inflammation develops and organisms usually contained within the bowel or another organ enter the normally sterile peritoneal space, a knowable series of events takes place. Intraabdominal infections occur in two stages: peritonitis and—if the patient survives this stage and goes untreated—abscess formation. The types of microorganisms predominating in each stage of infection are responsible for the pathogenesis of disease.

PERITONITIS

Peritonitis is a life-threatening event that is often accompanied by bacteremia and sepsis syndrome **(Chap. 315)**. The peritoneal cavity is large but is divided into compartments. The upper and lower peritoneal cavities are divided by the transverse mesocolon; the greater omentum extends from the transverse mesocolon and from the lower pole of the stomach to line the lower peritoneal cavity. The pancreas, duodenum, and ascending and descending colon are located in the anterior retroperitoneal space; the kidneys, ureters, and adrenals are found in the posterior retroperitoneal space. The other organs, including the liver, stomach, gallbladder, spleen, jejunum, ileum, transverse and sigmoid colon, cecum, and appendix, are within the peritoneal cavity. The cavity is lined with a serous membrane that can serve as a conduit for fluids—a property exploited in peritoneal dialysis **(Fig. 137-1)**. A small amount of serous fluid is normally present in the

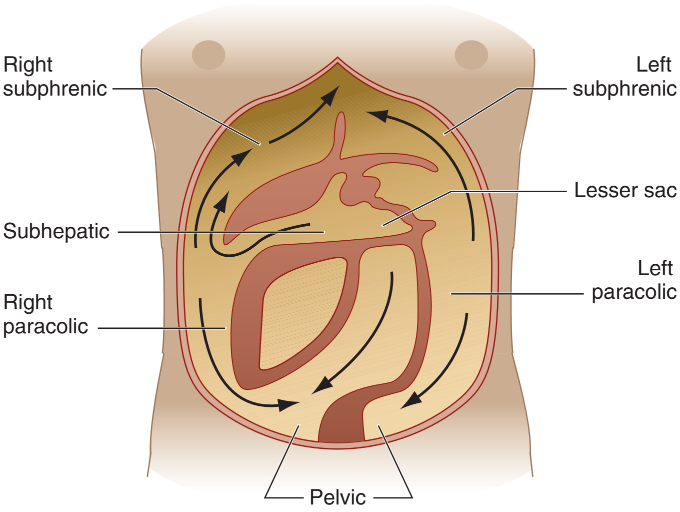

FIGURE 137-1 Diagram of the intraperitoneal spaces, showing the circulation of fluid and potential areas for abscess formation. Some compartments collect fluid or pus more often than others. These compartments include the pelvis (the lowest portion), the subphrenic spaces on the right and left sides, and Morrison's pouch, which is a posterosuperior extension of the subhepatic spaces and is the lowest part of the paravertebral groove when a patient is recumbent. The falciform ligament separating the right and left subphrenic spaces appears to act as a barrier to the spread of infection; consequently, it is unusual to find bilateral subphrenic collections. *(Reproduced with permission from B Lorber [ed]: Atlas of Infectious Diseases, vol VII: Intra-abdominal infections, hepatitis, and gastroenteritis. Springer; 1996.)*

peritoneal space, with a protein content (consisting mainly of albumin) of <30 g/L and <300 white blood cells (WBCs, generally mononuclear cells) per microliter. In bacterial infections, leukocyte recruitment into the infected peritoneal cavity consists of an early influx of polymorphonuclear leukocytes (PMNs) and a prolonged subsequent phase of mononuclear cell migration. The phenotype of the infiltrating leukocytes during the course of inflammation is regulated primarily by resident-cell chemokine synthesis.

■ PRIMARY (SPONTANEOUS) BACTERIAL PERITONITIS

Peritonitis is either primary (without an apparent source of contamination) or secondary. The types of organisms found and the clinical presentations of these two processes are different. In adults, primary bacterial peritonitis (PBP) occurs most commonly in conjunction with cirrhosis of the liver (frequently the result of alcoholism). However, the disease has been reported in adults with metastatic malignant disease, postnecrotic cirrhosis, chronic active hepatitis, acute viral hepatitis, congestive heart failure, systemic lupus erythematosus, and lymphedema as well as in patients with no underlying disease. Although PBP virtually always develops in patients with preexisting ascites, it is, in general, an uncommon event, occurring in ≤10% of cirrhotic patients. The cause of PBP has not been established definitively but is believed to involve hematogenous spread of organisms in a patient in whom a diseased liver and altered portal circulation result in a defect in the usual filtration function. Organisms multiply in ascites, a good medium for growth. Proteins of the complement cascade are found in peritoneal fluid, with lower levels in cirrhotic patients than in patients with ascites of other etiologies. The opsonic and phagocytic properties of PMNs are diminished in patients with advanced liver disease. Cirrhosis is associated with alterations in the gut microbiota, including an increased prevalence of potentially pathogenic bacteria such as Enterobacteriaceae. Small-intestinal bacterial overgrowth is frequently present in advanced stages of liver cirrhosis and has been linked with pathologic bacterial translocation and PBP. Factors promoting these changes in cirrhosis may include deficiencies in Paneth cell defensins, reduced intestinal motility, decreased pancreatobiliary secretions, and portal-hypertensive enteropathy.

The presentation of PBP differs from that of secondary peritonitis. The most common manifestation is fever, which is reported in up to 80% of patients. Ascites is found but virtually always predates infection. Abdominal pain, an acute onset of symptoms, and peritoneal irritation during physical examination can be helpful diagnostically, but the absence of any of these findings does not exclude this often-subtle diagnosis. Nonlocalizing symptoms (such as malaise, fatigue, or encephalopathy) or jaundice or acute kidney injury without another clear etiology also should prompt consideration of PBP in a susceptible patient. It is vital to sample the peritoneal fluid of any cirrhotic patient with ascites and fever. The finding of >250 PMNs/μL is diagnostic for PBP. This criterion does not apply to secondary peritonitis (see below). The microbiology of PBP also is distinctive. While enteric gram-negative bacilli such as *Escherichia coli* are commonly encountered, gram-positive organisms such as streptococci, enterococci, or even pneumococci are sometimes found. In an important development, widespread use of quinolones to prevent PBP in high-risk subgroups of patients, frequent hospitalizations, and exposure to broad-spectrum antibiotics have led to a change in the etiology of infections in patients with cirrhosis, with more gram-positive bacteria and extended-spectrum β-lactamase (ESBL)–producing Enterobacteriaceae in recent years. Risk factors for multidrug-resistant infections include nosocomial origin of infection, long-term norfloxacin prophylaxis, recent infection with multiresistant bacteria, and recent use of β-lactam antibiotics. In PBP, a single organism is typically isolated; anaerobes are found less frequently in PBP than in secondary peritonitis, in which a mixed flora including anaerobes is the rule. In fact, if PBP is suspected and multiple organisms including anaerobes are recovered from the peritoneal fluid, the diagnosis must be reconsidered and a source of secondary peritonitis sought.

The diagnosis of PBP is not easy. It depends on the exclusion of a primary intraabdominal source of infection. Contrast-enhanced CT is useful in identifying an intraabdominal source for infection. It may be difficult to recover organisms from cultures of peritoneal fluid, presumably because the burden of organisms is low. However, the yield can be improved if 10 mL of peritoneal fluid is placed directly into a blood culture bottle. Because bacteremia frequently accompanies PBP, blood should be cultured simultaneously. To maximize the yield, culture samples should be collected prior to administration of antibiotics. There is interest in identifying biomarkers in ascites that may be associated with PBP. No specific radiographic studies are helpful in the diagnosis of PBP. A plain film of the abdomen would be expected to show ascites. Chest and abdominal radiography should be performed when patients have abdominal pain to exclude free air, which signals a perforation **(Fig. 137-2)**.

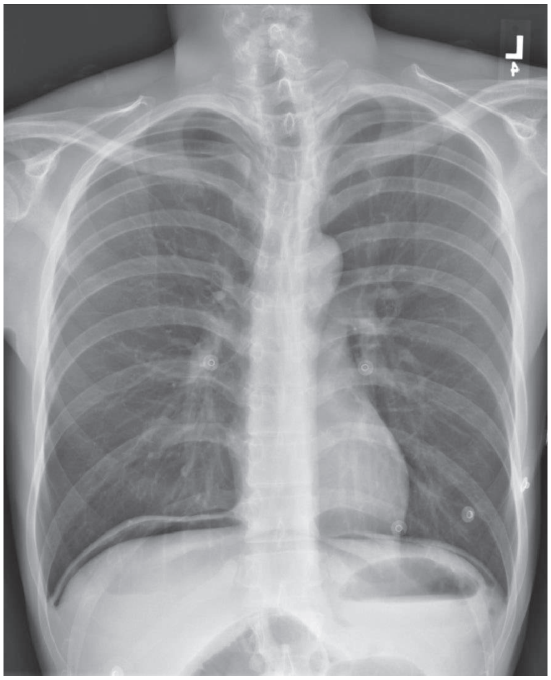

FIGURE 137-2 Pneumoperitoneum. Free air under the diaphragm on an upright chest film suggests the presence of a bowel perforation and associated peritonitis. *(Image courtesy of Dr. John Braver; with permission.)*

TREATMENT

Primary Bacterial Peritonitis

Treatment for PBP is directed at the isolate from blood or peritoneal fluid. Gram's staining of peritoneal fluid often gives negative results in PBP. Delays in antibiotic treatment of PBP are associated with increased mortality. Therefore, until culture results become available, prompt therapy should cover gram-negative aerobic bacilli and gram-positive cocci. Third-generation cephalosporins such as cefotaxime (2 g q8h, administered IV) provide reasonable initial coverage in moderately ill patients if the local prevalence of multidrug resistant organisms is low. Broad-spectrum antibiotics, such as β-lactam/β-lactamase inhibitor combinations (e.g., piperacillin/tazobactam, 3.375 g q6h IV for adults with normal renal function) or ceftriaxone (2 g q24h IV), also are options, with vancomycin for patients who have had prior infection or colonization with methicillin-resistant *Staphylococcus aureus*. Broader empirical coverage aimed at resistant hospital-acquired gram-negative bacteria (e.g., carbapenem or newer agents, such as ceftolozane-tazobactam or ceftazidime-avibactam) and vancomycin-resistant *Enterococcus* (e.g., daptomycin) may be appropriate for nosocomially acquired PBP or in patients with prior antibiotic-resistant infections or critical illness until culture results become available. Empirical coverage for anaerobes is not necessary. A mortality benefit from albumin (1.5 g/kg of body weight within 6 h of detection and 1.0 g/kg on day 3) has been demonstrated for patients with PBP, with greatest benefit among those who present with serum creatinine levels ≥1 mg/dL, blood urea nitrogen levels ≥30 mg/dL, or total bilirubin levels ≥5 mg/dL. After the infecting organism is identified, therapy should be narrowed to target the specific pathogen. Patients with PBP usually respond within 72 h to appropriate antibiotic therapy. Antimicrobial treatment can be administered for as little as 5 days if rapid improvement occurs and blood cultures are negative, but a course of up to 2 weeks may be required for patients with bacteremia and for those whose improvement is slow. Persistence of WBCs in the ascitic fluid after therapy should prompt a search for additional diagnoses.

Prognosis: PBP is associated with significant morbidity and mortality, perhaps reflecting the fact that PBP risk is highest among patients with advanced liver disease. In a 2018 study of hospitalized patients with PBP in the United States, in-hospital mortality was 17.6%. Another study in 2019 found that among patients at a U.S. tertiary academic center, mortality following PBP was 23% at 30 days and 37% at 90 days. The morbidity and mortality associated with PBP has led to interest in strategies for PBP prevention.

Prevention • **PRIMARY PREVENTION** Several observational studies and a meta-analysis raise the concern that gastric acid suppression may increase the risk of PBP. No prospective studies have yet addressed whether avoidance of such therapy may prevent PBP. A 2021 guideline from the American Association for the Study of Liver Diseases recommends chronic antibiotic prophylaxis be considered, after discussing risks and benefits, with a regimen described in the next section for patients who are at highest risk for PBP—that is, those with an ascitic-fluid total protein level <1.5 g/dL along with impaired renal function (creatinine, ≥1.2 mg/dL; blood urea nitrogen, ≥25 mg/dL; or serum sodium, ≤130 mEq/L) and/or liver failure (Child-Turcotte-Pugh score, ≥9; and bilirubin, ≥3 mg/dL). A course of antibiotic prophylaxis is recommended for patients with cirrhosis and gastrointestinal bleeding until bleeding has resolved and vasoactive drugs are stopped.

SECONDARY PREVENTION PBP has a high rate of recurrence. Up to 70% of patients experience a recurrence within 1 year. Antibiotic prophylaxis has been recommended for patients with a history of PBP to reduce this rate and improve short-term survival rates. Prophylactic regimens for adults with normal renal function include fluoroquinolones (ciprofloxacin, 500 mg weekly; or norfloxacin [not available in the United States], 400 mg/d) or trimethoprim-sulfamethoxazole (one double-strength tablet daily). However, these recommendations are being reconsidered in recent years as the risks and benefits evolve. Long-term administration of broad-spectrum antibiotics in this setting has been shown to increase the risk of severe staphylococcal infections and of infections with multidrug-resistant organisms (MDRO). Additionally, quinolone prophylaxis is less effective in patients who are colonized with MDRO. There is increased interest in using rifaximin, a broad-spectrum antibiotic that is used already for hepatic encephalopathy and is not absorbed, for PBP prophylaxis (1200 mg daily).

SECONDARY PERITONITIS

Secondary peritonitis develops when bacteria contaminate the peritoneum as a result of spillage from an intraabdominal viscus. The organisms found almost always constitute a mixed flora in which facultative gram-negative bacilli and anaerobes predominate, especially when the contaminating source is colonic. Early in the course of infection, when the host response is directed toward containment, exudate containing fibrin and PMNs is found. Early death in this setting is attributable to gram-negative bacillary sepsis and to potent endotoxins circulating in the bloodstream **(Chap. 315)**. Gram-negative bacilli, particularly *E. coli*, are common bloodstream isolates, but *Bacteroides fragilis* bacteremia also occurs. The severity of abdominal pain and the clinical course depend on the inciting process. The organisms isolated from the peritoneum also vary with the source of the initial process and the normal flora at that site. Secondary peritonitis can result primarily from chemical irritation and/or bacterial contamination. For example, as long as the patient is not achlorhydric, a ruptured gastric ulcer will release low-pH gastric contents that will serve as a chemical irritant. The normal flora of the stomach comprises the same organisms found in the oropharynx but in lower numbers. Thus, the bacterial burden in a ruptured ulcer is negligible compared with that in a ruptured appendix. The normal flora of the colon below the ligament of Treitz contains ~10^{11} anaerobic organisms/g of feces but only 10^8 aerobes/g; therefore, anaerobic species account for 99.9% of the bacteria **(Chap. 484)**. Leakage of colonic contents (pH 7–8) does not cause significant chemical peritonitis, but infection is intense because of the heavy bacterial load.

Depending on the inciting event, local symptoms may occur in secondary peritonitis—for example, epigastric pain from a ruptured gastric ulcer. In appendicitis **(Chap. 342)**, the initial presenting symptoms are often vague, with periumbilical discomfort and nausea followed in a number of hours by pain more localized to the right lower quadrant. Unusual locations of the appendix (including a retrocecal position) can complicate this presentation further. Once infection has spread to the peritoneal cavity, pain increases, particularly with infection involving the parietal peritoneum, which is innervated extensively. Patients usually lie motionless, often with knees drawn up to avoid stretching the nerve fibers of the peritoneal cavity. Coughing and sneezing, which increase pressure within the peritoneal cavity, are associated with sharp pain. There may or may not be pain localized to the infected or diseased organ from which secondary peritonitis has arisen. Patients with secondary peritonitis generally have abnormal findings on abdominal examination, with marked voluntary and involuntary guarding of the anterior abdominal musculature. Later findings include tenderness, especially rebound tenderness. In addition, there may be localized findings in the area of the inciting event. In general, patients are febrile, with marked leukocytosis and a left shift of the WBCs to band forms.

While recovery of organisms from peritoneal fluid is easier in secondary than in primary peritonitis, a tap of the abdomen is rarely the procedure of choice in secondary peritonitis. An exception is in cases involving trauma, where the possibility of a hemoperitoneum may need to be excluded early. Emergent studies (such as abdominal CT) to find the source of peritoneal contamination should be undertaken if the patient is hemodynamically stable; unstable patients may require surgical intervention without prior imaging. Results of cultures from chronic drain sites are not reliable for defining the etiology of infections.

TREATMENT

Secondary Peritonitis

Treatment for secondary peritonitis includes early administration of antibiotics aimed particularly at aerobic gram-negative bacilli and anaerobes (see below). The most appropriate regimen depends on the anticipated flora and the degree of illness. Community-acquired infections associated with mild to moderate disease can be treated with many drugs covering these organisms, including broad-spectrum β-lactam/β-lactamase inhibitor combinations (e.g., ticarcillin/clavulanate, 3.1 g q4–6h IV; or piperacillin/tazobactam, 3.375 g q6h IV) or a combination of either a fluoroquinolone (e.g., levofloxacin, 750 mg q24h IV) or a third-generation cephalosporin (e.g., ceftriaxone, 2 g q24h IV) plus metronidazole (500 mg q8h IV). Eravacycline is a newer antibiotic in the tetracycline class that has been approved by the U.S. Food and Drug Administration for treatment of complicated intraabdominal infections (1 mg/kg q12h IV). Patients in intensive care units and/or those with health care–associated infections should receive antibiotics targeting more resistant gram-negative organisms such as *Pseudomonas aeruginosa*—e.g., imipenem (500 mg q6h IV), meropenem (1 g q8h IV), higher-dose piperacillin/tazobactam (4.5 g IV q6h), or drug combinations such as cefepime (2 g IV q8h) or ceftazidime (2 g IV q8h) plus metronidazole. The role of enterococci and *Candida* species in mixed infections is controversial; however, because cephalosporin-based regimens lack activity against enterococci, ampicillin or vancomycin can be added to these regimens for enterococcal coverage in very ill patients until culture results are available. For patients known to be colonized with ampicillin-resistant, vancomycin-resistant enterococci (VRE), a VRE-active agent, such as linezolid or daptomycin, should be included. Antifungal coverage is warranted if there is growth of *Candida* species from a sterile site. Patients who are known to be colonized with highly resistant gram-negative organisms may require treatment with a newer agent such as ceftazidime/avibactam or ceftolozane/tazobactam. Secondary peritonitis usually requires both surgical intervention to address the inciting process and antibiotics to treat early bacteremia, to decrease the incidence of abscess formation and wound infection, and to prevent distant spread of infection. Although surgery is rarely indicated in PBP in adults, it may be life-saving in secondary peritonitis. In a recent study, the urgency and success of source control impacted the odds of survival, while the choice of empiric antibiotics did not. Recombinant human activated protein C (APC) was considered at one time for treatment of severe sepsis from causes including secondary peritonitis but was withdrawn from the market in 2011 after it was determined that the drug was associated with an increased risk of bleeding and that evidence for its beneficial effects was inadequate. Thus APC should not be used for sepsis or septic shock outside randomized clinical trials.

Peritonitis may develop as a complication of abdominal surgeries. These infections may be accompanied by localizing pain and/or nonlocalizing signs or symptoms such as fever, malaise, anorexia, and toxicity. As a nosocomial infection, postoperative peritonitis may be associated with organisms such as staphylococci, components of the gram-negative hospital microflora, and the microbes that cause PBP and secondary peritonitis, as described above.

■ PERITONITIS IN PATIENTS UNDERGOING CONTINUOUS AMBULATORY PERITONEAL DIALYSIS

A third type of peritonitis arises in patients who are undergoing continuous ambulatory peritoneal dialysis (CAPD). Unlike PBP and secondary peritonitis, which are caused by endogenous bacteria, CAPD-associated peritonitis usually involves skin organisms. The pathogenesis of infection is similar to that of intravascular device–related infection, in which skin organisms migrate along the catheter, which both serves as an entry point and exerts the effects of a foreign body. Exit-site or tunnel infection may or may not accompany CAPD-associated peritonitis. Like PBP, CAPD-associated peritonitis is usually caused by a single organism. Peritonitis is, in fact, the most common reason for discontinuation of CAPD. Improvements in equipment design, especially the Y-set connector, have resulted in a decrease from one case of peritonitis per 9 months of CAPD to one case per 24 months. Diabetes was reported to be a risk factor for CAPD-associated peritonitis in a study from Taiwan.

The clinical presentation of CAPD peritonitis resembles that of secondary peritonitis in that diffuse pain and peritoneal signs are common. The dialysate is usually cloudy and contains >100 WBCs/μL, >50% of which are neutrophils. However, the number of cells depends in part on dwell time. According to a guideline from the International Society for Peritoneal Dialysis (2016), for patients undergoing automated peritoneal dialysis who present during their nighttime treatment and whose dwell time is much shorter than with CAPD, the clinician should use the percentage of PMNs rather than the absolute number of WBCs to diagnose peritonitis. As the normal peritoneum has very few PMNs, a proportion above 50% is strong evidence of peritonitis even if the absolute WBC count does not reach 100/μL. Meanwhile, patients undergoing automated peritoneal dialysis without a daytime exchange who present with abdominal pain may have no fluid to withdraw, in which case 1 L of dialysate should be infused and permitted to dwell a minimum of 1–2 h, then drained, examined for turbidity, and sent for cell count with differential and culture. The differential (with a shortened dwell time) may be more useful than the absolute WBC count. In equivocal cases or in patients with systemic or abdominal symptoms in whom the effluent appears clear, a second exchange is performed, with a dwell time of at least 2 h. Clinical judgment should guide initiation of therapy.

The most common organisms are *Staphylococcus* species, which accounted for ~45% of cases in one series. Historically, coagulase-negative staphylococcal species were identified most commonly in these infections, but these isolates have more recently been decreasing in frequency. *Staphylococcus aureus* is more often involved among patients who are nasal carriers of the organism than among those who are not, and this organism is the most common pathogen in overt exit-site infections. Gram-negative bacilli and fungi such as *Candida* species also are found. Vancomycin-resistant enterococci and vancomycin-intermediate *S. aureus* have been reported to produce peritonitis in CAPD patients. The finding of more than one organism in dialysate culture should prompt evaluation for secondary peritonitis. As with PBP, culture of dialysate fluid in blood culture bottles improves the yield. To facilitate diagnosis, several hundred milliliters of removed dialysis fluid should be concentrated by centrifugation before culture.

TREATMENT

CAPD Peritonitis

Empirical therapy for CAPD peritonitis should be directed at *S. aureus*, coagulase-negative *Staphylococcus*, and gram-negative bacilli until the results of cultures become available. Guidelines suggest that agents should be chosen on the basis of local experience with resistant organisms. In some centers, a first-generation cephalosporin such as cefazolin (for gram-positive bacteria) and a fluoroquinolone or a third-generation cephalosporin such as ceftazidime (for gram-negative bacteria) may be reasonable; in areas with high rates of infection with methicillin-resistant *S. aureus*, vancomycin should be used instead of cefazolin, and gram-negative coverage may need to be broadened—e.g., with an aminoglycoside, ceftazidime, cefepime, or a carbapenem. Broad coverage including vancomycin should be particularly considered for patients with septic physiology or exit-site infections. Vancomycin should also be included in the regimen if the patient has a history of colonization or infection with methicillin-resistant *S. aureus* or has a history of severe allergy to penicillins and cephalosporins. Loading doses are administered intraperitoneally; doses depend on the dialysis method and the patient's renal function. Intraperitoneal antibiotics are given either continuously (i.e., with each exchange) or intermittently (i.e., once daily, with the dose allowed to remain in the peritoneal cavity for at least 6 h). If the patient is severely ill, IV antibiotics should be added at doses appropriate for the patient's degree of renal failure. The clinical response to an empirical treatment regimen should be rapid; if the patient has not responded after 48–96 h of treatment, new samples should be collected for cell counts and cultures, and catheter removal should be considered. For patients who lack exit-site or tunnel infection, the typical duration of antibiotic treatment is 14 days. For patients with exit-site or tunnel infection, catheter removal should be considered, and a longer duration of antibiotic therapy (up to 21 days) may be appropriate. In fungal infections, the catheter should be removed immediately.

■ TUBERCULOUS PERITONITIS

See Chap. 183.

INTRAABDOMINAL ABSCESSES

■ INTRAPERITONEAL ABSCESSES

Abscess formation is common in untreated peritonitis if overt gram-negative sepsis either does not develop or develops but is not fatal. In experimental models of abscess formation, mixed aerobic and anaerobic organisms have been implanted intraperitoneally. Without therapy directed at anaerobes, animals develop intraabdominal abscesses. As in humans, these experimental abscesses may stud the peritoneal cavity, lie within the omentum or mesentery, or even develop on the surface of or within viscera such as the liver.

Pathogenesis and Immunity There is often disagreement about whether an abscess represents a disease state or a host response. In a sense, it represents both: while an abscess is an infection in which viable infecting organisms and PMNs are contained in a fibrous capsule, it is also a process by which the host confines microbes to a limited space, thereby preventing further spread of infection. In any event, abscesses do cause significant symptoms, and patients with abscesses can be quite ill. Experimental work has helped to define both the host cells and the bacterial virulence factors responsible—most notably in the case of *B. fragilis*. This organism, although accounting for only 0.5% of the normal colonic flora, is the anaerobe most frequently isolated from intraabdominal infections, is especially prominent in abscesses, and is the most common anaerobic bloodstream isolate. On clinical grounds, therefore, *B. fragilis* appears to be uniquely virulent. Moreover, *B. fragilis* acts alone to cause abscesses in animal models of intraabdominal infection, whereas most other *Bacteroides* species must act synergistically with a facultative organism to induce abscess formation.

Of the several virulence factors identified in *B. fragilis*, one is critical: the capsular polysaccharide complex found on the bacterial surface. This complex comprises at least eight distinct surface polysaccharides. Structural analysis of these polysaccharides has shown an unusual motif of oppositely charged sugars. Polysaccharides having these *zwitterionic* characteristics, such as polysaccharide A, evoke a host response in the peritoneal cavity that localizes bacteria into abscesses. *B. fragilis* and polysaccharide A have been found to adhere to primary mesothelial cells in vitro; this adherence, in turn, stimulates the production of tumor necrosis factor α and intercellular adhesion molecule 1 by peritoneal macrophages. Although abscesses characteristically contain PMNs, the process of abscess induction depends on the stimulation of T lymphocytes by these unique zwitterionic polysaccharides. The stimulated CD4+ T lymphocytes secrete leukoattractant cytokines and chemokines. The alternative pathways of complement and fibrinogen also participate in abscess formation.

While antibodies to the capsular polysaccharide complex enhance bloodstream clearance of *B. fragilis*, CD4+ T cells are critical in immunity to abscesses. When administered experimentally, *B. fragilis* polysaccharide A has immunomodulatory characteristics and stimulates CD4+ T regulatory cells via an interleukin 2–dependent mechanism to produce interleukin 10. Interleukin 10 downregulates the inflammatory response, thereby preventing abscess formation.

Clinical Presentation Of all intraabdominal abscesses, 74% are intraperitoneal or retroperitoneal and are not visceral. Most intraperitoneal abscesses result from fecal spillage from a colonic source, such as an inflamed appendix. Abscesses can also arise from other processes. They usually form within weeks of the development of peritonitis and may be found in a variety of locations from omentum to mesentery, pelvis to psoas muscles, and subphrenic space to a visceral organ such as the liver, where they may develop either on the surface of the organ or within it. Periappendiceal and diverticular abscesses occur commonly. Diverticular abscesses are least likely to rupture. Infections of the female genital tract and pancreatitis also are among the more common causative events. When abscesses occur in the female genital tract—either as a primary infection (e.g., tuboovarian abscess) or as an infection extending into the pelvic cavity or peritoneum—*B. fragilis* figures prominently among the organisms isolated. *B. fragilis* is not found in large numbers in the normal vaginal flora. For example, it is encountered less commonly in pelvic inflammatory disease and endometritis without an associated abscess. In pancreatitis with leakage of damaging pancreatic enzymes, inflammation is prominent. Therefore, clinical findings such as fever, leukocytosis, and even abdominal pain do not distinguish pancreatitis itself from complications such as pancreatic pseudocyst, pancreatic abscess **(Chap. 359)**, or intraabdominal collections of pus. Especially in cases of necrotizing pancreatitis, in which the incidence of local pancreatic infection may be as high as 30%, needle aspiration under CT guidance is performed to sample fluid for culture. Traditionally, many centers have prescribed preemptive antibiotics for patients with necrotizing pancreatitis. Imipenem is frequently used for this purpose because it reaches high tissue levels in the pancreas (although it is not unique in this regard). Randomized controlled studies have not demonstrated a benefit from this practice, and guidelines no longer recommend preemptive antibiotics for patients with acute pancreatitis. If needle aspiration yields infected fluid in the setting of acute necrotizing pancreatitis, antibiotic treatment is appropriate in conjunction with surgical and/or percutaneous drainage of infected material. Infected pseudocysts that occur remotely from acute pancreatitis are unlikely to be associated with significant amounts of necrotic tissue and may be treated with either surgical or percutaneous catheter drainage in conjunction with appropriate antibiotic therapy.

Diagnosis Scanning procedures have considerably facilitated the diagnosis of intraabdominal abscesses. Abdominal CT probably has the highest yield, although ultrasonography is particularly useful for the right upper quadrant, kidneys, and pelvis. Both indium-labeled WBCs and gallium tend to localize in abscesses and may be useful in finding a collection. Because gallium is taken up in the bowel, indium-labeled WBCs may have a slightly greater yield for abscesses near the bowel. Neither indium-labeled WBC scans nor gallium scans serve as a basis for a definitive diagnosis, however; both need to be followed by other, more specific studies, such as CT, if an area of possible abnormality is identified. PET scanning should also be considered due to its ready availability and because it provides more resolution. Abscesses contiguous with or contained within diverticula are particularly difficult to diagnose with scanning procedures. Although barium should not be injected if a perforation is suspected, a barium enema occasionally may detect a diverticular abscess not diagnosed by other procedures. If one study is negative, a second study sometimes reveals a collection. Although exploratory laparotomy has been less commonly used since the advent of CT, this procedure still must be undertaken on occasion if an abscess is strongly suspected on clinical grounds.

TREATMENT

Intraperitoneal Abscesses

An algorithm for the management of patients with intraabdominal (including intraperitoneal) abscesses by percutaneous drainage is presented in **Fig. 137-3**. Treatment of intraabdominal infections involves determination of the initial focus of infection, administration of broad-spectrum antibiotics targeting the organisms involved, and performance of a drainage procedure if one or more definitive abscesses have formed. Antimicrobial therapy, in general, is adjunctive to drainage and/or surgical correction of an underlying lesion or process in intraabdominal abscesses. Results of cultures from drain sites are not reliable for defining the etiology of infections. Unlike the intraabdominal abscesses resulting from most causes, for which drainage of some kind is generally required,

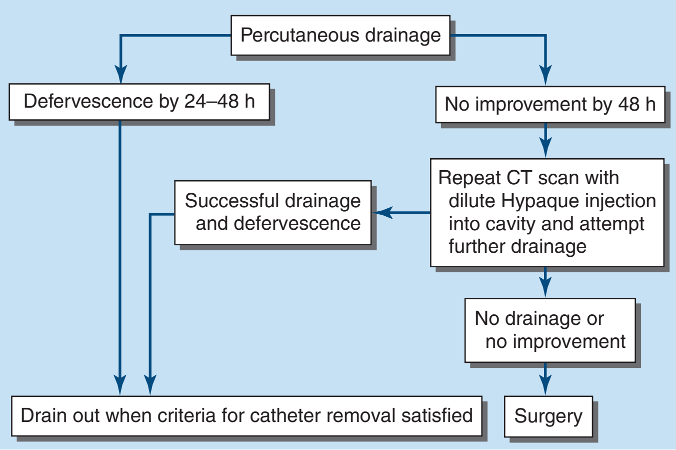
Percutaneous drainage
Defervescence by 24–48 h
No improvement by 48 h
Repeat CT scan with dilute Hypaque injection into cavity and attempt further drainage
Successful drainage and defervescence
No drainage or no improvement
Drain out when criteria for catheter removal satisfied
Surgery

FIGURE 137-3 Algorithm for the management of patients with intraabdominal abscesses by percutaneous drainage. Antimicrobial therapy should be administered concomitantly. *(Reproduced with permission from B Lorber [ed]: Atlas of Infectious Diseases, vol VII: Intra-abdominal infections, hepatitis, and gastroenteritis. Philadelphia, Current Medicine, 1996, p 1.30, as adapted from OD Rotstein, RL Simmons, in SL Gorbach et al [eds]: Infectious Diseases. Philadelphia, Saunders; 1992.)*

abscesses associated with diverticulitis usually wall off locally after rupture of a diverticulum, so that surgical intervention is not routinely required.

A number of agents exhibit excellent activity against aerobic gram-negative bacilli. Because death in intraabdominal sepsis is linked to gram-negative bacteremia, empirical therapy for intraabdominal infection always needs to include adequate coverage of gram-negative aerobic, facultative, and anaerobic organisms. Even if anaerobes are not cultured from clinical specimens, they still must be covered by the therapeutic regimen. Empirical antibiotic therapy should be the same as that discussed above for secondary peritonitis. Most clinical treatment failures are due to failure to drain the abscess and thereby achieve source control. The appropriate duration of antibiotic treatment for abdominal abscesses depends on whether the presumptive source of the intraabdominal infection has been controlled. With adequate source control, antibiotic treatment may be limited to 4 or 5 days.

VISCERAL ABSCESSES

Liver Abscesses The liver is the organ most subject to the development of abscesses. In one study of 540 intraabdominal abscesses, 26% were visceral. Liver abscesses made up 13% of the total number, or 48% of all visceral abscesses. Liver abscesses may be solitary or multiple; they may arise from hematogenous spread of bacteria or from local spread from contiguous sites of infection within the peritoneal cavity. In the past, appendicitis with rupture and subsequent spread of infection was the most common source for a liver abscess. Currently, associated disease of the biliary tract is most common. Pylephlebitis (suppurative thrombosis of the portal vein), usually arising from infection in the pelvis but sometimes from infection elsewhere in the peritoneal cavity, is another common source for bacterial seeding of the liver.

Fever is the most common presenting sign of liver abscess. Some patients, particularly those with associated disease of the biliary tract, have symptoms and signs localized to the right upper quadrant, including pain, guarding, punch tenderness, and even rebound tenderness. Nonspecific symptoms, such as chills, anorexia, weight loss, nausea, and vomiting, also may develop. Only 50% of patients with liver abscesses, however, have hepatomegaly, right-upper-quadrant tenderness, or jaundice; thus, one-half of patients have no symptoms or signs to direct attention to the liver. Fever of unknown origin may be the only manifestation of liver abscess, especially in the elderly. Diagnostic studies of the abdomen, especially the right upper quadrant, should be a part of any workup for fever of unknown origin. The single most reliable laboratory finding is an elevated serum concentration of alkaline phosphatase, which is documented in 70% of patients with liver abscesses. Other tests of liver function may yield normal results, but 50% of patients have elevated serum levels of bilirubin, and 48% have elevated concentrations of aspartate aminotransferase. Other laboratory findings include leukocytosis in 77% of patients, anemia (usually normochromic, normocytic) in 50%, and hypoalbuminemia in 33%. Concomitant bacteremia is found in one-third to one-half of patients. A liver abscess is sometimes suggested by chest radiography, especially if a new elevation of the right hemidiaphragm is seen; other suggestive findings include a right basilar infiltrate and a right pleural effusion.

Imaging studies are the most reliable methods for diagnosing liver abscesses. These studies include ultrasonography, CT **(Fig. 137-4)**, indium-labeled WBC or gallium scan, and MRI. More than one such study may be required.

Organisms recovered from liver abscesses vary with the source. In liver infection arising from the biliary tree, enteric gram-negative aerobic bacilli and enterococci are common isolates. *Klebsiella pneumoniae* liver abscess has been well described in Southeast Asia for more than 20 years and has become an emerging syndrome in North America and elsewhere. These community-acquired infections have been linked to a virulent hypermucoviscous *K. pneumoniae* phenotype and to a specific genotype. The typical syndrome includes liver abscess, bacteremia, and metastatic infection in the eye(s) and lung(s). Ampicillin/amoxicillin therapy started within the previous 30 days has been associated with increased risk for this syndrome, presumably because of selection for the causative strain. Unless previous surgery has been performed, anaerobes are not generally involved in liver abscesses arising from biliary infections. In contrast, in liver abscesses arising from pelvic and other intraperitoneal sources, a mixed flora including both aerobic and anaerobic species is common; *B. fragilis* is the species most frequently isolated. With hematogenous spread of infection, usually only a single organism is encountered; this species may be *S. aureus* or a streptococcal species such as one in the *Streptococcus milleri* group. Liver abscesses may also be caused by *Candida* species; such abscesses usually follow fungemia in patients receiving chemotherapy for cancer and often present when PMNs return after a period of neutropenia. Amebic liver abscesses are not an uncommon problem **(Chap. 230)**. Amebic serologic testing gives positive results in >95% of cases. In addition, polymerase chain reaction (PCR) testing has been used in recent years. Negative results from these studies help to exclude this diagnosis.

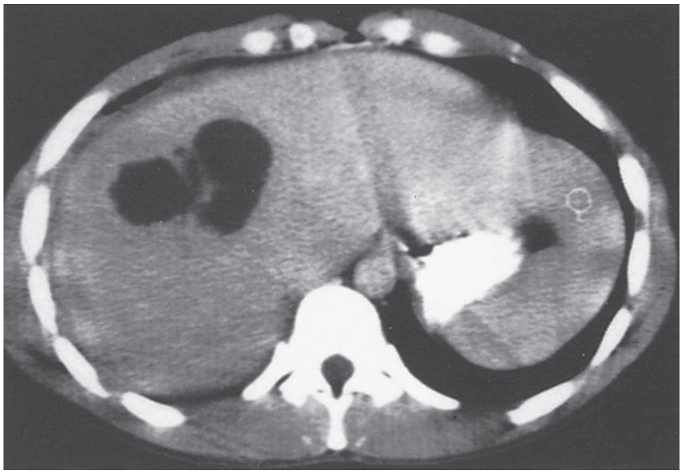

FIGURE 137-4 Multilocular liver abscess on CT scan. Multiple or multilocular abscesses are more common than solitary abscesses. *(Reprinted with permission from B Lorber [ed]: Atlas of Infectious Diseases, vol VII: Intra-abdominal Infections, Hepatitis, and Gastroenteritis. Philadelphia, Current Medicine, 1996, Fig. 1.22.)*

TREATMENT

Liver Abscesses

(Fig. 137-3) Drainage is the mainstay of therapy for intraabdominal abscesses, including liver abscesses; the approach can be either percutaneous (with a pigtail catheter kept in place or possibly with a device that can perform pulse lavage to fragment and evacuate the semisolid contents of a liver abscess), transluminal (with endoscopic ultrasound guidance), or surgical. However, there is growing interest in medical management alone for pyogenic liver abscesses. The drugs used for empirical therapy include the same ones used in intraabdominal sepsis and secondary bacterial peritonitis. Usually, blood cultures and a diagnostic aspirate of abscess contents should be obtained before the initiation of empirical therapy, with antibiotic choices adjusted when the results of Gram's staining and culture become available. Cases treated without definitive drainage generally require longer courses of antibiotic therapy. When percutaneous drainage was compared with open surgical drainage, the average length of hospital stay for the former was almost twice that for the latter, although both the time required for fever to resolve and the mortality rate were the same for the two procedures. The mortality rate was appreciable despite treatment, averaging 15%. Several factors predict the failure of percutaneous drainage and therefore may favor primary surgical intervention. These factors include the presence of multiple, sizable abscesses; viscous abscess contents that tend to plug the catheter; associated disease (e.g., disease of the biliary tract) requiring surgery; the presence of yeast; communication with an untreated obstructed biliary tree; or the lack of a clinical response to percutaneous drainage in 4–7 days.

Treatment of candidal liver abscesses often entails initial administration of liposomal amphotericin B (3–5 mg/kg IV daily) or an echinocandin, with subsequent fluconazole therapy **(Chap. 222)**. In some cases, therapy with fluconazole alone (6 mg/kg daily) may be used—e.g., in clinically stable patients whose infecting isolate is susceptible to this drug.

Splenic Abscesses Splenic abscesses are much less common than liver abscesses. The incidence of splenic abscesses has ranged from 0.14 to 0.7% in various autopsy series. The clinical setting and the organisms isolated usually differ from those for liver abscesses. The degree of clinical suspicion for splenic abscess needs to be high because this condition is frequently fatal if left untreated. Even in the most recently published series, diagnosis was made only at autopsy in 37% of cases. Although splenic abscesses may arise occasionally from contiguous spread of infection or from direct trauma to the spleen, hematogenous spread of infection is more common. Bacterial endocarditis is the most common associated infection **(Chap. 133)**. Splenic abscesses can develop in patients who have received extensive immunosuppressive therapy (particularly those with malignancy involving the spleen) and in patients with hemoglobinopathies or other hematologic disorders (especially sickle cell anemia).

Although ~50% of patients with splenic abscesses have abdominal pain, the pain is localized to the left upper quadrant in only one-half of these cases. Splenomegaly is found in ~50% of cases. Fever and leukocytosis are generally present; the development of fever preceded diagnosis by an average of 20 days in one series. Left-sided chest findings may include abnormalities to auscultation, and chest radiographic findings may include an infiltrate or a left-sided pleural effusion. CT scan of the abdomen has been the most sensitive diagnostic tool. Ultrasonography can yield the diagnosis but is less sensitive. Liver–spleen scan or gallium scan also may be useful. Streptococcal species are the most common bacterial isolates from splenic abscesses, followed by *S. aureus*—presumably reflecting the associated endocarditis. An increase in the prevalence of gram-negative aerobic isolates from splenic abscesses has been reported; these organisms often derive from a urinary tract focus, with associated bacteremia, or from another intraabdominal source. *Salmonella* species are seen fairly commonly, especially in patients with sickle cell hemoglobinopathy. Anaerobic species accounted for only 5% of isolates in the largest collected series, but the reporting of a number of “sterile abscesses” may indicate that optimal techniques for the isolation of anaerobes were not used.

TREATMENT

Splenic Abscesses

Because of the high mortality figures reported for splenic abscesses, splenectomy with adjunctive antibiotics has traditionally been considered standard treatment and remains the best approach for complex, multilocular abscesses or multiple abscesses. However, percutaneous drainage has worked well for single, small (<3-cm) abscesses in some studies and may also be useful for patients with high surgical risk. Patients undergoing splenectomy should be vaccinated against encapsulated organisms (*Streptococcus pneumoniae*, *Haemophilus influenzae*, *Neisseria meningitidis*). The most important factor in successful treatment of splenic abscesses is early diagnosis.

Perinephric and Renal Abscesses Perinephric and renal abscesses are not common. The former accounted for only ~0.02% of hospital admissions and the latter for ~0.2% in Altemeier’s series of 540 intraabdominal abscesses. Before antibiotics became available, most renal and perinephric abscesses were hematogenous in origin, usually complicating prolonged bacteremia, with *S. aureus* most commonly recovered. Now, in contrast, >75% of perinephric and renal abscesses arise from a urinary tract infection. Infection ascends from the bladder to the kidney, with pyelonephritis preceding abscess development. Bacteria may directly invade the renal parenchyma from medulla to cortex. Local vascular channels within the kidney may facilitate the transport of organisms. Areas of abscess developing within the parenchyma may rupture into the perinephric space. The kidneys and adrenal glands are surrounded by a layer of perirenal fat that, in turn, is surrounded by Gerota’s fascia, which extends superiorly to the diaphragm and inferiorly to the pelvic fat. Abscesses extending into the perinephric space may track through Gerota’s fascia into the psoas or transversalis muscles, into the anterior peritoneal cavity, superiorly to the subdiaphragmatic space, or inferiorly to the pelvis. Of the risk factors that have been associated with the development of perinephric abscesses, the most important is concomitant nephrolithiasis obstructing urinary flow. Of patients with perinephric abscess, 20–60% have renal stones. Other structural abnormalities of the urinary tract, prior urologic surgery, trauma, and diabetes mellitus also have been identified as risk factors.

The organisms most frequently encountered in perinephric and renal abscesses are *E. coli*, *Proteus* species, and *Klebsiella* species. *E. coli*, the aerobic species most commonly found in the colonic flora, seems to have unique virulence properties in the urinary tract, including factors promoting adherence to uroepithelial cells. The urease of *Proteus* species splits urea, thereby creating a more alkaline and more hospitable environment for bacterial proliferation. *Proteus* species are frequently found in association with large struvite stones caused by the precipitation of magnesium ammonium sulfate in an alkaline environment. These stones serve as a nidus for recurrent urinary tract infection. Although a single bacterial species is usually recovered from a perinephric or renal abscess, multiple species may be found. If a urine culture is not contaminated with periurethral flora and is found to contain more than one organism, a perinephric or renal abscess should be considered in the differential diagnosis. Urine cultures may also be polymicrobial in cases of bladder diverticulum.

Candida species can cause renal abscesses. Fungi of this genus may spread to the kidney hematogenously or by ascension from the bladder. The hallmark of the latter route of infection is ureteral obstruction with large fungal balls.

The presentation of perinephric and renal abscesses is quite nonspecific. Flank pain and abdominal pain are common. At least 50% of patients are febrile. Pain may be referred to the groin or leg, particularly with extension of infection. The diagnosis of perinephric abscess, like that of splenic abscess, is frequently delayed, and the mortality rate in some series is appreciable, although lower than in the past. Perinephric or renal abscess should be most seriously considered when a patient presents with symptoms and signs of pyelonephritis and remains febrile after 4 or 5 days of treatment. Moreover, when a urine culture yields a polymicrobial flora, when a patient is known to have renal stones, or when fever and pyuria coexist with a sterile urine culture, these diagnoses should be entertained.

Renal ultrasonography and abdominal CT are the most useful diagnostic modalities. If a renal or perinephric abscess is diagnosed, nephrolithiasis should be excluded, especially when a high urinary pH suggests the presence of a urea-splitting organism.

TREATMENT

Perinephric and Renal Abscesses

Treatment for perinephric and renal abscesses, like that for other intraabdominal abscesses, includes drainage of pus and antibiotic therapy directed at the organism(s) recovered. For perinephric abscesses, percutaneous drainage is usually successful.

Psoas Abscesses The psoas muscle is another location in which abscesses are encountered. Psoas abscesses may arise from a hematogenous source, by contiguous spread from an intraabdominal or pelvic process, or by contiguous spread from nearby bony structures (e.g., vertebral bodies). Associated osteomyelitis due to spread from bone to muscle or from muscle to bone is common in psoas abscesses. When Pott’s disease was common, *Mycobacterium tuberculosis* was a frequent cause of psoas abscess. Currently, either *S. aureus* or a mixture of enteric organisms including aerobic and anaerobic gram-negative bacilli is usually isolated from psoas abscesses in the United States. *S. aureus* is most likely to be isolated when a psoas abscess arises from hematogenous spread or a contiguous focus of osteomyelitis; a mixed enteric flora is the most likely etiology when the abscess has an intraabdominal or pelvic source. Patients with psoas abscesses frequently present with fever, lower abdominal or back pain, or pain referred to the hip or knee. CT is the most useful diagnostic technique.

TREATMENT

Psoas Abscesses

Treatment includes surgical drainage and the administration of an antibiotic regimen directed at the inciting organism(s).

Pancreatic Abscesses See Chap. 359.

FURTHER READING

Biggins SW et al: Diagnosis, evaluation, and management of ascites, spontaneous bacterial peritonitis, and hepatorenal syndrome: 2021 practice guidance by the American Association for the Study of Liver Diseases. Hepatology 74;1014.2021.

De Pascale G et al: Poor timing and failure of source control are risk factors for mortality in critically ill patients with secondary peritonitis. Intensive Care Med 48:1593, 2022.

Li PK et al: ISPD peritonitis recommendations: 2016 update on prevention and treatment. Perit Dial Int 36:481, 2016.

Oliver A et al: Role of rifaximin in spontaneous bacterial peritonitis prevention. South Med J 111:660 2018.

Roediger R, Lisker-Melman M: Pyogenic and amebic infections of the liver. Gastroenterol Clin North Am 49:361, 2020.

Ross JT et al: Secondary peritonitis: Principles of diagnosis and intervention. BMJ 361:k1407, 2018.

Van Hooste W et al: Infections caused by hypervirulent *Klebsiella pneumoniae* in non-endemic countries: Three case reports and review of the literature. Acta Clin Belg 78:229, 2023.

138

Acute Infectious Diarrheal Diseases and Bacterial Food Poisoning

Ana A. Weil, Regina C. LaRocque

Diarrheal disease mortality has decreased substantially in the past three decades. Nevertheless, acute diarrheal disease is still a leading cause of illness globally and is associated with an estimated 1.57 million deaths per year. Across low-income and middle-income countries, one in ten deaths among children <5 years of age is attributable to diarrhea, with substantial variation in incidence and mortality between countries. The morbidity from diarrhea also is significant. Recurrent intestinal infections are associated with physical and mental stunting, wasting, micronutrient deficiencies, and malnutrition. In short, diarrheal disease is a driving factor in global morbidity and mortality.

The wide range of clinical manifestations of acute gastrointestinal illnesses is matched by the wide variety of infectious agents involved, including viruses, bacteria, and parasites **(Table 138-1)**. This chapter discusses factors that enable gastrointestinal pathogens to cause disease, reviews host defense mechanisms, and delineates an approach to the evaluation and treatment of patients presenting with acute diarrhea. Individual organisms causing acute gastrointestinal illnesses are discussed in detail in subsequent chapters.

PATHOGENIC MECHANISMS

Enteric pathogens have developed a variety of tactics to overcome host defenses. Understanding the virulence factors employed by these organisms is important in the diagnosis and treatment of clinical disease.

INOCULUM SIZE

The number of microorganisms that must be ingested to cause disease varies considerably from species to species. For *Shigella*, enterohemorrhagic *Escherichia coli*, *Giardia lamblia*, or *Entamoeba*, as few as 10–100 bacteria or cysts can produce infection, while 10^5–10^8 *Vibrio cholerae* organisms must be ingested to cause disease. The infective dose of *Salmonella* varies widely, depending on the species, host, and food vehicle. The ability of organisms to overcome host defenses has important implications for transmission; *Shigella*, enterohemorrhagic *E. coli*, *Entamoeba*, and *Giardia* can spread by person-to-person contact, whereas under some circumstances, *Salmonella* may need to grow in food for several hours before reaching an effective infectious dose.

ADHERENCE

Many organisms must adhere to the gastrointestinal mucosa as an initial step in the pathogenic process; thus, organisms that can compete with the normal bowel flora and colonize the mucosa have an important advantage in causing disease. Specific cell-surface proteins involved in attachment of bacteria to intestinal cells are important virulence determinants. *V. cholerae*, for example, adheres to the brush border of small-intestinal enterocytes via specific surface adhesins, including the toxin-coregulated pilus and other accessory colonization factors. Enterotoxigenic *E. coli*, which causes watery diarrhea, produces an adherence protein called *colonization factor antigen* that is necessary for colonization of the upper small intestine by the organism prior to the production of enterotoxin. Enteropathogenic *E. coli*, an agent of diarrhea in young children, and enterohemorrhagic *E. coli*, which causes hemorrhagic colitis and the hemolytic-uremic syndrome, produce virulence determinants that allow these organisms to attach to and efface the brush border of the intestinal epithelium.

TOXIN PRODUCTION

The production of one or more exotoxins is important in the pathogenesis of numerous enteric organisms. Such toxins include *enterotoxins*, which cause watery diarrhea by acting directly on secretory mechanisms in the intestinal mucosa; *cytotoxins*, which cause destruction of mucosal cells and associated inflammatory diarrhea; and *neurotoxins*, which act directly on the central or peripheral nervous system.

The prototypical enterotoxin is cholera toxin, a heterodimeric protein composed of one A and five B subunits. The A subunit contains the enzymatic activity of the toxin, while the B subunit pentamer binds holotoxin to the enterocyte surface receptor, the ganglioside G_{M1}. After the binding of holotoxin, a fragment of the A subunit is translocated across the eukaryotic cell membrane into the cytoplasm, where it catalyzes the adenosine diphosphate ribosylation of a guanosine triphosphate (GTP)–binding protein and causes persistent activation of adenylate cyclase. The result is an increase of cyclic adenosine monophosphate in the intestinal cell, which increases Cl^- secretion and decreases Na^+ absorption, leading to a loss of fluid and the production of diarrhea.

Enterotoxigenic strains of *E. coli* may produce a protein called *heat-labile enterotoxin* (LT) that is like cholera toxin and causes secretory diarrhea by the same mechanism. Alternatively, enterotoxigenic strains of *E. coli* may produce *heat-stable enterotoxin* (ST), one form of which causes diarrhea by activation of guanylate cyclase and elevation of intracellular cyclic guanosine monophosphate. Some enterotoxigenic strains of *E. coli* produce both LT and ST.

Bacterial cytotoxins, in contrast, destroy intestinal mucosal cells and produce the syndrome of dysentery, with bloody stools containing

TABLE 138-1 Gastrointestinal Pathogens Causing Acute Diarrhea

MECHANISM	LOCATION	ILLNESS	STOOL FINDINGS	EXAMPLES OF PATHOGENS INVOLVED
Noninflammatory (enterotoxin)	Proximal small bowel	Watery diarrhea	No fecal leukocytes; mild or no increase in fecal lactoferrin	*Vibrio cholerae*, enterotoxigenic *Escherichia coli* (LT and/or ST), enteroaggregative *E. coli*, *Clostridium perfringens*, *Bacillus cereus*, *Staphylococcus aureus*, *Aeromonas hydrophila*, *Plesiomonas shigelloides*, rotavirus, norovirus, enteric adenoviruses, *Giardia lamblia*, *Cryptosporidium* spp., *Cyclospora* spp., microsporidia
Inflammatory (invasion or cytotoxin)	Colon or distal small bowel	Dysentery or inflammatory diarrhea	Fecal polymorphonuclear leukocytes; substantial increase in fecal lactoferrin	*Shigella* spp., *Salmonella* spp., *Campylobacter jejuni*, enterohemorrhagic *E. coli*, enteroinvasive *E. coli*, *Yersinia enterocolitica*, *Listeria monocytogenes*, *Vibrio parahaemolyticus*, *Clostridium difficile*, *A. hydrophila*, *P. shigelloides*, *Entamoeba histolytica*, *Klebsiella oxytoca*
Penetrating	Distal small bowel	Enteric fever	Fecal mononuclear leukocytes	*Salmonella* Typhi, *Y. enterocolitica*

Abbreviations: LT, heat-labile enterotoxin; ST, heat-stable enterotoxin.

inflammatory cells. Enteric pathogens that produce such cytotoxins include *Shigella dysenteriae* type 1, *Vibrio parahaemolyticus*, and *Clostridioides difficile*. *S. dysenteriae* type 1 and Shiga toxin–producing strains of *E. coli* produce potent cytotoxins and have been associated with outbreaks of hemorrhagic colitis and hemolytic-uremic syndrome.

Neurotoxins are usually produced by bacteria outside the host and therefore cause symptoms soon after ingestion. Included are the staphylococcal and *Bacillus cereus* toxins, which act on the central nervous system to produce vomiting.

■ INVASION

Dysentery may result not only from the production of cytotoxins but also from bacterial invasion and destruction of intestinal mucosal cells. Infections due to *Shigella* and enteroinvasive *E. coli* are characterized by the organisms' invasion of mucosal epithelial cells, intraepithelial multiplication, and subsequent spread to adjacent cells. *Salmonella* causes inflammatory diarrhea by invading the bowel mucosa but generally is not associated with the destruction of enterocytes or the full clinical syndrome of dysentery. *Salmonella enterica* serovar Typhi and *Yersinia enterocolitica* can penetrate intact intestinal mucosa, multiply intracellularly in Peyer patches and intestinal lymph nodes, and then disseminate through the bloodstream to cause enteric fever—a syndrome characterized by fever, headache, relative bradycardia, abdominal pain, splenomegaly, and leukopenia.

HOST DEFENSES

Given the enormous number of microorganisms ingested with every meal, the normal host must combat a constant influx of potential enteric pathogens. Studies of infections in patients with alterations in defense mechanisms have led to a greater understanding of the variety of ways in which the normal host can protect itself against disease.

■ INTESTINAL MICROBIOTA

The large numbers of bacteria that normally inhabit the intestine (*the intestinal microbiota*) act as an important host defense mechanism, preventing colonization by potential enteric pathogens. Mechanisms of colonization resistance are not fully understood but include geographic and nutritional exclusion at the mucosal surface. Persons with fewer intestinal bacteria, such as infants who have not yet developed normal enteric colonization or patients receiving antibiotics, are at greater risk of developing infections with enteric pathogens. The bacterial community composition of the intestinal microbiota is as important as abundance of organisms present. More than 99% of the normal colonic microbiota is made up of anaerobic bacteria, and the acidic pH and volatile fatty acids produced by these organisms appear to be critical elements in resistance to colonization.

■ GASTRIC ACID

The acidic pH of the stomach is an important barrier to enteric pathogens, and an increased frequency of infections due to *Salmonella*, *G. lamblia*, and a variety of helminths has been reported among patients who have undergone gastric surgery or are achlorhydric for some other reason. Neutralization of gastric acid with antacids, proton pump inhibitors, or H_2 blockers—a common practice in the management of hospitalized patients—similarly increases the risk of enteric colonization. In addition, some microorganisms can survive the extreme acidity of the gastric environment; rotavirus and *Shigella*, for example, are highly stable to acidity.

■ INTESTINAL MOTILITY

Peristalsis is the major mechanism for clearance of bacteria from the proximal small intestine. When intestinal motility is impaired (e.g., by treatment with opioids or other antimotility drugs, anatomic abnormalities, or hypomotility states), the frequency of bacterial overgrowth and infection of the small bowel with enteric pathogens is increased. Some patients whose treatment for *Shigella* infection consists of diphenoxylate hydrochloride with atropine (Lomotil) experience prolonged fever and shedding of organisms, while patients treated with opioids for mild *Salmonella* gastroenteritis have a higher frequency of bacteremia than those not treated with opioids.

■ INTESTINAL MUCIN

A complex layer of mucus, produced by specialized secretory cells, covers the stomach, small intestine, and large intestine, and separates the lumenal commensal microbiota from the epithelium. The thickness and constituents of this mucus barrier vary throughout the gastrointestinal tract. The mucus barrier turns over rapidly and comprises glycoproteins and a range of antimicrobial molecules and secreted immunoglobulins directed against specific microbial antigens. Enteric pathogens have evolved a wide range of strategies to overcome this barrier and thus to reach the underlying epithelium and cause disease. For example, pathogens can penetrate the mucus layer by secreting enzymes to degrade the mucus or through flagella-mediated motility. Some organisms, such as *Shigella*, secrete toxins that can diffuse through the mucus layer and disrupt the underlying epithelium. The resulting reduction of mucus production allows the pathogen to reach the cell surface.

■ IMMUNITY

Both cellular immune responses and antibody production play important roles in protection from enteric infections. Humoral immunity to enteric pathogens consists of systemic IgG and IgM as well as secretory IgA. The mucosal immune system may be the first line of defense against many gastrointestinal pathogens. The binding of bacterial antigens to the lumenal surface of M cells in the distal small bowel and the subsequent presentation of antigens to subepithelial lymphoid tissue leads to the proliferation of sensitized and specialized lymphocytes. These lymphocytes circulate and populate all of the mucosal tissues of the body as IgA-secreting plasma cells.

■ GENETIC DETERMINANTS

Host genetic variation influences susceptibility to diarrheal diseases. People with blood group O show increased susceptibility to disease due to *V. cholerae*, *Shigella*, *E. coli* O157, and norovirus. Polymorphisms in genes encoding inflammatory mediators have been associated with the outcome of infection with enteroaggregative *E. coli*, enterotoxin-producing *E. coli*, *Salmonella*, *C. difficile*, and *V. cholerae*.

APPROACH TO THE PATIENT

Infectious Diarrhea or Bacterial Food Poisoning

The approach to the patient with possible infectious diarrhea or bacterial food poisoning is shown in **Fig. 138-1**.

HISTORY

Diarrhea is defined as the passage of loose or watery stools that conform to the shape of a container three or more times in a 24-h period. The answers to questions with high discriminating value can quickly narrow the range of potential causes of diarrhea and help determine whether treatment is needed. Important elements of the narrative history are detailed in Fig. 138-1.

PHYSICAL EXAMINATION

The examination of patients for signs of dehydration provides essential information about the severity of the diarrheal illness and the need for rapid therapy. Mild dehydration is indicated by thirst, dry mouth, decreased axillary sweat, decreased urine output, and slight weight loss. Signs of moderate dehydration include an orthostatic fall in blood pressure, skin tenting, and sunken eyes (or, in infants, a sunken fontanelle). Signs of severe dehydration include lethargy, obtundation, feeble pulse, hypotension, and shock.

DIAGNOSTIC APPROACH

After assessing severity of illness, the clinician must distinguish between *inflammatory* and *noninflammatory* disease. Using the

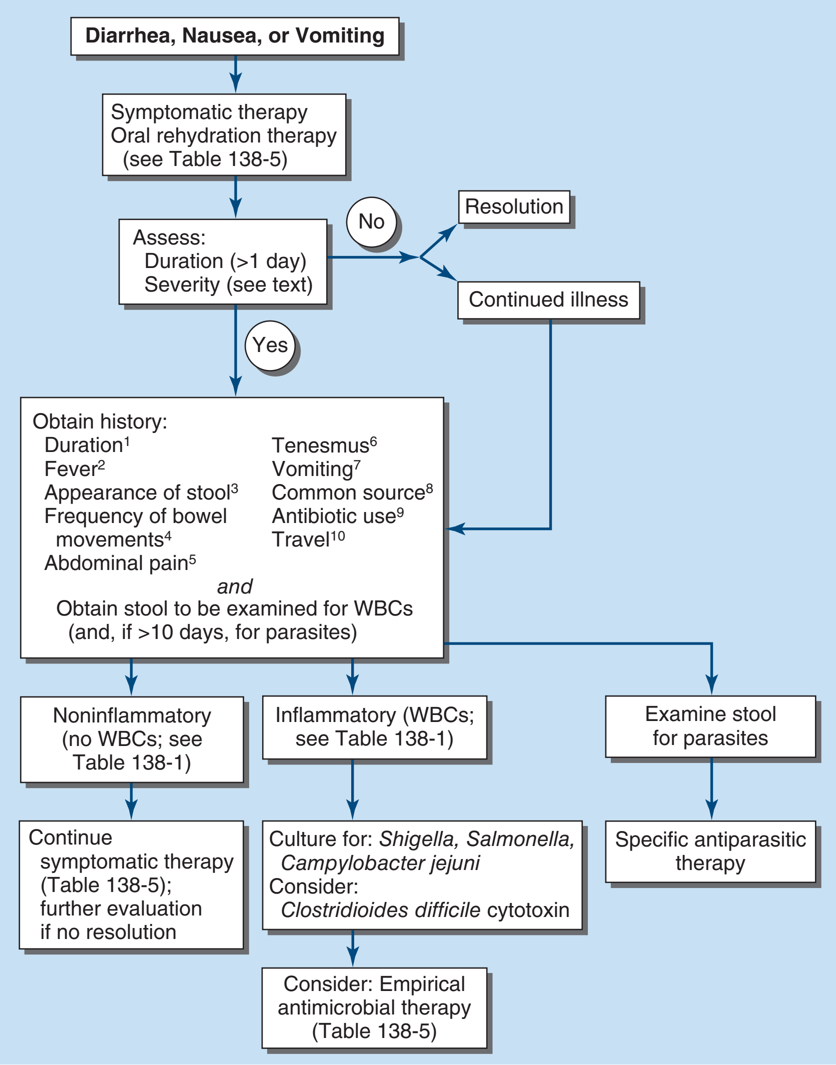
Diarrhea, Nausea, or Vomiting
Symptomatic therapy
Oral rehydration therapy
(see Table 138-5)
Assess:
Duration (>1 day)
Severity (see text)
No
Resolution
Continued illness
Yes
Obtain history:
Duration¹
Fever²
Appearance of stool³
Frequency of bowel movements⁴
Abdominal pain⁵
Tenesmus⁶
Vomiting⁷
Common source⁸
Antibiotic use⁹
Travel¹⁰
and
Obtain stool to be examined for WBCs
(and, if >10 days, for parasites)
Noninflammatory (no WBCs; see Table 138-1)
Inflammatory (WBCs; see Table 138-1)
Examine stool for parasites
Continue symptomatic therapy (Table 138-5); further evaluation if no resolution
Culture for: Shigella, Salmonella, Campylobacter jejuni
Consider: Clostridioides difficile cytotoxin
Specific antiparasitic therapy
Consider: Empirical antimicrobial therapy (Table 138-5)

FIGURE 138-1 Clinical algorithm for the approach to patients with community-acquired infectious diarrhea or bacterial food poisoning. Key to superscripts: 1. Diarrhea lasting >2 weeks is generally defined as chronic; in such cases, many of the causes of acute diarrhea are much less likely, and a new spectrum of causes needs to be considered. 2. Fever often implies invasive disease, although fever and diarrhea may also result from infection outside the gastrointestinal tract, as in malaria. 3. Stools that contain blood or mucus indicate ulceration of the large bowel. Bloody stools without fecal leukocytes should alert the laboratory to the possibility of infection with Shiga toxin–producing enterohemorrhagic *Escherichia coli*. Bulky white/gray stools suggest a small-intestinal process that is causing malabsorption. An acute illness with profuse "rice-water" stools, with an appearance of water after rice has been cooked, suggests cholera or a similar toxigenic process. 4. Frequent stools over a given period can provide the first warning of impending dehydration. 5. Abdominal pain may be most severe in inflammatory processes like those due to *Shigella*, *Campylobacter*, and necrotizing toxins. Painful abdominal muscle cramps, caused by electrolyte loss, can develop in severe cases of cholera. Bloating is common in giardiasis. An appendicitis-like syndrome should prompt a culture for *Yersinia enterocolitica* with cold enrichment. 6. Tenesmus (painful rectal spasms with a strong urge to defecate but little passage of stool) may be a feature of cases with proctitis, as in shigellosis or amebiasis. 7. Vomiting implies an acute infection (e.g., a toxin-mediated illness or food poisoning) but can also be prominent in a variety of systemic illnesses (e.g., malaria) and in intestinal obstruction. 8. Asking patients whether anyone they know is sick is a more efficient means of identifying a common source than is constructing a list of recently eaten foods. If a common source seems likely, specific foods can be investigated. See text for a discussion of bacterial food poisoning. 9. Current antibiotic therapy or a recent history of treatment suggests *Clostridioides difficile* diarrhea (**Chap. 139**). Stop antibiotic treatment if possible and consider tests for *C. difficile* toxins. Antibiotic use may increase the risk of chronic intestinal carriage following salmonellosis. 10. See text (and **Chap. 130**) for a discussion of traveler's diarrhea. *(From The New England Journal of Medicine. Bacterial and Protozoal Gastroenteritis. RL Guerrant, DA Bobak 325:327-340. Copyright @1991 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

history and epidemiologic features of the case, the clinician can then rapidly evaluate the need for further efforts to define a specific etiology and for therapeutic intervention. Examination of a stool sample may supplement the narrative history. Grossly bloody or mucoid stool suggests an inflammatory process. A test for fecal leukocytes (preparation of a thin smear of stool on a glass slide, addition of a drop of methylene blue, and examination of the wet mount) can suggest inflammatory disease in patients with diarrhea, although the predictive value of this test is still debated. A test for fecal lactoferrin, which is a marker of fecal leukocytes, is more sensitive and is available in latex agglutination and enzyme-linked immunosorbent assay formats. Stool culture or molecular testing is useful for diagnosing a causative organism, particularly in patients with severe illness or high-risk comorbidities. Causes of acute infectious diarrhea, categorized as inflammatory and noninflammatory, are listed in Table 138-1.

POST DIARRHEA COMPLICATIONS

Short- or long-term complications may follow the resolution of an acute diarrheal episode. The clinician should inquire about prior diarrheal illness if the conditions listed in **Table 138-2** are observed.

EPIDEMIOLOGY

TRAVEL HISTORY

Of the several million people who travel from temperate industrialized countries to tropical regions of Asia, Africa, and Central and South America each year, 20–50% experience a sudden onset of abdominal cramps, anorexia, and watery diarrhea; thus, *traveler's diarrhea* is the most common travel-related infectious illness **(Chap. 130)**. The time of onset is usually 3 days to 2 weeks after the traveler's arrival in a resource-poor area; most cases begin within the first 3–5 days. The illness is generally self-limited, lasting 1–5 days. The high rate of diarrhea among travelers is related to the ingestion of contaminated food or water.

The pathogens that cause traveler's diarrhea vary considerably with location **(Table 138-3)**, as does the pattern of antimicrobial resistance. In all areas, enterotoxigenic and enteroaggregative *E. coli* are the most common pathogens among persons with the classic secretory traveler's diarrhea syndrome. Infection with *Campylobacter jejuni* is especially common in areas of Asia.

TABLE 138-3 Causes of Traveler's Diarrhea

ETIOLOGIC AGENT	APPROXIMATE PERCENTAGE OF CASES	COMMENTS
Bacteria	**50–75**	
Enterotoxigenic *Escherichia coli*	10–45	Single most important agent
Enteroaggregative *E. coli*	5–35	Emerging enteric pathogen with worldwide distribution
Campylobacter jejuni	5–25	More common in Asia
Shigella	0–15	Major cause of dysentery
Salmonella	0–15	—
Others	0–5	Including *Aeromonas*, *Plesiomonas*, and *Vibrio cholerae*
Viruses	**0–20**	
Norovirus	0–10	Associated with cruise ships
Rotavirus	0–5	Particularly common among children
Parasites	**0–10**	
Giardia lamblia	0–5	Affects hikers and campers who drink from freshwater streams
Cryptosporidium	0–5	Resistant to chlorine treatment of water sources
Entamoeba histolytica	<1	—
Cyclospora	<1	—
Other	**0–10**	
Acute food poisoning[a]	0–5	—
No pathogen identified	10–50	—

[a] For etiologic agents, see Table 138-4.

Source: After DR Hill et al: The practice of travel medicine: Guidelines by the Infectious Diseases Society of America. Clin Infect Dis 43:1499, 2006.

LOCATION

Closed and semi-closed communities, including day-care centers, schools, residential facilities, and cruise ships, are important settings for outbreaks of enteric infections. Norovirus, which is highly contagious and robust in surviving on surfaces, is the most common etiologic agent associated with outbreaks of acute gastroenteritis. Other common organisms, often spread by fecal–oral contact in such communities, are *Shigella*, *C. jejuni*, and *Cryptosporidium*. Rotavirus is rarely a cause of pediatric diarrheal outbreaks in the United States since rotavirus vaccination was broadly recommended in 2006. Similarly, hospitals are sites in which enteric infections are concentrated. Diarrhea is one of the most common manifestations of nosocomial infections. *C. difficile* is the predominant cause of nosocomial diarrhea among adults in the United States, and outbreaks of norovirus infection are common in health care settings. *Klebsiella oxytoca* has been identified as a cause of antibiotic-associated hemorrhagic colitis. Enteropathogenic *E. coli* has been associated with outbreaks of diarrhea in nurseries for newborns. One-third of elderly patients in chronic-care institutions develop a significant diarrheal illness each year; more than one-half of these cases are caused by cytotoxin-producing *C. difficile*. Antimicrobial therapy can predispose to pseudomembranous colitis by altering the normal colonic microbiota, which then permits the multiplication of *C. difficile* **(Chap. 139)**.

TABLE 138-2 Postdiarrhea Complications of Acute Infectious Diarrheal Illness

COMPLICATION	COMMENTS
Chronic diarrhea (diarrhea lasting >4 weeks) • Lactase deficiency • Small-bowel bacterial overgrowth • Malabsorption syndromes (tropical and celiac sprue)	Occurs in ~1% of travelers with acute diarrhea Protozoa account for approximately one-third of cases
Initial presentation or exacerbation of inflammatory bowel disease	May be precipitated by traveler's diarrhea
Irritable bowel syndrome	Occurs in ~10% of travelers with traveler's diarrhea
Reactive arthritis	Particularly likely after infection with invasive organisms (*Shigella*, *Salmonella*, *Campylobacter*, *Yersinia*)
Hemolytic-uremic syndrome (hemolytic anemia, thrombocytopenia, and renal failure)	Follows infection with Shiga toxin–producing bacteria (*Shigella dysenteriae* type 1 and enterohemorrhagic *Escherichia coli*)
Guillain-Barré syndrome	Particularly likely after *Campylobacter* infection

AGE

Globally, most morbidity and mortality from enteric pathogens involve children <5 years of age. Breast-fed infants are protected from pathogens in contaminated food and water and derive some protection from maternal antibodies, but their risk of infection rises dramatically when they begin to eat solid foods. Exposure to rotavirus is universal, with most children experiencing their first infection in the first or second year of life if not vaccinated. Older children and adults are more commonly infected with norovirus. Other organisms with higher attack rates among children than among adults include enterotoxigenic, enteropathogenic, and enterohemorrhagic *E. coli*; *Shigella*; *C. jejuni*; and *G. lamblia*.

HOST IMMUNE STATUS

Immunocompromised hosts are at elevated risk of acute and chronic infectious diarrhea. Individuals with defects in cell-mediated immunity (including those with AIDS) are at particularly high risk of invasive enteropathies, including salmonellosis, listeriosis, and cryptosporidiosis. Individuals with hypogammaglobulinemia are at particular risk of *C. difficile* colitis and giardiasis and can develop chronic infections with viral pathogens such as norovirus. Patients with cancer are more likely to develop *C. difficile* infection due to chemotherapy-associated antibiotic exposure and frequent hospitalizations. Infectious diarrhea can be life-threatening in immunocompromised hosts, with complications including persistent infection, bacteremia, and metastatic seeding of infection. Furthermore, dehydration may compromise renal function and increase the toxicity of immunosuppressive drugs.

BACTERIAL FOOD POISONING

If the history and the stool examination indicate a noninflammatory etiology of diarrhea and there is evidence of a common-source outbreak, questions concerning the ingestion of specific foods and the time of onset of diarrhea after a meal can provide clues to the bacterial cause of the illness. Potential causes of bacterial food poisoning are shown in **Table 138-4**.

Bacterial disease caused by an enterotoxin elaborated outside the host, such as that due to *Staphylococcus aureus* or *B. cereus*, has the shortest incubation period (1–6 h) and generally lasts <12 h. Most cases of staphylococcal food poisoning are caused by contamination from *Staphylococcus*-infected human carriers. Staphylococci can multiply at a wide range of temperatures; thus, if food is left to cool slowly and remains at room temperature after cooking, the organisms will have the opportunity to form enterotoxin. Outbreaks following picnics where potato salad, mayonnaise, and cream pastries have been served offer classic examples of staphylococcal food poisoning. Diarrhea, nausea, vomiting, and abdominal cramping are common, while fever is rare.

B. cereus can produce either a syndrome with a short incubation period—the *emetic* form, mediated by a staphylococcal type of enterotoxin—or one with a longer incubation period (8–16 h)—the *diarrheal* form, caused by an enterotoxin resembling *E. coli* LT, in which diarrhea and abdominal cramps are characteristic but vomiting is uncommon. The emetic form of *B. cereus* food poisoning is associated with contaminated fried rice; the organism is common in uncooked rice, and its heat-resistant spores survive boiling. If cooked rice is not refrigerated, the spores can germinate and produce toxin. Frying before serving may not destroy the preformed, heat-stable toxin.

TABLE 138-4 Bacterial Food Poisoning

INCUBATION PERIOD, ORGANISM	SYMPTOMS	COMMON FOOD SOURCES
1–6 h		
Staphylococcus aureus	Nausea, vomiting, diarrhea	Ham, poultry, potato or egg salad, mayonnaise, cream pastries
Bacillus cereus	Nausea, vomiting, diarrhea	Fried rice
8–16 h		
Clostridium perfringens	Abdominal cramps, diarrhea (vomiting rare)	Beef, poultry, legumes, gravies
B. cereus	Abdominal cramps, diarrhea (vomiting rare)	Meats, vegetables, dried beans, cereals
>16 h		
Vibrio cholerae	Watery diarrhea	Shellfish, water
Enterotoxigenic *Escherichia coli*	Watery diarrhea	Salads, cheese, meats, water
Enterohemorrhagic *E. coli*	Bloody diarrhea	Ground beef, roast beef, salami, raw milk, raw vegetables, apple juice
Salmonella spp.	Inflammatory diarrhea	Beef, poultry, eggs, dairy products
Campylobacter jejuni	Inflammatory diarrhea	Poultry, raw milk
Shigella spp.	Dysentery	Potato or egg salad, lettuce, raw vegetables
Vibrio parahaemolyticus	Dysentery	Mollusks, crustaceans

Food poisoning due to *Clostridium perfringens* also has a slightly longer incubation period (8–14 h) and results from the survival of heat-resistant spores in inadequately cooked meat, poultry, or legumes. After ingestion, toxin is produced in the intestinal tract, causing moderately severe abdominal cramps and diarrhea; vomiting is rare, as is fever. The illness is self-limited, rarely lasting >24 h.

Not all food poisoning has a bacterial cause. Nonbacterial agents of short-incubation food poisoning include capsaicin, which is found in hot peppers, and a variety of toxins found in fish and shellfish **(Chap. 471)**.

LABORATORY EVALUATION

Many cases of noninflammatory diarrhea are self-limited or can be treated empirically, and in these instances, the clinician may not need to determine a specific etiology. Potentially pathogenic *E. coli* cannot be distinguished from normal fecal flora by routine culture, and tests to detect enterotoxins are not available in most clinical laboratories. In situations in which cholera is a concern, stool should be cultured on selective media such as thiosulfate–citrate–bile salts–sucrose (TCBS) or tellurite–taurocholate–gelatin (TTG) agar; rapid diagnostic tests are also available. A latex agglutination test has made the rapid detection of rotavirus in stool practical for many laboratories, while reverse-transcriptase polymerase chain reaction (PCR) and specific antigen enzyme immunoassays have been developed for the identification of norovirus. Stool specimens should be examined by immunofluorescence-based rapid assays, PCR, or (less sensitive) standard microscopy for *Giardia* cysts or *Cryptosporidium* if the level of clinical suspicion regarding the involvement of these organisms is high.

All patients with fever and evidence of inflammatory diarrhea acquired outside the hospital should be evaluated for *Salmonella*, *Shigella*, and *Campylobacter*. *Salmonella* and *Shigella* can be selected on MacConkey agar as non-lactose-fermenting (colorless) colonies or can be grown on *Salmonella–Shigella* agar or in selenite enrichment broth, both of which inhibit most organisms except these pathogens. Evaluation of nosocomial diarrhea should initially focus on *C. difficile*; stool culture for other pathogens in the hospital setting is extremely low yield and is not cost-effective. Toxins A and B produced by pathogenic strains of *C. difficile* can be detected by rapid enzyme immunoassays, latex agglutination tests, or PCR **(Chap. 139)**. Isolation of *C. jejuni* requires inoculation of fresh stool onto selective growth medium and incubation at 42°C in a microaerophilic atmosphere. In many laboratories in the United States, *E. coli* O157:H7 is among the most common pathogens isolated from visibly bloody stools. Shiga-toxin containing strains of this enterohemorrhagic serotype can be identified by PCR or in specialized laboratories by serotyping, but also can be identified presumptively as lactose-fermenting, indole-positive colonies of sorbitol nonfermenters (white colonies) on sorbitol MacConkey plates. If the clinical presentation suggests the possibility of intestinal amebiasis, stool should be examined by a rapid antigen detection assay or by (less sensitive and less specific) microscopy. Multiplex nucleic acid amplification methods for detection of many stool pathogens (viral, bacterial, and parasitic) are increasingly being used in clinical microbiology laboratories to decrease the time to detection of a pathogen. These tests are more sensitive and rapid than standard culture methods and may detect multiple pathogens, which may lead to challenges in interpretation. For bacterial enteric infections, the lack of a microbial isolate prevents determination of antimicrobial susceptibility and typing of strains by public health authorities in order to detect and respond to common-source outbreaks. For this reason, the Centers for Disease Control and Prevention suggests that diagnosis of an enteric bacterial infection by a nucleic acid amplification method should be followed by attempted isolation of the pathogen by culture.

TREATMENT

Infectious Diarrhea or Bacterial Food Poisoning

In many cases, a specific diagnosis is not necessary or not available to guide treatment. The clinician can proceed with the information obtained from the history, stool examination, and evaluation of dehydration severity. Empirical regimens for the treatment of traveler's diarrhea are listed in **Table 138-5**.

The mainstay of treatment is adequate rehydration. The treatment of cholera and other dehydrating diarrheal diseases was revolutionized by the development and promotion of oral rehydration solution (ORS), the efficacy of which depends on the fact that glucose-facilitated absorption of sodium and water in the small intestine remains intact in the presence of cholera toxin. The use of ORS has reduced cholera mortality rates from >50% (in untreated cases) to <1%. Several ORS formulas have been developed and tested. Initial preparations were based on the treatment of patients with cholera and included a solution containing 3.5 g of sodium chloride, 2.5 g of sodium bicarbonate (or 2.9 g of sodium citrate), 1.5 g of potassium chloride, and 20 g of glucose (or 40 g of sucrose) per liter of water. Such a preparation can still be used for the treatment of severe cholera. Many causes of secretory diarrhea, however, are associated with less electrolyte loss than occurs in cholera. Beginning in 2002, the World Health Organization recommended a reduced-osmolarity/reduced-salt ORS that is better tolerated and more effective than classic ORS. This preparation contains 2.6 g of sodium chloride, 2.9 g of trisodium citrate, 1.5 g of potassium chloride, and 13.5 g of glucose (or 27 g of sucrose) per liter of water. ORS formulations containing rice or cereal as the carbohydrate source may be even more effective than glucose-based solutions. Patients who are severely dehydrated, are unable to drink, or in whom vomiting precludes oral therapy should receive IV solutions such as Ringer's lactate in addition to ORS when the patient can ingest it safely.

TABLE 138-5 Treatment of Traveler's Diarrhea on the Basis of Clinical Features[a]

CLINICAL SYNDROME	SUGGESTED THERAPY
Watery diarrhea (no blood in stool, no fever), 1 or 2 unformed stools per day without distressing enteric symptoms	Oral fluids (oral rehydration solution, Pedialyte, Lytren, or flavored mineral water) and saltine crackers
Watery diarrhea (no blood in stool, no fever), 1 or 2 unformed stools per day with distressing enteric symptoms	Bismuth subsalicylate (for adults): 30 mL or 2 tablets (262 mg/tablet) every 30 min for 8 doses; or loperamide[b]: 4 mg initially followed by 2 mg after passage of each unformed stool, not to exceed 8 tablets (16 mg) per day (prescription dose) or 4 caplets (8 mg) per day (over-the-counter dose); drugs can be taken for 2 days. Antibacterial drug[c] can be considered in selected circumstances.
Dysentery (passage of bloody stools) or fever (>37.8°C)	Antibacterial drug[c]
Vomiting, minimal diarrhea	Bismuth subsalicylate (for adults; see dose above)
Diarrhea in infants (<2 years old)	Fluids and electrolytes (oral rehydration solution, Pedialyte, Lytren); continue feeding, especially with breast milk; seek medical attention for moderate dehydration, fever lasting >24 h, bloody stools, or diarrhea lasting more than several days

[a] All patients should take oral fluids (Pedialyte, Lytren, or flavored mineral water) plus saltine crackers. If diarrhea becomes moderate or severe, if fever persists, or if bloody stools or dehydration develops, the patient should seek medical attention. [b] Loperamide should not be used by patients with fever or dysentery; its use may prolong diarrhea in patients with infection due to *Shigella* or other invasive organisms. [c] The recommended antibacterial drugs are as follows:

If the level of suspicion is low for fluoroquinolone-resistant *Campylobacter*:

Adults: (1) A fluoroquinolone such as ciprofloxacin, 750 mg as a single dose or 500 mg bid for 3 days; levofloxacin, 500 mg as a single dose or 500 mg qd for 3 days; or norfloxacin, 800 mg as a single dose or 400 mg bid for 3 days. (2) Azithromycin, 1000 mg as a single dose or 500 mg qd for 3 days. (3) Rifaximin, 200 mg tid or 400 mg bid for 3 days (not recommended for use in dysentery).

Children: Azithromycin, 10 mg/kg on day 1, 5 mg/kg on days 2 and 3 if diarrhea persists.

If fluoroquinolone-resistant *Campylobacter* is suspected (for example, following travel to Southeast Asia):

Adults: Azithromycin (at above dose for adults). *Children:* Same as for children traveling to other areas (see above).

Source: After DR Hill et al: The practice of travel medicine: Guidelines by the Infectious Diseases Society of America. Clin Infect Dis 43:1499, 2006.

Most cases of traveler's diarrhea (usually due to enterotoxigenic or enteroaggregative *E. coli* or to *Campylobacter*) can be treated effectively with rehydration, bismuth subsalicylate, or antiperistaltic agents. Antimicrobial agents can shorten the duration of illness from 3–4 days to 24–36 h but may be associated with the acquisition of multidrug-resistant organisms; their use should therefore be reserved for severe cases. Changes in diet have not been shown to have an impact on the duration of illness, and the efficacy of probiotics to hasten recovery continues to be debated. Most individuals who present with dysentery (bloody diarrhea and fever) should be treated empirically with an antimicrobial agent (e.g., a fluoroquinolone or a macrolide, depending on local antimicrobial susceptibility patterns) pending diagnostic testing. Individuals with shigellosis should receive a 3- to 7-day course. Patients with severe or prolonged *Campylobacter* infection often benefit from antimicrobial treatment. Because of widespread resistance of *Campylobacter* to fluoroquinolones, especially in parts of Asia, a macrolide antibiotic such as erythromycin or azithromycin is preferred for this infection.

Treatment of salmonellosis must be tailored to the individual patient. Since administration of antimicrobial agents often prolongs intestinal colonization with *Salmonella*, these drugs are usually reserved for individuals at high risk of complications from disseminated salmonellosis, such as infants, patients with prosthetic devices, patients over age 50, and immunocompromised persons. Antimicrobial agents should not be administered to individuals (especially children) in whom enterohemorrhagic *E. coli* infection is suspected. Laboratory studies of enterohemorrhagic *E. coli* strains have demonstrated that many antibiotics induce replication of Shiga toxin–producing lambdoid bacteriophages, thereby significantly increasing toxin production by these strains. Clinical studies have supported these laboratory results, and antibiotics may increase by 20-fold the risk of hemolytic-uremic syndrome and renal failure during enterohemorrhagic *E. coli* infection. A clinical clue in the diagnosis of the latter infection is bloody diarrhea in a patient with a low-grade fever or who is afebrile.

PROPHYLAXIS

Improvements in safe water infrastructure to limit fecal–oral spread of enteric pathogens will be necessary if the prevalence of diarrheal diseases is to be significantly reduced in resource-poor countries. Travelers can reduce their risk of diarrhea by eating only hot, freshly cooked food; by avoiding raw vegetables, salads, and unpeeled fruit; and by drinking only boiled or treated water and avoiding ice. Historically, few travelers to tourist destinations adhere to these dietary restrictions, even after pretravel counseling. Bismuth subsalicylate is an inexpensive agent for the prophylaxis of traveler's diarrhea; it is taken at a dosage of 2 tablets (525 mg) four times a day. Treatment appears to be effective and safe for up to 3 weeks, but adverse events such as temporary darkening of the tongue, constipation, and tinnitus can occur. A meta-analysis suggests that probiotics may lessen the likelihood of traveler's diarrhea by ~15%, but further studies are needed. Prophylactic antimicrobial agents, although effective, are not generally recommended for the prevention of traveler's diarrhea except when travelers are immunosuppressed or have underlying illness that places them at high risk for morbidity from gastrointestinal infection. If prophylaxis is indicated, the nonabsorbed antibiotic rifaximin can be considered.

The possibility of exerting a major impact on the worldwide morbidity and mortality associated with diarrheal diseases has led to intensive efforts to develop effective vaccines against common bacterial and viral enteric pathogens. An effective rotavirus vaccine is available. Vaccines against *V. cholerae* are available and recommended in areas where active transmission is ongoing, although the protection they offer is incomplete and short lived. A typhoid conjugate vaccine is now recommended by the World Health Organization for use in countries where typhoid is endemic. At present, there are no effective commercially available vaccines against pathogenic *E. coli*, *Shigella*, *Campylobacter*, nontyphoidal *Salmonella*, norovirus, or intestinal parasites.

ACKNOWLEDGMENT

The authors thank Stephen B. Calderwood, MD, Edward T. Ryan, MD, and Richelle C. Charles, MD, for their significant contributions to this chapter in the previous editions.

FURTHER READING

Brown AB et al: Travel-related diagnoses among U.S. nonmigrant travelers or migrants presenting to U.S. GeoSentinel Sites – GeoSentinel Network, 2012-2021. MMWR Surveill Summ 72:1, 2023.

Global Burden of Disease Cause of Death Collaborators: Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet 392:1736, 2018.

Goldenberg JZ et al: Probiotics for the prevention of pediatric antibiotic-associated diarrhea. Cochrane Database Syst Rev 12:CD004827, 2015.

Guttman JA, Finlay BB: Subcellular alterations that lead to diarrhea during bacterial pathogenesis. Trends Microbiol 16:535, 2008.

Hyesuk S et al: Vaccines against gastroenteritis, current progress and challenges. Gut Microbes 11:1486, 2020.

Levine MM et al: Diarrhoeal disease and subsequent risk of death in infants and children residing in low-income and middle-income countries: Analysis of the GEMS case-controlled study and 12-month GEMS-1A follow-on study. Lancet Glob Health 8:e202, 2020.

Local Burden of Disease Diarrhoea Collaborators: Mapping geographical inequalities in childhood diarrhoeal morbidity and mortality in low-income and middle-income countries, 2000-17: Analysis for the Global Burden of Disease Study 2017. Lancet 395:1779, 2020.

Rodgers AP et al: Impact of enteric bacterial infections at and beyond the epithelial barrier. Nat Rev Microbiol 21:260, 2023.

Shane AL et al: Infectious Diseases Society of America. Clinical practice guidelines for the diagnosis and management of infectious diarrhea. Clin Infect Dis 65:e45, 2017.

Teh R et al: Review of the role of gastrointestinal multiplex polymerase chain reaction in the management of diarrheal illness. J Gastroenterol Hepatol 36:3286, 2021.

Tsolis RM, Baumler AJ: Gastrointestinal host-pathogen interaction in the age of microbiome research. Curr Opin Microbiol 53:78, 2020.

139

Clostridioides difficile Infection, Including Pseudomembranous Colitis

Dale N. Gerding, Stuart Johnson

DEFINITION

Clostridioides difficile infection (CDI) is a unique colonic disease that is acquired most commonly in association with antimicrobial use and the consequent disruption of the normal colonic microbiota. The most commonly diagnosed diarrheal illness acquired in the hospital, CDI results from the ingestion of spores of *C. difficile* that vegetate, multiply, and secrete toxins, causing diarrhea and, in the most severe cases, pseudomembranous colitis (PMC).

ETIOLOGY AND EPIDEMIOLOGY

C. difficile is an obligately anaerobic, gram-positive, spore-forming bacillus whose spores are found widely in nature, particularly in the environment of hospitals and chronic-care facilities. CDI occurs frequently in hospitals and nursing homes (or shortly after discharge from these facilities) where the level of antimicrobial use is high and the environment is contaminated by *C. difficile* spores.

Clindamycin, ampicillin, and cephalosporins were the first antibiotics associated with CDI. The second- and third-generation cephalosporins, particularly cefotaxime, ceftriaxone, cefuroxime, and ceftazidime, are agents frequently responsible for this condition, and the fluoroquinolones (ciprofloxacin, levofloxacin, and moxifloxacin) are the most recent drug class to be implicated in hospital outbreaks. Penicillin/β-lactamase-inhibitor combinations such as ticarcillin/clavulanate and piperacillin/tazobactam pose significantly less risk. However, all antibiotics, including vancomycin (the agent most commonly used to treat CDI), have been found to carry a risk of subsequent CDI. A minority of cases, especially in the community, are reported in patients without documentation of prior antibiotic exposure.

C. difficile is acquired exogenously—most often in the hospital or nursing home, but also in the outpatient setting—and is carried in the stool of both symptomatic and asymptomatic patients. The rate of fecal colonization increases in proportion to length of hospital stay and is often ≥20% among adult patients hospitalized for >2 weeks; in contrast, the rate is 1–3% among community residents. CDI is the most common health care–associated infection in the United States, with an estimated 462,100 cases in 2017. Between 2011 and 2017, the total burden of CDI in the United States decreased by 24%, which was due primarily to decreases in health care–associated CDI. The estimated burden of community-associated CDI was unchanged, but it now approximates the health care–associated rate in the United States.

Asymptomatic fecal carriage of *C. difficile* in healthy neonates is very common, with repeated colonization by multiple strains in infants <1–2 years of age, but associated disease in these infants is extremely rare if it occurs at all. Spores of *C. difficile* are found on environmental surfaces (where the organism can persist for months) and on the hands of hospital personnel who fail to practice good hand hygiene. Hospital epidemics of CDI have been attributed to a single *C. difficile* strain and to multiple different strains, introduced by patients on admission, that are present simultaneously. Other identified risk factors for CDI include older age, greater severity of underlying illness, gastrointestinal surgery, use of electronic rectal thermometers, enteral tube feeding, and antacid treatment. Use of proton pump inhibitors may be a risk factor, but this risk is probably modest, and no firm data have implicated these agents in patients who are not already receiving antibiotics.

PATHOLOGY AND PATHOGENESIS

Spores of toxigenic *C. difficile* are ingested, survive gastric acidity, germinate in the small bowel, and, in the presence of a disrupted microbiota, colonize the lower intestinal tract, where they elaborate two large toxins: toxin A (an enterotoxin) and toxin B (a cytotoxin). These toxins initiate processes resulting in the disruption of epithelial-cell barrier function, diarrhea, and pseudomembrane formation. Toxin A is a potent neutrophil chemoattractant, and both toxins glucosylate the GTP-binding proteins of the Rho subfamily that regulate the actin cell cytoskeleton. Data from studies using molecular disruption of toxin genes in isogenic mutants suggest that toxin B may be the more important virulence factor, which is consistent with the well-documented occurrence of clinical disease caused by toxin A–negative strains but not by toxin B–negative strains. Disruption of the cytoskeleton results in loss of cell shape, adherence, and tight junctions, with consequent fluid leakage. A third toxin, binary toxin CDT, was previously found in only ~6% of strains but is present in all isolates of the widely recognized epidemic NAP1/BI/027 strain (see "Global Considerations," below); this toxin is related to *C. perfringens* iota toxin. Its role in the pathogenesis of CDI has not yet been defined.

The pseudomembranes of PMC are confined to the colonic mucosa and initially appear as 1- to 2-mm whitish-yellow plaques. The intervening mucosa appears unremarkable, but, as the disease progresses,

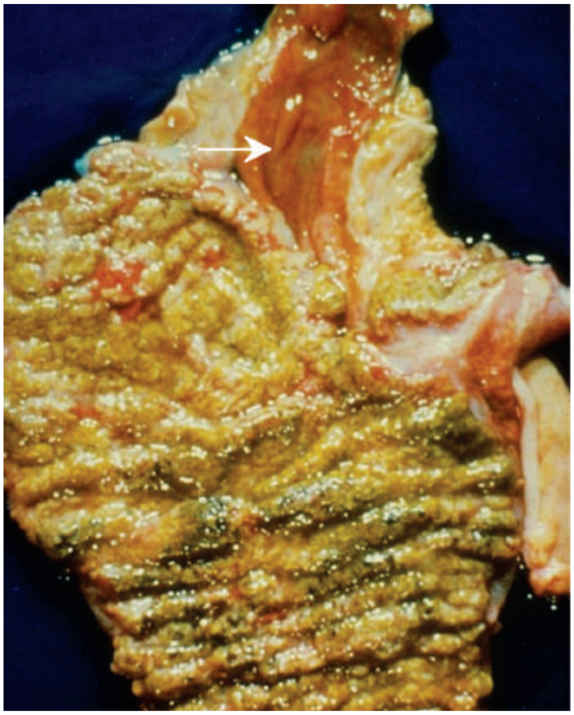

FIGURE 139-1 Autopsy specimen showing confluent pseudomembranes covering the cecum of a patient with pseudomembranous colitis. Note the sparing of the terminal ileum (*arrow*).

the pseudomembranes coalesce to form larger plaques and become confluent over the entire colon wall **(Fig. 139-1)**. The whole colon is usually involved, but 10% of patients have rectal sparing. Viewed microscopically, the pseudomembranes have a mucosal attachment point and contain necrotic leukocytes, fibrin, mucus, and cellular debris. The epithelium is eroded and necrotic in focal areas, with neutrophil infiltration of the mucosa.

Patients colonized with *C. difficile* were initially thought to be at high risk for CDI. However, four prospective studies have shown that colonized patients who have not previously had CDI actually have a decreased risk of CDI, possibly because many of these patients are colonized by nontoxigenic strains. At least three events are proposed as essential for the development of CDI **(Fig. 139-2)**. Exposure to antimicrobial agents is the first event and establishes susceptibility to CDI, most likely through disruption of the normal gastrointestinal microbiota. The second event is exposure to toxigenic *C. difficile*. Given that the majority of patients do not develop CDI after the first two events, a third event is clearly essential for its occurrence. Candidate third events include exposure to a *C. difficile* strain of particular virulence, exposure to antimicrobial agents especially likely to cause CDI, and an inadequate host immune response. The host anamnestic immune response as has been shown for serum IgG antibody response to toxin A of *C. difficile* is likely one factor in the third event that determines which patients develop diarrhea and which patients remain asymptomatic. The majority of humans probably first develop antibody to *C. difficile* toxins when colonized asymptomatically during the first year of life or after CDI in childhood. Infants are thought not to develop symptomatic CDI because they lack suitable mucosal toxin receptors that develop later in life. In adulthood, there is evidence that serum antibodies to both toxin A and B protect against recurrent CDI. Two large clinical trials in which intravenous monoclonal antibodies to toxin A and toxin B were used together and as single agents in addition to standard antibiotic therapy showed that rates of recurrent CDI were significantly lower with the combination of antibodies and with the toxin B antibody alone than with placebo plus standard therapy. Antibody to toxin A alone was ineffective.

■ GLOBAL CONSIDERATIONS

Rates and severity of CDI in the United States, Canada, and Europe increased markedly in the early 2000s. Rates in U.S. hospitals tripled between 2000 and 2005. Hospitals in Montreal, Quebec, reported rates in 2005 that were four times higher than the 1997 baseline, with directly attributable mortality of 6.9% (increased from 1.5%). An epidemic strain, variously known as toxinotype III, REA type BI, PCR ribotype 027, and pulsed-field type NAP1 (collectively designated NAP1/BI/027), likely accounted for much of the increase in incidence. Two clones of NAP1/BI/027 originated in the United States and Canada and spread to the United Kingdom, Europe, and Asia. This epidemic strain was characterized by (1) an ability to produce 16–23 times as much toxin A and toxin B as control strains in vitro, (2) the presence of binary toxin CDT, and (3) high-level resistance to all fluoroquinolones. National control policies instituted in England in 2006 resulted in a marked decline in CDI cases, and restriction of fluoroquinolones, in particular, was correlated with near elimination of fluoroquinolone-resistant strains of *C. difficile* (i.e., NAP1/BI/027) there by 2013. This epidemic strain has likewise decreased in the United States, with data from the Centers for Disease Control and Prevention showing a decrease among health care–associated isolates from 31% to 15% (and from 19% to 6% in community-associated isolates) between 2011 and 2017. New strains have been and will probably continue to be implicated in outbreaks, including a strain commonly found in food animals that also carries binary toxin and has been associated with high mortality rates in human infections (toxinotype V, ribotype 078). Currently, the most frequently isolated community-associated strain in the United States is ribotype 106 (REA group DH), which was previously found to be epidemic in the United Kingdom.

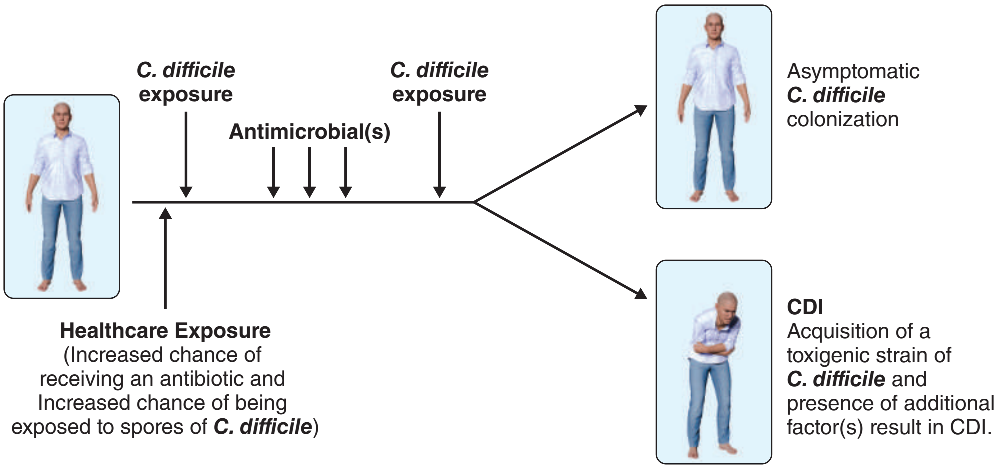

FIGURE 139-2 Pathogenesis model for *Clostridioides difficile* infection (CDI). At least three events are integral to *C. difficile* pathogenesis. Exposure to antibiotics establishes susceptibility to *C. difficile* infection. If susceptible, the patient may ingest nontoxigenic (nonpathogenic) or toxigenic strains of *C. difficile* as a second event. Acquisition of toxigenic *C. difficile* may be followed by asymptomatic colonization or CDI, depending on one or more additional factors, including the strain of *C. difficile* and an inadequate host anamnestic antibody response to *C. difficile* toxins.

CLINICAL MANIFESTATIONS

Diarrhea is the most common manifestation caused by *C. difficile*. Stools are almost never grossly bloody and range from soft and unformed to watery or mucoid in consistency, with a characteristic odor. Clinical and laboratory findings include fever in 28% of cases, abdominal pain in 22%, and leukocytosis in 50%. When adynamic ileus (which is seen on x-ray in ~20% of cases) results in cessation of stool passage, the diagnosis of CDI is frequently overlooked. A clue to the presence of unsuspected CDI in these patients is unexplained leukocytosis, with ≥15,000 white blood cells (WBCs)/μL. Such patients are at high risk for complications of CDI, particularly toxic megacolon and sepsis.

C. difficile diarrhea recurs after treatment in ~15–30% of cases and remains one of the most challenging treatment dilemmas. Recurrences may represent either relapses due to the same strain or reinfections with a new strain. Susceptibility to recurrence of clinical CDI is likely a result of continued disruption of the normal fecal microbiota caused by the antibiotic used to treat CDI.

DIAGNOSIS

The diagnosis of CDI is based on a combination of clinical criteria: (1) diarrhea (≥3 unformed stools per 24 h for ≥2 days) with no other recognized cause plus (2) detection of toxin A or B in the stool, detection of toxin-producing *C. difficile* in the stool by nucleic acid amplification testing (NAAT; e.g., polymerase chain reaction [PCR]) or by culture, or visualization of pseudomembranes in the colon. PMC is a more advanced form of CDI and is visualized at endoscopy in only ~50% of patients with diarrhea who have a positive stool culture and toxin assay for *C. difficile* **(Table 139-1)**. Endoscopy is a rapid diagnostic tool in seriously ill patients with suspected PMC and an acute abdomen, but a negative result in this examination does not rule out CDI.

Despite the array of tests available for *C. difficile* and its toxins (Table 139-1), no single test has high sensitivity, high specificity, and rapid turnaround. Most laboratory tests for toxins, including enzyme immunoassays (EIAs), lack sensitivity. NAATs (including PCR) are widely used diagnostically and are both rapid and sensitive; however, concern has been raised that PCR may detect colonization with toxigenic *C. difficile* in patients who have diarrhea for a reason other than CDI. Confirmation of the presence of toxin in the stool in addition to NAAT or glutamate dehydrogenase (GDH) positivity is recommended in the European CDI guidelines for diagnosis of CDI, and inclusion of a stool toxin test is recommended in the U.S. guidelines when there are no prior criteria for stool submission. Test algorithms that include NAAT followed by toxin EIA for NAAT+ results, and GDH plus toxin EIA arbitrated by NAAT when the two initial test results do not agree, have become widely used; however, when results of individual algorithm tests are discrepant (NAAT+ or GDH+/toxin EIA–), most patients are nonetheless treated for CDI. Empirical treatment is appropriate if stool testing is delayed. Testing of asymptomatic patients is not recommended except for epidemiologic study purposes. In particular, so-called tests of cure following treatment are not recommended because >50% of patients continue to harbor the organism and its toxin after diarrhea has ceased and test results are not predictive of recurrence of CDI. The results of such tests should not be used to restrict placement of patients in long-term care or nursing home facilities.

TREATMENT

Clostridioides difficile Infection

PRIMARY CDI

When possible, discontinuation of any ongoing antimicrobial administration is recommended as the first step in treatment of CDI. Earlier studies indicated that 15–23% of patients respond to this simple measure. However, with the advent of the NAP1/BI/027 epidemic strain and the associated rapid clinical deterioration of some patients, prompt initiation of specific CDI treatment has become the standard. General treatment guidelines include hydration and the avoidance of antiperistaltic agents and opiates, which may mask symptoms and possibly worsen disease. Nevertheless, antiperistaltic agents have been used safely with vancomycin or metronidazole treatment for mild to moderate CDI.

Oral administration of fidaxomicin was suggested as first-line treatment for CDI in the 2021 Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA) focused update guidelines on management of CDI in adults. However, because of resource availability issues, oral vancomycin remains an acceptable alternative. Oral metronidazole is recommended only for mild or moderate CDI when fidaxomicin or vancomycin is not available. IV vancomycin is ineffective for CDI. Fidaxomicin is available only for oral administration. Two large clinical trials comparing vancomycin and fidaxomicin indicated comparable clinical resolution of diarrhea in ~90% of patients, and the rate of recurrent CDI was significantly lower with fidaxomicin. The largest randomized controlled trial of vancomycin versus metronidazole showed that the vancomycin cure rate was superior to the metronidazole cure rate (81% vs 73%; $p = .034$) for all patients with CDI, regardless of severity. Although the mean time to resolution of diarrhea is 2–4 days, the response to metronidazole may be much slower. Treatment should not be deemed a failure until a drug has been given for at least 6 days. On the basis of data for shorter courses of vancomycin and the results of four large clinical trials, it is recommended that vancomycin or fidaxomicin be given for at least 10 days. Metronidazole was never approved for CDI by the

TABLE 139-1 Relative Sensitivity and Specificity of Diagnostic Tests for *Clostridioides difficile* Infection (CDI)

TYPE OF TEST	RELATIVE SENSITIVITY[a]	RELATIVE SPECIFICITY[a]	COMMENT
Stool culture for *C. difficile*	++++	+++	Most sensitive test; specificity of ++++ if the *C. difficile* isolate tests positive for toxin; turnaround time too slow for practical use
Cell culture cytotoxin test on stool	+++	++++	With clinical data, is diagnostic of CDI; highly specific but not as sensitive as stool culture; slow turnaround time
Enzyme immunoassay for toxins A and B in stool	++ to +++	+++	With clinical data, is diagnostic of CDI; rapid results, but not as sensitive as stool culture or cell culture cytotoxin test
Enzyme immunoassay for *C. difficile* common antigen in stool	+++ to ++++	+++	Detects glutamate dehydrogenase found in toxigenic and nontoxigenic strains of *C. difficile* and other stool organisms; more sensitive and less specific than enzyme immunoassay for toxins; requires confirmation with a toxin test; rapid results
Nucleic acid amplification tests for *C. difficile* toxin A or B gene in stool	++++	+++	Detects toxigenic *C. difficile* in stool; widely used in United States for clinical testing; more sensitive than enzyme immunoassay toxin testing; marked increase in CDI diagnoses when implemented
Colonoscopy or sigmoidoscopy	+	++++	Highly specific if pseudomembranes are seen; insensitive compared with other tests

[a] According to both clinical and test-based criteria.

Note: ++++, >90%; +++, 71–90%; ++, 51–70%; +, ~50%.

TABLE 139-2 Recommendations for the Treatment of *Clostridioides difficile* Infection (CDI)

CLINICAL SETTING	TREATMENT(S)	COMMENTS
Initial episode, mild to moderate	Fidaxomicin (200 mg bid × 10 d) ***or alternatively*** Oral vancomycin (125 mg qid × 10 d)	Oral metronidazole is less effective than the other options and may necessitate a longer treatment course for response. Metronidazole (500 mg tid × 10–14 d) is recommended only if vancomycin or fidaxomicin is not readily accessible and for mild to moderate disease only.
Initial episode, severe	Oral vancomycin (125 mg qid × 10 d) ***or alternatively*** Fidaxomicin (200 mg bid × 10 d)	Indicators of severe disease may include leukocytosis (≥15,000 white blood cells/μL) and a creatinine level ≥1.5 mg/dL.
Initial episode, fulminant	Vancomycin (500 mg PO or via nasogastric tube) plus metronidazole (500 mg IV q8h) ***plus consider*** Rectal instillation of vancomycin (500 mg in 100 mL of normal saline as a retention enema q6–8h)	Fulminant CDI is defined as severe CDI with the addition of hypotension, shock, ileus, or toxic megacolon. The duration of treatment may need to be >2 weeks and is dictated by response.
First recurrence	Fidaxomicin (200 mg bid × 10 d) ***or*** Oral vancomycin (125 mg qid × 10 d) ***or*** Oral vancomycin followed by a taper-and-pulse regimen[a]	Treatment for the initial episode may be considered when choosing treatment for the first recurrence.
Multiple recurrences	Oral vancomycin treatment followed by a taper-and-pulse regimen ***or*** Fidaxomicin (200 mg bid × 10 d or 200 mg bid × 5 d followed by every other day × 20 d) ***or*** Vancomycin (125 mg qid × 10 d), then stop vancomycin and start rifaximin (400 mg bid × 2 weeks) ***or*** Fecal microbiota replacement therapy (FMRT)	It is recommended that FMRT by fecal microbiota (Rebyota, live-jslm) given by enema or fecal microbiota spores (Vowst, live-brpk) given orally be considered only after appropriate antibiotic treatment for recurrent CDI.
Patients at high risk of recurrent CDI who are receiving vancomycin, fidaxomicin, or metronidazole	Bezlotoxumab 10 mg/kg given IV	Bezlotoxumab is adjuvant therapy (in addition to and during antibiotic treatment) for patients at high risk for recurrent CDI. Risk factors include age >65 years, immunocompromised host, severe CDI on presentation, and prior episode of CDI in the past 6 months.

[a] A typical taper-and-pulse vancomycin regimen following a 10-day treatment course includes: 125 mg bid × 1 week, then daily × 1 week, then q2–3d for 2–8 weeks.

U.S. Food and Drug Administration (FDA), and its use for CDI treatment declined markedly after publication of the 2017 IDSA/SHEA CDI guidelines. It is important to initiate treatment with oral vancomycin or fidaxomicin for patients who appear seriously ill, particularly if they have a high WBC count (>15,000/μL) or creatinine level (≥1.5 mg/dL) **(Table 139-2)**. Small randomized trials of nitazoxanide, bacitracin, rifaximin, and fusidic acid for treatment of CDI have been conducted. These drugs have not been extensively studied, shown to be superior, or approved by the FDA for CDI, but they provide potential alternatives to vancomycin and fidaxomicin.

RECURRENT CDI

Overall, ~15–30% of successfully treated patients experience recurrences of CDI following treatment. CDI recurrence is significantly lower in patients treated with fidaxomicin than in those treated with vancomycin. Vancomycin and metronidazole have comparable recurrence rates, and metronidazole is not recommended for treatment of recurrent CDI. Patients who have a first recurrence of CDI have an even higher rate of second recurrence. Fidaxomicin is superior to vancomycin in reducing further recurrences in patients who have had one CDI recurrence (Table 139-2). Recurrent disease, once thought to be relatively mild, has now been documented to pose a significant (11%) risk of serious complications (shock, megacolon, perforation, colectomy, or death within 30 days). There is no standard treatment for multiple recurrences, but the use of vancomycin in a tapering and pulsed dosing regimen every other day for 2–8 weeks has been used for years as a practical approach to treating these patients, and recent data suggest it is still effective. Other recommended treatment options for patients with multiple CDI recurrences include fidaxomicin in standard or extended/pulsed dosing regimens, vancomycin followed by rifaximin, or fecal microbiota transplantation (FMT) via nasoduodenal tube, colonoscope, enema, or oral capsules (Table 139-2). FMT has been widely used over the past decade, and the availability of stool banks and oral capsule formulations made this approach more practical. Recently, the FDA has approved two microbiota replacement therapies and discontinued enforcement discretion for centralized donor stool banks. A fecal microbiota suspension (Rebyota, live-jslm) and a suspension of live fecal microbiota spores for oral delivery (Vowst, live-brpk) are now approved for patients with recurrent CDI to prevent further recurrence. Both of these products are adjunctive treatments that are indicated for patients who have completed antibiotic treatment for CDI.

In addition to antibacterial therapies, another adjunctive treatment is now available for patients who are receiving standard-of-care antibacterial agents and who are at high risk for recurrent CDI (rCDI). Bezlotoxumab, a monoclonal antibody directed against *C. difficile* toxin B, has been shown to reduce the risk of rCDI by an absolute rate of ~10% when administered to patients currently receiving vancomycin, fidaxomicin, or metronidazole. Risk factors for rCDI in the clinical trials included age >65 years, immunocompromise, severe CDI on presentation, and prior episode of CDI in the past 6 months.

SEVERE COMPLICATED OR FULMINANT CDI

Fulminant (rapidly progressive and severe) CDI presents the most difficult treatment challenge. Patients with fulminant disease often do not have diarrhea, and their illness mimics an acute surgical abdomen. Sepsis (hypotension, fever, tachycardia, leukocytosis) may result from fulminant CDI. An acute abdomen (with or without toxic megacolon) may include signs of obstruction, ileus, colon-wall thickening and ascites on abdominal CT, and peripheral-blood leukocytosis (≥20,000 WBCs/μL). With or without diarrhea, the differential diagnosis of an acute abdomen, sepsis, or toxic megacolon should include CDI if the patient has received antibiotics in the past 2 months. Cautious sigmoidoscopy or colonoscopy to visualize PMC and an abdominal CT examination are the best diagnostic tests in patients without diarrhea.

Medical management of fulminant CDI is suboptimal because of the difficulty of delivering oral fidaxomicin, metronidazole, or vancomycin to the colon in the presence of ileus (Table 139-2). The combination of vancomycin (given orally or via nasogastric tube and by retention enema) plus IV metronidazole has been used with some success in uncontrolled studies, as has IV tigecycline in small-scale uncontrolled studies. Surgical colectomy may be lifesaving if there is no response to medical management. If possible, colectomy should be performed before the serum lactate level reaches 5 mmol/L. However, mortality and morbidity associated with colectomy may be reduced by performing instead a laparoscopic ileostomy followed by colon lavage with polyethylene glycol and vancomycin infusion into the colon via the ileostomy.

■ PROGNOSIS

The mortality rate attributed to CDI, previously found to be 0.6–3.5%, has reached 6.9% in recent outbreaks and is progressively higher with increasing age. Most patients recover, but recurrences are common.

■ PREVENTION AND CONTROL

Strategies for the prevention of CDI are of two types: those aimed at preventing transmission of the organism to the patient and those aimed at reducing the risk of CDI if the organism is transmitted. Transmission of *C. difficile* in clinical practice has been prevented by gloving of personnel, elimination of the use of contaminated electronic thermometers, and use of hypochlorite (bleach) solution for environmental decontamination of patients' rooms. Hand hygiene is critical; hand washing is recommended in CDI outbreaks because alcohol hand gels are not sporicidal. CDI outbreaks have been best controlled by restricting the use of specific antibiotics, such as clindamycin, second- and third-generation cephalosporins, and fluoroquinolones. Outbreaks of CDI due to antibiotic-resistant strains have resolved promptly when specific antibiotic use is restricted. Future primary prevention strategies include use of monoclonal antibodies, as well as biotherapeutics with live organisms to prevent colonization. Vaccines have to date been unsuccessful.

■ FURTHER READING

Dingle KE et al: Effects of control interventions on *Clostridium difficile* infection in England: An observational study. Lancet Infect Dis 17:411, 2017.

Feuerstadt P et al: SER-109, an oral microbiome therapy for recurrent *Clostridioides difficile* infection. N Engl J Med 386:220, 2022.

Guh AY et al: Trends in U.S. burden of *Clostridioides difficile* infection and outcomes. N Engl J Med 382:1320, 2020.

Guh AY et al: Potential underreporting of treated patients using a *Clostridioides difficile* testing algorithm that screens with a nucleic acid amplification test. Infect Cont Hosp Epidemiol 2024. Online ahead of print.

He M et al: Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. Nat Genet 45:109, 2013.

Johnson S et al: Vancomycin, metronidazole, or tolevamer for *Clostridium difficile* infection: Results from two multinational, randomized, controlled trials. Clin Infect Dis 59:345, 2014.

Johnson S et al: Clinical practice guideline by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA): 2021 Focused update guidelines on management of *Clostridioides difficile* infection in adults. Clin Infect Dis 73:755, 2021.

Khanna S et al: Efficacy and safety of RBX2660 in PUNCH CD3, a phase III, randomized, double-blind, placebo-controlled trial with a Bayesian primary analysis for the prevention of recurrent *Clostridioides difficile* infection. Drugs 82:1527, 2022.

Kociolek LK et al: Natural *Clostridioides difficile* toxin immunization in colonized infants. Clin Infect Dis 70:2095, 2020.

Louie TJ et al: Fidaxomicin versus vancomycin for *Clostridium difficile* infection. N Engl J Med 364:422, 2011.

McDonald LC et al: Clinical practice guidelines for *Clostridium difficile* infection in adults and children: 2017 update by the Infectious Diseases Society of America (IDSA) and the Society for Healthcare Epidemiology of America (SHEA). Clin Infect Dis 66:987, 2018.

Polage CR et al: Overdiagnosis of *Clostridium difficile* infection in the molecular test era. JAMA Intern Med 175:1792, 2015.

See I et al: NAP1 strain type predicts outcomes from *Clostridium difficile* infection. Clin Infect Dis 58:1394, 2014.

Wilcox MH et al: Bezlotoxumab for prevention of *Clostridium difficile* infection recurrence. N Engl J Med 376:305, 2017.

140 Urinary Tract Infections: Cystitis, Prostatitis, and Pyelonephritis

Barbara W. Trautner, Prathit A. Kulkarni,
Nicolás W. Cortés-Penfield, Kalpana Gupta

In the preantibiotic era, urinary tract infection (UTI) caused significant morbidity. Hippocrates described acute cystitis as sometimes lasting for a year before either resolving or worsening to involve the kidneys. Early twentieth-century chemotherapeutics for UTI were ineffective, and persistent infection was common. Nitrofurantoin, which became available during the 1950s, was the first tolerable and effective agent for UTI treatment. Today, UTI remains common, being one of the leading reasons for which antibiotics are used within and outside the hospital. Patients typically favorably and quickly to appropriately chosen modern antimicrobials.

Since the most common manifestation of UTI is acute cystitis, which is far more prevalent among women than among men, most clinical research on UTI has involved healthy young adult women recruited from college campuses or large U.S. health maintenance organizations. Therefore, clinicians must understand that UTI remains relatively understudied in other patient populations, and they must use expert judgment to decide when and to what extent UTI clinical trial data and treatment guidelines can be generalized to other groups of patients.

■ DEFINITIONS

The term *urinary tract infection* refers to an infection somewhere along the urinary tract that produces clinical symptoms. Infection can be caused either by bladder invasion by a new urinary pathogen or by a shift in the existing urinary bacteria's dynamic equilibrium with the host; both causes result in local tissue injury and inflammation. Thus, while UTI encompasses a broad range of clinical presentations, from acute simple cystitis to bacteremic emphysematous pyelonephritis, importantly, the presence alone of bacteria in the urinary tract in a patient *without* related symptoms does *not* constitute UTI. The presence of bacteria in the urine without associated symptoms is instead termed *asymptomatic bacteriuria* (ASB).

This distinction between UTI and ASB has major clinical implications. Both UTI and ASB connote the presence of bacteria in the urinary tract and are usually accompanied by white blood cells in the urine (termed *pyuria*). However, while antibiotic treatment of UTI is almost always indicated to relieve the symptoms attributable to urinary bacterial infection and to prevent progression of the infection, ASB typically poses no threat for most patients and usually does not require treatment. Limited data suggest inappropriate treatment of ASB might in fact *increase* risk for future UTI, perhaps by facilitating bladder colonization with more virulent and potentially more resistant pathogens. Two exceptions to the general rule that ASB should not be treated are patients who are pregnant or who are about to undergo certain urologic procedures. In pregnant patients, untreated ASB is associated with pyelonephritis, preterm delivery, and low birth weight, which justifies use of routine ASB screening and treatment. In patients undergoing procedures that cause urologic mucosal injury (e.g., lithotripsy or transurethral resection of the prostate), bacterial translocation from urine to blood can occur; therefore, screening for, and periprocedural treatment of, ASB is recommended in this setting.

In this chapter, the term *urinary tract infection* denotes symptomatic disease; *cystitis*, symptomatic infection of the bladder; *prostatitis*, symptomatic infection of the prostate; and *pyelonephritis*, symptomatic infection of the kidneys. *Uncomplicated urinary tract infection* refers to an infection confined to the bladder in a woman or man without a urinary catheter, whereas *complicated urinary tract infection* refers to infection that extends beyond the bladder (e.g., prostatitis, pyelonephritis, bacteremia, or UTI in the setting of intermittent or indwelling urinary catheterization, termed *catheter-associated UTI* [CAUTI]). *Recurrent urinary tract infection* is not necessarily complicated; individual episodes can be uncomplicated and treated as such. This approach to UTI categorization differs from the classic approach, in which UTI in men and other populations perceived to be at higher risk for poor outcome (e.g., patients with diabetes, anatomic abnormalities of the urinary tract, or immunocompromise) was automatically considered complicated. The new updated approach more closely reflects actual clinical practice. For the frontline clinician, the key considerations in diagnostic workup and therapy for UTI include whether the patient is stable for outpatient management, whether a source of recalcitrant and/or recurrent infection (e.g., obstructing renal calculi) needs to be identified, and whether the prescribed antimicrobial agents must achieve adequate levels in the blood and renal tissue in addition to the urine.

Prostatitis, usually categorized as either acute bacterial prostatitis (ABP) or chronic bacterial prostatitis (CBP), can complicate cystitis or arise hematogenously (e.g., as a metastatic focus of *Staphylococcus aureus* bacteremia) in men and transgender women. While prostatitis is certainly an infection of the genitourinary tract, it is largely unrepresented in UTI clinical trial data, and earlier ABP literature suggests that prolonged therapy with antimicrobials penetrating the prostate might be required. Therefore, clinical practice guidelines for complicated UTI cannot be assumed to be generalizable to patients with prostatitis. *Urethritis* is occasionally caused by *Escherichia coli* and other urinary pathogens but is predominantly due to *Neisseria gonorrhea*, *Chlamydia trachomatis*, and *Mycoplasma genitalium*. Hence, symptoms such as urethral pain, pruritus, or discharge should prompt evaluation for sexually transmitted infections (STIs), which are covered in **Chap. 141**.

■ EPIDEMIOLOGY AND RISK FACTORS

During the neonatal period, the incidence of UTI is slightly higher among males than among females because male infants more commonly have congenital urinary tract anomalies. After 50 years of age, obstruction from prostatic hypertrophy becomes common in men, and incidence of UTI is almost as high among men as among women. Between 1 year and ~50 years of age, however, UTI and recurrent UTI are predominantly conditions of females, with most true male UTI occurring in the context of urinary catheterization or anatomic abnormalities. The prevalence of ASB is ~5% among women between ages 20 and 40 and might be as high as 40–50% among elderly women and men. Up to 80% of women develop at least one UTI during their lifetime, predominantly acute simple cystitis. The lifetime prevalence of UTI in men in the United States is estimated to be 14%.

About 20–30% of women who have had one episode of UTI will have recurrent episodes. Early recurrence (within 2 weeks) is usually regarded as relapse rather than reinfection and might indicate the need to evaluate the patient for a sequestered focus of ongoing infection. The rate of UTI recurrence ranges from 0.3 to 7.6 infections per patient per year, with an average of 2.6 reinfections per year. It is not uncommon for multiple recurrences to follow an initial infection, resulting in clustering of episodes. Clustering may be related to the presence of a new risk factor, sloughing of the protective outer bladder epithelial layer in response to bacterial attachment during acute cystitis, or possibly antibiotic-related alteration of the normal bacterial flora. The likelihood of recurrence decreases with increasing time since the last infection. Risk factors for recurrent UTI include frequent sexual intercourse, use of spermicide, a new sexual partner, first UTI before 15 years of age, and maternal history of UTI; of these, frequent sexual intercourse and spermicide use are the most consistently documented risk factors.

Some component of UTI susceptibility is likely hereditary, as family and maternal history of UTI has been identified as a risk factor for cystitis, pyelonephritis, and recurrent UTI. In addition, colonization with specific bacterial strains predisposed to cause infection might also be an important factor, as suggested by studies indicating that new sexual partners are a risk factor for UTI in women and that uropathogenic *E. coli* are shared between partners via vaginal and oral sex. Beyond this, risk factors for UTI can be grouped broadly into three categories: (1) factors that impede the normal outflow of urine, (2) factors that introduce new bacteria into the genitourinary tract or serve as a nidus of infection, and (3) factors that disrupt or impair local mucosal defenses.

The flushing action of normal urination is a key mechanism by which the host controls urinary bacterial populations and mitigates bacterial ascent from the urethra to the bladder and kidneys. Anatomic factors that result in urinary obstruction or retention have been associated with cystitis, pyelonephritis, and recurrent UTI. These risk factors include congenital urinary tract abnormalities, cystoceles, pelvic-organ prolapse, bladder dysfunction leading to overflow urinary incontinence, and benign prostatic hypertrophy. Importantly, not all men with UTI have detectable urinary tract abnormalities; this point is particularly relevant for men ≤45 years of age.

Factors that promote bacterial entry into the urinary tract predispose to UTI. Sexual activity is likely the most common among these and is temporally associated with cystitis, with an increased relative risk in one study ranging from 1.4 with weekly intercourse to 4.8 with intercourse five times weekly. For men, insertive rectal intercourse and lack of circumcision are associated with increased UTI risk, the latter likely because *E. coli* is tends to colonize the glans and prepuce and subsequently migrate into the urinary tract. For men and women, urinary instrumentation (e.g., intermittent or chronic catheterization or existence of a draining nephrostomy tube) dramatically increases bacterial entry into the urinary tract and risk of UTI. In addition, foreign bodies in the urinary tract (e.g., stents and kidney stones) also can become colonized with bacteria and serve as niduses for recurrent UTI.

Factors that disrupt or impair local mucosal defenses, usually either by irritating the urogenital mucosa or by disrupting the normal female urogenital microbiome (i.e., symbiotic lactobacilli), predispose to UTI. Such factors include recent use of diaphragms or spermicide and menopause. Women with diabetes have two- to threefold higher rates of ASB and UTI than women without diabetes, and this risk increases further with a longer duration of having had diabetes and with the need for insulin rather than oral medications. The specific mechanisms for this risk are not clear, although impaired cytokine secretion or diabetes complications leading to bladder dysfunction likely contribute. Whether glycosuria is major contributor to UTI risk remains controversial. The U.S. Food and Drug Administration has issued a drug-safety warning for the observed association between use sodium–glucose cotransporter 2 (SGLT-2) inhibitors for treatment of diabetes and UTI. The postulated mechanism is that these drugs increase excretion of glucose in the urine.

■ ETIOLOGY

E. coli is the predominant pathogen across the spectra of UTI clinical syndromes and patient populations. In cases of acute cystitis in the United States, *E. coli* accounts for 75–90% of isolates, *Staphylococcus saprophyticus* for 5–15% (with more frequent isolation from younger women), and other Enterobacterales species (i.e., *Klebsiella*, *Proteus*, *Enterobacter*, and *Citrobacter*) for the majority of the remainder. The microbiology of pyelonephritis is similar to that of acute cystitis, as would be expected since pyelonephritis usually develops as ascending infection from the bladder. The microbiology of CAUTI is more diverse, but *E. coli* remains the predominant organism. Other important pathogens in CAUTI are *Pseudomonas aeruginosa*, enterococci, *Staphylococcus aureus*, and *Candida*. Genetic sequencing of the bladder microbiome has consistently demonstrated that more bacterial species are present than can be identified by routine culture methods, in both symptomatic and asymptomatic states. The clinical significance of these noncultivable organisms is unknown, but such work demonstrates that the healthy bladder is often not a sterile site.

Antimicrobial resistance among Enterobacterales species that cause UTI is increasing. One recent surveillance study of isolates from the United States demonstrated resistance rates of >20% to trimethoprim-sulfamethoxazole (TMP-SMX), fluroquinolones, and nitrofurantoin among ambulatory patients in many regions of the country. In addition, approximately 6% of isolates were resistant to three or more antibiotic classes, and approximately 9% of isolates had production of extended-spectrum β-lactamase. Surveillance studies conducted in South America and Europe have yielded similar findings. The increased prevalence of multidrug-resistant uropathogens has left few oral options for therapy in some cases and no single agent or agents that can confidently be recommended for empirical ambulatory treatment of UTI without regard to geographic region. Since resistance rates vary in different areas, local antibiogram data can help inform which drugs should be used as preferred empirical therapy in UTI, with the caveat that organisms are identified only in cases in which urine is sent for culture—typically, when complicated or recurrent UTI is suspected—and so may overestimate antimicrobial resistance in healthy ambulatory patients with a primary episode of UTI.

PATHOGENESIS

The urinary tract can be viewed as an anatomic unit linked by a continuous column of urine extending from the kidneys to the urethra. In the majority of UTIs, bacteria establish infection by ascending from the urethra to the bladder. Continuing ascent up the ureter to the kidney is the pathway for most renal parenchymal infections. The other main route of entry into the genitourinary tract is hematogenous in cases where bacteremia precedes UTI. Importantly, introduction of bacteria into the bladder does not inevitably lead to sustained and symptomatic infection. The interplay of host, pathogen, and environmental factors determines whether tissue invasion and symptomatic infection will ensue **(Fig. 140-1)**. For example, bacteria often enter the bladder after sexual intercourse, but normal voiding and innate host defense mechanisms in the bladder eliminate these organisms. Any foreign body in the urinary tract, such as a urinary catheter or stone, provides an inert surface for bacterial colonization. Additionally, abnormal micturition and/or significant residual urine volume promotes infection. In the simplest terms, anything that increases the likelihood of bacteria entering the bladder and staying there increases risk of developing UTI. Intracellular bacterial communities of infecting organisms within the bladder epithelium have been demonstrated in animal models of UTI and in exfoliated human urothelial cells, but the clinical impact of this phenomenon in humans is not yet clear.

Bacteria can also gain access to the urinary tract through the bloodstream. However, hematogenous spread only accounts for less than 2% of documented UTIs and usually results from bacteremia caused by relatively virulent organisms, such as *S. aureus* and *Salmonella*. The isolation of either of these pathogens from a urine culture warrants consideration of concomitant bacteremia. Hematogenous infections can also produce focal renal abscesses or pyelonephritis.

The pathogenesis of candiduria in particular is distinct because the hematogenous route is more common in that situation. The presence of *Candida* in the urine of a non-instrumented immunocompetent patient implies either genital contamination or potentially widespread visceral dissemination. By contrast, candiduria is common in catheterized patients, particularly following antimicrobial therapy for CAUTI or inappropriate treatment of ASB.

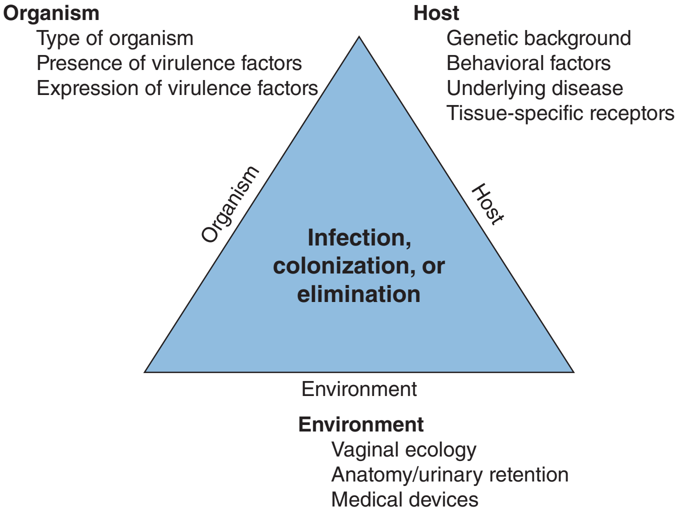
Organism
Type of organism
Presence of virulence factors
Expression of virulence factors
Host
Genetic background
Behavioral factors
Underlying disease
Tissue-specific receptors
Organism
Host
Infection, colonization, or elimination
Environment
Environment
Vaginal ecology
Anatomy/urinary retention
Medical devices

FIGURE 140-1 Pathogenesis of urinary tract infection. The relationship among specific host, pathogen, and environmental factors determines the clinical outcome.

Environmental Factors • **VAGINAL ECOLOGY** Vaginal ecology is an important environmental factor affecting risk of UTI in women. Colonization of the vaginal introitus and periurethral area with organisms from the intestinal flora (usually *E. coli*) is the critical initial step in the pathogenesis of UTI. Sexual intercourse is associated with an increased risk of vaginal colonization with *E. coli* and thereby increases risk of UTI. Nonoxynol-9 in spermicide is toxic to normal vaginal lactobacilli and thus is likewise associated with an increased risk of vaginal colonization and bacteriuria due to *E. coli*. In postmenopausal women, the previously predominant vaginal lactobacilli are replaced with colonizing Gram-negative bacteria because of vaginal atrophy. Topical estrogens have been demonstrated to reduce frequency of recurrent UTIs in postmenopausal women without altering systemic hormone levels. Given the side effects of systemic hormone replacement, oral estrogens should not be used to prevent UTI.

ANATOMIC AND FUNCTIONAL ABNORMALITIES Any condition that permits urinary stasis or obstruction predisposes the patient to development of UTI. Foreign bodies such as stones or urinary catheters provide an inert surface for bacterial colonization and formation of a persistent biofilm. Thus, vesicoureteral reflux, ureteral obstruction secondary to prostatic hypertrophy, neurogenic bladder, and urinary-diversion surgery all create an environment that can lead to UTI. Inhibition of ureteral peristalsis and decreased ureteral tone leading to vesicoureteral reflux are important in the pathogenesis of pyelonephritis in pregnant patients. Anatomic factors—in particular, the length of the urethra—are considered to be the primary reason why UTI is predominantly an illness of young women rather than of young men.

Host Factors The genetic background of the host influences the individual's susceptibility to recurrent UTI. In women, familial disposition to UTI is well documented. Additionally, women with recurrent UTI are more likely to have had their first UTI before the age of 15 years and to have a maternal history of UTI. A component of the underlying pathogenesis of this familial predisposition to recurrent UTI might be persistent vaginal colonization with *E. coli*, even during asymptomatic periods. Vaginal and periurethral mucosal cells from women with recurrent UTI bind uropathogenic bacteria threefold more than do mucosal cells from women without recurrent infection. This difference appears to be at least partly mediated by not expressing certain blood group antigens, facilitating adherence of *E. coli* to the urothelium. Mutations in host innate immune response genes (e.g., those coding for Toll-like receptors and the interleukin 8 receptor) also have been linked to recurrent UTI and pyelonephritis.

Microbial Factors Strains of *E. coli* that cause invasive symptomatic infection of the urinary tract in otherwise normal hosts often possess and express genetic virulence factors, including bacterial-surface adhesins that mediate microbial binding to specific receptors on the surface of uroepithelial cells. The best-studied adhesins are the P fimbriae, hairlike protein structures that interact with a specific receptor on renal epithelial cells. (The letter *P* denotes the ability of these fimbriae to bind to blood-group antigen P, which contains a D-galactose-D-galactose residue.) P fimbriae are important in the pathogenesis of pyelonephritis and subsequent bloodstream invasion from the kidney. Another adhesin is the type 1 pilus (fimbria), which all *E. coli* strains possess but which is not always expressed. Type 1 pili are thought to play a key role in initiating *E. coli* bladder infection; they mediate binding to mannose on the luminal surface of bladder uroepithelial cells. Toxins, metal (iron)-acquisition systems, biofilm formation, and bacterial capsules also can contribute to the ability of pathogenic *E. coli* to thrive in the bladder.

APPROACH TO THE PATIENT

Clinical Syndromes and Diagnostic Approaches

The key questions to be addressed when UTI is suspected are:

- Does the patient actually have UTI (rather than ASB or urinary symptoms not related to infection)?

- If the patient does have UTI, is it confined to the bladder (cystitis), or is there evidence that infection might be present beyond the bladder?
- If the diagnosis is uncertain, would it be safe to delay empirical antimicrobial therapy while attempting to achieve more diagnostic certainty?

The answers to these immediate clinical questions will shape the diagnostic and therapeutic approach.

ASYMPTOMATIC BACTERIURIA

A diagnosis of ASB should be strongly suspected when a patient with a positive urine culture does not have either localizing urinary symptoms or systemic symptoms that are unexplained. The clinical presentation is typically bacteriuria that is detected incidentally when a patient undergoes a screening urine culture. Accordingly, diagnostic stewardship (i.e., not sending urine cultures in patients who lack clinical signs and symptoms that suggest UTI except when screening patients who are pregnant or about to undergo urologic procedures) is the key intervention in reducing misdiagnosis of UTI and inappropriate treatment of ASB.

In the setting of pyuria (presence of white blood cells in the urine) and bacteriuria in a patient without symptoms localizing to the urinary tract, systemic symptoms and signs such as fever, altered mental status, and leukocytosis could be features of UTI but do not on their own merit a diagnosis of UTI unless other potential etiologic causes for such findings have been adequately explored and ruled out. Critically, the degree of pyuria in an asymptomatic patient should not influence the diagnosis of ASB versus true UTI. Very high levels of pyuria are common in urinary stasis, such as in patients with end-stage renal disease or obstructed urinary catheters.

In practice, one of the most challenging clinical scenarios to differentiate ASB from true UTI occurs when an older patient presents with isolated altered mentation, often superimposed on chronic cognitive deficits and potentially with little prior documentation. Accurately diagnosing the cause of such a patient's acute encephalopathy is challenging both because the patient might not be able to reliably report localizing urinary symptoms and because true UTI is an important diagnostic consideration on a long list of possible causes of the patient's symptoms. In such a patient, the treating clinician should first establish whether there is evidence of sepsis. If not, the next consideration is whether empirical therapy for potentially serious bacterial infection can be safely withheld while the diagnostic process is ongoing and symptomatic treatment for other noninfectious causes of delirium is given. The clinician should seek out collateral history, such as by calling the patient's caregivers if none are present at the bedside, to elicit information about what occurred prior to the patient's development of acute encephalopathy. A history of prior recurrent UTIs is might not be accurate and instead reflects a history of persistent ASB. The clinician must be vigilant to avoid premature diagnostic closure. Specifically, the clinician should rule out other potential causes of delirium, particularly those whose treatment is time-sensitive or that may worsen with if the diagnosis is not considered and addressed.

CYSTITIS

The typical symptoms of cystitis are dysuria, urinary frequency, and urinary urgency. Other common symptoms include suprapubic discomfort or tenderness and new nocturia, hesitancy, or gross hematuria. Fever, rigors, and unilateral back or flank pain are all inconsistent with uncomplicated cystitis and should provoke investigation for infection beyond the bladder (i.e., involving the kidneys, prostate, or bloodstream).

Cystitis in women can be treated on the basis of history alone. However, if the symptoms are not specific or if a reliable history cannot be obtained, then a urine dipstick (or, if rapidly available, urine microscopy) should be performed. In the scenario of a woman with symptoms potentially suggestive of cystitis, a negative dipstick result does not fully rule out UTI, and appropriate next steps would include formal urine microscopy with reflex to culture, possibly a pelvic examination or STI testing, and close clinical follow-up. In pregnant patients, in patients suspected to have a resistant organism, or in cases of recurrent UTI, a urine culture is specifically warranted to guide appropriate therapy. Urine dipstick and/or urine microscopy alone are not sufficient.

The signs and symptoms of cystitis in men are similar to those in women, but this diagnosis should be approached cautiously, as cystitis in men is less common, except in the setting of urinary catheterization/instrumentation or obstructive uropathy due to prostatic hypertrophy. Urine cultures should be obtained for all cases of male UTI, as the documentation of bacteriuria can differentiate the less common syndromes of acute and chronic bacterial prostatitis from the more common entity of chronic pelvic pain syndrome, which is not associated with bacteriuria and not improved by antibiotics.

PROSTATITIS

Prostatitis includes both infectious and noninfectious abnormalities of the prostate gland. Infections can be acute or chronic, are almost always bacterial in nature, and are far less common than the noninfectious entity *chronic pelvic pain syndrome* (formerly known as chronic prostatitis). ABP presents as dysuria, frequency, and pain in the prostatic pelvic or perineal area. It is typically a severe illness that can be accompanied by fever, rigors, and/or bladder outlet obstruction. CBP presents more insidiously as recurrent episodes of cystitis, sometimes with associated pelvic and perineal pain. Systemic signs of infection are typically absent, and many patients are evaluated for CBP in the ambulatory setting. Males with recurrent cystitis should be evaluated for a prostatic focus as well as for urinary retention.

Men with febrile UTI can have an elevated serum level of prostate-specific antigen and an enlarged prostate and enlarged seminal vesicles on ultrasound—findings suggestive of prostate involvement. In a study of 85 men with febrile UTI, symptoms of urinary retention, early recurrence of UTI, hematuria at follow-up, and voiding difficulties were predictive of surgically correctable disorders. Men with none of these symptoms had normal upper and lower urinary tracts on urologic workup. In general, men with febrile UTI, including at the first episode, should have imaging performed (CT or ultrasound) if it has not been performed previously. The purpose of such investigation is specifically to evaluate for urinary tract abnormalities or obstruction. If the diagnosis of UTI is unclear or if UTI is recurrent, referral for urologic consultation is appropriate.

PYELONEPHRITIS

Fever is one of the main features distinguishing cystitis from pyelonephritis, and while "febrile UTI" is sometimes used as a diagnostic category in UTI research, practically speaking, many of these patients have pyelonephritis or other infection that has extended beyond the bladder (hence the inclusion of fever as a criterion for complicated UTI).

Mild pyelonephritis can present with low-grade fever with or without lower back or costovertebral angle pain, whereas severe pyelonephritis often manifests with high fever, rigors, nausea, vomiting, and flank pain. Symptoms are generally acute in onset and may or may not be preceded by symptomatic cystitis. Obstructive uropathy related to acute papillary necrosis (i.e., when the sloughed papillae obstruct the ureter) can complicate pyelonephritis. Acute papillary necrosis can also occur in sickle cell disease and analgesic nephropathy. In the rare cases of bilateral papillary necrosis, a rapid rise in the serum creatinine level might be the first indication of the condition. *Emphysematous pyelonephritis* is a particularly severe form of pyelonephritis associated with production of gas in the renal and perinephric tissues. It occurs almost exclusively in patients with poorly controlled diabetes **(Fig. 140-2)**. *Xanthogranulomatous pyelonephritis* occurs when chronic urinary obstruction (often by staghorn calculi), together with chronic infection, leads to suppurative destruction of renal tissue **(Fig. 140-3)**. On pathologic

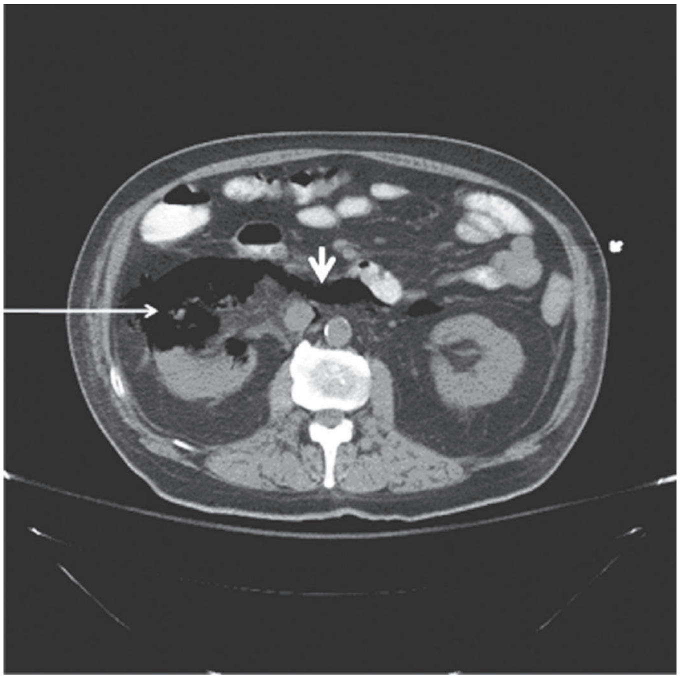

FIGURE 140-2 Emphysematous pyelonephritis. Infection of the right kidney of a diabetic man by *Escherichia coli*, a gas-forming, facultative anaerobic uropathogen, has led to destruction of the renal parenchyma (*arrow*) and tracking of gas through the retroperitoneal space (*arrowhead*).

examination, the residual renal tissue frequently has a yellow coloration, with infiltration by lipid-laden macrophages (hence the term *xanthogranulomatous*). Pyelonephritis can also be complicated by intraparenchymal abscess formation; this development should be suspected when a patient has continued fever and/or bacteremia despite antibacterial therapy.

CATHETER-ASSOCIATED URINARY TRACT INFECTION (CAUTI)

Clinically, CAUTI can be defined by the presence of symptoms along with presence of bacteriuria and pyuria. CAUTI may cause localizing urinary symptoms or otherwise unexplained systemic manifestations, such as fever. The accepted threshold for bacteriuria to meet the definition of CAUTI is $\geq 10^3$ colony-forming units per milliliter of urine.

The central diagnostic difficulty of CAUTI is that, because catheters provide a conduit for bacteria to enter the bladder, bacteriuria is inevitable with urinary catheterization. The typical signs and symptoms of UTI, including pain, urgency, dysuria, fever, and peripheral leukocytosis, also have less predictive value for diagnosis of UTI in catheterized patients. Since bacteriuria and pyuria are expected in this population, the diagnostic evaluation of a febrile patient with a urinary catheter should not end with an abnormal urinalysis. To make a diagnosis of CAUTI, first, signs and symptoms of UTI (localizing and/or systemic) should be present along with pyuria and bacteriuria. Additionally, other infectious and noninfectious causes of the patient's symptoms should be systematically ruled out before settling on the diagnosis of CAUTI.

DIAGNOSTIC TOOLS

History The diagnosis of any of the UTI syndromes or ASB begins with a detailed history **(Fig. 140-4)**. Patient-reported symptoms have high positive predictive value in uncomplicated cystitis, at least among adult women; data are lacking for other populations. Self-diagnoses among women with recurrent UTI are particularly accurate and account for the success of patient-initiated treatment for recurrent cystitis. A meta-analysis of the diagnostic utility of history and physical findings for UTI concluded that, in women presenting with at least one localizing urinary symptom (dysuria, frequency, hematuria, or back pain) and no complicating factors, the probability of acute cystitis or pyelonephritis is 50%. If vaginal discharge and complicating factors are absent and risk factors for UTI are present, then the probability of UTI is close to 90%. Further laboratory evaluation with dipstick testing or urine culture prior to antimicrobial treatment is not necessary in such patients unless there is concern for resistant pathogens (e.g., in patients presenting with recurrent UTI, prior treatment failure, known colonization with extensively resistant pathogens, or high local prevalence of resistance to the empirical antibiotic being given).

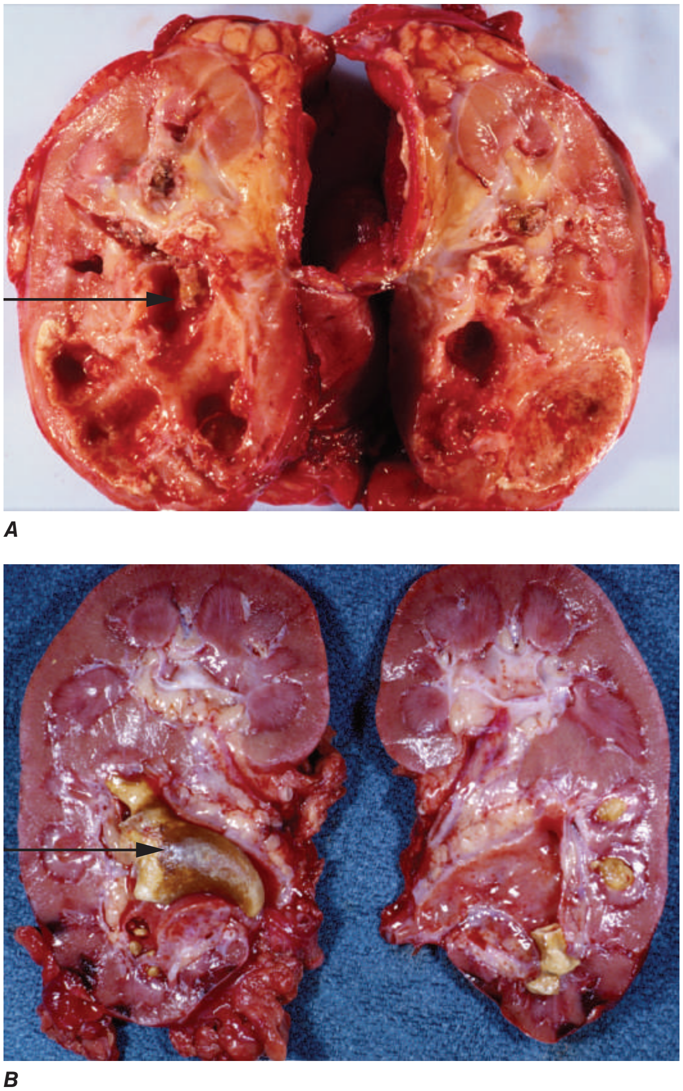

FIGURE 140-3 Xanthogranulomatous pyelonephritis. ***A.*** This photograph shows extensive destruction of renal parenchyma due to long-standing suppurative inflammation. The precipitating factor was obstruction by a staghorn calculus, which has been removed, leaving a depression (*arrow*). The mass effect of xanthogranulomatous pyelonephritis can mimic renal malignancy. ***B.*** A large staghorn calculus (*arrow*) is seen obstructing the renal pelvis and calyceal system. The lower pole of the kidney shows areas of hemorrhage and necrosis with collapse of cortical areas. *(Images courtesy of Dharam M. Ramnani, MD, Virginia Urology Pathology Laboratory, Richmond, VA.)*

In applying the patient's history as a diagnostic tool, one significant concern is that STIs—caused by *C. trachomatis* in particular—might be inappropriately treated as UTI. Importantly, women with multiple sexual partners and inconsistent use of condoms are at increased risk for both UTI and STIs, and symptoms alone do not always distinguish between these conditions. Dysuria might instead be due to cervicitis (*C. trachomatis*, *N. gonorrhoeae*), vaginitis (*Candida albicans*, *Trichomonas vaginalis*), herpetic urethritis, interstitial cystitis, or noninfectious vaginal or vulvar irritation. Vaginal discharge in particular should prompt consideration of STI. In this particular patient population, consideration should

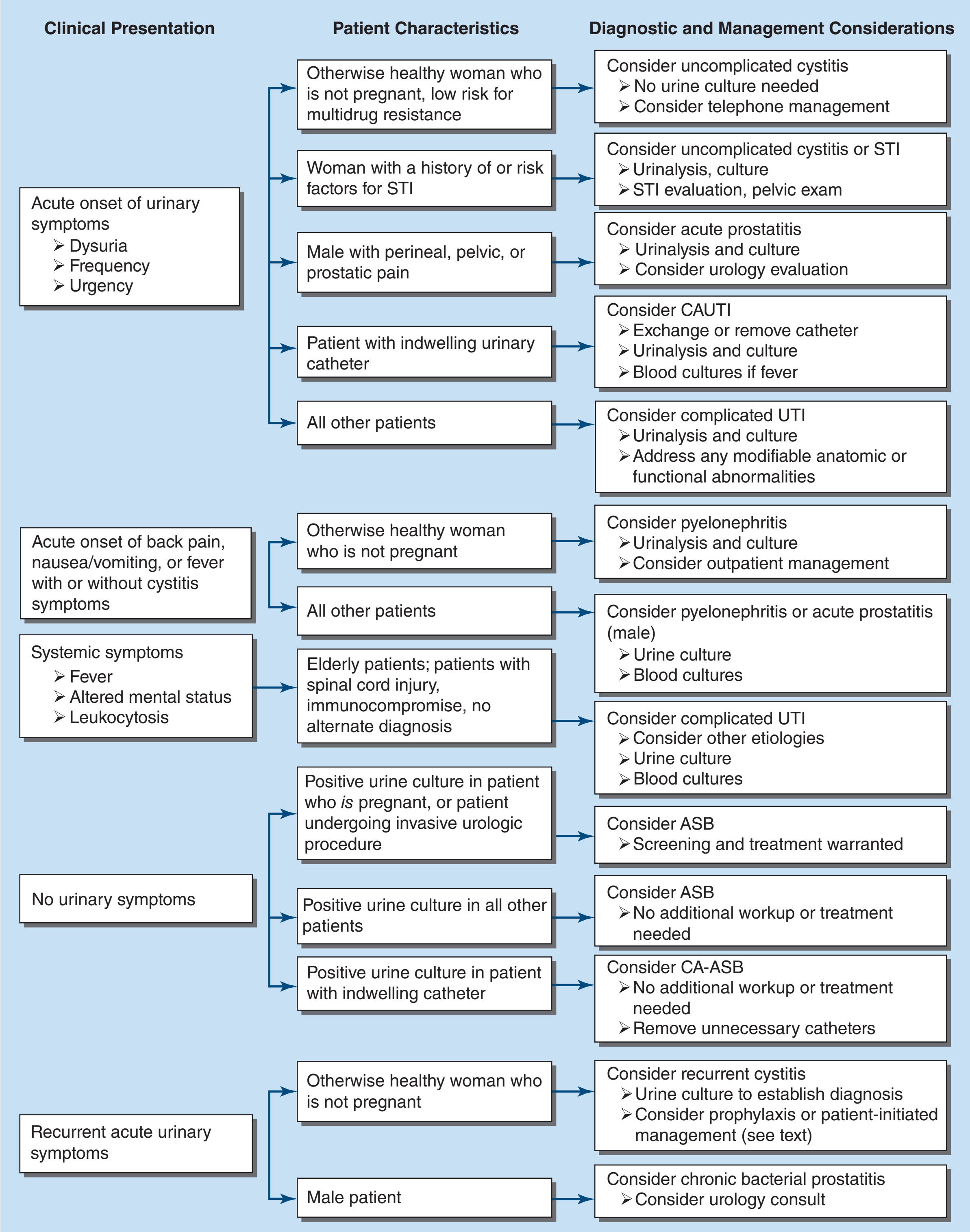
Clinical Presentation
Patient Characteristics
Diagnostic and Management Considerations
Acute onset of urinary symptoms ➢ Dysuria ➢ Frequency ➢ Urgency
Otherwise healthy woman who is not pregnant, low risk for multidrug resistance
Consider uncomplicated cystitis ➢ No urine culture needed ➢ Consider telephone management
Woman with a history of or risk factors for STI
Consider uncomplicated cystitis or STI ➢ Urinalysis, culture ➢ STI evaluation, pelvic exam
Male with perineal, pelvic, or prostatic pain
Consider acute prostatitis ➢ Urinalysis and culture ➢ Consider urology evaluation
Patient with indwelling urinary catheter
Consider CAUTI ➢ Exchange or remove catheter ➢ Urinalysis and culture ➢ Blood cultures if fever
All other patients
Consider complicated UTI ➢ Urinalysis and culture ➢ Address any modifiable anatomic or functional abnormalities
Acute onset of back pain, nausea/vomiting, or fever with or without cystitis symptoms
Otherwise healthy woman who is not pregnant
Consider pyelonephritis ➢ Urinalysis and culture ➢ Consider outpatient management
All other patients
Consider pyelonephritis or acute prostatitis (male) ➢ Urine culture ➢ Blood cultures
Systemic symptoms ➢ Fever ➢ Altered mental status ➢ Leukocytosis
Elderly patients; patients with spinal cord injury, immunocompromise, no alternate diagnosis
Consider complicated UTI ➢ Consider other etiologies ➢ Urine culture ➢ Blood cultures
No urinary symptoms
Positive urine culture in patient who *is* pregnant, or patient undergoing invasive urologic procedure
Consider ASB ➢ Screening and treatment warranted
Positive urine culture in all other patients
Consider ASB ➢ No additional workup or treatment needed
Positive urine culture in patient with indwelling catheter
Consider CA-ASB ➢ No additional workup or treatment needed ➢ Remove unnecessary catheters
Recurrent acute urinary symptoms
Otherwise healthy woman who is not pregnant
Consider recurrent cystitis ➢ Urine culture to establish diagnosis ➢ Consider prophylaxis or patient-initiated management (see text)
Male patient
Consider chronic bacterial prostatitis ➢ Consider urology consult

FIGURE 140-4 Diagnostic approach to urinary tract infection (UTI). NTF, nitrofurantoin; TMP-SMX, trimethoprim-sulfamethoxazole; UTI, urinary tract infection.

be given to STI screening either at the time of empirical treatment for UTI or if symptoms do not resolve with treatment.

Urine Dipstick Test, Urinalysis, and Urine Culture Useful tools in the diagnosis of UTI include the urine dipstick test (for nitrite and leukocyte esterase) and microscopic urinalysis (for counts of red and white blood cells), both of which provide point-of-care information that can primarily be used for ruling out UTI in patients who do not have a clear clinical diagnosis. Urine culture, another diagnostic tool, can retrospectively provide proof that bacteria were present in the urinary tract and can provide information on susceptibilities of the isolated organism(s). A key point is that patients with urinary catheterization will almost *always* have both pyuria and bacteriuria. *Therefore, presence of pyuria and bacteriuria alone in the absence of symptoms do not indicate a diagnosis of UTI.*

Understanding the parameters of the dipstick test is important in interpreting its results. Only members of the *Enterobacterales* family convert nitrate to nitrite, and enough nitrite must accumulate in the urine to reach the threshold of detection. This poses three problems limiting nitrites' sensitivity and specificity. First, nitrites could be positive in a patient with Enterobacterales ASB. Second, nitrites would be negative in a patient with UTI due to non-*Enterobacterales* organisms.

Third, nitrites might also be falsely negative in a woman with acute cystitis who is voiding frequently because of significant oral fluid intake.

The leukocyte esterase test detects this enzyme in neutrophils present in the host's urine, whether the cells are intact or lysed. Practically speaking, it is a less sensitive and specific surrogate for pyuria as measured by urine microscopy. Many reviews have attempted to describe the diagnostic accuracy of dipstick testing for UTI. The most important point for practicing clinicians is that a dipstick test negative for both nitrite and leukocyte esterase should prompt consideration of diagnoses other than UTI to explain the patient's symptoms, keeping in mind that the dipstick results can be falsely negative in a small percentage of cases. Importantly, a negative urine dipstick test is not sufficiently sensitive to rule out bacteriuria in pregnant women, for whom ASB screening should be performed via urine culture.

Urine microscopy reveals pyuria in nearly all cases of cystitis and hematuria in approximately 30% of cases. Pyuria has historically been defined as >10 leukocytes per high-powered microscopy field (HPF). Modern data indicate that, at least for older women, median urine white blood cells (WBCs) per HPF are far higher in UTI. As such, the cutoff of >10 WBCs/HPF has poor specificity (36%) for UTI, and a cutoff closer to 250 WBCs/HPF might better correlate with presence of urinary symptoms. However, at present, >10 WBCs/HPF remains the generally accepted standard for what defines presence of pyuria.

Detection of bacteria in a urine culture from a patient with symptoms of cystitis can confirm the diagnosis of UTI. However, culture results often do not become available until at least 24 h after a patient presents for care, with identification of individual organisms and their susceptibilities usually requiring an additional 24–48 h. Furthermore, identifying the presence of bacteria in the urine does not necessarily imply the presence of symptoms; therefore, a positive urine culture is consistent with both UTI and ASB. This issue is a perennial source of inaccuracy for retrospective studies of UTI, which are often flawed by defining UTI as the combination of pyuria and bacteriuria without requiring clinical documentation of symptoms. Studies of women with symptoms of cystitis have found that a colony count threshold of $>10^2$ bacteria/mL is more sensitive (95%) than a threshold of 10^5/mL for the diagnosis of acute cystitis in women. Fewer data are available in men about the threshold of bacteriuria to establish cystitis. When interpreting urine culture results, the clinician should consider that contamination with the normal microbial flora of the distal urethra, vagina, or skin is common and that these contaminants can grow to high numbers when the collected urine is allowed to stand at room temperature. In most instances, a culture that yields mixed bacterial species is contaminated except in settings of long-term catheterization, chronic urinary retention, or the presence of a fistula between the urinary tract and the gastrointestinal or genital tract.

TREATMENT

Urinary Tract Infections

The approach to diagnosis and treatment is influenced by which of the UTI clinical syndromes is suspected and presence of risk factors for resistant pathogens (Fig. 140-4).

GENERAL CONSIDERATIONS FOR ANTIMICROBIAL THERAPY

Studies indicate that treatment of UTI is usually the first or second most common indication for antimicrobials in ambulatory, inpatient, and long-term-care settings. Responsible use of antimicrobials for this common infection has broad implications for preserving antibiotic effectiveness into the future.

Importantly, antimicrobial therapy is warranted for any true (symptomatic) UTI. Studies of nonantibiotic treatment for UTI (e.g., with nonsteroidal anti-inflammatory drugs alone) suggest that this approach delays clinical response and predisposes patients to more invasive infections. Delayed therapy, in which a patient receives a prescription for antibiotics but fills it only if symptoms fail to resolve in a day or two, has the potential advantage of avoiding antibiotic use in those patients who do not have cystitis to begin with; however, this approach has the downsides of prolonging discomfort and a small increased risk of progression to pyelonephritis in patients who do have true UTI. Potentially higher-yield antimicrobial stewardship interventions in UTI management include encouraging clinicians to avoid unnecessary treatment of ASB, to avoid ordering urine cultures in patients who lack urinary or systemic symptoms suggestive of UTI, and to avoid prescribing overly long antibiotic durations for UTI.

In 1999, trimethoprim-sulfamethoxazole (TMP-SMX) was recommended as the treatment of choice for uncomplicated cystitis in the first published UTI guidelines of the Infectious Diseases Society of America. However, local uropathogen resistance patterns and national differences in the availability of certain antimicrobial agents are such that providing a global or even national recommendation for a single preferred empirical antibiotic for UTI is challenging. Instead, clinicians must use their knowledge of local resistance patterns and individual patient factors (e.g., allergies and potential for drug–drug interactions) to select from groups of antimicrobials based on their efficacy, toxicity risks, and potential for collateral damage (defined below). The choice of antimicrobial agent, the dose, and the duration of therapy also depend on the clinical syndrome (that is, which type of UTI specifically is being treated) and the rapidity of clinical improvement.

Collateral damage refers to the adverse ecologic effects of antimicrobial therapy, including killing of the patient's normal flora (predisposing to *Clostridioides difficile* infection) and selection of drug-resistant organisms. The implication of collateral damage for UTI management is that a drug that is highly efficacious for treatment of UTI is not necessarily the optimal first-line agent if it also has pronounced secondary effects on the normal flora or is likely to significantly adversely affect resistance patterns. Drugs used for UTI that have a minimal effect on fecal flora include pivmecillinam, fosfomycin, and nitrofurantoin. In contrast, TMP-SMX, quinolones, and β-lactams (particularly the late-generation cephalosporins) affect the fecal flora more significantly and are notably the agents for which rising resistance has been documented.

Choosing judiciously whether to initiate antibiotic therapy and then selecting the agent with the least potential for collateral damage to be given for the shortest effective duration are important factors in global efforts to stem the rise of antimicrobial-resistant organisms. Common risk factors for resistance to specific antimicrobials in UTI include prior urine or blood cultures with an organism resistant to that antimicrobial, recent exposure to that antimicrobial, or travel to an area where resistance to that antimicrobial is especially prevalent. For each of these factors, the magnitude of risk is greater with more recent exposures in the preceding year; the greatest increase in risk occurs with fluoroquinolones. More data is available documenting this risk of resistance with prior exposure to the antibiotic for fluoroquinolones than for other antibiotics. Other general risk factors for antimicrobial resistance in UTI include urinary catheterization, prior exposure to antibiotics of any class, and residence in a healthcare facility.

ASYMPTOMATIC BACTERIURIA

Critically, treatment of ASB does not decrease frequency of subsequent true UTIs. Clinical trial data from renal transplant recipients suggest that ASB treatment might even increase future risk of infection. Importantly, though, treatment of ASB in pregnant patients and patients undergoing high-risk urologic procedures is recommended to reduce risk of complications (see discussion above) and should be guided by urine culture results. In all other populations, screening for and treating ASB are discouraged. Specifically, a urine culture does not need to be obtained at the time of placement or removal of a urinary catheter if the patient is asymptomatic. The majority of cases of catheter-associated bacteriuria are asymptomatic and do not warrant antimicrobial therapy.

CYSTITIS

Many episodes of uncomplicated cystitis in women (i.e., cystitis without systemic symptoms, symptoms of pyelonephritis, or urinary

TABLE 140-1 Treatment Strategies for Acute Uncomplicated Cystitis

DRUG	DOSE	DURATION	COMMON SIDE EFFECTS
Nitrofurantoin	100 mg bid	Women or men: 5–7 days	Nausea, headache
TMP-SMX	1 DS tablet bid	Women: 3 days Men: 7 days	Rash, urticaria, nausea, vomiting, hematologic abnormalities
Fosfomycin	3-g sachet	Women: 1 day Men: qod × 3 doses	Diarrhea, nausea, headache
Pivmecillinam	400 mg bid	Women: 3–7 days	Nausea, vomiting, diarrhea
Fluoroquinolones	Dose varies by agent	Women: 3 days Men: 7 days	Nausea, vomiting, diarrhea, headache, drowsiness, insomnia
β-Lactams	Dose varies by agent	Women or men: 5–7 days	Diarrhea, nausea, vomiting, rash, urticaria

Abbreviations: DS, double-strength; TMP-SMX, trimethoprim-sulfamethoxazole; bid, twice a day; qod every other day.

catheterization) can be managed remotely (Fig. 140-4). Studies of telephone-management algorithms for UTI indicate low risk of serious complications, with the important caveat that such algorithms have generally involved otherwise healthy women who are at low risk of complications of UTI; thus, patients with other UTI syndromes or more complicated prior histories require further diagnostic evaluation and might not be as adequately managed remotely.

As described above, cystitis in men is unusual except in the setting of urinary catheterization or obstructive uropathy. Before settling on this diagnosis in men, however, the clinician should be confident that ABP, CBP, and pyelonephritis with or without urinary obstruction have been excluded. True uncomplicated cystitis in afebrile men can be treated with only 7 days of antimicrobial therapy; the durations shorter than 7 days used for uncomplicated cystitis in women have not been adequately studied in men.

Several effective therapeutic regimens are available for acute uncomplicated cystitis in women; data are very limited on therapy for cystitis in men **(Table 140-1)**. Preferred agents for treatment of acute uncomplicated cystitis include TMP-SMX, nitrofurantoin, fosfomycin, and pivmecillinam. Fosfomycin is somewhat less studied than TMP-SMX and nitrofurantoin, but it is also effective and is preferred in cystitis because of its selective concentration in urine, low risk for collateral damage, and potential to retain efficacy against MDR *E. coli*. Pivmecillinam was approved for treatment of cystitis in the United States in 2024. It also often retains activity against MDR *E. coli*, including some expressing extended-spectrum β-lactamases.

Alternative agents for treatment of cystitis include fluoroquinolones and β-lactams. β-lactams require longer durations of therapy (5–7 days) even in uncomplicated cystitis and are associated with higher clinical failure rates (compared with either fluoroquinolones or TMP-SMX) in some randomized controlled trials. Despite these limitations, β-lactams are favored over fluoroquinolones as an alternative agent for treatment of uncomplicated cystitis because of the bacterial resistance and side effects associated with fluoroquinolones. Fluoroquinolones are highly effective for treating UTI but cause high collateral damage as well as numerous rare but serious drug toxicities (e.g., collagen vascular pathology such as tendon rupture, retinal detachment, and aortic aneurysm), leading the U.S. Food and Drug Administration (FDA) to advise against using fluoroquinolones for uncomplicated cystitis in women unless no alternatives are available. Importantly, in men, if there is concern for prostatic involvement, fluoroquinolones or TMP-SMX are preferred over β-lactams because fluoroquinolones have superior penetration into the prostate. The pros and cons of each specific agent are discussed in more detail below.

TMP-SMX remains an excellent option for treatment of cystitis in areas where local uropathogen antibiogram data do not clearly favor other preferred agents or when the patient has recent microbiology demonstrating a TMP-SMX–susceptible uropathogen. Otherwise, the optimal setting for empirical use of TMP-SMX is uncomplicated cystitis in a patient who has an established relationship with the practitioner and has no barriers to seeking further care if their symptoms do not respond promptly.

Resistance to nitrofurantoin remains low despite >70 years of use, probably because several mutational steps are required for bacterial resistance to this drug. Nitrofurantoin remains highly active against *E. coli*. However, importantly, most *Proteus*, *Providencia*, *Morganella*, *Pseudomonas*, and *Serratia* and some *Klebsiella* and *Enterobacter* strains are intrinsically resistant to this drug. A 5-day course of nitrofurantoin is as effective as a 3-day course of TMP-SMX for treatment of acute uncomplicated cystitis in women. Nitrofurantoin should not be used in patients with a creatinine clearance <30 mL/min. It does not reach significant levels in tissue and so cannot be used to treat pyelonephritis or prostatitis. Additionally, we advise against long-term (i.e., multiple years) prescription of nitrofurantoin as prophylaxis for recurrent UTIs because of the risk of sometimes irreversible pulmonary toxicities.

Guidelines also recommend fosfomycin, which interferes with bacterial cell-wall formation, as an option for uncomplicated cystitis, given its easy dosing and low collateral damage. While fosfomycin is active against *E. coli*, its activity against other Enterobacterales species is less reliable. Moreover, fosfomycin susceptibility testing is difficult to perform, and the results of such testing are not typically included in standard automated microbiologic susceptibility reports. Oral fosfomycin is given as a single 3-g dose sachet (powder) that is dissolved in a glass of water and swallowed. Similar to nitrofurantoin, renal tissue levels of fosfomycin are low, and a different antibiotic should be used if treating pyelonephritis. While not extensively studied for this indication, fosfomycin has been used in the treatment of prostatitis to manage pathogens resistant to other oral agents.

Despite their often high rates of urinary secretion, β-lactams as a class have not performed as well as TMP-SMX or fluoroquinolones for treatment of acute cystitis. Rates of pathogen eradication are lower, and relapse rates are higher with β-lactam drugs compared with other agents. The generally accepted explanation for this observation is that β-lactams fail to eradicate uropathogens from the vaginal reservoir, which is specific to female patients with UTI. Supporting this hypothesis are limited indirect data indicating that treatment success with cephalexin is as good as or better than success with other classes in male UTI. Amoxicillin resistance has become widespread in *E. coli*, and many strains resistant to TMP-SMX are also resistant to some or all of the other oral β-lactams. Thus, as described for TMP-SMX above, β-lactams such as cefpodoxime or cefixime are optimally used for patients known to have susceptible uropathogens or in areas where local antibiograms demonstrate that resistance remains uncommon.

The fluoroquinolones ciprofloxacin and levofloxacin are highly effective as short-course therapy for cystitis due to susceptible flora. By contrast, moxifloxacin in particular might not reach adequate urinary levels. The two main concerns about using fluoroquinolones for acute cystitis are the propagation of fluoroquinolone resistance, not only among uropathogens but also among other organisms causing more serious and difficult-to-treat infections at other sites, and fluoroquinolones' rare but potentially serious adverse effects (increased risk of Achilles' tendon rupture, neuropathy, and aortic dissection). In light of these potential detrimental effects, the FDA issued an advisory against using fluoroquinolones to treat acute cystitis in patients who have other effective therapeutic options.

Urinary analgesics are appropriate in certain situations to help the patient with resolution of bladder discomfort in cystitis. The urinary tract analgesic phenazopyridine is widely used but can cause significant nausea. Combination analgesics containing urinary antiseptics (such as methenamine and methylene blue), a urine-acidifying agent (sodium phosphate), and an antispasmodic agent (hyoscyamine) also are available. There are no well-established criteria for when to use these agents, although as mentioned above, multiple clinical trials demonstrate that symptomatic treatments alone are not an adequate substitute for antimicrobials when patients have true UTI.

PROSTATITIS

The prostate is involved in the majority of cases of febrile UTI in men and can also be a source of recurrent cystitis. ABP should specifically be suspected in men presenting with UTI with systemic symptoms without another clear anatomic location of UTI, such as pyelonephritis. CBP should be suspected when patients have recurrent UTI, especially if not in the context of urinary catheterization). Both ABP and CBP should be suspected when patients have prostatic symptoms. If ABP is suspected, antimicrobial therapy should be promptly initiated after urine and blood cultures are obtained. For ABP, at least 2 weeks and potentially up to 4 weeks of a fluoroquinolone or TMP-SMX is recommended for susceptible uropathogens. Identification of a prostatic abscess should prompt consultation to urologic specialists. For CBP, a 4- to 6-week course of antibiotics is typically given, and recurrences (not uncommon) may warrant up to a 12-week course of therapy. These durations are based on older observational studies and historic norms of practice and need be to clarified with modern randomized controlled trials.

PYELONEPHRITIS

Patients with pyelonephritis have tissue-invasive disease and are often systemically ill. Therefore, empirical therapy should be reliable and rapidly reach therapeutic levels in both renal parenchyma and blood. In years past, increasing TMP-SMX resistance made fluoroquinolones the empirical treatment of choice for pyelonephritis, and quinolones remain among the best-studied agents for acute pyelonephritis. However, quinolone resistance is now also increasingly prevalent.

For hospitalized patients, we recommend empirical therapy with an IV β-lactam for pyelonephritis, with attention to choosing an agent to which the patient has not recently had a resistant uropathogen in urine or blood and appears highly reliable based on the local antibiogram. For patients with pyelonephritis who are not and do not need to be hospitalized, we recommend empirical therapy with TMP-SMX, ciprofloxacin, or levofloxacin based on local antibiogram data, along with urine culture collection and short-term follow-up. Optimally, an initial dose of IV or IM β-lactam or an aminoglycoside to ensure effective empirical therapy can be prescribed, where this is practical and feasible. Ambulatory patients with pyelonephritis should be hospitalized if they are unable to tolerate oral medications (e.g., due to nausea and vomiting) or are unlikely to adhere to short-term follow-up.

For susceptible isolates, clinical trial data indicate that 5–7 days of an oral fluroquinolone or 7–14 days of oral TMP-SMX (depending on rapidity of clinical improvement) are highly effective options for pyelonephritis. Oral β-lactam agents are less effective and should be used with caution and close follow-up. Nitrofurantoin, pivmecillinam, and fosfomycin should not be used for pyelonephritis. Options for parenteral therapy for uncomplicated pyelonephritis include late-generation cephalosporins (such as ceftriaxone or cefepime), piperacillin-tazobactam, a carbapenem, or fluoroquinolones (although the high bioavailability of this class makes IV administration superfluous for patients with who can tolerate and absorb oral medications). Alternative parenteral antibiotics include novel β-lactam/β-lactamase inhibitors (e.g., meropenem-vaborbactam) and aminoglycosides, which are considered alternative due to cost and toxicity profiles, respectively.

In general, the treatment of pyelonephritis should be guided by urine culture results and clinical resolution of symptoms. Pyelonephritis due to drug-resistant uropathogens requiring nonpreferred antibiotics should be treated in consultation with infectious disease specialists. Once the patient has responded clinically, oral therapy can be substituted for parenteral therapy. Treatment can be concluded at 7 days for patients who have clinically responded (i.e., resolution of fever and return to hemodynamic stability with adequate source control) within that timeframe and who can be treated with a fluoroquinolone or TMP-SMX. Oral β-lactams may necessitate a longer course of treatment for pyelonephritis.

Patients with pyelonephritis who have not had a clinical response within 72 h should undergo further evaluation for an uncontrolled source of infection. Pyelonephritis can be complicated by acute papillary necrosis causing urinary obstruction, which requires intervention to relieve the obstruction, overcome the infection, and preserve renal function. Perinephric and renal cortical abscesses may require drainage percutaneous drainage. Percutaneous drainage can also be used as the initial therapy in emphysematous pyelonephritis and can be followed by elective nephrectomy as needed depending upon the patient's clinical status in consultation with urologic specialists. Xanthogranulomatous pyelonephritis (XGP) is treated with close collaboration with urologic colleagues and often requires nephrectomy; antibiotic treatment duration should be individualized in cases of XGP.

CATHETER-ASSOCIATED UTI

The bacteriology of CAUTI is diverse, and urine culture results are essential to guide treatment. The catheter should be exchanged (or, if possible, removed entirely) during treatment for CAUTI, the purpose of which is to remove biofilm-associated organisms that could serve as a nidus for reinfection. When possible, this should be done prior to sending urine cultures to ensure that the culture results reflect the microbiology of the bladder rather than the catheter walls.

Most clinicians treat CAUTI for 7–14 days based on clinical response; the need for longer therapy than is used in uncomplicated cystitis is supported by a randomized controlled trial that identified more relapses after 3 versus 14 days of antimicrobials. Further studies on the optimal duration of therapy in CAUTI are needed.

The best strategy for prevention of CAUTI is to avoid insertion of unnecessary catheters and to remove catheters once they are no longer necessary. Quality-improvement collaboratives that have addressed technical aspects of CAUTI prevention (such as avoidance of inappropriate catheterization) and team-communication strategies have shown the benefit of such approaches in decreasing rates of CAUTI in both acute- and long-term-care settings. Antimicrobial catheters impregnated with silver or nitrofurazone have not been shown to provide significant clinical benefit with regard to reducing CAUTI rates and should not be used routinely. Evidence is insufficient to systematically recommend suprapubic catheters and condom urinary catheters as alternatives to indwelling urinary catheters as a way to prevent bacteriuria. Intermittent catheterization may be preferable to long-term indwelling urethral catheterization in certain populations (e.g., patients with spinal cord injury) to prevent both infectious and anatomic complications.

UTI IN PREGNANT PATIENTS

Aminopenicillins (ampicillin and amoxicillin) and cephalosporins have been used extensively in pregnancy and are the drugs of choice for treatment of ASB or UTI in pregnant patients. Importantly, though, widespread resistance to aminopenicillins precludes their use for pregnant patients with UTI if urine culture results are not already available. One retrospective case-control study suggested an association between nitrofurantoin use in the first trimester and birth defects, but this association has not been confirmed, and the

American College of Obstetrics and Gynecology (ACOG) suggests that nitrofurantoin can be used during all trimesters if β-lactam agents are not an option. Sulfonamides (including TMP-SMX) should be avoided both in the first trimester (because of possible teratogenic effects) and near term (because of a possible role in the development of neonatal kernicterus), although ACOG suggests TMP-SMX can be used as a potential treatment option at both times if no alternatives are available. Fluoroquinolones are avoided during pregnancy because of possible adverse effects on fetal cartilage development. Pregnant patients with ASB or cystitis are generally treated for 5–7 days because of the absence of evidence to support shorter durations of therapy in this patient population. For pregnant patients with pyelonephritis, guidelines recommend parenteral β-lactam therapy with or without aminoglycosides based on limited observational data and historic norms of practice.

CANDIDURIA

The appearance of *Candida* in the urine is an increasingly common complication of indwelling catheterization, particularly for patients in the intensive-care unit, those taking broad-spectrum antimicrobial drugs, and those with underlying diabetes mellitus. In many studies, >50% of urinary *Candida* isolates have been found to be non–*C. albicans* species. The clinical presentation of candiduria varies widely—from asymptomatic colonization to pyelonephritis and sepsis—thereby making management challenging in some instances, particularly for patients with multiple other ongoing medical conditions. Removal of the urinary catheter results in resolution of candiduria in more than one-third of asymptomatic cases. As with ASB, treatment of asymptomatic patients with candiduria has not been found to reduce recurrence of candiduria or subsequent infection. Therefore, treatment of asymptomatic patients is not routinely recommended for most patients.

By contrast, candiduria should be treated in patients who have symptomatic cystitis or pyelonephritis or those who are asymptomatic but are at high risk for disseminated disease (e.g., patients with neutropenia, patients who will undergo urologic manipulation, and low-birth-weight infants). Fluconazole (200–400 mg/d for 7–14 days) reaches high levels in the urine and is the first-line regimen for *Candida* infections of the urinary tract. Although instances of successful eradication of candiduria by some of the newer azoles and echinocandins have been reported, these agents are relatively unstudied and are not recommended as first-line agents but can be considered in certain cases. For *Candida* isolates with high levels of resistance to fluconazole, oral flucytosine and/or parenteral amphotericin B are additional options. Bladder irrigation with amphotericin B is generally not recommended.

PREVENTION OF RECURRENT UTI IN WOMEN

Recurrence of uncomplicated cystitis in women is common, and a preventive strategy is indicated if recurrent UTIs are interfering with a patient's lifestyle. The threshold of three or more symptomatic episodes per year to intervene or develop a specific preventive strategy is not absolute; clinicians are recommended to engage in shared decision-making with each individual patient to make determinations about a UTI preventive strategy.

Initial evaluation of a patient with recurrent cystitis should include confirmation of the diagnosis of recurrent UTI, imaging to identify renal calculi or another potential nidus of infection, and assessment for urinary retention. Referrals to urology, urogynecology, and/or pelvic physical therapy might be helpful when the diagnosis is unclear or an anatomic abnormality is suspected.

Nonantimicrobial prophylactic approaches to recurrent UTI are increasingly being studied. Observational data indicate that increasing fluid intake might reduce recurrent UTIs, and such an approach is generally well tolerated by healthy adults. Two recent clinical trials found that methenamine hippurate (converted to the antiseptic formaldehyde in the bladder) is well tolerated and might reduce recurrent UTIs as much as antibiotic prophylaxis. A recent Cochrane review of randomized clinical trials of cranberry products also found a significant benefit for adult women and children with recurrent UTI. Less is known about which products and what precise dosing are optimal. *Lactobacillus* probiotics are under investigation for UTI prevention, with one recent factorial trial suggesting that intravaginal but not oral lactobacilli products reduce recurrent UTIs. We advise caution, however, in recommending probiotics to immunocompromised patients, in whom probiotic-associated invasive infections have been described. Studies of mannose products for UTI prevention have produced mixed results.

Three antibiotic prophylactic strategies are available: continuous (using daily or thrice-weekly dosing), postcoital (one dose after intercourse), and patient-initiated therapy. For the continuous and postcoital strategies, low doses of TMP-SMX or nitrofurantoin are typically utilized. Such approaches are usually highly effective during the period of active antibiotic intake. A prophylactic regimen is typically prescribed for 6 months and then discontinued. If bothersome infections recur, the prophylactic program can be reinstituted for the same duration or for a longer period.

Selection of resistant strains in the fecal flora has been documented in studies of patients taking prophylactic antibiotics. An approach of alternating antibiotic prophylaxis with periods of nonantimicrobial prevention could hypothetically allow for intermittent recovery of the gut microbiome. The choice between methenamine and cranberry products should be based on shared decision-making and individual patient response. Some studies suggest that although the rate of UTI will be decreased when using nonantimicrobial prophylaxis compared with no prophylaxis, the rate of UTI might be slightly higher than during the antibiotic-suppression period.

Patient-initiated therapy involves supplying the patient with materials for urine culture and with a course of antibiotics for self-medication at the first symptoms of infection. The urine culture is refrigerated and delivered to the physician's office for confirmation of the diagnosis. When an established and reliable patient–provider relationship exists, the urine culture can be omitted as long as the symptomatic episodes respond completely to short-course therapy and are not followed by relapse.

PROGNOSIS

Cystitis is a risk factor for recurrent cystitis and pyelonephritis. ASB is common among elderly and catheterized patients but does not in itself increase risk of infection or death. The relationships among recurrent UTI, chronic pyelonephritis, and renal insufficiency have been widely studied. In the absence of anatomic abnormalities such as reflux, recurrent infections in children and adults do not lead to chronic pyelonephritis or renal failure. Additionally, infection does not play a primary role in chronic interstitial nephritis; the primary etiologic factors in this condition are analgesic overuse, obstruction, reflux, and toxin exposure. By contrast, in the presence of underlying renal abnormalities (particularly obstructing stones), infection as a secondary factor can accelerate renal parenchymal damage.

FURTHER READING

Bilsen M et al: Current pyuria cutoffs promote inappropriate urinary tract infection diagnosis in older women. Clin Infect Dis 76:2070, 2023.

Brehm TJ et al: Acute and chronic infectious prostatitis in older adults. Infect Dis Clin North Am 37:175, 2023.

Christmas MM et al: Menopause hormone therapy and urinary symptoms: A systematic review Menopause 30:672, 2023.

Drekonja DM et al: Effect of 7 vs 14 days of antibiotic therapy on resolution of symptoms among afebrile men with urinary tract infection: A randomized clinical trial. JAMA 326:324, 2021.

Gupta V et al: Effectiveness of prophylactic oral and/or vaginal probiotic supplementation in the prevention of recurrent urinary tract infections: A randomized, double-blind, placebo-controlled trial. Clin Infect Dis 78:1154, 2024.

Hooton TM et al: Asymptomatic bacteriuria and pyuria in premenopausal women. Clin Infect Dis 72:1332, 2021.

Lafaurie M et al: Antimicrobial for 7 or 14 days for febrile urinary tract Infection in men: A multicenter noninferiority double-blind, placebo-controlled, randomized clinical trial. Clin Infect Dis 76:2154, 2023.

Patel PK et al: Strategies to prevent catheter-associated urinary tract infections in acute-care hospitals: 2022 update. Infect Control Hosp Epidemiol 44:1209, 2023.

Roth R et al: The urobiome in men and women: A clinical review. Clin Microbiol Infect 29:1242, 2023.

Uitrakul S et al: The incidence and risk factors of urinary tract infection in patients with type 2 diabetes mellitus using SGLT2 inhibitors: A real-world observational study. Medicines (Basel) 9:59, 2022.

141

Sexually Transmitted Infections: Overview and Clinical Approach

Jeanne M. Marrazzo

CLASSIFICATION AND EPIDEMIOLOGY

Worldwide, most adults acquire at least one sexually transmitted infection (STI), and many remain at risk for complications. Each day, for example, more than 1 million STIs are acquired worldwide, placing many affected persons at risk for adverse reproductive health outcomes and neoplasia. Certain STIs, such as syphilis, gonorrhea, HIV infection, hepatitis B, and chancroid, often occur in highly interconnected sexual networks characterized by high rates of partner change or multiple concurrent partners. Such networks, for example, often include persons who engage in transactional sex, some men who have sex with men (MSM), and persons involved in the use of illicit drugs. Other STIs are distributed more evenly throughout populations. For example, chlamydial infections, genital human papillomavirus (HPV) infections, and genital herpes can spread efficiently even in relatively low-risk populations. Finally, modern technologies based on detection of nucleic acid have accelerated elucidation of the role of sexual transmission in the spread of some viruses, including Ebola virus and Zika virus, and have provided new evidence of apparent sexual transmission of several bacteria, including group C *Neisseria meningitidis* and anaerobes associated with bacterial vaginosis (BV).

In general, the product of three factors determines the initial rate of spread of any STI within a population: rate of sexual exposure of susceptible to infectious people, efficiency of transmission per exposure, and duration of infectivity of those infected. Accordingly, efforts to prevent and control STIs aim to decrease the rate of sexual exposure of susceptible to infected persons (e.g., through education and efforts to change sexual behavior norms and through control efforts aimed at reducing the proportion of the population infected, including postexposure prophylaxis [PEP]); to decrease the duration of infectivity (through early diagnosis and curative or suppressive treatment); and to decrease the efficiency of transmission (through promotion of condom use and safer sexual practices, use of effective vaccines, and male medical circumcision).

In all societies, STIs rank among the most common of all infectious diseases, with at least 40 microorganisms now classified as predominantly sexually transmitted or as frequently sexually transmissible **(Table 141-1)**. In developing countries, with three-quarters of the world's population and 90% of the world's STIs, factors such as population growth (especially in adolescent and young-adult age groups), rural-to-urban migration, wars, limited or no provision of reproductive health services for women, and poverty create exceptional vulnerability to disease resulting from unprotected sex. During the 1990s in China, Russia, the other states of the former Soviet Union, and South Africa, internal social structures changed rapidly as borders opened to the West, unleashing enormous new epidemics of HIV infection and other STIs; such patterns persist in the face of ongoing conflicts. Despite advances in the provision of highly effective antiretroviral therapy worldwide, HIV remains the leading cause of death in some developing countries, and HPV and hepatitis B virus (HBV) remain important causes of cervical and hepatocellular carcinoma, respectively—two of the most common (and preventable) malignancies in the developing world. Sexually transmitted herpes simplex virus (HSV) infection causes most genital ulcer disease throughout the world, and an increasing proportion of cases of genital herpes occur in developing countries with generalized HIV epidemics, where the positive feedback loop between HSV and HIV transmission remains intractable. Despite this consistent link, randomized trials evaluating the efficacy of antiviral therapy in suppressing HSV in both HIV-uninfected and HIV-infected persons have demonstrated no protective effect against acquisition or transmission of HIV. The World Health Organization estimated that 357 million new cases of four curable STIs—gonorrhea, chlamydial infection, syphilis, and trichomoniasis—occurred annually in recent years. Up to 50% of women of reproductive age in developing countries have BV (arguably acquired sexually). All of these curable STIs have been associated with increased risk of HIV transmission or acquisition.

TABLE 141-1 Sexually Transmitted and Sexually Transmissible Microorganisms

BACTERIA	VIRUSES	OTHER[a]
Transmitted in Adults Predominantly by Sexual Intercourse		
Neisseria gonorrhoeae *Chlamydia trachomatis* *Treponema pallidum* *Haemophilus ducreyi* *Klebsiella (Calymmatobacterium) granulomatis* *Ureaplasma urealyticum* *Mycoplasma genitalium*	HIV (types 1 and 2) Human T-cell lymphotropic virus type 1 Herpes simplex virus type 2 Human papillomavirus (multiple genital genotypes) Hepatitis B virus[b] Molluscum contagiosum virus	*Trichomonas vaginalis* *Pthirus pubis*
Sexual Transmission Repeatedly Described but Not Well Defined or Not the Predominant Mode		
Mycoplasma hominis *Gardnerella vaginalis* and other vaginal bacteria Group B *Streptococcus* *Mobiluncus* spp. *Helicobacter cinaedi* *Helicobacter fennelliae* Anaerobes associated with bacterial vaginosis *Leptotrichia/Sneathia* Group C *Neisseria meningitidis*	Cytomegalovirus Human T-cell lymphotropic virus type 2 Hepatitis C virus (?) Hepatitis D virus Herpes simplex virus type 1 Zika virus Ebola virus (?) Epstein-Barr virus Human herpesvirus type 8	*Candida albicans* *Sarcoptes scabiei*
Transmitted by Sexual Contact Involving Oral–Fecal Exposure; of Declining Importance in Men Who Have Sex with Men		
Shigella spp. *Campylobacter* spp.	Hepatitis A virus	*Giardia lamblia* *Entamoeba histolytica*

[a] Includes protozoa, ectoparasites, and fungi. [b] Among U.S. patients for whom a risk factor can be ascertained, most hepatitis B virus infections are transmitted sexually.

In the United States, the prevalence of antibody to HSV-2 began to fall in the late 1990s, especially among adolescents and young adults; the decline was presumably due to delayed sexual debut, increased condom use, and lower rates of multiple (four or more) sex partners—all well documented by the U.S. Youth Risk Behavior Surveillance System. The estimated annual incidence of HBV infection has also declined dramatically since the mid-1980s; this decrease is probably attributable to now-widespread administration of hepatitis B vaccine in infancy. Genital HPV remains the most common sexually transmitted pathogen in the United States, infecting 60% of a cohort of initially

HPV-negative, sexually active Washington state college women within 5 years in a study conducted from 1990 to 2000—i.e., during the pre–HPV immunization era. Global expansion of HPV vaccine coverage among young women has already shown promise in reducing the incidence of infection with the HPV types included in the vaccines and of conditions associated with these viruses, including invasive cervical cancer; however, gaps in immunization coverage remain.

In industrialized countries, fear of HIV infection in the mid-1980s and through the mid-2000s, coupled with widespread behavioral interventions and better-organized systems of care for the curable STIs, initially helped curb the transmission of several STDs. However, with current antiretroviral therapy, HIV has become for many a chronic disease associated with a normal life span and high quality of life, and HIV pre-exposure prophylaxis (PrEP) has proven highly effective in preventing HIV acquisition, further reducing motivation to use barrier protection such as condoms. Rates of gonorrhea and syphilis remain higher in the United States than in any other Western industrialized country.

In the United States, the Centers for Disease Control and Prevention (CDC) has compiled reported rates of STIs since 1941. The incidence of reported gonorrhea peaked at 468 cases per 100,000 population in the mid-1970s and fell to a low of 98 cases per 100,000 in 2012; in 2018, the case rate was 179.1 per 100,000 persons, which is more than an 80% increase since 2009 when the number of new cases reached an all-time low. With increased testing and more sensitive tests, the incidence of reported *Chlamydia trachomatis* infection has been increasing steadily since reporting began in 1984, reaching an all-time peak of 457.6 cases per 100,000 in 2011. The incidence of primary and secondary syphilis per 100,000 peaked at 71 cases in 1946, fell rapidly to 3.9 cases in 1956, ranged from ~10 to 15 cases through 1987 (with markedly increased rates among MSM and African Americans), and then fell to a nadir of 2.1 cases in 2000–2001 (with rates falling most rapidly among heterosexual African Americans). However, since 1996, with the introduction of highly active antiretroviral therapy, gonorrhea, syphilis, and chlamydial infection have had a remarkable resurgence among MSM in North America and Europe, where outbreaks of a rare type of chlamydial infection (lymphogranuloma venereum [LGV]) that had virtually disappeared during the AIDS era have occurred. In 2022, 207,255 cases of syphilis (all stages and congenital syphilis) were reported in the United States, which is the greatest number of cases reported since 1950 and an increase of 17.3% since 2021. In 2022, 45% of primary and secondary syphilis cases among males reported to the CDC were in MSM, but incidence has also increased in women, with a concomitant increase in congenital syphilis. In 2022, 3755 cases of congenital syphilis were reported, including 282 congenital syphilis-related stillbirths and infant deaths. The national congenital syphilis rate of 102.5 cases per 100,000 live births in 2022 represents a 30.6% increase relative to 2021 and is the highest reported rate since 1991.

Moreover, the uptake of oral PrEP for HIV-1 acquisition has increased among MSM since its initial approval for this purpose in 2012 and has been associated with reports of reduced condom-use frequency and concomitantly increased STI acquisition. These developments have resulted in a soaring incidence of STIs, with increasing co-infection with HIV and other sexually transmitted pathogens (particularly *Treponema pallidum*, the cause of syphilis; and *Neisseria gonorrhoeae*, the cause of gonorrhea), primarily among MSM.

TABLE 141-2 Major Sexually Transmitted Disease Syndromes and Sexually Transmitted Microbial Etiologies

SYNDROME	SEXUALLY TRANSMITTED MICROBIAL ETIOLOGIES
AIDS	HIV types 1 and 2
Urethritis: males	*Neisseria gonorrhoeae, Chlamydia trachomatis, Mycoplasma genitalium, Ureaplasma urealyticum* (subspecies *urealyticum*), *Trichomonas vaginalis,* HSV, some anaerobic bacteria, *Leptotrichia/Sneathia*
Epididymitis	*C. trachomatis, N. gonorrhoeae,* and (in older men or men who have sex with men) coliform bacteria
Lower genital tract infections: females	
Cystitis/urethritis	*C. trachomatis, N. gonorrhoeae,* HSV
Mucopurulent cervicitis	*C. trachomatis, N. gonorrhoeae, M. genitalium*
Vulvitis	*Candida albicans,* HSV
Bartholinitis	*C. albicans, T. vaginalis*
Vulvovaginitis	*C. albicans, T. vaginalis*
BV	BV-associated bacteria (see text)
Acute pelvic inflammatory disease	*N. gonorrhoeae, C. trachomatis,* BV-associated bacteria, *M. genitalium,* group B streptococci
Infertility	*N. gonorrhoeae, C. trachomatis,* BV-associated bacteria
Ulcerative lesions of the genitalia	HSV-1, HSV-2, *Treponema pallidum, Haemophilus ducreyi, C. trachomatis* (LGV strains), *Klebsiella (Calymmatobacterium) granulomatis*
Complications of pregnancy/puerperium	Several pathogens implicated
Intestinal infections	
Proctitis	*C. trachomatis, N. gonorrhoeae,* HSV, *T. pallidum*
Proctocolitis or enterocolitis	*Campylobacter* spp., *Shigella* spp., *Entamoeba histolytica, Helicobacter* spp., other enteric pathogens
Enteritis	*Giardia lamblia*
Acute arthritis with urogenital infection or viremia	*N. gonorrhoeae* (e.g., DGI), *C. trachomatis* (e.g., reactive arthritis), HBV
Genital and anal warts	HPV (30 genital types)
Mononucleosis syndrome	CMV, HIV, EBV
Hepatitis	Hepatitis viruses, *T. pallidum,* CMV, EBV
Neoplasias	
Squamous cell dysplasias and cancers of the cervix, anus, vulva, vagina, or penis	HPV (especially types 16, 18, 31, 45)
Kaposi's sarcoma, body-cavity lymphomas	HHV-8
T-cell leukemia	HTLV-1
Hepatocellular carcinoma	HBV
Tropical spastic paraparesis	HTLV-1
Scabies	*Sarcoptes scabiei*
Pubic lice	*Pthirus pubis*

Abbreviations: BV, bacterial vaginosis; CMV, cytomegalovirus; DGI, disseminated gonococcal infection; EBV, Epstein-Barr virus; HBV, hepatitis B virus; HHV-8, human herpesvirus type 8; HPV, human papillomavirus; HSV, herpes simplex virus; HTLV, human T-cell lymphotropic virus; LGV, lymphogranuloma venereum.

MANAGEMENT OF COMMON SEXUALLY TRANSMITTED DISEASE SYNDROMES

Although other chapters discuss management of specific STIs, most patients are managed (at least initially) on the basis of presenting symptoms and signs and associated risk factors, even in industrialized countries. **Table 141-2** lists some of the most common clinical sexually transmitted disease (STD) syndromes and their microbial etiologies. Strategies for their management are outlined below. **Chapters 207 and 208 address the management of infections with human retroviruses.**

STD care and management begin with risk assessment and proceed to clinical assessment, diagnostic testing or screening, treatment, and prevention. Risk assessment guides detection and interpretation of symptoms that could denote an STD; decisions on screening or prophylactic/preventive treatment; risk reduction counseling and intervention (e.g., hepatitis B vaccination); treatment of partners of patients with known infections; and behavioral risk reduction by the patient. Consideration of routine demographic data (e.g., identified gender, age, area of residence) is a simple first step in this risk assessment. For example, national guidelines strongly recommend routine screening of

TABLE 141-3 Eleven-Question Sexually Transmitted Disease (STD)/HIV Risk Assessment

<table>
<tr><th colspan="2">Framing Statement</th></tr>
<tr><td colspan="2">In order to provide the best care for you today and to understand whether we should consider certain infections, I'd like to talk about your sexual behavior.</td></tr>
<tr><th colspan="2">Screening Questions</th></tr>
<tr><td colspan="2">(1) Do you have any reason to think you might have a sexually transmitted infection? If so, what reason?
(2) For all adolescents <18 years old: Have you begun having any kind of sex yet?</td></tr>
<tr><th colspan="2">STD History</th></tr>
<tr><td colspan="2">(3) Have you ever had any sexually transmitted infections or any genital infections? If so, which ones?</td></tr>
<tr><th colspan="2">Sexual Preference</th></tr>
<tr><td colspan="2">(4) Have you had sex with men, women, or both?</td></tr>
<tr><th colspan="2">Injection Drug Use</th></tr>
<tr><td colspan="2">(5) Have you ever injected yourself with drugs? (If yes, have you ever shared needles or injection equipment?)
(6) Have you ever had sex with anyone who had ever injected drugs?</td></tr>
<tr><th colspan="2">Characteristics of Partner(s)</th></tr>
<tr><td colspan="2">(7) Have any of your sex partners had any sexually transmitted infections? If so, which ones?
(8) Have any of your sex partners had other sex partners during the time you've been together?</td></tr>
<tr><th colspan="2">STD Symptoms Checklist</th></tr>
<tr><td colspan="2">(9) Have you recently developed any of these symptoms?</td></tr>
<tr><th>For Men</th><th>For Women</th></tr>
<tr><td>(a) Discharge of pus (drip) from the penis
(b) Genital sores (ulcers) or rash</td><td>(a) Abnormal vaginal discharge (increased amount, abnormal odor, abnormal yellow color)
(b) Genital sores (ulcers), rash, or itching</td></tr>
<tr><th colspan="2">Sexual Practices, Past 2 Months (for patients answering yes to any of the above questions, to guide examination and testing)</th></tr>
<tr><td colspan="2">(10) Now I'd like to ask what parts of your body may have been sexually exposed to an STD (e.g., your penis, mouth, vagina, anus).</td></tr>
<tr><th colspan="2">Query About Interest in STD Screening Tests (for patients answering no to all of the above questions)</th></tr>
<tr><td colspan="2">(11) Would you like to be tested for HIV or any other STDs today? (If yes, clinician can explore which STD and why.)</td></tr>
</table>

Source: Adapted from JR Curtis, KK Holmes, in KK Holmes et al (eds): *Sexually Transmitted Diseases,* 4th ed. New York, McGraw-Hill, 2008.

sexually active females ≤25 years of age for *C. trachomatis* infection. Table 141-3 provides a set of 11 STD/HIV risk-assessment questions that clinicians can pose verbally or that health care systems can adapt (with yes/no responses) into a routine self-administered questionnaire. The initial framing statement gives permission to discuss topics that may be difficult for the patient to disclose.

Risk assessment is followed by clinical assessment (elicitation of information on specific current symptoms and signs of STDs). Confirmatory diagnostic tests (for persons with symptoms or signs) or screening tests (for those without symptoms or signs) may involve microscopic examination, culture, nucleic acid amplification tests (NAATs), or serology. Initial syndrome-based treatment should cover the most likely causes. For certain syndromes, results of rapid tests can narrow the spectrum of this initial therapy (e.g., pH of vaginal fluid for women with vaginal discharge, Gram's stain of urethral discharge for men with urethral discharge, rapid plasma reagin test for genital ulcer to assess the probability of syphilis). After the institution of treatment, STD management proceeds to the "4 Cs" of prevention and control: *c*ontact tracing (see "Prevention and Control of STIs," below), ensuring *c*ompliance with therapy, and *c*ounseling on risk reduction, including *c*ondom promotion and provision as well as motivational interviewing for risk reduction.

Consistent with current guidelines, all adults should be screened for infection with HIV-1 at least once, and more frequently if they are at elevated risk for acquisition of this infection.

URETHRITIS IN MEN

Urethritis in men produces urethral discharge, dysuria, or both, usually without frequency of urination. Causes include *N. gonorrhoeae, C. trachomatis, Mycoplasma genitalium, Ureaplasma urealyticum, Trichomonas vaginalis*, HSV, and (rarely) adenovirus.

Until recently, *C. trachomatis* caused ~30–40% of cases of nongonococcal urethritis (NGU), particularly in heterosexual men; however, the proportion of cases due to this organism has probably declined in some populations served by effective chlamydial control programs, and older men with urethritis appear less likely to have chlamydial infection. HSV and *T. vaginalis* each cause a small proportion of NGU cases in the United States. Recently, multiple studies have consistently implicated *M. genitalium* as a probable cause of many *Chlamydia*-negative cases. Fewer studies than in the past have implicated *Ureaplasma*; the ureaplasmas have been differentiated into *U. urealyticum* and *Ureaplasma parvum*, and a few studies suggest that *U. urealyticum*—but not *U. parvum*—is associated with NGU; for this reason, neither testing nor presumptive treatment for ureaplasmas in the setting of urethritis is recommended. Coliform bacteria can cause urethritis in men who engage in insertive anal intercourse. More recently, anaerobic bacteria that are characteristically involved in BV, especially *Leptotrichia*/*Sneathia* species, have occasionally been associated with urethritis in heterosexual men. Recommendations for the initial diagnosis of urethritis in men currently include specific tests only for *N. gonorrhoeae* and *C. trachomatis*; they do not yet include testing for *M. genitalium*, although a NAAT is now commercially available for the latter.

APPROACH TO THE PATIENT

Urethritis in Men

The following summarizes the approach to the male patient with suspected urethritis:

1. *Establish the presence of urethritis*. If proximal-to-distal "milking" of the urethra does not express a purulent or mucopurulent discharge, even after the patient has not voided for several hours (or preferably overnight), a Gram's-stained smear of an anterior urethral specimen obtained by passage of a small urethrogenital swab 2–3 cm into the urethra usually reveals ≥2 neutrophils per 1000× field when urethritis is present; in gonococcal infection, such a smear usually reveals gram-negative intracellular diplococci as well. Patients with symptoms who lack objective evidence of urethritis generally do not benefit from repeated courses of antibiotics, and other etiologies of such symptoms may be considered.
2. *Evaluate for complications or alternative diagnoses*. A brief history and examination can exclude epididymitis and systemic complications, such as disseminated gonococcal infection (DGI) and reactive arthritis. Although digital examination of the prostate gland seldom contributes to the evaluation of sexually active young men with urethritis, men with dysuria who lack evidence of urethritis as well as sexually inactive men with urethritis should undergo prostate palpation, urinalysis, and urine culture to exclude bacterial prostatitis and cystitis.
3. *Evaluate for gonococcal and chlamydial infection*. An absence of typical gram-negative diplococci on Gram's-stained smear of urethral exudate containing inflammatory cells warrants a preliminary diagnosis of NGU, as this test is 98% sensitive for the diagnosis of gonococcal urethral infection. However, most men with symptoms and/or signs of urethritis are simultaneously assessed for infection with *N. gonorrhoeae* and *C. trachomatis* by NAATs of first-catch urine. The urine specimen tested should consist of the first 10–15 mL of the stream, and if possible, patients should not have voided for the prior 2 h. Culture or NAAT for *N. gonorrhoeae* may yield positive results even when

Gram's staining is negative; certain strains of *N. gonorrhoeae* can result in negative urethral Gram's stains in up to 30% of cases of urethral infection. Results of tests for gonococcal and chlamydial infection predict the patient's prognosis (with greater risk for recurrent NGU if neither chlamydiae nor gonococci are found than if either is detected) and can guide both the counseling given to the patient and the management of the patient's sexual partner(s).

4. *Treat urethritis promptly while test results are pending.*

TREATMENT

Urethritis in Men

Table 141-4 summarizes the steps in management of urethral discharge and/or dysuria in sexually active men.

In practice, if Gram's stain does not reveal gonococci, urethritis is treated with a regimen effective for NGU, such as azithromycin or doxycycline. Both are generally effective. Although azithromycin has been more effective than doxycycline for *M. genitalium* infection, the efficacy of azithromycin for treatment of *M. genitalium* is rapidly declining. Alternatives include moxifloxacin and pristinamycin, a streptogramin antibiotic available in some countries. If gonococci are demonstrated by Gram's stain or if no diagnostic tests are performed to exclude gonorrhea definitively, treatment should include parenteral cephalosporin therapy for gonorrhea **(Chap. 161)**. Doxycycline is the preferred antibiotic for treating *C. trachomatis* infection, which can cause urethral co-infection in men with gonococcal urethritis. Sexual partners with contact to the index patient in the past 60 days should also be tested for gonorrhea and chlamydial infection. Regardless of whether they are tested for these infections, however, they should receive the same regimen given to the index case. Patients with confirmed persistence or recurrence of urethritis after treatment should be re-treated with the initial regimen if they did not comply with the original treatment or were reexposed to an untreated partner. Most persistent urethritis is due to *M. genitalium*, and prompt diagnostic testing and/or treatment for *M. genitalium* is recommended.

TABLE 141-4 Management of Urethral Discharge in Men

USUAL CAUSES	USUAL INITIAL EVALUATION
Chlamydia trachomatis *Neisseria gonorrhoeae* *Mycoplasma genitalium* *Ureaplasma urealyticum* *Trichomonas vaginalis* Herpes simplex virus	Demonstration of urethral discharge or pyuria Exclusion of local or systemic complications Urethral Gram's stain to confirm urethritis, detect gram-negative diplococci Test for *N. gonorrhoeae, C. trachomatis, M. genitalium* (if indicated and available)
Initial Treatment for Patient and Partners	
Treat gonorrhea (unless excluded): Ceftriaxone (500 mg IM[a]) For persons weighing ≥150 kg, 1 gram of ceftriaxone IM should be administered	
Management of Recurrence	
Confirm objective evidence of urethritis. If patient was reexposed to untreated or new partner, repeat treatment of patient and partner. If patient was not reexposed, consider infection with *T. vaginalis*[b] or antibiotic-resistant *M. genitalium*[c], and treat accordingly (metronidazole for trichomoniasis; azithromycin for *M. genitalium* followed by moxifloxacin if needed).	

[a]Neither oral cephalosporins nor fluoroquinolones are recommended for treatment of gonorrhea in the United States because of the emergence of increasing fluoroquinolone resistance in *N. gonorrhoeae,* especially (but not only) among men who have sex with men, and the decreasing susceptibility of a small proportion of gonococci to ceftriaxone **(Fig. 136-1)**. Updates on the emergence of antimicrobial resistance in *N. gonorrhoeae* can be obtained from the Centers for Disease Control and Prevention at *http://www.cdc.gov/std.* [b]In men, the diagnosis of *T. vaginalis* infection requires nucleic acid amplification testing of a urethral swab specimen obtained before voiding. [c]*M. genitalium* is often resistant to doxycycline and azithromycin but is usually susceptible to the fluoroquinolone moxifloxacin. Moxifloxacin can be considered for treatment of refractory nongonococcal, nonchlamydial urethritis.

National and international guidelines exist for treatment of gonococcal urethritis, typically with ceftriaxone. However, consensus is still lacking on treatment of urethritis that persists after treatment and cure of gonorrhea. Ideally, the approach would involve testing for potential causes of persistent urethritis (e.g., *M. genitalium*) and antimicrobial susceptibility testing in settings and populations where antimicrobial resistance is emerging. Currently, assays are available that can detect *M. genitalium*, and some experts believe it is time to integrate such testing into STD care. If *M. genitalium* is detected, the persistent urethritis can be treated with azithromycin or moxifloxacin in light of local patterns of antimicrobial susceptibility.

In heterosexual men with a high likelihood of exposure to trichomoniasis, an intraurethral swab specimen and a first-voided urine sample should be tested for *T. vaginalis* using NAAT. Presumptive treatment with metronidazole or tinidazole (2 g by mouth in a single dose) should be given. For MSM, trichomoniasis is unlikely, and consideration of a course of moxifloxacin is warranted. Because MSM also have the highest prevalence rates of antimicrobial-resistant *N. gonorrhoeae*, this possibility, even if apparently ruled out at the initial presentation, should be kept in mind.

■ EPIDIDYMITIS

Acute epididymitis, almost always unilateral, produces pain, swelling, and tenderness of the epididymis, with or without symptoms or signs of urethritis. This condition must be differentiated from testicular torsion, tumor, and trauma. Torsion, a surgical emergency, usually occurs in the second or third decade of life and produces a sudden onset of pain, elevation of the testicle within the scrotal sac, rotation of the epididymis from a posterior to an anterior position, and absence of blood flow on Doppler ultrasound. Persistence of symptoms after a course of therapy for epididymitis suggests the possibility of testicular tumor or of a chronic granulomatous disease, such as tuberculosis. In sexually active men under age 35, acute epididymitis is caused most frequently by *C. trachomatis* and less commonly by *N. gonorrhoeae* and is usually associated with overt or subclinical urethritis. Acute epididymitis occurring in older men or following urinary tract instrumentation is usually caused by urinary pathogens. These older men usually have no urethritis but do have bacteriuria. Similarly, epididymitis in MSM who have practiced insertive rectal intercourse is often caused by Enterobacteriaceae.

TREATMENT

Epididymitis

Ceftriaxone (500 mg as a single dose IM) followed by doxycycline (100 mg by mouth twice daily for 10 days) constitutes effective treatment for epididymitis caused by *N. gonorrhoeae* or *C. trachomatis*. Neither oral cephalosporins nor fluoroquinolones are recommended for treatment of gonorrhea in the United States because of resistance in *N. gonorrhoeae*, especially (but not only) among MSM **(Fig. 141-1)**. Given rapidly escalating rates of resistance in *N. gonorrhoeae* to azithromycin, this antibiotic is no longer recommended as co-therapy with a parenteral ceftriaxone for gonorrhea. When infection with Enterobacteriaceae is suspected, oral levofloxacin (500 mg once daily for 10 days) added to parenteral ceftriaxone (500 mg IM once) is effective for syndrome-based initial treatment.

■ URETHRITIS AND THE URETHRAL SYNDROME IN WOMEN

C. trachomatis, *N. gonorrhoeae*, and occasionally HSV cause symptomatic urethritis—known as *the urethral syndrome* in women—that is characterized by "internal" dysuria (usually without urinary urgency or frequency), pyuria, and an absence of *Escherichia coli* and other uropathogens at counts of $\geq 10^2$/mL in urine. In contrast, the dysuria associated with vulvar herpes or vulvovaginal candidiasis (and perhaps with trichomoniasis) is often described as "external," being caused

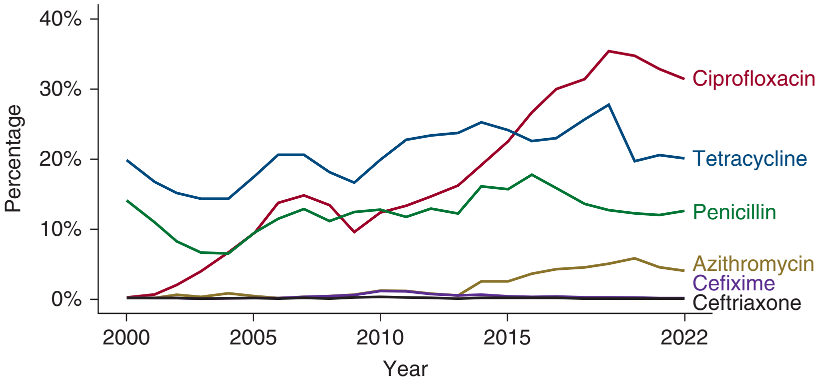

Antimicrobials	2000 n (%)	2001 n (%)	2002 n (%)	2003 n (%)	2004 n (%)	2005 n (%)	2006 n (%)	2007 n (%)	2008 n (%)	2009 n (%)	2010 n (%)
Azithromycin	19 (0.3)	15 (0.3)	33 (0.6)	26 (0.4)	57 (0.9)	35 (0.6)	14 (0.2)	27 (0.4)	11 (0.2)	12 (0.2)	27 (0.5)
Cefixime	10 (0.2)	12 (0.2)	9 (0.2)	4 (0.1)	6 (0.1)	6 (0.1)	5 (0.1)	N/A	N/A	45 (0.8)	77 (1.4)
Ceftriaxone	5 (0.1)	16 (0.3)	7 (0.1)	3 (0.0)	9 (0.1)	8 (0.1)	3 (0.0)	7 (0.1)	4 (0.1)	16 (0.3)	19 (0.3)
Ciprofloxacin	19 (0.3)	38 (0.7)	116 (2.2)	270 (4.1)	429 (6.8)	581 (9.4)	843 (13.8)	891 (14.8)	775 (13.5)	542 (9.6)	709 (12.5)
Penicillin	773 (14.2)	622 (11.4)	441 (8.2)	434 (6.6)	411 (6.5)	581 (9.4)	702 (11.5)	776 (12.9)	639 (11.2)	702 (12.5)	733 (12.9)
Tetracycline	1085 (19.9)	931 (17.0)	814 (15.2)	942 (14.4)	909 (14.4)	1073 (17.3)	1256 (20.6)	1233 (20.5)	1010 (18.2)	941 (16.7)	1149 (20.2)

Antimicrobials	2011 n (%)	2012 n (%)	2013 n (%)	2014 n (%)	2015 n (%)	2016 n (%)	2017 n (%)	2018 n (%)	2019 n (%)	2020 n (%)	2021 n (%)
Azithromycin	16 (0.3)	15 (0.3)	33 (0.6)	125 (2.5)	133 (2.6)	190 (3.6)	221 (4.4)	235 (4.6)	281 (5.1)	218 (5.8)	176 (4.6)
Cefixime	74 (1.4)	52 (0.9)	25 (0.4)	38 (0.7)	25 (0.5)	17 (0.3)	22 (0.4)	15 (0.3)	19 (0.3)	12 (0.3)	6 (0.2)
Ceftriaxone	21 (0.4)	15 (0.3)	3 (0.1)	7 (0.1)	14 (0.3)	14 (0.3)	10 (0.2)	9 (0.2)	8 (0.1)	3 (0.1)	4 (0.1)
Ciprofloxacin	726 (13.3)	809 (14.7)	955 (16.1)	978 (19.2)	1149 (22.3)	1409 (26.8)	1524 (30.1)	1611 (31.2)	1941 (35.4)	1301 (34.8)	1255 (32.8)
Penicillin	647 (11.8)	725 (13.2)	725 (12.2)	826 (16.2)	809 (15.7)	934 (17.8)	800 (15.8)	707 (13.7)	699 (12.8)	461 (12.3)	460 (12.0)
Tetracycline	1245 (22.8)	1288 (23.4)	1410 (23.7)	1287 (25.3)	1248 (24.2)	1187 (22.6)	1169 (23.1)	1322 (25.6)	1522 (27.8)	738 (19.7)	788 (20.6)

FIGURE 141-1 Percentage of Tetracycline, Penicillin, or Ciprofloxacin Resistance[*] or Elevated Cefixime, Ceftriaxone, or Azithromycin Minimum Inhibitory Concentrations (MICs)[†] by Year, Gonococcal Isolate Surveillance Project (GISP), 2001-2022.

[*] Resistance: ciprofloxacin MIC ≥1.0 µg/mL; penicillin MIC ≥2.0 µg/mL or β-lactamase positive; tetracycline MIC ≥2.0 µg/mL.

[†] Elevated MICs: azithromycin MIC ≥1.0 µg/mL (2000–2004), MIC ≥2.0 µg/mL (2005–2022); ceftriaxone MIC ≥0.125 µg/mL; cefixime MIC ≥0.25 µg/mL. Cefixime susceptibility was not tested in 2007 and 2008.

by painful contact of urine with the inflamed or ulcerated labia or introitus. Acute onset, association with urinary urgency or frequency, hematuria, or suprapubic bladder tenderness suggests bacterial cystitis. Among women with symptoms of acute bacterial cystitis, costovertebral pain and tenderness or fever suggest acute pyelonephritis. **The management of bacterial urinary tract infection (UTI) is discussed in Chap. 140.**

Signs of vulvovaginitis, coupled with symptoms of external dysuria, suggest vulvar infection (e.g., with HSV or *Candida albicans*). Among dysuric women without signs of vulvovaginitis, bacterial UTI must be differentiated from the urethral syndrome by assessment of risk, evaluation of the pattern of symptoms and signs, and specific microbiologic testing. An STI etiology of the urethral syndrome is suggested by young age, more than one current sexual partner, a new partner within the past month, a partner with urethritis, or coexisting mucopurulent cervicitis (see below). The finding of a single urinary pathogen, such as *E. coli* or *Staphylococcus saprophyticus*, at a concentration of ≥10^2/mL in a properly collected specimen of midstream urine from a dysuric woman with pyuria indicates probable bacterial UTI, whereas pyuria with <10^2 conventional uropathogens per milliliter of urine ("sterile" pyuria) suggests acute urethral syndrome due to *C. trachomatis* or *N. gonorrhoeae*. Gonorrhea and chlamydial infection should be sought by specific tests (e.g., NAATs of vaginal secretions collected with a swab). Among dysuric women with sterile pyuria caused by infection with *N. gonorrhoeae* or *C. trachomatis*, appropriate treatment alleviates dysuria. The role of *M. genitalium* in the urethral syndrome in women remains undefined.

■ VULVOVAGINAL INFECTIONS

Abnormal Vaginal Discharge If directly questioned about vaginal discharge during routine health checkups, many women acknowledge having nonspecific symptoms of vaginal discharge that do not correlate with objective signs of inflammation or with actual infection. However, unsolicited reporting of abnormal vaginal discharge often denotes BV or trichomoniasis. Specifically, an abnormally increased amount or an abnormal odor of the discharge is associated with one or both of these conditions. Cervical infection with *N. gonorrhoeae* or *C. trachomatis* does not often cause an increased amount or abnormal odor of discharge; however, when these pathogens cause cervicitis, they—like *T. vaginalis*—often result in an increased number of neutrophils in vaginal fluid, which thus takes on a yellow color. Vulvar conditions such as genital herpes or vulvovaginal candidiasis can cause vulvar pruritus, burning, irritation, or lesions as well as external dysuria (as urine passes over the inflamed vulva or areas of epithelial disruption) or vulvar dyspareunia.

Certain vulvovaginal infections may have serious sequelae. Trichomoniasis, BV, and vulvovaginal candidiasis have all been associated with increased risk of acquisition of HIV infection; BV promotes HIV transmission from HIV-infected women to their male sex partners. Vaginal trichomoniasis and BV early in pregnancy independently predict premature onset of labor. BV can also lead to anaerobic bacterial infection of the endometrium and salpinges. Vaginitis may be an early and prominent feature of toxic shock syndrome, and recurrent or chronic vulvovaginal candidiasis develops with increased frequency among women who have systemic illnesses, such as diabetes mellitus or HIV-related immunosuppression (although only a very small proportion of women with recurrent vulvovaginal candidiasis in industrialized countries actually have a serious predisposing illness).

Thus, vulvovaginal symptoms or signs warrant careful evaluation, including speculum and pelvic examination, diagnostic testing, and appropriate therapy specific for the infection identified. Unfortunately, clinicians do not always perform the tests required to establish the cause of such symptoms. Further, self-diagnosis of a specific type of infection—including vulvovaginal candidiasis—is often incorrect. The diagnosis and treatment of the three most common types of vaginal infection are summarized in Table 141-5.

Inspection of the vulva and perineum may reveal tender genital ulcerations or fissures (typically due to HSV infection or vulvovaginal candidiasis) or discharge visible at the introitus before insertion of a speculum (suggestive of BV or trichomoniasis). Speculum examination permits the clinician to discern whether the discharge appears abnormal and whether it emanates from the cervical os (mucoid and, if abnormal, yellow) or from the vagina (not mucoid, since the vaginal epithelium does not produce mucus). Symptoms or signs of abnormal vaginal discharge should prompt testing of vaginal fluid for pH, for a fishy odor when mixed with 10% KOH, and for certain microscopic

TABLE 141-5 Diagnostic Features and Management of Vaginal Infection

FEATURE	NORMAL VAGINAL EXAMINATION	VULVOVAGINAL CANDIDIASIS	TRICHOMONAL VAGINITIS	BACTERIAL VAGINOSIS (BV)
Etiology	Uninfected; lactobacilli predominant	*Candida albicans*	*Trichomonas vaginalis*	Associated with *Gardnerella vaginalis*, various anaerobic bacteria, and mycoplasmas
Typical symptoms	None	Vulvar itching and/or irritation	Profuse discharge; vulvar itching	Malodorous, slightly increased discharge
Discharge				
Amount	Variable; usually scant	Scant	Often profuse	Moderate
Color[a]	Clear or translucent	White	White or yellow	White or gray
Consistency	Nonhomogeneous, flocculent	Clumped; adherent plaques	Homogeneous	Homogeneous, low viscosity; uniformly coats vaginal walls
Inflammation of vulvar or vaginal epithelium	None	Erythema of vaginal epithelium, introitus; vulvar dermatitis, fissures common	Erythema of vaginal and vulvar epithelium; colpitis macularis	None
pH of vaginal fluid[b]	Usually ≤4.5	Usually ≤4.5	Usually ≥5	Usually >4.5
Amine ("fishy") odor with 10% KOH	None	None	May be present	Present
Microscopy[c]	Normal epithelial cells; lactobacilli predominant	Leukocytes, epithelial cells; mycelia or pseudomycelia in up to 80% of *C. albicans* culture–positive persons with typical symptoms	Leukocytes; motile trichomonads seen in 80–90% of symptomatic patients, less often in the absence of symptoms	Clue cells; few leukocytes; no lactobacilli or only a few outnumbered by profuse mixed microbiota, nearly always including *G. vaginalis* plus anaerobic species on Gram's stain (Nugent's score ≥7)
Other laboratory findings		Isolation of *Candida* spp.	Isolation of *T. vaginalis* or positive NAAT[d]	Diagnosis of BV by NAAT[d]
Usual treatment	None	Azole cream, tablet, or suppository—e.g., miconazole (100-mg vaginal suppository) or clotrimazole (100-mg vaginal tablet) once daily for 7 days Fluconazole, 150 mg orally (single dose)	Metronidazole or tinidazole, 2 g orally (single dose) Metronidazole, 500 mg PO bid for 7 days	Metronidazole, 500 mg PO bid for 7 days Metronidazole gel, 0.75%, one applicator (5 g) intravaginally once daily for 5 days Clindamycin, 2% cream, one full applicator vaginally each night for 7 days
Usual management of sexual partner	None	None; topical treatment if candidal dermatitis of penis is detected	Examination for sexually transmitted infection; treatment with metronidazole, 2 g PO (single dose)	None

[a]Color of discharge is best determined by examination against the white background of a swab. [b]A pH determination is not useful if blood is present or if the test is performed on endocervical secretions. [c]To detect fungal elements, vaginal fluid is digested with 10% KOH prior to microscopic examination; to examine for other features, fluid is mixed (1:1) with physiologic saline. Gram's stain is also excellent for detecting yeasts (less predictive of vulvovaginitis) and pseudomycelia or mycelia (strongly predictive of vulvovaginitis) and for distinguishing normal flora from the mixed flora seen in bacterial vaginosis, but it is less sensitive than the saline preparation for detection of *T. vaginalis*. [d]NAAT, nucleic acid amplification test (where available). NAAT for diagnosis of BV typically tests for combinations of BV-associated bacteria and absence of *Lactobacillus* species.

features when mixed with saline (motile trichomonads and/or "clue cells") and with 10% KOH (pseudohyphae or hyphae indicative of vulvovaginal candidiasis). Additional objective laboratory tests, described below, are useful for establishing the cause of abnormal vaginal discharge. Gram's staining of vaginal fluid can be used to characterize the vaginal bacteria using the Nugent score but is used primarily for research purposes and requires familiarity with the morphotypes and scale involved. Of note, NAATs that characterize relative concentrations of BV-associated bacteria and certain *Lactobacillus* species are now available and offer comparable performance to clinical diagnostic criteria.

TREATMENT

Vaginal Discharge

Patterns of treatment for abnormal vaginal discharge vary widely. In developing countries, where clinics or pharmacies often dispense treatment based on symptoms alone without examination or testing, oral treatment with metronidazole—particularly with a 7-day regimen—provides reasonable coverage against both trichomoniasis and BV, the usual causes of symptoms of vaginal discharge. Metronidazole treatment of sex partners prevents reinfection of women with *T. vaginalis*, although it does not help prevent the recurrence of BV. Guidelines for syndromic management promulgated by the World Health Organization suggest consideration of treatment for cervical infection and for trichomoniasis, BV, and vulvovaginal candidiasis in women with symptoms of abnormal vaginal discharge. However, it is important to note that the majority of chlamydial and gonococcal cervical infections produce no symptoms.

In industrialized countries, clinicians treating symptoms and signs of abnormal vaginal discharge should, at a minimum, differentiate between BV and trichomoniasis because optimal management of patients and partners differs for these two conditions.

Vaginal Trichomoniasis (See also Chap. 236) Symptomatic trichomoniasis characteristically produces a profuse, yellow, purulent, homogeneous vaginal discharge and vulvar irritation, sometimes with visible inflammation of the vaginal and vulvar epithelium and petechial lesions on the cervix (the so-called strawberry cervix, best visualized by colposcopy). The pH of vaginal fluid—normally <4.7—usually rises to ≥5. Microscopic examination of vaginal discharge mixed with saline reveals motile trichomonads in most culture-positive cases. However, saline microscopy detects fewer than one-half of all cases, and, especially in the absence of symptoms or signs, culture or NAAT is usually required for detection of the organism. NAAT for *T. vaginalis* is more sensitive than culture. Treatment of asymptomatic as well as symptomatic cases reduces rates of transmission and prevents later development of symptoms.

TREATMENT

Vaginal Trichomoniasis

Only nitroimidazoles (e.g., metronidazole and tinidazole) consistently cure trichomoniasis. A single 2-g oral dose of metronidazole has been the standard treatment for decades, but it is less effective than a weeklong course; the latter is preferred. Tinidazole has a longer half-life than metronidazole, causes fewer gastrointestinal symptoms, and may be useful in treating trichomoniasis that fails to respond to metronidazole. Treatment of sexual partners, facilitated by dispensing metronidazole to the female patient to give to her partner(s), significantly reduces both the risk of reinfection and the reservoir of infection; treating partners is the standard of care. Intravaginal treatment with 0.75% metronidazole gel is not reliable for vaginal trichomoniasis. Thus, systemic use of metronidazole is still recommended throughout pregnancy for treatment of trichomoniasis. In a large randomized trial, metronidazole treatment of trichomoniasis during pregnancy was associated with an increased frequency of perinatal morbidity. However, most studies, including randomized controlled trials, have shown no adverse effects of metronidazole use during pregnancy on preterm birth or birth defects.

Bacterial Vaginosis BV is a syndrome characterized by symptoms of vaginal malodor and increased white-gray discharge, which appears homogeneous, is low in viscosity, and uniformly covers the vaginal mucosa. BV has been associated with an increased risk of acquiring several other genital infections, including those caused by HIV, *C. trachomatis*, and *N. gonorrhoeae*. Other possible risk factors include recent unprotected vaginal intercourse, having a female sex partner, and vaginal douching. Although bacteria associated with BV have been detected under the foreskin of uncircumcised men and have been associated with urethritis, metronidazole treatment of male partners has not reduced the rate of recurrence of BV among affected women.

Among women with BV, culture of vaginal fluid has shown markedly increased prevalences and concentrations of *Gardnerella vaginalis*, *Mycoplasma hominis*, and several anaerobic bacteria (e.g., *Mobiluncus*, *Prevotella* [formerly *Bacteroides*], and some *Peptostreptococcus* species) as well as an absence of hydrogen peroxide–producing *Lactobacillus* species that constitute most of the normal vaginal microbiota and help protect against cervical and vaginal infections. Broad-range polymerase chain reaction (PCR) amplification of 16S rDNA in vaginal fluid, with subsequent identification of specific bacterial species by various methods, has documented even greater bacterial diversity, including several unique species not previously identified in culture (Fig. 141-2) and *Atopobium vaginae*, an organism that is strongly associated with BV and is resistant to metronidazole. Other genera newly implicated in BV include *Megasphaera, Leptotrichia, Eggerthella*, and *Dialister*.

BV is conventionally diagnosed clinically with the Amsel criteria, which include any three of the following four clinical abnormalities: (1) objective signs of increased white homogeneous vaginal discharge; (2) a vaginal discharge pH of >4.5; (3) liberation of a distinct fishy odor (attributable to volatile amines such as trimethylamine) immediately after vaginal secretions are mixed with a 10% solution of KOH; and (4) microscopic demonstration of "clue cells" (vaginal epithelial cells coated with coccobacillary organisms, which have a granular appearance and indistinct borders; Fig. 141-3) on a wet mount prepared by mixing vaginal secretions with normal saline in a ratio of ~1:1. More recently, NAAT targeting the absence of *Lactobacillus crispatus* and presence of BV-associated anaerobes have offered accurate options for diagnosing BV.

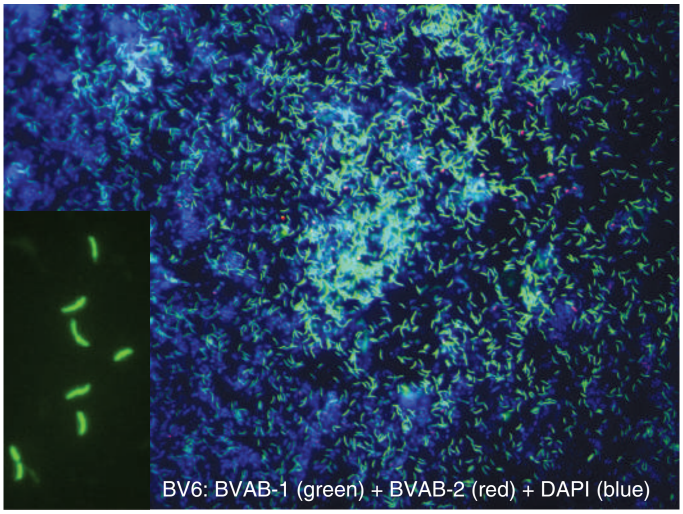

FIGURE 141-2 Broad-range polymerase chain reaction amplification of 16S rDNA in vaginal fluid from a woman with bacterial vaginosis (BV) shows a field of bacteria hybridizing with probes for BV-associated bacterium 1 (BVAB-1, *visible as a thin, curved green rod*) and for BVAB-2 (*red*). The *inset* shows that BVAB-1 has a morphology similar to that of *Mobiluncus* (curved rod). *(Adapted from DN Fredricks et al: Molecular identification of bacteria associated with bacterial vaginosis. N Engl J Med 353:1899, 2005.)*

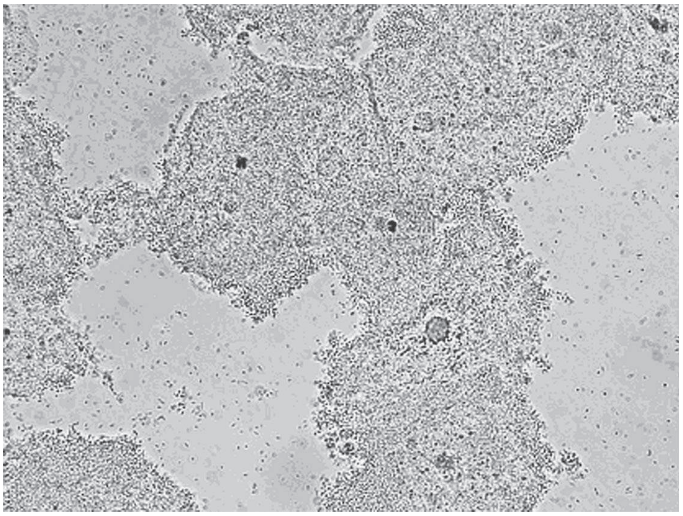

FIGURE 141-3 Wet mount of vaginal fluid showing typical clue cells from a woman with bacterial vaginosis. Note the obscured epithelial cell margins and the granular appearance attributable to many adherent bacteria (×400). *(Photograph provided by Lorna K. Rabe, reprinted with permission from S Hillier et al, in KK Holmes et al [eds]: Sexually Transmitted Diseases, 4th ed. New York, McGraw-Hill, 2008.)*

TREATMENT

Bacterial Vaginosis

The standard dosage of oral metronidazole for the treatment of BV is 500 mg twice daily for 7 days. Intravaginal treatment with 2% clindamycin cream (one full applicator [5 g containing 100 mg of clindamycin phosphate] each night for 7 nights) or with 0.75% metronidazole gel (one full applicator [5 g containing 37.5 mg of metronidazole] twice daily for 5 days) is also approved for use in the United States and does not elicit systemic adverse reactions; the response to both of these treatments is similar to the response to oral metronidazole. Another nitroimidazole given orally, secnidazole, is also effective (single 2-g dose). Other alternatives include oral clindamycin (300 mg twice daily for 7 days), clindamycin ovules (100 g intravaginally once at bedtime for 3 days), and oral tinidazole (1 g daily for 5 days or 2 g daily for 3 days). Unfortunately, recurrence over the long term (i.e., several months later) is distressingly common after either oral or intravaginal treatment. A randomized trial comparing intravaginal gel containing 37.5 mg of metronidazole with a suppository containing 500 mg of metronidazole plus nystatin (the latter not marketed in the United States) showed significantly higher rates of recurrence with the 37.5-mg regimen; this result suggests that higher metronidazole dosages may be important in topical intravaginal therapy. Recurrences can be significantly lessened with the twice-weekly use of suppressive intravaginal metronidazole gel. The goal of replenishing the vaginal lactobacilli that sustain vaginal health has recently been bolstered by a randomized trial that demonstrated that weekly vaginal administration of *L. crispatus* CTV-05 (LACTIN-V) reduced rates of recurrent BV by approximately one-third.

A meta-analysis of 18 studies concluded that BV during pregnancy substantially increased the risk of preterm delivery and of spontaneous abortion. However, in most studies, topical intravaginal treatment of BV with clindamycin during pregnancy has not reduced adverse pregnancy outcomes. Numerous trials of oral metronidazole treatment during pregnancy have given inconsistent results, and recent reviews have concluded that antenatal treatment of women with BV—including those with previous preterm delivery—did not reduce the risk of preterm delivery. The U.S. Preventive Services Task Force thus recommends against routine screening of pregnant women for BV.

Vulvovaginal Pruritus, Burning, or Irritation Vulvovaginal candidiasis produces vulvar pruritus, burning, or irritation, generally without symptoms of increased vaginal discharge or malodor. Genital herpes can produce similar symptoms, with lesions sometimes difficult to distinguish from the fissures and inflammation caused by candidiasis. Signs of vulvovaginal candidiasis include vulvar erythema, edema, fissures, and tenderness. With candidiasis, a white scanty vaginal discharge sometimes takes the form of white thrush-like plaques or cottage cheese–like curds adhering loosely to the vaginal epithelium. *C. albicans* accounts for nearly all cases of symptomatic vulvovaginal candidiasis, which probably arise from endogenous strains of *C. albicans* that have colonized the vagina or the intestinal tract. Complicated vulvovaginal candidiasis includes cases that recur four or more times per year; are unusually severe; are caused by non-*albicans Candida* species; or occur in women with uncontrolled diabetes, debilitation, immunosuppression, or pregnancy.

In addition to compatible clinical symptoms, the diagnosis of vulvovaginal candidiasis involves the demonstration of pseudohyphae or hyphae by microscopic examination of vaginal fluid mixed with saline or 10% KOH or subjected to Gram's staining. Microscopic examination is less sensitive than culture but correlates better with symptoms. Culture is typically reserved for cases that do not respond to standard first-line antimycotic agents and is undertaken to rule out imidazole or azole resistance (often associated with *Candida glabrata*) or before the initiation of suppressive antifungal therapy for recurrent disease.

TREATMENT

Vulvovaginal Pruritus, Burning, or Irritation

Symptoms and signs of vulvovaginal candidiasis warrant treatment, usually intravaginal administration of any of several imidazole antibiotics (e.g., miconazole or clotrimazole) for 3–7 days or of a single dose of oral fluconazole (Table 141-5). Over-the-counter marketing of such preparations has reduced the cost of care and made treatment more convenient for many women with recurrent yeast vulvovaginitis. However, most women who purchase these preparations do not have vulvovaginal candidiasis, whereas many have other vaginal infections that require different treatment. Therefore, only women with classic symptoms of vulvar pruritus and a history of previous episodes of yeast vulvovaginitis documented by an experienced clinician should self-treat. Short-course topical intravaginal azole drugs are effective for the treatment of uncomplicated vulvovaginal candidiasis (e.g., clotrimazole, two 100-mg vaginal tablets daily for 3 days; or miconazole, a 1200-mg vaginal suppository as a single dose). Single-dose oral treatment with fluconazole (150 mg) is also effective and is preferred by many patients. Management of complicated cases (see above) and those that do not respond to the usual intravaginal or single-dose oral therapy often involves prolonged or periodic oral therapy; this situation is discussed extensively in the 2015 CDC STD treatment guidelines (*https://www.cdc.gov/sti/hcp/clinical-guidance/?CDC_AAref_Val=https://www.cdc.gov/std/treatment/*). Treatment of sexual partners is not routinely indicated.

Other Causes of Vaginal Discharge or Vaginitis In the ulcerative vaginitis associated with staphylococcal toxic shock syndrome, *Staphylococcus aureus* should be promptly identified in vaginal fluid by Gram's stain and by culture. In desquamative inflammatory vaginitis, smears of vaginal fluid reveal neutrophils, massive vaginal epithelial cell exfoliation with increased numbers of parabasal cells, and gram-positive cocci; this syndrome may respond to treatment with 2% clindamycin cream, often given in combination with topical steroid preparations for several weeks. Additional causes of vaginitis and vulvovaginal symptoms include retained foreign bodies (e.g., tampons), cervical caps, vaginal spermicides, vaginal antiseptic preparations or douches, vaginal epithelial atrophy (in postmenopausal women or during prolonged breast-feeding in the postpartum period), allergic reactions to latex condoms, vaginal aphthae associated with HIV infection or Behçet's syndrome, and vestibulitis.

■ MUCOPURULENT CERVICITIS

Mucopurulent cervicitis (MPC) refers to inflammation of the columnar epithelium and subepithelium of the endocervix and of any contiguous columnar epithelium that lies exposed in an ectopic position on the ectocervix. MPC in women represents the "silent partner" of urethritis in men, being equally common and often caused by the same agents (*N. gonorrhoeae*, *C. trachomatis*, *M. genitalium*); however, MPC is more difficult than urethritis to recognize, given the nonspecific nature of symptoms (e.g., abnormal vaginal discharge) and the need for visualization by pelvic examination. As the most common manifestation of these serious bacterial infections in women, MPC can be a harbinger or sign of upper genital tract infection, also known as *pelvic inflammatory disease* (PID; see below). In pregnant women, MPC can lead to obstetric complications. In the pre-NAAT era, more than one-third of cervicovaginal specimens tested for *C. trachomatis, N. gonorrhoeae, M. genitalium*, HSV, and *T. vaginalis* revealed no identifiable etiology for MPC **(Fig. 141-4)**. More recent studies employing NAATs for these pathogens have still failed to identify a microbiologic etiology in nearly one-half of women with MPC. Individual bacteria associated with BV may also elicit an inflammatory reaction at the cervix; thus, BV may be a cause of MPC.

The diagnosis of MPC rests on the detection of cardinal signs at the cervix, including yellow mucopurulent discharge from the cervical os, endocervical bleeding upon gentle swabbing, and edematous cervical ectopy (see below); the latter two findings are somewhat more common with MPC due to chlamydial infection, but signs alone do not allow a distinction among the causative pathogens. Unlike the endocervicitis produced by gonococcal or chlamydial infection, cervicitis caused by HSV produces ulcerative lesions on the stratified squamous epithelium of the ectocervix as well as on the columnar epithelium. Yellow cervical mucus on a white swab removed from the endocervix indicates the presence of polymorphonuclear leukocytes (PMNs). Gram's staining may confirm their presence, although it adds relatively little to the diagnostic value of assessment for cervical signs. The presence of ≥20 PMNs per 1000× microscopic field within strands of cervical mucus not contaminated by vaginal squamous epithelial cells or vaginal bacteria indicates endocervicitis. Detection of intracellular gram-negative diplococci in carefully collected endocervical mucus is quite specific but ≤50% sensitive for gonorrhea. Therefore, NAATs for *N. gonorrhoeae* and *C. trachomatis* are always indicated in the evaluation of MPC, as is a careful evaluation of vaginal discharge for the causes of vaginitis discussed above.

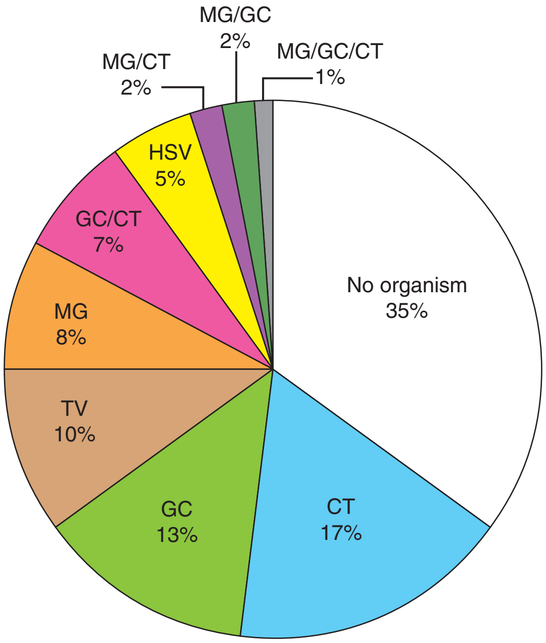

FIGURE 141-4 Organisms detected among female sexually transmitted disease clinic patients with mucopurulent cervicitis (n = 167). CT, *Chlamydia trachomatis*; GC, gonococcus; HSV, herpes simplex virus; MG, *Mycoplasma genitalium*; TV, *Trichomonas vaginalis*. *(Courtesy of Dr. Lisa Manhart.)*

TREATMENT

Mucopurulent Cervicitis

Although the above criteria for MPC are neither highly specific nor highly predictive of gonococcal or chlamydial infection in some settings, the 2015 CDC STD guidelines call for consideration of empirical treatment for MPC, pending test results, in most cases. Presumptive treatment with antibiotics active against *C. trachomatis* should be provided for women at increased risk for this common STI (risk factors: age <25 years, new or multiple sex partners, and unprotected sex), especially if follow-up cannot be ensured. Concurrent therapy for gonorrhea is indicated if the prevalence of this infection is substantial in the relevant patient population (e.g., young adults, a clinic with documented high prevalence). In this situation, therapy should include a single-dose regimen effective for gonorrhea plus treatment for chlamydial infection, as outlined in Table 141-4 for the treatment of urethritis. In settings where gonorrhea is much less common than chlamydial infection, initial therapy for chlamydial infection alone suffices, pending test results for gonorrhea. The etiology and potential benefit of treatment for endocervicitis not associated with gonorrhea or chlamydial infection have not been established. Although the antimicrobial susceptibility of *M. genitalium* is not yet well defined, the organism frequently persists after doxycycline therapy, and it currently seems reasonable to use azithromycin to treat possible *M. genitalium* infection in such cases. With resistance of *M. genitalium* to azithromycin now recognized, moxifloxacin may be a reasonable alternative. The sexual partner(s) of a woman with MPC should be examined and given a regimen similar to that chosen for the woman unless results of tests for gonorrhea or chlamydial infection in either partner warrant different therapy or no therapy.

■ CERVICAL ECTOPY

Cervical ectopy, often mislabeled "cervical erosion," is easily confused with infectious endocervicitis. Ectopy represents the presence of the one-cell-thick columnar epithelium extending from the endocervix out onto the visible ectocervix. In ectopy, the cervical os may contain clear or slightly cloudy mucus but usually not yellow mucopus. Colposcopy shows intact epithelium. Normally found during adolescence and early adulthood, ectopy gradually recedes through the second and third decades of life, as squamous metaplasia replaces the ectopic columnar epithelium. Oral contraceptive use favors the persistence or reappearance of ectopy, while smoking apparently accelerates squamous metaplasia. Cauterization of ectopy is not warranted. Ectopy may render the cervix more susceptible to infection with *N. gonorrhoeae, C. trachomatis*, or HIV.

■ PELVIC INFLAMMATORY DISEASE

The term *pelvic inflammatory disease* (PID) usually refers to infection that ascends from the cervix or vagina to involve the endometrium and/or fallopian tubes. Infection can extend beyond the reproductive tract to cause pelvic peritonitis, generalized peritonitis, perihepatitis, perisplenitis, or pelvic abscess. Rarely, infection not related to specific sexually transmitted pathogens extends secondarily to the pelvic organs (1) from adjacent foci of inflammation (e.g., appendicitis, regional ileitis, or diverticulitis) or BV, (2) as a result of hematogenous dissemination (e.g., of tuberculosis or staphylococcal bacteremia), or (3) as a complication of certain tropical diseases (e.g., schistosomiasis). Intrauterine infection can be primary (spontaneously occurring and usually sexually transmitted) or secondary to invasive intrauterine surgical procedures (e.g., dilation and curettage, termination of pregnancy, insertion of an intrauterine device [IUD], or hysterosalpingography) or to parturition.

Etiology The agents most often implicated in acute PID include the primary causes of endocervicitis (*N. gonorrhoeae, C. trachomatis*, and *M. genitalium*) and anaerobes associated with BV. In general, PID is most often caused by *N. gonorrhoeae* in settings where there is a high incidence of gonorrhea. *M. genitalium* has also been significantly associated with histopathologic diagnoses of endometritis and with salpingitis.

Anaerobic and facultative organisms (especially *Prevotella* species, peptostreptococci, *E. coli, Haemophilus influenzae*, and group B streptococci) as well as genital mycoplasmas have been isolated from the peritoneal fluid or fallopian tubes in a varying proportion (typically one-fourth to one-third) of women with PID studied in the United States. The difficulty of determining the exact microbial etiology of an individual case of PID—short of using invasive procedures for specimen collection—has implications for the approach to empirical antimicrobial treatment of this infection.

Epidemiology In the United States, the estimated annual number of initial visits to physicians' offices for PID by women 15–44 years of age fell from an average of 400,000 during the 1980s to 250,000 in 1999 and then to 51,000 in 2014. Hospitalizations for acute PID in the United States also declined steadily throughout the 1980s and early 1990s but have remained fairly constant at 70,000–100,000 per year since 1995. Important risk factors for acute PID include the presence of endocervical infection or BV, a history of salpingitis or of recent vaginal douching, and recent insertion of an IUD. Certain other iatrogenic factors, such as dilation and curettage or cesarean section, can increase the risk of PID, especially among women with endocervical gonococcal or chlamydial infection or BV. Symptoms of *N. gonorrhoeae*–associated and *C. trachomatis*–associated PID often begin during or soon after the menstrual period; this timing suggests that menstruation is a risk factor for ascending infection from the cervix and vagina. Experimental inoculation of the fallopian tubes of nonhuman primates has shown that repeated exposure to *C. trachomatis* leads to the greatest degree of tissue inflammation and damage; thus, immunopathology probably contributes to the pathogenesis of chlamydial salpingitis. Women using oral contraceptives appear to be at decreased risk of symptomatic PID, and tubal sterilization reduces the risk of salpingitis by preventing intraluminal spread of infection into the tubes.

Clinical Manifestations • **ENDOMETRITIS: A CLINICAL PATHOLOGIC SYNDROME** A study of women with clinically suspected PID who were undergoing both endometrial biopsy and laparoscopy showed that those with endometritis alone differed from those who also had salpingitis in significantly less often having lower quadrant, adnexal, or cervical motion or abdominal rebound tenderness; fever; or elevated C-reactive protein levels. In addition, women with endometritis alone differed from those with neither endometritis nor salpingitis in more often having gonorrhea, chlamydial infection, and risk factors such as douching or IUD use. Thus, women with endometritis alone were intermediate between those with neither endometritis nor salpingitis and those with salpingitis with respect to risk factors, clinical manifestations, cervical infection prevalence, and elevated C-reactive protein level. Women with endometritis alone are at lower risk of subsequent tubal occlusion and resulting infertility than are those with salpingitis.

SALPINGITIS Symptoms of nontuberculous salpingitis classically evolve from a yellow or malodorous vaginal discharge caused by MPC and/or BV to midline abdominal pain and abnormal vaginal bleeding caused by endometritis and then to bilateral lower abdominal and pelvic pain caused by salpingitis, with nausea, vomiting, and increased abdominal tenderness if peritonitis develops.

The abdominal pain in nontuberculous salpingitis is usually described as dull or aching. In some cases, pain is lacking or atypical, but active inflammatory changes are found in the course of an unrelated evaluation or procedure, such as a laparoscopic evaluation for infertility. Abnormal uterine bleeding precedes or coincides with the onset of pain in ~40% of women with PID, symptoms of urethritis (dysuria) occur in 20%, and symptoms of proctitis (anorectal pain, tenesmus, and rectal discharge or bleeding) are occasionally seen in women with gonococcal or chlamydial infection.

Speculum examination shows evidence of MPC (yellow endocervical discharge, easily induced endocervical bleeding) in the majority of women with gonococcal or chlamydial PID. Cervical motion tenderness is produced by stretching of the adnexal attachments on the side toward which the cervix is pushed. Bimanual examination reveals uterine fundal tenderness due to endometritis and abnormal adnexal tenderness due to salpingitis that is usually, but not necessarily, bilateral. Adnexal swelling is palpable in about one-half of women with acute salpingitis, but evaluation of the adnexae in a patient with marked tenderness is not reliable. The initial temperature is >38°C in only about one-third of patients with acute salpingitis. Laboratory findings include elevation of the erythrocyte sedimentation rate (ESR) in 75% of patients with acute salpingitis and elevation of the peripheral white blood cell count in up to 60%.

Unlike nontuberculous salpingitis, genital tuberculosis often occurs in older women, many of whom are postmenopausal. Presenting symptoms include abnormal vaginal bleeding, pain (including dysmenorrhea), and infertility. About one-quarter of these women have had adnexal masses. Endometrial biopsy shows tuberculous granulomas and provides optimal specimens for culture.

PERIHEPATITIS AND PERIAPPENDICITIS Pleuritic upper abdominal pain and tenderness, usually localized to the right upper quadrant (RUQ), develop in 3–10% of women with acute PID. Symptoms of perihepatitis arise during or after the onset of symptoms of PID and may overshadow lower abdominal symptoms, thereby leading to a mistaken diagnosis of cholecystitis. In perhaps 5% of cases of acute salpingitis, early laparoscopy reveals perihepatic inflammation ranging from edema and erythema of the liver capsule to exudate with fibrinous adhesions between the visceral and parietal peritoneum. When treatment is delayed and laparoscopy is performed late, dense "violin-string" adhesions can be seen over the liver; chronic exertional or positional RUQ pain ensues when traction is placed on the adhesions. Although perihepatitis, also known as the *Fitz-Hugh–Curtis syndrome*, was for many years specifically attributed to gonococcal salpingitis, most cases are now attributed to chlamydial salpingitis. In patients with chlamydial salpingitis, serum titers of microimmunofluorescent antibody to *C. trachomatis* are typically much higher when perihepatitis is present than when it is absent.

Physical findings include RUQ tenderness and usually include adnexal tenderness and cervicitis, even in patients whose symptoms do not suggest salpingitis. Results of liver function tests and RUQ ultrasonography are nearly always normal. The presence of MPC and pelvic tenderness in a young woman with subacute pleuritic RUQ pain and normal ultrasonography of the gallbladder points to a diagnosis of perihepatitis.

Periappendicitis (appendiceal serositis without involvement of the intestinal mucosa) has been found in ~5% of patients undergoing appendectomy for suspected appendicitis and can occur as a complication of gonococcal or chlamydial salpingitis.

Among women with salpingitis, HIV infection is associated with increased severity of salpingitis and with tuboovarian abscess requiring hospitalization and surgical drainage. Nonetheless, among women with HIV infection and salpingitis, the clinical response to conventional antimicrobial therapy (coupled with drainage of tuboovarian abscess, when found) has usually been satisfactory.

Diagnosis Treatment appropriate for PID must not be withheld from patients who have an equivocal diagnosis; it is better to err on the side of overdiagnosis and overtreatment. On the other hand, it is essential to differentiate between salpingitis and other pelvic pathology, particularly surgical emergencies such as appendicitis and ectopic pregnancy or the chronic syndrome of endometriosis.

Nothing short of laparoscopy definitively identifies salpingitis, but routine laparoscopy to confirm suspected salpingitis is generally impractical. Most patients with acute PID have lower abdominal pain of <3 weeks' duration, pelvic tenderness on bimanual pelvic examination, and evidence of lower genital tract infection (e.g., MPC). Approximately 60% of such patients have salpingitis at laparoscopy, and perhaps 10–20% have endometritis alone. Among the patients with these findings, a rectal temperature >38°C, a palpable adnexal mass, and elevation of the ESR to >15 mm/h also raise the probability of salpingitis, which has been found at laparoscopy in 68% of patients with one of these additional findings, 90% of patients with two, and 96% of patients with three. However, only 17% of all patients with laparoscopy-confirmed salpingitis have had all three additional findings.

In a woman with pelvic pain and tenderness, increased numbers of PMNs (30 per 1000× microscopic field in strands of cervical mucus) or leukocytes outnumbering epithelial cells in vaginal fluid (in the absence of trichomonal vaginitis, which also produces PMNs in vaginal discharge) increase the predictive value of a clinical diagnosis of acute PID, as do onset with menses, history of recent abnormal menstrual bleeding, presence of an IUD, history of salpingitis, and sexual exposure to a male with urethritis. Appendicitis or another disorder of the gut is favored by the early onset of anorexia, nausea, or vomiting; the onset of pain later than day 14 of the menstrual cycle; or unilateral pain limited to the right or left lower quadrant. Whenever the diagnosis of PID is being considered, serum assays for human β-chorionic gonadotropin should be performed; these tests are usually positive with ectopic pregnancy. Ultrasonography and magnetic resonance imaging (MRI) can be useful for the identification of tuboovarian or pelvic abscess. MRI of the tubes can also show increased tubal diameter, intratubal fluid, or tubal wall thickening in cases of salpingitis.

The primary value of laparoscopy in women with lower abdominal pain is for the exclusion of other surgical problems that cannot be resolved with noninvasive imaging. Some of the most common or serious problems that may be confused with salpingitis (e.g., acute appendicitis, ectopic pregnancy, corpus luteum bleeding, ovarian tumor) are unilateral. Unilateral pain or pelvic mass, although not incompatible with PID, is a strong indication for laparoscopy unless the clinical picture warrants laparotomy instead. Atypical clinical findings such as the absence of lower genital tract infection, a missed menstrual period, a positive pregnancy test, or failure to respond to appropriate therapy are other common indications for laparoscopy. Endometrial biopsy is relatively sensitive and specific for the diagnosis of endometritis, which correlates well with the presence of salpingitis.

Vaginal or endocervical swab specimens should be obtained for NAATs for *N. gonorrhoeae* and *C. trachomatis*. At a minimum, vaginal fluid should be evaluated for the presence of PMNs, and endocervical secretions ideally should be assessed by Gram's staining for PMNs and gram-negative diplococci, which indicate gonococcal infection. The clinical diagnosis of PID made by expert gynecologists is confirmed by laparoscopy or endometrial biopsy in ~90% of women who also have cultures positive for *N. gonorrhoeae* or *C. trachomatis*. Even among women with no symptoms suggestive of acute PID who were attending an STD clinic or a gynecology clinic in Pittsburgh, endometritis was significantly associated with endocervical gonorrhea or chlamydial infection or with BV, being detected in 26%, 27%, and 15% of women with these conditions, respectively.

TREATMENT

Pelvic Inflammatory Disease

Recommended combination regimens for ambulatory or parenteral management of PID are presented in **Table 141-6**. Women managed as outpatients should receive a combined regimen with broad activity, such as ceftriaxone (to cover possible gonococcal infection) followed by doxycycline (to cover possible chlamydial infection). Metronidazole should be added to enhance activity against anaerobes; in a randomized trial, the addition of metronidazole to ceftriaxone and doxycycline effected reduction in endometrial anaerobes, *M. genitalium,* and pelvic tenderness.

The CDC STD treatment guidelines recommend initiation of empirical treatment for PID in sexually active young women and other women at risk for PID if they are experiencing pelvic or lower abdominal pain, if no other cause for the pain can be identified, and if pelvic examination reveals one or more of the following criteria for PID: cervical motion tenderness, uterine tenderness, or adnexal tenderness. Women with suspected PID can be treated as either outpatients or inpatients. In the multicenter Pelvic Inflammatory Disease Evaluation and Clinical Health (PEACH) trial, 831 women with mild to moderately severe symptoms and signs of PID were randomized to receive either inpatient treatment with IV cefoxitin and doxycycline or outpatient treatment with a single IM dose of cefoxitin plus oral doxycycline. Short-term clinical and microbiologic outcomes and long-term outcomes were equivalent in the two groups. Nonetheless, hospitalization should be considered when (1) the diagnosis is uncertain and surgical emergencies such as appendicitis and ectopic pregnancy cannot be excluded, (2) the patient is pregnant, (3) pelvic abscess is suspected, (4) severe illness or nausea and vomiting preclude outpatient management, (5) the patient has HIV infection, (6) the patient is assessed as unable to follow or tolerate an outpatient regimen, or (7) the patient has failed to respond to outpatient therapy. Some experts also prefer to hospitalize adolescents with PID for initial therapy, although younger women do as well as older women on outpatient therapy.

TABLE 141-6 Combination Antimicrobial Regimens Recommended for Outpatient Treatment or for Parenteral Treatment of Pelvic Inflammatory Disease

OUTPATIENT REGIMENS[a]	PARENTERAL REGIMENS
Ceftriaxone (500 mg IM once) ***plus*** Doxycycline (100 mg PO bid for 14 days) ***plus***[b] Metronidazole (500 mg PO bid for 14 days)	Initiate parenteral therapy with either of the following regimens; continue parenteral therapy until 48 h after clinical improvement; then change to outpatient therapy, as described in the text **Regimen A** Cefotetan (2 g IV q12h) *or* cefoxitin (2 g IV q6h) ***plus*** Doxycycline (100 mg IV or PO q12h) **Regimen B** Clindamycin (900 mg IV q8h) ***plus*** Gentamicin (loading dose of 2 mg/kg IV or IM, then maintenance dose of 1.5 mg/kg q8h)

[a] See text for discussion of options in the patient who is intolerant of cephalosporins.
[b] The addition of metronidazole is recommended particularly if bacterial vaginosis or trichomoniasis is present.

Source: Adapted from Centers for Disease Control and Prevention: MMWR Recomm Rep 70(RR-04):1, 2021.

Currently, no agents other than parenteral cephalosporins provide reliable coverage for gonococcal infection. Thus, adequate oral treatment of women with serious intolerance to cephalosporins is a challenge. If penicillins are an option, amoxicillin/clavulanic acid combined with doxycycline has elicited a short-term clinical response in one trial. Clinical trials performed outside the United States support the effectiveness of oral moxifloxacin. For women whose PID involves quinolone-resistant *N. gonorrhoeae*, treatment is uncertain but could include parenteral gentamicin or ertapenem.

For hospitalized patients, the following two parenteral regimens (Table 141-6) have given nearly identical results in a multicenter randomized trial:

1. Doxycycline plus either cefotetan or cefoxitin: Administration of these drugs should be continued by the IV route for at least 48 h after the patient's condition improves and then followed with oral doxycycline (100 mg twice daily) to complete 14 days of therapy.
2. Clindamycin plus gentamicin in patients with normal renal function: Once-daily administration of gentamicin (with combination of the total daily dose into a single daily dose) has not been evaluated in PID but has been efficacious in other serious infections and could be substituted. Treatment with these drugs should be continued for at least 48 h after the patient's condition improves and then followed with oral doxycycline (100 mg twice daily) or clindamycin (450 mg four times daily) to complete 14 days of therapy. In cases with tuboovarian abscess, clindamycin rather than doxycycline for continued therapy provides better coverage for anaerobic infection.

FOLLOW-UP

Hospitalized patients should show substantial clinical improvement within 3–5 days. Women treated as outpatients should be clinically reevaluated within 72 h. A follow-up telephone survey of women seen in an emergency department and given a prescription for 10 days of oral doxycycline for PID found that 28% never filled the prescription and 41% stopped taking the medication early (after an average of 4.1 days), often because of persistent symptoms, lack of symptoms, or side effects. Women not responding favorably to ambulatory therapy should be hospitalized for parenteral therapy and further diagnostic evaluations, including a consideration of laparoscopy. Sex partners should be evaluated and treated empirically for gonorrhea and chlamydial infection. After completion of treatment, tests for persistent or recurrent infection with *N. gonorrhoeae* or *C. trachomatis* should be performed if symptoms persist or recur or if the patient has not complied with therapy or has been reexposed to an untreated sex partner.

SURGERY

Surgery is necessary for the treatment of salpingitis only in the face of life-threatening infection (such as rupture or threatened rupture of a tuboovarian abscess) or for drainage of an abscess. Conservative surgical procedures are usually sufficient. Pelvic abscesses can often be drained by posterior colpotomy, and peritoneal lavage can be used for generalized peritonitis.

Prognosis Late sequelae include infertility due to bilateral tubal occlusion, ectopic pregnancy due to tubal scarring without occlusion, chronic pelvic pain, and recurrent salpingitis. The overall post-salpingitis risk of infertility due to tubal occlusion in a large study in Sweden was 11% after one episode of salpingitis, 23% after two episodes, and 54% after three or more episodes. A University of Washington study found a sevenfold increase in the risk of ectopic pregnancy and an eightfold increase in the rate of hysterectomy after PID.

Prevention A randomized controlled trial designed to determine whether selective screening for chlamydial infection reduces the risk of subsequent PID showed that women randomized to undergo screening had a 56% lower rate of PID over the following year than did women receiving the usual care without screening. This report helped prompt U.S. national guidelines for risk-based chlamydial screening of young women to reduce the incidence of PID and the prevalence of post-PID sequelae, while also reducing sexual transmission of *C. trachomatis*. The CDC and the U.S. Preventive Services Task Force recommend that sexually active women ≤25 years of age be screened annually for genital chlamydial infection.

ULCERATIVE GENITAL OR PERIANAL LESIONS

Genital ulceration reflects a set of important STIs, most of which sharply increase the risk of sexual acquisition and shedding of HIV. In a 1996 study of genital ulcers in 10 of the U.S. cities with the highest rates of primary syphilis, PCR testing of ulcer specimens demonstrated HSV in 62% of patients, *T. pallidum* in 13%, and *Haemophilus ducreyi* (the cause of chancroid) in 12–20%. Today, genital herpes represents an even higher proportion of genital ulcers in the United States and other industrialized countries, even with a marked increase in early syphilis.

In Asia and Africa, chancroid **(Fig. 141-5)** was once considered the most common type of genital ulcer, followed in frequency by primary syphilis and then genital herpes **(Fig. 141-6)**. With increased efforts to control chancroid and syphilis and widespread use of broad-spectrum antibiotics to treat STI-related syndromes, together with more frequent recurrences or persistence of genital herpes attributable to HIV infection, PCR testing of genital ulcers now clearly implicates genital herpes as by far the most common cause of genital ulceration in most developing countries. LGV due to *C. trachomatis* **(Fig. 141-7)** and donovanosis (granuloma inguinale, due to *Klebsiella granulomatis*; **see Fig. 178-1**) continue to cause genital ulceration in some developing countries. LGV virtually disappeared in industrialized countries during the first 20 years of the HIV pandemic, but outbreaks are again occurring in Europe (including the United Kingdom), in North America, and in Australia. In these outbreaks, LGV typically presents as proctitis, with or without anal lesions, in men who report unprotected receptive anal intercourse, very often in association with HIV and/or hepatitis C virus infection; the latter may be an acute infection acquired through the same exposure. Other causes of genital ulcers include (1) candidiasis and traumatized genital warts; (2) lesions due to genital involvement by more widespread dermatoses; (3) cutaneous manifestations of systemic diseases such as genital mucosal ulceration in Stevens-Johnson syndrome or Behçet's disease; (4) superinfections of lesions that may originally have been sexually acquired (for example, methicillin-resistant *S. aureus* complicating a genital ulcer due to HSV-2); and (5) localized drug reactions, such as the ulcers occasionally seen with topical paromomycin cream or boric acid preparations.

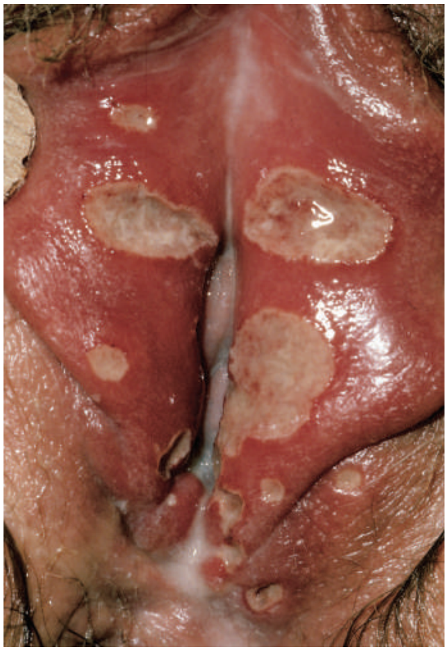

FIGURE 141-5 Chancroid: multiple, painful, punched-out ulcers with undermined borders on the labia occurring after autoinoculation.

Diagnosis Although most genital ulcerations cannot be diagnosed confidently on clinical grounds alone, clinical findings **(Table 141-7)** and epidemiologic considerations can guide initial management **(Table 141-8)** pending results of specific tests. Clinicians should order a rapid serologic test for syphilis in all cases of genital ulcer and treat presumptively while awaiting serology in a patient at increased

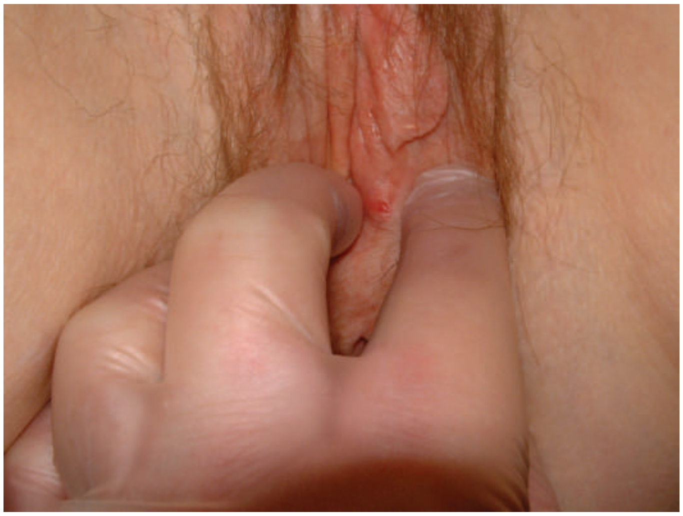

FIGURE 141-6 Genital herpes. A relatively mild, superficial ulcer is typically seen in episodic outbreaks. *(Courtesy of Michael Remington, University of Washington Virology Research Clinic.)*

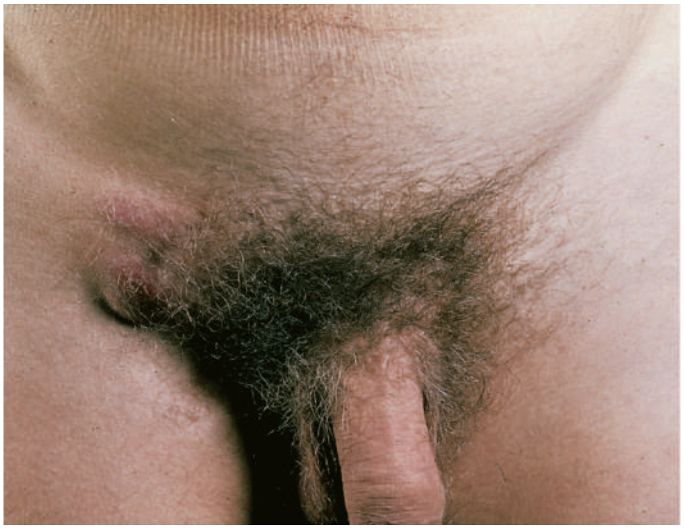

FIGURE 141-7 Lymphogranuloma venereum (LGV): striking tender lymphadenopathy occurring at the femoral and inguinal lymph nodes, separated by a groove made by Poupart's ligament. This "sign-of-the-groove" is not considered specific for LGV; for example, lymphomas may present with this sign.

epidemiologic risk (for example, MSM) or in pregnancy. To evaluate lesions except those highly characteristic of infection with HSV (i.e., those with herpetic vesicles), dark-field microscopy, direct immunofluorescence, and a NAAT for *T. pallidum* can be useful but are rarely available. It is important to note that 30% of syphilitic chancres—the primary ulcer of syphilis—are associated with an initially nonreactive syphilis serology. All patients presenting with genital ulceration should be counseled and tested for HIV infection.

Typical vesicles or pustules or a cluster of painful ulcers preceded by vesiculopustular lesions suggest genital herpes. These typical clinical manifestations make detection of the virus optional; however, many patients want confirmation of the diagnosis, and differentiation of HSV-1 from HSV-2 has prognostic implications, because the latter causes more frequent genital recurrences and is more infectious to vulnerable sex partners.

Painless, nontender, indurated ulcers with firm, nontender inguinal adenopathy suggest primary syphilis. If results of dark-field examination and a rapid serologic test for syphilis are initially negative, or if these tests are not available, presumptive therapy should be provided on the basis of the individual's risk. With historically high rates of syphilis among MSM in the United States, therapy for this infection should not be withheld pending watchful waiting and/or subsequent detection of seroconversion. Repeated serologic testing for syphilis 1 or 2 weeks after treatment of seronegative primary syphilis usually demonstrates seroconversion.

"Atypical" or clinically trivial ulcers may be more common manifestations of genital herpes than classic vesiculopustular lesions. Specific tests for HSV in such lesions are therefore indicated **(Chap. 197)**. Commercially available type-specific serologic tests for serum antibody to HSV-2 may give negative results, especially when patients present early with the initial episode of genital herpes or when HSV-1 is the cause of genital herpes. Furthermore, a positive test for antibody to HSV-2 does not prove that the current lesions are herpetic because nearly one-fifth of the general population of the United States becomes seropositive for HSV-2 during early adulthood. Although even "type-specific" tests for HSV-2 that are commercially available in the United States are not 100% specific, a positive HSV-2 serology does enable the clinician to tell the patient that they have probably had genital herpes, should learn to recognize symptoms, and should avoid sex during recurrences. In addition, because genital shedding and sexual transmission of HSV-2 often occur in the absence of symptoms and signs of recurrent herpetic lesions, persons who have a history of genital herpes or who are seropositive for HSV-2 should consider disclosure of serostatus to partners and the use of condoms or suppressive antiviral therapy, both of which can reduce the risk of HSV-2 transmission to a sexual partner.

Demonstration of *H. ducreyi* by culture (or by PCR, where available) is most useful when ulcers are painful and purulent, especially if inguinal lymphadenopathy with fluctuance or overlying erythema is noted; if chancroid is prevalent in the community; or if the patient has recently had a sexual exposure elsewhere in a chancroid-endemic area (e.g., a developing country). Enlarged, fluctuant lymph nodes should be aspirated for culture or PCR to detect *H. ducreyi* as well as for Gram's staining and culture to rule out the presence of other pyogenic bacteria.

When genital ulcers persist beyond the natural history of initial episodes of herpes (2–3 weeks) or of chancroid or syphilis (up to 6 weeks) and do not resolve with syndrome-based antimicrobial therapy, then—in addition to the usual tests for herpes, syphilis, and chancroid—biopsy is indicated to exclude donovanosis as well as carcinoma and other nonvenereal dermatoses.

TREATMENT

Ulcerative Genital or Perianal Lesions

Immediate syndrome-based treatment for acute genital ulcer (after collection of all necessary diagnostic specimens at the first visit) is often appropriate before all test results become available because patients with typical initial or recurrent episodes of genital or

TABLE 141-7 Clinical Features of Genital Ulcers

FEATURE	SYPHILIS	HERPES	CHANCROID	LYMPHOGRANULOMA VENEREUM	DONOVANOSIS
Incubation period	9–90 days	2–7 days	1–14 days	3 days–6 weeks	1–4 weeks (up to 6 months)
Early primary lesions	Papule	Vesicle	Pustule	Papule, pustule, or vesicle	Papule
Number of lesions	Usually one	Multiple	Usually multiple, may coalesce	Usually one; often not detected, despite lymphadenopathy	Variable
Diameter	5–15 mm	1–2 mm	Variable	2–10 mm	Variable
Edges	Sharply demarcated, elevated, round, or oval	Erythematous	Undermined, ragged, irregular	Elevated, round, or oval	Elevated, irregular
Depth	Superficial or deep	Superficial	Excavated	Superficial or deep	Elevated
Base	Smooth, nonpurulent, relatively nonvascular	Serous, erythematous, nonvascular	Purulent, bleeds easily	Variable, nonvascular	Red and velvety, bleeds readily
Induration	Firm	None	Soft	Occasionally firm	Firm
Pain	Uncommon	Frequently tender	Usually very tender	Variable	Uncommon
Lymphadenopathy	Firm, nontender, bilateral	Firm, tender, often bilateral with initial episode	Tender, may suppurate, loculated, usually unilateral	Tender, may suppurate, loculated, usually unilateral	None; pseudobuboes

Source: Reproduced with permission from RM Ballard, in KK Holmes et al (eds): *Sexually Transmitted Diseases,* 4th ed. New York, McGraw-Hill, 2008.

TABLE 141-8 Initial Management of Genital or Perianal Ulcer

Causative Pathogens
HSV
Treponema pallidum (primary syphilis)
Haemophilus ducreyi (chancroid)
Usual Initial Laboratory Evaluation
Dark-field examination (if available), direct FA, or PCR for *T. pallidum*
RPR, VDRL, or EIA serologic test for syphilis[a]
Culture, direct FA, ELISA, or PCR for HSV
HSV-2-specific serology (consider)
In chancroid-endemic area: PCR or culture for *H. ducreyi*
Initial Treatment
Herpes confirmed or suspected (history or sign of vesicles):
Treat for genital herpes with acyclovir, valacyclovir, or famciclovir.
Syphilis confirmed (dark-field, FA, or PCR showing *T. pallidum*, or RPR reactive):
Benzathine penicillin (2.4 million units IM once to patient, to recent [e.g., within 3 months] seronegative partner[s], and to all recent partners)[b]
Chancroid confirmed or suspected (diagnostic test positive, or HSV and syphilis excluded, and persistent lesion):
Ciprofloxacin (500 mg PO as single dose) *or*
Ceftriaxone (250 mg IM as single dose) *or*
Azithromycin (1 g PO as single dose)

[a]If results are negative but primary syphilis is suspected, treat presumptively when indicated by epidemiologic and sexual risk assessment; repeat in 1 week. [b]The same treatment regimen is also effective in HIV-infected persons with early syphilis.

Abbreviations: EIA, enzyme immunoassay; ELISA, enzyme-linked immunosorbent assay; FA, fluorescent antibody; HSV, herpes simplex virus; PCR, polymerase chain reaction; RPR, rapid plasma reagin; VDRL, Venereal Disease Research Laboratory.

anorectal herpes can benefit from prompt oral antiviral therapy **(Chap. 197)**; because early treatment of sexually transmitted causes of genital ulcers decreases further transmission; and because some patients do not return for test results and treatment. A thorough assessment of the patient's sexual-risk profile and medical history is critical in determining the course of initial management. The patient who has risk factors consistent with exposure to syphilis (e.g., a male patient who reports sex with other men or who has HIV infection) should generally receive initial treatment for syphilis. Empirical therapy for chancroid should be considered if there has been an exposure in an area of the world where chancroid occurs or if regional lymph node suppuration is evident. Finally, empirical antimicrobial therapy may be indicated if ulcers persist and the diagnosis remains unclear after a week of observation despite attempts to diagnose herpes, syphilis, and chancroid.

PROCTITIS, PROCTOCOLITIS, ENTEROCOLITIS, AND ENTERITIS

Sexually acquired *proctitis*, with inflammation limited to the rectal mucosa (the distal 10–12 cm), results from direct rectal inoculation of typical STD pathogens. In contrast, inflammation extending from the rectum to the colon (*proctocolitis*), involving both the small and the large bowel (*enterocolitis*), or involving the small bowel alone (*enteritis*) can result from ingestion of typical intestinal pathogens through oral–anal exposure during sexual contact. Anorectal pain and mucopurulent, bloody rectal discharge suggest proctitis or proctocolitis. Proctitis commonly produces tenesmus (causing frequent attempts to defecate, but not true diarrhea) and constipation, whereas proctocolitis and enterocolitis more often cause true diarrhea. In all three conditions, anoscopy usually shows mucosal exudate and easily induced mucosal bleeding (i.e., a positive "wipe test"), sometimes with petechiae or mucosal ulcers. Exudate should be sampled for Gram's staining and other microbiologic studies. Sigmoidoscopy or colonoscopy shows inflammation limited to the rectum in proctitis or disease extending at least up into the sigmoid colon in proctocolitis.

Acquisition of HSV, *N. gonorrhoeae*, or *C. trachomatis* (including LGV strains of *C. trachomatis*) during receptive anorectal intercourse causes most cases of infectious proctitis in women and MSM. Primary and secondary syphilis can also produce anal or anorectal lesions, with or without symptoms. Gonococcal or chlamydial proctitis typically involves the most distal rectal mucosa and the anal crypts and is clinically mild, without systemic manifestations. In contrast, primary proctitis due to HSV and proctocolitis due to the strains of *C. trachomatis* that cause LGV usually produce severe anorectal pain and often cause fever. Perianal ulcers and inguinal lymphadenopathy, most commonly due to HSV, can also occur with LGV or syphilis. Sacral nerve root radiculopathies, usually presenting as urinary retention, laxity of the anal sphincter, or constipation, may complicate primary herpetic proctitis. In LGV, rectal biopsy typically shows crypt abscesses, granulomas, and giant cells—findings resembling those in Crohn's disease; such findings should always prompt rectal culture and serology for LGV, which is a curable infection. Syphilis can also produce rectal granulomas, usually in association with infiltration by plasma cells or other mononuclear cells. Syphilis, LGV, and HSV infection involving the rectum can produce perirectal adenopathy that is sometimes mistaken for malignancy; syphilis, LGV, HSV infection, and chancroid involving the anus can produce inguinal adenopathy because anal lymphatics drain to inguinal lymph nodes.

Diarrhea and abdominal bloating or cramping pain without anorectal symptoms and with normal findings on anoscopy and sigmoidoscopy occur with inflammation of the small intestine (enteritis) or with proximal colitis. In MSM without HIV infection, enteritis is often attributable to *Giardia lamblia*. Sexually acquired proctocolitis is most often due to *Campylobacter* or *Shigella* species.

TREATMENT

Proctitis, Proctocolitis, Enterocolitis, and Enteritis

Acute proctitis in persons who have practiced receptive anorectal intercourse is usually sexually acquired. Such patients should undergo anoscopy to detect rectal ulcers or vesicles and petechiae after swabbing of the rectal mucosa; to examine rectal exudates for PMNs and gram-negative diplococci; and to obtain rectal swab specimens for testing for rectal gonorrhea, chlamydial infection, herpes, and syphilis. Pending test results, patients with proctitis should receive empirical syndromic treatment—e.g., with ceftriaxone (a single IM dose of 500 mg for gonorrhea) plus doxycycline (100 mg by mouth twice daily for 7 days for possible chlamydial infection) plus treatment for herpes or syphilis if indicated. If LGV proctitis is proven or suspected, the recommended treatment is doxycycline (100 mg by mouth twice daily for 21 days); alternatively, 1 g of azithromycin once a week for 3 weeks is likely to be effective but is little studied.

PREVENTION AND CONTROL OF STIs

Prevention and control of STIs require the following:

1. Reduction of the average rate of sexual exposure to STIs through alteration of sexual risk behaviors and behavioral norms among both susceptible and infected persons in all population groups. The necessary changes include reduction in the total number of sexual partners and the number of concurrent sexual partners. The U.S. Preventive Services Task Force recommends intensive behavioral counseling for all sexually active adolescents and adults who are at increased risk for STIs (grade B recommendation). Motivational interviewing is one approach that has elicited behavioral changes, including safer sex practices and more consistent contraception, that contribute to these goals.
2. Reduction of the efficiency of transmission through the promotion of safer sexual practices, the use of condoms during casual or commercial sex, vaccination against HBV and HPV infection, male circumcision (which reduces risk of acquisition of HIV infection, chancroid, and perhaps other STIs), and a growing number of other approaches (e.g., early detection and treatment of other STIs to reduce the efficiency of sexual transmission of HIV; provision of

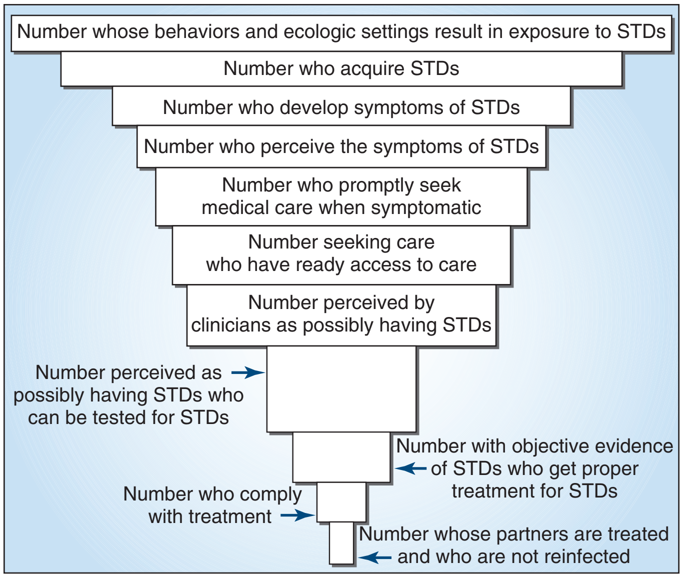
Number whose behaviors and ecologic settings result in exposure to STDs
Number who acquire STDs
Number who develop symptoms of STDs
Number who perceive the symptoms of STDs
Number who promptly seek medical care when symptomatic
Number seeking care who have ready access to care
Number perceived by clinicians as possibly having STDs
Number perceived as possibly having STDs who can be tested for STDs
Number with objective evidence of STDs who get proper treatment for STDs
Number who comply with treatment
Number whose partners are treated and who are not reinfected

FIGURE 141-8 Critical control points for preventive and clinical interventions against sexually transmitted diseases (STDs). *(Adapted from HT Waller and MA Piot: Bull World Health Organ 41:75, 1969 and 43:1, 1970; and from "Resource allocation model for public health planning—a case study of tuberculosis control," Bull World Health Organ 48[Suppl], 1973.)*

doxycycline for PEP). Among MSM and transgender women, doxycycline PEP (200 mg taken once orally within 72 h of condomless sex, with a maximum dose of 200 mg each day) can reduce the risk of chlamydia, syphilis, and, in some studies, gonorrhea. However, studies have failed to demonstrate efficacy among cisgender women.

3. Shortening of the duration of infectivity of STIs through early detection and curative or suppressive treatment of patients and their sexual partners.

Financial and time constraints imposed by many clinical practices, along with the reluctance of some clinicians to ask questions about stigmatized sexual behaviors, often curtail screening and prevention services. As outlined in **Fig. 141-8**, the success of clinicians' efforts to detect and treat STIs depends in part on societal efforts to teach young people how to recognize symptoms of STIs; to motivate individuals with symptoms to seek care promptly; to educate persons who are at risk but have no symptoms about what tests they should undergo routinely; and to make high-quality, appropriate care accessible, affordable, and acceptable, especially to the young indigent patients most likely to acquire an STI.

STI RISK ASSESSMENT

Because many infected individuals develop no symptoms or fail to recognize and report symptoms, clinicians should routinely perform an STI risk assessment for teenagers and young adults as a guide to selective screening. As stated earlier, the U.S. Preventive Services Task Force recommends screening sexually active female patients ≤25 years of age for *C. trachomatis* whenever they present for health care (at least once a year); older women should be tested if they have more than one sexual partner, have begun a new sexual relationship since the previous test, or have another STI diagnosed. In women 25–29 years of age, chlamydial infection is uncommon but still may reach a prevalence of 3–5% in some settings; information provided by women in this age group on a sex partner's concurrency (whether a male partner has had another sex partner during the time they have been together) is helpful in identifying women at increased risk. In some regions of the United States, widespread selective screening and treatment of young women for cervical *C. trachomatis* infection have been associated with a 50–60% drop in prevalence. Such screening and treatment also protect the individual woman from PID. Sensitive urine-based genetic amplification tests permit expansion of screening to men, teenage boys, and girls in settings where examination is not planned or is impractical (e.g., during preparticipation sports examinations or during initial medical evaluation of adolescent girls). Vaginal swabs—collected either by the health care provider at a pelvic examination or by the woman herself—are highly sensitive and specific for the diagnosis of chlamydial and gonococcal infection; they are now the preferred type of specimen for screening and diagnosis of these infections.

Although gonorrhea is now substantially less common than chlamydial infection in women in industrialized countries, screening tests for *N. gonorrhoeae* are still appropriate for women and teenage girls attending STD clinics and for sexually active teens and young women from areas of high gonorrhea prevalence. Multiplex NAATs that combine screening for *N. gonorrhoeae* and *C. trachomatis*—and, more recently, for *T. vaginalis*—in a single low-cost assay now facilitate the prevention and control of these infections for populations at high risk.

All patients who have newly detected STIs or are at high risk for STIs according to routine risk assessment as well as all pregnant women should be encouraged to undergo serologic testing for syphilis and HIV infection, with appropriate counseling. Randomized trials have shown that risk-reduction counseling of patients with STIs significantly lowers subsequent risk of acquiring an STI; such counseling should now be considered a standard component of STI management. Preimmunization serologic testing for antibody to HBV is indicated for unvaccinated persons who are known to be at high risk, such as MSM and people who use injection drugs. In most young persons, however, it is more cost-effective to vaccinate against HBV without serologic screening. It is important to recognize that, while immunization against HBV has contributed to marked reductions in the incidence of infection with this virus, the majority of new cases that occur are acquired through sex. In 2006, the CDC's Advisory Committee on Immunization Practices (ACIP) recommended the following: (1) Universal hepatitis B vaccination should be implemented for all unvaccinated adults in settings in which a high proportion of adults have risk factors for HBV infection (e.g., STD clinics, HIV testing and treatment facilities, drug-abuse treatment and prevention settings, health care settings targeting services to injection drug users or MSM, and correctional facilities). (2) In other primary care and specialty medical settings that provide care to adults at risk for HBV infection, health care providers should inform all patients about the health benefits of vaccination, the risk factors for HBV infection, and the persons for whom vaccination is recommended; they should vaccinate adults who report risk factors for HBV infection as well as any adult who requests protection from HBV infection. To promote vaccination in all settings, health care providers should implement standing orders to identify adults recommended for hepatitis B vaccination, should administer hepatitis B vaccine as part of routine clinical services, should not require acknowledgment of an HBV infection risk factor for adult vaccination, and should use available reimbursement mechanisms to remove financial barriers to hepatitis B vaccination.

In 2007, the ACIP made its first recommendation for routine immunization of 9- to 26-year-old girls and women with the quadrivalent HPV vaccine (against HPV types 6, 11, 16, and 18). In 2011, the ACIP recommended routine administration of quadrivalent HPV vaccine to boys at 11 or 12 years of age and to males 13–21 years of age who have not yet been vaccinated or who have not completed the three-dose vaccine series; HBV vaccination of men 22–26 years of age has also been recommended. Since that time, a nonavalent HPV vaccine has become available and has largely replaced the earlier vaccines. Moreover, single-dose HPV vaccine has great promise to simplify implementation without compromising efficacy. The optimal age for recommended vaccination is 11–12 years because of the very high risk of HPV infection after sexual debut.

Partner management is the process of identifying and informing partners of infected patients about possible exposure to an STI and of examining, testing, vaccinating, and treating partners as appropriate. In a series of 22 reports concerning partner notification during the 1990s, index patients with gonorrhea or chlamydial infection named a mean of 0.75–1.6 partners, of whom one-fourth to one-third were infected; those with syphilis named 1.8–6.3 partners, with one-third to one-half infected; and those with HIV infection named 0.76–5.31 partners, with up to one-fourth infected. Persons who transmit infection or who have recently been infected and are still in the incubation period usually have no symptoms or only mild symptoms and seek medical attention only when notified of their exposure. Therefore, the clinician must encourage patients to participate in partner notification, must ensure that exposed persons are notified and treated, and must guarantee confidentiality to all involved. In the United States, local health departments often offer assistance in partner notification, treatment, and/or counseling. It seems both feasible and most useful to notify those partners exposed within the patient's likely period of infectiousness, which is often considered the preceding 1 month for gonorrhea, 1–2 months for chlamydial infection, and up to 3 months for early syphilis.

Persons with a new-onset STI have a *source* contact who gave them the infection; in addition, they may have a *secondary* (*spread* or *exposed*) contact with whom they had sex after becoming infected. The identification and treatment of these two types of contacts have different objectives. Treatment of the source contact (often a casual contact) benefits the community by preventing further transmission and benefits the source contact; treatment of the recently exposed secondary contact (typically a spouse or another steady sexual partner) prevents the development of serious complications (such as PID) in the partner, reinfection of the index patient, and further spread of infection. A randomized trial compared patients' delivery of therapy to partners exposed to gonorrhea or chlamydial infection with conventional notification and advice to partners to seek evaluation for STD; patients' delivery of partners' therapy, also known as *expedited partner therapy* (EPT), significantly reduced combined rates of reinfection of the index patient with *N. gonorrhoeae* or *C. trachomatis*. EPT, which is now commonly used by many practicing physicians, is currently permissible in 46 states and potentially allowable in the remaining four. (Updated information on the legal status of EPT is available at *https://www.cdc.gov/sti/hcp/clinical-guidance/expedited-partner-therapy.html?CDC_AAref_Val=https://www.cdc.gov/std/ept/*.)

In summary, clinicians and public health agencies share responsibility for the prevention and control of STIs. In the current health care environment, the role of primary care clinicians has become increasingly important in STI prevention as well as in diagnosis and treatment, and the resurgence of bacterial STIs like congenital syphilis—particularly in the setting of HIV co-infection—emphasizes the need for risk assessment and routine screening.

Acknowledgment

The author wishes to acknowledge King K. Holmes, MD, PhD, for his valuable contributions to this chapter in prior editions.

FURTHER READING

Cannon CA et al: On the horizon: Novel approaches to sexually transmitted infection prevention. Med Clin North Am 108:403, 2024.

Gottlieb SL et al: Advancing vaccine development for gonorrhoea and the Global STI Vaccine Roadmap. Sex Health 16:426, 2019.

Johnston C, Corey L: Current concepts for genital herpes simplex virus infection: Diagnostics and pathogenesis of genital tract shedding. Clin Microbiol Rev 29:149, 2016.

Kirkcaldy RD et al: *Neisseria gonorrhoeae* antimicrobial resistance among men who have sex with men and men who have sex exclusively with women: The Gonococcal Isolate Surveillance Project, 2005–2010. Ann Intern Med 158:321, 2013.

Mlisana K et al: Symptomatic vaginal discharge is a poor predictor of sexually transmitted infections and genital tract inflammation in high-risk women in South Africa. J Infect Dis 206:6, 2012.

Price MJ et al: Risk of pelvic inflammatory disease following *Chlamydia trachomatis* infection: Analysis of prospective studies with a multistate model. Am J Epidemiol 178:484, 2013.

Tuddenham S et al: Diagnosis and treatment of sexually transmitted infections: A review. JAMA 327:161, 2022.

Unemo M et al: Sexually transmitted infections: Challenges ahead. Lancet Infect Dis 17:e235, 2017.

U.S. Preventive Services Task Force: Screening for chlamydia and gonorrhea: US Preventive Services Task Force Recommendation Statement. JAMA 326:949, 2021.

U.S. Preventive Services Task Force: Behavioral counseling interventions to prevent sexually transmitted infections. JAMA 324:674, 2020.

Wiesenfeld HC et al: A randomized controlled trial of ceftriaxone and doxycycline, with or without metronidazole, for the treatment of acute pelvic inflammatory disease. Clin Infect Dis 13:ciaa101, 2020.

Workowski KA, Bachmann L: Sexually transmitted disease treatment guidelines, 2021. MMWR Recomm Rep 70:1, 2021.

142 Encephalitis

Karen L. Roos, Michael R. Wilson, Kenneth L. Tyler

DEFINITION

Encephalitis is defined as an inflammation of the brain caused either by infection, usually with a virus, or from a primary autoimmune process. This chapter will focus on infectious causes of encephalitis. Many patients with encephalitis also have evidence of associated meningitis (meningoencephalitis) and, in some cases, involvement of the spinal cord or nerve roots (encephalomyelitis, encephalomyeloradiculitis).

CLINICAL MANIFESTATIONS

Similar to meningitis, encephalitis is typically an acute febrile illness. The patient with encephalitis commonly has an altered state of consciousness (confusion, behavioral abnormalities), or a depressed level of consciousness ranging from mild lethargy to coma, and evidence of either focal or diffuse neurologic signs and symptoms. Patients with encephalitis may have hallucinations, agitation, personality change, behavioral disorders, and, at times, a frankly psychotic state. Focal or generalized seizures occur in many patients with encephalitis. Virtually every possible type of focal neurologic disturbance has been reported in viral encephalitis; the signs and symptoms reflect the sites of infection and inflammation. The most commonly encountered focal findings are aphasia, ataxia, upper or lower motor neuron patterns of weakness, involuntary movements (e.g., myoclonic jerks, tremor), and cranial nerve deficits (e.g., ocular palsies, facial weakness). Involvement of the hypothalamic-pituitary axis may result in temperature dysregulation, diabetes insipidus, or the development of the syndrome of inappropriate secretion of antidiuretic hormone (SIADH). Even though neurotropic viruses typically cause injury in distinct regions of the central nervous system (CNS), variations in clinical presentations make it impossible to reliably establish the etiology of a specific case of encephalitis on clinical grounds alone (see "Differential Diagnosis," below).

ETIOLOGY

In the United States, there are an estimated ~20,000 cases of encephalitis per year, although the actual number of cases is likely to be significantly higher. Despite comprehensive diagnostic efforts, most cases of acute encephalitis with a suspected viral etiology remain of unknown cause. Hundreds of viruses are capable of causing encephalitis, although only a limited subset is responsible for most cases in which a specific cause is identified **(Table 142-1)**. The most commonly identified viruses causing sporadic cases of acute encephalitis in immunocompetent adults are herpesviruses (herpes simplex virus [HSV] **[Chap. 197]**, varicella-zoster virus [VZV] **[Chap. 198]**, and Epstein-Barr virus [EBV] **[Chap. 199]**). Epidemics of encephalitis are caused by arboviruses **(Chap. 215)**, which belong to several different viral taxonomic groups including *Alphaviruses* (e.g., eastern equine encephalitis [EEE] virus and chikungunya virus), *Flaviviruses* (e.g., West Nile virus [WNV], St. Louis encephalitis virus, Japanese encephalitis virus,

TABLE 142-1 Viruses Causing Acute Encephalitis in North America

COMMON	LESS COMMON
Herpesviruses	Rabies
Cytomegalovirus[a]	Eastern equine encephalitis virus
Herpes simplex virus 1[b]	Powassan virus
Herpes simplex virus 2	Cytomegalovirus[a]
Human herpesvirus 6	Colorado tick fever virus
Varicella-zoster virus	Mumps
Epstein-Barr virus	Jamestown Canyon virus
Arthropod-borne viruses	
La Crosse virus	
West Nile virus[c]	
St. Louis encephalitis virus	
Zika	
Enteroviruses	

[a] Immunocompromised host. [b] The most common cause of sporadic encephalitis. [c] The most common cause of epidemic encephalitis.

Powassan virus, Zika virus, dengue virus, and tick-borne encephalitis virus), and *Bunyaviruses* (e.g., California encephalitis virus serogroup, La Crosse virus, Jamestown Canyon virus). Historically, the largest number of cases of arbovirus encephalitis in the United States has been due to St. Louis encephalitis virus and the California encephalitis virus serogroup. However, since 2002, WNV has been responsible for the majority of arbovirus meningitis and encephalitis cases in the United States. WNV caused 28,684 confirmed cases of neuroinvasive disease (encephalitis, meningitis, or myelitis) in the years 1999–2022 with 2641 deaths. In 2023, there were 1599 reported cases of neuroinvasive disease (encephalitis, meningitis, acute flaccid paralysis). The majority of cases occur in August and September. It is important to recognize that WNV epidemics are unpredictable and that cases have occurred in every state in the continental United States. Since 2006, there have been increasing numbers of cases of the tick-borne Powassan virus primarily in the northeastern United States and Minnesota and Wisconsin. New causes of viral CNS infections are constantly appearing, as evidenced by multiple outbreaks of cases of encephalitis in Southeast Asia caused by Nipah virus, a member of the Paramyxoviridae family; meningitis in Europe caused by Toscana virus, an arbovirus belonging to the Bunyavirus family; neurologic disorders associated with Zika virus, a flavivirus, in South America; and neurologic disorders associated with major epidemics of chikungunya virus, a togavirus, in Africa, India, and Southeast Asia. Dengue virus is common in >100 countries worldwide with cases on the rise in the Caribbean and Puerto Rico and rare cases reported in the United States in Florida and in southern Texas. Parechoviruses including human parechovirus 3 (HPeV3), members of the Picornavirus family, have been reported as causes of fever, sepsis, and meningitis in infants (age <3 months) in the United States and abroad.

■ LABORATORY DIAGNOSIS

CSF Examination Cerebrospinal fluid (CSF) examination should be performed in all patients with suspected viral encephalitis unless contraindicated by the presence of severely increased intracranial pressure (ICP). Ideally, at least 20 mL of the initial CSF sample should be collected, with 5–10 mL stored frozen for later studies, including additional direct detection tests like virus-specific polymerase chain reaction (PCR) or metagenomic next-generation sequencing, since many neuroinvasive viruses are only transiently present in the CSF. The characteristic CSF profile is indistinguishable from that of viral meningitis (Chap. 143) and typically consists of a lymphocytic pleocytosis, a mildly elevated protein concentration, and a normal glucose concentration. A CSF pleocytosis (>5 cells/μL) occurs in >95% of immunocompetent patients with documented viral encephalitis. In rare cases, a pleocytosis may be absent on the initial lumbar puncture (LP) but present on subsequent LPs. Patients who are severely immunocompromised by HIV infection, glucocorticoid or other immunosuppressant drugs, chemotherapy, or lymphoreticular malignancies may fail to mount a CSF inflammatory response. CSF cell counts exceed 500/μL in only about 10% of patients with encephalitis. Infections with certain arboviruses (e.g., EEE virus or California encephalitis virus), mumps, and lymphocytic choriomeningitis virus (LCMV) may occasionally result in cell counts >1000/μL, but this degree of pleocytosis should suggest the possibility of nonviral infections or other inflammatory processes. Atypical lymphocytes in the CSF may be seen in EBV infection and less commonly with other viruses, including cytomegalovirus (CMV), HSV, and enteroviruses. Increased numbers of plasmacytoid or Mollaret-like large mononuclear cells have been reported in WNV encephalitis. Polymorphonuclear pleocytosis occurs in ~45% of patients with WNV encephalitis and is also a common feature in CMV myeloradiculitis in immunocompromised patients. Large numbers of CSF polymorphonuclear leukocytes may be present in patients with encephalitis due to EEE virus, echovirus 9, and, more rarely, other enteroviruses. However, persisting CSF neutrophilia should prompt consideration of bacterial infection, leptospirosis, amebic infection, and noninfectious processes such as acute hemorrhagic leukoencephalitis (Chap. 456). About 20% of patients with encephalitis will have a significant number of red blood cells (>500/μL) in the CSF in a nontraumatic tap. The pathologic correlate of this finding may be punctate microhemorrhages of the type seen with HSV; however, CSF red blood cells occur with similar frequency and in similar numbers in patients with nonherpetic focal encephalitides. A decreased CSF glucose concentration is distinctly unusual in viral encephalitis and should suggest the possibility of bacterial, fungal, tuberculous, parasitic, leptospiral, syphilitic, sarcoid, or neoplastic meningitis. Rare patients with mumps, LCMV, VZV, or advanced HSV encephalitis and many patients with CMV myeloradiculitis have low CSF glucose concentrations.

■ CSF POLYMERASE CHAIN REACTION

CSF PCR has become the primary diagnostic test for CNS infections caused by HSV, CMV, EBV, HHV-6, and enteroviruses. In the case of VZV CNS infection, CSF PCR and detection of virus-specific IgM or intrathecal antibody synthesis both provide important aids to diagnosis. The sensitivity and specificity of CSF PCRs vary with the virus being tested. The sensitivity (~96%) and specificity (~99%) of HSV CSF PCR are equivalent to or exceed those of brain biopsy. It is important to recognize that HSV CSF PCR results need to be interpreted after considering the likelihood of disease in the patient being tested, the timing of the test in relationship to onset of symptoms, and the prior use of antiviral therapy. A negative HSV CSF PCR test performed by a qualified laboratory at the appropriate time during illness in a patient with a high likelihood of HSV encephalitis based on clinical and laboratory abnormalities significantly reduces the likelihood of HSV encephalitis but does not exclude it. For example, in a patient with a pretest probability of 35% of having HSV encephalitis, a negative HSV CSF PCR reduces the posttest probability to ~2%, and for a patient with a pretest probability of 60%, a negative test reduces the posttest probability to ~6%. In both situations, a positive test makes the diagnosis almost certain (98–99%). There have been reports of initially negative HSV CSF PCR tests that were obtained early (≤72 h) following symptom onset and that became positive when repeated 1–3 days later. The frequency of positive HSV CSF PCRs in patients with herpes encephalitis also decreases as a function of the duration of illness, with only ~20% of cases remaining positive after ≥14 days. PCR results are generally not affected by ≤1 week of antiviral therapy. In one study, 98% of CSF specimens remained PCR positive during the first week of antiviral therapy, but the numbers fell to ~50% by 8–14 days and to ~21% by >15 days after initiation of antiviral therapy.

The sensitivity and specificity of CSF PCR tests for viruses other than HSV have not been definitively characterized. Enteroviral (EV) CSF RT-PCR appears to have a sensitivity and specificity of >95%. EV RT-PCR sensitivity for EV-A71 may be considerably lower (~30% in some reports). Patients with EV-D68-associated acute flaccid myelitis (AFM) only rarely have a positive CSF RT-PCR (<3%) but may have a positive test on nasopharyngeal swab specimens. Parechoviruses are also not detected by standard EV RT-PCRs. The specificity of EBV

CSF PCR has not been established. Positive EBV CSF PCRs associated with positive tests for other pathogens have been reported and may reflect reactivation of EBV latent in lymphocytes that enter the CNS as a result of an unrelated infectious or inflammatory process. In patients with CNS infection due to VZV, CSF antibody and PCR studies should be considered complementary because patients may have evidence of intrathecal synthesis of VZV-specific antibodies and negative CSF PCRs. In the case of WNV infection, CSF PCR appears to be less sensitive than detection of WNV-specific CSF IgM, although PCR testing remains useful in immunocompromised patients who may not mount an effective anti-WNV antibody response. The recent pandemic due to SARS-CoV-2 (COVID-19) has been associated with cases of encephalopathy due to the indirect effects on the nervous system of multiorgan system failure and/or to a hyperinflammatory syndrome and disseminated intravascular coagulation, but also with rare cases of true encephalitis caused by viral CNS invasion. In both sets of patients, nasopharyngeal reverse transcriptase (RT)-PCR tests for SARS-CoV-2 are positive, but only cases with encephalitis have a positive CSF RT-PCR for SARS-CoV-2. Rare cases of neuroinvasion by SARS-CoV-2 has also been detected by RT-PCR of brain tissue.

Unbiased metagenomic sequencing technologies capable of identifying infectious genomes in CSF, brain, and other tissues have recently shown great promise for rapid diagnosis of obscure cases of encephalitis and other brain infections, especially in immunocompromised patients.

CSF Culture CSF culture is generally of limited utility in the diagnosis of acute viral encephalitis. Culture may be insensitive (e.g., >95% of patients with HSV encephalitis have negative CSF cultures, as do virtually all patients with EBV-associated CNS disease) and often takes too long to significantly affect immediate therapy.

Serologic Studies and Antigen Detection For many arboviruses including WNV, serologic studies remain important diagnostic tools. Serum antibody determination is less useful for viruses with high seroprevalence rates in the general population such as HSV, VZV, CMV, and EBV. For viruses with low seroprevalence rates, diagnosis of acute viral infection can be made by documenting seroconversion between acute-phase and convalescent sera (typically obtained after 2–4 weeks) or by demonstrating the presence of virus-specific IgM antibodies. For viruses with high seroprevalence such as VZV and HSV, demonstration of synthesis of virus-specific antibodies in CSF, as shown by an increased IgG index or the presence of CSF IgM antibodies, may be useful and can provide presumptive evidence of CNS infection. Unfortunately, the delay between onset of infection and the host's generation of a virus-specific antibody response often means that serologic data are useful mainly for the retrospective establishment of a diagnosis, rather than in aiding acute diagnosis or management.

In patients with HSV encephalitis, antibodies to HSV-1 glycoproteins and HSV glycoprotein antigens have been detected in the CSF. Optimal detection of both HSV antibodies and antigen typically occurs after the first week of illness, limiting the utility of these tests in acute diagnosis. Nonetheless, HSV CSF antibody testing is of value in selected patients whose illness is >1 week in duration and who are CSF PCR negative for HSV. In the case of VZV infection, CSF IgM antibody tests may be positive when PCR fails to detect viral DNA, and both tests should be considered complementary rather than mutually exclusive.

Demonstration of CSF WNV IgM antibodies is diagnostic of WNV encephalitis because the high molecular weight of IgM antibodies restricts their passage from serum to CSF through the blood-brain barrier and their presence in CSF is therefore indicative of intrathecal synthesis. Timing of antibody testing may be important because the rate of CSF WNV IgM seropositivity increases during the first week after illness onset, reaching 80% or higher on day 7 after symptom onset. Although serum and CSF IgM antibodies generally persist for only a few months after acute infection, there are exceptions to this rule, and WNV serum IgM has been shown to persist in some patients for >1 year following acute infection.

MRI, CT, and EEG Patients with suspected encephalitis almost invariably undergo neuroimaging studies and often electroencephalogram (EEG). These tests help identify or exclude alternative diagnoses and assist in the differentiation between a focal and a diffuse encephalitic process. Specific focal findings in a patient with encephalitis should always raise the possibility of HSV encephalitis. Examples of focal findings found in HSV encephalitis include: (1) areas of increased signal intensity in the frontotemporal, cingulate, or insular regions of the brain on T2-weighted, fluid-attenuated inversion recovery (FLAIR), or diffusion-weighted magnetic resonance imaging (MRI) (Fig. 142-1); (2) focal areas of low absorption, mass effect, and contrast enhancement in frontotemporal areas on computed tomography (CT); or (3) periodic focal temporal lobe spikes on a background of slow or low-amplitude ("flattened") activity on EEG. Approximately 10% of patients with PCR-documented HSV encephalitis may have a normal MRI, although nearly 80% will have asymmetric abnormalities in the temporal lobe, and an additional 10% in extratemporal regions. The addition of FLAIR and diffusion-weighted images to the standard MRI sequences enhances sensitivity. Children with HSV encephalitis may have atypical patterns of MRI lesions and often show involvement of brain regions outside the frontotemporal areas. CT is less sensitive than MRI and is normal in up to 20–35% of patients. EEG abnormalities occur in >75% of PCR-documented cases of HSV encephalitis; they typically involve the temporal lobes but are often nonspecific. Some patients with HSV encephalitis have a distinctive EEG pattern consisting of periodic, stereotyped, sharp-and-slow complexes originating in one or both temporal lobes and repeating at regular intervals of 2–3 s. The periodic complexes are typically noted between days 2 and 15 of the illness and are present in two-thirds of pathologically proven cases of HSV encephalitis.

Significant MRI abnormalities are found in only approximately two-thirds of patients with WNV encephalitis, a frequency less than that found with HSV encephalitis. When present, abnormalities often involve deep brain structures, including the thalamus, basal ganglia, and brainstem, rather than the cortex, and may only be apparent on T2/FLAIR images. Similar MRI patterns can be observed in patients infected with other arboviruses, including other flaviviruses such as Japanese encephalitis virus and St. Louis encephalitis virus, as well the *Alphavirus* EEE virus. EEGs in patients with WNV encephalitis typically show generalized slowing that may be more anteriorly prominent rather than the temporally predominant pattern of sharp or periodic discharges more characteristic of HSV encephalitis. Patients with VZV encephalitis may show multifocal areas of hemorrhagic and ischemic infarction, reflecting the tendency of this virus to produce a CNS vasculopathy rather than a true encephalitis. Immunocompromised

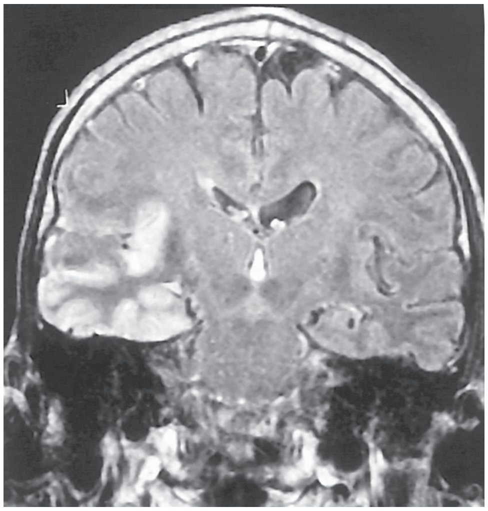

FIGURE 142-1 Coronal fluid-attenuated inversion recovery (FLAIR) magnetic resonance image from a patient with herpes simplex encephalitis. Note the area of increased signal in the right temporal lobe (*left side of image*) confined predominantly to the gray matter. This patient had predominantly unilateral disease; bilateral lesions are more common but may be quite asymmetric in their intensity.

TABLE 142-2 Use of Diagnostic Tests in Encephalitis

The best test for WNV encephalitis is the *CSF IgM antibody test.* The prevalence of positive CSF IgM tests increases by about 10% per day after illness onset and reaches 70–80% by the end of the first week. Serum WNV IgM can provide evidence for recent WNV infection, but in the absence of other findings does not establish the diagnosis of neuroinvasive disease (meningitis, encephalitis, acute flaccid paralysis).

Approximately 80% of patients with proven HSV encephalitis have *MRI* abnormalities involving the temporal lobes. This percentage likely increases to >90% when FLAIR and diffusion-weighted MRI sequences are also used. The absence of temporal lobe lesions on MRI reduces the likelihood of HSV encephalitis and should prompt consideration of other diagnostic possibilities.

The *CSF HSV PCR* test may be negative in the first 72 h of symptoms of HSV encephalitis. A repeat study should be considered in patients with an initial early negative PCR in whom diagnostic suspicion of HSV encephalitis remains high and no alternative diagnosis has yet been established.

Detection of *intrathecal synthesis* (increased CSF/serum HSV antibody ratio corrected for breakdown of the blood-brain barrier) of *HSV-specific antibody* may be useful in diagnosis of HSV encephalitis in patients in whom only late (>1 week after onset) CSF specimens are available and PCR studies are negative. Serum serology alone is of no value in diagnosis of HSV encephalitis due to the high seroprevalence rate in the general population.

Negative *CSF viral cultures* are of no value in excluding the diagnosis of HSV or EBV encephalitis.

VZV CSF IgM antibodies may be present in patients with a negative VZV CSF PCR. Both tests should be performed in patients with suspected VZV CNS disease.

The specificity of *EBV CSF PCR* for diagnosis of CNS infection is unknown. Positive tests may occur in patients with a CSF pleocytosis due to other causes. Detection of EBV CSF IgM or intrathecal synthesis of antibody to VCA supports the diagnosis of EBV encephalitis. Serologic studies consistent with acute EBV infection (e.g., IgM VCA, presence of antibodies against EA but not against EBNA) can help support the diagnosis.

In addition to broad-based PCR assays for bacterial and fungal infections, metagenomic next-generation sequencing (mNGS) allows for unbiased detection of nucleic acids from the whole range of infectious agents (except prions), which can then be confirmed by independent pathogen-specific techniques. Due to the sensitivity of this technology, there is a risk of false-positive results. As this technology becomes refined and the turnaround time faster, mNGS is likely to become a routine test on CSF for the diagnosis of encephalitis.

Abbreviations: CNS, central nervous system; CSF, cerebrospinal fluid; DWI, diffusion-weighted imaging; EA, early antigen; EBNA, EBV-associated nuclear antigen; EBV, Epstein-Barr virus; FLAIR, fluid-attenuated inversion recovery; HSV, herpes simplex virus; IgM, immunoglobulin M; MRI, magnetic resonance imaging; PCR, polymerase chain reaction; VCA, viral capsid antibody; VZV, varicella-zoster virus; WNV, West Nile virus.

adult patients with CMV often have enlarged ventricles with areas of increased T2 signal on MRI outlining the ventricles and subependymal enhancement on T1-weighted postcontrast images. Prominent cerebellar T2/FLAIR abnormalities have been observed with Powassan virus encephalitis and in children with herpesviruses like EBV and VZV. **Table 142-2** highlights specific diagnostic test results in encephalitis that can be useful in clinical decision-making.

Brain Biopsy Brain biopsy is now generally reserved for patients in whom CSF PCR studies fail to lead to a specific diagnosis and who have focal abnormalities on MRI, no serologic evidence of autoimmune disease, and continue to show progressive clinical deterioration despite treatment with acyclovir and supportive therapy.

DIFFERENTIAL DIAGNOSIS

Infection by a variety of other organisms can mimic viral encephalitis. In studies of biopsy-proven HSV encephalitis, common infectious mimics of focal viral encephalitis included mycobacteria, fungi, rickettsiae, *Listeria, Mycoplasma*, and other bacteria (including *Bartonella* sp.) as well as neurosyphilis. There are an increasing number of antibodies reported that cause autoimmune encephalitis and mimic those caused by viral infection, including those associated with antibodies against *N*-methyl-D-aspartate (NMDA) receptor, two components of the voltage-gated potassium channels/leucine-rich glioma inactivated protein-1 (LGI-1) and contracting-associated protein-like 2 (CASPR2), α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA), γ-aminobutyric acid (GABA) receptors, and glutamic acid decarboxylase (GAD 65) **(Chap. 99)**. In most cases, diagnosis is made by detection of the specific autoantibodies in serum and/or CSF. NMDA receptor antibodies have been reported in up to 25% of patients who have recovered from HSV encephalitis, and their presence should not exclude appropriate testing and treatment for HSV encephalitis. The development of NMDA receptor antibodies in patients with HSV encephalitis may contribute to new or worsening symptoms in the weeks following recovery from HSV encephalitis. Autoimmune encephalitis may also be associated with specific cancers (paraneoplastic) and onconeuronal antibodies (e.g., anti-Hu, Yo, Ma2, amphiphysin, CRMP5, CV2). Subacute or chronic forms of encephalitis may occur in association with autoantibodies against thyroglobulin and thyroperoxidase (Hashimoto's encephalopathy) and with prion diseases.

Infection caused by the ameba *Naegleria fowleri* can also cause acute meningoencephalitis (primary amebic meningoencephalitis), whereas that caused by *Acanthamoeba* and *Balamuthia* more typically produces subacute or chronic granulomatous amebic meningoencephalitis. *Naegleria* thrive in warm, iron-rich pools of water, including those found in drains, canals, and both natural and human-made outdoor pools **(Chap. 230)**. Infection has typically occurred in immunocompetent children with a history of swimming in potentially infected water. The CSF, in contrast to the typical profile seen in viral encephalitis, often resembles that of bacterial meningitis with a neutrophilic pleocytosis and hypoglycorrhachia. Motile trophozoites can be seen in a wet mount of warm, fresh CSF. There have been an increasing number of cases of *Balamuthia mandrillaris* amebic encephalitis in children and immunocompetent adults, mimicking acute viral encephalitis. This organism has also been associated with encephalitis in recipients of transplanted organs from a donor with unrecognized infection. No effective treatment has been approved, and mortality approaches 100%. There have been a few case reports of patients who recovered with nitroxoline therapy.

Encephalitis can be caused by the raccoon pinworm *Baylisascaris procyonis*. Clues to the diagnosis include a history of raccoon exposure, especially of playing in or eating dirt potentially contaminated with raccoon feces. Most patients are children, and many have an associated eosinophilia.

Once nonviral causes of encephalitis have been excluded, the major diagnostic challenge is to distinguish HSV from other viruses that cause encephalitis. This distinction is particularly important because in virtually every other instance the therapy is supportive, whereas specific and effective antiviral therapy is available for HSV, and its efficacy is enhanced when it is instituted early in the course of infection. HSV encephalitis should be considered when clinical features suggest involvement of the inferomedial frontotemporal regions of the brain, including prominent olfactory or gustatory hallucinations, anosmia, unusual or bizarre behavior or personality alterations, or memory disturbance. HSV encephalitis should always be suspected in patients with signs and symptoms consistent with acute encephalitis who have focal findings on clinical examination, neuroimaging studies, or EEG. The diagnostic procedure of choice in these patients is CSF PCR analysis for HSV. A positive CSF PCR establishes the diagnosis, and a negative test dramatically reduces the likelihood of HSV encephalitis (see above).

The anatomic distribution of lesions may provide an additional clue to diagnosis. Patients with rapidly progressive encephalitis and prominent brainstem signs, symptoms, or neuroimaging abnormalities may be infected by flaviviruses (WNV, St. Louis encephalitis virus, Japanese encephalitis virus), HSV, enterovirus A71 (EV-A71), rabies, or *Listeria monocytogenes*. Significant involvement of deep gray matter structures, including the basal ganglia and thalamus, should also suggest possible flavivirus infection. These patients may present clinically with prominent movement disorders (tremor, myoclonus) or other parkinsonian features. Patients with WNV infection can also present with a poliomyelitis-like AFM, as can patients infected with EV-A71, EV-D68, and less commonly, other enteroviruses. Acute flaccid paralysis is characterized by the acute onset of a lower motor neuron type of weakness with flaccid tone, reduced or absent reflexes, and relatively preserved sensation. Patients often have multisegmental increased FLAIR and T2 signal in the anterior horns of the spinal cord and a CSF lymphocytic pleocytosis.

Epidemiologic factors may provide important clues to the diagnosis of viral encephalitis. Particular attention should be paid to the season of the year; the geographic location and travel history; and possible exposure to animal bites or scratches, rodents, and ticks. Although transmission from the bite of an infected dog remains the most common cause of rabies **(Chap. 214)** worldwide, in the United States very few cases of dog rabies occur, and the most common risk factor is exposure to bats—although a clear history of a bite or scratch is often lacking. The classic clinical presentation of encephalitic (furious) rabies is fever, fluctuating consciousness, and autonomic hyperactivity. Phobic spasms of the larynx, pharynx, neck muscles, and diaphragm can be triggered by attempts to swallow water (*hydrophobia*) or by inspiration (*aerophobia*). Patients may also present with paralytic (dumb) rabies characterized by acute ascending paralysis. Rabies due to the bite of a bat has a different clinical presentation than classic rabies due to a dog or wolf bite. Patients present with focal neurologic deficits, myoclonus, seizures, and hallucinations; phobic spasms are not a typical feature. Patients with rabies have a CSF lymphocytic pleocytosis and may show areas of increased T2 signal abnormality in the brainstem, hippocampus, and hypothalamus. Diagnosis can be made by finding rabies virus antigen in brain tissue or in the neural innervation of hair follicles at the nape of the neck. PCR amplification of viral nucleic acid from CSF and saliva or tears may also enable diagnosis. Serology is frequently negative in both serum and CSF in the first week after onset of infection, which limits its acute diagnostic utility. No specific therapy is available, and cases are almost invariably fatal, with isolated survivors having devastating neurologic sequelae.

State public health authorities provide a valuable resource concerning isolation of particular agents in individual regions. Regular updates concerning the number, type, and distribution of cases of arboviral encephalitis can be found on the Centers for Disease Control and Prevention and U.S. Geological Survey (USGS) websites (*http://www.cdc.gov* and *http://diseasemaps.usgs.gov*).

TREATMENT

Viral Encephalitis

Specific antiviral therapy should be initiated when appropriate. Vital functions, including respiration and blood pressure, should be monitored continuously and supported as required. In the initial stages of encephalitis, many patients will require care in an intensive care unit. Basic management and supportive therapy should include careful monitoring of ICP, fluid restriction, avoidance of hypotonic intravenous solutions, and suppression of fever. Seizures should be treated with standard anticonvulsant regimens, and prophylactic therapy should be considered in view of the high frequency of seizures in severe cases of encephalitis. As with all seriously ill, immobilized patients with altered levels of consciousness, encephalitis patients are at risk for aspiration pneumonia, stasis ulcers and decubiti, contractures, deep venous thrombosis and its complications, and infections of indwelling lines and catheters.

Acyclovir is of benefit in the treatment of HSV and should be started empirically in patients with suspected viral encephalitis, especially if focal features are present, while awaiting viral diagnostic studies. Treatment should be discontinued in patients found not to have HSV encephalitis, with the possible exception of patients with severe encephalitis due to VZV or EBV. HSV, VZV, and EBV all encode an enzyme deoxypyrimidine (thymidine) kinase that phosphorylates acyclovir to produce acyclovir-5′-monophosphate. Host cell enzymes then phosphorylate this compound to form a triphosphate derivative. It is the triphosphate that acts as an antiviral agent by inhibiting viral DNA polymerase and by causing premature termination of nascent viral DNA chains. The specificity of action depends on the fact that uninfected cells do not phosphorylate significant amounts of acyclovir to acyclovir-5′-monophosphate. A second level of specificity is provided by the fact that the acyclovir triphosphate is a more potent inhibitor of viral DNA polymerase than of the analogous host cell enzymes.

Adults should receive a dose of 10 mg/kg of acyclovir intravenously every 8 h (30 mg/kg per day total dose) for 21 days. Neonatal HSV CNS infection is less responsive to acyclovir therapy than HSV encephalitis in adults; it is recommended that neonates with HSV encephalitis receive 20 mg/kg of acyclovir every 8 h (60 mg/kg per day total dose) for a minimum of 21 days.

Prior to intravenous administration, acyclovir should be diluted to a concentration ≤7 mg/mL. (A 70-kg person would receive a dose of 700 mg, which would be diluted in a volume of 100 mL.) Each dose should be infused slowly over 1 h, rather than by rapid or bolus infusion, to minimize the risk of renal dysfunction. Care should be taken to avoid extravasation or intramuscular or subcutaneous administration. The alkaline pH of acyclovir can cause local inflammation and phlebitis (9%). Dose adjustment is required in patients with impaired renal glomerular filtration. Penetration into CSF is excellent, with average drug levels ~50% of serum levels. Complications of therapy include elevations in blood urea nitrogen and creatinine levels (5%), thrombocytopenia (6%), gastrointestinal toxicity (nausea, vomiting, diarrhea) (7%), and neurotoxicity (lethargy or obtundation, disorientation, confusion, agitation, hallucinations, tremors, seizures) (1%). Acyclovir resistance may be mediated by changes in either the viral deoxypyrimidine kinase or DNA polymerase. To date, acyclovir-resistant isolates have not been a significant clinical problem in immunocompetent individuals. It is now appreciated that some patients with worsening symptoms in the weeks following recovery from HSV encephalitis have developed NMDA receptor encephalitis requiring immunosuppression rather than having developed an acyclovir-resistant isolate. However, there have been reports of clinically virulent acyclovir-resistant HSV isolates from sites outside the CNS in immunocompromised individuals, including those with AIDS.

Oral antiviral drugs with efficacy against HSV, VZV, and EBV, including acyclovir, famciclovir, and valacyclovir, have not been evaluated in the treatment of encephalitis as primary therapy. Additional oral valaciclovir following a 14- to 21-day course of intravenous acyclovir does not improve outcomes in adult patients with HSV encephalitis. The role of adjunctive intravenous glucocorticoids in treatment of HSV and VZV infection remains unclear. Experimental models and case reports of HSV encephalitis suggest that glucocorticoids may be efficacious, although no data from randomized controlled human trials are available. Ganciclovir and foscarnet, as combination therapy, are used in the treatment of CMV-related CNS infections. Cidofovir (see below) may provide an alternative in patients who fail to respond to ganciclovir and foscarnet, although data concerning its use in CMV CNS infections are extremely limited.

Ganciclovir is a synthetic nucleoside analogue of 2′-deoxyguanosine. The drug is preferentially phosphorylated by virus-induced cellular kinases. Ganciclovir triphosphate acts as a competitive inhibitor of the CMV DNA polymerase, and its incorporation into nascent viral DNA results in premature chain termination. Following intravenous administration, CSF concentrations of ganciclovir are 25–70% of coincident plasma levels. The usual dose for treatment of severe neurologic illnesses is 5 mg/kg every 12 h given intravenously at a constant rate over 1 h. Induction therapy is followed by maintenance therapy of 5 mg/kg every day for an indefinite period. Induction therapy should be continued until patients show a decline in CSF pleocytosis and a reduction in CSF CMV DNA copy number on quantitative PCR testing (where available). Doses should be adjusted in patients with renal insufficiency. Treatment is often limited by the development of granulocytopenia and thrombocytopenia (20–25%), which may require reduction in or discontinuation of therapy. Gastrointestinal side effects, including nausea, vomiting, diarrhea, and abdominal pain, occur in ~20% of patients. Some patients treated with ganciclovir for CMV retinitis have developed retinal detachment, but the causal relationship to ganciclovir treatment is unclear. Valganciclovir is an orally bioavailable prodrug that can generate high serum levels of ganciclovir, although studies of its efficacy in treating CMV CNS infections are limited.

Foscarnet is a pyrophosphate analogue that inhibits viral DNA polymerases by binding to the pyrophosphate-binding site. Following intravenous infusion, CSF concentrations range from 15 to 100% of coincident plasma levels. The usual dose for serious CMV-related neurologic illness is 60 mg/kg every 8 h administered by constant infusion over 1 h. Induction therapy for 14–21 days is followed by maintenance therapy (60–120 mg/kg per day). Induction therapy may need to be extended in patients who fail to show a decline in CSF pleocytosis and a reduction in CSF CMV DNA copy number on quantitative PCR tests (where available). Approximately one-third of patients develop renal impairment during treatment, which is reversible following discontinuation of therapy in most, but not all, cases. This is often associated with elevations in serum creatinine and proteinuria and is less frequent in patients who are adequately hydrated. Many patients experience fatigue and nausea. Reductions in serum calcium, magnesium, and potassium occur in ~15% of patients and may be associated with tetany, cardiac rhythm disturbances, or seizures.

Cidofovir is a nucleotide analogue that is effective in treating CMV retinitis and equivalent to or better than ganciclovir in some experimental models of murine CMV encephalitis, although data concerning its efficacy in human CMV CNS disease are limited. The usual dose is 5 mg/kg intravenously once weekly for 2 weeks, then biweekly for two or more additional doses, depending on clinical response. Patients must be prehydrated with normal saline (e.g., 1 L over 1–2 h) prior to each dose and treated with probenecid (e.g., 1 g 3 h before cidofovir and 1 g 2 and 8 h after cidofovir). Nephrotoxicity is common; the dose should be reduced if renal function deteriorates.

Intravenous ribavirin (15–25 mg/kg per day in divided doses given every 8 h) has been reported to be of benefit in isolated cases of severe encephalitis due to California encephalitis (La Crosse) virus. Ribavirin might be of benefit for the rare patients, typically infants or young children, with severe adenovirus or rotavirus encephalitis and in patients with encephalitis due to LCMV or other arenaviruses. However, clinical trials are lacking. Hemolysis, with resulting anemia, has been the major side effect limiting therapy.

No specific antiviral therapy of proven efficacy is currently available for treatment of WNV encephalitis. Patients have been treated with interferon-α, ribavirin, an Israeli IVIg preparation that contains high-titer anti-WNV antibody (Omr-IgG-am), and humanized monoclonal antibodies directed against the viral envelope glycoprotein (*www.clinicaltrials.gov*, identifiers NCT00927953 and 00515385). Omr-IgG-am did not improve outcomes in patients with WNV neuroinvasive disease, but the study design was potentially flawed as some patients received drug up to a week after symptom onset, when expected benefit may have been minimal. Of the six West Nile virus human vaccines that advanced into phase I clinical trials, only two live attenuated virus vaccines have advanced to phase II clinical trials. There has been success with four equine vaccines, but all require multiple primary doses and annual boosters. The ideal human WNV vaccine needs to provide complete and long-lasting protective immunity after the administration of a single dose. Effective vaccines are already in human use for prevention of other flavivirus infections including Japanese encephalitis and yellow fever.

The Centers for Disease Control and Prevention (CDC) Clinical Considerations for COVID-19 treatment in outpatients as of January 2024 state that the preferred treatment for mild to moderate COVID-19 infection in adults is oral ritonavir-boosted nirmatrelvir. This antiviral is U.S. Food and Drug Administration (FDA) approved in adults with mild-to-moderate COVID-19 with symptoms of <5 days in duration who are at high risk of developing severe COVID-19 due to older age (>50 years) or other risk factors. It has also been approved under emergency use authorization (EUA) for 12- to 17-year-olds. A 3-day course of intravenous remdesivir is the second preferred treatment option after ritonavir-boosted nirmatrelvir for adults and pediatric patients as young as 28 days. Immunocompromised patients may be treated with longer or additional courses. The FDA has issued an EUA for the use of COVID-19 convalescent plasma with high titers of anti–SARS-CoV-2 antibodies for the treatment of COVID-19 in immunocompromised patients from disease or immunosuppressive therapy. The antiviral molnupiravir can be used for therapy according to the CDC but has been less effective in clinical trials than ritonavir-boosted nirmatrelvir or remdesivir.

SEQUELAE

There is considerable variation in the incidence and severity of sequelae in patients surviving viral encephalitis. In the case of EEE virus infection, nearly 80% of survivors have severe neurologic sequelae. At the other extreme are infections due to EBV, California encephalitis virus, and Venezuelan equine encephalitis virus, where severe sequelae are unusual. For example, ~5–15% of children infected with La Crosse virus have a residual seizure disorder, and 1% have persistent hemiparesis. Detailed information about sequelae in patients with HSV encephalitis treated with acyclovir is available from the NIAID-Collaborative Antiviral Study Group (CASG) trials. Of 32 acyclovir-treated patients, 26 survived (81%). Of the 26 survivors, 12 (46%) had no or only minor sequelae, 3 (12%) were moderately impaired (gainfully employed but not functioning at their previous level), and 11 (42%) were severely impaired (requiring continuous supportive care). The incidence and severity of sequelae were directly related to the age of the patient and the level of consciousness at the time of initiation of therapy. Patients with severe neurologic impairment (Glasgow Coma Scale score 6) at initiation of therapy either died or survived with severe sequelae. Young patients (<30 years) with good neurologic function at initiation of therapy did substantially better (100% survival, 62% with no or mild sequelae) compared with their older counterparts (>30 years; 64% survival, 57% no or mild sequelae). Many patients with WNV infection have sequelae, including cognitive impairment; weakness; and hyper- or hypokinetic movement disorders, including tremor, myoclonus, and parkinsonism. In a large longitudinal study of prognosis in 156 patients with WNV infection, the mean time to achieve recovery (defined as 95% of maximal predicted score on specific validated tests) was 112–148 days for fatigue, 121–175 days for physical function, 131–139 days for mood, and 302–455 days for mental function (the longer interval in each case representing patients with invasive CNS disease).

CHRONIC ENCEPHALITIS

PROGRESSIVE MULTIFOCAL LEUKOENCEPHALOPATHY

Clinical Features and Pathology Progressive multifocal leukoencephalopathy (PML) is characterized pathologically by multifocal areas of demyelination of varying size distributed throughout the brain but sparing the spinal cord and optic nerves. In addition to demyelination, there are characteristic cytologic alterations in both astrocytes and oligodendrocytes. Astrocytes are enlarged and contain hyperchromatic, deformed, and bizarre nuclei and frequent mitotic figures. Oligodendrocytes have enlarged, densely staining nuclei that contain viral inclusions formed by crystalline arrays of JC virus (JCV) particles. Patients often present with visual deficits (45%), typically a homonymous hemianopia; mental impairment (38%) (dementia, confusion, personality change); weakness, including hemi- or monoparesis; and ataxia. Seizures occur in ~20% of patients, predominantly in those with lesions abutting the cortex.

Almost all patients have an underlying immunosuppressive disorder or are receiving immunomodulatory therapy. The most common immunosuppressive disorder associated with PML is AIDS, followed by hematologic malignancies, solid organ and hematopoietic stem cell transplant, and chronic inflammatory diseases, including sarcoidosis. It has been estimated that up to 5% of AIDS patients will develop PML. There has been considerable progress in the development of disease-modifying therapies (DMTs) for multiple sclerosis and inflammatory bowel disease. Of the DMTs, the highest risk of PML is associated with natalizumab, a humanized monoclonal antibody that inhibits lymphocyte trafficking into CNS and bowel mucosa by binding to α_4 integrins. Overall risk in these patients has been estimated at ~4 PML cases per 1000 treated patients, but the risk depends on a variety of factors including anti-JCV antibody serostatus and the magnitude of the JCV antibody response, prior immunosuppressive therapy use, and duration of natalizumab therapy. Patients who lack detectable JCV antibody have a risk of developing PML of <0.1 case/1000 patients, whereas those who are JCV seropositive and have been exposed to prior immunosuppressive therapy and have received >24 months of natalizumab therapy have a risk of >1.3 cases/100 treated patients. Some recent studies suggest that extended dosing interval regimens of natalizumab (at 6- to 8-week intervals rather than the conventional 4-week interval) may significantly reduce the risk of PML. Among JCV-seropositive individuals, those with higher JCV antibody index values, presumably due to the "immunizing" effects of more frequent JCV reactivations, appear to be at higher risk than those with low antibody indices. Alternative therapies are preferred in patients who are JCV seropositive. PML cases have also been reported in patients receiving other immunomodulatory agents including rituximab, ocrelizumab, fingolimod, and dimethyl fumarate, although the relative risks have not been clearly established, and many individual cases are complicated by previous exposure to other therapies including natalizumab. Prolonged lymphopenia, a side effect of dimethyl fumarate, is associated with an increased risk of PML. The basic clinical and diagnostic features appear to be similar in HIV-associated PML and PML associated with immunomodulatory drugs with the exception of an increased likelihood of MRI enhancement of PML lesions in immunomodulatory cases. In natalizumab-associated PML, patients will also almost invariably develop clinical and radiographic worsening of lesions with discontinuation of therapy, attributed to development of immune reconstitution inflammatory syndrome (IRIS).

Diagnostic Studies The diagnosis of PML is frequently suggested by MRI. MRI reveals multifocal asymmetric, coalescing white matter lesions located periventricularly, in the centrum semiovale, in the parietal-occipital region, and in the cerebellum. These lesions have increased signal on T2 and FLAIR images and decreased signal on T1-weighted images. HIV-PML lesions are classically nonenhancing (90%), but patients with immunomodulatory drug-associated PML may have peripheral ring enhancement. PML lesions are not typically associated with edema or mass effect. CT scans, which are less sensitive than MRI for the diagnosis of PML, often show hypodense nonenhancing white matter lesions. JCV infection may also induce rare cases of encephalitis and cerebellitis in immunocompromised patients that are distinct from PML and have differing neuroimaging features.

The CSF is typically normal, although mild elevation in protein and/or IgG may be found. Pleocytosis occurs in <25% of cases, is predominantly mononuclear, and rarely exceeds 25 cells/μL. PCR amplification of JCV DNA from CSF has become an important diagnostic tool. The presence of a positive CSF PCR for JCV DNA in association with typical MRI lesions in the appropriate clinical setting is diagnostic of PML, reflecting the assay's relatively high specificity (92–100%); however, sensitivity is variable, and a negative CSF PCR does not exclude the diagnosis. In HIV-negative patients and HIV-positive patients not receiving antiretroviral therapy (ART), sensitivity is likely 70–90%. In ART-treated patients, sensitivity may be closer to 60%, reflecting the lower JCV CSF viral load in this relatively more immunocompetent group. Patients with natalizumab-associated PML have highly variable amounts of JCV DNA in CSF. Some patients may have negative CSF PCRs performed in commercial laboratories, where assay detection thresholds are typically >100 JCV DNA copies/μL, but positive results in reference laboratories using supersensitive techniques (detection of 10 JCV copies/μL or less). CSF studies with quantitative JCV PCR indicate that patients with low JCV loads (<100 copies/μL) have a generally better prognosis than those with higher viral loads. Patients with negative CSF PCR studies may require brain biopsy for definitive diagnosis. In biopsy or necropsy specimens of brain, JCV antigen and nucleic acid can be detected by immunocytochemistry, in situ hybridization, or PCR amplification.

Serologic studies of JCV antibody are of modest value in diagnosis of PML due to the high basal seroprevalence level, although the absence of detectable JCV antibody may be useful in reducing the likelihood of PML in the differential diagnosis, as PML results from viral reactivation in previously infected individuals and virtually all confirmed cases have been JCV seropositive at diagnosis. Antibody testing may also be useful in risk stratification of patients receiving immunomodulatory therapies.

TREATMENT

Progressive Multifocal Leukoencephalopathy

No consistently effective therapy for PML is available. There are case reports of potential beneficial effects of the 5-HT_{2a} receptor antagonist mirtazapine, which may inhibit binding of JCV to its receptor on oligodendrocytes. Retrospective noncontrolled studies have also suggested a possible beneficial effect of treatment with interferon-α. Neither of these agents has been tested in randomized controlled clinical trials. A prospective multicenter clinical trial to evaluate the efficacy of the antimalarial drug mefloquine failed to show benefit. Intravenous and/or intrathecal cytarabine were not shown to be of benefit in a randomized controlled trial in HIV-associated PML, although some experts suggest that cytarabine may have therapeutic efficacy in situations where breakdown of the blood-brain barrier allows sufficient CSF penetration. A randomized controlled trial of cidofovir in HIV-associated PML also failed to show significant benefit. Because PML almost invariably occurs in immunocompromised individuals, any therapeutic interventions designed to enhance or restore immunocompetence should be considered; a small series of patients treated with the PD-1 inhibitor pembrolizumab demonstrated clinical improvement and stabilization. Positive results in small case series have also been reported in patients receiving infusions of BK or JC virus–specific cytotoxic T lymphocytes. Perhaps the most dramatic demonstration of the benefit of restoring immune competence is disease stabilization and, in rare cases, improvement associated with an improved immune status of HIV-positive patients with AIDS following institution of ART. In HIV-positive PML patients treated with ART, 1-year survival is ~50%, although up to 80% of survivors may have significant neurologic sequelae. HIV-positive PML patients with higher CD4 counts (>300/μL) and low or nondetectable HIV viral loads have a better prognosis than those with lower CD4 counts and higher viral loads. Although institution of ART enhances survival in HIV-positive PML patients, the associated immune reconstitution in patients with an underlying opportunistic infection such as PML may also result in a severe CNS inflammatory syndrome (IRIS) associated with clinical worsening, CSF pleocytosis, and the appearance of new enhancing MRI lesions. Patients receiving natalizumab or other immunomodulatory therapies who are suspected of having PML should have therapy immediately halted. Patients should be closely monitored for development of IRIS, which is generally treated with intravenous glucocorticoids, although controlled clinical trials of efficacy remain lacking.

■ SUBACUTE SCLEROSING PANENCEPHALITIS

Subacute sclerosing panencephalitis (SSPE) is a rare, chronic, progressive demyelinating disease of the CNS associated with a chronic nonpermissive infection of brain tissue with measles virus. The frequency has been estimated at 1 in 100,000–500,000 measles cases. An average of five cases per year is reported in the United States. The incidence has declined dramatically since the introduction of a measles vaccine, but we may expect a rise in cases over the coming decades with increasing vaccine hesitancy and rising measles cases in the United States and

Europe in children from countries where routine childhood immunizations are not available. Most patients give a history of primary measles infection at an early age (2 years), which is followed after a latent interval of 6–8 years by the development of a progressive neurologic disorder. Some 85% of patients are between 5 and 15 years old at diagnosis. Initial manifestations include poor school performance and mood and personality changes. Typical signs of a CNS viral infection, including fever and headache, do not occur. As the disease progresses, patients develop progressive intellectual deterioration, focal and/or generalized seizures, myoclonus, ataxia, and visual disturbances. In the late stage of the illness, patients are unresponsive, quadriparetic, and spastic, with hyperactive tendon reflexes and extensor plantar responses.

Diagnostic Studies MRI is often normal early, although areas of increased T2 signal develop in the white matter of the brain and brainstem as disease progresses. The EEG may initially show only nonspecific slowing, but with disease progression, patients develop a characteristic periodic pattern with bursts of high-voltage, sharp, slow waves every 3–8 s, followed by periods of attenuated ("flat") background. The CSF is acellular with a normal or mildly elevated protein concentration and a markedly elevated gamma globulin level (>20% of total CSF protein). CSF antimeasles antibody levels are invariably elevated, and oligoclonal antimeasles antibodies are often present. Measles virus can be cultured from brain tissue using special cocultivation techniques. Viral antigen can be identified immunocytochemically, and viral genome can be detected by in situ hybridization or PCR amplification.

TREATMENT

Subacute Sclerosing Panencephalitis

No definitive therapy for SSPE is available. Treatment with isoprinosine (Inosiplex, 100 mg/kg per day), alone or in combination with intrathecal or intraventricular interferon-α, has been reported to prolong survival and produce clinical improvement in some patients but has never been subjected to a controlled clinical trial.

PROGRESSIVE RUBELLA PANENCEPHALITIS

This is an extremely rare disorder that primarily affects males with congenital rubella syndrome, although isolated cases have been reported following childhood rubella. After a latent period of 8–19 years, patients develop progressive neurologic deterioration. The manifestations are similar to those seen in SSPE. CSF shows a mild lymphocytic pleocytosis, slightly elevated protein concentration, markedly increased gamma globulin, and rubella virus–specific oligoclonal bands. No therapy is available. Universal prevention of both congenital and childhood rubella through the use of the available live attenuated rubella vaccine would be expected to eliminate the disease.

FURTHER READING

Bernard-Valnet R et al: Advances in treatment of progressive multifocal leukoencephalopathy. Ann Neurol 90:865, 2021.

Hodzic E et al: Steroids for the treatment of viral encephalitis: A systematic literature review and meta-analysis. Neurology 270:3603, 2023.

Kaiser JA et al: Twenty years of progress toward West Nile virus vaccine development. Viruses 11:823, 2019.

Ramachandran PS, Wilson MR: Metagenomics for neurological infections: Expanding our imagination. Nat Rev Neurol 16:547, 2020.

Tunkel AR et al: The management of encephalitis: Clinical practice guidelines by the Infectious Diseases Society of America. Clin Infect Dis 47:303, 2008.

Tyler KL: Acute viral encephalitis. N Engl J Med 379:557, 2018.

Venkatesan A et al: International Encephalitis Consortium. Case definitions, diagnostic algorithms, and priorities in encephalitis: Consensus statement of the International Encephalitis Consortium. Clin Infect Dis 57:1114, 2013.

143

Acute Meningitis

Karen L. Roos, Kenneth L. Tyler

BACTERIAL MENINGITIS

DEFINITION

Bacterial meningitis is an acute purulent infection within the subarachnoid space (SAS). It is associated with a central nervous system (CNS) inflammatory reaction that may result in decreased consciousness, seizures, raised intracranial pressure (ICP), and stroke. The meninges, SAS, and brain parenchyma are all frequently involved in the inflammatory reaction (*meningoencephalitis*).

EPIDEMIOLOGY

Bacterial meningitis is the most common form of suppurative CNS infection, with an annual incidence in the United States of ~1.4 cases/100,000 population. The organisms most often responsible for community-acquired bacterial meningitis are *Streptococcus pneumoniae* (~50%), *Neisseria meningitidis* (~25%), group B streptococci (~15%), and *Listeria monocytogenes* (~10%). *Haemophilus influenzae* type b accounts for <10% of cases of bacterial meningitis in most series. *N. meningitidis* is the causative organism of recurring epidemics of meningitis every 8–12 years.

ETIOLOGY

S. pneumoniae **(Chap. 151)** is the most common cause of meningitis in adults >20 years of age, accounting for nearly half the reported cases (1.1 per 100,000 persons per year). There are a number of predisposing conditions that increase the risk of pneumococcal meningitis, the most important of which is pneumococcal pneumonia. Additional risk factors include coexisting acute or chronic pneumococcal sinusitis or otitis media, alcoholism, diabetes, splenectomy, hypogammaglobulinemia, complement deficiency, and head trauma with basilar skull fracture and cerebrospinal fluid (CSF) rhinorrhea. The mortality rate remains ~20% despite antibiotic therapy.

The incidence of meningitis due to *N. meningitidis* **(Chap. 160)** has decreased with the routine immunization of 11- to 18-year-olds with the quadrivalent (serogroups A, C, W-135, and Y) meningococcal glycoconjugate vaccine, and adolescents and young adults (16–23 years old) with the serogroup B meningococcal vaccine. A pentavalent meningococcal vaccine (serogroups A, B, C, W-135, and Y) has recently become available. Individuals being treated with complement inhibitors are at increased risk of meningococcal disease and should receive either the quadrivalent vaccine and a serogroup B meningococcal vaccine, or the pentavalent meningococcal vaccine, prior to beginning therapy. Individuals with complement component deficiencies are at increased risk for meningococcal disease and similarly should receive either the quadrivalent and a serogroup B meningococcal vaccine or the pentavalent meningococcal vaccine. The meningococcal vaccines use outer membrane proteins as the vaccine antigens. The serogroup B polysaccharide capsule is poorly immunogenic. The serogroup B meningococcal vaccines do not reduce the risk of bacterial spread of group B meningococcus from vaccinated persons to *unimmunized* persons as the vaccines do not significantly reduce nasopharyngeal carriage of meningococci, and this remains the major source of person-to-person bacterial transmission. In contrast, nasopharyngeal carriage is reduced in vaccinated individuals who have received the conjugate vaccines that cover groups A, C, W, and Y. The presence of petechial or purpuric skin lesions can provide an important clue to the diagnosis of meningococcal infection. In some patients, the disease is fulminant, progressing to death within hours of symptom onset. Infection may be initiated by nasopharyngeal colonization, which can result in either an asymptomatic carrier state or invasive meningococcal disease. The risk of invasive disease following nasopharyngeal colonization depends on both bacterial virulence factors and host immune defense mechanisms, including the host's capacity to produce antimeningococcal antibodies and to lyse meningococci by both classic and alternative complement pathways. Individuals with deficiencies of any of the complement components, including properdin, are highly susceptible to meningococcal infections.

Gram-negative bacilli cause meningitis in individuals with chronic and debilitating diseases such as diabetes, cirrhosis, or alcoholism and in those with urinary tract infections. Gram-negative meningitis can also complicate neurosurgical procedures, particularly craniotomy, and head trauma associated with CSF rhinorrhea or otorrhea.

Otitis, mastoiditis, and sinusitis are predisposing and associated conditions for meningitis due to *Streptococcus* spp., gram-negative anaerobes, *Staphylococcus aureus, Haemophilus* spp., and Enterobacteriaceae. Meningitis complicating endocarditis may be due to viridans streptococci, *S. aureus, Streptococcus bovis,* the HACEK group (*Haemophilus* spp., *Actinobacillus actinomycetemcomitans, Cardiobacterium hominis, Eikenella corrodens, Kingella kingae*), or enterococci.

Group B streptococcus., or *Streptococcus agalactiae* **(Chap. 153)**, was previously responsible for meningitis predominantly in neonates, but it has been reported with increasing frequency in individuals aged >50 years, particularly those with underlying diseases.

L. monocytogenes **(Chap. 156)** is an increasingly important cause of meningitis in neonates (<1 month of age), pregnant women, individuals >60 years, and immunocompromised individuals of all ages. Infection is acquired by ingesting foods contaminated by *Listeria*. Foodborne human listerial infection has been reported from contaminated coleslaw, milk, soft cheeses, and several types of "ready-to-eat" foods, including delicatessen meat and uncooked hotdogs.

The frequency of *H. influenzae* type b (Hib) meningitis in children has declined dramatically since the introduction of the Hib conjugate vaccine, although rare cases of Hib meningitis in vaccinated children have been reported. More frequently, *H. influenzae* causes meningitis in unvaccinated children and older adults, and non-b *H. influenzae* is an emerging pathogen **(Chap. 162)**.

S. aureus and coagulase-negative staphylococci **(Chap. 152)** are important causes of meningitis that occurs following invasive neurosurgical procedures, particularly shunting procedures for hydrocephalus, or as a complication of the use of subcutaneous Ommaya reservoirs for administration of intrathecal chemotherapy.

■ PATHOPHYSIOLOGY

The most common bacteria that cause meningitis, *S. pneumoniae* and *N. meningitidis*, initially colonize the nasopharynx by attaching to nasopharyngeal epithelial cells. Bacteria are transported across epithelial cells in membrane-bound vacuoles to the intravascular space or invade the intravascular space by creating separations in the apical tight junctions of columnar epithelial cells. Once in the bloodstream, bacteria are able to avoid phagocytosis by neutrophils and classic complement-mediated bactericidal activity because of the presence of a polysaccharide capsule. Bloodborne bacteria can reach the intraventricular choroid plexus, directly infect choroid plexus epithelial cells, and gain access to the CSF. Some bacteria, such as *S. pneumoniae*, can adhere to cerebral capillary endothelial cells and subsequently migrate through or between these cells to reach the CSF. Bacteria are able to multiply rapidly within CSF because of the absence of effective host immune defenses. Normal CSF contains few white blood cells (WBCs) and relatively small amounts of complement proteins and immunoglobulins. The paucity of the latter two prevents effective opsonization of bacteria, an essential prerequisite for bacterial phagocytosis by neutrophils. Phagocytosis of bacteria is further impaired by the fluid nature of CSF, which is less conducive to phagocytosis than a solid tissue substrate.

A critical event in the pathogenesis of bacterial meningitis is the inflammatory reaction induced by the invading bacteria. Many of the neurologic manifestations and complications of bacterial meningitis result from the immune response to the invading pathogen rather than from direct bacteria-induced tissue injury. As a result, neurologic injury can progress even after the CSF has been sterilized by antibiotic therapy.

The lysis of bacteria with the subsequent release of cell-wall components into the SAS is the initial step in the induction of the inflammatory response and the formation of a purulent exudate in the SAS **(Fig. 143-1)**. Bacterial cell-wall components, such as the lipopolysaccharide (LPS) molecules of gram-negative bacteria and teichoic acid and peptidoglycans of *S. pneumoniae*, induce meningeal inflammation by stimulating the production of inflammatory cytokines and chemokines by microglia, astrocytes, monocytes, microvascular endothelial cells, and CSF leukocytes. In experimental models of meningitis, cytokines including tumor necrosis factor alpha (TNF-α) and interleukin 1β (IL-1β) are present in CSF within 1–2 h of intracisternal inoculation of LPS. This cytokine response is quickly followed by an increase in CSF protein concentration and leukocytosis. Chemokines (cytokines that induce chemotactic migration in leukocytes) and a variety of other proinflammatory cytokines are also produced and secreted by leukocytes and tissue cells that are stimulated by IL-1β and TNF-α. In addition, bacteremia and the inflammatory cytokines induce the production of excitatory amino acids, reactive oxygen and nitrogen species (free oxygen radicals, nitric oxide, and peroxynitrite), and other mediators that can induce death of brain cells, especially in the dentate gyrus of the hippocampus.

Much of the pathophysiology of bacterial meningitis is a direct consequence of elevated levels of CSF cytokines and chemokines. TNF-α and IL-1β act synergistically to increase the permeability of the blood-brain barrier, resulting in induction of vasogenic edema and the leakage of serum proteins into the SAS (Fig. 143-1). The subarachnoid exudate of proteinaceous material and leukocytes obstructs the flow of CSF through the ventricular system and diminishes the resorptive capacity of the arachnoid granulations in the dural sinuses, leading to obstructive and communicating hydrocephalus and concomitant interstitial edema.

Inflammatory cytokines upregulate the expression of selectins on cerebral capillary endothelial cells and leukocytes, promoting leukocyte adherence to vascular endothelial cells and subsequent migration into the CSF. The adherence of leukocytes to capillary endothelial cells increases the permeability of blood vessels, allowing for the leakage of plasma proteins into the CSF, which adds to the inflammatory exudate. Neutrophil degranulation results in the release of toxic metabolites that contribute to cytotoxic edema, cell injury, and death. Contrary to previous beliefs, CSF leukocytes probably do little to contribute to the clearance of CSF bacterial infection.

During the very early stages of meningitis, there is an increase in cerebral blood flow, soon followed by a decrease in cerebral blood flow and a loss of cerebrovascular autoregulation **(Chap. 318)**. Narrowing of the large arteries at the base of the brain due to encroachment by the purulent exudate in the SAS and infiltration of the arterial wall by inflammatory cells with intimal thickening (*vasculitis*) also occur and may result in ischemia and infarction, obstruction of branches of the middle cerebral artery by thrombosis, thrombosis of the major cerebral venous sinuses, and thrombophlebitis of the cerebral cortical veins. The combination of interstitial, vasogenic, and cytotoxic edema leads to raised ICP and coma. Cerebral herniation usually results from the effects of cerebral edema, either focal or generalized; hydrocephalus and dural sinus or cortical vein thrombosis may also play a role.

■ CLINICAL PRESENTATION

Meningitis can present as either an acute fulminant illness that progresses rapidly in a few hours or as a subacute infection that progressively worsens over several days. The classic clinical triad of meningitis is fever, headache, and nuchal rigidity, and these features each occur in >80% of adult cases of acute bacterial meningitis, although the complete classic triad is not always present. A decreased level of consciousness occurs in >75% of patients and can vary from lethargy to coma. Nausea, vomiting, and photophobia are also common complaints.

Nuchal rigidity ("stiff neck") is the pathognomonic sign of meningeal irritation and is present when the neck resists passive flexion. Kernig's and Brudzinski's signs are also classic signs of meningeal irritation. *Kernig's sign* is elicited with the patient in the supine position. The thigh is flexed on the abdomen, with the knee flexed; attempts to passively extend the knee elicit pain when meningeal irritation is present. *Brudzinski's sign* is elicited with the patient in the supine

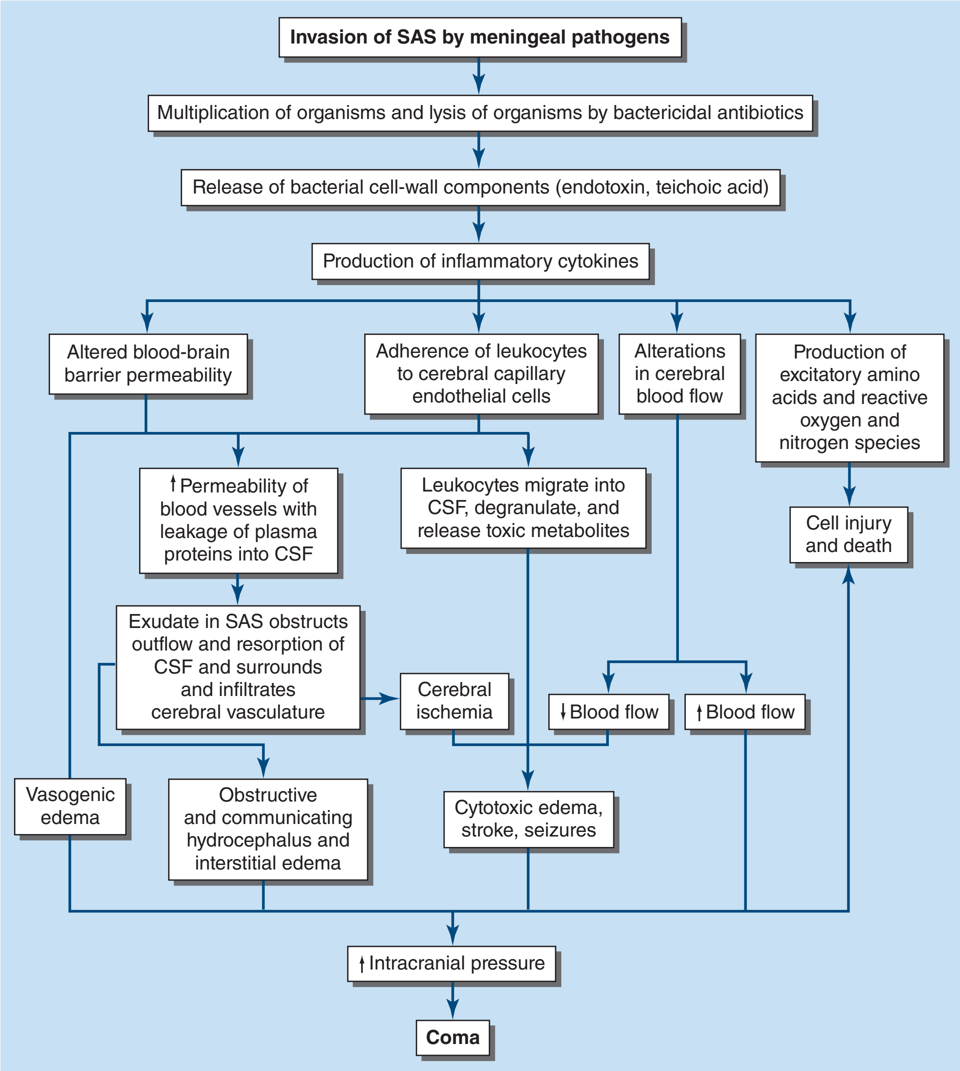
Invasion of SAS by meningeal pathogens
Multiplication of organisms and lysis of organisms by bactericidal antibiotics
Release of bacterial cell-wall components (endotoxin, teichoic acid)
Production of inflammatory cytokines
Altered blood-brain barrier permeability
Adherence of leukocytes to cerebral capillary endothelial cells
Alterations in cerebral blood flow
Production of excitatory amino acids and reactive oxygen and nitrogen species
↑ Permeability of blood vessels with leakage of plasma proteins into CSF
Leukocytes migrate into CSF, degranulate, and release toxic metabolites
Cell injury and death
Exudate in SAS obstructs outflow and resorption of CSF and surrounds and infiltrates cerebral vasculature
Cerebral ischemia
↓ Blood flow
↑ Blood flow
Vasogenic edema
Obstructive and communicating hydrocephalus and interstitial edema
Cytotoxic edema, stroke, seizures
↑ Intracranial pressure
Coma

FIGURE 143-1 The pathophysiology of the neurologic complications of bacterial meningitis. CSF, cerebrospinal fluid; SAS, subarachnoid space.

position and is positive when passive flexion of the neck results in spontaneous flexion of the hips and knees. Although commonly tested on physical examinations, the sensitivity and specificity of Kernig's and Brudzinski's signs are uncertain. Both may be absent or reduced in very young or elderly patients, immunocompromised individuals, or patients with a severely depressed mental status. The high prevalence of cervical spine disease in older individuals may result in false-positive tests for nuchal rigidity.

Seizures occur as part of the initial presentation of bacterial meningitis or during the course of the illness in 15–40% of patients. Focal seizures are usually due to focal arterial ischemia or infarction, cortical venous thrombosis with hemorrhage, or focal edema. Generalized seizure activity and status epilepticus may be due to hyponatremia, cerebral anoxia, or, less commonly, the toxic effects of antimicrobial agents.

Raised ICP is an expected complication of bacterial meningitis and the major cause of obtundation and coma in this disease. More than 90% of patients will have a CSF opening pressure >180 mmH$_2$O, and 20% have opening pressures >400 mmH$_2$O. Signs of increased ICP include a deteriorating or reduced level of consciousness, papilledema, dilated poorly reactive pupils, sixth nerve palsies, decerebrate posturing, and the Cushing reflex (bradycardia, hypertension, and irregular respirations). The most disastrous complication of increased ICP is cerebral herniation. The incidence of herniation in patients with bacterial meningitis has been reported to occur in as few as 1% to as many as 8% of cases.

Specific clinical features may provide clues to the diagnosis of individual organisms and are discussed in more detail in specific chapters devoted to individual pathogens. The most important of these clues is the rash of meningococcemia, which begins as a diffuse erythematous maculopapular rash resembling a viral exanthem; however, the skin lesions of meningococcemia rapidly become petechial. Petechiae are found on the trunk and lower extremities, in the mucous membranes and conjunctiva, and occasionally on the palms and soles.

■ DIAGNOSIS

When bacterial meningitis is suspected, blood cultures should be immediately obtained and empirical antimicrobial and adjunctive dexamethasone therapy initiated without delay **(Table 143-1)**. Therapy with dexamethasone should ideally be started 20 min before, or not later than concurrent with, the first dose of antibiotics. The diagnosis of bacterial meningitis is made by examination of the CSF **(Table 143-2)**. The need to obtain neuroimaging studies (computed tomography [CT] or magnetic resonance imaging [MRI]) prior to lumbar puncture (LP) requires clinical judgment. In an immunocompetent patient with no known history of recent head trauma, a normal level of consciousness, and no evidence of papilledema or focal neurologic deficits, it is considered safe to perform LP without prior neuroimaging studies. If LP is delayed in order to obtain neuroimaging studies, empirical antibiotic therapy should be initiated after blood cultures are obtained. Antibiotic therapy initiated a few hours prior to LP will not significantly alter the CSF WBC count or glucose concentration, nor is it likely to prevent visualization of organisms by Gram's stain or detection of bacterial nucleic acid by polymerase chain reaction (PCR) assay.

The classic CSF abnormalities in bacterial meningitis (Table 143-2) are (1) polymorphonuclear (PMN) leukocytosis (>100 cells/μL in 90%), (2) decreased glucose concentration (<2.2 mmol/L [<40 mg/dL] and/or CSF/serum glucose ratio of <0.4 in ~60%), (3) increased protein concentration (>0.45 g/L [>45 mg/dL] in 90%), and (4) increased opening pressure (>180 mmH$_2$O in 90%). CSF bacterial cultures are positive in >70% of patients, and CSF Gram's stain demonstrates organisms in >60%.

CSF glucose concentrations <2.2 mmol/L (<40 mg/dL) are abnormal, and a CSF glucose concentration of zero can be seen in bacterial meningitis. Use of the CSF/serum glucose ratio corrects for hyperglycemia that may mask a relative decrease in the CSF glucose concentration. The CSF glucose concentration is low when the CSF/serum glucose ratio is <0.6. A CSF/serum glucose ratio <0.4 is highly suggestive of bacterial meningitis but may also be seen in other conditions, including fungal, tuberculous, and carcinomatous meningitis. It takes from 30 min to several hours for the concentration of CSF glucose to reach equilibrium with blood glucose levels; therefore, administration of 50 mL of 50% glucose (D50) prior to LP, as commonly occurs in emergency room settings, is unlikely to alter CSF glucose concentration significantly unless more than a few hours have elapsed between glucose administration and LP.

The presently available CSF multiplex PCR pathogen assays, the most common of which is the FilmArray Meningitis/Encephalitis panel (BioFire Diagnostics), detect the nucleic acid of *S. pneumoniae,*

TABLE 143-1 Antibiotics Used in Empirical Therapy of Bacterial Meningitis and Focal Central Nervous System Infections[a]

INDICATION	ANTIBIOTIC
Preterm infants to infants <1 month	Ampicillin + cefotaxime
Infants 1–3 months	Ampicillin + cefotaxime or ceftriaxone
Immunocompetent children >3 months and adults <55	Cefotaxime, ceftriaxone, or cefepime + vancomycin
Adults >55 and adults of any age with alcoholism or other debilitating illnesses	Ampicillin + cefotaxime, ceftriaxone, or cefepime + vancomycin
Hospital-acquired meningitis, posttraumatic or postneurosurgery meningitis, neutropenic patients, or patients with impaired cell-mediated immunity	Ampicillin + ceftazidime or meropenem + vancomycin

	TOTAL DAILY DOSE AND DOSING INTERVAL	
ANTIMICROBIAL AGENT	CHILD (>1 MONTH)	ADULT
Ampicillin	300 (mg/kg)/d, q6h	12 g/d, q4h
Cefepime	150 (mg/kg)/d, q8h	6 g/d, q8h
Cefotaxime	225–300 (mg/kg)/d, q6h	12 g/d, q4h
Ceftriaxone	100 (mg/kg)/d, q12h	4 g/d, q12h
Ceftazidime	150 (mg/kg)/d, q8h	6 g/d, q8h
Gentamicin	7.5 (mg/kg)/d, q8h[b]	7.5 (mg/kg)/d, q8h
Meropenem	120 (mg/kg)/d, q8h	6 g/d, q8h
Metronidazole	30 (mg/kg)/d, q6h	1500–2000 mg/d, q6h
Nafcillin	200 (mg/kg)/d, q6h	12 g/d, q4h
Penicillin G	400,000 (U/kg)/d, q4h	20–24 million U/d, q4h
Vancomycin	45–60 (mg/kg)/d, q6h	45–60 (mg/kg)d, q6–12h[b]

[a]All antibiotics are administered intravenously; doses indicated assume normal renal and hepatic function. [b]Doses should be adjusted based on serum peak and trough levels: gentamicin therapeutic level: peak: 5–8 μg/mL; trough: <2 μg/mL; vancomycin therapeutic level: peak: 25–40 μg/mL; trough: 5–15 μg/mL.

N. meningitidis, Escherichia coli, L. monocytogenes, H. influenzae, and *S. agalactiae* (group B streptococci). Although these PCR assays have a rapid turnaround time, the sensitivity and specificity for the bacterial meningeal pathogens they test for are not known. The CSF multiplex PCR pathogen assays do not include *S. aureus*, coagulase-negative staphylococci, and many gram-negative organisms. The PCR assays cannot replace CSF bacteria culture, as culture is required for antimicrobial susceptibility testing. Almost all patients with bacterial meningitis will have neuroimaging studies performed during the course of their illness. MRI is preferred over CT because of its superiority in demonstrating areas of cerebral edema and ischemia. In patients with bacterial meningitis, diffuse meningeal enhancement is often seen after the administration of gadolinium. Meningeal enhancement is not diagnostic of meningitis but occurs in any CNS disease associated with increased blood-brain barrier permeability.

TABLE 143-2 Cerebrospinal Fluid (CSF) Abnormalities in Bacterial Meningitis

Opening pressure	>180 mmH_2O
White blood cells	10/μL to 10,000/μL; neutrophils predominate
Red blood cells	Absent in nontraumatic tap
Glucose	<2.2 mmol/L (<40 mg/dL)
CSF/serum glucose	<0.4
Protein	>0.45 g/L (>45 mg/dL)
Gram's stain	Positive in >60%
Culture	Positive in >80%
PCR	Detects bacterial DNA

Abbreviation: PCR, polymerase chain reaction.

Petechial skin lesions, if present, should be biopsied. The rash of meningococcemia results from the dermal seeding of organisms with vascular endothelial damage, and biopsy may reveal the organism on Gram's stain.

DIFFERENTIAL DIAGNOSIS

Viral meningoencephalitis, and particularly herpes simplex virus (HSV) encephalitis **(Chap. 142)**, can mimic the clinical presentation of bacterial meningitis (encephalitis). HSV encephalitis typically presents with headache, fever, altered consciousness, focal neurologic deficits (e.g., dysphasia, hemiparesis), and focal or generalized seizures. The findings on CSF studies, neuroimaging, and electroencephalogram (EEG) distinguish HSV encephalitis from bacterial meningitis. The typical CSF profile with viral CNS infections is a lymphocytic pleocytosis with a normal glucose concentration, in contrast to the PMN pleocytosis and hypoglycorrhachia characteristic of bacterial meningitis. The CSF HSV PCR has a 96% sensitivity and a 99% specificity when CSF is examined 72 h following symptom onset and in the first week of antiviral therapy. MRI abnormalities (other than meningeal enhancement) are not seen in uncomplicated bacterial meningitis. By contrast, in HSV encephalitis, on T2-weighted, fluid-attenuated inversion recovery (FLAIR), and diffusion-weighted MRI images, high-signal-intensity lesions are seen in the orbitofrontal, anterior, and medial temporal lobes in the majority of patients within 48 h of symptom onset. Some patients with HSV encephalitis have a distinctive periodic pattern on EEG.

Rickettsial disease **(Chap. 192)** can resemble bacterial meningitis. Rocky Mountain spotted fever (RMSF) is transmitted by a tick bite and caused by the bacteria *Rickettsia rickettsii*. The disease may present acutely with high fever, prostration, myalgia, headache, nausea, and vomiting. Most patients develop a characteristic rash within 96 h of the onset of symptoms. The rash is initially a diffuse erythematous maculopapular rash that may be difficult to distinguish from that of meningococcemia. It progresses to a petechial rash, then to a purpuric rash, and if untreated, to skin necrosis or gangrene. The color of the lesions changes from bright red to very dark red, then yellowish-green to black. The rash typically begins in the wrist and ankles and then spreads distally and proximally within a matter of a few hours, involving the palms and soles. Diagnosis is made by immunofluorescent staining of skin biopsy specimens. Ehrlichioses are also transmitted by a tick bite. These are small gram-negative coccobacilli of which two species cause human disease. *Anaplasma phagocytophilum* causes human granulocytic ehrlichiosis (anaplasmosis), and *Ehrlichia chaffeensis* causes human monocytic ehrlichiosis. The clinical and laboratory manifestations of the infections are similar. Patients present with fever, headache, confusion, nausea, and vomiting. Twenty percent of patients have a maculopapular or petechial rash. There is laboratory evidence of leukopenia, thrombocytopenia, and anemia, and mild to moderate elevations in alanine aminotransferases, alkaline phosphatase, and lactate dehydrogenase. Patients with RMSF and those with ehrlichial infections may have an altered level of consciousness ranging from mild lethargy to coma, confusion, focal neurologic signs, cranial nerve palsies, hyperreflexia, and seizures.

Focal suppurative CNS infections, including subdural and epidural empyema and brain abscess, should also be considered **(Chap. 145)**, especially when focal neurologic findings are present. MRI should be performed promptly in all patients with suspected meningitis who have focal features, both to detect the intracranial infection and to search for associated areas of infection in the sinuses or mastoid bones.

A number of noninfectious CNS disorders can mimic bacterial meningitis. Subarachnoid hemorrhage (SAH; **Chap. 440**) is generally the major consideration. Other possibilities include medication-induced hypersensitivity meningitis; chemical meningitis due to rupture of tumor contents into the CSF (e.g., from a cystic glioma, craniopharyngioma, epidermoid or dermoid cyst); carcinomatous or lymphomatous meningitis; meningitis associated with inflammatory disorders such as sarcoid, systemic lupus erythematosus (SLE), and

Behçet's syndrome; pituitary apoplexy; and uveomeningitic syndromes (Vogt-Koyanagi-Harada syndrome).

On occasion, subacutely evolving meningitis (see below and **Chap. 144**) may be considered in the differential diagnosis of acute meningitis. The principal causes include *Mycobacterium tuberculosis* **(Chap. 183)**, *Cryptococcus neoformans* **(Chap. 221)**, *Histoplasma capsulatum* **(Chap. 218)**, *Coccidioides immitis* **(Chap. 219)**, and *Treponema pallidum* **(Chap. 187)**.

TREATMENT

Acute Bacterial Meningitis

EMPIRICAL ANTIMICROBIAL THERAPY

(Table 143-1) Bacterial meningitis is a medical emergency. The goal is to begin antibiotic therapy within 60 min of a patient's arrival in the emergency room. Empirical antimicrobial therapy is initiated in patients with suspected bacterial meningitis before the results of CSF multiplex PCR assays, Gram's stain, and culture are known. *S. pneumoniae* **(Chap. 151)** and *N. meningitidis* **(Chap. 160)** are the most common etiologic organisms of community-acquired bacterial meningitis. Due to the emergence of penicillin- and cephalosporin-resistant *S. pneumoniae*, empirical therapy of community-acquired suspected bacterial meningitis in children and adults should include a combination of dexamethasone, a third- or fourth-generation cephalosporin (e.g., ceftriaxone, cefotaxime, or cefepime), and vancomycin, plus acyclovir, as HSV encephalitis is the leading disease in the differential diagnosis, and doxycycline during tick season to treat tick-borne bacterial infections. Ceftriaxone or cefotaxime provides good coverage for susceptible *S. pneumoniae*, group B streptococci, and *H. influenzae* and adequate coverage for *N. meningitidis*. Cefepime is a broad-spectrum fourth-generation cephalosporin with in vitro activity similar to that of cefotaxime or ceftriaxone against *S. pneumoniae* and *N. meningitidis* and greater activity against *Enterobacter* species and *Pseudomonas aeruginosa*. In clinical trials, cefepime has been demonstrated to be equivalent to cefotaxime in the treatment of penicillin-sensitive pneumococcal and meningococcal meningitis, and it has been used successfully in some patients with meningitis due to *Enterobacter* species and *P. aeruginosa*. Cefepime has been associated with seizures, myoclonus, and encephalopathy, any of which may limit its use in critically ill patients. Ampicillin should be added to the empirical regimen for coverage of *L. monocytogenes* in individuals <3 months of age, those >55, or those with suspected impaired cell-mediated immunity because of chronic illness, organ transplantation, pregnancy, malignancy, or immunosuppressive therapy. Metronidazole is added to the empirical regimen to cover gram-negative anaerobes in patients with otitis, sinusitis, or mastoiditis. In hospital-acquired meningitis, and particularly meningitis following neurosurgical procedures, staphylococci and gram-negative organisms including *P. aeruginosa* are the most common etiologic organisms. In these patients, empirical therapy should include a combination of vancomycin and ceftazidime or meropenem. Ceftazidime or meropenem should be substituted for ceftriaxone or cefotaxime in neurosurgical patients and in neutropenic patients because ceftriaxone and cefotaxime do not provide adequate activity against CNS infection with *P. aeruginosa*. Meropenem is a carbapenem antibiotic that is highly active in vitro against *L. monocytogenes*, has been demonstrated to be effective in cases of meningitis caused by *P. aeruginosa*, and shows good activity against penicillin-resistant pneumococci. In experimental pneumococcal meningitis, meropenem was comparable to ceftriaxone and inferior to vancomycin in sterilizing CSF cultures. When *S. pneumoniae*, *H. influenzae*, *L. monocytogenes*, or aerobic gram-negative bacilli (including *P. aeruginosa* and *E. coli*) are possible meningeal pathogens, based on predisposing and associated conditions, the combination of vancomycin plus meropenem can be recommended as empiric therapy for bacterial meningitis in children and adults. Meropenem should not be used as monotherapy.

TABLE 143-3 Antimicrobial Therapy of Central Nervous System Bacterial Infections Based on Pathogen[a]

ORGANISM	ANTIBIOTIC
Neisseria meningitides	
Penicillin-sensitive	Penicillin G or ampicillin
Penicillin-resistant	Ceftriaxone or cefotaxime
Streptococcus pneumoniae	
Penicillin-sensitive	Penicillin G
Penicillin-intermediate	Ceftriaxone or cefotaxime or cefepime
Penicillin-resistant	Ceftriaxone (or cefotaxime or cefepime) + vancomycin
Gram-negative bacilli (except *Pseudomonas* spp.)	Ceftriaxone or cefotaxime
Pseudomonas aeruginosa	Ceftazidime or cefepime or meropenem
Staphylococci spp.	
Methicillin-sensitive	Nafcillin
Methicillin-resistant	Vancomycin
Listeria monocytogenes	Ampicillin + gentamicin
Haemophilus influenzae	Ceftriaxone or cefotaxime if β-lactamase positive; ampicillin if β-lactamase negative
Streptococcus agalactiae	Penicillin G or ampicillin
Bacteroides fragilis	Metronidazole
Fusobacterium spp.	Metronidazole

[a]Doses are as indicated in Table 143-1.

SPECIFIC ANTIMICROBIAL THERAPY

Meningococcal Meningitis (Table 143-3) Although ceftriaxone and cefotaxime provide adequate empirical coverage for *N. meningitidis*, penicillin G remains the antibiotic of choice for meningococcal meningitis caused by susceptible strains. Isolates of *N. meningitidis* with moderate resistance to penicillin have been identified and are increasing in incidence worldwide. CSF isolates of *N. meningitidis* should be tested for penicillin and ampicillin susceptibility, and if resistance is found, cefotaxime or ceftriaxone should be substituted for penicillin. A 7-day course of intravenous antibiotic therapy is adequate for uncomplicated meningococcal meningitis. The index case and all close contacts should receive chemoprophylaxis with a 2-day regimen of rifampin (600 mg every 12 h for 2 days in adults and 10 mg/kg every 12 h for 2 days in children >1 year). Rifampin is not recommended in pregnant women. Alternatively, adults can be treated with one dose of azithromycin (500 mg) or one intramuscular dose of ceftriaxone (250 mg). Close contacts are defined as those individuals who have had contact with oropharyngeal secretions, either through kissing or by sharing toys, beverages, or cigarettes.

Pneumococcal Meningitis Antimicrobial therapy of pneumococcal meningitis is initiated with a cephalosporin (ceftriaxone, cefotaxime, or cefepime) and vancomycin. All CSF isolates of *S. pneumoniae* should be tested for sensitivity to penicillin and the cephalosporins. Once the results of antimicrobial susceptibility tests are known, therapy can be modified accordingly (Table 143-3). For *S. pneumoniae* meningitis, an isolate of *S. pneumoniae* is considered to be susceptible to penicillin with a minimal inhibitory concentration (MIC) <0.06 μg/mL and to be resistant when the MIC is >0.12 μg/mL. Isolates of *S. pneumoniae* that have cephalosporin MICs ≤0.5 μg/mL are considered sensitive to the cephalosporins (cefotaxime, ceftriaxone, cefepime). Those with MICs of 1 μg/mL are considered to have intermediate resistance, and those with MICs ≥2 μg/mL are considered resistant. For meningitis due to pneumococci, with cefotaxime or ceftriaxone MICs ≤0.5 μg/mL, treatment with cefotaxime or ceftriaxone is usually adequate. For MIC >1 μg/mL, vancomycin is the antibiotic of choice. Rifampin can be added to vancomycin for its synergistic effect but is inadequate as monotherapy because resistance develops rapidly when it is used alone.

A 2-week course of intravenous antimicrobial therapy is recommended for pneumococcal meningitis.

Patients with *S. pneumoniae* meningitis should have a repeat LP performed 24–36 h after the initiation of antimicrobial therapy to document sterilization of the CSF. Failure to sterilize the CSF after 24–36 h of antibiotic therapy should be considered presumptive evidence of antibiotic resistance. Patients with penicillin- and cephalosporin-resistant strains of *S. pneumoniae* who do not respond to intravenous vancomycin alone may benefit from the addition of intraventricular vancomycin. The intraventricular route of administration is preferred over the intrathecal route because adequate concentrations of vancomycin in the cerebral ventricles are not always achieved with intrathecal administration.

Listeria Meningitis Meningitis due to *L. monocytogenes* is treated with ampicillin for at least 3 weeks (Table 143-3). Gentamicin is added in critically ill patients (2 mg/kg loading dose, then 7.5 mg/kg per day given every 8 h and adjusted for serum levels and renal function). The combination of trimethoprim (10–20 mg/kg per day) and sulfamethoxazole (50–100 mg/kg per day) given every 6 h may provide an alternative in penicillin-allergic patients.

Staphylococcal Meningitis Meningitis due to susceptible strains of *S. aureus* or coagulase-negative staphylococci is treated with nafcillin (Table 143-3). Vancomycin is the drug of choice for methicillin-resistant staphylococci and for patients allergic to penicillin. In these patients, the CSF should be monitored during therapy. If the CSF is not sterilized after 48 h of intravenous vancomycin therapy, then either intraventricular or intrathecal vancomycin, 20 mg once daily, can be added.

Gram-Negative Bacillary Meningitis The third-generation cephalosporins—cefotaxime, ceftriaxone, and ceftazidime—are equally efficacious for the treatment of gram-negative bacillary meningitis, with the exception of meningitis due to *P. aeruginosa*, which should be treated with ceftazidime or meropenem (Table 143-3). A 3-week course of intravenous antibiotic therapy is recommended for meningitis due to gram-negative bacilli.

■ ADJUNCTIVE THERAPY

The release of bacterial cell-wall components by bactericidal antibiotics leads to the production of the inflammatory cytokines IL-1β and TNF-α in the SAS. Dexamethasone exerts its beneficial effect by inhibiting the synthesis of IL-1β and TNF-α at the level of mRNA, decreasing CSF outflow resistance, and stabilizing the blood-brain barrier. The rationale for giving dexamethasone 20 min before antibiotic therapy is that dexamethasone inhibits the production of TNF-α by macrophages and microglia only if it is administered before these cells are activated by endotoxin. Dexamethasone does not alter TNF-α production once it has been induced. The results of clinical trials of dexamethasone therapy in meningitis due to *H. influenzae, S. pneumoniae*, and *N. meningitidis* have demonstrated its efficacy in decreasing meningeal inflammation and neurologic sequelae such as the incidence of sensorineural hearing loss.

A prospective European trial of adjunctive therapy for acute bacterial meningitis in adults found that dexamethasone reduced the number of unfavorable outcomes (15 vs 25%, $p = .03$) including death (7 vs 15%, $p = .04$). The benefits were most striking in patients with pneumococcal meningitis. Dexamethasone (10 mg intravenously) was administered 15–20 min before the first dose of an antimicrobial agent, and the same dose was repeated every 6 h for 4 days. These results were confirmed in a second trial of dexamethasone in adults with pneumococcal meningitis. Therapy with dexamethasone should ideally be started 20 min before, or not later than concurrent with, the first dose of antibiotics. It is unlikely to be of significant benefit if started >6 h after antimicrobial therapy has been initiated. Dexamethasone may decrease the penetration of vancomycin into CSF, and it delays the sterilization of CSF in experimental models of *S. pneumoniae* meningitis. As a result, to assure reliable penetration of vancomycin into the CSF, children and adults are treated with vancomycin in a dose of 45–60 mg/kg per day. Alternatively, vancomycin can be administered by the intraventricular route. In clinical trials, dexamethasone has also been shown to reduce rates of death and hearing loss with no adverse effects in patients with meningococcal meningitis.

One of the concerns for using dexamethasone in adults with bacterial meningitis is that in experimental models of meningitis, dexamethasone therapy increased hippocampal cell injury and reduced learning capacity. This has not been the case in clinical series. The efficacy of dexamethasone therapy in preventing neurologic sequelae is also different between high- and low-income countries. Three large randomized trials in low-income countries (sub-Saharan Africa, Southeast Asia) failed to show benefit in subgroups of patients. The lack of efficacy of dexamethasone in these trials has been attributed to late presentation to the hospital with more advanced disease, antibiotic pretreatment, malnutrition, infection with HIV, and treatment of patients with probable, but not microbiologically proven, bacterial meningitis. The results of these clinical trials suggest that patients in sub-Saharan Africa and those in low-income countries with negative CSF Gram's stain and culture should not be treated with dexamethasone.

■ INCREASED INTRACRANIAL PRESSURE

Emergency treatment of increased ICP includes elevation of the patient's head to 30–45°, intubation, hyperventilation (Pa_{CO_2} 25–30 mmHg), and mannitol. Patients with increased ICP should be managed in an intensive care unit; accurate ICP measurements are best obtained with an ICP monitoring device.

Treatment of increased ICP is discussed in detail in Chap. 318.

■ PROGNOSIS

Mortality rate is 3–7% for meningitis caused by *H. influenzae, N. meningitidis*, or group B streptococci; 15% for that due to *L. monocytogenes*; and 20% for *S. pneumoniae*. In general, the risk of death from bacterial meningitis increases with (1) decreased level of consciousness on admission, (2) onset of seizures within 24 h of admission, (3) signs of increased ICP, (4) young age (infancy) and age >50, (5) the presence of comorbid conditions including shock and/or the need for mechanical ventilation, and (6) delay in the initiation of treatment. Decreased CSF glucose concentration (<2.2 mmol/L [<40 mg/dL]) and markedly increased CSF protein concentration (>3 g/L [> 300 mg/dL]) have been predictive of increased mortality and poorer outcomes in some series. Moderate or severe sequelae occur in ~25% of survivors, although the exact incidence varies with the infecting organism. Common sequelae include decreased intellectual function, memory impairment, seizures, hearing loss and dizziness, and gait disturbances.

VIRAL MENINGITIS

■ CLINICAL MANIFESTATIONS

Immunocompetent adult patients with viral meningitis usually present with headache, fever, and signs of meningeal irritation coupled with an inflammatory CSF profile (see below). Headache is almost invariably present and often characterized as frontal or retroorbital and frequently associated with photophobia and pain on moving the eyes. Nuchal rigidity is present in most cases but may be mild and present only near the limit of neck anteflexion. Constitutional signs can include malaise, myalgia, anorexia, nausea and vomiting, abdominal pain, and/or diarrhea. Patients often have mild lethargy or drowsiness; however, profound alterations in consciousness, such as stupor, coma, or marked confusion, do not occur in viral meningitis and suggest the presence of encephalitis or other alternative diagnoses. Similarly, seizures or focal neurologic signs or symptoms or neuroimaging abnormalities indicative of brain parenchymal involvement are not typical of viral meningitis and suggest the presence of encephalitis or another CNS infectious or inflammatory process.

■ ETIOLOGY

Using a variety of diagnostic techniques, including CSF PCR, culture, and serology, a specific viral cause can be found in 60–90% of cases of

TABLE 143-4 Viruses Causing Acute Meningitis in North America

COMMON	LESS COMMON
Enteroviruses (coxsackieviruses, echoviruses, and the numbered enteroviruses)	Herpes simplex virus 1
Varicella-zoster virus	Human herpesvirus 6
Herpes simplex virus 2	Cytomegalovirus
Epstein-Barr virus	Lymphocytic choriomeningitis virus
Arthropod-borne viruses (notably WNV)	Mumps
HIV	Zika and other non-WNV arboviruses

Abbreviation: WNV, West Nile virus.

viral meningitis. The most important agents are enteroviruses (including echoviruses and coxsackieviruses in addition to numbered enteroviruses), varicella-zoster virus (VZV), HSV (HSV-2 > HSV-1), HIV, and arboviruses **(Table 143-4)**. CSF cultures are positive in 30–70% of patients, with the frequency of isolation depending on the specific viral agent. Approximately two-thirds of culture-negative cases of “aseptic” meningitis have a specific viral etiology identified by CSF PCR testing (see below).

EPIDEMIOLOGY

Viral meningitis is not a nationally reportable disease; however, it has been estimated that the incidence is at least 75,000 cases per year in the United States. In temperate climates, there is a substantial increase in cases during the nonwinter months, reflecting the seasonal predominance of enterovirus and arthropod-borne virus (arbovirus) infections in the summer and fall, with a peak monthly incidence of about 1 reported case per 100,000 population.

LABORATORY DIAGNOSIS

CSF Examination The most important laboratory test in the diagnosis of viral meningitis is examination of the CSF. The typical profile is a pleocytosis, a normal or slightly elevated protein concentration (0.2–0.8 g/L [20–80 mg/dL]), a normal glucose concentration, and a normal or mildly elevated opening pressure (100–350 mmH_2O). Organisms are *not* seen on Gram’s stain of CSF. The total CSF cell count in viral meningitis is typically 25–500/μL, although cell counts of several thousand/μL are occasionally seen, especially with infections due to lymphocytic choriomeningitis virus (LCMV) and mumps virus. Lymphocytes are typically the predominant cell. Rarely, PMNs may predominate in the first 48 h of illness, especially with infections due to echovirus 9, West Nile virus (WNV), eastern equine encephalitis (EEE) virus, or mumps. A PMN pleocytosis occurs in 45% of patients with WNV meningitis and can persist for a week or longer before shifting to a lymphocytic pleocytosis. PMN pleocytosis with low glucose may also be a feature of cytomegalovirus (CMV) infections in immunocompromised hosts. Despite these exceptions, the presence of a CSF PMN pleocytosis in a patient with suspected viral meningitis in whom a specific diagnosis has not been established should prompt consideration of alternative diagnoses, including bacterial meningitis or parameningeal infections. The CSF glucose concentration is typically normal in viral infections, although it may be decreased in 10–30% of cases due to mumps or LCMV. Rare instances of decreased CSF glucose concentration occur in cases of meningitis due to echoviruses and other enteroviruses, HSV-2, and VZV. As a rule, a lymphocytic pleocytosis with a low glucose concentration should suggest fungal or tuberculous meningitis, *Listeria* meningoencephalitis, or noninfectious disorders (e.g., sarcoid, neoplastic meningitis).

A number of tests measuring levels of various CSF proteins, enzymes, and mediators—including C-reactive protein, lactic acid, lactate dehydrogenase, neopterin, quinolinate, IL-1β, IL-6, soluble IL-2 receptor, β_2-microglobulin, and TNF—have been proposed as potential discriminators between viral and bacterial meningitis or as markers of specific types of viral infection (e.g., infection with HIV), but they remain of uncertain sensitivity and specificity and are not widely used for diagnostic purposes.

Polymerase Chain Reaction Amplification of Viral Nucleic Acid Amplification of viral-specific DNA or RNA from CSF using PCR amplification has become the single most important method for diagnosing CNS viral infections. In both enteroviral and HSV infections of the CNS, CSF PCR has become the diagnostic procedure of choice and is substantially more sensitive than viral cultures. HSV CSF PCR is also an important diagnostic test in patients with recurrent episodes of “aseptic” meningitis, many of whom have amplifiable HSV DNA in CSF despite negative viral cultures. The FilmArray Meningitis/Encephalitis panel (BioFire Diagnostics) includes HSV-1, HSV-2, enteroviruses, VZV, human herpesvirus 6 (HHV-6), and human *Parechovirus*. The panel does not include CMV, WNV, LCMV, or Epstein-Barr virus (EBV). CSF PCR tests are available for WNV but are not as sensitive as detection of WNV-specific CSF IgM. PCR is also useful in the diagnosis of CNS infection caused by *Mycoplasma pneumoniae*, which can mimic viral meningitis and encephalitis. PCR of throat washings may assist in diagnosis of enteroviral and mycoplasmal CNS infections. PCR of stool specimens may also assist in diagnosis of enteroviral infections (see below).

Viral Culture The sensitivity of CSF cultures for the diagnosis of viral meningitis and encephalitis, in contrast to its utility in bacterial infections, is generally poor. In addition to CSF, specific viruses may also be isolated from throat swabs, stool, blood, and urine. Enteroviruses and adenoviruses may be found in feces; arboviruses, some enteroviruses, and LCMV in blood; mumps and CMV in urine; and enteroviruses, mumps, and adenoviruses in throat washings. During enteroviral infections, viral shedding in stool may persist for several weeks. The presence of enterovirus in stool is not diagnostic and may result from residual shedding from a previous enteroviral infection; it also occurs in some asymptomatic individuals during enteroviral epidemics.

Serologic Studies The basic approach to the serodiagnosis of viral meningitis is identical to that for viral encephalitis **(see Chap. 142)**. Serologic studies are important for the diagnosis of arboviruses such as WNV; however, these tests are less useful for viruses such as HSV, VZV, CMV, and EBV that have a high seroprevalence in the general population.

CSF oligoclonal γ globulin bands occur in association with a number of viral infections. The associated antibodies are often directed against viral proteins. Oligoclonal bands also occur commonly in certain noninfectious neurologic diseases (e.g., multiple sclerosis) and may be found in nonviral infections (e.g., neurosyphilis, Lyme neuroborreliosis).

Other Laboratory Studies All patients with suspected viral meningitis should have a complete blood count and differential, liver and renal function tests, erythrocyte sedimentation rate (ESR), C-reactive protein, electrolytes, glucose, creatine kinase, aldolase, amylase, and lipase. Neuroimaging studies (MRI preferable to CT) are not absolutely necessary in patients with uncomplicated viral meningitis but should be performed in patients with altered consciousness, seizures, focal neurologic signs or symptoms (see “Differential Diagnosis” below), atypical CSF profiles, or underlying immunocompromising treatments or conditions.

DIFFERENTIAL DIAGNOSIS

The most important issue in the differential diagnosis of viral meningitis is to consider diseases that can mimic viral meningitis, including (1) untreated or partially treated bacterial meningitis; (2) early stages of meningitis caused by fungi, mycobacteria, or *Treponema pallidum* (neurosyphilis), in which a lymphocytic pleocytosis is common, cultures may be slow growing or negative, and hypoglycorrhachia may not be present early; (3) meningitis caused by agents such as *Mycoplasma, Listeria* spp., *Brucella* spp., *Coxiella* spp., *Leptospira* spp., and *Rickettsia* spp.; (4) parameningeal infections; (5) neoplastic meningitis;

and (6) meningitis secondary to noninfectious inflammatory diseases, including medication-induced hypersensitivity meningitis, SLE and other rheumatologic diseases, sarcoidosis, Behçet's syndrome, and the uveomeningitic syndromes. Studies in children >28 days of age suggest that the presence of CSF protein >0.5 g/L (sensitivity 89%, specificity 78%) and elevated serum procalcitonin levels >0.5 ng/mL (sensitivity 89%, specificity 89%) were clues to the presence of bacterial as opposed to "aseptic" meningitis. A variety of clinical algorithms for differentiating bacterial from aseptic meningitis have been developed. One such prospectively validated system, the *bacterial meningitis score*, suggests that the probability of bacterial meningitis is 0.3% or less (negative predictive value 99.7%, 95% confidence interval 99.6–100%) in children with CSF pleocytosis who have (1) a negative CSF Gram's stain, (2) CSF neutrophil count <1000 cells/μL, (3) CSF protein <80 mg/dL, (4) peripheral absolute neutrophil count of <10,000 cells/μL, and (5) no prior history or current presence of seizures.

SPECIFIC VIRAL ETIOLOGIES

Enteroviruses (EVs) **(Chap. 210)** are the most common cause of viral meningitis, accounting for >85% of cases in which a specific etiology can be identified. Cases may either be sporadic or occur in clusters. EV71 has produced large epidemics of neurologic disease outside the United States, especially in Southeast Asia, but most recently reported cases in the United States have been sporadic. Enteroviruses are the most likely cause of viral meningitis in the summer and fall months, especially in children (<15 years), although cases occur at reduced frequency year-round. Although the incidence of enteroviral meningitis declines with increasing age, some outbreaks have preferentially affected older children and adults. Meningitis outside the neonatal period is usually benign. Patients present with sudden onset of fever; headache; nuchal rigidity; and often constitutional signs, including vomiting, anorexia, diarrhea, cough, pharyngitis, and myalgias. The physical examination should include a careful search for stigmata of EV infection, including exanthems, hand-foot-mouth disease, herpangina, pleurodynia, myopericarditis, and hemorrhagic conjunctivitis. The CSF profile is typically a lymphocytic pleocytosis (100–1000 cells/μL) with normal glucose and normal or mildly elevated protein concentration. However, up to 15% of patients, most commonly young infants rather than older children or adults, have a normal CSF leukocyte count. In rare cases, PMNs may predominate during the first 48 h of illness. CSF reverse transcriptase PCR (RT-PCR) is the diagnostic procedure of choice and is both sensitive (>95%) and specific (>100%). CSF RT-PCR has the highest sensitivity if performed within 48 h of symptom onset, with sensitivity declining rapidly after day 5 of symptoms. RT-PCR of throat washings or stool specimens may be positive for several weeks, and positive results can help support the diagnosis of an acute enteroviral infection. The sensitivity of routine enteroviral RT-PCRs for detecting EV71 is low, and specific testing may be required. Treatment is supportive, and patients usually recover without sequelae. Chronic and severe infections can occur in neonates and in individuals with hypo- or agammaglobulinemia.

Arbovirus infections **(Chap. 215)** occur predominantly in the summer and early fall. Arboviral meningitis should be considered when clusters of meningitis and encephalitis cases occur in a restricted geographic region during the summer or early fall. In the United States, the most important causes of arboviral meningitis and encephalitis are WNV, St. Louis encephalitis virus, and the California encephalitis group of viruses. In WNV epidemics, avian deaths may serve as sentinel infections for subsequent human disease. A history of tick exposure or travel or residence in the appropriate geographic area should suggest the possibility of Colorado tick fever virus or Powassan virus infection, although nonviral tick-borne diseases, including RMSF and Lyme neuroborreliosis, may present similarly. Arbovirus meningitis is typically associated with a CSF lymphocytic pleocytosis, normal glucose concentration, and normal or mildly elevated protein concentration. However, ~45% of patients with WNV meningitis have CSF neutrophilia, which can persist for a week or more. The rarity of hypoglycorrhachia in WNV infection, the absence of positive Gram's stains, and the negative cultures help distinguish these patients from those with bacterial meningitis. Definitive diagnosis of arboviral meningitis is based on demonstration of viral-specific IgM in CSF or seroconversion. The prevalence of CSF IgM increases progressively during the first week after infection, peaking at >80% in patients with neuroinvasive disease; as a result, repeat studies may be needed when disease suspicion is high and an early study is negative. CSF RT-PCR tests are available for some viruses in selected diagnostic laboratories and at the Centers for Disease Control and Prevention (CDC), but in the case of WNV, sensitivity (~70%) of CSF RT-PCR is less than that of CSF serology. WNV CSF RT-PCR may be useful in immunocompromised patients who may have absent or reduced antibody responses.

HSV meningitis **(Chap. 197)** has been increasingly recognized as a major cause of viral meningitis in adults, and overall, it is probably second in importance to enteroviruses as a cause of viral meningitis, accounting for 5% of total cases overall and undoubtedly a higher frequency of those cases occurring in adults and/or outside of the summer-fall period when enterovirus infections are increasingly common. In adults, the majority of cases of uncomplicated meningitis are due to HSV-2, whereas HSV-1 is responsible for 90% of cases of HSV encephalitis. HSV meningitis occurs in ~25–35% of women and ~10–15% of men at the time of an initial (primary) episode of genital herpes. Of these patients, 20% go on to have recurrent attacks of meningitis. Diagnosis of HSV meningitis is usually by HSV CSF PCR because cultures may be negative, especially in patients with recurrent meningitis. Demonstration of intrathecal synthesis of HSV-specific antibody may also be useful in diagnosis, although antibody tests are less sensitive and less specific than PCR and may not become positive until after the first week of infection. Although a history of or the presence of HSV genital lesions is an important diagnostic clue, many patients with HSV meningitis give no history and have no evidence of active genital herpes at the time of presentation. Most cases of recurrent viral or "aseptic" meningitis, including cases previously diagnosed as Mollaret's meningitis, are due to HSV.

VZV meningitis **(Chap. 198)** should be suspected in the presence of concurrent chickenpox or shingles. However, it is important to recognize that VZV is being increasingly identified as an important cause of both meningitis and encephalitis in patients without rash. The frequency of VZV as a cause of meningitis is extremely variable, ranging from as low as 3% to as high as 20% in different series. Diagnosis is usually based on CSF PCR, although the sensitivity of this test is not as high as for the other herpesviruses. VZV serologic studies complement PCR testing, and the diagnosis of VZV CNS infection can be made by the demonstration of VZV-specific intrathecal antibody synthesis and/or the presence of VZV CSF IgM antibodies, or by positive CSF cultures.

EBV infections **(Chap. 199)** may also produce aseptic meningitis, with or without associated infectious mononucleosis. The presence of atypical lymphocytes in the CSF or peripheral blood is suggestive of EBV infection but may occasionally be seen with other viral infections. EBV is almost never cultured from CSF. Serum and CSF serology help establish the presence of acute infection, which is characterized by IgM viral capsid antibodies (VCAs), antibodies to early antigens (EAs), and the absence of antibodies to EBV-associated nuclear antigen (EBNA). CSF PCR is another important diagnostic test, although false-positive results may reflect viral reactivation associated with other infectious or inflammatory processes or the presence of latent viral DNA in lymphocytes recruited due to other inflammatory conditions.

HIV meningitis should be suspected in any patient presenting with a viral meningitis with known or suspected risk factors for HIV infection. Meningitis may occur following primary infection with HIV in 5–10% of cases and less commonly at later stages of illness. Cranial nerve palsies, most commonly involving cranial nerves V, VII, or VIII, are more common in HIV meningitis than in other viral infections. Diagnosis can be confirmed by detection of HIV genome in blood or CSF. Seroconversion may be delayed, and patients with negative HIV serologies who are suspected of having HIV meningitis should be monitored for delayed seroconversion. **For further discussion of HIV infection, see Chap. 208.**

Mumps **(Chap. 213)** should be considered when meningitis occurs in the late winter or early spring, especially in males (male-to-female ratio 3:1). With the widespread use of the live attenuated mumps vaccine in the United States since 1967, the incidence of mumps meningitis has fallen by >95%; however, mumps remains a potential source of infection in nonimmunized individuals and populations, but also in those who have been vaccinated but may have waning immunity. Rare cases (10–100/100,000 vaccinated individuals) of vaccine-associated mumps meningitis have been described, with onset typically 2–4 weeks after vaccination. The presence of parotitis, orchitis, oophoritis, pancreatitis, or elevations in serum lipase and amylase is suggestive of mumps meningitis; however, their absence does not exclude the diagnosis. Clinical meningitis has been estimated to occur in 10–30% of patients with mumps parotitis. Mumps infection confers lifelong immunity, so a documented history of previous infection excludes this diagnosis. A CSF pleocytosis that can exceed 1000 cells/μL is present in 25% of patients with mumps meningitis. Lymphocytes predominate in 75%, although CSF neutrophilia occurs in 25%. Hypoglycorrhachia occurs in 10–30% of patients and may be a clue to the diagnosis when present. Diagnosis is typically made by culture of virus from CSF or by detecting IgM antibodies or seroconversion. CSF PCR is available in some diagnostic and research laboratories.

LCMV infection **(Chap. 215)** should be considered when aseptic meningitis occurs in the late fall or winter and in individuals with a history of exposure to house mice (*Mus musculus*), pet or laboratory rodents (e.g., hamsters, rats, mice), or their excreta. Some patients have an associated rash, pulmonary infiltrates, alopecia, parotitis, orchitis, or myopericarditis. Laboratory clues to the diagnosis of LCMV, in addition to the clinical findings noted above, may include the presence of leukopenia, thrombocytopenia, or abnormal liver function tests. Some cases present with a marked CSF pleocytosis (>1000 cells/μL) and hypoglycorrhachia (<30%). Diagnosis is based on serology and/or culture of virus from CSF.

TREATMENT

Acute Viral Meningitis

Treatment of almost all cases of viral meningitis is primarily symptomatic and includes use of analgesics, antipyretics, and antiemetics. Fluid and electrolyte status should be monitored. Patients with suspected bacterial meningitis should receive appropriate empirical therapy pending culture results (see above). Hospitalization may not be required in immunocompetent patients with presumed viral meningitis and no focal signs or symptoms, no significant alteration in consciousness, and a classic CSF profile (lymphocytic pleocytosis, normal glucose, negative Gram's stain) if adequate provision for monitoring at home and medical follow-up can be ensured. Immunocompromised patients; patients with significant alteration in consciousness, seizures, or the presence of focal signs and symptoms suggesting the possibility of encephalitis or parenchymal brain involvement; and patients who have an atypical CSF profile should be hospitalized. Oral or intravenous acyclovir may be of benefit in patients with meningitis caused by HSV-1 or -2 and in cases of severe EBV or VZV infection. Data concerning treatment of HSV, EBV, and VZV meningitis are extremely limited. Seriously ill patients should probably receive intravenous acyclovir (15–30 mg/kg per day in three divided doses), which can be followed by an oral drug such as acyclovir (800 mg five times daily), famciclovir (500 mg tid), or valacyclovir (1000 mg tid) for a total course of 7–14 days. Patients who are less ill can be treated with oral drugs alone. Patients with HIV meningitis should receive antiretroviral therapy **(Chap. 208)**. There is no specific therapy of proven benefit for patients with arboviral encephalitis, including that caused by WNV.

Patients with viral meningitis who are known to have deficient humoral immunity (e.g., X-linked agammaglobulinemia) and who are not already receiving either intramuscular γ globulin or intravenous immunoglobulin (IVIg) should be treated with these agents. Intraventricular administration of immunoglobulin through an Ommaya reservoir has been tried in some patients with chronic enteroviral meningitis who have not responded to intramuscular or intravenous immunoglobulin.

Vaccination is an effective method of preventing the development of meningitis and other neurologic complications associated with poliovirus, mumps, measles, rubella, and varicella infection. A recombinant zoster vaccine (RSV, Shingrix) contains recombinant VZV glycoprotein E in combination with an adjuvant (ASO1$_B$) and has greater efficacy in preventing zoster in adults aged ≥70 years than the previously recommended live attenuated vaccine (Zostrax). The Advisory Committee on Immunization Practices recommends the use of the recombinant zoster vaccine in immunocompetent adults aged ≥50 years and in immunodeficient or immunosuppressed adults ≥19 years of age.

PROGNOSIS

In adults, the prognosis for full recovery from viral meningitis is excellent. Rare patients complain of persisting headache, mild mental impairment, incoordination, or generalized asthenia for weeks to months. The outcome in infants and neonates (<1 year) is less certain; intellectual impairment, learning disabilities, hearing loss, and other lasting sequelae have been reported in some studies.

SUBACUTE MENINGITIS

CLINICAL MANIFESTATIONS

Patients with subacute meningitis typically have an unrelenting headache, stiff neck, low-grade fever, and lethargy for days to several weeks before they present for evaluation. Cranial nerve abnormalities and night sweats may be present. **This syndrome overlaps that of chronic meningitis, discussed in detail in Chap. 144 but is included here because the meningeal pathogens of subacute meningitis can also present as an acute meningitis.**

ETIOLOGY

Common causative organisms include *M. tuberculosis, C. neoformans, H. capsulatum, C. immitis*, and *T. pallidum*. Initial infection with *M. tuberculosis* is acquired by inhalation of aerosolized droplet nuclei. Tuberculous meningitis in adults does not develop acutely from hematogenous spread of tubercle bacilli to the meninges. Rather, millet seed–sized (miliary) tubercles form in the parenchyma of the brain during hematogenous dissemination of tubercle bacilli in the course of primary infection. These tubercles enlarge and are usually caseating. The propensity for a caseous lesion to produce meningitis is determined by its proximity to the SAS and the rate at which fibrous encapsulation develops. Subependymal caseous foci cause meningitis via discharge of bacilli and tuberculous antigens into the SAS. Mycobacterial antigens produce an intense inflammatory reaction that leads to the production of a thick exudate that fills the basilar cisterns and surrounds the cranial nerves and major blood vessels at the base of the brain.

Fungal infections are typically acquired by the inhalation of airborne fungal spores. The initial pulmonary infection may be asymptomatic or present with fever, cough, sputum production, and chest pain. The pulmonary infection is often self-limited. A localized pulmonary fungal infection can then remain dormant in the lungs until there is an abnormality in cell-mediated immunity that allows the fungus to reactivate and disseminate to the CNS. The most common pathogen causing fungal meningitis is *C. neoformans*. This fungus is found worldwide in soil and bird excreta. *H. capsulatum* is endemic to the Ohio and Mississippi River valleys of the central United States and to parts of Central and South America. *C. immitis* is endemic to the desert areas of the southwest United States, northern Mexico, and Argentina.

Syphilis is a sexually transmitted disease that is manifest by the appearance of a painless chancre at the site of inoculation. *T. pallidum* invades the CNS early in the course of syphilis. Cranial nerves VII and VIII are most frequently involved.

LABORATORY DIAGNOSIS

The classic CSF abnormalities in tuberculous meningitis are as follows: (1) elevated opening pressure, (2) lymphocytic pleocytosis (10–500 cells/μL), (3) elevated protein concentration in the range of 1–5 g/L, and (4) decreased glucose concentration in the range of 1.1–2.2 mmol/L (20–40 mg/dL). *The combination of unrelenting headache, stiff neck, fatigue, night sweats, and fever with a CSF lymphocytic pleocytosis and a mildly decreased glucose concentration is highly suspicious for tuberculous meningitis.* The last tube of fluid collected at LP is the best tube to send for a smear for acid-fast bacilli (AFB). If there is a pellicle in the CSF or a cobweb-like clot on the surface of the fluid, AFB can best be demonstrated in a smear of the clot or pellicle. Positive smears are typically reported in only 10–40% of cases of tuberculous meningitis in adults. Cultures of CSF take 4–8 weeks to identify the organism and are positive in ~50% of adults. Culture remains the gold standard to make the diagnosis of tuberculous meningitis. Nucleic acid amplification tests for the detection of *M. tuberculosis* DNA should be sent on CSF if available. The XpertMTB/RIF Ultra and the Xpert MTB/RIF are used most commonly, but the consensus of opinion is that the sensitivity, and thus the risk of false-negative results, has not been defined.

The characteristic CSF abnormalities in fungal meningitis are a mononuclear or lymphocytic pleocytosis, an increased protein concentration, and a decreased glucose concentration. There may be eosinophils in the CSF in *C. immitis* meningitis. Large volumes of CSF are often required to demonstrate the organism on India ink smear or grow the organism in culture. If spinal fluid examined by LP on two separate occasions fails to yield an organism, CSF should be obtained by high-cervical or cisternal puncture.

The cryptococcal polysaccharide antigen test is a highly sensitive and specific test for cryptococcal meningitis. A reactive CSF cryptococcal antigen test establishes the diagnosis. The detection of the *Histoplasma* polysaccharide antigen in CSF establishes the diagnosis of a fungal meningitis but is not specific for meningitis due to *H. capsulatum*. It may be falsely positive in coccidioidal meningitis. The CSF complement fixation antibody test is reported to have a specificity of 100% and a sensitivity of 75% for coccidioidal meningitis.

The diagnosis of syphilitic meningitis is made when a reactive serum treponemal test (fluorescent treponemal antibody absorption test [FTA-ABS] or microhemagglutination assay–*T. pallidum* [MHA-TP]) is associated with a CSF lymphocytic or mononuclear pleocytosis and an elevated protein concentration, or when the CSF Venereal Disease Research Laboratory (VDRL) test is positive. A reactive CSF FTA-ABS is not definitive evidence of neurosyphilis. The CSF FTA-ABS can be falsely positive from blood contamination. A negative CSF VDRL does not rule out neurosyphilis. A negative CSF FTA-ABS or MHA-TP rules out neurosyphilis.

TREATMENT

Subacute Meningitis

Empirical therapy of tuberculous meningitis is often initiated based on a high index of suspicion without adequate laboratory support. Initial therapy is a combination of isoniazid (300 mg/d), rifampin (10 mg/kg per day), pyrazinamide (30 mg/kg per day in divided doses), ethambutol (15–25 mg/kg per day in divided doses), and pyridoxine (50 mg/d). When the antimicrobial sensitivity of the *M. tuberculosis* isolate is known, ethambutol can be discontinued. If the clinical response is good, pyrazinamide can be discontinued after 8 weeks and isoniazid and rifampin continued alone for the next 6–12 months. A 6-month course of therapy is acceptable, but therapy should be prolonged for 9–12 months in patients who have an inadequate resolution of symptoms of meningitis or who have positive mycobacterial cultures of CSF during the course of therapy. Dexamethasone therapy is recommended for HIV-negative patients with tuberculous meningitis. The dose is 12–16 mg/d for 3 weeks, and then tapered over 3 weeks.

Meningitis due to *C. neoformans* in non-HIV, nontransplant patients is treated with induction therapy with amphotericin B (AmB) (0.7 mg/kg IV per day) plus flucytosine (100 mg/kg per day in four divided doses) for at least 4 weeks if CSF culture results are negative after 2 weeks of treatment. Therapy should be extended for a total of 6 weeks in the patient with neurologic complications. Induction therapy is followed by consolidation therapy with fluconazole 400 mg/d for 8 weeks. Organ transplant recipients are treated with liposomal AmB (3–4 mg/kg per day) or AmB lipid complex (ABLC; 5 mg/kg per day) plus flucytosine (100 mg/kg per day in four divided doses) for at least 2 weeks or until CSF culture is sterile. Follow CSF yeast cultures for sterilization rather than the cryptococcal antigen titer. This treatment is followed by an 8- to 10-week course of fluconazole (400–800 mg/d [6–12 mg/kg] PO). If the CSF culture is sterile after 10 weeks of acute therapy, the dose of fluconazole is decreased to 200 mg/d for 6 months to a year. Patients with HIV infection are treated with AmB or a lipid formulation plus flucytosine for at least 2 weeks, followed by fluconazole for a minimum of 8 weeks. HIV-infected patients may require indefinite maintenance therapy with fluconazole 200 mg/d. Meningitis due to *H. capsulatum* is treated with AmB (0.7–1.0 mg/kg per day) for 4–12 weeks. A total dose of 30 mg/kg is recommended. Therapy with AmB is not discontinued until fungal cultures are sterile. After completing a course of AmB, maintenance therapy with itraconazole 200 mg two or three times daily is initiated and continued for at least 9 months to a year. *C. immitis* meningitis is treated with either high-dose fluconazole (1000 mg daily) as monotherapy or intravenous AmB (0.5–0.7 mg/kg per day) for >4 weeks. Intrathecal AmB (0.25–0.75 mg/d three times weekly) may be required to eradicate the infection. Lifelong therapy with fluconazole (200–400 mg daily) is recommended to prevent relapse. AmBisome (5 mg/kg per day) or amphotericin B lipid complex (ABLC; 5 mg/kg per day) can be substituted for AmB in patients who have or who develop significant renal dysfunction. The most common complication of fungal meningitis is hydrocephalus. Patients who develop hydrocephalus should receive a CSF diversion device. A ventriculostomy can be used until CSF fungal cultures are sterile, at which time the ventriculostomy is replaced by a ventriculoperitoneal shunt.

Syphilitic meningitis is treated with aqueous penicillin G in a dose of 3–4 million units intravenously every 4 h for 10–14 days. An alternative regimen is 2.4 million units of procaine penicillin G intramuscularly daily with 500 mg of oral probenecid four times daily for 10–14 days. Either regimen is followed with 2.4 million units of benzathine penicillin G intramuscularly once a week for 3 weeks. The standard criterion for treatment success is reexamination of the CSF. The CSF should be reexamined at 6-month intervals for 2 years. The cell count is expected to normalize within 12 months, and the VDRL titer to decrease by two dilutions or revert to nonreactive within 2 years of completion of therapy. Failure of the CSF pleocytosis to resolve or an increase in the CSF VDRL titer by two or more dilutions requires retreatment.

FURTHER READING

Gundamraj V et al: Viral meningitis and encephalitis: An update. Curr Opin Infect Dis 36:177, 2023.

Krett JD et al: Neurology of acute viral infections. Neurohospitalist 12:632, 2022.

Pfefferle S et al: Implementation of the FilmArray ME panel in laboratory routine using a simple sample selection strategy for diagnosis of meningitis and encephalitis. BMC Infect Dis 20:170, 2020.

Roos KL et al: Acute bacterial meningitis, in *Infections of the Central Nervous System*, 4th ed. Scheld WM, Whitley RJ, Marra (eds). Philadelphia, Wolters Kluwer Health, 2014, pp. 365–419.

Van de Beek D et al: Community acquired bacterial meningitis. Lancet 398:1171, 2021.

144 Chronic and Recurrent Meningitis

Avindra Nath, Walter J. Koroshetz, Michael R. Wilson

Chronic inflammation of the meninges (pia, arachnoid, and dura) can produce profound neurologic disability and may be fatal if not successfully treated. Chronic meningitis is diagnosed when a characteristic neurologic syndrome exists for >4 weeks and is associated with a persistent inflammatory response in the cerebrospinal fluid (CSF) (white cell count >5/μL). The causes are varied, and appropriate treatment depends on identification of the etiology. Five categories of disease account for most cases of chronic meningitis: (1) meningeal infections, (2) malignancy, (3) autoimmune inflammatory disorders, (4) chemical meningitis, and (5) parameningeal infections. In addition, there is increasing recognition that some patients with recurrent meningitis may have monogenic autoinflammatory disorders.

CLINICAL PATHOPHYSIOLOGY

Neurologic manifestations of chronic meningitis **(Table 144-1)** are determined by the anatomic location of the inflammation and its consequences. Persistent headache, clinical signs of hydrocephalus, cranial neuropathies, radiculopathies, and cognitive or personality changes are the cardinal features. Meningeal signs are uncommon in chronic meningitis. These manifestations can occur alone or in combination. When they appear in combination, it may be indicative of widespread dissemination of the inflammatory process along CSF pathways. In some cases, the presence of an underlying systemic illness points to the probable cause of the meningitis **(Fig. 144-1)**. The diagnosis of chronic meningitis is usually made when the clinical presentation prompts the physician to examine the CSF for signs of inflammation. CSF is produced by the choroid plexus of the cerebral ventricles, exits through the foramina in the fourth ventricle into the subarachnoid space surrounding the brain and spinal cord, circulates around the base of the brain and over the cerebral hemispheres, and is resorbed by arachnoid villi projecting into the superior sagittal sinus where it mixes with blood in the venous sinuses. Recently, a cerebral lymphatic system has been identified that drains the dura mater **(Chap. 435)**; however, its role in chronic meningitis has not been studied. CSF flow provides a pathway for rapid spread of infectious and other infiltrative processes over the brain, spinal cord, cranial, and spinal nerve roots. Spread from the subarachnoid space into brain parenchyma may occur via the arachnoid cuffs that surround blood vessels that penetrate brain tissue (Virchow-Robin spaces). Microorganisms can travel along these perivascular spaces to enter the brain.

TABLE 144-1 Symptoms and Signs of Chronic Meningitis

SYMPTOM	SIGN
Chronic headache	± Papilledema
Neck or back pain/stiffness	Brudzinski's or Kernig's sign of meningeal irritation
Change in personality	Altered mental status—drowsiness, inattention, disorientation, memory loss, frontal release signs (grasp, suck, snout), perseveration
Facial weakness	Peripheral seventh CN paresis
Double vision	Paresis of CNs III, IV, and/or VI
Diminished vision	Papilledema, CN II (optic atrophy/inflammation)
Hearing loss	Eighth CN paresis
Arm or leg weakness	Myelopathy or radiculopathy
Numbness in arms or legs	Myelopathy or radiculopathy
Urinary retention/ incontinence	Myelopathy or radiculopathy Frontal lobe dysfunction (hydrocephalus)
Clumsiness	Ataxia

Abbreviation: CN, cranial nerve.

Intracranial Meningitis In intracranial meningitis, nociceptive nerve fibers of the meninges are stimulated by the inflammatory process, resulting in headache, neck pain, or back pain. Obstruction of CSF pathways at the cerebral aqueduct or arachnoid villi may produce hydrocephalus and signs and symptoms of raised intracranial pressure (ICP), including headache, vomiting, apathy or drowsiness, gait instability, papilledema, visual loss, impaired upgaze, or palsy of the sixth cranial nerve (CN VI). Cognitive and behavioral changes during the course of chronic meningitis may also result from vascular damage due to inflammation around the blood vessels in the subarachnoid space, causing infarction. Inflammatory deposits seeded via the CSF circulation are often prominent around the brainstem and cranial nerves and along the undersurface of the frontal and temporal lobes. Such cases, termed *basal meningitis*, often present as multiple cranial neuropathies **(Chap. 452)**, with some combination of decreased vision (CN II), facial weakness (CN VII), decreased hearing (CN VIII), diplopia (CNs III, IV, and VI), sensory or motor abnormalities of the oropharynx (CNs IX, X, and XII), decreased olfaction (CN I), or decreased facial sensation and masseter weakness (CN V). Involvement of the lower CNs is more common because the inflammatory exudate tends to collect at the base of the brain.

Spinal Meningitis In spinal meningitis, injury may occur to motor and sensory nerve roots as they traverse the subarachnoid space and penetrate the meninges. These cases present as multiple radiculopathies with combinations of radicular pain, sensory loss, motor weakness, and urinary or fecal incontinence. In some cases, chronic inflammation causes arachnoiditis with clumping of the lower nerve roots and thickening of the meninges. Preferential involvement of the lower nerve roots results from inflammatory cells that gravitate to the bottom of the intrathecal space. Meningeal inflammation can encircle and damage the spinal cord, resulting in a myelopathy. Slow progressive involvement of multiple CNs and/or spinal nerve roots is likely due to chronic meningitis. Electrophysiologic testing (electromyography, nerve conduction studies, and evoked response testing) may be helpful in determining whether there is involvement of cranial and spinal nerve roots.

Systemic Manifestations In some patients, evidence of systemic disease provides clues to the underlying cause of chronic meningitis. A complete history of travel, sexual exposure, insect bites, and other modes of exposure to infectious agents should be sought. Infectious causes are often associated with fever, malaise, anorexia, and signs of localized or disseminated infection outside the nervous system. Infectious causes are of major concern in immunosuppressed patients and especially in patients with untreated HIV infection, in whom chronic meningitis is most often caused by *Mycobacterium tuberculosis* or *Cryptococcus neoformans* and may present without headache, fever, or meningeal signs. In this population, a high index of clinical suspicion needs to be maintained even when there is only mild confusion or a nonspecific headache syndrome even in the context of a pauicellular CSF profile. Noninfectious inflammatory disorders most often produce systemic manifestations first, but meningitis may be the initial manifestation. Carcinomatous meningitis, caused by CSF seeding with metastatic cancer cells, may or may not be accompanied by clinical evidence of the primary neoplasm.

APPROACH TO THE PATIENT

Chronic Meningitis

The occurrence of chronic or worsening headache, typically constant and nonfocal, clinical signs of hydrocephalus, cranial neuropathy, radiculopathy, and/or cognitive decline in a patient with or without fever should prompt consideration of a lumbar puncture for evidence of meningeal inflammation. On occasion, the diagnosis is made when a contrast-enhanced imaging study

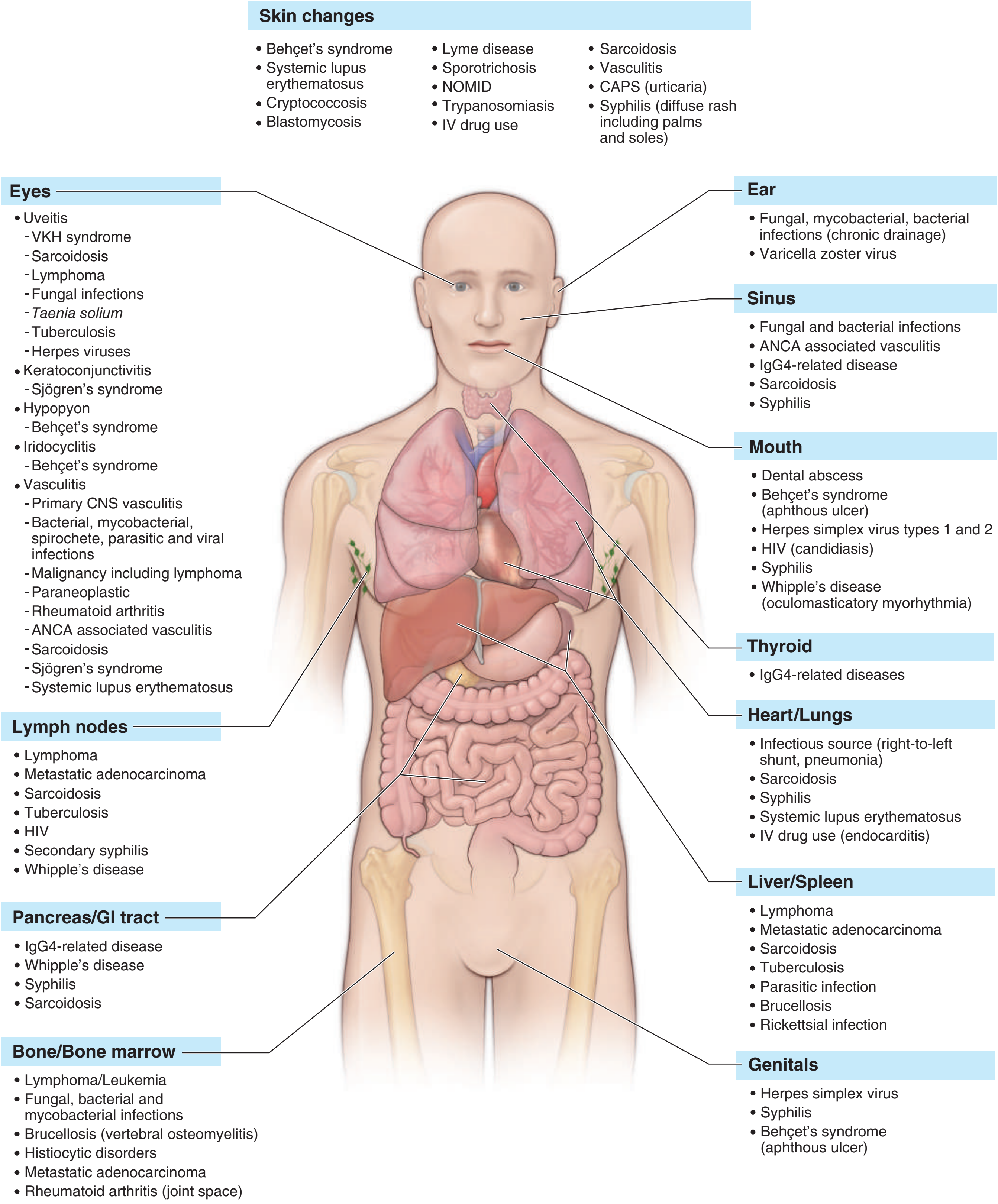
Skin changes
• Behçet's syndrome
• Systemic lupus erythematosus
• Cryptococcosis
• Blastomycosis
• Lyme disease
• Sporotrichosis
• NOMID
• Trypanosomiasis
• IV drug use
• Sarcoidosis
• Vasculitis
• CAPS (urticaria)
• Syphilis (diffuse rash including palms and soles)
Eyes
• Uveitis
-VKH syndrome
-Sarcoidosis
-Lymphoma
-Fungal infections
-*Taenia solium*
-Tuberculosis
-Herpes viruses
• Keratoconjunctivitis
-Sjögren's syndrome
• Hypopyon
-Behçet's syndrome
• Iridocyclitis
-Behçet's syndrome
• Vasculitis
-Primary CNS vasculitis
-Bacterial, mycobacterial, spirochete, parasitic and viral infections
-Malignancy including lymphoma
-Paraneoplastic
-Rheumatoid arthritis
-ANCA associated vasculitis
-Sarcoidosis
-Sjögren's syndrome
-Systemic lupus erythematosus
Lymph nodes
• Lymphoma
• Metastatic adenocarcinoma
• Sarcoidosis
• Tuberculosis
• HIV
• Secondary syphilis
• Whipple's disease
Pancreas/GI tract
• IgG4-related disease
• Whipple's disease
• Syphilis
• Sarcoidosis
Bone/Bone marrow
• Lymphoma/Leukemia
• Fungal, bacterial and mycobacterial infections
• Brucellosis (vertebral osteomyelitis)
• Histiocytic disorders
• Metastatic adenocarcinoma
• Rheumatoid arthritis (joint space)
Ear
• Fungal, mycobacterial, bacterial infections (chronic drainage)
• Varicella zoster virus
Sinus
• Fungal and bacterial infections
• ANCA associated vasculitis
• IgG4-related disease
• Sarcoidosis
• Syphilis
Mouth
• Dental abscess
• Behçet's syndrome (aphthous ulcer)
• Herpes simplex virus types 1 and 2
• HIV (candidiasis)
• Syphilis
• Whipple's disease (oculomasticatory myorhythmia)
Thyroid
• IgG4-related diseases
Heart/Lungs
• Infectious source (right-to-left shunt, pneumonia)
• Sarcoidosis
• Syphilis
• Systemic lupus erythematosus
• IV drug use (endocarditis)
Liver/Spleen
• Lymphoma
• Metastatic adenocarcinoma
• Sarcoidosis
• Tuberculosis
• Parasitic infection
• Brucellosis
• Rickettsial infection
Genitals
• Herpes simplex virus
• Syphilis
• Behçet's syndrome (aphthous ulcer)

FIGURE 144-1 Systemic manifestations that may provide clues to the etiology of chronic meningitis.

(magnetic resonance imaging [MRI] or computed tomography [CT]) shows leakage of contrast agent into the meninges. Meningeal enhancement is always concerning with the exception of dural enhancement after lumbar puncture, neurosurgical procedures, concussion, or spontaneous CSF leakage. Once chronic meningitis is confirmed by CSF examination, effort is focused on identifying the cause **(Tables 144-2 and 144-3)** by (1) further analysis of the CSF, (2) diagnosis of an underlying systemic infection or noninfectious inflammatory condition, or (3) pathologic examination of meningeal biopsy specimens.

Two clinical forms of chronic meningitis exist. In the first, the symptoms are chronic and persistent, whereas in the second, there

TABLE 144-2 Infectious Causes of Chronic Meningitis

CAUSATIVE AGENT	CSF FORMULA	HELPFUL DIAGNOSTIC TESTS	RISK FACTORS AND SYSTEMIC MANIFESTATIONS
Common Bacterial Causes			
Partially treated suppurative meningitis	Mononuclear or mixed mononuclear-polymorphonuclear cells	CSF culture and Gram's stain; CSF 16s rRNA PCR	History consistent with acute bacterial meningitis and incomplete treatment
Parameningeal infection	Mononuclear or mixed mononuclear-polymorphonuclear cells	Contrast-enhanced CT or MRI to detect parenchymal, subdural, epidural, or sinus infection	Otitis media, pleuropulmonary infection, right-to-left cardiopulmonary shunt for brain abscess; focal neurologic signs; neck, back, ear, or sinus tenderness
Mycobacterium tuberculosis	Mononuclear cells except polymorphonuclear cells in early infection (commonly <500 WBC/μL); low CSF glucose; high protein	Tuberculin skin test may be negative; interferon gamma release assay; PCR and AFB culture of CSF (sputum, urine, gastric contents if indicated); identify tubercle bacillus on acid-fast stain of CSF or protein pellicle	Exposure history; previous tuberculous illness; immunosuppressed, anti-TNF therapy or AIDS; young children; fever, meningismus, night sweats, miliary TB on x-ray or liver biopsy; stroke due to arteritis
Lyme disease (Bannwarth's syndrome) *Borrelia burgdorferi*	Mononuclear cells; elevated protein	Serum Lyme antibody titer; western blot confirmation; (patients with syphilis may have false-positive Lyme titer)	History of tick bite or appropriate exposure history; erythema chronicum migrans skin rash; arthritis, radiculopathy, Bell's palsy and other cranial neuropathies, meningoencephalitis–multiple sclerosis-like syndrome
Syphilis (secondary, tertiary) *Treponema pallidum*	Mononuclear cells; elevated protein	CSF VDRL; serum VDRL (or RPR); fluorescent treponemal antibody-absorbed (FTA) or MHA-TP; serum VDRL and RPR may be negative in tertiary syphilis due to waning antibody levels or earlier in the disease course due to very elevated antibody levels (prozone effect)	Appropriate exposure history; HIV-seropositive individuals at increased risk of aggressive infection; fever; lymphadenopathy; generalized, nonpruritic, mucocutaneous rash; "dementia"; cerebral infarction due to endarteritis; myelopathy
Uncommon Bacterial Causes			
Actinomyces	Polymorphonuclear cells	Anaerobic culture	Parameningeal abscess or sinus tract (oral or dental focus); pneumonitis
Nocardia	Polymorphonuclear; occasionally mononuclear cells; often low glucose	Isolation may require weeks; weakly acid fast	Associated brain abscess may be present
Brucella	Mononuclear cells (rarely polymorphonuclear); elevated protein; often low glucose	CSF antibody detection; serum antibody detection	Intake of unpasteurized dairy products; exposure to goats, sheep, cows; fever, arthralgia, myalgia, vertebral osteomyelitis
Whipple's disease *Tropheryma whipplei*	Mononuclear cells	Biopsy of small bowel or lymph node; CSF PCR for *T. whipplei;* brain and meningeal biopsy (with PAS stain and EM examination)	Diarrhea, weight loss, arthralgias, fever; dementia, ataxia, paresis, ophthalmoplegia, oculomasticatory myoclonus
Rare Bacterial Causes			
Leptospirosis (occasionally if left untreated may last 3–4 weeks)			
Fungal Causes			
Cryptococcus neoformans and *var. gattii*	Mononuclear cells; count not elevated in some patients with AIDS	India ink or fungal wet mount of CSF (budding yeast); blood and urine cultures; antigen detection in CSF (false negatives can occur in the setting of high antigen titers [prozone effect])	AIDS and immune suppression; pigeon exposure for *C. neoformans*, decaying wood exposure for *C. var. gattii*; skin and other organ involvement due to disseminated infection
Coccidioides immitis	Mononuclear cells (sometimes 10–20% eosinophils); often low glucose	Antibody detection in CSF and serum, antigen detection in CSF	Exposure history—southwestern United States
Candida species	Polymorphonuclear or mononuclear	Fungal stain and culture of CSF	IV drug abuse; postsurgery; prolonged IV therapy; disseminated candidiasis, recent epidural injection
Histoplasma capsulatum	Mononuclear cells; low glucose	Fungal stain and culture of large volumes of CSF; antigen detection in CSF, serum, and urine; antibody detection in serum, CSF	Exposure history—Ohio and central Mississippi River Valley; AIDS; mucosal lesions
Blastomyces dermatitidis	Mononuclear or polymorphonuclear	Fungal stain and culture of CSF; biopsy and culture of skin, lung lesions; antibody detection in serum	Midwestern and southeastern United States; usually systemic infection; abscesses, draining sinus, ulcers
Aspergillus species	Mononuclear or polymorphonuclear	CSF culture	Sinusitis; granulocytopenia or immunosuppression
Sporothrix schenckii	Mononuclear cells	Antibody detection in CSF and serum; CSF culture	Traumatic inoculation; IV drug use; ulcerated skin lesion
Rare Fungal Causes			
Xylohypha (formerly *Cladosporium*) *trichoides* and other dark-walled (dematiaceous) fungi such as *Curvularia; Drechslera; Mucor*, and, after water aspiration, *Pseudallescheria boydii;* iatrogenic *Exserohilum rostratum* infection following spinal blocks and *Fusarium solani* following epidural anesthesia in two clinics in Durango, Mexico, during 2022–2023			

(*Continued*)

TABLE 144-2 Infectious Causes of Chronic Meningitis (Continued)

CAUSATIVE AGENT	CSF FORMULA	HELPFUL DIAGNOSTIC TESTS	RISK FACTORS AND SYSTEMIC MANIFESTATIONS
Protozoal Causes			
Toxoplasma gondii	Mononuclear cells	Biopsy or response to empirical therapy in clinically appropriate context (including presence of antibody in serum)	Usually with intracerebral abscesses; common in HIV-seropositive patients; fever
Trypanosomiasis *Trypanosoma gambiense*, *T. rhodesiense*	Mononuclear cells; elevated protein	Elevated CSF IgM; identification of trypanosomes in CSF and blood smear	Endemic in Africa; chancre, lymphadenopathy; prominent sleep disorder
Rare Protozoal Causes			
Acanthamoeba sp. causing granulomatous amebic encephalitis and meningoencephalitis in immunocompromised and debilitated individuals; *Balamuthia mandrillaris* causing chronic meningoencephalitis in immunocompetent hosts			
Helminthic Causes			
Cysticercosis (infection with cysts of *Taenia solium*)	Mononuclear cells; may have eosinophils; glucose level may be low	Indirect hemagglutination assay in CSF; Serum serology with enzyme-linked immunoelectrotransfer blot preferred over crude extract ELISA; antigen or PCR testing in CSF	Usually with multiple cysts in basal meninges and hydrocephalus; cerebral cysts, ocular involvement; muscle calcification
Gnathostoma spinigerum	Eosinophils, mononuclear cells	Peripheral eosinophilia	History of eating raw fish; common in Thailand and Japan; ocular involvement; subarachnoid hemorrhage; painful radiculopathy
Angiostrongylus cantonensis	Eosinophils, mononuclear cells	Recovery of worms from CSF	History of eating raw shellfish; common in tropical Pacific regions; often benign; ocular involvement (rare)
Baylisascaris procyonis (raccoon ascarid)	Eosinophils, mononuclear cells	Immunoblot in CSF (Centers for Disease Control and Prevention)	Infection follows accidental ingestion of *B. procyonis* eggs from raccoon feces; ocular involvement; fatal meningoencephalitis
Rare Helminthic Causes			
Trichinella spiralis (trichinosis); *Fasciola hepatica* (liver fluke), *Echinococcus* cysts; *Schistosoma* spp. The former may produce a lymphocytic pleocytosis, whereas the latter two may produce an eosinophilic response in CSF associated with cerebral cysts (*Echinococcus*) or granulomatous lesions of brain or spinal cord.			
Viral Causes			
Mumps	Mononuclear cells	Antibody in serum	No prior mumps or immunization; orchitis; may produce meningoencephalitis; may persist for 3–4 weeks
Lymphocytic choriomeningitis	Mononuclear cells; may have low glucose	Antibody in serum; PCR for LCMV in CSF	Contact with rodents or their excreta; may persist for 3–4 weeks
Enteroviruses	Mononuclear cells; may have low glucose	Virus isolation and/or PCR for enteroviruses from CSF	Congenital hypogammaglobulinemia; history of recurrent meningitis
HIV (acute retroviral syndrome)	Mononuclear cells	PCR for HIV in blood and CSF	HIV risk factors; rash, fever, lymphadenopathy; lymphopenia in peripheral blood; syndrome may persist long enough to be considered as "chronic meningitis"; or chronic meningitis may develop in later stages (AIDS) due to HIV
Human herpes viruses	Mononuclear cells	PCR for HSV, EBV, CMV DNA; CSF antibody for HSV, EBV	Recurrent meningitis due to HSV-2 (rarely HSV-1) often associated with genital recurrences; EBV associated with myeloradiculopathy, CMV with polyradiculopathy

Abbreviations: AFB, acid-fast bacillus; CMV, cytomegalovirus; CSF, cerebrospinal fluid; CT, computed tomography; EBV, Epstein-Barr virus; ELISA, enzyme-linked immunosorbent assay; EM, electron microscopy; FTA, fluorescent treponemal antibody absorption test; HSV, herpes simplex virus; LCMV, lymphocytic choriomeningitis virus; MHA-TP, microhemagglutination assay–*T. pallidum*; MRI, magnetic resonance imaging; PAS, periodic acid–Schiff; PCR, polymerase chain reaction; RPR, rapid plasma reagin test; TB, tuberculosis; VDRL, Venereal Disease Research Laboratory test.

are recurrent discrete episodes of illness. In the latter group, all symptoms, signs, and CSF parameters of meningeal inflammation resolve completely between episodes either spontaneously or in response to a specific therapy. In such patients, the likely etiologies include Mollaret's meningitis, which is most often due to herpes simplex virus (HSV) type 2; chemical meningitis due to episodic leakage from an epidermoid tumor, craniopharyngioma, or cholesteatoma into CSF; primary autoimmune inflammatory conditions, including Vogt-Koyanagi-Harada syndrome, Behçet's syndrome, systemic lupus erythematosus (SLE), rheumatoid arthritis, neurosarcoidosis, IgG4-related disease, granulomatosis with polyangiitis, and primary central nervous system (CNS) vasculitis; and drug hypersensitivity with repeated administration of the offending agent. With the wider availability of whole genome sequencing, there is also a growing recognition that patients with inherited autoinflammatory syndromes like tumor necrosis factor (TNF) receptor–associated periodic syndrome (TRAPS), cryoporin-associated periodic fever syndrome (CAPS), complement factor I deficiency, and neonatal-onset multisystem inflammatory disorder can have recurrent meningitis.

The epidemiologic history is of considerable importance in diagnosis of chronic meningitis and may provide direction for selection of laboratory studies. Pertinent features include a history of tuberculosis or exposure; past travel to areas endemic for fungal infections (the San Joaquin Valley in California and southwestern states for coccidioidomycosis, midwestern states for histoplasmosis, southeastern states for blastomycosis); travel to the Mediterranean region or ingestion of imported unpasteurized dairy products (*Brucella*); time spent in wooded areas endemic for Lyme disease; exposure to sexually transmitted disease (syphilis, HSV-2); exposure of an immunocompromised host to pigeons and their droppings (*Cryptococcus neoformans*); exposure to decaying wood (*Cryptococcus gattii*); gardening (*Sporothrix schenckii*); ingestion of poorly cooked meat or contact with a household cat (*Toxoplasma gondii*); residence in Thailand or Japan (*Gnathostoma spinigerum*), Latin America (*Paracoccidioides brasiliensis*), or the South Pacific

TABLE 144-3 Noninfectious Causes of Chronic Meningitis

CAUSATIVE AGENTS	CSF FORMULA	HELPFUL DIAGNOSTIC TESTS	RISK FACTORS AND SYSTEMIC MANIFESTATIONS
Malignancy	Mononuclear cells; elevated protein; low glucose	Repeated cytologic examination of large volumes of CSF; CSF exam by polarizing microscopy; clonal lymphocyte markers; deposits on nerve roots or meninges seen on myelogram or contrast-enhanced MRI; meningeal biopsy	Metastatic cancer of breast, lung, stomach, or pancreas; melanoma, lymphoma, leukemia; meningeal gliomatosis; sarcoma; cerebral dysgerminoma
Chemical compounds (may cause recurrent meningitis)	Mononuclear or PMNs; low glucose, elevated protein; xanthochromia from subarachnoid hemorrhage in week prior to presentation with "meningitis"	Contrast-enhanced CT scan or MRI; cerebral angiogram to detect aneurysm. Enhancement and clumping of nerve roots of the cauda equina in arachnoiditis/pachymeningitis	History of recent injection into the subarachnoid space; history of sudden onset of headache; recent resection of acoustic neuroma or craniopharyngioma; epidermoid tumor of brain or spine, sometimes with dermoid sinus tract; pituitary apoplexy
Primary Inflammation			
CNS sarcoidosis	Mononuclear cells; elevated protein; often low glucose	Serum and CSF angiotensin-converting enzyme levels (insensitive); biopsy of extraneural affected tissues or brain lesion/meningeal biopsy, nodular meningeal and parenchymal enhancement	CN palsy, especially CN VII and CN II, including optic chiasm; hypothalamic dysfunction, especially diabetes insipidus; abnormal chest radiograph; peripheral neuropathy or myopathy; longitudinally extensive transverse myelitis
Vogt-Koyanagi-Harada syndrome (recurrent meningitis)	Mononuclear cells		Recurrent meningoencephalitis with uveitis, retinal detachment, alopecia, lightening of eyebrows and lashes, dysacousia, cataracts, glaucoma
Isolated granulomatous angiitis of the nervous system	Mononuclear cells; elevated protein	Angiography (often normal with small vessel angiitis); meningeal biopsy may be necessary if confined to small vessels. VZV PCR in blood, CSF, and biopsy tissue; microhemorrhages with amyloid beta–related angiitis	Subacute dementia; multiple cerebral infarctions; recent zoster ophthalmicus
Systemic lupus erythematosus	Mononuclear or PMNs	Anti-dsDNA antibody, antinuclear antibodies	Encephalopathy; seizures; stroke; transverse myelopathy; rash; arthritis; thromboembolism; renal and/or pulmonary complications
Behçet's syndrome (recurrent meningitis)	Mononuclear or PMNs; elevated protein	Rhombencephalitis	Oral and genital aphthous ulcers; iridocyclitis; retinal hemorrhages; pathergic lesions at site of skin puncture
Chronic benign lymphocytic meningitis	Mononuclear cells		Recovery in 2–6 months, diagnosis by exclusion
Mollaret's meningitis (recurrent meningitis)	Large endothelial cells and PMNs in first hours, followed by mononuclear cells	PCR for HSV; MRI/CT to rule out epidermoid tumor or dural cyst	Recurrent meningitis; exclude HSV-2; rare cases due to HSV-1; occasional case associated with dural cyst
Drug hypersensitivity	PMNs; occasionally mononuclear cells or eosinophils	Complete blood count (eosinophilia)	Exposure to nonsteroidal anti-inflammatory agents, sulfonamides, isoniazid, tolmetin, ciprofloxacin, penicillin, carbamazepine, lamotrigine, IV immunoglobulin, OKT3 antibodies, phenazopyridine; improvement after discontinuation of drug; recurrence with repeat exposure
Granulomatosis with polyangiitis (Wegener's)	Mononuclear cells	Chest and sinus radiographs; urinalysis; ANCA antibodies in serum; pachymeningitis on contrast-enhanced MRI	Associated sinus, pulmonary, or renal lesions; CN palsies; skin lesions; peripheral neuropathy
Neonatal-onset multisystem inflammatory disorder	Mononuclear and PMNs	Gain-of-function mutation in *NLRP3* gene leading to elevated IL-1β	Recurrent fever, urticaria, arthralgia, sensorineural hearing loss, papilledema, increased ICP
IgG4-related hypertrophic pachymeningitis	Mild lymphocytic pleocytosis in some cases; normal to mildly increased protein; normal glucose	Serum IgG4 levels frequently elevated; ESR and C-reactive protein; pachymeningitis on contrast-enhanced MRI; meningeal biopsy shows swirling "storiform" fibrosis with lymphocytic infiltrates, obliterative phlebitis and IgG4+ plasma cells	Headache; seizures; focal symptoms from dural involvement in spinal cord/nerve roots, clivus, periorbital, vestibular, and brainstem structures. Systemic IgG4-related disease can involve many tissues including pancreas, thyroid, lungs, retroperitoneum, lacrimal, parotid and submandibular glands, orbits, kidney, aorta, liver
TNF receptor–associated periodic fever syndrome (TRAPS)	Mononuclear cells	Mutation in *TNFRSF1A* gene leading to elevated TNF	Headache, seizures, tinnitus, skin rash, abdominal pain, lymphadenopathy, periorbital edema, joint pain, myalgia
Complement factor I deficiency	PMNs	Mutation in complement factor I gene leading to low serum levels of factor I (or dysfunctional factor I) and C3	Recurrent, steroid-responsive, aseptic, neutrophilic meningitis with or without encephalitis; increased risk for systemic infections with encapsulated bacteria, glomerulonephritis, systemic lupus erythematosus and leukocytoclastic vasculitis
Cryoporin-associated periodic fever syndrome (CAPS)	Mononuclear cells	Heterozygous gain-of-function mutations within the *NLRP3* gene	Fever, urticaria, amyloidosis, arthralgia, sensorineural hearing loss, myalgias, papilledema, vision changes
Other: multiple sclerosis, Sjögren's syndrome, and rarer forms of vasculitis (e.g., Cogan's syndrome)			

Abbreviations: ANCA, antineutrophil cytoplasmic antibodies; CN, cranial nerve; CNS, central nervous system; CSF, cerebrospinal fluid; CT, computed tomography; HSV, herpes simplex virus; ICP, intracranial pressure; IL, interleukin; MRI, magnetic resonance imaging; PCR, polymerase chain reaction; PMNs, polymorphonuclear cells; TNF, tumor necrosis factor; VZV, varicella-zoster virus.

(*Angiostrongylus cantonensis*); rural residence and raccoon exposure (*Baylisascaris procyonis*); and residence in Latin America, the Philippines, sub-Saharan Africa, or Southeast Asia (*Taenia solium/cysticercosis, schistosomiasis*). CNS meliodosis caused by *Burkholderia pseudomallei* is endemic in South Asia and Australia. Individuals with agammaglobulinemia or those receiving B cell–depleting therapy may be susceptible to chronic enterovirus meningitis.

Focal cerebral signs in a patient with chronic meningitis suggest the possibility of a brain abscess, parameningeal infection, or infarct; identification of a potential source of infection (chronic draining ear, sinusitis, dental abscess, right-to-left cardiac or pulmonary shunt, chronic pleuropulmonary infection) supports this diagnosis. In some cases, diagnosis may be established by recognition and biopsy of unusual skin lesions (Behçet's syndrome, SLE, cryptococcosis, blastomycosis, Lyme disease, sporotrichosis, trypanosomiasis, IV drug use) or enlarged lymph nodes (lymphoma, sarcoid, tuberculosis, HIV, secondary syphilis, or Whipple's disease). Ophthalmologic examination may reveal uveitis (Vogt-Koyanagi-Harada syndrome, sarcoidosis, or CNS lymphoma), keratoconjunctivitis sicca (Sjögren's syndrome), or iridocyclitis (Behçet's syndrome) and is essential to assess visual loss from papilledema. *M. tuberculosis* can cause a wide spectrum of ophthalmologic pathology. If neurocysticercosis is suspected, it is important to rule out an intraocular infection requiring surgical treatment before initiating antihelminthic therapy. Aphthous oral lesions, genital ulcers, and hypopyon (inflammatory cells in the anterior chamber of the eye) suggest Behçet's syndrome. Hepatosplenomegaly suggests lymphoma, sarcoid, tuberculosis, or brucellosis. Arthralgias could be indicative of Lyme disease or Whipple's disease. The latter can also cause gastrointestinal symptoms, including diarrhea and abdominal pain. Herpetic lesions in the genital area or on the thighs suggest HSV-2 infection. A breast nodule; a suspicious hyperpigmented skin lesion; focal bone pain; hard, fixed lymph nodes; or an abdominal mass suggests possible carcinomatous meningitis.

IMAGING

Once the clinical syndrome is recognized as a potential manifestation of chronic meningitis, proper analysis of the CSF is essential. However, if the possibility of raised ICP exists, a brain imaging study should be performed before lumbar puncture. If ICP is elevated because of a mass lesion, brain swelling, or a block in ventricular CSF outflow (obstructive hydrocephalus), then lumbar puncture carries the potential risk of brain herniation. Obstructive hydrocephalus usually requires direct ventricular drainage. In patients with open CSF flow pathways, elevated ICP can still occur due to impaired resorption of CSF by arachnoid villi. In such patients, lumbar puncture is usually safe and may be therapeutic. Indeed, repetitive or continuous lumbar drainage may be necessary to prevent abrupt deterioration and death from raised ICP. In some patients, especially those with cryptococcal meningitis, fatal levels of raised ICP can occur without enlarged ventricles.

Contrast-enhanced MRI or CT studies of the brain and spinal cord can identify meningeal enhancement, parameningeal infections (including brain abscess), encasement of the spinal cord (malignancy, inflammation, or infection), or nodular deposits on the meninges or nerve roots (malignancy, sarcoidosis, or schistosomiasis) **(Fig. 144-2)**. Imaging studies are also useful to guide biopsy of affected meninges. Lastly, a cyst that recurrently causes a chemical meningitis due to a leak may be better visualized in between clinical episodes when a recent leak has not shrunken the cyst volume.

The patterns of enhancement of the different layers of the meninges can be very informative and can be divided into two types: leptomeningeal (pia and arachnoid), when enhancement of the meninges follows the convolutions of the gyri and/or involves the meninges around the basal cisterns; and pachymeningeal (dura), when the enhancement is thick and linear or nodular along the inner surface of the calvarium, falx, or tentorium without extension into the cortical gyri or basal cistern involvement. For example, infectious meningitis presents mostly as leptomeningitis, while lymphomatous meningitis can present as pachymeningitis.

Angiographic studies can identify evidence of cerebral arteritis in patients with chronic meningitis and stroke.

CEREBROSPINAL FLUID ANALYSIS

The CSF pressure should be measured and samples sent for bacterial, fungal, and mycobacterial culture; Venereal Disease Research Laboratory (VDRL) test; cell count and differential; Gram's stain; and measurement of glucose and protein. CSF VDRL is a highly specific, but not particularly sensitive, test for syphilis. If CSF VDRL is negative in an otherwise high-risk patient with positive treponemal antibodies in the serum and an otherwise unexplained CSF pleocytosis, empiric treatment for neurosyphilis may still be appropriate. Wet mount for fungus and parasites, India ink preparation, culture for fastidious bacteria and fungi, assays for cryptococcal antigen and oligoclonal immunoglobulin bands, and cytology should be performed. Other specific CSF tests (Tables 144-2 and 144-3) or blood tests and cultures should be ordered as indicated based on the history, physical examination, or preliminary CSF results (i.e., eosinophilic, mononuclear, or polymorphonuclear meningitis). Rapid diagnosis may be facilitated by serologic tests and polymerase chain reaction (PCR) testing to identify DNA sequences in the CSF that are specific for the suspected pathogen. 16s ribosomal RNA (rRNA) PCR can be used to detect a broad range of bacterial causes of meningitis and can be particularly useful in partially treated meningitis when the yield of culture is low. 18s and 28s rRNA PCR can similarly be useful for detecting a broad range of fungal species. In patients with suspected fungal infections, when other tests are negative, CSF assays for beta-glucans may be a useful adjunct in establishing the diagnosis. Building on progress in parallel deep sequencing and informatics, unbiased metagenomic next-generation sequencing is becoming generally available, representing an efficient and powerful method for diagnosis of challenging infectious cases.

In most categories of chronic (not recurrent) meningitis, mononuclear cells predominate in the CSF. When neutrophils predominate after 3 weeks of illness, the principal etiologic considerations are *Nocardia asteroides*, *Actinomyces israelii*, *Brucella*, *M. tuberculosis* (5–10% of early cases only), various fungi (*Blastomyces dermatitidis*, *Candida* spp., *Histoplasma capsulatum*, *Aspergillus* spp., *Pseudallescheria boydii*, *Cladophialophora bantiana*), and noninfectious causes (SLE, exogenous chemical meningitis). When eosinophils predominate or are present in limited numbers in a primarily mononuclear cell response in the CSF, the differential diagnosis includes parasitic diseases (*A. cantonensis*, *G. spinigerum*, *B. procyonis*, or *Toxocara canis* infection; cysticercosis; schistosomiasis; echinococcal disease; *T. gondii* infection), fungal infections (6–20% eosinophils along with a predominantly lymphocyte pleocytosis, particularly with coccidioidal meningitis), neoplastic disease (lymphoma, leukemia, metastatic carcinoma), or other inflammatory processes (sarcoidosis, hypereosinophilic syndrome).

It is often necessary to broaden the number of diagnostic tests if the initial workup does not reveal the cause. In addition, repeated samples (three or more) of large volumes of lumbar CSF may be required to diagnose certain infectious and malignant causes of chronic meningitis. Lymphomatous or carcinomatous meningitis may be diagnosed by examination of sections cut from a cell block formed by spinning down the sediment from a large volume of CSF. Flow cytometry for malignant cells may also be useful in patients with suspected carcinomatous meningitis. The diagnosis of fungal meningitis may also require large volumes of CSF for culture of sediment. If standard lumbar puncture is unrewarding, a cervical cisternal tap to sample CSF near to the basal meninges may be fruitful. Ventricular fluid may appear sterile in cases with active infection in the lower lumbar space.

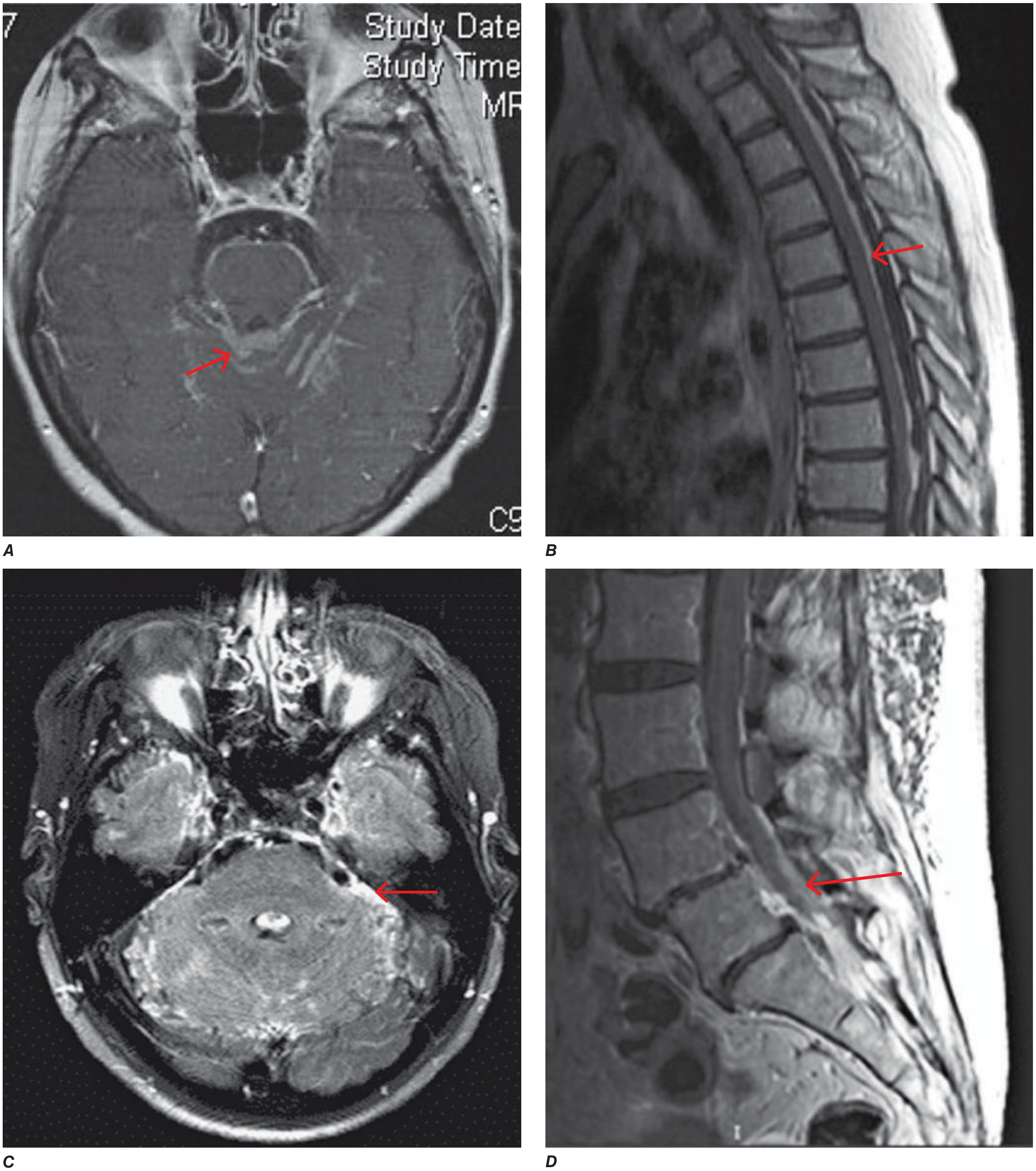

FIGURE 144-2 Chronic meningitis illustrating meningeal enhancement on contrast magnetic resonance imaging scan. ***A*** and ***B*** are images from a patient with chronic meningitis due to carcinoma. ***C*** and ***D*** are from a patient with chronic meningitis due to *Cryptococcus* infection. *Arrows* point to the most prominent areas of meningeal inflammation around the brainstem and cerebellar folia (***A***), cerebellum (***C***), along the dorsal spinal cord (***B***), and clumping of roots in the cauda equina (***D***).

LABORATORY INVESTIGATION

In addition to the CSF examination, an attempt should be made to uncover pertinent underlying illnesses. Tuberculin skin test, chest radiograph, urine analysis and culture, blood count and differential, renal and liver function tests, alkaline phosphatase, sedimentation rate, antinuclear antibody, anti-Ro antibody, anti-La antibody, rheumatoid factor, and IgG4 level are often indicated. In some cases, a thorough search for a systemic site of infection is indicated. Pulmonary foci of infection may be present, particularly with fungal or tuberculous disease. Hence, a CT or MRI of the chest and a sputum examination may be helpful. Abnormalities can be pursued by bronchoscopy or transthoracic needle biopsy. A tuberculin skin test is often placed, although the test has limited specificity and sensitivity for diagnosis of active disease. Where available, gamma interferon release assays may be used to diagnose latent tuberculosis. Liver, bone marrow, or lymph node biopsy may be diagnostic in some cases of miliary tuberculosis, disseminated fungal infection, sarcoidosis, or metastatic malignancy. Positron emission tomography with fluorodeoxyglucose may be useful in identifying a systemic site for biopsy in patients with suspected carcinomatous meningitis or sarcoidosis when other tests are unrevealing. Genetic testing can identify mutations that cause rare monogenic autoinflammatory disorders or underlying immunocompromised states.

MENINGEAL BIOPSY

If CSF is not diagnostic, then a meningeal biopsy should be strongly considered in patients who are severely disabled, who need chronic ventricular decompression, or whose illness is progressing rapidly. The activities of the surgeon, pathologist, microbiologist, and cytologist should be coordinated so that a large enough sample is obtained and the appropriate cultures and histologic and molecular studies, including electron-microscopic and PCR studies, are performed. The diagnostic yield of meningeal biopsy can be increased by targeting regions that enhance with contrast on MRI or CT. With current microsurgical techniques, most areas of the basal meninges can be accessed for biopsy via a limited craniotomy. In one series, MRI demonstrated meningeal enhancement in 47% of patients undergoing meningeal biopsy; biopsy of an enhancing region was diagnostic in 80% of cases, biopsy of nonenhancing regions was diagnostic in only 9%, and sarcoidosis (31%) and metastatic adenocarcinoma (25%) were the most common conditions identified. Tuberculosis is the most common condition identified in many reports from outside the United States.

APPROACH TO THE ENIGMATIC CASE

In approximately one-third of cases, the diagnosis is not known despite careful evaluation of CSF and potential extraneural sites of disease. A number of the organisms that cause chronic meningitis may take weeks to be identified by cultures. In enigmatic cases, several options are available, determined by the extent of the clinical deficits and rate of progression. It is prudent to wait until cultures are finalized if the patient is asymptomatic or symptoms are mild and not progressive. Unfortunately, in many cases, progressive neurologic deterioration occurs, and rapid treatment is required. Ventricular-peritoneal shunts may be placed to relieve hydrocephalus, but the risk of disseminating the undiagnosed inflammatory process into the abdomen must be considered.

Empirical Treatment Diagnosis of the causative agent is essential because effective therapies exist for many etiologies of chronic meningitis, but if the condition is left untreated, progressive damage to the CNS and cranial nerves and roots is likely to occur. Occasionally, empirical therapy must be initiated when all attempts at diagnosis fail. In general, empirical therapy in the United States consists of antimycobacterial agents, amphotericin B (often combined with flucytosine) for fungal infection, and/or glucocorticoids for noninfectious inflammatory causes. It is important to direct empirical therapy of lymphocytic meningitis at tuberculosis, particularly if the condition is associated with low CSF glucose, since untreated disease can be devastating within weeks. Prolonged anti-TNF therapy and anti–programmed cell death 1 (PD-1) inhibitors can cause reactivation of TB, and such patients who develop chronic meningitis should be treated empirically with antituberculous therapy if the etiology is uncertain. Even with treatment, TB meningitis can carry high rates of morbidity. In the Mayo Clinic series, the most useful empirical therapy was administration of glucocorticoids rather than antituberculous therapy. When proceeding with empiric glucocorticoids, caution should be maintained whenever a transient response to treatment is noted, as some infectious (e.g., tuberculosis and cysticercosis) and noninfectious (e.g., lymphoma) etiologies may temporarily respond to glucocorticoid monotherapy. Carcinomatous or lymphomatous meningitis may be difficult to diagnose initially, but the diagnosis becomes evident with time.

THE IMMUNOSUPPRESSED PATIENT

Chronic meningitis is not uncommon in the course of HIV infection. Pleocytosis and mild meningeal signs often occur at the onset of HIV infection, and occasionally, low-grade meningitis persists. In worldwide populations, *M. tuberculosis* is the most common cause of chronic meningitis, followed by *C. neoformans*. Toxoplasmosis commonly presents as intracranial abscesses and also may be associated with meningitis. Other important causes of chronic meningitis in AIDS include infection with *Nocardia*, *Candida*, or other fungi; syphilis; and lymphoma (Fig. 144-2). With HIV infection, primary CNS lymphomas may arise, which are typically positive for EBV infection. Toxoplasmosis, nocardiosis, cryptococcosis and other fungal infections are important etiologic considerations in individuals with immunodeficiency states other than AIDS, including those due to immunosuppressive medications. Because of the increased risk of chronic meningitis and the attenuation of clinical signs of meningeal irritation in immunosuppressed individuals, CSF examination should be performed for any persistent headache or unexplained change in mental state.

FURTHER READING

Aksamit AJ: Chronic meningitis. N Engl J Med 385:930, 2021.

Baldwin K, Avila JD: Diagnostic approach to chronic meningitis. Neurol Clin 36:831, 2018.

Chang CC et al: Global guideline for the diagnosis and management of cryptococcosis: An initiative of the ECMM and ISHAM in cooperation with the ASM. Lancet Infect Dis 24:e495, 2024.

Johnson TP, Nath A: Biotypes of HIV associated neurocognitive disorders. Curr Opin Inf Dis 335:223, 2022.

Kacar M et al: Hereditary systemic autoinflammatory diseases and Schnitzler's syndrome. Rheumatology 58:vi31, 2019.

Lu LX et al: IgG4-related hypertrophic pachymeningitis: Clinical features, diagnostic criteria and treatment. JAMA Neurol 71:785, 2014.

Saifon W et al: Distinguishing clinical characteristics of central nervous system tuberculosis in immunodeficient and non-immunodeficient individuals: A 12-year retrospective study. Ann Clin Microbiol Antimicrob 22:69, 2023.

Wilson MR et al: Chronic meningitis investigated via metagenomic next-generation sequencing. JAMA Neurol 75:947, 2018.

145

Brain Abscess and Empyema

Karen L. Roos, Kenneth L. Tyler

BRAIN ABSCESS

DEFINITION

A brain abscess is a focal, suppurative infection within the brain parenchyma, typically surrounded by a vascularized capsule. The term *cerebritis* is often employed to describe a nonencapsulated brain abscess.

EPIDEMIOLOGY

A bacterial brain abscess is a relatively uncommon intracranial infection, with an incidence of ~0.3–1.3:100,000 persons per year. Predisposing conditions include otitis media and mastoiditis, paranasal sinusitis, pyogenic infections in the chest or other body sites, penetrating head trauma or neurosurgical procedures, and dental infections. In immunocompetent individuals, the most important pathogens are *Streptococcus* spp. (anaerobic, aerobic, and viridans [40%]), Enterobacteriaceae (*Proteus* spp., *Escherichia coli* sp., *Klebsiella* spp. [25%]), anaerobes (e.g., *Bacteroides* spp., *Fusobacterium* spp. [30%]), and staphylococci (10%). In immunocompromised hosts with underlying HIV infection, organ transplantation, cancer, or immunosuppressive therapy, most brain abscesses are caused by *Nocardia* spp., *Toxoplasma* *gondii, Aspergillus* spp., *Candida* spp., and *Cryptococcus neoformans*. In Latin America and in immigrants from Latin America, the most common cause of brain abscess is *Taenia solium* (neurocysticercosis). In India and East Asia, mycobacterial infection (tuberculoma) remains a major cause of focal CNS mass lesions.

■ ETIOLOGY

A brain abscess may develop (1) by direct spread from a contiguous cranial site of infection, such as paranasal sinusitis, otitis media, mastoiditis, or dental infection; (2) following head trauma or a neurosurgical procedure; or (3) as a result of hematogenous spread from a remote site of infection. In up to 25% of cases, no obvious primary source of infection is apparent (cryptogenic brain abscess).

Approximately one-third of brain abscesses are associated with otitis media and mastoiditis, often with an associated cholesteatoma. Otogenic abscesses occur predominantly in the temporal lobe (55–75%) and cerebellum (20–30%). In some series, up to 90% of cerebellar abscesses are otogenic. Common organisms include streptococci, *Bacteroides* spp., *Pseudomonas* spp., *Haemophilus* spp., and Enterobacteriaceae. Abscesses that develop as a result of direct spread of infection from the frontal, ethmoidal, or sphenoidal sinuses and those that occur due to dental infections are usually located in the frontal lobes. Approximately 10% of brain abscesses are associated with paranasal sinusitis, and this association is particularly strong in young males in their second and third decades of life. The most common pathogens in brain abscesses associated with paranasal sinusitis are streptococci (especially *Streptococcus milleri*), *Haemophilus* spp., *Bacteroides* spp., *Pseudomonas* spp., and *Staphylococcus aureus*. Dental infections are associated with ~2% of brain abscesses, although it is often suggested that many "cryptogenic" abscesses are in fact due to dental infections. The most common pathogens in this setting are streptococci, staphylococci, *Bacteroides* spp., and *Fusobacterium* spp.

Hematogenous abscesses account for ~25% of brain abscesses. Hematogenous abscesses are often multiple, and multiple abscesses often (50%) have a hematogenous origin. These abscesses show a predilection for the territory of the middle cerebral artery (i.e., posterior frontal or parietal lobes). Hematogenous abscesses are often located at the junction of the gray and white matter and are often poorly encapsulated. The microbiology of hematogenous abscesses is dependent on the primary source of infection. For example, brain abscesses that develop as a complication of infective endocarditis are often due to viridans streptococci or *S. aureus*. Abscesses associated with pyogenic lung infections such as lung abscess or bronchiectasis are often due to streptococci, staphylococci, *Bacteroides* spp., *Fusobacterium* spp., or Enterobacteriaceae. Enterobacteriaceae and *Pseudomonas aeruginosa* are important causes of abscesses associated with urinary sepsis. Congenital cardiac malformations that produce a right-to-left shunt allow bloodborne bacteria to bypass the pulmonary capillary bed and reach the brain. Similar phenomena can occur with pulmonary arteriovenous malformations. The decreased arterial oxygenation and saturation from the right-to-left shunt and polycythemia may cause focal areas of cerebral ischemia, thus providing a nidus for microorganisms that bypassed the pulmonary circulation to multiply and form an abscess. Streptococci are the most common pathogens in this setting.

Abscesses that follow penetrating head trauma or neurosurgical procedures are frequently due to methicillin-resistant *S. aureus* (MRSA), *Staphylococcus epidermidis*, Enterobacteriaceae, *Pseudomonas* spp., and *Clostridium* spp.

■ PATHOGENESIS AND HISTOPATHOLOGY

Results of experimental models of brain abscess formation suggest that for bacterial invasion of brain parenchyma to occur, there must be preexisting or concomitant areas of ischemia, necrosis, or hypoxemia in brain tissue. The intact brain parenchyma is relatively resistant to infection. Once bacteria have established infection, brain abscess frequently evolves through a series of stages, influenced by the nature of the infecting organism and by the immunocompetence of the host. The early cerebritis stage (days 1–3) is characterized by a perivascular infiltration of inflammatory cells, which surround a central core of coagulative necrosis. Marked edema surrounds the lesion at this stage. In the late cerebritis stage (days 4–9), pus formation leads to enlargement of the necrotic center, which is surrounded at its border by an inflammatory infiltrate of macrophages and fibroblasts. A thin capsule of fibroblasts and reticular fibers gradually develops, and the surrounding area of cerebral edema becomes more distinct than in the previous stage. The third stage, early capsule formation (days 10–13), is characterized by the formation of a capsule that is better developed on the cortical than on the ventricular side of the lesion. This stage correlates with the appearance of a ring-enhancing capsule on neuroimaging studies. The final stage, late capsule formation (day 14 and beyond), is defined by a well-formed necrotic center surrounded by a dense collagenous capsule. The surrounding area of cerebral edema has regressed, but marked gliosis with large numbers of reactive astrocytes has developed outside the capsule. This gliotic process may contribute to the development of seizures as a sequela of brain abscess.

■ CLINICAL PRESENTATION

A brain abscess typically presents as an expanding intracranial mass lesion rather than as an infectious process. Although the evolution of signs and symptoms is extremely variable, ranging from hours to weeks or even months, most patients present to the hospital 11–12 days following onset of symptoms. The classic clinical triad of headache, fever, and a focal neurologic deficit is present in <50% of cases. The most common symptom in patients with a brain abscess is headache, occurring in >75% of patients. The headache is often characterized as a constant, dull, aching sensation, either hemicranial or generalized, and it becomes progressively more severe and refractory to therapy. Fever is present in only 50% of patients at the time of diagnosis, and its absence should not exclude the diagnosis. The new onset of focal or generalized seizure activity is a presenting sign in 15–35% of patients. Focal neurologic deficits including hemiparesis, aphasia, or visual field defects are part of the initial presentation in >60% of patients.

The clinical presentation of a brain abscess depends on its location, the nature of the primary infection if present, and the level of the intracranial pressure (ICP). Hemiparesis is the most common localizing sign of a frontal lobe abscess. A temporal lobe abscess may present with a disturbance of language (dysphasia) or an upper homonymous quadrantanopia. Nystagmus and ataxia are signs of a cerebellar abscess. Signs of raised ICP—papilledema, nausea and vomiting, and drowsiness or confusion—can be the dominant presentation of some abscesses, particularly those in the cerebellum. Meningismus is not present unless the abscess has ruptured into the ventricle or the infection has spread to the subarachnoid space.

■ DIAGNOSIS

Diagnosis is made by neuroimaging studies. Magnetic resonance imaging (MRI) **(Fig. 145-1)** is better than computed tomography (CT) for demonstrating abscesses in the early (cerebritis) stages and is superior to CT for identifying abscesses in the posterior fossa. Cerebritis appears on MRI as an area of low signal intensity on T1-weighted images with irregular postgadolinium enhancement and as an area of increased signal intensity on T2-weighted images. Cerebritis is often not visualized by CT scan, but when present, appears as an area of hypodensity. On a contrast-enhanced CT scan, a mature brain abscess appears as a focal area of hypodensity surrounded by ring enhancement with surrounding edema (hypodensity). On contrast-enhanced T1-weighted MRI, a mature brain abscess has a capsule that enhances surrounding a hypodense center and surrounded by a hypodense area of edema. On T2-weighted MRI, there is a hyperintense central area of pus surrounded by a well-defined hypointense capsule and a hyperintense surrounding area of edema. It is important to recognize that the CT and MRI appearance, particularly of the capsule, may be altered by treatment with glucocorticoids. The distinction between a brain abscess and other focal CNS lesions such as primary or metastatic tumors may be facilitated by the use of diffusion-weighted imaging sequences on which a brain abscess typically shows increased signal due to restricted diffusion of the abscess cavity with corresponding low signal on apparent diffusion coefficient images.

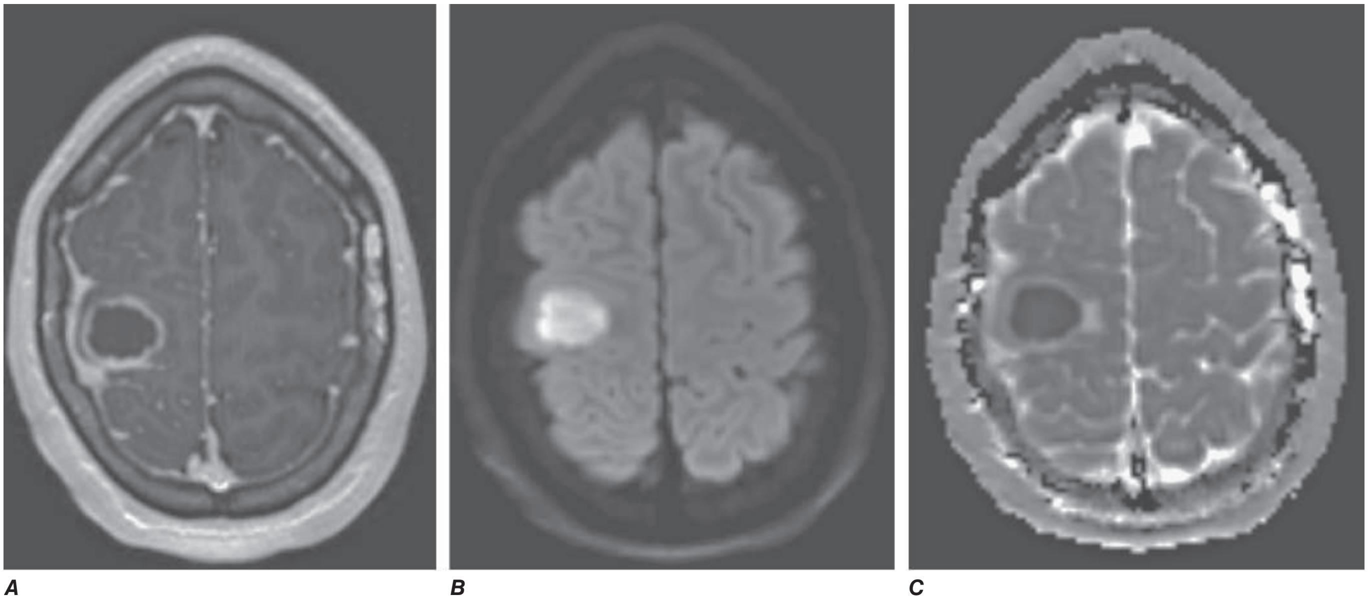

FIGURE 145-1 Pyogenic brain abscess. Note that the abscess wall enhances prominently after gadolinium administration on the magnetic resonance axial T1-weighted image (***A***). The abscess is hyperintense on the diffusion-weighted image (***B***) and dark on the apparent diffusion coefficient (ADC) image (***C***). *(Courtesy of Aaron Kamer, MD; with permission.)*

Microbiologic diagnosis of the etiologic agent is most accurately determined by Gram's stain and culture of abscess material obtained by CT-guided stereotactic needle aspiration. Aerobic and anaerobic bacterial cultures and mycobacterial and fungal cultures should be obtained. Blood cultures are positive in ~10% of cases overall but may be positive in >85% of patients with abscesses due to *Listeria*. About 50% of patients have a peripheral leukocytosis, 60% an elevated erythrocyte sedimentation rate, and 80% an elevated C-reactive protein. Lumbar puncture (LP) should not be performed in patients with known or suspected focal intracranial infections such as abscess or empyema; cerebrospinal fluid (CSF) analysis contributes nothing to diagnosis or therapy, and LP increases the risk of herniation.

DIFFERENTIAL DIAGNOSIS

Conditions that can cause headache, fever, focal neurologic signs, and seizure activity include brain abscess, subdural empyema, bacterial meningitis **(Chap. 143)**, viral meningoencephalitis **(Chap. 142)**, superior sagittal sinus thrombosis **(Chap. 438)**, and acute disseminated encephalomyelitis **(Chap. 456)**. When fever is absent, primary and metastatic brain tumors become the major differential diagnosis. Less commonly, cerebral infarction or hematoma can have an MRI or CT appearance resembling brain abscess.

TREATMENT

Brain Abscess

Optimal therapy of brain abscesses involves a combination of high-dose parenteral antibiotics and neurosurgical drainage. Empirical therapy of community-acquired brain abscess in an immunocompetent patient typically includes a third- or fourth-generation cephalosporin (e.g., cefotaxime, ceftriaxone, or cefepime) and metronidazole **(see Table 143-1 for antibiotic dosages)**. In patients with penetrating head trauma or recent neurosurgical procedures, treatment should include ceftazidime as the third-generation cephalosporin to enhance coverage of *Pseudomonas* spp. and vancomycin for coverage of staphylococci. Meropenem plus vancomycin also provides good coverage in this setting.

Aspiration and drainage of the abscess under stereotactic guidance are beneficial for both diagnosis and therapy. Empirical antibiotic coverage should be modified based on the results of Gram's stain and culture of the abscess contents. Complete excision of a bacterial abscess via craniotomy or craniectomy is generally reserved for multiloculated abscesses or those in which stereotactic aspiration is unsuccessful.

Medical therapy alone is not optimal for treatment of brain abscess and should be reserved for patients whose abscesses are neurosurgically inaccessible, for patients with small (<2–3 cm) or nonencapsulated abscesses (cerebritis), and for patients whose condition is too tenuous to allow performance of a neurosurgical procedure. All patients should receive a minimum of 6–8 weeks of parenteral antibiotic therapy. The role, if any, of supplemental oral antibiotic therapy following completion of a standard course of parenteral therapy has never been adequately studied.

In addition to surgical drainage and antibiotic therapy, patients should receive prophylactic anticonvulsant therapy because of the high risk (~35%) of focal or generalized seizures. Anticonvulsant therapy is continued for at least 3 months after resolution of the abscess, and decisions regarding withdrawal are then based on the electroencephalogram (EEG). If the EEG is abnormal, anticonvulsant therapy should be continued. If the EEG is normal, anticonvulsant therapy can be slowly withdrawn, with close follow-up and repeat EEG after the medication has been discontinued.

Glucocorticoids should not be given routinely to patients with brain abscesses. Intravenous dexamethasone therapy (10 mg every 6 h) is usually reserved for patients with substantial periabscess edema and associated mass effect and increased ICP. Dexamethasone should be tapered as rapidly as possible to avoid delaying the natural process of encapsulation of the abscess.

Serial MRI or CT scans should be obtained on a monthly or twice-monthly basis to document resolution of the abscess. More frequent studies (e.g., weekly) are probably warranted in the subset of patients who are receiving antibiotic therapy alone. A small amount of enhancement may remain for months after the abscess has been successfully treated.

PROGNOSIS

The mortality rate of brain abscess has declined in parallel with the development of enhanced neuroimaging techniques, improved neurosurgical procedures for stereotactic aspiration, and improved antibiotics. In modern series, the mortality rate is typically <15%. Significant sequelae, including seizures, persisting weakness, aphasia, or mental impairment, occur in ≥20% of survivors.

NONBACTERIAL CAUSES OF INFECTIOUS FOCAL CNS LESIONS

ETIOLOGY

Neurocysticercosis is the most common parasitic disease of the CNS worldwide. Humans acquire cysticercosis by the ingestion of food contaminated with the eggs of the parasite *Taenia solium* **(Chap. 242)**. Toxoplasmosis **(Chap. 235)** is a parasitic disease caused by *T. gondii* and acquired from the ingestion of undercooked meat and from handling cat feces.

CLINICAL PRESENTATION

The most common manifestation of neurocysticercosis is new-onset partial seizures with or without secondary generalization. Cysticerci may develop in the brain parenchyma and cause seizures or focal neurologic deficits. When present in the subarachnoid or ventricular spaces, cysticerci can produce increased ICP by interference with CSF flow. Spinal cysticerci can mimic the presentation of intraspinal tumors. When the cysticerci first lodge in the brain, they frequently cause little in the way of an inflammatory response. As the cysticercal cyst degenerates, it elicits an inflammatory response that may present clinically as a seizure. Eventually the cyst dies, a process that may take several years and is typically associated with resolution of the inflammatory response and, often, abatement of seizures.

Primary *Toxoplasma* infection is often asymptomatic. However, during this phase, parasites may spread to the CNS, where they become latent. Reactivation of CNS infection is almost exclusively associated with immunocompromised hosts, particularly those with HIV infection. During this phase, patients present with headache, fever, seizures, and focal neurologic deficits.

DIAGNOSIS

The lesions of neurocysticercosis are readily visualized by MRI or CT scans depending on the stage of the lesion. There are four stages of neurocysticercosis: (1) the vesicular stage, (2) the colloidal stage, (3) the granulonodular stage, and (4) the nodular-calcified stage. Lesions with viable parasites appear as cystic lesions, and the scolex can often be visualized on MRI. Cystic lesions with small nodules (scolex) within the cyst are in the vesicular stage of neurocysticercosis **(Fig. 145-2*A* and *B*)**. There is no significant edema surrounding a lesion in the vesicular stage. Lesions in the colloidal stage demonstrate peripheral enhancement on postcontrast imaging **(Fig. 145-2*C*)** with substantial surrounding edema on T2 images **(Fig. 145-2*D*)**. In the granulonodular stage, on postcontrast imaging, the lesion enhances in a homogenous fashion **(Fig. 145-2*E*)**. On fluid-attenuated inversion recovery (FLAIR) images, there is no surrounding edema **(Fig. 145-2*F*)**. Parenchymal brain calcifications are the most common finding and evidence that the parasite is no longer viable. These chronic lesions are best seen on CT **(Fig. 145-2*G*)** and can be difficult to detect on MRI. The most sensitive technique for the detection of these small calcific foci on MRI is susceptibility-weighted imaging (SWI). If a confirmatory test for neurocysticercosis is needed, the enzyme-linked immunotransfer blot is recommended. A funduscopic exam is also recommended for all patients with suspected neurocysticercosis.

MRI findings of toxoplasmosis consist of multiple lesions in the deep white matter, the thalamus, and basal ganglia and at the gray-white junction in the cerebral hemispheres. With contrast administration, the majority of the lesions enhance in a ringed, nodular, or homogeneous pattern and are surrounded by edema. In the presence of the characteristic neuroimaging abnormalities of *T. gondii* infection, serum IgG antibody to *T. gondii* should be obtained and, when positive, the patient should be treated.

TREATMENT

Infectious Focal CNS Lesions

Anticonvulsant therapy is initiated when the patient with neurocysticercosis presents with a seizure. Anthelmintic therapy is given to patients based on the stage of the lesion(s). Cysticerci appearing as cystic lesions in the brain parenchyma with or without pericystic edema or in the subarachnoid space at the convexity of the cerebral hemispheres should be treated with anticysticidal therapy. Cysticidal drugs accelerate the destruction of the parasites, resulting in a faster resolution of the infection. Albendazole monotherapy is recommended for patients with one to two parenchymal cysts. The dose of albendazole is 15 mg/kg per day in two daily doses for 10–14 days. A combination of albendazole plus praziquantel is recommended for patients with more than two viable cysts. Viable cysts are defined as those in the vesicular or colloidal stages (see above). The recommended dose of praziquantel is 50 mg/kg per day for 10–14 days. Prednisone or dexamethasone is begun prior to anticysticidal therapy to reduce the host inflammatory response to degenerating parasites. Only cysts in the vesicular stage, where the cyst contains living larva (scolex seen on CT or MRI), and cysts in the colloidal stage, as the larva degenerates (edema surrounds the lesion), are treated with anticysticidal therapy. Some, but not all, experts recommend anticysticidal therapy for lesions that are in the granulonodular stage (surrounded by a contrast-enhancing ring). There is universal agreement that calcified lesions do not need to be treated with anticysticidal therapy. Antiepileptic therapy can be stopped once a follow-up CT or MRI scan shows resolution of the lesion and the patient has had no seizures for 24 consecutive months. Long-term antiepileptic therapy is recommended when seizures occur after resolution of edema and resorption or calcification of the degenerating cyst.

CNS toxoplasmosis is treated with a combination of sulfadiazine, 1.5–2.0 g orally qid, plus pyrimethamine, 100 mg orally to load, then 75–100 mg orally qd, plus folinic acid, 10–15 mg orally qd. Folinic acid is added to the regimen to prevent megaloblastic anemia. Therapy is continued until there is no evidence of active disease on neuroimaging studies, which typically takes at least 6 weeks, and then the dose of sulfadiazine is reduced to 2–4 g/d and pyrimethamine to 50 mg/d. Clindamycin plus pyrimethamine is an alternative therapy for patients who cannot tolerate sulfadiazine, but the combination of pyrimethamine and sulfadiazine is more effective.

SUBDURAL EMPYEMA

A subdural empyema (SDE) is a collection of pus between the dura and arachnoid membranes **(Fig. 145-3)**.

EPIDEMIOLOGY

SDE is a rare disorder that accounts for 15–25% of focal suppurative CNS infections. Sinusitis is the most common predisposing condition and typically involves the frontal sinuses, either alone or in combination with the ethmoid and maxillary sinuses. Sinusitis-associated empyema has a striking predilection for young males, possibly reflecting sex-related differences in sinus anatomy and development. It has been suggested that SDE may complicate 1–2% of cases of frontal sinusitis severe enough to require hospitalization. As a consequence of this epidemiology, SDE shows an ~3:1 male/female predominance, with 70% of cases occurring in the second and third decades of life. SDE may also develop as a complication of head trauma or neurosurgery. Secondary infection of a subdural effusion may also result in empyema, although secondary infection of hematomas, in the absence of a prior neurosurgical procedure, is rare.

ETIOLOGY

Aerobic and anaerobic streptococci, staphylococci, Enterobacteriaceae, and anaerobic bacteria are the most common causative organisms of sinusitis-associated SDE. Staphylococci and gram-negative bacilli are often the etiologic organisms when SDE follows neurosurgical procedures or head trauma. Up to one-third of cases are culture-negative, possibly reflecting difficulty in obtaining adequate anaerobic cultures.

PATHOPHYSIOLOGY

Sinusitis-associated SDE develops as a result of either retrograde spread of infection from septic thrombophlebitis of the mucosal veins

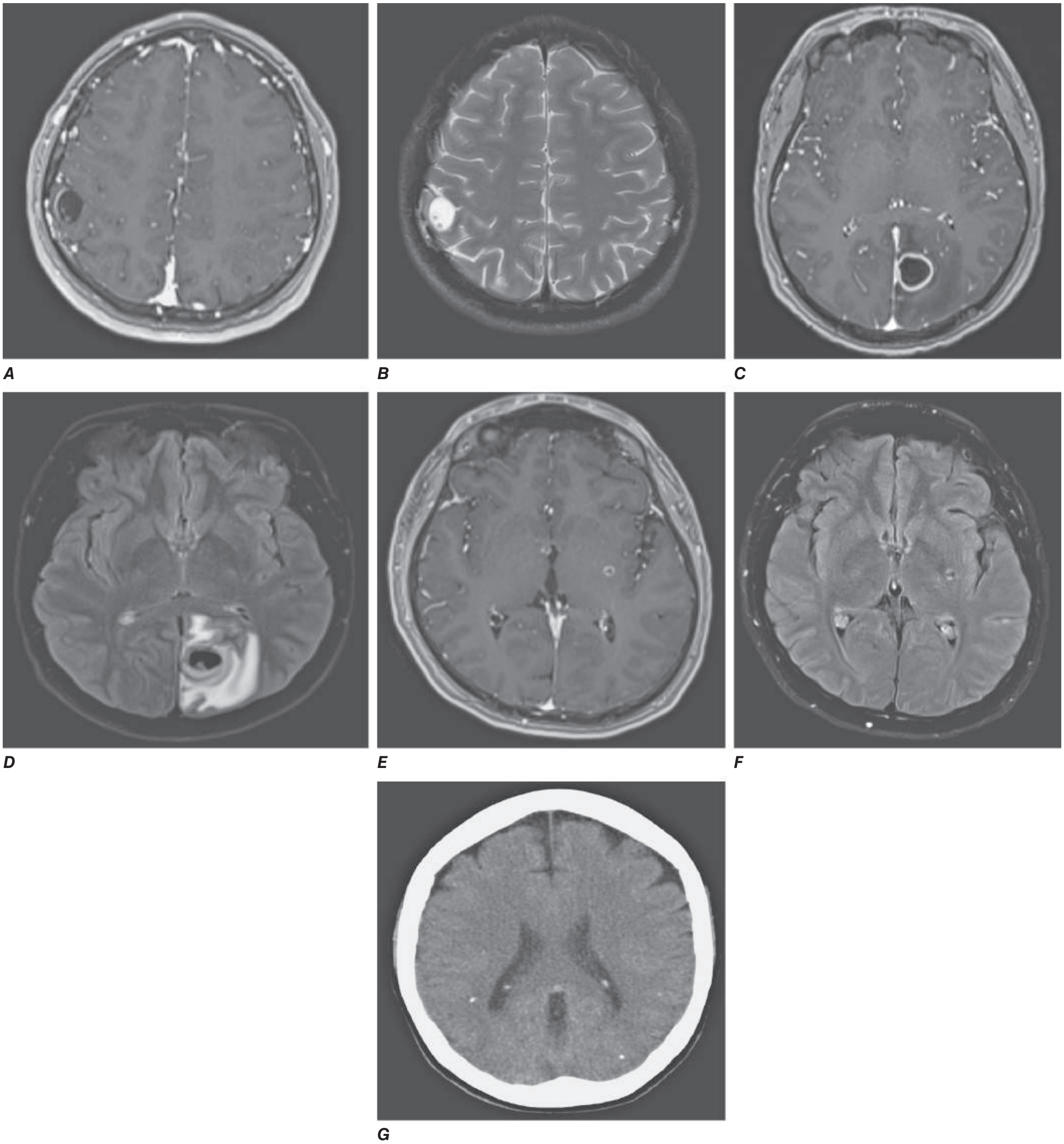

FIGURE 145-2 The four stages of neurocysticercosis. ***A, B.*** The vesicular stage. ***A.*** Postcontrast T1 magnetic resonance image (MRI). Note lesion in right parietal area. Small hypointense nodules within the cyst likely represent scolex. ***B.*** T2 MRI. The cyst is now visualized as a uniform hyperintense lesion with the small hypointense nodules likely representing scolex. No significant edema is present around the lesion on T2, typical for this stage of the disease. ***C, D.*** The colloidal stage. ***C.*** A medial left occipital lesion demonstrates peripheral enhancement on postcontrast imaging. ***D.*** On fluid-attenuated inversion recovery (FLAIR) MRI, the lesion has substantial surrounding hyperintense edema. ***E, F.*** The granulonodular stage. ***E.*** Postcontrast T1-weighted imaging demonstrates enhancing lesions in the left putamen and in the genu of the internal capsule near the foramen of Monro. ***F.*** These lesions demonstrate no surrounding edema on FLAIR imaging, typical for this stage of the disease. ***G.*** The nodular-calcified stage. Computed tomography scan demonstrates typical parenchymal brain calcifications. *(Courtesy of Aaron Kamer, MD; with permission.)*

draining the sinuses or contiguous spread of infection to the brain from osteomyelitis in the posterior wall of the frontal or other sinuses. SDE may also develop from direct introduction of bacteria into the subdural space as a complication of a neurosurgical procedure. The evolution of SDE can be extremely rapid because the subdural space is a large compartment that offers few mechanical barriers to the spread of infection. In patients with sinusitis-associated SDE, suppuration typically begins in the upper and anterior portions of one cerebral hemisphere and then extends posteriorly. SDE is often associated with other intracranial infections, including epidural empyema (40%), cortical thrombophlebitis (35%), and intracranial abscess or cerebritis (>25%). Cortical venous infarction produces necrosis of underlying cerebral cortex and subcortical white matter, with focal neurologic deficits and seizures (see below).

CLINICAL PRESENTATION

A patient with SDE typically presents with fever and a progressively worsening headache. The diagnosis of SDE should always be suspected

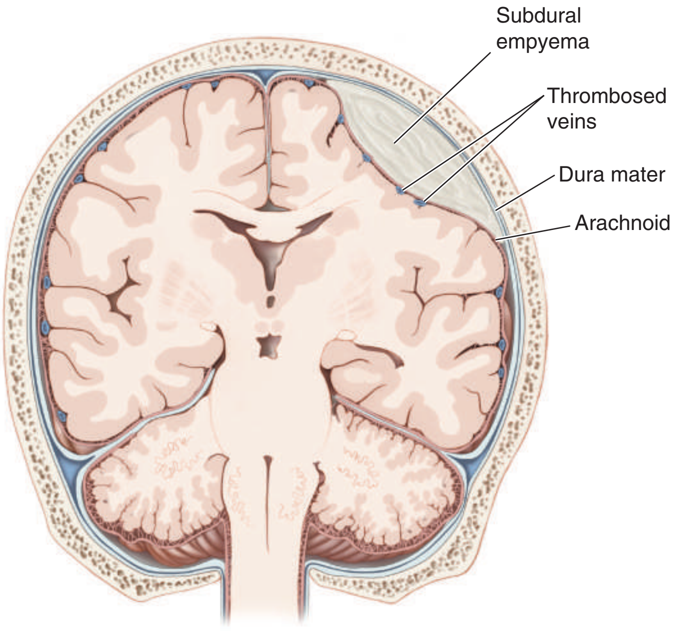

FIGURE 145-3 Subdural empyema.

in a patient with known sinusitis who presents with new CNS signs or symptoms. Patients with underlying sinusitis frequently have symptoms related to this infection. As the infection progresses, focal neurologic deficits, seizures, nuchal rigidity, and signs of increased ICP commonly occur. Headache is the most common complaint at the time of presentation; initially it is localized to the side of the subdural infection, but then it becomes more severe and generalized. Contralateral hemiparesis or hemiplegia is the most common focal neurologic deficit and can occur from the direct effects of the SDE on the cortex or as a consequence of venous infarction. Seizures begin as partial motor seizures that then become secondarily generalized. Seizures may be due to the direct irritative effect of the SDE on the underlying cortex or result from cortical venous infarction. In untreated SDE, the increasing mass effect and increase in ICP cause progressive deterioration in consciousness, leading ultimately to coma.

DIAGNOSIS

MRI **(Fig. 145-4)** is superior to CT in identifying SDE and any associated intracranial infections. The administration of gadolinium greatly improves diagnosis by enhancing the rim of the empyema and allowing the empyema to be clearly delineated from the underlying brain parenchyma. Cranial MRI is also extremely valuable in identifying sinusitis, other focal CNS infections, cortical venous infarction, cerebral edema, and cerebritis. CT may show a crescent-shaped hypodense lesion over one or both hemispheres or in the interhemispheric fissure. Frequently the degree of mass effect, exemplified by midline shift, ventricular compression, and sulcal effacement, is far out of proportion to the mass of the SDE.

CSF examination should be avoided in patients with known or suspected SDE because it adds no useful information and is associated with the risk of cerebral herniation.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of the combination of headache, fever, focal neurologic signs, and seizure activity that progresses rapidly to an altered level of consciousness includes subdural hematoma, bacterial meningitis, viral encephalitis, brain abscess, superior sagittal sinus thrombosis, and acute disseminated encephalomyelitis. The presence of nuchal rigidity is unusual with brain abscess or epidural empyema and should suggest the possibility of SDE when associated with significant focal neurologic signs and fever. Patients with bacterial meningitis also have nuchal rigidity but do not typically have focal deficits of the severity seen with SDE.

TREATMENT

Subdural Empyema

SDE is a medical emergency. Emergent neurosurgical evacuation of the empyema, either through craniotomy, craniectomy, or burr-hole drainage, is the definitive step in the management of this infection. Empirical antimicrobial therapy for community-acquired SDE should include a combination of a third-generation cephalosporin (e.g., cefotaxime or ceftriaxone), vancomycin, and metronidazole **(see Table 143-1 for dosages)**. Patients with hospital-acquired SDE may have infections due to *Pseudomonas* spp. or MRSA and should receive coverage with a carbapenem (e.g., meropenem) and vancomycin. Metronidazole is not necessary for antianaerobic therapy when meropenem is being used. Parenteral antibiotic therapy should be continued for a minimum of 3–4 weeks after SDE drainage. Patients with associated cranial osteomyelitis may require longer therapy. Specific diagnosis of the etiologic organisms is made based on Gram's stain and culture of fluid obtained via either burr holes or craniotomy; the initial empirical antibiotic coverage can be modified accordingly.

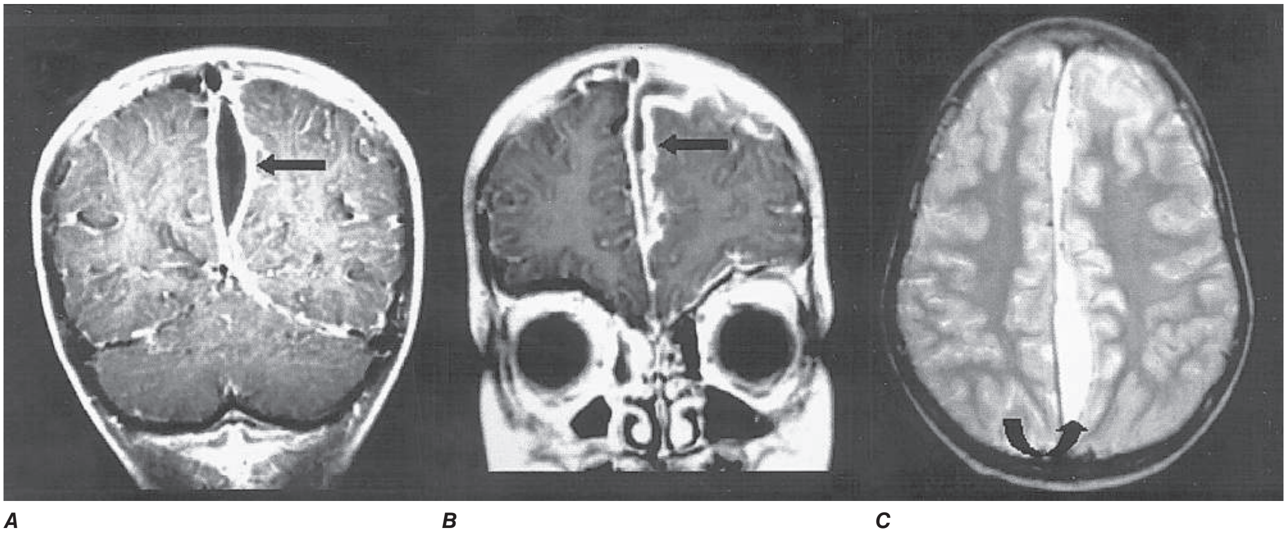

FIGURE 145-4 Subdural empyema. The purulent fluid collection along the left falx and left frontal convexity is hypointense on T1-weighted images (***A***) but markedly hyperintense on the diffusion-weighted (***B***) and the T2 fat-saturation (***C***) magnetic resonance images.

PROGNOSIS

Prognosis is influenced by the level of consciousness of the patient at the time of hospital presentation, the size of the empyema, and the speed with which therapy is instituted. Long-term neurologic sequelae, which include seizures and hemiparesis, occur in up to 50% of cases.

CRANIAL EPIDURAL ABSCESS

Cranial epidural abscess is a suppurative infection occurring in the potential space between the inner skull table and dura **(Fig. 145-5)**.

ETIOLOGY AND PATHOPHYSIOLOGY

Cranial epidural abscess is less common than either brain abscess or SDE and accounts for <2% of focal suppurative CNS infections. A cranial epidural abscess develops as a complication of a craniotomy or compound skull fracture or as a result of spread of infection from the frontal sinuses, middle ear, mastoid, or orbit. An epidural abscess may develop contiguous to an area of osteomyelitis, when craniotomy is complicated by infection of the wound or bone flap, or as a result of direct infection of the epidural space. Infection in the frontal sinus, middle ear, mastoid, or orbit can reach the epidural space through retrograde spread of infection from septic thrombophlebitis in the emissary veins that drain these areas or by way of direct spread of infection through areas of osteomyelitis. Unlike the subdural space, the epidural space is really a potential rather than an actual compartment. The dura is normally tightly adherent to the inner skull table, and infection must dissect the dura away from the skull table as it spreads. As a result, epidural abscesses are often smaller than SDEs. Cranial epidural abscesses, unlike brain abscesses, only rarely result from hematogenous spread of infection from extracranial primary sites. The bacteriology of a cranial epidural abscess is similar to that of SDE (see above). The etiologic organisms of an epidural abscess that arises from frontal sinusitis, middle-ear infections, or mastoiditis are usually streptococci or anaerobic organisms. Staphylococci or gram-negative organisms are the usual cause of an epidural abscess that develops as a complication of craniotomy or compound skull fracture.

CLINICAL PRESENTATION

Patients present with fever (60%), headache (40%), nuchal rigidity (35%), seizures (10%), and focal deficits (5%). Development of symptoms may be insidious, as the empyema usually enlarges slowly in the confined anatomic space between the dura and the inner table of the skull. Periorbital edema and Pott's puffy tumor, reflecting underlying associated frontal bone osteomyelitis, are present in ~40%. In patients with a recent neurosurgical procedure, wound infection is invariably present, but other symptoms may be subtle and can include altered mental status (45%), fever (35%), and headache (20%). The diagnosis should be considered when fever and headache follow recent head trauma or occur in the setting of frontal sinusitis, mastoiditis, or otitis media.

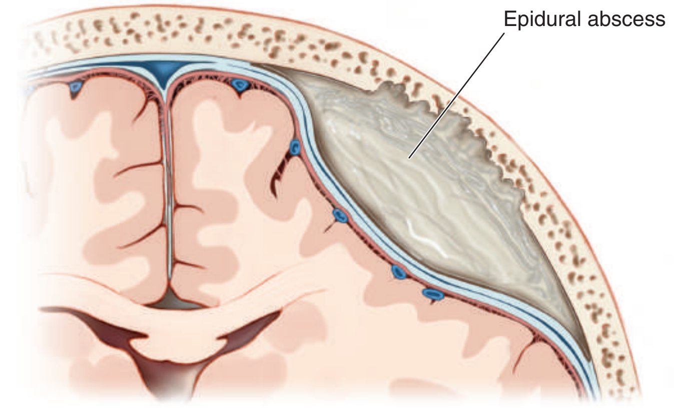

FIGURE 145-5 Cranial epidural abscess is a collection of pus between the dura and the inner table of the skull.

DIAGNOSIS

Cranial MRI with gadolinium enhancement is the procedure of choice to demonstrate a cranial epidural abscess. The sensitivity of CT is limited by the presence of signal artifacts arising from the bone of the inner skull table. The CT appearance of an epidural empyema is that of a lens or crescent-shaped hypodense extraaxial lesion. On MRI, an epidural empyema appears as a lentiform or crescent-shaped fluid collection that is hyperintense compared to CSF on T2-weighted images. On T1-weighted images, the fluid collection may be either isointense or hypointense compared to brain. Following the administration of gadolinium, there is linear enhancement of the dura on T1-weighted images. In distinction to subdural empyema, signs of mass effect or other parenchymal abnormalities are uncommon.

TREATMENT

Epidural Abscess

Immediate neurosurgical drainage is indicated. Empirical antimicrobial therapy, pending the results of Gram's stain and culture of the purulent material obtained at surgery, should include a combination of a third-generation cephalosporin, vancomycin, and metronidazole **(see Table 143-1)**. Ceftazidime or meropenem should be substituted for ceftriaxone or cefotaxime in neurosurgical patients. Metronidazole is not necessary for antianaerobic coverage in patients receiving meropenem. When the organism has been identified, antimicrobial therapy can be modified accordingly. Antibiotics should be continued for 3–6 weeks after surgical drainage. Patients with associated osteomyelitis may require additional therapy.

PROGNOSIS

The mortality rate is <5% in modern series, and full recovery is the rule in most survivors.

SUPPURATIVE THROMBOPHLEBITIS

DEFINITION

Suppurative intracranial thrombophlebitis is septic venous thrombosis of cortical veins and sinuses. This may occur as a complication of bacterial meningitis; SDE; epidural abscess; or infection in the skin of the face, paranasal sinuses, middle ear, or mastoid.

ANATOMY AND PATHOPHYSIOLOGY

The cerebral veins and venous sinuses have no valves; therefore, blood within them can flow in either direction. The superior sagittal sinus is the largest of the venous sinuses **(Fig. 145-6)**. It receives blood from the

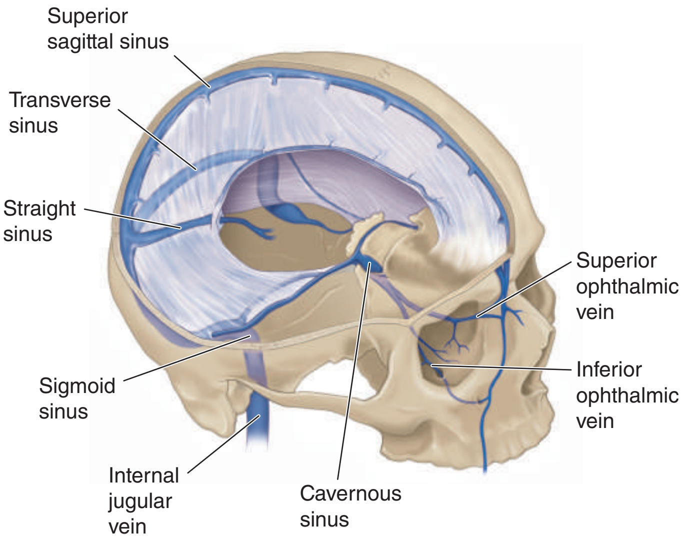

FIGURE 145-6 Anatomy of the cerebral venous sinuses.

frontal, parietal, and occipital superior cerebral veins and the diploic veins, which communicate with the meningeal veins. Bacterial meningitis is a common predisposing condition for septic thrombosis of the superior sagittal sinus. The diploic veins, which drain into the superior sagittal sinus, provide a route for the spread of infection from the meninges, especially in cases where there is purulent exudate near areas of the superior sagittal sinus. Infection can also spread to the superior sagittal sinus from nearby SDE or epidural abscess. Dehydration from vomiting, hypercoagulable states, and immunologic abnormalities, including the presence of circulating antiphospholipid antibodies, also contribute to cerebral venous sinus thrombosis. Thrombosis may extend from one sinus to another, and at autopsy, thrombi of different histologic ages can often be detected in several sinuses. Thrombosis of the superior sagittal sinus is often associated with thrombosis of superior cortical veins and small parenchymal hemorrhages.

The superior sagittal sinus drains into the transverse sinuses (Fig. 145-6). The transverse sinuses also receive venous drainage from small veins from both the middle ear and mastoid cells. The transverse sinus becomes the sigmoid sinus before draining into the internal jugular vein. Septic transverse/sigmoid sinus thrombosis can be a complication of acute and chronic otitis media or mastoiditis. Infection spreads from the mastoid air cells to the transverse sinus via the emissary veins or by direct invasion. The cavernous sinuses are inferior to the superior sagittal sinus at the base of the skull. The cavernous sinuses receive blood from the facial veins via the superior and inferior ophthalmic veins. Bacteria in the facial veins enter the cavernous sinus via these veins. Bacteria in the sphenoid and ethmoid sinuses can spread to the cavernous sinuses via the small emissary veins. The sphenoid and ethmoid sinuses are the most common sites of primary infection resulting in septic cavernous sinus thrombosis.

CLINICAL MANIFESTATIONS

Septic thrombosis of the superior sagittal sinus presents with headache, fever, nausea and vomiting, confusion, and focal or generalized seizures. There may be a rapid development of stupor and coma. Weakness of the lower extremities with bilateral Babinski's signs or hemiparesis is often present. When superior sagittal sinus thrombosis occurs as a complication of bacterial meningitis, nuchal rigidity and Kernig's and Brudzinski's signs may be present.

The oculomotor nerve, the trochlear nerve, the abducens nerve, the ophthalmic and maxillary branches of the trigeminal nerve, and the internal carotid artery all pass through the cavernous sinus **(see Fig. 452-7)**. The symptoms of *septic cavernous sinus thrombosis* are fever, headache, frontal and retroorbital pain, and diplopia. The classic signs are ptosis, proptosis, chemosis, and extraocular dysmotility due to deficits of cranial nerves III, IV, and VI; hyperesthesia of the ophthalmic and maxillary divisions of the fifth cranial nerve and a decreased corneal reflex may be detected. There may be evidence of dilated, tortuous retinal veins and papilledema.

Headache and earache are the most frequent symptoms of *transverse sinus thrombosis*. A transverse sinus thrombosis may also present with otitis media, sixth nerve palsy, and retroorbital or facial pain (*Gradenigo's syndrome*). Sigmoid sinus and internal jugular vein thrombosis may present with neck pain.

DIAGNOSIS

The diagnosis of septic venous sinus thrombosis is suggested by an absent flow void within the affected venous sinus on MRI and confirmed by contrast-enhanced magnetic resonance venography, CT venography, or the venous phase of cerebral angiography. The diagnosis of thrombophlebitis of intracerebral and meningeal veins is suggested by the presence of intracerebral hemorrhage but requires the venous phase of cerebral angiography for definitive diagnosis.

TREATMENT

Suppurative Thrombophlebitis

Septic venous sinus thrombosis is treated with antibiotics, hydration, and removal of infected tissue and thrombus in septic lateral or cavernous sinus thrombosis. The choice of antimicrobial therapy is based on the bacteria responsible for the predisposing or associated condition. Optimal duration of therapy is unknown, but antibiotics are usually continued for 6 weeks or until there is radiographic evidence of resolution of thrombosis. Anticoagulation with unfractionated or low-molecular-weight heparin is recommended for aseptic venous sinus thrombosis and in the treatment of septic venous sinus thrombosis complicating bacterial meningitis in patients who have progressive neurologic deterioration despite antimicrobial therapy and intravenous fluids. The presence of a small intracerebral hemorrhage from septic thrombophlebitis is not an absolute contraindication to heparin therapy. Successful management of aseptic venous sinus thrombosis has been reported with surgical thrombectomy, catheter-directed urokinase therapy, and a combination of intrathrombus recombinant tissue plasminogen activator (rtPA) and intravenous heparin, but there are not enough data to recommend these therapies in septic venous sinus thrombosis.

FURTHER READING

Bodilsen J et al: European Society of Clinical Microbiology and Infectious Diseases guidelines on diagnosis and treatment of brain abscess in children and adults. Clinical Microbiol Infect 30:66, 2024.

Prosty C et al: Revisiting the evidence base for modern-day practice of the treatment of toxoplasmic encephalitis: A systematic review and meta-analysis. Clin Infect Dis 76:e1302, 2023.

Ropper AH, Klein JP: Cerebral venous thrombosis. N Engl J Med 385:59, 2021.

White AC et al: Diagnosis of Neurocysticercosis: 2017 Clinical Practice Guidelines by the Infectious Diseases Society of America (IDSA) and the American Society of Tropical Medicine and Hygiene (ASTMH). Clin Infect Dis 66:e49, 2018.

146

Infectious Complications of Bites

Nongnooch Poowanawittayakom,
Lawrence C. Madoff

The skin is an essential component of nonspecific immunity, protecting the host from potential pathogens in the environment. Breaches in this protective barrier thus represent a form of immunocompromise that predisposes the patient to infection. Bites and scratches from animals and humans allow the inoculation of microorganisms past the skin's protective barrier into deeper, susceptible host tissues.

Each year in the United States, millions of animal-bite wounds are sustained. The vast majority are inflicted by pet dogs and cats, which number >100 million; the annual incidence of dog and cat bites has been reported as 300 bites per 100,000 population. Other bite wounds are a consequence of encounters with animals in the wild or in occupational settings. While many of these wounds require minimal or no therapy, a significant number result in infection, which may be life-threatening. The microbiology of bite-wound infections in general reflects the oropharyngeal flora of the biting animal, although organisms from the soil, the skin of the animal and the victim, and the animal's feces may also be involved.

DOG BITES

In the United States, dogs bite >4.7 million people each year and are responsible for 80% of all animal-bite wounds, an estimated 15–20% of which become infected. Each year, 800,000 Americans seek medical attention for dog bites; of those injured, 386,000 require treatment in an emergency department, with >1000 emergency department visits each day and ~43 deaths per year. Most dog bites are provoked and are inflicted by the victim's pet or by a dog known to the victim. These bites are frequently sustained during efforts to break up a dogfight. Children are more likely than adults to sustain canine bites, with the highest incidence of 6 bites per 1000 population among boys 5–9 years old. Victims are more often male than female, and bites most often involve an upper extremity. Among children <4 years old, two-thirds of all these injuries involve the head or neck. Infection typically manifests 8–24 h after the bite as pain at the site of injury with cellulitis accompanied by purulent, sometimes foul-smelling discharge. Septic arthritis and osteomyelitis may develop if a canine tooth penetrates synovium or bone. Systemic manifestations (e.g., fever, lymphadenopathy, and lymphangitis) also may occur. The microbiology of dog-bite wound infections is usually mixed and includes *Pasteurella* species, β-hemolytic streptococci, *Staphylococcus* species (including methicillin-resistant *Staphylococcus aureus* [MRSA] and *Staphylococcus intermedius*), *Neisseria* species (commonly *Neisseria weaveri*, formerly known as CDC group M-5), *Eikenella corrodens*, and *Capnocytophaga canimorsus*. Many wounds also include anaerobic bacteria such as *Bacteroides*, *Fusobacterium*, *Prevotella*, and *Porphyromonas* species.

While most infections resulting from dog-bite injuries are localized to the area of injury, many of the microorganisms involved are capable of causing systemic infection, including bacteremia, meningitis, brain abscess, endocarditis, and chorioamnionitis. These infections are particularly likely in hosts with edema or compromised lymphatic drainage in the involved extremity (e.g., after a bite on the arm in a woman who has undergone mastectomy) and in patients who are immunocompromised by medication or disease (e.g., glucocorticoid use, systemic lupus erythematosus, acute leukemia, or hepatic cirrhosis). In addition, dog bites and scratches may result in systemic illnesses such as rabies **(Chap. 214)** and tetanus **(Chap. 157)**.

Infection with *Capnocytophaga canimorsus* (and other *Capnocytophaga* species) following dog-bite wounds (or licking of preexisting wounds) may result in fulminant sepsis, disseminated intravascular coagulation, and renal failure, particularly in hosts who have impaired hepatic function, who have undergone splenectomy, or who are immunosuppressed. This thin gram-negative rod is difficult to culture on most solid media but grows in a variety of liquid media. It may require up to 14 days of incubation to grow on blood cultures. The bacteria are occasionally seen within polymorphonuclear leukocytes on Wright-stained smears of peripheral blood from septic patients. Tularemia **(Chap. 175)** also has been reported to follow dog bites.

CAT BITES

Although less common than dog bites, cat bites and scratches result in infection in more than half of all cases. Because the cat's narrow, sharp canine teeth penetrate deeply into tissue, cat bites are more likely than dog bites to cause septic arthritis and osteomyelitis; the development of these conditions is particularly likely when punctures are located over or near a joint, especially in the hand. Women sustain cat bites more frequently than do men. These bites most often involve the hands and arms. Both bites and scratches from cats are prone to infection from organisms in the cat's oropharynx. *Pasteurella multocida*, a normal component of the feline oral flora, is a small gram-negative coccobacillus implicated in the majority of cat-bite wound infections. Like that of dog-bite wound infections, however, the microflora of cat-bite wound infections is usually mixed. However, the median time from bite to the appearance of signs and symptoms of wound infection is much shorter when compared to dog bites. Other microorganisms causing infection after cat bites are similar to those causing dog-bite wound infections.

The same risk factors for systemic infection following dog-bite wounds apply to cat-bite wounds. *Pasteurella* infections tend to advance rapidly, often within hours, causing severe inflammation accompanied by purulent drainage with adenitis; *Pasteurella* may also be spread by respiratory droplets from animals, resulting in pneumonia or bacteremia. Like dog-bite wounds, cat-bite wounds may result in the transmission of rabies or in the development of tetanus. Infection with *Bartonella henselae* causes cat-scratch disease **(Chap. 177)** and is an important late consequence of cat bites and scratches. Tularemia **(Chap. 175)** also has been reported to follow cat bites. Occasionally, sporotrichosis **(Chap. 225)** has been associated with scratches or bites by animals, especially domestic cats.

OTHER ANIMAL BITES

Infections have been attributed to bites from many animal species. Often these bites are sustained as a consequence of occupational exposure (farmers, laboratory workers, veterinarians) or recreational exposure (hunters and trappers, wilderness campers, owners of exotic pets). Generally, the microflora of bite wounds reflects the oral flora of the biting animal. Most members of the cat family, including feral cats, harbor *P. multocida*. Bite wounds from aquatic animals such as alligators or piranhas may contain *Aeromonas hydrophila*. Shark, moray eel, and barracuda bites, like other injuries sustained in saltwater, are often associated with infections with marine *Vibrio* species. Venomous snakebites **(Chap. 471)** result in severe inflammatory responses and tissue necrosis—conditions that render these injuries prone to infection. The snake's oral flora includes many species of aerobes and anaerobes, such as *Pseudomonas aeruginosa, Serratia marcescens, Proteus* species, *Staphylococcus epidermidis, Salmonella* species, *Bacteroides fragilis*, and *Clostridium* species. Bites from nonhuman primates are highly susceptible to infection with pathogens similar to those isolated from human bites (see below). Bites from Old World monkeys (*Macaca*) may also result in the transmission of B virus (*Macacine herpesvirus 1, Herpesvirus simiae, Cercopithecine herpesvirus*), a cause of serious infection of the human central nervous system. *Actinobacillus lignieresii* has often been reported in infected wounds of humans bitten by horses, pigs, and sheep. Bites of seals, walruses, and polar bears may cause a chronic suppurative infection known as *seal finger*, which is probably due to one or more species of *Mycoplasma,* including *Mycoplasma phocacerebrale,* colonizing these animals.

Small rodents, including rats, mice, and gerbils, as well as animals that prey on rodents may transmit *Streptobacillus moniliformis* (a microaerophilic, pleomorphic gram-negative rod) or *Spirillum minor* (a spirochete); these organisms cause a clinical illness known as *rat-bite fever*. The vast majority of cases in the United States are streptobacillary, whereas *Spirillum* infection occurs mainly in Asia.

In the United States, the risk of rodent bites is usually greatest among laboratory workers or inhabitants of rodent-infested dwellings (particularly children). Rat-bite fever is distinguished from acute bite-wound infection by its typical manifestation after the initial wound has healed. Streptobacillary disease follows an incubation period of 3–10 days. Fever, chills, myalgias, headache, and severe migratory arthralgias are usually followed by a maculopapular rash, which characteristically involves the palms and soles and may become confluent or purpuric. Complications include endocarditis, myocarditis, meningitis, pneumonia, and abscesses in many organs. *Haverhill fever* is an *S. moniliformis* infection acquired from contaminated milk or drinking water and has similar manifestations. Streptobacillary rat-bite fever was frequently fatal in the preantibiotic era. The differential diagnosis includes Rocky Mountain spotted fever, Lyme disease, leptospirosis, and secondary syphilis. The diagnosis is made by direct observation of the causative organisms in tissue or blood, by culture of the organisms on enriched media, or by serologic testing with specific agglutinins.

Spirillum infection (referred to in Japan as *sodoku*) causes pain and purple swelling at the site of the initial bite, with associated lymphangitis and regional lymphadenopathy, after an incubation period of 1–4 weeks. The systemic illness includes fever, chills, and headache. The original lesion may eventually progress to an eschar. The infection is diagnosed by direct visualization of the spirochetes in blood or tissue or by animal inoculation.

HUMAN BITES

Human bites may be self-inflicted; may be sustained by medical personnel caring for patients; or may take place during fights, domestic abuse, or sexual activity. The risk of infection in human-bite wounds depends on the depth of the wound. The risk of wound infection ranges from about 2% in superficial wounds to over 25% in deep bite wounds such as clenched-fist injuries. Human-bite wounds become infected more frequently (~10–15% of the time) than do bites inflicted by other animals. These infections reflect the diverse oral microflora of humans, which includes multiple species of aerobic and anaerobic bacteria. Common aerobic isolates include viridans streptococci, *S. aureus*, *E. corrodens* (which is particularly common in clenched-fist injury; see below), and *Haemophilus influenzae*. Anaerobic species, including *Fusobacterium nucleatum* and *Prevotella*, *Porphyromonas*, and *Peptostreptococcus* species, are isolated from 50% of wound infections due to human bites; many of these isolates produce β-lactamases. The oral flora of hospitalized and debilitated patients often includes Enterobacteriaceae in addition to the usual organisms. Hepatitis B, hepatitis C, herpes simplex virus infection, syphilis, tuberculosis, actinomycosis, and tetanus have been reported to be transmitted by human bites; it is biologically possible to transmit HIV through human bites, although the risk is quite low. In general, postexposure prophylaxis should be considered for bites involving severe trauma with extensive tissue damage and the presence of blood in saliva. There is essentially no risk of transmission if the skin is intact.

Human bites are categorized as either *occlusional* injuries, which are inflicted by actual biting, or *clenched-fist* injuries, which are sustained when the fist of one individual strikes the teeth of another, causing traumatic laceration of the hand. For several reasons, clenched-fist injuries, which are sometimes referred to as "fight bite" and which are more common than occlusional injuries, result in particularly serious infections. The deep spaces of the hand, including the bones, joints, and tendons, are frequently inoculated with organisms in the course of such injuries. The clenched position of the fist during injury, followed by extension of the hand, may further promote the introduction of bacteria as contaminated tendons retract beneath the skin's surface Moreover, medical attention is often sought only after frank infection develops. Patients with clenched-fist injury should undergo careful physical examination of the area, including the extensor tendons.

APPROACH TO THE PATIENT

Animal or Human Bites

A careful history should be elicited, including the type of biting animal, the type of attack (provoked or unprovoked), and the amount of time elapsed since injury. Local and regional public-health authorities should be contacted to determine whether an individual species could be rabid and/or to locate and observe the biting animal when rabies prophylaxis may be indicated **(Chap. 214)**. Suspicious human-bite wounds should provoke careful questioning regarding domestic or child abuse. Details on antibiotic allergies, immunosuppression, splenectomy, liver disease, mastectomy, and immunization history should be obtained. The wound should be inspected carefully for evidence of infection, including redness, exudate, and foul odor. The type of wound (puncture, laceration, or scratch); the depth of penetration; and the possible involvement of joints, tendons, nerves, and bones should be assessed. It is often useful to include a diagram or photograph of the wound in the medical record. In addition, a general physical examination should be conducted and should include an assessment of vital signs as well as an evaluation for evidence of lymphangitis, lymphadenopathy, dermatologic lesions, and functional limitations. Injuries to the hand warrant consultation with a hand surgeon for the assessment of tendon, nerve, and muscular damage. Radiographs should be obtained in penetrating wounds to evaluate the evidence of fracture or retained foreign body such as a tooth fragment. Culture and Gram's staining of all infected wounds are essential; anaerobic cultures should be undertaken if abscesses, devitalized tissue, or foul-smelling exudate is present. A small-tipped swab may be used to culture deep punctures or small lacerations. It is also reasonable to culture samples from apparently uninfected wounds due to bites inflicted by animals other than dogs and cats, since the microorganisms causing disease are less predictable in these cases. The microbiology laboratory should be notified if fastidious organisms such as *E. corrodens* are under consideration in human bites. The white blood cell count should be determined, and the blood cultured if systemic infection is suspected.

TREATMENT

Bite-Wound Infections

WOUND MANAGEMENT

Wound closure is controversial in bite injuries. Many authorities prefer not to attempt primary closure of wounds that are or may become infected, choosing instead to irrigate these wounds copiously, debride devitalized tissue, remove foreign bodies, and approximate the wound edges. All abscesses should be drained. Delayed primary closure may be undertaken after the risk of infection is over. Small uninfected wounds may be allowed to close by secondary intention. Puncture wounds due to cat bites should be left unsutured because of the high rate at which they become infected. Facial wounds are usually sutured after thorough cleaning and irrigation because of the importance of a good cosmetic result in this area and because anatomic factors such as an excellent blood supply and the absence of dependent edema lessen the risk of infection. In general, wounds >12 h old (for bites to the arm or leg) or >24 h old (for bites to the face) should not be closed primarily and may require prophylactic antibiotics (see below).

ANTIBIOTIC THERAPY

Established Infection Antibiotics should be administered for all established bite-wound infections and should be chosen in light of the most likely potential pathogens, as indicated by the biting species and by Gram's stain and culture results **(Table 146-1)**. For dog and cat bites, antibiotics should be effective against *S. aureus, Pasteurella* species, *C. canimorsus*, streptococci, and oral anaerobes. For human bites, agents with activity against *S. aureus, H. influenzae*, and β-lactamase-positive oral anaerobes should be used. The combination of an extended-spectrum penicillin with a β-lactamase inhibitor (amoxicillin/clavulanic acid, ampicillin/sulbactam) appears to offer the most reliable coverage for these pathogens. Third-generation cephalosporins (ceftriaxone, cefpodoxime) also offer substantial coverage when given in conjunction with a drug that provides anaerobic coverage (clindamycin or metronidazole). The choice of antibiotics for penicillin-allergic patients (particularly those in whom immediate-type hypersensitivity makes the use of cephalosporins hazardous) is more difficult and is based primarily on in vitro sensitivity since data on clinical efficacy are inadequate. The combination of an antibiotic active against gram-positive cocci and anaerobes (such as clindamycin or metronidazole) with trimethoprim-sulfamethoxazole or a fluoroquinolone, which is active against many of the other potential pathogens, would appear reasonable. Moxifloxacin, a fluoroquinolone with anaerobic coverage, can also be considered as a single agent. In vitro data suggest that azithromycin alone provides coverage against most commonly isolated bite-wound pathogens; however, this agent has variable activity against *P. multocida*, *E. corrodens,* and fusobacteria and thus should be avoided unless no alternative agent is available. Empirical use of agents active against MRSA should be considered in high-risk situations while culture results are awaited.

Antibiotics are generally given for about 5–7 days and for no more than 14 days, but the response to therapy must be carefully monitored. Failure to respond should prompt a consideration of diagnostic alternatives and surgical evaluation for possible drainage or debridement. Complications such as osteomyelitis or septic arthritis mandate a longer duration of therapy.

TABLE 146-1 Management of Wound Infections Following Animal and Human Bites

BITING SPECIES	COMMONLY ISOLATED PATHOGENS	PREFERRED ANTIBIOTIC(S)[a]	ALTERNATIVE IN PENICILLIN-ALLERGIC PATIENT	PROPHYLAXIS ADVISED FOR EARLY UNINFECTED WOUNDS	OTHER CONSIDERATIONS
Dog	*Staphylococcus aureus, Pasteurella* spp. (mainly *P. multocida* and *P. canis*), anaerobes, *Capnocytophaga canimorsus*	Amoxicillin/clavulanate (875/125 mg PO q12h) or ampicillin/sulbactam (3.0 g IV q6h) or Ceftriaxone 2 g IV once daily plus metronidazole 500 mg q8h	Clindamycin[b] or metronidazole plus either TMP-SMX (1 DS tablet PO bid) or ciprofloxacin (500 mg PO bid)	Sometimes[c]	Consider rabies prophylaxis.
Cat	*P. multocida, S. aureus*, anaerobes	Amoxicillin/clavulanate, ampicillin/sulbactam, or ceftriaxone plus metronidazole as above	Clindamycin or metronidazole plus TMP-SMX as above or fluoroquinolone	Usually[d]	Consider rabies prophylaxis. Carefully evaluate for joint/bone penetration.
Human, occlusional	Viridans streptococci, *S. aureus, Haemophilus influenzae*, anaerobes *Eikenella corrodens*[e]	Amoxicillin/clavulanate plus TMX-SMX if consider including MRSA coverage or ampicillin/sulbactam or ceftriaxone plus metronidazole (consider adding vancomycin if MRSA coverage required)	TMP-SMX plus metronidazole	Always	
Monkey	As for human bite	As for human bite	As for human bite	Always	For macaque monkeys, consider B virus prophylaxis with acyclovir.
Snake	Snake oral flora including *Pseudomonas, Morganella* spp., *E. coli*, group D streptococci, *Salmonella* spp., anaerobic organisms including *Bacteroides fragilis, Clostridium* spp.	Piperacillin/tazobactam 3.375 g IV q6–8h	Clindamycin or metronidazole plus a fluoroquinolone	Evidence does not support the benefit. Can consider in regions with high rates of infection such as Brazil.	Administer antivenin for venomous snakebite. Tetanus prophylaxis.
Rodent	Rat bite fever; *Streptobacillus moniliformis, Spirillum minus, Streptobacillus notomytis, Leptospira* spp., *P. multocida*	Penicillin VK (500 mg PO qid) or ceftriaxone IV	Doxycycline (100 mg PO bid)	Sometimes	
Aquatic animal (alligator, piranha, shark, moray eel, barracuda)	*Aeromonas hydrophila*, marine *Vibrio* spp. (*Vibrio vulnificus*)	Third-generation cephalosporin (e.g., ceftriaxone, 1 g IV q24h) plus doxycycline (100 mg PO bid)	Clindamycin or metronidazole plus levofloxacin (750 mg PO qd) plus doxycycline	Always	Obtain prompt surgical consultation, as risk for necrotizing infection is high with *Aeromonas* and *Vibrio* spp.

[a] Antibiotic choices should be based on culture data when available. These suggestions for empirical therapy need to be tailored to individual circumstances and local conditions. MRSA empirical coverage is based on individual risk factors. IV regimens should be used for hospitalized patients. A single IV dose of antibiotics may be given to patients who will be discharged after initial management. [b] Avoid monotherapy for aerobic coverage due to poor activity against *Pasteurella* spp. Risk of *Clostridioides difficile* infection. [c] Cat bite may leave small external wounds but deep puncture wounds. [d] Prophylactic antibiotics are suggested for severe or extensive wounds, facial wounds, and crush injuries; when bone or joint may be involved; delayed wound care >8 h; and when comorbidity is present (see text). Prophylactic antibiotic duration is generally between 3 and 5 days. [e] *Eikenella corrodens* is resistant to penicillinase-resistant penicillin, first and second generation cephalosporins, clindamycin, metronidazole, and aminoglycosides in vitro.

Abbreviations: DS, double-strength; TMP-SMX, trimethoprim-sulfamethoxazole.

Management of *C. canimorsus* bacteremia requires an initial treatment with a 2-week course of IV antibiotics such as penicillin G (2 million units IV every 4 h) or IV ampicillin/sulbactam (3.0 g every 6 h) along with supportive measures. Once the patient is improved, then a switch to oral antibiotics can be considered. Alternative agents for the treatment of *C. canimorsus* infection include cephalosporins or carbapenems. Serious infection with *P. multocida* (e.g., pneumonia, sepsis, or meningitis) also should be treated with IV penicillin G. Alternative agents include a second- or third-generation cephalosporin or ciprofloxacin. Penicillin resistance is uncommon.

Bites by venomous snakes **(Chap. 471)** may not require antibiotic treatment. Because it is often difficult to distinguish signs of infection from tissue damage caused by the envenomation, many authorities continue to recommend treatment directed against the snake's oral flora—i.e., the administration of broadly active agents such as ceftriaxone (1–2 g IV every 24 h) or ampicillin/sulbactam (3.0 g IV every 6 h).

Seal finger appears to respond to doxycycline (100 mg twice daily for a duration guided by the response to therapy).

Presumptive or Prophylactic Therapy The use of antibiotics for patients presenting early (within 8 h) after bite injury is controversial. Although symptomatic infection frequently will not yet have manifested at this point, many early wounds will harbor pathogens, and many will become infected. Studies of antibiotic prophylaxis for wound infections are limited and have often included only small numbers of cases in which various types of wounds have been managed according to various protocols. A meta-analysis of eight randomized trials of prophylactic antibiotics in patients with dog-bite wounds demonstrated a reduction in the rate of infection by 50% with prophylaxis. However, in the absence of sound clinical trials, many clinicians base the decision to treat bite wounds with empirical antibiotics on the species of the biting animal; the location, severity, and extent of the bite wound; and the existence of comorbid conditions in the host. All human- and monkey-bite wounds should be treated presumptively because of the high rate of infection. Most cat-bite wounds, particularly those involving the hand, should consider prophylactic antibiotics. Other factors favoring treatment for bite wounds include severe injury, as in crush wounds; potential bone or joint involvement; involvement of the hands or genital region; host immunocompromise, including that due to diabetes mellitus, liver disease, or splenectomy; involvement of extremities with underlying venous and/or lymphatic compromise; and prior mastectomy on the side of an involved upper extremity. When prophylactic antibiotics are administered, they are usually given for 3–5 days.

Rabies and Tetanus Prophylaxis Rabies prophylaxis, consisting of both passive administration of rabies immune globulin (with as much of the dose as possible infiltrated into and around the wound) and active immunization with rabies vaccine, should be given in consultation with local and regional public-health authorities for some animal bites and scratches as well as for certain nonbite exposures **(Chap. 214)**. Rabies is endemic in a variety of animals, including dogs and cats, in many areas of the world. In the United States, although the majority (90%) of rabid animals reported each year are wild (including raccoons, skunks, foxes, and bats), most rabies prophylaxis is given because of close contact with domestic animals. More cats than dogs are reported rabid each year. Many local health authorities require the reporting of all animal bites.

A tetanus booster immunization should be given if the patient has undergone primary immunization but has not received a booster dose in the past 5 years. Patients who have not previously completed primary immunization should be immunized and should also receive tetanus immune globulin. Elevation of the site of injury is an important adjunct to antimicrobial therapy. Immobilization of the infected area, especially the hand, also is beneficial.

Hepatitis B Prophylaxis Hepatitis B virus can be transmitted, albeit rarely, by exposure of nonintact skin to blood-free saliva. The mainstay of postexposure prophylaxis is active immunization with hepatitis B vaccine, but, in certain circumstances, hepatitis B immune globulin is recommended in addition to vaccine for added protection **(Chap. 350)**.

Acknowledgment

The authors would like to acknowledge Drs. Sandeep S. Jubbal and Florencia Pereyra for their prior contributions to this chapter.

FURTHER READING

Abrahamian FM, Goldstein EJC: Microbiology of animal bite wound infections. Clin Microbiol Rev 24:231, 2011.

Brook I: Management of human and animal bite wounds: An overview. Adv Skin Wound Care 18:197, 2005.

Bystritsky R, Chambers H: Cellulitis and soft tissue infections. Ann Intern Med 168:ITC17, 2018.

Ellis R, Ellis C: Dog and cat bites. Am Fam Phys 90:239, 2014.

Fallouji MA: Traumatic love bites. Br J Surg 77:100, 1990.

Fleisher GR: The management of bite wounds. N Engl J Med 340:138, 1999.

Kullberg BJ et al: Purpura fulminans and symmetrical peripheral gangrene caused by *Capnocytophaga canimorsus* (formerly DF-2) septicemia—a complication of dog bite. Medicine (Baltimore) 70:287, 1991.

Lohiya GS et al: Human bites: Bloodborne pathogen risk and postexposure follow-up algorithm. J Natl Med Assoc 105:92, 2013.

Martino R et al: Bacteremia caused by *Capnocytophaga* species in patients with neutropenia and cancer: Results of a multicenter study. Clin Infect Dis 33:e20, 2001.

Morgan M, Palmer J: Dog bites. BMJ 334:413, 2007.

Oehler RL et al: Bite-related and septic syndromes caused by cats and dogs. Lancet Infect Dis 9:439, 2009.

Stevens DL et al: Practice guidelines for the diagnosis and management of skin and soft tissue infections. 2014 update by the Infectious Diseases Society of America. Clin Infect Dis 59:e10, 2014.

Weber DJ et al: Infections resulting from animal bites. Infect Dis Clin North Am 5:663, 1991.

World Health Organization, Regional Office for South-East Asia: Guidelines for the management of snakebites, 2nd ed, 2016. Available at *https://iris.who.int/handle/10665/249547*. Accessed February 13, 2024.

Section 3 Clinical Syndromes: Health Care–Associated Infections

147 Infections Acquired in Health Care Facilities

Mini Kamboj, Tara N. Palmore

Health care–associated infections affect at least 3% of hospitalized patients at any given time. Through concerted efforts, national rates of some nosocomial infections were declining before the onset of the COVID-19 pandemic, but infection control challenges related to the pandemic reversed years of progress. The past few years have also seen a rise in incidence of multidrug-resistant infections, which are challenging to treat and contain. However, newer tools combined with evidence-based methods of infection prevention and control are robust and can succeed. This chapter reviews the epidemiology, prevention, and control of health care–associated infections and recent challenges faced by health care epidemiologists.

ORGANIZATION, RESPONSIBILITIES, AND OVERSIGHT OF INFECTION PREVENTION AND CONTROL PROGRAMS

Infection prevention and control programs are composed of infection preventionists supervised by an experienced team lead. These typically include a doctoral-level (MD/DO/PhD) health care epidemiologist who may report to the chief medical officer or chief quality officer. The number of staff required in an infection prevention and control program depends on the size and complexity of the health care facility and its patients.

Infection prevention and control programs are responsible for a broad range of activities, including surveillance and reporting of nosocomial infections; preventing and thwarting transmission of nosocomial pathogens through use of isolation and education; reducing device-associated infections through evidence-based interventions; collaborating with occupational health to manage infectious exposures; preparing for and managing emerging infectious diseases; and investigating and controlling outbreaks. The team collects and analyzes infection data and reports those data to institutional stakeholders, such as the multidisciplinary Infection Control Committee. Infection preventionists usually perform the mandatory reporting of select nosocomial infection data to the National Healthcare Safety Network that is managed by the Centers for Disease Control and Prevention (CDC). Such reporting is required by the U.S. Centers for Medicare and Medicaid Services and affects facilities' reimbursement for the care they have provided, i.e., nonpayment for care related to preventable nosocomial infections.

SURVEILLANCE

Surveillance to detect and prevent health care–associated infections focuses on outcomes, processes, and other related measures that directly or indirectly influence the risk of contracting them. Examples of outcomes include surgical site infections and hospital-onset *Clostridioides difficile* infections. Key process measures include compliance with evidence-based practices that reduce the risk of infection, such as hand hygiene, central line insertion care, and maintenance practices for indwelling devices. Finally, health care personnel influenza immunization rates are an example of a related measure that is tracked at a local, regional, and national level to gauge efforts toward reducing the risk of nosocomial influenza in acute and long-term care settings.

Detecting health care–associated infections using a case-finding strategy is a labor- and resource-intensive process. Most U.S. hospitals rely on laboratory-based surveillance as the fundamental data collection methodology, supplemented with clinical reviews by infection preventionists.

Widespread adoption of automated surveillance systems to collect, analyze, and combine infection and antimicrobial prescription data from diverse sources within electronic records has improved surveillance efficiency and scalability. Challenges with reliability, interfacility standardization, and the need for considerable human input to refine and ensure comprehensive data flow and adjudication of infection definitions continue to leave considerable room for improvement in electronic surveillance. Leveraging natural language processing models for health care–associated infection surveillance should further enhance existing systems. Surveillance data interpretation can be complex as hospitals apply health care–associated infection metrics for diverse purposes, including monitoring disease trends, detecting outbreaks, driving quality improvement, and meeting mandated reporting requirements to state and federal agencies, and as a vital part of value-based care measures.

Because of its role as the nation's mandatory reporting mechanism for health care–associated infections, the CDC's National Healthcare Safety Network benchmarks and tracks health care–associated infections and provides the measures and analytical tools that allow comparison across facilities. Infection rates are expressed as standardized infection ratios, calculated by dividing the number of observed infections by the number of predicted infections.

EPIDEMIOLOGIC BASIS AND GENERAL MEASURES FOR PREVENTION AND CONTROL

Patients in health care facilities are vulnerable to transmission of pathogens from other patients, visitors, staff, or the inanimate health care environment. Health care personnel may convey multidrug-resistant organisms between patients and the environment on their hands or on shared patient care equipment (e.g., stethoscopes or portable imaging machines). Spread can also occur from contaminated plumbing, person to person via a respiratory route (e.g., influenza or group A *Streptococcus* from health care personnel), or via contamination of food, water, or medications.

Infections that arise within the first 48–72 h after admission are generally considered community acquired or, if the patient was transferred from another facility, attributed to the transferring institution. Those occurring henceforth are considered health care associated, or nosocomial. Patient factors that increase vulnerability to nosocomial infection include presence of an invasive device, immune deficiency (congenital or acquired), renal insufficiency, diabetes, and other major comorbidities.

When patients' own microbiota have been altered by exposure to antibiotics or other medications, they become more susceptible to colonization with multidrug-resistant organisms. Colonized patients serve as unintentional reservoirs for transmission of resistant organisms to other patients or to the hospital environment, including shared equipment. Patients can develop nosocomial infections from these transmitted pathogens (cross-transmission) or from their own microbiota.

Hand hygiene and environmental cleaning are essential interventions to interrupt transmission. Hand hygiene can be performed either with soap and water or alcohol-based hand gel, with special requirements for care of patients who have diarrhea (see "Health Care–Associated Diarrhea," below). Environmental cleaning and disinfection are focused on frequently touched surfaces, such as doorknobs, bed rails, and objects in the bathroom, and are similarly adjusted for disinfectant-resistant pathogens when necessary.

■ COLONIZATION

Nosocomial bacteria or yeast that are transmitted to a patient may be unintentionally ingested, leading to colonization with the organism, a carrier state. Patients whose microbiota have been altered by antibiotics are more susceptible to colonization with nosocomial organisms. Colonization pressure, the proportion of patients in a ward who are colonized, is a risk factor for spread of resistant organisms. Patients who are immunocompromised or whose immune barriers have been breached by surgery or invasive devices, are at increased risk of infection from the nosocomial organisms with which they are colonized.

■ OUTBREAK INVESTIGATION AND RESPONSE

Transmission in a health care facility may be detected via careful surveillance or by notification from a clinician or laboratory. Investigation typically involves molecular typing of bacterial or fungal (less commonly viral) isolates to ascertain whether they are, in fact, closely related and represent an outbreak. The gold standard for these comparisons is whole genome sequencing, which has high resolution for determining clonality and, in the case of bacteria, conveys valuable information about plasmid-carried resistance genes.

Once transmission has been established, outbreak investigation involves a series of steps including drafting a case definition, finding cases, reviewing medical records, performing descriptive epidemiology, and developing a hypothesis regarding the source of the outbreak. Control measures are implemented while surveillance (including environmental cultures in some cases) is conducted. Communicating with patients, staff, facility leaders, and public health is integral to the outbreak response.

NOSOCOMIAL AND DEVICE-RELATED INFECTIONS

Invasive procedures and indwelling devices deliver supportive care and can save lives, but they provide potential portals of entry for pathogens. Infections in patients who undergo surgery or have indwelling devices are important targets for prevention in health care facilities.

Postsurgical infections occur more frequently when surgery is performed in the setting of emergencies, tissue damage from trauma or radiation, uncontrolled infection elsewhere in the body, malnutrition, chemotherapy, and other conditions that impair wound healing.

Routine perioperative checklists enumerate activities that reduce the risk of surgical site infections **(Table 147-1)**.

After years of decline due to prevention efforts, rates of device-related and other nosocomial infections rose during the COVID-19 pandemic. These infections are a substantial cause of morbidity and mortality for hospitalized patients and are often due to antimicrobial-resistant infections.

■ CATHETER-ASSOCIATED URINARY TRACT INFECTIONS

Catheter-associated urinary tract infections occur more frequently in patients with critical illness, older age, and female sex. Sterile and atraumatic insertion of catheters and preservation of a closed drainage system help prevent contamination. Infections are preventable largely by concerted efforts to avoid the placement of indwelling urinary catheters and by reevaluating daily the duration of their use.

The presence of asymptomatic bacteriuria, or bacteria without symptoms of urinary tract infection, often leads to improper antimicrobial treatment in patients with indwelling urinary catheters. Apart from pregnant patients, most others need not undergo urine culture or treatment for asymptomatic bacteriuria. Diagnostic stewardship efforts include guidelines to reduce collection of urinalysis and culture from patients in whom these studies are not indicated to avoid subsequent inappropriate treatment of asymptomatic bacteriuria. In the elderly, this often occurs in the setting of nonspecific symptoms, such as confusion, that are not related to asymptomatic bacteriuria. Improper treatment of asymptomatic bacteriuria is often followed by improper testing for "clearance" of bacteriuria, which results in a vicious cycle of escalating antimicrobial use and resistance.

■ HEALTH CARE–ASSOCIATED PNEUMONIA

Hospital-acquired pneumonia is often due to aspiration of oral or gastric contents or, less commonly, hematogenous transmission from a remote source. Because of changes in patients' microbiota in medical facilities, bacterial causes of nosocomial pneumonia are frequently antimicrobial resistant. The most likely pathogens involved in nosocomial pneumonia and its treatment are discussed in **Chaps. 131 and 147**.

Ventilator-associated pneumonia is a subset of nosocomial pneumonia that occurs in up to 10% of ventilated patients. Mechanical ventilation can be complicated by a variety of infectious and noninfectious problems that are encompassed under the term *ventilator-associated* *events*. Ventilator-associated pneumonia can prolong the duration of ventilation and intensive care unit (ICU) stay or even prove fatal. Like other device-related infections, ventilator-associated pneumonia is preventable by limiting use of the invasive device (Table 147-1).

TABLE 147-1 Evidence-Based Bundled Measures to Reduce the Risk of Select Device- and Procedure-Related Infections

Prevention of Surgical Site Infections
Treat active infections prior to surgery
Administer prophylactic antibiotics within 1 h before surgery and discontinue immediately after surgery
Wear surgical attire that covers hair, mouth, and nose
Perform hand hygiene with an antiseptic agent
Perform antisepsis of the surgical site with chlorhexidine/alcohol solutions whenever possible
If hair must be removed, clip rather than shave (to avoid skin microtrauma)
Implement operating room asepsis, minimizing movement into and out of the room
Maintain operating room air at positive pressure, with moderate humidity (20–60%) and at least 20 air changes per hour
Prevention of Ventilator-Associated Pneumonia
Attempt to avoid intubation using high-flow nasal oxygen or noninvasive positive-pressure ventilation
Elevate head of bed to 30–45 degrees to reduce risk from gastric contents
Brush teeth daily
Use aseptic care of all respiratory care equipment
Avoid use of tap water for rinsing any respiratory care equipment
Follow strategies to promote earlier extubation: • Minimize sedation • Provide early exercise and mobilization • Provide early enteral feeding • Assess readiness for extubation on a daily basis
Prevention of Catheter-Associated Urinary Tract Infection
Place indwelling catheters only when strictly necessary, e.g., to relieve obstruction, and not for convenience
Use aseptic equipment and technique for catheter insertion and urinary tract instrumentation
Minimize urinary tract instrumentation
Minimize manipulation of or entry into urinary catheter systems, and avoid catheter irrigation
Reevaluate daily the need for continued use of an indwelling urinary catheter in each patient
Use alternative methods to avoid indwelling catheters, e.g., bladder scans, condom catheters, intermittent catheterization
Prevention of Catheter-Associated Central Line Infections
Optimize nurse-to-patient ratio
Consider alternatives to central line (e.g., peripheral IV, midline catheter)
Catheter insertion • Use a checklist to ensure adherence to insertion bundle • Use hand hygiene • Use maximum sterile barrier precautions and aseptic technique • Use chlorhexidine-alcohol antisepsis to prepare the site • Use ultrasound guidance • Use an all-inclusive catheter insertion kit • The subclavian vein is least prone and the femoral vein most prone to infection in the intensive care unit (ICU) setting
Catheter maintenance • Administer daily chlorhexidine gluconate baths to ICU patients with central lines • Scrub the hub with alcohol • Cleanse the catheter site with chlorhexidine-based antiseptic with every dressing change, at a minimum every 7 days (earlier if site is soiled or dressing disrupted) • Use disinfectant caps • Apply chlorhexidine-impregnated dressings • Reevaluate daily the need for continued use of each central line in each patient

CATHETER-ASSOCIATED BLOODSTREAM INFECTIONS

Central venous catheters represent another major target for infection prevention. Although they are often necessary for patient care, central lines are foreign objects that lie in direct contact with the bloodstream. Without meticulous care, each health care personnel interaction with the catheter serves as a potential contamination event. Catheter-associated bloodstream infections triple a patient's risk of in-hospital death. Similar to other device-related infections, the rate of catheter-related bloodstream infections had declined in the years leading up to 2020, when infection control gains were lost with the onset of the pandemic.

Preventing catheter-associated bloodstream infections begins with aseptic insertion technique using a bundle, or checklist, of evidence-based steps (Table 147-1). Catheter maintenance practices are important for ongoing prevention of infection and include meticulous, aseptic technique in handling the catheter. Daily full-body cleansing with 2% chlorhexidine gluconate is effective for preventing bloodstream infections, particularly central line infections, in the ICU and are standard of care. Although studies outside the ICU show mixed results, some hospitals utilize the baths for all patients with central lines because of the treatment's low side effect profile and the potential impact of the intervention.

TRANSMISSION-BASED PRECAUTIONS

Standard precautions are the basic set of practices that health care personnel should follow to minimize exposure to potentially infectious material and prevent pathogen transmission while caring for all patients in all health care settings. The key elements of standard precautions are applied based on the situational risk and include hand hygiene; respiratory etiquette; personal protective equipment use depending on the potential for exposure to blood, body fluid, or infectious material; and safe injection practices.

Additional pathogen-specific prevention measures are called transmission-based precautions. When patients harbor a suspected or confirmed communicable disease, transmission-based precautions are used to prevent spread in the health care setting. The fundamental modes of pathogen transmission inform the recommended prevention methods categorized by the World Health Organization (WHO) and CDC into contact, droplet, and airborne precautions.

Contact precautions utilize hand hygiene and barrier methods (gown and gloves) to prevent the spread of ubiquitous pathogens in a colonized or infected person's immediate environment, for example, *C. difficile* and multidrug-resistant organisms. Institutions differ in use of contact precautions for ubiquitous resistant organisms such as methicillin-resistant *Staphylococcus aureus*.

Droplet precautions are implemented when a patient has respiratory symptoms or a confirmed respiratory infection with a pathogen that is transmitted from person to person. Protective equipment includes a gown, gloves, eye protection, and a mask or respirator, given recent evidence that respiratory viruses are spread by both droplets (respiratory particles ≥ 100 microns in size) and aerosols (sprays of particles <100 microns). The principal mode of respiratory pathogen spread is by inhalation of respiratory particles emitted by an infected person during sneezing, coughing, talking, and breathing. Host factors such as infection stage, symptoms, immune status, and environmental conditions, including ventilation and humidity, influence pathogen transmissibility. Respiratory droplets remain suspended in the air only briefly and travel short distances.

Airborne transmission occurs via tiny particles 100 that can travel larger distances and remain suspended in the air for an extended duration. Measles, tuberculosis, varicella, and COVID-19 are infectious agents that pose a major risk of airborne transmission. *Airborne precautions* require respirators and placement of the infected individual in a negative-pressure airborne isolation room with a minimum air flow rate of 12 air changes per hour with direct exhaust outside the building or recirculation through a high-efficiency particulate air (HEPA) filter and respirator use.

Some conditions may merit combinations of these precautions (e.g., patients with varicella and SARS-CoV-2 are placed in contact and airborne precautions).

EMERGING INFECTIOUS DISEASES

■ EPIDEMICS, EMERGING INFECTIOUS DISEASES, AND AGENTS OF BIOTERRORISM

Preparedness for newly evolved and reemerging human pathogens is critical to infection prevention and control activities. The 21st century has witnessed the emergence of several infectious threats and two major pandemics: 2009 H1N1 swine influenza and 2019 SARS-CoV-2, with their vast worldwide impact and, in the case of SARS-CoV-2, millions of deaths. The changing of global ecosystems from deforestation, climate change, and extensive urbanization with a rising world population has increased the proximity between humans and wildlife. The confluence of these factors can generate conditions that favor mutational viral evolution and, ultimately, cross-species spillover of otherwise geographically confined zoonotic illnesses, as was seen with the outbreaks of Ebola in 2014, Zika in 2016, SARS-CoV-2, and the 2022–2023 international monkeypox (mpox) epidemic.

Another emerging challenge relates to socio-behavioral perceptions and beliefs linked to vaccines; the perpetuation of misinformation through social media around childhood and adult immunizations has amplified vaccine hesitancy, threatening the maintenance of safe vaccination rates and herd immunity. The most urgent threat is from highly contagious illnesses such as measles, mumps, polio, and varicella. The current global rise in measles cases has disproportionately affected unvaccinated or partially vaccinated individuals. Therefore, addressing vaccine access, misinformation, and declining immunization rates is critical to avert outbreaks of vaccine-preventable illnesses with a high public health impact.

Finally, health care personnel awareness of intentional human threats from potential agents of bioterrorism such as *Bacillus anthracis* and variola virus is essential. Natural smallpox is no longer a hazard; routine vaccination against smallpox ended in the United States in 1971. However, there is a possibility of laboratory-maintained variola outside of the WHO-approved repositories. In the event of an accidental or intentional release, vaccination is the primary strategy to stop smallpox spread. Two currently licensed vaccines for smallpox prevention are the replication-competent vaccinia virus vaccine, ACAM2000, and the live, nonreplicating modified vaccinia Ankara vaccine. Both vaccines are also recommended for prevention of mpox in high-risk groups. *B. anthracis* is a category A pathogen with a high case fatality rate and potential for mass casualty if aerosolized. After the U.S. inhalational anthrax cases in 2001 from mailed letters containing anthrax spores that resulted in five deaths, protocols for a mass disaster related to *B. anthracis* exposure have been a biodefense preparedness priority.

■ VIRAL RESPIRATORY INFECTIONS WITH PANDEMIC POTENTIAL

Safety protocols for patients and health care personnel from new and reemerging contagious respiratory illnesses including novel influenza A viruses, Middle East respiratory syndrome coronavirus (MERS-CoV), and other novel coronaviruses are vital. Adopting an integrated all-hazard screening approach to identify and implement infection control measures, initiate diagnostic testing, and safely deliver care is critical. Lessons from the COVID-19 pandemic underscore the importance of maintaining par levels of personal protective equipment, testing supplies, and essential therapeutics. Regular training of frontline health care personnel on appropriate donning and doffing of personal protective equipment ensures readiness to handle patients with contagious illnesses safely. Testing for novel pathogens may only be available through public health laboratories; however, developing the capability for laboratory-developed tests and existing infrastructures to rapidly adopt testing protocols ensures timely diagnosis, contact investigations, and postexposure management, which is critical for effective containment. Health care facilities should leverage a structured information-sharing approach through hospital incident command system activation. From a public health standpoint, robust surveillance mechanisms, clear and concise guidelines for clinicians and the public, and alignment among federal agencies, state and local public health, and commercial laboratories are critical to ensure testing capacity and vaccine and therapeutics access.

■ HIGHLY PATHOGENIC AVIAN INFLUENZA

Among the potential pandemic pathogens, avian influenza A (H5 and H7 strains) has an exceptionally high impact due to its constant viral evolution and potential for adaptation to transmit effectively in humans. The most immediate pandemic threat is posed by highly pathogenic H5N1 avian influenza, with ~900 known human infections worldwide since 2003 and a case fatality rate of ~50%. Sustained person-to-person transmission of this group of viruses has not occurred thus far. H5N1 began circulating in North American wild birds and poultry in late 2021, with subsequent detections in mammals, including U.S. dairy cattle herds and barn cats, in multiple U.S. states and numerous cases of animal-to-human transmission.

The recommended prevention measures against novel influenza in hospitals are similar to those against other high-consequence pathogens, including contact and airborne precautions and eye protection. Close contacts should receive antiviral prophylaxis with oseltamivir. Regarding diagnostics, influenza targets on commercial multiplex reverse transcriptase polymerase chain reaction (PCR) panels are insufficient for strain identification and additional, dedicated testing in public health laboratories from nasopharyngeal swabs, washes, and conjunctival swabs is required to establish the diagnosis in the appropriate clinical and epidemiologic context.

■ MPOX

Previously a geographically constrained virus endemic to West Africa mpox Clade II virus caused s global outbreak in 2022. On its heels, in central Africa, cases of Clade I, endemic to that region, soared. The virus is zoonotic, and spreads from person to person through close contact of mucosal surfaces or nonintact skin with open lesions or via contaminated inanimate objects such as bed linens. The role of respiratory transmission without close physical contact is less well established. Despite initial concerns, health care–associated transmission of mpox in nonendemic countries is rare, with multiple reported occupational transmissions after direct inoculation from needlestick injuries. Postexposure prophylaxis with the JYNNEOS or ACAM2000 vaccine series is recommended for high-risk exposures.

■ VIRAL HEMORRHAGIC FEVER PREPAREDNESS

During the 2014–2015 West African Ebola epidemic, symptomatic returning travelers presented to U.S. hospitals, resulting in occupational Ebola infections in two nurses at one facility. These and other events led to formation of a national network of Regional Emerging Special Pathogen Treatment Centers at large hospitals that have built specially engineered containment wards and trained staff to care for patients with suspected or confirmed Ebola, Marburg, or other viral hemorrhagic fever diseases. A series of additional hospitals are prepared to receive and provide temporary care for such patients prior to transferring them to these treatment centers. Finally, all medical facilities and emergency services should have plans for managing patients with highly contagious diseases who may present unexpectedly.

HEALTH CARE–ASSOCIATED DIARRHEA

■ *CLOSTRIDIOIDES DIFFICILE* INFECTION

Diarrhea that begins in health care facilities or is attributable to recent care in a health care facility is considered health care associated. *C. difficile* infection **(Chap. 139)** is not only the most frequent cause but also the most frequent hospital-acquired infection in the United States. *C. difficile* colitis has symptoms that can range in severity from mild diarrhea to life-threatening toxic megacolon. Receipt of antibiotics is the most common cause of *C. difficile* infection, making antimicrobial stewardship the first-line measure for its prevention. Restricting use of antibiotics that are highly associated with subsequent *C. difficile* infection, such as quinolones and clindamycin, has been shown to reduce the incidence of the disease. Longitudinal genomic surveillance shows that patients who enter the hospital already colonized with *C. difficile* are by far the most likely to develop the infection during their hospital stay.

Patients who have suspected or confirmed *C. difficile* infection are placed in contact isolation to contain spores that infected patients shed in high numbers and that heavily contaminate their skin and environment. These spores are hardy, survive for a prolonged time in the health care environment, and can be transmitted on hands of health care personnel, on surfaces, or on shared patient care equipment. Because alcohol-based hand gel and standard hospital cleaners do not kill *C. difficile* spores, handwashing with soap and water is preferred to mechanically remove them, and sporicidal disinfectants such as bleach are used to clean the rooms of infected patients.

■ NOROVIRUS

Norovirus is another important cause of health care–associated diarrhea that can be transmitted among patients and staff and cause outbreak in facilities. A nonenveloped virus that is poorly inactivated by alcohol-based hand gel and standard hospital cleaners, norovirus is also addressed with handwashing with soap and water and bleach cleaning.

Another potential cause of health care–associated diarrhea is a foodborne outbreak, which can start with symptomatic food handlers or point-source food contamination that originates outside the facility.

TUBERCULOSIS

Tuberculosis **(Chap. 183)** requires special infection prevention and occupational health measures. Early identification, isolation, and testing of patients who may have active tuberculosis are critical steps in tuberculosis control in health care facilities. Patients with suspected or confirmed tuberculosis should be managed with airborne precautions (see “Transmission-Based Precautions”). Health care personnel managing patients with infectious tuberculosis use fit-tested particulate respirators, known in the United States and Canada as N95 respirators because they filter 95% of airborne particles, or powered air-purifying respirators that draw air through a HEPA filter into a hood.

Health care personnel should undergo preemployment screening for tuberculosis, preferably with an interferon-γ release assay rather than a tuberculin skin test due to its higher positive predictive value. Staff with positive tests should be evaluated further with symptom screening and chest imaging to assess for active or latent tuberculosis. Staff who have latent tuberculosis should be encouraged to undergo treatment to reduce the risk of reactivation and transmission to patients. In hospitals with high caseloads and local prevalence of tuberculosis, follow-up testing may be performed routinely among staff in certain disciplines. In most facilities, follow-up testing is event driven, such as following exposure to a patient with active tuberculosis.

FUNGAL INFECTIONS

■ MOLD INFECTIONS

Some immunocompromised hospitalized patients are highly susceptible to nosocomial mold infections, which are usually acquired through inhalation. Patients who will experience prolonged neutropenia during treatment of leukemia and stem cell transplant recipients require protective isolation rooms that are specially engineered with positive-pressure, HEPA-filtered airflow and sealed seams to repel mold spores, which, at 2–4 microns in size, are ubiquitous in dust and air currents. Vulnerable patients should wear masks when leaving their rooms to go to areas of the hospital that lack such measures. Exposure to hospital construction is a well-described risk factor for invasive fungal infections among immunosuppressed patients; thus, all construction sites within the hospital must employ negative airflow, HEPA filtration, sealed walls, and sticky doormats to minimize leakage of particles containing mold spores into patient care areas.

■ *CANDIDA* INFECTIONS

Candida infections represent the vast majority of health care–associated fungal infections. Many of these arise from endogenous sources in patients who are immunosuppressed, have undergone surgery, or have invasive devices in place. Although *Candida albicans* represent a majority of *Candida* infections, antifungal prophylaxis may select out non-albicans species such as *Candida glabrata.*

Candida auris Some *Candida* infections are attributed to exogenous transmission. *Candida auris* was first isolated in Japan in 2009 and subsequently emerged as a global nosocomial pathogen. Since the first U.S. case was identified in 2015, the incidence has increased steadily, with a sharp acceleration during the COVID-19 pandemic. *C. auris* has become an increasing cause of candidiasis in hospitals and in long-term care facilities where patients are mechanically ventilated. A multidrug-resistant species that is adapted to the health care environment, *C. auris* survives for extended periods on surfaces, and rapid recontamination from colonized persons makes sustained disinfection challenging.

Shared medical equipment such as glucometers, ultrasound machines, blood pressure cuffs, and axillary temperature probes are potential sources of transmission between patients, and thorough disinfection after each use is vital to prevent the spread of *C. auris*. Hand hygiene and contact precautions should be followed strictly when caring for a colonized or infected patient. The organism is resistant to some hospital disinfectant cleaners and requires use of select disinfectants. *C. auris* is of substantial public health concern, such that efforts to contain its spread require collaboration among infection control, public health, and laboratory experts on a facility, local, and regional basis. Patients transferred from long-term care facilities are often screened for *C. auris* by PCR or culture and isolated pending the results of screening **(Fig. 147-1)**.

THE HEALTH CARE BUILT ENVIRONMENT

Patients, staff, and visitors continually shed microorganisms into the built environment of a health care facility (surfaces and plumbing). Organisms are also introduced via the external environment, including potable water. Patients who are immunologically vulnerable may acquire pathogens from the built environment. Health care facilities must reduce their risk through adherence to current guidelines and through surveillance for such infections.

■ ENVIRONMENTAL CLEANING

Floors and frequently touched surfaces in health care facilities should be cleaned often with disinfectant cleaners in order to reduce the burden of communicable bacteria, viruses, and fungi. Patient care equipment (e.g., stethoscopes, portable x-ray cartridges) should be disinfected between patients to avoid becoming point sources for transmission of multidrug-resistant pathogens. Sporicidal cleaners and ultraviolet C light are often used as adjunctive methods for inactivating *C. difficile* spores and other tenacious pathogens. Clinical trial data suggest that standard cleaning plus adjunctive ultraviolet C may reduce nosocomial transmission of pathogens over cleaning alone.

■ WATER SAFETY

Waterborne infections arising from wastewater (drains and toilets) as well as potable water (sinks, showers, and fountains) pose a potential risk to vulnerable patients in health care facilities. Because many of these infections are preventable, hospitals are required to have a water management plan to anticipate and reduce risk to patients.

Tap water is not sterile, and contamination of potable hospital water and plumbing occurs most frequently when waterborne organisms (e.g., *Legionella pneumophila,* nontuberculous mycobacteria, and many others) colonize the biofilm within the pipes or cooling towers of a health care facility. Low disinfectant levels, warm temperature ranges, and stagnation of water promote organism growth, which can lead to transmission to patients via aerosols or droplets. Immunosuppressed patients are the most susceptible to waterborne infections; pneumonia in this patient population should be thoroughly investigated to include testing for waterborne infection.

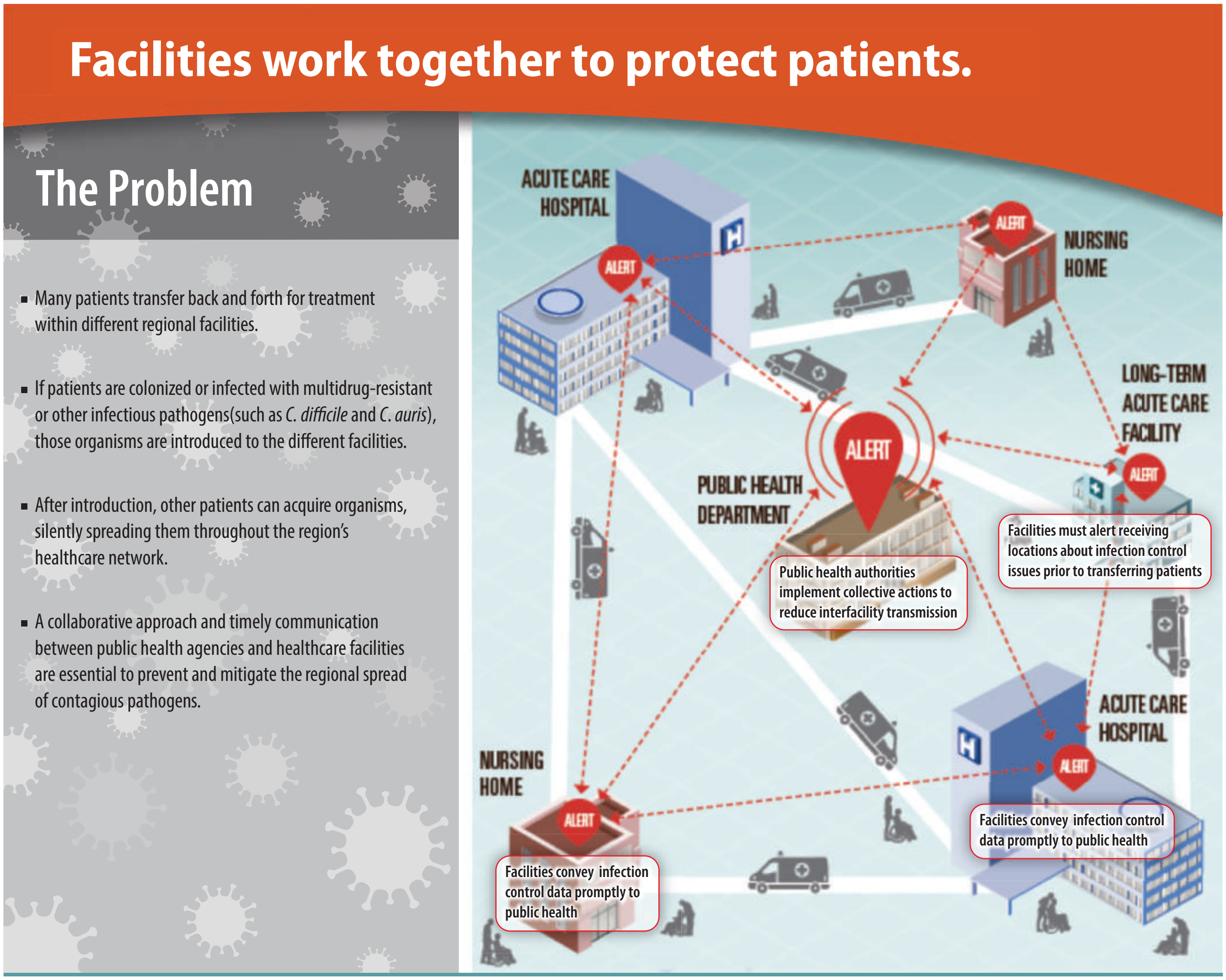

Facilities work together to protect patients.

The Problem

- Many patients transfer back and forth for treatment within different regional facilities.
- If patients are colonized or infected with multidrug-resistant or other infectious pathogens(such as *C. difficile* and *C. auris*), those organisms are introduced to the different facilities.
- After introduction, other patients can acquire organisms, silently spreading them throughout the region's healthcare network.
- A collaborative approach and timely communication between public health agencies and healthcare facilities are essential to prevent and mitigate the regional spread of contagious pathogens.

Public health authorities implement collective actions to reduce interfacility transmission

Facilities must alert receiving locations about infection control issues prior to transferring patients

Facilities convey infection control data promptly to public health

Facilities convey infection control data promptly to public health

FIGURE 147-1 Regional spread and control of antimicrobial resistance. *(Modified from https://archive.cdc.gov/#/details?url=https://www.cdc.gov/vitalsigns/stop-spread/infographic.html.)*

In the case of nosocomial Legionnaires disease, detection of a single case should prompt a thorough investigation. In facilities with no history of an outbreak, routine hospital water testing for *L. pneumophila* is not strictly necessary if robust clinical surveillance is in place. Institutions that have had nosocomial cases should conduct regular water testing for the organism as part of their water management plan.

Wastewater drains in health care facility sinks, tubs, and showers, as well as toilets, serve as silent reservoirs for multidrug-resistant gram-negative organisms and can cause protracted outbreaks. These bacteria reach patients via transmission in droplets or aerosols of contaminated water from splashback, plumbing fixtures, or toilet flushing. Maneuvers to reduce risk include installing lids on toilets and angling sink faucets so water does not land directly on the drain.

Contamination of medications and medical devices can also lead to nosocomial outbreaks. In the past decade, outbreaks of nontuberculous mycobacterial infections among cardiac surgery patients have been linked to operative heater-cooler units that aerosolized the organisms from contaminated water tanks.

ANTIBIOTIC-RESISTANT BACTERIA

Implementing measures to stop the spread of multidrug-resistant organisms is a key priority for public health agencies and health care epidemiologists. Globally, 1.27 million deaths were attributed to antibiotic-resistant infections in 2019, with the highest rates in sub-Saharan Africa. According to recent estimates, infections due to methicillin-resistant *S. aureus* (MRSA) and extended-spectrum β-lactamase (ESBL) infections are the most common multidrug-resistant organism infections in U.S. hospitals. While improved infection prevention measures have led to a substantial reduction in MRSA, vancomycin-resistant *Enterococcus* (VRE), carbapenem-resistant *Acinetobacter*, and multidrug-resistant *Pseudomonas* infections, the prevalence of carbapenem-resistant Enterobacterales and ESBL-producing Enterobacterales has not declined. Infections due to multidrug-resistant organisms, which frequently colonize the human gastrointestinal and respiratory tracts, are associated with higher morbidity and mortality and other adverse clinical consequences such as extended hospital stays, transmission risk, and added health care costs. Without meticulous environmental disinfection, many of these organisms can establish enduring reservoirs in the hospital environment, effectively transmitting between patients and posing the risk of horizontal gene transfer among co-colonizing strains. Coordination of care during interfacility transfer of known patients with MDRO colonization is paramount to ensure timely implementation of infection prevention and control measures (Fig. 147-1).

Regardless of pathogen, frequent health care–related exposure, extended hospital stays, residence in long-term care facilities, multiple comorbid conditions, short- or long-term indwelling devices, and antibiotic exposure are among the key risk factors for multidrug-resistant organism acquisition. Additionally, community determinants can contribute substantially to spread of certain organisms. For example, people who use drugs, prison inmates, and those without stable housing conditions are at a greater risk of MRSA infections.

METHICILLIN-RESISTANT *STAPHYLOCOCCUS AUREUS*

Among hospitalized patients, up to a third of MRSA-colonized individuals develop MRSA infection, and the risk in the pediatric ICU can be as high as 47%. MRSA is the most common health care–associated infection in neonatal ICUs. Transmission within facilities occurs through indirect contact with contaminated environments and sometimes from contaminated or colonized health care personnel. Screening, contact precautions, and education can reduce MRSA transmission, but most U.S. hospitals selectively deploy MRSA screening only in high-risk settings, including ICUs; dialysis, burn, and transplant units; and post–acute care facilities. Screening is also the fundamental approach to combat MRSA outbreaks.

Among the preventative practices assessed in ICUs, universal gowns and gloves and decolonization have reduced MRSA acquisition and infections. Universal decolonization is more effective than screening and contact precautions, with or without concurrent targeted decolonization with chlorhexidine bathing and 5-day nasal application of mupirocin. The advantages of decolonizing MRSA carriers extend beyond the hospital stay, and post–hospital discharge measures can reduce the short-term risk of MRSA infections in carriers by 30%. Still, this approach may not achieve sustained benefits over longer time periods.

Preventing postoperative and device-related MRSA infections is an essential surgical quality and safety goal for health care settings. A meta-analysis of 39 studies evaluated nasal decolonization and glycopeptide antibacterial prophylaxis before cardiac and orthopedic surgery, demonstrating a reduced risk of *S. aureus* infections with universal nasal decolonization and all gram-positive surgical site infections with a combined approach of decolonization and targeted antibacterial prophylaxis. Similarly, universal decolonization methods may be effective in non-ICU settings where there is high utilization of invasive devices to lower the risk of associated infections, including MRSA.

VANCOMYCIN-RESISTANT *ENTEROCOCCUS* (VRE)

Soon after VRE emerged in the U.S. in 1987, identifying carriers by screening for gastrointestinal carriage to institute contact precautions proved to be an influential early approach. This commonly applied method successfully controlled the regional spread of this pathogen while its prevalence in health care settings was still low. Initial animal studies and clinical observations established the critical role of antibiotic exposure to vancomycin, third-generation cephalosporins, and anti-anaerobic agents as risk drivers of VRE acquisition and overgrowth. Subsequently, microbiome analysis elucidated that a higher VRE colonization burden often precedes mucosal translocation in susceptible hosts, also influencing the number of bacteria shed in the environment that can facilitate person-to-person spread. Immunocompromised patients with hematologic malignancy, as well as stem cell and liver transplant recipients, are especially vulnerable to VRE colonization and infection. Over the years, VRE has become ubiquitous in many U.S. health care settings. Active surveillance and contact precautions offer minimal benefit in hyperendemic settings. However, the judicious use of antibiotics through robust stewardship measures and bundles incorporating chlorhexidine baths remain effective in reducing the impact of VRE by rendering colonized patients less infectious and decreasing the risk of invasive infection.

MULTIDRUG-RESISTANT GRAM-NEGATIVE BACTERIA

Multidrug-resistant gram-negative bacteria are increasingly common in health care settings. Chromosomally encoded resistance mechanisms and plasmid-carried resistance genes make them particularly challenging therapeutic targets. Multidrug resistance is defined as nonsusceptibility to at least one agent from three distinct antibacterial drug classes. Difficult-to-treat resistance is defined as resistance to all first-line antibacterial agents. Extensively drug-resistant organisms exhibit nonsusceptibility to two or fewer antimicrobials in all categories. Finally, pan-drug resistance is defined by the noneffectiveness of all antimicrobial agents. The rising incidence of drug-resistant gram-negative bacteria has been noted across all types of health care settings, carrying decreased survival compared with susceptible infections from the same species. Carbapenem-resistant Enterobacterales and *Acinetobacter baumanii*, multidrug-resistant *Pseudomonas aeruginosa*, and ESBL-producing *Escherichia coli* and *Klebsiella pneumoniae* are among the top threats.

Extended hospital stays, long-term care facility exposure, antimicrobials, medical devices, mechanical ventilation, and impaired immunity are frequent risk factors for acquisition of multidrug-resistant gram-negative bacteria. Outbreaks mainly occur in long-term care facilities, and hospital-based clusters tend to occur in intensive care, hematology-oncology, and liver transplant units. Inadequate disinfection of duodenoscopes due to their complex design has made them vulnerable to sterilization lapses, resulting in numerous outbreaks of multidrug-resistant gram-negative bacteria.

Aggressive control measures are among the cornerstones of prevention. Given an association between antimicrobial overuse and the emergence of multidrug-resistant gram-negative organisms, antimicrobial stewardship is a pivotal component in the strategies to prevent their emergence and spread. Targeted screening and contact precautions for rapid identification and isolation of gastrointestinal carriers have proven essential for outbreak management. Universal gown and glove use failed to reduce acquisition. Local epidemiology and international travel to hyper endemic areas, especially for medical care, are other important risk factors to consider when deciding on targeted screening practices. Finally, there are no proven decolonization strategies for multidrug-resistant gram-negative bacteria.

DIAGNOSTIC STEWARDSHIP

Diagnostic advances, including molecular technologies and high-throughput automated systems, have vastly enhanced clinical laboratory efficiencies and reduced the time to actionable information. However, rapid, convenient, and unrestricted testing can become problematic when used excessively for medically unnecessary investigations. From a health care–associated infection standpoint, inappropriate testing for *C. difficile* infection and excessive pursuit of urine cultures are pervasive problems in the current health care environment. Estimates show that as much as 40–60% of testing in hospitalized patients may not be medically indicated. Consequently, overtesting for these two common health care–associated infections leads to antibiotic overuse, patient distress,

TABLE 147-2 Diagnostic Stewardship Examples

	KEY FOCUS AREAS
Urine culture	• Educate clinicians on asymptomatic bacteriuria and appropriate indications to suspect urinary tract infection. • Require documentation of patient symptoms at the time of urine culture order placement. • Implement alerts to discourage testing when patient is asymptomatic, with exceptions (pregnancy, urologic procedures, etc.). • Consider algorithms that incorporate pyuria threshold (>10 white blood cells per high-power field) to proceed to culture in nonneutropenic patients. • Educate clinicians on appropriate methods for urine collection, storage, and transport for voided and catheterized urine. Avoid collection from the bag or a catheter that has been in place for an extended time.
Clostridioides difficile testing	• Set clinical criteria for testing: three or more unformed stools in 24 h in the absence of alternate etiology for the diarrhea. • Implement alerts that discourage an ordering clinician from testing when a patient is on laxatives. • Empower laboratory to reject specimens based on Bristol stool scale. • Apply auto-cancellation of test if sample is not received within 24 h or repeated within 7 days after a negative test or 14 days after a positive test. • Educate clinicians on the lack of utility of a test of cure. • Implement a two-step algorithm for testing including a toxin detection method.

TABLE 147-3 Health Care Personnel Immunizations

VACCINE	VACCINE TYPE	ELIGIBILITY	RECOMMENDATIONS
COVID-19	Non-live	All HCP	Per latest CDC recommendations
Influenza	Non-live	All HCP	Once annually
Hepatitis B	Non-live	HCP without serologic evidence of immunity or past infection	Two-dose series (Heplisav-B) or three-dose series (Engerix-B, PreHevbrio, or Recombivax HB)
MMR	Live	HCP without serologic evidence of immunity or past infection	Two doses at a 4-week interval
Varicella	Live	HCP without serologic evidence of immunity, clinician-verified history, or serologic evidence of past infection	Two doses at a 4-week interval
Tetanus-diphtheria-pertussis (TdaP)	Non-live	HCP without immunization in past 10 years Pregnant HCP	Single dose (booster every 10 years) With every pregnancy
Meningococcal	Non-live	Laboratory workers with potential exposure	Men ACWY (booster every 5 years) and MenB (booster at 1 year and every 2–3 years thereafter)

Abbreviations: CDC, Centers for Disease Control and Prevention; HCP, healthcare personnel; MMR, measles, mumps, rubella.

spurious inflation of publicly reported infection rates, and excess health care costs. Clinicians should use interventions implemented through the computerized clinical decision support systems. Lab measures such as reflex or multistep testing algorithms can address the potential harms of overtesting. Interventions based on the electronic health record yield better results than human interventions. Key strategies to reduce inappropriate testing are included in Table 147-2.

OCCUPATIONAL HEALTH

VACCINATION OF HEALTH CARE PERSONNEL

Health care personnel are at higher risk for acquiring certain vaccine-preventable illnesses and for transmitting infection to patients. The goals of vaccination are to provide personal safety, reduce the risk of occupational acquisition or transmission to preserve the workforce during periods of high community transmission, and protect vulnerable patients. Pathogens that pose a substantial risk of occupational transmission for which licensed vaccines are available include respiratory viruses such as influenza, COVID-19, measles and mumps, varicella, pertussis, hepatitis B, and meningococcus.

Vaccine eligibility is assessed at the time of employment, and the schedules for recommended immunizations are shown in Table 147-3. Additional risk-based vaccines may be recommended (e.g., Ebola, vaccinia, hepatitis A). Highly contagious illnesses such as measles and varicella have caused outbreaks in health care settings, disproportionately affecting underimmunized or unimmunized patients and health care personnel. Even with vaccines that have only moderate effectiveness against infection, higher vaccination uptake is associated with a lower risk of respiratory illness among personnel and a lower risk of nosocomial acquisition of these infections, especially in vulnerable patients. Numerous medical professional societies endorse mandatory health care personnel influenza immunization, and many facilities require it.

Health care personnel who provide direct care to patients should be tested for hepatitis B surface antibody 1–2 months after the last dose of the series and reimmunized if antibody levels are <10 mIU/mL. Nonresponders after repeat series are considered susceptible to hepatitis B and should be treated as such in the event of a bloodborne pathogen exposure.

BLOODBORNE PATHOGEN EXPOSURE

Since 1991, the U.S. Occupational Safety and Health Administration has promulgated a standard to regulate exposure to bloodborne pathogens. Advancements in engineering and workplace controls, such as needleless catheter systems, have made the use of medical devices safer, prevented sharps-related injuries, and reduced bloodborne pathogen transmission events. Despite this, exposure to potentially infectious agents in blood and other bodily fluids through contact with skin, eyes, and other mucous membranes remains a significant problem during health care delivery. HIV and hepatitis B and C acquisition after blood exposure are the most common concerning risks; semen, vaginal and rectal fluid, and breast milk can also pose a transmission risk. The risk of bloodborne pathogen transmission after contact with feces, urine, gastric and respiratory secretions, saliva, and sweat that is not contaminated with blood is exceedingly low. Significant exposures involve percutaneous injury or skin puncture with a much lower risk from contamination of nonintact skin and mucous membranes.

TABLE 147-4 Estimated Risk of Bloodborne Pathogen Transmission from Percutaneous Injury

PATHOGEN	RISK
Hepatitis B[a]	6–30%
Hepatitis C	1–3%
HIV[b]	0.3%

[a]Highest for hepatitis B e antigen positive. [b]If on effective treatment and with an undetectable viral load, the risk is negligible.

Among the three pathogens, hepatitis B transmits most effectively, with the estimated risks after percutaneous injury from needlesticks or other sharps shown in Table 147-4.

When managing personnel with a bloodborne pathogen exposure, a thorough history and risk evaluation and baseline serologic testing should be conducted. For significant potential hepatitis B exposures in nonimmune persons, hepatitis B immune globulin should ideally be given within 24 h and can effectively prevent transmission when administered up to 7 days from exposure. HIV postexposure prophylaxis must be started within 72 after possible exposure, and likely has diminishing efficacy the longer initiation is delayed. Serial testing for HIV, hepatitis B, and hepatitis C should be performed 6 weeks and 3 months after exposure, followed by testing for hepatitis C and HIV at 6 months if the exposed person is found to be immune to hepatitis B or received immune globulin.

With widespread vaccination, transmission of hepatitis B from health care personnel to patients is rare. Cases have occurred in the setting of occult hepatitis B with high-grade viremia and invasive procedures. Guidelines set out recommendations for management of clinicians infected with HIV, hepatitis B, and hepatitis C to minimize risks to patients. Unsafe medical device and medication management can spread infection between patients. Outbreaks and transmission events arise from improper disinfection of blood glucose monitors between patient uses, infection control breaches during dialysis, single-dose medication reuse, improper practices with multidose vials, and drug diversion. Such occurrences are entirely preventable with adherence to basic hygiene and patient safety practices.

FURTHER READING

Harris AD et al: Acquisition of antibiotic-resistant gram-negative bacteria in the benefits of universal glove and gown (BUGG) cluster randomized trial. Clin Infect Dis 72:431, 2021.

Huang SS et al: Decolonization to reduce postdischarge infection risk among MRSA carriers. N Engl J Med 380:638, 2019.

Lyman M et al: Worsening spread of *Candida auris* in the United States, 2019 to 2021. Ann Intern Med 176:489, 2023.

Miles-Jay A et al: Longitudinal genomic surveillance of carriage and transmission of *Clostridioides difficile* in an intensive care unit. Nat Med 29:2526, 2023.

O'Grady NP: Prevention of central line-associated bloodstream infections. N Engl J Med 389:1121, 2023.

Rankin DA et al: Concurrent transmission of multiple carbapenemases in a long-term acute-care hospital. Infect Control Hosp Epidemiol 45:292, 2024.

Singh HK et al: Diagnostic stewardship to improve patient outcomes and healthcare-associated infection (HAI) metrics. Infect Control Hosp Epidemiol 45:405, 2024.

148

Infections in Transplant Recipients

Jennifer M. Cuellar-Rodriguez,
Juan C. Gea-Banacloche

GENERAL CONCEPTS

This chapter discusses infections in recipients of solid organ transplant (SOT), vascular composite allografts (VCAs), and hematopoietic stem cell transplantation (HCT). Despite the notable differences between these procedures, they share the fact that recipients receive immunosuppression (in the case of SOT and VCAs, for life). They become immunocompromised and, consequently, are at increased risk for infection.

Infections in immunocompromised patients may be caused by the common pathogens that affect everybody but also by opportunistic microorganisms that usually do not cause significant illness in healthy people. The physician taking care of transplant recipients should be aware of the risk of unusual pathogens, uncommon presentation of infections, and multiple concomitant infections. In addition, response to treatment may be suboptimal due to the effect of immunosuppressive medication and corticosteroids as well as coexisting medical or surgical complications. Professional societies issue and regularly update guidelines for the diagnosis and management of infections in these patients. For SOT, the reader is referred to the guidelines from the American Society of Transplantation Infectious Diseases Community of Practice and, for HCT, to the guidelines published by the American Society for Transplantation and Cellular Therapy and the European Conference on Infections in Leukemia (ECIL).

In this textbook, there are chapters dedicated to SOT **(Chaps. 271, 309, 325, and 356)**, HCT **(Chap. 119)**, and cellular therapy **(Chap. 483)**, as well as chapters about every one of the pathogens that will be mentioned here. Our goal is to focus on the particulars of infections as they pertain to transplant, as well as to discuss some relevant topics that are not included elsewhere. These include pretransplant evaluation of transplant candidates, assessment of potential donors, infections acquired through the transplant, and prophylaxis of infections.

Understanding the similarities between transplants regarding increased risk of infection (e.g., immunosuppressive agents) is important, as is understanding some of the basic differences. An obvious and relevant difference is that SOT and VCA transplants include surgical procedures, and the operations are extremely challenging. Most early infections are related to complications of the surgery, and sometimes even late infections may have their origin in the operating room. In contrast, early infections after HCT are related to the neutropenia induced by the conditioning regimen. Another difference is that both SOT and VCA transplants require lifelong immunosuppression to prevent rejection. This means the risk of opportunistic infection never goes away. Conversely, immunosuppression is not typically administered after autologous HCT (auto-HCT), and it is only given for a few months after allogeneic HCT (allo-HCT). Finally, both SOT and allo-HCT carry a risk of unfavorable immune reactions: rejection of the transplanted organ and graft-versus-host disease (GHVD), respectively. Both rejection and GVHD are treated by increasing the immunosuppression, with consequent increased risk of infection.

INFECTIONS AFTER HEMATOPOIETIC STEM CELL TRANSPLANTATION

HCT is a procedure in which a conditioning regimen that eliminates (partially or completely) the recipient's myeloid and lymphoid immune systems is followed by the infusion of hematopoietic stem cells, with subsequent recovery of bone marrow and immune function. It includes two distinct categories: auto-HCT and allo-HCT. Of the approximately 20,000 HCTs performed in the United States every year, 12,000 are autologous and 8000 allogeneic. In both, a conditioning regimen (variable combinations of chemotherapy, radiation, and serotherapy) is administered. This conditioning regimen eliminates the myeloid progenitors and the lymphoid immune system of the patient. Then, hematopoietic stem cells are administered to replenish both the myeloid and the lymphoid immune systems. In the case of autologous transplantation, the stem cells are from the recipient, harvested in advance and cryopreserved until the day of the transplant. This procedure does not correct defects present in the stem cells (e.g., hemoglobinopathies, primary immunodeficiencies) and, in the treatment of cancer, is not expected to cure a malignancy that is refractory to chemotherapy because the autologous stem cells have no anticancer potential. For malignancies still susceptible to chemotherapy, however, the intensive myeloablative conditioning regimen could be curative, while the infusion of autologous cells would effectively "rescue" the bone marrow. Conversely, in allo-HCT, the donor is another person, and their stem cells are free of whatever genetic defect the recipient's stem cells had. This means allo-HCT could, in principle, cure stem cell disorders. The reconstituted myeloid system, for instance, will be free of hemoglobinopathy. The reconstituted immune system should ultimately become immunocompetent. If a malignancy is being treated, there is potential for the new immune system to fight it and even eradicate it, through a now well-defined phenomenon called *graft-versus-leukemia* (GVL) or *graft-versus-tumor* (GVT) effect. From this standpoint, allo-HCT may be considered a form of immunotherapy, and it has been used to cure chemotherapy-refractory malignancies. There is also, unfortunately, the potential for the regenerated immune system to recognize the recipient as "alien" and to mount an immune response against their organs, which is called *graft-versus-host disease* (GVHD) and can be severe and even fatal. GVHD is common, affecting between 35 and 50% of allo-HCT recipients with varying severity (from mild to life-threatening). To prevent GVHD, allo-HCT recipients receive a few months of immunosuppression, which is then slowly tapered. GVHD is not expected to occur after auto-HCT, so no immunosuppression is administered after the stem cell infusion.

The biological difference between auto-HCT and allo-HCT dictates the different indications of each one. The vast majority of auto-HCTs performed are for multiple myeloma and lymphoma, malignancies that do not involve the stem cells. In this situation, the therapeutic intervention is the conditioning regimen, and the function of the infusion of stem cells (the transplant) is to replenish the myeloid progenitors eliminated by the conditioning. A minute fraction of auto-HCT is used or is being studied in select nonmalignant conditions that could potentially benefit from intense immunoablation. Auto-HCT was superior to standard immunosuppression for the treatment of severe systemic sclerosis in a randomized controlled trial and is also being used in selected patients with multiple sclerosis. It is also being investigated in systemic lupus erythematosus, rheumatoid arthritis, polymyositis/dermatomyositis, and Crohn disease.

The most common indication for allo-HCT is acute leukemia, followed by myelodysplastic and myeloproliferative disorders. It is also increasingly used for nonmalignant conditions including aplastic anemia, hemoglobinopathies (e.g., sickle cell disease, thalassemia), and primary immunodeficiency disorders (PIDs; also called inborn errors of immunity [IEI]) including severe combined immunodeficiencies (SCID), phagocytic disorders such as chronic granulomatous disease, diseases of immune dysregulation such as familial hemophagocytic lymphohistiocytosis, and others.

Different conditioning regimens result in varying degrees of organ toxicity, including mucositis, and variable durations of neutropenia. In the case of allo-HCT, additional factors that impact the risk of infection include the source of stem cells (bone marrow, peripherally collected stem cells, or umbilical cord blood [UCB]) and the degree of human leukocyte antigen (HLA) matching and relatedness between the donor and the recipient (in increasing infection risk: matched sibling donor, matched unrelated donor, haploidentical donor, mismatched donor).

Infections are common after transplant and account for between 15 and 25% of the deaths in the first 100 days after auto-HCT and between 20 and 40% after allo-HCT, depending on the type of transplant. The risk factors for infection include preexisting conditions (intimately associated with the indication for transplant), conditioning regimen with its attendant neutropenia, mucositis and immunosuppression, and (in the case of allo-HCT) immunosuppression to prevent GVHD and, if this occurs, GVHD and its treatment. As mentioned earlier, however, and in distinction to SOT, most allo-HCT recipients do not receive lifelong immunosuppression. A few years after allo-HCT, a patient without chronic GVHD has an essentially normal immune system.

TIMELINE OF INFECTIONS AFTER HCT

Considerable experience over decades of transplantation has resulted in learning what the expected infections are and when they are supposed to occur. Anti-infective prophylaxis derives from this knowledge. Nomenclature varies, but we will consider very early infections (from initiation of the conditioning regimen until resolution of neutropenia), early infections (from neutrophil engraftment until day +100), and late infections (after day +100) **(Table 148-1)**.

Very Early Infections: Preexisting Infections, Baseline Condition, and Neutropenia We consider the "very early" phase of HCT to be the period between the initiation of the conditioning regimen and recovery from neutropenia (neutrophil engraftment). The intensity of the conditioning (which is classified as myeloablative, reduced intensity, and nonmyeloablative) results in different degrees and durations of neutropenia, mucositis, and immunosuppression.

During this time, bacterial infections related to neutropenia and mucosal barrier injury predominate, but one must also consider preexisting infections, both common and frequently known in patients undergoing transplant for PID (e.g., known diarrhea and cholangiopathy caused by cryptosporidiosis in patients with SCID) and potentially unknown and acquired during therapies received prior to transplant (e.g., pulmonary aspergillosis developed during induction therapy of acute myeloid leukemia but asymptomatic pulmonary nodules at the time of transplant). A pretransplant infectious diseases consult is recommended, when possible, as a review of previous records and imaging may be extremely useful for prevention, diagnosis, and management of infection. An etiologic diagnosis for abnormal findings should be pursued. A review of past infection or colonization with resistant pathogens may also provide actionable information, for instance, determining the choice of empirical antibiotics during neutropenic fever.

Fever during neutropenia (neutropenic fever) occurs in a large proportion of HCT recipients, depending on the conditioning regimen. Neutropenic fever is discussed in depth in **Chap. 79**, but a summary follows. Fever during neutropenia is considered infectious in origin even if no infection is identified in most cases. The management centers on the fact that bacterial infections may progress very quickly in the absence of neutrophils, so expediency is critical. A thorough physical exam should be performed, with emphasis on possible bacterial portals of entry (e.g., mouth, perianal region, catheter exit site); blood cultures should be obtained; and broad-spectrum antibiotics including activity against *Pseudomonas aeruginosa* should be initiated. Antibiotic choice varies with prior history (e.g., previous infection with multidrug-resistant [MDR] gram-negative bacteria or methicillin-resistant *Staphylococcus aureus*), the clinical presentation (e.g., localizing symptoms or signs, hemodynamic instability), and the local patterns of resistance. Once antibiotics are started, they are modified depending on new microbiologic data or changes in the clinical status. In general, persistent fever alone is not an indication to modify the antibacterial regimen. However, persistent fever while receiving broad-spectrum antibiotics identifies a subset of patients with higher likelihood of harboring invasive fungal disease (IFD), and a thorough search for IFD or modification of the antifungal regimen is indicated when neutropenic fever continues for >4 days.

Bacterial infections during this time are related to the degree and duration of neutropenia and the extent of the mucositis. Consequently, milder, nonmyeloablative conditioning regimens may result in less very early infections. Antibacterial prophylaxis (typically with a fluoroquinolone such as levofloxacin) decreases the frequency of neutropenic fever and is used by some groups, but there is growing concern about the long-term consequences of this practice (potential for worsening transplant outcome due to microbiome disruption), as well as colonization with resistant bacteria and *Clostridioides difficile* colitis. Currently, antibacterial prophylaxis during HCT is not universal.

Neutropenia during HCT is usually shorter than after induction of remission of acute leukemia. The most common early fungal infection after HCT used to be candidiasis, but now antifungal prophylaxis is almost universal. Prophylaxis with fluconazole decreases the likelihood of invasive candidiasis and improves survival in the allo-HCT setting. Fluconazole prophylaxis is most often used, but if the likelihood of aspergillosis is considered high, an agent with activity against mold, typically posaconazole, is used instead. *Aspergillus* and other molds are uncommon during short neutropenia, but there may be other factors that put the patient at risk. It should be noted that antimicrobial prophylaxis is

TABLE 148-1 Common Sources of Infection After Hematopoietic Stem Cell Transplantation

	PERIOD AFTER TRANSPLANTATION		
INFECTION SITE	**VERY EARLY (<1 MONTH)**	**EARLY (BEFORE 100 DAYS)**	**LATE (>100 DAYS)**
Disseminated	Aerobic bacteria (gram-negative, gram-positive)	*Candida, Aspergillus,* CMV, EBV, *Toxoplasma*	Encapsulated bacteria (*Streptococcus pneumoniae, Haemophilus influenzae, Neisseria meningitidis*)
Skin and mucous membranes	HSV, *Candida*	HSV, VZV	VZV, HPV (warts)
Lungs	Aerobic bacteria (gram-negative, gram-positive), *Aspergillus,* other molds	CMV, community-acquired respiratory viruses, *Pneumocystis, Toxoplasma,* molds, adenovirus	Community-acquired respiratory viruses, *Pneumocystis, Nocardia, S. pneumoniae, H. influenzae* NTM (cGVHD)
Gastrointestinal tract	*Clostridioides difficile*	CMV, adenovirus, norovirus	EBV, CMV, norovirus
Genitourinary tract	BK virus	BK virus, adenovirus	
Brain		HHV-6, *Toxoplasma*	*Toxoplasma,* JC virus (rare)
Bone marrow		CMV, *Toxoplasma*	CMV

Abbreviations: cGVHD, chronic graft-versus-host disease; CMV, cytomegalovirus; EBV, Epstein-Barr virus; HHV-6, human herpesvirus type 6; HPV, human papillomavirus; HSV, herpes simplex virus; NTM, nontuberculous mycobacteria; VZV, varicella-zoster virus.

never completely effective. Breakthrough infections may occur, such as non-albicans *Candida* when fluconazole is used or *Fusarium*, *Mucorales*, or other azole-resistant molds when posaconazole is used.

Risk of early *Pneumocystis jirovecii* pneumonia (PJP) varies with the indication for transplant and prior treatments, as well as with the conditioning regimen. For instance, a patient receiving allo-HCT for sickle cell disease is not at risk for PJP before receiving the conditioning regimen, whereas patients with immunodeficiency or heavily treated for acute lymphoblastic leukemia carry considerable risk. In most cases, prophylaxis against PJP is used. It is well accepted that trimethoprim-sulfamethoxazole (TMP-SMX) is most effective, and it can be used until the infusion of stem cells. After the stem cells are given, the potential for myelotoxicity and delayed engraftment makes other options (e.g., inhaled or intravenous pentamidine) preferable. Atovaquone may be used, but it is expensive, unpalatable, and requires a fatty meal for optimal absorption, which may be difficult during this phase of significant mucositis.

Herpes simplex virus (HSV) reactivation occurs early, associated with the mucositis caused by the conditioning regimen. Acyclovir prophylaxis is standard of care, and it effectively prevents HSV and varicella-zoster virus (VZV). Breakthrough HSV must be differentiated from mucositis caused by chemotherapy, and viral polymerase chain reaction (PCR) should be used. If HSV breaks through acyclovir prophylaxis and absorption is not an issue, acyclovir-resistant HSV should be considered. Decisions must be made clinically, since drug resistance can only be determined definitively by phenotypic testing, which takes weeks. Some instances of low-level resistance may be overcome by switching to high-dose (10 mg/kg per 8 h) IV acyclovir, but if this does not result in clinical improvement, foscarnet 60 mg/kg per 12 h or cidofovir 5 mg/kg weekly should be used.

Cytomegalovirus (CMV) infection is uncommon before engraftment and will be discussed in detail later, but in many centers, HCT recipients who are CMV seropositive start prophylaxis with the antiviral letermovir at the time of transplant. Of note, letermovir has no activity against HSV or VZV, so acyclovir should be administered concomitantly.

Early Infections: Between Neutrophil Engraftment and Day 100—Reactivation of Latent Infections and Complications of GVHD and Its Treatment During this phase, the patient has neutrophils but not a functioning lymphoid immune system. Potential infections include exogenous pathogens (e.g., community-acquired respiratory viruses), acquired when the patient is discharged from the hospital and reenters the community, as well as the reactivation of latent pathogens (e.g., herpesviruses, adenovirus, JC virus, and parasites such as *Toxoplasma gondii*). Despite neutrophil engraftment, the patient's lymphoid immune system was eliminated by the conditioning regimen, and the infused stem cells have not had time to create a functioning immune system. Other factors may impact the risk of infection. In the case of allo-HCT, there is usually ongoing immunosuppression with a calcineurin inhibitor (cyclosporine A or tacrolimus), sirolimus, or mycophenolate mofetil. The conditioning regimen may have included particularly immunosuppressive strategies such as serotherapy with antithymocyte globulin or alemtuzumab, which have a long half-life and may result in prolonged T-cell depletion that persists after engraftment. The source of the stem cells administered also plays a role: peripherally collected stem cells include more mature T cells than bone marrow, and these may offer some protection against latent pathogens. Ex vivo T-cell depletion (CD34 selection) is sometimes used and results in a product with even fewer mature T cells than bone marrow. UCB also lacks effective cellular immunity against infections latent in the recipient. Finally, if the patient develops acute GVHD, high-dose corticosteroids may be added to the immunosuppressive regimen, increasing the risk not only of latent infection reactivation but also of newly acquired bacterial and fungal infections. If GVHD does not respond to steroids, second-line agents may be added, compounding the immunocompromise of the patient.

BACTERIAL INFECTIONS Mucositis and neutropenia have resolved during this phase. Some patients still have a central venous catheter, with the attendant risk of catheter-related bloodstream infection. However, in the absence of GVHD, bacterial infections are relatively uncommon. Conversely, if acute GVHD of the bowel occurs, the patient is at risk of mucosal barrier injury–related bloodstream infection with bowel flora, particularly when high-dose corticosteroids are initiated. *C. difficile* infection is also common during this period and should be aggressively treated.

FUNGAL INFECTIONS Because neutropenia is no longer a factor, fungal infection during this time is usually associated with corticosteroid use, although other immunosuppressive agents used to treat acute GVHD may also increase the risk. Aspergillosis is the most common mold infection, and patients with GVHD on high-dose corticosteroids (i.e., >0.5 mg/kg of prednisone every other day) or on a combination of immunosuppressive agents frequently receive posaconazole prophylaxis, based on the results of a randomized controlled trial. Many authorities recommend monitoring serum posaconazole levels, particularly when absorption may be compromised by severe acute GVHD of the bowel. A level <700 ng/μL is considered subtherapeutic, but experts disagree on what levels are desirable for effective prophylaxis and/or treatment. The use of serologic markers of fungal infection (galactomannan and/or beta-D-glucan) is not recommended in this setting, as the sensitivity of the tests is markedly reduced when on mold-active azoles.

HERPESVIRUSES Herpesviruses may reactivate during this phase, particularly in the setting of T-cell depletion, UCB transplant, or GVHD and its treatment.

HSV and VZV are usually effectively prevented by acyclovir prophylaxis, but acyclovir-resistant HSV may occur (see the previous section). CMV disease may be prevented by prophylaxis or preemptive management and is discussed separately.

Human herpesvirus 6 (HHV-6), the cause of exanthem subitum (sixth disease) in children, is carried latently by most adults and reactivates frequently after HCT. Most of the time, reactivation is asymptomatic and goes undetected, since routine monitoring of HHV-6 is not recommended. However, it can sometimes cause encephalitis, which may be severe and result in significant morbidity and even death. It presents as limbic encephalitis, typically without a fever and initially with a clear sensorium but with short-term memory loss accompanied by tremor or other involuntary movements. It may progress to abnormal mental status, seizures, and coma. The diagnosis is made in the presence of signs and symptoms of encephalitis with detectable HHV-6 in the cerebrospinal fluid (CSF) and no other explanation for the disease. Brain magnetic resonance imaging (MRI) may be normal initially but later shows bilateral involvement of the amygdala on the temporal lobes. It is not clear whether treatment modifies the outcome, as HHV-6 tends to get under control in most patients without any intervention and affected patients frequently suffer significant long-term sequelae despite antiviral treatment. Guidelines recommend treatment with ganciclovir or foscarnet, although some experts recommend combining both antivirals. Other less well-established disease associations with HHV-6 include fever, GVHD, pneumonia, poor graft function, liver enzyme abnormalities, and thrombocytopenia.

Human herpesvirus 7 (HHV-7) is, like CMV and HHV-6, a beta-herpesvirus. It may be detected by PCR in the blood of HCT recipients, but it is not known whether it causes any disease or is a co-factor in other infections such as CMV and HHV-6.

BK VIRUS BK virus, a polyomavirus that infects 80–90% of people, is associated with hemorrhagic cystitis after allo-HCT and, very seldomly, nephropathy. The clinical presentation includes dysuria and bladder spasms and hematuria, usually macroscopic. High-level BK viruria is usually present, but this may also happen without symptoms. The increasing use of posttransplant cyclophosphamide as prophylaxis for GVHD may make diagnosis difficult, as cyclophosphamide can cause hemorrhagic cystitis. There is no known effective treatment for BK virus after HCT, but hemorrhagic cystitis, albeit debilitating, is almost always self-limited. In severe cases, intravesical clots may cause obstructive uropathy and require continuous bladder irrigations.

Rarely, BK virus may cause nephropathy after allo-HCT. This entity is discussed in detail in the section on infections after SOT.

RESPIRATORY VIRUSES Besides the reactivation of latent viral infections expected during this period, once the patient is discharged after transplant, they are exposed to community-acquired respiratory viruses, which have significant mortality when they progress from upper respiratory tract infection to pneumonia.

Many respiratory viruses can cause significant morbidity after HCT. SARS-CoV-2, influenza, respiratory syncytial virus (RSV), metapneumovirus, and parainfluenza 3 may cause severe disease more often, but multiple enteroviruses and non–severe acute respiratory syndrome (SARS) coronaviruses may occasionally result in pneumonia. Only a few of these infections are amenable to antiviral treatment. Diagnosis usually relies on molecular methods (PCR of nasopharyngeal swab or nasal wash), but rapid viral culture (shell vial) is still used in some centers.

A general principle of viral infections is that controlling them (especially long term) requires the host's specific immunity, to a degree that bacterial infections do not. In this regard, decreasing the immunosuppression, if possible, is a standard part of the management. Unfortunately, this may not be possible in allo-HCT, particularly if the viral infection occurs in the setting of GVHD. If there is no GVHD, some consideration should be given to decreasing the dose of immunosuppressive drugs such as calcineurin inhibitors or sirolimus, which are usually carefully monitored through blood levels. Specific antiviral treatment recommendations follow.

Coronavirus Disease 2019 (COVID-19) U.S. Food and Drug Administration (FDA)-approved antiviral treatments for SARS-CoV-2 include remdesivir and nirmatrelvir-ritonavir (Paxlovid). In nonvaccinated, never-infected, non-immunocompromised patients with comorbid conditions or older than 60 years with COVID-19, nirmatrelvir-ritonavir given within 3 days of the onset of symptoms resulted in a 90% decrease on the risk of hospitalization or death. More recent studies including people who have been vaccinated and are not at high risk of progression have not confirmed these results. However, many experts still recommend nirmatrelvir-ritonavir for immunocompromised patients. Unfortunately, transplant recipients are frequently receiving medications such as tacrolimus or sirolimus that may make nirmatrelvir-ritonavir use impractical due to drug–drug interactions. The antiviral molnupiravir has an FDA emergency use authorization for the treatment of adults with mild-to-moderate COVID-19 who are at high risk for progression to severe disease and for whom other COVID-19 treatment options are not accessible or clinically appropriate. Most experts would at least consider antiviral treatment of COVID-19 in HCT recipients. Depending on the clinical setting, outpatient administration of IV remdesivir may be a more practical option. Bacterial superinfection is uncommon after COVID-19, but aspergillosis has been described. Risk factors for COVID-19-associated pulmonary aspergillosis include high-dose corticosteroids and hematologic malignancy but not necessarily transplant.

Influenza Virus Influenza may be treated with neuraminidase inhibitors (oseltamivir, zanamivir, or peramivir) or baloxavir, an inhibitor of the polymerase acidic (PA) protein. There are limited data to choose between these two options, but more experience with the former. Amantadine and rimantadine, only active against influenza A, are not currently recommended in the United States because of widespread resistance.

Respiratory Syncytial Virus RSV may be treated off-label with ribavirin, either orally or by inhalation. In RSV, lymphopenia (absolute lymphocyte count [ALC]) has been found to be a risk factor for the development of pneumonia, and some experts recommend using ALC <1000/μL as a guide to initiate early treatment with ribavirin.

Adenovirus Adenovirus occupies a particular place within this group, as it can represent a newly acquired community-acquired infection or be reactivation of latent virus from prior infection. Because adenovirus may be more common in children, some pediatric centers monitor high-risk patients by PCR of stool and/or blood and attempt intervention before established adenovirus disease ensues. Drugs with activity against adenovirus include cidofovir, brincidofovir, and (for some species) ribavirin. The response of established adenovirus disease to antivirals is poor.

As mentioned above, all respiratory viral infections may occasionally cause lower respiratory tract disease. When there is lower respiratory tract with any respiratory virus, a search for co-pathogens (bacterial or fungal) is indicated, as these are frequently present.

TOXOPLASMOSIS *Toxoplasma gondii* infects >2.5 billion people. The prevalence and routine monitoring after HCT vary significantly between countries. Studies from Europe suggest that allo-HCT recipients who have positive *Toxoplasma* serology reactivate frequently (8–21%) and early (95% within the first 6 months). The presentation may be just fever (this, in the presence of a positive *Toxoplasma* PCR, is often called toxoplasmosis infection), but also may include pneumonia (small pulmonary nodules and interstitial pneumonia predominantly at the lung bases), heart disease, brain abscess (single, multiple, or diffuse encephalitis), and disseminated disease sometimes associated with hemophagocytic lymphohistiocytosis. Diagnosis is by blood PCR and/or immunohistochemistry. The standard treatment is with pyrimethamine and sulfadiazine, but TMP-SMX may be equally effective. Atovaquone is a second-line agent, and there are anecdotal reports in advanced HIV of successful use of the combination of azithromycin and clindamycin.

Late Infections (After Day 100) Day 100 post-HCT remains an arbitrary boundary, as this is when tapering of immunosuppression usually begins. Conceptually, immune reconstitution takes place unimpeded by immunosuppressants, so the risk of infection should become progressively less. Most infections at this time are caused by community-acquired pathogens, although late CMV may also occur. The main risk factor for infection is the presence of GVHD and its treatment. Chronic GVHD (cGVHD) results in functional asplenia, and *Streptococcus pneumoniae* is a particularly important pathogen. Late fatal infections are predominantly bacterial, but respiratory viral infections such as influenza are still more common than in age-matched populations. The group at highest risk are patients with cGVHD receiving immunosuppression.

INFECTIONS AFTER SOLID ORGAN TRANSPLANT

RISK FACTORS FOR INFECTION AFTER SOLID ORGAN TRANSPLANT

Infectious complications remain a major cause of morbidity and mortality after SOT. The risk of infection is determined by the interaction of various factors, such as age of the recipient; type of transplant and technical complexities associated with it; other invasive procedures; dose, duration, and temporal sequence of immunosuppressive drugs; epidemiologic exposures of both donor and recipient; use of antimicrobial prophylaxis; donor-recipient serostatus to certain infections (e.g., CMV, Epstein-Barr virus [EBV], toxoplasmosis); and ongoing viral replication (so-called indirect effects). Assessment of a recipient's risk for infection helps tailor specific preventive strategies and workup when an infection is suspected.

Induction immunosuppression is usually administered immediately prior to transplantation. This is followed by maintenance immunosuppression, which frequently consists of triple therapy with a calcineurin inhibitor (cyclosporine A or tacrolimus) or an mTOR inhibitor (sirolimus or everolimus), mycophenolate mofetil, and corticosteroids. Maintenance immunosuppression may be intensified during periods of rejection, and it may be decreased with time, but it is usually continued for life. The amount of immunosuppression the patient is receiving is important in terms of risk of infection.

Age is also an important determinant of susceptibility to infections; it impacts the likelihood of prior exposures to microbial pathogens, either by primary infection or vaccination. History of exposure can have either positive or negative effects. Older patients are more likely to have encountered pathogens that can remain latent and reactivate at the time of transplant (e.g., tuberculosis [TB]); younger patients have a higher risk of acquiring primary infections after transplant, sometimes from the transplanted organ, which tend to be more severe than those secondary to reactivation of a latent infection (e.g., CMV). In addition, preexisting immunity can have a protective effect against clinical disease (e.g., EBV-associated posttransplant lymphoproliferative disorder [PTLD]).

The type of allograft affects the specific infectious risk due to technical factors associated with the transplant procedure but also inherent to the transplanted organ and risk of rejection. For example, urinary tract infections are most common after kidney transplant, either from catheter placement or ureter stenting. BK virus is ubiquitous, but associated nephropathy (BK virus-associated nephropathy) is most common after kidney transplant and is an important cause of allograft loss. Infection after liver transplant frequently results from leaks of biliary and gastrointestinal anastomoses. Cardiac assist devices in heart transplant recipients are a frequent source of infection. Tracheal anastomotic infections, particularly due to fungi, are a significant complication of lung transplantation. In single-lung transplants, recurrent infections of the native lung, such as gram-negative bacilli or fungi, can extend to the transplanted lung. In small intestine transplant, opportunistic and nonopportunistic viral infections of the gastrointestinal tract are common; these can be severe and even life-threatening (e.g., norovirus).

■ TIME COURSE OF INFECTIONS AFTER SOLID ORGAN TRANSPLANT (TABLE 148-2)

Despite all the differences in individual risk of infection, some general patterns of infection in the absence of antimicrobial intervention are similar among SOT recipients. This predictable temporal pattern has enabled the institution of specific prophylactic strategies and is useful when developing an infection differential diagnosis.

Early Period: Infections in the First Month After Transplant

Infections in this period are generally associated with technical complications of the transplant surgery and are nosocomial. Bacterial infections are most common during this time across all organs. Infections with prior colonizers (e.g., MDR bacteria) can occur, and bacterial prophylactic strategies should take these into account. Antimicrobial prophylaxis is usually started promptly after transplant. Hepatitis C occurs during this period when hepatitis C–positive donors are used. Opportunistic infections can happen in recipients with comorbid immunodeficiencies or those who received immunosuppression before transplant, but in the absence of these risk factors, the development of classic opportunistic infections during this early posttransplant period should alert to the possibility of donor allograft-transmitted infections (e.g., invasive coccidioidomycosis).

Intermediate Period: Infections 1–6 Months After Transplant

In the absence of routine antimicrobial prophylaxis, this period is characterized by the presence of classic opportunistic infections, such as *P. jirovecii* and CMV. The incidence of these infections has been significantly reduced or delayed with the use of prophylaxis or preemptive therapy. Reactivation of latent infections such as TB, Chagas' disease, endemic mycoses, and cryptococcosis can occur, so it is important to ascertain the risk before transplant and implement specific prophylactic strategies. Viral infections such as BK virus, adenovirus, RSV, hepatitis B virus, and EBV are common. Invasive fungal infections, specifically *Aspergillus* and other mold infections, can be problematic during this period of heightened immune suppression, but risk varies significantly by type of organ transplanted. Donor-derived infections may still occur.

Late Period: Infections 6 Months After Transplant This period is less well defined. Patients with satisfactory allograft function may develop more severe manifestations of community-acquired infections. Prophylaxes are frequently stopped, since the risk of opportunistic infections is lower than in the earlier period. Patients with poor allograft function due to rejection receive increased immunosuppression and are at higher risk of opportunistic infections for a variable period thereafter. Frequently, antimicrobial prophylaxes are "reset" when an episode of acute rejection ensues. The type of rejection (e.g., cellular or humoral) and its specific management may influence the infection risk. Given that all SOT recipients need lifelong immunosuppression, increased risk of infection persists. Late-onset CMV, nocardiosis, PJP, listeriosis, invasive fungal infections, and EBV-associated PTLD are some examples of opportunistic infections that can be expected in these patients. Patients with ongoing chronic viral reactivation, particularly CMV, are also at increased risk of other opportunistic infections. In addition, lung transplant recipients with chronic graft dysfunction are at high risk of recurrent bacterial pneumonia. Similarly, liver transplant recipients with chronic graft dysfunction frequently develop biliary strictures and recurrent cholangitis. Over time, MDR bacteria can become a problem.

VASCULAR COMPOSITE ALLOGRAFTS

VCAs refer to transplantation as a single functional unit of multiple tissues, such as muscle, blood vessel and nerve bundles, and associated viscera or bone. Face, hand, and uterus are common examples. These procedures are increasing in frequency and complexity, but data regarding infection risk are still scarce and inferred from the known risk of SOT. Due to the highly complex anatomy of the allografts and their natural microbial colonization, bacterial infections are most common. Relevant infection prevention strategies are like those used in SOT, such as updated immunizations pretransplant, donor/recipient serologic screening, and targeted antimicrobial prophylaxis. Planning and in-depth communication between the organ procurement team, the surgical team, and an infectious disease consultant are important so that relevant cultures are obtained and results reported back. Surveillance cultures for targeted microbiologic diagnosis will vary depending on the transplanted tissue. For example, for face transplants, in addition to cultures of respiratory pathogens, sinus cultures for bacteria and fungi are routinely obtained. Similarly, in penile or uterine transplants, sexually transmitted infections such *Chlamydia* or human papillomavirus need to be explicitly ruled out.

Opportunistic infection risk is largely driven by the induction, maintenance, and intensification of immunosuppression due to the frequent rejection episodes. Systemic viral infections, in particular CMV and EBV, have been a significant problem after VCA; similar prevention strategies, as in other high-risk SOT recipients, are recommended. Fungal infections, such as cryptococcus or mucocutaneous candidiasis, may mimic rejection, which underscores the importance of establishing a histopathologic and/or microbiologic diagnosis and reassessing frequently. As the field of VCA continues to evolve, the risk of infectious complications needs to be continually assessed so that in the future we can better define its similarities and differences from other types of transplants.

DONOR-DERIVED INFECTIONS

Donor-derived infections (DDIs) can be expected or unexpected. Expected transmission to the recipient occurs with pathogens that are known to be present in the transplanted organ. Viruses such as CMV, EBV, hepatitis B virus, hepatitis C virus, and BK polyomavirus are examples. Preventive strategies such as targeted antimicrobial prophylaxis and access to highly specific molecular assays minimize the risk of end-organ disease and the overall impact on the recipient despite transmission. Global and regional health and transplantation societies provide detailed recommendations for adequate screening for certain blood and tissue donor-derived pathogens, such as HIV and viral hepatitis, and provide specific guidance for continued surveillance in recipients of donors with a history of IV drug use, unprotected sexual activity, multiple sexual partners, or incarceration within 30 days of potential donation. Recommendations often vary between organizations and countries. Transmission of unexpected infections occurs when pathogens not detected in an organ donor prior to organ recovery are transmitted to recipients. Screening donors for transmissible pathogens requires clinical and epidemiologic history as well as laboratory evaluation. Despite clinical and laboratory screening of potential donors for transmissible diseases, unexpected transmission of infection from donor to recipient remains an inherent risk of transplantation, predominantly of solid organ and composite tissue transplantation. Unexpected donor-derived diseases affect <1%

TABLE 148-2 Common Infections After Solid Organ Transplantation by Site of Infection and Transplant Type

INFECTED SITE	TYPE OF INFECTION	RISK FACTOR	MANAGEMENT
Transplanted organ	Bacterial and fungal infections of the graft, anastomotic site, and surgical wound	Prior colonization of the donor with MDR organisms	Institute targeted antimicrobial treatment as soon as prior colonization is known (peritransplant or immediately posttransplant). Source control (surgical or interventional radiology) and continued directed antimicrobials. Optimal duration after source control is unknown. Usual correlates of adequate treatment are clinical, laboratory, and radiologic resolution.
Surgical incision infection	Bacterial, yeast infections are most common; NTM (rare, mostly thoracic transplants)	Pain, erythema, discharge, or dehiscence of wounds, typically within the first 30 days posttransplant	Collection of microbiologic specimens for culture. Thorough washout and debridement. Imaging to evaluate for the presence of deeper infection. Antimicrobials should be adjusted based on microbiologic results, and duration guided by clinical response.
Intra-abdominal infections in liver transplant recipients	Bacterial and yeast infections	Secondary to biliary anastomoses leaks, or Roux-en-Y hepaticojejunostomy or other anastomoses that increase the risk of intestinal reflux into the biliary system; hepatic necrosis, from hepatic vascular thrombosis	Obtain blood cultures. Deep tissue or fluid cultures obtained intraoperatively or by IR. Source control. Antimicrobials adjusted based on microbiologic results (polymicrobial infections are the norm). Optimal duration of antimicrobials is unknown. Usual correlates of adequate treatment are clinical, laboratory, and radiologic resolution.
Intrahepatic infection in liver transplant	Recurrent cholangitis (bacteria and yeast)	Biliary strictures: anastomotic (acute) or nonanastomotic (chronic allograft rejection)	Biliary drainage and antimicrobials.
Urinary tract infections in kidney transplant	Recurrent cystitis and pyelonephritis (bacteria, in particular Enterobacteriaceae); *Mycobacterium tuberculosis* and yeasts are less frequent causes of UTI but should be suspected when there is pyuria and/or hematuria and negative cultures	Duration of indwelling urinary tract catheters, female recipients, as well as recurrent UTI prior to the transplant	Prompt catheter and ureteral stent removal. Targeted antimicrobials. If recurrent UTI, evaluate for anatomic abnormalities (ureteral reflux, ureterovesical junction stenosis, neurogenic bladder, bladder diverticulum).
Polyomavirus-associated nephropathy in kidney transplant	BK, rarely JC virus	Heightened immunosuppression, male recipient, older age, and a high BK viremia (>10,000 copies/mL)	The only effective intervention is reduction in immunosuppression. An allograft kidney biopsy is necessary when there is suspicion of rejection or a different infection.
Pneumonia and tracheobronchitis in lung transplant	Bacterial infections: severe or recalcitrant community-acquired viral pathogens (RSV, influenza, SARS-CoV-2); invasive mold infections; fungal tracheobronchial infection; NTM pulmonary disease	Pretransplant colonization with resistant organisms, requiring endotracheal intubation prior to the transplant, and mucociliary dysfunction, are significant risk factors for bacterial infections; CLAD is a risk factor for and a result of recurrent viral, bacterial, fungal, and mycobacterial infections	Early antimicrobial therapy if possible. For viral respiratory pathogens, prevention with seasonal annual vaccination is recommended, with monoclonal antibodies when available. In patients with CLAD, respiratory rehabilitation and pulmonary toileting.
Mediastinitis in thoracic organ transplants	Bacterial or *Candida* mediastinitis, with or without sternal osteomyelitis; less common *Nocardia* spp., NTM, or molds	Diabetes mellitus, surgical reoperation, and acute rejection	Deep tissue cultures (intraoperatively). Duration of antimicrobials is unknown. Correlates of adequate treatment are clinical, laboratory, and radiologic resolution.
Myocarditis, pericarditis, and cardiomyopathy in heart transplant	*Toxoplasma gondii* infection	Toxoplasma pretransplant serology D+/R–	Diagnosis is established by the identification of tachyzoites in the endomyocardial biopsy or PCR in tissue and/or pericardial fluid. Preemptive monitoring via blood PCR and TMP-SMX prophylaxis reduce incidence of disease.
	Trypanosoma cruzi	Pretransplant recipient seropositivity (hearts from seropositive donors are not used)	Clinical manifestations are fever, myocarditis, and painful skin lesions. Preemptive monitoring via blood PCR to detect early reactivation is indicated in all previously known seropositive recipients. Antitrypanosomal treatment (benznidazole or nifurtimox) for all patients with reactivation and tissue-invasive disease.
Infectious hyperammonemia syndrome in lung transplant (less commonly in other transplants)	*Ureaplasma* spp. systemic infection (less commonly *Mycoplasma* spp.) that results in encephalopathy due to high ammonia levels >200 μmol/L	Donor lung colonization	Targeted antimicrobial treatment and direct ammonia removal through renal replacement therapy, lactulose, and reducing enteral protein intake. Prevalence in transplanted organs is roughly 10%, and screening should be considered.

Abbreviations: CLAD, chronic lung allograft dysfunction; D+/R–, donor positive, recipient negative (serology); IR, interventional radiology; MDR, multidrug-resistant; NTM, nontuberculous mycobacteria; PCR, polymerase chain reaction; RSV, respiratory syncytial virus; TMP-SMX, trimethoprim-sulfamethoxazole; UTI, urinary tract infection.

of all transplant recipients, and most are infections. (Besides infection, donors may also transmit unsuspected malignancies and other conditions [e.g., hemochromatosis], but these are beyond the scope of this chapter and will not be discussed.) Unexpected DDIs occur predominantly in cadaveric donation, where screening of potentially transmissible infection is hampered by the incomplete assessment of the donor's risk. Almost all become apparent within 90 days of the transplant surgery. Overall mortality is around 15%, but it is much higher (30–50%) in some fungal (e.g., aspergillosis, coccidioidomycosis) and parasitic infections (e.g., strongyloidiasis, toxoplasmosis). The risk of DDI needs to be balanced against the growing number of patients on the transplant waitlist and their associated risks of mortality. The early diagnosis of a DDI requires a high index of suspicion. Clinicians should be on alert to the possibility of an unexpected transmission. DDI should be considered when opportunistic infections happen outside the expected at-risk period, in the case of infections with no clear epidemiologic risk factors in the recipient, with unusual clinical syndromes, and with persistent undiagnosed fever, especially during the first 90 days. Transplant and infectious disease physicians should be alerted to clusters of infections in recipients from the same donor. When an unexpected DDI is suspected, besides prompt diagnosis and treatment in the recipient, immediate communication with public health authorities and the organ procurement organization is essential to decrease the risk of transmission/disease in other recipients.

Donors with a known or suspected active infection at the time of donation, such as pneumonia, urinary tract infection or bloodstream infection, may transmit the infection to the recipient. Living donors with an identified active infection should ideally be treated and transplantation delayed until infection resolves. In deceased donors, targeted antimicrobial prophylaxis for recipients is highly effective in preventing disease transmission when the infection is due to susceptible bacteria. Bacteremia in deceased donors is not uncommon, and studies have shown that, in adult recipients, donor bacteremia is not associated with adverse outcomes if a feasible treatment option exists. Bacterial meningitis carries a similar risk of transmission. Blood cultures should be obtained to rule out occult bacteremia in deceased donors, but bacteremia is generally not considered a contraindication to organ procurement and transplantation, unless no effective antimicrobial therapy is available. However, if the organ selected for transplantation is not infected or colonized and the donor is not bacteremic, the risk of transmission is low. These scenarios require careful analysis, and decisions should be individualized. The patient should be involved in deciding the acceptance of donors who may entail a higher risk for the recipient.

Meningoencephalitis in SOT recipients can be devastating. Viruses cause most cases, and effective antiviral treatments exist for only a few of these. Parasitic and fungal meningoencephalitis are also associated with significant morbidity and mortality. Development of meningoencephalitis or other neurologic syndrome in an SOT recipient, especially early after transplantation, should raise suspicion for DDI. Meningoencephalitis of unknown etiology in a donor is one of the few remaining absolute contraindications of transplantation. Even in patients in whom the etiology is suspected, caution is advised, and a comprehensive evaluation for infectious causes, with special consideration for opportunistic infection when the donor is from an endemic area, should be undertaken.

TB and endemic mycoses have been well-documented donor-transmitted infections and are associated with significant risk of mortality or graft loss. Routine screening is recommended in areas of endemicity or in other recognized high-risk scenarios for TB, such as incarceration, displacement, or homelessness. Screening in deceased donors is frequently limited by the extent of the medical history and time constraints.

IMPORTANT PATHOGENS IN TRANSPLANT RECIPIENTS

CYTOMEGALOVIRUS

CMV is the most important virus in both SOT and HCT recipients. CMV is usually acquired in early childhood and establishes latency for life in many different cell types, including endothelial cells and dendritic cells. Seropositivity for CMV (defined as positive anti-CMV IgG) varies by geographic region and ranges from 30 to 97%, increasing with age. CMV infection is defined as detection of the virus in the blood (typically by PCR, which has replaced the old pp65 antigenemia test). CMV DNA may be quantified in whole blood or in plasma, with results expressed in international units per microliter (IU/μL), and only results using the same test should be compared. CMV viremia is usually asymptomatic, but SOT recipients may develop a syndrome of fever, malaise, myalgias, moderate leukopenia with occasional atypical peripheral blood lymphocytes, and mild transaminitis concomitant with the detection of CMV in the peripheral blood. This is called CMV syndrome. Tissue-invasive CMV disease is diagnosed by histopathologic proof of invasion by the virus and tissue damage. Pneumonitis is the most feared end-organ disease after allo-HCT, and gastrointestinal disease is the most common, both after SOT and allo-HCT, but many other organs may be affected. Typically, CMV infection precedes development of disease by a few days, but sometimes CMV gastrointestinal disease occurs before the virus may be detected in the blood, or simultaneously. Clinically significant CMV infection is a term used in some clinical trials of CMV prophylaxis and includes invasive disease and/or a viremia level that triggers treatment with antivirals. Before widespread use of interventions aimed at reducing the incidence of CMV end-organ disease (i.e., pneumonitis, carditis, gastroenteritis, hepatitis), this occurred frequently during the first 1–6 months after transplantation and was associated with high morbidity and mortality. In addition, active CMV replication is associated with several "indirect effects," including increased rates of bacterial and fungal infection, increased graft dysfunction, transplant vascular disease, acute rejection, and, after HCT, increased GVHD and increased nonrelapse transplant-related mortality.

Avoiding CMV disease is an important part of transplant medicine. The process starts with donor selection, as the serostatuses of both donor and recipient are the main determinants of the risk of disease. In SOT, the highest risk is seen with a seropositive donor going into a seronegative recipient (D+/R–). In this setting, up to 50% of recipients may become viremic after prophylaxis is stopped, with lung, small bowel, pancreas, and combined kidney-pancreas having higher risks. After HCT, the highest risk is for seropositive recipients (who harbor the virus in many of their tissues) receiving stem cells from seronegative donors (D–/R+), 60–70% of whom may develop CMV viremia. When several donors are available, avoiding high-risk CMV serostatus combinations is part of the selection process. Once the transplant takes place, prevention of CMV disease can be achieved by two different strategies: universal prophylaxis or preemptive management. There are two strategies to prevent CMV disease after transplant: universal prophylaxis, in which every recipient at high risk is given antivirals (usually valganciclovir after SOT or letermovir after allo-HCT) for an arbitrary period (usually 3–6 months after SOT and 100 days after allo-HCT), or preemptive management, in which CMV is monitored in the blood weekly and anti-CMV treatment initiated when a particular level of viremia is detected. Of note, valganciclovir (a prodrug of ganciclovir) is active against HSV and VZV, but letermovir is not, so if letermovir is used, acyclovir or valaciclovir must be added for HSV/VZV prophylaxis. The tolerable level of viremia differs based on type of transplant, timing of the infection, and other factors that increase the risk of CMV disease. In UCB transplant or T-cell-depleted transplants, lower levels of viremia may trigger treatment, as the kinetics of viral replication and risk of progression to end-organ disease may be significantly higher in these patients. Randomized trials comparing both strategies have been performed only after allo-HCT for kidney and liver transplant but are not available for SOT involving other organs. Prophylaxis and preemptive treatment seem similarly effective in terms of overall mortality, and transplant centers may choose one over the other based on considerations such as logistics (more complicated with preemptive therapy) and cost (potentially higher with letermovir). Treatment of significant viremia is with ganciclovir or foscarnet. Cidofovir and maribavir are inferior, and the former is potentially more toxic. Ganciclovir and foscarnet are considered equally effective but differ in toxicity: ganciclovir has significant myelotoxicity, and foscarnet causes electrolyte abnormalities that usually cannot be managed in the outpatient setting, as well as nephrotoxicity. If viremia increases after 2 weeks of optimal antiviral treatment or if it does not decrease, the infection is considered "refractory" or "possibly refractory" CMV, respectively, and testing for genetic resistance to antivirals is recommended. The antiviral maribavir is FDA approved for resistant/refractory CMV, but its efficacy is only moderate, with <20% of patients with resistant/refractory CMV maintaining control of the virus at 16 weeks.

Invasive disease due to CMV varies with the type of transplant. After allo-HCT, CMV pneumonia is the most severe, but it is now uncommon with the use of prophylaxis or preemptive management.

Involvement of the gastrointestinal tract (most commonly CMV colitis, but any segment of the bowel may be affected) is more frequent, in part because the time between asymptomatic viremia and CMV colitis is shorter. After SOT, CMV reactivation may start in the transplanted organ in CMV-seropositive donors (e.g., CMV pneumonia in lung transplant, enteritis in small-bowel transplant). Hepatitis, retinitis, and, rarely, central nervous system (CNS) involvement (myelopathy, encephalopathy) may also occur.

Frequency of CMV monitoring after prophylaxis or treatment is completed is not well established. If available, immune monitoring, via commercially available CMV-specific interferon γ (IFN-γ) release assays that detect CMV-specific T cells in whole blood, may allow for more precise targeting of monitoring and prophylactic strategies.

EPSTEIN-BARR VIRUS

Most adults are infected with EBV, which remains latent in B cells. When an EBV-seropositive donor provides an allograft to an EBV-seronegative recipient (D+/R–), EBV infection (detection of EBV in blood or plasma) is very common. This primary EBV infection may be asymptomatic, present with end-organ disease such as hepatitis or enteritis, or cause a mononucleosis-like febrile illness with or without lymphadenopathy, atypical lymphocytes, and culture-negative exudative pharyngitis. EBV-associated PTLD is a heterogeneous group of clinical syndromes associated with uncontrolled lymphoproliferation (typically of B cells, but cases of natural killer– and T-cell EBV-positive PTLD occur), which ranges from polyclonal proliferation to true lymphoma containing clonal chromosomal abnormalities. The diagnosis requires tissue examination, which may show a polymorphic or monomorphic proliferation of B cells or even frank lymphoma. Risk of PTLD after SOT is greater in EBV-seronegative transplant recipients, who acquire a primary infection through the allograft. Consequently, this scenario is most common in pediatric transplant recipients. Viral burden within the lymphoid tissue of the transplanted organ appears to be a risk factor for PTLD, with intestinal transplant recipients at the highest risk and kidney lowest. EBV-positive PTLD is most frequent in the first year after SOT, but late PTLD incidence increases after 7–10 years and is related to duration of immunosuppression. Late PTLD can present as primary CNS lymphoma and gastrointestinal tract disease. EBV viral load monitoring and radiologic evaluation can assist in early diagnosis. Positron emission tomography/computed tomography (CT) imaging is useful in establishing lymph node or end-organ disease involvement, thus allowing for targeted tissue sampling to establish the diagnosis. In SOT, reduction of immunosuppression is the preferred initial treatment strategy. Rituximab is recommended next if the PTLD cells express CD20. Cytotoxic chemotherapy may be required. For refractory cases, adoptive immunotherapy with EBV-specific T cells, including chimeric antigen receptor T cells and third-party trained T cells, such as tabelecleucel, have been used. All donors and recipients should undergo EBV serologic testing before transplant. Preemptive monitoring of EBV viral load in high-risk (EBV D+/R–) SOT recipients is suggested, and reduction of immunosuppression is recommended in EBV viremic patients. Withdrawal of immunosuppression may trigger graft rejection. Specific viral load cutoffs that should trigger an intervention have not been established; there is considerable interlaboratory variability even with commercial EBV PCR assays. Serial monitoring with the same assay (e.g., whole blood or plasma) and the same laboratory is necessary to establish a trend. Rapidly rising EBV viral load is a significant risk factor for PTLD.

EBV-positive PTLD is an early complication of allo-HCT, occurring in the first 2–4 months. In this case, the virus is usually present in the recipient, and the risk is highest in EBV D–/R+ pair and cord blood stem cell source, although any D/R mismatch (D+/R– or D–/R+) in the setting of T-cell-depleted grafts is recognized as a significant risk. EBV D/R matching is recommended whenever possible. The overall frequency is ~3%, varying between 1% in matched sibling donor transplant to up to 10% in unrelated UCB transplants. The risk depends on the degree of T-cell depletion because it is the T-cell function after transplant (derived from the donor) that will control EBV. The management is like that described above, with the caveats that decreasing immunosuppression after allo-HCT may be difficult or impossible if GVHD is present and that, if the stem cell donor is available and is EBV positive, administration of lymphocytes from the donor (donor lymphocyte infusion) may be a successful strategy, as a significant fraction of the T-cell repertoire is directed against EBV. Donor-derived adoptive immunotherapy with EBV-specific T cells or off-the-shelf commercially available EBV-specific T cells, such as tabelecleucel, are increasingly used as second-line therapy.

POLYOMAVIRUSES

Polyomaviruses have been identified as frequent causes of allograft dysfunction in kidney transplant recipients. BK virus (BKV) and JC virus (JCV) are the most common pathogenic poliomaviruses that infect humans. BKV (most commonly) and JCV have been associated with tubulointerstitial nephritis and nephropathy, usually referred to as polyomavirus-associated nephropathy (PVAN). Ureteric stenosis and hemorrhagic cystitis can also occur. BKV infection is most common in kidney transplant recipients. Viruria is common, but some patients develop viremia, which in the absence of an intervention results in PVAN and allograft loss. The presumptive diagnosis of BKV nephropathy can be established with renal dysfunction and viremia, but acute rejection and PVAN can coexist and can be clinically indistinguishable. Definitive histopathologic diagnosis requires identification of viral cytopathic changes and immunohistochemistry for SV40 antigen or BKV-specific antigen detection in the tissue. Reduction of immunosuppression is the mainstay of treatment. Preemptive monitoring of BKV viral load can aid in the early detection of BK viremia; increasing viral loads are predictive of PVAN, and careful reduction of immunosuppression can prevent renal dysfunction and graft loss.

BKV after allo-HCT was discussed earlier, in the section on early infections after HCT, above.

After PVAN, PML caused by JCV CNS infection is the most common disease caused by polyomaviruses in SOT, although it is significantly less common than in other immunocompromised patients, such as those with advanced HIV. Diagnosis is established by a compatible neurologic syndrome, molecular detection of JC virus in the CSF, and radiologic evidence of demyelinating disease. As with other polyomavirus infections, no effective antiviral therapy exists, and reduction of immunosuppression is commonly employed, with overall poor response rates. Directed T-cell therapies are under development, and immune checkpoint inhibition (e.g., pembrolizumab) has been used successfully in some cases.

FUNGAL INFECTIONS

Invasive fungal infections are a common infectious complication in transplant recipients. Newer antifungals and corticosteroid-sparing immunosuppressive regimens have decreased the overall incidence of invasive fungal infections and improved outcomes. Invasive candidiasis is the most common invasive fungal infection in SOT recipients. Other fungal infections such as invasive aspergillosis, cryptococcosis, and endemic mycoses are also prevalent. In HCT, almost universal antifungal prophylaxis with activity against *Candida* has made mold infections relatively more frequent.

Candidiasis Invasive candidiasis is most often an early complication after SOT. Length of hospitalization, frequent use of broad-spectrum antimicrobials, intravascular catheters, renal replacement therapy, and critical illness are all risk factors for increased incidence of invasive candidiasis. Additionally, the complex nature of abdominal organ transplant surgical procedures increases the risk of invasive candidiasis. In liver transplant recipients, choledochojejunostomy, large-volume transfusion of cellular blood products (including platelets and packed red blood cells), reoperation, and retransplantation are all risk factors for invasive candidiasis. Pancreas transplantation with enteric drainage and small-bowel transplantation are particularly prone to be complicated by candidemia or invasive candidiasis. Targeted antifungal prophylaxis in liver, pancreas, and small-bowel transplant recipients is routine and decreases the incidence of invasive candidiasis.

All forms of invasive candidiasis in transplant recipients merit antifungal treatment, and echinocandins are the initial treatment of choice, except in CNS disease or urinary tract infection. In addition to antifungal treatment, source control, such as catheter removal, and mechanical drainage of abdominal or surgical collections are the mainstay of successful treatment. Lung transplant recipients are at risk of invasive disease at the site of tracheal anastomosis, which can lead to anastomotic leak. Mycotic aneurisms of the main vessel of the allograft are associated with a high risk of rupture and usually represent a DDI or result from contamination of the preservation fluid. Early diagnosis and aggressive medical and surgical treatment are indispensable for overall survival.

Aspergillus and Other Molds Invasive pulmonary aspergillosis is the most common invasive mold infection in transplant recipients and the overall most common invasive fungal infection in lung transplant recipients. Airway and allograft colonization with *Aspergillus* spp. is a common but also invasive disease. Lung transplant recipients experience the full spectrum of *Aspergillus* spp. pathology, including tracheobronchitis, bronchial anastomotic infections, invasive pulmonary aspergillosis, and disseminated disease. Tracheobronchitis and bronchial anastomotic infections are characterized by ulceration and cartilage invasion. Dissemination, including to the brain, can occur and should be ruled out in all transplant recipients. Risk factors are single-lung transplant, lung transplant for cystic fibrosis, airway ischemia, rejection with increased immunosuppression, and chronic lung allograft dysfunction (CLAD). Risk factors for invasive aspergillosis (IA) in other SOTs include CMV infection and disease and degree of immunosuppression, such as high-dose corticosteroids for rejection. Retransplantation and posttransplant liver or renal hepatic failure, as well as prolonged neutropenia, increase the overall risk of IA.

Targeted antimold prophylaxis in those at high risk of IA is recommended. Monitoring serum galactomannan (GMN) does not perform well in SOT recipients. Targeted diagnostic bronchoalveolar lavage (BAL) GMN and serum GMN are preferred when there is suspicion of invasive disease. CT imaging can assist in the early diagnosis of IA. The CT findings are variable **(Fig. 148-1)**. Whereas dense, well-circumscribed pulmonary nodules, sometimes with halo sign, are characteristic in patients with hematologic malignancies, SOT recipients more frequently present with nonspecific pulmonary findings such as alveolar and centrilobular infiltrates, and the halo sign is almost never seen. Despite early diagnosis and treatment, the mortality associated with IA is still high. Other mold infections in SOT are less frequent but significantly more common than in nontransplant populations.

Aspergillosis after HCT is a complication of prolonged neutropenia or the use of high-dose corticosteroids for the treatment of GVHD. Prophylaxis with posaconazole is frequently used in patients receiving >0.5 mg/kg of prednisone every day or being treated for GVHD with two or more immunosuppressants. Some of the newer treatments for GVHD, such as ibrutinib and ruxolitinib, seem to significantly increase the risk of IA. Presentation is nonspecific with dyspnea, cough, and chest pain. Fever may be absent. Lack of systemic toxicity is frequent in mold infections. Establishing an accurate diagnosis is essential, since most mold infections can have similar presentations, and antifungal susceptibility varies significantly between molds.

Cryptococcosis With prolonged immunosuppression, SOTs are at risk of cryptococcal infection and disease. These tend to occur later after SOT, with most cases occurring between 1 and 3 years after the transplant. Earlier presentation should alert to the possibility of donor-derived cryptococcosis. Advanced liver disease is a risk factor for cryptococcosis, and liver transplant recipients may present earlier after transplant, especially if mental status changes peri-transplant were attributed to hepatic encephalopathy but no microbiologic workup was completed. Other presentations include skin papules or cellulitis, pneumonia, and asymptomatic or minimally symptomatic lung nodules. Both *Cryptococcus neoformans* and *Cryptococcus gattii* are known pathogens in SOT recipients. Serum cryptococcal antigen assay allows early noninvasive diagnosis but requires a high index of suspicion. In patients with CNS disease, both serum and CSF cryptococcal antigen are extremely useful in the early diagnosis of cryptococcal disease. In the evaluation of patients with mental status changes with or without fever or with cryptococcal disease outside of the CNS, patients should undergo a lumbar puncture with CSF opening pressure, as well as cell count and culture. These are useful for the diagnosis of CSF involvement, but also to evaluate the appropriate treatment response. Monitoring of cryptococcal antigen to determine the duration of therapy is not recommended because it may persist despite microbiologic and clinical resolution. Immune reconstitution inflammatory syndrome has been described in SOT recipients when immunosuppression is tapered aggressively in cryptococcal meningitis and may result in worsening of the initial symptoms despite microbiologic cure.

Cryptococcosis is less common after allo-HCT, probably due to the prevalence of antifungal prophylaxis.

Endemic Mycoses SOT recipients are second only to people with advanced HIV in terms of risk of disseminated endemic mycoses. Given its worldwide endemicity, histoplasmosis is the most common, but disseminated coccidioidomycosis, in particular CNS disease, is associated with higher mortality, and patients who survive usually require lifelong antifungal therapy. Invasive forms of all endemic mycoses have been described in transplant recipients.

Pneumocystis jirovecii PJP is an important complication of SOT and allo-HCT, and prophylaxis early after transplant is used routinely. Breakthroughs are uncommon if patients are receiving TMP-SMX but may be seen with second-line prophylaxis such as inhaled pentamidine or atovaquone. Lung transplant recipients retain a lifelong risk of infection. Acute or subacute shortness of breath, with significant hypoxemia, is the most common presentation. Radiologic changes may be subtle and out of proportion to the hypoxemia, with a broad alveolar-arterial gradient. Many radiologic presentations have been described, but the most common are ground-glass opacities. Nodules and atypical

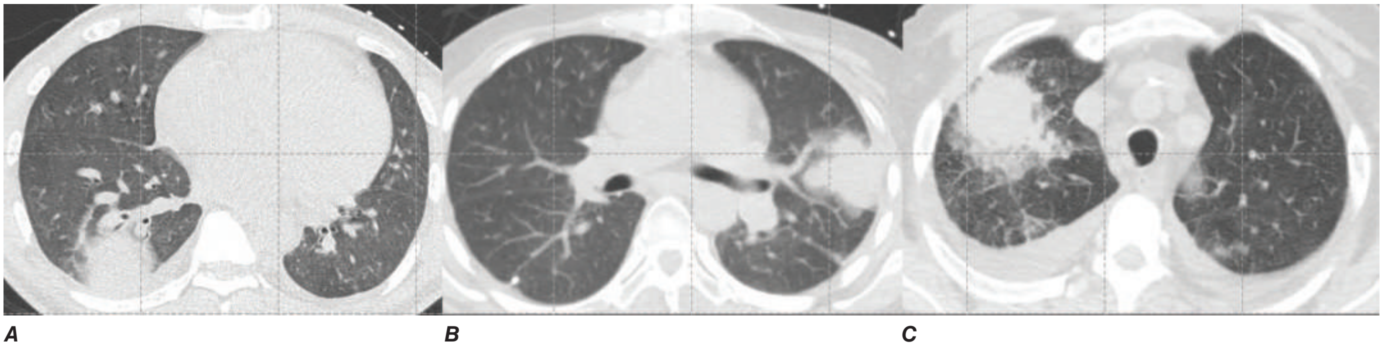

FIGURE 148-1 Radiologic findings are often nonspecific in transplant recipients. ***A.*** *Nocardia* 3 months after kidney transplant in a patient who was not receiving trimethoprim-sulfamethoxazole. A fine-needle aspirate showed only necrosis, but the bronchoalveolar lavage (BAL) grew *Nocardia cyriacigeorgica.* ***B.*** Tuberculosis 3 months after heart transplant in a patient with known latent tuberculosis infection who was receiving isoniazid prophylaxis. The isolate of *Mycobacterium tuberculosis* was resistant to isoniazid. ***C.*** Aspergillosis in a neutropenic patient receiving allogeneic hematopoietic stem cell transplant for acute myelogenous leukemia. The BAL was positive for galactomannan, but the culture was negative. The mold was later identified as a non-fumigatus *Aspergillus* by next-generation sequencing of a surgical specimen.

consolidations have also been described, especially in patients who receive less effective prophylactic agents, such as inhaled pentamidine or atovaquone. Most patients with PJP have elevated β-D-glucan in serum. The definitive diagnosis is established by identification of *Pneumocystis* trophozoites in a respiratory sample from an induced sputum, a BAL, or a lung biopsy, but microscopy has low sensitivity, and increasingly, PCR in the BAL has become the diagnostic modality of choice. A positive PCR requires clinical interpretation because some patients may be merely colonized. During treatment, an immune reconstitution inflammatory syndrome (IRIS)-like condition can occur, so tapering of immunosuppression is usually not advised in transplant recipients newly diagnosed with PJP. The use of adjuvant corticosteroids in non-HIV patients with severe PJP remains controversial.

MYCOBACTERIAL INFECTIONS

Mycobacterium tuberculosis TB remains a major challenge and public health threat in transplant recipients. Incidence of active TB among SOT recipients is estimated to be 20–75 times higher than in the general population. Incidence varies significantly among different geographic regions and ethnic groups. Most cases occur between 6 and 12 months after the transplant and are associated with very high mortality; many more cases result in graft loss. Donor-derived TB has been also well documented after SOT, and these cases can present earlier after transplant. In contrast, TB after HCT is relatively uncommon and presents early within the first year after allo-HCT. In areas of medium or high endemicity, cases of active TB develop most commonly from reactivation of old foci of infection, but primary infection after transplant can also occur. Disseminated disease is common, and fever is the most common presenting symptom (>90%). In those with only pulmonary involvement, fever is present in ~60%. Involvement of every organ and atypical presentations have been reported (Fig. 148-1). Good-quality data regarding the optimal combinations of drugs and length of therapy are lacking in organ transplant recipients. Decisions are based on case series, recommendations for the general population and other immunocompromised individuals such as people living with HIV, and expert opinion. Most cases require 6–9 months of treatment, but in disseminated disease with CNS involvement, extrapulmonary TB, cavitary TB, or culture positivity after 2 months of treatment, 12–18 months of treatment are recommended. Daily therapy is recommended over intermittent dosing, given signals of higher relapse rates with intermittent dosing and concerns of less reliable immunosuppressive drug levels due to drug–drug interactions that may ultimately result in higher rejection rates or graft loss. Rifamycin-containing regimens are preferred for all types of TB due to the potent sterilizing activity of such regimens. Rifampin, the most used rifamycin, is a potent inducer of the cytochrome P450 enzyme, which increases the metabolism of many drugs, including calcineurin inhibitors (cyclosporine and tacrolimus), mTOR inhibitors (sirolimus and everolimus), and corticosteroids; its use has been frequently associated with acute rejection. Rifampin-based treatment is extremely problematic. Even though there is less experience with the use of rifabutin and it is not available worldwide, rifabutin is a weaker inducer of cytochrome p450, and drug interactions are significantly easier to manage than with rifampin. Other rifamycin-sparing treatment protocols can be used in organ transplant recipients to avoid drug–drug interactions; however, when a rifamycin is not used, shorter courses (6–9 months) should not be used. Use of a rifamycin is also highly desirable in patients with severe, disseminated, or cavitary disease. The initial treatment regimen should follow local recommendations based on local resistance patterns and tailored to available susceptibility data. Newer antituberculous antimicrobials have not been studied in transplant recipients but are promising. Reduction of immunosuppressive therapies, which help control the infection, should be done slowly if at all, as IRIS is a well-known complication of mycobacterial disease. If symptoms worsen with lowering immunosuppression, IRIS should be considered.

Prevention of TB after transplantation relies on adequately identifying those at risk of reactivation or recipients of an organ from a donor who has latent TB infection (LTBI) and treating LTBI. Screening relies on known epidemiologic exposures and probing the individual's immune response to TB antigens. There are two methods available for screening for prior TB exposure (for additional information, see Chap. 183), the tuberculin skin test (TST) and the IFN-γ release assays (IGRAs); both assays are less sensitive in immunocompromised hosts. In transplant candidates or recipients, a TST >5 mm is consistent with LTBI. Patients with prior positive bacillus Calmette-Guérin vaccination can have a positive TST from prior vaccination. In these patients, IGRA is preferable. End-stage renal disease, end-stage liver disease, and receipt of prior chemotherapies are all well-known risk factors that can lead to a false-negative TST or indeterminate results in the IGRAs. In patients with indeterminate results, the decision to treat LTBI should be made clinically, according to epidemiologic risk factors based on history. In patients with a negative TST, a second TST should be performed at about 2 weeks to look for a booster effect. An IGRA test can be performed after an initial negative TST in those at risk. In patients with evidence of LTBI by TST or IGRA, active TB should be ruled out before proceeding to transplant. Live donors should undergo a similar screening as recipients. TST and IGRAs may not be available on deceased donors; therefore, risk should be assessed based on a detailed history from the donor's family regarding previous active or latent TB and any treatment received and chest imaging.

All transplant recipients with LTBI should receive treatment. Consideration should be given to time to transplantation, drug–drug interactions with ongoing treatments, and potential toxicities. In patients in whom the transplant is unlikely to occur until at least 4–6 months, shorter courses are suggested (e.g., rifampin for 4 months or isoniazid plus rifapentine for 12 weeks). If the transplant is likely to occur sooner, isoniazid for 9 months is recommended. When possible, LTBI should be treated prior to transplantation; however, the transplant should not be delayed to complete treatment. For liver transplant candidates or HCT candidates receiving chemotherapy, treating prior to transplantation is challenging due to potential hepatoxicity or significant drug–drug interactions. In these patients, it may be safer to begin therapy after transplantation once liver function has normalized or the conditioning regimen has been completed. SOT recipients who received an organ from an untreated donor with known LTBI should also be treated. If a living donor has LTBI, treatment of the donor should prioritize their risk of progression to active TB following national guidelines for the treatment of LTBI.

Nontuberculous Mycobacteria Nontuberculous mycobacteria (NTM) are ubiquitous in the environment, and some species are important opportunistic human pathogens. Transplant recipients are at increased risk of NTM infection. The lungs are the most common site of infection and the most frequently colonized organ. Disease due to NTM can be separated into pulmonary, disseminated, catheter- and cardiac implantable electronic device–associated infection, surgical site deep-seated infection, and skin and soft tissue (SST) infections. Lung followed by heart transplant recipients are at highest risk of post-transplant NTM disease. Because structural lung disease increases the risk of colonization, allo-HCT recipients with cGVHD of the lungs are also at risk. Lung transplant candidates with cystic fibrosis or chronic obstructive pulmonary disease are at increased risk of being colonized or infected by NTMs, and pretransplant isolation of NTM is associated with increased risk of NTM disease after transplant. NTM disease is associated with increased mortality and CLAD. Most common infections are due to *Mycobacterium abscessus* and *M. avium* complex. Diagnosis of NTM disease from nonsterile sites, especially the respiratory tract, can be a challenge. Validation of criteria suggested for the routine diagnosis of pulmonary NTM in immunocompetent individuals is lacking in transplant recipients; however, these provide a useful framework for the diagnosis of pulmonary NTM in transplant recipients. In lung transplant recipients, the threshold for treatment of infection should be lower, given the high rate of infection, secondary dissemination, or development of CLAD. Treatment of NTM is complex and often compromised due to drug–drug interaction and drug toxicities. Species-level identification is necessary to design a treatment regimen. For detailed treatment strategies, the reader is referred to Chap. 185 and American Thoracic Society/British Thoracic Society guidelines. The length of treatment should be guided by the specific clinical scenario; localized SST infections require shorter treatment durations, whereas pulmonary or disseminated NTMs require at least 1 year (frequently longer).

All patients with NTM infection pretransplant should start treatment before transplantation. Routine screening for NTM infection in lung transplant candidates is recommended, and those found to be colonized or infected should be treated, ideally for at least 3 months prior to proceeding to transplant. Rapidly growing NTMs, especially *M. abscessus,* have been associated with a high risk of deep surgical site infection and with frequent bone involvement. Treatment regimens should be based on species identification and susceptibility testing when appropriate, as well as stabilization of disease before transplant and ability to tolerate the designed regimen. Therapeutic drug monitoring can help to reduce toxicity. Bilateral lung transplant is preferred in those already colonized or infected with NTMs prior to transplant. Despite these measures, the risk of posttransplant infection is high, and the patient should be counseled about this risk.

Some PIDs that result in susceptibility to mycobacterial infection present with disseminated NTM or TB. Without successful replacement of the deficient immune system via allo-HCT, these infections are difficult to treat. Management of these infections is challenging; aggressive antimicrobial treatment is required before, during, and early after transplant. Ideally, infection and associated inflammatory response should be partially controlled prior to transplantation, which usually requires at least 3 months of treatment before the starting conditioning. Drug–drug interactions and drug toxicities in the peri- and posttransplant period are a challenge. In the first few months after engraftment, IRIS-like phenomena are common in patients being treated for active mycobacterial infection at the time of transplantation.

Definitive diagnosis of IRIS is a challenge, as there are no clear diagnostic criteria for this disease. Classically, there is worsening of symptoms or imaging at the site of prior infection that coincides with tapering of immune suppression. Adequate management requires ongoing treatment of the underlying infection, and augmented immunosuppression or anti-inflammatory therapy. Once symptoms are controlled, tapering of immunosuppression should be done slowly, taking into consideration other potential toxicities of ongoing immunosuppression and predisposition to other opportunistic infections.

NOCARDIOSIS

Nocardia is an aerobic gram-positive bacillus found in the soil that infects predominantly immunocompromised people. It is more frequently acquired by inhalation, although skin inoculation also occurs (typical with *Nocardia braziliensis*). The presentation varies from acute with fever, cough, and dyspnea, to subacute or chronic with night sweats and weight loss. Radiologic findings are commonly dense nodules, sometimes with cavitation, but a variety of pulmonary infiltrates may be seen (Fig. 148-1). Dissemination to the CNS occurs in 25% of cases and is frequently asymptomatic. Brain MRI is recommended in all patients with *Nocardia* to rule out brain abscess. Antibiotic susceptibility can be predicted based on species identification, but this may take time. All *Nocardia* species are susceptible to linezolid, so this antibiotic may be administered empirically until identification and susceptibilities are obtained. Most isolates are also susceptible to TMP-SMX, amikacin, and imipenem or meropenem. Combination of two or more antibiotics and long duration of treatment (usually 6–12 months) is recommended by many experts, but comparative trials are not available.

PREVENTION OF INFECTIONS IN TRANSPLANT RECIPIENTS

There are three different and complementary approaches to prevent infections after transplant: minimizing exposures (lifestyle modifications), immunizations, and chemoprophylaxis. Hand hygiene and avoiding sick contacts are some of the most impactful ways to avoid infection. Other suggested lifestyle modifications include recommendations for food and water safety, specifically making sure the drinking water is safe, avoiding raw or poorly cooked eggs and meat, and peeling or carefully washing fruits and vegetables. Hobbies and pets are important components of healthy rehabilitation after transplant, but some specific recommendations can make these safer and reduce the risks of infection. Examples of these are to avoid landscaping, gardening, and

TABLE 148-3 Prophylactic Regimens Commonly Used to Decrease Risk of Infection in Transplant Recipients

RISK FACTOR	ORGANISM	PROPHYLACTIC DRUG	EXAMINATION(S)[a]
Neutropenia, mucositis	*Candida*	Fluconazole	Candida is part of the normal GI flora; everyone with disruption of the mucosal integrity is at risk
Prolonged neutropenia, high-dose corticosteroids, enhanced immunosuppression	*Aspergillus* and other molds	Posaconazole	Blood tests (GMN, β-D-glucan) are less sensitive if triazoles are being administered and are always negative in mucormycosis
Travel to or residence in an area with known risk of endemic fungal infection	*Histoplasma*, *Blastomyces*, *Coccidioides*, *Talaromyces marneffei*	Triazoles considered in context of clinical and laboratory assessment	Chest imaging, antigen testing, serology
Chronic hepatitis B	HBV	Entecavir	HBV serology, HBV DNA
Chronic hepatitis C	HCV	Prophylaxis not used; treat to achieve persistent virological response	HCV serology, HCV RNA
Latent herpesviruses	HSV, VZV, significantly lower efficacy for CMV, EBV	Acyclovir or valacyclovir	Serologic tests for HSV, VZV, CMV, HHV-6, EBV, KSHV (HHV-8); PCR
	CMV	Letermovir (more experience in HCT) or valganciclovir (active against HSV and VZV)	
Exposure (unknown reservoir) or colonization	*Pneumocystis jirovecii*	Trimethoprim-sulfamethoxazole (TMP-SMX) Second line: dapsone, pentamidine, atovaquone	
Parasites	*Toxoplasma gondii* *Strongyloides stercoralis*	TMP-SMX or atovaquone for toxoplasma Ivermectin for *Strongyloides*	Serologic test for *Toxoplasma* and *Strongyloides*
LTBI	*Mycobacterium tuberculosis*	Isoniazid or rifampin in patients with recent seroconversion, positive chest imaging, or recent known exposure and no previous treatment, once active TB is ruled out	TST and/or IGRA; if indeterminate, clinical assessment of exposure and risk

[a] Serologic examination, TST, and IGRAs may be less reliable after transplantation.

Abbreviations: CMV, cytomegalovirus; EBV, Epstein-Barr virus; GI, gastrointestinal; GMN, galactomannan; HBV, hepatitis B virus; HCT, hematopoietic stem cell transplantation; HCV, hepatitis C virus; HHV, human herpesvirus; HSV, herpes simplex virus; IGRA, interferon γ release assay; KSHV, Kaposi's sarcoma–associated herpesvirus; LTBI, latent TB infection; PCR, polymerase chain reaction; TB, tuberculosis; TST, tuberculin skin test; VZV, varicella-zoster virus.

construction sites or to wear protective gear if this is not possible. Using a mask in encounters with high risk of aerosol transmission is recommended. There is a general recommendation against getting new pets for the first 6–12 months after transplant. Avoiding tick and mosquito bites is particularly important. Different professional societies provide advice to transplant recipients. In the United States, the American Society of Transplantation (AST) (*myast.org*) and the Centers for Disease Control and Prevention (*cdc.gov*) are useful resources. Available and required immunizations vary by country, and recommendations are modified as new vaccines and new data become available. Articles published in 2023 are provided in the "Further Reading" section, but updated information may be found in the AST website, and the CDC also presents the guidelines of the Advisory Committee on Immunization Practices. Specific prophylaxes against many infections have been presented in the text; a summary is presented in **Table 148-3**.

■ FURTHER READING

Amengual JE, Pro B: How I treat posttransplant lymphoproliferative disorder. Blood 142:1426, 2023.

Dadwal SS et al: American Society of Transplantation and Cellular Therapy Series, 2: Management and prevention of aspergillosis in hematopoietic cell transplantation recipients. Transplant Cell Ther 27:201, 2021.

Fishman JA: Infection in organ transplantation. Am J Transplant 17:856, 2017.

Hakki M et al: American Society for Transplantation and Cellular Therapy Series, 3: Prevention of cytomegalovirus infection and disease after hematopoietic cell transplantation. Transplant Cell Ther 27:707, 2021.

Kaul DR et al: Ten years of donor-derived disease: A report of the disease transmission advisory committee. Am J Transplant 21:689, 2021.

Reynolds G et al: Vaccine schedule recommendations and updates for patients with hematologic malignancy post-hematopoietic cell transplant or CAR T-cell therapy. Transpl Infect Dis 25(suppl 1):e14109, 2023.

Stewart AG, Kotton CN: What's new: Updates on cytomegalovirus in solid organ transplantation. Transplantation 108:884, 2024.

Timsit JF et al: Diagnostic and therapeutic approach to infectious diseases in solid organ transplant recipients. Intensive Care Med 45:573, 2019.

Viganò M et al: Vaccination recommendations in solid organ transplant adult candidates and recipients. Vaccines (Basel) 11:1611, 2023.

Wolfe CR et al: Donor-derived guidelines: Guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. Clin Transplant 33:e13547, 2019.

Section 4 Therapy for Bacterial Diseases

149

Treatment and Prophylaxis of Bacterial Infections

David C. Hooper, Erica S. Shenoy,
Alyssa R. Letourneau, Ramy H. Elshaboury

Antimicrobial agents have had a major impact on human health. Together with vaccines, they have contributed to reduced mortality, extended life span, and enhanced quality of life. Among drugs used in human medicine, however, they are distinctive in that their use promotes the occurrence of drug resistance in the pathogens they are designed to treat as well as in other "bystander" organisms. Indeed, the history of antimicrobial development has been driven in large part by the medical need engendered by the emergence of resistance to each generation of agents. Thus, the careful and appropriate use of antimicrobial drugs is particularly important not only for optimizing efficacy and minimizing adverse effects but also for minimizing the risk of resistance and preserving the value of existing agents. Although this chapter focuses on antibacterial agents, the optimal use of all antimicrobials depends on an understanding of each drug's mechanism of action, spectrum of activity, mechanisms of resistance, pharmacology, and adverse effect profile. This information is applied in the context of the patient's clinical presentation, underlying conditions, and epidemiology to define the site and likely nature of the infection or other condition and thus to choose the best therapy. Gathering of microbiologic information is especially important for refining therapeutic choices based on the documented pathogen and susceptibility data whenever possible; this information also makes it possible to choose more targeted therapy, thereby reducing the risk of selection of resistant bacteria that can occur with use of agents with a broader spectrum of activity than needed for the patient. Durations of therapy are chosen according to the nature of the infection and the patient's response to treatment and are informed by clinical studies when they are available, with the understanding that shorter courses are less likely than longer courses to promote the emergence of resistance. This chapter and the one that follows provide specific information that is necessary for making informed choices among antibacterial agents. The mechanisms of action of antibacterial agents are discussed in detail in the text of this chapter, and mechanisms of resistance are discussed in detail in **Chap. 150**. Both types of mechanisms, which are related to each other, are summarized for the most commonly used groups of agents in **Table 150-1**. A schematic of antibacterial targets is provided in **Fig. 150-1**.

MECHANISMS OF ACTION (SEE TABLE 150-1)

Multiple essential components of bacterial cell structures and metabolism have been the targets of antibacterial agents used in clinical medicine, and the interaction of an agent with its target results in either inhibition of bacterial growth and replication (*bacteriostatic effect*) or bacterial killing (*bactericidal effect*). In general, targets have been chosen because they either do not exist in mammalian cells and physiology or are sufficiently different from their mammalian counterparts to allow selective bacterial targeting. Treatment with bacteriostatic agents is effective when the patient's host defenses are sufficient to contribute to eradication or sufficient reduction of the infecting pathogen. In patients with impaired host defenses (e.g., neutropenia) or infections at body sites with impaired or limited host defenses (e.g., meningitis and endocarditis), bactericidal agents are generally preferred.

■ INHIBITION OF CELL WALL SYNTHESIS

The bacterial cell wall, which is external to the cytoplasmic membrane and has no counterpart in mammalian cells, protects bacterial cells from lysis under low osmotic conditions. The cell wall is a cross-linked peptidoglycan composed of a polymer of alternating units of *N*-acetylglucosamine (NAG) and *N*-acetylmuramic acid (NAM), four-amino-acid stem peptides linked to each NAM, and a peptide cross-bridge that links adjacent stem peptides to form a net-like structure. Several steps in peptidoglycan synthesis are targets of antibacterial agents. Inhibition of cell-wall synthesis generally results in a bactericidal effect that is linked to cell lysis. This effect results not only from the blocking of new cell-wall formation but from the uninhibited action of cell wall–remodeling enzymes called *autolysins*, which cleave peptidoglycan as part of normal cell-wall growth and cell division.

In gram-positive bacteria, the peptidoglycan is the most external cell structure, but in gram-negative bacteria, an asymmetric lipid outer membrane is external to the peptidoglycan and contains diffusion channels called *porins*. The space between the outer membrane and the peptidoglycan and cytoplasmic membrane is referred to as the *periplasmic space*. Most antibacterial drugs enter the gram-negative bacterial cell through a porin channel, since the outer membrane is a major diffusion barrier. Although the peptidoglycan layer is thicker in gram-positive (20–80 nm) than in gram-negative (1 nm) bacteria, peptidoglycan itself constitutes only a limited diffusion barrier for antibacterial agents.

β-Lactams The β-lactam drugs, including penicillins, cephalosporins, monobactams, and carbapenems, target transpeptidase enzymes (also called *penicillin-binding proteins* [PBPs]) involved in the stem-peptide cross-linking step. Inhibitors of β-lactamases—bacterial enzymes that can degrade β-lactams—are used in combination with some β-lactams to expand their spectrum of activity.

Glycopeptides and Lipoglycopeptides The glycopeptides, including vancomycin and teicoplanin, and the lipoglycopeptides, including telavancin, dalbavancin, and oritavancin, bind the two terminal D-alanine residues of the stem peptide, hindering the glycosyltransferase involved in polymerizing NAG–NAM units as well as transpeptidases. Vancomycin also binds to the lipid II intermediate that delivers cell-wall precursor subunits. The additional binding of teicoplanin, telavancin, dalbavancin, and oritavancin to the bacterial cytoplasmic membrane contributes to their increased potency. Both β-lactams and glycopeptides interact with their targets external to the cytoplasmic membrane.

Bacitracin (Topical) and Fosfomycin These agents interrupt enzymatic steps in the production of peptidoglycan precursors in the cytoplasm.

INHIBITION OF PROTEIN SYNTHESIS

Most inhibitors of bacterial protein synthesis target bacterial ribosomes, whose differences from eukaryotic ribosomes allow selective antibacterial action. Some inhibitors bind to the 30S ribosomal subunit and others to the 50S subunit. Most protein synthesis–inhibiting agents are bacteriostatic; aminoglycosides are an exception and are bactericidal.

Aminoglycosides Aminoglycosides (amikacin, gentamicin, kanamycin, netilmicin, streptomycin, tobramycin, and plazomicin) bind irreversibly to 16S ribosomal RNA (rRNA) of the 30S ribosomal subunit, blocking the translocation of peptidyl transfer RNA (tRNA) from the A (aminoacyl) to the P (peptidyl) site and, at low concentrations, causing misreading of messenger RNA (mRNA) codons, and thus cause the introduction of incorrect amino acids into the peptide chain; at higher concentrations, translocation of the peptide chain is blocked. Cellular uptake of aminoglycosides is dependent on the electrochemical gradient across the bacterial membrane. Under anaerobic conditions, this gradient is reduced, with a consequent reduction in the uptake and activity of the aminoglycosides. Spectinomycin is a related aminocyclitol antibiotic that also binds to 16S rRNA of the 30S ribosomal subunit but at a different site. This drug inhibits translocation of the growing peptide chain but does not trigger codon misreading and produces only a bacteriostatic effect.

Tetracyclines Tetracyclines (doxycycline, minocycline, tetracycline) bind reversibly to the 16S rRNA of the 30S ribosomal subunit and block the binding of aminoacyl tRNA to the ribosomal A site, thereby inhibiting peptide elongation. Active transport of tetracyclines into bacterial but not mammalian cells contributes to the selectivity of these agents. Tigecycline, a derivative of minocycline and the only available glycylcycline, acts similarly to the tetracyclines but is distinctive for its ability to circumvent the most common mechanisms of resistance to the tetracyclines. Other new tetracycline derivatives—eravacycline, a fluorocycline, and omadacycline, an aminomethylcycline—like tigecycline are notable for being little affected by prior common tetracycline resistance mechanisms.

Macrolides and Ketolides In contrast to the aminoglycosides and tetracyclines, the macrolides (azithromycin, clarithromycin, and erythromycin) and ketolides (telithromycin) bind to the 23S rRNA of the 50S ribosomal subunit. These agents block translocation of the growing peptide chain by binding to the tunnel from which the chain exits the ribosome.

Lincosamides Clindamycin is the only lincosamide in clinical use. It binds to the 23S rRNA of the 50S ribosomal subunit, interacting with both the ribosomal A and P sites and blocking peptide bond formation.

Streptogramins The only streptogramin in clinical use is a combination of quinupristin, a group B streptogramin, and dalfopristin, a group A streptogramin. Both components bind to 23S rRNA of the 50S ribosome: dalfopristin binds to both the A and P sites of the peptidyl transferase center, and quinupristin binds to a site that overlaps the macrolide-binding site, blocking the emergence of nascent peptide from the ribosome. The combination is bactericidal, but macrolide-resistant bacteria exhibit cross-resistance to quinupristin, and the remaining activity of dalfopristin alone is only bacteriostatic.

Chloramphenicol Chloramphenicol binds reversibly to the 23S rRNA of the 50S subunit in a manner that interferes with the proper positioning of the aminoacyl component of tRNA in the A site. This site of binding is near those of the macrolides and lincosamides.

Oxazolidinones Linezolid and tedizolid are the only oxazolidinones in clinical use. They bind directly to the A site in the 23S rRNA of the 50S ribosomal subunit and block binding of aminoacyl tRNA, inhibiting the initiation of protein synthesis.

Pleuromutilins Lefamulin is the only systemic pleuromutilin in clinical use. It binds to the peptidyl transferase center of the 50S ribosomal subunit and prevents the correct positioning of tRNAs, thereby inhibiting peptide bond formation and protein synthesis.

Mupirocin Mupirocin (pseudomonic acid) is used topically. It competes with isoleucine for binding to isoleucyl tRNA synthetase, depleting stores of isoleucyl tRNA and thereby inhibiting protein synthesis.

INHIBITION OF BACTERIAL METABOLISM

Available inhibitors (antimetabolites) target the pathway for synthesis of folate, which is a cofactor in a number of one-carbon transfer reactions involved in the synthesis of some nucleic acids, including the pyrimidine thymidine and all purines (adenine and guanine), as well as some amino acids (methionine and serine) and acetyl coenzyme A. Two sequential steps in folate synthesis are targeted. The selective antibacterial effect stems from the inability of mammalian cells to synthesize folate; they depend instead on exogenous sources. Antibacterial activity, however, may be reduced in the presence of high exogenous concentrations of the end products of the folate pathway (e.g., thymidine and purines) that may occur in some infections, resulting from local breakdown of leukocytes and host tissues.

Sulfonamides Sulfonamides, including sulfadiazine, sulfisoxazole, and sulfamethoxazole, inhibit dihydropteroate synthetase (DHPS), which adds *p*-aminobenzoic acid (PABA) to pteridine, producing dihydropteroate. Sulfonamides are structural analogues of PABA and act as competing enzyme substrates.

Trimethoprim Subsequent steps in folate synthesis are catalyzed by dihydrofolate synthase, which adds glutamate to dihydropteroate, and dihydrofolate reductase (DHFR), which then generates the final product, tetrahydrofolate. Trimethoprim is a structural analogue of pteridine and inhibits DHFR. Trimethoprim is available alone but is most often used in combination products that also contain sulfamethoxazole and thus block two sequential steps in folate synthesis.

INHIBITION OF DNA AND RNA SYNTHESIS OR ACTIVITY

A variety of antibacterial agents act on these processes.

Quinolones The quinolones include nalidixic acid, the first agent in the class, and newer, more widely used fluorinated derivatives (fluoroquinolones), including norfloxacin, ciprofloxacin, levofloxacin, moxifloxacin, gemifloxacin, and delafloxacin. The quinolones are synthetic compounds that inhibit bacterial DNA synthesis by interacting with the DNA complexes of two essential enzymes, DNA gyrase and DNA topoisomerase IV, which alter DNA topology. Quinolones trap enzyme–DNA complexes in such a way that they block movement of the DNA replication apparatus and can generate lethal double-strand breaks in DNA, resulting in bactericidal activity. Although mammalian cells also have type II DNA topoisomerases related to gyrase and topoisomerase IV, the structures of the mammalian enzymes are sufficiently different from those of the bacterial enzymes that quinolones have substantially selective antibacterial activity.

Rifamycins Rifampin, rifabutin, and rifapentine are semisynthetic derivatives of rifamycin B and bind the β subunit of bacterial RNA polymerase, thereby blocking elongation of mRNA. Their action is highly selective for the bacterial enzyme over mammalian RNA polymerases.

Nitrofurantoin The reduction of nitrofurantoin, a nitrofuran compound, by bacterial enzymes produces highly reactive derivatives that are thought to cause DNA strand breakage. Nitrofurantoin is used only for the treatment of lower urinary tract infections.

Metronidazole Metronidazole is a synthetic nitroimidazole with activity limited to anaerobic bacteria and certain anaerobic protozoa. Reduction of its nitro group by the electron-transport system in anaerobic bacteria produces reactive intermediates that damage DNA and result in bactericidal activity. Both nitrofurantoin and metronidazole have selective antibacterial activity because the reducing activity needed to produce active derivatives is generated only by bacterial and not mammalian enzymes.

■ DISRUPTION OF MEMBRANE INTEGRITY

The integrity of the bacterial cytoplasmic membrane—and, in gram-negative bacteria, the outer membrane—is important for bacterial viability. Two bactericidal drugs have membrane targets.

Polymyxins The polymyxins, including polymyxin B and polymyxin E (colistin), are cationic cyclic polypeptides that disrupt the cytoplasmic membrane and the outer membrane (the latter by binding lipopolysaccharides, which are negatively charged).

Daptomycin Daptomycin is a lipopeptide that binds the cytoplasmic membrane of gram-positive bacteria in the presence of calcium, generating a channel that leads to leakage of cytoplasmic potassium ions and membrane depolarization.

PHARMACOKINETICS AND PHARMACODYNAMICS

The term *pharmacokinetics* describes the disposition of a drug in the human body, whereas *pharmacodynamics* describes the drug action on the pathogen in relation to pharmacokinetic factors. An understanding of the principles governing these two areas is required for effective drug selection, dosing, and prevention of toxicities.

■ PHARMACOKINETICS

The process of drug disposition consists of four principal phases: absorption, distribution, metabolism, and excretion. These phases determine the time course of drug concentrations in serum, tissues, and body fluids.

Absorption When a drug is administered, *absorption* is defined as the percentage of the dose that reaches the vasculature. The fraction of a drug, however, that reaches the systemic circulation or the pharmacologic site of action is termed *bioavailability*. The bioavailability is more relevant when non-IV routes are used—e.g., the oral, IM, SC, and topical routes. For example, since IV administration provides direct access to the systemic circulation, the bioavailability is therefore 100%. IV and oral dosing for highly bioavailable agents can result in equivalent systemic and tissue concentrations; examples of such agents include metronidazole, fluoroquinolones, tetracyclines, and linezolid. Additionally, many factors can influence oral bioavailability, including the timing of food consumption relative to drug administration, drug-metabolizing enzymes, efflux transporters, concentration-dependent solubility, and acid degradation. Underlying conditions such as diarrhea or ileus can also affect the site of drug absorption and thereby alter its bioavailability. Certain orally administered drugs may have lower bioavailability because of the *first-pass effect*—the process by which drugs are absorbed in the small intestine through the portal circulation and directly transported to the liver for metabolism before reaching their intended site of action.

Distribution *Distribution* describes the process of drug transfer reversibly between the general circulation and body tissues and fluids. After absorption into the systemic circulation and the central compartment (the extensively perfused organs), the drug also distributes into the peripheral compartment (less well-perfused tissues). The volume of distribution (Vd) is a pharmacokinetic parameter that describes the amount of drug in the body at a given time relative to the measured serum concentration. Properties such as the drug's lipophilicity, partition coefficient within different body tissues, protein binding, blood flow, penetration of the blood-brain barrier, and pH can affect the Vd and subsequently the concentration in various tissues. Drugs with a small Vd are limited to certain areas within the body (typically extracellular fluid), whereas those with a higher Vd penetrate extensively into tissues and organs. Some antibacterial drugs can bind to serum proteins, resulting in typically lower Vd as only the unbound (free) fraction of the drug distributes into body tissues and fluids. Furthermore, only the unbound fraction of the drug is considered therapeutically active and available to exert antibacterial effects.

Metabolism *Metabolism* is the chemical transformation of a drug by the body. This modification can occur within several areas; the liver is the organ most commonly involved. Drugs are metabolized by enzymes, but enzyme systems have a finite capacity to metabolize a substrate. If a drug is given in a dose at which the concentration does not exceed the rate of metabolism, the metabolic process is generally linear. If the dose exceeds the amount that can be metabolized, drug accumulation and potential toxicity may occur over time. Drugs are metabolized through phase I or phase II reactions. In phase I reactions, the drug is made more polar through dealkylation, hydroxylation, oxidation, and deamination. Polarity increases water solubility and facilitates removal from the body (e.g., renal elimination). Phase II reactions, which include glucuronidation, sulfation, and acetylation, result in larger and more polar compounds than the parent drug. Both phases usually inactivate the parent drug, although some drugs are rendered more active. The hepatic cytochrome P450 (CYP) enzyme system is mostly responsible for phase I reactions. CYP3A4 is a common subfamily within this system that is responsible for the majority of phase I metabolism. Antibacterial drugs can be substrates, inhibitors, or inducers of a particular CYP enzyme. Inducers, e.g., rifampin, can increase the production of CYP enzymes and consequently increase the metabolism of other drugs. Inhibitors, such as macrolides, cause a decrease in enzyme activity and therefore an increase in the concentration of the interacting drug by decreasing the rate of its metabolism.

Excretion *Excretion* describes the body's mechanisms of drug elimination. Drugs can be eliminated through more than one mechanism. Renal clearance is the most common route and includes elimination through glomerular filtration, tubular secretion, and/or passive diffusion. Some agents undergo nonrenal clearance and rely on the biliary tract or the intestine for excretion. Rate of excretion affects the half-life of a drug—i.e., defined as the time it takes for the blood concentration of a drug to decrease by one-half. This value can range from minutes to days. Half-life and overall drug clearance time can be extended if the organ responsible for clearance is impaired. For example, patients with renal or hepatic impairment may require dose adjustments that take delayed clearance into account to prevent drug accumulation and toxicity. For example, the majority of β-lactam agents are cleared predominantly through glomerular filtration, and in the presence of renal impairment, the dosing interval is typically increased to account for the increased half-life.

PHARMACODYNAMICS

The term *pharmacodynamics* describes the relationship between the drug concentrations that determine its efficacy and those that may produce toxic effects. For an antibacterial agent, the pharmacodynamic focus is the level of drug exposure needed for optimal antibacterial effect in relation to the minimal inhibitory concentration (MIC)—the lowest drug concentration that inhibits the growth of a microorganism under standardized laboratory conditions. Antibacterial effect usually correlates with one or more of the following parameters: (1) concentration-dependent killing (defined as the ratio of peak drug concentration to the MIC), (2) time-dependent killing (defined as duration of drug concentrations above the MIC), or (3) the area under the concentration–time curve to the MIC (AUC/MIC), a measure of the overall drug exposure during the dosing interval **(Fig. 149-1)**.

For *concentration-dependent* killing agents, as the designation implies, the higher the drug peak concentration (C_{max}) at the site of action, the higher the rate and extent of bacterial killing. Aminoglycosides fit into the C_{max}/MIC model of pharmacodynamics activity, and a particular measured peak serum concentration is often targeted to achieve optimal killing. In contrast, *time-dependent* killing agents reach a ceiling at which higher concentrations do not result in increased effect(s). Rather, these agents are active against bacteria when the drug concentration exceeds the MIC for a desired period of time. The T > MIC predicts clinical efficacy for all β-lactams. The longer the concentration of the β-lactam remains above the MIC for an infecting pathogen during the dosing interval, the greater the killing effect. Fluoroquinolones and vancomycin exemplify agents for which the AUC/MIC is a predictor of efficacy. For example, studies have found that an AUC/MIC ratio of >30 will maximize killing of *Streptococcus pneumoniae* by fluoroquinolones, whereas AUC/MIC ratios >125 are required to exert their optimal effects against gram-negative pathogens. Finally, for some antibacterial drugs such as aminoglycosides, a *postantibiotic effect*—the delayed regrowth of surviving bacteria after exposure to an antibiotic—supports less frequent dosing, increased drug-free intervals, and likely decreased drug-related toxicities.

APPROACH TO THERAPY

The approach to antibiotic therapy is driven by host factors, site of infection, and local resistance profiles of suspected or known pathogens. Further, national and local drug shortages and formulary restrictions can affect available therapies. Regular monitoring of the patient and collection of laboratory data should be undertaken to streamline antibacterial therapy as appropriate and to investigate the possibility of treatment failure if the patient fails to respond appropriately.

EMPIRICAL AND DIRECTED THERAPY

Therapy is considered *empirical* when the causative agent has yet to be determined and therapeutic decisions are based on the severity of illness, the clinician's assessment of likely pathogens in light of the clinical syndrome, the patient's medical conditions and prior therapy, and relevant epidemiologic factors. For patients with severe illness, empirical therapy often takes the form of an antibacterial combination that provides broad coverage of diverse agents and thus ensures adequate treatment of possible pathogens while additional data are being collected. *Directed* therapy is predicated on identification of the pathogen, determination of its susceptibility profile, and establishment of the extent of the infection. Directed therapy generally allows the use of more targeted and narrower-spectrum antibacterial agents than does empirical therapy.

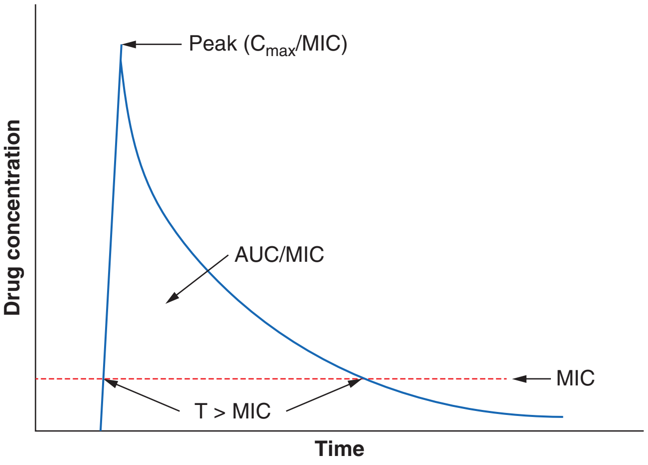

FIGURE 149-1 Pharmacokinetic and pharmacodynamic model predicting efficacy of antibacterial drugs. AUC, area under the time–concentration curve; C_{max}, peak serum concentration of drug; MIC, minimal inhibitory concentration; T > MIC, duration of drug concentrations above the MIC.

Information on epidemiology, exposures, and local antibacterial susceptibility patterns can help guide empirical therapy. When empirical treatment is clinically appropriate, care should be taken to obtain clinical specimens for microbiologic analysis before the initiation of therapy and to adjust therapy as new information is obtained about the patient's clinical condition and the causal pathogens. Change to directed therapy can limit unnecessary risks of drug side effects as well as selection for antibacterial resistance and the risk of *Clostridioides difficile* infection and disease.

SITE OF INFECTION

The site of infection is a consideration in antibacterial therapy, largely because of the differing abilities of drugs to penetrate and achieve adequate concentrations at particular body sites. For example, to be effective in the treatment of meningitis, an agent must (1) be able to cross the blood-brain barrier and reach adequate concentrations in the cerebrospinal fluid (CSF) and (2) be active against the relevant pathogen(s). Dexamethasone, administered with or 15–20 min before the first dose of an antibacterial drug, has been shown to improve outcomes in patients with some types of acute bacterial meningitis, but its use may reduce penetration of some antibacterial agents, such as vancomycin, into the CSF. In this case, rifampin is added because its penetration is not reduced by dexamethasone. Infections at sites where pathogens are protected from normal host defenses, penetration of an antibacterial drug is limited, or local conditions (e.g., low pH) limit activity of some agents include, in addition to meningitis, osteomyelitis, prostatitis, intraocular infections, and abscesses. In such cases, consideration must be given to the route of drug delivery (e.g., intravitreal injections) as well as to interventions to drain, debride, or otherwise reduce bacterial load and necrotic material that can reduce antibacterial activity.

HOST FACTORS

Host factors, including immune function, pregnancy, allergies, age, renal and hepatic function, drug–drug interactions, comorbid conditions, and occupational or social exposures, should be considered.

Immune Dysfunction Patients with deficits in immune function that blunt the response to bacterial infection, including neutropenia, deficient humoral immunity, and asplenia (either surgical or functional), are all at increased risk of severe bacterial infection. Such patients should be treated aggressively and often broadly in the early stages of suspected infection pending results of microbiologic tests. For asplenic patients, treatment should include coverage of encapsulated organisms, particularly *S. pneumoniae*, that may cause rapidly life-threatening infection. For neutropenic patients, initial treatment typically includes antibacterial agents with broad activity against gram-negative bacteria.

Pregnancy Pregnancy affects decisions regarding antibacterial therapy in two respects. First, pregnancy is associated with an increased risk of particular infections (e.g., those caused by *Listeria*). Second, the potential risks to the fetus that are posed by specific drugs must be considered. As for other drugs, the safety of the vast majority of antibacterial agents in pregnancy has not been established, and such agents are grouped in categories B and C by the U.S. Food and Drug Administration (FDA). Drugs in categories D and X are contraindicated in pregnancy or lactation due to established risks. Note that in accordance with the Pregnancy and Lactation Labeling Final Rule (PLLR), drugs submitted to the FDA for approval after 2015 do not use the pregnancy risk categories. The risks associated with antibacterial use in pregnancy and during lactation are summarized in **Table 149-1**.

TABLE 149-1 Risks Associated with Use of Antibacterial Drugs in Pregnancy and Lactation

PREGNANCY CATEGORY[a]	ANTIBACTERIAL DRUG	FETAL RISK RECOMMENDATION[b]	BREAST-FEEDING RISK RECOMMENDATION[b]
B	Azithromycin	Limited human data. Animal data suggest low risk.	Limited human data; probably compatible
	Cephalosporins (including cephalexin, cefuroxime, cefixime, cefpodoxime, cefotaxime, ceftriaxone)	Compatible	Compatible
	Ceftazidime-avibactam	No human data; no fetal harm in animal studies	Ceftazidime is excreted into human milk in low concentrations. Avibactam is excreted into the milk of lactating rats; no human studies have been conducted.
	Ceftolozane-tazobactam	Compatible	Unknown
	Clindamycin	Compatible	Compatible
	Ertapenem	No human data; probably compatible	Limited human data; probably compatible
	Erythromycin	Compatible (except for estolate salt)	Compatible
	Meropenem and meropenem-vaborbactam	No human data. Animal data suggest low risk.	No human data; probably compatible
	Metronidazole	Human data suggest low risk.	Interrupt breast-feeding for 12–24 h after single 2-g dose. Limited human data; potential toxicity in divided doses
	Nitrofurantoin	Human data suggest risk in third trimester.	Limited human data; probably compatible. Higher risk associated with younger infants and those with G6PD deficiency
	Penicillins (including amoxicillin, ampicillin, cloxacillin)	Compatible	Compatible
	Quinupristin-dalfopristin	Compatible. Maternal benefit must far outweigh risk to embryo/fetus.	No human data; potential toxicity
	Vancomycin	Compatible	Limited human data; probably compatible
C	Chloramphenicol	Compatible	Limited human data; potential toxicity
	Fluoroquinolones	Human data suggest low risk.	Limited human data; probably compatible
	Clarithromycin	Limited human data. Animal data suggest high risk.	No human data; probably compatible
	Imipenem-cilastatin	Limited human data. Animal data suggest low risk.	Limited human data; probably compatible
	Linezolid	Compatible. Maternal benefit must far outweigh risk to embryo/fetus.	No human data; potential toxicity
	Telavancin	No human data. Animal studies have revealed evidence of teratogenicity.[c]	No human data. Animal studies have revealed evidence of teratogenicity[c]
	Tedizolid	Limited data. Embryo-fetal studies in mice, rats, and rabbits have demonstrated fetal developmental toxicities. Use only if benefit outweighs risk.	Excreted in the breast milk of rats; unknown in humans; caution use
	Dalbavancin	Limited human data. At high doses in animal studies, delayed fetal maturation, increased embryo and offspring death. Use only if benefit outweighs risk.	Excreted in the breast milk of animals; unknown in humans; caution use
	Oritavancin	Limited human data. Studies in rats and rabbits demonstrated no harm at 25% of recommended human dose. Use only if benefit outweighs risk.	Excreted in the breast milk of rats; unknown in humans; caution use
C/D	Amikacin	Human data suggest low risk.	Compatible
	Gentamicin	Human data suggest low risk.	Compatible
D	Kanamycin	Human data suggest risk.	Limited human data; probably compatible
	Streptomycin	Human data suggest risk.	Compatible
	Sulfonamides	Human data suggest risk in third trimester.	Limited human data; potential toxicity. Avoid in ill, stressed, premature infants and in infants with hyperbilirubinemia or G6PD deficiency.
	Tetracyclines	Contraindicated in second and third trimesters	Compatible
	Tigecycline	Human data suggest risk in second and third trimesters.	No human data; potential toxicity
Not assigned[d]	Cefiderocol	No controlled data in human pregnancy; animal studies have not provided evidence of fetal harm.	Unknown if excreted in human milk; excreted in animal milk.
	Eravacycline	No controlled data in human pregnancy; animal data indicate drug crosses placenta and is associated with risk at higher doses,	Unknown if excreted in human milk; excreted in animal milk. Not recommended during and for a period after treatment.
	Imipenem-cilastatin-relebactam	No controlled data in human pregnancy; animal studies have not revealed teratogenicity but have shown evidence of increased fetal loss.	Imipenem and cilastatin are excreted into human milk; no data on relebactam in human milk. Relebactam is excreted in animal milk. No human data regarding potential effect on infant.
	Lefamulin[e]	No controlled data in human pregnancy; animal studies have revealed evidence of fetal harm.	Unknown if excreted in human milk; excreted in animal milk. Breast-feeding is not recommended during use and for 2 days afterward.

(*Continued*)

TABLE 149-1 Risks Associated with Use of Antibacterial Drugs in Pregnancy and Lactation (Continued)

PREGNANCY CATEGORY[a]	ANTIBACTERIAL DRUG	FETAL RISK RECOMMENDATION[b]	BREAST-FEEDING RISK RECOMMENDATION[b]
	Meropenem-vaborbactam	No controlled data in human pregnancy; animal studies have revealed evidence of fetal harm (related to vaborbactam component).	Meropenem is excreted in human milk; it is unknown if vaborbactam is excreted in human milk. Data on excretion of vaborbactam in animal milk are unknown.
	Omadacycline	No controlled data in human pregnancy; however, as this is a tetracycline class antibiotic, may cause deciduous tooth discoloration and bone growth inhibition in second and third trimester of pregnancy; animal data have demonstrated embryofetal lethality, teratogenicity, and embryofetal toxicity.	Unknown if excreted in human milk; data not available regarding excretion in animal milk. Not recommended during and for a period after treatment.
	Plazomicin	No controlled data in human pregnancy; however, aminoglycoside antibiotics are known to cause fetal harm in pregnancy.	Unknown if excreted in human milk; excreted in animal milk.

[a] *Category B: Either* animal reproduction studies have failed to demonstrate a risk to the fetus, and there are no adequate and well-controlled studies in pregnant women; *or* animal studies have shown an adverse effect, but adequate and well-controlled studies in pregnant women have failed to demonstrate a risk to the fetus in any trimester. *Category C:* Animal reproduction studies have shown an adverse effect on the fetus, and there are no adequate and well-controlled studies in humans, but potential benefits may warrant use of the drug in pregnant women despite potential risks. *Category D:* There is positive evidence of human fetal risk based on adverse-reaction data from investigational or marketing experience or studies in humans, but potential benefits may warrant use of the drug in pregnant women despite potential risks. [b] Fetal risk recommendation and breast-feeding risk recommendation adapted from GG Briggs et al (eds): *Drugs in Pregnancy and Lactation*, 9th ed. Philadelphia, Lippincott Williams and Wilkins, 2011; and the U.S. Food and Drug Administration (Drugs@FDA). [c] A registry has been established to monitor pregnancy outcomes of pregnant women exposed to telavancin. Physicians are encouraged to register pregnant patients, or pregnant women may enroll themselves by calling 1-855-633-8479. [d] The U.S. Food and Drug Administration is phasing out use of pregnancy categories A, B, C, D, and X. [e] A pregnancy pharmacovigilance program is available: If this drug is inadvertently administered during pregnancy or if a patient becomes pregnant while receiving this drug, healthcare providers or patients should report drug exposure by calling 1-855-5NABRIVA (1-855-56227482) to enroll.

Abbreviation: G6PD, glucose-6-phosphate dehydrogenase.

Allergies Allergies to antibiotics are among the most common allergies reported, and an allergy history should be obtained whenever possible before therapy is chosen. A detailed allergy history can shed light on the type of reaction experienced previously and on whether rechallenge with the same or a related medication is advisable (and, if so, under what circumstances). Allergies to the penicillins are most common. Although as many as 10% of patients may report an allergy to penicillin, studies suggest that >90% of these patients could tolerate a penicillin or cephalosporin. Adverse effects and intolerances (Table 149-2) should be distinguished from true allergies to ensure appropriate selection of antibacterial therapy.

Drug–Drug Interactions Patients commonly receive other drugs that may interact with antibacterial agents. A summary of the most common drug–drug interactions, by antibacterial class, is provided in Table 149-3.

Exposures Exposures, both occupational and nonoccupational, may provide clues to likely pathogens. When relevant, inquiries about exposure to ill contacts, animals, insects, and water should be included in the history, along with sites of residence and travel within a relevant time frame.

Other Host Factors Age, renal and hepatic function, and comorbid conditions are all considerations in the choice of, and schedule for, therapy. Dose adjustments should be made accordingly. In patients with decreased or unreliable oral absorption, intravenous therapy may be preferred to ensure adequate blood levels of drug and delivery of the antibacterial agent to the site of infection. In general, initial treatment for severe and life-threatening infections is given by intravenous injection to assure prompt and adequate drug delivery.

DURATION OF THERAPY

Whether empirical or directed, the duration of therapy should be established in most clinical situations. Guidelines that synthesize available literature and expert opinion provide recommendations on therapy duration that are based on infecting organism, organ system, and patient factors. For example, the American Heart Association has published guidelines endorsed by the Infectious Diseases Society of America (IDSA) on diagnosis, antibacterial therapy, and management of complications of infective endocarditis. Additional guidelines from the IDSA in collaboration with other specialty societies, such as the Society of Healthcare Epidemiology of America, the American Thoracic Society, and the European Society for Clinical Microbiology and Infectious Diseases, exist for a range of infectious syndromes and specific pathogens. In general, where data on adequate durations of therapy exist, shorter courses are preferred to reduce the likelihood of drug adverse effects and selection of resistant bacteria.

FAILURE OF THERAPY

If a patient does not respond to therapy, investigations often should include imaging and the collection of additional specimens for microbiologic testing as indicated. Failure to respond can be the result of an antibacterial regimen that does not address the underlying causative organism, the development of resistance during therapy, or the existence of a focus of infection at a site poorly penetrated by systemic therapy. Some infections may also require surgical interventions (e.g., large abscesses, myonecrosis). Fever due to allergic drug reactions may complicate assessment of the patient's response to antibacterial treatment.

EXPERT GUIDANCE

Selected websites with the most up-to-date information and guidance for the clinician include the following:

- Johns Hopkins ABX Guide (*www.hopkins-abxguide.org*)
- IDSA Practice Guidelines (*https://www.idsociety.org/practice-guideline/practice-guidelines*)
- OneHealth Trust Resistance Map (*https://resistancemap.onehealthtrust.org/index.php*)
- Centers for Disease Control and Prevention Antibiotic/Antimicrobial Resistance (*www.cdc.gov/drugresistance/*)

CLINICAL USE OF ANTIBACTERIAL AGENTS

The clinical application of antibacterial therapy is guided by the spectrum of the agent and the suspected or known target pathogen. Infections for which specific antibacterial agents are among the drugs of choice are listed, along with associated pathogens and susceptibility data, in Table 149-4. Resistance rates of specific organisms are dynamic and should be taken into account in the approach to antibacterial therapy. While national resistance rates can serve as a reference, the most useful reference for the clinician is the most recent local laboratory antibiogram, which provides details on local resistance patterns, often on an annual or semiannual basis.

β-LACTAMS

The β-lactam class of antibiotics consists of penicillins, cephalosporins, carbapenems, and monobactams. The term *β-lactam* reflects

TABLE 149-2 Common Adverse Reactions to Antibacterial Agents

ANTIBACTERIAL(S)	POTENTIAL ADVERSE EFFECTS	COMMENTS
β-Lactams	Hypersensitivity reactions	Ranges from rash to anaphylaxis. Cross-reactivity among β-lactams is related to chemical structure and side chain similarity.
	Neurotoxicity	More commonly described with cefepime and imipenem, but likely a class effect. Risk is increased in patients with history of seizures, renal impairment, and advanced age.
	Neutropenia/hematologic reactions	May be related to high doses and prolonged duration
Vancomycin	Nephrotoxicity	Risk increases with vancomycin trough levels >20 μg/mL or concomitant administration with other potentially nephrotoxic agents. The effect is usually reversible.
	"Red man syndrome"	Can be managed with a slower vancomycin infusion and pretreatment with antihistamine
Telavancin	QT prolongation	
	Interference with coagulation tests	May falsely affect INR, PT, aPTT. Perform these tests before the next dose of telavancin (when serum drug levels are at their nadir).
	Taste disturbances	
	Nephrotoxicity	
Oritavancin	Interference with coagulation tests	May falsely affect INR, PT, aPTT. Perform these tests at least 24 h after the dose is administered.
	Gastrointestinal distress	
Dalbavancin	Gastrointestinal distress	
Daptomycin	Myopathy	Monitor CPK levels during therapy. Rhabdomyolysis has been reported but appears to be rare.
	Eosinophilic pneumonia	
Aminoglycosides	Nephrotoxicity	Associated with prolonged use; usually reversible
	Ototoxicity	Can cause both vestibular and cochlear toxicity. Ototoxicity may be irreversible.
Fluoroquinolones	QT_c prolongation	Moxifloxacin appears more likely than other quinolones to exert this effect. Risk of arrhythmia increases when these drugs are given concomitantly with other QT_c-prolonging agents.
	Tendinitis	Risk is greater among the elderly and patients receiving steroids.
	Dysglycemia	
	Exacerbation of myasthenia gravis	
Rifampin	Hepatotoxicity	Risk is greater when drug is given with other antituberculosis agents. When rifampin is given alone, LFT values may be transiently elevated without symptoms.
	Orange discoloration of body fluids	
Tetracyclines, including tigecycline, eravacycline, and omadacycline	Photosensitivity	
	Gastrointestinal distress	High incidence of diarrhea, nausea, vomiting
Macrolides	Gastrointestinal distress	Erythromycin is occasionally used as a therapeutic agent for some gastric motility disorders.
	QT_c prolongation	Azithromycin use is associated with an increased risk of death from cardiovascular causes among patients at high baseline risk.
Metronidazole	Peripheral neuropathy	Associated with prolonged use
Clindamycin	Diarrhea and pseudomembranous colitis	
Linezolid, tedizolid	Myelosuppression	Associated with prolonged use
	Optic and peripheral neuropathy	Associated with prolonged use
	Lactic acidosis	
TMP-SMX	Hypersensitivity reactions	Allergy usually associated with sulfonamide moiety
	Nephrotoxicity	Associated with high doses
	Hematologic effects	Associated with prolonged use
Nitrofurantoin	Pneumonitis and other pulmonary reactions	Associated with prolonged use
	Peripheral neuropathy	Associated with accumulation of nitrofurantoin in renal failure. Avoid use in renal impairment.
Fosfomycin	Gastrointestinal effects	
Polymyxins	Nephrotoxicity	Associated with high dose
	Neurotoxicity	Neuromuscular blockade and muscle weakness are well described and usually reversible.
Quinupristin-dalfopristin	Arthralgias and myalgias	
Chloramphenicol	Bone marrow suppression	Aplastic anemia or hematopoietic toxicity
Pleuromutilin	Gastrointestinal	Diarrhea
	QT_c prolongation	When used in conjunction with CYP3A4 substrates

Note: All systemic antibiotics have the potential to alter abdominal flora and induce *Clostridioides difficile* infection.

Abbreviations: aPTT, activated partial thromboplastin time; CPK, creatine phosphokinase; INR, international normalized ratio; LFT, liver function test; PT, prothrombin time; TMP-SMX, trimethoprim-sulfamethoxazole.

TABLE 149-3 Important Antibacterial Drug Interactions

ANTIBACTERIAL(S)	INTERACTING AGENT(S)	POTENTIAL EFFECT AND MANAGEMENT
Nafcillin	Warfarin, cyclosporine, tacrolimus	Decreased effects of interacting drug via CYP3A4 induction. Monitor levels of affected drug closely if drugs are given concomitantly.
Ceftriaxone	Calcium-containing IV solutions	Concomitant use is contraindicated in neonates (<28 days); the combination can lead to precipitation of ceftriaxone-calcium particulate. Ceftriaxone and calcium-containing solutions can be given to infants >28 days of age provided they are given sequentially and the lines are thoroughly flushed between infusions, or infused via separate lines.
Carbapenems	Valproic acid	Diminished levels of valproic acid. Monitor valproic acid levels closely if drugs are given concomitantly and consider alternative therapies.
Linezolid, tedizolid	Serotonergic and adrenergic agents (e.g., SSRIs, vasopressors)	Increased levels of serotonergic and adrenergic agents. Monitor for serotonin syndrome. Tedizolid may have less potential than linezolid to cause this drug interaction.
Quinupristin-dalfopristin	Substrates of CYP3A4 (e.g., warfarin, ritonavir, cyclosporine, diazepam, verapamil)	Can result in increased levels of interacting drug
Fluoroquinolones	Theophylline[a]	Can result in theophylline toxicity
	Sucralfate; antacids containing aluminum, calcium, or magnesium; ferrous sulfate– and zinc-containing multivitamins	Can result in decreased oral absorption of fluoroquinolones. Administer fluoroquinolone 2 h before or 6 h after interacting drug.
	Tizanidine[a]	Can result in increased levels of tizanidine and hypotensive, sedative effects. Monitor for side effects if drugs are given concomitantly.
	QT_c-prolonging drugs (e.g. azoles, sotalol, amiodarone, dofetilide, fluoxetine)	Increased risk of cardiotoxicity and arrhythmias. Monitor QT_c.
Rifampin	Substrates of CYP3A4 (e.g., warfarin, ritonavir, cyclosporine, diazepam, verapamil, protease inhibitors, voriconazole)	Can result in decreased levels of interacting drug. Avoid concomitant use if possible. If giving drugs concomitantly, monitor drug levels if possible.
	Substrates of CYP2C19 (e.g., omeprazole, lansoprazole)	
	Substrates of CYP2C9 (e.g., warfarin, tolbutamide)	
	Substrates of CYP2C8 (e.g., repaglinide, rosiglitazone)	
	Substrates of CYP2B6 (e.g., efavirenz)	
	Hormone therapy (e.g., norethindrone)	Can result in decreased levels of hormone. If oral contraceptive and rifampin are given concomitantly, use alternative or additional forms of birth control.
Tetracyclines	Antacids or drugs containing calcium, magnesium, iron, or aluminum	Can result in decreased oral absorption of tetracyclines. Administer tetracycline 2 h before or 6 h after interacting drug.
	Warfarin	Increased effect of warfarin. Monitor levels closely if drugs are given concomitantly.
	Eravacycline: Strong CYP3A4 inducers (e.g., rifampin)	Reduced eravacycline efficacy
Macrolides[b]	Substrates of CYP3A4 (e.g., warfarin, ritonavir, cyclosporine, diazepam, verapamil, amiodarone)	Avoid concomitant administration if possible.
	QT_c-prolonging agents (e.g., fluoroquinolones, sotalol)	Increased risk of cardiotoxicity and arrhythmias. Monitor QT_c.
	Protease inhibitors (e.g., ritonavir)	Can result in increased levels of both macrolides and protease inhibitors. Avoid concomitant use if possible.
	Cimetidine	Cimetidine can increase levels of macrolides.
Metronidazole	Ethanol	Can result in disulfiram-like reaction. Ethanol may be present in some formulations of oral drug suspensions (e.g., ritonavir).
	Warfarin	Can increase warfarin effects. Monitor INR closely if drugs are given concomitantly.
TMP-SMX	Warfarin	Increased effect of warfarin. Monitor levels closely if drugs are given concomitantly.
	Phenytoin	Increased levels of phenytoin. Monitor levels closely if drugs are given concomitantly.
	Methotrexate	Increased levels of methotrexate and prolonged exposure. Monitor levels closely if drugs are given concomitantly.
Oritavancin	Substrates of CYP3A4 (e.g., cyclosporine, warfarin) and CYP2D6 (e.g., aripiprazole) Substrates of CYP2C19 (e.g., omeprazole) and CYP2C9 (e.g., warfarin)	Can result in decreased levels of interacting drug. Avoid concomitant use if possible. If giving drugs concomitantly, monitor drug levels if possible.
Lefamulin	QT_c-prolonging drugs (e.g., azoles, sotalol, amiodarone, dofetilide, fluoxetine)	Increased risk of cardiotoxicity and arrhythmias. Monitor QT_c.
	Strong CYP3A4 inducers (e.g., rifampin)	Reduced lefamulin efficacy
	Strong CYP3A4 strong inhibitors (e.g., ritonavir)	Increased lefamulin exposure

[a]Drug reaction described with ciprofloxacin only. [b]Clarithromycin and erythromycin are potent CYP3A4 inhibitors; the probability of a drug interaction with azithromycin is lower.

Abbreviations: INR, international normalized ratio; SSRI, selective serotonin-reuptake inhibitor; TMP-SMX, trimethoprim-sulfamethoxazole.

TABLE 149-4 Drug Indications for Specific Infections, Associated Pathogens, and Sample Susceptibility Rates

ANTIMICROBIAL(S)	INFECTIONS	COMMON PATHOGENS (% SUSCEPTIBLE); RESISTANCE AS NOTED[a]
Penicillin G	Syphilis; yaws; leptospirosis; streptococcal infections; pneumococcal infections; actinomycosis; oral and periodontal infections; meningococcal meningitis and meningococcemia; viridans streptococcal endocarditis; clostridial myonecrosis; tetanus; rat-bite fever; *Pasteurella multocida* infections; erysipeloid (*Erysipelothrix rhusiopathiae*)	*Neisseria meningitidis*; viridans streptococci (60%); *Streptococcus pneumoniae* (97% nonmeningitis; 74% meningitis)
Ampicillin, amoxicillin	Salmonellosis; acute otitis media; *Haemophilus influenzae* meningitis and epiglottitis; *Listeria monocytogenes* meningitis; *Enterococcus faecalis* UTI	*Escherichia coli* (52%); *H. influenzae* (72%); *Salmonella* spp. (89%)
Nafcillin, oxacillin	MSSA bacteremia and endocarditis	*Staphylococcus aureus* (72%); coagulase-negative staphylococci (49%)
Piperacillin-tazobactam	Intraabdominal infections (facultative enteric gram-negative bacilli and obligate anaerobes); infections caused by mixed flora (aspiration pneumonia, diabetic foot ulcers); infections caused by *Pseudomonas aeruginosa*	*P. aeruginosa* (78%)
Cefazolin	*E. coli* UTI; surgical prophylaxis; MSSA bacteremia and endocarditis	*E. coli* (80%)
Cefoxitin, cefotetan	Intraabdominal infections and pelvic inflammatory disease	*Bacteroides fragilis* (60%)[b]
Ceftriaxone	Gonococcal infections; pneumococcal meningitis; viridans streptococcal endocarditis; salmonellosis and typhoid fever; health care–associated infections caused by nonpseudomonal facultative gram-negative enteric bacilli	*S. pneumoniae* (87% meningitis; 99% nonmeningitis); *E. coli* (88%); *Klebsiella pneumoniae* (85%)
Ceftazidime, cefepime	Health care–associated infections caused by facultative gram-negative bacilli and *Pseudomonas* spp.	*P. aeruginosa* (82% for ceftazidime, 85% for cefepime)
Ceftaroline	CAP caused by *S. pneumoniae*, MSSA, *H. influenzae, K. pneumoniae, Klebsiella oxytoca*, and *E. coli*; acute bacterial skin and skin-structure infections caused by MSSA, MRSA, *Streptococcus pyogenes, Streptococcus agalactiae, E. coli, K. pneumoniae*, and *K. oxytoca*	Mostly susceptible; four strains of MRSA with ceftaroline MICs >4 μg/mL reported in isolates from a single Greek hospital[c]; additional case reports, including in patients without prior exposure to ceftaroline[d,e]
Ceftazidime-avibactam, meropenem-vaborbactam	Complicated UTIs (ceftazidime-avibactam and meropenem-vaborbactam) and complicated intraabdominal infections (ceftazidime-avibactam in combination with metronidazole) caused by resistant gram-negative organisms, including *Pseudomonas*, and some anaerobes	*P. aeruginosa* (84–97%)[f] MDR Enterobacterales, including carbapenem-resistant Enterobacterales that produce KPCs No activity against metallo-β-lactamases (e.g., NDM)
Ceftolozane-tazobactam	Complicated UTIs and complicated intraabdominal infections (in combination with metronidazole) caused by resistant gram-negative organisms, including *Pseudomonas*, and some anaerobes	*P. aeruginosa* (>86% overall; 60–80% of ceftazidime- and meropenem-resistant strains)[f] MDR Enterobacterales No activity against KPC-producing organisms
Imipenem, meropenem	Intraabdominal infections, infections caused by *Enterobacter* spp. and ESBL-producing gram-negative bacilli	*P. aeruginosa* (84%); *Acinetobacter calcoaceticus-baumannii* complex (84%) (meropenem susceptibilities reported)
Ertapenem	CAP; complicated UTIs, including pyelonephritis; acute pelvic infections; complicated intraabdominal infections; complicated skin and skin-structure infections, excluding diabetic foot infections accompanied by osteomyelitis or caused by *P. aeruginosa*	*Enterobacter cloacae* (88%); *K. pneumoniae* (98%)
Aztreonam	Infections caused by facultative gram-negative bacilli and *Pseudomonas* in penicillin-allergic patients	
Vancomycin	Bacteremia, endocarditis, and other invasive disease caused by MRSA; pneumococcal meningitis; oral formulation for CDAD	*S. aureus* (100%); *E. faecalis* (93%); *E. faecium* (41%)
Telavancin	Health care– and ventilator-associated pneumonia or skin and soft tissue infections caused by MRSA	*S. aureus*: none reported
Dalbavancin, oritavancin	Complicated skin and soft tissue infections	*S. aureus*: rarely reported for dalbavancin.[g] Rarely reported for oritavancin.[h]
Daptomycin	VRE infections; MRSA bacteremia	*E. faecalis* (99.9%);[i] *E. faecium* (99.7%);[i] *S. aureus* (99.9%)[g]
Gentamicin, tobramycin, amikacin	Combined with penicillin for staphylococcal, enterococcal, or streptococcal endocarditis; combined with β-lactam for gram-negative bacteremia; pyelonephritis	*E. coli* (gentamicin, 90%); *P. aeruginosa* (amikacin, 93%; gentamicin, 89%); *A. calcoaceticus-baumannii* complex (gentamicin, 83%)
Azithromycin, clarithromycin, erythromycin	*Legionella, Campylobacter*, and *Mycoplasma* infections; CAP; GAS pharyngitis in penicillin-allergic patients; bacillary angiomatosis; gastric infections due to *Helicobacter pylori*; MAI infections	*S. pneumoniae* (60%); group A streptococci (82%); *H. pylori* (75%)[j]
Clindamycin	Severe, invasive GAS infections (with a β-lactam); infections caused by obligate anaerobes; infections caused by susceptible staphylococci	*S. aureus* (73%)
Doxycycline, minocycline	Acute bacterial exacerbations of chronic bronchitis; granuloma inguinale; brucellosis (with streptomycin); tularemia; glanders; melioidosis; spirochetal infections caused by *Borrelia* (Lyme disease and relapsing fever; doxycycline); infections caused by *Vibrio vulnificus*; some *Aeromonas* infections; infections due to *Stenotrophomonas* (minocycline); plague; ehrlichiosis; chlamydial infections (doxycycline); granulomatous infections due to *Mycobacterium marinum* (minocycline); rickettsial infections; mild CAP; skin and soft tissue infections caused by gram-positive cocci (e.g., CA-MRSA infections); leptospirosis; syphilis; and actinomycosis in the penicillin-allergic patient	*S. pneumoniae* (54%); *S. aureus* (97%)

(*Continued*)

TABLE 149-4 Drug Indications for Specific Infections, Associated Pathogens, and Sample Susceptibility Rates (Continued)

ANTIMICROBIAL(S)	INFECTIONS	COMMON PATHOGENS (% SUSCEPTIBLE); RESISTANCE AS NOTED[a]
Tigecycline	CAP caused by *S. pneumoniae, H. influenzae*, or *Legionella pneumophila*; complicated skin infections caused by *E. coli*, MRSA, MSSA, *S. pyogenes, Streptococcus anginosus, S. agalactiae, B. fragilis*; complicated intraabdominal infections caused by *E. coli*, vancomycin-susceptible *E. faecalis, Citrobacter freundii, E. cloacae, K. pneumoniae, K. oxytoca, Bacteroides* spp., *Clostridium perfringens*, and *Peptostreptococcus* spp.	Mostly susceptible, although case reports of resistance in *A. baumannii* and *K. pneumoniae*
TMP-SMX	Community-acquired UTI; CA-MRSA skin and soft tissue infections	*E. coli* (73%); *S. aureus* (94%)
Sulfonamides	Nocardial infections; leprosy (dapsone); toxoplasmosis (sulfadiazine)	Unknown
Ciprofloxacin, levofloxacin, moxifloxacin, delafloxacin	CAP (levofloxacin and moxifloxacin); UTI; bacterial gastroenteritis; health care–associated gram-negative enteric infections; *Pseudomonas* infections (ciprofloxacin and levofloxacin); skin and skin-structure infections (delafloxacin)	*S. pneumoniae* (100% levofloxacin); *E. coli* (75% for ciprofloxacin and 70% for levofloxacin); *P. aeruginosa* (78% for ciprofloxacin and 70% for levofloxacin); *Salmonella* spp. (79% for ciprofloxacin and 77% for levofloxacin)
Rifampin	Staphylococcal foreign body infections (in combination with other antistaphylococcal agents); *Legionella* pneumonia; *Mycobacterium tuberculosis*; atypical nontuberculous mycobacterial infection; pneumococcal meningitis when organisms are susceptible or response is delayed	*S. aureus* (99%), although staphylococci rapidly develop resistance with monotherapy
Metronidazole	Obligate anaerobic gram-negative bacteria (e.g., *Bacteroides* spp.); abscess in lung, brain, or abdomen; bacterial vaginosis; CDAD	Mostly susceptible; resistance very rare
Linezolid, tedizolid	VRE; uncomplicated and complicated skin and soft tissue infections caused by MSSA and MRSA; CAP with concurrent bacteremia; health care–associated pneumonia	Mostly susceptible; resistance occasionally seen in VRE
Chloramphenicol	Infections due to gram-positive and gram-negative organisms resistant to standard alternatives (e.g., *Burkholderia*)	Unknown
Colistin	Infections due to gram-negative bacilli resistant to all other chemotherapy (e.g., *P. aeruginosa, Acinetobacter* spp., and *Stenotrophomonas maltophilia*)	*P. aeruginosa* (case reports, outbreaks)
Quinupristin-dalfopristin	VRE; complicated skin and skin-structure infections due to MSSA and *S. pyogenes*	*E. faecalis* (<20%);[k] *E. faecium* (>90%)[k]
Mupirocin	Topical application to nares for *S. aureus* decolonization	*S. aureus* (74–100%)[l]
Nitrofurantoin	UTI caused by most gram-negative bacilli and some gram-positive organisms; prophylaxis in recurrent cystitis	*E. coli* (97%); *E. faecalis* (99%)
Fosfomycin	UTI caused by most gram-negative bacilli and some gram-positive organisms; prophylaxis in recurrent cystitis	Considered low prevalence[m]
Cefiderocol	Complicated UTIs and/or pyelonephritis caused by multidrug-resistant gram-negative bacteria, including extended-spectrum β-lactamase- or carbapenemase-producing organisms and multidrug-resistant *P. aeruginosa, A. baumannii, Stenotrophomonas maltophilia*, and *Burkholderia cepacia*	Very low resistance rates in initial studies
Eravacycline	Complicated intraabdominal infections caused by *E. coli, K. pneumoniae, C. freundii, E. cloacae, K. oxytoca, E. faecalis, E. faecium, S. aureus, S. anginosus* group, *C. perfringens, Bacteroides* spp., and *Parabacteroides distasonis*	Resistance noted in both gram-negative and gram-positive bacteria[n]
Imipenem-cilastatin-relebactam	Complicated intraabdominal infections, pneumonia, and complicated UTI including pyelonephritis caused by multidrug-resistant organisms including Enterobacterales and against some imipenem-nonsusceptible *P. aeruginosa*	Low resistance rates in initial studies
Lefamulin[e]	Community acquired pneumonia caused by MRSA, *S. pneumoniae*, and atypical CAP pathogens	Low resistance rates in target pathogens in initial studies
Meropenem-vaborbactam	Complicated UTI caused by KPC-producing *Enterobacterales*	Identified in KPC-producing strains of *K. pneumoniae*[o]
Omadacycline	Community-acquired bacterial pneumonia caused by *S. pneumoniae*, *S. aureus* (methicillin-susceptible isolates), *H. influenzae*, *Haemophilus parainfluenzae*, *K. pneumoniae*, *L. pneumophila*, *M. pneumoniae*, and *Chlamydophila pneumoniae*	Broad spectrum overall, but resistance can occur in gram-negative isolates; not active against *Pseudomonas*
Plazomicin	Complicated UTIs caused by carbapenemase-producing Enterobacteriaceae	Resistance uncommon except in infrequent isolates with plasmid-encoded ribosome modifying methylases

[a] Unless otherwise noted, susceptibility rates are based on isolates from the Massachusetts General Hospital Clinical Microbiology Laboratory collected between January and December 2022. Local rates will vary. [b] JA Karlowsky et al: Antimicrob Agents Chemother 56:1247, 2012. [c] RE Mendes et al: J Antimicrob Chemother 67:1321, 2012. [d] SW Long et al: Antimicrob Agents Chemother 58:6668, 2014. [e] M Nigo et al: Antimicrob Agents and Chemother 61:e01235, 2017. [f] D Van Duin, RA Bonomo: Clin Infect Dis 63:234, 2016. [g] SP McCurdy et al: Antimicrob Agents Chemother 59:5007, 2015. [h] *JA Karlowsky et al: Diag Microbiol Infect Dis 87:349, 2017.* [i] HS Sader et al: J Chemother 23:200, 2011. [j] J Torres et al: J Clin Microbiol 39:2677, 2001. [k] WS Oh et al: Antimicrob Agents Chemother 49:5176, 2005. [l] AE Simor et al: Antimicrob Agents Chemother 51:3880, 2007. [m] S Demirci-Duarte et al: Diagn Microbiol Infect Dis 98:115098, 2020. [n] LJ Scott: Drugs 79:315, 2019. [o] Sun et al Antimicrob Agents Chemother 61:e01694, 2017.

Abbreviations: CA-MRSA, community-acquired MRSA; CAP, community-acquired pneumonia; CA-UTI, community-acquired UTI; CDAD, *Clostridioides difficile*–associated diarrhea; ESBL, extended-spectrum β-lactamase; GAS, group A streptococcal; KPCs, *Klebsiella pneumoniae* carbapenemases; MAI, *Mycobacterium avium-intracellulare*; MDR, multidrug-resistant; MIC, minimal inhibitory concentration; MRSA, methicillin-resistant *S. aureus*; MSSA, methicillin-susceptible *S. aureus*; NDM, New Delhi metallo-β-lactamase; TMP-SMX, trimethoprim-sulfamethoxazole; UTI, urinary tract infection; VRE, vancomycin-resistant *Enterococcus*.

the drugs' four-membered lactam ring, which is their core structure. The differing side groups among the agents of this family determine the spectrum of activity. All β-lactams exert a bactericidal effect by inhibiting bacterial cell-wall synthesis. The β-lactams are classified as time-dependent killing agents; therefore, their clinical efficacy is best correlated with the proportion of the dosing interval during which drug levels remain above the MIC for the targeted pathogen.

Penicillins and β-Lactamase Inhibitors Penicillin, the first β-lactam, was discovered in 1928 by Alexander Fleming. Natural penicillins, such as penicillin G, are active against non-β-lactamase-producing gram-positive and gram-negative bacteria, anaerobes, and some gram-negative cocci. Penicillin G is used for penicillin-susceptible streptococcal infections, pneumococcal and meningococcal meningitis, enterococcal endocarditis, and syphilis. The antistaphylococcal penicillins, which have potent activity against methicillin-susceptible *Staphylococcus aureus* (MSSA), include nafcillin, oxacillin, dicloxacillin, and flucloxacillin. Aminopenicillins, such as ampicillin and amoxicillin, provide added coverage beyond penicillin against gram-negative cocci, such as *Haemophilus influenzae*, and some Enterobacterales, including *Escherichia coli*, *Proteus mirabilis*, *Salmonella*, and *Shigella*. The aminopenicillins are hydrolyzed by many common β-lactamases. These drugs are commonly used for infections caused by susceptible enterococcal and streptococcal species. IV ampicillin is commonly used in meningitis and endocarditis. Oral amoxicillin may be an option for otitis media, respiratory tract infections, and urinary tract infections. The antipseudomonal penicillins include ticarcillin and piperacillin. These penicillin groups generally offer adequate anaerobic coverage; the exceptions are *Bacteroides* species (such as *Bacteroides fragilis*), which produce β-lactamases and are generally resistant. The rising prevalence of β-lactamase-producing bacteria has led to the increased use of β-lactam–β-lactamase inhibitor combinations, such as ampicillin-sulbactam, amoxicillin-clavulanate, ticarcillin-clavulanate, piperacillin-tazobactam, ceftolozane-tazobactam, ceftazidime-avibactam, meropenem-vaborbactam, and imipenem-relebactam. The β-lactamase inhibitors themselves do not have antibacterial activity (with the exception of sulbactam, which has activity against *Acinetobacter baumannii*) but typically inhibit the *S. aureus* class A β-lactamase, β-lactamases of *H. influenzae* and *Bacteroides* species, and a number of plasmid-encoded β-lactamases. These combination agents are typically used when broader-spectrum coverage is needed—e.g., in pneumonia and intraabdominal infections. Piperacillin-tazobactam is a useful agent for febrile neutropenia, when local *P. aeruginosa* susceptibility rates are high. Avibactam, vaborbactam, and relebactam inhibit a broader spectrum of β-lactamases than the other inhibitors, including extended-spectrum β-lactamases (ESBLs), AmpC β-lactamases, and some carbapenemases **(see Chap. 150)**. Sulbactam combined with the inhibitor durlobactam is limited to treatment for *Acinetobacter* infections.

Cephalosporins The cephalosporin drug class encompasses several generations distinguished by spectrum of antibacterial activity. The first generation (cefazolin, cefadroxil, and cephalexin) largely has activity against gram-positive bacteria, with some additional activity against *E. coli*, *P. mirabilis*, and *Klebsiella pneumoniae*. First-generation cephalosporins are commonly used for infections caused by MSSA and streptococci (e.g., skin and soft tissue infections). Cefazolin is recommended for surgical prophylaxis against skin organisms. The second generation (cefamandole, cefuroxime, cefaclor, cefprozil, cefuroxime axetil, cefoxitin, and cefotetan) has additional activity against *H. influenzae* and *Moraxella catarrhalis*. Cefoxitin and cefotetan have potent activity against anaerobes as well, although there is some increased resistance seen for *B. fragilis*. Second-generation cephalosporins have been used to treat community-acquired pneumonia because of their activity against *S. pneumoniae*, *H. influenzae*, and *M. catarrhalis*. They are also used for other mild or moderate infections, such as acute otitis media and sinusitis. The third-generation cephalosporins are characterized by greater potency against gram-negative bacilli and reduced potency against gram-positive cocci. These cephalosporins, which include cefoperazone, cefotaxime, ceftazidime, ceftriaxone, cefdinir, cefixime, and cefpodoxime, are used for infections caused by Enterobacterales, although resistance is an increasing concern. Ceftriaxone penetrates the CSF and can be used to treat meningitis caused by *H. influenzae*, *N. meningitidis*, and susceptible strains of *S. pneumoniae*. It is also used for the treatment of later-stage Lyme disease, gonococcal infections, and streptococcal endocarditis. It is noteworthy that ceftazidime is the only third-generation cephalosporin with activity against *Pseudomonas aeruginosa*, but it lacks activity against gram-positive bacteria. This drug is frequently used for pulmonary infections in cystic fibrosis, postneurosurgical meningitis, and febrile neutropenia. The fourth generation of cephalosporins includes cefepime and cefpirome, broad-coverage agents with potent activity against both gram-negative bacilli, including *P. aeruginosa*, and gram-positive cocci. The fourth generation has clinical applications similar to those of the third generation and may offer additional activity over the first, second, and third generations in the presence of certain β-lactamases. These agents can be used in bacteremia, febrile neutropenia, and intraabdominal and urinary tract infections. Ceftaroline, a fifth-generation cephalosporin, differs from the other cephalosporins in its added activity against methicillin-resistant *S. aureus* (MRSA), which is resistant to all other β-lactams. Ceftaroline's gram-negative activity is similar to that of the third-generation cephalosporins but does not include *P. aeruginosa*. Ceftaroline may be used in community-acquired pneumonia and skin infections, and emerging data support its use in more severe infections such as bacteremia. Adverse reactions to ceftaroline have included hypersensitivity reactions and neutropenia. Ceftolozane-tazobactam and ceftazidime-avibactam are novel cephalosporin–β-lactamase inhibitor combinations with activity against gram-negative bacteria, including *Pseudomonas*, and some anaerobes. Both agents have been studied in complicated intraabdominal infections and complicated urinary tract infections. Ceftolozane-tazobactam is thought to be stable against many ESBL-producing organisms because of the tazobactam component. The addition of avibactam to ceftazidime yields a combination agent with activity against AmpC-, ESBL-, and *K. pneumoniae* carbapenemase (KPC)–producing organisms. These cephalosporin–β-lactamase inhibitor combinations may be of clinical benefit in multidrug-resistant gram-negative infections. Cefiderocol is a novel cephalosporin with enhanced uptake through bacterial iron uptake pathways and stability to a broad range of β-lactamases. It has been studied in complicated urinary tract infections and hospital-acquired and ventilator-associated pneumonia and may provide most benefit for multidrug-resistant gram-negative bacterial infections.

Carbapenems Carbapenems, including doripenem, imipenem, meropenem, and ertapenem, offer the most reliable coverage for strains containing ESBLs. All carbapenems have broad activity against gram-positive cocci, gram-negative bacilli, and anaerobes. None is active against MRSA, but all are active against MSSA, *Streptococcus* species, and Enterobacterales. Ertapenem is the only carbapenem that has poor activity against *P. aeruginosa* and *Acinetobacter*. Imipenem is active against penicillin-susceptible *Enterococcus faecalis* but not *Enterococcus faecium*. Carbapenems are not active against Enterobacterales containing carbapenemases. *Stenotrophomonas maltophilia* and some *Bacillus* species are intrinsically resistant to carbapenems because of a zinc-dependent carbapenemase. Addition of vaborbactam to meropenem and relebactam to imipenem results in inhibition of AmpC β-lactamases, ESBLs, and KPCs but not metallo-carbapenemases, such as NDM (New Delhi metallo-β-lactamases). Carbapenems may decrease the concentration of valproate products, and caution should be used in patients on this combination of therapy.

Monobactams Aztreonam is the sole monobactam in clinical use. Its activity is limited to gram-negative bacteria and includes *P. aeruginosa* and most other Enterobacterales. It is inactivated by ESBLs and carbapenemases. The principal use for aztreonam is as an alternative to penicillins, cephalosporins, or carbapenems in patients with a serious β-lactam allergy. Aztreonam is structurally related to ceftazidime and should be used cautiously in individuals with a serious ceftazidime allergy. It is used in febrile neutropenia and intraabdominal infections when other β-lactams cannot be used. Aztreonam is sometimes used in combination with avibactam for the treatment of NDM carbapenemase gram-negative bacterial infections.

Adverse Reactions to β-Lactam Drugs Agents within the β-lactam class are known for several adverse effects. Gastrointestinal side effects, mainly diarrhea, are common, but hypersensitivity reactions constitute the most common adverse effect of β-lactams. The reactions' severity can range from rash to anaphylaxis, but the rate of true anaphylactic reactions is only 0.05%. An individual with an accelerated IgE-mediated reaction to one β-lactam agent may still receive another agent within the class, but caution should be used and a β-lactam that has a dissimilar side chain and a low level of cross-reactivity would be the preferred choice. For example, the second-, third-, and fourth-generation cephalosporins and the carbapenems display very low cross-reactivity in patients with penicillin allergy. Aztreonam is the only β-lactam that has no cross-reactivity with the penicillin group. In cases of severe allergy, desensitization (a graded challenge) to the indicated β-lactam, with close monitoring, may be warranted if other antibacterial options are not suitable.

β-Lactams can rarely cause serum sickness, Stevens-Johnson syndrome, nephropathy, hematologic reactions, and neurotoxicity. Neutropenia appears to be related to high doses or prolonged use. Neutropenia and interstitial nephritis caused by β-lactams generally resolve upon discontinuation of the agent. Imipenem and cefepime are associated with an increased risk of seizure, but this risk is likely a class effect and related to high doses or doses that are not adjusted in renal impairment.

SULBACTAM-DURLOBACTAM

Durlobactam is a novel diazabicyclooctane non-β-lactam β-lactamase inhibitor combined with the β-lactamase inhibitor sulbactam for treatment of multidrug-resistant *Acinetobacter* infections. It was approved in the United States in 2023 for treatment of hospital-acquired and ventilator-associated pneumonia caused by *Acinetobacter baumannii-calcoaceticus* complex. Adverse reactions include abnormal liver function tests, diarrhea, anemia, and hypokalemia.

GLYCOPEPTIDES AND LIPOGLYCOPEPTIDES

Vancomycin is a glycopeptide antibiotic with activity against staphylococci (including MRSA and coagulase-negative staphylococci), streptococci (including *S. pneumoniae*), and enterococci. It is not active against gram-negative organisms. Vancomycin also displays activity against *Bacillus* species, *Corynebacterium jeikeium*, and gram-positive anaerobes such as *Peptostreptococcus*, *Actinomyces*, *Clostridium*, and *Propionibacterium* species. Vancomycin has several important clinical uses. It is used for serious infections caused by MRSA, including health care–associated pneumonia, bacteremia, osteomyelitis, and endocarditis. It is also commonly used for skin and soft tissue infections. Oral vancomycin is not absorbed systemically and is reserved for the treatment of *C. difficile* infection. Vancomycin is also an alternative for the treatment of infections caused by MSSA in patients who cannot tolerate β-lactams. Resistance to vancomycin is a rising concern. Strains of vancomycin-intermediate *S. aureus* (VISA) and vancomycin-resistant enterococci (VRE) are not uncommon. Vancomycin appears to be a concentration-dependent killer, with the AUC/MIC ratio being the best predictor of efficacy (Fig. 149-1). Guidelines recommend targeting a vancomycin trough level of 15–20 μg/mL in MRSA infections in order to maintain an AUC/MIC ratio >400. When using vancomycin, clinicians should monitor for nephrotoxicity. The risk of nephrotoxicity increases when trough levels are >20 μg/mL. Concomitant therapy with other nephrotoxic agents, such as aminoglycosides, also increases the risk. Ototoxicity was reported with early formulations of vancomycin but is currently uncommon because purer formulations are available. Both of these adverse effects are reversible upon discontinuation of vancomycin. Clinicians should be aware of vancomycin infusion reaction (formerly known as "red man syndrome"), a common reaction that presents as a rapid onset of erythematous rash or pruritus on the head, face, neck, and upper trunk. This reaction is caused by histamine release from basophils and mast cells and can be treated with diphenhydramine and slowing of the vancomycin infusion.

Telavancin, dalbavancin, and oritavancin are structurally similar to vancomycin and are referred to as *lipoglycopeptides*. They have antibacterial activity against *S. aureus* (including MRSA and some strains of VISA and vancomycin-resistant *S. aureus* [VRSA]), streptococci, and enterococci. Oritavancin may have activity against some strains of VRE. These lipoglycopeptide agents also provide coverage against anaerobic gram-positive organisms except for *Lactobacillus* and some *Clostridium* species. The clinical efficacy of telavancin has been demonstrated in both skin and soft tissue infections and nosocomial pneumonia, and the efficacy of dalbavancin and oritavancin has been shown in skin and soft tissue infections. The vancomycin resistance phenotype may reduce the potency of all three lipoglycopeptides, but the rate of resistance to these drugs among *S. aureus* and enterococcal isolates has been low. Adverse effects of telavancin include nephrotoxicity, metallic taste, and gastrointestinal side effects. Clinicians should be aware of the potential for electrocardiographic QT_c prolongation that can increase the risk of cardiac arrhythmias when telavancin is used concomitantly with other QT_c-prolonging agents. Telavancin may interfere with certain coagulation tests (e.g., causing false elevations in prothrombin time). Dalbavancin and oritavancin have safety profiles similar to that of vancomycin, with common effects reported as headache and gastrointestinal side effects. These glycolipopeptides should be used cautiously in patients with hypersensitivity reactions to vancomycin, as cross-allergy may be possible.

LIPOPEPTIDES

Daptomycin is a lipopeptide antibiotic with activity against a broad range of gram-positive organisms. It is active against staphylococci (including MRSA and coagulase-negative staphylococci), streptococci, and enterococci. Daptomycin remains active against enterococci that are resistant to vancomycin. In addition, it exhibits activity against *Bacillus*, *Corynebacterium*, *Peptostreptococcus*, and *Clostridium* species. Daptomycin's pharmacodynamic parameter for efficacy is concentration-dependent killing. Resistance to daptomycin is rare, but MICs may be higher for some VISA strains. Daptomycin can be used in skin and soft tissue infections, bacteremia, endocarditis, and osteomyelitis. It is an important alternative for MRSA and other gram-positive infections when bactericidal therapy is needed and vancomycin cannot be used. Daptomycin is generally well tolerated, and its main toxicity consists of elevation of creatine phosphokinase (CPK) levels and myopathy. CPK should be monitored during daptomycin treatment, and the drug should be discontinued if muscular toxicities occur. There have also been case reports of reversible eosinophilic pneumonia associated with daptomycin use.

AMINOGLYCOSIDES

The aminoglycosides are a class of antibacterial agents with concentration-dependent activity against most gram-negative organisms. The most commonly used aminoglycosides are gentamicin, tobramycin, and amikacin, although others, such as streptomycin, kanamycin, neomycin, and paromomycin, may be used in special circumstances. Plazomicin is a new aminoglycoside that is less affected by common resistance mechanisms and is approved for treatment of complicated urinary tract infections and acute pyelonephritis. Aminoglycosides have a significant dose-dependent postantibiotic effect; i.e., they have an antibacterial effect even after serum drug levels fall below inhibitory concentrations. The postantibiotic effect and concentration-dependent killing form the rationale behind extended-interval aminoglycoside dosing, in which a larger dose is given once daily rather than smaller doses multiple times daily. Aminoglycosides are active against gram-negative bacilli, such as Enterobacterales, *P. aeruginosa*, and *Acinetobacter*. They also enhance the activity of cell wall–active agents such as β-lactams or vancomycin against some gram-positive bacteria, including staphylococci and enterococci. This combination therapy is termed *synergistic* because the effect of both agents provides a killing effect greater than would be predicted from the effects of either agent alone. Amikacin and streptomycin have activity against *Mycobacterium tuberculosis*, and amikacin has activity against nontuberculous mycobacteria, including *Mycobacterium avium* complex and *Mycobacterium abscessus*. The aminoglycosides do not have activity against anaerobes, *S. maltophilia*, or *Burkholderia cepacia* complex. Aminoglycosides are used in clinical practice in a variety of infections caused by gram-negative organisms, including bacteremia and urinary tract infections. They are frequently used in combination for the treatment of *P. aeruginosa* infection. When used in combination with a cell wall–active agent, gentamicin and streptomycin are also important for the treatment of gram-positive bacterial endocarditis. All aminoglycosides can cause nephrotoxicity and ototoxicity. The risk of nephrotoxicity is not well defined; however, some studies have indicated that the effect may be related to the duration of therapy as well as to the concomitant use of other nephrotoxic agents. Nephrotoxicity is usually reversible, but ototoxicity can be irreversible.

MACROLIDES AND KETOLIDES

The macrolides (azithromycin, clarithromycin, and erythromycin) and ketolides (telithromycin) are classes of antibiotics that inhibit protein synthesis. Compared with erythromycin (the older antibiotic), azithromycin and clarithromycin have better oral absorption and tolerability. Azithromycin, clarithromycin, and telithromycin all have broader spectra of activity than erythromycin, which is less frequently used. These agents are commonly used in the treatment of upper and lower respiratory tract infections caused by *S. pneumoniae*, *H. influenzae*, *M. catarrhalis*, and atypical organisms (e.g., *Chlamydophila pneumoniae*, *Legionella pneumophila*, and *Mycoplasma pneumoniae*); group A streptococcal pharyngitis in penicillin-allergic patients; and nontuberculous mycobacterial infections (e.g., caused by *Mycobacterium marinum* and *Mycobacterium chelonae*) as well as in the prophylaxis and treatment of *M. avium* complex infection in patients with HIV/AIDS and in combination therapy for *Helicobacter pylori* infection and bartonellosis. Enterobacterales, *Pseudomonas* species, and *Acinetobacter* species are intrinsically resistant to macrolides as a result of decreased membrane permeability, although azithromycin is active against gram-negative diarrheal pathogens. The major adverse effects of this drug class include nausea, vomiting, diarrhea and abdominal pain, prolongation of QT_c interval, exacerbation of myasthenia gravis, and tinnitus and reversible deafness, especially in the elderly. Azithromycin specifically has been associated with an increased risk of death, especially among patients with underlying heart disease, because of the risk of QT_c interval prolongation and torsades de pointes arrhythmia. Erythromycin, clarithromycin, and telithromycin inhibit the CYP3A4 hepatic drug-metabolizing enzyme and can result in increased levels of coadministered drugs, including benzodiazepines, statins, warfarin, cyclosporine, and tacrolimus. Azithromycin does not inhibit CYP3A4 and therefore does not interact with these drugs.

CLINDAMYCIN

Clindamycin is a lincosamide antibiotic and is bacteriostatic against some organisms and bactericidal against others. It is used most often to treat bacterial infections caused by anaerobes (e.g., *B. fragilis*, *Clostridium perfringens*, *Fusobacterium* species, *Prevotella melaninogenicus*, and *Peptostreptococcus* species) and susceptible staphylococci and streptococci. Clindamycin is used for treatment of dental infections, anaerobic lung abscess, and skin and soft tissue infections. It is used together with bactericidal agents (penicillins or vancomycin) to inhibit new toxin synthesis in the treatment of streptococcal or staphylococcal toxic shock syndrome. Other uses include treatment of infections caused by *Capnocytophaga canimorsus*, combination therapy for babesiosis and occasionally malaria, and therapy for toxoplasmosis. Clindamycin has excellent oral bioavailability. Adverse effects include nausea, vomiting, diarrhea, *C. difficile*–associated diarrhea and pseudomembranous colitis, maculopapular rash, and rarely Stevens-Johnson syndrome.

TETRACYCLINES

The older (doxycycline, minocycline, and tetracycline) and newer (tigecycline, eravacycline, and omadacycline) tetracyclines inhibit protein synthesis and are bacteriostatic. These drugs have wide clinical uses. They are used in the treatment of skin and soft tissue infections caused by gram-positive cocci (including MRSA), spirochetal infections (e.g., Lyme disease, syphilis, leptospirosis, and relapsing fever), rickettsial infections (e.g., Rocky Mountain spotted fever, scrub typhus), atypical pneumonia, sexually transmitted infections (e.g., *Chlamydia trachomatis* infection, lymphogranuloma venereum, and granuloma inguinale), infections with *Nocardia* and *Actinomyces*, brucellosis, tularemia, Whipple's disease, and malaria. Tigecycline, eravacycline, and omadacycline are also used in combination with other agents for treatment of *M. abscessus*. Tigecycline is a glycylcycline derived from minocycline and is available only in IV formulation. It is indicated in the treatment of complicated skin and soft tissue infections, complicated intraabdominal infections, and community-acquired bacterial pneumonia in adults. Tigecycline has activity against MRSA, vancomycin-sensitive enterococci, many Enterobacterales, and *Bacteroides* species; it has no activity against *P. aeruginosa*. This drug has been used in combination with colistin for the treatment of serious infections with multidrug-resistant gram-negative organisms. A pooled analysis of 13 clinical trials found an increased risk of death and treatment failure among patients given tigecycline alone; as a result, the FDA mandated a black box warning. Eravacycline is a fluorocycline derivative available in IV formulation with a similar spectrum but more potent than tigecycline in vitro. It has been approved for complicated intraabdominal infections. Omadacycline is an aminomethylcycline derivative available in both IV and oral formulations. It has activity similar to that of tigecycline against gram-positive pathogens but is less active against gram-negative pathogens. Omadacycline has been approved for treatment of bacterial skin and skin structure infections and community-acquired bacterial pneumonia. Tetracyclines have reduced absorption when orally coadministered with calcium- and iron-containing compounds, including milk, and doses should be spaced at least 2 h apart. The major adverse reactions to old and new tetracyclines are nausea, vomiting, diarrhea, and photosensitivity. Tetracyclines have been associated with fetal bone-growth abnormalities and should be avoided during pregnancy and in the treatment of children <8 years old.

TRIMETHOPRIM-SULFAMETHOXAZOLE

Trimethoprim-sulfamethoxazole (TMP-SMX) is an antibiotic with two components that each inhibit a separate step in folate synthesis and produce antibacterial activity. TMP-SMX is active against gram-positive bacteria such as staphylococci and streptococci; however, its use against MRSA is usually limited to community-acquired infections, and its activity against *Streptococcus pyogenes* may not be reliable. This drug is also active against many gram-negative bacteria, including *H. influenzae*, *E. coli*, *P. mirabilis*, *Neisseria gonorrhoeae*, and *S. maltophilia*. TMP-SMX is not active against anaerobes or *P. aeruginosa*. It has many uses because of its wide spectrum of activity and high oral bioavailability. Urinary tract infections, skin and soft tissue infections, and respiratory tract infections are among the common uses. Another important indication is for both prophylaxis and treatment of *Pneumocystis jirovecii* infections in immunocompromised patients. Resistance to TMP-SMX has limited its use against many Enterobacterales. Resistance rates among urinary isolates of *E. coli* are almost 25% in the United States. The most common adverse reactions associated with TMP-SMX are gastrointestinal effects such as nausea, vomiting, and diarrhea. In addition, rash is a common allergic reaction and may preclude the subsequent use of other sulfonamides. With prolonged use, leukopenia, thrombocytopenia, and granulocytopenia can develop. TMP-SMX can also cause nephrotoxicity, hyperkalemia, and hyponatremia, which are more common at high doses. TMP-SMX has several important interactions with other drugs (Table 149-3), including warfarin, phenytoin, and methotrexate.

FLUOROQUINOLONES

The fluoroquinolones include norfloxacin, ciprofloxacin, ofloxacin, levofloxacin, moxifloxacin, gemifloxacin, and delafloxacin. Ciprofloxacin and levofloxacin have the broadest spectrum of activity against gram-negative bacteria, including *P. aeruginosa* (similar to that of third-generation cephalosporins). Because of the risk of selection of resistance during fluoroquinolone treatment of serious pseudomonal infections, these agents are usually used in combination with an antipseudomonal β-lactam. Levofloxacin, moxifloxacin, gemifloxacin, and delafloxacin have additional gram-positive activity, including that against *S. pneumoniae* and some strains of MSSA, and with the exception of delafloxacin, these agents are used for treatment of community-acquired pneumonia. Strains of MRSA are commonly resistant to all fluoroquinolones except delafloxacin. Moxifloxacin is used as one component of second-line regimens for multidrug-resistant tuberculosis. Fluoroquinolones are no longer used for treatment of gonorrhea because of common resistance in *N. gonorrhoeae.* Fluoroquinolones exhibit concentration-dependent killing, are well absorbed orally, and have elimination half-lives that usually support once- or twice-daily dosing. Oral coadministration with compounds containing high concentrations of aluminum, magnesium, or calcium can reduce fluoroquinolone absorption. The penetration of fluoroquinolones into prostate tissue supports their use for bacterial prostatitis. Fluoroquinolones are generally well tolerated but can cause central nervous system (CNS) stimulatory effects, including seizures; peripheral neuropathy; glucose dysregulation; and tendinopathy associated with Achilles tendon rupture, particularly in older patients, organ transplant recipients, and patients taking glucocorticoids. Other potential effects on connective tissues include an association with increased risk of aortic aneurysm. Worsening of myasthenia gravis also has been associated with quinolone use. Moxifloxacin causes modest prolongation of the QT_c interval and should be used with caution in patients receiving other QT_c-prolonging drugs.

RIFAMYCINS

The rifamycins include rifampin, rifabutin, and rifapentine. Rifampin is the most commonly used rifamycin. For almost all therapeutic indications, it is used in combination with other agents to reduce the likelihood of selection of high-level rifampin resistance. Rifampin is used foremost in the treatment of mycobacterial infections—specifically, as a mainstay of combination therapy for *M. tuberculosis* infection or as a single agent in the treatment of latent *M. tuberculosis* infection. In addition, it is often used in the treatment of some nontuberculous mycobacterial infections. Rifampin is used in combination regimens for the treatment of staphylococcal infections, particularly prosthetic-valve endocarditis and bone infections with retained hardware. It is a component of combination therapy for brucellosis (with doxycycline) and leprosy (with dapsone for tuberculoid leprosy and with dapsone and clofazimine for lepromatous disease). Rifampin can be used alone for prophylaxis in close contacts of patients with *H. influenzae* or *N. meningitidis* meningitis. The drug has high oral bioavailability, which is further enhanced when it is taken on an empty stomach. Rifampin has several adverse effects, including elevated aminotransferase levels (14%), rash (1–5%), and gastrointestinal events such as nausea, vomiting, and diarrhea (1–2%). Its many clinically relevant interactions with other drugs (Table 149-3) mandate the clinician's careful review of the patient's medications before rifampin initiation to assess safety and the need for additional monitoring, including monitoring of drug levels.

METRONIDAZOLE

Metronidazole is used in the treatment of anaerobic bacterial infections as well as infections caused by protozoa (e.g., amebiasis, giardiasis, trichomoniasis). It is the agent of choice as a component of combination therapy for polymicrobial abscesses in the lung, brain, or abdomen, the etiology of which often includes anaerobic bacteria, and for bacterial vaginosis, pelvic inflammatory disease, and anaerobic infections, such as those due to *Bacteroides*, *Fusobacterium*, and *Prevotella* species. This drug is an alternative agent for treatment of mild to moderate *C. difficile*–associated diarrhea. Metronidazole is bactericidal against anaerobic bacteria and exhibits concentration-dependent killing. It has high oral bioavailability and tissue penetration, including penetration of the blood-brain barrier. The majority of *Actinomyces*, *Propionibacterium*, and *Lactobacillus* species are intrinsically resistant to metronidazole. The major adverse effects include nausea, diarrhea, and a metallic taste. Concomitant ingestion of alcohol may result in a disulfiram-like reaction, and patients are usually instructed to avoid alcohol during treatment. Long-term treatment carries the risk of leukopenia, neutropenia, peripheral neuropathy, and CNS toxicity manifesting as confusion, dysarthria, ataxia, nystagmus, and ophthalmoparesis. Through metronidazole's effect on the CYP2C9 drug-metabolizing enzyme, its coadministration with warfarin can result in decreased metabolism and enhanced anticoagulant effects that require close monitoring. Concomitant administration of metronidazole with lithium can result in increased serum levels of lithium and associated toxicity; coadministration with phenytoin can result in phenytoin toxicity and possibly decreased levels of metronidazole.

OXAZOLIDINONES

Linezolid is a bacteriostatic agent and is indicated for serious infections due to resistant gram-positive bacteria, such as MRSA and VRE. The intrinsic resistance of gram-negative bacteria is mediated primarily by endogenous efflux pumps. Linezolid has excellent oral bioavailability. Adverse effects include myelosuppression and ocular and peripheral neuropathy with prolonged therapy. Peripheral neuropathy may be irreversible. Linezolid is a weak, reversible monoamine oxidase inhibitor, and coadministration with sympathomimetics and foods rich in tyramine should be avoided. Linezolid has been associated with serotonin syndrome when coadministered with selective serotonin reuptake inhibitors. Tedizolid has properties similar to those of linezolid, but with lower dosing due to greater potency, it may be less likely to cause adverse hematologic and neuropathic effects.

PLEUROMUTILINS

Lefamulin is the only member of the pleuromutilin class approved for systemic use; it is available in IV and oral formulations. Lefamulin has in vitro activity against *S. aureus* (including MRSA), *S. pneumoniae*, *H. influenzae*, and atypical respiratory pathogens, including *L. pneumophila*, *M. pneumoniae*, and *C. pneumoniae*, and has been approved for treatment of community-acquired bacterial pneumonia. Adverse effects are most commonly gastrointestinal, including diarrhea (12%), nausea (5%), and vomiting (3%). Prolongation of QT_c interval and hepatic transaminase elevations occur in some patients. There can be interactions with drugs that are either inducers or inhibitors of CYP3A4 or P-glycoprotein transporter.

NITROFURANTOIN

Nitrofurantoin's antibacterial activity results from the drug's conversion to highly reactive intermediates that can damage bacterial DNA and other macromolecules. Nitrofurantoin is bactericidal, and its action is concentration dependent. It displays activity against a range of gram-positive bacteria, including *S. aureus*, *Staphylococcus epidermidis*, *Staphylococcus saprophyticus*, *E. faecalis*, *Streptococcus agalactiae*, group D streptococci, viridans streptococci, and corynebacteria, as well as gram-negative organisms, including *E. coli, Enterobacter*, *Salmonella*, and *Shigella* species. Nitrofurantoin is used primarily in the treatment of urinary tract infections and is preferred in the treatment of such infections in pregnancy. It may be used for the prevention of recurrent cystitis. Recently, there has been interest in the use of nitrofurantoin for treatment of urinary tract infections caused by ESBL-producing Enterobacterales such as *E. coli*, although resistance has been growing in Latin America and parts of Europe. Coadministration with magnesium should be avoided because of decreased absorption, and patients should be encouraged to take the drug with food to increase its bioavailability and decrease the risk of adverse effects, which include nausea, vomiting, and diarrhea. Nitrofurantoin may also cause pulmonary fibrosis and drug-induced hepatitis. Because the risk of adverse reactions increases with age, the use of nitrofurantoin in elderly patients is not recommended. Patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency are at elevated risk for nitrofurantoin-associated hemolytic anemia.

POLYMYXINS

Colistin and polymyxin B act by disrupting bacterial cell membrane integrity and are active against the nonenteric pathogens *P. aeruginosa* and *A. baumannii* but not against *Burkholderia*. These drugs also exhibit activity against many Enterobacterales, with the exceptions of *Proteus*, *Providencia*, and *Serratia* species. They lack activity against gram-positive bacteria. Polymyxins are bactericidal and are available in IV formulations. Colistimethate is converted to the active form (colistin) in plasma. Polymyxins are most often used for infections due to pathogens resistant to multiple other antibacterial agents, including urinary tract infections, hospital-acquired pneumonia, and bloodstream infections. Nebulized formulations have been used for adjunctive treatment of refractory ventilator-associated pneumonia as well as a prevention strategy for patients with bronchiectasis and/or cystic fibrosis. The most important adverse effect is dose-dependent reversible nephrotoxicity. Neurotoxicity, including paresthesias, muscle weakness, and confusion, is reversible and less common than nephrotoxicity.

■ QUINUPRISTIN-DALFOPRISTIN

Quinupristin-dalfopristin contains two members of the streptogramin class of antibiotics and kills bacteria by inhibiting protein synthesis. The antibacterial spectrum of quinupristin-dalfopristin includes staphylococci (including MRSA), streptococci, and *E. faecium* (but not *E. faecalis*). This drug combination is also active against *Corynebacterium* species and *L. monocytogenes*. Quinupristin-dalfopristin is not reliably active against gram-negative organisms. It exhibits concentration-dependent killing, with an AUC/MIC ratio predicting efficacy. The clinical use of quinupristin-dalfopristin is largely for infections due to vancomycin-resistant *E. faecium* and other gram-positive bacterial infections. The drug has demonstrated efficacy in a variety of infections, including urinary tract infections, bone and joint infections, and bacteremia. Adverse effects associated with quinupristin-dalfopristin include infusion-related reactions, arthralgias, and myalgias. The arthralgias and myalgias may be severe enough to warrant drug discontinuation. Quinupristin-dalfopristin inhibits the CYP3A4 drug-metabolizing enzyme, with consequent drug interactions (Table 149-3).

■ FOSFOMYCIN

Fosfomycin is a phosphonic acid antibiotic that has greater activity in acidic environments and is excreted in its active form in the urine. The oral formulation is primarily for prophylaxis and treatment of uncomplicated cystitis and should be avoided if there is concern about pyelonephritis. The drug is administered as a single 3-g dose that results in high urine concentrations for up to 48 h. Fosfomycin is active against *S. aureus*, vancomycin-susceptible enterococci and VRE, and a wide range of gram-negative organisms, including *E. coli*, *Enterobacter* species, *Serratia marcescens*, *P. aeruginosa*, and *K. pneumoniae*. Notably, the vast majority of ESBL-producing Enterobacterales are susceptible to fosfomycin. *A. baumannii* and *Burkholderia* species are resistant. The emergence of resistance to fosfomycin has not been observed during treatment of cystitis but has been documented during IV treatment of respiratory tract infections and osteomyelitis. The few adverse effects that have been reported include nausea and diarrhea.

■ CHLORAMPHENICOL

The use of chloramphenicol is limited by its potentially serious toxicities. When other agents are contraindicated or ineffective, chloramphenicol represents an alternative treatment for infections, including meningitis caused by susceptible bacteria such as *N. meningitidis*, *H. influenzae*, and *S. pneumoniae*. It has also been used for the treatment of anthrax, brucellosis, *Burkholderia* infections, chlamydial infections, clostridial infections, ehrlichiosis, rickettsial infections, and typhoid fever. Adverse reactions include aplastic anemia, myelosuppression, and gray baby syndrome. Chloramphenicol inhibits the CYP2C19 and CYP3A4 drug-metabolizing enzymes and consequently increases levels of many classes of drugs.

APPROACH TO PROPHYLAXIS OF INFECTION

Antibacterial prophylaxis is indicated only in selected circumstances **(Table 149-5)** and should be supported by well-designed studies or expert panel recommendations. In all cases, the risk or severity of the infection to be prevented should be greater than the adverse consequences of antibacterial therapy, including the potential for selection of resistance. In addition, the timing and duration of antibacterial treatment should be targeted for maximal effect and minimal required exposure. Prophylaxis of surgical infections targets bacteria that may contaminate the wound during the surgical procedure, including the skin flora of the patient or operating team and the air in the operating room. Delivery of the antibacterial drug within 1 h before the surgical incision is most effective because it maximizes tissue concentrations. For prolonged procedures, redosing may be necessary to maintain effective blood and tissue levels until the wound is closed. Additional dosing is not recommended after the incision is closed, and antimicrobials should be discontinued after incisional closure in the operating room. In patients with nasal carriage of *S. aureus*, preoperative decolonization with nasal mupirocin reduces the rate of *S. aureus* surgical-site infections and is generally recommended for high-risk procedures such as cardiac surgery and orthopedic implantation of prosthetic devices. For dental procedures, preprocedure antibacterial drugs are recommended for select patient populations to prevent transient bacteremia during the procedure and the seeding of certain high-risk cardiac lesions. Prophylaxis is also used in nonprocedural settings in certain patients who have recurrent infections or who are at risk of serious infection from a specific exposure (e.g., close contact with a patient with meningococcal meningitis). Extension of prophylaxis beyond the period of infection risk does not add further benefit and may increase the risk of resistance selection or *C. difficile* disease.

ANTIMICROBIAL STEWARDSHIP

In an era of increasing prevalence of multidrug-resistant bacteria and with a substantial amount of inappropriate antimicrobial use, the need for rational antimicrobial prescribing has never been greater **(Chap. 150)**. *Antimicrobial stewardship* describes the practice of promoting the selection of the appropriate drug, dosage, route, and duration of therapy. Antimicrobial stewardship programs implement a variety of strategies to (1) improve patient care through appropriate antimicrobial use; (2) preserve a vital health care resource by curbing the development of resistance within patient populations; (3) reduce the incidence of adverse effects; and (4) control costs. The Centers for Disease Control and Prevention (CDC) guidelines, The Joint Commission (TJC) Medication Management Standards, and the Centers for Medicare and Medicaid Services (CMS) Conditions of Participation, as well as the 2015 National Action Plan for Combating Antibiotic-Resistant Bacteria, have all supported *antimicrobial stewardship* in various health care settings. Antimicrobial stewardship programs are typically multidisciplinary and often include infectious disease physicians, clinical pharmacists (usually with special training in infectious disease), clinical microbiologists, information systems specialists, infection prevention and control practitioners, and epidemiologists. These teams employ a variety of approaches to achieve the program's goals.

Established strategies of antimicrobial stewardship programs include (1) prospective audit of antimicrobial use, with intervention and feedback; (2) formulary restriction; and (3) preauthorization. *Prospective audit and feedback* are usually undertaken by an infectious disease physician or a pharmacist. In this process, orders for broad-spectrum antimicrobials (e.g., carbapenems) or agents for which more cost-effective alternatives may exist (e.g., daptomycin, ceftazidime-avibactam) are reviewed on a regular basis for appropriateness. In circumstances when antimicrobial use can be further optimized, the stewardship program team can intervene to recommend an alternative. This process has been successful in several quasi-experimental studies, resulting in declines in use of broad-spectrum drugs when unnecessary and decreases in adverse events, such as *C. difficile* infection. *Formulary restriction* is the inclusion of a limited set of antimicrobial agents in a hospital formulary for the purpose of limiting indiscriminate use of antimicrobials in the absence of demonstrated benefit. Such restriction also serves to avoid unnecessary drug expenditure. *Preauthorization* is the practice of requiring clinicians to obtain approval before using

TABLE 149-5 Prophylaxis of Bacterial Infections in Adults

CONDITION	ANTIBACTERIAL AGENTS[a]	TIMING OR DURATION OF PROPHYLAXIS
Surgical		
Clean (cardiac, thoracic, neurologic, orthopedic, vascular, plastic)	Cefazolin (vancomycin,[b] clindamycin)	1 h before incision; re-dose with long procedures
Clean (ophthalmic)	Topical neomycin–polymyxin B–gramicidin, topical moxifloxacin	Every 5–15 min for 5 doses immediately prior to procedure
Clean-contaminated (head and neck)	Cefazolin + metronidazole, ampicillin-sulbactam[c] (clindamycin)	1 h before incision; re-dose with long procedures
Clean-contaminated (hysterectomy, gastroduodenal, biliary, unobstructed small intestine, urologic)	Cefazolin, ampicillin-sulbactam[c] (clindamycin + aminoglycoside, aztreonam, or fluoroquinolone)	1 h before incision; re-dose with long procedures
Clean-contaminated (colorectal, appendectomy)	Cefazolin + metronidazole, ampicillin-sulbactam,[c] ertapenem (clindamycin + aminoglycoside, aztreonam, or fluoroquinolone)	1 h before incision; re-dose with long procedures
Dirty (ruptured viscus)	Therapeutic regimen directed at anaerobes and gram-negative bacteria (e.g., ceftriaxone + metronidazole)	1 h before incision; re-dose with long procedures; continue for 3–5 days after procedure
Dirty (traumatic wound)	Therapeutic regimen: cefazolin (clindamycin ± aminoglycoside, aztreonam, or fluoroquinolone)	1 h before incision; re-dose with long procedures; continue for 3–5 days after procedure
Nonsurgical		
Dental, oral, or upper respiratory procedures in patients with high-risk cardiac lesions (prosthetic valves, congenital heart defects, prior endocarditis)	Amoxicillin PO, ampicillin IM (clindamycin PO, IV)	Oral agents 1 h before procedure; injection 30 min before procedure
Recurrent *S. aureus* skin infections[d]	Mupirocin[e]	Intranasal application for 5 days
Recurrent cellulitis associated with lymphatic disruption[d]	Benzathine penicillin IM monthly, oral penicillin or erythromycin twice daily	Undefined
Recurrent cystitis in women[d]	Nitrofurantoin, TMP-SMX, fluoroquinolone	After sexual intercourse *or* 3 times weekly for up to 1 year
Bite wounds	Amoxicillin-clavulanate (doxycycline, moxifloxacin)	3–5 days
Recurrent spontaneous bacterial peritonitis in cirrhotic patients[d]	Fluoroquinolone[f]	Undefined
Recurrent pneumococcal meningitis in patient with CSF leak or humoral immune defect[d]	Penicillin	Undefined
Exposure to patient with meningococcal meningitis	Rifampin, ciprofloxacin	2 days (rifampin), single dose (ciprofloxacin)
High-risk neutropenia (ANC, ≤100/μL for >7 days)[d]	Levofloxacin or ciprofloxacin[f]	Until neutropenia resolves or fever dictates use of other antibacterials

[a]Regimens in parentheses are alternatives for patients allergic to β-lactams. [b]Vancomycin may be given together with cefazolin to patients known to be colonized with methicillin-resistant *Staphylococcus aureus*. [c]Cefoxitin or cefotetan may also be considered. [d]Not considered routine for all patients, but an acceptable consideration among alternative approaches. [e]Usually coupled with bathing with chlorhexidine-containing skin antiseptic. [f]Choice of fluoroquinolone prophylaxis must be balanced against the risk of selection of resistance.

Abbreviations: ANC, absolute neutrophil count; CSF, cerebrospinal fluid; TMP-SMX, trimethoprim-sulfamethoxazole.

selected antimicrobials. Approval may be provided electronically with sophisticated Computerized Provider Order Entry (CPOE) software, after specific criteria for use are met, or after communication with an infectious disease specialist as designated by the stewardship program. These strategies have led to a decrease in *C. difficile* infections and to improvements in drug susceptibility patterns.

Additional strategies used in specific health care settings are guidelines and pathways, dose optimization, parenteral-to-oral conversion, antibiotic time-out, and de-escalation of therapy. Documentation of the indication for which each antimicrobial is prescribed is also encouraged. Finally, antimicrobial stewardship teams provide ongoing education of best practices. An evolving and increasingly active area of clinical research to identify best practices, antimicrobial stewardship continues to grow as an essential service in various health care settings. The IDSA, in collaboration with several other professional organizations, has published guidelines for developing institutional antimicrobial stewardship programs (*www.idsociety.org/Antimicrobial_Agents/*).

ACKNOWLEDGMENT

The authors thank Christy A. Varughese for her significant contributions to this chapter in the previous editions.

FURTHER READING

Barlam TF et al: Implementing an antibiotic stewardship program: Guidelines by the Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America. Clin Infect Dis 62:e51, 2016.

Bratzler DW et al: Clinical practice guidelines for antimicrobial prophylaxis in surgery. Surg Infect (Larchmt) 14:73, 2013.

Calderwood MS et al: Strategies to prevent surgical site infections in acute care hospitals: 2022. Update. Infect Control Hospital Epidemiol 44:695, 2023.

Grayson ML et al (eds): *Kucers' The Use of Antibiotics. A Clinical Review of Antibacterial, Antifungal, Antiparasitic and Antiviral Drugs*, 7th ed. Boca Raton, CRC Press, 2018.

Infectious Diseases Society of America: Practice guidelines by organ system. Available at *https://www.idsociety.org/practice-guideline/practice-guidelines/*.

Jeffries MN et al: Consequences of avoiding β-lactams in patients with β-lactam allergies. J Allergy Clin Immunol 137:1148, 2016.

Labreche MJ et al: Recent updates on the role of pharmacokinetics–pharmacodynamics in antimicrobial susceptibility testing as applied to clinical practice. Clin Infect Dis 61:1446, 2015.

Rotschafer J et al (eds): *Antibiotic Pharmacodynamics. Methods in Pharmacodynamics and Toxicology*. New York, Humana Press, 2016.

Shenoy ES et al: Evaluation and management of penicillin allergy: A review. JAMA 321:188, 2019.

Tamma PD et al: Infectious Diseases Society of America 2023 guidance on the treatment of antimicrobial resistant gram-negative infections. Clin Infect Dis 18:ciae403, 2024.

150

Bacterial Resistance to Antimicrobial Agents

David C. Hooper

DEFINITION OF RESISTANCE

The action of antimicrobial agents on a range of targets within the bacterial cell can result in inhibition of bacterial growth or in killing of the bacterial cell **(Chap. 149)**. Reduction in or loss of an agent's antibacterial effect is referred to as *resistance*, and the properties of or alterations in the bacterium that result in reduced antimicrobial activity are termed *resistance mechanisms*. Bacteria can be resistant to single or multiple antimicrobials, as detailed in the sections that follow. The occurrence and magnitude of resistance are often assessed in clinical microbiology laboratories by measurement of the lowest drug concentration that inhibits growth of a bacterium (minimal inhibitory concentration, or MIC) with a standardized inoculum and growth conditions. MIC values are generally interpreted as representing bacterial susceptibility, intermediate susceptibility, or resistance; the interpretation is based on correlations of the MIC values with the pharmacokinetics and delivery of a drug to the site of infection in the body as well as with data from clinical trials. Thus, a clinical laboratory result of "susceptible" for a bacterium predicts a likely clinical response to an appropriately dosed antimicrobial drug by a patient infected with that organism, whereas a result of "resistant" predicts poor or no clinical response to that drug. Breakpoint MIC values for categorization of bacteria as susceptible, intermediate, or resistant are generally developed by regulatory and advisory groups and are often based on the distribution of MIC values from a large collection of recent clinical bacterial isolates. Research studies on the mechanisms and epidemiology of resistance may in some cases use different and less rigid definitions of resistance based on determination of a reproducible increase in an MIC value relative to a baseline reference MIC, independent of clinical breakpoints.

MECHANISMS OF RESISTANCE

Bacteria use a wide variety of mechanisms to interrupt or circumvent the activity of antibacterial agents **(Table 150-1 and Fig. 150-1)**. Although myriad, these mechanisms can generally be grouped into three categories: (1) alteration or bypassing of targets that exhibit reduced binding of the drug, (2) altered access of the drug to its target by reductions in uptake or increases in active efflux, and (3) a modification of the drug that reduces its activity. These mechanisms result from either mutations in bacterial chromosomal genes occurring spontaneously during bacterial DNA replication or the acquisition of new genes by DNA transfer from other bacteria or uptake of exogenous DNA. New genes are most often acquired on self-replicating plasmids or other DNA elements transferred from other bacteria. However, some bacteria, such as *Streptococcus pneumoniae* and *Neisseria gonorrhoeae*, can also take up fragments of environmental DNA from related bacterial species and recombine that DNA directly into their own chromosomes, a process called *transformation*. Not uncommonly, resistant bacteria have combinations of resistance mechanisms either within one category or among categories; many plasmids contain more than one resistance gene; and bacteria can acquire multiple plasmids. Thus, plasmid acquisition itself can in many cases confer resistance to multiple antibacterial agents. Resistance to multiple, structurally unrelated antibiotics can also occur by mutations that cause increased expression or, less often, expanded substrate profiles of certain bacterial efflux pumps, some of which are able to transport multiple antibacterial agents as well as other compounds out of the cell.

Many antibacterial drugs are derived from natural products of environmental microbial species. Some genes encoding resistance to these drugs originate in the drug-producer organism to protect it from its own product and have then been mobilized onto plasmids that spread to other organisms. Surviving nonproducer bacteria in the exposed natural environment may also have evolved resistance under selection pressure that adds to the reservoir of resistance mechanisms. Exposure to antibacterial agents either in nature or during human, animal, or other use then results in the selection of resistant strains within an otherwise susceptible bacterial population. In some cases, resistance mechanisms may confer disadvantages that render bacterial growth or survival fitness inferior to that of susceptible strains. In a number of examples, however, fitness defects are often mitigated over time by compensatory mutational mechanisms that make the bacteria both resistant and fit and thereby more likely to persist in a reservoir even in the absence of continued antimicrobial selection pressures. Discussed below are the major classes of antimicrobial agents currently in clinical use and the most important mechanisms of resistance encountered in clinical infections.

β-Lactams β-lactams, the largest class of antibiotics, inhibit bacterial cell-wall synthesis by binding to cell-wall transpeptidases, cross-linking enzymes that are also called penicillin-binding proteins (PBPs); PBPs are targets that are unique to bacteria and have no mammalian counterpart. The most common mechanism of resistance to β-lactams, particularly in gram-negative bacteria, is their degradation by β-lactamases, enzymes that break down the core β-lactam ring and destroy drug activity. β-Lactamases differ in the spectrum of β-lactams they can degrade. Some β-lactamases are encoded on the bacterial chromosome, and their activity contributes to the intrinsic susceptibility profile of a particular species. Chromosomally encoded β-lactamases can be produced in varying amounts that affect the degree of resistance. In some cases, enzyme expression is physiologically induced by exposure to certain β-lactams; in other cases, enzyme expression can become constant or constitutive through mutations in genes that encode the regulators of expression of a β-lactamase gene. Other β-lactamases are encoded by genes on acquired plasmids and are usually constitutively expressed. The resistance profiles due to plasmids may be present in some strains of a species but not others, depending on which plasmids the strain has acquired. In gram-positive bacteria, β-lactamases are secreted into the extracellular environment, whereas in gram-negative bacteria these enzymes are secreted into the periplasmic space between the cytoplasmic and outer membranes—a limited space that enables the presence of high concentrations of β-lactamase. In gram-negative bacteria, access of β-lactams both to their target PBPs and to β-lactamases requires diffusion across the outer membrane, generally through the porin diffusion channels. Reductions in outer-membrane diffusion channels due to mutation can further augment the efficiency of β-lactamase degradation of β-lactams: slow diffusion acts together with the high enzyme concentrations in the periplasmic space to enhance drug degradation and resistance.

Most strains of *Staphylococcus aureus* produce a plasmid-encoded β-lactamase that degrades penicillin but not semisynthetic penicillins, such as oxacillin and nafcillin. The greatest diversity among β-lactamases, however, is found in gram-negative bacteria. The most common and earliest identified plasmid-encoded β-lactamases of gram-negative bacteria can inactivate all penicillins and most early-generation cephalosporins. Multiple extended-spectrum β-lactamase (ESBL) variants of these early enzymes have emerged and are now widely disseminated. These ESBLs can degrade later-generation cephalosporins (ceftriaxone, cefotaxime, ceftazidime) as well as the monobactam aztreonam, and some ESBLs also degrade the fourth-generation cephalosporin cefepime. Carbapenems (imipenem, meropenem, ertapenem, doripenem) generally are not degraded by ESBLs, but additional β-lactamases, called *carbapenemases*, which degrade carbapenems and most if not all other β-lactams, have emerged and are increasing in prevalence. In the United States, *Klebsiella pneumoniae* carbapenemases (KPCs), which are usually found in strains of *Escherichia coli* and *K. pneumoniae*, are most widespread, but New Delhi metallo-β-lactamases (NDM carbapenemases), which were found initially on the Indian subcontinent, have now appeared and increased in the United States, as has an OXA group carbapenemase, OXA-48. In some cases, high levels of expression of an ESBL or an AmpC chromosomally encoded enzyme (see below), together

TABLE 150-1 The Most Common Mechanisms of Resistance to Antibacterial Agents

ANTIBACTERIAL AGENT(S)	MAJOR TARGET	MECHANISM(S) OF ACTION	MECHANISM(S) OF RESISTANCE
β-Lactams (penicillins, cephalosporins, monobactams, carbapenems)	Cell-wall synthesis	Bind cell-wall cross-linking enzymes (PBPs, transpeptidases)	Drug inactivation by β-lactamases Altered PBP targets Reduced diffusion through porin channels Altered iron uptake proteins (cefiderocol)
Glycopeptides and lipoglycopeptides (vancomycin, teicoplanin, telavancin, dalbavancin, oritavancin)	Cell-wall synthesis	Block cell wall glycosyltransferases by binding D-Ala-D-Ala stem-peptide terminus Teicoplanin, telavancin, dalbavancin, and oritavancin: affect membrane function	Altered D-Ala-D-Ala target (D-Ala-D-Lac) Increased D-Ala-D-Ala target binding at sites distant from cell wall synthesis enzymes
Bacitracin	Cell-wall synthesis	Blocks lipid carrier of cell wall precursors	Active drug efflux
Fosfomycin	Cell-wall synthesis	Blocks linkage of stem peptide to NAG by enoyltransferase	Target enzyme overexpression Drug-modifying enzymes
Aminoglycosides (gentamicin, tobramycin, amikacin, plazomicin)	Protein synthesis	Bind 30S ribosomal subunit Block translocation of peptide chain Cause misreading of mRNA	Drug-modifying enzymes Methylation at ribosome binding site Decreased permeation to target due to active efflux
Tetracyclines (tetracycline, doxycycline, minocycline)	Protein synthesis	Bind 30S ribosomal subunit Inhibit peptide elongation	Active drug efflux Ribosomal protection proteins
Tigecycline, eravacycline, omadacycline	Protein synthesis	Same as tetracyclines	Active drug efflux (pumps different from those affecting tetracyclines)
Macrolides (erythromycin, clarithromycin, azithromycin) and the ketolide telithromycin	Protein synthesis	Bind 50S ribosomal subunit Block peptide chain exit	Methylation at ribosome binding site Active drug efflux
Lincosamides (clindamycin)	Protein synthesis	Bind 50S ribosomal subunit Block peptide bond formation	Methylation at ribosome binding site
Streptogramins (quinupristin, dalfopristin)	Protein synthesis	Same as macrolides	Same as macrolides Drug-modifying enzymes
Chloramphenicol	Protein synthesis	Binds 50S ribosomal subunit Blocks aminoacyl tRNA positioning	Drug-modifying enzymes
Oxazolidinones (linezolid, tedizolid)	Protein synthesis	Bind 50S ribosomal subunit Inhibit initiation of peptide synthesis	Altered rRNA binding site Methylation of ribosome binding site
Pleuromutilins (lefamulin)	Protein synthesis	Bind 50S ribosomal subunit Blocks peptidyl transferase center	Altered L3 and L4 protein binding site Methylation of ribosome binding site
Mupirocin	Protein synthesis	Blocks isoleucyl tRNA synthetase	Acquired resistant tRNA synthetase (drug bypass) Altered native tRNA synthetase target
Sulfonamides (sulfadiazine, sulfisoxazole, and sulfamethoxazole)	Folate synthesis	Inhibit dihydropteroate synthetase	Acquired resistant dihydropteroate synthetase (drug bypass)
Trimethoprim	Folate synthesis	Inhibits dihydrofolate reductase	Acquired resistant dihydrofolate reductase (drug bypass)
Quinolones (norfloxacin, ciprofloxacin, ofloxacin, levofloxacin, moxifloxacin, gemifloxacin, delafloxacin)	DNA synthesis	Inhibit DNA gyrase and DNA topoisomerase IV Enzyme–DNA–drug complex: blocks DNA replication apparatus	Altered target(s) Active efflux Protection of target from drug Drug-modifying enzyme (ciprofloxacin)
Rifamycins (rifampin, rifabutin, rifapentine)	RNA synthesis	Inhibit RNA polymerase	Altered target
Nitrofurantoin	Nucleic acid synthesis	Reduces reactive drug derivatives that damage DNA	Altered drug-activating enzymes
Metronidazole	Nucleic acid synthesis	Reduces reactive drug derivatives that damage DNA	Altered drug-activating enzyme Acquired detoxifying enzymes Active efflux
Polymyxins (polymyxin B and polymyxin E [colistin])	Cell membrane	Bind LPS and disrupt both outer and cytoplasmic membranes	Altered cell-membrane charge with reduced drug binding
Daptomycin	Cell membrane	Produces membrane channel and membrane leakage	Altered cell-membrane charge with reduced drug binding

Abbreviations: LPS, lipopolysaccharide; NAG, *N*-acetylglucosamine; PBP, penicillin-binding protein.

with reduced porin diffusion channels, can also result in resistance to carbapenems. In *Pseudomonas aeruginosa*, resistance to carbapenems can occur by mutations that cause reductions in the OprD diffusion channel for imipenem or increased expression of efflux pumps that can remove meropenem from the bacterial cell.

The chromosomal β-lactamase of *K. pneumoniae* preferentially degrades penicillins over cephalosporins. In contrast, the chromosomal β-lactamase of *Enterobacter* and related genera, AmpC, can degrade almost all cephalosporins but is normally expressed in only small amounts. Mutations in regulatory genes that cause increased

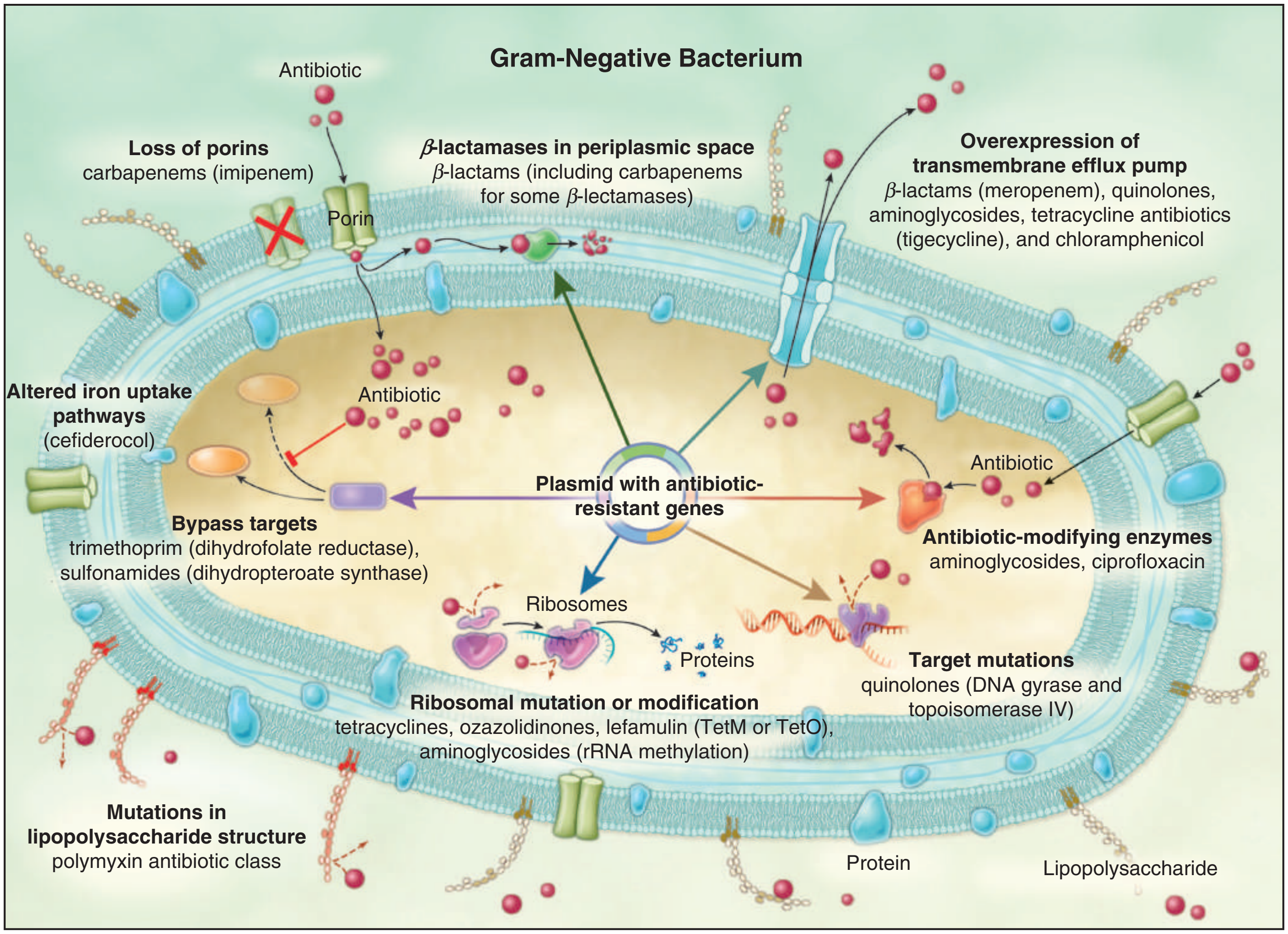
Gram-Negative Bacterium
Antibiotic
Loss of porins
carbapenems (imipenem)
Porin
β-lactamases in periplasmic space
β-lactams (including carbapenems for some β-lectamases)
Overexpression of transmembrane efflux pump
β-lactams (meropenem), quinolones, aminoglycosides, tetracycline antibiotics (tigecycline), and chloramphenicol
Altered iron uptake pathways
(cefiderocol)
Antibiotic
Antibiotic
Plasmid with antibiotic-resistant genes
Bypass targets
trimethoprim (dihydrofolate reductase), sulfonamides (dihydropteroate synthase)
Antibiotic-modifying enzymes
aminoglycosides, ciprofloxacin
Ribosomes
Proteins
Target mutations
quinolones (DNA gyrase and topoisomerase IV)
Ribosomal mutation or modification
tetracyclines, ozazolidinones, lefamulin (TetM or TetO), aminoglycosides (rRNA methylation)
Mutations in lipopolysaccharide structure
polymyxin antibiotic class
Protein
Lipopolysaccharide

FIGURE 150-1 Antibacterial targets and mechanisms of resistance to antibacterial agents, as illustrated in a gram-negative bacterium. Similar mechanisms are found in gram-positive bacteria, but their lack of an outer membrane causes β-lactamases to be excreted outside the cell, rather than into the periplasmic space between the inner and outer membranes, and reduces the efficiency of efflux pumps because exported drugs can re-enter the cell after crossing a single membrane, rather than the two membranes in gram-negative bacteria. Red spheres indicate antibiotics. *(From AY Peleg, DC Hooper: Hospital-acquired infections due to gram-negative bacteria. N Engl J Med 362:1084, 2010. Copyright © 2010 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

amounts of AmpC to be produced can cause full resistance to penicillins and cephalosporins; the exceptions are cefoxitin and cefepime, which are relatively stable to AmpC. Resistance to cefepime can develop, however, through the combined effects of mutations that cause increased AmpC production and decreased porin diffusion channels. Genes encoding AmpC have also been found on plasmids but are less common than plasmid-encoded ESBLs. A recent novel cephalosporin, cefiderocol, has enhanced stability to β-lactamases and, due to a catechol side group, is actively taken up into the bacterial cell by siderophore iron uptake pathways, rather than diffusing passively through porin channels. It is active against many gram-negative bacteria that are resistant to other β-lactams, including carbapenems. Reduced susceptibility has been reported to occur in strains with mutations in multiple iron transport genes.

Inhibitors of β-lactamases such as clavulanate, sulbactam, tazobactam, avibactam, vaborbactam, relebactam, and durlobactam have been developed and paired with amoxicillin and ticarcillin (clavulanate), ampicillin (sulbactam), piperacillin and ceftolozane (tazobactam), ceftazidime (avibactam), meropenem (vaborbactam), imipenem (relebactam), and sulbactam (durlobactam), which has selective activity against *Acinetobacter*. These inhibitors, with the exception of sulbactam, have little or no antibacterial activity of their own but inhibit plasmid-mediated β-lactamases, including ESBLs. Only avibactam, vaborbactam, and relebactam inhibit AmpC enzymes and some carbapenemases (KPCs but not metallo-carbapenemases, such as NDM).

Resistance to β-lactams also occurs through alterations in the drugs' target transpeptidase enzymes (PBPs) involved in cross-linking of the bacterial cell-wall peptidoglycan structure. In *S. pneumoniae*, *N. gonorrhoeae*, and *Neisseria meningitidis*, resistance to penicillin occurs by recombination of transformed DNA from related species that results in mosaic PBPs with lower affinity for penicillin. A combination of increased expression of an efflux pump and a porin mutation also causes penicillin resistance in *N. gonorrhoeae*. In staphylococci, resistance to methicillin and other β-lactams occurs by acquisition of the *mec* gene, which encodes a PBP2a with reduced drug affinity. PBP2a is a bypass target that can function in cell wall cross-linking in the presence of β-lactams, bypassing their effect on other PBPs. Ceftaroline is the only β-lactam that has affinity for PBP2a and is thus active against methicillin-resistant staphylococcal strains. Resistance to ceftaroline can occur, however, by mutations in the gene encoding PBP2a that reduce its affinity for the drug.

Glycopeptides and Lipoglycopeptides Glycopeptides and lipoglycopeptides inhibit bacterial cell-wall synthesis by binding to the terminal two D-alanine amino acids on the cell-wall peptidoglycan stem peptides, which are involved in peptidoglycan cross-links. In doing so, these drugs block the transpeptidase cross-linking enzymes and glycosyl transferases necessary for cell-wall synthesis. Resistance to vancomycin in enterococci is due to the acquisition of a set of *van* genes that result in (1) the production of D-alanine-D-lactate—instead of the normal D-alanine-D-alanine—at the end of the peptidoglycan stem peptide and (2) the reduction of existing D-alanine-D-alanine–terminated peptides. Vancomycin binds D-alanine-D-lactate with a 1000-fold lower affinity than D-alanine-D-alanine. The *van* genes originated in the organisms that naturally produce vancomycin and have been mobilized and reorganized in transposon mobile genetic elements and onto plasmids, which can be transferred between enterococci. In rare cases, the *van* gene cassettes have been transferred from enterococci to *S. aureus*, with the consequent generation of full vancomycin resistance. In *S. aureus*, intermediate resistance to vancomycin is more common than full vancomycin resistance and is due to a different mechanism that results from a series of several chromosomal mutations leading to a thickened and poorly cross-linked cell wall. This modified cell wall contains additional D-alanine-D-alanine–terminated stem peptides that bind vancomycin at a site distant from the cell membrane, adjacent to which new peptidoglycan is synthesized and where vancomycin binding blocks transpeptidase and transglycosylase enzymes. Thus, vancomycin's binding to these distant termini impedes its access to the proximal binding sites that result in inhibition of peptidoglycan synthesis. This intermediate-resistance phenotype was first recognized in patients receiving prolonged courses of vancomycin that created an opportunity for selection of the multiple mutations needed to produce the modified cell wall. Because of the energy costs of a thickened cell wall, this intermediate-resistance phenotype may be unstable, with strains returning to susceptibility in the absence of vancomycin selection pressure. Susceptibility to telavancin, dalbavancin, and oritavancin is also reduced in strains that exhibit resistance or intermediate susceptibility to vancomycin, although in some cases, the drugs remain sufficiently active that the strains may still be classified as susceptible based on standard clinical laboratory interpretive criteria.

Aminoglycosides Aminoglycosides are one of several classes of antimicrobials that inhibit protein synthesis by binding to either the 30S or the 50S bacterial ribosomal subunit (both of which differ from eukaryotic ribosomal subunits), with consequent selective antibacterial activity. The aminoglycosides bind to the 30S subunit of the bacterial ribosome. The most common mechanism of resistance to aminoglycosides in gram-negative bacteria is due to acquisition of plasmids with genes encoding transferase enzymes that modify aminoglycosides by the addition of acetyl, adenyl, or phosphate groups; these added groups decrease the drugs' binding affinity to their ribosomal target site. Various transferases differ in which aminoglycosides they modify, and amikacin resistance occurs less often than resistance to gentamicin or tobramycin by these mechanisms. Plazomicin, a recently developed aminoglycoside, is distinctive in that it remains active and is not modified by most transferases. Another mechanism of plasmid-mediated aminoglycoside resistance is due to methylase enzymes that can methylate the site of aminoglycoside binding on the 16S ribosomal RNA of the 30S ribosomal subunit and reduce drug binding to its ribosome target, resulting in resistance to all aminoglycosides, including plazomicin. For streptomycin, a single ribosomal protein mutation may also cause resistance. In *P. aeruginosa*, aminoglycoside resistance can also occur through mutations in regulatory genes causing increased expression of a chromosomally encoded efflux pump, MexXY, which reduces intracellular drug concentrations.

Tetracyclines These antibiotics bind the 16S ribosomal RNA of the 30S ribosomal subunit at a site distinct from the binding site of the aminoglycosides and inhibit bacterial protein synthesis. For tetracyclines, including doxycycline and minocycline, resistance is often plasmid-mediated and due either to active efflux pumps, which in some cases are specific for tetracyclines, or to proteins that protect the ribosome from tetracycline action. Some broad-spectrum, chromosomally encoded efflux pumps may also include tetracyclines among their substrates, and regulatory mutations that cause pump overexpression may confer tetracycline resistance together with resistance to other agents that are pump substrates. There have been recent derivatives of the tetracyclines, including the glycylcycline, tigecycline, the fluorocycline, eravacycline, and the aminomethylcycline, omadacycline, which have modifications on the core tetracycline ring structure rendering them less affected or unaffected by the common tetracycline resistance mechanisms. Resistance to the newer tetracyclines can occur, however, through mutations that cause overexpression of some broad-spectrum efflux pumps, particularly in *Proteus* species. An uncommon plasmid-encoded tetracycline modification mechanism can also cause resistance to the newer agents.

Macrolides, Ketolides, Lincosamides, and Streptogramins These antibiotics are also inhibitors of bacterial protein synthesis, in this case through their binding to the 23S RNA of the 50S ribosomal subunit. They are generally active against gram-positive bacteria. Resistance to macrolides, clindamycin, and quinupristin is most often due to acquired Erm methylases that modify the drug-binding site on the ribosome, reducing drug binding. Resistance to quinupristin by this mechanism renders the quinupristin-dalfopristin combination bacteriostatic rather than bactericidal. Telithromycin, a ketolide structurally related to macrolides, has an additional binding site on the ribosome and remains active in the presence of some methylases. Methylase gene expression can be induced by exposure to most macrolides but generally not ketolides (e.g., telithromycin); however, bacterial strains constitutively expressing methylase genes can display resistance to both macrolides and ketolides. Acquired genes encoding active efflux pumps can also contribute to resistance to macrolides in streptococci and to resistance to macrolides, clindamycin, and dalfopristin in staphylococci. Plasmid-acquired, drug-modifying enzymes in staphylococci can also cause resistance to quinupristin and dalfopristin. Macrolide resistance due to 23S rRNA mutations at the site of drug binding is uncommon in staphylococci and streptococci because of the multiple copies of the rRNA genes on the chromosomes of these species; such resistance mechanisms may occur more frequently, however, in mycobacteria, *Helicobacter pylori*, and *Treponema* species, which have only one or two chromosomal copies of these rRNA genes, thus requiring fewer mutations to generate a resistance phenotype. Among gram-negative bacteria, many of which are not susceptible to current macrolides because of inadequate drug permeation across the outer membrane, some strains with acquired genes for macrolide-modifying enzymes have been described.

Chloramphenicol Chloramphenicol inhibits bacterial protein synthesis by binding to the 23S rRNA of the 50S subunit at a site that overlaps the macrolide-binding site. Chloramphenicol is uncommonly used in human medicine because of infrequent but potentially severe bone marrow toxicity. Resistance to chloramphenicol is most often due to plasmid-encoded, drug-modifying acetyltransferases that have been found in both gram-positive and gram-negative bacteria and whose expression can be induced by drug exposure. Among staphylococci, some resistant strains have been found to have a plasmid-encoded ribosomal methylase that confers resistance to chloramphenicol, clindamycin, and oxazolidinones. As is the case for macrolides, ribosomal mutations causing resistance to chloramphenicol are uncommon because of multiple copies of rRNA genes in the common gram-positive human pathogens. Plasmid-encoded efflux pumps affecting chloramphenicol specifically have been found in gram-negative bacteria, and other pumps affecting chloramphenicol and oxazolidinones have been found in gram-positive bacteria.

Oxazolidinones Linezolid and tedizolid are the only members of the oxazolidinone class of antimicrobials in clinical use, and both are active against gram-positive bacteria only; lack of sufficient activity in gram-negative bacteria results from the ability of native efflux pumps in these bacteria to limit drug access to their cytoplasmic ribosome targets. Oxazolidinones target the bacterial ribosome and inhibit protein synthesis by binding to 23S rRNA of the 50S subunit at a distinct site that overlaps with the chloramphenicol-binding site. Resistance has been seen in enterococci more often than in staphylococci and, in both organisms, is most often due to mutations in multiple copies of the 23S rRNA genes that reduce drug binding to the ribosome. A plasmid-acquired ribosomal methylase gene that enables ribosomal alteration at a site that confers resistance to both linezolid and chloramphenicol has also been found in some strains of both *S. aureus* and coagulase-negative staphylococci but is not yet widespread. A plasmid-encoded active efflux pump conferring resistance to oxazolidinones (both linezolid and tedizolid) and chloramphenicol has been described in animal isolates and a small number of human isolates of *Enterococcus faecalis*.

Pleuromutilins Lefamulin is the only systemic pleuromutilin in clinical use. Retapamulin has been available for topical use in skin infections. Pleuromutilins inhibit bacterial protein synthesis by binding to the peptidyl transferase center in the 50S ribosomal subunit, and lefamulin is generally active against gram-positive bacteria, *Haemophilus influenzae*, *Moraxella catarrhalis*, and atypical respiratory pathogens, such a *Mycoplasma pneumoniae* and *Legionella* spp. Although there is partial overlap in the site of binding of lefamulin and those of other antibacterials binding the 50S ribosomal subunit, cross-resistance with macrolides, oxazolidinones, lincosamides, and streptogramins is uncommon. Resistance to lefamulin can occur by mutations in L3 and L4 proteins of the 50S subunit that alter the lefamulin binding site. In addition, the plasmid-encoded Cfr methylase, which confers resistance to chloramphenicol and oxazolidinones, can also cause resistance to lefamulin by disrupting its binding site. Vga efflux transporters, which cause resistance to lincosamides and streptogramins, also affect pleuromutilins.

Mupirocin Mupirocin is used only in topical formulations, most often for elimination of nasal carriage of *S. aureus*. It targets bacterial leucyl-tRNA synthetase and inhibits protein synthesis. Resistance to mupirocin occurs by either mutation in the target leucyl-tRNA synthetase (low-level resistance) or the acquisition of a plasmid-encoded resistant tRNA synthetase (high-level resistance), which bypasses drug inhibition of the native, sensitive synthetase.

Sulfonamides and Trimethoprim These agents inhibit the folate biosynthesis pathway at different steps. Sulfonamides are structurally similar to *para*-aminobenzoic acid (PABA) and competitively inhibit dihydropteroate synthetase, which, in an early step in the pathway, uses PABA to synthesize dihydropteroate, a precursor of dihydrofolate. Trimethoprim inhibits dihydrofolate reductase at a later step in the pathway that generates tetrahydrofolate. Clinical use of folate pathway inhibitors most often consists of the combination of sulfamethoxazole and trimethoprim; on occasion, however, trimethoprim or various sulfonamides are used individually. Resistance to both of these antimetabolites can result from mutation in their target enzymes or can be due to plasmid-acquired genes encoding resistant enzymes that bypass the inhibition of the native sensitive enzymes—a resistant dihydropteroate synthetase in the case of sulfonamides and a resistant dihydrofolate reductase in the case of trimethoprim. Resistance to the combination of sulfamethoxazole and trimethoprim requires that the bacterial strain have resistance mechanisms for both agents and yet is relatively common. Resistance due to drug efflux or drug modification has been limited for both sulfonamides and trimethoprim.

Quinolones Quinolones are synthetic inhibitors of bacterial DNA synthesis. They bind to two enzymes required for DNA synthesis: DNA gyrase and DNA topoisomerase IV, which alter DNA conformation and the interlinking of replicated molecules. In addition to inhibiting the enzymes' catalytic functions of altering DNA topology, they stabilize enzyme–DNA complexes that form a barrier to the DNA replication machinery and are a precursor to lethal double-strand DNA breaks. Although related topoisomerase enzymes are involved in mammalian DNA synthesis, the mammalian and bacterial enzymes are sufficiently different from each other for quinolones to have selective activity against bacteria. Resistance to quinolones is most often due either to chromosomal mutations altering the target enzymes DNA gyrase and DNA topoisomerase IV, with consequent reduction in drug binding, or to mutations that increase the expression of native broad-spectrum efflux pumps for which quinolones (among other compounds) are substrates. In addition, three types of acquired genes can confer reduced susceptibility or low-level resistance by protecting the target enzymes, modifying some quinolones (particularly ciprofloxacin and norfloxacin) to reduce their activity, or generating an efflux of quinolones. These genes are usually located on multidrug-resistance plasmids that have spread worldwide. Their presence can promote higher levels of quinolone resistance by enhancing selection of the mutations in chromosomal target genes with exposure to quinolones and can then link quinolone resistance to resistance to other antibacterial drugs that are encoded by the same plasmid.

Rifampin, Rifabutin, and Rifapentine Antimicrobials of the rifamycin class target bacterial RNA polymerase and thereby inhibit transcription of messenger RNA and gene expression. Their activity is generally limited to gram-positive bacteria because native efflux pumps in most gram-negative bacteria reduce drug access to the cytoplasmic enzyme target. Single mutations in the β subunit of RNA polymerase constitute the principal mechanism of acquired rifampin resistance, which is high-level. Thus, rifampin and other rifamycins are used for treatment of infections only in combination with other antibacterial drugs in order to reduce the likelihood of selection of high-level resistance.

Metronidazole Metronidazole is actively taken up by most anaerobic bacteria and then converted to reactive drug derivatives that nonspecifically damage cytoplasmic proteins and nucleic acids. Thus, metronidazole lacks a specific cellular target. Acquired resistance to metronidazole in *Bacteroides* species is rare. Such resistance has been reported in strains that lack the endogenous activating nitroreductase or that have acquired *nim* genes responsible for further reduction of the DNA-damaging nitroso intermediates to an inactive derivative. Active efflux and enhanced DNA repair mechanisms also have been associated with resistance.

Nitrofurantoin Nitrofurantoin is used only for treatment of lower urinary tract infections because adequate drug concentrations are found only in urine. Its mechanism of action is not fully understood but is thought to involve generation of reactive derivative molecules (as occurs with metronidazole) that damage DNA and ribosomes. Resistance to nitrofurantoin in *E. coli* can emerge through a series of mutations that progressively decrease the nitroreductase activity required for generating active nitrofuran metabolites. These mutants are also impaired in growth; this impairment possibly explains the infrequent occurrence of resistance with clinical use of nitrofurantoin.

Polymyxins Because of emerging multidrug resistance in gram-negative bacteria, colistin and polymyxin B have been used for infections due to resistant Enterobacterales, *P. aeruginosa*, and *Acinetobacter* species. Polymyxins are cationic cyclic peptide molecules that bind negatively charged lipopolysaccharides on the gram-negative bacterial outer membrane, with subsequent disruption and permeabilization of both outer-membrane and cytoplasmic-membrane structure. Thus, the polymyxins are bactericidal. Resistance is so far uncommon but can emerge during therapy through mutations that cause reductions in the negative charge of the gram-negative bacterial cell surface, thereby reducing binding of the positively charged colistin. Transferable plasmid-mediated colistin resistance has also been found to be due to *mcr-1* and other *mcr* variants that encode a gene for a phosphoethanolamine transferase that also reduces the negative charge on the cell surface. *mcr*-containing enteric bacteria have now been identified in Asia, Europe, and the United States.

Daptomycin Daptomycin is active against gram-positive bacteria and interacts with and disrupts the cytoplasmic membrane in a calcium-dependent manner, resulting in bactericidal activity. The mechanisms of resistance to daptomycin are complex and involve mutations in several genes that can alter cell membrane charge and structure and reduce daptomycin binding. Resistance to daptomycin is relatively infrequent but has emerged in some *S. aureus* strains with intermediate vancomycin susceptibility from patients treated with vancomycin but not exposed to daptomycin. In some strains of methicillin-resistant *S. aureus*, daptomycin resistance has been linked to acquired susceptibility to β-lactams; combinations of daptomycin with nafcillin or ceftaroline have been successful for treatment of patients infected with resistant strains when daptomycin alone or in combination with other agents has failed. The mechanism of this effect is not yet clear but may involve alteration in surface charge and increased daptomycin binding in the presence of β-lactams. Daptomycin resistance has also been reported in enterococci.

EPIDEMIOLOGY OF RESISTANCE AND REDUCTION OF ITS OCCURRENCE

Multidrug resistance in human bacterial infections has been increasing overall in recent years, substantially limiting the number of antibiotics that can be used to treat some infections. The prevalence of resistance to various antimicrobials among human pathogens can, however, vary greatly in different geographic areas and even at different institutions in the same area. Thus, specific local data on the occurrence of various types of resistance are an important component of the choice of antimicrobials for empirical treatment of infection until the responsible pathogen is identified and its specific susceptibilities are determined by the clinical microbiology laboratory. Prompt adjustment of the initially chosen antimicrobial on the basis of species and susceptibility data to best target therapy is equally important. These principles emphasize the importance of obtaining appropriate samples for culture or other diagnostic modalities and susceptibility testing—whenever possible, prior to administration of antimicrobials. They also highlight the importance of rapid and sensitive diagnostic methods and the prompt communication of their results to clinicians to inform best choices of antimicrobials.

The overall prevalence of resistance can be affected by a number of factors, including (1) the extent of resistance reservoirs in the patient population; (2) the selection pressures from use of antimicrobials that favor resistant strains over susceptible ones; and (3) the extent by which resistance is amplified by transmission of resistant strains to patients from their environment or other persons, either directly or indirectly via the contaminated hands of health care workers when hand hygiene and other infection control practices are inadequately followed. The likelihood that an individual patient will be infected with a resistant pathogen is likewise affected by his or her history. Studies have shown that prior antibiotic treatment, prior infection with resistant pathogens, and prior hospitalizations all increase this likelihood.

These factors emphasize the importance of the appropriate use of antimicrobials (particularly, the avoidance of their use in clinical conditions in which they are not needed), the use of the shortest courses of therapy sufficient for a successful clinical outcome, and the implementation of antimicrobial stewardship programs **(Chap. 149)** as well as careful and consistent infection control practices in short-term and long-term-care institutions. Antimicrobial agents are distinct among drug classes in human medicine in that—despite their clear clinical value when used appropriately—the extent of their use can compromise their future utility because of resistance. The remarkable ability of pathogens to acquire resistance is inherent in their biology and emphasizes the necessity for clinicians and institutions to pay careful attention to those factors that can be controlled through judicious antimicrobial use and rigorous infection control and prevention practices.

Efforts to address the problems caused by resistance are now being made worldwide. The U.S. Centers for Disease Control and Prevention (CDC) has recently estimated that >2.8 million resistant bacterial infections occur in the United States each year, with 35,900 deaths, and has identified particular resistant pathogens that are of greatest concern because of their overall effects on public health **(Table 150-2)**. Enteric bacteria (such as *E. coli*, *K. pneumoniae*, and *Enterobacter* spp.) and *Acinetobacter* spp. that are resistant to carbapenems are included in the "urgent" category because of their increasing occurrence worldwide and because they are often highly resistant to multiple drugs, with few if any active antimicrobials available for treatment. Resistant *N. gonorrhoeae* is included in this category as well because of the ease with which gonorrhea can be spread from person to person and because limited active agents are now available. Other resistances are common and also affect clinical care, often requiring use of alternatives to first-line agents that can be less effective and less well tolerated. Also affecting clinical care and considered urgent are infections due to *Clostridioides difficile*. Although not directly due to acquired resistance, *C. difficile* disease, like resistant bacterial infections, is linked to antibacterial use (by the disruption of the normal microbiome of the gastrointestinal tract rather than direct resistance selection) and to its ability as a spore-forming bacterium to be spread in health care environments. To address the problems posed by resistance and *C. difficile* disease, the CDC has emphasized a set of five core actions. (1) *Infection prevention and control:* These efforts focus on implementation of evidence-based activities to reduce the risks and incidence of device-related infections overall and on improvement of compliance with infection control practices that prevent transmission of resistant pathogens from one person to another, such as hand hygiene and isolation precautions in health care and long-term care settings. (2) *Tracking and data:* Efforts aim to increase the reporting and sharing of the occurrence of resistance to enhance epidemiologic data and inform targeting of preventive interventions. (3) *Antibiotic use and access:* Antimicrobial stewardship programs with specific components to track usage and educate clinicians on appropriate use have become required in hospitals, and the CDC has implemented efforts to reduce inappropriate use in outpatient settings, with particular attention to upper respiratory illnesses that often do not require antimicrobials because of their common self-limited viral causes. (4) *Vaccines, therapeutics, and diagnostics:* The U.S. Congress and the U.S. Food and Drug Administration, as well as agencies in other countries, have recently developed incentives and enhanced regulatory pathways for drug approval that pharmaceutical companies can use for development of antimicrobials that specifically address particular resistant pathogens. Both small and large companies have undertaken efforts in this area. New technologies for rapid detection of resistance and susceptibility are also being developed by multiple diagnostics companies in order to facilitate the appropriate choice of antimicrobials earlier in the course of illness, providing an important tool for antimicrobial stewardship programs. (5) *Environment and sanitation:* Reservoirs of resistant bacteria and resistance genes on mobile genetic elements can exist in agriculture and food production and domestic animals and have the potential for introduction into humans. Thus, antibiotic use in these environments can amplify resistance reservoirs and increase the chance of human exposure. Therefore, public health interventions addressing these issues in a One Health approach (*https://www.cdc.gov/onehealth/index.html*) are an important component of managing resistance risks.

TABLE 150-2 Antibiotic Resistance Threats in the United States, 2019

THREAT CATEGORY	ORGANISMS
Urgent	Carbapenem-resistant *Acinetobacter* *Candida auris* *Clostridioides difficile* Carbapenem-resistant Enterobacterales Drug-resistant *Neisseria gonorrhoeae*
Serious	Drug-resistant *Campylobacter* Drug-resistant *Candida* Extended-spectrum β-lactamase–producing Enterobacterales Vancomycin-resistant *Enterococcus* Multidrug-resistant *Pseudomonas aeruginosa* Drug-resistant nontyphoidal *Salmonella* Drug-resistant *Salmonella* serotype Typhi Drug-resistant *Shigella* Methicillin-resistant *Staphylococcus aureus* Drug-resistant *Streptococcus pneumoniae* Drug-resistant *Mycobacterium tuberculosis*
Concerning	Erythromycin-resistant group A *Streptococcus* Clindamycin-resistant group B *Streptococcus*
Watch List	Azole-resistant *Aspergillus fumagatis* Drug-resistant *Mycoplasma genitalium* Drug-resistant *Bordetella pertussis*

Source: U.S. Centers for Disease Control and Prevention.

FURTHER READING

Bush K, Bradford PA: Interplay between β-lactamases and new β-lactamase inhibitors. Nat Rev Microbiol 17:295, 2019.

Centers for Disease Control and Prevention: Antibiotic resistance threats in the United States, 2019. Available at *https://www.cdc.gov/antimicrobial-resistance/media/pdfs/2019-ar-threats-report-508.pdf*. Accessed June 23, 2020.

French GL: Antimicrobial resistance and healthcare-associated infections, in *Hospital Epidemiology and Infection Control*, 4th ed. GC Mayhall (ed). Philadelphia, Lippincott Williams & Wilkins, 2012, pp. 1297–1310.

Rice LB: Mechanisms of resistance and clinical relevance of resistance to β-lactams, glycopeptides, and fluoroquinolones. Mayo Clin Proc 87:198, 2012.

Silver LL, Bush K (eds): *Antibiotics and Antibiotic Resistance*. Cold Spring Harbor Perspectives in Medicine. New York, Cold Spring Harbor Laboratory Press, 2016.

Section 5 Diseases Caused by Gram-Positive Bacteria

151

Pneumococcal Infections

David Goldblatt, Katherine L. O'Brien

In the late nineteenth century, pairs of micrococci were first recognized in the blood of rabbits injected with human saliva by both Louis Pasteur, working in France, and George Sternberg, an American army physician. The important role of these micrococci in human disease was not appreciated at that time. By 1886, when the organism was designated "pneumokokkus" and *Diplococcus pneumoniae*, it had been isolated by many independent investigators, and its role in the etiology of pneumonia was well known. In the 1930s, pneumonia was the third leading cause of death in the United States (after heart disease and cancer) and was responsible for ~7% of all deaths both in the United States and in Europe. While pneumonia was caused by a host of pathogens, lobar pneumonia—a pattern more likely to be caused by the pneumococcus—accounted for approximately one-half of all pneumonia deaths in the United States in 1929. In 1974, the organism was reclassified as *Streptococcus pneumoniae*.

MICROBIOLOGY

Etiologic Agent Pneumococci are spherical gram-positive bacteria of the genus *Streptococcus*. Within this genus, cell division occurs along a single axis, and bacteria grow in chains or pairs—hence the name *Streptococcus*, from the Greek *streptos*, meaning "twisted," and *kokkos*, meaning "berry." At least 22 streptococcal species are recognized and are divided further into groups based on their hemolytic properties. *S. pneumoniae* belongs to the α-hemolytic group that characteristically produces a greenish color on blood agar because of the reduction of iron in hemoglobin **(Fig. 151-1)**. The bacteria are fastidious and grow best in 5% CO_2 but require a source of catalase (e.g., blood) for growth on agar plates, where they develop mucoid (smooth/shiny) colonies. Pneumococci without a capsule produce colonies with a rough surface. Unlike that of other α-hemolytic streptococci, their growth is inhibited in the presence of optochin (ethylhydrocupreine hydrochloride), and they are bile soluble.

In common with other gram-positive bacteria, pneumococci have a cell membrane beneath a cell wall, which in turn is covered by a polysaccharide capsule. Pneumococci are divided into serogroups or serotypes based on capsular polysaccharide structure, as distinguished with rabbit polyclonal antisera; capsules swell in the presence of specific antiserum (the Quellung reaction). The most recently discovered serotypes—6C, 6D, 6F, 6G, 6H, 10D, 11E, 20A, 20B, and 35D—have been identified with monoclonal antibodies and by serologic, genetic, and biochemical means. The currently recognized 100 serotypes fall into 21 serogroups, and each serogroup contains two to eight serotypes with closely related capsules. Detailed genetic analysis of the locus coding for the polysaccharide capsule, the *cps* locus, continues to reveal putative novel capsular polysaccharides, variants within existing serogroups that are designated with an "X." In the absence of type-specific antibody, the capsule protects the bacteria from phagocytosis by host cells and is arguably the most important determinant of pneumococcal virulence. Unencapsulated variants are occasionally identified in cases of invasive pneumococcal disease; however, when their genotype is assessed, they often contain capsular genes. Thus it is likely that they were encapsulated in vivo and have stopped producing capsule during the laboratory steps of pathogen isolation.

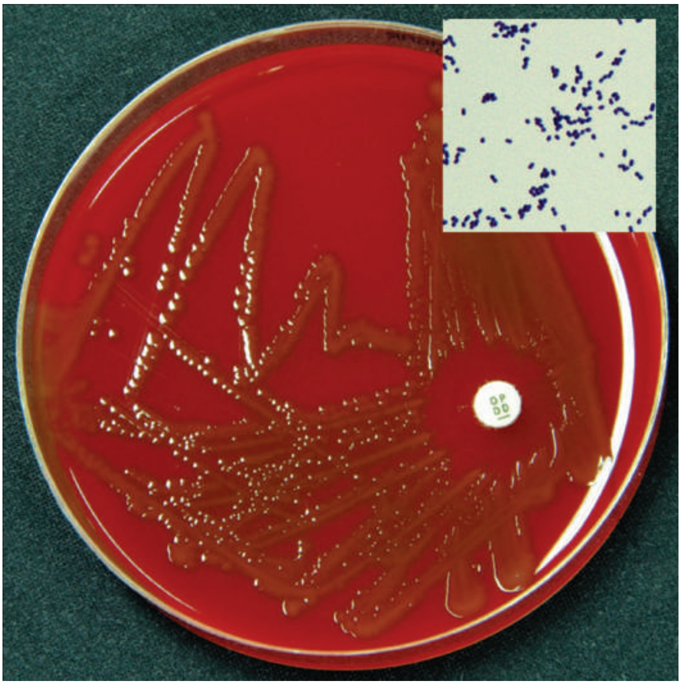

FIGURE 151-1 Pneumococci growing on blood agar, illustrating α hemolysis and optochin sensitivity (zone around optochin disk). *Inset:* Gram's stain, illustrating gram-positive diplococci. *(Photographs courtesy of Paul Turner, University of Oxford, United Kingdom.)*

Virulence Factors Within the cytoplasm, cell membrane, and cell wall, many molecules that may play a role in pneumococcal pathogenesis and virulence have been identified **(Fig. 151-2)**. These proteins are often involved in direct interactions with host tissues or in concealment of the bacterial surface from host defense mechanisms. Pneumolysin (PLY) is a secreted cytotoxin thought to result in cytolysis of cells and tissues, and LytA enhances pathogenesis. A number of cell wall proteins interfere with the complement pathway, thus inhibiting complement deposition and preventing lysis and/or opsonophagocytosis. The pneumococcal H inhibitor (Hic) impedes the formation of C3 convertase, while pneumococcal surface protein C (PspC), also known as choline-binding protein A (CbpA), binds factor H and is thought to accelerate the breakdown of C3. PspA and CbpA inhibit the deposition of or degrade C3b. To avoid clearance by the mucus, pneumococci utilize the matrix metalloprotease ZmpA, which cleaves mucosal IgA to evade complement activation, preventing agglutination and thus clearance by the mucociliary flow.

The numerous pneumococcal proteins thought to be involved in adhesion include pneumococcal surface adhesin A (PsaA) and the exoglycosidases such as neuraminidase (NanA), β-galactosidase (BgaA), and β-N-Acetylglucosaminidase (StrH), which deglycosylate host glycoproteins releasing sugars as a nutrient source and exposing hidden receptors for adhesion. Once through the epithelial barrier, pneumococci utilize PLY and mannose receptor C type lectin 1 (MRC-1/CD206) on the surface of dendritic cells and macrophages to enter cells, where they may survive intracellularly in vacuoles thus

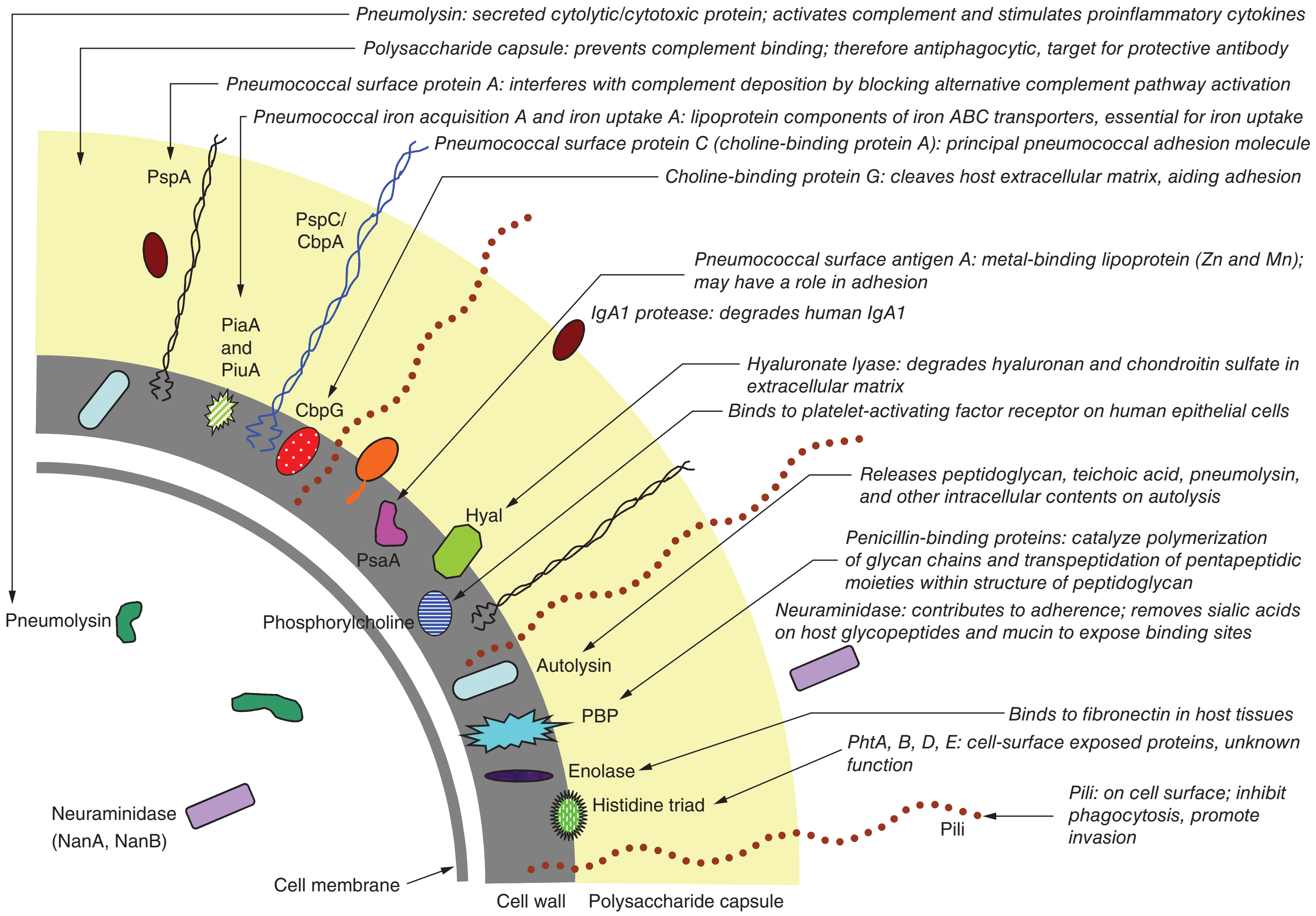
Pneumolysin: secreted cytolytic/cytotoxic protein; activates complement and stimulates proinflammatory cytokines
Polysaccharide capsule: prevents complement binding; therefore antiphagocytic, target for protective antibody
Pneumococcal surface protein A: interferes with complement deposition by blocking alternative complement pathway activation
Pneumococcal iron acquisition A and iron uptake A: lipoprotein components of iron ABC transporters, essential for iron uptake
Pneumococcal surface protein C (choline-binding protein A): principal pneumococcal adhesion molecule
Choline-binding protein G: cleaves host extracellular matrix, aiding adhesion
PspA
PspC/
CbpA
Pneumococcal surface antigen A: metal-binding lipoprotein (Zn and Mn); may have a role in adhesion
IgA1 protease: degrades human IgA1
PiaA
and
PiuA
Hyaluronate lyase: degrades hyaluronan and chondroitin sulfate in extracellular matrix
CbpG
Binds to platelet-activating factor receptor on human epithelial cells
Releases peptidoglycan, teichoic acid, pneumolysin, and other intracellular contents on autolysis
Hyal
PsaA
Penicillin-binding proteins: catalyze polymerization of glycan chains and transpeptidation of pentapeptidic moieties within structure of peptidoglycan
Pneumolysin
Phosphorylcholine
Neuraminidase: contributes to adherence; removes sialic acids on host glycopeptides and mucin to expose binding sites
Autolysin
PBP
Binds to fibronectin in host tissues
Enolase
PhtA, B, D, E: cell-surface exposed proteins, unknown function
Histidine triad
Pili: on cell surface; inhibit phagocytosis, promote invasion
Neuraminidase
(NanA, NanB)
Pili
Cell membrane
Cell wall
Polysaccharide capsule

FIGURE 151-2 Schematic diagram of the pneumococcal cell surface, with key antigens and their roles highlighted.

facilitating spread. To outcompete the other co-colonizing bacteria, the pneumococcus produces bacteriocins called pneumocins that mediate intraspecific competition. Some of the antigens mentioned above are potential vaccine candidates (see "Prevention," below). Biofilm production by pneumococci is now well recognized and is likely to be an important mechanism aiding survival of pneumococci in the upper respiratory tract and contributing to local disease manifestations such as otitis media.

Although the capsule surrounding the cell wall of *S. pneumoniae* is the basis for categorization by serotype, the disease potential of a serotype is also related to the genetic composition of the strain. Molecular genotyping and epidemiology are therefore essential. Conventionally, multilocus sequence typing (MLST) was the gold-standard technique for epidemiologic analyses due to its simplicity and effectiveness. Alleles at seven loci are sequenced, compared with all of the known alleles at that locus and a unique sequence type (ST) assigned using the pneumococcal MLST website (*pubmlst.org/organisms/streptococcus-pneumoniae/*). With the advent of high-throughput and relatively inexpensive sequencing techniques, whole-genome sequencing has facilitated even more precise molecular epidemiology: enhanced genomic epidemiology can be performed using ribosomal MLST (rMLST; >50 ribosomal genes) or core genome MLST (cgMLST; >1300 core genes) with assignment via the PubMLST website, while *k*-mer–based methods fully exploit core and accessory genomic variation (PopPUNK, *github.com/bacpop/PopPUNK*) to assign Global Pneumococcal Sequence Clusters (GPSCs) via the Global Pneumococcal Sequencing project website (*pneumogen.net*).

Twenty-three years after the first pneumococcal genome was sequenced, >50,000 pneumococcal genomes are now available on public nucleotide sequence databases. There are two curated pneumococcal genome databases: the PubMLST Pneumococcal Genome Library contains ~33,000 curated, published, assembled genomes and isolate provenance data (*pubmlst.org/organisms/streptococcus-pneumoniae/pgl/*), and the GPS Database Monocle Dataviewer, which contains ~21,000 high-quality genomes with epidemiologic data (*www.pneumogen.net/gps/gps-database-overview/*). These databases are publicly available and free to access. Over the past decade, web applications such as Pathogenwatch (*pathogen.watch/*) have provided a user-friendly platform to enable fast and reliable analysis of pneumococcal genome data without the requirement for bioinformatic expertise. Users can simply drag and drop genome data into a browser to obtain details such as serotype, genotype (including GPSC and MLST), and antimicrobial resistance profiles. In recent years, genome sequence analyses have made major contributions to the understanding of pneumococcal molecular epidemiology, biology, diversity, pathogenicity, and vaccine impact.

EPIDEMIOLOGY

(See also "Global Health," below.) Pneumococcal infections remain a significant global cause of morbidity and death, particularly among children and the elderly. Rapid and dramatic changes in the epidemiology of this disease during the past 20 years in several developed countries followed the licensure and routine childhood administration of pneumococcal polysaccharide–protein conjugate vaccine (PCV). With PCV introduction in low- and middle-income countries (LMIC), additional profound changes in pneumococcal ecology and disease epidemiology are occurring. The disease burden and serotype distribution in the PCV era are influenced not only by the reduction in disease caused by serotypes included in PCV but also by serotype replacement as a result of reductions in vaccine serotypes, concomitant secular trends in pneumococcal strains unrelated to vaccine use, the impact of antibiotic use on pneumococcal strain ecology, and surveillance system attributes that can themselves affect analysis of epidemiologic features of pneumococcal strains and disease.

Serotype Distribution Not all pneumococcal serotypes are equally likely to cause disease; observed serotype distributions vary by age category, disease syndrome, and geography. Geographic differences may be driven by variations in the relative prevalence of syndromes causing disease rather than by true serotype distribution differences, as certain serotypes are more common causes of some syndromes than others (e.g., pneumonia and meningitis). Most data on serotype distribution come from pediatric invasive pneumococcal disease (IPD, defined as infection of a normally sterile site); much less information on global or regional serotype distributions is available for disease in adults. In the era before PCV use, five to seven serotypes caused >60% of IPD cases among children <5 years of age in most parts of the world; seven serotypes (1, 5, 6A, 6B, 14, 19F, and 23F) accounted for ~60% of such cases in all areas of the world, but in any given region these seven serotypes may not all rank as the most common disease strains. Some serotypes (e.g., types 1 and 5) not only tend to cause disease in areas with a high disease burden but also cause waves of disease in lower-burden areas (e.g., Europe) or outbreaks (e.g., in military barracks; meningitis in sub-Saharan Africa). The widespread use of pneumococcal conjugate vaccines has significantly altered serotype-specific epidemiology, with some of the serotypes identified above now causing little invasive disease in countries with mature vaccine programs and emerging serotypes, not prominent in the pre–conjugate vaccine era, appearing as important causes of invasive disease. These include serotypes such as 15BC, 22F, 10A, 23B, 12F, 15A, and 8 while some serotypes included in PCVs such as 3 and 19A continue to cause IPD.

Nasopharyngeal Carriage Pneumococci are intermittent inhabitants of the healthy human nasopharynx and are transmitted by respiratory droplets. In children, pneumococcal nasopharyngeal ecology varies by geographic region, socioeconomic status, climate, degree of crowding, and particularly intensity of exposure to other children, with children in day-care settings having higher rates of colonization. In developed-world settings, children serve as the major vectors of pneumococcal transmission. By 1 year of age, ~50% of children have had at least one episode of pneumococcal colonization. Cross-sectional prevalence data show rates of pneumococcal carriage ranging from 20% to 50% among children <5 years of age and from 5% to 15% among young and middle-aged adults; **Fig. 151-3** shows relevant data from the United Kingdom. Data on colonization rates among healthy elderly individuals are limited. In LMICs, pneumococcal acquisition occurs much earlier, sometimes within the first few days after birth, and nearly all infants have had at least one episode of colonization by 2 months of age. Cross-sectional studies show that up to the age of 5 years, 70–90% of children carry *S. pneumoniae* in the nasopharynx, and a significant proportion of adults (sometimes >40%) also are colonized. Their high rates of colonization make adults an important source of transmission and may affect community transmission dynamics.

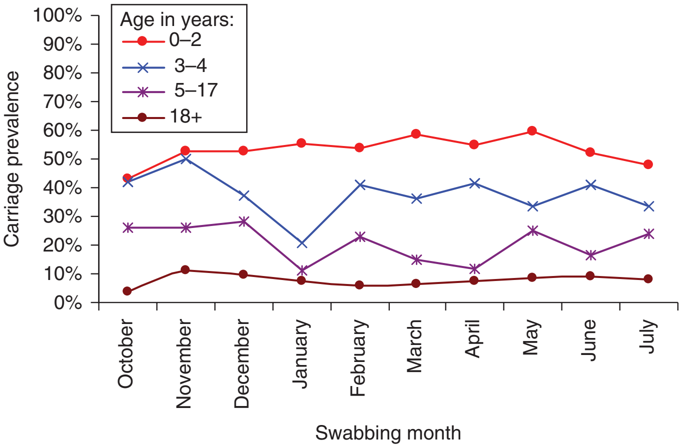

FIGURE 151-3 Prevalence of pneumococcal carriage in adults and children resident in the United Kingdom who had nasopharyngeal swabs collected monthly for 10 months (no seasonal trend; *t* test trend, >.05). *(Data adapted from D Goldblatt et al: J Infect Dis 192:387, 2005.)*

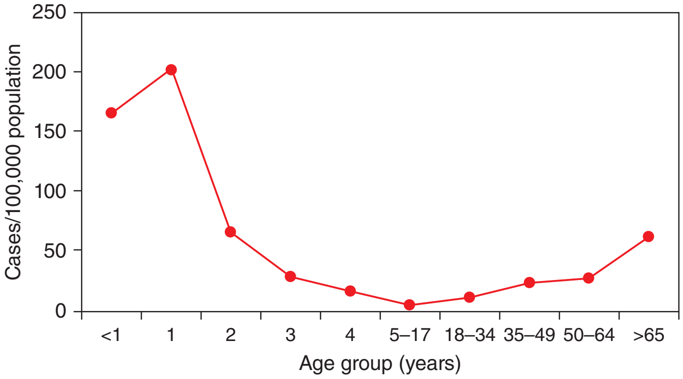

FIGURE 151-4 Rates of invasive pneumococcal disease before the introduction of pneumococcal conjugate vaccine, by age group: United States, 1998. *(Source: CDC, Active Bacterial Core Surveillance/Emerging Infectious Program Network, 2000. Data adapted from MMWR 49[RR-9], 2000.)*

Invasive Disease and Pneumonia IPD develops when *S. pneumoniae* invades the bloodstream and seeds other organs or directly reaches the cerebrospinal fluid (CSF) by local extension. Pneumonia may follow aspiration of pneumococci, although only 10–30% of pneumococcal pneumonia cases are associated with a positive blood culture (and thus contribute to the measured burden of IPD). The substantial variation of IPD rates with age is illustrated by data from the United States for 1998–1999, a period prior to PCV introduction. Rates of IPD were highest among children <2 years of age and among adults ≥65 years of age (188 and 60 cases/100,000, respectively; **Fig. 151-4**). Since the introduction of PCV, IPD rates among infants and children in the United States have fallen by >75%, a decrease driven by the near elimination of vaccine-serotype IPD. A similar impact of PCV on vaccine-serotype IPD rates has been consistently observed in countries where PCV has been introduced into the routine pediatric vaccination schedule. However, the magnitudes of change in the non-vaccine-serotype IPD rate in various countries have been heterogeneous; the interpretation of this heterogeneity is a complex issue. In the United States, Canada, and Australia, rates of non-vaccine-serotype IPD have increased but the magnitude of the increase is generally small relative to the substantial reductions in vaccine-serotype IPD. In contrast, in other settings (e.g., Alaska Native communities and adults in the United Kingdom), the reduction in vaccine-serotype IPD has been offset by notable increases in rates of disease caused by non-vaccine serotypes. Explanations for the heterogeneity of findings include replacement disease resulting from vaccine pressure, changes in clinical case investigation, secular trends unrelated to PCV use, antibiotic pressure selecting for resistant organisms, changes in surveillance or reporting systems, rapidity of PCV introduction, and inclusion of a catch-up campaign. Serotype replacement in IPD follows the use of PCV7, PCV10, and PCV13, but the magnitude of this phenomenon may be small relative to the reduction in disease from vaccine serotypes in vaccinated populations. In adults in the UK, however, where rates of IPD due to vaccine serotypes fell following PCV introduction, the increase in IPD secondary to non- vaccine serotypes is eroding the original impact of PCV. Furthermore, not all vaccine serotypes have declined and persistent disease due to serotypes 3 and 19A in particular has been noted in many settings.

Pneumonia is the most common of the serious pneumococcal disease syndromes and poses special challenges from a clinical and public health perspective. Most cases of pneumococcal pneumonia are not associated with bacteremia, and in these cases a definitive etiologic diagnosis is difficult or impossible. As a result, estimates of disease burden focus primarily on IPD rates and fail to include the major portion of the burden of serious pneumococcal disease. Among children, PCV trials designed to collect efficacy data on syndrome-based outcomes (e.g., radiographically confirmed pneumonia, clinically diagnosed pneumonia) have revealed the burden of culture-negative pneumococcal pneumonia. These trials have provided the means to infer that only ~5–20% of pneumococcal pneumonia cases result in bacteremia. An important randomized controlled trial of PCV among the elderly in the Netherlands (the CAPiTA trial) has revealed the small fraction of adult pneumococcal pneumonia patients who also have bacteremia. Use of high-quality sputum specimens and, in the case of adults with a low likelihood of colonization absent disease, urine antigen detection both contribute to the diagnosis of nonbacteremic pneumococcal pneumonia. Furthermore, accruing evidence continues to indicate that pneumococcal pneumonia events are often the result of co-infection with viral or other bacterial pathogens. Thus a pneumonia case resulting from a pulmonary infection with a single pathogen is probably an uncommon event; rather, most cases of pneumonia likely result from the sequential or contemporaneous co-infection of a host with multiple pathogens, often both viruses and bacteria.

The case–fatality ratios (CFRs) for pneumococcal pneumonia and IPD vary by age, underlying medical condition, and access to care. In addition, the CFR for pneumococcal pneumonia varies with the severity of disease at presentation (rather than according to whether the pneumonia episode is associated with bacteremia) and with the patient's age (from <5% among hospitalized patients 18–44 years old to >12% among those >65 years old, even when appropriate and timely management is available). Notably, the likelihood of death in the first 24 h of hospitalization did not change substantially with the introduction of antibiotics; this surprising observation highlights the fact that the pathophysiology of severe pneumococcal pneumonia among adults reflects a rapidly progressive cascade of events that often unfolds irrespective of antibiotic administration. Management in an intensive care unit can provide critical support for the patient through the acute period, with lower CFRs, while antibiotics address the underlying infection.

Rates of pneumococcal disease vary by season, with higher rates in colder than in warmer months in temperate climates; by sex, with males more often affected than females; and by risk group, with risk factors including underlying medical conditions, behavioral issues (e.g., smoking), and ethnic group. In the United States, some Native American populations (including Alaska Natives) and African Americans have higher rates of disease than the general population; the increased risk is probably attributable to socioeconomic conditions and the prevalence of underlying risk factors for pneumococcal disease. Medical conditions that increase the risk of pneumococcal infection are listed in **Table 151-1**. Outbreaks of disease are well recognized in crowded settings with susceptible individuals, such as infant day-care facilities, military barracks, and nursing homes. Furthermore, there is a clear association between preceding viral respiratory disease (especially but not exclusively influenza) and risk of secondary pneumococcal infections. The significant role of pneumococcal pneumonia in the morbidity and mortality associated with seasonal and pandemic influenza is increasingly well recognized.

Antibiotic Resistance Reduced pneumococcal susceptibility to penicillin was first noted in 1967, but not until the 1990s did reduced antibiotic susceptibility emerge as a significant clinical and public health issue, with an increasing prevalence of pneumococcal isolates resistant to single or multiple classes of antibiotics and a rising absolute magnitude of minimal inhibitory concentrations (MICs). Strains with reduced susceptibility to penicillin G, cefotaxime, ceftriaxone, macrolides, and other antibiotics are now found worldwide and account for a significant proportion of disease-causing strains in many locations, especially among children. Vancomycin resistance has not yet been observed in clinical pneumococcal strains. Lack of antimicrobial susceptibility is clearly related to a subset of serotypes, many of which disproportionately cause disease among children. Resistance phenotypes are based on a diverse array of mutational events and inter- and intraspecies gene-transfer phenomena carried out by several types of mobile genetic elements, with consequent dissemination of successful resistant clones. The vicious cycle of antibiotic exposure, selection of resistant organisms in the nasopharynx, and transmission of these organisms within the community, leading to difficult-to-treat infections and increased antibiotic exposure, has been interrupted to some extent by the introduction and routine use of PCV. The clinical implications of pneumococcal antimicrobial nonsusceptibility are addressed in "Treatment," below.

TABLE 151-1 Clinical Risk Groups for Pneumococcal Infection

CLINICAL RISK GROUP	EXAMPLES
Asplenia or splenic dysfunction	Sickle cell disease and other hemoglobinopathies, celiac disease
Chronic respiratory disease	Chronic obstructive pulmonary disease, bronchiectasis, cystic fibrosis, interstitial lung fibrosis, pneumoconiosis, bronchopulmonary dysplasia, aspiration risk, neuromuscular disease (e.g., cerebral palsy), severe asthma
Chronic heart disease	Ischemic heart disease, congenital heart disease, hypertension with cardiac complications, chronic heart failure
Chronic kidney disease	Nephrotic syndrome, chronic renal failure, renal transplantation
Chronic liver disease	Cirrhosis, biliary atresia, chronic hepatitis
Diabetes mellitus	Diabetes mellitus requiring insulin or oral hypoglycemic drugs
Immunocompromise/ immunosuppression	HIV infection, primary immunodeficiency (including B cell, T cell, complement, and some phagocytic disorders), leukemia, lymphoma, Hodgkin's disease, multiple myeloma, generalized malignancy, chemotherapy, organ or bone marrow transplantation, systemic glucocorticoid treatment for >1 month at a dose equivalent to ≥20 mg/d (children, ≥1 mg/kg per day)
Cochlear implants	. . .
Cerebrospinal fluid leaks	. . .
Miscellaneous	Infancy and old age; prior hospitalization; alcoholism; malnutrition; cigarette smoking; day-care center attendance; residence in military training camps, prisons, homeless shelters

Note: Groups for whom pneumococcal vaccines are recommended by the Advisory Committee on Immunization Practices can be found at *www.cdc.gov/vaccines/schedules/*.

PATHOGENESIS

Pneumococci colonize the human nasopharynx from an early age; colonization acquisition events are generally described as asymptomatic, but evidence exists to associate acquisition with mild respiratory symptoms, especially in the very young. Bacteria survive in the nasopharynx protected by a variety of factors, including their bacterial capsule and the formation of a biofilm. From the nasopharynx, the bacteria spread either via the bloodstream to distant sites (e.g., brain, joint, bones, peritoneal cavity) or locally to mucosal surfaces where they can cause otitis media or pneumonia. Direct spread from the nasopharynx to the central nervous system (CNS) can occur in rare cases of skull-base fracture, although most cases of pneumococcal meningitis are secondary to hematogenous spread. The pneumococcus is not a static bacterium; rather, it modifies its expression of capsule in adaptation to the external environment. In the nasopharynx, the pneumococcus downregulates capsular expression, averting protective immunologic mechanisms that recognize capsule; rough colonies are the phenotype on culture. Upon invasion by traversal of the epithelium, the pneumococcus upregulates its capsular expression, transforming its appearance on culture to smooth colonies—a change illustrating the dynamic nature of the organism in response to the local environment. Pneumococci can cause disease in almost any organ or part of the body; however, otitis media, pneumonia, bacteremia, and meningitis are most common. Colonization is a relatively frequent event, yet disease is rare. In the nasopharynx, pneumococci survive in mucus secreted by epithelial cells and in a biofilm they create, where they can avoid local immune factors such as leukocytes and complement. The mucus itself is a component of local defense mechanisms, and the flow of mucus (driven in part by cilia in what is known as the *mucociliary escalator*) effects mechanical clearance of pneumococci. While many colonization episodes are of short duration, longitudinal studies in adults and children have revealed persistent colonization with a specific serotype over many months. Colonization eventually results in the development of capsule- and protein-specific serum IgG antibodies, which are thought to play a role in mediating clearance of bacteria from the nasopharynx. IgG antibodies to surface-exposed cell-wall or secreted proteins also appear in the circulation in an age-dependent fashion or after colonization; these antibodies are likely to have a disease-modifying and/or protective role. Recent acquisition of a new colonizing serotype is more likely to be associated with subsequent invasion, presumably as a result of the absence of type-specific immunity. Intercurrent viral infections make the host more susceptible to pneumococcal colonization, and pneumococcal disease in a colonized individual often follows perturbation of the nasopharyngeal mucosa by such infections. Local cytokine production after a viral infection is thought to upregulate adhesion factors in the respiratory epithelium, allowing pneumococci to adhere via a variety of surface adhesin molecules, including PsaA, PspA, CbpA, PspC, Hyl, pneumolysin, and the neuraminidases (Fig. 151-2). Adhesion coupled with inflammation induced by pneumococcal factors such as peptidoglycans and teichoic acids results in invasion. It is the inflammation induced by various bacterium-derived factors that is responsible for the pathology associated with pneumococcal infection. Pneumococcal cell wall–derived teichoic acids and peptidoglycans induce a variety of cytokines, including the proinflammatory cytokines interleukin (IL) 1, IL-6, and tumor necrosis factor, and activate complement via the alternative pathway. Polymorphonuclear leukocytes are thus attracted, and an intense inflammatory response is initiated. Pneumolysin also is important in local pathology, inducing proinflammatory cytokine production by local monocytes.

The pneumococcal capsule, consisting of polysaccharides with antiphagocytic properties (i.e., the capacity to resist complement deposition in the absence of type-specific antibody), plays an important role in pathogenesis. While most capsular types can cause human disease, certain capsular types are more commonly isolated from sites of infection. The reason for the dominance of some serotypes over others in IPD is unclear.

HOST DEFENSE MECHANISMS

Innate Immunity As described above, intact respiratory epithelium and a host of nonspecific or innate immune factors (e.g., mucus, splenic function, complement, neutrophils, and macrophages) constitute the first line of defense against pneumococci. Physical factors such as the cough reflex and the mucociliary escalator are important in clearing bacteria from the lungs. Immunologic factors are critical as well: C-reactive protein (CRP) binds phosphorylcholine in the pneumococcal cell wall, inducing complement activation and leading to bacterial clearance; Toll-like receptor 2 (TLR2) recognizes pneumococcal-derived lipoproteins. In animal models, the absence of host TLR2 leads to more severe infection and impaired clearance of nasopharyngeal colonization. TLR4 appears to be necessary for the proinflammatory effect of pneumolysin on macrophages. The importance of TLR recognition is underlined by descriptions of an inherited deficiency of human IL-1 receptor–associated kinase 4 (IRAK-4) that manifests as an unusual susceptibility to infection with bacteria, including *S. pneumoniae*. IRAK-4 is essential for the normal functioning of several TLRs. Other factors that interfere with these nonspecific mechanisms (e.g., viral infections, cystic fibrosis, bronchiectasis, complement deficiency, chronic obstructive pulmonary disease) all predispose to the development of pneumococcal pneumonia. Patients who lack a spleen or have abnormal splenic function (e.g., persons with sickle cell disease) are at high risk of developing overwhelming pneumococcal disease.

Acquired Immunity Acquired immunity induced following colonization or through exposure to cross-reactive antigens rests largely on the development of serum IgG antibody specific for the pneumococcal capsular polysaccharide. Nearly all polysaccharides are T cell–independent antigens; B cells can make antibodies to such antigens without T cell help. However, in children <1–2 years old, such B cell responses are poorly developed. This delayed ontogeny of capsule-specific IgG in young children is associated with susceptibility to pneumococcal infection (Fig. 151-4). The extremely high risk of pneumococcal infection in the absence of serum immunoglobulin (i.e., in conditions such as agammaglobulinemia) highlights the important role of capsular antibody in protection against disease. Each serotype's capsule is chemically distinct, even though for some serotypes the chemical distinction from another type may be a minor one; thus immunity tends to be serotype specific, although some cross-immunity exists. For example, conjugate vaccine–induced antibodies to serotype 6B prevent infection due to serotype 6A but not 6C; cross-protection against serotype 6C requires the administration of vaccines containing 6A. Antibodies to surface-exposed or secreted pneumococcal proteins (such as pneumolysin, PsaA, and PspA) also appear in the circulation with increasing age of the host and are likely to contribute to protection. Data from murine models suggest that CD4+ T cells may play a role in preventing pneumococcal colonization and disease, and experimental data derived from humans suggest that IL-17-secreting CD4+ T cells may be relevant.

APPROACH TO THE PATIENT

Pneumococcal Infections

There is no pathognomonic presentation of pneumococcal disease; patients may present with one or more clinical syndromes (e.g., pneumonia, meningitis, sepsis). *S. pneumoniae* can infect nearly any body tissue, manifesting as disease ranging in severity from mild and self-limited to life-threatening. The differential diagnosis of common clinical syndromes such as pneumonia, otitis media, fever of unknown origin, and meningitis should always include pneumococcal infection. A microbiologically confirmed diagnosis is made in only a minority of pneumococcal cases since, in most circumstances (and especially in pneumonia and otitis media), fluid from the site of infection is not available for etiologic determination, and infection of body fluids distant from the site of infection (e.g., blood in the case of pneumonia) occurs in only a minority of true pneumococcal cases. Empirical therapy that includes appropriate treatment for *S. pneumoniae* is often indicated.

Algorithms for assessment and management of ill children (Integrated Management of Childhood Illness; IMCI) have been developed for use in the developing world or in other settings where evaluation by a trained physician may not be feasible. No such algorithms for the management of adults with suspected disease exist. Children who present with signs associated with increased risk of serious disease, such as an inability to drink, convulsions, lethargy, and severe malnutrition, are categorized as having very severe disease without further evaluation by the community health care worker; are given antibiotics; and are immediately referred to a hospital for diagnosis and management. Children who present with cough and tachypnea (the latter defined according to specific age strata) are further stratified into severity categories based on the presence or absence of lower chest wall indrawing and are managed accordingly either with antibiotics alone or with antibiotics and referral to a hospital facility. Children with cough but no tachypnea are categorized as having a nonpneumonia respiratory illness.

CLINICAL MANIFESTATIONS

The clinical manifestations of pneumococcal disease depend on the site of infection and the duration of illness. Clinical syndromes are classified as noninvasive (e.g., otitis media) or invasive (e.g., bacteremic pneumonia, meningitis) according to whether a normally sterile site is infected. The pathogenesis of noninvasive illness involves contiguous spread from the nasopharynx or skin; invasive disease involves infection of a normally sterile body fluid or follows bacteremia. Regardless of the mechanism, all pneumococcal infections result from nasopharyngeal acquisition of the organism.

Pneumonia Pneumonia is the most common serious pneumococcal syndrome and is considered invasive when associated with a positive blood culture. Whether to categorize nonbacteremic pneumococcal pneumonia as invasive or noninvasive remains debatable.

Pneumococcal pneumonia can present as a mild community-acquired infection at one extreme and as a life-threatening disease requiring intubation and intensive support at the other.

PRESENTING MANIFESTATIONS The presentation of pneumococcal pneumonia does not reliably distinguish it from pneumonia of other etiologies. In a subset of cases, pneumococcal pneumonia is recognized at the outset as associated with a viral upper respiratory infection and is characterized by the abrupt onset of cough and dyspnea accompanied by fever, shaking chills, and myalgias. The cough evolves from nonpurulent to productive of sputum that is purulent and sometimes tinged with blood. Patients may describe stabbing pleuritic chest pain and significant dyspnea indicating involvement of the parietal pleura. Among the elderly, the presenting clinical symptoms may be less specific, with confusion or malaise but without fever or cough. In such cases, a high index of suspicion is required because failure to treat pneumococcal pneumonia promptly in an elderly patient is likely to result in rapid evolution of the infection, with increased severity, morbidity, and risk of death.

FINDINGS ON PHYSICAL EXAMINATION The clinical signs associated with pneumococcal pneumonia among adults include tachypnea (defined as >20 breaths/min) and tachycardia, hypotension in severe cases, and fever in most cases (although not in all elderly patients). Respiratory signs are varied, including dullness to percussion in areas of the chest with significant consolidation, crackles on auscultation, reduced expansion of the chest in some cases as a result of splinting to reduce pain, bronchial breathing in a minority of cases, pleural rub in occasional cases, and cyanosis in cases with significant hypoxemia. Among infants with severe pneumonia, chest wall indrawing and nasal flaring are common. Nonrespiratory findings can include upper abdominal pain if the diaphragmatic pleura is involved as well as mental status changes, particularly confusion in elderly patients.

DIFFERENTIAL DIAGNOSIS The differential diagnosis of pneumococcal pneumonia includes cardiac conditions such as myocardial infarction and heart failure with atypical pulmonary edema; pulmonary conditions such as atelectasis; and pneumonia caused by viral pathogens, mycoplasmas, *Haemophilus influenzae*, *Klebsiella pneumoniae*, *Staphylococcus aureus*, *Legionella*, or (in HIV-infected and otherwise immunocompromised hosts) *Pneumocystis jirovecii*. In cases with abdominal symptoms, the differential diagnosis includes cholecystitis, appendicitis, perforated peptic ulcer disease, and subphrenic abscesses. The challenge in cases with abdominal symptoms is to remember to include pneumococcal pneumonia—a nonabdominal process—in the differential diagnosis.

DIAGNOSIS Some authorities advocate treating uncomplicated, nonsevere, community-acquired pneumonia without determining the microbiologic etiology, given that this information is unlikely to alter clinical management. However, efforts to identify the cause of pneumonia are important when the disease is more severe and when the diagnosis of pneumonia is not clearly established. The gold standard for etiologic diagnosis of pneumococcal pneumonia is pathologic examination of lung tissue. In lieu of that procedure, evidence of an infiltrate on chest radiography warrants a diagnosis of pneumonia. However, cases of pneumonia without radiographic evidence do occur. An infiltrate can be absent either early in the course of the illness or with dehydration; upon rehydration, an infiltrate usually appears. The radiographic appearance of pneumococcal pneumonia is varied; it classically consists of lobar or segmental consolidation **(Fig. 151-5)** but in some cases is patchy. More than one lobe is involved in ~30% of cases. Consolidation may be associated with a small pleural effusion or empyema in complicated cases. In children, "round pneumonia," a distinctly spherical consolidation on chest radiography, is associated with a pneumococcal etiology. Round pneumonia is uncommon in adults. *S. pneumoniae* is not the only cause of such lesions; other causes, especially cancer, should be considered.

Blood drawn from patients with suspected pneumococcal pneumonia can be used for supportive or definitive diagnostic tests. Blood cultures are positive for pneumococci in a minority (<30%) of cases of pneumococcal pneumonia, as evidenced especially by vaccine clinical trials, which provide an independent method to reveal the contribution of the pneumococcus to pneumonia cases. Nonspecific findings include an elevated polymorphonuclear leukocyte count (>15,000/μL in most cases and upward of 40,000/μL in some), leukopenia in <10% of cases (a poor prognostic sign associated with a fatal outcome), and elevated values in liver function tests (e.g., both conjugated and unconjugated hyperbilirubinemia). Anemia, low serum albumin levels, hyponatremia, and elevated serum creatinine levels are all found in ~20–30% of patients.

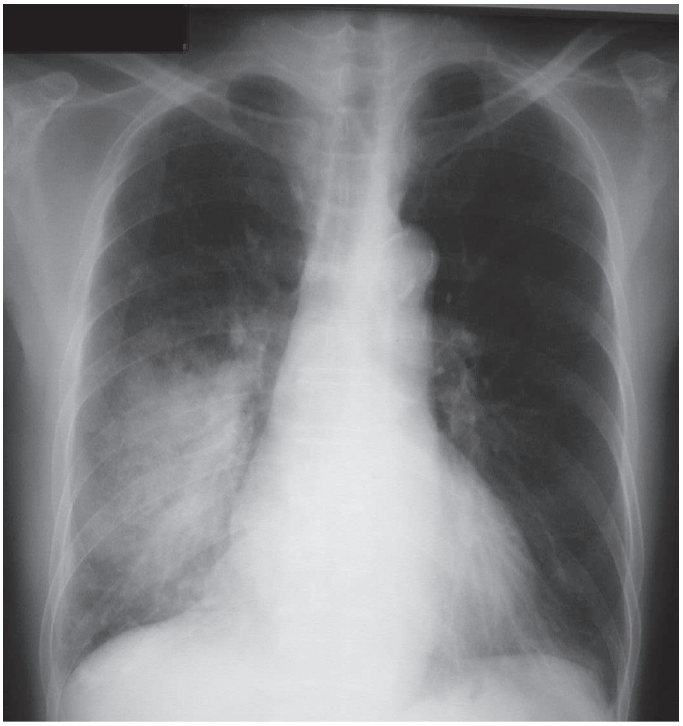

FIGURE 151-5 Chest radiograph depicting classic lobar pneumococcal pneumonia in the right lower lobe of an elderly patient's lung.

Urinary pneumococcal antigen assays, based on identifying a ubiquitous common cell wall polysaccharide, have facilitated etiologic diagnosis, but the application of the results is confounded by the fact that nasopharyngeal colonization with the pneumococcus, in the absence of disease, also results in a positive test. In adults, therefore, a positive pneumococcal urinary antigen test has a predictive value for etiologic attribution of pneumonia because the prevalence of pneumococcal nasopharyngeal colonization is relatively low, although the sensitivity of the assay is modest. In communities, particularly those in low-income countries, where colonization rates among adults are high, urine antigen assays may be less useful. The same issue holds for children, in whom a positive urinary antigen test is usually uninformative for etiologic attribution of their pneumonia illness because colonization rates are generally high. A recent advance is the development of quantitative serotype-specific urinary antigen detection assays for up to 24 pneumococcal antigens; their application for adults and children holds promise, especially in detecting serotypes that are rarely identified in asymptomatic carriage (e.g., serotype 1), even among children.

Most cases of pneumococcal pneumonia in adults are diagnosed by Gram's staining and culture of sputum. The utility of a sputum specimen is directly related to its quality and the patient's antibiotic treatment status.

COMPLICATIONS Empyema is the most common focal complication of pneumococcal pneumonia, occurring in <5% of cases. When fluid in the pleural space is accompanied by fever and leukocytosis (even low-grade) after 4–5 days of appropriate antibiotic treatment for pneumococcal pneumonia, empyema should be considered. Parapneumonic effusions are more common than empyema, representing a self-limited inflammatory response to pneumonia. Pleural fluid with frank pus, bacteria (detected by microscopic examination), or a pH of ≤7.1 indicates empyema and demands aggressive and complete drainage, usually through chest tube insertion.

Meningitis Pneumococcal meningitis usually presents as a pyogenic condition that is clinically indistinguishable from meningitis of other bacterial etiologies. Meningitis can be the primary presenting pneumococcal syndrome or a complication of other conditions such as skull fracture, otitis media, bacteremia, or mastoiditis. Now that *H. influenzae* type b vaccine is routinely used in children, *S. pneumoniae* and *Neisseria meningitidis* are the most common bacterial causes of meningitis in both adults and children. Pyogenic meningitis, including that due to *S. pneumoniae*, is associated clinically with findings that include severe, generalized, gradual-onset headache, fever, and nausea as well as specific CNS manifestations such as stiff neck, photophobia, seizures, and confusion. Clinical signs include a toxic appearance, altered consciousness, bradycardia, and hypertension indicative of increased intracranial pressure. A small proportion of adult patients have Kernig's or Brudzinski's sign or cranial nerve palsies (particularly of the third and sixth cranial nerves).

A definitive diagnosis of pneumococcal meningitis rests on the examination of CSF for (1) evidence of turbidity (visual inspection); (2) elevated protein level, elevated white blood cell count, and reduced glucose concentration (quantitative measurement); and (3) specific identification of the etiologic agent (culture, Gram's staining, antigen testing, or polymerase chain reaction [PCR]). A blood culture positive for *S. pneumoniae* in conjunction with clinical manifestations of meningitis also is considered confirmatory. As discussed in "Pneumonia," above, detection of pneumococcal antigen in urine is considered highly specific among adults because of the low prevalence of nasopharyngeal colonization in this age group.

The mortality rate for pneumococcal meningitis is ~20%. In addition, up to 50% of survivors experience acute or chronic complications, including deafness, hydrocephalus, and mental retardation in children and diffuse brain swelling, subarachnoid bleeding, hydrocephalus, cerebrovascular complications, and hearing loss in adults.

Other Invasive Syndromes *S. pneumoniae* can cause other invasive syndromes involving virtually any body site. These syndromes include primary bacteremia without other sites of infection (bacteremia without a source; occult bacteremia), osteomyelitis, septic arthritis, endocarditis, pericarditis, and peritonitis. The essential diagnostic approach is collection of fluid from the site of infection by sterile technique and examination by Gram's staining, culture, and—when relevant—capsular antigen assay or PCR. Hemolytic-uremic syndrome can complicate invasive pneumococcal disease.

Noninvasive Syndromes The major noninvasive syndromes caused by *S. pneumoniae* are sinusitis, bacterial bronchitis, and otitis media; the latter is the most common pneumococcal syndrome and most often affects young children. The manifestations of otitis media include the acute onset of severe pain, fever, deafness, and tinnitus, most frequently in the setting of a recent upper respiratory tract infection. Clinical signs include a red, swollen, often bulging tympanic membrane with reduced movement on insufflation or tympanography. Redness of the tympanic membrane is not sufficient for the diagnosis of otitis media.

Pneumococcal sinusitis is also a complication of upper respiratory tract infections and presents with facial pain, congestion, fever, and—in many cases—persistent nighttime cough. A definitive diagnosis is made by aspiration and culture of sinus material; however, presumptive treatment is most commonly initiated after application of a strict set of clinical diagnostic criteria. Pneumococcal bronchitis is usually seen in the context of pre-existing lung conditions such as bronchiectasis or chronic obstructive pulmonary disease (COPD) and may be caused by nontypeable strains.

TREATMENT

Pneumococcal Infections

Historically, the activity of penicillin against pneumococci made parenteral penicillin G the drug of choice for disease caused by susceptible organisms, including community-acquired pneumonia. Today, parenteral β-lactam drugs such as ampicillin, cefotaxime, ceftriaxone, and cefuroxime are often used as first-line agents for community-acquired infections. Macrolides and cephalosporins are alternatives for penicillin-allergic patients. While agents such as clindamycin, tetracycline, and trimethoprim-sulfamethoxazole exhibit some activity against pneumococci, resistance to these agents is frequently encountered in different parts of the world.

Penicillin-resistant pneumococci were first described in the mid-1960s, at which point tetracycline- and macrolide-resistant strains had already been reported. Multidrug-resistant strains were first described in the 1970s, but it was during the 1990s that pneumococcal drug resistance reached pandemic proportions. The use of antibiotics selects for resistant pneumococci, and strains resistant to β-lactam agents and to multiple drugs are now found all over the world. The emergence of high rates of macrolide and fluoroquinolone resistance also has been described. Drug-resistant pneumococci are considered a serious threat by the Centers for Disease Control and Prevention.

The molecular basis of penicillin resistance in *S. pneumoniae* is the alteration of penicillin-binding protein (PBP) genes by transformation and horizontal transfer of DNA from related streptococcal species. Such alteration of PBPs results in lower affinity for penicillins. Depending on the specific PBP(s) and the number of PBPs altered, the level of resistance ranges from intermediate to high. For many years, penicillin susceptibility breakpoints have been defined by MICs as follows: susceptible, ≤0.06 μg/mL and resistant, ≥2.0 μg/mL. However, in vitro results often were not predictive of the response of a patient to treatment for pneumococcal diseases other than meningitis. Revised recommendations have been based on the penicillin G breakpoints established in 2008 by the Clinical and Laboratory Standards Institute. For IV treatment of meningitis with at least 24 million units per day in 8 divided doses, the susceptibility breakpoint remains ≤0.06 μg/mL, and MICs of ≥0.12 μg/mL indicate resistance. For IV treatment of nonmeningeal infections with 12 million units per day in 6 divided doses, the breakpoints are ≤2 μg/mL for susceptible organisms and ≥8 μg/mL for resistant organisms; a dosage of 18–24 million units per day is recommended for strains with MICs in the intermediate category.

Although guidelines for antibiotic therapy should be driven in part by local patterns of resistance, guidelines from national organizations in many countries (e.g., the Infectious Diseases Society of America/American Thoracic Society, the British Thoracic Society, the European Respiratory Society) lay out evidence-based approaches. The following guidelines for the treatment of individual sepsis syndromes are based on those advocated by the American Academy of Pediatrics Red Book and updated in 2023.

MENINGITIS LIKELY OR PROVEN TO BE DUE TO *S. PNEUMONIAE*

In areas of the world with an increased prevalence of resistant pneumococci, first-line therapy for persons ≥1 month of age is a combination of vancomycin (adults, 30–60 mg/kg per day; infants and children, 60 mg/kg per day) and cefotaxime (adults, 8–12 g/d in 4–6 divided doses; children, 225–300 mg/kg per day in 1 dose or 2 divided doses) or ceftriaxone (adults, 4 g/d in 1 dose or 2 divided doses; children, 100 mg/kg per day in 1 dose or 2 divided doses). In low-prevalence areas and where the patient has not recently traveled, vancomycin is not included in first-line therapy. If children are hypersensitive to β-lactam agents (penicillins and cephalosporins), rifampin (adults, 600 mg/d; children, 20 mg/d in 1 dose or 2 divided doses) can be substituted for cefotaxime or ceftriaxone and added as a second agent. A repeat lumbar puncture should be considered after 48 h if the organism is not susceptible to penicillin and information on cephalosporin sensitivity is not yet available, if the patient's clinical condition does not improve or deteriorates, or if dexamethasone has been administered interfering with the ability to interpret clinical responses in the deteriorating patient. When antibiotic sensitivity data become available, treatment should be modified accordingly. If the isolate is sensitive to penicillin, vancomycin can be discontinued and penicillin can replace the cephalosporin, or cefotaxime or ceftriaxone can be continued alone. If the isolate displays any resistance to penicillin but is susceptible to the cephalosporins, vancomycin can be discontinued and cefotaxime or ceftriaxone continued. If the isolate exhibits any resistance to penicillin and is not susceptible to cefotaxime and ceftriaxone, vancomycin and high-dose cefotaxime or ceftriaxone can be continued and rifampin may be added. Data support the use of corticosteroids in high-income countries but do not appear to have a beneficial effect in low-income countries. This discrepancy in the efficacy of corticosteroids may be related to differences in availability of appropriate and timely medical care. Glucocorticoids significantly reduce rates of mortality, severe hearing loss, and neurologic sequelae in adults and should be administered to those with community-acquired bacterial meningitis. If dexamethasone is given to either adults or children, it should be administered before or in conjunction with the first antibiotic dose.

SEPSIS (EXCLUDING MENINGITIS)

In previously well children with noncritical illness, therapy with a recommended antibiotic should be instigated at the usually recommended dosages: ampicillin 200 mg/kg/day (doses 6 h apart), cefotaxime, 75–225 mg/kg/day (doses 8 h apart), ceftriaxone, 50–75 mg/kg/day (doses 12–24 h apart) or penicillin G, 250,000–400,000 units/kg per day (in divided doses 4–6 h apart). For critically ill children, including those who have myocarditis or multilobular pneumonia with hypoxia or hypotension, vancomycin may be added if the isolate may possibly be resistant to β-lactam drugs, with its use reviewed once susceptibility data become available. If the organism is resistant to β-lactam agents, therapy should be modified on the basis of clinical response and susceptibility to other antibiotics. Clindamycin or vancomycin can be used as a first-line agent for children with severe β-lactam hypersensitivity, but vancomycin should not be continued if the organism is shown to be sensitive to other non-β-lactam antibiotics.

For outpatient management, oral amoxicillin (45–90 mg/kg/day, doses 8 h apart) provides effective treatment for virtually all cases of pneumococcal pneumonia. Cephalosporins, which are far more expensive, offer no advantages over amoxicillin. Levofloxacin (500–750 mg/d as a single dose) and moxifloxacin (400 mg/d as a single dose) also are highly likely to be effective in the United States except in patients who come from closed populations where these drugs are used widely or who have themselves been treated recently with a quinolone. Clindamycin (600–1200 mg/d every 6 h) is effective in 90% of cases and azithromycin (500 mg on day 1 followed by 250–500 mg/d) or clarithromycin (500–750 mg/d as a single dose) in 80% of cases. Treatment failure resulting in bacteremic disease due to macrolide-resistant isolates has been amply documented in patients given azithromycin empirically. As noted above, rates of resistance to all these antibiotics are relatively low in some countries and much higher in others; high-dose amoxicillin remains the best option worldwide.

The optimal duration of treatment for pneumococcal pneumonia is uncertain, but its continuation for at least 5 days once the patient becomes afebrile appears to be a prudent approach—although in adults, 5 days in total will usually suffice. Cases with a second focus of infection (e.g., empyema or septic arthritis) require longer therapy.

ACUTE OTITIS MEDIA

Amoxicillin (80–90 mg/kg per day) is recommended for infants <6 months of age and those 6–23 months of age with bilateral disease. Observation and symptom-based treatment without antibiotics are advocated for nonsevere illness and an uncertain diagnosis in children 6 months to 2 years of age and nonsevere illness (even if the diagnosis seems certain) in children >2 years of age. Although the optimal duration of therapy has not been conclusively established, a 10-day course is recommended for younger children and for children with severe disease at any age. For children >6 years old who have mild or moderate disease, a course of 5–7 days is considered adequate. Patients whose illness fails to respond should be reassessed at 48–72 h. If acute otitis media is confirmed and antibiotic treatment has not been started, administration of amoxicillin should be commenced. If antibiotic therapy fails, a change is indicated. Failure to respond to second-line antibiotics (such as high-dose amoxicillin-clavulanate) as well indicates that myringotomy or tympanocentesis may need to be undertaken in order to obtain samples for culture.

The above recommendations can also be followed for the treatment of sinusitis. Detailed information on the further management of these conditions in children has been published by the American Academy of Pediatrics, the American Academy of Family Physicians, the Pediatric Infectious Diseases Society, and the Infectious Diseases Society of America.

PREVENTION

Measures to prevent pneumococcal disease include vaccination against *S. pneumoniae* and influenza viruses, reduction of comorbidities that increase the risk of pneumococcal disease, and prevention of antibiotic overuse, which fuels pneumococcal resistance.

Capsular Polysaccharide Vaccines The 23-valent pneumococcal polysaccharide vaccine (PPSV23), containing 25 μg of each capsular polysaccharide, has been licensed for use since 1983. Recommendations for its use vary by country with an age-based recommendation for those ≥65 years of age most commonly found. However, the utility of PPSV23 is likely to be limited in the future as more countries introduce the newly licensed (2021) extended-valency pneumococcal conjugate vaccines (see "Polysaccharide–Protein Conjugate Vaccines," below).

The effectiveness of PPSV23 against IPD, pneumococcal pneumonia, all-cause pneumonia, and death is controversial, with wide variation in observations. The many published meta-analyses of PPSV efficacy have often reached opposing conclusions with regard to a given clinical entity. Generally, observational studies cite greater effectiveness than do controlled clinical trials. The consensus is that PPSV is effective against IPD but is less effective against nonbacteremic pneumococcal pneumonia. However, the results of some published trials, observational studies, and meta-analyses contradict this view. Effectiveness is often lower in the elderly and in immunodeficient patients whose condition is associated with reduced antibody responses to vaccines than in younger, healthier populations. When PPSV is effective, the duration of protection following a single dose of vaccine is estimated to be ~5 years.

Even in the setting of routine pneumococcal conjugate vaccination of infants (which indirectly protects adults from vaccine-serotype strains), disease caused by serotypes not represented in the conjugate vaccine continues to be a significant burden among adults.

Polysaccharide–Protein Conjugate Vaccines Infants and young children respond poorly to PPSV, which contains T cell–independent antigens. Consequently, another class of pneumococcal vaccines, the PCVs, were developed specifically for infants and young children. The first product, a 7-valent PCV, was licensed in 2000 in the United States and two PCV products—containing 10 and 13 serotypes, respectively—were licensed in 2009 and 2010, respectively. The serotypes included in these original PCV formulations were important causes of IPD and antibiotic resistance among young children. Randomized controlled trials and real-world evidence accumulated following their widespread use have demonstrated a high degree of efficacy of PCVs against vaccine-serotype IPD as well as efficacy against pneumonia, otitis media, nasopharyngeal colonization, and all-cause mortality. PCVs are recommended by the World Health Organization for inclusion in routine childhood immunization schedules worldwide, especially in countries with high infant mortality rates. To date 146 countries (75%) have PCV in their National Immunization program, 15 are planning introduction, and 33 have no national decision.

In 2021, two new PCVs were licensed for adults: a 15-valent pneumococcal conjugate vaccine, VAXNEUVANCE (Merck), and a

20-valent pneumococcal conjugate vaccine, PREVNAR 20 (Pfizer). Both vaccines are now also licensed for infant immunization and are likely to gradually replace PCV10 and PCV13. The additional serotypes in the new vaccines account for a variable fraction of the residual IPD in children (additional 15–45%) and adults (additional 11–33%) for PCV15 and PCV20, respectively, in countries where PCV13 coverage is high. For adults the United States Advisory Committee on Immunization Practices (ACIP) now recommends PCV20 instead of PPSV23 for all persons ≥65 years of age and for those 2–64 years of age who have underlying medical conditions that put them at increased risk for pneumococcal disease or, if infected, disease of increased severity (Table 151-1; see also *www.cdc.gov/vaccines/schedules*). United States recommendations for PPSV23 are now limited to sequential use in the above age groups receiving the PCV15 vaccine.

The introduction of PCV in high-income settings has resulted in a >90% reduction in vaccine-serotype IPD among the whole population. This decline has been noted not only in those age groups immunized but also in adults and is attributable to the near elimination of vaccine-serotype nasopharyngeal colonization in immunized infants, which reduces spread to adults. This protection of unimmunized community members through vaccination of a subset of the community is termed *the indirect effect*. Increases in colonization with—and concomitantly in disease due to—non-vaccine-serotype strains (i.e., replacement colonization and disease) have been seen. The scale of replacement disease has varied geographically with the impact eroding vaccine impact significantly in the elderly in the United Kingdom while having relatively little impact in the United States (see “Epidemiology,” above). Since vaccine-serotype strains are more commonly resistant to antibiotics than are non-vaccine serotypes, use of PCV has also resulted in substantial declines in the proportion and absolute rates of drug-resistant pneumococcal disease. The ACIP recommendations for the use of conjugate vaccines can be found at *www.cdc.gov/vaccines/hcp/acip-recs/vacc-specific/pneumo.html*. PCV has been shown to prevent pneumococcal infection in HIV-infected adults.

Other Prevention Strategies Pneumococcal disease can be averted through the prevention of illnesses that predispose individuals to pneumococcal infections. Relevant measures include smoking cessation and influenza vaccination, as well as improved management and control of diabetes, HIV infection, heart disease, and lung disease. Finally, the reduction of antibiotic misuse is a strategy for the prevention of pneumococcal disease in that antimicrobial resistance directly and indirectly perpetuates organism transmission and disease in the community.

■ GLOBAL HEALTH

In 2015, pneumococcal infections were estimated to have caused ~317,000 annual deaths worldwide among children 1–59 months of age, accounting for 9.7% of the 3.2 million all-cause deaths and 38% of all pneumonia deaths in this age group. Updated estimates of residual pneumococcal disease burden since the widespread introduction of PCV have not been published. Reliable estimates of adult cases and deaths globally are more difficult to establish because of limited data from parts of the world where most disease occurs. Rates of pneumococcal disease and mortality vary substantially across geographic settings, with the highest rates in selected countries of sub-Saharan Africa and southern Asia, where risk factors for pneumococcal disease—including HIV infection, lack of breast feeding of infants and children, malnutrition, sickle cell disease, and limited access to medical care—are prevalent. Serotypes causing disease exhibit some heterogeneity across geographic settings, but a small number of serotypes universally account for the preponderance of disease in the absence of vaccination; accordingly, vaccine development and vaccination programs are globally relevant. Reductions in disease from pneumococcal infections are anchored in prevention through the inclusion of pneumococcal vaccines in infant immunization programs, timely assessment and appropriate treatment of persons with pneumococcal infections, and reduction of risk factors for pneumococcal disease. The use of vaccines for the prevention of adult pneumococcal disease, particularly among the elderly, is currently implemented in high-income countries, with virtually no use in low-income countries where most cases of disease exist.

■ FURTHER READING

Krone CL et al: Immunosenescence and pneumococcal disease: An imbalance in host–pathogen interactions. Lancet Respir Med 2:141, 2014.

Lees JA et al: Fast and flexible bacterial genomic epidemiology with PopPUNK. Genome Res 29:304, 2019.

Mackenzie GA et al: The impact of the introduction of pneumococcal conjugate vaccination in invasive pneumococcal disease and pneumonia in The Gambia: 10 years of population-based surveillance. Lancet Infect Dis 21:1293, 2021.

Subramanian K et al: Pneumolysin binds to the mannose receptor C type 1 (MRC-1) leading to anti-inflammatory responses and enhanced pneumococcal survival. Nat Microbiol 4:62, 2019.

Van Der Poll T, Opal SM: Pathogenesis, treatment, and prevention of pneumococcal pneumonia. Lancet 374:1543, 2009.

■ WEBSITES

American Academy of Pediatrics: Red Book: The report of the Committee on Infectious Diseases. Available at: *aapredbook.aappublications.org*.

Cochrane: Corticosteroids for Bacterial Meningitis. Available at: *www.cochrane.org/CD004405/ARI_corticosteroids-bacterial-meningitis*.

U.S. Department of Health and Human Services: Antibiotic Resistance Threats in the United States 2019. Available at: *www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf*.

World Health Organization: Summary of WHO Position Paper on Pneumococcal conjugate vaccines in infants and children under 5 years of age, February 2019. Available at: *www.who.int/publications/i/item/10665-310968*.

152

Staphylococcal Infections

Franklin D. Lowy, Anne-Catrin Uhlemann

Staphylococcus aureus, the most virulent of the many (≥40) staphylococcal species, has demonstrated its versatility by remaining a major cause of morbidity and mortality worldwide despite the availability of numerous effective antistaphylococcal antibiotics. *S. aureus* is a pluripotent pathogen, causing disease through both toxin- and non-toxin-mediated mechanisms. It is responsible for numerous nosocomial and community-based infections that range from relatively minor skin and soft tissue infections (SSTIs) to life-threatening systemic infections.

The “other” staphylococci, coagulase-negative staphylococci, are less virulent than *S. aureus* but remain important pathogens in select settings, such as infections involving prosthetic devices.

MICROBIOLOGY AND TAXONOMY

Staphylococci, gram-positive cocci in the family Micrococcaceae, form grapelike clusters on Gram’s stain **(Fig. 152-1)**. These organisms (~1 μm in diameter) are catalase-positive (unlike streptococcal species), nonmotile, aerobic, and facultatively anaerobic. They are capable of prolonged survival on environmental surfaces under varying conditions. Some species have a relatively broad host range, including mammals and birds, whereas the host range for others is quite narrow—i.e., limited to one or two closely related animals.

S. aureus is generally distinguished from other staphylococcal species by coagulase production, a surface enzyme that converts fibrinogen to fibrin. However, several of the “coagulase-negative staphylococci,”

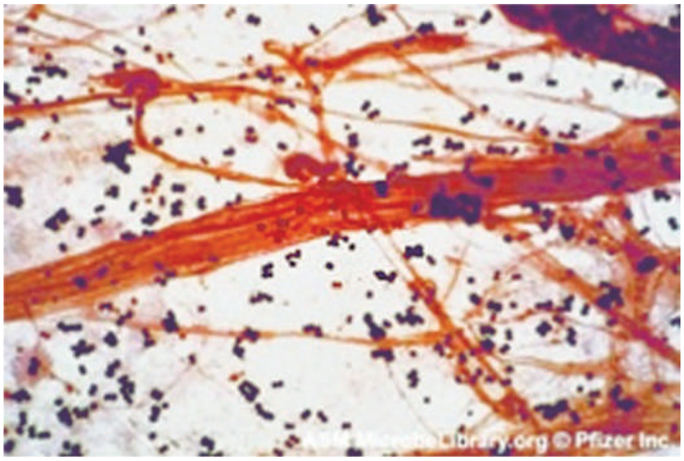

FIGURE 152-1 Gram's stain of *S. aureus* in a sputum sample, illustrating staphylococcal clusters. *(From ASM MicrobeLibrary.org. © Pfizer, Inc.)*

including *S. pseudintermedius* and *S. argenteus*, are coagulase-positive. As a result, description of these other staphylococci as non–*S. aureus* staphylococci (NSaS) is more accurate.

S. aureus ferments mannitol, is positive for protein A, and produces DNAse. On blood agar plates, *S. aureus* forms golden β-hemolytic colonies; in contrast, most NSaS form small, white nonhemolytic colonies. Latex kits that detect both protein A and clumping factor can distinguish *S. aureus* from most other staphylococcal species. Point-of-care tests targeting these two proteins also are used for the rapid detection of staphylococcal colonization. Newer methods such as matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) are increasingly being used for staphylococcal speciation.

Determining whether multiple staphylococcal isolates from different patients are the same or different is often relevant when there is concern that a nosocomial outbreak of staphylococcal infections is due to a common point source (e.g., a contaminated medical instrument). Molecular typing methods, such as pulsed-field gel electrophoresis and sequence-based techniques (e.g., staphylococcal protein A [SpA] typing), have been used for this purpose. More recently, whole-genome sequencing has emerged as the gold standard for discrimination among different isolates.

S. AUREUS INFECTIONS

EPIDEMIOLOGY

S. aureus is both a commensal and an opportunistic pathogen. Approximately 20–40% of healthy persons are colonized with *S. aureus*, with a smaller percentage (~10%) persistently colonized with the same strain. The rate of colonization is elevated among type 1 diabetics, HIV-infected patients, patients undergoing hemodialysis, injection drug users, and individuals with damaged skin. The anterior nares and oropharynx are frequent sites of human colonization, although the skin (especially when damaged), axilla, vagina, and perineum also are often colonized. These colonization sites serve as potential reservoirs for future infections.

Most individuals who develop *S. aureus* infections become infected with a strain that is already a part of their own commensal flora. Breaches of the skin or mucosal membrane allow *S. aureus* to initiate infection. Person-to-person transmission of *S. aureus* also occurs, most frequently from direct personal contact with an infected body site. Spread of staphylococci in aerosols of respiratory or nasal secretions from heavily colonized individuals, although rare, has been reported.

Some diseases increase the risk of *S. aureus* infection. Diabetes, for example, combines an increased rate of *S. aureus* colonization and the use of injectable insulin with the possibility of impaired leukocyte function. Individuals with congenital or acquired qualitative or quantitative defects of polymorphonuclear leukocytes (PMNs) are at increased risk of *S. aureus* infections; this group includes neutropenic patients (e.g., those receiving chemotherapeutic agents), those with chronic granulomatous disease, and those with autosomal dominant hyperimmunoglobulin E (Job syndrome) or Chédiak-Higashi syndrome. Other groups at risk include individuals with end-stage renal disease, HIV infection, skin abnormalities, or prosthetic devices.

S. aureus is a leading cause of health care–associated infections **(Chap. 147)**. It is the most common cause of surgical wound infections and is second only to NSaS as a cause of primary bacteremia. These isolates are often resistant to multiple antibiotics; thus, available therapeutic options may be limited. In the community, *S. aureus* remains an important cause of SSTIs, respiratory infections, and, especially among injection drug users, infective endocarditis. The increasing use of home infusion therapy also poses a risk of community-acquired staphylococcal infections.

In the past three decades, there has been a dramatic change in the epidemiology of infections due to methicillin-resistant *S. aureus* (MRSA). In addition to its major role as a nosocomial pathogen, MRSA has become an established community-based pathogen. Numerous outbreaks of community-associated MRSA (CA-MRSA) infections have been reported in both rural and urban settings in widely separated regions throughout the world.

This trend appears to be due in part to the dramatic increase in MRSA colonization found in the community in different parts of the world. Outbreaks of CA-MRSA infections have occurred among such diverse groups as children, prisoners, athletes, Native Americans, and drug users. Risk factors common to these outbreaks include poor hygienic conditions, close contact, contaminated material, and damaged skin. In some geographic regions of the world, the infections have been caused by a single CA-MRSA strain, while in others, a variety of CA-MRSA strains have been responsible. In the United States, strain sequence type 8 (PFGE type USA300) has been the predominant clone **(Fig. 152-2)**. Although most infections caused by these strains have involved the skin and soft tissue, 5–10% have been invasive and potentially life-threatening. CA-MRSA strains have also been responsible for an increasing number of nosocomial infections. Of concern has been the enhanced capacity of CA-MRSA to cause disease in immunocompetent individuals.

PATHOGENESIS

General Concepts *S. aureus* is a pyogenic pathogen known for its capacity to induce abscess formation at both local and distant sites (i.e., metastatic infections). This classic pathologic response to *S. aureus* defines the framework within which infections will progress. The bacteria elicit an inflammatory response characterized by an initial intense infiltration of PMNs and a subsequent infiltration of macrophages and fibroblasts. Either the host cellular response (including the deposition of fibrin and collagen) contains the infection with the formation of a fibrinous capsule, or infection spreads to the adjoining tissue or into the bloodstream.

In toxin-mediated staphylococcal disease, infection is not invariably present. For example, in staphylococcal food poisoning, once the heat-stable enterotoxin has been released into food, symptoms can develop in the absence of viable bacteria. In staphylococcal toxic shock syndrome (TSS), conditions allowing toxin elaboration at colonization sites (e.g., the presence of a superabsorbent tampon) suffice for initiation of clinical illness.

The *S. aureus* Genome The complete genomes of *S. aureus* strains are now readily available. Among the interesting revelations are (1) the high degree of nucleotide sequence similarity of the core genomes of different strains; (2) the acquisition of a relatively large amount of genetic information by horizontal transfer from other bacterial species; and (3) the presence of unique "pathogenicity" or "genomic" islands—mobile genetic elements that contain clusters of enterotoxin and exotoxin genes and/or antimicrobial resistance determinants. Among the genes in these islands is *mecA*, the gene responsible for methicillin resistance. Methicillin resistance–containing islands have been designated staphylococcal cassette chromosome *mec* (SCC*mec*). There are different SCC*mec* types that range in size from ~20 to 60 kb. Among the more common SCC*mec* types, types 1–3 are

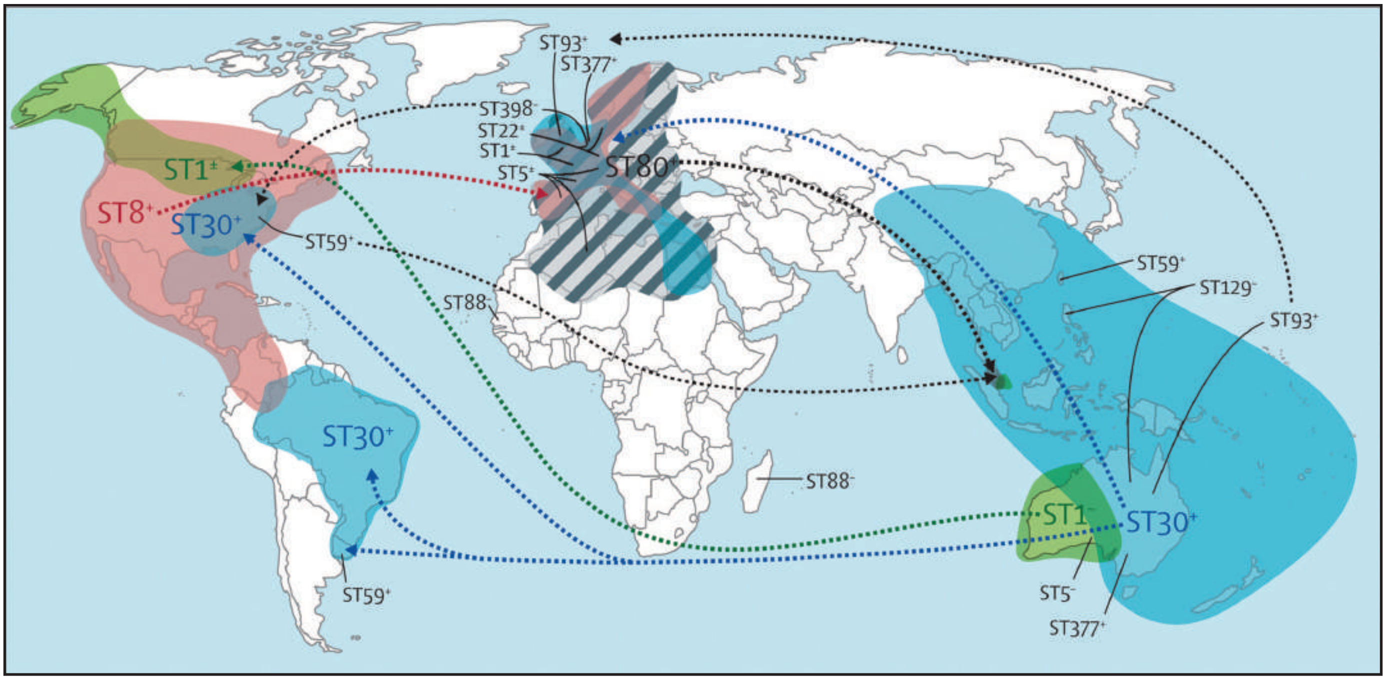

FIGURE 152-2 Global distribution of community-associated MRSA. Dotted lines indicate possible route of dissemination. Estimates of the areas are shown where infection with the main strains—i.e., ST1 *(green)*, ST8 *(red)*, ST30 *(blue)*, and ST80 *(gray hatched)*—have been reported. +, Panton-Valentine leukocidin (PVL)-positive strains; –, PVL-negative strains; ±, PVL-positive and -negative strains. *(Reproduced with permission from FR DeLeo, M Otto, BN Kreiswirth, HF Chambers: Community-associated methicillin-resistant Staphylococcus aureus. Lancet 375:1557, 2010.)*

traditionally associated with nosocomial MRSA isolates, whereas types 4–6 have been associated with epidemic CA-MRSA strains.

A relatively limited number of MRSA clones have been responsible for most community- and hospital-associated infections worldwide. A comparison of these strains with those from earlier outbreaks (e.g., the phage 80/81 strains from the 1950s) has revealed preservation of much of the nucleotide sequence over time. This observation suggests that these strains possess determinants that facilitate survival and spread.

Regulation of Virulence Gene Expression In both toxin-mediated and non-toxin-mediated diseases due to *S. aureus*, the expression of virulence determinants associated with infection depends on a series of regulatory genes (e.g., accessory gene regulator [*agr*] and staphylococcal accessory regulator [*sar*]) that coordinately control the expression of many virulence genes. The regulatory gene *agr* is part of a quorum-sensing signal transduction pathway that senses and responds to bacterial density. Staphylococcal surface proteins are synthesized during the bacterial exponential growth phase in vitro. In contrast, many secreted proteins, such as α toxin, the enterotoxins, and assorted enzymes, are released during the post–exponential growth phase in response to transcription of the effector molecule of *agr*, RNAIII.

These regulatory genes appear to serve a similar function in vivo. Successful invasion requires the sequential expression of these different bacterial elements. Bacterial adhesins are needed to initiate colonization of host tissue surfaces. The subsequent release of various enzymes enables the colony to obtain nutritional support and permits bacteria to spread to adjacent tissues. Studies with strains in which these regulatory genes are inactivated show reduced virulence in several animal models of *S. aureus* infection.

Pathogenesis of Invasive *S. aureus* Infection Staphylococci are opportunists. For these organisms to invade the host and cause infection, some or all of the following steps are necessary: contamination and colonization of host tissue surfaces, breach of cutaneous or mucosal barriers, establishment of a localized infection, invasion, evasion of the host response, and metastatic spread. Colonizing strains or strains transferred from other exposures are introduced into damaged skin, a wound, or the bloodstream. Recurrences of *S. aureus* infections are common, apparently because of the capacity of these pathogens to persist in a quiescent state in various tissues, and then to cause recrudescent infections when suitable conditions arise.

***S. AUREUS* COLONIZATION OF BODY SURFACES** The anterior nares and oropharynx are primary sites of staphylococcal colonization. In the nares, colonization appears to involve the attachment of *S. aureus* to keratinized epithelial cells. Other factors that contribute to colonization include the influence of other resident nasal flora and their bacterial density, host factors, and nasal mucosal damage (e.g., that resulting from inhalational drug use). Other colonized body sites, such as damaged skin, the groin, and the oropharynx, may be particularly important reservoirs for CA-MRSA strains.

INOCULATION AND COLONIZATION OF TISSUE SURFACES Staphylococci may be introduced into tissue as a result of minor abrasions (e.g., mosquito bites), administration of medications such as insulin, or establishment of IV access with catheters. After introduction into a tissue site, bacteria replicate and colonize the host tissue surface. A family of structurally related *S. aureus* surface proteins referred to as MSCRAMMs (microbial surface components recognizing adhesive matrix molecules) play an important role as mediators of adherence to these different sites. By adhering to exposed matrix molecules (e.g., fibrinogen, collagen, fibronectin), MSCRAMMs, such as clumping factor and collagen-binding protein, enable the bacteria to colonize different host tissue surfaces; these proteins contribute to the pathogenesis of invasive infections such as endocarditis and septic arthritis by facilitating the adherence of *S. aureus* to surfaces with exposed fibrinogen or collagen.

Although NSaS are classically known for their ability to elaborate biofilms and to colonize prosthetic devices, *S. aureus* also possesses the genes responsible for biofilm formation, such as the intercellular adhesion (*ica*) locus. Binding to these devices occurs in a stepwise fashion, involving staphylococcal adherence to serum constituents that have coated the device surface and subsequent biofilm elaboration. *S. aureus* is thus a frequent cause of biomedical device–related infections.

INVASION After colonization, staphylococci replicate at the initial site of infection, elaborating enzymes that include serine proteases, hyaluronidases, thermonucleases, and lipases. These enzymes facilitate bacterial survival and local spread across tissue surfaces. The lipases may facilitate survival in lipid-rich areas such as the hair follicles, where *S. aureus* infections are often initiated.

Constitutional findings may result from either localized or systemic infections. The staphylococcal cell wall—consisting of alternating *N*-acetyl muramic acid and *N*-acetyl glucosamine units in combination with an additional cell wall component, lipoteichoic acid—can initiate an inflammatory response that includes the sepsis syndrome. Staphylococcal alpha (α) toxin is a critical staphylococcal toxin. It causes pore formation in various eukaryotic cells and can also initiate an inflammatory response with findings suggestive of sepsis. The *S. aureus* toxin Panton-Valentine leukocidin is cytolytic to PMNs, macrophages, and monocytes. Strains elaborating this toxin have been epidemiologically linked with cutaneous and more serious infections (i.e., pneumonia) caused by strains of CA-MRSA.

EVASION OF HOST DEFENSE MECHANISMS Staphylococci have many host immune evasion strategies that are crucial to their survival. They possess an antiphagocytic polysaccharide microcapsule. Most human *S. aureus* infections are due to strains with capsular types 5 and 8. The zwitterionic (both negatively and positively charged) *S. aureus* capsule also plays a critical role in the induction of abscess formation. Protein A, an MSCRAMM unique to *S. aureus*, acts as an Fc receptor, binding the Fc portion of IgG subclasses 1, 2, and 4 and preventing opsonophagocytosis by PMNs. Both chemotaxis inhibitory protein of staphylococci (CHIPS, a secreted protein) and extracellular adherence protein (EAP, a surface protein) interfere with PMN migration to sites of infection. There are several cytolytic toxins, including α toxin and Panton-Valentine toxin, that are secreted by staphylococci that cause lysis of different host cells and contribute to host tissue damage.

An additional potential mechanism of *S. aureus* evasion is its capacity for intracellular survival. Both professional and nonprofessional phagocytes internalize staphylococci. Internalization by these cells may provide a sanctuary that protects bacteria against the host's defenses. This phenomenon appears to be especially relevant for hepatic Kupffer cells during staphylococcal bacteremias. The intracellular environment favors the phenotypic expression of *S. aureus* small-colony variants, which are found in patients receiving antimicrobial therapy (e.g., with aminoglycosides) and in those with cystic fibrosis or osteomyelitis. These variants, whether intra- or extracellular, may facilitate prolonged staphylococcal survival in different tissue sites and enhance the likelihood of recurrences. Finally, *S. aureus* can survive within PMNs and may use these cells to spread and seed other tissue sites.

PATHOGENESIS OF COMMUNITY-ACQUIRED MRSA INFECTIONS A number of specific virulence determinants contribute to the pathogenesis of CA-MRSA infections. A strong epidemiologic association links the presence of the gene for the Panton-Valentine leukocidin with SSTIs and with necrotizing postinfluenza pneumonia. Other determinants that play a role in the pathogenesis of these infections and contribute to the unique virulence of these clones include the arginine catabolic mobile element (ACME), a cluster of unique genes that may facilitate evasion of host defense mechanisms; phenol-soluble modulins, a family of cytolytic peptides; and α toxin.

Host Response to *S. aureus* Infection The primary host response to *S. aureus* infection is the recruitment of PMNs. These cells are attracted to infection sites by bacterial components such as formylated peptides or peptidoglycan as well as by the cytokines tumor necrosis factor (TNF) and interleukins (ILs) 1 and 6, which are released by activated macrophages and endothelial cells.

Although most individuals have antibodies to staphylococci, it is not clear that antibody levels are qualitatively or quantitatively sufficient to protect against infection. Anticapsular and anti-MSCRAMM antibodies facilitate opsonization in vitro and have been protective against infection in several animal models; however, vaccines with these components have not yet successfully prevented staphylococcal infections in clinical trials.

Pathogenesis of Toxin-Mediated Disease *S. aureus* produces three types of toxins: cytotoxins, pyrogenic toxin superantigens, and exfoliative toxins. Both epidemiologic data and studies in animals suggest that antitoxin antibodies are protective against illness in TSS, staphylococcal food poisoning, and staphylococcal scalded-skin syndrome (SSSS). Illness develops after toxin synthesis and absorption and the subsequent toxin-initiated host response.

ENTEROTOXIN AND TOXIC SHOCK SYNDROME TOXIN 1 (TSST-1) The pyrogenic toxin superantigens are a family of small-molecular-size, structurally similar proteins that are responsible for two diseases: TSS and food poisoning. TSS results from the ability of TSST-1 and enterotoxins to function as T-cell mitogens. In the normal process of antigen presentation, the antigen is first processed within the cell, and peptides are then presented in the major histocompatibility complex (MHC) class II groove, initiating a measured T-cell response. In contrast, TSST-1 and enterotoxins bind directly to the invariant region of MHC—outside the MHC class II groove. TSST-1 and the enterotoxins can then bind T-cell receptors via the vβ chain; this binding results in a dramatic overexpansion of T-cell clones (up to 20% of the total T-cell population). The consequence of this T-cell expansion is a cytokine storm, with the release of inflammatory mediators that include interferon γ, IL-1, IL-6, TNF-α, and TNF-β. The resulting multisystem disease produces a constellation of findings that mimic those found in endotoxin shock; however, the pathogenic mechanisms differ.

A different region of the enterotoxin molecule is responsible for the symptoms of food poisoning. The enterotoxins are heat stable and can survive conditions that kill the bacteria. Illness results from the ingestion of preformed toxin; as a result, the incubation period is short (1–6 h). The toxin stimulates the vagus nerve and the vomiting center of the brain. It also appears to stimulate intestinal peristaltic activity.

EXFOLIATIVE TOXINS AND SSSS The exfoliative toxins are responsible for SSSS, most commonly seen in newborns. The toxins that produce disease in humans are of two serotypes: ETA and ETB. These toxins are serine proteases that cleave desmosomal cadherins in the superficial layer of the skin, triggering exfoliation. The result is a split in the epidermis at the granular level, which is responsible for the superficial desquamation of the skin that typifies this illness.

DIAGNOSIS

Staphylococcal infections are readily diagnosed by Gram's stain (Fig. 152-1) and microscopic examination of abscess contents or of infected tissue. Routine cultures of infected material usually are positive; blood cultures are sometimes positive even when infections are localized to extravascular sites. *S. aureus* is rarely a blood culture contaminant. Polymerase chain reaction (PCR)–based assays are now often used for the rapid diagnosis of *S. aureus* infection. A number of point-of-care tests are available to screen patients for colonization with MRSA. Determining whether patients with documented *S. aureus* bacteremia also have infective endocarditis or a metastatic focus of infection remains a diagnostic challenge. Uniformly positive cultures of blood collected over time suggest an endovascular infection such as endocarditis (see "Bacteremia, Sepsis, and Infective Endocarditis," below).

CLINICAL SYNDROMES

(Table 152-1)

Skin and Soft Tissue Infections *S. aureus* causes a variety of cutaneous infections. Common factors predisposing to *S. aureus* cutaneous infection include chronic skin conditions (e.g., eczema), skin damage (e.g., insect bites, minor trauma), injections (e.g., in diabetes, injection drug use), and poor personal hygiene. These infections are characterized by the formation of pus-containing blisters, which often begin in hair follicles and spread to adjoining tissues. *Folliculitis* is a superficial infection that involves the hair follicle, with a central area of purulence (pus) surrounded by induration and erythema. *Furuncles* (boils) are more extensive, painful lesions that tend to occur in hairy, moist regions of the body and extend from the hair follicle to become a true abscess with an area of central purulence. *Carbuncles* are most often located in the lower neck and are even more severe and painful, resulting from the coalescence of other lesions that extend to a deeper layer of the subcutaneous tissue. In general, furuncles and carbuncles are readily apparent, with pus often expressible or discharging from the abscess. Other cutaneous *S. aureus* infections include impetigo

TABLE 152-1 Common Illnesses Caused by *Staphylococcus aureus*

Skin and Soft Tissue Infections
Folliculitis
Abscess, furuncle, carbuncle
Cellulitis
Impetigo
Mastitis
Surgical wound infections
Musculoskeletal Infections
Septic arthritis
Osteomyelitis (hematogenous or contiguous spread)
Pyomyositis
Psoas abscess
Respiratory Tract Infections
Ventilator-associated or nosocomial pneumonia
Septic pulmonary emboli
Postviral pneumonia (e.g., influenza)
Empyema
Bacteremia and Its Complications
Sepsis, septic shock
Metastatic foci of infection (kidney, joints, bone, lung)
Infective endocarditis
Infective Endocarditis
Injection drug use–associated
Native-valve
Prosthetic-valve
Nosocomial
Device-Related Infections (e.g., intravascular catheters, prosthetic joints)
Toxin-Mediated Illnesses
Toxic shock syndrome
Food poisoning
Staphylococcal scalded-skin syndrome
Invasive Infections Associated with Community-Acquired Methicillin-Resistant *S. aureus*
Necrotizing fasciitis
Waterhouse-Friderichsen syndrome
Necrotizing pneumonia
Purpura fulminans

and cellulitis. *S. aureus* is one of the most common causes of surgical wound infections.

Mastitis develops in 1–3% of nursing mothers. This infection of the breast, which generally presents within 2–3 weeks after delivery, is characterized by findings that range from cellulitis to abscess formation. Systemic signs, such as fever and chills, are often present in more severe cases.

Musculoskeletal Infections *S. aureus* is a common cause of bone infections—both those resulting from hematogenous dissemination and those arising from contiguous spread from a soft tissue site. *Hematogenous osteomyelitis* in children most often involves the long bones. Infections present with fever and bone pain or with a child's reluctance to bear weight. The white blood cell count and erythrocyte sedimentation rate are often elevated. Blood cultures are positive in ~50% of cases. When necessary, bone biopsies for culture and histopathologic examination are usually diagnostic.

In adults, hematogenous osteomyelitis involving the long bones is less common. However, *vertebral osteomyelitis* is among the more common clinical presentations. Vertebral bone infections are most often seen in patients with endocarditis, those undergoing hemodialysis, diabetics, and injection drug users. These infections may present with intense back pain and fever but may also be clinically occult, presenting as chronic back pain with low-grade fever. *S. aureus* is the most common cause of epidural abscess, a complication that can result in neurologic compromise. Patients report difficulty voiding or walking and radicular pain in addition to the symptoms associated with their osteomyelitis. Surgical intervention in this setting often constitutes a medical emergency.

Magnetic resonance imaging (MRI) is the most reliable imaging modality to help establish the diagnosis of osteomyelitis **(Fig. 152-3)**. Routine x-rays are an appropriate first step, but findings may be normal for up to 14 days after the onset of symptoms. If an MRI is not possible, computed tomography (CT) is an acceptable alternative.

Bone infections that result from contiguous spread tend to develop from soft tissue infections, such as those associated with diabetic or vascular ulcers, surgery, or trauma. Exposure of bone, a draining fistulous tract, failure to heal, or continued drainage suggests involvement of underlying bone. Bone involvement is established by bone culture and histopathologic examination (revealing evidence of PMN infiltration). Contamination of culture material from adjacent tissue can make the diagnosis of osteomyelitis difficult in the absence of pathologic confirmation. Samples obtained during surgery are the most reliable. An MRI is the most reliable radiologic test to distinguish between osteomyelitis and overlying soft tissue infection with underlying osteitis.

In both children and adults, *S. aureus* is the most common cause of *septic arthritis* in native joints. If left untreated, this infection is rapidly progressive and may be associated with extensive joint destruction. It presents with intense pain on motion of the affected joint, swelling, and fever. Aspiration of the joint reveals turbid fluid, with >50,000 PMNs/μL and gram-positive cocci in clusters seen on Gram's stain (Fig. 152-1). In adults, septic arthritis may result from trauma, surgery, or hematogenous dissemination. The most commonly involved joints include the knees, shoulders, hips, and phalanges. Infection frequently develops in joints previously damaged by osteoarthritis or rheumatoid

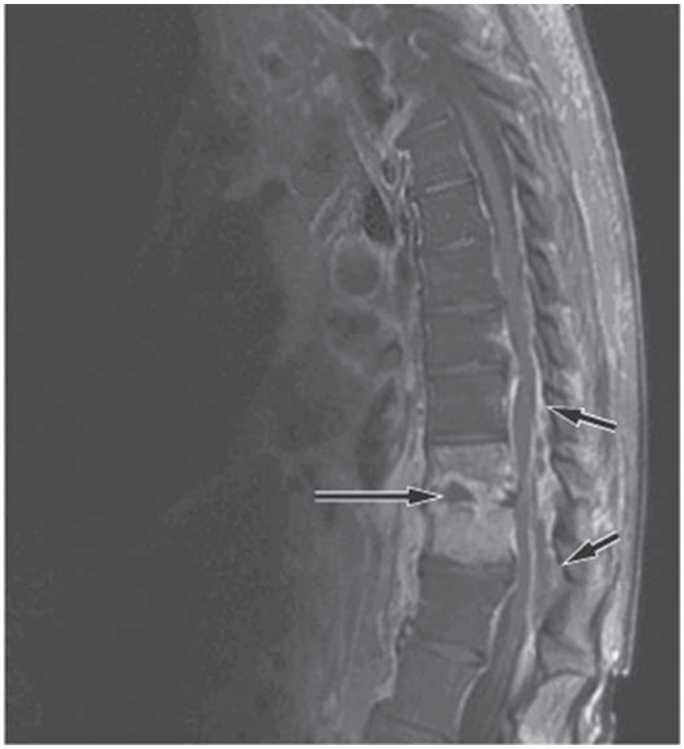

FIGURE 152-3 *S. aureus* vertebral osteomyelitis and epidural abscess involving the thoracic disk between T9 and T10. Sagittal postcontrast magnetic resonance imaging of the spine illustrates destruction of the T9–T10 intervertebral space with enhancement (*long arrow*). There is impingement on the thoracic cord and an epidural collection extending from T9 through T11 (*short arrows*).

arthritis. Iatrogenic infections resulting from aspiration or injection of agents into the joint also occur. In these settings, the patient experiences increased pain and swelling in the involved joint in association with fever.

Pyomyositis is an unusual infection of skeletal muscles that is seen primarily in tropical climates but also occurs in immunocompromised (e.g., HIV-infected) patients. It is believed to arise from occult bacteremia. Pyomyositis presents as fever, swelling, and pain overlying the involved muscle. Aspiration of fluid from the involved tissue yields pus. Although a history of trauma may be associated with the infection, its pathogenesis is poorly understood.

Respiratory Tract Infections Respiratory tract infections caused by *S. aureus* occur in selected clinical settings. *S. aureus* is a cause of serious respiratory tract infections in newborns and infants; these infections present with shortness of breath, fever, and respiratory failure. Chest x-ray may reveal pneumatoceles (shaggy, thin-walled cavities). Pneumothorax and empyema are recognized complications.

In adults, nosocomial *S. aureus* pulmonary infections are common among intubated patients in intensive care units. Nasally colonized patients are at increased risk of these infections. The clinical presentation is no different from pulmonary infections caused by other bacterial pathogens. Patients produce increased volumes of purulent sputum and develop respiratory distress, fever, and new pulmonary infiltrates. Distinguishing bacterial pneumonia from respiratory failure or other causes of new pulmonary infiltrates in critically ill patients is difficult and relies on a constellation of clinical, radiologic, and laboratory findings.

Community-acquired respiratory tract infections due to *S. aureus* often follow viral infections—most commonly influenza. Patients may present with fever, bloody sputum production, and midlung-field pneumatoceles or multiple, patchy pulmonary infiltrates. Diagnosis is made by sputum Gram's stain and culture. Blood cultures, although useful, are usually negative.

Bacteremia, Sepsis, and Infective Endocarditis *S. aureus* bacteremia may be complicated by sepsis, endocarditis, vasculitis, or metastatic seeding (establishment of suppurative collections at other tissue sites). Among the more commonly seeded tissue sites are bones, joints, kidneys, and lungs. The frequency of metastatic seeding during bacteremia has been estimated to be as high as 31%. The incidence of these complications increases with the duration of the bacteremia.

Recognition of these complications by clinical criteria alone is challenging. Comorbid conditions that are frequently seen in association with *S. aureus* bacteremia and that increase the risk of complications include diabetes, HIV infection, and renal insufficiency. Other host factors that increase the risk of complications include presentation with community-acquired *S. aureus* bacteremia, lack of an identifiable primary focus of infection, and the presence of prosthetic devices or material.

Clinically, *S. aureus* sepsis presents in a manner similar to that documented for sepsis due to other bacteria. The well-described progression of hemodynamic changes—beginning with respiratory alkalosis and clinical findings of hypotension and fever—is commonly seen. The microbiologic diagnosis is established by positive blood cultures.

The overall incidence of *S. aureus* endocarditis has increased over the past 20 years. *S. aureus* is now the leading cause of endocarditis worldwide, accounting for 25–35% of cases. This increase is due, at least in part, to the increased use of intravascular devices and, more recently, the upsurge in injection drug use. Studies using transesophageal echocardiography found an endocarditis incidence of ~25% among patients with intravascular catheter–associated *S. aureus* bacteremia. Other factors associated with an increased risk of endocarditis are hemodialysis, the presence of intravascular prosthetic devices at the time of bacteremia, and immunosuppression. Patients with implantable cardiac devices (e.g., permanent pacemakers) are at increased risk of endocarditis or device-related infections. Despite the availability of effective antibiotics, mortality rates from these infections continue to range from 20 to 40%, depending on both the host and the nature of the infection. Complications of *S. aureus* endocarditis include cardiac valvular insufficiency, peripheral emboli, metastatic seeding, vasculitis, and central nervous system (CNS) involvement (e.g., mycotic aneurysms, embolic strokes).

S. aureus endocarditis is encountered in four clinical settings: (1) right-sided endocarditis in association with injection drug use; (2) left-sided native-valve endocarditis; (3) prosthetic-valve endocarditis; and (4) nosocomial endocarditis. In each of these settings, the diagnosis is suspected from the patient's history and the recognition of physical signs suggestive of endocarditis. These findings include cardiac manifestations, such as new or changing cardiac valvular murmurs; cutaneous evidence, such as vasculitic lesions, Osler's nodes, or Janeway lesions; evidence of right- or left-sided embolic disease; and a history suggesting a risk for *S. aureus* bacteremia. In the absence of antecedent antibiotic therapy, blood cultures are almost uniformly positive. Transthoracic echocardiography, while less sensitive than transesophageal echocardiography, is less invasive and often identifies valvular vegetations. The Duke criteria **(Chap. 133)** are commonly used to help establish this diagnosis.

Acute right-sided tricuspid valvular *S. aureus* endocarditis is most often seen in patients who inject drugs. The classic presentation includes a high fever, a toxic clinical appearance, pleuritic chest pain, and the production of purulent, sometimes bloody, sputum. Chest x-rays or CT scans reveal evidence of septic pulmonary emboli (small, peripheral, circular lesions that may cavitate with time) **(Fig. 152-4)**. A high percentage of affected patients have no history of antecedent valvular damage. At the outset of their illness, patients may present with fever alone, without cardiac or other localizing findings. As a result, a high index of clinical suspicion is essential for diagnosis.

Individuals with antecedent cardiac valvular damage more commonly present with left-sided native-valve endocarditis involving the damaged valve. These patients tend to be older than those with right-sided endocarditis, their prognosis is worse, and their incidence of complications (including peripheral emboli, cardiac decompensation, cerebrovascular events, and metastatic seeding) is increased.

S. aureus is one of the more common causes of prosthetic-valve endocarditis. This infection is especially fulminant in the early postoperative period and is associated with increased morbidity and mortality. In most instances, medical therapy alone is not sufficient and urgent valve replacement is necessary. Patients are prone to develop valvular insufficiency or myocardial abscesses originating from the region of valve implantation.

The increased frequency of nosocomial endocarditis (15–30% of cases, depending on the series) reflects in part the increased use of intravascular devices. This form of endocarditis is most commonly caused by *S. aureus*. These patients are often critically ill, are receiving antibiotics for various other indications, and have comorbid conditions. As a result, blood cultures may be negative, and the diagnosis missed.

Prosthetic Device–Related Infections *S. aureus* accounts for a large proportion of prosthetic device–related infections. These infections include intravascular and peritoneal catheters, prosthetic valves, orthopedic devices, pacemakers, left ventricular assist devices, or vascular grafts. In contrast with the more indolent presentation of

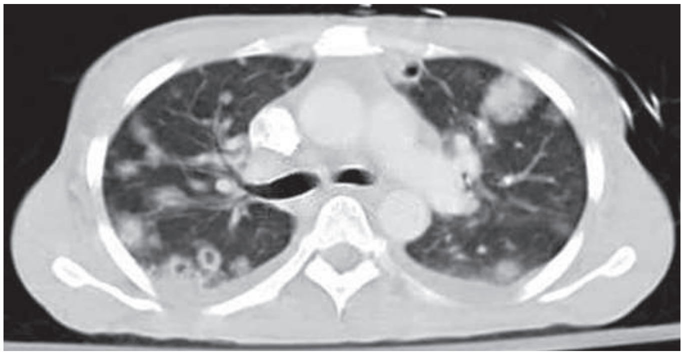

FIGURE 152-4 Computed tomography scan illustrating septic pulmonary emboli in a patient with methicillin-resistant *Staphylococcus aureus* bacteremia.

NSaS infections, *S. aureus* device-related infections are often acute, have both local and systemic manifestations, and tend to progress more rapidly. It is relatively common for a pyogenic collection to be present at the device site. Aspiration of these collections and performance of blood cultures are important components in establishing a diagnosis. *S. aureus* infections tend to occur more commonly soon after implantation unless the device is used for access (e.g., intravascular or hemodialysis catheters). In the latter instance, infections can occur at any time. As in most prosthetic-device infections, successful therapy usually involves removal of the device. Left in place, the device serves as a potential nidus for either persistent or recurrent infections.

Urinary Tract Infections Urinary tract infections (UTIs) are infrequently caused by *S. aureus*. The presence of *S. aureus* in the urine often suggests hematogenous dissemination. Ascending *S. aureus* infections occasionally result from instrumentation of the genitourinary tract.

Infections Associated with Community-Acquired MRSA Although skin and soft tissues are by far the most common sites of infection associated with CA-MRSA, 5–10% of these infections are invasive and can be life-threatening. The latter unique infections, including necrotizing fasciitis, necrotizing pneumonia, and sepsis with Waterhouse-Friderichsen syndrome or purpura fulminans, were rarely associated with *S. aureus* prior to the emergence of CA-MRSA. These life-threatening infections reflect the increased virulence of CA-MRSA strains.

Toxin-Mediated Diseases • **FOOD POISONING** *S. aureus* is among the most common causes of foodborne outbreaks in the United States. Staphylococcal food poisoning results from the inoculation of toxin-producing *S. aureus* into food by colonized food handlers. Toxin is then elaborated in such growth-promoting food as custards, potato salad, or processed meats. Even if the bacteria are killed by warming, the heat-stable toxin is not destroyed. The onset of illness is rapid, occurring within 1–6 h of ingestion; it is characterized by nausea and vomiting, although diarrhea, hypotension, and dehydration may occur. The differential diagnosis includes diarrhea of other etiologies, especially that caused by similar toxins (e.g., the toxins elaborated by *Bacillus cereus*). The rapidity of onset, the absence of fever, and the epidemic nature of the presentation (without secondary spread) should arouse suspicion of staphylococcal food poisoning. Symptoms generally resolve within 8–10 h. The diagnosis can be established by the demonstration of bacteria or the documentation of enterotoxin in the implicated food. Treatment is entirely supportive.

TOXIC SHOCK SYNDROME TSS gained attention in the early 1980s, when a nationwide outbreak occurred among young, otherwise healthy, menstruating women. Epidemiologic investigation demonstrated that these cases were associated with the use of a highly absorbent tampon recently introduced to the market. Subsequent studies established the role of TSST-1 in these illnesses. Withdrawal of the tampon from the market resulted in a rapid decline in the incidence of this disease. However, menstrual and nonmenstrual cases continue to be reported. Nonmenstrual cases are seen in patients with surgical or postpartum wound infections, especially when packing of the wound occurs.

The clinical presentation is similar in menstrual and nonmenstrual TSS. Evidence of clinical *S. aureus* infection is not a prerequisite. TSS results from the elaboration of an enterotoxin or the structurally related enterotoxin-like TSST-1. More than 90% of menstrual cases are caused by TSST-1, whereas a high percentage of nonmenstrual cases are caused by enterotoxins (e.g., enterotoxin B). TSS begins with relatively nonspecific flulike symptoms. In menstrual cases, the onset usually comes 2 or 3 days after the start of menstruation. Patients present with fever, hypotension, and erythroderma of variable intensity. Mucosal involvement is common (e.g., conjunctival hyperemia). The illness can rapidly progress to symptoms that include vomiting, diarrhea, confusion, myalgias, and abdominal pain. These symptoms reflect the multisystemic nature of the disease, with involvement of the liver, kidneys, gastrointestinal tract, and/or CNS. Desquamation of the skin occurs during convalescence, usually 1–2 weeks after the onset of illness. Laboratory findings may include azotemia, leukocytosis, hypoalbuminemia, thrombocytopenia, and liver function abnormalities.

TABLE 152-2 Case Definition of *Staphylococcus aureus* Toxic Shock Syndrome

Clinical Criteria

An illness with the following clinical manifestations:

- Fever: temperature ≥102.0°F (≥38.9°C)
- Rash: diffuse macular erythroderma
- Desquamation: 1–2 weeks after rash onset
- Hypotension: systolic blood pressure ≤90 mmHg for adults or less than the fifth percentile, by age, for children <16 years old
- Multisystem involvement (≥3 of the following organ systems)
    - Gastrointestinal: vomiting or diarrhea at illness onset
    - Muscular: severe myalgia or creatine phosphokinase level at least twice ULN
    - Mucous membrane: vaginal, oropharyngeal, or conjunctival hyperemia
    - Renal: blood urea nitrogen or creatinine level at least twice ULN for laboratory or urinary sediment with pyuria (≥5 leukocytes per high-power field) in the absence of urinary tract infection
    - Hepatic: total bilirubin or aminotransferase level at least twice ULN for laboratory
    - Hematologic: platelet count $<10^5$/mL
    - Central nervous system: disorientation or alterations in consciousness without focal neurologic signs in the absence of fever and hypotension

Laboratory Criteria

Negative results in the following tests, if obtained:

- Blood or cerebrospinal fluid cultures for another pathogen[a]
- Serologic tests for Rocky Mountain spotted fever, leptospirosis, or measles

Case Classification

Probable: a case that meets the laboratory criteria and in which four of the five clinical criteria are fulfilled

Confirmed: a case that meets the laboratory criteria and in which all five of the clinical criteria are fulfilled, including desquamation (unless the patient dies before desquamation occurs)

[a] Blood cultures may be positive for *S. aureus*.

Abbreviation: ULN, upper limit of normal.

Source: Centers for Disease Control and Prevention (*www.cdc.gov/nndss/conditions/toxic-shock-syndrome-other-than-streptococcal/case-definition/2011/*).

Diagnosis of TSS still depends on a constellation of findings rather than one specific finding and on a lack of evidence of other possible infections (Table 152-2). These other diagnoses include drug toxicities, viral exanthems, Rocky Mountain spotted fever, sepsis, and Kawasaki disease. Illness occurs only in persons who lack antibody to TSST-1. Recurrences are possible if antibody fails to develop after the illness.

STAPHYLOCOCCAL SCALDED-SKIN SYNDROME SSSS primarily affects newborns and children. The illness may vary from a localized blister to exfoliation of much of the skin surface. The skin is usually fragile and often tender, with thin-walled, fluid-filled bullae (Fig. 152-5). Gentle pressure results in rupture of the lesions, leaving denuded underlying skin. The mucous membranes are usually spared. In more generalized infection, there are often constitutional symptoms, including fever, lethargy, and irritability with poor feeding. Significant amounts of fluid can be lost in more extensive cases. Illness usually follows localized infection at one of several possible sites. SSSS is much less common among adults but can follow infections caused by exfoliative toxin–producing strains.

NON–*S. AUREUS* STAPHYLOCOCCAL INFECTIONS

Although less virulent than *S. aureus*, NSaS are among the most common causes of prosthetic-device infections, including endocarditis. They also are increasingly a cause of native-valve endocarditis and life-threatening bloodstream infections in neonates and in neutropenic patients. Approximately half of the identified NSaS species have been associated with human infections. Of these species, *Staphylococcus*

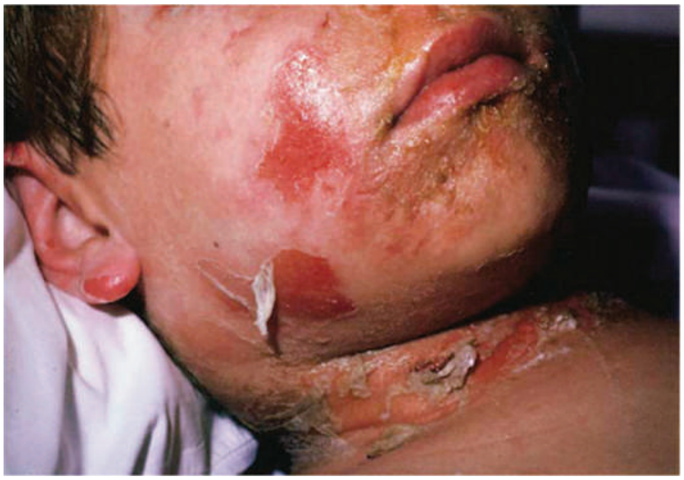

FIGURE 152-5 Staphylococcal scalded skin syndrome in a 6-year-old boy. Nikolsky's sign, with separation of the superficial layer of the outer epidermal layer, is visible. *(Adapted from LA Schenfeld: Staphylococcal scalded skin syndrome: N Engl J Med 342:1178, 2000.)*

epidermidis is the most common human pathogen. It is part of the normal human flora and is found on the skin (where it is the most abundant bacterial species) as well as in the oropharynx and vagina. *Staphylococcus saprophyticus*, a novobiocin-resistant species, is a common pathogen in UTIs.

PATHOGENESIS

S. epidermidis is the NSaS species most often associated with prosthetic-device infections. Infection is a two-step process, with initial adhesion to the device followed by colonization. *S. epidermidis* is uniquely adapted to colonize these devices because of its capacity to elaborate the extracellular polysaccharide (glycocalyx or slime) that facilitates formation of a protective biofilm on the device surface.

Implanted prosthetic material is rapidly coated with host matrix molecules such as fibrinogen or fibronectin. These molecules serve as potential bridging ligands, facilitating initial bacterial attachment to the device surface. A number of staphylococcal surface-associated proteins, such as autolysin (AtlE), fibrinogen-binding protein, and accumulation-associated protein (AAP), appear to play a role in attachment to either modified or unmodified prosthetic surfaces. The polysaccharide intercellular adhesin facilitates subsequent staphylococcal colonization, aggregation, and accumulation on the device surface. Intercellular adhesin (*ica*) genes are more commonly found in strains of *S. epidermidis* that are associated with device infections than in strains associated with colonization of mucosal surfaces. Biofilm acts as a barrier, protecting bacteria from host defense mechanisms as well as from antibiotics while providing a suitable environment for bacterial maturation, survival, and potential spread to other tissue sites.

Two additional NSaS species, *Staphylococcus lugdunensis* and *Staphylococcus schleiferi*, produce more serious infections (native-valve endocarditis and osteomyelitis) than do other NSaS. The basis for this enhanced virulence is not known, although both species appear to share more virulence determinants with *S. aureus* (e.g., clumping factor and lipase) than do other NSaS.

The capacity of *S. saprophyticus* to cause UTIs in young women appears related to the presence of adhesins that facilitate adherence to uroepithelial cells. A 160-kDa hemagglutinin/adhesin may contribute to this affinity.

DIAGNOSIS

Although the detection of NSaS at sites of infection or in the bloodstream by standard microbiologic culture methods is not difficult, interpretation of these results is frequently problematic. Because these organisms are present in large numbers on the skin, they often contaminate cultures. It has been estimated that only 10–20% of blood cultures positive for NSaS reflect true bacteremia. Similar problems arise with cultures obtained from other sites. Among the clinical findings suggestive of true bacteremia are fever, evidence of local infection (e.g., erythema or purulent drainage at the IV catheter site), leukocytosis, and systemic signs of sepsis. Laboratory findings suggestive of true bacteremia include repeated isolation of the same strain (i.e., the same species with the same antibiogram or with a closely related DNA fingerprint) from separate cultures, growth of the strain within 48 h, and bacterial growth in both aerobic and anaerobic bottles.

CLINICAL SYNDROMES

NSaS cause a variety of prosthetic device–related infections, including those that involve prosthetic cardiac valves and joints, vascular grafts, intravascular devices, and CNS shunts. In all of these settings, the clinical presentation is similar. The signs of localized infection are often subtle, the rate of disease progression is slow, and the systemic findings are often limited. Signs of infection, such as purulent drainage, pain at the site, or loosening of prosthetic implants, are sometimes evident. Fever is frequently but not always present, and there may be mild leukocytosis. Acute-phase reactant levels, erythrocyte sedimentation rate, and C-reactive protein concentration may be elevated.

Infections that are not associated with prosthetic devices include, as noted, native-valve endocarditis due to NSaS, which accounts for ~5% of cases. Infections in preterm infants and neutropenic patients are often associated with the need for intravascular devices. *S. lugdunensis* appears to be a more aggressive pathogen in this setting, causing greater mortality and rapid valvular destruction with abscess formation than other NSaS.

TREATMENT

Staphylococcal Infections

GENERAL PRINCIPLES OF THERAPY

Source control (e.g., incision and drainage of suppurative collections or removal of infected prosthetic devices), coupled with rapid institution of appropriate antimicrobial therapy, is essential for the management of all staphylococcal infections. The emergence of MRSA as a community-based pathogen has increased the importance of culturing all sites of infection to determine antimicrobial susceptibility and optimize oral treatment regimens.

DURATION OF ANTIMICROBIAL THERAPY

Therapy for *S. aureus* bacteremia is generally prolonged (4–6 weeks) because of the high risk of complications (e.g., endocarditis, metastatic foci of infection). Among the findings associated with complicated bacteremias are (1) persistently positive blood cultures 96 h after institution of therapy, (2) failure to promptly remove or drain an identified focus of infection (i.e., an intravascular catheter), (3) the presence of deep-seated infections, and (4) acquisition of the infection in the community. Patients with uncomplicated bacteremias are defined by a removable focus of infection, prompt response to antimicrobial therapy (i.e., no fever or positive blood cultures after 3–4 days), no evidence of metastatic foci of infection, and no implanted prostheses. In these latter infections, short-course therapy (2 weeks) can be given; however, these findings are not always predictive of an uncomplicated bacteremia. Given these concerns, caution is therefore needed in instituting a short course of therapy. Transesophageal echocardiography to rule out endocarditis is generally necessary because neither clinical nor laboratory findings can reliably detect cardiac involvement. A thorough radiologic investigation to identify potential metastatic collections is also indicated. All symptomatic body sites must be carefully evaluated.

Recent studies have demonstrated that parenteral therapy is not always necessary to complete a course of treatment for invasive staphylococcal infections such as endocarditis or osteomyelitis for carefully selected patients. These include patients with uncomplicated staphylococcal bacteremia.

NSaS treatment is complicated by the possibility that a single isolate may be a contaminant. Therapy for 7–14 days is recommended for documented infections (i.e., blood cultures of the same strain ≥24 h apart) in the absence of endocarditis or additional sites of infection.

CHOICE OF ANTIMICROBIAL AGENTS

The choice of antimicrobial agents to treat both coagulase-positive and coagulase-negative staphylococcal infections is often difficult because of the prevalence of multidrug-resistant strains and the limited number of clinical trials that have compared the available agents. Staphylococcal resistance to most antibiotic families, including β-lactams, aminoglycosides, fluoroquinolones, and (to a lesser extent) glycopeptides, has increased. This trend is even more apparent with NSaS; >80% of nosocomial isolates are resistant to methicillin, and these methicillin-resistant strains are often resistant to other antibiotics. Because the selection of antimicrobial agents for *S. aureus* infections is similar to that for NSaS infections, treatment options for these pathogens are discussed together and are summarized in **Table 152-3**.

TABLE 152-3 Antimicrobial Therapy for Staphylococcal Infections[a]

SENSITIVITY/ RESISTANCE OF ISOLATE	DRUG OF CHOICE	ALTERNATIVE(S)	COMMENTS
Parenteral Therapy for Serious Infections			
Sensitive to penicillin	Penicillin G (4 mU q4h)	Nafcillin or oxacillin (2 g q4h), cefazolin (2 g q8h), vancomycin (15–20 mg/kg q8h[b])	Fewer than 5% of isolates are sensitive to penicillin. The clinical microbiology laboratory must verify that the strain is not a β-lactamase producer.
Sensitive to methicillin; resistant to penicillin	Nafcillin or oxacillin (2 g q4h), cefazolin (2 g q8h)	Daptomycin (6–10 mg/kg IV q24h[b,d]), vancomycin (15–20 mg/kg q8h[b]), ceftobiprole (500 mg IV q6h[g])	Patients with a penicillin allergy can be treated with a cephalosporin if the allergy does not involve an anaphylactic or accelerated reaction; desensitization to β-lactams may be indicated in selected cases of serious infection when maximal bactericidal activity is needed (e.g., prosthetic-valve endocarditis[c]). Vancomycin is a less effective option than a β -lactam.
Resistant to methicillin	Vancomycin (15–20 mg/kg q8–12h[b]), daptomycin (6–10 mg/kg IV q24h[b,d]) for bacteremia, endocarditis, osteomyelitis, and complicated skin infections	Linezolid (600 mg q12h PO or IV), ceftaroline (600 mg IV q8–12h), telavancin (7.5–10 mg/kg IV q24h)[b], TMP-SMX (5 mg [based on TMP]/kg IV q8–12h)[f] Additional agents include tedizolid (200 mg once daily IV), oritavancin (single dose of 1200 mg), dalbavancin (single dose of 1500 mg), delafloxacin (300 mg q 12 h IV), omadacycline 100 mg OD). Ceftobiprole (500 mg IV q6h[g])	Sensitivity testing is necessary before an alternative drug is selected. The efficacy of adjunctive therapy is not well established in many settings. Linezolid, ceftaroline, and telavancin have in vitro activity against most VISA and VRSA strains. See footnote for treatment of prosthetic-valve endocarditis.[c]
Resistant to methicillin with intermediate or complete resistance to vancomycin[e]	Daptomycin (6–10 mg/kg q24h[b,d]) for bacteremia, endocarditis, osteomyelitis, and complicated skin infections	Same as for methicillin-resistant strains (check antibiotic susceptibilities) *or*	Same as for methicillin-resistant strains; check antibiotic susceptibilities. Ceftaroline is used either alone or in combination with daptomycin.
		Ceftaroline (600 mg IV q8–12h) Newer agents include tedizolid (200 mg once daily IV or PO), oritavancin (single dose of 1200 mg), and dalbavancin (single dose of 1500 mg). These drugs are approved only for the treatment of skin and soft tissue infections.	
Not yet known (i.e., empirical therapy)	Vancomycin (15–20 mg/kg q8–12h[b]), daptomycin (6–10 mg/kg q24h[b,d]) for bacteremia, endocarditis, osteomyelitis, and complicated skin infections	—	Empirical therapy is given when the susceptibility of the isolate is not known. Vancomycin with or without a β-lactam is recommended for suspected community- or hospital-acquired *Staphylococcus aureus* infections because of the increased frequency of methicillin-resistant strains in the community. If isolates with an elevated MIC to vancomycin (≥1.5 μg/mL) are common in the community, daptomycin may be preferable.
Oral Therapy for Skin and Soft Tissue Infections			
Sensitive to methicillin	Dicloxacillin (500 mg qid), cephalexin (500 mg qid), or cefadroxil (1 g q12h)	Minocycline or doxycycline (100 mg q12h[b]), TMP-SMX (1 or 2 DS tablets bid), clindamycin (300–450 mg tid), linezolid (600 mg PO q12h), tedizolid (200 mg PO q24h)	It is important to know the antibiotic susceptibility of isolates in the specific geographic region. All collections should be drained, and drainage should be cultured.
Resistant to methicillin	Clindamycin (300–450 mg tid), TMP-SMX (1 or 2 DS tablets bid), minocycline or doxycycline (100 mg q12h[b]), linezolid (600 mg bid), or tedizolid (200 mg once daily)	Delafloxacin 450 mg q12 h, omadacycline 300 mg once a day	It is important to know the antibiotic susceptibility of isolates in the specific geographic region. All collections should be drained, and drainage should be cultured.

[a] Recommended dosages are for adults with normal renal and hepatic function. [b] The dosage must be adjusted for patients with reduced creatinine clearance. [c] For the treatment of prosthetic-valve endocarditis, the addition of gentamicin (1 mg/kg q8h) and rifampin (300 mg PO q8h) is recommended, with adjustment of the gentamicin dosage if the creatinine clearance rate is reduced. [d] Daptomycin cannot be used for the treatment of pneumonia. [e] Vancomycin-resistant *S. aureus* isolates from clinical infections have been reported. [f] TMP-SMX may be less effective than vancomycin. [g] Additional studies are needed.

Abbreviations: DS, double-strength; TMP-SMX, trimethoprim-sulfamethoxazole; VISA, vancomycin-intermediate *S. aureus;* VRSA, vancomycin-resistant *S. aureus.*

Source: Modified from C Liu et al: Clin Infect Dis 52:285, 2011; DL Stevens et al: Clin Infect Dis 59:148, 2014; DL Stevens et al: Med Lett Drugs Ther 56:39, 2014; and LM Baddour et al: Circulation 132:1435, 2015.

Few strains of staphylococci (≤5%) remain susceptible to penicillin. This is a result of the widespread dissemination of plasmids containing the enzyme penicillinase. Penicillin-resistant isolates are treated with semisynthetic penicillinase-resistant penicillins (SPRPs), such as oxacillin or nafcillin. Methicillin, the first of the SPRPs, is no longer used. Cephalosporins are alternative therapeutic agents for these infections. In patients with a history of serious β-lactam allergies, alternatives to SPRPs for the treatment of invasive infections should be used only after careful consideration. Desensitization to β-lactams remains an option for life-threatening infections. Second- and third-generation cephalosporins offer no therapeutic advantage over first-generation cephalosporins for the treatment of staphylococcal infections, and some third-generation cephalosporins (e.g., ceftazidime, ceftriaxone) have considerably less activity and should be avoided. The carbapenems have excellent activity against methicillin-sensitive *S. aureus* but not against MRSA.

The isolation of MRSA was reported within 1 year of the introduction of methicillin. Since then, the prevalence of MRSA has steadily increased. In many U.S. hospitals and elsewhere, 40–50% of *S. aureus* isolates are resistant to methicillin. Resistance to methicillin indicates resistance to all SPRPs as well as to all cephalosporins (except ceftaroline). Production of a novel penicillin-binding protein (PBP2a) is responsible for methicillin resistance. This protein is synthesized by the *mecA* gene, which (as stated above) is part of a large mobile genetic element—a pathogenicity or genomic island—called SCC*mec*. It is hypothesized that *mecA* was acquired via horizontal transfer from related staphylococcal species. Phenotypic expression of methicillin resistance may be constitutive (i.e., expressed in all cells in a population) or heterogeneous (i.e., displayed by only a proportion of the total cell population). Detection of methicillin resistance is enhanced by growth of cultures at reduced temperatures (≤35°C for 24 h) and with increased concentrations of salt in the medium. Culture techniques are increasingly being replaced by PCR-based or other methods (e.g., latex agglutination) that allow for the rapid detection of methicillin resistance.

Either vancomycin or daptomycin is recommended as the drug of choice for the treatment of invasive MRSA infections. MRSA susceptibility to vancomycin has decreased in many areas of the world. It is important to note that vancomycin is less effective than SPRPs for the treatment of infections due to methicillin-susceptible strains.

Three types of staphylococcal resistance to vancomycin have emerged. (1) Minimal inhibitory concentration (MIC; an in vitro measure of susceptibility) "creep" refers to the incremental increase in vancomycin MICs that has been detected in various geographic areas. Studies suggest that morbidity and mortality may be increased in infections due to *S. aureus* strains with vancomycin MICs of ≥1.5 μg/mL. (2) In 1997, an *S. aureus* strain with reduced susceptibility to vancomycin (vancomycin-intermediate *S. aureus* [VISA]) was reported from Japan. Subsequently, additional VISA clinical isolates were reported. These strains were resistant to methicillin and many other antimicrobial agents. The VISA strains appear to evolve (under vancomycin selective pressure) from strains that are susceptible to vancomycin but are heterogeneous, with a small proportion of the bacterial population expressing the resistance phenotype. The mechanism of VISA resistance is in part due to an abnormally thick cell wall. Vancomycin is trapped by the abnormal peptidoglycan cross-linking and is unable to gain access to its target site. Regulatory genes involved in cell wall metabolism appear to play an important role in this type of resistance. (3) In 2002, the first clinical isolate of fully vancomycin-resistant *S. aureus* (VRSA) was reported. Resistance in this and several additional clinical isolates was due to the presence of *vanA*, the gene responsible for expression of vancomycin resistance in enterococci. This observation suggested that resistance was acquired as a result of horizontal conjugal transfer from a vancomycin-resistant strain of *Enterococcus faecalis.* Several of the patients infected with the VRSA strain had both MRSA and vancomycin-resistant enterococci cultured from infection sites. The *vanA* gene is responsible for the synthesis of the dipeptide D-Ala-D-Lac in place of D-Ala-D-Ala. Vancomycin cannot bind to the altered peptide. While isolates with MICs of ≥1.5 μg/mL have been relatively common in some areas, VISA and VRSA isolates are uncommon.

Daptomycin, a parenteral bactericidal agent with antistaphylococcal activity, is approved for the treatment of bacteremia (including right-sided endocarditis) and complicated skin infections. It is not effective in respiratory infections. This drug has a unique mechanism of action: it disrupts the cytoplasmic membrane. Staphylococcal resistance to daptomycin has been reported. Resistance can emerge during therapy; patients previously treated with vancomycin may have elevated daptomycin MICs. Patients need to be monitored for rhabdomyolysis with creatine phosphokinase measurement and for eosinophilic pneumonia.

Linezolid—the first oxazolidinone—is bacteriostatic against staphylococci; it offers the advantage of comparable bioavailability after oral or parenteral administration. Cross-resistance with other inhibitors of protein synthesis has not been detected. Resistance to linezolid is rare but has been reported. Serious adverse reactions to linezolid include thrombocytopenia, occasional cases of neutropenia, and rare instances of lactic acidosis or peripheral and optic neuropathy. These reactions tend to occur after relatively prolonged courses of therapy.

Tedizolid, a second oxazolidinone, is available as both oral and parenteral preparations. It exhibits enhanced in vitro activity against antibiotic-resistant gram-positive bacteria, including staphylococci. Tedizolid is administered once a day. Data on its efficacy for the treatment of deep-seated infections are limited.

Ceftaroline is a fifth-generation cephalosporin with bactericidal activity against MRSA (including strains with reduced susceptibility to vancomycin and daptomycin). It is generally well tolerated. Ceftaroline is approved for use in nosocomial pneumonias and for SSTIs. It has increasingly been used to treat invasive MRSA infections with or without glycopeptides.

Telavancin is a parenteral lipoglycopeptide derivative of vancomycin that is approved for the treatment of complicated SSTIs and for nosocomial pneumonias. The drug has two targets: the cell wall and the cell membrane. It remains active against VISA strains. Because of its potential nephrotoxicity, telavancin should be avoided in patients with renal disease.

Dalbavancin and oritavancin are long-acting, parenterally administered lipoglycopeptides that have been used to treat complicated SSTIs. Because of their long half-lives, they can be administered on a weekly basis. Both have been used as single-dose regimens for the treatment of SSTIs. Emerging data support their use for the treatment of invasive staphylococcal infections.

Although the quinolones are active against staphylococci in vitro, the frequency of staphylococcal resistance to these agents has increased, especially among methicillin-resistant isolates. Of particular concern in MRSA is the possibility of quinolone resistance emerging during therapy. Therefore, quinolones are not recommended for the treatment of MRSA infections. Resistance to the quinolones is most commonly chromosomal and results from mutations of the topoisomerase IV or DNA gyrase genes, although multidrug efflux pumps also may contribute. Although the newer quinolones exhibit increased in vitro activity against staphylococci, it is uncertain whether this increase translates into enhanced in vivo activity. Delafloxacin, a fluoroquinolone with broad-spectrum activity, has excellent activity against MRSA, retaining activity against some isolates resistant to other fluoroquinolones.

Tigecycline, a broad-spectrum minocycline analogue, has bacteriostatic activity against MRSA and is approved for use in SSTIs as well as intraabdominal infections caused by *S. aureus.* It is not recommended for the treatment of invasive infections.

Other older antibiotics, such as minocycline, doxycycline, clindamycin, and trimethoprim-sulfamethoxazole, continue to be successfully used to treat MRSA infections.

Ceftobiprole is a new antibiotic with excellent activity against both MRSA and methicillin-susceptible *S. aureus* (MSSA). It has been shown in clinical trials to be effective in treating complicated staphylococcal bacteremias, SSTIs, and pneumonia.

The benefit of antistaphylococcal combinations to enhance bactericidal activity in the treatment of deep-seated infections remains controversial. Clinical studies have not documented a therapeutic benefit from the addition of gentamicin to single-drug regimens; recent reports have raised concern about the potential nephrotoxicity of gentamicin and adverse reactions from, or drug interactions with, rifampin. As a result, the use of gentamicin in combination with β-lactams or other antimicrobial agents is no longer routinely recommended for the treatment of invasive infections such as native-valve endocarditis. Rifampin continues to be used for the treatment of prosthetic device–related infections and for osteomyelitis.

Omadacycline and eravacycline are broad-spectrum semisynthetic tetracycline derivatives with activity against MRSA. They are currently approved for the treatment of SSTIs.

The use of bacteriophages with activity against staphylococci is now being investigated in clinical trials as adjunctive therapy in invasive infections.

ANTIMICROBIAL THERAPY FOR SELECTED SETTINGS

Empirical Therapy Empirical coverage for MRSA is indicated when antibiotic susceptibility is not known. Vancomycin or daptomycin is generally recommended. It remains uncertain whether daptomycin is preferable when elevated vancomycin MICs (>1.5 μg/mL) are common in a specific locale.

Salvage Therapy Salvage therapy for complicated *S. aureus* infections is sometimes needed when the bacteremia persists (i.e., for 3 days) despite appropriate treatment. The risk of a poor outcome (i.e., increased mortality, metastatic infections) is increased with the duration of bacteremia. Prolonged bacteremia can occur with both MRSA and MSSA. There is limited high-quality evidence to serve as a guide to salvage therapy. The combination of daptomycin or vancomycin with a β-lactam antibiotic (e.g., ceftaroline) has been successfully used to treat patients with persistent MRSA bacteremia, even those patients with isolates displaying reduced susceptibility to these antimicrobial agents. This combination appears to enhance the bactericidal activity of daptomycin by reducing the bacterial cell-surface charge and thus allowing enhanced daptomycin binding. For vancomycin, the combination may allow more strategic binding to the target site with reduced cell-wall thickness. Other combinations have included trimethoprim-sulfamethoxazole or rifampin combined with daptomycin. Linezolid and ceftaroline have also been used as single alternative agents.

Endocarditis *S. aureus* endocarditis is usually an acute, life-threatening infection. Thus, prompt collection of blood for cultures should be followed by immediate institution of empirical antimicrobial therapy. For native-valve endocarditis, therapy with a β-lactam is recommended. If a MRSA strain is isolated, vancomycin (15–20 mg/kg every 8–12 h, given in equal doses up to a total of 2 g, with the dose adjusted in the case of renal disease) or daptomycin (6–10 mg/kg every 24 h) is recommended. The vancomycin dose should be adjusted based on area under the curve (AUC)-based dosing, although measurement of trough levels may also be used. Patients are generally treated for 6 weeks. For prosthetic-valve endocarditis, surgery in addition to antibiotic therapy is often necessary. The combination of a β-lactam agent—or, if the isolate is β-lactam-resistant, vancomycin or daptomycin—with an aminoglycoside (gentamicin, 1 mg/kg IV every 8 h) for 2 weeks and rifampin (300 mg orally or IV every 8 h) for ≥6 weeks is recommended. Infectious diseases and, if necessary, surgical consultation should be considered.

Bone and Joint Infections For hematogenous osteomyelitis or septic arthritis in children, a 4-week course of therapy is usually adequate. In adults, treatment is often more prolonged. For chronic forms of osteomyelitis, surgical debridement is necessary in combination with antimicrobial therapy. For joint infections, a critical component of therapy is the repeated aspiration or arthroscopy of the affected joint to prevent damage from leukocytes. The combination of rifampin with ciprofloxacin has been used successfully to treat or suppress prosthetic-joint infections, especially when the device cannot be removed. The efficacy of this combination may reflect enhanced activity against staphylococci in biofilms as well as the attainment of effective intracellular concentrations.

Skin and Soft Tissue Infections The increase in SSTIs caused by CA-MRSA has drawn attention to the need for initiation of appropriate empirical therapy. Even small abscesses appear to benefit from antibiotic therapy in addition to incision and drainage. Antibiotics are selected depending on local antibiotic susceptibility data; several oral agents have been used to treat these infections, including clindamycin, trimethoprim-sulfamethoxazole, doxycycline, linezolid, and tedizolid. Parenteral therapy is reserved for more complicated infections.

Toxic Shock Syndrome Treatment of shock is the mainstay of therapy for TSS. Both fluids and pressors may be necessary. Tampons or other packing material should be promptly removed. Some investigators recommend therapy with a combination of clindamycin and a semisynthetic penicillin or (if the isolate is resistant to methicillin) vancomycin. Clindamycin is advocated because, as a protein synthesis inhibitor, it reduces toxin production. Linezolid also appears to be effective. A semisynthetic penicillin or a glycopeptide is recommended to eliminate any potential focus of infection as well as to eradicate persistent carriage that might increase the possibility of recurrence. Intravenous immunoglobulin to treat TSS is of uncertain benefit. Glucocorticoids are not recommended for the treatment of this disease.

Other Toxin-Mediated Diseases Therapy for staphylococcal food poisoning is entirely supportive. For SSSS, antistaphylococcal therapy targets the primary site of infection.

NONTRADITIONAL APPROACHES TO ANTISTAPHYLOCOCCAL THERAPY

In addition to the development of new antibiotics, new and nontraditional approaches to therapy are currently being investigated. These include the use of phages or phage-derived peptides, as well as probiotics and antivirulence strategies that target selected virulence determinants.

■ PREVENTION

Primary prevention of *S. aureus* infections in the hospital setting involves hand washing and careful attention to appropriate isolation procedures. Through careful screening for MRSA carriage and strict isolation practices, several Scandinavian countries have been remarkably successful at preventing the introduction and dissemination of MRSA in hospitals.

Decolonization strategies, using both universal and targeted approaches with topical agents (e.g., mupirocin) to eliminate nasal colonization and/or chlorhexidine to eliminate colonization of additional body sites with *S. aureus*, have been successful in some clinical settings where the risk of infection is high (e.g., intensive care units). An analysis of clinical trials suggests that decolonization can reduce the incidence of postsurgical infections among people nasally colonized with *S. aureus*. The risk of recurrent admissions among patients with *S. aureus* bacteremia following discharge is high (~22% within 30 days). Decolonization following discharge with mupirocin and chlorhexidine can lower the incidence of recurrent infections.

"Bundling" (the application of selected medical interventions in a sequence of prescribed steps) has reduced rates of nosocomial infections related to procedures such as the insertion of intravenous catheters, in which staphylococci are among the most common pathogens **(see Table 147-1)**. A number of immunization strategies to prevent *S. aureus* infections—both active (e.g., capsular polysaccharide–protein conjugate vaccine) and passive (e.g., clumping factor antibody)—have been investigated. However, to date, none has been successful for either prophylaxis or therapy in clinical trials.

Strategies to prevent recurrent *S. aureus* infections in the community have had limited success. Decolonization with intranasal mupirocin and chlorhexidine washes of the infected individual and the additional decolonization of household members combined with environmental cleaning of surfaces and personal items have all been studied. For individuals with extensive skin disease and recurrent infections, the use of bleach baths (e.g., one-half cup of household bleach in a half-filled bathtub) for 15 minutes three times weekly may be useful.

FURTHER READING

Becker K et al: Coagulase-negative staphylococci. Clin Microbiol Rev 27:870, 2014.

Cheung GYC et al: Pathogenicity and virulence of *Staphylococcus aureus*. Virulence 12:547, 2021.

DeLeo FR et al: Community-associated methicillin-resistant *Staphylococcus aureus*. Lancet 375:1557, 2010.

Holland TL et al: Ceftobiprole for treatment of complicated *Staphylococcus aureus* bacteremia. N Engl J Med 389:1390, 2023.

Lee AS et al: Methicillin-resistant *Staphylococcus aureus*. Nat Rev Dis Primers 4:18033:1, 2018.

Minter DJ et al: Contemporary management of *Staphylococcus aureus* bacteremia—Controversies in clinical practice. Clin Infect Dis 77:e57, 2023.

Rose W et al: Current paradigms of combination therapy in methicillin-resistant Staphylococcus aureus (MRSA) bacteremia: Does it work, which combination, and for which patients? Clin Infect Dis 73:2353, 2021.

Siciliano V et al: Difficult-to-treat pathogens: A review on the management of multidrug- resistant *Staphylococcus epidermidis*. Life 13:1126, 2023.

Tong SY et al: *Staphylococcus aureus* infections: Epidemiology, pathophysiology, clinical manifestations, and management. Clin Microbiol Rev 28:603, 2015.

153

Streptococcal Infections

Michael R. Wessels

Many varieties of streptococci are found as part of the normal flora colonizing the human respiratory, gastrointestinal, and genitourinary tracts. Several species are important causes of human disease. Group A *Streptococcus* (GAS; *Streptococcus pyogenes*) is responsible for streptococcal pharyngitis, one of the most common bacterial infections of school-age children, and for the postinfectious syndromes of acute rheumatic fever (ARF) and poststreptococcal glomerulonephritis (PSGN). Group B *Streptococcus* (GBS; *Streptococcus agalactiae*) is a leading cause of bacterial sepsis and meningitis in newborns and a major cause of endometritis and fever in parturient women. Viridans streptococci are a common cause of bacterial endocarditis. Enterococci, which are morphologically similar to streptococci, are now considered a separate genus on the basis of DNA homology studies. Thus, the species previously designated as *Streptococcus faecalis* and *Streptococcus faecium* have been renamed *Enterococcus faecalis* and *Enterococcus faecium*, respectively. **The enterococci are discussed in Chap. 154.**

Streptococci are gram-positive, spherical to ovoid bacteria that characteristically form chains when grown in liquid media. Most streptococci that cause human infections are facultative anaerobes, although some are strict anaerobes. Streptococci are relatively fastidious organisms, requiring enriched media for growth in the laboratory. Clinicians and clinical microbiologists identify streptococci by several classification systems, including hemolytic pattern, Lancefield group, species name, and common or trivial name. Many streptococci associated with human infection produce a zone of complete (β) hemolysis around the bacterial colony when cultured on blood agar. The β-hemolytic streptococci that form large (≥0.5-mm) colonies on blood agar can be classified by the Lancefield system, a serologic grouping based on the reaction of specific antisera with bacterial cell-wall carbohydrate antigens. With rare exceptions, organisms belonging to Lancefield groups A, B, C, and G are all β-hemolytic, and each is associated with characteristic patterns of human infection. Other streptococci produce a zone of partial (α) hemolysis, often imparting a greenish appearance to the agar. These α-hemolytic streptococci are further identified by biochemical testing and include *Streptococcus pneumoniae* **(Chap. 151)**, an important cause of pneumonia, meningitis, and other infections, and the several species referred to collectively as the *viridans streptococci*, which are part of the normal oral flora and are important agents of subacute bacterial endocarditis. Finally, some streptococci are nonhemolytic, a pattern sometimes called γ *hemolysis*. Among the organisms classified serologically as group D streptococci, the enterococci are assigned to a distinct genus **(Chap. 154)**. The classification of the major streptococcal groups causing human infections is outlined in **Table 153-1**.

GROUP A STREPTOCOCCI

Lancefield group A consists of a single species, *S. pyogenes*. As its species name implies, this organism is associated with a variety of suppurative infections. In addition, GAS can trigger the postinfectious syndromes of ARF (which is uniquely associated with *S. pyogenes* infection; **Chap. 371**) and PSGN **(Chap. 326)**.

Worldwide, GAS infections and their postinfectious sequelae (primarily ARF and rheumatic heart disease) account for an estimated 500,000 deaths per year. Although data are incomplete, the incidence of all forms of GAS infection and that of rheumatic heart disease are thought to be tenfold higher in resource-limited countries than in

TABLE 153-1 Classification of Streptococci

LANCEFIELD GROUP	REPRESENTATIVE SPECIES	HEMOLYTIC PATTERN	TYPICAL INFECTIONS
A	*S. pyogenes*	β	Pharyngitis, impetigo, cellulitis, scarlet fever
B	*S. agalactiae*	β	Neonatal sepsis and meningitis, puerperal infection, urinary tract infection, diabetic ulcer infection, endocarditis
C, G	*S. dysgalactiae* subsp. *equisimilis*	β	Cellulitis, bacteremia, endocarditis
D	Enterococci[a]: *E. faecalis*, *E. faecium*	Usually nonhemolytic	Urinary tract infection, nosocomial bacteremia, endocarditis
	Nonenterococci: *S. gallolyticus* (formerly *S. bovis*)	Usually nonhemolytic	Bacteremia, endocarditis
Variable or nongroupable	Viridans streptococci: *S. sanguis*, *S. mitis*	α	Endocarditis, dental abscess, brain abscess
	Intermedius or *milleri* group: *S. intermedius*, *S. anginosus*, *S. constellatus*	Variable	Brain abscess, visceral abscess
	Anaerobic streptococci[b]: *Peptostreptococcus magnus*	Usually nonhemolytic	Sinusitis, pneumonia, empyema, brain abscess, liver abscess

[a] See **Chap. 154**. [b] See **Chap. 182.**

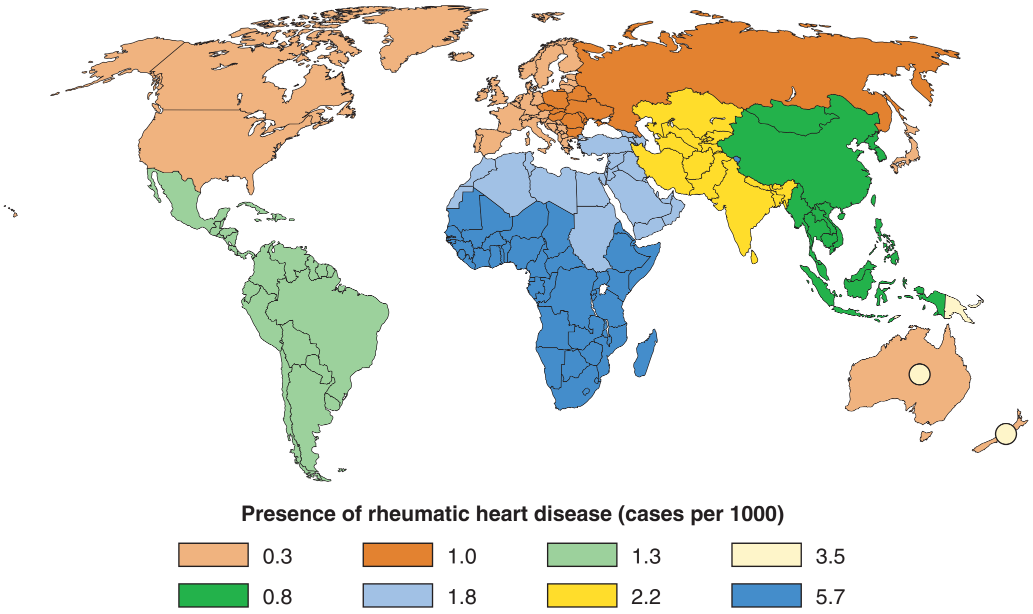

FIGURE 153-1 Prevalence of rheumatic heart disease in children 5–14 years old. The *circles* within Australia and New Zealand represent indigenous populations (and also Pacific Islanders in New Zealand). *(Reproduced with permission from JR Carapetis et al: The global burden of group A streptococcal diseases. Lancet Infect Dis 5:685, 2005.)*

developed countries **(Fig. 153-1)**. As has been observed with infections due to other pathogens associated with the human respiratory tract, the incidence of streptococcal pharyngitis and of invasive GAS infection fell during 2020–2022 in association with COVID-19-related social distancing restrictions. A rebound in GAS infection has been observed in many countries beginning in the last quarter of 2022. Unusually high rates of invasive GAS infections in both children and adults were reported in 2023 and 2024 from centers in Europe and the United States. Prominent among these cases have been severe pneumonia and empyema, sometimes as a co-infection with influenza or another respiratory virus, as well as necrotizing soft tissue infections and streptococcal toxic shock. A GAS strain designated M1_{UK} first emerged in the United Kingdom in association with an upsurge in scarlet fever and has since been implicated as a dominant strain in invasive GAS infections in the United Kingdom and other European countries. Multiple other lineages of GAS have been identified in invasive infections elsewhere in Europe and around the world.

PATHOGENESIS

GAS elaborates a number of cell-surface components and extracellular products important in both the pathogenesis of infection and the human immune response. The cell wall contains a carbohydrate antigen that may be released by acid treatment. The reaction of such acid extracts with group A–specific antiserum is the basis for definitive identification of a streptococcal strain as *S. pyogenes*. Rarely, the group A antigen may be present on isolates of *S. dysgalactiae ssp. equisimilis*, which usually express the group C or G antigen (see "Streptococci of Groups C and G," below). The major surface protein of GAS is M protein, which is the basis for the serotyping of strains with specific antisera. The M protein molecules are fibrillar structures anchored in the cell wall of the organism that extend as hairlike projections away from the cell surface. The amino acid sequence of the distal or amino-terminal portion of the M protein molecule is variable, accounting for the antigenic variation of the different M types, while more proximal regions of the protein are relatively conserved. Traditional M-typing by serologic methods has been largely supplanted by use of the polymerase chain reaction to amplify the variable region of the *emm* gene, which encodes M protein. DNA sequence analysis of the amplified gene segment can be compared with an extensive database (developed at the Centers for Disease Control and Prevention [CDC]) for assignment of *emm* type. Use of *emm* typing has increased the number of identified *emm* types to more than 200. This method eliminates the need for typing sera, which are available in only a few reference laboratories. The presence of M protein on a GAS isolate correlates with its capacity to resist phagocytic killing in fresh human blood. This phenomenon appears to be due, at least in part, to the binding of plasma fibrinogen to M protein molecules on the streptococcal surface, which interferes with complement activation and deposition of opsonic complement fragments on the bacterial cell. This resistance to phagocytosis may be overcome by M protein–specific antibodies; thus, individuals with antibodies to a given M type acquired as a result of prior infection are protected against subsequent infection with organisms of the same M type but not against infection with different M types.

GAS also elaborates, to varying degrees, a polysaccharide capsule composed of hyaluronic acid. While most clinical isolates of GAS produce a hyaluronic acid capsule, strains of M type 4 or 22 lack a capsule, as do some isolates of M type 89. The fact that acapsular strains have been associated with pharyngitis and invasive infection implies that the capsule is not essential for virulence. The production of large amounts of capsule by certain strains imparts a characteristic mucoid appearance to the colonies. The capsular polysaccharide plays an important role in protecting GAS from ingestion and killing by phagocytes. In contrast to M protein, the hyaluronic acid capsule is a weak immunogen, and antibodies to hyaluronate have not been shown to be important in protective immunity. The presumed explanation is the apparent structural identity between streptococcal hyaluronic acid and the hyaluronic acid of mammalian connective tissues. The capsular polysaccharide may also play a role in GAS colonization of the pharynx by binding to CD44, a hyaluronic acid–binding protein expressed on human pharyngeal epithelial cells.

GAS produces a large number of extracellular products that may be important in local and systemic toxicity and in the spread of infection through tissues. These products include streptolysins S and O, toxins that damage cell membranes and account for the hemolysis produced by the organisms; streptokinase; DNAses; SpyCEP, a serine protease that cleaves and inactivates the chemoattractant cytokine interleukin 8, thereby inhibiting neutrophil recruitment to the site of infection; and several pyrogenic exotoxins. Previously known as erythrogenic toxins, the pyrogenic exotoxins cause the rash of scarlet fever. Since the mid-1980s, pyrogenic exotoxin–producing strains of GAS have been linked to unusually severe invasive infections, including necrotizing fasciitis and the streptococcal toxic shock syndrome (TSS). Several extracellular products stimulate specific antibody responses useful for serodiagnosis of recent streptococcal infection. Tests for antibodies to streptolysin O and DNase B are used most commonly for detection of preceding streptococcal infection in cases of suspected ARF or PSGN.

Pharyngitis Although seen in patients of all ages, GAS pharyngitis is one of the most common bacterial infections of childhood, accounting for 20–40% of all cases of exudative pharyngitis in children; it is rare among those under the age of 3. Younger children may manifest streptococcal infection with a syndrome of fever, malaise, and lymphadenopathy without exudative pharyngitis. Infection is acquired through contact with another individual carrying the organism. Respiratory droplets are the usual mechanism of spread, although other routes, including food-borne outbreaks, have been well described. The incubation period is 1–4 days. Symptoms include sore throat, fever and chills, malaise, and sometimes abdominal complaints and vomiting, particularly in children. Both symptoms and signs are variable, ranging from mild throat discomfort with minimal physical findings to high fever and severe sore throat associated with intense erythema and swelling of the pharyngeal mucosa and the presence of purulent exudate over the posterior pharyngeal wall and tonsillar pillars. Enlarged, tender anterior cervical lymph nodes commonly accompany exudative pharyngitis.

The differential diagnosis of streptococcal pharyngitis includes the many other bacterial and viral etiologies (Table 153-2). Streptococcal infection is an unlikely cause when symptoms and signs suggestive of viral infection are prominent (conjunctivitis, coryza, cough, hoarseness, or discrete ulcerative lesions of the buccal or pharyngeal mucosa). Because of the range of clinical presentations of streptococcal pharyngitis and the large number of other agents that can produce the same clinical picture, diagnosis of streptococcal pharyngitis on clinical grounds alone is not reliable. The throat culture remains the diagnostic gold standard. Culture of a throat specimen that is properly collected (i.e., by vigorous rubbing of a sterile swab over both tonsillar pillars) and properly processed is the most sensitive and specific means of definitive diagnosis. A rapid diagnostic test for latex agglutination or enzyme immunoassay of swab specimens is a useful adjunct to throat culture. Rapid diagnostic tests typically have a specificity of >95%. Thus, a positive result can be relied upon for definitive diagnosis and eliminates the need for throat culture. In settings in which the incidence of rheumatic fever is low, a confirmatory throat culture is not recommended for routine evaluation of most adults with a negative rapid test. However, because rapid diagnostic tests are less sensitive than throat culture (relative sensitivity in comparative studies, 70–90%), a negative result should be confirmed by throat culture for individuals at higher risk such as those with a history of rheumatic fever or immunocompromise or a family member with such a history; patients living in congregate settings of young adults such as dormitories or military facilities where the incidence of GAS pharyngitis may be elevated; individuals with household exposure to someone with proven GAS infection; and those living in an area in which rheumatic fever is endemic. National or professional organization guidelines in some countries with a very low incidence of ARF recommend symptomatic treatment for patients with mild symptoms and restricting diagnostic testing and/or antibiotic treatment to those with suggestive clinical features (e.g., fever, tonsillar swelling or exudate, tender anterior cervical adenopathy, and/or absence of cough) or special risk factors.

TABLE 153-2 Infectious Etiologies of Acute Pharyngitis

ORGANISM	ASSOCIATED CLINICAL SYNDROME(S)
Viruses	
Rhinovirus	Common cold
Coronavirus	Common cold, COVID-19
Adenovirus	Pharyngoconjunctival fever
Influenza virus	Influenza
Parainfluenza virus	Cold, croup
Coxsackievirus	Herpangina, hand-foot-and-mouth disease
Herpes simplex virus	Gingivostomatitis (primary infection)
Epstein-Barr virus	Infectious mononucleosis
Cytomegalovirus	Mononucleosis-like syndrome
HIV	Acute (primary) infection syndrome
Bacteria	
Group A streptococci	Pharyngitis, scarlet fever
Group C or G streptococci	Pharyngitis
Mixed anaerobes	Vincent's angina
Arcanobacterium haemolyticum	Pharyngitis, scarlatiniform rash
Neisseria gonorrhoeae	Pharyngitis
Treponema pallidum	Secondary syphilis
Francisella tularensis	Pharyngeal tularemia
Corynebacterium diphtheriae	Diphtheria
Yersinia enterocolitica	Pharyngitis, enterocolitis
Yersinia pestis	Plague
Chlamydiae	
Chlamydia pneumoniae	Bronchitis, pneumonia
Chlamydia psittaci	Psittacosis
Mycoplasmas	
Mycoplasma pneumoniae	Bronchitis, pneumonia

TREATMENT

GAS Pharyngitis

In the usual course of uncomplicated streptococcal pharyngitis, symptoms resolve after 3–5 days. The course is shortened little by treatment, which is given primarily to prevent suppurative complications and ARF. Prevention of ARF depends on eradication of the organism from the pharynx, not simply on resolution of symptoms, and requires 10 days of penicillin treatment (Table 153-3). A first-generation cephalosporin, such as cephalexin or cefadroxil, may be substituted for penicillin in cases of penicillin allergy if the nature of the allergy is not an immediate hypersensitivity reaction (anaphylaxis or urticaria) or another potentially life-threatening manifestation (e.g., severe rash and fever).

Alternative agents are erythromycin and azithromycin. Azithromycin offers the advantages of better gastrointestinal tolerability, once-daily dosing, and a 5-day treatment course. Resistance to erythromycin and other macrolides is common among isolates from several countries, including Spain, Italy, Finland, Japan, and Korea. Macrolide resistance may be becoming more prevalent

TABLE 153-3 Treatment of Group A Streptococcal Infections

INFECTION	TREATMENT[a]
Pharyngitis	Benzathine penicillin G (1.2 mU IM) *or* penicillin V (250 mg PO tid or 500 mg PO bid) × 10 days
	(Children <27 kg: Benzathine penicillin G [600,000 units IM] *or* penicillin V [250 mg PO bid or tid] × 10 days)
Impetigo	Same as pharyngitis
Erysipelas/cellulitis	Severe: Penicillin G (1–2 mU IV q4h)
	Mild to moderate: Procaine penicillin (1.2 mU IM bid)
Necrotizing fasciitis/ myositis	Surgical debridement *plus* penicillin G (2–4 mU IV q4h) *plus* clindamycin[b] (600–900 mg IV q8h)
Pneumonia/empyema	Penicillin G (2–4 mU IV q4h) *plus* drainage of empyema
Streptococcal toxic shock syndrome	Penicillin G (2–4 mU IV q4h) *plus* clindamycin[b] (600–900 mg IV q8h) *plus* IV immunoglobulin (2 g/kg as a single dose)

[a] Penicillin allergy: A first-generation cephalosporin, such as cephalexin or cefadroxil, may be substituted for penicillin in cases of penicillin allergy if the nature of the allergy is not an immediate hypersensitivity reaction (anaphylaxis or urticaria) or another potentially life-threatening manifestation (e.g., severe rash and fever). Alternative agents for oral therapy are erythromycin (10 mg/kg PO qid, up to a maximum of 250 mg per dose) and azithromycin (a 5-day course at a dose of 12 mg/kg once daily, up to a maximum of 500 mg/d). Vancomycin is an alternative for parenteral therapy. [b] Linezolid (600 mg IV q12h) may be substituted for isolates resistant to clindamycin. See text for discussion.

elsewhere with the increasing use of this class of antibiotics. Data from the CDC and the Active Bacterial Core surveillance program indicate a rise in macrolide resistance among invasive GAS isolates in the United States to >30% since 2021. In areas with resistance rates exceeding 5–10%, macrolides should be avoided unless results of susceptibility testing are known.

Follow-up culture after treatment is not routinely recommended but may be warranted in selected cases, such as those involving patients or families with frequent streptococcal infections or those occurring in situations in which the risk of ARF is thought to be high (e.g., when cases of ARF have recently been reported in the community).

Complications Suppurative complications of streptococcal pharyngitis have become uncommon with the widespread use of antibiotics for most symptomatic cases. These complications result from the spread of infection from the pharyngeal mucosa to deeper tissues by direct extension or by the hematogenous or lymphatic route and may include cervical lymphadenitis, peritonsillar or retropharyngeal abscess, sinusitis, otitis media, meningitis, bacteremia, endocarditis, and pneumonia. Local complications, such as peritonsillar or parapharyngeal abscess formation, should be considered in a patient with unusually severe or prolonged symptoms or localized pain associated with high fever and a toxic appearance. Nonsuppurative complications include ARF **(Chap. 370)** and PSGN **(Chap. 326)**, both of which are thought to result from immune responses to streptococcal infection. Penicillin treatment of streptococcal pharyngitis reduces the likelihood of ARF but not that of PSGN.

BACTERIOLOGIC TREATMENT FAILURE AND THE ASYMPTOMATIC CARRIER STATE

Surveillance cultures have shown that up to 20% of individuals in certain populations may have asymptomatic pharyngeal colonization with GAS. There are no definitive guidelines for management of these asymptomatic carriers or of asymptomatic patients who still have a positive throat culture after a full course of treatment for symptomatic pharyngitis. A reasonable course of action is to give a single 10-day course of penicillin for symptomatic pharyngitis and, if positive cultures persist, not to re-treat unless symptoms recur. Studies of the natural history of streptococcal carriage and infection have shown that the risk both of developing ARF and of transmitting infection to others is substantially lower among asymptomatic carriers than among individuals with symptomatic pharyngitis. Therefore, aggressive attempts to eradicate carriage probably are not justified under most circumstances. An exception is the situation in which an asymptomatic carrier is a potential source of infection to others. Outbreaks of food-borne infection and nosocomial puerperal infection have been traced to asymptomatic carriers who may harbor the organisms in the throat, vagina, or anus or on the skin.

TREATMENT

Asymptomatic Pharyngeal Colonization with GAS

When a carrier is transmitting infection to others, attempts to eradicate carriage are warranted. Data are limited on the best regimen to clear GAS after penicillin alone has failed. Regimens reported to have efficacy superior to that of penicillin alone for eradication of carriage include (1) a first-generation cephalosporin such as cephalexin (30 mg/kg; 500 mg maximum) twice daily for 10 days or (2) oral clindamycin (7 mg/kg; 300 mg maximum) three times daily for 10 days. A 10-day course of oral vancomycin (250 mg four times daily) and rifampin (600 mg twice daily) has eradicated rectal colonization. Single-dose azithromycin (20 mg/kg; 1000 mg maximum) has been used for mass prophylaxis/eradication of colonization in outbreak situations. In vitro susceptibility to clindamycin or azithromycin should be confirmed before prescribing either of these antibiotics for eradication of colonization.

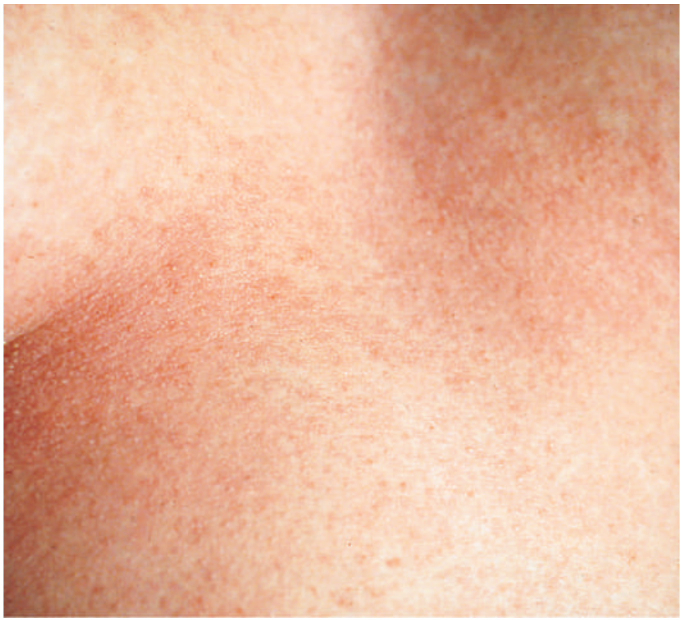

FIGURE 153-2 Scarlet fever exanthem. Finely punctate erythema has become confluent (scarlatiniform); petechiae can occur and have a linear configuration within the exanthem in body folds (Pastia's lines). *(From TB Fitzpatrick, RA Johnson, K Wolff: Color Atlas and Synopsis of Clinical Dermatology, 4th ed, New York, McGraw-Hill, 2001, with permission.)*

Scarlet Fever Scarlet fever consists of streptococcal infection, usually pharyngitis, accompanied by a characteristic rash **(Fig. 153-2)**. The rash arises from the effects of one of several toxins, currently designated *streptococcal pyrogenic exotoxins* and previously known as *erythrogenic* or *scarlet fever toxins*. In the past, scarlet fever was thought to reflect infection of an individual lacking toxin-specific immunity with a toxin-producing strain of GAS. Susceptibility to scarlet fever was correlated with results of the Dick test, in which a small amount of erythrogenic toxin injected intradermally produced local erythema in susceptible individuals but elicited no reaction in those with specific immunity. Subsequent studies have suggested that development of the scarlet fever rash may reflect a hypersensitivity reaction requiring prior exposure to the toxin. For reasons that are not clear, scarlet fever has become less common in recent years, although large outbreaks have occurred recently in China and the United Kingdom. The symptoms of scarlet fever are the same as those of pharyngitis alone. The rash typically begins on the first or second day of illness over the upper trunk, spreading to involve the extremities but sparing the palms and soles. The rash is made up of minute papules, giving a characteristic "sandpaper" feel to the skin. Associated findings include circumoral pallor, "strawberry tongue" (enlarged papillae on a coated tongue, which later may become denuded), and accentuation of the rash in skin folds (*Pastia's lines*). Subsidence of the rash in 6–9 days is followed after several days by desquamation of the palms and soles. The differential diagnosis of scarlet fever includes other causes of fever and generalized rash, such as measles and other viral exanthems, Kawasaki disease, toxic shock syndrome, and systemic allergic reactions (e.g., drug eruptions).

Skin and Soft Tissue Infections GAS—and occasionally other streptococcal species—can cause a variety of infections involving the skin, subcutaneous tissues, muscles, and fascia. While several clinical syndromes offer a useful means for classification of these infections, not all cases fit exactly into one category. The classic syndromes are general guides to predicting the level of tissue involvement in a particular patient, the probable clinical course, and the likelihood that surgical intervention or aggressive life support will be required.

IMPETIGO (PYODERMA)

Impetigo, a superficial infection of the skin, is caused primarily by GAS and occasionally by other streptococci or *Staphylococcus aureus*. Impetigo is seen most often in young children, tends to occur during

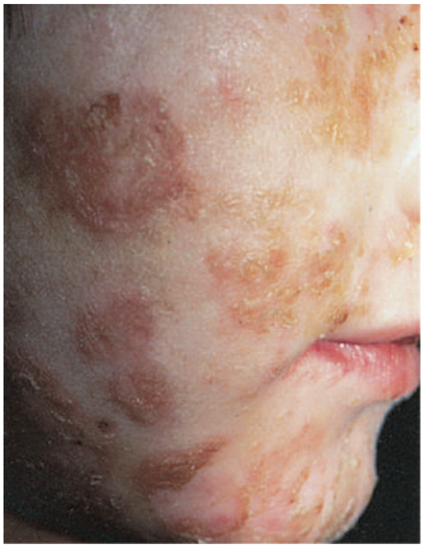

FIGURE 153-3 Impetigo is a superficial streptococcal or *Staphylococcus aureus* infection consisting of honey-colored crusts and erythematous weeping erosions. Occasionally, bullous lesions may be seen. *(Courtesy of Mary Spraker, MD; with permission.)*

warmer months, and is more common in semitropical or tropical climates than in cooler regions. Infection is more common among children living under conditions of poor hygiene. Prospective studies have shown that colonization of unbroken skin with GAS precedes clinical infection. Minor trauma, such as a scratch or an insect bite, may then serve to inoculate organisms into the skin. Impetigo is best prevented, therefore, by attention to adequate hygiene. The usual sites of involvement are the face (particularly around the nose and mouth) and the legs, although lesions may occur at other locations. Individual lesions begin as red papules, which evolve quickly into vesicular and then pustular lesions that break down and coalesce to form characteristic honeycomb-like crusts **(Fig. 153-3)**. Lesions generally are not painful, and patients do not appear ill. Fever is not a feature of impetigo and, if present, suggests either infection extending to deeper tissues or another diagnosis. The classic presentation of impetigo usually poses little diagnostic difficulty. Cultures of impetiginous lesions often yield *S. aureus* as well as GAS. In almost all cases, streptococci are isolated initially, and staphylococci appear later, presumably as secondary colonizing flora. In the past, penicillin was nearly always effective against these infections. However, an increasing frequency of penicillin treatment failure suggests that *S. aureus* may have become more prominent as a cause of impetigo. *Bullous impetigo* due to *S. aureus* is distinguished from typical streptococcal infection by more extensive, bullous lesions that break down and leave thin paper-like crusts instead of the thick amber crusts of streptococcal impetigo. Other skin lesions that may be confused with impetigo include herpetic lesions—either those of orolabial herpes simplex or those of chickenpox or zoster. Herpetic lesions can generally be distinguished by their appearance as more discrete, grouped vesicles and by a positive Tzanck test or by herpes simplex virus- or varicella-zoster virus-specific PCR. In difficult cases, cultures of vesicular fluid should yield GAS (or *Staphylococcus aureus*) in impetigo and the responsible virus in herpesvirus infections.

TREATMENT

Streptococcal Impetigo

Treatment of streptococcal impetigo is the same as that for streptococcal pharyngitis. In view of evidence that *S. aureus* has become a relatively frequent cause of impetigo, empirical regimens should cover both streptococci and *S. aureus*. For example, either dicloxacillin or cephalexin can be given at a dose of 250 mg four times daily for 10 days. Topical mupirocin ointment also is effective. Culture may be indicated to rule out methicillin-resistant *S. aureus*, especially if the response to empirical treatment is unsatisfactory. In most areas of the world, ARF is not a sequela to streptococcal skin infections, although PSGN may follow either skin or throat infection. The reason for this difference is not known. One hypothesis is that the immune response necessary for development of ARF occurs only after infection of the pharyngeal mucosa. In addition, the strains of GAS that cause pharyngitis are generally of different M protein types than those associated with skin infections; thus, the strains that cause pharyngitis may have rheumatogenic potential, while the skin-infecting strains may not. An exception to this general rule may occur among indigenous people in northern Australia and in certain Pacific Island groups. Acute rheumatic fever and rheumatic heart disease are prevalent in these populations as is streptococcal impetigo/pyoderma, but not pharyngitis. This epidemiologic pattern has led investigators to suggest that skin infection may trigger acute rheumatic fever in this setting.

CELLULITIS

Inoculation of organisms into the skin may lead to *cellulitis*: infection involving the skin and subcutaneous tissues. The portal of entry may be a traumatic or surgical wound, an insect bite, or any other break in skin integrity. Often, no entry site is apparent. One form of streptococcal cellulitis, *erysipelas*, is characterized by a bright red appearance of the involved skin, which forms a plateau sharply demarcated from surrounding normal skin **(Fig. 153-4)**. The lesion is warm to the touch, may be tender, and appears shiny and swollen. The skin often has a *peau d'orange* texture, which is thought to reflect involvement of superficial lymphatics; superficial blebs or bullae may form, usually 2–3 days after onset. The lesion typically develops over a few hours and is associated with fever and chills. Erysipelas tends to occur on the malar area of the face (often with extension over the bridge of the nose to the contralateral malar region) or on the lower extremities. After one episode, recurrence at the same site—sometimes years later—is not uncommon. Classic cases of erysipelas, with typical features, are almost always due to β-hemolytic streptococci, usually GAS and occasionally group C or G. Often, however, the appearance of streptococcal cellulitis is not sufficiently distinctive to permit a specific diagnosis on clinical grounds. The anatomic area involved may not be typical for erysipelas, the lesion may be less intensely red than usual and may fade into surrounding skin, and/or the patient may appear only mildly ill. In such cases, it is prudent to broaden the spectrum of empirical antimicrobial therapy to include other pathogens, particularly *S. aureus*, that can produce cellulitis with the same

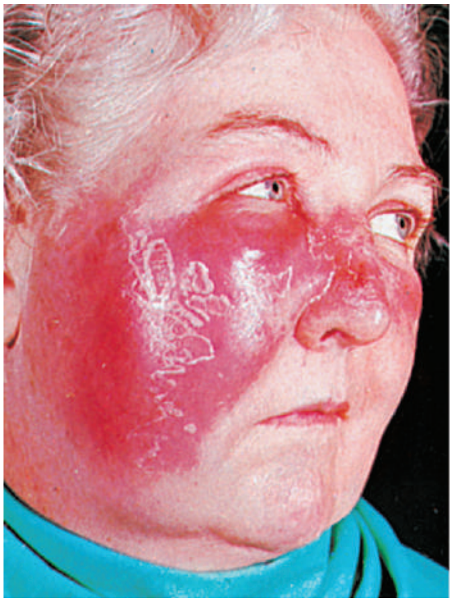

FIGURE 153-4 Erysipelas is a streptococcal infection of the superficial dermis and consists of well-demarcated, erythematous, edematous, warm plaques.

appearance. Staphylococcal infection should be suspected if cellulitis develops around a wound or an ulcer.

Streptococcal cellulitis tends to develop at anatomic sites in which normal lymphatic drainage has been disrupted, such as sites of prior cellulitis, the arm ipsilateral to a mastectomy and axillary lymph node dissection, a lower extremity previously involved in deep venous thrombosis or chronic lymphedema, or the leg from which a saphenous vein has been harvested for coronary artery bypass grafting. The organism may enter via a dermal breach some distance from the eventual site of clinical cellulitis. For example, some patients with recurrent leg cellulitis following saphenous vein removal may stop having recurrent episodes only after treatment of tinea pedis on the affected extremity. Fissures in the skin presumably serve as a portal of entry for streptococci, which then produce infection more proximally in the leg at the site of previous injury. Streptococcal cellulitis may also involve recent surgical wounds. GAS is among the few bacterial pathogens that typically produce signs of wound infection and surrounding cellulitis within the first 24 h after surgery. These wound infections are usually associated with a thin exudate and may spread rapidly, either as cellulitis in the skin and subcutaneous tissue or as a deeper tissue infection (see below). Streptococcal wound infection or localized cellulitis may also be associated with *lymphangitis*, manifested by red streaks extending proximally along superficial lymphatics from the infection site.

TREATMENT

Streptococcal Cellulitis

See Table 153-3 and **Chap. 134.**

DEEP SOFT TISSUE INFECTIONS

Necrotizing fasciitis (*hemolytic streptococcal gangrene*) involves the superficial and/or deep fascia investing the muscles of an extremity or the trunk. The source of the infection is either the skin, with organisms introduced into tissue through trauma (sometimes trivial), or the bowel flora, with organisms released during abdominal surgery or from an occult enteric source, such as a diverticular or appendiceal abscess. The inoculation site may be inapparent and is often some distance from the site of clinical involvement; e.g., the introduction of organisms via minor trauma to the hand may be associated with clinical infection of the tissues overlying the shoulder or chest. Cases associated with the bowel flora are usually polymicrobial, involving a mixture of anaerobic bacteria (such as *Bacteroides fragilis* or anaerobic streptococci) and facultative organisms (usually gram-negative bacilli). Cases unrelated to contamination from bowel organisms are most commonly caused by GAS alone or in combination with other organisms (most often *S. aureus*). Overall, GAS is implicated in ~60% of cases of necrotizing fasciitis. The onset of symptoms is usually quite acute and is marked by severe pain at the site of involvement, malaise, fever, chills, and a toxic appearance. The physical findings, particularly early on, may not be striking, with only minimal erythema of the overlying skin. Pain and tenderness are usually severe. In contrast, in more superficial cellulitis, the skin appearance is more abnormal, but pain and tenderness are only mild or moderate. As the infection progresses (often over several hours), the severity and extent of symptoms worsen, and skin changes become more evident, with the appearance of dusky or mottled erythema and edema. The marked tenderness of the involved area may evolve into anesthesia as the spreading inflammatory process produces infarction of cutaneous nerves.

Although myositis is more commonly due to *S. aureus* infection, GAS occasionally produces abscesses in skeletal muscles (*streptococcal myositis*), with little or no involvement of the surrounding fascia or overlying skin. The presentation is usually subacute, but a fulminant form has been described in association with severe systemic toxicity, bacteremia, and a high mortality rate. The fulminant form may reflect the same basic disease process seen in necrotizing fasciitis, but with the necrotizing inflammatory process extending into the muscles themselves rather than remaining limited to the fascial layers.

TREATMENT

Deep Soft Tissue Streptococcal Infections

Once necrotizing fasciitis is suspected, early surgical exploration is both diagnostically and therapeutically indicated. Surgery reveals necrosis and inflammatory fluid tracking along the fascial planes above and between muscle groups, without involvement of the muscles themselves. The process usually extends beyond the area of clinical involvement, and extensive debridement is required. Drainage and debridement are central to the management of necrotizing fasciitis; antibiotic treatment is a critical adjunct (Table 153-3), but surgery is life-saving. Treatment for streptococcal myositis consists of surgical drainage—usually by an open procedure that permits evaluation of the extent of infection and ensures adequate debridement of involved tissues—and high-dose penicillin (Table 153-3). Infection models in animals and multiple retrospective clinical studies support the addition of clindamycin to the antibiotic treatment regimen for severe GAS necrotizing soft tissue infections. In vitro and animal models suggest that the growth of GAS slows after the organism reaches a high density in infected tissues and that penicillin and other beta-lactam antibiotics have reduced efficacy under these circumstances as they target cell wall biosynthesis. By contrast, protein synthesis inhibitors retain activity regardless of growth phase of the bacteria. If the infecting strain is not known to be susceptible to clindamycin, linezolid is an alternative.

Pneumonia and Empyema GAS is an occasional cause of pneumonia, generally in previously healthy individuals. The onset of symptoms may be abrupt or gradual. Pleuritic chest pain, fever, chills, and dyspnea are the characteristic manifestations. Cough is usually present but may not be prominent. Approximately one-half of patients with GAS pneumonia have an accompanying pleural effusion. In contrast to the sterile parapneumonic effusions typical of pneumococcal pneumonia, those complicating streptococcal pneumonia are almost always infected. The empyema fluid is usually visible by chest radiography on initial presentation, and its volume may increase rapidly. These pleural collections should be drained early, as they tend to become loculated rapidly, resulting in a chronic fibrotic reaction that may require thoracotomy for removal.

Bacteremia, Puerperal Sepsis, and Streptococcal Toxic Shock Syndrome In adults, GAS bacteremia is usually associated with an identifiable local infection, whereas children may have bacteremia without an associated focal infection. Bacteremia occurs rarely with otherwise uncomplicated pharyngitis, occasionally with cellulitis or pneumonia, and relatively frequently with necrotizing fasciitis. Bacteremia without an identified source raises the possibility of endocarditis, an occult abscess, or osteomyelitis. A variety of focal infections may arise secondarily from streptococcal bacteremia, including endocarditis, meningitis, septic arthritis, osteomyelitis, peritonitis, and visceral abscesses. GAS is occasionally implicated in infectious complications of childbirth, usually endometritis and associated bacteremia. In the preantibiotic era, puerperal sepsis was commonly caused by GAS; currently, it is more often caused by GBS. Several nosocomial outbreaks of puerperal GAS infection have been traced to an asymptomatic carrier, usually someone present at delivery. The site of carriage may be the skin, throat, anus, or vagina.

Beginning in the late 1980s, several reports described patients with GAS infections associated with shock and multisystem organ failure. This syndrome was called *streptococcal toxic shock syndrome* (*TSS*) because it shares certain features with staphylococcal TSS. A case definition for streptococcal TSS was formulated in 1993 and updated in 2010 **(Table 153-4)**. The general features of the illness include fever, hypotension, renal impairment, and respiratory distress syndrome. Various types of rash have been described, but rash usually does not develop. Laboratory abnormalities include a marked shift to the left in the white blood cell differential, with many immature granulocytes; hypocalcemia; hypoalbuminemia; and thrombocytopenia, which usually becomes more pronounced on the second or third day of illness.

TABLE 153-4 Case Definition for Streptococcal Toxic Shock Syndrome[a]

- I. Isolation of group A streptococci (*Streptococcus pyogenes*)
  - A. From a normally sterile site
  - B. From a nonsterile site
- II. Clinical signs of severity
  - A. Hypotension and
  - B. ≥2 of the following signs
    - 1. Renal impairment
    - 2. Coagulopathy
    - 3. Liver function impairment
    - 4. Adult respiratory distress syndrome
    - 5. A generalized erythematous macular rash that may desquamate
    - 6. Soft tissue necrosis, including necrotizing fasciitis or myositis; *or* gangrene

[a] An illness fulfilling criteria IA, IIA, and IIB is defined as a *definite* case. An illness fulfilling criteria IB, IIA, and IIB is defined as a *probable* case if no other etiology for the illness is identified.

Source: Modified from Streptococcal Toxic Shock Syndrome (STSS) (*Streptococcus pyogenes*) 2010 Case Definition: Centers for Disease Control and Prevention (*https://ndc.services.cdc.gov/case-definitions/streptococcal-toxic-shock-syndrome-2010/*).

In contrast to patients with staphylococcal TSS, the majority with streptococcal TSS are bacteremic. The most common associated infection is a soft tissue infection—necrotizing fasciitis, myositis, or cellulitis—although a variety of other associated local infections have been described, including pneumonia, peritonitis, osteomyelitis, and myometritis. Streptococcal TSS is associated with a mortality rate of ≥30%, with most deaths secondary to shock and respiratory failure. Because of its rapidly progressive and lethal course, early recognition of the syndrome is essential. Patients should receive aggressive supportive care (fluid resuscitation, pressors, and mechanical ventilation) in addition to antimicrobial therapy and, in cases associated with necrotizing fasciitis, should undergo surgical debridement. Exactly why certain patients develop this fulminant syndrome is not known. Early studies of the streptococcal strains isolated from these patients demonstrated a strong association with the production of pyrogenic exotoxin A. This association has been inconsistent in subsequent case series. Pyrogenic exotoxin A and several other streptococcal exotoxins act as superantigens to trigger release of inflammatory cytokines from T lymphocytes. Fever, shock, and organ dysfunction in streptococcal TSS may reflect, in part, the systemic effects of superantigen-mediated cytokine release.

TREATMENT

Streptococcal Toxic Shock Syndrome

High-dose penicillin is the agent of choice for streptococcal TSS. For empiric treatment of sepsis before the etiology is known, a broad-spectrum antibiotic regimen should include an agent with excellent activity against GAS such as an extended-spectrum penicillin, cephalosporin, or a carbapenem. In light of the possible role of pyrogenic exotoxins or other streptococcal toxins in streptococcal TSS, adjunctive treatment with clindamycin has been advocated by some authorities (Table 153-3), who argue that, through its direct action on protein synthesis, clindamycin is more effective in rapidly terminating toxin production than is penicillin—a cell-wall agent. As discussed previously (see "Treatment of Deep Soft Tissue Streptococcal Infections," above), support for this view comes from studies of an experimental model of streptococcal myositis, in which mice given clindamycin had a higher rate of survival than those given penicillin. Comparable data on the treatment of human infections are not available, although retrospective analysis has suggested a better outcome when patients with invasive soft tissue infection and streptococcal TSS are treated with clindamycin rather than with cell wall–active antibiotics alone. As clindamycin resistance occurs in 30% or more of invasive GAS isolates in the United States and many other countries, susceptibility to clindamycin should not be assumed in the absence of in vitro testing of the infecting strain. Linezolid is an appropriate alternative in the setting of clindamycin resistance or unknown susceptibility. IV immunoglobulin also has been used as adjunctive therapy for streptococcal TSS (Table 153-3). Pooled immunoglobulin preparations contain antibodies capable of neutralizing the effects of streptococcal toxins and may have opsonic activity against GAS. Intravenous immunoglobulin (IVIG) also has immunomodulatory properties, which may mitigate the pathophysiologic effects of inflammatory cytokines. Anecdotal reports and case series have suggested favorable clinical responses to IVIG, but no adequately powered, prospective, controlled trials have been reported. A meta-analysis of five studies of streptococcal TSS patients treated with clindamycin found that IVIG use was associated with a reduction in mortality rate from 33.7% to 15.7%.

PREVENTION

No vaccine against GAS is commercially available. A formulation that consists of recombinant peptides containing epitopes of 26 M-protein types has undergone phase 1 and 2 testing in volunteers. Early results indicate that the vaccine is well tolerated and elicits type-specific antibody responses. Vaccines based on a conserved region of M protein or on a mixture of other conserved GAS protein antigens are in earlier stages of development.

Household contacts of individuals with invasive GAS infection (e.g., bacteremia, necrotizing fasciitis, or streptococcal TSS) are at greater risk of invasive infection than the general population. Asymptomatic pharyngeal colonization with GAS has been detected in up to 25% of persons with >4 h/d of same-room exposure to an index case. However, the CDC does not recommend antibiotic prophylaxis routinely for contacts of patients with invasive disease because such an approach (if effective) would require treatment of hundreds of contacts to prevent a single case. Prophylaxis may be considered for contacts of unusually severe cases or for individuals at increased risk for invasive infection.

STREPTOCOCCI OF GROUPS C AND G

Group C and group G streptococci are β-hemolytic bacteria that occasionally cause human infections similar to those caused by GAS. Strains that form small colonies on blood agar (<0.5 mm) are generally members of the *Streptococcus milleri* group (*Streptococcus intermedius*, *Streptococcus anginosus*; see "Viridans Streptococci," below). Large-colony group C and G streptococci of human origin are now considered a single species, *Streptococcus dysgalactiae* subspecies *equisimilis* (SDSE). Genomic studies have demonstrated extensive overlap between the genomes of SDSE and GAS and evidence of cross-species recombination. These organisms have been associated with pharyngitis, cellulitis and soft tissue infections, pneumonia, bacteremia, endocarditis, and septic arthritis. Puerperal sepsis, meningitis, epidural abscess, intraabdominal abscess, urinary tract infection, and neonatal sepsis also have been reported. SDSE bacteremia most often affects elderly or chronically ill patients and, in the absence of obvious local infection, is likely to reflect endocarditis. Septic arthritis, sometimes involving multiple joints, may complicate endocarditis or develop in its absence. Distinct streptococcal species of Lancefield group C cause infections in domesticated animals, especially horses and cattle; some human infections are acquired through contact with animals or consumption of unpasteurized milk. These zoonotic organisms include *Streptococcus equi* subspecies *zooepidemicus* and *S. equi* subspecies *equi*.

TREATMENT

Group C or G Streptococcal Infection

Penicillin is the drug of choice for treatment of group C or G streptococcal infections. Antibiotic treatment is the same as for similar syndromes due to GAS (Table 153-3). Patients with bacteremia or septic arthritis should receive IV penicillin (2–4 mU every 4 h). All group C and G streptococci are sensitive to penicillin; nearly all are inhibited in vitro by concentrations of ≤0.03 μg/mL. Occasional isolates exhibit tolerance: although inhibited by low concentrations of penicillin, they are killed only by significantly higher concentrations. The clinical significance of tolerance is unknown. Because of the poor clinical response of some patients to penicillin alone, the addition of gentamicin (1 mg/kg every 8 h for patients with normal renal function) is recommended by some authorities for treatment of endocarditis or septic arthritis due to group C or G streptococci; however, combination therapy has not been shown to be superior to penicillin treatment alone. Patients with joint infections often require repeated aspiration or open drainage and debridement for cure; the response to treatment may be slow, particularly in debilitated patients and those with involvement of multiple joints. Infection of prosthetic joints almost always requires prosthesis removal in addition to antibiotic therapy.

GROUP B STREPTOCOCCI

Identified first as a cause of mastitis in cows, streptococci belonging to Lancefield group B have since been recognized as a major cause of sepsis and meningitis in human neonates. GBS is also a frequent cause of peripartum fever in women and an occasional cause of serious infection in nonpregnant adults. Since the widespread institution of prenatal screening for GBS in the 1990s, the incidence of neonatal infection per 1000 live births has fallen from ~2–3 cases to ~0.6 case. During the same period, GBS infection in adults with underlying chronic illnesses has become more common; adults now account for a larger proportion of invasive GBS infections than do newborns. Lancefield group B consists of a single species, *S. agalactiae*, which is definitively identified with specific antiserum to the group B cell wall–associated carbohydrate antigen. A streptococcal isolate can be classified presumptively as GBS on the basis of biochemical tests, including hydrolysis of sodium hippurate (in which 99% of isolates are positive), hydrolysis of bile esculin (in which 99–100% are negative), bacitracin susceptibility (in which 92% are resistant), and production of CAMP factor (in which 98–100% are positive). CAMP factor is a phospholipase produced by GBS that causes synergistic hemolysis with β lysin produced by certain strains of *S. aureus*. Its presence can be demonstrated by cross-streaking of the test isolate and an appropriate staphylococcal strain on a blood agar plate. GBS organisms causing human infections are encapsulated by 1 of 10 antigenically distinct polysaccharides. The capsular polysaccharide is an important virulence factor. Antibodies to the capsular polysaccharide afford protection against GBS of the same (but not of a different) capsular type.

■ INFECTION IN NEONATES

Two general types of GBS infection in infants are defined by the age of the patient at presentation. *Early-onset infections* occur within the first week of life, with a median age of 20 h at onset. Approximately half of these infants have signs of GBS disease at birth. The infection is acquired during or shortly before birth from the colonized maternal genital tract. Surveillance studies have shown that 5–40% of women are vaginal or rectal carriers of GBS. Approximately 50% of infants delivered vaginally by carrier mothers become colonized, although only 1–2% develop clinically evident infection. Prematurity, prolonged labor, obstetric complications, and maternal fever are risk factors for early-onset infection. The presentation of early-onset infection is the same as that of other forms of neonatal sepsis. Typical findings include respiratory distress, lethargy, and hypotension. Essentially all infants with early-onset disease are bacteremic, one-third to one-half have pneumonia and/or respiratory distress syndrome, and one-third have meningitis.

Late-onset infections occur in infants 1 week to 3 months old and, in rare instances, in older infants (mean age at onset, 3–4 weeks). The infecting organism may be acquired during delivery (as in early-onset cases) or during later contact with a colonized mother, nursery personnel, or another source. Meningitis is the most common manifestation of late-onset infection and in most cases is associated with a strain of capsular type III. Infants present with fever, lethargy or irritability, poor feeding, and seizures. The various other types of late-onset infection include bacteremia without an identified source, osteomyelitis, septic arthritis, and facial cellulitis associated with submandibular or preauricular adenitis.

TREATMENT

Group B Streptococcal Infection in Neonates

Penicillin is the agent of choice for all GBS infections. Empirical broad-spectrum therapy for suspected bacterial sepsis, consisting of ampicillin and gentamicin, is generally administered until culture results become available. If cultures yield GBS, many pediatricians continue to administer gentamicin, along with ampicillin or penicillin, for a few days until clinical improvement becomes evident. Infants with bacteremia or soft tissue infection should receive penicillin at a dosage of 200,000 units/kg per day in divided doses. For meningitis, infants ≤7 days of age should receive 250,000–450,000 units/kg per d in three divided doses; infants >7 days of age should receive 450,000–500,000 units/kg per day in four divided doses. Meningitis should be treated for at least 14 days because of the risk of relapse with shorter courses.

Prevention The incidence of GBS infection is unusually high among infants of women with risk factors: preterm delivery, early rupture of membranes (>24 h before delivery), prolonged labor, fever, or chorioamnionitis. Because the usual source of the organisms infecting a neonate is the mother's birth canal, efforts have been made to prevent GBS infections by the identification of high-risk carrier mothers and their treatment with various forms of antibiotic prophylaxis or immunoprophylaxis. Prophylactic administration of ampicillin or penicillin to such patients during delivery reduces the risk of infection in the newborn. This approach has been hampered by logistical difficulties in identifying colonized women before delivery; the results of vaginal cultures early in pregnancy are poor predictors of carrier status at delivery. The CDC recommends screening for anogenital colonization at 35–37 weeks of pregnancy by a swab culture of the lower vagina and anorectum; intrapartum chemoprophylaxis is recommended for culture-positive women and for women who, regardless of culture status, have previously given birth to an infant with GBS infection or have a history of GBS bacteriuria during pregnancy. Women whose culture status is unknown and who develop premature labor (<37 weeks), prolonged rupture of membranes (>18 h), or intrapartum fever or who have a positive intrapartum nucleic acid amplification test for GBS also should receive intrapartum chemoprophylaxis. The recommended regimen for chemoprophylaxis is a loading dose of 5 million units of penicillin G followed by 2.5 million units every 4 h until delivery. Cefazolin is an alternative for women with a history of penicillin allergy who are thought not to be at high risk for anaphylaxis. For women with a history of immediate hypersensitivity, clindamycin may be substituted, but only if the colonizing isolate has been demonstrated to be susceptible. If susceptibility testing results are not available or indicate resistance, vancomycin should be used in this situation.

Treatment of all pregnant women who are colonized or have risk factors for neonatal infection will result in exposure of up to one-third of pregnant women and newborns to antibiotics, with the attendant risks of allergic reactions and selection for resistant organisms. Although still in the developmental stages, a GBS vaccine may ultimately offer a better solution to prevention. Because transplacental passage of maternal antibodies produces protective antibody levels in newborns, efforts are underway to develop a vaccine against GBS that can be given to childbearing-age women before or during pregnancy. Results of phase 1 clinical trials of GBS capsular polysaccharide–protein conjugate vaccines suggest that a multivalent conjugate vaccine would be safe and highly immunogenic.

■ INFECTION IN ADULTS

The majority of GBS infections in otherwise healthy adults are related to pregnancy and parturition. Peripartum fever, the most common manifestation, is sometimes accompanied by symptoms and signs of endometritis or chorioamnionitis (abdominal distention and uterine or adnexal tenderness). Blood and vaginal swab cultures are often positive. Bacteremia is usually transitory but occasionally results in meningitis or endocarditis. Infections in adults that are not associated with the peripartum period generally involve individuals who are elderly or have an underlying chronic illness, such as diabetes mellitus or a malignancy. Among the infections that develop with some frequency in adults are cellulitis and soft tissue infection (including infected diabetic skin ulcers), urinary tract infection, pneumonia, endocarditis, and septic arthritis. Other reported infections include meningitis, osteomyelitis, and intraabdominal or pelvic abscesses. Relapse or recurrence of invasive infection weeks to months after a first episode is documented in ~4% of cases.

TREATMENT

Group B Streptococcal Infection in Adults

GBS is less sensitive to penicillin than GAS, requiring somewhat higher doses. Adults with serious localized infections (pneumonia, pyelonephritis, abscess) should receive divided doses of ~12 million units of penicillin G daily; patients with endocarditis or meningitis should receive 18–24 million units per day in divided doses. Vancomycin is an acceptable alternative for penicillin-allergic patients.

NONENTEROCOCCAL GROUP D STREPTOCOCCI

The main nonenterococcal group D streptococci that cause human infections were previously considered a single species, *Streptococcus bovis*. The organisms encompassed by *S. bovis* have been reclassified into two species, each of which has two subspecies: *Streptococcus gallolyticus* subspecies *gallolyticus*, *S. gallolyticus* subspecies *pasteurianus*, *Streptococcus infantarius* subspecies *infantarius*, and *S. infantarius* subspecies *coli*. Endocarditis caused by these organisms is often associated with neoplasms of the gastrointestinal tract—most frequently, a colon carcinoma or polyp—but is also reported in association with other bowel lesions. When occult gastrointestinal lesions are carefully sought, abnormalities are found in >60% of patients with endocarditis due to *S. gallolyticus* or *S. infantarius*. In contrast to the enterococci, nonenterococcal group D streptococci like these organisms are reliably killed by penicillin as a single agent, and penicillin is the agent of choice for the infections they cause.

VIRIDANS AND OTHER STREPTOCOCCI

■ VIRIDANS STREPTOCOCCI

Consisting of multiple species of α-hemolytic streptococci, the viridans streptococci are a heterogeneous group of organisms that are important agents of bacterial endocarditis **(Chap. 133)**. Several species of viridans streptococci, including *Streptococcus salivarius*, *Streptococcus mitis*, *Streptococcus sanguis*, and *Streptococcus mutans*, are part of the normal flora of the mouth, where they live in close association with the teeth and gingiva. Some species contribute to the development of dental caries.

Previously known as *Streptococcus morbillorum*, *Gemella morbillorum* has been placed in a separate genus, along with *Gemella haemolysans*, on the basis of genetic-relatedness studies. These species resemble viridans streptococci with respect to habitat in the human host and associated infections.

The transient viridans streptococcal bacteremia induced by eating, toothbrushing, flossing, and other sources of minor trauma, together with adherence to biologic surfaces, is thought to account for the predilection of these organisms to cause endocarditis **(see Fig. 133-1)**. Viridans streptococci are also isolated, often as part of a mixed flora, from sites of sinusitis, brain abscess, and liver abscess.

Viridans streptococcal bacteremia occurs relatively frequently in neutropenic patients, particularly after bone marrow transplantation or high-dose chemotherapy for cancer. Some of these patients develop a sepsis syndrome with high fever and shock. Risk factors for viridans streptococcal bacteremia include chemotherapy with high-dose cytosine arabinoside, prior treatment with trimethoprim-sulfamethoxazole or a fluoroquinolone, treatment with antacids or histamine antagonists, mucositis, and profound neutropenia.

The *S. milleri* group (also referred to as the *S. intermedius* or *S. anginosus* group) includes three species that cause human disease: *S. intermedius*, *S. anginosus*, and *Streptococcus constellatus*. These organisms are often considered viridans streptococci, although they differ somewhat from other viridans streptococci in their small colony size on solid medium (they typically form colonies <0.5 mm in diameter), their hemolytic pattern (they may be α-, β-, or nonhemolytic), and the disease syndromes they cause. This group commonly produces suppurative infections, particularly abscesses of brain and abdominal viscera, and infections related to the oral cavity or respiratory tract, such as peritonsillar abscess, sinusitis and complications such as orbital cellulitis, subdural and epidural abscess, and cerebral venous sinus thrombosis, lung abscess, and empyema.

TREATMENT

Infection with Viridans Streptococci

Isolates from neutropenic patients with bacteremia are often resistant to penicillin; thus, these patients should be treated presumptively with vancomycin until the results of susceptibility testing become available. Viridans streptococci isolated in other clinical settings usually are sensitive to penicillin. Susceptibility testing should be performed to guide treatment of serious infections.

■ *ABIOTROPHIA* AND *GRANULICATELLA* SPECIES (NUTRITIONALLY VARIANT STREPTOCOCCI)

Occasional isolates cultured from the blood of patients with endocarditis fail to grow when subcultured on solid media. These *nutritionally variant streptococci* require supplemental thiol compounds or active forms of vitamin B_6 (pyridoxal or pyridoxamine) for growth in the laboratory. The nutritionally variant streptococci are generally grouped with the viridans streptococci because they cause similar types of infections. However, they have been reclassified on the basis of 16S ribosomal RNA sequence comparisons into two separate genera: *Abiotrophia*, with a single species (*Abiotrophia defectiva*), and *Granulicatella*, with three species associated with human infection (*Granulicatella adiacens*, *Granulicatella para-adiacens*, and *Granulicatella elegans*).

TREATMENT

Infection with Nutritionally Variant Streptococci

Treatment failure and relapse appear to be more common in cases of endocarditis due to nutritionally variant streptococci than in those due to the usual viridans streptococci. Thus, the addition of gentamicin (1 mg/kg every 8 h for patients with normal renal function) to the penicillin regimen is recommended for endocarditis due to the nutritionally variant organisms.

■ OTHER STREPTOCOCCI

Streptococcus suis is an important pathogen in swine and has been reported to cause meningitis in humans, usually in individuals with occupational exposure to pigs. *S. suis* has been reported to be the most common cause of bacterial meningitis in Vietnam, and it has been responsible for outbreaks in China. Strains of *S. suis* associated with human infections have generally reacted with Lancefield group R typing serum and sometimes with group D typing serum as well. Isolates may be α- or β-hemolytic and are sensitive to penicillin. *Streptococcus iniae*, a pathogen of fish, has been associated with infections in humans who have handled live or freshly killed fish. Cellulitis of the hand is the most common form of human infection, although bacteremia and endocarditis have been reported. *Anaerobic streptococci*, or *peptostreptococci*, are part of the normal flora of the oral cavity, bowel, and vagina. **Infections caused by the anaerobic streptococci are discussed in Chap. 182.**

FURTHER READING

Bruckner L, Gigliotti F: Viridans group streptococcal infections among children with cancer and the importance of emerging antibiotic resistance. Semin Pediatr Infect Dis 17:153, 2006.

Parks T et al: Polyspecific intravenous immunoglobulin in clindamycin-treated patients with streptococcal toxic shock syndrome: A systematic review and meta-analysis. Clin Infect Dis 67:1434, 2018.

Raabe V, Shane A: Group B *Streptococcus* (*Streptococcus agalactiae*), in *Gram-Positive Pathogens*, 3rd ed, Fischetti V et al (eds). Washington, DC, ASM Press, 2019, pp 228–238.

Shulman ST et al: Clinical practice guideline for the diagnosis and management of group A streptococcal pharyngitis: 2012 update by the Infectious Diseases Society of America. Clin Infect Dis 55:1279, 2012.

Stevens DL, Bryant AE: Necrotizing soft tissue infections. N Engl J Med 377:2253, 2017.

Xie O et al: Inter-species gene flow drives ongoing evolution of *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subsp. *equisimilis*. Nat Commun 15:2286, 2024.

154

Enterococcal Infections

William R. Miller, Cesar A. Arias,
Barbara E. Murray*

Enterococci have been recognized as potential human pathogens for well over a century, but only in recent years have these organisms acquired prominence as important causes of nosocomial infections. The ability of enterococci to survive and/or disseminate in the hospital environment and to acquire antibiotic resistance determinants makes the treatment of some enterococcal infections in critically ill patients a difficult challenge. Enterococci were first mentioned in the French literature in 1899; the "entérocoque" was found in the human gastrointestinal tract. The first pathologic description of an enterococcal infection dates to the same year. A clinical isolate from a patient who died as a consequence of endocarditis was initially designated *Micrococcus zymogenes*, was later named *Streptococcus faecalis* subspecies *zymogenes*, and would now be classified as *Enterococcus faecalis*. The ability of this isolate to cause severe disease in both rabbits and mice illustrated its potential lethality in the appropriate settings.

MICROBIOLOGY AND TAXONOMY

Enterococci are gram-positive organisms. In clinical specimens, they are usually observed as single cells, diplococci, or short chains **(Fig. 154-1)**, although long chains are noted with some strains. Enterococci were originally classified as streptococci because organisms of the two genera share many morphologic and phenotypic characteristics, including a generally negative catalase reaction. Only DNA hybridization studies and later 16S rRNA sequencing clearly demonstrated that enterococci should be grouped as a genus distinct from the streptococci. Unlike the majority of streptococci, enterococci hydrolyze esculin in the presence of 40% bile salts and grow at high salt concentrations (e.g., 6.5%) and at high temperatures (46°C). Enterococci are usually reported by the clinical laboratory to be nonhemolytic on the basis of their inability to lyse the ovine or bovine red blood cells (RBCs) commonly used in agar plates; however, some strains of *E. faecalis* do lyse RBCs from humans, horses, and rabbits due to the presence of an acquired hemolysin/cytolysin gene. The majority of clinically relevant enterococcal species hydrolyze pyrrolidonyl-β-naphthylamide (PYR); this characteristic is helpful in differentiating enterococci from organisms of the *Streptococcus gallolyticus* group (formerly known as *S. bovis*, which includes *S. gallolyticus*, *S. pasteurianus*, and *S. infantarius*) and from *Leuconostoc* species. Although many species of enterococci have been isolated from human infections, the overwhelming majority of cases are caused by two species, *E. faecalis* and *E. faecium*. Less frequently isolated species include *Enterococcus gallinarum*, *E. durans*, *E. hirae*, and *E. avium*.

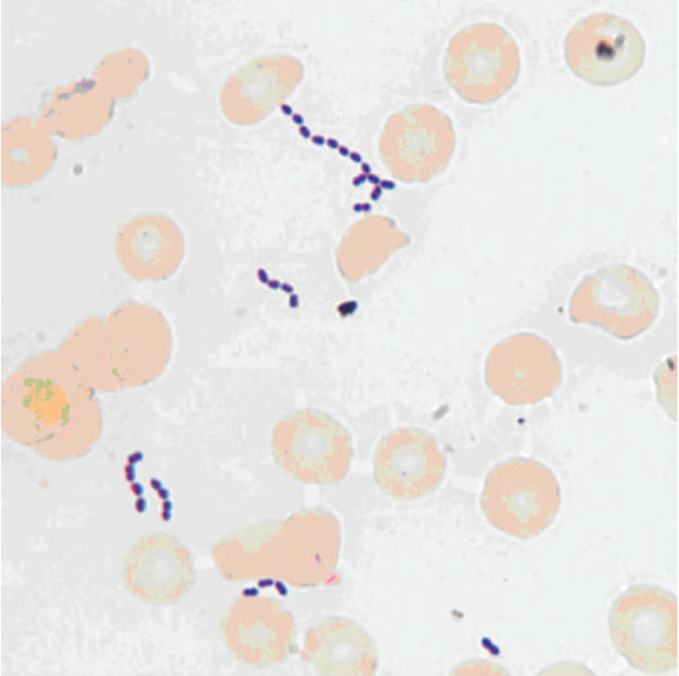

FIGURE 154-1 Gram's stain of cultured blood from a patient with enterococcal bacteremia. Oval gram-positive bacterial cells are arranged as diplococci and short chains. *(Courtesy of Audrey Wanger, PhD.)*

PATHOGENESIS

Enterococci are normal inhabitants of the large bowel of human adults, although they usually make up <1% of the culturable intestinal microbiota. In the healthy human host, enterococci are typical symbionts that coexist with other gastrointestinal bacteria; in fact, the utility of certain enterococcal strains as probiotics in the treatment of diarrhea suggests their possible role in maintaining the homeostatic equilibrium of the bowel. These commensals play a role in *colonization resistance*, or the ability of a healthy gastrointestinal microbiota to impede the establishment of a population of drug-resistant bacteria such as vancomycin-resistant enterococci (VRE). Colonization resistance arises from a complex set of metabolic and immunologic interactions between the host, pathogen, and intestinal microbiota, many of which are disrupted in hospitalized or chronically ill patients.

Several studies have shown that a higher level of gastrointestinal colonization is a critical factor in the pathogenesis of enterococcal infections. Physical factors, such as stomach pH and the mucin layer on the interior of the intestinal lumen, provide a barrier and limit pathogen access to the intestinal epithelium. In the hospital setting, administration of medications that suppress stomach acid secretion, or degradation of the mucin layer by gut commensals during periods of decreased oral intake, can disrupt these protective layers. The availability of specific carbon sources also has been shown to influence the ability of enterococci to colonize the intestine. Depletion of fructose by members of the genus *Olsenella* leads to impaired growth and a decrease in the enteric burden of VRE in a mouse-colonization model, while the presence of lactose favors the outgrowth of enterococci. Further, enterococci may potentiate the fitness and virulence of pathogens such as *Clostridioides difficile* by altering the availability of fermentable amino acids, by promoting *C. difficile* fitness, and by depletion of arginine, which leads to increased *C. difficile* toxin production.

---

*Deceased.

One of the most important factors that promotes increased gastrointestinal colonization by enterococci is the administration of antimicrobial agents since enterococci are intrinsically resistant to a variety of commonly used antibacterial drugs. In particular, antibiotics that are excreted in the bile and have broad-spectrum activity (e.g., certain cephalosporins and antimicrobials with anaerobic activity) are usually associated with the recovery of higher numbers of enterococci from feces. However, the increased colonization by hospital-associated strains of *E. faecium* in the presence of antimicrobial agents appears to be due to more than the simple filling of a "biologic niche." Studies of colonization dynamics in mouse intestines suggest host-derived antimicrobial peptides produced by the innate immune system (such as the lectin RegIIIγ) and compounds such as lantibiotics or bacteriocins produced by members of the microbiota itself are important mediators of colonization resistance. Activation of Toll-like receptors by lipopolysaccharide (an important component of the gram-negative cell envelope) leads, in mice, to increased production of RegIIIγ, and loss of this stimulation by antibiotic-induced disruptions of commensal gram-negative bacteria impairs clearance of VRE from the intestines. Similarly, antimicrobial lantibiotics produced by commensal bacteria (such as *Blautia producta*) are active against VRE in vitro, but this organism may require a cooperative partnership with other members of the microbiota to effectively provide colonization resistance. Disruption of these partnerships by antibiotic administration can lead to an environment where VRE can flourish. Another factor that may contribute to enterococcal survival in the gastrointestinal tract is the production of bacteriocins (molecules that kill competing bacteria). Strains of *E. faecalis* harboring pheromone-producing plasmids that code for bacteriocins are capable of outcompeting enterococci lacking such plasmids, and conjugative transfer of these elements enhances the survival of recipient enterococci. In the absence of antibiotics, hospital-associated lineages of *E. faecium* seem to be less adapted for survival in the gastrointestinal tract than are commensal *E. faecium* strains. Studies examining the rate of carriage of VRE in patients after discharge from the hospital document a median time to clearance between 2 and 4 months in patients without ongoing risk factors, such as continued antibiotic use, residence in a long-term care facility, or need for hemodialysis.

Colonization by enterococci may also have important effects on the treatment of noninfectious syndromes. In hematopoietic stem cell transplant patients, domination of the gastrointestinal tract by enterococci is associated with increased inflammation and a greater severity of illness and mortality in graft-versus-host disease. Certain enterococcal species, such as *E. faecium*, *E. durans*, and *E. hirae*, modulate the host immune response and can potentiate the activity of anti-PD-L1 immune checkpoint inhibitor therapy. This effect is mediated by the presence of *salA*, a gene which encodes a peptidoglycan hydrolase with D,L-endopeptidase activity. The muropeptides generated by SalA induce host activation of NF-κB via NOD2, leading to an enhanced antitumor immune response.

Several vertebrate, worm, and insect models have been developed to study the role of possible pathogenic determinants in both *E. faecalis* and *E. faecium*. Three main groups of virulence factors may increase the ability of enterococci to colonize the gastrointestinal tract and/or cause disease. The first group, *enterococcal secreted factors*, are molecules released outside the bacterial cell that contribute to the process of infection. The best studied of these molecules include enterococcal hemolysin/cytolysin and two enterococcal proteases (gelatinase and serine protease). Enterococcal cytolysin is a heterodimeric toxin produced by some strains of *E. faecalis* that is capable of lysing human (as well as equine but not ovine) RBCs as well as polymorphonuclear leukocytes and macrophages. A newly described pore-forming toxin, Epx, specifically targets human leukocyte antigen type 1, and has been shown to be active against peripheral blood mononuclear cells and intestinal organoids in vitro. *E. faecalis* gelatinase and serine protease are thought to mediate virulence by several mechanisms, including the degradation of host tissues and the modification of critical components of the immune system. Mutants lacking the genes corresponding to these proteins are highly attenuated in experimental animal models of peritonitis, endocarditis, and endophthalmitis.

A second group of virulence factors, *enterococcal surface components*, includes adhesins and is thought to contribute to bacterial attachment to extracellular matrix molecules in the human host. Several molecules on the surface of enterococci have been characterized and shown to play a role in the pathogenesis of enterococcal infections. Among the characterized adhesins is *aggregation substance* of *E. faecalis*, which mediates the attachment of bacterial cells to each other, thereby facilitating conjugative plasmid exchange. Several lines of evidence indicate that aggregation substance and enterococcal cytolysin act synergistically to increase the virulence potential of *E. faecalis* strains in experimental endocarditis. The surface protein adhesin of collagen of *E. faecalis* (Ace) and its *E. faecium* homologue (Acm) are microbial surface components adhering to matrix molecules (MSCRAMMs); they recognize adhesive matrix molecules involved in bacterial attachment to host proteins such as collagen, fibronectin, and fibrinogen. Both Ace and Acm are collagen adhesins that are important in the pathogenesis of experimental endocarditis. Pili of both *E. faecalis* and *E. faecium* are important mediators of attachment to and invasion of host tissues. Mutants of *E. faecalis* lacking pili are attenuated in biofilm production, experimental endocarditis, and urinary tract infections (UTIs). Other surface proteins that share structural homology with MSCRAMMs and appear to play a role in enterococcal attachment to the host and in virulence include the *E. faecalis* surface protein Esp and its *E. faecium* homologue Esp_{fm}, the second collagen adhesin of *E. faecium* (Scm), the surface proteins of *E. faecium* (Fms), SgrA (which binds to components of the basal lamina), and EcbA (which binds to collagen type V). Additional surface components apparently associated with pathogenicity include the Erl protein (a protein from the WxL family) and polysaccharides, which are thought to interfere with phagocytosis of the organism by host immune cells. Teichoic acids on the enterococcal surface appear to be immunogenic, and antibodies to these molecules are protective in some animal models.

The third group of virulence factors has not been well characterized but includes the *E. faecalis* stress protein Gls24, which has been associated with enterococcal resistance to bile salts and appears to be important in the pathogenesis of endocarditis, and the $\textit{hyl}_{\text{Efm}}$-containing plasmids of *E. faecium*, which are transferable between strains and increase gastrointestinal colonization by *E. faecium*. In mouse peritonitis, acquisition of these plasmids increased the lethality of a commensal strain of *E. faecium* and enhanced colonization of the uroepithelium. A gene encoding a regulator of oxidative stress (AsrR) has been identified as an important virulence factor of *E. faecium.*

EPIDEMIOLOGY

According to data collected from 2018 to 2021 by the National Healthcare Safety Network of the Centers for Disease Control and Prevention, enterococci are the third most common isolates (after *Escherichia coli* and staphylococci) from hospital-associated infections in the United States. Although *E. faecalis* remains the predominant species recovered from nosocomial infections, the isolation of *E. faecium* has increased substantially in the past 20 years and accounts for approximately one-third of all enterococcal infections identified to the species level. This point is important, since *E. faecium* is by far the most resistant and challenging enterococcal species to treat. More than 90% of *E. faecium* isolates are resistant to ampicillin (historically the most effective β-lactam agent against enterococci), while ampicillin resistance in *E. faecalis* is uncommon. Vancomycin resistance in *E. faecium* isolates ranges from 50–70% in acute care hospitals in the United States to up to 80% in long-term care facilities. Resistance to vancomycin in *E. faecalis* isolates is less common, with a higher prevalence in long-term care facilities (10–12%) than in acute care hospitals (2–5%).

The dynamics of enterococcal transmission and dissemination in the hospital environment have been extensively studied, with a focus on VRE. These studies have revealed that VRE colonization of the gastrointestinal tract is a critical step in the development of enterococcal disease and that a substantial proportion of patients colonized with VRE remain colonized for prolonged periods (sometimes >1 year) and are more likely than patients without VRE colonization to develop an *Enterococcus*-related illness (e.g., bacteremia). Important factors associated with VRE colonization and persistence in the gut include prolonged hospitalization; long courses of antibiotic therapy; hospitalization in long-term care facilities, surgical units, and/or intensive care units; organ transplantation; renal failure (particularly in patients undergoing hemodialysis) and/or diabetes; high APACHE (Acute Physiology and Chronic Health Evaluation) scores; and physical proximity to patients infected or colonized with VRE or these patients' rooms. Once a patient becomes colonized with VRE, several key factors are involved in the organisms' dissemination in the hospital environment. VRE can survive exposure to heat and certain disinfectants and have been found on numerous inanimate objects in the hospital, including bed rails, medical equipment, doorknobs, gloves, telephones, and computer keyboards. Thus, health care workers and the environment play pivotal roles in enterococcal transmission from patient to patient, and infection control measures are crucial in breaking the chain of transmission. Moreover, two meta-analyses have found that, independent of the patient's clinical status, VRE infection increases the risk of death over that among individuals infected with a glycopeptide-susceptible enterococcal strain.

The epidemiology of enterococcal disease and the emergence of VRE have followed slightly different trends in other parts of the world than in the United States. In Europe, the emergence of VRE in the mid-1980s was seen primarily in isolates recovered from animals and healthy humans rather than from hospitalized patients. The presence of VRE was associated with the use of the glycopeptide avoparcin as a growth promoter in animal feeds; this association prompted the European Union to ban the use of this compound in animal husbandry in 1996. However, after an initial decrease in the isolation of VRE from animals and humans, the prevalence of hospital-associated VRE infections has slowly increased, from 8.1% in 2012 to 17.6% in 2022. Furthermore, there are important regional differences in rates of vancomycin resistance among *E. faecium*, with rates below 5–10% in most of western Europe and Scandinavia, and 25–50% across much of southern and eastern Europe. Despite regional differences, Europe has seen a general trend of increasing rates of VRE over the past decade, although these rates continue to be much lower than in the United States. The reasons are not totally understood, but it has been postulated that this difference is related to the higher levels of human antibiotic use in the United States. Recent data have also shown increasing rates of enterococcal resistance to vancomycin in Latin American countries, with 34% of clinical *E. faecium* isolates found to be resistant in a multicenter study including hospitals from Colombia, Venezuela, Ecuador, and Peru. In Asia, rates of vancomycin resistance among enterococci appear to be similar to those in U.S. hospitals.

The ability to sequence bacterial genomes has increased our understanding of bacterial diversity, evolution, pathogenesis, and mechanisms of antibiotic resistance. Ongoing collection of enterococcal isolates as a part of epidemiologic surveillance has allowed researchers to trace the evolutionary trajectory of enterococci from their origin to the emergence of hospital-adapted clones. Sequence analysis suggests the genus appeared ~400 million years ago with the advent of terrestrial animals. Several key features aided in this transition, including the ability to recombine large portions of chromosomal DNA from the *core genome* and a malleable *accessory genome* consisting of plasmids, phages, and mobile genetic elements. This genomic plasticity contributes to the rising rates of antibiotic resistance seen within the genus and, in particular, in *E. faecium*.

A large proportion of the genomes available for analysis belong to *E. faecium*, due to its importance as a nosocomial pathogen and the epidemiologic surveillance projects to track the spread of vancomycin-resistant strains. The population can be divided into two large groups, or clades, of organisms: a hospital-associated clade A and a community-associated clade B. The hospital-associated clade appears to be evolving rapidly via a series of historic and ongoing recombination events. Certain regions of the chromosome show a convergence of genomic diversity both within and between clades, suggesting there are several "hot spots" for recombination. Importantly, these regions contain the polysaccharide biosynthesis cluster, which has been implicated in altering the antigenic capsular polysaccharide in streptococci, and the LiaFSR locus, a major cell envelope stress system that mediates resistance to membrane-active antibiotics and host antimicrobial peptides. Strains belonging to clade A are more frequently identified as isolates causing invasive disease and are more likely to carry drug resistance determinants, whereas clade B isolates largely retain a susceptible phenotype.

One reason for the propensity of clade A strains to acquire resistance determinants is that they more frequently lack a functional CRISPR-Cas system (short for *c*lustered *r*egularly *i*nterspaced *s*hort *p*alindromic *r*epeats). These systems serve as a primitive "immune system" and provide a genome defense for bacteria to protect them from foreign DNA, such as phages, but they also serve to reduce the frequency of acquisition of resistance genes borne on mobile genetic elements. Another reason for their survival in the hospital environment is that clade A isolates tend to possess alleles of penicillin-binding protein 5 (PBP5) associated with high-level β-lactam resistance in *E. faecium* and may express higher levels of this enzyme than commensal strains.

A notable feature of the distribution of strains in clade A in some studies is that they share a relatively recent common ancestor with *E. faecium* of livestock origin. Use of antibiotics in animal husbandry as both therapeutics and growth promoters has been linked to resistance in several important contexts, including glycopeptides, as mentioned above. This phenomenon suggests that continued surveillance, and an expanding understanding of the population structure of enterococci, may help identify potential reservoirs of resistance and inform policy to limit their spread.

■ CLINICAL SYNDROMES

Urinary Tract Infection and Prostatitis Enterococci are well-known causes of nosocomial UTI—the most common infection caused by these organisms **(Chap. 140)**. Enterococcal UTIs are usually associated with indwelling catheterization, instrumentation, or anatomic abnormalities of the genitourinary tract, and it is often challenging to differentiate between true infection and colonization (particularly in patients with chronic indwelling catheters). Their role as pathogens in otherwise healthy premenopausal woman with acute cystitis is less clear, with data from one study suggesting that enterococci recovered from midstream urine cultures were not predictive of bacteriuria in a subsequent catheterized specimen. The presence of leukocytes in the urine in conjunction with systemic manifestations (e.g., fever) or local signs and symptoms of infection with no other explanation and a positive urine culture ($\geq 10^5$ CFU/mL) suggests the diagnosis. Moreover, enterococcal UTIs often occur in critically or chronically ill patients whose comorbidities may obscure the diagnosis. In many cases, removal of the indwelling catheter may suffice to eradicate the organism without specific antimicrobial therapy. In rare circumstances, UTIs caused by enterococci may run a complicated course, with the development of pyelonephritis and perinephric abscesses that may be a portal of entry for bloodstream infections (see below). Enterococci are also known causes of chronic prostatitis, particularly in men whose urinary tract has been manipulated surgically or endoscopically. These infections can be difficult to treat since the agents most potent against enterococci (i.e., aminopenicillins and glycopeptides) penetrate prostatic tissue poorly. Chronic prostatic infection can be a source of recurrent enterococcal bacteremia.

Bacteremia and Endocarditis Bacteremia without endocarditis is another frequently encountered presentation of enterococcal disease. Intravascular catheters and other devices are commonly associated with these bacteremic episodes **(Chap. 147)**. Other well-known sources of enterococcal bacteremia include the gastrointestinal and hepatobiliary tracts; pelvic and intraabdominal foci; and, less frequently, wound infections, UTIs, and bone infections. In the United States, enterococci are ranked second (after staphylococci) as etiologic agents of central line–associated bacteremia. Patients with enterococcal bacteremia usually have comorbidities and have been in the hospital for prolonged periods; they commonly have received several courses of antibiotics. Several studies indicate that the isolation of *E. faecium* from the blood may lead to worse outcomes and higher mortality rates than when other enterococcal species are isolated; this finding may be related to the higher prevalence of vancomycin and ampicillin resistance in *E. faecium* than in other enterococcal species, with the consequent reduction of therapeutic options. In some cases (usually when the gastrointestinal tract is the source), enterococcal bacteremia may be polymicrobial, with gram-negative organisms isolated at the same time. In addition, several cases have been documented in which enterococcal bacteremia was associated with *Strongyloides stercoralis* hyperinfection syndrome in immunocompromised patients.

Enterococci are important causes of community- and health care–associated endocarditis, ranking second after staphylococci in the latter infections, and recent studies have implicated *E. faecalis* as the leading cause of endocarditis after transcatheter aortic valve implantation (TAVI). The presumed initial source of bacteremia leading to endocarditis is the gastrointestinal or genitourinary tract—e.g., in patients who have malignant and inflammatory conditions of the gut or have undergone procedures in which these tracts are manipulated. The affected patients tend to be male and elderly and to have other debilitating diseases and heart conditions. Both prosthetic and native valves can be involved; mitral and aortic valves are affected most often. Community-associated endocarditis (usually caused by *E. faecalis*) also occurs in patients with no apparent risk factors or cardiac abnormalities. Endocarditis in women of childbearing age was well described in the past. The typical presentation of enterococcal endocarditis is a subacute course of fever, weight loss, malaise, and cardiac murmur; typical stigmata of endocarditis (e.g., petechiae, Osler's nodes, Roth's spots) are found in only a minority of patients. Atypical manifestations include arthralgias and manifestations of metastatic disease (splenic abscesses, hiccups, pain in the left flank, pleural effusion, and spondylodiscitis). Embolic complications are variable and can affect the brain. Heart failure is a common complication of enterococcal endocarditis, and valve replacement may be critical in curing this infection, particularly when multidrug-resistant organisms or major complications are involved. A clinical scoring system (DENOVA) has been proposed to help differentiate enterococcal bacteremia from true endocarditis and determine the need for echocardiography. The duration of therapy is usually 4–6 weeks, with more prolonged courses suggested for multidrug-resistant isolates in the absence of valvular replacement. Risk of relapse for endocarditis due to *E. faecalis* varies across studies, but it is higher in patients who are treated with antibiotics alone (2–11%) than in those who undergo surgery (0–3%); relapse may occur out to 1-year post-treatment.

Meningitis Enterococcal meningitis is an uncommon disease (accounting for only ~4% of meningitis cases) that is usually associated with neurosurgical interventions and conditions such as shunts, central nervous system (CNS) trauma, and cerebrospinal fluid (CSF) leakage. In some instances—usually in patients with a debilitating condition, such as cardiovascular or congenital heart disease, chronic renal failure, malignancy, receipt of immunosuppressive therapy, or HIV/AIDS—presumed hematogenous seeding of the meninges is seen in infections such as endocarditis or bacteremia. Fever and changes in mental status are common, whereas overt meningeal signs are less so. CSF findings are consistent with bacterial infection—i.e., pleocytosis, with a predominance of polymorphonuclear leukocytes (average, ~500/μL), an elevated serum protein level (usually >100 mg/dL), and a decreased glucose concentration (average, 28 mg/dL). Gram's staining yields a positive result in about half of cases, with a high rate of organism recovery from CSF cultures; the most common species isolated are *E. faecalis* and *E. faecium*. Complications include hydrocephalus, brain abscesses, and stroke. As mentioned before for bacteremia, an association with *Strongyloides* hyperinfection also has been documented.

Intraabdominal, Pelvic, and Soft Tissue Infections As mentioned above, enterococci are part of the commensal microbiota of the gastrointestinal tract and can produce spontaneous peritonitis in cirrhotic individuals and in patients undergoing chronic ambulatory peritoneal dialysis **(Chap. 137)**. These organisms are commonly found (usually along with other bacteria, including enteric gram-negative species and anaerobes) in clinical samples from intraabdominal and pelvic collections. The presence of enterococci in intraabdominal infections is sometimes considered to be of little clinical relevance. Several studies have shown that the role of enterococci in intraabdominal infections originating in the community and involving previously healthy patients is minor since surgery and broad-spectrum antimicrobial drugs that do not target enterococci are often sufficient to treat these infections successfully. In the past few decades, however, these organisms have become prominent as a cause of intraabdominal infections in hospitalized patients because of the emergence and spread of vancomycin resistance among enterococci and the increase in rates of nosocomial infections due to multidrug-resistant *E. faecium* isolates. In fact, several studies have now documented treatment failures due to enterococci, with consequently increased rates of postoperative complications and death among patients with intraabdominal infections. Thus, anti-enterococcal therapy is recommended for nosocomial peritonitis in immunocompromised and severely ill patients who have had a prolonged hospital stay, have undergone multiple procedures, have persistent abdominal sepsis and collections, or have risk factors for the development of endocarditis (e.g., prosthetic or damaged heart valves). Conversely, specific treatment for enterococci in the first episode of intraabdominal infection originating in the community and affecting previously healthy patients with no important cardiac risk factors for endocarditis does not appear to be beneficial.

Enterococci are commonly isolated from soft tissue infections **(Chap. 134)**, particularly those involving surgical wounds **(Chap. 147)**. In fact, these organisms rank third as agents of nosocomial surgical-site infections, with *E. faecalis* the most frequently isolated species. The clinical relevance of enterococci in some of these infections—as in intraabdominal infections—is a matter of debate; differentiating between colonization and true infection is sometimes challenging, although in some cases, enterococci have been recovered from lung, liver, and skin abscesses. Diabetic foot and decubitus ulcers are often colonized with enterococci and may be the portal of entry for bone infections.

Other Infections Enterococci are well-known causes of neonatal infections, including sepsis (mostly late-onset), bacteremia, meningitis, pneumonia, and UTI. Outbreaks of enterococcal sepsis in neonatal units have been well documented. Risk factors for enterococcal disease in newborns include prematurity, low birth weight, indwelling devices, and abdominal surgery. Enterococci have also been described as etiologic agents of bone and joint infections, including vertebral osteomyelitis, usually in patients with underlying conditions such as diabetes or endocarditis. Similarly, enterococci have been isolated from bone infections in patients who have undergone arthroplasty or reconstruction of fractures with the placement of hardware. Since enterococci can produce a biofilm that is likely to alter the efficacy of anti-enterococcal agents, treatment of infections that involve foreign material is challenging, and removal of the hardware may be necessary to eradicate the infection. Rare cases of enterococcal pneumonia, lung abscess, and spontaneous empyema have been described.

TREATMENT

Enterococcal Infections

GENERAL PRINCIPLES

Enterococci are intrinsically resistant and/or tolerant to several antimicrobial agents. (*Tolerance* is defined as lack of killing by drug concentrations 32 times higher than the minimal inhibitory concentration [MIC].) Monotherapy for endocarditis with a β-lactam antibiotic (to which many enterococci are tolerant) has produced disappointing results, with high relapse rates after the end of therapy. However, the addition of an aminoglycoside to a cell wall–active agent (a β-lactam or a glycopeptide) increases cure rates and eradicates the organisms; moreover, this combination is synergistic and bactericidal in vitro. Therefore, for many decades, combination therapy with a cell wall–active agent and an aminoglycoside was the standard of care for endovascular infections caused by enterococci. This synergistic effect can be explained, at least in part, by the increased penetration of the aminoglycoside into the bacterial cell, presumably as a result of cell-wall alterations produced by the β-lactam (or glycopeptide). Nonetheless, attaining synergistic bactericidal activity in the treatment of severe enterococcal infections—particularly those caused by *E. faecium*—has become increasingly difficult because of the development of resistance to virtually all antibiotics available for this purpose.

The treatment of *E. faecalis* differs substantially from that of *E. faecium* **(Tables 154-1 and 154-2)**, mainly because of differences in resistance profiles (see below). For example, resistance to ampicillin and vancomycin is rare in *E. faecalis*, whereas these antibiotics are only infrequently useful against current isolates of *E. faecium*. Moreover, as a consequence of the challenges and therapeutic limitations posed by the emergence of drug resistance in enterococci, valve replacement may need to be considered in the treatment of endocarditis caused by multidrug-resistant enterococci. Less severe infections are often related to indwelling intravascular catheters; removal of the catheter increases the likelihood of enterococcal eradication by a subsequent short course of appropriate antimicrobial therapy.

CHOICE OF ANTIMICROBIAL AGENTS

Among the β-lactams, the most active are the aminopenicillins (ampicillin, amoxicillin); next most active are penicillin G, imipenem, and ureidopenicillins (i.e., piperacillin). Cephalosporins are not active as monotherapy. The only two aminoglycosides recommended for synergistic therapy in severe enterococcal infections are gentamicin and streptomycin. This is because the most common acquired enzyme conferring high-level resistance to gentamicin also is active against tobramycin and amikacin but not streptomycin, and the resistance mechanisms causing high-level resistance to streptomycin do not affect gentamicin. The use of amikacin is strongly discouraged because it is infrequently active, and tobramycin should never be used for the treatment of *E. faecium* infections due to the presence of a chromosomally encoded, species-specific, tobramycin-modifying enzyme. Aminoglycoside monotherapy should not be employed. Vancomycin is an alternative to β-lactam drugs for the treatment of *E. faecalis* infections but is less useful against *E. faecium* because resistance is common.

As mentioned above, use of the aminoglycoside–ampicillin combination for *E. faecalis* infections has become increasingly problematic because of toxicity in critically ill patients and increased rates of high-level resistance to aminoglycosides. An observational, nonrandomized, comparative study encompassing a multicenter cohort was conducted in 17 Spanish hospitals and 1 Italian hospital; the results indicated that a 6-week course of ampicillin plus ceftriaxone is as effective as ampicillin plus gentamicin in the treatment of *E. faecalis* endocarditis, with less risk of toxicity. Therefore, this regimen should be considered in patients at risk for aminoglycoside toxicity or those with isolates displaying high-level resistance to aminoglycosides, and it is now recommended as first-line therapy for *E. faecalis* endocarditis. Use of dual β-lactam regimens for ampicillin-susceptible isolates of *E. faecium* has not been studied in the clinical setting. Limited in vitro data suggest that synergism between ampicillin and ceftriaxone is not reliably active against these isolates.

Linezolid is the only agent approved by the U.S. Food and Drug Administration (FDA) for the treatment of VRE infections (Table 154-2). Linezolid is not bactericidal, and its use in severe endovascular infections has produced mixed results; therefore, it is recommended only as an alternative to other agents for such infections. In addition, linezolid may cause significant toxicities (thrombocytopenia, peripheral neuropathy, optic neuritis, and lactic acidosis) when used in regimens given for >2 weeks. Nonetheless, linezolid may play a role in the treatment of enterococcal meningitis and other CNS infections, although clinical data are limited.

The lipopeptide daptomycin is a bactericidal antibiotic with in vitro activity against all enterococci. Although daptomycin is not approved by the FDA for the treatment of VRE or *E. faecium*

TABLE 154-1 Suggested Regimens for the Management of Infections Caused by *Enterococcus faecalis*

CLINICAL SYNDROME	SUGGESTED THERAPEUTIC OPTIONS[a]
Endovascular infections (including endocarditis)	• Ampicillin[b] (12 g/d IV in divided doses q4h) plus ceftriaxone (2 g IV q12h) • Ampicillin[b] (12 g/d IV in divided doses q4h or by continuous infusion) or penicillin (18–30 mU/d IV in divided doses q4h or by continuous infusion) *plus* an aminoglycoside[c] • Vancomycin[d] (15 mg/kg IV per dose) plus an aminoglycoside[c] • High-dose daptomycin[e] ± another active agent[f] • Ampicillin[b] plus imipenem
Nonendovascular bacteremia[g]	• Ampicillin[b] (12 g/d IV in divided doses q4h) or penicillin (18 mU/d IV in divided doses q4h) ± an aminoglycoside[c] or ceftriaxone • Vancomycin[d] (15 mg/kg IV per dose) • High-dose daptomycin[e] ± another active agent[f] • Linezolid (600 mg IV/PO q12h)
Meningitis	• Ampicillin (20–24 g/d IV in divided doses q4h) or penicillin (24 mU/d IV in divided doses q4h) *plus* an aminoglycoside[c,h] and consider adding ceftriaxone (2 g IV q12h) • Vancomycin (500–750 mg IV q6h)[d] *plus* an aminoglycoside[c] or rifampin • Linezolid • High-dose daptomycin[e] (plus intrathecal daptomycin) ± another active agent[f]
Urinary tract infections (uncomplicated)	• Amoxicillin (500 mg PO q8h) • Fosfomycin (3 g PO, one dose)[i] • Ampicillin (500 mg IV or PO q6h) • Nitrofurantoin (100 mg PO q6h)

[a] Authors’ preferences are underlined for each category; many of the regimens are off-label. [b] In rare cases, β-lactamase-producing isolates may be present. Because these isolates are not detected by conventional determination of the minimal inhibitory concentration, additional tests (e.g., the nitrocefin disk) are recommended for isolates from endocarditis. The use of ampicillin/sulbactam (12–24 g/d) is suggested in these cases. If choosing a β-lactam other than ampicillin for an invasive infection, susceptibility testing for penicillin should be performed to identify penicillin-resistant, ampicillin-susceptible (PRASEF) isolates. Limited in vitro data suggest that synergistic killing by dual β-lactams (i.e., ampicillin and ceftriaxone) may be compromised in PRASEF isolates. Some suggest caution with the combination of ampicillin and ceftriaxone in PRASEF isolates and would consider an alternate therapy (such as daptomycin plus ampicillin) guided by susceptibility testing in cases of persistent bacteremia or signs of clinical failure. [c] Only if the organism does not exhibit high-level resistance (HLR) to aminoglycosides. This test is performed by the clinical microbiology laboratory only for gentamicin or streptomycin (growth of enterococci on agar containing gentamicin [500 μg/mL] or streptomycin [2000 μg/mL]). If HLR is documented, the aminoglycoside will not act synergistically with the other agent in the combination. However, HLR to one of these aminoglycosides does not indicate resistance to the other agent (as reported individually). HLR to gentamicin implies lack of synergism with tobramycin and with amikacin. Gentamicin (1–1.5 mg/kg IV q8h) and streptomycin (15 mg/kg per day IV/IM in two divided doses) are the only two recommended aminoglycosides. [d] Vancomycin is recommended only as an alternative to β-lactam agents in cases of allergy or toxicity plus the inability to desensitize. Specific pharmacologic targets for trough concentrations have not been clinically evaluated in enterococcal bacteremia; trough concentrations of 15–20 mg/L have been associated with increased rates of nephrotoxicity. Vancomycin-resistant strains of *E. faecalis* have been reported. [e] Consider doses of 10–12 mg/kg once daily, adjusted for renal function. Monitoring of creatine phosphokinase levels is recommended throughout therapy because of possible rhabdomyolysis. [f] Potentially active agents may include an aminoglycoside (if HLR is not detected), ampicillin, ceftaroline, tigecycline, eravacycline, or a fluoroquinolone (which, if the isolate is susceptible, may be favored in meningitis). The presence of mutations in LiaFSR seems to increase susceptibility to ampicillin and ceftaroline, and combinations of daptomycin with these compounds are bactericidal in vitro against such strains. [g] In selected cases of catheter-associated bacteremia, removal of the catheter and a short course of therapy (~5–7 days) may be sufficient. A single positive blood culture that is likely to be associated with a catheter in a patient who is otherwise doing well may not require therapy after removal of the catheter. Patients at high risk for endovascular infections or with severe disease may benefit from synergistic combination therapy. [h] The addition of intrathecal or intraventricular therapy with gentamicin (2–10 mg/d) if the organism does not exhibit HLR or with vancomycin (10–20 mg/d) when the isolate is susceptible has been suggested by some authorities. The addition of systemic rifampin (a good CSF-penetrating agent) may be considered. The combination of ampicillin and ceftriaxone may have clinical benefit (by analogy with endocarditis), but no cases treated with this combination have been reported; the authors would use this combination. [i] Approved by the U.S. Food and Drug Administration only for uncomplicated urinary tract infections caused by vancomycin-susceptible *E. faecalis.*

TABLE 154-2 Suggested Regimens for the Management of Infections Caused by Vancomycin- and Ampicillin-Resistant *Enterococcus faecium*

CLINICAL SYNDROME	SUGGESTED THERAPEUTIC OPTIONS[a]
Endovascular infections (including endocarditis)	• High-dose daptomycin[b] plus another agent[c] ± an aminoglycoside[d] • Linezolid (600 mg IV q12h) • High-dose ampicillin (if MIC is ≤64 μg/mL) ± an aminoglycoside[d]
Nonendovascular bacteremia[e]	• High-dose daptomycin[b] ± another agent[c] ± an aminoglycoside[d] • Linezolid (600 mg IV q12h)
Meningitis	• Linezolid (600 mg IV q12h) ± another CSF-penetrating active agent[f] • High-dose daptomycin[b] (plus intraventricular daptomycin) ± another CSF-penetrating active agent[f,g]
Urinary tract infections	• Fosfomycin (3 g PO, one dose)[h] • Nitrofurantoin (100 mg PO q6h) • Ampicillin or amoxicillin (2 g IV/PO q4–6h)[i]

[a] Authors' preferences are underlined for each category; many of these regimens are off-label. [b] Daptomycin at doses of 10–12 mg/kg once daily is suggested (off-label). Close monitoring of creatine phosphokinase levels is recommended throughout therapy because of possible rhabdomyolysis. [c] Potentially active agents may include ampicillin or ceftaroline (even if the infecting strain is resistant in vitro), tigecycline, eravacycline, or a fluroquinolone. In vitro synergism of daptomycin with ampicillin or ceftaroline has been observed against some isolates that subsequently become nonsusceptible to daptomycin during therapy. The synergism of daptomycin and β-lactams is associated with mutations in LiaFSR. The authors prefer use of combination therapy for all endovascular infections or for bacteremic patients in whom source control cannot be achieved, as in vitro data suggest combination therapy may prevent or delay the emergence of daptomycin resistance. [d] Only if the organism does not exhibit high-level resistance to aminoglycosides (see Table 154-1, footnote *c*). [e] In selected cases of catheter-associated bacteremia, removal of the catheter and a short course of therapy (~5–7 days) may be sufficient. A single positive blood culture that is likely to be associated with a catheter in a patient who is otherwise doing well may not require therapy after removal of the catheter. [f] Fluoroquinolones (e.g., moxifloxacin) and rifampin (if the isolate is susceptible to each agent) reach therapeutic levels in the cerebrospinal fluid. [g] Intrathecal gentamicin (2–10 mg/d) if high-level resistance is not detected. Intraventricular daptomycin has been used in two cases of meningitis. [h] Approved by the U.S. Food and Drug Administration (FDA) only for uncomplicated urinary tract infections caused by vancomycin-susceptible *E. faecalis*. [i] Concentrations of amoxicillin and ampicillin in urine far exceed those in serum and may be potentially effective even against isolates with high MICs. Doses up to 12 g/d are suggested for isolates with MICs of ≥64 μg/mL. Of note, quinupristin-dalfopristin (Q-D) lost FDA approval for infections due to vancomycin-resistant *Enterococcus*, and it was discontinued by the manufacturer in 2022.

infections, it has been used alone (at high dosage) or in combination with other agents (ampicillin, ceftaroline, and tigecycline) with apparent success against multidrug-resistant enterococcal infections (Tables 154-1 and 154-2). The main adverse reactions to daptomycin are elevated creatine phosphokinase levels and eosinophilic pneumonitis (rare). Daptomycin is not useful against pulmonary infections because the pulmonary surfactant inhibits its antibacterial activity.

Several meta-analyses have examined the question of which agent should be preferred for VRE bacteremia—linezolid or daptomycin. These studies concluded either no difference between the two drugs or favored linezolid due to lower all-cause and infection-related mortality, but they were limited by small patient numbers and heterogenous outcomes. A subsequent large retrospective observational study from the Veterans Affairs database reported lower rates of all-cause mortality at 30 days and less microbiologic failure (i.e., positive cultures despite therapy) with daptomycin compared to linezolid. One important observation from these investigations is that the efficacy of daptomycin is dependent on the dose, with improved outcomes seen with high-dose daptomycin therapy (≥10 mg/kg) compared to standard-dose therapy (6 mg/kg). Genome sequencing of clinical isolates has revealed that mutations in genes associated with daptomycin resistance are not uncommon (see "Antimicrobial Resistance," below) and were associated with the emergence of resistance to daptomycin at lower simulated dosing regimens (6 mg/kg) in experimental models of infection. These data led the Clinical Laboratory and Standards Institute (CLSI) to change the daptomycin breakpoints in 2019. For *E. faecium*, all isolates with an MIC of ≤4 mg/L are placed in a "susceptible dose-dependent" category based on a dosing regimen of 8–12 mg/kg, while those with an MIC ≥8 mg/L are considered resistant. For all other enterococci, isolates are considered susceptible with an MIC of ≤2 mg/L, intermediate with an MIC of 4 mg/L, and resistant with an MIC ≥8 mg/L.

The glycylcycline drug tigecycline is active in vitro against all enterococci, regardless of the isolates' vancomycin susceptibility. However, its use as monotherapy for endovascular or severe enterococcal infections is not recommended because of low attainable blood levels. Newer-generation tetracyclines, such as eravacycline and omadacycline, also display in vitro activity, but their role in the treatment of enterococcal infections remains to be evaluated.

Telavancin, a lipoglycopeptide approved by the FDA for the treatment of skin and soft tissue infections as well as hospital-associated pneumonia, is active against vancomycin-susceptible enterococci but not VRE. Likewise, dalbavancin, a lipoglycopeptide antibiotic with a long terminal half-life, has FDA approval for skin and soft tissue infections due to susceptible strains of *E. faecalis*, but no activity against VRE. Oritavancin, a novel glycopeptide with activity against VRE, has been approved for the treatment of acute bacterial skin and soft tissue infections caused by susceptible organisms, including vancomycin-susceptible *E. faecalis*. The MICs of oritavancin against VRE are low, and this compound may be a promising drug for VRE treatment in the future.

Lastly, tedizolid—a new oxazolidinone now available for clinical use—is approved only for the treatment of *E. faecalis* infections. Tedizolid is more potent than linezolid in vitro against VRE strains; however, its role in severe VRE infections remains to be determined.

ANTIMICROBIAL RESISTANCE

Resistance to β-lactam agents continues to be observed only infrequently in *E. faecalis* but is characteristic of *E. faecium*. The mechanism of ampicillin resistance in *E. faecium* is related to a penicillin-binding protein (PBP) designated PBP5, which is the critical target of β-lactam antibiotics. PBP5 exhibits low affinity for ampicillin and can synthesize cell wall in the presence of this antibiotic, even when other PBPs are inhibited. The version of this protein found in ampicillin-resistant hospital-associated strains has multiple amino-acid differences that even further decrease the affinity of PBP5 for ampicillin; recent structural work has demonstrated that these changes lead to a more dynamic binding pocket with an increase in the hydrolysis of ampicillin at the active site, thus allowing for increased turnover of the enzyme. Increased production of PBP5 is also seen in high-level ampicillin-resistant (e.g., MIC >32 μg/mL) clinical strains. In *E. faecalis*, ampicillin resistance is extremely rare; however, an emerging phenotype of penicillin-resistant but ampicillin-susceptible *E. faecalis* (PRASEF) has been described. Similar to *E. faecium*, changes in the promoter of PBP4 (the *E. faecalis* homologue of PBP5) or the amino-acid sequence of the enzyme itself are associated with increases in the MIC of ampicillin (although still below the susceptible breakpoint), penicillin, piperacillin, and imipenem; in vitro data suggest that these changes may also impair the synergistic activity of ampicillin and ceftriaxone against some PRASEF isolates.

Vancomycin is a glycopeptide antibiotic that inhibits cell-wall peptidoglycan synthesis in susceptible enterococci and has been widely used against enterococcal infections in clinical practice when the utility of β-lactams is limited by resistance, allergy, or adverse reactions. This effect is mediated by binding of the antibiotic to peptidoglycan precursors (UDP-MurNAc-pentapeptides) upon their exit from the bacterial cell cytoplasm. The interaction of vancomycin with the peptidoglycan is specific and involves the last two D-alanine residues of the precursor. The first isolates of VRE were documented in 1986, and vancomycin resistance (particularly in *E. faecium*) has since increased considerably around the world.

The mechanism involves the replacement of the last D-alanine residue of peptidoglycan precursors with D-lactate (e.g., VanA and VanB) or D-serine (e.g., VanC), with consequent high- and low-level resistance, respectively. There is significant heterogeneity among isolates, but either substitution substantially decreases the affinity of vancomycin for the peptidoglycan; with the D-lactate substitution, the affinity for binding to the pentapeptide precursor is decreased by ~1000-fold. Vancomycin-resistant organisms also produce enzymes that destroy the D-alanine-D-alanine ending precursors, ensuring that additional binding sites for vancomycin are not available.

The genes encoding the machinery responsible for vancomycin resistance are located in the *van* operon and likely originated in soil bacteria. Several variants of the operon have been described, but VanA is the most common in clinical isolates in the United States, Latin America, and Europe, whereas VanB isolates are more frequent in Australia. Two enterococcal species, *E. gallinarum* and *E. casseliflavus*, have intrinsic low-level resistance to vancomycin due to the presence of the VanC operon in the chromosome.

High-level resistance to aminoglycosides (of which gentamicin and streptomycin are the only two tested by clinical laboratories) abolishes the synergism observed between cell wall–active agents and the aminoglycoside. This important phenotype is routinely sought by the clinical laboratory in isolates from serious infections (Tables 154-1 and 154-2). Genes encoding aminoglycoside-modifying enzymes are usually the cause of high-level resistance to these compounds and are widely disseminated among enterococci, decreasing the options for the treatment of severe enterococcal infections. Additionally, ribosomal methyltransferases, enzymes that methylate rRNA and, as a consequence, disrupt the binding site for aminoglycosides, also can lead to high-level resistance.

Resistance to daptomycin has now been well documented in both *E. faecalis* and *E. faecium*. Daptomycin exerts its action by complexing with calcium and binding to phosphatidylglycerol in the bacterial membrane. After binding, daptomycin forms oligomers, with recent data suggesting that it displaces enzymes important for cell envelope synthesis (MurG and PlsX) and that it can form a complex with lipid II molecules critical for cell-wall synthesis, among other effects on the membrane. Resistance to this antibiotic in enterococci arises via two main pathways. The first involves mutations in genes that coordinate the cell-wall and cell-membrane stress response, most commonly a three-component system designated LiaFSR (for *l*ipid II *i*nterfering *a*ntibiotics). These mutations lead to activation of the system, with increased expression of an extracellular protein known as LiaX capable of binding daptomycin and enhancing the signaling response. In clinical isolates, mutations in LiaFSR may lead to tolerance (loss of bactericidal activity)—usually in isolates with MICs near the daptomycin breakpoint (i.e., 3–4 mg/L). The second pathway involves changes in genes involved in phospholipid metabolism. It is thought that mutations priming the stress response system occur first, with the subsequent accrual of phospholipid changes leading to a fully resistant phenotype. Prior exposure to daptomycin has been identified as a risk factor for the emergence of daptomycin-resistant *E. faecium* in cancer patients. Resistance in the absence of exposure to the drug has also been well described, possibly due to the similarity of this antibiotic to antimicrobial peptides of the innate immune system. Thus, careful consideration of patient characteristics, bacterial phenotype, and daptomycin dose is warranted, and it is advisable to obtain infectious diseases consultation in complicated VRE infections.

The oxazolidinones (linezolid and tedizolid) act by binding to the ribosome and inhibiting the binding of aminoacyl-tRNAs, thus preventing protein synthesis. Resistance to this class of antibiotics is usually due to alterations of the binding site, either via mutations in the 23S rRNA genes or via the presence of an rRNA methylase. Since enterococci carry multiple copies of the gene encoding the 23S rRNA, prolonged exposure to oxazolidinones can select for increasing levels of resistance by favoring propagation of the resistance allele via recombination. Changes in accessory ribosomal proteins have also been associated with linezolid resistance and may act to mitigate the fitness defects of mutations in the rRNA. More concerning is the emergence of transferable resistance genes, which can readily move between enterococcal strains. Several of these genes were first recognized in bacterial isolates of animal origin, likely under the selective pressure of antibiotics such as florfenicol. The *cfr* (*c*hloramphenicol-*f*lorfenicol *r*esistance) gene encodes an rRNA methylase that modifies the 23S rRNA, leading to increases in the MIC of linezolid. Tedizolid tends to exhibit lower MICs in the presence of *cfr*; however, animal models suggest that some variants of the enzyme may compromise the activity of this drug. Two other transmissible resistance genes, *optrA* and *poxtA*, encode a ribosomal protection factor that has been implicated in linezolid resistance in enterococcal strains of human and animal origin. While still relatively rare to encounter in clinical practice, these determinants have been identified across the globe and could be an emerging source of resistance.

Newer tetracycline agents, such as tigecycline, omadacycline, and eravacycline, retain activity in the presence of typical tetracycline resistance determinants, including drug efflux pumps and ribosomal protection factors. However, resistance has been documented and appears to be related to changes in the S10 ribosomal protein, which is situated near the binding site for the drug.

Acknowledgment

The authors dedicate this chapter to the memory of Dr. Barbara Murray, a trailblazer in infectious diseases and a fearless leader. Dr. Murray was an exceptional scientist and an avid adventurer. She will be remembered as an everlasting mentor who touched many lives with her wisdom. Her indelible legacy will remain with us forever.

FURTHER READING

Berge A et al: The DENOVA score efficiently identifies patients with monomicrobial *Enterococcus faecalis* bacteremia where echocardiography is not necessary. Infection 47:45, 2019.

Contreras GA et al: Contemporary clinical and molecular epidemiology of vancomycin-resistant enterococcal bacteremia: A prospective multicenter cohort study (VENOUS I). Open Forum Infect Dis 9: ofab616, 2021.

Lebreton F et al: Tracing the enterococci from Paleozoic origins to the hospital. Cell 169:849, 2017.

Rogers R, Rice LB: State-of-the-art review: Persistent enterococcal bacteremia. Clin Infect Dis 78:e1, 2024.

Satlin MJ et al: Development of daptomycin susceptibility breakpoints for *Enterococcus faecium* and revision of the breakpoints for other enterococcal species by the Clinical and Laboratory Standards Institute. Clin Infect Dis 70:1240, 2020.

155 Diphtheria and Other Corynebacterial Infections

William R. Bishai, John R. Murphy

DIPHTHERIA

Diphtheria is a nasopharyngeal and skin infection caused by *Corynebacterium diphtheriae*. Toxigenic strains of *C. diphtheriae* produce a protein toxin that causes systemic toxicity, myocarditis, and polyneuropathy. The toxin is associated with the formation of pseudomembranes in the pharynx during respiratory diphtheria. While toxigenic strains most frequently cause pharyngeal diphtheria, nontoxigenic strains commonly cause cutaneous disease.

ETIOLOGY

C. diphtheriae is a gram-positive bacillus that is unencapsulated, nonmotile, and nonsporulating. The organism was first identified microscopically in 1883 by Klebs and a year later was isolated in pure culture by Löffler in Robert Koch's laboratory. The bacteria have a characteristic club-shaped bacillary appearance and typically form clusters of parallel rays, or *palisades*, that are referred to as "Chinese characters." The specific laboratory media recommended for the cultivation of *C. diphtheriae* rely upon tellurite, colistin, or nalidixic acid for the organism's selective isolation from other autochthonous pharyngeal microbes. *C. diphtheriae* may be isolated from individuals with both nontoxigenic (*tox*$^-$) and toxigenic (*tox*$^+$) phenotypes. Uchida and Pappenheimer demonstrated that corynebacteriophage beta carries the structural gene *tox*, which encodes diphtheria toxin, and that a family of closely related corynebacteriophages are responsible for toxigenic conversion of *tox*$^-$ *C. diphtheriae* to the *tox*$^+$ phenotype. Moreover, lysogenic conversion from a nontoxigenic to a toxigenic phenotype has been shown to occur in situ. Growth of toxigenic strains of *C. diphtheriae* under iron-limiting conditions leads to the optimal expression of diphtheria toxin and is believed to be a pathogenic mechanism during human infection. Less commonly, diphtheria-like disease may be caused by *Corynebacterium ulcerans* and *Corynebacterium pseudotuberculosis*, which express the same toxin and are considered members of the *C. diphtheriae* group (discussed below).

EPIDEMIOLOGY

While in many areas diphtheria has been controlled in recent years with effective vaccination, there have been sporadic outbreaks throughout the 1970s in the United States and the 1990s in Europe. Diphtheria is still common in parts of Africa, Asia, Latin America, and the Caribbean where mass immunization programs are not enforced. Large-scale epidemics of diphtheria have occurred in the post–Soviet Union independent states in the late 1990s and, more recently, in Nigeria and Yemen during 2022–2023. In temperate regions, respiratory diphtheria occurs year-round but is most common during winter months.

C. diphtheriae is transmitted via the aerosol route, usually during close contact with an infected person. Untreated individuals with respiratory diphtheria are thought to be infectious for ~18.5 days and the R_0 (basic reproduction number) is 1.7–4.3. There are no significant reservoirs other than humans. The mean incubation period for respiratory diphtheria is 1.4 days, but disease onset has occurred as late as 10 days after exposure. Prior to the vaccination era, most individuals over the age of 10 were immune to *C. diphtheriae*; infants were protected by maternal IgG antibodies but became susceptible after ~6 months of age. Thus, the disease primarily affected children and nonimmune young adults.

Cutaneous diphtheria is usually a secondary infection that follows a primary skin lesion due to trauma, allergy, or autoimmunity. Most often, these isolates lack the *tox* gene and thus do not express diphtheria toxin. In tropical latitudes, cutaneous diphtheria is more common than respiratory diphtheria. Toxin-producing diphtheria in symptomatic individuals, regardless of site (respiratory or cutaneous), is a reportable disease in the United States, while nontoxigenic disease is not. Nontoxigenic strains of *C. diphtheriae* have been associated with pharyngitis in Europe, causing outbreaks among men who have sex with men and persons who use illicit IV drugs.

The development of diphtheria antitoxin in 1898 by von Behring and of the diphtheria toxoid vaccine in 1924 by Ramon led to the near elimination of diphtheria in Western countries. The annual incidence rate in the United States peaked in 1921, with 206,000 cases (191 cases per 100,000) and 15,520 deaths. In contrast, current U.S. rates are exceeding low, with only 14 cases reported from 1996–2018, with the last case of respiratory diphtheria occurring in 2003 in a returning traveler from Haiti. Nevertheless, pockets of colonization persist in North America, and groups or individuals who resist vaccination remain at risk. Immunity to diphtheria induced by childhood vaccination gradually decreases in adulthood. An estimated 30% of men 60–69 years old have antitoxin titers below the protective level. In addition to older age and lack of vaccination, risk factors for diphtheria outbreaks include alcoholism, low socioeconomic status, crowded living conditions, and Native American ethnic background. An outbreak of diphtheria in Seattle, Washington, between 1972 and 1982 comprised 1100 cases, most of which were cutaneous. During the 1990s in the states of the former Soviet Union, a much larger diphtheria epidemic included more than 140,000 cases and more than 4000 deaths; at its peak in 1995, more than 50,412 cases were reported. Clonally related toxigenic *C. diphtheriae* strains of the ET8 complex were associated with this outbreak. Beginning in 1998, this epidemic was controlled by mass vaccination programs, and between 2000 and 2009, the diphtheria incidence fell by >95%, with high-burden countries such as Latvia reporting fewer than 10 cases.

Despite the World Health Organization (WHO) estimate that ~84% of the global population of children have been adequately vaccinated, 8638 diphtheria cases were reported globally in 2021, and many more cases are likely to have gone unreported. The recent coronavirus disease 2019 (COVID-19) pandemic, socioeconomic instability, migration, vaccine hesitancy, and other factors remain as threats to diphtheria control. For example, the WHO reported that the percentage of children who received three doses of diphtheria, tetanus, and pertussis immunization fell by 5% (86% to 81%) during 2019–2021 due to health system strains from COVID-19. Additionally, multiple European nations have reported high rates of diphtheria (often cutaneous disease) among unvaccinated asylum seekers, with West Africa and Afghanistan as common source regions.

PATHOGENESIS AND IMMUNOLOGY

Diphtheria toxin produced by *tox*$^+$ strains of *C. diphtheriae* is the primary virulence factor in clinical disease. The toxin is synthesized in precursor form; is released as a 535-amino-acid, single-chain protein; and, in sensitive species (e.g., guinea pigs and humans, but not mice or rats), has a 50% lethal dose of ~100 ng/kg of body weight. The toxin is produced in the pseudomembranous lesion in the pharynx and is taken up in the bloodstream with subsequent distribution to all organs. Once bound to its cell surface receptor (a heparin-binding epidermal growth factor–like precursor), the toxin is internalized by receptor-mediated endocytosis and enters the cytosol from an acidified early endosomal compartment. In vitro, the toxin may be separated into two chains by digestion with serine proteases: the N-terminal A fragment and the C-terminal B fragment. Delivery of the A fragment into the eukaryotic cell cytosol results in irreversible inhibition of protein synthesis by NAD$^+$-dependent ADP-ribosylation of elongation factor 2 and subsequent cell death.

In 1926, Ramon at the Institut Pasteur found that formalinization of diphtheria toxin resulted in the production of a nontoxic but highly immunogenic diphtheria toxoid. Subsequent studies showed that immunization with diphtheria toxoid elicited antibodies that neutralized the toxin and prevented most disease manifestations. In the 1930s, mass immunization of children and susceptible adults with diphtheria toxoid commenced in the United States and Europe.

Individuals with a diphtheria antitoxin titer of >0.01 U/mL are at low risk of disease. In populations where a majority of individuals have protective antitoxin titers, the carrier rate for toxigenic strains of *C. diphtheriae* decreases, and the overall risk of diphtheria among susceptible individuals is reduced. Nevertheless, individuals with nonprotective titers may contract diphtheria either through travel or exposure to individuals who have recently returned from regions where the disease is endemic.

Characteristic pathologic findings of diphtheria include mucosal ulcers with a pseudomembranous coating composed of an inner band of fibrin and a luminal band of neutrophils. Initially white and firmly adherent, in advanced diphtheria the pseudomembranes turn gray or even green or black as necrosis progresses. Mucosal ulcers result from toxin-induced necrosis of the epithelium accompanied by edema, hyperemia, and vascular congestion of the submucosal base. A significant fibrinosuppurative exudate from the ulcer develops into the pseudomembrane. Ulcers and pseudomembranes in severe respiratory diphtheria may extend from the pharynx into medium-sized bronchial airways. Expanding and sloughing membranes may result in fatal airway obstruction.

APPROACH TO THE PATIENT

Diphtheria

Diphtheria, although rare in the United States and other developed countries, should be considered when a patient has severe pharyngitis, particularly when there is difficulty swallowing, respiratory compromise, or signs of systemic disease (e.g., myocarditis or generalized weakness). The leading causes of pharyngitis are respiratory viruses (rhinoviruses, influenza viruses, parainfluenza viruses, coronaviruses, adenoviruses; ~25% of cases), group A streptococci (15–30%), group C streptococci (~5%), atypical bacteria such as *Mycoplasma pneumoniae* and *Chlamydia pneumoniae* (15–20% in some series), and other viruses such as herpes simplex virus (~4%) and Epstein-Barr virus (<1% in infectious mononucleosis). Less common causes are acute HIV infection, gonorrhea, fusobacterial infection (e.g., Lemierre's syndrome), thrush due to *Candida albicans* or other *Candida* species, and diphtheria. The presence of a pharyngeal pseudomembrane or an extensive exudate should prompt consideration of diphtheria (Fig. 155-1).

CLINICAL MANIFESTATIONS

Respiratory Diphtheria The clinical diagnosis of diphtheria is based on the constellation of sore throat; adherent tonsillar, pharyngeal, or nasal pseudomembranous lesions; and low-grade fever. In addition, diagnosis requires the isolation of *C. diphtheriae* or histopathologic isolation of compatible gram-positive organisms. The Centers for Disease Control and Prevention (CDC) recognizes *confirmed* respiratory diphtheria (laboratory proven or epidemiologically linked to a culture-confirmed case) and *probable* respiratory diphtheria (clinically compatible but not laboratory proven or epidemiologically linked). Carriers are defined as individuals who have positive cultures for *C. diphtheriae* and who either are asymptomatic or have symptoms but lack pseudomembranes. Most patients seek medical care for sore throat and fever several days into the illness. Occasionally, weakness, dysphagia, headache, and voice change are the initial manifestations. Neck edema and difficulty breathing are evident in more advanced cases and carry a poor prognosis.

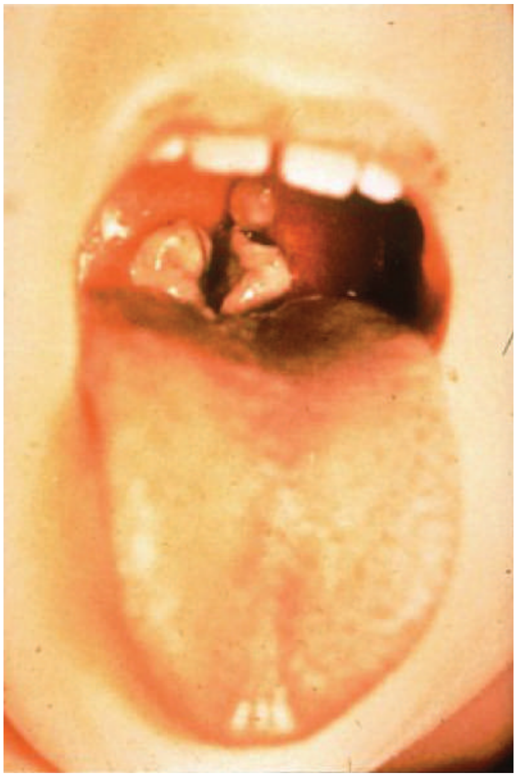

FIGURE 155-1 Respiratory diphtheria due to toxigenic *C. diphtheriae* producing exudative pharyngitis in a child displaying a pseudomembrane extending from the uvula to the pharyngeal wall. The characteristic white pseudomembrane is caused by diphtheria toxin–mediated necrosis of the respiratory epithelial layer, producing a fibrinous coagulative exudate. Submucosal edema adds to airway narrowing. The pharyngitis is acute in onset, and respiratory obstruction from the pseudomembrane may occur in severe cases. Inoculation of pseudomembrane fragments or submembranous swabs onto Löffler's or tellurite selective medium reveals *C. diphtheriae. (Photograph courtesy of the Centers for Disease Control and Prevention and Immunization Action Coalition, used by permission.)*

The systemic manifestations of diphtheria stem from the effects of diphtheria toxin and include weakness as a result of neurotoxicity and cardiac arrhythmias or congestive heart failure due to myocarditis. Most commonly, the pseudomembranous lesion is located in the tonsillopharyngeal region. Less commonly, the lesions are located in the larynx, nares, and trachea or bronchial passages. Large pseudomembranes are associated with severe disease and a poor prognosis. A few patients develop massive swelling of the tonsils and present with "bull-neck" diphtheria, which results from edema of the submandibular and paratracheal region and is further characterized by foul breath, thick speech, and stridorous breathing. The diphtheritic pseudomembrane is gray or whitish and sharply demarcated. Unlike the exudative lesion associated with streptococcal pharyngitis, the pseudomembrane in diphtheria is tightly adherent to the underlying tissues. Attempts to dislodge the membrane may cause bleeding. Hoarseness suggests laryngeal diphtheria, in which laryngoscopy may be diagnostically helpful.

Cutaneous Diphtheria This dermatosis is characterized by punched-out ulcerative lesions with necrotic sloughing or pseudomembrane formation (Fig. 155-2). The diagnosis requires cultivation of *C. diphtheriae* from lesions, which most commonly occur on the lower and upper extremities, head, and trunk.

Infections Due to Non-*diphtheriae Corynebacterium* Species and Nontoxigenic *C. diphtheriae* Non-*diphtheriae* species of *Corynebacterium* and related genera (discussed below) as well as nontoxigenic strains of *C. diphtheriae* itself have been found in bloodstream and respiratory infections, often in individuals with immunosuppression or chronic respiratory disease. These organisms can cause disease manifestations and should not necessarily be dismissed as colonizers.

Other Clinical Manifestations *C. diphtheriae* causes rare cases of endocarditis and septic arthritis, most often in patients with preexisting risk factors, such as abnormal cardiac valves, injection drug use, or cirrhosis.

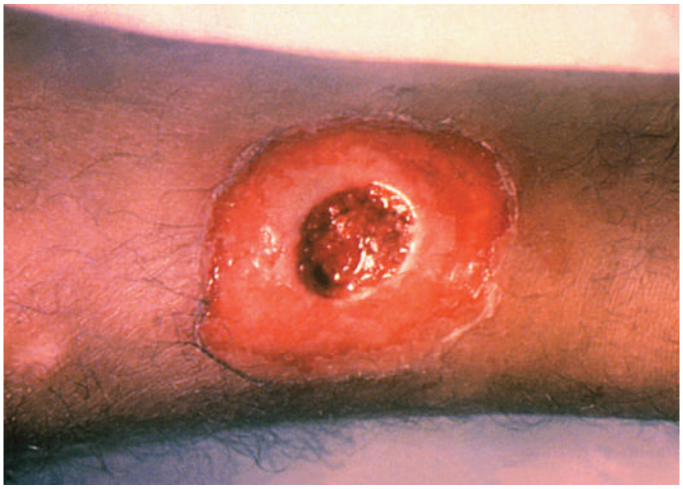

FIGURE 155-2 Cutaneous diphtheria due to nontoxigenic *C. diphtheriae* on the lower extremity. *(From the Centers for Disease Control and Prevention, Public Health Image Library [PHIL]. #1941.)*

COMPLICATIONS

Airway obstruction poses a significant early risk in patients presenting with advanced diphtheria and accounts for 60–65% of deaths typically in the first 1–2 weeks after symptom onset. Pseudomembranes may slough and obstruct the airway or may advance to the larynx or into the tracheobronchial tree. Children are particularly prone to obstruction because of their small airways.

Cardiomyopathy and polyneuropathy are late toxic manifestations of diphtheria arising 1 week or more after respiratory symptoms. Based on systematic reviews, toxic cardiomyopathy accounted for 20–25% of deaths and is typically associated with arrhythmias and dilated cardiomyopathy.

Polyneuropathy is seen 3–5 weeks after the onset of diphtheria and has a slow indolent course but may account for 15% of deaths. Patients may develop severe and prolonged neurologic abnormalities. The disorders typically occur in the mouth and neck, with lingual or facial numbness as well as dysphonia, dysphagia, and cranial nerve paresthesias. More ominous signs include weakness of respiratory and abdominal muscles and paresis of the extremities. Sensory manifestations and sensory ataxia also are observed. Cranial nerve dysfunction typically precedes disturbances of the trunk and extremities because of proximity to the site of infection. Autonomic dysfunction also is associated with polyneuropathy and can lead to hypotension. Polyneuropathy is typically reversible in patients who survive the acute phase.

Other complications of diphtheria include pneumonia, renal failure, encephalitis, cerebral infarction, pulmonary embolism, and serum sickness from antitoxin therapy.

DIAGNOSIS

The diagnosis of diphtheria is based on clinical signs and symptoms plus laboratory confirmation. Respiratory diphtheria should be considered in patients with sore throat, pharyngeal exudates, and fever. Other symptoms may include hoarseness, stridor, or palatal paralysis. The presence of a pseudomembrane should prompt strong consideration of diphtheria. Once a clinical diagnosis of diphtheria is made, diphtheria antitoxin should be obtained and administered as rapidly as possible.

Laboratory diagnosis of diphtheria is based either on cultivation of *C. diphtheriae* or toxigenic *C. ulcerans* from the site of infection or on the demonstration of local lesions with characteristic histopathology. *Corynebacterium pseudodiphtheriticum*, a nontoxigenic organism, is a common component of the normal throat flora and does not pose a significant risk. Throat samples should be submitted to the laboratory for culture with the notation that diphtheria is being considered. This information should prompt cultivation on special selective medium and subsequent biochemical testing to differentiate *C. diphtheriae* from other nasopharyngeal commensal corynebacteria. All laboratory isolates of toxigenic *C. diphtheriae* should be reported to the state health department.

A diagnosis of cutaneous diphtheria requires laboratory confirmation since the lesions are not characteristic and are indistinguishable from other dermatoses. Diphtheritic ulcers occasionally—but not consistently—have a punched-out appearance (Fig. 155-2). Patients in whom cutaneous diphtheria is identified should have the nasopharynx cultured for *C. diphtheriae*. The laboratory medium for cutaneous diphtheria specimens is the same as that used for respiratory diphtheria: Löffler's or Tinsdale's selective medium in addition to nonselective medium such as blood agar. As has been mentioned, isolation of toxigenic strains of *C. diphtheriae* in symptomatic individuals is notifiable disease in the United States, regardless of the body site of origin.

TREATMENT

Diphtheria

DIPHTHERIA ANTITOXIN

Prompt administration of diphtheria antitoxin is critical in the management of respiratory diphtheria. Diphtheria antitoxin, a horse antiserum, is effective in reducing the extent of local disease as well as the risk of complications of myocarditis and neuropathy. Rapid institution of antitoxin therapy is also associated with a significant reduction in mortality risk. Because diphtheria antitoxin cannot neutralize cell-bound toxin, prompt initiation is important. This product, which is no longer commercially available in the United States, can be obtained from the CDC Emergency Operations Center at 770-488-7100 (*www.cdc.gov/diphtheria/hcp/dat/index.html*) after first contacting the state health department. The current protocol for the use of diphtheria antitoxin involves a test dose to rule out immediate hypersensitivity. Patients who demonstrate hypersensitivity require desensitization before a full therapeutic dose of antitoxin is administered.

Given that the world supply of equine anti–diphtheria toxin is limited, a human monoclonal antibody with the potential to provide a safer alternative to equine antitoxin therapy is being developed.

ANTIMICROBIAL THERAPY

Antibiotics are used in the management of diphtheria primarily to prevent transmission to susceptible contacts. Antibiotics also prevent further toxin production and may reduce the severity of local infection. Recommended treatment options for patients with respiratory diphtheria are as follows:

- Erythromycin, 500 mg IV q6h (for children: 40–50 mg/kg per day IV in two or four divided doses) until the patient can swallow comfortably; then 500 mg PO qid to complete a 14-day course
- Procaine penicillin G, 600,000 U IM q12h (for children: 12,500–25,000 U/kg IM q12h) until the patient can swallow comfortably; then oral penicillin V, 125–250 mg qid to complete a 14-day course

A clinical study in Vietnam found that penicillin was associated with a more rapid resolution of fever and a lower rate of bacterial resistance than erythromycin; however, relapses were more common in the penicillin group. Erythromycin therapy targets protein synthesis and thus offers the presumed benefit of stopping toxin synthesis more quickly than a cell wall–active β-lactam agent. Alternative therapeutic agents for patients who are allergic to penicillin or cannot take erythromycin include rifampin and clindamycin. Other reasonable antibiotics are clarithromycin, azithromycin, linezolid, and vancomycin, although they have not been studied in comparison to the agents above.

Eradication of *C. diphtheriae* should be documented after antimicrobial therapy is complete. A repeat throat culture 2 weeks later is recommended. For patients in whom the organism is not eradicated after a 14-day course of erythromycin or penicillin, an additional 10-day course followed by repeat culture is recommended. Drug-resistant strains of *C. diphtheriae* exist and are appearing at higher frequency; several reports have described multidrug-resistant strains. Drug resistance should be considered when efforts at pathogen eradication fail.

Cutaneous diphtheria should be treated as described above for respiratory disease. Individuals infected with toxigenic strains should receive antitoxin. It is important to treat the underlying cause of the dermatoses in addition to the superinfection with *C. diphtheriae*.

Patients who recover from respiratory or cutaneous diphtheria should have antitoxin levels measured. If diphtheria antitoxin has been administered, this test should be performed 6 months later. Patients who recover from respiratory or cutaneous diphtheria should receive the appropriate vaccine to ensure the development of protective antibody titers.

MANAGEMENT STRATEGIES

Patients in whom diphtheria is suspected should be hospitalized in respiratory isolation rooms with close monitoring of cardiac and respiratory function. A cardiac workup is recommended to assess the possibility of myocarditis. In patients with extensive pseudomembranes, an anesthesiology or an ear, nose, and throat consultation is recommended because of the possible need for tracheostomy or intubation. In some settings, pseudomembranes can be removed surgically. Treatment with glucocorticoids has not been shown to reduce the risk of myocarditis or polyneuropathy.

PROGNOSIS

A systematic review of over 20 reported outbreaks found the diphtheria case fatality ratio among unvaccinated, untreated individuals to be 29%, with children under the age of 5 being at a 1.5-fold higher risk of mortality. Fatal pseudomembranous diphtheria typically occurs in patients with nonprotective antibody titers and in unimmunized patients. The pseudomembrane may actually increase in size from the time it is first noted. Risk factors for death include bullneck diphtheria; myocarditis with ventricular tachycardia; atrial fibrillation; complete heart block; an age of >60 years or <6 months; alcoholism; extensive pseudomembrane elongation; and laryngeal, tracheal, or bronchial involvement. Another important predictor of fatal outcome is the interval between the onset of local disease and the administration of antitoxin. Cutaneous diphtheria has a low mortality rate and is rarely associated with myocarditis or peripheral neuropathy.

PREVENTION

Vaccination Diphtheria toxoid-based vaccine efficacy is estimated to be 87% in the prevention of symptomatic disease, and sustained campaigns for vaccination of children and adequate boosting vaccination of adults are responsible for the exceedingly low incidence of diphtheria in most developed nations. Diphtheria toxoid vaccine has typically been coadministered with tetanus vaccine (with or without acellular pertussis). DTaP (full-level diphtheria toxoid, tetanus toxoid, and acellular pertussis vaccine) is currently recommended for children in five doses up to the age of 6 years; DTaP replaced the earlier whole-cell pertussis vaccine DTP in 1997. Tdap is a tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis vaccine formulated for adolescents and adults and is the recommended booster for children 11–12 years old. Tdap is recommended for all adults if they have not received it previously regardless of the interval since the last dose of Td (tetanus and reduced-dose diphtheria toxoids, adsorbed). Tdap vaccination is a priority for health care workers, pregnant women, adults anticipating contact with infants, and adults not previously vaccinated for pertussis. Adults who have received acellular pertussis vaccine should continue to receive decennial Td booster vaccinations. In 2018, a hexavalent vaccine that combined diphtheria-tetanus toxoids, acellular pertussis adsorbed (DTaP), inactivated poliovirus (IPV), *Haemophilus influenzae* type b (Hib) conjugate, and recombinant hepatitis B (HepB) known as DTaP-IPV-Hib-HepB was approved by the U.S. Food and Drug Administration; this product may be used to replace the initial three childhood doses of DTaP (2, 4, and 6 months) or as part of a catch-up schedule in children under 5 years old. **The vaccine schedule is detailed in Chap. 129.**

Prophylaxis Administration to Contacts Close contacts of diphtheria patients should undergo throat culture to determine whether they are carriers. After samples for throat culture are obtained, antimicrobial prophylaxis should be considered for all contacts, even those whose cultures are negative. The options are 7–10 days of oral erythromycin or one dose of IM benzathine penicillin G (1.2 million units for persons ≥6 years of age or 600,000 units for children <6 years of age).

Contacts of diphtheria patients whose immunization status is uncertain should receive the appropriate diphtheria toxoid–containing vaccine. The Tdap vaccine (rather than Td) is now the booster vaccine of choice for adults who have not recently received an acellular pertussis–containing vaccine. Carriers of *C. diphtheriae* in the community should be treated and vaccinated when identified.

OTHER CORYNEBACTERIAL AND *RHODOCOCCUS* INFECTIONS

Nondiphtherial corynebacteria, referred to as *diphtheroids* or *coryneforms*, are frequently considered colonizers or contaminants; however, they have been associated with invasive disease, particularly in immunocompromised patients. Importantly, even though they are termed nondiphtherial corynebacteria, *C. ulcerans* and *C. pseudotuberculosis* may produce diphtheria toxin and therefore cause severe human illness. These organisms have been isolated from the bloodstream, especially in association with catheter infection, endocarditis, prosthetic valve infection, meningitis, brain abscess, osteomyelitis, and peritonitis. Risk factors include indwelling intravenous or peritoneal catheters and neurosurgical shunts. Patients infected with these organisms are often immunosuppressed or have significant medical comorbidities. The nondiphtherial coryneforms are a collection of bacteria that are taxonomically grouped together in the genus *Corynebacterium* on the basis of their 16S rDNA signature nucleotides. Despite the shared rDNA signatures, these isolates are quite diverse. For example, their guanine-cytosine content ranges from 45 to 70%. Several nondiphtheroid corynebacteria, including *Corynebacterium jeikeium* and *Corynebacterium urealyticum*, are associated with resistance to multiple antibiotics. *Rhodococcus equi* is associated with necrotizing pneumonia and granulomatous infection, particularly in immunocompromised individuals.

MICROBIOLOGY AND LABORATORY DIAGNOSIS

These organisms are non-acid-fast, catalase-positive, aerobic or facultatively anaerobic rods. Their colonial morphologies on blood agar vary widely; some species are small and α-hemolytic (similar to lactobacilli), whereas others form large white colonies (similar to yeasts). Many nondiphtherial coryneforms require special media, such as Löffler's, Tinsdale's, or tellurite medium. These cultivation idiosyncrasies have led to a complex taxonomic categorization of the organisms.

EPIDEMIOLOGY

Humans are the natural reservoirs for several nondiphtherial coryneforms, including *C. xerosis, C. pseudodiphtheriticum*, *C. striatum*, *C. minutissimum*, *C. jeikeium*, *C. urealyticum*, and *Arcanobacterium haemolyticum*. Animal reservoirs including milk are responsible for carriage of *C. ulcerans* and *C. pseudotuberculosis*. Soil is the natural reservoir for *R. equi*.

CLINICAL MANIFESTATIONS

C. ulcerans This organism causes a diphtheria-like illness and produces both diphtheria toxin and a dermonecrotic toxin. The organism is a commensal in horses and cattle and has been isolated from cow's milk. In contrast to diphtheria, this infection is considered a zoonosis, and cases have been traced to contact with animal carriers, including dogs and pigs. *C. ulcerans* causes exudative pharyngitis, primarily during summer months, in rural areas, and among individuals exposed to animals. Treatment with antitoxin and antibiotics should be initiated when respiratory *C. ulcerans* is identified, and a contact investigation should be conducted (including throat cultures to determine the need for antimicrobial prophylaxis and, in unimmunized contacts, administration of the appropriate diphtheria toxoid–containing vaccine). The organism grows on Löffler's, Tinsdale's, and tellurite agars as well as blood agar. In addition to exudative pharyngitis, cutaneous disease due to *C. ulcerans* has been reported. *C. ulcerans* is susceptible to a wide panel of antibiotics. Erythromycin and other macrolides appear to be the first-line agents.

C. pseudotuberculosis Infection caused by *C. pseudotuberculosis* is an important animal pathogen (most notably of sheep) that rarely causes human disease. *C. pseudotuberculosis* causes suppurative granulomatous lymphadenitis and an eosinophilic pneumonia syndrome among individuals who handle sheep; horses, cattle, goats, deer, and raw milk has also been implicated. Surgical excision of affected lymph nodes should be performed when feasible, and successful treatment with erythromycin or tetracycline has been reported. Some strains express diphtheria toxin and produce a diphtheria-like disease, which should be treated with antitoxin.

***C. jeikeium* (Group JK)** Originally described in American hospitals, *C. jeikeium* infection was subsequently reported in Europe. After a 1976 survey of diseases caused by nondiphtherial corynebacteria, CDC group JK emerged as an important opportunistic pathogen among neutropenic and HIV-infected patients. The organism has now been designated a separate species. *C. jeikeium* forms small, gray to white, glistening, nonhemolytic colonies on blood agar. It lacks urease and nitrate reductase and does not ferment most carbohydrates. The predominant syndrome associated with *C. jeikeium* is sepsis, sometimes with associated pneumonia, endocarditis, meningitis, osteomyelitis, or epidural abscess. Risk factors for *C. jeikeium* infection include hematologic malignancy, neutropenia from comorbid conditions, prolonged hospitalization, exposure to multiple antibiotics, and skin disruption. There is evidence that *C. jeikeium* is part of the inguinal, axillary, genital, and perirectal flora of hospitalized patients.

Broad-spectrum antimicrobial therapy appears to select for colonization. The organisms appear as gram-positive coccobacillary forms slightly resembling streptococci. *C. jeikeium* is resistant to the majority of antibiotic classes except oxazolidinones (e.g., linezolid) and glycopeptides (e.g., vancomycin). Effective therapy involves removal of the infectious source, whether a catheter, prosthetic joint, or prosthetic valve. Efforts have been made to prevent *C. jeikeium* infection with strict institution of infection control protocols for high-risk patients, particularly those in intensive care units.

***C. urealyticum* (Group D2)** Identified as a urease-positive nondiphtherial *Corynebacterium* in 1972, *C. urealyticum* is an opportunistic pathogen causing sepsis and urinary tract infection. *C. urealyticum* appears to be the etiologic agent of a severe urinary tract syndrome known as *alkaline-encrusted cystitis*, a chronic inflammatory bladder infection associated with deposition of ammonium magnesium phosphate on the surface and walls of ulcerating lesions in the bladder. In addition, *C. urealyticum* has been associated with pneumonia, peritonitis, endocarditis, osteomyelitis, and wound infection. It is similar to *C. jeikeium* in its resistance to most antibiotics except oxazolidinones and glycopeptides. Vancomycin therapy has been used successfully in severe infections.

***C. minutissimum* (Erythrasma)** Erythrasma is a cutaneous infection producing reddish-brown, macular, scaly, pruritic intertriginous patches. The dermatologic presentation under the Wood's lamp is of coral red fluorescence. *C. minutissimum* appears to be a common cause of erythrasma, although there is evidence for a polymicrobial etiology in certain settings. This microbe has also been associated with bacteremia in patients with hematologic malignancy. Erythrasma responds to topical erythromycin, clarithromycin, clindamycin, or fusidic acid, although more severe infections may require oral macrolide therapy.

Other Nondiphtherial Corynebacteria *C. xerosis* is a human commensal found in the conjunctiva, nasopharynx, and skin. This nontoxigenic organism is occasionally identified as a source of invasive infection in immunocompromised or postoperative patients and prosthetic joint recipients. *C. amycolatum* is a closely related species but tends to demonstrate more antibiotic resistance. *C. striatum* is found in the anterior nares, skin, face, and upper torso of healthy individuals. Also nontoxigenic, this organism has been associated with invasive opportunistic infections in severely ill or immunocompromised patients. *C. glucuronolyticum* is a nonlipophilic species that causes male genitourinary tract infections such as prostatitis and urethritis. These infections may be successfully treated with a wide variety of antibacterial agents, including β-lactams, rifampin, aminoglycosides, or vancomycin; however, the organism appears to be resistant to fluoroquinolones, macrolides, and tetracyclines. *C. imitans* has been identified in eastern Europe as a nontoxigenic cause of pharyngitis. *C. auris* has been identified in children with otitis media; it is susceptible to fluoroquinolones, rifampin, tetracycline, and vancomycin but resistant to penicillin G and variably susceptible to macrolides. *C. pseudodiphtheriticum* is a nontoxigenic species that is part of the normal human flora. Human infections—particularly endocarditis of either prosthetic or natural valves and invasive pneumonia—have been reported only rarely. Although *C. pseudodiphtheriticum* may be isolated from the nasopharynx of patients with suspected diphtheria, it is part of the normal flora and does not produce diphtheria toxin. *C. propinquum*, a close relative of *C. pseudodiphtheriticum*, is part of CDC group D-1 and has been isolated from the human respiratory tract and blood. *C. afermentans* and subspecies belong to CDC group ANF-1; it is a rare human pathogen that has been isolated from human blood and abscesses.

Rhodococcus *Rhodococcus* species are phylogenetically related to the corynebacteria. These gram-positive coccobacilli have been associated with tuberculosis-like infections in humans with granulomatous pathology. While *R. equi* is best known, other near-relative species have been identified in human infections including *R. fascians*, *R. erythropolis*, *R. rhodochrous*, *Gordonia bronchialis*, *G. sputi*, *G. terrae*, and *Tsukamurella paurometabola*.

R. equi has been recognized as a cause of pneumonia in horses since the 1920s and as a cause of related infections in cattle, sheep, and swine. It is found in soil as an environmental microbe. The organisms vary in length; appear as spherical to long, curved, clubbed rods; and produce large irregular mucoid colonies. *R. equi* cannot ferment carbohydrates or liquefy gelatin and is often acid fast. An intracellular pathogen of macrophages, *R. equi* can cause granulomatous necrosis and caseation. This organism has most commonly been identified in pulmonary infection, but infections of brain, bone, and skin also have been reported. Most commonly, *R. equi* disease manifests as nodular and/or cavitary pneumonia of the upper lobe—a picture similar to that seen in tuberculosis or nocardiosis. Most patients are immunocompromised, often by HIV infection. Subcutaneous nodular lesions also have been identified. The involvement of *R. equi* should be considered when any patient presents with a tuberculosis-like syndrome.

Infection due to *R. equi* has been treated successfully with antibiotics that penetrate intracellularly, including macrolides, clindamycin, rifampin, and trimethoprim-sulfamethoxazole. β-Lactam antibiotics have not been useful. The organism is routinely susceptible to vancomycin, which is considered the drug of choice.

Arcanobacteria *Arcanobacterium haemolyticum* was identified as an agent of wound infections in U.S. soldiers in the South Pacific during World War II. It appears to be a human commensal of the nasopharynx and skin, but it is known to cause true pharyngitis as well as chronic skin ulcers. In contrast to the much more common pharyngitis caused by *Streptococcus pyogenes*, *A. haemolyticum* pharyngitis is associated with a scarlatiniform rash on the trunk and proximal extremities in about half of cases; this illness is occasionally confused with toxic shock syndrome. Because *A. haemolyticum* pharyngitis primarily affects teenagers, it has been postulated that the rash–pharyngitis syndrome may represent co-pathogenicity, synergy, or opportunistic secondary infection with Epstein-Barr virus. *A. haemolyticum* has also been reported as a cause of bacteremia, soft tissue infections, osteomyelitis, and cavitary pneumonia, predominantly in the setting of underlying diabetes mellitus. The organism is susceptible to most β-lactams, macrolides, fluoroquinolones, clindamycin, vancomycin, and doxycycline. However, resistance to trimethoprim-sulfamethoxazole as well as tetracycline is common.

FURTHER READING

Kim R, Reboli AC: Other coryneform bacteria and *Rhodococcus*, in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 9th ed. JE Bennett et al (eds). Philadelphia, Elsevier, 2020, pp 2532–2542.

Moore LS et al: *Corynebacterium ulcerans* cutaneous diphtheria. Lancet Infect Dis 15:1100, 2015.

Saleeb PG: *Corynebacterium diphtheriae* (diphtheria), in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 9th ed. JE Bennett et al(eds). Philadelphia, Elsevier, 2020, pp 2526–2531.

Sharma NC et al: Diphtheria. Nat Rev Dis Primers 5:81, 2019.

Truelove SA et al: Clinical and epidemiologic aspects of diphtheria: A systematic review and pooled analysis. Clin Infect Dis 71:89, 2020.

156

Listeria monocytogenes Infections

Karen P. Neil, Jennifer P. Collins,
Patricia M. Griffin

Listeria monocytogenes is a ubiquitous environmental saprophyte and an intracellular pathogen in several animals. Humans develop *L. monocytogenes* infection—listeriosis—primarily through foodborne transmission. The clinical spectrum of listeriosis ranges from febrile gastroenteritis in healthy persons to invasive disease, including bacteremia and meningoencephalitis. The major risk groups for invasive disease are pregnant women and their neonates, older adults, and immunocompromised persons. Febrile gastroenteritis is rarely diagnosed because *Listeria* are not detected by routine stool cultures or in rapid diagnostic tests for stool specimens, so the term listeriosis generally refers to invasive infection.

■ MICROBIOLOGY

L. monocytogenes lives in soil and decaying vegetable matter. Numerous bird and mammal species are reservoirs. *L. monocytogenes* is a nonsporulating, facultatively anaerobic, short, gram-positive rod that grows well on blood agar, demonstrating small zones of β-hemolysis. Organisms sometimes appear gram-variable and resemble cocci, diplococci, or diphtheroids; this appearance can obscure the diagnosis. On light microscopy, *L. monocytogenes* demonstrates characteristic tumbling motility. It grows optimally at 30–37°C but can grow at refrigerator temperatures as low as 4°C. In addition to its ability to grow at cold temperatures, *Listeria*'s tolerance to low-pH and high-salt environments facilitates its environmental survival. Serotypes are usually determined on the basis of somatic (O) and flagellar (H) antigens. Nearly all human illness is caused by serotypes 1/2a, 1/2b, and 4b.

■ PATHOGENESIS

Infection with *L. monocytogenes* typically occurs through ingestion of contaminated food. The infectious dose has not been well established but is likely to be very low for persons with severely impaired cellular immunity. Increased gastric pH, such as that due to proton pump inhibitors, probably promotes the organism's survival in the gastrointestinal tract. After transcytosis across the intestinal epithelium, the bacteria travel via mesenteric lymph nodes and the bloodstream to the liver and spleen, its target organs; dissemination to other organs can occur. *L. monocytogenes* can also migrate across the blood-brain barrier and the placenta.

Virulence factors include internalins (InlA, InlB), which promote invasion into normally nonphagocytic host cells. A pore-forming cytolysin (listeriolysin O; LLO) and phospholipases facilitate evasion of intracellular killing by mediating escape from the internalization vacuole into the cell cytosol. The surface protein ActA facilitates direct cell-to-cell movement within the cytosol, allowing *L. monocytogenes* to avoid encountering components of the host immune system, such as antibodies and complement, during dissemination. Iron promotes growth of the organism in vitro, an effect that explains why listeriosis has been associated with iron-overload conditions, including hemochromatosis.

■ IMMUNE RESPONSE

Although *L. monocytogenes* is ubiquitous in the environment, infection is rare because of both innate and adaptive host immune responses. Studies of mice have contributed to a detailed understanding of the immune response to infection. Activation of innate immunity is important for host survival. Interferon γ and tumor necrosis factor α (TNF-α) are among the key cytokines involved in this response. T cells are the primary drivers of the adaptive immune response, furthering the clearance of infected cells. Cytotoxic (CD8+) T cells are the main contributors to long-term immunity.

These immune mechanisms explain the association between invasive listeriosis and immunocompromising conditions, particularly impaired cellular immunity. In light of numerous reports of invasive listeriosis in patients treated with TNF-α inhibitors, the U.S. Food and Drug Administration added listeriosis to the boxed warning for this drug class. Because *L. monocytogenes* induces a vigorous cell-mediated immune response, attenuated strains that express foreign antigens are undergoing clinical trials as a cancer immunotherapy.

■ EPIDEMIOLOGY

L. monocytogenes was first reported as a human pathogen in 1929. Foodborne transmission was not identified until over 50 years later, when a 1983 outbreak investigation implicated coleslaw. It is now known that *L. monocytogenes* transmission is almost always foodborne. Listeriosis is a nationally notifiable disease in the United States. According to the Centers for Disease Control and Prevention's (CDC's) Foodborne Diseases Active Surveillance Network (FoodNet), the incidence of invasive listeriosis was 2.1–3.3 cases per million persons during 2011–2022 **(Fig. 156-1)**. Although illness is uncommon, listeriosis is a substantial contributor to morbidity from foodborne illness because virtually all patients are hospitalized and 18% die.

Only about 3% of reported cases of listeriosis are part of a recognized outbreak—i.e., have a source determined; however, outbreak investigations provide data on major foods linked to illness. Strains of *L. monocytogenes* have survived in production facilities for years. Refrigerated, ready-to-eat foods are of particular concern because the organism's ability to grow at refrigerator temperatures can result in a small amount of contamination during production growing to much higher levels. PulseNet, the laboratory network for bacterial foodborne disease surveillance, implemented routine subtyping for *L. monocytogenes* in the United States in 1998 and transitioned to whole-genome sequencing in 2013. Improvements in the ability to detect and investigate outbreaks using laboratory-based surveillance has provided more information on known food sources and facilitated recognition of new food sources.

Hot dogs and deli meats were the major sources of U.S. outbreaks until after 2002, when an outbreak linked to turkey deli meat resulted in eight deaths and the recall of >30 million pounds of meat. After that outbreak, the U.S. Department of Agriculture's Food Safety and Inspection Service (FSIS) issued new regulations requiring validated *L. monocytogenes* control programs and intensified testing for *L. monocytogenes* in ready-to-eat meat and poultry plants. Isolations of *L. monocytogenes* from ready-to-eat meats continued to decline during 2003–2012 (Fig. 156-1), and outbreaks linked to these products decreased. However, the incidence of listeriosis, which had declined markedly from 1989 through 2001, has not declined significantly since the measures were implemented (Fig. 156-1). Dairy products remain an important source, especially soft cheese made with raw (unpasteurized) milk or produced from pasteurized milk in unsanitary facilities; ice cream and raw milk also have caused outbreaks. Raw produce has been increasingly linked to outbreaks; implicated items include packaged salad, sprouts, enoki mushrooms, cantaloupe and other fruit, caramel apples, and frozen vegetables.

Most people with invasive listeriosis are older adults, whose risk increases with each decade over 59 years of age. Most other patients have impaired cellular immunity associated with hematologic malignancy, solid organ or bone marrow transplantation, HIV infection, or receipt of glucocorticoid or other immunosuppressive drugs. The group at highest risk of infection is pregnant women, who almost always have only mild flulike symptoms but who transmit the infection to the fetus through the placenta. Some neonates may acquire infection in the hospital, as illustrated by an outbreak associated with contaminated mineral oil. Rarely, children and adults with no recognized risk factors develop invasive listeriosis, probably through highly contaminated food.

The diagnosis of listeriosis in a hospitalized patient with new symptoms should prompt investigation into the food provided during hospitalization as a source. In fact, outbreaks have been traced to food served to hospitalized patients, especially those with immunocompromising

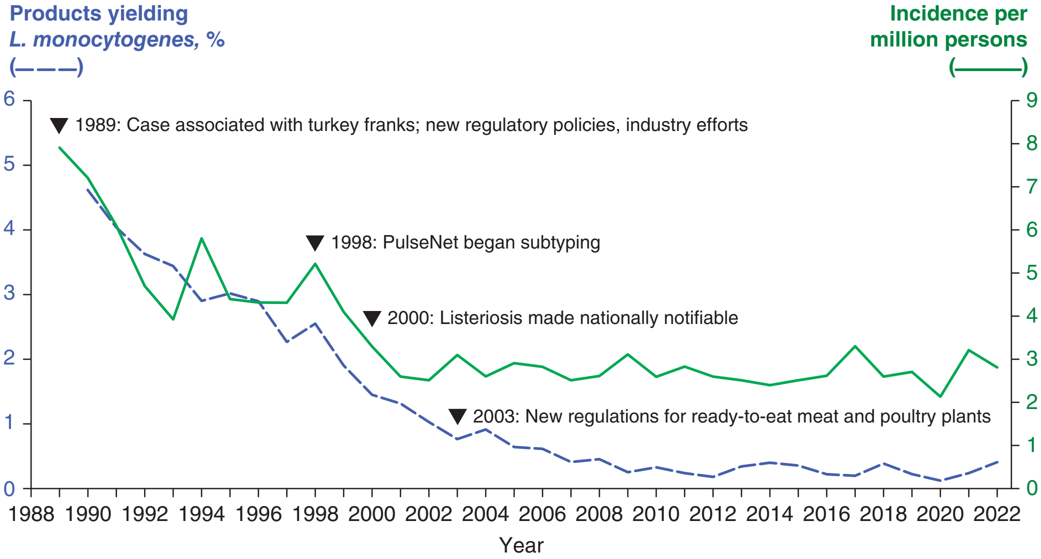

FIGURE 156-1 Incidence of listeriosis and percentage of ready-to-eat meat and poultry products from Food Safety and Inspection Service (FSIS) testing with cultures that yielded *L. monocytogenes*, United States, 1989–2022. Incidence data are from the Centers for Disease Control and Prevention's active sentinel site surveillance and include data from an early surveillance system (1986–1995) and from the Foodborne Diseases Active Surveillance Network (FoodNet) database (1996–2022). The incidence was 7.3 cases per million in 1986. Product data for 1990–2022 are from the U.S. Department of Agriculture Food Safety and Inspection Service Microbiological Testing Program for Ready-to-Eat Meat and Poultry Products and related publications (P Levine et al: J Food Prot 64:1188, 2001; SW Mamber et al: J Food Prot 83:1598, 2020; and the 2010 FSIS Comparative Risk Assessment for *L. monocytogenes* in RTE Meat and Poultry Deli Meats).

conditions; implicated foods include sandwiches, butter, precut celery, Camembert cheese, sausage, tuna salad, and ice cream. Although very rare, transmission of *L. monocytogenes* has also occurred through transfusion of platelet products, and febrile illness with bacteremia has occurred in persons who received an attenuated strain as a vaccine vector.

Listeriosis is an important problem worldwide. A large 2017–2018 outbreak of listeriosis in South Africa, linked to a ready-to-eat processed meat product, disproportionately affected people living with HIV and pregnant women. No outbreak-associated cases were detected in the 15 countries that imported the product; this discrepancy suggests that listeriosis is underrecognized in low- to middle-income countries, particularly those with a high prevalence of HIV infection.

CLINICAL MANIFESTATIONS

L. monocytogenes infection can manifest in several ways. The incubation period differs according to host factors and dose consumed: on average, this interval is <24 h for gastroenteritis and ~11 days for invasive disease, although it can be much longer. Data from outbreak investigations suggest that the incubation period is longer in pregnant women than in nonpregnant adults.

Febrile Gastroenteritis *Listeria* organisms typically pass through healthy people without causing symptoms, but acute febrile gastroenteritis can occur. Outbreak investigations of *L. monocytogenes* febrile gastroenteritis have identified high organism density in implicated foods, suggesting that a large inoculum must be ingested to cause illness. Major manifestations are fever, diarrhea, headache, and constitutional symptoms. Illness is usually self-limited, with symptoms lasting an average of 1–3 days.

Bacteremia Bacteremia without a focus is the most common manifestation of invasive listeriosis. Major features are fever, chills, myalgias, and arthralgias, sometimes preceded by nausea or diarrhea—markers of the initial gut infection. Bacteremia can cause neurolisteriosis or localized infection at other sites, in which case the diagnosis may be suggested by neurologic or other focal findings. In a large French cohort study, the 3-month mortality rate for *L. monocytogenes* bacteremia was 46%; factors associated with death included multiorgan failure, neoplasia, worsening of preexisting comorbidities, monocytopenia (<200 cells/μL), older age, and female sex.

Neurolisteriosis *L. monocytogenes* has an affinity for the central nervous system. Neurolisteriosis is the second most common manifestation of invasive listeriosis. Signs of meningitis along with altered mental status, seizures, or focal neurologic findings suggest meningoencephalitis. A French cohort study found that 84% of patients with neurolisteriosis presented with meningoencephalitis.

The onset of neurologic disease can be sudden or subacute and occur over several days. Patients typically present with signs and symptoms similar to those in other bacterial meningitides. Compared with other bacterial meningitides, cerebrospinal fluid (CSF) pleocytosis and CSF neutrophil predominance are typically less pronounced; most patients (~60–75%) have CSF white blood cell counts of <1000/μL, but some have high levels. Approximately 30–40% of patients have low CSF glucose levels. Gram's staining of CSF sediment can show the expected gram-positive rods but commonly shows no organisms and sometimes shows gram-positive cocci, diplococci, or diphtheroids. In a U.S. study, *L. monocytogenes* caused <5% of cases of community-acquired bacterial meningitis in adults.

Uncommon neurolisteriosis manifestations include cerebritis, focal abscess, and rhombencephalitis (encephalitis of the cerebellum and brainstem). Patients with macroscopic abscesses often have positive blood cultures, but CSF findings may be normal in the absence of concurrent meningitis. Abscesses may be misdiagnosed as a primary or metastatic malignancy; they rarely occur in the cerebellum or spinal cord. Rhombencephalitis disproportionately affects otherwise healthy older adults. The classic presentation is biphasic, beginning with fever and headache and continuing after several days with signs of brainstem or cerebellar involvement, such as asymmetric cranial nerve palsies, ataxia, tremor, hemiparesis, or hemisensory deficits. Nearly half of patients with rhombencephalitis experience respiratory failure. The diagnosis may be delayed by the subacute course and by CSF findings, which are often only minimally abnormal. Magnetic resonance imaging (MRI) is superior to computed tomography (CT) for the diagnosis of neurolisteriosis, including rhombencephalitis.

Overall, the 3-month mortality rate for neurolisteriosis was 30% in a French cohort study; death was associated with the same risk factors as those documented for bacteremia. Neurolisteriosis-associated mortality was also higher among patients with a positive blood culture and among those treated with dexamethasone. Nearly half of survivors had long-term neurologic impairment.

Focal Infections Hematogenous dissemination of *L. monocytogenes* infrequently causes endocarditis, pneumonia, localized abscesses in the liver or other internal organs, peritonitis, septic arthritis, osteomyelitis, urinary tract infection, or skin lesions. Direct inoculation has been reported as a rare cause of ocular infection, skin infection, and lymphadenitis.

Infection in Pregnant Women and Neonates The risk of *Listeria* infection during pregnancy is increased due to impaired maternal cell-mediated immunity. Pregnancy-associated listeriosis is most common in the third trimester, although this may in part reflect underdiagnosis of earlier infections. Typically, pregnant women are either asymptomatic or have a mild, flulike illness with fever, headache, myalgias, or arthralgias. Neurolisteriosis and death are rare in pregnant women without other risk factors. Although nearly all infected women fully recover, only a minority (~5%) have a normal delivery and postpartum course, and three-quarters of the live-born infants are ill. In a French cohort study with 107 pregnancies in which *L. monocytogenes* was isolated from the mother, fetus, or neonate, 24% ended with fetal loss, 45% with premature birth, 6% with late-onset listeriosis in the neonate, and 21% with term delivery with fever or signs of fetal distress (i.e., meconium release into amniotic fluid, abnormal fetal heart rate). Fetal loss is uncommon after 29 weeks of gestation. Granulomatosis infantiseptica is a severe in utero infection caused by *L. monocytogenes* and characterized by disseminated microabscesses and granulomas in the skin, liver, and spleen; most infants with this condition are stillborn or die soon after birth. Neonatal listeriosis usually manifests in one of two ways: early-onset sepsis is hypothesized to result from in utero infection because it is typically diagnosed within 48 h after birth and is often associated with prematurity, whereas late-onset meningitis is thought to result from infection acquired at or soon after birth because it is typically diagnosed at ~2 weeks of age in full-term infants. A recent update from the above study that included additional years of data found that 70% of the 189 live-born infants developed early-onset listeriosis and 6% developed late-onset listeriosis; the other infants appeared uninfected (i.e., had negative cultures) and were well (14%) or had nonspecific signs (10%) that the authors attributed likely to prematurity. Overall, 3% of the 189 live-born infants died.

DIAGNOSIS

Because symptoms of listeriosis overlap with those of other infections, a high index of suspicion can facilitate timely diagnosis. Pregnant women with suspected listeriosis should have blood drawn for cultures, although blood cultures are positive only about half the time. Isolation of *L. monocytogenes* from a normally sterile site, such as blood, CSF, amniotic fluid, placental tissue, or fetal tissue, is diagnostic. *Listeria* must be distinguished from other gram-positive rods, especially diphtheroids. *L. monocytogenes* can be isolated from sterile specimens on routine medium; selective enrichment medium (such as PALCAM *Listeria* Selective Agar or Oxford Agar) enhances the capacity for isolation of the organism from nonsterile specimens, such as stool. Stool culture is not indicated in the evaluation of invasive listeriosis; culture on selective medium can be helpful for outbreak investigations of febrile gastroenteritis. Commercially available multiplex polymerase chain reaction panels for CSF specimens include *L. monocytogenes* as a target and may be a useful adjunct to culture. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry can rapidly identify an isolate as *L. monocytogenes*. Whole-genome sequencing has been a valuable tool for solving outbreaks, including a nosocomial outbreak associated with ice cream served in hospital milkshakes.

TREATMENT

Infections Caused by *Listeria monocytogenes*

L. monocytogenes treatment has not been evaluated in clinical trials. Recommendations are based on in vitro animal studies and observational clinical data. High-dose ampicillin (adult dose, 2 g IV every 4 h) or penicillin G (adult dose, 4 million units IV every 4 h) constitutes first-line therapy. Because penicillins are only weakly bactericidal against *L. monocytogenes*, many experts recommend adding gentamicin for synergy (1.0–1.7 mg/kg every 8 h if renal function is normal), particularly if the infection is severe. Small studies have had varying results with regard to the benefits of gentamicin. A large study provided evidence favoring amoxicillin-gentamicin as first-line therapy. Patients who are allergic to penicillin should undergo desensitization or be treated with trimethoprim-sulfamethoxazole (TMP-SMX; 5 mg/kg per dose of the trimethoprim component, given IV every 6–12 h). TMP-SMX should be avoided during the first trimester because it has been associated with neural tube and cardiovascular defects and in the perinatal period because it may increase the risk of kernicterus. Resistance to TMP-SMX has been reported; thus, antibiotic susceptibility testing should be performed if this drug is considered. Treatment failures have been reported with meropenem despite in vitro susceptibility of the organism. *L. monocytogenes* is susceptible in vitro to several other drugs, including vancomycin, linezolid, tetracycline, macrolides, and fourth-generation fluoroquinolones (e.g., moxifloxacin), but relevant clinical reports are limited. Cephalosporins are not effective. A study with 252 patients with neurolisteriosis found significantly reduced survival rates among the 32 patients treated with adjunctive dexamethasone; the authors suggest avoiding this drug in neurolisteriosis. Prepartum antibiotic treatment of pregnant women with listeriosis, initiated at least 1 day before delivery, enhances the chance of delivering a healthy infant.

The optimal duration of antibiotic therapy has not been established. Treatment duration usually depends on the clinical syndrome, disease severity, patient attributes, and response to treatment. The typical minimal treatment duration is 2 weeks for bacteremia, 2 weeks for early-onset neonatal disease, 3 weeks for meningitis, 4–6 weeks for endocarditis, and 6–8 weeks for brain abscess or encephalitis. Longer courses may be needed when patients are immunocompromised or are not improving as expected. Patients with neurolisteriosis do not routinely require a follow-up lumbar puncture if they are improving clinically during antibiotic therapy.

PREVENTION

Implementation of general precautions to prevent foodborne illness can help prevent listeriosis. These measures include fully cooking meats; washing fresh produce; cleaning hands, utensils, and kitchen surfaces after handling uncooked foods; and avoiding unpasteurized dairy products. Persons at increased risk for listeriosis should take additional precautions, including avoiding unpasteurized milk, soft cheeses (particularly those made with unpasteurized milk), refrigerated pâté or meat spreads, and premade deli salads (e.g., coleslaw; potato, tuna, or chicken salad). Hot dogs, fermented or dry sausages, and refrigerated, ready-to-eat delicatessen foods (e.g., smoked fish, deli meat, cheese sliced at a deli) should be avoided or heated until the internal temperature is 165°F or until they are steaming hot. Additional CDC recommendations can be found at *www.cdc.gov/listeria/prevention.html*. Testing and treatment are not indicated for an asymptomatic person who has eaten a product recalled because of *L. monocytogenes* contamination, even if the person has risk factors for invasive listeriosis. Hospital dietary services should implement safe food-preparation procedures for immunocompromised patients and should not serve these patients higher-risk foods. Care of patients with listeriosis includes only standard precautions because person-to-person transmission is rare. TMP-SMX given as prophylaxis for *Pneumocystis jirovecii* infection (e.g., to people infected with HIV or organ transplant recipients) helps prevent listeriosis.

FURTHER READING

Charlier C et al: Clinical features and prognostic factors of listeriosis: The MONALISA National Prospective Cohort Study. Lancet Infect Dis 17:510, 2017.

Charlier C et al: Neonatal listeriosis presentation and outcome: A prospective study of 189 cases. Clin Infect Dis 74:8, 2022.

Conrad AR et al: *Listeria monocytogenes* illness and deaths associated with ongoing contamination of a multiregional brand of ice cream products, United States, 2010–2015. Clin Infect Dis 76:89, 2023.

Farley MM: *Listeria monocytogenes*, in *Principles and Practice of Pediatric Infectious Diseases*, 6th ed, Long SS et al (eds). Philadelphia, Elsevier, 2023, pp 797–802.

Gottlieb SL et al: Multistate outbreak of listeriosis linked to turkey deli meat and subsequent changes in US regulatory policy. Clin Infect Dis 42:29, 2006.

Hof H: An update on the medical management of listeriosis. Expert Opin Pharmacother 5:1727, 2004.

Koopmans MM et al: Human listeriosis. Clin Microbiol Rev 36:e0006019, 2023.

Thomas J et al: Outbreak of listeriosis in South Africa associated with processed meat. N Engl J Med 382:632, 2020.

157

Tetanus

C. Louise Thwaites, Hai Duong Ha Thi

Tetanus is a preventable disease manifested by skeletal muscle spasms and autonomic nervous system disturbance. It is caused by a powerful neurotoxin produced by the bacterium *Clostridium tetani*, which is found globally. Tetanus commonly occurs in areas with low vaccination coverage. In developed countries, the disease is seen occasionally in individuals who are incompletely vaccinated. Even though the mortality rate has decreased significantly over the past two decades, the disease causes around 50,000 deaths annually worldwide. The mortality rate of tetanus varies depending on staff skills, clinical practices, and equipment capacity.

DEFINITION

Tetanus diagnosis is based on clinical manifestations with limited supportive laboratory confirmation. Case definitions are often used to facilitate clinical and epidemiologic assessments. The Centers for Disease Control and Prevention (CDC) defines probable tetanus as "in the absence of a more likely diagnosis, an acute illness with muscle spasms or hypertonia, and diagnosis of tetanus by a health care provider." *Neonatal* tetanus is defined by the World Health Organization (WHO) as "an illness occurring in a child who has the normal ability to suck and cry in the first 2 days of life but who loses this ability between days 3 and 28 of life and becomes rigid and has spasms." Given the unique presentation of neonatal tetanus, the history generally permits accurate classification of the illness with a high degree of probability. *Maternal* tetanus is defined by the WHO as "tetanus during pregnancy or within 6 weeks after the end of pregnancy (whether with birth, miscarriage, or abortion)."

ETIOLOGY

C. tetani is an anaerobic, gram-positive, spore-forming rod whose spores are highly resilient and can survive readily in the environment throughout the world. The spores resist boiling and many disinfectants. *C. tetani* spores and bacilli survive in the intestinal systems of many animals, and fecal carriage is common. The spores or bacteria enter the body through abrasions, wounds, or the umbilical stump (in the case of neonates). Once in a suitable anaerobic environment, the organisms grow, multiply, and release tetanus toxin that enters the nervous system and causes disease.

Approximately 20–30% tetanus cases have unclear entry wounds. Superficial abrasions to the limbs are the most common entry sites in adults. Deeper injuries or infections (e.g., open fractures, drug injection) are associated with more severe clinical presentations. Home delivery with unhygienic umbilical cutting and inadequate umbilical-cord care are the main causes of neonatal tetanus. Circumcision or ear-piercing also can result in neonatal tetanus.

Tetanus occurs when immunity is lacking. Unvaccinated individuals or those with incomplete vaccination history are at increased risk of developing tetanus.

EPIDEMIOLOGY

The incidence of tetanus has decreased gradually over the past few decades as vaccine coverage has increased. This disease is rare in developed countries. In 2022, 28 tetanus cases were reported in the United States. Most tetanus cases occur in individuals who have not received recommended tetanus vaccinations and booster shots every 10 years.

In the United States between 2009 and 2017, tetanus cases occurred primarily in those between the ages of 20 and 64 (64%), with only 13% of cases occurring in those younger than 20. Three cases of neonatal tetanus were reported in this period. Diabetes and intravenous drug use were associated with increased tetanus risk.

The global incidence of neonatal tetanus has reduced significantly following a concerted elimination program by WHO partnering with the United Nations Children's Fund (UNICEF) and the United Nations Population Fund (UNFPA); however, approximately 25,000 neonates with tetanus died in 2018. The incidence of tetanus among older children and adults (who have a high risk of tetanus due to the deterioration of immunity and lack of booster shots) is unknown. As few countries have good surveillance systems, in 2015, there were estimated to be between 30,000 and 62,000 deaths from tetanus in older children and adults.

PATHOGENESIS

Genome sequencing of *C. tetani* has allowed identification of several exotoxins and virulence factors. Only those bacteria producing tetanus toxin can cause tetanus. Tetanus toxin undergoes retrograde transport into the central nervous system (CNS) and thus produces clinical effects.

Tetanus toxin is intra-axonally transported to motor nuclei of the cranial nerves or ventral horns of the spinal cord. This toxin is produced as a single 150-kDa protein that is cleaved to produce heavy (100-kDa) and light (50-kDa) chains linked by a disulfide bond and noncovalent forces. The carboxy terminal of the heavy chain binds to specific membrane components in presynaptic α-motor nerve terminals; evidence suggests binding to both polysialogangliosides and membrane proteins. This binding results in toxin internalization and uptake into the nerves. Once inside the neuron, the toxin enters a retrograde transport pathway, whereby it is carried proximally to the motor neuron body. It is known that tetanus toxin exhibits several different pH-dependent conformations and therefore can interact with a variety of different receptors. During its passage from the periphery to the central nervous system, tetanus toxin can access neuronal trafficking systems and evade degradation.

Following retrograde transport in the motor neuron, the tetanus toxin undergoes translocation across the synapse to the GABA-ergic presynaptic inhibitory interneuron terminals. Here the light chain, which is a zinc-dependent endopeptidase, cleaves vesicle-associated membrane protein 2 (VAMP2, also known as *synaptobrevin*). This molecule is necessary for presynaptic binding and release of neurotransmitter; thus, tetanus toxin prevents transmitter release and effectively blocks inhibitory interneuron discharge. The result is unregulated activity in the motor nervous system. Similar activity in the autonomic system accounts for the characteristic features of skeletal muscle spasm and autonomic system disturbance. The increased circulating catecholamine levels in severe tetanus are associated with cardiovascular complications.

Relatively little is known about the processes of recovery from tetanus. Recovery can take several weeks. Peripheral nerve sprouting is involved in recovery from botulism, and similar CNS sprouting may occur in tetanus. Other evidence suggests toxin degradation as a mechanism of recovery.

APPROACH TO THE PATIENT

Tetanus

The clinical manifestations of tetanus occur only after tetanus toxin has reached presynaptic inhibitory nerves. Treatment should not be delayed once the diagnosis of tetanus is confirmed. Management strategies aim to neutralize remaining unbound toxin, support vital functions, treat symptoms, and control complications until the effects of the toxin have worn off. Patients usually recover after 4–6 weeks (see "Treatment," below).

CLINICAL MANIFESTATIONS

Clinical presentations of tetanus are diverse and are divided into the following categories: generalized, localized, cephalic, and neonatal tetanus. In the mild form of local tetanus, only isolated areas of the body are affected and only small areas of local muscle spasm may be apparent. Localized tetanus can progress to generalized tetanus. If the cranial nerves are involved in localized cephalic tetanus, the pharyngeal or laryngeal muscles may spasm, with consequent aspiration, respiratory failure, or airway obstruction. Generalized tetanus is the most common form of clinical presentation, characterized by muscle rigidity and generalized spasms. Neonates with tetanus typically present with an inability to suck, poor feeding, and generalized spasms.

The most common initial symptoms are trismus (lockjaw), which progresses to neck and body muscle rigidity, difficulty swallowing, and pharyngeal and laryngeal spasms. As the disease progresses, generalized muscle spasms develop and cause pain. Laryngeal spasm can cause respiratory failure or apnea and is a life-threatening event; without immediate respiratory support, this is the most common cause of death in tetanus.

Autonomic nervous system disturbance (ANSD) occurs during the second week of severe tetanus, and death due to cardiovascular events becomes the major risk. Clinical symptoms of ANSD include fluctuated heart rate (tachycardia or bradycardia) accompanied by a fluctuation in blood pressure (hypertension or hypotension) resulting from alteration in the activities of the sympathetic or parasympathetic nervous systems. Autonomic involvement is evidenced by gastrointestinal stasis, sweating, and increased tracheal secretions.

The Ablett score classifications of the clinical presentation of tetanus are outlined in **Table 157-1**.

DIAGNOSIS

The diagnosis of tetanus is based on clinical findings. Confirmatory laboratory tests are limited. Positive culture of *C. tetani* from wounds cannot confirm a diagnosis of tetanus. Serum antitetanus immunoglobulin G may also be measured in a sample taken before the administration of antitoxin or immunoglobulin; levels >0.1 IU/mL (measured by standard enzyme-linked immunosorbent assay) are deemed protective and do not support the diagnosis of tetanus. If levels are below this threshold, a bioassay for serum tetanus toxin may be helpful, but a negative result does not exclude the diagnosis, and these levels are not generally performed. Polymerase chain reaction and recombinase polymerase amplification have been developed to detect the tetanus neurotoxin gene; these techniques have proven promising in recent years.

A diagnosis of tetanus requires differentiation from other diseases. The few conditions that mimic generalized tetanus include strychnine poisoning, dystonic reactions to antidopaminergic drugs, somatic symptom disorder, stiff person syndrome, and neuroleptic malignant syndrome. Abdominal muscle rigidity is characteristically continuous in tetanus and should be differentiated from peritonitis or acute abdominal emergency. Cephalic tetanus can be confused with trismus of other etiologies, such as oropharyngeal infection, cranial nerve diseases, intracranial hemorrhage, submaxillary lymphadenitis, or peritonsillar infections. Hypocalcemia and meningoencephalitis are included in the differential diagnosis of neonatal tetanus.

TABLE 157-1 Ablett Classification of Severity of Tetanus

GRADE	SEVERITY	SYMPTOMS
I	Mild	Mild trismus, general spasticity, no respiratory compromise, no spasms, no dysphagia
II	Moderate	Moderate trismus, rigidity, short spasms, mild dysphagia, moderate respiratory involvement, respiratory rate >30 breaths/min
III	Severe	Severe trismus, generalized rigidity, prolonged spasms, severe dysphagia, apneic spells, pulse >120 beats/min, respiratory rate >40 breaths/min
IV	Very severe	Grade 3 with autonomic dysfunction

TREATMENT

Tetanus

If possible, the entry wound should be identified, cleaned, and debrided of necrotic material in order to remove anaerobic foci of infection and prevent further toxin production. Wound care should be performed several hours after anti-toxin administration. Failure to remove devitalized tissue and treat infection may result in recurrent or prolonged tetanus. Metronidazole (400 mg rectally or 500 mg IV every 6 h for 7 days) is preferred for antibiotic therapy. Although not a first choice for therapy, an alternative is penicillin (100,000–200,000 IU/kg per day); this drug theoretically may exacerbate spasms due to its ability to bind to the GABA receptor and, in one study, was associated with increased mortality.

Antitoxin should be given early in an attempt to deactivate any circulating tetanus toxin and prevent its uptake into the nervous system. Two preparations are available: human tetanus immunoglobulin (HTIG) and equine antitoxin. HTIG is the preparation of choice, as it is less likely to be associated with anaphylactoid reactions. A single IM dose (500–5000 IU) is given. Equine-derived antitoxin is available widely and is used in low- and middle-income countries; after hypersensitivity testing, 10,000–20,000 U is administered IM. According to a recent randomized controlled trial in Vietnam, there were no significant differences in the outcome and complications between groups using either intramuscular equine antitoxin or HTIG. Additional intrathecal antitoxin (HTIG) showed no benefit.

Benzodiazepines are commonly used to control spasms, and patients can tolerate them in high doses. The use of intermittent or continuous sedation depends on the severity of spasms and the availability of appropriate resources, e.g., mechanical ventilators. High-dose diazepam may cause hyperosmolarity and lactic acidosis. Midazolam is another option for controlling spasms in tetanus; it has fewer side effects and can be used continuously. Infusions of propofol can be used to control spasms and provide sedation; however, consideration should be given to the likely long duration of therapy. Sedatives should be reduced in the elderly and patients who have liver diseases due to drug accumulation and slow excretion.

When sedatives alone cannot control spasms, a combination of neuromuscular blocking agents is recommended. Nondepolarizing neuromuscular blockers are used in clinical practice, depending upon availability. IV magnesium sulfate has been used as a muscle relaxant. Patients who are using these agents require mechanical ventilation support. In those settings with limited availability of mechanical ventilators, controlling spasms while maintaining adequate ventilation is problematic. Patients may require ventilator support for several weeks.

It is important to establish a secure airway early in severe tetanus. Ideally, patients should be nursed in calm, quiet environments because light and noise can trigger spasms. Dysphagia due to pharyngeal involvement combined with hyperactivity of laryngeal muscles makes endotracheal intubation difficult. Tracheostomy is the better option for securing the airway in severe tetanus.

Magnesium sulfate has been used in autonomic nervous system dysfunction to control high blood pressure and tachycardia; maintaining a plasma concentration of 2–4 mmol/L is recommended. If magnesium alone cannot control tachycardia, short-acting beta-blockers (labetalol, propranolol, esmolol) or calcium antagonists can be used with strict monitoring. When the parasympathetic nervous system predominates—resulting in prolonged low blood pressure and bradycardia—vasopressors are required.

A complication arising from treatment with diazepam injection is thrombophlebitis. Long-term hospitalization with high-dose sedation creates a high risk of hospital-acquired infections (ventilator-associated pneumonia, bloodstream infection, urinary tract infection, sepsis), deep vein thrombosis, pneumonia emboli, myocardiopathy, myocardial infarction, stress ulcers, muscle weakness, and pressure sores. Many of these patients require long-term rehabilitation.

Patients must be given a full primary course of immunization as tetanus toxin is poorly immunogenic and the immune response following natural infection is inadequate.

PROGNOSIS

Rapid development of tetanus is associated with more severe disease; it is important to note time of onset and length of incubation period. More sophisticated modeling has revealed other important predictors of prognosis. In many adults, particularly in the elderly, surviving tetanus is associated with reduced long-term functional outcome measures. Studies of children and neonates have suggested a higher incidence of neurologic sequelae. Neonates may be at increased risk of learning disabilities, behavioral problems, cerebral palsy, and deafness. Tetanus has a high survival rate if complications can be controlled and interventions conducted promptly.

In assessing prognosis, the incubation period (time from wound to first symptom) and the period of onset (time from first symptom to first generalized spasm or pharyngeal or laryngeal spasm) are of particular significance. The shorter these periods, the worse is the prognosis. Among the three main scales available to predict the severity and outcome of tetanus, the Tetanus Severity Score (TSS) **(Table 157-2)** is superior to the Dakar and Philips scores, with a sensitivity of 66% and a specificity of 91%.

PREVENTION

Tetanus is prevented by good wound management and vaccination **(Chap. 129)**. Safe delivery, hygienic umbilical-cord care, and maternal vaccination are recommended to prevent neonatal tetanus. Individuals sustaining wounds should undergo passive immunization (see "Treatment of Tetanus," above) if their vaccination status is incomplete or unknown or if their last booster was given >10 years earlier. Vaccination programs and recommended prevention measures vary somewhat according to individual countries.

The rate of primary vaccination coverage in infancy (three doses of DTP [diphtheria, pertussis, and tetanus]) is 86%, but rates for the subsequent boosters necessary for long-term protection are unknown. WHO guidelines for tetanus vaccination consist of a primary course of three doses and should start at 6 weeks of age; the interval between doses is 4 weeks. The first booster is given from 12–23 months of age. The second and third boosters are at 4–7 and 9–15 years of age, respectively. There should be at least 4 years between booster doses.

In 2022, the CDC reported the rate of infants who had received a full primary vaccination course of three doses of the diphtheria, tetanus toxoid (DTP3), and pertussis vaccine was 94%. There is a lack of data about boosters in older children, adolescents, and adults in many countries. The CDC recommends five doses of tetanus vaccine for infants and children at 2 months, 4 months, 6 months, 15 through 18 months, and 4 through 6 years of age, an additional booster dose at 11–12 years of age, and every 10 years thereafter. For those with delayed primary vaccination, catch-up immunization schedules are recommended. There are separate schedules for children ages 4 months through 6 years of age and from ages 7 through 18 years of age.

Complete maternal vaccination reduces the incidence of neonatal tetanus by an estimated 94%. WHO recommends that pregnant women who have not been vaccinated with the tetanus vaccine should receive at least two doses, with an interval of 4 weeks between doses. The second dose should be given at least 2 weeks before delivery. The third dose should be received at least 6 months later, with the fourth and fifth doses following an interval of 1 year or during subsequent pregnancies. A total of five doses can provide long-term immunity. The CDC provides a special schedule for those with partial vaccination. In high-risk areas, women of childbearing age should receive a primary course of vaccination and education on safe delivery and postnatal practices.

TABLE 157-2 Tetanus Severity Score (TSS) (Sensitivity 66%, Specificity 91%), Cutoff Point to Predict Death ≥8

VARIABLES	SCORE
Age (year)	
≤70	0
71–80	5
>80	10
Time from first symptom to admission (days)	
≤2	0
3–5	–5
>5	–6
Difficulty breathing on admission	
No	0
Yes	4
Coexisting medical condition	
Fit and well	0
Minor illness or injury	3
Moderately severe illness	5
Severe illness not immediately life threatening	5
Immediately life-threatening illness	9
Entry sites	
Internal or injection	7
Other (including unknown)	0
Highest systolic blood pressure recorded during first day in hospital (mmHg)	
≤130	0
131–140	2
>140	4
Highest heart rate recorded during first day in hospital (beats/min)	
≤100	0
101–110	1
111–120	2
>120	4
Lowest heart rate recorded during first day in hospital (beats/min)	
≤110	0
>110	–2
Highest temperature recorded during first day in hospital (°C)	
≤38.5	0
38.6–39	4
39.1–40	6
>40	8

Since March 2022, 47 countries have achieved maternal and neonatal tetanus elimination. Despite this relative success, immunization programs need to be continued and promoted to maintain individual long-term protective immunity and to eliminate the incidence of tetanus gradually. Dedicated public health initiatives still need to be improved, and the continuing reports of sizable case series in the medical literature suggest that tetanus continues to pose a significant global health burden.

ACKNOWLEDGMENT

The authors wish to thank Dr. Lam Minh Yen for her contributions to this chapter in previous editions.

FURTHER READING

BORROW R et al: The immunological basis for immunization series. Module 3: Tetanus update 2018. Edited by Vaccines and Biologicals Immunization. World Health Organization, 2018.

KYU HH et al: Mortality from tetanus between 1990 and 2015: Findings from the global burden of disease study 2015. BMC Public Health 17:179, 2017.

RODRIGO C et al: Pharmacological management of tetanus: An evidence-based review. Crit Care 18:217, 2014.

Thwaites CL et al: Magnesium sulfate for treatment of severe tetanus: A randomised controlled trial. Lancet 368:1436, 2006.

Thwaites CL et al: Predicting the clinical outcome of tetanus: The tetanus severity score. Trop Med Int Health 11:279, 2006.

Van Hao N et al: Human versus equine intramuscular antitoxin, with or without human intrathecal antitoxin, for the treatment of adults with tetanus: A 2 x 2 factorial randomised controlled trial. Lancet Glob Health 10:e862, 2022.

Yen LM, Thwaites CL: Tetanus. Lancet 393:1657, 2019.

WEBSITES

Centers for Disease Control and Prevention: Pink Book. Tetanus. 2021. *www.cdc.gov/vaccines/pubs/pinkbook/downloads/tetanus.pdf*.

Centers for Disease Control and Prevention: Recommended catch-up immunization schedule for children and adolescents who start late or who are more than 1 month behind. *https://www.cdc.gov/vaccines/schedules/hcp/imz/catchup.html*.

Health Protection Agency: Tetanus: Information for health professionals. 2013. *www.gov.uk/government/publications/tetanus-advice-for-health-professionals*.

Statista: Reported cases of tetanus in the United States from 1980 to 2022, by year. *www.statista.com/statistics/1122819/tetanus-cases-us-by-year/*.

World Health Organization: Immunization coverage. *www.who.int/news-room/fact-sheets/detail/immunization-coverage*.

World Health Organization: Maternal and neonatal tetanus (MNT) elimination. *www.who.int/immunization/diseases/MNTE_initiative/en/*.

World Health Organization: Tetanus. *www.who.int/news-room/fact-sheets/detail/tetanus*.

158

Botulism

Carolina Lúquez, Jeremy Sobel

Botulism is a rare, life-threatening disease characterized by cranial nerve palsies and symmetric descending flaccid paralysis. Four forms of naturally occurring botulism have been described: foodborne botulism, infant botulism, wound botulism, and adult intestinal colonization. Other forms of botulism include iatrogenic botulism and inhalational botulism. Effective treatment depends on early clinical diagnosis.

ETIOLOGY AND PATHOGENESIS

Botulism is caused by botulinum neurotoxins (BoNTs), which are produced by *Clostridium botulinum*. Rare strains of *Clostridium butyricum* and *Clostridium baratii* can also produce BoNTs. Seven distinct serotypes of BoNT (A through G) are well characterized; serotypes A, B, E, and F reportedly cause disease in humans. Novel serotypes—BoNT/FA (or H or HA), BoNT/En, and BoNT/X—have been proposed, but the scientific community has not yet reached a consensus as to whether each represents a new serotype or a combination of known serotypes, as in the case of BoNT/FA (or H or HA), or whether they represent true toxins or botulinum-like proteins, as in the case of BoNT/En and BoNT/X. BoNTs are encoded by the *bont* gene, which is also diverse in its DNA sequence. At least 40 unique subtypes of BoNT have been identified within serotypes A, B, E, and F. By definition, a variant of BoNT represents a new subtype when its amino acid sequence differs by at least 2.6% from those of all known subtypes within that particular serotype. Although 2.6% is an arbitrary threshold, this figure has provided the basis for genetic subtype designations for the past decade, aiding in the classification of BoNTs as new DNA or amino acid sequences become publicly available. In addition, *bont* genes typically reside within two types of gene clusters. One type includes *ha* genes encoding hemagglutinin proteins, which facilitate the absorption of toxins across the epithelial barrier. The other type of cluster includes *orfX* genes that encode proteins with unknown functions. Both cluster types include an *ntnh* gene, which encodes for a nontoxic nonhemagglutinin protein. It has been proposed that these accessory proteins form a complex with BoNTs and protect them from external proteolytic activity.

Despite their structural variability, BoNTs all have a similar mechanism of action: they target neurons and block neurotransmission by cleaving SNARE-family proteins in the host, with consequent inhibition of acetylcholine release. BoNTs are metalloproteases composed of a light chain and a heavy chain. The light chain has catalytic activity, and the heavy chain contains a translocation domain and a receptor-binding domain. The receptor-binding domain of the heavy chain mediates the neurospecific binding of BoNTs, which leads to its internalization within endocytic compartments. Interaction of the translocation domain of the heavy chain with the membrane of endocytic vesicles leads to the translocation of the light chain into the cytosol. Once in the cytosol, the light chain cleaves specific SNARE-family proteins. Serotypes A and E cleave SNAP-25; serotypes B, D, F, and G cleave VAMP; and serotype C cleaves SNAP-25 and syntaxin. Cleavage of any of these proteins disrupts the assembly of synaptic fusion complexes, and this disruption inhibits the fusion of the membrane of the synaptic vesicle containing acetylcholine with the neuronal cell membrane. Clinically, the result is flaccid paralysis of voluntary muscles. The irreversible binding of BoNTs to their targets has a clinical consequence: once toxin binding has occurred, the resulting paralysis persists for weeks or months, until nerve endings have been regenerated.

BoNTs are produced by *C. botulinum* and a few strains of *C. butyricum* and *C. baratii*, which are all gram-positive, rod-shaped, spore-forming, anaerobic bacteria. Under most environmental conditions, *C. botulinum* exists as spores that are heat-resistant and ubiquitous in soil. In general, *C. botulinum* spores require temperatures above boiling to ensure destruction; their thermal resistance increases with higher pH and lower salt content. Spores present in foods can survive most preservation methods and, if the conditions allow it, can germinate and produce BoNTs in significant amounts to cause disease. Commercial retort canning reliably destroys botulinal spores.

BoNTs are among the most toxic substances known. Extremely small amounts of BoNT can cause severe disease and death. Severity of disease varies with dose, serotype, and route of exposure. The lethal dose of BoNT in humans is not known but can be estimated by extrapolation of toxicity data from animal studies. The estimated human lethal dose of BoNT acquired via the IV or IM route is 0.1–1 ng/kg of body weight. The human lethal dose of BoNT acquired by inhalation of aerosolized toxin is estimated at 1–75 ng/kg. The degree of toxicity of BoNT acquired by the oral route is estimated to be much lower: 0.1–1 μg/kg.

As stated above, four naturally occurring and two non–naturally occurring forms of botulism are known. Foodborne botulism is caused by the ingestion of foods contaminated with BoNT. Wound botulism occurs when spores of BoNT-producing species of *Clostridium* contaminate a wound and then germinate, multiply, and produce toxin. Infant botulism is caused by BoNT-producing species of *Clostridium* colonizing the intestinal tract of infants ≤1 year of age. Adult intestinal colonization is similar to infant botulism but affects persons >1 year of age. Iatrogenic botulism occurs when a patient given injections of BoNT experiences signs of systemic botulism. BoNTs can also be aerosolized and used as a bioweapon, entering the human body by inhalation.

Foodborne Botulism Foodborne botulism is the most common form reported in many countries. Every case of foodborne botulism represents a public health emergency because of the potential for causing outbreaks. Foodborne botulism is an intoxication in which food containing preformed toxin is ingested. Spores of BoNT-producing species of *Clostridium* are ubiquitous in soil and can be found on vegetables and other foodstuffs. *C. botulinum* type E is commonly found in aquatic environments and in aquatic animals. Because the spores are found in many foods, improper preparation or storage may produce the confluence of conditions that allow germination and growth of BoNT-producing species of *Clostridium*, which in turn result in production of BoNT. Both historically and at the present time, canned foods are of concern because they create anaerobic environments. To render these foods safe, proper processing procedures in conditions of enough heat and pressure to inactivate *Clostridium* spores, along with sufficient acidity, salinity, or other preservative methods to limit the organism's growth and its production of BoNT, are required. Low-acidity foods, such as corn, peppers, potatoes, and beets, represent a higher risk. A series of botulism outbreaks from commercially canned foods in the early twentieth century resulted in standardization of retort canning methods and promulgation and enforcement of production safety codes. Consumption of fish or other foods of marine origin can cause botulism if prepared or conserved improperly. Most foodborne botulism cases in the United States are caused by home-canned vegetables such as green beans; however, commercially prepared foods, including chicken broth, carrot juice, hot dog chili sauce, and nacho cheese, also have been implicated in recent outbreaks. Marine mammal and fish products traditionally prepared by Alaskan Natives and First Peoples are the main source of botulism in Alaska and Canada.

Wound Botulism Wound botulism is caused by germination and growth of *C. botulinum* spores in a wound or necrotic tissue where they produce BoNT, which then enters circulation and produces systemic disease. Few cases of wound botulism were described in the United States until 1981, when the first case associated with injection drug use was reported. Since then, botulism cases due to injection drug use, especially in association with subcutaneous or tissue injection (skin popping) of black tar heroin, have substantially increased in the United States. Black tar heroin was introduced into the United States in the 1970s and, since the late 1980s, has become the predominant form of heroin west of the Mississippi River. Black tar heroin is contaminated with by-products of the manufacturing process, adulterants, and diluents and therefore is considered the most probable source of *C. botulinum* spores. In recent decades, the few cases of wound botulism not associated with injection drug use have been associated with vehicle crashes, gunshot wounds, open-fracture wounds, and penetrating wounds caused by contaminated objects.

Infant Botulism Infant botulism is the most common form of botulism in the United States. It affects infants ≤1 year old, with a mean age at onset of 14 weeks. It has been suggested that the intestinal microbiota in infants may induce susceptibility to botulism; animal models seem to support this claim. Spores of BoNT-producing species of *Clostridium* can enter the body by ingestion. The highly resistant spores survive passage through the stomach and colonize the intestine, where they germinate, grow, and produce BoNT in situ. Infants can continue excreting *C. botulinum* for weeks after clinical recovery. Spores of BoNT-producing species of *Clostridium* have been found in honey. Consumption of honey has been epidemiologically implicated in infant botulism; therefore, honey should not be fed to babies ≤1 year of age. Honey exposure, however, explains only a small proportion of cases. As spores are found in dust and soil, most infant botulism patients probably acquire BoNT-producing species of *Clostridium* by swallowing dust particles. Why only a few dozen infants are affected each year when presumably most infants regularly ingest clostridial spores remains unknown.

Adult Intestinal Botulism Similar to infant botulism, adult intestinal colonization is caused by spores of BoNT-producing species of *Clostridium* colonizing the large intestine, growing, and producing BoNT in situ. Although spores are routinely ingested and excreted by humans, the adult intestinal tract does not support spore germination and toxin production under normal circumstances. Adult intestinal colonization is usually associated with inborn anatomic abnormalities, gastrointestinal surgery, or prolonged use of antibiotics, which may alter the normal intestinal microbiota and facilitate colonization by BoNT-producing species of *Clostridium*. Although these associated conditions are relatively common, fewer than 30 cases of adult intestinal colonization have been reported worldwide.

Iatrogenic Botulism Iatrogenic botulism occurs in patients injected with large doses of BoNT for treatment of muscle complications related to such conditions as cerebral palsy and spastic dystonia. The small doses of botulinum toxin used for wrinkle elimination in dermatologic practice are usually insufficient to cause systemic disease. In 2004, an outbreak of four cases caused by the injection of an unlicensed, highly concentrated BoNT product for cosmetic purposes occurred in the United States. Similarly, in 2017, an outbreak of nine cases occurred in Egypt in association with an unlicensed, highly concentrated BoNT preparation.

Weaponized Inhalational Botulism BoNTs were weaponized by the biological weapons programs of several countries in the twentieth century. Aerosolized BoNTs can be used as a bioweapon, exerting their effect by entering the body through inhalation. In the United States, BoNTs are designated as Tier 1 select agents—i.e., agents that present the greatest risk of deliberate misuse with significant potential for mass casualties or devastating effects on the economy, critical infrastructure, or public confidence. Tier 1 agents pose a severe threat to public health and safety. Terrorists have attempted to use BoNT as a bioweapon: Aum Shinrikyo, a Japanese cult, tried unsuccessfully to aerosolize BoNT in terrorism attacks at multiple sites in Japan between 1990 and 1995.

■ EPIDEMIOLOGY

Foodborne Botulism In the United States, foodborne botulism is the third most common form of botulism. From 2001 to 2019, 362 foodborne botulism cases were reported, with a mean of 19 cases per year. Most cases (64%) were caused by serotype A BoNT, which was followed in frequency by serotype E (26%). Serotypes B and F caused 7% and 2% of foodborne botulism cases, respectively. Outbreaks caused by serotype E usually had a shorter incubation period, those caused by type A had higher numbers of patients who required mechanical ventilation, and those caused by type B had lower numbers of deaths.

Foodborne botulism cases are usually sporadic (i.e., cases occur singly), but small and large outbreaks also can occur. From 2001 to 2019, 5 foodborne botulism outbreaks affecting 10 or more people were reported in the United States **(Table 158-1)**. Every case of foodborne botulism is considered a public health emergency because it may be the first in an outbreak involving additional patients.

Most foodborne botulism cases in the United States are due to a wide variety of home-canned vegetables and pickled vegetables (e.g., beets, green beans, carrots, mushrooms, asparagus, peppers, beans, mustard greens, corn, tomato sauce, olives, pumpkin butter), vegetables baked in aluminum foil (e.g., potatoes, beets), home-canned meat-based foods (e.g., tuna, pickled pigs' feet, stew, pasta in meat sauce),

TABLE 158-1 Total Foodborne Botulism Outbreaks of 10 or More Cases Reported in the United States Between 2001 and 2019

YEAR	STATE	FOOD SOURCE	NO. OF CONFIRMED CASES
2001	Texas	Chili	16
2007	Multistate	Commercially canned hot dog chili sauce	10
2015	Ohio	Home-canned potatoes used to prepare a potato salad, served at a church potluck	27
2016	Mississippi	Pruno, illegal alcoholic beverage consumed by persons who were incarcerated at a federal facility	19
2017	California	Commercially produced nacho cheese, sold at a convenience store	10

oil-based foods (e.g., pasta and jarred pesto, homemade garlic-infused oil), herbal deer antler tea, home-prepared fermented tofu, commercial clam chowder, or commercial grain and vegetable products. In Alaska, traditional Alaskan Native foods linked to foodborne botulism cases have included seal oil, seal blubber, dried herring in seal oil, fermented seal flipper, stinkheads and other fermented fish heads, stinkfish, salmon eggs, beaver tail, whitefish, fish eggs, fermented beluga, and whale blubber.

Commercial food manufacturing processes include retort canning, in which high temperature and pressure destroy the highly resistant clostridial spores, and manipulations that inhibit bacterial growth, such as acidification or addition of growth inhibitors that prevent germination and growth of BoNT-producing species of *Clostridium* and the production of BoNT. However, commercial foods occasionally still cause botulism if safe manufacturing processes are not followed or fail or if foods are stored or used inappropriately by the retailer or consumer. For instance, an outbreak of 10 cases associated with commercially canned hot dog chili sauce occurred in 2007 as a result of deficiencies in the canning process. Other commercial food–associated outbreaks that occurred in the United States between 2001 and 2019 include a 2001 outbreak of 16 cases linked to chili that was stored at inappropriate temperatures and later served at a church event in Texas and a 2006 outbreak linked to commercial carrot juice, which included four cases in the United States and two cases in Canada. The investigation of the latter outbreak led to an international product recall. The juice, which had no added sugar, salt, or preservatives, was stored at inappropriate temperatures.

Pruno, an illicit prison-brewed alcoholic beverage, first caused a botulism outbreak in a California prison in 2004, affecting 4 prisoners. In 2011, a second outbreak due to pruno was reported and involved 8 patients at a prison in Utah. In 2012, 2 outbreaks associated with pruno occurred in a single prison in Arizona, with 4 and 8 cases, respectively. The largest outbreak from pruno occurred in 2016 in a Mississippi prison; 31 cases were identified, including 19 confirmed and 12 suspected.

Wound Botulism Wound botulism was once rare in the United States, but its frequency has been increasing for decades, and it is now the second most common form of botulism. Between 2001 and 2019, 452 cases of wound botulism were reported, with an average of 24 cases per year. Most cases (92%) were caused by BoNT serotype A and by serotype B (5%). Most cases (96%) were among persons who injected drugs (mainly black tar heroin), and the remaining 4% of cases were due to traumatic injuries.

Infant Botulism Infant botulism is the most common form of botulism in the United States. Between 2001 and 2019, 2172 infant botulism cases were reported. BoNT serotypes A and B caused most cases (40% and 58%, respectively). Only 2 cases were due to serotype E. Of these 2 cases, 1 was due to *C. botulinum* type E and the other to *C. butyricum* type E; both cases represented the first report anywhere in this country of infant botulism due to those respective organisms. A small fraction (<1%) of cases were caused by serotype F. Of note, 22 infant botulism cases were due to strains of *C. botulinum* that can produce two BoNT serotypes (A and B or B and F).

Botulism of Other Etiologies Between 2001 and 2019, 51 cases were reported as being of "unknown or other etiology." This category includes laboratory-confirmed botulism cases that do not meet the definition of foodborne, infant, or wound botulism. Most of these cases were caused by serotype A (63%) and serotype F (27%). Many were thought to be cases of adult intestinal colonization, although confirmation of this form of botulism is not always possible.

■ CLINICAL MANIFESTATIONS

Botulism produces a syndrome characterized by bilateral cranial nerve palsies that may be followed by symmetric, descending flaccid paralysis that may cause respiratory arrest. There are no sensory deficits; patients are fully conscious, with normal intellectual function, although cranial nerve palsies may give a mistaken impression of altered consciousness. The typical incubation period (based on data for foodborne botulism cases, where exposure can be identified) is 1 or 2 days, but a range of 6 h to >7 days has been reported. Several recent systematic reviews substantiate long-known observations that the syndrome is essentially identical for all types of botulism in patients of all ages, although elicitation of the typical signs and symptoms may be challenging in infants and young children. A recent systematic review of 16 cases of botulism in pregnant women reported the same clinical syndrome as in nonpregnant individuals. In all botulism syndromes, the first neurologic manifestation usually is ptosis, which can be striking. Ocular findings of fuzzy vision or frank diplopia are caused by extraocular muscle paralysis due to palsies of cranial nerves III, IV, and VI. Flat, youthfully unlined, expressionless facies are produced by cranial nerve VII (facial nerve) palsy. Dysarthria is also a prominent manifestation. Oral and nasal regurgitation of foods or beverages is caused by cranial nerve IX (glossopharyngeal nerve) palsy. The autonomic system may be affected, producing anhidrosis manifesting as severe pharyngeal pain and erythema that has been mistaken for pharyngitis; paradoxically, other patients experience an inability to manage copious oral secretions. Autonomic dysfunction may produce hemodynamic instability requiring monitoring. Cranial nerve palsy may produce pharyngeal muscle flaccidity, causing airway collapse and respiratory arrest early in the course of illness, while reduction in diaphragmatic and accessory muscle function may cause respiratory compromise hours or days later. Cranial nerve palsies may be followed by descending symmetric flaccid paralysis of the muscles of the neck, shoulders, upper limbs, and lower limbs; proximal muscle groups of each limb are affected before distal muscle groups.

A recent analysis of 332 U.S. botulism cases found the following frequencies for patient-reported symptoms: difficulty swallowing, 86%; fatigue, 85%; blurred vision, 80%; slurred speech, 78%; double vision, 76%; shortness of breath, 65%; and dry mouth, 62%. The analysis also reported the following frequencies of observed signs: afebrile body temperature, 99%; descending paralysis, 93%; alert and oriented status, 93%; ptosis, 81%; limb weakness, 78%; decreased palatal reflex, 54%; facial palsy, 47%; and dilated pupils, 20%. Sixty-six percent of patients were intubated and received mechanical ventilation. These findings are similar to those reported in many smaller series. Rarely, asymmetry of cranial nerve palsies or distal muscle paralysis is reported and, at least in some cases (especially those described in reports based on chart abstractions), may reflect an incomplete or incompletely recorded neurologic examination. Despite intact sensorium, symptoms such as ptosis, dysarthria, and gait instability may be mistaken for diminished consciousness and lack of coordination and may be erroneously attributed to intoxication from alcohol or other substances. Paresthesias have been reported in some patients; these sensations are not explained by the known activity of botulinum toxin. Paralysis of the diaphragm and accessory muscles of respiration may occur, producing respiratory compromise. Distal tendon reflexes diminish symmetrically. Constipation due to intestinal paralysis develops in almost all patients. Nausea and vomiting may occur in foodborne botulism, preceding neurologic symptoms. Whether these manifestations are due to BoNT, other products of BoNT-producing species of *Clostridium*, or other contaminants of spoiled food is unknown. These gastrointestinal symptoms have not been reported in wound botulism.

Death in untreated patients during the first hours to days of illness is caused by airway obstruction resulting from pharyngeal muscle paralysis and inadequate tidal volume resulting from paralysis of diaphragmatic and accessory respiratory muscles. The combination of expressionless facies from cranial nerve paralysis and immobility from voluntary muscle paralysis may give patients with botulism a placid appearance that masks the agitation expected with respiratory distress. Respiratory compromise occurs early in the course of disease in a substantial proportion of patients: the largest systematic literature review to date of foodborne and wound botulism cases (402 patients) reported that the average time from symptom onset to hospitalization was 2 days and that, at hospital admission, 42% of patients had respiratory symptoms; of these patients, 42% presented with no extremity weakness. In the same review, 87% of patients who required mechanical ventilation were intubated during the first 2 days of hospitalization. The severity of disease varies greatly between patients and is probably governed by the dose of toxin to which they have been exposed. Without treatment, some patients do not progress beyond ptosis and mild palsy in one or two cranial nerves; others experience fulminant cranial nerve palsies and rapidly progressive descending flaccid paralysis eventually affecting most or all voluntary muscles as well as respiratory failure requiring intubation and mechanical ventilation within hours.

The different BoNT serotypes are associated with variations in the botulism syndrome. BoNT type A is associated with more rapid disease progression, more frequent respiratory compromise and mechanical ventilation, and longer duration of paralysis. Type B is associated with a milder syndrome, with less severe and shorter-duration paralysis. Intoxication with the rarely occurring type F produces a syndrome of rapidly progressing paralysis that often leads to respiratory failure, with more rapid recovery than occurs with other toxin types. However, all toxin types causing human illness can cause severe disease; the clinical approach is the same for all.

The paralysis of botulism can last for weeks or months—the time required for regeneration of affected nerve endings and recovery of voluntary muscle function. For severely affected patients with extensive paralysis, management consists of protracted intensive care, with detection and treatment of attendant risks not specific to botulism, such as ventilator-associated pneumonia, decubitus ulcers, and psychological trauma. More than 95% of noninfant botulism patients in the United States recover; hospital discharge is often followed by protracted rehabilitative care. The survival rate for infant botulism is near 100%.

CLINICAL DIAGNOSIS AND LABORATORY CONFIRMATION

Rapid clinical diagnosis is essential. A diagnostic aid for botulism, "Clinical Criteria to Trigger Suspicion of Botulism," has been published by botulism consultants at the Centers for Disease Control and Prevention (CDC; accessible at *academic.oup.com/cid/article/66/suppl_1/S38/4780423*). The paralysis of botulism lasts for weeks or months, and administration of equine-source botulinum antitoxin (BAT)—the specific therapy to arrest the progression of paralysis—depends on the correct diagnosis. At this time, laboratory confirmation of botulism, which may require ≥24 h, must take place at a specialized public health laboratory. Therefore, effective, timely treatment relies on rapid clinical diagnosis of botulism in a patient with clinically compatible findings, without awaiting laboratory confirmation. A clinician suspecting noninfant botulism in a patient should immediately contact the state health department's emergency 24-h line. The state will connect the clinician with a botulism clinical consultant at the CDC (or, in Alaska, California, and Colorado, at the state health department), who will review the case with the clinician, assist in the shipping of appropriate specimens to a public health laboratory for definitive diagnosis, and, when indicated, arrange for immediate shipping of BAT from the federal stockpile at no charge. A clinician suspecting infant botulism in a patient should immediately contact the Infant Botulism Treatment and Prevention Program's on-call physician at (510)231-7600, who will provide consultation, assist with specimen collection, and, when indicated, assist with the provision of human-derived botulinum antitoxin (BabyBIG), a specific treatment licensed for treatment of infant botulism.

The neurologic examination is the key to clinical diagnosis of botulism, as it readily uncovers the cranial nerve deficits that are invariably present in botulism and focuses the differential diagnosis. In principle, the distinct syndrome of bilateral cranial palsies and descending flaccid paralysis in a fully conscious patient should render the diagnosis and prompt treatment of botulism straightforward. The presentation of two or more patients with this syndrome is almost pathognomonic, since other illnesses considered during the differential diagnosis of botulism do not produce outbreaks. In practice, however, sporadic (lone) cases of botulism are misdiagnosed, and sometimes the diagnosis is missed even in the setting of an outbreak. In part, these failures may be due to the rarity of botulism and the clinician's unfamiliarity with its presentation. A possible cause of misdiagnosis is failure to perform a complete neurologic examination; indeed, review of some botulism patients' charts reveals documentation of the first neurologic examination, which suggested the correct diagnosis, days after hospital admission. As stated earlier, the combination of ptosis, dysarthria, and perceived gait instability from muscle paralysis in some cases may be misinterpreted as intoxication from alcohol or other substances. In other cases, rapidly progressing botulism may result in pharyngeal collapse and respiratory distress relatively early in the course, leading the clinical team to focus on airway management and primary respiratory diagnoses and thus delaying the neurologic evaluation.

Standard clinical studies, including bloodwork and radiology, are not useful in diagnosing botulism. In contrast to the findings in Guillain-Barré syndrome (GBS; see below), lumbar-puncture cerebrospinal fluid (CSF) values—and specifically the protein level—are usually normal in botulism. The CSF protein level may be very slightly elevated in a minority of botulism cases. The fact that botulism produces no abnormal findings on brain imaging may help rule out rare basilar strokes that produce nonlateralizing symptoms. Serological antibody tests and rapid nerve stimulation tests can rule out myasthenia gravis. Electromyography, when performed by an experienced practitioner, can provide support for the diagnosis. Botulism is indicated by findings consistent with neuromuscular junction blockage, normal axonal conduction, and potentiation with rapid repetitive stimulation in affected muscles.

Once a neurologic examination reveals the cranial nerve palsies of botulism and any additional bilateral flaccid paralysis, the differential diagnosis may include GBS, myasthenia gravis, Lambert-Eaton syndrome, and tick paralysis. Less likely conditions include tetrodotoxin or shellfish poisoning, antimicrobial-associated paralysis, and rarer poisonings. A careful history and physical examination can further narrow the range of diagnoses. GBS is a rare (~1 case per 100,000 population per year in the United States) autoimmune demyelinating polyneuropathy that follows acute infection by *Campylobacter jejuni*, certain viruses, and other bacteria. In 95% of cases, GBS presents as an ascending paralysis. Recent reports from Peru indicate massive outbreaks of GBS of unknown cause, challenging the previously held notion that conditions causing flaccid paralysis other than botulism occur only as sporadic cases. The 5% of GBS cases presenting as the Miller Fisher variant are characterized by the triad of ophthalmoplegia, ataxia, and areflexia, which may resemble early descending paralysis. The CSF protein level is elevated in GBS, but the increase may take place days after symptom onset; thus, normal CSF levels should be taken into account along with the duration of symptoms, and lumbar puncture may need to be repeated. Electromyography performed by an experienced operator may yield findings indicative of GBS and not botulism. Serological antibody tests or rapid nerve stimulation tests can confirm myasthenia gravis; borderline positive Tensilon tests have been reported in botulism patients. In most stroke patients, the physical examination should reveal asymmetric paralysis and upper motor neuron signs; brain imaging can help reveal rare basilar strokes that can produce symmetric bulbar palsies. The history and physical examination should rule out Lambert-Eaton syndrome, which is characterized by proximal limb weakness in patients with advanced cancer.

Laboratory testing confirms clinically diagnosed botulism cases and determines the BoNT serotype causing the disease. In addition, laboratory testing can confirm epidemiologic data by demonstrating presence of BoNT in the suspected food. Botulism cases are confirmed by the laboratory when BoNT is identified in serum or stool specimens or when a BoNT-producing species of *Clostridium* is isolated from stool specimens or wound cultures. Identification of preformed BoNT in food consumed by patients also confirms foodborne botulism.

The gold standard for identification and serotyping of BoNT in clinical or food specimens is the mouse bioassay. The drawback is that this highly sensitive and specific method requires the use of animals. Specimens are injected IP into the mice with and without antitoxin; the mice are then observed for up to 96 h for signs of botulism. If the specimen contains BoNT at levels sufficient to affect the mice quickly, results may be available within 24 h of injection. Low levels of toxin may produce signs later, so mice should be monitored for 4 days after injection. Many in vitro methods have been developed for detection of BoNT and BoNT-producing species of *Clostridium* in clinical and food specimens. For instance, public health laboratories in the United States can use a real-time polymerase chain reaction test that detects *bont* genes encoding serotypes A through G. This test is a useful screening method to determine whether BoNT-producing species of *Clostridium* are present in cultures of clinical specimens, but positive results must be confirmed. Another in vitro method, the Endopep mass spectrometry (Endopep-MS) assay, is highly sensitive and specific and can detect BoNT in clinical specimens and foods. The advantage of Endopep-MS is that it detects active BoNT and therefore represents an ideal alternative to the mouse bioassay. Immune-based assays can provide rapid and sensitive results; their main limitation is that they detect antigens, which may not necessarily represent active BoNT. Cell-based in vitro assays are also a possible alternative to the mouse bioassay as they detect biologic activity of BoNT.

TREATMENT

Botulism

Treatment for botulism consists of two components: meticulous monitoring and supportive care, including admittance to the intensive care unit when indicated, and administration of botulinum antitoxin, the only specific therapy for botulism, as quickly as possible. Paralysis from botulism can be rapidly progressive. Vital capacity, and often hemodynamic parameters, should be frequently monitored and mechanical ventilation instituted immediately if needed. Paralysis induced by BoNT lasts weeks or months, and patients with extensive paralysis require painstaking care to avoid complications associated with protracted immobilization, including respirator-dependent pneumonia, decubitus ulcers, and psychological trauma. Patients who have recovered from severe botulism report that their appearance and immobility often led caregivers to assume they were unconscious; as a consequence, patients were sometimes subjected to painful procedures without warning and to insensitive comments. Signage should remind all caregivers that botulism patients are conscious but “locked in.” Psychological support should be instituted for intubated botulism patients from the outset. With proper supportive care, >95% of botulism patients in the United States recover, even without antitoxin therapy; however, antitoxin, if promptly administered, can substantially reduce the extent and duration of illness (see below).

Botulinum antitoxin is the only specific treatment for botulism. The antitoxin prevents the progression of paralysis but does not reverse existing paralysis. If given early enough in the course of disease, it may avert respiratory compromise, obviate mechanical intubation, and forestall protracted paralysis and hospitalization along with associated complications. Accordingly, it is essential to administer antitoxin as soon as possible. A recent systematic literature review and meta-analysis covering nearly a century of the published literature in noninfant botulism patients confirmed long-known findings from smaller studies by showing significantly reduced mortality rates among patients treated with equine antitoxin, especially when treatment was administered within 48 h of symptom onset. Another large systematic literature review of pediatric noninfant botulism recently showed significantly reduced mortality risk among children treated with equine antitoxin. Published studies have demonstrated a substantial reduction in the duration and severity of illness among patients with infant botulism who are treated with human-derived botulinum antitoxin.

The equine botulinum antitoxin used to treat noninfant botulism consists of antibodies produced in horses immunized with botulinum toxoids (inactivated toxins) and toxins. The antibodies are type-specific (anti-A neutralizes BoNT type A and so forth). The currently licensed antitoxin product in the United States, heptavalent botulinum antitoxin (BAT), contains antibodies to BoNT types A, B, C, D, E, F, and G. These equine antibodies have undergone despeciation to reduce antigenicity and the risk of anaphylaxis to foreign protein. A recent systematic literature review, along with studies of BAT use, indicated that <2% of recipients experience serious adverse reactions. Administration of one vial of BAT elicits circulating antitoxin concentrations sufficient to neutralize toxin levels one to two orders of magnitude higher than those found in the serum of most botulism patients. As noted earlier, clinicians suspecting botulism in a patient should immediately call their state health department’s emergency contact to be put in touch with a botulism clinical consultant who will review the case and assist in its management, including shipment of BAT from the federal stockpile at no charge. The botulinum antitoxin used to treat infants, BabyBIG, consists of human antibodies obtained from hyperimmunized volunteers. The product is licensed for treatment of infant botulism due to BoNT types A and B and, as noted earlier, can be obtained through the Infant Botulism Treatment and Prevention Program.

There is no prophylactic treatment for botulism. Persons who may have been exposed to botulinum toxin should be evaluated by a physician and carefully observed for the development of symptoms of botulism. If symptoms appear, the patient should be treated immediately with botulinum antitoxin.

PREVENTION

No vaccine is licensed for the prevention of botulism. In the United States, a botulinum toxoid vaccine was available through the CDC until 2011, but it was discontinued because of a decline in immunogenicity of some serotypes and an increase in occurrence of moderate local reactions. Several vaccine candidates are currently in clinical trials.

Because most foodborne botulism cases are caused by home-canned or home-preserved foods, the prevention of foodborne botulism depends mainly on proper preparation and preservation that ensures the destruction of spores of BoNT-producing species of *Clostridium* that may be present in the food or on the creation of an environment that will not allow the germination and growth of these spores, such as low pH or low water activity. Water activity is a measure of how much water is free, unbound, and thus available to microorganisms to use for growth. If foods have low water activity, it means they do not have much free water, and growth of *C. botulinum* will be limited or inhibited. Using pressure canners and properly cleaning items employed in the canning process can reduce the risk of foodborne botulism. Among other resources, the *USDA Complete Guide to Home Canning* provides a detailed description of safe home-canning practices. Other ways of preventing foodborne botulism include refrigerating homemade oils infused with garlic or herbs and discarding any of these oils that have not been used after 4 days; maintaining baked potatoes or similar foods wrapped in aluminum foil at temperatures above 140°F until served and then refrigerating leftovers; refrigerating canned or pickled foods after opening; and boiling home-canned foods before eating, especially those foods that are low in acid.

Wound botulism largely affects people who inject drugs, especially black tar heroin. Using safe injection practices may help prevent wound botulism and many other infections, such as HIV and hepatitis C virus infections. Thus, educating injection drug users on the prevention of wound botulism and other infections is vital in protecting their health. As wound botulism can also follow traumatic injuries, keeping wounds clean is key.

The risk factors for infant botulism are not fully understood, but possible sources of spores of BoNT-producing species of *Clostridium* include foods and dust. In most cases of infant botulism, no source of spores is identified. Honey is the only food that has been identified as an epidemiologically associated reservoir of spores of BoNT-producing species of *Clostridium*. Honey should not be fed to infants ≤1 year of age.

GLOBAL CONSIDERATIONS

Botulism has been reported from all parts of the world. The European Centre for Disease Prevention and Control has reported an average of 110 botulism cases each year from 2007 to 2018. During that period, 1315 botulism cases were reported from 25 countries, with the most cases in Italy (311 cases), Romania (239 cases), and Poland (202 cases). Foodborne botulism is the most common form of botulism in Europe. Most laboratory-confirmed cases reported from Italy, Romania, and Poland were due to BoNT serotype B. The country of Georgia has a high incidence of botulism (0.9 case per 100,000 persons) relative to rates in the European Union (<0.1/100,000) and the United States (0.01/100,000). From 1980 to 2002, a total of 879 cases of botulism were reported in Georgia; all of them were foodborne, most were associated with home-preserved vegetables, and the majority were due to serotype B. From 1958 to 1983, 986 foodborne botulism outbreaks affecting 4377 individuals were reported from China. Most cases were due to serotype A and were associated with bean products. Botulism in Thailand has been associated with fermented bamboo shoots and fermented soybeans. In 2006, a large foodborne botulism outbreak associated with bamboo shoots occurred in Thailand and affected 209 people who attended a local festival. In South America, Brazil and Argentina have reported several outbreaks of foodborne botulism. For instance, between 2001 and 2008, Brazil reported 18 outbreaks, most of which were associated with meat-based foods such as home-canned meat, homemade pork liver pâté, and commercially canned liver pâté. From 1994 to 2007, Argentina reported 36 outbreaks, most frequently involving home-canned vegetables. Although reports of foodborne botulism in Africa are rare, 5 outbreaks were reported in South Africa between 1959 and 2002, with the majority due to serotype B and associated with noncommercial foods. In addition, 1 outbreak of 91 cases was reported in Egypt in 1991 and was due to serotype E associated with a traditional salted fish.

Wound botulism cases have been reported most frequently from the United States, next most frequently from the United Kingdom, and occasionally from Italy, France, and Australia. Clusters of wound botulism are rare, but, according to a report from the European Centre for Disease Prevention and Control, 23 cases of wound botulism among people who had injected heroin were reported in Norway and Scotland between December 2014 and February 2015. Other countries that have reported wound botulism cases include Argentina, China, and Ecuador.

Although rarely reported, infant botulism cases have been noted on all continents except Africa. Outside the United States (where there were 2419 cases), Argentina reported the largest number of cases (366) and Australia the next largest number (32) between 1976 and 2006. Canada, Italy, and Japan also reported a relatively large number of cases (27, 26, and 22, respectively).

FURTHER READING

Centers for Disease Control and Prevention: Botulism in the United States, 1899–1996. *Handbook for Epidemiologists, Clinicians, and Laboratory Workers*. Atlanta, Centers for Disease Control and Prevention, 1998.

Centers for Disease Control and Prevention: National Botulism Surveillance. Available at *https://www.cdc.gov/botulism/php/national-botulism-surveillance/*. Accessed December 19, 2023.

Dorner MB et al: A large travel-associated outbreak of iatrogenic botulism in four European countries following intragastric botulinum neurotoxin injections for weight reduction, Türkiye, February to March 2023. Euro Surveill 28:2300203, 2023.

European Centre for Disease Prevention and Control: Botulism. Available at *www.ecdc.europa.eu/en/botulism*. Accessed September 27, 2020.

Fleck-Derderian S et al: The epidemiology of foodborne botulism outbreaks: A systematic review. Clin Infect Dis 66:S73, 2017.

Koepke R et al: Global occurrence of infant botulism, 1976–2006. Pediatrics 122:e73, 2008.

National Center for Home Food Preservation: *USDA Complete Guide to Home Canning, 2015 Revision*. Available at *nchfp.uga.edu/papers/guide/INTRO_HomeCanrev0715.pdf*. Accessed March 18, 2024.

Peck M et al: Historical perspectives and guidelines for botulinum neurotoxin subtype nomenclature. Toxins (Basel) 9:38, 2017.

Pirazzini M et al: Botulinum neurotoxins: Biology, pharmacology, and toxicology. Pharmacol Rev 69:200, 2017.

Rao AK et al: Clinical criteria to trigger suspicion for botulism: An evidence-based tool to facilitate timely recognition of suspected cases during sporadic events and outbreaks. Clin Infect Dis 66:S38, 2018.

Rossetto O et al: Botulinum neurotoxins: Mechanism of action. Handb Exp Pharmacol 263:35, 2021.

Yu PA et al: Safety and improved clinical outcomes in patients treated with new equine-derived heptavalent botulinum antitoxin. Clin Infect Dis 66:S57, 2017.

159

Gas Gangrene and Other Clostridial Infections

Amy E. Bryant, Dennis L. Stevens

The genus *Clostridium* encompasses >60 species that may be commensals of the gut microflora or may cause a variety of infections in humans and animals through the production of a plethora of proteinaceous exotoxins. *C. tetani* and *C. botulinum*, for example, cause specific clinical disease by elaborating single but highly potent toxins. In contrast, *C. perfringens* and *C. septicum* cause aggressive necrotizing infections that are attributable to multiple toxins, including bacterial proteases, phospholipases, and cytotoxins.

ETIOLOGIC AGENT

Vegetative cells of *Clostridium* species are pleomorphic, rod-shaped, and arranged singly or in short chains **(Fig. 159-1)**; the cells have rounded or sometimes pointed ends. Although clostridia stain gram-positive in the early stages of growth, they may appear to be gram-negative or gram-variable later in the growth cycle or in infected tissue specimens. Most strains are motile by means of peritrichous flagella; *C. septicum* swarms on solid media. Nonmotile species include *C. perfringens*, *C. ramosum*, and *C. innocuum*. Most species are obligately anaerobic, although clostridial tolerance to oxygen varies widely; some species (e.g., *C. septicum*, *C. tertium*) will grow but will not sporulate in air.

Clostridia produce more protein toxins than any other bacterial genus, and more than 25 clostridial toxins lethal to mice have been identified. These proteins include neurotoxins, enterotoxins, cytotoxins,

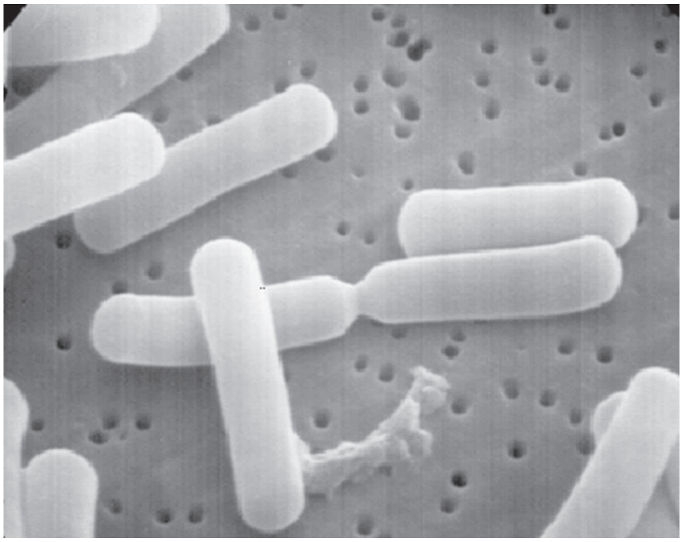

FIGURE 159-1 Scanning electron micrograph of *C. perfringens*.

collagenases, permeases, necrotizing toxins, lipases, lecithinases, hemolysins, proteinases, hyaluronidases, DNases, ADP-ribosyltransferases, and neuraminidases. Botulinum and tetanus neurotoxins are the most potent toxins known, with lethal doses of 0.2–10 ng/kg for humans. Epsilon toxin, a 33-kDa protein produced by *C. perfringens* types B and D, causes edema and hemorrhage in the brain, heart, spinal cord, and kidneys of animals. It is among the most lethal of the clostridial toxins and is considered a potential agent of bioterrorism **(Chap. S4)**. The genomic sequences of some pathogenic clostridia are now available and are likely to facilitate a comprehensive approach to understanding the virulence factors involved in clostridial pathogenesis.

EPIDEMIOLOGY AND TRANSMISSION

Clostridium species are widespread in nature, forming endospores that are commonly found in soil, feces, sewage, and marine sediments. The ecology of *C. perfringens* in soil is greatly influenced by the degree and duration of animal husbandry in a given location and is relevant to the incidence of gas gangrene caused by contamination of wounds with soil. For example, the incidence of clostridial gas gangrene is higher in agricultural regions of Europe than in the Sahara Desert of Africa. Similarly, the incidences of tetanus and food-borne botulism are clearly related to the presence of clostridial spores in soil, water, and many foods. Clostridia are present in large numbers in the indigenous microbiota of the intestinal tract of humans and animals, in the female genital tract, and on the oral mucosa. It should be noted that not all commensal clostridia are toxigenic.

Clostridial infections remain a serious public health concern worldwide. In developing nations, food poisoning, necrotizing enterocolitis, and gas gangrene are common because large portions of the population are poor and have little or no immediate access to health care. These infections remain prevalent in developed countries as well. Gas gangrene commonly follows knife or gunshot wounds or vehicular accidents or develops as a complication of surgery or gastrointestinal carcinoma. Severe clostridial infections have emerged as a health threat to injection drug users and to women undergoing childbirth or abortion. Historically, clostridial gas gangrene has been the scourge of the battlefield. The global political situation portends another possible scenario involving mass casualties of war or terrorism, with extensive injuries conducive to gas gangrene. Therefore, there is an ongoing need to develop novel strategies to prevent or attenuate the course of clostridial infections in both civilians and military personnel. Vaccination against exotoxins important in pathogenesis would be of great benefit in developing nations and could also be used safely in at-risk populations such as the elderly, patients with diabetes who may require lower-limb surgery due to trauma or poor circulation, and those undergoing intestinal surgery. Moreover, a hyperimmune globulin would be a valuable tool for prophylaxis in victims of acute traumatic injury or for attenuation of the spread of infection in patients with established gas gangrene.

CLINICAL SYNDROMES

Life-threatening clostridial infections range from intoxications (e.g., food poisoning, tetanus) to necrotizing enteritis/colitis, bacteremia, myonecrosis, and toxic shock syndrome (TSS). **Tetanus and botulism are discussed in Chaps. 157 and 158, respectively. Colitis due to *C. difficile* is discussed in Chap. 139.**

■ CLOSTRIDIAL WOUND CONTAMINATION

Of open traumatic wounds, 30–80% reportedly are contaminated with clostridial species. In the absence of devitalized tissue, the presence of clostridia does not necessarily lead to infection. In traumatic injuries, clostridia are isolated with equal frequency from both suppurative and well-healing wounds. Thus, diagnosis and treatment of clostridial infection should be based on clinical signs and symptoms and not solely on bacteriologic findings.

■ POLYMICROBIAL INFECTIONS INVOLVING CLOSTRIDIA

Clostridial species may be found in polymicrobial infections also involving microbial components of the endogenous flora. In these infections, clostridia often appear in association with non-spore-forming anaerobes and facultative or aerobic organisms. Head and neck infections, conjunctivitis, brain abscess, sinusitis, otitis, aspiration pneumonia, lung abscess, pleural empyema, cholecystitis, septic arthritis, and bone infections all may involve clostridia. These conditions are often associated with severe local inflammation but may lack the characteristic systemic signs of toxicity and rapid progression seen in other clostridial infections. In addition, clostridia are isolated from ~66% of intraabdominal infections in which the mucosal integrity of the bowel or respiratory system has been compromised. In this setting, *C. ramosum*, *C. perfringens*, and *C. bifermentans* are the most commonly isolated species. Their presence does not invariably lead to a poor outcome. Clostridia have been isolated from suppurative infections of the female genital tract (e.g., ovarian or pelvic abscess) and from diseased gallbladders. Although the most frequently isolated species is *C. perfringens*, gangrene is not typically observed; however, gas formation in the biliary system can lead to emphysematous cholecystitis, especially in diabetic patients. *C. perfringens* in association with mixed aerobic and anaerobic microbes can cause aggressive life-threatening type I necrotizing fasciitis or Fournier's gangrene.

The treatment of mixed aerobic/anaerobic infection of the abdomen, perineum, or gynecologic organs should be based on Gram staining, culture, and antibiotic sensitivity information. Reasonable empirical treatment consists of ampicillin or ampicillin/sulbactam combined with either clindamycin or metronidazole **(Table 159-1)**.

TABLE 159-1 Treatment of Clostridial Infections

CONDITION	ANTIBIOTIC TREATMENT	PENICILLIN ALLERGY	ADJUNCTIVE TREATMENT/NOTE
Wound contamination	None	—	Treatment should be based on clinical signs and symptoms as listed below and not solely on bacteriologic findings.
Polymicrobial anaerobic infections involving clostridia (e.g., abdominal wall, gynecologic)	Ampicillin (2 g IV q4h) ***plus*** Clindamycin (600–900 mg IV q6–8h) ***plus*** Ciprofloxacin (400 mg IV q6–8h)	Vancomycin (1 g IV q12h) ***plus*** Metronidazole (500 mg IV q6h) ***plus*** Ciprofloxacin (400 mg IV q6–8h)	Empirical therapy should be initiated. Therapy should be based on Gram stain and culture results and on sensitivity data when available. Add gram-negative coverage if indicated (see text).
Clostridial sepsis	Penicillin (3–4 mU IV q4–6h) ***plus*** Clindamycin (600–900 mg IV q6–8h)	Clindamycin alone ***or*** Metronidazole (as above) ***or*** Vancomycin (as above)	Transient bacteremia without signs of systemic toxicity may be clinically insignificant.
Gas gangrene	Penicillin G (4 mU IV q4–6h) ***plus*** Clindamycin (600–900 mg IV q6–8h)	Cefoxitin (2 g IV q6h) ***plus*** Clindamycin (600–900 mg IV q6–8h)	Emergent surgical exploration and thorough debridement are extremely important. Hyperbaric oxygen therapy may be considered after surgery and antibiotic initiation.

Broader gram-negative coverage may be necessary if the patient has recently been hospitalized or treated with antibiotics. Such coverage can be obtained by substituting ticarcillin/clavulanic acid, piperacillin/sulbactam, or a carbapenem antibiotic for ampicillin or by adding a fluoroquinolone or an aminoglycoside to the regimen. Empirical treatment should be given for 10–14 days or until the patient's clinical condition improves.

ENTERIC CLOSTRIDIAL INFECTIONS

C. perfringens type A is one of the most common bacterial causes of food-borne illness in the United States and Canada. The foods typically implicated include improperly cooked meat and meat products (e.g., gravy) in which residual spores germinate and proliferate during slow cooling or insufficient reheating. Illness results from the ingestion of food containing at least $\sim 10^8$ viable vegetative cells, which sporulate in the alkaline environment of the small intestine, producing *C. perfringens* enterotoxin in the process. The diarrhea that develops within 7–30 h of ingestion of contaminated food is generally mild and self-limiting; however, in the very young, the elderly, and the immunocompromised, symptoms are more severe and occasionally fatal. Enterotoxin-producing *C. perfringens* has been implicated as an etiologic agent of persistent diarrhea in elderly patients in nursing homes and tertiary-care institutions and has been considered to play a role in antibiotic-associated diarrhea without pseudomembranous colitis.

C. perfringens strains associated with food poisoning possess the gene (*cpe*) coding for enterotoxin, which acts by forming pores in host cell membranes. *C. perfringens* strains isolated from non-food-borne diseases, such as antibiotic-associated and sporadic diarrhea, carry *cpe* on a plasmid that may be transmitted to other strains. Several methods have been described for the detection of *C. perfringens* enterotoxin in feces, including cell culture assay (Vero cells), enzyme-linked immunosorbent assay, reversed-phase latex agglutination, and polymerase chain reaction (PCR) amplification of *cpe*. Each method has its advantages and limitations. Interestingly, spores from these strains are particularly resistant to heat, cold, and chemical preservatives. In addition, the extracellular sialidase produced by *C. perfringens* facilitates pathogenesis.

Enteritis necroticans (gas gangrene of the bowel) is a fulminating clinical illness characterized by extensive necrosis of the intestinal mucosa and wall. Cases can occur sporadically in adults or as epidemics in people of all ages. Enteritis necroticans is caused by α toxin– and β toxin–producing strains of *C. perfringens* type C; β toxin is located on a plasmid and is mainly responsible for pathogenesis. This life-threatening infection causes ischemic necrosis of the jejunum. In Papua New Guinea during the 1960s, enteritis necroticans (known in that locale as *pigbel*) was found to be the most common cause of death in childhood; it was associated with pig feasts and occurred both sporadically and in outbreaks. Intramuscular immunization against the β toxin resulted in a decreased incidence of the disease in Papua New Guinea, although the condition remains common. Enteritis necroticans has also been recognized in the United States, the United Kingdom, Germany (where it is known as *darmbrand*), and other developed nations; especially affected are adults who are malnourished or who have diabetes, alcoholic liver disease, or neutropenia.

Necrotizing enterocolitis, a disease resembling enteritis necroticans but associated with *C. perfringens* type A, has been found in North America in previously healthy adults. It is also a serious gastrointestinal disease of low-birth-weight (premature) infants hospitalized in neonatal intensive care units. The etiology and pathogenesis of this disease have remained enigmatic for more than four decades. Pathologic similarities between necrotizing enterocolitis and enteritis necroticans include the pattern of small-bowel necrosis involving the submucosa, mucosa, and muscularis; the presence of gas dissecting the tissue planes; and the degree of inflammation. In contrast to enteritis necroticans, which most commonly involves the jejunum, necrotizing enterocolitis affects the ileum and frequently the ileocecal valve. Both diseases may manifest as intestinal gas cysts, although this feature is more common in necrotizing enterocolitis. The sources of the gas, which contains hydrogen, methane, and carbon dioxide, are probably the fermentative activities of intestinal bacteria, including clostridia. Epidemiologic data support an important role for *C. perfringens* or other gas-producing microorganisms (e.g., *C. neonatale*, certain other clostridia, or *Klebsiella* species) in the pathogenesis of necrotizing enterocolitis.

Patients with suspected clostridial enteric infection should undergo nasogastric suction and receive IV fluids. Pyrantel is given by mouth, and the bowel is rested by fasting. Benzylpenicillin (1 mU) is given IV every 4 h, and the patient is observed for complications requiring surgery. Patients with mild cases recover without surgical intervention. However, if surgical indications are present (gas in the peritoneal cavity, absent bowel sounds, rebound tenderness, abdominal rigidity), the mortality rate ranges from 35 to 100%; a fatal outcome is due in part to perforation of the intestine.

As pigbel continues to be a common disease in Papua New Guinea, consideration should be given to the use of a *C. perfringens* type C β toxoid vaccine in local areas. Two doses given 3–4 months apart are preventive.

CLOSTRIDIAL BACTEREMIA

Clostridium species are important causes of bloodstream infections. Molecular epidemiologic studies of anaerobic bacteremia have identified *C. perfringens* and *C. tertium* as the two most frequently isolated species; these organisms cause up to 79 and 5%, respectively, of clostridial bacteremias. Occasionally, *C. perfringens* bacteremia occurs in the absence of an identifiable infection at another site. When associated with myonecrosis, bacteremia has a grave prognosis.

C. septicum is also commonly associated with bacteremia. This species is isolated only rarely from the feces of healthy individuals but may be found in the normal appendix. More than 50% of patients whose blood cultures are positive for this organism have some gastrointestinal anomaly (e.g., diverticular disease) or underlying malignancy (e.g., carcinoma of the colon). In addition, a clinically important association of *C. septicum* bacteremia with neutropenia of any origin—and, more specifically, with neutropenic enterocolitis involving the terminal ileum or cecum—has been observed. Patients with diabetes mellitus, severe atherosclerotic cardiovascular disease, or anaerobic myonecrosis (gas gangrene) also may develop *C. septicum* bacteremia. *C. septicum* has been recovered from the bloodstream of cirrhotic patients, as have *C. perfringens*, *C. bifermentans*, and other clostridia. Infections of the bloodstream by *C. sordellii* and *C. perfringens* have been associated with TSS. Of note, Clostridium sordellii has been recently renamed Paeniclostridium sordellii. However, throughout this text, the authors have used the original nomenclature for this pathogen.

Bloodstream infection by *C. tertium*, either alone or in combination with *C. septicum* or *C. perfringens*, can be found in patients with serious underlying disease such as malignancy or acute pancreatitis, with or without neutropenic enterocolitis; the frequency has not been systematically studied. *C. tertium* may present special problems in terms of both identification and treatment. This organism may stain gram-negative; is aerotolerant; and is resistant to metronidazole, clindamycin, and cephalosporins.

Other clostridia from the *C. clostridioforme* group (including *C. clostridioforme*, *C. hathewayi*, and *C. bolteae*) can cause bacteremia.

The clinical importance of recognizing clostridial bacteremia—especially that due to *C. septicum*—and starting appropriate treatment immediately (Table 159-1) cannot be overemphasized. Patients with this condition usually are gravely ill, and infection may metastasize to distant anatomic sites, resulting in spontaneous myonecrosis (see next section). Alternative methods to identify bacteremia-causing clostridial species, such as PCR or other rapid diagnostic tests, are not currently available. Anaerobic blood cultures and Gram's stain interpretation remain the best diagnostic tests at this point.

CLOSTRIDIAL SKIN AND SOFT TISSUE INFECTIONS

Histotoxic clostridial species such as *C. perfringens*, *C. histolyticum*, *C. septicum*, C. *novyi*, and *C. sordellii* cause aggressive necrotizing infections of the skin and soft tissues. These infections are attributable in part to the elaboration of bacterial proteases, phospholipases, and cytotoxins. Necrotizing clostridial soft tissue infections are rapidly progressive and are characterized by marked tissue destruction, gas in the tissues, and shock; they frequently end in death. Severe pain, crepitus, brawny induration with rapid progression to skin sloughing, violaceous bullae, and marked tachycardia are characteristics found in the majority of patients.

Clostridial Myonecrosis (Gas Gangrene) • TRAUMATIC GAS GANGRENE

C. perfringens myonecrosis (gas gangrene) is one of the most fulminant gram-positive bacterial infections of humans. Even with appropriate antibiotic therapy and management in an intensive care unit, tissue destruction can progress rapidly. Gas gangrene is accompanied by bacteremia, hypotension, and multiorgan failure and is invariably fatal if untreated. Gas gangrene is a true emergency and requires immediate surgical debridement.

The development of gas gangrene requires an anaerobic environment and contamination of a wound with spores or vegetative organisms. Devitalized tissue, foreign bodies, and ischemia reduce locally available oxygen levels and favor outgrowth of vegetative cells and spores. Thus, conditions predisposing to traumatic gas gangrene include crush-type injury, laceration of large or medium-sized arteries, and open fractures of long bones that are contaminated with soil or bits of clothing containing the bacterial spores. Gas gangrene of the abdominal wall and flanks follows penetrating injuries such as knife or gunshot wounds that are sufficient to compromise intestinal integrity, with resultant leakage of the bowel contents into the soft tissues. Proximity to fecal sources of bacteria is a risk factor for cases following hip surgery, adrenaline injections into the buttocks, or amputation of the leg for ischemic vascular disease. In addition, cutaneous gas gangrene caused by *C. perfringens*, *C. novyi*, and *C. sordellii* has been described in the United States and northern Europe among persons injecting black-tar heroin subcutaneously.

The incubation period for traumatic gas gangrene can be as short as 6 h and is usually <4 days. The infection is characterized by the sudden onset of excruciating pain at the affected site and the rapid development of a foul-smelling wound containing a thin serosanguineous discharge and gas bubbles. Brawny edema and induration develop and give way to cutaneous blisters containing bluish to maroon-colored fluid. Such tissue later may become liquefied and slough. The margin between healthy and necrotic tissue often advances several inches per hour despite appropriate antibiotic therapy, and radical amputation remains the single best life-saving intervention. Shock and organ failure frequently accompany gas gangrene; when patients become bacteremic, the mortality rate exceeds 50%.

Diagnosis of traumatic gas gangrene is not difficult because the infection always begins at the site of significant trauma, is associated with gas in the tissue, and is rapidly progressive. Gram staining of drainage or tissue biopsy is usually definitive, demonstrating large gram-positive (or gram-variable) rods, an absence of inflammatory cells, and widespread soft tissue necrosis.

SPONTANEOUS (NONTRAUMATIC) GAS GANGRENE

Spontaneous gas gangrene generally occurs via hematogenous seeding of normal muscle with histotoxic clostridia—principally *C. perfringens*, *C. septicum*, and *C. novyi* and occasionally *C. tertium*—from a gastrointestinal tract portal of entry (as in colonic malignancy, inflammatory bowel disease, diverticulitis, necrotizing enterocolitis, cecitis, or distal ileitis or after gastrointestinal surgery, including colonoscopic polypectomy). These gastrointestinal pathologies permit bacterial access to the bloodstream; consequently, aerotolerant *C. septicum* can proliferate in normal tissues. Patients surviving bacteremia or spontaneous gangrene due to *C. septicum* should undergo aggressive diagnostic studies to rule out gastrointestinal pathology.

Additional predisposing host factors include leukemia, lymphoproliferative disorders, cancer chemotherapy, radiation therapy, and AIDS. Cyclic, congenital, or acquired neutropenia also is strongly associated with an increased incidence of spontaneous gas gangrene due to *C. septicum*; in such cases, necrotizing enterocolitis, cecitis, or distal ileitis is common, particularly among children.

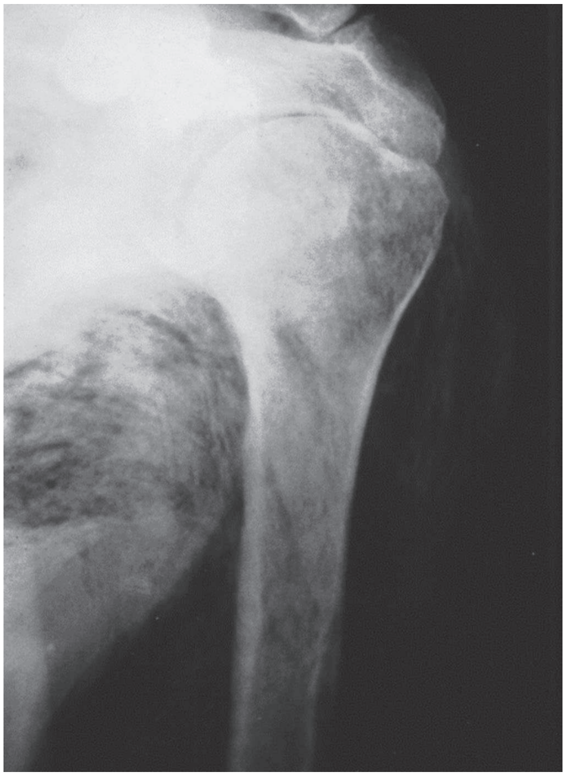

FIGURE 159-2 Radiograph of patient with spontaneous gas gangrene due to *C. septicum*, demonstrating gas in the affected arm and shoulder.

The first symptom of spontaneous gas gangrene may be confusion followed by the abrupt onset of excruciating pain in the absence of trauma. These findings, along with fever, should heighten suspicion of spontaneous gas gangrene. However, because of the lack of an obvious portal of entry, the correct diagnosis is frequently delayed or missed. The infection is characterized by rapid progression of tissue destruction with demonstrable gas in the tissue **(Fig. 159-2)**. Swelling increases, and bullae filled with clear, cloudy, hemorrhagic, or purplish fluid appear. The surrounding skin has a purple hue, which may reflect vascular compromise resulting from the diffusion of bacterial toxins into surrounding tissues. Invasion of healthy tissue rapidly ensues, with quick progression to shock and multiple-organ failure. Mortality rates in this setting range from 67 to 100% among adults; among children, the mortality rate is 59%, with the majority of deaths occurring within 24 h of onset.

PATHOGENESIS OF GAS GANGRENE

In traumatic gas gangrene, organisms are introduced into devitalized tissue. It is important to recognize that for *C. perfringens* and *C. novyi*, trauma must be sufficient to interrupt the blood supply and thereby to establish an optimal anaerobic environment for growth of these species. These conditions are not strictly required for the more aerotolerant species such as *C. septicum* and *C. tertium*, which can seed normal tissues from gastrointestinal lesions. Once introduced into an appropriate niche, the organisms proliferate locally and elaborate exotoxins.

The major *C. perfringens* extracellular toxins implicated in gas gangrene are α toxin and θ toxin. A lethal hemolysin that has both phospholipase C and sphingomyelinase activities, α toxin has been implicated as the major virulence factor of *C. perfringens*: immunization of mice with the C-terminal domain of α toxin provides protection against lethal challenge with *C. perfringens*, and isogenic α toxin–deficient mutant strains of *C. perfringens* are not lethal in a murine model of gas gangrene. Recently, a human single chain recombinant antibody against α toxin has been developed having significant preventative and therapeutic efficacy in mice.

It has been shown in experimental models that the severe pain, rapid progression, marked tissue destruction, and absence of neutrophils in *C. perfringens* gas gangrene are attributable in large part to

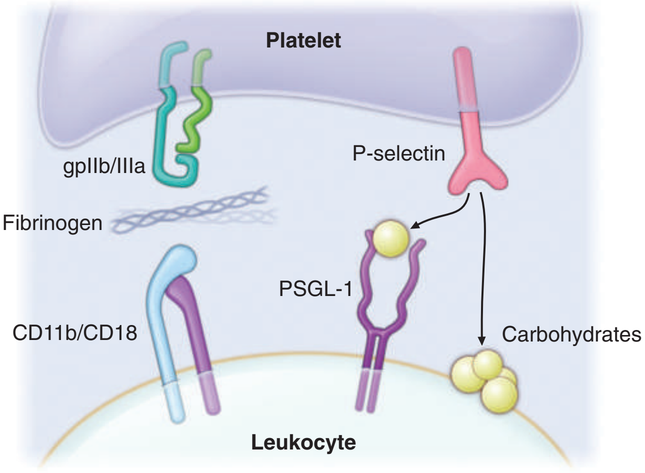

FIGURE 159-3 Schematic illustration of the molecular mechanisms of *C. perfringens* α toxin–induced platelet/neutrophil aggregates. Homotypic aggregates of platelets (not shown) and heterotypic aggregates of platelets and leukocytes are due to α toxin–induced activation of the platelet fibrinogen receptor gpIIb/IIIa and upregulation of leukocyte CD11b/CD18. Binding of fibrinogen (*red*) bridges the connection between these adhesion molecules on adjacent cells. An auxiliary role for α toxin–induced upregulation of platelet P-selectin and its binding to leukocyte P-selectin glycoprotein ligand 1 (PSGL-1) or other leukocyte surface carbohydrates also has been demonstrated.

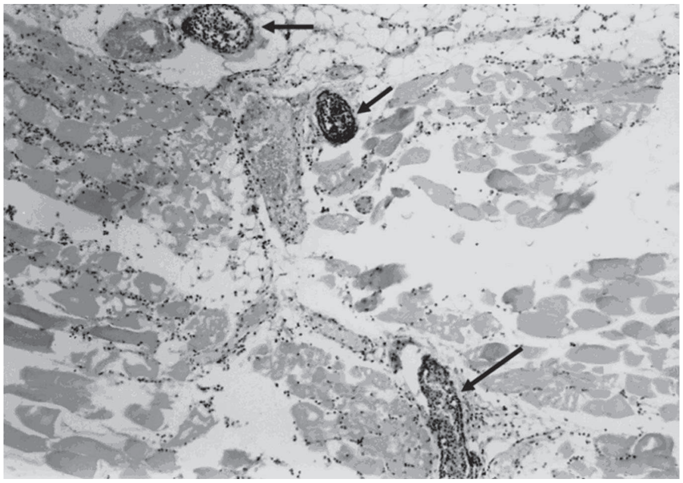

FIGURE 159-4 Histopathology of experimental gas gangrene due to *C. perfringens*, demonstrating widespread muscle necrosis, a paucity of leukocytes in infected tissues, and accumulation of leukocytes in adjacent vessels (*arrows*). These features are due to the effects of α and θ toxins on muscle cells, platelets, leukocytes, and endothelial cells.

α toxin–induced occlusion of blood vessels by heterotypic aggregates of platelets and neutrophils. The formation of these aggregates, which occurs within minutes, is largely mediated by α toxin's ability to activate the platelet adhesion molecule gpIIb/IIIa **(Fig. 159-3)**; the implication is that platelet glycoprotein inhibitors (e.g., eptifibatide, abciximab) may be therapeutic for maintaining tissue blood flow.

C. perfringens θ toxin (*perfringolysin,* PFO) is a member of the thiol-activated cytolysin family known as cholesterol-dependent cytolysins, which includes streptolysin O from group A *Streptococcus*, pneumolysin from *Streptococcus pneumoniae*, and several other toxins. Cholesterol-dependent cytolysins bind as oligomers to cholesterol in host cell membranes. At high concentrations, these toxins form ring-like pores resulting in cell lysis. At sublytic concentrations, PFO hyperactivates phagocytes and vascular endothelial cells. PFO-mediated activation of the macrophage inflammasome, with production of IL-1β, has also been reported.

Cardiovascular collapse and end-organ failure occur late in the course of *C. perfringens* gas gangrene and are largely attributable to both direct and indirect effects of α and θ toxins. In experimental models, θ toxin causes markedly reduced systemic vascular resistance but increased cardiac output (i.e., "warm shock"), probably via induction of endogenous mediators (e.g., prostacyclin, platelet-activating factor) that cause vasodilation. This effect is similar to that observed in gram-negative sepsis. In sharp contrast, α toxin directly suppresses myocardial contractility; the consequence is profound hypotension due to a sudden reduction in cardiac output. The roles of other endogenous mediators, such as cytokines (e.g., tumor necrosis factor, interleukins 1 and 6) and vasodilators (e.g., bradykinin) have not been fully elucidated.

C. septicum produces four main toxins—α toxin (lethal, hemolytic, necrotizing activity), β toxin (DNase), γ toxin (hyaluronidase), and Δ toxin (septicolysin, an oxygen-labile hemolysin)—as well as a protease and a neuraminidase. Unlike the α toxin of *C. perfringens*, that of *C. septicum* does not possess phospholipase activity. The mechanisms remain to be fully elucidated, but it is likely that each of these toxins contributes uniquely to *C. septicum* gas gangrene.

TREATMENT

Gas Gangrene

Patients with suspected gas gangrene (either traumatic or spontaneous) should undergo prompt surgical inspection of the infected site. Direct examination of a gram-stained smear of the involved tissues is of major importance. Characteristic histologic findings in clostridial gas gangrene include widespread tissue destruction, a paucity of leukocytes in infected tissues in conjunction with an accumulation of leukocytes in adjacent vessels **(Fig. 159-4)**, and the presence of gram-positive rods (with or without spores). Computed tomography (CT) and magnetic resonance imaging (MRI) are invaluable for determining whether the infection is localized or is spreading along fascial planes, and needle aspiration or punch biopsy may provide an etiologic diagnosis in at least 20% of cases. However, these techniques should not replace surgical exploration, Gram's staining, and histopathologic examination. When spontaneous gas gangrene is suspected, blood should be cultured since bacteremia usually precedes cutaneous manifestations by several hours.

For patients with evidence of clostridial gas gangrene, thorough emergent surgical debridement is of extreme importance. All devitalized tissue should be widely resected back to healthy viable muscle and skin so as to remove conditions that allow anaerobic organisms to continue proliferating. Closure of traumatic wounds or compound fractures should be delayed for 5–6 days until it is certain that these sites are free of infection.

Antibiotic treatment of traumatic or spontaneous gas gangrene (Table 159-1) consists of the administration of penicillin and clindamycin for 10–14 days. Penicillin is recommended on the basis of in vitro sensitivity data; clindamycin is recommended because of its superior efficacy over penicillin in animal models of *C. perfringens* gas gangrene and in some clinical reports. Controlled clinical trials comparing the efficacy of these agents in humans have not been performed. In the penicillin-allergic patient, clindamycin may be used alone. The superior efficacy of clindamycin is probably due to its ability to inhibit bacterial protein toxin production, its insensitivity to the size of the bacterial load or the stage of bacterial growth, and its ability to modulate the host immune response.

Although *C. perfringens* remains largely susceptible to first-line antibiotics, antibiotic resistance has been reported. Case reports from the United Kingdom and from Spain found clindamycin-resistant *C. perfringens* in cellulitis and in a spontaneous abscess, respectively. Larger studies from Canada and Taiwan also showed increasing resistance to clindamycin among bloodstream isolates. In 2014, Marchand-Austin et al. published a 2-year prospective Canadian study that examined antimicrobial susceptibility of anaerobic bacteria isolated from blood, body fluids, and abscesses. Of 1412 isolates submitted for susceptibility testing, 68 were *C. perfringens*. Of these, all were universally susceptible to penicillin, but 3.8% were clindamycin-resistant. Notably, for *Clostridium* species other than *C. perfringens* (n = 289), 14.2% were penicillin-resistant and 21.6% clindamycin-resistant. A more recent study from Iran found that 21.2% of *C. perfringens* isolates were resistant to penicillin. Lastly, a 2019 study from Hungary found resistance to penicillin (2.6%) and clindamycin (3.8%) among *C. perfringens* isolates (n = 313) from tissues with gas gangrene. Among the non-perfringens gas gangrene isolates (n = 59), higher resistance to penicillin and clindamycin was observed (6.8% and 8.5%, respectively). These findings, though not universal, highlight the importance of good anaerobic microbiology susceptibility testing to provide up-to-date information to guide optimal clinical management decisions for clostridial infections.

C. tertium is resistant to penicillin, cephalosporins, and clindamycin. Appropriate antibiotic therapy for *C. tertium* infection is vancomycin (1 g every 12 h IV) or metronidazole (500 mg every 8 h IV).

The value of adjunctive treatment with hyperbaric oxygen (HBO) for gas gangrene remains controversial. Basic science studies suggest that HBO can inhibit the growth of *C. perfringens* but not that of the more aerotolerant *C. septicum*. In vitro, blood and macerated muscle inhibit the bactericidal potential of HBO. Numerous studies in animals demonstrate little efficacy of HBO alone, whereas antibiotics alone—especially those that inhibit bacterial protein synthesis—confer marked benefits. Addition of HBO to the therapeutic regimen provides some additional benefit, but only if surgery and antibiotic administration precede HBO treatment.

In conclusion, gas gangrene is a rapidly progressive infection whose outcome depends on prompt recognition, emergent surgery, and timely administration of antibiotics that inhibit toxin production. Gas gangrene associated with bacteremia probably represents a later stage of illness and is associated with the worst outcomes. Emergent surgical debridement is crucial to ensure survival, and ancillary procedures (e.g., CT or MRI) or transport to HBO units should not delay this intervention. Some trauma centers associated with HBO units may have special expertise in managing these aggressive infections, but proximity and speed of transfer must be carefully weighed against the need for haste.

PROGNOSIS OF GAS GANGRENE The prognosis for patients with gas gangrene is more favorable when the infection involves an extremity rather than the trunk or visceral organs, since debridement of the latter sites is more difficult. Gas gangrene is most likely to progress to shock and death in patients with associated bacteremia and intravascular hemolysis. Mortality rates are highest for patients in shock at the time of diagnosis. Mortality rates are relatively high among patients with spontaneous gas gangrene, especially that due to *C. septicum*. Survivors of gas gangrene may undergo multiple debridements and face long periods of hospitalization and rehabilitation.

PREVENTION OF GAS GANGRENE Initial aggressive debridement of devitalized tissue can reduce the risk of gas gangrene in contaminated deep wounds. Interventions to be avoided include prolonged application of tourniquets and surgical closure of traumatic wounds; patients with compound fractures are at significant risk for gas gangrene if the wound is closed surgically. Vaccination against α toxin is protective in experimental animal models of *C. perfringens* gas gangrene but has not been investigated in humans. In addition, as mentioned above, a hyperimmune globulin would represent a significant advance for prophylaxis in victims of acute traumatic injury or for attenuation of the spread of infection in patients with established gas gangrene.

Toxic Shock Syndrome Clostridial infection of the endometrium, particularly that due to *C. sordellii*, can develop after gynecologic procedures, childbirth, or abortion (spontaneous or elective, surgical, or medical) and, once established, proceeds rapidly to TSS and death. Systemic manifestations, including edema, effusions, profound leukocytosis, and hemoconcentration, are followed by the rapid onset of hypotension and multiple-organ failure. Elevation of the hematocrit to 75–80% and leukocytosis of 50,000–200,000 cells/μL, with a left shift, are characteristic of *C. sordellii* infection. Pain may not be a prominent feature, and fever is typically absent. In one series, 18% of 45 cases of *C. sordellii* infection were associated with normal childbirth, 11% with medically induced abortion, and 0.4% with spontaneous abortion; the case-fatality rate was 100% in these groups. Of the infections in this series that were not related to gynecologic procedures or childbirth, 22% occurred in injection drug users, and 50% of these patients died. Other infections followed trauma or surgery (42%), mostly in healthy persons, and 53% of these patients died. Overall, the mortality rate was 69% (31 of 45 cases). Of patients who succumbed, 85% died within 2–6 days after infection onset or following procedures. Rapidly fatal spontaneous *C. bifermentans* necrotizing endometritis with toxic shock, leukemoid reaction, and capillary leak has also been described.

Early diagnosis of *C. sordellii* infections often proves difficult for several reasons. First, the prevalence of these infections is low. Second, the initial symptoms are nonspecific and frankly misleading. Early in the course, the illness resembles any number of infectious diseases, including viral syndromes. Given these vague symptoms and an absence of fever, physicians usually do not aggressively pursue additional diagnostic tests. The absence of local evidence of infection and the lack of fever make early diagnosis of *C. sordellii* infection particularly problematic in patients who develop deep-seated infection following childbirth, therapeutic abortion, gastrointestinal surgery, or trauma. Such patients are frequently evaluated for pulmonary embolization, gastrointestinal bleeding, pyelonephritis, or cholecystitis. Unfortunately, such delays in diagnosis increase the risk of death, and as in most necrotizing soft tissue infections, patients are hypotensive with evidence of organ dysfunction by the time local signs and symptoms become apparent. In contrast, infection is more readily suspected in injection drug users presenting with local swelling, pain, and redness at injection sites; early recognition probably contributes to the lower mortality rates in this group.

Physicians should suspect *C. sordellii* infection in patients who present within 2–7 days after injury, surgery, drug injection, childbirth, or abortion and who report pain, nausea, vomiting, and diarrhea but are afebrile. There is little information regarding appropriate treatment for *C. sordellii* infections. In fact, the interval between onset of symptoms and death is often so short that there is little time to initiate empirical antimicrobial therapy. Indeed, anaerobic cultures of blood and wound aspirates are time-consuming, and many hospital laboratories do not routinely perform antimicrobial sensitivity testing on anaerobes. Antibiotic susceptibility data from older studies suggest that *C. sordellii*, like most clostridia, is susceptible to β-lactam antibiotics, clindamycin, tetracycline, and chloramphenicol but is resistant to aminoglycosides and sulfonamides. Antibiotics that suppress toxin synthesis (e.g., clindamycin) may possibly prove useful as therapeutic adjuncts since they are effective in necrotizing infections due to other toxin-producing gram-positive organisms. With the adoption of restrictive legislation that reduces or prohibits access to medically supervised abortions, the incidence of these deadly clostridial infections could increase as patients undergo unsafe pregnancy termination.

Other Clostridial Skin and Soft-Tissue Infections *Crepitant cellulitis* (also called *anaerobic cellulitis*) occurs principally in diabetic patients and characteristically involves subcutaneous tissues or retroperitoneal tissues, whereas the muscle and fascia are not involved. This infection can progress to fulminant systemic disease.

Cases of *C. histolyticum* infection with cellulitis, abscess formation, or endocarditis have also been documented in injection drug users. Endophthalmitis due to *C. sordellii* or *C. perfringens* has been described. *C. ramosum* is also isolated frequently from clinical specimens, including blood and both intraabdominal and soft tissues. This species may be resistant to clindamycin and multiple cephalosporins.

FURTHER READING

Aldape MJ et al: *Clostridium sordellii* infection: Epidemiology, clinical findings, and current perspectives on diagnosis and treatment. Clin Infect Dis 43:1436, 2006.

Aronoff DM et al: Infections caused by *Clostridium perfringens* and *Paeniclostridium sordellii* after unsafe abortion. Lancet Infect Dis 23:e48, 2023. Erratum in: Lancet Infect Dis 22:e310, 2022.

Bodey GP et al: Clostridial bacteremia in cancer patients. A 12-year experience. Cancer 67:1928, 1991.

Bos J et al: Fatal necrotizing colitis following a foodborne outbreak of enterotoxigenic *Clostridium perfringens* type A infection. Clin Infect Dis 40:e78, 2005.

Bryant AE et al: Clostridial gas gangrene II: Phospholipase C–induced activation of platelet gpIIb/IIIa mediates vascular occlusion and myonecrosis in *C. perfringens* gas gangrene. J Infect Dis 182:808, 2000.

Li J et al: *Clostridium perfringens* sporulation and sporulation-associated toxin production. Microbiol Spectr 4:10.1128/microbiolspec.TBS-0022-2015, 2016.

Li J et al: NanJ is the major sialidase for *Clostridium perfringens* Type F food poisoning strain 01E809. Infect Immun 91:e0005323, 2023.

Marchand-Austin A et al: Antimicrobial susceptibility of clinical isolates of anaerobic bacteria in Ontario, 2010-2011. Anaerobe 28:120, 2014.

Obladen M: Necrotizing enterocolitis—150 years of fruitless search for the cause. Neonatology 96:203, 2009

Sayeed S et al: Beta toxin is essential for the intestinal virulence of *Clostridium perfringens* type C disease isolate CN3685 in a rabbit ileal loop model. Mol Microbiol 67:15, 2008.

Smith LDS, Williams BL: *The Pathogenic Anaerobic Bacteria*, 3rd ed. Springfield, IL, Charles C Thomas, 1984.

Stevens DL, Bryant AE: Necrotizing soft tissue infections. N Engl J Med 377:2253, 2017.

Stevens DL et al: Clostridium, in *Manual of Clinical Microbiology*, 11th ed, J Versalovic (ed). ASM Press, 2014, pp. 940–966.

Stevens DL et al: Practice guidelines for the diagnosis and management of skin and soft tissue infections: 2014 update by the Infectious Diseases Society of America. *Clin Infect Dis* 59:e10, 2014.

Wang C et al: Hyperbaric oxygen for treating wounds: A systematic review of the literature. Arch Surg 138:272, 2003.

Section 6 Diseases Caused by Gram-Negative Bacteria

160 Meningococcal Infections

Manish Sadarangani, Andrew J. Pollard

DEFINITION

Infection with *Neisseria meningitidis* most commonly manifests as asymptomatic colonization in the nasopharynx of healthy adolescents and adults. Invasive disease occurs rarely, usually presenting as either bacterial meningitis or meningococcal septicemia. Patients may also present with occult bacteremia, pneumonia, septic arthritis, conjunctivitis, and chronic meningococcemia.

ETIOLOGY AND MICROBIOLOGY

N. meningitidis is a gram-negative aerobic diplococcus that colonizes humans only and causes disease after transmission to a susceptible individual. Several related neisserial organisms have been recognized, including the pathogen *N. gonorrhoeae* and the commensals *N. lactamica*, *N. flavescens*, *N. mucosa*, *N. sicca*, and *N. subflava*. *N. meningitidis* is a catalase- and oxidase-positive organism that utilizes glucose and maltose to produce acid.

Meningococci associated with invasive disease are usually encapsulated with polysaccharide, and the antigenic nature of the capsule determines an organism's capsular group (serogroup) **(Table 160-1)**.

TABLE 160-1 Structure of the Polysaccharide Capsule of Common Disease-Causing Meningococci

MENINGOCOCCAL CAPSULAR GROUP	CHEMICAL STRUCTURE OF OLIGOSACCHARIDE	CURRENT DISEASE EPIDEMIOLOGY
A	2-Acetamido-2-deoxy-D-mannopyranosyl phosphate	Epidemic disease mainly in sub-Saharan Africa; sporadic cases worldwide
B	α-2,8-*N*-acetylneuraminic acid	Sporadic cases worldwide; propensity to cause hyperendemic disease
C	α-2,9-*O*-acetylneuraminic acid	Small outbreaks and sporadic disease
Y	4-*O*-α-D-glucopyranosyl-*N*-acetylneuraminic acid	Sporadic disease and occasional small institutional outbreaks
W	4-*O*-α-D-galactopyranosyl-*N*-acetylneuraminic acid	Sporadic disease; outbreaks of disease associated with mass gatherings; epidemics in sub-Saharan Africa
X	(α1→4) *N*-acetyl-D-glucosamine-1-phosphate	Sporadic disease and large outbreaks in the meningitis belt of Africa

In total, 12 capsular groups have been identified (A–C, X–Z, E, W, H–J, and L), but just six of these—A, B, C, X, Y, and W (formerly W135)—account for the majority of cases of invasive disease. Group D is often listed as the thirteenth capsular group but has been identified as an unencapsulated variant of group C. Meningococci are commonly isolated from the nasopharynx in studies of carriage; the lack of capsule often is a result of phase variation of capsule expression, but as many as 16% of isolates lack the genes for capsule synthesis and assembly. These "capsule-null" meningococci and those that express capsules other than A, B, C, X, Y, and W are only rarely associated with invasive disease and are most commonly identified in the nasopharynx of asymptomatic carriers.

Beneath the capsule, meningococci are surrounded by an outer phospholipid membrane containing lipopolysaccharide (LPS, endotoxin) and multiple outer-membrane proteins **(Figs. 160-1 and 160-2)**. Antigenic variability in porins expressed in the outer membrane defines the serotype (PorB) and serosubtype (PorA) of the organism, and structural differences in LPS determine the immunotype. Serologic methods for typing meningococci are restricted by the limited availability of serologic reagents that can distinguish among the organisms' highly variable surface proteins. Where available, high-throughput antigen gene sequencing has superseded serology for

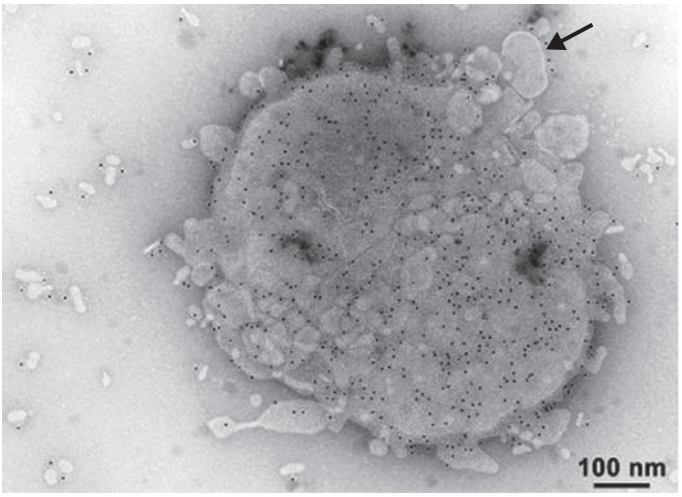

FIGURE 160-1 Electron micrograph of *Neisseria meningitidis*. Black dots are gold-labeled polyclonal antibodies binding surface opacity proteins. Blebs of outer membrane can be seen being released from the bacterial surface (*arrow*). *(Photo courtesy of D. Ferguson, Oxford University.)*

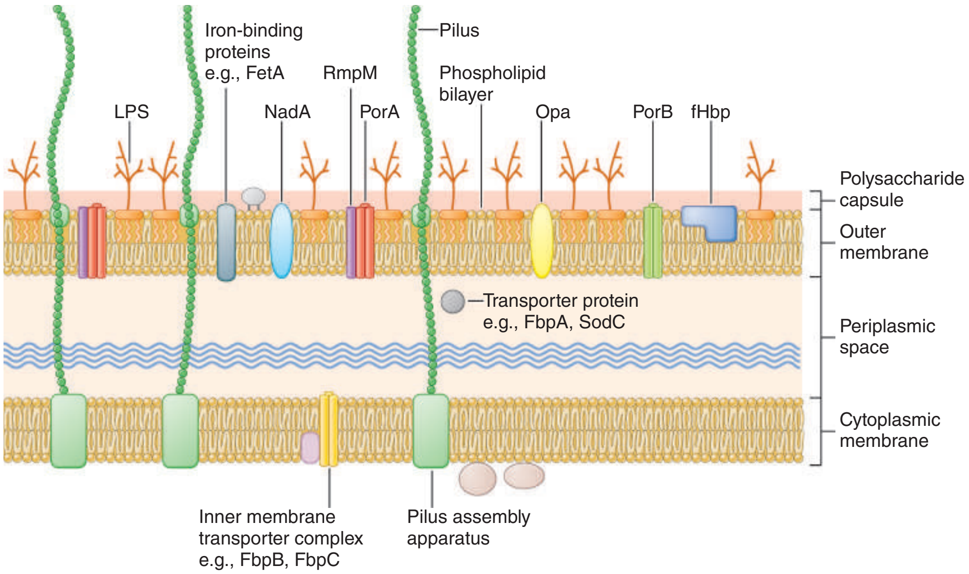

FIGURE 160-2 Cross-section through surface structures of *Neisseria meningitidis*. LPS, lipopolysaccharide. *(Reproduced with permission from M Sadarangani, AJ Pollard: Serogroup B meningococcal vaccines–an unfinished story. Lancet Infect Dis 10:112, 2010.)*

meningococcal typing. A large database of antigen gene sequences for the outer-membrane proteins PorA, PorB, FetA, Opa, NadA, neisserial heparin binding antigen (NHBA), and factor H–binding protein (fHbp) is available online (*pubmlst.org/organisms/neisseria-spp*). The number of specialized iron-regulated proteins found in the meningococcal outer membrane (e.g., FetA and transferrin-binding proteins) highlights the organisms' dependence on iron from human sources. A thin peptidoglycan cell wall separates the outer membrane from the cytoplasmic membrane.

The structure of meningococcal populations involved in local and global spread was first studied with multilocus enzyme electrophoresis (MLEE), which characterizes isolates according to differences in the electrophoretic mobility of cytoplasmic enzymes, followed by multilocus sequence typing (MLST), in which meningococci are characterized by sequence types assigned on the basis of sequences of internal fragments of seven housekeeping genes and, more recently, whole genome sequencing (>60,000 genomes are listed in PubMLST [*https://pubmlst.org/organisms/neisseria-spp/*]). While many distinct genotypes exist, a limited number of hyperinvasive lineages of *N. meningitidis* have been recognized, persist over decades, and are responsible for the majority of cases of invasive meningococcal disease worldwide. Hyperinvasive lineages may be associated with more than one capsular group. The apparent genetic stability of these meningococcal clones over decades and during wide geographic spread indicates that they are well adapted to the nasopharyngeal environment of the host and to efficient transmission.

The group B meningococcal genome is >2 megabases in length and contains 2158 coding regions. Many genes undergo phase variation that makes it possible to control their expression; this capacity is likely to be important in meningococcal adaptation to the host environment and evasion of the immune response. Meningococci can obtain DNA from their environment and can acquire new genes—including the capsular operon—such that *capsule switching* from one capsular group to another can occur.

■ EPIDEMIOLOGY

Patterns of Disease Up to 500,000 cases of meningococcal disease are thought to occur worldwide each year, although the numbers have been declining recently as a result of both immunization programs and secular trends. About 10% of affected individuals die. There are several patterns of disease: epidemic, outbreak (small clusters of cases), hyperendemic, and sporadic or endemic.

Epidemics have continued since the original descriptions of meningococcal disease, especially affecting the sub-Saharan meningitis belt of Africa, where tens to hundreds of thousands of cases (caused mainly by capsular group A but also by capsular groups C, W, and X) may be reported over a season and rates may be as high as 1000 cases per 100,000 population. Capsular group A epidemics took place in Europe and North America after the First and Second World Wars, and capsular group A outbreaks have been documented over the past 40 years in New Zealand, China, Nepal, Mongolia, India, Pakistan, Poland, and Russia. However, 65% of outbreaks reported in the meningitis belt between 2010 and 2017 were caused by capsular group C and 35% by capsular group W meningococci, following an immunization campaign to control capsular group A outbreaks. New vaccines covering A, C, W, Y, and X are becoming available globally to extend control of outbreaks.

Clusters of cases occur where there is an opportunity for increased transmission—i.e., in closed or semi-closed communities such as schools, colleges, universities, military training centers, and refugee camps. Over the past 4 decades, such clusters have been especially strongly linked with a particular clone (sequence type 11) that is mainly associated with capsular group C or W but was first described in association with capsular group B. Clusters of capsular group W disease associated with the Hajj pilgrimage in 2000/2001 led to a requirement for vaccination against meningococcal disease for travel to Saudi Arabia. Wider and more prolonged community outbreaks (hyperendemic disease) due to single clones of capsular group B meningococci account for ≥10 cases per 100,000. Regions affected in the past 35 years include the U.S. Pacific Northwest, New Zealand (both islands), and the province of Normandy in France.

Most countries experience predominantly sporadic cases (0.0–2.8 cases per 100,000 population, but recently rates up to 10 cases per 100,000 have been reported in Africa), with many different disease-causing clones involved and usually no clear epidemiologic link between one case and another. The disease rate and the distribution of meningococcal strains vary in different regions of the world and also in any one location over time. For example, in the United States, the rate of meningococcal disease fell from 1.2 cases per 100,000 population in 1997 to 0.06 cases per 100,000 in 2021 **(Fig. 160-3)**. Meningococcal disease in the United States was previously dominated by capsular groups B and C; however, in 2011–2021, group B alone was predominant in children age <5 years, whereas disease in children over 11 years of age, adolescents, and adults was dominated by capsular groups C, W, and Y **(Fig. 160-4)**. There has been a sharp, but small, increase in capsular group Y starting in 2022, with the highest rates in Americans of black ethnicity, individuals over 30 years of age, and those living with HIV. In contrast, rates of disease in England and Wales rose to >5 cases per 100,000 during the 1990s because of an increase in cases caused by the ST11 capsular group C clone. A mass immunization program against capsular group C was undertaken beginning in 1999 and resulted in a large impact against the disease, leaving capsular group B meningococci as the predominant cause of infection in the past quarter century (87% of United Kingdom cases in 2021–2022). Introduction of a group B meningococcal (MenB) vaccine for infants in the United Kingdom in 2015 also led to a significant reduction in group B cases. A hyperinvasive ST11 clone bearing a W capsule emerged in South America and spread to various countries in Europe and in Australia with cases peaking in 2016 in the United Kingdom. During the same decade, increases in capsular group Y disease were noted in various countries including Europe, Canada, and South Africa, highlighting the continuing emergence and re-emergence of capsular groups and genotypes over time. Nevertheless, over the past 15 years, most industrialized nations have observed a decrease in meningococcal disease linked to the introduction of immunization against capsular group C meningococci

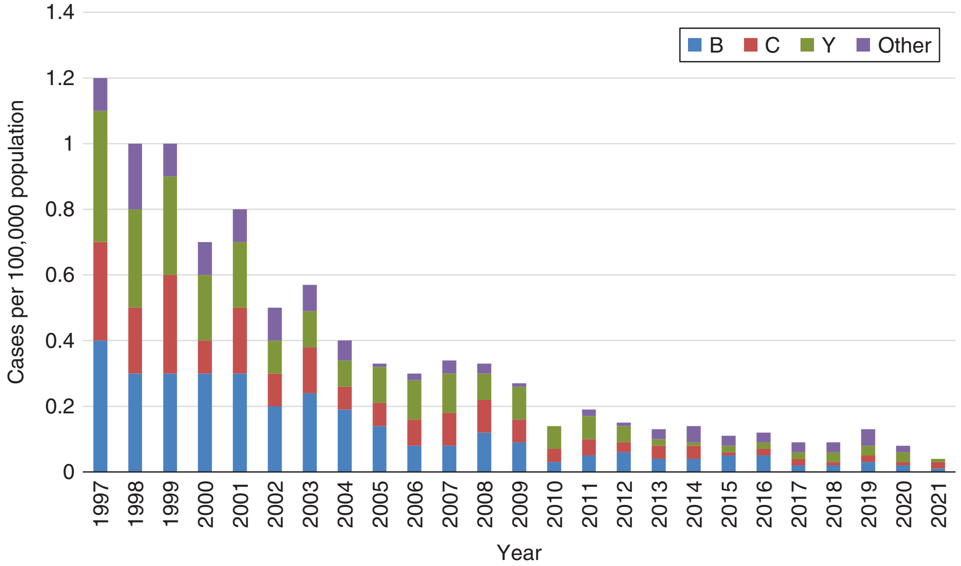

FIGURE 160-3 Meningococcal disease in the United States, 1997–2021. ABCs, active bacterial cores. *(Adapted from ABC Surveillance data, Centers for Disease Control and Prevention; https://www.cdc.gov/abcs/reports-findings/surv-reports.html.)*

in young children or teenagers and the use of adolescent immunization programs for capsular groups A, C, Y, and W. However, other factors, including changes in natural population immunity (induced by exposure from nasopharyngeal colonization) and prevalent clones of meningococci (factors that, in combination, probably explain the historical cyclic nature of meningococcal disease rates) as well as a reduction in smoking and passive exposure to tobacco smoke (driven by bans on smoking in buildings and public spaces) across wealthy countries, are likely to have contributed to the fall in cases. There are also data from the United Kingdom indicating that reductions in contacts between individuals during the COVID-19 pandemic led to further substantial declines in meningococcal disease and that contact patterns remain altered and could have a sustained effect on transmission.

Factors Associated with Disease Risk and Susceptibility The principal determinant of disease susceptibility is age, with the peak incidence in the first year of life **(Fig. 160-5)**. The susceptibility of the very young presumably results from an absence of specific adaptive immunity in combination with very close contact with colonized individuals, including parents. Compared with other age groups, infants appear to be particularly susceptible to capsular group B disease: >30% of capsular group B cases in the United States occur during the first year of life. In the early 1990s prior to use of vaccines in North America, the median ages for patients with disease due to capsular groups B, C, Y, and W were 6, 17, 24, and 33 years, respectively. In populations where capsular group A, C, W, and Y vaccines are being used with good coverage, disease from these capsular groups has become very rare in children.

After early childhood, a second peak of disease occurs among adolescents and young adults (15–25 years of age) in Europe and North America. It is thought that this peak relates to social behaviors and environmental exposures in this age group, as discussed below. Most cases of infection with *N. meningitidis* in developed countries today are sporadic, and the rarity of the disease suggests that individual

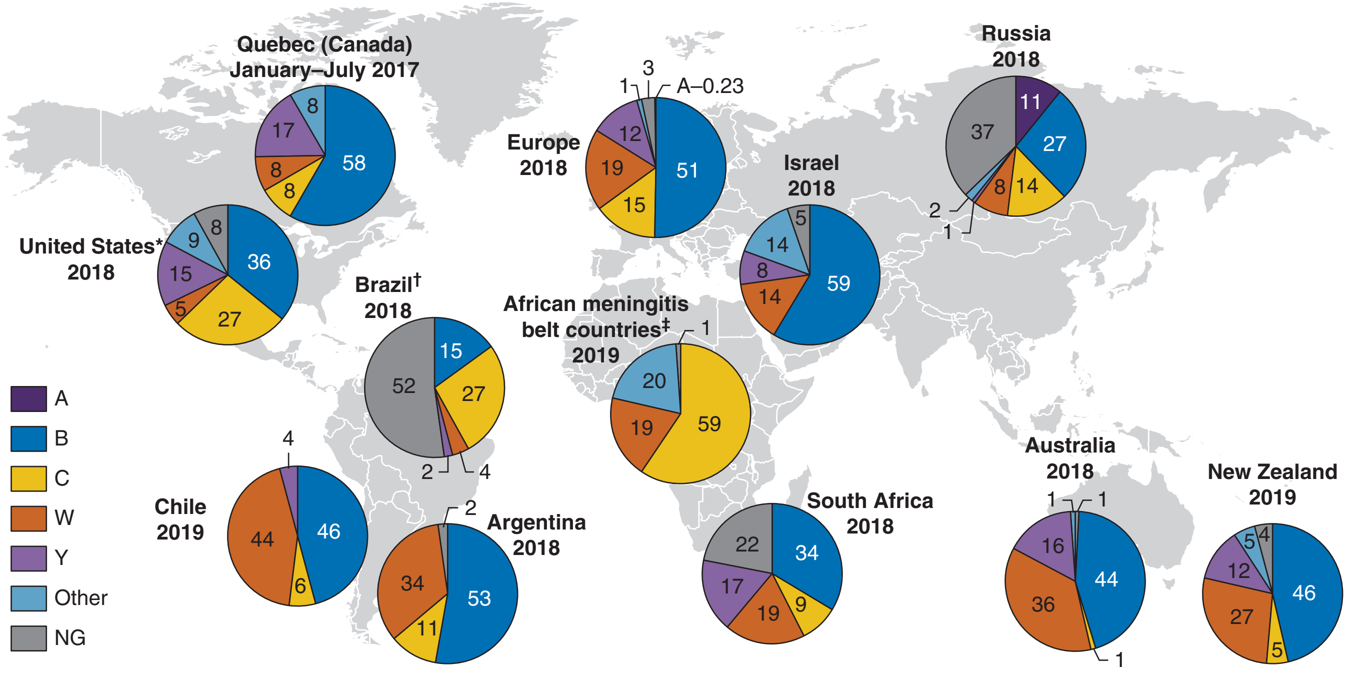
Quebec (Canada)
January–July 2017
8
17
8
8
58
Europe
2018
3
1
A–0.23
12
19
15
51
Russia
2018
11
37
27
8
14
2
1
Israel
2018
5
14
8
14
59
United States*
2018
8
9
15
5
27
36
Brazil†
2018
15
52
27
2
4
African meningitis
belt countries‡
2019
1
20
19
59
A
B
C
W
Y
Other
NG
Chile
2019
4
44
6
46
Argentina
2018
2
34
11
53
South Africa
2018
22
34
17
19
9
Australia
2018
1
1
16
36
44
1
New Zealand
2019
5
4
12
27
5
46

FIGURE 160-4 Global percentage distribution of meningococcal capsular groups causing invasive meningococcal disease, 2017–2019. NG, non-groupable plus capsular groups other than B, C, W, and Y. *(Adapted from C Pardo de Santayana et al: Epidemiology of invasive meningococcal disease worldwide from 2010–2019: A literature review. Epidemiol Infect 151:e57, 2023.)*

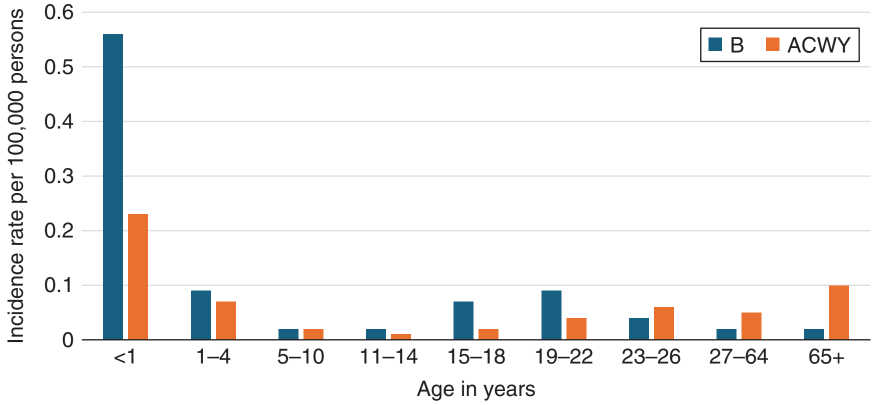

FIGURE 160-5 Age distribution of capsular groups B and ACWY meningococcal disease United States, 2012–2021. *(Adapted from https://www.cdc.gov/meningococcal/php/surveillance/.)*

susceptibility may be important. A number of factors probably contribute to individual susceptibility, including the host's genetic constitution, environment, and contact with a carrier or a case.

The best-documented genetic association with meningococcal disease is complement deficiency, chiefly of the terminal complement components (C5–9), properdin, or factor D or those treated with complement inhibitors such as eculizumab; such a deficiency increases the risk of disease by up to 600-fold and may result in recurrent attacks. Complement components are believed to be important for the bactericidal activity of serum, which is considered the principal mechanism of immunity against invasive meningococcal disease. However, when investigated, complement deficiency is found in only a very small proportion of individuals with meningococcal disease (0.3%). Conversely, 7–20% of persons whose disease is caused by the less common capsular groups (W, X, Y, Z, E) have a complement deficiency. Complement deficiency appears to be associated with capsular group B disease only rarely. Individuals with recurrences of meningococcal disease, particularly those caused by non-B capsular groups, should be assessed for complement deficiency by measurement of total hemolytic complement activity. There is also limited evidence that hyposplenism (through reduction in phagocytic capacity), hypogammaglobulinemia (through absence of specific antibody), and HIV increase the risk of meningococcal disease. Genetic studies have revealed various associations with disease susceptibility, including complement and mannose-binding lectin deficiency, single-nucleotide polymorphisms in Toll-like receptor (TLR) 4 and complement factor H, and variants of Fc gamma receptors.

Factors that increase the chance of a susceptible individual's acquiring *N. meningitidis* via the respiratory route also increase the risk of meningococcal disease. Acquisition occurs through close contact with carriers as a result of overcrowding (e.g., in poor socioeconomic settings, in refugee camps, during the Hajj pilgrimage to Mecca, during freshman-year residence in college dormitories), recruitment into the military, and certain social behaviors (e.g., attendance at bars and nightclubs, kissing). Secondary cases may occur in close contacts of an index case (e.g., household members, persons kissing the infected individual); the risk to these contacts may be as high as 1000 times the background rate in the population. Factors that damage the nasopharyngeal epithelium also increase the risk of both colonization with *N. meningitidis* and invasive disease. The most important of these factors are tobacco smoking (odds ratio, 4.1) and passive exposure to tobacco smoke. In addition, recent viral respiratory tract infection, infection with *Mycoplasma* species, and winter or the dry season (in sub-Saharan Africa) have been associated with meningococcal disease; all of these factors presumably either increase the expression of adhesion molecules in the nasopharynx, thus enhancing meningococcal adhesion, or facilitate meningococcal invasion of the bloodstream.

PATHOGENESIS

N. meningitidis has evolved as an effective colonizer of the human nasopharynx, with asymptomatic infection rates of >25% described in some series of adolescents and young adults and among residents of crowded communities. Point-prevalence studies reveal widely divergent rates of carriage for different types of meningococci. This variation suggests that some types may be adapted to a short duration of carriage with frequent transmission to maintain the population, while others may be less efficiently transmitted but may overcome this disadvantage by colonizing for a long period. Despite the high rates of carriage among adolescents and young adults, only ~10% of adults carry meningococci, and colonization is very rare in early childhood. Many of the same factors that increase the risk of meningococcal disease also increase the risk of carriage. Colonization of the nasopharynx involves a series of interactions of meningococcal adhesins (e.g., Opa proteins and pili) with their ligands on the epithelial mucosa. *N. meningitidis* produces an IgA1 protease that is likely to reduce interruption of colonization by mucosal IgA.

Colonization should be considered the normal state of meningococcal infection, with an increased risk of invasion being the unfortunate consequence (for both host and organism) of adaptations of hyperinvasive meningococcal lineages to favor their survival. The meningococcal capsule is an important virulence factor: acapsular strains only very rarely cause invasive disease. The capsule provides resistance to phagocytosis and may be important in preventing desiccation during transmission between hosts. Antigenic diversity in surface structures and an ability to vary levels of their expression probably have evolved as important factors in maintaining meningococcal populations within and between individual hosts.

Invasion through the mucosa into the bloodstream occurs rarely, usually within a few days of acquisition of an invasive strain by a susceptible individual. Only occasional cases of prolonged colonization prior to invasion have been documented. Once the organism is in the bloodstream, its growth may be limited if the individual is partially immune, although bacteremia may allow seeding of another site, such as the meninges or the joints. Alternatively, unchecked proliferation may continue, resulting in high bacterial counts in the circulation. During growth, meningococci release blebs of outer membrane (Fig. 160-1) containing outer-membrane proteins and LPS. Endotoxin binds cell-bound CD14 in association with TLR4 to initiate an inflammatory cascade with the release of high levels of various mediators, including tumor necrosis factor (TNF) α, soluble TNF receptor, interleukin (IL) 1, IL-1 receptor antagonist, IL-1β, IL-6, IL-8, IL-10, plasminogen-activator inhibitor 1 (PAI-1), and leukemia inhibitory factor. Soluble CD14-bound endotoxin acts as a mediator of endothelial activation. The severity of meningococcal disease is related both to the levels of endotoxin in the blood and to the magnitude of the inflammatory response. The latter is determined to some extent by polymorphisms in the inflammatory response genes (and their inhibitors), and the release of the inflammatory cascade heralds the development of meningococcal septicemia (meningococcemia). Endothelial injury is central to many clinical features of meningococcemia, including increased vascular permeability, pathologic changes in vascular tone, loss of thromboresistance, intravascular coagulation, and myocardial dysfunction. Endothelial injury leads to increased vascular permeability (attributed to loss of glycosaminoglycans and endothelial proteins), with subsequent gross proteinuria. Leakage of fluid and electrolytes into the tissues from capillaries ("capillary leak syndrome") leads to hypovolemia, tissue edema, and pulmonary edema. Initial compensation results in vasoconstriction and tachycardia, although cardiac output eventually falls. While resuscitation fluids may restore circulating volume, tissue edema will continue to increase, and, in the lung, the consequence may be respiratory failure.

Intravascular thrombosis (caused by activation of procoagulant pathways in association with upregulation of tissue factor on the endothelium) occurs in some patients with meningococcal disease and results in purpura fulminans and infarction of areas of skin or even of whole limbs. At the same time, multiple anticoagulant pathways are downregulated through loss of endothelial thrombomodulin and protein C receptors and decreases in levels of antithrombin III, protein C, protein S, and tissue factor pathway inhibitor. Thrombolysis is also profoundly impaired in meningococcal sepsis through the release of high levels of PAI-1.

Shock in meningococcal septicemia appears to be attributable to a combination of factors, including hypovolemia, which results from the capillary leak syndrome secondary to endothelial injury, and myocardial depression, which is driven by hypovolemia, hypoxia, metabolic derangements (e.g., hypocalcemia), and cytokines (e.g., IL-6). Decreased perfusion of tissues as a result of intravascular thrombosis, vasoconstriction, tissue edema, and reduced cardiac output in meningococcal septicemia can cause widespread organ dysfunction, including renal impairment and—later in the disease—a decreased level of consciousness due to central nervous system involvement.

Bacteria that reach the meninges cause a local inflammatory response—with release of a spectrum of cytokines similar to that seen in septicemia—that presents clinically as meningitis and is thought to determine the severity of neuronal injury. Local endothelial injury may result in cerebral edema and rapid onset of raised intracranial pressure in some cases.

■ CLINICAL MANIFESTATIONS

As discussed above, the most common form of infection with *N. meningitidis* is asymptomatic carriage of the organism in the nasopharynx. Despite the location of infection in the upper airway, meningococcal pharyngitis is rarely reported; however, upper respiratory tract symptoms are common prior to presentation with invasive disease. It is not clear whether these symptoms relate to preceding viral infection (which may promote meningococcal acquisition and/or invasion) or to meningococcal acquisition itself. After acquiring the organism, susceptible individuals develop disease manifestations in 1–10 days (usually <4 days, although colonization for 11 weeks has been documented).

Along the spectrum of presentations of meningococcal disease, the most common clinical syndromes are meningitis and meningococcal septicemia. In fulminant cases, death may occur within hours of the first symptoms. Occult bacteremia is also recognized and, if untreated, progresses in two-thirds of cases to focal infection, including meningitis or septicemia. Meningococcal disease may also present as pneumonia, pyogenic arthritis or osteomyelitis, purulent pericarditis, endophthalmitis, conjunctivitis, primary peritonitis, or (rarely) urethritis. Perhaps because it is difficult to diagnose, meningococcal pneumonia is not commonly reported but is associated with capsular groups Y, W, and Z and appears most often to affect individuals >10 years of age.

Rash A nonblanching rash (petechial or purpuric) develops in >80% of cases of meningococcal disease; however, the rash is often absent early in the illness. Usually initially blanching in nature (macules, maculopapules, or urticaria) and indistinguishable from more common viral rashes, the rash of meningococcal infection becomes petechial or frankly purpuric over the hours after onset. In the most severe cases, large purpuric lesions develop (purpura fulminans; **Fig. A1-41**). Some patients (including those with overwhelming sepsis) may have no rash. While petechial rash and fever are important signs of meningococcal disease, <10% of children (and, in some clinical settings, <1% of patients) with this presentation are found to have meningococcal disease. Most patients presenting with a petechial or purpuric rash have a viral infection **(Table 160-2)**. The skin lesions exhibit widespread endothelial necrosis and occlusion of small vessels in the dermis and subcutaneous tissues, with a neutrophilic infiltrate.

TABLE 160-2 Common Causes of Petechial or Purpuric Rashes

Enteroviruses
Influenza and other respiratory viruses
Measles virus
Epstein-Barr virus
Cytomegalovirus
Parvovirus
Deficiency of protein C or S (including post-varicella protein S deficiency)
Platelet disorders (e.g., idiopathic thrombocytopenic purpura, drug effects, bone marrow infiltration)
Henoch-Schönlein purpura, connective tissue disorders, trauma (including nonaccidental injuries in children)
Pneumococcal, streptococcal, staphylococcal, or gram-negative bacterial sepsis

Meningitis Meningococcal meningitis commonly presents as nonspecific manifestations, including fever, vomiting, and (especially in infants and young children) irritability, and is indistinguishable from other forms of bacterial meningitis unless there is an associated petechial or purpuric rash, which occurs in two-thirds of cases. Headache is rarely reported in early childhood but is more common in later childhood and adulthood. When headache is present, the following features, in association with fever or a history of fever, are suggestive of bacterial meningitis: neck stiffness, photophobia, decreased level of consciousness, seizures or status epilepticus, and focal neurologic signs. Classic signs of meningitis, such as neck stiffness and photophobia, are often absent in infants and young children with bacterial meningitis, who more usually present with fever and irritability and may have a bulging fontanelle.

While 30–50% of patients present with a meningitis syndrome alone, up to 40% of meningitis patients also present with some features of septicemia. Most deaths from meningococcal meningitis alone (i.e., without septicemia) are associated with raised intracranial pressure presenting as a reduced level of consciousness, relative bradycardia and hypertension, focal neurologic signs, abnormal posturing, and signs of brainstem involvement—e.g., unequal, dilated, or poorly reactive pupils; abnormal eye movement; and impaired corneal responses **(Chap. 30)**.

Septicemia Meningococcal septicemia alone accounts for up to 20% of cases of meningococcal disease. The condition may progress from early nonspecific symptoms to death within hours. Mortality rates among children with this syndrome have been high (25–40%), but early aggressive management (as discussed below) may reduce the figure to <10%. Early symptoms are nonspecific and suggest an influenza-like illness with fever, headache, and myalgia accompanied by vomiting and abdominal pain. As discussed above, the rash, if present, may appear to be viral early in the course until petechiae or purpuric lesions develop. Purpura fulminans occurs in severe cases **(Fig. A1-41)**, with multiple large purpuric lesions and signs of peripheral ischemia. Surveys of patients have indicated that limb pain, pallor (including a mottled appearance and cyanosis), and cold hands and feet may be prominent. Shock is manifested by tachycardia, poor peripheral perfusion, tachypnea, and oliguria. Decreased cerebral perfusion leads to confusion, agitation, or decreased level of consciousness. With progressive shock, multiorgan failure ensues; hypotension is a late sign in children, who more commonly present with compensated shock (tachycardia, poor peripheral perfusion, and normal blood pressure). Poor outcome is associated with an absence of meningism, hypotension, young age, coma, relatively low temperature (<38°C), leukopenia, and thrombocytopenia. Spontaneous hemorrhage (pulmonary, gastric, or cerebral) may result from consumption of coagulation factors and thrombocytopenia.

Chronic Meningococcemia Chronic meningococcemia, which is rarely recognized, presents as repeated episodes of petechial rash **(Fig. A1-42)** associated with fever, joint pain, features of arthritis, and splenomegaly that may progress to acute meningococcal septicemia if untreated. During the relapsing course, bacteremia characteristically clears without treatment and then recurs. The differential diagnosis includes bacterial endocarditis, acute rheumatic fever, Henoch-Schönlein purpura, infectious mononucleosis, disseminated gonococcal infection, and immune-mediated vasculitis. This condition has been associated with complement deficiencies in some cases and with inadequate sulfonamide therapy in others.

A study from the Netherlands found that half of isolates from patients with chronic meningococcemia had an underacylated lipid A (part of the surface LPS molecule) due to an *lpxL1* gene mutation, which markedly reduces the inflammatory response to endotoxin.

Postmeningococcal Reactive Disease In a small proportion of patients, an immune complex disease develops ~4–10 days after the onset of meningococcal disease, with manifestations that include a maculopapular or vasculitic rash (2% of cases), arthritis (up to 8% of cases), iritis (1%), pericarditis, and/or polyserositis associated with fever. The immune complexes involve meningococcal polysaccharide antigen and result in immunoglobulin and complement deposition with an inflammatory infiltrate. These features resolve spontaneously without sequelae. It is important to recognize this condition since a new onset of fever and rash, and/or arthritis, can lead to concerns about relapse of meningococcal disease and unnecessarily prolonged antibiotic treatment.

DIAGNOSIS

Like other invasive bacterial infections, meningococcal disease may produce elevations of the white blood cell (WBC) count and of values for inflammatory markers (e.g., C-reactive protein and procalcitonin levels or the erythrocyte sedimentation rate). Values may be normal or low in rapidly progressive disease, and a lack of rise in these laboratory test values does not exclude the diagnosis. However, in the presence of fever and a petechial rash, these elevations are suggestive of meningococcal disease. In patients with severe meningococcal septicemia, common laboratory findings include hypoglycemia, acidosis, hypokalemia, hypocalcemia, hypomagnesemia, hypophosphatemia, anemia, and coagulopathy.

Although meningococcal disease is often diagnosed on clinical grounds, in suspected meningococcal meningitis or meningococcemia, blood should routinely be sent for culture to confirm the diagnosis and to facilitate public health investigations; blood cultures are positive in up to 75% of cases. Culture media containing sodium polyanethol sulfonate, which may inhibit meningococcal growth, should be avoided. Meningococcal viability is reduced if there is a delay in transport of the specimen to the microbiology laboratory for culture or in plating of cerebrospinal fluid (CSF) samples. In countries where treatment with antibiotics before hospitalization is recommended for meningococcal disease, the majority of clinically suspected cases are culture negative. Real-time polymerase chain reaction (PCR) analysis of whole-blood samples increases the diagnostic yield by >40%, and results obtained with this method may remain positive for several days after administration of antibiotics.

Unless contraindications exist (raised intracranial pressure, uncorrected shock, disordered coagulation, thrombocytopenia, respiratory insufficiency, local infection, ongoing convulsions), lumbar puncture should be undertaken to identify and confirm the etiology of suspected meningococcal meningitis, whose presentation cannot be distinguished from that of meningitis of other bacterial causes. Some authorities have recommended a computed tomography (CT) brain scan prior to lumbar puncture because of the risk of cerebral herniation in patients with raised intracranial pressure. However, a normal CT scan is not uncommon in the presence of raised intracranial pressure in meningococcal meningitis, and the decision to perform a lumbar puncture should be made on clinical grounds. CSF features of meningococcal meningitis (elevated protein level and WBC count, decreased glucose level) are indistinguishable from those of other types of bacterial meningitis unless a gram-negative diplococcus is identified. (Gram's staining is up to 80% sensitive for meningococcal meningitis.) CSF should be submitted for culture (sensitivity, 90%) and (where available) PCR analysis. CSF antigen testing with latex agglutination is insensitive and should be replaced by molecular diagnosis when possible.

Lumbar puncture should generally be avoided in meningococcal septicemia, as positioning for the procedure may critically compromise the patient's circulation in the context of hypovolemic shock. Delayed lumbar puncture may still be useful when the diagnosis is uncertain, particularly if molecular diagnostic technology is available.

In other types of focal infection, culture and PCR analysis of normally sterile body fluids (e.g., synovial fluid) may aid in the diagnosis. Although some authorities have recommended cultures of scrapings or aspirates from skin lesions, this procedure adds little to the diagnostic yield when compared with a combination of blood culture and PCR analysis. Urinary antigen testing also is insensitive, and serologic testing for meningococcal infection has not been adequately studied. Because *N. meningitidis* is a component of the normal human nasopharyngeal flora, identification of the organism on throat swabs has limited diagnostic value, but strains identified in the nasopharynx in the context of a probable case are likely to be those responsible for disease.

TREATMENT

Meningococcal Infections

Death from meningococcal disease is associated most commonly with hypovolemic shock (meningococcemia) and occasionally with raised intracranial pressure (meningococcal meningitis). Therefore, management should focus on the treatment of these urgent clinical issues in addition to the administration of specific antibiotic therapy. Delayed recognition of meningococcal disease or its associated physiologic derangements, together with inadequate emergency management, is associated with poor outcome. Since the disease is rare, protocols for emergency management have been developed (see *https://www.meningitis.org/healthcare-professionals/resources*).

Airway patency may be compromised if the level of consciousness is depressed as a result of shock (impaired cerebral perfusion) or raised intracranial pressure; this situation may require intervention. In meningococcemia, pulmonary edema and pulmonary oligemia (presenting as hypoxia) require oxygen therapy or elective endotracheal intubation. In cases with shock, aggressive fluid resuscitation (with replacement of the circulating volume several times in severe cases) and inotropic support may be necessary to maintain cardiac output. If shock persists after volume resuscitation at 40 mL/kg, the risk of pulmonary edema is high, and elective intubation is recommended to improve oxygenation and decrease the work of breathing. Metabolic derangements, including hypoglycemia, acidosis, hypokalemia, hypocalcemia, hypomagnesemia, hypophosphatemia, anemia, and coagulopathy, should be anticipated and corrected. However, aggressive fluid resuscitation with unbuffered electrolyte solutions was found to increase mortality in febrile African children. Studies of the effects of lower volumes of buffered solutions and similar studies in resource-rich settings are required. In the presence of raised intracranial pressure, management includes correction of coexistent shock and neurointensive care to maintain cerebral perfusion.

Empirical antibiotic therapy for suspected meningococcal disease consists of a third-generation cephalosporin such as ceftriaxone (75–100 mg/kg per d [maximum, 4 g/d] in one or two divided IV doses) or cefotaxime (200 mg/kg per day [maximum, 8 g/d] in four divided IV doses) to cover the various other (potentially penicillin-resistant) bacteria that may produce an indistinguishable clinical syndrome. In many settings, vancomycin (usually 4–60 mg/kg per d in two to four divided IV doses) is also recommended for the empiric management of sepsis and/or suspected bacterial meningitis. Although unusual in most isolates, reduced meningococcal sensitivity to penicillin (a minimal inhibitory concentration of 0.12–1.0 μg/mL) has been reported. Use of penicillin is appropriate once information on antimicrobial resistance patterns is available.

Both meningococcal meningitis and meningococcal septicemia are conventionally treated for 7 days, although courses of 3–5 days may be equally effective. Furthermore, a single dose of ceftriaxone or an oily suspension of chloramphenicol has been used successfully in resource-poor settings. No data are available to guide the duration of treatment for meningococcal infection at other foci (e.g., pneumonia, arthritis); antimicrobial therapy is usually continued until clinical and laboratory evidence of infection has resolved. Cultures usually become sterile within 24 h of initiation of appropriate antibiotic chemotherapy. Eye infections (including keratoconjunctivitis and endophthalmitis) should be treated with a combination of topical and systemic IV therapy, with some small studies suggesting an increased risk of bacteremia when treated with topical therapy alone.

The use of glucocorticoids for adjunctive treatment of meningococcal meningitis remains controversial since no relevant studies have had sufficient power to determine true efficacy in this condition. One large study in adults did indicate a trend toward benefit, and in clinical practice, a decision to use glucocorticoids would best precede a definite microbiologic diagnosis. Therapeutic doses of glucocorticoids are not recommended in meningococcal septicemia, but many intensivists recommend replacement glucocorticoid doses for patients who have refractory shock in association with impaired adrenal gland responsiveness, management that is supported by limited evidence.

Various other adjunctive therapies for meningococcal disease have been considered, but few have been subjected to clinical trials and none can currently be recommended. An antibody to LPS (HA1A) failed to confer a demonstrable benefit. Recombinant bactericidal/permeability-increasing protein (which is not currently available) was tested in a study that had inadequate power to show an effect on mortality rates; however, there were trends toward lower mortality rates among patients who received a complete infusion, and this group also had fewer amputations, fewer blood-product transfusions, and a significantly improved functional outcome. Given that protein C concentrations are reduced in meningococcal disease, the use of activated protein C has been considered. A survival benefit was demonstrated in adult sepsis trials; however, trials in pediatric sepsis (of particular relevance for meningococcal disease) found no benefit and indicated a potential risk of bleeding complications with use of activated protein C.

The postmeningococcal immune-complex inflammatory syndrome has been treated with nonsteroidal anti-inflammatory agents until spontaneous resolution occurs.

COMPLICATIONS

About 10% of patients with meningococcal disease die despite the availability of antimicrobial therapy and other intensive medical interventions. The most common complication of meningococcal disease (10% of cases) is scarring after necrosis of purpuric skin lesions, for which skin grafting may be necessary. The lower limbs are most often affected; next in frequency are the upper limbs, the trunk, and the face. On average, 13% of the skin surface area is involved. Amputations are necessary in 1–2% of survivors of meningococcal disease because of a loss of tissue viability after peripheral ischemia or compartment syndromes. Unless there is local infection, amputation should usually be delayed to allow the demarcation between viable and nonviable tissue to become apparent. Approximately 5% of patients with meningococcal disease suffer hearing loss, and 7% have neurologic complications. In one study, pain was reported by 21% of survivors, and in an analysis of capsular group B meningococcal disease (the MOSAIC study), as many as one-quarter of survivors had psychological disorders. In some investigations, the rate of complications is higher for capsular group C disease (mostly associated with the ST11 clone) than for capsular group B disease. In patients with severe hypovolemic shock, renal perfusion may be impaired and prerenal failure is common, but permanent renal replacement therapy is rarely needed.

Several studies suggest adverse psychosocial outcomes after meningococcal disease, with reduced quality of life, lowered self-esteem, and poorer neurologic development, including increased rates of attention deficit/hyperactivity disorder and special educational needs. Other studies have not found evidence of such outcomes.

PROGNOSIS

Several prognostic scoring systems have been developed to identify patients with meningococcal disease who are least likely to survive. Factors associated with a poorer prognosis are shock; young age (infancy), old age, and adolescence; coma; purpura fulminans; disseminated intravascular coagulation; thrombocytopenia; leukopenia; absence of meningitis; metabolic acidosis; low plasma concentrations of antithrombin and proteins S and C; high blood levels of PAI-1; and a low erythrocyte sedimentation rate or C-reactive protein level. The Glasgow Meningococcal Septicaemia Prognostic Score (GMSPS) performs well and may be clinically useful for severity assessment in meningococcal disease. However, scoring systems do not direct the clinician to specific interventions, and the priority in management should be recognition of compromised airways, breathing, or circulation and direct, urgent intervention. Most patients improve rapidly with appropriate antibiotics and supportive therapy. Fulminant meningococcemia is more likely to result in death or ischemic skin loss than is meningitis; optimal emergency management may reduce mortality rates among the most severely affected patients.

PREVENTION

Since mortality rates in meningococcal disease remain high despite improvements in intensive care management, immunization is the only rational approach to prevention at a population level. Secondary cases are common among household and "kissing" contacts of cases, and secondary prophylaxis with antibiotics is widely recommended for these contacts (see below).

Polysaccharide Vaccines Purified meningococcal capsular polysaccharide was first used for immunization in the 1960s. Meningococcal polysaccharide vaccines were formulated as either bivalent (capsular groups A and C) or quadrivalent (capsular groups A, C, Y, and W), with 50 μg of each polysaccharide per dose. Local reactions (erythema, induration, and tenderness) occurred in up to 40% of vaccinees, but serious adverse events (including febrile convulsions in young children) are very rarely reported. In adults, the vaccines are immunogenic, but immunity is relatively short-lived (with antibody levels above baseline for only 2–10 years), and booster doses do not induce a further rise in antibody concentration. Indeed, a state of immunologic hyporesponsiveness has been widely reported to follow booster doses of plain polysaccharide vaccines. The repeating sugar units of these vaccines cross-link B-cell receptors to drive specific memory B cells to become plasma cells and produce protective antibody. Because meningococcal polysaccharides are T cell–independent antigens, no memory B cells are produced after immunization, and the memory B-cell pool is depleted such that fewer polysaccharide-specific cells are available to respond to a subsequent dose of vaccine **(Fig. 160-6)**. The clinical relevance of hyporesponsiveness is unknown. Plain polysaccharide vaccines generally are not immunogenic in early childhood, possibly because marginal-zone B cells are involved in polysaccharide responses and maturation of the splenic marginal zone is not complete until 18 months to 2 years of age. The efficacy of the meningococcal capsular group C component is >90% in young adults, but there is no protection in infants; no efficacy data are available for the capsular group Y and W polysaccharides in this age group.

Group A meningococcal polysaccharides are exceptional in that they are effective in preventing disease at all ages. Two doses administered 2–3 months apart to children 3–18 months of age or a single dose administered to older children or adults has a protective efficacy rate of >95%. The vaccine was previously used widely in the control of outbreaks of meningococcal disease in the African meningitis belt, providing protection for 3–5 years. The plain polysaccharide vaccines have been largely superseded by protein–polysaccharide conjugate vaccines.

There is no meningococcal capsular group B plain polysaccharide vaccine because α-2,8-*N*-acetylneuraminic acid is expressed on the surface of neural cells in the fetus such that the B polysaccharide is perceived as "self" and therefore is not immunogenic in humans.

Conjugate Vaccines The poor immunogenicity of plain polysaccharide vaccines in infancy has been overcome by chemical conjugation of the polysaccharides to a carrier protein (CRM_{197}, tetanus toxoid, or diphtheria toxoid). Conjugates that contain monovalent capsular group C polysaccharide and quadrivalent vaccines with A, C, Y, and W polysaccharides were developed, and vaccines with other antigen

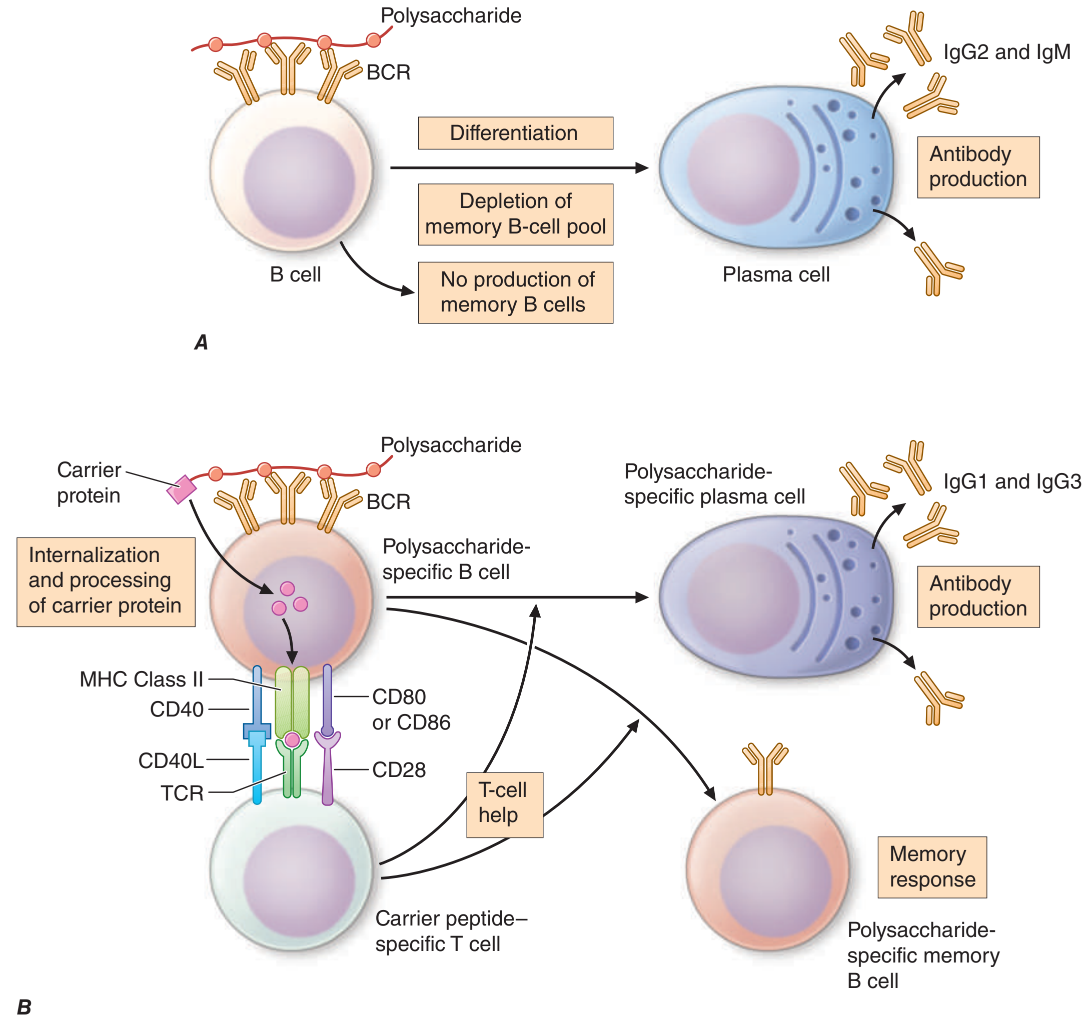
Polysaccharide
BCR
B cell
Differentiation
Depletion of memory B-cell pool
No production of memory B cells
Plasma cell
IgG2 and IgM
Antibody production
A
Carrier protein
Polysaccharide
BCR
Internalization and processing of carrier protein
Polysaccharide-specific B cell
MHC Class II
CD40
CD40L
TCR
CD80 or CD86
CD28
T-cell help
Carrier peptide–specific T cell
Polysaccharide-specific plasma cell
IgG1 and IgG3
Antibody production
Memory response
Polysaccharide-specific memory B cell
B

FIGURE 160-6 ***A.*** Polysaccharides from the encapsulated bacteria that cause disease in early childhood stimulate B cells by cross-linking the B-cell receptor (BCR) and driving the production of immunoglobulins. There is no production of memory B cells, and the B-cell pool may be depleted by this process such that subsequent immune responses are decreased. ***B.*** The carrier protein from protein–polysaccharide conjugate vaccines is processed by the polysaccharide-specific B cell, and peptides are presented to carrier peptide–specific T cells, with the consequent production of both plasma cells and memory B cells. MHC, major histocompatibility complex; TCR, T-cell receptor. *(Reproduced with permission from AJ Pollard: Maintaining protection against invasive bacteria with protein–polysaccharide conjugate vaccines. Nat Rev Immunol 9:213, 2009.)*

combinations were produced for some markets (e.g., tetanus conjugates with capsular group C and/or Y polysaccharide and *Haemophilus influenzae* type b polysaccharide). After immunization, peptides from the carrier protein are conventionally thought to be presented by polysaccharide-specific B cells to peptide-specific T cells in association with major histocompatibility complex (MHC) class II molecules. (Some data suggest that carrier protein peptide may actually be presented in association with an oligosaccharide and MHCII.) The result is a T cell–dependent immune response that allows production of antibody and generation of an expanded B-cell memory pool. Unlike responses to booster doses of plain polysaccharides, responses to booster doses of conjugate vaccines have the characteristics of memory responses. Indeed, conjugate vaccines overcome the hyporesponsiveness induced by plain polysaccharides by replenishing the memory pool. The reactogenicity of conjugate vaccines is similar to that of plain polysaccharide vaccines.

The first widespread use of capsular group C meningococcal conjugate vaccine (MenC) came in 1999 in the United Kingdom after a rise in capsular group C disease. A mass vaccination campaign involving all individuals <19 years of age was undertaken, and the number of laboratory-confirmed capsular group C cases fell from 955 in 1998–1999 to just 29 in 2011–2012. The effectiveness of the immunization program was attributed both to direct protection of immunized persons and to reduced transmission of the organism in the population as a result of decreased rates of colonization among the immunized (i.e., herd immunity). Data on immunogenicity and effectiveness have shown that the duration of protection is short when the vaccine is administered in early childhood; thus, booster doses are needed to maintain population immunity. In contrast, immunity after a dose of vaccine given in adolescence appears to be more prolonged.

In 2005, the first quadrivalent conjugate meningococcal vaccine containing A, C, Y, and W polysaccharides conjugated to diphtheria toxoid was initially recommended for all children >11 years of age in the United States and for persons 2–55 years of age in Canada. Such vaccines are now recommended by the Advisory Committee on Immunization Practices (ACIP) for routine administration to individuals 11–12 years of age, with a booster dose at 16 years of age; only a single dose is given to persons >16 years of age. These vaccines are also recommended for high-risk persons from 2 months to 55 years of age (see *https://www.cdc.gov/mmwr/volumes/69/rr/rr6909a1.htm*).

Uptake was slow initially, but U.S. data show an efficacy rate of 82% in the first year after vaccination and 69% at 8 years (diphtheria conjugate vaccine). Early reports of an increase in the risk of Guillain-Barré syndrome after immunization with the diphtheria conjugate vaccine have not been substantiated with further observation. Quadrivalent conjugate vaccines with tetanus or CRM$_{197}$ as carrier protein are now available in many countries and are used for high-risk groups and in routine programs for toddlers and adolescents. Use of the A, C, W, and Y conjugate vaccine for adolescents since 2015 in the United Kingdom has led to a large reduction in meningococcal disease caused by these capsular groups. This conjugate vaccine provided direct protection for vaccinated individuals (combined vaccine effectiveness of 94% against C, W, and Y disease) and marked reductions in carriage—a 36% reduction in carriage of capsular groups C, W, and Y combined at 2 months postvaccination—along with evidence of herd immunity. Indeed, modeling of the decline of cases and carriage in the United Kingdom indicates elimination of these capsular groups over the coming decade.

A monovalent capsular group A vaccine, manufactured in India, was licensed in 2010 and rolled out to countries in the sub-Saharan African meningitis belt in a mass immunization campaign. There is strong evidence that this vaccine has been highly effective in controlling epidemic meningococcal disease in the region, with >90% reduction in disease in vaccinated populations. New combination vaccines covering A, C, W, X, and Y have been developed and are set to replace monovalent MenA vaccines in Africa as part of the WHO's global roadmap "Defeating Meningitis by 2030" (*https://www.who.int/publications/i/item/9789240026407*).

Vaccines Based on Subcapsular Antigens The lack of immunogenicity of the group B capsule has led to the development of vaccines based on subcapsular antigens. Various surface components have been studied in early-phase clinical trials. Outer-membrane vesicles (OMVs) containing outer-membrane proteins, phospholipid, and LPS can be extracted from cultures of *N. meningitidis* by detergent treatment **(Fig. 160-7)**. OMVs prepared in this way were used in efficacy trials with a Norwegian outbreak strain and reduced the incidence of group B disease among 14- to 16-year-old schoolchildren by 53%. Similarly, OMV vaccines constructed from local outbreak strains in Cuba and New Zealand have had reported efficacy rates of >70%. These OMV vaccines appear to produce strain-specific immune responses, with only limited cross-protection, and are therefore best suited to clonal outbreaks (e.g., those in Cuba and New Zealand as well as others in Norway and the province of Normandy in France).

Several purified surface proteins have been evaluated in phase 1 clinical trials but have not yet been developed further because of antigenic variability or poor immunogenicity (e.g., transferrin-binding proteins, neisserial surface protein A). Other vaccine candidates have been identified since sequencing of the meningococcal genome. The combination vaccine 4CMenB, which includes the New Zealand OMV vaccine and three recombinant proteins (neisserial adhesin A, factor H–binding protein, and neisserial heparin-binding antigen), is immunogenic from infancy and has been licensed for use in the United States, Canada, Europe, and Australia. This vaccine has been used with apparent success in the control of several university outbreaks in the United States and in a community outbreak in an area of Quebec, Canada. The 4CMenB vaccine has an acceptable safety profile, with fever prominent among infants and injection-site pain frequently reported among older children and adults. In September 2015, 4CMenB was recommended for routine use in the United Kingdom for all infants born from May 2015 onward; a recent analysis reported a 75% reduction in age groups that were fully eligible for vaccination, with a high coverage rate of 95%. The licensed schedule is three priming doses before 6 months of age and a booster dose at 12 months of age (but is used in the United Kingdom as two priming doses under 6 months of age and a booster at 1 year). A nonsignificant vaccine effectiveness of 53% was seen after two doses, and 59% effectiveness was found after the booster dose at 1 year of age. As of 2022, protein-based MenB vaccines were authorized in 58 countries. Of these countries, 15 have a universal program in at least one age group, 21 have a recommendation for high-risk groups based on medical conditions, and 13 have a recommendation based on increased risk of exposure (e.g., laboratory staff).

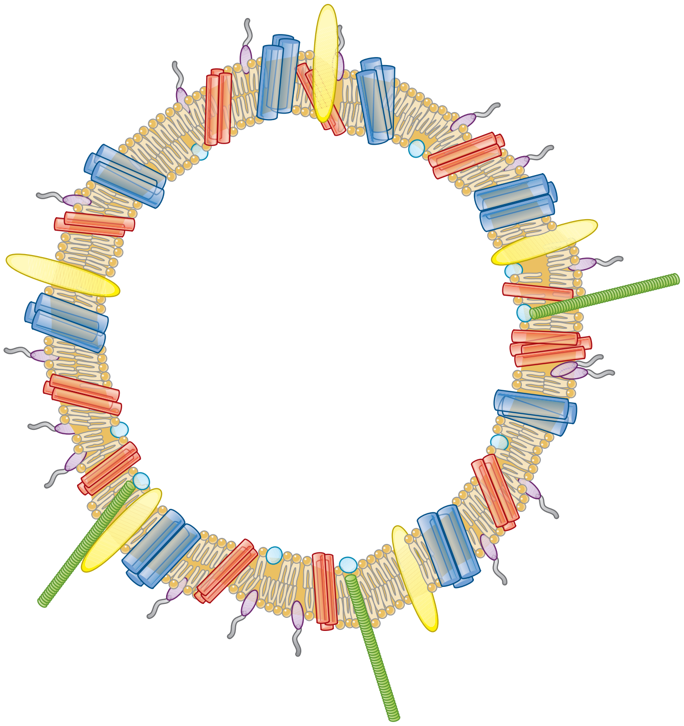

FIGURE 160-7 Illustration of meningococcal outer-membrane vesicle containing outer-membrane structures.

Because the disease is so rare, the cost-effectiveness of capsular group B vaccine in infant immunization programs, as assessed with conventional thresholds, is borderline in the United Kingdom. Since infants are not commonly colonized with capsular group B meningococci, any impact on the total population burden of carried organisms will be small. It is therefore unlikely that an infant immunization program will provide additional value through induction of herd immunity. Rates of capsular group B carriage are higher among teenagers and young adults than at other ages (apart from infancy). A large cluster-randomized trial in Australia found no effect of 4CMenB on carriage of disease-causing meningococci, highlighting that the benefit of this vaccine is likely to be via direct protection only.

An immunogenic vaccine based on two variants of the lipoprotein factor H–binding protein (MenB-fHBP) has been developed for use in adolescents and is licensed in the United States and Europe. The vaccine is immunogenic against representative indicator strains, inducing fourfold rises in bactericidal antibody titer in 50–92% of individuals. MenB-fHBP has an acceptable safety profile, with pain at the injection site, fatigue, and headache commonly reported. This vaccine can be used with a range of vaccines routinely administered in adolescence, including Tdap (tetanus–diphtheria–acellular pertussis), human papillomavirus, and MenACWY vaccines. MenB-fHBP has been used to control outbreaks of meningococcal disease in educational institutions in the United States, but no formal studies of its effectiveness have yet been undertaken due to the absence of any public health programs with this vaccine. Studies in the United Kingdom are currently evaluating the impact of both 4CMenB and MenB-fHBP against meningococcal carriage among teenagers.

Both of the new capsular group B meningococcal vaccines are licensed for use in the United States for persons 10–25 years of age but are not recommended for routine use. ACIP advises that the vaccines can be used in a two-dose schedule following shared decision-making between the doctor, patient, and/or family. In addition, ACIP recommends their administration to individuals at high risk of capsular group B disease in a two-dose schedule (4CMenB) or a three-dose schedule (MenB-fHBP).

MANAGEMENT OF CONTACTS

Close (household and kissing) contacts of individuals with meningococcal disease are at increased risk of developing secondary disease (up to 1000 times the rate for the general population); a secondary case follows as many as 3% of sporadic cases. About one-fifth of secondary cases are actually co-primary cases—i.e., cases that occur soon after the primary case and in which transmission is presumed to have originated from the same third party. The rate of secondary cases is highest during the week after presentation of the index case. The risk falls rapidly but remains above baseline for up to 1 year after the index case; 30% of secondary cases occur in the first week, 20% in the second week, and most of the remainder over the next 6 weeks. In outbreaks of meningococcal disease, mass prophylaxis has been used; however, limited data support population intervention, and significant concerns have arisen about adverse events and the development of resistance. For these reasons, prophylaxis is usually restricted to (1) persons at greatest risk who are intimate and/or household contacts of the index case and (2) health care workers who have been directly exposed to respiratory secretions. In most cases, members of wider communities (e.g., at schools or colleges) are not offered prophylaxis.

The aim of prophylaxis is to eradicate colonization of close contacts with the strain that has caused invasive disease in the index case. Prophylaxis should be given to all contacts at the same time to avoid recolonization by meningococci transmitted from untreated contacts and should also be used as soon as possible to treat early disease in secondary cases. If the index patient is treated with an antibiotic that does not reliably clear colonization (e.g., penicillin), the patient should be given a prophylactic agent at the end of treatment to prevent relapse or onward transmission. Although rifampin has been most widely used and studied, it is not the optimal agent because it fails to eradicate carriage in 15–20% of cases, rates of adverse events have been high, compliance is affected by the need for four doses, and emerging resistance has been reported. Ceftriaxone as a single IM or IV injection is highly (97%) effective in carriage eradication and can be used at all ages and in pregnancy. Reduced susceptibility of isolates to ceftriaxone has occasionally been reported. Ciprofloxacin or ofloxacin is preferred in some countries; these agents are highly effective and can be administered by mouth but are not recommended in pregnancy. Resistance to fluoroquinolones has been reported in some meningococci in North America, Europe, and Asia.

In documented capsular group A, B, C, Y, or W disease, contacts may be offered immunization (with either the MenACWY conjugate vaccine or the MenB vaccine, as appropriate) in addition to chemoprophylaxis to provide protection beyond the duration of antibiotic therapy. Mass vaccination has been used successfully to control disease during outbreaks in closed communities (educational and military establishments) as well as during epidemics in open communities.

FURTHER READING

Carr JP et al: Impact of meningococcal ACWY conjugate vaccines on pharyngeal carriage in adolescents: Evidence for herd protection from the UK MenACWY programme. Clin Microbiol Infect 28:1649.e1, 2022.

Castilla J et al: Effectiveness of a meningococcal group B vaccine (4CMenB) in children. N Engl J Med 388:427, 2023.

Christensen H et al: Meningococcal carriage by age: A systematic review and meta-analysis. Lancet Infect Dis 10:853, 2010.

Haidara FC et al: Meningococcal ACWYX conjugate vaccine in 2-to-29-year-olds in Mali and Gambia. N Engl J Med 388:1942, 2023.

Ladhani SN et al: Vaccination of infants with meningococcal group B vaccine (4CMenB) in England. N Engl J Med 382:309, 2020.

Marshall HS et al: Meningococcal B vaccine and meningococcal carriage in adolescents in Australia. N Engl J Med 382:318, 2020.

Mbaeyi SA et al: Meningococcal vaccination: Recommendations of the Advisory Committee on Immunization Practices, United States, 2020. MMWR Recomm Rep 69:1, 2020.

Pardo de Santayana C et al: Epidemiology of invasive meningococcal disease worldwide from 2010–2019: A literature review. Epidemiol Infect 151:e57, 2023.

Pollard AJ et al: Maintaining protection against invasive bacteria with protein–polysaccharide conjugate vaccines. Nat Rev Immunol 9:213, 2009.

Read RC et al: Effect of a quadrivalent meningococcal ACWY glycoconjugate or a serogroup B meningococcal vaccine on meningococcal carriage: An observer-blind, phase 3 randomised clinical trial. Lancet 384:2123, 2014.

Vieusseux M: Memoire sur le maladie qui a regne a Geneva au printemps de 1805. J Med Clin Pharm 11:163, 1805.

Wang B et al: 4CMenB sustained vaccine effectiveness against invasive meningococcal B disease and gonorrhoea at three years post programme implementation. J Infect 87:95, 2023.

161

Gonococcal Infections

Sanjay Ram, Peter A. Rice

DEFINITION

Gonorrhea is a sexually transmitted infection (STI) of epithelium and commonly manifests as cervicitis, urethritis, proctitis, and conjunctivitis. If untreated, infections at these sites can lead to local complications such as endometritis, salpingitis, tuboovarian abscess, bartholinitis, peritonitis, and perihepatitis in female patients; periurethritis and epididymitis in male patients; and ophthalmia neonatorum in newborns. Disseminated gonococcemia is an uncommon event whose manifestations include skin lesions, tenosynovitis, septic arthritis, and (in rare cases) endocarditis or meningitis.

MICROBIOLOGY

Neisseria gonorrhoeae is a gram-negative, nonmotile, non-spore-forming organism that grows singly and in pairs (i.e., as monococci and diplococci, respectively). Exclusively a human pathogen, the gonococcus contains, on average, three genome copies per coccal unit; this polyploidy permits a high level of antigenic variation and the survival of the organism in its host. Gonococci, like all other *Neisseria* species, are oxidase positive. They are distinguished from other neisseriae by their ability to grow on selective media and to use glucose but not maltose, sucrose, or lactose.

EPIDEMIOLOGY

The incidence of gonorrhea had been declining steadily in the United States, but in 2021, there were ~710,000 newly reported cases—up 136% since their historic low in 2009. With 82.4 million cases estimated by the World Health Organization to have occurred globally in 2020 among adults aged 15 to 49 years, gonorrhea remains a major public health problem worldwide, is a significant cause of morbidity in developing countries, and may play a role in enhancing transmission of HIV.

Gonorrhea predominantly affects young, nonwhite, unmarried, less educated members of urban populations. The number of reported cases probably represents half of the true number of cases—a discrepancy resulting from underreporting, self-treatment, nonspecific treatment without a laboratory-proven diagnosis, and asymptomatic infection. The number of reported new cases of gonorrhea in the United States rose from ~250,000 in the early 1960s to a high of 1.01 million in 1978. The recorded incidence of gonorrhea in modern times peaked in 1975, with 468 reported new cases per 100,000 population in the United States. This peak was attributable to the interaction of several variables, including improved accuracy of diagnosis, changes in patterns of contraceptive use, and changes in sexual behavior. A decline in the overall incidence of gonorrhea in the United States over the past quarter-century may have reflected increased condom use resulting from public health efforts to curtail HIV transmission. Nevertheless, in 2019, 214 new cases per 100,000 population were reported in this country, representing a 1-year increase of 4.6%; this figure is the highest among industrialized countries. Simultaneously, antibiotic resistance is increasing in the United States and other countries, prompting the U.S. Centers for Disease Control and Prevention (CDC) to name antibiotic-resistant *N. gonorrhoeae* as one of the three most urgent threats of its kind. At present, the attack rate in the United States is highest among 15- to 24-year-old women (730.5 per 100,000) and 20- to 29-year-old men (813.6 per 100,000); >70% of all reported cases occur in these two groups. From the standpoint of ethnicity, rates are highest among African Americans (652.9 per 100,000) and lowest among persons of Asian descent (37.8 per 100,000).

The incidence of gonorrhea is higher in developing countries than in industrialized nations. The exact incidence of any STI is difficult to ascertain in developing countries because of limited surveillance and variable diagnostic criteria. Extremely high rates of gonorrhea have been reported among aboriginal populations in Namibia and Australia. Studies in Africa have clearly demonstrated that nonulcerative STIs such as gonorrhea (in addition to ulcerative STIs) are an independent risk factor for the transmission of HIV **(Chap. 208)**.

Gonorrhea is transmitted from males to females more efficiently than in the opposite direction. The rate of transmission to a woman during a single unprotected sexual encounter with an infected man is ~50–80%. Oropharyngeal gonorrhea occurs in ~20% of women who practice fellatio with infected partners. Transmission in either direction by cunnilingus is rare.

In any population, there exists a small minority of individuals who have high rates of new-partner acquisition. These "core-group members" or "high-frequency transmitters" are vital in sustaining STI transmission at the population level. Another instrumental factor in sustaining gonorrhea in the population is the large number of infected individuals who are asymptomatic or have minor symptoms that are ignored. These persons, unlike symptomatic individuals, may not cease sexual activity and therefore may continue to transmit the infection. This situation underscores the importance of contact tracing and empirical treatment of the sex partners of index cases.

PATHOGENESIS, IMMUNOLOGY, AND ANTIMICROBIAL RESISTANCE

Outer-Membrane Proteins • **PILI** Fresh clinical isolates of *N. gonorrhoeae* initially form piliated (fimbriated) colonies distinguishable on translucent agar. Pilus expression is rapidly switched off with unselected subculture because of rearrangements in pilus genes. This change is a basis for antigenic variation of gonococci. Piliated strains adhere better to cells derived from human mucosal surfaces and are more virulent in organ culture models and human inoculation experiments than nonpiliated variants. In a fallopian tube explant model, pili mediate gonococcal attachment to nonciliated columnar epithelial cells. This event initiates gonococcal adherence, invasion and transport through these cells to intercellular spaces near the basement membrane or directly into the subepithelial tissue. Pili are also essential for genetic competence and transformation of *N. gonorrhoeae*, which permit horizontal transfer of genetic material between different gonococcal lineages in vivo.

OPACITY-ASSOCIATED PROTEIN Another gonococcal surface protein that is important in adherence to epithelial cells is opacity-associated protein (Opa; formerly called protein II). Opa contributes to intergonococcal adhesion, which is responsible for the opaque nature of gonococcal colonies on translucent agar and the organism's adherence to a variety of eukaryotic cells, including polymorphonuclear leukocytes (PMNs). Certain Opa variants promote invasion of epithelial cells, and this effect has been linked with the ability of Opa to bind vitronectin, heparan sulfate proteoglycans, and several members of the carcinoembryonic antigen–related cell adhesion molecule (CEACAM) receptor family. Epithelial CEACAM-binding gonococci prevent exfoliation of epithelium through a mechanism that involves nitric oxide that is produced during anaerobic bacterial metabolism and upregulation of CD105 (a member of the transforming growth factor-beta receptor family), which may interfere with bacterial clearance. *N. gonorrhoeae* Opa proteins that bind CEACAM1, which is expressed by primary CD4+ T lymphocytes, suppress the activation and proliferation of these lymphocytes. Select Opa proteins can engage CEACAM3, which is expressed on neutrophils, with consequent nonopsonic phagocytosis (i.e., phagocytosis independent of antibody and complement) and killing of bacteria.

PORIN Porin (previously designated protein I) is the most abundant gonococcal surface protein. Porin molecules exist as trimers that provide anion-transporting aqueous channels through the otherwise hydrophobic outer membrane. Porin exhibits stable interstrain antigenic variation and forms the basis for gonococcal serotyping. Two main serotypes have been identified; PorB.1A strains are often associated with disseminated gonococcal infection (DGI), whereas PorB.1B strains usually cause local genital infections only. DGI strains are generally resistant to the killing action of normal human serum and do not incite a significant local inflammatory response; therefore, they may not cause symptoms at genital sites. These characteristics may be related to the ability of PorB.1A strains to bind to complement-inhibitory molecules, resulting in a diminished inflammatory response. Porin can translocate to the cytoplasmic membrane of host cells—a process that could initiate gonococcal endocytosis and invasion. PorB.1B present in outer membrane vesicles shed during bacterial growth inhibits the ability of dendritic cells to induce T-cell proliferation and may contribute to the ability of gonococci to subvert adaptive immunity.

OTHER OUTER-MEMBRANE PROTEINS Other notable outer-membrane proteins include H.8, a lipoprotein that is present in high concentration on the surface of all gonococcal strains and is an excellent target for antibody-based diagnostic testing. Transferrin-binding proteins (Tbp1 and Tbp2), lactoferrin-binding proteins (LbpA and LbpB), and hemoglobin/haptoglobin binding proteins (HpuA and HpuB) are required for scavenging iron from transferrin, lactoferrin, and heme in vivo. Transferrin and iron have been shown to enhance the attachment of iron-deprived *N. gonorrhoeae* to human endometrial cells. TdfH and TdfJ enable gonococci to scavenge host zinc from calprotectin and S100 calcium binding protein A7 (psoriasin). IgA1 protease is produced by *N. gonorrhoeae* and may protect the organism from the action of mucosal IgA.

Lipooligosaccharide Gonococcal lipooligosaccharide (LOS) consists of a lipid A and a core oligosaccharide that lacks the repeating O-carbohydrate antigenic side chain seen in many other gram-negative bacteria. Gonococcal LOS possesses marked endotoxic activity and contributes to the local cytotoxic effect in a fallopian tube model. LOS core sugars undergo a high degree of phase variation under different conditions of growth; this variation reflects genetic regulation and expression of glycotransferase genes that dictate the carbohydrate structure of LOS. These phenotypic changes may affect interactions of *N. gonorrhoeae* with elements of the humoral immune system (antibodies and complement) and may also influence direct binding of organisms to both professional phagocytes and nonprofessional phagocytes (epithelial cells). For example, gonococci that are sialylated at their LOS sites inhibit the classic pathway of complement by reducing binding of IgG and also bind complement factor H to inhibit the alternative pathway of complement. LOS sialylation may also decrease nonopsonic Opa-mediated association with neutrophils and inhibit the oxidative burst in PMNs. The binding of the unsialylated terminal lactosamine residue of LOS to an asialoglycoprotein receptor on male epithelial cells facilitates adherence and subsequent gonococcal invasion of these cells. Moreover, oligosaccharide structures in LOS can modulate host immune responses. For example, the terminal monosaccharide expressed by LOS determines the C-type lectin receptor on dendritic cells that is targeted by the bacteria. In turn, the specific C-type lectin receptor engaged influences whether a T_H1- or T_H2-type response is elicited; the latter response may be less favorable for clearance of gonococcal infection.

Host Factors In addition to gonococcal structures that interact with epithelial cells, host factors seem to be important in mediating entry of gonococci into nonphagocytic cells. Activation of phosphatidylcholine-specific phospholipase C and acidic sphingomyelinase by *N. gonorrhoeae*, which results in the release of diacylglycerol and ceramide, is a requirement for the entry of *N. gonorrhoeae* into epithelial cells. Ceramide accumulation within cells leads to apoptosis, which may disrupt epithelial integrity and facilitate entry of gonococci into subepithelial tissue. Release of chemotactic factors as a result of complement activation contributes to inflammation, as does the toxic effect of LOS and peptidoglycan in provoking the release of inflammatory cytokines.

The importance of humoral immunity in host defenses against neisserial infections is best illustrated by the predisposition of persons deficient in terminal complement components (C5 through C9) to have recurrent bacteremic gonococcal infections and recurrent meningococcal meningitis or meningococcemia. Gonococcal porin induces

T cell–proliferative responses in persons with urogenital gonococcal disease. A significant increase in porin-specific interleukin (IL) 4–producing CD4+ as well as CD8+ T lymphocytes is seen in individuals with mucosal gonococcal disease. A portion of these lymphocytes that show a porin-specific T_H2-type response could traffic to mucosal surfaces and play a role in immune protection against the disease. Few data clearly indicate that protective immunity is acquired from a previous gonococcal infection, although bactericidal and opsonophagocytic antibodies to porin and LOS may offer partial protection. On the other hand, women who are infected and acquire high levels of antibody to another outer-membrane protein, Rmp (reduction modifiable protein, formerly called protein III), may be especially likely to become reinfected with *N. gonorrhoeae* because Rmp antibodies block the effect of bactericidal antibodies to porin and LOS. Rmp shows little, if any, interstrain antigenic variation; therefore, Rmp antibodies potentially may block antibody-mediated killing of all gonococci. The mechanism of blocking has not been fully characterized, but Rmp antibodies may noncompetitively inhibit binding of porin and LOS antibodies because of the proximity of these structures in the gonococcal outer membrane. In male volunteers who have no history of gonorrhea, the net effect of these events may influence the outcome of experimental challenge with *N. gonorrhoeae*. Because Rmp bears extensive homology to enterobacterial OmpA and meningococcal class 4 proteins, it is possible that these blocking antibodies result from prior exposure to cross-reacting proteins from these species and also play a role in first-time infection with *N. gonorrhoeae*.

Gonococcal Resistance to Antimicrobial Agents It is no surprise that *N. gonorrhoeae*, with its remarkable capacity to alter its antigenic structure and adapt to changes in the microenvironment, has become resistant to numerous antibiotics. The first effective agents against gonorrhea were the sulfonamides, which were introduced in the 1930s and became ineffective within a decade. Penicillin was then used as the drug of choice for the treatment of gonorrhea. By 1965, 42% of gonococcal isolates had developed low-level resistance to penicillin G. Resistance due to the production of penicillinase arose later.

Gonococci become fully resistant to antibiotics either by chromosomal mutations or by acquisition of R factors (plasmids). Two types of chromosomal mutations have been described. The first type, which is drug specific, is a single-step mutation leading to high-level resistance. The second type involves mutations at several chromosomal loci that combine to determine the level as well as the pattern of resistance. Strains with mutations in chromosomal genes were first observed in the late 1950s. As recently as 2007, chromosomal mutations accounted for resistance to penicillin, tetracycline, or both in ~16% of strains surveyed in the United States.

β-Lactamase (penicillinase)–producing strains of *N. gonorrhoeae* (PPNG) carrying β-lactamase plasmids had rapidly spread worldwide by the early 1980s. *N. gonorrhoeae* strains with plasmid-borne tetracycline resistance (TRNG) can mobilize some β-lactamase plasmids, and PPNG and TRNG occur together, sometimes along with strains exhibiting chromosomally mediated resistance (CMRNG). Penicillin, ampicillin, and tetracycline are no longer reliable for the treatment of gonorrhea and should not be used.

Quinolone-containing regimens also were recommended for treatment of gonococcal infections; the fluoroquinolones offered the advantage of antichlamydial activity when administered for 7 days. However, quinolone-resistant *N. gonorrhoeae* (QRNG) appeared soon after these agents were first used to treat gonorrhea. QRNG is particularly common in the Pacific Islands (including Hawaii) and Asia, where, in certain areas, all gonococcal strains are now resistant to quinolones. At present, QRNG is also common in parts of Europe and the Middle East. In the United States, QRNG has been identified in all areas but predominantly in states on the Pacific coast, where resistant strains were first seen. Alterations in DNA gyrase and topoisomerase IV have been implicated as mechanisms of fluoroquinolone resistance.

Third-generation cephalosporins have remained highly effective as single-dose therapy for gonorrhea, but the recent isolation of strains highly resistant to ceftriaxone (minimal inhibitory concentrations [MICs], 2 μg/mL) in Asia, some European countries and recently, in the United States, is cause for concern. Even though the MICs of ceftriaxone against certain strains may reach 0.015–0.125 μg/mL (higher than the MICs of 0.0001–0.008 μg/mL for fully susceptible strains), these levels are greatly exceeded in the blood, the urethra, and the cervix when the routinely recommended parenteral dose of ceftriaxone is administered. The rising MICs of oral cefixime (the previously recommended alternative oral third-generation cephalosporin) against *N. gonorrhoeae*, combined with this drug's limited capacity to reach levels sufficiently higher than MICs in the blood, the urethra, the cervix, and especially the pharynx, have resulted in the removal of cefixime from the list of first-line agents for treatment of uncomplicated gonorrhea. *N. gonorrhoeae* strains with reduced susceptibility to ceftriaxone and cefixime (i.e., cephalosporin-intermediate/resistant strains) contain mutations in (1) the *penA* allele, which is the principal resistance determinant and encodes a penicillin-binding protein (PBP2) whose sequence can differ in up to 60–70 amino acids from that of wild-type PBP2; (2) the *m*ultiple *t*ransferable *r*esistance *r*egulator (*mtrR*) gene that results in increased drug efflux through the MtrCDE efflux pump; and (3) *penB*, which decreases drug influx through PorB.

Spectinomycin has been used as an alternative agent because it is not associated with resistance to other antibiotics and can be reserved for use against multidrug-resistant strains of *N. gonorrhoeae*. In China, spectinomycin is recommended as an alternative agent for primary treatment of urogenital gonorrhea and is often used there instead of ceftriaxone. Nevertheless, outbreaks caused by strains resistant to spectinomycin have been documented in Korea and England when the drug had been used for primary treatment of gonorrhea.

Resistance to azithromycin can result from alterations of the ribosomal binding target by azithromycin and—as with cephalosporins—the over- and underexpression of efflux and influx systems. Combined resistance to cephalosporins and azithromycin has been reported in several instances throughout the world.

■ CLINICAL MANIFESTATIONS

Gonococcal Infections in Men Acute urethritis is the most common clinical manifestation of gonorrhea in male patients. However, there is a reservoir of infected men who remain asymptomatic. The usual incubation period after exposure is 2–7 days before symptoms develop although the interval can be longer. Strains of the PorB.1A serotype tend to cause a greater proportion of cases of mild and asymptomatic urethritis than do PorB.1B strains. When they occur, urethral discharge and dysuria, usually without urinary frequency or urgency, are the major symptoms. The discharge initially is scant and mucoid but becomes profuse and purulent within a day or two. Gram staining of the urethral discharge may reveal PMNs and gram-negative intracellular monococci and diplococci **(Fig. 161-1)**. The clinical manifestations of gonococcal urethritis are usually more severe and overt than

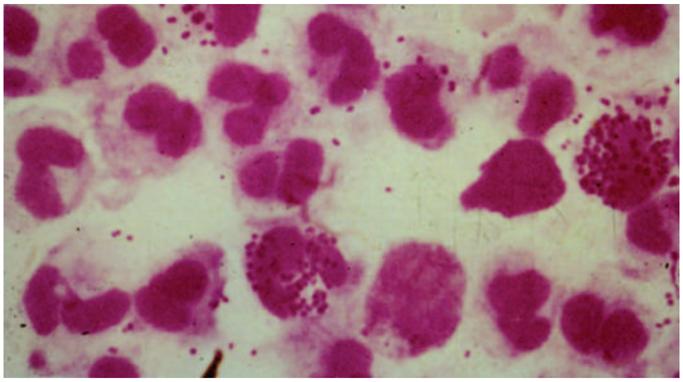

FIGURE 161-1 Gram stain of urethral discharge from a male patient with gonorrhea shows gram-negative intracellular monococci and diplococci. *(Source: © All rights reserved. Canadian Guidelines on Sexually Transmitted Infections. Public Health Agency of Canada, modified 2020. Adapted and reproduced with permission from the Minister of Health, 2021.)*

those of nongonococcal urethritis, including urethritis caused by *Chlamydia trachomatis* **(Chap. 194)**; however, exceptions are common, and it is often impossible to differentiate the causes of urethritis on clinical grounds alone. The majority of cases of urethritis seen in the United States today are not caused by *N. gonorrhoeae* and/or *C. trachomatis*. Although a number of other organisms may be responsible, many cases do not have a specific etiologic agent identified. Certain clones of *Neisseria meningitidis*, the second member of the pathogenic *Neisseria* species, have been associated with urethritis in men who have sex with men (MSM) in Europe and in heterosexual men in the southern and midwestern United States.

Most symptomatic men with gonorrhea seek treatment and cease to be infectious. The remaining men, who are largely asymptomatic, accumulate in number over time and constitute the majority of all infected men at any point in time; together with men incubating the organism who shed the organism but have yet to become symptomatic, they serve as the source of spread of infection. Before the antibiotic era, symptoms of urethritis persisted for ~8 weeks. Epididymitis is now an uncommon complication, and gonococcal prostatitis occurs rarely, if at all. Other unusual local complications of gonococcal urethritis include edema of the penis due to dorsal lymphangitis or thrombophlebitis, submucous inflammatory "soft" infiltration of the urethral wall, periurethral abscess or fistula, inflammation or abscess of Cowper's gland, and seminal vesiculitis. Balanitis may develop in uncircumcised men.

Gonococcal Infections in Women • **GONOCOCCAL CERVICITIS** Mucopurulent cervicitis is a common STI diagnosis in American women and may be caused by *N. gonorrhoeae*, *C. trachomatis*, and other organisms, including *Mycoplasma genitalium* **(Chap. 193)**. Cervicitis is often associated with more that one bacterial species, for example, coinfection with *C. trachomatis*, and may also coexist with candidal or trichomonal vaginitis. *N. gonorrhoeae* primarily infects the columnar epithelium of the cervical os. Bartholin's glands occasionally become infected.

About one-half of women infected with *N. gonorrhoeae* develop symptoms. Women who either remain asymptomatic or have only minor symptoms may delay seeking medical attention. These minor symptoms may include scant vaginal discharge issuing from the inflamed cervix (without vaginitis or vaginosis per se) and dysuria (often without urgency or frequency) that may be associated with gonococcal urethritis. Although the incubation period of gonorrhea is less well defined in women than in men, symptoms usually develop within 10 days of infection and are more acute and intense than those of chlamydial cervicitis.

The speculum examination reveals a mucopurulent discharge (mucopus) issuing from the cervical os or a reddened (inflamed) cervix even in the absence of reported symptoms. Because Gram stain is not sensitive for the diagnosis of gonorrhea in women, specimens should be submitted for culture or a nonculture assay (see "Laboratory Diagnosis," below). Edematous and friable cervical ectopy and endocervical bleeding induced by gentle swabbing are more often seen in chlamydial infection. Gonococcal infection may extend deep enough to produce dyspareunia and lower abdominal or back pain. In such cases, it is imperative to consider a diagnosis of pelvic inflammatory disease (PID) and to administer treatment for that disease **(Chaps. 141 and 194)**.

N. gonorrhoeae may also be recovered from the urethra and rectum of women with cervicitis, but these are rarely the only infected sites. Urethritis in women may produce symptoms of internal dysuria, which is often attributed to "cystitis." Pyuria in the absence of bacteriuria visible on Gram stain of unspun urine, accompanied by urine cultures that fail to yield $>10^2$ colonies of bacteria usually associated with urinary tract infection, signifies the possibility of urethritis usually due to *C. trachomatis*. Urethral infection with *N. gonorrhoeae* also may occur in this context, but in this instance, urethral cultures are usually positive.

GONOCOCCAL VAGINITIS The vaginal mucosa of healthy women is lined by stratified squamous epithelium and is rarely infected by *N. gonorrhoeae*. However, gonococcal vaginitis can occur in anestrogenic women (e.g., prepubertal girls and postmenopausal women), in whom the vaginal stratified squamous epithelium is often thinned down to the basilar layer, which can be infected by *N. gonorrhoeae*. The intense inflammation of the vagina makes the physical (speculum and bimanual) examination extremely painful. The vaginal mucosa is red and edematous, and an abundant purulent discharge is often present. Infection in the urethra and in Skene's and Bartholin's glands often accompanies gonococcal vaginitis. Inflamed cervical erosion or abscesses in nabothian cysts may also occur. Coexisting cervicitis may result in pus in the cervical os.

Anorectal Gonorrhea Because the female anatomy permits the spread of cervical exudate to the rectum, *N. gonorrhoeae* is sometimes recovered from the rectum of women with uncomplicated gonococcal cervicitis. The rectum is the sole site of infection in only 5% of women with gonorrhea. Such women are usually asymptomatic but occasionally have acute proctitis manifested by anorectal pain or pruritus, tenesmus, purulent rectal discharge, and rectal bleeding. Among MSM, the frequency of gonococcal infection, including rectal infection, fell by ≥90% throughout the United States in the early 1980s. A resurgence of gonorrhea among MSM has been documented in several cities since the 1990s; the estimated rates of reported cases having more than doubled in a recent 3-year period. Gonococcal isolates from the rectum of MSM tend to be more resistant to antimicrobial agents than are gonococcal isolates from other sites. Gonococcal isolates with a mutation in *mtrR* or in the promoter region of the gene that encodes for this transcriptional regulator develop increased resistance to antimicrobial hydrophobic agents such as bile acids and fatty acids in feces and thus are found with increased frequency in MSM. This situation may have been responsible for higher rates of failure of treatment for rectal gonorrhea with older regimens consisting of penicillin or tetracyclines.

Pharyngeal Gonorrhea Pharyngeal gonorrhea is usually mild or asymptomatic, although symptomatic pharyngitis does occasionally occur with cervical lymphadenitis. The mode of acquisition is oral–genital sexual exposure, with fellatio being a more efficient means of transmission than cunnilingus. In certain female adolescent populations in the United States, pharyngeal gonorrhea has become as common as genital gonorrhea. Most cases resolve spontaneously and transmission from the pharynx to sexual contacts is rare. Pharyngeal infection almost always coexists with genital infection. Swabs from the pharynx should be plated directly onto gonococcal selective media. Pharyngeal colonization with *N. meningitidis* needs to be differentiated from that with other *Neisseria* species. Because commensal oropharyngeal neisseriae are often resistant to antimicrobials, horizontal gene transfer between these organisms and *N. gonorrhoeae* may be important in the development of antimicrobial resistance of *N. gonorrhoeae.*

Ocular Gonorrhea in Adults Ocular gonorrhea in an adult usually results from autoinoculation of *N. gonorrhoeae* from an infected genital site. As in genital infection, the manifestations range from severe to occasionally mild or asymptomatic disease. The variability in clinical manifestations may be attributable to differences in the ability of the infecting strain to elicit an inflammatory response. Infection may result in a markedly swollen eyelid, severe hyperemia and chemosis, and a profuse purulent discharge. The massively inflamed conjunctiva may be draped over the cornea and limbus. Lytic enzymes from the infiltrating PMNs occasionally cause corneal ulceration and rarely cause perforation.

Prompt recognition and treatment of this condition are of paramount importance. Gram stain and culture of the purulent discharge establish the diagnosis. Genital cultures also should be performed.

Gonorrhea in Pregnant Women, Neonates, and Children Gonorrhea in pregnancy can have serious consequences for both the mother and the infant. Recognition of gonorrhea early in pregnancy also identifies a population at risk for other STIs, particularly chlamydial infection, syphilis, and trichomoniasis. The risks of salpingitis and PID—conditions associated with a high rate of fetal loss—are highest during the first trimester. Pharyngeal infection, most often asymptomatic, may be more common during pregnancy because of altered sexual practices. Prolonged rupture of the membranes, premature delivery, chorioamnionitis, funisitis (infection of the umbilical cord stump), and sepsis in the infant (with *N. gonorrhoeae* detected in the newborn's gastric aspirate during delivery) are common complications of maternal gonococcal infection at term. Other conditions and microorganisms, including *Mycoplasma hominis*, *Mycoplasma genitalium*, *Ureaplasma urealyticum*, *C. trachomatis*, and bacterial vaginosis (often accompanied by infection with *Trichomonas vaginalis*), have been associated with similar complications.

The most common form of gonorrhea in neonates is ophthalmia neonatorum, which results from exposure to infected cervical secretions during parturition. Ocular neonatal instillation of 0.5% erythromycin ophthalmic ointment, a prophylactic agent, prevents ophthalmia neonatorum but is not effective for its treatment, which requires systemic antibiotics. One-percent (1%) silver nitrate eye drops and ointments that contain silver nitrate or tetracycline can also be used for prophylaxis but are no longer available in the United States. The clinical manifestations are acute and usually begin 2–5 days after birth. An initial nonspecific conjunctivitis with a serosanguineous discharge is followed by tense edema of the eyelids, chemosis, and a profuse, thick, purulent discharge. Corneal ulcerations that result in nebulae or perforation may lead to anterior synechiae, anterior staphyloma, panophthalmitis, and blindness. Infections described at other mucosal sites in infants, including vaginitis, rhinitis, and anorectal infection, are likely to be asymptomatic. Pharyngeal colonization has been demonstrated in 35% of infants with gonococcal ophthalmia, and coughing is the most prominent symptom in these cases. Septic arthritis (see below) is the most common manifestation of systemic infection or disseminated gonococcal infection (DGI) in the newborn. The onset usually comes at 3–21 days of age, and polyarticular involvement is common. Sepsis, meningitis, and pneumonia are seen in rare instances.

Any STI in children beyond the neonatal period raises the possibility of sexual abuse. Gonococcal vulvovaginitis is the most common manifestation of gonococcal infection in children beyond infancy. Anorectal and pharyngeal infections are common in these children and are frequently asymptomatic. The urethra, Bartholin's and Skene's glands, and upper genital tract are rarely involved. All children with gonococcal infection should also be evaluated for chlamydial infection, syphilis, and possibly HIV infection.

Gonococcal Arthritis Disseminated gonococcal infection (DGI; gonococcal arthritis) results from gonococcal bacteremia. In the 1970s, DGI occurred in ~0.5–3% of persons with untreated gonococcal mucosal infection. The lower incidence of DGI at present is probably attributable to a decline in the prevalence of particular strains that are likely to disseminate. Nonetheless, sporadic outbreaks of DGI still occur in North America. DGI strains resist the bactericidal action of human serum and generally do not incite inflammation at genital sites, probably because of limited generation of chemotactic factors. Strains recovered from DGI cases in the 1970s were often of the PorB.1A serotype, were highly susceptible to penicillin, and had special growth requirements—including arginine, hypoxanthine, and uracil—that made the organism more fastidious and more difficult to isolate.

Menstruation is a risk factor for dissemination, and approximately two-thirds of cases of DGI are in women. In about half of affected women, symptoms of DGI begin within 7 days of onset of menses. Complement deficiencies, especially of the components involved in the assembly of the membrane attack complex (C5 through C9), predispose to neisserial bacteremia, and persons with more than one episode of DGI should be screened with an assay for total hemolytic complement activity. DGI is also associated with the use of the complement C5–blocking monoclonal antibody eculizumab.

The clinical manifestations of DGI have sometimes been classified into two stages: a bacteremic stage, which is less common today, and a joint-localized stage with suppurative arthritis. A clear-cut progression usually is not evident. Patients in the bacteremic stage have higher temperatures, and chills more frequently accompany their fever. Painful joints are common and often occur together with tenosynovitis and skin lesions. Polyarthralgias usually include the knees, elbows, and more distal joints; the axial skeleton is generally spared. Skin lesions are seen in ~75% of patients and include papules and pustules, often with a hemorrhagic component **(Fig. 161-2; see also Fig. A1-43 in the Atlas of Rashes Associated with Fever)**. Other manifestations of noninfectious dermatitis, such as nodular lesions, urticaria, and erythema multiforme, have been described. These lesions are usually on the extremities and number between 5 and 40. The differential diagnosis of the bacteremic stage of DGI includes reactive arthritis, acute rheumatoid arthritis, sarcoidosis, erythema nodosum, drug-induced arthritis, and viral infections (e.g., hepatitis B and acute HIV infection). The distribution of joint symptoms in reactive arthritis differs from that in DGI **(Fig. 161-3)**, as do the skin and genital manifestations **(Chap. 374)**.

Suppurative arthritis involves one or two joints, most often the knees, wrists, ankles, and elbows (in decreasing order of frequency); other joints occasionally are involved. Most patients who develop gonococcal septic arthritis do so without prior polyarthralgias or skin lesions; in the absence of symptomatic genital infection, this disease cannot be distinguished from septic arthritis caused by other pathogens. The differential diagnosis of acute arthritis in young adults is discussed in **Chap. 135**. Rarely, osteomyelitis complicates septic arthritis involving small joints of the hand.

Gonococcal endocarditis, although rare today, was a relatively common complication of DGI in the preantibiotic era, accounting for about one-quarter of reported cases of endocarditis. Another unusual complication of DGI is meningitis.

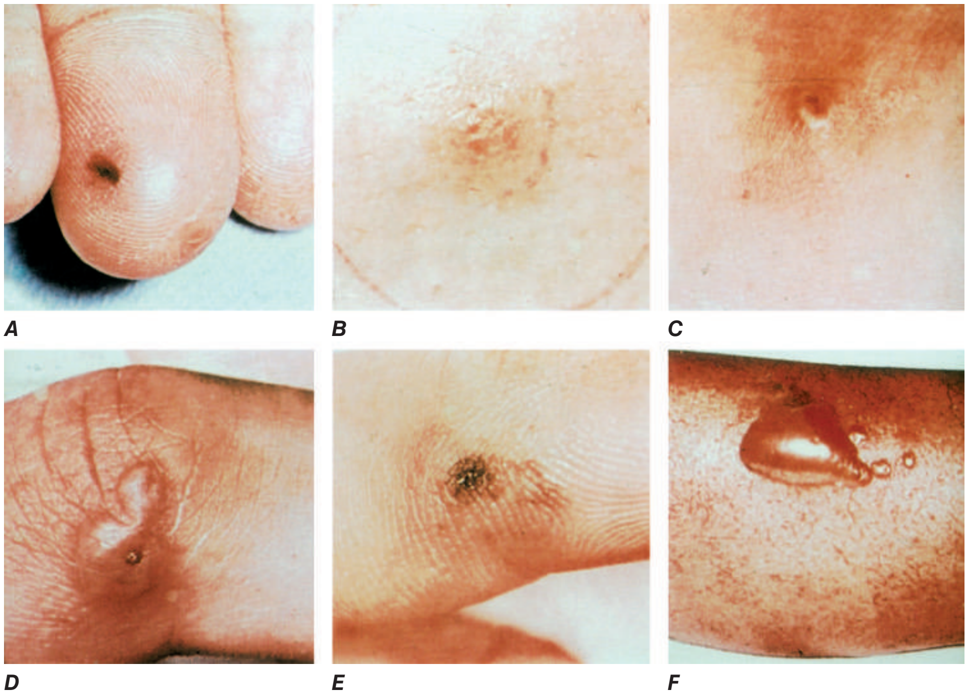

FIGURE 161-2 Characteristic skin lesions in patients with proven gonococcal bacteremia. The lesions are in various stages of evolution. ***A.*** Very early petechia on finger. ***B.*** Early papular lesion, 7 mm in diameter, on lower leg. ***C.*** Pustule with central eschar resulting from early petechial lesion. ***D.*** Pustular lesion on finger. ***E.*** Mature lesion with central necrosis (*black*) on hemorrhagic base. ***F.*** Bullae on anterior tibial surface. *(Reprinted with permission from KK Holmes et al: Disseminated gonococcal infection. Ann Intern Med 74:979, 1971.)*

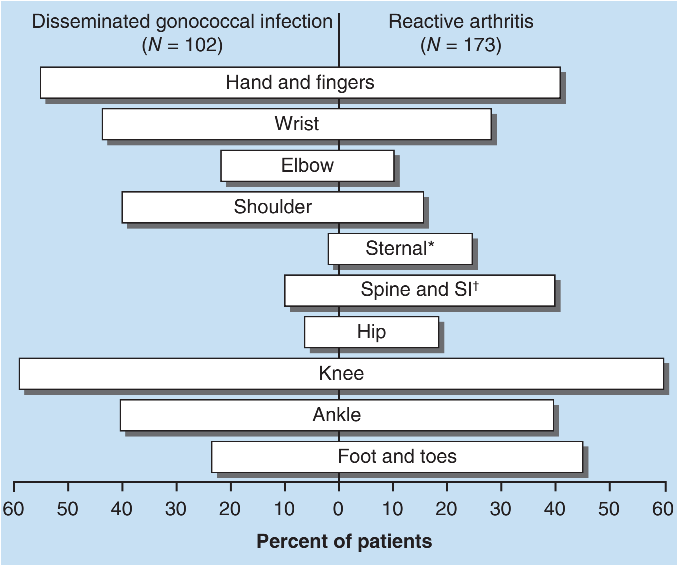

FIGURE 161-3 Distribution of joints with arthritis in 102 patients with disseminated gonococcal infection and 173 patients with reactive arthritis. *Includes the sternoclavicular joints. †SI, sacroiliac joint.

Gonococcal Infections in HIV-Infected Persons The association between gonorrhea and the acquisition of HIV has been demonstrated in several well-controlled studies, mainly in Kenya and Zaire. The nonulcerative STIs enhance the transmission of HIV three- to fivefold; transmission of HIV-infected immune cells and increased viral shedding by persons with urethritis or cervicitis may contribute **(Chap. 208)**. HIV has been detected by polymerase chain reaction (PCR) more commonly in ejaculates from HIV-positive men with gonococcal urethritis than in those from HIV-positive men with nongonococcal urethritis. PCR positivity diminishes twofold after appropriate therapy for urethritis. Not only does gonorrhea enhance the transmission of HIV, but it may also increase the individual's risk for acquisition of HIV. A proposed mechanism is the significantly greater number of CD4+ T lymphocytes and dendritic cells that can be infected by HIV in endocervical secretions from women with nonulcerative STIs than in those from women with ulcerative STIs.

■ LABORATORY DIAGNOSIS

A rapid diagnosis of gonococcal infection in men may be obtained by Gram staining of urethral exudates (Fig. 161-1). The detection of gram-negative intracellular monococci and diplococci is usually highly specific and sensitive in diagnosing gonococcal urethritis in symptomatic males but is only ~50% sensitive in diagnosing gonococcal cervicitis. Samples should be collected with Dacron or rayon swabs. Part of the sample should be inoculated onto a plate of modified Thayer-Martin or other gonococcal selective medium for culture. It is important to process all samples immediately because gonococci do not tolerate drying. If plates cannot be incubated immediately, they can be held safely for several hours at room temperature in candle extinction jars prior to incubation. If processing is to occur within 6 h, transport of specimens may be facilitated by the use of nonnutritive swab transport systems such as Stuart or Amies medium. For longer holding periods (e.g., when specimens for culture are to be mailed), culture media with self-contained CO_2-generating systems (such as the JEMBEC or Gono-Pak systems) may be used. Specimens should also be obtained for the diagnosis of chlamydial infection **(Chap. 194)**.

PMNs are often seen in the endocervix on a Gram stain, and an abnormally increased number (≥30 PMNs per field in five 1000× oil-immersion microscopic fields) establishes the presence of an inflammatory discharge. Unfortunately, the presence or absence of gram-negative intracellular monococci or diplococci in cervical smears does not accurately predict which patients have gonorrhea, and the diagnosis in this setting should be made by culture or another suitable nonculture diagnostic method. The sensitivity of a single endocervical culture is ~80–90%. If a history of rectal sex is elicited, a rectal wall swab (uncontaminated with feces) should be cultured. A presumptive diagnosis of gonorrhea cannot be made on the basis of gram-negative diplococci in Gram-stained smears from the pharynx, where other *Neisseria* species are also components of the pharyngeal flora.

Several nucleic acid amplification tests (NAATs), including the Roche COBAS AMPLICOR, Gen-Probe Aptima Combo 2, Cepheid Xpert® CT/NG Assay and BD ProbeTec ET, are now widely available on semiautomated or fully automated platforms and are commonly employed diagnostic tests for gonorrhea. These tests also detect *C. trachomatis* and are more sensitive than culture for identification of either *N. gonorrhoeae* or *C. trachomatis*. These tests offer the advantage that urine samples can be tested with a sensitivity similar to or greater than that obtained when urethral or cervical swab samples are assessed by other non-NAATs or culture, respectively. NAAT tests performed on self-collected vaginal swabs are as sensitive and specific as physician-collected samples and may be used in women to facilitate sample collection when a pelvic exam is not indicated as part of their clinic evaluation. A point-of-care NAAT-based test (Binx *io*®) for gonorrhea and chlamydia with a 30-minute turnaround time is now approved by the U.S. Food and Drug Administration (FDA). In MSM, it is important to screen the rectum and pharynx because screening urine alone will miss the majority of cases. A disadvantage of non-culture-based assays is that *N. gonorrhoeae* cannot be grown from the transport systems. Thus, a culture-confirmatory test and formal antimicrobial susceptibility testing, if needed, cannot be performed.

Because of the legal implications, the preferred method for the diagnosis of gonococcal infection in children is a standardized culture. Two positive NAATs, each targeting a different nucleic acid sequence, may be substituted for culture of the cervix or the urethra as legal evidence of infection in children. Although nonculture tests for gonococcal infection have not been approved by the FDA for use with specimens obtained from the pharynx and rectum of infected children, NAATs from these sites are preferred for diagnostic evaluation in adult victims of suspected sexual abuse, especially if the NAATs have been evaluated by the local laboratory and found to be superior. Cultures should be obtained from the pharynx and anus of both girls and boys, the urethra of boys, and the vagina of girls; cervical specimens are not recommended for prepubertal girls. For boys with a urethral discharge, a meatal specimen of the discharge is adequate for culture. Presumptive colonies of *N. gonorrhoeae* should be identified definitively by at least two independent methods.

Blood should be cultured in suspected cases of DGI. The use of Isolator blood culture tubes may enhance the yield. The probability of positive blood cultures decreases after 48 h of illness. Synovial fluid should be inoculated into blood culture broth medium and plated onto chocolate agar rather than selective medium because this fluid is not likely to be contaminated with commensal bacteria. Gonococci are infrequently recovered from early joint effusions containing <20,000 leukocytes/μL but may be recovered from effusions containing >80,000 leukocytes/μL. The organisms are seldom recovered from blood and synovial fluid of the same patient.

TREATMENT

Gonococcal Infections

Treatment failure can lead to continued transmission and the emergence of antibiotic resistance. The importance of adequate treatment with a regimen that the patient will adhere to cannot be overemphasized. Single-dose regimens have been developed for uncomplicated gonococcal infections. The Centers for Disease Control and Prevention (CDC) Gonorrhea Treatment Recommendations in 2020 and the 2021 Sexually Transmitted Infections (STI) Treatment Guidelines are summarized in **Table 161-1**. The third-generation cephalosporin ceftriaxone is now recommended as the first-line regimen for use at twice the previous dose (now,

TABLE 161-1 Recommended Treatment for Gonococcal Infections: Adapted from the 2021 Guidelines for Gonococcal Infection of the Centers for Disease Control and Prevention

DIAGNOSIS	TREATMENT OF CHOICE[a]
Uncomplicated gonococcal infection of the cervix, urethra, pharynx,[b] or rectum	
First-line regimen	Ceftriaxone (500 mg IM, single dose) ***plus*** Doxycycline (100 mg orally twice a day for 7 days) for treatment of chlamydial infection if chlamydial infection cannot be excluded
Alternative regimens if ceftriaxone is not available	Gentamicin (240 mg IM, single dose) plus azithromycin (2 g orally as a single dose)[c] ***or*** Cefixime (800 mg PO, single dose) *or* spectinomycin (2 g IM, single dose)[d,e] ***plus*** Doxycycline (100 mg orally twice a day for 7 days) for treatment of chlamydial infection if chlamydial infection cannot be excluded
Epididymitis	**See Chap. 141**
Pelvic inflammatory disease	**See Chap. 141**
Gonococcal conjunctivitis in an adult	Ceftriaxone (1 g IM, single dose)[f]
Ophthalmia neonatorum[g]	Ceftriaxone (25–50 mg/kg IV, single dose, not to exceed 125 mg)
Disseminated gonococcal infection[h]	
Initial therapy[i]	
Patient tolerant of β-lactam drugs	Ceftriaxone (1 g IM or IV q24h; recommended) *or* cefotaxime (1 g IV q8h) *or* ceftizoxime (1 g IV q8h)
Patients allergic to β-lactam drugs	Spectinomycin (2 g IM q12h)[d]
Continuation therapy[j]	Cefixime (400 mg PO bid)
Meningitis or endocarditis	See text for specific recommendations[k]

[a]True failure of treatment with a recommended regimen is rare and should prompt an evaluation for reinfection, infection with a drug-resistant strain, or an alternative diagnosis. [b]Ceftriaxone is the most reliable agent recommended for treatment of pharyngeal infection. [c]In vitro synergistic killing of *N. gonorrhoeae* by gentamicin plus azithromycin is mild to moderate; azithromycin is for treatment chlamydial infection, primarily. [d]Spectinomycin is unavailable in the United States; in uncomplicated gonococcal infection it should be used at a higher dose (4 g IM, single dose) in areas of the world where increased resistance to spectinomycin exists. [e]Spectinomycin may be ineffective for the treatment of pharyngeal gonorrhea. [f]Plus lavage of the infected eye with saline solution (once). [g]Prophylactic regimens are discussed in the text. [h]Hospitalization is indicated if the diagnosis is uncertain, if the patient has the joint-localized stage with suppurative arthritis, or if the patient cannot be relied on to adhere to treatment. [i]All initial regimens should also include doxycycline (100 mg orally twice a day for 7 days) for treatment of chlamydial infection if chlamydial infection cannot be excluded; [j]gonococcal therapy should be continued for 24–48 h after clinical improvement begins, at which time the switch may be made to an oral agent (e.g., cefixime) if antimicrobial susceptibility can be documented by culture of the causative organism. If no organism is isolated and the diagnosis is secure, then treatment with ceftriaxone should be continued for at least 1 week. [k]Hospitalization is indicated to exclude suspected meningitis or endocarditis.

500 mg IM, single dose) based on doubling of mean inhibitory concentrations (MICs) of current strains compared with MICs over 20 years ago. The development of decreased sensitivity to ceftriaxone throughout the world will require the development of new effective regimens. Azithromycin, which had been recommended to provide additional treatment of gonorrhea (also to include treatment of chlamydial infection) is no longer recommended as part of a first line regimen. Resistance to azithromycin of U.S. isolates of *N. gonorrhoeae*, which had been less than 0.6% over a number of years, has increased more than sevenfold to 4.7% in 2021. If chlamydial infection cannot be excluded, concurrent treatment with doxycycline (100 mg orally twice a day for 7 days) is recommended. The recommendations for uncomplicated gonorrhea apply to HIV-infected as well as HIV-uninfected patients.

The currently recommended regimen for the treatment of uncomplicated gonococcal infection of the urethra, cervix, rectum, or pharynx (a single IM dose of ceftriaxone) almost always results in an effective cure. Quinolone-containing regimens are no longer recommended in the United States as first-line treatment because of widespread resistance. Rising MICs of cefixime worldwide have led the CDC to discontinue its recommendation of this agent as first-line treatment for uncomplicated gonorrhea. Multicenter trials of treatment for uncomplicated gonorrhea in the United States have shown ≥99.5% efficacy of two combination regimens and 96% efficacy in one single-agent regimen: gemifloxacin (320 mg, single oral dose) plus azithromycin (2 g, single oral dose); gentamicin (a single IM dose of 240 mg or, in individuals who weigh ≤45 kg, 5 mg/kg) plus azithromycin (2 g, single oral dose), and zoliflodacin (2 or 3 g, single oral dose). At this time, however, none of these regimens is recommended by CDC as first-line treatment; gentamicin plus azithromycin is recommended as an alternative regimen.

Co-infection with *C. trachomatis* occurs frequently; concurrent treatment with doxycycline (100 mg orally twice daily for 7 days) is effective against chlamydial infection. Spectinomycin has been used as an alternative agent for the treatment of uncomplicated gonococcal infections in penicillin-allergic persons outside the United States but is not currently available in the United States. Of note, the limited effectiveness of spectinomycin and gentamicin for the treatment of pharyngeal infection reduces the utility of this regimen in populations among whom gonococcal infection is common, such as MSM.

Persons with uncomplicated genital or rectal infections who receive ceftriaxone or an alternative regimen do not need a test of cure; however, cultures for *N. gonorrhoeae* should be performed if symptoms persist after therapy with an established regimen, and any gonococci isolated should be tested for antimicrobial susceptibility. Persons with pharyngeal infection should undergo a test of cure regardless of the treatment regimen, 7–14 days after treatment to ensure eradication or detection of a possible treatment failure. Symptomatic gonococcal pharyngitis is more difficult to eradicate than genital infection. Persons who cannot tolerate cephalosporins may be treated with an alternative regimen. Treatment with spectinomycin results in a cure rate of ≤52%; persons given spectinomycin should have a subsequent pharyngeal sample cultured early (3–5 days) following treatment as a test of cure. A single 2-g dose of azithromycin may be used if the infecting organism is known to be sensitive or in areas where rates of resistance to azithromycin are low. Quinolones may be used if the infecting organism is known to be sensitive. If culture is not readily available and NAAT is positive, every effort should be made to perform a confirmatory culture. All isolates from test-of-cure cultures should undergo antimicrobial susceptibility testing. Because of high rates of reinfection with *N. gonorrhoeae* (and *C. trachomatis*) within 6–12 months, persons previously treated for gonorrhea should be retested 3 months after treatment.

Treatments for gonococcal epididymitis and PID are discussed in **Chap. 141**. Ocular gonococcal infections in older children and adults should be managed with a single dose of ceftriaxone combined with saline irrigation of the conjunctivae (both undertaken expeditiously), and patients should undergo a careful ophthalmologic evaluation that includes a slit-lamp examination.

DGI, particularly the joint-localized stage with suppurative arthritis, may require higher dosages and longer durations of therapy (Table 161-1). Hospitalization is indicated if the diagnosis is uncertain, if the patient has localized suppurative arthritis that requires aspiration, or if the patient cannot be relied on to comply with treatment. Open drainage is necessary only occasionally—e.g., for management of hip infections that may be difficult to drain percutaneously. Nonsteroidal anti-inflammatory agents may be indicated to alleviate pain and hasten clinical improvement of affected joints.

Gonococcal meningitis and endocarditis should be treated in the hospital with high-dose IV ceftriaxone (1–2 g IV every 12–24 h); therapy should continue for 10–14 days for meningitis and for at least 4 weeks for endocarditis. All persons who experience more than one episode of DGI should be evaluated for complement deficiency.

PREVENTION AND CONTROL

Condoms, if properly used, provide effective protection against the transmission and acquisition of gonorrhea as well as other infections that are transmitted to and from genital mucosal surfaces. Spermicidal preparations used with a diaphragm or cervical sponges impregnated with nonoxynol-9 offer some protection against gonorrhea and chlamydial infection. However, the frequent use of preparations that contain nonoxynol-9 is associated with mucosal disruption that paradoxically may enhance the risk of HIV infection in the event of exposure. All patients should be instructed to refer sex partners for evaluation and treatment. All sex partners of persons with gonorrhea should be evaluated and treated for *N. gonorrhoeae* and *C. trachomatis* infections if their last contact with the patient took place within 60 days before the onset of symptoms or the diagnosis of infection in the patient. If the patient's last potential sexual exposure to infection was >60 days before onset of symptoms or diagnosis, the patient's most recent sex partner should be treated. Partner-delivered medications or prescriptions for medications to treat gonorrhea and chlamydial infection diminish the likelihood of reinfection (or relapse) in the infected patient. This approach is permissible (or potentially allowable) in all 50 states and is an option for partner management. Patients should be instructed to abstain from sexual intercourse until therapy is completed and until they and their sex partners no longer have symptoms. Greater emphasis must be placed on prevention by public health education, individual patient counseling, and behavior modification, particularly the use of condoms. Sexually active persons, especially adolescents, should be offered screening for STIs. For most male patients, NAAT of urine or a urethral swab may be used for screening. Preventing the spread of gonorrhea may help reduce the transmission of HIV. No effective vaccine for gonorrhea is yet available, but efforts to test several candidates are underway including a field trial of a licensed group B meningococcal vaccine (4CMenB or Bexsero®), which in retrospective epidemiologic analyses has been associated with reduced rates of gonorrhea.

FURTHER READING

Bolan GA et al: The emerging threat of untreatable gonococcal infection. N Engl J Med 366:485, 2012.

Bolan GA et al: Morb Mortal Wkly Rep 70 (No. 4), 2021. Gonococcal Infections, in Sexually Transmitted Infections Treatment Guidelines, 2021. U.S. Centers for Disease Control and Prevention (CDC), 2021, pp 71-80.

Golden MR et al: Effect of expedited treatment of sex partners on recurrent or persistent gonorrhea or chlamydial infections. N Engl J Med 352:676, 2005.

Petousis-Harris EH et al: Effectiveness of a group B outer membrane vesicle meningococcal vaccine against gonorrhoea in New Zealand: A retrospective case-control study. Lancet 390:1603, 2017.

Rice PA: Gonococcal arthritis (disseminated gonococcal infection). Infect Dis Clin North Am 19:853, 2005.

Su X et al: *Neisseria gonorrhoeae* infection in women increases with rising gonococcal burdens in male partners: Chlamydia coinfection in women increases their gonococcal burdens. J Infect Dis 226:2192, 2022.

Taylor SN et al: Single-dose zoliflodacin (ETX0914) for treatment of urogenital gonorrhea. N Engl J Med 379:1835, 2018.

Unemo M et al: Antimicrobial resistance expressed by *Neisseria gonorrhoeae*: A major global public health problem in the 21st century. Microbiol Spectr 4:10.1128/microbiolspec.EI10-0009-2015, 2016.

Unemo MM et al: Gonorrhoea. Nat Rev Dis Primers 5:80, 2019.

162

Haemophilus and *Moraxella* Infections

Kristian Riesbeck, Timothy F. Murphy

HAEMOPHILUS INFLUENZAE

MICROBIOLOGY

Haemophilus influenzae was first recognized in 1892 by Pfeiffer, who erroneously concluded that the bacterium was the cause of influenza. *H. influenzae* is a small (1- × 0.3-μm) gram-negative organism of variable shape; thus, it is often described as a pleomorphic coccobacillus. In clinical specimens such as cerebrospinal fluid (CSF) and sputum, *H. influenzae* frequently stains only faintly with safranin and therefore can easily be overlooked.

H. influenzae grows both aerobically and anaerobically. As the Latin name implies, *Haemophilus* is the "blood-loving bacterium." A prerequisite for growth is a requirement for the media to contain blood, but the blood substrate needs to be preheated; hence, *H. influenzae* grows on chocolate agar plates (slowly heated to 80°C [176°F]), whereas ordinary blood plates (without preheating) will not sustain growth. Its aerobic growth requires two factors: hemin (designated X factor) and nicotinamide adenine dinucleotide (V factor). While MALDI-ToF (Matrix-Assisted Laser Desorption/Ionization - Time of Flight) is now widely used for bacterial identification, these requirements are still sometimes applied in clinical laboratories. However, using phenotypic methods for differentiating among *Haemophilus* species has limitations, as the growing number of whole-genome sequences of *Haemophilus* isolates from the human respiratory tract is revealing complex genetic relationships among *Haemophilus* species (see "Diagnosis," below).

Six major serotypes of *H. influenzae* have been identified; designated *a* through *f*, they are based on antigenically distinct polysaccharide capsules. In addition, some strains lack a polysaccharide capsule and are referred to as *nontypeable* strains. Type b and nontypeable strains are the most relevant strains clinically **(Table 162-1)**, although encapsulated strains other than type b can cause disease. An *H. influenzae* type b isolate was the first free-living organism to have its entire genome sequenced in the early 1990s.

The antigenically distinct type b capsule is a linear polymer composed of ribosyl-ribitol phosphate. Strains of *H. influenzae* type b (Hib) cause disease primarily in infants and children <6 years of age, although it is nowadays rare due to effective vaccine campaigns. Nontypeable strains are primarily mucosal pathogens but occasionally cause invasive disease, particularly in immunocompromised hosts and the elderly.

EPIDEMIOLOGY AND TRANSMISSION

H. influenzae, an exclusively human pathogen, is spread by airborne droplets or by direct contact with secretions or fomites. Between

TABLE 162-1 Characteristics of Type b and Nontypeable Strains of *Haemophilus influenzae*

FEATURE	TYPE b STRAINS	NONTYPEABLE STRAINS
Capsule	Ribosyl-ribitol phosphate	Unencapsulated
Pathogenesis	Invasive infections due to hematogenous spread	Mucosal infections due to contiguous spread
Clinical manifestations	Meningitis and invasive infections in incompletely immunized infants and children	Otitis media in infants and children; lower respiratory tract infections in adults with chronic bronchitis
Evolutionary history	Basically clonal	Genetically diverse
Vaccine	Highly effective conjugate vaccines	Protein D used as carrier protein in pneumococcal vaccine approved in Europe: GSK Synflorix. Others under development

30 and 60% of healthy preschool children are intermittent carriers. Colonization with nontypeable *H. influenzae* is a dynamic process; new strains are acquired and other strains are replaced periodically. This is particularly evident in patients with chronic obstructive pulmonary disease (COPD).

The widespread use of Hib conjugate vaccines in many industrialized countries has resulted in striking decreases in the rate of nasopharyngeal colonization by Hib and in the incidence of Hib infection **(Fig. 162-1)**. On a global basis, invasive Hib disease occurs predominantly in unimmunized children and in those who have not completed the primary immunization series. Most World Health Organization member countries have introduced Hib conjugate vaccination, but a large number of the world's children remain unimmunized, principally in countries without national vaccine programs. In addition to immunocompromised hosts, certain groups have a higher incidence of invasive disease with encapsulated *H. influenzae* (along with capsular serotypes other than Hib) than the general population, including African Americans, Australian Aboriginal children, and Native American groups. Although this increased incidence has not yet been accounted for, several factors may be relevant, including age at exposure to the bacterial species, socioeconomic conditions, and genetic differences.

PATHOGENESIS

Hib strains cause systemic disease by invasion and hematogenous spread from the respiratory tract to distant sites such as the meninges, bones, and joints. The type b polysaccharide capsule is an important virulence factor affecting the bacterium's ability to avoid opsonization and cause systemic disease.

Nontypeable strains cause disease by local invasion of mucosal surfaces. Otitis media results when bacteria reach the middle ear by way of the eustachian tube. Adults with COPD experience recurrent lower respiratory tract infection due to nontypeable strains. In addition, nontypeable *H. influenzae* persist in the lower airways of adults with COPD in both extracellular and intracellular locations, contributing to the airway inflammation that is a hallmark of the disease. Nontypeable strains that cause infection in adults with COPD differ in pathogenic potential and genome content from strains that cause otitis media. In the middle ear, nontypeable strains form biofilms. More resistant to host clearance mechanisms and to antibiotics than are planktonic bacteria, biofilms are associated with chronic and recurrent otitis media. Nontypeable *H. influenzae* persist in the human respiratory tract and cause infection by altering expression of genes through slipped-strand mispairing and through phase-variable expression of DNA methylase genes that control the expression of multiple genes that play a role in virulence.

The incidence of invasive disease caused by nontypeable strains is low but appears to have increased over the past decade; one possible explanation for this increase might be that unencapsulated and hence nontypeable *H. influenzae* were overlooked in the past during the prevaccine era due to the high incidence of Hib. Most strains that cause invasive disease are genetically and phenotypically diverse.

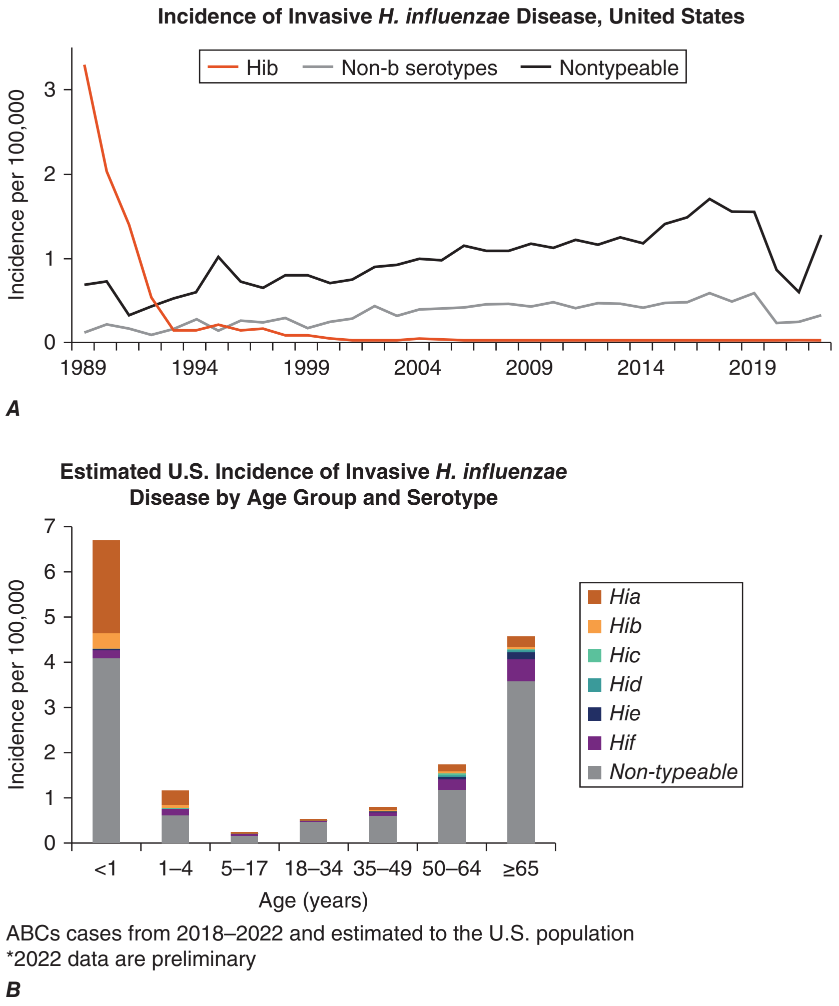

ABCs cases from 2018–2022 and estimated to the U.S. population
*2022 data are preliminary

FIGURE 162-1 Estimated incidence rates (per 100,000) of invasive disease due to *Haemophilus influenzae* caused by Hib, non-b serotypes, and nontypeable bacteria in the United States, 1994–2022. ***A.*** Disease caused by Hib remains low. In 2020–2021, disease caused by non-b serotypes and nontypeable bacteria increased. ***B.*** Estimated incidence rates (per 100,000) of invasive *H. influenzae* disease by serotype and age in the United States, 2018–2022. *(Source: Active Bacterial Core surveillance, Centers for Disease Control and Prevention, https://www.cdc.gov/hi-disease/surveillance.html.)*

IMMUNE RESPONSE

Antibody to the capsule is important in protection from infection by Hib strains. The level of (maternally acquired) serum antibody to the capsular polysaccharide, which is a polymer of polyribitol ribose phosphate (PRP), declines from birth to 6 months of age and, in the absence of vaccination, remains low until ~2 or 3 years of age. The age at the antibody nadir correlates with that of the peak incidence of type b disease. Antibody to PRP then appears partly due to exposure to Hib or cross-reacting antigens. Systemic Hib disease is unusual after the age of 6 years because of the presence of protective antibody. Vaccines in which PRP is conjugated to protein carrier molecules have been developed and are now used widely. These vaccines generate an antibody response to PRP in infants and effectively prevent invasive infections in infants and children.

Since nontypeable strains lack a capsule, the immune response to infection is directed at noncapsular antigens. These antigens have generated considerable interest as immune targets and potential vaccine components. The human immune response to nontypeable strains appears to be strain-specific, a characteristic that accounts in part for the propensity of these strains to cause recurrent otitis media and recurrent exacerbations of chronic bronchitis in immunocompetent hosts.

CLINICAL MANIFESTATIONS

Hib The most serious manifestation of infection with Hib is *meningitis* **(Chap. 143)**, which primarily affects children <2 years of age. The clinical manifestations of Hib meningitis are similar to those of meningitis caused by other bacterial pathogens. Fever and altered central nervous system function are the most common features at presentation. Nuchal rigidity may or may not be evident. Subdural effusion, the most common complication, is suspected when, despite 2 or 3 days of appropriate antibiotic therapy, the infant has seizures, hemiparesis, or continued obtundation. The overall mortality rate from Hib meningitis is ~5%, and the morbidity rate is high. Of survivors, 6% have permanent sensorineural hearing loss, and about one-fourth have a significant disability of some type. If more subtle disabilities are sought, up to half of survivors are found to have some neurologic sequelae, such as partial hearing loss and delayed language development.

Epiglottitis **(Chap. 37)** is a life-threatening Hib infection involving cellulitis of the epiglottis and supraglottic tissues. It can lead to acute upper-airway obstruction. Its unique epidemiologic features are its occurrence in an older age group (2–7 years old) than other Hib infections and its absence among Navajo Native Americans and Alaskan Eskimos. Sore throat and fever rapidly progress to dysphagia, drooling, and airway obstruction. In countries with Hib vaccine child immunization programs, epiglottitis now occurs primarily in adults.

Cellulitis **(Chap. 134)** due to Hib occurs in young children. The most common location is on the head or neck; the involved area sometimes takes on a characteristic bluish-red color. Most patients have bacteremia, and 10% have an additional focus of infection.

Hib causes *pneumonia* in infants. The infection is clinically indistinguishable from other types of bacterial pneumonia (e.g., pneumococcal pneumonia) except that Hib is more likely to involve the pleura. Several less common invasive conditions can be important clinical manifestations of Hib infection in children. These include osteomyelitis, septic arthritis, pericarditis, orbital cellulitis, endophthalmitis, urinary tract infection, abscesses, and bacteremia without an identifiable focus.

Non–type b encapsulated strains of *H. influenzae* (types a, c, d, e, and f) are unusual causes of invasive infection manifested predominantly by bacteremia and pneumonia. *H. influenzae* type a infections are seen with increased frequency in indigenous populations of North America, and these strains are predominantly clonal. Most infections due to non–type b encapsulated strains occur in the setting of underlying conditions including immunosuppression.

Nontypeable *H. influenzae* Nontypeable *H. influenzae* is the most common bacterial cause of exacerbations of COPD; these exacerbations are characterized by increased cough, sputum production, and shortness of breath. Fever is low-grade, and no infiltrates are evident on chest x-ray. Nontypeable strains also cause community-acquired bacterial pneumonia in adults, especially among patients with COPD or HIV or other immunocompromised hosts. The clinical features of *H. influenzae* pneumonia are similar to those of other types of bacterial pneumonia, including pneumococcal pneumonia. In recent years, pneumonia due to nontypeable *H. influenzae* has been more common than pneumonia caused by *Streptococcus pneumoniae* **(Fig. 162-2)**.

Nontypeable *H. influenzae* is one of the three most common causes of childhood otitis media (the other two being *S. pneumoniae* and *Moraxella catarrhalis*) **(Chap. 37)**. Infants are febrile and irritable, while older children report ear pain. Symptoms of viral upper-respiratory infection often precede otitis media. The diagnosis is made by pneumatic otoscopy. An etiologic diagnosis, although not routinely sought, can be established by tympanocentesis and culture of middle-ear fluid. Clinical features associated with *H. influenzae* otitis media include a history of recurrent episodes, treatment failure, concomitant conjunctivitis, bilateral otitis media, and recent antimicrobial therapy. The increasing use of pneumococcal polysaccharide conjugate vaccines in most countries has resulted in an overall decrease in otitis media and its complications. However, a relative increase in the proportion of otitis media caused by *H. influenzae* in children failing initial antimicrobial therapy or with recurrent episodes has occurred. Continued monitoring of the incidence and etiology of otitis media will be important.

Nontypeable *H. influenzae* also causes puerperal sepsis and is an important cause of neonatal bacteremia. These nontypeable strains, provisionally named *Haemophilus quentini*, are closely related to but distinct from *H. haemolyticus*, tend to be of biotype IV, and cause invasive disease after colonizing the female genital tract.

Nontypeable *H. influenzae* causes sinusitis **(Chap. 37)** in adults and children. In addition, the bacterium is a less common cause of various invasive infections. These infections include bacteremia, empyema, adult epiglottitis, pericarditis, cellulitis, septic arthritis, osteomyelitis, endocarditis, cholecystitis, intraabdominal infections, urinary tract infections, mastoiditis, and aortic graft infection. Most *H. influenzae* invasive infections in countries where Hib vaccines are used widely are caused by nontypeable strains, and a recent increased incidence of such infections has been observed. Although most strains of nontypeable *H. influenzae* that cause invasive infections are genetically diverse, recent localized clusters of infections have been caused by clonally related strains. Continued monitoring will be important. Many patients with *H. influenzae* bacteremia have an underlying condition, such as HIV infection, cardiopulmonary disease, alcoholism, or cancer.

DIAGNOSIS

The most reliable method for establishing a diagnosis of invasive *H. influenzae* infection is recovery of the organism in culture in a normally sterile body site, such as blood, CSF, or joint fluid. Isolation of bacteria is important for antimicrobial susceptibility testing. Rapid polymerase chain reactions (PCR) and other DNA-based methods are commonly used in modern clinical laboratories.

H. influenzae isolated from the respiratory tract must be distinguished from a complex flora and from other *Haemophilus* species. Particular caution must be used to distinguish *H. influenzae* from *Haemophilus haemolyticus*, a respiratory tract commensal with identical growth requirements. *H. haemolyticus* has classically been distinguished from *H. influenzae* by the hemolysis of the former species on horse blood agar. However, a significant proportion of isolates of *H. haemolyticus* have now been recognized as nonhemolytic. Analysis of

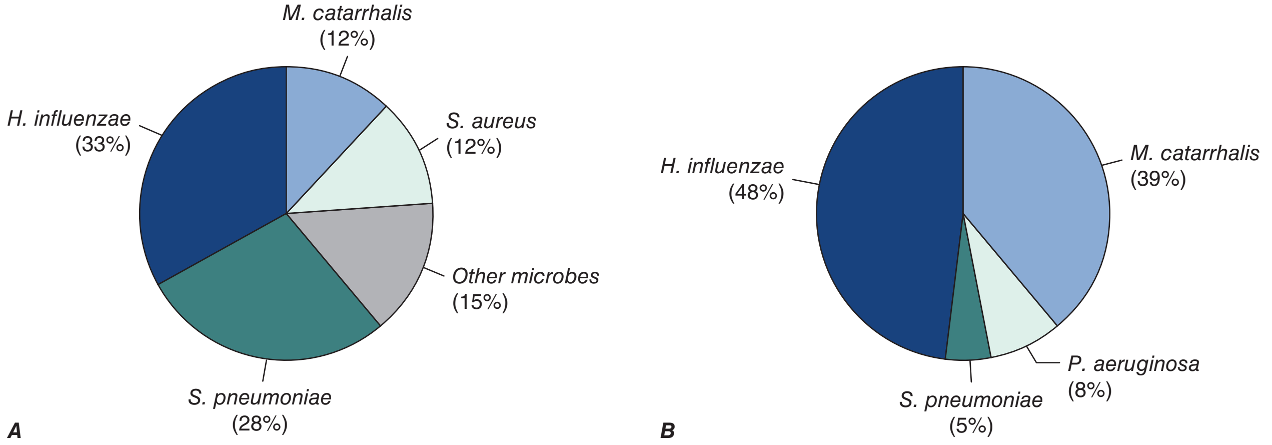

FIGURE 162-2. Importance of nontypeable *Haemophilus influenzae* and *Moraxella catarrhalis* in community-acquired pneumonia (CAP) and in exacerbations of COPD patients. ***A.*** Meta-analysis of bacterial detection in lower respiratory tract specimens using molecular panels or quantitative assays from patients with CAP. ***B.*** Cumulative results of a prospective study (1994–2004) of bacterial infection in chronic obstructive pulmonary disease (COPD) showing etiology of exacerbations. Numbers of exacerbations shown indicate acquisition of a new strain simultaneous with clinical symptoms of an exacerbation. *(Data source for A: NJ Gadsby et al: Clin Infect Dis 62:817, 2016; DN Gilbert et al: Diagn Microbiol Infect Dis 99:115246, 2021; S Serigstad et al: Sci Rep 12:326, 2022; M Fally et al: Infect Dis (Lond) 53:122, 2021.)*

various genotypic markers, including 16S ribosomal sequences, superoxide dismutase, outer-membrane protein P6, protein D, and fuculose kinase, can be used to distinguish these two species. The availability of whole genome sequences of an increasing number of *Haemophilus* isolates from the human upper respiratory tract has revealed complex genomic relationships among *Haemophilus* species, suggesting a genetic continuum between some *Haemophilus* species.

The presence of gram-negative coccobacilli in Gram-stained CSF is strong evidence for Hib meningitis. Recovery of the organism or detection of *H. influenzae* DNA from CSF confirms the diagnosis. The same applies for other normally sterile body fluids, such as blood, joint fluid, pleural fluid, pericardial fluid, and subdural effusion, and are thus also confirmatory in other infections.

Because nontypeable *H. influenzae* is primarily a mucosal pathogen, it is a component of a mixed flora; thus, etiologic diagnosis is challenging. Nontypeable *H. influenzae* infection is strongly suggested by the predominance of gram-negative coccobacilli among abundant polymorphonuclear leukocytes in a Gram-stained sputum specimen from a patient in whom pneumonia is suspected. Although bacteremia is detectable in a small proportion of patients with pneumonia due to nontypeable *H. influenzae*, most such patients have negative blood cultures.

A diagnosis of otitis media is based on the detection by pneumatic otoscopy of fluid in the middle ear. An etiologic diagnosis requires tympanocentesis but is not routinely sought. An invasive procedure is also required to determine the etiology of sinusitis; thus, treatment is often empirical once the diagnosis is suspected in light of clinical symptoms and sinus radiographs.

TREATMENT

Haemophilus influenzae

Initial therapy for meningitis due to Hib should consist of a cephalosporin such as ceftriaxone or cefotaxime. For children, the dosage of ceftriaxone is 75–100 mg/kg daily given in two doses 12 h apart. The pediatric dosage of cefotaxime is 200 mg/kg daily given in four doses 6 h apart. Adult dosages are 2 g every 12 h for ceftriaxone and 2 g every 4–6 h for cefotaxime. An alternative regimen for initial therapy is ampicillin (200–300 mg/kg daily in four divided doses) plus chloramphenicol (75–100 mg/kg daily in four divided doses). Therapy should continue for a total of 1–2 weeks.

Administration of glucocorticoids to patients with Hib meningitis reduces the incidence of neurologic sequelae. The presumed mechanism is reduction of the inflammation induced by bacterial cell-wall mediators of inflammation when cells are killed by antimicrobial agents. Dexamethasone (0.6 mg/kg per day intravenously in four divided doses for 2 days) is recommended for the treatment of Hib meningitis in children >2 months of age.

Invasive infections other than meningitis are treated with the same antimicrobial agents. For epiglottitis, the dosage of ceftriaxone is 50 mg/kg daily, and the dosage of cefotaxime is 150 mg/kg daily, given in three divided doses 8 h apart. Epiglottitis constitutes a medical emergency, and maintenance of an airway is critical. The duration of therapy is determined by the clinical response. A course of 1–2 weeks is usually appropriate.

Many infections caused by nontypeable strains of *H. influenzae*, such as otitis media, sinusitis, and exacerbations of COPD, can be treated with oral antimicrobial agents. Approximately 30–40% of nontypeable strains produce β-lactamase (with the exact proportion depending on geographic location), and these strains are resistant to ampicillin as well as amoxicillin. Several agents have excellent activity against nontypeable *H. influenzae*, including amoxicillin/clavulanic acid, various extended-spectrum cephalosporins, and the macrolides azithromycin and clarithromycin. Fluoroquinolones are highly active against *H. influenzae* and are useful in adults with exacerbations of COPD. However, fluoroquinolones are not currently recommended for the treatment of children or pregnant women because of possible effects on articular cartilage.

In addition to β-lactamase production, alteration of penicillin-binding proteins—a second mechanism of ampicillin resistance—has been detected in isolates of *H. influenzae*. Although still rare in the United States, these β-lactamase-negative ampicillin-resistant strains are common in Japan and are increasing in prevalence in Europe. Resistance to macrolides is also being observed with increasing frequency globally. Continued monitoring of the evolving antimicrobial susceptibility patterns of *H. influenzae* will be important.

PREVENTION

Vaccination **(See also Chap. 129)** Three monovalent conjugate vaccines that prevent invasive infections with Hib in infants and children are licensed in the United States. In addition to eliciting protective antibodies, these vaccines prevent disease by reducing rates of pharyngeal colonization with Hib. The widespread use of conjugate vaccines has dramatically reduced the incidence of Hib disease in developed countries. Even though the manufacture of Hib vaccines is costly, vaccination is cost-effective. The Global Alliance for Vaccines and Immunizations has recognized the underuse of Hib conjugate vaccines.

The disease burden has been reduced in developing countries that have implemented routine vaccination (e.g., The Gambia, Chile). An important obstacle to more widespread vaccination is the lack of data on the epidemiology and burden of Hib disease in many developing countries.

All children should be immunized with a Hib conjugate vaccine, receiving the first dose at ~2 months of age, the rest of the primary series at 2–6 months of age, and a booster dose at 12–15 months of age. Specific recommendations vary for the different conjugate vaccines. The reader is referred to the recommendations of the American Academy of Pediatrics (**Chap. 129** and *https://www.aap.org/en/patient-care/immunizations/vaccination-recommendations-by-the-aap/*).

Currently, no vaccines are available specifically for the prevention of disease caused by nontypeable *H. influenzae*. However, a vaccine that contains protein D—a surface protein of *H. influenzae*—conjugated to pneumococcal polysaccharides is licensed in other countries and is used widely throughout the world. The vaccine has shown partial efficacy in preventing *H. influenzae* otitis media in clinical trials. Vaccine formulations that include surface protein antigens are currently in clinical trials, and additional progress in the development of vaccines against nontypeable *H. influenzae* is anticipated.

Chemoprophylaxis The risk of secondary disease is greater than normal among household contacts of patients with Hib disease. Therefore, all children and adults (except pregnant women) in households with an index case and at least one incompletely immunized contact <4 years of age should receive prophylaxis with oral rifampin. When two or more cases of invasive Hib disease have occurred within 60 days at a child-care facility attended by incompletely vaccinated children, administration of rifampin to all attendees and personnel is indicated, as it is for household contacts. Chemoprophylaxis is not indicated in nursery and child-care contacts of a single index case. The reader is referred to the recommendations of the American Academy of Pediatrics.

HAEMOPHILUS DUCREYI

Haemophilus ducreyi is the etiologic agent of chancroid **(Chap. 141)**, a sexually transmitted disease characterized by genital ulceration and inguinal adenitis. In addition to being a cause of morbidity in itself, chancroid is associated with HIV infection because of the role played by genital ulceration in HIV transmission. Chancroid increases the efficiency of transmission of and the degree of susceptibility to HIV infection. *H. ducreyi* has also been recognized as an important cause of non–sexually transmitted cutaneous ulcers.

MICROBIOLOGY

H. ducreyi is a highly fastidious coccobacillary gram-negative bacterium whose growth requires X factor (hemin). Although in light of this requirement, the bacterium has been classified in the genus *Haemophilus*, DNA homology and chemotaxonomic studies have established substantial differences between *H. ducreyi* and other *Haemophilus* species. Taxonomic reclassification of the organism is likely in the future but awaits further study. Ulcers contain predominantly T cells. The fact that patients who have had chancroid may have repeated infections indicates that infection does not confer protection.

EPIDEMIOLOGY AND PREVALENCE

The prevalence of chancroid has steadily declined in the United States and worldwide over the past decade and a half. The infection appears to be more common in developing countries. Transmission is predominantly heterosexual, and cases in males have outnumbered those in females by ratios of 3:1 to 25:1 during outbreaks. Contact with commercial sex workers and illicit drug use are strongly associated with chancroid. Most cases in developed countries are sporadic.

H. ducreyi has emerged as a major cause of cutaneous ulcers in children in developing countries, particularly in the South Pacific and Africa. Strains that cause cutaneous ulcers have genome sequences that are nearly identical to class I strains (of two related classes) of *H. ducreyi* that cause genital ulcers.

CLINICAL MANIFESTATIONS AND DIFFERENTIAL DIAGNOSIS

Infection is acquired as the result of a break in the epithelium during sexual contact with an infected individual. After an incubation period of 4–7 days, the initial lesion—a papule with surrounding erythema—appears. In 2 or 3 days, the papule evolves into a pustule, which spontaneously ruptures and forms a sharply circumscribed ulcer that generally is not indurated **(Fig. 162-3)**. The ulcers are painful and bleed easily; little or no inflammation of the surrounding skin is evident. Approximately half of patients develop enlarged, tender inguinal lymph nodes, which frequently become fluctuant and spontaneously rupture. Patients usually seek medical care after 1–3 weeks of painful symptoms.

The presentation of chancroid does not usually include all of the typical clinical features and is sometimes atypical. Multiple ulcers can coalesce to form giant ulcers. Ulcers can appear and then resolve, with inguinal adenitis (Fig. 162-3) and suppuration following 1–3 weeks later; this clinical picture can be confused with that of lymphogranuloma venereum **(Chap. 194)**. Multiple small ulcers can resemble folliculitis. Other differential diagnostic considerations include the various infections causing genital ulceration, such as primary syphilis, secondary syphilis (condyloma latum), genital herpes, and donovanosis. In rare cases, chancroid lesions become secondarily infected with bacteria; the result is extensive inflammation.

Non–sexually transmitted cutaneous ulcers caused by *H. ducreyi* resemble those of yaws caused by *Treponema pallidum* subspecies *pertenue*, which is endemic in regions where *H. ducreyi* cutaneous ulcers are seen. Ulcers caused by *H. ducreyi* are less likely than those of yaws to show central granulating tissue and less likely to have indurated edges, but substantial overlap in clinical characteristics exists.

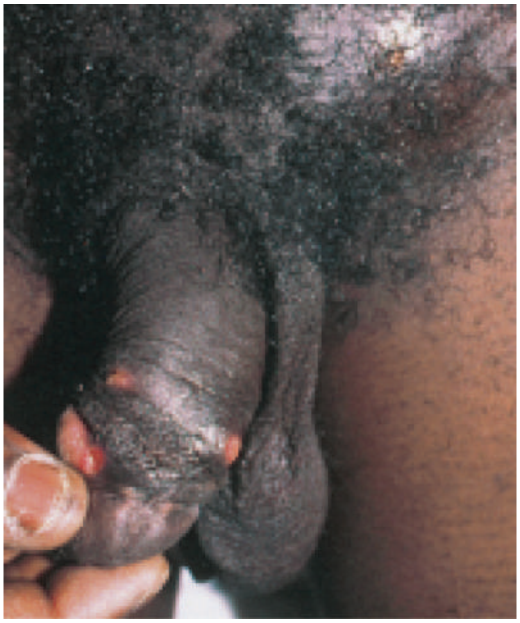

FIGURE 162-3 Chancroid with characteristic penile ulcers and associated left inguinal adenitis (bubo).

DIAGNOSIS

Clinical diagnosis of chancroid is often inaccurate, and laboratory confirmation should be attempted in suspected cases. An accurate diagnosis of chancroid relies on culture of *H. ducreyi* from the lesion or from an aspirate of suppurative lymph nodes. Since the organism can be difficult to grow, the use of selective and supplemented media is necessary. DNA detection by PCR is in use in larger clinical laboratories.

A probable diagnosis of sexually transmitted chancroid can be made when the following criteria are met: (1) one or more painful genital ulcers; (2) no evidence of *T. pallidum* infection by dark-field examination of ulcer exudate or by a negative serologic test for syphilis performed at least 7 days after ulcer onset; (3) a typical clinical presentation for chancroid; and (4) a negative test for herpes simplex virus in the ulcer exudate. A serologic test for syphilis does not distinguish cutaneous ulcers due to *H. ducreyi* from those due to yaws, but PCR should be performed.

TREATMENT

Haemophilus ducreyi

Treatment regimens for both genital and cutaneous infections include either (1) a single 1-g oral dose of azithromycin; (2) ceftriaxone (250 mg intramuscularly in a single dose); (3) ciprofloxacin (500 mg by mouth twice a day for 3 days); or (4) erythromycin base (500 mg by mouth three times a day for 7 days). Isolates from patients who do not respond promptly to treatment should be tested for antimicrobial resistance. In patients with HIV infection, healing may be slow and longer courses of treatment may be necessary. Clinical treatment failure in HIV-seropositive patients may reflect co-infection, especially with herpes simplex virus. Contacts of patients with chancroid should be identified and treated, whether or not symptoms are present, if they have had sexual contact with the patient during the 10 days preceding the patient's onset of symptoms.

MORAXELLA CATARRHALIS

MICROBIOLOGY

M. catarrhalis is an unencapsulated gram-negative diplococcus whose ecologic niche is the human respiratory tract. The organism was initially designated *Micrococcus catarrhalis*. Its name was changed to *Neisseria catarrhalis* in 1970 because of phenotypic similarities to commensal *Neisseria* species. On the basis of more rigorous analysis of genetic relatedness, *Moraxella catarrhalis* is now the widely accepted name for this species.

EPIDEMIOLOGY

Nasopharyngeal colonization by *M. catarrhalis* is common in infancy, with colonization rates ranging between 33 and 100% and depending on geographic location. Several factors probably account for this geographic variation, including living conditions, day-care attendance, hygiene, household smoking, and population genetics. The prevalence of colonization decreases steadily with age.

The widespread use of pneumococcal conjugate vaccines in some countries has resulted in alterations in patterns of nasopharyngeal colonization in resident populations. A relative increase in colonization by nonvaccine pneumococcal serotypes, nontypeable *H. influenzae*, and *M. catarrhalis* has occurred. These changes in colonization patterns may be altering the distribution of pathogens of both otitis media and sinusitis in children.

PATHOGENESIS

M. catarrhalis causes mucosal infections of the respiratory tract by contiguous spread from its colonizing site in the upper airway. A preceding viral upper respiratory tract infection is a common inciting event for otitis media. In exacerbations of COPD, the acquisition of new strains is critical for pathogenesis. Strains exhibit substantial genetic diversity and differences in virulence properties.

The expression of several adhesin molecules with differing specificities for various host cell receptors reflects the importance of adherence to the respiratory epithelial surface in the pathogenesis of infection. *M. catarrhalis* invades multiple cell types. Its intracellular residence in lymphoid tissue provides a potential reservoir for persistence in the human respiratory tract. Like many gram-negative bacteria, *M. catarrhalis* sheds vesicles into the surrounding environment. The vesicles are internalized by host cells and mediate several virulence mechanisms, including induction of inflammation and delivery of β-lactamase, that can promote the survival of co-pathogens such as group A streptococci and *H. influenzae*.

CLINICAL MANIFESTATIONS

In children, *M. catarrhalis* causes predominantly mucosal infections when the bacterium migrates from the nasopharynx to the middle ear or the sinuses **(Chap. 37)**. The inciting event for both otitis media and sinusitis is often a preceding viral infection. It has, for example, been shown that rhinovirus promotes adhesion of *M. catarrhalis*. Overall, cultures of middle-ear fluid obtained by tympanocentesis indicate that *M. catarrhalis* causes 15–20% of cases of acute otitis media. More sensitive molecular analysis of middle ear fluid detects *M. catarrhalis* alone or with other pathogens in 30–50% of middle ear fluid samples from children with otitis media. Acute otitis media caused by *M. catarrhalis* or nontypeable *H. influenzae* is clinically milder than otitis media caused by *S. pneumoniae*, with less fever and a lower prevalence of a red bulging tympanic membrane. However, substantial overlap makes it impossible to predict etiology in an individual child on the basis of clinical features.

A small proportion of viral upper respiratory tract infections are complicated by bacterial sinusitis. Cultures of sinus puncture aspirates show that *M. catarrhalis* accounts for ~20% of cases of acute bacterial sinusitis in children and for a smaller proportion in adults.

M. catarrhalis is a common cause of exacerbations in adults with COPD. The bacterium used to be overlooked in this clinical setting because it had long been considered a commensal and because it was easily mistaken for commensal *Neisseria* species in cultures of respiratory secretions (see "Diagnosis," below). Several independent lines of evidence have established *M. catarrhalis* as a pathogen in COPD. These include (1) the demonstration of *M. catarrhalis* in the lower airways during exacerbations, (2) the association of exacerbation with acquisition of new strains, (3) elevations of inflammatory markers in association with *M. catarrhalis*, and (4) the development of specific immune responses following infection. *M. catarrhalis* is the second most common bacterial cause of COPD exacerbations (after *H. influenzae*), as shown in a 10-year prospective study; the distribution of exacerbations associated with new-strain acquisitions is shown in Fig. 162-2. Not included are culture-negative cases or cases from which a pathogen had been previously isolated. With the application of rigorous clinical criteria for defining the etiology of exacerbations (both culture-positive and culture-negative), ~10% of all exacerbations in the same study were caused by *M. catarrhalis*. The clinical features of an exacerbation due to *M. catarrhalis* are similar to those of exacerbations due to other bacterial pathogens, including *H. influenzae* and *S. pneumoniae*. The cardinal symptoms are cough with increased sputum production, sputum purulence, and dyspnea in comparison with baseline symptoms.

Pneumonia due to *M. catarrhalis* occurs in the elderly, particularly in the setting of underlying cardiopulmonary disease, but is infrequent. Invasive infections, such as bacteremia, endocarditis, neonatal meningitis, and septic arthritis, are rare.

DIAGNOSIS

Tympanocentesis is required for etiologic diagnosis of otitis media, but this procedure is not performed routinely. Therefore, treatment of otitis media is generally empirical. Similarly, an etiologic diagnosis of sinusitis requires an invasive procedure and thus is usually not available to the clinician. Isolation of *M. catarrhalis* from an expectorated sputum sample from an adult experiencing clinical symptoms of an exacerbation is suggestive, but not diagnostic, of *M. catarrhalis* as the cause.

Upon culture, colonies of *M. catarrhalis* resemble those of commensal neisseriae that are part of the normal upper airway flora. As mentioned above, the difficulty in distinguishing colonies of *M. catarrhalis* from neisserial colonies in cultures of respiratory secretions explains in part why *M. catarrhalis* has been overlooked as a pathogen. In contrast to these *Neisseria* species, *M. catarrhalis* colonies can be slid across the agar surface without disruption (the "hockey puck sign"). In addition, after 48 h of growth, *M. catarrhalis* colonies take on a pink color and tend to be larger than neisserial colonies.

TREATMENT

Moraxella catarrhalis

M. catarrhalis rapidly acquired β-lactamases during the 1970s and 1980s, the fastest documented spread of β-lactamases encoding plasmids. Antimicrobial susceptibility patterns have remained relatively stable since that time, with >90% of strains now producing β-lactamase and thus resistant to amoxicillin. Otitis media in children and exacerbations of COPD in adults are generally managed empirically with antimicrobial agents that are active against *S. pneumoniae*, *H. influenzae*, and *M. catarrhalis*. Most strains of *M. catarrhalis* are susceptible to amoxicillin/clavulanic acid, extended-spectrum cephalosporins, newer macrolides (azithromycin, clarithromycin), trimethoprim-sulfamethoxazole, and fluoroquinolones. However, recent reports from several centers in Asia show substantial resistance to macrolides and fluoroquinolones, indicating emerging resistance. Continued monitoring of global antimicrobial susceptibility patterns of *M. catarrhalis* will be critical.

FURTHER READING

Ahearn CP et al: Insights on persistent airway Infection by nontypeable *Haemophilus influenzae* in chronic obstructive pulmonary disease. Pathog Dis 75:ftx042, 2017.

Jalalvand F, Riesbeck K: Update on non-typeable *Haemophilus influenzae*-mediated disease and vaccine development. Expert Rev Vaccines 17:503, 2018.

Lewis DA, Mitja O: *Haemophilus ducreyi*: From sexually transmitted infection to skin ulcer pathogen. Curr Opin Infect Dis 29:52, 2016.

Perez AC, Murphy TF: Potential impact of a *Moraxella catarrhalis* vaccine in COPD. Vaccine 37:5551, 2019.

Su YC et al: The interplay between immune response and bacterial infection in COPD: Focus upon non-typeable *Haemophilus influenzae*. Front Immunol 9:2530, 2018.

163

Infections Due to the HACEK Group and Miscellaneous Gram-Negative Bacteria

Tamar F. Barlam

THE HACEK GROUP

HACEK organisms are a group of fastidious, slow-growing, gram-negative bacteria whose growth requires an atmosphere of carbon dioxide. These organisms do not grow on media routinely used for enteric bacteria (e.g., MacConkey agar). Species belonging to this group include several *Haemophilus* species, *Aggregatibacter* (formerly

Actinobacillus) species, *Cardiobacterium* species, *Eikenella corrodens*, and *Kingella kingae*. HACEK bacteria normally reside in the oral cavity and have been associated with local infections in the mouth. They are also known to cause severe systemic infections—such as bacteremia and bacterial endocarditis, which can develop on either native or prosthetic valves **(Chap. 133)**. In a nationwide survey in Denmark, HACEK bacteremias were most often due to *Haemophilus* species, followed by *Aggregatibacter* species. HACEK bacteremia is strongly predictive of underlying infective endocarditis (overall positive predictive value, 60%). However, this association varies significantly by organism. For example, infective endocarditis was diagnosed in 62–100% of patients with *Aggregatibacter actinomycetemcomitans* bacteremia, 17–55% with *Haemophilus parainfluenzae* bacteremia, 50% with *Cardiobacterium hominis* bacteremia, but only 0–6% of patients with *Eikenella* bacteremia.

In large series, 0.8–6% of cases of infective endocarditis are attributable to HACEK organisms, most often *Aggregatibacter* species, *Haemophilus* species, and *Cardiobacterium hominis*. Invasive infection typically occurs in patients with a history of cardiac valvular disease or prosthetic valves, often in the setting of a recent dental procedure or nasopharyngeal infection. The aortic and mitral valves are most commonly affected. The clinical course of HACEK endocarditis tends to be subacute, particularly with *Aggregatibacter* or *Cardiobacterium*. However, *K. kingae* endocarditis may have a more aggressive presentation. Compared with non-HACEK endocarditis, HACEK endocarditis occurs in younger patients and has been more frequently associated with embolic, vascular, and immunologic manifestations. Systemic embolization is common. The overall prevalence of major emboli associated with HACEK endocarditis ranges from 28 to 71% in different series. On echocardiography, valvular vegetations are seen in up to 85% of patients. *Aggregatibacter* and *Haemophilus* species cause mitral valve vegetations most often; *Cardiobacterium* is associated with aortic valve vegetations.

The microbiology laboratory should be alerted when a HACEK organism is being considered. Most cultures that ultimately yield a HACEK organism become positive within the first week, especially with improved culture systems such as BACTEC. Studies have not shown that prolonged incubation increases laboratory recovery of clinically significant HACEK isolates. Polymerase chain reaction (PCR) techniques, such as gene amplification of 16S rRNA, can facilitate the diagnosis of HACEK infection of blood or cardiac valves. Other tools, such as matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometry performed directly on agar colonies, can increase the accuracy and speed of diagnosis of HACEK infections.

Because of HACEK organisms' slow growth, antimicrobial susceptibility testing may be difficult, and β-lactamase production may not be detected. Resistance is most commonly noted in *Haemophilus* and *Aggregatibacter* species. Etest methodology may increase the accuracy of susceptibility testing. In recent studies, ceftriaxone and levofloxacin have been active against all isolates. The overall prognosis in both native-valve and prosthetic-valve HACEK endocarditis is excellent and is significantly better than that in endocarditis caused by non-HACEK pathogens.

***Haemophilus* Species** *Haemophilus parainfluenzae* is the most common *Haemophilus* species isolated from cases of HACEK endocarditis. Of patients with HACEK endocarditis due to *Haemophilus* species, 60% have been ill for <2 months before presentation, and 19–50% develop congestive heart failure. The mitral valve is most frequently involved, and large vegetations resulting in septic emboli, including central nervous system emboli, are common. *H. parainfluenzae* has been isolated from other infections, such as pneumonia; meningitis; brain, dental, pelvic, and liver abscess; urinary tract infection; and septicemia that is often polymicrobial.

***Aggregatibacter* Species** *Aggregatibacter* species are the most common cause of HACEK endocarditis; the species most frequently involved are *A. actinomycetemcomitans*, *A.* (formerly *Haemophilus*) *aphrophilus*, and *A. paraphrophilus*. *Aggregatibacter* is associated with prosthetic-valve endocarditis more often than are *Haemophilus* species. *A. actinomycetemcomitans* can be isolated from soft tissue infections and abscesses in association with *Actinomyces israelii*. Typically, patients who develop *Aggregatibacter* endocarditis have periodontal disease or have recently undergone dental procedures in the setting of underlying cardiac valvular damage. The disease is insidious; patients may be sick for several months before diagnosis. Frequent complications include embolic phenomena, congestive heart failure, and renal failure.

A. actinomycetemcomitans has been isolated from patients with brain abscess, meningitis, endophthalmitis, parotitis, osteomyelitis, urinary tract infection, pneumonia, and empyema, among other infections. *A. aphrophilus* is often associated with bone and joint infection and is an important cause of brain abscess. In one series, *A. aphrophilus* was isolated from brain abscesses in 10% of cases—a rate that is disproportionate to its isolation from oral flora. This species has also been described as a cause of abscess in other organ systems.

***Cardiobacterium* Species** *Cardiobacterium* species, most often *C. hominis*, cause endocarditis primarily in patients with underlying valvular heart disease or with prosthetic valves. These organisms most frequently affect the aortic valve. Many patients have signs and symptoms of long-standing infection before diagnosis, with evidence of arterial embolization, vasculitis, cerebrovascular accidents, immune complex glomerulonephritis, or arthritis at presentation. Embolization, mycotic aneurysms, and congestive heart failure are common complications. A second species, *C. valvarum*, has been described in association with endocarditis.

Eikenella corrodens *E. corrodens* is most frequently recovered from sites of infection in conjunction with other bacterial species. Clinical sources of *E. corrodens* include sites of human bite wounds (clenched-fist injuries), endocarditis, soft tissue infections, osteomyelitis, head and neck infections, respiratory infections, chorioamnionitis, gynecologic infections associated with intrauterine devices, meningitis, brain abscesses, and visceral abscesses. This organism is the least common cause of HACEK endocarditis.

Kingella kingae More than half of cases of *K. kingae* infection are bone and joint infections; the majority of the remaining infections are infective endocarditis, bacteremia, and meningitis. Invasive *K. kingae* infections with bacteremia are associated with upper respiratory tract infections and stomatitis in 80% of cases. Rates of oropharyngeal colonization with *K. kingae* are highest in the first 3 years of life (detected in ~5–10% of children) and appear higher in children regularly attending daycare; colonization coincides with an increased incidence of skeletal infections and other invasive infections due to this organism from the age of 6 months to 4 years. *K. kingae* can be transmitted from child to child and has been the cause of outbreaks among young children. Isolates associated with outbreaks appear to be more virulent and transmissible compared with nonoutbreak strains. *K. kingae* bacteremia can present with a petechial rash similar to that seen in *Neisseria meningitidis* sepsis.

Because of improved microbiologic methodology and molecular methods such as real-time PCR, the isolation of *K. kingae* is increasingly common. Inoculation of clinical specimens (e.g., synovial fluid) into aerobic blood culture bottles enhances recovery of this organism. PCR studies of blood or joint fluid can identify *K. kingae* in culture-negative cases. Some studies have demonstrated that *K. kingae* has surpassed *Staphylococcus aureus* as the leading cause of septic arthritis and osteomyelitis in children.

Infective endocarditis, unlike other infections with *K. kingae*, occurs in older children and adults. The majority of patients have preexisting valvular disease. There is a high incidence of complications, including arterial emboli, cerebrovascular accidents, tricuspid insufficiency, and congestive heart failure with cardiovascular collapse.

TREATMENT

HACEK Endocarditis

(Table 163-1) Ceftriaxone (2 g/d) is first-line therapy for HACEK endocarditis, with a favorable outcome in 80–90% of cases. Data on the use of levofloxacin (750 mg/d) for HACEK endocarditis remain

TABLE 163-1 Treatment of Infections Caused by HACEK-Group and Other Fastidious Gram-Negative Organisms

ORGANISMS	PREFERRED THERAPY	ALTERNATIVE AGENTS	COMMENTS
Haemophilus spp. *Aggregatibacter* spp. *Cardiobacterium* spp. *Eikenella corrodens* *Kingella kingae*	Ceftriaxone (2 g/d)	Ampicillin/sulbactam (3 g of ampicillin q6h) Levofloxacin (750 mg/d)	Ampicillin/sulbactam resistance has been described in *Haemophilus* and *Aggregatibacter* spp. Data on use of levofloxacin for endocarditis therapy are limited. Fluoroquinolones are not recommended for treatment of patients <18 years of age. Penicillin (16–18 million units q4h) or ampicillin (2 g q4h) can be used if the organism is susceptible. However, because of the slow growth of HACEK bacteria, antimicrobial testing may be difficult, and β-lactamase production may not be detected.
Capnocytophaga spp.	Ampicillin/sulbactam (1.5–3 g of ampicillin q6h)	Ceftriaxone (2 g/d q12–24h)	Penicillin (12–18 million units q4h) should be used if the isolate is known to be susceptible.
Pasteurella multocida	Ampicillin/sulbactam (1.5–3 g of ampicillin q6h)	Ceftriaxone (1–2 g/d q12–24h)	Penicillin should be used if the isolate is known to be susceptible. *P. multocida* is also susceptible to tetracyclines and fluoroquinolones.

limited, but this drug can be considered an alternative for treatment of patients intolerant of β-lactam therapy. Of note, *Eikenella* is resistant to clindamycin, metronidazole, and aminoglycosides.

Native-valve endocarditis should be treated for 4 weeks with antibiotics, whereas prosthetic-valve endocarditis requires 6 weeks of therapy. The cure rates for HACEK prosthetic-valve endocarditis appear to be high. Unlike prosthetic-valve endocarditis caused by other gram-negative organisms, HACEK endocarditis is often cured with antibiotic treatment alone—i.e., without surgical intervention. In a recent case-control study, 1-year mortality was significantly lower than infective endocarditis caused by viridans group *Streptococcus*.

OTHER FASTIDIOUS GRAM-NEGATIVE BACTERIA

***Capnocytophaga* Species** Like HACEK organisms, this genus of fastidious, fusiform, gram-negative coccobacilli is facultatively anaerobic and requires an atmosphere enriched in carbon dioxide for optimal growth. *Capnocytophaga* species such as *C. ochracea*, *C. gingivalis*, *C. haemolytica*, and *C. sputigena* are part of the oral flora; most infections are contiguous with the oropharynx (e.g., periodontal disease, respiratory tract infections, cervical abscesses). These organisms have also been associated with sepsis in immunocompromised hosts, particularly neutropenic patients with oral ulcerations, meningitis, endocarditis, cellulitis, osteomyelitis, and septic arthritis. *Capnocytophaga* species have been isolated from many other sites as well, usually as part of a polymicrobial infection. There is a high prevalence of resistance to β-lactams and macrolides in *Capnocytophaga*.

C. canimorsus and *C. cynodegmi* are endogenous to the canine and feline mouth **(Chap. 146)**. Patients infected with these species frequently have a history of dog or cat bites or of exposure without scratches or bites. Asplenia, glucocorticoid therapy, and alcohol abuse are predisposing conditions that can be associated with severe sepsis with shock and disseminated intravascular coagulation. Patients typically have a petechial rash that can progress from purpuric lesions to gangrene.

TREATMENT

Capnocytophaga Infections

(Table 163-1) Because of increasing β-lactamase production, a penicillin derivative plus a β-lactamase inhibitor—such as ampicillin/sulbactam (1.5–3.0 g of ampicillin every 6 h)—is currently recommended for empirical treatment of infections caused by *Capnocytophaga* species. If the isolate is known to be susceptible, infections with *C. canimorsus* should be treated with penicillin (12–18 million units every 4 h). *Capnocytophaga* is also susceptible to third-generation cephalosporins such as ceftriaxone (2 g every 12–24 h), doxycycline (100 mg every 12 h), and clindamycin (600–900 mg every 6–8 h). Antibiotics should be given prophylactically to asplenic patients who have sustained dog-bite injuries.

Pasteurella multocida *P. multocida* is a fastidious, bipolar-staining, gram-negative coccobacillus that colonizes the respiratory and gastrointestinal tracts of domestic animals; oropharyngeal colonization rates are 50–90% in cats and 50–70% in dogs. *P. multocida* can be transmitted to humans through bites or scratches, via the respiratory tract from contact with contaminated dust or infectious droplets, or via deposition of the organism on injured skin or mucosal surfaces during licking. Most human infections affect skin and soft tissue; almost two-thirds of these infections are caused by cats. Patients at the extremes of age or with serious underlying disorders (e.g., cirrhosis, diabetes, cancer) are at increased risk for invasive infection with systemic manifestations, including bacteremia, meningitis, peritonitis, osteomyelitis and septic arthritis, endocarditis, septic shock, ecthyma, necrotizing fasciitis, and purpura fulminans, and are more likely not to have evidence of an animal bite. However, cases have also occurred in healthy individuals of all ages. Mortality has been reported to be as high as 30% in invasive disease, despite active antimicrobial therapy. If inhaled, *P. multocida* can cause acute respiratory tract infection, particularly in patients with underlying sinus and pulmonary disease.

TREATMENT

Pasteurella multocida Infections

(Table 163-1) *P. multocida* is susceptible to penicillin, ampicillin, ampicillin/sulbactam, second- and third-generation cephalosporins, tetracyclines, and fluoroquinolones. β-Lactamase–producing strains have been reported.

OTHER GRAM-NEGATIVE BACTERIA

Achromobacter xylosoxidans *Achromobacter* (previously *Alcaligenes*) *xylosoxidans* is an aerobic nonfermenting gram-negative organism that is probably part of the endogenous intestinal flora. It has been isolated from a variety of water sources, including well water, IV fluids, and humidifiers. Immunocompromised hosts, including patients with cancer and postchemotherapy neutropenia, cirrhosis, chronic renal failure, and cystic fibrosis, are at increased risk for infection. Nosocomial outbreaks and pseudo-outbreaks of infection with *A. xylosoxidans,* and other *Achromobacter* species, have been attributed to contaminated fluids, and clinical illness has been associated with isolates from many sites, including blood (often in the setting of intravascular devices). Community-acquired *A. xylosoxidans* bacteremia usually occurs in the setting of pneumonia. Metastatic skin lesions are present in one-fifth of cases. The reported mortality rate is as high as 67%—a figure similar to rates for other bacteremic gram-negative pneumonias.

TREATMENT

Achromobacter xylosoxidans Infections

(Table 163-2) Treatment is based on in vitro susceptibility testing of all clinically relevant isolates; multidrug resistance is common.

TABLE 163-2 Treatment Options for Other Selected Gram-Negative Bacteria[a]

ORGANISM	TREATMENT OPTIONS
Achromobacter xylosoxidans	Carbapenems, piperacillin-tazobactam, ceftazidime, trimethoprim-sulfamethoxazole
Aeromonas spp.	Fluoroquinolones, third- and fourth-generation cephalosporins, carbapenems, minocycline, aminoglycosides
Elizabethkingia/Chryseobacterium spp.	Fluoroquinolones, minocycline, trimethoprim-sulfamethoxazole, piperacillin/tazobactam
Rhizobium radiobacter	Fluoroquinolones, third- and fourth-generation cephalosporins, carbapenems
Shewanella spp.	Fluoroquinolones, third- and fourth-generation cephalosporins, β-lactam/β-lactamase inhibitors, carbapenems, aminoglycosides
Chromobacterium violaceum	Carbapenems, fluoroquinolones, trimethoprim-sulfamethoxazole

[a]Treatment should be based on in vitro susceptibility testing; multidrug resistance is common among these organisms.

Carbapenems, piperacillin-tazobactam, ceftazidime, trimethoprim-sulfamethoxazole, minocycline, and colistin are typically the most active agents.

***Aeromonas* Species** *Aeromonas* is a facultative anaerobic gram-negative bacterium. *Aeromonas* infections are most often caused by *A. hydrophila*, *A. caviae*, *A. veronii*, and *A. dhakensis*. *Aeromonas* proliferates in potable water, seawater, freshwater, raw seafood, and soil. It remains controversial whether *Aeromonas* is a cause of bacterial gastroenteritis; asymptomatic colonization of the intestinal tract with *Aeromonas* occurs frequently, and no clonally related diarrheal outbreak has been documented. However, rare cases of hemolytic-uremic syndrome following bloody diarrhea have been shown to be secondary to the presence of *Aeromonas*.

Aeromonas causes health care–associated sepsis and bacteremia in infants with multiple medical problems and in immunocompromised hosts, particularly those with cancer or hepatobiliary disease, including cirrhosis. *A. caviae* is associated with health care–related bacteremia. Community-acquired infections include bacteremia, spontaneous bacterial peritonitis, biliary tract infections, and skin and soft tissue infections. Severe soft tissue infections such as necrotizing fasciitis are more common in Taiwan than in Western countries; *Aeromonas* was the most common pathogen associated with skin and soft tissue infections after the tsunami in Thailand. Along with other gram-negative organisms such as *Shewanella* and *Chromobacterium*, *Aeromonas* infections are associated with floods and other hydrologic disasters. *Aeromonas* infection and sepsis can occur in patients with trauma (including severe trauma with myonecrosis), patients with seawater-contaminated wounds, and burn patients exposed to the organism by environmental (freshwater or soil) wound contamination. Reported mortality rates range from 25% among immunocompromised adults with sepsis to >90% among patients with myonecrosis. Patients with *A. dhakensis* bacteremia have higher 14-day mortality rates than do those whose bacteremia is attributable to other species. *Aeromonas* can produce ecthyma gangrenosum (hemorrhagic vesicles surrounded by a rim of erythema with central necrosis and ulceration; **see Fig. A1-34**) resembling the lesions seen in *Pseudomonas aeruginosa* infection. This organism causes nosocomial infections related to catheters, surgical incisions, or use of leeches. Other manifestations include meningitis, peritonitis, pneumonia, and ocular infections.

TREATMENT

Aeromonas Infections

(Table 163-2) *Aeromonas* species are generally susceptible to fluoroquinolones (e.g., ciprofloxacin at a dosage of 500 mg every 12 h PO or 400 mg every 12 h IV), third- and fourth-generation cephalosporins, carbapenems, minocycline, and aminoglycosides, but resistance to all those agents has been described. Because *Aeromonas* can produce various β-lactamases, including carbapenemases, susceptibility testing must be used to guide therapy. Antibiotic prophylaxis (e.g., with ciprofloxacin) is indicated when medicinal leeches are used.

***Elizabethkingia/Chryseobacterium* Species** *Elizabethkingia meningoseptica* (formerly *Chryseobacterium meningosepticum*) is a nonfastidious aerobic nonfermentative gram-negative bacillus, found in soil, water, and hospital environments. It is an important cause of nosocomial infections, including outbreaks due to contaminated fluids (e.g., contaminated sinks, disinfectants, and aerosolized antibiotics) and sporadic infections due to indwelling devices, feeding tubes, and other fluid-associated apparatuses. Outbreaks due to this organism have persisted until extensive cleaning of environmental surfaces and equipment has been performed. Nosocomial *E. meningoseptica* infection usually involves preterm neonates, patients with underlying immunosuppression (e.g., related to malignancy or diabetes), or patients exposed to antibiotics in intensive care. *E. meningoseptica* has been reported to cause meningitis (primarily in neonates), pneumonia, sepsis, endocarditis, bacteremia, eye infections, and soft tissue infections. Other species of *Elizabethkingia* are emerging, identified using MALDI-TOF and 16s rRNA sequencing. In a report of 86 clinical isolates, only 12 (19.8%) of the isolates were *E. meningoseptica;* the majority were *E. anophelis* (51 isolates; 59.3%). More than three-quarters of the isolates were from the lower respiratory tract, and 9.3% were from blood cultures. In a U.S. study of 11 patients with *E. anophelis* infection, all had bloodstream infections. In that series, the patients had underlying comorbidities and recent health care exposure; there was a mortality rate of 18.2%. *E. anophelis* has also been associated with sepsis and meningitis in neonates.

Chryseobacterium indologenes has caused bacteremia, sepsis, peritonitis, meningitis, and pneumonia, typically in immunocompromised patients with indwelling devices. Mortality rates have been as high as 50% in some reports; it is unclear whether a poor prognosis is related to underlying comorbidities or to the multidrug-resistant phenotype of the organism.

TREATMENT

Elizabethkingia/Chryseobacterium Infections

(Table 163-2) These organisms are often susceptible to fluoroquinolones, minocycline, and trimethoprim/sulfamethoxazole. They may be susceptible to β-lactam/β-lactamase inhibitor combinations such as piperacillin/tazobactam but are resistant to carbapenems. Multidrug-resistant isolates are increasing and can possess extended-spectrum β-lactamases and metallo-β-lactamases. Combination therapy may be needed for successful treatment. Susceptibility testing should be performed to guide the choice of optimal agents.

MISCELLANEOUS ORGANISMS

Rhizobium (formerly *Agrobacterium) radiobacter* has usually been associated with infection in the presence of medical devices, including intravascular catheter–related infections, prosthetic-joint and prosthetic-valve infections, peritonitis caused by dialysis catheters, and endophthalmitis after cataract surgery. Most *R. radiobacter* infections occur in immunocompromised hosts, especially individuals with malignancy or HIV infection. Elderly patients with septic shock, watery diarrhea, and acute renal failure also have been reported. Strains are usually susceptible to fluoroquinolones, third- and fourth-generation cephalosporins, and carbapenems (Table 163-2).

Shewanella species are ubiquitous nonfermentative gram-negative organisms found in seawater and marine environments. Human disease is caused primarily by *S. putrefaciens* and *S. algae*; *S. algae* may be the more virulent species. Most infections involve skin and soft tissue, ranging from impetigo to necrotizing fasciitis. Patients are exposed to the organism through contact of bites, open wounds, or devitalized tissue with seawater, marine animals, or through ingestion of seawater or of raw or undercooked seafood, especially shellfish. Nosocomial outbreaks have been reported. *Shewanella* species also cause chronic ulcers of the lower extremities, bacteremia, osteomyelitis, biliary tract infections, pneumonia, sepsis, and potentially chronic otitis media. A fulminant course is associated with cirrhosis, hemochromatosis, diabetes mellitus, malignancy, or other severe underlying conditions. These organisms are often susceptible to fluoroquinolones, third- and fourth-generation cephalosporins, β-lactam/β-lactamase inhibitors, carbapenems, and aminoglycosides (Table 163-2), but multidrug resistance is increasing.

Chromobacterium violaceum is a facultative anaerobic organism found in soil and water in tropical or subtropical regions. After exposure, it can cause rare but serious—often fatal—skin and soft tissue infections of limbs, although several recent reports suggest a more benign course with lower mortality. Life-threatening infections with severe sepsis and metastatic abscesses occur most often in patients with underlying illness, particularly in children with defective neutrophil function (e.g., those with chronic granulomatous disease). *C. violaceum* is frequently resistant to multiple drugs; carbapenems are most often used empirically. Fluoroquinolones and trimethoprim-sulfamethoxazole also can be active (Table 163-2).

Organisms that can cause nosocomial infections of medical devices in compromised hosts, such as central venous catheters, include *Ochrobactrum anthropi*, *Pseudomonas* (formerly *Flavimonas*) *oryzihabitans*, and *Sphingobacterium*. *Alcaligenes faecalis* has been associated with hospital-acquired infections, such as bloodstream infections, due to contaminated hemodialysis and IV fluids or respirators. *Sphingomonas koreensis* was associated with a small cluster of nosocomial cases at one hospital and was traced to a reservoir in the plumbing system. *Ralstonia* species also can contaminate water supplies, including hospital water systems. Cases of bacteremia, osteomyelitis, pneumonia, and meningitis have been described. *Pandoraea commovens* primarily infects patients with cystic fibrosis. However, an outbreak described in Germany resulted in critical illness. Patients were more likely to be receiving mechanical ventilation and have had prior surgery or antimicrobial exposure. *Sphingomonas paucimobilis,* a rare cause of infection in both healthy and immunocompromised patients, can cause bloodstream infections, respiratory distress, and sepsis. It has a predilection for bone and soft tissue infection, osteomyelitis, and septic arthritis. Other organisms can cause rare human infections such as *Weeksella* species; *Bergeyella* species; various Centers for Disease Control and Prevention (CDC) groups; and *Oligella urethralis.* The reader is advised to consult subspecialty texts and references for further guidance on these and other organisms.

FURTHER READING

Bläckberg A et al: Infective endocarditis caused by HACEK group bacteria—a registry-based comparative study. Eur J Clin Microbiol 40:1919, 2021.

Dernoncourt A et al: Prognostic factors of *Pasteurella* infections: A single-center retrospective cohort study over a 14-year period (2005–2018). Int J Infect Dis 116:197, 2022.

Spencer HK et al: An overview of the treatment of less common non-lactose-fermenting gram-negative bacteria. Pharmacotherapy 40:936, 2020.

164

Legionella Infections

Steven A. Pergam, Thomas R. Hawn

Bacteria of *Legionella* species cause two primary human diseases: *Legionella* pneumonia (often referred to as Legionnaires' disease) and Pontiac fever; collectively, these diseases are referred to as *legionellosis*. Legionnaires' disease was first described in 1976 in an outbreak among members of the American Legion participating in a conference at a hotel in Philadelphia, Pennsylvania. Since their original description, *Legionella*-related infections have increased in frequency throughout the world as techniques to diagnose them have improved, clinical awareness has increased, cities have grown, and water systems have both aged and become more complex. Most cases of legionellosis are linked to waterborne exposures. These infections can be either sporadic or due to common-source community or nosocomial exposures. Outbreaks of legionellosis are well described. After exposure, legionellosis occurs primarily among persons with risk factors for disease, including older adults and those with primary organ dysfunction, immunocompromise, or other chronic illnesses. Clinical awareness is important, as the similarity of signs and symptoms of legionellosis to those of other respiratory illnesses can lead to delayed treatment. Despite appropriate therapy, *Legionella* pneumonia is associated with significant morbidity and mortality.

PATHOGEN AND PATHOGENICITY

Legionellae are aerobic gram-negative bacteria that are ubiquitous in aquatic environments, damp soil, and compost. Of the more than 60 *Legionella* species, approximately half have been documented to lead to clinical disease, but most clinical disease is driven by *Legionella pneumophila*, primarily serotype 1. The primary habitats for growth and replication of *Legionella* are amoebae and other free-living protozoa, in which these bacterial species can thrive intracellularly; humans are accidental hosts. Legionellae are reliant on host-derived amino acids and nutrients for intracellular replication. The organisms have a biphasic life cycle: a replicative phase in nutrient-rich conditions (e.g., in their protozoal hosts) and a noninfective transmissive phase under scarcity of resources. Therefore, they can persist in complex biofilms in both natural and engineered water systems (e.g., premise plumbing—a building's hot and cold water piping systems) and are phagocytized by waterborne protozoa. In premise plumbing systems, where temperature and nutrients support the protozoal hosts of legionellae, the bacteria can replicate to concentrations sufficient to cause human infection.

After exposure to *Legionella* through inhalation or aspiration of small aerosol particles, the organisms attach to immune cells and are phagocytized. After phagocytosis, they can evade intracellular defenses and replicate in human alveolar macrophages and monocytes. Pathogenic *Legionella* species have numerous virulence systems that they use to evade the human immune system, including the development of *Legionella*-containing vacuoles within immune cells, downregulation of cytokine receptors, inhibition of host protein synthesis, and avoidance of lysosomal degradation. Despite their ability to replicate and persist in an intracellular environment, innate immune components that target intracellular pathogens—specifically, pattern recognition receptors, including Toll-like receptors and nucleotide-binding oligomerization domain–like receptors—activate immune responses. Adaptive CD4 and CD8 cytotoxic T-cell involvement and these innate immune responses eventually lead to the production of interferon γ and tumor necrosis factor, the promotion of neutrophil recruitment into the lung, and other proinflammatory responses. This cascade can be beneficial and result in clearance of the pathogen. However, these inflammatory responses can also cause immunopathology and adverse outcomes. *L. pneumophila* is more cytopathogenic than most non-*pneumophila Legionella* species, a characteristic that may be partially responsible for its association with severe disease.

EPIDEMIOLOGY

Legionella species are responsible for >50% of all waterborne outbreaks and >10% of disease related to drinking water in the United States. A National Academies of Sciences, Engineering, and Medicine report estimates that 50,000–70,000 Americans develop Legionnaires' disease per year. Incidence rates of legionellosis in the United States are reportedly 2–3 cases per 100,000 persons, but higher rates have been reported in other parts of the world. Numerous global epidemiologic studies assessing legionellosis have shown an increasing prevalence over the past few decades; this increase has been hypothesized to be due to a variety of causes, including an aging population, improved diagnostics, global temperature changes, and an aging water infrastructure. Legionellosis is associated with substantial health care costs.

Legionella species are found throughout the world, but most epidemiologic data focus on legionellosis in large metropolitan areas in Australia/New Zealand, Europe, and North America. Rates of infection in other parts of the world are unknown, as surveillance systems and laboratory testing are less readily available in large portions of Africa and Asia. More than 80% of cases of Legionnaires' disease are linked to *L. pneumophila*—in particular to serotype 1, which is the most frequently isolated *Legionella* pathogen. Although *L. pneumophila* predominates as a cause of disease, species predilection varies regionally. In Australia and New Zealand, for example, the rate of disease due to *Legionella longbeachae* approaches or exceeds that for *L. pneumophila*.

As previously mentioned, most reported cases are due to *L. pneumophila* serotype 1—a reflection of its pathogenicity. However, this predominance is also due to the frequency and ease of use of urinary antigen testing that targets this pathogen and allows more effective diagnosis in the community. It is unclear how large a role non-*pneumophila* species and non–serotype 1 *L. pneumophila* play in disease. However, in studies in Europe, where respiratory cultures are more frequently collected, nearly 10% of Legionnaires' disease patients were infected with species other than *L. pneumophila*. In the United States, nearly 10% of culture-confirmed cases are due to non–serogroup 1 *L. pneumophila*. Immunosuppressed patients, such as cancer patients and transplant recipients, may be more likely to develop pneumonia caused by non-*pneumophila* species such as *Legionella micdadei, Legionella bozemanii,* and *L. longbeachae.*

Despite increases in cases in the United States **(Fig. 164-1)** and throughout the world, incident cases are still thought to be underreported. Many cohort studies of community-acquired pneumonia do not require routine testing for *Legionella* or assess only for *L. pneumophila* serotype 1 (by urinary antigen testing) and therefore may underestimate true prevalence. For example, a large administrative database of studies shows that, of patients with clinically proven community-acquired pneumonia, only 26% underwent *Legionella*-specific testing; even patients with documented risk factors for legionellosis are not always tested for *Legionella.* In studies that routinely assess for legionellosis, the prevalence of *Legionella* pneumonia ranges between 2 and 10% of all community-acquired pneumonia cases. In addition, extrapulmonary presentations and Pontiac fever are less likely to be identified or to result in presentation for health care, and this trend leads to further underestimation of the true burden of legionellosis.

Seasonality and Climate Geoclimatic changes, storms, and seasonality are thought to be important components of *Legionella*'s epidemiology. The incidence of *Legionella* disease increases in the summer and fall—specifically, in warmer weather and with increased rain and humidity. Studies that screen all respiratory samples for *Legionella* find that legionellosis is indeed diagnosed most frequently in the United States during warmer summer/fall months and periods of greater humidity. Furthermore, seasonal storms, which may disrupt water pipes or cause increased flooding, can result in contamination of water systems with soil and lead to *Legionella* exposures. There is concern that, with ongoing climate shifts and rising global temperatures, cases of legionellosis may continue to increase.

Community and Health Care–Associated Outbreaks Small and large clusters and point-source outbreaks of *Legionella* cases lead to public health investigations, but these situations account for only ~5–10% of all *Legionella* cases yearly. Outbreaks occur when two or more people become ill after shared exposures in a community. In health care systems, a single proven case should trigger a *Legionella* investigation. The Centers for Disease Control and Prevention (CDC) recommends an outbreak investigation if a single patient with *Legionella* is identified who did not leave the facility/campus for the 10 days prior to illness onset. Additionally, an outbreak investigation within a health care system is warranted if there are at least two possible *Legionella* patients who spent any time in the hospital/long-term care facility within 12 months of each other (see "Clinical Presentations" below).

Most common outbreaks are linked to water sources dispersing aerosol droplets that increase the area of particle spread (e.g., cooling towers or fountains) or to large building structural water systems that cause multiple prolonged exposures (e.g., those in hospitals, hotels, or apartments). The most commonly reported sources include not only cooling towers and fountains but also water misters; centralized heating, ventilation, and air-conditioning systems; hot tubs/spas; pools; ice machines; and showerheads and sinks in large premise plumbing structures **(Fig. 164-2)**. When used as primary sources of water, groundwater and wells have also been associated with *Legionella* exposures. The majority of exposures are related to engineered hot-water systems, which are often maintained at temperatures that limit scalding but are ideally suited for *Legionella*'s growth. *Legionella* can also be found in cold water, particularly in warmer summer months, as a consequence of the warming water temperature; engineering issues (e.g., heating lamps in fountains); or unexpected breaks in plumbing systems (e.g., malfunctioning thermostatic mixing valves), which can lead to hot-water contamination of cold-water systems.

Buildings with inconsistent use patterns, such as hotels in seasonal travel destinations, can be linked to outbreaks of legionellosis, as water stagnation leads to low chlorine/disinfectant levels and organism proliferation can reach high enough levels to cause disease. Outbreaks have also been linked to cruise ships and boats. Following stay-at-home orders during the SARS-CoV-2 pandemic, when buildings (e.g., hotels) were reopened, limited water movement and stagnation led to increases in cases of legionellosis. Modern buildings with water-saving devices, which aim to limit water and energy use, may increase the risk of legionellosis, as they can decrease water temperatures and limit water flow.

Outbreaks in health care and long-term care facilities are identified more frequently than outbreaks in other facilities, as they often bring together at-risk patients, prolonged water exposures, accessible testing, elevated awareness, and regulations that help ensure that cases are more easily linked to common sources. The outbreak examples listed in **Table 164-1** demonstrate the wide variety of common sources and the number of cases associated with such factors. As previously mentioned, most large outbreaks involve cooling towers, which can spread aerosol droplets over a wide area. The largest outbreak reported to date involved a cooling tower in Spain that was linked to 449 documented cases of Legionnaires' disease. Outbreaks are increasingly discussed in the media, such as outbreaks linked to cooling towers in the Bronx neighborhood of New York City, a large hotel outbreak in Atlanta, and the outbreaks associated with the Flint, Michigan, water crisis that

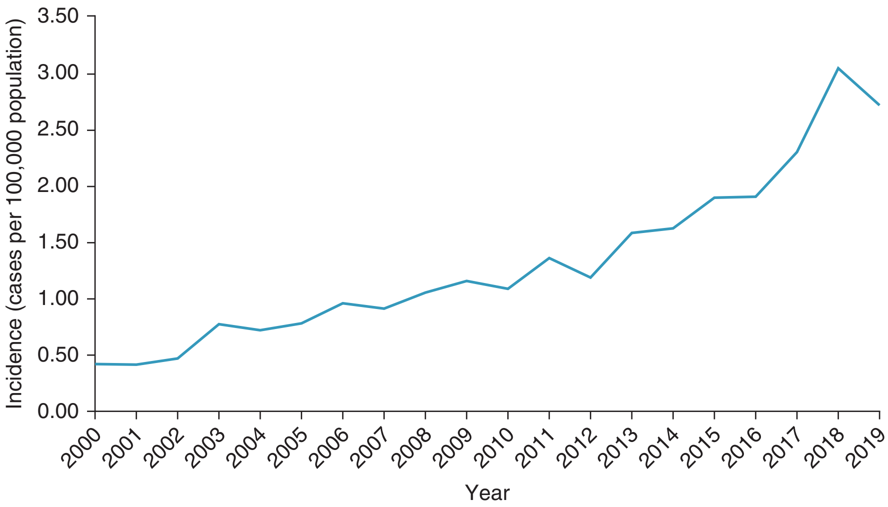

FIGURE 164-1 Increasing *Legionella* disease incidence in the United States over the past two decades (2000–2019). *(From https://www.cdc.gov/legionella/about/history.html.)*

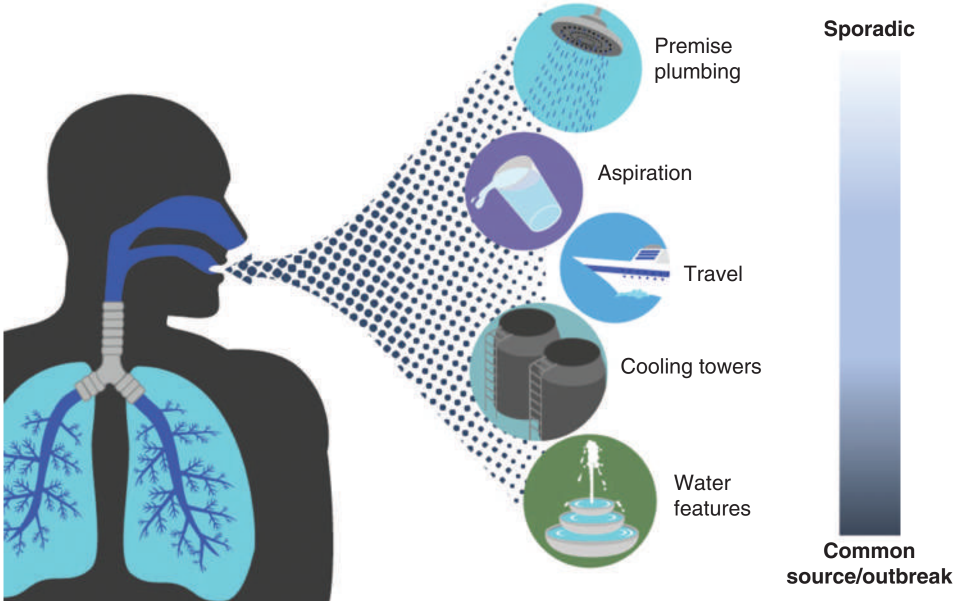

FIGURE 164-2 Sources of waterborne *Legionella* exposures and spectrum of presentation. The spectrum of sporadic to common-source outbreaks is a continuum. For example, premise plumbing in a large office building can lead to a large outbreak, and travel exposures can be related to large outbreaks. Most sporadic cases have no documented source of exposure, while outbreaks often involve mechanisms that spread water aerosol droplets over long distances (e.g., cooling towers), with a consequent ability to infect more individuals. *(Reproduced with permission from Kyoko Kurosawa.)*

led to numerous deaths. It is not uncommon for lawsuits to be initiated when deaths are linked to outbreaks.

Sporadic Cases The vast majority of cases of Legionnaires' disease occur sporadically in the community, manifesting as community-acquired pneumonia. Identification of the transmission source is more difficult in community-acquired cases than in nosocomial cases, despite reporting and review by local public health jurisdictions. In nearly 90% of all cases of legionellosis, a source of exposure is never identified. Since the spectrum of water exposures in the community is so broad and incubation periods can be long, identifying individual exposures often is not possible. Transient exposures to common sources, travel-related exposures, and exposures to less commonly linked sources (e.g., potting soil and compost) may also be hard to identify. Furthermore, studies of domestic hot water have demonstrated that 5–30% of households may have *Legionella* species detected, but the role that households play in clinical legionellosis is hard to determine, as home water testing is infrequently a part of usual contact investigations. Because of underdiagnosis, it is likely that diagnosed sporadic community-acquired cases represent only patients who are ill enough to present to health care for evaluation.

Risk Factors A number of epidemiologic and demographic risk factors are associated with legionellosis. Older age is a risk factor; most studies suggest that risk begins to increase at an age of ~40 years. Furthermore, elderly patients are at the highest risk for major complications. Males are at approximately three times greater risk for *Legionella* disease than are females in most large epidemiologic studies. Children are thought to be less likely to develop severe infections. However, since routine testing is less common among children, cases may be underreported.

Smoking has been strongly linked to legionellosis. Inhalation of smoke leads to anatomic changes in the airway epithelium, impairs neutrophil and monocyte phagocytosis, and has negative effects on airway ciliary clearance—all of which can increase the risk of pneumonia. Studies have shown that cigarette smoking is a dose-dependent risk factor. Smoking cannabis has also been associated with increased risk. Risk and severity of illness are further associated with smoking-related pulmonary diseases such as chronic obstructive pulmonary disease or emphysema, which in turn lead to increased risk for complications. Patients with other organ dysfunction/failure, such as those with renal disease (including those on dialysis), hepatic disease, nonsmoking pulmonary disease, and cardiac disease, are at increased risk for legionellosis, although it is unclear whether these factors are related to disease severity or to greater awareness and consequent recognition by health care providers.

Immunosuppressed patients are at increased risk for legionellosis and *Legionella*-related complications. Patients undergoing treatment for cancer (including recipients of hematopoietic cell transplantation) and solid organ transplant recipients are at high risk for legionellosis due to immunosuppression as well as disease- and treatment-related comorbidities. Use of prednisone and other glucocorticoids is strongly associated with legionellosis; however, in light of the heterogeneity of immunosuppressive agents and their use, it remains unclear whether

TABLE 164-1 Examples of *Legionella* Common Source Outbreaks, Indicating the Wide Variety of Sources and Cases[a]

SITE	YEAR	ORGANISM[a]	REPORTED SOURCE(S)	CASES
Hotel[b]	2012	*L. pneumophila* serotype 1	Potable water, fountain, spa	85 (29 suspect)
Hospital[c]	2012	*L. pneumophila*	Potable water	22 cases
Community[d]	2014	*L. pneumophila* serotype 1	Cooling tower	334 cases
Hospital/community[e]	2014-15	*L. pneumophila*	Potable water, household, cooling towers	86 cases
Long-term care facility[f]	2015	*L. pneumophila*	Potable water	74 cases
Community[g]	2015	*L. pneumophila*	Hotel cooling towers	128 cases
Hospital[h]	2018	*L. pneumophila* serotype 1	Potable water, showers	13 cases
Hotel[i]	2019	*L. pneumophila*	Fountain	13 LP (66 suspect)
Community[j]	2019	*L. pnuemophila*[k]	Hot-tub display	141 cases
Community[l]	2022	*L. pneumophila* serotype 1	Cooling tower	17 cases

[a]Large community outbreaks most commonly linked to cooling towers. Cases not noted to be a specific serotype, were not reported. [b]Smith SS et al: Open Forum Infect Dis 2:ofv164, 2015. [c]Department of Veteran's Affairs Inspector General. *https://www.va.gov/oig/pubs/VAOIG-13-00994-180.pdf.* [d]Shivaji T et al: Eurosuveillance 19:20991, 2014. [e]Smith AF et al: Environ Health Perspect 127:127001, 2019. [f]State of Illinois, Auditor General. *https://auditor.illinois.gov/Audit-Reports/Performance-Special-Multi/Performance-Audits/2019_Releases/19-Quincy-Legionnaires-Disease-Perf-Digest.pdf.* [g]New York City Department of Health and Mental Hygiene. *https://www1.nyc.gov/assets/doh/downloads/pdf/han/alert/legionella-in-bronx-source-identified.pdf.* [h]Kessler MA et al: Am J Infect Control. 49:1014, 2021. [i]Brown E: New York Times. 2019. *https://www.nytimes.com/2019/08/16/us/legionnaires-disease-atlanta-hotel-reopen.html.* [j]North Carolina Department of Health. *https://epi.dph.ncdhhs.gov/cd/legionellosis/MSFOutbreakReport_FINAL.pdf.* [k]Non–serotype 1. [l]Grossman NV et al: Morbid Mortal Wkly Rep 72:1315, 2023. *https://www.cdc.gov/mmwr/volumes/72/wr/mm7249a1.htm.*

most other single agents are as strongly associated with the disease. Combination immunosuppressive regimens increase risk. Patients treated with these regimens are more likely to develop non-*pneumophila* legionellosis and non–serotype 1 *L. pneumophila* infections that may be missed by routine urinary antigen testing. Patients with autoimmune diseases receiving tumor necrosis factor inhibitors, either with or without concomitant glucocorticoid use, are also at increased risk for legionellosis. Furthermore, studies suggest a possible association of legionellosis with genetic polymorphisms in components of the innate immune system that are important in recognizing and responding to intracellular pathogens (e.g., Toll-like receptors and interferon genes). There has also been transmission to lung transplant recipients from donor lungs.

Transmission The *Legionella* species involved in human disease are usually waterborne pathogens. However, disease development requires sufficient levels of the organism at the exposure site, the formation of small particles that can be inhaled or aspirated into pulmonary alveoli, and an at-risk host. *Legionella*-containing aerosol particles must be <10 μm in diameter for deposition into the alveoli. The infective dose during exposures is unknown but likely depends on the host: disease development in at-risk individuals may require a more limited exposure. Strain virulence is also thought to be important in disease development: *L. pneumophila* serotype 1 is more apt to lead to outbreaks and disease than, for example, *Legionella anisa*, which has only rarely been associated with disease in high-risk patients. Because of the necessity for these various factors, estimated attack rates during an exposure are only ~5% for pneumonic presentations. Attack rates for Pontiac fever (see below) are thought to be higher—up to 90% among those exposed.

Most exposures occur through the inhalation of contaminated aerosols from mists, sprays, or other mechanisms that produce small water droplets that can be inhaled into the distal alveoli. In homes, the most common sites of exposure are showerheads and sinks, which are especially apt to produce particles small enough for inhalation. The role played by aspiration or microaspiration in exposures is more controversial but is hypothesized to be a secondary route for developing pneumonia. Although human-to-human transmission is not a common pathway, a single presumptive case has been reported. After exposure, *L. pneumophila* has an incubation period of ~2–10 days; this period has been reported to be longer in immunosuppressed hosts. In contrast, symptoms of Pontiac fever occur within 24–48 h after exposure.

CLINICAL PRESENTATIONS

***Legionella* Pneumonia** *Legionella* pneumonia is the most common manifestation of legionellosis. In clinical practice, *Legionella* pneumonia is often referred to by clinicians as an "atypical pneumonia" (i.e., pneumonia that lacks the classic signs and symptoms of bronchopneumonia). Other bacterial pathogens, such as *Chlamydia pneumoniae* and *Mycoplasma pneumoniae*, are also considered as etiologic agents of atypical pneumonia. Initial symptoms of *Legionella* pneumonia are nonspecific and include fever, myalgias, headache, shortness of breath, and either a dry or a productive cough **(Table 164-2)**. Patients with pneumonia who present with neurologic or gastrointestinal symptoms such as anorexia, nausea, or vomiting may be more likely than others to have legionellosis. Immunosuppressed patients may present without typical symptoms such as fever. Patients who have recently traveled, who present during a known or possible *Legionella* outbreak, or who develop pneumonia while hospitalized should undergo testing for legionellosis. Patients with severe pneumonia presentations, including acute respiratory failure, and those with pneumonia and sepsis-like presentations should undergo testing for *Legionella* as per current community-acquired pneumonia guidelines.

TABLE 164-2 Clinical and Epidemiologic Features of *Legionella* Pneumonia (Legionnaires' Disease) and Pontiac Fever

FEATURE	*LEGIONELLA* PNEUMONIA	PONTIAC FEVER
Incubation period	2–10 days[a]	24–72 h
Pathogenesis	*Legionella* infection	*Legionella* infection or exposure
Common symptoms	Abdominal or chest pain Anorexia Cough, sputum production Confusion[b] Diarrhea[b] Fatigue Fever/chills Headache Myalgias Nausea/vomiting[b] Shortness of breath	Cough Diarrhea Fatigue Fever/chills Headache Myalgias Nausea/vomiting Vertigo
Risk factors	Age >40 years Male Smoker Immunosuppressed host Neurologic disease Chronic lung disease Organ dysfunction/ chronic illness	Factors associated with increased exposure
Attack rate among exposed individuals	~5%[c]	~90%
Hospitalization rate	>90%	<1%
ICU admission rate	30–50%	Extremely low
Treatment	Antibiotics (macrolide or fluoroquinolone)	Supportive care
Case-fatality rate[d]	10%	Extremely low

[a] Incubation period in immunosuppressed hosts may be longer than 14 days. [b] This symptom is strongly associated with *Legionella* pneumonia. [c] Attack rates are highly dependent on method of exposure, level of the pathogen in source water, and host's level of risk. [d] Case-fatality rates are much higher among immunosuppressed patients and those with severe underlying lung disease, ranging from 30 to 50%.

Abbreviation: ICU, intensive care unit.

Source: Modified from *https://www.cdc.gov/legionella/clinicians/clinical-features.html.*

Patients with *Legionella* pneumonia classically present with rales, rhonchi, and—when consolidation is present—egophony and dullness to percussion. Not all patients, particularly immunosuppressed patients, present with pulmonary findings on clinical examination. Initial laboratory findings in patients with *Legionella* pneumonia include leukocytosis or leukopenia, thrombocytopenia, and elevated liver enzyme levels; hyponatremia and/or renal dysfunction are frequent. Levels of nonspecific laboratory markers of inflammation, such as C-reactive protein, can also be elevated; however, procalcitonin levels may not be as useful as a diagnostic tool. Although clinical symptoms and laboratory findings tend to be nonspecific, a number of clinical prediction tools, such as the Winthrop-University Hospital Criteria and the *Legionella* Score, have been developed to assist with the diagnosis of *Legionella* pneumonia. These scoring systems may be more useful for their negative than for their positive predictive value.

An important subset of *Legionella* pneumonia cases are those that are linked to health care systems—i.e., nosocomial cases. Although cases of hospital-acquired legionellosis are rare, their identification is necessary as they may be harbingers of contamination of water systems, devices, and/or potable water sources. Because of the rarity of nosocomial cases, outbreaks have sometimes occurred over years before the source is identified within the health care system. In this regard, the CDC offers the following definitions: (1) A *presumptive* health care–associated case of Legionnaires' disease is one developing in a patient with *Legionella* pneumonia after ≥10 days of continuous stay at a health care facility during the 14 days before onset of symptoms. (2) A *possible* case is one that develops in a patient with *Legionella* pneumonia who has spent a portion of the 14 days before symptom onset in one or more health care facilities but not enough time to meet the criteria for a presumptive case. To ensure that singular cases lead to more system-wide evaluations, the CDC also recommends an investigation if a health care system detects one or more cases of presumptive health care–associated Legionnaires' disease at any time or two or more possible cases within 12 months of one another.

PONTIAC FEVER

Pontiac fever is an influenza-like illness whose primary symptoms are fever, headache, myalgias, chills, vertigo, nausea, vomiting, and diarrhea (Table 164-2). Compared with *Legionella* pneumonia, Pontiac fever is a milder, self-limited illness that is defined by the absence of pneumonia. Although studies have shown that Pontiac fever is associated with exposure to higher counts of colony-forming units in water sources, the role of the pathogen in the disease is not clear. Symptoms usually develop 24–48 h after exposure and can last for 2–5 days. Since many other illnesses resemble Pontiac fever, the diagnosis usually relies on the recognition of typical clinical features during an outbreak situation; therefore, sporadic cases are likely to be missed even when patients present for health care. Studies documenting specific *Legionella* species as the cause of Pontiac fever clusters find that most are due to *L. pneumophila* exposure; however, non-*pneumophila* species such as *L. anisa* have also been associated with this presentation.

Extrapulmonary Disease A number of rare presentations for legionellosis have been described. Skin and soft tissue infections that resemble cellulitis, including cases due to tap water contamination of postsurgical wounds, have been reported. Endocarditis, primarily culture-negative prosthetic valve endocarditis, and myocarditis and pericarditis have also been reported. Rarely, *Legionella* species have been associated with septic arthritis and sinusitis.

DIAGNOSIS

The diagnosis of legionellosis on the basis of clinical findings alone is difficult. Additional workup is needed to make a definitive diagnosis, even when cases are potentially linked to a possible outbreak. To make a diagnosis, laboratory confirmation is needed, and invasive procedures may be required—e.g., bronchoscopy, particularly for patients whose results on urinary antigen testing are negative and who cannot produce sufficient sputum for testing or for patients with severe disease requiring intensive care unit (ICU) admission. As current treatment guidelines for community-acquired pneumonia recommend empirical coverage that includes antibiotics active against *Legionella* species, diagnostic testing is not routine even among persons who meet the criteria for *Legionella*-specific testing. Furthermore, not all currently available diagnostic laboratory assays, including urinary antigen testing, are accessible or rapidly available in primary care clinics, urgent care facilities, and emergency rooms where patients may present with their initial symptoms.

Radiologic Findings On chest radiography, *Legionella* pneumonia presents as focal infiltrates or consolidations, most frequently in the lower lobes, that are indistinguishable from those due to other causes of pneumonia **(Fig. 164-3)**. On computed tomography (CT), air-space disease in one or more lobes is often with associated ground-glass opacities **(Fig. 164-4)**; pleural effusions and lymphadenopathy are less frequently seen. In immunocompromised patients, *Legionella* can present with similar lower-lobe consolidations or atypically as pulmonary nodules—with or without cavitation—that mimic fungal infections **(Fig. 164-5)** or even as lung abscesses. Progression during early therapy is not uncommon in immunosuppressed patients.

Laboratory Diagnostics • CULTURE Cultures—of sputum, bronchoalveolar lavage fluid, lung tissue, or extrapulmonary sites—are the gold standard for diagnosis of *Legionella* pneumonia because they are critical for epidemiologic investigations. *Legionella* species require special nutrients, such as cysteine, for growth and therefore require specialized media, such as buffered charcoal yeast extract (BCYE) agar. Legionellae grow slowly, usually over 3–5 days, with non-*pneumophila* species often requiring longer incubation times. Once growth is seen, *Legionella* can be stained with standard Gram stain, and colonies often fluoresce blue or white under ultraviolet light. *L. micdadei* is the only *Legionella* species that is also modified-acid-fast positive. Sensitivity varies with the sample but is highest among lower respiratory tract samples. At some referral centers, lower-tract samples from high-risk immunosuppressed patient populations are routinely sent for culture. Unfortunately, because of current community-acquired pneumonia guidelines, patients are often treated empirically, and many either never have samples sent for *Legionella*-specific cultures or have such samples collected only after antibiotic administration, which decreases sensitivity. Respiratory cultures from patients with legionellosis are crucial during outbreak investigations, as clinical and environmental cultures can be compared molecularly to help identify common-source outbreaks; cultures are also used for serotyping of *L. pneumophila*.

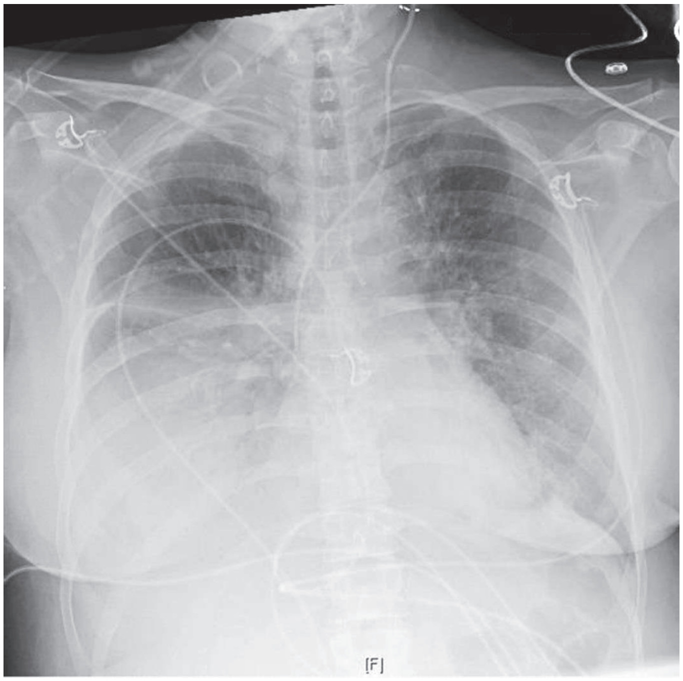

FIGURE 164-3 Chest x-ray of a patient with *Legionella* pneumonia and right-lower-lobe consolidation. A 64-year-old woman presented with fever, dry cough, and shortness of breath 7 days after returning from international travel. *Legionella* urinary antigen testing was positive for *L. pneumophila* serotype 1.

URINARY ANTIGEN TESTING *Legionella* urinary antigen tests are widely available at many hospitals and commercial laboratories and are characterized by ease of use, simple specimen collection, rapid turnaround time, high sensitivity, and the ability to detect the most prevalent *Legionella* species associated with clinical disease—*L. pneumophila* serotype 1. Urinary antigen testing has limitations, however: it detects only *L. pneumophila* serotype 1 and gives false-negative results in most cases caused by clinically important non–serotype 1 *L. pneumophila* and non-*pneumophila* species. Sensitivity for *L. pneumophila* serotype 1 is ~70% for most assays, but specificity is very high. The urinary antigen test can be negative very early in the disease and can remain positive for months after an infection, particularly in immunosuppressed patient populations; it cannot be used for patients who are anuric. Urinary antigen testing is not recommended for routine use in screening for exposures among asymptomatic patients in outbreak investigations.

SEROLOGY Acute- and convalescent-phase titers of antibody to *Legionella* have limited sensitivity in diagnosing acute Legionnaires' disease but can be useful during outbreak investigations. A case is *confirmed* by documenting a fourfold or greater rise in titer of specific serum antibody to *L. pneumophila* serogroup 1. A case is *suspected* in tests using pooled antigens by (1) a fourfold or greater rise in antibody titer to specific species (e.g., *L. longbeachae*) or non–serogroup 1 *L. pneumophila* or (2) a fourfold or greater rise in antibody titer to multiple species of *Legionella*. Some experts think that a single antibody level of ≥1:256 may be an adequate basis for diagnosing a presumptive case, but most prefer paired serology for confirmation. Serology is an imperfect tool; data suggest that as many as 20–30% of patients with

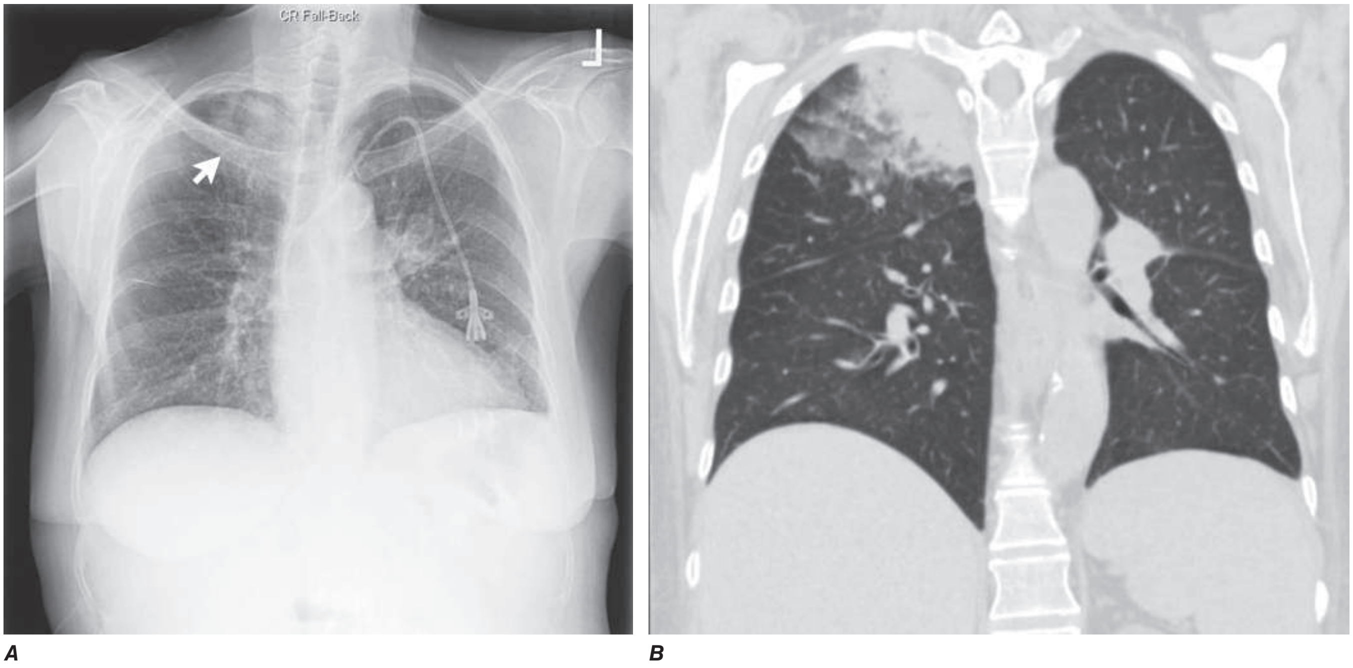

FIGURE 164-4 Right-upper-lobe infiltrate in a patient with *L. pneumophila* pneumonia on chest x-ray and computed tomography (CT). An immunosuppressed patient from a long-term care facility presented with cough, sputum production, fever, and chills. New renal insufficiency and hyponatremia were documented. A chest x-ray (***A***) was consistent with a small right-upper-lobe infiltrate (*white arrow*), which was confirmed by CT (***B***). Urinary antigen testing for *L. pneumophila* serotype 1 was negative, but polymerase chain reaction on bronchoalveolar lavage fluid was positive for *L. pneumophila*.

proven legionellosis may not mount an antibody response that is sufficient for diagnosis, and the sensitivity and specificity of seroconversion with regard to non-*pneumophila Legionella* species are unclear among patients with altered immunity. Serology can provide important information for epidemiologic investigations, helping to identify additional cases missed by other diagnostic methods. In addition, the use of serologic testing during outbreak studies allows the investigation of patients without severe disease (e.g., those with Pontiac fever).

DIRECT FLUORESCENT ANTIBODY TESTING The sensitivity of direct fluorescent antibody (DFA) testing of sputum is lower than that of other testing modalities, ranging from 20 to 70% depending on the assay used. Most available assays target specific species (e.g., *L. pneumophila*) or serotypes. DFA testing may have a higher positive predictive value in patients with severe pneumonia or symptoms consistent with Legionnaires' disease, but it is not recommended for screening of low-risk patients because of the frequency of false-positive results.

MOLECULAR TESTING Polymerase chain reaction (PCR), loop-mediated isothermal amplification (LAMP), and other nucleic acid amplification tests are highly sensitive for lower respiratory tract specimens

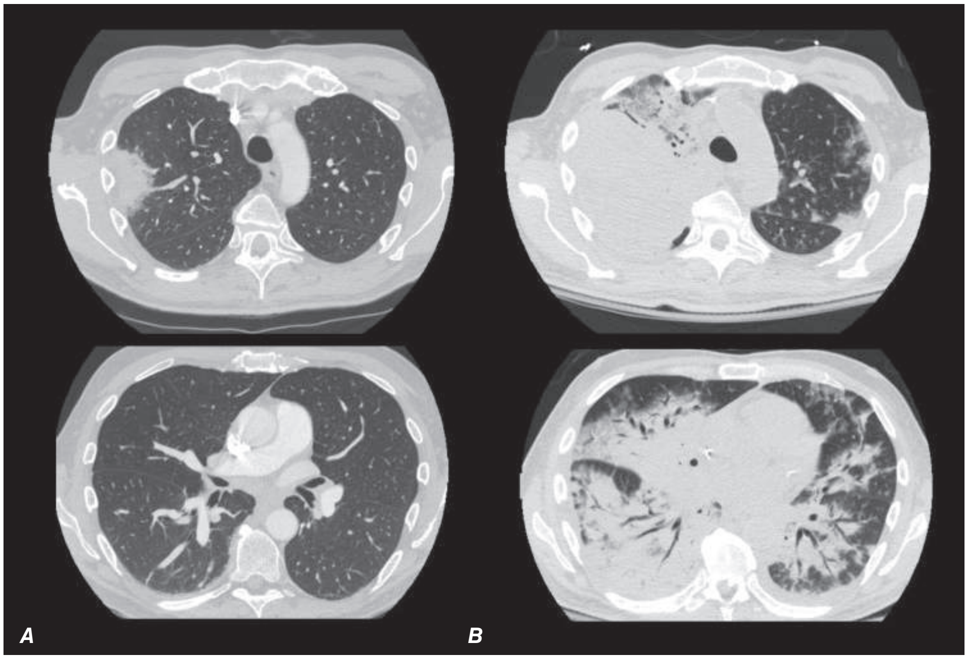

FIGURE 164-5 Nodular disease presentation on computed tomography (CT) in an immunosuppressed patient infected with *L. micdadei*. *A.* CT scan in a hematopoietic cell transplant recipient presenting with fever and cough. A pulmonary nodule was noted in the right upper lobe. Bronchoscopy was performed; cultures were positive on day 5 for small white colonies on buffered charcoal yeast extract plates, and these colonies were eventually identified as *L. micdadei*. ***B.*** Repeat CT scan at day 12 demonstrated an enlarging nodule, diffuse infiltrates, and possible cavitation. The patient required intensive care unit admission and intubation despite appropriate targeted antibiotic therapy.

(e.g., sputum) and are becoming more widely available. Molecular methods can detect *Legionella* from multiple sources but are most commonly used for respiratory specimens such as sputum and bronchoalveolar lavage fluid. PCR is more sensitive than culture; in some studies, up to two to four times as many cases of lower tract disease were detected only by molecular methods. Molecular techniques also are useful in diagnosing infection in patients during antibiotic therapy. However, PCR methods are not used to determine *L. pneumophila* serotypes—information that is needed for epidemiologic investigations—and most commercially available assays target only *L. pneumophila*. Multiplex PCR tests for pneumonia and other respiratory pathogens are increasingly available and may include *L. pneumophila*.

TREATMENT

Legionella Pneumonia

Treatment of *Legionella* pneumonia involves antibiotics that target intracellular pathogens, whereas patients with Pontiac fever do not require antibiotic therapy. Macrolides and fluoroquinolones are the first-line agents for *Legionella* pneumonia according to guidelines in the United States and Europe **(Table 164-3)**. Macrolides disrupt protein production critical for survival of the organism. Although erythromycin and clarithromycin are both effective, azithromycin is the preferred agent, as it is easier to tolerate and has fewer drug-drug interactions. Azithromycin and clarithromycin also reach higher intracellular concentrations than erythromycin.

Fluoroquinolones are potent agents against *Legionella* species. Data from both in vitro and in vivo models of infection suggest that fluoroquinolones may be more effective than macrolides, but no randomized clinical trials have yet compared the two drug classes for treatment of legionellosis. In nonrandomized observational studies, fluoroquinolones have been shown to be more effective than macrolides (erythromycin and clarithromycin) in terms of fever resolution and decreased duration of hospitalization; other such studies have shown no difference in outcome.

Both macrolides and fluoroquinolones are available as IV and oral formulations. Most experts prefer IV therapy during the first few days of treatment for patients with severe *Legionella* pneumonia. Secondary agents, such as rifampin, doxycycline, minocycline, and, less frequently, trimethoprim-sulfamethoxazole, have also been used, with mixed responses. Tigecycline, a third-generation glycylcycline related to tetracyclines, has been used for treatment of patients with significant antibiotic allergies. The novel aminomethylcycline antibiotic omadacycline appears to be efficacious in vitro, but its clinical efficacy has not been studied to date, and it is not currently recommended for routine use. Although data are limited, combination therapy does not appear to improve outcomes.

The optimal duration of treatment for Legionnaire's disease is unknown. For community-acquired pneumonia, guidelines recommend treatment until the patient achieves clinical stability and for a minimum of 5 days. In the absence of data, a similar duration of treatment for *Legionella* is a reasonable approach. For immunosuppressed patients and patients with severe disease, a more protracted course of therapy is recommended. The duration of therapy for extrapulmonary manifestations of *Legionella* infection is unknown and depends on the site involved and clinical improvement. Resistance to macrolides and fluoroquinolones has been reported only rarely. Susceptibility testing is not routinely performed but is available in specialized laboratories and public health departments.

TABLE 164-3 ***Legionella* Treatment Options**

	DISEASE SEVERITY	
DISEASE	**MILD**	**MODERATE/SEVERE**[a,b]
Pontiac fever	None	N/A
Legionella pneumonia	**A) Fluoroquinolone:** **Levofloxacin, 750 mg PO once daily** *or* Ciprofloxacin, 500 mg PO twice daily *or* Moxifloxacin, 400 mg PO once daily ***or*** **B) Macrolide:** **Azithromycin, 500 mg PO once daily (day 1), followed by 250 mg PO once daily (for minimum of 4 days)** *or* Clarithromycin, 400 mg PO daily	**A) Fluoroquinolone:**[c] **Levofloxacin, 750 mg IV once daily** *or* Ciprofloxacin, 500 mg IV twice daily *or* Moxifloxacin, 400 mg PO twice daily ***or*** **B) Macrolide:**[c] **Azithromycin, 500 mg IV daily** *or* Clarithromycin, 400 mg IV twice daily

Note: Agents in bold type are considered first-line treatment.

[a]All immunosuppressed patients should be considered to have moderate or severe disease and started on IV therapy if possible. [b]All patients requiring inpatient care should receive IV therapy until improving, when they can be switched to an oral agent. [c]Consider dual therapy, either with dual fluoroquinolone and macrolide therapy, or either agent with another secondary agent (e.g., rifampin). Secondary agents include doxycycline, minocycline, rifampin, and trimethoprim-sulfamethoxazole, all with varying efficacy for treatment.

Abbreviations: IV, intravenous; N/A, not applicable; PO, oral.

OUTCOMES

Legionella infections are associated with significant morbidity and mortality, leading to hospitalization and ICU admission of most patients who develop pneumonia. Case-fatality rates of *Legionella* pneumonia are reported to be ~10%, with death more likely among patients who are admitted to the ICU or have major comorbidities. Among patients in whom antibiotic treatment is delayed, mortality rates are approximately three times higher than among those treated earlier. Patients who develop nosocomial pneumonia attributable to health care–associated exposures, particularly those due to *L. pneumophila*, have case-fatality rates of ~25%. Death is a much more common outcome among immunocompromised hosts, whose mortality rates can reach ~30–50%. Assessment of long-term follow-up among patients who recover from severe *Legionella* pneumonia demonstrates that more than one-quarter have ongoing complications after recovery, including recurrent hospitalizations, acute renal failure, respiratory complications, and recurrent pneumonias. In contrast, recovery from Pontiac fever usually takes place within 3–5 days, as the disease is self-limiting; hospitalization, complications, and death related to Pontiac fever are extremely rare.

PREVENTION

Prevention of legionellosis starts with addressing water systems. Large municipal water systems provide water throughout the globe, but the quality of these systems varies regionally; many areas have limited access to potable water. Only limited regions have the resources to address *Legionella* water contamination; most water-monitoring agencies focus on control of enteric pathogens, such as *Escherichia coli* and other coliform bacteria, and do not have an adequate infrastructure to address *Legionella*. Even in countries and cities with more complex water systems, there is wide variation in how waterborne pathogens are addressed, and rules and regulations are often country dependent. In the Netherlands, for example, chlorination is not routine, whereas the United Kingdom and most countries in the European Union use chlorine routinely as the primary mode of disinfection for public water systems. Although regulated by the Environmental Protection Agency, management and treatment strategies in the United States vary by state and, in some instances, by city.

Prevention in the United States focuses on health care organizations and hospitals, where water-based exposures are more often linked to case fatalities. Federal requirements to reduce *Legionella* risk in the United States were first established in June 2017, when the Centers for Medicare and Medicaid Services required that all health care organizations develop and adhere to water management plans. These plans require the development of multidisciplinary teams, an understanding of the organization's water system, identification of high-risk areas (e.g., transplant units, oncology floors), identification of at-risk structures for *Legionella* growth, implementation and monitoring of control measures, methods for intervention if control measures fail, and procedures to assure documentation that policies are followed. All medical centers are required to have an awareness of water quality and to have systems in place to help prevent nosocomial *Legionella* pneumonia. Such policies leave water quality assessment, including testing for *Legionella*, up to the individual facility. In addition to hospitals, an increasing number of cities, including New York City, require similar water-management plans for cooling towers, with registration, testing, and mitigation options.

Even if detected in regional water systems, *Legionella* becomes a human pathogen only after replication in premise plumbing systems. In buildings, *Legionella* finds the ideal environment for logarithmic growth, which leads to exposures and subsequent disease. An important first step in prevention within hospitals is a review of plumbing systems to identify areas of concern and a review of impact areas such as dental clinics, ICUs, rehabilitation units, and units that house high-risk patients. Specific water features, such as therapy pools, ice machines, and decorative fountains, need policies for cleaning and disinfection. Targeted approaches to management of cooling towers, such as high-efficiency drift eliminators and routine maintenance, are important considerations. In addition, areas that have undergone recent construction or renovation should be flagged, with prevention policies in place to address the associated risks. New construction or structural updates can lead to water stagnation, while modifications to plumbing can disrupt biofilms. Units with older premise plumbing are thought to be at higher risk, but even brand new facilities can become colonized during construction, with consequent outbreaks.

Testing for *Legionella* is an important step when presumptive or possible nosocomial pneumonia cases occur and can help address a facility's potential risks. There are a number of methods for environmental testing for *Legionella*, but environmental cultures are used in most hospitals because they quantify *Legionella* levels, allow species identification/serotyping, and can link environmental sources to nosocomial outbreaks. Testing usually focuses on locations where the index patient(s) may have had potential waterborne exposures (e.g., at showers and sinks). Other adjacent areas, along with those noted to be high-risk locations within the hospital, should be considered for additional testing; positive results should widen the testing area. Proactive testing is increasingly being used to preclude nosocomial cases; however, if testing is planned, it should be coupled with a management plan that addresses how *Legionella* will be dealt with if it is found in the water system and where and how frequently testing should be done; we recommend biannual or quarterly testing of select sites within hospital systems.

If a common-source outbreak is discovered, a number of approaches can be used to address *Legionella*. Regardless of source, immediate limitation of ongoing water exposures for patients in the affected room, unit, or floor is a crucial step in avoiding additional cases. Removing or replacing water features associated with exposures, such as decorative fountains and affected equipment or plumbing devices, may be needed. Immediate interventions such as heat shock (increasing water temperatures for a limited period) and hyperchlorination may also be useful as short-term steps in addressing an outbreak.

The addition of a disinfectant to the water system is one of the most common ways to address the presence of *Legionella*. Chemical disinfection with agents such as chlorine or monochloramine and copper and silver ionization are commonly used for secondary disinfection. Use of disinfectants requires routine maintenance and monitoring of chemical or ion levels to assure that they are sufficient for prevention. Lack of monitoring and system failures have led to breakthrough nosocomial *Legionella* cases. Another option is water filtration, which either can serve as a primary method for prevention or can be used in combination with secondary disinfection. Filters—either in-line with plumbing or at point-of-use sites—can be considered for either short- or long-term prevention during an outbreak. However, filters have a limited life span, can weaken water pressure, and are costly to maintain.

FURTHER READING

Cassell K et al: Estimating the true burden of Legionnaires' disease. Am J Epidemiol 188:1686, 2019.

Centers for Disease Control and Prevention: Developing a water management program to reduce *Legionella* growth and spread in buildings: A practical guide to implementing industry standards. June 24, 2021. Available at *https://www.cdc.gov/legionella/wmp/toolkit/*. Accessed October 10, 2023.

Centers for Disease Control and Prevention: Legionnaires' disease surveillance summary report, United States 2018–2019. Available at *https://www.cdc.gov/legionella/health-depts/surv-reporting/2018-19-surv-report-508.pdf*. Accessed October 20, 2023.

Kato H et al: Meta-analysis of fluoroquinolones versus macrolides for treatment of legionella pneumonia. J Infect Chemother 27:424, 2021.

National Academies of Sciences, Engineering, and Medicine: *Management of Legionella in Water Systems*. Washington, DC, The National Academies Press, 2020.

Pierre DM et al: Diagnostic testing for Legionnaires' disease. Ann Clin Microbiol Antimicrob 16:1, 2017.

165

Pertussis and Other *Bordetella* Infections

May S. ElSherif, Scott A. Halperin

Pertussis is an acute infection of the respiratory tract caused by *Bordetella pertussis*. The word *pertussis* means "violent cough," which aptly describes the most consistent and prominent feature of the illness. The inspiratory sound made at the end of an episode of paroxysmal coughing gives rise to the common name for the illness, "whooping cough." However, this feature is variable: it is uncommon among infants ≤6 months of age and is frequently absent in older children and adults. The Chinese name for pertussis is "the 100-day cough," which describes the clinical course of the illness accurately. The identification of *B. pertussis* was first reported by Bordet and Gengou in 1906, and vaccines were produced over the following two decades.

MICROBIOLOGY

Of the 10 identified species in the genus *Bordetella*, only four are of major medical significance. *B. pertussis* infects only humans and is the most important *Bordetella* species causing human disease. *B. parapertussis* causes an illness in humans that is similar to pertussis but is typically milder; co-infections with *B. parapertussis* and *B. pertussis* have been documented. With improved polymerase chain reaction (PCR) diagnostic methodology, up to 20% of patients with a pertussis-like syndrome have been found to be infected with *B. holmesii*, formerly thought to be an unusual cause of bacteremia. *B. bronchiseptica* is an important pathogen of domestic animals that causes kennel cough in dogs, atrophic rhinitis and pneumonia in pigs, and pneumonia in cats. Both respiratory infection and opportunistic infection due to *B. bronchiseptica* are reported occasionally in humans. *B. petrii*, *B. hinzii*, and *B. ansorpii* have been isolated from patients who are immunocompromised.

Bordetella species are gram-negative pleomorphic aerobic bacilli that share common genotypic characteristics. *B. pertussis* and *B. parapertussis* are the most similar of the species, but *B. parapertussis* does not express the gene coding for pertussis toxin. *B. pertussis* is a slow-growing fastidious organism that requires selective medium and forms small, glistening, bifurcated colonies. Suspicious colonies are presumptively identified as *B. pertussis* by direct fluorescent antibody testing or by agglutination with species-specific antiserum. *B. pertussis* is further differentiated from other *Bordetella* species by biochemical and motility characteristics.

B. pertussis produces a wide array of toxins and biologically active products that are important in its pathogenesis and in immunity. Most of these virulence factors are under the control of a single genetic locus that regulates their production, resulting in antigenic modulation and phase variation. Although these processes occur both in vitro and in vivo, their importance in the pathobiology of the organism is unknown; they may play a role in intracellular persistence and person-to-person spread. The organism's most important virulence factor is *pertussis toxin*, which is composed of a B oligomer–binding subunit and an enzymatically active A protomer that ADP-ribosylates a guanine nucleotide–binding regulatory protein (G protein) in target cells, producing a variety of biologic effects. Pertussis toxin has important mitogenic activity, affects the circulation of lymphocytes, and serves as an adhesin for bacterial binding to respiratory ciliated cells. Other important virulence factors and adhesins are *filamentous hemagglutinin*, a component of the cell wall, and *pertactin*, an outer-membrane protein. *Fimbriae*, bacterial appendages that play a role in bacterial attachment, are the major antigens against which agglutinating antibodies are directed. These agglutinating antibodies have historically been the primary means of serotyping *B. pertussis* strains. Other virulence factors include tracheal cytotoxin, a peptidoglycan fragment, which causes inflammatory respiratory epithelial damage; adenylate cyclase-hemolysin toxin, which impairs host phagocytic cell function; dermonecrotic toxin, which may contribute to respiratory mucosal damage; and lipooligosaccharide, which has properties similar to those of other gram-negative bacterial endotoxins. Since 2010, the emergence of pertactin-negative strains worldwide has been attributed to immune pressure resulting from the use of pertactin-containing acellular pertussis vaccines, reaching >80% dominance in some countries.

EPIDEMIOLOGY

Pertussis is a highly communicable disease, with attack rates of 80–100% among unimmunized household contacts and 20% within households in well-immunized populations. The infection has a worldwide distribution, with cyclical outbreaks every 3–5 years (a pattern that has persisted despite widespread immunization). Pertussis occurs in all months; however, in North America, its activity peaks in autumn and winter.

In developing countries, pertussis remains an important cause of infant morbidity and mortality despite the reported worldwide decrease in incidence following improved vaccine coverage **(Fig. 165-1)**. Monitored as a surrogate to evaluate immunization programs, 2022 DTP3 (diphtheria, tetanus, and pertussis) coverage rates are still <50% in many developing nations while globally recovering to 84% from 2021's 81%; the World Health Organization (WHO) estimates that 90% of the burden of pertussis is in developing regions. The overreporting of immunization coverage and underreporting of disease result in substantial underestimation of the global burden of pertussis. WHO estimates are 40 million cases globally each year, with 400,000 deaths, mostly among infants <3 months of age.

Before the institution of widespread immunization programs in the developed world, pertussis was one of the most common infectious causes of morbidity and death. In the United States before the 1940s, between 115,000 and 270,000 cases of pertussis were reported annually, with an average yearly rate of 150 cases per 100,000 population. With universal childhood immunization, the number of reported cases fell by >90%, and mortality rates decreased even more dramatically. Only 1010 cases of pertussis were reported in 1976 **(Fig. 165-2)**. After that historic low, rates of pertussis increased slowly. In recent years, pertussis epidemics have been reported with increasing frequency in high-income countries, including Australia, the United Kingdom, and the United States. The United States experienced widespread pertussis outbreaks in 2005, 2010, 2012, 2014, and 2015 at levels not seen in 40–50 years (48,000 reported cases in 2012).

Although thought of as a disease of childhood, pertussis can affect people of all ages and is a known cause of prolonged coughing illness in adolescents and adults. In unimmunized populations, pertussis incidence peaks during the preschool years, and well over half of children have the disease before reaching adulthood. In highly immunized populations such as those in North America, the peak incidence is among infants <1 year of age who have not completed the three-dose primary immunization series. An increase in pertussis incidence among adolescents and adults began in the late 1990s and led to the introduction of an adolescent booster dose across North America by 2006. While the disease burden among adolescents decreased initially, children 7–10 years of age emerged as a high-risk group during a major outbreak in 2010. Most of the affected children were fully immunized. Subsequent outbreaks in 2012, 2014, and 2015 showed a shift in epidemiology, with pertussis incidence increasing among adolescents while still remaining elevated among 10-year-olds. The most highly affected cohorts were those who received acellular pertussis vaccines in infancy. Although adults contribute a smaller proportion of reported cases of pertussis than do children and adolescents, this difference may be related to a greater degree of underrecognition and underreporting. Several studies of prolonged

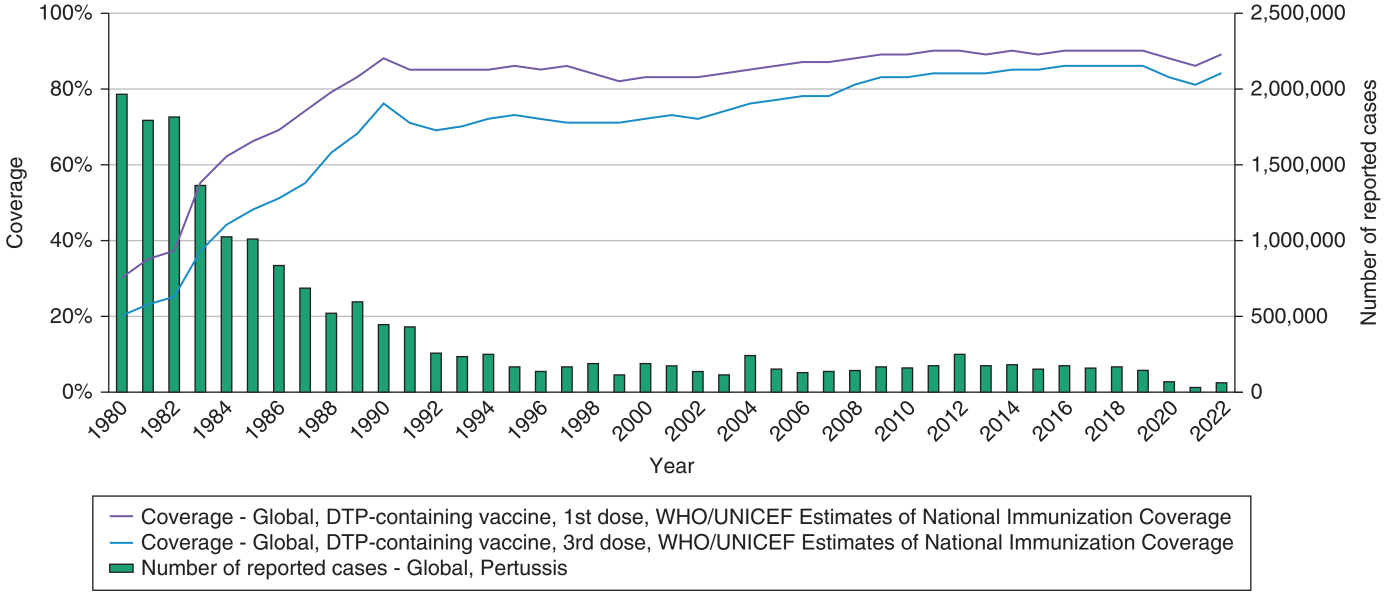

FIGURE 165-1 Global annual reported cases of pertussis and rate of coverage with DTP3 (diphtheria toxoid, tetanus toxoid, and pertussis vaccine; one and three doses), 1980–2022. *(Reproduced with permission from Diphtheria tetanus toxoid and pertussis (DTP) vaccination coverage (who.int). World Health Organization Immunization Data Portal; WHO 2023.)*

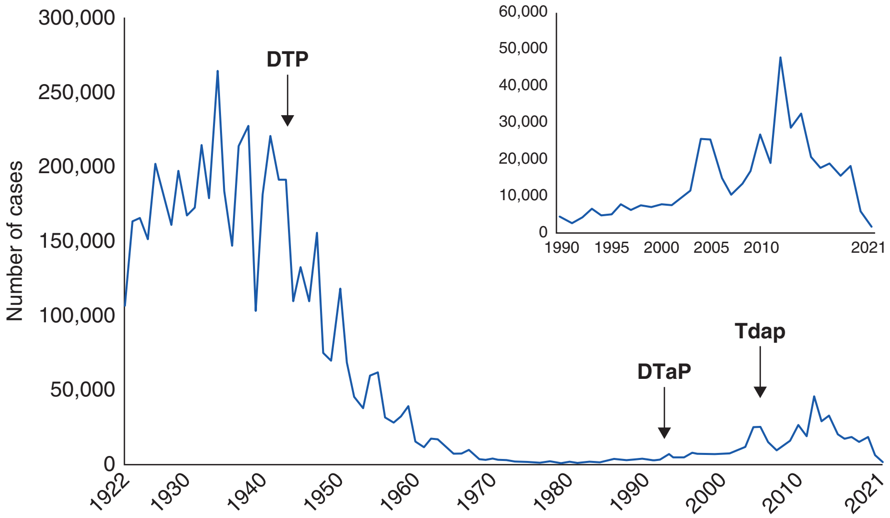

FIGURE 165-2 Reported cases of pertussis by year: United States, 1922–2021, marking the introduction of different vaccine types. *(From the Centers for Disease Control and Prevention, Pertussis Surveillance Trend Reporting and Case Defintion | CDC. Available at https://www.cdc.gov/pertussis/surv-reporting.html. Accessed December 26, 2023.)*

coughing illness suggest that *B. pertussis* may be the etiologic agent in 12–30% of adults with cough that does not improve within 2 weeks. In one study of the efficacy of an acellular pertussis vaccine in adolescents and adults, the incidence of pertussis in the placebo group was 3.7–4.5 cases per 1000 person-years. Although this prospective cohort study yielded a lower estimate than the studies of cough illness, its results still translate to ~1 million cases of pertussis annually among adults in the United States. In addition, asymptomatic pertussis infection, estimated at 56% by a systemic review of tested household contacts, is common and appears to contribute to disease transmission.

Severe morbidity and high mortality rates, however, are restricted almost entirely to infants. In the United States between 2000 and 2017, infants <2 months old accounted for 84% of pertussis deaths with annual mean number of cases, hospitalizations, and deaths at 2957 (range 1803–4994), 1122 (range 544–1938), and 15 (range 3–35), respectively. Although school-age children are the source of infection for most households, adults are often the source for cases in high-risk infants and may serve as the reservoir of infection between epidemic years.

PATHOGENESIS

Infection with *B. pertussis* is initiated by attachment of the organism to the ciliated epithelial cells of the nasopharynx. Attachment is mediated by surface adhesins (e.g., pertactin and filamentous hemagglutinin), which bind to the integrin family of cell-surface proteins, probably in conjunction with pertussis toxin. The role of fimbriae in adhesion and in maintenance of infection has not been fully delineated. Perhaps the result of redundancy of adhesins, no differences in virulence or clinical manifestations have been detected with the emergence of pertactin-negative strains. At the site of attachment, the organism multiplies, producing a variety of other toxins that cause local mucosal damage (tracheal cytotoxin, dermonecrotic toxin). Impairment of host defense by *B. pertussis* is mediated by pertussis toxin and adenylate cyclase-hemolysin toxin. There is local cellular invasion, with intracellular bacterial persistence; however, systemic dissemination does not occur. Systemic manifestations (lymphocytosis) result from the effects of the toxins.

The pathogenesis of the clinical manifestations of pertussis is poorly understood. It is not known what causes the hallmark paroxysmal cough. A pivotal role for pertussis toxin has been proposed but has not been confirmed. It is thought that neurologic events in pertussis, such as seizures and encephalopathy, are due to hypoxia from coughing paroxysms or apnea rather than to the effects of specific bacterial products. *B. pertussis* pneumonia, which occurs in up to 22% of infants with pertussis, is usually a diffuse bilateral primary infection. In older children and adults with pertussis, pneumonia is often due to secondary bacterial infection with streptococci or staphylococci. Deaths from pertussis among young infants are frequently associated with very high levels of leukocytosis and pulmonary hypertension.

IMMUNITY

Both humoral and cell-mediated immunity are thought to be important in pertussis. Although immunity after natural infection was thought to be lifelong, seroepidemiologic evidence demonstrates that it is not and that subsequent episodes of clinical pertussis are prevented by intermittent subclinical infection. Pertussis agglutinins were correlated with protection in early studies of whole-cell pertussis vaccines. Antibodies to pertussis toxin, filamentous hemagglutinin, pertactin, and fimbriae are all protective in animal models. Serologic correlates of protection conferred by acellular pertussis vaccines have not been established, although antibody to pertactin, fimbriae, and (to a lesser degree) pertussis toxin correlated best with protection in two efficacy trials. The duration of immunity after whole-cell pertussis vaccination is short-lived, with little protection remaining after 10–12 years. Waning of immunity is even more rapid in adolescents and children who have received all their immunizations with acellular vaccines—i.e., within 2–4 years after the fifth or sixth dose. The type of immune response elicited may affect duration of protection; natural infection and whole-cell pertussis vaccine elicit a Th1/Th17-predominant response, whereas acellular pertussis vaccines stimulate a Th2-biased response. Controlled Human Infection Models of pertussis are under development to facilitate better understanding of immunity after infection and after vaccination.

CLINICAL MANIFESTATIONS

Pertussis is a prolonged coughing illness with clinical manifestations that vary by age **(Table 165-1)**. Although not uncommon among adolescents and adults, classic pertussis is most often seen in preschool and

TABLE 165-1 Clinical Features of Pertussis, by Age Group and Diagnostic Status

	PERCENTAGE OF PATIENTS	
FEATURE	ADOLESCENTS AND ADULTS	INFANTS AND CHILDREN
Cough		
Paroxysmal	70–99	89–93
Worse at night	61–87	41
Whoop	8–82	69–92
Post-tussive vomiting	17–65	48–60

Source: Reproduced with permission from PE Kilgore et al: Pertussis: Microbiology, disease, treatment, and prevention. Clin Microbiol Rev 29:449, 2016.

school-age children. After an incubation period averaging 7–10 days, an illness develops that is indistinguishable from the common cold and is characterized by coryza, lacrimation, mild cough, low-grade fever, and malaise. After 1–2 weeks, this *catarrhal phase* evolves into the *paroxysmal phase*: the cough becomes more frequent and spasmodic with repetitive bursts of 5–10 coughs, often within a single expiration. Post-tussive vomiting is frequent, with a mucous plug occasionally expelled at the end of an episode. The episode may be terminated by an audible whoop, which occurs upon rapid inspiration against a closed glottis at the end of a paroxysm. During a spasm, there may be impressive neck-vein distension, bulging eyes, tongue protrusion, and cyanosis. Paroxysms may be precipitated by noise, eating, or physical contact. Between attacks, the patient's appearance is normal, but increasing fatigue is evident. The frequency of paroxysmal episodes varies widely, from several per hour to 5–10 per day. Episodes are often worse at night and interfere with sleep. Most complications occur during the paroxysmal stage. Fever is uncommon and suggests bacterial superinfection.

After 2–4 weeks, the coughing episodes become less frequent and less severe—changes heralding the onset of the *convalescent phase*. This phase can last 1–3 months and is characterized by gradual resolution of coughing episodes. For 6–12 months, intercurrent viral infections may be associated with a recrudescence of paroxysmal cough.

Not all individuals who develop pertussis have classic disease. The clinical manifestations in adolescents and adults are more often atypical. The cough is severe, prolonged, and often paroxysmal. Though uncommon, a whoop and vomiting with cough are more specific signs of pertussis in adults with prolonged cough. Other suggestive features are a cough at night, sweating episodes between paroxysms of coughing, and exposure to other individuals with a prolonged coughing illness.

COMPLICATIONS

Complications are frequently associated with pertussis and are more common among infants than among older children or adults. Subconjunctival hemorrhages, abdominal and inguinal hernias, pneumothoraces, and facial and truncal petechiae can result from increased intrathoracic pressure generated by severe fits of coughing. Weight loss can follow decreased caloric intake. Urinary incontinence, rib fracture, carotid artery aneurysm, and cough syncope have also been reported in adolescents and adults with pertussis. In a series of >1100 children <2 years of age who were hospitalized with pertussis, 27.1% had apnea, 9.4% had pneumonia, 2.6% had seizures, and 0.4% had encephalopathy; 10 children (0.9%) died. Pneumonia is reported in <5% of adolescents and adults and increases in frequency after 50 years of age. In contrast to the primary *B. pertussis* pneumonia that develops in infants, pneumonia in older children, adolescents, and adults with pertussis is usually caused by a secondary infection with encapsulated organisms such as *Streptococcus pneumoniae* or *Haemophilus influenzae*.

DIAGNOSIS

If the classic symptoms of pertussis are present, clinical diagnosis is not difficult. However, particularly in older children and adults, it is difficult to differentiate infections caused by *B. pertussis* and *B. parapertussis* from other respiratory tract infections on clinical grounds. Therefore, laboratory confirmation should be attempted in all cases. Lymphocytosis (absolute lymphocyte count $>10^8$–10^9/L) is common among young children, in whom it is unusual with other infections, but not among adolescents and adults. Culture of nasopharyngeal secretions remains the gold standard of diagnosis because of its 100% specificity, although DNA detection by PCR has replaced culture in most laboratories because of substantially increased sensitivity and quicker results. Multitarget real-time PCR methodology includes primers that differentiate between *B. pertussis*, *B. parapertussis*, and *B. holmesii*. The best specimen is collected by nasopharyngeal aspiration, in which a fine flexible plastic catheter attached to a 10-mL syringe is passed into the nasopharynx and withdrawn while gentle suction is applied. Since *B. pertussis* is highly sensitive to drying, secretions for culture should be inoculated without delay onto appropriate medium (Bordet-Gengou or Regan-Lowe), or the catheter should be flushed with a phosphate-buffered saline solution for culture and/or PCR. An alternative to the aspirate is a Dacron or rayon nasopharyngeal swab; again, inoculation of culture plates should be immediate or an appropriate transport medium (e.g., Regan-Lowe charcoal medium) should be used. Results of PCR can be available within hours; cultures become positive by day 5 of incubation.

Nasopharyngeal cultures in untreated pertussis remain positive for a mean of 3 weeks after the onset of illness; these cultures become negative within 5 days of the institution of appropriate antimicrobial therapy. The duration of a positive PCR in untreated pertussis or after therapy is not known but exceeds that of positive cultures. Since much of the period during which the organism can be recovered from the nasopharynx falls into the catarrhal phase, when the etiology of the infection is not suspected, there is only a small window of opportunity for culture- or PCR-proven diagnosis. Cultures and PCR from infants and young children are more frequently positive than those from older children and adults; this difference may reflect earlier presentation of the former age group for medical care.

As a result of the difficulties with laboratory diagnosis of pertussis in adolescents, adults, and patients who have been symptomatic for >4 weeks, increasing attention is being given to serologic diagnosis. Enzyme immunoassays detecting IgA and IgG antibodies to pertussis toxin, filamentous hemagglutinin, pertactin, and fimbriae have been developed and assessed for reproducibility. Two- or fourfold increases in antibody titer are suggestive of pertussis, although cross-reactivity of some antigens (such as filamentous hemagglutinin and pertactin) among *Bordetella* species makes it difficult to depend diagnostically on seroconversion involving a single type of antibody. Criteria for serologic diagnosis based on comparison of results for a single serum specimen with established population values are gaining acceptance, and serologic measurement of antibody to pertussis toxin is becoming more widely standardized and available for diagnostic purposes, particularly in outbreak settings and for surveillance.

DIFFERENTIAL DIAGNOSIS

A child presenting with paroxysmal cough, post-tussive vomiting, and whoop is likely to have an infection caused by *B. pertussis* or *B. parapertussis*; lymphocytosis increases the likelihood of a *B. pertussis* etiology. Viruses such as respiratory syncytial virus, rhinovirus, and adenovirus have been isolated from patients with clinical pertussis but probably represent co-infection, particularly in children <1 year of age.

In adolescents and adults, who often do not have paroxysmal cough or whoop, the differential diagnosis of a prolonged coughing illness is more extensive. Pertussis should be suspected when any patient has a cough that does not improve within 14 days, a paroxysmal cough of any duration, a cough followed by vomiting (adolescents and adults), or any respiratory symptoms after contact with a laboratory-confirmed case of pertussis. Other etiologies to consider include infections caused by *Mycoplasma pneumoniae*, *Chlamydia pneumoniae*, adenovirus, influenza virus, and other respiratory viruses. Use of angiotensin-converting enzyme (ACE) inhibitors, reactive airway disease, and gastroesophageal reflux disease are well-described noninfectious causes of prolonged cough in adults.

TREATMENT

Pertussis

ANTIBIOTICS

The purpose of antibiotic therapy for pertussis is to eradicate the infecting bacteria from the nasopharynx; therapy does not substantially alter the clinical course unless given early in the catarrhal phase. Macrolide antibiotics are the drugs of choice for treatment of pertussis **(Table 165-2)**; macrolide-resistant *B. pertussis* strains have been reported but are rare. Trimethoprim-sulfamethoxazole is recommended as an alternative for individuals allergic to macrolides.

SUPPORTIVE CARE

Young infants have the highest rates of complication and death from pertussis; therefore, most infants (and older children with

TABLE 165-2 Antimicrobial Therapy for Pertussis

PATIENT AGE GROUP	PRIMARY AGENTS			ALTERNATE AGENT
	AZITHROMYCIN (AZ)	ERYTHROMYCIN (ER)	CLARITHROMYCIN	TRIMETHOPRIM (TMP)-SULFAMETHOXAZOLE (SMZ)[a]
<1 month	AZ is the recommended agent for this age group, at 10 mg/kg per day in a single dose for 5 days	ER is not preferred, only use if AZ unavailable; 40–50 mg/kg per day in 4 divided doses for 14 days	Not recommended in this age group	Contraindicated in infants <2 months
1-5 months	10 mg/kg per day in a single dose for 5 days	40–50 mg/kg per day in 4 divided doses for 14 days	15 mg/kg per day in 2 divided doses for 7 days	Only for infants ≥2 months (otherwise contraindicated); TMP 8 mg/kg per day, SMZ 40 mg/kg per day in 2 divided doses for 14 days
Infants (≥6 months) and children	10 mg/kg on day 1, 5 mg/kg mg per day (maximum 500 mg) on days 2–5	40–50 mg/kg per day (maximum 2 g/d) in 4 divided doses for 14 days	15 mg/kg per day (maximum 1 g/d) in 2 divided doses for 7 days	TMP 8 mg/kg per day, SMZ 40 mg/kg per day in 2 divided doses for 14 days
Adults	500 mg on day 1, 250 mg/d on days 2–5	2 g/d in 4 divided doses for 14 days	1 g/d in 2 divided doses for 7 days	TMP 320 mg/d, SMZ 1600 mg/d in 2 divided doses for 14 days
Comments	Abdominal discomfort, prescribed with caution to cardiac patients	Frequent gastrointestinal side effects, hypersensitivity reactions	Epigastric distress and cramps	For patients allergic to macrolides, data on effectiveness are limited

[a]Trimethoprim-sulfamethoxazole (TMP-SMZ) replaces macrolides only in the event of a macrolide-resistant strain of *Bordetella pertussis* or when patients older than 2 months are allergic to or cannot tolerate macrolides.

Source: T Tiwari et al: Recommended antimicrobial agents for the treatment and postexposure prophylaxis of pertussis: 2005 CDC guidelines. MMWR Recomm Rep 54(RR-14):1, 2005.

severe disease) should be hospitalized. A quiet environment may decrease the stimulation that can trigger paroxysmal episodes. Use of β-adrenergic agonists and/or glucocorticoids has been advocated by some authorities but has not been proven to be effective. Cough suppressants are not effective and play no role in the management of pertussis.

INFECTION CONTROL MEASURES

Hospitalized patients with pertussis should be placed in respiratory isolation, with the use of precautions appropriate for pathogens spread by large respiratory droplets. Isolation should continue for 5 days after initiation of macrolide therapy or, in untreated patients, for 3 weeks (i.e., until nasopharyngeal cultures are consistently negative).

PREVENTION

Chemoprophylaxis Because the risk of transmission of *B. pertussis* within households is high, chemoprophylaxis is widely recommended for household contacts of pertussis cases regardless of their immunization status and should be initiated within 21 days of cough onset in the index case. The effectiveness of chemoprophylaxis is supported by several epidemiologic studies of institutional and community outbreaks. In the only randomized, placebo-controlled study, erythromycin estolate (50 mg/kg per day; maximum dose, 1 g/d) was effective in reducing the incidence of bacteriologically confirmed pertussis by 67%; however, there was no decrease in the incidence of clinical disease. Despite these results, authorities continue to recommend chemoprophylaxis, particularly in households with members at high risk of severe disease (children <1 year of age, pregnant women). Data on the use of the newer macrolides for chemoprophylaxis are not available, but these drugs are commonly used because of their increased tolerability and their effectiveness.

Immunization (See also Chap. 129) The mainstay of pertussis prevention is active immunization. Pertussis vaccine became widely used in North America after 1940; the reported number of pertussis cases subsequently fell by >90%. Whole-cell pertussis vaccines are prepared through the heating, chemical inactivation, and purification of whole *B. pertussis* organisms. Despite their efficacy (average estimate, 85%; range for different products, 30–100%), whole-cell pertussis vaccines are associated with adverse events—both common (fever; injection-site pain, erythema, and swelling; irritability) and uncommon (febrile seizures, hypotonic-hyporesponsive episodes). Alleged associations of whole-cell pertussis vaccine with encephalopathy, sudden infant death syndrome, and autism, although not substantiated, spawned an active anti-immunization lobby. The development of acellular pertussis vaccines, which are effective and less reactogenic, has greatly alleviated concerns about the inclusion of pertussis vaccine in the combined infant immunization series.

Although a wide variety of acellular pertussis vaccines were developed, only a few are still marketed widely; all contain pertussis toxoid and filamentous hemagglutinin. One acellular pertussis vaccine also contains pertactin, and another contains pertactin and two types of fimbriae. Adult formulations of acellular pertussis vaccines have been shown to be safe, immunogenic, and efficacious in clinical trials in adolescents and adults and are now recommended for routine immunization of these groups in several countries.

Although whole-cell vaccines are still used extensively in developing regions of the world, acellular pertussis vaccines are used exclusively for childhood immunization in much of the developed world. In light of evidence of early waning of immunity among children who received acellular pertussis vaccine in infancy, the WHO Strategic Advisory Group of Experts (SAGE) recommends that countries using whole-cell pertussis vaccine for the primary infant immunization series continue to do so. In countries using acellular pertussis vaccines in infancy, additional booster immunizations in older children, adolescents, and adults are recommended to prevent pertussis in high-risk infants. Pertussis immunization is also recommended during pregnancy to increase passive transfer of maternal antibodies to the fetus. Studies in high-income countries demonstrate that immunization of women during pregnancy is 90–93% effective at preventing pertussis in infants <2 months of age and is safe. In North America, acellular pertussis vaccines for children are given as a three-dose primary series at 2, 4, and 6 months of age, with a reinforcing dose at 15–18 months of age and a booster dose at 4–6 years of age. Adolescents (11–18 years of age) and all unvaccinated adults should receive a dose of the adult-formulation diphtheria–tetanus–acellular pertussis vaccine. Immunization is specifically recommended for health care providers, individuals in close contact with infants, and women at 27–36 weeks of every pregnancy, preferably in the earlier weeks of this range. Pertussis vaccine coverage among U.S. adolescents was 89.9% in 2022; among pregnant women, it was 55.4% during October 2022 to January 2023; and it was low among adults, at 30.1% in 2019. Further improvements in adult vaccine coverage may permit better control of pertussis across the age spectrum, with collateral protection of infants too young to be immunized. However, more effective vaccines with longer-lasting protection will ultimately be needed to control this disease.

FURTHER READING

Craig R et al: Asymptomatic infection and transmission of pertussis in households: A systematic review. Clin Infect Dis 70:152, 2020.

Forsyth KD et al: Recommendations to control pertussis prioritized relative to economies: A Global Pertussis Initiative update. Vaccine 36:7270, 2018.

Havers FP et al: Use of tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis vaccines: Updated recommendations of the Advisory Committee on Immunization Practices—United States, 2019. MMWR Morb Mortal Wkly Rep 69:77, 2020.

Ma L et al: Pertactin-deficient *Bordetella pertussis*, vaccine-driven evolution, and reemergence of pertussis. Emerg Infect Dis 27:1561, 2021.

Macina D et al: Estimating the pertussis burden in adolescents and adults in the United States between 2007 and 2019. Hum Vaccin Immunother 19:2208514, 2023.

Miguelena Chamorro B et al: *Bordetella bronchiseptica* and *Bordetella pertussis*: Similarities and differences in infection, immuno-modulation, and vaccine considerations. Clin Microbiol Rev 36:e0016422, 2023.

Skoff TH: Sources of infant pertussis infection in the United States. Pediatrics 136:635, 2015.

Tatti KM et al: Novel multitarget real-time PCR assay for rapid detection of *Bordetella* species in clinical specimens. J Clin Microbiol 49:4059, 2011.

Winter K et al: Pertussis in California: A tale of 2 epidemics. Pediatr Infect Dis J 37:324, 2018.

Wright J et al: Uptake of pertussis immunization in pregnancy and determinants of vaccination in Toronto, Canada. Vaccine 41: 6895, 2023.

166

Diseases Caused by Gram-Negative Enteric Bacilli

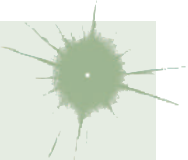

Thomas A. Russo, Yohei Doi

GENERAL FEATURES AND PRINCIPLES

The postantibiotic era has begun. For most patients, this is the first time in their lives when an effective treatment for a bacterial infection may not exist. Species in the order *Enterobacterales* are at the forefront of this evolving public health crisis. For example, the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) have designated carbapenem-resistant *Enterobacterales* (CRE) as representing a threat level of "urgent" and "priority one, critical." CRE are estimated to have caused more than 100,000 deaths in 2019 globally, and the disease burden is especially high in low- and middle-income countries (e.g., Indian Subcontinent). These pathogens cause a wide variety of infections involving diverse anatomic sites, mostly in compromised hosts but also in healthy individuals. Therefore, a thorough knowledge of clinical presentations and appropriate therapeutic choices is necessary for optimal outcomes. *Escherichia coli*, *Klebsiella*, *Proteus*, *Enterobacter*, *Serratia*, *Citrobacter*, *Morganella*, *Providencia*, *Cronobacter*, and *Edwardsiella* are enteric gram-negative bacilli (GNB) within the order *Enterobacterales* that commonly cause extraintestinal infections. *Salmonella*, *Shigella*, and *Yersinia*, which also are in the order *Enterobacterales* but more commonly cause gastrointestinal infections, are discussed in **Chaps. 171, 172, and 176**, respectively.

EPIDEMIOLOGY

E. coli, *Klebsiella*, *Proteus*, *Enterobacter*, *Serratia*, *Citrobacter*, *Morganella*, *Providencia*, *Cronobacter*, and *Edwardsiella* are components of the normal animal and human colonic microbiota and/or the microbiota in various environmental habitats, including long-term-care facilities (LTCFs) and hospitals. As a result, except for certain pathotypes of intestinal pathogenic *E. coli*, these genera are global pathogens. The incidence of infection due to these agents is increasing because of the combination of an aging population and increasing antimicrobial resistance. In healthy humans, *E. coli* is the predominant species of GNB in the colonic microbiota, followed by *Klebsiella* and *Enterobacter*. GNB can also colonize the oropharynx and intact skin but, in healthy individuals, tend to do so only transiently. By contrast, in LTCFs and hospital settings, a variety of GNB emerge as the dominant colonizers of both mucosal and skin epithelial surfaces, particularly in association with antimicrobial use, severe illness, and extended length of stay. LTCFs are emerging as an important reservoir for resistant GNB. Such colonization with GNB may lead to subsequent extraintestinal infection; for example, oropharyngeal colonization may lead to pneumonia, and colonic/perineal colonization may lead to urinary tract infection (UTI). The use of ampicillin or amoxicillin was associated with an increased risk of subsequent infection due to the hypervirulent pathotype of *Klebsiella pneumoniae* in Taiwan; this association suggests that changes in the quantity or prevalence of colonizing bacteria may significantly influence the risk of infection. *Serratia*, *Enterobacter*, and, less commonly, *Citrobacter* infection may be acquired directly through a variety of contaminated infusates (e.g., medications, blood products, non–U.S. Food and Drug Administration [FDA]-approved stem cell products). A multistate outbreak of highly resistant *Serratia* due to contaminated eyedrops has occurred. *Edwardsiella* infections are acquired through freshwater and marine environment exposures and are most common in Southeast Asia.

STRUCTURE AND FUNCTION

Enteric GNB possess an extracytoplasmic outer membrane consisting of a lipid bilayer with associated proteins, lipoproteins, and polysaccharides (capsule, lipopolysaccharide). The outer membrane interfaces with the external environment, including the human host. A variety of components of the outer membrane are critical determinants in pathogenesis (e.g., capsule, lipopolysaccharide) and antimicrobial resistance (e.g., permeability barrier, efflux pumps). In addition, secreted products play an important role in both host infection (e.g., iron acquisition molecules) and environmental niche survival and colonization (e.g., type VI secretion systems).

PATHOGENESIS

Multiple bacterial virulence factors are required for the pathogenesis of infections caused by GNB. Possession of specialized virulence genes defines pathogens and enables them to infect the host efficiently. Hosts and their cognate pathogens have been co-adapting throughout evolutionary history. During the host–pathogen "chess match" over time, various and redundant strategies have emerged in both the pathogens and their hosts **(Table 166-1)**.

Intestinal pathogenic (diarrheagenic) mechanisms are discussed below. The members of the order *Enterobacterales* that cause extraintestinal infections are primarily extracellular pathogens and therefore share certain pathogenic features. The two principal components of host defense against *Enterobacterales*, regardless of species, are innate immunity (including intact skin and mucosal barriers; the withholding of nutrients; and the activities of complement, antimicrobial peptides, and professional phagocytes) and humoral immunity. Both susceptibility to and severity of infection are increased with dysfunction or deficiencies of these host components. By contrast, the virulence traits of intestinal pathogenic *E. coli*—i.e., the distinctive strains that can cause diarrheal disease—are for the most part different from those of extraintestinal pathogenic *E. coli* (ExPEC) and other GNB that cause extraintestinal infections. This distinction reflects site-specific differences in host environments, defense mechanisms, and physiologic derangements that lead to disease.

A given enterobacterial strain usually possesses multiple adhesins for binding to a variety of host cells (e.g., in *E. coli*: type 1, S, and F1C fimbriae; P pili). Nutrient acquisition (e.g., of iron via

TABLE 166-1 Interactions of Extraintestinal Pathogenic *Escherichia coli* with the Human Host: A Paradigm for Extracellular, Extraintestinal, Gram-Negative Bacterial Pathogens

BACTERIAL GOAL	HOST OBSTACLE	BACTERIAL SOLUTION
Extraintestinal attachment	Flow of urine, mucociliary escalator	Multiple adhesins (e.g., type 1, S, and F1C fimbriae; P pili)
Nutrient acquisition for growth	Nutrient sequestration (e.g., iron via intracellular storage and extracellular scavenging via lactoferrin and transferrin)	Cellular lysis (e.g., hemolysin), multiple mechanisms for competing for iron (e.g., siderophores) and other nutrients
Initial avoidance of host bactericidal activity	Complement, phagocytic cells, antimicrobial peptides	Capsular polysaccharide, lipopolysaccharide
Dissemination (within host and between hosts)	Intact tissue barriers	Irritant tissue damage resulting in increased excretion (e.g., toxins such as hemolysin), invasion of brain endothelium
Late avoidance of host bactericidal activity	Acquired immunity (e.g., specific antibodies), treatment with antibiotics	Cell entry, acquisition of antimicrobial resistance

siderophores) requires many genes that are necessary but not sufficient for pathogenesis. The ability to resist the bactericidal activity of complement and phagocytes in the absence of antibody (e.g., as conferred by capsule or the O antigen component of lipopolysaccharide) is one of the defining traits of an extracellular pathogen. Tissue damage (e.g., as mediated by *E. coli* hemolysin) may facilitate nutrient acquisition and spread within the host. Without doubt, many important virulence genes await identification.

The ability to induce septic shock is another defining feature of these genera. GNB are the most common causes of this potentially lethal syndrome. Pathogen-associated molecular pattern molecules (PAMPs; e.g., the lipid A moiety of lipopolysaccharide) stimulate a proinflammatory host response via pattern recognition receptors (e.g., Toll-like or C-type lectin receptors) that activate host defense signaling pathways; if overly exuberant, this response results in shock **(Chap. 315)**. Direct bacterial damage of host tissue (e.g., by toxins) or collateral damage from the host response can result in the release of damage-associated molecular pattern molecules (DAMPs; e.g., HMGB1) that can propagate a detrimental proinflammatory host response.

Many antigenic variants (serotypes) exist in most genera of GNB. For example, *E. coli* has >150 O (somatic) antigens, 80 K (capsular) antigens, and 53 H (flagellar) antigens. This antigenic variability, which permits immune evasion and allows recurrent infection by different strains of the same species, has impeded vaccine development **(Chap. 129)**.

INFECTIOUS SYNDROMES

Depending on both the host and the pathogen, GNB can infect nearly every organ or body cavity. *E. coli* can cause either intestinal or extraintestinal infection, depending on the pathotype, and *Edwardsiella tarda* can cause both intestinal and extraintestinal infection. *Klebsiella* causes primarily extraintestinal infection, but a toxin-producing variant of *Klebsiella oxytoca* has been associated with hemorrhagic colitis, and *Providencia alcalifaciens* and *Escherichia albertii* have been associated with gastroenteritis.

E. coli and—to a lesser degree—*Klebsiella* account for most extraintestinal infections due to GNB. These species (for *K. pneumoniae*, primarily its hypervirulent pathotype) are the most virulent pathogens within this group, as demonstrated by their ability to cause severe infections in healthy, ambulatory hosts from the community. However, the other genera of GNB are also important extraintestinal pathogens, especially among LTCF residents and hospitalized patients, in large part because of the intrinsic or acquired antimicrobial resistance of these organisms and the increasing number of individuals with compromised host defenses. The mortality rate is substantial in many GNB infections and correlates with severity of illness, underlying host status, and in some cases the antimicrobial resistance of the infecting pathogen, which can result in suboptimal therapy. Especially problematic are pneumonia, sepsis, and septic shock (arising from any site of infection), for which the associated mortality rates are 20–60%.

DIAGNOSIS

Isolation of GNB from sterile sites almost always implies infection, whereas their isolation from nonsterile sites, particularly open wounds, the respiratory tract, and urine in the presence of an indwelling catheter, requires careful clinical correlation to differentiate colonization from infection. Clinical microbiology laboratories are increasingly replacing identification by biochemical tests with newer diagnostic methodologies such as matrix-assisted laser desorption–ionization–time-of-flight mass spectrometry (MALDI-TOF-MS), nucleic acid amplification tests (NAATs) and sequencing, and immunoassays to enhance the sensitivity, accuracy, and rapidity of reporting on pathogen identification and resistance genes. This information can be used to increase the timeliness of initiation and/or the accurate selection of empirical antimicrobial therapy, thereby improving outcomes. These new diagnostic modalities have also resulted in the identification from clinical specimens of unfamiliar species (e.g., *K. grimontii*, *Enterobacter hormaechei*). It is best to assume such isolates possess a similar pathogenic potential as their more familiar counterparts until more data become available.

TREATMENT

Principles Guiding Treatment in the Era of Increasing Antimicrobial Resistance

(See also Chap. 149) Initiation of appropriate empirical antimicrobial therapy early in the course of infections due to GNB (particularly the more serious ones) leads to improved outcomes. The ever-increasing prevalence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) GNB; the lag between published and current resistance rates; and variations in antimicrobial susceptibility by species, geographic location, regional antimicrobial use, and hospital site (e.g., intensive care units [ICUs] vs wards) necessitate familiarity with evolving patterns of antimicrobial resistance for the selection of appropriate empirical therapy.

Patient factors predictive of resistance in a given isolate include recent antimicrobial use, a health care association (e.g., recent or ongoing hospitalization, dialysis, residence in an LTCF, transplant, hematologic malignancy), or international travel (e.g., to Asia, Latin America, Africa, Eastern Europe). Of concern are an increasing number of reports of resistant *Enterobacterales* causing infections in ambulatory patients without known risk factors.

In this era of increasing antimicrobial resistance, it is critical to culture the primary site of infection before initiating antimicrobial therapy and, for systemically ill patients, to obtain blood cultures. In vitro testing may not always detect antimicrobial resistance; therefore, it is important to assess the patient's clinical response to treatment. Moreover (see discussion of AmpC β-lactamases below), resistance may emerge during therapy. In addition, drainage of abscesses, resection of necrotic tissue, and removal of infected foreign bodies, sometimes referred to collectively as "source control," are often required for cure.

For appropriately selected patients, it may be prudent initially, pending antimicrobial susceptibility results, to use two potentially active agents to increase the likelihood that at least one agent will be active against the patient's organism. If broad-spectrum treatment has been initiated, it is important to switch to the most appropriate narrower-spectrum agent once antimicrobial susceptibility results become available. Such responsible antimicrobial stewardship should help disrupt the ever-escalating cycle of selection for increasingly resistant bacteria, plus decrease the likelihood of *Clostridioides difficile* infection, decrease costs, and maximize the useful longevity of available antimicrobial agents. Likewise, it is important to avoid treatment of patients who are colonized but not infected (e.g., who have a positive sputum culture without evidence of pneumonia, or a positive urine culture without clinical manifestations of UTI).

At present, the most reliably and broadly active antimicrobial agents in vitro against *Enterobacterales* are the carbapenems (excepting imipenem, to which *Proteeae* [*Proteus*, *Morganella*, *Providencia*] are intrinsically resistant); the aminoglycoside amikacin (excepting the *Proteeae*); the fourth-generation cephalosporin cefepime; the β-lactamase inhibitor combination agents piperacillin-tazobactam, ceftolozane-tazobactam, ceftazidime-avibactam, meropenem-vaborbactam, and imipenem/cilastatin-relebactam; and the novel cephalosporin-siderophore cefiderocol. A limitation of imipenem/cilastatin-relebactam; the tetracycline derivatives tigecycline, omadacycline, and eravacycline; and the polymyxins B and E (colistin) (which are otherwise very active) is their poor activity against *Proteeae* and *Serratia*. Furthermore, the tetracycline derivatives achieve suboptimal concentrations at several anatomic sites (including urine and blood). Clinical data are limited for cefiderocol outside of UTIs and hospital-acquired ventilator-associated bacterial pneumonias; thus, caution is in order for serious infections at other sites.

The number of antimicrobial agents active against certain strains of *Enterobacterales* is shrinking, and truly pandrug-resistant GNB exist. Accordingly, the currently available antimicrobial drugs must be used judiciously. Extensive resistance to available agents may leave the clinician with few or no ideal therapeutic options. However, use of a regimen that considers the site of infection, achievable drug levels at that site (e.g., higher concentrations of many agents in urine), and pharmacodynamically guided administration strategies (e.g., prolonged infusion of β-lactam agents to maintain drug levels above the minimal inhibitory concentration [MIC]) may increase the chance for a successful outcome. Point-of-care, NAAT-based identification of resistance mechanisms in GNB is becoming available and will enable a strain-specific, patient-specific, precision medicine–based treatment approach that would be predicted to improve outcome.

GNB are commonly involved in polymicrobial infections, in which the role of each individual pathogen is uncertain **(Chap. 182)**. Although some GNB are more pathogenic than others, it is usually prudent, if possible, to design an antimicrobial regimen active against all the GNB identified, because each is typically capable of pathogenicity in its own right. For patients treated initially with a broad-spectrum empirical regimen, the regimen should be de-escalated as expeditiously as possible once susceptibility results are known and the patient has responded to therapy.

Treatment duration is best individualized based on underlying host status and site of infection. However, for selected non–critically ill patients with source control and a satisfactory clinical response to therapy, 7 days of treatment may suffice for many infections.

ANTIMICROBIAL TREATMENT AND RESISTANCE MECHANISMS

The most common resistance mechanisms possessed by *Enterobacterales* are summarized in **Table 166-2**. Enzymatic hydrolysis (e.g., β-lactamases, of which >3000 variants have been described) and modification of antimicrobials are the major mediators of resistance in GNB and will be discussed below. Importantly, it is becoming increasingly recognized that MDR and XDR GNB often possess multiple plasmids and genes that encode for multiple β-lactamases.

Broad-spectrum β-lactamases mediate resistance to many penicillins and first-generation cephalosporins and are frequently expressed in enteric GNB. These enzymes are inhibited by all available β-lactamase inhibitors (e.g., clavulanate, sulbactam, tazobactam, avibactam, relebactam, vaborbactam). In their wild-type form, they do not hydrolyze third- and fourth-generation cephalosporins or cephamycins (e.g., cefoxitin).

Extended-spectrum β-lactamases (ESBLs) are modified broad-spectrum enzymes that hydrolyze third-generation cephalosporins, aztreonam, and (in some instances) fourth-generation cephalosporins, in addition to the drugs hydrolyzed by broad-spectrum β-lactamases. GNB that produce ESBLs may also exhibit porin mutations that result in decreased uptake of relevant β-lactam agents (cephalosporins, β-lactam/β-lactamase inhibitor combinations, and carbapenems), further reducing susceptibility to these agents. The prevalence of acquired ESBL production, particularly of CTX-M-type enzymes, is increasing in GNB worldwide, largely due to the presence of the corresponding genes on transferable plasmids, which also variably confer or are associated with resistance to fluoroquinolones, trimethoprim-sulfamethoxazole (TMP-SMX), aminoglycosides, tetracyclines, and (more recently) fosfomycin. To date, ESBLs are most prevalent in *E. coli* (especially ST131), *K. pneumoniae*, and *K. oxytoca*, but these enzymes can occur in all species of *Enterobacterales*. The approximate regional prevalence of ESBL-producing GNB currently follows a descending gradient as follows: China > Eastern Europe > other parts of Asia (e.g., India) > Latin America and Africa > Western Europe, the United States, Canada, and Australia. Travel to high-prevalence regions increases the likelihood of colonization with these strains. The incidence of community-acquired infections due to ESBL-producing *Enterobacterales* has increased worldwide, including in the United States.

TABLE 166-2 Common Antimicrobial Resistance Mechanisms Possessed by the *Enterobacterales*

MECHANISM	ANTIMICROBIALS MOST SIGNIFICANTLY AFFECTED	COMMON MEDIATORS OF RESISTANCE
Efflux	Tetracyclines, fluoroquinolones (FQ)	Efflux pumps
Decreased permeability	Fosfomycin	Alterations in uptake system
Target site alteration or overproduction	FQ, trimethoprim-sulfamethoxazole (TMP-SMX), and polymyxins	DNA gyrase or topoisomerase IV for FQ; enzymes for folic acid synthesis for TMP-SMX Lipid A for polymyxins
Enzymatic hydrolysis of antimicrobials	Penicillins, cephalosporins, cephamycins, carbapenems	Broad-spectrum β-lactamases (e.g., TEM, SHV) ESBLs (e.g., CTX-M, modified TEM and SHV) AmpC β-lactamases Carbapenemases (e.g., serine-based KPC, OXA; metallo-based NDM, VIM, IMP)
Enzymatic modification of antimicrobials	Aminoglycosides	AAC, ANT, APH

Abbreviations: AAC, *N*-acetyltransferases; ANT, *O-adenylyltransferases*; APH, *O*-phosphotransferases; CTX, cefotaxime β-lactamase; ESBL, extended-spectrum β-lactamase; IMP, active on imipenem; KPC, *Klebsiella pneumoniae* carbapenemase; NDM, New Delhi metallo-β-lactamase; OXA, oxacillinase; SHV, sulfhydryl reagent variable β-lactamase; TEM, Temoniera β-lactamase; VIM, Verona integron-mediated metallo-β-lactamase.

Carbapenems are the most reliably active β-lactam agents against ESBL-producing strains. Piperacillin-tazobactam, when active in vitro, has been used as a carbapenem-sparing alternative, but recent data from the MERINO trial do not support its use for bloodstream infections. Ceftazidime-avibactam, ceftolozane-tazobactam (less active against *Klebsiella*, *Enterobacter*, and *Citrobacter*) are also active against most ESBL-producing strains but have limited clinical data that support potential utility. The roles for tigecycline, eravacycline, and omadacycline are unclear despite these agents' excellent in vitro activity against most *Enterobacterales*; however, they are inactive against *Proteus*, *Morganella*, *Providencia*, and *Serratia*.

Oral options for the treatment of ESBL-producing strains are very limited. Fosfomycin, nitrofurantoin (for *E. coli*, 75–90% susceptible), pivmecillinam (recently approved in the United States), and omadacycline are the most reliably active agents. Older tetracyclines (e.g., doxycycline and minocycline) also are often active, although urine levels may be insufficient and clinical experience with gram-negative infections is limited.

AmpC β-lactamases, when induced or stably derepressed to high levels of expression, confer resistance to the same substrates as do ESBLs plus to the cephamycins (e.g., cefoxitin and cefotetan), except to the fourth-generation cephalosporins. The genes encoding these enzymes are primarily chromosomal and therefore may not exhibit the linked resistance to TMP-SMX, aminoglycosides, and tetracyclines that is common with ESBLs. These enzymes are problematic for the clinician: resistance may develop during therapy with third-generation cephalosporins and result in clinical failure, particularly in the setting of bacteremia. Although chromosomal AmpC β-lactamases are present in nearly all members of the order *Enterobacterales* (with the notable exceptions of *K. pneumoniae*, *K. oxytoca*, and *Proteus mirabilis*), the risk of clinically significant induction of high-level expression or selection of stably derepressed mutants with cephalosporin treatment is not uniform across species, being greatest with *Enterobacter cloacae*, *Klebsiella* (formerly *Enterobacter*) *aerogenes*, *Citrobacter freundii*, and *Hafnia alvei*, and less with *Serratia marcescens*, *Providencia*, and *Morganella morganii.* In addition, rare strains of *E. coli*, *K. pneumoniae*, and other *Enterobacterales* have acquired plasmids that contain AmpC β-lactamase genes.

For AmpC-producing strains, carbapenems are an appropriate treatment option, especially for severely ill patients. Meta-analyses support piperacillin-tazobactam as a possible option. The fourth-generation cephalosporin cefepime may be an appropriate option if the concomitant production of an ESBL can be excluded (a task that currently exceeds the capability of most clinical microbiology laboratories) and source control is achieved. Ceftazidime-avibactam and cefiderocol are active in vitro, but clinical data are limited. Other carbapenem-sparing alternatives to consider if isolates are susceptible in vitro include fluoroquinolones, TMP-SMX, and aminoglycosides. Tigecycline, eravacycline, and omadacycline are active in vitro (except against *Proteus*, *Morganella*, *Providencia*, and *Serratia*).

Carbapenemases of Ambler class A (serine-β-lactamases; e.g., *K. pneumoniae* carbapenemase [KPC]) and class B (metallo-β-lactamases [MBLs]; e.g., New Delhi metallo-β-lactamase [NDM], Verona integron-mediated metallo-β-lactamase [VIM], imipenemase [IMP]) confer resistance to the same drugs as do ESBLs, plus to cephamycins and carbapenems. By contrast, Ambler class D carbapenemases (serine-β-lactamases; e.g., oxacillinase-48 [OXA-48]) hydrolyze carbapenems and penicillins, but they have minimal activity against extended-spectrum cephalosporins. As with ESBLs, carbapenemase-encoding genes may be present on transferable plasmids, which often encode linked resistance to fluoroquinolones, TMP-SMX, tetracyclines, and aminoglycosides. Transposon-mediated spread (e.g., Tn*4401* for KPC) also is important. Although all major carbapenemases have been described around the globe, KPC is most common in the Americas, NDM in Asia, and OXA in Europe. Asymptomatic intestinal carriage of producing bacteria may facilitate spread.

Carbapenemase-producing *Enterobacterales* (CPE) are most prevalent in *K. pneumoniae*, followed by *Enterobacter* spp. and *E. coli*, but have been described in nearly all members of the order. *M. morganii*, *Proteus*, and *Providencia* exhibit intrinsic low-level imipenem resistance. A variety of genotypic and phenotypic methods can detect carbapenemase genes or activity, which could inform epidemiologic surveillance, infection control efforts, antimicrobial stewardship, and treatment decisions, especially if susceptibility data for selected agents are not available.

For the treatment of infections due to *Enterobacterales* that produce class A or D carbapenemases (serine-β-lactamases; KPC, OXA), ceftazidime-avibactam is emerging as a first-line agent particularly for bacteremia, but suboptimal efficacy has been observed with pneumonia and in patients on renal replacement therapy, and resistance has developed in up to 10% of cases. Clinical success against KPC-producing CRE has also been reported for meropenem-vaborbactam and, to a lesser extent, imipenem/cilastatin-relebactam; unlike ceftazidime-avibactam, however, neither of these agents is active against OXA-producing CRE. Ceftazidime, cefepime, and aztreonam are active against OXA-48-producing CRE, unless other enzymes such as ESBL an AmpC are co-produced.

Treatment of infections due to class B MBL–producing CRE is more challenging. The polymyxins B and E currently constitute one of the last lines of defense against strains that produce MBLs (e.g., NDM). However, these agents' nephrotoxicity and neurotoxicity potential, their limited clinical efficacy, and the recent emergence of the polymyxin resistance threaten their utility.

Aztreonam is stable against MBLs but is hydrolyzed by ESBLs and AmpC β-lactamases, which are often coproduced in XDR strains. Ongoing clinical trials are assessing aztreonam plus avibactam, a promising combination with in vitro activity against class A, B, and D enzymes, for the treatment of CRE strains that produce MBLs like NDM. A currently available workaround involving approved drugs is co-administration of ceftazidime-avibactam and aztreonam; avibactam protects aztreonam from hydrolysis from ESBLs and AmpC β-lactamases.

Cefiderocol is active in vitro against most strains producing KPC, MBLs, and OXA-48 (i.e., classes A, B, and D enzymes); clinical trials suggest that it may be as efficacious for serious CRE infections as the standard of care, but real-world data are limited. Although tigecycline, eravacycline, and omadacycline are active in vitro, pharmacokinetic-pharmacodynamic limitations exist, and along with the polymyxins, they exhibit poor activity against the tribe *Proteeae* and *Serratia*. Aminoglycosides, especially amikacin, may have some utility for combination therapy. Fosfomycin is often active in vitro, but clinical data in the treatment of serious infections due to CPE are limited and resistance may develop with monotherapy.

Carbapenem resistance in the absence of carbapenemases can occur in the presence of ESBLs or AmpC β-lactamase production in combination with porin mutations (non-CP-CRE); however, most laboratories will not be able to differentiate CPE from non-CP-CRE. The non-CP-CRE phenotype is most commonly seen in *E. coli* and *Enterobacter* spp. In general, resistance to noncarbapenem antimicrobial classes is less, but data are limited on the optimal management approach for non-CP-CRE.

Resistance to classic β-Lactamase inhibitors is an uncommon (4% of *E. coli/K. pneumoniae* blood isolates) but increasingly recognized phenotype that is characterized by resistance to β-lactamase inhibitors but not to third-generation cephalosporins. This mechanism of resistance is distinct from production of ESBLs, AmpC β-lactamases, and carbapenemases, and it is still being delineated. Limited evidence suggests that ceftriaxone is an appropriate treatment option for such strains. Resistance to newer β-lactamase inhibitors, especially avibactam, is increasingly reported in KPC-producing CPE and is due to mutations leading to structural modifications of the KPC enzyme. These strains remain susceptible to meropenem-vaborbactam.

Fluoroquinolone resistance is usually due to alterations in or protection of the target sites in DNA gyrase and topoisomerase IV, with or without decreased permeability and active efflux. Fluoroquinolone resistance is increasingly prevalent among GNB and is associated with resistance to other antimicrobial classes; for example, 20–80% of ESBL-producing enteric GNB are also resistant to fluoroquinolones. At present, fluoroquinolones should be considered unreliable as empirical therapy for GNB infections in critically ill patients.

Aminoglycoside resistance in *Enterobacterales* is conferred via enzymatic modification by *N*-acetyltransferases, *O*-adenylyltransferases, or *O*-phosphotransferases, which in turn affects ribosomal binding. Amikacin is less affected by these transferases than gentamicin and tobramycin and therefore is generally more active. A yet uncommon resistance mechanism involves acquired 16S ribosomal RNA methyltransferases, which prevent all parenterally administered aminoglycosides from binding to their target ribosomes. To date, these methyltransferases are most common in strains that produce MBLs (e.g., NDM).

PREVENTION

(See also Chap. 147) Certain measures are broadly applicable for decreasing infection risk. Antimicrobial stewardship programs should be instituted to facilitate appropriate antimicrobial use, which will minimize the development of resistance. Diligent adherence to hand hygiene protocols by health care personnel and cleaning/disinfection or single-patient use of objects that come into contact with patients (e.g., stethoscopes and blood pressure cuffs) are essential. Indwelling devices (e.g., urinary and intravascular catheters) should be used only when necessary and inserted according to an appropriate protocol; protocols for daily-use evaluation and prompt removal should be implemented. Multiuse medication vials should be avoided if possible. Oral application of chlorhexidine decreases the incidence of pneumonia among patients on ventilators. Increasing data support the implementation of universal decolonization (e.g., chlorhexidine bathing) to prevent infection in ICU patients or nursing home residents. The public health threat from CRE has resulted in additional recommendations, especially for carbapenemase-producing CRE, which are an even greater concern. These recommendations include contact precautions for patients colonized or infected with CRE, notification to the receiving facility from facilities transferring such a patient, and daily environmental cleaning. Screening of contacts and active surveillance for these bacteria also may be appropriate.

ESCHERICHIA COLI INFECTIONS

All *E. coli* strains share a core genome of ~2,000 genes. In contrast, an *E. coli* strain's ability to cause infection and the nature of such infections are defined largely by accessory (i.e., noncore, nonessential) genes that encode various virulence factors. The composition of the *E. coli* accessory genome is continuously in flux, as demonstrated by the recent evolution of Shiga toxin–producing enteroaggregative *E. coli.*

COMMENSAL STRAINS

Commensal *E. coli* variants are an important constituent of the normal intestinal microbiota that confer benefits to the host (e.g., resistance to colonization with pathogenic organisms). Such strains generally lack the specialized virulence traits that enable extraintestinal and intestinal pathogenic *E. coli* strains to cause disease outside and within the gastrointestinal tract, respectively. However, even commensal *E. coli* strains can be involved in extraintestinal infections in the presence of an aggravating factor, such as a foreign body (e.g., a urinary catheter), host compromise (e.g., local anatomic or functional abnormalities [including urinary or biliary tract obstruction] or systemic immunocompromise), or an inoculum that is large or contains a mixture of bacterial species (e.g., fecal contamination of the peritoneal cavity).

EXTRAINTESTINAL PATHOGENIC STRAINS

ExPEC strains are the most common enteric GNB to cause community-acquired and health care–associated bacterial infections. The emerging propensity of these strains to acquire new mechanisms of antimicrobial resistance (e.g., FQ resistance mutations, ESBLs, carbapenemases) poses novel challenges in managing ExPEC infection. Several ExPEC clonal groups (e.g., sequence types [STs] ST131, ST95, ST69, ST73) are recognized to have undergone global dissemination. The mechanisms underlying the epidemiologic success of such disseminated lineages remain an area of active investigation. In the case of ST131, efficient human-to-human transmission followed by colonization and long-term persistence within the intestinal microbiota appears to be a critical factor. Although acquisition of ESBL-producing *E. coli* from the food chain has been described, this appears to occur relatively uncommonly.

Like commensal *E. coli* (but unlike intestinal pathogenic *E. coli*), ExPEC strains are often found in the intestinal microbiota of healthy individuals and, except for rare chimeric ExPEC/intestinal pathogenic *E. coli* strains, do not cause gastroenteritis in humans. Entry from their site of colonization (e.g., the colon, vagina, or oropharynx) into a normally sterile extraintestinal site (e.g., the urinary tract, peritoneal cavity, or lungs) is the rate-limiting step for infection. ExPEC strains have acquired accessory genes encoding diverse virulence factors that enable the bacteria to cause infections outside the gastrointestinal tract in both normal and compromised hosts (Table 166-1). These virulence genes define ExPEC and, for the most part, are distinct from the virulence genes that enable intestinal pathogenic strains to cause diarrheal disease **(Table 166-3)**. All age groups, all types of hosts, and nearly all organs and anatomic sites are susceptible to infection by ExPEC. Even previously healthy hosts can become severely ill or die when infected with ExPEC; however, adverse outcomes are more common among hosts with comorbid illnesses and host defense abnormalities. The diversity and the medical and economic impact of ExPEC infections are evident from consideration of the following specific syndromes.

Extraintestinal Infectious Syndromes • URINARY TRACT INFECTION The urinary tract is the site most frequently infected by ExPEC. UTI is an exceedingly common infection among ambulatory patients, accounting for 1% of ambulatory care visits in the United States and second only to lower respiratory tract infection among infections responsible for hospitalization. UTIs are best considered by clinical syndrome (e.g., cystitis, pyelonephritis, catheter-associated UTI) and within the context of specific hosts (e.g., premenopausal women, immunocompromised hosts; **Chap. 140**). *E. coli* is the single most common pathogen for all UTI syndrome/host group combinations.

TABLE 166-3 Intestinal Pathogenic *Escherichia coli*

PATHOTYPE	EPIDEMIOLOGY	CLINICAL SYNDROME[a]	DEFINING MOLECULAR TRAIT	RESPONSIBLE GENETIC ELEMENT[b]
STEC/EHEC/ST-EAEC	Food, water, person-to-person; all ages, industrialized countries	Hemorrhagic colitis, hemolytic-uremic syndrome	Shiga toxin	Lambda-like Stx1- or Stx2-encoding bacteriophage
ETEC	Food, water; young children in and travelers to developing countries	Traveler's diarrhea	Heat-stable and labile enterotoxins, colonization factors	Virulence plasmid(s)
EPEC	Person-to-person; young children and neonates in developing countries	Watery diarrhea, persistent diarrhea	Localized adherence, attaching and effacing lesion on intestinal epithelium	EPEC adherence factor plasmid pathogenicity island (locus for enterocyte effacement [LEE])
EIEC	Food, water; children in and travelers to developing countries	Watery diarrhea, occasionally dysentery	Invasion of colonic epithelial cells, intracellular multiplication, cell-to-cell spread	Multiple genes contained primarily in a large virulence plasmid
EAEC	?Food, water; children in and travelers to developing countries; all ages, industrialized countries	Traveler's diarrhea, acute diarrhea, persistent diarrhea	Aggregative/diffuse adherence, virulence factors regulated by AggR	Chromosomal or plasmid-associated adherence and toxin genes

[a] Classic syndromes; see text for details on disease spectrum. [b] Pathogenesis involves multiple genes, including genes in addition to those listed.

Abbreviations: EAEC, enteroaggregative *E. coli*; EHEC, enterohemorrhagic *E. coli*; EIEC, enteroinvasive *E. coli*; EPEC, enteropathogenic *E. coli*; ETEC, enterotoxigenic *E. coli*; ST-EAEC, Shiga toxin–producing enteroaggregative *E. coli*; STEC, Shiga toxin–producing *E. coli.*

Each year in the United States, *E. coli* causes 80–90% of the estimated 6–8 million episodes of cystitis that occur in ambulatory, premenopausal women with an anatomically and functionally normal urinary tract (i.e., uncomplicated cystitis). Furthermore, 20% of women with an initial cystitis episode develop frequent recurrences.

Uncomplicated cystitis, the most common acute UTI syndrome, is characterized by dysuria, urinary frequency and urgency, and suprapubic pain. Progression to more severe infection is rare; the natural history is slow spontaneous symptom resolution, which antimicrobial therapy hastens. Fever and/or back pain suggest progression to pyelonephritis. Even when pyelonephritis is treated effectively, fever may take 5–7 days to resolve completely. Persistently elevated or increasing fever, flank pain, and neutrophil counts should prompt evaluation for intrarenal or perinephric abscess and/or obstruction. Pyelonephritis uncommonly causes renal parenchymal damage and loss of renal function, primarily in association with urinary obstruction, which can be preexisting or, rarely, occurs de novo in diabetic patients who develop renal papillary necrosis due to kidney infection. Pregnant women are at unusually high risk for developing pyelonephritis, which can adversely affect the outcome of pregnancy. As a result, prenatal screening for and treatment of asymptomatic bacteriuria during pregnancy are standard. Prostatic infection (prostatitis), a potential complication of UTI in men, can present either acutely (severe), which is rare, or in a chronic manner (recurrent cystitis), which is much more common. Acute pyelonephritis, acute prostatitis, and other systemic illnesses due to UTI can be designated collectively as *urosepsis*, *febrile UTI*, or *systemic UTI*, and may or may not be accompanied by bacteremia. The diagnosis and treatment of UTI, as detailed in **Chap. 140**, should be tailored to the individual host, the nature and site of infection, and local patterns of antimicrobial susceptibility.

ABDOMINAL AND PELVIC INFECTION The abdomen/pelvis is the second most common site of extraintestinal infection due to *E. coli*. A wide variety of clinical syndromes occur in this location, including acute peritonitis secondary to fecal contamination, spontaneous bacterial peritonitis, dialysis-associated peritonitis, diverticulitis, appendicitis, intraperitoneal or visceral abscesses (hepatic, pancreatic, splenic), infected pancreatic pseudocysts, and septic cholangitis and/or cholecystitis. In intraabdominal infections, *E. coli* can be isolated either alone or, as occurs more often, in combination with other facultative and/or anaerobic members of the intestinal microbiota **(Chap. 137)**.

PNEUMONIA *E. coli* is not usually considered an important cause of pneumonia **(Chap. 131)**. Indeed, enteric GNB account for only 1–3% of cases of community-acquired pneumonia, in part because these organisms colonize the oropharynx only transiently in a minority of healthy individuals. However, rates of oral colonization with *E. coli* and other GNB increase with severity of illness and antibiotic use. Consequently, GNB are a more common cause of pneumonia among residents of LTCFs and of hospital-acquired pneumonia **(Chap. 147)**, particularly among postoperative and ICU patients (e.g., ventilator-associated pneumonia).

Pulmonary infection is usually acquired by small-volume aspiration but occasionally occurs via hematogenous spread, in which case multifocal nodular infiltrates can be seen. Tissue necrosis, probably due in part to bacterial cytotoxins, is common. Despite significant institutional variation, *E. coli* is generally the third or fourth most commonly isolated type of GNB in hospital-acquired pneumonia, accounting for 5–8% of episodes in both U.S.-based and Europe-based studies. Regardless of the host, pneumonia due to ExPEC is a serious disease, with high crude and attributable mortality rates (20–60% and 10–20%, respectively).

MENINGITIS **(See also Chap. 143)** *E. coli* is one of the leading causes of neonatal meningitis, together with group B *Streptococcus*. Most *E. coli* strains that cause neonatal meningitis possess the K1 capsular antigen and derive from a limited number of meningitis-associated clonal groups (ST95, ST59, ST62). Ventriculomegaly occurs commonly. After the first month of life, *E. coli* meningitis is uncommon and usually accompanies surgical or traumatic disruption of the meninges or hepatic cirrhosis. In patients with cirrhosis who develop meningitis, the meninges are presumably seeded due to poor hepatic clearance of portal vein bacteremia.

CELLULITIS/MUSCULOSKELETAL INFECTION *E. coli* contributes frequently to infections of decubitus ulcers and occasionally to infections of lower-extremity ulcers and wounds in diabetic patients and other hosts with neurovascular compromise. Osteomyelitis secondary to contiguous spread can occur in these settings. *E. coli* also causes cellulitis or infections of burn sites and surgical wounds (accounting for ~10% of surgical site infections), particularly when the infection originates close to the perineum. *E. coli* causes hematogenously acquired osteomyelitis, especially of vertebral discs and bodies, accounting for up to 10% of cases in some series **(Chap. 136)**. *E. coli* occasionally causes orthopedic device–associated infection or septic arthritis and rarely causes hematogenous myositis. Myositis or fasciitis of the thigh due to *E. coli* should prompt an evaluation for an abdominal source with contiguous spread.

ENDOVASCULAR INFECTION Despite being one of the most common causes of bacteremia, *E. coli* rarely seeds native heart valves. When the organism does infect native valves, it usually does so in the setting of prior valvular disease. *E. coli* infections of aneurysms, the portal vein (*pylephlebitis*), and vascular grafts are uncommon.

MISCELLANEOUS INFECTIONS *E. coli* can cause infection in nearly every organ and anatomic site. It occasionally causes postoperative mediastinitis or complicated sinusitis and uncommonly causes endophthalmitis, ecthyma gangrenosum, or brain abscess.

BACTEREMIA *E. coli* bacteremia can arise from infection at any extraintestinal site. In addition, *E. coli* bacteremia can arise from percutaneous intravascular devices, transrectal prostate biopsy, and the increased intestinal mucosal permeability seen in neonates and patients with advanced cirrhosis, neutropenia, chemotherapy-induced mucositis, trauma, and extensive burns. *E. coli* bacteremia due to an ESBL-producing strain also has been reported after fecal microbiota transplant in patients with increased mucosal permeability. Roughly equal proportions of *E. coli* bacteremia cases originate in the community and in health care settings. Isolation of *E. coli* from the blood is almost always clinically significant and may be accompanied by the sepsis syndrome (dysfunction of at least one organ or system) or septic shock **(Chap. 315)**.

The urinary tract is the most common source for *E. coli* bacteremia, accounting for one-half to two-thirds of episodes. Bacteremia from a urinary tract source is particularly common among patients with pyelonephritis, urinary tract obstruction, or urinary instrumentation in the presence of infected urine. The abdomen is the second most common source, accounting for ~25% of episodes. Although many of these episodes result from biliary obstruction (stones, tumor) and overt bowel disruption, which typically are readily apparent, some abdominal sources (e.g., abscesses) are remarkably silent clinically and require identification via imaging studies (e.g., computed tomography). Therefore, especially given the high prevalence of asymptomatic bacteriuria among elderly and functionally compromised individuals, the physician should be cautious in attributing *E. coli* bacteremia to a urinary source in the absence of characteristic signs and symptoms of UTI. Soft tissue, bone, pulmonary infections, and intravascular catheter infections are other sources of *E. coli* bacteremia.

Diagnosis Strains of *E. coli* that cause extraintestinal infections usually grow both aerobically and anaerobically within 24 h on standard diagnostic media and are identified readily by the clinical microbiology laboratory according to routine biochemical criteria. More than 90% of ExPEC strains are rapid lactose fermenters and are indole-positive. However, MALDI-TOF MS is increasingly replacing biochemical methods.

TREATMENT

Extraintestinal *E. coli* Infections

E. coli does not possess clinically significant intrinsic resistance to antimicrobials; however, increasing acquired resistance is making treatment problematic. Although geographic differences exist, in general, the prevalence of resistance is >20% for ampicillin, amoxicillin-clavulanate, ampicillin-sulbactam, cefazolin, TMP-SMX, and fluoroquinolones, even in community-acquired infections. This resistance precludes empirical use of these agents for serious infections. Travel outside of the United States, prior exposure to an antimicrobial agent, or exposure to a health care setting further increases the likelihood of resistance. Fortunately, >90% of isolates that cause uncomplicated cystitis remain susceptible to nitrofurantoin and fosfomycin.

From 2015 to 2017, the U.S. National Healthcare Safety Network (USNHSN) identified 24% of *E. coli* clinical isolates as ESBL-producers. Higher prevalences are reported from Asia, Eastern Europe, South America, and Africa; prevalence is also greater in isolates from health care settings, especially LTCFs. Unfortunately, community-acquired UTIs caused by *E. coli* strains that produce CTX-M ESBLs are increasingly common. Oral treatment options for ESBL-producers are limited. However, in vitro and limited clinical data indicate that fosfomycin, pivmecillinam, and nitrofurantoin are most active and can be used for cystitis (but not pyelonephritis); omadacycline is an option for pulmonary of soft-tissue infection.

For parenteral therapy of carbapenem-susceptible strains, the most predictably active agents (>90%) include carbapenems, amikacin, ceftazidime-avibactam, ceftolozane-tazobactam, piperacillin-tazobactam, polymyxins, cefiderocol, tigecycline, eravacycline, and omadacycline with the caveat that site-specific concentration and potential efficacy are agent dependent. Treatment of carbapenemase-producing strains is dependent on the class of enzyme produced (see "Carbapenemase" above). Uncertainty exists on the optimal treatment for non-CP-CR *E. coli*.

Empirical treatment decisions for critically ill patients should be dictated by local susceptibility patterns and patient-specific risk factors (1.2% prevalence from the USNHSN 2015–2017 data). Equally important as prompt institution of effective empirical therapy for seriously ill patients is use of appropriate narrower-spectrum agents for definitive therapy whenever possible and avoidance of treatment for patients who are colonized but not infected.

INTESTINAL PATHOGENIC STRAINS

Pathotypes Certain strains of *E. coli* are capable of causing diarrheal disease. **(Other important intestinal pathogens are discussed in Chaps. 138, 139, and 171–174.)** At least in the industrialized world, intestinal pathogenic *E. coli* strains are rarely encountered in the fecal flora of healthy persons, and instead appear to be essentially obligate pathogens. These strains have evolved a special ability to cause enteritis, enterocolitis, and colitis when ingested in sufficient quantities by a naïve host. At least five distinct pathotypes of intestinal pathogenic *E. coli* exist: (1) Shiga toxin–producing *E. coli* (STEC), which includes the subsets enterohemorrhagic *E. coli* (EHEC) and the recently evolved Shiga toxin–producing enteroaggregative *E. coli* (ST-EAEC); (2) enterotoxigenic *E. coli* (ETEC); (3) enteropathogenic *E. coli* (EPEC); (4) enteroinvasive *E. coli* (EIEC); and (5) enteroaggregative *E. coli* (EAEC). Diffusely adherent *E. coli* (DAEC) and cytodetaching *E. coli* are additional putative pathotypes. Lastly, a variant termed *adherent invasive E. coli* (AIEC) has been associated with Crohn disease (although a causal role remains unproven) but does not cause acute diarrheal disease.

Contaminated food and water are the primary transmission vehicles for ETEC, STEC/EHEC/ST-EAEC, EIEC, and EAEC, whereas person-to-person spread (direct or indirect) is the primary transmission route for EPEC and a secondary transmission route for STEC/EHEC/ST-EAEC. Gastric acidity confers some protection against infection; therefore, persons with decreased stomach acid levels are especially susceptible. Humans are the major reservoir for such strains (except for STEC/EHEC, for which bovines are the main carriers); host range appears to be dictated by species-specific attachment factors. Although some overlap exists, each pathotype possesses a distinctive combination of virulence traits that results in a pathotype-specific pathogenic mechanism (Table 166-3). With rare exceptions (e.g. DAEC), these strains are largely incapable of causing disease outside the intestinal tract. Whereas disease due to STEC/EHEC/ST-EAEC occurs primarily in high-income countries, disease due to ETEC, EPEC, and EIEC occurs primarily in low- and middle-income countries in Asia, Africa, and Latin America, and disease due to EAEC occurs globally.

SHIGA TOXIN–PRODUCING *E. COLI* STEC/EHEC/ST-EAEC strains are pathogens that can cause hemorrhagic colitis and the hemolytic-uremic syndrome (HUS). In contrast to other intestinal pathotypes, STEC/EHEC/ST-EAEC causes infections more frequently in high-income countries than in low and middle-income countries (LMICs). Several large outbreaks resulting from the consumption of fresh produce (e.g., lettuce, spinach, sprouts) and of undercooked ground beef have received significant media attention. In addition, a dramatic 2011 outbreak—mainly in Germany—involved an EAEC strain that acquired a Shiga toxin–encoding phage, resulting in a novel genotype, ST-EAEC (O104:H4). This strain was transmitted to the primary cases by sprouted fenugreek seeds, with subsequent human-to-human transmission, and resulted in >4000 cases and 54 deaths.

STEC strains are the fourth most commonly reported cause of bacterial diarrhea in the United States (after *Campylobacter*, *Salmonella*, and *Shigella*). O157:H7 is the most prominent serotype among STEC strains, but many other serogroups have been described, including O6, O26, O45, O55, O91, O103, O111, O113, O121, and O145. Domesticated ruminant animals, particularly cattle and young calves, serve as the major reservoir for STEC/EHEC. Ground or mechanically tenderized beef—the most common food source of STEC/EHEC strains—is often contaminated with intestinal bacteria from the source animals during processing. Furthermore, manure from cattle or other animals (including in the form of fertilizer) can contaminate produce (potatoes, lettuce, spinach, sprouts, fallen fruits, nuts, strawberries), and fecal runoff from manure can contaminate water systems. Dairy products and petting zoos are additional sources of infection.

It is estimated that $<10^2$ colony-forming units (CFU) of STEC/EHEC/ST-EAEC can cause disease. Therefore, not only can low levels of food or environmental contamination (e.g., in water swallowed while swimming) result in disease, but person-to-person transmission (e.g., at day-care centers and in institutions) is an important route for secondary spread. Laboratory-associated infections also occur. Illness due to this group of pathogens peaks in the summer months and occurs both as outbreaks and as sporadic cases.

For STEC/EHEC/ST-EAEC, production of Shiga toxin (Stx2a–g and/or Stx1a,c,d) is a critical factor for occurrence of clinical disease, as demonstrated by the 2011 ST-EAEC outbreak. The *stx* gene is present on chromosomally integrated prophages, and various combinations of *stx* types and subtypes can occur in a given strain. *Shigella dysenteriae* strains that produce the closely related Shiga toxin Stx can also cause hemorrhagic colitis and HUS. Stx2 (especially Stx2a,c,d) appears to be more important than Stx1 in the development of HUS. All Shiga toxins studied to date are multimers; they comprise one A subunit that is enzymatically active and five identical B subunits that mediate binding to globosyl ceramides, which are membrane-associated glycolipids expressed on certain host cells. As in ricin, the Stx A subunit cleaves an adenine from the host cell's 28S rRNA, thereby irreversibly inhibiting ribosomal function (i.e., protein synthesis) and potentially leading to apoptosis.

For full pathogenicity, STEC strains require additional properties such as acid tolerance and epithelial cell adherence. Most disease-causing isolates possess the chromosomal locus for enterocyte effacement (LEE). This pathogenicity island was first described in EPEC strains; it contains genes that mediate adherence to intestinal epithelial cells and a system that subverts host cells by the translocation of bacterial proteins (type III secretion system). EHEC strains make up the subgroup of STEC strains that possess stx_1 and/or stx_2, as well as LEE. By contrast, the 2011 ST-EAEC outbreak strain lacked LEE yet was associated with a higher proportion of patients developing HUS (22%) than the historic average for STEC/EHEC outbreaks (2–8%). Data support the essential role of the 2011 outbreak strain's EAEC-associated virulence factors (e.g., AAF/I fimbriae, serine proteases SigA, SepA) in adherence, increased inflammation, and disruption of the intestinal epithelial barrier, which in turn increased the systemic translocation of Stx2a.

After exposure to STEC/EHEC/ST-EAEC and a 3- to 4-day incubation period, colonization of the colon and perhaps the ileum results in symptoms. Colonic edema and an initial nonbloody secretory diarrhea may progress to the hallmark syndrome of grossly bloody diarrhea (identified by history or examination). Significant abdominal pain and fecal leukocytes are common (70% of cases), whereas fever is not; absence of fever can incorrectly lead to consideration of noninfectious conditions (e.g., intussusception and inflammatory or ischemic bowel disease). Occasionally, infections caused by *C. difficile*, *K. oxytoca* (see "*Klebsiella* Infections," below), *Campylobacter*, and *Salmonella* present in a similar fashion. STEC/EHEC disease is usually self-limited, lasting 5–10 days.

A feared complication of infection with STEC/EHEC strains is HUS, which occurs 2–14 days after diarrhea, most often in young children (estimated to occur in 15% of infected children <10 years of age) or elderly patients. It is estimated that in the United States >50% of all HUS cases—and 90% of HUS cases in children, which is a leading cause of acute renal failure in this latter population—are caused by STEC/EHEC. By contrast, with ST-EAEC infection, HUS occurs more commonly among nonelderly adults, especially young women. HUS is mediated by the systemic translocation of Shiga toxins. Erythrocytes may serve as carriers of Stx to endothelial cells located in the small vessels of the kidney and brain. The subsequent development of thrombotic microangiopathy (perhaps with direct toxin-mediated effects on various nonendothelial cells) commonly produces some combination of fever, hemolytic anemia (hematocrit <30%), thrombocytopenia (<150,000/mm^3), renal failure, and encephalopathy. Stx-mediated complement activation also plays a role in the development of HUS. Although with dialysis support the mortality rate of HUS is <10%, survivors often have persisting renal and neurologic dysfunction.

ENTEROTOXIGENIC *E. COLI* ETEC is a major cause of endemic diarrhea in low- and middle-income countries and is responsible for an estimated 800 million cases annually. After weaning, children in these locales commonly experience several episodes of ETEC infection during the first 3 years of life. The incidence of disease diminishes with age, a pattern that correlates with the development of mucosal immunity to colonization factors (i.e., adhesins).

In industrialized countries, ETEC is the most common agent of traveler's diarrhea, causing 25–75% of cases. The incidence of infection may be decreased by prudent avoidance of potentially contaminated fluids and foods, particularly items that are raw, insufficiently cooked, peeled, or unrefrigerated **(Chap. 130)**. ETEC infection is uncommon in the United States, but outbreaks secondary to consumption of food products imported from endemic areas have occurred. A large inoculum (10^6–10^8 CFU) is needed to produce disease, which usually develops after an incubation period of 12–72 h.

After adherence of ETEC to enterocytes via colonization factors (e.g., CFA/I, CS), disease is mediated, primarily by a heat-labile toxin (LT) and/or a heat-stable toxin (ST), leading to diarrheal disease. Disease is less severe with strains that produce only LT. Both LT and ST cause net fluid secretion via activation of adenylate cyclase and/or guanylate cyclase C (ST) in the jejunum and ileum. The result is watery diarrhea accompanied by cramps.

LT consists of an A and a pentameric B subunit and is structurally and functionally similar to cholera toxin. Strong binding of the B subunit to the GM$_1$ ganglioside on intestinal epithelial cells leads to the intracellular translocation of the A subunit, which functions as an ADP-ribosyltransferase. Mature ST is an 18- or 19-amino-acid secreted peptide that leads to increased intracellular concentrations of cGMP. Characteristically absent in ETEC-mediated disease are histopathologic changes within the small bowel; mucus, blood, and inflammatory cells in stool; and fever.

The disease spectrum of ETEC infection ranges from mild illness to a life-threatening, cholera-like syndrome. Although symptoms are usually self-limited (typically lasting for 3–5 days), infection may result in significant morbidity and mortality (>250,000 deaths annually, mostly from profound volume depletion) when access to health care or suitable rehydration fluids is limited and when small and/or undernourished children are affected.

ENTEROPATHOGENIC *E. COLI* EPEC causes disease primarily in young children, including neonates. The first *E. coli* pathotype recognized as an agent of diarrheal disease, EPEC was responsible for outbreaks of infantile diarrhea (including in hospital nurseries) in industrialized countries in the 1940s and 1950s. At present, EPEC infection is uncommon in high-income countries, but among infants in low- and middle-income countries, it is an important cause of diarrhea (both sporadic and epidemic), often accompanied by vomiting and fever. Breast-feeding diminishes the incidence of EPEC infection. Rapid person-to-person spread may occur.

Symptoms develop after colonization of the small bowel and a brief incubation period (1 or 2 days). Initial localized adherence to enterocytes via type IV bundle-forming pili leads to a characteristic effacement of microvilli, with the formation of cuplike, actin-rich pedestals mediated by factors in the LEE. Diarrhea production is a complex and regulated process in which host cell modulation by a type III secretion system plays an important role. Strains lacking bundle-forming pili have been categorized as atypical EPEC (aEPEC); increasing data support a role for these strains as intestinal pathogens in all age groups and among HIV-infected individuals. Diarrheal stool often contains mucus but not blood. Although EPEC diarrhea is usually self-limited (lasting 5–15 days), it may persist for weeks.

ENTEROINVASIVE *E. COLI* EIEC, a relatively uncommon (or perhaps underrecognized) cause of diarrhea, is rarely identified in the United States, although a few food-related outbreaks have been described. In low- and middle-income countries, sporadic disease is recognized infrequently in children and travelers.

EIEC shares many genetic and clinical features, as well as a common ancestor, with *Shigella*. Both are intracellular pathogens for which virulence is mediated by the presence of specific factors and by the loss or inactivation of other factors (antivirulence genes), which presumably occurred during these organisms' transition from an extracellular to an intracellular lifestyle.

Colonization and invasion of the colonic mucosa, followed by replication therein and cell-to-cell spread (in part via a type III secretion system), result in the development of inflammatory colitis. However, unlike *Shigella*, EIEC produces disease only with a large inoculum (10^8–10^{10} CFU) and is less virulent, typically causing only mild, self-limited (7–10 days), watery diarrhea. Onset generally follows an incubation period of 1–3 days. Occasionally, EIEC can cause a shigellosis-like (dysentery) syndrome characterized by fever, abdominal pain, tenesmus, and scant stool containing mucus, blood, and inflammatory cells.

ENTEROAGGREGATIVE AND DIFFUSELY ADHERENT *E. COLI* EAEC has been described primarily in low- and middle-income countries and in young children. However, recent studies indicate that it may also be a relatively common cause of diarrhea in all age groups in industrialized countries. EAEC has been recognized increasingly as an important cause of traveler's diarrhea. It is highly adapted to humans—the probable reservoir. A large inoculum is required for infection, which usually manifests as watery and sometimes persistent diarrhea in healthy but also malnourished or HIV-infected hosts.

In vitro, EAEC cells exhibit a diffuse or "stacked-brick" pattern of adherence to small-intestine epithelial cells. Virulence factors that probably are necessary for disease are regulated in large part by the transcriptional activator AggR. The pathogenesis of EAEC disease begins with intestinal adherence, which results from stimulation of epithelial mucus production and bacterial biofilm formation, the latter mediated by fimbriae and possibly the mucinase Pic and dispersin. Inflammation ensues, resulting in epithelial cell exfoliation and intestinal secretion, which is mediated by the enterotoxins Pet, EAST-1, ShET1, and HlyE.

An additional enteric pathotype, DAEC, is a heterogenous group associated with diarrheal disease, primarily in children 2–6 years of age in some LMICs, and may cause traveler's diarrhea. DAEC can also cause UTI. Diffuse adherence is observed on epithelial cells. The Afa/Dr adhesins may contribute to the pathogenesis of such infections.

Diagnosis Acute infectious diarrhea can be classified as noninflammatory or inflammatory; the latter is suggested by grossly bloody or mucoid stools or a positive test for fecal leukocytes, lactoferrin, or calprotectin **(Chap. 138)**. ETEC, EPEC, DAEC, and EAEC cause noninflammatory diarrhea. Identification of these agents can be achieved with commercial multiplex molecular panels (e.g., the BioFire FilmArray Gastrointestinal Panel can detect STEC, ETEC, EPEC, EAEC, and EIEC). However, organism identification is rarely needed because the associated diseases are self-limited. ETEC causes the majority and EAEC a minority of cases of noninflammatory traveler's diarrhea; here again, however, definitive diagnosis generally is not necessary for management (as discussed below). If diarrhea persists for >10 days despite treatment, *Giardia* or *Cryptosporidium* (or, in immunocompromised hosts, certain opportunistic pathogens) should be sought.

Because of the considerable public-health importance of STEC/EHEC/ST-EAEC infections, including the threat of HUS, the CDC now recommends that all patients with community-acquired diarrhea, whether inflammatory or not, be evaluated for these pathogens by simultaneous culture (to provide an isolate for strain typing and for outbreak detection and control) and detection of Shiga toxin or the corresponding genes. The rationale for testing all cases of community-acquired diarrhea, regardless of clinical features, is that bloody stool and fecal white blood cells (or lactoferrin) are not reliably present with STEC/EHEC/ST-EAEC infection. In addition, the use of both tests increases diagnostic sensitivity over that with either test alone.

O157 STEC/EHEC may be identified via culture by screening for *E. coli* strains that do not ferment sorbitol, with subsequent serotyping and testing for Shiga toxin. Selective or screening media are not available for culture-based detection of non-O157 STEC/EHEC/ST-EAEC strains. Detection of Shiga toxins or toxin genes via DNA-based, enzyme-linked immunosorbent, and cytotoxicity assays offers the advantages of rapidity and detection of non-O157 STEC/EHEC/ST-EAEC strains. Specimens positive for toxin but culture-negative for O157 should be forwarded to the local or state public-health laboratory for specialized testing.

TREATMENT

Intestinal *E. coli* Infections

The mainstay of treatment for all diarrheal syndromes is replacement of water and electrolytes. This measure is especially important for STEC/EHEC/ST-EAEC infection because appropriate volume expansion may protect against renal injury and improve outcome.

The use of prophylactic antibiotics to prevent traveler's diarrhea generally should be discouraged, especially in light of high rates of antimicrobial resistance. However, in selected patients (e.g., those who cannot afford a brief illness or are predisposed to infection), the use of rifaximin, which is nonabsorbable and is well tolerated, is reasonable.

When stools are free of mucus and blood, early patient-initiated treatment of traveler's diarrhea with a fluoroquinolone or azithromycin decreases the duration of illness, and the use of loperamide may halt symptoms within a few hours. Although dysentery caused by EIEC is self-limited, antimicrobial therapy hastens the resolution of symptoms, particularly in severe cases. By contrast, antimicrobial therapy for STEC/EHEC/ST-EAEC infection (the presence of which is suggested by grossly bloody diarrhea without fever) should be avoided because antibiotics may increase the incidence of HUS (possibly via increased production/release of Stx). In the treatment of HUS, plasmapheresis is not recommended and the use of eculizumab (inhibition of C5) should be limited to clinical trials.

KLEBSIELLA INFECTIONS

K. pneumoniae is the most important *Klebsiella* species from a medical standpoint, causing community-acquired, LTCF-acquired, and nosocomial infections. *K. oxytoca* complex and *K.* (formerly *Enterobacter*) *aerogenes* are primarily pathogens in LTCFs and hospitals. *Klebsiella* species are broadly prevalent in the environment and colonize the mucosal surfaces of mammals. In healthy humans, the prevalence of *K. pneumoniae* colonization is 5–35% in the colon and 1–5% in the oropharynx; skin is usually colonized only transiently.

Most *Klebsiella* infections in Western countries are caused by "classic" *K. pneumoniae* (cKp) and occur in hospitals and LTCFs. The most common clinical syndromes due to cKp are pneumonia, UTI, abdominal infection, intravascular device infection, surgical site infection, soft tissue infection, and secondary bacteremia. cKp strains have gained notoriety because of their propensity for acquiring treatment-confounding antimicrobial resistance determinants and causing both localized and widespread outbreaks, such as with the global spread of cKp strains producing NDM-group MBLs. Clonal groups 11, 15, 101, 307, and 258, many members of which produce carbapenemases, are undergoing international dissemination. Transmission within or between institutions is common. *K. pneumoniae* is nearly fourfold more transmissible than *E. coli*, and, disconcertingly, carbapenemase-producing strains are associated with increased spread compared with carbapenem-susceptible strains.

In addition, hypervirulent *K. pneumoniae* (hvKp) strains that are phenotypically and clinically distinct from cKp have emerged recently, after their initial recognition in Taiwan in 1986. Although hvKp infections have occurred globally in all ethnic groups, most cases have been reported in individuals of Asian ethnicity residing in countries from the Asian Pacific Rim, but also in Asians living in other countries. Affected individuals often have diabetes mellitus. These demographics raise the possibility of a locale-specific distribution of the organism or an increased susceptibility of Asian hosts, especially those who are diabetic. In contrast to the usual health care–associated context for cKp infections in the West, hvKp can cause serious life- and organ-threatening infections in younger, healthy individuals from the community and can spread metastatically from the primary site of infection or present with multiple sites of infection. Of concern, recent reports from Asian countries have demonstrated that hvKp is responsible for an increasing number of health care–associated or hospital-acquired infections.

hvKp infection initially was characterized and distinguished from traditional infections caused by cKp strains by its (1) presentation as community-acquired monomicrobial pyogenic liver abscess **(Fig. 166-1, *top*)**, (2) occurrence in patients lacking a history of hepatobiliary disease, and (3) propensity for metastatic spread to distant sites. Subsequently, the hvKp pathotype has been recognized as the cause of extrahepatic abscesses and infections with or without liver involvement, including pneumonia; meningitis (in the absence of trauma or neurosurgery); endophthalmitis **(Fig. 166-1, *middle*)**; splenic, psoas, prostatic, epidural, and brain abscesses; and necrotizing fasciitis. Survivors often suffer catastrophic morbidity, such as vision loss and major neurologic sequelae. Most recently, clinicians are faced with an even greater challenge—the confluence of antimicrobial resistance determinants such as carbapenemase and ESBL genes possessed by cKp and the virulence factors possessed by hvKp on the same or coexisting plasmids. The result is the evolution of MDR and XDR hvKp.

K. pneumoniae subspecies *rhinoscleromatis* is the causative agent of rhinoscleroma, a granulomatous mucosal upper-respiratory infection that progresses slowly (over months or years) and causes necrosis and occasionally obstruction of the nasal passages. *K. pneumoniae* subspecies *ozaenae* has been implicated as a cause of chronic atrophic rhinitis and rarely of invasive disease in compromised hosts. *K.* (*Calymmatobacterium*) *granulomatis*, a sexually transmitted pathogen, is the causative agent of granuloma inguinale (donovanosis) that results in chronic genital ulcers **(Chap. 178)**. These *Klebsiella* pathotypes are usually isolated from patients in tropical climates and are genomically distinct from both cKp and hvKp.

■ INFECTIOUS SYNDROMES

Pneumonia Although cKp accounts for only a small proportion of cases of community-acquired pneumonia in Western countries

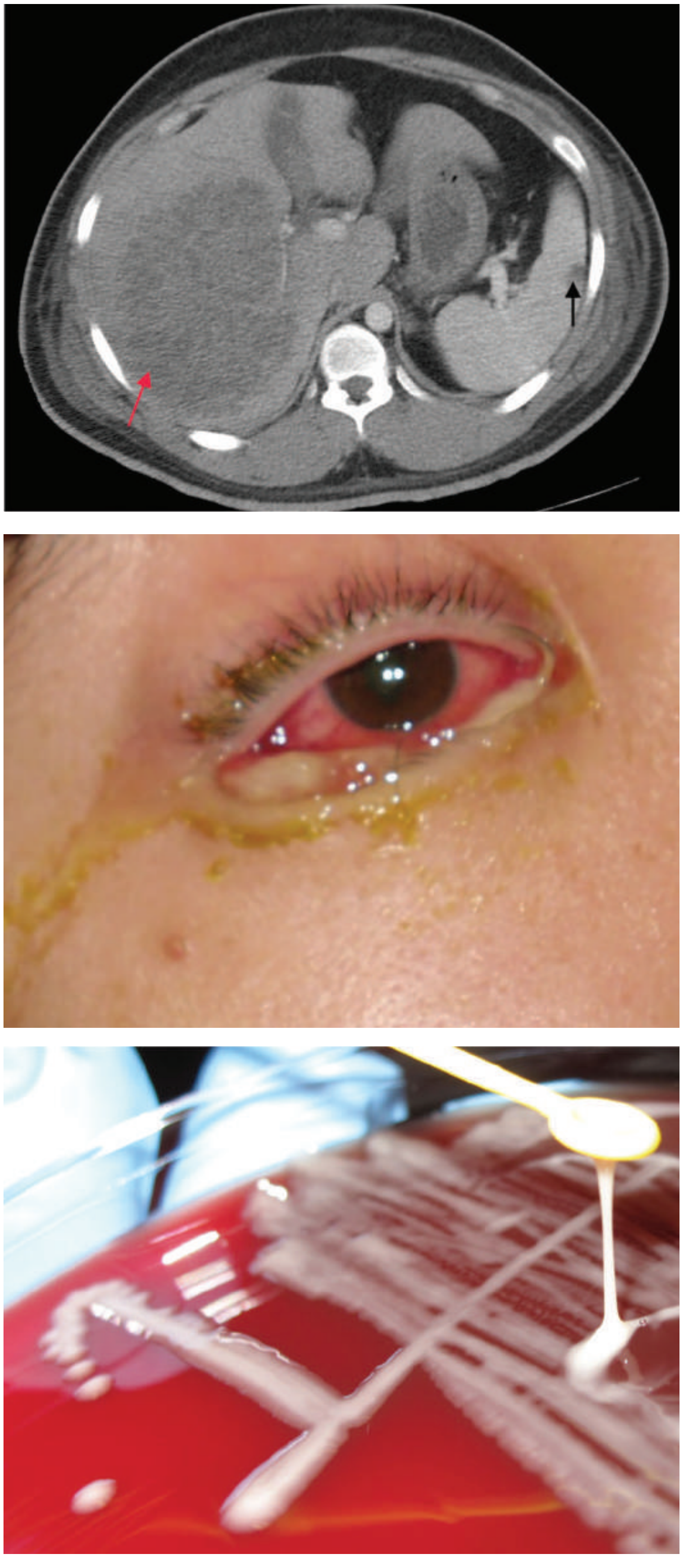

FIGURE 166-1 Hypervirulent pathotype of *K. pneumoniae* (hvKp). ***Top:*** Abdominal CT scan of a previously healthy 24-year-old Vietnamese man shows a primary liver abscess (*red arrow*) with metastatic spread to the spleen (*black arrow*). *(Courtesy of Drs. Chiu-Bin Hsaio and Diana Pomakova.)* ***Middle:*** A previously healthy 33-year-old Chinese man presented with endophthalmitis. *(AS Shon, RP Bajwa, TA Russo: Hypervirulent (hypermucoviscous) Klebsiella pneumoniae: A new and dangerous breed. Virulence 4:107, 2013.)* ***Bottom:*** A hypermucoviscous phenotype (which does not necessarily equate with a mucoid phenotype) has been associated with hvKp strains. A positive string test is shown. However, this test is not optimally sensitive or specific. Identification of all 5 of the biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2* is presently the most accurate means to identify hvKp.

(Chap. 131), cKp and *K. oxytoca* are common causes of pneumonia among LTCF residents and hospitalized patients because of increased rates of oropharyngeal colonization with these organisms in such individuals. Mechanical ventilation is an important risk factor. In Asia and South Africa, community-acquired pneumonia due to hvKp is becoming increasingly common, rivaling *Streptococcus pneumoniae*, and may occur in younger patients with no underlying disease. *Klebsiella* is also a common cause of pneumonia in severely malnourished children in LMICs.

As in all pneumonias due to enteric GNB, typical manifestations include production of purulent sputum and evidence of airspace disease. Presentation with earlier, less extensive infection is now more common than is the classically described lobar infiltrate, bulging fissure, and currant-jelly sputum. Pulmonary infection due to hvKp that has spread metastatically (e.g., from a hepatic abscess) usually includes nodular bilateral densities, more commonly in the lower lobes. Pulmonary necrosis, pleural effusion, and empyema can occur with disease progression.

UTI cKp accounts for only 1–2% of UTI episodes among otherwise healthy adults but for 5–17% of episodes of UTI in patients with anatomic and functional abnormalities of the urinary tract, including indwelling urinary catheter use (complicated UTI). UTI due to hvKp presents more commonly as renal or prostatic abscess due to bacteremic spread than as ascending infection from the urethra and bladder.

Abdominal Infection cKp causes a spectrum of abdominal infections similar to that caused by *E. coli* but is less frequently isolated than *E. coli*. hvKp is a common cause of monomicrobial community-acquired pyogenic liver abscess; in the Asian Pacific Rim, it has been recovered with steadily increasing frequency over the past two decades, replacing *E. coli* as the most common pathogen causing this syndrome. hvKp also is increasingly described as a cause of spontaneous bacterial peritonitis and splenic abscess.

Other Infections When cKp and *K. oxytoca* cause cellulitis or soft tissue infection, the process most frequently involves devitalized tissue (e.g., decubitus and diabetic ulcers, burn wounds) and immunocompromised hosts. cKp and *K. oxytoca* cause some cases of surgical site infection and nosocomial sinusitis as well as occasional cases of osteomyelitis contiguous to soft tissue infection, nontropical myositis, and meningitis (during the neonatal period and after neurosurgery). By contrast, hvKp has become an important cause of community-acquired monomicrobial necrotizing fasciitis, meningitis, endophthalmitis (Fig. 166-1, *middle*), and abscesses within the brain, subdural space, and epidural space, particularly in the Asian Pacific Rim but also globally. Cytotoxin-producing strains of *K. oxytoca* have been implicated as a cause of non–*C. difficile* antibiotic-associated hemorrhagic colitis.

Bacteremia *Klebsiella* infection at any site can produce bacteremia. Infections of the urinary tract, respiratory tract, and abdomen (especially hepatic abscess) each account for 15–30% as the source of *Klebsiella* bacteremia. Intravascular device–related infections account for another 5–15% of episodes, and surgical site and miscellaneous infections account for the rest. *Klebsiella* is an occasional cause of sepsis in neonates and of bacteremia in neutropenic patients. However, like other enteric GNB, *Klebsiella* rarely causes endocarditis or other endovascular infections, although endocarditis can involve extensive valvular destruction when it occurs.

■ DIAGNOSIS

Klebsiellae are readily isolated and identified in the laboratory. These organisms usually ferment lactose, although the subspecies *rhinoscleromatis* and *ozaenae* are nonfermenters and are indole-negative. hvKp frequently possesses a hypermucoviscous phenotype **(Fig. 166-1, *bottom*)**, although the sensitivity and specificity of the string test are less than optimal. Identification of all 5 of the biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2* is presently the most accurate means to identify hvKp in both antimicrobial sensitive and MDR isolates, although currently, this test is not routinely available.

TREATMENT
Klebsiella Infections

K. (formerly *Enterobacter*) *aerogenes* has a similar resistance profile to *E. cloacae*, the treatment of which is discussed below. *K. pneumoniae* and *K. oxytoca* have similar antibiotic resistance profiles; both are intrinsically resistant to ampicillin. The prevalence of acquired resistance in *K. pneumoniae* and *K. oxytoca* is generally >30% for amoxicillin-clavulanate, ampicillin-sulbactam, nitrofurantoin, and TMP-SMX and ~10–20% for fluoroquinolones, piperacillin-tazobactam, fosfomycin, and omadacycline.

USNHSN data from 2015–2017 identified 25% of *K. pneumoniae* as ESBL-producing strains; higher rates are reported from Asia, South America, and Africa. Although prevalence of ESBL-producing strains is greatest in LTCF, isolates of cKp that produce CTX-M ESBLs are increasingly described from the community. Oral treatment for infection due to ESBL-producing strains is more challenging with *Klebsiella* than with *E. coli* because of the comparatively poor activity of nitrofurantoin, the lesser activity of fosfomycin (~80%), and limited available data regarding pivmecillinam (>80%) and omadacycline (75–100% susceptible for ESBL-producing isolates, but 60% if resistant to tetracycline).

Predictably, the ESBL-driven use of carbapenems has selected for strains of cKp and *K. oxytoca* that produce carbapenemases (8–18% based on the study and locale, 8.6% prevalence from 2015–2017 USNHSN data). Treatment can be problematic for such organisms, especially those producing MBLs (e.g., NDM), for which the highest prevalences are in cKp and *K. oxytoca* isolates from Eastern Europe and Asia and among health care–associated isolates. Likewise, hvKp strains from Asia are also increasingly reported to produce ESBLs and carbapenemases.

Treatment options for carbapenem-resistant *Klebsiella* are similar to those described for *E. coli* and depend on the class of carbapenemase produced (see "Carbapenemase," above); consultation with infectious disease experts is advised. For carbapenem-susceptible strains, the most predictably active agents include carbapenems, amikacin, ceftazidime-avibactam, ceftolozane-tazobactam, polymyxins, cefiderocol, tigecycline, eravacycline, and omadacycline. Empirical treatment decisions for the critically ill patient should be dictated by local susceptibility patterns, patient-specific risk factors, and the site of infection.

PROTEUS INFECTIONS

Proteus species are part of the colonic flora of a wide variety of mammals, birds, fish, and reptiles. The ability of these GNB to generate histamine from contaminated fish has implicated them in the pathogenesis of scombroid (fish) poisoning **(Chap. 471)**.

P. mirabilis causes 90% of *Proteus* infections, which occur in the community, LTCFs, and hospitals. By contrast, *Proteus vulgaris* and *Proteus penneri* are associated primarily with infections acquired in LTCFs or hospitals. Correspondingly, *P. mirabilis* colonizes healthy humans (up to 50%), whereas *P. vulgaris* and *P. penneri* are isolated primarily from individuals with underlying disease. By far the most common site of *Proteus* infection is the urinary tract, where the principal known urovirulence factors of *Proteus* include adhesins, flagella, IgA-IgG protease, iron acquisition systems, and urease. *Proteus* less commonly causes infection at a variety of other extraintestinal sites.

■ INFECTIOUS SYNDROMES

UTI *P. mirabilis* causes only 1–2% of UTIs in healthy women, and *Proteus* species collectively cause only 5% of hospital-acquired UTIs. However, *Proteus* is responsible for 10–15% of cases of complicated UTI, primarily those associated with catheterization; indeed, *Proteus* accounts for 20–45% of urine isolates from chronically catheterized patients. This high prevalence is due in part to bacterial production of urease, which hydrolyzes urea to ammonia and results in alkalization of the urine. In alkaline urine, organic and inorganic compounds precipitate, contributing to the formation of struvite and carbonate–apatite crystals, biofilms on catheters, and/or frank calculi. *Proteus* becomes associated with the stones and biofilms; thereafter, it usually cannot be eradicated without removal of the stones or catheter. Over time, staghorn calculi may form within the renal pelvis and lead to obstruction and renal failure. Although biologically plausible, clinical support is lacking for the concept that urine samples exhibiting unexplained alkalinity should be cultured, and that isolation of a *Proteus* species (or other urea-splitting organism) should prompt consideration of an evaluation for urolithiasis.

Other Infections *Proteus* occasionally causes pneumonia (primarily in LTCF residents or hospitalized patients), nosocomial sinusitis, intraabdominal abscesses, biliary tract infection, surgical site infection, soft tissue infection (especially decubitus and diabetic ulcers), and osteomyelitis (primarily contiguous); in rare cases, it causes nontropical myositis. In addition, *Proteus* uncommonly causes neonatal meningitis, with the umbilicus frequently implicated as the source; this disease is often complicated by development of a cerebral abscess. Otogenic brain abscess also occurs.

Bacteremia Most episodes of *Proteus* bacteremia originate from the urinary tract, although intravascular devices and any of the less common sites of *Proteus* infection also are potential sources. Endovascular infection is rare. *Proteus* species are occasional agents of sepsis in neonates and of bacteremia in neutropenic patients.

■ DIAGNOSIS

Proteus is readily isolated and identified in the laboratory. Most strains are lactose-negative, produce H_2S, and demonstrate characteristic swarming motility and distinct odor on agar plates. *P. mirabilis* and *P. penneri* are indole-negative, whereas *P. vulgaris* is indole-positive. The inability to produce ornithine decarboxylase differentiates *P. penneri* from *P. mirabilis*.

TREATMENT
Proteus Infections

Intrinsic resistance occurs in all *Proteus* spp. to nitrofurantoin, polymyxins, imipenem, and the tetracycline derivatives (e.g., tigecycline, eravacycline, omadacycline) and, in *P. vulgaris* and *P. penneri*, also to ampicillin and the first- and second-generation cephalosporins. Acquired resistance (% of isolates) occurs in *P. mirabilis* to ampicillin (15–65%), and in *Proteus* spp. to fluoroquinolones (10–55%), fosfomycin (7–22%), and TMP-SMX (20–50%). In *P. mirabilis*, ampicillin-sulbactam is more active than ampicillin, with resistance rates of 6–18%, but the prevalence of ESBL production (which confers ampicillin-sulbactam resistance) is increasing in the United States (5–10%) and Asia (up to 60%). Isolates of *P. mirabilis* that produce CTX-M ESBLs have been recovered from ambulatory patients with no recent health care contact (see the section on the treatment of extraintestinal *E. coli* infections for treatment considerations). Acquired carbapenem resistance remains relatively infrequent (<10%). However, production of MBLs (e.g., NDM) limits treatment options due to the inherent resistance of *Proteus* spp. to polymyxins and tetracycline derivatives (see "Carbapenemase," above). For critically ill patients, agents with excellent activity overall against *Proteus* spp. (90–100% of isolates susceptible) include carbapenems (excepting imipenem), amikacin, piperacillin-tazobactam, aztreonam, cefepime, ceftazidime-avibactam, ceftolozane-tazobactam.

ENTEROBACTER AND *CRONOBACTER* INFECTIONS

The *E. cloacae* complex is responsible for most *Enterobacter* infections, whereas *Cronobacter sakazakii* (formerly *Enterobacter sakazakii*), *Cronobacter malonaticus*, *E. cancerogenus*, E. *asburiae*, *E. hormaechei*, *E. kobei*, *E. ludwigii*, and *E. gergoviae* are less commonly isolated (<1% for each). *Enterobacter bugandensis* has been recently described as an agent of sepsis in neonates and was isolated from the International Space Station. *Enterobacter* spp. cause primarily health care–related infections. The organisms are widely prevalent in foods, environmental sources (including equipment at health care facilities), and a variety of animals.

Colonization with these organisms is uncommon among healthy humans but increases significantly with LTCF residence or hospitalization. Although colonization is an important prelude to infection, direct introduction via IV lines (e.g., contaminated IV fluids or pressure monitors) or contaminated non-FDA-approved stem cell products also occurs. Extensive antibiotic resistance has developed in *Enterobacter* spp. and probably has contributed to these organisms' emergence as prominent nosocomial pathogens. Risk factors for *Enterobacter* infection include prior antibiotic treatment, comorbid disease, and ICU residency. *Enterobacter* spp. causes a spectrum of extraintestinal infections similar to those described for other GNB.

■ INFECTIOUS SYNDROMES

The most commonly encountered syndromes include pneumonia, UTI (particularly catheter-associated), intravascular device–related infection, surgical site infection, and abdominal infection (primarily postoperative or related to devices such as biliary stents). Nosocomial sinusitis, meningitis related to neurosurgical procedures (including use of intracranial pressure monitors), osteomyelitis, and endophthalmitis after eye surgery are less frequent. Neonates (particularly if low-birth-weight) are at risk for *C. sakazakii* infection, including neonatal bacteremia, necrotizing enterocolitis, and meningitis (which is often complicated by brain abscess or ventriculitis). Contaminated powdered infant formula has been implicated as a source for such neonatal infections. The WHO recommends that, to reduce the initial number of bacteria, powdered infant formula should be reconstituted with hot water (>70°C) and, to limit replication of residual bacteria, the reconstituted formula should be stored at <5°C or its storage time minimized.

Enterobacter bacteremia can result from primary infection at any anatomic site. In bacteremia of unclear origin, particularly in an outbreak setting, sources for consideration should include contaminated IV fluids or medications, blood components or plasma derivatives, catheter-flushing fluids, pressure monitors, and dialysis equipment. *Enterobacter* can also cause bacteremia in neutropenic patients. *Enterobacter* endocarditis is rare, occurring primarily in association with IV drug use or prosthetic valves.

■ DIAGNOSIS

Enterobacter is readily isolated and identified in the laboratory. Most strains are lactose-positive and indole-negative.

TREATMENT

Enterobacter Infections

E. cloacae is intrinsically resistant to ampicillin, ampicillin-sulbactam, ampicillin-clavulanate, the first-generation cephalosporins, and the cephamycins. The prevalence of acquired resistance has ranged from 15 to 40% for piperacillin-tazobactam, 5 to 23% for polymyxins, 15 to 17% for fosfomycin, 15 to 30% for TMP-SMX, and 5 to 20% for fluoroquinolones and is ~10% for omadacycline (53% if tetracycline resistant). USNHSN data from 2015–2017 identified at least 9% of *E. cloacae* isolates as presumptively ESBL-producing, based on cefepime resistance. The prevalence of ESBLs in *E. cloacae* outside of the United States is 20–50%. The use of third-generation cephalosporins can induce or select for stable derepression of AmpC β-lactamase. Because resistance may emerge during therapy (in one study, this phenomenon was documented in 20% of clinical isolates), these agents should be avoided in the treatment of severe *Enterobacter* infection.

Cefepime is stable to hydrolysis by AmpC β-lactamases; thus, it is a suitable option for treatment of *Enterobacter* infections so long as ESBL is not co-produced. Overall, resistance prevalence generally ranges from 10 to 25% for cefepime and 25 to 50% for aztreonam and the third-generation cephalosporins. Carbapenem resistance remains relatively uncommon (USNHSN data from 2015–2017 identified a 5% prevalence) and is more commonly associated with a combination of increased AmpC expression and decreased permeability due to porin mutations rather than carbapenemase production, although acquisition of carbapenemase genes is increasing (see "Carbapenemase," above). Uncertainty exists on the optimal treatment for non-CP-CR-*Enterobacter* spp. Fortunately, overall, the percentage of susceptibility is high (90–99%) for carbapenems, amikacin, ceftazidime-avibactam, cefiderocol, tigecycline, eravacycline, and omadacycline (the latter three for tetracycline-susceptible isolates). Once susceptibility data for a patient's isolate become available, de-escalation of the antimicrobial regimen is advisable whenever possible.

SERRATIA INFECTIONS

S. marcescens causes >90%, and *Serratia liquefaciens* complex <10%, of *Serratia* infections. Serratiae are found primarily in the environment (including in health care institutions), particularly in moist settings. Serratiae have been isolated from a variety of animals, insects, and plants, but only infrequently from healthy humans. In LTCFs and hospitals, reservoirs for the organisms include the hands and fingernails of health care personnel, food, milk (on neonatal units), sinks, medical equipment or devices, IV solutions or parenteral medications (particularly those generated by compounding pharmacies), prefilled syringes and multiple-access medication vials (e.g., for heparin, propofol, saline), blood products (e.g., platelets), hand soaps and lotions, irrigation solutions, and even disinfectants such as chlorhexidine.

Infection results from either direct inoculation (e.g., via contaminated injected substances [IV fluids, medications, or recreational drugs] or snake bite) or colonization (primarily of the respiratory tract). Sporadic infection is most common, but outbreaks (often involving MDR strains in adult and neonatal ICUs) also occur. Hygiene, medication-compounding standards, sterile technique, and infection control programs are critical measures to prevent infection.

The spectrum of extraintestinal infections caused by *Serratia* is similar to that for other GNB. *Serratia* species are usually considered to cause mainly health care–associated infections; they account for 1–3% of hospital-acquired infections. However, population-based laboratory surveillance studies in Canada and Australia have demonstrated that community-acquired *Serratia* infections occur more commonly than was previously appreciated, and case reports have documented serious infection in otherwise healthy hosts. *Serratia* also is one of the pathogens associated with chronic granulomatous disease.

■ INFECTIOUS SYNDROMES

The most common primary sites of *Serratia* infection are the respiratory and genitourinary tracts, intravascular devices, the eye (contact lens–associated keratitis and other ocular infections), surgical wounds, and the bloodstream (from contaminated infusions), although most episodes of *Serratia* bacteremia arise from one of the listed focal infections rather than contaminated infusate. Less common syndromes are soft tissue infections (including myositis, fasciitis, mastitis), osteomyelitis, abdominal and biliary tract infections (usually postprocedural), and septic arthritis (primarily from intraarticular injections). Serratiae are uncommon causes of neonatal meningitis; postsurgical meningitis, endophthalmitis, or breast implant infection; and bacteremia in neutropenic patients. Endocarditis is rare, occurring most commonly in IV drug users.

■ DIAGNOSIS

Serratiae are readily cultured and identified by the laboratory and are usually lactose- and indole-negative. The red pigmentation of some *S. marcescens* strains and *Serratia rubidaea* can produce distinctive clinical findings (e.g., pink breast milk or hypopyon; pseudohemoptysis).

TREATMENT

Serratia Infections

Most *Serratia* strains (>80%) are intrinsically resistant to ampicillin, amoxicillin-clavulanate, ampicillin-sulbactam, first- and second-generation cephalosporins, cephamycins, nitrofurantoin, and polymyxins; likewise, tetracycline derivatives are poorly active. By contrast, fluoroquinolones, TMP-SMX, piperacillin-tazobactam, fosfomycin, and omadacycline are active against 85–95% of U.S. and European isolates, including those resistant to tetracycline. Both in the United States and globally, the prevalence of ESBL-producing isolates is generally low (<10%), but rates of 20–30% have been reported in Asia and Latin America. Induction or selection of variants with stable de-repression of chromosomal AmpC β-lactamases during therapy with third-generation cephalosporins is considered to be uncommon. Resistance prevalence generally ranges from 10 to 20% for aztreonam and the third-generation cephalosporins. Acquisition of carbapenemase-encoding genes is uncommon but increasing. Production of MBL (e.g., NDM) limits treatment options due to *Serratia*'s predictable resistance to polymyxins and tetracycline derivatives (see "Carbapenemase," above). For critically ill patients, the most active agents overall (>90% susceptible) are carbapenems, piperacillin-tazobactam, cefepime, amikacin, ceftazidime-avibactam, and ceftolozane-tazobactam.

CITROBACTER INFECTIONS

C. freundii and *Citrobacter koseri* cause most human *Citrobacter* infections, which are epidemiologically and clinically similar to *Enterobacter* infections. *Citrobacter* species are commonly present in water, food, soil, and certain animals. Colonization with these organisms is uncommon among healthy humans but increases significantly with LTCF residence or hospitalization. *Citrobacter* species account for 1–2% of nosocomial infections. The affected hosts are usually immunocompromised and/or have comorbid disease or disruption of skin or mucosal barriers. Infection from treatment with contaminated, non-FDA-approved stem cell products has been described. *Citrobacter* causes extraintestinal infections like those described for other GNB.

INFECTIOUS SYNDROMES

The urinary tract accounts for 40–50% of *Citrobacter* infections. Less commonly involved sites include the biliary tree (particularly with stones or obstruction), the respiratory tract, surgical sites, soft tissue (e.g., decubitus ulcers), the peritoneum, and intravascular devices. Osteomyelitis (usually from a contiguous focus), central nervous system infection in adults (from neurosurgical or other types of meningeal disruption), and myositis occur rarely. *Citrobacter* (primarily *C. koseri*) also causes 1–2% of neonatal meningitis cases, of which 50–80% are complicated by brain abscess. Further, case reports in adults suggest that *C. koseri* infection has a predilection for abscess formation. *Citrobacter* bacteremia is most often due to UTI, biliary/abdominal infection, or intravascular device infection, and occurs in some neutropenic patients. Endocarditis and other endovascular infections are rare.

DIAGNOSIS

Citrobacter species are readily isolated and identified; 35–50% of isolates are lactose-positive, and 100% are oxidase-negative. *C. freundii* is indole-negative, whereas *C. koseri* is indole-positive.

TREATMENT

Citrobacter Infections

C. freundii is more antibiotic-resistant than is *C. koseri*. Most *C. freundii* isolates are intrinsically resistant to ampicillin, ampicillin-sulbactam, amoxicillin-clavulanate, first-generation cephalosporins, and cephamycins. *C. koseri* exhibits intrinsic resistance to ampicillin and ampicillin-sulbactam. Overall, the prevalence of acquired resistance generally ranges from 15 to 35% for third-generation cephalosporins, piperacillin-tazobactam, fluoroquinolones, and TMP-SMX and is ~10% for nitrofurantoin and omadacycline (but 39% for omadacycline if tetracycline-resistant). The prevalence of ESBL production ranges from 5 to 30%. The use of third-generation cephalosporins may result in the induction or selection of variants with stable de-repression of chromosomal AmpC β-lactamases during therapy. Presently, <10% of isolates have acquired carbapenemases (see "Carbapenemase," above). Carbapenems, amikacin, fosfomycin, polymyxins, cefepime, ceftolozane-tazobactam, ceftazidime-avibactam, cefiderocol, tigecycline, eravacycline, and omadacycline (the latter three if tetracycline-susceptible) are the most active agents against *Citrobacter* isolates (>90% susceptible).

MORGANELLA AND *PROVIDENCIA* INFECTIONS

M. morganii, *Providencia stuartii*, and (less frequently) *Providencia rettgeri* are the members of their respective genera that cause systemic human infections. *P. alcalifaciens* has been implicated as a cause of food-borne gastroenteritis. These organisms' epidemiologic associations, pathogenic properties, and clinical manifestations resemble those of *Proteus* species. *Morganella* and *Providencia* occur more commonly among LTCF residents than among hospitalized patients, largely resulting from chronic urinary catheter use. Because of these organisms' intrinsic resistance to polymyxins and tigecycline, they may become more common in settings with extensive use of these agents.

INFECTIOUS SYNDROMES

These species are primarily urinary tract pathogens, causing UTIs that are most often associated with long-term (>30-day) catheterization. Such infections commonly lead to biofilm formation and catheter encrustation (sometimes causing catheter obstruction) or the development of struvite bladder or renal stones (sometimes causing renal obstruction, abscess, and extrarenal extension, and serving as foci for relapse). They can cause purple urine ("purple bag syndrome"), as can *P. mirabilis*, *K. pneumoniae*, *E. coli*, and *P. aeruginosa*. *Morganella* is also commonly isolated from snakebite infection.

Other, less common infectious syndromes due to *Morganella* and *Providencia* include surgical site infection, soft tissue infection (primarily involving decubitus and diabetic ulcers), burn site infection, pneumonia (particularly ventilator-associated), intravascular device infection, and intraabdominal infection. Rarely, the other extraintestinal infections described for GNB also occur. Bacteremia is uncommon; when it does occur, any infected site can serve as the source, but the urinary tract accounts for most cases, followed by surgical site, soft tissue, and hepatobiliary infections.

DIAGNOSIS

M. morganii and *Providencia* are readily isolated and identified. Nearly all isolates are lactose-negative and indole-positive.

TREATMENT

Morganella and *Providencia* Infections

Morganella and *Providencia* are intrinsically resistant to ampicillin, ampicillin-clavulanate, ampicillin-sulbactam, first-generation cephalosporins, nitrofurantoin, tetracyclines and derivatives (e.g., tigecycline), imipenem (but not the other carbapenems), and the polymyxins. *P. stuartii* additionally exhibits intrinsic resistance to gentamicin and tobramycin, as does *M. morganii* to second-generation cephalosporins. Fosfomycin is poorly active (>50% resistance). The prevalence of resistance generally ranges from 10 to 30% for the third-generation cephalosporins, from 10 to 40% for fluoroquinolones, and from 20 to 40% for TMP-SMX; the prevalence is more widely variable for piperacillin-tazobactam. The prevalence of ESBL production is generally <10%. The prevalence of acquired carbapenemase production is <10%. Production of MBL (e.g., NDM) limits treatment options due to the inherent resistance of *Proteeae* to polymyxins and tetracycline derivatives (see "Carbapenemase," above). Overall, the most active agents (>90% of isolates susceptible) are carbapenems (excepting imipenem), amikacin, cefepime, ceftazidime-avibactam, ceftolozane-tazobactam, and cefiderocol. Removal of a colonized urinary catheter or stone is critical for eradication of UTI.

EDWARDSIELLA INFECTIONS

E. tarda is the only member of the genus *Edwardsiella* that is associated with human disease. This organism is found predominantly in freshwater and marine environments and in the associated aquatic animal species. Human acquisition occurs primarily from interaction with these reservoirs or ingestion of raw or inadequately cooked aquatic animals. *E. tarda* infection is rare in the United States, where acquisition occurs mainly along the Gulf of Mexico; recently reported cases are mostly from Asia. This pathogen shares clinical features with *Salmonella* species (as an intestinal pathogen; **Chap. 171**), *Vibrio vulnificus* (as an extraintestinal pathogen; **Chap. 173**), and *Aeromonas hydrophila* (as both an intestinal and an extraintestinal pathogen; **Chap. 173**).

INFECTIOUS SYNDROMES

Gastroenteritis is the predominant *Edwardsiella*-associated infectious syndrome (50–80% of reported cases). Self-limiting watery diarrhea is most common, but severe colitis also occurs. The most common extraintestinal infection is wound infection due to direct inoculation, which is often associated with brackish or freshwater injuries, snakebites, or fish-related trauma. A case of pneumonia occurred after a near-drowning incident. Cholecystitis, cholangitis, and hepatic abscess may be due to ascending infection via the biliary tree. Other infectious syndromes result from invasion of the gastrointestinal tract and subsequent bacteremia. A primary bacteremic syndrome, sometimes complicated by meningitis, has a 40% case–fatality rate; hematogenous seeding may result in hepatic and intra- and extraperitoneal abscesses, endocarditis, mycotic aneurysm, septic arthritis, osteomyelitis, necrotizing fasciitis, and empyema. Most hosts who develop systemic *Edwardsiella* infection have significant comorbidities (e.g., hepatobiliary disease, iron overload, cancer, or diabetes mellitus).

DIAGNOSIS

Although *E. tarda* can readily be isolated and identified, most laboratories do not routinely screen for or identify it in stool samples. Production of hydrogen sulfide is a characteristic biochemical property.

TREATMENT

Edwardsiella Infections

E. tarda is susceptible to most antimicrobial agents appropriate for use against GNB. Gastroenteritis is generally self-limiting, but treatment with a fluoroquinolone may hasten resolution. In the setting of severe sepsis, fluoroquinolones, third- and fourth-generation cephalosporins, carbapenems, and amikacin—either alone or in combination—are the safest choices pending susceptibility data.

INFECTIONS CAUSED BY MISCELLANEOUS GENERA

Other gram-negative organisms such as *Hafnia*, *Kluyvera*, *Cedecea*, *Pantoea*, *Ewingella*, *Leclercia*, *Raoultella*, and *Photorhabdus* spp. are occasionally isolated from diverse clinical specimens, including blood, sputum, urine, cerebrospinal fluid, joint fluid, bile, and wounds. Such organisms cause infection predominantly in compromised hosts or in association with an invasive procedure or foreign body. Cephalosporinases from *Kluyvera* have been implicated as the progenitors of CTX-M ESBLs. *Kluyvera* and *Raoultella* may produce carbapenemases.

FURTHER READING

Antimicrobial Resistance Collaborators: Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. Lancet 399:629, 2022. [Erratum in Lancet 400:1102, 2022.]

Bonten M et al: Epidemiology of *Escherichia coli* bacteremia: A systematic literature review. Clin Infect Dis 72:1211, 2021.

Cheng MP et al: Beta-lactam/beta-lactamase inhibitor therapy for potential AmpC-producing organisms: A systematic review and meta-analysis. Open Forum Infect Dis 6:ofz248, 2019.

David S et al: Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. Nat Microbiol 4:1919, 2019.

Harris PNA et al: Effect of piperacillin-tazobactam vs meropenem on 30-day mortality for patients with *E coli* or *Klebsiella pneumoniae* bloodstream infection and ceftriaxone resistance: A randomized clinical trial [published correction appears in JAMA 321:2370, 2019]. JAMA 320:984, 2018.

Holy O, Forsythe S: *Cronobacter spp*. as emerging causes of healthcare-associated infection. J Hosp Infect 86:169, 2014.

Kamiyama S et al: *Edwardsiella tarda* bacteremia, Okayama, Japan, 2005–2016. Emerg Infect Dis 25:1817, 2019.

Russo TA, Marr CM: Hypervirulent *Klebsiella pneumoniae*. Clin Microbiol Rev 32:e00001, 2019.

Tamma PD et al: Infectious Diseases Society of America 2022 Guidance on the Treatment of Extended-Spectrum β-lactamase Producing Enterobacterales (ESBL-E), Carbapenem-Resistant Enterobacterales (CRE), and *Pseudomonas aeruginosa* with Difficult-to-Treat Resistance (DTR-*P. aeruginosa*). Clin Infect Dis 75:187, 2022.

van Duin D et al: Molecular and clinical epidemiology of carbapenem-resistant Enterobacterales in the USA (CRACKLE-2): A prospective cohort study Lancet Infect Dis 20:731, 2020. [Erratum in Lancet Infect Dis 19:30755, 2020.]

Weiner-Lastinger LM et al: Antimicrobial-resistant pathogens associated with adult healthcare-associated infections: Summary of data reported to the National Healthcare Safety Network, 2015–2017. Infect Control Hosp Epidemiol 41:1, 2020.

167

Acinetobacter Infections

Rossana Rosa, Christine A. Vu

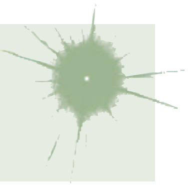

DEFINITION

Acinetobacter species were first described in 1911 and named *Micrococcus calcoaceticus*. Thereafter, the genus was renamed multiple times; since 1950, it has been known as *Acinetobacter*. *Acinetobacter* species are gram-negative, oxidase-negative, nonmotile, nonfermenting coccobacilli that are easily recovered on standard culture media. Differentiation among *Acinetobacter* species on the basis of phenotypic characteristics alone is very difficult. Molecular-based methods such as matrix-assisted laser desorption–ionization–time-of-flight mass spectrometry (MALDI-TOF-MS) and quantitative real-time polymerase chain reaction (PCR) are usually necessary to identify *Acinetobacter baumannii*, the most clinically relevant species of the genus.

ETIOLOGY AND EPIDEMIOLOGY

Acinetobacter species are naturally encountered in water and soil and have also been recovered from fruits and vegetables. In humans, *Acinetobacter* can be found on the skin and in the respiratory and gastrointestinal tracts. *A. baumannii* is capable of surviving environmental desiccation for weeks; this characteristic is important from an infection-control perspective as it allows this organism to persist in the hospital environment and on equipment.

Acinetobacter was historically considered a pathogen of hot and humid climates. In recent years, however, hospital outbreaks caused by *A. baumannii* have been reported worldwide, even in temperate climates. In the United States, the Centers for Disease Control and Prevention (CDC) estimates that 12,000 *Acinetobacter* infections occur every year, 7300 of which are caused by multidrug-resistant strains, with 500 attributable deaths. The increase in the number of infections with *A. baumannii* is suspected to be due to the rapid spread of certain genetically distinct lineages; of the three international clonal lineages (ICLs), ICL I and ICL II are multidrug resistant. The predominance of these lineages remains unexplained, although it has been proposed that this population structure is the result of two waves of expansion.

The first wave followed a bottleneck (possibly linked to a restricted ecologic niche) that occurred in the distant past. The second wave is ongoing and is being driven by the rapid expansion of a limited number of multidrug-resistant clones. The COVID-19 pandemic resulted in a setback to the efforts to control the spread of multidrug-resistant organisms, with significant increases in the rates of infections with carbapenem-resistant *Acinetobacter* reported worldwide.

Analysis of the *A. baumannii* pangenome (the sum of the core and dispensable genomes) has shown that its organization is characterized by a small core genome and a large accessory or disposable genome. This organization reflects *A. baumannii's* high plasticity, which enables it to acquire exogenous genetic material. With few exceptions, gene functions associated with virulence are found in the core genome; this observation suggests a limited role for the acquisition of new virulence traits in the recent nosocomial expansion of *A. baumannii* clones. Genes associated with resistance to antimicrobial agents are found in both the species core genome and the accessory genome. In the accessory genome, these genes have been found in alien islands, often flanked by integrases, transposases, or insertion sequences. This pattern suggests possible acquisition by horizontal gene transfer from other *Acinetobacter* strains or even from different bacterial species present in the immediate environment. Acquisition of these antimicrobial resistance genes is hypothesized to have led to the recent rapid expansion of highly homogeneous clonal lineages, whose main difference from nonclonal *A. baumannii* appears to be their antimicrobial resistance.

Health Care–Associated Infections Infections caused by *A. baumannii* occur frequently among patients admitted to intensive care units (ICUs). Risk factors for colonization and infection with this pathogen include nursing home residence, prolonged ICU stay, central venous catheterization, tracheostomy, mechanical ventilation, enteral feedings, and treatment with third-generation cephalosporins, fluoroquinolones, and carbapenems. Acquisition of carbapenem-resistant *A. baumannii* is most common among patients exposed to carbapenems. Spread of *A. baumannii* across different regions is facilitated by the movement of patients between health care systems and throughout the continuum of health care. Within the hospital, environmental spread of *A. baumannii* occurs as a result of inappropriate hand hygiene among workers providing health care for infected or colonized patients and the contamination of hospital equipment, such as respiratory therapy and ventilation equipment. The air surrounding the patient may also play a role in environmental colonization with *A. baumannii*, especially in inpatient areas without physical barriers between patients and with an inadequate number of air exchanges.

A. baumannii strains identified during hospital outbreaks are typically resistant to more antibiotic classes than strains from the community. The prevalence of colonization with *A. baumannii* at the time of admission or during a stay in a long-term acute-care hospital (LTACH) or nursing home is variable and depends on regional flora. Outbreaks of *A. baumannii* in acute-care hospitals and LTACHs that "share" patients have been described in Ohio, Michigan, Illinois, and Indiana.

Community-Acquired Infections Community-acquired infections caused by *Acinetobacter* have been described in Australia and Asia. Few cases have been reported in regions with a temperate climate, and even those few cases have taken place during warm and humid months. Risk factors for community-acquired pneumonia due to this organism include a history of alcohol abuse, diabetes mellitus, smoking, and chronic lung disease.

War Zone–Associated Infections Infections caused by *Acinetobacter* in war zones include skin and soft tissue infections associated with traumatic injuries and bloodstream infections. Outbreak investigations of *A. baumannii* infections among military personnel returning from Iraq and Afghanistan suggested the acquisition of *A. baumannii* in field hospitals rather than colonization of the skin before an injury. This view is supported by the recovery of *A. baumannii* isolates with similar genetic characteristics from inanimate surfaces in field hospitals and from patients.

Disaster Medicine *A. baumannii* is linked to infections among victims of trauma during tsunamis, earthquakes, and terrorist attacks. The types of infections most frequently observed in these settings are soft tissue injuries, but bloodstream infections and pneumonia have also been reported. In addition, outbreaks of *A. baumannii* infection in ICUs caring for disaster victims have been described.

PATHOGENESIS

Mechanisms of pathogenesis and virulence in *Acinetobacter* species have not been fully elucidated. However, *A. baumannii* seems to have greater virulence potential than other *Acinetobacter* species, as evidenced by its ability to grow at 37°C and to resist uptake by macrophages.

Initial *A. baumannii* colonization of the host and the environment is facilitated by the organism's ability to adhere to surfaces and human cells and to create biofilms. The ability to form a biofilm is phenotypically associated with exopolysaccharide production and pilus formation. A quorum-sensing molecule encoded by the *abaI* autoinducer synthase gene has been implicated in *A. baumannii* biofilm formation on abiotic surfaces. Outer-membrane porins appear to mediate cell apoptosis. *A. baumannii* can survive in harsh environments within the host and on inanimate surfaces by modifying the structure of its lipid A, with a consequent decrease in susceptibility to antibiotics and antimicrobial peptides and an increase in survival upon desiccation.

Acinetobacter species produce an extracellular capsule that protects the bacteria from external threats, including complement-mediated killing. Studies of mouse models showed that *Acinetobacter* species can increase capsule production in the presence of subinhibitory levels of antibiotic—an ability that leads to increased resistance to complement-mediated killing and a hypervirulent phenotype.

Phospholipase C and phospholipase D have been identified as virulence factors in *A. baumannii*. These enzymes exert cytotoxic effects on epithelial cells and facilitate their invasion.

Iron-acquisition systems are also important virulence mechanisms in *A. baumannii*. Through secretion of siderophores (low-molecular-mass ferric-binding compounds), *A. baumannii* is able to grow despite iron deficiencies in the surrounding environment (e.g., in the human host).

Several protein-secretion systems have been identified in *A. baumannii*. The most recently described is a type II secretion system. The substrate for this system, the LipA lipase, is required for growth on medium containing lipids as a sole carbon source. Mutants lacking the genes for the type II secretion system or its substrate exhibit defective in vivo growth in a neutropenic murine model of bacteremia. *A. baumannii* also has a type VI secretion system whose primary function seems to be to secrete antibacterial toxins that kill competing bacteria, including other strains in the same species.

The type V autotransporter system has been characterized in *A. baumannii*. In a murine systemic model of *Acinetobacter* infection, the *Acinetobacter* trimeric autotransporter mediates biofilm formation and maintenance; adherence to extracellular matrix components such as collagen I, II, and IV; and virulence.

Outer-membrane vesicles (OMVs) play a special role in protein secretion. Many *A. baumannii* strains secrete OMVs containing various virulence factors, including outer-membrane protein A (OmpA), proteases, and phospholipases. The membrane proteins in OMVs are responsible for eliciting a potent innate immune response. Several studies have shown that *A. baumannii* OMVs could be used as an acellular vaccine to effectively control *A. baumannii* infections.

Nosocomial strains of *Acinetobacter* can deploy multiple mechanisms of resistance, including alterations in porins and efflux pumps and expression of β-lactamases. More specifically, *Acinetobacter* species can reduce the expression of porins, thus hindering the passage of β-lactam antibiotics into the periplasmic space. These species can overexpress bacterial efflux pumps and decrease the concentration of β-lactam antibiotics in the periplasmic space. Efflux pumps can also actively remove quinolones, tetracyclines, chloramphenicol, disinfectants, and tigecycline. *Acinetobacter* species possess chromosomally encoded cephalosporinases and are capable of acquiring β-lactamases, including serine and metallo-β-lactamases. AmpC β-lactamases are class C β-lactamases intrinsic to all *A. baumannii* strains. Although these enzymes are expressed at low levels and are not inducible, the addition of the insertion sequence *ISAba1* next to the AmpC gene increases β-lactamase production, with resulting resistance to most cephalosporins.

Carbapenem resistance in *Acinetobacter* species is mostly tied to the emergence of Ambler class D oxacillinases of group 2d, some of which are intrinsic and chromosomal (e.g., OXA-51-like) while others are acquired and are found in plasmids or are chromosomally encoded (e.g., OXA-23-like, 24 [33-like, 40-like], 58-like, 143-like, and 235-like).

CLINICAL MANIFESTATIONS

Pneumonia *A. baumannii* is a notorious cause of nosocomial pneumonia, most frequently among patients requiring prolonged mechanical ventilation. The onset of disease tends to be later than that caused by other gram-negative bacilli; however, clinical symptoms of hospital-acquired or ventilator-associated pneumonia due to *A. baumannii* are similar to those of nosocomial or ventilator-associated pneumonia due to other nosocomial pathogens. Thus, the most common indicators of infection include fever and increased sputum production. The positivity of respiratory cultures in most cases may present a challenge for the clinician since airway colonization with *A. baumannii* may not always indicate a diagnosis of pneumonia, but it is a known risk factor for infection itself. In addition, radiologic findings are nonspecific and can include lobar consolidations and pleural effusions, with cavitations being rarely seen. The crude mortality rates associated with nosocomial pneumonia due to *A. baumannii* are reported as high as 65%. However, since these infections occur in debilitated patients, their attributable mortality has been difficult to establish.

Community-acquired pneumonia due to *A. baumannii* is relatively rare. Its clinical presentation is characterized by fever, severe respiratory symptoms, and multiple-organ dysfunction. Patients frequently have a cough productive of purulent sputum, shortness of breath, and chest pain. Imaging studies usually show lobar consolidation. Mortality rates associated with this process are >50%.

Bloodstream Infections Bloodstream infections due to *A. baumannii* are most frequent among ICU patients and usually occur in the presence of a central venous catheter or as a secondary complication of hospital-acquired or ventilator-associated pneumonia. Polymicrobial growth has been reported in 20–36% of bacteremia episodes. Fever is the most common sign of infection (developing in >95% of cases), and presentation with septic shock and disseminated intravascular coagulopathy has been described in as many as 25 to 30% of patients, respectively. *A. baumannii* bloodstream infections often result in higher hospitalization costs and longer ICU stays. Crude mortality rates from this infection are as high as 40%; however, rates can be as high as 70% from infections caused by carbapenem-resistant isolates. In patients with infections caused by extremely drug-resistant strains, poor outcomes are thought to be driven by delays in the initiation of adequate antimicrobial therapy.

Skin and Soft Tissue Infections *Acinetobacter* species have been described as part of the skin flora, yet the majority of the organisms from this genus that colonize the skin are not those associated with nosocomial infections. Discerning infection from wound colonization is challenging. Gunshot wounds and the presence of orthopedic external-fixation devices are common among patients with combat trauma–associated *A. baumannii* skin and soft tissue infections. The report on a case series of eight U.S. military patients described the clinical presentation of their infections as evolving from an edematous *peau d'orange* appearance to a sandpaper appearance with overlying vesicles and then to a necrotizing process with hemorrhagic bullae. Other case series have also included necrotizing fasciitis. *A. baumannii* is an important pathogen in burn units worldwide. Large burns provide ideal conditions for *A. baumannii* and facilitate patient-to-patient transmission. The presence of *A. baumannii* in wounds contributes to healing delays and graft loss. In addition, wound colonization is a risk factor for bloodstream infections among patients with extensive burn injuries.

A. baumannii infections resulting from trauma to soft tissues in the setting of natural disasters, such as tsunamis and earthquakes, have been reported. The implication is that *A. baumannii* should be considered in the differential diagnosis of soft tissue infections following exposure to tropical and subtropical environments.

Urinary Tract Infections *A. baumannii* is an infrequent cause of urinary tract infections. The majority of cases reported are catheter-associated infections, reflecting the ability of *A. baumannii* to form biofilms on these devices. A few reports have described community-acquired infections occurring in the setting of nephrolithiasis and after renal transplantation.

Meningitis Central nervous system infections with *A. baumannii* have been reported in the context of outbreaks, traumatic injuries, neurosurgical procedures, and external ventricular drains. One case series described a petechial rash in up to 30% of patients. *Acinetobacter* species may look similar to *Neisseria meningitidis* on a Gram stain of cerebrospinal fluid; both appear as gram-negative paired cocci. Eradication of *A. baumannii* from the cerebrospinal fluid can be challenging and requires careful selection of antibiotics that adequately penetrate the site of infection.

Other Miscellaneous Infections A few cases of *A. baumannii* keratitis associated with the use of contact lenses have been reported. Cases of native- and prosthetic-valve endocarditis have also been described.

TREATMENT

Acinetobacter Infections

Treatment of *Acinetobacter* infections is challenging due to difficulties in differentiating colonization versus infection and because *Acinetobacter* can develop resistance to most available antibiotics. Therefore, the choice of empirical therapy should be based on local epidemiology and, if available, the patient's colonization status with a carbapenem-resistant isolate. Definitive therapy should be determined by antimicrobial susceptibility testing. Antimicrobial options for the management of infections caused by *A. baumannii* are displayed in **Table 167-1**.

Acinetobacter species possess intrinsic β-lactamases that inactivate first- and second-generation cephalosporins. Through acquisition of extended-spectrum β-lactamases, these organisms can also become resistant to third- and fourth-generation cephalosporins, along with carbapenems. Nevertheless, when the isolate is susceptible, β-lactam agents should be used. Ampicillin-sulbactam (due to its sulbactam component) is the treatment of choice, with cefepime, meropenem, and imipenem as alternative options based on in vitro susceptibility testing.

Currently, there is no antibiotic regimen that has been proven superior for the treatment of carbapenem-resistant *A. baumannii*. The 2024 Infectious Diseases Society of America (IDSA) "Guidance on the Treatment of Antimicrobial Resistant Gram-Negative Infections" recommends sulbactam-durlobactam in combination with a carbapenem as their preferred regimen, and high-dose ampicillin-sulbactam in combination with either polymyxin B, minocycline, tigecycline, or cefiderocol is an alternative. Pairing of sulbactam with durlobactam makes a novel diazabicyclooctane non-β-lactam β-lactamase inhibitor with activity against the *Acinetobacter*-derived cephalosporinases and class D β-lactamases including carbapenemases of the OXA family. High-dose ampicillin-sulbactam is recommended in combination with either polymixn B, minocycline, tigecycline, or cefiderocol. The use of high-dose over standard dose ampicillin-sulbactam increases binding of sulbactam to its penicillin-binding proteins (PBP) targets (PBP2 and PBP3) in order to optimize inhibition of cell wall synthesis. This recommendation is based on two meta-analyses of small clinical trials and observational data. In a randomized clinical trial including 125 patients with carbapenem-resistant *A. baumannii*, patients treated

TABLE 167-1 Therapeutic Options for the Management of Multidrug-Resistant *Acinetobacter baumannii* Infections

ANTIBIOTIC	DOSING[a]	COMMENTS
Sulbactam	6–9 g/d	Unavailable as single drug in many countries (including the United States). Different dosing strategies proposed if administered with ampicillin.
Ampicillin-sulbactam	3 g q4h 9 g q8h 27 g q24h	Infuse over 30 min Infuse over 4 h Infuse as continuous infusion
Sulbactam-durlobactam	1 g/1 g q6h	Infuse over 3 h
Meropenem	2 g q8h	Carbapenem-susceptible isolates only; infuse over 3 h
Imipenem-cilastatin	500 mg q6h	Carbapenem-susceptible isolates only; infuse over 3 h
Cefiderocol	2g q8h	Use in combination therapy; infuse over 3 h
Colistin	Dosing per the international consensus guidelines on polymixins (Tsuji BT et al, Pharmacotherapy 39:10, 2019)	Colistin is preferred for urinary tract infections.
Polymyxin B	Dosing per the international consensus guidelines on polymixins (Tsuji BT et al, Pharmacotherapy 39:10, 2019)	Polymyxin B is preferred over colistin for bloodstream infections.
Tigecycline	200-mg loading dose followed by 100 mg q12h	Use in combination therapy
Minocycline	200 mg q12h	IV/PO. Use in combination therapy

[a]All drugs are given by the IV route unless otherwise stated.

with sulbactam-durlobactam had a 28-day all-cause mortality of 19% compared to 32% in patients treated with colistin, and with lower rates of nephrotoxicity in the sulbactam-durlobactam arm.

Cefiderocol is a siderophore cephalosporine with in vitro stability against the *Acinetobacter*-derived cephalosporinase and other extended-spectrum β-lactamases. However, *A. baumannii* isolates with reduced cefiderocol susceptibility have been described, and in a randomized clinical trial that included 54 critically ill patients with carbapenem-resistant *A. baumannii*, the end-of-study mortality was 50% in the cefiderocol arm, compared to 18% in the best available therapy arm (mostly consisting of colistin).

Polymyxins are cationic detergents that have become less popular as a result of nephrotoxicity and neurotoxicity. Additionally, polymyxins are difficult to dose, have a narrow therapeutic window, and do not reach optimal tissue concentration in the lungs, which is a common site of infection for *A. baumannii*. Despite its disadvantages, polymyxin B and polymyxin E (colistin) have been reintroduced in clinical practice as they retain in vitro activity against carbapenem-resistant *A. baumannii*. In a randomized study of patients with pneumonia due to carbapenem-resistant *Acinetobacter,* patients receiving colistin in combination with high-dose ampicillin-sulbactam had a higher rate of clinical improvement by day 5 compared to those receiving colistin monotherapy. The combination of colistin plus meropenem was long favored due to in vitro synergy; however, two randomized controlled trials showed this strategy had comparable outcomes to colistin monotherapy.

Several tetracycline derivatives have in vitro activity against *A. baumannii*. Of them, tigecycline and minocycline could be considered as part of a combination regimen, used at high doses when minimum inhibitory concentrations are low. Although doxycycline is a widely available agent with established breakpoints against *A. baumannii*, it is usually less active than minocycline. Eravacycline is a newer tetracycline with promising activity against *A. baumannii*; however its use has been limited; it is currently being marketed to treat more resistant strains.

Bacteriophage therapy against multidrug-resistant *A. baumannii* has been reported with varied success rates. Furthermore, dosing and duration of therapy vary by syndrome and resistance can also arise during treatment.

COMPLICATIONS AND PROGNOSIS

Infections caused by *A. baumannii* can be associated with high mortality rates. Factors contributing to higher mortality are thought to include severity of the patient's underlying illness and drug resistance in the infecting strain.

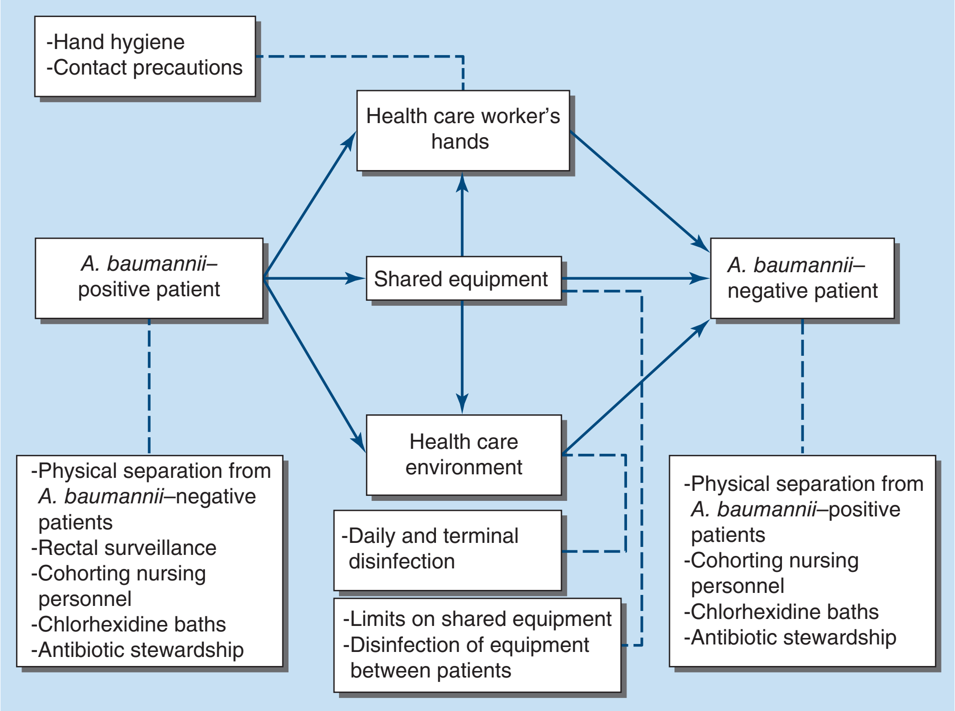
-Hand hygiene
-Contact precautions
Health care worker's hands
A. baumannii–positive patient
Shared equipment
A. baumannii–negative patient
Health care environment
-Physical separation from A. baumannii–negative patients
-Rectal surveillance
-Cohorting nursing personnel
-Chlorhexidine baths
-Antibiotic stewardship
-Daily and terminal disinfection
-Limits on shared equipment
-Disinfection of equipment between patients
-Physical separation from A. baumannii–positive patients
-Cohorting nursing personnel
-Chlorhexidine baths
-Antibiotic stewardship

FIGURE 167-1 Strategies for the prevention of dissemination of *Acinetobacter baumannii* in health care facilities.

INFECTION CONTROL AND PREVENTION

Acinetobacter species are capable of surviving on hospital surfaces for prolonged periods. In the hospital environment, *A. baumannii* has been associated with establishment of a *fecal patina*; this term refers to a coating of enteric organisms that can cover the skin of colonized patients and extend to their surrounding environment. Concentrations of enteric organisms are highest in the colonized patient's rectum, with spread in a target-like concentric pattern covering the patient's body and the surrounding environment. High-frequency touch areas in rooms occupied by patients colonized with *A. baumannii* are more likely to be contaminated. The hands, gloves, and gowns of health care workers can be contaminated after entry into the room of a patient colonized with *A. baumannii* **(Fig. 167-1)**.

Outbreaks caused by *A. baumannii* are frequently mono- or oligoclonal. A common source of infection has been identified in ~50% of outbreaks. These sources include respiratory therapy equipment, the hands of health care workers, bedside humidifiers, warm bathwater, hospital-prepared distilled water, bedpans, urine jugs, heparinized saline solution, mattresses, reusable pressure transducers in arterial lines, and fluids used for pressure lavage of wounds.

Control of multidrug-resistant *Acinetobacter* outbreaks starts with early recognition, with subsequent halting of the spread of infection throughout a facility and prevention of the establishment of an endemic strain. It is important to identify the outbreak strain and differentiate it from nonoutbreak strains so that infection control activities can be better targeted. Traditionally, the strain was identified with phenotypic typing systems (biotyping) or by determination of antimicrobial susceptibility patterns. Whole genome sequencing (WGS) has been proven an effective tool during outbreak investigations, usually deployed in response to an increase in the number of patients infected or colonized with a multidrug-resistant organism. However, an emerging body of evidence supports the prospective use of WGS for infection prevention, as it leads to early detection of transmission pathways and the potential for halting spread sooner and limiting cluster sizes through targeted infection prevention interventions.

During outbreaks, the simultaneous introduction of multiple ("bundled") measures makes it difficult to assess the impact of each individual measure. These interventions include aggressive cleaning of the general environment, active surveillance, contact isolation of colonized or infected patients, cohorting of medical staff, reinforcement of compliance with hand hygiene by health care workers, and use of aseptic care devices.

Colonization with *A. baumannii* is a strong predictor of subsequent clinical infection by this organism. Exposure to carbapenems is a risk factor for initial acquisition of this pathogen; therefore, efforts to curtail unnecessary use of antibiotics are fundamental to the prevention of *A. baumannii* colonization of patients and the organism's establishment in health care facilities.

FURTHER READING

Adams-Haduch JM et al: Molecular epidemiology of carbapenem-nonsusceptible *Acinetobacter baumannii* in the United States. J Clin Microbiol 49:3849, 2011.

Antunes LC et al: *Acinetobacter baumannii:* Evolution of a global pathogen. Pathog Dis 71:292, 2014.

Chen W: Host innate immune responses to *Acinetobacter baumannii* infection. Front Cell Infect Microbiol 10:486, 2020.

Dexter C et al: Community-acquired *Acinetobacter baumannii*: Clinical characteristics, epidemiology and pathogenesis. Expert Rev Anti Infect Ther 13:567, 2015.

Lee CR et al: Biology of *Acinetobacter baumannii*: Pathogenesis, antibiotic resistance mechanisms, and prospective treatment options. Front Cell Infect Microbiol 7:55, 2017.

Munoz-Price LS: Controlling multi-drug resistant gram-negative bacilli in your hospital: A transformational journey. J Hosp Infect 89:254, 2015.

Ong SWX et al: Genomic surveillance uncovers ongoing transmission of carbapenem-resistant *Acinetobacter baumannii* (CRAB) and identifies actionable routes of transmissions in an endemic setting. Infect Control Hosp Epidemiol 44:460, 2023.

Peleg AY et al: *Acinetobacter baumannii*: Emergence of a successful pathogen. Clin Microbiol Rev 21:538, 2008.

Shields RK et al: Navigating available treatment options for carbapenem-resistant *Acinetobacter baumannii-calcoaceticus* complex infections. Clin Infect Dis 76:S179, 2023.

Tamma PD et al: Infectious Diseases Society of America 2024 guidance on the treatment of antimicrobial-resistant gram-negative infections. Clin Infect Dis 2024:ciae403. *Available at https://doi.org/10.1093/cid/ciae403*. Accessed February 14, 2025.

Tal-Jasper R et al: Clinical and epidemiological significance of carbapenem resistance in *Acinetobacter baumannii* infections. Antimicrob Agents Chemother 60:3127, 2016.

Wong D et al: Clinical and pathophysiological overview of *Acinetobacter* infections: A century of challenges. Clin Microbiol Rev 30:409, 2017.

168

Helicobacter pylori Infections

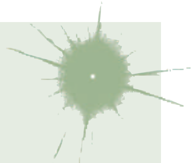

Emad M. El-Omar, Martin J. Blaser

Helicobacter pylori colonizes the stomach in ~40% of the world's human population, essentially for life unless eradicated by antibiotic treatment. Colonization with this organism is the main risk factor for peptic ulceration **(Chap. 335)** as well as for gastric adenocarcinoma and gastric mucosa-associated lymphoid tissue (MALT) lymphoma **(Chap. 85)**. Treatment for *H. pylori* has revolutionized the management of peptic ulcer disease, providing a permanent cure in most cases. Such treatment also represents first-line therapy for patients with low-grade gastric MALT lymphoma. Treatment of *H. pylori* is of no benefit in the treatment of gastric adenocarcinoma, but prevention of *H. pylori* colonization or eradicative treatment could potentially prevent gastric malignancy and peptic ulceration. In contrast, increasing evidence indicates that lifelong *H. pylori* colonization may offer some protection against complications of gastroesophageal reflux disease (GERD), including esophageal adenocarcinoma, although this point is still uncertain. Recent research has focused on whether *H. pylori* colonization is also a risk factor for some extragastric diseases and whether it is protective against some recently emergent medical problems, such as childhood-onset asthma and other allergic and metabolic conditions.

ETIOLOGIC AGENT

Helicobacter pylori *H. pylori* is a gram-negative bacillus that has naturally colonized humans for at least 100,000 years, and probably throughout human evolution. It lives in gastric mucus, with a proportion of the bacteria adherent to the mucosa and possibly a very small number of the organisms entering cells or penetrating the mucosa; the organism's distribution is mucosal rather than systemic. Its spiral shape and flagella render *H. pylori* motile in the mucus environment. The organism has several acid-resistance mechanisms, most notably a highly expressed urease that catalyzes urea hydrolysis to produce buffering ammonia. *H. pylori* is microaerophilic (i.e., grows in low levels of oxygen), is slow-growing, and requires complex growth media in vitro.

Other *Helicobacter* Species A small proportion of gastric *Helicobacter* infections are due to species other than *H. pylori*, possibly acquired as zoonoses. These non-*pylori* gastric helicobacters are associated with low-level inflammation and occasionally with disease. In immunocompromised hosts, several nongastric (intestinal) *Helicobacter* species can cause disease with clinical features resembling those of *Campylobacter* infections; these species are covered in **Chap. 169**.

EPIDEMIOLOGY

Prevalence and Risk Factors The prevalence of *H. pylori* among adults is <30% in most parts of the United States, Europe, and Oceania as opposed to >60% in many parts of Africa, South America, and West Asia. In the United States, prevalence varies with age but has been dropping; an estimated 18% of adults and 14% of children and adolescents are positive. *H. pylori* is usually acquired in childhood. The age association is due mostly to a birth-cohort effect whereby current 60-year-olds were more commonly colonized as children than are current children. Spontaneous acquisition or loss of *H. pylori* in adulthood is uncommon. Childhood acquisition explains why the main risk factors for infection are markers of crowding and social deprivation in childhood. Longitudinal studies have shown declining prevalences over the past half-century, concomitant with socioeconomic development and widespread antibacterial treatments.

Transmission Humans are the only important reservoir of *H. pylori*. Children may acquire the organism from their parents (most often the primary caregiver) or from other children. The former is more common in developed countries and the latter in less developed countries. Whether transmission takes place more often by the fecal–oral or the oral–oral route is unknown, but *H. pylori* is easily cultured from vomitus and gastroesophageal refluxate and is much less easily cultured from stool. Most acquisition of *H. pylori* is during the early years of childhood.

PATHOLOGY AND PATHOGENESIS

Long-term *H. pylori* colonization induces *chronic superficial gastritis*, a tissue response in the stomach that includes infiltration of the mucosa by both mononuclear and polymorphonuclear cells. (The term *gastritis* should be used specifically to describe histologic features; it has also been used to describe endoscopic appearances and even symptoms, but only magnification endoscopy correlates with microscopic findings or even with the presence of *H. pylori*, and even this is insufficient for diagnosis.) Although *H. pylori* is capable of numerous adaptations that prevent excessive stimulation of the immune system, colonization is accompanied by a considerable persistent local and systemic immune response, including the production of antibodies and cell-mediated responses. However, these responses are ineffective in clearing the bacterium. This inefficient clearing appears to be due in part to *H. pylori*'s downregulation of the immune system, which fosters its own persistence.

Most *H. pylori*–colonized persons do not develop clinical sequelae. That some persons develop overt disease whereas others do not is related to a combination of factors: bacterial strain differences, host susceptibility to disease, and environmental factors.

Bacterial Virulence Factors Several *H. pylori* virulence factors are more common among strains that are associated with disease than among those that are not. The *cag* island is a group of genes that encodes a bacterial type IV secretion system. Through this system, an effector protein, CagA, is translocated into epithelial cells, where it may be activated by phosphorylation and induces host cell signal transduction; proliferative, cytoskeletal, and inflammatory changes in the cell result. The protein at the tip of the secretory apparatus, CagL, binds to integrins on the cell surface, transducing further signaling. Finally, soluble components of the peptidoglycan cell wall enter the cell, mediated by the same secretory system. These components are recognized by the intracellular bacterial receptor Nod1, which stimulates a proinflammatory cytokine response resulting in an enhanced tissue response. Carriage of *cag*-positive strains increases the risk of both peptic ulcer and gastric adenocarcinoma. A second major host-interaction factor is the vacuolating cytotoxin VacA, which forms pores in cell membranes. VacA is polymorphic, and carriage of more active forms also increases the risk of ulcer disease and gastric cancer. Other bacterial factors that are associated with increased disease risk include adhesins, such as BabA (which binds to blood group antigens on epithelial cells).

Host Genetic and Environmental Factors There are currently no host genetic factors that have been reproducibly associated with risk of acquisition of *H. pylori* infection. The best-characterized host determinants of disease are genetic polymorphisms leading to enhanced activation of the innate immune response, including polymorphisms in cytokine genes and in genes encoding bacterial recognition proteins such as Toll-like receptors. For example, colonized people with polymorphisms in the interleukin 1 gene that increase the production of this cytokine in response to *H. pylori* infection are at increased risk of gastric adenocarcinoma. Another host genetic factor that has been linked with increased risk of gastric cancer in multiple populations is a single-nucleotide polymorphism (SNP) in the prostate stem cell antigen (*PSCA*) rs2294008. In addition to host genetics, environmental cofactors are important in pathogenesis. Smoking increases the risks of duodenal ulcers and gastric cancer in *H. pylori*–positive individuals. Diets high in salt and preserved foods increase cancer risk, whereas diets high in antioxidants and vitamin C are modestly protective.

Distribution of Gastritis and Differential Disease Risk The pattern of gastric tissue response is associated with disease risk: antral-predominant gastritis is most closely linked with duodenal ulceration, whereas pan-gastritis and corpus-predominant gastritis are linked with gastric ulceration and adenocarcinoma. This difference probably explains why patients with duodenal ulceration are not at high risk of developing gastric adenocarcinoma later in life, despite being colonized by *H. pylori*.

PATHOGENESIS OF DUODENAL ULCERATION How gastric colonization causes duodenal ulceration is now becoming clearer. *H. pylori*–induced tissue responses in the gastric antrum diminish the number of somatostatin-producing D cells. Because somatostatin inhibits gastrin release, gastrin levels are higher than in *H. pylori*–negative persons, and these higher levels lead to increased meal-stimulated acid secretion from the relatively spared gastric corpus. How this situation increases duodenal ulcer risk remains controversial, but the increased acid secretion may contribute to the formation of potentially acid-protective gastric metaplasia in the duodenum. Gastric metaplasia in the duodenum may become colonized by *H. pylori* and subsequently inflamed and ulcerated.

PATHOGENESIS OF GASTRIC ULCERATION AND GASTRIC ADENOCARCINOMA The pathogenesis of these conditions is less well understood, although both arise in association with pan- or corpus-predominant gastritis. The hormonal changes described above still occur, but the tissue responses in the gastric corpus mean that it produces less acid (hypochlorhydria) despite hypergastrinemia. Gastric ulcers commonly occur at the junction of antral and corpus-type mucosa, an area that is often particularly inflamed. Gastric cancer usually arises in stomachs with extensive atrophic gastritis and hypo- or achlorhydria, and probably stems from progressive DNA damage and the survival of abnormal epithelial cell clones. The DNA damage is thought to be due principally to reactive oxygen and nitrogen species arising from inflammatory cells, perhaps in relation to other bacteria that survive in a hypochlorhydric stomach. Longitudinal analyses of gastric biopsy specimens taken years apart from the same patient show that the common *intestinal* type of gastric adenocarcinoma follows stepwise changes from simple gastritis to gastric atrophy, metaplasia, and dysplasia. A second, *diffuse* type of gastric adenocarcinoma found more commonly in younger adults may arise directly from chronic gastritis without atrophic changes. In recent years, there has been a progressive rise in gastric cancers centered on the gastric corpus and occurring in younger adults (<50 years old) and disproportionately in females; this appears to be in the absence of *H. pylori.*

PATHOGENESIS OF GASTRIC MALT LYMPHOMA Low-grade B-cell MALT lymphomas are rare malignancies, reported at a rate of ~1 per million population per year prior to the discovery of *H. pylori.* Since then, reported rates have increased substantially, possibly reflecting overdiagnosis. These tumors arise from the substrate of chronic stimulation of lymphocyte populations by the persistent *H. pylori* colonization. Importantly, there have been numerous reports of these low-grade tumors responding dramatically to *H. pylori* eradication therapies. However, the boundary between true malignancy and benign lymphoid hypertrophy is uncertain. Among responders to *H. pylori* eradication, most do not have the characteristic t(11;18)(q21;q21) translocation of the malignancy and may not have true malignancies but rather benign polyclonal lymphoid proliferation. CagA-positive *H. pylori* strains have been significantly associated with the t(11;18)(q21;q21)–positive gastric MALT lymphoma compared with translocation-negative cases.

CLINICAL MANIFESTATIONS

Essentially all *H. pylori*–colonized persons have histologic gastritis, but only ~10–15% develop associated illnesses such as peptic ulceration, gastric adenocarcinoma, or gastric lymphoma **(Fig. 168-1)**. Despite similar rates of *H. pylori* colonization, rates of these diseases among women are less than half of those among men.

PEPTIC ULCER DISEASE

Worldwide, ~70% of duodenal ulcers and ~50% of gastric ulcers are related to *H. pylori* colonization **(Chap. 335)**. However, in particular, the proportion of gastric ulcers caused by aspirin and nonsteroidal anti-inflammatory drugs (NSAIDs) is increasing, and in many developed countries, these drugs have overtaken *H. pylori* as a cause of gastric ulceration. The main lines of evidence supporting an ulcer-promoting role for *H. pylori* are that (1) the presence of the organism is a risk factor for the development of ulcers, (2) non-NSAID-induced ulcers rarely develop in the absence of *H. pylori*, (3) eradication of *H. pylori* virtually abolishes long-term ulcer relapse, and (4) experimental *H. pylori* infection of gerbils can cause gastric ulceration. Thus, *H. pylori* is neither necessary nor sufficient for the development of peptic ulcer disease, but it is a very strong risk factor for its occurrence, and removal of *H. pylori* changes the natural history of ulcer disease.

Gastric Adenocarcinoma and Lymphoma Prospective nested case–control studies have shown that *H. pylori* colonization is a risk factor for adenocarcinomas of the distal (noncardia) stomach **(Chap. 85)**. Long-term experimental infection of gerbils also may result in gastric adenocarcinoma. Moreover, *H. pylori* may induce primary gastric lymphoma, although this condition is much less common, and the approaches to histopathologic and cytogenetic evaluations are not standardized. Many of the diagnosed low-grade gastric B-cell lymphomas are dependent on *H. pylori* for continuing growth and proliferation, and these tumors may regress either fully or partially after *H. pylori* eradication. However, they require careful short- and long-term monitoring; any that are not confined to the superficial mucosa (and, indeed, some that are) require additional treatment with chemotherapeutic agents or radiotherapy.

Functional Dyspepsia Many patients have upper gastrointestinal symptoms but have normal results on upper gastrointestinal endoscopy (so-called functional or nonulcer dyspepsia; **Chap. 335**). Because *H. pylori* is common, some of these patients will be colonized with the organism. *H. pylori* eradication leads to symptom resolution up to 15% more commonly than does placebo treatment. Whether such patients have peptic ulcers in remission at the time of endoscopy or whether a small subgroup of patients with "true" functional dyspepsia respond to *H. pylori* treatment is unclear. Either way, because functional dyspepsia is often persistent and difficult to treat, most consensus conference guidelines recommend *H. pylori* eradication in these patients. If this advice is followed, it is important to realize that only a small subgroup of patients who are treated will benefit.

Protection Against Peptic Esophageal Disease, Including Esophageal Adenocarcinoma Much interest has focused on a protective role for *H. pylori* against GERD **(Chap. 334)**, Barrett's esophagus **(Chap. 334)**, and adenocarcinoma of the esophagus and gastric cardia **(Chap. 85)**. The main lines of evidence for this role are (1) that there is a temporal relationship between a falling prevalence of gastric *H. pylori* colonization and a rising incidence of these conditions; (2) that, in most studies, the prevalence of *H. pylori* colonization (especially with proinflammatory *cagA*$^+$ strains) is significantly lower among patients with these esophageal diseases than among control participants; and (3) that, in prospective nested studies (see above), the presence of *H. pylori* is inversely related to these cancers. The mechanism underlying this protective effect is likely *H. pylori*–induced hypochlorhydria. Because, at the individual level, GERD severity may decrease, worsen, or remain unchanged after *H. pylori* treatment, concerns about GERD should not affect decisions about whether to treat *H. pylori* in an individual patient if a clear-cut indication exists; if there is no clear indication, clinicians should carefully balance considerations of benefit and harm.

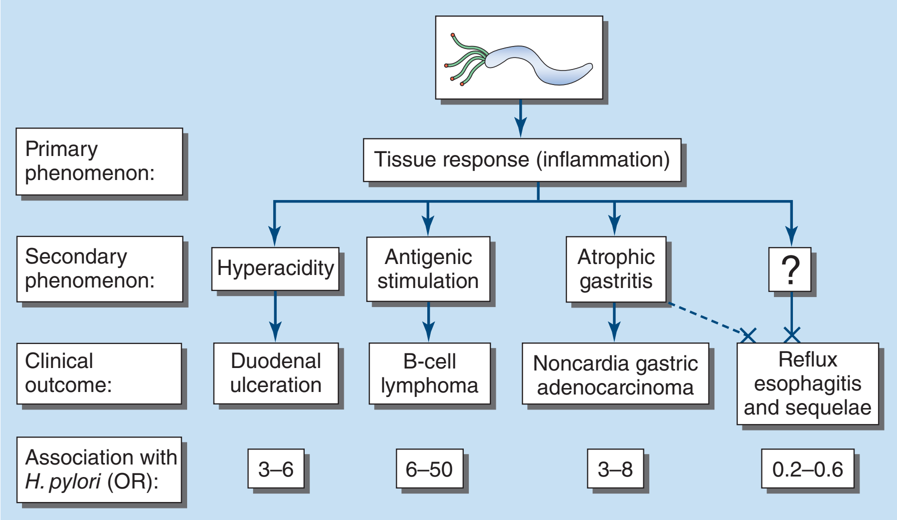

FIGURE 168-1 Schematic of the relationships between colonization with *Helicobacter pylori* and diseases of the upper gastrointestinal tract. Essentially all persons colonized with *H. pylori* develop a host response, which is generally termed *chronic gastritis*. The nature of the host's interaction with the particular bacterial population determines the clinical outcome. *H. pylori* colonization increases the lifetime risk of peptic ulcer disease, noncardia gastric cancer, and B-cell non-Hodgkin's gastric lymphoma (odds ratios [ORs] for all, >3). In contrast, a growing body of evidence indicates that *H. pylori* colonization (especially with *cagA*$^+$ strains) protects against adenocarcinoma of the esophagus (and the sometimes related gastric cardia) and premalignant lesions such as Barrett's esophagus (ORs, <1). Although the incidences of peptic ulcer disease (cases not due to nonsteroidal anti-inflammatory drugs) and noncardia gastric cancer are declining in developed countries, the incidence of adenocarcinoma of the esophagus is increasing. *(Reproduced with permission from MJ Blaser: Hypothesis: The changing relationships of Helicobacter pylori and humans: Implications for health and disease. J Inf Dis 179:1523, 1999.)*

Other Pathologies *H. pylori* has an increasingly recognized role in other gastric pathologies. It may predispose some patients to iron deficiency through occult blood loss and/or hypochlorhydria and reduced iron absorption. In addition, several extragastrointestinal pathologies have been linked with *H. pylori* colonization, although evidence of causality is less strong. Studies of *H. pylori* treatment in idiopathic thrombocytopenic purpura have consistently described improvement in or even normalization of platelet counts. Potentially important but even more controversial (protective) associations are with ischemic heart disease and cerebrovascular disease. However, the strength of the latter associations is reduced if confounding factors are taken into account, and our present knowledge is incomplete. Most authorities consider the associations to be non-causal. An increasing number of studies have shown an inverse association of *cagA*$^+$ *H. pylori* with childhood-onset asthma, hay fever, and atopic disorders. These associations have been shown to be causal in animal models, but the effect size in humans has not been established.

DIAGNOSIS

Tests for *H. pylori* fall into two groups: tests that require upper gastrointestinal endoscopy and simpler tests that can be performed in the clinic **(Table 168-1)**.

TABLE 168-1 Tests Commonly Used to Detect *Helicobacter pylori*

TEST	ADVANTAGES	DISADVANTAGES
Tests Based on Endoscopic Biopsy		
Biopsy urease test	Quick, simple	Some commercial tests not fully sensitive before 24 h
Histology	May give additional histologic information	Sensitivity dependent on experience and use of special stains
Culture	Permits determination of antibiotic susceptibility	Sensitivity dependent on experience
Noninvasive Tests		
Serology	Inexpensive and convenient; not affected by recent antibiotics or proton pump inhibitors to the same extent as breath and stool tests	Cannot be used to monitor treatment success; some commercial kits inaccurate, and most less accurate than urea breath test
^{13}C urea breath test	Inexpensive and simpler than endoscopy; useful for follow-up after treatment	Requires fasting; not as convenient as blood or stool tests
Stool antigen test	Inexpensive and convenient; useful for follow-up after treatment; may be particularly useful in children	Stool-based tests disliked by people from some cultures

Endoscopy-Based Tests Endoscopy is usually unnecessary in the initial management of young patients with simple dyspepsia but is commonly used to exclude malignancy and make a positive diagnosis in older patients or those with "alarm" symptoms. If endoscopy is performed, the most convenient biopsy-based test is the biopsy urease test, in which one large or two small gastric biopsy specimens are placed into a gel containing urea and an indicator. The *H. pylori* urease enzyme hydrolyses urea in gastric juice and generates ammonia. This leads to a rise in pH and therefore to a color change, which often occurs within minutes but can require up to 24 h. Histologic examination of biopsy specimens for *H. pylori* also is accurate, provided that a special stain (e.g., a modified Giemsa, silver, or immuno-stain) permitting optimal visualization of the organism is used. If biopsy specimens are obtained from both antrum and corpus, histologic study yields additional information, including the degree and pattern of inflammation and the presence of any atrophy, metaplasia, or dysplasia. Microbiologic culture is most specific but may be insensitive because of difficulty with *H. pylori* isolation. Once the organism is cultured, its identity as *H. pylori* can be confirmed by its typical appearance on Gram's stain and its positive reactions in oxidase, catalase, and urease tests. Moreover, the organism's susceptibility to antibiotics can be determined, and this information can be clinically useful in difficult cases. The occasional biopsy specimens containing the less common non-*pylori* gastric helicobacters give weakly positive results in the biopsy urease test. Positive identification of these bacteria requires visualization of the characteristic long, tight spirals in histologic sections; they cannot easily be cultured.

Noninvasive Tests Noninvasive *H. pylori* testing is the norm if gastric cancer does not need to be excluded by endoscopy. The longest-established test (and a very accurate one) is the *urea breath test*. In this simple test, the patient drinks a solution of urea labeled with the nonradioactive isotope ^{13}C and then blows into a tube. If *H. pylori* urease is present, the urea is hydrolyzed, and labeled carbon dioxide is detected in breath samples. The *stool antigen test*, a simple and accurate test using monoclonal antibodies specific for *H. pylori* antigens, is more convenient and less expensive than the urea breath test, but some patients dislike sampling stool. The simplest tests for ascertaining *H. pylori* status are *serologic assays* measuring specific IgG levels in serum by enzyme-linked immunosorbent assay or immunoblot. The best of these tests are nearly as accurate as other diagnostic methods, but many commercial tests—especially rapid office tests—do not perform well.

Use of Tests to Assess Treatment Success The urea breath test, the stool antigen test, and biopsy-based tests can all be used to assess the success of treatment (**Fig. 168-2**). However, because these tests are dependent on *H. pylori* load, their use <4 weeks after treatment may yield false-negative results. Early suppression of bacterial numbers may lead to false-negative results since regrowth of the organism can result in its detection weeks later. For the same reason, these tests are unreliable if performed within 4 weeks of intercurrent treatment with antibiotics or bismuth compounds or within 2 weeks of the discontinuation of proton pump inhibitor (PPI) treatment. In the assessment of treatment success, noninvasive tests are normally preferred. However, after gastric ulceration, endoscopy should be repeated to ensure healing and exclude gastric carcinoma by further histologic sampling; if PPIs have been stopped for at least 2 weeks and no antibiotics or bismuth compounds have been given for at least 6 weeks, there is an opportunity to assess treatment success with biopsy-based tests. Serologic tests are not used to monitor treatment success, as the gradual drop in titer of *H. pylori*–specific antibodies is too slow (requiring >14 weeks) to be of practical use.

TREATMENT

Helicobacter pylori Infection

INDICATIONS

The most clear-cut indications for treatment are *H. pylori*–related duodenal or gastric ulceration or low-grade gastric B-cell MALT lymphoma. Whether or not the ulcers are currently active, *H. pylori* should be eradicated in patients with documented ulcer disease to prevent relapse (Fig. 168-2). Guidelines have recommended *H. pylori* treatment for colonized patients with functional dyspepsia in case they are among the small percentage who will benefit from such therapy (beyond placebo effects). *H. pylori* eradication in the treatment of conditions not definitively known to respond has also been recommended but is not universally supported; such conditions include idiopathic thrombocytopenic purpura, vitamin B_{12} deficiency, and iron-deficiency anemia where other causes have been carefully excluded. For individuals with a strong family history of gastric cancer, treatment to eradicate *H. pylori* in the hope of reducing cancer risk is reasonable but of unproven value: it slightly reduces future cancer incidence, but there is no evidence it reduces all-cause mortality. For older dyspeptic patients in the community or those who have "alarm" symptoms (e.g., weight loss) associated with their dyspepsia, upper gastrointestinal endoscopy is indicated to seek a diagnosis and test for *H. pylori*; the decision regarding whether to eradicate the organism can then be based on indication. Endoscopy is usually considered unnecessary for young dyspeptic patients in the community who have no alarm symptoms (with the precise age cutoff dependent on local guidelines). If the community prevalence of *H. pylori* is below ~20%, such patients are treated with a short course of acid suppression using a PPI. If these patients do not respond or relapse when treatment is stopped, or if the *H. pylori* community prevalence is >20%, many national guidelines recommend a strategy of testing for *H. pylori* noninvasively and eradicating it if it is found. This strategy will benefit patients who have peptic ulcers and the ~5–10% of patients who have functional dyspepsia responsive to *H. pylori* eradication, but most patients will be treated unnecessarily. Currently, widespread community screening for and treatment of *H. pylori* as primary prophylaxis for gastric cancer and peptic ulcers are not recommended in most countries, mainly because the extent of the consequent reduction in cancer risk is not known. Several studies have found a modestly reduced cancer risk after treatment, but the period of follow-up is still fairly short, and the magnitude of the effect in different populations remains unclear. Other reasons not to treat *H. pylori* in asymptomatic populations at present include (1) the adverse side effects (which are common and can be severe in rare cases) of the multiple-antibiotic regimens used; (2) antibiotic resistance, which may emerge in *H. pylori* or other incidentally carried bacteria; (3) the anxiety that may arise in otherwise healthy people, especially if treatment is unsuccessful; and (4) the existence of a

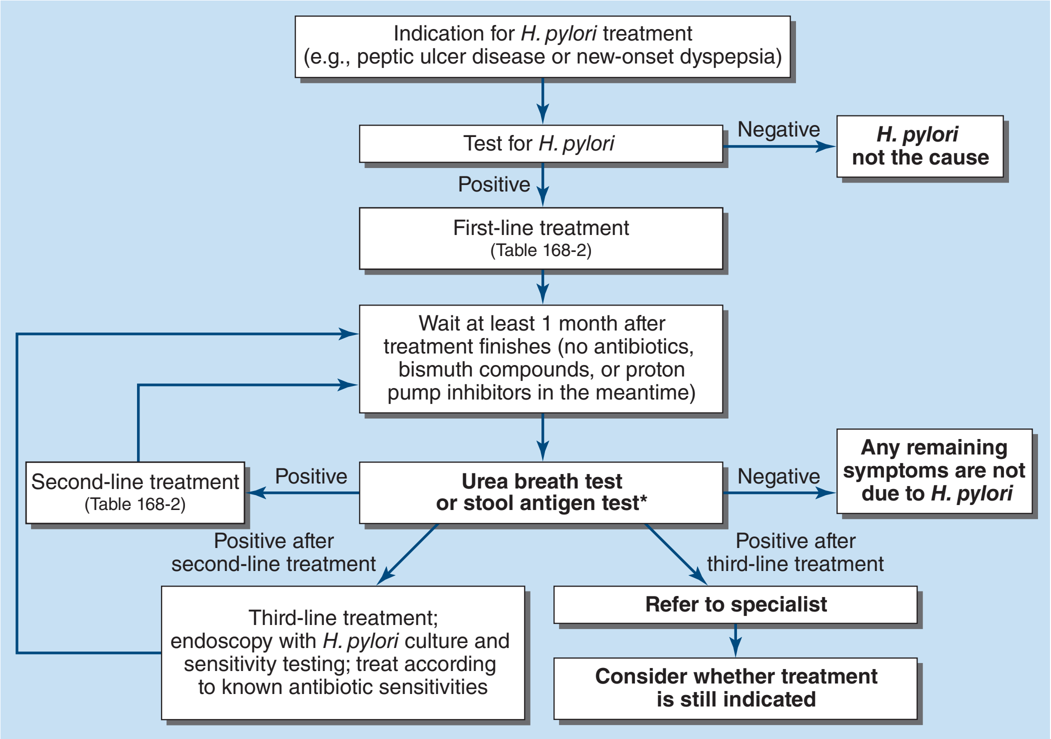
Indication for *H. pylori* treatment
(e.g., peptic ulcer disease or new-onset dyspepsia)
Test for *H. pylori*
Negative
***H. pylori* not the cause**
Positive
First-line treatment
(Table 168-2)
Wait at least 1 month after treatment finishes (no antibiotics, bismuth compounds, or proton pump inhibitors in the meantime)
Second-line treatment
(Table 168-2)
Positive
Urea breath test or stool antigen test*
Negative
Any remaining symptoms are not due to *H. pylori*
Positive after second-line treatment
Positive after third-line treatment
Third-line treatment; endoscopy with *H. pylori* culture and sensitivity testing; treat according to known antibiotic sensitivities
Refer to specialist
Consider whether treatment is still indicated

FIGURE 168-2 Algorithm for the management of *Helicobacter pylori* infection. *Note that either the urea breath test or the stool antigen test can be used in this algorithm. Occasionally, endoscopy and a biopsy-based test are used instead of either of these tests in follow-up after treatment. The main indication for these invasive tests is in follow-up after gastric ulceration; in this condition, as opposed to duodenal ulceration, it is important to check healing and exclude underlying gastric adenocarcinoma. However, even in this situation, patients undergoing endoscopy may still be receiving proton pump inhibitor therapy, which precludes *H. pylori* testing. Thus, a urea breath test or a stool antigen test is still required at a suitable interval after the end of therapy to determine whether treatment has been successful (see text). Some authorities use empirical third-line regimens, of which several have been described.

subset of people who will develop GERD symptoms after treatment. Despite the absence of screening strategies, many doctors treat *H. pylori* if it is known to be present (particularly in children and younger adults), even when the patient is asymptomatic. The rationale is that it reduces patient concern and may reduce future gastric cancer risk and that any reduction in risk is likely to be greater in younger patients. However, such practices do not factor in any potential benefits of *H. pylori* colonization. Overall, despite widespread clinical activity in this area, most treatment of persons with asymptomatic *H. pylori* carriage is given with no firm evidence base. Because a proportion of patients (up to 70%) of those diagnosed with gastric low-grade B-cell MALT lymphomas respond to *H. pylori* eradication, it should be used in all cases, regardless of whether *H. pylori* can be detected by the diagnostic modalities used since there may be falsely negative results. However, not all of these cases represent true malignancies, so the reported success rate may reflect the eradication of benign processes. Examination of tissues for the characteristic chromosomal translocations should be done to help distinguish benign and malignant processes and to guide further therapeutic approaches. These generally are slowly progressive tumors, so the time needed for *H. pylori* eradication and subsequent evaluation will not interfere with the use of subsequent chemotherapy and/or radiotherapy, if needed.

REGIMENS

Although *H. pylori* is susceptible to a wide range of antibiotics in vitro, monotherapy is not usually successful, probably because of inadequate active antibiotic delivery to the colonization niche. Clinical failure of monotherapy prompted the development of multidrug regimens. Current regimens consist of a PPI and two or three antimicrobial agents given for 10–14 days (Table 168-2). The optimal regimens vary in different parts of the world, depending on the known

TABLE 168-2 Commonly Recommended Treatment Regimens for *Helicobacter pylori*

REGIMEN[a] (DURATION)	DRUG 1	DRUG 2	DRUG 3	DRUG 4
Regimen 1: OCM (14 days)[b]	Omeprazole (20 mg bid[c])	Clarithromycin (500 mg bid)	Metronidazole (500 mg bid)	—
Regimen 2: OCA (14 days)[b]	Omeprazole (20 mg bid[c])	Clarithromycin (500 mg bid)	Amoxicillin (1 g bid)	—
Regimen 3: OBTM (14 days)[d]	Omeprazole (20 mg bid[c])	Bismuth subsalicylate (2 tabs qid)	Tetracycline HCl (500 mg qid)	Metronidazole (500 mg tid)
Regimen 4: concomitant (14 days)[e]	Omeprazole (20 mg bid[c])	Amoxicillin (1 g bid)	Clarithromycin (500 mg bid)	Tinidazole (500 mg bid[f])
Regimen 5: OAL (10 days)[g]	Omeprazole (20 mg bid[c])	Amoxicillin (1 g bid)	Levofloxacin (500 mg bid or 200 mg bid)	—

[a] The recommended first-line regimens for most of the world are shown in **bold** type. [b] This regimen should be used only for populations in which the prevalence of clarithromycin-resistant strains is known to be <20%. In practice, this restriction limits the regimens' appropriate range mainly to northern Europe. [c] Many authorities and some guidelines recommend doubling this dose of omeprazole as trials show resultant increased efficacy with some antibiotic combinations. Omeprazole may be replaced with any proton pump inhibitor (PPI) at an equivalent dosage. Because extensive metabolizers of PPIs are prevalent among Caucasian populations, many authorities recommend esomeprazole (40 mg bid) or rabeprazole (20 mg bid), particularly for regimens 4 and 5. [d] Data supporting this regimen come mainly from Europe and are based on the use of bismuth subcitrate (1 tablet qid) and metronidazole (400 mg tid). This is a recommended first-line regimen in most countries and is the recommended second-line regimen in northern Europe. [e] This regimen may be used as an alternative to regimen 3. [f] Metronidazole (500 mg bid) may be used as an alternative. [g] These regimens are used as second-line treatment in many countries (particularly where quadruple or concomitant therapy is used as the first-line regimen) and as third-line treatment in others. They may be less effective where rates of fluoroquinolone use are high and are more likely to be ineffective if there is a personal history of fluoroquinolone use for previous treatment of other infections.

rates of primary antibiotic resistance in most *H. pylori* strains in a particular locale. For this reason, guidelines on optimal regimens for *H. pylori* eradication in individual countries are evolving, and physicians should refer to the most up-to-date local guideline.

The two most important factors in successful *H. pylori* treatment are the patient's close compliance with the regimen and the use of drugs to which the patient's strain of *H. pylori* has not acquired resistance. Treatment failure following minor lapses in compliance is common and often leads to acquired resistance. To stress the importance of compliance, written instructions should be given to the patient, and minor side effects of the regimen should be explained. Increasing levels of primary *H. pylori* resistance to macrolides (azithromycin and clarithromycin are primarily used), levofloxacin, and—to a lesser extent—metronidazole are of growing concern. In most parts of the world (the main exception being northwestern Europe), the rate of primary macrolide resistance is sufficiently high that regimens containing clarithromycin plus one other antibiotic often fail; regimens with clarithromycin and two other antibiotics remain an option as the other two antibiotics are likely to eradicate *H. pylori* even if the strain is macrolide -resistant. When a patient is known to have been exposed—even remotely in time—to clarithromycin or a fluoroquinolone, these antibiotics usually should be avoided. Resistance to amoxicillin or tetracycline is unusual, even if these antibiotics have been given previously, and resistance to metronidazole is only partial; thus, there is no need to avoid using these antibiotics whether or not they have been previously prescribed. Whichever antibiotic regimen is used, meta-analyses show that using high rather than moderate doses of acid-suppressive PPIs with the antibiotics increases the effectiveness of the regimen. Similarly, use of vonoprazan, a highly effective potassium-competitive acid blocker, originally licensed in Japan and now in several countries, including the United States, was associated with higher eradication rates in conjunction with amoxicillin and clarithromycin, than when a PPI was used for acid suppression.

Assessment of antibiotic susceptibilities before treatment would be optimal but is not usually undertaken because endoscopy and mucosal biopsy are necessary to obtain *H. pylori* for culture and because most microbiology laboratories are inexperienced in *H. pylori* culture. If initial *H. pylori* treatment fails, the usual approach is empirical re-treatment with another drug regimen (Table 168-2). The third-line approach ideally should be endoscopy, biopsy, and culture plus treatment based on documented antibiotic sensitivities. However, empirical third-line therapies are often used.

Non-*pylori* gastric helicobacters are treated in the same way as *H. pylori*. However, in the absence of trials, it is unclear whether a positive outcome always represents successful treatment or whether it is sometimes due to natural clearance of the bacteria.

PREVENTION

Carriage of *H. pylori* has considerable public health significance in economically richer countries, where it is associated with peptic ulcer disease and gastric adenocarcinoma, and in some, but not all, economically poorer countries, where gastric adenocarcinoma may be an even more common cause of cancer death late in life. If mass prevention were contemplated, vaccination would be the most obvious method: experimental immunization of animals has given promising results, but vaccines in humans have thus far not been successful. However, given that *H. pylori* has co-evolved with its human host over millennia, preventing colonization on a population basis may have biological and clinical costs. For example, lifelong absence of *H. pylori* may be a risk factor for GERD complications, including esophageal adenocarcinoma. We have speculated that the disappearance of *H. pylori* may also be associated with an increased risk of other emergent diseases reflecting aspects of the current Western lifestyle, such as childhood-onset asthma and allergy, as supported by both epidemiologic and animal model studies.

Acknowledgment

The authors wish to thank John C. Atherton, MD, FRCP, for his prior contributions to this chapter.

FURTHER READING

Amieva M, Peek RM: Pathobiology of *Helicobacter pylori*–induced gastric cancer. Gastroenterology 150:64, 2016.

Anderson WF et al: The changing face of noncardia gastric cancer incidence among US non-Hispanic whites. J Natl Cancer Inst 110:608, 2018.

Arnold IC et al: *Helicobacter pylori* infection prevents allergic asthma in mouse models through the induction of regulatory T cells. J Clin Invest 121:3088, 2011.

Chen MJ et al: Molecular testing-guided therapy versus susceptibility testing-guided therapy in first-line and third-line *Helicobacter pylori* eradication: Two multicentre, open-label, randomised controlled, non-inferiority trials. Lancet Gastroenterol Hepatol 8:623, 2023.

Chen Y, Blaser MJ: Inverse associations of *Helicobacter pylori* with asthma and allergies. Arch Intern Med 167:821, 2007.

Chen Y et al: Association between *Helicobacter pylori* and mortality in the NHANES II study. Gut 62:1262, 2013.

Chen YC et al: Global prevalence of *Helicobacter pylori* infection and incidence of gastric cancer between 1980 and 2022. Gastroenterology 166:605, 2024.

Chow WH et al: An inverse relation between *cagA*+ strains of *Helicobacter pylori* infection and risk of esophageal and gastric cardia adenocarcinoma. Cancer Res 58:588, 1998.

Ford AC et al: *Helicobacter pylori* eradication therapy to prevent gastric cancer in healthy asymptomatic infected individuals: Systematic review and meta-analysis of randomized controlled trials. BMJ 348:g3174, 2014.

Graham DY et al: Rifabutin-based triple therapy (RHB-105) for *Helicobacter pylori* eradication: A double-blind, randomized, controlled trial. Ann Intern Med 172:795, 2020.

Hooi JKY et al: Global prevalence of *Helicobacter pylori* infection: Systematic review and meta-analysis. Gastroenterology 153:420, 2017.

Liou JM et al: Second-line levofloxacin-based quadruple therapy versus bismuth-based quadruple therapy for *Helicobacter pylori* eradication and long-term changes to the gut microbiota and antibiotic resistome: A multicentre, open-label, randomised controlled trial. Lancet Gastroenterol Hepatol 8:228, 2023.

Marshall BJ, Warren JR: Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet 1:1311, 1984.

Plummer M et al: Global burden of gastric cancer attributable to *Helicobacter pylori*. Int J Cancer 136:487, 2015.

Tshibangu-Kabamba E, Yamaoka Y: *Helicobacter pylori* infection and antibiotic resistance: From biology to clinical implications. Nat Rev Gastroenterol Hepatol 18:613, 2021.

169

Infections Due to *Campylobacter* and Related Organisms

Martin J. Blaser

DEFINITION

Bacteria of the genus *Campylobacter* and of the related genera *Arcobacter* and *Helicobacter* **(Chap. 168)** cause a variety of inflammatory conditions. Although acute diarrheal illnesses are most common, these organisms may cause infections in virtually all parts of the body, especially in compromised hosts, and these infections may have late nonsuppurative sequelae. The designation *Campylobacter* comes from the Greek for "curved rod" and refers to the organism's vibrio-like morphology.

ETIOLOGY

Campylobacters are motile, non-spore-forming, curved, gram-negative rods. Originally known as *Vibrio fetus*, these bacilli were reclassified as a new genus in 1973 after their dissimilarity to other vibrios was recognized. More than 20 species have since been identified. These species are currently divided into three genera: *Campylobacter*, *Arcobacter*, and *Helicobacter*. Not all of the species are pathogens of humans. The human pathogens fall into two major groups: those that primarily cause diarrheal disease and those that cause extraintestinal infection. The principal diarrheal pathogen is *Campylobacter jejuni*, which accounts for 80–90% of all cases of recognized illness due to campylobacters and related genera. Other organisms that can cause diarrheal disease include *Campylobacter coli, Campylobacter upsaliensis, Campylobacter lari, Campylobacter hyointestinalis, Campylobacter fetus, Arcobacter butzleri, Arcobacter cryaerophilus, Helicobacter cinaedi*, and *Helicobacter fennelliae*. The two *Helicobacter* species causing diarrheal disease, *H. cinaedi* and *H. fennelliae*, are intestinal rather than gastric organisms; in terms of the clinical features of the illnesses they cause, these species most closely resemble *Campylobacter* rather than *Helicobacter pylori* **(Chap. 168)** and thus are considered in this chapter. The pathogenic roles of *Campylobacter concisus* (a member of the normal human oral microbiome), *Campylobacter ureolyticus*, and *Campylobacter troglodytis* are uncertain. A new subspecies—*C. fetus* subspecies *testudinum*—has been described, chiefly in Asian patients; the very close resemblance of human isolates to strains isolated from reptiles suggests a food source.

The major species causing extraintestinal illnesses is *C. fetus*. However, any of the diarrheal agents listed above may cause systemic or localized infection as well, especially in compromised hosts. Neither aerobes nor strict anaerobes, these microaerophilic organisms are adapted for survival in the gastrointestinal mucous layer. This chapter focuses on *C. jejuni* and *C. fetus* as the major pathogens and prototypes for their groups. The key features of infection are listed by species (excluding *C. jejuni*, described in detail in the text below) in **Table 169-1**.

EPIDEMIOLOGY

Campylobacters are found in the gastrointestinal tract of many animals used for food (including poultry, cattle, sheep, and swine) and many household pets (including birds, dogs, and cats). These microorganisms often do not cause illness in their animal hosts, but occasionally this can occur (especially in puppies); in such circumstances, puppy-to-human transmission may be detected. In most cases, campylobacters are transmitted to humans in raw or undercooked food products or through direct contact with infected animals. In the United States and other developed countries, ingestion of contaminated poultry that has not been sufficiently cooked is the most common mode of acquisition (30–70% of cases). Other modes include ingestion of raw (unpasteurized) milk or untreated water, contact with infected household pets, ingestion of contaminated seafood, travel to developing countries (campylobacters being a leading cause of traveler's diarrhea; **Chaps. 130 and 138**), oral–anal sexual contact, cross-contamination from any of these sources, and (occasionally) contact with an index case who is incontinent of stool.

Campylobacter infections are common. Active surveillance of foodborne infections in the United States estimates the incidence of diarrheal disease due to campylobacters at ~20 cases per 100,000 persons—similar in incidence to *Salmonella* and more common than *Shigella*. Infections occur throughout the year, but the incidence peaks during summer and early autumn. Persons of all ages are affected; however, attack rates for *C. jejuni* are highest among young children and young adults, whereas those for *C. fetus* are highest at the extremes of age. Systemic infections due to *C. fetus* (and to other *Campylobacter* and related species) are most common among compromised hosts. Persons at increased risk include those with AIDS, immunoglobulin deficiencies, neoplasia, liver disease, diabetes mellitus, and generalized atherosclerosis as well as neonates and pregnant women; proton pump inhibitor use also increases risk. However, apparently healthy nonpregnant persons occasionally develop transient *Campylobacter* bacteremia as part of a gastrointestinal illness (0.1–1% of cases).

In contrast, in many developing countries where sanitation is poor, *C. jejuni* infections are hyperendemic, with the highest rates among children <2 years old. According to large prospective cohort studies in low- to middle-income countries, *C. jejuni* infections—even when asymptomatic—are associated with short stature (stunting) and with a particular microbiome signature. Rates of clinically apparent infection fall with age, as does the illness-to-infection ratio, consistent with progressive development of immunity.

PATHOLOGY AND PATHOGENESIS

C. jejuni infections may be subclinical, especially in hosts in developing countries who have had multiple prior infections and may be partially

TABLE 169-1 Clinical Features Associated with Infection Due to "Atypical" *Campylobacter* and Related Species Implicated as Causes of Human Illness

SPECIES	COMMON CLINICAL FEATURES	LESS COMMON CLINICAL FEATURES	ADDITIONAL INFORMATION
Campylobacter coli	Fever, diarrhea, abdominal pain	Bacteremia[a]	Clinically indistinguishable from *C. jejuni*
Campylobacter fetus	Bacteremia[a], sepsis, meningitis, vascular infections	Diarrhea, relapsing fevers	Not usually isolated from media containing cephalothin or incubated at 42°C
Campylobacter upsaliensis	Watery diarrhea, low-grade fever, abdominal pain	Bacteremia[a], abscesses	Difficult to isolate because of cephalothin susceptibility
Campylobacter lari	Abdominal pain, diarrhea	Colitis, appendicitis	Seagulls frequently colonized; organism often transmitted to humans via contaminated water
Campylobacter hyointestinalis	Watery or bloody diarrhea, vomiting, abdominal pain	Bacteremia[a]	Causes proliferative enteritis in swine
Helicobacter fennelliae	Chronic mild diarrhea, abdominal cramps, proctitis	Bacteremia[a]	Best treated with fluoroquinolones
Helicobacter cinaedi	Chronic mild diarrhea, abdominal cramps, proctitis	Bacteremia[a]	Best treated with fluoroquinolones; identified in healthy hamsters
Campylobacter jejuni subspecies *doylei*	Diarrhea	Chronic gastritis, bacteremia[b]	Uncertain role as human pathogen
Arcobacter cryaerophilus	Diarrhea	Bacteremia[a]	Poultry, seafood sources. Cultured under aerobic conditions
Arcobacter butzleri	Fever, diarrhea, abdominal pain, nausea; or asymptomatic	Bacteremia[a], appendicitis	Cultured under aerobic conditions; enzootic in nonhuman primates
Campylobacter sputorum	Pulmonary, perianal, groin, and axillary abscesses; diarrhea	Bacteremia	Three clinically relevant biovars: *sputorum, faecalis*, and *paraureolyticus*

[a] In compromised hosts, especially including the elderly, and patients with immunodeficiencies, diabetes, and infection with HIV. [b] In children.

Source: Adapted from BM Allos, MJ Blaser: Clin Infect Dis 20:1092, 1995.

immune. Symptomatic infections mostly occur within 2–4 days (range, 1–7 days) of exposure to the organism. The sites of tissue injury include the jejunum, ileum, and colon. Biopsies show an acute nonspecific inflammatory reaction, with neutrophils, monocytes, and eosinophils in the lamina propria, as well as damage to the epithelium, including loss of mucus, glandular degeneration, and crypt abscesses. Biopsy findings may be consistent with Crohn's disease or ulcerative colitis, but these "idiopathic" chronic inflammatory diseases should not be diagnosed unless infectious colitis, *specifically including* that due to infection with *Campylobacter* species and related organisms, has been ruled out.

The components of protective immunity to *Campylobacter* in humans are poorly understood. The high frequency of *C. jejuni* infections and their severity and recurrence among immunoglobulin-deficient patients suggest that antibodies are important in protective immunity. Experience from field studies and human experimental infection models suggests that immune protection may be short-lived or incomplete in the absence of continuous exposure. Knowledge of the pathogenesis of infection is also incomplete. Both the motility of the strain and its capacity to adhere to host tissues appear to favor disease, but classic enterotoxins and cytotoxins (including cytolethal distending toxin) appear not to play substantial roles in tissue injury or disease production. The organisms have been visualized within the epithelium, albeit in low numbers. The documentation of a significant tissue response and occasionally of *C. jejuni* bacteremia further suggests that tissue invasion is clinically significant, and in vitro studies are consistent with this pathogenic feature.

The pathogenesis of *C. fetus* infections is better defined. Virtually all clinical isolates of *C. fetus* possess a proteinaceous capsule-like structure (an S-layer) that renders the organisms resistant to complement-mediated killing and opsonization. As a result, *C. fetus* can cause bacteremia and can seed sites beyond the intestinal tract. The ability of the organism to switch the S-layer proteins expressed—a phenomenon that results in antigenic variability—may contribute to the chronicity and high rate of recurrence of *C. fetus* infections in compromised hosts.

CLINICAL MANIFESTATIONS

The clinical features of infections due to *Campylobacter* and the related *Arcobacter* and intestinal *Helicobacter* species causing enteric disease appear to be highly similar. *C. jejuni* can be considered the prototype, in part because it is by far the most common enteric pathogen in the group. A prodrome of fever, headache, myalgia, and/or malaise often occurs 12–48 h before the onset of diarrheal symptoms. The most common signs and symptoms of the intestinal phase are diarrhea, abdominal pain, and fever. The degree of diarrhea varies from several loose watery stools to visibly bloody stools (~10% of cases in adults); most patients presenting for medical attention have ≥10 bowel movements on the worst day of illness. Abdominal pain usually consists of cramping and may be the most prominent symptom. Pain is usually generalized but may become localized; *C. jejuni* infection may cause pseudoappendicitis. Fever may be the only initial manifestation of *C. jejuni* infection, a situation mimicking the early stages of typhoid fever. Febrile young children may develop convulsions, and both myocarditis and pericarditis have been observed, especially in young men. *Campylobacter* enteritis is generally self-limited; however, symptoms persist for >1 week in 10–20% of patients seeking medical attention, and clinical relapses occur in 5–10% of untreated patients. Studies of common-source epidemics indicate that milder illnesses or asymptomatic infections may commonly occur.

C. fetus may cause a diarrheal illness similar to that due to *C. jejuni*, especially in immunocompetent hosts. This organism also may cause either intermittent diarrhea or nonspecific abdominal pain without localizing signs. Sequelae are uncommon, and the outcome is benign. *C. fetus* may also cause a prolonged relapsing systemic illness (with fever, chills, and myalgias) with bacteremia that has no obvious primary source; this manifestation is especially common among compromised hosts. Secondary seeding of an organ (e.g., meninges, brain, bone, urinary tract, or soft tissue) complicates the course, which may be fulminant. *C. fetus* infections have a tropism for vascular sites: endocarditis, mycotic aneurysm, and septic thrombophlebitis may all occur. Infection during pregnancy often leads to fetal death. A variety of *Campylobacter* species and *H. cinaedi* can cause recurrent cellulitis with fever and bacteremia in immunocompromised hosts.

COMPLICATIONS

About 90% of bacteremic infections are caused by *C. jejuni* and *C. fetus,* in roughly equal proportions, despite the manyfold greater incidence of *C. jejuni* infections. Overall, mortality has been reported at ~12%. Except in infection with *C. fetus*, bacteremia is uncommon, developing most often in compromised hosts—including those who are immunocompromised and diabetic—and at the extremes of age. Three patterns of extraintestinal infection have been noted: (1) transient bacteremia in a normal host with enteritis (benign course, no specific treatment needed); (2) sustained bacteremia or focal infection in a normal host (bacteremia originating from enteritis, with patients responding well to antimicrobial therapy); and (3) sustained bacteremia or focal infection in a compromised host. Enteritis may not be clinically apparent. Immediate antimicrobial therapy, possibly prolonged, is necessary for suppression or cure of these infections.

Campylobacter, Arcobacter, and intestinal *Helicobacter* infections in the elderly as well as in patients with AIDS or immunoglobulin-deficient patients (most often common variable immunodeficiency) may be severe, persistent, and extraintestinal; relapse after cessation of therapy is common. Immunoglobulin-deficient patients also may develop osteomyelitis and an erysipelas-like rash or cellulitis.

Local suppurative complications of infection include cholecystitis, pancreatitis, and cystitis; distant complications include meningitis, endocarditis, and endovasculitis (leading to mycotic aortic aneurysm), osteoarticular infection, peritonitis, cellulitis, and septic abortion. All these complications are rare, except in immunocompromised or elderly hosts. Hepatitis, interstitial nephritis, and the hemolytic-uremic syndrome occasionally complicate acute infection. The two most common postinfectious sequelae are reactive arthritis and Guillain-Barré syndrome. Reactive arthritis has been reported in up to 2.5% of cases, although nonspecific rheumatologic symptoms are more common (~10%). Reactive arthritis may develop several weeks after infection; population-based analysis shows that there is no association with the HLA-B27 phenotype. The knees are most frequently involved, but involvement of the ankles, wrists, and small joints of the hands is common, with an average of 3.2 joints affected. Guillain-Barré syndrome or its Miller Fisher (cranial polyneuropathy) variant follow either symptomatic or asymptomatic *Campylobacter* infections uncommonly—i.e., in 1 of every 1000–2000 cases or, for certain *C. jejuni* serotypes (such as O19), in 1 of every 100–200 cases. Despite the low frequency of this complication, it is estimated that *Campylobacter* infections, because of their high incidence, may trigger 20–40% of all cases of Guillain-Barré syndrome. The presence of sialylated lipopolysaccharides on *C. jejuni* strains prompts a form of molecular mimicry that promotes autoimmune recognition of sialylated cell-surface molecules on axons. Immunoproliferative small-intestinal disease (*alpha chain disease*), a form of lymphoma that originates in small-intestinal mucosa-associated lymphoid tissue (MALToma), has been associated with *C. jejuni*; antimicrobial therapy has led to marked clinical improvement.

DIAGNOSIS

In patients with *Campylobacter* enteritis, peripheral leukocyte counts reflect the severity of the inflammatory process. In addition, stools from nearly all *Campylobacter*-infected patients presenting for medical attention in the United States contain leukocytes or erythrocytes. Gram- or Wright-stained fecal smears should be examined in all suspected cases. When the diagnosis of *Campylobacter* enteritis is suspected on the basis of findings indicating inflammatory diarrhea (fever, fecal leukocytes), clinicians can ask the microbiology laboratory to attempt the visualization of organisms with characteristic vibrioid morphology by direct microscopic examination of stools with Gram's staining or to use phase-contrast or dark-field microscopy to identify the organisms' characteristic "darting" motility. Confirmation of the diagnosis of *Campylobacter* infection is based on identification of an isolate from cultures of stool, blood, or another site; specific species can be identified by matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) mass spectrometry. *Campylobacter*-specific media should be used to culture stools from all patients with inflammatory or bloody diarrhea. Because all *Campylobacter* species are fastidious, they will not be isolated unless selective media or other selective techniques are used. Failure to isolate campylobacters from stool by culture does not entirely rule out their presence. Although culture remains the diagnostic gold standard, species-specific real-time polymerase chain reaction (PCR) techniques appear more sensitive than culture. Although PCR and other culture-independent diagnostic test (CIDTs), including antigen detection tests, may detect nonviable bacteria and may be falsely positive, they are now used frequently to diagnose infection with *Campylobacter* and other enteric bacteria in clinical microbiology laboratories. The detection of the organisms in stool in the United States by culture almost always implies active or recent infection, but CIDT positivity is more questionable.

In any event, follow-up testing after the clinical resolution of an acute infection is rarely needed. *Campylobacter sputorum, C. concisus,* and related organisms ubiquitously found in the oral cavity are commensals that only rarely have pathogenic significance. In patients with microscopic colitis, *C. concisus* can be more frequently detected than in controls, but whether detection of this common oral organism is a marker for colonic disease or is involved in pathogenesis is uncertain. Because of the low levels of metabolic activity of *Campylobacter* species in standard blood culture media, *Campylobacter* bacteremia may be difficult to detect.

DIFFERENTIAL DIAGNOSIS

The symptoms of *Campylobacter* enteritis are not sufficiently unusual to distinguish this illness from that due to *Salmonella, Shigella, Yersinia*, enterohemorrhagic *Escherichia coli*, and other pathogens. The combination of fever and fecal leukocytes or erythrocytes is indicative of inflammatory diarrhea, and definitive diagnosis is based on culture, CIDTs, or demonstration of the characteristic organisms on stained fecal smears. Extraintestinal *Campylobacter* illness is diagnosed by culture. Infection due to *Campylobacter* should be suspected in the setting of septic abortion, and that due to *C. fetus* should be suspected specifically in the setting of septic thrombophlebitis. It is important to reiterate that (1) the presentation of *Campylobacter* enteritis may mimic that of ulcerative colitis or Crohn's disease, (2) *Campylobacter* enteritis is much more common than either of the latter (especially among young adults), and (3) biopsy may not distinguish among these entities. Thus, a diagnosis of inflammatory bowel disease should not be made until *Campylobacter* infection has been ruled out, especially in persons with a history of foreign travel, significant animal contact, immunodeficiency, or exposure incurring a high risk of transmission.

TREATMENT

Campylobacter Infection

Fluid and electrolyte replacement is central to the treatment of diarrheal illnesses **(Chap. 138)**. Even among patients presenting for medical attention with *Campylobacter* enteritis, not all clearly benefit from specific antimicrobial therapy. Indications for therapy include high fever, bloody diarrhea, severe diarrhea, persistence for >1 week, and worsening of symptoms. A 3-day course of azithromycin (500 mg once daily) is the regimen of choice. A 1-day regimen of azithromycin (1000 mg given as two 500-mg tablets) can also be used. Alternative regimens for adults consist of fluoroquinolones—ciprofloxacin (500 mg by mouth twice daily for 3 days) or levofloxacin (750 mg daily for 3 days)—but resistance to this class of agents as well as to tetracyclines is substantial; ~27% of U.S. human isolates of *Campylobacter* in 2014 were resistant to ciprofloxacin, and rates are higher in many other countries; thus, travel-related *Campylobacter* infections should be considered a priori to be fluoroquinolone-resistant. Because macrolide resistance usually is much less common (<10%), these drugs are the empirical agents of choice. Patients infected with antibiotic-resistant strains are at increased risk of adverse outcomes. Use of antimotility agents, which may prolong the duration of symptoms and have been associated with toxic megacolon and with death, is not recommended. Of note, *C. jejuni* and *C. coli* are resistant to trimethoprim and β-lactam antibiotics, including penicillin and most cephalosporins.

For patients with immunocompromising conditions and uncomplicated enteritis caused by *C. jejuni*, therapy duration should be extended to 7–14 days. Bacteremic infections should always be treated with appropriate antibiotics to reduce the high rate of mortality associated with these infections. For bacteremic or other systemic infections, treatment with a carbapenem (imipenem, 500 mg IV every 6 h; or meropenem, 1–2 g IV every 8 h) should be started empirically, and susceptibility testing should always be performed. For life-threatening illness, gentamicin (1.0–1.7 mg/kg IV every 8 h after a loading dose of 1.5–2 mg/kg) can be added. In the absence of endovascular involvement, therapy for systemic infections should be administered for 7–14 days. For immunocompromised patients with systemic infections due to *C. fetus* and for patients with endovascular infections due to any species, prolonged therapy (up to 4 weeks) is usually necessary. For recurrent infections in immunocompromised hosts, lifelong therapy/prophylaxis is sometimes necessary.

PROGNOSIS

Nearly all patients recover fully from *Campylobacter* enteritis, either spontaneously or after antimicrobial therapy. Volume depletion probably contributes to the few deaths that are reported. As stated above, occasional patients develop reactive arthritis or Guillain-Barré syndrome or its variants. Systemic infection with *C. fetus* is much more often fatal than that due to related species; this higher mortality rate reflects in part the population affected. Prognosis depends on the rapidity with which appropriate therapy is begun. Otherwise healthy hosts usually survive *C. fetus* infections without sequelae. Compromised hosts often have recurrent and/or life-threatening infections due to a variety of *Campylobacter* species.

FURTHER READING

Amour C et al: Epidemiology and impact of *Campylobacter* infection in children in 8 low-resource settings: Results from the MAL-ED Study. Clin Infect Dis 63:1171, 2016.

Costa D, Iraola G: Pathogenomics of emerging *Campylobacter* Species. Clin Microbiol Rev 32:e00072, 2019.

Dai L et al: New and alternative strategies for the prevention, control, and treatment of antibiotic-resistant *Campylobacter*. Transl Res 223:76, 2020.

Man SM: The clinical importance of emerging *Campylobacter* species. Nat Rev Gastroenterol Hepatol 8:669, 2011.

Marder EP et al: Incidence and trends of infections with pathogens transmitted commonly through food and the effect of increasing use of culture-independent diagnostic tests on surveillance—Foodborne Diseases Active Surveillance Network, 10 U.S. sites, 2013–2016. Morb Mortal Wkly Rep 66:397, 2017.

Nielsen HL: High risk of microscopic colitis after *Campylobacter concisus* infection: Population-based cohort study. Gut 69:1952, 2020.

Riddle MS et al: ACG clinical guideline: Diagnosis, treatment, and prevention of acute diarrheal infections in adults. Am J Gastroenterol 111:602, 2016.

Rouhani S et al: Gut microbiota features associated with *Campylobacter* burden and postnatal linear growth deficits in a Peruvian birth cohort. Clin Infect Dis 71:1000, 2020.

Same RG, Tamma PD: *Campylobacter jejuni* infections in children. Pediatr Rev 39:533, 2018.

Ternhag A et al: A meta-analysis of the effects of antibiotic treatment on duration of symptoms caused by infection with *Campylobacter* species. Clin Infect Dis 44:696, 2007.

Tinévez C et al: Retrospective multicentric study on *Campylobacter* spp. bacteremia in France: The campylobacteremia study. Clin Infect Dis 75:702, 2022.

Watkins LKF et al: Ongoing outbreak of extensively drug-resistant *Campylobacter jejuni* infections associated with US pet store puppies, 2016-2020. JAMA Netw Open 4:e2125203, 2021.

170

Infections Due to *Pseudomonas, Burkholderia,* and *Stenotrophomonas* Species

Kartikeya Cherabuddi, Reuben Ramphal

The pseudomonads are a heterogeneous group of gram-negative bacteria that have in common an inability to ferment lactose. Formerly classified in the genus *Pseudomonas*, the members of this group have been assigned to three medically important genera—*Pseudomonas, Burkholderia*, and *Stenotrophomonas*—whose biologic behaviors encompass both similarities and marked differences and whose genetic repertoires differ in many respects. The pathogenicity of most pseudomonads is based on opportunism; the exceptions are *Burkholderia pseudomallei* and *Burkholderia mallei*, which are primary pathogens.

The genus *Pseudomonas* now contains >140 species. *Pseudomonas aeruginosa*, the major pathogen of the group, is a significant cause of infections in hospitalized patients and in patients with cystic fibrosis (CF; **Chap. 302**). Cytotoxic chemotherapy, mechanical ventilation, chronic lung diseases, and broad-spectrum antibiotic therapy set up conditions that predispose to colonization and infection of increasing numbers of hospitalized patients by this pathogen. Other significant members of the genus—*Pseudomonas putida, Pseudomonas fluorescens, Pseudomonas oryzihabitans*, and *Pseudomonas stutzeri*—infect humans infrequently and are generally opportunists that are always present in the environment.

The genus *Burkholderia* comprises >20 species, of which *Burkholderia cepacia* is most frequently encountered in Western countries. Similar to *P. aeruginosa, B. cepacia* (now referred to as the *B. cepacia* complex species) is both an opportunistic nosocomial pathogen and a cause of infection in CF. The other medically important members of this genus are *B. pseudomallei* and *B. mallei*, the etiologic agents of melioidosis and glanders, respectively.

The genus *Stenotrophomonas* contains one species of medical significance, *Stenotrophomonas maltophilia*. This organism is strictly an opportunist that "overgrows" in the setting of broad-spectrum antibiotic use.

PSEUDOMONAS AERUGINOSA

EPIDEMIOLOGY

P. aeruginosa is found in most moist environments. Soil, plants, vegetables, tap water, and countertops are all potential reservoirs for this microbe, as it has simple nutritional needs. Given the ubiquity of *P. aeruginosa*, it is clear that simple contact with the organism is not sufficient for colonization or infection. Clinical and experimental observations suggest that infection by *P. aeruginosa* occurs concomitantly with compromised host defenses, mucosal trauma, physiologic derangement, and antibiotic-mediated suppression of normal flora. Thus, it comes as no surprise that the majority of *P. aeruginosa* infections occur in intensive care units (ICUs), where these factors frequently converge. Although the organism is initially acquired from environmental sources, patient-to-patient spread occurs in CF clinics and may occur in closed hospital units.

In the past, patients with burns appeared to be unusually susceptible to *P. aeruginosa*. For example, in 1959–1963, *Pseudomonas* burn-wound sepsis was the principal cause of death in 60% of patients with burns dying at the U.S. Army Institute of Surgical Research. For reasons that are unclear, *P. aeruginosa* infection in burns is no longer the major problem that it was during the 1950s and 1960s. Similarly, in the 1960s, *P. aeruginosa* appeared as a common pathogen in patients receiving cytotoxic chemotherapy at many institutions in the United States, but it has subsequently diminished in importance. Despite this subsidence, *P. aeruginosa* remains one of the most feared pathogens in this population because of its high attributable mortality.

In some parts of Asia and Latin America, *P. aeruginosa* continues to be the most common cause of gram-negative bacteremia in neutropenic patients.

In contrast to the trends for patients with burns or neutropenia in the United States, the incidence of *P. aeruginosa* infections among patients with CF has not changed. *P. aeruginosa* remains the most common contributing factor to respiratory failure in CF and is responsible for the majority of deaths among CF patients.

LABORATORY FEATURES

P. aeruginosa is a nonfastidious, motile, gram-negative rod that grows on most common laboratory media, including blood and MacConkey agars. It is easily identified in the laboratory on primary-isolation agar plates by pigment production that confers a yellow to dark green or even bluish appearance. Colonies have a shiny "gun-metal" appearance and a characteristic fruity odor. Two of the identifying biochemical characteristics of *P. aeruginosa* are an inability to ferment lactose on MacConkey agar and a positive reaction in the oxidase test. Most strains are identified on the basis of these readily detectable laboratory features even before extensive biochemical testing is done. Some isolates from CF patients are easily identified by their mucoid appearance, which is due to the production of large amounts of the mucoid exopolysaccharide or alginate. Recently, there has been increasing use of molecular testing with multiplex polymerase chain reaction (PCR) platforms, which rapidly identify *P. aeruginosa* in respiratory and blood samples much earlier than the classical methods.

PATHOGENESIS

Unraveling the mechanisms that underlie disease caused by *P. aeruginosa* has proved challenging. Of the common gram-negative bacteria, no other species produces such a large number of putative virulence factors **(Table 170-1)**. Yet *P. aeruginosa* rarely initiates an infectious process in the absence of host injury or compromise, and few of its putative virulence factors have been shown definitively to be involved in disease in humans. Despite its metabolic versatility and possession of multiple colonizing factors, *P. aeruginosa* exhibits no competitive advantage over enteric bacteria in the human gut; it is not a normal inhabitant of the healthy human gastrointestinal tract, despite the host's continuous environmental exposure to the organism.

Virulence Attributes Involved in Acute *P. aeruginosa* Infections • **MOTILITY AND COLONIZATION** A general tenet of bacterial pathogenesis is that most bacteria must adhere to surfaces or colonize a host niche in order to initiate disease. Most gram-negative bacteria examined thus far possess adherence factors called *adhesins*. *P. aeruginosa* is no exception. Among its many adhesins are its pili, which demonstrate adhesive properties for a variety of cells and adhere best to injured cell surfaces. In the organism's flagellum, the flagellin molecule binds to cells, and the flagellar cap attaches to mucins through the recognition of glycan chains. Other *P. aeruginosa* adhesins

TABLE 170-1 Main Putative Virulence Factors of *Pseudomonas aeruginosa*

SUBSTANCE/ ORGANELLE	FUNCTION	VIRULENCE IN ANIMAL DISEASE
Pili	Adhesion to cells	?
Flagella	Adhesion, motility, inflammation	Yes
Lipopolysaccharide	Antiphagocytic activity, inflammation	Yes
Type III secretion system	Toxic activity (ExoU, ExoS)	Yes
Type II secretion system	Toxic activity	Yes
Proteases	Proteolytic activity	?
Phospholipases	Cytotoxicity	?
Exotoxin A	Cytotoxicity	?
Pyocyanin	Cytotoxicity	Yes

include the outer core of the lipopolysaccharide (LPS) molecule, which binds to the cystic fibrosis transmembrane conductance regulator (CFTR) and aids in internalization of the organism, and the alginate coat of mucoid strains, which enhances adhesion to cells and mucins. In addition, membrane proteins and lectins have been proposed as colonization factors. The deletion of any given adhesin by itself is not sufficient to abrogate the ability of *P. aeruginosa* to colonize surfaces probably because of the redundancy of adhesins. Motility is important in host invasion via mucosal surfaces or burned skin in animal models of infection; however, nonmotile strains are not uniformly avirulent. It has been well demonstrated that nonmotile strains of *P. aeruginosa* are poorly phagocytosed in vitro, possibly leading to enhancement of the virulence of this organism in vivo.

EVASION OF HOST DEFENSES The transition from bacterial colonization to disease requires the evasion of host defenses followed by invasion by the microorganism. *P. aeruginosa* appears to be well equipped for evasion. Attached bacteria inject four known toxins (ExoS or ExoU, ExoT, and ExoY) via a type III secretion system that allows the bacteria to evade phagocytic cells either by direct cytotoxicity or by inhibition of phagocytosis. Clinical studies suggest that the mortality rate is higher among patients infected by strains that secrete the ExoU toxin. Another secretion system—the type II system—secretes toxins that can kill animals, and some of its secreted toxins, such as exotoxin A, have the potential to kill phagocytic cells. Multiple proteases secreted by this system may degrade host effector molecules, such as cytokines and chemokines, that are released in response to infection and appear to play a role in corneal infections in mice.

TISSUE INJURY Among gram-negative bacteria, *P. aeruginosa* probably produces the largest number of substances that are toxic to cells and thus have the potential to injure tissues. The toxins secreted by the organism's type III secretion system are capable of injuring tissue. However, their delivery requires the adherence of the organism to cells. Thus, the effects of these toxins are likely to be local or to depend on the presence of large numbers of bacteria at the site of an infection or in the bloodstream. On the other hand, diffusible toxins, secreted by the organism's type II secretion system, can act freely wherever they come into contact with cells. Possible effectors of this system include exotoxin A, at least four different proteases, and at least two phospholipases. In addition to these secreted toxins, rhamnolipids, pyocyanins—the pigments that confer the characteristic color and odor of *P. aeruginosa* colonies—and hydrocyanic acid, are produced by *P. aeruginosa* and are all capable of causing host tissue injury and even neutrophil death.

INFLAMMATORY COMPONENTS The inflammatory responses to the lipid A component of *Pseudomonas* LPS and to its flagellin, mediated through the Toll-like receptor (TLR) system (principally TLR4 and TLR5, respectively), are thought to represent important factors in disease causation. Although these inflammatory responses are required for successful defense against *P. aeruginosa* (i.e., in their absence, animals are defenseless against *P. aeruginosa* infection), florid responses are likely to result in severe disease. Thus, when the sepsis syndrome and septic shock develop in *P. aeruginosa* infection, they are probably the result of the host response to one or both of these substances, but injury to the lung by *Pseudomonas* toxins may also result in sepsis syndromes, possibly by causing cell death and the release of cellular components (e.g., heat-shock proteins) that may activate the TLR or another proinflammatory system. Thus, the virulence of this bacterium in acute infections is likely to be multifactorial with a great redundancy of effector molecules being produced.

Chronic *P. aeruginosa* Infections Chronic infection due to *P. aeruginosa* occurs mainly in the lungs in the setting of structural pulmonary diseases. The classic example is CF; others include bronchiectasis and chronic relapsing panbronchiolitis, a disease seen in Japan and some Pacific Islands. A hallmark of these illnesses is severely defective mucociliary clearance leading to mucus stasis and mucus accumulation in the lungs. There is probably a common factor that selects for *P. aeruginosa* colonization in these lung diseases—perhaps the adhesiveness of *P. aeruginosa* for mucus, a phenomenon that is not noted for most other common gram-negative bacteria, and/or the ability of *P. aeruginosa* to evade host defenses in mucus. Furthermore, *P. aeruginosa* undergoes evolutionary adaptations and diversification in ways that allow its prolonged survival in the lung without an early fatal outcome for the host. The strains found in CF patients exhibit minimal production of virulence factors. Many strains lose the ability to produce pili and flagella, and most become complement-sensitive because of the loss of the O side chain of their LPS molecules. In addition, most strains found in CF patients overproduce a mucoid exopolysaccharide. These changes probably dampen the host response, allowing the organism to survive in CF mucus. *P. aeruginosa* is also believed to lose its ability to secrete many of its injectable toxins during growth in mucus. Although the alginate coat is thought to play a role in the organism's survival, alginate is not essential as nonmucoid strains may predominate for long periods. In short, virulence in chronic infections may be mediated by the chronic but attenuated host inflammatory response, which injures the lungs over decades.

CLINICAL MANIFESTATIONS

P. aeruginosa causes infections at almost all sites in the body but shows a rather marked predilection for the lungs. The infections encountered most commonly in hospitalized patients are described below.

Bacteremia Crude mortality rates exceeding 50% have been reported among patients with *P. aeruginosa* bacteremia. Consequently, this clinical entity has been much feared, and its management has been attempted with the use of multiple antibiotics. Recent publications report attributable mortality rates of 28–44%, with the precise figure depending on the adequacy and timing of treatment and the seriousness of the underlying disease. In the past, the patient with *P. aeruginosa* bacteremia classically was neutropenic or had a burn injury. Today, however, a minority of such patients have bacteremic *P. aeruginosa* infections. Rather, *P. aeruginosa* bacteremia is seen most often in patients in ICUs, with the lungs, the urinary tract, central venous lines, or wounds being the most important portals for systemic invasion.

The clinical presentation of *P. aeruginosa* bacteremia rarely differs from that of sepsis in general **(Chap. 315)**. Patients are usually febrile, but those who are most severely ill may be in shock or even hypothermic. The only point differentiating this entity from gram-negative sepsis due to other bacteria may be the distinctive skin lesions (ecthyma gangrenosum) of *Pseudomonas* infection, which occur almost exclusively in markedly neutropenic patients and patients with AIDS. These small or large, painful, reddish, maculopapular lesions have a geographic margin; they are initially pink, then darken to purple, and finally become black and necrotic **(Fig. 170-1)**. Histopathologic studies indicate that the lesions are due to vascular invasion and are teeming with bacteria. Although similar lesions may occur in aspergillosis, mucormycosis, and occasionally *Staphylococcus aureus* bacteremia, their presence in a neutropenic patient generally suggests *P. aeruginosa* bacteremia as the most likely cause.

TREATMENT

P. aeruginosa Bacteremia

(Table 170-2) Antimicrobial treatment of *P. aeruginosa* bacteremia has been controversial. Combination therapy with an

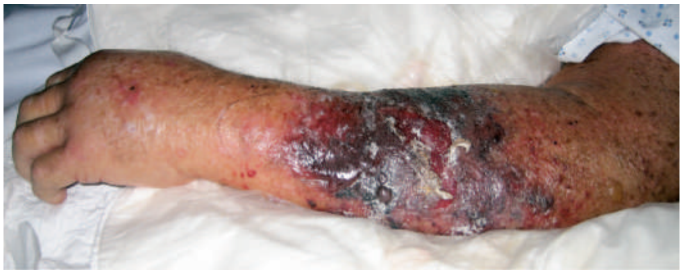

FIGURE 170-1 Ecthyma gangrenosum in a neutropenic patient 3 days after onset.

TABLE 170-2 Antibiotic Treatment of Infections Due to *Pseudomonas aeruginosa* and Related Species

INFECTION	ANTIBIOTICS AND DOSAGES	OTHER CONSIDERATIONS
Bacteremia		
Nonneutropenic host	Ceftazidime (2 g q8h IV) *or* cefepime (2 g q8h IV) *or* piperacillin/tazobactam (4.5 g q6h IV) *or* imipenem (500 mg q6h IV) *or* meropenem (1–2 g q8h IV) *or* doripenem (500 mg q8h IV) ***Optional:*** Amikacin (7.5 mg/kg q12h or 15 mg/kg q24h IV)	Add an aminoglycoside empirically for patients in shock and in regions or hospitals where rates of resistance to the primary β-lactam agents are high. Tobramycin may be used instead of amikacin (susceptibility permitting). A duration of 6–10 days of therapy can be used for uncomplicated bacteremia.
Neutropenic host	Cefepime (2 g q8h IV) *or* any of the other agents above (except doripenem) in the above dosages	Febrile neutropenic patients should be treated until no longer neutropenic.
Endocarditis	Antibiotic regimens as for bacteremia for 6–8 weeks	Resistance during therapy is common. Surgery is required for relapse.
Pneumonia	Drugs and dosages as for bacteremia, except that the available carbapenems should not be the sole primary drugs because of high rates of resistance during therapy.	Add aminoglycoside or ciprofloxacin, as for bacteremia, until sensitivities available. The duration of therapy is 7 days.
Bone infection, malignant otitis externa	Cefepime or ceftazidime at the same dosages as for bacteremia; aminoglycosides not a necessary component of therapy; ciprofloxacin (500–750 mg q12h PO) may be used	Duration of therapy varies with the drug used and type of infection (e.g., 6 weeks for a β-lactam agent; except in puncture-wound osteomyelitis, for which the duration should be 2–4 weeks; oral therapy can be used).
Central nervous system infection	Ceftazidime or cefepime (2 g q8h IV) *or* meropenem (2 g q8h IV)	Abscesses or other closed-space infections may require drainage. The duration of therapy is ≥2 weeks.
Eye infection		
Keratitis/ulcer	Topical therapy with tobramycin/ciprofloxacin/levofloxacin eyedrops	Use maximal strengths available or compounded by pharmacy. Therapy should be administered for 2 weeks or until the resolution of eye lesions, whichever is shorter.
Endophthalmitis	Ceftazidime or cefepime as for central nervous system infection ***plus*** Topical therapy	
Urinary tract infection (UTI)	Ciprofloxacin (500 mg q12h PO) *or* levofloxacin (750 mg q24h) *or* any aminoglycoside (total daily dose given once daily). Cefepime or ceftazidime (1g q8h) *or* piperacillin/tazobactam (3.375 g q6h)	Uncomplicated cystitis may be treated for 3 days with oral agents. Relapse may occur if an obstruction or a foreign body is present. The duration of therapy for complicated cystitis and uncomplicated pyelonephritis is 5–7 days.
Multidrug- and extreme drug-resistant *P. aeruginosa* infection	Ceftazidime/avibactam (2.5 g q8h, infused over 2 h) *or* ceftolozane/tazobactam (1.5–3 g q8h) *or* imipenem/relebactam (500 mg q6h) *or* cefiderocol (2 g q8h) *or* colistin (100 mg q12h IV for the shortest possible period to obtain a clinical response)	Use 3-g dose of ceftolozane/tazobactam for pneumonia. Meropenem-vaborbactam offers minimal benefit in carbapenem resistant strains. Alternatives to colistin are preferred, if available. Colistin dosing requires renal adjustment and expertise in its use. Inhaled colistin may be added for pneumonia (100 mg q12h).
Burkholderia cepacia complex infection	Meropenem (2 g q8h IV) *or* TMP-SMX (1600/320 mg q12h IV) for 14 days	Resistance to both agents is increasing. Do not use them in combination because of possible antagonism.
Melioidosis (*B. pseudomallei*), glanders (*B. mallei*)	Ceftazidime (2 g q6h) *or* meropenem (1 g q8h) *or* imipenem (500 mg q6h) for 2 weeks ***followed by*** TMP-SMX (1600/320 mg q12h PO) for 3 months	
Stenotrophomonas maltophilia infection	TMP-SMX (1600/320 mg q12h IV) *plus* either levofloxacin (750 mg q24h) *or* minocycline (100-200 mg q12h) or ticarcillin/clavulanate (3.1 g q4h IV) for 7 to 14 days	Broad-spectrum antibiotic therapy leads to respiratory tract colonization and often warrants no treatment. Ceftazidime-avibactam plus aztreonam, cefiderocol, or tigecycline are alternatives for XDR strains. Combination therapy should be used for bacteremia, especially in immunosuppressed patients.

Abbreviations: MIC, minimum inhibitory concentration; TMP-SMX, trimethoprim-sulfamethoxazole.

antipseudomonal β-lactam and an aminoglycoside became the standard of care because of the dismal outcome of single-drug therapy, mainly with aminoglycosides and polymixins, prior to 1971—first for *P. aeruginosa* bacteremia in febrile neutropenic patients and then extrapolated to all *P. aeruginosa* bacteremic infections in both neutropenic and nonneutropenic patients.

Following the introduction of new antipseudomonal drugs, a number of studies have revisited the choice between combination treatment and monotherapy for *Pseudomonas* bacteremia. Although some clinicians still favor combination therapy, most recent observational studies indicate that a single modern antipseudomonal β-lactam agent to which the isolate is sensitive is as efficacious as a combination. Even in patients at greatest risk of early death from *P. aeruginosa* bacteremia (i.e., those with fever and neutropenia), empirical antipseudomonal monotherapy is deemed to be as efficacious as empirical combination therapy by the practice guidelines of the Infectious Diseases Society of America (IDSA). One firm conclusion is that monotherapy with an aminoglycoside is not optimal.

There are, of course, institutions and countries where rates of susceptibility of *P. aeruginosa* to first-line antibiotics are <80%. Thus, when a septic patient with a high probability of *P. aeruginosa* infection is encountered in such settings, empirical combination therapy should be administered until the pathogen is identified and susceptibility data become available. Thereafter, whether one or two agents should be continued remains a matter of individual preference. Recent studies suggest that extended or continuous infusions of β-lactams such as cefepime, piperacillin-tazobactam, or meropenem may result in better outcomes of *Pseudomonas* bacteremia and possibly of *Pseudomonas* pneumonia. The duration of antibiotic therapy has now become an important consideration due to the increasing isolation of multiple drug-resistant (MDR) and extensively drug-resistant (XDR) *P. aeruginosa* strains. Recently published studies now strongly support the use of shorter courses of therapy (7 days) rather than the longer duration (10–14 days) that is commonly recommended for many cases of *Pseudomonas* bacteremia. As in *S. aureus* bacteremia, catheter removal is important.

Acute Pneumonia Respiratory infections are the most common of all infections caused by *P. aeruginosa*. *P. aeruginosa* is common in both hospital-acquired pneumonia (HAP) and ventilator-associated pneumonia (VAP). This organism appears first or second among the causes of VAP. However, much debate centers on the actual role of *P. aeruginosa* in VAP. Many of the relevant data are based on cultures of sputum or endotracheal tube aspirates and may represent nonpathogenic colonization of the tracheobronchial tree, biofilms on the endotracheal tube, or simple tracheobronchitis. The increasing use of PCR testing on endotracheal samples has further compounded this issue. PCR coupled with the widespread use of computed tomography with poor specificity for pneumonia, and overlap in appearance with fluid overload in patients with low to intermediate clinical suspicion, may contribute to the overdiagnosis of HAP and VAP. This scenario emphasizes the importance of clinical judgment in adjudicating *P. aeruginosa* colonization versus infection.

Older reports of *P. aeruginosa* pneumonia described patients with an acute clinical syndrome of fever, chills, cough, and necrotizing pneumonia indistinguishable from other gram-negative bacterial pneumonias. The traditional accounts described a fulminant infection. Chest radiographs demonstrated bilateral pneumonia, often with nodular densities with or without cavities. This picture is now remarkably rare. Today, the typical patient is on a ventilator, has a slowly progressive infiltrate, and has been colonized with *P. aeruginosa* for days. While some cases may progress rapidly over 48–72 h, they are the exceptions. Nodular densities are not commonly seen. However, infiltrates may go on to necrosis. Necrotizing pneumonia has also been seen in the community (e.g., after inhalation of hot-tub water contaminated with *P. aeruginosa*). The typical patient has fever, leukocytosis, tachypnea, hypoxemia, and purulent sputum, and the chest radiograph shows a new infiltrate or the expansion of a preexisting infiltrate. A sputum Gram's stain showing mainly polymorphonuclear leukocytes (PMNs) in conjunction with a culture positive for *P. aeruginosa* in this setting suggests a diagnosis of acute *P. aeruginosa* pneumonia.

There have been increasing reports of the occurrence of community-acquired *P. aeruginosa* pneumonia among patients with underlying lung diseases. While this undoubtedly occurs, it is difficult to make this diagnosis with a great degree of certainty with the use of sputum cultures in a population prone to airway colonization by multiple strains of bacteria. The patient population in whom the possibility of a community-acquired *P. aeruginosa* pneumonia should be considered is the neutropenic patient, given the pivotal role that neutrophils play in defense against this bacterium. Such a patient, whether hospitalized or admitted from the community with a pneumonia, should be treated empirically for *P. aeruginosa*.

TREATMENT

Acute Pneumonia

(Table 170-2) Therapy for *P. aeruginosa* pneumonia remains unsatisfactory. Reports suggest mortality rates of 40–80%, but how many of these deaths are attributable to underlying disease remains unknown. The drugs of choice for *P. aeruginosa* pneumonia are similar to those given for bacteremia. A potent antipseudomonal β-lactam drug is the mainstay of therapy. Failure rates were high when aminoglycosides were used as single agents, possibly because of their poor penetration into the airways and their binding to airway secretions. Nonetheless, for the treatment of patients at high risk of death, some experts suggest the combination of a β-lactam agent and an antipseudomonal fluoroquinolone or aminoglycoside. As for the duration of therapy, recent IDSA/American Thoracic Society (ATS) guidelines recommend 7 days of treatment for HAP or VAP, even when *P. aeruginosa* is the offending organism. However, the outcome in neutropenic patients is poor, especially if accompanied by bacteremia; thus, therapy needs to be extended until neutropenia resolves. In addition, therapy longer than 7 days should be used in patients with *P. aeruginosa* necrotizing pneumonia as, functionally, this entity is similar to a lung abscess.

Chronic Respiratory Tract Infections *P. aeruginosa* is responsible for chronic infections of the airways associated with a number of underlying or predisposing conditions—most commonly CF **(Chap. 302)**. A state of chronic colonization beginning early in childhood is seen in some Asian populations with chronic or diffuse panbronchiolitis, a disease of unknown etiology. *P. aeruginosa* is one of the organisms that colonizes damaged bronchi in bronchiectasis, a disease secondary to multiple causes in which profound structural abnormalities of the airways result in mucus stasis.

TREATMENT

Chronic Respiratory Tract Infections

Optimal management of chronic *P. aeruginosa* lung infection has not been determined. Patients respond clinically to antipseudomonal therapy, but the organism is rarely eradicated. Because eradication is unlikely, the aim of treatment for chronic infection is to quell exacerbations of inflammation. The regimens used are similar to those used for pneumonia, but an aminoglycoside is almost always added because resistance is common in chronic disease. However, it may be more appropriate to use an inhaled aminoglycoside preparation in order to maximize airway drug levels. MDR strains are now commonly found in such patients given their increased life span and the repeated courses of antibiotics they receive.

Endovascular Infections Infective endocarditis of native valves due to *P. aeruginosa* is most commonly seen in those who use IV drugs. This organism has also been reported to cause prosthetic-valve endocarditis. Sites of prior native-valve injury due to the injection of foreign material such as talc or fibers probably serve as niduses for bacterial attachment to the heart valve. The manifestations of *P. aeruginosa* endocarditis resemble those of other forms of endocarditis in those who use IV drugs except that the disease is more indolent than *S. aureus* endocarditis. While most disease involves the right side of the heart, left-sided involvement is not rare, and multivalvular disease is common. Fever is a common manifestation, as is pulmonary involvement (due to septic emboli to the lungs). Thus, patients may also experience chest pain and hemoptysis. Involvement of the left side of the heart may lead to signs of cardiac failure, systemic emboli, and local cardiac involvement with sinus of Valsalva abscesses and conduction defects. Skin manifestations other than injection site infections are rare in this disease, and ecthyma gangrenosum is not common. Vertebral osteomyelitis and sternoclavicular joint septic arthritis are uncommon but pathognomic complications of this disease. The diagnosis is based on positive blood cultures along with clinical signs of endocarditis.

TREATMENT

Endovascular Infections

(Table 170-2) It has been customary to use synergistic antibiotic combinations in treating *P. aeruginosa* endocarditis because of the development of resistance during therapy with a single antipseudomonal β-lactam agent. Which combination therapy is preferable is unclear, as all combinations have failed. Treatment is likely to more often be successful in cases of right-sided endocarditis. Cases of *P. aeruginosa* endocarditis that relapse during or fail to respond to therapy are often caused by resistant organisms and may require surgical therapy. Other considerations for valve replacement are similar to those in other forms of endocarditis **(Chap. 133)**.

Bone and Joint Infections *P. aeruginosa* is an infrequent cause of bone and joint infections. However, *Pseudomonas* bacteremia or infective endocarditis caused by the injection of contaminated illicit drugs has been documented to result in vertebral osteomyelitis and sternoclavicular joint arthritis. The clinical presentation of vertebral *P. aeruginosa* osteomyelitis is more indolent than that of staphylococcal osteomyelitis. The duration of symptoms in IV drug users with vertebral osteomyelitis due to *P. aeruginosa* varies from weeks to months. Fever is not uniformly present; when present, it tends to be low grade. There may be mild tenderness at the site of involvement. Blood cultures are usually negative unless there is concomitant endocarditis. The erythrocyte sedimentation rate (ESR) is generally elevated. Vertebral osteomyelitis due to *P. aeruginosa* has also been reported in the elderly, in whom it originates from urinary tract infections (UTIs). The infection generally involves the lumbosacral area because of a shared venous drainage (Batson's plexus) between the lumbosacral spine and the pelvis. Sternoclavicular septic arthritis due to *P. aeruginosa* is seen almost exclusively in persons who use IV drugs. This disease may occur with or without endocarditis, and a primary site of infection often is not found. Plain radiographs show joint or bone involvement. Treatment of these forms of disease is generally successful.

Pseudomonas osteomyelitis of the foot most often follows puncture wounds through sneakers and mostly affects children. The main manifestation is pain in the foot, sometimes with superficial cellulitis around the puncture wound and tenderness on deep palpation of the wound. Multiple joints or bones of the foot may be involved. Systemic symptoms are generally absent, and blood cultures are usually negative. Radiographs may or may not be abnormal, but the bone scan is usually positive, as are magnetic resonance imaging (MRI) studies. Needle aspiration usually yields a diagnosis. Prompt surgery, with exploration of the nail puncture tract and debridement of the involved bones and cartilage, is generally recommended in addition to antibiotic therapy.

Osteomyelitis due to *P. aeruginosa* is also seen following trauma and with decubitus ulcers. In these settings, the cause of osteomyelitis is often polymicrobial, and the role of *P. aeruginosa* can be questioned. It is therefore critical that deep bone biopsies be requested to ascertain its significance prior to starting treatment that targets *P. aeruginosa*.

TREATMENT

Bone and Joint Infections

The treatment of bone and joint infections due to *P. aeruginosa* is often governed by the primary *Pseudomonas* infection. Since endocarditis is often the primary infection, the agents used for endocarditis will dictate treatment. In other situations, a 6-week course of therapy with an antipseudomonal β-lactam is recommended, and in case of puncture-wound osteomyelitis, oral ciprofloxacin may be used.

Central Nervous System (CNS) Infections CNS infections due to *P. aeruginosa* are relatively rare. Involvement of the CNS is almost always secondary to a surgical procedure, head trauma, implanted devices, temporary external ventricular drains, and rarely bacteremia. The entity seen most often is postoperative or posttraumatic meningitis. Subdural or epidural infection occasionally results from contamination of these areas. Embolic disease arising from endocarditis in users of IV drugs and leading to brain abscesses has also been described. The cerebrospinal fluid (CSF) profile of *P. aeruginosa* meningitis is no different from that of pyogenic meningitis of any other etiology.

TREATMENT

Central Nervous System Infections

(Table 170-2) Treatment of *Pseudomonas* meningitis is difficult; little information has been published. However, the general principles involved in the treatment of meningitis apply, including the need for high doses of bactericidal antibiotics to attain high drug levels in the CSF. The agent with which there is the most published experience in *P. aeruginosa* meningitis is ceftazidime, but other antipseudomonal β-lactam drugs that attain reasonable CSF concentrations, such as cefepime, piperacillin/tazobactam, and meropenem, have also been used successfully. Other forms of *P. aeruginosa* CNS infection, such as brain abscesses and epidural and subdural empyema, generally require surgical drainage and removal of devices, in addition to antibiotic therapy.

Eye Infections Eye infections due to *P. aeruginosa* occur mainly as a result of direct inoculation into the tissue during trauma or surface injury by contact lenses. Keratitis and corneal ulcers are the most common types of eye disease and are often associated with contact lenses (especially the extended-wear variety). Keratitis can be slowly or rapidly progressive, but the classic description is disease progressing over 48 h to involve the entire cornea, with opacification and sometimes perforation. *P. aeruginosa* keratitis should be considered a medical emergency because of the rapidity with which it can progress to loss of sight. *P. aeruginosa* endophthalmitis secondary to bacteremia is the most devastating of *P. aeruginosa* eye infections. The disease is fulminant, with severe pain, chemosis, decreased visual acuity, anterior uveitis, vitreous involvement, and panophthalmitis. It is also a rare complication of cataract removal with lens insertion. In the United States, a recent outbreak associated with artificial tears with carbapenem-resistant *P. aeruginosa* led to serious eye infections and, in some cases, vision loss, enucleation, or death.

TREATMENT

Eye Infections

(Table 170-2) The usual therapy for keratitis is the administration of topical antibiotics. Therapy for endophthalmitis includes the use of high-dose local and systemic antibiotics (to achieve higher drug concentrations in the eye) and vitrectomy.

Ear Infections *P. aeruginosa* infections of the ears vary from mild swimmer's ear to serious life-threatening infections with neurologic sequelae. Swimmer's ear is common among children and results from infection of moist macerated skin of the external ear canal. Most cases resolve with treatment, but some patients develop chronic drainage. Swimmer's ear is managed with topical antibiotic agents (otic solutions). The use of hearing aids may also predispose to this type of infection. The most serious form of *Pseudomonas* infection involving the ear has been given various names: two of these designations, *malignant otitis externa* and *necrotizing otitis externa*, are now used for the same entity. This disease was originally described in elderly patients with diabetes, in whom the majority of cases still occur. However, it has also been described in patients with AIDS and in elderly patients without underlying diabetes or immunocompromise. The usual presenting symptoms are decreased hearing and ear pain, which may be severe and lancinating. The pinna is usually painful, and the external canal may be tender. The ear canal almost always shows signs of inflammation, with granulation tissue and exudate. Tenderness anterior to the tragus may extend as far as the temporomandibular joint and mastoid process. A small minority of patients have systemic symptoms. Patients in whom the diagnosis is made late may present with cranial nerve palsies, most commonly cranial nerve VII, or even with cavernous venous sinus thrombosis. The ESR is invariably elevated (≥100 mm/h). The diagnosis is made on clinical grounds in severe cases; however, the "gold standard" is a positive technetium-99 bone scan in a patient with otitis externa due to *P. aeruginosa*. In diabetic patients, a positive bone scan constitutes presumptive evidence for this diagnosis and should prompt biopsy or empirical therapy. However, it should be kept in mind that *S. aureus* and *Aspergillus* spp. can also cause this entity.

TREATMENT

Ear Infections

(Table 170-2) Given the infection of the ear cartilage, sometimes with mastoid or petrous ridge involvement, patients with malignant (necrotizing) otitis externa are treated as for osteomyelitis.

Urinary Tract Infections UTIs due to *P. aeruginosa* generally occur as a complication of a catheter in the urinary tract, an obstruction or stone in the genitourinary system, urinary tract instrumentation, or surgery. A *P. aeruginosa* UTI occurring in the community often signals the presence of an abnormality in the urinary tract. It has been reported that the urinary tract is the second most important site of infection leading to *Pseudomonas* bacteremia.

TREATMENT

Urinary Tract Infections

(Table 170-2) Most *P. aeruginosa* UTIs are considered complicated infections that must be treated longer than uncomplicated cystitis. In general, a 7- to 10-day course of treatment suffices. Urinary catheters, stents, or stones should be removed to prevent relapse, which is common and may not be due to antibiotic resistance but rather to factors such as a foreign body that has been left in place or an ongoing obstruction. Removal of a urinary catheter will allow shorter courses of antibiotic therapy if that is the only predisposing factor.

Skin and Soft Tissue Infections Besides pyoderma (ecthyma) gangrenosum in neutropenic patients, folliculitis and other papular or vesicular lesions due to *P. aeruginosa* have been extensively described and are collectively referred to as *dermatitis*. Multiple outbreaks have been linked to whirlpools, spas, and swimming pools. To prevent such outbreaks, the growth of *P. aeruginosa* in the home and in recreational environments must be controlled by proper chlorination of water. Most cases of hot-tub folliculitis are self-limited, requiring only the avoidance of exposure to the contaminated source of water. Approximately one-third of reported outbreaks are associated with hotel facilities. Patients may also have ear pain, skin rashes, and eye irritation.

Toe-web infections occur especially often in the tropics, and the "green-nail syndrome" is caused by *P. aeruginosa* paronychia, which results from frequent submersion of the hands in water. In the latter entity, the green discoloration results from diffusion of pyocyanin into the nail bed. *P. aeruginosa* remains a prominent cause of burn wound infections in some parts of the world. The management of these infections is best left to specialists in burn wound care.

Infections in Febrile Neutropenic Patients In febrile neutropenia, *P. aeruginosa* has historically been the organism against which empirical coverage is always essential. Although in Western countries these infections are now less common, their importance has not diminished because of persistently high mortality rates. In other parts of the world, *P. aeruginosa* continues to be a significant problem in febrile neutropenia, causing a larger proportion of infections in febrile neutropenic patients than any other single organism. For example, *P. aeruginosa* was responsible for 28% of documented infections in 499 febrile neutropenic patients in one study from the Indian subcontinent and for 31% of such infections in another. In a large study of infections in leukemia patients from Japan, *P. aeruginosa* was the most frequently documented cause of bacterial infection. In studies performed in North America, northern Europe, and Australia, the incidence of *P. aeruginosa* bacteremia in febrile neutropenia was quite variable. In a review of 97 reports published between 1987 and 1994, the incidence was reported to be 1–2.5% among febrile neutropenic patients given empirical therapy and 5–12% among patients with microbiologically documented infections. The most common clinical syndromes encountered were bacteremia, pneumonia, and soft tissue infections manifesting mainly as ecthyma gangrenosum.

TREATMENT

Infections in Febrile Neutropenic Patients

(Table 170-2) Compared with rates three decades ago, improved rates of response to antibiotic therapy have been reported in many studies. A study of 127 patients demonstrated a reduction in the mortality rate from 71 to 25% with the introduction of ceftazidime and imipenem. Because neutrophils—the normal host defenses against this organism—are absent in febrile neutropenic patients, maximal doses of antipseudomonal β-lactam antibiotics should be used for the management of *P. aeruginosa* bacteremia in this setting.

Infections in Patients with AIDS *P. aeruginosa* infections were well documented in patients with AIDS before the advent of antiretroviral therapy. Since the introduction of protease inhibitors, *P. aeruginosa* infections in patients with AIDS have been seen less frequently but still occur, particularly in the form of sinusitis. While this entity is now uncommon in developed nations, there are still large numbers of patients with untreated HIV infection or poorly controlled disease in developing nations who are likely to suffer from *P. aeruginosa* infections. The clinical presentation of *Pseudomonas* infection (especially pneumonia and bacteremia) in patients with AIDS is remarkable in that, although the illness may appear not to be severe, the infection may nonetheless be fatal. Patients with bacteremia may have only a low-grade fever and may present with ecthyma gangrenosum. Pneumonia, with or without bacteremia, is perhaps the most common type of *P. aeruginosa* infection. Patients with *P. aeruginosa* pneumonia exhibit the classic clinical signs and symptoms of pneumonia, such as fever, productive cough, and chest pain. The infection may be lobar or multilobar and shows no predisposition for any particular location. The most striking feature is the high frequency of cavitary disease.

TREATMENT

Infections in Patients with AIDS

Therapy for any of these conditions in AIDS patients is no different from that in other patients. However, relapse is the rule unless the patient's CD4+ T-cell count rises to >50/μL or suppressive antibiotic therapy is given. In attempts to achieve cures and prevent relapses, therapy tends to be more prolonged.

Gastrointestinal Infections A poorly understood syndrome caused by *P. aeruginosa* has been described in the Far East and has been called *Shanghai fever* and *Pseudomonas enterocolitis*. This syndrome occurs in young children; its occurrence in adults appears to be rare. Shanghai fever manifests as severe enteric disease, sepsis with invasive disease, and complications, whereas *Pseudomonas* enterocolitis is characterized by prolonged fever with bloody or mucoid diarrhea mimicking bacterial enterocolitis. The mortality rate ranges between 23 and 89%, with ecthyma gangrenosum occurring in >50% of cases. Early recognition and treatment have led to a reduction in the mortality rate. There is an above-average occurrence of the *exoU* gene among *Pseudomonas* isolates from patients with this syndrome.

Multidrug-Resistant Infections (Table 170-2) *P. aeruginosa* has a notorious propensity to develop antibiotic resistance. Over three decades, the impact of resistance was minimized by the rapid development of several potent antipseudomonal β-lactams and fluoroquinolones. However, rates of resistance to these agents that revolutionized the treatment of *P. aeruginosa* have risen to the point where some are almost unusable empirically because of the worldwide emergence of strains carrying determinants that mediate resistance. Extremely high rates of MDR strains have been reported from eastern and southern Europe, Latin America, India, and China, especially in ICUs. Physicians have had to resort to drugs such as colistin and polymyxin B, which were discarded decades ago. This surge in resistance is mediated by multiple mechanisms sometimes converging in individual strains. Chief among these are chromosomal or plasmid-borne penicillinases, extended-spectrum β-lactamases, cephalosporinases, and carbapenemases. Any of these may be combined with permeability mutations and efflux pump overexpression. The greatest nemesis in this regard is the worldwide presence of carbapenemases in *P. aeruginosa* leading to resistance to most β-lactams except some of the newest agents recently developed. These new agents are generally combinations of a cephalosporin or a carbapenem most often with a novel β-lactamase inhibitor. Several have been approved for clinical use, and all are active against

MDR *P. aeruginosa* to varying degrees. Currently approved agents include ceftolozane-tazobactam, ceftazidime-avibactam, meropenem-vaborbactam, and imipenem-relebactam. A novel cephalosporin, cefiderocol, which uses the iron uptake pathway of *P. aeruginosa*, also demonstrates activity against MDR strains. Since MDR and XDR *P. aeruginosa* are unpredictable in regard to the underlying mechanisms of resistance, laboratory testing is absolutely required before the use of any of these agents. Most academic institutions restrict the use of these agents as there are increasing reports of resistance even to these agents, as well the cost implications of misuse.

BURKHOLDERIA SPECIES

BURKHOLDERIA CEPACIA COMPLEX

The *B. cepacia* complex (BCC) gained notoriety as the cause of a rapidly fatal syndrome of respiratory distress and septicemia (the "cepacia syndrome") in CF patients. Of the more than 20 species of this complex, the three most frequently seen in CF patients are *B. cenocepacia*, *B. multivorans*, and *B. stabilis*. In addition to their occurrence in CF, members of this complex were not uncommonly encountered in ICU patients (previously designated *Pseudomonas cepacia*) and patients with chronic granulomatous disease, in whom they caused lung disease. BCC organisms are environmental organisms that inhabit moist environments and are found in the rhizosphere. They possess multiple virulence factors that may play roles in disease as well as colonizing factors that are capable of binding to lung mucus—an ability that may explain the predilection of *B. cepacia* for the lungs in CF. *B. cenocepacia* is motile, secretes elastase, and possesses components of an injectable toxin-secretion system like that of *P. aeruginosa*; its LPS is among the most potent of all LPSs in stimulating an inflammatory response in the lungs. Inflammation may be the major cause of the lung disease seen in the "cepacia" syndrome. Besides infecting the lungs in CF, the BCC organisms appear as airway colonizers during broad-spectrum antibiotic therapy and are causes of VAP, catheter-associated infections, and wound infections. *B. cenocepacia* has emerged as a barrier to lung transplantation in CF, with relatively high mortality rates after a year compared to infection with other members of this complex.

TREATMENT

B. cepacia Complex Infections

BCC organisms are intrinsically resistant to many antibiotics, rendering empiric treatment difficult. Therefore, treatment must be tailored according to sensitivities. Trimethoprim-sulfamethoxazole (TMP-SMX), meropenem, and minocycline are the most active agents in vitro and may be started as first-line agents (Table 170-2). However, recent reports indicate that there has been increasing resistance to these agents especially in CF patients. Some strains are susceptible to third-generation ureidopenicillins, advanced cephalosporins, and fluoroquinolones, and these agents may be used against isolates known to be susceptible. Newer antibiotics such as ceftolozane-tazobactam and ceftazidime-avibactam show good activity against MDR strains in vitro. However, there is very limited clinical experience with these agents.

BURKHOLDERIA PSEUDOMALLEI

B. pseudomallei is the causative agent of melioidosis, a disease of humans and animals that is geographically restricted to Southeast Asia and northern Australia, with occasional cases in countries such as India and China. This organism may be isolated from individuals returning directly from these endemic regions and from military personnel who have served in endemic regions. Symptoms of this illness may develop only at a later date because of the organism's ability to cause latent infections, which has been attributed to its ability to survive within cells. *B. pseudomallei* is found in soil and water. Humans and animals are infected by inoculation, inhalation, or ingestion; only rarely is the organism transmitted from person to person. Humans are not colonized without being infected. Among the pseudomonads, *B. pseudomallei* is perhaps the most virulent species. Host compromise is not an essential prerequisite for disease, although many patients have common underlying medical diseases (e.g., diabetes, renal failure, or alcohol abuse). *B. pseudomallei* is a facultative intracellular organism whose replication in PMNs and macrophages may be aided by the possession of a polysaccharide capsule. The organism also possesses elements of a type III secretion system that plays a role in its intracellular survival. During infection, there is a florid inflammatory response whose role in disease is unclear.

B. pseudomallei causes a wide spectrum of conditions, ranging from asymptomatic infection to abscesses, pneumonia, and disseminated disease. It is a significant cause of fatal community-acquired pneumonia and septicemia in endemic areas, with mortality rates as high as 44% reported in Thailand. Acute pulmonary infection is the most commonly diagnosed form of melioidosis. Pneumonia may be asymptomatic (with routine chest radiographs showing mainly upper-lobe infiltrates) or may present as severe necrotizing disease. *B. pseudomallei* also causes chronic pulmonary infections with systemic manifestations that mimic those of tuberculosis, including chronic cough, fever, hemoptysis, night sweats, and cavitary lung disease. Besides pneumonia, the other principal form of *B. pseudomallei* disease is skin ulceration with associated lymphangitis and regional lymphadenopathy. Spread from the lungs or skin, which is most often documented in debilitated individuals, gives rise to septicemic forms of melioidosis that carry a high mortality rate.

TREATMENT

B. pseudomallei Infections

B. pseudomallei is susceptible to advanced penicillins, cephalosporins, and carbapenems (Table 170-2). Treatment is divided into two stages: an intensive 2-week phase of therapy with ceftazidime or a carbapenem followed by at least 12 weeks of oral TMP-SMX to eradicate the organism and prevent relapse. Australian guidelines for treating this condition recommend longer periods of intensive therapy—4–8 weeks for severe infections, osteomyelitis, and CNS infections. The recognition of this bacterium as a potential agent of biologic warfare has stimulated interest in the development of a vaccine.

BURKHOLDERIA MALLEI

B. mallei causes the equine disease glanders in Africa, Asia, and South America. The organism was eradicated from Europe and North America decades ago. The last case seen in the United States occurred in 2001 in a laboratory worker; before that, *B. mallei* had last been seen in this country in 1949. In contrast to the other organisms discussed in this chapter, *B. mallei* is not an environmental organism and does not persist outside its equine hosts. Consequently, *B. mallei* infection is an occupational risk for handlers of horses, equine butchers, and veterinarians in areas of the world where it still exists. Diabetics are thought to be especially susceptible to infection by this organism. The polysaccharide capsule is a critical virulence determinant. The organism is transmitted from animals to humans by inoculation into the skin, where it causes local infection with nodules and lymphadenitis. Regional lymphadenopathy is common. Respiratory secretions from infected horses are extremely infectious. Inhalation results in clinical signs of typical pneumonia but may also cause an acute febrile illness with ulceration of the trachea. The organism may disseminate from the skin or lungs to cause septicemia with signs of sepsis. The septicemic form is frequently associated with shock and a high mortality rate. The infection may also enter a chronic phase and present as disseminated abscesses. *B. mallei* infection may present as early as 1–2 days after inhalation or (in cutaneous disease) may not become evident for months.

TREATMENT

B. mallei Infections

The antibiotic susceptibility pattern of *B. mallei* is similar to that of *B. pseudomallei*; in addition, the organism is susceptible to the macrolides azithromycin and clarithromycin. *B. mallei* infection should be treated with the same drugs and for the same duration as melioidosis.

STENOTROPHOMONAS MALTOPHILIA

S. maltophilia is the only potential human pathogen among a genus of ubiquitous organisms found in the rhizosphere (i.e., the soil that surrounds the roots of plants). The organism is an opportunist that is acquired from the environment but is even more limited than *P. aeruginosa* in its ability to colonize patients or cause infections. Immunocompromise is not sufficient to permit these events; rather, major perturbations of the human flora are usually necessary for the establishment of *S. maltophilia*. Accordingly, most cases of human infection occur in the setting of very broad-spectrum antibiotic therapy with agents such as advanced cephalosporins and carbapenems, which eradicate the normal flora and other pathogens. The remarkable ability of *S. maltophilia* to resist virtually all classes of antibiotics is attributable to the possession of antibiotic efflux pumps and of two β-lactamases (L1 and L2) that mediate β-lactam resistance, including that to carbapenems. It is fortunate that the virulence of *S. maltophilia* appears to be limited. Although a serine protease is present in some strains, virulence is probably a result of the host's inflammatory response to components of the organism such as LPS and flagellin. *S. maltophilia* is most commonly found in the respiratory tract of ventilated patients, where the distinction between its roles as a colonizer and as a pathogen is often difficult to make. However, *S. maltophilia* does cause pneumonia and bacteremia in such patients, and these infections have led to septic shock. A severe form of pneumonia with significant mortality has been reported in neutropenic patients. Also common is central venous line–associated infection (with or without bacteremia), which has been reported most often in patients with cancer. *S. maltophilia* is a rare cause of ecthyma gangrenosum in neutropenic patients. It has been isolated from ~5% of CF patients but is not believed to be a significant pathogen in this setting. There are also less common sites of infection such as the endocardium, meninges, urinary tract, and skin; however, the sites of major concern remain the blood and lungs.

TREATMENT

S. maltophilia Infections

The intrinsic resistance of *S. maltophilia* to most antibiotics renders infection difficult to treat. The antibiotics to which it is most often (although not uniformly) susceptible are TMP-SMX, ticarcillin-clavulanate, levofloxacin, minocycline, cefiderocol and tigecycline (Table 170-2). Ceftazidime is no longer recommended for treatment. A few retrospective studies show utility of minocycline, but relevant clinical outcomes have not been thoroughly evaluated. Consequently, combination therapy with TMP-SMX plus ticarcillin-clavulanate, levofloxacin, or high-dose minocycline is recommended for initial therapy pending susceptibility testing. Catheters must be removed early in the treatment of bacteremia. The treatment of proven, high-inoculum infections due to *S. maltophilia* is associated with frequent development of resistance during therapy. The newest β-lactam/β-lactamase inhibitor combinations show mixed results against this organism probably because of the presence of two different β-lactamases, a penicillinase and a cephalosporinase, compounded by the presence of efflux pumps. Resistance to one or the other of the newer agents has already been noted, including to cefiderocol. The most controversial aspect of therapy of *S. maltophilia* is whether pulmonary cultures in a nonneutropenic patient with lung consolidation require treatment or represent colonization. Many clinicians opt not to treat such patients given the fact that the development of resistance is quite common, compromising a serious subsequent infection such as a bacteremia.

FURTHER READING

Bauer KA et al: Extended-infusion cefepime reduces mortality in patients with *Pseudomonas aeruginosa* infections. Antimicrob Agents Chemother 57:2907, 2013.

Bowers DR et al: Outcomes of appropriate empiric combination versus monotherapy for *Pseudomonas aeruginosa* bacteremia. Antimicrob Agents Chemother 157:1270, 2013.

Brooke JS: *Stenotrophomonas maltophilia*: An emerging global opportunistic pathogen. Clin Microbiol Rev 25:2, 2012.

Cattaneo C et al: *P. aeruginosa* bloodstream infections among hematological patients: An old or new question? Ann Hematol 91:1299, 2012.

Fabre V et al: Antibiotic therapy for *Pseudomonas aeruginosa* bloodstream infections: How long is long enough? Clin Infect Dis 69:2011, 2019.

Gupte A et al: High pyocyanin production and non-motility of *Pseudomonas aeruginosa* isolates are correlated with septic shock or death in bacteremic patients. PLoS One 16:e0253259, 2021.

Horcajada JP et al: Epidemiology and treatment of multidrug-resistant and extensively drug-resistant *Pseudomonas aeruginosa* infections. Clin Microbiol Rev 32:e00031, 2019.

Kalil AC et al: Executive summary: Management of adults with hospital-acquired and ventilator-associated pneumonia: 2016 clinical practice guidelines by the Infectious Diseases Society of America and the American Thoracic Society. Clin Infect Dis 63:575, 2016.

Peña C et al: Influence of virulence genotype and resistance profile in the mortality of *Pseudomonas aeruginosa* bloodstream infections. Clin Infect Dis 60:539, 2015.

Wunderlink RG et al: Cefiderocol versus high-dose, extended-infusion meropenem for the treatment of Gram-negative nosocomial pneumonia (APEKS-NP): A randomized, double-blind, phase 3, non-inferiority trial. Lancet Infect Dis 21:213, 2021.

171

Salmonellosis

David A. Pegues, Samuel I. Miller

Bacteria of the genus *Salmonella* are highly adapted for growth in both humans and animals and cause a wide spectrum of disease. The growth of serotypes *Salmonella* Typhi and *Salmonella* Paratyphi is restricted to human hosts, in whom these organisms cause enteric (typhoid) fever. The remaining serotypes (nontyphoidal *Salmonella*, or NTS) can colonize the gastrointestinal tracts of a broad range of animals, including mammals, reptiles, birds, and insects. More than 200 serotypes of *Salmonella* are pathogenic to humans, in whom they often cause gastroenteritis and can be associated with localized infections and/or bacteremia.

ETIOLOGY

This large genus of gram-negative bacilli within the family Enterobacteriaceae consists of two species: *Salmonella enterica*, which contains six subspecies, and *Salmonella bongori*. *S. enterica* subspecies I includes almost all the serotypes pathogenic for humans. Members of the seven *Salmonella* subspecies are classified into >2600 serotypes (serovars); for simplicity, *Salmonella* serotypes (most of which are named for the city where they were identified) are often used as the species designation. For example, the full taxonomic designation *S. enterica* subspecies *enterica* serotype Typhimurium can be shortened to *Salmonella* serotype Typhimurium or simply S. Typhimurium. Serotyping is based on antigenically diverse surface structures: the somatic O antigen (lipopolysaccharide cell-wall components), the surface Vi antigen (restricted to S. Typhi and S. Paratyphi C), and the flagellar H antigen.

Salmonellae are gram-negative, non-spore-forming, facultatively anaerobic bacilli that measure 2–3 µm by 0.4–0.6 µm. The initial identification of salmonellae in the clinical microbiology laboratory is based on growth characteristics. Salmonellae, like other Enterobacteriaceae, produce acid on glucose fermentation, reduce nitrates, and do not produce cytochrome oxidase. In addition, all salmonellae except *Salmonella* Gallinarum-Pullorum are motile by means of peritrichous flagella, and all but *S.* Typhi produce gas (H_2S) on sugar fermentation. Notably, only 1% of clinical isolates ferment lactose; a high level of suspicion must be maintained to detect these rare clinical lactose-fermenting isolates.

Although serotyping of all surface antigens can be used for formal identification, most laboratories perform a few simple agglutination reactions that define specific O-antigen serogroups, designated A, B, C_1, C_2, D, and E. Strains in these six serogroups cause ~99% of *Salmonella* infections in humans and other warm-blooded animals. Whole-genome sequencing is used in epidemiologic investigations to identify the source of foodborne outbreaks and to explore the international transmission of multidrug-resistant *Salmonella* strains.

■ PATHOGENESIS

All *Salmonella* infections begin with ingestion of organisms, most commonly in contaminated food or water. The infectious dose ranges from 200 colony-forming units (CFU) to 10^6 CFU, and the ingested dose is an important determinant of incubation period and disease severity. Conditions that decrease either stomach acidity (an age of <1 year, acid suppression therapy, or achlorhydric disease) or intestinal integrity (inflammatory bowel disease, cytotoxic chemotherapy, prior gastrointestinal surgery, or alteration of the intestinal microbiome by antibiotic administration) increase susceptibility to *Salmonella* infection.

Once *S.* Typhi and *S.* Paratyphi reach the small intestine, they penetrate the mucus layer of the gut and traverse the intestinal layer through phagocytic microfold (M) cells that reside within Peyer's patches. Salmonellae can trigger the formation of membrane ruffles in normally nonphagocytic epithelial cells. These ruffles reach out and enclose adherent bacteria within large vesicles by *bacterium-mediated endocytosis*. This process is dependent on the direct delivery of *Salmonella* proteins into the cytoplasm of epithelial cells by the specialized bacterial type III secretion system. These bacterial proteins mediate alterations in the actin cytoskeleton that are required for *Salmonella* uptake.

After crossing the epithelial layer of the small intestine, *S.* Typhi and *S.* Paratyphi, which cause enteric (typhoid) fever, are phagocytosed by macrophages. These salmonellae survive the antimicrobial environment of the macrophage by sensing environmental signals that trigger alterations in regulatory systems of the phagocytosed bacteria. For example, PhoP/PhoQ (the best-characterized regulatory system) triggers the alteration of the outer membrane by increasing the synthesis and transport of different outer-membrane proteins, lipopolysaccharides, and glycerophospholipids, so that the altered bacterial surface can resist microbicidal activities and potentially alter host cell signaling. In addition, salmonellae encode a second type III secretion system that directly delivers bacterial proteins across the phagosome membrane into the macrophage cytoplasm. This secretion system functions to remodel the *Salmonella*-containing vacuole, promoting bacterial survival and replication.

Once phagocytosed, typhoidal salmonellae disseminate throughout the body in macrophages via the lymphatics and colonize reticuloendothelial tissues (liver, spleen, lymph nodes, and bone marrow). Patients have relatively few or no signs and symptoms during this initial incubation stage. Signs and symptoms, including fever and abdominal pain, probably result from secretion of cytokines by macrophages and epithelial cells in response to bacterial products that are recognized by innate immune receptors when a critical number of organisms have replicated. Over time, the development of hepatosplenomegaly is likely to be related to the recruitment of mononuclear cells and the development of a specific acquired cell-mediated immune response to *S.* Typhi colonization. The recruitment of additional mononuclear cells and lymphocytes to Peyer's patches during the several weeks after initial colonization/infection can result in marked enlargement and necrosis of the Peyer's patches, which may be mediated by bacterial products that promote cell death as well as the inflammatory response. In the case of *S.* Typhi and *S.* Paratyphi A, many strains produce a toxin when localized within host cells. This toxin is then transported extracellularly and probably contributes to systemic symptoms as well as the unusual neuropsychiatric states that can be seen in severe typhoidal illness.

In contrast to enteric fever, which is characterized by an infiltration of mononuclear cells into the small-bowel mucosa, NTS gastroenteritis is characterized by massive polymorphonuclear leukocyte infiltration into both the large- and small-bowel mucosa. This response appears to depend on the induction of interleukin 8, a strong neutrophil chemotactic factor, which is secreted by intestinal cells because of nontyphoidal *Salmonella* colonization and translocation of bacterial proteins and LPS into host cell cytoplasm with subsequent activation of inflammasomes. The degranulation and release of toxic substances by neutrophils may result in damage to the intestinal mucosa, causing the inflammatory diarrhea observed with nontyphoidal gastroenteritis. An additional important factor in the persistence of NTS in the intestinal tract and the organism's capacity to compete with endogenous flora is the ability to utilize the sulfur-containing compound tetrathionate for metabolism in a microaerophilic environment. In the presence of intestinal inflammation, tetrathionate is generated from thiosulfate produced by epithelial cells through inflammatory cell production of reactive oxygen species. In contrast to nontyphoidal Salmonellae, typhoidal Salmonellae do not effectively colonize the intestinal tract but have evolved as systemic pathogens, largely through gene loss and perhaps by the loss of the ability to utilize butyrate within the human intestine.

ENTERIC (TYPHOID) FEVER

Enteric (typhoid) fever is a systemic disease characterized by fever and abdominal pain and caused by dissemination of *S.* Typhi or *S.* Paratyphi. The disease was initially called *typhoid fever* because of its clinical similarity to typhus. In the early 1800s, typhoid fever was clearly defined pathologically as a unique illness based on its association with enlarged Peyer's patches and mesenteric lymph nodes. In 1869, given the anatomic site of infection, the term *enteric fever* was proposed as an alternative designation to distinguish typhoid fever from typhus. However, to this day, the two designations are used interchangeably.

■ EPIDEMIOLOGY

In contrast to other *Salmonella* serotypes, the etiologic agents of enteric fever—*S.* Typhi and *S.* Paratyphi serotypes A, B, and C—have no known hosts other than humans. Most commonly, food-borne or waterborne transmission results from fecal contamination by ill or asymptomatic chronic carriers. Sexual transmission between male partners has been described. Health care workers occasionally acquire enteric fever after exposure to infected patients or during processing of clinical specimens and cultures.

With improvements in food handling and water/sewage treatment, enteric fever has become rare in developed nations. An estimated 9.2–21 million cases of typhoid fever, 5 million cases of paratyphoid fever, and 110,000–280,000 deaths occur each year. The highest estimated annual incidence rates of typhoid fever are in the Indian subcontinent, including Pakistan, Bangladesh, Nepal, India, and Eastern Mediterranean and African regions, and exceed 1000 cases per 100,000 children in some urban areas **(Fig. 171-1)**. A high incidence of enteric fever correlates with mixing of drinking water with human sewage. In endemic regions, enteric fever is more common in poor neighborhoods in large cities than rural areas and among young children and adolescents than among other age groups. Risk factors include fecally contaminated drinking water or ice, flooding, food and drinks purchased from street vendors, raw fruits and vegetables grown in fields fertilized with sewage, ill household contacts, lack of hand washing and toilet access, and evidence of prior *Helicobacter pylori* infection (an association probably related to chronically reduced gastric acidity).

Multidrug-resistant (MDR) strains of *S.* Typhi emerged in the 1980s in China and Southeast Asia and have since disseminated widely. These strains contain plasmids encoding resistance to chloramphenicol,

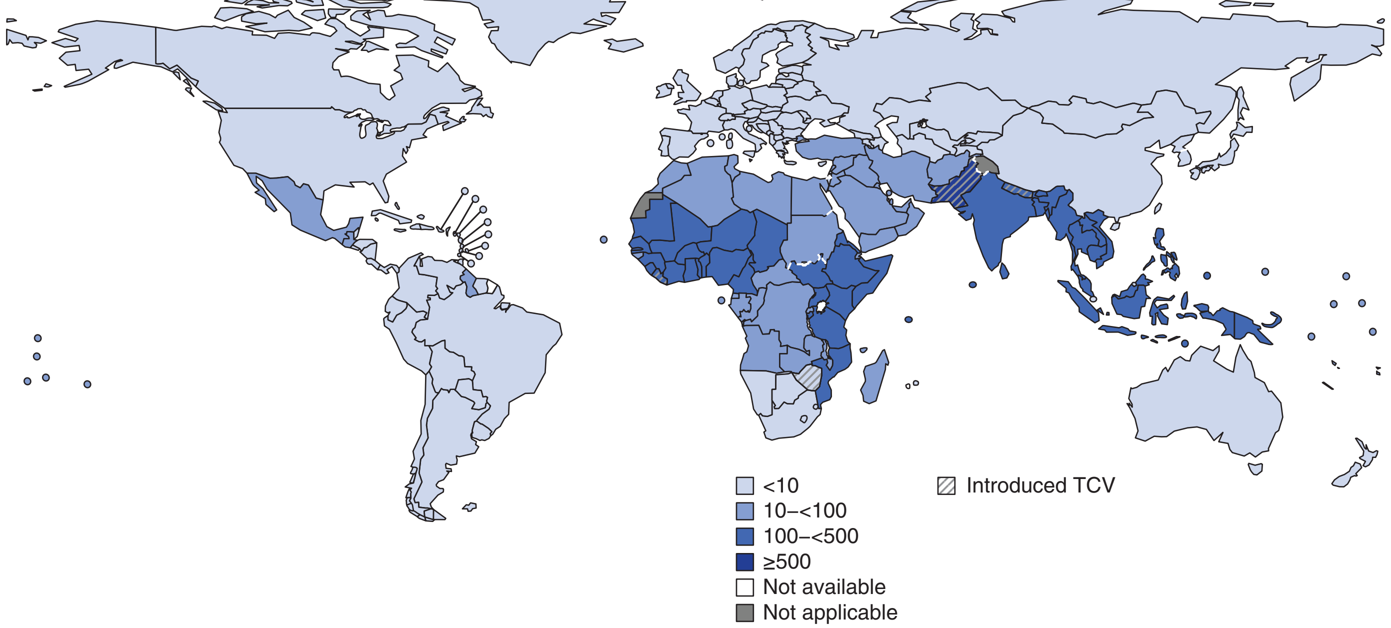

FIGURE 171-1 Estimated national typhoid fever incidence and worldwide typhoid conjugate vaccine introduction, 2019–2022. *(Reproduced from M Hancuh et al: Typhoid fever surveillance incidence estimates, and progress toward typhoid conjugate vaccine introduction – Worldwide, 2018–2022.* MMWR Morb Mortal Wkly Rep 72:171, 2023.)

ampicillin, and trimethoprim—antibiotics long used to treat enteric fever. With the increased use of fluoroquinolones to treat MDR enteric fever in the 1990s, MDR strains of *S.* Typhi and *S.* Paratyphi with decreased susceptibility to ciprofloxacin (DSC; minimal inhibitory concentration [MIC], ≥0.125 μg/mL) or ciprofloxacin resistance (MIC, ≥1 μg/mL) emerged on the Indian subcontinent and have spread with human migration first to Southern Asia and more recently to Eastern and Southern Africa. These strains represent clone H58, which is increasingly associated with clinical treatment failure of fluoroquinolones. Since emerging in 2016 in urban slums of Sindh province in Southeastern Pakistan, an extensively drug-resistant (XDR) *S.* Typhi H58 clone with plasmid-mediated extended-spectrum beta-lactamase (ESBL) resistance has now become the dominant cause of typhoid fever in Pakistan. Air travel from Pakistan has facilitated the international spread of this XDR strain. Azithromycin resistance has emerged in multiple countries where azithromycin is used for first-line treatment of enteric fever and for mass administration for trachoma.

In the United States, the Centers for Disease Control and Prevention (CDC) estimates that typhoid fever affects 5700 persons each year, a number far in excess of the ~350 cases of typhoid fever and ~90 cases of paratyphoid fever reported annually. In 2015, the median age of patients with typhoid fever was 23 years, and it was 29 years for paratyphoid fever. Most cases of enteric fever were associated with international travel (78%), predominantly to Indian, Pakistan, and Bangladesh, and visiting friends and family. Only 3% of travelers diagnosed with typhoid fever had received *S.* Typhi vaccine within the previous 5 years. In 2015, 66% of *S.* Typhi in the United States were DSC, and ~10% were resistant to ampicillin, chloramphenicol, and trimethoprim-sulfamethoxazole (TMP-SMX). Infection with DSC *S.* Typhi was associated with travel to the Indian subcontinent. In the United States, domestically acquired cases of enteric fever are less often DSC or MDR compared with travel-associated cases and are most often sporadic, although outbreaks linked to contaminated food products and previously unrecognized chronic carriers continue to occur.

CLINICAL COURSE

Enteric fever is a misnomer, in that the hallmark features of this disease—fever and abdominal pain—are variable. While fever is documented at presentation in >75% of cases, abdominal pain is reported in only 30–40%. Thus, a high index of suspicion for this potentially fatal systemic illness is necessary when a person presents with fever and a history of recent travel to a developing country.

The mean incubation period for *S.* Typhi is 10–14 days but ranges from 5 to 21 days, depending on the inoculum size and the host's health and vaccination status. The most prominent symptom is prolonged fever (38.8°–40.5°C [101.8°–104.9°F]), which can continue for up to 4 weeks if untreated. *S.* Paratyphi A is thought to cause milder disease than *S.* Typhi, with predominantly gastrointestinal symptoms. However, a prospective study of 669 consecutive cases of enteric fever in Kathmandu, Nepal, found that the infections caused by these organisms were clinically indistinguishable. In this series, symptoms reported on initial medical evaluation included headache (80%), chills (35–45%), cough (30%), sweating (20–25%), myalgias (20%), malaise (10%), and arthralgia (2–4%). Gastrointestinal manifestations included anorexia (55%), abdominal pain (30–40%), nausea (18–24%), vomiting (18%), and diarrhea (22–28%) more commonly than constipation (13–16%). Physical findings included coated tongue (51–56%), splenomegaly (5–6%), and abdominal tenderness (4–5%).

Early physical findings of enteric fever include rash ("rose spots"; 30%), hepatosplenomegaly (3–6%), epistaxis, and relative bradycardia at the peak of high fever (<50%). Rose spots **(Fig. 171-2; see also Fig. A1-9)** make up a faint, salmon-colored, blanching, maculopapular rash located

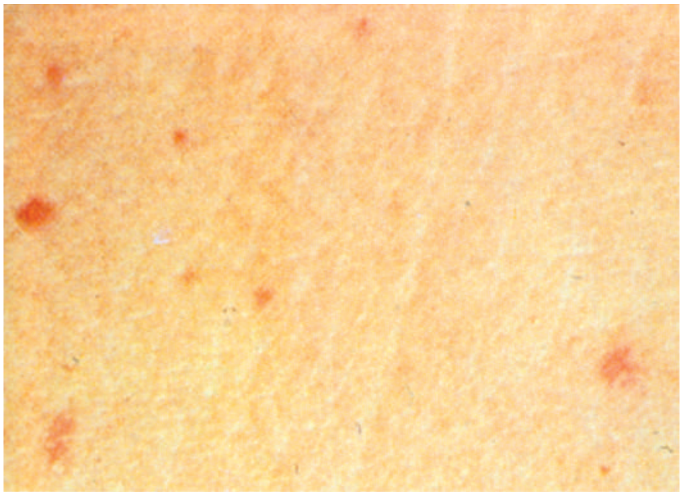

FIGURE 171-2 "Rose spots," the rash of enteric fever due to *Salmonella* Typhi or *Salmonella* Paratyphi.

primarily on the trunk and chest. The rash is evident in ~30% of patients at the end of the first week and resolves without a trace after 2–5 days. Patients can have two or three crops of lesions, and *Salmonella* can be cultured from punch biopsies of these lesions. The faintness of the rash makes it difficult to detect in highly pigmented patients.

Complications of typhoid fever are estimated to occur in ~27% of hospitalized patients and correlate with a longer duration of symptoms before hospitalization, host factors (host genetics, immunosuppression, acid suppression therapy, previous exposure, and vaccination status), strain virulence and inoculum, and choice of antibiotic therapy. Gastrointestinal bleeding (6%) and intestinal perforation (1%) most commonly occur in the third and fourth weeks of illness and result from hyperplasia, ulceration, and necrosis of the ileocecal Peyer's patches at the initial site of *Salmonella* infiltration **(Fig. 171-3)**. Both complications are life-threatening and require immediate fluid resuscitation and surgical intervention, with broadened antibiotic coverage for polymicrobial peritonitis **(Chap. 137)** and treatment of gastrointestinal hemorrhages, including bowel resection. Neurologic manifestations occur in 2–40% of patients and include meningitis, Guillain-Barré syndrome, neuritis, and neuropsychiatric symptoms (described as "muttering delirium" or "coma vigil"), with picking at bedclothes or imaginary objects.

Uncommon complications whose incidences are reduced by prompt antibiotic treatment include disseminated intravascular coagulation, hemophagocytic syndrome, pancreatitis, hepatitis, hepatic and splenic abscesses and granulomas, endocarditis, pericarditis, myocarditis, orchitis, glomerulonephritis, pyelonephritis and hemolytic-uremic syndrome, severe pneumonia, arthritis, osteomyelitis, endophthalmitis, and parotitis. Up to 10% of patients develop mild relapse, usually within 2–3 weeks of fever resolution and associated with the same strain and susceptibility profile.

Up to 10% of untreated patients with typhoid fever excrete *S.* Typhi in the feces for up to 3 months, and 2–5% develop chronic asymptomatic carriage, shedding *S.* Typhi in either urine or stool for >1 year. Chronic carriage is more common among women, infants, and persons who have biliary abnormalities or concurrent bladder infection with *Schistosoma haematobium*. *S.* Typhi and other salmonellae are adapted to survive in the gallbladder environment by forming biofilms on gallstones and invading gallbladder epithelial cells. Chronic carriage is associated with an increased risk of gallbladder cancer, which is much more common in locales where *S.* Typhi is common, such as the Indian subcontinent.

DIAGNOSIS

Because the clinical presentation of enteric fever is relatively nonspecific, the diagnosis needs to be considered in any febrile traveler returning from a developing region, especially the Indian subcontinent, and the Southeast Asian or African region. Other diagnoses that should be considered in these travelers include malaria, viral hepatitis, bacterial enteritis, dengue fever, rickettsial infections, leptospirosis, amebic liver abscesses, and acute HIV infection **(Chap. 130)**. Other than a positive culture, no specific laboratory test is diagnostic for enteric fever. In 15–25% of cases, leukopenia and neutropenia are detectable. Leukocytosis is more common among children, during the first 10 days of illness, and in cases complicated by intestinal perforation or secondary infection. Other nonspecific laboratory findings include moderately elevated values in liver function tests and muscle enzyme levels.

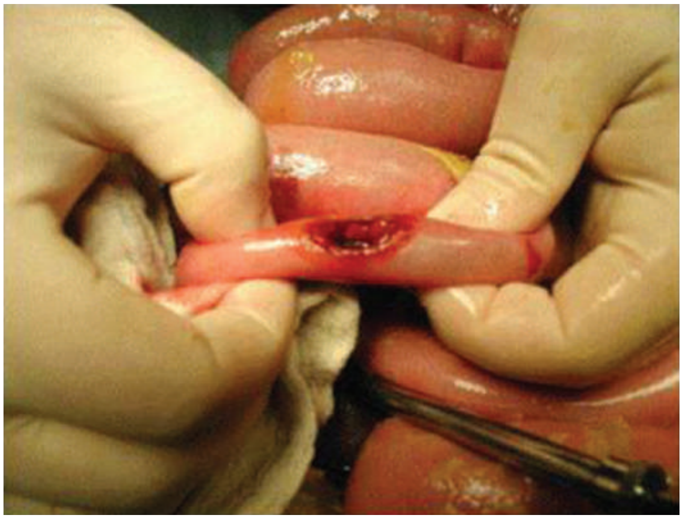

FIGURE 171-3 Typical ileal perforation associated with *Salmonella* Typhi infection. *(From JM Saxe, R Cropsey: Is operative management effective in treatment of perforated typhoid? Am J Surg 189:342, 2005.)*

The definitive diagnosis of enteric fever requires the isolation of *S.* Typhi or *S.* Paratyphi from blood, bone marrow, other sterile sites, rose spots, stool, or intestinal secretions. The diagnostic sensitivity of blood culture is only ~40–60% and is lower with low blood sample volume and among patients with prior antimicrobial use or in the first week of illness, reflecting the small number of *S.* Typhi organisms (i.e., <15/mL) typically present in the blood. Because almost all *S.* Typhi organisms in blood are associated with the mononuclear cell/platelet fraction, centrifugation of blood and culture of the buffy coat can substantially reduce the time to isolation of the organism but do not increase sensitivity.

Bone marrow culture is ~80% sensitive for detection of *S.* Typhi or *S.* Paratyphi, and, unlike that of blood culture, its yield is not reduced by up to 5 days of prior antibiotic therapy. Culture of intestinal secretions (best obtained by a noninvasive duodenal string test) can be positive despite a negative bone marrow culture. If blood, bone marrow, and intestinal secretions are all cultured, the yield is >90%. Stool cultures, although negative in 60–70% of cases during the first week, can become positive during the third week of infection in untreated patients.

Rapid immunodiagnostic commercial tests, including Tubex and Typhidot, mainly focus on detection of IgM and IgG antibodies to O and H antigens and are widely used at point of care to diagnose typhoid and paratyphoid fever because they are simple and low cost. In a 2017 systematic review, these rapid diagnostic tests had sensitivities ranging from ~70% to 80% and specificities ranging from ~80% to 90% and thus are not sufficiently accurate to replace blood cultures as the main approach to diagnose enteric fever. PCR detection of *S.* Typhi and *S.* Paratyphi in the blood have sensitivities of ~40–100%, depending on the gene targets. Molecular-based test platforms were scarce in resource-limited settings, but advancements and investments in molecular diagnostics arising from the SARS-CoV-2 pandemic have made feasible the widespread use of molecular tests for diagnosis of enteric fever.

TREATMENT

Enteric (Typhoid) Fever

Enteric fever is associated with an overall case-fatality rate of 2.5%, but mortality rates rise to 4.5% among hospitalized patients and to 10–30% if untreated. Prompt administration of appropriate antibiotic therapy prevents severe complications of enteric fever and reduces mortality to <1%. The initial choice of antibiotics depends on the susceptibility of the *S.* Typhi and *S.* Paratyphi strains in the area of residence or travel **(Table 171-1)**. A 2022 systematic review of 27 randomized clinical trials for the treatment of enteric fever found no difference between ceftriaxone, fluoroquinolone, or azithromycin in comparative risk of treatment failure, microbiologic failure, relapse, convalescent carriage, or adverse events. Oral cefixime also can be used to treat enteric fever but may increase the risk of clinical failure and time to defervescence compared with fluoroquinolones. Of note, most of the trials included in this review were small and conducted >20 years ago, and data on antimicrobial resistance could not be analyzed. Because of the high prevalence of strains of *S.* Typhi and *S.* Paratyphi with decreased susceptibility to ciprofloxacin (MIC >0.125 μg/mL) on the Indian subcontinent and in some locales in Africa, fluoroquinolones should no longer

TABLE 171-1 Antibiotic Therapy for Enteric Fever in Adults

INDICATION	AGENT	DOSAGE (ROUTE)	DURATION, DAYS
Empirical Treatment			
	Ceftriaxone[a]	2 g/d (IV)	10–14
	Ciprofloxacin[b]	500 mg bid (PO) or 400 mg q12h (IV)	5–7
	Azithromycin[c]	1 g/d (PO)	5
Fully Susceptible			
Optimal treatment	Ceftriaxone	2 g/d (IV)	10–14
	Ciprofloxacin	500 mg bid (PO) or 400 mg q12h (IV)	5–7
	Azithromycin	1 g/d (PO)	5
Alternative treatment	Amoxicillin	1 g tid (PO) or 2 g q6h (IV)	14
	Chloramphenicol	25 mg/kg tid (PO or IV)	14–21
	Trimethoprim-sulfamethoxazole	160/800 mg bid (PO)	7–14
Multidrug-Resistant, Depending on the Susceptibility Pattern			
Optimal treatment	Ceftriaxone	2 g/d (IV)	10–14
	Ciprofloxacin	500 mg bid (PO) or 400 mg q12h (IV)	5–7
	Azithromycin	1 g/d (PO)	5
Ceftriaxone-Resistant			
Optimal treatment	Meropenem[d]	1 g q8h (IV)	10–14
	Azithromycin	1 g/d (PO)	5
Eradication of Carriage			
Optimal treatment	Ciprofloxacin	500–750 mg bid (PO)	28
Alternative treatment	Azithromycin	500 mg (PO)	28

[a] Or another third-generation cephalosporin (e.g., cefotaxime, 2 g q8h IV; or cefixime, 400 mg bid PO). [b] Or 1 g on day 1 followed by 500 mg/d PO for 6 days. [c] Or ofloxacin, 400 mg bid PO for 2–5 days. [d] Or imipenem 500 mg q6h IV.

be used for empirical treatment of enteric fever in these regions. Patients with concern for ceftriaxone-resistant *S.* Typhi infection based on a history of travel to Pakistan should be treated empirically with a carbapenem or azithromycin.

Ceftriaxone, cefotaxime, and (oral) cefixime are effective for treatment of MDR enteric fever in adults and children, including that caused by DSC and fluoroquinolone-resistant strains. These agents clear fever in ~1 week, with failure rates of ~5–10%, fecal carriage rates of <3%, and relapse rates of 3–6%. Fluoroquinolones are effective against susceptible strains, with cure rates of ~98% and relapse and fecal carriage rates of <2%. Oral azithromycin is recommended for the treatment of uncomplicated enteric fever, including ESBL, DSC, and fluoroquinolone-resistant strains, and results in defervescence in 4–6 days, with rates of relapse and convalescent stool carriage of <3%. Against DSC strains, azithromycin is associated with lower rates of treatment failure and shorter durations of hospitalization than are fluoroquinolones. Carbapenems are increasingly being used to treat complicated XDR *S.* Typhi infections, but cost and IV route of administration are significant barriers.

Most patients with uncomplicated enteric fever can be managed at home with oral antibiotics and antipyretics. Patients with persistent vomiting, diarrhea, and/or abdominal distension should be hospitalized and given supportive therapy as well as a parenteral third-generation cephalosporin, a fluoroquinolone, or carbapenem, depending on the susceptibility profile. Therapy should be administered for at least 10 days or for 5 days after fever resolution.

In a randomized, prospective, double-blind study of critically ill patients with enteric fever (i.e., those with shock and obtundation) in Indonesia in the early 1980s, the administration of dexamethasone (an initial dose of 3 mg/kg followed by eight doses of 1 mg/kg every 6 h) with chloramphenicol was associated with a substantially lower mortality rate than was treatment with chloramphenicol alone (10% vs 55%). Although this study has not been repeated in the "post-chloramphenicol era," severe enteric fever remains one of the few indications for glucocorticoid treatment of an acute bacterial infection. Steroid treatment beyond 48 hours may increase the relapse rate.

The 2–5% of patients who develop chronic carriage of fluoroquinolone-susceptible *S.* Typhi can be treated for 4 weeks with oral ciprofloxacin or other fluoroquinolones, with an eradication rate of ~80%. A 4-week course of oral azithromycin can potentially be used to treat carriers with fluoroquinolone-resistant strains, but clinical experience is limited. Oral amoxicillin is associated with lower eradication rates than fluoroquinolones but can be considered in persons with fluoroquinolone-resistant strains that are susceptible to ampicillin. In cases of anatomic abnormality (e.g., biliary, kidney, or bladder stones), eradication often requires both antibiotic therapy and surgical correction.

PREVENTION AND CONTROL

Theoretically, it is possible to eliminate the salmonellae that cause enteric fever because they survive only in human hosts and are spread by contaminated food and water. However, given the high prevalence of the disease in developing countries that lack adequate sewage disposal and water treatment, this goal is currently unrealistic. Thus, travelers to developing countries should be advised to monitor their food and water intake carefully and to strongly consider immunization against *S.* Typhi.

Two unconjugated typhoid vaccines are commercially available in the United States: (1) Ty21a, an oral live attenuated *S.* Typhi vaccine (given on days 1, 3, 5, and 7, with revaccination with a full four-dose series every 5 years); and (2) Vi CPS, a parenteral vaccine consisting of purified Vi polysaccharide from the bacterial capsule (given in a single dose, with a booster every 2 years). The minimal age for vaccination is 6 years for Ty21a and 2 years for Vi CPS. In a 2018 meta-analysis of 18 randomized clinical trials of vaccines for preventing typhoid fever in populations in endemic areas, the cumulative efficacy was 50% for Ty21a at 2.5 to 3 years and 55% for Vi CPS at 3 years. Although data on typhoid vaccines in travelers are limited, recent evidence suggests that typhoid vaccines are moderately effective (80%) in U.S. travelers. Currently, there is no licensed vaccine for paratyphoid fever.

Unconjugated typhoid vaccines are poorly immunogenic in children <5 years of age because of limited ability to elicit T cell–dependent immune responses and immunologic memory. Compared with unconjugated vaccines, typhoid Vi polysaccharide conjugated vaccines are effective in children <2 years of age and elicit substantially longer duration of protection. The World Health Organization has recommended two typhoid conjugate vaccines (TCV) prioritized to prevent typhoid fever in countries with high incidence rates—Typbar TCV (manufactured by Bharat Biotech) in 2018 and TYPHIBEV (manufactured by Biological E) in 2020. A single intramuscular 0.5-mL dose of either TCV is safe and 79–95% effective, with antibody response persisting up to 7 years. As of 2023, TCV has been routinely introduced in national immunization programs in Pakistan, Nepal, Liberia, Zimbabwe, Malawi, and Samoa (Fig. 171-1).

Typhoid vaccine is not required for international travel, but it is recommended for travelers to areas where there is a moderate to high risk of exposure to *S.* Typhi, especially those who are traveling to southern Asia and other developing regions of Asia, Africa, the Caribbean, and Central and South America and who will be exposed to potentially contaminated food and drink. Typhoid vaccine should be considered even for persons planning <2 weeks of travel to high-risk areas. In addition, clinical microbiology or research laboratory staff at risk of occupational exposure to *S.* Typhi and household contacts of known *S.* Typhi carriers should be vaccinated. Because the protective efficacy of vaccine can be overcome by the high inocula that are commonly encountered in food-borne exposures, immunization is an adjunct and not a substitute for the avoidance of high-risk foods and beverages. Immunization is not recommended for the management of persons who may have been exposed in a common-source outbreak.

Enteric fever is a notifiable disease in the United States. Individual health departments have their own guidelines for allowing ill or colonized food handlers or health care workers to return to their jobs. The reporting system enables public health departments to identify potential source patients and to treat chronic carriers in order to prevent further outbreaks. In addition, because 1–4% of patients with *S.* Typhi infection become chronic carriers, it is important to monitor patients (especially child-care providers and food handlers) for chronic carriage and to treat this condition if indicated.

NONTYPHOIDAL SALMONELLOSIS

■ EPIDEMIOLOGY

Worldwide, NTS causes ~93–150 million enteric infections and ~60,000–155,000 deaths annually. In the United States, the CDC estimates that NTS causes ~1.35 million illnesses, 26,500 hospitalizations, and 420 deaths each year. In 2022, the incidence of NTS infection in the United States was 14.5 cases per 100,000 persons—the second highest rate after *Campylobacter* (17.4 cases per 100,000 persons) among the 8 food-borne enteric pathogens under active surveillance. Although declining modestly since 2017, the incidence rate remains above the U.S. Healthy People 2030 goal of 11.5 cases per 100,000 persons. Globally, *S.* Typhimurium and *S.* Enteritidis are the most common serotypes causing human disease, with large differences in serotype distributions between regions but lesser differences between countries within the same region.

The incidence of NTS infection is highest during the rainy season in tropical climates and during the warmer months in temperate climates—a pattern coinciding with the peak in food-borne outbreaks. Rates of morbidity and mortality associated with NTS are highest among the elderly, infants, and immunocompromised individuals, including those with hemoglobinopathies, HIV infection, or infections that cause blockade of the reticuloendothelial system (e.g., bartonellosis, malaria, schistosomiasis, histoplasmosis).

Over the past three decades, bloodstream infection caused by invasive NTS, predominantly associated with closely related lineages of *S.* Typhimurium sequence type (ST) 313, as well as *S.* Enteritidis ST11, have emerged in sub-Saharan Africa and have spread to South Asia, the United Kingdom, and Brazil. These invasive NTS strains are adapted to person-to-person transmission through stepwise loss-of-function mutations and typically present with nonspecific febrile illness similar to enteric fever and uncommonly cause diarrhea. In 2017, there were ~535,000 invasive NTS cases and ~77,500 deaths, most of which (~80%) occurred in sub-Saharan Africa. Most (75%) *S.* Typhimurium ST313 isolates are MDR to ampicillin, trimethoprim-sulfamethoxazole, and chloramphenicol, and some are also resistant to ceftriaxone or ciprofloxacin, especially in South Asia. Recently, a sublineage of *S.* Typhimurium ST313 combining MDR with *both* ceftriaxone and azithromycin resistance has emerged in the Democratic Republic of the Congo. The incidence of invasive NTS infection is highest in children <5 years of age, exceeding 100 per 100,000 person-years in several West African countries, and risk is associated with malaria, HIV, malnutrition, and unclean drinking water sources. Transmission from asymptomatic stool carriers in the household and food or water sources has been proposed, but the sources of invasive NTS infection remain uncertain.

Unlike *S.* Typhi and *S.* Paratyphi, whose only reservoir is humans, NTS can be acquired from multiple animal and plant reservoirs that are part of the typical food supply. Transmission is most commonly associated with food products of animal origin (especially eggs, poultry, undercooked ground meat, and dairy products), fresh produce contaminated with animal waste, and contact with animals or their environments. In the United States, NTS are the second most common cause of food-borne outbreaks after norovirus, causing 30% of outbreaks and 35% of outbreak-associated illnesses.

S. Enteritidis infection associated with chicken eggs emerged as a major cause of food-borne disease during the 1980s and 1990s. *S.* Enteritidis infection of the ovaries and upper oviduct tissue of hens results in contamination of egg contents before shell deposition. Infection is spread to egg-laying hens from breeding flocks and through contact with rodents and manure. The number of *S.* Enteritidis outbreaks and the proportion attributable to egg-containing foods have continued to decline since the mid-1990s; these declines have coincided with interventions in the egg-producing and food service industries. Despite these control efforts, outbreaks of *S.* Enteritidis infection associated with shell eggs continue to occur. Transmission via contaminated eggs can be prevented by cooking eggs until the yolk is solidified and pasteurizing egg products. Increasingly, outbreaks of *S.* Enteritidis infection are associated with other foods, including meat, chicken, vegetables, dairy, and baked goods.

Salmonella serotype 4,[5],12:i:–, an antigenic variant of *S.* Typhimurium that lacks the second-stage flagellar antigen, has emerged since the 1990s as a foodborne pathogen primarily associated with pigs and pork products. This serotype is the second most common NTS in Europe and the fifth most common in the United States. These strains are MDR, with resistance to ampicillin, streptomycin, sulfonamides, and tetracycline. Increasing reports of plasmid-mediated colistin resistance in these strains have been linked to international travel to Thailand and swine farm isolates.

Centralization of food processing and widespread food distribution have contributed to the increased incidence of NTS in developed countries. NTS accounts for a significant majority of illnesses and hospitalizations associated with multistate foodborne outbreaks and retail food establishment outbreaks in the United States. Manufactured foods to which recent multistate *Salmonella* outbreaks have been traced include peanut butter; milk products, including powdered infant formula; and various processed foods, including packaged breakfast cereal, salsa, frozen prepared meals, and snack foods. Large outbreaks have also been linked to fresh produce, including alfalfa sprouts, nuts/seeds, cantaloupe, mangoes, papayas, tomatoes, and the herbal substance kratom consumed for its stimulant effect; these items become contaminated by manure or water at a single site and then are widely distributed.

In the United States, NTS infection associated with exotic pets is an ongoing clinical and public health problem, especially from contact with reptiles or amphibians, including iguanas, snakes, turtles, and lizards. Other pets, including hedgehogs, rodents, birds, baby chicks, ducklings, dogs, and cats, also are potential sources of NTS. Compared to foodborne outbreaks, outbreaks of NTS linked to animal contact more commonly affect young children (<1–4 years of age), result in hospitalization, and are more sustained.

Increasing antibiotic resistance in NTS species is a global problem and has been linked to the widespread use of antimicrobial agents in food animals and especially in animal feed. In the United States, clinically important resistant NTS infections, defined as resistance to ampicillin or ceftriaxone or nonsusceptibility to ciprofloxacin, increased an estimated 40% during 2015–2016 (annual incidence ~220,000) compared with 2004–2008 (~159,000 infections). The incidence (51.0 per 100,000 persons per year) and proportion of invasive NTS infections that are multidrug-resistant (75% with co-resistance to ampicillin, trimethoprim-sulfamethoxazole, and chloramphenicol) is dramatically higher across all sub-Saharan African regions. In the United States, infections caused by NTS with any antimicrobial resistance compared to NTS with no resistance are less likely to be associated with an outbreak and more likely to be associated with international travel, an increased risk of hospitalization, hospital length of stay, and death.

Outbreaks and sporadic cases of NTS resistant to third-generation cephalosporins have been reported, and international travel and adoption may have contributed to the global spread. Resistance is most commonly mediated by a transferable plasmid containing the *ampC* (bla_{CMY}) gene and has been linked to the widespread use of the veterinary cephalosporin ceftiofur. In 2021, 3.0% of NTS isolates from humans in the United States were ceftriaxone resistant (MIC, ≥4 μg/mL). Ceftriaxone resistance is more common among invasive NTS isolates from sub-Saharan Africa (>5%). Carbapenem-producing NTS have been reported sporadically in Europe, North Africa, and South Asia but remain extremely rare in the United States.

Since the early 2000s, strains of DSC NTS (MIC, ≥0.125 μg/mL) have emerged and have been associated with delayed response and treatment failure. In 2021, 10.6% of NTS isolates in the United States were DSC but only 0.5% of isolates were resistant to ciprofloxacin. These strains have diverse resistance mechanisms, including single and multiple mutations in the DNA gyrase genes *gyr*A and *gyr*B, mutations in the chromosomally encoded quinolone resistance–determining region, and plasmid-encoded quinolone resistance genes that are not reliably detected by nalidixic acid susceptibility testing or standard ciprofloxacin disk diffusion. In 2012, the U.S. Clinical Laboratory Standards Institute proposed a lower ciprofloxacin susceptibility breakpoint (≥0.06 μg/mL) for all *Salmonella* species to address this issue. Because some commercial test systems do not contain ciprofloxacin concentrations as low as the revised ≥0.06 μg/mL susceptibility breakpoint, laboratories can also determine the ciprofloxacin MIC by Etest or another alternative method.

In the United States, azithromycin-resistant NTS strains (MIC, ≥32 μg/mL) remain uncommon (1.1% in 2021), but most such isolates were also MDR. Azithromycin resistance is conferred by plasmid-encoded macrolide resistance genes, and their emergence may be related to international travel, especially to Thailand, and azithromycin overuse. Sporadic cases of carbapenemase-resistant NTS have been reported in Europe, North Africa, and southern Asia.

CLINICAL MANIFESTATIONS

Gastroenteritis Infection with NTS most often results in gastroenteritis indistinguishable from that caused by other enteric pathogens. Nausea, vomiting, and diarrhea occur 6–48 h after the ingestion of contaminated food or water. Patients often experience abdominal cramping and fever (38–39°C [100.5–102.2°F]). Diarrheal stools are usually loose, nonbloody, and of moderate volume. However, large-volume watery stools, bloody stools, or symptoms of dysentery may occur. Rarely, NTS causes pseudoappendicitis or an illness that mimics inflammatory bowel disease.

Gastroenteritis caused by NTS is usually self-limited. Diarrhea resolves within 3–7 days and fever within 72 h. Stool cultures remain positive for 4–5 weeks after infection—and in rare cases of chronic carriage (<1%) for >1 year. Persistent NTS infection and relapsing diarrhea have been described in a small fraction of Israeli patients and were associated with in-host single-nucleotide mutations in key virulence regulators. For acute NTS gastroenteritis, antibiotic treatment usually is not recommended and may prolong fecal carriage. Neonates, the elderly, and immunosuppressed patients (e.g., transplant recipients, HIV-infected persons) with NTS gastroenteritis are especially susceptible to dehydration and invasive infection and may require hospitalization and antibiotic therapy. Acute NTS gastroenteritis was associated with a threefold increased risk of dyspepsia and irritable bowel syndrome at 1 year in a study from Spain.

Bacteremia and Endovascular Infections Up to 8% of patients with NTS gastroenteritis develop bacteremia; of these, 5–10% develop localized infections. Bacteremia and metastatic infection are most common with *Salmonella* Choleraesuis and *Salmonella* Dublin and among infants, the elderly, and immunocompromised patients, especially those with HIV infection. NTS endovascular infection should be suspected in high-grade or persistent bacteremia, especially with preexisting valvular heart disease, atherosclerotic vascular disease, prosthetic vascular graft, or aortic aneurysm. Arteritis should be suspected in elderly patients with prolonged fever and back, chest, or abdominal pain developing after an episode of gastroenteritis. Endocarditis and arteritis are rare (<1% of cases) but are associated with serious and potentially fatal complications, including valve perforation, endomyocardial abscess, infected mural thrombus, pericarditis, mycotic aneurysms, aneurysm rupture, aortoenteric fistula, and vertebral osteomyelitis.

Invasive NTS disease is among the most common causes of bacteremia in children and in HIV-infected adults in sub-Saharan Africa and Southeast Asia, causing 39% of community-acquired bloodstream infection in one study. NTS bacteremia among these children is not associated with diarrhea and has been associated with poor nutritional status, malaria, sickle cell disease, and HIV infection. *S.* Typhimurium ST 131, the most common cause of invasive NTS disease in sub-Saharan Africa, forms a specific clade that is associated with genome reduction and loss of traits required for environmental stress resistance, likely contributing to making this strain more human adapted.

Localized Infections • **INTRAABDOMINAL INFECTIONS** Intraabdominal infections due to NTS are rare and usually manifest as hepatic or splenic abscesses or as cholecystitis. Risk factors include hepatobiliary anatomic abnormalities (e.g., gallstones), abdominal malignancy, and sickle cell disease (especially with splenic abscesses). Eradication of the infection often requires surgical correction of abnormalities and percutaneous drainage of abscesses.

CENTRAL NERVOUS SYSTEM INFECTIONS NTS meningitis most commonly develops in infants 1–4 months of age and in adults with HIV infection. It often results in severe sequelae (including seizures, hydrocephalus, brain infarction, and mental retardation), with death in up to 60% of cases. Other rare central nervous system infections include ventriculitis, subdural empyema, and brain abscesses.

PULMONARY INFECTIONS NTS pulmonary infections usually present as lobar pneumonia, and complications include lung abscess, empyema, and bronchopleural fistula formation. The majority of cases occur in patients with lung cancer, structural lung disease, sickle cell disease, or glucocorticoid use.

URINARY AND GENITAL TRACT INFECTIONS Urinary tract infections caused by NTS present as either cystitis or pyelonephritis. Risk factors include malignancy, urolithiasis, structural abnormalities, HIV infection, and renal transplantation. NTS genital infections are rare and include ovarian and testicular abscesses, prostatitis, and epididymitis. Like other focal infections, both genital and urinary tract infections can be complicated by abscess formation.

BONE, JOINT, AND SOFT TISSUE INFECTIONS *Salmonella* osteomyelitis most commonly affects the femur, tibia, humerus, or lumbar vertebrae and is most often seen in association with sickle cell disease, hemoglobinopathies, or preexisting bone disease (e.g., fractures). Prolonged antibiotic treatment is recommended to decrease the risk of relapse and chronic osteomyelitis. Septic arthritis occurs in the same patient population as osteomyelitis and usually involves the knee, hip, or shoulder joints. Reactive arthritis can follow NTS gastroenteritis and is seen most frequently in persons with the HLA-B27 histocompatibility antigen. NTS rarely can cause soft tissue infections, usually at sites of local trauma in immunosuppressed patients.

DIAGNOSIS

The diagnosis of NTS infection is based on isolation of the organism from freshly passed stool or from blood or another ordinarily sterile body fluid. *Salmonella* is increasingly identified by culture-independent molecular diagnostic tests due to increased sensitivity, rapid turnaround, and ability to detect multiple enteric pathogens in one test. Culture-independent positive specimens should have primary isolation performed to replicate results and recover NTS isolates. All NTS isolates should be referred to local public health departments for serotyping. Blood cultures should be obtained whenever a patient has prolonged or recurrent fever. Endovascular infection should be suspected if there is high-grade bacteremia (>50% of three or more sets of blood cultures positive). Echocardiography, CT, and indium-labeled white cell scanning are used to identify localized infection. When a localized infection is suspected, joint fluid aspiration, bone biopsy, abscess drainage, or cerebrospinal fluid should be cultured, as clinically indicated.

TREATMENT

Nontyphoidal Salmonellosis

Antibiotics should not be used routinely to treat uncomplicated NTS gastroenteritis. The symptoms are usually self-limited, and the duration of fever and diarrhea is not significantly decreased by

TABLE 171-2 Antibiotic Therapy for Nontyphoidal *Salmonella* Infection in Adults

INDICATION	AGENT	DOSAGE (ROUTE)	DURATION, DAYS
Preemptive Treatment[a]			
	Ciprofloxacin[b]	500 mg bid (PO)	2–3
Severe Gastroenteritis[c]			
	Ciprofloxacin	500 mg bid (PO) or 400 mg q12h (IV)	7
	Azithromycin	500 mg once daily	5
	Trimethoprim-sulfamethoxazole	160/800 mg bid (PO)	7
	Amoxicillin	1 g tid (PO)	7
	Ceftriaxone	1–2 g/d (IV)	7
Bacteremia			
	Ceftriaxone[d]	2 g/d (IV)	7–14
	Ciprofloxacin	400 mg q12h (IV), then 500 mg bid (PO)	
Endocarditis or Arteritis			
	Ceftriaxone	2 g/d (IV)	42
	Ciprofloxacin	400 mg q8h (IV), then 750 mg bid (PO)	
	Ampicillin	2 g q4h (IV)	
Meningitis			
	Ceftriaxone	2 g q12h (IV)	14–21
	Ampicillin	2 g q4h (IV)	
Other Localized Infection			
	Ceftriaxone	2 g/d (IV)	14–28
	Ciprofloxacin	500 mg bid (PO) or 400 mg q12h (IV)	
	Ampicillin	2 g q6h (IV)	

[a]Consider for neonates; persons >50 years of age with possible atherosclerotic vascular disease; and patients with immunosuppression, endovascular graft, or joint prosthesis. [b]Or ofloxacin, 400 mg bid (PO). [c]Consider on an individualized basis for patients with severe diarrhea and high fever who require hospitalization.
[d]Or cefotaxime, 2 g q8h (IV).

antibiotic therapy. In addition, antibiotic treatment has been associated with increased rates of relapse, prolonged gastrointestinal carriage, and adverse drug reactions. Dehydration secondary to diarrhea should be treated with fluid and electrolyte replacement.

Preemptive antibiotic treatment **(Table 171-2)** should be considered for patients at increased risk for invasive NTS infection, including neonates (probably up to 3 months of age); persons >50 years of age with known or suspected atherosclerosis; and patients with immunosuppression, cardiac valvular or endovascular abnormalities, or significant joint disease. Treatment should consist of an oral or IV antibiotic administered for 48–72 h or until the patient becomes afebrile. Immunocompromised persons may require up to 7–14 days of therapy. The <1% of persons who develop chronic carriage of NTS should receive a prolonged antibiotic course, as described above for chronic carriage of *S.* Typhi.

Because of the increasing prevalence of antibiotic resistance, empirical therapy for life-threatening NTS bacteremia or focal NTS infection should include a third-generation cephalosporin or a fluoroquinolone (Table 171-2). If the bacteremia is low-grade (<50% of blood cultures positive), the patient should be treated for 7–14 days. Patients with HIV/AIDS and NTS bacteremia should receive 1–2 weeks of IV antibiotic therapy followed by 4 weeks of oral therapy with a fluoroquinolone. Patients whose infections relapse after this regimen should receive long-term suppressive therapy with a fluoroquinolone, azithromycin, or TMP-SMX, as indicated by bacterial sensitivities.

If the patient has endocarditis or arteritis, treatment for 6 weeks with an IV β-lactam antibiotic (such as ceftriaxone or ampicillin) is indicated. IV ciprofloxacin followed by prolonged oral therapy is an option. Early surgical resection of infected aneurysms or other infected endovascular sites is recommended. Patients with infected prosthetic vascular grafts that cannot be resected have been maintained successfully on chronic suppressive oral therapy. For extraintestinal nonvascular infections, a 2- to 4-week course of antibiotic therapy (depending on the infection site) is usually recommended. In chronic osteomyelitis, abscess, or urinary or hepatobiliary infection associated with anatomic abnormalities, surgical resection or drainage may be required in addition to prolonged antibiotic therapy for eradication of infection.

PREVENTION AND CONTROL

Immunization against Invasive NTS Despite an urgent need for safe and effective vaccines, there is no available vaccine to protect against invasive NTS disease, and few candidate vaccines have progressed into early clinical trials. Candidate human NTS vaccines are O-antigen based and contain *S.* Typhimurium and *S.* Enteritidis bivalent components. NTS vaccines will need to target young children and protect those in at-risk groups, including persons with HIV infection and with malaria and those with sickle cell disease. In addition, rapid, point-of-care diagnostics are critically needed to reduce the morbidity and mortality associated with invasive NTS infection.

Despite widespread efforts to prevent or reduce bacterial contamination of animal-derived food products and to improve food-safety education and training, recent declines in the incidence of NTS in the United States have been modest compared with those of other food-borne pathogens. This observation probably reflects the complex epidemiology of NTS. Identifying effective risk-reduction strategies requires monitoring of every step of the food supply chain, including farm sources, slaughter and processing of raw animal or plant products, storage and transport, and preparation of finished foods. Contaminated food can be made safe for consumption by pasteurization, irradiation, or proper cooking. All cases of NTS infection should be reported to local public health departments because tracking and monitoring of these cases can identify the source(s) of infection and help authorities anticipate large outbreaks. Prudent use of antimicrobial agents in both humans and animals is needed to limit the emergence of MDR and XDR *Salmonella*.

FURTHER READING

Cruz Espinoza LM et al: Occurrence of typhoid fever complications and their relation to duration of illness preceding hospitalization: A systematic literature review and meta-analysis. Clin Infect Dis 69:S435, 2019.

Kuehn R et al: Treatment of enteric fever (typhoid and paratyphoid fever) with cephalosporins. Cochrane Database Syst Rev 11:CD010452, 2022.

Marchello CS et al: Complications and mortality of non-typhoidal salmonella invasive disease: A global systematic review and meta-analysis. Lancet Infect Dis 22:692, 2022.

Medalla F et al: Increased Incidence of Antimicrobial-Resistant Nontyphoidal Salmonella Infections, United States, 2004-2016. Emerg Infect Dis 27:1662-1672, 2021.

Milligan R et al: Vaccines for preventing typhoid fever. Cochrane Database Syst Rev 5:CD001261, 2018.

Onwuezobe IA et al: Antimicrobials for treating symptomatic non-typhoidal *Salmonella* infection. Cochrane Database Syst Rev 11:CD001167, 2012.

Pulford CV et al: Stepwise evolution of Salmonella Typhimurium ST313 causing bloodstream infection in Africa. Nat Microbiol 6:327, 2021.

Watkins LKF et al: Clinical outcomes of patients with nontyphoidal *Salmonella* infections by isolate resistance—foodborne diseases active surveillance network, 10 US sites, 2004 – 2018. Clin Infect Dis 78:535, 2024.

172

Shigellosis

Philippe J. Sansonetti, Jean Bergounioux

The discovery of *Shigella* as the etiologic agent of dysentery—a clinical syndrome of fever, intestinal cramps, and frequent passage of small, bloody, mucopurulent stools—is attributed to the Japanese microbiologist Kiyoshi Shiga, who isolated the Shiga bacillus (now known as *Shigella dysenteriae* type 1) from patients' stools in 1897 during a large and devastating dysentery epidemic. *Shigella* cannot be distinguished from *Escherichia coli* by genome comparison and remains a separate species only on historical and clinical grounds.

ETIOLOGIC AGENT

Shigella is a non-spore-forming, gram-negative bacterium that, unlike *E. coli*, is nonmotile and does not produce gas from sugars, decarboxylate lysine, or hydrolyze arginine. Some serovars produce indole, and occasional strains utilize sodium acetate. *Shigella dysenteriae*, *Shigella flexneri*, *Shigella boydii*, and *Shigella sonnei* (serogroups A, B, C, and D, respectively) can be differentiated on the basis of biochemical and serologic characteristics.

Genome sequencing of *E. coli* K12, *S. flexneri* 2a, *S. sonnei*, *S. dysenteriae* type 1, and *S. boydii* has revealed that these species have ~93% of genes in common. The three major genomic "signatures" of *Shigella* are (1) a 215-kb virulence plasmid that carries most of the genes required for pathogenicity (particularly invasive capacity); (2) the lack or alteration of genetic sequences encoding products (e.g., lysine decarboxylase) that, if expressed, would attenuate pathogenicity; and (3) in *S. dysenteriae* type 1, the presence of genes encoding Shiga toxin, a potent cytotoxin.

EPIDEMIOLOGY

The human intestinal tract is the major reservoir of *Shigella*, which is also found (albeit rarely) in the higher primates. Because excretion of shigellae is greatest in the acute phase of disease, the bacteria are transmitted most efficiently by the fecal–oral route via hand carriage; however, some outbreaks reflect foodborne or waterborne transmission. In impoverished areas, *Shigella* can be transmitted by flies. The high-level infectivity of *Shigella* is reflected by the very small inoculum required for experimental infection of volunteers (100 colony-forming units [CFU]), by the very high attack rates during outbreaks in day-care centers (33–73%), and by the high rates of secondary cases among family members of sick children (26–33%). Shigellosis can also be transmitted sexually.

Throughout history, bacillary dysentery epidemics have often occurred in settings of human crowding under conditions of poor hygiene—e.g., among soldiers in campaigning armies, inhabitants of besieged cities, groups on pilgrimages, and refugees in camps. Epidemics followed a cyclical pattern in areas such as the Indian subcontinent and sub-Saharan Africa. These patterns of devastating epidemics, which were most often caused by *S. dysenteriae* type 1, were characterized by high attack and mortality rates. In Bangladesh, for instance, an epidemic caused by *S. dysenteriae* type 1 was associated with a 42% increase in mortality rate among children 1–4 years of age. This pattern of severe cyclic disease has steeply declined over the last few decades. Along with the increasingly rare isolation of *S. dysenteriae 1*, the current epidemiology of shigellosis is generally attributed to a significant global reduction of extreme poverty; this coincides with the tremendous decrease in child mortality.

Nevertheless, apart from these severe epidemics of the past, shigellosis remains an endemic disease, with 99% of cases occurring in developing countries, with the highest prevalence in the most impoverished areas, where personal and general hygiene is below standard. *S. flexneri* isolates predominate in the least developed areas, whereas *S. sonnei* is more prevalent in economically emerging countries and in the industrialized world.

Prevalence in the Developing World In a review published under the auspices of the World Health Organization (WHO), the total annual number of cases in 1966–1997 was estimated at 165 million, and 69% of these cases occurred in children <5 years of age. In this review, the annual number of deaths was calculated to range between 500,000 and 1.1 million. Data (2000–2004) from six Asian countries indicate that, even though the incidence of shigellosis remains stable, mortality rates associated with this disease may have decreased significantly, possibly as a result of improved nutritional status. However, extensive and essentially uncontrolled use of antibiotics, which may also account for declining mortality rates, has increased the emergence of multidrug-resistant *Shigella* strains.

A 2013 prospective, matched, case-control study of children <5 years of age emphasized the importance of *Shigella* in the burden and etiology of diarrheal diseases in developing countries. *Shigella* was one of the top four pathogens associated with moderate to severe diarrhea and ranked first among children 12–59 months of age. These moderate to severe cases accounted for an 8.5-fold increase in mortality incidence over the average diarrheal disease-related mortality. The study concluded that *Shigella* remained a major pathogen to be targeted by health care programs.

Global reevaluation of the *Shigella* disease burden is warranted because an often-overlooked complication of shigellosis is the short- and long-term impairment of the nutritional status of infected children in endemic areas. Combined with anorexia, the exudative enteropathy resulting from mucosal abrasions contributes to rapid deterioration of the patient's nutritional status. Shigellosis thus now appears as a major contributor to stunted growth among children in endemic regions.

Peaking in incidence in the pediatric population, endemic shigellosis is rare among young and middle-aged adults, probably because of naturally acquired immunity. Incidence then increases again in the elderly population.

Prevalence in the Industrialized World In pediatric populations, local outbreaks occur when proper and adapted hygiene policies are not implemented in group facilities such as day-care centers and institutions for the developmentally disabled. In adults, as in children, sporadic cases occur among travelers returning from endemic areas, and rare outbreaks of varying size can follow waterborne or foodborne infections.

PATHOGENESIS AND PATHOLOGY

Shigella infection occurs essentially through oral contamination via direct fecal–oral transmission, the organism being poorly adapted to survive in the environment. Resistance to low-pH conditions allows *Shigella* to survive passage through the gastric barrier.

The watery diarrhea that usually precedes dysenteric symptoms is attributable to active secretion and abnormal water reabsorption in the jejunum, as described in experimentally infected rhesus monkeys. This initial purge is possibly due to the combined action of an enterotoxin (ShET-1) and mucosal inflammation. The dysenteric syndrome, manifested by bloody and mucopurulent stools, reflects invasion of the colonic mucosa.

The pathogenesis of *Shigella* is essentially determined by a large virulence plasmid of 214 kb comprising ~100 genes, of which 25 encode a type III secretion system that inserts into the membrane of the host cell to allow effectors to transit from the bacterial cytoplasm to the host cell cytoplasm **(Fig. 172-1)**. Bacteria are thereby able to invade intestinal epithelial cells by inducing their own uptake either directly at the opening of colonic crypts or following the initial crossing of the epithelial barrier through M cells (the specialized translocating epithelial cells in the follicle-associated epithelium that covers mucosal lymphoid nodules). *Shigella* induces apoptosis of subepithelial resident macrophages. Once inside the cytoplasm of intestinal epithelial cells, *Shigella* effectors trigger the cytoskeletal rearrangements necessary to direct uptake of the organism into the epithelial cell. The *Shigella*-containing vacuole is then quickly lysed, releasing bacteria into the cytosol.

Intracellular shigellae next use cytoskeletal components to propel themselves inside the infected cell; when the moving organism and the host cell membrane come into contact, cellular protrusions form and

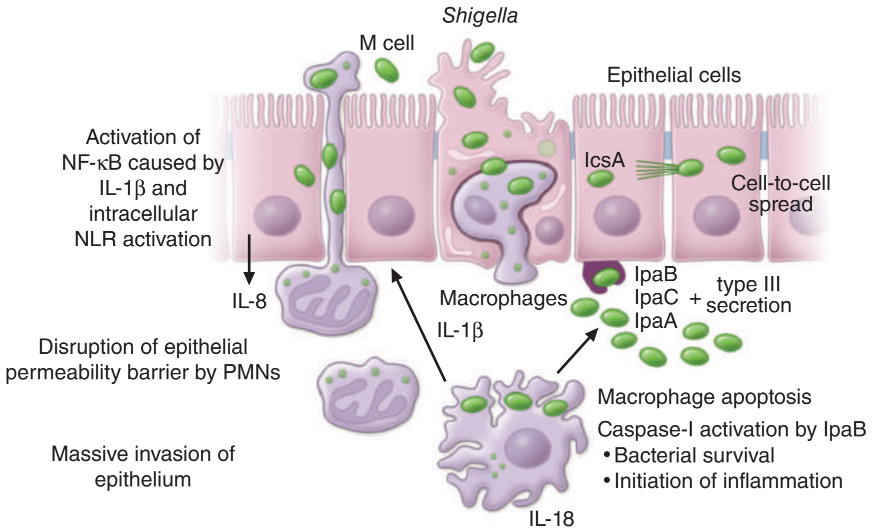
Shigella
M cell
Epithelial cells
Activation of NF-κB caused by IL-1β and intracellular NLR activation
IcsA
Cell-to-cell spread
IL-8
IpaB
IpaC
IpaA
+ type III secretion
Macrophages
IL-1β
Disruption of epithelial permeability barrier by PMNs
Macrophage apoptosis
Caspase-I activation by IpaB
• Bacterial survival
• Initiation of inflammation
Massive invasion of epithelium
IL-18

FIGURE 172-1 Invasive strategy of *Shigella flexneri*. IL, interleukin; NF-κB, nuclear factor κB; NLR, NOD-like receptor; PMN, polymorphonuclear leukocyte.

are engulfed by neighboring cells. This series of events permits bacterial cell-to-cell spread.

Cytokines released by a growing number of infected intestinal epithelial cells massively attract immune cells—particularly polymorphonuclear leukocytes [PMNs]—to the infected site, thus further destabilizing the epithelial barrier, exacerbating inflammation, and leading to the acute colitis that characterizes shigellosis. Evidence indicates that some type III secretion system–injected effectors can control the extent of inflammation, thus facilitating bacterial survival.

Shiga toxin produced by *S. dysenteriae* type 1 increases disease severity. This toxin belongs to a group of A1-B5 protein toxins whose B subunit binds to the receptor globotriaosylceramide on the target cell surface and whose catalytic A subunit is internalized by receptor-mediated endocytosis and interacts with the subcellular machinery to inhibit protein synthesis by expressing RNA *N*-glycosidase activity on 28S ribosomal RNA. This process leads to inhibition of binding of the amino-acyl-tRNA to the 60S ribosomal subunit and thus to a general shutoff of cell protein biosynthesis. Shiga toxins are translocated from the bowel into the circulation. After binding of the toxins to target cells in the kidney, pathophysiologic alterations may result in hemolytic-uremic syndrome (HUS; see below).

CLINICAL MANIFESTATIONS

The presentation and severity of shigellosis depend to some extent on the infecting serotype but even more on the age and the immunologic and nutritional status of the host. Poverty and poor standards of hygiene are strongly related to the number and severity of diarrheal episodes, especially in children <5 years following weaning.

Shigellosis typically evolves through four phases: incubation, watery diarrhea, dysentery, and the postinfectious phase. The incubation period usually lasts 1–4 days but may be as long as 8 days. Typical initial manifestations are transient fever, limited watery diarrhea, malaise, and anorexia. Signs and symptoms may range from mild abdominal discomfort to severe cramps, diarrhea, fever, vomiting, and tenesmus. The manifestations are usually exacerbated in children, with temperatures up to 40°–41°C (104.0°–105.8°F) and more severe anorexia and watery diarrhea. This initial phase may represent the only clinical manifestation of shigellosis, especially in developed countries. Otherwise, dysentery follows within hours or days and is characterized by uninterrupted excretion of small volumes of bloody mucopurulent stools with increased tenesmus and abdominal cramps. At this stage, *Shigella* produces acute colitis involving mainly the distal colon and the rectum. Unlike most diarrheal syndromes, dysenteric syndromes rarely present with dehydration as a major feature. Endoscopy, if performed, shows an edematous and hemorrhagic mucosa, with ulcerations and possibly overlying exudates resembling pseudomembranes. The extent of the lesions correlates with the number and frequency of stools and with the degree of protein loss by exudative mechanisms. Most episodes are self-limited and resolve without treatment in 1 week. With appropriate treatment, recovery takes place within a few days to a week, with no sequelae.

Acute life-threatening complications are seen most often in children <5 years (particularly those who are malnourished) and in elderly patients. Risk factors for death in a clinically severe case include nonbloody diarrhea, moderate to severe dehydration, bacteremia, absence of fever, abdominal tenderness, and rectal prolapse. Major complications are predominantly intestinal (e.g., toxic megacolon, intestinal perforations, rectal prolapse) or metabolic (e.g., hypoglycemia, hyponatremia, dehydration). Bacteremia is rare and is reported most frequently in severely malnourished and HIV-infected patients. Alterations of consciousness, including seizures, delirium, and coma, may occur, especially in children <5 years, and are associated with a poor prognosis; fever and severe metabolic alterations are more often the major causes of altered consciousness than is meningitis or the Ekiri syndrome (toxic encephalopathy associated with bizarre posturing, cerebral edema, and fatty visceral degeneration), which has been reported mostly in Japanese children. Pneumonia, vaginitis, and keratoconjunctivitis due to *Shigella* are rarely reported. In the absence of serious malnutrition, severe and very unusual clinical manifestations, such as meningitis, may be linked to genetic defects in innate immune functions (i.e., deficiency in interleukin 1 receptor–associated kinase 4 [IRAK-4]) and may require genetic investigation.

Two complications of particular importance are toxic megacolon and HUS. Toxic megacolon is a consequence of severe inflammation extending to the colonic smooth-muscle layer and causing paralysis and dilation. The patient presents with abdominal distention and tenderness, with or without signs of localized or generalized peritonitis. The abdominal x-ray characteristically shows marked dilation of the transverse colon (with the greatest distention in the ascending and descending segments); thumbprinting caused by mucosal inflammatory edema; and loss of the normal haustral pattern associated with pseudopolyps, often extending into the lumen. Pneumatosis coli is an occasional finding. If perforation occurs, radiographic signs of pneumoperitoneum may be apparent. Predisposing factors (e.g., hypokalemia and use of opioids, anticholinergics, loperamide, psyllium seeds, and antidepressants) should be investigated.

Shiga toxin produced by *S. dysenteriae* type 1 has been linked to HUS in developing countries but rarely in industrialized countries, where enterohemorrhagic *E. coli* (EHEC) predominates as the etiologic agent of this syndrome. HUS is an early complication that most often develops after several days of diarrhea. Clinical examination shows pallor, asthenia, and irritability and, in some cases, bleeding of the nose and gums, oliguria, and increasing edema. HUS is a nonimmune (Coombs-negative) hemolytic anemia defined by a diagnostic triad: microangiopathic hemolytic anemia (hemoglobin level typically <80 g/L [<8 g/dL]), thrombocytopenia (mild to moderate in severity; typically <60,000 platelets/μL), and acute renal failure due to thrombosis of the glomerular capillaries (with markedly elevated creatinine levels). Anemia is severe, with fragmented red blood cells (*schizocytes*) in the peripheral smear, high serum concentrations of lactate dehydrogenase and free circulating hemoglobin, and elevated reticulocyte counts. Acute renal failure occurs in 55–70% of cases; however, renal function recovers in most of these cases (up to 70% in various series). Leukemoid reactions, with leukocyte counts of 50,000/μL, are sometimes present in association with HUS.

The postinfectious immunologic complication known as *reactive arthritis* can develop weeks or months after shigellosis, especially in patients expressing the histocompatibility antigen HLA-B27. About 3% of patients infected with *S. flexneri* later develop this syndrome, with arthritis, ocular inflammation, and urethritis—a condition that can last for months or years and can progress to difficult-to-treat chronic arthritis. Postinfectious arthritis occurs only after infection with *S. flexneri* and not after infection with the other *Shigella* serotypes.

LABORATORY DIAGNOSIS

The differential diagnosis in patients with a dysenteric syndrome depends on the clinical and environmental context. In developing areas, infectious diarrhea caused by other invasive pathogenic bacteria (*Salmonella*, *Campylobacter jejuni*, *Clostridium difficile*, *Yersinia enterocolitica*) or parasites (*Entamoeba histolytica*) should be considered. Only bacteriologic and parasitologic examinations of stool can truly differentiate among these pathogens. A first flare of inflammatory bowel disease, such as Crohn's disease or ulcerative colitis **(Chap. 337)**, should be considered in patients in industrialized countries. Despite the similarity in symptoms, anamnesis discriminates between shigellosis, which usually follows recent travel in an endemic zone, and these other conditions.

Microscopic examination of stool smears shows erythrophagocytic trophozoites with very few PMNs in *E. histolytica* infection, whereas bacterial enteroinvasive infections (particularly shigellosis) are characterized by high PMN counts in each microscopic field. However, because shigellosis often manifests only as watery diarrhea, systematic attempts to isolate *Shigella* are necessary.

The "gold standard" diagnosis of *Shigella* infection remains the isolation and identification of the pathogen from fecal material. One major difficulty, particularly in endemic areas where laboratory facilities are not immediately available, is the fragility of *Shigella* and its common disappearance during transport, especially with rapid changes in temperature and pH. In the absence of a reliable enrichment medium, buffered glycerol saline or Cary-Blair medium can be used as a holding medium, but prompt inoculation onto isolation medium is essential. The probability of isolation is higher if the portion of stools that contains bloody and/or mucopurulent material is directly sampled. Rectal swabs can be used, as they offer the highest rate of successful isolation during the acute phase of disease. Blood cultures are positive in fewer than 5% of cases but should be done when a patient presents with a clinical picture of severe sepsis.

In addition to quick processing, the use of several media increases the likelihood of successful isolation: a nonselective medium such as bromocresol-purple agar lactose; a low-selectivity medium such as MacConkey or eosin-methylene blue; and a high-selectivity medium such as Hektoen, *Salmonella-Shigella*, or xylose-lysine-deoxycholate agar. After incubation on these media for 12–18 h at 37°C (98.6°F), shigellae appear as non-lactose-fermenting colonies that measure 0.5–1 mm in diameter and have a convex, translucent, smooth surface. Suspected colonies on nonselective or low-selectivity medium can be subcultured on a high-selectivity medium before being specifically identified or can be identified directly by standard commercial systems on the basis of four major characteristics: glucose positivity (usually without production of gas), lactose negativity, H_2S negativity, and lack of motility. The four *Shigella* serogroups (A–D) can then be differentiated by additional characteristics. This approach adds time and difficulty to the identification process; however, after presumptive diagnosis, the use of serologic methods (e.g., slide agglutination, with group- and then type-specific antisera) should be considered. Group-specific antisera are widely available; in contrast, because of the large number of serotypes and subserotypes, type-specific antisera are rare and more expensive and thus are often restricted to reference laboratories.

Molecular methods of diagnostics, including polymerase chain reaction based on *Shigella*-specific virulence gene sequences or mass spectrometry, are not yet standardized for global use. However, serotyping may soon be replaced by genome-based techniques.

TREATMENT

Shigellosis

ANTIBIOTIC SUSCEPTIBILITY OF *SHIGELLA*

As an enteroinvasive disease, shigellosis requires antibiotic treatment. Since the mid-1960s, however, increasing resistance to multiple drugs has been a dominant factor in treatment decisions. Resistance rates are highly dependent on the geographic area. Clonal spread of particular strains and horizontal transfer of resistance determinants, particularly via plasmids and transposons, contribute to multidrug resistance. The current global status—i.e., high rates of resistance to classic first-line antibiotics such as amoxicillin—has led to a rapid switch to quinolones such as nalidixic acid. However, resistance to such early-generation quinolones has also emerged and spread quickly as a result of chromosomal mutations affecting DNA gyrase and topoisomerase IV; this resistance has necessitated the use of later-generation quinolones as first-line antibiotics in many areas. For instance, a review of the antibiotic resistance history of *Shigella* in India found that, after their introduction in the late 1980s, the second-generation quinolones norfloxacin, ciprofloxacin, and ofloxacin were highly effective in the treatment of shigellosis, including cases caused by multidrug-resistant strains of *S. dysenteriae* type 1. However, investigations of subsequent outbreaks in India and Bangladesh detected resistance to norfloxacin, ciprofloxacin, and ofloxacin in 5% of isolates. In the United States, the resistance rate of *Shigella* to fluoroquinolones reached 87% during 2014–2015. The incidence of multidrug resistance parallels the widespread, uncontrolled use of antibiotics and calls for the rational use of effective drugs. Despite the alarming proportion of resistant *Shigella*, there is a lack of studies assessing the resistance of community-acquired strains.

ANTIBIOTIC TREATMENT OF SHIGELLOSIS (TABLE 172-1)

With effective antibiotic therapy, clinical improvement occurs within 48 h, resulting in a decreased risk of complications and death, shorter duration of symptoms, and elimination of *Shigella* from the stools. Because of the transmissibility of *Shigella*, current public health recommendations in the United States are that every case be treated with antibiotics. The use of fluoroquinolones (first-line, preferably ciprofloxacin) and cephalosporins and β-lactams (second-line) for 7–10 days is recommended for the treatment of shigellosis. Whereas infections caused by non-*dysenteriae Shigella* in immunocompetent individuals are routinely treated with a 3-day course of antibiotics, it is recommended that *S. dysenteriae* type 1 infections be treated for 5 days and that *Shigella* infections in immunocompromised patients be treated for 7–10 days.

TABLE 172-1 Recommended Antimicrobial Therapy for Shigellosis

ANTIMICROBIAL AGENT	TREATMENT SCHEDULE		LIMITATIONS
	CHILDREN	ADULTS	
First-Line			
Ciprofloxacin	15 mg/kg	500 mg	
	2 times per day for 3 days, PO		
Second-Line			
Pivmecillinam	20 mg/kg	100 mg	Cost
	4 times per day for 5 days PO		No pediatric formulation Frequent administration Emerging resistance
Ceftriaxone	50–100 mg/kg	—	Efficacy not validated Must be injected
	Once a day IM for 2–5 days		
Azithromycin	6–20 mg/kg	1–1.5 g	Cost
	Once a day for 1–5 days PO		Efficacy not validated Minimum inhibitory concentration near serum concentration Rapid emergence of resistance and spread to other bacteria

Source: Reproduced with permission from World Health Organization: Guidelines for the control of shigellosis, including epidemics due to *Shigella dysenteriae* type 1. *https://www.who.int/publications/i/item/9241592330.*

Treatment for shigellosis must be adapted to the clinical context, with the recognition that the most fragile patients are children <5 years, who represent two-thirds of all cases worldwide and are the group most at risk of severe complications. There are few data on the use of quinolones in children, but *Shigella* dysentery is an accepted indication. The half-life of ciprofloxacin is longer in infants than in older individuals. The ciprofloxacin dose generally recommended for children is 30 mg/kg per day in two divided doses. Adults living in areas with high standards of hygiene are likely to develop milder, shorter-duration disease, whereas infants in endemic areas can develop severe, sometimes fatal, dysentery. In the former setting, treatment will remain minimal and bacteriologic proof of infection will often come after symptoms have resolved; in the latter setting, antibiotic treatment and more aggressive measures, possibly including resuscitation, are often required.

REHYDRATION AND NUTRITION

Shigella infection rarely causes significant dehydration, particularly in industrialized countries. It is recommended that rehydration should be oral unless the patient is comatose or in shock. Because of the improved effectiveness of reduced-osmolarity oral rehydration solutions, WHO and UNICEF now recommend a standard solution of 245 mOsm/L (sodium, 75 mmol/L; chloride, 65 mmol/L; glucose [anhydrous], 75 mmol/L; potassium, 20 mmol/L; citrate, 10 mmol/L). In shigellosis, the coupled transport of sodium and glucose may be variably affected, but oral rehydration therapy remains the easiest and most efficient form of rehydration, especially in severe cases.

Nutrition should be started as soon as possible, keeping in mind that in developing countries, malnutrition remains the primary indicator of the risk of diarrhea-related death. Early refeeding is safe, well tolerated, and clinically beneficial. Because breast-feeding reduces diarrheal losses and the need for oral rehydration in infants, it should be maintained in the absence of contraindications (e.g., maternal HIV infection).

NONSPECIFIC, SYMPTOM-BASED THERAPY

Antimotility agents have been implicated in prolonged fever in volunteers with shigellosis. These agents are suspected of increasing the risk of toxic megacolon and are thought to have been responsible for HUS in children infected by EHEC strains. For safety reasons, it is better to avoid antimotility agents in bloody diarrhea.

TREATMENT OF COMPLICATIONS

There is no consensus regarding the best treatment for toxic megacolon. The patient should be assessed frequently by both medical and surgical teams. Anemia, dehydration, and electrolyte deficits (particularly hypokalemia) may aggravate colonic atony and should be actively treated. Nasogastric aspiration helps to deflate the colon. Parenteral nutrition has not been proven to be beneficial. Fever persisting beyond 48–72 h raises the possibility of local perforation or abscess. Most studies recommend colectomy if, after 48–72 h, colonic distention persists. However, some physicians recommend continuation of medical therapy for up to 7 days if the patient seems to be improving clinically despite persistent megacolon without perforation. Intestinal perforation, either isolated or complicating toxic megacolon, requires surgical treatment and intensive medical support.

Rectal prolapse must be treated as soon as possible. With the health care provider using surgical gloves or a soft warm wet cloth and the patient in the knee-chest position, the prolapsed rectum is gently pushed back into place. If edema of the rectal mucosa is evident, making reduction difficult, it can be osmotically reduced by application of gauze impregnated with a warm solution of saturated magnesium sulfate. Rectal prolapse often relapses but usually resolves along with the resolution of dysentery.

HUS must be treated by water restriction, including discontinuation of oral rehydration solutions and potassium-rich nutrition. Hemofiltration or peritoneal dialysis is often required.

PREVENTION

Hand washing after defecation or handling of children's feces and before handling of food is recommended. Stool decontamination (e.g., with sodium hypochlorite), together with a cleaning protocol for medical staff as well as for patients, has proven useful in limiting the spread of infection during *Shigella* outbreaks. Ideally, patients should have a negative stool culture before their infection is considered cured. Recurrences are rare if therapeutic and preventive measures are correctly implemented. Protection against fly intrusion in commonly infested sites such as kitchens and latrines is strongly advised in endemic areas.

Although several live attenuated oral and subunit parenteral vaccine candidates have been produced and are undergoing clinical trials, no vaccine against shigellosis is currently available. Especially given the rapid progression of antibiotic resistance in *Shigella* and its increasing recognition as a cause of child stunting, a vaccine is urgently needed. Most recent evidence indicates that synthetic polysaccharide conjugate vaccines (based on chemically synthesized, concatenated lipopolysaccharide O-side chain mimics) elicit high levels of protective IgG antibodies, hence promising protection. Protection would be particularly important for children <3 years, the population most susceptible to severe forms of shigellosis.

FURTHER READING

Arena ET et al: Bioimage analysis of *Shigella* infection reveals targeting of colonic crypts. Proc Natl Acad Sci USA 112:E3282, 2015.

Bennish ML, Wojtyniak BJ: Mortality due to shigellosis: Community and hospital data. Rev Infect Dis 13(Suppl 4):S245, 1991.

Cohen D et al: Safety and immunogenicity of a synthetic carbohydrate conjugate vaccine against *Shigella flexneri* 2a in healthy adult volunteers: A phase 1, dose-escalating, single-blind, randomised, placebo-controlled study. Lancet Infect Dis 21:546, 2021.

Cossart P, Sansonetti PJ: Bacterial invasion: The paradigms of enteroinvasive pathogens. Science 304:242, 2004.

Kotloff KL et al: The incidence, aetiology, and adverse clinical consequences of less severe diarrhoeal episodes among infants and children residing in low-income and middle-income countries: A 12-month case-control study as a follow-on to the Global Enteric Multicenter Study (GEMS). Lancet Glob Health 7:E568, 2019.

Mani S et al: Status of vaccine research and development for *Shigella*. Vaccine 34:2887, 2016.

Niyogi SK: Shigellosis. J Microbiol 43:133, 2005.

Phalipon A, Sansonetti PJ: *Shigella's* ways of manipulating the host intestinal innate and adaptive immune system: A tool box for survival? Immunol Cell Biol 85:119, 2007.

Traa BS et al: Antibiotics for the treatment of dysentery in children. Int J Epidemiol 39:i70, 2010.

World Health Organization: Guidelines for the control of shigellosis, including epidemics due to *Shigella dysenteriae* type 1. WHO Library Cataloguing-in-Publication Data. *www.who.int/cholera/publications/shigellosis/en/*.

173

Cholera and Other Vibrioses

Matthew K. Waldor, Edward T. Ryan

Members of the genus *Vibrio* cause several important infectious syndromes. Classic among them is cholera, a devastating diarrheal disease caused by *Vibrio cholerae* that has been responsible for seven global pandemics and much suffering over the past two centuries. Epidemic cholera remains a significant public-health concern in the developing world today. Other vibrioses caused by other *Vibrio* species include syndromes of diarrhea, soft tissue infection, or primary sepsis. All *Vibrio* species are highly motile, facultatively anaerobic, curved, gram-negative rods with one or more flagella. In nature, vibrios most commonly reside in tidal rivers and bays under conditions of moderate salinity. They proliferate in the summer months when water temperatures exceed 20°C, and the illnesses they cause also increase in frequency during the warm months.

CHOLERA

DEFINITION

Cholera is an acute diarrheal disease that can, in a matter of hours, result in profound, rapidly progressive dehydration and death. Accordingly, *cholera gravis* (the severe form) is a much-feared disease, particularly in its epidemic presentation. Fortunately, prompt aggressive fluid repletion and supportive care can obviate the high mortality that is historically associated with cholera. Although the term *cholera* has occasionally been applied to any severely dehydrating secretory diarrheal illness, whether infectious in etiology or not, it now refers to disease caused by *V. cholerae* serogroup O1 or O139—i.e., the serogroups with epidemic potential.

MICROBIOLOGY AND EPIDEMIOLOGY

The species *V. cholerae* is classified into >200 serogroups based on the carbohydrate constituents of their lipopolysaccharide (LPS) O antigens. Although some non-O1 *V. cholerae* serogroups (strains that do not agglutinate in antisera to the O1 group antigen) have occasionally caused sporadic outbreaks of diarrhea, serogroup O1 was, until the emergence of serogroup O139 in 1992 (see below), the exclusive cause of epidemic cholera. The O1 serogroup is further subdivided into two serotypes, termed *Inaba* and *Ogawa*. Two biotypes of *V. cholerae* O1, classic and El Tor, have been described, but the former is thought to be extinct.

The natural habitat of *V. cholerae* is coastal salt water and brackish estuaries, where the organism lives in close relation to plankton. *V. cholerae* can also exist in freshwater in the presence of adequate nutrients and warmth. Humans become infected incidentally but, once infected, can act as vehicles for spread. Ingestion of water contaminated by human feces is thought to be the most common means of acquisition of *V. cholerae*. However, consumption of contaminated food and human-to-human transmission also contribute to spread. There is no known animal reservoir. Although the infectious dose is relatively high, it is markedly reduced in hypochlorhydric persons, in those using antacids, and when gastric acidity is buffered by a meal. Cholera is predominantly a pediatric disease in endemic areas, but it affects adults and children equally when newly introduced into a population. In endemic areas, the burden of disease is often greatest during "cholera seasons" associated with high temperatures, heavy rainfall, and flooding, but cholera can occur year-round.

Cholera is native to the Ganges delta on the Indian subcontinent. Since 1817, seven global pandemics have occurred. The current (seventh) pandemic—the first due to the El Tor biotype—began in Indonesia in 1961 and spread in serial waves throughout Asia as *V. cholerae* El Tor displaced the endemic classic biotype, which is thought to have caused the previous six pandemics. In the early 1970s, El Tor cholera erupted in Africa, causing major epidemics before becoming a persistent endemic problem. Currently, >95% of cholera cases reported annually to the World Health Organization (WHO) are from Africa and Asia **(Fig. 173-1)**, but the true burden and distribution of cholera are unknown because the diagnosis is often syndromic and many countries with endemic cholera do not report cholera to the WHO. In 2022 and 2023, large outbreaks involving more than 10,000 cases per country were being reported every few weeks to the World Health Organization. It is possible that >1–4 million cases of cholera occur yearly (of which only ~400,000 are reported to the WHO) and that these cases result in >20,000–140,000 deaths annually (of which <2000 are reported to the WHO).

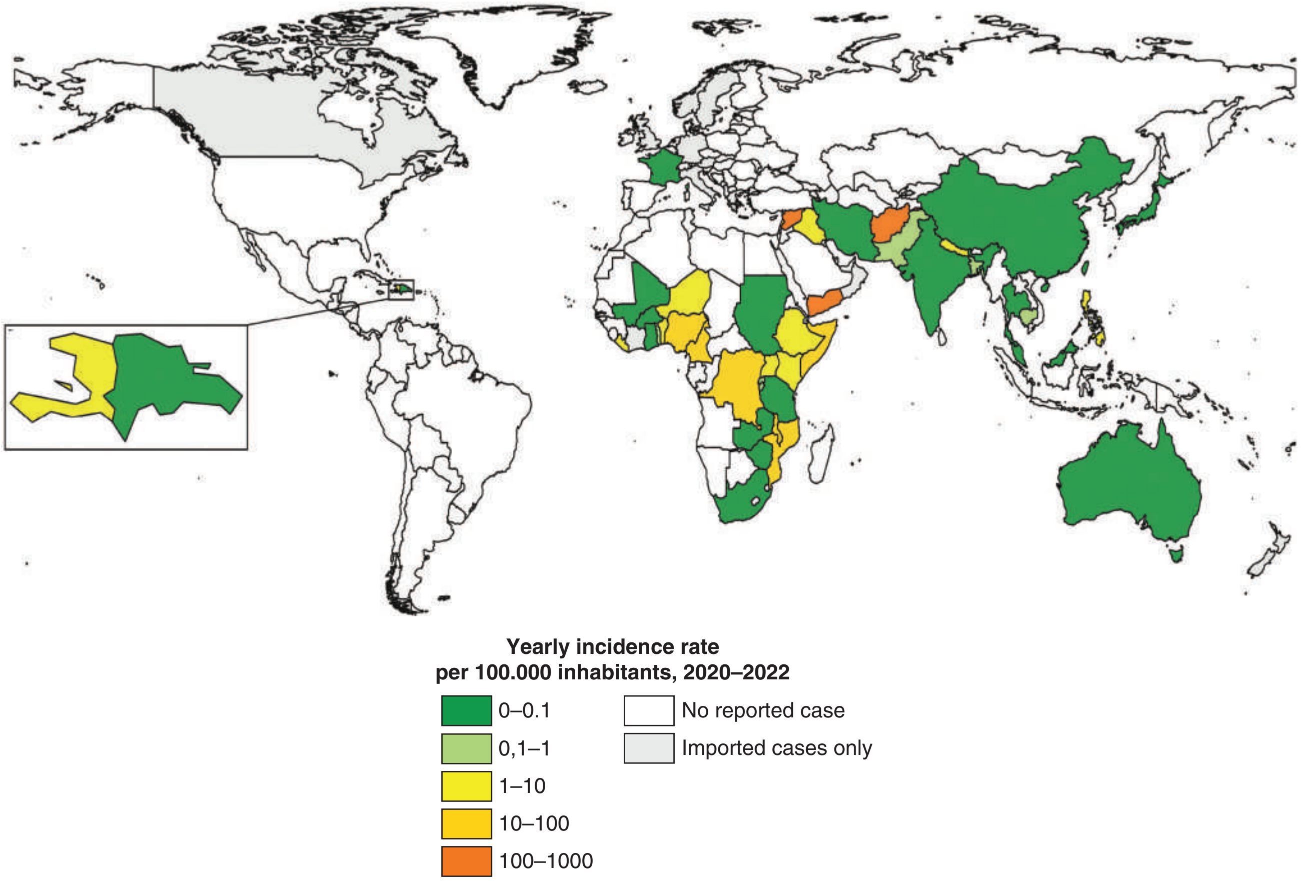

FIGURE 173-1 World distribution of cholera in 2020–2022. WHO, World Health Organization. *(Reproduced with permission from Dr. M. Piarroux, Université de la Méditerranée, France.)*

After a century without cholera in Latin America, the current cholera pandemic reached Central and South America in 1991. Following an initial explosive spread that affected millions, the burden of disease has markedly decreased in Latin America. In 2010, a severe cholera outbreak began in Haiti, a country with no recorded history of this disease. Several lines of evidence indicate that cholera was likely introduced into Haiti by United Nations security forces from Asia, raising the possibility that asymptomatic carriers of *V. cholerae* play an important role in transmitting cholera over long distances. In 2016, an outbreak of cholera began in Yemen in the setting of a civil war and population displacement and the breakdown of health infrastructure. The recent history of cholera has been punctuated by such severe outbreaks, especially among impoverished or displaced persons. These outbreaks are often precipitated by war or other circumstances that lead to the breakdown of public-health measures. Such was the case in the camps for Rwandan refugees set up in 1994 around Goma, Zaire; in 2008–2009 in Zimbabwe; in 2015 in South Sudan and the Democratic Republic of the Congo; and in 2022 in Lebanon and the Syrian Arab Republic.

Sporadic endemic infections due to *V. cholerae* O1 strains related to the seventh-pandemic strain have been recognized along the U.S. Gulf Coast of Louisiana and Texas. These infections are typically associated with the consumption of contaminated, locally harvested shellfish. Occasionally, cases in U.S. locations remote from the Gulf Coast have been linked to shipped-in Gulf Coast seafood.

In October 1992, a large-scale outbreak of clinical cholera caused by a new serogroup, O139, occurred in southeastern India. The organism appears to be a derivative of El Tor O1 but has a distinct LPS and an immunologically related O-antigen polysaccharide capsule. (O1 organisms are not encapsulated.) After an initial spread across 11 Asian countries, *V. cholerae* O139 has once again been almost entirely replaced by O1 strains. The clinical manifestations of disease caused by *V. cholerae* O139 are indistinguishable from those of O1 cholera. Immunity to one, however, is not protective against the other.

PATHOGENESIS

In the final analysis, cholera is a toxin-mediated disease. The watery diarrhea characteristic of cholera is due to the action of cholera toxin, a potent protein enterotoxin elaborated by the organism in the small intestine. The toxin-coregulated pilus (TCP), so named because its synthesis is regulated in parallel with that of cholera toxin, is essential for *V. cholerae* to survive and multiply in (colonize) the small intestine. Production of cholera toxin, TCP, and several other virulence factors are coordinately regulated by ToxR. This protein modulates the expression of genes coding for virulence factors in response to environmental signals via a cascade of regulatory proteins. Additional regulatory processes, including bacterial responses to the density of the bacterial population (in a phenomenon known as *quorum sensing*), modulate the virulence of *V. cholerae*.

Once established in the human small bowel, the organism produces cholera toxin, which consists of a monomeric enzymatic moiety (the A subunit) and a pentameric binding moiety (the B subunit). The B pentamer binds to GM_1 ganglioside, a glycolipid on the surface of epithelial cells that serves as the toxin receptor and makes possible the delivery of the A subunit to its cytosolic target. The activated A subunit (A_1) irreversibly transfers ADP-ribose from nicotinamide adenine dinucleotide to its specific target protein, the GTP-binding regulatory component of adenylate cyclase. The ADP-ribosylated G protein upregulates the activity of adenylate cyclase; the result is the intracellular accumulation of high levels of cyclic adenosine monophosphate (AMP). In intestinal epithelial cells, cyclic AMP inhibits the absorptive sodium-transport system in villus cells and activates the secretory chloride-transport system in crypt cells, and these events lead to the accumulation of sodium chloride in the intestinal lumen. Because water moves passively to maintain osmolality, isotonic fluid accumulates in the lumen. When the volume of that fluid exceeds the capacity of the rest of the gut to resorb it, watery diarrhea results. Unless the wasted fluid and electrolytes are adequately replaced, shock (due to profound dehydration) and acidosis (due to loss of bicarbonate) follow. Although perturbation of the adenylate cyclase pathway is the primary mechanism by which cholera toxin causes excess fluid secretion, cholera toxin also enhances intestinal secretion via prostaglandins and/or neural histamine receptors.

The *V. cholerae* genome is composed of two circular chromosomes. Lateral gene transfer has played a key role in the evolution of epidemic *V. cholerae*. The genes encoding cholera toxin (*ctxAB*) are part of the genome of a bacteriophage, CTXΦ. The receptor for this phage on the *V. cholerae* surface is the intestinal colonization factor TCP. Because *ctxAB* is part of a mobile genetic element (CTXΦ), horizontal transfer of this bacteriophage may account for the emergence of new toxigenic *V. cholerae* strains. Many of the other genes important for *V. cholerae* pathogenicity, including the genes encoding the biosynthesis of TCP, accessory colonization factors, and virulence regulators, are clustered together in the *V. cholerae* pathogenicity island. Similar clustering of virulence genes is found in other bacterial pathogens. It is believed that pathogenicity islands are acquired by horizontal gene transfer. *V. cholerae* O139 is probably derived from an El Tor O1 strain that acquired the genes for O139 O-antigen synthesis by horizontal gene transfer.

CLINICAL MANIFESTATIONS

Individuals infected with *V. cholerae* O1 or O139 exhibit a range of clinical manifestations. Some individuals are asymptomatic or have only mild diarrhea; others present with the sudden onset of explosive and life-threatening diarrhea (*cholera gravis*). The reasons for the range in signs and symptoms of disease are incompletely understood but include the level of preexisting immunity, blood type (persons with type O blood are at greatest risk of severe disease if infected, whereas those with type AB are at least risk), and nutritional status. In a nonimmune individual, after a 24- to 48-h incubation period, cholera characteristically begins with the sudden onset of painless watery diarrhea that may quickly become voluminous. Patients often vomit. In severe cases, volume loss can exceed 250 mL/kg in the first 24 h. If fluids and electrolytes are not replaced, hypovolemic shock and death may ensue. Fever is usually absent. Muscle cramps due to electrolyte disturbances are common. The stool has a characteristic appearance: a nonbilious, gray, slightly cloudy fluid with flecks of mucus, no blood, and a somewhat fishy, inoffensive odor. It has been called "rice-water" stool because of its resemblance to the water in which rice has been washed **(Fig. 173-2)**. Clinical symptoms parallel volume contraction:

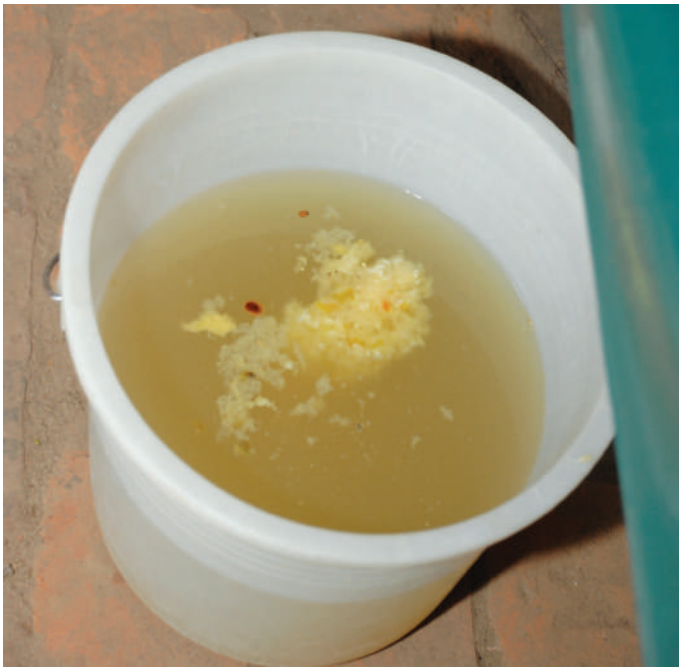

FIGURE 173-2 Rice-water cholera stool. Note floating mucus and gray watery appearance. *(Courtesy of Dr. A. S. G. Faruque, International Centre for Diarrhoeal Disease Research, Dhaka; with permission.)*

at losses of <5% of normal body weight, thirst develops; at 5–10%, postural hypotension, weakness, tachycardia, and decreased skin turgor are documented; and at >10%, oliguria, weak or absent pulses, sunken eyes (and, in infants, sunken fontanelles), wrinkled ("washerwoman") skin, somnolence, and coma are characteristic. Complications derive exclusively from the effects of volume and electrolyte depletion and include renal failure due to acute tubular necrosis. Thus, if the patient is adequately treated with fluid and electrolytes, complications are averted and the process is self-limited, resolving in a few days.

Laboratory data usually reveal an elevated hematocrit (due to hemoconcentration) in nonanemic patients; mild neutrophilic leukocytosis; elevated levels of blood urea nitrogen and creatinine consistent with prerenal azotemia; normal sodium, potassium, and chloride levels; a markedly reduced bicarbonate level (<15 mmol/L); and an elevated anion gap (due to increases in serum lactate, protein, and phosphate). Arterial pH is usually low (~7.2).

DIAGNOSIS

Cholera should be suspected when a patient ≥5 years of age develops acute watery diarrhea in an area known to have cholera or develops severe dehydration or dies from acute watery diarrhea, even in an area where cholera is not known to be present. The clinical suspicion of cholera can be confirmed by the identification of *V. cholerae* in stool; however, the organism must be specifically sought. With experience, it can be detected directly by dark-field microscopy on a wet mount of fresh stool, and its serotype can be discerned by immobilization with specific antisera. Laboratory isolation of the organism requires the use of a selective medium such as taurocholate–tellurite–gelatin (TTG) agar or thiosulfate–citrate–bile salts–sucrose (TCBS) agar. If a delay in sample processing is expected, Cary-Blair transport medium and/or alkaline-peptone water-enrichment medium may be used as well. In endemic areas, there is little need for biochemical confirmation and characterization, although these tasks may be worthwhile in places where *V. cholerae* is an uncommon isolate. Standard microbiologic biochemical testing for Enterobacteriaceae will suffice for identification of *V. cholerae*. All vibrios are oxidase-positive. Point-of-care antigen-detection cholera dipstick assays are commercially available for use in the field or where laboratory facilities are lacking.

TREATMENT

Cholera

Death from cholera is due to hypovolemic shock; thus, treatment of individuals with cholera first and foremost requires fluid resuscitation and management. In light of the level of dehydration **(Table 173-1)** and the patient's age and weight, euvolemia should first be rapidly restored, and adequate hydration should then be maintained to replace ongoing fluid losses **(Table 173-2)**. Administration of oral rehydration solution (ORS) takes advantage of the hexose-Na^+ co-transport mechanism to move Na^+ across the gut mucosa together with an actively transported molecule such as glucose (or galactose); Cl^- and water follow. This transport mechanism remains intact even when cholera toxin is active. ORS may be made by adding safe water to prepackaged sachets containing salts and sugar or by adding 0.5 teaspoon (i.e., half a small spoonful) of table salt and 6 level teaspoons (i.e., 6 small spoonfuls) of table sugar to 1 L of safe water. Potassium intake in bananas or green coconut water should be encouraged. A number of ORS formulations are available, and the WHO now recommends "low-osmolarity" ORS for treatment of individuals with dehydrating diarrhea of any cause **(Table 173-3)**. If available, rice-based ORS is considered superior to standard ORS in the treatment of cholera. ORS can be administered via a nasogastric tube to individuals who cannot ingest fluid; however, optimal management of individuals with severe dehydration includes the administration of IV fluid and electrolytes. Because profound acidosis (pH <7.2) is common in this group, Ringer's lactate is the best choice among commercial products **(Table 173-4)**; it must be used with additional potassium supplements, preferably given by mouth. The total fluid deficit in severely dehydrated patients (>10% of body weight) can be replaced safely within the first 3–4 h of therapy, half within the first hour. Transient muscle cramps and tetany are common. Thereafter, oral therapy can usually

TABLE 173-1 Assessing the Degree of Dehydration in Patients with Cholera

DEGREE OF DEHYDRATION	CLINICAL FINDINGS
None or mild, but diarrhea	Thirst in some cases; <5% loss of total body weight
Moderate	Thirst, postural hypotension, weakness, tachycardia, decreased skin turgor, dry mouth/tongue, no tears; 5–10% loss of total body weight
Severe	Unconsciousness, lethargy, or "floppiness"; weak or absent pulse; inability to drink; sunken eyes (and, in infants, sunken fontanelles); >10% loss of total body weight

TABLE 173-2 Treatment of Cholera, Based on Degree of Dehydration[a]

DEGREE OF DEHYDRATION, PATIENT'S AGE (WEIGHT)	TREATMENT[b]
None or Mild, but Diarrhea[c]	
<2 years	1/4–1/2 cup (50–100 mL) of ORS, to a maximum of 0.5 L/d
2–9 years	1/2–1 cup (100–200 mL) of ORS, to a maximum of 1 L/d
≥10 years	As much ORS as desired, to a maximum of 2 L/d
Moderate[c,d]	
<4 months (<5 kg)	200–400 mL of ORS
4–11 months (5–<8 kg)	400–600 mL of ORS
12–23 months (8–<11 kg)	600–800 mL of ORS
2–4 years (11–<16 kg)	800–1200 mL of ORS
5–14 years (16–<30 kg)	1200–2200 mL of ORS
≥15 years (≥30 kg)	2200–4000 mL of ORS
Severe[c]	
All ages and weights	Undertake IV fluid replacement with Ringer's lactate (or, if not available, normal saline). Give 100 mL/kg in the first 3-h period (or the first 6-h period for children <12 months old); start rapidly, then slow down. Give a total of 200 mL/kg in the first 24 h. Continue until the patient is awake, can ingest ORS, and no longer has a weak pulse.

[a]Adapted from World Health Organization: Outbreak Response Field Manual, Global Task on Cholera Control; *https://www.gtfcc.org/wp-content/uploads/2020/05/gtfcc-cholera-outbreak-response-field-manual-2024.pdf*. [b]Continue normal feeding during treatment. [c]Reassess regularly; monitor stool and vomit output. [d]Volumes of ORS listed should be given within the first 4 h.

Abbreviation: ORS, oral rehydration solution.

TABLE 173-3 Composition of World Health Organization Reduced-Osmolarity Oral Rehydration Solution (ORS)[a,b]

CONSTITUENT	CONCENTRATION, mmol/L
Na^+	75
K^+	20
Cl^-	65
Citrate[c]	10
Glucose	75
Total osmolarity	245

[a]Contains (per package, to be added to 1 L of drinking water): NaCl, 2.6 g; $Na_3C_6H_5O_7 \cdot 2H_2O$, 2.9 g; KCl, 1.5 g; and glucose (anhydrous), 13.5 g. [b]If prepackaged ORS is unavailable, a simple homemade alternative can be prepared by combining 3.5 g (~1/2 teaspoon) of NaCl with either 50 g of precooked rice cereal or 6 teaspoons of table sugar (sucrose) in 1 L of drinking water. In that case, potassium must be supplied separately (e.g., in orange juice or coconut water). [c]10 mmol of citrate per liter, which supplies 30 mmol of HCO_3/L.

TABLE 173-4 Electrolyte Composition of Cholera Stool and of Intravenous Rehydration Solution

	CONCENTRATION, mmol/L			
SUBSTANCE	**NA$^+$**	**K$^+$**	**CL$^-$**	**BASE**
Stool				
Adult	135	15	100	45
Child	100	25	90	30
Ringer's lactate	130	4[a]	109	28

[a] Potassium supplements, preferably administered by mouth, are required to replace the usual potassium losses from stool.

be initiated, with the goal of maintaining fluid intake equal to fluid output. However, patients with continued large-volume diarrhea may require prolonged IV treatment to match gastrointestinal fluid losses. Severe hypokalemia can develop but will respond to potassium given either IV or orally. In the absence of adequate staff to monitor the patient's progress, the oral route of rehydration and potassium replacement is safer than the IV route.

Although not necessary for cure, the use of an antibiotic to which the organism is susceptible diminishes the duration and volume of fluid loss and hastens clearance of the organism from the stool. Adjunctive antibiotics should therefore be administered to patients with moderate or severe dehydration due to cholera. In many areas, macrolides such as erythromycin (adults, 250 mg orally four times a day for 3 days; children, 12.5 mg/kg per dose four times a day for 3 days) or azithromycin (adults, a single 1-g dose; children, a single 20-mg/kg dose) are the agents of choice. Increasing resistance to tetracyclines is widespread; however, in areas with confirmed susceptibility, tetracycline (nonpregnant adults, 500 mg orally four times a day for 3 days; children >8 years old, 12.5 mg/kg per dose four times a day for 3 days) or doxycycline (nonpregnant adults, a 300-mg single dose; children >8 years old, a single dose of 4–6 mg/kg) may be used. Similarly, increasing resistance to fluoroquinolones is being reported, but in areas with confirmed susceptibility, a fluoroquinolone such as ciprofloxacin may be used (adults, 500 mg twice a day for 3 days; children, 15 mg/kg twice a day for 3 days). Oral administration of supplemental zinc is associated with decreased volume and severity of diarrhea in young children, including in those with cholera. Children <6 months of age with cholera should be treated with 10 mg of zinc daily for 10 days; children from 6 to <60 months of age should be treated with 20 mg of oral zinc daily for 10 days.

■ PREVENTION

Provision of safe water and of facilities for sanitary disposal of feces, improved nutrition, and attention to food preparation and storage in the household can significantly reduce the incidence of cholera. In addition, precautions should be taken to prevent the spread of cholera via infected and potentially asymptomatic persons from endemic to nonendemic regions of the world (as was probably the case in the outbreak in Haiti; see "Microbiology and Epidemiology," above).

Much effort has been devoted to the development of an effective cholera vaccine over the past few decades, with a particular focus on oral vaccine strains. In an attempt to maximize mucosal responses, two types of oral cholera vaccine have been developed: oral killed vaccines and live attenuated vaccines. Currently, three oral killed cholera vaccines have been prequalified by the WHO. BivWC (Shanchol™; Shantha Biotechnics, Hyderabad, India) contains both biotypes and serotypes of *V. cholerae* O1 and *V. cholerae* O139 without supplemental cholera toxin B subunit, but its manufacture is being terminated. A related vaccine is produced in South Korea (Euvichol™, Euvichol-Plus™; Eubiologics, Seoul) and currently accounts for the vast majority of cholera vaccine available to global health control programs. WC-rBS (Dukoral®; Valneva, Sweden AB, Stockholm,) contains both biotypes and serotypes of *V. cholerae* O1 supplemented with 1 mg of recombinant cholera toxin B subunit per dose. The vaccines are administered as a two- or three-dose regimen, with doses usually separated by 14 days. They provide ~60–85% protection for the first few months. Booster immunizations of WC-rBS are recommended after 2 years for individuals ≥6 years of age and after 6 months for children 2–5 years of age. For BivWC, no formal recommendation regarding booster immunizations exists. However, BivWC was associated with ~60% protection over 5 years among recipients of all ages in a study in Kolkata, India; the rate of protection among children ≤5 years of age approximated 40%. In outbreak situations, even a single dose of BivWC can provide some protection: 40% and 63% adjusted protection for 6 months for all and severely dehydrating cholera, respectively—although there was no evidence of protection in children younger than 5 years of age. Models predict significant herd immunity when vaccination coverage rates exceed 50%. The killed vaccines have been safely administered among populations with high rates of HIV infection.

Oral live attenuated vaccines for *V. cholerae* O1 also are in development. These strains have in common their lack of the genes encoding cholera toxin. One such vaccine, CVD 103-HgR (Vaxchora™; Emergent BioSolutions, Washington, DC), is derived from a classic strain of *V. cholerae* and has been approved by the U.S. Food and Drug Administration for use in travelers to cholera-endemic regions but is not prequalified by the WHO. The vaccine was 90% and 80% efficacious against severe cholera after experimental infection of North American volunteers 10 days and 90 days after vaccination, respectively. Vaxchora is approved for use in individuals 2–64 years of age; no recommendations concerning the timing or need for booster vaccinations are currently available. Other live attenuated vaccine candidate strains have been prepared from El Tor and O139 *V. cholerae* and have been tested in studies of volunteers. An advantage of live attenuated cholera vaccines is that they may induce more potent protection after a single oral dose. Conjugate and subunit cholera vaccines also are being developed.

Recognizing that it may be decades before safe water and adequate sanitation become a reality for those most at risk of cholera, the WHO has recommended incorporation of cholera vaccination into comprehensive control strategies and has established an international stockpile of oral killed cholera vaccine to assist in outbreak responses. A global strategy on cholera control was launched in 2017 with the WHO hosting the secretariat of the Global Task Force on Cholera Control (GTFCC). This country-by-country approach aims to reduce cholera deaths by 90% and to eliminate cholera in as many as 20 countries by 2030. Integral components of this strategy are advancing water, sanitation, and hygiene (WASH) programs, as well as use of cholera vaccine. From 2016 to 2023, almost 300 million doses of cholera vaccine have been requested from the Global Vaccine Stockpile, and approximately 150 million doses have been shipped to requesting countries for use in control programs. Due to vaccine shortages, cholera vaccines have most recently been almost solely used in reactive and not preventative campaigns, and are often given as single doses.

OTHER *VIBRIO* SPECIES

The genus *Vibrio* includes several human pathogens that do not cause cholera. Abundant in coastal waters throughout the world, noncholera vibrios can reach high concentrations in the tissues of filter-feeding mollusks. As a result, human infection commonly follows the ingestion of seawater or of raw or undercooked shellfish **(Table 173-5)**. Most noncholera vibrios can be cultured on blood or MacConkey agar, which contains enough salt to support the growth of these halophilic species. In the microbiology laboratory, the species of noncholera vibrios are distinguished by standard biochemical tests. The most important of these organisms are *Vibrio parahaemolyticus* and *Vibrio vulnificus*. Vibriosis causes an estimated 80,000 illnesses and 100 deaths in the United States every year.

The two major types of syndromes for which these noncholera vibrios are responsible are gastrointestinal illness (due to *V. parahaemolyticus*, non-O1/O139 *V. cholerae*, *V. mimicus*, *V. fluvialis*, *V. hollisae*, and *V. furnissii*) and soft tissue infections (due to *V. vulnificus*, *V. alginolyticus*, and *V. damselae*). *V. vulnificus* is also a cause of primary sepsis in some immunocompromised individuals.

TABLE 173-5 Features of Selected Noncholera Vibrioses

ORGANISM	VEHICLE OR ACTIVITY	HOST AT RISK	SYNDROME
Vibrio parahaemolyticus	Shellfish, seawater	Normal	Gastroenteritis
	Seawater	Normal	Wound infection
Non-01/0139 *Vibrio cholerae*	Shellfish, travel	Normal	Gastroenteritis
	Seawater	Normal	Wound infection, otitis media
Vibrio vulnificus	Shellfish	Immunosuppressed[a]	Sepsis, secondary cellulitis
	Seawater	Normal, immunosuppressed[a]	Wound infection, cellulitis
Vibrio alginolyticus	Seawater	Normal	Wound infection, cellulitis, otitis
	Seawater	Burned, other immunosuppressed	Sepsis

[a] Especially with liver disease or hemochromatosis.

Source: Table 161-3 in *Harrison's Principles of Internal Medicine*, 14th edition.

SPECIES ASSOCIATED PRIMARILY WITH GASTROINTESTINAL ILLNESS

V. parahaemolyticus Widespread in marine environments, the halophilic *V. parahaemolyticus* is the leading seafood-borne bacterial cause of enteritis worldwide. This species was originally implicated in enteritis in Japan in 1953, accounting for 24% of reported cases in one study—a rate that presumably was due to the common practice of eating raw seafood in that country. In the United States, common-source outbreaks of diarrhea caused by this organism have been linked to the consumption of undercooked or improperly handled seafood or of other foods contaminated by seawater. Since the mid-1990s, the incidence of *V. parahaemolyticus* infections has increased in several countries, including the United States. Serotypes O3:K6, O4:K68, and O1:K-untypable, which are genetically related to one another, account in part for this increase. Recent reports from China and Thailand suggest that serotype O10:K4 may be an emerging serotype. The enteropathogenicity of *V. parahaemolyticus* is associated with its ability to cause hemolysis via a thermostable direct hemolysin (Vp-TDH). Although the mechanisms by which the organism causes diarrhea are not fully defined, most *V. parahaemolyticus* genomes encode two type III secretion systems, which directly inject toxic bacterial proteins into host cells. The activity of one of these secretion systems is required for intestinal colonization and virulence in animal models. *V. parahaemolyticus* should be considered a possible etiologic agent in all cases of diarrhea that can be linked epidemiologically to seafood consumption or to the sea itself. The incidence of *V. parahaemolyticus* infection in the United States has increased from 0.06 per 100,000 persons in 1996 to 0.9 cases per 100,000 in 2019.

Infections with *V. parahaemolyticus* can result in two distinct gastrointestinal presentations. The more common of the two presentations (including nearly all cases in North America) is characterized by watery diarrhea, usually occurring in conjunction with abdominal cramps, nausea, and vomiting and accompanied in ~25% of cases by fever and chills. After an incubation period of 4 h to 4 days, symptoms develop and persist for a median of 3 days. Dysentery, the less common presentation, is characterized by severe abdominal cramps, nausea, vomiting, and bloody or mucoid stools. *V. parahaemolyticus* also causes rare cases of wound infection and otitis and very rare cases of sepsis.

Most cases of *V. parahaemolyticus*–associated gastrointestinal illness, regardless of the presentation, are self-limited. Fluid replacement should be stressed. Antimicrobial agents may be of benefit in moderate or severe disease. Doxycycline, fluoroquinolones, macrolides, or third-generation cephalosporins are usually used. Deaths are extremely rare among immunocompetent individuals. Severe infections are associated with underlying diseases, including diabetes, preexisting liver disease, iron-overload states, or immunosuppression.

Non-O1/O139 (Noncholera) *V. cholerae* The heterogeneous non-O1/O139 *V. cholerae* organisms cannot be distinguished from *V. cholerae* O1 or O139 by routine biochemical tests but do not agglutinate in O1 or O139 antiserum. Non-O1/O139 strains have caused several well-studied food-borne outbreaks of gastroenteritis and have also been responsible for sporadic cases of otitis media, wound infection, and bacteremia. Generally, non-O1/O139 *V. cholerae* strains do not produce cholera toxin and do not cause large epidemics of diarrheal disease. Like other vibrios, non-O1/O139 *V. cholerae* organisms are widely distributed in marine environments. In most instances, recognized cases in the United States have been associated with the consumption of raw oysters or with recent travel. The broad clinical spectrum of diarrheal illness caused by these organisms is probably due to the group's heterogeneous virulence attributes.

The typical incubation period for gastroenteritis due to these organisms is <2 days, and the illness lasts for ~2–7 days. Patients' stools may be copious and watery or may be partly formed, less voluminous, and bloody or mucoid. Diarrhea can result in severe dehydration. Many cases include abdominal cramps, nausea, vomiting, and fever. Like those with cholera, patients who are seriously dehydrated should receive oral or IV fluids; the value of antibiotics is not clear.

Extraintestinal infections due to non-O1/O139 *V. cholerae* commonly follow occupational or recreational exposure to seawater. Around 10% of non-O1/O139 *V. cholerae* isolates come from cases of wound infection, 10% from cases of otitis media, 20% from cases of bacteremia (which is particularly likely to develop in patients with liver disease), and approximately 40% from stool. Extraintestinal infections should be treated with antibiotics. Information to guide antibiotic selection and dosing is limited, but most strains are sensitive in vitro to tetracycline, ciprofloxacin, and third-generation cephalosporins.

SPECIES ASSOCIATED PRIMARILY WITH SOFT TISSUE INFECTION OR BACTEREMIA

(See also Chap. 134)

V. vulnificus Infection with *V. vulnificus* is rare, but this organism is the most common cause of severe *Vibrio* infections in the United States. Like most vibrios, *V. vulnificus* proliferates in the warm summer months and requires a saline environment for growth. In the United States, infections in humans typically occur in coastal states between May and October and most commonly affect men >40 years of age. *V. vulnificus* has been linked to two distinct syndromes: primary sepsis, which usually occurs in patients with underlying liver disease, and primary wound infection, which generally affects people without underlying disease. (*Vulnificus* is Latin for "wound maker.") Some authors have suggested that *V. vulnificus* also causes gastroenteritis independent of other clinical manifestations. *V. vulnificus* is endowed with a number of virulence attributes, including a capsule that confers resistance to phagocytosis and to the bactericidal activity of human serum as well as a cytolysin. Measured as the 50% lethal dose in mice, the organism's virulence is considerably increased under conditions of iron overload; this observation is consistent with the propensity of *V. vulnificus* to infect patients who have hemochromatosis.

Primary sepsis most often develops in patients who have cirrhosis or hemochromatosis. However, *V. vulnificus* bacteremia can also affect individuals who have hematopoietic disorders or chronic renal insufficiency, those who are using immunosuppressive medications or alcohol, or (in rare instances) those who have no known underlying disease. After a median incubation period of 16 h, the patient develops malaise, chills, fever, and prostration. One-third of patients develop hypotension, which is often apparent at admission. Cutaneous manifestations develop in most cases (usually within 36 h of onset) and characteristically involve the extremities (the lower more often than the upper). In a common sequence, erythematous patches are followed by ecchymoses, vesicles, and bullae. In fact, sepsis and hemorrhagic bullous skin lesions suggest the diagnosis in appropriate settings. Necrosis and sloughing also may be evident. Laboratory studies reveal leukopenia more often than leukocytosis, thrombocytopenia, or elevated levels of fibrin-split products. *V. vulnificus* can be cultured from blood or cutaneous lesions. The mortality rate approaches 50%, with most deaths due to uncontrolled sepsis **(Chap. 315)**. Accordingly, prompt treatment is critical and should include empirical antibiotic administration, aggressive debridement, and general supportive care. *V. vulnificus* is sensitive in vitro to a number of antibiotics, including tetracycline, fluoroquinolones, and third-generation cephalosporins. Optimal treatment usually involves administration of a third-generation cephalosporin such as ceftazidime with doxycycline or a fluoroquinolone and wound debridement.

V. vulnificus–associated soft tissue infection can complicate either a fresh or an old wound that comes into contact with seawater; the patient may or may not have underlying disease. After a short incubation period (4 h to 4 days; mean, 12 h), the disease begins with swelling, erythema, and (in many cases) intense pain around the wound. These signs and symptoms are followed by cellulitis, which spreads rapidly and is sometimes accompanied by vesicular, bullous, or necrotic lesions. Metastatic events are uncommon. Most patients have fever and leukocytosis. *V. vulnificus* can be cultured from skin lesions and occasionally from the blood. Prompt antibiotic therapy and debridement are usually curative.

V. alginolyticus First identified as a pathogen of humans in 1973, *V. alginolyticus* occasionally causes eye, ear, and wound infections. This species is the most salt-tolerant of the vibrios and can grow in salt concentrations of >10%. Most clinical isolates come from superinfected wounds that presumably become contaminated at the beach. Although its severity varies, *V. alginolyticus* infection tends not to be serious and generally responds well to antibiotic therapy and drainage. Cases of otitis externa, otitis media, and conjunctivitis due to this pathogen have been described. Tetracycline treatment usually results in cure. *V. alginolyticus* is a rare cause of bacteremia in immunocompromised hosts.

FURTHER READING

Brumfield KD et al: Environmental parameters associated with incidence and transmission of pathogenic *Vibrio* spp. Environ Microbiol 23:7314; 2021.

Collins JP et al: Preliminary incidence and trends of infections with pathogens transmitted commonly through food—Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2016–2021. MMWR Morb Mortal Wkly Rep 71:1260, 2022.

Domman D et al: Integrated view of *Vibrio cholerae* in the Americas. Science 358:789, 2017.

Global Task Force on Cholera Control: Early detection and response to contain outbreaks. *www.gtfcc.org*. Accessed December 6, 2023.

Huang Y et al: New variant of *Vibrio parahaemolyticus*, sequence type 3, serotype O10:K4, China, 2020. Emerg Infect Dis 28:1261; 2022.

Kayembe HC et al: Drivers of the dynamics of the spread of cholera in the Democratic Republic of the Congo, 2000–2018: An eco-epidemiological study. PLoS Negl Trop Dis 17:e0011597, 2023.

Mavian CN et al: Ancestral origin and dissemination dynamics of reemerging toxigenic *Vibrio cholerae*, Haiti. Emerg Infect Dis 29:2073, 2023.

Weill FX et al: Genomic history of the seventh pandemic of cholera in Africa. Science 358:785, 2017.

World Health Organization: Cholera vaccines: WHO position paper. Wkly Epidemiol Rec 92:477, 2017.

World Health Organization: Cholera, 2022. Wkly Epidemiol Rec 38:432, 2023.

World Health Organization: Outbreak response field manual, Global Task Force on Cholera Control. *https://www.gtfcc.org/wp-content/uploads/2020/05/gtfcc-cholera-outbreak-response-field-manual-2024.pdf*. Accessed March 21, 2024.

174

Brucellosis

Nicholas J. Beeching

DEFINITION

Brucellosis is a bacterial zoonosis transmitted directly or indirectly to humans from infected animals, predominantly domesticated ruminants and swine. The disease is known colloquially as *undulant fever* because of its remittent character. Although brucellosis commonly presents as an acute febrile illness, its clinical manifestations vary widely, and definitive signs indicative of the diagnosis may be lacking. Thus, the clinical diagnosis usually must be supported by the results of bacteriologic and/or serologic tests.

ETIOLOGIC AGENTS

Human brucellosis is caused by strains of the genus *Brucella*, with several species groups defined by differences in chromosomal structure, host preference, and epidemiologic patterns of infection. *Brucella melitensis* is the most common cause of symptomatic disease in humans, for which the main sources are sheep, goats, and camels; *B. abortus* is usually acquired from cattle or buffalo; *B. suis* is generally acquired from swine but has one variant that is enzootic in reindeer and caribou and another in rodents; *B. canis* is acquired from dogs. *B. ovis* causes reproductive disease in sheep but has not been clearly implicated in human disease. Rare human infections have been reported with *B. neotomae*, which is found in desert rodents. Two species, *B. ceti* and *B. pinnipedialis*, have been identified in marine mammals, including seals and dolphins. At least one case of laboratory-acquired human disease due to one of these species has been described, and several cases of natural human infection have been reported. As infections in marine mammals appear to be widespread, more cases of zoonotic infection in humans may be identified. Other reported species include *B. microti* (isolated from field voles), *B. papionis* (from baboons), *B. vulpis* (from foxes), and *B. inopinata* (from a patient with a breast implant). Additional novel strains have been described in diverse species, including frogs, bats, and various rodents, and the genus likely will expand further in forthcoming years. The genus *Brucella* is closely related to the genus *Ochrobactrum*, which includes free-living environmental bacteria that can occasionally cause opportunistic infections. Recent changes in taxonomy now place *Ochrobactrum* spp. as subspecies of *Brucella* on genetic grounds, although the ecology, physiology, clinical niche, and antimicrobial sensitivity of these organisms are completely different.

All brucellae are small, gram-negative, unencapsulated, nonsporulating rods or coccobacilli. They grow aerobically on peptone-based medium incubated at 37°C; the growth of some types is improved by supplementary CO_2. In vivo, brucellae behave as facultative intracellular parasites. The organisms are sensitive to sunlight, ionizing radiation, and moderate heat; they are killed by boiling and pasteurization but are resistant to freezing and drying. Their resistance to drying renders brucellae stable in aerosol form, facilitating airborne transmission. The organisms can survive for up to 2 months in soft cheeses made from goat's or sheep's milk; for at least 6 weeks in dry soil contaminated with infected urine, vaginal discharge, or placental or fetal tissues; and for at least 6 months in damp soil or liquid manure kept in cool dark conditions. Brucellae are easily killed by a wide range of common disinfectants used under optimal conditions but are likely to be much more resistant at low temperatures or in the presence of heavy organic contamination.

EPIDEMIOLOGY

Brucellosis is a zoonosis whose occurrence and control are closely related to its prevalence in domesticated animals. Its distribution is worldwide apart from the few countries where it has been eradicated from the animal reservoir. The true global prevalence of human brucellosis is unknown because of the imprecision of diagnosis and the inadequacy of reporting and surveillance systems in many countries.

Recent estimates suggest there may be more than 2 million cases of human infection a year worldwide. There has been increased recognition of brucellosis in India, Pakistan, Sri Lanka, and China, and of importations to countries in Oceania, such as Fiji, and in Asia, such as Thailand and Vietnam. In Europe, the incidence of brucellosis in a country is inversely related to gross domestic product, and in both developed and less well-resourced settings, human brucellosis is related to rural poverty and inadequate access to medical care. Failure of veterinary control programs due to conflicts or for economic reasons contributes further to the emergence and re-emergence of disease, as seen currently in some eastern Mediterranean countries.

Even in well-resourced settings, the true incidence of brucellosis in domesticated animals may be 10–20 times higher than the reported figures. Bovine brucellosis has been the target of control programs in many parts of the world and has been eradicated from the cattle populations of much of northern Europe, Australia, New Zealand, and Canada, among other nations. Its incidence has been reduced to a low level in the United States and most western European countries, with a varied picture in other parts of the world. Efforts to eradicate *B. melitensis* infection from sheep and goat populations have been much less successful. These efforts have relied heavily on vaccination programs, which have tended to fluctuate with changing economic and political conditions. In some countries (e.g., Israel), *B. melitensis* has caused serious outbreaks in cattle. Infections with *B. melitensis* still pose a major public health problem in Mediterranean countries; in western, central, and southern Asia; and in parts of Africa and South and Central America. Infections with *B. abortus* are common in cattle-rearing communities in African countries such as Kenya and Uganda. Canine infection with *B. canis* is present on most continents—the incidence appears to be increasing in North America and in several European countries, often associated with importation of dogs from an endemic area.

Human brucellosis is usually associated with occupational or domestic exposure to infected animals or their products. Farmers, shepherds, goatherds, veterinarians, and employees in slaughterhouses and meat-processing plants in endemic areas are occupationally exposed to infection. Feral pig hunters are at risk of infection with *B. suis* in several countries, including Australia. Family members of individuals involved in animal husbandry may be at risk, although it is often difficult to differentiate food-borne infection from environmental contamination under these circumstances. Laboratory workers who handle cultures or infected samples also are at risk. Travelers and urban residents usually acquire the infection through consumption of contaminated foods. In countries that have eradicated the disease, new cases are most commonly acquired abroad. Dairy products, especially soft cheeses, unpasteurized milk, and ice cream, are the most frequently implicated sources of infection; raw meat and bone marrow may be sources under exceptional circumstances. Infections acquired through cosmetic treatments using materials of fetal origin have been reported. Person-to-person transmission is extremely rare, as is transfer of infection by blood or tissue donation. Although brucellosis is a chronic intracellular infection, there is no evidence for increased prevalence or severity among individuals with HIV infection or with immunodeficiency or immunosuppression of other etiologies.

Brucellosis may be acquired by ingestion, inhalation, or mucosal or percutaneous exposure. Accidental injection or ingestion of the live vaccine strains of *B. abortus* (S19 and RB51) and *B. melitensis* (Rev 1) can cause disease. *B. melitensis* and *B. suis* have historically been developed as biological weapons by several countries and could be exploited for bioterrorism **(Chap. S4)**. This possibility should be borne in mind in the event of sudden unexplained outbreaks.

IMMUNITY AND PATHOGENESIS

Exposure to brucellosis elicits both humoral and cell-mediated immune responses. The mechanisms of protective immunity against human brucellosis are presumed to be similar to those documented in laboratory animals, but such generalizations must be interpreted with caution. The response to infection and its outcome are influenced by the virulence, phase, and species of the infecting strain. Differences have been reported between *B. abortus* and *B. suis* in modes of cellular entry and subsequent compartmentalization and processing. Antibodies promote clearance of extracellular brucellae by bactericidal action and by facilitation of phagocytosis by polymorphonuclear and mononuclear phagocytes; however, antibodies alone cannot eradicate infection. Organisms taken up by macrophages and other cells can establish persistent intracellular infections. The key target cell is the macrophage, and bacterial mechanisms for suppressing intracellular killing and apoptosis result in very large intracellular populations. Opsonized bacteria are actively phagocytosed by neutrophilic granulocytes and by monocytes. In these and other cells, initial attachment takes place via specific receptors, including Fc, C3, fibronectin, and mannose-binding proteins. Opsonized—but not unopsonized—bacteria trigger an oxidative burst inside phagocytes. Unopsonized bacteria are internalized via similar receptors but at much lower efficiency. Smooth strains enter host cells via lipid rafts. Smooth lipopolysaccharide (LPS), β-cyclic glucan, and possibly an invasion–attachment protein (Alb) are involved in this process. Tumor necrosis factor α (TNF-α) produced early in the course of infection stimulates cytotoxic lymphocytes and activates macrophages, which can kill intracellular brucellae (probably mainly through production of reactive oxygen and nitrogen intermediates) and may clear infection. However, virulent *Brucella* cells can suppress the TNF-α response, and control of infection in this situation depends on macrophage activation and interferon γ (IFN-γ) responses. Cytokines such as interleukin (IL) 12 promote production of IFN-γ, which drives T_H1-type responses and stimulates macrophage activation. Inflammatory cytokines, including IL-4, IL-6, and IL-10, downregulate the protective response. As in other types of intracellular infection, it is assumed that initial replication of brucellae takes place within cells of the lymph nodes draining the point of entry. Subsequent hematogenous spread may result in chronic localizing infection at almost any site, although the reticuloendothelial system, musculoskeletal tissues, and genitourinary system are most frequently targeted. Both acute and chronic inflammatory responses develop in brucellosis, and the local tissue response may include granuloma formation with or without necrosis and caseation. Abscesses may also develop, especially in chronic localized infection.

The determinants of pathogenicity of *Brucella* have not been fully characterized, and the mechanisms underlying the manifestations of brucellosis are incompletely understood. The organism is a “stealth” pathogen whose survival strategy is centered on several processes that avoid triggering innate immune responses and that permit survival within monocytic cells. These processes include evasion of intracellular destruction by restricting the fusion of type IV secretion system–dependent *Brucella*-containing vacuoles with lysosomal compartments, inhibition of apoptosis of infected mononuclear cells, and prevention of dendritic cell maturation, antigen presentation, and activation of naïve T cells. The smooth *Brucella* LPS, which has an unusual O-chain and core-lipid composition, has relatively low endotoxin activity and plays a key role in pyrogenicity and in resistance to phagocytosis and serum killing in the nonimmune host. In addition, LPS is believed to play a role in suppressing phagosome–lysosome fusion and diverting the internalized bacteria into vacuoles located in endoplasmic reticulum, where intracellular replication takes place. Specific exotoxins have not been isolated, but a type IV secretion system (Vir) that regulates intracellular survival and trafficking has been identified. In *B. abortus*, this system can be activated extracellularly, but in *B. suis*, it is activated (by low pH) only during intracellular growth. Brucellae then produce acid-stable proteins that facilitate the organisms’ survival in phagosomes and may enhance their resistance to reactive oxygen intermediates. A type III secretion system based on modified flagellar structures also has been inferred, although not yet confirmed. Virulent brucellae are resistant to defensins and produce a Cu-Zn superoxide dismutase that increases their resistance to reactive oxygen intermediates. A hemolysin-like protein may trigger the release of brucellae from infected cells.

CLINICAL FEATURES

Brucellosis almost invariably causes fever, which may be associated with profuse sweats, especially at night. In endemic areas, brucellosis may be difficult to distinguish from other causes of fever. However, two features recognized in the nineteenth century distinguish brucellosis from other tropical fevers, such as typhoid and malaria: (1) Left untreated, the fever of brucellosis shows an undulating pattern that persists for weeks before the commencement of an afebrile period that may be followed by relapse. (2) The fever of brucellosis is associated with musculoskeletal symptoms and signs in about one-half of all patients.

The clinical syndromes caused by the different species are similar, although *B. melitensis* tends to be associated with a more acute and aggressive presentation and *B. suis* with focal abscess induction. *B. abortus* infections may have a more insidious onset and are more likely to become chronic. *B. canis* infections are generally regarded as less severe but, like other species, can cause serious disease such as endocarditis.

The incubation period varies from 1 week to several months, and the onset of fever and other symptoms may be abrupt or insidious. In addition to experiencing fever and sweats, patients become increasingly apathetic and fatigued; lose appetite and weight; and have nonspecific myalgia, headache, and chills. Overall, the presentation of brucellosis often fits one of three patterns: febrile illness that resembles typhoid but is less severe; fever and acute monoarthritis, typically of the hip or knee, in a young child; and long-lasting fever, misery, and low-back or hip pain in an older person (especially men). In an endemic area (e.g., much of the Middle East), a patient with fever and difficulty walking into the clinic would be suspected to have brucellosis until it was proven otherwise.

Diagnostic clues in the patient's history include travel to an endemic area, employment in a diagnostic microbiology laboratory, consumption of unpasteurized milk products (including soft cheeses), contact with animals, accidental inoculation with veterinary *Brucella* vaccines, and—in an endemic setting—a history of similar illness in the family (documented in almost 50% of cases). Focal features are present in the majority of patients. The most common are musculoskeletal pain and physical findings in the peripheral and axial skeleton (~40% of cases). Osteomyelitis more commonly involves the lumbar and low thoracic vertebrae than the cervical and high thoracic spine. Individual joints that are most commonly affected by septic arthritis are the knee, hip, sacroiliac, shoulder, and sternoclavicular joints; the pattern may be one of monoarthritis or polyarthritis. Osteomyelitis may also accompany septic arthritis.

In addition to the usual causes of vertebral osteomyelitis or septic arthritis, the most important disease in the differential diagnosis is tuberculosis. This point influences the therapeutic approach as well as the prognosis, given that several antimicrobial agents used to treat brucellosis are also used to treat tuberculosis. Septic arthritis in brucellosis progresses slowly, starting with small pericapsular erosions. In the vertebrae, anterior erosions of the superior end plate are typically the first features to become evident, with eventual involvement and sclerosis of the whole vertebra. Anterior osteophytes eventually develop, but vertebral destruction or impingement on the spinal cord is rare and usually suggests tuberculosis **(Table 174-1)**.

Other systems may be involved in a manner that resembles typhoid. About one-quarter of patients have a dry cough, usually with few changes visible on the chest x-ray, although pneumonia, empyema, intrathoracic adenopathy, or lung abscess can occur. Sputum or pleural effusion cultures are rarely positive in such cases, which respond well to standard brucellosis treatment. One-quarter of patients have hepatosplenomegaly, and 10–20% have significant lymphadenopathy; the differential diagnosis includes glandular fever–like illness such as that caused by Epstein-Barr virus, *Toxoplasma*, cytomegalovirus, HIV, or *Mycobacterium tuberculosis*. Up to 10% of men have acute epididymo-orchitis, which must be distinguished from mumps and from surgical problems such as torsion. Prostatitis, inflammation of the seminal vesicles, salpingitis, and pyelonephritis all occur. There is an increased incidence of fetal loss among infected pregnant women, although teratogenicity has not been described and the tendency toward abortion is much less pronounced in humans than in animals.

Neurologic involvement is common, with depression and lethargy whose severity may not be fully appreciated by either the patient or the physician until after treatment. A small proportion of patients develop lymphocytic meningoencephalitis that mimics mild neurotuberculosis, atypical leptospirosis, or noninfectious conditions. Rare manifestations include intracerebral abscess, a variety of cranial nerve deficits, or ruptured mycotic aneurysms.

TABLE 174-1 Radiology of the Spine: Differentiation of Brucellosis from Tuberculosis

	BRUCELLOSIS	TUBERCULOSIS
Site	Lumbar and others	Dorsolumbar
Vertebrae	Multiple or contiguous	Contiguous
Diskitis	Late	Early
Body	Intact until late	Morphology lost early
Canal compression	Rare	Common
Epiphysitis	Anterosuperior	General: upper and lower disk regions, central, subperiosteal
Osteophyte	Anterolateral (parrot beak)	Unusual
Deformity	Wedging uncommon	Anterior wedge, gibbus
Recovery	Sclerosis, whole-body	Variable
Paravertebral abscess	Small, well-localized	Common and discrete loss, transverse process
Psoas abscess	Rare	More likely

Endocarditis occurs in ~1% of cases, most often affecting the aortic valve (natural or prosthetic). Any site in the body may be involved in metastatic abscess formation or inflammation; the female breast and the thyroid gland are affected particularly often. Nonspecific maculopapular rashes and other skin manifestations are uncommon and are rarely noticed by the patient even if they develop.

■ DIAGNOSIS

Because the clinical picture of brucellosis is not distinctive, the diagnosis must be based on a history of potential exposure, a presentation consistent with the disease, and supporting laboratory findings. Results of routine biochemical assays are usually within normal limits, although serum levels of hepatic enzymes and bilirubin may be elevated. Peripheral leukocyte counts are usually normal or low, with relative lymphocytosis. Mild anemia may be documented. Thrombocytopenia and disseminated intravascular coagulation with raised levels of fibrinogen degradation products can develop. The erythrocyte sedimentation rate and C-reactive protein levels are often normal but may be raised.

In body fluids such as cerebrospinal fluid (CSF) or joint fluid, lymphocytosis and low glucose levels are the norm. Elevated CSF levels of adenosine deaminase cannot be used to distinguish tubercular meningitis, as they may also be found in brucellosis. Biopsied samples of tissues such as lymph node or liver may show noncaseating granulomas without acid-fast bacilli. The radiologic features of bony disease develop late and are much more subtle than those of tuberculosis or septic arthritis of other etiologies, with less bone and joint destruction. Isotope scanning is more sensitive than plain x-ray and continues to give positive results long after successful treatment.

Isolation of brucellae from blood, CSF, bone marrow, or joint fluid or from a tissue aspirate or biopsy sample is definitive, and attempts at isolation are usually successful in 50–70% of cases. Blood culture using modern nonradiometric or similar signaling systems (e.g., Bactec) usually become positive within 7 days. Clinicians should alert the laboratory to the possibility of brucellosis if suspected, as all cultures should be handled under containment conditions appropriate for dangerous pathogens. *Brucella* species may be misidentified as *Agrobacterium*, *Ochrobactrum*, or *Psychrobacter (Moraxella) phenylpyruvicus* by the gallery identification strips that may still be used in the diagnostic laboratory. In recent years, matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) has emerged as a powerful tool for bacterial identification in well-resourced laboratories. Earlier implementations of MALDI-TOF often failed to identify *Brucella* species correctly as a genus or at species level. Following recent changes in taxonomy, *Ochrobactrum* isolates may now be reported as a subspecies of *Brucella*. This can lead to confusion and have important clinical and public health consequences, as management differs completely. Clinicians should confirm the meaning of any culture report suggestive of brucellosis with laboratory colleagues before making therapeutic decisions.

The peripheral blood–based polymerase chain reaction (PCR) has enormous potential to detect bacteremia, to predict relapse, and to exclude "chronic brucellosis." This method is more sensitive and quicker than blood culture, and it does not carry the attendant biohazard risk posed by culture. However, it is not perfect, and false-negative results are sometimes observed in patients with positive blood cultures. Nucleic acid amplification techniques are now quite widely used, although no single standardized procedure has been adopted. Primers for the spacer region between the genes encoding the 16S and 23S ribosomal RNAs (*rrs-rrl*), various outer-membrane protein–encoding genes, the insertion sequence *IS711*, and the protein BCSP31 are sensitive and specific. Blood and other tissues are the most suitable samples for analysis. The clinical significance of prolonged PCR positivity, commonly seen in blood after successful treatment, remains controversial.

Serologic examination often provides the only positive laboratory findings in brucellosis. In acute infection, IgM antibodies appear early and are followed by IgG and IgA. All these antibodies are active in agglutination tests, whether performed by tube, plate, or microagglutination methods. The majority of patients have detectable agglutinins at this stage. As the disease progresses, IgM levels decline, and the avidity and subclass distribution of IgG and IgA change. The result is reduced or undetectable agglutinin titers. However, the antibodies are detectable by alternative tests, including the complement fixation test, Coomb's antiglobulin test, and enzyme-linked immunosorbent assays. There is no clear cutoff value for a diagnostic titer. Rather, serology results must be interpreted in the context of exposure history and clinical presentation. In endemic areas or in settings of potential occupational exposure, agglutinin titers of 1:320–1:640 or higher are considered diagnostic; in nonendemic areas, a titer of ≥1:160 is considered significant. Repetition of tests after 2–4 weeks may demonstrate a rising titer.

In most centers, the standard agglutination test (or a derivative such as the microagglutination test) is still the mainstay of serologic diagnosis. In an endemic setting, >90% of patients with acute bacteremia have standard agglutination titers of at least 1:320 at the time of clinical presentation. Some centers rely on the Rose Bengal test, particularly for screening, but it has been only partially validated for human diagnostic use. A variety of near-patient or point-of-care tests are still in developmental stages.

Antibody to the *Brucella* LPS O chain—the dominant antigen—is detected by all the conventional tests that employ smooth *B. abortus* cells as antigen. Because *B. abortus* cross-reacts with *B. melitensis* and *B. suis*, there is no advantage in replicating the tests with these antigens. Cross-reactions also occur with the O chains of some other gram-negative bacteria, including *Yersinia enterocolitica* O:9, *Escherichia coli* O157, *Francisella tularensis*, *Salmonella enterica* group N, *Stenotrophomonas maltophilia*, and *Vibrio cholerae*. The cell-surface antigens of rough *Brucella* strains such as *B. canis* or *B. ovis* do not react in these serologic tests, which will give false-negative results. Specific antigen-based tests used in veterinary practice to detect *B. canis* and *B. ovis* produce inconsistent results in human diagnosis and are not validated for this purpose. Clinicians should consult with the laboratory about test selection if *B. canis* infection is suspected. Similarly, live *B. abortus* vaccine strain RB51 does not elicit antibody responses in serologic tests that use smooth antigens, and this fact must be taken into account if serologic tests are employed in attempts to identify or follow the course of infections in persons accidentally exposed to the vaccine.

TREATMENT

Brucellosis

The broad aims of antimicrobial therapy are to treat and relieve the symptoms of current infection and to prevent relapse. Focal disease presentations may require specific intervention in addition to more prolonged and tailored antibiotic therapy. In addition, tuberculosis must always be excluded, or—to prevent the emergence of resistance—therapy should be tailored to specifically exclude drugs active against tuberculosis (e.g., rifampin used alone) or to include a full antituberculous regimen.

Early experience with streptomycin or tetracycline monotherapy showed that relapse was common; thus dual therapy with both agents became the norm. This is still the most effective combination, but alternatives may be used, with the options depending on local or national policy about the use of rifampin for the treatment of nonmycobacterial infection. For the several antimicrobial agents that are active in vivo, efficacy can usually be predicted by in vitro testing. However, numerous *Brucella* strains show in vitro sensitivity to a whole range of antimicrobials that are therapeutically ineffective, including assorted β-lactams. The use of fluoroquinolones remains controversial despite the good in vitro activity and white-cell penetration of most agents of this class. Low intravacuolar pH is probably a factor in the poor performance of these drugs.

For adults with acute nonfocal brucellosis (duration <1–2 months), a 6-week course of therapy incorporating at least two antimicrobial agents is required. Complex or focal disease may necessitate ≥3 months of therapy. Adherence to the therapeutic regimen is very important, and poor adherence underlies almost all cases of apparent treatment failure; such failure is rarely due to the emergence of drug resistance, although increasing resistance to trimethoprim-sulfamethoxazole (TMP-SMX) is being reported.

The gold standard for the treatment of brucellosis in adults is IM streptomycin (0.75–1 g daily for 14–21 days) together with doxycycline (100 mg twice daily for 6 weeks). In both clinical trials and observational studies, relapse follows such treatment in 5–10% of cases. The usual alternative regimen favored by the World Health Organization and many national guidelines consists of rifampin (600–900 mg/d) plus doxycycline (100 mg twice daily) for 6 weeks. This is easier for the patient and health care provider but requires sustained treatment adherence. The relapse/failure rate of this regimen is ~10% in trial conditions but can rise to >20% in many nontrial situations, possibly because doxycycline levels are reduced and clearance rates increased by concomitant rifampin administration. Patients who cannot tolerate or receive tetracyclines (children, pregnant women) can be given high-dose TMP-SMX instead (two or three standard-strength tablets twice daily for adults, depending on weight).

Good evidence supports the use of an aminoglycoside such as gentamicin (5–6 mg/kg per day for 1–2 weeks) instead of streptomycin, and this is recommended in U.S. and U.K. guidelines. Shorter courses of gentamicin have been associated with high failure rates in adults. Early experience with fluoroquinolone monotherapy was disappointing, but it has been suggested that ofloxacin or ciprofloxacin, given together with rifampin for 6 weeks, might be an acceptable alternative to the other 6-week regimens for adults. A substantial meta-analysis did not support the use of fluoroquinolones in first-line treatment regimens, and these drugs were not recommended by an expert consensus group (the Ioannina Recommendations) except in the context of well-designed clinical trials. However, a more recent meta-analysis is more supportive of the efficacy of these drugs, and adequately powered prospective studies will be needed to resolve their role in standard combination therapy. Although triple-drug regimens are superior to double-drug regimens in meta-analyses, they are not indicated for uncomplicated brucellosis. A triple-drug regimen such as doxycycline and rifampin, enhanced by initial aminoglycoside, should be considered for all patients with complicated disease and for those for whom treatment adherence is likely to be a problem.

Focal neurologic disease due to *Brucella* species requires prolonged treatment (i.e., for 3–6 months), usually with ceftriaxone supplementation of a standard regimen. *Brucella* endocarditis is treated with at least three drugs (an aminoglycoside, a tetracycline, and rifampin), and many experts add ceftriaxone and/or a fluoroquinolone to reduce the need for valve replacement. Treatment is usually given for at least 4–6 months, and clinical endpoints for its discontinuation are often difficult to define. Surgery is still required for the majority of cases of infection of prosthetic heart valves and prosthetic joints.

There is good evidence for the benefit of antimicrobial prophylaxis to reduce the risk of infection after occupational or other exposure to *Brucella* organisms, inadvertent immunization with live vaccine intended for use in animals, or exposure to deliberately released brucellae. However, there is little evidence to inform the duration of prophylaxis or preference for dual or single therapy. Based on historical practices, many national guidelines recommend the administration of rifampin plus doxycycline for 3 weeks. However, such regimens are poorly tolerated, and doxycycline monotherapy of the same duration may be used instead. Monotherapy is the standard recommendation in the United Kingdom but not in the United States. Rifampin should be omitted after exposure to vaccine strain RB51, which is resistant to rifampin, and replaced by another agent such as TMP-SMX in combination with doxycycline. After significant brucellosis exposure, expert consultation is advised for women who are (or may be) pregnant.

■ PROGNOSIS AND FOLLOW-UP

Relapse occurs in up to 30% of poorly compliant patients. Thus, patients should ideally be followed clinically for up to 2 years to detect relapse, which responds to a prolonged course of the same therapy used originally. The general well-being and the body weight of the patient are more useful guides than serology to lack of relapse. IgG antibody levels detected by the standard agglutination test and its variants can remain in the diagnostic range for >2 years after successful treatment. Complement fixation titers usually fall to normal within 1 year of cure. Immunity is not solid; patients can be reinfected after repeated exposures. Fewer than 1% of patients die of brucellosis. When the outcome is fatal, death is usually a consequence of cardiac involvement; more rarely, it results from severe neurologic disease. Despite the low mortality rate, recovery from brucellosis is slow, and the illness can cause prolonged inactivity, with domestic and economic consequences.

The existence of a prolonged chronic brucellosis state after successful treatment remains controversial. Evaluation of patients in whom this state is considered (often those with work-related exposure to brucellae) includes careful exclusion of malingering, nonspecific chronic fatigue syndromes, and other causes of excessive sweating, such as alcohol abuse and obesity. In the future, the availability of more sensitive assays to detect *Brucella* antigen or DNA may help to identify patients with ongoing infection.

■ PREVENTION

Vaccines based on live attenuated *Brucella* strains, such as *B. abortus* strain 19BA or 104M, have been used in some countries to protect high-risk populations but have displayed only short-term efficacy and high reactogenicity. Subunit vaccines have been developed but are of uncertain value and cannot be recommended at present. Research in this area has been stimulated by interest in biodefense **(Chap. S4)** and may eventually yield new products. The mainstay of veterinary prevention is a national commitment to testing and slaughter of infected herds/flocks (with compensation for owners), control of animal movement, and active immunization of animals. These measures are usually sufficient to control human disease as well. In their absence, pasteurization of all milk products before consumption is sufficient to prevent nonoccupational animal-to-human transmission. All cases of brucellosis in animals and humans should be reported to the appropriate public health authorities.

■ FURTHER READING

Beeching NJ et al: Brucellosis. BMJ Best Practice, 2023. *https://bestpractice.bmj.com/topics/en-us/911.*

Bosilkovski M et al: The current therapeutical strategies in human brucellosis. Infection 49:823, 2021.

Centers for Disease Control and Prevention: Brucellosis. *https://www.cdc.gov/brucellosis/index.html.*

Dean AS et al: Clinical manifestations of human brucellosis: A systematic review and meta-analysis. PLoS Negl Trop Dis 6:e1929, 2012.

Norman FF et al: Imported brucellosis: A case series and literature review. Travel Med Infect Dis 14:182, 2016.

Zagorsky P et al: Laboratory diagnosis of human brucellosis. Clin Microbial Rev 33:e00073, 2019.

175

Tularemia

Anders F. Johansson

■ DEFINITION

Tularemia is a zoonotic disease that can naturally be transmitted from vertebrate animals to humans, causing a febrile illness. Synonyms include deer-fly fever, rabbit fever, market men's disease, water-rat trappers' disease, wild hare disease (yato-byo), Francis' disease, and Ohara's disease.

■ ETIOLOGY

Tularemia is caused by *Francisella tularensis*, a small (0.2–0.7 × 0.2 μm), facultatively intracellular, aerobic, gram-negative, pleomorphic bacterium. *F. tularensis* is highly virulent for humans and numerous mammals, including rodents, hares, and rabbits. It is known to cause airborne laboratory-acquired infection; therefore, laboratory work, including culturing of the bacterium, must be performed by trained staff in biologic safety cabinets under biosafety level 3 conditions. The disease tularemia was described by the American researchers McCoy and Chapin as a "plague-like disease of rodents" in California in 1911, 1 year after they cultivated and described the causative bacterium as "Bacterium tularense." *Francisella* was later named after Edward Francis, an American bacteriologist who extensively studied the etiologic agent and the pathogenesis of tularemia over several decades. There are two *F. tularensis* subspecies of high clinical importance: *F. tularensis* subspecies *tularensis* (aka Jellison type A) with human disease documentation exclusively from the North American continent, and subspecies *holarctica* (Jellison type B) with human disease documentation both from North America and from other parts of the world. A third subspecies, *F. tularensis* subspecies *mediasiatica*, has been isolated from vertebrate animals in Central Asia and Russia, but there is no documented disease in humans. The division into type A and type B strains of *F. tularensis* is based on minor biochemical test differences, with type A strains being significantly more lethal in experimentally infected rabbits. Type A but not type B cultures can ferment glycerol and are positive in a citrulline ureidase assay. Microbial genome sequencing has revealed an average nucleotide identity >99% between type A and type B strains of *F. tularensis,* but current and historic epidemiologic studies mirror the experimental findings in rabbits—tularemia mortality among humans is higher with type A in the United States. However, recent research using high-resolution genetic subtyping, combined with analyses of disease outcomes, shows there is heterogeneity in mortality risks among different type A genetic varieties, with some being less virulent for humans than the type B varieties.

Genetic neighbors of *F. tularensis*, including *Francisella novicida* (sometimes referred to as a fourth subspecies of *F. tularensis* with a genomic average nucleotide identity of ~97% with other subspecies) and the genetically more distant *Francisella philomiragia*, are opportunistic pathogens that may cause disease in humans with symptoms varying according to the immune status of the patient. Moreover, modern microbiology methods and intense research efforts have resulted in the identification of additional *Francisella* bacteria, including pathogens of fish, endosymbiont bacteria of ticks, soil bacteria, and bacteria isolated from sea and freshwater, some of which may be opportunistic human pathogens. Taxonomists now recognize at least nine distinct *Francisella* species, including *Francisella tularensis*, the causative agent of tularemia.

EPIDEMIOLOGY

F. tularensis infects humans through bites of arthropods functioning as disease vectors (e.g., mosquitoes, ticks, tabanid flies), inhaling infectious aerosols, handling infected animals, or ingesting contaminated water. The reported human cases represent bridging between endemic *F. tularensis* maintenance in nature, or spillover events from massive amplification in infected rodents, hares, or rabbits during disease outbreaks among animals (epizootics). The disease is endemic to the Northern Hemisphere but has a patchy and geographically uneven distribution, which is believed to be the result of *F. tularensis* being maintained long-term in smaller geographic areas experiencing repeated outbreaks, often at intervals of several years. In the United States, reported human cases have dropped sharply from several thousand annual cases before 1950 to 200–300 cases annually from 2010–2020, with most of these cases being associated with tick bites in Arkansas, Kansas, Missouri, and Oklahoma according to Centers for Disease Control and Prevention (CDC) statistics. Surveillance data from 2006–2021 with available *F. tularensis* culture data showed 66% type A and 34% type B cases in humans. In some parts of Europe and Asia where type B only causes human disease, tularemia was relatively more common in 2010–2022, with several hundred to >2000 cases in peak years in Finland, Sweden, and Turkey, respectively. In the Northern Boreal Forest regions of Sweden, Finland, and Russia, tularemia is primarily associated with mosquito bites, while tick bites are more common in Western and Eastern Europe. Most cases in the United States, as well as in Finland and Sweden, occur during summer, likely reflecting increased human exposure to blood-feeding arthropods, infected wildlife mammals, or aerosols from carcasses of infected animals during outdoor activities. In contrast to this seasonal pattern, most cases occur in the winter months in Turkey, where the predominant infection route to humans in some rural areas is nonchlorinated drinking water from natural springs that may be contaminated by infected rodents. Colder temperatures enhance the survival of the bacterium in water during winter. In the Southern Hemisphere, rare human and animal infections have been documented from Australia. It is uncertain whether the lack of reporting from other areas of the Southern Hemisphere is due to lack of disease or from underreporting.

Based on extensive historic field studies performed in the United States and the former Soviet Union during the 1950s–1970s, it is believed that distinct disease ecologies characterize *F. tularensis* subspecies *tularensis* (type A) and *F. tularensis* subspecies *holarctica* (type B), respectively. This concept recognizes the association of type A tularemia with rabbits, ticks, domestic cats, and sheep in comparatively dry environments, in contrast to the association of type B tularemia with streams, ponds, lakes, rivers, and semiaquatic animals such as muskrats and beavers **(Fig. 175–1)**.

It remains unknown exactly how the bacterium is maintained in nature. *F. tularensis* is a generalist eukaryotic host-associated microbe that can infect many animals and cell types, including amphibians, birds, rodents, rabbits, carnivores, and ruminant mammals. Most of these animals are accidental hosts or rapidly die from septic disease. It is unclear if any can maintain the bacterium for a long duration. However, when an animal dies from a tularemia infection with a high level of bacteria in their bodies, this can potentially help temporarily maintain the bacteria in the environment. Over the long term, decaying infectious tissues might also contribute to the bacteria's persistence

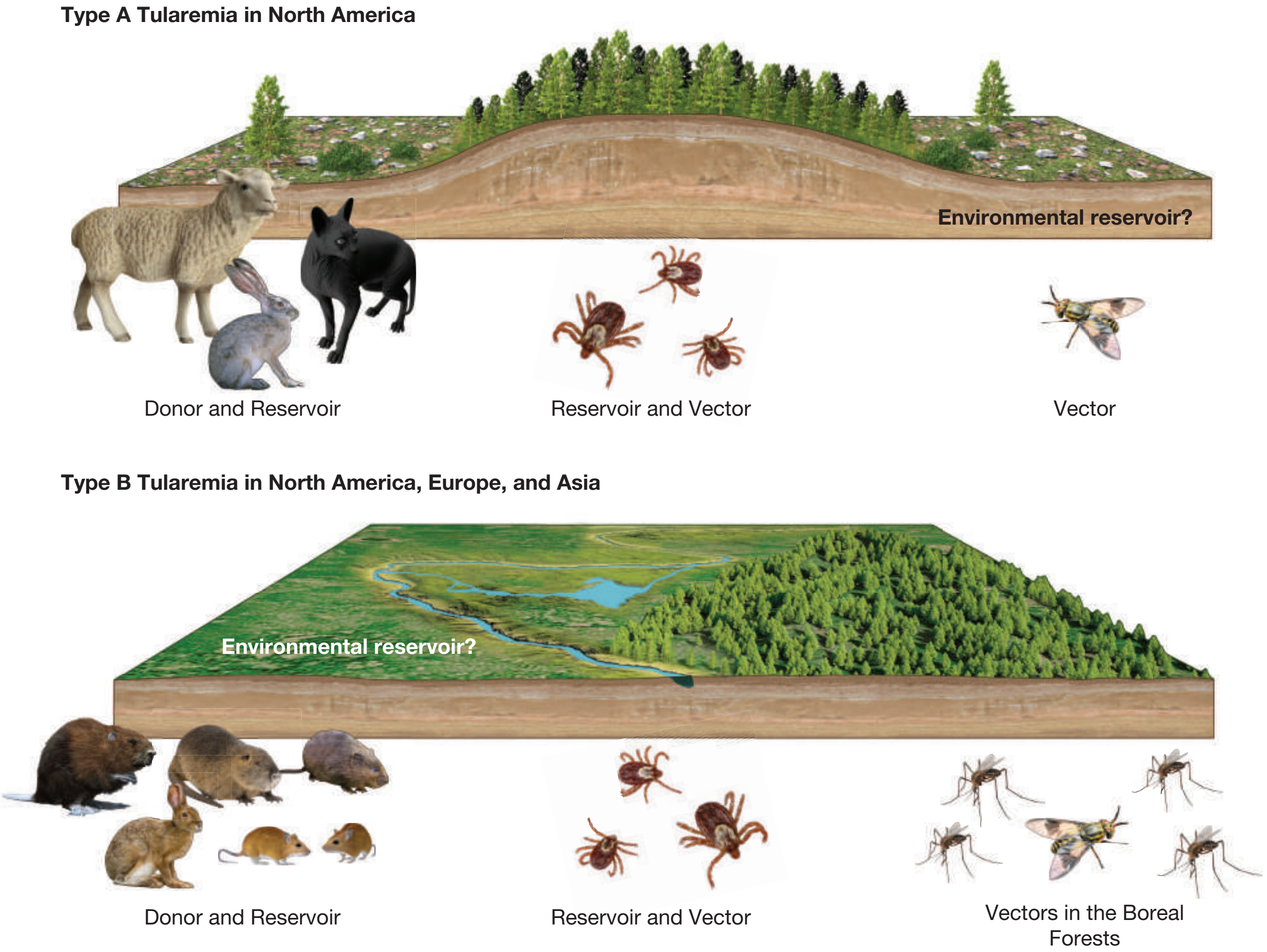

FIGURE 175–1 Schematic illustration of the type A and type B disease ecology concepts. *(Adapted from P. Keim et al: Molecular epidemiology, evolution, and ecology of Francisella. Ann N Y Acad Sci 1105:30, 2007, Figure 11.)*

under certain conditions. Another hypothesis for maintenance of *F. tularensis* is persistence in water with protozoa such as amoeba serving as host cells.

Since the 1940s, research has shown that *F. tularensis* can infect multiple species of soft and hard ticks, supporting the idea of tularemia as an obligate vector-borne disease with ticks as reservoirs. In North America, tularemia in humans is mainly tick-borne, but the role of ticks in maintaining and spreading the disease geographically is more debated compared with other pathogens like *Babesia* and *Borrelia*; these latter pathogens are better adapted for transmission between the different life stages of a tick. Contemporary well-designed experiments involving mice and American dog ticks have tested *F. tularensis* type A strains from three main genetic clades and a type B strain. They demonstrated rapid *F. tularensis* acquisition by adult ticks from bacteremic mice, an interrupted blood-feeding behavior, and high transmission efficiency to uninfected mice (58–89%), indicating that adult ticks can serve as highly effective disease vectors to humans during outbreaks. However, ticks of the more immature nymph life stage showed low survival after *F. tularensis* ingestion and low transmission efficiency (0–13.5%), indicating that these are ineffective disease vectors. Overall, both older and newer studies suggest that *F. tularensis* infection poses a high burden for immature tick stages, and the literature is conflicted on the presence of transovarial transmission in ticks. It has been suggested that ticks, like other blood-feeding arthropods including tabanid flies in North America and mosquitoes in the Boreal Forest regions of Northern Europe, are important for transmitting disease to humans during tularemia epizootics but may not be obligate for maintaining *F. tularensis*. To better understand the zoonotic potential and the epidemiology seen in humans, a better understanding of how *F. tularensis* is maintained between outbreaks will be required.

PATHOGENESIS

Natural transmission to humans usually occurs through the skin or by inhalation of aerosols containing *F. tularensis*, but ingesting infected food or water also may cause infection. An extremely low dose of 10–50 bacteria inoculated through microtrauma of the skin or inhaled into the lungs suffices to cause disease in humans. In contrast, a dose in the range of 10^6–10^8 bacteria is required to infect humans and other primates by the oral route. Tissue-resident macrophages, including alveolar macrophages in the lung and Langerhans cells in the skin, are believed to be important in the transport of *F. tularensis* by blood or lymph vessels to lymph nodes, spleen, and bone marrow. In the skin, *F. tularensis* undergoes phagocytosis by neutrophils at initial stage of infection. After the bacteria have been taken up by tissue-resident macrophages and other local phagocytic cells, *F. tularensis* escapes from the phagosome into the cell cytosol, replicates, leaves the cell, and disseminates throughout the host. *F. tularensis* is a "stealth pathogen," implying that the bacterium is able to manipulate host-cell signaling, thus rendering itself a safe milieu inside the host cell suitable for its replication. In experimental oral infection of mice, there is minimal local gastrointestinal pathology despite the large infection dose required. It remains unknown what cell types mediate *F. tularensis* to cross the intestinal epithelium to cause fatal systemic infection.

Gross and histopathology findings in nonhuman primates experimentally infected by the inhalation route have shown 0.5- to 1-mm necrotic lesions with live bacteria in the spleen at day 4 after the infection, and subsequent larger foci in liver and lymph nodes at days 5–7, which likely mirror disease events and progression in more severe human tularemia cases. At day 6 after infection of nonhuman primates, randomly scattered subacute abscesses and pyogranulomas 6–20 mm in diameter have been observed on the lung surfaces. From day 7 and onward, granuloma formation and necrotic lesions in tissues surrounded by layers of macrophages, lymphocytes, and giant cells are typical reaction patterns in lymph nodes, lungs, spleen, kidney, and liver, corroborated by historic reports of similar autopsy findings in fatal human tularemia.

The immune invasion strategy of *F. tularensis*—delaying the immune response and permitting rapid systemic distribution—has made studies of early immune responses challenging. The mechanisms behind cell-mediated responses against *F. tularensis* after a primary infection or vaccination by the live vaccine strain of *F. tularensis*, however, are well described and require both B and T cells, resulting in antigen-specific recall T cell responses persisting 25–30 years after infection or vaccination. Nevertheless, although the bacterium can manipulate and delay them, effective early immune responses occur and involve complement, antibodies, neutrophils, inducible NO synthase, phagocyte oxidase, and cytokines such as interferon γ, tumor necrosis factor α, and interleukin 12.

APPROACH TO THE PATIENT

Clinical recognition of tularemia may be challenging, especially if the patient does not spontaneously report fever. Sporadic cases of patients seeking health care can easily be misdiagnosed with a diversity of other conditions, including lung cancer, head and neck cancer, lymphoma, tuberculosis, or another disease with lymphoid tissue engagement or a suspected tumor. Tularemia may not be suspected before extensive and costly medical investigations have already been performed. Awareness of the disease and a high index of suspicion is critical. Even in high-endemic areas, patients with tularemia often initially are mistakenly managed and treated for other conditions, including unspecified viral infection, common skin and soft tissue infection, common community-acquired pneumonia, or undifferentiated fever or sepsis. A careful disease history of a patient with tularemia typically reveals an acute fever onset episode with influenza-like illness. If the patient seeks care immediately, an acute fever onset can be easily recollected; if care is sought several weeks after initial symptoms, acute onset of fever may not be mentioned spontaneously. The acute illness may be described as "a severe virus" with high fever, muscle pain, headache, and other constitutional symptoms. Fever may be persistent or intermittent. A tularemia-focused disease history should include risk factors for exposure to *F. tularensis* and epidemiologic information (e.g., household member or friend having similar symptoms and sharing the same exposure). Exposure risk factors or activities include wild animal contact (e.g., rodents, rabbits, hares), contact with or inhalation of dust contaminated with their urine or feces, recent arthropod bites (tick, fly, or mosquitoes), performing brush cutting or grass trimming, handling of wood, handling of hay, and walking or hunting in the forest or other natural habitats of wildlife species.

When tularemia is suspected it is important to define patient factors with relevance for coping with infection, to define the infection syndrome, and to decide on an appropriate level for patient management:

- Consider factors that may increase the risk of severe infection (e.g., geriatric frailty, immunosuppression, severe lung and/or heart function impairment, other significant comorbidities).
- By physical examination, determine any anatomic location and extent of inflammation (e.g., presence of pneumonitis, lymph node inflammation, skin inflammation including minute ulcers or pustules, pharyngitis). Collect information on the rate of progression and judge the severity of infection.
- Determine an appropriate level of care (e.g., localized symptoms likely suitable for outpatient care, multiorgan involvement/hemodynamic instability suggesting inpatient care).

CLINICAL MANIFESTATIONS

The most common initial manifestation of tularemia is an influenza-like illness with high fever. The incubation period is usually 3–5 days but may range from 1 to 21 days. Inflamed swollen lymphoid tissue is typical for tularemia. The initial infection route determines what lymphoid tissue is engaged **(Fig. 175–2)**. If bacteria are acquired through skin or oral mucous membranes of the pharynx, palpable enlarged tender regional lymph node(s) can result. If bacteria are inhaled, the result will be lymph node enlargement in the mediastinum. Different recognized clinical forms of tularemia are the same for type A and type B tularemia and depend on the initial *F. tularensis* infection route as outlined in **Table 175-1**.

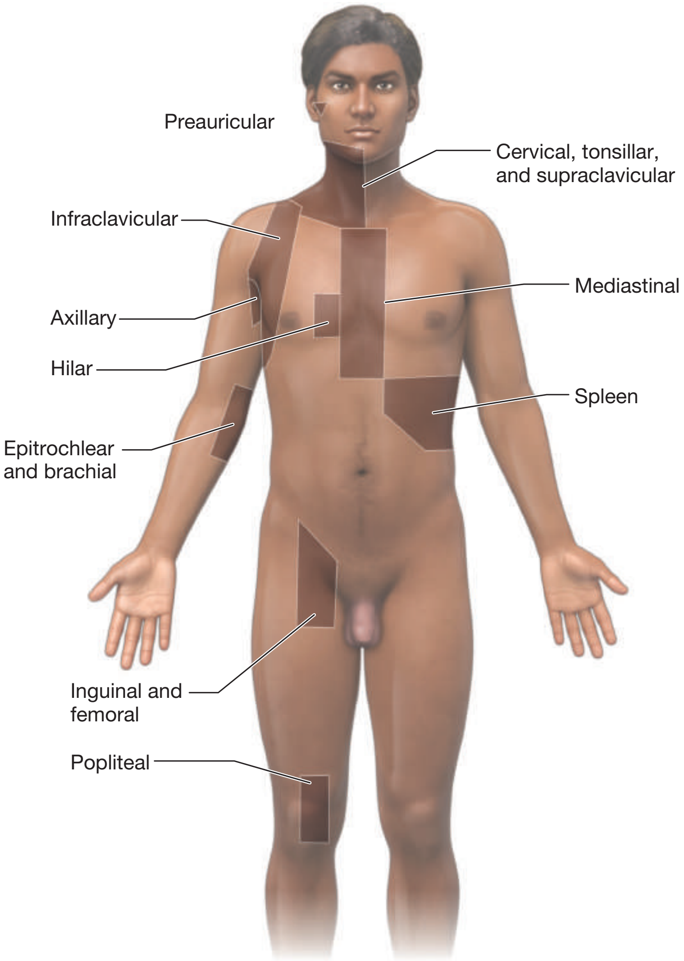

FIGURE 175–2 Characteristic sites of lymphoid tissue engagement in tularemia. *(From https://training.seer.cancer.gov/lymphoma/anatomy/lymph-nodes.html.)*

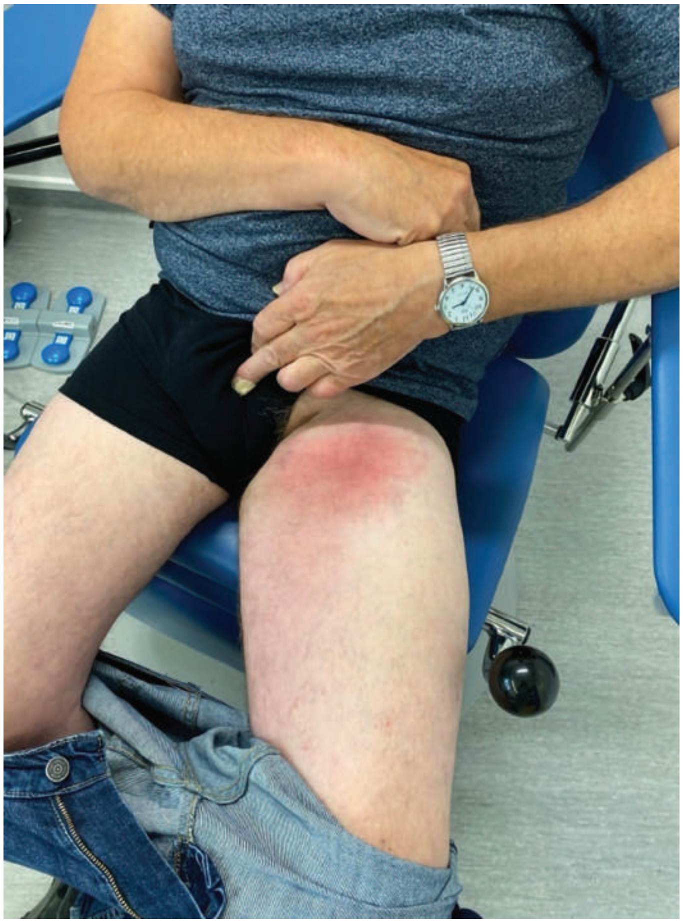

FIGURE 175–3 Patient presenting with ulceroglandular tularemia 3 weeks after onset of high fever and influenza-like illness. Lymph nodes of the left thigh were red, tender, and swollen. A tularemia ulcer was present on the left ankle (not shown). *(Photo from patient care at Region Västerbotten, Sweden.)*

The primary tularemia ulcer is often a small break in the skin, with limited signs of inflammation, and may require careful clinical examination to be detected. In some cases, it may not be detectable and the disease form will then be classified as glandular. Red, tender, swollen lymph regional nodes are often the most remarkable clinical examination findings in patients with acute ulceroglandular tularemia **(Fig. 175–3)**.

In oropharyngeal and oculoglandular tularemia, there is more often pronounced local inflammation at the primary site of infection with acute pharyngitis or conjunctivitis, respectively. The respiratory or pneumonic form of tularemia acquired by inhalation of *F. tularensis* may, more often than other clinical forms, be bacteremic and progress to severe disease requiring inpatient care, particularly in patients aged 65 and over and those with comorbidities. In the respiratory/pneumonic clinical form of tularemia, onset of dry cough may occur more than 7 days after onset of the influenza-like illness. Bronchoscopy findings may include signs of local tracheitis and pneumonitis, and the lymphoid tissue engaged will be in the mediastinum or lung hila. "Typhoidal tularemia" is an older medical classification term referring to a disease syndrome like typhoid fever with prolonged high fever, fatigue, headache, and nausea. It is recommended to avoid using this term, if possible, as it is believed that historically, most cases of typhoidal tularemia were acquired by inhalation and were, in fact, respiratory/pneumonic tularemia. For type A tularemia, historic disease descriptions suggest that respiratory/pneumonic tularemia or meningitis may also occur as secondary manifestations in severe cases of ulceroglandular disease. Data from the United States in 2006–2021 on 1046 cases with primary disease manifestations available showed 47% ulceroglandular, 18% glandular, 17% pneumonic or respiratory, 14% typhoidal, 2% oculoglandular, and 2% oropharyngeal disease according to CDC statistics.

TABLE 175-1 Clinical Forms of Tularemia

FORM[a]	*F. TULARENSIS* INFECTION ROUTE
Ulceroglandular or glandular	Skin inoculation by blood-feeding arthropods or direct contact (touching infected animal or *F. tularensis* contaminated material)
Oropharyngeal	Ingestion of contaminated water or food
Oculoglandular	Touching the eye with contaminated fingers or exposure to contaminated aerosol
Respiratory/pneumonic	Inhalation of contaminated aerosol
Typhoidal	Unknown (likely by inhalation, rarely by ingestion)

[a] Additional less common but well-described clinical invasive manifestations of tularemia exist, including primary tularemia meningitis, endocarditis, and bone and joint infections. In patients with immunosuppression, tularemia manifestations may be indolent with few focal symptoms and signs.

Source: Reproduced with permission from A Tärnvik (ed): *WHO Guidelines on Tularaemia: Epidemic and Pandemic Alert and Response*. Geneva, World Health Organization Press, 2007.

Routine biochemical testing of blood generally is not helpful in tularemia because findings are unspecific and may or may not reveal elevated liver enzymes or red blood cells in the urine. In a recent more extensive case series from Sweden of type B tularemia of primarily outpatients, white blood cell counts remained within normal limits for 70% of patients, and the median C-reactive protein level was moderately elevated, peaking at day 7–9 after disease onset at a median of 78 mg/L (range, 30–130). A recent analysis of 33 patients hospitalized with severe respiratory type B tularemia showed peak C-reactive protein levels at 100–400 mg/L among patients with *F. tularensis* growth in blood cultures and 50–220 mg/L among nonbacteremic patients.

■ DIAGNOSIS

Laboratory confirmation is recommended and may be based on recovery of an *F. tularensis* isolate, *F. tularensis*–specific antigens or nucleic acids, or the detection of *F. tularensis*–specific antibodies in blood by serology testing. Appropriate clinical specimens to collect for direct detection of *F. tularensis* by nucleic acids or culture include ulcer specimens, lower respiratory tract specimens (sputum/tracheal washing/bronchoalveolar lavage), lymph node biopsies or aspirates, and blood cultures. Additional clinical specimens with *F. tularensis* growth reported include pharyngeal and ventricular washings. Importantly, sampling from ulcers should be performed from the edge of the wound, preferably selecting a swab with a slightly stiffer stick and some rubbing of the tissue so that it becomes somewhat blood-tinged. If there is a scab, this should be removed before sampling. If culture specimens are sent to a clinical microbiology laboratory, a suspicion of tularemia must be clearly communicated to the laboratory in advance to minimize the risk of laboratory-acquired infections in lab workers.

Direct detection of *F. tularensis*–specific nucleic acid in clinical specimens by polymerase chain reaction or another detection method is a rapid and reliable diagnostic method. For ulcer specimens, the method is comparable with culture in sensitivity and specificity. There is clinical experience of *F. tularensis* nucleic acid detection from other specimen types, including sputum, bronchoalveolar lavage, and lymph node biopsies/aspirates, but formal scientific evaluation of the diagnostic performance is lacking. With increased availability of molecular methods in some clinical microbiology laboratories, nucleic acid–based subtyping without a prior culture step can be performed directly from clinical specimens and may help clinicians to identify *F. tularensis* varieties rapidly. Identification of the subspecies and genetic variety involved in tularemia infections is currently limited in most laboratories but may be of greatest importance in North America, where type A and type B tularemia exist side by side. Cultivation of *F. tularensis* requires Biosafety Level 3 (BSL-3) laboratories, and because *F. tularensis* is a U.S. Department of Health and Human Services (HHS) Tier 1 Select Agent, an accurate culture diagnosis in the clinic is challenging. However, culture identification provides a definitive diagnosis and allows for subsequent genetic methods to identify different *F. tularensis* varieties with varying risks of mortality. Unexpected findings in blood cultures with matrix-assisted laser desorption/ionization–time-of-flight (MALDI-TOF) mass spectrometry identification of *F. tularensis* is a scenario that has become more common in recent years with the improvement of blood culture systems and MALDI-TOF reference libraries. *F. tularensis* is a slow-growing bacterium, and an incubation period of 10 days is recommended for optimal sensitivity in blood-culture diagnostics. Serology remains commonly used to confirm tularemia and is the mainstay diagnostic in many clinical microbiology laboratories. Serology is particularly useful in patients when tularemia is suspected as a differential diagnosis at a later stage—the antibody responses against *F. tularensis* are generally detectable in the blood of patients 10–20 days post-infection using an enzyme-linked immunosorbent assay or a microagglutination test. Some commercial serologic tests may detect *F. tularensis*–specific antibodies as early as 1–2 weeks after disease onset. Importantly, the persistence of *F. tularensis* antibodies is prolonged, meaning that up to at least 1 year after acute infection, single titers of antibodies remain above the cutoff set by laboratories to detect acute infections. Therefore, a reliable serologic diagnosis of tularemia in the acute phase should rely on the demonstration of a significant increase of antibody titers between two samples taken 2–4 weeks apart.

Chest imaging is useful in respiratory tularemia. Single or multifocal consolidation may be present, accompanied by pleural effusion with hilar and mediastinal lymphadenopathy **(Fig. 175–4)**.

Computed tomography (CT) scans in the acute phase may reveal multiple rounded consolidations, often in the subpleural region and sometimes with signs of inclusion necrosis with a diffuse marginal zone. Lymph node enlargements in the mediastinal and hilar regions on a CT scan may also contain necrotic inclusions. A relatively common scenario for sporadic cases of tularemia is pulmonary nodules or dense consolidations, which can mimic lung cancer, resulting in more extensive clinical investigation paths to exclude cancer, including CT and positron emission tomography (PET) scans **(Fig. 175–5)**. Endobronchial ultrasound bronchoscopy procedures with biopsy may be performed in some cases before tularemia is suspected. The resolution of tularemia pathology on chest imaging may be slow and take months.

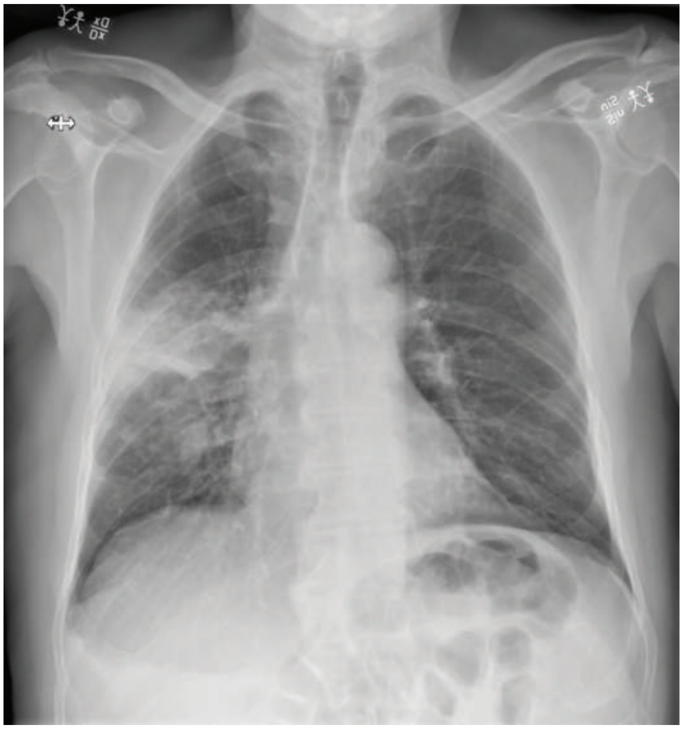

FIGURE 175–4 Frontal projection chest x-ray with acute respiratory/pneumonic tularemia (type B) with consolidation at the right side of the lungs with hilar enlargement. Fever and chills started 10 days before. *(Photo from patient care at Region Västerbotten, Sweden.)*

PATHOLOGY

Analyses of tularemia skin lesions of humans infected by skinning of rabbits with tularemia have suggested a scenario with accumulation of lymphocytes, plasma cells, and neutrophils at the infection site during the first week after *F. tularensis* infection of the skin. Microabscesses with neutrophils occur in subcutis along with histiocytes and Langerhans cells in dilated lymph vessels. A skin ulcer and granuloma formation occur only after the first week, along with lymph node engagement. Histopathology findings in other tissues of patients with more advanced tularemia and a more extended disease history include irregular microabscesses and granulomas in samples from liver, spleen, kidney, and lymph nodes. Autopsy samples from the lungs of humans have shown necrotizing pneumonia characterized by abundant fibrin, cellular debris, and neutrophils within alveolar walls and alveolar spaces.

TREATMENT

Specific antimicrobial treatment is highly recommended for tularemia as outlined in **Table 175-2**. The antibiotics of choice are ciprofloxacin, levofloxacin, gentamicin, or doxycycline. Early specific antimicrobial treatment is essential to avoid complications irrespective of targeting type A or type B tularemia. Several commonly used antimicrobials, including all the β-lactam drugs and clindamycin, lack treatment efficacy against *F. tularensis*. Macrolide antibiotics, including erythromycin and azithromycin, show an in vitro effect against *F. tularensis* type A and type B varieties known to be present in the United States, but globally there are very scarce treatment outcome data for these drugs in humans. Generally, macrolide antibiotics are not recommended for tularemia because macrolide-resistant type B strains are common in Europe and Asia, resulting in total lack of treatment effect and overall little experience using these drugs for tularemia treatment. Azithromycin, however, may be an alternative treatment choice for patients with verified type A infection according to recent CDC guidelines.

Streptomycin, classically used to treat tularemia, has been the drug of choice due to its high cure rate in a U.S. case series reported from 1949–1988, but is now considered a less attractive

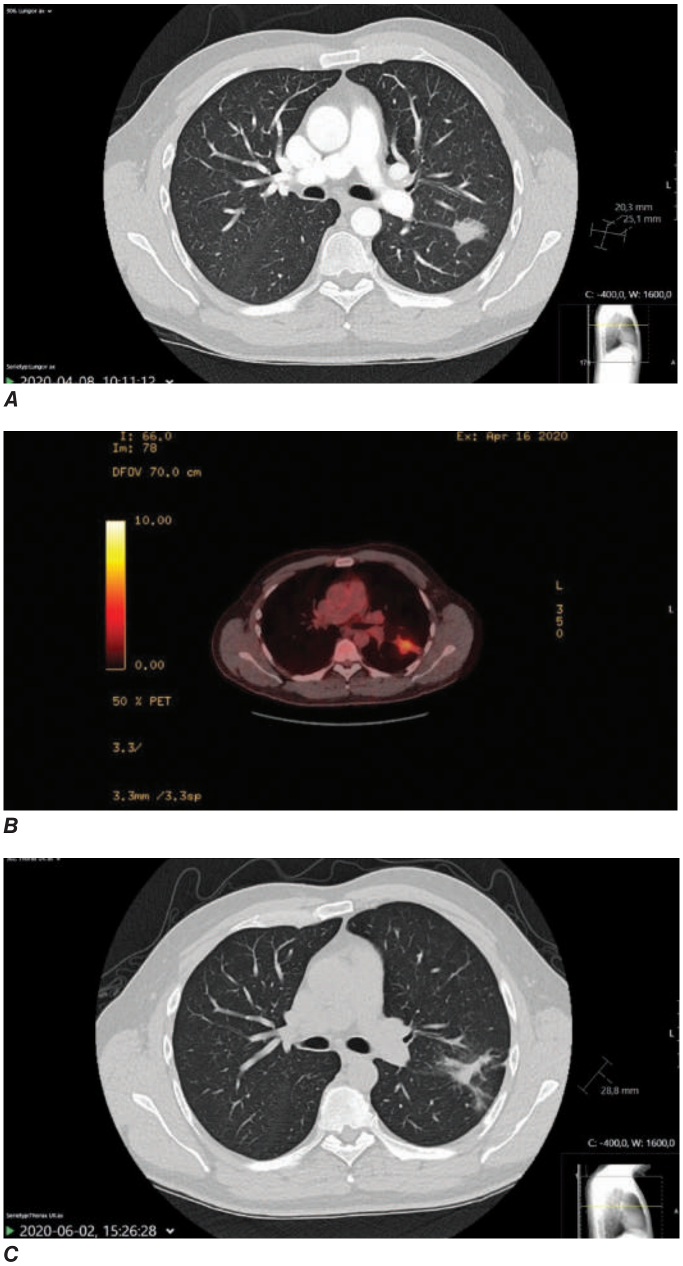

FIGURE 175–5 Dense consolidation in the left lung of patient with a history of weight loss and cough for 3–4 months. ***A.*** CT scan showed dense consolidation (2.5 × 2.5 × 2.0 cm) in the basal left upper lobe extending to the apical part of the lower lobe and increased hilar lymphoid tissue. ***B.*** Eight days later, PET scan showed increased metabolic activity and progress of the consolidation to 5.0 × 4.5 × 3.5 cm. Tissue biopsy revealed granuloma formation in the adjacent pleura and no malignant cells. ***C.*** A tularemia diagnosis was confirmed by serology, the patient was treated with ciprofloxacin, and a CT scan performed 2 months later showed resolution of the consolidation. *(Photos from patient care at Region Västerbotten, Sweden.)*

alternative to gentamicin due to its drawbacks, including the need for intramuscular administration, limited clinical availability, and a higher risk of vestibular and renal toxicity compared with newer aminoglycoside antibiotics. Chloramphenicol is another drug with a higher risk of adverse effects that, for this reason, is seldom used systemically in modern medicine. Chloramphenicol has an in vitro effect against *F. tularensis* and was reportedly effective in treating tularemia meningitis in the past.

TABLE 175-2 Treatment and Postexposure Prophylaxis Recommendations for Adult Patients with Tularemia or with Significant Exposure to an *F. tularensis* Aerosol

ADULTS, INCLUDING PREGNANT WOMEN			
	DRUG	DOSE[a]	DURATION (DAYS)
Preferred choices	Ciprofloxacin	400 mg every 8 hrs IV or 750 mg every 12 hrs PO	10
	Levofloxacin	750 mg every 24 hrs IV or PO	10
	Gentamicin	5–7 mg/kg IV or IM per day, given every 12 hrs or 24 hrs	10
Alternative choice[b]	Doxycycline[c]	200 mg loading dose, then 100 mg every 12 hrs IV or PO	14–21
Postexposure prophylaxis[d]	Doxycycline	100 mg PO twice daily	14
	Ciprofloxacin	500 mg every 12 hrs PO	10

[a]Patients beginning with IV treatment of ciprofloxacin, levofloxacin, or doxycycline can switch to oral administration when clinically indicated. [b]Other alternative choices having less supporting clinical data are high-dose moxifloxacin, ofloxacin, amikacin, tobramycin, or plazomycin for 10 days. In verified type A-infection, azitromycin may be considered based on in vitro susceptibility analysis. [c]Doxycycline may safely be included as a preferred choice for patients with nonsevere disease but is associated with a higher frequency of fever relapse and needs a more extended treatment duration. Tetracycline 500 mg every 6 hrs IV or PO for 14-21 days is equivalent to doxycycline. [d]For low-risk exposures, preparedness for early treatment may suffice using daily "fever watch": the patient monitors temperature with instructions to seek immediate treatment if developing a fever. Postexposure prophylaxis is recommended for high-risk exposures, including performing aerosol-generating procedures with *F. tularensis* cultures outside a biologic safety cabinet or handling infected animals. There are extensive human data showing that doxycycline is effective.

Source: Adapted from J Terriquez, C Nelson (eds): Clinical infectious diseases: Tularemia: Update on treatment and clinical findings. Clin Infect Dis 78:s1, 2024.

COMPLICATIONS

Inflamed lymph nodes can progress into an abscess with local and/or systemic reaction and may occur in about one-third of patients, especially if there were delays in *F. tularensis*–specific antibiotic treatment. An antibiotic treatment delay of >2–3 weeks after disease onset has been statistically associated with lymph node abscess formation. In some cases, the progression to a lymph node abscess may occur despite ongoing appropriate antibiotic treatment initiated late in the disease course. It is unclear if extending the antibiotic treatment duration is of clinical benefit. Culture specimens from an abscess at this disease stage are typically *F. tularensis* negative, but an additional 10–day course of appropriate antibiotic treatment is often used to avoid further disease progression and may be successful. If the clinician considers that there is a clear and immediate risk of spontaneous rupture of the skin or the mucus membranes in the pharynx with suppuration of pus, open surgical drainage or needle aspiration is recommended for treatment. A more invasive procedure with total surgical excision may be needed in more complicated and long-standing cases. Another complication of uncertain genesis occasionally reported by patients includes postinfectious persisting fatigue and disability. Additional complications or disease manifestations of tularemia have been reported, including adult respiratory distress syndrome, myocarditis, pericarditis, endocarditis, meningitis, osteomyelitis, hepatitis, and renal failure.

PROGNOSIS

Prior to the use of antibiotics to treat tularemia, case fatality rates in the United States were typically 5%–15% and could be as high as 60% for patients with respiratory/pneumonic tularemia. Recent surveillance data of tularemia in the United States between 2006 and 2021 show that of 903 patients with illness outcome available, 27 patients (3.0%) died. Culture of *F. tularensis* from a clinical specimen was present in approximately 50% of cases and associated with significantly reduced odds of survival (odds ratio [OR], 0.1; 95% confidence interval [CI], 0.04–0.4). Treatment with at least one high-efficacy antimicrobial drug class (aminoglycoside, fluoroquinolone, or tetracycline) was independently associated with increased odds for survival (adjusted OR, 10.4; 95% CI, 4.4–24.5 after controlling for disease severity). Notably, previous surveillance data from 1964–2004 considering the modern genetic distinction of three epidemiologically important type A genetic varieties in the United States found a 24% fatality outcome for the most lethal type A variety. For the type B *F. tularensis* variety, the same surveillance data from 1964–2004 showed a 7% fatality outcome. Caution in interpreting these latter data is warranted, as they likely are overestimates of mortality due to selected reporting and disease severity bias from the inclusion of culture-positive cases only. Recent analyses of patients with respiratory/pneumonic type B tularemia in Sweden revealed a 1.5% fatality. Outcome analyses considering all clinical forms of tularemia suggest an overall type B tularemia fatality rate of <1% in Sweden. The frequency of disease complications, including lymphadenopathy, among 327 patients was <10%, much lower than previously reported for tularemia, likely because appropriate antibiotics were often given early, at a median of 7 days after disease onset.

PREVENTION

Standard precautions are recommended for infection control and prevention in the clinic. Face protection using goggles or a face shield and a fluid-resistant medical mask should be used in procedures with a risk of splashes from potentially infected body fluids. Only procedures with a very high risk of aerosol formation that may contain live *F. tularensis*, such as the surgical evacuation of pus from an abscess by applying high pressure or autopsy work with rotating tools, warrant respiratory protection with an N95 mask. Tularemia patients do not need to be isolated, given there is no human-to-human transmission. Because *F. tularensis* is a known risk for laboratory-acquired infection, suspected isolates should, at a minimum, be manipulated at BSL-2 using BSL-3 precautions in a biosafety cabinet. In the United States, specialized reference laboratories should always be consulted and involved in attempts for complete identification and antimicrobial susceptibility testing of *F. tularensis*. Vaccination is a potentially important measure for preventing tularemia. However, no U.S. Food and Drug Administration (FDA)-approved *F. tularensis* vaccines exist. The efficacy of the only available vaccine, LVS, is uncertain given concerns about adverse reactions to the vaccine, potential for reversion, unknown correlates of protection, and variable immunogenicity. Prevention strategies for the public include measures to avoid *F. tularensis* exposure. It is advised to avoid tick-infested areas and mosquito bites in tularemia-endemic areas, wear trousers and long-sleeved shirts, use arthropod repellants, and remove attaching ticks promptly. Exposure to and touching dead or sick wild mammals should be avoided. If touching the mammal is necessary, gloves should be worn and hands should be washed thoroughly after removing the gloves. In addition to avoiding direct contact, exposure to potentially *F. tularensis*–contaminated aerosols should also be avoided, e.g., from handling dry hay or grain where carcasses of rodents are found.

GLOBAL CONSIDERATIONS

Due to its past development as a biologic weapon, there is a risk that *F. tularensis* could be used as a biothreat agent. *F. tularensis* is considered an HHS Tier 1 Select Agent in the United States because it presents the greatest risk of deliberate misuse with the most significant potential for mass casualties or devastating effects to the economy, critical infrastructure, or public confidence. From 1932 to 1945, Japanese research units examined the utility of *F. tularensis* as a biologic weapon. After World War II, there were continuous military studies of tularemia. In the United States, there were federally funded biologic warfare programs from 1943 to 1969, with the development of weapons to disseminate type A *F. tularensis* aerosols. By 1973, the entire *F. tularensis* weapon arsenal of the United States had been destroyed. The former Soviet Union, as part of the civilian component of its offensive Biopreparat program, also incorporated *F. tularensis* into weapons. In 1970, the World Health Organization published a report estimating that an aerosol dispersal of 50 kg of virulent type A *F. tularensis* over a metropolitan area with 5 million inhabitants would result in 250,000 incapacitating causalities, including 19,000 deaths.

Efficacy trials of tularemia vaccines in humans against highly virulent type A tularemia are not feasible, and animal studies are judged to be the best option for screening and evaluating new vaccine candidates. There is ongoing research on the matter, including evaluating vaccines under the FDA "Animal Rule," aimed at bridging the outcomes in animals to use in humans.

New tularemia treatment options with promising results, tested in experimentally infected animals, include the use of passive immunization by *F. tularensis*–specific antibodies and a new triazaacenaphthylene antibiotic, gepotidacin. This new antibiotic class blocks bacterial DNA replication via inhibition of DNA gyrase and topoisomerase IV and is claimed to have no cross-resistance with the fluoroquinolones. Importantly, experience from treatment of humans with tularemia is lacking for both these treatments.

Considerations for tularemia in pregnant women include a recommendation to use antibiotic treatment to avoid a risk of loss of pregnancy. Literature is scarce, but the available data suggest that the risk of adverse outcomes, including lymph node complications in the mother, maternal bleeding, spontaneous abortion, intrauterine fetal death, and preterm birth, is lowered with prompt institution of antibiotic treatment. *F. tularensis* findings in transplacental villi and transplacental transmission to the fetus have been reported without treatment.

FURTHER READING

Baskerville A et al: The pathology of untreated and antibiotic-treated experimental tularemia in monkeys. Br J Exp Pathol 59:615, 1978.

Enderlin G et al: Streptomycin and alternative agents for the treatment of tularemia: Review of the literature. Clin Infect Dis 19:42, 1994.

Glynn AR et al: Comparison of experimental respiratory tularemia in three nonhuman primate species. Comp Immunol Microbiol Infect Dis 39:13, 2015.

Harrell JE et al: Current vaccine strategies and novel approaches to combatting *Francisella* infection. Vaccine 42:2171, 2024.

Lamps LW et al: Histologic and molecular diagnosis of tularemia: A potential bioterrorism agent endemic to North America. Mod Pathol 17:489, 2004.

Lindgren H et al: Kinetics of the serological response up to one year after tularemia. Front Cell Infect Microbiol 12:1072703, 2023.

Maurin M et al: Tularemia treatment: Experimental and clinical data. Front Microbiol 14:1348323, 2024.

Roberts LM et al: Identification of early interactions between *Francisella* and the host. Infect Immun 82:2504, 2014.

Tärnvik A (ed): *WHO Guidelines on Tularaemia: Epidemic and Pandemic Alert and Response*. Geneva, Switzerland; World Health Organization Press, 2007.

Telford SR 3rd, Goethert HK: Ecology of *Francisella tularensis*. Annu Rev Entomol 65:351, 2020.

Widerström M et al: Treatment outcome of severe respiratory Type B tularemia using fluoroquinolones. Clin Infect Dis 78:S38, 2024.

Wu HJ et al: Tularemia clinical manifestations, antimicrobial treatment, and outcomes: An analysis of US surveillance data, 2006–2021. Clin Infect Dis 78:S29, 2024.

Yanes H et al: Evaluation of in-house and commercial serological tests for diagnosis of human tularemia. J Clin Microbiol 56:e01440, 2017.

176

Plague and Other *Yersinia* Infections

Michael B. Prentice

PLAGUE

Plague is a systemic zoonosis caused by *Yersinia pestis*. It predominantly affects small rodents in rural areas of Africa, Asia, and the Americas and is usually transmitted to humans by an arthropod vector (the flea). Less often, infection follows contact with animal tissues or respiratory droplets. Plague is an acute febrile illness that is treatable with antimicrobial agents, but mortality rates among untreated patients are high.

Ancient DNA studies have confirmed that both the fourteenth-century Black Death and the sixth-century Plague of Justinian in Europe were due to *Y. pestis* infection. Patients can present with the bubonic, septicemic, or pneumonic form of the disease. Although there is concern about epidemic spread of plague by the respiratory route, this is not the most common route of plague transmission, and established infection-control measures for respiratory plague exist. However, the fatalities associated with plague and the capacity for infection via the respiratory tract mean that *Y. pestis* fits the profile of a potential agent of bioterrorism **(Chap. S4)**. Consequently, measures have been taken to restrict access to the organism, including legislation affecting diagnostic and research procedures in some countries (e.g., the United States).

ETIOLOGY

The genus *Yersinia* comprises gram-negative bacteria of the order *Enterobacterales* (class *Gammaproteobacteria*). Overwhelming taxonomic and paleogenomic evidence shows *Y. pestis* recently evolved from *Yersinia pseudotuberculosis*, an enteric pathogen of mammals spread by the fecal–oral route, and thus has a phenotype distinctly different from that of *Y. pestis*. When grown in vivo or at 37°C, *Y. pestis* forms an amorphous capsule made from a plasmid-specified fimbrial protein, Caf or fraction 1 (F1) antigen, which is an immunodiagnostic marker of infection.

EPIDEMIOLOGY

Human plague generally follows an outbreak in a host rodent population (epizootic). Mass deaths among the rodent primary hosts lead to a search by fleas for new hosts, with consequent incidental infection of other mammals. The precipitating cause for an epizootic may ultimately be related to climate or other environmental factors. The reservoir for *Y. pestis* causing enzootic plague in natural endemic foci between epizootics (i.e., when the organism may be difficult to detect in rodents or fleas) is a topic of ongoing research and may not be the same in all regions. The enzootic/epizootic pattern may be the result of complex dynamic interactions of host rodents that have different plague susceptibilities with different flea vectors; alternatively, an environmental reservoir may be important.

GLOBAL FEATURES

In general, the enzootic areas for plague are lightly populated regions of Africa, Asia, and the Americas **(Fig. 176-1)**. Between January 2013 and December 2018, 2886 cases of plague with a global case-fatality rate of 17% were notified to the World Health Organization (WHO) under the International Health Regulations. More than 97% of these cases were in Africa. The majority of cases in each year were from the island of Madagascar, which in 2017 experienced an urban outbreak of over 2400 clinically suspected cases, with an unusually high proportion of pneumonic plague (78%). A decline in reports from the Democratic Republic of the Congo (DRC) may reflect ongoing conflict in that country affecting surveillance rather than a true decrease. In the past decade, outbreaks of pneumonic plague have been recorded in the DRC, Uganda, Algeria, Madagascar, China, and Peru.

Plague was introduced into North America via the port of San Francisco in 1900 as part of the Third Pandemic, which spread around the world from Hong Kong. The disease is presently enzootic on the western side of the continent from southwestern Canada to Mexico. Most human cases in the United States occur in two regions: "Four Corners" (the junction point of New Mexico, Arizona, Colorado, and Utah), especially northern New Mexico, northern Arizona, and southern Colorado; and further west in California, southern Oregon, and western Nevada (*https://www.cdc.gov/plague/maps-statistics/*). From 1970 to 2020, 496 cases of plague were reported in the United States; in recent decades, incidence has fallen to an average of seven cases per year. Most cases occur from May to October—the time of year when people are outdoors and rodents and their fleas are most plentiful. Prior animal contact occurs in at least 50% of cases, and about 60% of these include domestic animals (usually dogs or cats) that brought wild animals or plague-infected fleas home. Infected cats or dogs may transmit plague directly to humans by the respiratory route. A slightly lower percentage of prior animal contacts involve direct handling of living or dead wild small mammals (e.g., rabbits, hares, prairie dogs) or wild carnivores (e.g., wildcats, coyotes, or mountain lions). In 2014, an outbreak of nonfatal pneumonic plague in Colorado affected four people exposed to an infected dog, with possible interhuman transmission in one case. Prior to this report, the most recent case of person-to-person transmission in the United States occurred in the Los Angeles pneumonic plague outbreak of 1924.

Plague most often develops in areas with poor sanitary conditions and infestations of rats—in particular, the widely distributed roof rat *Rattus rattus* and the brown rat *Rattus norvegicus* (which serves as a laboratory model of plague). Rat control in warehouses and shipping facilities has been recognized as important in preventing the spread of plague since the early twentieth century and features in the current WHO International Health Regulations. Urban rodents acquire infection from wild rodents, and the proximity of the former to humans increases the risk of transmission. The oriental rat flea *Xenopsylla cheopis* is the most efficient vector for transmission of plague among rats and onward to humans in Asia, Africa, and South America.

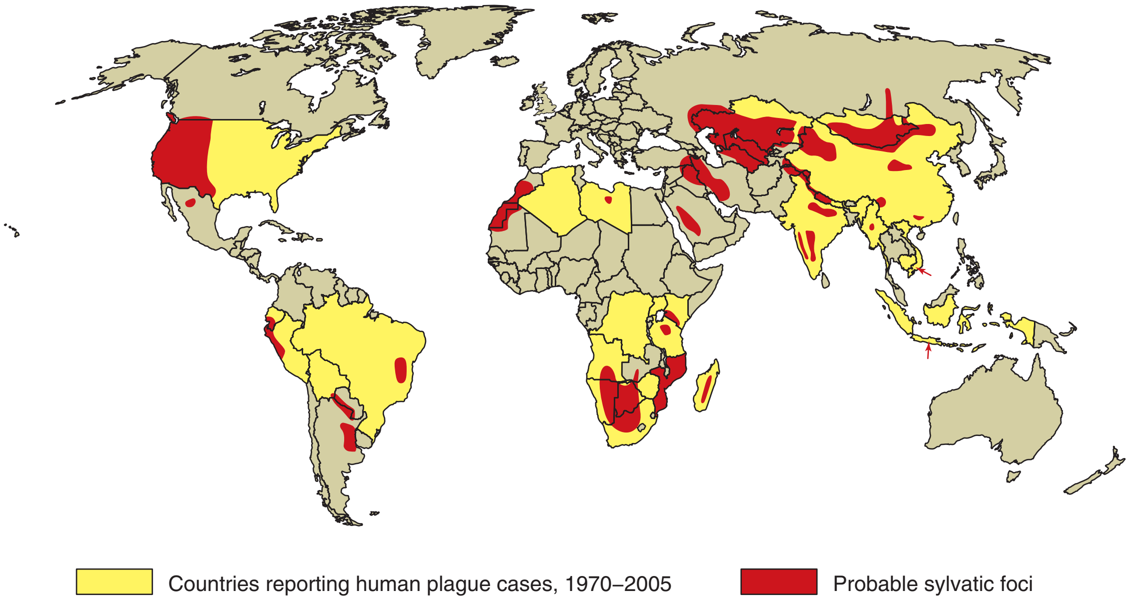

FIGURE 176-1 Approximate global distribution of *Yersinia pestis*. *(Compiled from WHO, CDC, and country sources. Reprinted with permission from DT Dennis, GL Campbell: Plague and other Yersinia infections, in Harrison's Principles of Internal Medicine, 17th ed, AS Fauci et al [eds]. New York, McGraw-Hill, Chap. 152, 2008.)*

Worldwide, bubonic plague is the predominant form reported. A minority of patients (10–20%) present with primary septicemic plague (i.e., systemic *Y. pestis* sepsis with no bubo; see "Clinical Manifestations," below) or pneumonic disease. Primary pneumonic plague is generally the least common and most fatal of the main plague presentations, but, as in the 2017 Madagascar outbreak, it is occasionally predominant. Rare outbreaks of pharyngeal plague following consumption of raw or undercooked camel or goat meat have been reported.

A total of 744 (82%) of the 913 plague cases with clinically documented features (out of 1006 cases reported in total) in the United States from 1900 to 2012 were bubonic disease, 87 (10%) were septicemic disease, and 74 (8%) were pneumonic disease; 6 cases (1%) were pharyngeal. Sixteen percent of cases were fatal in the postantibiotic era from 1942 onward compared with 66% in the period 1900–1941. This century (2000–2018), mortality in the United States has been 8% in patients receiving any antimicrobial and 4% in patients receiving one of the high-efficacy antimicrobials currently recommended as therapy. A systematic review and meta-analysis of the worldwide literature estimated a death rate for treated pneumonic plague of 17%, compared with 98% for untreated pneumonic plague.

PATHOGENESIS

As mentioned earlier, genetic evidence shows *Y. pestis* is a clone derived from the enteric pathogen *Y. pseudotuberculosis* in the recent evolutionary past (7000–50,000 years ago). The change from infection by the fecal–oral route to a two-stage life cycle, with alternate parasitization of arthropod and a wide range of mammalian hosts, occurred as a result of two plasmid gene acquisitions (*pla* on pPCP1/pPst and *ymt* on pFra/pMT1), and the inactivation of a handful of *Y. pseudotuberculosis* genes, in conjunction with preexisting properties of the *Y. pseudotuberculosis* ancestor, including the presence of a virulence plasmid, pYV, and the capacity to cause septicemia. In the arthropod-parasitizing portion of its life cycle, *Y. pestis* multiplies and forms biofilm-embedded aggregates in the flea midgut after ingestion of a blood meal containing bacteria. In some fleas, biofilm-embedded bacteria eventually fill the proventriculus (a valve connecting the esophagus to the midgut) and block normal blood feeding. Both "blocked" fleas and those containing masses of biofilm-embedded *Y. pestis* without complete blockage inoculate *Y. pestis* into each bite site. The ability of *Y. pestis* to colonize and multiply in fleas fed on infected blood from most hosts requires phospholipase D encoded by the *ymt* gene on the pFra (pMT1) plasmid, and biofilm synthesis requires the chromosomal *hms* locus shared with *Y. pseudotuberculosis*. Recently, it has been shown *Y. pestis* does not require an intact *ymt* gene to infect and be transmitted from fleas fed on blood from the brown rat (*Rattus norvegicus*), although this is required for blood from the black rat (*Rattus rattus*). Three *Y. pseudotuberculosis* genes inhibiting biofilm formation or promoting its degradation are inactivated in *Y. pestis*, together with urease (urease activity otherwise causes acute flea gastrointestinal toxicity). Blockage takes days or weeks to come about after initial infection of the flea and is followed by the flea's death. Many flea vectors (including *X. cheopis*) are also able to transmit plague in an early-phase unblocked state for up to a week after feeding, but 10 fleas in this state are required to infect a mammalian host (mass transmission).

Y. pestis disseminates from the site of inoculation in the mammalian host in a process initially dependent on plasminogen activator Pla, which is encoded by the small pPCP1 (pPst) plasmid. This surface protease activates mammalian plasminogen, degrades complement, and adheres to the extracellular matrix component laminin. Pla is essential for the high-level virulence of *Y. pestis* in mice by subcutaneous or intradermal injection (laboratory proxies for fleabites) and for the development of primary pneumonic plague. When actual fleabite inoculation is used in mouse models, the fimbrial capsule-forming protein (Ca1 or fraction 1; F1 antigen) encoded on pFra increases the efficiency of transmission, and plasminogen activator is required for the formation of buboes.

Paleogenomics (sequencing of DNA extracts from teeth of ancient human remains) shows that the 14th-century Black Death and the 6th-century Plague of Justinian were caused by *Y. pestis* and suggests Black Death mortality selected for protective immune response gene variants now associated with autoimmune disease. It has also revealed that *Y. pestis* infection was a common cause of death in Eurasia in the Bronze Age and Neolithic period. Remarkably, the *ymt* gene is absent from the pFra (pMT1) plasmid in *Y. pestis* sequences from some remains >4000 years old, whereas *pla* is present with intact urease and biofilm regulatory genes. This suggests that plague was a common fatal human infection before flea-borne transmission was fully optimized, possibly spread by the pneumonic or gastrointestinal route.

Macrophages, neutrophils, and dendritic cells are all involved in the innate immune response to flea-transmitted *Y. pestis*. The organism is taken up by macrophages but avoids being killed by autophagy and can also survive and replicate in neutrophils. Rapid transport of the bacteria to regional lymph nodes occurs. *Y. pestis* then undergoes extracellular replication with full expression of its antiphagocytic systems: the type III secretion machines and their effectors encoded by pYV as well as the F1 capsule. These factors prevent neutrophil uptake, and the type III secretion effectors also block extrusion of microbicidal DNA by neutrophils and trigger apoptotic cell death. Immune cell targeting follows binding of the N-formylpeptide receptor (FPR1) on phagocytic cells by LcrV, the needle cap protein of the type III secretion system. Overproduction of LcrV also exerts an anti-inflammatory effect, reducing host immune responses. Likewise, *Y. pestis* lipopolysaccharide is modified to minimize stimulation of host Toll-like receptor 4, thereby reducing protective host inflammatory responses during peripheral infection and prolonging host survival with high-grade bacteremia—an effect that probably enhances the pathogen's subsequent transmission by fleabite.

Replication of *Y. pestis* in a regional lymph node results in the local swelling of the lymph node and periglandular region known as a *bubo*. On histology, the node is found to be hemorrhagic or necrotic, with thrombosed blood vessels, and the lymphoid cells and normal architecture are replaced by large numbers of bacteria and fibrin. Periglandular tissues are inflamed and also contain large numbers of bacteria in a serosanguineous, gelatinous exudate.

Continued spread through the lymphatic vessels to contiguous lymph nodes produces second-order primary buboes. Infection is initially contained in the infected regional lymph nodes, although transient bacteremia can be detected. As infection progresses, spread via efferent lymphatics to the thoracic duct produces high-grade bacteremia. Hematogenous spread to the spleen, liver, and secondary buboes follows, with subsequent uncontrolled septicemia leading to death. In some patients, this septicemic phase occurs without obvious prior bubo development or lung disease (septicemic plague). Hematogenous spread to the lungs results in secondary plague pneumonia, with bacteria initially more prominent in the interstitium than in the air spaces (the reverse being the case in primary plague pneumonia). Hematogenous spread to other organs, including the meninges, can occur.

CLINICAL MANIFESTATIONS

Bubonic Plague After an incubation period of 2–6 days, the onset of bubonic plague is sudden and is characterized by fever (>38°C), malaise, myalgia, dizziness, and increasing pain due to progressive lymphadenitis in the regional lymph nodes near the fleabite or other inoculation site. Lymphadenitis manifests as a tense, tender swelling (bubo) that, when palpated, has a boggy consistency with an underlying hard core. Generally, there is one painful and erythematous bubo with surrounding periganglionic edema. The bubo is most commonly inguinal but can also be crural, axillary (Fig. 176-2), cervical, or submaxillary, depending on the site of the bite. Abdominal pain from intraabdominal node involvement can occur without other visible signs. Children are most likely to present with cervical or axillary buboes.

The differential diagnosis includes acute focal lymphadenopathy of other etiologies, such as streptococcal or staphylococcal infection, tularemia, cat-scratch disease, tick typhus, infectious mononucleosis, or lymphatic filariasis. These infections do not progress as rapidly, are not as painful, and are associated with visible cellulitis or ascending lymphangitis—both of which are absent in plague.

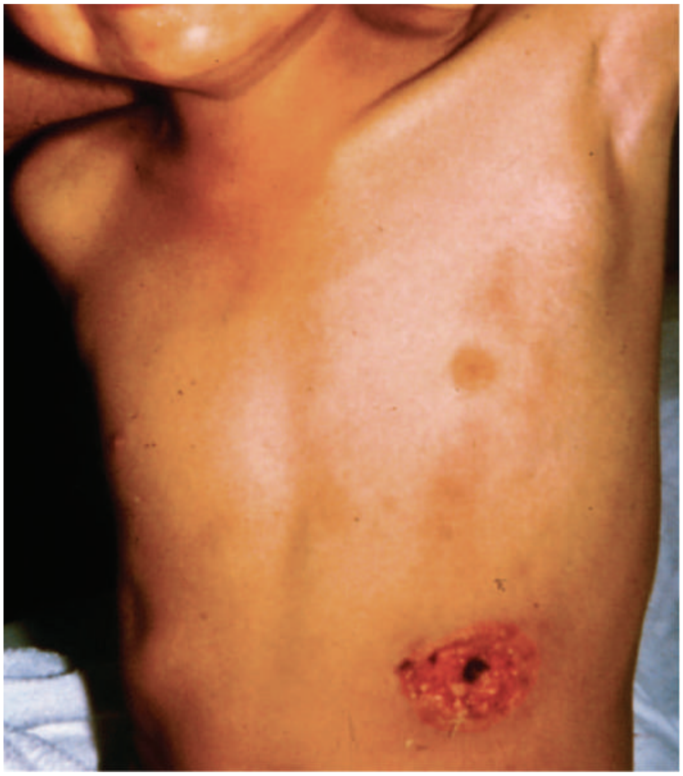

FIGURE 176-2 Plague patient in the southwestern United States with a left axillary bubo and an unusual plague ulcer and eschar at the site of the infective flea bite. *(Reprinted with permission from DT Dennis, GL Campbell: Plague and other Yersinia infections, in Harrison's Principles of Internal Medicine, 17th ed, AS Fauci et al [eds]. New York, McGraw-Hill, Chap. 152, 2008.)*

Without treatment, *Y. pestis* dissemination occurs and causes serious illness, including pneumonia (secondary pneumonic plague) and meningitis. Secondary pneumonic plague can be the source of person-to-person transmission of respiratory infection by productive cough (droplet infection), with the consequent development of primary plague pneumonia. Appropriate treatment of bubonic plague results in fever resolution within 2–5 days, but buboes may remain enlarged for >1 week after initial treatment and can become fluctuant.

Primary Septicemic Plague A minority (10–25%) of infections with *Y. pestis* present as gram-negative septicemia (hypotension, shock) without preceding lymphadenopathy. Septicemic plague occurs in all age groups, but persons >40 years of age are at elevated risk. Some chronic conditions may predispose to septicemic plague: in 2009 in the United States, a fatal laboratory-acquired infection with an attenuated *Y. pestis* strain manifested as septicemic plague in a 60-year-old researcher with diabetes mellitus and undiagnosed hemochromatosis. These conditions also carry an increased risk of septicemia with other pathogenic *Yersinia* species. The term *septicemic plague* can be confusing since most patients with buboes have detectable bacteremia at some stage, with or without systemic signs of sepsis. In laboratory experiments, however, septicemic disease without histologic changes in lymph nodes is seen in a minority of mice infected via fleabites.

Pneumonic Plague Primary pneumonic plague results from inhalation of infectious bacteria in droplets expelled from another person or an animal with primary or secondary plague pneumonia. This syndrome has a short incubation period, averaging from a few hours to 2–3 days (range, 1–7 days), and is characterized by a sudden onset of fever, headache, myalgia, weakness, nausea, vomiting, and dizziness. Respiratory signs—cough, dyspnea, chest pain, and sputum production with hemoptysis—typically arise after 24 h. Progression of initial segmental pneumonitis to lobar pneumonia and then to bilateral lung involvement may occur **(Fig. 176-3)**. The possible release of aerosolized *Y. pestis* bacteria in a bioterrorist attack, manifesting as an outbreak of primary pneumonic plague in nonendemic regions or in an urban setting where plague is rarely seen, has been a source of public health concern. Secondary pneumonic plague is a consequence of bacteremia occurring in ~10–15% of patients with bubonic plague. Bilateral alveolar infiltrates are seen on chest x-ray, and diffuse interstitial pneumonitis with scanty sputum production is typical.

Meningitis Meningeal plague is uncommon, occurring in ≤6% of plague cases reported in the United States. Presentation with headache and fever typically occurs >1 week after the onset of bubonic or septicemic plague and may be associated with suboptimal antimicrobial therapy (delayed therapy, penicillin administration, or low-dose tetracycline treatment) and cervical or axillary buboes.

Pharyngitis Symptomatic plague pharyngitis can follow the consumption of contaminated meat from an animal dying of plague or contact with persons or animals with pneumonic plague. This condition can resemble tonsillitis, with peritonsillar abscess and cervical lymphadenopathy. Asymptomatic pharyngeal carriage of *Y. pestis* can also occur in close contacts of patients with pneumonic plague.

LABORATORY DIAGNOSIS

Because of the scarcity of laboratory facilities in regions where human *Y. pestis* infection is most common, and because of the potential significance of *Y. pestis* isolation in a nonendemic area or an area from which human plague has been absent for many years, the WHO recommends an initial presumptive diagnosis followed by reference laboratory confirmation **(Table 176-1)**. In the United States, comprehensive national diagnostic facilities for plague have been in place since 1999 (Laboratory Response Network for Biological Threats [LRN-B]; *https://emergency.cdc.gov/lrn/index.asp*) to detect possible use of biological terrorism agents, including *Y. pestis*. Routine diagnostic clinical

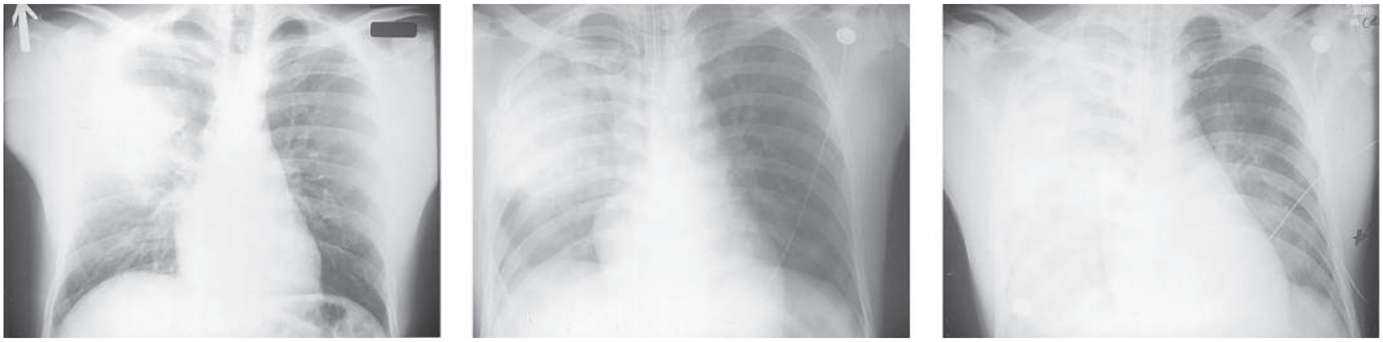

FIGURE 176-3 Sequential chest radiographs of a patient with fatal primary plague pneumonia. ***Left:*** Upright posteroanterior film taken at admission to hospital emergency department on third day of illness, showing segmental consolidation of right upper lobe. ***Center:*** Portable anteroposterior film taken 8 h after admission, showing extension of pneumonia to right middle and right lower lobes. ***Right:*** Portable anteroposterior film taken 13 h after admission (when patient had clinical acute respiratory distress syndrome), showing diffuse infiltration throughout right lung and patchy infiltration of left lower lung. A cavity later developed at the site of initial right-upper-lobe consolidation. *(Reprinted with permission from DT Dennis, GL Campbell: Plague and other Yersinia infections, in Harrison's Principles of Internal Medicine, 17th ed. AS Fauci et al [eds]. New York, McGraw-Hill, Chap. 152, 2008.)*

TABLE 176-1 World Health Organization Case Definitions of Plague

Suspected Case
Clinical presentation suggestive of plague
And
Epidemiological context suggesting possible exposure to plague (exposure to infected humans or animals, or residence in or travel to a known endemic focus within 10 days prior to onset of the disease)
Probable Case
Meeting the definition of a suspected case
***plus* 1 of the following**
• F1 antigen detected in bubo aspirate, sputum, blood, or postmortem tissues by rapid antigen test or direct immunofluorescence
• A single anti-F1 serology without evidence of previous *Yersinia pestis* infection or immunization
• Direct microscopy in a clinical sample: gram-negative coccobacilli that display bipolar staining with Wayson or Wright-Giemsa stain
Confirmed Case
Meeting the definition of a suspected case
***Plus* at least 1 of the following**
• Isolation of *Y. pestis* from a clinical sample (based on appropriate colonial morphology and at least 2 of the following tests positive: phage lysis of cultures at 20–25°C and 37°C; biochemical profile; F1 antigen detection)
• Seroconversion or a fourfold difference in anti-F1 antibody titer in paired serum samples drawn at least 2 weeks apart
• *Y. pestis* DNA positive by species-specific PCR on either clinical sample or culture
Not a case (exclusion of diagnosis)
Meeting the definition of a suspected case
***And* either of the following**
• At least two laboratory tests (rapid antigen test or direct immunofluorescence against F1 antigen, direct microscopy, convalescent serology, culture, PCR) are conducted AND they are negative
• When no confirmatory tests can be performed, 2 negative rapid antigen tests for F1 antigen on 2 clinical specimens collected at 24-h intervals

Abbreviation: PCR, polymerase chain reaction.

Source: Reproduced with permission from E Bertherat, S Jullien. Revision of the international definition of plague cases. Wkly Epidemiol Rec 24:238, 2021.

microbiology laboratories that are included in this network as sentinel-level laboratories use joint protocols from the Centers for Disease Control and Prevention (CDC) and the American Society for Microbiology (*https://asm.org/Articles/Policy/Laboratory-Response-Network-LRN-Sentinel-Level-C*) to identify suspected *Y. pestis* isolates and to refer these specimens to LRN-B reference laboratories for confirmatory tests. *Y. pestis* is designated a "Tier 1 select agent" under the Public Health Security and Bioterrorism Preparedness and Response Act of 2002 and subsequent executive orders; the provisions of this act, the Patriot Act of 2001, and related executive orders apply to all U.S. laboratories and individuals working with *Y. pestis*. Details of the applicable regulations are available from the CDC (*www.selectagents.gov*).

Yersinia species are gram-negative coccobacilli (short rods with rounded ends) 1–3 μm in length and 0.5–0.8 μm in diameter. *Y. pestis* in particular appears bipolar (with a "closed safety pin" appearance) and pleomorphic when stained with a polychromatic stain (Wayson or Wright-Giemsa; **Fig. 176-4**). Its lack of motility distinguishes *Y. pestis* from other *Yersinia* species, which are motile at 25°C and nonmotile at 37°C. Transport medium (e.g., Cary-Blair medium) preserves the viability of *Y. pestis* if transport is delayed.

The appropriate specimens for diagnosis of bubonic, pneumonic, and septicemic plague are bubo aspirate, bronchoalveolar lavage fluid or sputum, and blood, respectively. Culture of postmortem organ biopsy samples also can be diagnostic. A bubo aspirate is obtained by injection of 1 mL of sterile normal saline into a bubo under local anesthetic and aspiration of a small amount of (usually blood-stained) fluid. The WHO has provided guidance on how to aspirate buboes and collect sputum from patients with suspected pneumonic plague (*https://www.who.int/emergencies/outbreak-toolkit/disease-outbreak-toolboxes/plague-outbreak-toolbox*; *https://www.who.int/publications/i/item/operational-guidelines-on-plague-surveillance-diagnosis-prevention-and-control*). Gram's staining of these specimens may reveal gram-negative rods, which are shown by Wayson or Wright-Giemsa staining to be bipolar. These bacteria may even be visible in direct blood smears in septicemic plague (Fig. 176-4); this finding indicates very high numbers of circulating bacteria and a poor prognosis.

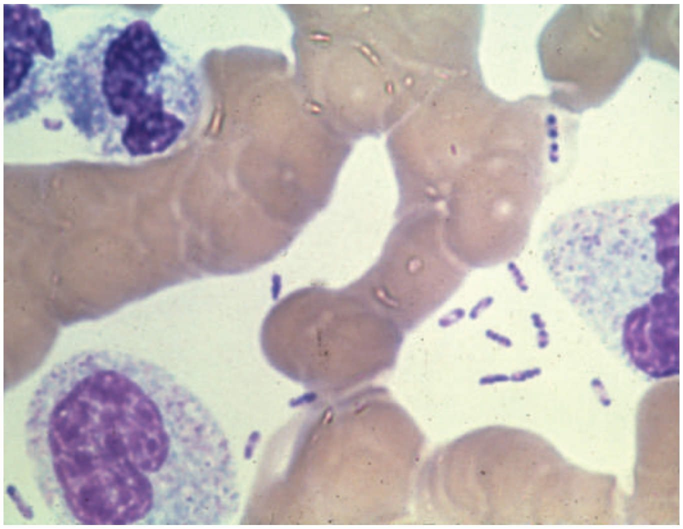

FIGURE 176-4 Peripheral-blood smear from a patient with fatal plague septicemia and shock, showing characteristic bipolar-staining *Yersinia pestis* bacilli (Wright's stain, oil immersion). *(Reprinted with permission from DT Dennis, GL Campbell: Plague and other Yersinia infections, in Harrison's Principles of Internal Medicine, 17th ed, AS Fauci et al [eds]. New York, McGraw-Hill, Chap. 152, 2008.)*

Y. pestis grows on nutrient agar and other standard laboratory media but forms smaller colonies than do other Enterobacteriaceae. Specimens should be inoculated onto nutrient-rich media such as sheep blood agar (SBA), into nutrient-rich broth such as brain-heart infusion broth, and onto selective agar such as MacConkey or eosin methylene blue (EMB) agar. *Yersinia*-specific CIN (cefsulodin, triclosan [Irgasan], novobiocin) agar can be useful for culture of contaminated specimens, such as sputum. Blood should be cultured in a standard blood culture system. The optimal growth temperature is <37°C (25–29°C), with pinpoint colonies only on SBA at 24 h. Slower growth occurs at 37°C. *Y. pestis* is oxidase-negative, catalase-positive, urease-negative, indole-negative, and lactose-negative. Automated biochemical or mass spectrometry identification systems can misidentify *Y. pestis* as *Y. pseudotuberculosis* or other bacterial species.

Reference laboratory tests for definitive identification of isolates include direct immunofluorescence for F1 antigen; polymerase chain reaction (PCR) for specific *Y. pestis* targets (see below); and specific bacteriophage lysis. PCR can also be applied to diagnostic specimens, as can direct immunofluorescence for F1 antigen (produced in large amounts by *Y. pestis*) by slide microscopy. An immunochromatographic test strip for F1 antigen detection by monoclonal antibodies in clinical specimens (rapid diagnostic test based on the F1 antigen [F1RDT]) has been devised in Madagascar. This method is effective for both laboratory and near-patient use on bubo aspirates and sputum and is now widely used in endemic countries. A similar test strip for Pla antigen has been developed and could be used to detect wild-type or engineered F1-negative virulent strains. Recent clinical experience, including the 2017 Madagascar outbreak, found F1RDT to be at least as sensitive as laboratory culture for diagnosis of bubonic or pneumonic plague. However, there was a low diagnostic yield of culture in this outbreak, possibly due to widespread prehospital use of antimicrobials. The specificity of F1RDT for pneumonic plague diagnosed by culture was 71%; thus, while a positive sputum test is useful in determining whether a symptomatic patient in an endemic area is a probable case requiring treatment, additional laboratory tests are required to confirm a diagnosis of pneumonic plague. F1RDT had a high negative predictive value in the limited studies available; thus, when other laboratory tests are unavailable, two negative F1RDT tests on clinical specimens 1 day apart can help exclude a diagnosis of plague in a symptomatic patient. This use of F1RDT has been incorporated in the revised WHO international definition of plague introduced following the 2017 Madagascar outbreak.

The WHO requires diagnostic PCR or real-time PCR (RT-PCR) assays to show positivity for at least two different targets specific for *Y. pestis* from a short list comprising the plasmid-based *caf1* (specifying the F1 antigen) and *pla* (specifying plasminogen activator) genes and three other chromosomal genes. *Y. pestis* is included in the U.S. Food and Drug Administration (FDA)-authorized Biofire FilmArray Next Generation Diagnostic System (NGDS) Warrior Panel for use with the FilmArray 2.0 system (Biomérieux) as a medical diagnostic device suitable for whole blood (ethylenediaminetetraacetic acid [EDTA]), blood cultures, and sputum specimens used by U.S. Department of Defense laboratories and laboratories in the CDC-managed LRN-B network. It is also one of 14 viral, bacterial, and protozoan pathogens diagnosable from blood samples with the BioFire Global Fever Special Pathogens Panel recently authorized by the FDA. Detailed phylogeographic DNA sequence data based on culture collections have been accumulated to trace plague evolution, and this approach could be adapted in the future to real-time clinical plague epidemiology.

In the absence of other positive laboratory diagnostic tests, a retrospective serologic diagnosis may be made on the basis of rising titers of hemagglutinating antibody to F1 antigen. Enzyme-linked immunosorbent assays (ELISAs) for IgG and IgM antibodies to F1 antigen are also available.

The white blood cell (WBC) count is generally raised (to 10,000–20,000/μL) in plague, with neutrophilic leukocytosis and a left shift (numerous immature neutrophils); in some cases, however, the WBC count is normal or leukopenia develops. WBC counts are occasionally very high, especially in children (>100,000/μL). Levels of fibrinogen degradation products are elevated in a majority of patients, but platelet counts are usually normal or low-normal. However, disseminated intravascular coagulation, with low platelet counts, prolonged prothrombin times, reduced fibrinogen, and elevated fibrinogen degradation product levels, occurs in a significant minority of patients.

TREATMENT

Plague

Guidelines for the treatment of plague are given in **Table 176-2**. A 10- to 14-day course of antimicrobial therapy (or a course continued until 2 days after fever subsides) is recommended. Streptomycin has historically been the parenteral treatment of choice for plague and is approved for this indication by the FDA. Although not yet approved by the FDA for plague, gentamicin has proved safe and effective in clinical trials in Tanzania and Madagascar and in retrospective reviewed cases in the United States. In view of streptomycin's adverse-reaction profile and limited availability, some experts now recommend gentamicin over streptomycin. The FDA has approved levofloxacin, moxifloxacin, and ciprofloxacin for prophylaxis and treatment of plague (including septicemic and pneumonic plague) under a regulatory approach based on animal studies alone, known as the Animal Rule. Levofloxacin has more efficacy than ciprofloxacin in postexposure prophylaxis of inhalational anthrax in animal models and has also received FDA approval for this indication **(Chap. S4)**; thus, it is a suitable agent for prophylaxis against two diseases in possible bioterrorism exposures. The WHO issued new guidelines in 2021 adding these three fluoroquinolones to the recommended list of first-line medicines for treating bubonic, pneumonic, or septicemic plague. It also recommended moxifloxacin and ofloxacin for plague meningitis, and ciprofloxacin for postexposure prophylaxis.

TABLE 176-2 Guidelines for the Treatment of Plague

DRUG	DAILY DOSE	DOSING INTERVAL, h	ROUTE
Gentamicin			
Adult	5 mg/kg[a]	24	IM/IV
Child	4.5–7.5 mg/kg[a]	24	IM/IV
Streptomycin			
Adult	2 g	12	IM
Child	30 mg/kg (maximum 1 g per dose)	12	IM
Levofloxacin			
Adult (child >50 kg)	750 (500–750) mg	24	PO/IV
Child <50 kg and ≥6 months of age	16 mg/kg (maximum, 250 mg/dose)	12	PO/IV
Ciprofloxacin			
Adult	1500 mg	12	PO
	1200 mg	8	IV
Child	30–45 mg/kg (maximum, 500 mg/dose)	8–12	PO
	20–30 mg/kg (maximum, 400 mg/dose)	8–12	IV
Moxifloxacin			
Adult	400 mg	24	PO/IV
Doxycycline			
Adult and child ≥45 kg	200 mg (200 mg loading dose)	12	PO/IV
Child <45 kg	4.4 mg/kg (maximum, 100 mg/dose, 4.4 mg/kg loading dose)	12	PO/IV
Tetracycline			
Adult	2 g	6	PO/IV
Child >8 y	40–50 mg/kg	6	PO/IV
Chloramphenicol			
Adult	50–100 mg/kg	6	PO/IV
Child >2 y	50–100 mg/kg (maximum, 4 g)	6	PO/IV

[a] Aminoglycoside dose is adjusted with impaired renal function. No trial data have been published for once-daily gentamicin therapy for plague in adults or children, but this regimen is efficacious in gram-negative sepsis of other etiologies and has been successful in a recent outbreak of pneumonic plague in the Democratic Republic of the Congo. Neonates (up to 1 week of age) and premature infants should receive gentamicin at 4 mg/kg IV once daily.

Source: CA Nelson et al: Antimicrobial treatment and prophylaxis of plague: Recommendations for naturally acquired infections and bioterrorism response. MMWR Recomm Rep 70(No. RR-3):1, 2021. Provides detailed guidelines on recommended regimens for pneumonic versus bubonic plague, plague meningitis, treatment during pregnancy and lactation, and neonatal infection. Recommends dual therapy with two different classes of antimicrobials for initial treatment of patients with severe pneumonic or septicemic plague and patients infected after intentional release of *Yersinia pestis.*

While systemic chloramphenicol therapy is available in the resource-poor countries primarily affected by plague, it is less likely to be available or used in high-income countries because of its adverse effect profile. Tetracyclines are also effective and can be given by mouth but are not generally recommended for children age <7 years because of tooth discoloration. Doxycycline is the tetracycline of choice; at an oral dosage of 100 mg twice daily, this drug was as effective as intramuscular gentamicin (2.5 mg/kg twice daily) in a trial in Tanzania. There is recent evidence that doxycycline does not cause dental staining in children because it binds calcium less readily than other tetracyclines.

Although *Y. pestis* is sensitive to β-lactam drugs in vitro and these drugs have been efficacious against plague in some animal models, the response to penicillins has been poor in some clinical cases; thus β-lactams and macrolides are not generally recommended as first-line therapy. Chloramphenicol, alone or in combination, is recommended for some focal complications of plague (e.g., meningitis, endophthalmitis, myocarditis) because of its tissue penetration properties. Fluoroquinolones, effective in vitro and in animal models, are recommended in guidelines for possible bioterrorism-associated pneumonic plague and are increasingly used in plague therapy.

PREVENTION

In endemic areas, the control of plague in humans is based on reduction of the likelihood of being bitten by infected fleas or exposed to infected droplets from either humans or animals with plague pneumonia. In the United States, residence and outdoor activity or contact with wild or pet animals in rural areas of western states where epizootics occur are the main risk factors for infection. To assess potential risks to humans in specific areas, surveillance for *Y. pestis* infection among animal plague hosts and vectors is carried out regularly as well as in response to observed animal die-offs. Personal protective measures include avoidance of areas where a plague epizootic has been identified and publicized (e.g., by warning signs or closure of campsites). Sick or dead animals should not be handled by the general public. Hunters, zoologists, and pet owners should wear gloves if handling wild-animal carcasses in endemic areas. General measures to avoid rodent fleabite during outdoor activity are appropriate and include the use of insect repellent, insecticide, and protective clothing. General measures to reduce peridomestic and occupational human contact with rodents are advised and include rodent-proofing of buildings and food-waste stores and removal of potential rodent habitats (e.g., woodpiles and junk heaps). Flea control by insecticide treatment of wild rodents is an effective means of minimizing human contact with plague if an epizootic is identified in an area close to human habitation. Any attempt to reduce rodent numbers must be preceded by flea suppression to reduce the migration of infected fleas to human hosts. An oral F1-V subunit vaccine using raccoon poxvirus (RCN) as a vector (sylvatic plague vaccine) is partially protective against plague when administered to wild prairie dogs in field trials and may in the future provide a means of reducing the risk of human exposure to *Y. pestis*.

Patients in whom pneumonic plague is suspected should be managed in isolation (with negative pressure, if available), with droplet precautions observed until pneumonia is excluded or effective antimicrobial therapy has been given for 48 h. Review of the literature published before the advent of antimicrobial agents suggests that the main infective risk is posed by patients in the final stages of disease who are coughing up sputum with plentiful visible blood and/or pus. Cotton and gauze masks were protective in these circumstances. Current surgical masks capable of barrier protection against droplets, including large respiratory particles, are probably protective, but the differential diagnosis of fever and hemoptysis in plague-endemic areas includes small airborne particle-transmitted infections such as tuberculosis. In addition, WHO guidance recommends that personal protective equipment for potential aerosol-generating procedures (e.g., collection of respiratory samples from patients with suspected or confirmed plague) or handling the remains of someone who was infected with plague should include a fit-tested N95 face mask, a gown, gloves, and a face shield or goggles.

Antimicrobial Prophylaxis Postexposure antimicrobial prophylaxis lasting 7 days is recommended following household, hospital, or other close contact with persons with untreated pneumonic plague. (*Close contact* is defined as contact with a patient at <2 m.) In animal aerosol-infection studies, levofloxacin and ciprofloxacin are associated with higher survival rates than doxycycline **(Table 176-3)**.

TABLE 176-3 Guidelines for Plague Prophylaxis

DRUG	DAILY DOSE	DOSING INTERVAL, h	ROUTE
Doxycycline			
Adult	200 mg	12 or 24	PO
Child ≥8 y	≥45 kg: adult dose	12	PO
	≤45 kg: 4.4 mg/kg (maximum, 200 mg)	12	PO
Tetracycline			
Adult	2 g	6 or 12	PO
Child ≥8 y	40 mg/kg (maximum 500 mg dose)	6 or 12	PO
Levofloxacin			
Adult and child >50 kg	500–750 mg	24	PO
Child <50 kg and ≥6 months of age	16 mg/kg (maximum, 250 mg/dose)	12	PO
Ciprofloxacin			
Adult	1–1.5 g	12	PO
Child	30 mg/kg (maximum 750 mg dose)	12	PO

Source: TV Inglesby et al: Plague as a biological weapon: Medical and public health management. Working Group on Civilian Biodefense. JAMA 283:2281, 2000; *https://www.cdc.gov/plague/healthcare/clinicians.html*; CA Nelson et al: Antimicrobial treatment and prophylaxis of plague: Recommendations for naturally acquired infections and bioterrorism response. MMWR Recomm Rep 70(No. RR-3):1, 2021.

Immunization Studies with candidate plague vaccines in animal models show that neutralizing antibody provides protection against exposure but that cell-mediated immunity is critical for protection and clearance of *Y. pestis* from the host. A killed whole-cell vaccine used in humans required multiple doses, caused significant local and systemic reactions, and was not protective against pneumonic plague; this vaccine is not currently available. A live attenuated vaccine based on strain EV76 is still used in countries of the former Soviet Union and China but has significant side effects. Different subunit vaccines devised by governmental agencies in the United States, United Kingdom, and China all comprising recombinant F1 (rF1) and various recombinant V (rV) proteins produced in *Escherichia coli*, combined either as a fusion protein or as a mixture, purified, and adsorbed to aluminum hydroxide for injection are close to licensing. This combination protects mice and various nonhuman primates in laboratory models of bubonic and pneumonic plague and has been evaluated in phase 2 clinical trials. Prelicensing field-efficacy studies (phase 3 trials) are difficult to devise because of plague epidemiology. In the United States, the FDA will assess plague vaccines for human use under the Animal Rule, using efficacy data from animal studies and antibodies and other correlates of immunity from human vaccinees (*https://www.fda.gov/emergency-preparedness-and-response/mcm-regulatory-science/animal-rule-information*), and the rF1-V subunit vaccine has orphan drug status. The WHO has produced a Target Product Profile (TPP) for phase 3 trial design and prioritization of the vaccine candidates. Candidate vaccines include protein subunit, live-attenuated, and bacterial, viral, and bacteriophage vectors, DNA, and mRNA vaccines. Antigens other than F1 and V are being investigated because of the recovery of F1-negative *Y. pestis* strains from natural sources and the observation that F1 antigen is not required for virulence in primate models of pneumonic plague.

YERSINIOSIS

Yersiniosis is a zoonotic infection with an enteropathogenic *Yersinia* species, usually *Y. enterocolitica* or *Y. pseudotuberculosis*. The usual hosts for these organisms are pigs and other wild and domestic animals; humans are usually infected by the oral route, and outbreaks from contaminated food occur. Yersiniosis is most common in childhood and in colder climates. Patients present with abdominal pain and sometimes with diarrhea (which may not occur in up to 50% of cases). *Y. enterocolitica* is more closely associated with terminal ileitis and *Y. pseudotuberculosis* with mesenteric adenitis, but both organisms may cause mesenteric adenitis and symptoms of abdominal pain and tenderness that result in pseudoappendicitis, with the surgical removal of a normal appendix. Diagnosis was historically based on culture of the organism or convalescent serology, but some proprietary multiplex PCR systems for gastrointestinal infection diagnosis now include *Y. enterocolitica* (but not *Y. pseudotuberculosis*). *Y. pseudotuberculosis* and some rarer strains of *Y. enterocolitica* are especially likely to cause systemic infection, which is also more likely in patients with diabetes or iron overload. Systemic sepsis is treatable with antimicrobial agents, but postinfective arthropathy responds poorly to such therapy. Over twenty other *Yersinia* species lacking the virulence plasmid pYV common to *Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica* are now recognized, primarily from genome sequencing. These are, at most, opportunistic pathogens of humans (including *Y. aldovae*, *Y. aleksiciae*, *Y. bercovieri*, *Y. entomophaga*, *Y. frederiksenii*, *Y. hibernica*, *Y. intermedia*, *Y. kristensenii*, *Y. massiliensis*, *Y. mollaretii*, *Y. nurmii*, *Y. pekkanenii*, *Y. rohdei*, *Y. similis*, *Y. ruckeri*, and *Y. wautersii*). Molecular phylogeny shows that *Y. enterocolitica* is more distantly related to *Y. pseudotuberculosis* than these other *Yersinia* species, and the similar virulence plasmid they share has probably been acquired independently by at least one of the two since the species diverged.

EPIDEMIOLOGY

Y. enterocolitica *Y. enterocolitica* is found worldwide and has been isolated from a wide variety of wild and domestic animals and environmental samples, including samples of food and water. In vitro, *Y. enterocolitica* is resistant to predation by the protozoon *Acanthamoeba castellani* and can survive inside it, suggesting a possible mode of environmental persistence. Strains are classically differentiated by biochemical reactions (biovar or biotype) combined with serogroup and, increasingly, by whole genome sequence data used for core genome multilocus sequence typing (cgMLST) (see "Laboratory Diagnosis," below).

Yersiniosis, >99% due to *Y. enterocolitica*, remains the third most common bacterial food-borne zoonosis reported in Europe, especially prevalent in Germany and Scandinavia. The incidence is highest among children; children <4 years of age are more likely to present with diarrhea than are older children. Abdominal pain with mesenteric adenitis and terminal ileitis is more prominent among older children and adults. Septicemia is more likely in patients with preexisting conditions such as diabetes mellitus, liver disease, any condition involving iron overload (including thalassemia and hemochromatosis), advanced age, malignancy, or HIV/AIDS. As in enteritis of other bacterial etiologies, postinfective complications such as reactive arthritis occur mainly in individuals who are HLA-B27 positive. Erythema nodosum **(Fig. A1-39)** following *Yersinia* infection is not associated with HLA-B27 and is more common among women than among men.

Consumption or preparation of raw pork products (such as chitterlings) and some processed pork products is strongly linked with infection because a high percentage of pigs carry pathogenic *Y. enterocolitica* strains. Outbreaks of *Y. enterocolitica* infection have been associated with consumption of milk (pasteurized, unpasteurized, and chocolate-flavored) and various ready-to-eat vegetables including fresh spinach that were processed or washed with water. Person-to-person transmission is suspected in a few cases (e.g., in nosocomial and familial outbreaks) but is much less likely with *Y. enterocolitica* than with other causes of gastrointestinal infection, such as *Salmonella*. A multivariate analysis indicates that contact with companion animals is a risk factor for *Y. enterocolitica* infection among children in Sweden, and low-level colonization of dogs and cats with *Y. enterocolitica* has been reported. Transfusion-associated septicemia due to *Y. enterocolitica*, while recognized as a very rare but frequently fatal event for >30 years, has been difficult to eradicate.

Y. pseudotuberculosis *Y. pseudotuberculosis* is much less frequently reported as a cause of human disease than *Y. enterocolitica*, and infection with *Y. pseudotuberculosis* is more likely to present as fever and abdominal pain due to mesenteric lymphadenitis and be identified from a blood culture isolate. This organism is associated with wild mammals (rodents, rabbits, and deer), birds, and domestic pigs. Although outbreaks are generally rare, several have recently occurred associated by food culture or dietary history with consumption of lettuce or raw carrots (Finland, New Zealand), tomatoes (France), or unpasteurized milk (Finland). Strains have historically been differentiated by combined biochemical reactions (biovar) and serogroup. cgMLST is now used for investigation of both *Y. enterocolitica* and *Y. pseudotuberculosis* outbreaks in several countries.

PATHOGENESIS

The usual route of infection is oral. Studies with both *Y. enterocolitica* and *Y. pseudotuberculosis* in animal models suggest that initial replication in the small intestine is followed by invasion of Peyer's patches of the distal ileum via M cells, with onward spread to mesenteric lymph nodes. The liver and spleen can also be involved after oral infection. The characteristic histologic appearance of enteropathogenic *Yersinia* after invasion of host tissues is as extracellular microabscesses surrounded by an epithelioid granulomatous lesion.

Experiments involving oral infection of mice with tagged *Y. enterocolitica* show that only a very small proportion of bacteria in the gut invade tissues. Individual bacterial clones from an orally inoculated pool give rise to each microabscess in a Peyer's patch, and the host restricts the invasion of previously infected Peyer's patches. A prior model positing progressive bacterial spread from Peyer's patches and mesenteric lymph nodes to the liver and spleen appears to be inaccurate: spread of *Y. pseudotuberculosis* and *Y. enterocolitica* to the liver and spleen of mice occurs independently of regional lymph node colonization and in mice lacking Peyer's patches.

Invasion requires the expression of several nonfimbrial adhesins, such as invasin (Inv) and—in *Y. pseudotuberculosis*—*Yersinia* adhesin A (YadA). Inv interacts directly with β1 integrins, which are expressed on the apical surfaces of M cells but not enterocytes. YadA of *Y. pseudotuberculosis* interacts with extracellular matrix proteins such as collagen and fibronectin to facilitate host cell integrin association and invasion. YadA of *Y. enterocolitica* lacks a crucial N-terminal region and binds collagen and laminin but not fibronectin and does not cause invasion. Inv is chromosomally encoded, whereas YadA is encoded on the virulence plasmid pYV. YadA also helps to confer serum resistance in *Y. enterocolitica* by binding host complement regulators such as factor H and C4-binding protein. Another chromosomal gene, *ail* (attachment and invasion locus), encodes the extracellular protein Ail, which is the main factor conferring serum resistance in *Y. pseudotuberculosis* by binding these complement regulators.

By binding to host cell surfaces, YadA allows targeting of immune effector cells by the pYV plasmid–encoded type III secretion system (injectisome). As a consequence, the host's innate immune response is altered; toxins (*Yersinia* outer proteins, or Yops) are injected into host macrophages, neutrophils, and dendritic cells, affecting signal transduction pathways, resulting in reduced phagocytosis and inhibited production of reactive oxygen species by neutrophils, and triggering apoptosis of macrophages. Other factors functional in invasive disease include yersiniabactin (Ybt), a siderophore produced by some strains of *Y. pseudotuberculosis* and *Y. enterocolitica* as well as other Enterobacterales. Ybt allows bacteria to access iron from saturated lactoferrin during infection and reduces production of reactive oxygen species by innate immune effector cells, thereby decreasing bacterial killing. *Y. pseudotuberculosis* and *Y. pestis* make other siderophores apart from Ybt.

CLINICAL MANIFESTATIONS

Self-limiting diarrhea is the most common reported presentation in infection with pathogenic *Y. enterocolitica*, especially in children <4 years of age, who form the single largest group in most case series. Blood may be detected in diarrheal stool. Older children and adults are more likely than younger children to present with abdominal pain, which can be localized to the right iliac fossa—a situation that often leads to laparotomy for presumed appendicitis (pseudoappendicitis). Appendectomy is not indicated for *Yersinia* infection causing pseudoappendicitis. Thickening of the terminal ileum and cecum is seen on endoscopy and ultrasound, with elevated round or oval lesions that may overlie Peyer's patches. Mesenteric lymph nodes are enlarged. Ulcerations of the mucosa are noted on endoscopy. Gastrointestinal complications include granulomatous appendicitis, a chronic inflammatory condition affecting the appendix that is responsible for ≤2% of cases of appendicitis; *Yersinia* is involved in a minority of cases. *Y. enterocolitica* infection can present as acute pharyngitis with or without other gastrointestinal symptoms. Fatal *Y. enterocolitica* pharyngitis has been recorded. Mycotic aneurysm can follow *Y. enterocolitica* bacteremia, as can focal infection (abscess) in many other sites and body compartments (liver, spleen, kidney, bone, meninges, endocardium).

Y. pseudotuberculosis infection is more likely to present as abdominal pain and fever than as diarrhea. A superantigenic toxin—

Y. pseudotuberculosis mitogen (YPM)—is produced by strains seen in eastern Russia in association with Far Eastern scarlet-like fever (FESLF), a childhood illness with desquamating rash, arthralgia, and toxic shock. A similar illness is recognized in Japan (Izumi fever) and Korea. Similarities have been noted with Kawasaki disease, the idiopathic acute systematic vasculitis of childhood. There is an epidemiologic link between exposure of populations to superantigen-positive *Y. pseudotuberculosis* and an elevated incidence of Kawasaki disease.

Y. enterocolitica or *Y. pseudotuberculosis* septicemia presents as a severe illness with fever and leukocytosis, often without localizing features, and is significantly associated with predisposing conditions such as diabetes mellitus, liver disease, and iron overload. Hemochromatosis combines several of these risk factors. Administration of iron chelators like desferrioxamine, which provide iron accessible to *Yersinia* (and have an inhibitory effect on neutrophil function), may result in *Yersinia* septicemia in patients with iron overload who presumably have an otherwise mild gastrointestinal infection. HIV/AIDS has been associated with *Y. pseudotuberculosis* septicemia. The unusual phenomenon of transfusion-associated septicemia is linked to the ability of *Y. enterocolitica* to multiply at refrigerator temperature (psychrotrophy). Typically, the transfused unit has been stored for >20 days, and it is believed that small numbers of yersiniae from an apparently healthy donor with subclinical bacteremia are amplified to very high numbers by growth inside the bag at ≤4°C, with consequent septic shock after transfusion. Complete prevention of this very rare event (one case in several million transfused units in countries such as the United States and France) without unacceptable restriction in the blood supply has not yet been devised.

POSTINFECTIVE PHENOMENA

As in other invasive intestinal infections (salmonellosis, shigellosis), reactive arthritis (articular arthritis of multiple joints developing within 2–4 weeks of a preceding infection) occurs as a result of autoimmune activity initiated by the deposition of bacterial components (not viable bacteria) in joints in combination with the immune response to invading bacteria. The majority of individuals affected by reactive arthritis due to *Yersinia* are HLA-B27 positive. Myocarditis with electrocardiographic ST-segment abnormalities may occur with *Yersinia*-associated reactive arthritis. Most *Yersinia*-associated cases follow *Y. enterocolitica* infection (presumably because it is more common than infection with other species), but *Y. pseudotuberculosis*–associated reactive arthritis is also well documented in Finland, where sporadic and outbreak infections with *Y. pseudotuberculosis* are more common than in other countries. Of infected individuals identified in a recent *Y. pseudotuberculosis* serotype O:3 outbreak in Finland, 12% developed reactive arthritis affecting the small joints of the hands and feet, knees, ankles, and shoulders and lasting >6 months in most cases. Erythema nodosum **(Fig. A1-39)** occurs after *Yersinia* infection (more commonly in women) with no evidence of HLA-B27 linkage.

There is a long-standing association between antithyroid and anti-*Yersinia* antibodies. Antibody evidence of prior *Y. enterocolitica* infection in Graves' disease and increased levels of antithyroid antibody in patients with *Y. enterocolitica* antibodies were first noted in the 1970s. *Y. enterocolitica* contains a thyroid-stimulating hormone (TSH)–binding site that is recognized by antibodies to TSH from Graves' disease patients. Raised titers of antibodies to *Y. enterocolitica* whole cells and Yops have been found in some series of Graves' disease patients but not in others. It remains unclear whether this cross-reactivity is significant in the etiology of Graves' disease.

LABORATORY DIAGNOSIS

Standard laboratory culture methods can be used to isolate enteropathogenic *Yersinia* species from sterile samples, including blood and cerebrospinal fluid. Culture on specific selective media (CIN agar), with or without pre-enrichment in broth or phosphate-buffered saline at either 4°C or 16°C, is the basis of most schema for isolation of yersiniae from stool or other nonsterile samples. Outside known high-incidence areas, specific culture may only be carried out by laboratories on request, or if a multiplex PCR screen detects *Y. enterocolitica*–specific DNA in feces. Several CE-marked, FDA-approved kits for enteric pathogens now offer *Y. enterocolitica* detection (the precise assay targets are not disclosed), and their use has increased detection of *Y. enterocolitica*. These kits generally do not detect *Y. pseudotuberculosis*. A standard for PCR detection of pathogenic *Y. enterocolitica* and *Y. pseudotuberculosis* in food samples is available from the International Organization for Standardization.

Matrix-assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry systems can speciate isolates of *Y. enterocolitica* and *Y. pseudotuberculosis* (but cannot separate *Y. pestis* from *Y. pseudotuberculosis*). Most clinical infections from typical presentations described above are associated with virulence plasmid-containing low-pathogenic strains, which genome sequencing has assigned to four phylogroups that correspond to the classical groups of Biotype 2/3 serogroup O:9 (BT2/3 O:9), BT4 O:3, BT2/3 O:5,27, and BT5. A highly pathogenic (mouse-lethal) virulence plasmid-containing phylogroup (BT1B O:8) used to be commonly reported from North America, where it is now rare, but occasional cases are now reported from Europe and Japan. A current area under active research is the clinical significance of a further phylogroup comprising Biotype 1A *Y. enterocolitica* strains of various serotypes. These comprise >50% of *Y. enterocolitica* fecal isolates in recent clinical studies from England, France, and China and >20% in New Zealand. They have generally been regarded as nonpathogenic because they lack the virulence plasmid and are not pathogenic in mouse-infection models. However, they do contain other genes associated with *Yersinia* pathogenesis, invade epithelial cells and macrophages in vitro, and are pathogenic in an insect model. They form the majority of *Y. enterocolitica* cultures from food sampled in the United Kingdom and France.

Because of the frequency with which the virulence plasmid is lost on laboratory subculture, combined biochemical identification (with biotyping according to a standard schema) and serologic identification was usually required to interpret the significance of an isolate of *Y. enterocolitica* from a nonsterile site. Whole genome DNA sequencing applying a *Yersinia* genus wide seven-gene multilocus sequence typing (MLST) scheme can now speciate *Y. enterocolitica, Y. pestis,* and *Y. pseudotuberculosis* and differentiate *Y. enterocolitica* biotypes. A cgMLST scheme provides a more detailed population structure and has been used for outbreak tracing and revealing novel, as yet phenotypically undefined *Yersinia* species.

Agglutinating or ELISA antibody titers to specific O-antigen types are used in the retrospective diagnosis of both *Y. enterocolitica* and *Y. pseudotuberculosis* infections. IgA and IgG antibodies persist in patients with reactive arthritis. Serologic cross-reactions between *Y. enterocolitica* serogroup O:9 and *Brucella* are due to the similarity of their lipopolysaccharide structures. Multiple assays are required to cover even the predominant serogroups (*Y. enterocolitica* O:3, O5,27, and O:9; *Y. pseudotuberculosis* O:1a, O:1b, and O:3), and these assays are generally available only in reference laboratories. ELISA and western blot tests for antibodies to Yops, which are expressed by all pathogenic strains of *Y. enterocolitica* and *Y. pseudotuberculosis*, are also available; most of the positivity in these assays probably relates to previous infection with *Y. enterocolitica*.

TREATMENT

Yersiniosis

Most cases of diarrhea caused by enteropathogenic *Yersinia* are self-limiting. Data from clinical trials do not support antimicrobial treatment for adults or children with *Y. enterocolitica* diarrhea. Systemic infections with bacteremia or focal infections outside the gastrointestinal tract generally require antimicrobial therapy. Infants <3 months of age with documented *Y. enterocolitica* infection may require antimicrobial treatment because of the increased likelihood of bacteremia in this age group. *Y. enterocolitica* strains nearly always express β-lactamases. Because of the relative rarity of systemic *Y. enterocolitica* infection, there are no clinical trial data to guide antimicrobial choice or to suggest the optimal dose and duration of therapy. On the basis of retrospective case series and in vitro sensitivity data, fluoroquinolone therapy is effective for bacteremia in adults; for example, ciprofloxacin is given at a typical dose of 500 mg twice daily by mouth or 400 mg twice daily IV for at least 2 weeks (longer if positive blood cultures persist). A third-generation cephalosporin is an alternative—e.g., cefotaxime (typical dose, 6–8 g/d in 3 or 4 divided doses) or ceftriaxone. In children, third-generation cephalosporins are effective; for example, cefotaxime is given to children ≥1 month of age at a typical dose of 75–100 mg/kg per day in 3 or 4 divided doses, with an increase to 150–200 mg/kg per day in severe cases (maximal daily dose, 8–10 g). Amoxicillin and amoxicillin-clavulanate have shown poor efficacy in case series. Trimethoprim-sulfamethoxazole, gentamicin, and imipenem are all active in vitro. *Y. pseudotuberculosis* strains do not express β-lactamase but are intrinsically resistant to polymyxin. Because human infection with *Y. pseudotuberculosis* is less common than that with *Y. enterocolitica*, less case information is available; however, studies in mice suggest that ampicillin is ineffective. Drugs similar to those used against *Y. enterocolitica* should be used. The best results have been obtained with a quinolone.

Some trials of treatment for reactive arthritis (with a large proportion of cases due to *Yersinia*) found that 3 months of oral ciprofloxacin therapy did not affect outcome. One trial in which the same therapy was given specifically for *Y. enterocolitica*–reactive arthritis found that, while outcome indeed was not affected, there was a trend toward faster remission of symptoms in the treated group. Follow-up 4–7 years after initial antibiotic treatment of reactive arthritis (predominantly following *Salmonella* and *Yersinia* infections) demonstrated apparent efficacy in the prevention of chronic arthritis in HLA-B27-positive individuals. A trial showing that azithromycin therapy did not affect outcome in reactive arthritis included cases thought to have followed yersiniosis, although no breakdown of cases was provided.

■ PREVENTION AND CONTROL

Current control measures are similar to those used against other enteric pathogens like *Salmonella* and *Campylobacter*, which colonize the intestine of food animals. The focus is on safe handling and processing of food. No vaccine is effective in preventing intestinal colonization of food animals by enteropathogenic *Yersinia*. Consumption of food made from raw pork (which is popular in Germany and Belgium) should be discouraged at present because it is not possible to eliminate contamination with the enteropathogenic *Yersinia* strains found worldwide in pigs. Exposure of infants to raw pig intestine during domestic preparation of chitterlings is inadvisable. Modification of abattoir technique in Scandinavian countries from the 1990s onward included the removal of pig intestines in a closed plastic bag; levels of carcass contamination with *Y. enterocolitica* were reduced, but such contamination was not eliminated. Experimental pig herds free of pathogenic *Y. enterocolitica* O:3 (and also of *Salmonella*, *Campylobacter*, *Toxoplasma*, and *Trichinella*) have been established by selective breeding in Norway but remain rare. In the food industry, vigilance is required because of the potential for large outbreaks if small numbers of enteropathogenic yersiniae contaminate any ready-to-eat food whose safe preservation is based on refrigeration before consumption.

The rare phenomenon of contamination of blood for transfusion has proved impossible to eradicate. However, leukodepletion is now practiced in most blood transfusion centers, primarily to prevent nonhemolytic febrile transfusion reactions and alloimmunization against HLA antigens. This measure reduces but does not eliminate the risk of *Yersinia* blood contamination.

Notification of yersiniosis is now obligatory in some countries.

■ FURTHER READING

Plague

Campbell SB et al: Animal exposure and human plague, United States, 1970–2017. Emerg Infect Dis 25:2270, 2019.

Demeure C et al: *Yersinia pestis* and plague: An updated view on evolution, virulence determinants, immune subversion, vaccination and diagnostics. Microbes Infect 21:202, 2019.

Klunk J et al: Evolution of immune genes is associated with the Black Death. Nature 611:312, 2022.

Nelson CA et al: Antimicrobial treatment and prophylaxis of plague: Recommendations for naturally acquired infections and bioterrorism response. MMWR Recomm Rep 70(No. RR-3):1, 2021.

Salam AP et al: Deaths associated with pneumonic plague, 1946–2017. Emerg Infect Dis 26: 2432, 2020.

World Health Organization: WHO guidelines for plague management: Revised recommendations for the use of rapid diagnostic tests, fluoroquinolones for case management and personal protective equipment for prevention of post-mortem transmission, May 27, 2021. Available at: *https://www.who.int/publications/i/item/9789240015579*. Accessed February 16, 2024.

Yang R et al: *Yersinia pestis* and plague: Some knowns and unknowns. Zoonoses (Burlingt) 3:5, 2023.

Yersiniosis

Francis MS et al: The pathogenic Yersiniae–advances in the understanding of physiology and virulence. Front Cell Infect Microbiol 9:119. 2019.

Savin C et al: Genus-wide *Yersinia* core-genome multilocus sequence typing for species identification and strain characterization. Microb Genom 5:e000301, 2019.

Šumilo D et al: Forgotten but not gone: *Yersinia* infections in England, 1975 to 2020. Eurosurveillance 28:2200516, 2023.

177

Bartonella Infections, Including Cat-Scratch Disease

Michael Giladi, Moshe Ephros

Bartonella species are fastidious, facultative intracellular, slow-growing, gram-negative bacteria that cause a broad spectrum of diseases in humans. This genus includes >40 distinct species or subspecies, of which at least 20 have been recognized as confirmed or potential human pathogens; *Bartonella bacilliformis*, *Bartonella quintana*, and *Bartonella henselae* are most commonly identified **(Table 177-1)**. Most *Bartonella* species have successfully adapted to survival in specific domestic or wild mammals. Prolonged intraerythrocytic and endothelial cell infection in these animals creates a niche where the bacteria are protected from both innate and adaptive immunity and which serves as a reservoir for human infections. Chronic bacteremia, which may persist even following prolonged antimicrobial therapy, is a typical finding in *Bartonella*-hosting mammals. Adhesion and entry to endothelial cells and erythrocytes is facilitated by the *Bartonella* adhesin A or variably expressed outer-membrane proteins. *Bartonella* characteristically evades the host immune system by modification of its virulence factors (e.g., lipopolysaccharides or flagella) and by attenuation of the immune response. Arthropod vectors are often involved. Humans are incidental hosts for most *Bartonella* species with the exception of *B. quintana* and *B. bacilliformis*, which are not considered zoonotic agents. Isolation and characterization of *Bartonella* species are difficult and require special techniques. Clinical presentation generally depends on both the infecting *Bartonella* species and the immune status of the infected individual. *Bartonella* species are susceptible to many antibiotics in vitro; however, clinical responses to therapy and studies in animal

TABLE 177-1 *Bartonella* Species Known or Suspected to Be Human Pathogens

BARTONELLA SPECIES[a]	DISEASE(S)[b]	RESERVOIR HOST(S)[c]	ARTHROPOD VECTOR
B. alsatica	Endocarditis, lymphadenitis, vascular graft infection	Rabbits	Fleas
B. ancashensis	Verruga peruana, bacteremia	Unknown	Unknown
B. bacilliformis	Carrión's disease	Humans	Sandflies (*Lutzomyia verrucarum*)
B. clarridgeiae	Possibly cat-scratch disease	Cats, dogs	Unknown
B. doshiae	Bacteremia, arthralgia, blurred vision	Voles, rats	Unknown
B. elizabethae	Endocarditis	Rats, dogs	Fleas
B. grahamii	Lymphadenopathy, possibly retinitis and retinal vascular occlusion	Mice, voles, rats, deer	Fleas
B. henselae	Cat-scratch disease, bacillary angiomatosis, bacillary peliosis, bacteremia, endocarditis	Cats, other felines	Cat fleas (*Ctenocephalides felis*): associated with cat-to-cat, but not with cat-to-human, transmission
B. koehlerae	Endocarditis	Cats	Unknown
B. kosoyi	Lymphadenopathy	Rats	Fleas
"Candidatus Bartonella *mayotimonensis"*[d]	Endocarditis	Bats	Unknown
"Candidatus Bartonella *melophagi"*[d]	Various clinical manifestations	Sheep	Sheep keds
B. quintana	Trench fever, chronic bacteremia, bacillary angiomatosis, endocarditis	Humans	Human body lice (*Pediculus humanus corporis*)
B. rochalimae	Bacteremia, fever, splenomegaly	Foxes, coyotes, dogs, skunks, rats, other rodents	Fleas
B. schoenbuchensis	Bacteremia, fever, myalgia, deer ked dermatitis?	Cattle, roe deer, moose	Unknown
B. tamiae	Bacteremia, fever, myalgia, rash	Rats	Unknown
B. tribocorum	Bacteremia, myalgia, headache	Rats, mice	Unknown
B. vinsonii subsp. *arupensis*	Endocarditis, febrile illness	Mice, dogs	Ticks
B. vinsonii subsp. *berkhoffii*	Endocarditis	Domestic dogs, coyotes, gray foxes	Ticks
"Candidatus Bartonella *washoensis"*[d]	Endocarditis, myocarditis, meningitis	Squirrels, rabbits, prairie dogs, possibly other rodents	Fleas, ticks

[a]Many other *Bartonella* species exist but are not recognized as human pathogens. [b]Animal-associated *Bartonella* species (*B. henselae, B. doshiae, B. schoenbuchensis*, and *B. tribocorum*) were isolated from blood of patients who reported tick bites and chronic symptoms such as fatigue and myalgia. DNA of *B. henselae, B. vinsonii* subsp. *berkhoffii, B. koehlerae*, or *B. melophagi* or co-infection with more than one *Bartonella* species was detected by polymerase chain reaction in blood samples from patients with extensive arthropod and animal exposure who presented with chronic neurologic or neurocognitive syndromes. The causal relationship between bacteremia with these pathogens, tick bites, and clinical manifestations needs to be established. [c]Animals are implicated when existing evidence supports their infection with *Bartonella* species. Data supporting animal-to-human transmission may be lacking. [d]*Candidatus* is a taxonomic status for bacteria that cannot be described in sufficient detail to warrant establishment of a novel taxon or cannot be cultured or propagated in culture media. The phylogenetic relatedness of these bacteria has been determined by gene amplification and sequence analysis.

models suggest that the minimal inhibitory concentrations of many antimicrobial agents correlate poorly with the drugs' in vivo efficacies in patients with *Bartonella* infections.

CAT-SCRATCH DISEASE

DEFINITION AND ETIOLOGY

Usually a self-limited illness, cat-scratch disease (CSD) has two general clinical presentations. *Typical* CSD, the more common, is characterized by subacute regional lymphadenopathy; *atypical* or *extranodal* CSD is the collective designation for numerous extranodal manifestations involving various organs. *B. henselae* is the principal etiologic agent of CSD. Rare cases have been associated with *Afipia felis* and other *Bartonella* species.

EPIDEMIOLOGY

CSD occurs worldwide, favoring warm and humid climates. In temperate climates, incidence peaks during fall and winter. Adults are affected nearly as frequently as children. Intrafamilial clustering is rare, and person-to-person transmission does not occur. Apparently healthy bacteremic cats constitute the major reservoir of *B. henselae*, and cat fleas (*Ctenocephalides felis*) may be responsible for cat-to-cat transmission. CSD usually follows contact with cats (especially kittens), but other animals (e.g., dogs) have been implicated as possible reservoirs in rare instances. In the United States, the estimated annual disease incidence is 4–9 outpatient cases per 100,000 population with a higher incidence among those who lived in the southern United States and among children 5–9 years of age. About 5–10% of patients in the United States are hospitalized. A higher proportion of hospitalization has been reported in Israel.

PATHOGENESIS

Although cat fleas are likely responsible for cat-to-cat transmission, the mode of cat-to-human transmission is undetermined. *B. henselae*–infected cats' saliva spreads to claws by self-licking, and *B. henselae*–contaminated flea feces can be inoculated by a scratch or a bite. Infection of mucous membranes or conjunctivae via droplets or licking may occur as well. With lymphatic drainage to one or more regional lymph nodes in immunocompetent hosts, a T_H1 response can result in necrotizing granulomatous lymphadenitis. Dendritic cells, along with their associated chemokines, play a role in the host inflammatory response and granuloma formation.

CLINICAL MANIFESTATIONS AND PROGNOSIS

Of patients with CSD, 85–90% have *typical* disease. The primary lesion, a small (0.3- to 1-cm) painless erythematous papule or pustule, develops at the inoculation site within days to 2 weeks in about one-third to two-thirds of patients **(Fig. 177-1*A, B*)**. Lymphadenopathy develops 1–3 weeks or longer after cat contact. The affected lymph node(s) are enlarged and usually painful, sometimes have overlying erythema, and suppurate in ~10% of cases **(Fig. 177-1*C, D*, and *E*)**. Axillary/epitrochlear nodes are most commonly involved, followed by head/neck nodes and inguinal/femoral nodes. Approximately 50% of patients have fever, malaise, and anorexia. A smaller proportion experience weight loss and night sweats mimicking the presentation of lymphoma. Fever is usually low-grade but infrequently rises to ≥39°C. Resolution is slow, requiring

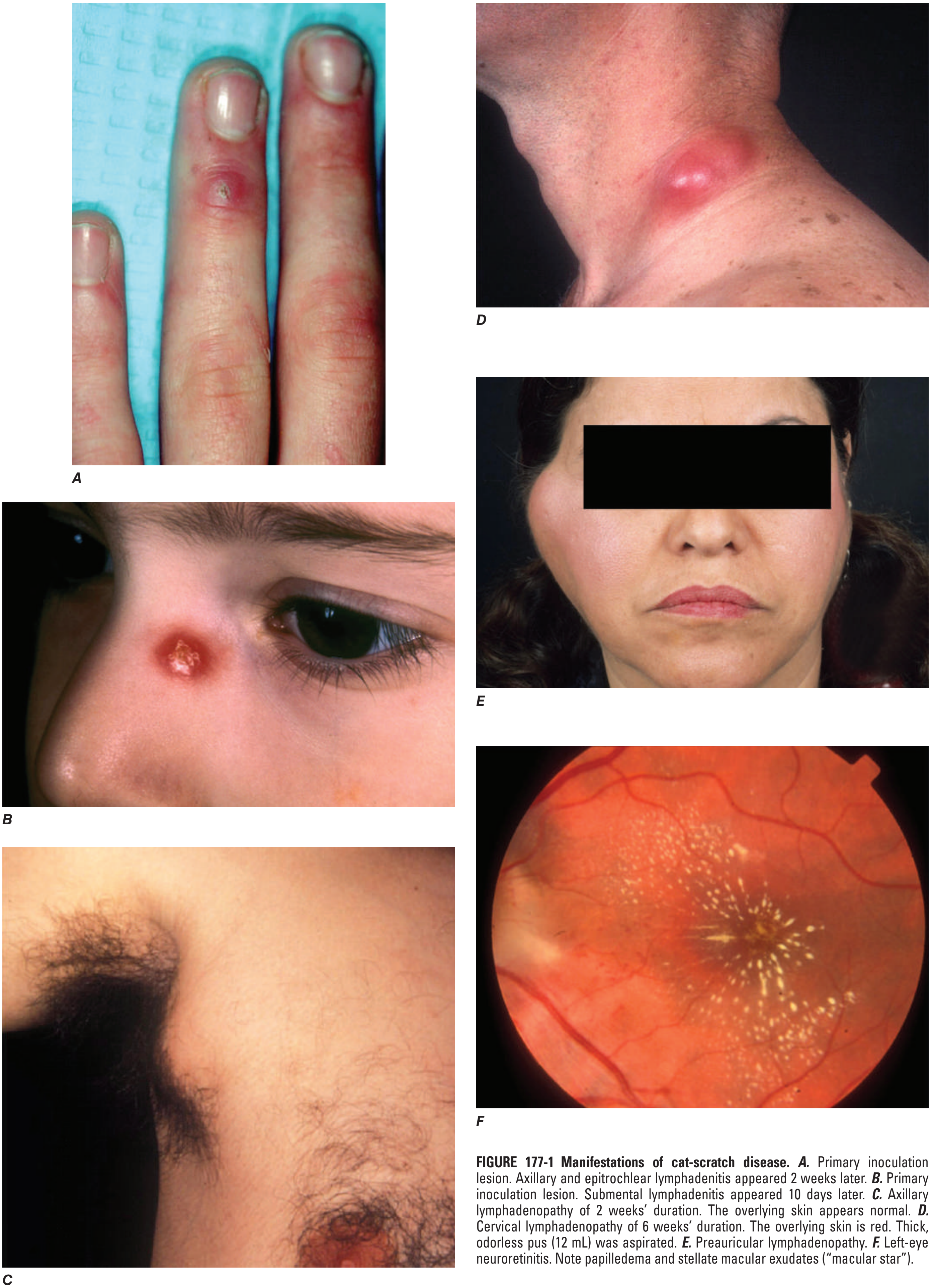

[redacted]

FIGURE 177-1 Manifestations of cat-scratch disease. ***A.*** Primary inoculation lesion. Axillary and epitrochlear lymphadenitis appeared 2 weeks later. ***B.*** Primary inoculation lesion. Submental lymphadenitis appeared 10 days later. ***C.*** Axillary lymphadenopathy of 2 weeks' duration. The overlying skin appears normal. ***D.*** Cervical lymphadenopathy of 6 weeks' duration. The overlying skin is red. Thick, odorless pus (12 mL) was aspirated. ***E.*** Preauricular lymphadenopathy. ***F.*** Left-eye neuroretinitis. Note papilledema and stellate macular exudates ("macular star").

weeks (for fever, pain, and accompanying signs and symptoms) to months (for node shrinkage).

Atypical CSD occurs in 10–15% of patients in the absence or presence of lymphadenopathy. Atypical disease includes Parinaud's oculoglandular syndrome (conjunctivitis with ipsilateral preauricular lymphadenitis; Fig. 177-1*E*), hepatosplenic disease, neuroretinitis (often presenting as unilateral deterioration of vision; **Fig. 177-1*F***) and other ophthalmologic manifestations, neurologic manifestations (encephalitis, seizures, myelitis, facial and other cranial nerve palsies, peripheral neuritis), fever of unknown origin (FUO), pneumonitis, debilitating myalgia, arthritis or arthralgia (affecting mostly women >20 years old), osteomyelitis (including multifocal disease), and dermatologic manifestations (including erythema nodosum **[see Fig. A1-39]**, sometimes accompanying arthropathy). CSD-associated FUO is a unique syndrome that may be severe and debilitating, often mimics malignancy, and may present with multiorgan involvement, including hepatosplenic space-occupying lesions, abdominal/mediastinal lymphadenopathy, ocular disease, and multifocal osteomyelitis. Fever may be continuous or relapsing. Other manifestations and syndromes (e.g., idiopathic thrombocytopenic purpura, Henoch-Schönlein purpura, erythema multiforme **[see Fig. A1-24*A*]**, myocarditis) also have been associated with CSD; however, a causal role for *B. henselae* has yet to be established. In elderly patients (>60 years old), lymphadenopathy is more often absent but encephalitis and FUO are more common than in younger patients. In immunocompetent individuals, CSD—whether typical or atypical—usually resolves without treatment and without sequelae, although some of the ophthalmologic manifestations may occasionally result in moderate to severe vision loss. Lifelong immunity is the rule.

DIAGNOSIS

Routine laboratory tests usually yield normal or nonspecific results. Histopathology initially shows lymphoid hyperplasia and later demonstrates stellate granulomata with necrosis, coalescing microabscesses, and occasional multinucleated giant cells—findings that, although nonspecific, may narrow the differential diagnosis. Serologic testing (immunofluorescence or enzyme immunoassay) is the most commonly used laboratory diagnostic approach, with variable sensitivity and specificity. CSD serodiagnosis is often based on the presence of IgG alone (i.e., in the absence of IgM), and seroconversion may take a few weeks; these two factors may pose difficulties in the interpretation of serologic results. Other tests are of low sensitivity (culture, Warthin-Starry silver staining), of low specificity (cytology, histopathology), or of limited availability in routine diagnostic laboratories (polymerase chain reaction [PCR], immunohistochemistry). PCR of pus aspirated from lymph nodes or the primary inoculation lesion is highly sensitive and specific and is particularly useful for definitive and rapid diagnosis in seronegative patients. PCR of a lymph node biopsy specimen may be less sensitive, perhaps because of sampling error.

APPROACH TO THE PATIENT

Cat-Scratch Disease

A history of cat contact, a primary inoculation lesion, and regional lymphadenopathy—especially axillary/epitrochlear lymphadenopathy—are highly suggestive of CSD. A characteristic clinical course and corroborative laboratory tests make the diagnosis very likely. Conversely, when acute- and convalescent-phase sera are negative (as is the case in 10–20% of CSD patients), when spontaneous regression of lymph node size does not occur, and particularly when constitutional symptoms persist, malignancy must be ruled out. Pyogenic lymphadenitis, mycobacterial infection, brucellosis, syphilis, tularemia, plague, toxoplasmosis, sporotrichosis, and histoplasmosis also should be considered. In clinically suspected CSD in a seronegative individual, fine-needle aspiration may be adequate and PCR can confirm the diagnosis. When data are less supportive of CSD, lymph node biopsy rather than fine-needle aspiration is preferred. In seronegative CSD patients with lymphadenopathy and severe complications (e.g., encephalitis or neuroretinitis), early biopsy is important to establish a specific diagnosis.

TREATMENT

Cat-Scratch Disease

(Table 177-2) Treatment regimens are based on only minimal data. Suppurative nodes should be drained by large-bore needle aspiration and not by incision and drainage to avoid chronic draining tracts. Systemic antibiotics are recommended in immunocompromised patients.

TABLE 177-2 Antimicrobial Therapy for Disease Caused by *Bartonella* Species in Adults

DISEASE	ANTIMICROBIAL THERAPY
Typical cat-scratch disease	Not routinely indicated; for patients with extensive lymphadenopathy, consider azithromycin (500 mg PO on day 1, then 250 mg PO once a day for 4 days)
Cat-scratch disease neuroretinitis	Value of systemic antibiotics is controversial, particularly when visual acuity is not significantly compromised. For more severe cases, doxycycline (100 mg PO bid) *plus* rifampin (300 mg PO bid) for 4–6 weeks is given. Consider adding systemic glucocorticoids.
Other atypical cat-scratch disease manifestations[a]	As per neuroretinitis. Treatment duration should be individualized.
Trench fever or chronic bacteremia with *B. quintana*	Gentamicin (3 mg/kg IV once a day for 14 days) *plus* doxycycline (200 mg PO once a day or 100 mg PO bid for 6 weeks)
Suspected *Bartonella* endocarditis	Gentamicin[b] (1 mg/kg IV q8h for ≥14 days) *plus* doxycycline (100 mg PO/IV bid for 6 weeks[c]) *plus* ceftriaxone (2 g IV once a day for 6 weeks)
Confirmed *Bartonella* endocarditis	As for suspected *Bartonella* endocarditis *minus* ceftriaxone
Bacillary angiomatosis	Erythromycin[d] (500 mg PO qid for 3 months) *or* Doxycycline (100 mg PO bid for 3 months)
Bacillary peliosis	Erythromycin[d] (500 mg PO qid for 4 months) *or* Doxycycline (100 mg PO bid for 4 months)
Carrión's disease	
Oroya fever	Chloramphenicol (500 mg PO/IV qid for 14 days) *plus* another antibiotic (β-lactam preferred) *or* Ciprofloxacin (500 mg PO bid for 10 days) ± ceftriaxone (1–2 g IV once a day for 10 days)
Verruga peruana	Azithromycin (500 mg PO once a day for 7 days) *or* Ciprofloxacin (500 mg PO bid for 7–10 days) *or* Rifampin (10 mg/kg PO once a day, to a maximum of 600 mg, for 14 days) *or* Streptomycin (15–20 mg/kg IM once a day for 10 days)

[a] Data on treatment efficacy for encephalitis and hepatosplenic cat-scratch disease are lacking. Therapy similar to that given for neuroretinitis is reasonable. [b] Some experts recommend gentamicin at 3 mg/kg IV once a day. If gentamicin is contraindicated, rifampin (300 mg PO bid) can be added to doxycycline for documented *Bartonella* endocarditis. [c] Some experts recommend extending oral doxycycline therapy for 3–6 months and measuring anti-*Bartonella* IgG titers every 8–10 weeks as a guide to duration of therapy, particularly when the infected valve was not resected. [d] Other macrolides are probably effective and may be substituted for erythromycin or doxycycline.

Source: Recommendations are modified from JM Rolain et al: Antimicrob Agents Chemother 48:1921, 2004.

PREVENTION

Avoiding cats (especially kittens) and instituting flea control are options for immunocompromised patients and for patients with valvular heart disease.

TRENCH FEVER AND CHRONIC BACTEREMIA

DEFINITION AND ETIOLOGY

Trench fever, also known as *5-day fever* or *quintan fever*, is a febrile illness caused by *B. quintana*. It was first described as an epidemic in the trenches of World War I; however, recent paleomicrobiological studies have provided evidence that *B. quintana* has been associated with human infection for 4000 years. This infection recently reemerged as chronic bacteremia seen most often in homeless people, also referred to as *urban* or *contemporary trench fever*.

EPIDEMIOLOGY

In addition to epidemics in World Wars I and II, sporadic outbreaks of trench fever have been reported in many regions of the world. The human body louse has been identified as the vector and humans as the only known reservoir for human infections. Recently, macaque monkeys from China and Japan also were found to be natural hosts of *B. quintana,* without evidence for macaque-to-human transmission. After a hiatus of several decades during which trench fever was almost forgotten, small clusters of cases of *B. quintana* chronic bacteremia were reported sporadically, primarily from the United States and France, in homeless people without HIV infection. Alcoholism and louse infestation were identified as risk factors.

CLINICAL MANIFESTATIONS

The typical incubation period is 15–25 days (range, 3–38 days). "Classic" trench fever, as described in 1919, ranges from a mild febrile illness to a recurrent or protracted and debilitating disease. Fever is often periodic, lasting 4–5 days with 5-day (range, 3- to 8-day) intervals between episodes. Other symptoms and signs include headache, back and limb pain, profuse sweating, shivering, myalgia, arthralgia, splenomegaly, a maculopapular rash in occasional cases, and nuchal rigidity in some cases. Untreated, the disease usually lasts 4–6 weeks. Death is rare. The clinical spectrum of *B. quintana* bacteremia in homeless people ranges from asymptomatic infection to a febrile illness with headache, severe leg pain, and thrombocytopenia. Endocarditis sometimes develops.

DIAGNOSIS

Definitive diagnosis requires isolation of *B. quintana* by blood culture. Some patients have positive blood cultures for several weeks. Patients with acute trench fever typically develop significant titers of antibody to *Bartonella*, whereas those with chronic *B. quintana* bacteremia may be seronegative. Patients with high titers of IgG antibodies should be evaluated for endocarditis. In epidemics, trench fever should be differentiated from epidemic louse-borne typhus and relapsing fever, which occur under similar conditions and share many features.

TREATMENT

B. quintana Bacteremia

(Table 177-2) In a small, randomized, placebo-controlled trial involving homeless people with *B. quintana* bacteremia, therapy with gentamicin and doxycycline was superior to administration of placebo in eradicating bacteremia. Treatment of bacteremia is important, even in clinically mild cases, to prevent endocarditis. Optimal therapy for trench fever without documented bacteremia is uncertain.

BARTONELLA ENDOCARDITIS

DEFINITION AND ETIOLOGY

Coxiella burnetii **(Chap. 192)** and *Bartonella* species are the most common pathogens of culture-negative endocarditis **(Chap. 133)**. In France, for example, *Bartonella* species were identified as the etiologic agents in 28% of 348 cases of culture-negative endocarditis. Prevalence, however, varies by geographic location and epidemiologic setting. *B. quintana* and *B. henselae* are the most common *Bartonella* species implicated in endocarditis, and other *Bartonella* species have reportedly caused rare cases (Table 177-1).

EPIDEMIOLOGY

Bartonella endocarditis has been reported worldwide. Most patients are adults; more are male than female. Risk factors associated with *B. quintana* endocarditis include homelessness, alcoholism, and body louse infestation; however, individuals with no risk factors have had *Bartonella* endocarditis diagnosed as well. *B. henselae* endocarditis is associated with exposure to cats. Most cases involve native rather than prosthetic valves; the aortic valve accounts for ~60% of cases. Patients with *B. henselae* endocarditis usually have preexisting valvulopathy, whereas *B. quintana* often infects normal valves.

CLINICAL MANIFESTATIONS

Clinical manifestations are usually characteristic of subacute endocarditis of any etiology. However, a substantial number of patients have a prolonged, minimally febrile or even afebrile indolent illness, with mild nonspecific symptoms lasting weeks or months before the diagnosis is made. Initial echocardiography may not show vegetations. Acute, aggressive disease is rare.

DIAGNOSIS

Blood cultures, even with use of special techniques (lysis centrifugation or EDTA-containing tubes), are positive in only ~25% of cases—mostly those caused by *B. quintana* and only rarely those caused by *B. henselae*. Prolonged incubation of cultures (up to 6 weeks) is required. Serologic tests—either immunofluorescence or enzyme immunoassay—usually demonstrate high-titer (≥1:800) IgG antibodies to *Bartonella*. Because of cross-antigenicity, routine serology does not distinguish between *B. quintana* and *B. henselae* and may also be cross-reactive with other pathogens, such as *C. burnetii* and *Chlamydia* species. Approximately 40–50% of patients with Q fever endocarditis have antibodies that significantly react with *B. henselae* and/or *B. quintana;* however, in the majority of cases, IgG titers against the true pathogen are several dilutions higher than those for the pathogen causing the false-positive serology, thus allowing the correct diagnosis to be made. Identification of *Bartonella* to the species level is usually accomplished by application of PCR and DNA sequencing methods to valve tissue or, infrequently, blood specimens.

TREATMENT

Bartonella Endocarditis

(Table 177-2) For patients with culture-negative endocarditis suspected to be due to *Bartonella* species, empirical treatment consists of gentamicin, doxycycline, and ceftriaxone; the major role of ceftriaxone in this regimen is to adequately treat other potential causes of culture-negative endocarditis, including members of the HACEK group **(Chap. 163)**. Once a diagnosis of *Bartonella* endocarditis has been established, ceftriaxone is discontinued. Aminoglycosides, the only antibiotics known to be bactericidal against *Bartonella*, should be included in the regimen for ≥2 weeks. Indications for valvular surgery are the same as in subacute endocarditis due to other pathogens; however, the proportion of patients who undergo surgery (~60%) is high, probably as a consequence of delayed diagnosis.

BACILLARY ANGIOMATOSIS AND PELIOSIS

DEFINITION AND ETIOLOGY

Bacillary angiomatosis (sometimes called *bacillary epithelioid angiomatosis* or *epithelioid angiomatosis*) is a disease of severely immunocompromised patients, is caused by *B. henselae* or *B. quintana*, and is characterized by neovascular proliferative lesions involving various organs. Both species cause cutaneous lesions; hepatosplenic lesions are caused only by *B. henselae*, whereas subcutaneous and lytic bone lesions are more frequently associated with *B. quintana*. *Bartonella* angiogenic factor A, a proangiogenic autotransporter identified in *B. henselae*, activates the host vascular endothelial growth factor pathway to drive angiogenesis. Bacillary peliosis is a closely related angioproliferative disorder caused by *B. henselae* and involving primarily the liver (peliosis hepatis) but also the spleen and lymph nodes. Bacillary peliosis is characterized by blood-filled cystic structures whose size ranges from microscopic to several millimeters.

■ EPIDEMIOLOGY

Bacillary angiomatosis and bacillary peliosis occur primarily in HIV-infected persons **(Chap. 208)** with CD4+ T-cell counts of <100/μL but also affect other immunosuppressed patients and, in rare instances, immunocompetent patients. The incidence has decreased since the introduction of effective antiretroviral therapy and the routine use of rifabutin and macrolides to prevent *Mycobacterium avium* complex infection in AIDS patients. Contact with cats or cat fleas increases the risk of *B. henselae* infection. Risk factors for *B. quintana* infection are low income, homelessness, and body louse infestation.

■ CLINICAL MANIFESTATIONS

Bacillary angiomatosis presents most commonly as one or more cutaneous lesions that are not painful and may be tan, red, or purple in color. Subcutaneous, often tender nodules, superficial ulcerated plaques **(Fig. 177-2)**, and verrucous growths are also seen. Nodular forms resemble those seen in fungal or mycobacterial infections. Painful osseous lesions, most often involving long bones, may underlie cutaneous lesions and occasionally develop in their absence. Other organs are rarely involved. Patients usually have constitutional symptoms, including fever, chills, malaise, headache, anorexia, weight loss, and night sweats. In patients with advanced immunodeficiency, *B. henselae* and *B. quintana* are important causes of FUO. In osseous disease, lytic lesions are generally seen on radiography, and technetium scan shows focal uptake. The differential diagnosis of cutaneous bacillary angiomatosis includes Kaposi's sarcoma, pyogenic granuloma, subcutaneous tumors, and verruga peruana. In bacillary peliosis, hypodense hepatic areas are usually evident on imaging.

■ PATHOLOGY

Bacillary angiomatosis consists of lobular proliferations of small blood vessels lined by enlarged endothelial cells interspersed with mixed infiltrates of neutrophils and lymphocytes, with predominance of the former. Histologic examination of organs with bacillary peliosis reveals small blood-filled cystic lesions partially lined by endothelial cells that can be several millimeters in size. Peliotic lesions are surrounded by fibromyxoid stroma containing inflammatory cells, dilated capillaries, and clumps of granular material. Warthin-Starry silver staining of bacillary angiomatosis and peliosis lesions reveals clusters of bacilli. Cultures are usually negative.

■ DIAGNOSIS

Bacillary angiomatosis and bacillary peliosis are diagnosed by histologic examination. Blood cultures may be positive.

TREATMENT

Bacillary Angiomatosis and Peliosis

(Table 177-2) Prolonged therapy with a macrolide or doxycycline is recommended for both bacillary angiomatosis and bacillary peliosis.

■ PREVENTION

Reasonable strategies for HIV-infected persons consist of control of cat-flea infestation and avoidance of cat scratches (for prevention of

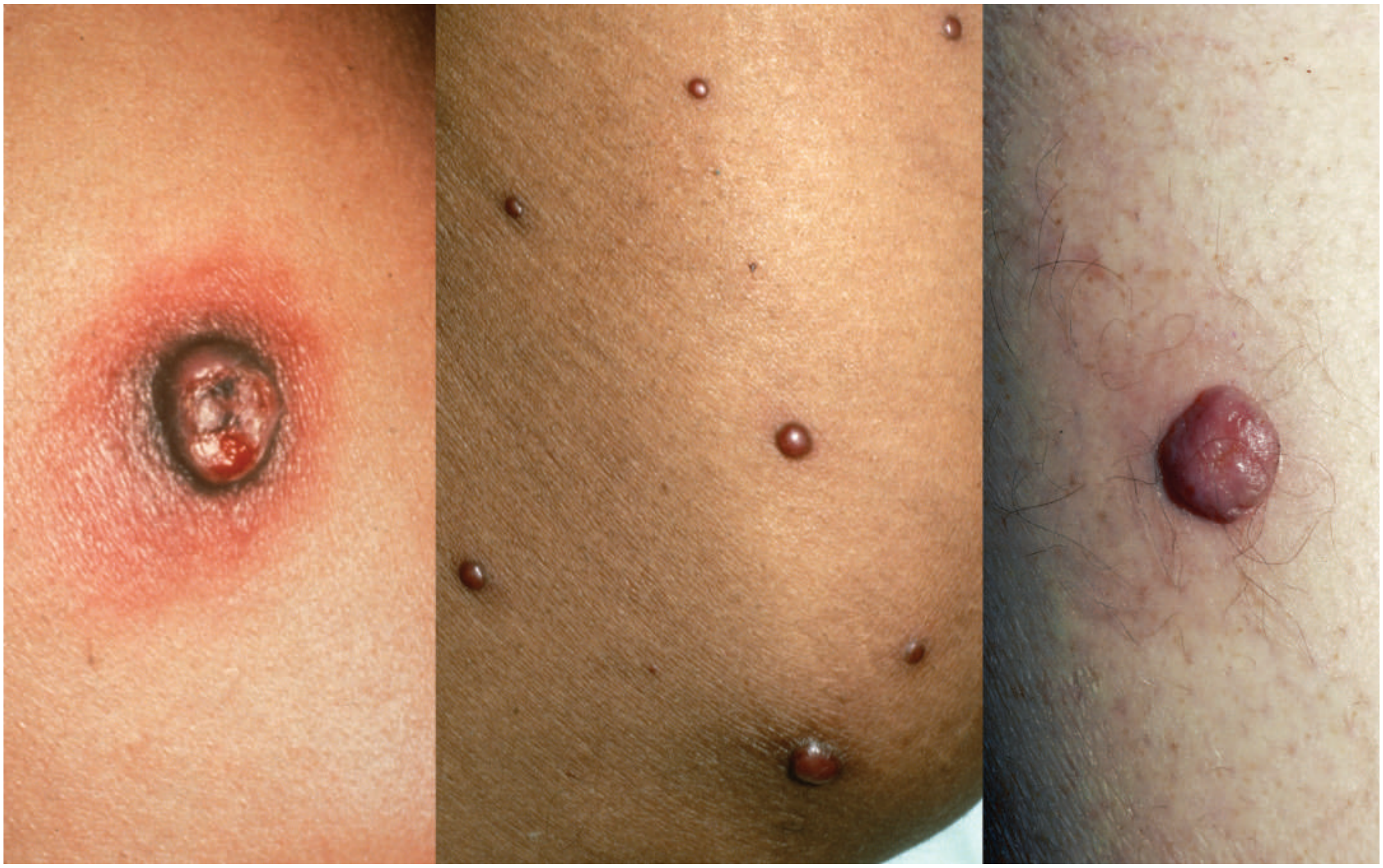

FIGURE 177-2 Lesions of cutaneous bacillary angiomatosis (BA) in three severely immunocompromised AIDS patients. Left panel shows a 1.5-cm ulcerated, bleeding BA lesion with an erythematous base; middle panel shows numerous small, 2-mm, scattered angiomatous BA lesions; right panel shows a 2.0-cm friable BA lesion on the thigh. *(Photos courtesy of Timothy Berger, MD; Jordan Tappero, MD, MPH; and Jane Koehler, MA, MD.)*

B. henselae) and avoidance and treatment of body louse infestation (for prevention of *B. quintana*). Primary prophylaxis is not recommended, but suppressive therapy with a macrolide or doxycycline is indicated in HIV-infected patients with bacillary angiomatosis or bacillary peliosis until CD4+ T-cell counts are >200/μL. Relapse may necessitate lifelong suppressive therapy in individual cases.

CARRIÓN'S DISEASE (OROYA FEVER AND VERRUGA PERUANA)

■ DEFINITION AND ETIOLOGY

Carrión's disease is a biphasic disease caused by *B. bacilliformis*. Oroya fever is the initial, bacteremic, systemic form, and verruga peruana—or "Peruvian warts"—is its late-onset, eruptive manifestation.

■ EPIDEMIOLOGY AND PREVENTION

Infection is endemic to the geographically restricted Andes valleys of Peru, Ecuador, and Colombia (~500–3200 m above sea level). Sporadic epidemics occur. The disease is transmitted by the phlebotomine sandfly *Lutzomyia verrucarum*. Maternal-fetal transmission as well as transmission by blood transfusion have been reported. Humans are the only known reservoir of *B. bacilliformis*. Sandfly control measures (e.g., insecticides) and personal protection measures (e.g., repellents, screening, bed nets) may decrease the risk of infection.

■ PATHOGENESIS

After inoculation by the sandfly, bacteria invade the blood vessel endothelium and proliferate; the reticuloendothelial system and various organs also may be involved. Upon reentry into blood vessels, *B. bacilliformis* invades, replicates, and ultimately destroys erythrocytes, with consequent massive hemolysis and sudden, severe anemia. Microvascular thrombosis results in end-organ ischemia. Survivors sometimes develop cutaneous hemangiomatous lesions characterized by various inflammatory cells, endothelial proliferation, and the presence of *B. bacilliformis* (verruga peruana).

■ CLINICAL MANIFESTATIONS

The incubation period is 3 weeks (range, 2–14 weeks). Oroya fever may present as a nonspecific bacteremic febrile illness without anemia or as an acute, severe hemolytic anemia with hepatomegaly and jaundice of rapid onset leading to vascular collapse and clouded sensorium. Myalgia, arthralgia, lymphadenopathy, and abdominal pain may develop. Temperature is elevated but not extremely so; high fever may suggest intercurrent infection. Subclinical asymptomatic infection also occurs. Secondary infection with *Salmonella* (*typhi* and non-*typhi*) is common, though other secondary infections have been reported and should be suspected when fever persists or recurs after defervescence. In verruga peruana, red, hemangioma-like, cutaneous vascular lesions of various sizes appear either weeks to months after systemic illness or with no previous suggestive history. These lesions persist for months up to 1 year. Mucosal and internal lesions also may develop.

■ DIAGNOSIS AND APPROACH TO THE PATIENT

Systemic illness (with or without anemia) or the development of cutaneous lesions in a person who has been to an endemic area raises the possibility of *B. bacilliformis* infection. Severe anemia with exuberant reticulocytosis—and sometimes thrombocytopenia—can occur. In systemic illness, Giemsa-stained blood films may show typical intraerythrocytic bacilli. Blood and bone marrow cultures may be positive, but growth is slow (1–6 weeks) and requires lower incubation temperature. Serologic assays may be helpful. Diagnosis of verruga peruana is largely clinical, although biopsy may be required to confirm the diagnosis. Several PCR assays have been described; however, their role in diagnosis remains to be clinically validated. Differential diagnosis includes coendemic systemic febrile illnesses (e.g., typhoid fever, malaria, brucellosis) and diseases producing cutaneous vascular lesions (e.g., hemangiomata, bacillary angiomatosis, Kaposi's sarcoma).

TREATMENT

Carrión's Disease

(Table 177-2) Antibiotic therapy for systemic *B. bacilliformis* infection usually results in rapid defervescence. Additional antibiotic treatment of intercurrent infection is often required. Blood transfusion may be necessary. Treatment of verruga peruana is not always required. Patients with numerous lesions, especially lesions that have been present for only a short period, may respond well to antibiotic therapy.

■ COMPLICATIONS AND PROGNOSIS

Mortality rates associated with Oroya fever have been reported to be as high as 40% without treatment but are considerably lower (~10%) with treatment. Complications such as bacterial superinfection and neurologic and cardiac manifestations occur frequently. Generalized massive edema (anasarca) and petechiae are associated with poor outcome. Permanent immunity usually develops.

■ FURTHER READING

Fournier PE et al: Epidemiologic and clinical characteristics of *Bartonella quintana* and *Bartonella henselae* endocarditis: A study of 48 patients. Medicine (Baltimore) 80:245, 2001.

Gomes C, Ruiz J: Carrion's disease: The sound of silence. Clin Microbiol Rev 31:e56, 2018.

Koehler JE et al: Molecular epidemiology of *Bartonella* infections in patients with bacillary angiomatosis-peliosis. N Engl J Med 337:1876, 1997.

Landes M et al: Cat scratch disease presenting as fever of unknown origin is a unique clinical syndrome. Clin Infect Dis 71:2818, 2020.

Rolain JM et al: Recommendations for treatment of human infections caused by *Bartonella* species. Antimicrob Agents Chemother 48:1921, 2004.

Rose SR, Koehler JE: *Bartonella* including cat scratch disease, in *Principles and Practice of Infectious Diseases*, 9th ed, GL Mandell et al (eds). Philadelphia, Elsevier, Inc. 2020, pp 2824–2843.

Wagner A, Dehio C: Role of distinct type-IV-secretion systems and secreted effector sets in host adaptation by pathogenic *Bartonella* species. Cell Microbiol 21:e13004, 2019.

178

Donovanosis

Nigel O'Farrell

Donovanosis is a chronic, progressive bacterial infection that usually involves the genital region. The condition is generally regarded as a sexually transmitted infection of low infectivity. This infection has been known by many other names, the most common being *granuloma inguinale*.

■ ETIOLOGY

The causative organism has been reclassified as *Klebsiella granulomatis comb nov* on the basis of phylogenetic analysis, although there is ongoing debate about this decision. Some authorities consider the original nomenclature (*Calymmatobacterium granulomatis*) to be more appropriate in light of analysis of 16S rRNA gene sequences.

Donovanosis was first described in Calcutta in 1882, and the causative organism was recognized by Charles Donovan in Madras in 1905. He identified the characteristic Donovan bodies, measuring 1.5 × 0.7 μm, in macrophages and the stratum malpighii. The organism was not reproducibly cultured until the mid-1990s, when its isolation in peripheral-blood monocytes and human epithelial cell lines was reported.

EPIDEMIOLOGY

Donovanosis has an unusual geographic distribution that has included Papua New Guinea, parts of southern Africa, India, the Caribbean, French Guyana, Brazil, and Aboriginal communities in Australia. In Australia, donovanosis has been almost entirely eliminated through a sustained program backed by strong political commitment and resources at the primary health care level. In South Africa, donovanosis is also very close to elimination. Although few cases are now reported in the United States, donovanosis was once prevalent in this country, with 5000–10,000 cases recorded in 1947. The largest epidemic recorded was in Dutch South Guinea, where 10,000 cases were identified in a population of 15,000 (the Marind-anim) between 1922 and 1952.

Donovanosis is associated with poor hygiene and is more common in lower socioeconomic groups than in those who are better off and in men than in women. Infection in sexual partners of index cases occurs to a limited extent. Donovanosis is a risk factor for HIV infection **(Chap. 208)**.

Globally, the incidence of donovanosis has decreased significantly in recent times. This decline probably reflects a greater focus on effective management of genital ulcers because of their role in facilitating HIV transmission.

CLINICAL FEATURES

A lesion starts as a papule or subcutaneous nodule that later ulcerates after trauma. The incubation period is uncertain, but experimental infections in humans indicate a duration of ~50 days. Four types of lesions have been described: (1) the classic ulcerogranulomatous lesion **(Fig. 178-1)**, a beefy red ulcer that bleeds readily when touched; (2) a hypertrophic or verrucous ulcer with a raised irregular edge; (3) a necrotic, offensive-smelling ulcer causing tissue destruction; and (4) a sclerotic or cicatricial lesion with fibrous and scar tissue.

The genitals are affected in 90% of patients and the inguinal region in 10%. The most common sites of infection are the prepuce, coronal sulcus, frenum, and glans in men and the labia minora and fourchette in women. Cervical lesions may mimic cervical carcinoma. In men, lesions are associated with lack of circumcision. Lymphadenitis is uncommon. Extragenital lesions occur in 6% of cases and may involve the lip, gums, cheek, palate, pharynx, larynx, and chest. Hematogenous spread with involvement of liver and bone has been reported. During pregnancy, lesions tend to develop more quickly and respond more slowly to treatment. Polyarthritis and osteomyelitis are rare complications. In newborn infants, donovanosis may present with ear infection. Cases in children have been attributed to sitting on the laps of infected adults. As the incidence of donovanosis has decreased, the number of unusual case reports has appeared to be increasing.

Complications include neoplastic changes, pseudoelephantiasis, and stenosis of the urethra, vagina, or anus.

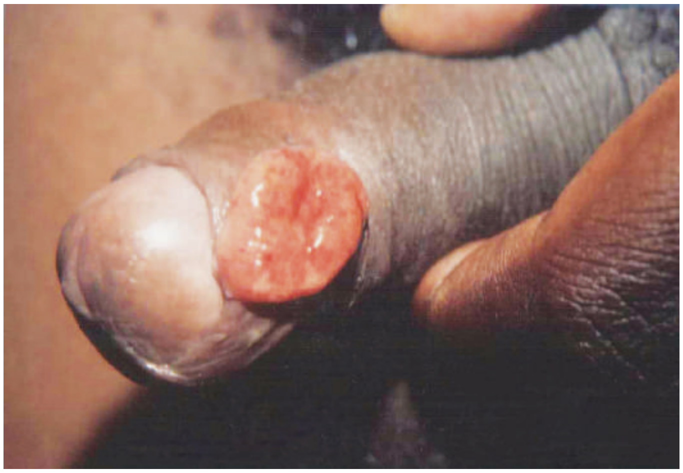

FIGURE 178-1 Ulcerogranulomatous penile lesion of donovanosis, with some hypertrophic features.

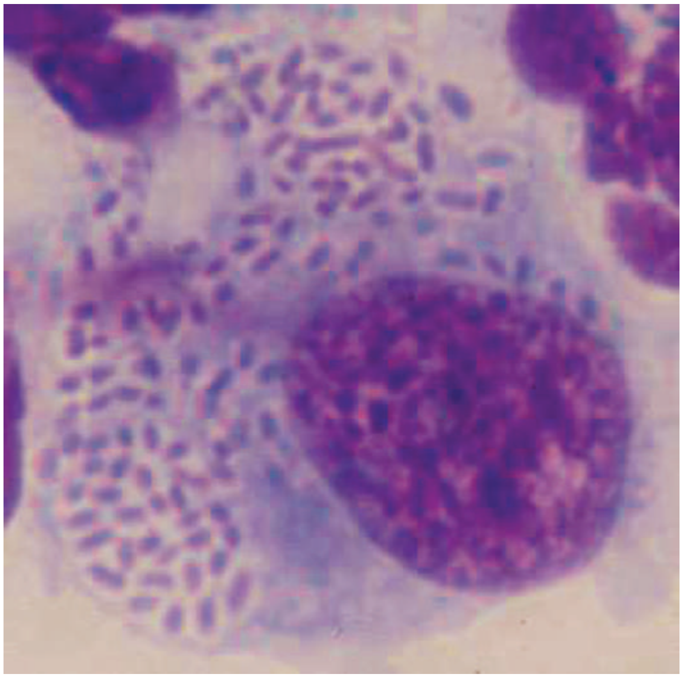

FIGURE 178-2 Pund cell stained by rapid Giemsa (RapiDiff) technique. Numerous Donovan bodies are visible.

DIAGNOSIS

A clinical diagnosis of donovanosis made by an experienced practitioner on the basis of the lesion's appearance usually has a high positive predictive value. The diagnosis is confirmed by microscopic identification of Donovan bodies **(Fig. 178-2)** in tissue smears. Preparation of a good-quality smear is important. If donovanosis is suspected on clinical grounds, the smear for Donovan bodies should be taken before swab samples are collected to be tested for other causes of genital ulceration so that enough material can be collected from the ulcer. A swab should be rolled firmly over an ulcer previously cleaned with a dry swab to remove debris. Smears can be examined in a clinical setting by direct microscopy with a rapid Giemsa or Wright's stain. Alternatively, a piece of granulation tissue crushed and spread between two slides can be used. Donovan bodies can be seen in large, mononuclear (Pund) cells as gram-negative intracytoplasmic cysts filled with deeply staining bodies that may have a safety-pin appearance. These cysts eventually rupture and release the infective organisms. Histologic changes include chronic inflammation with infiltration of plasma cells and neutrophils. Epithelial changes include ulceration, microabscesses, and elongation of rete ridges.

A diagnostic polymerase chain reaction (PCR) test was based on the observation that two unique base changes in the *phoE* gene eliminate Hae111 restriction sites, enabling differentiation of *K. granulomatis comb nov* from related *Klebsiella* species. PCR analysis with a colorimetric detection system can now be used in routine diagnostic laboratories. A genital ulcer multiplex PCR that includes *K. granulomatis* has been developed. Serologic tests are only poorly specific and are not currently used.

The differential diagnosis of donovanosis includes primary syphilitic chancres, secondary syphilis (condylomata lata), chancroid, lymphogranuloma venereum, genital herpes, neoplasm, and amebiasis. Mixed infections are common. Histologic appearances should be distinguished from those of rhinoscleroma, leishmaniasis, and histoplasmosis.

TREATMENT

Donovanosis

Many patients with donovanosis present quite late with extensive ulceration. They may be embarrassed and have low self-esteem related to their disease. Reassurance that they have a treatable

TABLE 178-1 Effective Antibiotics for the Treatment of Donovanosis

ANTIBIOTIC	ORAL DOSE
Azithromycin	1 g on day 1, then 500 mg daily for 7 days or 1 g weekly for 4 weeks
Trimethoprim-sulfamethoxazole	960 mg bid for 14 days
Doxycycline	100 mg bid for 14 days
Erythromycin	500 mg qid for 14 days (in pregnant women)
Tetracycline	500 mg qid for 14 days

condition is important, as are the administration of antibiotics and the monitoring of patients for an adequate interval (see below). Epidemiologic treatment of sexual partners and advice about how to improve genital hygiene are recommended.

The recommended drug regimens for donovanosis are shown in Table 178-1. Gentamicin can be added if the response is slow. Ceftriaxone, chloramphenicol, and norfloxacin also are effective. Patients treated for 14 days should be monitored until lesions have healed completely. Those treated with azithromycin probably do not need such rigorous follow-up.

Surgery may be indicated for very advanced lesions.

CONTROL AND PREVENTION

Donovanosis is probably the cause of genital ulceration that is most readily recognizable clinically. Donovanosis is now limited to a few specific locations, and its global eradication is a distinct possibility.

FURTHER READING

Muller EE, Kularatne R: The changing epidemiology of genital ulcer disease in South Africa: Has donovanosis been eliminated? Sex Transm Infect 96:596, 2020.

O'Farrell N: Donovanosis, in *Sexually Transmitted Diseases*, 4th ed. KK Holmes et al (eds). McGraw-Hill, 2008, pp 701–708.

Rajam RV, Rangiah PN: Donovanosis (granuloma inguinale, granuloma venereum). Monogr Ser World Health Organ 24:1, 1954.

Sehgal VN, Prasad AL: Donovanosis. Current concepts. Int J Dermatol 5:8, 1986.

Section 7 Miscellaneous Bacterial Infections

179 Nocardiosis

Michael S. Abers, Gregory A. Filice

INTRODUCTION

Bacteria of the genus *Nocardia* are saprophytic filamentous aerobes ubiquitous in soil and water worldwide. In the past, the majority of isolates associated with pneumonia and systemic disease were identified biochemically as *Nocardia asteroides*, but the development of genome sequencing has demonstrated that at least 53 of the more than 100 species of *Nocardia* are associated with human disease. Most cases of systemic nocardiosis are caused by *N. farcinica*, *N. cyriacigeorgica*, *N. nova*, *N. abscessus*, *N. otitidiscaviarum*, *N. transvalensis*, *N. brasiliensis*, *N. pseudobrasiliensis*, *N. paucivorans*, or *N. brevicatena* **(Fig. 179-1)**. *N. brasiliensis* is usually associated with disease limited to the skin. Nocardiosis is usually an opportunistic infection, occurring primarily in individuals with impairments in host defenses. Infections following local inoculation and pulmonary or systemic disease have distinct pathogenesis, microbiology, and management.

MICROBIOLOGY

Nocardiae are Gram-positive, weakly acid-fast, and catalase positive. Like other members of the Mycobacteriales order, the cell walls of nocardiae contain mycolic acids (45–65 carbon atoms) and trehalose. *Nocardia* tend to clump together when grown in liquid media. On solid agar, *Nocardia* species typically form chalky, wrinkled colonies with a whitish-yellow or orange-brown hue. Growth is optimal at 37°C but can occur at temperatures up to 45°C.

EPIDEMIOLOGY AND RISK FACTORS

Nocardiae are ubiquitous environmental saprophytes found in soil, water, and decaying organic matter worldwide. Humans are frequently exposed via inhalation or direct inoculation of the skin or an eye. Most immunocompetent individuals readily clear nocardiae without developing clinical disease. Certain impairments in host defenses predispose to invasive infection after inhalation. Primary cutaneous infection usually remains local. Nearly all cases are sporadic, but outbreaks have been reported in nosocomial settings among immunocompromised patients and in immunocompetent individuals related to surgical procedures or intravenous drug use. Person-to-person spread is not well documented. There is no known seasonality.

The incidence of nocardiosis in the general population, estimated on three continents (North America, Europe, and Australia), is approximately ~0.375 cases per 100,000 persons per year. *Nocardia* infections are more common among adults than among children and more common among males than females. In contrast to the other forms of nocardiosis that occur worldwide, cases of actinomycetoma have been reported mostly in tropical and subtropical regions, especially in Mexico, Sudan, and India. The most important risk factors are lower socioeconomic status and frequent contact with soil or vegetable matter. Many cases are in laborers.

Most cases of systemic nocardiosis occur in patients with host defense defects, including cell-mediated immunity and specific phagocyte defects. Those with nocardiosis typically possess one or more of the following risk factors: solid organ transplantation, hematopoietic stem cell transplantation, systemic corticosteroid use or Cushing syndrome, immunosuppressive therapy, or HIV infection. Among transplant recipients, risk factors for nocardiosis include corticosteroid dose, recent augmentation of immunosuppression for rejection or graft-versus-host disease, elevated calcineurin inhibitor levels, and patient age. Rare but well-described syndromes associated with nocardiosis include pulmonary alveolar proteinosis (PAP), neutralizing autoantibodies to granulocyte-macrophage colony-stimulating factor (GM-CSF), chronic granulomatous disease (CGD), and interleukin 12 (IL-12) or IL-12R deficiency. In the absence of any major risk factor for *Nocardia* infection, children and adults with nocardiosis should be tested for CGD.

PATHOGENESIS

Pulmonary and systemic nocardiosis both occur following inhalation of fragmented bacterial mycelia. In healthy individuals, a variety of host defense mechanisms, including both innate and adaptive immune responses, help control and eliminate nocardiae from the respiratory tract without causing clinical disease. The importance of neutrophils and macrophages in antinocardial host defense is suggested by the frequency of nocardiosis in patients with CGD. Nocardiae have evolved a number of properties that enable survival within phagocytes, including neutralization of oxidants, prevention of phagosome–lysosome fusion, and prevention of phagosome acidification. Neutrophils phagocytose nocardiae and limit their growth but do not kill them efficiently. Neutralizing autoantibodies against GM-CSF have been found in the majority of patients with autoimmune PAP and appear to be central to the pathogenesis of this disease. Nocardiae stimulate the production of GM-CSF in phagocytes in vitro, and extrapulmonary nocardiosis has been observed in several patients with autoantibodies to GM-CSF, most of whom had not had pulmonary alveolar proteinosis. The relationships between pulmonary alveolar proteinosis, nocardiosis, and antibodies to GM-CSF remain incompletely defined.

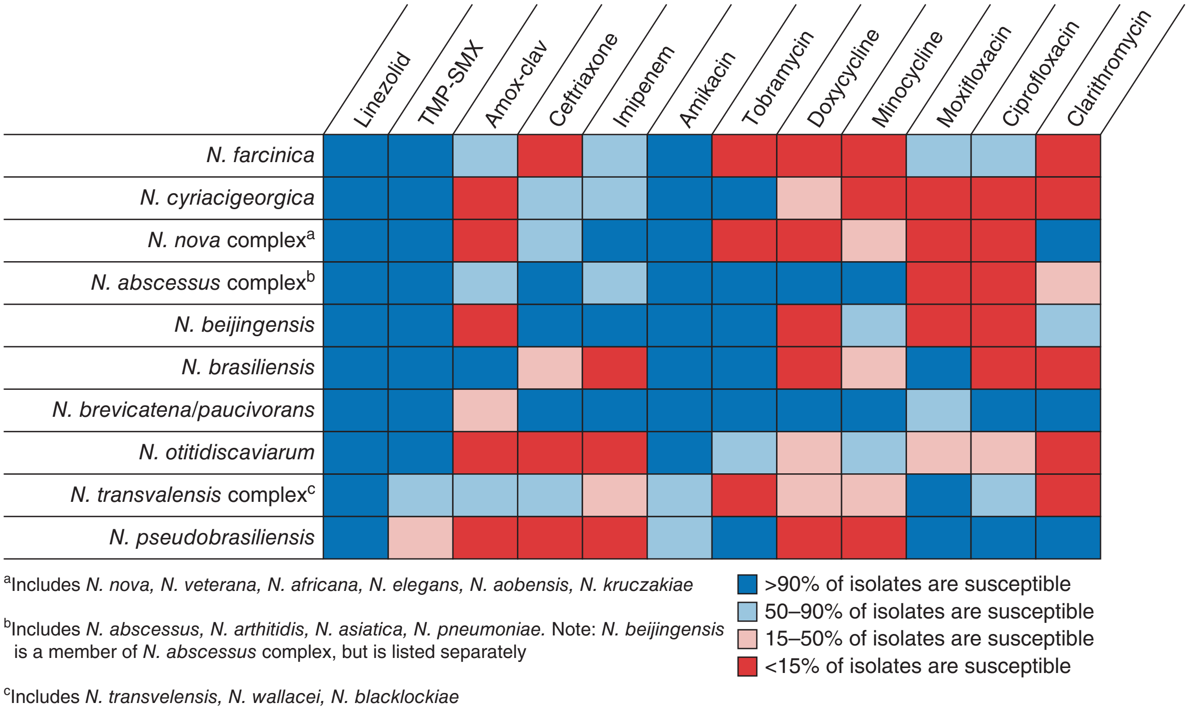

	Linezolid	TMP-SMX	Amox-clav	Ceftriaxone	Imipenem	Amikacin	Tobramycin	Doxycycline	Minocycline	Moxifloxacin	Ciprofloxacin	Clarithromycin
N. farcinica	🟦	🟦	🔹	🟥	🔹	🟦	🟥	🟥	🟥	🔹	🔹	🟥
N. cyriacigeorgica	🟦	🟦	🟥	🔹	🔹	🟦	🟦	🔸	🟥	🟥	🟥	🟥
N. nova complex[a]	🟦	🟦	🟥	🔹	🟦	🟦	🟥	🟥	🔸	🟥	🟥	🟦
N. abscessus complex[b]	🟦	🟦	🔹	🟦	🔹	🟦	🟦	🟦	🟦	🟥	🟥	🔸
N. beijingensis	🟦	🟦	🟥	🟦	🟦	🟦	🟦	🟥	🔹	🟥	🟥	🔹
N. brasiliensis	🟦	🟦	🟦	🔸	🟥	🟦	🟦	🟥	🔸	🟦	🟥	🟥
N. brevicatena/paucivorans	🟦	🟦	🔸	🟦	🟦	🟦	🟦	🟦	🟦	🔹	🟦	🟦
N. otitidiscaviarum	🟦	🟦	🟥	🟥	🟥	🟦	🔹	🔸	🔹	🔸	🔸	🟥
N. transvalensis complex[c]	🟦	🔹	🔹	🔹	🔸	🔹	🟥	🔸	🔸	🟦	🔹	🟥
N. pseudobrasiliensis	🟦	🔸	🟥	🟥	🟥	🔹	🟦	🟥	🟥	🟦	🟦	🟦

[a]Includes *N. nova, N. veterana, N. africana, N. elegans, N. aobensis, N. kruczakiae*

[b]Includes *N. abscessus, N. arthitidis, N. asiatica, N. pneumoniae.* Note: *N. beijingensis* is a member of *N. abscessus* complex, but is listed separately

[c]Includes *N. transvelensis, N. wallacei, N. blacklockiae*

🟦 >90% of isolates are susceptible
🔹 50–90% of isolates are susceptible
🔸 15–50% of isolates are susceptible
🟥 <15% of isolates are susceptible

FIGURE 179-1. ***Nocardia* species most commonly associated with human disease and their in vitro susceptibility patterns.** TMP-SMX: trimethoprim-sulfamethoxazole. *(Adapted from multiple sources.)*

Primary cutaneous nocardiosis typically follows penetrating injury of the skin by an object contaminated with soil or vegetable matter. Thorn-associated injuries, such as those associated with gardening, are a particularly common source of infection. Nocardiae may spread to draining lymph nodes, but dissemination to anatomically distant sites is uncommon. The ocular epithelium serves as a barrier that excludes nocardiae from the deeper tissues. Traumatic injury of the corneal epithelium provides a portal of entry for nocardiae to enter the ocular tissues, resulting in keratitis. The infectious process is limited to the superficial tissues.

PATHOLOGY

The characteristic histologic feature of nocardiosis is an abscess with extensive infiltration of neutrophils and prominent necrosis. Granulation tissue may surround the lesion, but extensive fibrosis and encapsulation are uncommon. Pyogranulomatous inflammation occurs in a minority of cases. In contrast, actinomycetoma is characterized by granulomatous inflammation and the formation of sinus tracts.

CLINICAL MANIFESTATIONS

Pleuropulmonary Disease The onset of pulmonary nocardiosis is typically subacute, with symptoms evolving over days to weeks prior to presentation. Among severely immunocompromised patients, the onset may be more acute. Cough is prominent and produces small amounts of thick, purulent sputum that is not malodorous. Fever, anorexia, and malaise are common; weight loss, dyspnea, pleuritic chest pain, and hemoptysis are uncommon. Without definitive treatment, patients frequently experience remissions and exacerbations over a period of weeks to months.

Typical radiographic patterns include infiltrates, nodules, and masses in one or more lobes **(Figs. 179-2 and 179-3)**. Nodules frequently cavitate (Fig. 179-3). Pleural effusion or empyema occurs in about one-third of cases. In some cases, empyema may occur in the absence of lung involvement.

Extrapulmonary Disease Dissemination to anatomically distant sites occurs in up to one-third of patients with pneumonia. While blood cultures are frequently negative, nocardiae likely reach metastatic sites of infection via hematogenous spread. Virtually any organ can be involved, but the central nervous system (CNS) and skin and subcutaneous tissues are the most common sites.

The CNS is involved in up to two-thirds of cases of disseminated nocardiosis. The vast majority of patients present with one or more brain abscesses. One-third of patients have no symptoms of CNS involvement, and one-fourth of patients do not have apparent pulmonary disease at presentation. Brain imaging demonstrates one or more ring-enhancing lesions that characteristically have a multilobulated appearance **(Fig. 179-4)**. Multiple abscesses are seen in two-thirds of patients. Many abscesses can be detected with contrast-enhanced computed tomography (CT), but magnetic resonance imaging (MRI) with intravenous gadolinium is more sensitive. *N. farcinica* appears to have a special predilection for CNS involvement. Meningitis is uncommon and is usually due to spread from a nearby brain abscess.

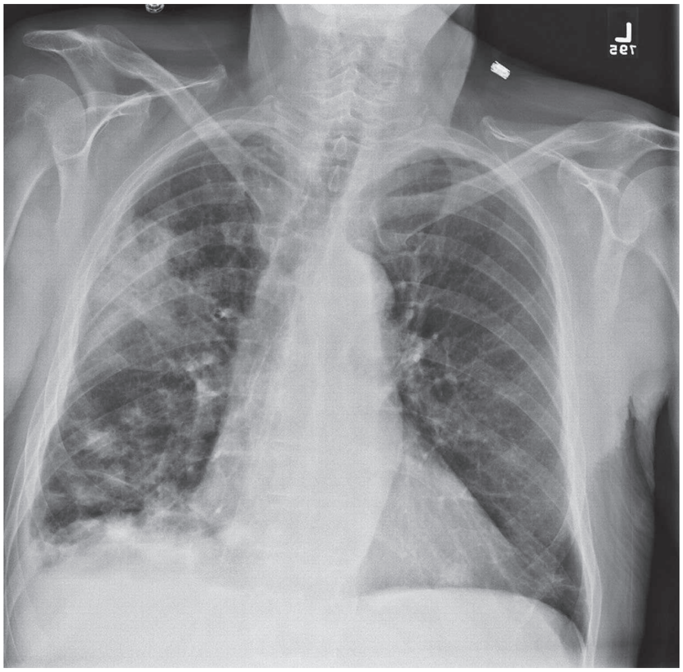

FIGURE 179-2. ***Nocardia* pneumonia.** A dense infiltrate with a possible cavity and several nodules are apparent in the right lung.

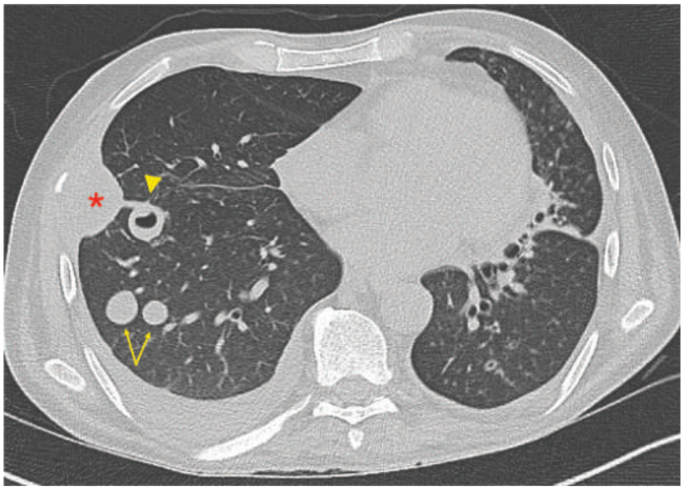

FIGURE 179-3. Pulmonary nocardiosis. A computed tomography scan shows solid nodules *(yellow arrows)* and a cavitated nodule *(yellow arrowhead)* in the right lung as well as a right-sided pleural effusion and an abscess of the right chest wall *(red asterisk)*. Bronchiectasis, which was unrelated to nocardiosis, is present in the left lung.

A neutrophilic pleocytosis is typically seen in the cerebrospinal fluid, but cultures are rarely positive.

Involvement of the skin and subcutaneous tissues occurs in one-fifth of patients with disseminated disease. Typical manifestations include subcutaneous abscesses, papules, nodules, and ulcers. Skin abscesses involve adjacent muscle in a minority of cases. Abscesses involving virtually every organ have been described with the most common sites being kidney, adrenal gland, muscle, bone, joint, eye, lymph node, and thyroid gland. Endocarditis has been reported and can affect native or prosthetic valves. Central venous catheter infections have been reported. Peritonitis has been reported in patients undergoing peritoneal dialysis.

Primary Cutaneous Nocardiosis Primary cutaneous nocardiosis follows transcutaneous nocardial inoculation and takes one of three forms: cellulitis, lymphocutaneous syndrome, or actinomycetoma.

Cellulitis begins 1–3 weeks after a recognized breach of the skin, often with soil contamination. Pain, swelling, erythema, and warmth develop over days to weeks. The lesions are usually firm and not fluctuant. Depending on the nature of the inoculation injury, disease may spread to nearby muscles, tendons, bones, and joints. Dissemination is rare.

Lymphocutaneous disease usually begins as a pyodermatous nodule at the site of inoculation, followed by central ulceration and purulent or honey-colored drainage. Subcutaneous nodules often appear along lymphatics that drain the primary lesion. Most cases follow inoculation of a limb, but cases involving the face have been reported, especially in children. Most cases of nocardial lymphocutaneous syndrome are associated with *N. brasiliensis*. Similar disease occurs with other pathogens, most notably *Sporothrix schenckii* **(Chap. 225)** and *Mycobacterium marinum* **(Chap. 185)**.

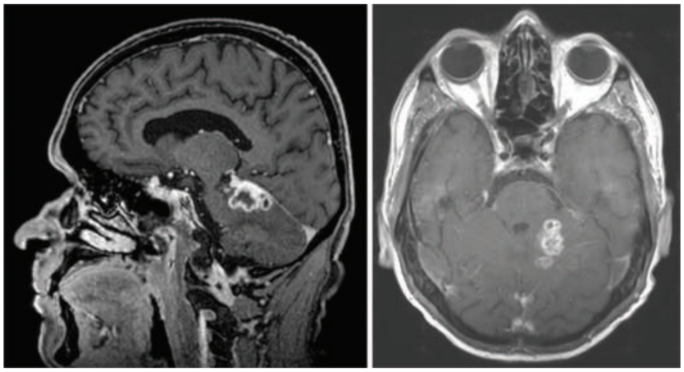

FIGURE 179-4. Multiloculated *Nocardia* brain abscess in the left cerebellum.

Mycetoma is an indolent, slowly progressive infection of the skin and subcutaneous tissues with nodular swellings and draining sinuses. Actinomycetoma refers to cases of mycetoma caused by filamentous bacteria. Nearly all cases of *Nocardia*-associated actinomycetoma are caused by *N. brasiliensis*. Clinical manifestations usually begin with a nodular swelling, sometimes at a site of local trauma. Lesions **(Fig. 179-5)** typically develop on the feet or lower legs but may involve the hands, the posterior part of the neck, the upper back, the head, and other sites. The nodule eventually breaks down, and a fistula appears, typically followed by others. The fistulae tend to come and go, with new ones forming as old ones disappear. The discharge is serous or purulent, may be bloody, and often contains 0.1 to 2-mm white granules consisting of masses of mycelia (Fig. 179-5). The lesions spread slowly along fascial planes to involve adjacent areas of skin, subcutaneous tissue, and bone. Over months or years, there may be extensive deformation of the affected part. Lesions involving soft tissues are only mildly painful; those affecting bones or joints are more so (Fig. 179-5). Systemic symptoms are absent or minimal, but mycetoma cases are often associated with prolonged, severe disability. Infection rarely disseminates from actinomycetoma, but lesions on the head, neck, and trunk can invade locally to involve deep organs. The World Health Organization has designated mycetoma a neglected tropical disease.

Primary Ocular Disease Primary ocular nocardiosis includes keratitis, scleritis, and exogenous endophthalmitis. These infections typically occur in otherwise healthy individuals. Patients may report recent exposure to soil. Ocular trauma, surgery, and the use of contact lenses are common risk factors. The onset of disease is typically subacute. Signs and symptoms alone cannot distinguish nocardiosis from other ocular diseases. Spread beyond the eye does not occur.

DIAGNOSIS

The first step in diagnosis is examination of sputum or pus for branching, beaded, Gram-positive filaments 1 μm wide and up to 50 μm long **(Fig. 179-6)**. These filaments tend to fragment; this may produce apparent cocci and coccobacilli forms. Most nocardiae are acid-fast in direct smears if a weak acid is used for decolorization (e.g., in the modified Kinyoun, Ziehl-Neelsen, and Fite-Faraco methods). The organisms often take up silver stains. Recovery from specimens containing mixed flora can be improved with selective media (colistin–nalidixic acid agar, modified Thayer-Martin agar, or buffered charcoal–yeast extract agar). Nocardiae grow well on most fungal and mycobacterial media, but procedures used for decontamination of specimens for mycobacterial culture can kill nocardiae and should not be used when nocardiosis is on the differential diagnosis.

Nocardiae grow relatively slowly; colonies may take up to 2 weeks to appear and may not develop their characteristic appearance—white, yellow, or orange, with aerial hyphae—for up to 4 weeks. When the diagnosis of nocardiosis is being considered, clinicians should inform the microbiology laboratory so that cultures can be incubated for prolonged periods of time.

Sputum smears are frequently negative in patients with pulmonary nocardiosis. In such cases, bronchoalveolar lavage fluid or a lung biopsy specimen should be obtained. Isolation of *Nocardia* from a respiratory specimen does not always indicate pulmonary infection. *Nocardia* may colonize the respiratory tract of patients with underlying bronchiectasis or other structural lung diseases. In such cases, isolation of *Nocardia* from a respiratory sample must be considered in the greater clinical context. When nocardiae are present in respiratory specimens from patients with intact host defenses, pulmonary nocardiosis should be diagnosed only when clinical and radiographic features are supportive. Nocardiosis is less likely if Gram-stained specimens are negative and cultures are not consistently positive. In contrast, a positive culture in an immunosuppressed patient usually reflects clinical disease, and empirical antimicrobial therapy should be started.

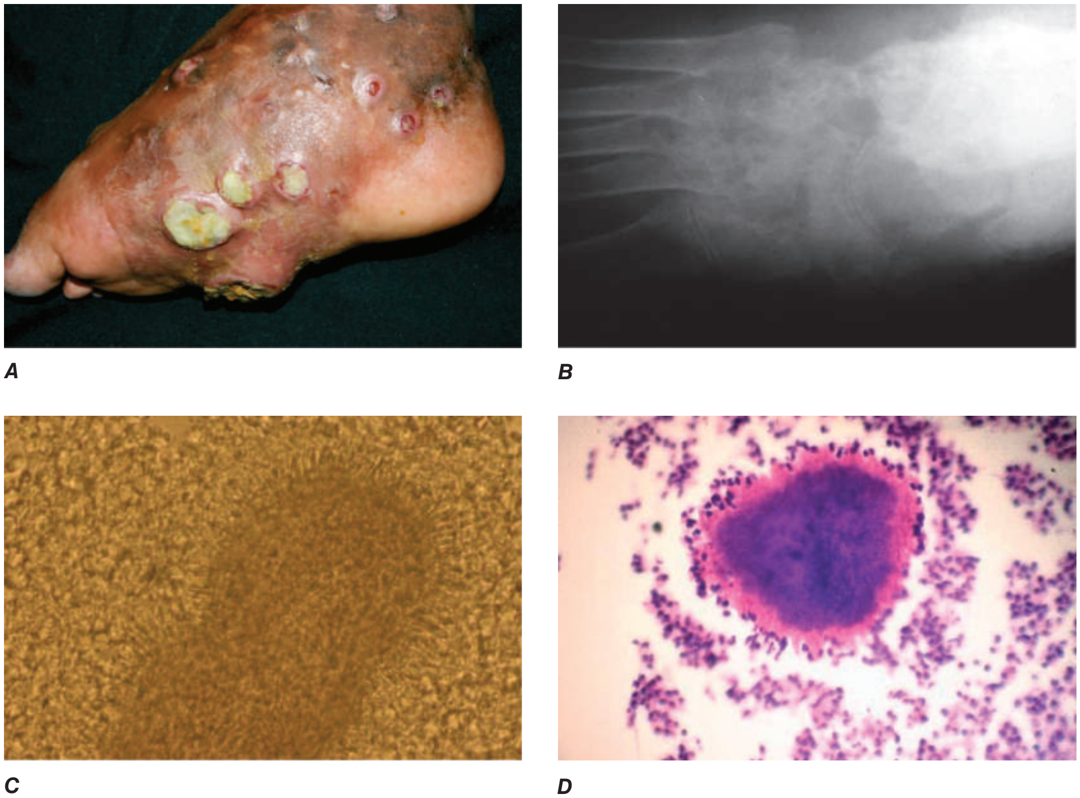

FIGURE 179-5. ***Nocardia brasiliensis* mycetoma.** ***A.*** Draining sinuses and giant white grains with a seropurulent discharge. ***B.*** Radiography of the foot showing marked soft tissue enlargement and bony lytic lesions. ***C.*** Direct microscopy of grains stained with Lugol's iodine (×40). ***D.*** Periodic acid–Schiff stain of skin biopsy (×40). *(Images provided by Roberto Arenas and Mahreen Ameen, St. John's Institute of Dermatology, Guy's & St Thomas' NHS Trust, London, UK. Reprinted from R Arenas, M Ameen: Lancet Infect Dis 10:66, 2010, with permission from Elsevier.)*

Occasionally, *Nocardia* isolated from blood cultures may represent contamination in individuals without risk factors for nocardiosis, particularly those who lack clinical manifestations suggestive of nocardial disease.

Actinomycetoma, eumycetoma (cases involving fungi; **Chap. 225**), and botryomycosis (cases involving cocci or bacilli, often *Staphylococcus aureus*) are difficult to distinguish clinically but are readily distinguished with microbiologic testing or biopsy. Granules should be sought in any discharge. Suspect particles should be washed in saline, examined microscopically, and cultured. Granules in actinomycetoma are usually white, pale yellow, pink, or red. They consist of tight masses of fine filaments (0.5–1 μm wide) radiating outward from a central core (Fig. 179-5). Granules in eumycetoma cases are white, yellow, brown, black, or green; under the microscope, they appear as masses of broader filaments (2–5 μm wide) encased in a matrix. Granules of botryomycosis consist of loose masses of cocci or bacilli. Organisms can also be seen in wound discharge or histologic specimens. Culture is the most reliable method for determining the causative organism in cases of mycetoma.

When possible, isolates should be speciated at least to a complex level, by either matrix-assisted laser desorption/ionization–time of flight mass spectrometry (commonly referred to as MALDI-TOF) or genomic sequencing (typically 16S rRNA, *secA1*, *hsp65, or gyrB*). When possible, antimicrobial susceptibility testing should be performed, ideally with a Clinical Laboratory Standards Institute–approved broth dilution test. E-tests are less definitive. Nocardiae grow more slowly than most clinically important bacteria, and they tend to clump in suspension so that susceptibility test endpoints are difficult to interpret. Thus, experience is required for interpretation of susceptibility testing results.

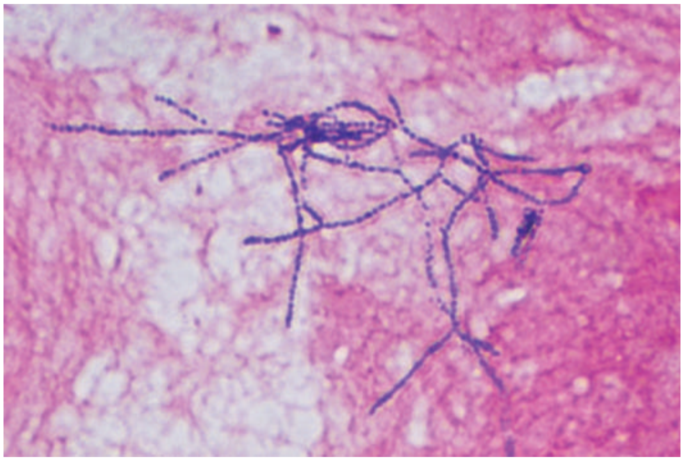

FIGURE 179-6. Gram-stained brain biopsy specimen from a patient with CNS nocardiosis. *(Reproduced with permission from N Hauser et al: An immunocompromised woman with a brain lesion. Am J Med 133:e516-e517.)*

■ EVALUATION OF PATIENTS WITH ESTABLISHED NOCARDIOSIS

Timely identification of metastatic sites of infection is a critical component of managing patients with nocardiosis. Disseminated nocardiosis is exceedingly uncommon in patients with primary cutaneous or ocular disease that develops following traumatic injury. Patients with other forms of nocardiosis, regardless of symptoms, should undergo contrast-enhanced imaging of the CNS to identify undiagnosed brain abscesses. MRI is the preferred imaging modality, with CT reserved for scenarios when MRI cannot be performed.

The presence of a previously undiagnosed immunologic defect should be considered in patients with disseminated nocardiosis who lack typical risk factors for infection. Such patients should undergo testing for HIV infection. Further testing for underlying immunologic disorders should be guided by a careful history and physical examination.

TABLE 179-1 Treatment Duration for Nocardiosis

DISEASE	DURATION
Pulmonary or systemic	
Intact host defenses	6–12 months
Deficient host defenses	12 months[a]
CNS disease	12 months[b]
Primary cutaneous disease (cellulitis, lymphocutaneous disease)	2 months
Osteomyelitis, arthritis, laryngitis, sinusitis	4 months
Actinomycetoma	6–12 months after clinical cure
Keratitis	Topical: until clinical cure Systemic: until 2–4 months after clinical cure

[a] In some patients with AIDS and CD4$^+$ T lymphocyte counts of <200/μL or with chronic granulomatous disease, therapy for pulmonary or systemic disease may be continued indefinitely.

[b] If all apparent central nervous system (CNS) disease has been excised, the duration of therapy may be reduced to 6 months.

TREATMENT

Nocardiosis

Trimethoprim-sulfamethoxazole (TMP-SMX) is the drug of choice for most cases of nocardiosis (Fig. 179-1 and **Table 179-1**). At the outset, 10–20 mg/kg of TMP and 50–100 mg/kg of SMX are given each day in two divided doses. Later, daily doses can be reduced to as little as 5 mg/kg and 25 mg/kg, respectively. In persons with sulfonamide allergies, desensitization usually allows continuation of therapy with these effective and inexpensive drugs.

Linezolid is highly bioavailable and uniformly active against all *Nocardia* species. Treatment-limiting toxicity frequently develops after 2–3 weeks of standard-dose therapy. Tedizolid is associated with less toxicity, but clinical experience is limited.

Amikacin is active against nearly all *Nocardia* species with the notable exceptions of *N. transvalensis* and some *N. pseudobrasiliensis* isolates. Doses of 5–7.5 mg/kg every 12 h or 15 mg/kg every 24 h are typically used. Serum drug levels should be monitored during prolonged therapy, especially in patients with diminished renal function and in the elderly.

Among the β-lactams, ceftriaxone and imipenem have activity against many *Nocardia* species and are frequently used when combination therapy is indicated. Ceftriaxone is not active against *N. farcinica*, which is among the most common species isolated in nocardiosis. In patients with CNS involvement, meropenem is preferred to imipenem as the latter is associated with a higher risk of seizures. Amoxicillin (875 mg) combined with clavulanate (125 mg), given twice a day, has been effective in treating *N. brasiliensis* and some cases of *N. farcinica*.

Among the quinolones, moxifloxacin appears to be most active. Minocycline (100–200 mg twice daily) is often effective; other tetracyclines are usually less effective. Tigecycline appears to be active in vitro against some species, but clinical experience is limited.

Empirical treatment of patients with severe disease should consist of combination therapy with two or three of the following agents: TMP-SMX, linezolid, amikacin, and imipenem (or meropenem if the CNS is involved). Clinical improvement is usually noticeable after 1–2 weeks of therapy but may take longer. After definite clinical improvement, therapy can be continued with a single oral drug, usually TMP-SMX. Some experts use two or more drugs for the entire course of therapy, but whether multiple drugs are better than a single agent is not known, and additional drugs increase the risk of toxicity.

Surgical management of nocardial disease is similar to that of other bacterial diseases. Brain abscesses should be aspirated, drained, or excised if the diagnosis is unclear, if an abscess is large, or if an abscess fails to respond to antimicrobial therapy. Surgical options include needle aspiration and open drainage/excision. Needle aspiration is less invasive, but patients frequently require multiple aspirations. Small or inaccessible brain abscesses can be treated medically. Brain imaging should be repeated to document the resolution of lesions, although radiographic improvement typically lags behind clinical improvement.

With appropriate treatment, the mortality rate for localized pulmonary nocardiosis is <10%. Disseminated nocardiosis carries a higher mortality rate, especially among patients with CNS involvement.

Primary cutaneous and ocular nocardiosis often respond to antimicrobial therapy. Occasionally, surgical management is required for patients with actinomycetoma or scleritis that fails to response to antimicrobial therapy. The mortality rate for actinomycetoma is exceedingly low, but patients suffer often from substantial morbidity, disfigurement, disability, and/or stigmatization. Patients with nocardial keratitis should not receive adjunctive topical corticosteroids.

PREVENTION

Use of TMP-SMX in high-risk populations to prevent *Pneumocystis* disease may reduce but does not eliminate the risk of nocardiosis. The incidence of nocardiosis is low enough that prophylaxis solely to prevent a first episode of *Nocardia* infection (i.e., primary prophylaxis) is not recommended. The role of secondary prophylaxis to prevent recurrent nocardiosis has not been adequately studied. Some experts recommend secondary prophylaxis for patients who remain at high risk for *Nocardia* infection after the completion of antimicrobial therapy.

FURTHER READING

Averbuch D et al: *Nocardia* infections in hematopoietic cell transplant recipients: A multicenter international retrospective study of the Infectious Diseases Working Party of the European Society for Blood and Marrow Transplantation. Clin Infect Dis 75:88, 2022.

Corsini Campioli C et al: Clinical presentation, management, and outcomes of patients with brain abscess due to *Nocardia* species. Open Forum Infect Dis 8:ofab067, 2021.

Hamdi AM et al: Retrospective analysis of antimicrobial susceptibility profiles of *Nocardia* species from a tertiary hospital and reference laboratory, 2011 to 2017. Antimicrob Agents Chemother 64(3):e01868, 2020.

Lebeaux D et al: Outcome and treatment of nocardiosis after solid organ transplantation: New insights from a European study. Clin Infect Dis 64:1396, 2017.

Margalit I et al: How do I manage nocardiosis? Clin Microbiol Infect 27:550, 2021.

Passerini M et al: Trimethoprim-sulfamethoxazole significantly reduces the risk of nocardiosis in solid organ transplant recipients: Systematic review and individual patient data meta-analysis. Clin Microbiol Infect 30:170, 2024.

Restrepo A et al: *Nocardia* infections in solid organ transplantation: Guidelines from the Infectious Diseases Community of Practice of the American Society of Transplantation. Clin Transplant 33:e13509, 2019.

Rosen LB et al: *Nocardia*-induced granulocyte macrophage colony-stimulating factor is neutralized by autoantibodies in disseminated/extrapulmonary nocardiosis. Clin Infect Dis 60:1017, 2015.

Traxler RM et al: Updated review on *Nocardia* species: 2006–2021. Clin Microbiol Rev 35:e0002721, 2022.

Wang H et al: Epidemiology and antimicrobial resistance profiles of the *Nocardia* species in China, 2009 to 2021. Microbiol Spectr 10:2, 2022

Yetmar ZA et al: Mortality after nocardiosis: Risk factors and evaluation of disseminated infection. Open Forum Infect Dis 10:ofad409, 2023.

180 Actinomycosis

Thomas A. Russo, John C. Hu

Actinomycosis is uncommon, and most physicians' personal experience with its clinical presentations is limited. Laboratory identification of the etiologic agents from the order Actinomycetales is not routine. Thus, actinomycosis remains a diagnostic challenge, even for a skilled clinician. However, this infection is usually curable with medical therapy alone. Therefore, an awareness of the full spectrum of clinical syndromes can expedite diagnosis and treatment and minimize unnecessary surgical interventions, morbidity, and mortality.

Classical actinomycosis is an indolent, slowly progressive infection caused by anaerobic or microaerophilic bacteria, primarily of the genus *Actinomyces*, that colonize the mouth, colon, and vagina. Mucosal disruption may lead to infection at virtually any site in the body. In vivo growth of actinomycetes usually results in the formation of characteristic clumps called *grains* or *sulfur granules*. The clinical presentations of actinomycosis are myriad. Common in the preantibiotic era, actinomycosis has diminished in incidence, as has its timely recognition. Actinomycosis has been called the most misdiagnosed disease, and it has been said that no disease is so often missed by experienced diagnosticians.

Three "classic" clinical presentations that should prompt consideration of this unique infection are (1) the combination of chronicity, progression across tissue boundaries, and mass-like features (mimicking malignancy, with which it is often confused); (2) the development of a sinus tract, which may spontaneously resolve and recur; and (3) a refractory or relapsing infection after a short course of therapy, since cure of established actinomycosis requires prolonged treatment.

ETIOLOGIC AGENTS

Actinomycosis is most commonly caused by *A. israelii, A. naeslundii, Schaalia (Actinomyces) odontolyticus, A. viscosus, Schaalia (Actinomyces) meyeri, A. graevenitzii*, and *A. gerencseriae*. Infections due to *Winkia* (*Actinomyces) neuii* have been increasingly recognized. Most if not all actinomycotic infections are polymicrobial. *Aggregatibacter (Actinobacillus) actinomycetemcomitans, Eikenella corrodens*, Enterobacteriaceae, and species of *Fusobacterium, Bacteroides, Capnocytophaga, Staphylococcus*, and *Streptococcus* are commonly isolated with actinomycetes in various combinations, depending on the site of infection. Their contribution to the pathogenesis of actinomycosis is uncertain.

Genome-based analysis and comparative 16S rRNA gene sequencing have led to the identification of an ever-expanding list of *Actinomyces* species and a reclassification of some species to other genera. In recent years, many prior *Actinomyces* species have been placed into new genera including *Schaalia, Winkia, Gleimia,* and *Pauljensenia,* though most publications have yet to adapt these new taxonomic changes. At present, 33 species remain in the *Actinomyces* genus with at least 26 species implicated as causes of human disease. *Gleimia (Actinomyces) europaeus, A. radingae, Schaalia (Actinomyces) turicensis, Schaalia (Actinomyces) cardiffensis, A. urogenitalis, Pauljensenia (Actinomyces) hongkongensis, Schaalia (Actinomyces) georgiae, Schaalia (Actinomyces) massiliensis, A. timonensis, Schaalia* (*Actinomyces) funkei*, *Trueperella* (*Arcanobacterium*) *pyogenes*, *Trueperella* (*Arcanobacterium*) *bernardiae*, and *Propionibacterium propionicum* are additional causes of human actinomycosis, albeit not always with a "classic" presentation.

EPIDEMIOLOGY

Actinomycosis has no geographic boundaries and occurs throughout life, with a peak incidence in the middle decades. Males have a threefold higher incidence than females, possibly because of poorer dental hygiene and/or more frequent trauma. Improved dental hygiene and the initiation of antimicrobial treatment before actinomycosis fully develops have probably contributed to a decrease in incidence since the advent of antibiotics. Individuals who do not seek or have access to health care, those who have an intrauterine contraceptive device (IUD) in place for a prolonged period (see "Pelvic Disease," below), and those who receive bisphosphonate treatment (see "Oral–Cervicofacial Disease," below) are probably at higher risk.

PATHOGENESIS AND PATHOLOGY

The etiologic agents of actinomycosis are members of the normal oral flora and are often cultured from the bronchi, the gastrointestinal tract, and the female genital tract. The critical step in the development of actinomycosis is disruption of the mucosal barrier. Local infection may ensue. Once established, actinomycosis spreads contiguously in a slow, progressive manner, ignoring tissue planes. Although acute inflammation may initially develop at the infection site, the hallmark of actinomycosis is the characteristic chronic, indolent phase manifested by lesions that usually appear as single or multiple indurations. Central necrosis consisting of neutrophils and sulfur granules develops and is virtually diagnostic. The fibrotic walls of the mass are typically described as "wooden." The responsible bacterial and/or host factors have not been identified. Over time, sinus tracts to the skin, adjacent organs, or bone may develop. In rare instances, distant hematogenous seeding may occur; lymphatic spread and associated lymphadenopathy are uncommon. As mentioned above, these unique features of actinomycosis mimic malignancy, with which it is often confused.

Foreign bodies appear to facilitate infection. This association most frequently involves IUDs. Reports have described an association of actinomycosis with HIV infection; transplantation; common variable immunodeficiency; chronic granulomatous disease; treatment with anti–tumor necrosis factor α agents, glucocorticoids, or bisphosphonates; and radio- or chemotherapy. Actinomycosis after SARS-CoV-2 infection is reported but the association is not well-established. Ulcerative mucosal infections (e.g., by herpes simplex virus or cytomegalovirus) may facilitate disease development.

CLINICAL MANIFESTATIONS

Oral–Cervicofacial Disease Actinomycosis occurs most frequently at an oral, cervical, or facial site, usually as a soft tissue swelling, abscess, mass, or ulcerative lesion that is often mistaken for a neoplasm. Dental diseases or procedures are common precipitating factors. The angle of the jaw is generally involved, but a diagnosis of actinomycosis should be considered with any mass lesion or relapsing infection in the head and neck. Radiation therapy and medication-related osteonecrosis of the jaw (MRONJ) due to antiresorptive therapy with bisphosphonates and anti–receptor activator of nuclear factor-κβ ligand (RANKL) such as denosumab, angiogenesis inhibitors, and tyrosine kinase inhibitors have all been recognized as contributing to an increasing incidence of actinomycotic infection of the mandible and maxilla **(Fig. 180-1)**. Canaliculitis (commonly due to *P. propionicum*),

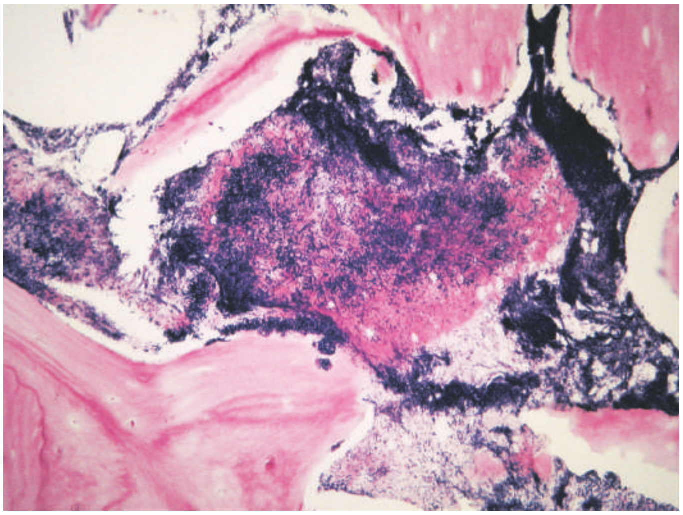

FIGURE 180-1 Bisphosphonate-associated maxillary osteomyelitis due to *Actinomyces viscosus*. A sulfur granule is seen within the bone. *(Reprinted with permission from NH Naik, TA Russo: Bisphosphonate related osteonecrosis of the jaw: The role of Actinomyces. Clin Infect Dis 49:1729, 2009. © 2009 Oxford University Press.)*

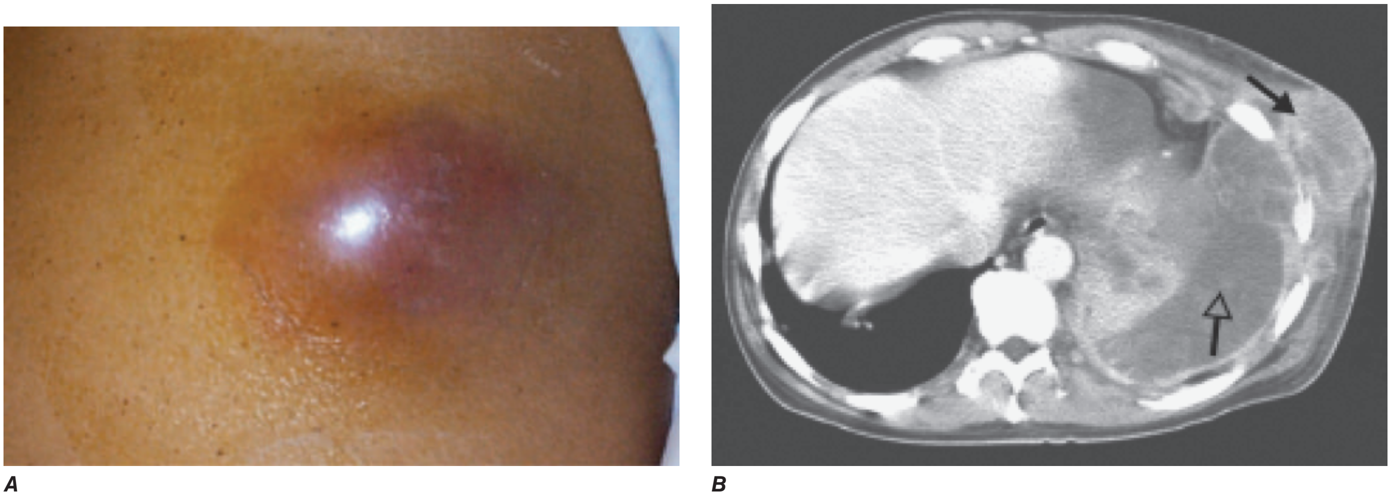

FIGURE 180-2 Thoracic actinomycosis. ***A.*** A chest wall mass from extension of pulmonary infection. ***B.*** Pulmonary infection is complicated by empyema (*open arrow*) and extension to the chest wall (*closed arrow*). *(Courtesy of Dr. C. B. Hsiao, Division of Infectious Diseases, Department of Medicine, State University of New York at Buffalo.)*

otitis, sinusitis, and laryngeal disease also can develop. Pain, fever, and leukocytosis are variably reported. Contiguous extension to the cranium, cervical spine, or thorax is a potential sequela.

Thoracic Disease Thoracic actinomycosis, which may be facilitated by aspirated foreign material such as animal bones or teeth, usually follows an indolent progressive course, with involvement of the pulmonary parenchyma and/or the pleural space. Chest pain, fever, and weight loss are common. A cough, when present, is variably productive. The usual radiographic finding is either a mass lesion or pneumonia. On computed tomography (CT), central areas of low attenuation and ring-like rim enhancement may be seen; cavitary disease may develop. More than 50% of cases include pleural thickening, effusion, or empyema **(Fig. 180-2)**. Rarely, pulmonary nodules or endobronchial lesions occur. Lesions suggestive of actinomycosis include those that cross fissures or pleura; extend into the mediastinum, contiguous bone, or chest wall (*empyema necessitatis or empyema necessitans*); or are associated with a sinus tract. In the absence of these findings, thoracic actinomycosis is usually mistaken for a neoplasm or pneumonia due to more usual causes.

Mediastinal infection is uncommon, usually arising from thoracic extension but rarely from perforation of the esophagus, trauma, or extension of head and neck or abdominal disease. The structures within the mediastinum and the heart can be involved in various combinations; consequently, the possible presentations are diverse. Primary endocarditis (in which *W. neuii* has been increasingly described), esophageal infection, and isolated disease of the breast occur.

Abdominal Disease Abdominal actinomycosis poses a great diagnostic challenge. Months or years usually pass from the inciting event (e.g., appendicitis, diverticulitis, peptic ulcer disease, spillage of gallstones or bile during cholecystectomy, foreign-body perforation, bowel surgery, or ascension from IUD-associated pelvic disease) to clinical recognition. Because of the flow of peritoneal fluid and/or the direct extension of primary disease, virtually any abdominal organ, region, or space can be involved. The disease usually presents as an abscess, a mass, or a mixed lesion that is often fixed to underlying tissue and mistaken for a tumor. On CT, enhancement is most often heterogeneous and adjacent bowel is thickened. Sinus tracts to the abdominal wall, to the perianal region, or between the bowel and other organs may develop and mimic inflammatory bowel disease **(Chap. 337)**. Recurrent disease or a wound or fistula that fails to heal suggests actinomycosis.

Hepatic infection usually presents as one or more abscesses or masses **(Fig. 180-3)**. Isolated disease presumably develops via hematogenous seeding from cryptic foci. Imaging and percutaneous techniques have resulted in improved diagnosis and treatment.

All levels of the urogenital tract can be infected. Renal disease usually presents as pyelonephritis and/or renal and perinephric abscess. Bladder involvement, usually due to extension of pelvic disease, may result in ureteral obstruction or fistulas to bowel, skin, or uterus. *Actinomyces* can be detected in urine with appropriate stains and cultures.

Pelvic Disease Actinomycotic involvement of the pelvis occurs most commonly in association with an IUD but can also be associated with other foreign bodies, such as surgical mesh. When an IUD is in place or

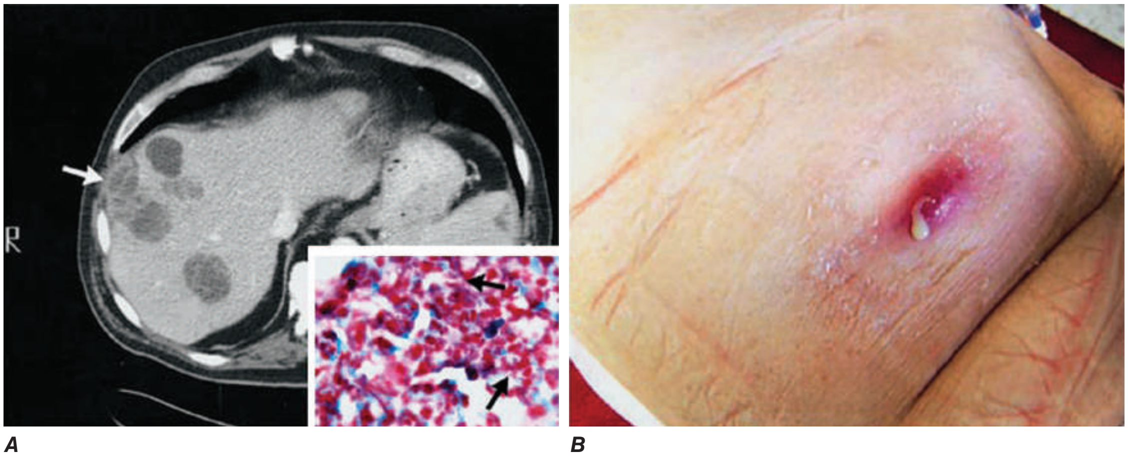

FIGURE 180-3 Hepatic–splenic actinomycosis. ***A.*** Computed tomogram showing multiple hepatic abscesses and a small splenic lesion due to *Actinomyces israelii. Arrow* indicates extension outside the liver. *Inset:* Gram stain of abscess fluid demonstrating beaded filamentous gram-positive rods. ***B.*** Subsequent formation of a sinus tract. *(Reprinted with permission from M Saad: Actinomyces hepatic abscess with cutaneous fistula. N Engl J Med 353:e16, 2005. © 2005 Massachusetts Medical Society. All rights reserved.)*

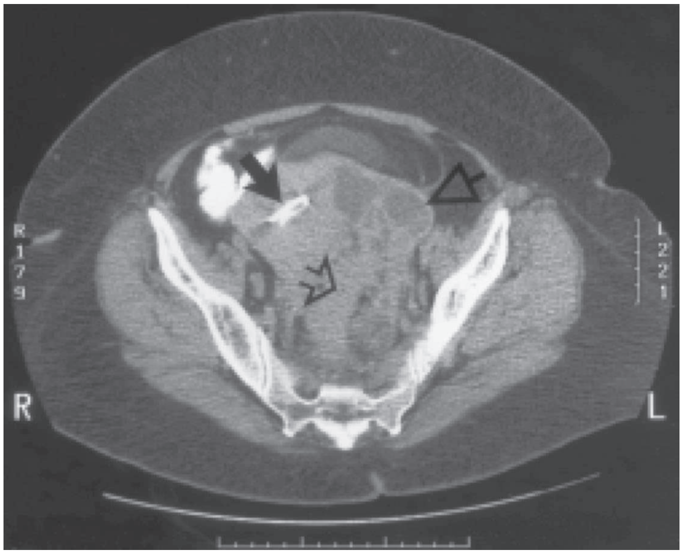

FIGURE 180-4 Computed tomogram showing pelvic actinomycosis associated with an intrauterine contraceptive device. The device is encased by endometrial fibrosis (*solid arrow*); also visible are paraendometrial fibrosis (*open triangular arrowhead*) and an area of suppuration (*open arrow*).

has been used but removed, pelvic symptoms should prompt consideration of actinomycosis. The risk, although not quantified, appears small. The disease rarely develops when the IUD has been in place for <1 year, but the risk increases with time. Symptoms are typically indolent; fever, weight loss, abdominal pain, and abnormal vaginal bleeding or discharge are the most common. The earliest stage of disease—often endometritis—commonly progresses to pelvic masses or a tuboovarian abscess **(Fig. 180-4)**. Unfortunately, because the diagnosis is often delayed, a "frozen pelvis" mimicking malignancy or endometriosis can develop by the time of recognition, which may lead to unnecessary surgery. Cancer antigen 125 levels may be elevated, further contributing to misdiagnosis. In contrast to malignancy and tuberculosis, pelvic actinomycosis only uncommonly includes ascites and lymphadenopathy. An endometrial biopsy may enable diagnosis in a minimally invasive fashion.

Actinomyces-like organisms (ALOs), which are identified in Papanicolaou-stained specimens in (on average) 7% of women using an IUD, have a low positive predictive value for diagnosis. The detection of ALOs in an asymptomatic patient warrants education and close follow-up but not removal of the IUD unless a suitable contraceptive alternative is agreed on. In the presence of symptoms that cannot be accounted for, it seems prudent to remove the IUD and—if advanced disease is excluded—to initiate a 14-day course of empirical treatment for possible early endometritis.

Central Nervous System Disease Actinomycosis of the central nervous system (CNS) is rare. Single or multiple brain abscesses are most common. Individuals with hereditary hemorrhagic telangiectasia are at increased risk for brain abscess with *Actinomyces* as the potential etiologic agent. An abscess usually appears on CT as a ring-enhancing lesion with a thick wall that may be irregular or nodular. Magnetic resonance perfusion and spectroscopy findings have also been described, as have primary meningitis, epidural or subdural space infection, and cavernous sinus syndrome.

Musculoskeletal and Soft Tissue Infection Actinomycotic infection of bones and joints is usually due to adjacent soft tissue infection but may be associated with trauma, injections, surgery (e.g., prostheses), osteoradionecrosis and bisphosphonate osteonecrosis (limited to mandibular and maxillary bones), or hematogenous spread. Because of slow disease progression, new bone formation and bone destruction can be seen concomitantly. Infection of soft tissue is uncommon and is usually a result of trauma. Actinomycetoma is a slowly progressive infection of the skin and subcutaneous tissue that is usually seen in warm climates. Despite the name being suggestive of *Actinomyces* as a causative agent, it is most commonly caused by *Nocardia* or *Actinomadura* species **(Chap. 179)**.

Disseminated Disease Hematogenous dissemination of disease from any location rarely results in multiple-organ involvement. *S. meyeri* is most commonly involved. The lungs and liver are most often affected, with the presentation of multiple nodules mimicking disseminated malignancy. The clinical presentation may be surprisingly indolent given the extent of disease.

■ DIAGNOSIS

The diagnosis of actinomycosis is rarely considered. All too often, actinomycosis is first mentioned by the pathologist after extensive surgery. Since medical therapy alone is frequently sufficient for cure, the challenge for the clinician is to consider the possibility of actinomycosis, to diagnose it in the least invasive fashion, and to avoid unnecessary surgery. The clinical and radiographic presentations that suggest actinomycosis are discussed above. Of note, hypermetabolism has been demonstrated by ^{18}F-fluorodeoxyglucose positron emission tomography (FDG-PET) in actinomycotic disease. Aspirations and biopsies (with or without CT or ultrasound guidance) are being used successfully to obtain clinical material for diagnosis, although surgery may be required. The microscopic identification of sulfur granules (an in vivo matrix of bacteria, calcium phosphate, and host material) in pus or tissues, which increases with the examination of additional histopathologic sections and the use of positively charged slides to optimize adhesion, is the most common means of diagnosis. Occasionally, these granules are identified grossly from draining sinus tracts or pus. Periodic acid–Schiff (PAS), Grocott methenamine silver (GMS), and Gram stains may be helpful to identify actinomycotic aggregates in surgical specimens. On hematoxylin-eosin stain, the granules may be eosinophilic or variably surrounded by a radiating fringe of eosinophilic clubs called the Splendore-Hoeppli phenomenon **(Fig. 180-5)**. Although sulfur granules are a defining characteristic of actinomycosis, granules also are found in mycetoma **(Chaps. 179 and 225)** and botryomycosis (a chronic suppurative bacterial infection of soft tissue or, in rare cases, visceral tissue that produces clumps of bacteria resembling granules). These entities can easily be differentiated from actinomycosis with appropriate histopathologic and microbiologic studies. Microbiologic identification of actinomycetes is often precluded by prior antimicrobial therapy or failure to perform appropriate microbiologic cultures. For optimal yield, the avoidance of even a single dose of antibiotics is mandatory. Although some species can grow aerobically, isolation is maximized under anaerobic conditions, usually requiring 5–7 days but potentially up to 2–4 weeks. The use of 16S rRNA gene amplification and sequencing by clinical microbiology laboratories is increasing and is enhancing diagnostic sensitivity and specificity. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) holds similar promise, but databases are still being optimized. Because actinomycetes are components of the normal oral and genital-tract flora, their identification in the absence of sulfur granules in sputum, bronchial washings, and cervicovaginal secretions may reflect colonization rather than infection.

TREATMENT

Actinomycosis

Decisions about treatment are based on the collective clinical experience of the past 75 years. Actinomycosis requires prolonged treatment with high doses of antimicrobial agents; suitable antimicrobial agents and those deemed unreliable are listed in **Table 180-1**. The need for intensive treatment is presumably due to the drugs' poor penetration of the thick-walled masses common in this infection and/or the sulfur granules themselves, which may represent a biofilm. Although therapy must be individualized, the IV administration of 18–24 million units of penicillin daily for 2–6 weeks, followed by oral therapy with penicillin or amoxicillin (total duration, 6–12 months), is a reasonable guideline for serious infections and bulky disease. For penicillin-allergic patients, tetracyclines, ceftriaxone, or carbapenems are reasonable alternatives. Less extensive disease, particularly that involving the oral–cervicofacial region or

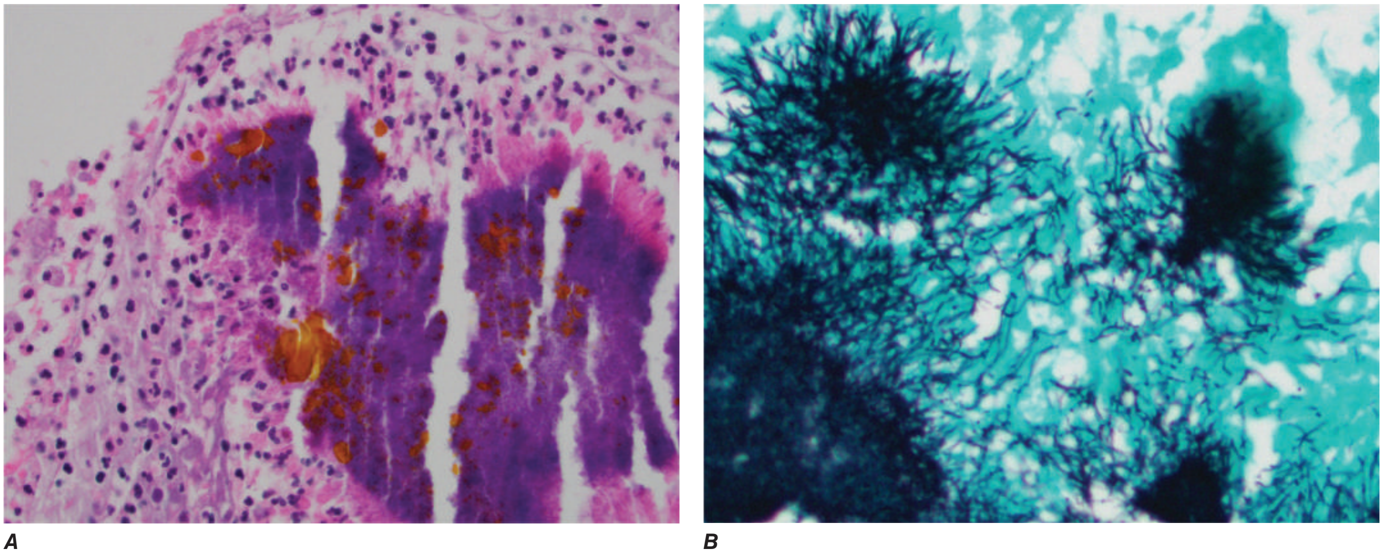

FIGURE 180-5 Microscopic evaluation of actinomycotic sulfur granules. ***A.*** Actinomycotic sulfur granule with gram-positive *Actinomyces* organisms surrounded by eosinophilic, proteinaceous coating called the Splendore-Hoeppli phenomenon. ***B.*** Actinomycotic granule appearance with Grocott methenamine silver stain. *(Courtesy of Ayesha Arshad, MD, VA Western New York Healthcare System.)*

TABLE 180-1 Appropriate and Inappropriate Antibiotic Therapy for Actinomycosis[a]

CATEGORY	AGENT
Extensive successful clinical experience[b]	Penicillin: 3–4 million units IV q4h[c,d] Amoxicillin: 500 mg PO q6h Erythromycin: 500–1000 mg IV q6h or 500 mg PO q6h[c] Tetracycline: 500 mg PO q6h Doxycycline: 100 mg IV or PO q12h Minocycline: 100 mg IV or PO q12h Clindamycin: 900 mg IV q8h or 300–450 mg PO q6h[c]
Anecdotal successful clinical experience	Ceftriaxone[d] Imipenem-cilastatin Piperacillin-tazobactam
Agents predicted to be efficacious on the basis of in vitro activity	Vancomycin Dalbavancin Linezolid Rifampin Ertapenem[d] Meropenem Tigecycline[d] Eravacycline Azithromycin[d]
Agents that should be avoided	Metronidazole Aminoglycosides Oxacillin, dicloxacillin Cephalexin Ceftazidime Daptomycin Fluoroquinolones

[a] Additional coverage for concomitant "companion" bacteria may be required. [b] Controlled evaluations have not been performed. Dose and duration require individualization depending on the host, site, and extent of infection. As a general rule, a maximal parenteral antimicrobial dose for 2–6 weeks followed by oral therapy, for a total duration of 6–12 months, is required for serious infections and bulky disease, whereas a shorter course may suffice for less extensive disease, particularly in the oral–cervicofacial region. Monitoring the impact of therapy with computed tomography or magnetic resonance imaging is advisable when appropriate. [c] Recent in vitro data have demonstrated resistance in up to 33% of isolates. [d] This agent can be considered for at-home parenteral therapy; penicillin requires a continuous infusion pump.

the isolation of *Actinomyces* in the absence of tissue changes associated with actinomycosis, may be cured with a shorter course. For home IV therapy, the ease of once-a-day dosing makes ceftriaxone appealing in certain circumstances; however, a greater body of literature supporting its efficacy would be desirable. The availability of portable infusion pumps for home therapy allows for both the appropriate dosing and practical administration of IV penicillin. For infections in critical sites (e.g., CNS), this approach remains the safest until more information is available on other agents. The pharmacokinetic properties, availability of oral and parenteral formulations, and potential efficacy of azithromycin also make this agent appealing. Unfortunately, few in vitro and no clinical data exist on its use to treat actinomycosis. If therapy is extended beyond the resolution of measurable disease, the risk of relapse—a clinical hallmark of this infection—will be minimized; CT and magnetic resonance imaging (MRI) are generally the most sensitive and objective techniques by which to accomplish this goal. A similar approach is reasonable for immunocompromised patients, although refractory disease has been described in HIV-infected individuals. While the role played by "companion" microbes in actinomycosis is unclear, many isolates are pathogens in their own right, and a regimen covering these organisms during the initial treatment course is reasonable. Isolation of *Actinomyces* from blood cultures in the absence of defined infection may represent contamination or transient bacteremia from a mucosal site of colonization, in which case treatment may not be necessary.

Combined medical–surgical therapy is still advocated in some reports. However, an increasing body of literature now supports an initial attempt at cure with medical therapy alone, even in extensive disease. CT and MRI should be used to monitor the response to therapy. In most cases, either surgery can be avoided or a less extensive procedure can be used. This approach is particularly valuable in sparing critical organs, such as the bladder or the reproductive organs in women of childbearing age. For a well-defined abscess, percutaneous drainage in combination with medical therapy is a reasonable approach. When a critical location is involved (e.g., the epidural space, the CNS), when there is significant hemoptysis, or when suitable medical therapy fails, surgical intervention may be appropriate. In the absence of optimal data, the combination of a prolonged course of antimicrobial therapy and resection—at least of necrotic bone for radiation- and medication-related osteonecrosis of the jaw—is a reasonable approach.

FURTHER READING

Barberis C et al: Antimicrobial susceptibility of clinical isolates of *Actinomyces* and related genera reveals an unusual clindamycin resistance among *Actinomyces urogenitalis* strains. J Glob Antimicrob Resist 8:115, 2017.

Bonnefond S et al: Clinical features of actinomycosis: A retrospective, multicenter study of 28 cases of miscellaneous presentations. Medicine 95:e3923, 2016.

Brody A et al: Targeted histological evaluation shows high incidence of *Actinomyces* infection in medication-related osteonecrosis of the jaws. Sci Rep 12:3406, 2022.

Fong P et al: Identification and diversity of *Actinomyces* species in a clinical microbiology laboratory in the MALDI-TOF MS era. Anaerobe 54:151, 2018.

Heo SH et al: Imaging of actinomycosis in various organs: A comprehensive review. Radiographics 34:19, 2014.

Jeffery-Smith A et al: Is the presence of *Actinomyces* spp. in blood culture always significant? J Clin Microbiol 54:1137, 2016.

Karanfilian KM et al: Cervicofacial actinomycosis. Int J Dermatol 59:1185, 2020.

Kononen E, Wade WG: *Actinomyces* and related organisms in human infections. Clin Microbiol Rev 28:419, 2015.

Lo Muzio L et al: The contribution of histopathological examination to the diagnosis of cervico-facial actinomycosis: A retrospective analysis of 68 cases. Eur J Clin Microbiol Infect Dis 33:1915, 2014.

Lynch T et al: Species-level identification of *Actinomyces* isolates causing invasive infections: Multiyear comparison of Vitek MS (matrix-assisted laser desorption ionization-time of flight mass spectrometry) to partial sequencing of the 16S rRNA gene. J Clin Microbiol 54:712, 2016.

Qiu L et al: Pulmonary actinomycosis imitating lung cancer on (18) F-FDG PET/CT: A case report and literature review. Korean J Radiol 16:1262, 2015.

Yang WT, Grant M: *Actinomyces neuii*: A case report of a rare cause of acute infective endocarditis and literature review. BMC Infect Dis 19:511, 2019.

181

Whipple Disease

Thomas A. Russo, Seth R. Glassman

Whipple disease (WD), described by George Whipple in 1907, is a chronic infection caused by *Tropheryma whipplei*. Most commonly, years pass from the onset of symptoms to the recognition of the disease because of its rarity, its various manifestations mimicking other conditions, and the need to perform nonroutine diagnostic tests. The long-held belief that WD is an infection was supported by observations on its responsiveness to antimicrobial therapy in the 1950s and the identification of bacilli via electron microscopy in small-bowel biopsy specimens in the 1960s. This hypothesis was finally confirmed by amplification and sequencing of a partial 16S rRNA polymerase chain reaction (PCR)–generated amplicon from duodenal tissue in 1991. The subsequent successful cultivation of *T. whipplei* enabled whole genome sequencing and the development of additional diagnostic tests. The development of PCR-based diagnostics has broadened our understanding of both the epidemiology of and the clinical syndromes attributable to *T. whipplei*. Exposure to *T. whipplei*, which appears to be much more common than has been appreciated, can be followed by asymptomatic carriage, acute disease, or chronic infection. Chronic infection—WD—is a rare development after exposure. "Classic" WD is manifested by some combination of arthralgias/arthritis, weight loss, chronic diarrhea, abdominal pain, and fever. Variable involvement at other sites also occurs; neurologic and cardiac disease are most common. Acute infection and chronic organ disease in the absence of intestinal involvement (see "Isolated Infection," below) are described with increasing frequency. Since untreated WD is often fatal and delayed diagnosis may lead to irreparable organ damage (e.g., in the central nervous system [CNS]), knowledge of the clinical scenarios in which WD should be considered and of an appropriate diagnostic strategy is mandatory.

ETIOLOGIC AGENT

T. whipplei is a weakly staining gram-positive bacillus. Genomic sequence data have revealed that the organism has a small (<1-megabase) chromosome, with many biosynthetic pathways absent or incomplete. This finding is consistent with a host-dependent intracellular pathogen or a pathogen that requires a nutritionally rich extracellular environment. It is one of the slowest growing human pathogens, with a doubling time of 18 days. A genotyping scheme based on a variable region has disclosed >100 genotypes to date. All genotypes appear to be capable of causing similar clinical syndromes.

EPIDEMIOLOGY

WD is rare but has been increasingly recognized since the advent of PCR-based diagnostic tools. Prevalence had been previously estimated at 1–3 cases per 1 million population, although a recent U.S. epidemiologic survey places the number closer to 10 cases per million. Seroprevalence studies indicate that ~50% of Western Europeans and ~75% of Africans from rural Senegal have been exposed to *T. whipplei*. Higher prevalence may be attributable to differences in sanitation. Humans are the only known host. In most studies, males more commonly develop WD; WD is more common in Caucasians and increases with age. To date, no clear animal or environmental reservoir has been demonstrated. However, the organism has been identified by PCR in sewage water and human feces. Workers with direct exposure to sewage are more likely to be asymptomatically colonized than controls, a pattern suggesting fecal–oral spread. Fecal PCR detection rates of 38% among family members of carriers or patients with infection support oral–oral or fecal–oral spread, although a common environmental exposure cannot be excluded. Further, the development of acute *T. whipplei* pneumonia in children raises the possibility of droplet or airborne transmission.

PATHOGENESIS AND PATHOLOGY

Rates of asymptomatic carriage of *T. whipplei* are far higher than rates of chronic infection (<0.01% of those exposed). Both decreased host pathogen-specific inflammatory response and pathogen-driven modulation of host inflammatory response likely play a role in establishing chronic infection. The human leukocyte antigen (HLA) alleles DRB1(*)13 and DQB1(*)06, which stimulate humoral rather than cell-mediated immune responses, are associated with an increased risk of infection. However, only a minority of infected patients possess these haplotypes, suggesting a role for other host factors. IRF4, a transcription factor involved with the immune response, could be such a factor as evidenced by four related family members with WD who possessed IFR4 haploinsufficiency due to a loss-of-function mutation; the distribution of WD in this extended family was consistent with an autosomal dominant trait with incomplete penetrance.

Flow cytometry performed in WD patients demonstrates B-cell subset abnormalities when compared to matched controls. Chronic infection is associated with an impaired T_H1 response, enhanced production of anti-inflammatory cytokines, increased activity of regulatory T cells, M2 polarization of macrophages with diminished antimicrobial activity and impaired phagosome–lysosome fusion and ensuing apoptosis, and blunted development of *T. whipplei*–specific T cells. Therapies that blunt cell-mediated host immune responses (e.g. systemic glucocorticoids or anti–tumor necrosis factor α [TNF-α] agents) may accelerate progression of chronic disease. Impaired cell-mediated immunity may play a role in establishing chronic carriage of *T. whipplei*, as is evidenced by higher rates of detection in the secretions of HIV-infected persons.

T. whipplei has a tropism for myeloid cells, which it invades and in which it can avoid being killed. Infiltration of infected tissue by large numbers of foamy macrophages containing periodic acid–Schiff (PAS)–staining inclusions (representing ingested bacteria) is a characteristic and most common finding. With gastrointestinal disease progression, villus atrophy, lymphangiectasia, crypt hyperplasia, and apoptosis of surface epithelial cells are observed in the small intestine, with resultant diarrhea due to decreased absorption and increased leak of water and solutes. Occasionally, involvement of lymphatic or hepatic tissue may manifest as noncaseating granulomas that can mimic sarcoid or granulomatous vasculitis.

CLINICAL MANIFESTATIONS

Asymptomatic Colonization/Carriage Studies using primarily PCR have detected *T. whipplei* sequence in stool, saliva, duodenal tissue, and (rarely) blood in the absence of symptoms. Although prevalence rates are still being defined, in Western European countries, detection in saliva (0.2%) is less common than that in stool (1–11%) and appears to occur only with concomitant fecal carriage. The prevalence of fecal carriage is elevated among individuals with exposure to waste water or sewage (12–26%) and among children living in tropical Africa and Asia (20–48%). A duration of carriage of 7 years for the same strain has been described in a sewer worker. Evolution of the carrier state into chronic disease is uncommon. Bacterial loads are lighter in asymptomatic carriage than in active disease.

Acute Infection *T. whipplei* has been implicated as a cause of acute gastroenteritis in children. It was also detected via PCR in the blood of 4.6% of febrile patients (75% of whom were <15 years of age) from two rural villages in Senegal as opposed to 0.25% of healthy controls. Further, *T. whipplei* has been implicated as a cause of acute pneumonia. These data suggest that primary acquisition may result in symptomatic pulmonary or intestinal infection or a febrile syndrome, which perhaps are more common than is generally appreciated.

Chronic Infection • **"CLASSIC" WD** So-called classic WD was the initial clinical syndrome recognized, with consequent identification of *T. whipplei*. This chronic infection is defined by involvement of the duodenum and/or jejunum that develops over years. In most individuals, the initial phase of disease manifests primarily as intermittent, often symmetrical, occasionally chronic, and rarely destructive migratory oligo- or polyarthralgias/seronegative arthritis involving the knees, wrists, ankles, and metacarpal-interphalangeal joints most commonly. Less frequently, spondylitis, sacroiliitis, discitis, tenosynovitis, bursitis, and prosthetic hip infection also have been described. Intermittent fever, myalgias, and skin nodules may accompany joint symptoms. Tests for rheumatoid factor and antinuclear antibody are usually negative. This initial stage is often confused with a variety of rheumatologic disorders and, on average, lasts 6–8 years before gastrointestinal symptoms commence. Treatment of presumed inflammatory arthritis with immunosuppressive agents (e.g., glucocorticoids, anti-TNF-α, anakinra) can accelerate progression of the disease process; thus, screening for WD prior to initiation of immunosuppressant therapy may be appropriate, depending on the clinical scenario. Alternatively, antimicrobial therapy for another indication may reduce symptoms, and this situation should also prompt consideration of WD. The intestinal symptoms that develop in the majority of cases are characterized by diarrhea with accompanying weight loss and may be associated with fever and abdominal pain. Occult gastrointestinal blood loss, vitamin deficiencies, hepatosplenomegaly (10–15%), and ascites (10%) are less common. Anemia and hypereosinophilia may be detected. The most common finding on abdominal computed tomography is mesenteric and/or retroperitoneal lymphadenopathy (usually raising concern about lymphoma). The endoscopic or video-capsule observation of pale, yellow, or shaggy mucosa with erythema or ulceration past the first portion of the duodenum suggests WD **(Fig. 181-1)**. When endoscopy with duodenal biopsy is nondiagnostic, a video-capsule study may assist in identifying more distal lesions for subsequent biopsy. ^{18}F-Fluorodeoxyglucose positron emission tomography (FDG-PET) studies in patients with WD suggest the entire small bowel can be involved. Diagnostic misdirection can be caused by co-infection with *Giardia lamblia*, which is occasionally identified. The intestinal phase can also be confused with Crohn or celiac disease. In addition to rheumatologic and intestinal disease, neurologic (6–63%), cardiac (17–55%), pulmonary (10–50%), lymphatic (10–55%), ocular (5–10%), dermal (5–30%), and less commonly other sites are variably involved in classic WD.

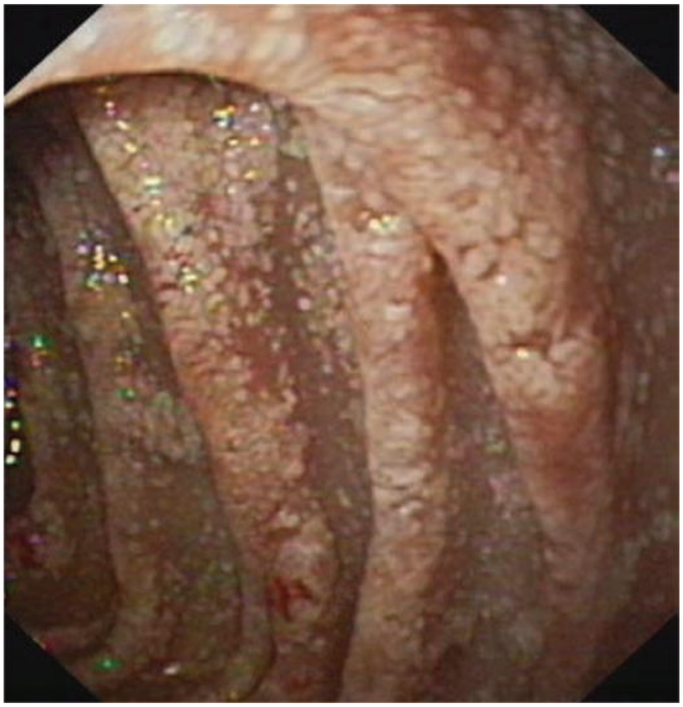

FIGURE 181-1 Endoscopic view of the jejunal mucosa demonstrating a thickened, granular mucosa and "white spots" due to dilated lacteals. *(Reprinted with permission from J Bureš et al: Whipple's disease: Our own experience and review of the literature. Gastroenterol Res Pract, 2013.)*

Neurologic Disease CNS disease, defined by PCR-based detection of *T. whipplei* in cerebrospinal fluid (CSF), develops in ~50% of patients, many of whom are asymptomatic. A variety of neurologic manifestations have been reported and portend a poor prognosis. The most common are cognitive changes including memory impairment progressing to dementia, personality and mood alterations, hypothalamic involvement (e.g., polyuria/polydipsia, sleep-cycle disorders), and supranuclear ophthalmoplegia. In addition, neuro-ophthalmologic manifestations of WD include supranuclear gaze palsy (usually vertical), oculomasticatory and oculofacial myorhythmia (highly suggestive of WD), nystagmus, and retrobulbar neuritis. Focal neurologic presentations (dependent on lesion location), seizures, ataxia, meningitis, encephalitis, rhombo- or limbic encephalitis, hydrocephalus, myelopathy, myoclonus, choreiform movements, and distal polyneuropathy also have been described. Neurologic sequelae occur with CNS disease, and the mortality risk is significant.

Magnetic resonance imaging (MRI) results may be normal. Identified lesions (solitary or multifocal) are usually T2 and fluid-attenuated inversion recovery (FLAIR) hyperintense and may enhance with gadolinium. All sites can be involved, and the nature of lesions is variable (e.g., nodular, infiltrative, tumor-like). Although imaging findings are myriad and are not diagnostic, the median temporal lobe, midbrain, hypothalamus, and thalamus are commonly affected. FDG-PET may reveal increased uptake. CSF analysis may be normal; when abnormal, leukocytosis (generally lymphocyte-predominant) and an elevated protein concentration are common. A low CSF glucose level has been reported.

Cardiac Disease Endocarditis is increasingly recognized in WD (85% of cases in males), causes 2.6–6.3% of culture-negative endocarditis cases, and may be complicated by congestive heart failure (40% of cases), embolic events, arrhythmias, mycotic aneurysm, or rarely hypotension. Fever is often absent, and the Duke clinical criteria are rarely met.

Vegetations are identified by echocardiography in 50–75% of cases. All valves, alone or in combination, can be affected; most commonly involved are the aortic and mitral valves. Preexisting valvular disease is found in only a minority of cases, although infection of bioprosthetic valves has been described. Mural, myocardial, aortic (aortitis), or pericardial disease also occurs alone or in combination with valvular involvement. Constrictive pericarditis develops infrequently. The diagnosis of cardiac disease is rarely made prior to surgical intervention.

Pulmonary Disease Some combination of interstitial disease, nodules, parenchymal infiltrate, and pleural effusion is observed. An association with pulmonary hypertension has also been reported. The clinical significance of *T. whipplei* sequence identified in bronchoalveolar lavage fluid (BALF) from asymptomatic HIV-infected individuals or in a case of interstitial lung disease is unresolved but suggests caution in diagnosing "isolated" pneumonia based on sequence alone. Notably, while the bacterium seems to exist in the airways of HIV-infected persons at higher rates, its presence is not clearly associated with increased inflammation or a discernible decrease in lung function.

Lymphatic Disease Mesenteric and retroperitoneal lymphadenopathy are common with intestinal disease, and mediastinal adenopathy may be associated with pulmonary infection. Peripheral adenopathy is less common.

Ocular Disease (Non–Neuro-Ophthalmologic) Uveitis is the most common form of ocular disease, usually presenting as a change in vision or "floaters." Anterior (anterior chamber), intermediate (vitreous), and posterior (retina/choroid) uveitis can occur alone or in combination. Treatment with glucocorticoids alone can worsen uveitis and unmask extraocular disease. Likewise, use of local or systemic glucocorticoids after ocular surgery can precipitate ocular infection, likely as a result of asymptomatic or subclinical disease. Keratitis, crystalline keratopathy, and optic neuritis also have been reported. Patients may be misdiagnosed with sarcoid or Behçet's disease prior to the recognition of WD.

Dermatologic Disease Skin hyperpigmentation (melanoderma), particularly in light-exposed areas in the absence of adrenal dysfunction, is suggestive of WD. A variety of other cutaneous manifestations have been described, including erythematous macular lesions, nonthrombocytopenic purpura, subcutaneous nodules, and hyperkeratosis.

Miscellaneous Sites Thyroid, renal, testicular, epididymal, gallbladder, skeletal muscle, and bone marrow involvement and membranous nephropathy have all been described. In fact, almost any organ can be involved in classic WD, with varying frequency, variable combinations, and myriad signs and symptoms. As a result, WD should be considered in the setting of a chronic multisystemic process. Despite its rarity, the combination of rheumatologic and intestinal disease with weight loss, with or without neurologic and cardiac involvement, warrants heightened suspicion.

ISOLATED INFECTION This entity has been defined as infection in the absence of intestinal symptoms, although an occasional small-bowel biopsy may be PAS-positive or more commonly PCR-positive in this setting. "Isolated infection" is something of a misnomer since multiple nonintestinal sites of *T. whipplei* infection are not uncommon. Infection at the same nonintestinal sites (single or multiple) that are variably involved in classic WD may also present as "isolated infection." Further, intestinal disease can subsequently develop. Endocarditis, neurologic disease, uveitis, rheumatologic manifestations, and pulmonary involvement are most commonly described. Signs and symptoms are similar to those described for *T. whipplei* infection of these sites in classic WD. With enhanced PCR-based diagnostic capabilities, *T. whipplei* infection without concomitant intestinal involvement (of which endocarditis is the best example) will probably be diagnosed increasingly often.

REINFECTION/RELAPSING DISEASE/IMMUNE RECONSTITUTION INFLAMMATORY SYNDROME (IRIS) It has been suggested that, if an underlying host immune defect places an individual at risk for chronic infection, then that person may be at risk for reinfection due to occupational exposure or contact with family members who are asymptomatically colonized. One case of apparent reinfection that was due to a different genotype supports this contention.

Optimal treatment regimens and durations are still being defined. However, it is clear, especially in the setting of occult or overt CNS disease, that treatment with oral tetracycline or trimethoprim-sulfamethoxazole (TMP-SMX) alone may result in disease relapse. Relapses or perhaps reinfections occurring years to decades after initial therapy have been described.

As in patients treated for HIV or mycobacterial disease, IRIS has been described in up to 17% of patients treated for *T. whipplei* infection. Prior immunosuppressive therapy increases the likelihood of IRIS, in which inflammation recurs after an initial clinical response to treatment and loss of PCR detection of *T. whipplei*. In patients with chronic WD, IRIS may be related to sustained disruption of the epithelial barrier, leading to increased translocation of gut-derived microbial products and dysbalanced T-cell restoration. Manifestations include the development of fever, arthritis, skin lesions, subcutaneous nodules, pleuritis, uveitis, and orbital and periorbital inflammation; some cases have been fatal.

■ DIAGNOSIS

Considering *T. whipplei* infection and ensuring that the appropriate tests are performed are the critical steps in making the diagnosis, which otherwise will likely be missed. Serology is of little value since patients with active infection usually mount a poor IgM/IgG response to *T. whipplei* and a positive result most likely reflects prior exposure and clearance. The clinical presentation will in part dictate which clinical specimens are most likely to enable the diagnosis. In the presence (and perhaps the absence) of gastrointestinal symptoms, postbulbar duodenal biopsies should be performed, although a normal macroscopic appearance is common. As a general rule, the diagnostic yield is greater for tissue specimens than for body fluids. Biopsy of normal-appearing skin may detect *T. whipplei* in the setting of classic WD and serve as a minimally invasive means to establish the diagnosis. It is prudent to collect CSF even in the absence of CNS symptoms; asymptomatic disease is common, the CNS is the most common site for relapse, and thus the information gained by CSF examination could influence the design and duration of the treatment regimen.

The diagnosis of classic WD was originally based on histologic findings in intestinal biopsy specimens. Although this diagnostic procedure remains important, it is not optimally sensitive. Infiltration of the lamina propria with macrophages containing PAS-positive inclusions that are resistant to diastase is observed. However, PAS is nonspecific, also yielding positive results with mycobacteria as well as other microorganisms. Staining of other tissues or fluids (e.g., ocular aspirations) for PAS-positive inclusions in macrophages can be performed to support the diagnosis. The sensitivity of identification of PAS-positive inclusions in WD may be decreased by anti-TNF-α therapy. Electron microscopy can be used to identify the trilaminar cell wall of *T. whipplei*. When available, immunohistochemistry has greater specificity and sensitivity than PAS staining and can be performed on archived fixed tissue. Alternatively, the use of fluorescence in situ hybridization (FISH) has been reported as a complementary diagnostic tool with various tissue samples.

The development and implementation of specific PCR-based diagnostics have significantly increased the sensitivity and specificity of *T. whipplei* identification. PCR can be applied to affected tissues (with greater sensitivity for non-formalin-fixed than for formalin-fixed tissue) in support of histologic findings and to various body fluids. It is important to note that the interpretation of a PCR-based diagnostic approach must take into account limitations such as false-positive results due to sample contamination, false-negative results due to low organism load, poor sample quality, inadequate DNA extraction, and variability in performance of various PCR assays. Quantitative comparisons from different sites can add specificity to PCR-based diagnostics of WD and distinguish between WD patients and asymptomatic carriers. In patients suspected of having WD, PCR testing of duodenal biopsy specimens with a cycle threshold value of ≤30 can help confirm the diagnosis, even in cases with negative PAS staining. As with all diagnostic tests, consideration of pretest probability is critical for interpretation, and a negative result does not exclude WD. Urine PCR for *T. whipplei* infection may hold promise for the noninvasive diagnosis of classic and isolated WD. In one study of 12 cases, urine PCR was positive in nine cases (75%) prior to treatment compared to zero (0%) of 110 controls, including 11 controls that were presumed carriers in whom feces PCR was positive, although there was no evidence of disease. In addition, urine PCR is a potential tool to evaluate the success of WD therapy. Saliva and fecal PCR are inappropriate as the sole diagnostic tools for WD due to low positive predictive values, which more commonly identify colonization, not disease; a positive result requires confirmation from an appropriate end-organ tissue or body fluid.

Next-generation sequencing techniques to evaluate for cell-free DNA (cfDNA) in plasma may lead to increased recognition of *T. whipplei* as a cause of endocarditis.

T. whipplei has been successfully cultured from blood, CSF, synovial fluid, BALF, valve tissue, duodenal tissue, skeletal muscle, and lymph nodes, but culture is not practical since it takes months to obtain a positive result.

Affected anatomic sites in WD patients may demonstrate uptake on FDG-PET, which in turn could guide tissue sampling for use in specific tests.

TREATMENT

Whipple Disease

Data on treatment are emerging, but the optimal regimen and duration for chronic infection, which may depend on the sites involved (e.g., CNS and heart valve), are unclear. Appropriate treatment usually results in a rapid—and at times remarkable—clinical response (e.g., in CNS disease), but eradication requires prolonged treatment. Maintenance of a durable response has been more challenging because of both relapse and host predisposition to reinfection.

Rates of relapse, particularly of CNS disease, were unacceptable with oral tetracycline or TMP-SMX monotherapy. Sequence data now indicate that TMP is not active against *T. whipplei* (given the absence of dihydrofolate reductase in *T. whipplei*) and that resistance to SMX and sulfadiazine can occur. However, a randomized controlled trial in 40 patients, who received either ceftriaxone (2 g IV q24h) or meropenem (1 g IV q8h) for 2 weeks followed by oral TMP-SMX (160/800 mg) twice a day for 1 year, demonstrated outstanding efficacy. The only case in which therapy failed—an asymptomatic CNS infection that was not eradicated by either regimen—was subsequently cured with oral minocycline and chloroquine (250 mg/d after a loading dose). A follow-up trial reported similar efficacy with a regimen of ceftriaxone (2 g IV q24h) for 2 weeks followed by oral TMP-SMX for 3 months. One issue in these trials was that the doses—and perhaps the duration of ceftriaxone and meropenem treatment as well—were not optimal for CNS infection. By contrast, in a small retrospective series, outcome was better in patients treated with oral doxycycline (100 mg twice a day) plus hydroxychloroquine (200 mg three times a day to raise phagosome pH and increase drug activity in vitro) than in patients initially treated with TMP-SMX.

Until more data become available, it seems prudent—at least in asymptomatic/symptomatic CNS disease (which is present in many cases of WD)—first to administer CNS-optimized doses of IV ceftriaxone (2 g q12h) or meropenem (2 g q8h) for 2–4 weeks and then to treat with oral doxycycline, or minocycline plus hydroxychloroquine for at least 1 year, if tolerated. Although TMP-SMX has been frequently used as the oral alternative with reported success, a number of relapses or reinfections with TMP-SMX treatment have been reported, thereby suggesting caution for its use in patients with infection in critical locations such as the CNS and the heart. Although data on the use of PCR to guide therapy do not exist, it seems reasonable that continued *T. whipplei* detection by PCR, especially in the CSF and perhaps urine, should dictate at least continuation of therapy or perhaps consideration of an alternative regimen when in conjunction with a poor clinical response.

Timely recognition may result in cure with medical management alone. Surgery may be needed in the setting of endocarditis with significant valve dysfunction or myocardial abscess. Current European guidelines for the treatment of endocarditis caused by *T. whipplei* recommend oral doxycycline plus hydroxychloroquine for ≥18 months or, alternatively, ceftriaxone (2 g q24h IV) or penicillin (2 million units q4h IV) plus streptomycin (1 g q24h IV) for 2–4 weeks followed by oral TMP-SMX (800 mg q12h); a small study from Spain reported that treatment durations of 12–13 months with these regimens or variations were efficacious.

Data on isolated infection and certain site-specific treatment issues are even more limited. Anecdotal reports describe successful treatment of uveitis with oral TMP-SMX with or without rifampin, whereas treatment with tetracycline alone has resulted in relapse. Although a role for adjunctive intraocular therapy has been reported, the data are unclear on this point. There is a single case report of clearance of infection in a chronically relapsing patient by the addition of interferon gamma to antimicrobials; supplementation to antimicrobials may be a consideration to address refractory disease or potential issues with antibiotic resistance.

Although data on the treatment of foreign body–associated infection are virtually nonexistent, medical treatment for a prosthetic hip infection was apparently successful; however, follow-up was limited.

The occurrence of a Jarisch-Herxheimer reaction within 24 h of treatment initiation has been described, with rapid resolution. The addition of glucocorticoids may be beneficial in the management of IRIS, and thalidomide has been used in steroid-refractory cases.

Importantly, although data are lacking, due to the inherent risk of relapse or reinfection, lifelong suppressive therapy with doxycycline after completion of the initial treatment regimen has been advocated. Regardless of the therapeutic approach chosen, an effort to ensure compliance and close follow-up for potential relapse or reinfection, which can occur many years after an apparent cure, will maximize the chances for a good outcome.

PREVENTION

Attempts are underway to develop a peptide-based vaccine against *T. whipplei*. However, the role for this may be limited given the relative rarity of infection.

■ FURTHER READING

Bally JF et al: Systematic review of movement disorders and oculomotor abnormalities in Whipple's disease. Mov Disord 33:1700, 2018.

Boumaza A et al: Whipple's disease and *Tropheryma whipplei* infections: From bench to bedside. Lancet Infect Dis 22:e280, 2022.

Crews NR et al: Diagnostic approach for classic compared with localized Whipple disease. Open Forum Infect Dis 5:ofy136, 2018.

Damaraju D et al: Clinical problem-solving: A surprising cause of chronic cough. N Engl J Med 373:561, 2015.

de Vries CR et al: 710. Non-invasive diagnosis of Whipple endocarditis using next-generation sequencing for microbial cell-free DNA in plasma. Open Forum Infect Dis 7:S407, 2020.

Guérin A et al: IRF4 haploinsufficiency in a family with Whipple's disease. Elife 7:e32340, 2018.

Gunther U et al: Gastrointestinal diagnosis of classical Whipple disease: Clinical, endoscopic, and histopathologic features in 191 patients. Medicine 94:e714, 2015.

Ioannou P et al: Whipple's disease-associated infective endocarditis: A systematic review. Infect Dis (Lond) 55:447, 2023.

Lagier JC, Raoult D: Whipple's disease and *Tropheryma whipplei* infections: When to suspect them and how to diagnose and treat them. Curr Opin Infect Dis 31:463, 2018.

Meyer S et al: Contribution of PCR to differential diagnosis between patients with Whipple disease and *Tropheryma whipplei* carriers. J Clin Microbiol 61:e0145722, 2023.

Moter A et al: Potential role for urine polymerase chain reaction in the diagnosis of Whipple's Disease. Clin Infect Dis 68:1089, 2019.

182 Infections Due to Mixed Anaerobic Organisms

Neeraj K. Surana, Dennis L. Kasper

Anaerobes constitute the predominant class of bacteria of the normal human microbiota that reside on mucous membranes and predominate in many infectious processes, particularly those arising from mucosal surfaces. These organisms generally cause disease subsequent to the breakdown of mucosal barriers and the leakage of the microbiota into normally sterile sites. Infections resulting from contamination by the microbiota are usually polymicrobial and involve both aerobic and anaerobic bacteria. However, the difficulties encountered in handling specimens in which anaerobes may be important and the technical challenges entailed in cultivating and identifying these organisms in clinical microbiology laboratories continue to leave the anaerobic etiology of an infectious process unproven in many cases. Therefore, an understanding of the types of infections in which anaerobes can play a role is crucial in selecting appropriate microbiologic tools to identify the organisms in clinical specimens and in choosing the most appropriate treatment, including antibiotics and surgical drainage or debridement of the infected site. This chapter focuses on infections caused by anaerobic bacteria other than *Clostridium* species, which are covered elsewhere **(Chaps. 139 and 159)**.

HISTORIC PERSPECTIVE

Anaerobic organisms were first identified by Antonie van Leeuwenhoek in 1680—nearly a century before oxygen itself was discovered. Leeuwenhoek set up culture medium (crushed pepper powder and clean rainwater) in two glass tubes—one open to ambient air and the other sealed closed—that he incubated for several days. Although he did not expect to observe anything in the sealed tube, he was surprised to find "animalcules" in both tubes. He noted that these bacteria in the sealed tube were "bigger than the biggest sort" in the tube left open to air. It was not until the mid- to late nineteenth century that Leeuwenhoek's findings were confirmed by Pasteur and others. However, these principles described by Leeuwenhoek underlie the basic pathogenesis of anaerobic infections: development of an anaerobic environment in a closed space is due to consumption of oxygen by aerobic organisms and results in the outgrowth of anaerobic organisms.

DIFFERENCES BETWEEN ANAEROBIC AND AEROBIC ORGANISMS

Anaerobic bacteria can be categorized as *obligate anaerobes* (killed in the presence of ≥0.5% oxygen), *aerotolerant organisms* (can tolerate the presence of oxygen but cannot use it for growth), and *facultative anaerobes* (can grow in the presence or absence of oxygen). Most clinically relevant anaerobes, such as *Bacteroides fragilis*, *Prevotella melaninogenica*, and *Fusobacterium nucleatum*, are relatively aerotolerant. These organisms contrast with *obligate aerobes*, which require high concentrations of oxygen for growth, and *microaerophilic organisms*, which are damaged by atmospheric concentrations of oxygen (~21%) but require low concentrations of oxygen (typically 2–10%) for growth. Given that molecular oxygen can reduce to superoxide (O_2^-) and hydrogen peroxide (H_2O_2), which are damaging to cells, the ability to tolerate the presence of oxygen is due, in part, to the expression of superoxide dismutase and catalase. The variation in anaerobic organisms tolerating anywhere from <0.5 to 8% O_2 may reflect the amount of these enzymes that is produced.

Furthermore, aerobic and anaerobic organisms differ in their energy metabolism. Cellular respiration requires establishment of an electrochemical gradient across the membrane, resulting in an electric potential (often related to a proton gradient) across the membrane. In aerobic respiration, electrons are shuttled through an electron transport chain, with oxygen as the final electron acceptor. Anaerobic organisms can metabolize energy by either anaerobic respiration or fermentation. Given that the final electron acceptor in anaerobic respiration (e.g., sulfate, nitrate, carbon dioxide, or fumarate) is not as highly oxidizing as oxygen, this pathway is less efficient than aerobic respiration and produces less ATP per glucose molecule. In contrast, fermentation does not use an electrochemical gradient. Rather, it releases energy from an organic molecule (e.g., pyruvate and its derivatives) via substrate-level phosphorylation and is therefore a less efficient process than either aerobic or anaerobic respiration; for comparison, fermentation results in ~5% of the energy released by aerobic respiration. For these reasons, facultative anaerobes will preferentially utilize oxygen if it is available; in oxygen-limiting situations, organisms will use anaerobic respiration rather than fermentative processes, if possible.

ANAEROBES OF THE HUMAN MICROBIOTA

Most human mucocutaneous surfaces harbor a rich indigenous normal microbiota composed of aerobic and anaerobic bacteria. These surfaces are dominated by anaerobic bacteria, which often account for 99.0–99.9% of the cultivable microbiota and range in concentration from 10^3/mL in the nose to 10^{12}/mL in gingival scrapings and the colon **(Table 182-1)**. It is interesting that anaerobes inhabit many areas of the body that are exposed to air: skin, nose, mouth, and throat. Anaerobes are thought to reside in the portions of these sites that either are relatively well protected from oxygen (e.g., gingival crevices) or have a local anaerobic environment conferred by neighboring aerobic organisms (e.g., tooth surfaces). The ability to cultivate these organisms is improving, and—with strict attention to anaerobic conditions—more than 80% of the microscopic counts in fecal samples can be cultured. However, culture-independent approaches (e.g., sequencing of the 16S rDNA gene) show that the overwhelmingly diverse low-abundance bacterial species present in the microbiota remain uncultivated. Several projects, including the Human Microbiome Project (funded by the U.S. National Institutes of Health) and MetaHIT (financed by the European Commission), have characterized the normal microbiota of healthy individuals and have demonstrated the presence of >10,000 different bacterial species in the collective human microbiota. The human gut

TABLE 182-1 The Anaerobic Human Microbiota: An Overview

ANATOMIC SITE	TOTAL BACTERIA[a]	ANAEROBIC/ AEROBIC RATIO	POTENTIAL PATHOGEN(S)
Nose	10^3–10^4	2:1	*Peptostreptococcus* spp., *Prevotella* spp.
Oral cavity			
Saliva	10^8–10^9	10:1	*Fusobacterium nucleatum*, *Prevotella melaninogenica*, *Prevotella oralis* group, *Bacteroides ureolyticus* group, *Peptostreptococcus* spp.
Tooth surface	10^{10}–10^{11}	1:1	
Gingival crevices	10^{11}–10^{12}	10^3:1	
Gastrointestinal tract			
Stomach	10^0–10^3	1:10	*Lactobacillus* spp.
Duodenum	10^1–10^5	1:1	*Lactobacillus* spp., *Streptococcus* spp.
Jejunum	10^3–10^6	1:1	*Streptococcus* spp., *Lactobacillus* spp., *Peptostreptococcus* spp.
Ileum	10^4–10^9	10:1	*Bacteroides* spp., *Streptococcus* spp., *Enterococcus* spp.
Cecum and colon	10^{11}–10^{12}	10^3:1	*Bacteroides* spp. (principally members of the *B. fragilis* group), *Prevotella* spp., *Clostridium* spp.
Female genital tract	10^7–10^9	10:1	*Peptostreptococcus* spp., *Bacteroides* spp., *Prevotella bivia*
Skin	10^4–10^6	100:1	*Cutibacterium acnes*

[a]Per gram or milliliter.

alone harbors >1000 bacterial species, with 100–200 species present in any given individual.

The major reservoir of anaerobic bacteria is the lower gastrointestinal tract, but these organisms are also present in considerable numbers in the oral cavity, skin, and female genital tract (Table 182-1). In the oral cavity, the ratio of anaerobic to aerobic bacteria ranges from 1:1 on the surface of a tooth to 1000:1 in the gingival crevices. *Prevotella* and *Porphyromonas* species make up much of the indigenous oral anaerobic microbiota. *Fusobacterium* and *Bacteroides* (non–*B. fragilis* group) species are present in lower numbers. Anaerobic bacteria are not found in appreciable numbers in the normal stomach and proximal small intestine. In the distal ileum, the microbiota begins to resemble that of the colon, where the ratio of anaerobes to aerobic species is high (~1000:1). The predominant anaerobes in the human intestine belong to the phyla Bacteroidetes and Firmicutes and include a number of *Prevotella* and *Bacteroides* species (e.g., members of the *B. fragilis* group, such as *B. fragilis*, *B. thetaiotaomicron*, *B. ovatus*, *B. vulgatus*, *B. uniformis*, and *Parabacteroides distasonis*) as well as various *Clostridium*, *Peptostreptococcus*, *Blautia*, and *Fusobacterium* species. In the female genital tract, there are ~10^9 organisms/mL of secretions, with an anaerobe-to-aerobe ratio of ~10:1. The predominant anaerobes in the female genital tract are *Prevotella*, *Bacteroides*, *Fusobacterium*, *Clostridium*, and the anaerobic *Lactobacillus* species. The skin microbiota contains anaerobes as well; *Cutibacterium acnes* (which was previously *Propionibacterium acnes* and will be considered as one of the *Propionibacterium* species for the remainder of this chapter) is the predominant species, and other species of propionibacteria and peptostreptococci are present in lower numbers.

ANAEROBES AND HUMAN HEALTH

Commensal anaerobes have been implicated as crucial mediators of physiologic, metabolic, and immunologic functions in the mammalian host. The intestinal microbiota is essential for fermenting dietary carbohydrates into forms that are more usable by the host, among which polysaccharides are the most abundant biologic source of energy. Of the organisms found within the intestines, *Bacteroides* species express the widest array of polysaccharide-degrading enzymes, providing important nutrients for both the host and other commensal organisms. For example, *B. thetaiotaomicron* expresses 172 glycosyl hydrolases. The anaerobic intestinal microbiota is also responsible for the production of secreted products that promote human health (e.g., vitamin K and bile acids useful in fat absorption and cholesterol regulation).

One of the most important roles that anaerobes serve as components of the normal colonic microbiota is the promotion of resistance to colonization. The presence of commensal bacteria effectively interferes with colonization by potentially pathogenic bacterial species through the depletion of oxygen and nutrients, the production of enzymes and toxic end products, and the modulation of the host's intestinal innate immune response. For example, the normal intestinal microbiota plays an important role in protection against enteric infections, including those due to *Salmonella enterica* serotype Typhimurium and *Clostridium difficile*.

The anaerobic intestinal microbiota also has immunomodulatory properties that help regulate the immune system. The first example of this role was demonstrated with *B. fragilis*, which can balance the effector functions of T cells in the peripheral immune system and induce colonic regulatory T cells via expression of polysaccharide A (PSA). Moreover, *B. fragilis* expresses a glycosphingolipid that regulates the number of colonic invariant natural killer T cells. There are now numerous examples of commensal anaerobes that can modulate different aspects of the intestinal and extraintestinal immune system—everything from specific effector T cells to dendritic cells to antimicrobial peptides.

Clearly, the gut microbiota confers many benefits, and its dysregulation may play a role in the pathogenesis of diseases characterized by inflammation and aberrant immune responses, such as inflammatory bowel disease, rheumatoid arthritis, multiple sclerosis, asthma, and type 1 diabetes. Furthermore, the gut microbiota has been associated with obesity and metabolic syndrome. A more complete discussion of the intersection between the microbiota and human health is covered elsewhere **(Chap. 484)**.

ETIOLOGY

There are >10,000 species of bacteria—the overwhelming majority of which are anaerobes—in the human microbiota, with each individual colonized by hundreds of species. Anaerobic infections occur when the harmonious relationship between the host and the host's microbiota is disrupted. Any site in the body is susceptible to infection with these indigenous organisms if they are introduced into otherwise sterile tissue, either through disruption of mucosal surfaces (e.g., intestinal perforation, ischemia, surgery) or via direct inoculation of organisms into tissue (e.g., bite wounds, trauma). Because the sites that are colonized by anaerobes contain many species of bacteria, the resulting infections are often polymicrobial, involving multiple species of anaerobes in combination with synergistically acting facultative and/or microaerophilic organisms.

Despite the complex array of bacteria in the normal microbiota, relatively few genera are isolated commonly from human infections **(Fig. 182-1)**. While the specific organisms identified vary with the site and source of infection, the etiologic agents typically reflect the neighboring microbiota. For example, organisms normally found in the oro- and nasopharyngeal microbiota (e.g., *P. melaninogenica*, *Fusobacterium necrophorum*, *F. nucleatum*, *Peptostreptococcus* species, *Porphyromonas gingivalis*, *Porphyromonas asaccharolytica*, and *Actinomyces* species) can cause disease in contiguous areas, including odontogenic infections, peripharyngeal space infections, chronic sinusitis, and pleuropulmonary infections. In female genital tract infections, organisms normally colonizing the vagina (e.g., *Prevotella bivia* and *Prevotella disiens*) are the most common isolates. *Escherichia coli* and *B. fragilis*, both of which are components of the intestinal microbiota, are the most commonly identified isolates from intraabdominal abscesses. Indeed, the *B. fragilis* group, which encompasses 25 species and includes *B. thetaiotaomicron*, *B. vulgatus*, *B. uniformis*, and *B. ovatus*, contains the anaerobic organisms among the most frequently isolated from clinical infections.

It is useful to think about anaerobic infectious etiologies with regard not only to their anatomic location but also to their microbiologic features. While many anaerobic gram-negative bacilli cause disease (e.g., *Prevotella*, *Bacteroides*, *Fusobacterium*, and *Porphyromonas* species), *Veillonella* species, which are part of the oral and intestinal microbiota, are among the few anaerobic gram-negative cocci that have been implicated in human disease. Similarly, the peptostreptococci (e.g., *P. micros*, *P. asaccharolyticus*, and *P. anaerobius*) and *Finegoldia magnus* (which was previously *Peptostreptococcus magnus* and will be considered as part of the peptostreptococci for the remainder of this chapter) are the chief anaerobic gram-positive cocci that have pathogenic potential. *Clostridium* species are the primary anaerobic spore-forming gram-positive rods that produce human disease **(Chap. 159)**. Uncommonly, anaerobic gram-positive non-spore-forming bacilli cause infection;

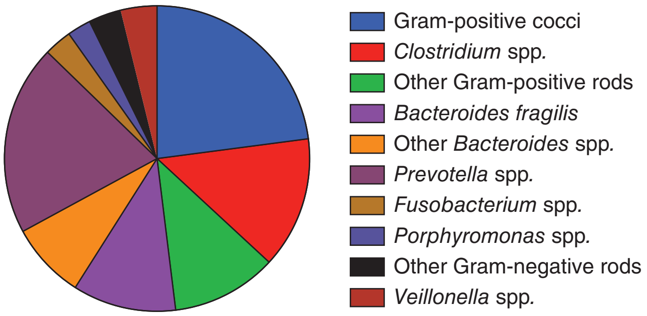

FIGURE 182-1 Distribution of anaerobic organisms isolated from clinical materials. *(Data combined from Y Park et al: Clinical features and prognostic factors of anaerobic infections: A 7-year retrospective study. Korean J Intern Med 24:13, 2009; and Japanese Association for Anaerobic Infections Research: Anaerobic infections (general): Epidemiology of anaerobic infections. J Infect Chemother 17:4, 2011.)*

C. acnes, a component of the skin microbiota and a cause of foreign-body infections, and *Actinomyces* species are relevant examples.

PATHOGENESIS

First and foremost, anaerobic infections require an anaerobic environment with a lowered oxidation-reduction potential. In some circumstances, this environment can occur directly—e.g., in tissue ischemia, trauma, surgery, or a perforated viscus. In many other situations, the infection is polymicrobial, and the facultative organisms maintain a lowered oxidation-reduction potential in the local microenvironment that allows for the propagation of obligate anaerobes. Once the proper anaerobic environment is established, the organisms must still contend with the host's immune defenses. Similar to aerobic organisms, anaerobes express an array of virulence factors that help evade host defenses, they can form abscesses as a protective measure, and they can act synergistically with other bacteria to better persist in the host.

Virulence factors associated with anaerobes typically confer the ability to evade host defenses, adhere to cell surfaces, produce toxins and/or enzymes, or display surface structures such as capsular polysaccharides and lipopolysaccharides that contribute to pathogenic potential. The ability of an organism to adhere to host tissues is often critical to the establishment of infection. Some oral species adhere to the epithelium in the oral cavity. *P. gingivalis*, a common isolate in periodontal disease, has fimbriae that facilitate attachment. In supragingival plaque, many oral anaerobes are able to attach directly to aerobic bacteria (e.g., *Streptococcus* species) that are adherent to the tooth's surface. *F. nucleatum* is a notable example of these secondary colonizers: it expresses receptors to which almost all oral bacteria can bind and serves as an important bridge between the primary colonizers and subsequent layers of bacteria. *B. fragilis* synthesizes pili, fimbriae, and hemagglutinins that aid in attachment to host cell surfaces in the intestine.

Anaerobic bacteria produce several exoproteins that can enhance the organisms' virulence. *P. gingivalis* produces a collagenase that enhances tissue destruction. Exotoxins produced by clostridial species, including botulinum toxins, tetanus toxin, *C. difficile* toxins A and B, and five toxins produced by *Clostridium perfringens*, are among the most virulent bacterial toxins in mouse lethality assays. Anaerobic gram-negative bacteria, such as *B. fragilis*, *P. gingivalis*, and *Prevotella intermedia*, possess lipid A molecules (endotoxins) that are 100–1000 times less biologically potent than endotoxins associated with aerobic gram-negative bacteria; these differences relate to variations in acylation status, length of fatty acids, and number of phosphate groups. This relative biologic inactivity may account for the lower frequency of disseminated intravascular coagulation and purpura in anaerobic gram-negative bacteremia than in facultative and aerobic gram-negative bacillary bacteremia. An exception is the lipopolysaccharide from *Fusobacterium*, which may account for the severity of Lemierre's syndrome (see "Complications of Anaerobic Head and Neck Infections," below).

The most extensively studied virulence factor of the nonsporulating anaerobes is the capsular polysaccharide complex of *B. fragilis*. This organism is unique among anaerobes in its potential for virulence during growth at normally sterile sites. Although it constitutes only 0.5–1% of the normal colonic microbiota, *B. fragilis* is the anaerobe most commonly isolated from intraabdominal infections and bacteremia. In an animal model of intraabdominal sepsis, the capsular polysaccharide was identified as the major virulence factor of *B. fragilis*; this polymer plays a specific, central role in the induction of abscesses. A series of detailed biologic and molecular studies of this virulence factor showed that *B. fragilis* produces at least eight distinct capsular polysaccharides, far more than the number reported for any other encapsulated bacterium. *B. fragilis* can exhibit distinct surface polysaccharides either alone or in combination by regulating the expression of these different capsules in an on–off manner through a reversible inversion of DNA segments within the promoters for operons containing the genes required for polysaccharide synthesis. Structural analysis of two of these polysaccharides, PSA and polysaccharide B (PSB), revealed that each polymer consists of repeating units with positively charged free amino groups and negatively charged groups. This structural feature is rare among bacterial polysaccharides, and the ability of PSA—and, to a lesser extent, PSB—to induce abscesses in animals depends on this zwitterionic charge motif. Intraabdominal abscess induction is related to the capacity of PSA to stimulate macrophages to release cytokines and chemokines—in particular, interleukin (IL) 8, IL-17, and tumor necrosis factor α (TNF-α)—from resident peritoneal cells through a Toll-like receptor 2–dependent mechanism. The release of cytokines and chemokines results in the chemotaxis of polymorphonuclear neutrophils (PMNs) into the peritoneum, where they adhere to mesothelial cells induced by TNF-α to upregulate their expression of intercellular adhesion molecule 1 (ICAM-1). PMNs adherent to ICAM-1-expressing cells probably represent the nidus for an abscess. PSA also activates T cells to produce certain cytokines, including IL-17 and interferon γ, that are necessary for abscess formation.

These virulence factors not only promote persistence of the anaerobe that produces them but also aid in the survival of bystander organisms and result in bacterial synergies. Clinically, these synergies are evidenced by the fact that anaerobic infections typically involve three to six different organisms. Examples of this synergistic pathogenesis include creation of a favorable environment for growth (e.g., establishment and maintenance of an anaerobic environment by facultative organisms), inhibition of host defenses (e.g., production of short-chain fatty acids and succinic acid that inhibit the ability of phagocytes to clear facultative organisms), provision of necessary growth factors for other organisms (e.g., oral diphtheroids that produce vitamin K, which is needed by *P. melaninogenica*), and creation of tissue damage that promotes spread of the infection. In these ways, facultative and obligate anaerobes synergistically potentiate abscess formation.

APPROACH TO THE PATIENT

Infections Due to Anaerobic Bacteria

The physician must consider several points when approaching the patient with a possible infection due to anaerobic bacteria.

1. The organisms colonizing mucosal sites are commensals, very few of which typically cause disease. When these organisms do cause disease, it often occurs in proximity to the mucosal site they colonize.
2. For anaerobes to cause tissue infection, they must spread beyond the normal mucosal barriers.
3. Conditions favoring the propagation of anaerobic bacteria, particularly a lowered oxidation-reduction potential, are necessary. These conditions exist at sites of trauma, tissue destruction, compromised vascular supply, and necrosis.
4. Frequently, a complex array of infecting microbes can be found, occasionally with >10 different species isolated from a suppurative site.
5. Anaerobic organisms tend to be found in abscess cavities or in necrotic tissue. The failure of an abscess to yield organisms on routine culture is a clue that the abscess is likely to contain anaerobic bacteria. Often smears of this "sterile pus" are found to be teeming with bacteria when Gram's stain is applied. Although some facultative organisms (e.g., *Staphylococcus aureus*) are also capable of causing abscesses, abscesses in organs or deeper body tissues should call anaerobic infection to mind.
6. Gas is found in many anaerobic infections of deep tissues but is not diagnostic because it can be produced by aerobic bacteria as well.
7. Although a putrid-smelling infection site or discharge is considered diagnostic for anaerobic infection, this manifestation usually develops late in the course and is present in only 30–50% of cases.
8. Some species (the best example being the *B. fragilis* group) require specific therapy. However, many synergistic infections can be cured with antibiotics directed at some but not all of the organisms involved. Antibiotic therapy, combined with debridement and drainage, disrupts the interdependent relationship

among the bacteria, and some species that are resistant to the antibiotic do not survive without the co-infecting organisms.

9. Manifestations of severe sepsis and disseminated intravascular coagulation are unusual in patients with purely anaerobic infection.

EPIDEMIOLOGY

Difficulties in the performance of appropriate cultures, contamination of cultures by components of the normal microbiota, and the lack of readily available, reliable culture techniques have made it challenging to obtain accurate data on the incidence or prevalence of anaerobic infections. However, anaerobic infections are encountered frequently, with anaerobes constituting 7–8% and 13–15% of bacteria isolated from inpatients and outpatients, respectively. Bacteremia and soft tissue infections are the most common types of anaerobic infection **(Fig. 182-2)**. Typically, anaerobic bacteria account for <1% of all cases of bacteremia.

CLINICAL MANIFESTATIONS

Although anaerobes can cause infection anywhere in the body, certain clinical findings and characteristics are commonly found. These include abscess formation, putrid purulence (due to volatile fatty acid by-products), septic thrombophlebitis, tissue necrosis, and failure to respond clinically to broad-spectrum antibiotics that lack activity against anaerobes.

Anaerobic Infections of the Mouth, Head, and Neck Anaerobic bacteria are commonly involved in infections of the mouth, head, and neck **(Chap. 37)**. The predominant isolates are components of the normal microbiota of the upper airways—mainly *Prevotella* species, *P. asaccharolytica*, *Fusobacterium* species, peptostreptococci, and microaerophilic streptococci.

OROFACIAL INFECTIONS The most common oral infections are odontogenic and include dental caries and periodontal disease (gingivitis and periodontitis). While dental caries usually manifest with pain, sensitivity, and discoloration of the tooth, periodontal disease involves inflammation of the gums and underlying tissue. In its more severe forms, periodontitis can result in difficulty chewing, loose teeth, and occasionally tooth loss. Severe orofacial infections typically develop as a consequence of dental infection, and the infection can spread from the tooth to different anatomic areas that provide the least resistance, resulting in periapical, periodontal, or pericoronal infections. If the dental surface is completely breached, an endodontic infection (pulpitis) can occur. In late stages of pulpitis, the tooth is generally very sensitive to heat, but cold stimuli may provide relief. Left untreated, pulpitis can progress to invade the alveolar bone and develop into a periapical abscess. The abscesses, particularly those involving the second and third molars, can occasionally extend into the submandibular, sublingual, and submental spaces (*Ludwig's angina*). This infection results in marked local swelling of tissues, with pain, trismus, and superior and posterior displacement of the tongue. Submandibular swelling of the neck and obstruction by the tongue can impair swallowing and cause respiratory obstruction. In some cases, tracheotomy is lifesaving.

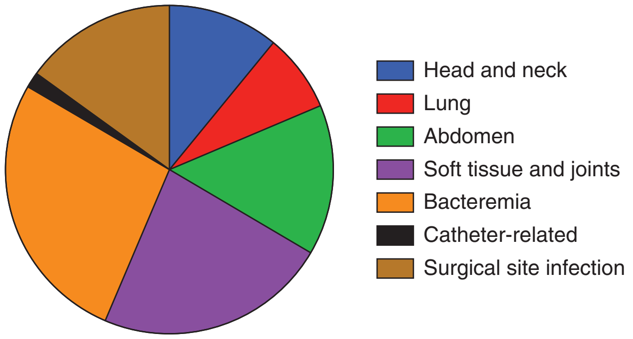

FIGURE 182-2 Distribution of types of infection from which anaerobic organisms were cultured at a single hospital over a 7-year period. Head and neck infections included sinusitis, otitis media, and retropharyngeal abscess; abdominal infections included liver abscess, biliary tract infection, bowel obstruction, and intraabdominal abscess; catheter-related infections included those related to peritoneal dialysis catheters and ventriculoperitoneal shunts. *(Data from Y Park et al: Clinical features and prognostic factors of anaerobic infections: A 7-year retrospective study. Korean J Intern Med 24:13, 2009.)*

Microbiologically, dental caries begin with the binding of *Streptococcus mutans* and *Streptococcus sanguis* to the tooth surface, with subsequent further colonization by anaerobes. In contrast, periodontitis is typically associated with *P. gingivalis*, *Tannerella forsythia*, *Aggregatibacter actinomycetemcomitans*, and *Treponema denticola*. *Fusobacterium*, *Actinomyces*, *Peptostreptococcus*, and *Bacteroides* species (other than *B. fragilis*) are the organisms most commonly isolated from periapical abscesses.

ACUTE NECROTIZING ULCERATIVE GINGIVITIS Gingivitis may become a necrotizing infection (*trench mouth*, *Vincent's stomatitis*). The onset of disease is usually sudden and is associated with painful bleeding gums, foul breath, and a bad taste. The gingival mucosa, especially the papillae between the teeth, becomes ulcerated and may be covered by a yellowish-white or gray "pseudomembrane," which is removable with gentle pressure. Patients may become systemically ill, developing fever, malaise, cervical lymphadenopathy, and leukocytosis. The infection can sometimes extend into the pharynx, resulting in an extremely sore throat, foul breath, and tonsillar pillars that are swollen, red, ulcerated, and covered by a pseudomembrane. *Prevotella*, *Treponema*, and *Fusobacterium* species have been implicated.

In some cases, acute necrotizing gingivitis can rapidly progress to noma (*cancrum oris*), a gangrenous infection that destroys the soft and hard tissues related to the oral cavity. Noma occurs most frequently in young children (1–4 years of age) who have immune dysfunction related to malnutrition and endemic infections (particularly measles). This infection occurs worldwide but is most common in sub-Saharan Africa, where the incidence is 1–7 cases per 1000 children. Although the pathogenesis is not fully understood, infections with *F. necrophorum* and *P. intermedia* are thought to be key drivers of this disease. Without treatment, the mortality rate is 70–90%.

PERIPHARYNGEAL SPACE INFECTIONS These infections arise from the spread of organisms from the upper airways to potential spaces formed by the fascial planes of the head and neck. The etiology is typically polymicrobial and represents the normal microbiota of the mucosa of the originating site.

Peritonsillar abscess (*quinsy*) is the most common peripharyngeal infection and occurs as a complication of acute tonsillitis. Consistent with its association with tonsillitis, adolescents are most commonly affected. Patients present with a sore throat, dysphagia, peritonsillar swelling, muffled voice, and uvular deviation to the contralateral side. The abscess material typically grows group A *Streptococcus* in conjunction with obligate anaerobes (e.g., *Bacteroides*, *Prevotella*, and *Peptostreptococcus* species) **(Chap. 37)**. Retropharyngeal abscesses typically occur in children 2–4 years of age, although they can occur at any age. Although a suppurative infection of the retropharyngeal lymph nodes is the usual precursor to these abscesses in children, foreign-body ingestion and/or local trauma is more commonly the inciting factor in adults. The clinical presentation shares many features with peritonsillar abscesses, but difficulty extending the neck and torticollis are more common with retropharyngeal abscesses. The etiologic agents are the same as in peritonsillar abscesses, with additional aerobic organisms (e.g., *S. aureus*, viridans streptococci) also playing a role.

SINUSITIS AND OTITIS Anaerobic bacteria have been implicated in chronic sinusitis but play little role in acute sinusitis. Numerous studies related to the microbiology of chronic sinusitis have been conducted; on average, anaerobic bacteria have been found in two-thirds of patients, with many studies demonstrating their presence in >90% of patients. Anaerobic bacteria represent ~40% of all bacteria cultured, with *Peptostreptococcus*, *Prevotella*, and *Porphyromonas* species the most commonly isolated anaerobes. *S. aureus* and Enterobacteriaceae are the aerobes most commonly recovered in chronic sinusitis. Anaerobic bacteria have been isolated in ~60% of cases of chronic suppurative otitis media in children, but they are not involved in acute otitis media.

COMPLICATIONS OF ANAEROBIC HEAD AND NECK INFECTIONS Direct extension of these infections into contiguous areas can result in additional disease manifestations. Cranial spread of these infections can result in osteomyelitis of the skull or mandible or in intracranial infections, such as brain abscess and subdural empyema. Caudal spread can produce mediastinitis or pleuropulmonary infection. Hematogenous complications can also result from anaerobic infections of the head and neck. Bacteremia, which occasionally is polymicrobial, can lead to endocarditis or other distant infections. Lemierre's syndrome **(Chap. 37)**, which is usually due to *F. necrophorum*, is an acute oropharyngeal infection with secondary septic thrombophlebitis of the internal jugular vein and frequent septic emboli, most commonly to the lung. This infection typically begins with pharyngitis, which is followed by local invasion in the lateral pharyngeal space, with resultant internal jugular vein thrombophlebitis.

Central Nervous System (CNS) Infections CNS infections associated with anaerobic bacteria are brain abscess, epidural abscess, and subdural empyema, in which anaerobes are recovered in up to 30, 20, and 10% of cases, respectively. The frequency with which anaerobes are recovered depends in large part on the underlying reason for the infection. For example, brain abscesses are typically due to hematogenous seeding, contiguous spread, penetrating head trauma, or recent surgical intervention. Anaerobic bacteria are most commonly associated with brain abscesses resulting from contiguous spread (related to otogenic, odontogenic, and sinus infections), and the pathogens recovered are the same as in these antecedent infections. Facultative or microaerophilic streptococci and coliforms are often part of a mixed infecting flora in brain abscesses. The location of the abscess may also provide insight into the pathogens. Abscesses in the frontal lobe (often associated with sinusitis) are due to anaerobes, streptococci, and staphylococci; temporal lobe and cerebellar abscesses are often related to the oral microbiota and middle-ear pathogens.

Obligate anaerobes rarely cause meningitis. Only one obligate anaerobe was identified in a seminal study of 188 bacterial meningitis isolates, and a U.S. national surveillance study of 18,642 such isolates collected between 1977 and 1981 found only five obligate anaerobes. This low incidence may be due, in part, to the fact that many clinical microbiology laboratories do not routinely culture cerebrospinal fluid (CSF) for anaerobes.

Pleuropulmonary Infections The lungs are constantly seeded with organisms from the oral microbiota via subclinical microaspiration that normally occurs in all people. Even though the lung is the site of oxygen exchange and is therefore an overwhelmingly aerobic environment, the organisms most abundant in the lower respiratory tract (as assessed by culture-independent methods) include anaerobes such as *Prevotella* and *Veillonella* species, with oral microaerophilic streptococcal species (e.g., the *Streptococcus milleri* group) also present in significant abundances. In patients who have impaired bacterial clearance (due to decreased cough, dysfunctional mucociliary transport, or alcohol intoxication) and/or increased rates of aspiration (due to neurologic disorders, impaired consciousness, or swallowing dysfunction), these anaerobic bacteria can establish an infection and result in aspiration pneumonia, lung abscess, or empyema. These anaerobic infections have an indolent course that may serve as a clinical clue differentiating them from conditions with other etiologies (e.g., chemical pneumonitis, pneumococcal pneumonia) that often present more acutely.

ASPIRATION PNEUMONIA Bacterial aspiration pneumonia must be distinguished from two other clinical syndromes associated with aspiration that are not of bacterial etiology. One syndrome results from aspiration of food or, rarely, other foreign bodies. Obstruction of major airways typically results in difficulty breathing, atelectasis, and moderate nonspecific inflammation. Therapy consists of removal of the foreign body. The second aspiration syndrome relates to chemical pneumonitis caused by inhalation or aspiration of alveolar irritants. Perhaps the most common cause of chemical pneumonitis is *Mendelson syndrome*, which results from regurgitation and aspiration of acidic gastric juices. Pulmonary inflammation—including the destruction of the alveolar lining, with transudation of fluid into the alveolar space—occurs with remarkable rapidity. This syndrome typically develops within 4–6 h, often following anesthesia when the gag reflex is depressed. The patient becomes tachypneic, tachycardic, and hypoxic, often in the absence of fever. The leukocyte count may rise, and the chest x-ray may evolve from normal to a complete bilateral "whiteout" within 8–24 h. Sputum production is minimal. The pulmonary signs and symptoms often resolve quickly with symptom-based therapy, but this condition can culminate in respiratory failure due, in part, to pulmonary edema. Antibiotic therapy is not indicated unless bacterial superinfection occurs.

In contrast to these syndromes, bacterial aspiration pneumonia develops over a period of several days or weeks rather than hours. The pathogenesis includes some combination of an increased bacterial burden, increased virulence of the organisms aspirated, and potential airway damage related to aspiration of gastric fluid. Patients generally report fever, malaise, and sputum production. In some patients, weight loss and anemia reflect a more chronic process. Usually the history reveals factors predisposing to aspiration, such as significant alcohol consumption or neurologic impairment due to a previous stroke. Severe dental disease is often associated with aspiration pneumonia, but it is not clear whether this association relates to an increased number of oral microbes and/or the presence of organisms with increased virulence. Sputum characteristically is not malodorous unless the process has been ongoing for at least a week. Chest x-rays show consolidation in dependent pulmonary segments: in the basilar segments of the lower lobes if the patient has aspirated while upright and in either the posterior segment of the upper lobe (usually on the right side, given that the right mainstem bronchus has a more vertical orientation) or the superior segment of the lower lobe if the patient has aspirated while supine.

A mixed bacterial population with many PMNs is evident on Gram's staining of sputum. Expectorated sputum is unreliable for anaerobic cultures because of inevitable contamination by the normal oral microbiota. Reliable specimens for culture can be obtained by transtracheal or transthoracic aspiration—techniques that are rarely used at present. Although the culture of protected-brush specimens or bronchoalveolar lavage fluid obtained by bronchoscopy is controversial, more recent data suggest that these approaches can be used without oropharyngeal contamination and can recover anaerobic organisms from the lower respiratory tract in a site-directed manner. Further research is needed to determine how these approaches compare with the previous gold standards.

ANAEROBIC LUNG ABSCESSES **(See also Chap. 132)** These abscesses result from subacute anaerobic pulmonary infection. The clinical presentation typically involves a history of constitutional signs and symptoms (including malaise, weight loss, fever, night sweats, and foul-smelling sputum) that have typically persisted for 1–3 weeks prior to hospitalization. Patients who develop lung abscesses often, but not always, have an antecedent dental infection. Abscess cavities may be single or multiple and generally occur in dependent pulmonary segments **(Fig. 182-3)**. The differential diagnosis for lung abscesses includes pneumonia (including necrotizing pneumonia), a purulent pleural effusion with a bronchopleural fistula, and a pneumatocele. Of note, infection with some aerobic organisms, particularly *S. aureus*, can develop into a lung abscess without an anaerobic component. Approximately 90% of cases have an anaerobe identified—usually three to six isolates per sample—if careful attention is paid to handling and processing of the abscess sample. The most common isolates include peptostreptococci, *Prevotella* and *Porphyromonas* species, and *F. nucleatum*. An important finding is that ~90% of cultures also demonstrate the presence of aerobic organisms, such as *S. aureus*, *Streptococcus pneumoniae*, and *Klebsiella pneumoniae*. Consistent with the notion that anaerobes are contributing to disease, patients often do not improve clinically until they receive an antibiotic regimen that includes anaerobic coverage.

EMPYEMA Empyema is a manifestation of long-standing anaerobic pulmonary infection and manifests with thick, purulent material in

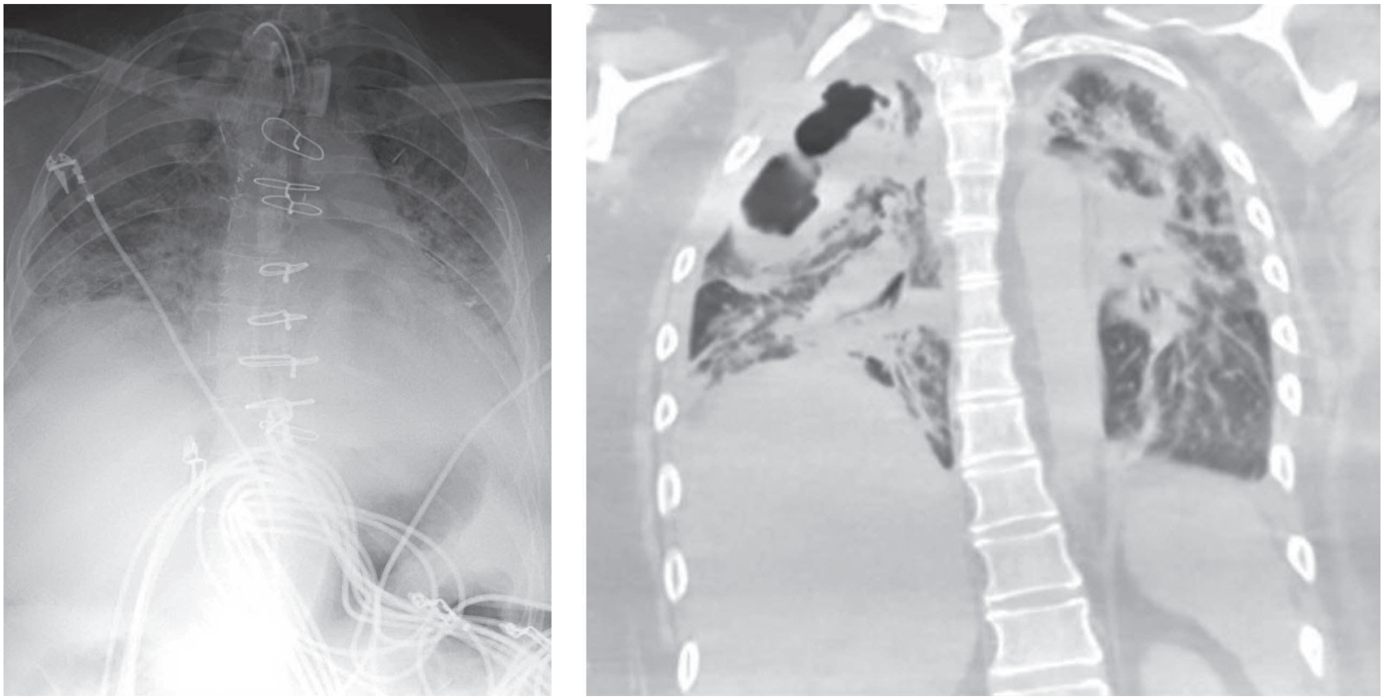

FIGURE 182-3 Chest radiograph (*left*) and CT image (*right*) of a lung abscess. The patient aspirated while supine and developed an abscess in the posterior segment of the right upper lobe. Cultures were pretreated and grew only *Klebsiella pneumoniae*. *(Images provided by Gita N. Mody, MD, MPH, Division of Cardiothoracic Surgery, Department of Surgery, The University of North Carolina at Chapel Hill.)*

the pleural space, often in association with a bronchopleural fistula. Alternatively, a subdiaphragmatic infection may extend into the pleural space and similarly result in an empyema. The clinical presentation resembles that of other anaerobic pulmonary infections and may include foul-smelling sputum, pleuritic chest pain, and marked chest-wall tenderness. This disease process must be differentiated from a parapneumonic effusion resulting from more routine causes of pneumonia (e.g., *S. pneumoniae*). In the latter instance, the fluid is a thin exudate that has a mean white blood cell (WBC) count of ~5000 cells/mL, a lactate dehydrogenase level of >200 IU/L, and a pH of ~7.4. In contrast, empyema is characterized by foul-smelling thick pus with a mean WBC count of ~55,000 cells/mL, a lactate dehydrogenase level of >1000 IU/L, and a pH of <7.2 as well as loculations and a thick pleural peel on imaging. Drainage and occasionally decortication of the visceral and parietal pleura are required. Defervescence, a return to a feeling of well-being, and resolution of the process may require several months, particularly in the absence of surgical intervention.

Intraabdominal Infections Breach of the gut mucosal surface (e.g., due to trauma, intestinal perforation, or malignancy) allows members of the microbiota to enter the normally sterile peritoneum. Accordingly, the offending organisms reflect the microbiota in the affected intestinal region. For example, recovery of *Candida* species from intraabdominal infections should prompt evaluation of the stomach and proximal small bowel for potential perforation. Furthermore, a study of patients with perforated and gangrenous appendicitis demonstrated that virtually all samples yielded *E. coli* and members of the *B. fragilis* group; peptostreptococci and *Bilophila wadsworthia*—additional components of the appendiceal and colonic microbiota—also were recovered from >50% of samples. Notably, some studies have identified an average of 10 different bacterial species, with an anaerobe-to-aerobe ratio of ~3:1. Given that >1000 bacterial species are present in the colonic microbiota, the dominance of such a limited repertoire of bacterial genera and species recovered in intraabdominal infections reflects a combination of two factors: the increased propensity of these organisms to result in intraabdominal abscesses and the difficulty faced by clinical microbiology laboratories in culturing the diverse organisms present in these samples. **See Chap. 137 for a complete discussion of intraabdominal infections.**

Neutropenic enterocolitis (typhlitis) involves marked thickening of the bowel wall (typically >4 mm) in the setting of neutropenia, abdominal pain, and fever. This condition most commonly affects the cecum and may extend to the neighboring terminal ileum and/or proximal colon, but any intestinal region may be involved. Typhlitis generally occurs after 1–2 weeks of chemotherapy-induced neutropenia associated with treatment of hematologic or, less commonly, solid tumor malignancies, but it can occur regardless of the cause of neutropenia. At least 5% of adults hospitalized for malignancy are thought to develop typhlitis, but this is likely an underestimate. Although the right lower quadrant is the most common location of abdominal pain and tenderness, these symptoms are absent in nearly half of cases; moreover, some patients, particularly those taking glucocorticoids, may not experience abdominal pain at all. Given the weakened integrity of the bowel wall and the associated neutropenia, patients often develop bacteremia due to one or more organisms related to the microbiota of the affected intestinal segment. Patients who develop bacteremia due to *Clostridium septicum* often have relatively severe disease, and identification of this organism is highly associated with the presence of malignancy—notably, colon cancer. Medical management including bowel rest, intestinal decompression, and broad-spectrum antibiotic administration is generally successful, although surgical intervention may be required in cases of persistent intestinal bleeding, necrotic bowel, or clinical deterioration suggestive of an ongoing intestinal process.

Pelvic Infections Anaerobes are frequently encountered in pelvic inflammatory disease, pelvic abscess, endometritis, tubo-ovarian abscess, septic abortion, and postoperative or postpartum infections. These infections are often of mixed etiology, involving both anaerobes and coliforms; pure anaerobic infections without coliform or other facultative bacterial species occur more often in pelvic than in intraabdominal sites. The major anaerobic pathogens in pelvic abscesses are *P. bivia*, *P. disiens*, and the *B. fragilis* group, but many other anaerobes also have been implicated. **See Chap. 141 for a complete discussion of pelvic inflammatory disease.**

Anaerobic bacteria have been thought to be contributing factors in bacterial vaginosis. This syndrome of unknown etiology is characterized by a profuse malodorous discharge and a change in bacterial ecology that results in replacement of the *Lactobacillus*-dominated normal microbiota with an overgrowth of anaerobic bacterial species.

Culture-based and culture-independent approaches have identified numerous organisms, including *Gardnerella vaginalis*, peptostreptococci, genital mycoplasmas, and species within the genera *Prevotella*, *Mobiluncus*, *Atopobium*, *Leptotrichia*, *Megasphaera*, and *Eggerthella*. This wide array of implicated bacteria may reflect differences in the overall disease spectrum of bacterial vaginosis and/or a shared physiologic response to these different organisms.

Skin and Soft Tissue Infections Similar to other anatomic sites, skin or soft tissue injured by trauma, ischemia, or surgery creates a suitable environment for anaerobic infections. The infecting bacteria either are introduced directly (e.g., wounds associated with intestinal surgery, decubitus ulcers, or human bites) or originate in contiguous areas (e.g., cutaneous abscesses, rectal abscesses, and axillary sweat gland infections [*hidradenitis suppurativa*]). Anaerobes also are often cultured from foot ulcers of diabetic patients. The most common locations for anaerobic cellulitis include the neck, trunk, groin (including the genitalia), and legs. The deep soft tissue infections associated with anaerobic bacteria are gas gangrene, synergistic cellulitis (both progressive and necrotizing), necrotizing fasciitis, and myositis **(Chaps. 134 and 159)**.

Gas gangrene (crepitus cellulitis) is most often due to *C. perfringens*, although other clostridial species have been implicated as well. This infection involves extensive gas formation in the tissue leading to crepitus and a thin, dark, occasionally malodorous discharge. True gas gangrene typically presents with fever and tenderness around the lesion and can rapidly spread; in contrast, there are somewhat more indolent forms of anaerobic cellulitis that may involve some gas formation but often present without fever or extensive local pain and can spread over the course of days rather than minutes.

Progressive bacterial synergistic gangrene (*Meleney gangrene*) is characterized by an area of exquisite pain, redness, and swelling followed by ulceration. As the ulcer enlarges, it is surrounded by a violaceous ring that fades into a pink edematous border. If it is not promptly treated, the ulcer continues to enlarge, and new, distant ulcers may emerge. Symptoms are limited to pain; fever is not typical. Peptostreptococci and microaerophilic streptococci are commonly found in the leading edge of the lesions, and *S. aureus* and *Proteus* species can be isolated from the ulcerated lesion. Treatment includes surgical removal of necrotic tissue and antimicrobial administration. In contrast, synergistic necrotizing cellulitis involves the deep fascia and occurs near the point of bacterial entry. Pain, fever, and systemic symptoms are common. If this form of cellulitis involves the scrotum, perineum, and anterior abdominal wall, it is referred to as *Fournier gangrene*. *S. aureus*, the *B. fragilis* group, *Peptostreptococcus* species, *Clostridium* species, *Fusobacterium* species, and members of the family Enterobacteriaceae are the predominant organisms identified.

Necrotizing fasciitis, a rapidly spreading destructive disease of the fascia, is usually attributed to group A streptococci **(Chap. 153)** but can also be a mixed infection involving anaerobes and aerobes. Polymicrobial necrotizing fasciitis differs from stereotypical group A streptococcal necrotizing fasciitis in that the initial erythematous, swollen, tender lesions progress over 3–5 days (as opposed to 1–3 days), with consequent skin breakdown and cutaneous gangrene. Fever, subcutaneous gas, development of anesthesia (often before skin necrosis), and a foul-smelling discharge are common. The particular clinical findings sometimes suggest the causative agent: regional lymphadenopathy suggests the *B. fragilis* group; necrosis and gangrene suggest *Clostridium* species, peptostreptococci, the *B. fragilis* group, and Enterobacteriaceae; bullous lesions suggest Enterobacteriaceae; a foul-smelling odor suggests *Bacteroides* and *Clostridium* species; and subcutaneous gas suggests peptostreptococci, *Clostridium* species, and the *B. fragilis* group. Moreover, diabetic infections are often associated with *Bacteroides* species, Enterobacteriaceae, and *S. aureus*, and infections related to trauma are associated with *Clostridium* species.

Although *S. aureus* is the typical cause of myositis, anaerobes—particularly *C. perfringens*—are often recovered from patients with pyogenic myositis. In anaerobic streptococcal myonecrosis, peptostreptococci are often identified along with group A streptococci or *S. aureus*. Patients typically present with fever, muscle pain, fatigue, and an elevated creatinine kinase level suggestive of muscle inflammation.

Bone and Joint Infections A comprehensive review of the world literature on anaerobic bone infections through 1975 included >650 cases. Of these, ~400 cases were caused by *Actinomyces* species; anaerobic cocci and *Bacteroides*, *Fusobacterium*, and *Clostridium* species were most commonly identified in the remaining cases. Actinomycotic involvement of the jaw was the most common bone infection, with the mandible involved four times as frequently as the maxilla. Patients with cervicofacial actinomycosis **(Chap. 180)** are often described as having a "lumpy jaw" because of the prominent soft tissue swelling that is sometimes mistaken for malignancy or granulomatous disease. These infections can be chronic in nature, can include the development of sinus tracts, can progress across normal tissue boundaries, and can require prolonged antibiotic treatment to prevent relapse. The vertebrae are the second most common location for *Actinomyces* infection; involvement of the thorax, abdomen, or pelvis is much less frequent.

Osteomyelitis involving anaerobes other than *Actinomyces* species most commonly develops by extension of an adjacent infection (e.g., soft tissue, paranasal sinus, or middle-ear infection). For example, diabetic foot ulcers and decubitus ulcers may be complicated by mixed aerobic–anaerobic osteomyelitis **(Chap. 136)**. Hematogenous seeding of bone by anaerobes is uncommon and is thought to occur in fewer than 10% of cases. The most common sites of anaerobic osteomyelitis are the head (skull and jaw) and the extremities. Fusobacteria have been isolated in pure culture from infections of the mastoid, mandible, and maxilla. *Clostridium* species have been reported as anaerobic pathogens in cases of osteomyelitis of the long bones following trauma. Anaerobic and microaerophilic cocci are most frequently isolated from infections involving the skull or mastoid; usually, these organisms are present in mixed cultures.

In contrast to anaerobic osteomyelitis, anaerobic arthritis **(Chap. 135)** is uncommon, typically involving a single isolate, and most cases are secondary to hematogenous spread. Although *Fusobacterium* species accounted for nearly one-third of cases in the preantibiotic era, *C. acnes*, peptostreptococci, and *B. fragilis* are now among the more frequent causes of anaerobic septic arthritis. Peptostreptococci and *C. acnes* are often found in association with prosthetic joints, *Fusobacterium* species have a predilection for the sternoclavicular and sacroiliac joints, and clostridial arthritis is especially common after trauma. As a frequent cause of bacteremia, *B. fragilis* is a common cause of anaerobic septic arthritis; however, arthritis occurs in fewer than 5% of patients with *B. fragilis* bacteremia.

Bacteremia *B. fragilis* is the anaerobe most commonly isolated from blood cultures. Although the frequency of positive cultures appeared to be decreasing in the 1980s, more recent evidence suggests that the rate is now increasing and that the increase may be related to changing demographics, with more patients who are elderly, immunocompromised, and/or receiving medications that may disrupt the mucosal barrier (e.g., chemotherapy). The source of bacteremia is most often an abscess in the abdomen, female genital tract, or soft tissue. At a large tertiary-care U.S. hospital, 0.8% of all positive blood cultures yielded an anaerobic gram-negative bacillus, with 0.5% yielding *B. fragilis*. A similar study in France revealed that 0.6% of all positive blood cultures yielded an anaerobic organism; 60% of these isolates were *Bacteroides* species, and 22% were *Clostridium* species. *Peptostreptococcus* and *Fusobacterium* species are also recovered with significant frequency.

Once the organism in the blood has been identified, both the portal of bloodstream entry and the underlying problem that probably led to seeding of the bloodstream can often be deduced from an understanding of the organism's normal site of residence. For example, mixed anaerobic bacteremia including *B. fragilis* usually implies a colonic pathology, with mucosal disruption from neoplasia, diverticulitis, or some other inflammatory lesion. The initial manifestations are determined by the portal of entry and reflect the localized condition. Although the clinical manifestations of *B. fragilis* bacteremia (e.g., rigors, hectic fevers) are similar to those of aerobic gram-negative bacillary bacteremia, the incidence of septic shock is lower with *B. fragilis*. This difference may be due to differences in the immunostimulatory effects of the different endotoxin structures.

In virtually all cases, isolation of a member of the *B. fragilis* group from blood indicates underlying infection that is associated with a mortality rate of 60% if untreated. It has been suggested that the mortality rate depends in part on the species recovered (*B. thetaiotaomicron* > *P. distasonis* > *B. fragilis*), but it is unclear whether differences in mortality rates relate to intrinsic differences in the virulence of these organisms, in their antimicrobial susceptibility profiles, and/or in the host's immune response. Case–fatality rates appear to increase with the increasing age of the patient (with reported rates of >66% among patients >60 years old), with the isolation of multiple species from the bloodstream, and with the failure to surgically remove a focus of infection.

Endocarditis **(See also Chap. 133)** Although gram-negative anaerobic bacteria only rarely cause endocarditis, their involvement is associated with significant mortality rates (21–43%). Members of the *B. fragilis* group are the most commonly identified gram-negative anaerobes in endocarditis. Anaerobic streptococci, which are often classified incorrectly, are likely responsible for this disease more frequently than is generally appreciated. Compared with aerobic bacterial endocarditis, endocarditis due to *Bacteroides* species is less likely to be associated with a history of cardiovascular disease and more likely to involve thromboembolic complications.

DIAGNOSIS

There are three critical steps in the diagnosis of anaerobic infection: (1) proper collection of specimens; (2) rapid transport of the specimens to the microbiology laboratory, preferably in anaerobic transport media; and (3) proper handling of the specimens by the laboratory. Specimens must be collected by meticulous sampling of infected sites, with avoidance of contamination by the normal microbiota. Samples from sites known to harbor numerous anaerobes (e.g., the mouth, nose, vagina, feces) are not acceptable for anaerobic culture as the presence of the normal microbiota will complicate interpretation of the results in a clinically meaningful manner. In contrast, samples from normally sterile locations (e.g., blood, pleural fluid, peritoneal fluid, CSF, and aspirates or biopsy samples from normally sterile sites) are appropriate for anaerobic culture in clinical microbiology laboratories. As a general rule, liquid or tissue specimens are preferred; if swab specimens must be used, special anaerobic swab systems should be used to help maintain persistence of anaerobes. Liquid samples should be collected in an air-free syringe that is then capped, injected into anaerobic transport bottles, or quickly transported to the clinical microbiology laboratory for immediate culture.

Because of the time and difficulty involved in the isolation of anaerobic bacteria, the diagnosis of anaerobic infections must frequently be based on presumptive evidence. As mentioned previously, anaerobic infections are sometimes suggested by specific clinical factors, such as origins from a site with an anaerobic-rich microbiota (e.g., the intestinal tract, oropharynx), the presence of an abscess, involvement of sites with lowered oxidation-reduction potential (e.g., avascular necrotic tissues), a foul odor, and the presence of gas in tissues. None of these features is necessarily pathognomonic or required for the diagnosis of an anaerobic infection, but these are helpful clues to keep in mind when constructing a differential diagnosis.

When cultures of obviously infected sites or purulent material yield no growth, streptococci only, or a single aerobic species (such as *E. coli*) and Gram's staining reveals a mixed bacterial population, the involvement of anaerobes should be suspected; the implication is that the anaerobic microorganisms have failed to grow because of inadequate transport and/or culture techniques. It is also important to remember that prior antibiotic therapy reduces the cultivability of these bacteria. Failure of an infection to respond to antibiotics that are not active against anaerobes (e.g., aminoglycosides and—in some circumstances—penicillin, cephalosporins, or tetracyclines) suggests an anaerobic etiology.

TREATMENT

Anaerobic Infections

Similar to successful therapy for other types of infection, treatment for anaerobic infections requires the administration of appropriate antibiotics, surgical debridement of devitalized tissues, and drainage of any large abscess. Any mucosal breach must be closed promptly to prevent ongoing infection.

ANTIBIOTIC THERAPY AND RESISTANCE

The antibiotics used to treat anaerobic infections should be active against both aerobic and anaerobic organisms because many of these infections are of mixed etiology. Antibiotic regimens can usually be selected empirically on the basis of the location of infection (which provides insight into the likely species involved), the severity of infection, and knowledge of local antimicrobial resistance patterns. Other factors influencing the selection of antibiotics include need for penetration into certain organs (such as the brain) and associated toxicity **(Chap. 149)**. As with all infections, the general maxim is to use the the narrowest-spectrum agent(s) possible so as to minimize the impact on the normal microbiota and the development of resistance.

Because of the slow growth rate of many anaerobes, the lack of standardized testing methods and of clinically relevant standards for resistance, and the generally good results obtained with empirical therapy, the role of antibiotic susceptibility testing of these organisms has been limited in most clinical microbiology laboratories. Instead, isolates are sent to reference laboratories for susceptibility testing when an infection is serious (e.g., brain abscess, meningitis, joint infection), is refractory, or requires prolonged therapy (e.g., osteomyelitis, prosthetic joint infection, endocarditis). Such testing should also be considered when a patient is not responding to antimicrobial therapy as expected; multidrug-resistant anaerobes have been reported. Antimicrobial susceptibility testing is also helpful in monitoring the activity of new drugs and recording current resistance patterns among anaerobic pathogens.

The need for susceptibility testing of anaerobic organisms is highlighted by increasing rates of antimicrobial resistance, geographic and institutional differences in susceptibility profiles, species-specific antibiograms, and the potential for worse clinical outcomes when ineffective antibiotics are used. These differences preclude making any sweeping generalizations regarding antibiotic therapy for anaerobic infections. For example, rates of resistance to piperacillin-tazobactam have remained low (≤1%) for all *Bacteroides* species in the United States, but *B. thetaiotaomicron* isolates in Korea have a notably higher resistance rate (17%). Clindamycin was historically effective against members of the *B. fragilis* group, but rates of resistance have increased to 30–43% in the United States and are >80% in some parts of the world. Furthermore, metronidazole is effective against many different anaerobic organisms and is considered a first-line agent for many anaerobic infections worldwide, but, in a population of Colombian patients with refractory periodontitis, 45% of *Fusobacterium* isolates and 25% of *Prevotella* and *Porphyromonas* strains were resistant to metronidazole; this finding underscores the importance of understanding the local antibiogram and of assessing susceptibility profiles in refractory disease.

Empirical Therapy Not every anaerobe isolated must be specifically targeted by the antibiotic regimen. Given that infections involving anaerobes are typically polymicrobial, that the cultivation and identification of anaerobes are challenging (i.e., not all organisms may be recovered), and that organisms often depend on one another for persistence, clinical resolution of the infection is often achieved with empirical antibiotics targeting the bulk of the organisms recovered. Antibiotics that demonstrate no useful activity against anaerobes include aminoglycosides, monobactams, and trimethoprim-sulfamethoxazole. With the caveat that susceptibility profiles may change with time and geography, the antibiotics that are commonly used

TABLE 182-2 Antimicrobial Therapy That Is Typically Active against Commonly Encountered Anaerobes

ANTIBIOTIC(S)	CAVEATS
Metronidazole	This drug is clinically unreliable against gram-positive non-spore-forming anaerobes (e.g., *Actinomyces* spp., *Propionibacterium* spp., *Peptostreptococcus* spp.).
β-Lactam/β-lactamase inhibitor combinations (ampicillin-sulbactam, ticarcillin–clavulanic acid, piperacillin-tazobactam)	Rates of resistance are increasing in some gram-negative anaerobes. The newer cephalosporin/β-lactamase combinations have limited anaerobic activity.
Clindamycin	Rates of resistance are increasing in *Bacteroides* spp.
Carbapenems (meropenem, imipenem, ertapenem, doripenem)	Rates of resistance are currently very low (<5%), although some carbapenemase-producing strains have been identified.
Chloramphenicol	Some clinical failures have been noted, even when the isolate is found to be susceptible by in vitro testing.

as empirical therapy against anaerobic bacteria include metronidazole, β-lactam/β-lactamase inhibitor combinations, clindamycin, carbapenems, and chloramphenicol (Table 182-2).

Metronidazole is active against gram-negative anaerobes, including nearly all isolates of *Bacteroides* species, and gram-positive spore-forming organisms, such as *C. difficile* (Chap. 139) and other *Clostridium* species. Given intrinsically reduced susceptibility, metronidazole is clinically unreliable against gram-positive non-spore-forming organisms, such as *Actinomyces*, *Propionibacterium*, *Lactobacillus*, *Bifidobacterium*, *Eubacterium*, and *Peptostreptococcus*. Of note, a few metronidazole-resistant *Bacteroides* isolates have been identified in the United States, and rates of such resistance have been increasing in Europe. Moreover, the rate of resistance to metronidazole has probably been greatly underestimated in some countries (e.g., the United Kingdom) that use metronidazole susceptibility to discriminate between obligate and facultative anaerobes (with obligate anaerobes defined by their susceptibility). Although the majority of metronidazole-resistant isolates have been identified in patients who have been exposed to the drug, resistant organisms have also been found in metronidazole-naïve patients.

More than 90% of clinical isolates from the *B. fragilis* group produce β-lactamases that are predominantly active against cephalosporins and that are highly active, cell associated, and produced constitutively. Thus, members of the *B. fragilis* group are presumed to be resistant to penicillin and ampicillin but may remain susceptible to extended-spectrum penicillins, particularly in combination with a β-lactamase inhibitor (e.g., ampicillin-sulbactam, piperacillin-tazobactam). Rates of resistance to ampicillin-sulbactam are increasing, particularly in *P. distasonis*, which has a reported resistance rate of 21% in the United States. Because β-lactamase production is not common in *Clostridium* species, these combination agents are usually effective. Of note, the newer cephalosporin/β-lactamase inhibitors (e.g., ceftolozane-tazobactam, ceftazidime-avibactam) have limited anaerobic activity.

Clindamycin is active against many anaerobes. However, rates of resistance to clindamycin among *Bacteroides* species increased in the United States from 7% in 1981 to 33% in 2010–2012. Resistance to clindamycin among non-*Bacteroides* gram-negative anaerobes is much less common (<10%). Some *Clostridium* species are resistant to clindamycin, although *C. perfringens* typically is not.

Carbapenems (ertapenem, doripenem, meropenem, and imipenem) are active against anaerobes, with fewer than 3% of *Bacteroides* isolates resistant. There is little difference among resistance rates for specific species, and, of the carbapenems, imipenem typically has the lowest resistance rate. Although the β-lactamase produced by most *Bacteroides* species is unable to inactivate carbapenems, rare *B. fragilis* strains have been reported to produce a carbapenemase.

Resistance to chloramphenicol is rare in *Bacteroides* species. Nationwide surveys in the United States have identified no resistant organisms, but some isolates with elevated minimal inhibitory concentrations (MICs)—i.e., 16 μg/mL—have been noted. Although chloramphenicol has excellent in vitro activity against all clinically relevant anaerobes, some clinical failures have been documented. Therefore, this drug may be less preferable if other active agents are available.

Other antibiotics with more variable activity against anaerobes include the fluoroquinolones and tigecycline. Although many fluoroquinolones (e.g., ciprofloxacin, levofloxacin, ofloxacin) display reasonable activity against anaerobic organisms other than *Bacteroides* species, these agents exhibit poor activity against the *B. fragilis* group. Rates of resistance to moxifloxacin are relatively high (39–83%) among *Bacteroides* isolates obtained in the United States but are much lower among *B. fragilis* and *B. thetaiotaomicron* isolates collected in Korea (8 and 2%, respectively) or Taiwan (8 and 15%, respectively). Tigecycline is active against most anaerobic bacteria, although MICs are somewhat higher for *Clostridium* species. Tigecycline's efficacy for treatment of complicated intraabdominal infections is comparable to that of imipenem, and it is therefore recommended as single-agent therapy for these infections.

Infections at Specific Sites In clinical situations, specific antibiotic regimens and durations must be tailored to the initial site of infection; the reader is referred to specific chapters on infections at specific sites for recommendations. In general, anaerobic infections are often broadly categorized as originating above or below the diaphragm. This distinction is clinically useful in that the predominant pathogens—and therefore the empirical antibiotic regimens—differ between these two categories of infection.

Infections above the diaphragm usually reflect the orodental microbiota, which includes *Prevotella*, *Porphyromonas*, *Fusobacterium*, and *Bacteroides* species other than the *B. fragilis* group along with streptococci (both aerobic and microaerophilic). Accordingly, antibiotic regimens should cover both aerobic and anaerobic bacteria. Given that >70% of these infections include a β-lactamase-producing organism, β-lactam drugs (penicillins and cephalosporins) are poor options as monotherapy. The recommended regimens include clindamycin, a β-lactam/β-lactamase inhibitor combination, or metronidazole in combination with a drug active against microaerophilic and aerobic streptococci (e.g., penicillin).

Anaerobic infections arising below the diaphragm (e.g., colonic and intraabdominal infections) must be treated specifically with agents active against *Bacteroides* species, including *B. fragilis*. Single agents suitable for this purpose include cefoxitin, moxifloxacin, a β-lactam/β-lactamase inhibitor combination, or a carbapenem. A two-drug regimen is an alternative, with one drug active against anaerobes and the other against coliforms (e.g., metronidazole with either a cephalosporin or a fluoroquinolone). In addition, if the clinician suspects that gram-positive facultative organisms such as enterococci are involved, therapeutic regimens should include ampicillin or vancomycin. Although clindamycin and cefotetan were previously considered acceptable options for intraabdominal infections involving anaerobes, these drugs are no longer recommended because of escalating rates of resistance in the *B. fragilis* group. Ampicillin-sulbactam is not recommended because of high rates of resistance among community-acquired strains of *E. coli* rather than because of resistance in anaerobic bacteria.

CNS infections involving anaerobic organisms may be treated with metronidazole, a carbapenem, chloramphenicol, or—if only gram-positive anaerobes are involved—penicillin. Clindamycin and cefoxitin have poor penetration into the CSF and should not be used. Cases of osteomyelitis in which a polymicrobial infection is identified from a bone biopsy specimen should be treated with a regimen that covers both aerobes and anaerobes, as some organisms that are often regarded as a contaminant (e.g., *C. acnes*) may have a pathogenic role. When an anaerobic organism is recognized as a major or sole pathogen infecting a joint, the duration of treatment should be similar to that used for arthritis caused by aerobic bacteria **(Chap. 135)**.

Although not every anaerobe needs to be covered with pathogen-directed therapy in most polymicrobial infections, several studies of *Bacteroides* bacteremia have clearly demonstrated that patients receiving effective therapy have lower mortality rates and more rapid sterilization of blood cultures than patients receiving ineffective therapy.

FAILURE OF THERAPY

Anaerobic infections that fail to respond to treatment or that relapse should be reassessed. Potential causes include an uncontrolled source of infection (e.g., ongoing intestinal leak into the peritoneum), superinfection with a new organism, and/or antibiotic failure. Additional imaging may be useful to discern whether surgical drainage or debridement is warranted. Obtaining additional culture specimens will help identify whether an organism resistant to the antibiotics being used is present. Strong consideration should be given to obtaining susceptibility profiles for the isolates.

■ FURTHER READING

Cooley L, Teng J: Anaerobic resistance: Should we be worried? Curr Opin Infect Dis 32:523, 2019.

Debreuil LJ: Fifty years devoted to anaerobes: Historical, lessons, and highlights. Eur J Clin Microbiol Infect Dis 43:1, 2024.

Finegold SM: Anaerobes: Problems and controversies in bacteriology, infections, and susceptibility testing. Rev Infect Dis 12:S223, 1990.

Reissier S et al: Recent trends in antimicrobial resistance among anaerobic clinical isolates. Microorganisms 11:1474, 2023.

Styrt B, Gorbach SL: Recent developments in the understanding of the pathogenesis and treatment of anaerobic infections (2). N Engl J Med 321:240, 1989.

Wexler HM: *Bacteroides*: The good, the bad, and the nitty-gritty. Clin Microbiol Rev 20:593, 2007.

Section 8 Mycobacterial Diseases

183 Tuberculosis

Mario C. B. Raviglione, Andrea Gori

Tuberculosis (TB), which is caused by bacteria of the *Mycobacterium tuberculosis* complex, is one of the oldest diseases known to affect humans. In 2023, after being replaced by COVID-19 during the three previous years, TB probably returned to be the top cause of infectious death from a single infectious agent. Population genomic studies suggest that *M. tuberculosis* may have emerged ~70,000 years ago in Africa and subsequently disseminated along with anatomically modern humans, expanding globally during the Neolithic Age as human density started to increase. This disease most often affects the lungs, although other organs are involved in up to one-third of cases. If properly treated, TB caused by drug-susceptible strains is curable in the vast majority of cases. If untreated, the disease may be fatal in more than 70% of people. Transmission usually takes place through the airborne spread of droplet nuclei produced by patients with infectious pulmonary TB. Through pharmacologic prophylaxis, the development of the disease can be prevented in those who have contracted TB infection.

ETIOLOGIC AGENT

Mycobacteria belong to the family Mycobacteriaceae and the order Actinomycetales. Of the pathogenic species belonging to the *M. tuberculosis* complex, which comprises eight distinct subgroups, the most common and important agent of human disease by far is *M. tuberculosis* (*sensu stricto*). A closely related organism isolated from cases in West, Central, and East Africa is *M. africanum*. The complex includes some zoonotic members, such as *M. bovis* (the bovine tubercle bacillus—characteristically resistant to pyrazinamide, once an important cause of TB transmitted by unpasteurized milk, and currently responsible for 140,000 human cases worldwide in 2020, half of them in Africa) and *M. caprae* (related to *M. bovis*). In addition, other organisms that have been reported rarely as causing TB include *M. pinnipedii* (a bacillus infecting seals and sea lions in the southern hemisphere and recently isolated from humans), *M. mungi* (isolated from banded mongooses in southern Africa), *M. orygis* (described in oryxes and other Bovidae in Africa and Asia and a potential cause of infection in humans), and *M. microti* (the "vole" bacillus, a less virulent organism). Finally, *M. canetti* is a rare isolate from East African cases that produces unusual smooth colonies on solid media and is considered closely related to a supposed progenitor type. There is no known environmental reservoir for any of these organisms.

M. tuberculosis is a rod-shaped, non-spore-forming, thin aerobic bacterium measuring 0.5 μm by 3 μm. Mycobacteria, including *M. tuberculosis*, are often neutral on Gram staining. However, once stained, the bacilli cannot be decolorized by acid alcohol; this characteristic justifies their classification as acid-fast bacilli (AFB; **Fig. 183-1**). Acid fastness is due mainly to the organisms' high content of mycolic acids, long-chain cross-linked fatty acids, and other cell-wall lipids. Microorganisms other than mycobacteria that display some acid fastness include species of *Nocardia* and *Rhodococcus*, *Legionella micdadei*, and the protozoa *Isospora* and *Cryptosporidium*. In the mycobacterial cell wall, lipids (e.g., mycolic acids) are linked to underlying arabinogalactan and peptidoglycan. This structure results in very low

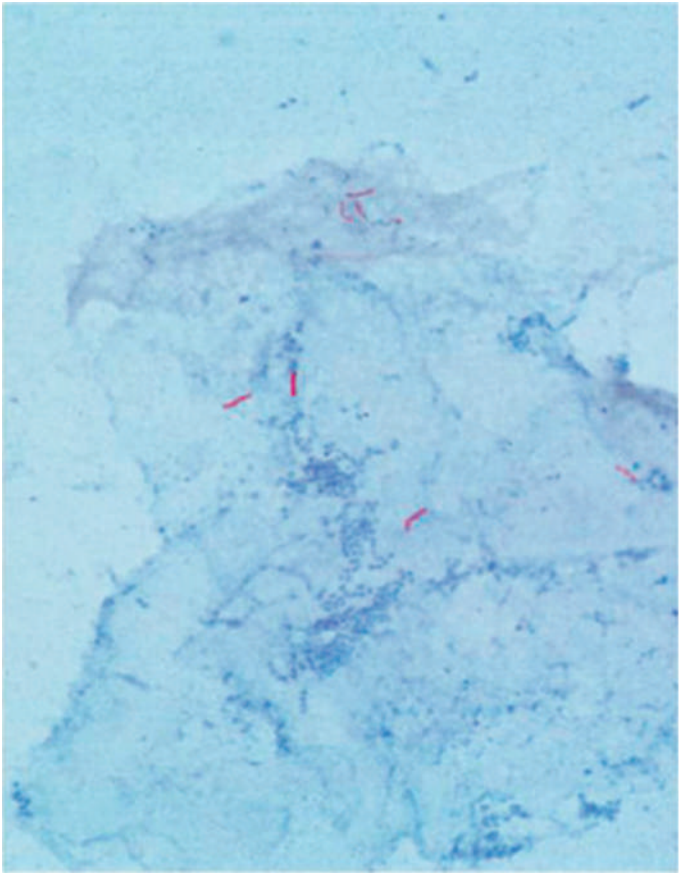

FIGURE 183-1 Acid-fast bacillus smear showing *M. tuberculosis* bacilli. *(Courtesy of the Centers for Disease Control and Prevention, Atlanta.)*

permeability of the cell wall, thus reducing the effectiveness of most antibiotics.

The complete genome sequence of *M. tuberculosis* comprises 4.4 million base pairs, 4043 genes encoding 3993 proteins, and 50 genes encoding stable RNAs; its high guanine-plus-cytosine content (65.6%) is indicative of an aerobic "lifestyle." A large proportion of genes are devoted to the production of enzymes involved in cell wall metabolism. Substantial genetic variability exists among the innumerable *M. tuberculosis* strains from different parts of the world. Based on such genetic variability it is possible to distinguish and compare different strains. Their distinction is important to study transmission dynamics and identify outbreaks. Starting in the 1990s, reproducible genotyping methods were developed to type the bacterium. Initially, they included insertion sequence 6110 (IS6110), restriction fragment length polymorphism (RFLP) typing, and spoligotyping. Lately, most studies utilize mycobacterial interspersed repetitive unit variable number tandem repeats (MIRU-VNTRs) and whole genome sequencing analysis.

EPIDEMIOLOGY

In 2023, 8.2 million new cases of TB (all forms, both pulmonary and extrapulmonary) were reported to the World Health Organization (WHO), around 97% of which were reported from low- and middle-income countries. However, because of insufficient case detection and incomplete notification, reported cases represent only about two-thirds of the total estimated cases. The WHO estimated that 10.8 million (range, 10.1–11.7 million, rate 134 per 100,000 persons) new (incident) cases of TB occurred worldwide in 2023. Eight countries accounted for two thirds of all cases: India (26%), Indonesia (10%), China (6.8%), the Philippines (6.8%), Pakistan (6.3%), Nigeria (4.6%), Bangladesh (3.5%) and the Democratic Republic of the Congo (3.1%). Of all cases, 55% occurred in male patients, 33% in female patients, and 12% in children. It is further estimated that 1.25 million (range 1.13–1.37 million) deaths from TB, including 160,000 among persons with HIV co-infection, occurred in 2023; 98% of these deaths were in low- and middle-income countries. Estimates of TB incidence and mortality rates (per 100,000 population) in 2022 are depicted in **Figs. 183-2 and 183-3**, respectively).

During the past few years, numbers of reported cases have stabilized or are slowly declining in most high-income countries. In the United States, TB cases and incidence rates steadily decreased from 1992 to 2021. In 2022, 8300 cases of TB (2.5 cases per 100,000 population) were reported to the U.S. Centers for Disease Control and Prevention (CDC), a slight increase from the 7882 cases reported in 2021. In the United States, TB is uncommon among young white adults of European descent, who have only rarely been exposed to *M. tuberculosis* infection during recent decades. In contrast, because of a high risk of transmission in the past, the prevalence of *M. tuberculosis* infection is relatively high among elderly whites; overall, 13 million persons are estimated to be "latently" infected. In general, adults ≥65 years of age have the highest incidence rate per capita and children <14 years of age the lowest. Of the total 7882 cases in 2021, 28% were among U.S.-born persons and 71% among non-U.S.-born persons. Non-Hispanic Black or African American persons accounted for the highest proportion of cases (34%), followed by non-Hispanic White persons (29%), and Hispanic/Latinos (24%). Among non-U.S.-born persons in the United States in 2021, 48% occurred in persons born in Asia, 33% among Hispanic/Latino persons, and 12% among African American persons. Overall, the highest rates per capita were among non-U.S.-born persons of more than one ethnicity (30 cases per 100,000 population). In 2020 in the United States 600 deaths were caused by TB, the highest number since 2006. In 2021 in Canada, TB cases and rates per 100,000 population were 1829 and 4.8, respectively. Of these cases, 76% occurred in foreign-born persons. However, the highest rates per 100,000 population were among Inuit (135) and First Nation persons (16). Similarly, in Europe, TB has reemerged as an important public health problem, mainly as a result of cases among immigrants from high-incidence countries and among marginalized populations, often in large urban settings. In most Western European countries, there are more cases annually among foreign-born than native populations.

In 2023 TB incidence was decreasing in Africa, resuming a slow decline in the European, Eastern Mediterranean, and Southeast Asia

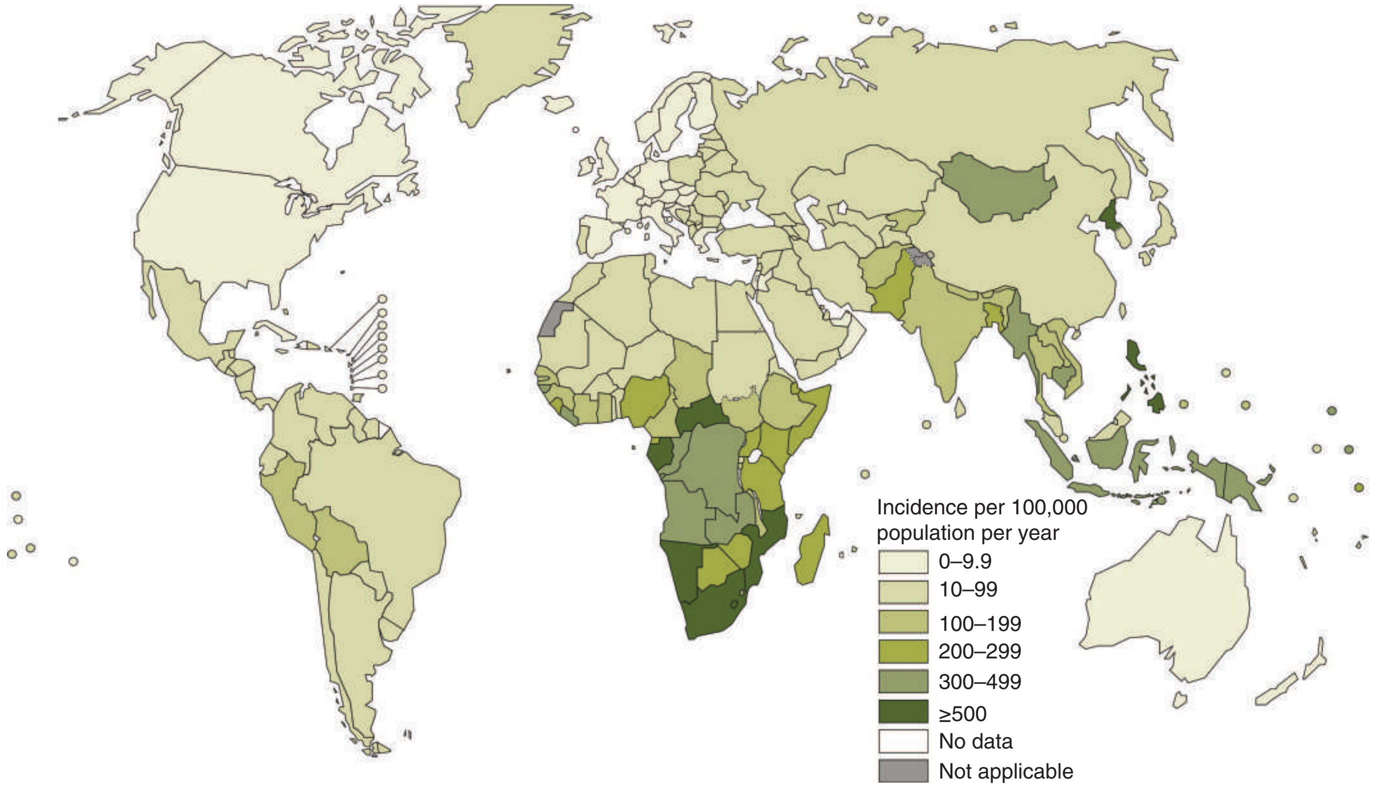

FIGURE 183-2 Estimated tuberculosis (TB) incidence rates (per 100,000 population) in 2022. The designations used and the presentation of material on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization (WHO) concerning the legal status of any country, territory, city, or area or of its authorities or concerning the delimitation of its frontiers or boundaries. *Dotted*, *dashed*, and *white lines* represent approximate border lines for which there may not yet be full agreement. *(Reproduced with permission from Global Tuberculosis Report 2023. Geneva, World Health Organization; 2023.)*

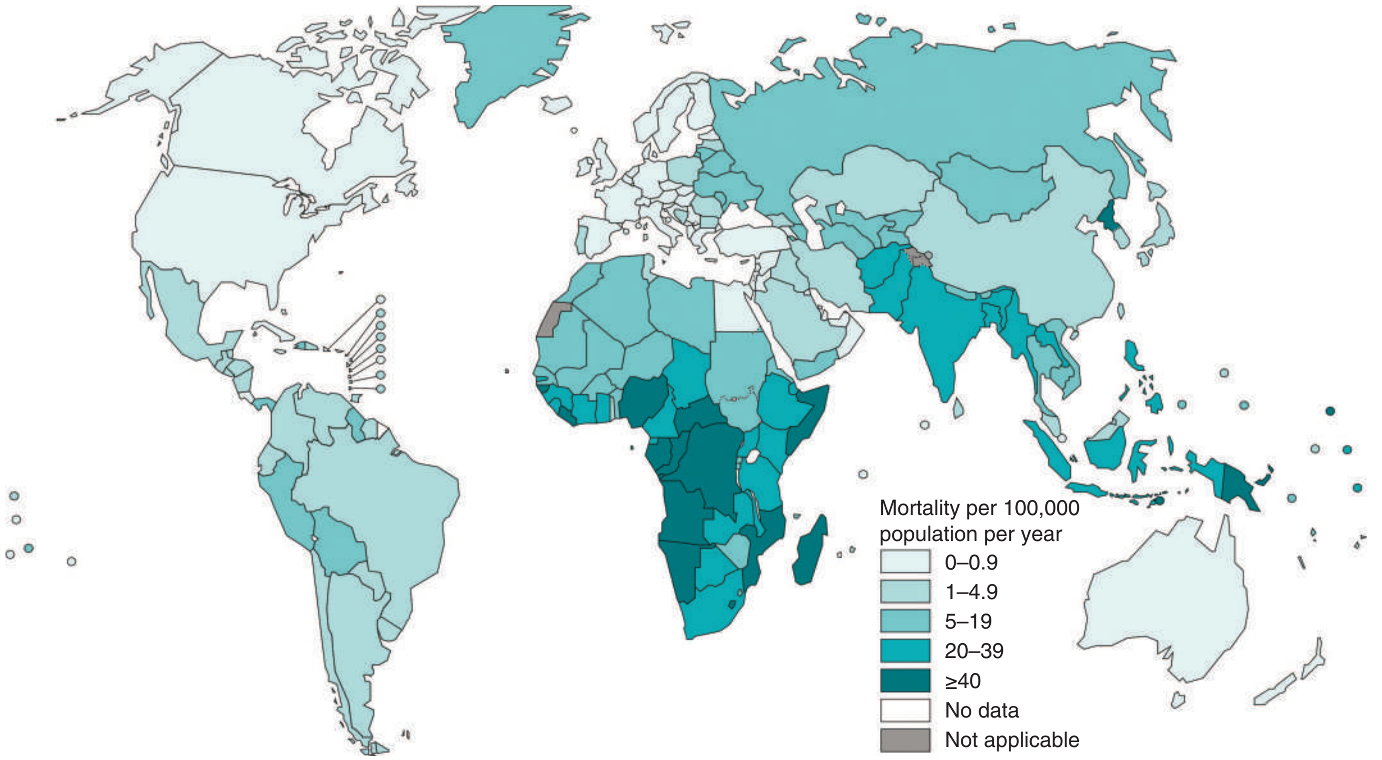

FIGURE 183-3 Estimated tuberculosis (TB) mortality rates in HIV-negative persons in 2022. *(See disclaimer in Fig. 183-2. Reproduced with permission from Global Tuberculosis Report 2023. Geneva, World Health Organization; 2023.)*

regions following the COVID-19 pandemic, and still increasing in the Americas and the Western Pacific. Globally, TB incidence has increased by nearly 5% since 2020 as a result of essential service disruptions during the worst period of the pandemic.

Of the estimated 10.8 million new cases of TB in 2023, 6.1% were in people living with HIV, with the majority of them occurring in Africa. Furthermore, an estimated 400,000 (range, 360,000-440,000) cases of rifampin-resistant (also called rifampicin-resistant) TB (RR-TB) and multidrug-resistant TB (MDR-TB)—a form of the disease caused by bacilli resistant at least to isoniazid and rifampin—occurred in 2023, representing 3.2% and 16%, respectively, of all new and previously treated cases. Only 189,000 MDR/RR-TB cases were enrolled on treatment in 2023 because of a lack of culture and drug susceptibility testing (DST) capacity in many settings worldwide. The countries of the former Soviet Union remain those with the highest proportions of MDR/RR disease among new TB cases, reaching up to one-third of the total. Overall, more than 50% of all MDR/RR-TB cases occur in India (27%), the Russian Federation (7.4%), Indonesia (7.4%), China (7.3%) and the Philippines (7.2%). Cases of extensively drug-resistant TB (XDR-TB), in which MDR-TB is compounded by additional resistance to any fluoroquinolones and at least one additional group A drug (e.g., bedaquiline and linezolid), and cases of pre-XDR-TB, a form in which MDR/RR-TB strains are also resistant to any fluoroquinolone, occur worldwide. However, the vast majority of XDR-TB cases remain undiagnosed because reliable methods for DST are still lacking and laboratory capacity in low-income countries is limited.

FROM EXPOSURE TO INFECTION

M. tuberculosis is most commonly transmitted from a person with infectious pulmonary TB by droplet nuclei containing *M. tuberculosis* bacteria, which are aerosolized by coughing, sneezing, or speaking. The tiny droplets dry rapidly; the smallest (<5–10 μm in diameter) may remain suspended in the air for several hours and may reach the terminal air passages when inhaled. Other routes of transmission of tubercle bacilli (e.g., through the skin or the placenta) are uncommon and of no epidemiologic significance. The risk of transmission and of subsequent acquisition of *M. tuberculosis* infection is determined mainly by exogenous factors, although endogenous factors also may play a role. The probability of contact with a person who has an infectious form of TB, the intimacy and duration of that contact, the degree of infectiousness of the case, and the shared environment in which the contact takes place are all important determinants of the likelihood of transmission. The most infectious patients whose sputum contains AFB visible by microscopy have cavitary pulmonary disease or, much less commonly, laryngeal TB and produce sputum containing as many as 10^5–10^7 AFB/mL. Patients with sputum smear–negative/culture-positive TB are less infectious, although they have been responsible for up to 20% of transmission in some studies in the United States. Those with culture-negative pulmonary TB and extrapulmonary TB are essentially noninfectious. Because persons with both HIV infection and TB are less likely to have cavitations, they may be less infectious than those without HIV co-infection. Crowding in poorly ventilated rooms is one of the most important factors in the transmission of tubercle bacilli because it increases the intensity of contact with a case. The virulence of the transmitted organism is also an important factor in establishing infection. Endogenous factors such as the degree of immune competence also are important. In particular, HIV-infected patients, persons undergoing cancer treatment, or those administered immunosuppressive drugs may be at higher risk of TB infection acquisition. Attempts to estimate the basic reproductive number R_0 for TB have resulted in a wide range of values depending on environmental conditions and social behaviors of populations: from 0.24 in the Netherlands during the period 1933–2007 to 4.3 in China in 2012, reflecting the status of disease control.

FROM INFECTION TO DISEASE

Unlike the risk of acquiring infection with *M. tuberculosis*, the risk of developing disease after being infected depends largely on endogenous factors, such as the individual's innate immunologic and nonimmunologic defenses and the level at which the individual's cell-mediated immunity is functioning. Clinical illness directly following infection is classified as *primary TB* and is common among children in the first few years of life and among immunocompromised persons. Although primary TB may be severe and disseminated, it generally is not associated with high-level transmissibility. When infection is acquired later in life, the chance is greater that the mature immune system will

TABLE 183-1 Risk Factors for Active Tuberculosis in Persons Who Have Been Infected with Tubercle Bacilli

FACTOR	RELATIVE RISK/ODDS[a]
Recent infection (<1 year)	12.9
Fibrotic lesions (spontaneously healed)	2–20
Comorbidities and iatrogenic causes	
HIV infection	21–>30
Silicosis	30
Chronic renal failure/hemodialysis	10–25
Diabetes	2–4
IV drug use	10–30
Excessive alcohol use	3
Immunosuppressive treatment	10
Tumor necrosis factor α inhibitors	4–5
Gastrectomy	2–5
Jejunoileal bypass	30–60
Posttransplantation period (renal, cardiac)	20–70
Tobacco smoking	2–3
Malnutrition and severe underweight	2

[a] Old infection = 1.

contain it at least temporarily. Bacilli, however, may persist for years before reactivating to produce *secondary* (or *postprimary*) *TB*, which, because of frequent cavitation, is more often infectious than is primary disease. Overall, it is estimated that up to 10% of infected persons will eventually develop active TB in their lifetime—half of them during the first 18 months after infection. The risk is much higher among immunocompromised individuals and, particularly, HIV-infected persons. Reinfection of a previously infected individual, which is common in areas with high rates of TB transmission, may also favor the development of disease. At the height of the TB resurgence in the United States in the early 1990s, molecular typing and comparison of strains of *M. tuberculosis* suggested that up to one-third of cases of active TB in some inner-city communities were due to recent transmission rather than to reactivation of old infection. Age is an important determinant of the risk of disease after infection. Among infected persons, the incidence of TB is highest during late adolescence and early adulthood; the reasons are unclear. The incidence among women peaks at 25–34 years of age. In this age group, rates among women may be higher than those among men, whereas at older ages the opposite is true. The risk increases in the elderly, possibly because of waning immunity and comorbidity.

A variety of diseases and conditions favor the development of active TB **(Table 183-1)**. In absolute terms, the most potent risk factor for TB among infected individuals is HIV co-infection, which suppresses cellular immunity. The risk that infection will proceed to active disease is directly related to the patient's degree of immunosuppression. In a study of people living with HIV (PLWH), those who were tuberculin skin test (TST)–positive had risks varying from 2.6 to 13.3 cases per 100 person-years, which increased as the CD4+ T cell count decreased.

NATURAL HISTORY OF DISEASE

Studies conducted in various countries before the advent of antimicrobial TB therapy showed that untreated TB is often fatal. About one-third of patients died within 1 year after diagnosis. Historic data also show that 55% of sputum smear-positive cases were dead within 5 years and up to 86% (weighted mean, 70%) within 10 years. A lower case fatality rate, around 20%, was estimated for untreated paucibacillary smear-negative cases at 5 years. Of the survivors at 5 years, ~60% had undergone spontaneous remission, while the remainder were still excreting tubercle bacilli. With effective, timely, and proper antimicrobial TB treatment, patients have a very high chance of being cured. However, improper use of anti-TB drugs, while reducing mortality rates, may also result in large numbers of chronic infectious cases, often with drug-resistant bacilli.

PATHOGENESIS AND IMMUNITY

INFECTION AND MACROPHAGE INVASION

The interaction of *M. tuberculosis* with the human host begins when droplet nuclei containing viable microorganisms, propelled into the air by infectious patients, are inhaled by a close bystander. Although the majority of inhaled bacilli are trapped in the upper airways and expelled by ciliated mucosal cells, a fraction (usually <10%) reach the alveoli, a unique immunoregulatory environment. There, in the very early phases of infection, the predominant cells infected by *M. tuberculosis* are myeloid dendritic cells. Subsequently, alveolar macrophages that have not yet been activated (prototypic alternatively activated macrophages) phagocytose the bacilli. Adhesion of mycobacteria to macrophages results largely from binding of the bacterial cell wall to a variety of macrophage cell-surface receptor molecules, including complement receptors, the mannose receptor, the immunoglobulin G $Fc\gamma$ receptor, and type A scavenger receptors. Surfactants may also play a role in the early phase of interaction between the host and the pathogen, and surfactant protein D can prevent phagocytosis. Phagocytosis is enhanced by complement activation, leading to opsonization of bacilli with C3 activation products such as C3b and C3bi. Concomitantly, binding of certain receptors, such as the mannose receptor, regulates postphagocytic events like phagosome–lysosome fusion and inflammatory cytokine production. After a phagosome forms, the survival of *M. tuberculosis* in the cell seems to depend in part on reduced acidification due to lack of assembly of a complete vesicular proton-adenosine triphosphatase. A complex series of events is generated by the bacterial cell-wall lipoglycan lipoarabinomannan, which inhibits the intracellular increase of Ca^{2+}. Thus, the Ca^{2+}/calmodulin pathway (leading to phagosome–lysosome fusion) is impaired, and the bacilli survive within the phagosomes by blocking fusion. The *M. tuberculosis* phagosome inhibits the production of phosphatidylinositol 3-phosphate, which normally earmarks phagosomes for membrane sorting and maturation (including phagolysosome formation), which would destroy the bacteria. Bacterial factors block the host defense of autophagy, in which the cell sequesters the phagosome in a double-membrane vesicle (*autophagosome*) that is destined to fuse with lysosomes. If the bacilli are successful in arresting phagosome maturation, then bacterial replication begins and the macrophage eventually ruptures and releases its bacillary contents. This process is mediated by the ESX-1 secretion system that is encoded by genes contained in the region of difference 1 (RD1). Other uninfected phagocytic cells are then recruited to continue the infection cycle by ingesting dying macrophages and their bacillary content, thus, in turn, becoming infected themselves and expanding the infection.

VIRULENCE OF TUBERCLE BACILLI

M. tuberculosis must be viewed as a complex formed by a multitude of strains that differ in virulence and are capable of producing a variety of manifestations of disease. Since the elucidation of the *M. tuberculosis* genome in 1998, large mutant collections have been generated, and many bacterial genes that contribute to *M. tuberculosis* virulence have been found. Moreover, different patterns of virulence defects have been defined in various animal models—predominantly mice but also guinea pigs, rabbits, and nonhuman primates. The *katG* gene encodes for a catalase/peroxidase enzyme that protects against oxidative stress and is required for isoniazid activation and subsequent bactericidal activity. RD1 is a 9.5-kb locus that encodes two key small protein antigens—the 6-kDa early secretory antigen (ESAT-6) and culture filtrate protein 10 (CFP-10)—as well as a putative secretion apparatus that may facilitate their egress; the absence of this locus in the vaccine strain *M. bovis* bacille Calmette-Guérin (BCG) is a key attenuating mutation. Mutants lacking key enzymes of bacterial biosynthesis become auxotrophic for the missing substrate and often are totally unable to proliferate in animals; these include the *leuCD* and *panCD* mutants, which require leucine and pantothenic acid, respectively. The isocitrate lyase gene (*icl1*) encodes a key step in the glyoxylate shunt that facilitates bacterial growth on fatty acid substrates; this gene is required for long-term persistence of *M. tuberculosis* infection in mice with chronic TB. *M. tuberculosis* mutants in regulatory genes such as sigma factor C and sigma factor H (*sigC* and *sigH*) are associated with normal bacterial growth in mice, but they fail to elicit full tissue pathology. Finally, the mycobacterial protein CarD (expressed by the *carD* gene) seems essential for the control of rRNA transcription that is required for mycobacterial replication and persistence in the host cell. Its loss exposes mycobacteria to oxidative stress, starvation, DNA damage, and ultimately sensitivity to killing by a variety of host mutagens and defense mechanisms.

INNATE RESISTANCE TO INFECTION

Several observations suggest that genetic factors play a key role in innate resistance to infection with *M. tuberculosis* and the development of disease. The existence of this resistance, which is polygenic in nature, is suggested by the differing degrees of susceptibility to TB in different populations. This mechanism of elimination of the pathogen may be accompanied by negative results in the TST and interferon γ (IFN-γ) release assays (IGRAs). In mice, a gene called *Nramp1* (natural resistance–associated macrophage protein 1) plays a regulatory role in resistance/susceptibility to mycobacteria. The human homologue *NRAMP1*, which maps to chromosome 2q, may play a role in determining susceptibility to TB, as is suggested by a study among West Africans. Studies of mice identified a novel host resistance gene, *ipr1*, that is encoded within the *sst1* locus; *ipr1* encodes an IFN-inducible nuclear protein that interacts with other nuclear proteins in macrophages primed with IFNs or infected by *M. tuberculosis*. In addition, polymorphisms in multiple genes, such as those encoding for various major histocompatibility complex alleles, IFN-γ, T cell growth factor β, interleukin (IL) 10, mannose-binding protein, IFN-γ receptor, Toll-like receptor 2, vitamin D receptor, and IL-1, have been associated with susceptibility to TB.

THE HOST RESPONSE, GRANULOMA FORMATION, AND "LATENCY"

In the initial stage of host–bacterium interaction, prior to the onset of an acquired cell-mediated immune (CMI) response, *M. tuberculosis* disseminates widely through the lymph vessels, spreading to other sites in the lungs and other organs, and undergoes a period of extensive growth within naïve inactivated macrophages; additional naïve macrophages are recruited to the early granuloma. How the bacillus accesses the parenchymal tissue still needs to be elucidated: it may directly infect epithelial cells or transmigrate through infected macrophages across the epithelium. Infected dendritic cells or monocytes then begin to transport bacilli to the lymphatic system. Studies suggest that *M. tuberculosis* uses specific virulence mechanisms to subvert host cellular signaling and to elicit an early regulated proinflammatory response that promotes granuloma expansion and bacterial growth during this key early phase. A study of *M. marinum* infection in zebrafish has delineated one molecular mechanism by which mycobacteria induce granuloma formation. The mycobacterial protein ESAT-6 induces secretion of matrix metalloproteinase 9 (MMP9) by nearby epithelial cells that are in contact with infected macrophages. MMP9 in turn stimulates recruitment of naïve macrophages, thus inducing granuloma maturation and bacterial growth. Disruption of MMP9 function results in reduced bacterial growth. Another study has shown that *M. tuberculosis*–derived cyclic AMP is secreted from the phagosome into host macrophages, subverting the cell's signal transduction pathways and stimulating an elevation in the secretion of tumor necrosis factor α (TNF-α) as well as further proinflammatory cell recruitment. Ultimately, the chemoattractants and bacterial products released during the repeated rounds of cell lysis and infection of newly arriving macrophages enable dendritic cells to access bacilli; these cells migrate to the draining lymph nodes and present mycobacterial antigens to T lymphocytes. At this point, the development of cell-mediated and humoral immunity begins. These initial stages of infection are usually asymptomatic.

About 2–4 weeks after infection, two host responses to *M. tuberculosis* develop: a macrophage-activating CMI response and a tissue-damaging response. The *macrophage-activating response* is a T cell–mediated phenomenon resulting in the activation of macrophages that are capable of killing and digesting tubercle bacilli. The *tissue-damaging response* is the result of a delayed-type hypersensitivity reaction to various bacillary antigens; it destroys inactivated macrophages that contain multiplying bacilli but also causes caseous necrosis of the involved tissues (see below). Although both of these responses can inhibit mycobacterial growth, it is the balance between the two that determines the forms of TB that will develop subsequently. With the development of specific immunity and the accumulation of large numbers of activated macrophages at the site of the primary lesion, granulomatous lesions (*tubercles*) are formed. These lesions consist of accumulations of lymphocytes and activated macrophages that evolve toward epithelioid and giant cell morphologies. Initially, the tissue-damaging response can limit mycobacterial growth within macrophages. As stated above, this response, mediated by various bacterial products, not only destroys macrophages but also produces early solid necrosis in the center of the tubercle. Although *M. tuberculosis* can survive, its growth is inhibited within this necrotic environment by low oxygen tension and low pH. At this point, some lesions may heal by fibrosis, with subsequent calcification, whereas inflammation and necrosis occur in other lesions. Some observations have challenged the traditional view that any encounter between mycobacteria and macrophages results in chronic infection. It is possible that an immune response capable of eradicating early infection may sometimes develop as a consequence, for instance, of disabling mutations in mycobacterial genomes rendering their replication ineffective. Individual granulomas that are formed during this phase of infection can vary in size and cell composition; some can contain the spread of mycobacteria, while others cannot. TB infection ensues as a result of this dynamic balance between the microorganism and the host. For many years, TB infection has been called "latent TB infection (LTBI)." This terminology was used to define a state of persistent immune response to stimulation by *M. tuberculosis* antigens with no evidence of clinically manifest, active TB. The qualification "latent" may offer some convenience of distinguishing infection from disease, albeit an inaccurate description of a process that encompasses bacterial generations that are not dormant. *Latency* may be an inaccurate term because bacilli may remain active during this "latent" stage, forming biofilms in necrotic areas within which they temporarily hide. Therefore, some have proposed the term *persister* as a more accurate descriptor of the behavior of the bacilli in this phase. It is important to recognize that infection and disease do not represent a binary state but rather a continuum along which infection will eventually move in the direction of full containment or disease. The ability to predict, through systemic biomarkers, which infected individuals will progress toward disease would be of immense value in devising prophylactic interventions at scale.

MACROPHAGE-ACTIVATING RESPONSE

Cell-mediated immunity is critical at this early stage. In the majority of infected individuals, local macrophages are activated when bacillary antigens processed by macrophages stimulate T lymphocytes to release a variety of lymphokines. These activated macrophages aggregate around the lesion's center and effectively neutralize tubercle bacilli without causing further tissue destruction. In the central part of the lesion, the necrotic material resembles soft cheese (*caseous necrosis*)—a phenomenon that may also be observed in other conditions, such as neoplasms. Even when healing takes place, viable bacilli may remain within macrophages or in the necrotic material for many years. These "healed" lesions in the lung parenchyma and hilar lymph nodes may later undergo calcification.

DELAYED-TYPE HYPERSENSITIVITY

In a minority of cases, the macrophage-activating response is weak, and mycobacterial growth can be inhibited only by intensified delayed hypersensitivity reactions, which lead to lung tissue destruction. The lesion tends to enlarge further, and the surrounding tissue is progressively damaged. At the center of the lesion, the caseous material liquefies. Bronchial walls and blood vessels are invaded and destroyed, and cavities are formed. The liquefied caseous material, containing large amounts of bacilli, is drained through bronchi. Within the cavity, tubercle bacilli multiply, spill into the airways, and are discharged into the environment through expiratory maneuvers such as coughing and talking. In the early stages of infection, bacilli are usually transported by macrophages to regional lymph nodes, from which they gain access to the central venous return; from there they reseed the lungs and may also disseminate beyond the pulmonary vasculature throughout the body via the systemic circulation. In young children with poor natural immunity, hematogenous dissemination may result in fatal miliary TB or tuberculous meningitis.

ROLE OF MACROPHAGES AND MONOCYTES

While cell-mediated immunity confers partial protection against *M. tuberculosis*, humoral immunity plays a less well-defined role in protection (although evidence is accumulating on the existence of antibodies to lipoarabinomannan, which may prevent dissemination of infection in children). In cell-mediated immunity, two types of cells are essential: macrophages, which directly phagocytose tubercle bacilli, and T cells (mainly CD4+ T lymphocytes, although the role of CD8+ T cells has recently been the subject of much research), which induce protection through the production of cytokines, especially IFN-γ. After infection with *M. tuberculosis*, alveolar macrophages secrete various cytokines responsible for several events (e.g., the formation of granulomas) as well as systemic effects (e.g., fever and weight loss). However, alternatively activated alveolar macrophages may be particularly susceptible to *M. tuberculosis* growth early on, given their more limited proinflammatory and bactericidal activity, which is related in part to being bathed in surfactant. New monocytes and macrophages attracted to the site are key components of the immune response. Their primary mechanism is probably related to production of oxidants (such as reactive oxygen intermediates or nitric oxide) that have antimycobacterial activity and increase the synthesis of cytokines such as TNF-α and IL-1, which in turn regulate the release of reactive oxygen intermediates and reactive nitrogen intermediates. In addition, macrophages can undergo apoptosis—a defensive mechanism to prevent the release of cytokines and bacilli via their sequestration in the apoptotic cell. Recent work also describes the involvement of neutrophils in the host response, although the timing of their appearance and their effectiveness remain uncertain.

ROLE OF T LYMPHOCYTES

Alveolar macrophages, monocytes, and dendritic cells are also critical in processing and presenting antigens to T lymphocytes, primarily CD4+ and CD8+ T cells; the result is the activation and proliferation of CD4+ T lymphocytes, which are crucial to the host's defense against *M. tuberculosis*. Qualitative and quantitative defects of CD4+ T cells explain the inability of HIV-infected individuals to contain mycobacterial proliferation. Activated CD4+ T lymphocytes can differentiate into cytokine-producing T_H1 or T_H2 cells. T_H1 cells produce IFN-γ—an activator of macrophages and monocytes—and IL-2. T_H2 cells produce IL-4, IL-5, IL-10, and IL-13 and may also promote humoral immunity. The interplay of these various cytokines and their cross-regulation determine the host's response. The role of cytokines in promoting intracellular killing of mycobacteria, however, has not been entirely elucidated. IFN-γ may induce the generation of reactive oxygen and nitrogen intermediates and regulate genes involved in bactericidal effects. TNF-α also is important. Although its precise mechanisms are complex and not yet fully clarified, there is a suggested model that foresees an ideal setting for TNF-α between excessive activation—with consequent worsening of immunopathologic reactions—and insufficient activation—with resulting lack of containment—in the control of TB infection. Observations made originally in transgenic knockout mice and more recently in humans suggest that other T cell subsets, especially CD8+ T cells, may play an important role. CD8+ T cells have been associated with protective activities via cytotoxic responses and lysis of infected cells as well as with production of IFN-γ and TNF-α. Finally, natural killer cells act as co-regulators of CD8+ T-cell lytic activities, and γδ T cells are increasingly thought to be involved in protective responses in humans.

MYCOBACTERIAL LIPIDS AND PROTEINS

Lipids are involved in mycobacterial recognition by the innate immune system, and lipoproteins (such as 19-kDa lipoprotein) trigger potent signals through Toll-like receptors present in blood dendritic cells. *M. tuberculosis* possesses various protein antigens. Some are present in the cytoplasm and cell wall; others are secreted. That the latter are more important in eliciting a T lymphocyte response is suggested by experiments documenting the appearance of protective immunity in animals after immunization with live, protein-secreting mycobacteria. Among the antigens that may play a protective role are some proteins of the antigen 85 complex are the 30-kDa major secretory protein (85B) and ESAT-6 antigens. Protective immunity is probably the result of reactivity to many different mycobacterial antigens. These antigens are being incorporated into newly designed vaccines on various platforms.

SKIN-TEST REACTIVITY

Coincident with the appearance of immunity, delayed-type hypersensitivity to *M. tuberculosis* develops. This reactivity is the basis of the TST, which is used primarily for the diagnosis of *M. tuberculosis* infection in persons without symptoms. The cellular mechanisms responsible for TST reactivity are related mainly to previously sensitized CD4+ T lymphocytes, which are attracted to the skin-test site. There, they proliferate and produce cytokines. Although delayed hypersensitivity is associated with protective immunity (TST-positive persons are less susceptible to a new *M. tuberculosis* infection than are TST-negative persons), it by no means guarantees protection against reactivation. In fact, cases of active TB are often accompanied by strongly positive skin-test reactions. There is also evidence of reinfection with new strains of *M. tuberculosis* in patients previously treated for active disease. This evidence underscores the fact that previous infection or active TB does not necessarily confer fully protective immunity.

CLINICAL MANIFESTATIONS

TB is classified as pulmonary, extrapulmonary, or both. Depending on several factors linked to host immunologic status and bacterial strains, extrapulmonary TB may occur in 10–40% of patients. Furthermore, up to two-thirds of HIV-infected patients with TB may have both pulmonary and extrapulmonary TB or extrapulmonary TB alone.

PULMONARY TB

Pulmonary TB is traditionally categorized as primary or postprimary (adult-type, secondary). This distinction has been challenged by molecular evidence from TB-endemic areas indicating that a large percentage of cases of adult pulmonary TB result from recent infection (either primary infection or reinfection) and not from reactivation.

Primary Disease Primary pulmonary TB occurs soon after the initial infection. It may be asymptomatic or may present with fever and occasionally pleuritic chest pain. In areas of high TB transmission, this form of disease is often seen in children. Because most inspired air is distributed to the middle and lower lung zones, these areas are most commonly involved in primary TB. The lesion forming after initial infection (*Ghon focus*) is usually peripheral and accompanied by transient hilar or paratracheal lymphadenopathy, which may or may not be visible on standard chest radiography (CXR) **(Fig. 183-4)**. Some patients develop erythema nodosum on the legs **(see Fig. A1-39)** or phlyctenular conjunctivitis. In the majority of cases, the lesion heals spontaneously and becomes evident only as a small calcified nodule. Pleural reaction overlying a subpleural focus also is common. The Ghon focus, with or without overlying pleural reaction, thickening, and regional lymphadenopathy, is referred to as the *Ghon complex*.

In young children with immature cell-mediated immunity and in persons with impaired immunity (e.g., those with malnutrition or HIV infection), primary pulmonary TB may progress rapidly to clinical illness. The initial lesion increases in size and can evolve in different ways. Pleural effusion, which is found in up to two-thirds of cases, results from the penetration of bacilli into the pleural space from an adjacent subpleural focus. In severe cases, the primary site rapidly enlarges, its central portion undergoes necrosis, and cavitation develops (*progressive primary TB*). TB in young children is almost

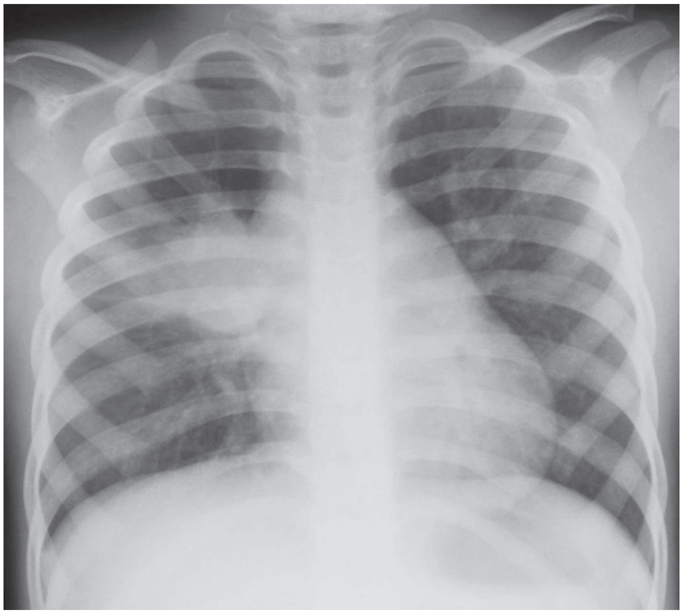

FIGURE 183-4 Chest radiograph showing right hilar lymph node enlargement with infiltration into the surrounding lung tissue in a child with primary tuberculosis. *(Courtesy of Prof. Robert Gie, Department of Paediatrics and Child Health, Stellenbosch University, South Africa; with permission.)*

invariably accompanied by hilar or paratracheal lymphadenopathy due to the spread of bacilli from the lung parenchyma through lymphatic vessels. Enlarged lymph nodes may compress bronchi, causing total obstruction with distal collapse, partial obstruction with large-airway wheezing, or a ball-valve effect with segmental/lobar hyperinflation. Lymph nodes may also rupture into the airway with development of pneumonia, often including areas of necrosis and cavitation, distal to the obstruction. Bronchiectasis **(Chap. 301)** may develop in any segment/lobe damaged by progressive caseating pneumonia. Occult hematogenous dissemination commonly follows primary infection. However, in the absence of a sufficient acquired immune response, which usually contains the infection, disseminated or miliary disease may result **(Fig. 183-5)**. Small granulomatous lesions develop in multiple organs and may cause locally progressive disease or result in tuberculous meningitis; this is a particular concern in very young children and immunocompromised persons (e.g., PLWH).

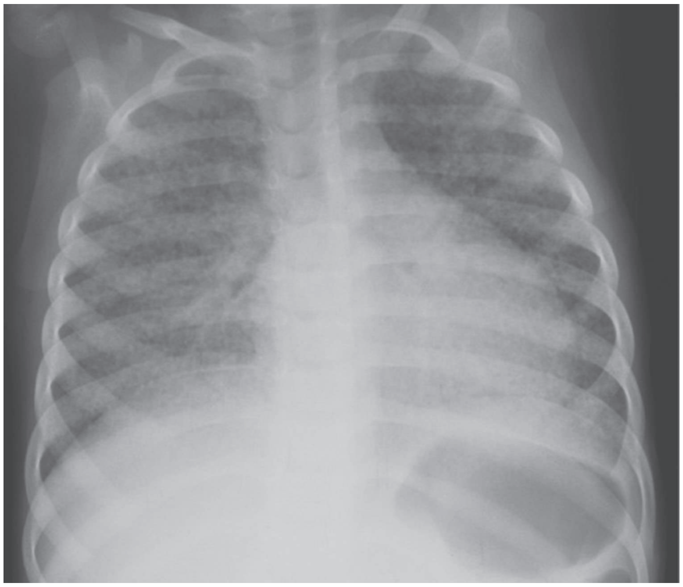

FIGURE 183-5 Chest radiograph showing bilateral miliary (millet-sized) infiltrates in a child. *(Courtesy of Prof. Robert Gie, Department of Paediatrics and Child Health, Stellenbosch University, South Africa; with permission.)*

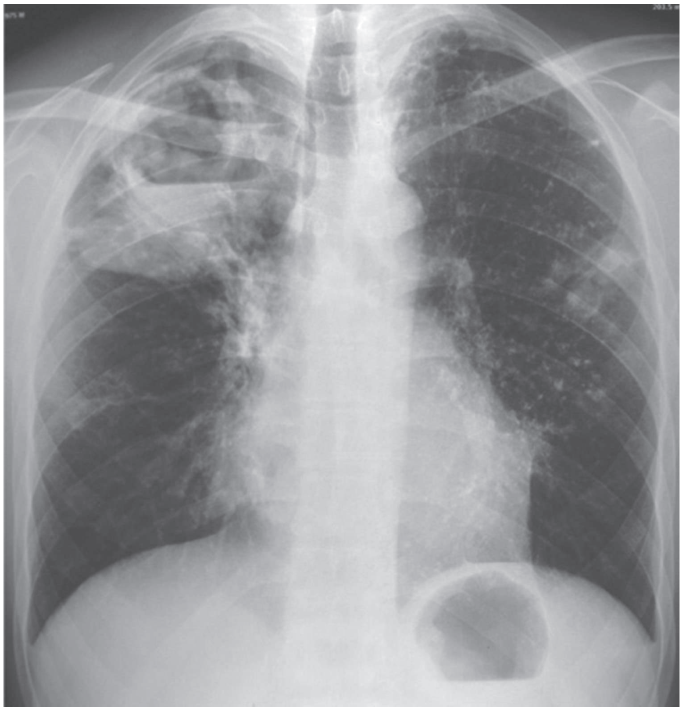

FIGURE 183-6 Chest radiograph showing a right-upper-lobe infiltrate and a cavity with an air-fluid level in a patient with active tuberculosis. *(Courtesy of Dr. Andrea Gori, Infectious Diseases Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, University of Milan, Milan, Italy; with permission.)*

Postprimary (Adult-Type) Disease Also referred to as *reactivation* or *secondary TB*, postprimary TB is probably most accurately termed *adult-type TB* because it may result from endogenous reactivation of distant or recent infection (primary infection or reinfection). It is usually localized to the apical and posterior segments of the upper lobes, where the substantially higher mean oxygen tension (compared with that in the lower zones) favors mycobacterial growth. The superior segments of the lower lobes also are frequently involved. The extent of lung parenchymal involvement varies greatly, from small infiltrates to extensive cavitary disease. With cavity formation, liquefied necrotic contents are ultimately discharged into the airways and may undergo bronchogenic spread, resulting in satellite lesions within the lungs that may in turn undergo cavitation **(Figs. 183-6 and 183-7)**. Massive involvement of pulmonary segments or lobes, with coalescence of lesions, produces caseating pneumonia. While up to one-third of untreated patients reportedly succumb to severe pulmonary TB

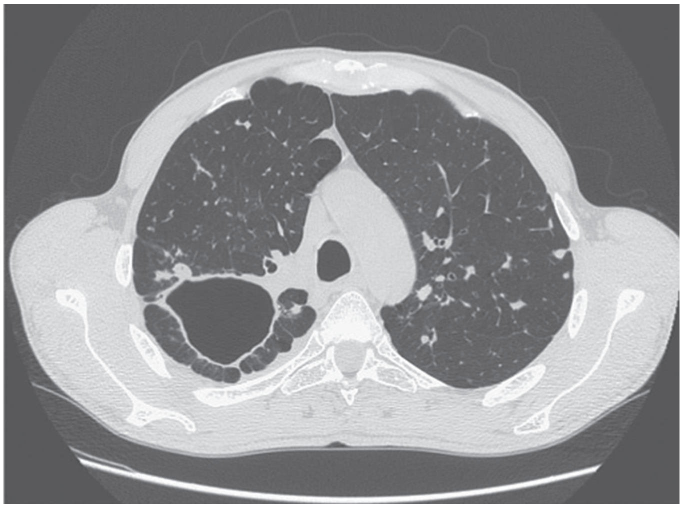

FIGURE 183-7 CT scan showing a large cavity in the right lung of a patient with active tuberculosis. *(Courtesy of Dr. Elisa Busi Rizzi, National Institute for Infectious Diseases, Spallanzani Hospital, Rome, Italy; with permission.)*

within a few months after onset (the classic "galloping consumption" of the past), others may undergo a process of spontaneous remission or proceed along a chronic, progressively debilitating course ("consumption" or *phthisis*). Under these circumstances, some pulmonary lesions become fibrotic and may later calcify, but cavities persist in other parts of the lungs. Individuals with such chronic disease continue to discharge tubercle bacilli into the environment. Most patients respond to treatment, with defervescence, decreasing cough, weight gain, and a general improvement in well-being within several weeks.

Early in the course of disease, symptoms and signs are often nonspecific and insidious, consisting mainly of fever, often diurnal, and night sweats due to defervescence, weight loss, anorexia, general malaise, and weakness. However, in up to 90% of cases, cough eventually develops—often initially nonproductive and limited to the morning and subsequently accompanied by the production of purulent sputum, sometimes with blood streaking. Hemoptysis develops in 20–30% of cases, and massive hemoptysis may ensue as a consequence of the erosion of a blood vessel in the wall of a cavity. Hemoptysis, however, may also result from rupture of a dilated vessel in a cavity (*Rasmussen's aneurysm*) or from aspergilloma formation in an old cavity. Pleuritic chest pain sometimes develops in patients with subpleural parenchymal lesions or pleural disease. Extensive disease may produce dyspnea and, in rare instances, adult respiratory distress syndrome. Physical findings are of limited use in pulmonary TB. Many patients have no abnormalities detectable by chest examination, whereas others have detectable rales in the involved areas during inspiration, especially after coughing. Occasionally, rhonchi due to partial bronchial obstruction and classic amphoric breath sounds in areas with large cavities may be heard. Systemic features include fever (often low-grade and intermittent) in up to 80% of cases and wasting. Absence of fever, however, does not exclude TB. In some recurrent cases and among patients with low Karnofsky score, finger clubbing has been reported. The most common hematologic findings are mild anemia, leukocytosis, and thrombocytosis with a slightly elevated erythrocyte sedimentation rate and/or C-reactive protein level. None of these findings is consistent or sufficiently accurate for diagnostic purposes. Hyponatremia due to the syndrome of inappropriate secretion of antidiuretic hormone also has been reported.

EXTRAPULMONARY TB

In descending order of frequency, the extrapulmonary sites most involved in TB are the lymph nodes, pleura, genitourinary tract, bones and joints, meninges, peritoneum, and pericardium. However, virtually any organ system may be affected. As a result of hematogenous dissemination in PLWH, extrapulmonary TB is seen more commonly today than in the past in settings of high HIV prevalence.

Lymph Node TB (Tuberculous Lymphadenitis) The most common presentation of extrapulmonary TB in both HIV-seronegative individuals and PLWH (35% of cases worldwide and >40% of cases in the United States in recent series), lymph node disease is particularly frequent among PLWH and among children **(Fig. 183-8)**. In the United States, besides children, women (particularly non-Caucasians) seem to be especially susceptible. Once caused mainly by *M. bovis*, tuberculous lymphadenitis today is due largely to *M. tuberculosis*. Lymph node TB presents as painless swelling of the lymph nodes, most commonly at posterior cervical and supraclavicular sites (a condition historically referred to as *scrofula*). Lymph nodes are usually discrete in early disease but develop into a matted nontender mass over time; a fistulous tract draining caseous material may result. Associated pulmonary disease is present in fewer than 50% of cases, and systemic symptoms are uncommon except in HIV-infected patients. The diagnosis is established by fine-needle aspiration biopsy (with a yield of up to 80%) or surgical excision biopsy. Bacteriologic confirmation is achieved in most cases, granulomatous lesions with or without visible AFBs are typically seen, and cultures are positive in 70–80% of cases. Among PLWH, granulomas are less well organized and are frequently absent entirely, but bacterial loads are heavier than in HIV-seronegative patients, with higher yields from microscopy and culture. Differential diagnosis includes a variety of infectious conditions, neoplastic diseases such as lymphomas or metastatic carcinomas, and rare disorders like Kikuchi disease (necrotizing histiocytic lymphadenitis), Kimura disease, and Castleman disease.

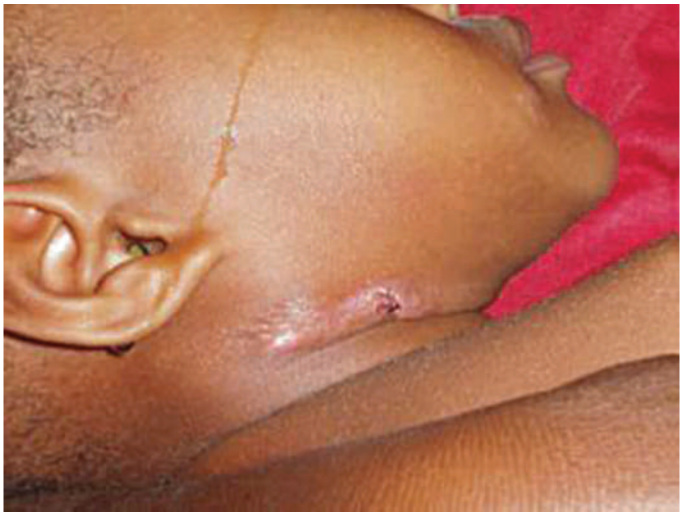

FIGURE 183-8 Tuberculous lymphadenitis affecting the cervical lymph nodes in a 2-year-old child from Malawi. *(Courtesy of Prof. S. Graham, Centre for International Child Health, University of Melbourne, Australia; with permission.)*

Pleural TB Involvement of the pleura accounts for ~20% of extrapulmonary cases in the United States and elsewhere. Isolated pleural effusion usually reflects recent primary infection, and the collection of fluid in the pleural space represents a hypersensitivity response to mycobacterial antigens. Pleural disease may also result from contiguous parenchymal spread, as in many cases pleurisy accompanies postprimary disease. Depending on the extent of reactivity, the effusion may be small, remain unnoticed, and resolve spontaneously or may be sufficiently large to cause symptoms such as fever, pleuritic chest pain, and dyspnea. Physical findings are those of pleural effusion: dullness to percussion and absence of breath sounds. CXR reveals the effusion and, in up to one-third of cases, also shows a parenchymal lesion. Thoracentesis is required to ascertain the nature of the effusion and to differentiate it from manifestations of other etiologies. The fluid is straw-colored and at times hemorrhagic; it is an exudate with a protein concentration >50% of that in serum (usually ~4–6 g/dL), a normal to low glucose concentration, a pH of ~7.3 (occasionally <7.2), and detectable white blood cells (usually 500–6000/μL). Neutrophils may predominate in the early stage, but lymphocyte predominance is typical later. Mesothelial cells are generally rare or absent. AFBs are rarely seen on direct smear, and cultures often may be falsely negative for *M. tuberculosis*; positive cultures are more common among postprimary cases. An elevated pleural concentration of adenosine deaminase is consistent with TB, while TB may be excluded if the value is very low. Lysozyme also is present in the pleural effusion. Measurement of IFN-γ, either directly or through stimulation of sensitized T cells with mycobacterial antigens, can be diagnostically helpful. Needle biopsy of the pleura is often required for diagnosis and is recommended over pleural fluid analysis; it reveals granulomas and/or yields a positive culture in up to 80% of cases. Pleural biopsy can yield a positive result in ~75% of cases when real-time automated nucleic acid amplification is used (the Xpert MTB/RIF assay [Cepheid; Sunnyvale, CA]; see "Nucleic Acid Amplification Technology," below); testing of pleural fluid with this assay is not recommended because of low sensitivity. This form of pleural TB responds rapidly to chemotherapy and may resolve spontaneously. Concurrent glucocorticoid administration may reduce the duration of fever and/or chest pain but is not of proven benefit.

Tuberculous empyema is a less common complication of pulmonary TB. It is usually the result of the rupture of a cavity, with spillage of a large number of organisms into the pleural space. This process may create a bronchopleural fistula with evident air in the pleural space.

CXR shows hydropneumothorax with an air-fluid level. The pleural fluid is purulent and thick and contains large numbers of lymphocytes. Acid-fast smears and mycobacterial cultures are often positive. Surgical drainage is usually required as an adjunct to chemotherapy. Tuberculous empyema may result in severe pleural fibrosis and restrictive lung disease. Removal of the thickened visceral pleura (*decortication*) is occasionally necessary to improve lung function.

TB of the Upper Airways Nearly always a complication of advanced cavitary pulmonary TB, TB of the upper airways may involve the larynx, pharynx, and epiglottis. Symptoms include hoarseness, dysphonia, and dysphagia in addition to chronic productive cough. Findings depend on the site of involvement, and ulcerations may be seen on laryngoscopy. Acid-fast smear of the sputum is often positive, but biopsy may be necessary in some cases to establish the diagnosis. Carcinoma of the larynx may have similar features but is usually painless.

Genitourinary TB Genitourinary TB, which accounts for ~10–15% of all extrapulmonary cases in the United States and elsewhere, may involve any portion of the genitourinary tract. Clinical manifestations are cryptic and protean. Patients may be asymptomatic and their disease discovered only after destructive lesions of the kidneys have developed. Symptoms are often nonspecific and include those of urinary tract infection with frequency, dysuria, nocturia and hematuria, and abdominal or flank pain. Without a high index of suspicion, this form of TB may result in delayed diagnosis with irreversible organ damage. Up to 75% of patients have abnormalities on CXR suggesting previous or concomitant pulmonary disease. Urinalysis gives abnormal results in 90% of cases, revealing pyuria and hematuria. The documentation of culture-negative pyuria in acidic urine should raise the suspicion of TB. IV pyelography, abdominal CT, or MRI (Fig. 183-9) may show deformities and obstructions; calcifications and ureteral strictures are suggestive findings. Culture of three morning urine specimens yields a definitive diagnosis in nearly 90% of cases. Severe ureteral strictures may lead to hydronephrosis, serious renal damage, and, ultimately, renal failure. Genital TB is diagnosed more commonly in female than in male patients. In female patients, it affects the fallopian tubes and the endometrium and may cause infertility, pelvic pain, and menstrual abnormalities. Diagnosis requires biopsy or culture of specimens obtained by dilation and curettage. In male patients, genital TB preferentially affects the epididymis, producing a slightly tender mass that may drain externally through a fistulous tract; orchitis and prostatitis also may develop. In almost half of cases of genitourinary TB, urinary tract disease is also present. Genitourinary TB responds well to chemotherapy.

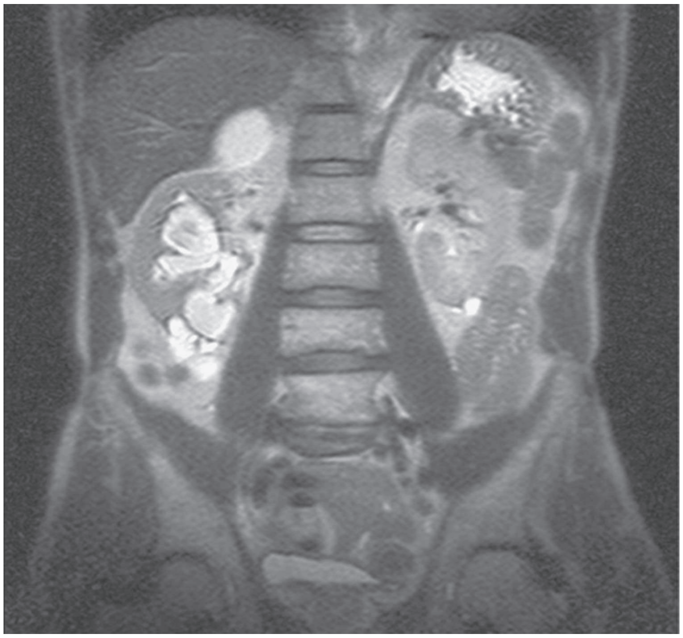

FIGURE 183-9 MRI of culture-confirmed renal tuberculosis. T2-weighted coronary plane: coronal sections showing several renal lesions in both the cortical and the medullary tissues of the right kidney. *(Courtesy of Dr. Alberto Matteelli, Department of Infectious Diseases, University of Brescia, Italy; with permission.)*

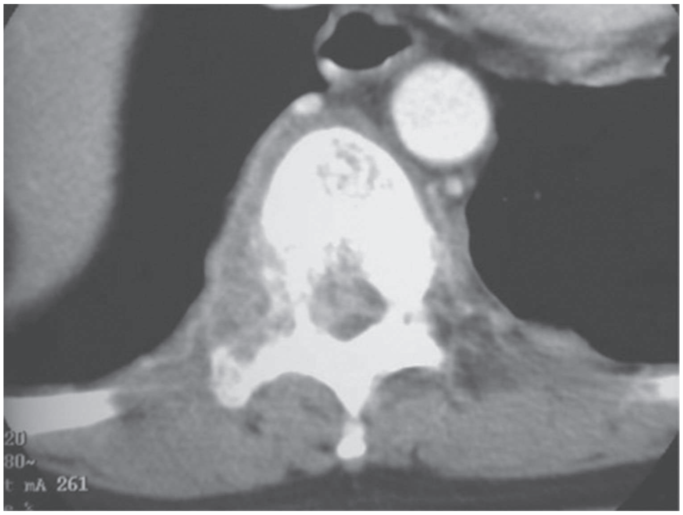

FIGURE 183-10 CT scan demonstrating destruction of the right pedicle of T10 due to Pott's disease. The patient, a 70-year-old Asian woman, presented with back pain and weight loss and had biopsy-proven tuberculosis. *(Courtesy of Charles L. Daley, MD, University of California, San Francisco; with permission.)*

Skeletal TB In the United States, TB of the bones and joints is responsible for ~10% of extrapulmonary cases. In bone and joint disease, pathogenesis is related to reactivation of hematogenous foci or to spread from adjacent paravertebral lymph nodes. Weight-bearing joints (the spine in 40% of cases, the hips in 13%, and the knees in 10%) are most commonly affected. Spinal TB (Pott's disease or tuberculous spondylitis; Fig. 183-10) often involves two or more adjacent vertebral bodies. Whereas the upper thoracic spine is the most common site of spinal TB in children, the lower thoracic and upper lumbar vertebrae are usually affected in adults. From the anterior superior or inferior angle of the vertebral body, the lesion slowly reaches the adjacent body, later affecting the intervertebral disk. With advanced disease, collapse of vertebral bodies results in kyphosis (*gibbus*). A paravertebral "cold" abscess also may form. In the upper spine, this abscess may track to and penetrate the chest wall, presenting as a soft tissue mass; in the lower spine, it may reach the inguinal ligaments or present as a psoas abscess. CT or MRI reveals the characteristic lesion and suggests its etiology. The differential diagnosis includes tumors and other infections. Pyogenic bacterial osteomyelitis involves the disk very early and produces rapid sclerosis. Aspiration of the abscess or bone biopsy confirms the tuberculous etiology, as cultures are usually positive and histologic findings highly typical. A catastrophic complication of Pott's disease is paraplegia, which is usually due to an abscess or a lesion compressing the spinal cord. Paraparesis due to a large abscess is a medical emergency and requires rapid drainage. TB of the hip joints, usually involving the head of the femur, causes pain; TB of the knee produces pain and swelling. If the disease goes unrecognized, the joints may be destroyed. Diagnosis requires examination of the synovial fluid, which is thick in appearance, with a high protein concentration and a variable cell count. Although synovial fluid culture is positive in a high percentage of cases, synovial biopsy and tissue culture may be necessary to establish the diagnosis. Skeletal TB responds to chemotherapy, but severe cases may require surgery.

Tuberculous Meningitis and Tuberculoma TB of the central nervous system (CNS) accounts for ~5% of extrapulmonary cases in the United States. It is seen most often in young children but also develops in adults, especially those infected with HIV. Tuberculous meningitis results from the hematogenous spread of primary or postprimary pulmonary TB or from the rupture of a subependymal tubercle into the subarachnoid space. In more than half of cases, evidence of old pulmonary lesions or a miliary pattern is found on CXR. The disease often presents subtly as headache and slight mental changes after a prodrome of weeks of low-grade fever, malaise, anorexia, and irritability. If not recognized, tuberculous meningitis may evolve acutely with severe headache, confusion, lethargy, altered sensorium, and neck rigidity. Typically, the disease evolves over 1–2 weeks, a course longer than that of typical bacterial meningitis. Because meningeal involvement is pronounced at the base of the brain, paresis of cranial nerves (ocular nerves in particular) is a frequent finding, and the involvement of cerebral arteries may produce focal ischemia. The ultimate evolution is toward coma, with hydrocephalus and intracranial hypertension.

Lumbar puncture is the cornerstone of diagnosis. In general, examination of cerebrospinal fluid (CSF) reveals a high leukocyte count (up to 1000/μL), usually with a predominance of lymphocytes but sometimes with a predominance of neutrophils in the early stage; a protein content of 1–8 g/L (100–800 mg/dL); and a low glucose concentration. However, any of these three parameters can be within the normal range. AFBs are infrequently seen on direct smear of CSF sediment, but repeated lumbar punctures increase the yield. Culture of CSF is diagnostic in up to 80% of cases and remains the gold standard. Real-time automated nucleic acid amplification (the Xpert MTB/RIF assay) has a sensitivity of up to 80% and is the preferred initial diagnostic option. Treatment should be initiated immediately upon a positive Xpert MTB/RIF result. A negative result does not exclude a diagnosis of TB and requires further diagnostic workup. Imaging studies (CT and MRI) may show hydrocephalus and abnormal enhancement of basal cisterns or ependyma. If unrecognized, tuberculous meningitis is uniformly fatal. This disease responds to chemotherapy; however, neurologic sequelae are documented in 25% of treated cases, in most of which the diagnosis has been delayed. Clinical trials have demonstrated that patients given adjunctive glucocorticoids may experience faster resolution of CSF abnormalities and elevated CSF pressure, resulting in lower rates of death or severe disability and relapse. In one study, adjunctive dexamethasone significantly enhanced the chances of survival among persons >14 years but did not reduce the frequency of neurologic sequelae. The dexamethasone schedule was (1) 0.4 mg/kg per day given intravenously with tapering by 0.1 mg/kg per week until the fourth week, when 0.1 mg/kg per day was administered, followed by (2) 4 mg/d given by mouth with tapering by 1 mg per week until the fourth week, when 1 mg/d was administered. The WHO now recommends that adjuvant glucocorticoid therapy with either dexamethasone or prednisolone, tapered over 6–8 weeks, should be used in CNS TB. However, among PLWH, a recent placebo-controlled study demonstrated no benefit with respect to survival or secondary endpoints from a 6- to 8-week tapering course of adjunctive dexamethasone.

Tuberculoma, an uncommon manifestation of TB of the CNS, presents as one or more space-occupying lesions and usually causes seizures and focal signs. CT or MRI reveals contrast-enhanced ring lesions, but biopsy is necessary to establish the diagnosis.

Gastrointestinal TB Gastrointestinal TB is uncommon, making up only 3.5% of extrapulmonary cases in the United States. Various pathogenetic mechanisms are involved: swallowing of sputum with direct seeding, hematogenous spread, or (largely in developing areas) ingestion of milk from cows affected by bovine TB. Although any portion of the gastrointestinal tract may be affected, the terminal ileum and the cecum are the sites most commonly involved. Abdominal pain (at times similar to that associated with appendicitis) and swelling, obstruction, hematochezia, and a palpable mass in the abdomen are common findings at presentation. Fever, weight loss, anorexia, and night sweats also are common. With intestinal wall involvement, ulcerations and fistulae may simulate Crohn disease; the differential diagnosis of this entity is always difficult. Anal fistulae should prompt an evaluation for rectal TB. Because surgery is required in most cases, the diagnosis can be established by histologic examination and culture of specimens obtained intraoperatively.

Tuberculous peritonitis follows either the direct spread of tubercle bacilli from ruptured lymph nodes and intraabdominal organs (e.g., genital TB in women) or hematogenous seeding. Nonspecific abdominal pain, fever, and ascites should raise the suspicion of tuberculous peritonitis. The coexistence of cirrhosis (Chap. 355) in patients with tuberculous peritonitis complicates the diagnosis. In tuberculous peritonitis, paracentesis reveals an exudative fluid with a high protein content and leukocytosis that is usually lymphocytic (although neutrophils occasionally predominate). The yield of direct smear and culture is relatively low; culture of a large volume of ascitic fluid can increase the yield, but peritoneal biopsy (with a specimen best obtained by laparoscopy) is often needed to establish the diagnosis.

Pericardial TB (Tuberculous Pericarditis) Due either to direct extension from adjacent mediastinal or hilar lymph nodes or to hematogenous spread, pericardial TB has often been a disease of the elderly in countries with low TB prevalence. However, it also develops frequently in PLWH. Case fatality rates are as high as 40% in some series. The onset may be subacute, although an acute presentation, with dyspnea, fever, dull retrosternal pain, and a pericardial friction rub, is possible. An effusion eventually develops in many cases; cardiovascular symptoms and signs of cardiac tamponade may ultimately appear (Chap. 281). In the presence of effusion, TB must be suspected if the patient belongs to a high-risk population (HIV-infected, originating in a high-prevalence country); if there is evidence of previous TB in other organs; or if echocardiography, CT, or MRI shows effusion and thickness across the pericardial space. A definitive diagnosis can be obtained by pericardiocentesis under echocardiographic guidance. The pericardial fluid must be submitted for biochemical, cytologic, and microbiologic evaluation. The effusion is exudative in nature, with a high count of lymphocytes and monocytes. Hemorrhagic effusion is common. Direct smear examination is very rarely positive. Culture of pericardial fluid reveals *M. tuberculosis* in up to two-thirds of cases, whereas pericardial biopsy has a higher yield. High levels of adenosine deaminase, lysozyme, and IFN-γ may suggest a tuberculous etiology.

Without treatment, pericardial TB is usually fatal. Even with treatment, complications may develop, including chronic constrictive pericarditis with thickening of the pericardium, fibrosis, and sometimes calcification, which may be visible on a chest radiograph. Systematic reviews and meta-analyses show a trend toward benefit from glucocorticoid treatment with regard to death and constrictive pericarditis. However, the largest and most recent study—the IMPI study—failed to show such a benefit. Of the patients enrolled in this trial, 67% were PLWH, and only a fraction was receiving antiretroviral treatment (ART). A supplemental analysis among HIV-negative patients showed a small mortality benefit, as did another small study among PLWH. The WHO currently recommends that, in patients with tuberculous pericarditis, initial adjuvant glucocorticoid therapy may be used. The 2016 guidelines of the American Thoracic Society (ATS), the CDC, and the Infectious Diseases Society of America (IDSA), on the other hand, suggest that glucocorticoid therapy should not be routinely administered.

Caused by direct extension from the pericardium or by retrograde lymphatic extension from affected mediastinal lymph nodes, tuberculous myocarditis is extremely rare. Usually, it is fatal and is diagnosed postmortem.

Miliary or Disseminated TB Miliary TB is due to hematogenous spread of tubercle bacilli. Although in children it is often the consequence of primary infection, in adults it may be due to either recent infection or reactivation of old disseminated foci. The lesions are usually yellowish granulomas 1–2 mm in diameter that resemble millet seeds (thus the term *miliary*, coined by nineteenth-century pathologists). Clinical manifestations are nonspecific and protean, depending on the predominant site of involvement. Fever, night sweats, anorexia, weakness, and weight loss are presenting symptoms in the majority of cases. At times, patients have a cough and other respiratory symptoms due to pulmonary involvement as well as abdominal symptoms. Physical findings include hepatomegaly, splenomegaly, and lymphadenopathy. Eye examination may reveal choroidal tubercles, which are pathognomonic of miliary TB, in up to 30% of cases. Meningismus occurs in fewer than 10% of cases.

A high index of suspicion is required for the diagnosis of miliary TB. Frequently, CXR (Fig. 183-5) reveals a miliary reticulonodular pattern (more easily seen on underpenetrated film), although no radiographic abnormality may be evident early in the course and among PLWH. Other radiologic findings include large infiltrates, interstitial infiltrates (especially in PLWH), and pleural effusion. Sputum-smear microscopy is negative in most cases. Various hematologic abnormalities may be seen, including anemia with leukopenia, lymphopenia, neutrophilic leukocytosis and leukemoid reactions, and polycythemia. Disseminated intravascular coagulation has been reported. Elevation of alkaline phosphatase levels and other abnormal liver function tests are detected in patients with severe hepatic involvement. TST results may be negative in up to half of cases, but reactivity may be restored during chemotherapy. Bronchoalveolar lavage and transbronchial biopsy are more likely to provide bacteriologic confirmation, and granulomas are evident in liver or bone-marrow biopsy specimens from many patients. If it goes unrecognized, miliary TB is lethal; with proper early treatment, however, it is amenable to cure. Glucocorticoid therapy has not proved beneficial.

A rare presentation seen in the elderly, *cryptic miliary TB* has a chronic course characterized by mild intermittent fever, anemia, and—ultimately—meningeal involvement preceding death. An acute septicemic form, *nonreactive miliary TB*, occurs very rarely and is due to massive hematogenous dissemination of tubercle bacilli. Pancytopenia is common in this form of disease, which is rapidly fatal. At postmortem examination, multiple necrotic but nongranulomatous ("nonreactive") lesions are detected.

Less Common Extrapulmonary Forms TB may cause chorioretinitis, uveitis, panophthalmitis, and painful hypersensitivity-related phlyctenular conjunctivitis. Tuberculous otitis is rare and presents as hearing loss, otorrhea, and tympanic membrane perforation. In the nasopharynx, TB may simulate granulomatosis with polyangiitis. Cutaneous manifestations of TB include primary infection due to direct inoculation, abscesses and chronic ulcers, scrofuloderma, lupus vulgaris (a smoldering disease with cutaneous nodules, plaques, and fissures), miliary lesions, and erythema nodosum. Tuberculous mastitis results from retrograde lymphatic spread, often from the axillary lymph nodes. Adrenal TB is a manifestation of disseminated disease presenting rarely as adrenal insufficiency. Finally, congenital TB results from transplacental spread of tubercle bacilli to the fetus or from ingestion of contaminated amniotic fluid. This rare disease affects the liver, spleen, lymph nodes, and various other organs.

Post-TB Lung Disease and Complications A substantial proportion of TB survivors considered cured experience chronic impairment of lung function, dyspnea, residual cough, bronchiectasis, and pulmonary hypertension. In addition, aspergillomas and chronic pulmonary aspergillosis **(Chap. 223)** have been seen post-TB. Chronic pulmonary aspergillosis may manifest as simple aspergilloma (fungal ball) or chronic cavitary aspergillosis. In the presence of large residual cavities, *Aspergillus fumigatus* may colonize the lesion and produce symptoms such as respiratory impairment, hemoptysis, persistent fatigue, and weight loss, often resulting in the erroneous diagnosis of TB recurrence. The detection of *Aspergillus* precipitins (IgG) in the blood suggests chronic pulmonary aspergillosis, as do radiographic abnormalities such as thickening of the pleura and cavitary walls or the presence of a fungal ball inside the cavity. Treatment is difficult. Recent preliminary studies on the use of triazoles for ≥6 months indicate improvement or stabilization of 60–75% of the radiologic and clinical manifestations. Surgical removal of lesions is risky except in simple aspergilloma.

HIV-Associated TB (See also Chap. 208) TB is one of the most common diseases among PLWH worldwide. Responsible for up to a quarter of all HIV-related mortality (167,000 deaths per year), TB is still one of the main causes of death in this population, especially in Africa **(Fig. 183-11)**.

Classic studies showed that a person with a positive TST who acquires HIV infection has a 3–13% annual risk of developing active TB, with the exact risk depending on the degree of immunosuppression when observation begins. Furthermore, a new TB infection acquired by a PLWH may evolve into active disease in a matter of weeks rather than months or years. TB can appear at any stage of HIV infection, and its presentation varies with the stage. When cell-mediated immunity is only partially compromised, pulmonary TB presents in a typical manner (Figs. 183-6 and 183-7), with upper-lobe infiltrates and cavitation and without significant lymphadenopathy or pleural effusion. In later stages of HIV infection, when the CD4+ T-cell count is <200/μL, a primary TB–like pattern, with diffuse interstitial and subtle infiltrates,

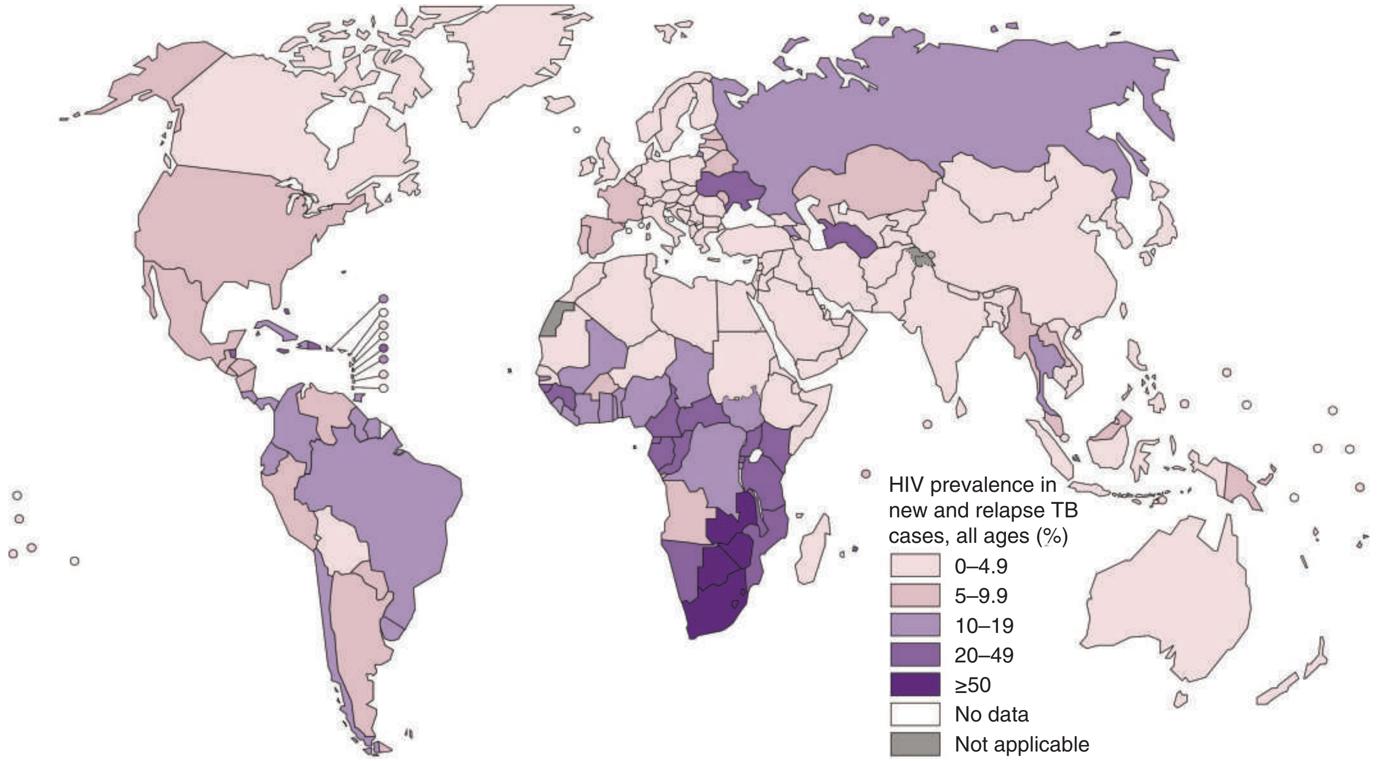

FIGURE 183-11 Estimated HIV prevalence in new and relapse tuberculosis (TB) cases in 2022. *(See disclaimer in Fig. 183-2. Reproduced with permission from Global Tuberculosis Report 2022. Geneva, World Health Organization; 2023.)*

little or no cavitation, pleural effusion, and intrathoracic lymphadenopathy, is more common. However, these forms are becoming less common because of the expanded use of ART. Extrapulmonary TB is common among PLWH. In various series, extrapulmonary TB—alone or in association with pulmonary disease—has been documented in 40–60% of all cases in PLWH. The most common forms are lymphatic, disseminated, pleural, and pericardial. Mycobacteremia and meningitis also are common, particularly in advanced HIV disease. The diagnosis of TB in PLWH can be complicated. The Xpert MTB/RIF assay is the preferred initial diagnostic option for pulmonary TB, ensuring a sensitivity of more than 80% and a specificity of 98%; therapy should be started on the basis of a positive result because treatment delays may be fatal. A negative Xpert MTB/RIF result, however, does not exclude a diagnosis of TB. Culture remains the gold standard. Detection of mycobacterial lipoarabinomannan antigen in urine has shown favorable results in assisting with the diagnosis of TB in PLWH (see "Additional Diagnostic Procedures," below).

The *immune reconstitution inflammatory syndrome* (*IRIS*) or *TB immune reconstitution disease* consists of exacerbations in systemic manifestations (lymphadenopathy, fever) or respiratory signs (worsening of pulmonary infiltrations, pleural effusion) as well as laboratory or radiographic manifestations of TB. This syndrome is rarely seen among non-HIV-infected persons, is associated with the administration of ART, and occurs in ~10% of HIV-infected TB patients. Usually developing 1–3 months after initiation of ART, IRIS is more common among patients with advanced immunosuppression and extrapulmonary TB. "Unmasking IRIS" may develop after the initiation of ART in patients with undiagnosed subclinical TB. The earlier ART is started after TB treatment is initiated and the lower the baseline CD4+ T-cell count, the greater the risk of IRIS. Death due to IRIS is relatively infrequent and occurs mainly among patients who have a high preexisting mortality risk. The presumed pathogenesis of IRIS consists of an immune response that is rapidly improved by HIV suppression and is stimulated by antigens released as bacilli are killed during effective chemotherapy. There is no diagnostic test for IRIS, and its confirmation relies heavily upon case definitions incorporating clinical and laboratory data; a variety of case definitions have been suggested. The first priority in the management of a possible case of IRIS is to ensure that the clinical syndrome does not represent a failure of TB treatment or the development of another infection. Mild paradoxical reactions can be managed with symptom-based treatment and do not worsen outcomes of treatment for TB. However, IRIS can result in serious neurologic complications or death in patients with CNS TB. Therefore, ART should not be initiated during the first 8 weeks of TB treatment in patients with TB meningitis. Glucocorticoids have been used for severe paradoxical reactions; prednisolone given for 4 weeks at a low dosage (1.5 mg/kg per day for 2 weeks and half that dose for the remaining 2 weeks) has reduced the need for hospitalization and therapeutic procedures and has hastened alleviation of symptoms, as reflected by Karnofsky performance scores, quality-of-life assessments, radiographic response, and C-reactive protein levels. The effectiveness of glucocorticoids in alleviating the symptoms of IRIS is probably linked to suppression of proinflammatory cytokine concentrations, as these medications reduce serum concentrations of IL-6, IL-10, IL-12p40, TNF-α, IFN-γ, and IFN-γ-inducible protein 10. Recommendations for the prevention and treatment of TB in PLWH are provided below.

DIAGNOSIS

The key to the early diagnosis of TB is a high index of suspicion. Diagnosis is not difficult in persons belonging to high-risk populations who present with typical symptoms and a classic chest radiograph showing upper-lobe infiltrates with cavities (Fig. 183-6). On the other hand, the diagnosis can easily be missed in an elderly nursing-home resident or a teenager with a focal infiltrate. Often, the diagnosis is first entertained when the chest radiograph of a patient being evaluated for respiratory symptoms is abnormal. If the patient has no complicating medical conditions that cause immunosuppression, the chest radiograph may show typical upper-lobe infiltrates with cavitation (Fig. 183-6). The longer the delay between the onset of symptoms and the diagnosis, the more likely is the finding of cavitary disease. In contrast, immunosuppressed patients, including PLWH, may have "atypical" findings on CXR—e.g., lower-zone infiltrates without cavity formation—or interstitial disease only.

The several approaches to the diagnosis of TB require, above all, a well-organized microbiology laboratory network with an appropriate distribution of tasks at different levels of the health care system. Besides clinical assessment and radiography, screening and referral are the principal tasks at the peripheral and community levels. Diagnosis at a secondary level (e.g., a traditional district hospital in a high-incidence setting) can be accomplished nowadays through real-time automated nucleic acid amplification technology (e.g., the Xpert MTB/RIF assay, which also allows detection of drug resistance) or through traditional AFB microscopy where new tools have not yet been introduced. At a tertiary level, molecular tests, rapid culture, and DST should be applied.

■ NUCLEIC ACID AMPLIFICATION TECHNOLOGY

Several test systems based on amplification of mycobacterial nucleic acid have become available in the past few years and are now the preferred first-line diagnostic tests. These tests are progressively replacing smear microscopy, as they ensure rapid confirmation of all types of TB. One system that permits rapid diagnosis of TB with high specificity and sensitivity (approaching that of liquid culture) is the fully automated, real-time nucleic acid amplification technology known as the Xpert MTB/RIF assay. Xpert MTB/RIF can simultaneously detect TB and rifampin resistance in <2 h and has minimal biosafety and training requirements. Therefore, it can be housed in nonconventional laboratory settings as long as a stable and uninterrupted power supply can be assured. The WHO recommends its use worldwide as the first-line diagnostic test in all adults and children with signs or symptoms of active TB. Given the test's high sensitivity, the WHO also recommends its use as the initial diagnostic test for PLWH in whom TB is suspected. In the diagnosis of pulmonary TB, this test has an overall sensitivity of 85% reaching 98% among AFB-positive cases and ~70% among AFB-negative specimens; its specificity is 98%. When compared with phenotypic drug susceptibility testing for simultaneous detection of rifampin resistance, Xpert MTB/RIF has an overall sensitivity of 96% and a specificity of 98%. The newer Xpert MTB/RIF Ultra assay (Ultra), which uses the same GeneXpert diagnostic platform, has an overall sensitivity of 90% including "trace calls" (i.e., the "noise" produced by detection of DNA from nonviable bacilli) as positive with the greatest increases among smear-negative, culture-positive cases (+17%) and among PLWH (+12%). If "trace calls" are excluded, sensitivity decreases to 86%. Because of this greater sensitivity and the capacity to also detect nonviable bacilli, the new Ultra cartridge has 2% lower specificity than the original test. However, excluding "trace calls," specificity increases to 98%. Among PLWH Ultra sensitivity is 88% and specificity 95%. Sensitivity and specificity for detection of rifampin resistance by Ultra are 94% and 99%, respectively, similar to those by the Xpert MTB/RIF assay.

In the diagnosis of extrapulmonary TB, Xpert MTB/RIF or Ultra should be the initial tests applied to CSF from patients in whom TB meningitis is suspected as well as a replacement test (preferable to conventional microscopy, culture, and histopathology) for selected nonrespiratory specimens—those obtained by gastric lavage, fine-needle aspiration, or pleural or other biopsies. Sensitivity varies according to specimen type with the lowest in pleural fluid (50% with Xpert MTB/RIF and 71% with Ultra) and the highest in synovial fluid (97%) and lymph node biopsy (100% with Ultra). "Trace calls" in specimens from persons with extrapulmonary TB, as well as for PLWH and children, should be considered true positives, given the high risk of severe morbidity and premature death, while among other cases they warrant additional tests to confirm the diagnosis of TB and prevent overtreatment. Among patients with a recent history of TB, "trace calls" may represent DNA from dead bacilli under degradation.

Truenat MTB and MTB Plus are two other newly introduced rapid adopted rapid molecular tests with sensitivities of 73% and 80%, respectively, if compared with bacteriological culture, and specificities of 98% and 96%, respectively. Truenat MTB-Rif Dx detects rifampin resistance with a sensitivity of 84% and a specificity of 97%. These rapid tests, being portable and battery-operated, can be used in peripheral care settings rather than smear microscopy, culture, and phenotypic DST. New high-throughput automated platforms for TB diagnosis and drug-resistant variants are now available (Abbott RealTime MTB and RIF/INH, FluoroType MTBDR and MTB, BD Max MDR-TB, cobas MTB and MTB-RIF/INH). These platforms are suitable for centralized laboratories and have the advantage of processing a large number of samples in a reasonable time. Sensitivity is higher than 91% and specificity ranges from 97 to 100%. Head-to-head studies with Xpert MTB/RIF have shown comparable performance. Another available molecular test for detection of *M. tuberculosis* is based on loop-mediated isothermal amplification (LAMP), a temperature-independent technology that amplifies DNA, is relatively simple to use, and is interpreted through a visual display. It may be used as a replacement for sputum-smear microscopy for the diagnosis of adult pulmonary TB and as a follow-up test to smear microscopy for the further investigation of smear-negative specimens from adults with suspected pulmonary TB. The TB-LAMP assay should not replace rapid molecular tests that detect both TB and rifampin resistance, and its usefulness in PLWH in whom TB is suspected remains unclear.

AFB MICROSCOPY

In some low- and middle-income settings, a presumptive diagnosis is still often based on the finding of AFB on microscopic examination, such as a smear of expectorated sputum or of tissue (e.g., a lymph node biopsy). Although inexpensive, AFB microscopy has relatively low sensitivity (40–60%) in culture-confirmed cases of pulmonary TB and does not distinguish TB from nontuberculous mycobacteria. The traditional method—light microscopy of specimens stained with Ziehl-Neelsen basic fuchsin dyes—is satisfactory, although time consuming and operator dependent. Most modern laboratories processing large numbers of diagnostic specimens use auramine–rhodamine staining and fluorescence microscopy; this approach is more sensitive than the Ziehl-Neelsen method. However, it is expensive because it requires high-cost mercury vapor light sources and a darkroom. Less expensive light-emitting diode (LED) fluorescence microscopes are now recommended by the WHO as the microscopy tool of choice. They are as sensitive as—or more sensitive than—traditional fluorescence microscopes. As a result, conventional light and fluorescence microscopes are being replaced with this more recent technology, especially in developing countries. For patients with signs or symptoms of pulmonary TB, it has been recommended that one or two sputum specimens, preferably collected early in the morning, should be submitted to the laboratory for AFB smear and mycobacterial culture. If tissue is obtained, it is critical that the portion of the specimen intended for culture not be put in preservation fluid such as formaldehyde. The use of AFB microscopy in examining urine or gastric lavage fluid is limited by the low numbers of organisms, which can cause false-negative results, or the presence of commensal mycobacteria, which can cause false-positive results.

MYCOBACTERIAL CULTURE

Definitive diagnosis depends on the isolation and identification of *M. tuberculosis* from a clinical specimen. Commercial liquid-culture systems are recommended by the WHO as the reference standard for culture. The MGIT (mycobacteria growth indicator tube) system uses a fluorescent compound sensitive to the presence of oxygen dissolved in the liquid medium. The appearance of fluorescence, detected by fluorometric technology, indicates active growth of mycobacteria. MGIT cultures usually become positive after a period ranging from 10 days to 2–3 weeks; the tubes are read weekly until the eighth week of incubation before the result is declared to be negative. Specimens may also be inoculated onto egg- or agar-based medium (e.g., Löwenstein-Jensen or Middlebrook 7H10 or 7H11) and incubated at 37°C (under 5% CO_2 for Middlebrook medium). Because most species of mycobacteria, including *M. tuberculosis*, grow slowly, 4–8 weeks may be required before growth is detected on these conventional culture media. Although *M. tuberculosis* may be identified presumptively on the basis of growth time and colony pigmentation and morphology, a variety of biochemical tests have traditionally been used to speciate mycobacterial isolates. In modern, well-equipped laboratories, commercial liquid culture for isolation and species identification by molecular methods or high-pressure liquid chromatography of mycolic acids has replaced isolation on solid media and identification by biochemical tests. A low-cost, rapid immunochromatographic lateral-flow assay based on detection of MTP64 antigen may also be used for species identification of the *M. tuberculosis* complex in culture isolates. These new methods, which are increasingly used in limited-resource settings, have decreased the time required for bacteriologic confirmation of TB to 2–3 weeks.

DRUG SUSCEPTIBILITY TESTING

Universal DST is considered by the WHO as the current standard of care for all TB patients and should consist of DST to at least rifampin for all initial isolates of *M. tuberculosis*, as rifampin resistance is an excellent proxy for MDR-TB diagnosis. Expanded and rapid susceptibility testing for isoniazid and key second-line anti-TB drugs (especially the fluoroquinolones and the injectable drugs) is mandatory when RR-TB is found in order to guide selection of the appropriate treatment regimens. Susceptibility testing may be conducted directly by molecular techniques (with the clinical specimen) or indirectly (with mycobacterial cultures) on solid or liquid medium. Results are obtained rapidly by direct susceptibility testing on liquid medium, with an average reporting time of 3 weeks. With indirect testing on solid medium, results may not be available for ≥8 weeks. Highly reliable genotypic methods for the rapid identification of genetic mutations in gene regions known to be associated with resistance to rifampin (such as those in *rpo*B) and isoniazid (such as those in *kat*G and *inh*A) have been developed and are being widely implemented for screening of patients at increased risk of drug-resistant TB. Apart from the Xpert MTB/RIF, Xpert MTB/RIF Ultra, and Truenat MTB-Rif Dx assays, which effectively detect rifampin resistance, the most widely used tests are molecular line probe assays (LPAs). LPAs are a family of DNA strip-based tests capable of detecting bacterial DNA and identifying drug resistance–associated mutations. After extraction of DNA from *M. tuberculosis* isolates or from clinical specimens, the resistance gene regions are amplified by polymerase chain reaction (PCR) and labeled and probe-hybridized PCR products are detected by colorimetric development. This assay confirms the presence of *M. tuberculosis* as well as mutations in target resistance-gene regions. Given the rapidity and accuracy of commercially available LPAs, the WHO recommends that they are used to detect resistance to isoniazid and rifampin when patients have sputum smear–positive specimens or a cultured isolate of *M. tuberculosis*. These recommendations do not eliminate the need for conventional phenotypic culture-based DST to identify resistance to other drugs and to monitor emergence of additional drug resistance. A similar approach has been developed for second-line anti-TB drugs, such as the fluoroquinolones. Therefore, second-line LPAs (instead of phenotypic culture-based DST) are now recommended by the WHO as the initial test for rapid detection of resistance to the fluoroquinolones or the second-line injectable drugs in isolates from patients with confirmed RR-TB or MDR-TB. As with first-line LPAs, these recommendations do not eliminate the need for conventional phenotypic, culture-based DST to identify resistance to other drugs and to monitor for the emergence of additional resistance. Detection of pyrazinamide resistance is important among persons with MDR/RR-TB. The WHO has recently recommended the use of a LPA with reverse hybridization-based technology in culture isolates rather than phenotypic culture-based DST.

Whole genome sequencing (WGS) of *M. tuberculosis* can provide comprehensive information on mutations conferring resistance, but it has been hampered by the requirement for a culture sample before DNA processing. Amplification and sequencing of relevant genomic targets directly from sputum samples have been successfully tested, and targeted new-generation sequencing (tNGS) is now recommended by the WHO to detect drug resistance after TB diagnosis in order to guide decisions on treatment. This class of diagnostics is particularly useful for patients requiring comprehensive DST with faster results than phenotypic DST. However, its suboptimal sensitivity for some new and repurposed drugs still requires confirmation by phenotypic DST.

RADIOGRAPHIC PROCEDURES

CXR is a rapid imaging technique that has historically been used as a primary tool to detect pulmonary TB. CXR has high sensitivity but poor specificity. Although TB may often present with typical patterns strongly suggesting the disease, some abnormalities seen in TB are also present in several other lung conditions. The initial suspicion of pulmonary TB is often based on abnormal CXR findings in a patient with respiratory symptoms. The presence of lesions suggestive of TB should prompt bacteriologic investigations in all cases, without exception. Although the "classic" picture is that of upper-lobe disease with infiltrates and cavities (Fig. 183-6), virtually any radiographic pattern—from a normal film or a solitary pulmonary nodule to diffuse alveolar infiltrates in a patient with adult respiratory distress syndrome—may be seen. In the era of HIV/AIDS, no radiographic pattern can be considered pathognomonic, but CXR can assist in diagnosing TB or ruling it out before initiation of any preventive treatment. CXR is also helpful as a screening test preceding rapid molecular assays to improve their predictive value. Digital CXR technology, which allows display of images in a digital format on a computer screen instead of on x-ray film, offers several advantages: the procedure time is reduced, the running costs are lower, the imaging is of superior quality, and telemedicine assistance is available, including computer-aided detection (CAD) and interpretation of findings using software programs that analyze digital imaging for abnormalities compatible with TB. However, limited evidence suggests that while sensitivity may be high, specificity is variable. A recent systematic review of CAD studies concluded that the diagnostic accuracy of this technology is still limited and that generalizability to low-prevalence settings is still uncertain.

CT (Fig. 183-7) may be useful in interpreting questionable findings on plain CXR and in diagnosing some forms of extrapulmonary TB (e.g., intrabdominal disease, Pott's disease; Fig. 183-10). MRI is useful in the diagnosis of bone lesions and intracranial TB.

ADDITIONAL DIAGNOSTIC PROCEDURES

Other diagnostic tests may be used when pulmonary TB is suspected. Sputum induction by ultrasonic nebulization of hypertonic saline may be useful for patients who cannot produce a sputum specimen spontaneously. Frequently, patients with radiographic abnormalities that are consistent with other diagnoses (e.g., bronchogenic carcinoma) undergo fiberoptic bronchoscopy with bronchial brushings and endobronchial or transbronchial biopsy of the lesion. Bronchoalveolar lavage of a lung segment containing an abnormality also may be performed. In all cases, it is essential that specimens be submitted for molecular testing with the Xpert MTB/RIF assay, mycobacterial culture, and AFB smear. For the diagnosis of primary pulmonary TB in children, who often do not expectorate sputum, induced sputum specimens and specimens from early-morning gastric lavage may yield positive results in the Xpert MTB/RIF assay or on culture. Recently, a promising artificial intelligence (AI)-powered monitoring of cough counts has been tested for prediction of TB disease and treatment monitoring, and AI-powered classification of cough sounds for TB screening. Once fully developed, these non-invasive tools may be useful complements to the diagnostic armamentarium.

Invasive diagnostic procedures are indicated for patients with suspected extrapulmonary TB. In addition to testing of specimens from involved sites (e.g., CSF for tuberculous meningitis, pleural fluid and biopsy samples for pleural disease), biopsy and culture of bone marrow and liver tissue have good diagnostic yields in disseminated (miliary) TB, particularly in PLWH, who also have a high frequency of positive blood cultures. Xpert MTB/RIF should always be the initial diagnostic test in patients in whom TB meningitis is suspected; any positive results should prompt immediate treatment initiation, while negative results should be followed up by additional testing. In some cases, the results of culture or Xpert MTB/RIF are negative but a clinical diagnosis of TB is supported by consistent epidemiologic evidence (e.g., a history of close contact with an infectious patient) and a compatible clinical and radiographic response to treatment. In the United States and other industrialized countries with low rates of TB, some patients with limited abnormalities on CXR and sputum positive for AFB are infected with nontuberculous mycobacteria, most commonly organisms of the *M. avium* complex or *M. kansasii* **(Chap. 185)**. Factors favoring the diagnosis of nontuberculous mycobacterial disease over TB include an absence of risk factors for TB and the presence of underlying chronic pulmonary disease.

Patients with HIV-associated TB pose several diagnostic problems (see "HIV-Associated TB," above). PLWH with sputum culture–positive, AFB-positive TB may present with a normal chest radiograph. The Xpert MTB/RIF assay is the preferred rapid diagnostic test in this population of patients because of its simplicity and increased sensitivity (~60–70% among AFB-negative, culture-positive cases and 97–98% among AFB-positive cases). With the advent of ART, the occurrence of disseminated *M. avium* complex disease that can be confused with TB has become much less common. A test based on the detection of mycobacterial lipoarabinomannan antigen in urine has emerged as a potentially useful point-of-care test for TB in PLWH with low CD4+ T-cell counts. The lateral-flow urine lipoarabinomannan (LF-LAM) assay can be performed manually and read by eye. The WHO recommends that this assay be used to assist in the diagnosis of TB only (1) in HIV-positive adults and children in inpatient settings who have signs and symptoms of TB, or who have advanced HIV disease or are seriously ill, or, irrespective of signs and symptoms of TB, who have a CD4+ T-cell count of ≤200 cells/μL or (2) in outpatient settings in HIV-positive adults and children who have signs and symptoms of TB, or irrespective of signs and symptoms, who have a CD4+ T cell count of <100 cells/μL.

BIOMARKERS

In view of the limitations of current diagnostics, research on TB biomarkers and multiple marker biosignatures that could be used as a point-of-care test for disease or triage is a high priority and has been crystallized in well-defined target product profiles by the WHO. Recent systematic reviews revealed that promising host biomarkers under study, such as antibodies, cytokines, chemokines, and RNA signatures, by far exceed pathogen biomarkers that can be obtained from urine or blood. However, currently, candidate biomarkers require additional studies to fully assess their performance.

DIAGNOSIS OF *M. TUBERCULOSIS* INFECTION

Two modalities currently exist for identification of individuals with TB infection: the skin tests and IGRA, both of which measure host immunologic response to TB antigens. These tests have limitations, especially in settings or populations with high TB and/or HIV prevalence.

Skin Testing In 1891, Robert Koch discovered that components of *M. tuberculosis* in a concentrated liquid-culture medium, subsequently named "old tuberculin," were capable of eliciting a skin reaction when injected subcutaneously into patients with TB. In 1932, Seibert and Munday purified this product by ammonium sulfate precipitation to produce an active protein fraction known as *tuberculin purified protein derivative* (*PPD*). In 1941, PPD-S, developed by Seibert and Glenn, was chosen as the international standard. Later, the WHO and UNICEF sponsored large-scale production of a master batch of PPD (RT23) and made it available for general use. The greatest limitation of PPD is its lack of mycobacterial species specificity, a property due to the large number of proteins in this product that are highly conserved in the various species. In addition, subjectivity of the skin-reaction interpretation that is dependent on the operator, deterioration of the product, and batch-to-batch variations limit the usefulness of PPD.

The skin test with tuberculin PPD (TST) is most widely used in screening for TB infection. It probably measures the response to antigenic stimulation by T cells that reside in the skin rather than the response of recirculating memory T cells. The test is of limited value in the diagnosis of active TB because of its relatively low sensitivity and specificity and its inability to discriminate between TB infection and active disease. False-negative reactions are common in immunosuppressed patients and in those with overwhelming TB. False-positive reactions may be caused by infections with nontuberculous mycobacteria (Chap. 185) and by BCG vaccination. A repeated TST can produce larger reaction sizes due to either boosting or true conversion. The "boosting phenomenon" is a spurious TST conversion resulting from boosting of reactivity on a subsequent TST 1–5 weeks after the initial test. Distinguishing boosting from true conversion is difficult yet important and can be based on clinical and epidemiologic considerations. For instance, true conversions are likely after BCG vaccination in a previously TST-negative person or in a close contact of an infectious patient.

Recently, new *Mycobacterium tuberculosis* antigen-based skin tests (TBSTs) using ESAT-6 and CFP-10 antigens have been introduced. Combining the simplicity of the TST approach with the specificity of IGRAs, three different TBSTs were assessed by the WHO and considered to have an accuracy similar to that of IGRAs and greater than that of TST, therefore being useful also for PLWH, children, and BCG-vaccinated persons.

IFN-γ Release Assays In-vitro assays that measure T cell release of IFN-γ in response to stimulation with the highly TB-specific RD1-encoded antigens ESAT-6 and CFP-10 were introduced in the early 2000s and are commercially available. The T-SPOT.TB test (Oxford Immunotec; Oxford, UK) is an enzyme-linked immunospot assay, and the QuantiFERON-TB Gold test (Qiagen GmbH; Hilden, Germany) is a whole-blood enzyme-linked immunosorbent assay for measurement of IFN-γ. The QuantiFERON-TB Gold In-Tube (QFT-GIT) assay, which facilitates blood collection and initial incubation, also contains another specific antigen, TB7.7. These tests mainly measure the response of recirculating memory CD4+ T cells—normally part of a reservoir in the spleen, bone marrow, and lymph nodes—to persisting bacilli-producing antigenic signals. However, CD8+ cells can also release IFN-γ in vitro in response to stimulation with TB antigens, and they seem to do so especially in the early phase of infection and in the phase of reactivation. Therefore, a new version of the QFT-GIT assay, called QuantiFERON-TB Gold Plus (QFT-Plus), has been developed and operates through two antigen tubes: TB1, containing long peptides from ESAT-6 and CFP-10 and inducing a CD4+ T cell response, and TB2, which also contains shorter peptides stimulating CD8+ cells.

Potential advantages of IGRAs include logistical convenience, the need for only one patient visit to complete testing, and the avoidance of somewhat subjective measurements (e.g., skin induration). However, IGRAs require that blood be drawn and then delivered to the laboratory in a timely fashion. IGRAs also require that testing be performed by specially trained technicians in a laboratory setting. These requirements pose challenges similar to those faced with the TST, including cold-chain requirements and batch-to-batch variations. Because of higher specificity and greater availability of resources, IGRAs have usually replaced the TST for TB infection diagnosis in low-incidence, high-income settings. However, in high-incidence TB and HIV settings and population groups, evidence regarding the performance and usefulness of IGRAs is still limited, and cost considerations may limit wider use.

A number of national guidelines on the use of IGRAs for TB infection testing have been issued. In the United States, an IGRA is preferred to the TST for most persons over the age of 5 years who are being screened for TB infection. However, for individuals at high risk of progression to active TB (e.g., PLWH), either test—or, to optimize sensitivity, both tests—may be used. Because of the paucity of data on the use of IGRAs in children, the TST is preferred for TB infection testing of children aged <5 years. In Canada and some European countries, a two-step approach for those with positive TSTs—i.e., an initial TST followed by an IGRA—is often recommended. However, a TST may boost an IGRA response if the interval between the two tests exceeds 3 days. Because the IGRA uses elaboration of IFN-γ for its readout, patients with anti-IFN-γ autoantibodies may have an indeterminate test result.

In conclusion, both the TST and IGRA, although useful as diagnostic aids, are imperfect tests for TB infection: while they can identify TB infected persons, they have low predictive value in identifying individuals with the highest risk of progression toward disease, cannot differentiate between active TB and TB infection, cannot distinguish new infections from reinfections, and display reduced sensitivity in immunocompromised patients.

TABLE 183-2 Recommended Dosage[a] for Initial Treatment of Tuberculosis in Adults and Children

	DAILY DOSE	
DRUG	**ADULT**	**PEDIATRIC**
Isoniazid	5 mg/kg, max 300 mg	10 (7–15) mg/kg, max 300 mg
Rifampin	10 mg/kg, max 600 mg	15 (10–20) mg/kg, max 600 mg
Pyrazinamide	25 mg/kg, max 2 g	35 (30–40) mg/kg
Ethambutol	15 mg/kg	20 (15–25) mg/kg

[a] The duration of treatment with individual drugs varies by regimen, as detailed in Table 183-3.

Source: Based on recommendations of the American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention and the World Health Organization.

TREATMENT

Tuberculosis

The two main aims of TB treatment are (1) to prevent morbidity and death by curing TB while preventing recurrences and emergence of drug resistance, and (2) to interrupt transmission by rendering patients noninfectious. Four major drugs are considered first-line agents for the treatment of TB: isoniazid, rifampin, pyrazinamide, and ethambutol. Table 183-2 presents currently recommended dosages in adults and children. An additional rifamycin, rifapentine, also is available and is recommended in some treatment and prevention regimens. For a detailed discussion of the drugs used for the treatment of TB, see Chap. 186.

Because of a lower degree of effectiveness and tolerability, several classes of second-line drugs are generally used mainly for the treatment of patients with drug-resistant TB. These agents have previously been classified in various manners to facilitate a standardized approach to their use. In the latest WHO guidance on the treatment of MDR-TB, they are now grouped in three ranked categories for the purpose of designing more individualized regimens of 18–20 months' duration (Table 183-3). Group A drugs include three

TABLE 183-3 Groups of Drugs Recommended for Use in Longer MDR-TB Regimens and Approach to the Design of a Longer Regimen for Adults and Children

GROUP	DRUG
Group A: All three drugs should be included to ensure that at least four likely effective agents (including one from group B below) are started and at least three are included for the rest of treatment if bedaquiline is stopped	Levofloxacin *or* moxifloxacin Bedaquiline Linezolid
Group B: At least one of these drugs should always be included	Clofazimine Cycloserine *or* terizidone
Group C: Drugs to be used to complete the regimen and when drugs from groups A and B cannot be used[a]	Ethambutol Delamanid Pyrazinamide Imipenem-cilastatin *or* meropenem Amikacin (*or* streptomycin if amikacin is not available) Ethionamide *or* prothionamide[b] *p*-Aminosalicylic acid[b]

[a] Kanamycin and capreomycin are not to be included in the longer regimen.

[b] To be included only if bedaquiline, linezolid, clofazimine, or delamanid are not used, or if better options are not possible.

Source: Adapted from the World Health Organization, 2022.

classes of oral agents: the fluoroquinolones levofloxacin and moxifloxacin; the oxazolidinone linezolid; and the diarylquinoline bedaquiline. Group B drugs include two other oral agents: clofazimine and cycloserine (or its analogue terizidone). Group C drugs include the nitroimidazole delamanid; imipenem-cilastatin or meropenem; the injectable aminoglycosides amikacin and streptomycin (the latter formerly a first-line agent, now rarely used for drug-resistant TB because of their toxicity and high resistance levels worldwide); ethionamide or prothionamide; and P-aminosalicylic acid (PAS).

In addition, the first-line anti-TB drugs ethambutol and pyrazinamide as well as high-dose isoniazid are used for MDR-TB treatment. Information about drugs used in the treatment of drug-resistant TB (including dosages) can be found in the following WHO Handbook: *https://iris.who.int/bitstream/handle/10665/365308/9789240063129-eng.pdf?sequence=1*. This classification scheme excludes the second-line injectable aminoglycoside kanamycin and the polypeptide capreomycin. Amithiozone, which has been associated with severe and at times fatal skin reactions—including Stevens-Johnson syndrome—among PLWH, is no longer recommended.

REGIMENS

Standard regimens are traditionally divided into an intensive (bactericidal) phase and a continuation (sterilizing) phase. During the intensive phase, the majority of tubercle bacilli are killed, symptoms resolve, and usually the patient becomes noninfectious. The continuation phase is required to eliminate persisting mycobacteria and prevent relapse.

Six-month regimen The treatment regimen of choice for virtually all forms of drug-susceptible TB in adults consists of a 2-month initial (intensive) phase of isoniazid, rifampin, pyrazinamide, and ethambutol followed by a 4-month continuation phase of isoniazid and rifampin (2HRZE/4HR) **(Table 183-4)**. This regimen can cure TB in >90% of patients. In children, most forms of TB in the absence of HIV infection or suspected isoniazid resistance can be safely treated without ethambutol in the intensive phase. Treatment should be given daily throughout the course, as systematic reviews have demonstrated that the use of intermittent (twice-weekly or thrice-weekly) regimens are associated with increased risk of treatment failure, relapse, and acquisition of drug resistance. The latest guidelines by the ATS, the CDC, and the IDSA, while recommending daily administration of drugs, include a provision for use of intermittent thrice-weekly supervised regimens mainly among patients who are not infected with HIV, do not have cavitary disease, and are at low risk of relapse.

Patients with cavitary pulmonary TB and delayed sputum-culture conversion (i.e., those who remain culture-positive at 2 months) should be retested immediately for drug-resistant TB, and a change of regimen should be considered. A full course of 6 months with four-drug therapy should be performed not including interruptions of >4 weeks.

Four-month regimens In children and adolescents between 3 months and 16 years of age with nonsevere TB, the WHO also recommends the use of a 4-month regimen consisting of a 2-month initial (intensive) phase of isoniazid, rifampin, pyrazinamide, and

TABLE 183-4 Recommended Antituberculosis Treatment Regimens

	INITIAL PHASE		CONTINUATION PHASE	
INDICATION	DURATION, MONTHS	DRUGS	DURATION, MONTHS	DRUGS
New drug-susceptible pulmonary or extrapulmonary[a] TB cases	2	HRZE[b]	4	HR[b,c]
New drug-susceptible pulmonary TB (12 years or older)	2	HPMZ[b,d]	2	HPM[b,d]
New non-severe[e] TB in children and adolescents aged 3 months to 16 years	2	HRZ(E)[b]	2	HR[b]
Pregnancy	2	HRE[f]	7	HR
Intolerance to Z	2	HRE	7	HR
Relapses, treatment default, failures	Tailored according to rapid drug susceptibility testing			
Resistance (or intolerance) to H	Throughout (6)	RZELfx		
MDR/RR-TB (see text for further details)	Throughout (6) BPaLM[g] for patients aged ≥14 years without previous exposure to B, L, and Pa. Throughout (6) BDLLfxCfz for patients with no previous exposure to B, D, and L, including children, adolescents, and pregnant and breastfeeding women. Either Lfx and Cfz may be omitted depending on fluoroquinolone drug susceptibility testing. This regimen may be used in place of 9-month or longer regimens described below. Throughout (9) BLMZ or BLLfxCfzZ or BDLLfxZ for patients with no previous exposure to B, D, and L, and in whom resistance to fluoroquinolones has been excluded. These regimens may be used in place of currently recommended longer (18-month) regimens described below. Alternatively: all-oral, B-containing, 9-month regimen: 4 months of B (used for up to 6 months), Lfx or M, Eto or Pto, E, Hh, Z, Cfz followed by 5 months of Lfx or M, E, Z, Cfz for adults and children with no previous exposure to second-line treatment including B, in whom resistance to fluoroquinolones has been excluded, and in the absence of extensive pulmonary TB or severe forms of extrapulmonary TB. In more complex forms of MDR/RR-TB (e.g., XDR-TB) longer (≥18 months) individualized regimens need to be formulated as per Table 183-3.			

[a] Except for TB of central nervous system, bone, or joint, for which longer therapy should be used.

[b] All drugs should be given daily.

[c] The American Thoracic Society, the Centers for Disease Control and Prevention, and the Infectious Diseases Society of America suggest that a 2-month continuation phase could be used in HIV-seronegative patients with sputum smear–negative and culture-negative TB.

[d] Rifapentine is given at the daily dose of 1200 mg.

[e] It is considered non-severe: TB of peripheral lymph nodes, intrathoracic lymph node without air obstruction, uncomplicated pleural effusion, or paucibacillary, non-cavitary disease confined to one lobe of the lungs and without a miliary pattern.

[f] The 6-month regimen with pyrazinamide can probably be used safely during pregnancy and is recommended by the WHO and the International Union Against Tuberculosis and Lung Disease. If pyrazinamide is not included in the initial treatment regimen, the minimal duration of therapy is 9 months.

[g] Linezolid is given at the dosage of 600 mg daily. This regimen may be used without moxifloxacin in case of documented resistance to fluoroquinolones.

Abbreviations: B, bedaquiline; Cfz, clofazimine; D, delamanid; E, ethambutol; Eto, ethionamide; H, isoniazid; Hh, high-dose isoniazid (900 mg); L, linezolid; Lfx, levofloxacin; M, moxifloxacin; MDR/RR-TB, multidrug-resistant and rifampin-resistant tuberculosis; Pa, pretomanid; Pto, prothionamide; R, rifampin; WHO, World Health Organization; XDR-TB, extensively drug-resistant tuberculosis; Z, pyrazinamide.

ethambutol followed by a 2-month continuation phase of isoniazid and rifampin.

The recent Tuberculosis Trials Consortium Study 31/AIDS Clinical Trials Group A5349 (Study 31/A5349) showed that a 4-month daily regimen that included isoniazid, pyrazinamide (for the first 2 months only), rifapentine at a daily dose of 1200 mg, and moxifloxacin at a daily dose of 400 mg was noninferior to the standard 6-month regimen and had a similar adverse-event profile. This option (2HPMZ/2HPM) is now recommended by the WHO and CDC as an alternative for patients aged 12 years or older, including PLWH, provided that rigorous antibacterial stewardship is ensured especially to prevent fluoroquinolone resistance. This regimen has not been evaluated, and therefore should not be used, in patients with body weight <40 kg; children aged <12 years; patients with severe extrapulmonary or disseminated forms of TB; PLWH with a CD4+ count of <100/μL, pregnant, breastfeeding or postpartum women; or those with history of prolonged QT syndrome or receiving medications with known clinically relevant drug–drug interactions or infected with an isolate known or suspected to be resistant to the regimen drugs. Alternative regimens for patients who exhibit drug intolerance or adverse reactions are listed in Table 183-4. However, severe side effects prompting discontinuation of any of the first-line drugs and use of these alternative regimens are uncommon. To prevent isoniazid-related neuropathy, pyridoxine (10–25 mg/d) should be added to the regimen given to persons at high risk of vitamin B_6 deficiency (e.g., those with alcohol use disorder; malnourished persons; pregnant and lactating women; and patients with conditions such as chronic renal failure, diabetes, and HIV infection, which are also associated with neuropathy). Finally, to facilitate absorption of rifampin, the drug should be taken on an empty stomach and without meals.

PATIENT CARE AND SUPPORT

Poor adherence to treatment is one of the most important impediments to cure. Moreover, the tubercle bacilli harbored by patients who do not fully adhere to the prescribed regimen are likely to become resistant to the drugs to which they are irregularly exposed. Both patient- and provider-related factors may affect adherence. Patient-related factors include a lack of belief that the illness is worth the cost of adherence; the existence of concomitant medical conditions (notably alcohol or substance abuse); lack of social support; fear of the stigma and discrimination associated with TB; and poverty, with attendant joblessness and homelessness. Provider-related factors that may prevent adherence include lack of support, education, and encouragement of patients and inconvenient clinical services.

A variety of interventions to increase the chances of completion of the months-long treatment course are available. First, a package of social support interventions that are complementary and not mutually exclusive, consisting of educational, psychological, and material goods and services, may enable patients with TB to address hurdles to treatment adherence. Health education and counseling on the disease's seriousness and solutions and on the importance of treatment adherence until cure should be provided to all patients at the start of and throughout the course of TB therapy. Psychological support (i.e., counseling sessions or peer-group support) can be particularly relevant in the context of the stigma and discrimination often affecting persons with TB and their families. Material support (e.g., food or financial support in forms such as meals, food baskets, food supplements, food vouchers, transport subsidies, living allowances, housing incentives, or financial bonuses) reduces indirect costs incurred by patients or their attendants in accessing health services and mitigates the consequences of income loss related to the disease. The recent RATIONS clinical trial in India showed that nutritional support provided to a cohort of undernourished TB patients resulted in rapid weight gain and a substantially decreased mortality. In the same trial, nutritional support of household contacts was associated with a 39–48% reduction in TB incidence in the household during 2 years of follow-up.

Second, it is paramount that health services be arranged to meet the needs and reasonable expectations of patients. Components of optimal health services include a suitable geographic location, a schedule responsive to patients' needs, functional channels of communication between patients and their health care providers (e.g., a telephone short-messaging system, audio/video call capability, home or workplace visits), and a staff willing and competent to care for patients with TB, to address their concerns, and to base the care they provide on sound ethical standards.

Third, it is crucial to offer the patient a suitable option for treatment administration that minimizes the chance of nonadherence. Such options traditionally include unsupervised, self-administered therapy; in-person directly observed therapy (DOT); and nondaily DOT (e.g., supervision not for every dose but weekly or a few times per week) at a location mutually agreed on by patient and health care provider, with supervisory responsibility delegated to a qualified person. Direct supervision of adherence is crucial in view of the lack of tools to accurately predict adherence to self-administered treatment and of the public health importance of TB. The WHO, along with the ATS, the CDC, and the IDSA, states that ideally all patients should have their therapy directly supervised, especially during the initial phase, with proper social support based on a patient-centered approach as described above. In several countries, personnel to supervise therapy are usually available through TB control programs of local public health departments, often involving members of the community who are accepted by the patient and who have been properly trained and educated by health workers to undertake the supervisory role. Direct supervision with social support has been shown to significantly increase the proportion of patients completing treatment in all settings and to lessen the chances of treatment failure, relapse, and default. In general, community- or home-based DOT is recommended over health facility–based DOT or unsupervised treatment; DOT administered by trained lay providers or health care workers is recommended over DOT administered by family members. Recently, comparison of video-observed therapy (VOT) with in-person DOT has shown similar outcomes. In a multicenter, analyst-blinded, randomized, controlled superiority trial of VOT through daily remote observation using a smartphone app versus DOT done 3–5 times weekly at home, community, or clinic settings, VOT was superior to DOT in ensuring scheduled observation of drug intake. Therefore, VOT can replace DOT when Internet access is good and video communication technology (e.g., smartphones, tablets, computers) is available. The system can be appropriately organized and operated by health care providers and patients. Other digital health tools can facilitate the monitoring of adherence, including digital medication monitors; these monitors can register when the pillbox is opened, with options to emit audio signals or a short message to remind patients to take medicines. These tools are customized to the needs and preferences of the individual patient and the provider.

In addition to the above measures promoting adherence, provision of fixed-dose combination products that reduce the number of tablets the patient needs to swallow is recommended by the WHO over separate drug formulations. Various fixed-dose combination products are available (e.g., isoniazid/rifampin, isoniazid/rifampin/pyrazinamide, and isoniazid/rifampin/pyrazinamide/ethambutol). Fixed-dose combinations increase patient satisfaction and minimize the likelihood of prescription error or of development of drug resistance resulting from monotherapy if a drug is out of stock or the patient prefers one drug over others. In addition, these combinations facilitate programmatic management of procurement and supply. In the past, the bioavailability of rifampin was found to be substandard in some formulations of fixed-dose combinations. Medical regulatory authorities should ensure that combination products are of good quality; however, top standards for drug quality assurance are not always operative, especially in limited-resource countries. Prescribers should be aware of this potential problem.

MONITORING TREATMENT RESPONSE AND DRUG TOXICITY

Bacteriologic evaluation through commercial liquid-culture systems (or—when liquid-culture capacity is not yet available—through smear microscopy) is essential in monitoring the response to TB treatment. In addition, the patient's weight should be monitored regularly and the drug dosage adjusted with any significant weight change. Patients with pulmonary disease should have their sputum examined monthly until cultures become negative to allow early detection of treatment failure. With the recommended 6-month standard first-line regimen, >80% of drug-susceptible TB patients will have negative sputum cultures at the end of the second month of treatment. By the end of the third month, the sputum of virtually all patients should be culture negative. In some patients, especially those with extensive cavitary disease and large numbers of organisms, AFB smear conversion may lag behind culture conversion as a result of the expectoration and microscopic visualization of dead bacilli. Therefore, as capacity is built, smear microscopy should be progressively abandoned as a monitoring tool in favor of liquid culture. As noted above, patients with cavitary disease in whom sputum culture conversion does not occur by 2 months require immediate testing or retesting for drug resistance. When a patient's sputum cultures or smears remain positive at ≥3 months despite good adherence, treatment failure caused by drug resistance is likely. Nontuberculous mycobacteria may confound and confuse AFB microscopy. The pattern of drug resistance should guide the choice of the best treatment option (see below). A sputum specimen should be collected at the end of treatment to document cure. In settings where mycobacterial cultures are not yet available, monitoring by AFB smear examination should be undertaken at 2, 5, and 6 months. Bacteriologic monitoring of patients with extrapulmonary TB is more difficult and often is not feasible. In these cases, the response to treatment must be assessed clinically with the help of medical imaging.

Monitoring of the response to chemotherapy by nucleic acid amplification technology, such as the Xpert MTB/RIF assay, is not suitable because these tests can produce positive results due to nonviable bacilli. Likewise, serial chest radiographs are not recommended because radiographic changes may lag behind bacteriologic response and are not highly sensitive. After the completion of treatment, neither sputum examination nor CXR is recommended for routine follow-up purposes. However, a chest radiograph obtained at the end of treatment may be useful for comparative purposes should the patient develop symptoms of recurrent TB months or years later. Patients should be instructed to report promptly for medical assessment if they develop any such symptoms.

During treatment, patients should be monitored for drug toxicity. The most common adverse reaction of significance among those treated for drug-susceptible TB is hepatitis. Patients should be carefully educated about the signs and symptoms of drug-induced hepatitis (e.g., dark urine, loss of appetite, nausea) and should be instructed to discontinue treatment promptly and see their health care provider if these manifestations occur. Although biochemical monitoring is not routinely recommended, all adult patients should undergo baseline assessment of liver function (e.g., measurement of serum levels of hepatic aminotransferases and bilirubin). Older patients, those with concomitant diseases, those with a history of hepatic disease (especially hepatitis C), and those using alcohol daily should be monitored especially closely (i.e., monthly), with repeated measurements of aminotransferases, during the initial phase of treatment. Up to 20% of patients have small increases (up to three times the upper limit of normal) in serum levels of aspartate aminotransferase that are not accompanied by symptoms and are of no consequence. Suspension of treatment should be considered for patients with symptomatic hepatitis, especially when accompanied by at least a three-fold increase in serum levels of AST and/or ALT, and for patients without symptoms of hepatic injury who have marked (at least fivefold) elevations in serum levels of AST and/or ALT. Drugs should be reintroduced one at a time after liver functions have returned to normal. Hypersensitivity reactions usually require the discontinuation of all drugs and rechallenge to determine which agent is the culprit. Because of the variety of regimens available, it usually is not necessary—although it is possible—to desensitize patients. Hyperuricemia and arthralgia caused by pyrazinamide can usually be managed by the administration of acetylsalicylic acid; however, pyrazinamide treatment should be stopped if the patient develops gouty arthritis. Individuals who develop autoimmune thrombocytopenia secondary to rifampin therapy should not receive the drug thereafter. Similarly, the occurrence of optic neuritis with ethambutol is an indication for permanent discontinuation of this drug. Other common manifestations of drug intolerance, such as pruritus and gastrointestinal upset, can generally be managed without the interruption of therapy. Treatment with second-line agents for drug-resistant TB is associated with a variety of adverse drug reactions that are more frequent and severe than in patients receiving first-line TB regimens (see below). The likelihood of drug–drug interactions also is higher when second-line regimens are used.

TREATMENT FAILURE AND RELAPSE

Treatment failure should be suspected when a patient's cultures (or sputum smears, when cultures are not available) remain positive after 3 months of treatment. In the management of such patients, it is imperative that the current isolate be urgently retested (or tested for the first time if, for some reason, rapid molecular susceptibility testing was not performed at the start of treatment) for susceptibility to first-line agents. If resistance to rifampin is detected, testing should be done to second-line agents as well. The treatment approach should start with molecular testing for—at the least—resistance to rifampin and isoniazid. Because results are expected to become available within a few days, changes in the regimen can be postponed until that time. However, if the patient's clinical condition is deteriorating rapidly, an earlier change in regimen may be indicated. A cardinal rule in the latter situation is always to add more than one drug, preferably two or three, at a time to a failing regimen; in practice, starting an empirical regimen for MDR-TB (see "Drug-Resistant TB," below) is warranted. The patient may continue to take isoniazid and rifampin along with these new agents pending the results of susceptibility tests.

Patients who experience a recurrence after apparently successful treatment (i.e., a relapse) are less likely to harbor drug-resistant strains than are patients in whom treatment has failed. Acquired resistance is uncommon among strains from patients in whom relapse follows the proper completion of a standard 6-month regimen. The treatment decision depends on a general assessment of the risk of drug resistance, the severity of the case, and the results of rapid susceptibility testing. Patients whose treatment has been interrupted and who have a high likelihood of MDR-TB should receive an MDR-TB regimen that includes second-line agents as soon as possible and according to DST results (Table 183-4).

DRUG-RESISTANT TB

Strains of *M. tuberculosis* resistant to individual drugs arise by spontaneous point mutations in the mycobacterial genome that occur at low but predictable rates (10^{-7}–10^{-10} for the key drugs). Resistance to rifampin is associated with mutations in the *rpoB* gene in 95% of cases, that to isoniazid with mutations mainly in the *katG* gene (50–95% of cases) and the *inhA* gene promoter region (up to 45%), that to pyrazinamide in the *pncA* gene (up to 98%), that to ethambutol in the *embB* gene (50–65%), that to the fluoroquinolones in the *gyrA*–*gyrB* genes (75–95%), and that to the aminoglycosides mainly in the *rrs* gene (up to 80%); the C-12T mutation is the most common mutation in the *eis* promoter region associated with aminoglycoside resistance, especially in Eastern European countries. Because there is no cross-resistance among the commonly used classes of drugs, the probability that a strain will be resistant to two drug classes is the product of the probabilities of resistance to each drug class and thus is low. The development of drug-resistant TB almost invariably follows monotherapy—i.e., the failure of the health care provider to prescribe at least two drugs to which tubercle bacilli are susceptible; of the patient to absorb or take properly prescribed therapy; or of the bioavailability of poor-quality drugs or preparations (e.g., due to crushing of tablets). Drug-resistant TB may be either primary or acquired. In primary drug resistance, the patient is infected from the start by a drug-resistant strain. Acquired resistance develops in the infecting strain during treatment. In North America, Western Europe, most of Latin America, and the Persian Gulf states, rates of primary resistance are generally low and isoniazid resistance is most common. In the United States, although rates of primary isoniazid resistance have been stable at ~7–8%, the rate of primary MDR-TB has declined from 2.5% in 1993 to <1% since 2000. MDR-TB is a serious problem in some regions, especially in the countries of the former Soviet Union and some countries of Asia **(Fig. 183-12)**. Even more serious is the occurrence of MDR strains that are also resistant to additional second-line agents used in treatment, such as the fluoroquinolones. Creation of drug-resistant TB can be prevented by adherence to the principles of sound treatment: inclusion of at least two quality-assured, bactericidal drugs to which the organism is susceptible; use of effective combination regimens; supervision of treatment with patient support; and verification that patients complete the prescribed course. The use of fixed-dose combination products may prevent selective drug intake and therefore possibly protect against the creation of drug resistance. Transmission of drug-resistant strains can be prevented by the implementation of respiratory infection-control measures (see below) and by early detection of persons with active TB followed by immediate initiation of treatment with an effective regimen.

Isoniazid-Resistant TB For the treatment of patients with isoniazid-resistant but rifampin-susceptible disease, a combination of rifampin, ethambutol, pyrazinamide, and levofloxacin for 6 months is recommended. This fluoroquinolone-containing regimen should not be used until rifampin resistance has been excluded by a reliable diagnostic test to avoid inadvertent treatment of MDR-TB with an inadequate regimen. Susceptibility should also be tested for the fluoroquinolones and pyrazinamide. If the fluoroquinolone is contraindicated because of intolerance or resistance, the patient can be given a 6-month regimen of rifampin, ethambutol, and pyrazinamide. Isoniazid probably does not contribute to a successful outcome in these regimens but may be retained (also to facilitate treatment with the four-drug fixed-dose formulation). Other drugs, such as the injectable aminoglycosides, are unlikely to play a role in the treatment of most isoniazid-resistant TB cases. However, they may be considered only in the presence of additional resistance or of drug intolerance.

RR-, MDR-TB MDR-TB, in which bacilli are resistant to (at least) isoniazid and rifampin, is more difficult to manage than is disease caused by drug-susceptible organisms because these two bactericidal drugs are potent first-line agents and because associated resistance to other first-line drugs as well (e.g., ethambutol) is not uncommon. Treatment for RR-TB and MDR-TB has traditionally been a topic of much debate, given its complexity, long duration, toxicity, and limited efficacy; the cost of most second-line drugs; and the lack of randomized controlled clinical trials to support combinations. Recent developments include the accrual of individual datasets for patients treated worldwide; the release of findings from randomized, controlled, phase 3 clinical trials (the STREAM stage 1 trial comparing a 9-month, shorter MDR-TB regimen with the previous optimized WHO background regimen, and Otsuka's phase 3 trial 213 comparing the addition of the new drug delamanid to the previous optimized WHO background regimen with the addition of placebo); results of the Nix-TB and ZeNix trials enrolling highly drug-resistant cases on a regimen composed of three oral drugs (bedaquiline, pretomanid, and linezolid [BPaL]) at different dosages; the TB-PRACTECAL trial testing BPaL-based regimens against multiple comparators; the NExT trial testing several drug combinations for 6–9 months; the assessment of programmatic data from South Africa on

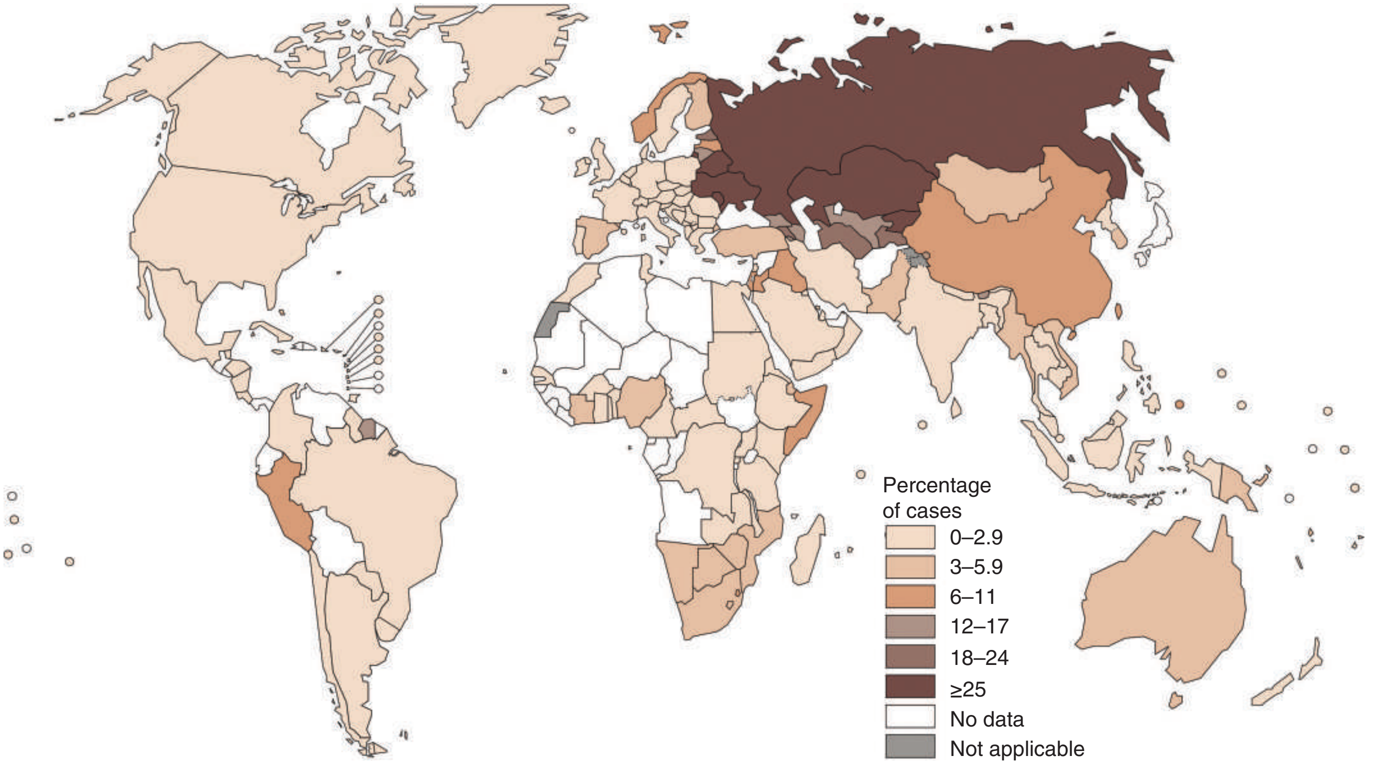

FIGURE 183-12 Percentage of new cases of multidrug-resistant/rifampin-resistant tuberculosis (TB) in all countries surveyed by the World Health Organization (WHO) Global Drug Resistance Surveillance Project during 1994–2022. Figures are based on the most recent year for which data have been reported, which varies among countries. Data reported before the year 2002 are not shown. *(See disclaimer in Fig. 183-2. Reproduced with permission from Global Tuberculosis Report 2023. Geneva, World Health Organization; 2023.)*

the large-scale use of shorter all-oral bedaquiline-containing regimens; the BEAT-Tuberculosis trial testing a new 6-month fully oral regimen consisting of BDLLfxCfz; and the endTB trial testing five different 9-month regimens consisting of different combinations of BLfx (or M) LCfzD. The assessment of this information resulted in a recent 2024 update of WHO guidance for the treatment of MDR-TB and all other RR-TB cases in which isoniazid resistance is absent or unknown (*https://iris.who.int/bitstream/handle/10665/378472/B09123-eng.pdf?sequence=1*).

As a result, the WHO is now recommending four approaches to treat MDR/RR-TB **(see Table 183-4)**: (1) a 6-month, fully oral regimen composed of bedaquiline, pretomanid, linezolid (600 mg) and moxifloxacin; (2) a 6-month oral regimen consisting of bedaquiline, delamanid, and linezolid (600 mg), plus levofloxacine and clofazimine; (3) different 9-months oral bedaquiline-containing regimens; and (4) an individualized longer regimen of at least 18 months' duration consisting of an optimal combination of oral drugs chosen according to a rational approach and using the WHO priority grouping of medicines (Table 183-4). All-oral regimens are now the preferred options, and the use of either a shorter or a longer regimen depends on the assessment of the severity of disease, knowledge of drug resistance pattern, and history of previous treatment.

Six-month, All-Oral, Bedaquiline, Pretomanid, Linezolid, and Moxifloxacin (BPaLM) Regimen New evidence generated through three recent clinical trials (Nix-TB, Ze-Nix, and TB-PRACTECAL) prompted WHO in 2022 to recommend that a regimen composed of the four drugs bedaquiline, the new nitroimidazole compound pretomanid, linezolid 600 mg daily, and moxifloxacin (BPaLM) should be preferred to previously recommended regimens for MDR/RR-TB. This regimen has shown treatment success rates of nearly 90% and, with a linezolid dosage of 600 mg daily, fewer adverse events than in the previous Nix-TB trial when the drug was administered at the dosage of 1200 mg daily. DST for fluoroquinolones should be obtained at the start of treatment and guide the decision to retain moxifloxacin in the regimen or administer only the other three drugs. This regimen is recommended for (1) patients with MDR/RR-TB or pre-XDR-TB aged 14 years or older regardless of HIV status, (2) all forms of disease except for disseminated, central nervous system, and osteoarticular TB, and (3) patients with no previous or <1 month exposure to bedaquiline, pretomanid and linezolid (with the possibility of usage also among those previously treated as long as susceptibility is confirmed). Pregnant and breast-feeding women should not be administered this regimen given the incomplete evidence on pretomanid safety in those conditions.

Six-Month Regimen, All-Oral (BDLLfxCfz) A 6-month regimen consisting of bedaquiline, delamanid, and linezolid (600 mg), plus levofloxacin and clofazimine has been recently shown to be non-inferior to the 9-month or longer regimens. Either levofloxacin or clofazimine can be omitted depending on fluoroquinolone drug susceptibility testing: in case of proven resistance to fluoroquinolone, clofazimine can be administered while in case of fluoroquinolone susceptibility, clofazimine can be avoided. This regimen can be used in all patients who have not been exposed (or who have been exposed for less than 1 month) to bedaquiline, delamanid, and linezolid. It can also be used in children, adolescent, and pregnant or breastfeeding women.

9-Month, All-Oral MDR-TB Regimens Although the 6-month regimens described above should be the first and second choice, in MDR/RR-TB patients without resistance to fluoroquinolones and who have not been exposed (or who have been exposed for less than 1 month) to bedaquiline, delamanid and linezolid, the use of 9-month, all-oral, bedaquiline-containing regimens has been proven to be non-inferior to longer, 18-month regimens. These regimens consist of, in order of preference: (i) bedaquiline, linezolid, moxifloxacin and pyrazinamide (BLMZ); (ii) bedaquiline, linezolid, levofloxacin, clofazimine, and pyrazinamide (BLLfxCfzZ); and (iii) bedaquiline, delamanid, linezolid, levofloxacin, and pyrazinamide (BDLLfxZ). In addition, in patients with no extensive pulmonary disease or severe extrapulmonary disease, with no or <1 month exposure to second-line drugs such as bedaquiline, fluoroquinolones, ethionamide, linezolid, and clofazimine (with the possibility of usage also among those previously treated as long as susceptibility is confirmed), an all-oral, bedaquiline-containing regimen may be used rather than the longer (18-month) regimens. Observational programmatic data from South Africa showed that a fully oral regimen starting with 6 months of bedaquiline accompanied by 4–6 months of levofloxacin or moxifloxacin, ethionamide, ethambutol, pyrazinamide, high-dose isoniazid (10–15 mg/kg per day), and clofazimine, and followed by 5 months of levofloxacin (or moxifloxacin), clofazimine, pyrazinamide, and ethambutol, was associated with low toxicity and better outcomes than the older, standardized, injectable-containing regimen. This regimen may also be used in children as well as in pregnant women for whom ethionamide is replaced with linezolid 600 mg daily. As in all TB cases, DST is essential to detect resistance, especially to rifampin and fluoroquinolones, before starting the 9-month regimen.

Longer MDR-TB Regimens In some MDR/RR-TB patients a longer regimen could be administered. Table 183-3 shows the priority grouping of drugs recommended by the WHO and the approach to the design of a longer regimen for both adults and children. As much as possible, the regimen is composed of all three group A agents and at least one group B agent to ensure that at least four active drugs likely to be effective are administered and at least three of them can be continued throughout the treatment if bedaquiline is discontinued. A second group B agent should be added if one or two group A agents cannot be used. Group C agents can be added to complete a regimen when group A or B agents alone cannot be used as described. Levofloxacin or moxifloxacin, bedaquiline, and linezolid should always be included in longer regimens. Clofazimine and cycloserine (or terizidone) are the two group B options to be added to group A drugs. Group C drugs can replace group A and B agents that cannot be used, and the choice should be based on drug susceptibility testing, drug resistance levels in the population, the patient's history of previous use of these drugs, and potential intolerance or toxicity. The injectable agents kanamycin and capreomycin should not be included in any regimen, since they have been associated with higher risks of failure and relapse when compared with longer regimens in which other agents were used instead. Amikacin (or streptomycin if the other aminoglycoside is not available) may be used among those aged ≥18 years as long as susceptibility is demonstrated and adverse reactions are strictly monitored. The use of ethionamide (or prothionamide) and PAS is restricted to situations in which bedaquiline, linezolid, clofazimine, and delamanid are not used, while clavulanic acid should not be included. A treatment course of at least 18–20 months is recommended, but duration may depend on patient response. In principle, a duration of 15–17 months after culture conversion is suggested. Most patients should receive an intensive course of 6–7 months. Important considerations when treating MDR-TB patients include the safety and effectiveness of several agents especially when used for long periods of time.

As in past recommendations, informed consent should be sought from patients treated with all MDR-TB regimens, and active TB drug safety monitoring is recommended. Patients taking QT interval–prolonging drugs (bedaquiline, delamanid, clofazimine, and fluoroquinolones) should be closely monitored, with electrocardiography performed at the start of treatment and repeated during treatment; patients with a QTc interval >500 ms or a history of ventricular arrhythmias should not be given these drugs. Patients taking amikacin should undergo serial audiometry to detect any hearing loss early on. Incentives and other forms of support can encourage patients not to interrupt treatment.

The design of regimens for complex patterns of MDR-TB, including XDR-TB, follows the same principles outlined in Table 183-3 through the selection of agents likely to be effective and tolerated. Observational studies have shown that aggressive management in such patients, with early drug susceptibility testing, use of a rational combination of effective drugs, strict adherence to directly observed therapy, monthly bacteriologic monitoring, and intensive patient support, may—besides interrupting transmission—increase the chances of cure and avert death. For patients with localized disease and sufficient pulmonary reserve, lobectomy or wedge resection may be considered as part of treatment.

Because the management of MDR-TB is complicated by both social and medical factors, care of seriously ill patients is ideally provided in specialized centers or, in their absence, in the context of programs with adequate resources and capacity, including community support. When patients are in stable condition, treatment and care on an ambulatory basis at a decentralized health care facility should be prioritized, as this approach may increase treatment success and reduce loss to follow-up. This approach should not, however, preclude hospitalization when it is necessary. Respiratory infection-control measures should be observed throughout. As part of a patient-centered approach, palliative and end-of-life care should be provided as a priority when all recommended treatment options have been exhausted.

HIV-ASSOCIATED TB

Several observational studies and randomized controlled trials have shown that treatment of HIV-associated TB with anti-TB drugs and simultaneous use of ART is associated with significant reductions in mortality risk and AIDS-related events. Evidence from randomized controlled trials shows that early initiation of ART during anti-TB treatment is associated with a 34–68% reduction in mortality rates, with especially good results in patients with CD4+ T cell counts of <50/μL. Therefore, the main aim in the management of HIV-associated TB is to initiate anti-TB treatment and to immediately consider initiating or continuing ART. All HIV-infected TB patients, regardless of CD4+ T cell count, are candidates for ART, which optimally is initiated as soon as possible after the diagnosis of TB and with the strong recommendation to start within the first 2 weeks of anti-TB therapy, especially for profoundly immunosuppressed patients with CD4+ T cell counts of <50/μL. Notably, patients affected by TB involving the CNS require special consideration. In these instances, it is recommended to delay the initiation of ART for the initial 8 weeks, irrespective of the CD4+ T cell count. In general, the standard 6-month daily regimen is equally effective in HIV-negative and HIV-positive patients with drug-susceptible TB. However, in the uncommon situation in which a PLWH cannot receive ART, prolongation of the continuation phase of TB treatment by 3 months can be considered. The new 4-month daily regimen including isoniazid, pyrazinamide, rifapentine at a daily dose of 1200 mg, and moxifloxacin at a daily dose of 400 mg (2HPMZ/2HPM) can also be used among PLWH and a CD4+ T cell counts of >100/μL. As in any other TB patient, intermittent regimens should not be used in PLWH. As for any other adult PLWH **(Chap. 208)**, first-line ART for TB patients consists of two nucleoside reverse transcriptase inhibitors plus a nonnucleoside reverse transcriptase inhibitor or an integrase or protease inhibitor. Recent guidelines have also considered a two-drug treatment consisting of one nucleoside reverse transcriptase inhibitor plus an integrase inhibitor. Although TB treatment modalities are similar to those in HIV-negative patients, adverse drug reactions may be more pronounced in PLWH. In this regard, three important considerations are relevant: an increased frequency of paradoxical reactions, interactions between ART components and rifamycins, and development of rifampin monoresistance with intermittent treatment. IRIS—i.e., the exacerbation of symptoms and signs of TB—has been described above. Rifampin, a potent inducer of enzymes of the cytochrome P450 system, lowers serum levels of many HIV protease inhibitors, some nonnucleoside reverse transcriptase inhibitors, and some integrase inhibitors—essential drugs used in ART. Typically, the ART regimens used alongside a rifampicin-based TB regimen include the integrase inhibitors dolutegravir or raltegravir (but not bictegravir, cabotegravir, or elvitegravir) at double the standard dose, combined with two nucleoside analogues. Otherwise, the nonnucleoside reverse transcriptase inhibitor efavirenz is a valid alternative to the integrase inhibitor. Alternatively, rifabutin, which has much less enzyme-inducing activity, has been used in place of rifampin. However, dosage adjustments for rifabutin and protease or integrase inhibitors are still being assessed. Several clinical trials have found that PLWH whose degree of immunosuppression is advanced (e.g., CD4+ T cell counts of <100/μL) are prone to treatment failure and relapse with rifampin-resistant organisms when treated with "highly intermittent" (i.e., once- or twice-weekly) rifamycin-containing regimens. Consequently, it is now recommended that all TB patients who are infected with HIV, like all other TB patients with rifampin-susceptible disease, receive a rifampin-containing regimen on a daily basis. Because recommendations are frequently updated, consultation of the following websites is advised: *www.who.int/health-topics/hiv-aids*, *www.who.int/health-topics/tuberculosis*, *www.cdc.gov/hiv*, and *www.cdc.gov/tb*.

SPECIAL CLINICAL SITUATIONS

Although comparative clinical trials of treatment for extrapulmonary TB are limited, the available evidence indicates that most forms of disease should be treated with a 6-month regimen recommended for patients with pulmonary disease. For TB meningitis, the ATS, the CDC, and the IDSA recommend extension of the continuation phase for 7–10 months. The WHO and the American Academy of Pediatrics recommend that children with bone and joint TB, tuberculous meningitis, or miliary TB receive up to 12 months of treatment (2-month induction treatment followed by 10-month consolidation treatment). Treatment for TB may be complicated by underlying medical problems that require special consideration. As a rule, patients with chronic renal failure should never receive aminoglycosides and should receive ethambutol only if serum drug levels can be monitored. Isoniazid, rifampin, and pyrazinamide may be given in the usual doses in cases of mild to moderate renal failure, but the dosage of pyrazinamide should be modified for patients with renal failure. Patients with hepatic disease pose a special problem because of the hepatotoxicity of isoniazid, rifampin, and pyrazinamide. Patients with severe hepatic disease may be treated with ethambutol, streptomycin, and possibly another drug (e.g., a fluoroquinolone); if required, isoniazid and rifampin may be administered under close supervision. The use of pyrazinamide in patients with liver failure should be avoided. Silicotuberculosis necessitates the extension of therapy by at least 2 months.

The regimen of choice for pregnant women (Table 183-4) is 9 months of treatment with isoniazid and rifampin supplemented by ethambutol for the first 2 months. Although the WHO has recommended routine use of pyrazinamide for pregnant women in combination with isoniazid and rifampin, this drug has not been recommended for pregnant women in the United States because of insufficient data documenting its safety in pregnancy. Streptomycin is contraindicated because it is known to cause eighth-cranial-nerve damage in the fetus. The thioamides, bedaquiline, and delamanid also should be avoided in the treatment of pregnant women with MDR-TB. Treatment for TB is not a contraindication to breast-feeding; most of the drugs administered will be present in small quantities in breast milk, albeit at concentrations far too low to provide any therapeutic or prophylactic benefit to the child.

Medical consultation on difficult-to-manage cases is provided by the US CDC Regional Training and Medical Consultation Centers (*https://www.cdc.gov/tb-programs/php/about/tb-coe.html*).

PREVENTION

The primary way to prevent TB is to diagnose and isolate infectious cases rapidly and to administer appropriate treatment until patients are rendered noninfectious (usually 2–4 weeks after the start of proper treatment) and the disease is cured. Additional strategies include BCG vaccination and preventive treatment of persons with TB infection who are at high risk of developing active disease.

BCG VACCINATION

One of the most used vaccines in the history of medicine, BCG was derived from an attenuated strain of *M. bovis* and was first administered to humans in 1921. Many BCG vaccines are available worldwide; all are derived from the original strain, but the vaccines vary in efficacy, ranging from 80% to nil in randomized, placebo-controlled trials. A similar range of efficacy was found in observational studies (case-control, historic cohort, and cross-sectional) in areas where infants are vaccinated at birth. These studies and a meta-analysis also found higher rates of efficacy in the protection of infants and young children from serious disseminated forms of childhood TB, such as tuberculous meningitis and miliary TB. BCG vaccine is safe and rarely causes serious complications except in those with underlying immunodeficiencies. The local tissue response begins 2–3 weeks after vaccination, with scar formation and healing within 3 months. Side effects—most commonly, ulceration at the vaccination site and regional lymphadenitis—occur in 1–10% of vaccinated persons. Some vaccine strains have caused osteomyelitis in ~1 case per million doses administered. Disseminated BCG infection ("BCGitis") and death have occurred in 1–10 cases per 10 million doses administered, although this problem is restricted almost exclusively to persons with impaired immunity, such as children with severe combined immunodeficiency syndrome or adults with HIV infection. BCG vaccination induces TST reactivity, which tends to wane with time. The presence or size of TST reactions after vaccination does not predict the degree of protection afforded.

BCG vaccine is recommended for routine use at birth in countries or among populations with high TB prevalence. However, because of the low risk of transmission of TB in the United States and other high-income countries, the variability in protection afforded by BCG, and its impact on the TST, the vaccine is not recommended for general use. HIV-infected adults and children should not receive BCG vaccine. Moreover, infants whose HIV status is unknown but who have signs and symptoms consistent with HIV infection or who are born to HIV-infected mothers should not receive BCG, nor should siblings of those with immunodeficiencies unless they are known to be unaffected.

Over the past decade, renewed research and development efforts have been made toward a new TB vaccine, and several candidates have been developed and tested. A promising candidate vaccine, M72/AS01$_E$, a subunit vaccine pairing two *M. tuberculosis* antigens (Mtb32A and Mtb39A) with the adjuvant system AS01$_E$, was tested in a randomized phase 2b trial showing an estimated efficacy of about 50% at 36 months. Adverse events were not different in the vaccine and placebo groups. This vaccine is now under development. The investigators aim to enroll up to 20,000 participants, including people living with HIV, at 60 trial sites in South Africa and six other countries including Zambia, Malawi, Mozambique, Kenya, Indonesia, and Vietnam. Participants in the double-blind trial will receive either M72/AS01E or a placebo. Completion of this trial is expected to take at least 5 years.

As of the end of 2024, 15 candidate vaccines were in the 4 stages of clinical trials. They included mycobacterial live attenuated or inactivated vaccines, viral vector vaccines, adjuvant recombinant protein vaccines and RNA vaccines. Several challenges must be faced in the development of a TB vaccine. For instance, the lack of predictive animal models and protection correlates renders trials long and expensive. Furthermore, the decision about whether a candidate vaccine should be developed for prevention of infection (preexposure) or prevention of reactivation (postexposure) without an exact understanding of its precise mechanism of action is complex.

TB PREVENTIVE TREATMENT (TPT)

It is estimated that nearly 2 billion individuals—a quarter of the human population—have been infected with *M. tuberculosis* in their lifetime. Although only a small fraction of these infections will progress toward active disease, new active cases will continue to emerge from this pool of infected individuals. Therefore, TPT (also called *chemoprophylaxis* or *preventive chemotherapy*, and previously referred to as *treatment of latent TB infection*) is a fundamental intervention in TB control and elimination strategies.

Infection can be tested using TST, the new *Mycobacterium tuberculosis* antigen-based skin tests (TBSTs), or IGRA, although these tests just measure host immune response to TB antigens. Unfortunately, at present, there is no gold-standard diagnostic test that can confirm true infection (as opposed to immunologic memory of previous exposure) or predict which infected individuals will develop active TB. As a result, decisions to treat infection should include consideration of the risk of progression in an individual. For skin testing, five tuberculin units of polysorbate-stabilized PPD should be injected intradermally into the volar surface of the forearm (i.e., the Mantoux method). Multipuncture tests are not recommended. Reactions are read at 48–72 h as the transverse diameter (in millimeters) of induration; the diameter of erythema is not considered. In some persons, TST reactivity wanes with time but can be recalled by a second skin test administered ≥1 week after the first (i.e., two-step testing). For persons periodically undergoing the TST, such as health care workers and individuals admitted to long-term-care institutions, initial two-step testing may preclude subsequent misclassification of those who have boosted reactions as TST converters. The cutoff for a positive TST (and thus for TPT) is related both to the probability that the reaction represents true infection and to the likelihood that the individual, if truly infected, will develop TB. **Table 183-5** suggests possible conventional cutoff by risk group. Thus, positive reactions for PLWH, recent close contacts of infectious cases, organ transplant recipients, previously untreated persons whose chest radiograph shows fibrotic lesions consistent with old

TABLE 183-5 Tuberculin Reaction Size and Cutoff for Tuberculosis (TB) Preventive Treatment

RISK GROUP	TUBERCULIN REACTION SIZE, mm
HIV-infected persons	≥5
Recent contacts of a patient with TB	≥5[a]
Organ transplant recipients	≥5
Persons with fibrotic lesions consistent with old TB on chest radiography	≥5
Persons who are immunosuppressed—e.g., due to the use of glucocorticoids or tumor necrosis factor α inhibitors	≥5
Persons with high-risk medical conditions[b]	≥5
Recent immigrants (≤5 years) from high-prevalence countries	≥10
Injection drug users	≥10
Mycobacteriology laboratory personnel; residents and employees of high-risk congregate settings[c]	≥10
Children <5 years of age; children and adolescents exposed to adults in high-risk categories	≥10
Low-risk persons[d]	≥15

[a] Tuberculin-negative contacts, especially children, should receive prophylaxis for 2–3 months after contact ends and should then undergo repeat tuberculin skin testing (TST). Those whose results remain negative should discontinue prophylaxis. HIV-infected contacts should receive a full course of treatment regardless of TST results.

[b] These conditions include silicosis and end-stage renal disease managed by hemodialysis.

[c] These settings include correctional facilities, nursing homes, homeless shelters, and hospitals and other health care facilities.

[d] Except for employment purposes where longitudinal TST screening is anticipated, TST is not indicated for these low-risk persons. A decision to treat should be based on individual risk/benefit considerations.

Source: Adapted from Centers for Disease Control and Prevention: *https://www.cdc.gov/tb/publications/factsheets/testing/Tuberculin_Skin_Testing_Information_for_Health_Care_Providers.pdf.*

TB, and persons receiving drugs that suppress the immune system are defined as an area of induration ≥5 mm in diameter. A 10-mm cutoff is used to define positive reactions in most other at-risk persons. For persons with a very low risk of developing TB if infected, a cutoff of 15 mm is used. (Except for employment purposes where longitudinal screening is anticipated, the TST is not indicated for these low-risk persons.) Criteria for positive TBSTs are based on manufacturer recommendations, and some require induration with a diameter cutoff of 5 mm. A positive IGRA also is based on manufacturer recommendations. Good clinical practice requires that, in addition to test results, epidemiologic and clinical factors also guide the decision to implement TPT and that active TB be definitively excluded before the initiation of a prophylactic regimen. The WHO recommends systematic testing for infection and TPT for the following groups at high risk of progression from infection to disease or of exposure and infection: adults, adolescents and children older than 12 months living with HIV; infants with HIV aged <12 months who are contacts of persons with TB; all household contacts of patients with infectious pulmonary TB including children <5 years of age; patients with silicosis, patients starting anti-TNF treatment, patients on dialysis, and patients preparing for organ or hematologic transplantation. In addition, testing and TPT may be considered for persons living or working in at-risk institutional or crowded settings, such as prisoners, health care workers, recent immigrants from high-TB-burden countries, and homeless persons who use drugs.

Some skin test- and IGRA-negative individuals are also candidates for TPT. Once an appropriate clinical evaluation has excluded active TB, infants and children <5 years of age who were in contact with infectious cases should be offered TPT even in the absence of a positive test for TB infection. PLWH >1 year of age who have been exposed to an infectious TB patient should receive TPT regardless of the results of a TB infection test. Any HIV-infected candidate for TPT must be screened carefully to exclude active TB, which would necessitate full disease treatment. The use of a clinical algorithm based on four signs/symptoms (current cough, fever, weight loss, and night sweats) helps to decide which PLWH can start TPT. The absence of all four symptoms tends to exclude active TB in PLWH. The presence of one of these four manifestations, on the other hand, warrants further investigation for active TB before TPT is started. Although a test for TB infection is prudent before starting TPT, this test is not an absolute requirement—given the logistical challenges—among contacts aged <5 years and PLWH in high-TB-incidence and low-resource settings.

Among PLWH receiving ART, conversion of the TST from negative to positive can occur during the first few months of TPT. Conversions (from negative to positive) and reversions (from positive to negative) are more common with IGRAs than with TSTs among serially tested health care workers in the United States.

TPT in selected persons at risk aims to prevent active disease and, in the absence of an immunizing vaccine, is a critical component of TB elimination strategies. This intervention is based on the results of a large number of randomized, placebo-controlled clinical trials demonstrating that a 6- to 9-month course of isoniazid reduces the risk of active TB in infected patients by up to 90%. Analysis of available data indicated that the optimal duration of treatment with this drug was ~9 months. In the absence of reinfection, the protective effect is believed to be lifelong. Clinical trials have shown that isoniazid reduces rates of TB among TST-positive PLWH. Studies in HIV-infected patients have also demonstrated the effectiveness of shorter TPT regimens containing a rifamycin. Several TPT regimens **(Table 183-6)** can be used. The most widely used has been that based on isoniazid alone at a daily dose of 5 mg/kg (up to 300 mg/d) for 9 months. On the basis of cost–benefit analyses and concerns about feasibility, a 6-month period of treatment at the same dose is considered adequate by the WHO. A clinical trial showed that a regimen of isoniazid (900 mg) and rifapentine (900 mg), given once weekly for 12 weeks, is as effective as the standard 9-month isoniazid regimen. This regimen was associated with higher rates of treatment completion (82% vs 69%) and less hepatotoxicity (0.4% vs 2.7%) than isoniazid alone, although the rate of permanent discontinuation due to an adverse event was higher (4.9% vs 3.7%). Currently, the isoniazid–rifapentine regimen is not recommended for children <2 years of age or pregnant women. A 3-month regimen of daily isoniazid and rifampin is used in some countries (e.g., the United Kingdom) for both adults and children who are known not to have HIV infection. An alternative regimen for adults is 4 months of daily rifampin, which should also be effective against isoniazid-resistant strains.

TABLE 183-6 Recommended Regimens and Drug Dosages for Tuberculosis Preventive Treatment[a]

REGIMEN	DOSE	ADVERSE EVENTS
Isoniazid alone for 6 or 9 months	Adults: 5 mg/kg (max, 300 mg) per day Children <10 years of age: 10 mg/kg per day (range, 7–15 mg)	Drug-induced liver injury, nausea, vomiting, abdominal pain, skin rash, peripheral neuropathy, dizziness, drowsiness, seizure
Rifapentine plus isoniazid for 3 months	Adults and children: Isoniazid: 15 mg/kg (900 mg) weekly Rifapentine: 15–30 mg/kg (900 mg) weekly	Hypersensitivity reactions, petechial skin rash, drug-induced liver injury Anorexia, nausea, abdominal pain Hypotensive reactions
Isoniazid plus rifampin for 3 months	As below	As above
Rifampin alone for 4 months	Adults: 10 mg/kg per day Children <10 years of age: 15 mg/kg (range, 10–20 mg) per day	Flulike syndrome, skin rash, drug-induced liver injury, anorexia, nausea, abdominal pain, neutropenia, thrombocytopenia, renal reactions (e.g., acute tubular necrosis and interstitial nephritis)
Rifapentine plus isoniazid for 1 month	Age >13 years only: isoniazid 300 mg and rifapentine 600 mg daily (28 doses)	Essentially similar to those of rifapentine plus isoniazid for 3 months with neutropenia more common and elevation in liver enzyme levels and neuropathy less common
Levofloxacin for 6 months	Daily: adults 10–15 mg/kg; children 15–20 mg/kg (maximum 750 mg)	No grade 3, 4 or serious adverse events reported

[a] See text for full description of evidence on and limitations of these regimens.

Source: Reproduced with permission from World Health Organization.

Recently, an open-label, randomized, phase 3 noninferiority trial has shown that among PLWH, a 1-month regimen of daily rifapentine plus isoniazid was noninferior to the 9-month daily isoniazid regimen and ensured a higher treatment completion. As a result, the WHO has included a 1-month regimen composed of daily isoniazid (300 mg) and rifapentine (600 mg) among the available options for patients aged 13 years or more. Rifampin and rifapentine are contraindicated in PLWH receiving protease inhibitors or most nonnucleoside reverse transcriptase inhibitors (e.g., nevirapine), as well as in those with chronic hepatitis B receiving tenofovir alafenamide. Efavirenz and tenofovir disoproxil can be used for simultaneous administration with a rifamycin without dose adjustment. However, the dose of the integrase inhibitor dolutegravir needs to be increased to 50 mg twice daily when given together with rifampin, a dose that is usually well tolerated and gives equivalent efficacy in viral suppression and recovery of CD4+ cell count compared with efavirenz. Administration of rifapentine with raltegravir was found to be safe and well tolerated. A recent phase 1/2 trial of a 3-month regimen of isoniazid plus rifapentine and dolutegravir in adults with HIV reported good tolerance and viral load suppression, reported no adverse events of grade >3, and did not indicate that rifapentine reduced dolutegravir levels sufficiently to require dose adjustment. Clinical trials to assess the efficacy of long-term isoniazid administration (i.e., for at least 3 years) among PLWH in high-TB-transmission settings have shown that this regimen can be more effective than 9 months of isoniazid and is therefore recommended under those circumstances. Studies looking at whether briefer treatment with rifapentine-based regimens could achieve similar efficacies have been undertaken. Isoniazid should not be given to persons with active liver disease. All isoniazid recipients at increased risk of hepatotoxicity (e.g., those abusing alcohol daily and those with a history of liver disease) should undergo baseline and then monthly assessment of liver function; they should be carefully educated about hepatitis and instructed to discontinue use of the drug immediately should any symptoms develop. Moreover, these patients should be seen and questioned monthly during therapy about adverse reactions and should be given no more than a 1-month supply of drug at each visit. Persons receiving high-dose isoniazid and who are at risk of vitamin B6 (pyridoxine) deficiency should receive pyridoxine to prevent peripheral neuropathy.

TPT among persons likely to have been infected by a multidrug-resistant strain is a challenge because no clinical trial results are available to guide treatment. Close observation for early signs of disease is one option. However, in selected high-risk household contacts of patients with MDR-TB (e.g., children, recipients of immunosuppressive therapy), TPT may be considered on the basis of individualized risk assessment and clinical criteria. In the absence of evidence of efficacy of any regimen, important factors in the decision to treat include intensity of exposure, certainty about a source case, information on the drug resistance pattern of the index case, and potential adverse events. Confirmation of infection with available testing is generally required. Drug selection should be based on the drug susceptibility profile of the index case. WHO recommends daily levofloxacin for 6 months in people exposed to MDR-TB. Some studies are currently investigating the role of delamanid and bedaquiline as TPT.

It may be more difficult to ensure adherence to TPT than when treating those with active TB. If family members of patients with active TB are being treated, adherence and monitoring may be easier. When feasible, supervised therapy may increase the likelihood of completion. As in active cases, the provision of incentives also may be helpful. Currently, no evidence shows that large-scale use of TPT leads to significant development of drug resistance. However, before TPT begins, it is mandatory to carefully exclude active TB in order to prevent undertreatment and development of drug resistance.

PRINCIPLES OF TB CONTROL

The highest priority in any TB control program is the early detection of all infectious cases and the provision of treatment under optimal case-management conditions with social support until cure. In addition, regular screening of high-risk groups, including immigrants from high-prevalence countries, migrant workers, prisoners, homeless individuals, substance abusers, and HIV-seropositive persons, is recommended to detect either infection or disease and treat promptly. Contact investigation is an important component of efficient TB control. A great deal of attention should been given to interrupt and prevent transmission of TB in institutional settings such as hospitals, homeless shelters, and prisons. Measures to limit such transmission include respiratory isolation of persons with suspected TB until they are proven to be noninfectious, proper ventilation in rooms of patients with infectious TB, use of ultraviolet irradiation in areas of increased risk of TB transmission, correct use of personal protective equipment, and periodic screening of personnel who may come into contact with known or unsuspected cases of TB.

In high-prevalence countries, the essential elements of good TB care and control consist of well-defined interventions including early detection of cases and bacteriologic confirmation of the diagnosis and drug resistance pattern; administration of the proper regimen ensuring adherence to treatment and social support to patients; availability of drugs of proven quality, with an effective supply and management system; and a monitoring and evaluation system, including assessment of treatment outcomes and measurement of the impact of control measures on indicators such as mortality, incidence, and prevalence of drug resistance. In the era of the United Nations (UN) Sustainable Development Goals (2016–2030) and of high political visibility of TB with international targets established by the UN General Assembly, the TB response is multisectoral and holistic. Engagement beyond dedicated programs and even the health sector is essential, as is multisectoral accountability. The "End TB" strategy promoted by the WHO since 2016 builds on three pillars and relies on increased investments and efforts by all governments, their national programs, and a multitude of partners within and beyond the health sector: (1) integrated, patient-centered care and prevention; (2) bold policies and supportive systems; and (3) intensified research and innovation. A "fourth" multisectoral pillar is however necessary to pursue TB elimination in the distant future. In fact, besides specific clinical care and control interventions as described in this chapter, elimination of TB in a society ultimately will require control and mitigation of the multitude of direct risk factors (e.g., HIV infection, smoking, alcohol abuse, diabetes) and socioeconomic determinants (e.g., extreme poverty, inadequate living conditions and poor housing, undernutrition, indoor air pollution) with clearly implemented policies within the health sector and other sectors linked to human development and welfare.

FURTHER READING

Bhargava A et al: Nutritional support for adult patients with microbiologically confirmed pulmonary tuberculosis: Outcomes in a programmatic cohort nested within the RATIONS trial in Jharkhand, India. Lancet Glob Health 11:1402, 2023.

Conradie F et al: Bedaquiline–pretomanid–linezolid regimens for drug-resistant tuberculosis. N Engl J Med 387:810, 2022.

Dorman SE et al: Four-month rifapentine regimens with or without moxifloxacin for tuberculosis. N Engl J Med 384:1705, 2021.

Migliori GB et al: Clinical standards for the assessment, management and rehabilitation of post-TB lung disease. Int J Tuberc Lung Dis 25:797, 2021.

Nahid P et al: An Official American Thoracic Society/Centers for Disease Control and Prevention/European Respiratory Society/Infectious Diseases Society of America clinical practice guideline: Treatment of drug-resistant tuberculosis. Am J Respir Crit Care Med 200:e93, 2019.

Nyang'wa BT et al: A 24-week, all-oral regimen for rifampin resistant tuberculosis. N Engl J Med 387:2331, 2022.

Reid M et al: Scientific advances and the end of tuberculosis: A report from the Lancet Commission on Tuberculosis. Lancet 402:1473, 2023.

Swindells S et al: One month of rifapentine plus isoniazid to prevent HIV-related tuberculosis. N Engl J Med 380:1001, 2019.

Uplekar M et al: WHO's new End TB strategy. Lancet 385:1799, 2015.

Walker TM et al: The 2021 WHO catalogue of *Mycobacterium tuberculosis* complex mutations associated with drug resistance: A genotypic analysis. Lancet Microbe 3:e265, 2022.

WEBSITES

United Nations General Assembly: Political Declaration of the High-Level Meeting on the Fight Against Tuberculosis. *https://documents.un.org/doc/undoc/gen/n23/306/91/pdf/n2330691.pdf*

World Health Organization: Global tuberculosis report 2024. Geneva: World Health Organization; 2024. *https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2024*

World Health Organization: WHO consolidated guidelines on tuberculosis: Module 3: diagnosis: rapid diagnostics for tuberculosis detection, 3rd ed. Geneva: World Health Organization; 2024. *https://www.who.int/publications/i/item/9789240089488*

World Health Organization: WHO consolidated guidelines on tuberculosis. Module 4: treatment - drug-resistant tuberculosis treatment, 2022 update. Geneva: World Health Organization; 2022 .*https://www.who.int/publications/i/item/9789240063129*

World Health Organization: WHO consolidated guidelines on tuberculosis. Module 4: treatment: drug-susceptible tuberculosis treatment. Geneva: World Health Organization; 2022. *https://www.who.int/publications/i/item/9789240048126*

184

Leprosy

Jan H. Richardus, Hemanta K. Kar, Zoica Bakirtzief, Wim H. van Brakel

Leprosy, also referred to as Hansen's disease, is a chronic infectious disease caused by *Mycobacterium leprae*. The clinical manifestations are largely confined to the skin, peripheral nervous system, eyes, and upper respiratory tract. The differing immune responses to *M. leprae* result in a spectrum of disease ranging from tuberculoid to lepromatous leprosy. *M. leprae* has a predilection for peripheral nerves, and immunologically mediated reactional states can cause nerve damage to the face, arms, and legs; this damage often results in disability, which in turn can lead to stigma and social exclusion. The physical disfigurement that accompanies leprosy has left marks on society that have endured long after the disease's disappearance in many countries. In everyday language, leprosy has become a metaphor for a horrible condition that warrants social exclusion. Leprosy is a neglected disease and is often thought no longer to exist. However, 174,087 new cases from 182 countries were reported in 2022. A general lack of awareness among both the public and medical practitioners often delays diagnosis and treatment and thus results in irreversible impairments. Early diagnosis and treatment of leprosy and leprosy reactions can cure the disease and prevent most chronic complications.

ETIOLOGY

M. leprae is an obligate, intracellular, acid-fast staining, rod-shaped bacterium, measuring 1–8 μm in length and 0.3 μm in diameter. *M. leprae* mostly appears irregularly stained and fragmented or granular, in which case the organism is usually considered to be dead. The few bacteria that are brightly and uniformly stained are thought to be solid, viable bacilli. The morphologic index is a measure of uniformly stained solid bacilli on slit-skin smear examination and is calculated as the percentage of viable bacilli among the total number of bacilli counted under oil-immersion microscopy. On slit-skin smear examination at the lepromatous end of the disease spectrum, *M. leprae* is predominantly found in clumps or globi within macrophages (*lepra cells*). Inside these cells, *M. leprae* multiplies in unrestricted fashion, and hundreds of bacilli may be present; the organisms are arranged in parallel arrays placed side by side as a result of the presence of surface lipids (*glial substances*). The bacteriologic index is a logarithmic-scaled measure of the density of bacilli of all forms found in the dermis upon slit-skin smear examination, varying from 0 to 6+ (with or without globi) from the tuberculoid to the lepromatous end of the disease spectrum. The bacteriological index falls an average of 1 log unit per year with multidrug therapy. *M. leprae* infects mainly macrophages and Schwann cells. It has never been grown in artificial media. Reproduction occurs by binary fission, and the organism grows slowly (over 12–14 days) in the footpads of mice. The temperature required for survival and proliferation—between 27°C and 30°C—explains the greater impact of the disease on surface areas such as the skin, peripheral nerves, testicles, and upper airways, with less inner visceral involvement. *M. leprae* remains viable for 9 days in the environment.

Ultrastructural Characteristics of *M. leprae* Electron microscopy reveals that *M. leprae* has a cytoplasm, plasma membrane, cell wall, and capsule. The cytoplasm contains structures common in gram-positive microorganisms. The plasma membrane has a permeable lipid bilayer containing interacting proteins—the protein surface antigens. Similar to that of other mycobacteria, *M. leprae*'s cell wall, which is attached to the plasma membrane, is composed of peptidoglycans bound to branched-chain polysaccharides; these peptidoglycans are arabinogalactans, which support mycolic acids, and lipoarabinomannan (LAM). The capsule—the outermost structure—contains lipids, particularly phthiocerol dimycocerosate and phenolic glycolipid (PGL-1), which has a trisaccharide bound to lipid by a molecule of phenol. Because this trisaccharide is antigenically specific for *M. leprae*, its detection is helpful in serologic diagnosis of leprosy.

Genome of *M. leprae* Comparative analysis of the genomics of single-nucleotide polymorphisms indicates that four distinct strains of *M. leprae* originated in East Africa or Central Asia. A mutation spread to Europe and subsequently underwent two separate mutations that were then followed by spread to West Africa and the Americas. The genome of *M. leprae* is circular. Its estimated molecular mass is 2.2×10^9 Da, with 3,268,203 base pairs and a guanine-plus-cytosine content of 57.8%.

Culture Difficulties Compared to the genome of *Mycobacterium tuberculosis,* that of *M. leprae* underwent reductive evolution, resulting in a smaller genome rich in inactive or entirely deleted genes. This reductive evolution, gene decay, and genome downsizing all may explain the unusually long generation time and may account for the inability to culture the leprosy bacillus in artificial media. As a result, propagation of *M. leprae* has been restricted to animal models, including the armadillo and normal, athymic, and gene-knockout mice. These systems have provided the basic resources for genetic, metabolic, and antigenic studies of the bacillus. Growth of *M. leprae* in mouse footpads also provides a tool for assessing the viability of the bacteria and testing the drug susceptibility of clinical isolates.

Immunologic Properties of *M. leprae* *M. leprae* induces both humoral and cell-mediated immune responses. The immunogenic components of *M. leprae* include polysaccharides and proteins. Polysaccharide components induce mainly a humoral immune response, whereas protein components induce both humoral and cell-mediated immune responses. The immunogens in *M. leprae* form two distinct groups: cytoplasmic antigens and antigens from the mycobacterial cell. As mentioned above, a species-specific phenolic glycolipid, PGL-1, has been identified in *M. leprae*. Other varieties of *M. leprae* antigens identified with monoclonal antibodies include antigens of 18, 28, 7, 14, 36, 65, and 70 kDa that may possibly induce an immune response.

Mycobacterium lepromatosis In 2008, a new mycobacterial species, *M. lepromatosis*, was isolated from patients with a special type of diffuse lepromatous leprosy known as *diffuse leprosy of Lucio and Latapí*. This clinical variety of leprosy is found mainly in Mexico and Central America. *M. lepromatosis* is very similar to *M. leprae* microbiologically and clinically. Microbiologically, both species are acid-fast and noncultivable and preferentially infect skin and peripheral nerves. Clinically, differentiation of *M. lepromatosis* from *M. leprae* in individual patients is not diagnostically necessary since both organisms respond well to the same antimycobacterial regimens.

EPIDEMIOLOGY

Incidence, Prevalence, and Disability The true incidence of leprosy is difficult to establish because the figure is very low and because the initial signs and symptoms are often insidious, and thus not all cases are detected as they occur. In 2022, as stated earlier, 174,087 new cases were reported to the World Health Organization (WHO) from 182 countries. New case detection per year is commonly used as a proxy for incidence, but operational factors, such as the intensity of case detection, the use of surveys, the use of contact tracing, the level of community awareness, and the quality and availability of health care, have a profound effect on case detection rates. In nonendemic countries around the world, leprosy is often misdiagnosed simply because it is not considered.

The registered prevalence of leprosy is defined as the number of patients receiving treatment at a point in time (usually at the end of a calendar year). The registered prevalence is a proxy measure for true prevalence, which would include existing cases that have not yet been detected. The two factors that determine the registered prevalence are the new case detection rate and the duration of treatment; changes in either factor will affect the registered prevalence.

The WHO leprosy disability grading system scores patients according to the presence of disabilities of the eyes, hands, and feet. For the

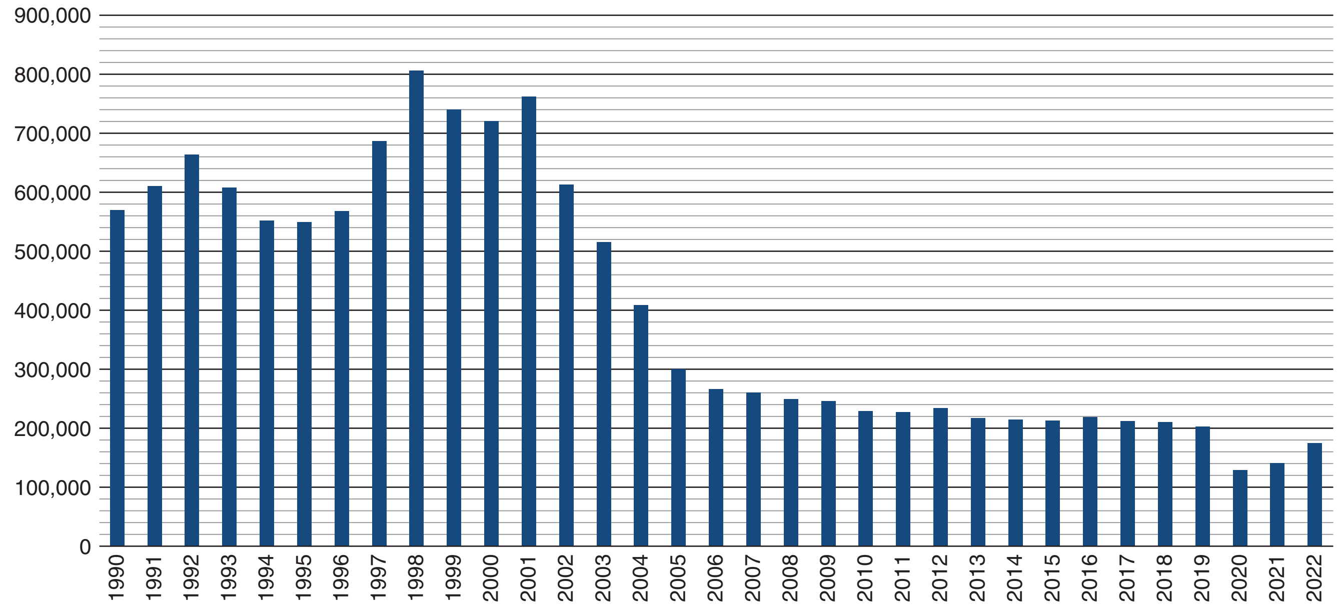

FIGURE 184-1 Global trend in leprosy new-case detection, 1990–2022.

hands and feet, grade 0 means no anesthesia and no visible impairment; grade 1 signifies anesthesia but no visible impairment; and grade 2 indicates visible impairment. For the eyes, grade 0 signifies no eye problems due to leprosy and no evidence of visual loss; grade 1 signifies eye problems due to leprosy without severe effects on vision; and grade 2 indicates severe visual impairment (vision score worse than 6/60; inability to count fingers at 6 meters) and also includes lagophthalmos, iridocyclitis, and corneal opacities. The sum score for these six body sites is called the Eye-Hand-Foot (EHF) score and is used as an overall indicator of the impairment status of a person with leprosy. Leprosy-related grade 2 disability is usually reported as the proportion of people with such disability at any site among patients newly diagnosed with leprosy in a specific year.

The global trend in new case detection since 1990 is presented in Fig. 184-1. The trend was remarkably static up to the year 2001, with a peak around the year 2000; fell dramatically between 2001 and 2005; and has leveled off from 2006 until 2016. Since 2017, a continuous decline has been observed, but the figures for 2020 and 2021 are unreliable due to underreporting during the COVID-19 pandemic. The most important factor contributing to the fast downward trend after the year 2000 was the decline in leprosy control activities following the declaration by the WHO in 2000 that leprosy was eliminated as a "public health problem." *Elimination* was defined as a prevalence of <1 case per 10,000 population at the global level. The decline in new case detection since 2016 to below 200,000 cases per year indicates that transmission of *M. leprae* is becoming less and that elimination of leprosy could become possible in the next 25 years or so.

Sex, Age, and Geographic Distribution Approximately 40% of all reported leprosy patients are women, but the low proportion in some countries raises concerns about underdiagnosis in women due to poor access to health services, illiteracy, low status, and other cultural factors. The age-specific incidence often shows a bimodal pattern, with peaks in the teenage years and in adulthood. Around 8% of all newly detected cases are found in children (<15 years of age), a measure that is often taken as an indicator of continued (recent) transmission. Leprosy is rare among children <5 years of age. Around 5% of all patients have a grade 2 disability.

There are large variations among world regions and countries in new case detection rates. Approximately 80% of global new case detection is reported from India, Brazil, and Indonesia. There are also distinct geographic variations within countries, with differences between urban and rural communities and clustering of cases at the village or neighborhood level. Geographic variations can be due to differences in health service provision, socioeconomic development, isolation, and poverty. Figure 184-2 depicts the geographic distribution of new leprosy cases in 2022.

Transmission Understanding of the transmission of *M. leprae* is limited. The existing evidence is largely circumstantial because of the long incubation period from exposure to disease, the inability to culture *M. leprae*, and the difficulty of diagnosing both infection and early disease. *M. leprae* organisms can be shed in large numbers from the mouth and nose of patients with untreated multibacillary leprosy (droplet infection) and sometimes from damaged skin, but it is unclear whether patients with paucibacillary leprosy can spread the bacillus. There is evidence for transmission between humans and—in southern U.S. states—for zoonotic transmission through wild armadillos. The main route of entry into the body is assumed to be the respiratory tract, but in patients with wounds or tattoos, transmission through the skin also is possible.

Reservoirs of Infection It is assumed that humans are the main reservoir of infection for *M. leprae*. The armadillo is also a reservoir for human infection. Certain species of monkeys and red squirrels are infected with *M. leprae* in the wild, but there is no evidence of transmission to humans through contact with these animals. Evidence is weak for the potential of water and soil as environmental sources of *M. leprae*. The higher incidence rate of leprosy among household contacts of multibacillary cases than among those of paucibacillary cases suggests that multibacillary cases represent an important reservoir for undetected and untreated cases in the community; that is, a prolonged period between the onset of signs of leprosy and treatment due to a delay in diagnosis and initiation of multidrug therapy increases exposure in the community. Persons with subclinical leprosy are likely to be a main source of infection, given that multidrug therapy for clinical leprosy apparently has not made an impact on transmission.

Incubation Period, the Role of Contacts, and Genetic Susceptibility The incubation period of leprosy is estimated to range from 2 to ≥10 years. The incubation period for multibacillary leprosy appears to be longer (5 to ≥15 years) than that for paucibacillary leprosy (~2–5 years). Poverty-associated factors such as low level of education, poor hygiene, and food shortages have been identified as risk factors for leprosy, but the most important risk factors are associated with intimacy and duration of contact with a leprosy patient, in particular with an index case with multibacillary leprosy, and the intensity of contact with and physical distance from the index patient. Increasing evidence from studies in twins and from observational studies supports host genetic susceptibility to leprosy. Ongoing studies are exploring the

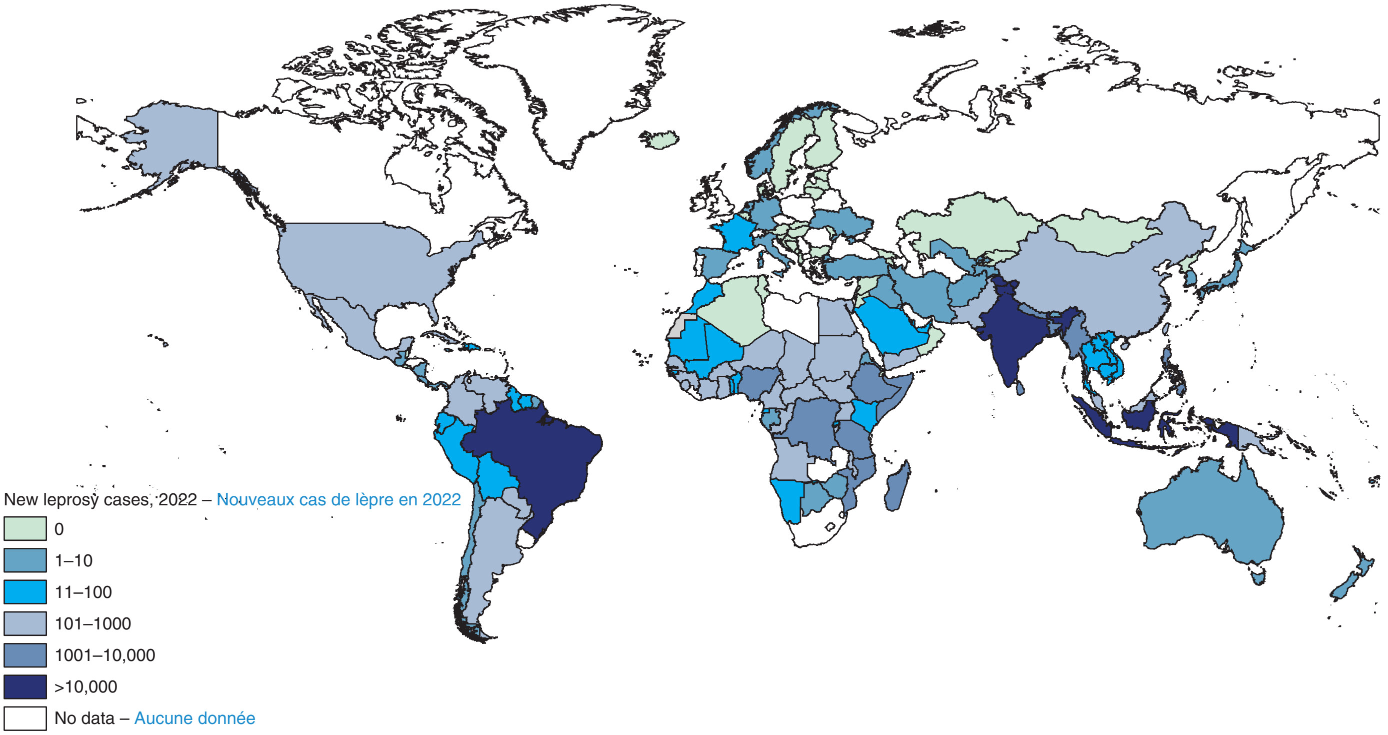

FIGURE 184-2 Geographic distribution of new leprosy cases, 2022. *(Reproduced with permission from Global leprosy (Hansen disease) update, 2022: new paradigm-control to elimination. Wkly Epidemiol Rec 98:409, 2023.)*

mechanism underlying genetic susceptibility to leprosy and its clinical manifestations.

PATHOGENESIS

Whatever the route of *M. leprae*'s entry into the human body, the pathogenic process usually starts in the peripheral nerves. Once bacilli are engulfed by Schwann cells, the histopathologic changes in nerve and skin—and thus the type of leprosy that develops—depend on the immunologic resistance of the person infected, in particular on the cell-mediated immune (CMI) response to the bacillus and its antigens.

Ridley-Jopling Classification of Leprosy In 1962, Ridley and Jopling described five overlapping categories of leprosy: tuberculoid (TT), borderline tuberculoid (BT), mid-borderline (BB), borderline lepromatous (BL), and lepromatous (LL). An early clinical manifestation is recognized and referred to as indeterminate leprosy (IL). Immunologic resistance is strong at the tuberculoid end of the spectrum, gradually diminishes through the borderline spectrum, and is weakest in lepromatous leprosy. The LL and TT types of leprosy are relatively stable, with little or no change in clinical disease expression over time, while the BL, BB, and BT types are unstable both clinically and immunologically. Further distinction indicates that subpolar types of TT and LL leprosy (TTs and LLs) are less stable than polar types (TTp and LLp). The immune reaction depends on predisposing genetic factors and the extent of exposure to *M. leprae*. The host tissue's reaction and related damage are largely due to delayed hypersensitivity. In response to the presence of *M. leprae*, a granuloma is formed either by macrophage–lymphocyte interaction when there is immunity or otherwise by macrophages only. The formation of a granuloma is preceded by a stage of infiltration by lymphocytes alone, as is seen in IL. Because of the strong immune response toward the tuberculoid end of the spectrum, macrophages, along with many lymphocytes, become fixed epithelioid cells, and groups of these cells become giant cells. The tuberculoid granuloma leads to nerve destruction resulting in anesthesia and muscle weakness. The cellular response is less focal and less destructive in the borderline portion of the spectrum; consequently, there is less damage to nerves and few bacilli are present. In BL leprosy, there are macrophage granulomas along with lymphocytes, but little nerve damage and more bacilli. In LL leprosy, bacilli multiply within Schwann cells and perineural cells. Liberated bacilli from these cells are engulfed by histiocytes, becoming wandering macrophages and traveling throughout the body to other nerves and tissues via blood, lymph, and tissue fluids. In addition, there are diffuse lepromas in LL leprosy that consist of histiocytes and/or macrophages, with very few lymphocytes and plasma cells. The bacilli are packed within macrophages called *globi* and outside macrophages either singly or in small groups.

WHO Simplified Clinical Classification of Leprosy Ridley-Jopling classification requires clinical and pathologic expertise that does not exist in many settings. The WHO has therefore introduced a simplified classification system based on slit-skin smear: patients with negative slit-skin smear results at all body sites are classified as having paucibacillary leprosy, whereas patients with positive smears at any body site are classified as having multibacillary leprosy. However, because slit-skin smear facilities are not available or dependable in many countries, most leprosy control programs use clinical criteria only for classifying leprosy and deciding on the appropriate treatment regimen for individual patients. In this circumstance, paucibacillary leprosy is defined as one to five skin lesions and no or only one involved peripheral nerve, while multibacillary leprosy is defined as six or more skin lesions and/or more than one involved peripheral nerve.

CLINICAL MANIFESTATIONS

Leprosy is a disease affecting mainly the skin, cutaneous and peripheral nerves, mucous membranes, and, less commonly, other sites such as joints, lymph nodes, eyes, and testes. Other systemic manifestations may occur, particularly in BL and LL disease, with or without leprosy reactions. Most dermal and cutaneous nerves feeding skin lesions are affected—e.g., the supraorbital, great auricular, radial cutaneous, infrapatellar, superficial fibular, and sural nerves and the cutaneous nerves of the thigh. The peripheral nerves involved include the ulnar, median, radial (in upper limbs), lateral popliteal, and posterior tibial (in lower limbs). The cranial nerves commonly involved are the trigeminal and facial.

Indeterminate Leprosy (IL) This early clinical type manifests as one or a few hypopigmented or faintly erythematous, ill-defined to well-defined macular lesions measuring 1–5 cm in diameter. These lesions invariably occur on the external aspects of the limbs, buttocks,

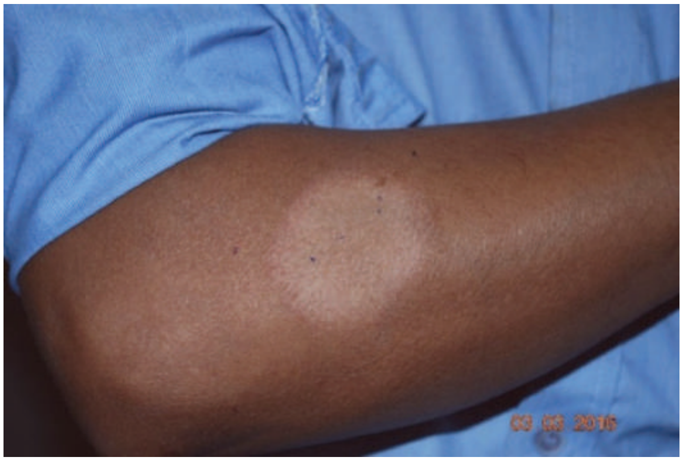

FIGURE 184-3 Tuberculoid (TT) leprosy. Hypopigmented macular lesion with a well-defined edge and loss of fine-touch sensation. *(From Dr. H. K. Kar, with permission.)*

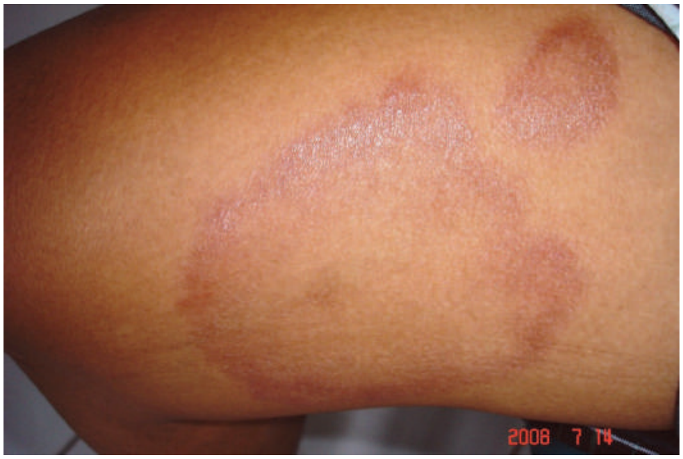

FIGURE 184-4 Borderline tuberculoid (BT) leprosy. Macular lesion with irregular, moderately defined edge and satellite lesion, with loss of sensation. *(From Dr. W. H. van Brakel, with permission from NLR.)*

and face, with mild to moderate impairment of touch and/or thermal sensations. There is no thickening of the corresponding cutaneous and peripheral nerves. IL is often, but not always, the first clinical sign of leprosy. This type either heals spontaneously or progresses to a determinate form of the disease (TT, BT, BB, BL, or LL), depending on CMI status.

Tuberculoid (TT) Leprosy TT leprosy (Fig. 184-3) presents either as a well-defined, hypopigmented macule or as a raised, erythematous/brown/copper-colored plaque with a well-defined edge. The lesions may be found on any part of the skin and are characterized by complete loss of fine touch and temperature sensations over their surface. Skin lesions are single or few (up to three) in number and can be of any size, but they seldom measure >10 cm in diameter. In plaque-type lesions, the raised clear-cut edge often slopes inward to a flattened and sometimes hypopigmented central area, acquiring an annular configuration. The skin surface of both macular and plaque lesions is dry, hairless, and anesthetic because of destruction of underlying superficial cutaneous nerves. Larger corresponding cutaneous nerves are thickened in a limited number of cases. On the face, sensory impairment may be difficult to demonstrate because of the generous and bilateral supply of sensory nerve endings. Autonomic nerve damage within the lesion is responsible for surface dryness and loss of sweating over the lesion. A solitary peripheral-nerve trunk in the vicinity of a lesion may be thickened, with sensory loss of the area supplied and with or without motor disfigurement. On slit-skin smear examination, no acid-fast bacilli (AFB) are normally found. The lepromin skin test is strongly positive, signifying good host CMI status.

Borderline Tuberculoid (BT) Leprosy BT leprosy (Fig. 184-4) is characterized by either macular or plaque-type lesions numbering three to nine or more and asymmetrically located on any part of the body, with variable sizes and contours. The margins of the lesions range from poorly defined to well defined; sometimes both forms of margin are seen in one lesion. There may be smaller satellite lesions around a larger one, especially on sides where the margin is less defined; this characteristic indicates downgrading of the lesion from TT to BT leprosy. The edges of plaque lesions may slope outward in contrast to TT lesions, which slope inward; plaques may gradually fade outward and eventually blend into normal-looking skin. Loss of sensation is less intense than it is in TT lesions and dryness on the surface less conspicuous. Several peripheral nerves are likely to be enlarged in an asymmetrical pattern, with sensory and motor deficits. One of the most striking features of BT leprosy is susceptibility to a type 1 leprosy reaction (T1R; see below) that exacerbates skin lesions and/or peripheral nerves. If not diagnosed and treated early, disease in these patients tends to downgrade across the spectrum to BB, BL, or LLs leprosy, with an increasing bacteriologic index and a regressed CMI response causing nerve damage along the way. Slit-skin smears show bacteriologic indices varying from negative to 1+.

Mid-Borderline (BB) Leprosy This form of leprosy is unstable. Many cases downgrade toward BL and LL disease, especially if not treated. There are multiple plaque lesions and, not infrequently, macular lesions; the lesions are of various shapes and sizes, are bilateral, and usually occur in a more or less symmetrical distribution. In annular lesions, the inner edge is well demarcated and "punched out," and the outer edge is ill defined and merges with normal-looking skin. The surface of the lesions is moderately shiny, and the central area looks pale. There is minimal loss of sensation over the lesions. Nerve damage is variable in BB leprosy. Many nerves may be thickened, and this effect may be asymmetrical. BB leprosy is not commonly observed and rapidly changes its spectrum—rarely to BT leprosy but more often to BL disease. The lepromin test is negative. Slit-skin smears of lesions show a moderate number of AFB (2+ to 3+).

Borderline Lepromatous (BL) Leprosy In BL leprosy (Fig. 184-5), there are numerous bilateral, round or oval, macular, diffusely infiltrated, erythematous or hypopigmented lesions with moderately defined borders. The lesions are usually 2–3 cm in diameter, may have a coppery hue, and tend to become symmetrical. Some loss of

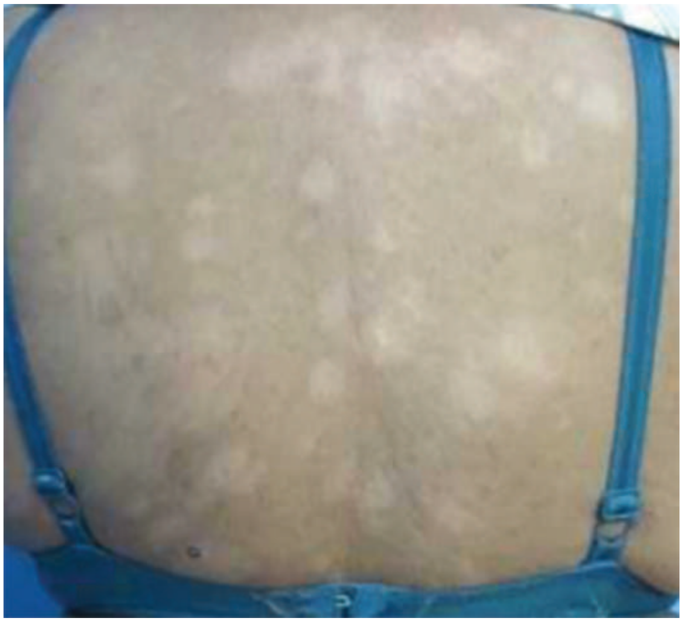

FIGURE 184-5 Borderline lepromatous (BL) leprosy. Numerous diffusely infiltrated erythematous and hypopigmented macules, downgrading from borderline tuberculoid to lepromatous leprosy. *(From Dr. C. L. M. van Hees, Department of Dermatology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands, with permission.)*

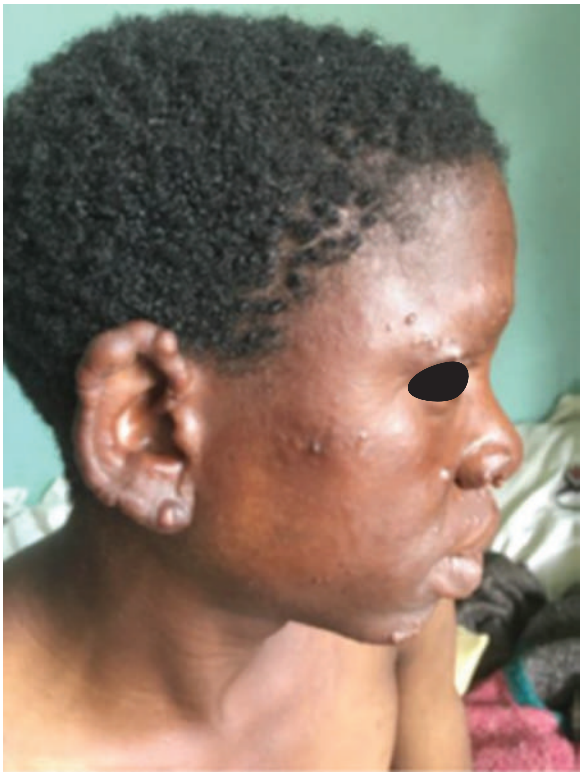

FIGURE 184-6 Lepromatous (LL) leprosy. Multiple nodules on ears and face and loss of eyebrows. *(From Dr. K. Mponda, Department of Dermatology, Queen Elisabeth Central Hospital, Blantyre, Malawi, with permission.)*

sensation may be detected, particularly over older lesions; however, no loss of sensation is observed over fresh lesions. With disease progression, papules, nodules, and plaques develop over the macular lesions. In untreated patients, new ill-defined skin lesions continue to develop. Widespread but asymmetrical thickening of peripheral nerves, with or without tenderness, leads to sensory and motor deficits. The lepromin test gives negative results, as it does in all degrees of lepromatous leprosy. Slit-skin smear examination of lesions shows a bacteriologic index varying from 3+ to 4+.

Lepromatous (LL) Leprosy LL leprosy (Fig. 184-6) presents with innumerable bilateral, symmetrically distributed, diffusely indurated, erythematous, copper-colored or skin-colored patches or plaques. There is no loss of sensation over these lesions, which have a smooth, shiny surface. The lesions spread over the face, earlobes, ears, extensor aspects of the upper and lower extremities, back, and buttocks. Induration can readily be recognized when lesions are viewed tangentially under natural sunlight. The induration initially is of a finer type but gradually becomes coarse, and lesions then progress to papules, plaques, and nodules. Bilateral earlobe thickening and eyebrow loss occur. Coarse induration on the face sometimes results in gross skin folds that lead to an appearance referred to as "lion face," particularly when associated with loss of eyebrows and thickening of earlobes. Of all cases of LL leprosy, 10–15% are of the polar type (LLp) from the time of lesion onset; the remaining cases downgrade from the untreated borderline spectrum to subpolar LLs leprosy. Patients with LLs disease develop nerve damage during the borderline stages. In LLp disease, involvement of peripheral nerves occurs late and is bilateral and symmetrical, with sensory loss in a "glove-and-stocking" distribution. Slit-skin smear examination shows a bacteriologic index of 4+ to 6+ with globi.

SYSTEMIC INVOLVEMENT In LL leprosy, AFB are found in the lymph nodes, spleen, liver, bone marrow, adrenal glands, smooth and striated muscles, tooth pulp, testes, oral cavity, nose, larynx, and eyes. Involvement of the testes leads first to sterility and then to gynecomastia and impotence. Eye involvement includes corneal anesthesia; early on, this manifestation is due to bacillary infiltration of corneal nerves, while later it arises from damage to the ophthalmic division of the trigeminal nerve. In addition, eye involvement includes episcleritis, iridocyclitis, iris atrophy, cataract and glaucoma, lagophthalmos, corneal ulceration and perforation, and blindness. The nose is a portal of entry for *M. leprae* and is the earliest site of involvement in LL leprosy. Edema and mucosal thickening occur in the inferior turbinate and nasal septum, with crusting and epistaxis. Later, patients develop chronic rhinitis with loss of smell sensation. Septal perforation due to bony destruction, with typical saddle-nose disfigurement, is common in advanced LL disease. In late-stage LL leprosy, ulceration of the tongue, pharynx, hard and soft palates (leading to palate perforation), tonsillar pillars, and uvula occurs. In the hands, slow resorption sets in, starting from the distal end of the terminal phalanx and proceeding proximally to involve the middle and proximal phalanges.

HISTOID LEPROSY Histoid leprosy is a rare form of LL leprosy in which waxy, shiny, firm, symmetrical or asymmetrical nodules and plaques are observed over normal-looking skin. Histologic examination of these lesions shows specific spindle-cell granulomas. Slit-skin smear examination reveals high bacteriologic and microbiologic indices without globi in most cases.

DIFFUSE LEPROSY OF LUCIO AND LATAPÍ This rare form of nonnodular LL leprosy occurring in Mexico and Central America is characterized by diffuse shiny infiltration of the skin and widespread sensory loss. The skin looks waxy and has a shiny appearance ("lepra bonita," or beautiful leprosy), with obvious diffuse induration of the earlobes and forehead as well as loss of eyebrows, sometimes eyelashes, and not infrequently all body hair. This form of leprosy can be complicated by an unusual reaction known as Lucio's phenomenon (see below).

Primary Neuritic Leprosy In some countries, such as India and Nepal, primary neuritic disease is observed in 2–10% of all leprosy cases, with only peripheral nerve involvement and no skin lesions. Nerve thickening and sensory loss occur in the affected area, with or without a motor deficit. Primary neuritic leprosy, even though not described by Ridley and Jopling, can manifest at different points along the disease spectrum. For practical purposes, primary neuritic leprosy is classified as paucibacillary or multibacillary on the basis of the absence or presence of AFB in nerve biopsy sections or the number of thickened nerves (single or multiple).

LEPROSY REACTIONS

Leprosy reactions are immunologic phenomena that occur before, during, or after treatment. They are severe complications that need to be diagnosed and treated early to prevent nerve function impairment and subsequent disfigurement as well as blindness.

Type 1 Leprosy Reaction (T1R) T1R is a delayed hypersensitivity reaction associated with sudden alteration of CMI status and leading to a shift in the patient's position on the leprosy spectrum. This reaction is marked by infiltration of lesions by activated CD4+ T lymphocytes, especially T helper cells. T1R is also called a *reversal reaction* because of the upgrading of CMI status. T1R is usually observed in the borderline portion of the spectrum. Skin lesions are characterized by acute swelling and redness (Fig. 184-7). Nerves may be painful and tender because of neuritis, with consequent nerve damage and disfigurement. In the severe form of T1R, nerve abscesses may be formed. Loss of nerve function can be much less obvious than usual when it occurs without other signs of inflammation. This "silent neuritis" may lead to sensory and motor impairment in the hands, feet, and face. Arthralgia or arthritis sometimes occurs. Rarely, the patient may develop fever and malaise, tenosynovitis, and edema of the feet and hands.

Type 2 Leprosy Reaction (T2R) T2R, also known as ENL (erythema nodosum leprosum), is an immune complex–mediated syndrome (i.e., an antigen–antibody reaction involving complement) that causes inflammation of the skin, nerves, and other organs as well as general malaise. ENL is an example of a type III hypersensitivity reaction (Coombs and Gell classification) or Arthus phenomenon.

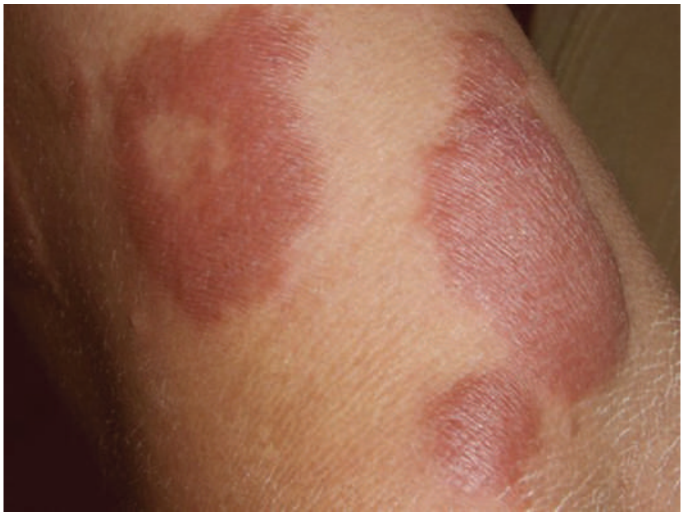

FIGURE 184-7 Type 1 leprosy reaction. Increased inflammation of existing lesions. *(From Dr. W. H. van Brakel, with permission from NLR.)*

This reaction occurs mostly during multidrug therapy but can also develop in untreated patients. Evanescent, pink-to-red, maculopapular, papular, nodular, or plaque lesions suddenly appear and are usually accompanied by constitutional symptoms like malaise and fever, with or without painful swelling in the joints **(Fig. 184-8)**. These crops of skin lesions present on the outer aspects of the thighs, legs, and face. They are painful or tender and warm, blanch with light finger pressure, and last for a few days. The lesions change in color from pink/red to bluish and brownish after 24–48 h and turn dark in a week. Rarely, ENL lesions become vesicular, pustular, bullous, and necrotic and break down to produce ulceration (erythema nodosum necroticans). The patient may have other associated signs such as lymph node enlargement, myositis, arthritis, synovitis, rhinitis, epistaxis, laryngitis, iridocyclitis, glaucoma, painful dactylitis, acute epididymoorchitis, nephritis and renal failure, hepatosplenomegaly, anemia, and—at a later stage—amyloidosis. Severe T2R may include swollen, painful, and tender nerve trunks with sensory and motor deficits.

Lucio's Phenomenon Lucio's phenomenon is observed in diffuse leprosy of Lucio and Latapí and may be a variant of erythema nodosum necroticans. Marked vasculitis and thrombosis of the superficial and deep vessels result in hemorrhage and infarction of the skin. Clinically, the skin reaction begins as slightly indurated, bluish-red, ill-defined, painful, and rarely palpable plaques with an erythematous halo, usually developing on one limb but sometimes on other areas of the body. The lesions are irregular or triangular. After a few days, they become purplish at the center; a central hemorrhagic infarct may develop with or without blister formation, and a necrotic eschar that detaches easily and leaves an ulcer of irregular shape may follow later. The ulcer heals, leaving a superficial scar. Patients remain afebrile throughout.

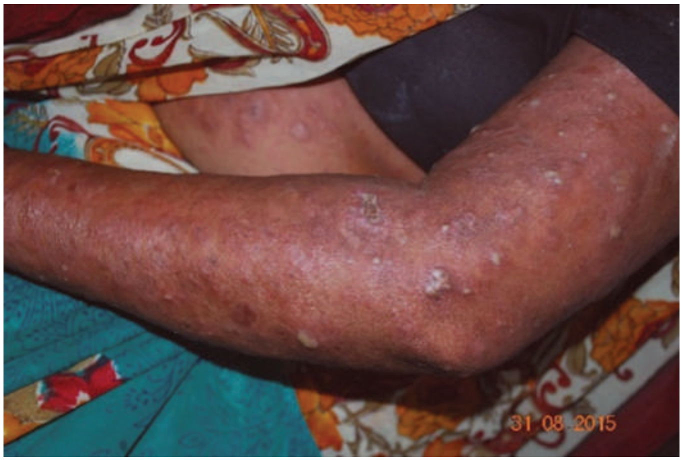

FIGURE 184-8 Type 2 leprosy reaction. Erythema nodosum leprosum, with pustular lesions. *(From Dr. H. K. Kar, with permission.)*

Nerve Function Impairment, Neuritis, and Disfigurement

The terms *nerve function impairment*, *nerve damage*, *neuropathy*, and *neuritis* are often used interchangeably for the sensory, motor, and/or autonomic nerve deficits that occur because of the pathologic processes resulting from *M. leprae* infection of the nerve. Neuritis (nerve inflammation) in leprosy is usually a subacute, demyelinating, and unremitting event involving cutaneous nerves and larger peripheral nerves. "Silent neuritis" or "quiet nerve paralysis" is defined as progressive sensory or motor impairment in the absence of symptoms such as pain, paresthesia, or tenderness of the nerve and with no obvious signs of leprosy reactions. Neuritis can occur at any time during leprosy but is more common and severe during leprosy reactions, mainly in T1R. Sensory and motor neuropathy can lead to secondary impairments in the upper and lower extremities, such as muscle atrophy, mobile- and fixed-joint contractures, bone absorption of digits, and cracks and wounds.

DIAGNOSIS

Clinical Diagnosis Three cardinal signs indicate a diagnosis of leprosy. The diagnosis can be established when at least one of these three signs are present:

1. Hypopigmented or erythematous skin lesion(s) with definite loss or impairment of sensation: The clinical presentation of skin patches or plaques is diagnostic when it is associated with a definite loss or impairment of sensation (light touch, pain, and/or temperature). Diagnostic dilemmas arise in the indeterminate stage of leprosy because of variable loss of sensation and the presence of facial lesions (i.e., because the density of innervation in the face can compensate for damage to certain nerve branches).
2. Involvement of the peripheral nerves, as demonstrated by definite thickening with sensory impairment: Thickening of a peripheral nerve should be assessed by palpation of the affected nerve and comparison with the corresponding contralateral nerve. In multibacillary leprosy, thickening of nerves is often bilateral. Nerve tenderness is established by the application of mild pressure on the nerve during palpation with the fingertips. The peripheral nerves commonly palpated in a leprosy patient are the greater auricular, ulnar, radial, radial cutaneous, median, lateral popliteal, posterior tibial, sural, and superficial peroneal nerves.
3. A positive result for AFB in slit-skin smears, establishment of the presence of AFB in a skin smear or biopsy sample, or a positive result in a biopsy polymerase chain reaction (PCR).

Diagnostic Tools • **TESTING OF SKIN SENSATION** Light-touch sensation is tested with cotton, wool, or a feather. Pain is assessed as the patient's ability to distinguish between the sharp and blunt ends of a wooden or bamboo toothpick. Thermal sensation thresholds are assessed with computer-assisted sensory testing equipment.

SLIT-SKIN SMEAR Normally a slit-skin smear is taken from four sites: the right earlobe, the forehead above the eyebrows, the chin, and the left buttock in men or the left upper thigh in women. The material is stained with Ziehl-Neelsen reagent and examined with a light microscope. The bacteriologic index is determined with a standard logarithmic scale and graded from 0 to 6. The microbiologic index is determined as the percentage of solid, stained AFB.

SKIN BIOPSY A skin biopsy is done to confirm the diagnosis of leprosy, to classify the disease, to support the diagnosis of reactions, and to determine cure after the completion of multidrug therapy. When macular lesions are suspected of reflecting IL, a biopsy sample should be taken from the middle of a lesion; with plaques, a sample should be obtained from the active indurated edge. When there are numerous skin lesions with different morphologies, more than one biopsy sample is required for proper evaluation of the disease spectrum. Identification of early lesions of leprosy by histopathologic techniques is enhanced by immunochemical staining, which reveals the presence of *M. leprae* antigens.

PGL-1 ANTIBODY TEST PGL-1 is a specific lipid on the *M. leprae* cell wall. A PGL-1 enzyme-linked immunosorbent assay (ELISA) has been used for serologic diagnosis of leprosy, yielding positive results in 90–95% of multibacillary cases and in 25–60% of paucibacillary cases. Using PGL-1 antigen and adopting an immunochromatographic technique, a rapid lateral-flow assay—the ML flow test—has been developed for detection of antibody to PGL-1. This assay gives positive results in 92–97% of patients with multibacillary leprosy and in 32–40% of patients with paucibacillary disease. Recently, a quantitative UPC-LFA (upconverting phosphor lateral flow assay) test has become available, which has a higher specificity in paucibacillary disease.

LEPROMIN TEST The lepromin (or Mitsuda) skin test measures cellular immunity against lepromin. A bacillary suspension standardized by the number of inactivated *M. leprae* it contains is injected just under the skin. The reaction to lepromin is measured as induration in millimeters 3–4 weeks after intradermal inoculation. The result provides information about the ability of an individual's T cells to respond to *M. leprae* and the likelihood of granuloma formation in that individual. A negative lepromin test is generally seen in patients with LL or BL leprosy, indicating the lack of a protective cellular response.

GENE AMPLIFICATION (PCR) TECHNIQUE Gene amplification significantly enhances the detection of *M. leprae*, especially in bacteriologic index–negative leprosy and cases that do not fulfill the criteria for the cardinal signs of leprosy. The several PCR methods developed to amplify different gene stretches in *M. leprae* include conventional DNA-based PCR, reverse-transcription PCR, and multiplex PCR. As major genes for detection of disease targets, PCR uses *M. leprae*–specific genes encoding 36-kDa antigen, 18-kDa antigen, 65-kDa antigen complex 85, 16S ribosomal RNA (rRNA), and repetitive sequences. These assays are sensitive to as few as 1–10 bacilli and yield positive results in 60–75% of smear-negative cases. Multiplex PCR employing the genes encoding the repetitive element RLEP, SodA, and 16S rRNA can be used for early diagnosis and for the diagnosis of subclinical infection among household contacts.

Differential Diagnosis Leprosy is often diagnosed late, with a consequent increase in the risk of nerve damage and its ensuing disabilities. The hypopigmented macules of leprosy must be differentiated from a variety of conditions, including pityriasis alba, vitiligo, progressive macular hypomelanosis, pityriasis versicolor, pityriasis rosea, postinflammatory hypopigmentation, sarcoidosis, post–kala-azar dermal leishmaniasis, and morphea. In the analysis of plaques and nodular lesions, conditions such as granuloma annulare, cutaneous sarcoidosis, cutaneous leishmaniasis, lupus miliaris disseminatus faciei, nodular histiocytosis, lupus erythematosus, cutaneous T-cell lymphomas (especially mycosis fungoides), and secondary syphilis should be kept in mind. ENL lesions must be differentiated from erythema nodosum of other etiologies, nodular vasculitis, and cutaneous polyarteritis nodosa. In the case of mononeuropathy lesions, diabetes, amyloidosis, and myxedema must be considered. With polyneuropathy lesions of acute onset, Guillain-Barré syndrome and toxic polyneuropathy must be given consideration.

Diagnostic Tools for Nerve Function Impairment All sensory modalities, autonomic function, and motor function of motor nerves may be affected in leprosy to varying degrees. The modalities mediated by small unmyelinated fibers, such as pain and warm temperature sensation and autonomic function, are often affected first. Clinically detectable impairment of touch sensation and motor function frequently follows after several months. Unfortunately, tools that allow reliable and safe testing of pain and temperature sensation and autonomic function often are not available at peripheral health facilities, but simple and reliable tests of touch sensation and motor function do provide a reflection of the underlying neuropathy.

TOUCH SENSATION TESTING The ulnar and median nerves and the posterior tibial nerve are usually tested for touch sensation. The most reliable test is the Semmes-Weinstein monofilament (SWM) test. If the impairment is of <6 months' duration and/or new nerve function impairment is diagnosed, glucocorticoid treatment should be given. Because filaments are not available in most peripheral health centers, the WHO recommends that a ballpoint pen be used instead. The testing protocol is the same as in the SWM test: the stimulus is delivered by touching the test sites with the tip of a ballpoint pen held at an angle of ~45° relative to the skin.

VOLUNTARY MUSCLE TESTING Motor function of the hands and feet should be evaluated by voluntary muscle testing. The muscle functions most affected in leprosy are eye closure (facial nerve), finger abduction (ulnar nerve), thumb opposition (median nerve), wrist extension (radial nerve), and ankle extension (common peroneal nerve). Strength is assessed with a WHO-recommended system as strong, weak, or paralyzed.

NERVE CONDUCTION TESTS Testing of nerve conduction parameters is sensitive in detecting early signs of peripheral neuropathy in leprosy. Sensory nerve conduction parameters are often affected several months ahead of clinical tests (e.g., the SWM test). However, a trial of glucocorticoid treatment of such early changes did not show improved long-term outcomes, perhaps suggesting that the glucocorticoids are unable to switch off or reverse the pathologic process.

ULTRASOUND TESTING OF NERVES Palpable enlargement of certain peripheral nerves is one of the cardinal signs of leprosy. Definite enlargement is easy to establish, but milder degrees are much harder to diagnose by palpation. Ultrasound imaging and measurement of nerve diameters—even with portable equipment—can detect nerve enlargement accurately. This technique may be used to support the diagnosis of leprosy and may indicate the onset of neuropathy that warrants anti-inflammatory treatment.

OTHER TESTS OF PERIPHERAL NERVE FUNCTION Pain and temperature sensation are commonly affected in leprosy neuropathy. However, these sensations are difficult to test safely and reliably under field conditions. Studies have shown that heat detection thresholds are often affected several months before touch sensation is impaired. Laser Doppler measurement of autonomic vasomotor reflexes is a sensitive method for detection of peripheral autonomic nerve damage in leprosy patients.

TREATMENT

Leprosy, Leprosy Reactions, and Other Major Manifestations

TREATMENT OF LEPROSY

Multidrug Therapy Only one multidrug regimen is recommended by the WHO for the treatment of leprosy. This regimen consists of a combination of two or three of the following drugs: rifampin, dapsone, and clofazimine **(Table 184-1)**. The keystone of WHO-recommended multidrug therapy for multibacillary leprosy is a monthly dose of rifampin together with daily doses of dapsone and daily and monthly doses of clofazimine. Patients with paucibacillary leprosy are treated with two drugs, receiving monthly doses of rifampin and daily doses of dapsone. The treatment duration is 12 months for multibacillary disease and 6 months for paucibacillary disease. Provided that patients complete therapy, treatment failure rates are very low.

Some studies have investigated a uniform regimen of three drugs for 6 months. In a recent systematic review of evidence on the potential benefits and risks of this shorter regimen, the WHO concluded that relevant evidence is limited and inconclusive, with a potential increase in the risk of relapse. Therefore, the WHO does not recommend a shortened treatment duration for multibacillary leprosy.

The WHO further recommends supervised intake, but actual practice varies among countries. Through the WHO, multidrug therapy is provided free of charge as blister packs for adults to all countries reporting leprosy. Blister packs are also provided for

TABLE 184-1 WHO-Recommended Multidrug Treatment for Leprosy

DRUG, AGE GROUP	PAUCIBACILLARY LEPROSY[a]	MULTIBACILLARY LEPROSY[b]
Dapsone		
Adult	100 mg/d	100 mg/d
Child age 10–14 years	50 mg/d	50 mg/d
Child <10 years	Dose adjusted to body weight	Dose adjusted to body weight
Rifampin		
Adult	600 mg monthly	600 mg monthly
Child 10–14 years	450 mg monthly	450 mg monthly
Child <10 years	Dose adjusted to body weight	Dose adjusted to body weight
Clofazimine[c]		
Adult	—	50 mg/d plus 300 mg monthly
Child 10–14 years	—	50 mg/d plus 150 mg monthly
Child <10 years	—	Dose adjusted to body weight

[a] Duration: 6 doses (6 blister packs). [b] Duration: 12 doses (12 blister packs). [c] In 2018, the World Health Organization (WHO) suggested including clofazimine in the multidrug therapy regimen for paucibacillary leprosy as well, but it is questionable whether this suggestion will be implemented because of the possibility that skin discoloration might compromise compliance. In addition, this alteration would involve a major change in the production of blister packs, which currently do not include clofazimine for paucibacillary leprosy patients (in line with the original WHO recommendation).

10- to 14-year-olds, while younger children are given doses adjusted according to body weight (Table 184-1).

***Adverse Events* • Rifampin** Rifampin acts by inhibiting DNA-dependent RNA polymerase, thereby interfering with bacterial RNA synthesis. Rifampin is well absorbed orally. Hepatotoxicity may occur with a mild transient elevation of hepatic aminotransferases, but this reaction is rare at the dosages and intervals recommended for leprosy and is not an indication for discontinuation of treatment. Because rifampin is given only monthly in WHO-recommended multidrug therapy regimens, the adverse effects recognized from its use in tuberculosis probably do not occur. A monthly dose of rifampin does not cause induction of hepatic cytochrome p450. Urine discoloration occurs but is harmless.

Dapsone Dapsone (4,4-diaminodiphenyl sulfone [DDS]) acts by blocking folic acid synthesis and is only weakly bactericidal. Oral absorption is good, and the drug has a long half-life averaging 28 h. Dapsone has a poor safety profile, and its use should be monitored carefully. In the doses recommended for leprosy, it can cause mild hemolysis and may cause anemia or, rarely, psychosis. Glucose-6-phosphate dehydrogenase deficiency seldom causes a problem, and enzyme levels are not routinely tested before the start of multidrug treatment. On the other hand, the "DDS syndrome" (also called the *dapsone hypersensitivity syndrome*) is a severe adverse event that is not uncommon in some countries. It usually develops 6 weeks after the commencement of dapsone administration and manifests as fever, skin rash, eosinophilia, lymphadenopathy, hepatitis, and encephalopathy. Other rare but severe cutaneous adverse reactions are erythema multiforme, Stevens-Johnson syndrome, toxic epidermal necrolysis, and exfoliative dermatitis. The fatality rate for DDS syndrome is 10%, with death occurring from liver failure, sepsis, and bone marrow failure. Most patients require treatment with systemic glucocorticoids. In all cases, dapsone treatment must be stopped. Agranulocytosis, hepatitis, and cholestatic jaundice occur rarely with dapsone therapy.

Clofazimine Clofazimine is a brick-red, fat-soluble crystalline dye. The mechanism of its weakly bactericidal action against *M. leprae* is not known. High drug concentrations are found in the intestinal mucosa, mesenteric lymph nodes, and body fat. The most noticeable adverse event is skin discoloration ranging from red to purple or black, with the degree of discoloration depending on the dosage. Clofazimine can accumulate in active leprosy skin lesions, thus making them more prominent. The abnormal pigmentation usually fades within 6–12 months of clofazimine discontinuation, although traces of discoloration may remain for up to 4 years. The skin discoloration associated with clofazimine is psychologically distressing for many people. Patients often stop taking the drug because the discoloration is socially disabling for them, alerting their social environment to the fact that they are taking anti-leprosy medication and thus breaking confidentiality about treatment. Urine, sputum, and sweat may become pink during clofazimine administration. Clofazimine also produces a characteristic ichthyosis on the shins and forearms. Adverse gastrointestinal events ranging from mild cramps to diarrhea and weight loss may result from clofazimine crystal deposition in the wall of the small bowel.

Relapse The cure rate for leprosy with multidrug therapy is 99%, but relapse is possible. In multibacillary leprosy, relapse is defined as the multiplication of *M. leprae*, with an increase of at least 2+ over the previous value in the bacteriologic index at any single site; this change usually occurs in conjunction with evidence of clinical deterioration (e.g., new skin patches or nodules and/or new nerve damage). Relapse rates are well below 1% except among a small proportion of patients who have a very high bacillary load at the start of treatment (bacteriologic index ≥4). In different studies, four to seven relapses were recorded per 100 person-years. These relapses usually occurred <5 years after the end of multidrug therapy. Since antimicrobial resistance to the combination of drugs used in multidrug treatment is rare, patients with relapse can be re-treated with the same multibacillary regimen.

Recognizing a relapse in paucibacillary leprosy can be difficult, as symptoms may resemble T1R. However, relapse of paucibacillary disease is very rare. Administration of a therapeutic trial with glucocorticoids to patients with new lesions may help distinguish between these two phenomena: a definite improvement within 4 weeks of initiation of glucocorticoid therapy indicates T1R, whereas a lack of response favors the diagnosis of a clinical relapse. Patients with multibacillary disease who present with a relapse are re-treated with the multidrug regimen regardless of any change in classification. Patients with paucibacillary disease require 2 years of monitoring after treatment and patients with multibacillary disease at least 5 years. Reinfection by different strains of *M. leprae* is possible and can be confused with relapse.

Rifampin Resistance and Second-Line Drugs Resistance to rifampin has been reported from several countries, although the number of patients involved is small. Evidence on the potential benefits and risks of using alternative regimens for drug-resistant leprosy is not available. Therefore, recommendations provided by the WHO for second-line regimens are based on expert opinion and the known activity of alternative drugs, including the likelihood of cross-resistance. For rifampin-resistant leprosy, the WHO guidelines recommend daily treatment with at least two second-line drugs—clarithromycin, minocycline, or a quinolone (ofloxacin, levofloxacin, or moxifloxacin)—plus clofazimine for 6 months, followed by clofazimine plus one of the second-line drugs daily for an additional 18 months. Leprosy patients infected with *M. leprae* resistant to both rifampin and ofloxacin may be treated daily with the following regimen: clarithromycin, minocycline, and clofazimine for 6 months, followed by clarithromycin or minocycline plus clofazimine for an additional 18 months.

TREATMENT OF LEPROSY REACTIONS

Type 1 Reactions Oral, short-acting glucocorticoids are the treatment of choice for T1R. Prednisolone is used most often in an initial dose of 1 mg/kg of body weight once a day, usually with a maximum of 60–80 mg. If standard treatment protocols are followed, as they are in most leprosy programs in endemic countries, an initial dose of 40 mg of prednisolone is recommended by the WHO. The dose is tapered slowly, usually by 5 mg every 2 weeks over a period of 20 weeks—a schedule that results in better outcomes and lower reaction relapse rates than the previously recommended 12-week glucocorticoid regimen. However, the clinical response should guide treatment. Patients should be examined every 2 weeks, and the examination should include a quick nerve function assessment. Not infrequently, the reaction flares up again once the daily glucocorticoid dose is tapered to <10–20 mg. The potential benefits of longer treatment should be balanced against the risks of prolonged glucocorticoid use, especially at higher doses.

Type 2 Reactions Mild first-time T2R (or ENL) reactions with localized skin nodules may be treated with aspirin and pentoxifylline. If a rapid effect is needed, the most effective drug to date is thalidomide, which rapidly suppresses clinical signs, including nerve impairment and iritis. However, the drug is blacklisted in many countries because of its teratogenicity. If available, it should be given with great caution to women of childbearing age—only after careful counseling and a negative pregnancy test and with strict adherence to contraception. A dose of 100–200 mg is given either once or twice daily. In acute first episodes, thalidomide treatment should be tapered down and stopped after 1–2 weeks. If tissues other than the skin are affected—e.g., the eyes (iritis/uveitis), testes (orchitis), kidneys (nephritis), or joints (arthritis)—longer treatment may be needed until signs and symptoms have resolved. In patients with severe recurrent ENL, a daily thalidomide maintenance dose of 50 mg may be effective in suppressing new episodes. Because of the restricted availability and use of thalidomide, patients with acute ENL are usually treated with glucocorticoids. T2R tends to be transient, often resolving in ~2 weeks. The treatment strategy is therefore to suppress the acute signs and symptoms with high-dose oral prednisolone, quickly tapering treatment in 2–3 weeks either to zero or to a low maintenance dose if the patient has had previous attacks. High-dose clofazimine also is effective in preventing recurrent ENL, but attainment of a maximal effect takes several weeks. The usual regimen is 300 mg daily for 1 month, followed by 200 mg daily for 1 month and, subsequently, 100 mg daily as a maintenance dose for as long as necessary. Prolonged use of high-dose clofazimine may cause significant adverse gastrointestinal effects. An important side effect of clofazimine is a dark discoloration of the skin. While discoloration resolves gradually after the drug is discontinued, it is one main reason that patients dislike or even refuse to take clofazimine.

TREATMENT AND PROGNOSIS OF NERVE FUNCTION IMPAIRMENT

Episodes of sensory or motor nerve function impairment without skin signs are common. Neuropathy may occur without obvious neuritis. Still, the treatment of such "silent neuropathy" is the same as that for T1R. High-dose prednisolone is the drug of choice. Some experts think that patients will benefit from nerve decompression surgery, but evidence from randomized controlled trials is lacking.

If glucocorticoid treatment is started shortly after the development of nerve function impairment, the prognosis for full recovery is good. Generally, some recovery can still be expected up to 6 months after onset, but the likelihood of recovery diminishes with every new episode. Generally, nerve function impairment that has persisted for >6 months does not benefit from glucocorticoid treatment.

TREATMENT OF (NEUROPATHIC) PAIN

Pain is common in people affected by leprosy and is often of neuropathic origin. Little evidence-based information is currently available on the origin and treatment of pain in leprosy. Generally, for the treatment of neuropathic pain, three classes of medication are available: tricyclic antidepressants, phenothiazines, and anticonvulsants (carbamazepine, oxcarbazepine, gabapentin, and pregabalin). These agents can be combined with analgesics and anti-inflammatory drugs according to the patient's needs.

DISEASE MANAGEMENT DURING TREATMENT

Leprosy can be cured effectively, but the long duration of multidrug therapy means that careful management is needed to help the patient complete treatment. Regular visits to a health center may invoke questions from community members that may threaten the patient's privacy, thus causing the patient mental distress and jeopardizing treatment adherence. Counseling is essential, as are patient-friendly arrangements for collecting treatment drugs. The disease, its treatment, and its possible complications should be discussed, including a consideration of disease prognosis, the resolution of skin patches, skin discoloration by clofazimine, the lack of contagiousness during multidrug therapy, and the capacity for unrestricted family relations, including marital life and sexual activity. Possible stigmatization, including self-stigmatization, also should be discussed.

Because of the diverse complications that are possible, especially in patients with multibacillary leprosy, a multidisciplinary approach to patient management is required. In low-income countries, the responsibility for treatment usually lies with a leprosy control officer or a general medical practitioner. In middle- and high-income countries, the main treatment responsibility usually falls to a dermatologist. Additional support should come from a neurologist or neurophysiologist for the diagnosis of nerve function impairment, and a rehabilitation physician, physiotherapist, infectious disease specialist, and/or psychologist may be needed. Occasionally, specialist support with regard to orthotics as well as in ophthalmology, occupational therapy, reconstructive surgery, and/or community-based rehabilitation is indicated.

Supervised Multidrug Therapy Regular treatment is important, especially the supervised 4-weekly dose of rifampin and clofazimine. However, treatment adherence can be facilitated by flexible arrangements; for example, patients can be allowed to take home more than one 4-week blister pack if they will be away for travel or seasonal labor. In such cases, a family member or another responsible person can be asked to supervise the monthly dose.

Monthly Nerve Function Assessment Since nerve damage can be insidious and silent, it is important to conduct a brief nerve function assessment at each clinic visit during multidrug therapy. This regular assessment is especially important in patients with known risk factors for nerve function impairment. At highest risk are patients with multibacillary disease, who already have nerve damage at the start of treatment. Their risk of additional nerve damage is as high as 65%. Multibacillary leprosy patients without nerve function impairment at diagnosis and paucibacillary leprosy patients *with* such impairment at diagnosis have a 16% chance of developing damage and additional damage, respectively. Patients with paucibacillary disease who do not have nerve function impairment at diagnosis are at lowest risk (3%); for them, an assessment at the start and completion of multidrug therapy can be sufficient. Leprosy reactions and new nerve damage may also occur after completion of multidrug treatment. While the risk diminishes with time, these manifestations can occur up to 3 years after the conclusion of therapy.

Health Education During treatment, patients will have questions that need to be addressed in order to ensure their treatment adherence. Sensitive questions may arise regarding everyday life within the family and at work that, if not addressed properly, could lead to social withdrawal and mental health issues. Crucial points for health education are at diagnosis and at completion of treatment. When communicating the diagnosis, the physician must explain that the disease is caused by a curable microbial infection and must cover the possible discomforts of drug intake, the interruption of disease transmission through drug intake, and the importance of adhering to treatment to achieve a cure. At the completion of multidrug therapy, the emphasis should be on separating the concept of cure (bacterial activity) from the sequelae of the disease (nerve function impairment, leprosy reactions, and disabilities) and explaining that the patient may need to continue receiving health care, including reconstructive surgery, for the sequelae. Patients often associate cure with the absence of symptoms, which is not accurate in leprosy. Some patients will experience discomfort during bacterial activity but will have no sequelae after treatment. In others, nerve function impairment or leprosy reactions may cause disfigurement with physical discomfort after cure. These sequelae will need further management and patients at risk should be warned that such posttreatment complications may occur and to report without delay should this happen. Disabilities such as claw hand or neuropathic foot require chronic care.

Guidelines After the Completion of Multidrug Therapy Patients should receive counseling at release from treatment. The topics covered should include reassurance that the person is no longer contagious, that in some patients hypopigmentation in skin lesions may not resolve for a long time, and that skin discoloration due to clofazimine will gradually disappear in the following months. Nerve impairment may continue to improve after release from treatment, but this is by no means certain. Most important, patients should be instructed to return to the clinic if any new skin signs or fresh nerve damage occurs. This situation is not uncommon, is usually due to a leprosy reaction, and should be managed carefully from both a medical and a social perspective, since patients and persons in their environment will interpret this development as "leprosy coming back." Patients at risk of further episodes of reaction and/or additional nerve function impairment (e.g., patients with preexisting nerve function impairment and multibacillary infection or patients who have experienced a reactional episode during therapy) should be asked to return for a check-up every 6 months for at least 3 years after being released from treatment.

REHABILITATION AND SOCIAL ASPECTS

Physical Rehabilitation Peripheral neuropathy and its secondary disabling consequences often require physical rehabilitation. This effort may include reconstructive surgery in the case of facial, ulnar, median, or posterior tibial paralysis. In this case, pre- and postoperative physical therapy is of crucial importance. Physical therapy is also indicated when muscles are not completely paralyzed or when contractures are too stiff to allow surgery. Since paralysis is usually accompanied by sensory and autonomic neuropathy, occupational therapy also is helpful; therapists teach patients how to minimize the risk of further injury and other techniques for prevention of disabilities. The key principle is teaching patients and former patients to self-manage their disabilities. In many programs, this teaching occurs in the setting of self-care groups. A well-tested and evidence-based self-care routine for hands and feet consists of inspection, soaking, scraping, and oiling (ISSO). Specifically, in ISSO, the person inspects the affected limbs for hotspots (evidence of too much stress on an area of skin): wounds, cracks, and calluses. Next, the affected limb(s) are soaked in plain water for 15 minutes. While the skin is wet, areas with excess calluses are scraped with a rough stone or another rough object. The skin is then rubbed with petroleum jelly or another nonfragrant oil in order to trap moisture in the skin. If this routine is performed daily, the skin can be kept supple and in good condition, despite sensory and autonomic damage. If sensation on the soles of the feet is impaired, the person must wear protective footwear. Simple footwear (e.g., sandals or sneakers) available at the local market is adequate as long as it has a strong sole and a soft insole of ethylene-vinyl acetate or microcellular rubber that distributes pressure—an especially important feature when foot muscles are weak or paralyzed or the architecture of the foot is damaged, as is often the case with neuropathy. In high-resource settings, tailor-made orthopedic shoes can be provided.

Mental and Social Support Like other chronic health conditions, leprosy requires patients to cope with the burden of new routines in everyday life. In addition to coping with stigma, they must organize themselves for prolonged treatment, prevention of disabilities, and rehabilitation activities. Moreover, like other neglected tropical diseases (see "Neglected Tropical Diseases," below), leprosy may lead to poor mental health. Such diseases are accompanied by social exclusion in the form of poor access to services such as health care, education, employment, and housing. This exclusion accounts for common mental health comorbidities in leprosy patients and their family members, including depression, anxiety, and suicidal thoughts. Leprosy is probably the most notorious of all stigmatized health conditions, and social stigmatization is the most common issue that triggers mental suffering. Other infectious diseases that raise this issue include HIV infection, tuberculosis, and neglected tropical diseases like lymphatic filariasis, Buruli ulcer, and dermal leishmaniasis. Although the reasons for stigmatization vary, the manifestations and interventions that effectively reduce stigma are similar across conditions and countries. Therefore, joint interventions addressing health-related stigmas for multiple conditions would be strategically and financially attractive. The need to introduce mental health care in leprosy services is pressing. Therapeutic group meetings among institutionalized patients and self-care groups at the community level, with a focus on prevention of disabilities and mental well-being, are known to ameliorate depression, encourage self-acceptance, and promote confidence.

NEGLECTED TROPICAL DISEASES Leprosy is one of a medically diverse group of 20 neglected tropical diseases (NTDs). This group includes infectious diseases caused by bacteria, viruses, fungi, and parasites as well as some noninfectious conditions, such as podoconiosis and snakebite. NTDs have been grouped together because they affect 1.5 billion of the poorest people on Earth and have been widely neglected in domains such as public policy, funding, and the development of diagnostics and treatments. Leprosy is the archetypical NTD, featuring all of the common characteristics: a treatable infectious disease, a known population at risk, available preventive chemotherapy, disease complications that may lead to severe disabilities, and a pervasive social stigma that leads to discrimination, social exclusion, and severe mental health consequences. Nevertheless, the priority accorded to leprosy on the public health agenda of most endemic countries is very low. By joining hands in advocacy, fundraising, and development of joint control strategies, health care organizations can substantially raise the priority profile of NTDs, benefiting each of the individual disease control programs. Such a joint approach serves the goal of universal health coverage and helps to strengthen health services more effectively than vertical programs are ever able to do on their own.

PREVENTION AND CONTROL

Interruption of Transmission and Novel Preventive Strategies Leprosy control was traditionally based on early case detection and multidrug treatment. Apart from health education and leprosy awareness campaigns, no preventive measures were available. In the 1990s, authorities hoped that the transmission of *M. leprae* in the community could be interrupted through timely detection of cases and provision of multidrug therapy, leading to a decline in leprosy incidence. Unfortunately, this has not been the case (Fig. 184-1). The inability to reduce leprosy incidence in many countries and the heightened interest in NTDs have invigorated research into new techniques for the diagnosis of disease and infection, leprosy vaccines, enhanced postexposure chemoprophylaxis regimens, epidemiologic tools (e.g., geographic information systems for identifying leprosy hotspots), surveillance of antimicrobial resistance, and alternative drugs and drug treatment regimens.

Vaccines Against Leprosy The bacille Calmette-Guérin (BCG) vaccine used against tuberculosis provides varying degrees of protection against leprosy and is used routinely as postexposure immunoprophylaxis for contacts of leprosy patients in Brazil. Two promising vaccine candidates are in the pipeline: the MIP vaccine from India, which is based on killed *Mycobacterium indicus pranii,* and the synthetic LepVax vaccine developed by the University of Washington's Infectious Disease Research Institute in the United States. If proven effective, these vaccines, like the BCG vaccine, will be used as postexposure prophylaxis for contacts of leprosy patients. Trials are in early stages, and sufficient proof of efficacy will take years.

Postexposure Chemoprophylaxis The introduction of postexposure chemoprophylaxis (PEP) for household and other close contacts of leprosy patients is an important innovation. A large randomized controlled trial has shown that single-dose rifampin, given once to household contacts, neighbors, and social contacts, reduces the recipients' risk of leprosy by ~60%. Implementation studies have shown that PEP with single-dose rifampin is feasible and well accepted by patients, contacts, and health workers in a variety of health care settings. Furthermore, modeling studies have indicated the potential impact of PEP on transmission of *M. leprae* in endemic populations. This intervention was included in the 2018 WHO Guidelines for the Diagnosis, Treatment, and Prevention of Leprosy and is currently being introduced in many countries. Research is ongoing into enhanced PEP regimens for those close contacts who are at increased risk of leprosy (e.g., blood-related household contacts and close contacts of multibacillary leprosy patients).

"Zero Leprosy" The WHO has formulated its new Global Leprosy Strategy 2021–2030. As in the organization's previous strategy, a holistic approach to leprosy control is advocated, focusing on zero infection and disease, zero disability, and zero stigma and discrimination. For 2030, the WHO is setting ambitious targets of achieving 120 countries with zero new autochthonous leprosy cases, reducing the annual number of new cases detected by 70%, reducing the rate of new cases with grade 2 disability per million population (as a proxy for detection delay) by 90%, and reducing the rate of new child cases with leprosy per million children (as a proxy for recent transmission) by 90%. Widespread implementation of PEP with single-dose rifampin is one of the key strategies to achieve these goals. The "Triple Zero Strategy" (*zeroleprosy.org*) has also been embraced by the partners united in the Global Partnership for Zero Leprosy, the International Federation of Anti-Leprosy Associations, the Novartis Foundation, the Sasakawa Health Foundation, and the International Association for Integration, Dignity, and Economic Advancement.

In July 2023, the WHO launched new "Technical Guidance on Interruption of Transmission and Elimination of Leprosy Disease" (*www.who.int/publications/i/item/9789290210467*). This contains clear milestones, definitions, and cutoffs for interruption of transmission and elimination of leprosy disease. Using the Leprosy Elimination Framework and the accompanying tools—the Leprosy Elimination Monitoring Tool and Leprosy Programme and Transmission Assessment—countries can track their progress toward these milestones in detail.

The outlook for achieving "zero leprosy" is better than ever before, but this goal is admittedly very ambitious. It can be reached only when all leprosy-endemic countries enhance their leprosy control activities to include (1) active case-finding strategies, including improved diagnosis; (2) contact screening; (3) implementation of PEP; (4) improved prevention of disability services; and (5) activities to reduce stigma and discrimination and to promote the social inclusion and mental well-being of affected patients and their families. Coincident with these efforts, an important threat must be confronted. With the waning of interest in leprosy and the integration of management of the disease into nonspecialized health systems, the number of medical doctors and health workers at the primary care level who have experience in diagnosing and treating leprosy has decreased substantially all over the world. Once lost, expertise is difficult to regain. Therefore, new energy and resources need to be invested in bolstering technical capacity for all aspects of leprosy services, with a view to strengthening the health system in an integrated way and leaving no one behind.

Acknowledgment

We thank Dr. Colette L.M. van Hees, dermatologist at Erasmus MC, University Medical Center Rotterdam, for critical review of this chapter.

FURTHER READING

Bratschi MW et al: Current knowledge on *Mycobacterium leprae* transmission: A systematic literature review. Lepr Rev 86:142, 2015.

Collin SM et al: Systematic review of Hansen disease attributed to *Mycobacterium lepromatosis*. Emerg Infect Dis 29:1376, 2023.

Fróes LAR Jr et al: Bacterial, fungal and parasitic co-infections in leprosy: A scoping review. PLoS Negl Trop Dis 17:e0011334, 2023.

Kumar B, Kar HK (eds): *IAL Textbook of Leprosy*, 2nd ed. New Delhi, Jaypee Brothers Medical Publishers (P) Ltd, 2017.

Scollard DM, Gillis TP (eds): *International Textbook of Leprosy*. Available at *https://internationaltextbookofleprosy.org*. Accessed February 17, 2024.

Smith WC et al: The missing millions: A threat to the elimination of leprosy. PLoS Negl Trop Dis 9:e0003658, 2015.

World Health Organization: Guidelines for the diagnosis, treatment and prevention of leprosy. New Delhi, WHO Regional Office for South-East Asia, 2018. Available at *https://apps.who.int/iris/handle/10665/274127*. Accessed February 17, 2024.

185 Nontuberculous Mycobacterial Infections

Steven M. Holland

Several terms—nontuberculous mycobacteria (NTM), atypical mycobacteria, mycobacteria other than tuberculosis, and environmental mycobacteria—all refer to mycobacteria other than *Mycobacterium tuberculosis*, its close relatives (*M. bovis*, *M. caprae*, *M. africanum*, *M. pinnipedii*, *M. canetti*), and *M. leprae*. The number of identified species of NTM is growing and will continue to do so because of the use of DNA sequence typing for speciation. The number of known species currently exceeds 199. NTM are highly adaptable and can inhabit hostile environments, including industrial solvents.

EPIDEMIOLOGY

NTM are ubiquitous in soil and water. Specific organisms have recurring niches, such as *M. simiae* in certain aquifers, *M. fortuitum* in pedicure baths, and *M. immunogenum* in metalworking fluids. Most NTM cause disease in humans only rarely unless some aspect of host defense is impaired, as in bronchiectasis, or breached, as by inoculation (e.g., liposuction, trauma, cardiac surgery). There are few instances of human-to-human transmission of NTM, which occurs almost exclusively in cystic fibrosis. Because infections due to NTM are rarely reported to health agencies and because their identification is sometimes problematic, reliable data on incidence and prevalence are lacking. Disseminated disease denotes significant immune dysfunction (e.g., advanced HIV infection), whereas pulmonary disease, which is much more common, is highly associated with pulmonary epithelial defects but not with systemic immunodeficiency.

In the United States, the incidence and prevalence of pulmonary infection with NTM, mostly in association with bronchiectasis **(Chap. 301)**, have for many years been severalfold higher than the corresponding figures for tuberculosis, and rates of the former are increasing among the elderly as rates of tuberculosis continue to fall. Among patients with cystic fibrosis, who often have bronchiectasis, rates of clinical infection with NTM range from 3 to 15%, with even higher rates among older patients. Although NTM may be recovered from the sputa of many individuals, it is critical to differentiate active disease from commensal harboring of the organisms. A scheme to help with the proper diagnosis of pulmonary infection caused by NTM has been developed by the American Thoracic Society and is widely used (*https://doi.org/10.1093/cid/ciaa241*). The bulk of nontuberculous mycobacterial disease in North America is due to *M. kansasii*, organisms of the *M. avium* complex (MAC), and organisms in the *M. abscessus* complex.

In Europe, Asia, and Australia, the distribution of NTM in clinical specimens is roughly similar to that in North America, with MAC species and rapidly growing organisms such as *M. abscessus* encountered frequently. *M. xenopi* and *M. malmoense* are especially prominent in northern Europe. *M. ulcerans* causes the distinct clinical entity Buruli ulcer, which occurs throughout tropical zones, especially in western Africa.

M. marinum is a common cause of cutaneous and tendon infections in coastal regions and among individuals exposed to fish tanks or swimming pools.

The true international epidemiology of infections due to NTM is hard to determine because the isolation of these organisms often is not reported and speciation often is not performed for *M. tuberculosis* or NTM. The latter issue poses an especially important problem during therapy for tuberculosis when smears positive for acid-fast bacilli are considered evidence of treatment failure. The increasing ease of identification and speciation of these organisms is already having a major impact on the description of the dynamic international epidemiology of tuberculosis and NTM infections.

PATHOBIOLOGY

Because exposure to NTM is essentially universal and disease is rare, it can be assumed that normal host defenses against these organisms must be strong and that otherwise healthy individuals in whom significant disease develops are highly likely to have specific susceptibility factors that permit NTM to become established, multiply, and cause disease. At the advent of HIV infection, CD4+ T lymphocytes were recognized as key effector cells against NTM; the development of disseminated MAC disease was highly correlated with a decline in CD4+ T lymphocyte numbers. Such a decrease has also been implicated in disseminated MAC infection in patients with idiopathic CD4+ T lymphocytopenia. Potent inhibitors of tumor necrosis factor α (TNF-α), such as infliximab, adalimumab, certolizumab, golimumab, and etanercept, neutralize this critical cytokine, with consequent inhibition of granuloma formation. The occasional result is severe mycobacterial or fungal infection; these associations indicate that TNF-α is a crucial element in mycobacterial control. However, in cases without the above risk factors, much of the basis of susceptibility to disseminated infection with NTM is accounted for by specific mutations in the interferon γ (IFN-γ)/interleukin 12 (IL-12) synthesis and response pathways or autoantibodies to IFN-γ itself.

Mycobacteria are typically phagocytosed by macrophages, which respond with the production of IL-12, a heterodimer composed of IL-12p35 and IL-12p40 moieties that together make up IL-12p70. IL-12 activates T lymphocytes and natural killer cells through binding to its receptor (composed of IL-12Rβ1 and IL-12Rβ2/IL-23R), with consequent phosphorylation of STAT4. IL-12 stimulation of STAT4 leads to secretion of IFN-γ, which activates neutrophils and macrophages to produce reactive oxidants, to increase expression of the major histocompatibility complex and Fc receptors, and to concentrate certain antibiotics intracellularly. Signaling by IFN-γ through its receptor (composed of IFN-γR1 and IFN-γR2) leads to phosphorylation of STAT1, which in turn regulates IFN-γ-responsive genes, such as those coding for IL-12 and TNF-α. TNF-α signals through its own receptor via a downstream complex containing the nuclear factor-κB (NF-κB) essential modulator (NEMO). Therefore, the positive feedback loop between IFN-γ and IL-12/IL-23 drives the immune response to mycobacteria and other intracellular infections. These genes are known to be the critical ones in the pathway of mycobacterial control: specific Mendelian mutations have been identified in *IFNG*, *IFNGR1*, *IFNGR2*, *STAT1*, *GATA2*, *ISG15*, *IRF8*, *IL-12A*, *IL-12RB1*, *IL-12RB2*, *CYBB* (which encodes the gp91*phox* protein of the NADPH oxidase), *SPP2A*, *MCTS1*, and *IKBKG* (which encodes NEMO) **(Fig. 185-1)**. Despite the identification of genes associated with disseminated disease, only ~70% of cases of disseminated nontuberculous mycobacterial infections that are not associated with HIV infection have a genetic diagnosis; the implication is that more mycobacterial susceptibility genes and pathways remain to be identified.

In contrast to the recognized genes and mechanisms associated with disseminated nontuberculous mycobacterial infection, the best-recognized underlying condition for pulmonary infection with NTM is bronchiectasis **(Chap. 301)**. Most of the well-characterized forms of bronchiectasis, including cystic fibrosis, primary ciliary dyskinesia, STAT3-dominant negative hyper-IgE syndrome (Job's syndrome), and idiopathic bronchiectasis, have high rates of association with nontuberculous mycobacterial infection. The precise mechanism by which bronchiectasis predisposes to locally destructive but not systemic involvement is unknown.

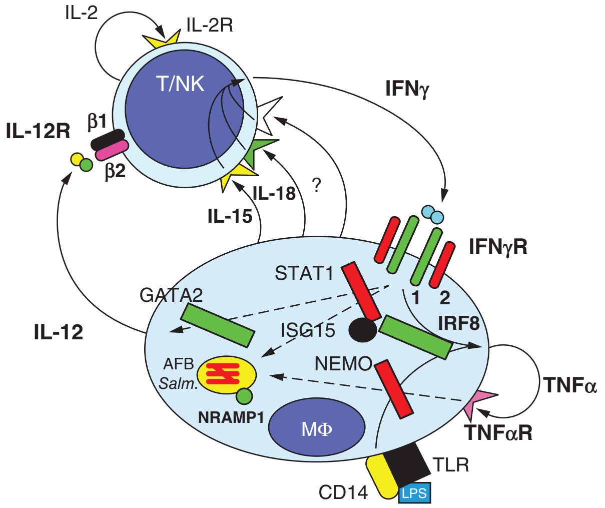

FIGURE 185-1 Cytokine interactions of infected macrophages (MΦ) with T and natural killer (NK) lymphocytes. Infection of macrophages by mycobacteria (AFB) leads to the release of heterodimeric interleukin 12 (IL-12). IL-12 acts on its receptor complex (IL-12R), with consequent STAT4 activation and production of homodimeric interferon γ (IFNγ). Through its receptor (IFNγR), IFNγ activates STAT1, stimulating the production of tumor necrosis factor α (TNFα) and leading to the killing of intracellular organisms such as mycobacteria, salmonellae (*Salm.*), and some fungi. Homotrimeric TNFα acts through its receptor (TNFαR) and requires nuclear factor-κB essential modulator (NEMO) to activate nuclear factor-κB, which also contributes to the killing of intracellular bacteria. Both IFNγ and TNFα lead to upregulation of IL-12. TNFα-blocking antibodies work either by blocking the ligand (infliximab, adalimumab, certolizumab, golimumab) or by providing soluble receptor (etanercept). Mutations in *IFNG*, *IFNGR1*, *IFNGR2*, *IL12B*, *IL12RB1*, *IL12RB2*, *STAT1*, *GATA2*, *ISG15*, *IRF8*, *CYBB*, *MCTS1*, and *IKBKG* (NEMO) have been associated with predisposition to mycobacterial infections. Other cytokines, such as IL-15 and IL-18, also contribute to IFNγ production. Signaling through the Toll-like receptor (TLR) complex and CD14 also upregulates TNFα production. IRF8, interferon regulatory factor 8; ISG15, interferon-stimulated gene 15; LPS, lipopolysaccharide; NRAMP1, natural resistance-associated macrophage protein 1.

Unlike disseminated or pulmonary infection, "hot-tub lung" represents pulmonary hypersensitivity to NTM—most commonly MAC organisms—growing in underchlorinated water, often in indoor hot tubs.

CLINICAL MANIFESTATIONS

Disseminated Disease Disseminated MAC or *M. kansasii* infections in people with advanced HIV infection are now uncommon in North America because of effective antimycobacterial prophylaxis and improved treatment of HIV infection. When such mycobacterial disease was common, the portal of entry was the bowel, with spread to bone marrow and the bloodstream. Surprisingly, disseminated infections with rapidly growing NTM (e.g., *M. abscessus*, *M. fortuitum*) are very rare in people with advanced HIV infection. Because these organisms are of low intrinsic virulence and disseminate only in conjunction with impaired immunity, disseminated disease can be indolent and progressive over weeks to months. Typical manifestations of malaise, fever, and weight loss are often accompanied by organomegaly, lymphadenopathy, and anemia. Because special cultures or stains are required to identify the organisms, the most critical step in diagnosis is to suspect infection with NTM. Blood cultures may be negative, but involved organs typically have significant organism burdens, sometimes with a grossly impaired granulomatous response.

Disseminated involvement (i.e., involvement of two or more organs) without an underlying iatrogenic cause should prompt a genetic investigation of the IFN-γ/IL-12 pathway. Recessive mutations in *IFNGR1* and *IFNGR2* typically ablate IFN-γ signaling and lead to severe infection with NTM. In contrast, dominant negative mutations in *IFNGR1*, which lead to overaccumulation of a defective interfering mutant receptor on the cell surface, inhibit but do not abolish normal IFN-γ signaling and cause nontuberculous mycobacterial osteomyelitis. Dominant negative mutations in *STAT1* and recessive mutations in *IL-12RB1* can produce variable phenotypes consistent with their residual capacities for IFN-γ synthesis and response. Male patients who have disseminated nontuberculous mycobacterial infections along with bacterial or viral infections; conical, peg, or missing teeth; or an abnormal hair pattern should be evaluated for defects in the pathway that activates NF-κB through NEMO (*IKBKG*). These patients may have associated immune globulin defects as well. Patients with myelodysplasia and mycobacterial disease should be investigated for GATA2 deficiency. A recently recognized group of patients who often develop disseminated infections with both MAC and rapidly growing NTM (predominantly *M. abscessus*) as well as other opportunistic infections such as *Talaromyces* have high-titer neutralizing autoantibodies to IFN-γ. This syndrome has been reported most frequently in East Asian female patients.

IV catheters can become infected with NTM, usually as a consequence of contaminated water. *M. abscessus* and *M. fortuitum* sometimes infect deep indwelling lines as well as fluids used in eye surgery, subcutaneous injections, and local anesthetics. Infected catheters should be removed.

Pulmonary Disease Lung disease is by far the most common form of nontuberculous mycobacterial infection in North America and the rest of the industrialized world. In North America, rates of NTM lung disease far exceed rates of tuberculosis. The clinical presentation typically consists of months or years of throat clearing, nagging cough, and slowly progressive fatigue. Patients will often have seen physicians multiple times and received symptom-based or transient therapy before the diagnosis is entertained and samples are sent for mycobacterial stains and cultures. Because not all patients can produce sputum, bronchoscopy may be required for diagnosis. The typical lag between onset of symptoms and diagnosis is ~5 years in older women. Predisposing factors include underlying lung diseases such as bronchiectasis **(Chap. 301)**, pneumoconiosis **(Chap. 300)**, chronic obstructive pulmonary disease **(Chap. 303)**, primary ciliary dyskinesia **(Chap. 301)**, α_1 antitrypsin deficiency **(Chap. 303)**, and cystic fibrosis **(Chap. 302)**. Bronchiectasis and nontuberculous mycobacterial infection often coexist and progress in tandem. This situation makes causality difficult to determine in a given index case, but bronchiectasis is certainly among the most critical predisposing factors that are exacerbated by infection.

MAC are the most common cause of pulmonary nontuberculous mycobacterial infection in North America, but rates vary somewhat by region. MAC infection most commonly develops during the sixth or seventh decade of life in women who have had months or years of nagging intermittent cough and fatigue, with or without sputum production or chest pain. The constellation of pulmonary disease due to NTM in a tall and thin woman who may have chest wall abnormalities is often referred to as "Lady Windermere syndrome," after an Oscar Wilde character of the same name. In fact, pulmonary MAC infection does afflict older nonsmoking white women more than men, with onset at ~60 years. Patients tend to be taller and thinner than the general population, with high rates of scoliosis, mitral valve prolapse, and pectus anomalies. Whereas male smokers with upper-lobe cavitary disease tend to carry the same single strain of MAC indefinitely, nonsmoking females with nodular bronchiectasis tend to carry several strains of MAC simultaneously, with changes over the course of their disease.

M. kansasii can cause a clinical syndrome that strongly resembles tuberculosis, consisting of hemoptysis, chest pain, and cavitary lung disease. The rapidly growing NTM, such as *M. abscessus*, have been associated with esophageal motility disorders such as achalasia. Patients with pulmonary alveolar proteinosis are prone to pulmonary nontuberculous mycobacterial and *Nocardia* infections; the underlying mechanism may be inhibition of alveolar macrophage function due to the autoantibodies to granulocyte-macrophage colony-stimulating factor found in many of these patients.

Cervical Lymph Nodes The most common form of nontuberculous mycobacterial infection among young children in North America is isolated cervical lymphadenopathy, caused most frequently by MAC organisms but also by other NTM. The cervical swelling is typically firm and relatively painless, with a paucity of systemic signs. Because the differential diagnosis of painless adenopathy includes malignancy, many children have infection with NTM diagnosed inadvertently at biopsy; cultures and special stains may not have been requested because mycobacterial disease was not ranked high in the differential. Local fistulae usually resolve completely with resection and/or antibiotic therapy. Likewise, the entity of isolated pediatric intrathoracic nontuberculous mycobacterial infection, which is probably related to cervical lymph node infection, is usually mistaken for cancer. In neither isolated cervical nor isolated intrathoracic infections with NTM have children with underlying immune defects been commonly identified, nor do the affected children usually go on to develop other opportunistic infections.

Skin and Soft Tissue Disease Cutaneous involvement with NTM usually requires a break in the skin for introduction of the bacteria. Pedicure bath–associated infection with *M. fortuitum* is more likely if skin abrasion (e.g., during leg shaving) has occurred just before the pedicure. Outbreaks of skin infection are often caused by rapidly growing NTM (especially *M. abscessus*, *M. fortuitum*, and *M. chelonae*) acquired via skin contamination from surgical instruments (especially in cosmetic surgery), injections, and other procedures. These infections are typically accompanied by painful, erythematous, draining subcutaneous nodules, usually without associated fever or systemic symptoms.

M. marinum lives in many water sources and can be acquired from fish tanks, swimming pools, barnacles, and fish scales. This organism typically causes papules or ulcers ("fish-tank granuloma"), but the infection can progress to tendinitis with significant impairment of manual dexterity. Lesions appear days to weeks after inoculation of organisms by a typically minor trauma (e.g., incurred during the cleaning of boats or the handling of fish). Tender nodules due to *M. marinum* can advance up the arm in a pattern also seen with *Sporothrix schenckii* (*sporotricoid spread*). The typical carpal tendon involvement may be the first presenting manifestation and may lead to surgical exploration or steroid injection. The index of suspicion for *M. marinum* infections must be high to ensure that proper specimens obtained during procedures are sent for culture.

M. ulcerans, another waterborne skin pathogen, is found mainly in the tropics, especially in tropical areas of Africa. Infection follows skin trauma or insect bites that allow admission to contaminated water. The skin lesions are typically painless, clean ulcers that slough and can cause osteomyelitis. The toxin mycolactone accounts for the diminished host inflammatory response and the painless ulcerations.

DIAGNOSIS

NTM can be detected on acid-fast or fluorochrome smears of sputum or other body fluids. When the organism burden is high, the organisms may appear as gram-positive beaded rods, but this finding is unreliable. (In contrast, nocardiae may appear as gram-positive and beaded but filamentous bacteria.) Again, the requisite and most sensitive step in the diagnosis of any mycobacterial disease is to think of including it in the differential. In almost all laboratories, mycobacterial sample processing, staining, and culture are conducted separately from routine bacteriologic tests; thus, many infections go undiagnosed because of physician failure to request the appropriate test. In addition, mycobacteria usually require separate blood culture media. NTM are broadly differentiated into rapidly growing (<7 days) and slowly growing (≥7 days) forms. Because *M. tuberculosis* typically takes ≥2 weeks to grow, many laboratories refuse to consider culture results final until 6 weeks have elapsed. Newer techniques using liquid culture media permit more rapid isolation of mycobacteria from specimens than is possible with traditional media. Species more readily detected with incubation at 30°C include *M. marinum*, *M. haemophilum*, and *M. ulcerans*. *M. haemophilum* prefers iron supplementation or blood, whereas *M. genavense* requires supplemented medium with the additive mycobactin J. Bacterial formation of pigment in light conditions (*photochromogenicity*) or dark conditions (*scotochromogenicity*) or a lack of bacterial pigment formation (*nonchromogenicity*) was historically used to help categorize NTM. In contrast to NTM colonies, *M. tuberculosis* colonies are beige, rough, dry, and flat. Current identification schemes reliably use biochemical, nucleic acid, or cell wall composition, as assessed by high-performance liquid chromatography or mass spectrometry, for speciation. With the remarkable decline in U.S. cases of tuberculosis over recent decades, NTM have become the mycobacteria most commonly isolated from humans in North America. However, not all isolations of NTM, especially from the lung, reflect pathology and require treatment. Whereas identification of an organism in a blood or organ biopsy specimen in a compatible clinical setting is considered diagnostic, the American Thoracic Society recommends that pulmonary infection due to NTM be diagnosed only when disease is clearly demonstrable—i.e., in an appropriate clinical and radiographic setting (nodules, bronchiectasis, cavities) and with repeated isolation of NTM from expectorated sputum or recovery of NTM from bronchoscopy or biopsy specimens. Given the large number of species of NTM and the importance of accurate diagnosis for the implementation of proper therapy, identification of these organisms is ideally taken to the species level.

The purified protein derivative (PPD) of tuberculin is delivered intradermally to evoke a memory T-cell response to mycobacterial antigens. This test is variously referred to as the PPD test, the tuberculin skin test, and the Mantoux test, among other designations. Unfortunately, the cutaneous immune response to these tuberculosis-derived filtrate proteins does not differentiate well between infection with some NTM and that with *M. tuberculosis*. Because intermediate reactions (~10 mm) to PPD in latent tuberculosis and nontuberculous mycobacterial infections can overlap significantly, the progressive decline in active tuberculosis in the United States means that NTM probably account for increasing proportions of PPD reactivity. In addition, *bacille Calmette-Guérin* (BCG) can cause some degree of cross-reactivity in PPD testing, posing problems of interpretation for patients who have received BCG vaccine. Assays to measure the elaboration of IFN-γ in response to the relatively tuberculosis-specific proteins ESAT6 and CFP10 form the basis for IFN-γ-release assays (IGRAs). These assays can be performed with whole blood or on membranes. It is important to note that *M. marinum*, *M. kansasii*, and *M. szulgai* also have ESAT6 and CFP10 and may cause false-positive reactions in IGRAs. Despite cross-reactivity with NTM, large PPD reactions (>15 mm) most commonly signify tuberculosis. Conversely, in the setting of anti-IFNγ autoantibodies, the IGRA test is indeterminate (failure of IFNγ detection in response to specific antigens and mitogens, due to neutralizing anti-IFNγ autoantibodies).

Isolation of NTM from blood specimens is clear evidence of disease. Whereas rapidly growing mycobacteria may proliferate in routine blood culture media, slow-growing NTM typically do not; therefore, it is imperative to suspect the diagnosis and to use the correct bottles for cultures. Isolation of NTM from a biopsy specimen constitutes strong evidence for infection, but cases of laboratory contamination do occur. Identification of organisms on stained sections of biopsy material confirms the authenticity of the culture. Certain NTM require lower incubation temperatures (*M. genavense*) or special additives (*M. haemophilum*) for growth. Some NTM (e.g., *M. tilburgii*) remain noncultivable but can be identified molecularly in clinical samples.

The radiographic appearance of nontuberculous mycobacterial disease in the lung depends on the underlying disease, the severity of the infection, and the imaging modality used. The advent and increase in the use of computed tomography (CT) scanning has allowed the identification of characteristic changes that are highly consistent with nontuberculous mycobacterial infection, such as the "tree-in-bud" pattern of bronchiolar inflammation **(Fig. 185-2)**. Involvement of the lingual and right-middle lobes is commonly seen on chest CT but is difficult to appreciate on plain film. Severe bronchiectasis and cavity formation are common in more advanced disease.

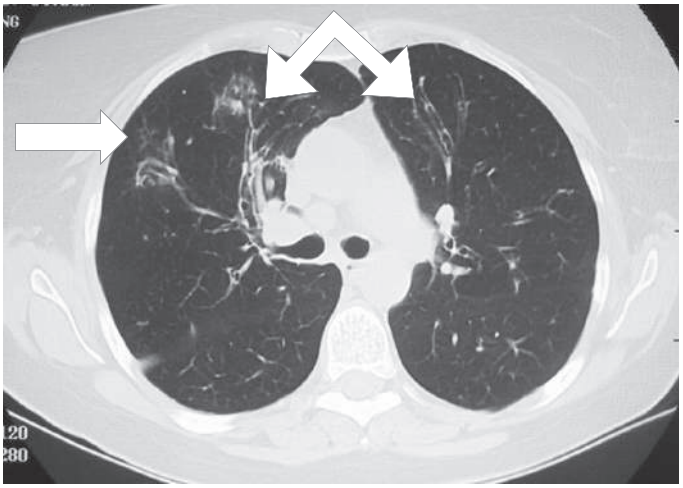

FIGURE 185-2 Chest computed tomography of a patient with pulmonary *Mycobacterium avium* complex infection. *Arrows* indicate the "tree-in-bud" pattern of bronchiolar inflammation (peripheral right lung) and bronchiectasis (central right and left lungs).

Isolation of NTM from respiratory samples can be confusing. *M. gordonae* is often recovered from respiratory samples but is not usually seen on smear and is almost never a pathogen. Patients with bronchiectasis occasionally have NTM recovered from sputum culture with a negative smear. The American Thoracic Society has developed guidelines for the diagnosis of infection with MAC, *M. abscessus*, and *M. kansasii*. A positive diagnosis requires the growth of NTM from two of three sputum samples, regardless of smear findings; a positive bronchoscopic alveolar sample, regardless of smear findings; or a pulmonary parenchyma biopsy sample with granulomatous inflammation or mycobacteria found on section and NTM found on culture. These guidelines probably apply to other NTM as well.

Although many laboratories use DNA probes to identify *M. tuberculosis*, MAC, *M. gordonae*, and *M. kansasii*, speciation of NTM helps determine the antimycobacterial therapy to be used. Only testing of MAC organisms for susceptibility to clarithromycin and of *M. kansasii* for susceptibility to rifampin is indicated; few data support other in vitro susceptibility tests, attractive though they appear. MAC isolates that have not been exposed to macrolides are almost always susceptible. NTM that have persisted beyond a course of antimicrobial therapy are often tested for antibiotic susceptibility, but the value and meaning of these tests are undetermined.

■ PREVENTION

Prophylaxis of MAC disease in patients infected with HIV is started when the CD4+ T lymphocyte count falls to <50/μL. Azithromycin (1200 mg weekly), clarithromycin (1000 mg daily), and rifabutin (300 mg daily) are effective. Macrolide prophylaxis in immunodeficient patients who are susceptible to NTM (e.g., those with defects in the IFN-γ/IL-12 axis) has not been prospectively validated but seems prudent.

TREATMENT

Nontuberculous Mycobacteria

NTM cause chronic infections that evolve relatively slowly over a period of weeks to years. Therefore, it is rarely necessary to initiate treatment on an emergent basis before the diagnosis is clear and the infecting species is known. Treatment of NTM is complex, often poorly tolerated, and potentially toxic. Just as in tuberculosis, inadequate single-drug therapy is almost always associated with the emergence of antimicrobial resistance and relapse.

MAC infection often requires multidrug therapy, the foundation of which is a macrolide (clarithromycin or azithromycin), ethambutol, and a rifamycin (rifampin or rifabutin). For disseminated nontuberculous mycobacterial disease in HIV-infected patients, the use of rifamycins poses special problems—i.e., rifamycin interactions with protease inhibitors. For pulmonary MAC disease, thrice-weekly administration of a macrolide, a rifamycin, and ethambutol has been successful. Therapy is prolonged, generally continuing for 12 months after culture conversion; typically, a course lasts for at least 18 months. Other drugs with activity against MAC organisms include IV and aerosolized aminoglycosides, fluoroquinolones, and clofazimine. In elderly patients, rifabutin can exert significant toxicity. However, with only modest efforts, most antimycobacterial regimens are well tolerated by most patients. Resection of cavitary lesions or severely bronchiectatic segments has been advocated for some patients, especially those with macrolide-resistant infections. The success of therapy for pulmonary MAC infections depends on whether disease is nodular or cavitary and on whether it is early or advanced, ranging from 20 to 80%.

M. kansasii lung disease is similar to tuberculosis in many ways and is also effectively treated with isoniazid (300 mg/d), rifampin (600 mg/d), and ethambutol (15 mg/kg per day). Other drugs with very high-level activity against *M. kansasii* include macrolides, fluoroquinolones, and aminoglycosides. Treatment should continue until cultures have been negative for at least 1 year. In most instances, *M. kansasii* infection is easily cured. Bulky, severe, necrotizing *M. kansasii* lymphadenopathy, especially in the mediastinum, is strongly associated with GATA2 deficiency.

Rapidly growing mycobacteria pose special therapeutic problems. Extrapulmonary disease in an immunocompetent host is usually due to inoculation (e.g., via surgery, injections, or trauma) or to line infection and is often treated successfully with a macrolide and another drug (with the choice based on in vitro susceptibility), along with removal of the offending focus. In contrast, pulmonary disease, especially that caused by *M. abscessus*, is extremely difficult to cure. Repeated courses of treatment are usually effective in reducing the infectious burden and symptoms. Therapy generally includes a macrolide along with an IV-administered agent such as amikacin, a carbapenem, cefoxitin, or tigecycline. Other oral agents (used according to in vitro susceptibility testing and tolerance) include fluoroquinolones, doxycycline, linezolid, and the newer tetracycline family drugs, omadacycline and eravacycline. Because nontuberculous mycobacterial infections are chronic, care must be taken in the long-term use of drugs with neurotoxicities, such as linezolid and ethambutol. Prophylactic pyridoxine has been suggested in these cases. Durations of therapy for *M. abscessus* lung disease are difficult to predict because so many cases are chronic and require intermittent therapy. Expert consultation and management are strongly recommended.

Once recognized, *M. marinum* infection is highly responsive to antimicrobial therapy and is cured relatively easily with any combination of a macrolide, ethambutol, and a rifamycin. Therapy should be continued for 1–2 months after clinical resolution of isolated soft-tissue disease; tendon and bone involvement may require longer courses in light of clinical evolution. Other drugs with activity against *M. marinum* include sulfonamides, trimethoprim-sulfamethoxazole, doxycycline, and minocycline.

Treatment of the other NTM is less well defined, but macrolides and aminoglycosides are usually effective, with other agents added as indicated. Expert consultation is strongly encouraged for difficult or unusual infections due to NTM.

PROGNOSIS

The outcomes of nontuberculous mycobacterial infections are closely tied to the underlying condition (e.g., IFN-γ/IL-12 pathway defect, cystic fibrosis) and can range from recovery to death. With no or inadequate treatment, symptoms and signs can be debilitating, including persistent cough, fever, anorexia, and severe lung destruction. With treatment, patients typically regain strength and energy. The optimal duration of therapy when NTM persist in sputum is unknown, but treatment in this situation can be prolonged. In general, for severe underlying immunodeficiencies, hematopoietic stem cell transplantation is recommended and may be helpful in the resolution of severe mycobacterial disease.

GLOBAL CONSIDERATIONS

In many countries, pulmonary tuberculosis is diagnosed by smear alone, which is also the method used for monitoring of response and relapse. However, examination of mycobacteria from the affected "relapsed" patients shows that a significant proportion of isolates are actually NTM. Overall, as rates of tuberculosis decline, the proportion of positive smears caused by NTM will increase. Advances in speciation will distinguish tuberculosis from nontuberculous mycobacterial infections and thereby affect rates of assumed relapse and resistance, leading to more targeted and appropriate therapy.

FURTHER READING

Blakney RA et al: Incidence of nontuberculous mycobacterial pulmonary infection, by ethnic group, Hawaii, USA, 2005-2019. Emerg Infect Dis 28:1543, 2022.

Daley CL et al: Treatment of nontuberculous mycobacterial pulmonary disease: An official ATS/ERS/ESCMID/IDSA Clinical Practice Guideline. Clin Infect Dis 71:905, 2020.

Holland SM et al: Case 28-2017. A 13-month-old girl with pneumonia and a 33-year-old woman with hip pain. N Engl J Med 377:1077, 2017.

Hong GH et al: Natural history and evolution of anti-interferon-γ autoantibody-associated immunodeficiency syndrome in Thailand and the United States. Clin Infect 71:53, 2020.

Lange C et al: Consensus management recommendations for less common non-tuberculous mycobacterial pulmonary diseases. Lancet Infect Dis 22:e178, 2022.

Marras TK et al: Relative risk of all-cause mortality in patients with nontuberculous mycobacterial lung disease in a US managed care population. Respir Med 145:80, 2018.

Marshall JE et al: Nontuberculous mycobacteria testing and culture positivity in the United States. BMC Infect Dis 24:288, 2024.

Prevots DR et al: Global epidemiology of nontuberculous mycobacterial pulmonary disease: A review. Clin Chest Med 44:675, 2023.

Spinner MA et al: GATA2 deficiency: A protean disorder of hematopoiesis, lymphatics, and immunity. Blood 123:809, 2014.

Szymanski EP et al: Pulmonary nontuberculous mycobacterial infection. A multisystem, multigenic disease. Am J Respir Crit Care Med 192:618, 2015.

186

Antimycobacterial Agents

Divya Reddy, Sebastian G. Kurz, Max R. O'Donnell

Agents used for the treatment of mycobacterial infections, including tuberculosis (TB), leprosy, and infections due to nontuberculous mycobacteria (NTM), are administered in multiple-drug regimens for prolonged courses. Currently, >180 species of mycobacteria have been identified, the majority of which do not cause disease in humans. While the overall incidence of disease caused by *Mycobacterium tuberculosis* has been declining, there has been a recent increase in incidence in the context of the COVID-19 pandemic, and TB remains a leading cause of morbidity and mortality in low- and middle-income countries—especially in sub-Saharan Africa where TB/HIV co-infection is common. Well-organized infrastructure for early diagnosis, treatment of TB infection and disease, and development of effective drug regimens and vaccines remain vital to the global strategies for TB control **(Chaps. 183 and 485)**. Infections with NTM have gained in clinical prominence in the United States and other developed countries. These largely environmental organisms often establish infection in immunocompromised patients or in persons with structural lung disease.

TABLE 186-1 Regimens for the Treatment of Latent Tuberculosis Infection in Adults

REGIMEN	SCHEDULE	DURATION	COMMENTS
Isoniazid plus rifapentine	900 mg (15 mg/kg) weekly plus 900 mg (for weight >50 kg) weekly	3 months	Directly observed therapy is recommended for once-weekly treatment in HIV-positive and -negative individuals. This regimen may be supplemented with pyridoxine (25–50 mg/d).
Rifampin	600 mg/d (10 mg/kg)	4 months	Recommended in HIV-negative individuals and in children. Data on effectiveness in HIV-positive patients are unavailable.
Isoniazid plus rifampin	300 mg/d (5 mg/kg) plus 600 mg/d (10 mg/kg)	3 months	Risk of hepatotoxicity may be higher with the combination regimen compared to that of the individual drugs.
Isoniazid	300 mg/d (5 mg/kg) Alternative: 900 mg twice weekly (15 mg/kg)	6–9 months (6 months acceptable)	Supplement with pyridoxine (25–50 mg daily) 6 months' duration strongly recommended for HIV-negative patients and conditional for HIV-positive patients. 9 months may be more effective but with higher risk of hepatic toxicity. Twice-weekly regimens require directly observed therapy.

Source: TR Sterling et al: Guidelines for the treatment of latent tuberculosis infection: Recommendations from the National Tuberculosis Controllers Association and CDC, 2020. MMWR Recomm Rep 69:1, 2020.

TUBERCULOSIS

GENERAL PRINCIPLES

The earliest recorded human case of TB dates back 9000 years. Early treatment modalities, such as bloodletting, were replaced by the sanatorium movement in the late nineteenth century, which focused on fresh air, nutrition, and bedrest to treat consumptive patients and came with the benefit of isolating infected individuals. The isolation of streptomycin from *Streptomyces griseus* in 1943 launched the era of antibiotic treatment for TB. Over subsequent decades, the discovery of additional agents and the use of multiple-drug regimens allowed progressive shortening of the treatment course from years to as little as 6 months for drug-susceptible TB. Latent TB infection (LTBI) and active TB disease are diagnosed by history, physical examination, radiographic imaging, tuberculin skin test, interferon-γ release assays, acid-fast staining, mycobacterial cultures, and/or new molecular diagnostics. LTBI is treated with isoniazid plus rifapentine (weekly for 3 months), rifampin (daily for 4 months), isoniazid plus rifampin (daily for 3 months), or isoniazid (optimally daily or twice weekly for 6–9 months) **(Table 186-1)**. The 3-month, weekly regimen of isoniazid with rifapentine is currently the regimen of choice in children >2 years of age and in all adults including HIV-positive individuals. The regimen is not recommended for pregnant women and for persons with hypersensitivity reactions to isoniazid or rifampin. Shorter duration rifamycin-based regimens (rifampin alone for 4 months or for 3 months in combination with isoniazid) are currently preferred for the treatment of LTBI over isoniazid for 6–9 months in adults and children due to their effectiveness, safety, and tolerability. Caution is advised in HIV-positive individuals due to potential for drug interactions, lack of definitive data on effectiveness, and the possibility of subclinical TB disease that could facilitate the development of rifampin resistance.

Completions rates of a self-administered, once-weekly regimen of isoniazid plus rifapentine for 3 months with monthly monitoring were found to be noninferior to those seen with directly observed therapy (DOT) in the United States, and thus, this regimen is considered an acceptable strategy for treating LTBI in countries with a focus on secondary prevention of TB disease. Recently, a 1-month daily regimen of rifapentine and isoniazid in HIV-positive individuals was found to be noninferior to 9 months of isoniazid; this regimen is included in the 2020 World Health Organization (WHO) LTBI treatment guidelines.

For active or suspected TB disease, clinical factors, including HIV co-infection, symptom duration, radiographic appearance, and public health concerns about TB transmission, drive diagnostic testing and treatment initiation. Confirmation of active TB relies on detection of *M. tuberculosis* via culture or molecular testing. A combination of drugs is used for the treatment of TB disease **(Table 186-2)**. For drug-susceptible disease, a standardized regimen is used with an intensive phase consisting of four drugs—isoniazid (H), rifampin (R), pyrazinamide (Z), and ethambutol (E)—given for 2 months, which is followed by a continuation phase of isoniazid and rifampin for 4 months, for a total treatment duration of 6 months. U.S. guidelines recommend extension of the continuation phase to 7 months (for a total treatment duration of 9 months) for patients with cavitary disease; if the 2-month course of pyrazinamide is not completed; or if sputum cultures remain positive beyond 2 months of treatment (delayed culture conversion), which also warrants evaluation for development of drug resistance.

In 2020, a large multinational randomized trial showed that a 4-month regimen composed of daily rifapentine, isoniazid, pyrazinamide, and moxifloxacin for 8 weeks, followed by rifapentine, isoniazid, and moxifloxacin for 9 weeks, was noninferior to the traditional 6-month HRZE regimen. Patients with HIV and a CD4 count >100 were included. This regimen now has conditional recommendation by

TABLE 186-2 Simplified Approach to Treatment of Active Tuberculosis (TB) in Adults

CULTURE RESULTS	INTENSIVE PHASE	CONTINUATION PHASE	EXTENSION OF TOTAL TREATMENT
Culture-positive, drug-susceptible	HRZE for 2 months, daily[a] or 3 times per week (with dose adjustment)	HR for 4 months, daily or 5 days per week *or* HR for 4 months, 3 times per week[b] (with dose adjustment)	Continuation phase extended to 7 months if 2 months of Z is not completed, if the patient is infected with HIV and is not receiving antiretroviral therapy, or if culture conversion is prolonged and/or cavitation is evident on chest radiography (U.S. guidelines)[c]
Culture-negative	HRZE for 2 months	HR for 2 months, daily or 2 or 3 times per week[d]	Continuation phase extended to 4 months if the patient is infected with HIV
Extrapulmonary, drug-susceptible	HRZE for 2 months	HR for 4–7 months, daily or 5 days per week[e]	Continuation phase extended to 10 months in TB meningitis; 7 months recommended by some authorities for bone/joint TB

[a] Daily treatment is preferred; however, thrice-weekly therapy in the intensive phase (with or without an initial 2 weeks of daily therapy) may be considered in patients who are not infected with HIV and are at low risk of relapse (i.e., in pulmonary tuberculosis caused by drug-susceptible organisms that, at the start of treatment, is noncavitary and/or smear negative). [b] Use regimen with caution in HIV patients and/or those with cavitary disease, as missed doses can lead to treatment failure, relapse, and acquired drug resistance. [c] Culture conversion is prolonged if it occurs beyond 2 months. [d] Twice-weekly treatment regimens are not recommended in patients infected with HIV and those with cavitary pulmonary disease suspected to be TB. [e] Standard daily 6-month TB treatment regimen is considered to be adequate for most forms of extrapulmonary TB, including miliary TB. For TB meningitis, the addition of glucocorticoids is recommended.

Abbreviations: E, ethambutol; H, isoniazid; R, rifampin; Z, pyrazinamide.

Sources: Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: Clinical practice guidelines: Treatment of drug-susceptible tuberculosis. Clin Infect Dis 63:e147, 2016.

the WHO and U.S. guidelines. Remaining concerns are higher daily pill burden, potential side effects from prolonged use of moxifloxacin, and need for fluoroquinolone resistance testing in areas where resistance is prevalent.

Treatment of TB in patients co-infected with HIV poses significant challenges, but some progress is being made. To improve survival, current recommendations include initiation of antiretroviral therapy (ART) in HIV patients co-infected with *M. tuberculosis* within 2 weeks of the initiation of treatment for TB (except TB meningitis) if the CD4+ T-cell count is ≤50/μL and by 8–12 weeks of TB treatment initiation if the CD4+ T-cell count is ≥50/μL. Interactions of rifampin with protease inhibitors or nonnucleoside reverse transcriptase inhibitors can be significant and require close monitoring and dose adjustments. Reassuringly, a recent study comparing the safety and efficacy of rifampin for 4 months in patients with LTBI showed that it was as effective as isoniazid for 9 months and was also well tolerated and safe for treatment in persons living with HIV. Rifabutin is an alternative drug of choice in HIV patients co-infected with *M. tuberculosis,* as it is a less potent cytochrome P3A inhibitor than rifampin. The TB immune reconstitution inflammatory syndrome (IRIS) may appear as early as 1 week after initiation of ART and manifests as paradoxical worsening or unmasking of existing TB infection. Conservative management consists of continued administration of ART and TB medications. However, severe or debilitating IRIS has been treated in reported case series with varying doses of glucocorticoids. A randomized, double-blind, placebo-controlled trial showed that a 4-week course of prednisone significantly reduced the need for hospitalization and hastened symptom improvement and quality of life in TB IRIS. Intermittent antimycobacterial therapy in patients infected with HIV and *M. tuberculosis* has been associated with low plasma levels of several key TB drugs and with higher rates of treatment failure or relapse; therefore, intermittent twice-weekly therapy for TB in HIV-co-infected individuals is not recommended.

Adherence to medications is critical in achieving a cure with antimycobacterial therapy. In addition to DOT by trained staff, either in the clinic or at home, case management interventions such as patient education/counseling, field/home visits, and patient reminders are also recommended to improve treatment adherence. Use of mobile health technologies, including video DOT, text messaging, and next-generation electronic pillboxes, shows promise in promoting TB adherence. In drug-susceptible TB, monthly dispensing of TB medications is also advised for all patients to allow essential clinical monitoring for hepatotoxicity due to these medications. Clinical monitoring includes at least monthly assessment for symptoms (nausea, vomiting, abdominal discomfort, and unexplained fatigue) and signs (jaundice, dark urine, light stools, diffuse pruritus) of hepatotoxicity, although the latter represent comparatively late manifestations **(Table 186-3)**. The presence of such symptoms and signs mandates provisional discontinuation of potentially hepatotoxic agents; discontinuation at the onset of hepatitis symptoms reduces the risk of progression to fatal hepatitis. Although biochemical monitoring is not routinely recommended, baseline assessment of liver function is recommended in adults including testing of at least serum alanine aminotransferase (ALT) and total bilirubin levels (Table 186-3). (See **Chap. 183** for further details.) For patients with active TB, monthly mycobacterial cultures of sputum are recommended until it is certain that the organisms have been cleared and the patient has responded to therapy or until no sputum is available for culture.

If significant clinical improvement does not occur or the patient's condition deteriorates over the course of therapy, possibilities include

TABLE 186-3 Monitoring and Clinical Management of Tuberculosis (TB) Treatment in Adults[a]

DRUG	ASSESSMENT	MANAGEMENT
LTBI Treatment		
With hepatic risk factors[b], check ALT and bilirubin at baseline. If ALT is ≥3× ULN or total bilirubin is >2× ULN, defer treatment and reevaluate.		
Isoniazid	Determine whether hepatic risk factors are present. If so, obtain baseline and periodic ALT and bilirubin values	If ALT is 5× ULN (or 3× ULN with symptoms)[c] or if bilirubin reaches jaundice levels (usually >2× ULN), interrupt treatment. With normalization, consider an alternative agent.
Rifampin	Same as above	Same as above
TB Treatment		
Check ALT, bilirubin, platelets, creatinine, and hepatitis panel for all patients at baseline. If hepatic risk factors are present, check ALT and bilirubin monthly.		
Isoniazid	If ALT is >5× ULN (or >3× ULN with hepatitis symptoms)[c]	Obtain history of alcohol consumption and concomitant drug use. In most instances, discontinue H, Z, R, and other hepatotoxic drugs. Consider alternative agents. Obtain viral hepatitis serologies. Rechallenge: With normalization of liver enzymes, R and H may be sequentially reintroduced. With no recurrence of hepatotoxicity, Z is not resumed in many cases. Alternative rechallenge protocols have been used.
Rifampin	If primary elevation is in bilirubin and alkaline phosphatase, most likely due to rifampin	Discontinue R if total bilirubin reaches jaundice levels (usually >2× ULN). May try to reintroduce; if not tolerated, may substitute Q.
Ethambutol	Decrease in visual acuity or color vision on monthly screening	Discontinue ethambutol and repeat ocular examination. Peripheral neuropathy may be a precursor of ocular toxicity; if it occurs, consider repeat ocular examination.
Pyrazinamide	If ALT is >5× ULN (or >3× ULN with symptoms)[c]	Same as for H.
Fluoroquinolone, bedaquiline, delamanid	QTc prolongation is a concern and should be monitored, especially if drugs are used in combination	Asymptomatic QTc prolongation should prompt consideration of stopping known QT-prolonging drugs and/or close monitoring, depending on the clinical situation and degree of prolongation. Symptomatic QTc prolongation (e.g., palpitations or arrhythmias) should prompt discontinuation of drugs.
Linezolid	Visual impairment; monitor for peripheral neuropathy and bone marrow suppression including anemia, thrombocytopenia, and leukopenia	Discontinue linezolid if visual toxicity develops. Rechallenge after complete resolution, especially with a lower dose, is an option. Stop if peripheral neuropathy or bone marrow suppression develops.

[a]All regimens require monthly clinical monitoring. [b]Hepatic risk factors: chronic alcohol use, viral hepatitis, preexisting liver disease, pregnancy or ≤3 months postpartum, hepatotoxic medications. [c]Relevant manifestations include nausea, vomiting, abdominal pain, jaundice, or unexplained fatigue.

Abbreviations: ALT, alanine aminotransferase; H, isoniazid; LTBI, latent tuberculosis infection; Q, fluoroquinolone; QTc, corrected QT interval; R, rifampin; ULN, upper limit of normal; Z, pyrazinamide.

Sources: JJ Saukkonen et al: An official ATS statement: Hepatotoxicity of antituberculosis therapy. Am J Respir Crit Care Med 174:935, 2006; American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: Treatment of tuberculosis. Am J Respir Crit Care Med 167:603, 2003; WHO consolidated guidelines on drug-resistant tuberculosis treatment. Geneva: World Health Organization; 2019. License: CC BY-NC-SA 3.0 IGO.

treatment failure due to incomplete adherence, poor medication absorption, or the development of resistance. For patients co-infected with HIV and *M. tuberculosis*, IRIS, which is a diagnosis of exclusion, should also be a consideration. Drug susceptibility testing should be repeated at this point. If resistance is documented or strongly suspected, at least two efficacious drugs to which the isolate is susceptible or which the patient has not already taken should be added to the therapeutic regimen.

Multidrug-resistant tuberculosis (MDR-TB) is defined as disease caused by a strain of *M. tuberculosis* that is resistant to both isoniazid and rifampin—the most efficacious of the first-line TB drugs. The risk of MDR-TB is elevated in patients presenting from geographic areas in which ≥5% of incident cases are MDR-TB and in patients previously treated for TB. Treatment regimens for MDR-TB are rapidly evolving, and in 2019, the WHO issued a new classification of second-line agents to treat drug-resistant disease **(Table 183-4)**. New 2022 WHO recommendations emphasize an all-oral bedaquiline-containing regimen with the goal to limit treatment duration to 6 months compared to conventional durations of 9 months or longer (Table 186-4).

Results from several recent large clinical trials have formed the basis of these recommendations. The "Bangladesh regimen" was the first short-course MDR-TB regimen systematically studied in the STREAM-1 trial and was able to reduce treatment duration to 9–12 months with favorable outcomes in up to 90% of patients. It consists of a seven-drug intensive phase (kanamycin, prothionamide, isoniazid, fluoroquinolone, ethambutol, pyrazinamide, and clofazimine) and a four-drug continuation phase (fluoroquinolone, ethambutol, pyrazinamide, and clofazimine). In 2018, a large meta-analysis, which pooled individual data from >12,000 patients enrolled in 50 trials, assessed the role of individual drugs to treat MDR-TB. This analysis showed an association of significantly better treatment outcomes with the use of linezolid, bedaquiline, clofazimine, carbapenems, and later generation fluoroquinolones and worse outcomes with kanamycin and capreomycin in these patients. As a result of this analysis, oral drug combinations are now prioritized, while several traditional second-line drugs, including kanamycin and capreomycin, are no longer recommended.

The shift toward all-oral regimens of shortened duration has been made possible by the introduction of novel drugs, most prominently bedaquiline and pretomanid, as well as the repurposing of existing agents for MDR-TB treatment (e.g., linezolid, clofazimine). A further step toward a shortened all-oral regimen was the Nix-TB study, which showed that a 6-month regimen of bedaquiline, pretomanid, and linezolid (BPaL regimen) for treatment of highly drug-resistant TB was associated with favorable outcomes (absence of clinical or bacteriologic treatment failure or relapse within 6 months of treatment completion) in 89% of patients. While a major breakthrough, caution has been raised regarding the higher rate of side effects, mostly due to linezolid, and lack of a control arm. The TB PRACTECAL trial was an open-label, multicenter, randomized, controlled, noninferiority trial that evaluated the safety and efficacy of a 24-week regimen of bedaquiline, pretomanid, linezolid, and moxifloxacin (BPaLM) compared to a 9- to 20-month standard-care regimen for the treatment of rifampicin-resistant TB. In this study, the 6-month BPaLM regimen was found to be noninferior in both the intention-to-treat and per-protocol analyses, with a lower percentage of serious adverse events compared to the standard-care regimen. High cost, limited access to these new drugs, and the threat of baseline and emergent resistance, especially to bedaquiline, are barriers that need to be addressed to facilitate global adaptation of these new regimens.

FIRST-LINE ANTITUBERCULOSIS DRUGS

The following discussion of individual anti-TB agents focuses on treatment of TB in adults, unless otherwise noted. Several agents are being actively investigated during the current remarkable period of drug development for TB treatment.

Isoniazid Isoniazid is a critical drug for treatment of both TB disease and LTBI. Isoniazid has excellent bactericidal activity against both intracellular and extracellular, actively dividing *M. tuberculosis*. This drug is bacteriostatic against slowly dividing organisms. In treatment of LTBI, isoniazid is generally well tolerated, has well-established efficacy, and is inexpensive. In this setting, the drug is taken daily, which is the preferred dosing schedule, or intermittently (i.e., twice weekly) using DOT for 6 months, which has been found to be equivalent to the traditional 9 months in most settings. A weekly isoniazid and rifapentine regimen, administered over 3 months under DOT, has been shown to be noninferior to daily isoniazid given for 9 months and had a higher treatment completion rate than the single-drug regimen. More recent evidence also suggests that completion rates of a self-administered 3-month regimen of weekly isoniazid and rifapentine are noninferior

TABLE 186-4 Simplified Approach to Treatment of Drug-Resistant Tuberculosis (TB) in Adults[a]

CULTURE RESULTS	INTENSIVE PHASE	CONTINUATION PHASE	EXTENSION OF TOTAL TREATMENT
Resistant to H	Lfx RZE[b] for 6 months	…	Prolonged culture conversion and/or evidence of cavitation on chest radiography.
Resistant to HR (MDR)[c]			
BPaLM Regimen[d]	Bdq, Pa, Lz, Mfx for 6 months (may drop Mfx if documented Q resistance)		Failed culture conversion at month 4–6, delayed clinical response. Treatment interruption of >7 days is made up by adding on to the total treatment duration.
WHO short-course regimen (9-month all-oral regimen)[e]	Bdq plus Lfx or Mfx, Eto, E, Z, Hh, Cfz for 4–6 months	Lfx or Mfx, Cfz, Z, E for 5 months	
WHO extended regimen[f]	At least four effective second-line agents, including all three group A and at least one group B; add group C if intolerant to A or B drugs for 5–7 months	At least 4 drugs for a total of 18–20 months or for 15–17 months after culture conversion	

[a]Drug-resistant TB treatment regimens should be constructed and care provided in close consultation with experienced drug-resistant TB clinicians. Surgical management should also be considered in appropriate cases. [b]Prolonged pyrazinamide duration may be associated with increased risk of liver toxicity. [c]Monoresistance to R is rare and should be treated as MDR. The BPaLM regimen is now the preferred MDR treatment regimen for patients without significant prior exposure to Bdq, Dlm, Pa, or Lz. [e]The WHO short-course regimen is recommended for patients not qualifying for BPaLM regimen (availability, medical comorbidities, drug resistance, prior exposure to Pa or Lz) with no prior exposure to second-line drugs and documented fluoroquinolone susceptibility only. Patients with treatment intolerance to antimycobacterial agents, disseminated TB, or pregnancy should be excluded from short-course regimens. [f]Patients who do not qualify for WHO short-course regimens should be treated using extended MDR-TB treatment regimens. The construction of extended regimens is guided by the requirement for selection of effective antimycobacterial agents, the need to combine sufficient medicines to maximize relapse-free survival, and the need to minimize toxicity.

Abbreviations: Bdq, bedaquiline; Cfz, clofazimine; E, ethambutol; Eto, ethionamide; H, isoniazid; Hh, high-dose isoniazid; Lfx, levofloxacin; Pa, Pretomanid; Dlm, Delamanid; Lz, Linezolid; MDR, multidrug resistant; Mfx, moxifloxacin; Pa, pretomanid; Q, fluoroquinolone; R, rifampin; WHO, World Health Organization; Z, pyrazinamide.

Sources: Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: Clinical practice guidelines: Treatment of drug-resistant tuberculosis. Am J Respir Crit Care Med 200:e93, 2019; World Health Organization consolidated guidelines on drug-resistant tuberculosis treatment. WHO 2022.

to those seen with DOT in the United States. It is expected that a 1-month daily regimen in combination with rifapentine will be added to new WHO guidelines.

For treatment of TB disease, isoniazid is used in combination with other agents to ensure killing of both actively dividing *M. tuberculosis* and slowly growing "persister" mycobacteria. Unless the organism is resistant, the standard regimen includes isoniazid, rifampin, ethambutol, and pyrazinamide (Table 186-2). Isoniazid is often given together with 25–50 mg of pyridoxine daily to prevent drug-related peripheral neuropathy.

MECHANISM OF ACTION Isoniazid is a prodrug activated by the mycobacterial KatG catalase-peroxidase; isoniazid is coupled with reduced nicotinamide adenine dinucleotide (NADH). The resulting isonicotinic acyl–NADH complex blocks the mycobacterial ketoenoylreductase known as InhA through binding to its substrate and inhibiting fatty acid synthase and ultimately mycolic acid synthesis. Mycolic acids are essential components of the mycobacterial cell wall. KatG activation of isoniazid also results in the release of free radicals that have antimycobacterial activity, including nitric oxide.

The minimal inhibitory concentrations (MICs) of isoniazid for wild-type (untreated) susceptible strains are <0.1 μg/mL for *M. tuberculosis* and 0.5–2 μg/mL for *M. kansasii*.

PHARMACOLOGY Isoniazid is the hydrazide of isonicotinic acid, a small, water-soluble molecule. The usual adult oral daily dose of 300 mg results in peak serum levels of 3–5 μg/mL within 30 min to 2 h after ingestion—well in excess of the MICs for most susceptible strains of *M. tuberculosis*. Both oral and IM preparations of isoniazid reach effective levels in the body, although antacids and high-carbohydrate meals may interfere with oral absorption. Isoniazid diffuses well throughout the body, reaching therapeutic concentrations in body cavities and fluids, with concentrations in cerebrospinal fluid (CSF) comparable to those in serum.

Isoniazid is metabolized in the liver via acetylation by *N*-acetyltransferase 2 (NAT2) and hydrolysis. Both fast- and slow-acetylation phenotypes occur; patients who are "fast acetylators" may have lower serum levels of isoniazid, whereas "slow acetylators" may have higher levels and experience more toxicity. Satisfactory isoniazid levels are attained in the majority of homozygous fast NAT2 acetylators given a dose of 6 mg/kg and in the majority of homozygous slow acetylators given only 3 mg/kg. Genotyping is increasingly being used to characterize isoniazid-related pharmacogenomic responses.

Isoniazid's interactions with other drugs are due primarily to its inhibition of the cytochrome P450 system. Among the drugs with significant isoniazid interactions are warfarin, carbamazepine, benzodiazepines, acetaminophen, clopidogrel, maraviroc, dronedarone, salmeterol, tamoxifen, eplerenone, and phenytoin.

DOSING The recommended daily dose of isoniazid for the treatment of TB is 5 mg/kg for adults and 10 mg/kg for children (U.S. guidelines recommend 10–15 mg), with a maximal daily dose of 300 mg for both. For intermittent therapy in adults (usually twice per week), the dose is 15 mg/kg, with a maximal daily dose of 900 mg. Isoniazid does not require dosage adjustment in patients with renal disease. When the 12-dose, 3-month weekly LTBI regimen is used, the dose of isoniazid is 15 mg/kg, with a maximal dose of 900 mg, and the drug is co-administered with rifapentine. The novel 1-month regimen uses isoniazid 300 mg in conjunction with rifapentine for people aged >13 years without weight adjustment.

RESISTANCE Although isoniazid, along with rifampin, is the mainstay of TB treatment regimens, ~7% of clinical *M. tuberculosis* isolates in the United States are resistant. Rates of primary isoniazid resistance among untreated patients are significantly higher in many populations born outside the United States. Five separate pathways for isoniazid resistance have been elucidated. Most strains have amino acid changes in either the catalase-peroxidase gene (*katG*) or the mycobacterial ketoenoylreductase gene (*inhA*). Less frequently, alterations in *kasA*, the gene for an enzyme involved in mycolic acid elongation, and loss of NADH dehydrogenase 2 activity confer isoniazid resistance. In 20–30% of isoniazid-resistant *M. tuberculosis* isolates, increased expression of efflux pump genes, such as *efpA*, *mmpL7*, *mmr*, *p55*, and the Tap-like gene *Rv1258c*, has been implicated as the underlying mechanism of resistance.

ADVERSE EFFECTS Although isoniazid is generally well tolerated, drug-induced liver injury and peripheral neuropathy are significant adverse effects associated with this agent. Isoniazid may cause asymptomatic transient elevation of aminotransferase levels (often termed *hepatic adaptation*) in up to 20% of recipients. Other adverse reactions include rash (2%), fever (1.2%), anemia, acne, arthritic symptoms, a systemic lupus erythematosus–like syndrome, optic atrophy, seizures, and psychiatric symptoms. Symptomatic hepatitis occurs in <0.1% of persons treated with isoniazid alone for LTBI, and fulminant hepatitis with hepatic failure occurs in <0.01%. Isoniazid-associated hepatitis is idiosyncratic, but its incidence increases with age, with daily alcohol consumption, and in women who are within 3 months postpartum.

In patients who have liver disorders or HIV infection, who are pregnant or in the 3-month postpartum period, who have a history of liver disease (e.g., hepatitis B or C, alcoholic hepatitis, or cirrhosis), who use alcohol regularly, who have multiple medical problems, or who have other risk factors for chronic liver disease, the risks and benefits of isoniazid treatment for LTBI should be weighed. If treatment is undertaken, these patients should have serum concentrations of ALT determined at baseline. Routine baseline hepatic ALT testing based solely on an age of >35 years is optional and depends on individual concerns. Monthly biochemical monitoring during isoniazid treatment is indicated for patients whose baseline liver function tests yield abnormal results and for persons at risk for hepatic disease, including the groups just mentioned. Guidelines recommend that isoniazid be discontinued in the presence of hepatitis symptoms or jaundice and an ALT or AST level three times the upper limit of normal or in the absence of symptoms with an ALT or AST level five times the upper limit of normal (Table 186-3).

Peripheral neuropathy associated with isoniazid occurs in up to 2% of patients given 5 mg/kg. Isoniazid appears to interfere with pyridoxine (vitamin B_6) metabolism. The risk of isoniazid-related neurotoxicity is greatest for patients with preexisting disorders that also pose a risk of neuropathy, such as HIV infection; for those with diabetes mellitus, alcohol abuse, or malnutrition; and for those simultaneously receiving other potentially neuropathic medications, such as stavudine. These patients should be given prophylactic pyridoxine (25–50 mg/d).

Rifampin Rifampin is a semisynthetic derivative of *Amycolatopsis rifamycinica* (formerly known as *Streptomyces mediterranei*). The most active antimycobacterial agent available, rifampin is the keystone of first-line treatment for TB. Introduced in 1968, this drug eventually permitted dramatic shortening of the TB treatment course. Rifampin has both sterilizing and bactericidal activity against dividing and nondividing *M. tuberculosis*. The drug is also active against an array of other organisms, including some gram-positive and gram-negative bacteria, *Legionella*, *M. kansasii*, and *Mycobacterium marinum*.

MECHANISM OF ACTION Rifampin exerts both intracellular and extracellular bactericidal activities. Like other rifamycins, rifampin specifically binds to and inhibits mycobacterial DNA-dependent RNA polymerase, blocking RNA synthesis. Susceptible strains of *M. tuberculosis* as well as *M. kansasii* and *M. marinum* are inhibited by rifampin concentrations of 1 μg/mL.

PHARMACOLOGY Rifampin is a fat-soluble, complex macrocyclic molecule readily absorbed after oral administration. Serum levels of 10–20 μg/mL are achieved 2.5 h after the usual adult oral dose of 10 mg/kg (given without food). Rifampin has a half-life of 1.5–5 h. The drug distributes well throughout most body tissues, including CSF. Rifampin turns body fluids such as urine, saliva, sputum, and tears a reddish-orange color—an effect that offers a simple means of assessing patients' adherence to this medication. Rifampin is excreted primarily through the bile and enters the enterohepatic circulation; <30% of a dose is renally excreted.

As a potent inducer of the hepatic cytochrome P450 system, rifampin can decrease the half-life of some drugs, such as digoxin, warfarin, phenytoin, prednisone, cyclosporine, methadone, oral contraceptives, clarithromycin, azole antifungal agents, quinidine, antiretroviral protease inhibitors, and nonnucleoside reverse transcriptase inhibitors. The Centers for Disease Control and Prevention (CDC) has issued guidelines for the management of drug interactions during treatment of HIV and *M. tuberculosis* co-infection (*www.cdc.gov/tb/*).

DOSING The daily dosage of rifampin is 10 mg/kg for adults and 10–20 mg/kg for children, with a maximum of 600 mg/d for both. The drug is given once daily, twice weekly, or three times weekly. No adjustments of dose or frequency are necessary in patients with renal insufficiency.

RESISTANCE Resistance to rifampin in *M. tuberculosis*, *M. leprae*, and other organisms is the consequence of spontaneous, mostly missense point mutations in a core region of the bacterial gene coding for the β subunit of RNA polymerase (*rpoB*). RNA polymerase altered in this manner is no longer subject to inhibition by rifampin. Most rapidly and slowly growing NTM harbor intrinsic resistance to rifampin, for which the mechanism has yet to be determined.

ADVERSE EFFECTS Adverse events associated with rifampin are infrequent and generally mild. Hepatotoxicity due to rifampin alone is uncommon in the absence of preexisting liver disease and often consists of isolated hyperbilirubinemia rather than aminotransferase elevation. Other adverse reactions include rash, pruritus, gastrointestinal symptoms, and pancytopenia. Rarely, a hypersensitivity reaction may occur with intermittent therapy, manifesting as fever, chills, malaise, rash, and—in some instances—renal and hepatic failure.

Pyrazinamide A nicotinamide analogue, pyrazinamide is an important bactericidal drug used in the initial phase of TB treatment. Its administration for the first 2 months of therapy with rifampin and isoniazid allows treatment duration to be shortened from 9 to 6 months and decreases rates of relapse.

MECHANISM OF ACTION Pyrazinamide's antimycobacterial activity is essentially limited to *M. tuberculosis*. The drug is more active against slowly replicating organisms than against actively replicating organisms. Pyrazinamide is a prodrug that is converted by the mycobacterial pyrimidase to the active form, pyrazinoic acid (POA). This agent is active only in acidic environments (pH <6.0), as are found within phagocytes or granulomas. The exact mechanism of action of POA is unclear, but fatty acid synthetase I may be the primary target in *M. tuberculosis*. Susceptible strains of *M. tuberculosis* are inhibited by pyrazinamide concentrations of 16–50 μg/mL at pH 5.5.

PHARMACOLOGY AND DOSING Pyrazinamide is well absorbed after oral administration, with peak serum concentrations of 20–60 μg/mL at 1–2 h after ingestion of the recommended adult daily dose of 15–30 mg/kg (maximum, 2 g/d). It distributes well to various body compartments, including CSF, and is an important component of treatment for tuberculous meningitis. The serum half-life of the drug is 9–11 h with normal renal and hepatic function. Pyrazinamide is metabolized in the liver to POA, 5-hydroxypyrazinamide, and 5-hydroxy-POA. A high proportion of pyrazinamide and its metabolites (~70%) is excreted in the urine. The dosage must be adjusted according to the level of renal function in patients with reduced creatinine clearance.

ADVERSE EFFECTS At the higher dosages used previously, hepatotoxicity was seen in as many as 15% of patients treated with pyrazinamide. However, at the currently recommended dosages, hepatotoxicity now occurs less commonly when this drug is administered with isoniazid and rifampin during the treatment of TB. Older age, active liver disease, HIV infection, and low albumin levels may increase the risk of hepatotoxicity. The use of pyrazinamide with rifampin for the treatment of LTBI is no longer recommended because of unacceptable rates of hepatotoxicity and death in this setting. Hyperuricemia is a common adverse effect of pyrazinamide therapy that usually can be managed conservatively. Clinical gout is rare.

Although pyrazinamide is recommended by international TB organizations for routine use in pregnancy, it is not recommended in the United States because of inadequate teratogenicity data.

RESISTANCE The basis of pyrazinamide resistance in *M. tuberculosis* is a mutation in the *pncA* gene coding for pyrazinamidase, the enzyme that converts the prodrug to active POA. Resistance to pyrazinamide is associated with loss of pyrazinamidase activity, which prevents conversion of pyrazinamide to POA. Of pyrazinamide-resistant *M. tuberculosis* isolates, 72–98% have mutations in *pncA*. Conventional methods of testing for susceptibility to pyrazinamide may produce both false-negative and false-positive results because the high-acidity environment required for the drug's activation also inhibits the growth of *M. tuberculosis*. There is some controversy as to the clinical significance of in vitro pyrazinamide resistance.

Ethambutol Ethambutol is a bacteriostatic antimycobacterial agent first synthesized in 1961. A component of the standard first-line regimen, ethambutol provides synergy with the other drugs in the regimen and is generally well tolerated. Susceptible species include *M. tuberculosis, M. marinum, M. kansasii*, and organisms of the *Mycobacterium avium* complex (MAC); however, among first-line drugs, ethambutol is the least potent against *M. tuberculosis*. This agent is also used in combination with other agents in the continuation phase of treatment when patients cannot tolerate isoniazid or rifampin or are infected with organisms resistant to either of the latter drugs.

MECHANISM OF ACTION Ethambutol is bacteriostatic against *M. tuberculosis*. Its primary mechanism of action is the inhibition of the arabinosyltransferases involved in cell wall synthesis, which probably inhibits the formation of arabinogalactan and lipoarabinomannan. The MIC of ethambutol for susceptible strains of *M. tuberculosis* is 0.5–2 μg/mL.

PHARMACOLOGY AND DOSING From a single dose of ethambutol, 75–80% is absorbed within 2–4 h of administration. Serum levels peak at 2–4 μg/mL after the standard adult daily dose of 15 mg/kg. Ethambutol is well distributed throughout the body except in the CSF; a dosage of 25 mg/kg is necessary for attainment of a CSF level half of that in serum. For intermittent therapy, the dosage is 25–35 mg/kg thrice weekly. To prevent toxicity, the dosage must be lowered and the frequency of administration reduced for patients with renal insufficiency.

ADVERSE EFFECTS Ethambutol is usually well tolerated and has no significant interactions with other drugs. Optic neuritis, the most serious adverse effect reported, typically presents as reduced visual acuity, central scotoma, and loss of the ability to see green (or, less commonly, red). The cause of this neuritis is unknown, but it may be due to an effect of ethambutol on the amacrine and bipolar cells of the retina. Symptoms typically develop several months after initiation of therapy, but ocular toxicity soon after initiation of ethambutol has been described. The risk of ocular toxicity is dose dependent, with occurrence in 1–5% of patients, and can be increased by renal insufficiency. The routine use of ethambutol in younger children is not recommended because monitoring for visual complications can be difficult. If drug-resistant TB is suspected, ethambutol can be used for treatment of children.

All patients starting therapy with ethambutol should have a baseline test for visual acuity, visual fields, and color vision and should undergo an examination of the optic fundus. Visual acuity and color vision should be monitored monthly or less often as needed. Cessation of ethambutol in response to early symptoms of ocular toxicity usually results in reversal of the deficit within several months. Recovery of all visual function may take up to 1 year. In the elderly and in patients whose symptoms are not recognized early, deficits may be permanent. Some experts think that supplementation with hydroxycobalamin (vitamin B_{12}) is beneficial for patients with ethambutol-related ocular toxicity. Other adverse effects of ethambutol are rare. Peripheral sensory neuropathy occurs in rare instances.

RESISTANCE Ethambutol resistance in *M. tuberculosis* and NTM is associated primarily with missense mutations in the *embB* gene that encodes for arabinosyltransferase. Mutations have been found in resistant strains at codon 306 in 50–70% of cases. Mutations at *embB*306 can cause significantly increased MICs of ethambutol, resulting in clinical resistance.

OTHER RIFAMYCIN DRUGS

Rifabutin Rifabutin, a semisynthetic derivative of rifamycin S, inhibits mycobacterial DNA-dependent RNA polymerase. Rifabutin is recommended in place of rifampin for the treatment of TB in HIV-co-infected individuals who are taking protease inhibitors or nonnucleoside reverse transcriptase inhibitors, particularly nevirapine. A study in India showed better TB treatment outcomes in HIV-co-infected patients given daily rifabutin plus atazanavir/ritonavir than in those given thrice-weekly rifabutin plus atazanavir/ritonavir. Rifabutin's effect on hepatic enzyme induction is less pronounced than that of rifampin. Protease inhibitors may cause significant increases in rifabutin levels through inhibition of hepatic metabolism. Rifabutin is more active in vitro than rifampin against MAC organisms and other NTM, but its clinical superiority has not been established.

PHARMACOLOGY Like rifampin, rifabutin is lipophilic and is absorbed rapidly after oral administration, reaching peak serum levels 2–4 h after ingestion. Rifabutin distributes best to tissues, reaching levels 5–10 times higher than those in plasma. Unlike rifampin, rifabutin and its metabolites are partially cleared by the hepatic microsomal system. Rifabutin's slow clearance results in a mean serum half-life of 45 h—much longer than the 3- to 5-h half-life of rifampin. Clarithromycin (but not azithromycin) and fluconazole appear to increase rifabutin levels by inhibiting hepatic metabolism.

ADVERSE EFFECTS The most common adverse effects of rifabutin treatment are gastrointestinal; other reactions include rash, headache, asthenia, chest pain, myalgia, and insomnia. Less common adverse reactions include fever, chills, a flulike syndrome, anterior uveitis, hepatitis, *Clostridium difficile*–associated diarrhea, a diffuse polymyalgia syndrome, and yellow skin discoloration ("pseudo-jaundice"). Laboratory abnormalities include neutropenia, leukopenia, thrombocytopenia, and increased levels of liver enzymes. Rifabutin appears to be better tolerated by the majority (72%) of adult TB patients who have developed rifampin-related adverse effects. Female patients, those co-infected with hepatitis B or hepatitis C, and those with rifampin-related arthralgias, dermatologic reactions, and cholestasis are more likely to develop mild to severe rifabutin-related adverse effects.

RESISTANCE Similar to rifampin resistance, rifabutin resistance is mediated by mutations in *rpoB*.

Rifapentine Rifapentine is a semisynthetic cyclopentyl rifamycin, sharing a mechanism of action with rifampin. Rifapentine is lipophilic and has a prolonged half-life that permits weekly or twice-weekly dosing. Therefore, rifapentine is the subject of intensive clinical investigation aimed at determining optimal dosing and frequency of administration. Currently, it is an alternative to rifampin in the continuation phase of treatment for noncavitary drug-susceptible pulmonary TB in HIV-seronegative patients who have negative sputum smears at completion of the initial phase of treatment. When administered in these specific circumstances, rifapentine (10 mg/kg, up to 600 mg) is given once weekly with isoniazid. Because of higher rates of relapse, this regimen is not recommended for patients with TB disease and HIV co-infection; moreover, it has not been approved for children <12 years of age. In a phase 2 study, substituting daily rifapentine for rifampin yielded higher rates of sputum sterilization after 2 months of intensive treatment. Higher doses of rifapentine (20 mg/kg vs 10 mg/kg) had better results and were safe and well tolerated. Regimens containing high doses of rifapentine are being evaluated to see whether they can shorten the TB treatment course to <6 months.

PHARMACOLOGY Rifapentine's absorption is improved when the drug is taken with food. After oral administration, rifapentine reaches peak serum concentrations in 5–6 h and achieves a steady state in 10 days. The half-life of rifapentine and its active metabolite, 25-desacetyl rifapentine, is ~13 h. The administered dose is excreted via the liver (70%).

ADVERSE EFFECTS The adverse effects profile of rifapentine is similar to that of other rifamycins. Rifapentine is teratogenic in animal models and is relatively contraindicated in pregnancy.

RESISTANCE Rifapentine resistance is mediated by mutations in *rpoB*. Mutations that cause resistance to rifampin also cause resistance to rifapentine.

SECOND-LINE ANTITUBERCULOSIS DRUGS

Second-line anti-TB agents are indicated for treatment of drug-resistant TB, for patients who are intolerant or allergic to first-line agents, and when first-line supplemental agents are unavailable. According to their usability, they are divided into three WHO groups.

Group A • FLUOROQUINOLONES Fluoroquinolones inhibit mycobacterial DNA gyrase and topoisomerase IV, preventing cell replication and protein synthesis, and are bactericidal. Given their excellent activity, they have been investigated for their potential to shorten the course of treatment for drug-susceptible TB from 6 to 4 months. In contrast to prior trials, a recent large, open-label, randomized controlled trial (TBTC Study 31) yielded promising results for shortening of TB treatment. Patients with drug-susceptible TB disease were randomized to receive either a standard 6-month TB regimen or a 4-month regimen containing rifapentine (8 weeks of once-daily rifapentine, isoniazid, pyrazinamide, and ethambutol followed by 9 weeks of once-daily rifapentine and isoniazid) or a 4-month regimen containing rifapentine and moxifloxacin (8 weeks of once-daily rifapentine, isoniazid, pyrazinamide, and moxifloxacin followed by 9 weeks of once-daily rifapentine, isoniazid, and moxifloxacin). The trial demonstrated that a 4-month regimen using daily rifapentine and moxifloxacin (but not the rifapentine-only regimen) was noninferior to the standard 6-month TB treatment regimen using an end point of TB-free survival 12 months after randomization. Combining once-daily rifapentine with moxifloxacin allows for synergistic action on sputum conversion in a compliance-friendly once-daily option. Current recommendations continue to be for a standard 6-month regimen, although it is anticipated that these results will inform future guidelines. Gatifloxacin has fallen out of favor because of significant dysglycemia. Ciprofloxacin and ofloxacin are no longer recommended for the treatment of TB because of poor efficacy. Despite documented resistance to early-generation fluoroquinolones (e.g., ofloxacin and ciprofloxacin), use of a later-generation fluoroquinolone in patients with drug-resistant TB has been associated with favorable outcomes. Fluoroquinolones are also considered safe alternatives for patients who develop treatment-limiting adverse effects from first-line agents. Levofloxacin and moxifloxacin have both been used effectively in the treatment of MDR-TB. The optimal dose of levofloxacin for this indication is being actively studied, but doses of at least 750 mg are commonly used. High-dose moxifloxacin (800 mg) is recommended for standardized shorter MDR-TB regimens.

The fluoroquinolones are well absorbed orally, reach high serum levels, and distribute well into body tissues and fluids. Their absorption is decreased by co-ingestion with products containing multivalent cations, such as antacids. Adverse effects are relatively infrequent (0.5–10% of patients) and include gastrointestinal intolerance, rashes, dizziness, and headache. Most studies of fluoroquinolone side effects have been based on relatively short-term administration for bacterial infections, but trials have now shown the relative safety and tolerability of fluoroquinolones administered for months during TB treatment in adults. Although the potential to prolong the QTc interval, leading to cardiac arrhythmias, has been a source of concern with fluoroquinolones, cessation of treatment due to this adverse effect is rare. Because the benefits may outweigh the risks in treatment of drug-resistant TB, there is increasing interest in the use of fluoroquinolones in children, which has traditionally been avoided because of the risks of tendon rupture and cartilage damage.

Multiple courses of empirical fluoroquinolone therapy for presumed community-acquired pneumonia are associated with delayed diagnosis of active pulmonary TB and increased fluoroquinolone resistance in *M. tuberculosis*. Mutations in the genes encoding for DNA gyrase (*gyrA* and *gyrB*) are implicated in the majority of cases—but not all cases—of clinical resistance to fluoroquinolones.

DIARYLQUINOLINES Bedaquiline (TMC207 or R207910) is a diarylquinoline with a novel mechanism of action: inhibition of the mycobacterial ATP synthetase proton pump. Bedaquiline is bactericidal for *M. tuberculosis*. Resistance has been reported due to point mutations in the *atpE* gene encoding for subunit c of ATP synthetase. Clinical bedaquiline resistance has also been reported due to non-target mutations in mmpR or Rv0678 (a negative repressor of the MmpS5–MmpL5 efflux pump) and PepQ (a cytoplasmic peptidase), both of which may cause cross-resistance to clofazimine. Bedaquiline is metabolized by the hepatic cytochrome CYP3A4. Rifampin lowers bedaquiline levels by 50%, and protease inhibitors also interact significantly with this drug. Because efavirenz induces CYP3A4, there is concern about lower bedaquiline levels with co-administration. In a study of co-treatment with bedaquiline and efavirenz in healthy volunteers, bedaquiline levels were reduced by only 20%; however, in a study modeling chronic co-administration of these two drugs, the reduction in bedaquiline levels was estimated to be 50%, leading many national TB programs to avoid efavirenz co-administration with bedaquiline.

The oral bioavailability of bedaquiline appears to be excellent. The dosage is 400 mg/d for the first 2 weeks and then 200 mg thrice weekly typically for 6 months total. The elimination half-life is long (>14 days). A single dose of this drug can inhibit the growth of *M. tuberculosis* for up to 1 week through a combination of long plasma half-life, high-level tissue penetration, and long tissue half-life. Bedaquiline added to a background regimen improved the 2-month sputum culture–conversion rate in multicenter, randomized, placebo-controlled trials, and these results led to approval by the U.S. Food and Drug Administration (FDA). However, the death rate in one trial was higher in the bedaquiline arm than in the control arm (11.4 vs 2.5%); the result was a "black box" warning from the FDA, which also included QT prolongation. Subsequent studies have not found an association with significant mortality. The CDC has made a provisional recommendation for the use of bedaquiline for 24 weeks in adults with laboratory-confirmed pulmonary MDR-TB when no other effective treatment regimen can be provided. Bedaquiline is an integral part of all shorter course, oral MDR treatment regimens endorsed by the WHO.

OXAZOLIDINONES Linezolid is an oxazolidinone used primarily for the treatment of drug-resistant gram-positive bacterial infections. However, this drug is active in vitro against *M. tuberculosis* and NTM. Several case series have suggested that linezolid may help clear mycobacteria relatively rapidly when included in a regimen for the treatment of complex cases of drug-resistant TB. Linezolid's mechanism of action is disruption of protein synthesis by binding to the 50S bacterial ribosome. Linezolid has nearly 100% oral bioavailability, with good penetration into tissues and fluids, including CSF. Clinical resistance to linezolid has been reported and is typically associated with mutations in the 23S rRNA and in two ribosomal proteins, L3 (*rplC*) and L4 (*rplD*). Adverse effects may include optic and peripheral neuropathy, pancytopenia, and lactic acidosis and are usually associated with higher doses. Linezolid is a weak monoamine oxidase inhibitor and can be associated with the serotonin syndrome when given concomitantly with serotonergic drugs (primarily antidepressants such as selective serotonin reuptake inhibitors). It has been shown that ~80% of patients with MDR-TB can be successfully treated with linezolid-containing, individualized anti-TB regimens based on drug sensitivity testing. Replacement of ethambutol with linezolid for 2–4 weeks during the intensive phase of treatment of drug-susceptible TB is currently being evaluated for possible faster sputum conversion and a shorter treatment regimen. For MDR-TB treatment, linezolid is usually administered at a dose of 600 mg (or less in some cases) once daily, which appears to be effective. A single daily dose is associated with fewer adverse events than twice-a-day dosing.

Sutezolid, a modified version of oxazolidinones and protein synthesis inhibitor, is found to have higher early bactericidal activity compared to linezolid and is currently undergoing phase 2A trials. It is currently FDA approved for complex skin infections and appears to have less frequent side effects compared to linezolid; the adverse effects profile of long-term exposure compared with that of linezolid needs further investigation.

Group B • CLOFAZIMINE Clofazimine is a fat-soluble riminophenazine dye used primarily in the treatment of leprosy worldwide. It is currently gaining popularity in the management of drug-resistant TB because of its low cost and its intracellular and extracellular activity. By increasing reactive oxidant species and causing membrane destabilization, clofazimine may promote killing of antibiotic-tolerant *M. tuberculosis* "persister" organisms. In addition to antimicrobial activity, the drug has other pharmacologic activities, such as anti-inflammatory, pro-oxidative, and immunopharmacologic properties. Clofazimine has a half-life of ~70 days in humans, and average steady-state concentrations are achieved at ~1 month. Intake with fatty meals can improve its low and variable rates of absorption (45–62%). Common side effects include gastrointestinal intolerance and reversible orange to brownish discoloration of skin, bodily fluids, and secretions. Dose adjustment may be necessary in patients with severe hepatic impairment. Clofazimine was studied as part of a regimen developed in Bangladesh for potential shortening of the MDR-TB treatment course. A meta-analysis suggested that inclusion of clofazimine in a multidrug regimen for treatment of MDR-TB was associated with a favorable outcome. Newer analogues with improved pharmacokinetics and alternative formulations of clofazimine (liposomal, nanosuspension, inhalational) are being studied.

CYCLOSERINE Cycloserine is an analogue of the amino acid D-alanine and prevents bacterial cell-wall synthesis. It inhibits the action of enzymes, including alanine racemase, that are involved in the production of peptidoglycans. Cycloserine is active against a range of bacteria, including *M. tuberculosis*. Mechanisms of mycobacterial resistance are not well understood, but overexpression of alanine racemase can confer resistance in *Mycobacterium smegmatis*. Cycloserine is well absorbed after oral administration and is widely distributed throughout body fluids, including CSF. The usual adult dosage is 250 mg two or three times per day. Serious potential side effects include seizures and psychosis (with suicide in some cases), peripheral neuropathy, headache, somnolence, and allergic reactions. Drug levels are monitored to achieve optimal dosing and to reduce the risk of adverse effects, especially in patients with renal failure. Cycloserine should be administered as DOT only with caution and with support from experienced TB physicians to patients with epilepsy, active alcohol abuse, severe renal insufficiency, or a history of depression or psychosis.

Group C • NITROIMIDAZOLES The prodrugs delamanid (OPC-67683) and pretomanid (PA 824) are novel nitro-dihydro-imidazooxazole derivatives that are activated by *M. tuberculosis*–specific flavin-dependent nitroreductases and whose antimycobacterial activity is attributable to inhibition of mycolic acid biosynthesis. Delamanid was shown in a randomized, placebo-controlled, multinational clinical trial to significantly improve the culture conversion rate at 2 months. QT prolongation occurred significantly more often in delamanid-treated patients, but no clinically relevant events were reported. In a subsequent randomized phase 3 trial, there was no significant difference in 6-month sputum conversion between delamanid and placebo among patients with an optimized background regimen. Currently, it is part of several ongoing clinical trials including combination with bedaquiline. It is recommended for use in children younger than 6 years with rifampicin-resistant TB. Usual adult dose is 100 mg twice daily.

Pretomanid, the second novel agent from this class, has shown promising results in the treatment of drug-resistant TB in combination with bedaquiline. A combination of pretomanid with moxifloxacin and pyrazinamide for treatment of drug-susceptible TB was found to have higher culture conversation rates at 8 weeks compared to HRZE; however, a subsequent phase 3 study raised concern for higher frequency of potentially fatal hepatotoxicity. It is currently being evaluated in several phase 3 clinical trials in various combinations, including with fluoroquinolones and pyrazinamide. Based on the previously mentioned results with the BPaL regimen (Nix-TB study), the FDA has granted approval for specific highly resistant TB cases. Adult treatment dose is 200 mg administered daily.

AMOXICILLIN-CLAVULANATE AND CARBAPENEMS β-Lactam agents are largely ineffective for the treatment of *M. tuberculosis* because of resistance conferred by a hydrolyzing class A β-lactamase, BlaC.

Carbapenems are poor substrates of BlaC, and clavulanic acid leads to irreversible inhibition. While the use of either amoxicillin-clavulanic acid or carbapenems alone for highly resistant forms of TB has been anecdotally reported with unclear results, the combination of meropenem and clavulanic acid turned out to be highly active in vitro. Recently, the combination was found to have effective early bactericidal activity, and in a large individual patient data meta-analysis, the combination was associated with positive outcomes. Nevertheless, the need to administer these carbapenems intravenously and the lack of information on the drugs' long-term side effects have restricted their use to certain severe cases only. Recommended daily doses are either imipenem-cilastatin 1 g (each component) IV twice daily or meropenem 1 g IV three times daily, each in combination with clavulanic acid 125 mg oral twice daily, which is only available in combination with amoxicillin.

AMINOGLYCOSIDES Aminoglycosides have played a time-honed role in the treatment of mycobacterial infections. Amikacin and streptomycin are aminoglycosides that exert mycobactericidal activity by binding to the 16S ribosomal subunit. The spectrum of antibiotic activity for amikacin and streptomycin includes *M. tuberculosis*, several NTM species, and aerobic gram-negative and gram-positive bacteria. Due to the need of intravenous or painful intramuscular injections and their serious side effect profile, the WHO recommends limiting their use with the increased availability of novel oral agents. Kanamycin and capreomycin, a cyclic polypeptide similar to aminoglycosides, are no longer recommended due to worse treatment outcomes and increased mortality. This recommendation is based on a large individual patient-level meta-analysis of observational cohort studies and is likely due to increased toxicity seen with these agents. Streptomycin was the first antimycobacterial agent used for the treatment of TB. Derived from *Streptomyces griseus*, streptomycin is bactericidal against dividing *M. tuberculosis* organisms but has only low-level early bactericidal activity. In developing countries, it continues to be widely used due to its low cost. The usual daily dose of streptomycin (given IM either daily or 5 days per week) is 15 mg/kg for adults and 20–40 mg/kg for children, with a maximum of 1 g/d for both with dose reduction recommended for patients ≥60 years of age or with renal impairment. Central nervous system penetration is poor.

Amikacin resistance is less widespread, and streptomycin-resistant strains may still be susceptible. The usual daily adult dosage is 15–30 mg/kg given IM or IV (maximal daily dose, 1 g). It is frequently used to treat severe NTM infections.

Mycobacterial resistance to aminoglycosides is due to mutations in the genes encoding the 16S ribosomal RNA gene (*rrs*). Adverse effects of both amikacin and streptomycin include ototoxicity (in up to 10% of recipients, with auditory dysfunction occurring more commonly than vestibulotoxicity), nephrotoxicity, and neurotoxicity.

ETHIONAMIDE Ethionamide is a derivative of isonicotinic acid. Its mechanism of action is through inhibition of the *inhA* gene product enoyl–acyl carrier protein (acp) reductase, which is involved in mycolic acid synthesis. Ethionamide is bacteriostatic against metabolically active *M. tuberculosis* and some NTM. It is used in the treatment of drug-resistant TB, but its use is limited by severe gastrointestinal reactions (including abdominal pain, nausea, and vomiting) as well as significant central and peripheral neurologic side effects, reversible hepatitis (in ~5% of recipients), hypersensitivity reactions, and hypothyroidism. Ethionamide should be taken with food to reduce gastrointestinal effects and with pyridoxine (50–100 mg/d) to limit neuropathic side effects.

PARA-AMINOSALICYLIC ACID Para-aminosalicylic acid (PAS; 4-aminosalicylic acid) is an oral agent used in the treatment of drug-resistant TB. Its bacteriostatic activity is due to inhibition of folate synthesis and of iron uptake. PAS has relatively little activity as an anti-TB agent. Adverse effects may include high-level nausea, vomiting, and diarrhea. PAS may cause hemolysis in patients with glucose-6-phosphate dehydrogenase deficiency. The drug should be taken with acidic foods to improve absorption. Enteric-coated PAS granules (4 g orally every 8 h) appear to be better tolerated than other formulations and produce higher therapeutic blood levels. PAS has a short half-life (1 h), and 80% of the dose is excreted in the urine.

■ DRUGS IN DEVELOPMENT

The pipeline of novel TB drugs is rapidly changing. We direct the reader to the Working Group on New TB Drugs for the most up-to-date information (*https://www.newtbdrugs.org/pipeline/clinical*).

NONTUBERCULOUS MYCOBACTERIA

More than 180 species of NTM have been identified. Only a minority of these environmental organisms, which are extensively found in soil and water, are important human pathogens. NTM cause extensive disease primarily in persons with preexisting pulmonary disease or immunocompromise but can also cause nodular/bronchiectatic disease in otherwise seemingly healthy hosts. Disseminated infections with NTM are common in immunocompromised individuals. NTM are also important causes of skin and soft tissue infections in surgical settings. The two major classes of NTM are the slow-growing and rapidly growing species; subcultures of the latter grow within 1 week. The growth characteristics of NTM have diagnostic, therapeutic, and prognostic implications. The rate of growth can provide useful preliminary information within a specific clinical context, in that growth within 2–3 weeks is much more likely to indicate an NTM than *M. tuberculosis*. When NTM do grow from cultures, colonization should be distinguished from active disease to optimize the risk and benefit of prolonged treatment with multiple medications. According to the recommendations of the American Thoracic Society and the Infectious Diseases Society of America, significant clinical manifestations and/or radiographic evidence of progressive disease consistent with NTM infection as well as either reproducible sputum culture results or a single positive culture from bronchoscopy are required for the diagnosis of NTM pulmonary disease. Isolation of NTM from blood or from an infected extrapulmonary site, such as soft tissue or bone, is usually indicative of disseminated or local NTM infection (**Chap. 185**). Treatment of NTM disease is prolonged and requires multiple medications. Side effects of the regimens employed are common, and intermittent therapy is often used to mitigate these adverse events. Treatment regimens depend on the NTM species, the extent or type of disease, and—to some degree—drug susceptibility test results.

■ THERAPEUTIC CONSIDERATIONS FOR SPECIFIC NTM

Slowly Growing Mycobacteria Slowly growing mycobacteria can be divided into three categories based on their pigment-producing capabilities and—if they do produce pigment—their requirement for light to do so. *Photochromogens*, including *M. marinum* and *M. kansasii*, can produce yellowish-orange pigment only when exposed to light. *Scotochromogens*, including *Mycobacterium gordonae* and *Mycobacterium scrofulaceum*, can make pigment regardless of light exposure. MAC organisms and *Mycobacterium ulcerans* are *nonchromogens*—i.e., are incapable of making pigment irrespective of light exposure.

***MYCOBACTERIUM AVIUM* COMPLEX** Among the NTM, MAC organisms most commonly cause human disease. In immunocompetent hosts, MAC species are most often found in conjunction with underlying significant lung disease, such as chronic obstructive pulmonary disease or bronchiectasis. For patients with nodular or bronchiectatic MAC lung disease, an initial regimen consisting of clarithromycin or azithromycin, rifampin or rifabutin (the latter is preferred for HIV patients receiving ART), and ethambutol is given three times per week for at least 12 months after culture conversion. A daily regimen of these three drugs, with consideration of amikacin or streptomycin in the initial treatment phase, is recommended for patients with fibrocavitary MAC lung disease or severe nodular/bronchiectatic disease. Routine initial testing for macrolide resistance is recommended, as is testing at 6 months with a failing regimen (i.e., with cultures persistently positive for NTM). Interpretation of susceptibility tests to drugs other than macrolides and aminoglycosides is hampered by poor correlation with clinical outcomes. Amikacin has been reformulated as a liposomal suspension for inhalation with increased penetration into airway biofilms. The CONVERT trial showed that addition of inhaled liposomal amikacin to standard three-drug regimen of azithromycin or clarithromycin, rifampin, and ethambutol in treatment-refractory (persistent sputum positivity after at least 6 months) MAC lung disease significantly increased culture conversion rates from 9 to 26% at 6 months. Respiratory adverse events (primarily dysphonia, cough, and dyspnea) were reported in 87.4% of patients receiving inhaled liposomal amikacin compared to 50% in the standard therapy group; however, rates of serious adverse events were not different between the regimens. Inhaled liposomal amikacin is now approved for use in refractory pulmonary MAC infections (persistent positive cultures after at least 6 months of treatment). It is currently being evaluated as a first-line agent and as a replacement for rifampin in the treatment of MAC lung disease.

Surgical resection should be considered for individuals whose infection is localized to one lung, who have adequate lung function to tolerate lung resection, who have had a poor response to medical therapy, and/or who have developed macrolide-resistant MAC disease.

Treatment of MAC in persons living with HIV should be initiated in consultation with an infectious diseases specialist. For HIV-infected patients with well-controlled HIV disease and CD4 T-cell counts in the normal range, MAC treatment is identical to patients without HIV disease except that drug-drug interactions between antimycobacterial agents and ART should be carefully considered. HIV-infected patients with low CD4 count (CD4+ T-cell count <100/μL) are at risk for disseminated MAC infection. MAC disease in these patients is generally treated with clarithromycin, ethambutol, and rifabutin. Azithromycin may be preferred to clarithromycin depending on adverse effects and patient tolerance. Amikacin and fluoroquinolones are often used in salvage regimens. Treatment for disseminated MAC infection in AIDS patients may be lifelong in the absence of immune reconstitution. Therapy is recommended for at least 12 months after culture conversion and at least 6 months of effective immune reconstitution with ART (CD4+ T-cell count >100/μL).

MYCOBACTERIUM KANSASII *M. kansasii* is the second most common NTM causing human disease in the United States. It is also the second most common cause of NTM pulmonary disease in the United States, where it is most commonly reported in the southeastern region. *M. kansasii* infection can be treated with rifampin, ethambutol, and either isoniazid or macrolide; therapy continues for at least 18 months or for 12 months after culture conversion. The American Thoracic Society and the Infectious Diseases Society of America recommend routine susceptibility testing to rifampin only. Resistance to isoniazid and ethambutol can be acquired during therapy but is usually associated with rifampin resistance as well. Rifampin-resistant *M. kansasii* is treated with a three-drug regimen including agents such as ciprofloxacin, azithromycin, ethambutol, rifabutin, amikacin, trimethoprim-sulfamethoxazole, and streptomycin after drug susceptibility testing.

MYCOBACTERIUM MARINUM *M. marinum* is an NTM found in salt water and freshwater, including swimming pools and fish tanks. It is a cause of localized soft tissue infections, which may require surgical management. Combination regimens include clarithromycin and either ethambutol or rifampin. Other agents with activity against *M. marinum* include doxycycline, minocycline, and trimethoprim-sulfamethoxazole. Drug susceptibility testing is recommended only if the swab remains culture positive after 3 months of appropriate therapy.

Rapidly Growing Mycobacteria Rapidly growing mycobacteria causing human disease include *Mycobacterium abscessus, Mycobacterium fortuitum*, and *Mycobacterium chelonae*. Treatment of these mycobacteria is complex and should be undertaken with input from experienced clinicians. It is important to note that testing rapidly growing mycobacteria for macrolide resistance is tricky, as an inducible *erm* gene may confer in vivo macrolide resistance to isolates that are susceptible in vitro.

M. abscessus is the third most common NTM pathogen in the United States. It is endemic in the southeastern states between Texas and Florida. Skin, soft tissue, and bone infections occur, usually after accidental trauma or surgery. This organism appears to have a predilection to cause lung infections in white nonsmoking women aged >60 who have no preexisting lung disease. *M. abscessus* isolates are usually resistant to standard anti-TB regimens. Skin and soft tissue infections are usually treated for a minimum of 4 months with a macrolide (clarithromycin or azithromycin) and a parenteral agent such as amikacin, cefoxitin, or imipenem. Bone infections are treated for at least 6 months. This regimen can be used for the treatment of lung infections but is often unsuccessful because of drug adverse effects and toxicities. A regimen comprising a combination of at least three active drugs (amikacin, linezolid, tigecycline, imipenem, azithromycin, provided the organism is macrolide susceptible) is recommended based on in vitro drug susceptibility testing. A recent meta-analysis has shown that overall therapeutic efficiency rates in *M. abscessus* lung infection are low at ~35%; however, incorporation of amikacin, imipenem, linezolid, and/or tigecycline was associated with improved outcomes. Conversely, macrolide resistance has been associated with worse outcomes. Surgical resection should be considered in all patients with good lung reserve and a localized infection.

DRUGS FOR THE TREATMENT OF NTM

Clarithromycin Clarithromycin is a macrolide antibiotic with broad activity against many gram-positive and gram-negative bacteria as well as NTM. This drug is active against MAC organisms and many other NTM species, inhibiting protein synthesis by binding to the 50S mycobacterial ribosomal subunit. NTM resistance to macrolides is probably caused by overexpression of the gene *ermB*, with consequent methylation of the binding site. Strains of *M. abscessus* subsp. *abscessus* harbor an inducible macrolide resistance mechanism coded by *erm41*, which leads to ribosomal methylation and becomes apparent after macrolide incubation of 3–5 days, significantly hampering treatment success. Twenty percent of strains have a nonfunctional *erm41* gene. Clarithromycin is well absorbed orally and distributes well to tissues. It is cleared both hepatically and renally; the dosage should be reduced in renal insufficiency. Clarithromycin is a substrate for and inhibits cytochrome 3A4 and should not be administered with cisapride, pimozide, or terfenadine because cardiac arrhythmias may occur. Numerous drugs interact with clarithromycin through the CYP3A4 metabolic pathway. Rifampin lowers clarithromycin levels; conversely, rifampin levels are increased by clarithromycin. However, the clinical relevance of this interaction does not appear to be great.

For patients with nodular/bronchiectatic MAC infection, the dosage of clarithromycin is 500 mg, given morning and evening three times a week. For the treatment of fibrocavitary or severe nodular/bronchiectatic MAC infection, a dose of 500–1000 mg is given daily. Disseminated MAC infection is treated with 1000 mg daily. Clarithromycin is used in combination regimens that typically include ethambutol and a rifamycin in order to avoid the development of macrolide resistance. Adverse effects include frequent gastrointestinal intolerance, hepatotoxicity, headache, rash, and rare instances of hypoglycemia. Clarithromycin is contraindicated during pregnancy because of its teratogenicity in animal models.

Azithromycin Azithromycin is a derivative of erythromycin. Although technically an azalide and not a macrolide, it works similarly to macrolides, inhibiting protein synthesis through binding to the 50S ribosomal subunit. Azithromycin is preferred over clarithromycin due to once-daily dosing, better tolerability, fewer drug interactions, and equal efficacy. Resistance to azithromycin is almost always associated with complete cross-resistance to clarithromycin. Azithromycin is well absorbed orally, with good tissue penetration and a prolonged half-life (~48 h). The usual dosage for treatment of MAC infection is 250 mg daily or 500 mg three times per week. Azithromycin is used in combination with other agents to avoid the development of resistance. For prophylaxis against disseminated MAC infection in immunocompromised individuals, a dose of 1200 mg once per week is given. Because azithromycin is not metabolized by cytochrome P450, it interacts with few drugs. Adjustment of the dosage on the basis of renal function is not necessary.

Amikacin Liposome Inhalation Suspension (ALIS) ALIS is a new formulation of the aminoglycoside amikacin, which allows for improved penetration in the lung with reduced toxicity. In the CONVERT study, treatment with amikacin liposome inhalation suspension in addition to standard background regimen was associated with significantly increased culture conversion (29 vs 8.9%; $p < .0001$) by month 6 in patients with treatment-refractory MAC lung disease compared to standard background regimen alone. It is now approved for treatment of refractory MAC lung infection with persistent sputum positivity at 6 months while on appropriate background regimen. The typical dose is 590 mg (one vial once a day) for 6 months along with the standard three-drug regimen of macrolide, rifampin, and ethambutol. Dosage adjustments in patients with hepatic and renal dysfunction are not required. Half-life elimination typically occurs in ~5.9–9.5 h. Respiratory side effects such as bronchospasm, cough, dysphonia, and dyspnea are common. Monitoring for systemic aminoglycoside toxicity should be considered.

Imipenem Imipenem primarily inhibits cell-wall biosynthesis by binding to the penicillin-binding proteins. It is rapidly gaining importance for the treatment of *M. abscessus*, with a meta-analysis showing improved outcomes with its inclusion in a multidrug regimen. It is dosed at 500 mg to 1 g twice to three times a day as part of a combination regimen for the treatment of *M. abscessus*. Half-life of imipenem is ~1 h, and because it is metabolized in the kidneys, dosing adjustment is needed with renal dysfunction. Adverse effects include anemia, thrombocythemia, and liver dysfunction.

Cefoxitin Cefoxitin is a second-generation parenteral cephalosporin with activity against rapidly growing NTM, particularly *M. abscessus* and *M. chelonae*. Its mechanism of action against NTM is unknown but may involve inactivation of cell-wall synthesis enzymes. High doses are used for treatment of NTM: 200 mg/kg IV three or four times per day, with a maximal daily dose of 12 g. The half-life of cefoxitin is ~1 h, with primary renal clearance that requires adjustment in renal insufficiency. Adverse effects are uncommon but include gastrointestinal manifestations, rash, eosinophilia, fever, and neutropenia.

Newer Drugs Three newer classes of drugs—the oxazolidinones, the glycylcyclines, and the ketolides—are currently being evaluated for possible use in the treatment of NTM infections, especially those caused by *M. abscessus*. Approximately 50% of *M. abscessus* isolates have shown some degree of susceptibility in vitro to linezolid, an oxazolidinone. Tigecycline, which is a glycylcycline and a tetracycline derivative, and telithromycin, a ketolide, also appear to have in vitro activity against *M. abscessus*. These drugs, however, have not yet been prospectively tested for NTM in patients.

In addition, some anti-TB drugs, including clofazimine and bedaquiline, are being evaluated as alternative agents for the treatment of refractory NTM infections. In particular, clofazimine appears to act synergistically in combination with amikacin, bedaquiline, or tigecycline. The exact role of these agents in the treatment of refractory NTM infections remains unclear. Suppressive therapy with periodic parenteral/oral drugs to limit disease progression and control symptoms may be an appropriate alternative to curative treatment.

CONCLUSION

Treatment of mycobacterial infections requires multiple-drug regimens that often exert significant side effects with the potential to limit tolerability. The prolonged duration of treatment has vastly improved results over those obtained in past decades, but drugs and regimens that will shorten treatment duration and limit adverse drug effects and interactions are needed.

FURTHER READING

Collaborative Group for the Meta-Analysis of Individual Patient Data in Mdr-Tb Treatment–2017: Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: An individual patient data meta-analysis. Lancet 392:821, 2018.

Daley CL et al: Treatment of nontuberculous mycobacterial pulmonary disease: An official ATS/ERS/ESCMID/IDSA clinical practice guideline. Clin Infect Dis 71:e1, 2020.

Nahid P et al: Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America clinical practice guidelines: Treatment of drug-susceptible tuberculosis. Clin Infect Dis 63:e147, 2016.

Sterling TR et al: Guidelines for the treatment of latent tuberculosis infection: Recommendations from the National Tuberculosis Controllers Association and CDC, 2020. MMWR Recomm Rep 69:1, 2020.

World Health Organization: WHO consolidated guidelines on tuberculosis Module 4: Treatment, Drug-resistant tuberculosis treatment Geneva: World Health Organization, 2022.

Section 9 Spirochetal Diseases

187 Syphilis

Sheila A. Lukehart

DEFINITION

Syphilis, a chronic systemic infection caused by *Treponema pallidum* subspecies *pallidum*, is usually sexually transmitted and is characterized by episodes of active disease interrupted by asymptomatic periods (latency). After an incubation period averaging 2–6 weeks, a primary lesion appears—often associated with regional lymphadenopathy—and then resolves without treatment. The secondary stage, with generalized mucosal and cutaneous lesions and generalized lymphadenopathy, also resolves spontaneously and is followed by a latent period of subclinical infection lasting years or decades. Central nervous system (CNS) invasion may occur early in infection, and CNS involvement may be symptomatic or asymptomatic. In the preantibiotic era, one-third of untreated patients developed tertiary syphilis, characterized by destructive mucocutaneous, skeletal, or parenchymal lesions; aortitis; or late CNS manifestations.

ETIOLOGY

The Spirochaetales include five genera that are pathogenic for humans and for a variety of other animals: *Leptospira* species (leptospirosis, **Chap. 189**); *Borrelia* and *Borreliella* species (relapsing fever and Lyme disease, respectively; **Chaps. 190 and 191**); *Brachyspira* species (gastrointestinal infections); and *Treponema* species (syphilis and the endemic treponematoses; **see also Chap. 188**). The *Treponema pallidum* subspecies include *T. pallidum* subsp. *pallidum* (venereal syphilis), *T. pallidum* subsp. *pertenue* (yaws), and *T. pallidum* subsp. *endemicum* (bejel). *T. carateum* (pinta), for which no extant strains are available for molecular studies, is still classified as a separate species. Historically, the pathogenic *Treponema* were distinguished by the clinical syndromes they produce, but phylogenetic analyses of whole genome sequences from several strains (excluding *T. carateum*) yield the three named subspecies groupings. Whether these groupings represent geographical variation or actual biological differences is unclear. The crossing of subspecies boundaries by some "molecular signatures" and the recent recognition of treponemes of the *endemicum* genotype in sexually acquired genital ulcers (chancres) and secondary rashes **(Chap. 188)** support the concept of a genetic and clinical "continuum" among strains and subspecies of the pathogenic treponemes.

T. pallidum subspecies are thin spiral organisms, with a cell body surrounded by a trilaminar cytoplasmic membrane, a delicate peptidoglycan layer, and a lipid-rich outer membrane. Endoflagella wind around the cell body in the periplasmic space and are responsible for motility.

Historically, *T. pallidum* subspecies could not be cultured in vitro, but long-term propagation of multiple strains of *T. pallidum* subsp. *pallidum* and one strain of subsp. *endemicum* in complex medium with eukaryotic cells is now possible. To date, the *pertenue* subspecies has not been cultured. All *T. pallidum* subspecies have severely limited metabolic capabilities and are highly dependent on host-derived amino acids, carbohydrates, and lipids. Genetic analyses have revealed the existence of a 12-member gene family (*tpr*) encoding outer-membrane antigens. One member, TprK, has discrete variable regions that undergo antigenic variation during infection, providing a mechanism for immune evasion and persistence.

The only known natural host for *T. pallidum* subsp. *pallidum* (referred to hereafter as *T. pallidum*) is the human. *T. pallidum* can infect many mammals, but only humans, higher apes, and a few laboratory animals develop syphilitic lesions. Rabbits are used to propagate *T. pallidum* and serve as the animal model that best reflects human disease and immunopathology.

TRANSMISSION AND EPIDEMIOLOGY

Nearly all cases of syphilis are acquired by sexual contact with infectious lesions (i.e., the chancre, mucous patch, skin rash, or condylomata lata; **see Fig. A1-20**). *T. pallidum* DNA has also been detected in swabs of normal-appearing oral mucosa, in saliva, in urine, and in semen, raising the possibility of transmission by these routes, but the infectivity of these organisms has not been assessed. Less common modes of transmission include nonsexual skin contact, infection in utero, blood transfusion, and organ transplantation.

SYPHILIS IN THE UNITED STATES

Following the introduction of penicillin therapy in the 1940s, the number of reported cases of syphilis of all stages in the United States declined 95% to a low of 31,575 cases in 2000, with 5979 reported cases of primary and secondary (P&S) syphilis. (P&S cases are infectious and are a better indicator of disease activity than total syphilis cases.) Since 2000, total cases have increased 6.6-fold to 207,255, and the number of P&S cases has increased tenfold, with 59,061 cases reported in 2022 **(Fig. 187-1)**. Nationally, ~45% of these cases were in men who have sex with men (MSM), ~40% of whom are co-infected with HIV. In two years, from 2020 to 2022, P&S cases rose 31.7% among all men and 85.4% among women, with increases in all racial and ethnic groups and in all geographic regions of the United States. Because the incidence of congenital syphilis parallels that of infectious syphilis in women, the striking increase in early syphilis in women has resulted in a dramatic increase in congenital syphilis. In 2022, 3755 cases of congenital syphilis were reported, resulting in 282 congenital syphilis-related stillbirths and deaths. In the last decade, the number of reported cases in infants <1 year of age has increased from 334 to 3755, or elevenfold. The vast majority of these cases resulted from late or no prenatal care, but some were in women who had tested positive for syphilis but were not treated prior to delivery. A recent study of substance use in pregnant women with syphilis found that illicit use of opioids and other illicit nonprescription substances was six and four times higher, respectively, in persons with a congenital syphilis outcome than in those without a congenital syphilis outcome. Other risk factors include homelessness and unstable housing, transactional sex, and incarceration.

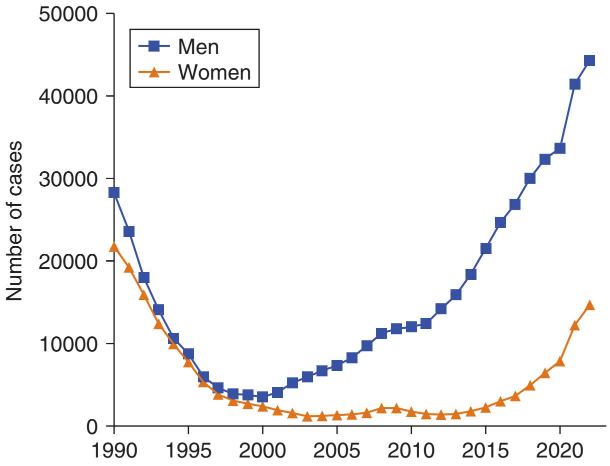

FIGURE 187-1 Primary and secondary syphilis in the United States, 1990–2022, by sex. *(Data from the Centers for Disease Control and Prevention.)*

The populations at highest risk for acquiring syphilis have changed over time, with outbreaks among MSM in the pre-HIV era of the late 1970s and early 1980s, as well as at present. The current dramatic decade-long increase in syphilis and other sexually transmitted infections in MSM may be due to unprotected sex between persons who are HIV concordant and to disinhibition facilitated by highly effective antiretroviral therapy (ART) or pre- and postexposure prophylaxis (PrEP and PEP). Many MSM diagnosed with syphilis have had syphilis previously, and reinfections may be asymptomatic in persons with multiple past episodes. Thus, more frequent (every 3 months) screening for syphilis and other sexually transmitted infections is warranted in high-risk populations. Cases of P&S syphilis among African Americans increased 5.5-fold between 2002 and 2022, and the rate (44.4 per 100,000 population) remains higher than rates for other racial/ethnic groups except American Indians/Alaska Natives (67 per 100,000).

Of individuals named as sexual contacts of persons with early syphilis, some will have developed manifestations of syphilis when they are first seen, and others will develop infectious syphilis if not treated; overall, ~30–60% of persons exposed to P&S syphilis will develop syphilis if not treated. Thus, identification and treatment of all recently exposed sexual contacts continue to be important aspects of syphilis control. Recent data suggest that PEP with 200 mg of doxycycline (Doxy-PEP), taken within 72 h of exposure, will significantly reduce the likelihood of syphilis (and other) infections in MSM. Data from a single study of Doxy-PEP in women have been disappointing, but these results may be due to lack of compliance in the test population. Concerns have been raised that widespread use of Doxy-PEP (perhaps misused as PrEP) may result in selection for resistance in sexually transmitted infection pathogens.

GLOBAL SYPHILIS

Syphilis remains a significant health problem globally; the number of new infections is estimated at ~8 million per year, with 22.3 million prevalent cases. The regions most affected include sub-Saharan Africa, South America, and central and south Asia. Rates of P&S syphilis have increased dramatically among MSM in many European, Asian, and South American countries. Globally, although efforts by the World Health Organization (WHO) reduced the incidence of congenital syphilis in the 2010s, that progress is being lost as recent increases are reported in nearly all regions of the world. In 2022, there were an estimated 700,000 cases of congenital syphilis and 390,000 adverse birth outcomes: 150,000 early fetal deaths and stillbirths, 70,000 neonatal deaths, 55,000 preterm or low-birth-weight infants, and 115,000 infants with clinical evidence of infection.

NATURAL COURSE AND PATHOGENESIS OF UNTREATED SYPHILIS

T. pallidum rapidly penetrates intact mucous membranes or microscopic abrasions in skin and, within a few hours, enters the lymphatics and blood to produce systemic infection and metastatic foci long before the appearance of a primary lesion. Blood from a patient with incubating or early syphilis is infectious. The generation time of *T. pallidum* during early disease is estimated to be ~33 h, and the incubation period of syphilis is inversely proportional to the number of organisms transmitted. The 50% infectious dose for intradermal inoculation in humans has been calculated to be 57 organisms, and the treponeme concentration generally reaches 10^7/g of tissue before a clinical lesion appears. The median incubation period in humans (~21 days) suggests an average inoculum of 500–1000 infectious organisms for naturally acquired disease; the incubation period rarely exceeds 6 weeks.

The primary lesion appears at the site of inoculation, usually persists for 4–6 weeks, and then heals spontaneously. Histopathologic examination shows perivascular infiltration, chiefly by CD4+ and CD8+ T lymphocytes, plasma cells, and macrophages, with capillary endothelial proliferation and subsequent obliteration of small blood vessels. The cellular infiltration produces a T_H1-type cytokine profile, consistent with the activation of macrophages. Phagocytosis of opsonized organisms by activated macrophages ultimately causes their destruction, resulting in spontaneous resolution of the chancre and later the secondary rash.

The generalized parenchymal, constitutional, mucosal, and cutaneous manifestations of secondary syphilis usually appear ~6–12 weeks after infection, although primary and secondary manifestations may occasionally overlap. In contrast, some patients may enter the latent stage without ever recognizing secondary lesions. The histopathologic features of secondary maculopapular skin lesions include hyperkeratosis of the epidermis, capillary proliferation with endothelial swelling in the superficial dermis, and—in the deeper dermis—perivascular infiltration by CD8+ and CD4+ T lymphocytes, macrophages, and variable numbers of plasma cells. *T. pallidum* disseminates during the first days to weeks of infection, invading many tissues, including the CNS; cerebrospinal fluid (CSF) abnormalities can be detected in as many as 40% of patients during the secondary stage. Clinical hepatitis and immune complex–induced glomerulonephritis are rare, but recognized, manifestations of secondary syphilis. Generalized nontender lymphadenopathy is noted in 85% of patients with secondary syphilis. The paradoxical appearance of secondary manifestations, even after the development of an immune response that clears primary lesions, likely results from immune evasion due to antigenic variation of exposed portions of the TprK surface protein. Secondary lesions generally subside within 2–6 weeks, and the infection enters the latent stage, which is detectable only by serologic testing. In the preantibiotic era, up to 25% of untreated patients experienced at least one cutaneous relapse of secondary lesions, usually during the first year. Therefore, identification and examination of sexual contacts are most important for patients with syphilis of <1 year in duration.

In the preantibiotic era, about one-third of patients with untreated latent syphilis developed clinically apparent tertiary disease, the most common types being the gumma (a usually benign granulomatous lesion); cardiovascular syphilis (usually involving the vasa vasorum of the ascending aorta and resulting in aneurysm); and late symptomatic neurosyphilis (tabes dorsalis and paresis). In Western countries today, specific treatment for early and latent syphilis and coincidental therapy (i.e., therapy with antibiotics active against treponemes, but given for other conditions) have nearly eliminated tertiary syphilis. Asymptomatic CNS involvement, however, is still demonstrable in up to 40% of persons with early syphilis and 25% of patients with late latent syphilis, and modern cases of general paresis and tabes dorsalis are being reported from China. The factors that contribute to the development and progression of tertiary disease are unknown.

The course of untreated syphilis was studied retrospectively in a group of nearly 2000 patients with primary or secondary disease diagnosed clinically (the Oslo Study, 1891–1951) and was assessed prospectively in 431 African-American men with seropositive latent syphilis of ≥3 years in duration (the notorious Tuskegee Study, 1932–1972). In the Oslo Study, serious late complications were nearly twice as common among men as among women. In the Tuskegee Study, untreated syphilis increased the death rate 17% compared to uninfected subjects, largely due to cardiovascular syphilis. The ethical issues eventually raised by the Tuskegee Study, begun in the pre-penicillin era but continuing into the early 1970s, had a major influence on the development of current guidelines for protection of human subjects and still contribute to a reluctance of some African Americans to participate in clinical research.

CLINICAL MANIFESTATIONS

PRIMARY SYPHILIS

The typical primary chancre usually begins as a single painless papule that rapidly erodes and becomes indurated, with a characteristic cartilaginous consistency on palpation of the edge and base of the ulcer. Multiple primary lesions are seen in a minority of patients. In heterosexual men, the chancre is usually located on the penis, where it is readily seen **(Fig. 187-2; see also Fig. A1-17)**, but in MSM, it may also be found in the anal canal, rectum, or mouth. Oral sex has been identified as the source of infection in some MSM. In women, common primary sites are the cervix, vaginal wall, and labia, as well as anal canal and mouth. Consequently, primary syphilis goes unrecognized in women and MSM more often than in heterosexual men.

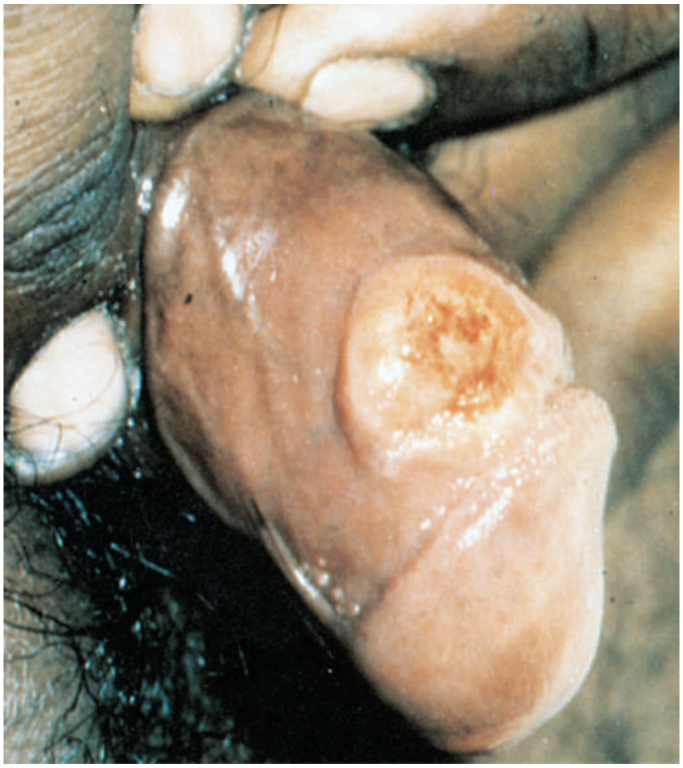

FIGURE 187-2 Primary syphilis with a firm, nontender chancre.

Atypical primary lesions are common, and may be multiple, small, or partially resolved. Therefore, syphilis should be considered in the evaluation of trivial or atypical dark-field-negative genital lesions. The lesions that most commonly must be differentiated from those of primary syphilis include those caused by herpes simplex virus infection **(Chap. 197)**, chancroid **(Chap. 162)**, traumatic injury, and donovanosis **(Chap. 178)**. Regional (usually inguinal) lymphadenopathy accompanies the primary syphilitic lesion, appearing within 1 week of lesion onset. The nodes are firm, nonsuppurative, and painless. Inguinal lymphadenopathy is bilateral and may occur with anal as well as with genital chancres. The chancre generally heals within 4–6 weeks (range, 2–12 weeks), but lymphadenopathy may persist for months.

SECONDARY SYPHILIS

The classical manifestations of the secondary stage include mucocutaneous or cutaneous lesions and generalized nontender lymphadenopathy. The healing primary chancre may still be present in ~15% of cases—more frequently in persons with concurrent HIV infection. The skin rash consists of macular, papular, papulosquamous, and occasionally pustular syphilides; often more than one form is present simultaneously. The eruption may be very subtle, and 25% of patients with a discernible rash may be unaware that they have dermatologic manifestations. Initial lesions are pale red or pink, nonpruritic, discrete macules distributed on the trunk and extremities; these macules progress to papular lesions that are distributed widely and that frequently involve the palms and soles **(Fig. 187-3; see also Figs. A1-18 and and A1-19)**. Rarely, severe necrotic lesions (*lues maligna*) may appear and are more commonly reported in HIV-infected individuals with low CD4+ T-cell counts. Involvement of the hair follicles may result in patchy alopecia of the scalp hair, eyebrows, or beard in up to 5% of cases.

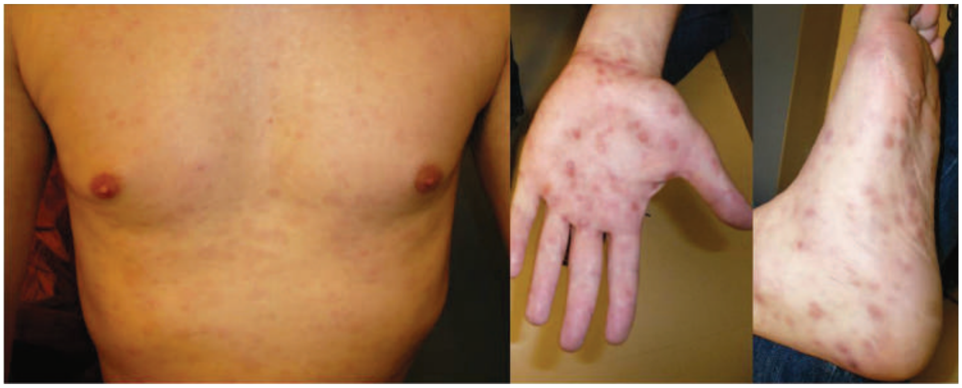

FIGURE 187-3 Secondary syphilis. ***Left:*** Maculopapular truncal eruption. ***Middle:*** Papules on the palms. ***Right:*** Papules on the soles. *(Photos courtesy of Jill McKenzie and Christina Marra.)*

In warm, moist, intertriginous areas (commonly the perianal region, vulva, and scrotum), papules can enlarge to produce broad, moist, pink or gray-white, highly infectious lesions (*condylomata lata*; **see Fig. A1-20**) in 10% of patients with secondary syphilis. Superficial mucosal erosions (*mucous patches*) occur in 10–15% of patients and commonly involve the oral or genital mucosa **(see Fig. A1-21)**. The typical mucous patch is a painless silver-gray erosion surrounded by a red periphery. *T. pallidum* DNA has been detected in oral mucosal swabs from persons with early and latent syphilis, but who have no visible oral lesions. The implications of this finding for transmission are unclear but warrant further research.

Constitutional signs and symptoms that may accompany or precede secondary syphilis include sore throat (15–30%), fever (5–8%), weight loss (2–20%), malaise (25%), anorexia (2–10%), headache (10%), and meningismus (5%). *Acute meningitis* occurs in only 1–2% of cases, but CSF cell and protein concentrations are increased in up to 40% of early syphilis cases, and viable *T. pallidum* organisms have been recovered from CSF during primary and secondary syphilis in 30% of cases, sometimes without other CSF abnormalities.

Persons with current or recent secondary syphilis may present with ocular or otic manifestations. Ocular findings include pupillary abnormalities and optic neuritis as well as the classic iritis or uveitis. The diagnosis of ocular syphilis is often considered in affected patients only after they fail to respond to topical steroid therapy. Anterior uveitis has been reported in 5–10% of patients with secondary syphilis, and *T. pallidum* has been demonstrated in aqueous humor from such patients. Permanent blindness may result without prompt diagnosis and treatment. Otic syphilis may present as sensorineural hearing loss, vertigo, or tinnitus, and deafness may result if untreated. The publication of several reports of ocular and otic syphilis reminds clinicians to inquire about neurologic manifestations in all stages of syphilis infection. In a recent study, 7.9% of patients with syphilis, when asked, reported recent vision or hearing changes, and more than half of those had abnormal CSF or ophthalmologic findings consistent with syphilis.

Less often recognized complications of secondary syphilis include hepatitis, nephropathy, gastrointestinal involvement (hypertrophic gastritis, patchy proctitis, or a rectosigmoid mass—sometimes mistakenly assumed to be malignant), arthritis, and periostitis. Hepatic involvement is common in syphilis; although it is usually asymptomatic, up to 25% of patients may have abnormal liver function tests. Frank syphilitic hepatitis is rare. Renal involvement usually results from immune complex deposition and produces proteinuria associated with an acute nephrotic syndrome. Like those of primary syphilis, most manifestations of the secondary stage resolve spontaneously, usually within 1–6 months.

■ LATENT SYPHILIS

Positive serologic tests for syphilis, together with a normal CSF examination and the absence of clinical manifestations of syphilis, indicate a diagnosis of latent syphilis in an untreated person. The diagnosis may be made following routine serologic screening or may be suspected due to a history of primary or secondary lesions, a history of exposure to syphilis, or the delivery of an infant with congenital syphilis. A previous nonreactive serologic test or clear history of lesions or exposure may help to establish the duration of infection, which is an important factor in the selection of appropriate therapy. *Early latent* syphilis is limited to the first year after infection, whereas *late latent* syphilis is defined as that of ≥1 year in duration or unknown duration. The classical definition of early latent syphilis could include an asymptomatic person whose secondary rash has resolved, as well as a person whose chancre has healed but who has not yet developed secondary manifestations. Accordingly, the Centers for Disease Control and Prevention (CDC) have revised the case definitions for surveillance and reporting purposes to better reflect the recognition that some clinical presentations may be seen at several stages of infection. These definitions include the traditional primary and secondary stages, as well as "syphilis, early nonprimary nonsecondary," describing infections of <12 months in duration, and "syphilis, unknown duration or late," encompassing the previous late latent and late (tertiary) classifications. In this new scheme, neurologic, ocular, otic, and late clinical manifestations are reported separately in the context of their separate primary, secondary, early nonprimary nonsecondary, and unknown duration or late categories.

It was previously thought that untreated late latent syphilis had three possible outcomes: (1) persistent lifelong infection; (2) development of tertiary syphilis; or (3) spontaneous cure, with reversion of serologic tests to negative. Although progression to clinically evident late syphilis is very rare today, the occurrence of spontaneous microbiologic cure is in doubt.

Because *T. pallidum* continues to be present throughout untreated infection, it may seed the bloodstream intermittently during the latent stage, and a pregnant woman with latent syphilis may infect the fetus in utero. Moreover, syphilis has been transmitted through blood transfusion or organ donation from patients with latent syphilis.

■ REINFECTION SYPHILIS

A growing number of individuals, particularly MSM, acquire multiple episodes of syphilis, with important implications for clinical presentation and serologic testing. Although no national data are available, 32% of enrollees (mostly MSM) in an 18-year longitudinal study of CNS involvement were known to have had multiple episodes of syphilis. It is well recognized that, after treatment, persons with past syphilis are less likely to revert to nonreactive in the Venereal Disease Research Laboratory (VDRL)/rapid plasma reagin (RPR) tests than persons with first episode syphilis, and treponemal tests will remain reactive. However, several recent studies also indicate that subsequent episodes of syphilis are more likely to be asymptomatic than initial episodes, less likely to have *T. pallidum* identified in blood or CSF, and less likely to have laboratory-defined neurosyphilis. These cases would be detectable only by serologic screening, reinforcing the utility of frequent screening in high-risk populations to identify reinfection.

■ INVOLVEMENT OF THE CNS

Traditionally, neurosyphilis has been considered a late manifestation of syphilis, but this view is inaccurate. CNS syphilis represents a continuum encompassing early invasion (usually within the first weeks of infection), months to years of asymptomatic involvement, and, in some cases, development of early or late neurologic manifestations. Early neurosyphilis includes asymptomatic or symptomatic meningitis and meningovascular syphilis; late neurosyphilis includes tabes dorsalis and general paresis.

Asymptomatic Neurosyphilis The diagnosis of asymptomatic neurosyphilis is made in patients who lack neurologic symptoms and signs but who have CSF abnormalities, including mononuclear pleocytosis, increased protein concentration, or reactivity in the CSF VDRL test. CSF abnormalities are demonstrated in up to 40% of cases of untreated primary or secondary syphilis and in 25% of cases of untreated latent syphilis. *T. pallidum* has been recovered by inoculation into rabbits of CSF from up to 30% of patients with primary or secondary syphilis but less frequently from patients with syphilis of >1 year in duration. The presence of *T. pallidum* in CSF is often associated with other CSF abnormalities, but organisms can be recovered from patients with otherwise normal CSF. Although the prognostic implications of these findings in early syphilis are uncertain, it may be appropriate to conclude that even patients with early syphilis who have CSF abnormalities do indeed have asymptomatic neurosyphilis and should be treated for neurosyphilis; such treatment is particularly important in patients with concurrent untreated HIV infection. Before the advent of penicillin, the risk of development of clinical neurosyphilis in untreated asymptomatic persons was roughly proportional to the intensity of CSF changes. In several large studies, neurosyphilis was associated with a serum RPR titer of ≥1:32, regardless of clinical stage or HIV infection status. Most experts agree that clinical neurosyphilis is more common among persons with untreated HIV infection, and that the immune reconstitution seen with effective ART may have a protective effect against development of clinical neurosyphilis in some HIV-infected persons with syphilis. Nonetheless, RPR titer ≥1:32 is still associated with reactive CSF VDRL, even in persons taking effective ART. HIV-uninfected persons with untreated latent syphilis and normal CSF probably run a very low risk of subsequent neurosyphilis.

Symptomatic Neurosyphilis The major clinical categories of symptomatic neurosyphilis include early meningeal and meningovascular and late parenchymatous syphilis. The last category includes general paresis and tabes dorsalis. The onset of symptoms usually occurs <1 year after infection for meningeal syphilis, up to 10 years after infection for meningovascular syphilis, at ~20 years for general paresis, and at 25–30 years for tabes dorsalis. Neurosyphilis is more frequently symptomatic in patients co-infected with untreated HIV, particularly those with low CD4+ T lymphocyte counts.

Meningeal syphilis may present as headache, nausea, vomiting, neck stiffness, cranial nerve involvement, seizures, and changes in mental status. This condition may be concurrent with or may follow the secondary stage. Patients presenting with uveitis, iritis, or hearing loss often have meningeal syphilis, but these clinical findings can also be seen in patients with normal CSF.

Meningovascular syphilis reflects meningitis together with inflammatory vasculitis of small, medium, or large vessels. The most common presentation is a stroke syndrome involving the middle cerebral artery of a relatively young adult. However, unlike the usual thrombotic or embolic stroke syndrome of sudden onset, meningovascular syphilis often becomes manifest after a subacute encephalitic prodrome (with headaches, vertigo, insomnia, and psychological abnormalities), which is followed by a gradually progressive vascular syndrome.

The manifestations of *general paresis* reflect widespread late parenchymal damage and include abnormalities corresponding to the mnemonic *paresis: p*ersonality, *a*ffect, *r*eflexes (hyperactive), *e*ye (e.g., Argyll Robertson pupils), *s*ensorium (illusions, delusions, hallucinations), *i*ntellect (a decrease in recent memory and in the capacity for orientation, calculations, judgment, and insight), and *s*peech. *Tabes dorsalis* is a late manifestation of syphilis that presents with symptoms and signs of demyelination of the posterior columns, dorsal roots, and dorsal root ganglia, including ataxia, foot drop, paresthesia, bladder disturbances, impotence, areflexia, and loss of positional, deep-pain, and temperature sensations. The small, irregular Argyll Robertson pupil, a feature of both tabes dorsalis and paresis, reacts to accommodation but not to light. *Optic atrophy* also occurs frequently in association with tabes.

OTHER MANIFESTATIONS OF LATE SYPHILIS

The slowly progressive inflammatory process leading to tertiary disease begins early during infection, although these manifestations may not become clinically apparent for years or decades. Early syphilitic aortitis first becomes evident soon after secondary lesions subside, and treponemes that trigger the development of gummas may have seeded the tissue years earlier.

Cardiovascular Syphilis Cardiovascular manifestations, usually appearing 10–40 years after infection, are attributable to endarteritis obliterans of the vasa vasorum, which provide the blood supply to large vessels; *T. pallidum* DNA has been detected by polymerase chain reaction (PCR) in aortic tissue. Cardiovascular involvement results in uncomplicated aortitis, aortic regurgitation, saccular aneurysm (usually of the ascending aorta), or coronary ostial stenosis. In the preantibiotic era, symptomatic cardiovascular complications developed in ~10% of persons with untreated late syphilis. Today, cardiovascular syphilis is rarely seen in the developed world.

Late Benign Syphilis (Gumma) Gummas are usually solitary lesions ranging from microscopic to several centimeters in diameter. Histologic examination shows a granulomatous inflammation, with a central area of necrosis due to endarteritis obliterans. *T. pallidum* in low numbers have been detected by PCR in these lesions, and penicillin treatment results in rapid resolution, confirming the treponemal stimulus for the inflammation. Common sites include the skin and skeletal system; however, any organ (including the brain) may be involved. Gummas of the skin produce indolent, painless, indurated nodular or ulcerative lesions that may resemble other chronic granulomatous conditions. Skeletal gummas may affect any bone or cartilage. Upper respiratory gummas can lead to perforation of the nasal septum or palate.

CONGENITAL SYPHILIS

Transmission of *T. pallidum* across the placenta from a pregnant person with syphilis to the fetus may occur at any stage of pregnancy, but fetal damage generally does not occur until after the fourth month of gestation when fetal immunologic competence begins to develop. This timing suggests that the pathogenesis of congenital syphilis, like that of adult syphilis, depends on the host immune response rather than on a direct toxic effect of *T. pallidum*. The risk of fetal infection during untreated early maternal syphilis is ~75–95%, decreasing to ~35% for maternal syphilis of >2 years in duration. Adequate treatment of the woman before the 16th week of pregnancy should prevent fetal damage, and treatment before the third trimester should adequately treat the infected fetus. Untreated maternal infection may result in a rate of fetal loss of up to 40% with second-trimester spontaneous abortion, stillbirth, prematurity, and neonatal death. Among infants born alive, only fulminant congenital syphilis is clinically apparent at birth, and these babies have a very poor prognosis. The most common clinical problem is the healthy-appearing baby born to a mother with a positive serologic test.

Routine serologic testing for syphilis in early pregnancy is cost-effective in virtually all populations, even in areas with a low prenatal prevalence of syphilis. Low-tech point-of-care tests have been developed and widely implemented to facilitate antenatal testing in resource-poor settings. Globally, congenital syphilis incidence has increased dramatically, particularly in Africa, South America, and the United States. Periodic lack of benzathine penicillin (BPG) availability in low- and middle-income countries and the current critical shortage in the United States and Europe complicate treatment of seropositive women. Globally, integration of programs to prevent congenital syphilis with programs to prevent maternal transmission of HIV would be highly cost-effective but is often hampered by the restrictions placed on HIV-focused funds.

All pregnant women should be serologically screened at their first antenatal visit. Where the prevalence of syphilis in women is high or when the patient is at high risk of reinfection, testing should be repeated at 28 weeks and at delivery. Those testing positive should be treated immediately, even before receiving results of confirmatory tests. During the current BPG shortage, many clinics are treating nonpregnant patients with doxycycline, thus reserving BPG for seropositive pregnant women. Neonatal congenital syphilis must be differentiated from other generalized congenital infections, including rubella, cytomegalovirus or herpes simplex virus infection, and toxoplasmosis, as well as from erythroblastosis fetalis.

Manifestations of congenital syphilis may appear early (within the first 2 years of life, often at 2–10 weeks of age) or late (after 2 years). The earliest manifestations of congenital syphilis include rhinitis, or "snuffles" (23%); mucocutaneous lesions (35–41%); bone changes (61%), including periostitis detectable by x-ray examination of long bones; hepatosplenomegaly (50%); lymphadenopathy (32%); anemia (34%); jaundice (30%); thrombocytopenia; and leukocytosis. CNS invasion by *T. pallidum* is detectable in 22% of infected neonates. Neonatal death is usually due to pulmonary hemorrhage, secondary bacterial infection, or severe hepatitis. Late congenital syphilis (untreated after 2 years of age) is subclinical in 60% of cases; the clinical spectrum in the remainder of cases may include interstitial keratitis (which occurs at 5–25 years of age), eighth-nerve deafness, and recurrent arthropathy. Neurosyphilis was documented in about one-quarter of untreated patients with late congenital syphilis in the preantibiotic era. Gummatous periostitis occurs at 5–20 years of age and, as in bejel, tends to cause destructive lesions of the palate and nasal septum. Classic stigmata include *Hutchinson's teeth* (centrally notched, widely spaced, peg-shaped upper central incisors), "mulberry" molars (sixth-year molars with multiple, poorly developed cusps), saddle nose, and saber shins.

LABORATORY EXAMINATIONS

DEMONSTRATION OF THE ORGANISM

Historically, dark-field microscopy and immunofluorescence antibody staining have been used to identify *T. pallidum* in moist lesions such as chancres or condylomata lata, but these tests are rarely available outside of research laboratories. Sensitive and specific PCR tests have been developed but are not commercially available, although a number of laboratories perform in-house validated PCR testing. The recent advances in cultivation of *T. pallidum* in a tissue culture system have not yet been implemented in clinical laboratories.

T. pallidum can be found in tissue by immunofluorescence or immunohistochemical methods using specific monoclonal antibodies to *T. pallidum*; some commercial polyclonal antibodies are cross-reactive with other spirochetes and should be avoided. Silver stains should be interpreted with caution because artifacts resembling *T. pallidum* are often seen. *T. pallidum* DNA has been detected by PCR in lesion swabs, tissue samples, blood, CSF, ocular fluid, urine, and oropharyngeal and rectal swabs.

SEROLOGIC TESTS FOR SYPHILIS

Treponemal and Lipoidal Tests There are two types of serologic tests for syphilis: lipoidal (formerly called "nontreponemal") and treponemal. Both are reactive in persons with any treponemal infection, including syphilis, yaws, pinta, and bejel.

The most widely used lipoidal antibody tests are the RPR and VDRL tests, which measure IgG and IgM directed against a cardiolipin-lecithin-cholesterol antigen complex. The RPR test is easier to perform and uses unheated serum or plasma; it is the test of choice for rapid serologic diagnosis in a clinical setting. The VDRL test remains the standard for examining CSF and is superior to the RPR for this purpose. Either test is recommended for screening and for quantitation of serum antibody. The titer generally reflects disease activity, rising during early syphilis, often exceeding 1:32 in secondary syphilis, and declining slowly thereafter without therapy. After treatment for early syphilis, a persistent fall by fourfold or more (e.g., a decline from 1:32 to 1:8) is considered an adequate response. VDRL titers do not correspond directly to RPR titers, and sequential quantitative testing (as for response to therapy) must employ a single test. A reactive VDRL/RPR screening test must be confirmed by a treponemal test to rule out a biological false-positive reaction.

Treponemal tests measure antibodies to native or recombinant *T. pallidum* antigens and include the fluorescent treponemal antibody–absorbed (FTA-ABS) test and the *T. pallidum* particle agglutination (TPPA) test, both of which are more sensitive for primary syphilis than the lipoidal tests. When used to confirm reactive lipoidal test results, treponemal tests have a very high positive predictive value for diagnosis of syphilis.

Treponemal enzyme or chemiluminescence immunoassays (EIAs/CIAs), based largely on reactivity to recombinant antigens, are automated and now used as screening tests by large laboratories. When these tests are used for screening, a high proportion of sera reactive by EIA/CIA are nonreactive by subsequent lipoidal tests. Such sera should be examined in the TPPA test, which includes different antigens and a different platform. If the TPPA test is nonreactive, the patient is unlikely to have syphilis; if it is reactive, the patient is likely to have current or past syphilis.

Both lipoidal and treponemal tests may be nonreactive in early primary syphilis, although treponemal tests are slightly more sensitive (85–90%) during this stage than lipoidal tests (~80%). All tests are reactive during secondary syphilis. (Fewer than 1% of patients with high titers have a lipoidal test that is nonreactive or weakly reactive with undiluted serum but is reactive with diluted serum—the *prozone phenomenon*.) VDRL and RPR sensitivity and titers may decline in untreated persons with late latent syphilis, but treponemal tests remain reactive in late syphilis. After treatment for early syphilis, lipoidal test titers will generally decline or the tests will become nonreactive, whereas treponemal tests often remain reactive after therapy and are not helpful in determining the infection status of persons with past syphilis. There is some concern in the literature about persons in whom the lipoidal test titer fails to become nonreactive or remains reactive in low titer after treatment; this is more commonly seen in persons with repeated episodes of syphilis. The implications in such cases are unclear, but re-treatment rarely achieves the desired goal and is not recommended in the absence of clinical findings.

False-Positive Serologic Tests for Syphilis The lipid antigens of lipoidal tests are similar to those found in human tissues, and these tests may be reactive (usually with titers ≤1:8) in persons without treponemal infection, largely limited to persons with autoimmune conditions or injection drug use. Among patients being screened for syphilis because of risk factors, clinical suspicion, or history of exposure, ~1% of reactive lipoidal tests are falsely positive. In a patient with a false-positive lipoidal test, syphilis is excluded by a nonreactive treponemal test.

False-positive reactions may also occur with treponemal tests, particularly the EIA/CIA tests. Screening a low-prevalence population for syphilis with a treponemal test may result in true-positive reactions being outnumbered by false-positive reactions, leading to unnecessary treatment. Thus, screening with lipoidal tests is highly recommended.

EVALUATION FOR NEURO-, OCULAR, AND OTIC SYPHILIS

Involvement of the CNS is detected by examination of CSF for mononuclear pleocytosis (>5 white blood cells/μL), increased protein concentration (>45 mg/dL), or CSF VDRL reactivity. Elevated CSF cell counts and protein concentrations are not specific for neurosyphilis and may be confounded by HIV co-infection. Because CSF pleocytosis may also be due to HIV, some studies have suggested using a CSF white cell cutoff of 20 cells/μL as diagnostic of neurosyphilis in HIV-infected patients with syphilis. The CSF VDRL test is highly specific and, when reactive, is considered diagnostic of neurosyphilis; however, this test is insensitive and may be nonreactive even in cases of symptomatic neurosyphilis. The RPR test should not be substituted for the VDRL test for CSF examination. The FTA-ABS test on CSF is reactive far more often than the CSF VDRL test in all stages of syphilis, but reactivity may reflect passive transfer of serum antibody into the CSF. A nonreactive FTA-ABS test on CSF, however, may be used to rule out asymptomatic neurosyphilis.

All *T. pallidum*–infected patients with signs or symptoms consistent with neurologic disease (e.g., meningitis, hearing loss) should have a CSF examination, regardless of disease stage. Persons with suspected ophthalmic disease (e.g., uveitis, iritis) should have a thorough ocular examination, with cranial nerve evaluation. Hearing loss, which can occur at any stage of syphilis, may be an isolated finding or may be due to neurosyphilis (involvement of the eighth cranial nerve). If there is no cranial nerve dysfunction in persons with ocular or otic manifestations, no CSF exam is required; conversely, a finding of cranial nerve involvement indicates the need for CSF examination. All persons with ocular or otic syphilis should nonetheless be treated as for neurosyphilis regardless of CSF findings.

The appropriate management of asymptomatic persons is less clear. Lumbar puncture on all asymptomatic patients with untreated syphilis is impractical and unnecessary. Even at high doses, penicillin G benzathine fails to result in treponemicidal drug levels in CSF, and viable *T. pallidum* have been isolated from the CSF of patients (with and without HIV infection) after penicillin G benzathine treatment for early syphilis. Therefore, it is important to identify persons at higher risk for having or developing neurosyphilis so that appropriate treatment may be given. Large-scale prospective studies have shown that patients with RPR titers of ≥1:32 are at higher risk of having neurosyphilis (11-fold and 6-fold higher in HIV-infected and HIV-uninfected persons, respectively), as are HIV-infected patients with CD4+ T-cell counts of ≤350/μL. Persons with active tertiary syphilis and those in whom treatment failure is suspected should also have their CSF examined to determine appropriate therapy.

EVALUATION OF HIV-INFECTED PATIENTS FOR SYPHILIS

Because persons at highest risk for syphilis are also at increased risk for HIV infection, these two infections frequently coexist. There is evidence that syphilis and other genital ulcer diseases are important risk factors for acquisition and transmission of HIV infection. Some manifestations of syphilis may be altered in patients with concurrent untreated HIV infection, and multiple cases of neurologic relapse after standard therapy have been reported in these patients.

Persons with newly diagnosed HIV infection should be tested for syphilis; conversely, all patients with newly diagnosed syphilis should be tested for HIV infection. Some authorities, persuaded by reports of persistent *T. pallidum* in CSF of HIV-infected persons after standard therapy for early syphilis, have recommended CSF examination for evidence of neurosyphilis for all co-infected patients, regardless of the stage of syphilis, with treatment for neurosyphilis if CSF abnormalities are found. Others, on the basis of their own clinical experience, think that standard therapy—without CSF examination—is sufficient for all cases of early syphilis in HIV-infected patients without neurologic signs or symptoms. All persons with HIV infection and early syphilis should receive careful neurologic, ophthalmic, and otologic examinations, including cranial nerve assessment; if cranial nerve dysfunction or neurologic signs or symptoms are found, a CSF examination is warranted to inform treatment.

As described above, RPR titer and CD4+ T-cell count can be used to identify patients at higher risk of neurosyphilis who might benefit from lumbar puncture, although some cases of neurosyphilis will be missed even when these criteria are used. Serologic testing after treatment is important for all patients with syphilis, particularly for those also infected with HIV.

TREATMENT

Syphilis

TREATMENT OF ACQUIRED SYPHILIS

The CDC's 2021 guidelines for the treatment of syphilis are summarized in **Table 187-1** and are discussed below. Penicillin G is the drug of choice for all stages of syphilis. *T. pallidum* is killed by very low concentrations of penicillin G, although a long period of exposure to penicillin is required because of the unusually slow rate of multiplication of the organism. Penicillin G benzathine is the preferred treatment for uncomplicated syphilis, with aqueous penicillin being used for ocular, otic, and neurosyphilis. The efficacy of penicillin against syphilis remains undiminished after 75 years of use, and there is no evidence of penicillin resistance in *T. pallidum*. The current (2023) extreme shortage of penicillin G benzathine has significantly complicated the management and control of the syphilis outbreak in the United States and elsewhere, increasing the need to use alternative antibiotics. Doxycycline, normally the second-line treatment, is now being used widely for men and nonpregnant women with uncomplicated syphilis, reserving penicillin G benzathine for pregnant women. Other antibiotics effective in syphilis include the tetracyclines and the cephalosporins. Aminoglycosides and spectinomycin inhibit *T. pallidum* only in very large doses, and the sulfonamides and most quinolones are inactive. Azithromycin showed significant promise as an effective oral agent against *T. pallidum*; however, strains harboring 23S rDNA mutations that confer macrolide resistance are widespread. Such strains represent >99% of recent isolates from large U.S., European, and Chinese cities, although the prevalence of resistant strains varies by geographic location. Routine treatment of syphilis with azithromycin is not recommended. Careful follow-up of any patient treated for syphilis with azithromycin must be assured. Several additional antibiotics are actively being tested for efficacy against human syphilis infection.

TABLE 187-1 Recommendations for the Treatment of Syphilis[a]

STAGE OF SYPHILIS	PATIENTS WITHOUT PENICILLIN ALLERGY	PATIENTS WITH CONFIRMED PENICILLIN ALLERGY
Primary, secondary, or early latent	*CSF normal or not examined:* Penicillin G benzathine (single dose of 2.4 mU IM) *CSF abnormal:* Treat as neurosyphilis	*CSF normal or not examined:* Doxycycline (100 mg PO bid) or tetracycline HCl (500 mg PO qid) for 2 weeks *CSF abnormal:* Treat as neurosyphilis
Late latent (or latent of unknown duration), cardiovascular, or benign tertiary	*CSF normal or not examined:* Penicillin G benzathine (2.4 mU IM weekly for 3 weeks) *CSF abnormal:* Treat as neurosyphilis	*CSF normal and patient not infected with HIV:* Doxycycline (100 mg PO bid) or tetracycline HCl (500 mg PO qid) for 4 weeks *CSF normal and patient infected with HIV:* Desensitize and treat with penicillin if compliance cannot be assured *CSF abnormal:* Treat as neurosyphilis
Neurosyphilis, ocular syphilis, or otic syphilis	Aqueous crystalline penicillin G (18–24 mU/d IV, given as 3–4 mU q4h or continuous infusion) for 10–14 days *or* Aqueous procaine penicillin G (2.4 mU/d IM) plus oral probenecid (500 mg qid), both for 10–14 days	Desensitize and treat with penicillin
Syphilis in pregnancy	According to stage	Desensitize and treat with penicillin

[a] See text for indications for CSF examination.

Abbreviations: CSF, cerebrospinal fluid; mU, million units.

Source: Adapted from the 2021 Sexually Transmitted Diseases Treatment Guidelines from the Centers for Disease Control and Prevention. Available from *https://www.cdc.gov/std/treatment-guidelines/default.htm.*

Early Syphilis Patients and Their Contacts Penicillin G benzathine is the most widely used agent for the treatment of early syphilis (2.4 million units; Table 187-1) and for preventive treatment of individuals exposed to infectious syphilis within the previous 3 months. *The regimens recommended for prevention are the same as those recommended for early syphilis*. Penicillin G benzathine cures >95% of cases of early syphilis, although clinical relapse can follow treatment, particularly in patients with untreated HIV infection. Because the risk of neurologic relapse may be higher in HIV-infected patients, CSF examination may be recommended for

HIV-seropositive individuals with syphilis of any stage, particularly those with a serum RPR titer of ≥1:32 or a CD4+ T-cell count of ≤350/μL. Therapy appropriate for neurosyphilis should be given if there is any evidence of CNS infection.

Late Latent Syphilis or Syphilis of Unknown Duration If the CSF is normal or is not examined, the recommended treatment is penicillin G benzathine (7.2 million units total; Table 187-1). If CSF abnormalities are found, the patient should be treated for neurosyphilis.

Tertiary Syphilis This category includes persons with gummas ("benign"), cardiovascular syphilis, or signs and symptoms of late neurosyphilis. CSF examination should be performed. If the CSF is normal, the recommended treatment is penicillin G benzathine (7.2 million units total; Table 187-1). If CSF is abnormal, the patient should be treated for neurosyphilis. The clinical response to treatment for benign tertiary syphilis is usually impressive, but responses in cardiovascular syphilis are not dramatic because aortic aneurysm and aortic regurgitation cannot be reversed by antibiotics.

Syphilis in Penicillin-Allergic Patients For penicillin-allergic patients with syphilis, a 2-week (early syphilis) or 4-week (late or late latent syphilis) course of therapy with doxycycline or tetracycline is recommended (Table 187-1). These regimens appear to be quite effective in early syphilis but have not been tested for late or late latent syphilis, and compliance may be problematic. Limited studies suggest that ceftriaxone (1 g/d, given IM or IV for 8–10 days) is effective for early syphilis, and there are reports of the success of 2 g/d for ocular syphilis. These nonpenicillin regimens have not been carefully evaluated in HIV-infected individuals and should be used with caution. If compliance and follow-up are not assured, penicillin-allergic HIV-infected persons with late latent or late syphilis should be desensitized and treated with penicillin.

Neurosyphilis Penicillin G benzathine, even at high doses, does not produce treponemicidal concentrations of penicillin G in CSF and should not be used for treatment of neurosyphilis. Asymptomatic neurosyphilis may relapse as symptomatic disease after treatment with benzathine penicillin, and the risk of relapse may be higher in immunosuppressed HIV-infected patients. Both symptomatic and asymptomatic neurosyphilis should be treated with aqueous penicillin (Table 187-1). Administration of either IV aqueous crystalline penicillin G or of IM aqueous procaine penicillin G plus oral probenecid in recommended doses is thought to ensure treponemicidal concentrations of penicillin G in CSF. The clinical response to penicillin therapy for meningeal syphilis is dramatic, but treatment of neurosyphilis with existing parenchymal damage may only arrest disease progression. No data suggest that additional therapy (e.g., penicillin G benzathine for 3 weeks) would be beneficial after treatment for neurosyphilis.

The use of antibiotics other than penicillin G for the treatment of neurosyphilis has not been studied, although limited data suggest that 1–2 g/d of IV ceftriaxone for 10–14 days and oral doxycycline, 200 mg twice daily for 21 days, may be used. Until further studies confirm these regimens, for patients with confirmed penicillin allergy, desensitization and treatment with penicillin are recommended.

Management of Syphilis in Pregnancy Every pregnant woman should undergo a lipoidal screening serologic test at the first prenatal visit and, if at high risk of re-exposure, again in the third trimester and at delivery. In the untreated pregnant patient with presumed syphilis, expeditious treatment appropriate to the stage of the disease is essential. Patients should be warned of the risk of a Jarisch-Herxheimer reaction, which may be associated with mild premature contractions but rarely results in premature delivery.

Penicillin is the only recommended agent for the treatment of syphilis in pregnancy. If the patient has a documented penicillin allergy, desensitization and penicillin therapy should be undertaken according to the CDC's 2021 guidelines. After treatment during early pregnancy, a quantitative lipoidal test should be repeated 8 weeks following treatment unless reinfection is suspected and again at delivery. Treated women may not achieve a fourfold decline in titer before delivery, but for those in whom antibody titers rise by fourfold, re-treatment is warranted, with careful evaluation of the neonate at birth.

EVALUATION AND MANAGEMENT OF CONGENITAL SYPHILIS

All infants born to women who are untreated or who were treated after a reactive lipoidal serologic test result should be carefully examined at birth for evidence of congenital syphilis. Samples from suspicious lesions and the placenta or umbilical cord should be examined (immunohistochemistry or Clinical Laboratory Improvement Amendments [CLIA]-validated PCR) for presence of *T. pallidum.* Long bone radiographs may provide evidence of the periostitis of congenital syphilis.

Whether or not they are infected, newborn infants of women with reactive serologic tests may themselves have reactive tests because of transplacental transfer of maternal IgG antibodies. If the neonatal titer is fourfold higher than the mother's, infection is indicated and treatment is warranted; CSF examination is recommended. For asymptomatic infants born to women treated adequately with penicillin during the first or second trimester of pregnancy, regular quantitative lipoidal tests may be performed to monitor for appropriate reduction in neonatal antibody titers. Rising or persistent titers indicate infection, and the infant should be treated. Detection of neonatal IgM antibody is insensitive, and no commercially available test is currently recommended.

An infant born to a seropositive mother should be treated at birth if (1) the treatment status of the seropositive mother is unknown; (2) the mother received inadequate or nonpenicillin therapy; (3) the mother received penicillin therapy in the third trimester; or (4) the infant may be difficult to follow. The CSF should be examined to obtain baseline values before treatment. Penicillin is the only recommended drug for the treatment of syphilis in infants. More detailed recommendations for the treatment of infants and older children are included in the CDC's 2021 treatment guidelines.

JARISCH-HERXHEIMER REACTION

A dramatic although self-limited reaction consisting of fever, chills, myalgia, headache, tachycardia, and increased respiratory rate may follow the initiation of treatment for syphilis. This reaction is thought to be a response to lipoproteins released by dying *T. pallidum* organisms. The Jarisch-Herxheimer reaction occurs in ~50% of patients with primary syphilis, 90% of those with secondary syphilis, and a lower proportion of persons with later-stage disease. Defervescence takes place within 12–24 h. In secondary syphilis, erythema and edema of cutaneous lesions may increase. Patients should be warned to expect such developments, which can be managed with symptom-based treatment; steroid therapy is not required for this mild transient reaction.

FOLLOW-UP EVALUATION OF RESPONSES TO THERAPY

Efficacy of treatment should be assessed by clinical evaluation and monitoring of the quantitative VDRL or RPR titer for a fourfold decline (e.g., from 1:32 to 1:8). Patients with primary or secondary syphilis should be examined 6 and 12 months after treatment, and persons with latent or late syphilis at 6, 12, and 24 months. More frequent clinical and serologic examination (3, 6, 9, 12, and 24 months) is recommended for patients concurrently infected with HIV, regardless of the stage of syphilis.

After successful treatment of seropositive first-episode primary or secondary syphilis, the VDRL or RPR titer progressively declines; the test becomes nonreactive by 12 months in 40–75% of seropositive primary cases and in 20–40% of secondary cases. A minority of patients treated for early syphilis may experience a one-dilution titer increase within 14 days after treatment; however, this early elevation does not significantly affect the serologic outcome at 6 months after treatment. In patients with HIV infection or a history of prior syphilis, VDRL and RPR tests are less likely to become nonreactive.

Rates of decline of serologic titers appear to be slower, and serologically defined treatment failures more common among HIV-infected patients than among those without HIV co-infection; however, effective ART may reduce these differences. Re-treatment should be considered if serologic responses are not adequate or if clinical signs persist or recur. The CSF may be examined, with treatment for neurosyphilis if CSF is abnormal and treatment for late latent syphilis if CSF is normal. Patients treated for late latent syphilis frequently have low initial VDRL or RPR titers and may not have a fourfold decline after therapy with penicillin. In such patients, re-treatment is not warranted unless the titer rises or signs and symptoms of syphilis appear. Because treponemal tests may remain reactive despite treatment for seropositive syphilis, these tests are not useful in following the response to therapy.

The activity of neurosyphilis (symptomatic or asymptomatic) correlates best with CSF pleocytosis, and this measure provides the most sensitive index of response to treatment. Repeat CSF examinations may be performed every 6 months until the cell count is normal. An elevated CSF cell count falls to normal in 3–12 months in adequately treated HIV-uninfected patients. The persistence of mild pleocytosis in HIV-infected patients may be due to the presence of HIV in CSF; this scenario may be difficult to distinguish from treatment failure. Elevated levels of CSF protein fall more slowly, and the CSF VDRL titer declines gradually over several years. In patients treated for neurosyphilis, a fourfold reduction in serum RPR titer has been positively correlated with normalization of CSF abnormalities; this correlation is stronger in HIV-uninfected patients and in HIV-infected patients receiving effective ART. Thus, multiple follow-up CSF examinations may not be necessary in patients treated for neurosyphilis in whom serologic titers are falling appropriately.

IMMUNITY TO SYPHILIS

The rate of development of acquired resistance to *T. pallidum* after natural or experimental infection depends on both the size of the infecting inoculum and the duration of infection before treatment. Both humoral and cellular responses are important in the healing of early lesions. Cellular infiltration, predominantly by T lymphocytes and macrophages, produces an interferon γ–dominated cytokine milieu and results in the clearance of organisms by activated macrophages. Specific antibodies to surface antigens enhance phagocytosis. Antigenic variation of the TprK protein contributes to development of subsequent stages of syphilis, persistence of infection, and susceptibility to reinfection with another strain. Comparative genomic studies have revealed genes with sequence variations among *T. pallidum* strains, leading to development of molecular typing methods used to examine syphilis outbreaks. Several laboratories are actively working to develop effective vaccines that stimulate the relevant protective immunologic functions to attenuate or prevent lesion development and reduce dissemination of *T. pallidum* to distant anatomic sites.

FURTHER READING

Hamill M et al: State of the art review: Neurosyphilis. Clin Infect Dis 78:e57, 2024.

Luetkemeyer A et al: Postexposure doxycycline to prevent bacterial sexually transmitted infections. N Engl J Med 388:1296, 2023.

Lukehart S et al: Immunization with a tri-antigen syphilis vaccine significantly attenuates chancre development, reduces bacterial load, and inhibits dissemination of *T. pallidum*. Vaccine 40:7676, 2022.

Tantalo L et al: Antimicrobial susceptibility of *Treponema pallidum* subspecies *pallidum*: An in-vitro study. Lancet Microbe 4:e994, 2023.

Workowski K et al: Sexually transmitted infections treatment guidelines, 2021. MMWR Recomm Rep 70:39, 2021.

188

Endemic Treponematoses

Sheila A. Lukehart, Lorenzo Giacani

The endemic treponematoses are chronic diseases that are generally transmitted by direct contact, usually during childhood and, like syphilis, can cause severe late manifestations years after initial infection. These diseases are caused by spirochete bacteria closely related to *Treponema pallidum* subspecies *pallidum*, the etiologic agent of syphilis **(Chap. 187)**. Yaws, pinta, and bejel (endemic syphilis) have traditionally been distinguished from venereal syphilis by mode of transmission, age of acquisition, geographic distribution, and clinical features; however, there is overlap for each of these factors. Most of our "knowledge" about these infections is based on observations by health care workers who have visited endemic areas during the past 70 years. Except for the ongoing programs of mass drug administration (MDA) for yaws eradication promoted by the World Health Organization (WHO), virtually no well-designed studies of the natural history, diagnosis, or treatment of these infections have been conducted. The four classically defined treponemal infections, modified by current knowledge, are compared in **Table 188-1**.

EPIDEMIOLOGY

Generally, yaws flourishes in moist tropical areas **(Fig. 188-1)**; bejel has been found primarily in arid climates of West Africa and the Middle East; and pinta has been found in temperate foci in the Americas. Because of the ongoing yaws eradication programs, some survey data are available for that disease, but no epidemiologic data are available for bejel and pinta, aside from sparse case reports. Thus, the current extent of these infections is unknown. The endemic treponematoses have traditionally been limited to rural areas of developing nations and have been seen in developed countries primarily among recent immigrants from endemic regions.

In a WHO-sponsored mass eradication campaign from 1952 to 1969, >160 million people in Africa, Asia, and South America were examined for treponemal infections, and >50 million cases, contacts, and persons with latent infections were treated. This campaign reduced the prevalence of active yaws from >20% to <1% in many areas. In subsequent decades, lack of focused surveillance and diversion of resources resulted in the documented resurgence of these infections in some regions. Of nearly 100 countries previously endemic for yaws, there are 16 countries with reported yaws cases during 2017–2022, and 3 others with suspected cases; there are no data for the remaining countries. In 2022, a total of 168,239 suspected cases were reported, primarily from Papua New Guinea, Côte d'Ivoire, Ghana, and Solomon Islands, all countries in which focused or integrated yaws detection programs and treatment trials are ongoing. Other areas of resurgent yaws morbidity in Africa include Cameroon, Togo, Benin, Central African Republic, Nigeria, Congo, Liberia, and Democratic Republic of the Congo. In Asia and the Pacific Islands, reports document active cases of yaws in Papua New Guinea, the Solomon Islands, Timor Leste, Vanuatu, the Philippines, and Indonesia. After years of focused programs, India was declared yaws-free in 2016. In the Americas, suspected yaws cases have been reported in Haiti, Colombia, and Ecuador, with no recent data for Peru, Brazil, Guyana, Suriname, and many Caribbean islands.

The prevalence of bejel is estimated to be >10% in some regions of northern Ghana, Mali, Niger, Burkina Faso, and Senegal, although data are scarce. No data are available from formerly endemic regions of the Middle East. Recent molecular studies, however, have reported genital lesions caused by *T. pallidum* subspecies *endemicum* in persons from Cuba and Japan, and in a person with sexual contact in Pakistan.

Pinta is thought to be limited to Central America and northern South America, where it is found rarely and only in very remote villages. The WHO lists 15 countries in Latin America where pinta was previously endemic; however, due to the lack of surveillance, the current prevalence of pinta is unknown. In the early 1980s, clinical evidence of pinta was discovered in 20% of the examined inhabitants of a

TABLE 188-1 Classic Comparison of the Agents of the Human Treponematoses and Their Associated Diseases

FEATURE	SYPHILIS	YAWS	BEJEL (ENDEMIC SYPHILIS)	PINTA
Organism	*T. pallidum* subsp. *pallidum*	*T. pallidum* subsp. *pertenue*	*T. pallidum* subsp. *endemicum*	*T. carateum*
Common modes of transmission	Sexual, transplacental, skin-to-skin	Skin-to-skin	Mouth-to-mouth or via shared drinking/eating utensils, skin to skin, sexual[a]	Skin-to-skin
Usual age of acquisition	Sexual maturity or in utero	Childhood	Early childhood, adulthood[a]	Late childhood
Primary lesion	Mucocutaneous ulcer (chancre)	Papilloma, often ulcerative	Mucosal papule, rarely seen	Nonulcerating papule with satellites, pruritic
Common location	Genital, oral, anal	Extremities	Oral, occasionally sexual[a]	Extremities, face
Secondary lesions	Cutaneous rash and mucosal lesions; condylomata lata, ocular and otic syphilis	Cutaneous papillomatous or ulcerative lesions; condylomata lata, osteoperiostitis	Mucocutaneous lesions (mucous patch, split papule, condylomata lata); osteoperiostitis	Pintides, pigmented, pruritic
Infectious relapses	~25%	Common	Unknown	Unknown
Late complications	Gummas, cardiovascular and central nervous system involvement	Destructive gummas of skin, bone, cartilage[b]	Destructive gummas of skin, bone, cartilage[b]	Nondestructive, dyschromic, achromic macules

[a]Sexual transmission has been recently postulated for bejel (see text). [b]Central nervous system involvement and congenital infection in the endemic treponematoses have been postulated by some investigators (see text).

remote village in Panama. In 1987 and 1993, pinta cases were reported in indigenous populations living in the Amazon border region of Brazil, Colombia, and Peru. More recently, in 1999, an active pinta lesion was identified in a Cuban tourist visiting Austria, and in 2021, a case of pinta was confirmed in southern Brazil. It is likely, therefore, that pinta is still endemic in some remote areas of Latin America.

It is important to note the vast majority of the data described for the endemic treponematoses are reported as "suspected cases" that are based on clinical diagnosis of ulcers and are not confirmed. Serologic confirmation is only occasionally available, and even reactive treponemal serologies may reflect past or latent treponemal infection that is unrelated to the etiology of the ulcer. Molecular studies of lesion swabs from Papua New Guinea indicate that ~30% of suspected yaws lesions contain *T. pallidum* DNA; most of these had *T. pallidum* DNA detected alone, while some contained both *T. pallidum* and *Haemophilus ducreyi* DNA (dual infection). Approximately 50% of lesions contained DNA from *H. ducreyi*, but not *T. pallidum*. Data from other sites confirm these approximate proportions. Of the 20% of lesions that had neither of these pathogens identified, *Streptococcus pyogenes* was the most abundant organism detected in a metagenomic analysis. Both *H. ducreyi* and *S. pyogenes* have been shown to be common skin and environmental contaminants in yaws-endemic areas but may also serve as primary pathogens or secondary infections. Thus, the accuracy of the diagnosis in "yaws cases" reported to WHO is unclear.

Evidence of yaws-like and genital lesions, with treponemal seroreactivity, has been found in several species of wild nonhuman primates (NHP) in sub-Saharan Africa, providing evidence that there is an animal reservoir for yaws treponemes. At the genomic level, these organisms are virtually identical to known human *T. pallidum* subspecies *pertenue* isolates. Although direct NHP-human transmission has not yet been confirmed, this finding likely has important implications for yaws eradication efforts.

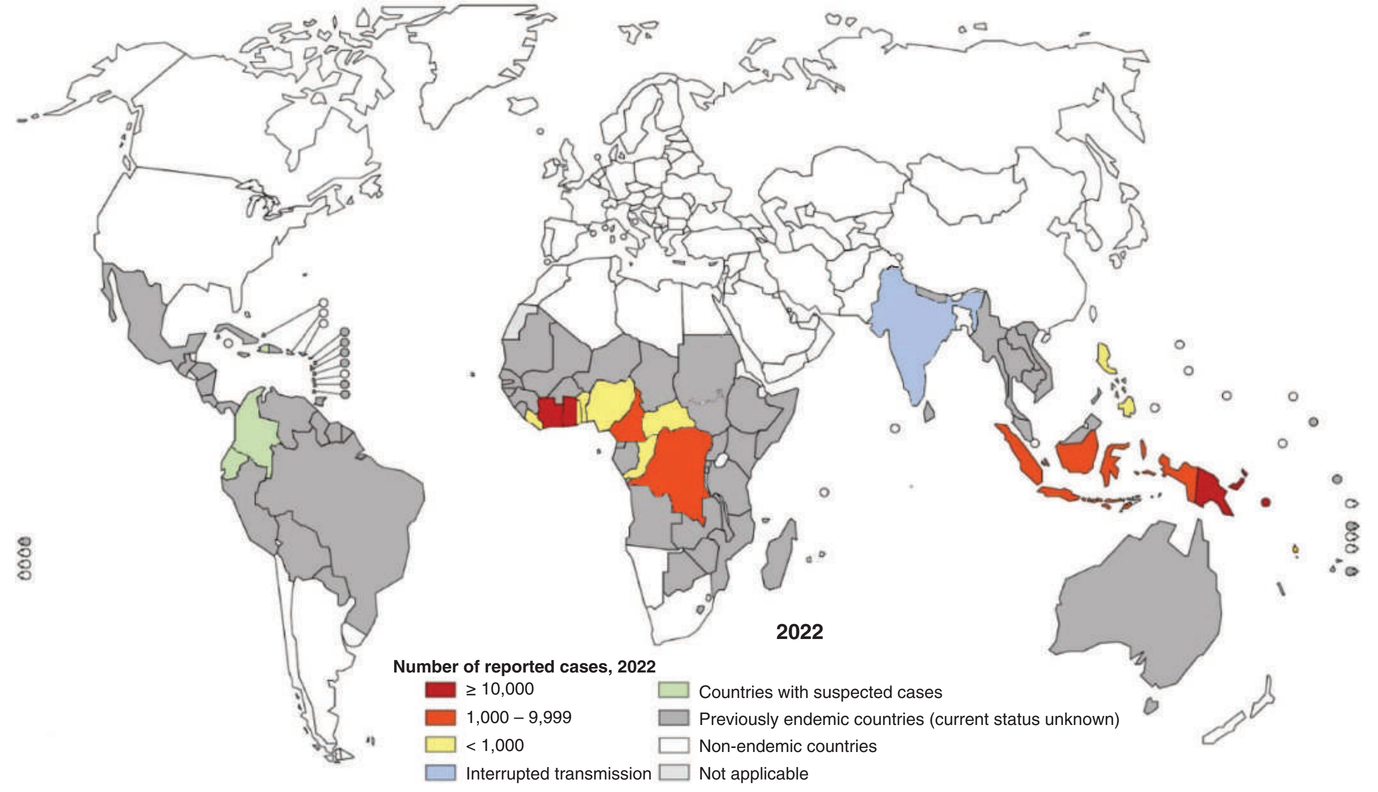
2022
Number of reported cases, 2022
≥ 10,000
1,000 – 9,999
< 1,000
Interrupted transmission
Countries with suspected cases
Previously endemic countries (current status unknown)
Non-endemic countries
Not applicable

FIGURE 188-1 Geographic distribution of yaws in 2022. *(Reproduced with permission from World Health Organization.)*

MICROBIOLOGY

The etiologic agents of the endemic treponematoses are listed in Table 188-1. These little-studied organisms are morphologically identical to *T. pallidum* subspecies *pallidum* (the agent of syphilis), and no definitive antigenic differences among them have been identified to date. A controversy has existed for decades about whether the pathogenic treponemes are truly separate organisms or represent a genetic continuum. Current genome sequencing indicates that, although yaws, bejel, and syphilis isolates clearly cluster in separate branches of phylogenetic trees, these organisms are 99.8% identical at the genomic level, and several studies support the ability of these pathogens to exchange DNA between subspecies. Genomic studies of many more isolates from differing geographic regions are needed to further understand the genomic relationship among the pathogenic treponemes to better inform nomenclature decisions. Currently, three of the four etiologic agents are classified as subspecies of *T. pallidum*; the fourth (*T. carateum*) remains a separate species simply because no pinta organisms have been available for genetic studies. Based on analysis of a limited number of strains and clinical samples available for genetic studies, molecular signatures that can differentiate the known strains of *T. pallidum* subspecies have been identified using approaches ranging from restriction fragment length polymorphism to whole genome sequencing. No obvious genetic polymorphisms have been identified that might be related to any distinct clinical characteristic of these diseases.

CLINICAL FEATURES

All of the treponemal infections, including syphilis, are chronic and are characterized by defined disease stages, with a localized primary lesion, disseminated secondary lesions, periods of latency, and possible late lesions. Primary and secondary stages are more frequently overlapping in yaws and bejel than in syphilis, and the late manifestations of pinta are very mild relative to the destructive lesions of the other treponematoses. The current preference is to divide the clinical course of the endemic treponematoses into "early" and "late" stages, and this terminology is increasingly used for syphilis as well.

Historically, the major clinical distinctions made between syphilis and the other human treponematoses are the apparent lack of congenital transmission and of central nervous system (CNS) involvement in the "nonvenereal" infections. It is not known whether these distinctions are entirely accurate. Because of the high degree of genetic relatedness among the organisms, there is little biologic reason to think that *T. pallidum* subspecies *endemicum* and *T. pallidum* subspecies *pertenue* would be unable to cross the blood–brain barrier or to invade the placenta. These organisms are like *T. pallidum* subspecies *pallidum* in that they obviously disseminate from the site of initial infection and can persist for decades. The lack of recognized congenital infection may be because childhood infections often reach the latent stage (characterized by a low to undetectable bacterial load) before girls reach sexual maturity, thus reducing the likelihood of fetal infection. Neurologic involvement may go unrecognized because of the lack of trained medical personnel in endemic regions, the delay of many years between infection and possible CNS manifestations, or a low rate of symptomatic CNS disease. Some published evidence supports congenital transmission as well as cardiovascular, ophthalmologic, and CNS involvement in yaws and endemic syphilis. Although the reported studies have been small, have failed to control for other causes of CNS abnormalities, and in some instances have not included serologic confirmation, it may be erroneous to accept unquestioningly the frequently repeated belief that these organisms fail to cause such manifestations.

Yaws Also known as *pian*, *framboesia*, or *bouba*, yaws is characterized by the development of one or several primary lesions ("mother yaw") followed by multiple disseminated skin lesions. All early skin lesions are infectious and may persist for many months; cutaneous relapses are common during the first 5 years. Late manifestations, affecting ~10% of untreated persons, are destructive lesions of skin, bone, and joints.

The infection is transmitted by direct contact with infectious lesions, often during play or group sleeping, and may be enhanced by disruption of the skin by insect bites or abrasions. While *T. pallidum* subspecies *pertenue* DNA has been detected on flies and fomites from endemic regions, there is not yet convincing evidence of insect or fomite transmission of infection. After an average of 3–4 weeks, the first lesion begins as a papule—usually on an extremity—and then enlarges (particularly during moist warm weather) to become ulcerated **(Fig 188-2*A*)** or papillomatous ("raspberry-like"—thus the name "framboesia"). Regional lymphadenopathy develops, and the lesion usually heals within 6 months; dissemination is thought to occur during the early weeks of infection. A generalized secondary eruption, accompanied by generalized lymphadenopathy, appears either concurrent with or after the primary lesion, and may take several forms—macular, papular, or papillomatous **(Fig. 188-2*B*)**. Painful papillomatous lesions on the soles of the feet result in a crablike gait ("crab yaws"), and periostitis **(Fig. 188-2*C*)** may result in nocturnal bone pain and polydactylitis **(Fig. 188-2*D*)**. Late yaws is manifested by gummas of the skin and long bones, hyperkeratoses of the palms and soles, osteitis and periostitis, and hydrarthrosis. The late gummatous lesions are characteristically extensive. Destruction of the nose, maxilla, palate, and pharynx is termed *gangosa* and is similar to the destructive lesions seen in leprosy and leishmaniasis.

Bejel The early lesions of bejel (*endemic syphilis*, *siti*, *dichuchwa*, *njovera*, *skerljevo*) are usually localized to mucocutaneous and mucosal surfaces. The infection is reportedly transmitted by direct contact, by kissing, by premastication of food, or by sharing of drinking and eating utensils. Recently, however, *T. pallidum* subspecies *endemicum* has been identified in genital lesions (clinically diagnosed as primary syphilitic chancres) and in secondary lesions in several settings, suggesting sexual transmission. The initial lesion, usually an intraoral papule, may go unrecognized and is followed by mucous patches on the oral mucosa **(Fig. 188-3*A*)** and mucocutaneous lesions resembling the condylomata lata of secondary syphilis. This eruption may last for months or even years, and treponemes can readily be demonstrated in early lesions. Periostitis and regional lymphadenopathy are common. After a variable period of latency, late manifestations may appear, including osseous and cutaneous gummas. Destructive gummas, osteitis, and gangosa are more common in bejel than in yaws.

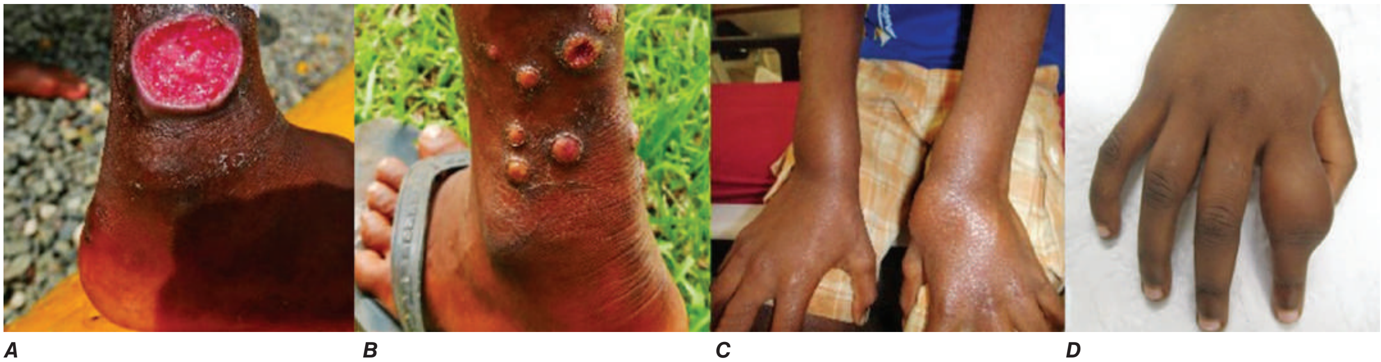

FIGURE 188-2 Clinical manifestations of early yaws. ***A.*** Primary ulcer. ***B.*** Secondary papillomata. ***C.*** Periostitis, ***D.*** Polydactylitis. *(Photos were taken during a yaws elimination trial in Papua New Guinea and are published with permission from Dr. Oriol Mitjà.)*

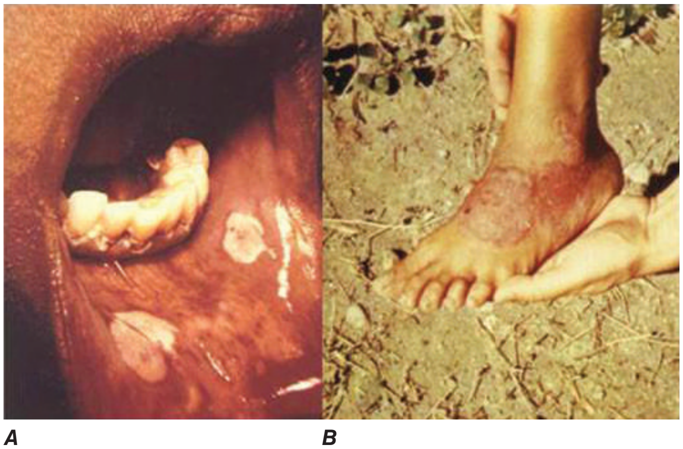

FIGURE 188-3 Clinical manifestations of bejel and pinta. ***A.*** Mucous patches of early bejel. ***B.*** Pigmented macules of early pinta. *(Photos reprinted with permission from the Handbook of Endemic Treponematoses, PL Perine et al, Geneva, World Health Organization, Color Plates 54, 60; 1984.)*

Pinta Pinta (*mal del pinto*, *carate*, *azul*, *purupuru*) is the most benign of the treponemal infections. This disease has three stages that are characterized by marked changes in skin color **(Fig. 188-3*B*)**, but pinta does not appear to cause destructive lesions or to involve tissues other than the skin. The initial papule is most often located on the extremities or face and is pruritic. After 1 to many months of infection, numerous disseminated secondary lesions (*pintides*) appear. These lesions are initially red but become deeply pigmented, ultimately turning a dark slate blue. The secondary lesions are infectious and highly pruritic and may persist for years. Late pigmented lesions are called *dyschromic macules* and contain treponemes. Over time, most pigmented lesions show varying degrees of depigmentation, becoming brown and eventually white and giving the skin a mottled appearance. White achromic lesions are characteristic of the late stage.

DIAGNOSIS

Diagnosis of the endemic treponematoses is based on clinical manifestations and, when available, dark-field microscopy and serologic testing. The same serologic tests—detecting antibodies to either lipoidal or treponemal antigens—that are used for syphilis **(Chap. 187)** become reactive during all treponemal infections. To date there is no antibody test that can discriminate among the treponemal infections. The nonsyphilis treponemal infections should also be considered in the evaluation of a reactive syphilis serology in any person who has emigrated from an endemic area. Sensitive nucleic acid amplification–based assays can be used to confirm treponemal infection and to identify the etiologic agent in research and selected clinical laboratories.

TREATMENT

Endemic Treponematoses

The current WHO-recommended therapy for patients and their contacts includes either azithromycin (30 mg/kg, up to a maximum of 2 g) or benzathine penicillin G (1.2 million units IM for adults; 600,000 units for children <10 years old); these two drugs have been shown to be equivalent for early yaws. The recommended dose of benzathine penicillin G is half of that recommended for early syphilis, yet no controlled efficacy studies have been conducted. Evidence of genetic resistance to penicillin is lacking, although relapsing lesions have been reported after penicillin treatment in Papua New Guinea.

The efficacy of single-dose azithromycin provided the WHO's revitalized yaws eradication program with a much easier regimen for use in mass treatment. Macrolide resistance has become common in circulating strains of *T. pallidum* subspecies *pallidum* in many parts of the world **(Chap. 187)**, and analyses of yaws samples from Papua New Guinea and elsewhere have yielded evidence of mutations conferring resistance to macrolide antibiotics, including azithromycin, in a small number of treated patients. Careful molecular surveillance is essential to monitor developing resistance in yaws-endemic areas. Limited data suggest the efficacy of tetracycline for treatment of yaws, but no data exist for other endemic treponematoses. Based solely on experience with syphilis, it is likely that doxycycline, tetracycline (at doses appropriate for syphilis; **Chap. 187**), or ceftriaxone are alternatives, in addition to azithromycin, for patients allergic to penicillin. A Jarisch-Herxheimer reaction **(Chap. 187)** may follow treatment of endemic treponematoses. Lipoidal serologic titers (in the Venereal Disease Research Laboratory [VDRL] slide test or the rapid plasma reagin [RPR] test) usually decline after effective therapy, but patients may not become seronegative.

CONTROL

Buoyed by the successful elimination of yaws in India and the availability of an inexpensive, single-dose oral drug for treatment, in 2012, the WHO renewed its efforts to eradicate yaws globally by 2020. Based on the results of several pilot programs of MDA, however, the target year for eradication was extended to 2030. Initial enthusiasm has been dampened by several factors: (1) Pilot studies have indicated that a very high level of MDA coverage must be achieved and that multiple rounds of MDA are needed in the affected areas. Treatment must be followed by careful case detection and targeted treatment of cases and contacts. (2) Azithromycin resistance has emerged during MDA studies in Papua New Guinea. Although subsequent treatment with benzathine penicillin G was able to contain the spread of resistant organisms, such evidence suggests that there may be only a short window of time during which countries can successfully use azithromycin for yaws eradication. Antibiotic resistance is of particular concern because multiple rounds of MDA are likely to be required. Further, given the ongoing campaigns against trachoma using low-dose azithromycin MDA, often in populations also at high risk for yaws, more widespread macrolide resistance seems inevitable. (3) Lastly, the possible animal reservoir needs to be evaluated, particularly in Africa. Yaws elimination will require rapid implementation and scale-up of high-level drug coverage in endemic areas, and continued careful surveillance by local health centers will be essential for success of this timely and important effort.

FURTHER READING

Giacani L, Lukehart SA: The endemic treponematoses. Clin Microbiol Rev 27:89, 2014.

Janecková K et al: The genomes of the yaws bacterium, *Treponema pallidum* subsp. *pertenue*, of nonhuman primate and human origin are not genomically distinct. PLoS Negl Trop Dis 17:e0011602, 2023.

John LN et al: Trial of three rounds of mass azithromycin administration for yaws eradication. N Engl J Med 386:47, 2022.

189 Leptospirosis

Marco Goeijenbier, Jiří F. P. Wagenaar

Leptospirosis is a globally important zoonotic disease whose apparent reemergence is illustrated by recent outbreaks on virtually all continents. The disease is caused by pathogenic *Leptospira* species and is characterized by a broad spectrum of clinical manifestations, varying from asymptomatic infection to fulminant, fatal disease. In its mild form, leptospirosis may present as nonspecific symptoms such as fever, headache, and myalgia. Severe leptospirosis, characterized by the triad of jaundice, renal dysfunction, and hemorrhagic diathesis, is often referred to as *Weil's syndrome*. With or without jaundice, severe pulmonary hemorrhage is increasingly recognized as an important presentation of severe disease.

ETIOLOGIC AGENT

Leptospira species are spirochetes belonging to the order Spirochaetales and the family Leptospiraceae. Traditionally, the genus *Leptospira* comprised two species: the pathogenic *L. interrogans* and the free-living *L. biflexa*, now designated *L. interrogans* sensu lato and *L. biflexa* sensu lato, respectively. Sixty-four *Leptospira* species with pathogenic (17 species), intermediate (21 species), and nonpathogenic (26 species) status have now been described based on phylogenetic analyses **(Fig. 189-1)**. Genome sequences of all *Leptospira* species have been published, and this will undoubtedly lead to a better understanding of the pathogenesis of leptospirosis. However, classification based on serologic differences better serves clinical, diagnostic, and epidemiologic purposes. Pathogenic *Leptospira* species are divided into serovars according to their antigenic composition. There are more than 260 known pathogenic serovars, which are arranged in 26 serogroups.

Leptospires are coiled, thin, highly motile organisms that have hooked ends and two periplasmic flagella, with polar extrusions from the cytoplasmic membrane that are responsible for motility **(Fig. 189-2)**. These organisms are 6–20 μm long and ~0.1 μm in diameter; they stain poorly but can be seen microscopically by dark-field examination and after silver impregnation staining of tissues. Leptospires require special media and conditions for growth; it may take weeks to months for cultures to become positive.

EPIDEMIOLOGY

Leptospirosis has a worldwide distribution. Infection occurs most commonly in the tropics and subtropics because the climate and occasionally poor hygienic conditions favor the pathogen's survival and distribution. In most countries, leptospirosis is an underappreciated problem. Most cases occur in men, with a peak incidence during the summer and fall in both the Northern and Southern Hemispheres and during the rainy season in the tropics.

Reliable data on morbidity and mortality from leptospirosis have gradually started to appear. Current information on global human leptospirosis varies but indicates that ~1 million severe cases occur per year, with a mean case–fatality rate of nearly 10%.

As a zoonosis, leptospirosis affects almost all mammalian species and represents a significant veterinary burden. Rodents, especially rats, are the most important reservoir, although other wild mammals as well as domestic and farm animals may also harbor these microorganisms. Leptospires establish a symbiotic relationship with their host and can persist in the urogenital tract for years. Some serovars are generally associated with specific animals—e.g., Icterohaemorrhagiae and Copenhageni with rats, Grippotyphosa with voles, Hardjo with cattle, Canicola with dogs, and Pomona with pigs—but may occur in other animals as well.

Leptospirosis presents as both an endemic and an epidemic disease. Transmission of leptospires may follow direct contact with urine, blood, or tissue from an infected animal or, more commonly, exposure to environmental contamination. The dogma that human-to-human transmission is very rare is challenged by recent findings on household clustering, asymptomatic renal colonization, and prolonged excretion of leptospires. Both of the latter features could imply human infection sources that are not recognized. Because leptospires can survive in a humid environment for many months, water is an important vehicle in their transmission. Epidemics of leptospirosis are not well understood. Outbreaks may result from exposure to floodwaters contaminated by urine from infected animals, as has been reported from several countries. However, it is also true that outbreaks may occur without floods, and floods often occur without outbreaks.

The vast majority of infections with *Leptospira* cause no or only mild disease in humans. A small percentage of infections (~1%) lead to severe, potentially fatal complications. The proportion of leptospirosis cases that are mild is unknown because patients either do not seek or do not have access to medical care or because the nonspecific symptoms are interpreted as an influenza-like illness. Reported cases surely represent a significant underestimation of the total number. Certain occupational groups are at especially high risk, including veterinarians, agricultural workers, sewage workers, slaughterhouse employees, and workers in the fishing industry. Risk factors include direct or indirect contact with animals, including exposure to water and soil contaminated with animal urine. Leptospirosis has also been recognized in deteriorating inner cities and suburban areas where rat and mouse populations are expanding.

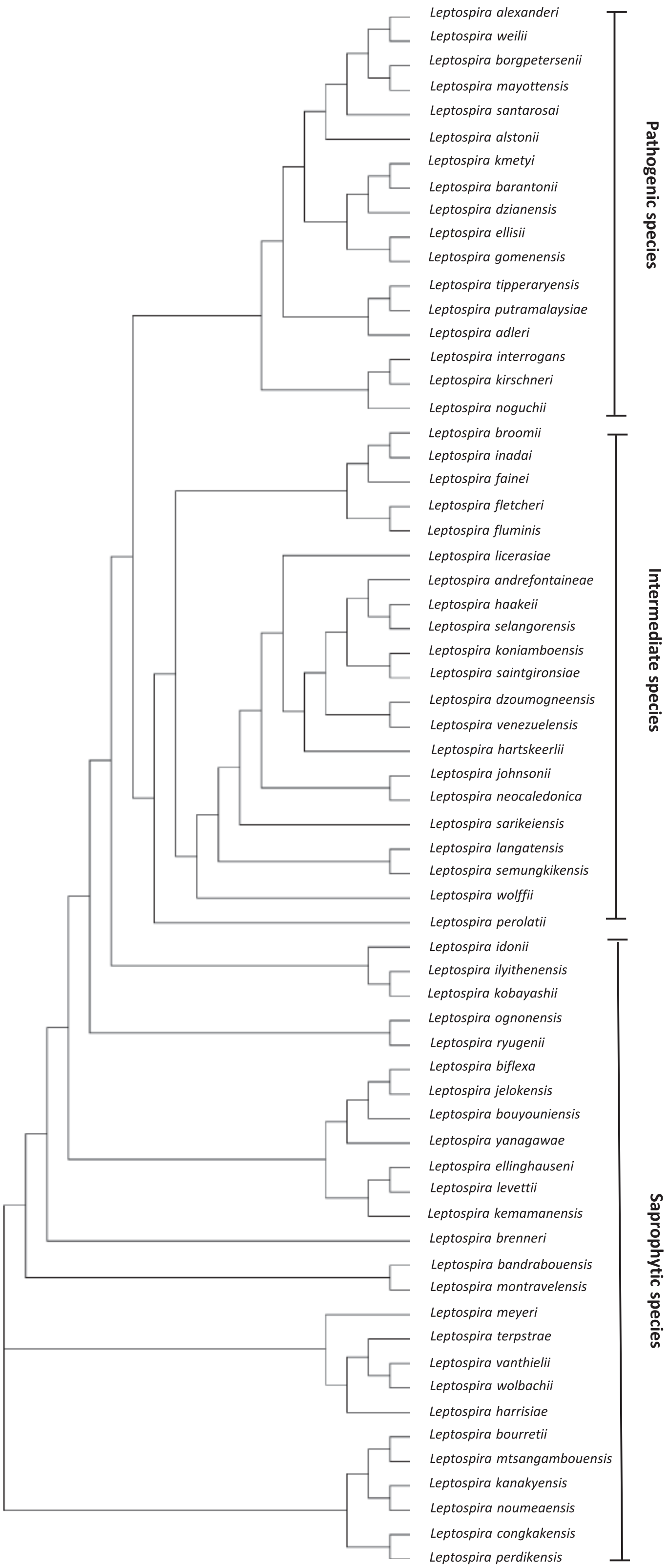
Leptospira alexanderi
Leptospira weilii
Leptospira borgpetersenii
Leptospira mayottensis
Leptospira santarosai
Leptospira alstonii
Leptospira kmetyi
Leptospira barantonii
Leptospira dzianensis
Leptospira ellisii
Leptospira gomenensis
Leptospira tipperaryensis
Leptospira putramalaysiae
Leptospira adleri
Leptospira interrogans
Leptospira kirschneri
Leptospira noguchii
Leptospira broomii
Leptospira inadai
Leptospira fainei
Leptospira fletcheri
Leptospira fluminis
Leptospira licerasiae
Leptospira andrefontaineae
Leptospira haakeii
Leptospira selangorensis
Leptospira koniamboensis
Leptospira saintgironsiae
Leptospira dzoumogneensis
Leptospira venezuelensis
Leptospira hartskeerlii
Leptospira johnsonii
Leptospira neocaledonica
Leptospira sarikeiensis
Leptospira langatensis
Leptospira semungkikensis
Leptospira wolffii
Leptospira perolatii
Leptospira idonii
Leptospira ilyithenensis
Leptospira kobayashii
Leptospira ognonensis
Leptospira ryugenii
Leptospira biflexa
Leptospira jelokensis
Leptospira bouyouniensis
Leptospira yanagawae
Leptospira ellinghauseni
Leptospira levettii
Leptospira kemamanensis
Leptospira brenneri
Leptospira bandrabouensis
Leptospira montravelensis
Leptospira meyeri
Leptospira terpstrae
Leptospira vanthielii
Leptospira wolbachii
Leptospira harrisiae
Leptospira bourretii
Leptospira mtsangambouensis
Leptospira kanakyensis
Leptospira noumeaensis
Leptospira congkakensis
Leptospira perdikensis
Pathogenic species
Intermediate species
Saprophytic species

FIGURE 189-1 Differentiation of pathogenic, intermediate, and nonpathogenic (saprophytic) *Leptospira* species by molecular phylogenetic analysis using core genomes comparison (CgMLST). *(Reproduced with permission from Dr. A Ahmed, Leptospirosis Reference Center, Academic Medical Center, Medical Microbiology, Amsterdam, The Netherlands.)*

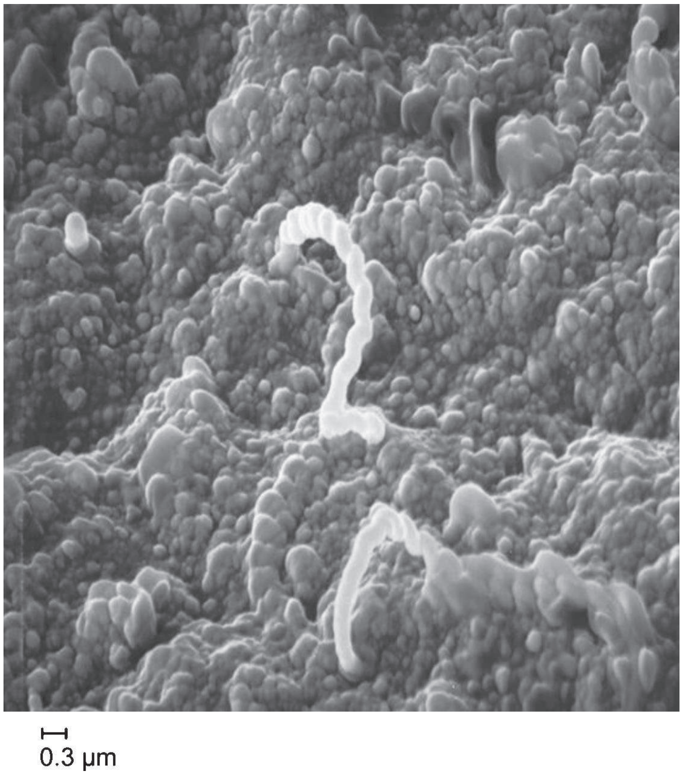

FIGURE 189-2 Transmission electron microscopic image of *Leptospira interrogans* invading equine conjunctival tissue. *(Image kindly provided by Dr. JE Nally, National Animal Disease Center, U.S. Department of Agriculture, Ames, IA.)*

Recreational exposure and domestic-animal contact are prominent sources of leptospirosis. Recreational freshwater activities, such as canoeing, windsurfing, swimming, and waterskiing, place persons at risk for infection. Also, several outbreaks have followed sporting events. For example, an outbreak took place in 1998 among athletes after a triathlon in Springfield, Illinois. Ingestion of one or more swallows of lake water during the swimming leg of the triathlon was a prominent risk factor for illness. Heavy rains that preceded the triathlon, with consequent agricultural runoff, are likely to have increased the level of leptospiral contamination in the lake water. In another outbreak, 42% of participants contracted leptospirosis during the 2000 Eco-Challenge-Sabah multisport endurance race in Malaysian Borneo. Swimming in the Segama River was shown to be an independent risk factor. Furthermore, outbreaks among athletes participating in the recently popular mud-runs are increasingly reported.

In addition, leptospirosis is a traveler's disease. Large proportions of patients acquire the infection while traveling in tropical countries, usually during adventurous activities such as whitewater rafting, jungle trekking, and caving or spelunking. Recent data from the GeoSentinel Global Surveillance Network described in detail 180 returned travelers (mostly male; 74%) with leptospirosis from January 1997 through December 2016. Infection was predominantly acquired in Southeast Asia (52% [n = 93]; mainly [n = 52] from Thailand); overall, 110 patients (59%) were hospitalized, and one patient died. Transmission via laboratory accidents has been reported but is rare. New data indicate that leptospirosis may develop after unanticipated immersion in contaminated water (e.g., in an automobile accident) more frequently than has generally been thought and can also result from an animal bite.

■ PATHOGENESIS

Transmission occurs through cuts, abraded skin, or mucous membranes, especially the conjunctival and oral mucosa. After entry, the highly motile organisms proliferate, cross tissue barriers, and disseminate hematogenously to all organs (leptospiremic phase). During this initial incubation period, leptospires can be isolated from the bloodstream **(Fig. 189-3)**. Clearly, *Leptospira* can survive in the nonimmune host by evading parts of the innate immune response such as complement-mediated killing and phagocytosis; however, earlier studies have highlighted the relation between an exaggerated proinflammatory immune response and mortality. During the immune phase, the appearance of antibodies coincides with the disappearance of leptospires from the blood. However, the bacteria persist in various organs, including liver, lung, kidney, heart, and brain. Autopsy findings illustrate the involvement of multiple organ systems in severe disease. Renal pathology shows both acute tubular damage and interstitial nephritis. Acute tubular lesions progress in time to interstitial edema and acute tubular necrosis. Severe nephritis is observed in patients who survive long enough to develop it and seems to be a secondary response to acute epithelial damage. The reported deregulation of the expression of several transporters along the nephron contributes to impaired sodium absorption, tubular potassium wasting, and polyuria. Histopathology of the liver shows focal necrosis (widespread hepatocellular necrosis is usually not found), foci of inflammation, and plugging of bile canaliculi. Hepatocyte apoptosis has also been documented. Experimental work showed infiltration of *Leptospira* in Disse space (perisinusoidal space) and migration between hepatocytes with detachment of the

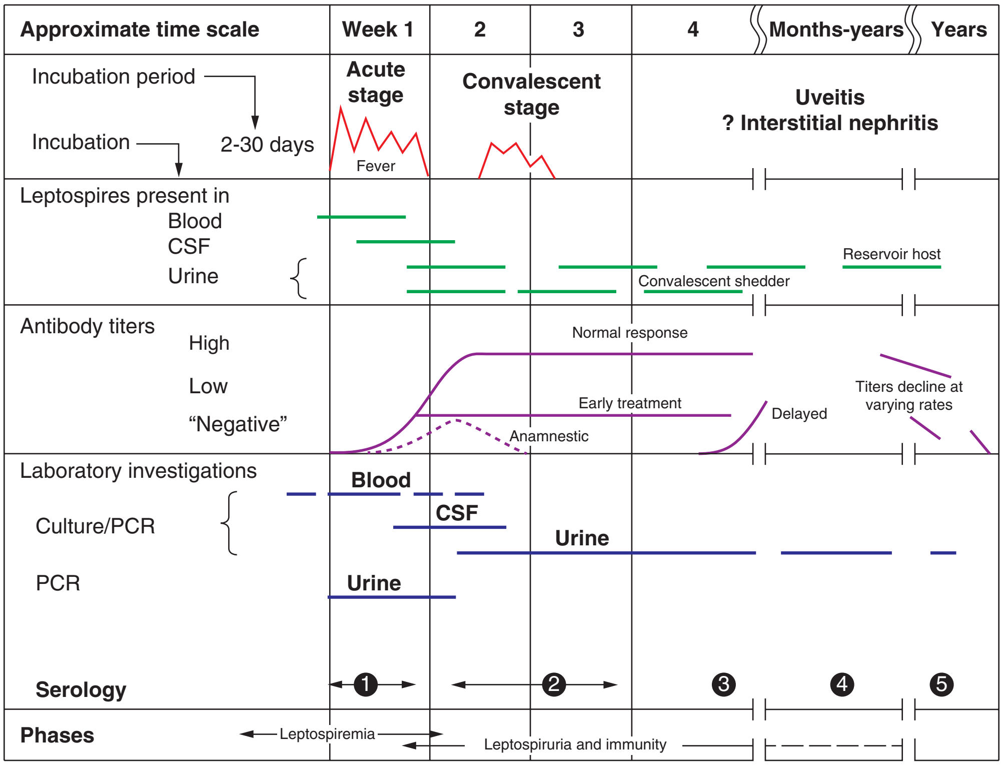

FIGURE 189-3 Biphasic nature of leptospirosis and relevant investigations at different stages of disease. Specimens 1 and 2 for serology are acute-phase serum samples; specimen 3 is a convalescent-phase serum sample that may facilitate detection of a delayed immune response; and specimens 4 and 5 are follow-up serum samples that can provide epidemiologic information, such as the presumptive infecting serogroup. CSF, cerebrospinal fluid. *(Used with permission of [ASM], from Leptospirosis, PN Levett, 14:296, 2001; permission conveyed through Copyright Clearance Center, Inc. and Turner, Leptospirosis I, Transactions of the Royal Society of Tropical Medicine & Hygiene, 61:842, 1967. Permission is granted as per the terms of the STM Permissions Guidelines. Reproduced by permission of Oxford University Press on behalf of the Royal Society of Tropical Medicine & Hygiene.)*

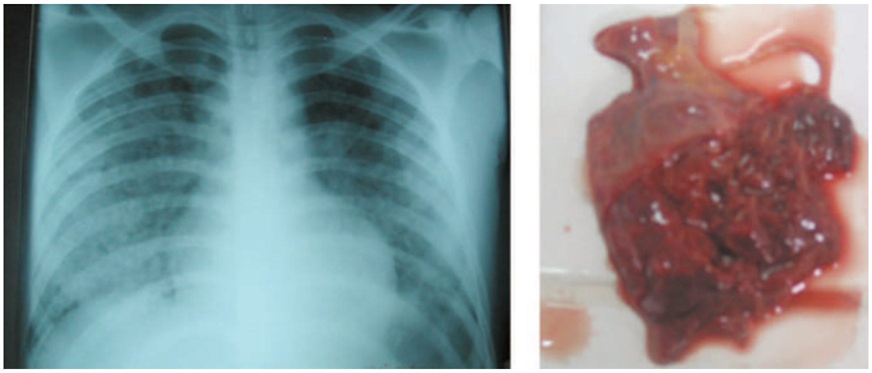

FIGURE 189-4 Severe pulmonary hemorrhage in leptospirosis. ***Left panel:*** Chest x-ray. ***Right panel:*** Gross appearance of right lower lobes of lung at autopsy. This patient, a 15-year-old from the Peruvian Amazonian city of Iquitos, died several days after presentation with acute illness, jaundice, and hemoptysis. Blood culture yielded *Leptospira interrogans* serovar Copenhageni/Icterohaemorrhagiae. *(Adapted with permission from E Segura et al: Clin Infect Dis 40:343, 2005. © 2005 by the Infectious Diseases Society of America.)*

intercellular junctions and disruption of bile canaliculi leading to bile leakage. Petechiae and hemorrhages are observed in the heart, lungs (Fig. 189-4), kidneys (and adrenals), pancreas, liver, gastrointestinal tract (including retroperitoneal fat, mesentery, and omentum), muscles, prostate, testes, and brain (subarachnoid bleeding). Several studies show an association between hemorrhage and thrombocytopenia. Although the underlying mechanisms of thrombocytopenia have not been elucidated, it seems likely that platelet consumption plays an important role. A consumptive coagulopathy may occur, with elevated markers of coagulation activation (thrombin–antithrombin complexes, prothrombin fragments 1 and 2, D-dimer), diminished anticoagulant markers (antithrombin, protein C), and deregulated fibrinolytic activity. Overt disseminated intravascular coagulation (DIC) has been documented in several clinical studies. Elevated plasma levels of soluble E-selectin and von Willebrand factor in patients with leptospirosis reflect endothelial cell activation. More specifically, markers of endothelial cell activation correlate to disease severity in patients with severe leptospirosis. Experimental models show that pathogenic leptospires or leptospiral proteins are able to activate endothelial cells in vitro and to disrupt endothelial-cell barrier function, thus increasing permeability and promoting dissemination. Platelets have been shown to aggregate on activated endothelium in the human lung, whereas histology reveals swelling of activated endothelial cells but no evident vasculitis or necrosis. Immunoglobulin and complement deposition have been demonstrated in lung tissue involved in pulmonary hemorrhage.

Leptospira species have a typical double-membrane cell wall structure harboring a variety of membrane-associated proteins, including an unusually high number of lipoproteins. The peptidoglycan layer is located close to the cytoplasmic membrane. The lipopolysaccharide (LPS) in the outer membrane has an unusual structure with relatively low endotoxic potency. However, host immunity depends on the production of circulating antibodies to serovar-specific LPS. It is unclear whether other antigens play a significant role in protective humoral immunity.

Pathogenic *Leptospira* contain a variety of genes coding for proteins involved in motility and in cell and tissue adhesion and invasion that represent (potential) virulence factors. Many of these are surface-exposed outer-membrane proteins (OMPs). It is likely that several surface-exposed proteins mediate pathogen–host cell interactions, and these proteins may represent candidate vaccine components. Although animal-model studies have shown various degrees of vaccine efficacy for various putative virulence-associated OMPs, it is not yet clear whether such proteins elicit acceptable levels of sterilizing immunity. Ongoing breakthroughs in genetic manipulation of *Leptospira* and whole genome sequencing will undoubtedly provide more insight into the biology and virulence of this pathogen.

■ CLINICAL MANIFESTATIONS

Although leptospirosis is a potentially fatal disease with bleeding and multiorgan failure as its clinical hallmarks, the majority of cases are thought to be relatively mild, presenting as the sudden onset of a febrile illness. The incubation period is usually 1–2 weeks but ranges from 2 to 30 days. Leptospirosis is classically described as biphasic. The acute *leptospiremic phase* is characterized by fever of 3–10 days' duration, during which time the organism can be cultured from blood and detected by polymerase chain reaction (PCR). During the *immune phase*, resolution of symptoms may coincide with the appearance of antibodies, and leptospires can be cultured from the urine. The distinction between the first and second phases is not always clear: milder cases do not always include the second phase, and severe disease may be monophasic and fulminant. The idea that distinct clinical syndromes are associated with specific serogroups has been refuted, although some serovars tend to cause more severe disease than others.

Mild Leptospirosis Most patients are asymptomatic or only mildly ill and do not seek medical attention. Serologic evidence of past inapparent infection is frequently found in persons who have been exposed but have not become ill. Mild symptomatic leptospirosis usually presents as a flulike illness of sudden onset, with fever, chills, headache, nausea, vomiting, abdominal pain, conjunctival suffusion (redness without exudate), and myalgia. Muscle pain is intense and especially affects the calves, back, and abdomen. The headache is intense, localized to the frontal or retroorbital region (resembling that occurring in dengue), and sometimes accompanied by photophobia. Aseptic meningitis may be present and is more common among children than among adults. Although *Leptospira* can be cultured from the cerebrospinal fluid (CSF) in the early phase, the majority of cases follow a benign course with regard to the central nervous system; symptoms disappear within a few days but may persist for weeks.

Physical examination may include any of the following findings, none of which is pathognomonic for leptospirosis: fever, conjunctival suffusion, pharyngeal injection, muscle tenderness, lymphadenopathy, rash, meningismus, hepatomegaly, and splenomegaly. If present, the rash is often transient; may be macular, maculopapular, erythematous, or hemorrhagic (petechial or ecchymotic); and may be misdiagnosed as due to scrub typhus or viral infection. Lung auscultation may reveal crackles. Mild jaundice may be present.

The natural course of mild leptospirosis usually involves spontaneous resolution within 7–10 days, but persistent symptoms have been documented. In the absence of a clinical diagnosis and antimicrobial therapy, the mortality rate in mild leptospirosis is low.

Severe Leptospirosis Although the onset of severe leptospirosis may be no different from that of mild leptospirosis, severe disease is often rapidly progressive and is associated with a case–fatality rate ranging from 1 to 50%. Higher mortality rates are associated with an age >40 years, altered mental status, acute renal failure, respiratory insufficiency, hypotension, and arrhythmias. The classic presentation, often referred to as *Weil's syndrome*, encompasses the triad of hemorrhage, jaundice, and acute kidney injury.

Patients die of multiorgan failure after septic shock and/or severe bleeding complications that most commonly involve the lungs (pulmonary hemorrhage), gastrointestinal tract (melena, hematemesis), urogenital tract (hematuria), and skin (petechiae, ecchymosis, and bleeding from venipuncture sites). Pulmonary hemorrhage (with or without jaundice) is now recognized as a widespread public health problem, presenting with cough, chest pain, respiratory distress, and hemoptysis that may not be apparent until patients are intubated.

Jaundice occurs in 5–10% of all patients with leptospirosis; it can be profound and give an orange cast to the skin but usually is not associated with fulminant hepatic necrosis. Physical examination may reveal an enlarged and tender liver.

Acute kidney injury is common in severe disease, presenting after several days of illness, and can be either nonoliguric or oliguric. Typical electrolyte abnormalities include hypokalemia and hyponatremia. Loss of magnesium in the urine is uniquely associated with leptospiral nephropathy. Hypotension is associated with acute tubular necrosis, oliguria, or anuria, requiring fluid resuscitation and sometimes vasopressor therapy. Hemodialysis can be lifesaving, with renal function typically returning to normal in survivors.

In severe leptospirosis, an altered mental status may reflect leptospiral meningitis. The diagnosis of leptospirosis meningitis may be challenging since patients may be anicteric or lack other diagnostic hallmarks of severe leptospirosis. Without proper antibiotic treatment, a mortality rate of 13% has been reported; in contrast, among patients treated with antibiotics, the mortality rate is 2%. Neurologic sequelae are described until months after acute illness.

Other syndromes include (necrotizing) pancreatitis, cholecystitis, skeletal muscle involvement, and rhabdomyolysis with moderately elevated serum creatine kinase levels. Cardiac involvement is commonly reflected on the electrocardiogram as nonspecific ST- and T-wave changes. Repolarization abnormalities and arrhythmias are considered poor prognostic factors. Myocarditis has been described. Rare hematologic complications include hemolysis, thrombotic thrombocytopenic purpura, and hemolytic-uremic syndrome.

Long-term symptoms following severe leptospirosis include fatigue, myalgia, malaise, and headache and may persist for years. Autoimmune-associated uveitis, a potentially chronic condition, is a recognized sequela of leptospirosis.

DIAGNOSIS

The clinical diagnosis of leptospirosis should be based on an appropriate exposure history combined with any of the protean manifestations of the disease. Returning travelers from endemic areas usually have a history of recreational freshwater activities or other mucosal or percutaneous contact with contaminated surface waters or soil. For nontravelers, recreational or accidental water/soil contact and especially occupational hazards that involve direct or indirect animal contact should be explored (see "Epidemiology," above).

Although biochemical, hematologic, and urinalysis findings in acute leptospirosis are nonspecific, certain patterns may suggest the diagnosis. Laboratory results usually show signs of a bacterial infection, including leukocytosis with a left shift and elevated markers of inflammation (C-reactive protein level, procalcitonin, and erythrocyte sedimentation rate). Thrombocytopenia (platelet count $\leq 100 \times 10^9/L$) is common and is associated with bleeding and renal failure. In severe disease, signs of coagulation activation may be present, varying from borderline abnormalities to a serious derangement compatible with DIC as defined by international criteria. The kidneys are invariably involved in leptospirosis. The absence of renal involvement does not rule out leptospirosis. However, when the kidneys are involved, related findings range from urinary sediment changes (leukocytes, erythrocytes, and hyaline or granular casts) and mild proteinuria in mild disease to renal failure and azotemia in severe leptospirosis. Nonoliguric hypokalemic renal insufficiency (see "Clinical Manifestations," above) is characteristic of early leptospirosis. Serum bilirubin levels may be high, whereas rises in aminotransferase and alkaline phosphatase levels are usually moderate. Although clinical symptoms of pancreatitis are not a common finding, amylase levels are often elevated. When symptoms of meningitis develop, examination of the CSF shows pleocytosis that can range from a few cells to >1000 cells/μL, with a predominance of lymphocytes. Predominant polymorphonuclear pleocytosis has been reported. This phenomenon may be related to the timing of the lumbar puncture: polymorphonuclear cells are thought to be found in early disease and are later replaced by lymphocytes. Although protein and glucose levels in the CSF are usually normal, protein levels may be slightly elevated.

In severe leptospirosis, pulmonary radiographic abnormalities are more common than would be expected based on physical examination (Fig. 189-4). The most common radiographic finding is a patchy bilateral alveolar pattern that corresponds to scattered alveolar hemorrhage. These abnormalities predominantly affect the lower lobes. Other findings include pleura-based densities (representing areas of hemorrhage) and diffuse ground-glass attenuation typical of acute respiratory distress syndrome (ARDS).

A definitive diagnosis of leptospirosis is based on isolation of the organism from the patient, on a positive result in the PCR, or on seroconversion or a rise in antibody titer. In cases with strong clinical evidence of infection, a single antibody titer of 1:200–1:800 (depending on whether the case occurs in a low- or high-endemic area) in the microscopic agglutination test (MAT) is required. Preferably, a fourfold or greater rise in titer is detected between acute- and convalescent-phase serum specimens. Antibodies generally do not reach detectable levels until the second week of illness. The antibody response can be affected by early treatment with antibiotics.

The MAT, which uses a battery of live leptospiral strains, and the enzyme-linked immunosorbent assay (ELISA), which uses a broadly reacting antigen, are the standard serologic procedures. The MAT usually is available only in specialized laboratories and is used for determination of the antibody titer and for tentative identification of the involved leptospiral serogroup—and, when epidemiologic background information is available, the putative serovar. This point underscores the importance of testing antigens representative of the serovars prevalent in the particular geographic area. However, cross-reactions occur frequently, and thus definitive identification of the infecting serovar or serogroup is not possible without isolation of the causative organism. Because serologic testing lacks sensitivity in the early acute phase of the disease (up to day 5), it cannot be used as the basis for a timely decision about whether to start treatment.

In addition to the MAT and the ELISA, various rapid tests with diagnostic value have been developed, and some of these are commercially available. These rapid tests mainly apply lateral flow, (latex) agglutination, or ELISA methodology and are reasonably sensitive and specific, although results reported in the literature vary, probably as a consequence of differences in test interpretation, (re)exposure risks, serovar distribution, and the use of biased serum panels. These methods do not require culture or MAT facilities and are useful in settings that lack a strong medical infrastructure. PCR methodologies, notably real-time PCR, have become increasingly widely implemented. Compared with serology, PCR offers a great advantage: the capacity to confirm the diagnosis of leptospirosis with a high degree of accuracy during the first 5 days of illness.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of leptospirosis is broad, reflecting the diverse clinical presentations of the disease. Although leptospirosis transmission is more common in tropical and subtropical regions, the absence of a travel history does not exclude the diagnosis. When fever, headache, and myalgia predominate, influenza, SARS-CoV-2, and other common and less common (e.g., dengue and chikungunya) viral infections should be considered. Malaria, typhoid fever, ehrlichiosis, viral hepatitis, and acute HIV infection may mimic the early stages of leptospirosis and are important to recognize. Rickettsial diseases, dengue, and hantavirus infections (hemorrhagic fever with renal syndrome or hantavirus cardiopulmonary syndrome) share epidemiologic and clinical features with leptospirosis. Dual infections have been reported. In this light, it is advisable to conduct serologic testing for rickettsiae, dengue virus, and hantavirus when leptospirosis is suspected. When bleeding is detected, dengue hemorrhagic fever and other viral hemorrhagic fevers, including hantavirus infection, yellow fever, Rift Valley fever, filovirus infections, and Lassa fever, should be considered.

TREATMENT

Leptospirosis

Severe leptospirosis should be treated with IV penicillin **(Table 189-1)** as soon as the diagnosis is considered. *Leptospira* are highly susceptible to a broad range of antibiotics, including the β-lactam antibiotics, cephalosporins, aminoglycosides, and macrolides, but are not susceptible to vancomycin, rifampicin, metronidazole, and chloramphenicol. Early intervention may prevent the development

TABLE 189-1 Treatment and Chemoprophylaxis of Leptospirosis in Adults[a]

INDICATION	REGIMEN
Treatment	
Mild leptospirosis	Doxycycline[b] (100 mg PO bid) *or* Amoxicillin (500 mg PO tid) *or* Ampicillin (500 mg PO tid)
Moderate/severe leptospirosis	Penicillin (1.5 million units IV or IM q6h) *or* Ceftriaxone (2 g/d IV) *or* Cefotaxime (1 g IV q6h) *or* Doxycycline[b] (loading dose of 200 mg IV, then 100 mg IV q12h)
Chemoprophylaxis	
	Doxycycline[b] (200 mg PO once a week) *or* Azithromycin (250 mg PO once or twice a week)

[a]All regimens are given for 7 days. [b]Doxycycline should not be given to pregnant women or children. [c]The efficacy of doxycycline prophylaxis in endemic or epidemic settings remains unclear. Experiments in animal models and a cost-effectiveness model indicate that azithromycin has a number of characteristics that may make it efficacious in treatment and prophylaxis.

of major organ-system failure or lessen its severity. Although studies supporting antibiotic therapy have produced conflicting results, clinical trials are difficult to perform in settings patients frequently present for medical care in the later stages of disease. Antibiotics are less likely to benefit patients in whom organ damage has already occurred. Two open-label randomized studies comparing penicillin with parenteral cefotaxime, parenteral ceftriaxone, and doxycycline showed no significant differences among the antibiotics with regard to complications and mortality risk. Thus ceftriaxone, cefotaxime, or doxycycline is a satisfactory alternative to penicillin for the treatment of severe leptospirosis. Antimicrobial susceptibility testing is not routine practice in individual cases of leptospirosis; to date, however, antibiotic resistance has not been reported in isolates from patients or the environment.

In mild cases, oral treatment with doxycycline, azithromycin, ampicillin, or amoxicillin is recommended. In regions where rickettsial diseases are coendemic, doxycycline or azithromycin is the drug of choice. In rare instances, a Jarisch-Herxheimer reaction develops within hours after the initiation of antimicrobial therapy.

Aggressive supportive care for leptospirosis is essential and can be life-saving. Patients with nonoliguric renal dysfunction require aggressive fluid and electrolyte resuscitation to prevent dehydration and precipitation of oliguric renal failure. Peritoneal dialysis or hemodialysis should be provided to patients with oliguric renal failure. Rapid initiation of hemodialysis has been shown to reduce mortality risk and typically is necessary only for short periods. Patients with pulmonary hemorrhage may have reduced pulmonary compliance (as seen in ARDS) and may benefit from mechanical ventilation with low tidal volumes to avoid high ventilation pressures. Evidence is contradictory for the use of glucocorticoids and desmopressin as adjunct therapy for pulmonary involvement associated with severe leptospirosis.

PROGNOSIS

Most patients with leptospirosis recover. However, post-leptospirosis symptoms, mainly of a depression-like nature, may occur and persist for years after the acute disease. Mortality rates are highest among patients who are elderly and those who have severe disease (pulmonary hemorrhage, Weil's syndrome). Leptospirosis during pregnancy is associated with high fetal mortality rates. Long-term follow-up of patients with renal failure and hepatic dysfunction has documented good recovery of both renal and hepatic function.

PREVENTION

Individuals who may be exposed to *Leptospira* through their occupations or their involvement in recreational freshwater activities should be informed about the risks. Measures for controlling leptospirosis include avoidance of exposure to urine and tissues from infected animals through proper eyewear, footwear, and other protective equipment. Targeted rodent control strategies have a potential benefit.

Vaccines for agricultural and companion animals are generally available, and their use should be encouraged. The veterinary vaccine used in each area should contain the serovars known to be present in that area in order to prevent vaccine serovar mismatch. Unfortunately, some vaccinated animals potentially still excrete leptospires in their urine. Commercial vaccines for human leptospirosis are available in Japan, China, Cuba, and France. These vaccines, made with bacterins—inactivated *Leptospira*—provide short-term, serovar-specific immunity but cause strong adverse side effects; both usability and availability remain problematic. One of the largest-scale trials of vaccine efficacy for leptospirosis in humans has been reported from Cuba. However, no conclusions can be drawn about efficacy and adverse reactions because of insufficient study design and outcome details. The efficacy of chemoprophylaxis with doxycycline (200 mg once a week) or azithromycin (in pregnant women and children) is being disputed, but focused pre- and postexposure administration is indicated in instances of well-defined short-term exposure (Table 189-1).

ACKNOWLEDGMENT

The authors gratefully acknowledge Dr. Marga G. A. Goris for her substantial contributions to this chapter in previous editions.

FURTHER READING

Adler A: *Leptospira and Leptospirosis*. Berlin Heidelberg, Springer-Verlag, 2015.

Azevedo IR et al: Human leptospirosis: In search for a better vaccine. Scand J Immunol 98:1, 2023.

de Vries SG et al: Leptospirosis among returned travelers: A GeoSentinel Site Survey and Multicenter Analysis–1997–2016. Am J Trop Med Hyg 99:127, 2018.

Haake DA, Levett PN: Leptospirosis in humans. Curr Top Microbiol Immunol 387:65, 2015.

Levett PN: Leptospirosis. Clin Microbiol Rev 14:296, 2001.

Petakh P et al: Current treatment options for leptospirosis: A mini-review. Front Microbiol 15:1403765, 2024.

Sykes JE et al: A global one health perspective on leptospirosis in humans and animals. J Am Vet Med Assoc 260:1589,2022.

van Samkar A et al: Suspected leptospiral meningitis in adults: Report of four cases and review of the literature. Neth J Med 73:464, 2015.

Vincent AT et al: Revisiting the taxonomy and evolution of pathogenicity of the genus *Leptospira* through the prism of genomics. PLoS Negl Trop Dis 13:e0007270, 2019.

190

Relapsing Fever

Alan G. Barbour

Relapsing fever is caused by infection with any of several species of *Borrelia* spirochetes. It occurs in three different clinical and epidemiologic forms: louse-borne relapsing fever (LBRF), soft tick relapsing fever (STRF), and hard tick relapsing fever (HTRF). Physicians in ancient Greece distinguished LBRF from other febrile disorders by its characteristic clinical presentation: two or more fever episodes separated by varying periods of well-being. In the nineteenth century, LBRF was one of the first diseases to be associated with a specific microbe by virtue of its characteristic laboratory finding: the presence of large numbers of spirochetes of the genus *Borrelia* in the blood.

The host responds with systemic inflammation that results in an illness ranging from a flulike syndrome to sepsis. Other manifestations are the consequences of central nervous system (CNS) involvement and disordered hemostasis. Antigenic variation of the spirochetes' surface proteins accounts for the infection's relapsing course. Acquired immunity follows the serial development of antibodies to each of the several variants appearing during an infection. Treatment with antibiotics results in rapid cure but at the risk of a moderate to severe Jarisch-Herxheimer reaction.

LBRF caused large epidemics well into the twentieth century and currently occurs in northeastern Africa and among migrants from that area. At present, however, most cases of relapsing fever are tick-borne in origin, with transmission from either soft-bodied ticks or hard-bodied ticks. Sporadic cases and small outbreaks of STRF are focally distributed on most continents, with Africa and Central Asia most affected. In North America, the majority of reports of relapsing fever have been from the western United States and Canada and northern Mexico.

Two other members of the genus, *Borrelia miyamotoi* and *B. lonestari*, are causes of HTRF, an acute febrile illness with nonspecific constitutional symptoms and occasionally meningoencephalitis.

ETIOLOGIC AGENT

Coiled, thin microscopic filaments that swim in one direction and then coil up before heading in another were first observed in the blood of patients with relapsing fever in the 1880s. These microbes were categorized as spirochetes and assigned to the genus *Borrelia*. The breakthrough cultivation medium was rich in ingredients, ranging from simple (e.g., *N*-acetylglucosamine) to more complex (e.g., serum). The limited biosynthetic capacity of *Borrelia* cells is accounted for by a genome content one-quarter that of *Escherichia coli*.

Like other spirochetes, the helix-shaped *Borrelia* cells have two membranes, the outer of which is more loosely secured than in other double-membrane bacteria, such as *E. coli*. As a consequence, fixed organisms with damaged membranes can assume a variety of morphologies in smears and histologic preparations. The flagella of spirochetes run between the two membranes and are not on the cell surface. Although technically gram-negative, the 10- to 20-μm-long *Borrelia* cells, with a diameter of 0.2–0.3 μm, are too narrow to be seen by microscopy of Gram-stained slides.

EPIDEMIOLOGY

Borrelia recurrentis, unlike the other relapsing fever *Borrelia* species, is acquired from an insect, the body louse (*Pediculus humanus corporis*), with humans serving as the sole reservoir in its life cycle **(Table 190-1)**. Acquisition occurs not from the bite itself but from either rubbing the crushed insect's hemolymph into the bite site or auto-inoculation into the conjunctivae or a wound. Although LBRF transmission is currently limited to Ethiopia, Eritrea, and Somalia, the disease has had a global distribution in the past, and that potential remains. Outbreaks of LBRF, often in association with typhus, can occur under circumstances of famine, refugee migration, war, homelessness, and incarceration. LBRF can occur in camps of migrants at a distance from their home countries.

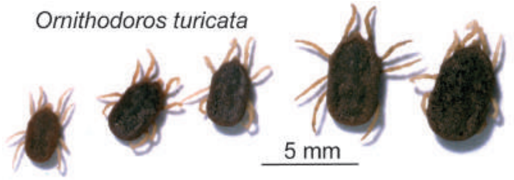

FIGURE 190-1 ***Ornithodoros turicata* soft ticks** of different ages.

The several species of *Borrelia* that cause STRF have geographic distributions that correspond with those of their vectors: soft ticks of the genus *Ornithodoros* **(Fig. 190-1)**. STRF is found on most continents but is absent in arctic environments. For most species, the reservoirs of infection are small to medium-sized mammals, usually rodents, but also pigs and other domestic animals living around human habitats. However, one species, *Borrelia duttonii* in sub-Saharan Africa, is largely maintained by tick transmission between human hosts. In North America, STRF occurs as single cases or small case clusters through transient exposure of persons to tick-infested buildings or caves where mammals have nests or sleep. The two main *Borrelia* species involved in North America are *Borrelia hermsii* and *Borrelia nietonii* in the mountainous west and *Borrelia turicatae* in arid southwestern and south-central regions. The soft tick vectors typically feed for no more than 30 min, usually while the victim is sleeping, and then leave undetected. Transovarial transmission from one generation of ticks to the next means that infection risk may persist in a dwelling long after incriminated mammalian reservoirs have been removed.

Borrelia miyamotoi is transmitted to humans from other mammals by different species of *Ixodes* hard ticks (e.g., *I. scapularis* in the eastern United States and *I. ricinus* in Europe) that also transmit Lyme disease, babesiosis, anaplasmosis, and a viral encephalitis. *B. miyamotoi* is acquired through outdoor activities and through contact with ticks in forested and shrubby areas during recreation, work, or activities around the home, similarly to Lyme disease **(Chap. 191)**. Among residents of most areas where *B. miyamotoi* and *Borreliella* (also called *Borrelia*) *burgdorferi* coexist, the prevalence of antibodies to the former is about one-fourth of that to the latter. In contrast to *B. burgdorferi*, *B. miyamotoi* is transmitted to the host soon after the tick begins to feed, and it may be acquired from tick larvae as well as nymphs and adults. A less common cause of HTRF is *B. lonestari*, which is transmitted by *Amblyomma americanum* ticks of the southern and eastern United States.

TABLE 190-1 Relapsing Fever *Borrelia* Species, by RF Type, Geographic Region, and Vector

SPECIES	RELAPSING FEVER TYPE	REGION(S)	ARTHROPOD VECTOR(S)
B. crocidurae	Soft tick RF (STRF)	West Africa	*Ornithodoros sonrai*
B. duttonii	STRF	East Africa, southern and central Africa	*O. moubata*
B. hermsii	STRF	Western North America	*O. hermsi*
B. hispanica	STRF	North Africa, southern Europe	*O. erraticus*
B. kalaharica	STRF	West Africa, southern Africa	*O. savignyi*
B. lonestari	Hard tick RF (HTRF)	Southern and eastern United States	*Amblyomma americanum*
B. mazzotti	STRF	Mexico, Central America	*O. talaje*
B. miyamotoi	HTRF	North America, Asia, Europe	*Ixodes pacificus, I. persulcatus, I. ricinus, I. scapularis*
B. nietonii	STRF	Western North America	*O. hermsii*
B. persica	STRF	Central Asia, Middle East	*O. tholozani*
B. puertoricensis	STRF	Central America	*O. puertoricensis*
B. recurrentis	Louse-borne RF	Africa, global[a]	*Pediculus humanus corporis* (human body louse)
B. turicatae	STRF	Southwestern United States, northern Mexico	*O. turicata*
B. venezuelensis	STRF	Central America, South America	*O. rudis*

[a] Although transmission is currently limited to Ethiopia and adjacent countries, *B. recurrentis* infection has had a global distribution in the past, and that potential remains.

PATHOGENESIS AND IMMUNITY

STRF and HTRF spirochetes enter the body in the tick's saliva with the onset of feeding. From an inoculum of a few cells, STRF spirochetes proliferate in the blood, doubling every 6 h to numbers of 10^6–10^7/mL or more. HTRF spirochetes grow more slowly in a mammalian host and attain lower peak concentrations in the blood. *Borrelia* species are extracellular pathogens; their presence inside cells connotes dead bacteria after phagocytosis. Binding of the spirochetes to erythrocytes leads to aggregation of red blood cells, their sequestration in the spleen and liver, and hepatosplenomegaly and anemia. A bleeding disorder is probably the consequence of thrombocytopenia, impaired hepatic production of clotting factors, and/or blockage of small vessels by aggregates of spirochetes, erythrocytes, and platelets. Some species (e.g., *B. turicatae*) are neurotropic and enter the brain, where they are comparatively sheltered from host immunity. Relapsing fever spirochetes can cross the maternal-fetal barrier and cause placental damage and inflammation, leading to intrauterine growth retardation and congenital infection.

Although *Borrelia* species do not have potent exotoxins or a lipopolysaccharide endotoxin, they have abundant lipoproteins that activate Toll-like receptors on host cells, which leads to a proinflammatory process similar to that in endotoxemia, with elevations of tumor necrosis factor α, interleukin 6, and interleukin 8 concentrations.

IgM antibodies specific for the serotype-defining surface lipoprotein appear after a few days of infection and soon reach a concentration that causes lysis of bacteria in the blood through either direct bactericidal action or opsonization. The release of lipoproteins and other bacterial products from dying bacteria provokes a "crisis," during which there can be an increase in temperature, hypotension, and other signs of shock. A similar phenomenon occurring in some patients soon after the initiation of antibiotic treatment is characterized by an abrupt worsening of the patient's condition, which is called a Jarisch-Herxheimer reaction (JHR).

CLINICAL MANIFESTATIONS

STRF and LBRF present with the sudden onset of fever. Febrile periods are punctuated by intervening afebrile periods of a few days; this pattern occurs at least twice in STRF. The patient's temperature is ≥39°C and may be as high as 43°C. The first fever episode often ends in a crisis lasting ~15–30 min and consisting of rigors, a further elevation in temperature, and increases in pulse and blood pressure. The crisis phase is followed by profuse diaphoresis, falling temperature, and hypotension, which usually persist for several hours. In LBRF, the first episode of fever is unremitting for 3–6 days; it is usually followed by a single milder episode. In STRF, multiple febrile periods last 1–3 days each. In both forms, the interval between fevers ranges from 4 to 14 days, sometimes with symptoms of malaise and fatigue.

The symptoms that accompany the fevers are usually nonspecific. Headache, neck stiffness, arthralgia, myalgia, and vomiting may accompany the first and subsequent febrile episodes. An enlarging spleen and liver cause abdominal pain. A nonproductive cough is common during LBRF and—in combination with fever and myalgias—may suggest influenza. Acute respiratory distress syndrome may occur during LBRF or STRF.

On physical examination, the patient with LBRF or STRF may be delirious or apathetic. There may be body lice in the patient's clothes or signs of insect bites. In regions with *B. miyamotoi* or *B. lonestari* infection, a hard tick may be embedded in the skin. Jaundice, epistaxis, and subconjunctival hemorrhages are common during LBRF but not in STRF or HTRF. Splenomegaly or spleen tenderness is common in LBRF and STRF.

Localizing neurologic findings are more common in STRF than in LBRF or HTRF. In North America, *B. turicatae* infection has neurologic manifestations, including aseptic meningitis and cranial neuritis, more often than *B. hermsii* or *B. nietonii* infection. Unilateral or bilateral Bell's palsy is the most common form of cranial neuritis and typically presents in the second or third febrile episode, not the first. Visual impairment from unilateral or bilateral iridocyclitis or panophthalmitis may be permanent. In LBRF, neurologic manifestations such as altered mental state are thought to be secondary to systemic inflammation or small hemorrhages in the brain rather than to direct invasion of the nervous system.

Myocarditis in STRF or LBRF is evidenced by gallops on cardiac auscultation, a prolonged QT_c interval, and cardiomegaly and pulmonary edema on chest radiography.

General laboratory studies in all forms of relapsing fever are not specific. Mild to moderate normocytic anemia is common, but frank hemolysis and hemoglobinuria do not develop. Leukocyte counts range from slightly elevated to leukopenic. Mild to moderate thrombocytopenia is common; platelet counts may fall below 50,000/μL in some cases. C-reactive protein and procalcitonin levels are elevated. Laboratory evidence of hepatitis can be found, with elevated serum concentrations of aminotransferases; the prothrombin and partial thromboplastin times may be moderately prolonged.

Analysis of the cerebrospinal fluid (CSF) is indicated in all cases of suspected relapsing fever with signs of meningitis or meningoencephalitis. The presence of mononuclear pleocytosis and mildly to moderately elevated protein levels justifies intravenous antibiotic therapy in relapsing fever.

The manifestations and course of *B. miyamotoi* disease are not as distinctive as those of relapsing fever. The most common presentation is acute onset of fever with chills, headache, and myalgias starting 1–2 weeks after a tick bite. Patients have been hospitalized with a presumptive diagnosis of undifferentiated sepsis. There may be a second fever episode in untreated patients. Meningoencephalitis or meningitis was documented in adults with deficiencies in humoral immunity, including the effects of anti–B cell antibodies such as rituximab. If the patient has coexisting early Lyme disease, there may be erythema migrans, the localized skin rash.

DIAGNOSIS

STRF or LBRF should be considered in a patient with the characteristic fever pattern and a history of recent exposure—i.e., within 1–2 weeks before illness onset—to body lice or soft ticks in geographic areas with documented current or past transmission. Because of the longevity of the ticks and the transovarial transmission of the pathogen in the ticks, a case of relapsing fever may be diagnosed many years after the last case reported in that locale. The lice may be on the clothes of a migrant, refugee, or unhoused person. While the epidemiologic risks for *B. miyamotoi* HTRF are similar to those for Lyme disease, prompt removal of an embedded tick upon discovery may not reduce the risk of infection of this pathogen.

The bedrock for laboratory diagnosis of LBRF and STRF remains direct detection of the spirochetes by microscopy of the blood. Manual differential counts of white blood cells by Wright, Giemsa, or Giemsa-Wright stain usually reveal spirochetes in thin blood smears if their concentration is ≥10^5/mL and several oil-immersion fields are examined **(Fig. 190-2)**. The peak density of *B. miyamotoi* or *B. lonestari* in the blood may not be high enough for use of a blood smear alone for diagnosis. For LBRF and STRF, the preferred time to obtain a blood specimen is at or just before fever's peak. Lower concentrations of spirochetes may be revealed by a thick blood smear that is treated with 0.5% acetic acid before staining. An alternative is a wet mount of citrated blood mixed with saline and examined by phase-contrast or dark-field microscopy for motile spirochetes.

Polymerase chain reaction (PCR) and similar nucleic acid amplification test (NAAT) procedures are increasingly used for examination of blood or CSF in cases of suspected relapsing fever. Overall, NAAT is more sensitive than thin blood smear, particularly for samples obtained between febrile episodes. PCR is the preferred procedure for direct detection of *B. miyamotoi* or *B. lonestari* in blood or CSF. Culture of blood or CSF in Barbour-Stoenner-Kelly broth medium or equivalent is an option for isolation of *Borrelia* species. However, visible growth takes several days, and few laboratories offer this service.

Options for serologic confirmation of infection are limited, and results may be misleading. Whole cell–based assays, such as enzyme-linked immunosorbent assay (ELISA) and the C6 peptide ELISA for Lyme disease, may be positive in relapsing fever through antigenic cross-reactivities among these spirochetes. A commercially available assay based on GlpQ—a protein antigen of STRF, HTRF, and LBRF *Borrelia* species but not of any Lyme disease species—has better

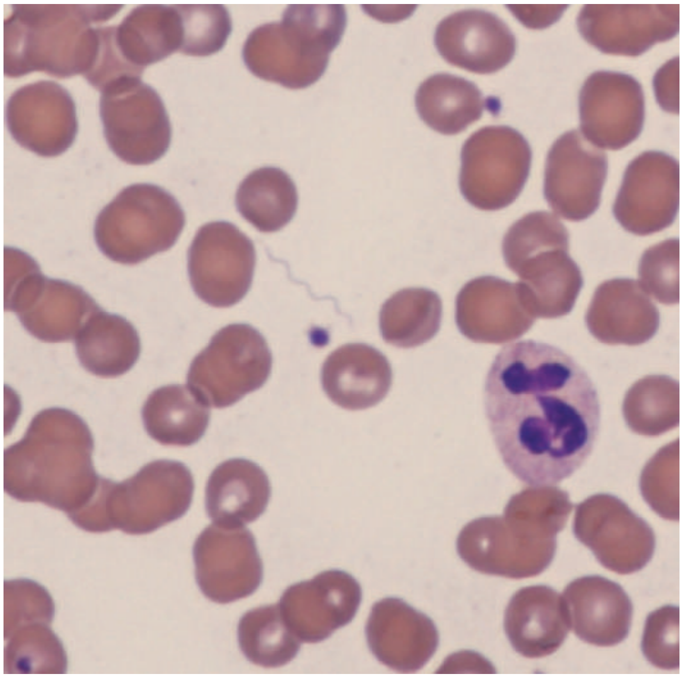

FIGURE 190-2 Photomicrograph of tick-borne relapsing fever spirochete (*Borrelia turicatae*) in a Giemsa-Wright–stained thin blood smear. Included in the figure are a polymorphonuclear leukocyte and two platelets.

specificity but commonly is negative at a time when a blood smear or PCR assay would be positive. The results of a GlpQ-based assay or an indirect immunofluorescence assay of whole cells cannot be used to differentiate between different *Borrelia* species as to etiology. A positive IgG immunoassay of a single specimen may be the consequence of a past infection and not the present illness.

DIFFERENTIAL DIAGNOSIS

Depending on the patient's history of residential, occupational, travel, and recreational exposures, the differential diagnosis of STRF includes one or more of the following infections that feature either periodicity in the fever pattern or an extended single febrile period with nonspecific constitutional symptoms: Rocky Mountain spotted fever and other rickettsioses, ehrlichiosis, anaplasmosis, tick-borne viral infection, and babesiosis in North America, Europe, Russia, and northeastern Asia. STRF with Bell's palsy may be misdiagnosed as Lyme disease in areas endemic for both diseases. Elsewhere in the Americas and Asia and in most of Africa, malaria, typhoid fever, typhus and other rickettsioses, dengue, and leptospirosis also may be considered. Other agents transmitted by the body louse are *Rickettsia prowazekii*, the cause of typhus, and *Bartonella quintana*, the cause of trench fever. There may be co-infections of malaria, typhus, or typhoid with STRF or LBRF. *B. miyamotoi* infection may coexist with Lyme disease.

TREATMENT

Relapsing Fever

Penicillin and tetracyclines have been the antibiotics of choice for LBRF and STRF for several decades. Erythromycin has been a long-standing alternative choice. There is no evidence of acquired resistance to these antibiotics. *Borrelia* species are also susceptible to second- and third-generation cephalosporins. These spirochetes are relatively resistant to rifampin, sulfonamides, and aminoglycosides. Spirochetes are no longer detectable in the blood within a few hours after the first dose of an effective antibiotic.

Under conditions of limited resources or in the midst of an epidemic, a single dose of antibiotic usually suffices for successful treatment of LBRF **(Fig. 190-3)**. For adults, a single dose of oral doxycycline (200 mg), oral tetracycline (500 mg), or intramuscular ceftriaxone 250 mg for adults and 125 mg for children is effective. The corresponding doses for children are oral tetracycline at 12.5 mg/kg, oral doxycycline at 5 mg/kg, and intramuscular penicillin G procaine at 200,000–400,000 units. When an adult patient is stuporous or nauseated, the intravenous dose of tetracycline is 250–500 mg. Tetracyclines are contraindicated in pregnant and nursing women; for individuals in these groups who are allergic to penicillin, oral erythromycin (500 mg for adults and 12.5 mg/kg for children) is an alternative. While there is little reported experience with other macrolides, such as azithromycin, these are likely to be as effective as erythromycin.

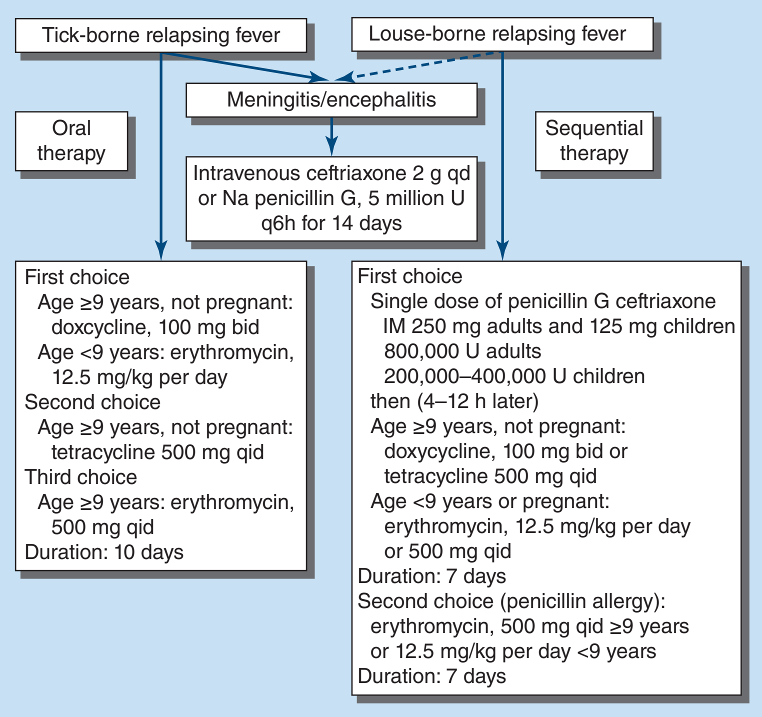
Tick-borne relapsing fever
Louse-borne relapsing fever
Meningitis/encephalitis
Oral therapy
Sequential therapy
Intravenous ceftriaxone 2 g qd or Na penicillin G, 5 million U q6h for 14 days
First choice
Age ≥9 years, not pregnant: doxcycline, 100 mg bid
Age <9 years: erythromycin, 12.5 mg/kg per day
Second choice
Age ≥9 years, not pregnant: tetracycline 500 mg qid
Third choice
Age ≥9 years: erythromycin, 500 mg qid
Duration: 10 days
First choice
Single dose of penicillin G ceftriaxone
IM 250 mg adults and 125 mg children
800,000 U adults
200,000–400,000 U children
then (4–12 h later)
Age ≥9 years, not pregnant:
doxycycline, 100 mg bid or
tetracycline 500 mg qid
Age <9 years or pregnant:
erythromycin, 12.5 mg/kg per day
or 500 mg qid
Duration: 7 days
Second choice (penicillin allergy):
erythromycin, 500 mg qid ≥9 years
or 12.5 mg/kg per day <9 years
Duration: 7 days

FIGURE 190-3 Algorithm for treatment of relapsing fever. If it is not known whether the patient has tick-borne or louse-borne relapsing fever, the patient should be treated for the tick-borne form. The *dashed line* indicates that central nervous system invasion in louse-borne relapsing fever is uncommon.

There are shortcomings of single-dose therapies for LBRF. With penicillin alone, recurrence may occur in up to 20% of patients, and the frequency of JHR was higher after tetracycline than penicillin. For treatment of LBRF in adults in Ethiopia, a regimen that reduces rates of both recurrence and JHR was a single dose of 400,000 units of intramuscular penicillin G procaine followed several hours later or the next day by doxycycline (100 mg orally twice daily) or tetracycline (500 mg or 12.5 mg/kg orally every 6 h) for 7 days.

The accumulated anecdotal reports on STRF therapy indicate a recurrence rate of ≥20% after single-dose treatment, plausibly due to the propensity of some tick-borne species to invade the CNS. Accordingly, multiple antibiotic doses are recommended. The preferred treatment for adults is a 10-day course of doxycycline (100 mg twice daily) or tetracycline (500 mg or 12.5 mg/kg orally every 6 h). When tetracyclines are contraindicated, the alternatives are oral penicillin V potassium (500 mg or 12.5 mg/kg every 6–8 h) or erythromycin (500 mg or 12.5 mg/kg orally every 6 h) for 10 days. If a β-lactam antibiotic is given and CNS involvement is confirmed or suspected, it is preferably administered intravenously rather than orally. For adults, the regimen is penicillin G (5 million units IV every 6 h) or ceftriaxone (2 g IV daily) for 10–14 days.

The JHR during treatment of LBRF or STRF can be severe and may end in death if precautions are not in place for close monitoring for at least 24 h and with provision of parenteral cardiovascular and volume support as needed. Apprehension, rigors, fever, and hypotension occur within 1–3 h of initiation of antibiotic treatment and may be accompanied by a further decrease in the platelet count. The incidence of the JHR is 20–60% in LBRF after the first antibiotic dose. JHR may also be encountered when a patient with unsuspected relapsing fever is treated with other types of antibiotics, such ciprofloxacin, that have suboptimal effects.

Experience with the treatment of *B. miyamotoi* or *B. lonestari* HTRF is limited, but these organisms likely have the same antibiotic susceptibilities as other *Borrelia* species. Therapy for *B. miyamotoi* disease follows the guidelines for Lyme disease. This would include parenteral therapy for CNS involvement. In absence of contraindications, doxycycline (100 mg twice daily) is the preferred choice for uncomplicated *B. miyamotoi* infection because of the antibiotic's efficacy for anaplasmosis and Lyme disease. If JHR occurs, it is generally milder than is observed in relapsing fever.

PROGNOSIS

The mortality rates for untreated LBRF and STRF are in the ranges of 10–70% and 4–10%, respectively, and are largely determined by coexisting conditions, such as malnutrition or another infection, and by the availability of medical support. With prompt antibiotic treatment, the mortality rate is 2–5% for LBRF and <2% for STRF. There are no reported deaths from HTRF. Features associated with a poor prognosis of LBRF or STRF include concurrence with malaria, typhus, or typhoid; pregnancy; stupor or coma on admission; diffuse bleeding; poor liver function; myocarditis; and bronchopneumonia. The mortality rate from the JHR in LBRF, in the absence of adequate monitoring and resuscitation measures, is ~5%. LBRF or STRF during pregnancy frequently leads to abortion, stillbirth, or perinatal death, but congenital malformations have not been reported. Although spirochetes or their remnants may persist in the CNS or other sequestered sites after bacteremia has resolved, posttreatment sequelae and prolonged disability have not been documented for any form of relapsing fever. Partial immunity against reinfection seems to develop in residents of areas with perennial elevated risk.

PREVENTION

There is no vaccine for LBRF, STRF, or HTRF. Reduction of exposure to lice and ticks is the key strategy for prevention. LBRF can be prevented through improved personal hygiene, reduction of crowding, better access to hot water (≥60° C) for clothes washing, and selected use of pesticides. Clothing is an important factor in maintaining the human body louse. The risk of STRF can be reduced by construction of houses with concrete or sealed plank floors and without thatched roofs or mud walls. Dwellings in forested areas pose a risk in western North America when rodents nest in the roof, attic, or wall spaces or under the structure. Buildings infested with *Ornithodoros* ticks can be treated with pesticides and then rodent-proofed. If residing in a high-risk environment, individuals should not sleep on the floor, and beds should be moved away from the wall. Individuals with recreational or occupational exposure to caves, where mammals may reside, merit advice about the risk of STRF. Following exposure at a site of STRF risk, treatment with doxycycline (a single dose of 100 mg or 200 mg on day 1 followed by 100 mg/d for 4 days) was efficacious in preventing infection in a placebo-controlled trial. Recommendations for preventing *B. miyamotoi* infection follow those for reducing risk of Lyme disease from exposure to the vector, hard ticks **(Chap. 191)**.

FURTHER READING

Barbour AG, Schwan TG: *Borrelia*, in *Bergey's Manual of Systematics of Archaea and Bacteria*, WB Whitman et al (eds). Hoboken, Wiley, 2015.

Beeson AM et al: Soft tick relapsing fever — United States, 2012–2021. MMWR Morb Mortal Wkly Rep 72:777, 2023.

Butler T: The Jarisch-Herxheimer reaction after antibiotic treatment of spirochetal infections: A review of recent cases and our understanding of pathogenesis. Am J Trop Med Hyg 96:46, 2017.

Isenring E et al: Infectious disease profiles of Syrian and Eritrean migrants presenting in Europe: A systematic review. Travel Med Infect Dis 25:65, 2018.

Kahlig E et al: Louse-borne relapsing fever-A systematic review and analysis of the literature: Part 1-Epidemiology and diagnostic aspects. PLoS Negl Trop Dis 15:e0008564, 2021.

Kahlig E et al: Louse-borne relapsing fever-A systematic review and analysis of the literature: Part 2-Mortality, Jarisch-Herxheimer reaction, impact on pregnancy. PLoS Negl Trop Dis 15:e0008656, 2021.

McCormick DW et al: Characteristics of hard tick relapsing fever caused by *Borrelia miyamotoi*, United States, 2013–2019. Emerg Infect Dis 29:1719, 2023.

Vazquez LJ et al: Relapsing fever caused by *Borrelia lonestari* after tick bite in Alabama. Emerg Infect Dis 29:441, 2023.

Warrell DA: Louse-borne relapsing fever (*Borrelia recurrentis* infection). Epidemiol Infect 147:e106, 2019.

Wormser GP et al: Aggregation of data from 4 clinical studies demonstrating efficacy of single-dose doxycycline postexposure for prevention of the spirochetal infections: Lyme disease, syphilis, and tick-borne relapsing fever. Diagn Microbiol Infect Dis 99:115293 2021.

191

Lyme Borreliosis

Allen C. Steere, Jacob E. Lemieux

DEFINITION

Lyme borreliosis is caused by a closely related group of spirochetes, *Borrelia burgdorferi sensu lato* (also called *Borreliella* spp.), transmitted by ticks of the *Ixodes ricinus* complex. The infection usually begins with a characteristic expanding skin lesion, erythema migrans (EM; stage 1, localized infection). After several days or weeks, the spirochete may spread to many different sites (stage 2, disseminated infection). Possible manifestations of disseminated infection include additional annular skin lesions, meningitis, cranial neuritis, radiculoneuritis, peripheral neuritis, carditis, atrioventricular nodal block, or migratory musculoskeletal pain. Months or years later (usually after periods of latent infection), intermittent or persistent arthritis, chronic encephalopathy or polyneuropathy, or acrodermatitis may develop (stage 3, persistent infection). Most patients experience early symptoms of the illness during the summer, but the infection may not become symptomatic until it progresses to stage 2 or 3.

Lyme disease was recognized as a separate entity in 1976 because of a geographic cluster of children in Lyme, Connecticut, who were thought to have juvenile rheumatoid arthritis. It became apparent that Lyme disease was a multisystem illness that affected primarily the skin, nervous system, heart, and joints. Epidemiologic studies of patients with EM implicated certain *Ixodes* ticks as vectors of the disease. Early in the twentieth century, EM had been described in Europe and attributed to *I. ricinus* tick bites. In 1982, a previously unrecognized spirochete, now called *Borrelia burgdorferi*, was recovered from *Ixodes scapularis* ticks and then from patients with Lyme disease. The entity is now called Lyme disease or Lyme borreliosis.

ETIOLOGIC AGENT

B. burgdorferi, the causative agent of Lyme disease, is a fastidious microaerophilic bacterium. The spirochete's genome is quite small (~1.5 Mb) and consists of a highly unusual genome organization with a linear chromosome of 950 kb and 17–21 linear and circular plasmids. The most remarkable aspect of the *B. burgdorferi* genome is that there are sequences for more than 100 known or predicted lipoproteins—a larger number than in almost any other organism. Most of these lipoproteins are encoded on plasmids and exported to the outer leaflet of the outer membrane, where they interact with the infected host or tick vector. The spirochete has few proteins with biosynthetic activity and depends on its host for most of its nutritional requirements. It has no sequences for recognizable toxins.

Currently, 20 closely related borrelial species are collectively referred to as *B. burgdorferi sensu lato* (i.e., "*B. burgdorferi* in the general sense") or *Borreliella* spp. The human infection Lyme borreliosis is caused primarily by four pathogenic genospecies: *B. burgdorferi sensu stricto* ("*B. burgdorferi* in the strict sense," hereafter referred to simply as *B. burgdorferi*), *Borrelia garinii*, *Borrelia bavariensis,* and *Borrelia afzelii*. *B. burgdorferi* is the major cause of the infection in the United States; all four genospecies are found in Europe, and *B. garinii, B. Afzelii,* and *B. bavariencis* are the major causes in Asia.

Strains of *B. burgdorferi* have been subdivided according to several typing schemes: one based on sequence variation of outer-surface protein C (OspC), a second based on differences in the 16S–23S rRNA intergenic spacer region (RST or IGS), a third called *multilocus sequence typing,* and a fourth based on whole genome sequencing (WGS). From these typing systems, it is apparent that strains of *B. burgdorferi* differ in pathogenicity. WGS type A, which includes OspC type A (RST1) strains, has the largest pangenome and the largest number of plasmid-encoded lipoproteins, which are often immunogenic, and is especially likely to disseminate. Thus, this strain is particularly virulent and may have played a role in the emergence of Lyme disease in epidemic form in the northeastern United States in the late twentieth century.

EPIDEMIOLOGY

The >20 known genospecies of *B. burgdorferi sensu lato* live in nature in enzootic cycles involving 14 species of ticks that are part of the *I. ricinus* complex. *I. scapularis* **(Fig. 472-1)** is the principal vector in the eastern United States from Maine to Georgia and in the midwestern states of Wisconsin, Minnesota, Indiana, and Michigan. *I. pacificus* is the vector in the western states of California and Oregon. The disease is acquired throughout Eurasia, from Ireland and Great Britain to Scandinavia to western Russia, where *I. ricinus* is the vector, and in eastern Russia, China, and Japan, where *I. persulcatus* is the vector. These ticks may transmit other agents as well. In the United States, *I. scapularis* also transmits *Babesia microti*, *Anaplasma phagocytophilum*, *Ehrlichia muris*–like agent, *Borrelia miyamotoi*, *Borrelia mayonii*, and Powassan virus (the deer tick virus) (see "Differential Diagnosis," below). In Europe and Asia, *I. ricinus* and *I. persulcatus* also transmit tick-borne encephalitis virus.

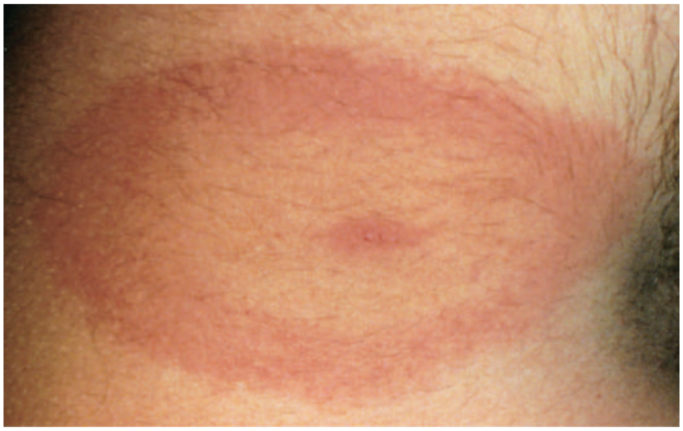

FIGURE 191-1 A classic erythema migrans lesion (9 cm in diameter) is shown near the right axilla. The lesion has partial central clearing, a bright red outer border, and a target center. *(Courtesy of Vijay K. Sikand, MD; with permission.)*

Ticks of the *I. ricinus* complex have larval, nymphal, and adult stages. They require a blood meal at each stage. The risk of infection in a given area depends largely on the density of these ticks as well as their feeding habits and animal hosts, which have evolved differently in different locations. For *I. scapularis* in the northeastern United States, the white-footed mouse and certain other rodents are the preferred hosts of the immature larvae and nymphs. It is critical that both of the tick's immature stages feed on the same host because the life cycle of the spirochete depends on horizontal transmission: in early summer from infected nymphs to mice and in late summer from infected mice to larvae, which then molt to become the infected nymphs that will begin the cycle again the following year. It is the tiny nymphal tick that is primarily responsible for transmission of the disease to humans, which peaks during the early summer months. White-tailed deer, which are not involved in the life cycle of the spirochete, are the preferred host for the adult stage of *I. scapularis* and seem to be critical to the tick's survival.

Lyme disease is now the most common vector-borne infection in the United States and Europe. Since surveillance was begun by the Centers for Disease Control and Prevention (CDC) in 1982, the number of cases in the United States has increased dramatically. More than 30,000 new cases are now reported each summer, but the number of true cases is estimated at 476,000 annually. In Europe, reported frequencies of the disease are highest in the middle of the continent and in Scandinavia.

PATHOGENESIS AND IMMUNITY

To maintain its complex enzootic cycle, *B. burgdorferi* must adapt to two markedly different environments: the tick and the mammalian host. The spirochete expresses outer-surface protein A (OspA) in the midgut of the tick, whereas OspC is upregulated as the organism travels to the tick's salivary gland. There, OspC binds a tick salivary-gland protein (Salp15), which is required for infection of the mammalian host. The tick usually must be attached for at least 24 h for transmission of *B. burgdorferi*.

After injection into the human skin, the spirochete downregulates OspC and upregulates the VlsE lipoprotein. This protein undergoes extensive antigenic variation, which is necessary for spirochetal survival. After several days to weeks, *B. burgdorferi* may migrate outward in the skin, producing EM, and may spread hematogenously or in the lymph to other organs. The only known virulence factors of *B. burgdorferi* are surface proteins that allow the spirochete to attach to mammalian proteins, integrins, glycosaminoglycans, or glycoproteins. For example, spread through the skin and other tissue matrices may be facilitated by the binding of human plasminogen and its activators to the surface of the spirochete. Several *Borrelia* strains bind components of complement, such as Factor H and other complement regulators, which help to protect spirochetes from complement-mediated lysis. Dissemination of the organism in the blood is facilitated by binding to the fibrinogen receptor ($\alpha_{IIb}\beta_3$) on activated platelets and the vitronectin receptor ($\alpha_v\beta_3$) on endothelial cells. As the name indicates, spirochetal decorin-binding proteins A and B bind decorin, a glycosaminoglycan on collagen fibrils, and *B. burgdorferi* also binds directly to native type 1 collagen lattices. This binding may explain why the organism is commonly aligned with collagen fibrils in the extracellular matrix in the heart, nervous system, or joints.

To control and eradicate *B. burgdorferi*, the host mounts both innate and adaptive immune responses, resulting in macrophage- and antibody-mediated killing of the spirochete. As part of the innate immune response, complement may lyse the spirochete in the skin. Cells at affected sites release potent proinflammatory cytokines, including interleukin 6, tumor necrosis factor α, interleukin 1β, and interferon γ (IFN-γ). Patients who are homozygous for a Toll-like receptor 1 polymorphism (1805GG), particularly when infected with highly inflammatory *B. burgdorferi* RST1 strains, have exceptionally high levels of proinflammatory cytokines. The purpose of the adaptive immune response appears to be the production of specific antibodies, which opsonize the organism—a step necessary for optimal spirochetal killing. Studies with protein arrays expressing ~1200 *B. burgdorferi* proteins detected antibody responses to a total of 120 spirochetal proteins (particularly outer-surface lipoproteins) in a population of patients with Lyme arthritis. Histologic examination of all affected tissues reveals an infiltration of lymphocytes, macrophages, and plasma cells with some degree of vascular damage, sometimes including obliterative microvascular lesions.

In enzootic infection, *B. burgdorferi* spirochetes must survive this immune assault only during the summer months before returning to larval ticks to begin the cycle again the following year. In contrast, infection of humans is a dead-end event for the spirochete. Within several weeks or months, innate and adaptive immune mechanisms—even without antibiotic treatment—control widely disseminated infection, and generalized systemic symptoms wane. Thus, immune mechanisms seem to succeed eventually in the near or total eradication of *B. burgdorferi* from selected niches, including the joints or nervous system, and symptoms resolve in most patients. However, without antibiotic therapy, spirochetes may survive in localized niches for several more years. For example, *B. burgdorferi* infection in the United States may cause persistent arthritis or, in rare cases, subtle encephalopathy or polyneuropathy, and *B. afzelii* may cause acrodermatitis.

CLINICAL MANIFESTATIONS

Early Infection: Stage 1 (Localized Infection) Because of the small size of nymphal ixodid ticks, most patients do not remember the preceding tick bite. After an incubation period of 3–32 days, EM usually begins as a red macule or papule at the site of the tick bite that expands slowly to form a large annular lesion **(Fig. 191-1)**. As the lesion increases in size, it often develops a bright red outer border and partial central clearing. The center of the lesion sometimes becomes intensely erythematous and indurated, vesicular, or necrotic. In other instances, the expanding lesion remains an even, intense red; several red rings are found within an outside ring; or the central area turns blue before the lesion clears. Although EM can be located anywhere, the thigh, groin, and axilla are particularly common sites. The lesion is warm but not often painful. Approximately 20% of patients do not exhibit this characteristic skin manifestation.

Early Infection: Stage 2 (Disseminated Infection) In cases in the United States, *B. burgdorferi* often spreads hematogenously to many sites within days or weeks after the onset of EM. In these cases, patients may develop secondary annular skin lesions similar in appearance to the initial lesion. Skin involvement is commonly accompanied by severe headache, mild stiffness of the neck, fever, chills, migratory musculoskeletal pain, arthralgias, and profound malaise and fatigue. Less common manifestations include generalized lymphadenopathy or splenomegaly, hepatitis, sore throat, nonproductive cough, conjunctivitis, iritis, or orchitis. Except for fatigue and lethargy, which are often constant, the early signs and symptoms of Lyme disease are typically intermittent and changing. Even in untreated patients, the early symptoms usually become less severe or disappear within several weeks. In ~15% of patients, the infection presents with these nonspecific systemic symptoms.

Symptoms suggestive of meningeal irritation may develop early in Lyme disease when EM is present but usually are not associated with cerebrospinal fluid (CSF) pleocytosis or an objective neurologic deficit. After several weeks or months, ~15% of untreated patients develop frank neurologic abnormalities, including meningitis, subtle encephalitic signs, cranial neuritis (most commonly involving the facial nerve, resulting in unilateral or bilateral facial palsy), motor or sensory radiculoneuropathy, peripheral neuropathy, mononeuritis multiplex, cerebellar ataxia, or myelitis—alone or in various combinations. In children, the optic nerve may be affected because of inflammation or increased intracranial pressure, and these effects may lead to blindness. In the United States, the usual pattern consists of fluctuating symptoms of meningitis accompanied by facial palsy and peripheral radiculoneuropathy. Lymphocytic pleocytosis (~100 cells/μL) is found in CSF, often along with elevated protein levels and normal or slightly low glucose concentrations. In Europe and Asia, the first neurologic sign is characteristically radicular pain, which is followed by the development of CSF pleocytosis (meningopolyneuritis or *Bannwarth's syndrome*); meningeal or encephalitic signs are frequently absent. These early neurologic abnormalities usually resolve completely within months, but in rare cases, chronic neurologic disease may occur later.

Within several weeks after the onset of illness, ~8% of patients develop cardiac involvement. The most common abnormality is a fluctuating degree of atrioventricular block (first-degree, Wenckebach, or complete heart block). Some patients have more diffuse cardiac involvement, including electrocardiographic changes indicative of acute myopericarditis, left ventricular dysfunction evident on radionuclide scans, or (in rare cases) cardiomegaly or fatal pancarditis. Cardiac involvement lasts for only a few weeks in most patients but may recur in untreated patients. A few cases of mitral or aortic valve endocarditis have been reported, in one case occurring years after acute cardiac involvement of Lyme disease. Chronic cardiomyopathy caused by *B. burgdorferi* has been reported in Europe.

During this stage, musculoskeletal pain is common. The typical pattern consists of migratory pain in joints, tendons, bursae, muscles, or bones (usually without joint swelling) lasting for hours or days and affecting one or two locations at a time.

Late Infection: Stage 3 (Persistent Infection) Months after the onset of infection, ~60% of patients in the United States who have received no antibiotic treatment develop frank arthritis. The typical pattern comprises intermittent attacks of oligoarticular arthritis in large joints (especially the knees), lasting for weeks or months in a given joint. A few small joints or periarticular sites also may be affected, primarily during early attacks. The number of patients who continue to have recurrent attacks decreases each year. However, in a small percentage of cases, involvement of large joints—usually one or both knees—is persistent and may lead to erosion of cartilage and bone.

White cell counts in joint fluid range from 500 to 110,000/μL (average, 25,000/μL); most of these cells are polymorphonuclear leukocytes. Tests for rheumatoid factor or antinuclear antibodies usually give negative results, but a low-titer antinuclear antibody value may occur. Examination of synovial biopsy samples reveals fibrin deposits, villous hypertrophy, vascular proliferation, microangiopathic lesions, and a heavy infiltration of lymphocytes and plasma cells.

Although most patients with Lyme arthritis respond well to antibiotic therapy, a small percentage in the northeastern United States have persistent postinfectious (also called postantibiotic or antibiotic-refractory) Lyme arthritis for months or even for several years after receiving oral and IV antibiotic therapy for 2 or 3 months. Although more often these patients are initially infected with OspA type A (RST1) strains of *B. burgdorferi*, this complication is not thought to result from persistent infection. Results of culture and polymerase chain reaction (PCR) for *B. burgdorferi* in synovial tissue obtained in the postantibiotic period have been uniformly negative. Rather, the basic pathogenetic feature of postinfectious Lyme arthritis is the development of an excessive, dysregulated proinflammatory immune response during the infection, characterized by exceptionally high IFN-γ levels, which persist in the postinfectious period. Risk factors for excessively high IFN-γ responses include presentation of an epitope of *B. burgdorferi* OspA ($\text{OspA}^{164\text{-}175}$) by certain class II major histocompatibility complex molecules (particularly HLA-DRBI*0401); a Toll-like receptor 1 polymorphism 1805GG in patients who were infected with OspC type A (RST1) *B. burgdorferi* strains; and an imbalance of the CD4+ T effector/regulatory cell ratio in which the majority of CD4+CD25+ T cells, which are ordinarily regulatory T cells, become IFN-γ-secreting T effector cells.

The consequences of this excessive proinflammatory response in Lyme synovia include vascular damage, autoimmune and cytotoxic processes, and tumor-like fibroblast proliferation and fibrosis. An important driver of innate immune responses may be persistence of *B. burgdorferi* peptidoglycan in synovial fluid, which may be especially difficult to clear. In addition, seven autoantigens that are targets of T- and B-cell responses in patients with Lyme disease, particularly those with postinfectious arthritis, have now been identified. These include three autoantigens associated with the vasculature (i.e., endothelial cell growth factor, apolipoprotein B-100, and annexin A2) and four extracellular matrix (ECM) proteins (i.e., matrix metalloproteinase 10, fibronectin-1, laminin B2, and collagen Vα1). Autoantibodies against vascular antigens are associated with obliterative microvascular lesions, and T-cell responses to epitopes of ECM proteins are associated with significantly longer durations of postinfectious arthritis.

Although rare, chronic neurologic involvement also may become apparent months to several years after the onset of infection, sometimes after long periods of latent infection. The most common form of chronic central nervous system involvement is subtle encephalopathy affecting memory, mood, or sleep, and the most common form of peripheral neuropathy is an axonal polyneuropathy manifested as either distal paresthesia or spinal radicular pain. Patients with encephalopathy frequently have evidence of memory impairment in neuropsychological tests and abnormal results in CSF analyses. In cases of polyneuropathy, electromyography generally shows extensive abnormalities of proximal and distal nerve segments. Encephalomyelitis or leukoencephalitis, a rare manifestation of Lyme borreliosis associated primarily with *B. garinii* infection in Europe, is a severe neurologic disorder that may include spastic paraparesis, upper motor neuron bladder dysfunction, and, rarely, lesions in the periventricular white matter.

Acrodermatitis chronica atrophicans, the late skin manifestation of Lyme borreliosis, has been associated primarily with *B. afzelii* infection in Europe and Asia. It has been observed especially often in elderly women. The skin lesions, which are usually found on the acral surface of an arm or leg, begin insidiously with reddish-violaceous discoloration; they become sclerotic or atrophic over a period of years.

The basic patterns of Lyme borreliosis are similar worldwide, but there are regional variations, primarily between the illness found in North America, which is caused exclusively by *B. burgdorferi*, and that found in Europe, which is caused primarily by *B. afzelii*, *B. garinii*, and *B. bavariensis*. With each of the *Borrelia* species, the infection usually begins with EM. However, *B. burgdorferi* strains in the eastern United States often disseminate widely; they are particularly arthritogenic, and especially OspC type A (RST1) strains may lead to postinfectious arthritis. *B. garinii* and *B. bavariensis* typically disseminate less widely, but are especially neurotropic, are more likely to cause typical neuroborreliosis (Bannwarth's syndrome) and rarely may cause borrelial encephalomyelitis. *B. afzelii* often infects only the skin but may persist in that site, where it may cause several different dermatoborrelioses, including acrodermatitis chronica atrophicans.

Posttreatment Lyme Disease Syndrome (PTLDS) Despite resolution of the objective manifestations of the infection with antibiotic therapy, ~10% of patients (although the reported percentages vary widely) continue to have subjective pain, neurocognitive manifestations, or fatigue symptoms. This disabling problem has been known for years following certain other infections but has been brought to fore recently with postacute sequelae of COVID-19. In Lyme disease, these symptoms usually improve and resolve within months but may last for years. At the far end of the spectrum, the symptoms may be similar to or indistinguishable from chronic fatigue syndrome **(Chap. 461)** and fibromyalgia **(Chap. 385)**. Compared with symptoms of active Lyme disease, post-Lyme symptoms tend to be more generalized or disabling. They include marked fatigue, severe headache, diffuse musculoskeletal pain, multiple symmetric tender points in characteristic locations, pain and stiffness in many joints, diffuse paresthesias, difficulty with concentration, and sleep disturbances. Patients with this condition lack evidence of joint inflammation, have normal neurologic test results, and may exhibit anxiety and depression. In contrast, late manifestations of Lyme disease, including arthritis, encephalopathy, and neuropathy, are usually associated with minimal systemic symptoms. Currently, no evidence indicates that persistent subjective symptoms after recommended courses of antibiotic therapy are caused by active infection. Randomized controlled trials have shown than repeated courses of antibiotics do not improve the symptoms of PTLDS and are not recommended.

DIAGNOSIS

The culture of *B. burgdorferi* in Barbour-Stoenner-Kelly (BSK) medium permits definitive diagnosis, but this method has been used primarily in research studies. Moreover, with a few exceptions, positive cultures have been obtained only early in the illness—particularly from biopsy samples of EM skin lesions, less often from plasma samples, and occasionally from CSF samples. Later in the infection, PCR is greatly superior to culture for the detection of *B. burgdorferi* DNA in joint fluid; this is the major use for PCR testing in Lyme disease. However, because *B. burgdorferi* DNA may persist for at least weeks after spirochetal killing with antibiotics, detection of spirochetal DNA in joint fluid is not an accurate test of active joint infection in Lyme disease and cannot be used reliably to determine the adequacy of antibiotic therapy. The sensitivity of PCR determinations in CSF from patients with neuroborreliosis has been much lower than that in joint fluid. With current methods, there seems to be little if any role for PCR in the detection of *B. burgdorferi* DNA in blood or urine samples.

Because of the problems associated with direct detection of *B. burgdorferi*, Lyme disease is usually diagnosed by the recognition of a characteristic clinical picture accompanied by serologic confirmation. Although serologic testing may yield negative results during the first several weeks of infection, almost all patients have a positive antibody response to *B. burgdorferi* after that time when a two-test approach of enzyme-linked immunosorbent assay (ELISA) and Western blot or a protocol of two enzyme immunoassays (EIAs) is used. The limitation of serologic tests is that they do not clearly distinguish between active and inactive infection. After antibiotic therapy, the amount of antibody declines but the results of Western blot, a nonquantitative test, do not change much (or very slowly). Thus, patients with previous Lyme disease—particularly in cases progressing to late stages—often remain seropositive for years, even after adequate antibiotic therapy. In addition, ~10% of patients are seropositive because of asymptomatic infection. If individuals with past or asymptomatic *B. burgdorferi* infection subsequently develop another illness, the positive serologic test for Lyme disease may cause diagnostic confusion. According to an algorithm published by the American College of Physicians **(Table 191-1)**, serologic testing for Lyme disease is recommended only for patients with at least an intermediate pretest probability of Lyme disease, such as those with oligoarticular arthritis. It should not be used as a screening procedure in patients with pain or fatigue syndromes. In such patients, the probability of a false-positive serologic result is higher than that of a true-positive result.

TABLE 191-1 Algorithm for Testing for and Treating Lyme Disease

PRETEST PROBABILITY	EXAMPLE	RECOMMENDATION
High	Patients with erythema migrans	Empirical antibiotic treatment without serologic testing
Intermediate	Patients with oligoarticular arthritis	Serologic testing and antibiotic treatment if test results are positive
Low	Patients with nonspecific symptoms (myalgias, arthralgias, fatigue)	Neither serologic testing nor antibiotic treatment

Source: Adapted from the recommendations of the American College of Physicians (G Nichol et al: Ann Intern Med 128:37, 1998).

For serologic analysis of Lyme disease in the United States, the CDC recommends a two-step approach in which samples are first tested by ELISA, and equivocal or positive results are then tested by Western blot. This is called the conventional two-test approach. During the first weeks of infection, both IgM and IgG responses to the spirochete should be determined, preferably in both acute- and convalescent-phase serum samples. Approximately 20–30% of patients have a positive response detectable in acute-phase samples (usually only a positive IgM response), whereas ~70–80% have a positive response during convalescence (2–4 weeks later). After 4–8 weeks of infection (by which time most patients with active Lyme disease have disseminated infection), the sensitivity and specificity of the IgG response to the spirochete are both very high—in the range of 99%—as determined by the two-test approach of ELISA and Western blot. At this point and thereafter, a single test (that for IgG) is usually sufficient. In persons with illness of >2 months' duration, a positive IgM test result alone is likely to be false-positive and therefore should not be used to support the diagnosis.

According to current criteria adopted by the CDC, an IgM Western blot is considered positive if two of the following three bands are present: 23, 39, and 41 kDa. However, the combination of two such bands may still represent a false-positive result. Misuse or misinterpretation of IgM blots has been a factor in the incorrect diagnosis of Lyme disease in patients with other illnesses. An IgG blot is considered positive if 5 of the following 10 bands are present: 18, 23, 28, 30, 39, 41, 45, 58, 66, and 93 kDa. In European cases, no single set of criteria for the interpretation of immunoblots results in high levels of sensitivity and specificity in all countries.

A new methodology called the modified two-test approach, which is now approved by the U.S. Food and Drug Administration, is a two-test approach using two EIAs, thereby dispensing with the Western blot. One such method employs a whole–*B. burgdorferi* sonicate ELISA followed by a VlsE C6 peptide IgG ELISA. This approach, which gives simply a positive or a negative result, increases sensitivity during the first several weeks of infection without compromising specificity. For more complex cases or in those with late infection, it is still valuable to determine antibody specificities to multiple spirochetal proteins, as is done with Western blots. More recently, line immunoblots or other multiplexed antibody platforms have been developed as substitutes for Western blots. These assays allow more objective interpretation, and some platforms can provide quantitative data about antibody responses to many spirochetal proteins. After successful antibiotic treatment, antibody titers decline slowly, but responses (including that to the VlsE C6 peptide) may persist for years. Moreover, not only the IgG but also the IgM response may persist for years after therapy. Therefore, even a positive IgM response cannot be interpreted as confirmation of recent infection or reinfection unless the clinical picture is appropriate.

■ DIFFERENTIAL DIAGNOSIS

Classic EM is a slowly expanding erythema, often with partial central clearing. If the lesion expands little, it may represent the red papule of an uninfected tick bite. If the lesion expands rapidly, it may represent cellulitis (e.g., streptococcal cellulitis) or an allergic reaction, perhaps to tick saliva. Patients with secondary annular lesions may be thought to have erythema multiforme, but neither the development of blistering mucosal lesions nor the involvement of the palms or soles is a feature of *B. burgdorferi* infection. In the eastern United States, an EM-like skin lesion, sometimes with mild systemic symptoms, may be associated with *Amblyomma americanum* tick bites. However, the cause of this southern tick-associated rash illness (STARI) has not yet been identified. This tick may also transmit *Ehrlichia chaffeensis*, a rickettsial agent **(Chap. 192)**.

As stated above, *I. scapularis* ticks in the United States may transmit not only *B. burgdorferi* but also *B. microti*, the red blood cell parasite causing babesiosis **(Chap. 232)**; *A. phagocytophilum*, the agent of human granulocytotropic anaplasmosis **(Chap. 192)**; *B. miyamotoi*, a relapsing fever spirochete **(Chap. 190)**; *B. mayonii* and *E. muris*–like agent, newly recognized species that occur in the upper midwestern United States; or less commonly, Powassan virus (the deer tick virus, which is closely related to European tick-borne encephalitis virus), which may cause fatal infection **(Chap. 215)**. Babesiosis, anaplasmosis, *B. miyamotoi,* and *B. mayonii* typically cause an influenza-like syndrome with fever, myalgias, and cytopenia, but symptoms may range from asymptomatic infection to severe or even fatal disease, particularly in the young or the elderly. Co-infected patients may have more severe or persistent symptoms than patients infected with a single agent. Standard blood counts may yield clues regarding the presence of co-infection. Anaplasmosis and *B. miyamotoi* may cause leukopenia or thrombocytopenia, and babesiosis may cause thrombocytopenia and hemolytic anemia. IgM serologic responses may confuse the diagnosis. For example, *A. phagocytophilum* may elicit a positive IgM response to *B. burgdorferi*. PCR of blood is the diagnostic test of choice for co-infection with *A. phagocytophilum, B. microti,* or *B. miyamotoi*. Alternatively, examination of a peripheral blood smear can be used to detect *B. microti,* but blood smear analysis is insensitive for *A. phagocytophilum* and *B. miyamotoi*. The frequency of co-infection in different studies has been variable. In one prospective study, 4% of patients with EM had evidence of co-infection, although this appears to be an increasing problem in early infection.

Facial palsy caused by *B. burgdorferi*, which occurs in the early disseminated phase of the infection (often in July, August, or September), is usually recognized by its association with EM. However, facial palsy without EM may be the presenting manifestation of Lyme disease. In such cases, both the IgM and the IgG responses to the spirochete are usually positive. The most common infectious agents that cause facial palsy are herpes simplex virus type 1 (Bell's palsy; **Chap. 197**) and varicella-zoster virus (Ramsay Hunt syndrome; **Chap. 198**).

Later in the infection, oligoarticular Lyme arthritis most resembles peripheral spondyloarthropathy in an adult or the pauciarticular form of juvenile idiopathic arthritis in a child. Patients with Lyme arthritis usually have the strongest IgG antibody responses seen in Lyme borreliosis, with reactivity to many spirochetal proteins.

The most common problem in the diagnosis of early Lyme disease is to miss the diagnosis, either because of an atypical morphology of EM or the erroneous assumption that a negative serologic test for *B. burgdorferi* excludes acute disease. Clinicians should be aware of the typical and atypical manifestations of EM, recognize the limited value of serologic testing in evaluating patients with early Lyme disease, and understand that the diagnosis of early Lyme disease is a clinical one.

The most common problem in diagnosis of late-stage Lyme disease is to mistake chronic fatigue syndrome **(Chap. 461)** or fibromyalgia **(Chap. 385)** for Lyme disease. This difficulty is compounded by the fact that a small percentage of patients with Lyme disease do in fact develop these chronic pain or fatigue syndromes in association with or soon after Lyme disease. Moreover, a counterculture has emerged that ascribes pain and fatigue syndromes to chronic Lyme disease when there is little or no evidence of *B. burgdorferi* infection. In such cases, the term *chronic Lyme disease*, which is equated with chronic *B. burgdorferi* infection, is a misnomer, and the use of repeated courses of antibiotic treatment is not warranted. Well-controlled randomized trials have found no benefit to antibiotic therapy for PTLDS, whereas there is well-documented risk of harm.

TREATMENT

Lyme Borreliosis

ANTIBIOTIC TREATMENT

As outlined in the algorithm in **Fig. 191-2**, the various manifestations of Lyme disease can usually be treated successfully with orally administered antibiotics; the exceptions are severe objective neurologic abnormalities and third-degree atrioventricular heart block, which are generally treated with IV antibiotics, and arthritis that does not respond to oral therapy. For early Lyme disease, doxycycline is effective and can be administered to men, nonpregnant women, and children older than age 8. An advantage of this regimen is that it is also effective against *A. phagocytophilum*, *B. miyamotoi*, and *B. mayonii*, which are transmitted by the same tick that transmits the Lyme disease agent. Amoxicillin, cefuroxime axetil, and erythromycin or its congeners are second-, third-, and

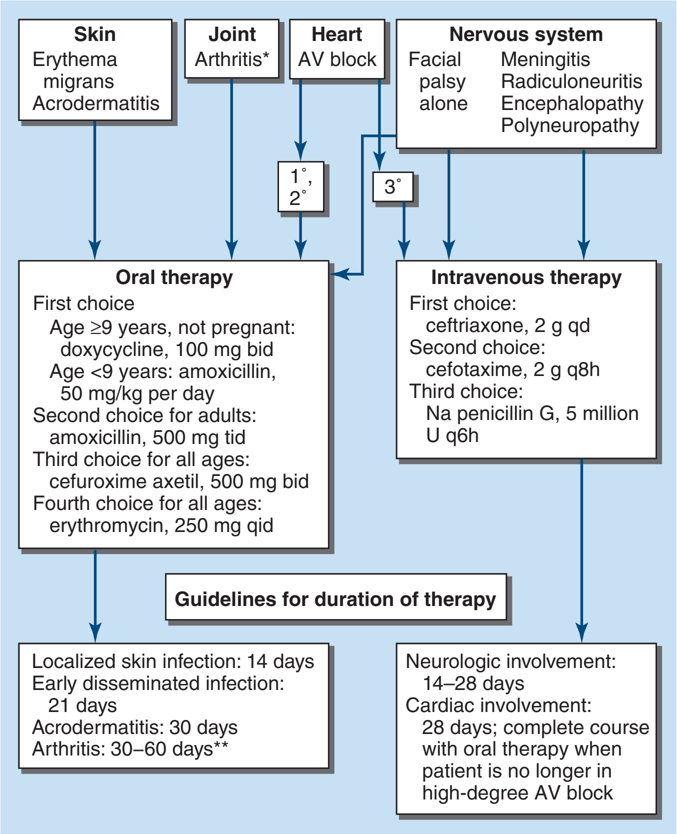
Skin
Erythema migrans
Acrodermatitis
Joint
Arthritis*
Heart
AV block
Nervous system
Facial palsy alone
Meningitis
Radiculoneuritis
Encephalopathy
Polyneuropathy
1°, 2°
3°
Oral therapy
First choice
Age ≥9 years, not pregnant: doxycycline, 100 mg bid
Age <9 years: amoxicillin, 50 mg/kg per day
Second choice for adults: amoxicillin, 500 mg tid
Third choice for all ages: cefuroxime axetil, 500 mg bid
Fourth choice for all ages: erythromycin, 250 mg qid
Intravenous therapy
First choice: ceftriaxone, 2 g qd
Second choice: cefotaxime, 2 g q8h
Third choice: Na penicillin G, 5 million U q6h
Guidelines for duration of therapy
Localized skin infection: 14 days
Early disseminated infection: 21 days
Acrodermatitis: 30 days
Arthritis: 30–60 days**
Neurologic involvement: 14–28 days
Cardiac involvement: 28 days; complete course with oral therapy when patient is no longer in high-degree AV block

FIGURE 191-2 Algorithm for the treatment of the various early or late manifestations of Lyme borreliosis. AV, atrioventricular. *For arthritis, oral therapy should be tried first; if arthritis is unresponsive, IV therapy should be administered. **For Lyme arthritis, IV ceftriaxone (2 g given once a day for 14–28 days) also is effective and is necessary for patients who do not respond to oral therapy. However, compared with oral treatment, this regimen is less convenient to administer, has more side effects, and is more expensive.

fourth-choice alternatives, respectively, for the treatment of Lyme disease. In children, amoxicillin is effective (not >2 g/d); in cases of penicillin allergy, cefuroxime axetil or erythromycin may be used. In contrast to second- or third-generation cephalosporin antibiotics, first-generation cephalosporins, such as cephalexin, are not effective. For patients with infection localized to the skin, a 10-day course of doxycycline or a 14-day course of amoxicillin is generally sufficient; in contrast, for patients with early disseminated infection, a 14- to 21-day course is recommended. Approximately 15% of patients experience a Jarisch-Herxheimer-like reaction during the first 24 h of therapy. In multicenter studies, >90% of patients whose early Lyme disease was treated with these regimens had satisfactory outcomes. Although some patients reported symptoms after treatment, objective evidence of persistent infection or relapse was rare, and re-treatment was usually unnecessary.

Oral administration of doxycycline or amoxicillin for 30 days is recommended for the initial treatment of Lyme arthritis in patients who do not have concomitant neurologic involvement. Among patients with arthritis who have an incomplete response to oral antibiotics, a second, 30-day course of oral antibiotics may be successful. However, among patients with arthritis who have minimal or no response to oral antibiotics, re-treatment with IV ceftriaxone for 28 days is appropriate. In patients with arthritis in whom joint inflammation persists for months or even several years after both oral and IV antibiotics, treatment with nonsteroidal anti-inflammatory agents, therapy with disease-modifying antirheumatic drugs, or synovectomy may be successful.

In the United States, parenteral antibiotic therapy is usually used for severe objective neurologic abnormalities. Patients with such abnormalities are most commonly treated with IV ceftriaxone for 14–28 days, but IV cefotaxime or IV penicillin G for the same duration also may be effective. In Europe, similar results have been obtained with oral doxycycline and IV antibiotics in the treatment of acute neuroborreliosis. Although systematic trials have not been conducted in the United States, oral doxycycline is now used by many clinicians in this country for the treatment of patients with less severe neurologic abnormalities, such as facial palsy alone or uncomplicated Lyme meningitis. In patients with high-degree atrioventricular block or a PR interval of >0.3 s, IV therapy for at least part of the course and cardiac monitoring are recommended, but the insertion of a permanent pacemaker is not necessary.

It is unclear how and whether asymptomatic infection should be treated, but patients with such infection are often given a course of oral antibiotics. Because maternal–fetal transmission of *B. burgdorferi* seems to occur rarely (if at all), standard therapy for the manifestations of the illness is recommended for pregnant women. Long-term persistence of *B. burgdorferi* has not been documented in any large series of patients after treatment with currently recommended regimens, but there are a few case reports of persistent infection after such regimens. Although an occasional patient requires a second course of antibiotics, there is no indication for multiple, repeated antibiotic courses in the treatment of Lyme disease.

CHRONIC LYME DISEASE

After appropriately treated Lyme disease, a small percentage of patients continue to have subjective symptoms, primarily musculoskeletal pain, neurocognitive difficulties, or fatigue. This syndrome, termed *posttreatment Lyme disease syndrome* (PTLDS), is sometimes a disabling condition that is similar to chronic fatigue syndrome or fibromyalgia. Five double-blind, placebo-controlled trials conducted in the United States and Europe have failed to show benefit of further antibiotic therapy in these patients. For example, in a large study, one group of patients with PTLDS received IV ceftriaxone for 30 days followed by oral doxycycline for 60 days, while another group received IV and oral placebo preparations for the same durations. No significant differences were found between groups in the numbers of patients reporting that their symptoms had improved, become worse, or stayed the same. Such patients are best treated for the relief of symptoms rather than with prolonged courses of antibiotics.

PROPHYLAXIS AFTER A TICK BITE

The risk of infection with *B. burgdorferi* after a recognized tick bite is so low that antibiotic prophylaxis is not routinely indicated. However, if an attached, engorged *I. scapularis* nymph is found or if follow-up is anticipated to be difficult, a single 200-mg dose of doxycycline, which usually prevents Lyme disease when given within 72 h after the tick bite, may be administered.

PROGNOSIS

The response to treatment is best early in the disease. Later treatment of Lyme borreliosis is still effective, but the period of convalescence may be longer. Eventually, most patients recover with minimal or no residual deficits.

REINFECTION

Reinfection may occur after EM when patients are treated with antimicrobial agents. In such cases, the immune response is not adequate to provide protection from subsequent infection. However, patients who develop an expanded immune response to the spirochete over a period of months (e.g., those with Lyme arthritis) have protective immunity for a period of years and rarely, if ever, acquire the infection again.

PREVENTION

Protective measures for the prevention of Lyme disease may include the avoidance of tick-infested areas, the use of repellents and acaricides, tick checks, and modification of landscapes in or near residential areas. Although a vaccine for Lyme disease used to be available, the manufacturer has discontinued its production. Another company is currently testing a similar vaccine in both the United States and Europe. However, no vaccine is currently available commercially for the prevention of this infection.

FURTHER READING

Arvikar SL, Steere AC: Diagnosis and treatment of Lyme arthritis. Infect Dis Clin North Am 36:563, 2022.

Aucott JN: Posttreatment Lyme disease syndrome. Infect Dis Clin North Am 29:309, 2015.

Branda JA, Steere AC: Laboratory diagnosis of Lyme borreliosis. Clin Micro Rev 34:e00018, 2021.

Branda JA et al: Two-tiered antibody testing for Lyme disease with use of 2 enzyme immunoassays, a whole-cell sonicate enzyme immunoassay followed by a VlsE C6 peptide enzyme immunoassay. Clin Infect Dis 53:541, 2011.

Klempner MS et al: Two controlled trials of antibiotic treatment in patients with persistent symptoms and a history of Lyme disease. N Engl J Med 345:85, 2001.

Lantos PM et al: Clinical practice guidelines by the Infectious Diseases Society of America (IDSA), American Academy of Neurology, and the American College of Rheumatology (ACR): 2020 guidelines for the prevention, diagnosis, and treatment of Lyme disease. Clin Infect Dis 72:1, 2021.

Lemieux JE et al: Whole genome sequencing of human *Borrelia burgdorferi* isolates reveals linked blocks of accessory genome elements located on plasmids and associated with human dissemination. PLoS Pathog 17:e1011243, 2023.

Li X et al: Burden and viability of *Borrelia burgdorferi* in skin or joints of patients with erythema migrans or Lyme arthritis. Arthritis Rheum 63:2238, 2011.

Lochhead RB et al: Lyme arthritis, linking infection, inflammation, and autoimmunity. Nat Rev Rheumatol 17:449, 2021.

Oschmann P et al: Stages and syndromes of neuroborreliosis. J Neurol 245:262, 1998.

Steere AC: Lyme disease. N Engl J Med 345:115, 2001.

Steere AC: Posttreatment Lyme disease syndromes: Distinct pathogenesis caused by maladaptive host responses. J Clin Invest 130:2148, 2020.

Steere AC et al: Prospective study of serologic tests for Lyme disease. Clin Infect Dis 47:188, 2008.

Steere AC et al: Lyme borreliosis. Nat Rev Dis Primers 2:16090, 2016.

Ursinus J et al: Prevalence of persistent symptoms after treatment for Lyme borreliosis: A prospective observational cohort study. Lancet Reg Health Eur 6:100142, 2021. Erratum in: Lancet Reg Health Eur 27:100622, 2023.

Section 10 Diseases Caused by Rickettsiae, Mycoplasmas, and Chlamydiae

192

Rickettsial Diseases

David H. Walker, J. Stephen Dumler,
Lucas S. Blanton, Chantal P. Rovers

Rickettsiae are a heterogeneous group of small, obligately intracellular, gram-negative coccobacilli and short bacilli, most of which are transmitted by a tick, mite, flea, or louse vector. Except in the case of louse-borne typhus, humans are incidental hosts. Among rickettsiae, *Coxiella burnetii*, *Rickettsia prowazekii*, and *Rickettsia typhi* have the well-documented ability to survive for an extended period outside the reservoir or vector and to be extremely infectious: inhalation of a single *Coxiella* microorganism can cause pneumonia. High-level infectivity and severe illness after inhalation make *R. prowazekii*, *R. rickettsii*, *R. typhi*, *R. conorii*, and *C. burnetii* bioterrorism threats **(Chap. S4)**.

Clinical infections with rickettsiae can be classified according to (1) the taxonomy and diverse microbial characteristics of the agents, which belong to seven genera (*Rickettsia*, *Orientia*, *Ehrlichia*, *Anaplasma*, *Neorickettsia*, "*Candidatus* Neoehrlichia," and *Coxiella*); (2) epidemiology; or (3) clinical manifestations. The clinical manifestations of all the acute presentations are similar during the first 5 days: fever, headache, and myalgias with or without nausea, vomiting, and cough. As the course progresses, clinical manifestations—including a macular, maculopapular, or vesicular rash; eschar; pneumonitis; and meningoencephalitis—vary from one disease to another. Given the many etiologic agents with varied mechanisms of transmission, geographic distributions, and associated disease manifestations, the consideration of rickettsial diseases as a single entity poses complex challenges **(Table 192-1)**.

Establishing the etiologic diagnosis of rickettsioses is very difficult during the acute stage of illness, and definitive diagnosis usually requires the examination of serum samples during the acute and convalescent phases of illness. Heightened clinical suspicion is based on epidemiologic data, history of exposure to vectors or reservoir animals, travel to endemic locations, clinical manifestations (sometimes including rash or eschar), and characteristic laboratory findings (including thrombocytopenia, normal or low white blood cell [WBC] counts, elevated hepatic enzyme levels, and hyponatremia). Such suspicion should prompt empirical treatment. Doxycycline is the empirical drug of choice for most of these infections. Only one agent, *C. burnetii*, has been documented to cause chronic illness. One other species, *R. prowazekii*, causes recrudescent illness (Brill-Zinsser disease) when latent infection is reactivated years after resolution of the acute illness.

Rickettsial infections dominated by fever may resolve without further clinical evolution. However, after nonspecific early manifestations, the illnesses can also evolve along one or more of several principal clinical lines: (1) development of a macular or maculopapular rash; (2) development of an eschar at the site of tick or mite feeding, which can occur during the incubation period; (3) development of a vesicular rash (often in rickettsialpox, *R. parkeri* infection, and African tick-bite fever); (4) development of pneumonitis with chest radiographic opacities and/or rales (Q fever and severe cases of Rocky Mountain spotted fever [RMSF], Mediterranean spotted fever [MSF], louse-borne typhus, human monocytotropic ehrlichiosis [HME], human granulocytotropic anaplasmosis [HGA], scrub typhus, and murine typhus); (5) development of meningoencephalitis (louse-borne typhus and severe cases of RMSF, scrub typhus, HME, murine typhus, MSF, and [rarely] Q fever); and (6) progressive hypotension and multiorgan failure as seen with sepsis or toxic shock syndromes (RMSF, MSF, louse-borne typhus, murine typhus, scrub typhus, HME, HGA, and neoehrlichiosis).

Epidemiologic clues to the transmission of a particular pathogen include (1) environmental exposure to ticks, fleas, or mites during the season of activity of the vector species for the disease in the appropriate geographic region (spotted fever and typhus rickettsioses, scrub typhus, ehrlichiosis, anaplasmosis); (2) travel to or residence in an endemic geographic region during the incubation period (Table 192-1); (3) exposure to parturient ruminants, cats, and dogs (Q fever); (4) exposure to flying squirrels (*R. prowazekii* infection); and (5) history of previous louse-borne typhus (recrudescent typhus).

Clinical laboratory findings such as thrombocytopenia (particularly in spotted fever and typhus rickettsioses, ehrlichiosis, anaplasmosis, and scrub typhus), normal or low WBC counts, mild to moderate serum elevations of hepatic aminotransferases, and hyponatremia suggest some common pathophysiologic mechanisms.

Application of these clinical, epidemiologic, and laboratory principles requires consideration of a rickettsial diagnosis and knowledge of the individual diseases.

TICK-, MITE-, LOUSE-, AND FLEA-BORNE RICKETTSIOSES

These diseases, caused by organisms of the genera *Rickettsia* and *Orientia* in the family Rickettsiaceae, result from endothelial cell infection and increased vascular permeability. Pathogenic rickettsial species are very closely related, have small genomes (as a result of reductive evolution, which eliminated many genes for biosynthesis of intracellularly available molecules), and are traditionally separated into typhus and spotted fever groups on the basis of lipopolysaccharide antigens. Some

TABLE 192-1 Features of Selected Rickettsial Infections

DISEASE	ORGANISM	TRANSMISSION	GEOGRAPHIC RANGE	INCUBATION PERIOD, DAYS	DURATION, DAYS	RASH, %	ESCHAR, %	LYMPHADENOPATHY[a]
Rocky Mountain spotted fever (RMSF)	*Rickettsia rickettsii*	Tick bite: *Dermacentor andersoni, D. variabilis, D. similis*	United States	2–14	10–20	90	<1	+
		Amblyomma cajennense, A sculptum, A. patinoi, A. mixtum, A. aureolatum	Central/ South America					
		Rhipicephalus sanguineus	Mexico, South America, United States					
Mediterranean spotted fever (MSF)	*R. conorii*	Tick bite: *R. sanguineus, R. pumilio*	Southern Europe, Africa, Middle East, central Asia	5–7	7–14	97	50	+
African tick-bite fever	*R. africae*	Tick bite: *A. hebraeum, A. variegatum*	Sub-Saharan Africa, West Indies	4–10	4–19	50	90	+++
Maculatum disease	*R. parkeri*	Tick bite: *A. maculatum, A. triste, A. tigrinum, A. ovale*	United States, South America	2–10	6–16	88	94	++
Pacific Coast tick fever	*Rickettsia ssp californica*	Tick bite: *D. occidentalis*	United States	3–9	5–14	14	100	+++
Rickettsialpox	*R. akari*	Mite bite: *Liponyssoides sanguineus*	United States, Ukraine, Turkey, Mexico, Croatia	10–17	3–11	100	90	+++
Tick-borne lymphadenopathy	*R. slovaca, R. raoultii*	Tick bite: *D. marginatus, D. reticularis*	Europe	7–9	17–180	5	100	++++
Flea-borne spotted fever	*R. felis*	Flea (mechanism undetermined): *Ctenocephalides felis*	Worldwide	8–16	8–16	80	15	—
Epidemic typhus	*R. prowazekii*	Louse feces: *Pediculus humanus humanus*, fleas and lice of flying squirrels, or recrudescence	Worldwide	7–14	10–18	80	None	—
Murine typhus	*R. typhi*	Flea feces: *Xenopsylla cheopis, C. felis*, others	Worldwide	8–16	9–18	50	None	—
Human monocytotropic ehrlichiosis (HME)	*Ehrlichia chaffeensis*	Tick bite: *A. americanum, D. variabilis*	United States	1–21	3–21	26	None	++
Ewingii ehrlichiosis	*E. ewingii*	Tick bite*: A. americanum*	United States	1–21	4–21	0	None	
Unnamed ehrlichiosis	*E. muris* ssp. *eauclairensis*	Tick bite*: Ixodes scapularis*	United States	Unknown	3–14	12	None	
Human granulocytotropic anaplasmosis (HGA)	*Anaplasma phagocytophilum*	Tick bite: *I. scapularis, I. ricinus, I. pacificus, I. persulcatus, Haemaphysalis concinna*	United States, Europe, Asia	4–8	3–14	Rare	None	—
Unnamed disease	*A. capra*	*I. persulcatus*	Northeastern China, France	Unknown	11–21	17	9	+
Neoehrlichiosis	"*Candidatus* Neoehrlichia mikurensis"	Tick bite: *I. ricinus, I. persulcatus, Haemaphysalis concinna*	Europe, China	≥8	11–75	10	None	
Scrub typhus	*Orientia tsutsugamushi*	Mite bite: *Leptotrombidium deliense*, others	Asia, Australia, Pacific and Indian Ocean islands	9–18	6–21	50	35	+++
	O. chuto	Mite bite	Africa and Western Asia					
	Candidatus O. chiloensis	Mite bite	Chile					
Q fever	*Coxiella burnetii*	Inhalation of aerosols of infected parturition material (goats, sheep, cattle, cats, others), ingestion of infected milk or milk products	Worldwide except New Zealand, Antarctica	3–30	5–57	4–18%	None	—

[a]++++, severe; +++, marked; ++, moderate; +, present in a small proportion of cases; —, not a noted feature.

diseases and their agents (e.g., *R. africae*, *R. parkeri*, and *R. sibirica*) are too similar to require separate descriptions. Indeed, the similarities among MSF (*R. conorii* [all strains] and *R. massiliae*), North Asian tick typhus (*R. sibirica*), Japanese spotted fever (*R. japonica*), and Flinders Island spotted fever (*R. honei*) far outweigh their minor variations. The Rickettsiaceae that cause life-threatening infections are, in order of decreasing case–fatality rate, *R. rickettsii* (RMSF); *R. prowazekii* (louse-borne typhus); *Orientia tsutsugamushi* (scrub typhus); *R. conorii* (MSF); *R. typhi* (murine typhus); and, in rare cases, other spotted fever–group (SFG) organisms. Some agents (e.g., *R. parkeri*, *R. africae*, *Rickettsia* 364D, *R. akari*, *R. slovaca*, *R. honei*, *R. felis*, *R. massiliae*, *R. helvetica*, *R. heilongjiangensis*, *R. aeschlimannii*, and *R. monacensis*) have never been documented to cause a fatal illness. The most prevalent SFG rickettsia in the United States, *R. amblyommatis*, has been circumstantially associated with asymptomatic seroconversion in most persons and with self-limited illness in others.

ROCKY MOUNTAIN SPOTTED FEVER

Epidemiology RMSF occurs in 47 states (with the highest prevalence in the south-central and southeastern states) as well as in Canada, Mexico, and Central and South America. The infection is transmitted by *Dermacentor variabilis*, the American dog tick, in the eastern two-thirds of the United States; by *D. andersoni*, the Rocky Mountain wood tick, in the western United States; by *D. similis* in California; by *Rhipicephalus sanguineus*, the brown dog tick, in Mexico, Arizona, and probably Colombia; and by *Amblyomma sculptum*, *A. mixtum*, *A. patinoi*, *A. cajennense*, *A. tonelliae*, and *A. aureolatum* in Central and/or South America. Maintained partially by transovarian transmission from one generation of ticks to the next, *R. rickettsii* can be acquired by uninfected ticks through the ingestion of a blood meal from rickettsemic small mammals or by co-feeding adjacent to an infected tick.

Humans become infected during tick season (in the Northern Hemisphere, from April to September), although some cases occur in winter. The mortality rate was 20–25% in the preantibiotic era and has been reported at ~3–5% in the postantibiotic era, principally because of delayed diagnosis and treatment. Recent reporting of a relatively low mortality rate (0.4%) for spotted fever rickettsiosis is likely an artifact related to the abundance of less pathogenic SFG rickettsial species likely causing subclinical or undiagnosed infection with cross-reactive anti-SFG antibodies and to a relatively low proportion of diagnostically confirmed cases. Indeed, the reported case–fatality rates in confirmed cases in the United States and in parts of Arizona, where *R. rickettsii* is the sole infecting SFG species, are 9% and 10%, respectively. The case–fatality rate is highest among children (<10 years of age) and in the later decades of life (>70 years). For unknown reasons, the case–fatality rate of RMSF in Mexico and Brazil approaches 50%.

Pathogenesis *R. rickettsii* organisms are inoculated into the dermis along with secretions of the tick's salivary glands after ≥6 h of feeding. The rickettsiae spread lymphohematogenously throughout the body and infect numerous foci of contiguous endothelial cells. The dose-dependent incubation period is ~1 week (range, 2–14 days). Occlusive thrombosis and ischemic necrosis are not the fundamental pathologic bases for tissue and organ injury. Instead, increased vascular permeability, with resulting edema, hypovolemia, and ischemia, is responsible. Consumption of platelets results in thrombocytopenia in 32–52% of patients, but disseminated intravascular coagulation (DIC) with hypofibrinogenemia is rare. Activation of platelets, generation of thrombin, and activation of the fibrinolytic system all appear to be homeostatic physiologic responses to endothelial injury by nonocclusive hemostatic plugs.

Clinical Manifestations Early in the illness, when medical attention usually is first sought, RMSF is difficult to distinguish from many self-limiting viral illnesses. Fever, headache, malaise, myalgia, nausea, vomiting, and anorexia are the most common symptoms during the first 3 days. The patient becomes progressively more ill as vascular infection and injury advance. In one large series, only one-third of patients were diagnosed with presumptive RMSF early in the clinical course and treated appropriately as outpatients. In the tertiary-care setting, RMSF is all too often recognized only when late severe manifestations, developing at the end of the first week or during the second week of illness in patients without appropriate treatment, prompt return to a physician or hospital and admission to an intensive care unit.

The progressive nature of the infection is clearly manifested in the skin. Rash is evident in only 14% of patients on the first day of illness and in only 49% during the first 3 days. Erythematous macules (1–5 mm) appear first on the wrists and ankles and then on the remainder of the extremities and the trunk. Later, more severe vascular damage results in frank hemorrhage at the center of the maculopapule, producing a petechia that does not disappear upon compression **(Fig. 192-1)**. This sequence of events is sometimes delayed or aborted by effective treatment. However, the rash is a variable manifestation, appearing on day 6 or later in 20% of cases and not appearing at all in ~10% of cases. Petechiae occur in 41–59% of cases, appearing on or after day 6 in 74% of cases that manifest a rash. Involvement of the palms and soles, often considered diagnostically important, usually develops relatively late in the course (after day 5 in 43% of cases) and does not develop at all in 18–64% of cases.

Hypovolemia leads to prerenal azotemia and (in 17% of cases) hypotension. Infection of the pulmonary microcirculation leads to noncardiogenic pulmonary edema; 12% of patients have acute respiratory distress syndrome, and 8% require mechanical ventilation. Cardiac involvement manifests as dysrhythmia in 7–16% of cases.

Besides respiratory failure, central nervous system (CNS) involvement is the other important determinant of the outcome of RMSF. Encephalitis, presenting as confusion or lethargy, is apparent in 26–28% of cases. Progressively severe encephalitis manifests as stupor or delirium in 21–26% of cases, ataxia in 18%, coma in 10%, and seizures in 8%. Numerous focal neurologic deficits have been reported.

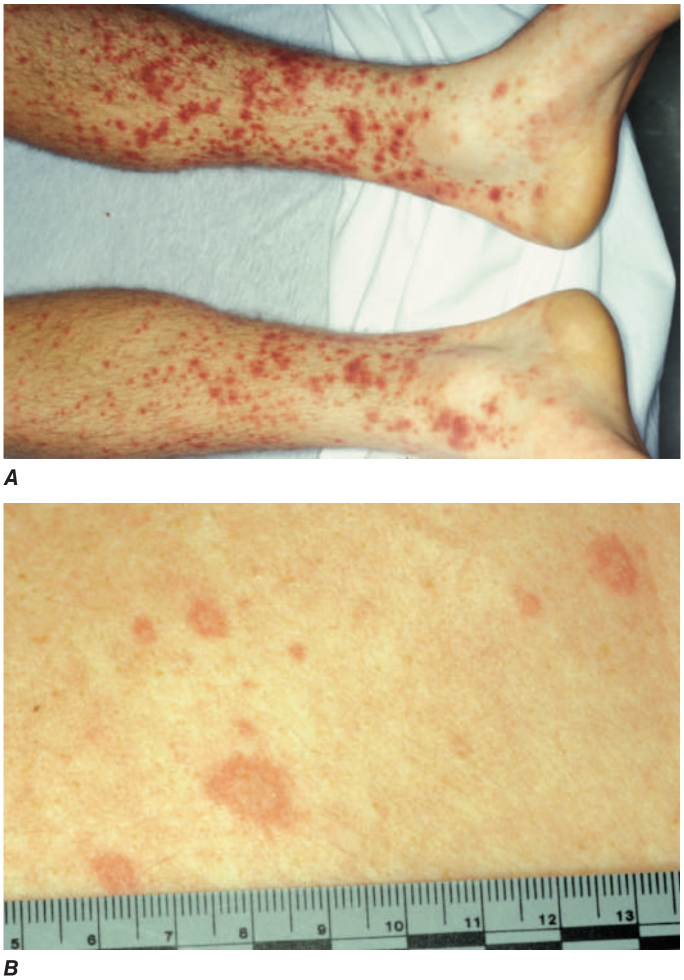

FIGURE 192-1 ***A.*** Petechial lesions of Rocky Mountain spotted fever on the lower legs and soles of a young, previously healthy patient. ***B.*** Close-up of lesions from the same patient. *(Photos courtesy of Dr. Lindsey Baden; with permission.)*

Meningoencephalitis results in cerebrospinal fluid (CSF) pleocytosis in 34–38% of cases; usually there are 10–100 cells/μL and a mononuclear predominance, but occasionally there are >100 cells/μL and a polymorphonuclear predominance. The CSF protein concentration is increased in 30–35% of cases, but the CSF glucose concentration is usually normal.

Acute kidney injury, often reversible with rehydration, is caused by acute tubular necrosis in severe cases with shock. Hepatic injury with increased serum aminotransferase concentrations (38% of cases) is due to multifocal death of individual hepatocytes without hepatic failure. Jaundice is recognized in 9% of cases and an elevated serum bilirubin concentration in 18–30%.

Life-threatening bleeding is rare. Anemia develops in 30% of cases and is severe enough to require transfusions in 11%. Blood is detected in the stool or vomitus of 10% of patients, and death has followed massive upper-gastrointestinal hemorrhage.

Other characteristic clinical laboratory findings include increased plasma levels of proteins of the acute-phase response (C-reactive protein, procalcitonin, fibrinogen, ferritin, and others), hypoalbuminemia, and hyponatremia (in 56% of cases) due to the appropriate secretion of antidiuretic hormone in response to the hypovolemic state. Myositis occurs occasionally, with marked elevations in serum creatine kinase levels and multifocal rhabdomyonecrosis. Ocular involvement includes conjunctivitis in 30% of cases and retinal vein engorgement, flame hemorrhages, arterial occlusion, and papilledema with normal CSF pressure in some instances.

Severe RMSF can present as sepsis or septic shock. In untreated fatal cases, death occurs 8–15 days after onset. A rare presentation, fulminant RMSF, is fatal within 5 days after onset. This fulminant presentation is seen most often in male black patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency and may be related to an undefined effect of hemolysis on the rickettsial infection. Although survivors of RMSF usually return to their previous state of health, permanent sequelae, including neurologic deficits and gangrene necessitating amputation of extremities, may follow severe illness.

Diagnosis The diagnosis of RMSF during the acute stage is more difficult than is generally appreciated. The most important epidemiologic factor is a history of exposure to a potentially tick-infested environment within the 14 days preceding disease onset during a season of possible tick activity. However, only 60% of patients actually recall being bitten by a tick during the incubation period.

The differential diagnosis for early clinical manifestations of RMSF (fever, headache, and myalgia without a rash) includes influenza, enteroviral infection, infectious mononucleosis, viral hepatitis, leptospirosis, typhoid fever, gram-negative or gram-positive bacterial sepsis, HME, HGA, murine typhus, sylvatic flying-squirrel typhus, and rickettsialpox. Enterocolitis may be suggested by nausea, vomiting, and abdominal pain; prominence of abdominal tenderness has resulted in exploratory laparotomy. CNS involvement can masquerade as bacterial or viral meningoencephalitis. Cough, pulmonary signs, and chest radiographic opacities can lead to a diagnostic consideration of bronchitis or pneumonia.

At presentation during the first 3 days of illness, only 3% of patients exhibit the classic triad of fever, rash, and history of tick exposure. When a rash appears, a diagnosis of RMSF should be considered. However, many illnesses considered in the differential diagnosis also can be associated with a rash, including rubeola, rubella, meningococcemia, disseminated gonococcal infection, secondary syphilis, toxic shock syndrome, drug hypersensitivity, immune thrombocytopenic purpura, thrombotic thrombocytopenic purpura, Kawasaki syndrome, and immune complex vasculitis. Conversely, any person in an endemic area with a provisional diagnosis of one of the above illnesses could have RMSF. Thus, if a viral infection is suspected during RMSF season in an endemic area, it should always be kept in mind that RMSF can mimic viral infection early in the course; if the illness worsens over the next couple of days after initial presentation, the patient should return for reevaluation.

The most common serologic test for confirmation of the diagnosis is the indirect immunofluorescence assay. Not until 7–10 days after onset is a reactive titer of ≥64 first detectable. The sensitivity and specificity of the indirect immunofluorescence IgG assay are 89–100% and 99–100%, respectively. Detection of IgM is no more sensitive in early illness and is subject to nonspecific cross-reactivity. It is important to understand that serologic tests for RMSF are usually negative at the time of presentation for medical care and that treatment should not be delayed while a positive serologic result is awaited. Because of significant seropositivity in the expanding geographic distribution of *Amblyomma americanum* ticks, a positive serologic titer in the first 5–6 days of illness represents preexisting antibody unrelated to the current illness.

The only diagnostic test that has proven useful during the acute illness is immunohistologic examination of a cutaneous biopsy sample from a rash lesion for *R. rickettsii*. Examination of a 3-mm punch biopsy from such a lesion is 70% sensitive and 100% specific, and polymerase chain reaction (PCR) on a rash biopsy would likely yield even higher sensitivity. PCR amplification for detection of *R. rickettsii* DNA in peripheral blood is not adequately sensitive. Although rickettsiae are present in large quantities in heavily infected foci of endothelial cells, there are relatively low quantities in the circulation. Cultivation of rickettsiae in cell culture is feasible but is seldom undertaken because of technical difficulty and biohazard concerns. The recent dramatic increase in the reported incidence of RMSF correlates with the use of single-titer SFG cross-reactive enzyme immunoassay serology. Few cases are specifically determined to be caused by *R. rickettsii*. Currently, many febrile persons who do not have RMSF present with cross-reactive antibodies, possibly because of previous exposure to the highly prevalent SFG rickettsia *R. amblyommatis*.

TREATMENT

Rocky Mountain Spotted Fever

The drug of choice for the treatment of both children and adults with RMSF is doxycycline. Because of the severity of RMSF, immediate empirical administration of doxycycline should be strongly considered for any patient with a consistent clinical presentation in the appropriate epidemiologic setting. Doxycycline is administered orally (or, with coma or vomiting, intravenously) at 100 mg twice daily. For children with suspected RMSF, up to five courses of doxycycline may be administered with minimal risk of dental staining. In patients with allergy to doxycycline, desensitization should be considered. Once considered an alternative during pregnancy, chloramphenicol is not readily available in the United States. Although available in much of the world, it is less effective than doxycycline. Fortunately, there is little evidence to support the occurrence of tetracycline-associated adverse events in mothers (hepatotoxicity) and fetuses (staining of deciduous teeth and teratogenicity) who receive doxycycline. The antirickettsial drug should be administered until the patient is afebrile and improving clinically—usually 3–5 days after defervescence. β-Lactam antibiotics, erythromycin, and aminoglycosides have no role in the treatment of RMSF, and sulfa-containing drugs are associated with more adverse outcomes than no treatment at all. There is little clinical experience with fluoroquinolones, clarithromycin, and azithromycin, which are not recommended. The most seriously ill patients are managed in intensive care units, with careful administration of fluids to achieve optimal tissue perfusion without precipitating noncardiogenic pulmonary edema. In some severely ill patients, hypoxemia requires intubation and mechanical ventilation; oliguric or anuric acute renal failure requires renal replacement therapy; seizures necessitate the use of antiseizure medication; anemia or severe hemorrhage necessitates transfusions of packed red blood cells; or bleeding with severe thrombocytopenia requires platelet transfusions.

Prevention Avoidance of tick bites is the only available preventive approach. Use of protective clothing and tick repellents, inspection of the body once or twice a day, and removal of ticks before they inoculate rickettsiae reduce the risk of infection. Prophylactic doxycycline treatment of tick bites has no proven role in preventing RMSF.

MEDITERRANEAN SPOTTED FEVER (BOUTONNEUSE FEVER), AFRICAN TICK-BITE FEVER, AND OTHER TICK-BORNE SPOTTED FEVERS

Epidemiology and Clinical Manifestations *R. conorii* is prevalent in southern Europe, Africa, and southwestern and south-central Asia. The disease is characterized by high fever, rash, and—in most geographic locales—an inoculation eschar (*tâche noire*) that appears before the onset of fever at the site of the tick bite. A severe form of the disease (mortality rate, 50%) occurs in patients with diabetes, alcoholism, or heart failure.

African tick-bite fever, caused by *R. africae*, occurs in rural areas of sub-Saharan Africa and in the Caribbean islands and is transmitted by *Amblyomma hebraeum* and *A. variegatum* ticks. The average incubation period is 4–10 days. The mild illness consists of headache, fever, eschar, and regional lymphadenopathy. *Amblyomma* ticks, a high portion of which are infected with *R. africae*, often feed in groups, with the consequent development of multiple eschars. Rash may be vesicular, sparse, or absent altogether. Because of tourism in sub-Saharan Africa, African tick-bite fever is the rickettsiosis most frequently imported into Europe and North America. Maculatum disease, a similar disease caused by the closely related species *R. parkeri*, is transmitted by *A. maculatum* and found in a low percentage of *A. americanum* ticks in the United States. It is also transmitted by *A. triste* in South America and Arizona as well as *A. tigrinum* and *A. ovale* in South America.

R. japonica causes Japanese spotted fever, which also occurs in Korea, Thailand, Laos, and China. Similar diseases in northern Asia are caused by *R. sibirica* and *R. heilongjiangensis*. Queensland tick typhus due to *R. australis* is transmitted by *Ixodes holocyclus* ticks. Flinders Island spotted fever, found on the island for which it is named as well as in Tasmania, mainland Australia, and Asia, is caused by *R. honei*. In Europe, patients infected with *R. slovaca* by a wintertime *Dermacentor* tick bite usually manifest an afebrile illness with an eschar (usually on the scalp) and painful regional lymphadenopathy; *R. raoultii* causes a similar syndrome in Europe and Asia.

Diagnosis Diagnosis of these tick-borne spotted fevers is based on clinical and epidemiologic findings and is confirmed by serology, immunohistochemical demonstration of rickettsiae in skin biopsy specimens, cell-culture isolation of rickettsiae, or PCR of skin biopsy, eschar biopsy or swab, or blood samples. Serologic diagnosis detects antibodies to antigens shared among SFG rickettsiae, hindering identification of the etiologic species. In an endemic area, a possible diagnosis of rickettsial spotted fevers should be considered when patients present with fever, rash, and/or a skin lesion consisting of a black necrotic lesion or a crust surrounded by erythema.

TREATMENT

Tick-Borne Spotted Fevers

As with RMSF, severe cases should be treated with doxycycline (100 mg bid orally) continued for 3–5 days after defervescence. Alternative agents for milder disease include doxycycline (100 mg bid orally for 1–5 days), chloramphenicol (500 mg qid orally for 7–10 days), ciprofloxacin (750 mg bid orally for 7 days), and josamycin (3 g/d orally for 5 days) where available. Data on the efficacy of treatment of mildly ill children with clarithromycin or azithromycin should not be extrapolated to adults or to patients with moderate or severe illness.

RICKETTSIALPOX

R. akari infects mice and their mites (*Liponyssoides sanguineus*), which maintain the organisms by transovarial transmission.

Epidemiology Rickettsialpox is recognized principally in New York City, but cases have also been reported in other urban and rural locations in the United States and in Ukraine, Croatia, Mexico, and Turkey. Investigation of eschars suspected of representing bioterrorism-associated cutaneous anthrax revealed that rickettsialpox occurs more frequently than previously realized.

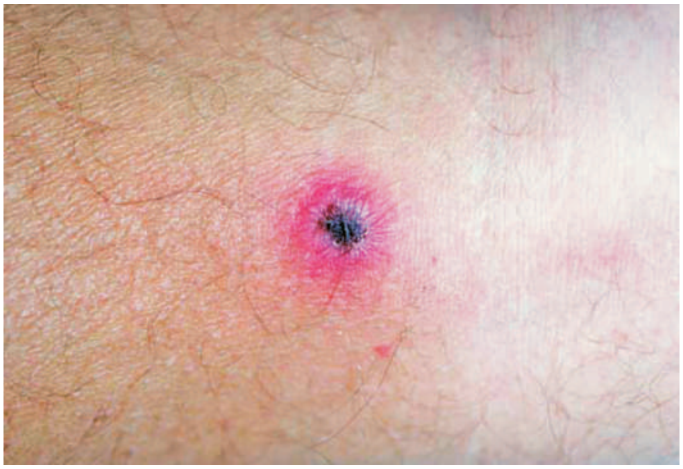

FIGURE 192-2 Eschar at the site of the mite bite in a patient with rickettsialpox. *(Reprinted from A Krusell et al: Emerg Infect Dis 8:727, 2002. Photo obtained by Dr. Kenneth Kaye.)*

Clinical Manifestations A papule forms at the site of the mite's feeding, develops a central vesicle, and becomes a 1- to 2.5-cm painless black crusted eschar surrounded by an erythematous halo **(Fig. 192-2)**. Enlargement of the regional lymph nodes draining the eschar suggests initial lymphogenous spread. After an incubation period of 10–17 days, during which the eschar and regional lymphadenopathy frequently go unnoticed, disease onset is marked by malaise, chills, fever, headache, and myalgia. A macular rash appears 2–6 days after onset and usually evolves sequentially into papules, vesicles, and crusts that heal without scarring **(Fig. 192-3)**; in some cases, the rash remains macular or maculopapular. Some patients develop nausea, vomiting, abdominal pain, cough, conjunctivitis, or photophobia. Without treatment, fever lasts 6–10 days.

Diagnosis and Treatment Clinical, epidemiologic, and convalescent serologic data establish the diagnosis of a SFG rickettsiosis that is seldom pursued further. Doxycycline is the drug of choice for treatment.

FLEA-BORNE SPOTTED FEVER

Rickettsia felis is suspected to cause an emerging rickettsiosis worldwide. Maintained transovarially in the geographically widespread cat flea, *Ctenocephalides felis*, the infection has been described as moderately severe, with fever, rash, and headache as well as CNS, gastrointestinal, and pulmonary symptoms on the basis of PCR, which often detects organisms in healthy persons. Patient isolates are lacking.

EPIDEMIC (LOUSE-BORNE) TYPHUS

Epidemiology The human body louse (*Pediculus humanus humanus*) lives in clothing under poor hygienic conditions and usually in impoverished cold areas. Lice acquire *R. prowazekii* when they ingest blood from a rickettsemic patient. The rickettsiae multiply in the louse's midgut epithelial cells and are shed in its feces. The infected louse leaves a febrile person and deposits infected feces on its subsequent host during its blood meal; the patient autoinoculates the organisms by scratching. The louse is killed by the rickettsiae and does not pass *R. prowazekii* to its offspring.

Epidemic typhus haunts regions afflicted by wars and disasters. An outbreak involved 100,000 people in refugee camps in Burundi in 1997. A small focus was documented in Russia in 1998, sporadic cases were reported from Algeria, and frequent outbreaks occurred in Peru, Rwanda and Burundi. Eastern flying squirrels (*Glaucomys volans*) and their lice and fleas maintain *R. prowazekii* in a zoonotic cycle and transmit infection to humans.

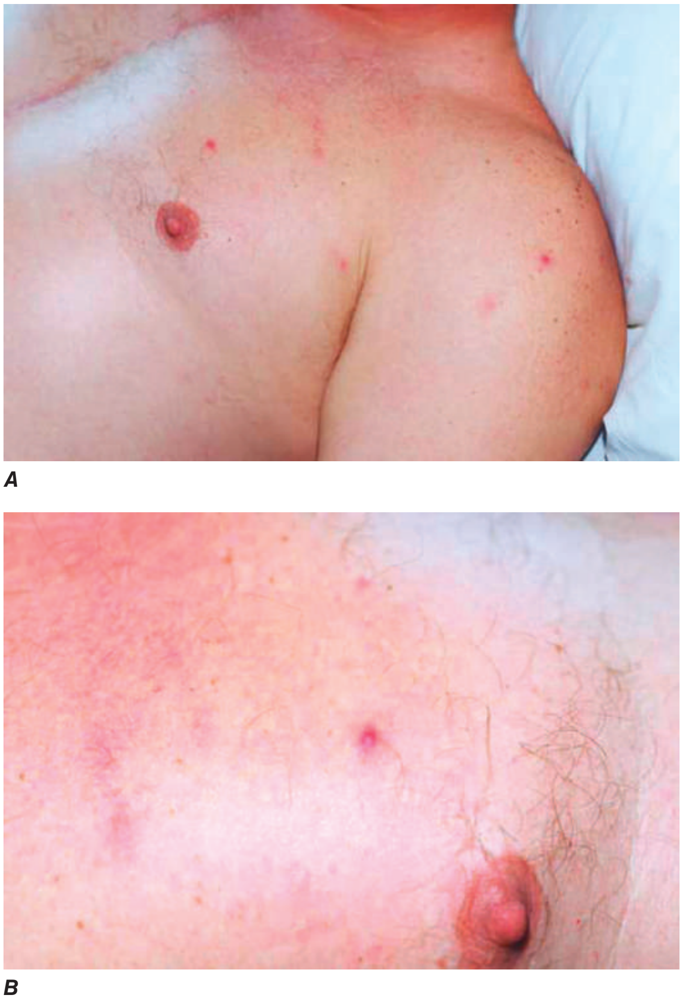

FIGURE 192-3 ***A***. Papulovesicular lesions on the trunk of the patient with rickettsialpox shown in Fig. 192-2. ***B***. Close-up of lesions from the same patient. *(Reprinted from A Krusell et al: Emerg Infect Dis 8:727, 2002. Photos obtained by Dr. Kenneth Kaye.)*

Brill-Zinsser disease is a recrudescent illness occurring years after acute epidemic typhus, probably as a result of waning immunity. *R. prowazekii* remains latent for years; its reactivation results in sporadic cases of disease in louse-free populations or in epidemics in louse-infested populations. Recrudescence has been documented after flying squirrel–associated typhus.

Rickettsiae are potential agents of bioterrorism **(Chap. S4)**. Infections with *R. prowazekii* and *R. rickettsii* have high case–fatality ratios. These organisms cause difficult-to-diagnose diseases and are highly infectious when inhaled as aerosols. Organisms resistant to tetracycline or chloramphenicol have been developed in the laboratory.

Clinical Manifestations After an incubation period of ~1–2 weeks, the onset of illness is abrupt, with prostration, severe headache, and fever rising rapidly to 38.8°–40.0°C (102°–104°F). Cough is prominent, developing in 70% of patients. Myalgias are usually severe. A rash begins on the upper trunk, usually on the fifth day, and then becomes generalized, involving the entire body except the face, palms, and soles. Initially, this rash is macular; without treatment, it becomes maculopapular, petechial, and confluent. The rash often goes undetected on black skin; 60% of African patients have spotless epidemic typhus. Photophobia, with considerable conjunctival injection, is common. The tongue may be dry, brown, and furred. Confusion and coma associated with meningoencephalitis are common. Skin necrosis and gangrene of the digits as well as interstitial pneumonia may occur in severe cases. Untreated disease is fatal in 7–40% of cases, with outcome depending primarily on the condition of the host. Patients with untreated infections develop renal insufficiency and multiorgan involvement in which neurologic manifestations are frequently prominent. Overall, 12% of patients with epidemic typhus have neurologic involvement. Infection associated with North American flying squirrels is a milder illness; whether this milder disease is due to host factors (e.g., better health status) or attenuated virulence is unknown.

Diagnosis and Treatment Epidemic typhus is sometimes misdiagnosed as typhoid fever in tropical countries **(Chap. 171)**. This means even serologic studies are often unavailable in settings of louse-borne typhus. Epidemics can be recognized by the serologic or immunohistochemical diagnosis of a single case or by detection of *R. prowazekii* in a louse found on a patient. Doxycycline (100 mg bid) is administered orally or—if the patient is comatose or vomiting—intravenously and continued until 3–5 days after defervescence. Under epidemic conditions, a single 200-mg oral dose can be tried but fails in some cases. Pregnant patients should be evaluated individually and treated with chloramphenicol early in pregnancy or with doxycycline late in pregnancy.

Prevention Prevention of epidemic typhus involves control of body lice. Clothes should regularly be changed and laundered in hot water, and insecticides can be used every 6 weeks to control the louse population.

■ ENDEMIC MURINE TYPHUS

Epidemiology *R. typhi* is maintained in mammalian host–flea cycles, with rats (*Rattus rattus* and *R. norvegicus*) and the Oriental rat flea (*Xenopsylla cheopis*) as the classic zoonotic niche. Fleas acquire *R. typhi* from rickettsemic rats and carry the organism throughout their life span. Nonimmune rats and humans are infected when rickettsia-laden flea feces contaminate pruritic bite lesions; less frequently, the flea bite transmits the organisms. Transmission can also occur via inhalation of aerosolized rickettsiae from flea feces. Infected rats appear healthy, although they are rickettsemic for ~2 weeks.

In the United States, murine typhus occurs mainly in Texas, where the disease is expanding northward, and in southern California, where it appears to affect people experiencing homelessness disproportionally. The incidence is steadily increasing in both of these states, where the classic rat–flea cycle is absent and an opossum–cat flea (*C. felis*) cycle is highly suspected. Globally, endemic typhus occurs mainly in warm (often coastal) areas throughout the tropics and subtropics, where it is highly prevalent though often unrecognized. The incidence peaks from April through July in Texas and during the warm months of summer and early fall in other geographic locations. Patients seldom recall exposure to fleas, although exposure to animals such as cats, opossums, and rats is reported in nearly 40% of cases.

Clinical Manifestations The incubation period of experimental murine typhus averages 11 days (range, 8–16 days). Headache, myalgia, arthralgia, nausea, and malaise develop 1–3 days before onset of chills and fever. Patients often experience nausea and vomiting.

The duration of untreated illness averages 12 days (range, 9–18 days). Rash occurs in approximately half of all patients. It is present in only 13% of patients at presentation for medical care (usually ~4 days after onset of fever), appearing an average of 2 days later in half of the remaining patients. The initial macular rash is often faint and detected by careful inspection of the axilla or the inner surface of the arm. Subsequently, the rash becomes maculopapular, involving the trunk more often than the extremities; it is seldom petechial and rarely involves the face, palms, or soles. A rash is detected in only 20% of patients with darkly pigmented skin.

Pulmonary involvement is frequently prominent; 35% of patients have a hacking, nonproductive cough, and 23% of patients who undergo chest radiography have pulmonary densities due to interstitial pneumonia, pulmonary edema, and pleural effusions. Bibasilar rales are the most common pulmonary sign. Less common clinical manifestations include abdominal pain, confusion, stupor, seizures, ataxia, coma, and jaundice. Clinical laboratory studies frequently reveal anemia and leukopenia early in the course, leukocytosis late in the course, thrombocytopenia, hyponatremia, hypoalbuminemia, increased serum levels of hepatic aminotransferases, and prerenal azotemia. Complications can include respiratory failure, hematemesis, cerebral hemorrhage, and hemolysis. Severe illness necessitates the admission of 10% of hospitalized patients to an intensive care unit. Greater severity is generally associated with old age, underlying disease, and treatment with a sulfonamide; the case–fatality rate is 0.4%.

Diagnosis and Treatment Serologic studies of acute- and convalescent-phase serum samples can provide a diagnosis, and an immunohistochemical method for identification of typhus group-specific antigens in biopsy samples has been developed. Cultivation is used infrequently and is not widely available. PCR of the blood is not adequately sensitive. When endemic typhus is suspected, patients should be treated empirically with doxycycline (100 mg twice daily by mouth for 7 days). Where available, chloramphenicol is an alternative. Ciprofloxacin and azithromycin are active in vitro, but clinically, they are less effective than doxycycline.

SCRUB TYPHUS

Epidemiology *O. tsutsugamushi* differs substantially from *Rickettsia* species both genetically and in cell-wall composition (i.e., it lacks lipopolysaccharide). *O. tsutsugamushi* is maintained by transovarial transmission in trombiculid mites. After hatching, infected larval mites (chiggers, the only stage that feeds on a host) inoculate organisms into the skin. Infected chiggers are particularly likely to be found in areas of heavy scrub vegetation during the wet season, when mites lay eggs.

Scrub typhus is endemic and reemerging in eastern and southern Asia, northern Australia, and islands of the western Pacific and Indian Oceans. Infections are prevalent in these regions; in some areas, >3% of the population is infected or reinfected each month. Immunity to the homologous strain wanes over 1–3 years, and the organisms exhibit remarkable antigenic diversity with loss of cross-protective immunity in as short a period as 1 month. Emerging cases in Chile, serologic evidence of infection in other parts of South and Central America and Africa, and the discovery of *Orientia chuto* infection in southwestern Asia challenge the classic epidemiology of scrub typhus.

Clinical Manifestations Illness varies from mild and self-limiting to fatal. After an incubation period of 6–21 days, onset is characterized by fever, headache, myalgia, cough, and gastrointestinal symptoms. Some patients recover spontaneously after a few days. The classic case description includes an eschar where the chigger has fed, regional lymphadenopathy, and a maculopapular rash—signs that are seldom seen in indigenous patients. In fact, <50% of returning travelers develop an eschar, and <40% develop a rash (on day 4–6 of illness). Severe cases typically manifest with encephalitis and interstitial pneumonia due to vascular injury. The case–fatality rate for untreated classic cases is 6% but would probably be lower if all mild cases were diagnosed.

Diagnosis and Treatment Serologic assays (indirect fluorescent antibody, indirect immunoperoxidase, enzyme immunoassays, and rapid diagnostic assays using lateral flow chromatographic tests) are the mainstays of laboratory diagnosis. PCR amplification of *Orientia* genes from eschars is effective, but less so for blood. Patients are treated with oral doxycycline (100 mg twice daily for 7–15 days), azithromycin (500 mg for 3 days), or chloramphenicol (500 mg four times daily for 7–15 days). In severe scrub typhus, the combination of doxycycline and azithromycin is beneficial compared to monotherapy of either agent.

Some cases of scrub typhus in Thailand are poorly responsive to doxycycline or chloramphenicol but respond to azithromycin and rifampin.

EHRLICHIOSES AND ANAPLASMOSES

Ehrlichioses are acute febrile infections caused by members of the family Anaplasmataceae, which is made up of obligately intracellular organisms of more than five genera, including those with established human pathogens: *Ehrlichia*, *Anaplasma*, *Wolbachia*, "*Candidatus* Neoehrlichia," and *Neorickettsia*. The bacteria reside in vertebrate reservoirs and target vacuoles of hematopoietic and, for some species, endothelial cells **(Fig. 192-4)**. Four *Ehrlichia* species, two *Anaplasma* species, and one *Neoehrlichia* species are transmitted by ticks to humans and cause infection that can be severe and prevalent. *E. chaffeensis*, the agent of HME, and *E. muris* subsp. *eauclairensis* infect predominantly mononuclear phagocytes; *E. ewingii* and *A. phagocytophilum* infect neutrophils. Infections with "*Candidatus* Neoehrlichia mikurensis," *A. capra*, and *A. bovis* are less well characterized but have been reported to grow in endothelium, erythrocytes, and monocytes, respectively.

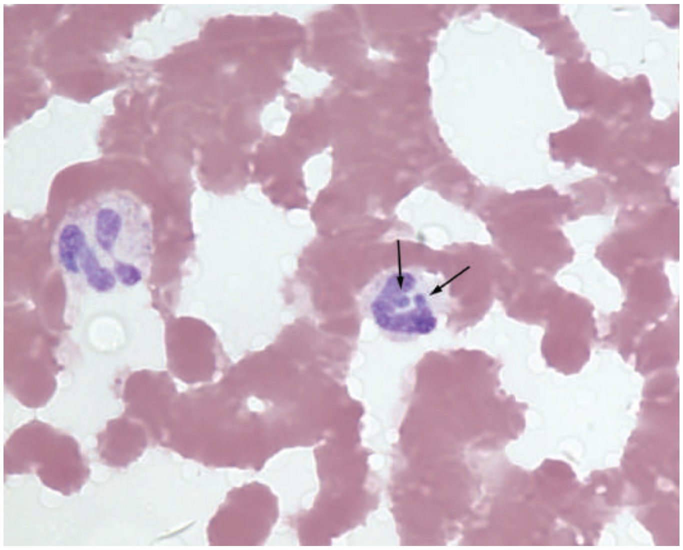

FIGURE 192-4 Peripheral-blood smear from a patient with human granulocytotropic anaplasmosis. A neutrophil contains two morulae (vacuoles filled with *Anaplasma phagocytophilum*). *(Photo courtesy of Dr. J. Stephen Dumler.)*

Ehrlichia, "*Candidatus* Neoehrlichia," and *Anaplasma* are maintained by horizontal tick–mammal–tick transmission, and humans are only inadvertently infected. Wolbachiae are associated with human filariasis, since they are important for filarial viability and pathogenicity; antibiotic treatment targeting wolbachiae is a strategy for filariasis control. Neorickettsiae parasitize flukes (trematodes) that in turn parasitize aquatic snails, fish, and insects. Only a single pathogen of human neorickettsiosis has been described: *N. sennetsu* causes sennetsu fever, an infectious mononucleosis–like illness first identified in 1953 in association with the ingestion of raw fish containing *N. sennetsu*–infected flukes.

HUMAN MONOCYTOTROPIC EHRLICHIOSIS

Epidemiology More than 25,480 cases of *E. chaffeensis* infection had been reported to the U.S. Centers for Disease Control and Prevention (CDC) as of November 2023. However, active prospective surveillance documented an incidence as high as 414 cases per 100,000 population in some U.S. regions. Most *E. chaffeensis* infections are identified in the south-central, southeastern, and mid-Atlantic states, but cases have also been recognized in California, New York, New England, and midwestern states. All stages of the Lone Star tick (*A. americanum*), which is expanding its geographic range, feed on white-tailed deer—a major reservoir. Dogs and coyotes also serve as reservoirs and often lack clinical signs. Tick bites and exposures are frequently reported by patients in rural areas, and 64% of infections occur in May through July. Transfusion-related infections are documented, as are infections in patients with organ transplants. The incidence of HME is highest in patients 60–84 years of age; however, 9% of infections occur in children ≤19 years of age, and these include severe and fatal infections. Of patients with HME, 59% are male.

E. chaffeensis has been detected in South and Central America, Africa, and Asia.

Clinical Manifestations *E. chaffeensis* disseminates hematogenously from the dermal blood pool created by the feeding tick. After a median incubation period of 8 days, illness develops. Clinical manifestations are undifferentiated and include fever (97% of cases), headache (70%), myalgia (68%), and malaise (77%). Less frequently observed are nausea, vomiting, and diarrhea (28–57%); cough (30%); rash (29% overall, 6% at presentation); and confusion (20%). HME can be severe:

77% of patients with confirmed cases are hospitalized, and 2% die. Life-threatening complications include renal failure, meningoencephalitis, acute respiratory distress syndrome, a DIC-like syndrome, pneumonia, septic shock, cardiac failure, hepatitis, hemorrhage, hemophagocytic lymphohistiocytosis and—in immunocompromised patients—overwhelming ehrlichial infection; patients with diabetes, cancer, organ transplantation, asplenia, hepatitis C, or HIV infection have a 2.3 relative risk for death. Laboratory findings are valuable in the differential diagnosis of HME; 66% of patients have leukopenia (initially lymphopenia, later neutropenia), 86% have thrombocytopenia, and 89% have elevated serum levels of hepatic aminotransferases. Despite low blood cell counts, the bone marrow is hypercellular, and noncaseating granulomas can be present. Vasculitis is not a component of HME.

Diagnosis HME can be fatal. If not given empirical doxycycline treatment, 39% and 40% of patients with HME require admission to an intensive care unit and mechanical ventilation, respectively; these measures are not necessary in patients receiving prompt empirical treatment. In addition, hospital stay and illness duration are lengthened in untreated patients by 8 and 12 days, respectively. The diagnosis is suggested by fever, known tick exposure in the preceding 3 weeks, thrombocytopenia and/or leukopenia, and increased serum aminotransferase activities. Morulae are demonstrated in <10% of peripheral-blood smears. HME can be confirmed during active infection by PCR amplification of *E. chaffeensis* nucleic acids in blood obtained before the start of doxycycline therapy. Retrospective serodiagnosis requires a consistent clinical picture and a fourfold increase in *E. chaffeensis* antibody titer to ≥128 in paired serum samples obtained ~3 weeks apart. Separate specific diagnostic tests are necessary for HME and HGA (see below).

■ EWINGII EHRLICHIOSIS AND *EHRLICHIA MURIS EAUCLAIRENSIS* INFECTIONS

Ehrlichia ewingii resembles *E. chaffeensis* in its tick vector (*A. americanum*) and vertebrate reservoirs (white-tailed deer and dogs). *E. muris eauclairensis* causes human infections after *Ixodes scapularis* tick exposure in Wisconsin and Minnesota. *E. ewingii* and *E. muris* illnesses are similar to but less severe than HME. Many cases occur in immunocompromised patients. Human infections with *E. canis* have been documented as subclinical ehrlichemia. No specific serologic diagnostic tests for these other ehrlichiae are readily available, and *E. chaffeensis* serologic tests can be positive when the infecting agent is actually a different species of *Ehrlichia*.

■ "*CANDIDATUS* NEOEHRLICHIA MIKURENSIS" INFECTION

"*Candidatus* Neoehrlichia mikurensis," a bacterium in a phylogenetic clade between *Ehrlichia* and *Anaplasma*, was originally identified in *Ixodes ricinus* ticks from the Netherlands and in mice and *Ixodes ovatus* ticks from Japan. By means of broad-range 16S rRNA gene amplification and sequence analysis, this organism was identified as the cause of severe and sometimes prolonged febrile illnesses in European immunocompromised patients with tick bites or exposures and in Chinese patients developing a mild febrile illness after being bitten by *Ixodes persulcatus* and *Haemaphysalis concinna* ticks. The clinical presentation is similar to those of HME and HGA except for an often prolonged course and a predilection for venous thrombosis. Specific diagnostic methods have been developed but are not widely available.

TREATMENT

Ehrlichioses

Doxycycline is effective for HME as well as other ehrlichioses; the use of this drug in "*Candidatus* N. mikurensis" infection is associated with disease resolution. Therapy with doxycycline (100 mg given PO or IV twice daily) or tetracycline (250–500 mg given PO every 6 h) lowers hospitalization rates and shortens fever duration. *E. chaffeensis* is not susceptible to chloramphenicol in vitro, and the use of this drug is controversial. While a few reports document *E. chaffeensis* persistence in humans, this finding is rare; most infections are cured by short courses of doxycycline continuing for 3–5 days after defervescence. Although poorly studied for this indication, rifampin may be suitable when doxycycline is contraindicated.

Prevention HME, *E. ewingii* ehrlichiosis, *E. muris* ehrlichiosis, and "*Candidatus* N. mikurensis" infection can be prevented by the avoidance of ticks in endemic areas. The use of protective clothing and tick repellents, careful postexposure tick searches, and prompt removal of attached ticks probably diminish infection risk.

■ HUMAN GRANULOCYTOTROPIC ANAPLASMOSIS

Epidemiology As of November 2023, 64,225 cases of HGA had been reported to the CDC, most in the upper Midwest and northeastern United States. The global geographic distribution is similar to that of Lyme disease because of the shared *Ixodes* tick vectors. Natural reservoirs for *A. phagocytophilum* are small mammals, such as white-footed mice and squirrels, and large animals such as white-tailed deer in the United States and red deer in Europe. HGA incidence peaks in May through July, but the disease can occur throughout the year with exposure to *Ixodes* ticks. HGA often affects males (59%), and the incidence of infection in persons older than 60 years is 7 times higher than in those younger than 15 years.

Clinical Manifestations Seroprevalence rates are high in endemic regions; thus, it seems likely that most individuals develop subclinical infections. The incubation period for HGA is 4–8 days, after which the disease manifests as fever (75–100% of cases), myalgia (73%), headache (82%), and malaise (97%). A minority of patients develop nausea, vomiting, or diarrhea (22–40%); cough (27%); or confusion (17%). A rash in HGA (5%) most often reflects co-infection with *Borrelia*, resulting in erythema migrans. Most patients develop thrombocytopenia (80%) and/or leukopenia (63%) with increased serum hepatic aminotransferase levels (80%).

Life-threatening complications occur most often in the elderly and include renal failure, acute respiratory distress syndrome, a toxic shock–like syndrome, pneumonia, a DIC- or sepsis-like syndrome and hemophagocytic lymphohistiocytosis. Meningoencephalitis is rare in documented cases of HGA. Other documented neurologic sequelae include brachial plexopathy, cranial nerve involvement, and demyelinating polyneuropathy. Infection of patients with a preexisting immunocompromising condition (diabetes, immunosuppressive medications, asplenia, arthritis) is associated with a 3.0 relative risk for life-threatening complications. Of patients with HGA, 31% are hospitalized, and 7% require intensive care. The case–fatality rate is 0.3%, but the relative risk for death is 16 if infection occurs with an immunosuppressive condition. Neither vasculitis nor granulomas are components of HGA. While patients can be co-infected with *Borrelia burgdorferi* and *Babesia microti* (transmitted by the same tick vector[s]), there is little evidence that these infections increase the severity or persistence of HGA. HGA transmitted by transfusion (including the transfusion of leukoreduced blood or platelets) has now been reported in at least nine cases, including a fatality.

Diagnosis HGA should be included in the differential diagnosis of influenza-like illnesses during seasons with *Ixodes* tick activity (May through December), especially in the context of a known tick bite or exposure. Concurrent thrombocytopenia, leukopenia, or elevated serum levels of alanine or aspartate aminotransferase further increase the likelihood of HGA. Many HGA patients develop Lyme disease antibodies in the absence of clinical findings consistent with that diagnosis. Thus, HGA should be considered in the differential diagnosis of atypical severe Lyme disease presentations. Peripheral-blood film examination for neutrophil morulae can yield a diagnosis in 20–75% of infections. PCR testing of blood from patients with active disease before doxycycline therapy is sensitive and specific. Serodiagnosis is retrospective, requiring a fourfold increase in *A. phagocytophilum* antibody titer (to ≥128) in paired serum samples obtained 1 month apart. Since seroprevalence is high in some regions, a single acute-phase titer should not be used for diagnosis.

Anaplasma capra and A. bovis Infection Human infection by *A. capra*, first isolated from goat blood, was identified in 28 patients from northeastern China. Patients presented with fever, headache, malaise, dizziness, myalgias, and chills, but these manifestations were less severe than in HGA. Hospitalization was recorded for 18% of patients, and 14% had underlying disorders, including hyperglycemia, hypertension, coronary heart disease, diabetes, and cancer. Five patients had severe manifestations, including one with encephalitic signs and *A. capra* DNA present in CSF. *A. capra* is found most often in *Haemaphysalis longicornis* ticks, and at least seven additional species can carry it over a distribution across Asia and eastern Europe. All patients responded to doxycycline treatment and survived.

Anaplasma bovis was recently identified in four human patients in a surveillance study of suspected Lyme borreliosis, HGA, HME, and babesiosis in the United States and in two patients with fever, myalgias, and rash in China. All six patients were between the ages of 35 and 71 years; four were male, and two were female. The apparent tick vectors are *Dermacentor variabilis* in the United States and *Haemaphysalis longicornis* in China. Both Chinese patients recovered after antimicrobial treatment.

TREATMENT

Human Granulocytotropic Anaplasmosis

No prospective studies of therapy for HGA have been conducted. However, doxycycline (100 mg PO twice daily) is effective. Rifampin therapy is associated with improvement of HGA in pregnant women and children. Most treated patients defervesce within 24–48 h.

Prevention HGA prevention requires tick avoidance. Transmission can be documented as few as 4 h after a tick bite.

Q FEVER

The agent of Q fever is *C. burnetii*, a small pleomorphic coccobacillus with a gram-negative cell wall, that was first isolated in 1935 and called a rickettsia due to its presence in ticks, intracellular replication, small size, and staining characteristics, but it is now known to be genetically quite distinct from Rickettsiaceae and to have a number of unique features. It survives in harsh environments, escapes intracellular killing in macrophages by inhibiting the final step in phagosome maturation, and has adapted to the acidic phagolysosome.

Epidemiology Q fever is a zoonosis: transmission of *C. burnetii* to humans typically occurs by inhalation after it has been shed by animals. The primary sources of human infection are infected sheep, goats, and cattle. At parturition, when large amounts of *C. burnetii* are present in the fetus, placenta, membranes, and fluids, the bacterium readily contaminates the environment. Smaller amounts can be shed in milk for weeks to months after parturition, but also in urine and feces. Once shed, *C. burnetii* can remain viable in manure, hay, soil, etc., for many years, after which it can be aerosolized and inhaled, even after traveling miles from the source by wind. A variety of other vertebrate animals can be hosts of *C. burnetii*, including birds, cats, dogs, rabbits, skunks, raccoons, deer, bears, sloths, kangaroos, and marine animals. *C. burnetii* has also been found in several tick species, which could be important for maintenance of the agent in veterinary populations, but the majority of human Q fever cases are associated with aerosol transmission from infected livestock. Infections in animals are usually asymptomatic, but abortions and stillbirth have been observed in pregnant goats and sheep. Because it is easily dispersed as an aerosol and because of the extremely low infectious dose required for human infection (probably between 1 and 10 viable bacteria), *C. burnetii* is a potential agent of bioterrorism **(Chap. S4)**, with a high infectivity rate and pneumonia as the major manifestation.

Persons most at risk for Q fever include abattoir workers, veterinarians, farmers, and other individuals who have contact with infected animals (particularly newborn animals), but outbreaks often affect people without any livestock contact. In Canada and the Netherlands, 65% and 72%, respectively, of persons living and/or working on dairy cattle farms were seropositive, and in the United States, 22% of veterinarians were seropositive, compared to ~3% of the population overall. Seroprevalence data confirm that Q fever is underreported in many countries. An outbreak of Q fever associated with ingestion of raw milk confirmed the oral route of transmission, although this route is uncommon. In rare instances, person-to-person transmission follows childbirth in an infected woman, surgery, autopsy, blood transfusion, or organ transplantation of an infected individual. Multiple outbreaks involving laboratory staff have been reported in the past. Some evidence suggests that *C. burnetii* can be sexually transmitted among humans. Some unusual modes of *C. burnetii* transmission to humans include treatment with live fetal sheep cells and percutaneous infection after crushing an infected tick between the fingers.

Infections due to *C. burnetii* occur in most geographic locations except New Zealand and Antarctica. Several factors influence the epidemiology: environmental conditions such as high concentrations of animals, high animal pregnancy rates, dry weather, and the strength and direction of winds. In addition to differences between strains of *C. burnetii*, the variability in human susceptibility to *C. burnetii* can influence transmission and development of disease. Some people become sick after exposure, whereas others have only mild symptoms that are not sufficient to lead them to seek medical assistance, and ~60% have asymptomatic seroconversion. Q fever continues to be endemic in Australia and France. In Cayenne, French Guiana, Q fever is hyperendemic: 40% of all community-acquired pneumonias are caused by *C. burnetii*. The largest known outbreak occurred between 2007 and 2010 in the Netherlands. Over 4000 cases were reported, and over 40,000 people were infected. The outbreak was due to a combination of high-density goat farming in areas with large urban populations and environmental factors. Farms where spread did not occur had high vegetation density and lower groundwater concentrations.

Young age seems to be protective against disease caused by *C. burnetii*. In a large outbreak in Switzerland, symptomatic infection occurred five times more often among persons >15 years of age than among younger individuals. In many outbreaks, men are affected more commonly than women.

Clinical Manifestations • **ACUTE Q FEVER** The incubation period is 3–30 days. The primary manifestations of acute Q fever differ geographically. During the Dutch outbreak, but also in Canada and Croatia, pneumonia is the more common presentation. In some countries where Q fever is endemic, such as France and Israel, hepatitis is more common. These differences could reflect the route of infection (i.e., ingestion of contaminated milk for hepatitis and inhalation of contaminated aerosols for pneumonia) or strain differences. In the Dutch outbreak, sequelae of infection in pregnant women were rare; this was not the case among pregnant women elsewhere. Pericarditis, myocarditis, acalculous cholecystitis, pancreatitis, lymphadenitis, spontaneous rupture of the spleen, transient hypoplastic anemia, hemolytic anemia, hemophagocytic lymphohistiocytosis, meningitis, encephalitis, optic neuritis, and erythema nodosum are less common manifestations.

The symptoms of acute Q fever are nonspecific; common among them are fever, extreme fatigue, photophobia, and severe headache that is frequently retro-orbital. Other symptoms include chills, sweats, nausea, vomiting, and diarrhea. Cough develops in about half of patients with Q fever pneumonia. A nonspecific rash may be evident in 4–18% of patients. The WBC count is usually normal. Thrombocytopenia occurs in ~25% of patients, and reactive thrombocytosis frequently develops during recovery. Biochemical markers of autoimmunity, such as anticytoplasmic antibodies (ANCA), antinuclear antibodies (ANA), anti–smooth muscle antibodies, or antiphospholipid antibodies, are often present in acute Q fever. Chest radiography can show opacities similar to those seen in pneumonia caused by other pathogens. Case-fatality rate of acute Q fever is ~1–2%.

Acute Q fever occasionally complicates pregnancy. In one series, it resulted in premature birth in 35% of cases and in abortion or neonatal death in 43%. Neonatal death and lower infant birth weight are reported up to three times more often among women seropositive for *C. burnetii* in some areas but not others.

Q FEVER FATIGUE SYNDROME Prolonged fatigue follows acute Q fever in up to 20% of cases and can be accompanied by a constellation of symptoms, including headaches, sweats, arthralgia, and myalgias. Several hypotheses regarding the etiology exist, including host and genetic factors, cytokine dysregulation, and a partly biopsychological etiology with *C. burnetii* acting as a trigger for fatigue development. A randomized controlled trial including 155 patients with strictly diagnosed Q fever fatigue syndrome showed that long-term treatment with doxycycline did not reduce fatigue severity compared to placebo, so antibiotics should not be prescribed for these patients. Cognitive behavioral therapy aimed at fatigue-related cognitions and behaviors thought to perpetuate symptoms was effective in reducing fatigue severity in the short term. The beneficial effect of this treatment, however, was not maintained after 1 year.

CHRONIC Q FEVER Following symptomatic or asymptomatic primary infection, 1–5% of all patients develop chronic Q fever. Chronic Q fever endocarditis, infected aneurysms, or infected vascular prostheses are most frequently observed. The primary infection often has not been recognized or was asymptomatic, and the median duration between primary infection and manifestation of chronic infection is 13 months, with the largest observed interval between acute infection and diagnosis of chronic Q fever being >9 years. Risk factors for the development of chronic Q fever include male sex, older age, valvulopathy or prior valve surgery, aneurysms, vascular prostheses, renal insufficiency, and immunocompromised state. Diagnosing chronic Q fever is difficult, as patients are sometimes asymptomatic until complications occur and often present with nonspecific symptoms, such as fever, night sweats, weight loss, fatigue, and malaise. Fever can be absent and is frequently low grade. C-reactive protein is often low or even normal. Chronic Q fever endocarditis differs from endocarditis caused by other bacteria, manifesting as endothelium-covered nodules on the valves, aortic root abscess, or new or rapidly increasing valve insufficiency. A high index of suspicion is necessary for timely diagnosis. Patients with chronic Q fever are often ill for >1 year before the diagnosis is made. The disease should be suspected in all patients with culture-negative endocarditis. In addition, all patients with valvular heart disease, an aneurysm or vascular prosthesis and unexplained weight loss, fever, elevated inflammatory markers, stroke, unexpected aneurysm growth, and/or progressive heart failure should be tested for *C. burnetii* infection. Other manifestations of chronic Q fever include lymphadenitis and bone infection including vertebral osteomyelitis and prosthetic joint infection. Of 313 patients with proven chronic Q fever in the Netherlands, 65% developed complications. The most frequently observed complications were acute aneurysms, heart failure, and noncardiac abscesses. One in six patients with vascular chronic Q fever develops arterial fistula, including aortoenteric fistula, aortocaval fistula, aortobronchial fistula, and arteriocutaneous fistula. PCR positivity at any time during the disease, presence of prosthetic material, and older age were associated with complications. Q fever–related mortality was 25% in patients diagnosed with proven chronic Q fever after the Dutch outbreak. Chronic Q fever–related mortality was highest in patients with both endocarditis and vascular infection (33%), followed by patients with vascular infection only (25%), and was lowest in endocarditis patients (12%).

Diagnosis Culture of *C. burnetii* from buffy-coat blood samples or tissue specimens is possible but requires a biosafety level 3 laboratory and is not used in clinical practice. PCR detects *C. burnetii* DNA in blood and tissue specimens, including paraffin-embedded samples. The detection of antibodies to *C. burnetii* is the most commonly used method for the diagnosis of Q fever. Available serologic assays are indirect fluorescent antibody (IFA) assay, enzyme-linked immunosorbent assay (ELISA), and complement fixation test (CFT), with IFA being the gold standard. IFA tests are useful for the detection of and discrimination between acute and chronic infection and have excellent sensitivity and specificity. The diagnosis of acute Q fever is dependent on seroconversion, defined as a fourfold increase in IgG titer for phase II antigens between acute- and convalescent-phase samples. In the first 1–2 weeks of illness, PCR on blood or serum can also be positive.

TABLE 192-2 Diagnostic Criteria for Chronic Q Fever as Defined by the Dutch Q Fever Consensus Group

PROVEN CHRONIC Q FEVER	PROBABLE CHRONIC Q FEVER	POSSIBLE CHRONIC Q FEVER
1. Positive *Coxiella burnetii* PCR in blood or tissue[a] *or* 2. IFA ≥1:800 or 1:1024 for *C. burnetii* phase I IgG *and* Definite endocarditis according to the modified Duke criteria *or* Proven large vessel or prosthetic infection by imaging studies (^{18}FDG-PET, CT, MRI, or AUS)	IFA ≥1:800 or 1:1024 for *C. burnetii* phase I IgG *AND AT LEAST ONE OF THE FOLLOWING:* Valvulopathy not meeting the major criteria of the modified Duke criteria Known aneurysm and/or vascular or cardiac valve prosthesis without signs of infection by means of TEE/TTE, ^{18}FDG-PET, CT, MRI, or AUS Suspected osteomyelitis or hepatitis as manifestation of chronic Q fever Pregnancy Symptoms and signs of chronic infection, such as fever, weight loss, night sweats, hepatosplenomegaly, and persistently raised ESR and CRP Granulomatous tissue inflammation, proven by histologic examination Immunocompromised state	IFA ≥1:800 or 1:1024 for *C. burnetii* phase I IgG *without* manifestations meeting the criteria for proven or probable chronic Q fever

[a] In absence of acute infection.

Abbreviations: AUS, abdominal ultrasound; CRP, C-reactive protein; CT, computed tomography; ^{18}FDG-PET, fluorodeoxyglucose positron emission tomography; ESR, erythrocyte sedimentation rate; IFA, indirect fluorescent antibody assay; MRI, magnetic resonance imaging; PCR, polymerase chain reaction; TEE, transesophageal echocardiography; TTE, transthoracic echocardiography.

A high phase I IgG titer (e.g., >512) is suggestive of chronic Q fever, but alone, it is not enough for a definite diagnosis. A positive PCR for *C. burnetii* in blood or tissue in the absence of an acute infection confirms the diagnosis, but PCR on blood is negative in the majority of patients and tissue samples are often very difficult to obtain. The diagnosis of chronic Q fever should be based on a combination of clinical, laboratory, and imaging criteria. There has been debate on the optimal set of criteria, but the Dutch literature-based consensus guideline **(Table 192-2)** appears to be the most sensitive and is easy to use.

Valvular vegetations are detected in only 12% of patients with Q fever endocarditis by transthoracic echocardiography, but the rate of detection is higher (21–50%) with transesophageal echocardiography. Fluorodeoxyglucose positron emission tomography combined with computed tomography (FDG-PET/CT) can detect not only prosthetic valvular infection but also intravascular infection elsewhere, osteomyelitis, and lymphadenitis. In native valve endocarditis, specificity is very high but sensitivity is low, so a normal FDG-PET/CT scan cannot exclude native valve endocarditis. A study including 273 FDG-PET/CT scans performed at diagnosis in patients suspected of chronic Q fever showed that, even after serology, PCR, and often ultrasound or CT had been performed, FDG-PET/CT led to a change in diagnosis or treatment in 20% of patients. Adding FDG uptake in a heart valve as a major criterion to the Duke criteria led to a 1.9-fold increase of definite endocarditis diagnoses. Of 218 scans performed during follow-up, 57% resulted in treatment adjustment. In case of suspected chronic Q fever, FDG-PET/CT should be considered.

TREATMENT

Q Fever

ANTIBIOTICS

In vitro, *C. burnetii* is susceptible to doxycycline, quinolones, trimethoprim-sulfamethoxazole (TMP-SMX), macrolides, and rifampin. Although antimicrobial susceptibility testing is not routinely performed and resistance to doxycycline does not appear to be a common problem in clinical practice, doxycycline-resistant isolates do exist.

Treatment of acute Q fever with doxycycline (100 mg twice daily for 14 days) is usually successful. Quinolones also are effective. When Q fever is diagnosed during pregnancy, treatment with TMP-SMX is recommended for the duration of the pregnancy.

Treatment with doxycycline and hydroxychloroquine for 6–12 months following acute infection should be considered in patients with valve abnormalities, a prosthetic heart valve, an aneurysm, or vascular prosthesis. This appeared to be effective in preventing progression to chronic Q fever in patients with valvulopathy. The exact indications and duration of prophylaxis should be based on a careful consideration of possible benefits and side effects.

Decisions on treatment of chronic Q fever are challenging, so consultation with an infectious diseases expert is recommended. There is no indication for antibiotic therapy in those with possible chronic Q fever (only elevated phase I IgG without symptoms or an infectious focus). Addition of hydroxychloroquine (to alkalinize the phagolysosome) renders doxycycline bactericidal against *C. burnetii* in vitro, and the combination of doxycycline 100 mg twice daily with 200 mg hydroxychloroquine three times daily is the favored regimen. It is advised to determine serum levels of doxycycline aiming for concentrations between 5 and 10 mg/L, often requiring higher doses than 200 mg per day. Patients treated with this regimen must be advised about photosensitivity, but side effects should not lead to cessation of doxycycline too easily since it appears to be the most effective approach for this serious infection that has a high mortality despite treatment. Patients treated with hydroxychloroquine are at risk for developing retinopathy, so they should be evaluated by an ophthalmologist before starting treatment and every 6–12 months during the course of therapy. If doxycycline-hydroxychloroquine cannot be used, the regimen chosen should include at least two antibiotics active against *C. burnetii*. In a study including 322 patients with chronic Q fever, treatment with doxycycline combined with a quinolone appeared to be a safe alternative.

Minimum treatment duration is 18 months after PCR on blood had become negative (if positive before) and adequate doxycycline levels have been reached for native valve endocarditis and other manifestations without prosthetic material and 24 months for patients with prosthetic valve endocarditis or infected vascular prostheses. Many patients with vascular infection need prolonged treatment before the infection resolves, and surgical intervention is often necessary to remove an infected graft if the patient does not respond to antibiotic therapy. Abscesses need drainage for antibiotic therapy to be successful. Defining cure of chronic Q fever after the minimum treatment duration should be based on a combination of imaging (if abnormal at diagnosis), decline of serologic titers, negativity of PCR on blood or serum, and improvement of symptoms.

FOLLOW-UP

After acute Q fever, patients without risk factors for developing chronic Q fever should be evaluated clinically and serologically after 6 months. When IgG phase I is <1024 and clinical symptoms do not suggest chronic infection, follow-up can be stopped. For patients with a very high risk of developing chronic Q fever who have received antibiotics for 6–12 months, patients with immunosuppression or other risk factors not treated with antibiotics for a prolonged period of time, or patients with possible chronic Q fever (only phase I IgG ≥1024), follow-up with serology and PCR every 3–6 months for 2 years is recommended. During treatment of chronic Q fever, patients should be followed every 3 months to evaluate symptoms, side effects, serology, and PCR. When new complications are suspected, imaging should be repeated. After the end of treatment, relapse has been described up to 5 years later. It is therefore recommended to continue monitoring with serology and PCR until a minimum of 5 years after end of treatment.

Prevention A whole-cell vaccine (Q-Vax) licensed in Australia effectively prevents Q fever in abattoir workers. Vaccine is given only to people without a history of Q fever and negative results in both serologic and skin testing that is performed with intradermal diluted *C. burnetii* vaccine to prevent side effects. Cases among abattoir workers in Australia declined dramatically as a result of a vaccination program, but the vaccine has not been approved outside Australia.

Good animal-husbandry practices are important in preventing widespread contamination of the environment by *C. burnetii*. These practices include isolating aborting animals for up to 14 days, raising feed bunks to prevent contamination of feed by excreta, destroying aborted materials (by burning and burying fetal membranes and stillborn animals), and wearing masks and gloves when handling aborted materials. Vaccination of sheep and goats and a culling program were effective in the Dutch outbreak.

During an outbreak of Q fever and for 4 weeks after it ceases, blood donations should not be accepted from individuals who live in the affected area.

ACKNOWLEDGMENT

The authors thank Thomas Marrie, MD, for his significant contributions to this chapter in the previous editions.

FURTHER READING

Biggs HM et al: Diagnosis and management of tickborne rickettsial diseases: Rocky Mountain spotted fever and other spotted fever group rickettsioses, ehrlichioses, and anaplasmosis—United States. MMWR 65:1, 2016.

Buijs SB et al: Still new chronic Q fever cases diagnosed 8 years after a large Q fever outbreak. Clin Infect Dis 73:1476, 2021.

Gillespie J, Salje J: *Orientia* and *Rickettsia*: Different flowers from the same garden. Curr Opin Microbiol 74:102318, 2023.

Ismail N, Mcbride JW: Tick-borne emerging infections: Ehrlichiosis and anaplasmosis. Clin Lab Med 37:317, 2017.

Varghese GM et al: Intravenous doxycycline, azithromycin, or both for severe scrub typhus. N Engl J Med 388:792, 2023.

193

Infections Due to Mycoplasmas

R. Doug Hardy

Mycoplasmas are prokaryotes of the class Mollicutes. Their size (150–350 nm) is closer to that of viruses than to that of typical bacteria. Unlike viruses, however, mycoplasmas grow in cell-free culture media; in fact, they are the smallest organisms capable of independent replication. The entire genomes of many *Mycoplasma* species have been sequenced and have been found to be among the smallest of all prokaryotic genomes. Sequencing information for these genomes has helped define the minimal set of genes necessary for cellular life. The absence of genes related to the synthesis of amino acids, fatty acid metabolism, and cholesterol dictates the mycoplasmas' parasitic or saprophytic dependence on a host for exogenous nutrients and necessitates the use of complex fastidious media to culture these organisms. Mycoplasmas lack a cell wall and are bound only by a cell membrane. The absence of a cell wall explains the inactivity of β-lactam antibiotics (penicillins and cephalosporins) against infections caused by these organisms.

At least 13 *Mycoplasma* species, 2 *Acholeplasma* species, and 2 *Ureaplasma* species have been isolated from humans. Most of these species are thought to be normal inhabitants of oral and urogenital mucous membranes. *M. pneumoniae*, *M. hominis*, *M. genitalium*, *U. urealyticum*, and *U. parvum* have been shown conclusively to be pathogenic in immunocompetent humans. *M. pneumoniae* primarily infects the respiratory tract, while *M. hominis*, *M. genitalium*, *U. urealyticum*, and *U. parvum* are associated with a variety of genitourinary tract disorders and neonatal infections. Other mycoplasmas may cause disease in immunocompromised persons.

MYCOPLASMA PNEUMONIAE

PATHOGENESIS

M. pneumoniae is generally thought to act as an extracellular pathogen. Although the organism has been shown to exist and replicate within human cells, it is not known whether these intracellular events contribute to the pathogenesis of disease. *M. pneumoniae* attaches to ciliated respiratory epithelial cells by means of a complex terminal organelle at the tip of one end of the organism. Cytoadherence is mediated by interactive adhesins and accessory proteins clustered on this organelle. After extracellular attachment, *M. pneumoniae* causes injury to host respiratory tissue. The mechanism of injury is thought to be mediated by the production of hydrogen peroxide and of an ADP-ribosylating and vacuolating cytotoxin of *M. pneumoniae* that has many similarities to pertussis toxin. Because mycoplasmas lack a cell wall, they also lack cell wall–derived stimulators of the innate immune system, such as lipopolysaccharide, lipoteichoic acid, and murein (peptidoglycan) fragments. However, lipoproteins from the mycoplasmal cell membrane appear to have inflammatory properties, probably acting through Toll-like receptors (primarily TLR2) on macrophages and other cells. Lung biopsy specimens from patients with *M. pneumoniae* respiratory tract infection reveal an inflammatory process involving the trachea, bronchioles, and peribronchial tissue, with a monocytic infiltrate that coincides with a luminal exudate of polymorphonuclear leukocytes.

Experimental evidence indicates that innate immunity provides most of the host's defense against mycoplasmal infection in the lungs, whereas cellular immunity may actually play an immunopathogenic role, exacerbating mycoplasmal lung disease. Humoral immunity appears to provide protection against dissemination of *M. pneumoniae* infection; patients with humoral immunodeficiencies do not have more severe lung disease than do immunocompetent patients in the early stages of infection but more often develop disseminated infection resulting in syndromes such as arthritis, meningitis, and osteomyelitis. The immunity that follows severe *M. pneumoniae* infections is more protective and longer-lasting than that following mild infections. Genuine second attacks of *M. pneumoniae* pneumonia have been reported infrequently.

EPIDEMIOLOGY

M. pneumoniae infection occurs worldwide. It is likely that the incidence of upper respiratory illness due to *M. pneumoniae* is up to 20 times that of pneumonia caused by this organism. Infection is spread from one person to another by respiratory droplets expectorated during coughing and results in clinically apparent disease in an estimated 80% of cases. The incubation period for *M. pneumoniae* is 2–4 weeks; therefore, the time-course of infection in a specific population may be several weeks long. Intrafamilial attack rates are as high as 84% among children and 41% among adults. Outbreaks of *M. pneumoniae* illness often occur in institutional settings such as military bases, boarding schools, and summer camps. Infections tend to be endemic, with sporadic epidemics every 4–7 years.

Most significantly, *M. pneumoniae* is a major cause of community-acquired respiratory illness in both children and adults and is often grouped with *Chlamydia pneumoniae* and *Legionella* species as one of the most important bacterial causes of "atypical" community-acquired pneumonia. For community-acquired pneumonia in adults, *M. pneumoniae* is the most frequently detected "atypical" organism. Analysis of 13 studies of community-acquired pneumonia published between 1996 and 2001 (which included 6207 ambulatory and hospitalized adults) showed that the overall prevalence of *M. pneumoniae* was 22.7%; by comparison, the prevalence of *C. pneumoniae* was 11.7%, and that of *Legionella* species was 4.6%. The summation of 26 more recent investigations of "atypical" organisms in community-acquired pneumonia in adults published between 2002 and 2015 found the overall prevalence of *M. pneumoniae* was 7.2%; by comparison, the prevalence of *C. pneumoniae* was 4.3%, and that of *Legionella* species was 2.8%. *M. pneumoniae* pneumonia is also referred to as Eaton agent pneumonia (the organism having first been isolated in the early 1940s by Monroe Eaton), primary atypical pneumonia, and "walking" pneumonia.

CLINICAL MANIFESTATIONS

Upper Respiratory Tract Infections and Pneumonia Acute *M. pneumoniae* infections generally manifest as pharyngitis, tracheobronchitis, reactive airway disease/wheezing, or a nonspecific upper respiratory syndrome. Little evidence supports the commonly held belief that this organism is an important cause of otitis media, with or without bullous myringitis. Pneumonia develops in 3–13% of infected individuals; its onset is usually gradual, occurring over several days, but may be more abrupt. Although *Mycoplasma* pneumonia may begin with a sore throat, the most common presenting symptom is cough. The cough is typically nonproductive, but some patients produce sputum. Headache, malaise, chills, and fever are noted in the majority of patients.

On physical examination, wheezes or rales are detected in ~80% of patients with *M. pneumoniae* pneumonia. In many patients, however, pneumonia can be diagnosed only by chest radiography. The most common radiographic pattern is that of peribronchial pneumonia with thickened bronchial markings, streaks of interstitial infiltration, and areas of subsegmental atelectasis. Segmental or lobar consolidation is not uncommon. While clinically evident pleural effusions are infrequent, lateral decubitus views reveal that up to 20% of patients have pleural effusions.

Overall, the clinical presentation of pneumonia in an individual patient is not useful for differentiating *M. pneumoniae* pneumonia from other types of community-acquired pneumonia. The possibility of *M. pneumoniae* infection deserves particular consideration when community-acquired pneumonia fails to respond to treatment with a penicillin or a cephalosporin—antibiotics that are ineffective against mycoplasmas. Symptoms usually resolve within 2–3 weeks after the onset of illness. Although *M. pneumoniae* pneumonia is generally self-limited, appropriate antimicrobial therapy significantly shortens the duration of clinical illness. Infection uncommonly results in critical illness and only rarely in death. In some patients, long-term recurrent wheezing or reactive airway disease may follow the resolution of acute pneumonia. The significance of chronic infection, especially as it relates to asthma, is an area of active investigation.

Extrapulmonary Manifestations An array of extrapulmonary manifestations may develop during *M. pneumoniae* infection. The most significant are neurologic, dermatologic, cardiac, rheumatologic, and hematologic in nature. Extrapulmonary manifestations can be a result of disseminated infection, especially in patients with humoral immunodeficiencies (e.g., septic arthritis); postinfectious autoimmune phenomena (e.g., Guillain-Barré syndrome); or possibly ADP-ribosylating toxin. Overall, these manifestations are uncommon, given the frequency of *M. pneumoniae* infection. Notably, many patients with extrapulmonary *M. pneumoniae* disease do not have respiratory disease.

Skin eruptions described with *M. pneumoniae* infection include erythematous (macular or maculopapular), vesicular, bullous, petechial, and urticarial rashes. In some reports, 17% of patients with *M. pneumoniae* pneumonia have had an exanthem. Erythema multiforme major (Stevens-Johnson syndrome) is the most clinically significant skin eruption associated with *M. pneumoniae* infection; it appears to occur more commonly with *M. pneumoniae* than with other infectious agents.

A wide spectrum of neurologic manifestations has been reported with *M. pneumoniae* infection. The most common are meningoencephalitis, encephalitis, Guillain-Barré syndrome, and aseptic meningitis. *M. pneumoniae* has been implicated as a likely etiologic agent in 5–7% of cases of encephalitis. Other neurologic manifestations may include cranial neuropathy, acute psychosis, cerebellar ataxia, acute demyelinating encephalomyelitis, cerebrovascular thromboembolic events, and transverse myelitis.

TABLE 193-1 Diagnostic Tests for Respiratory *Mycoplasma pneumoniae* Infection[a]

TEST	SENSITIVITY, %	SPECIFICITY, %
Respiratory culture	≤60	100
Respiratory PCR	65–90	90–100
Serologic studies[b]	55–100	55–100

[a] A combination of PCR and serology is suggested for routine diagnosis. If macrolide resistance is suspected, resistance testing by culture and/or PCR is available.
[b] Acute- and convalescent-phase serum samples are recommended.

Abbreviation: PCR, polymerase chain reaction.

TABLE 193-2 Antimicrobial Agents of Choice for *Mycoplasma* Infections[a]

ORGANISM(S)	DRUGS
Mycoplasma pneumoniae	Azithromycin, clarithromycin, erythromycin, doxycycline, levofloxacin, moxifloxacin, gemifloxacin (*not* ciprofloxacin or ofloxacin)
Ureaplasma urealyticum, Ureaplasma parvum	Azithromycin, clarithromycin, erythromycin, doxycycline
Mycoplasma hominis	Doxycycline, clindamycin
Mycoplasma genitalium	Azithromycin, moxifloxacin, doxycycline

[a] Antimicrobial resistance has been reported in mycoplasmas, as described in the text.

Hematologic manifestations of *M. pneumoniae* infection include hemolytic anemia, aplastic anemia, cold agglutinins, disseminated intravascular coagulation, and hypercoagulopathy. When anemia does occur, it generally develops in the second or third week of illness.

In addition, hepatitis, glomerulonephritis, pancreatitis, myocarditis, pericarditis, rhabdomyolysis, and arthritis (septic and reactive) have been convincingly ascribed to *M. pneumoniae* infection. Septic arthritis has been described most commonly in hypogammaglobulinemic patients.

DIAGNOSIS

Clinical findings, nonmicrobiologic laboratory tests, and chest radiography are not useful for differentiating *M. pneumoniae* pneumonia from other types of community-acquired pneumonia. In addition, since *M. pneumoniae* lacks a cell wall, it is not visible on Gram stain. Although of historical interest, the measurement of cold agglutinin titers is no longer recommended for the diagnosis of *M. pneumoniae* infection because the findings are nonspecific and assays specific for *M. pneumoniae* are now available.

Acute *M. pneumoniae* infection can be diagnosed by polymerase chain reaction (PCR) detection of the organism in respiratory tract secretions or by isolation of the organism in culture **(Table 193-1)**. Oropharyngeal, nasopharyngeal, and pulmonary specimens are all acceptable for diagnosing *M. pneumoniae* pneumonia. Other bodily fluids, such as cerebrospinal fluid, are acceptable for extrapulmonary infection. *M. pneumoniae* culture (which requires special media) is not recommended for routine diagnosis because the organism may take weeks to grow and is often difficult to isolate from clinical specimens. In contrast, PCR allows rapid, specific diagnosis earlier in the course of clinical illness.

The diagnosis can also be established by serologic tests for IgM and IgG antibodies to *M. pneumoniae* in paired (acute- and convalescent-phase) serum samples; enzyme-linked immunoassay is the recommended serologic method. An acute-phase sample alone is not adequate for diagnosis, as antibodies to *M. pneumoniae* may not develop until 2 weeks into the illness; therefore, it is important to test paired samples. In addition, IgM antibody to *M. pneumoniae* can persist for up to 1 year after acute infection. Thus, its presence may indicate recent rather than acute infection.

The combination of PCR of respiratory tract secretions and serologic testing constitutes the most sensitive and rapid approach to the diagnosis of *M. pneumoniae* infection.

TREATMENT

Mycoplasma pneumoniae Infections

Although in the majority of untreated cases symptoms resolve within 2–3 weeks without significant associated morbidity, *M. pneumoniae* pneumonia can be a serious illness that responds to appropriate antimicrobial therapy **(Table 193-2)**. Randomized, double-blind, placebo-controlled trials in adults have demonstrated that antimicrobial treatment significantly decreases the duration of fever, cough, malaise, hospitalization, and radiologic abnormalities in *M. pneumoniae* pneumonia. Treatment options for acute *M. pneumoniae* infection include macrolides (e.g., oral azithromycin, 500 mg on day 1, then 250 mg/d on days 2–5), tetracyclines (e.g., oral doxycycline, 100 mg twice daily for 7–14 days), and respiratory fluoroquinolones. However, ciprofloxacin and ofloxacin are *not* recommended because of their high minimal inhibitory concentrations against *M. pneumoniae* isolates and their poor performance in experimental studies. A 7- to 14-day course of quinolone therapy appears adequate. Even though appropriate antibiotic therapy significantly reduces the duration of respiratory illness, it does not appear to shorten the duration of detection of *M. pneumoniae* by culture or PCR; therefore, a test of cure for eradication is not suggested.

In Asian countries, a high prevalence (range 34–76%) of *M. pneumoniae* resistant to macrolides has been reported. In Europe and in the United States, macrolide-resistant *M. pneumoniae* is less common. In the United States, national surveillance from 2018 found that 10.2% of isolates demonstrated macrolide resistance. Furthermore, national surveillance from 2015–2018 found macrolide resistance of 15.2–21.7% in the eastern United States and 1.9–2.8% in the western United States. Clinical studies have demonstrated that, when treated with macrolides, patients with community-acquired pneumonia due to macrolide-resistant *M. pneumoniae* experience a significantly longer duration of symptoms than do patients infected with macrolide-sensitive organisms; thus, macrolide resistance in *M. pneumoniae* does appear to have clinical significance. In addition, clinical investigations have indicated that for macrolide-refractory *M. pneumoniae* pneumonia, tetracycline class therapy results in shorter duration of fever and hospital length of stay compared with macrolide therapy. If macrolide resistance is prominent in a particular geographic locale or is suspected, then a nonmacrolide antibiotic should be considered for treatment; in addition, in these instances, a respiratory sample may be sent to a *Mycoplasma* reference laboratory for the detection of macrolide resistance by culture or PCR.

While the 2019 Infectious Diseases Society of America and American Thoracic Society guidelines do not recommend routinely using corticosteroids in community-acquired pneumonia, growing clinical literature suggests that the addition of glucocorticoids to an antibiotic regimen may be of value for the treatment of severe or refractory *M. pneumoniae* pneumonia. A 2019 meta-analysis of 24 randomized controlled trials in children found that use of corticosteroids in macrolide-refractory *M. pneumoniae* pneumonia significantly reduced hospital days and duration of fever. Clinical literature in adults also shows benefit, but these data are limited and more observational in nature.

The roles of antimicrobial drugs, glucocorticoids, and IV immunoglobulin in the treatment of neurologic disease due to *M. pneumoniae* remain unknown.

UROGENITAL MYCOPLASMAS

EPIDEMIOLOGY

M. hominis, *M. genitalium*, *U. urealyticum*, and *U. parvum* can cause urogenital tract disease. The significance of isolation of these organisms in a variety of other syndromes is unknown and in some cases is being investigated. *M. fermentans* has not been shown convincingly to cause human disease.

While urogenital mycoplasmas may be transmitted to a fetus during passage through a colonized birth canal, sexual contact is the major mode of transmission, and the risk of colonization increases dramatically with increasing numbers of sexual partners. In asymptomatic women, these mycoplasmas may be found throughout the lower urogenital tract. The vagina yields the largest number of organisms; next most densely colonized are the periurethral area and the cervix. Ureaplasmas are isolated less often from urine than from the cervix, but *M. hominis* is found with approximately the same frequency at these two sites. Ureaplasmas are isolated from the vagina of 40–80% of sexually active, asymptomatic women and *M. hominis* from 21–70%. The two microorganisms are found concurrently in 31–60% of women. In men, colonization with each organism is less prevalent. A large U.S. prospective multicenter study of a nucleic acid amplification test for *M. genitalium* including male and female patients seeking care in diverse geographic regions found overall prevalence to be 10.3%. Mycoplasmas have been isolated from urine, semen, and the distal urethra of asymptomatic men.

CLINICAL MANIFESTATIONS

Urethritis, Pyelonephritis, and Urinary Calculi In many episodes of *Chlamydia*-negative nongonococcal urethritis, ureaplasmas may be the causative agent. These organisms may also cause chronic voiding symptoms in women. The common presence of ureaplasmas in the urethra of asymptomatic men may suggest either that only certain serovars are pathogenic or that predisposing factors, such as lack of immunity, must exist in persons who develop symptomatic infection. Alternatively, disease may develop only upon initial exposure to ureaplasmas. Ureaplasmas have been implicated in epididymitis. *M. genitalium* also appears to cause urethritis. *M. genitalium* and ureaplasmas do not have a known role in prostatitis. *M. hominis* does not appear to play a primary etiologic role in urethritis, epididymitis, or prostatitis.

Evidence suggests that *M. hominis* causes up to 5% of cases of acute pyelonephritis. Ureaplasmas have not been associated with this disease.

Ureaplasmas play a limited role in the production of urinary calculi. The frequency with which ureaplasmas reach the kidney, the predisposing factors that allow them to do so, and the relative frequency of urinary tract calculi induced by this organism (compared with other organisms) are not known.

Pelvic Inflammatory Disease *M. hominis* can cause pelvic inflammatory disease. In most episodes, *M. hominis* occurs as part of a polymicrobial infection, but the organism may play an independent role in a limited number of cases. Data also support an association of *M. genitalium* with pelvic inflammatory disease. Ureaplasmas are not thought to cause pelvic inflammatory disease.

Postpartum and Postabortal Infection Studies implicate *M. hominis* as the primary pathogen in ~5–10% of women who have postpartum or postabortal fever; ureaplasmas have been implicated to a lesser degree. These infections are generally self-limited; however, if symptoms persist, specific antimicrobial therapy should be given. Ureaplasmas also appear to play a role in occasional postcesarean wound infections.

Nonurogenital Infection In rare instances, *M. hominis* causes nonurogenital infections, such as brain abscess, wound infection, poststernotomy mediastinitis, endocarditis, and neonatal meningitis. These infections are most common among immunocompromised and hypogammaglobulinemic patients. Ureaplasmas and *M. hominis* can cause septic arthritis in immunodeficient patients. Ureaplasmas probably cause neonatal pneumonitis; their possible causal role in the development of bronchopulmonary dysplasia—the chronic lung disease of premature infants—has been extensively investigated, with most studies indicating at least a significant association. It is unclear whether ureaplasmas and *M. hominis* cause infertility, spontaneous abortion, premature labor, low birth weight, or chorioamnionitis.

DIAGNOSIS

Culture and PCR are both appropriate methods for the isolation of urogenital mycoplasmas. Culture of these organisms, however, requires special techniques and media that generally are available only at larger medical centers and reference laboratories. Serologic testing is not recommended for the clinical diagnosis of urogenital *Mycoplasma* infections. Antibiotic resistance testing to guide appropriate therapy is becoming more available to clinical practice.

TREATMENT

Urogenital *Mycoplasma* Infections

Because colonization with urogenital mycoplasmas is common, it appears at present that their isolation from the urogenital tract in the absence of disease generally does not warrant treatment.

Patients with recurrent urethritis or cervicitis after treatment failure of appropriate antibiotic therapy for sexually transmitted infections should receive *M. genitalium* testing. It is not recommended to screen for *M. genitalium* in those who are asymptomatic. However, sex partners of patients with symptomatic *M. genitalium* infection should receive testing. Those with positive test results should receive treatment with antibiotics to possibly reduce the risk for reinfection.

Macrolides and doxycycline are considered the antimicrobial agents of choice for *Ureaplasma* infections (Table 193-2). *Ureaplasma* resistance to macrolides, doxycycline, quinolones, and chloramphenicol has been reported. *M. hominis* is resistant to macrolides. Doxycycline is generally the drug of choice for *M. hominis* infections, although resistance has been reported. Clindamycin is generally active against *M. hominis*. Quinolones are active in vitro against *M. hominis*.

For *M. genitalium*, treatment failure and/or antibiotic resistance has increased for both azithromycin and moxifloxacin. The Centers for Disease Control and Prevention's 2021 Sexually Transmitted Infections Treatment Guidelines recommend two-stage therapy based on antibiotic resistance testing, if available: doxycycline 100 mg orally 2 times per day for 7 days, followed by either azithromycin 1 g orally initial dose, followed by 500 mg orally once daily for 3 additional days (if macrolide sensitive), or moxifloxacin 400 mg orally once daily for 7 days (if macrolide resistant). If *M. genitalium* antibiotic resistance testing is not available, then doxycycline followed by moxifloxacin is recommended. Test of cure for *M. genitalium* infection is not recommended for asymptomatic persons who received treatment with a recommended regimen.

FURTHER READING

Waites KB et al: *Mycoplasma pneumoniae* from the respiratory tract and beyond. Clin Microbiol Rev 30:747, 2017.

Waites KB et al: Macrolide-resistant *Mycoplasma pneumoniae* in the United States as determined from a national surveillance program. J Clin Microbiol 57:e00968, 2019.

Wang G et al: Global prevalence of resistance to macrolides in *Mycoplasma pneumoniae*: A systematic review and meta-analysis. J Antimicrob Chemother 77:2353, 2022.

Workowski KA et al: Sexually transmitted diseases treatment guidelines, 2021. MMWR Recomm Rep 70:1, 2021.

194

Chlamydial Infections

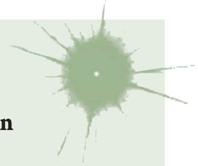

Matthew M. Hamill, Thomas C. Quinn

Chlamydiae are obligate intracellular bacteria that cause a wide variety of diseases in humans and nonhuman animals.

ETIOLOGIC AGENTS

The chlamydiae were originally classified as four species in the genus *Chlamydia*: *C. trachomatis, C. pneumoniae*, *C. psittaci*, and *C. pecorum* (the last species being found in ruminants). The *C. psittaci* group has been separated into three species: *C. psittaci*, *C. felis*, and *C. abortus*. The mouse pneumonitis strain (MoPn) is now classified as

C. muridarum, and the guinea pig inclusion conjunctivitis strain (GPIC) is now designated *C. caviae*.

C. trachomatis is divided into two biovars: trachoma and LGV (lymphogranuloma venereum). The trachoma biovar causes two major types of disease in humans: ocular trachoma, the leading infectious cause of preventable blindness in the developing world; and urogenital infections, which are predominantly sexually transmitted but can be neonatally transmitted. The 18 serovars of *C. trachomatis* fall into three groups: the trachoma serovars A, B, Ba, and C; the oculogenital serovars D–K; and the LGV serovars L_1–L_3. Serovars can be distinguished by serologic typing with monoclonal antibodies or by molecular gene typing. However, serovar identification usually is not important clinically, since the antibiotic susceptibility pattern is the same for all three groups. The one exception applies when LGV is suspected on clinical grounds; in this situation, serovar determination is important because a longer treatment duration is typically required for LGV strains.

BIOLOGY, GROWTH CYCLE, AND PATHOGENESIS

BIOLOGY

During their intracellular growth, chlamydiae produce characteristic intracytoplasmic inclusions that can be visualized by direct fluorescent antibody or Giemsa staining of infected clinical material, such as conjunctival scrapings or cervical or urethral epithelial cells. Chlamydiae are nonmotile, gram-negative, obligate intracellular bacteria that replicate within the cytoplasm of host cells, forming the characteristic membrane-bound inclusions that are the basis for some diagnostic tests. Originally considered to be large viruses, chlamydiae differ from viruses in possessing RNA and DNA as well as a cell wall that is quite similar in structure to the cell wall of typical gram-negative bacteria. However, chlamydiae lack peptidoglycan; their structural integrity depends on disulfide binding of outer-membrane proteins.

GROWTH CYCLE

Among the defining characteristics of chlamydiae is a unique growth cycle that involves alternation between two highly specialized morphologic forms **(Figs. 194-1 and 194-2)**: the elementary body, which is the infectious form and is specifically adapted for extracellular survival, and the metabolically active and replicating reticulate body, which is not infectious, is adapted for an intracellular environment, and does not survive well outside the host cell. The biphasic growth cycle begins with attachment of the elementary body (diameter, 0.25–0.35 μm) at specific sites on the surface of the host cell. The elementary body enters the cell through a process similar to receptor-mediated endocytosis and resides in an inclusion, where the entire growth cycle is completed. The chlamydiae prevent phagosome–lysosome fusion. The inclusion membrane is modified by insertion of chlamydial antigens. Once the elementary body has entered the cell, it reorganizes into a reticulate body, which is larger (0.5–1 μm) and contains more RNA. After ~8 h, the reticulate body starts to divide by binary fission. The intracytoplasmic, membrane-bound inclusion body containing the reticulate bodies increases in size as the reticulate bodies multiply. Approximately 18–24 h after infection of the cell, these reticulate bodies begin to become elementary bodies by a reorganization or condensation process that is poorly understood. After rupture of the inclusion body, the elementary bodies are released to initiate another cycle of infection.

Chlamydiae are susceptible to many broad-spectrum antibiotics and possess a number of enzymes, but they have a very restricted metabolic capacity. None of these metabolic reactions result in the production of energy. Chlamydiae have thus been considered to be energy parasites that use the ATP produced by the host cell for their own metabolic functions. Many aspects of chlamydial molecular biology are not well understood, but the sequencing of several chlamydial genomes and new proteomics research have provided researchers with many relevant tools for elucidating the biology of the life cycle.

PATHOGENESIS

Genital infections are primarily caused by *C. trachomatis* serovars D–K, with serovars D, E, and F involved most frequently. Molecular

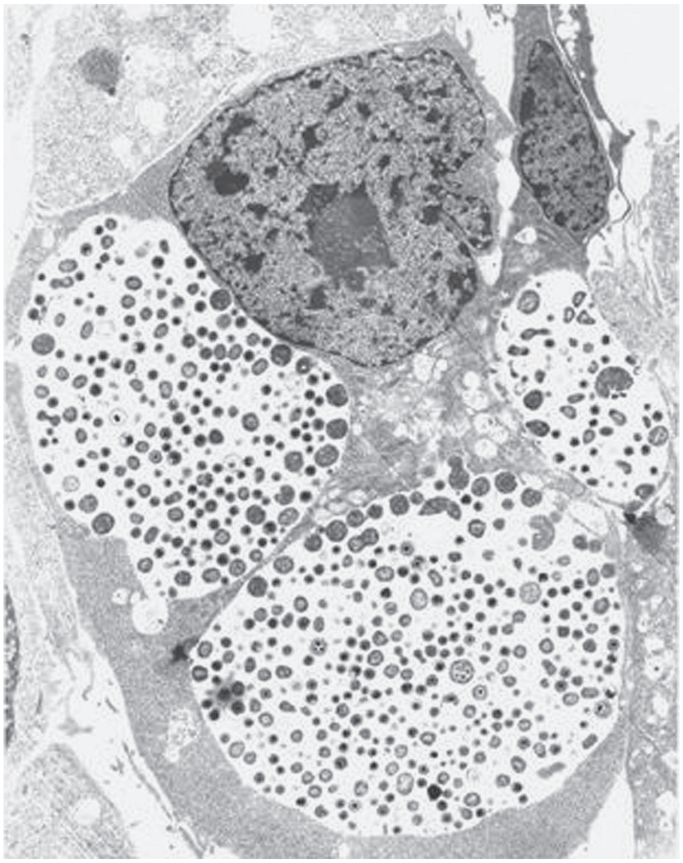

FIGURE 194-1 Chlamydial intracellular inclusions filled with smaller dense elementary bodies and larger reticulate bodies. *(Reprinted with permission from WE Stamm: Chlamydial infections, in Harrison's Principles of Internal Medicine, 17th ed. AS Fauci et al [eds]. New York, McGraw-Hill, 2008, p 1070.)*

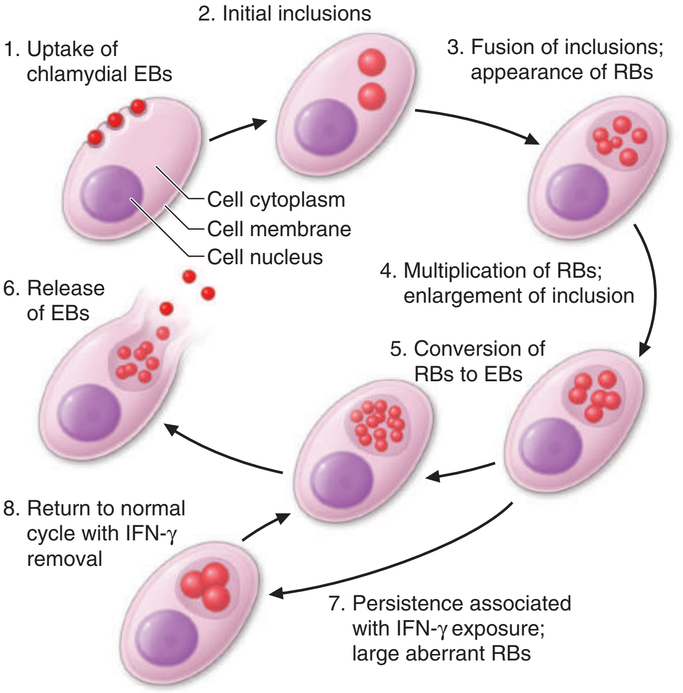
1. Uptake of chlamydial EBs
2. Initial inclusions
3. Fusion of inclusions; appearance of RBs
Cell cytoplasm
Cell membrane
Cell nucleus
4. Multiplication of RBs; enlargement of inclusion
5. Conversion of RBs to EBs
6. Release of EBs
7. Persistence associated with IFN-γ exposure; large aberrant RBs
8. Return to normal cycle with IFN-γ removal

FIGURE 194-2 Chlamydial life cycle. EBs, elementary bodies; IFN-γ, interferon γ; RBs, reticulate bodies. *(Reproduced with permission from WE Stamm: Chlamydial infections, in AS Fauci et al [eds]: Harrison's Principles of Internal Medicine, 17th ed. New York, McGraw-Hill, 2008.)*

typing of the major outer-membrane protein gene (*omp1*) from which serovar differences arise has been used to demonstrate that polymorphisms can occur in isolates from patients who are exposed frequently to multiple infections, while less variation is observed in isolates from less sexually active populations. Polymorphisms in the major outer-membrane protein may provide antigenic variation, and the different forms allow persistence in the community because immunity to one is not protective against the others.

The trachoma biovar is essentially a parasite of squamocolumnar epithelial cells; the LGV biovar is more invasive and involves lymphoid cells. As is typical of chlamydiae, *C. trachomatis* strains are capable of causing chronic, clinically inapparent, asymptomatic infections. Because the duration of the chlamydial growth cycle is ~48–72 h, the incubation period of sexually transmitted chlamydial infections is relatively long—generally 1–3 weeks. *C. trachomatis* causes cell death as a result of its replicative cycle and can induce cell damage whenever it persists. However, few toxic effects are demonstrated, and cell death because of chlamydial replication is not sufficient to account for disease manifestations, the majority of which are due to immunopathologic mechanisms or nonspecific host responses to the organism or its by-products.

In recent years, the entire genomes of various chlamydial species have been sequenced, the field of proteomics has become established, host innate immunity has been more precisely delineated, and innovative host cell–chlamydial interaction studies have been conducted. As a result, many insights have been gained into how chlamydiae adapt and replicate in their intracellular environment and produce disease. These insights into pathogenesis include information on the regulation of gene expression, protein localization, the type III secretion system, the roles of CD4+ and CD8+ T lymphocytes in the host response, and T lymphocyte trafficking.

The chlamydial heat-shock protein, which shares antigenic epitopes with similar proteins of other bacteria and with human heat-shock protein, may sensitize the host, and repeated infections may cause host cell damage. Persistent or recurrent chlamydial infections are associated with fibrosis, scarring, and complications following simple squamocolumnar epithelial infections. A common endpoint of these late consequences is scarring of mucous membranes. Genital complications can lead to pelvic inflammatory disease (PID) and its late consequences of tubal factor infertility, ectopic pregnancy, and chronic pelvic pain, while ocular infections may lead to blinding trachoma. High levels of antibody to human heat-shock protein have been associated with tubal factor infertility and ectopic pregnancy. Without adequate therapy, chlamydial infections may persist for several years, although symptoms—if present—usually abate.

Pathogenic mechanisms of *C. pneumoniae* have yet to be completely elucidated. The same is true for *C. psittaci*, except that this agent infects cells very efficiently and causes disease that may reflect direct cytopathic effects.

C. TRACHOMATIS INFECTIONS

GENITAL INFECTIONS (SEE ALSO CHAP. 141)

Spectrum Although chlamydiae cause a number of human diseases, localized lower genital tract infections caused by *C. trachomatis* and the sequelae of such infections are the most important in terms of medical and economic impact. Oculogenital infections due to *C. trachomatis* serovars D–K are transmitted during sexual contact or from mother to baby during childbirth and are associated with many syndromes, including cervicitis, salpingitis, acute urethral syndrome, endometritis, ectopic pregnancy, infertility, and PID in female patients; urethritis, proctitis, and epididymitis in male patients; and reactive arthritis, conjunctivitis, and pneumonia in infants. Women bear the greatest burden of morbidity because of the serious sequelae of these infections. Untreated infections may lead to PID, and multiple episodes of PID can lead to tubal factor infertility and chronic pelvic pain. Studies estimate that up to 80–90% of women and >50% of men with *C. trachomatis* genital infections lack symptoms; other patients have very mild symptoms. Additionally, extragenital infection in the pharynx and rectum is typically asymptomatic and may persist for years without treatment. Thus, a large reservoir of infected persons continues to transmit infection to sexual partners.

As their designations reflect, the LGV serovars (L_1, L_2, and L_3) cause LGV, an invasive sexually transmitted infection (STI) characterized by a transient and typically painless genital ulcer followed by acute lymphadenitis with bubo formation and/or acute hemorrhagic proctitis (see “Lymphogranuloma Venereum,” below).

Epidemiology • **GLOBAL EPIDEMIOLOGY** *C. trachomatis* genital infections are global in distribution. The World Health Organization (WHO) estimated that in 2020, 50 million (95% uncertainty interval [UI] 36-67) prevalent cases in men and 77 million [UI 67-90] prevalent cases in women and 59 million [95% UI 34-90] new cases of chlamydia in men and 70[95% UI 44-98] new cases in women. These figures make chlamydial infection the most prevalent bacterial STI in the world. There are large differences by WHO region; in 2020 in the Americas, the male incidence (per 1000) was 48 (95% UI, 23–86), compared with 12 (95% UI, 5–25) in Southeast Asia. In women in the Americas, there were 68 (95% UI, 39–104) incident cases (per 1000) compared with 16 (95% UI, 7–28) in Southeast Asia. The associated morbidity is substantial, and the economic cost is high.

U.S. EPIDEMIOLOGY In the United States, these infections are the most commonly reported of all infectious diseases. Chlamydia is notifiable nationally and is reported to the U.S. Centers for Disease Control and Prevention (CDC). In 2022, 1,649,716 cases were reported to the CDC, representing a 5-year decline of 6.2%. As chlamydial infections are typically asymptomatic, case reports are profoundly influenced by screening coverage. The disruptions in STI-related screening, prevention, and care activities during the COVID-19 pandemic necessitate that the trend for chlamydia surveillance data collected during the pandemic should be interpreted with caution. Higher rates among women than among men reflect the focus on expansion of screening programs for women during the past 25 years. Use of increasingly sensitive diagnostic nucleic acid amplification tests, an increased emphasis on case reporting, and improvements in the information systems used have elevated the number of cases reported every year. The CDC and other professional organizations recommend annual screening of all sexually active women <25 years of age and women aged >25 years who are at increased risk of infection, as well as rescreening of previously infected individuals at 3 months. Additionally, the CDC recommends that all pregnant women aged <25 years and those >25 years who are at increased risk be routinely screened for chlamydia during the first prenatal visit. Women who remain at risk for infection should be retested in the third trimester to prevent maternal complications and neonatal infection. In pregnancy, a test-of-cure is recommended 4 weeks after treatment. The 2022 U.S. total case count corresponds to 495 cases per 100,000 population. Women have the highest infection rates (621.2 cases per 100,000) compared to the rate among men (363.7 cases per 100,000). With the increased availability of urine testing and extragenital testing, men—including gay, bisexual, and other men who have sex with men (MSM)—are increasingly being tested for chlamydial infection. From 2021 to 2022, rates of chlamydial infection in men increased by 1.8%, whereas rates in women fell by 1.2% during this period. Chlamydia rates remain highest among adolescents and young adults; in 2022, 57.7% of all cases were reported among persons aged 15–24 years. Chlamydial infection rates vary considerably among different racial and ethnic populations as well as between states. For example, the rate in Louisiana per 100,000 population in 2022 was 788.6, compared with 198.0 in Vermont; the rate among blacks/African Americans was 1113.3 per 100,000 population compared with 100.6 in Asians.

The aforementioned statistics are based on case reporting; in 2018, the CDC estimated there were 4 million cases of chlamydia in the United States. Although screening programs have demonstrated reductions in PID rates, evidence is lacking to support general screening of young sexually active men. In contrast, sexually active MSM should undergo at least annual screening at anatomically exposed sites. Studies based on screening surveys estimate that the U.S. prevalence of *C. trachomatis* cervical infection was 5.9% and 8.8% among asymptomatic male and female college students, respectively; between 3.3 and 8.3% among prenatal patients; 7% for women seen in family planning clinics; and >10% for women seen in STI clinics. In a cohort of female college students, incident chlamydial infection was also associated with bacterial vaginosis and high-risk human papillomavirus infection. The prevalence of *C. trachomatis* in the cervix of pregnant women is approximately six times higher than that of *Neisseria gonorrhoeae*. The prevalence of genital infection with either agent is highest among women who are between the ages of 20 and 24. Recurrent infections are common in groups with greater vulnerability to STIs and are often acquired from untreated sexual partners. The use of hormonal contraception and the presence of cervical ectopy are also associated with an increased risk. The proportion of infections that are asymptomatic appears to be higher for *C. trachomatis* (83.9% in men, 87.1% in women) than for *N. gonorrhoeae*, (41.3% in men, 68.4% in women), and symptomatic *C. trachomatis* infections are clinically less severe. Mild or asymptomatic *C. trachomatis* infections of the fallopian tubes nonetheless cause ongoing tubal damage and infertility. The costs of *C. trachomatis* infections and their complications to the U.S. health care system have recently been estimated to be >$516.7 million annually.

Clinical Manifestations • **NONGONOCOCCAL AND POSTGONOCOCCAL URETHRITIS** *C. trachomatis* is the most common cause of nongonococcal urethritis (NGU) and postgonococcal urethritis (PGU). The designation *PGU* refers to NGU developing in men 2–3 weeks after treatment of gonococcal urethritis with single doses of agents such as penicillin or cephalosporins, which lack antimicrobial activity against chlamydiae. Current treatment regimens for gonorrhea have evolved and now include combination therapy with ceftriaxone and doxycycline unless chlamydia has been excluded; if chlamydia has been excluded, then gonorrhea is treated with ceftriaxone monotherapy. Thus, both the incidence of PGU and the causative role of *C. trachomatis* in this syndrome have declined.

In the United States, most of the estimated 2 million cases of acute urethritis are NGU, and *C. trachomatis* is implicated in 30–50% of these cases. The cause of most of the remaining cases of NGU is due to *Mycoplasma genitalium; Trichomonas vaginalis* and herpes simplex virus (HSV) cause some cases. Other etiologies include *Haemophilus* species, *N. meningitidis*, and adenovirus. The rate of involvement of *C. trachomatis* in urethral infection ranges from 3–7% among asymptomatic men to 15–20% among symptomatic men attending STI clinics. One multisite study of men in Baltimore, Seattle, Denver, and San Francisco reported an overall chlamydial prevalence of 7% in urine samples assessed by nucleic acid amplification tests (NAATs)—molecular tests that amplify the nucleic acids in clinical specimens. As in women, infection in men is age related, with young age most strongly associated with chlamydial urethritis. The prevalence among men is highest at 20–24 years of age. In STI clinics, urethritis is usually less prevalent among MSM than among heterosexual men.

NGU is diagnosed by documentation of leukocytes on urethral exudate and by exclusion of gonorrhea by Gram's staining, NAAT, or culture. *C. trachomatis* urethritis is generally less severe than gonococcal urethritis, although in any individual patient, these two forms of urethritis cannot reliably be differentiated solely on clinical grounds. Symptoms include urethral discharge (often whitish and mucoid rather than frankly purulent), dysuria, and urethral itching. Physical examination may reveal meatal erythema and tenderness as well as a urethral exudate that is often demonstrable only by stripping or milking of the urethra.

At least one-third of male patients with *C. trachomatis* urethral infection have no evident signs or symptoms of urethritis. The availability of NAATs for first-void urine specimens has facilitated broader-based testing for asymptomatic infection in male patients. As a result, asymptomatic chlamydial infection has been demonstrated in 5–10% of sexually active male adolescents screened at school-based clinics or community centers. Such patients generally have pyuria (≥15 leukocytes per 400× microscopic field in the sediment of first-void urine), a positive leukocyte esterase test, or an increased number of leukocytes on a Gram-stained smear prepared from a urogenital swab inserted 1–2 cm into the anterior urethra. When specific diagnostic tests for chlamydiae are not available, the examination of an endourethral specimen for increased leukocytes is useful in differentiating between true urethritis and functional symptoms in symptomatic patients. Alternatively, urethritis can be assayed noninvasively by examination of a first-void urine sample for pyuria, either by microscopy or by the leukocyte esterase test. Urine (or a urethral swab) can also be tested directly for chlamydiae by DNA amplification methods (NAATs), as described below (see "Detection Methods"). Urine testing for urethral STIs in men is much more acceptable than endourethral swab collection.

EPIDIDYMITIS Chlamydial urethritis may be followed by acute epididymitis (<6 weeks' duration), but this condition is rare, generally occurring in sexually active patients <35 years of age; in older men, epididymitis is usually associated with gram-negative bacterial infection and/or instrumentation procedures. An estimated 50–70% of cases of acute epididymitis are caused by *C. trachomatis*. The condition usually presents as unilateral scrotal pain with tenderness, swelling, and fever in a young man, often occurring in association with chlamydial urethritis. The illness may be mild enough to treat with oral antibiotics on an outpatient basis or severe enough to require hospitalization and parenteral therapy. Testicular torsion should be excluded promptly by imaging or surgical exploration in a teenager or young adult who presents with acute unilateral testicular pain without urethritis. The possibility of testicular tumor or chronic infection (>6 weeks' duration) (e.g., tuberculosis) should be excluded when a patient with unilateral intrascrotal pain and swelling does not respond to appropriate antimicrobial therapy.

REACTIVE ARTHRITIS Reactive arthritis consists of conjunctivitis, urethritis (or, in female patients, cervicitis), arthritis, and characteristic mucocutaneous lesions. It may develop in 1–2% of cases of NGU; the rate of chlamydia-induced arthritis has been estimated to be 4–15% and is thought to be the most common type of peripheral inflammatory arthritis in young men. When reactive arthritis is triggered by an STI, it may also be referred to as sexually acquired reactive arthritis (SARA). *C. trachomatis* has been recovered from the urethra of 16–44% of patients with reactive arthritis and 69% of men with signs of urogenital inflammation at the time of examination. Antibodies to *C. trachomatis* have also been detected in 46–67% of patients with reactive arthritis, and *Chlamydia*-specific cell-mediated immunity has been documented in 72%. In addition, *C. trachomatis* has been isolated from synovial biopsy samples from 15 of 29 patients in a number of small series and from a smaller proportion of synovial fluid specimens. Chlamydial nucleic acids have been identified in synovial membranes and chlamydial elementary bodies in joint fluid. The pathogenesis of reactive arthritis is unclear, but this condition probably represents an abnormal host response to a number of infectious agents, including those associated with bacterial gastroenteritis (e.g., *Salmonella*, *Shigella*, *Yersinia*, or *Campylobacter*), or to infection with *C. trachomatis* or *N. gonorrhoeae*. An association between reactive arthritis and the HLA-B27 phenotype in white individuals has been described, but not in patients from sub-Saharan Africa, where the prevalence of HLA-B27 is much lower. Since other mucosal infections produce an identical syndrome, chlamydial infection is thought to initiate an aberrant hyperreactive immune response that produces inflammation of the involved target organs in these genetically predisposed individuals. Evidence of exaggerated cell-mediated and humoral immune responses to chlamydial antigens in reactive arthritis supports this hypothesis. The finding of chlamydial elementary bodies and DNA in joint fluid and synovial tissue from patients with reactive arthritis suggests that chlamydiae may actually spread from genital to joint tissues in these patients—perhaps in macrophages.

NGU is the initial manifestation of reactive arthritis in 80% of patients, typically occurring within 14 days after sexual exposure. The urethritis may be mild and may even go unnoticed by the patient. Similarly, gonococcal urethritis may precede reactive arthritis, but co-infection with an agent of NGU is difficult to rule out. The urethral discharge may be purulent or mucopurulent, and patients may or may not report dysuria. Accompanying prostatitis, usually asymptomatic, has been described. Arthritis usually begins ~4 weeks after the onset of urethritis but may develop sooner or, in a small percentage of cases, may actually precede urethritis. The knees are most frequently involved; next most commonly affected are the ankles and small joints of the feet. Sacroiliitis, either symmetrical or asymmetrical, is documented in two-thirds of patients. Mild bilateral conjunctivitis, iritis, keratitis, or uveitis is sometimes present but lasts for only a few days. Finally, dermatologic manifestations occur in up to 50% of patients. The initial lesions—usually papules with a central yellow spot—most often involve the soles and palms and, in ~25% of patients, eventually epithelialize and thicken to produce keratoderma blenorrhagicum. Circinate balanitis is usually painless and occurs in fewer than half of patients. The initial episode of reactive arthritis usually lasts 2–6 months.

PROCTITIS Primary anal or rectal infections with *C. trachomatis* have been described in women and MSM who practice receptive anal intercourse. Symptoms are characterized by anorectal pain, a bloody mucopurulent discharge, and tenesmus. Oculogenital serovars D–K and LGV serovars L_1, L_2, and L_3 have been found to cause proctitis. The LGV serovars are far more invasive and typically cause more severely symptomatic disease, including severe ulcerative proctocolitis that can be clinically confused with HSV proctitis. However, asymptomatic or paucisymptomatic LGV infection is increasingly recognized in MSM. Histologically, LGV proctitis may resemble Crohn's disease in that giant cell formation and granulomas are detected. In the United States and Europe, cases of LGV proctitis occur almost exclusively in MSM, many of whom have HIV infection and other STI coinfections.

The less invasive non-LGV serovars of *C. trachomatis* cause mild proctitis. Many infected individuals are asymptomatic, and in these cases, infection is diagnosed only by routine NAAT of rectal swabs. The number of fecal leukocytes is usually abnormal in both asymptomatic and symptomatic cases. Sigmoidoscopy may yield normal findings or may reveal mild inflammatory changes or small erosions or follicles in the lower 10 cm of the rectum. Histologic examination of rectal biopsies generally shows anal crypts and prominent follicles as well as neutrophilic infiltration of the lamina propria. Chlamydial proctitis is best diagnosed by isolation of *C. trachomatis* from the rectum and documentation of a response to appropriate therapy. NAATs are the diagnostic test of choice.

MUCOPURULENT CERVICITIS Although most women with chlamydial infections of the cervix have no symptoms, almost half generally have local signs of infection on examination. Cervicitis is usually characterized by the presence of a mucopurulent discharge, with >20 neutrophils per microscopic field visible in strands of cervical mucus in a thinly smeared, Gram-stained preparation of endocervical exudate. Hypertrophic ectopy of the cervix may also be evident as an edematous area near the cervical os that is congested and bleeds easily on minor trauma (e.g., when a specimen is collected with a swab). A Papanicolaou smear shows increased numbers of neutrophils as well as a characteristic pattern of mononuclear inflammatory cells, including plasma cells, transformed lymphocytes, and histiocytes. Cervical biopsy shows a predominantly mononuclear cell infiltrate of the subepithelial stroma. Clinical experience and collaborative studies indicate that a cutoff of >30 polymorphonuclear leukocytes (PMNs)/1000× field in a Gram-stained smear of cervical mucus correlates best with chlamydial or gonococcal cervicitis. However, the wide availability of NAATs and limited access to microscopy outside of specialist centers has decreased the role of cervical mucus examination.

Clinical recognition of chlamydial cervicitis depends on a high index of suspicion and careful cervical examination. No genital symptoms are specifically correlated with chlamydial cervical infection. The differential diagnosis of a mucopurulent discharge from the endocervical canal in a young, sexually active woman includes gonococcal endocervicitis, salpingitis, endometritis, and intrauterine contraceptive device–induced inflammation. Diagnosis of cervicitis is based on the presence of PMNs on a cervical swab as noted above; the presence of chlamydiae is confirmed by NAAT.

PELVIC INFLAMMATORY DISEASE Inflammation of sections of the fallopian tube is often referred to as salpingitis or PID. The proportion of acute salpingitis cases caused by *C. trachomatis* varies geographically and with the population studied. It has been estimated that *C. trachomatis* causes up to 50% of PID cases in the United States. Recent studies report that the proportions of PID cases attributable to *N. gonorrhoeae* or *C. trachomatis* is decreasing; approximately 50% of women diagnosed with acute PID tested positive for either organism. PID occurs via ascending intraluminal spread of *C. trachomatis* or *N. gonorrhoeae* from the lower genital tract. Mucopurulent cervicitis is often followed by endometritis, endosalpingitis, and finally pelvic peritonitis. Evidence of mucopurulent cervicitis is often found in women with laparoscopically verified salpingitis. Similarly, endometritis, demonstrated by an endometrial biopsy showing plasma cell infiltration of the endometrial epithelium, is documented in most women with laparoscopy-verified chlamydial (or gonococcal) salpingitis. Chlamydial endometritis can also occur in the absence of clinical evidence of salpingitis. Histologic evidence of endometritis has been correlated with a syndrome consisting of vaginal bleeding, lower abdominal pain, and uterine tenderness in the absence of adnexal tenderness. Chlamydial salpingitis produces milder symptoms than gonococcal salpingitis and may be associated with less marked adnexal tenderness. Thus, mild adnexal or uterine tenderness in a sexually active woman with cervicitis may suggest chlamydial PID.

Screening and treating sexually active women for chlamydia and gonorrhea reduce their risk for PID. Chronic untreated endometrial and tubal inflammation can result in tubal scarring, impaired tubal function, tubal occlusion, and infertility, even among women who report no prior treatment for chlamydial infection. *C. trachomatis* has been particularly implicated in "subclinical" PID on the basis of a lack of history of PID among *Chlamydia*-seropositive women with tubal damage and detection of chlamydial DNA or antigen among asymptomatic women with tubal infertility. These data suggest that the best method to prevent PID and its sequelae is surveillance and control of lower genital tract infections along with diagnosis and treatment of sex partners and prevention of reinfections. Promotion of early symptom recognition and health care presentation may reduce the frequency and severity of sequelae of PID.

PERIHEPATITIS Fitz-Hugh–Curtis syndrome was originally described as a complication of gonococcal PID. However, studies over the past several decades have suggested that chlamydial infection is more commonly associated with perihepatitis than is *N. gonorrhoeae*. Perihepatitis should be suspected in young, sexually active women who develop right-upper-quadrant pain, fever, or nausea. Evidence of salpingitis may or may not be found on examination. Frequently, perihepatitis is strongly associated with extensive tubal scarring, adhesions, and inflammation observed at laparoscopy, and high titers of antibody to the 57-kDa chlamydial heat-shock protein have been documented. Culture and/or serologic evidence of *C. trachomatis* has been found in three-fourths of women with this syndrome.

URETHRAL SYNDROME IN WOMEN In the absence of infection with uropathogens such as coliforms or *Staphylococcus saprophyticus*, *C. trachomatis* is the pathogen most commonly isolated from college women with dysuria, frequency, and pyuria. Screening studies can recover *C. trachomatis* at both the cervix and the urethra; in up to 25% of infected women, the organism is isolated only from the urethra. The urethral syndrome in women consists of dysuria and frequency in conjunction with chlamydial urethritis, pyuria, and no bacteriuria or urinary pathogens. Although symptoms of the urethral syndrome may develop in some women with chlamydial infection, the majority of women attending STI clinics for urethral chlamydial infection do not have dysuria or frequency. Even in women with chlamydial urethritis causing the acute urethral syndrome, signs of urethritis such as urethral discharge, meatal redness, and swelling are uncommon. However, mucopurulent cervicitis in a woman presenting with dysuria and frequency strongly suggests *C. trachomatis* urethritis. Other correlates of chlamydial urethral syndrome include a duration of dysuria of >7–10 days, lack of hematuria, and lack of suprapubic tenderness. Abnormal urethral

Gram's stains showing >10 PMNs/1000× field in women with dysuria but without coliform bacteriuria support the diagnosis of chlamydial urethritis. Other possible diagnoses include gonococcal or trichomonal infection of the urethra.

INFECTION IN PREGNANCY AND THE NEONATAL PERIOD Infections during pregnancy can be transmitted to infants during delivery. Approximately 20–30% of infants exposed to *C. trachomatis* in the birth canal develop conjunctivitis, and 10–15% subsequently develop pneumonia. All newborn infants receive ocular prophylaxis at birth to prevent ophthalmia neonatorum. The agents used to prevent gonococcal eye infection are not effective against chlamydial conjunctivitis. The most effective measure to prevent neonatal chlamydial conjunctivitis is through screening and treatment of chlamydial infections in pregnant women. Without treatment, conjunctivitis usually develops at 5–19 days of life and often results in a profuse mucopurulent discharge. Roughly half of infected infants develop clinical evidence of inclusion conjunctivitis. However, it is impossible to differentiate chlamydial conjunctivitis from other forms of neonatal conjunctivitis (e.g., that due to *N. gonorrhoeae*, *Haemophilus influenzae*, *Streptococcus pneumoniae*, or HSV) on clinical grounds; thus, laboratory diagnosis is required. Inclusions within epithelial cells are often detected in Giemsa-stained conjunctival smears, but these smears are considerably less sensitive than cultures or NAATs for chlamydiae. Gram-stained smears may show gonococci or occasional small gram-negative coccobacilli in *Haemophilus* conjunctivitis, but smears should be accompanied by cultures or NAATs for these agents.

C. trachomatis has also been isolated frequently and persistently from the nasopharynx, rectum, and vagina of infected infants—occasionally for >1 year in the absence of treatment. In some cases, otitis media results from perinatally acquired chlamydial infection. Pneumonia may develop in infants from 2 weeks to 4 months of age. *C. trachomatis* is estimated to cause 20–30% of pneumonia cases in infants <6 months of age. Epidemiologic studies have linked chlamydial pulmonary infection in infants with increased occurrence of subacute lung disease (bronchitis, asthma, wheezing) in later childhood.

LYMPHOGRANULOMA VENEREUM *C. trachomatis* serovars L_1, L_2, and L_3 cause LGV, an invasive systemic STI. The peak incidence of LGV corresponds with the age of greatest sexual activity: the second and third decades of life. The worldwide incidence of LGV is falling, but the disease is still endemic and a major cause of morbidity in parts of Asia, Africa, South America, and the Caribbean. LGV is rare in industrialized countries; for more than a decade, the reported incidence in the United States has been only 0.1 case per 100,000 population. In the Bahamas, an apparent outbreak of LGV was described in association with a concurrent increase in heterosexual infection with HIV. Reports of outbreaks with the newly identified variant L_{2b} in Europe, Australia, and the United States indicate that LGV is becoming more prevalent among MSM. These cases have usually presented as hemorrhagic proctocolitis in HIV-positive men. However, LGV is increasingly described as an asymptomatic rectal infection in 30–50% of cases and is frequently seen in MSM without HIV. More widespread use of NAATs for identification of rectal infections may have enhanced case recognition.

LGV begins as a small painless papule that tends to ulcerate at the site of inoculation, often escaping attention. This primary lesion heals in a few days without scarring and is usually recognized as LGV only in retrospect. LGV strains of *C. trachomatis* have occasionally been recovered from genital ulcers and from the urethra of men and the endocervix of women who present with inguinal adenopathy; these areas may be the primary sites of infection in some cases. Proctitis is more common among people who practice receptive anal intercourse, and an elevated white blood cell count in anorectal smears may predict LGV in these patients. Ulcer formation may facilitate transmission of HIV infection and other sexually transmitted and blood-borne infections.

As NAATs for *C. trachomatis* are standard of care in high-income settings, increasing numbers of cases of LGV proctitis are being recognized in MSM. Such patients present with anorectal pain and mucopurulent, bloody rectal discharge. Sigmoidoscopy reveals ulcerative proctitis or proctocolitis, with purulent exudate and mucosal bleeding. Histopathologic findings in the rectal mucosa include granulomas with giant cells, crypt abscesses, and extensive inflammation. These clinical, sigmoidoscopic, and histopathologic findings may closely resemble those of Crohn's disease of the rectum.

The most common presenting picture in heterosexual men and women is the *inguinal syndrome*, which is characterized by painful inguinal lymphadenopathy beginning 2–6 weeks after presumed exposure; in rare instances, the onset comes after a few months. The inguinal adenopathy is unilateral in two-thirds of cases, and palpable enlargement of the iliac and femoral nodes is often evident on the same side as the enlarged inguinal nodes. The nodes are initially discrete, but progressive periadenitis results in a matted mass of nodes that becomes fluctuant and suppurative. The overlying skin becomes fixed, inflamed, and thin, and multiple draining fistulas finally develop. Extensive enlargement of chains of inguinal nodes above and below the inguinal ligament ("the groove sign") is not specific and, although not uncommon, is documented in only a minority of cases. Spontaneous healing usually takes place after several months; inguinal scars or granulomatous masses of various sizes persist for life. Massive pelvic lymphadenopathy may lead to exploratory laparotomy.

Constitutional symptoms are common during the stage of regional lymphadenopathy and, in cases of proctitis, may include fever, chills, headache, meningismus, anorexia, myalgias, and arthralgias. Other systemic complications are infrequent but include arthritis with sterile effusion, aseptic meningitis, meningoencephalitis, conjunctivitis, hepatitis, and erythema nodosum **(Fig. A1-39)**. Complications of untreated anorectal infection include perirectal abscess; anal fistulas; and rectovaginal, rectovesical, and ischiorectal fistulas. Secondary bacterial infection probably contributes to these complications. Rectal stricture is a late complication of anorectal infection and usually develops 2–6 cm from the anal orifice—i.e., at a site within reach on digital rectal examination. A small percentage of cases of LGV in men present as chronic progressive infiltrative, ulcerative, or fistular lesions of the penis, urethra, or scrotum. Associated lymphatic obstruction may produce elephantiasis. When urethral stricture occurs, it usually involves the posterior urethra and causes incontinence or difficulty with urination. In women, esthiomene of the genitalia is a late, rare complication of untreated LGV infection resulting in massive labial elephantiasis and ulceration followed by scarring.

Diagnosis • DETECTION METHODS

Historically, chlamydiae were cultivated in the yolk sac of embryonated eggs. The organisms can be grown more easily in tissue culture, but cell culture—once considered the diagnostic gold standard—has been replaced by nonculture assays **(Table 194-1)**. In general, culture for chlamydiae in clinical specimens is now performed only in specialized laboratories. The first nonculture assays, such as direct fluorescent antibody staining of clinical material and enzyme immunoassay (EIA), have been replaced by NAATs, which are currently recommended by the CDC as the diagnostic assays of choice. At present, numerous NAAT assays cleared by the U.S. Food and Drug Administration (FDA) are commercially available, some of which are available as high-throughput robotic platforms. Three point-of-care rapid diagnostic assays are available; they are of increasing interest since patients can potentially be treated before leaving the clinic, thus preventing forward transmission while patients wait for results from tests with longer turnaround times. However, turnaround time and cost of goods currently limit their widespread adoption. Over-the-counter tests are also in development.

CHOICE OF SPECIMEN Cervical and urethral swabs have traditionally been used for the diagnosis of STIs in female and male patients, respectively. However, given the greatly increased sensitivity and specificity of NAATs, less invasive samples (e.g., urine for both sexes and vaginal swabs for women) can be used. For screening of women, the CDC now recommends that self-collected or clinician-collected vaginal swabs, which are slightly more sensitive than urine, be used; a recent analysis demonstrated the superiority of vaginal swabs over urine in women. Urine screening tests are often used in outreach screening programs, however. For symptomatic women undergoing a pelvic examination, cervical swab samples may be used because they have slightly higher

TABLE 194-1 Diagnostic Tests for Sexually Transmitted and Perinatal *Chlamydia trachomatis* Infection

INFECTION	SUGGESTIVE SIGNS/SYMPTOMS	PRESUMPTIVE DIAGNOSIS[a]	CONFIRMATORY TEST OF CHOICE
Men			
NGU, PGU	Discharge, dysuria	Gram's stain with $\geq$2 WBCs/high power field (HPF) in high-prevalence settings (e.g., STI clinics) or $\geq$5 WBCs/HPF in lower-prevalence settings. No gram-negative intracellular diplococci (GNID). In the absence of microscopy: positive leukocyte esterase test on first catch urine	Urine NAAT for *C. trachomatis*
Epididymitis	Unilateral intrascrotal swelling, pain, tenderness; fever; NGU	Gram's stain with $\geq$2 (or $\geq$5) WBCs/HPF; no GNID; urinalysis with pyuria	Urine NAAT for *C. trachomatis*
Women			
Cervicitis	Mucopurulent cervical discharge, sustained endocervical bleeding easily induced by nontraumatic passage of a swab through the cervical os	Leukorrhea, defined as >10 WBCs/HPF on microscopic examination of vaginal fluid, might be a sensitive indicator of cervical inflammation with a high negative predictive value	Vaginal (or cervical) NAAT for *C. trachomatis*
Salpingitis	Lower abdominal pain, cervical motion tenderness, adnexal tenderness or masses	*C. trachomatis* always potentially present in salpingitis	Vaginal (or cervical) NAAT for *C. trachomatis*
Urethritis	Dysuria and frequency without hematuria	MPC; sterile pyuria; negative routine urine culture	Urine NAAT for *C. trachomatis*
Adults of Either Sex			
Proctitis	Rectal pain, discharge, tenesmus, bleeding; history of receptive anorectal intercourse	Negative gonococcal NAAT and Gram's stain; at least 1 WBC/HPF in rectal Gram's stain	Rectal NAAT for *C. trachomatis*
Reactive arthritis	NGU, arthritis, conjunctivitis, typical skin lesions	Gram's stain with $\geq$5 WBC/HPF; lack of GNID	Urine or vaginal NAAT for *C. trachomatis*
LGV	Regional adenopathy, primary lesion, proctitis, systemic symptoms	None	NAAT for *C. trachomatis*. Molecular testing for LGV is not widely available or not FDA cleared. If available, molecular PCR testing for *C. trachomatis* serovars L1, L2, or L3 can confirm diagnosis LGV
Neonates			
Conjunctivitis	Purulent conjunctival discharge 5–12 days after birth	Negative culture and Gram's stain for gonococci, *Haemophilus* spp., pneumococci, staphylococci	Tissue culture, conjunctival NAAT for *C. trachomatis* (not FDA cleared); DFA-stained scraping of conjunctival material
Infant pneumonia	Subacute, afebrile pneumonia in infants aged 1–3 months	None	Chlamydial tissue culture, DFA, or NAAT (not FDA cleared) of nasopharyngeal specimen

[a] A presumptive diagnosis of chlamydial infection is often made in the syndromes listed when gonococci are not found. A positive test for *Neisseria gonorrhoeae* does not exclude the involvement of *C. trachomatis*, which often is present in patients with gonorrhea.

Abbreviations: DFA, direct fluorescent antibody; FDA, U.S. Food and Drug Administration; HPF, high-power field; LGV, lymphogranuloma venereum; MPC, mucopurulent cervicitis; NAAT, nucleic acid amplification test; NGU, nongonococcal urethritis; PGU, postgonococcal urethritis; WBC, white blood cell.

Source: Based on Centers for Disease Control and Prevention Sexually Transmitted Infections Treatment Guidelines, 2021. *https://www.cdc.gov/std/treatment-guidelines/STI-Guidelines-2021.pdf.*

chlamydial counts. For male patients, a first-catch urine specimen is the sample of choice, but self-collected penile-meatal swabs have been shown to be very effective.

ALTERNATIVE SPECIMEN TYPES Ocular samples from babies and adults can be assessed by NAATs. However, NAATs are not cleared by the FDA for detecting chlamydia from conjunctival swabs, and clinical laboratories should verify the procedure according to Clinical Laboratory Improvement Amendments regulations. Samples from extragenital rectal and pharyngeal sites are used to detect chlamydiae by NAATs; in 2019, the CDC cleared the first NAATS for use on extragenital samples.

OTHER DIAGNOSTIC ISSUES Because NAATs detect nucleic acids instead of live organisms, they should be used with caution as test-of-cure assays. Residual nucleic acid from cells rendered noninfective by antibiotics may continue to yield a positive result in NAATs for as long as 3 weeks after therapy when viable organisms have actually been eradicated. Therefore, clinicians should not use NAATs for test of cure until after 3 weeks. The CDC currently does not recommend a test of cure after treatment for infection with *C. trachomatis* except in pregnancy. However, because incidence studies have demonstrated that previous chlamydial infection increases the probability of becoming reinfected, the CDC does recommend that previously infected individuals be rescreened 3 months after treatment.

SEROLOGY Serologic testing may be helpful in the diagnosis of LGV and neonatal pneumonia caused by *C. trachomatis*. The serologic test of choice is the microimmunofluorescence (MIF) test, in which high-titer purified elementary bodies mixed with embryonated chicken yolk sac material are affixed to a glass microscope slide to which dilutions of sera are applied. After incubation and washing, fluorescein-conjugated IgG or IgM antibody is applied. The test is read with an epifluorescence microscope, with the highest dilution of serum producing visible fluorescence designated as the titer. The MIF test is not widely available except in research laboratories and is highly labor intensive. Although the complement fixation (CF) test can also be used, it employs lipopolysaccharide (LPS) as the antigen and therefore identifies the pathogen only to the genus level. Single-point titers of >1:64 support a diagnosis of LGV, for which it is difficult to demonstrate rising antibody titers—i.e., paired serum samples are difficult to obtain since the disease often results in the patient's being seen by the physician after the acute stage. Any antibody titer of >1:16 is considered significant evidence of exposure to chlamydiae. However, serologic testing is never recommended for diagnosis of uncomplicated genital infections of the cervix, urethra, and lower genital tract or for *C. trachomatis* screening of asymptomatic individuals.

TREATMENT

C. trachomatis Genital Infections

A 7-day course of oral doxycycline (100 mg twice daily) is the recommended regimen of treatment for uncomplicated chlamydial infections. A single 1-g oral dose of azithromycin or oral levofloxacin 500 mg once daily for 7 days are alternatives. Doxycycline has slightly better efficacy than azithromycin in treatment of genital chlamydia infection and has demonstrated significantly higher microbiological cure rates in the treatment of rectal infection. The single-dose regimen of azithromycin has great appeal for the treatment of patients with uncomplicated chlamydial infection (especially those without symptoms and those with a likelihood of adherence challenges) and of the sexual partners of infected patients. These advantages must be weighed against the lower efficacy and greater cost of azithromycin. Whenever possible, the single 1-g dose should be given as directly observed therapy. Although not approved by the FDA for use in pregnancy, this regimen appears to be safe and effective for this purpose. Amoxicillin (500 mg three times daily for 7 days) can also be given as an alternative to pregnant women. The fluoroquinolones are contraindicated in pregnancy. A 2-week course of treatment is recommended for complicated chlamydial infections (e.g., PID, epididymitis). Traditionally, a 3-week course of doxycycline (100 mg orally twice daily), azithromycin 1 g orally once weekly for 3 weeks, or erythromycin base (500 mg orally four times daily) is recommended treatment for both bubonic and anogenital LGV. However, several studies have demonstrated the efficacy of shorter-course treatment for LGV. Failure of treatment with a tetracycline in genital infections usually indicates limited adherence or reinfection rather than involvement of a drug-resistant strain. To date, clinically significant drug resistance has not been observed in *C. trachomatis*.

Treatment or testing for chlamydiae should be considered among patients with *N. gonorrhoeae* because of the frequency of co-infection. Systemic treatment with erythromycin base or ethylsuccinate has been recommended for ophthalmia neonatorum and for *C. trachomatis* pneumonia in infants. Data on azithromycin efficacy are limited, but it may be effective for use in neonates. For the treatment of adult inclusion conjunctivitis, a single 1-g dose of azithromycin was as effective as standard 10-day treatment with doxycycline.

SEX PARTNERS

The continued high prevalence of chlamydial infections in most parts of the United States is due primarily to the failure to diagnose—and therefore treat—patients with symptomatic or asymptomatic infection and their sex partners. Urethral or cervical infection with *C. trachomatis* has been well documented in a high proportion of the sex partners of patients with NGU, epididymitis, reactive arthritis, salpingitis, and endocervicitis. If possible, confirmatory laboratory tests for chlamydiae should be undertaken in these individuals, but even persons without positive tests or evidence of clinical disease who have recently been exposed to proven or possible chlamydial infection (e.g., NGU) should be offered therapy. A novel approach is partner-delivered therapy, in which index patients receive treatment and are also provided with treatment to give to their sex partner(s).

NEONATES AND INFANTS

In neonates with conjunctivitis or infants with pneumonia, erythromycin ethylsuccinate or base can be given orally at a dosage of 50 mg/kg per day, preferably in four divided doses, for 2 weeks. Careful attention must be given to adherence to therapy—a frequent problem. Relapses of eye infection may follow oral erythromycin therapy. Thus, careful follow-up is required after treatment; the efficacy of erythromycin treatment for ophthalmia neonatorum is ~80%; therefore, a second course of therapy might be required. Both parents should be examined for *C. trachomatis* infection and, if diagnostic testing is not readily available, should be treated with doxycycline or azithromycin.

Prevention Since many chlamydial infections are asymptomatic, effective control and prevention must involve periodic screening of individuals at risk. Selective cost-effective screening criteria have been developed. Among women, young age (generally <25 years) is a critical risk factor for chlamydial infections in nearly all studies. Other risk factors include mucopurulent cervicitis; multiple, new, or symptomatic male sex partners; and lack of barrier contraceptive use. In some settings, screening based on young age may be as sensitive as criteria that incorporate behavioral and clinical measures. Another strategy is universal testing of all patients in high-prevalence clinic populations (e.g., STI clinics, juvenile detention facilities, and family planning clinics).

The effectiveness of selective screening in reducing the prevalence of chlamydial infection among women has been demonstrated in several studies. In the Pacific Northwest, where extensive screening began in family planning clinics in 1998 and in STI clinics in 1993, the prevalence declined from 10% in the 1980s to <5% in 2000. Similar trends have occurred in association with screening programs elsewhere. In addition, screening can reduce upper genital tract disease. In Seattle, women at a large health maintenance organization who were screened for chlamydial infection on a routine basis had a lower incidence of symptomatic PID than did women who received standard care and underwent more selective screening.

In settings with low to moderate prevalence, the prevalence at which selective screening becomes more cost-effective than universal screening must be defined. Most studies have concluded that universal screening is preferable in settings with a chlamydial prevalence of >3–7%. Depending on the criteria used, selective screening is likely to be more cost-effective when prevalence falls below 3%. The availability of highly sensitive and specific diagnostic NAATs using urine specimens and self-obtained vaginal swabs makes it feasible to mount an effective nationwide *Chlamydia* control program, with screening of individuals with increased vulnerability for STI acquisition in traditional health care settings and in novel outreach and community-based settings. The U.S. Preventive Services Task Force has named *Chlamydia* screening as a Grade B recommendation, which means that private insurance and Medicare will cover the cost of screening under the Affordable Care Act. More recently, home-based, self-collected, mail-in testing platforms have expanded the testing for chlamydia and other STIs outside of clinical settings. Since 2018, data have emerged on the efficacy of doxycycline postexposure prophylaxis. Efficacy data exist for MSM and transgender women, where a 200-mg single oral dose of doxycycline taken within 24–72 h of condomless sex reduced the risk of chlamydia acquisition by ~70–80%.

TRACHOMA

Epidemiology Trachoma—a sequela of ocular disease predominantly in resource-limited areas—continues to be a leading cause of preventable infectious blindness worldwide. The WHO estimates that ~6 million people have been blinded by trachoma and that ~1.3 million people in developing countries still suffer from preventable blindness due to trachoma; certainly, hundreds of millions live in trachoma-endemic areas. Foci of trachoma persist in Africa, Asia, Latin America, the Middle East, and the Pacific Rim. *C. trachomatis* serovars A, B, Ba, and C are isolated from patients with clinical trachoma in areas of endemicity in countries in Africa, the Middle East, Asia, and South America.

The trachoma-hyperendemic areas of the world are in northern and sub-Saharan Africa, the Middle East, drier regions of the Indian subcontinent, and Southeast Asia. In hyperendemic areas, the prevalence of trachoma is essentially 100% by the second or third year of life. Active disease is most common among young children, who are the reservoir for trachoma. By adulthood, active infection is infrequent, but sequelae result in blindness. In such areas, trachoma constitutes the major cause of blindness.

Trachoma is transmitted through contact with discharges from the eyes of infected patients. Transmission is most common under poor hygienic conditions and most often takes place between family members or between families with shared facilities. Eye-seeking flies (e.g., *Musca sorbens*) can also transfer the mucopurulent ocular discharges, carrying the organisms on their legs from one person to another. The International Trachoma Initiative founded by the WHO in 1998 aimed to eliminate blinding trachoma globally by 2020. The Neglected Tropical Disease road map 2021–2030 has set 2030 as the new target year for global elimination of trachoma as a public health problem.

Clinical Manifestations The clinical manifestations of trachoma include two phases: active trachoma (conjunctivitis) and cicatricial disease (conjunctival scarring). Both endemic trachoma and adult inclusion conjunctivitis present initially as conjunctivitis characterized by small lymphoid follicles in the conjunctiva. In regions with hyperendemic classic blinding trachoma, the disease usually starts insidiously before the age of 2 years. Reinfection is common and probably contributes to the pathogenesis of trachoma. Studies using polymerase chain reaction (PCR) or other NAATs indicate that chlamydial DNA is often present in the ocular secretions of patients with trachoma, even in the absence of positive cultures. Thus, persistent infection may be more common than was previously thought.

The cornea becomes involved, with inflammatory leukocytic infiltrations and superficial vascularization (pannus formation). As the inflammation continues, conjunctival scarring eventually distorts the eyelids, causing them to turn inward so that the lashes constantly abrade the eyeball (trichiasis and entropion); eventually the corneal epithelium is abraded and may ulcerate, with subsequent corneal scarring, opacification, and blindness. Destruction of the conjunctival goblet cells, lacrimal ducts, and lacrimal gland may produce a “dry-eye” syndrome, with resultant corneal opacity due to drying (xerosis) or secondary bacterial corneal ulcers.

Communities with blinding trachoma often experience seasonal epidemics of conjunctivitis due to *H. influenzae* that contribute to the intensity of the inflammatory process. In such areas, the active infectious process usually resolves spontaneously in affected persons at 10–15 years of age, but conjunctival scars continue to shrink, producing trichiasis and entropion with subsequent corneal scarring in adults. In areas with milder and less prevalent disease, the process may be much slower, with active disease continuing into adulthood; blindness is rare in these cases.

Eye infection with oculogenital *C. trachomatis* strains in sexually active young adults presents as an acute onset of unilateral follicular conjunctivitis and preauricular lymphadenopathy similar to that seen in acute conjunctivitis caused by adenovirus or HSV. If untreated, the disease may persist for 6 weeks to 2 years. It is frequently associated with corneal inflammation in the form of discrete opacities (“infiltrates”), punctate epithelial erosions, and minor degrees of superficial corneal vascularization. Very rarely, conjunctival scarring and eyelid distortion occur, particularly in patients treated for many months with topical glucocorticoids. Recurrent eye infections develop most often in patients whose sexual partners are not treated with antimicrobial agents.

Diagnosis The clinical diagnosis of classic trachoma can be made if two of the following signs are present: (1) lymphoid follicles on the upper tarsal conjunctiva; (2) typical conjunctival scarring; (3) vascular pannus; or (4) limbal follicles or their sequelae, Herbert pits. The clinical diagnosis of endemic trachoma should be confirmed by laboratory tests in children with relatively marked degrees of inflammation. Intracytoplasmic chlamydial inclusions are found in 10–60% of Giemsa-stained conjunctival smears in such populations, but chlamydial NAATs are more sensitive and are often positive when smears or cultures are negative. While NAATS are the most sensitive and specific diagnostic assays they are prohibitively expensive and require laboratory infrastructure that are not available in most endemic regions. Follicular conjunctivitis in European or American adults living in trachomatous regions is rarely due to trachoma.

TREATMENT
Trachoma

Adult inclusion conjunctivitis responds well to treatment with the same regimens used in uncomplicated genital infections—namely, azithromycin (a 1-g single oral dose) or doxycycline (100 mg twice daily for 7 days). Chlamydial resistance to azithromycin has not been documented. Topical tetracycline (1% eye ointment twice daily for 6 weeks) can be used as an alternative, but adherence rates are likely to be suboptimal. In the setting of eye infection with oculogenital strains, simultaneous treatment of all sexual partners is necessary to prevent ocular reinfection and chlamydial genital disease. Topical antibiotic treatment is not required for patients who receive systemic antibiotics.

PSITTACOSIS

Psittacine birds and many other avian species act as natural reservoirs for *C. psittaci*–type organisms, common pathogens in domestic mammals and birds. The species *C. psittaci,* which now includes only avian strains, affects humans only as a zoonosis. (The other strains previously included in this species have been placed into different species that reflect the animals they infect: *C. abortus*, *C. muridarum*, *C. suis, C. felis*, and *C. caviae.*) Although all birds are susceptible, pet birds (parrots, parakeets, macaws, and cockatiels) and poultry (turkeys and ducks) are most frequently involved in transmission of *C. psittaci* to humans. Exposure is greatest in poultry-processing workers and in owners of pet birds. Infectious forms of the organisms are shed from both symptomatic and apparently healthy birds and may remain viable for several months. *C. psittaci* can be transmitted to humans by direct contact with infected birds or by inhalation of aerosols from avian nasal discharges and from infectious avian fecal or feather dust. Transmission from person to person has never been demonstrated.

The diagnosis is usually established serologically. Psittacosis in humans may present as acute primary atypical pneumonia (which can be fatal in up to 10% of untreated cases); as severe chronic pneumonia; or as a mild illness or asymptomatic infection in persons exposed to infected birds.

■ EPIDEMIOLOGY

True incidence and prevalence of psittacosis is difficult to ascertain, in part due to lack of routine testing and the varying performance of commonly used diagnostic tests. Since 2010, the CDC has typically received 10 reports annually of confirmed cases of psittacosis, although many more cases probably occur than are reported. Outbreaks have been observed in, for example, poultry plant workers. Control of psittacosis depends on control of avian sources of infection. A pandemic of psittacosis was once stopped by banning shipment or importation of psittacine birds. Birds can receive prophylaxis in the form of a tetracycline-containing feed. Imported birds are currently quarantined for 30 days of treatment.

■ CLINICAL MANIFESTATIONS

Typical symptoms include fever, chills, muscular aches and pains, severe headache, hepato- and/or splenomegaly, and gastrointestinal symptoms. Hepatitis and neurologic complications can occur. Cardiac complications may involve endocarditis and myocarditis. Fatal cases were common in the preantibiotic era. As a result of quarantine of imported birds and improved veterinary-hygienic measures, outbreaks and sporadic cases of psittacosis are now rare. Severe pneumonia requiring management in an intensive care unit may develop. The incubation period is usually 5–19 days but can last as long as 28 days.

DIAGNOSIS

Previously, the most widely used serologic test for diagnosing chlamydial infections was the genus-specific CF test, in which assay of paired serum specimens often shows fourfold or greater increases in antibody titer. The CF test remains useful, but the gold standard of serologic tests is now the MIF test, which is not widely available (see section on diagnosis of *C. trachomatis* genital infection, above). Any antibody titer above 1:16 is considered significant evidence of exposure to chlamydiae, and a fourfold titer rise in paired sera in combination with a clinically compatible syndrome can be used to diagnose psittacosis. Some commercially available serologic tests based on measurement of antibodies to LPS can be useful when the clinical diagnosis is consistent with bird exposure; however, since these tests are reactive for all chlamydiae (i.e., all chlamydiae contain LPS), caution must be used in their interpretation. NAATs from respiratory samples are highly sensitive and specific but are not widely available in laboratories. *C. psittaci* is now considered a biohazard category B biothreat agent and has been associated with laboratory-acquired infections.

TREATMENT

Psittacosis

The antibiotic of choice is doxycycline; the dosage for adults is 100 mg twice a day, continued for 10–21 days; it is unclear if longer courses prevent relapse. Severely ill patients may need intravenous doxycycline and cardiovascular and respiratory support. Tetracycline (500 mg four times a day by mouth) and azithromycin 250–500 mg by mouth daily for 7 days are alternative therapies. Erythromycin (500 mg four times a day by mouth) appears to be inferior to other agents in animal models.

C. PNEUMONIAE INFECTIONS

C. pneumoniae is a common cause of human respiratory diseases, such as pneumonia and bronchitis. This organism reportedly accounts for as many as 10% of cases of community-acquired pneumonia, but more typically 1–2%. PCR is the preferred method for diagnosis of acute infections. Serologic studies have linked *C. pneumoniae* to atherosclerosis; isolation and PCR detection in cardiovascular tissues have also been reported. These findings suggest an expanded range of diseases and syndromes for *C. pneumoniae*. Large-scale case–cohort studies have demonstrated some association of *C. pneumoniae* with lung cancer, as evaluated by serology.

EPIDEMIOLOGY

Primary infection occurs mainly in school-aged children and reinfection in adults. Seroprevalence rates of 40–70% show that *C. pneumoniae* is widespread in both industrialized and developing countries. Seropositivity usually is first detected at school age, and rates generally increase by ~10% per decade. About 50% of individuals have detectable antibody at 30 years of age, and most have detectable antibody by the eighth decade of life. Although, as mentioned, serologic evidence suggests that *C. pneumoniae* may be associated with up to 10% of cases of community-acquired pneumonia, most of this evidence is based not on paired serum samples but rather on a single high IgG titer. Some doubt exists about the true prevalence and etiologic role of *C. pneumoniae* in atypical pneumonia, especially since reports of cross-reactivity have raised questions about the specificity of serology when only a single serum sample is used for diagnosis.

PATHOGENESIS

Little is known about the pathogenesis of *C. pneumoniae* infection. It begins in the upper respiratory tract and may persist as a prolonged asymptomatic condition of the upper respiratory mucosal surfaces. However, evidence of replication within vascular endothelium and synovial membranes of joints shows that, in at least some individuals, the organism is transported to distant sites, perhaps within macrophages. A *C. pneumoniae* outer-membrane protein may induce host immune responses whose cross-reactivity with human proteins results in an autoimmune reaction.

The role of *C. pneumoniae* in the etiology of atherosclerosis has been discussed since 1988, when Finnish researchers presented serologic evidence of an association of this organism with coronary heart disease and acute myocardial infarction. Subsequently, the organism was identified in atherosclerotic lesions by culture, PCR, immunohistochemistry, and transmission electron microscopy; however, discrepant study results (including those of animal studies) and failure of large-scale treatment studies have raised doubts as to the etiologic role of *C. pneumoniae* in atherosclerosis. Epidemiologic studies have demonstrated an association between serologic evidence of *C. pneumoniae* infection and atherosclerotic disease of the coronary and other arteries. In addition, *C. pneumoniae* has been identified in atherosclerotic plaques by electron microscopy, DNA hybridization, and immunocytochemistry. The organism has been recovered in culture from atheromatous plaques—a result indicating the presence of viable replicating bacteria in vessels. Evidence from animal models supports the hypothesis that *C. pneumoniae* infection of the upper respiratory tract is followed by recovery of the organism from atheromatous lesions in the aorta and that the infection accelerates the process of atherosclerosis, especially in hypercholesterolemic animals. Antimicrobial treatment of the infected animals reverses the increased risk of atherosclerosis. In humans, two small trials in patients with unstable angina or recent myocardial infarction suggested that antibiotics reduce the likelihood of subsequent cardiac events. However, larger-scale trials have not documented an effect of various antichlamydial regimens on the risk of these events.

CLINICAL MANIFESTATIONS

C. pneumoniae was first reported as the etiologic agent of mild atypical pneumonia in military recruits and college students. The clinical spectrum of *C. pneumoniae* infection includes acute pharyngitis, sinusitis, bronchitis, and pneumonitis, primarily in young adults. The clinical manifestations of primary infection appear to be more severe and prolonged than those of reinfection. The pneumonitis of *C. pneumoniae* pneumonia resembles that of *Mycoplasma* pneumonia in that leukocytosis is frequently lacking and patients often have prominent antecedent upper respiratory tract symptoms, fever, nonproductive cough, mild to moderate illness, minimal findings on chest auscultation, and small segmental infiltrates on chest x-ray. In elderly patients, pneumonia due to *C. pneumoniae* can be especially severe and may necessitate hospitalization and respiratory support.

Chronic infection with *C. pneumoniae* has been reported among patients with chronic obstructive pulmonary disease and may play a role in the natural history of asthma, including exacerbations. The clinical symptoms of respiratory infections caused by *C. pneumoniae* are nonspecific and do not differ from those caused by other agents of atypical pneumonia, such as *Mycoplasma pneumoniae*.

DIAGNOSIS

PCR amplification of respiratory secretions is the preferred method of diagnosis; several commercial assays are available. Serology and culture can be used to diagnose *C. pneumoniae* infection. Serology was the traditional diagnostic method. Serology is not FDA approved because of its poor predictive value and is not routinely used clinically. Cell culture, which can perform similarly to PCR and allows for antimicrobial susceptibility testing, is seldom used as most clinical laboratories are not equipped to culture *Chlamydia* spp. The organism is very difficult to grow in tissue cultures but has been cultivated in HeLa cells, HEp-2 cells, and HL cells. Previously, the gold standard serologic test was the MIF test (see section on diagnosis of *C. trachomatis* genital infection, above). Any antibody titer >1:16 is considered significant evidence of exposure to chlamydiae. According to a CDC-sponsored expert working group, the diagnosis of acute *C. pneumoniae* infection requires demonstration of a fourfold rise in titer in paired serum samples. There are no official recommendations for diagnosis of chronic infections, although many research studies have used high titers of IgA as an indicator. The older CF tests and EIAs for LPS are not recommended, as they are not specific for *C. pneumoniae* but identify the chlamydiae only to the genus level.

TREATMENT

C. pneumoniae Infections

Although few controlled trials of treatment have been reported, *C. pneumoniae* is inhibited in vitro by erythromycin, tetracycline, azithromycin, clarithromycin, gatifloxacin, and gemifloxacin. Directed therapies include azithromycin 500 mg orally once followed by 250 mg on days 2–5; doxycycline 100 mg orally twice daily; or clarithromycin 500 mg twice daily. The fluoroquinolones levofloxacin (750 mg orally once daily) and moxifloxacin (400 mg orally once daily) are alternatives, but their role is limited by increasing safety concerns due to serious side effects. For most patients, a 5-day course of therapy is sufficient. Beta-lactams and trimethoprim/sulfamethoxazole are not active.

Acknowledgment

The authors wish to thank Dr. Charlotte A. Gaydos for her contributions to this chapter in previous editions.

FURTHER READING

Centers for Disease Control and Prevention: *Sexually Transmitted Infections Surveillance, 2022*. Atlanta, GA: U.S. Department of Health and Human Services, 2024. *https://www.cdc.gov/std/statistics/2022/default.htm*.

Centers for Disease Control and Prevention: Sexually transmitted infections treatment guidelines, 2021. MMWR Recomm Rep 70:1, 2021.

Elwell C et al: Chlamydia cell biology and pathogenesis. Nat Rev Microbiol 14:385, 2016.

Gaydos CA, Essiq A: Chlamydiaceae, in *Manual of Clinical Microbiology*, 11th ed. JH Jorgensen et al (eds). Washington, DC, ASM Press, 2015, pp 1106–1121.

Goller JL et al: Population attributable fraction of pelvic inflammatory disease associated with chlamydia and gonorrhoea: A cross-sectional analysis of Australian sexual health clinic data. Sex Transm Infect 92:525, 2016.

Gregory ECW, Ely DM: Trends and characteristics of sexually transmitted infections during pregnancy: United States, 2016–2018. National Vital Statistics Reports 69:1, 2020.

Hammerschlag MR et al: *Chlamydia pneumoniae*, in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 9th ed. JE Bennett, R Dolin, MJ Blaser (eds). Philadelphia, Elsevier, 2020, Chapter 182.

Hughes Y et al: Universal lymphogranuloma venereum (LGV) testing of rectal chlamydia in men who have sex with men and detection of asymptomatic LGV. Sex Transm Infect 98:582, 2022.

Kuypers J et al: Principles of laboratory diagnosis of STIs, in *Sexually Transmitted Diseases*, 4th ed. KK Holmes et al (eds). New York, McGraw-Hill, 2008, pp 937–948.

Luetkemeyer AF et al: Postexposure doxycycline to prevent bacterial sexually transmitted infections. N Engl J Med 388:1296, 2023.

Papp JR et al: Recommendations for the laboratory-based detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae*, 2014. MMWR 63:1, 2014.

Rowley J et al: Chlamydia, gonorrhea, trichomoniasis and syphilis: Global prevalence and incidence estimates, 2016. Bull World Health Organ 97:548, 2019.

Schachter J, Stephens RS: Biology of *Chlamydia trachomatis*, in *Sexually Transmitted Diseases*, 4th ed. KK Holmes et al (eds). New York, McGraw-Hill, 2008, pp 555–574.

Taylor HR: *Trachoma: A Blinding Scourge from the Bronze Age to the Twenty-First Century*. East Melbourne, Victoria, Australia, Centre for Eye Research Australia/Haddington Press, 2008.

Section 11 Viral Diseases: General Considerations

195

Principles of Medical Virology

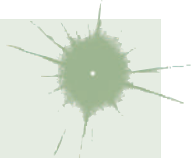

David M. Knipe, Max L. Nibert

Viruses are obligate intracellular parasites that must enter cells to replicate and propagate themselves to spread to other cells. Infection often injures the host cell—hence the name "virus," derived from the Latin word *virus* for poison or toxin. Viruses are one of the simplest life forms and, at the minimum, have a nucleic acid genome with a protein coat. They do not divide by division, as do cells; instead, viruses are programmed to disassemble inside cells, to use their nucleic acid genome to encode viral proteins that replicate their genomic nucleic acid, and then to assemble the progeny genomes into viral particles. The progeny viruses are secreted or released from the host cell as extracellular *virions* that infect surrounding cells. Viruses depend on the host cell for many of the enzymes and organelles that synthesize carbohydrates, lipids, nucleic precursors and nucleic acids, and high-energy molecules, including the host cell's ribosomes, which are used to make viral proteins. In the process of taking over the host cell, viruses inhibit normal cell metabolic pathways and cause damage to the cell in a process that results in the *cytopathic effect* (CPE). Injury to cells and cell death can cause tissue damage and contribute to virus-induced disease.

Viruses are distinct from other intracellular parasites such as viroids, virusoids, prions, and intracellular bacteria. *Viroids* are small, circular, single-stranded RNA infectious pathogens of plants that do not have a protein coat, while *virusoids* are small, circular-RNA, infectious pathogens that depend on viruses to provide the proteins for their replication and protein coat. *Prions* are misfolded proteins that spread from one cell to another, causing the same protein molecules to misfold in the new cell. The misfolded proteins in prions cause cellular damage **(Chap. 449)**.

VIRUS STRUCTURE

There are many different virus structures, but nearly all are formed from a few fundamental structural elements. The minimal virion particle is composed of a complex of nucleic acids (the *genome*) and a protein shell (the *capsid*) **(Fig. 195-1)**. The combination of the genome and the capsid is called the *nucleocapsid*. The genome is protected within the capsid. The external surface of virions can consist of either the protein capsid or a lipid envelope around the capsid (Fig. 195-1).

Viral genomes can consist of single- or double-stranded RNA or DNA and can comprise one or more genome segments. Single-stranded (ss) genomes are designated as *positive strand* (+) if they contain the sequences encoding the open reading frames for viral proteins, while they are designated as *negative strand* (–) if they contain only complementary sequences. Thus, a positive-strand RNA viral genome can be translated into a viral protein upon entry into the host cell, while a negative-strand genome must be copied into complementary RNA molecules for translation. This dilemma is solved in negative-strand viruses by the loading of transcriptases onto the viral genome prior to encapsidation; these enzymes transcribe the genome into viral mRNA upon entry into and uncoating within the cell.

Viral capsids are made of repeating protein subunits because their genomes have limited coding capacity. The capsids are constructed with a few structural units or capsomers packed into a symmetrical arrangement. Capsids are usually organized in one of two ways: (1) an *icosahedral* or *spherical* symmetry based on an icosahedron with two-, three-, and fivefold axes of symmetry formed from 20 triangular faces or (2) a *helical* symmetry. However, viruses occasionally have more complex structures (e.g., the poxviruses) **(Fig. 195-2)**.

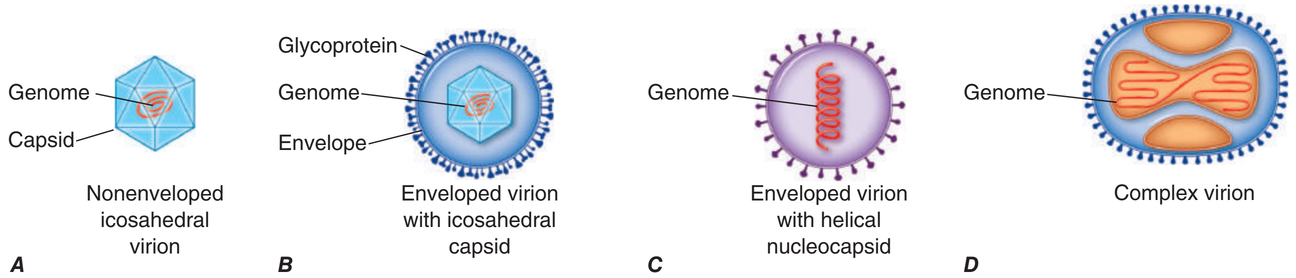
Genome
Capsid
Nonenveloped icosahedral virion
A
Glycoprotein
Genome
Envelope
Enveloped virion with icosahedral capsid
B
Genome
Enveloped virion with helical nucleocapsid
C
Genome
Complex virion
D

FIGURE 195-1 Schematic diagrams of the major forms of human viruses. ***A.*** Icosahedral capsid without an envelope. ***B.*** Icosahedral capsid with a lipid envelope. ***C.*** Helical capsid with a lipid envelope. ***D.*** Complex virion.

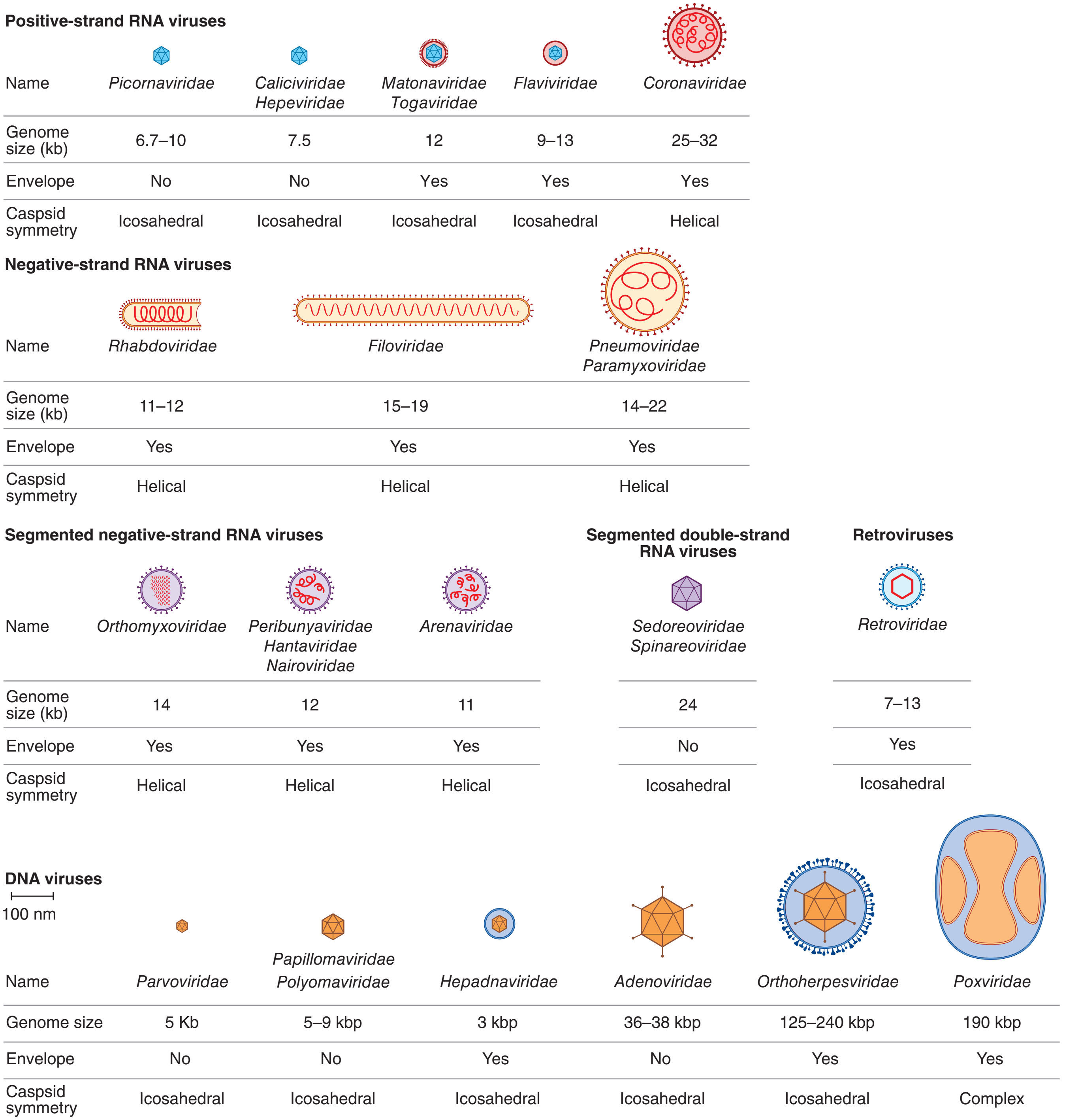

Positive-strand RNA viruses

Name	*Picornaviridae*	*Caliciviridae* *Hepeviridae*	*Matonaviridae* *Togaviridae*	*Flaviviridae*	*Coronaviridae*
Genome size (kb)	6.7–10	7.5	12	9–13	25–32
Envelope	No	No	Yes	Yes	Yes
Caspsid symmetry	Icosahedral	Icosahedral	Icosahedral	Icosahedral	Helical

Negative-strand RNA viruses

Name	*Rhabdoviridae*	*Filoviridae*	*Pneumoviridae* *Paramyxoviridae*
Genome size (kb)	11–12	15–19	14–22
Envelope	Yes	Yes	Yes
Caspsid symmetry	Helical	Helical	Helical

Segmented negative-strand RNA viruses

Name	*Orthomyxoviridae*	*Peribunyaviridae* *Hantaviridae* *Nairoviridae*	*Arenaviridae*
Genome size (kb)	14	12	11
Envelope	Yes	Yes	Yes
Caspsid symmetry	Helical	Helical	Helical

Segmented double-strand RNA viruses

Name	*Sedoreoviridae* *Spinareoviridae*
Genome size (kb)	24
Envelope	No
Caspsid symmetry	Icosahedral

Retroviruses

Name	*Retroviridae*
Genome size (kb)	7–13
Envelope	Yes
Caspsid symmetry	Icosahedral

DNA viruses

Name	*Parvoviridae*	*Papillomaviridae* *Polyomaviridae*	*Hepadnaviridae*	*Adenoviridae*	*Orthoherpesviridae*	*Poxviridae*
Genome size	5 Kb	5–9 kbp	3 kbp	36–38 kbp	125–240 kbp	190 kbp
Envelope	No	No	Yes	No	Yes	Yes
Caspsid symmetry	Icosahedral	Icosahedral	Icosahedral	Icosahedral	Icosahedral	Complex

FIGURE 195-2 Schematic diagrams of viruses of the major families that infect humans. The viruses are grouped by genotype, and the virions are drawn approximately to scale. Prototype viruses of each family are listed in Table 195-1. *(Source: Modified from Fig. 185-2 in Harrison's Principles of Internal Medicine, 20th ed.)*

Enveloped viruses (e.g., measles virus) are efficient in infecting cells because the viral lipid membrane fuses easily with the plasma membrane of the host cell or with internal membranes to deliver the nucleocapsid to the cytoplasm of the host cell. Thus, these viruses are highly transmissible. The lipid envelope is susceptible to disruption by detergents or organic solvents; thus, enveloped viruses such as measles virus, coronaviruses, and influenza viruses can be inactivated by soap and water or alcohol-based hand sanitizers. In contrast, unenveloped viruses (e.g., norovirus or poliovirus) have a tough protein shell whose resistance to small-intestine bile salts—a surfactant that emulsifies lipids—allows them to infect the intestine. Unenveloped viruses, especially those that infect the gastrointestinal tract, are not inactivated by detergents or organic solvents and must be inactivated by peroxide or hypochlorite or removed by washing with soap and water.

CLASSIFICATION OF VIRUSES

Viruses are classified as a free-standing groups because they are not formally related to organisms within any of the major kingdoms. The highest level of viral classification was originally the family, but there have been efforts to classify viruses into higher ranks, culminating in kingdoms and realms. This higher classification is largely not that relevant to medical virology because the major viruses of clinical interest can be conveniently classified into a number of families **(Table 195-1)**, each of which has characteristic virion and genome structures (Fig. 195-2). Classification of viruses into families, genera, and species was previously based on multiple criteria, including type of genomic nucleic acid (i.e., RNA or DNA; ss positive or negative strand or double strand), capsid symmetry (helical, icosahedral, or complex), presence or absence of an envelope, mode of replication, and tropism (preferred cell type for replication) or type of disease it causes. Recent sequence analysis of viral genomes has refined and revised some of the original virus classifications. The International Committee on Taxonomy of Viruses specifies both formal and common names for viruses. For example, herpes simplex virus (HSV) is the common name for species *simplex virus human alpha 1*.

VIRAL REPLICATION IN CELLS

Viral replication takes place in the host cell by the following steps: binding, entry, uncoating, transport to the site of replication, transcription of mRNA, translation of viral proteins, replication of the input genome, assembly of progeny viral particles, and egress from the cell. All viruses must enter cells by mechanisms that allow virus binding to the cell surface and subsequent crossing of the plasma membrane and/or other membranes to gain entry into the cytoplasm. After entry, the mechanisms of replication diverge for the different viruses, depending on the nature of the viral genome.

■ VIRAL ENTRY

Viruses bind to specific receptors on the cell surface and generally enter cells by one of three pathways: (1) fusion of the envelope with the surface plasma membrane; (2) endocytosis followed by fusion with the endosome membrane; or (3) lysis of the endosome or formation of pores in the endosome. Viruses often bind to a charged molecule on the surface of cells to concentrate themselves thereon. They then bind more specifically to a protein or carbohydrate molecule, and this binding triggers endocytosis or fusion of the viral envelope with the cellular plasma membrane. Endocytosis can occur by any of several mechanisms, including clathrin-mediated endocytosis, macropinocytosis, micropinocytosis, and caveolar endocytosis. After viral entry into endocytic vesicles, acidification of the vesicles leads to conformational changes in the viral glycoproteins, fusion of the viral envelope with the endocytic membrane, and release of the nucleocapsid into the cytoplasm. At the entry stage or later, the genome must be uncoated or the capsid opened sufficiently to allow transcription, translation, and/or replication.

■ VIRAL REPLICATION STRATEGIES

Positive-Strand RNA Viruses The RNA genomes of the picornaviruses, caliciviruses, hepeviruses, togaviruses, flaviviruses, and coronaviruses can be translated in the cytoplasm directly after removal of the capsid coat or uncoating. The picornaviral and flaviviral genomic RNA is translated into a polyprotein that is cleaved by viral and cellular proteases to generate (1) nonstructural proteins that replicate the genomic RNA to complementary negative-strand molecules and then back to positive-strand RNA molecules and (2) structural proteins that assemble capsids for progeny virions. Replication of positive-strand viral RNA takes place in replication complexes associated with cytoplasmic membranes, often in membrane sacs that concentrate the components, protect them from host responses, and provide the redox environment needed for optimal replication. Progeny virions are released when the host cell lyses. The positive-strand genome RNA of the caliciviruses, hepatitis E virus (a hepevirus), and the togaviruses is translated to generate a polyprotein, which, when cleaved by viral and cellular proteases, yields the nonstructural proteins that replicate the viral genome to a negative-strand copy and then synthesize new full-length positive strands and a subgenomic mRNA that encodes the structural proteins. Progeny virions are released by budding or cell lysis, depending on whether the virus is enveloped or not. Replication of the genome to the negative strand is followed by a transition back to the positive-strand genome for translation and encapsidation. Progeny virions are released by budding.

Negative-Strand RNA Viruses The rhabdoviruses, filoviruses, and paramyxoviruses have a single negative strand of genome RNA that is transcribed by a virion-associated RNA-dependent RNA polymerase (transcriptase) to yield subgenomic mRNAs that encode the replicase and structural proteins. The replicase copies the full-length negative-strand RNA to a full-length positive-strand RNA and then back to a full-length negative strand, which is assembled into nucleocapsids that bud out of the cell to form progeny virions.

The influenza viruses, peri bunyaviruses, and arenaviruses have segmented negative RNA genomes that are transcribed by virion-associated transcriptases to yield mRNAs that encode nonstructural and structural proteins. The replicase enzyme complex copies the negative-strand RNA genomes to full-length positive-strand copies and back to full-length negative-strand RNA molecules. The peri bunyaviruses and arenaviruses replicate entirely in the cytoplasm. In contrast, influenza viral transcription takes place in the nucleus, with nascent cellular transcripts serving as primers to yield mRNAs that are transported to the cytoplasm for translation. Viral proteins are transported into the nucleus to promote genome replication, and progeny negative-strand RNAs are transported to the cytoplasm to bud into progeny virions. Some of the bunyaviruses and the arenaviruses have open reading frames on the "negative strand." Thus, these viruses use both negative- and positive-sense or ambisense coding of their RNA genomes. The full-length negative strands are assembled in the correct assortment in capsid proteins and then bud to yield infectious progeny virions.

Double-Stranded RNA Viruses The reovirus and rotavirus genomes consist of multiple double-stranded (ds) RNA molecules that are transcribed by virion-associated, RNA-dependent RNA polymerases (transcriptases) to yield mRNAs encoding nonstructural and structural proteins. Following viral protein synthesis, replication of positive-strand RNAs to form dsRNA molecules and assembly into viral capsids occur in cytoplasmic viral factories. Progeny viruses are released when infected cells lyse.

Double-Stranded DNA Viruses Most dsDNA viral genomes are transported to the infected cell's nucleus for transcription and replication. The host cell recognizes foreign DNA that is not fully loaded with histone nucleosomes with a normal pattern and tries to epigenetically silence these molecules; DNA viruses have evolved mechanisms to overcome these epigenetic silencing mechanisms. The dsDNA genomes of the papovaviruses and papillomaviruses are coated with nucleosomal chromatin in the virion and therefore are delivered to the nucleus in a form that is not recognized as foreign. Viral early gene expression is promoted by an enhancer adjacent to the early gene promoter, which is transcribed by host cell RNA polymerase II to yield the early mRNAs. The early proteins promote viral DNA replication by host enzymes, and late genes are then transcribed. The late proteins encode the capsid proteins to assemble progeny virions.

TABLE 195-1 Major Families of Human Pathogenic Viruses

FAMILY	REPRESENTATIVE VIRUSES	TYPE OF RNA/DNA	LIPID ENVELOPE
Picornaviridae	Coxsackievirus Echovirus Enteroviruses, including poliovirus Rhinoviruses Hepatitis A virus	(+) RNA	No
Caliciviridae	Norovirus	(+) RNA	No
Hepeviridae	Hepatitis E virus	(+) RNA	No
Matonaviridae	Rubella virus	(+) RNA	Yes
Togaviridae	Eastern equine encephalitis virus Western equine encephalitis virus	(+) RNA	Yes
Flaviviridae	Yellow fever virus Dengue virus St. Louis encephalitis virus West Nile virus Zika virus Hepatitis C virus Hepatitis G virus	(+) RNA	Yes
Coronaviridae	SARS-CoV-1 SARS-CoV-2 Middle East respiratory syndrome virus	(+) RNA	Yes
Rhabdoviridae	Rabies virus Vesicular stomatitis virus	(–) RNA	Yes
Filoviridae	Marburg virus Ebola virus	(–) RNA	Yes
Pneumoviridae	Respiratory syncytial virus	(–) RNA	Yes
Paramyxoviridae	Parainfluenza virus Newcastle disease virus Mumps virus Rubeola (measles) virus	(–) RNA	Yes
Orthomyxoviridae	Influenza A, B, and C viruses	(–) RNA, 8 segments	Yes
Peribunyaviridae	California encephalitis virus	(–) RNA, 3 segments	Yes
Hantaviridae	Hantavirus	(–) RNA, 3 segments	Yes
Nairoviridae	Crimean–Congo hemorrhagic fever virus	(–) RNA, 3 segments	Yes
Arenaviridae	Lymphocytic choriomeningitis virus Lassa fever virus South American hemorrhagic fever virus	(–) RNA, 2 segments	Yes
Sedoreoviridae	Rotavirus	dsRNA, 11 segments	No
Spinareoviridae	Reovirus Colorado tick fever virus	dsRNA, 10–12 segments	No
Retroviridae	Human T lymphotropic virus 1 and 2 Human immunodeficiency virus 1 and 2	(+) RNA, 2 identical segments	Yes
Hepadnaviridae	Hepatitis B virus	dsDNA with ss portions	Yes
Parvoviridae	Parvovirus B19	ssDNA	No
Papillomaviridae	Human papillomaviruses	dsDNA	No
Polyomaviridae	JC virus BK virus Merkel cell polyoma virus	…	…
Adenoviridae	Human adenoviruses	dsDNA	No
Orthoherpesviridae	Herpes simplex virus 1 and 2 Varicella-zoster virus Epstein-Barr virus Cytomegalovirus Human herpesvirus 6 Human herpesvirus 7 Kaposi's sarcoma–associated herpesvirus	dsDNA	Yes
Poxviridae	Variola (smallpox) virus Orf virus Molluscum contagiosum virus	dsDNA	Yes

Abbreviations: ds, double-stranded; ss, single-stranded.

The dsDNA genomes of adenoviruses are delivered to the infected cell's nucleus coated with a viral protein that hides the viral genomes from the host's epigenetic silencing mechanisms. Viral DNA genomes are transported to and released through the nuclear pores and are transcribed by host cell RNA polymerase II to yield pre-early mRNAs. The pre-early proteins promote the transcription of early mRNAs, whose proteins promote viral DNA replication. The late proteins encode structural proteins of the virion.

The dsDNA genomes of the herpesviruses, which are not coated with histones in the virion, are transported to the infected cell's nuclear pores and released into the nucleus. The naked DNA is rapidly loaded with histones bearing silencing modifications by host cell mechanisms; however, a viral enhancer and a virion protein that uses host enzymes to drive chromatin reorganization allow immediate-early gene transcription and expression. Immediate-early proteins promote early gene transcription. Among the E proteins, eight or nine viral proteins including the viral DNA polymerase are essential for viral DNA synthesis. Late genes then encode proteins for virion assembly.

In contrast, the poxviruses replicate entirely in the cytoplasm—an unusual site for replication of a dsDNA virus. As a result, they encode many of the enzymes and factors needed for viral transcription and genome replication. A virus-encoded, virion-associated, DNA-dependent RNA polymerase transcribes the viral genome in the infected cell's cytoplasm to yield early mRNAs. The early mRNAs encode additional transcription factors and DNA replication factors, including a viral DNA polymerase. After DNA replication, the full set of viral proteins needed for viral progeny assembly is generated by intermediate and late transcription.

Single-Stranded DNA Viruses The ssDNA genomes of the parvoviruses are delivered to the infected cell's nucleus, and host cell enzymes copy the ssDNA into dsDNA. The dsDNA is then transcribed by the cell's RNA polymerase II to yield mRNAs encoding proteins that promote viral DNA replication and assemble progeny capsids. How the parvoviruses deal with host epigenetic silencing mechanisms is not known.

Retroviruses The retrovirus genome consists of two identical positive-strand ssRNA molecules, which are not translated but instead copied into dsDNA by the virion RNA-dependent DNA polymerase or *reverse transcriptase* upon entry into the host cell's cytoplasm. The dsDNA is transported with the reverse transcriptase–integrase complex into the nucleus, where the viral integrase catalyzes the integration of the viral DNA molecule into the host cell's chromosomes to yield the provirus. Transcription of the provirus by host RNA polymerase II yields mRNA for translation of viral proteins and for viral full-length transcripts for assembly of progeny virions.

VIRAL EFFECTS ON THE HOST CELL

Many viruses inhibit cellular macromolecular processes, such as host cell transcription and protein synthesis, in an attempt to optimize their own replication by usurping the host cell's machinery and biochemical precursors. These inhibitory events can lead to cell injury and ultimately to cell death, or necrosis. The effects are often manifest by progressive changes in cell structure, detachment from the substrate and rounding up, and eventual lysis. Collectively, these changes are referred to as the CPE. Cells may detect infection as described below and initiate a pathway called *programmed cell death*, or apoptosis, in an attempt to limit viral infection.

Some viruses induce host cell growth to optimize their own replication or to amplify the host cells. Papovaviruses, papillomaviruses, and adenoviruses induce the cellular S phase to activate functions needed for viral DNA replication. These viruses also target cellular proteins that control cell growth, inactivating or degrading them to allow the cell cycle to progress to the S phase. Studies of the mechanisms of these viral effects on host cells have identified cellular tumor-suppressor genes such as the *p53* and retinoblastoma *pRB* genes. Epstein-Barr virus induces proliferation to amplify its latent-infection host cell, a B cell. However, the viral mechanisms sometimes induce immortalization of a cell that has already undergone or later undergoes the oncogenic transformation leading to a cancer cell. Some retroviruses encode altered versions of host genes that can induce transformation. Collectively, these DNA viruses and retroviruses are called *tumor viruses*.

HOST ANTIVIRAL RESPONSES AND VIRAL ANTAGONISTIC MECHANISMS

Host cells have evolved numerous mechanisms for resisting viral infection. They encode constitutively expressed proteins that inhibit viral replication in a process called *intrinsic resistance*. One well-known host resistance factor is the rhesus macaque Trim5α protein, which inhibits human immunodeficiency virus (HIV) type 1 infection soon after the viral core enters the cytoplasm.

Viruses have in turn evolved mechanisms by which to evade or neutralize resistance factors in cells of their host species. The promyelocytic leukemia (PML) protein and its associated proteins in nuclear domain 10 (ND-10) structures in the nucleus of human cells restrict HSV replication, but HSV has evolved a gene product—infected cell protein 0 (ICP0), an E3 ubiquitin ligase—that promotes the degradation of the PML protein and thwarts this antiviral mechanism. Similarly, IFI16 can restrict HSV infection, but the viral ICP0 protein promotes its degradation. Nevertheless, PML and IFI16 protein expression are increased by interferon (IFN) signaling, and the elevated levels of these interferon-stimulated genes (ISGs) are sufficient to reduce wild-type HSV infection. Thus, during HSV infection, there is a race between cellular IFN and viral ICP0 expression.

■ TYPES OF CELLULAR INFECTIONS

The balance of proviral and antiviral factors in a cell defines whether it is permissive or nonpermissive for viral replication. An infection in which progeny virus is produced is a *productive* infection. If a cell becomes infected but does not die, a virus may establish a *persistent* infection. A *chronic* infection can result if infectious virus is continually produced. An *abortive* infection occurs when infection begins but is not completed. In abortive infections, the cell may (1) die, if enough CPEs are exerted, as described above; (2) undergo oncogenic transformation; or (3) harbor a latent infection in which no infectious virus is found but the virus can reactivate at a later time. Examples of these outcomes are the abortive oncogenic infection of cells by Merkel cell polyomavirus, chronic infection of liver cells by hepatitis B virus, and latent infection of neurons by HSV.

■ STAGES OF INFECTION OF A HOST

The stages of viral infection are (1) entry into the host, (2) primary replication and disease at the site of entry, (3) spread through the host, (4) secondary replication and disease at new sites, (5) persistence or clearance by the host immune response, and (6) transmission or release from the host. Infection of a host can be acute, chronic, or latent.

Entry Keratinized skin cells are not viable and therefore are not good host cells for viral replication. Thus, viruses must enter the host at a mucosal surface (e.g., at oral, respiratory, and nasal sites), through a body opening (e.g., by inhalation or ingestion), or through a break in the skin (e.g., the sites of mosquito or other insect bites). For example, papillomaviruses and HSV enter at breaks in the skin, while Zika and dengue viruses can be introduced via insect bites.

Primary Replication and Disease Viruses replicate at the site of entry into the body (i.e., the primary site of infection), are shed back into the environment, and may cause entry-site disease and/or spread to cause systemic illness. For example, influenza viruses can infect the respiratory mucosa. Noroviruses and rotaviruses can infect epithelial cells in the gastrointestinal tract. Dengue and Zika viruses can infect dendritic cells in the tissues after a mosquito bite. If viral infection injures cells and tissues and causes disease at the entry site, the incubation period between exposure and disease can be as short as 1 or 2 days.

Viral Spread Although some viral infections remain localized at the primary site, others spread from the primary site to secondary sites where the viruses infect new cells and cause disease. This spread may take place through the lymph and the bloodstream (*viremia*). Measles virus, for example, replicates initially in the respiratory epithelium, and infected dendritic cells spread through the lymph to lymph nodes where T cells and monocytes are infected and transmit virus through the bloodstream to organs and lymph nodes throughout the body. Systemic disease can result from the disseminated infection, and viral spread into the skin causes the classic measles rash. The incubation period of 10–14 days from exposure to clinical symptoms reflects the time involved for multiple rounds of viral replication and spread within the body before the classic rash symptoms appear. Similarly, dendritic cells and macrophages infected with dengue virus can travel through the circulatory system and transmit virus to secondary sites where infection and disease can follow.

Alternatively, viral spread may occur via neuronal pathways by transsynaptic spread of virions. Rabies virus spreads transsynaptically from the periphery to the central nervous system to cause encephalitis. HSV-1 causes a primary infection at mucosal surfaces and then enters the axon of a sensory neuron and establishes latent infection in the neuron's cell body. Reactivation usually leads to a recurrent infection at the site of primary infection, but occasionally, the virus can move along nerve tracts to the central nervous system and cause encephalitis.

Host Immune Responses Acute viral infection is blunted by the rapid host innate immune response and then controlled by the later adaptive immune response.

INNATE IMMUNITY The first arm of the host's immune response—the innate immune response—is rapid, with recognition of general patterns of viral molecules but not of specific antigens, whose recognition occurs during the later adaptive response. Using pattern recognition receptors, host cells recognize foreign molecules with patterns contained in microbes—i.e., pathogen-associated molecular patterns (PAMPs). Recognition of the foreign molecules leads to activation of innate signaling pathways that induce the expression of IFNs, cytokines, and other host gene products, including those attributable to IFN-stimulated genes, which serve as antiviral effector molecules. Viral ssRNA is recognized by Toll-like receptor 7 (TLR7) and TLR8, which induce transcription of type I IFN genes and IFN-stimulated genes. IFNs act on the producing cell in an autocrine manner and on surrounding cells in a paracrine manner to induce expression of antiviral genes and to activate antiviral mechanisms. dsRNA is recognized by TLR3, which activates expression of type I IFNs. ssRNA and dsRNA are recognized by retinoic acid–inducible gene I (RIG-I) and melanoma differentiation-associated antigen 5 (MDA5), which induce type I IFN expression. Viral glycoproteins are recognized by TLR2 and TLR4. Viral DNA is recognized by the cytoplasmic cGAS receptor, which activates type I IFN expression, and by the nuclear IFN-inducible protein 16 (IFI16) receptor, which activates IFN expression in some cell types and epigenetic silencing of the viral DNA genome in many cell types. IFI16 can therefore act as a constitutively expressed resistance factor or as an IFN-stimulated gene. Innate responses also direct the induction of the later, more specific adaptive immune responses.

ADAPTIVE IMMUNITY Viral antigens are presented as peptides to both CD4+ and CD8+ T cells by antigen-presenting cells to induce these T cells to develop into antigen-specific T cells. Viral antigens are also presented to B cells, which induce differentiation of antibody-producing B cells. Antibodies can bind to virions and neutralize their infectivity by preventing their binding to receptors, their entry, their uncoating, or other steps in infection (Fig. 195-3). Antibodies can also bind to viral antigens on the surface of virions and infected cells and promote phagocytosis, antibody-dependent cytotoxicity, and complement-mediated lysis. T cells recognize viral peptides bound to major histocompatibility complex molecules on the surface of infected cells and produce cytokines that exert an antiviral effect or activate cell-killing mechanisms. Thus, the host's adaptive immune responses can target either virions or infected cells and can clear infection.

Long-Term Effects of Infection Persistent infections can lead to continued pathology due to the ongoing immune response, but even acute infection can lead to long-term effects on the host. The chronic sequelae following SARS-CoV-2 infection or long COVID brought attention to this puzzling aspect of viral infection, but this type of viral pathogenesis had been long recognized in various forms of *post-acute infection syndromes* (PAIS). These included the chronic symptoms following infectious mononucleosis, or acute Epstein-Barr infection, post-dengue fatigue syndrome, and post-polio syndrome, among others. The persisting symptoms may be the result of (1) persisting viral replication or antigens; (2) activation of autoimmune responses; (3) alteration of endogenous bacteria or viruses; and/or (4) irreparable tissue damage. The large disease burden of long COVID and other forms of PAIS make this a priority for future studies of viral pathogenesis.

VIRAL EVOLUTION

Because viral RNA-dependent RNA polymerases are error-prone and most do not have editing functions, sequence changes are frequently introduced into their genomes. These alterations can lead to populations or swarms of viruses with divergent sequences among a viral population in an individual. Upon drug selection, immune pressure, or host restriction, preexisting variants can emerge as the new major form of a virus. Differences in replicative ability can lead to enrichment

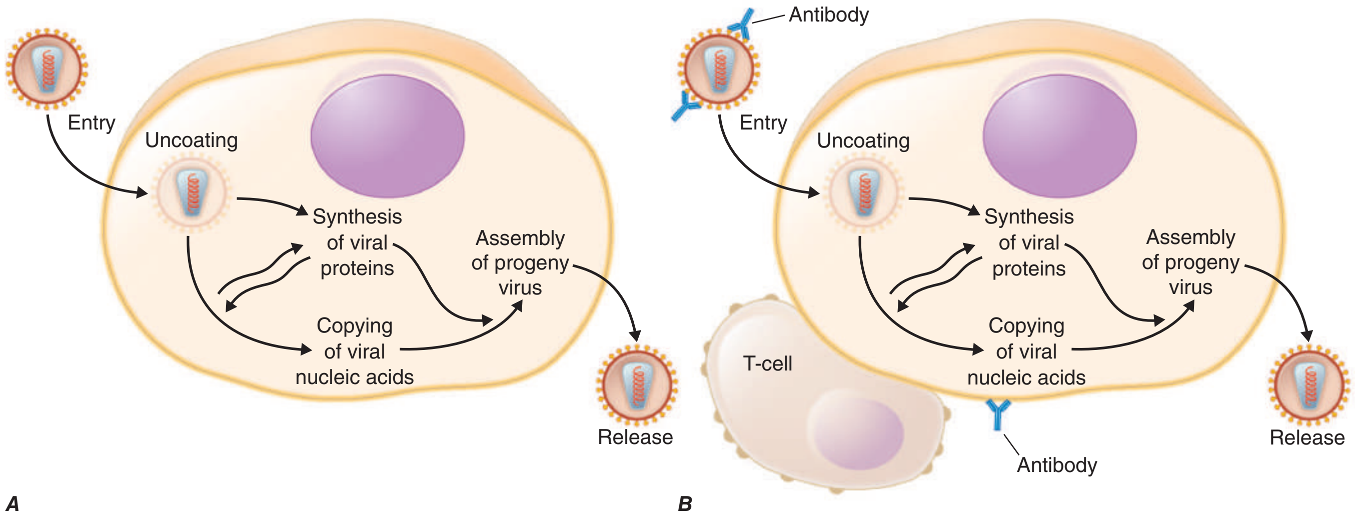
Entry
Uncoating
Synthesis of viral proteins
Assembly of progeny virus
Copying of viral nucleic acids
Release
A
Antibody
Entry
Uncoating
Synthesis of viral proteins
Assembly of progeny virus
T-cell
Copying of viral nucleic acids
Release
Antibody
B

FIGURE 195-3 Steps in viral infection of a host cell and effects of immune effector mechanisms. ***A.*** Steps in viral infection of a host cell. The steps include entry into the cell, uncoating of the viral genomic nucleic acid, synthesis of viral proteins, copying of viral nucleic acids, assembly of progeny virus, egress, and release from the host cell. ***B.*** Mechanisms of immune effector mechanisms. Antibodies can bind to the extracellular virion and neutralize infectivity by preventing binding to the cellular receptor, preventing entry at other steps, preventing uncoating, or preventing other steps of infection. T cells recognize antigenic peptides presented on the surface of infected cells and produce antiviral cytokines and/or activate cell killing.

of more fit viruses and loss of less fit variants. This trend was observed in the COVID-19 pandemic as more fit variants became the dominant forms of SARS-CoV-2 in the population.

Viruses with segmented genomes can undergo genome reassortment in cells co-infected with two viral strains, the result being a new genetic composition for a given virus. For example, new segments can arise in influenza virus isolates thought to be reassortants between the extant human strains and animal or avian strains, such as those from porcine or avian species. This type of event is the cause of the major shifts in influenza viruses that occur periodically over a decade. These major changes due to reassortment and acquisition of a new genome segment are referred to as *antigenic shift*, as opposed to the small changes due to sequence variation, which are designated *antigenic drift*.

Especially in DNA viruses but—under special circumstances—also in RNA viruses such as coronaviruses, viral genomes can undergo recombination between two strains of virus and generate recombinant genomes with new combinations of genes that may be more or less fit.

Viral variants can acquire the ability to infect cells of new host species or to jump species barriers. *Zoonotic* infection occurs when a virus spreads from animals to humans, as is thought to have occurred with both SARS-CoV-1 and SARS-CoV-2. The original viral ancestor of these viruses—probably endemic in bats—is thought to have spread to other animals sold in the markets of China, and viral variants then arose that could efficiently infect humans. Evolution of variants that could efficiently infect and be transmitted by humans as agents of respiratory infection led to the COVID-19 pandemic.

MOLECULAR EPIDEMIOLOGY OF VIRUSES

Several molecular techniques allow easy genotyping of virus isolates. Direct sequencing, analysis of polymorphisms in restriction endonuclease cleavage sites, and polymerase chain reaction (PCR) analysis allow a search for genotypic markers in isolates, with sequencing being the most precise definition of a viral strain. When these types of tests are applied, some viruses (e.g., influenza virus and measles virus) are found to have mainly one strain prevalent in the population at a given time. Thus, only one virus strain spreads through the population. For other viruses, such as HIV or HSV, nearly every unrelated isolate can be differentiated by these tests, and many strains are latent and spreading within the population and are evolving in parallel. With these molecular techniques, genotypic markers can be used to determine whether a virus has been transmitted from one individual to another.

Genomic sequencing studies of SARS-CoV-2 have identified a number of major strains circulating at any given time. As new variants have arisen, each has become the dominant circulating strain.

DETECTION AND QUANTIFICATION OF VIRUSES

Viruses and viral infections need to be detected and quantified for both clinical and scientific purposes. Diagnostic virology employs the scientific principles described above to detect viruses and evidence of infection in clinical samples, to define the type of virus present in a sample, and in some cases to quantify the amount of virus or the viral load in a patient. Scientific studies use these principles for detection and quantification of viruses in laboratory stocks and for measurement of viral replication.

■ DETECTION OF INFECTIOUS VIRUS

Biologic assays must be used to detect and measure infectious virus. Infectivity can be measured as either the ability to infect animals and cause disease or the ability to infect cultured cells and cause CPE. For example, SARS-CoV-1 virus was first isolated by the introduction of an oropharyngeal swab sample into Vero cell cultures and detection of CPE.

■ DETECTION OF VIRAL PARTICLES, THEIR COMPONENTS, AND VIRAL GENE PRODUCTS

Viral Particles Electron microscopy (EM) must be used to visualize virions directly, because viruses (other than the poxviruses) are smaller than the resolution of the light microscope. Virions can be visualized by EM with negative staining of the virions themselves or by transmission EM of infected cells. As stated above, SARS viral particles were first visualized in sections of Vero cells infected with samples from patients. The cell culture supernatant showed coronavirus particles by negative-staining EM. The latter method has also been used to detect viral particles in stool during outbreaks of gastroenteritis. Antibodies specific for viral capsid proteins are often used in this assay to concentrate the virus and enhance its detection.

Viral Nucleic Acids Viral nucleic acids are detected by amplification methods involving PCR with specific primers, which amplifies very small numbers of viral nucleic acid molecules. These methods can use direct amplification of DNA in clinical samples to detect and quantify viral DNA genomes; alternatively, they can use reverse transcription of RNA followed by PCR to detect a DNA product in clinical samples as a means to detect viral RNA sequences. Multiple primers can be used in a multiplex reaction to detect multiple pathogens. The process of nucleic acid isolation, reverse transcription, and PCR has been automated, and high-throughput instruments measure the HIV load in serum samples. HSV-1 DNA can be measured in cerebrospinal fluid as a rapid assay for HSV encephalitis. These methods have also been transferred to rapid assays for point-of-care detection of viral genomes.

Viral Antigens Viral antigens can be detected by immunologic methods such as immunofluorescence and enzyme immunosorbent assay (EIA). Immunofluorescence involves fixation and permeabilization of cells or tissues from clinical specimens and reaction with either (1) an antiviral antibody conjugated to a fluorophore (direct immunofluorescence) or (2) an antiviral antibody followed by an anti-immunoglobulin antibody conjugated to a fluorophore (indirect immunofluorescence), with detection of the fluorophore by fluorescence microscopy in either case.

The EIA entails the immobilization of an antiviral antibody on a substrate such as a microtiter well, incubation of the patient's sample in the well, and further incubation with an antibody linked to an enzyme. The bound enzyme is then measured by production of a colored substrate that can be read spectrophotometrically or detected in a rapid antigen test kit.

Hemagglutination Some viruses have the ability to cross-link and agglutinate red blood cells of specific species, a process called *hemagglutination*. Viral titer is measured by the inverse of the last dilution of the sample that causes hemagglutination.

Quantitative Assays of Viruses Viruses can be quantified in terms of virion particle numbers and/or infectivity. The number of virion particles in a sample can be determined by negative staining and observation by EM. The numbers of viral DNA genomes can be determined by PCR, and RNA genomes can be determined by reverse transcriptase PCR (RT-PCR), as described above. Alternatively, purified viral particles can be quantified biochemically by spectrophotometric assays that measure viral protein.

The number of infectious particles can be quantified by an endpoint dilution assay in which the virus is diluted until only one-half of cultures are infected; this concentration is designated the tissue culture infectious dose for 50% of cultures, or $TCID_{50}$. An alternative assay can determine at what dose one-half of experimental animals die of viral disease (lethal dose for 50% of test animals, or LD_{50}). A more quantitative assay of infectivity is the plaque assay. A plaque is an area of visualized localized CPE. In the plaque assay, dilutions of the virus sample are placed on cells attached to a culture dish, and after adsorption of the virus to cells, the cells are overlaid with semisolid medium or medium containing antibody, which prevents virus diffusion through the medium. Virus then spreads only cell to cell, causing a restricted area of CPE—a plaque—on the cellular monolayer. The number of plaques formed by each dilution of virus defines the titer in plaque-forming units (PFUs) per volume of virus stock.

For viruses that infect humans, the ratio of viral particles to infectious units, or the particle-to-PFU ratio, is always much greater than 1—usually 10–1000. This result signifies a large excess of particles that are defective and/or that do not score as infectious in laboratory assays. Thus, for experimental purposes, following input virus particles, either visually or biochemically, does not guarantee that the observer is following the real infection pathway. In addition, clinical preparations of viruses used for vaccines, vaccine vectors, gene therapy vectors, and oncolytic viruses need to be defined precisely and specifically in terms of particles versus infectious units for accurate and safe dosing. As an example, a recent adenovirus-based COVID vaccine was quantified on the basis of spectrophotometric measurement of purified virions. After the trial was initiated, lower than expected immune responses led to a reexamination of the vaccine dose. An excipient discovered in the vaccine was found to cause errors in spectrophotometric measurement that led to an overestimate of the virus concentration. Parallel measurements of viral genomes with RT-PCR allowed a more accurate measurement of the vaccine vector batches, and the dose was revised to one-half of the original level. This example illustrates the importance of precise measurements of viral particles and infectious particles in preparations of viruses for clinical use.

DETECTION OF VIRUS-SPECIFIC ANTIBODIES

The presence of virus-specific antibodies provides evidence of prior infection with a virus or prior exposure to viral antigens through immunization; thus, antibody tests are extremely important clinically. The most common tests for antibodies are the enzyme-linked immunosorbent assay (ELISA) and the Western blot or immunoblot assay. An ELISA involves the immobilization of viral antigen on a substrate such as a microtiter well, its incubation with the patient's serum, and further incubation with an antibody to human IgG coupled to an enzyme. The amount of bound antibody is measured by detection of a colored product made by the bound enzyme. The Western blot assay involves the resolution of viral proteins in a polyacrylamide gel, their transfer to a membrane, incubation with the patient's serum, and further incubation with antibody to human IgG coupled to an enzyme. Proteins with bound antibodies are detected as a colored product made by the bound antibody. The Western blot detects antigen of a specific size and therefore is more specific than ELISA. For example, HIV serologic testing involves high-throughput ELISA screening followed by a Western blot assay to confirm the specificity of any positive ELISA result.

In a hemagglutination inhibition assay, antibodies specific for viral surface proteins are detected by their ability to block hemagglutination.

IMMUNIZATION AGAINST VIRAL DISEASES

Viral vaccines are among the most effective biomedical and public health measures that have been implemented: millions of deaths have been prevented by their use. These vaccines are safe because extensive protocols have been developed for monitoring vaccine safety both before and after licensure. Historically, viral vaccines were based on either inactivated virus or live attenuated viruses, as exemplified by the Salk polio vaccine and the Sabin live attenuated polio vaccine, respectively. Both of these vaccines were quite successful, offering individual advantages. Further vaccine types have been developed, including those based on recombinant proteins, viral vectors, and, most recently, mRNA. For each virus, the optimal antigen and immunization strategy must be developed on the basis of the virus-specific immune correlates, antibodies, or T cells needed for immunologic protection against infection and disease. These concepts are discussed in greater detail in Chap. 129.

ANTIVIRAL THERAPEUTICS

ANTIVIRAL DRUGS

Viruses replicate in human cells and use much of the host cell's machinery. Therefore, antiviral drugs must target virus-specific events to optimize safety. Viral targets for drugs have been identified in studies of the mechanisms of viral infection and replication **(Chap. 196)**. Many of the most successful antiviral drugs target viral enzymes; examples include the anti-HSV drugs that target the virus DNA polymerases and thymidine kinase **(Chap. 196)** and the HIV drugs that target the virus reverse transcriptase, protease, and integrase **(Chap. 208)**.

VIRUSES AS THERAPEUTICS

Viruses have been engineered for a number of medical purposes, including gene delivery and tumor cell killing. As described above, viruses have been developed as vaccines and vaccine vectors. For example, vesicular stomatitis virus–based vectors have been employed as Ebola vaccines. Adenovirus-based vectors have been used as AIDS vaccine vectors and have been used as COVID-19 vaccine vectors. Viral recombinants, including those of retroviruses and adeno-associated viruses, have been approved as vectors for delivery of genes to cells for treatment of single-gene defects. Retroviruses integrate into the cell's chromosomes and are maintained with stable expression of the transgene, although some concerns have arisen about possible activation of neighboring promoters and adverse effects due to that activation. Adeno-associated viruses are not integrated but are stably maintained and capable of durable expression of the transgene. Adenoviruses and herpesviruses are also being tested as gene therapy vectors. Finally, an attenuated strain of HSV expressing granulocyte-macrophage colony-stimulating factor has been approved for treatment of melanoma because of its oncolytic and immunotherapeutic properties. Many additional studies are assessing viruses for use as vectors and for immunotherapeutic and oncolytic applications.

SUMMARY

As obligate intracellular parasites, viruses enter host cells, replicate, and spread in the form of progeny viruses. Injury to the host cell resulting from viral entry may lead to tissue and organ damage. Basic knowledge of the mechanisms underlying infection by and replication of viruses that infect humans is the foundation for medical studies of viral pathogenesis, viral vaccines, antiviral drugs, and the use of viruses as therapeutics. There are many interactions between different viruses; thus, a broad knowledge of all viruses is essential to our preparedness for the next viral epidemic or pandemic.

FURTHER READING

Choutka J et al: Unexplained post-acute infection syndromes. Nat Med 28:911, 2022.

Howley PM et al (eds): *Fields Virology: Vol. 4: Fundamentals*, 7th ed. Philadelphia, Wolters Kluwer/Lippincott Williams & Wilkins Health, 2023.

Knipe DM et al: Ensuring vaccine safety. Science 370:1274, 2020.

Ksiazek TG et al: A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 348:1953, 2003.

Sodroski CN, Knipe DM: Nuclear interferon-stimulated gene product maintains heterochromatin on the herpes simplex viral genome to limit lytic infection. Proc Natl Acad Sci USA 120:e2310996120, 2023.

Voysey M et al: Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: An interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. Lancet 397:99, 2021.

Zhou P et al: A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270, 2020.

196

Antiviral Chemotherapy, Excluding Antiretroviral Drugs

Jeffrey I. Cohen, Marc G. Ghany

Most antiviral drugs inhibit viral DNA or RNA replication, but other activities, such as virus entry, viral RNA transcription, cleavage of proteins by the viral protease, virus uncoating after infection, and virus release from cells, are all targeted by different licensed antiviral agents. Inhibition of viral replication does not eliminate the virus in the cell; host cell immune responses are important for viral clearance. Antiviral drugs usually do not eradicate latent viral infections but instead inhibit viral replication; thus, when treatment is stopped, the virus can reactivate and replicate again. Resistance to antiviral agents due to mutations in viral proteins is not uncommon and is more common for RNA viruses with a higher mutation rate than for DNA viruses. This difference may explain the observation that drug-resistant DNA viruses are a greater problem in immunocompromised patients, whereas drug-resistant RNA viruses can be found in healthy persons as well. Patients may harbor a mixture of drug-resistant and drug-sensitive viruses that is dynamic and changes under pressure from the drug. Combination therapy with more than one antiviral agent, each with a different mechanism of action, may be more effective than monotherapy, particularly against RNA viruses, which may be present as mixtures with different resistance patterns. Antiviral testing can be performed in patients who do not respond to antiviral drugs or whose response diminishes. For some viruses, such testing involves the sequencing of selected viral genes; however, in many cases, it involves the growth of virus in the presence of different concentrations of the drug, which is a laborious, time-consuming process. Response to antiviral therapy has traditionally been assessed clinically, but quantitative PCR has been useful in monitoring the response to therapy for viruses that circulate in the blood (e.g., cytomegalovirus [CMV], hepatitis B and C viruses [HBV and HCV, respectively]). Systemic therapy with antivirals is usually more effective than topical therapy but is more commonly associated with side effects.

ANTIVIRAL DRUGS FOR HERPESVIRUS INFECTIONS

■ ACYCLOVIR, VALACYCLOVIR, FAMCICLOVIR, AND PENCICLOVIR

Acyclovir is an analogue of deoxyguanosine and is phosphorylated to the monophosphate form by viral thymidine kinase in cells infected with herpes simplex virus (HSV) or varicella-zoster virus (VZV). Cellular protein kinases further phosphorylate the drug to the active triphosphate form, which inhibits viral DNA polymerase; the drug is incorporated into viral DNA to terminate its replication. Valacyclovir, a valine ester of acyclovir, is absorbed much better than acyclovir; its rapid conversion to acyclovir in the liver and intestine results in plasma acyclovir levels approximately four times higher than are attained with oral acyclovir. Acyclovir and valacyclovir are approved by the U.S. Food and Drug Administration (FDA) for treatment of initial episodes of genital herpes, recurrent genital herpes, varicella, and zoster **(Table 196-1)**. Valacyclovir is also approved for treatment of herpes labialis (cold sores), for suppression of recurrences of genital herpes, and for reduction of transmission of genital HSV. The doses of acyclovir and valacyclovir used for treating VZV infections are higher than those used for HSV infections since VZV is less susceptible to inhibition by these drugs. Both drugs exhibit poor activity against CMV. Intravenous acyclovir is used for severe disease requiring hospitalization; oral acyclovir or valacyclovir is used for outpatient therapy; and topical acyclovir, penciclovir, and docosanol are approved for treatment of orolabial herpes but are much less effective than the oral drugs.

Acyclovir is excreted by the kidneys. Thus the dose of acyclovir or valacyclovir needs to be reduced with renal insufficiency. Central nervous system (CNS) side effects that occur with IV acyclovir or oral valacyclovir are more common with the higher drug levels seen in persons with renal insufficiency. Reversible renal insufficiency due to crystallization of the drug in renal tubules can occur with IV acyclovir, especially in persons who are dehydrated. Headache, nausea, rash, and diarrhea have been reported with acyclovir. Mutations in the HSV or VZV thymidine kinase or, less commonly, in viral DNA polymerase can result in resistance to acyclovir or valacyclovir. Viruses lacking thymidine kinase activity are also resistant to famciclovir and ganciclovir. Acyclovir- and valacyclovir-resistant HSV and VZV are rare in immunocompetent persons. Resistant virus is treated with foscarnet or, less commonly, cidofovir. Mucosal disease due to resistant virus in immunocompromised persons is sometimes treated with topical foscarnet, trifluridine, or cidofovir.

Famciclovir is a diacetyl ester of penciclovir that is converted to penciclovir in the intestine and liver. Penciclovir is a guanosine analogue that is less potent than acyclovir, but, because of its longer intracellular half-life, its activity is similar to that of acyclovir. Penciclovir is phosphorylated by HSV and VZV thymidine and cellular kinases and has activity similar to that of acyclovir for HSV and VZV infections. Famciclovir is approved for treatment of zoster, suppression of genital herpes, and treatment of recurrent mucocutaneous herpes in patients with HIV infection. Famciclovir is excreted by the kidneys, and the dose is adjusted for renal insufficiency. Side effects are uncommon and can include headache, nausea, and diarrhea. Resistance due to mutations in viral thymidine kinase or DNA polymerase can occur.

Oral acyclovir reduces the duration of pain and other symptoms, time to healing, and shedding in patients with their first episode of genital herpes when treatment is begun within 6 days of infection. Acyclovir, valacyclovir, and famciclovir are all effective for treatment of primary and recurrent genital and orolabial herpes as well as for suppressive therapy for these conditions. Topical acyclovir cream reduces shedding and time to healing by 1–2 days if given within 1 day of symptom onset in persons with recurrent genital or orolabial herpes. Oral acyclovir or valacyclovir reduces the severity of varicella when given within 1 day of onset of the rash. Oral acyclovir, famciclovir, or valacyclovir shortens the duration of pain and rash associated with zoster if given within 3 days of onset. Oral valacyclovir is more effective than oral acyclovir and is generally preferred since it has better oral bioavailability and does not need to be given as frequently. Suppressive valacyclovir therapy for genital herpes reduces transmission to uninfected partners by 50%. Intravenous acyclovir is used for herpes encephalitis and disseminated HSV or VZV disease.

■ GANCICLOVIR AND VALGANCICLOVIR

Ganciclovir is a deoxyguanosine analog that is phosphorylated by UL97 protein kinase in cells infected with CMV and converted to its active form, ganciclovir triphosphate, by cellular protein kinases. Ganciclovir triphosphate inhibits both viral DNA polymerase and incorporation of guanosine triphosphate into viral DNA. Valganciclovir is a valine ester of ganciclovir and is converted to ganciclovir in the liver and intestine. Valganciclovir has much better oral bioavailability than ganciclovir; plasma levels of oral valganciclovir and IV ganciclovir are similar. Ganciclovir and valganciclovir are used for treatment and prevention of CMV disease in immunocompromised patients and are approved for prevention of CMV infection in transplant recipients and for treatment of CMV retinitis. Ganciclovir is effective against HSV, VZV, human herpesvirus type 6 (HHV-6), and herpes B virus. This drug is excreted by the kidneys, and dose adjustment is required in renal insufficiency. Ganciclovir therapy often results in neutropenia and thrombocytopenia after 1 week. Less commonly, ganciclovir has been associated with CNS symptoms, particularly at high plasma drug levels. Mutations in CMV UL97 protein kinase or, less commonly, UL54 viral DNA polymerase can result in resistance to ganciclovir or valganciclovir. CMV with mutations in protein kinase is usually sensitive to foscarnet and cidofovir, while CMV with mutations in both protein kinase and DNA polymerase is usually sensitive only to foscarnet. Mutations are more common among persons who are highly immunocompromised and who have been taking the drug for a long time. Resistant virus is treated with foscarnet or cidofovir.

Ganciclovir and valganciclovir are used for treating severe CMV infections in immunocompromised patients, including colitis, pneumonitis, retinitis, and encephalitis. Induction therapy, given two or three times daily, is usually followed by less frequently administered maintenance therapy. Oral valganciclovir has activity similar to that of intravenous ganciclovir. Ganciclovir and valganciclovir are used for prevention of CMV infection in transplant recipients when given either preemptively (on the basis of viremia) or prophylactically. Ganciclovir reduces developmental delay in infants with congenital CMV disease involving the CNS and reduces hearing loss in infants with asymptomatic congenital CMV infection. Ganciclovir and valganciclovir are used for treatment of HHV-6 encephalitis, HHV-8–associated Castleman

TABLE 196-1 Antiviral Drugs for Herpesvirus Treatment and Prophylaxis in Adults

DISEASE	DRUG	ROUTE	ADULT DOSE	COMMENTS
Orolabial herpes, primary episode	Acyclovir	Oral	400 mg tid × 7–10 d	Reduces duration of fever, lesions, and virus shedding
	Valacyclovir	Oral	1 g bid × 7–10 d	
	Famciclovir	Oral	500 mg bid or 250 mg tid × 7–10 d	
Orolabial herpes, recurrence	Acyclovir	Oral	400 mg 5 times daily × 5 d	Reduces duration of lesions by 1–2 d if given during prodrome
	Valacyclovir	Oral	2 g bid × 1 d	
	Famciclovir	Oral	1500 mg × 1 d	
Orolabial herpes, suppression	Acyclovir	Oral	400 mg bid	In patients with >6 recurrences per year, reduces number of recurrences by ~50% and increases time to first recurrence
	Valacyclovir	Oral	500 mg or 1 g once daily	
	Famciclovir	Oral	500 mg bid	
Genital herpes, primary episode	Acyclovir	Oral	400 mg tid or 200 mg 5 times daily × 7–10 d	Reduces duration of symptoms, genital lesions, and virus shedding by 2, 4, and 7 d, respectively
	Valacyclovir	Oral	1 g bid × 7–10 d	
	Famciclovir	Oral	250 mg tid × 7–10 d	
Genital herpes, recurrence	Acyclovir	Oral	800 mg tid × 2 d or 400 mg tid × 5 d	Reduces duration of symptoms, genital lesions, and virus shedding by 1–2 d
	Valacyclovir	Oral	500 mg bid × 3 d or 1 g daily × 5 d	
	Famciclovir	Oral	500 mg once, then 250 mg bid × 2 d	
Genital herpes suppression	Acyclovir	Oral	400 mg bid	In patients with >6 recurrences per year, reduces recurrence rates from 80–85% to 25–30%, reduces virus shedding and transmission
	Valacyclovir	Oral	250 mg bid	
	Famciclovir	Oral	500 mg to 1 g daily	
HSV encephalitis	Acyclovir	IV	10–15 mg/kg q8h × 14–21 d	Reduces mortality and sequelae
HSV keratitis	Acyclovir	Topical	3% ophthalmic ointment, 5 times daily	Shortens duration of disease; acyclovir better tolerated, especially with prolonged treatment
	Trifluridine	Topical	1% ophthalmic solution, 1 drop q2h when awake (9 drops daily max)	
	Vidarabine	Topical	3% ointment, 0.5-inch ribbon 5 times daily	
Mucocutaneous herpes in immunocompromised patient	Acyclovir	IV	5 mg/kg q8h × 7–14 d	IV acyclovir reduces time to healing, duration of pain, and duration of virus shedding
	Valacyclovir	Oral	500 mg to 1 g bid × 7–10 d	
	Famciclovir	Oral	500 mg bid × 7–10 d	
Varicella	Acyclovir	Oral	20 mg/kg (800 mg max) 5 times daily × 5 d	Has modest effect on symptoms, reduces fever duration by 1 day
	Valacyclovir	Oral	20 mg/kg (1 g max) tid × 5 d	
Zoster	Acyclovir	Oral	800 mg 5 times daily × 7 d	Reduces time for last new lesion formation, virus shedding, and pain duration
	Valacyclovir	Oral	1 g tid × 7 d	
	Famciclovir	Oral	500 mg tid × 7 d	
Varicella or zoster, disseminated	Acyclovir	IV	10 mg/kg q8h × 7 d	Reduces time for last new lesion formation and virus shedding; reduces cutaneous dissemination
Cytomegalovirus disease	Ganciclovir	IV	5 mg/kg q12h × 14–21 d, then 5 mg/kg daily (maintenance dose)	Neutropenia and thrombocytopenia common after 1 week
	Valganciclovir	Oral	900 mg bid × 14–21 d, then 90 mg daily (maintenance dose)	Levels and side effects similar to ganciclovir
	Foscarnet	IV	60 mg/kg q8h × 14–21 d, then 90–120 mg daily (maintenance dose)	Nephrotoxicity, electrolyte abnormalities; give with additional saline
	Cidofovir	IV	5 mg/kg once weekly twice, then every other week	Nephrotoxicity; give with probenecid and saline
	Maribavir	Oral	400 mg bid	Used for disease refractory to ganciclovir, foscarnet, or cidofovir Antagonizes activity of ganciclovir Numerous drug interactions
Cytomegalovirus prophylaxis	Letermovir	Oral	480 mg qd	Numerous drug interactions
		IV	480 mg qd	Given 100 days after stem cell transplant, 200 days after kidney transplant

disease in patients with poorly controlled HIV infection, and severe HSV or VZV disease when acyclovir is unavailable.

FOSCARNET

Foscarnet is a pyrophosphate analogue that directly inhibits herpesvirus DNA polymerases by blocking the pyrophosphate binding site in the enzyme. Foscarnet does not require additional phosphorylation (unlike acyclovir, cidofovir, or ganciclovir) in virus-infected cells for its activity. This drug is approved for treatment of CMV retinitis and mucocutaneous acyclovir-resistant HSV disease. It is also used to treat ganciclovir-resistant CMV and acyclovir-resistant VZV. Foscarnet is given intravenously and is excreted by the kidneys; dose adjustment is required in renal insufficiency. Up to one-third of patients receiving foscarnet develop nephrotoxicity with elevated levels of creatinine and blood urea nitrogen, and proteinuria. Renal tubular acidosis and interstitial nephritis also have been reported. Renal insufficiency is more common among persons who are dehydrated, given other nephrotoxic drugs, or given high doses or rapid infusions of foscarnet. Administering IV saline before and after each foscarnet dose and giving the drug over an adequate period can reduce nephrotoxicity. Renal insufficiency is often reversible after treatment when the drug is stopped. Other side effects include hypomagnesemia and hypocalcemia, which can be associated with arrhythmias, paresthesias, and seizures. Other metabolic abnormalities include hypokalemia, hypophosphatemia, or hyperphosphatemia. Foscarnet can also cause headache, fever, rash, diarrhea, acute dystonia, tremors, hemorrhagic cystitis, genital ulcerations, anemia, and abnormal liver function values. Mutations in CMV DNA polymerase (UL54) or in HSV or VZV DNA polymerase can result in resistance to foscarnet. CMV, HSV, and VZV can become resistant to foscarnet; some strains of CMV are resistant to foscarnet, ganciclovir, and cidofovir; and HSV can become resistant to acyclovir and foscarnet. Foscarnet is typically used to treat CMV retinitis, HHV-6 encephalitis, or drug-resistant severe CMV, HSV, or VZV infections in immunocompromised patients. Topical foscarnet has been used to treat acyclovir-resistant mucosal infections due to HSV.

■ CIDOFOVIR

Cidofovir is an analogue of deoxycytidine monophosphate and is phosphorylated in cells to its active diphosphate form. The diphosphate form of cidofovir competes with deoxycytidine triphosphate for incorporation into herpesvirus DNA. The drug inhibits replication of all human herpesviruses as well as poxviruses, papillomaviruses, polyomaviruses, and adenoviruses. Cidofovir is approved for treatment of CMV retinitis in patients with AIDS; it is also used for treatment of infections caused by CMV exhibiting ganciclovir resistance due to mutations in UL97 protein kinase and those caused by HSV or VZV displaying mutations in thymidine kinase. Because cidofovir is excreted by the kidneys, dose adjustment is required in renal insufficiency. About one-fifth of patients receiving cidofovir develop nephrotoxicity, and the drug is associated with metabolic acidosis and glucosuria. Cidofovir therapy is preceded by at least 1 L of saline, and probenecid is given 3 h before, 2 h after, and 8 h after each dose to reduce nephrotoxicity. An additional 1 L of saline is recommended during treatment or immediately thereafter. About one-fourth of patients receiving cidofovir develop neutropenia; additional side effects include ocular hypotony, uveitis, iritis, headache, nausea, vomiting, diarrhea, and rash. Mutations in CMV DNA polymerase (UL54) or HSV DNA polymerase can result in resistance to cidofovir. Some strains of CMV exhibiting ganciclovir resistance due to mutations in viral DNA polymerase are resistant to cidofovir, whereas many CMV and HSV strains exhibiting foscarnet resistance due to mutations in DNA polymerase may retain sensitivity to cidofovir. Cidofovir is typically used to treat ganciclovir- and/or foscarnet-resistant severe CMV disease or acyclovir- and/or foscarnet-resistant HSV disease in immunocompromised patients. Cidofovir has been used as preemptive therapy against CMV infection in transplant recipients. It has also been used to treat severe adenovirus infections, adenovirus or BK virus hemorrhagic cystitis, BK nephropathy, and severe molluscum contagiosum, although controlled studies have not been performed. Topical cidofovir has been used to treat acyclovir-resistant HSV mucosal infections and anogenital warts.

■ LETERMOVIR

Letermovir is a dihydroquinazolin that inhibits the CMV DNA terminase complex (UL51, UL59), which is required for cleavage and packaging of CMV into nucleocapsids. The drug has no activity against other human herpesviruses. Letermovir is approved for prophylaxis of CMV infection and disease in adult CMV-seropositive recipients of an allogeneic hematopoietic stem cell transplant or donor CMV-positive, recipient CMV-negative kidney transplant recipients. Letermovir is metabolized by the liver and excreted in the feces; dose adjustment is not required if the creatinine clearance rate (CrCl) is >10 mL/min. The dose of letermovir must be decreased in persons taking cyclosporine. Letermovir therapy results in reduced levels of voriconazole and increased levels of sirolimus, tacrolimus, cyclosporine, and other drugs metabolized by CYP2C8 or transported by OAT1B1/3. Side effects of letermovir include headache, nausea, diarrhea, and peripheral edema. Letermovir does not cause nephrotoxicity and is not myelosuppressive. Resistance to letermovir occurs more frequently in vitro than resistance to ganciclovir or foscarnet, and clinically significant letermovir resistance due to mutations in UL56 in patients with CMV disease has been reported; resistance may be less common when the drug is used for prophylaxis in patients with low or undetectable CMV levels. When given to CMV-seropositive patients, starting a median of 8 days after hematopoietic stem cell transplantation and continuing for 14 weeks, letermovir reduced the incidence of clinically significant CMV infection by 38% compared with placebo. While anecdotes describe the use of letermovir for treatment of CMV disease, resistance may develop quickly.

■ MARIBAVIR

Maribavir is a benzimidazole that inhibits the CMV UL97 protein kinase and CMV replication, and reduces the egress of viral particles from the nucleus. Maribavir is approved for treating adults and children with posttransplant CMV infection/disease that is refractory to treatment (with or without proven genotype resistance) with ganciclovir, cidofovir, or foscarnet. Since maribavir inhibits the UL97 protein kinase, it can antagonize ganciclovir and should not be given with the latter drug. Resistance to maribavir has been reported, and such CMV strains are also resistant to ganciclovir. Maribavir can increase concentrations of several drugs including tacrolimus, sirolimus, everolimus, and cyclosporine. The most common side effect of maribavir is taste disturbance.

■ TRIFLURIDINE AND VIDARABINE

Trifluridine is a thymidine analogue that is incorporated into viral DNA and inhibits its synthesis. Vidarabine is approved for topical therapy of herpes keratitis and has also been used topically to treat acyclovir-resistant mucosal HSV infections. Trifluridine is active against acyclovir-resistant HSV, CMV, and vaccinia virus. Vidarabine is an adenosine analogue that is incorporated into viral DNA and inhibits viral DNA polymerase. Both trifluridine and vidarabine are used for topical therapy only.

■ INVESTIGATIONAL AND OTHER AGENTS

Brincidofovir is a phospholipid conjugate of cidofovir that is rapidly taken up by cells and converted into cidofovir. It is active against herpesviruses (including most strains of ganciclovir-resistant CMV), poxviruses, adenovirus, and polyomaviruses. It does not cause nephrotoxicity and is not myelosuppressive. Diarrhea is the most common side effect. The drug has been associated with intestinal toxicity and acute graft-versus-host disease of the gastrointestinal tract. The drug did not meet its primary endpoints in trials for adenovirus disease or CMV prophylaxis. At the time of this writing, it is not available as part of an expanded access program. An intravenous formulation that, it is hoped, will cause less gastrointestinal toxicity is being tested for adenovirus viremia.

Pritelivir inhibits the helicase–primase complex required for replication of HSV. This drug has reduced viral shedding in patients with recurrent genital herpes and is being tested for use against acyclovir-resistant HSV mucocutaneous infection. Pritelivir is available as an expanded access drug for acyclovir-resistant HSV infection.

Amenamevir is a helicase–primase inhibitor under development for HSV and VZV infections.

ANTIVIRAL DRUGS FOR RESPIRATORY VIRUS INFECTIONS

■ INFLUENZA

Neuraminidase Inhibitors Oseltamivir, zanamivir, and peramivir are neuraminidase inhibitors that inhibit cleavage of sialic acid, which is required for the release of influenza virus from infected cells and its spread to other cells.

Oseltamivir phosphate is an oral prodrug that is cleaved by esterases in the liver, gastrointestinal tract, and blood to oseltamivir carboxylate, the more active form. It is approved for treatment of uncomplicated influenza A or B disease when given ≤48 h after symptom onset and for prophylaxis of influenza A and B in persons ≥1 year of age (**Table 196-2**). Oseltamivir is much less active against influenza B than against influenza A. The drug is excreted by the kidneys, and the dose is adjusted in renal insufficiency. The most common side effects are nausea, abdominal pain, and vomiting. Although CNS side effects have been reported, particularly in children, it is unclear whether they are due to the drug or to influenza virus infection itself. Resistance to oseltamivir can develop as a result of mutations in the viral neuraminidase or in the hemagglutinin. Oseltamivir-resistant virus has been transmitted from person to person. Resistance has been reported in ~15% of healthy children and ~1% of adults; resistance is more common among immunocompromised persons.

TABLE 196-2 Antiviral Drugs for Respiratory Virus Treatment and Prophylaxis in Adults

DISEASE	DRUG	ROUTE	ADULT DOSE	COMMENTS
Influenza A, B	Oseltamivir	Oral	Treatment: 75 mg bid × 5 d Prophylaxis: 75 mg/d	Shortens duration of symptoms by 1 d when given within 2 d of onset; reduces complications; considered drug of choice for patients with complications of influenza
Influenza A, B	Zanamivir	Inhaled	Treatment: 10 mg bid × 5 d Prophylaxis: 10 mg/d	Shortens duration of symptoms by 1–2 d when given within 2 d of onset; requires patient training for use; can cause bronchospasm; not recommended for persons with asthma or chronic obstructive pulmonary disease
Influenza A, B	Peramivir	IV	600 mg once	Shortens duration of symptoms by 1–2 d when given within 2 d of onset
Influenza A, B	Baloxavir	Oral	Treatment or postexposure prophylaxis: 40 mg once; if >80 kg, 80 mg once	Shortens duration of symptoms by 1 d when given within 2 d of onset; active against virus resistant to neuraminidase inhibitors
Influenza A	Amantadine	Oral	Treatment: 100 mg bid × 5 d Prophylaxis: 200 mg/d	Most influenza virus strains are resistant; use only if virus is known to be sensitive.
Influenza A	Rimantadine	Oral	Treatment: 100 mg bid × 5 d Prophylaxis: 200 mg/d	Most influenza virus strains are resistant; use only if virus is known to be sensitive.
Respiratory syncytial virus	Ribavirin	Inhaled	Aerosol from reservoir containing 20 mg/mL for 12–18 h/d × 3–7 d	Reduces severity of symptoms in hospitalized infants with lower respiratory tract disease; anecdotal reports of reduced progression to lower respiratory tract disease and mortality in stem cell transplant patients
SARS-CoV-2	Nirmatrelvir/ritonavir	Oral	300 mg/100 mg bid × 5 days	Reduces rate of hospitalization by 50% within 30 days after diagnosis
SARS-CoV-2	Remdesivir	IV	200 mg on day 1, then 100 mg qd × 2 d for outpatients, × 4 days for inpatients	Reduces duration of hospitalization in some studies. Duration of treatment extended up to 10 days if no improvement.
SARS-CoV-2	Molnupiravir	Oral	800 mg q12h × 5 days	Recommended for patients when nirmatrelvir or remdesvir are not available, unable to be used, or not appropriate. Approved under EUA.

Abbreviation: EUA, emergency use authorization by the U.S. Food and Drug Administration (FDA).

Zanamivir is approved for treatment of uncomplicated influenza A and B in adults and children ≥7 years of age who have had symptoms for ≤2 days and for prophylaxis in persons ≥5 years of age. Because zanamivir has poor oral bioavailability, it is given as a powder through an inhaler. Thus, use of the drug can be difficult for young children and some elderly patients. Inhalation of zanamivir may cause bronchospasm, particularly in persons with underlying lung disease; it is not recommended for persons with asthma, chronic obstructive pulmonary disease, or other airway disease. Zanamivir is more active than oseltamivir against influenza B. It is also active against some isolates of influenza virus that are resistant to oseltamivir; resistance to zanamivir is less common than that to oseltamivir.

Peramivir is approved for treating uncomplicated influenza in patients ≥2 years of age who have had symptoms for ≤2 days. Because of its long half-life, it is given as a single IV dose. Peramivir is highly active against both influenza A and B. The drug is excreted by the kidneys, and the dose is adjusted in renal insufficiency. The most common side effect is diarrhea. While peramivir-resistant virus is rare in healthy persons, peramivir-resistant virus has been isolated from immunocompromised persons.

Oseltamivir, zanamivir, and peramivir are effective for treatment of uncomplicated influenza A and B, including disease caused by avian influenza viruses (e.g., H5N1, H7N9, and H9N2). None of the neuraminidase inhibitors is approved by the FDA for complicated influenza or for persons requiring hospitalization for the disease. While not licensed for the treatment of persons with complicated disease, inpatients, and pregnant women, oseltamivir is considered the drug of choice in these settings. The efficacy of zanamivir is similar to that of oseltamivir in hospitalized patients. Treatment is most effective when begun within 2 days of symptom onset and should be started as early as possible; such early treatment reduces symptoms by ~1 day in persons with uncomplicated disease. For persons with influenza requiring hospitalization and with pneumonia, treatment with oseltamivir or zanamivir is recommended even later. Treatment may reduce the risk of complications and death in hospitalized patients with influenza.

Oseltamivir and zanamivir (but not peramivir) are approved for prophylaxis of influenza, especially in institutions where outbreaks can be severe, and for prophylaxis in persons who have been exposed to the virus, are at high risk for disease complications, and have not recently been vaccinated. The efficacy of oseltamivir and zanamivir for prophylaxis is estimated to be ~70–90%. For persons at institutions, prophylaxis is given for at least 2 weeks and for up to 1 week after outbreaks resolve. For other high-risk persons, prophylaxis is given within 2 days of exposure and continued for 1 week after exposure. Since neuraminidase inhibitors reduce virus release from cells, they should not be given 2 days before or within 2 weeks after receipt of live, attenuated influenza vaccine. Resistance has been reported during treatment with oseltamivir or peramivir, especially in immunocompromised persons; oseltamivir-resistant viruses are usually sensitive to zanamivir.

Baloxavir Baloxavir inhibits the cap-dependent endonuclease that is important in initiating synthesis of influenza virus mRNA. This drug is approved by the FDA as a single oral dose for postexposure prophylaxis of influenza and for treatment of uncomplicated influenza in persons ≥12 years of age who have had symptoms for ≤48 h. Baloxavir inhibits influenza A and B viruses, including avian strains and strains that are resistant to neuraminidase inhibitors. The drug's efficacy is similar to that of the neuraminidase inhibitors in persons with uncomplicated influenza and reduces symptoms by ~1 day. In addition, baloxavir exhibits efficacy similar to that of oseltamivir for reducing symptoms in high-risk patients. However, its effectiveness in patients hospitalized with complications of influenza is unknown. Reduced sensitivity of influenza virus to baloxavir has been associated with mutations in the viral polymerase acidic protein after one dose. The incidences of nausea and vomiting are lower with baloxavir than with oseltamivir. Levels of the drug are lower if it is taken with dairy products, polyvalent cation–containing laxatives or antacids, or oral supplements containing calcium, iron, magnesium, selenium, or zinc. Since baloxavir reduces virus replication, it should not be given 2 days before or within 2 weeks after receipt of live, attenuated influenza vaccine.

Adamantanes Amantadine and rimantadine inhibit the influenza virus's M2 protein and its uncoating and membrane fusion. While these drugs are active against influenza A, resistance is widespread and can develop rapidly; thus, the adamantanes are not recommended as treatment or prophylaxis for influenza unless the virus is known to be sensitive.

■ RESPIRATORY SYNCYTIAL VIRUS

Ribavirin Ribavirin is an analogue of guanosine and inhibits replication of numerous RNA and DNA viruses. The drug inhibits viral RNA synthesis and capping of viral mRNA and in some cases increases the viral RNA mutation rate to lethal levels for some viruses. Ribavirin inhibits replication of respiratory syncytial virus (RSV), influenza virus, parainfluenza virus, and many other RNA viruses in vitro. While the drug has been used to treat numerous viral infections, including Lassa fever and hepatitis E, it is approved by the FDA only for use against RSV and as a component of combination therapy for hepatitis C. Aerosolized ribavirin is approved for treatment of hospitalized infants and young children with severe lower respiratory tract infections due to RSV; it is given for 12–18 h per day and is most effective when used early in the course of these severe infections. Ribavirin is given in a generator that yields an aerosol of particles small enough to reach the lower respiratory tract; the level of systemic absorption is low. The aerosolized form of the drug can induce bronchospasm, sudden deterioration of respiratory function (especially in infants), and rash and can precipitate in ventilators, interfering with their function. Ribavirin is mutagenic and teratogenic in animals; accordingly, it is not recommended for use in pregnant women, and the exposure of health care workers should be minimized with personal protective equipment. In early studies, ribavirin reduced the shedding of RSV and the severity of symptoms in hospitalized infants with lower respiratory tract disease who were not on mechanical ventilation, the duration of oxygen supplementation, and the duration of time on mechanical ventilation in infants. More recent analyses of the literature suggest that the efficacy of the drug in these settings is much less certain, and the drug is not recommended for routine use by the American Academy of Pediatrics. In retrospective studies, ribavirin has been reported to reduce the risk of progression of RSV from upper to lower respiratory tract disease in stem cell transplant recipients and to reduce mortality rates in these patients. In a retrospective study, the outcome of treatment with oral ribavirin was similar to that obtained with the aerosolized drug in hematopoietic stem cell transplant recipients with RSV disease. Ribavirin has not been shown to affect the clinical course of patients with parainfluenza and is not recommended for their treatment. Aerosolized ribavirin costs more than $25,000 per day.

Nirsevimab-alip Nirsevimab-alip, a human monoclonal antibody that targets the prefusion form of the RSV F protein, is approved for prevention of RSV lower respiratory tract disease in neonates and infants born during or entering their first RSV season, and for children up to 24 months who are vulnerable to severe RSV disease through their second RSV season.

Palivizumab Palivizumab, a humanized monoclonal antibody to RSV F protein, is approved for prevention of lower respiratory tract disease due to RSV in pediatric patients at high risk of RSV disease, including premature infants and children with bronchopulmonary dysplasia.

■ SARS-COV-2 (SEE CHAP. 204)

Remdesivir is converted in cells to an adenosine triphosphate analogue that inhibits the RNA-dependent RNA polymerase of several viruses. The drug is approved by the FDA for treatment of persons ≥12 years of age with SARS-CoV-2 requiring hospitalization; it shortens the duration of hospitalization in persons with lower respiratory tract disease. While the results of studies with the drug vary, it is recommended by the National Institutes of Health (NIH) for patients with SARS-CoV-2 who require supplemental oxygen while hospitalized. The drug is given intravenously and is not recommended in persons with a glomerular filtration rate (GFR) <30 mL/min. Serum transaminase elevations have been reported in healthy persons receiving remdesivir, and liver enzymes should be monitored before and during treatment. Chloroquine inhibits the activity of remdesivir in vitro; hydroxychloroquine or chloroquine phosphate should not be given with remdesivir.

Nirmatrelvir, a SARS-CoV-2 main protease inhibitor, boosted with ritonavir, a CYP3 and HIV protease inhibitor, is approved for treatment of mild to moderate COVID-19 in adults at high risk for progression to severe COVID-19. The drug should be given as soon as possible after infection and within 5 days of onset of symptoms. The dose should not be given with drugs highly dependent on CYP3A in which elevated levels can be associated with severe reactions, such as statins, sirolomus, tacrolimus, colchicine, and many other drugs. Use of the drug with medications that induce CYP3A can reduce the levels of nirmatrelvir or ritonavir with loss of effectiveness. The dose should be reduced for renal impairment.

Molnupiravir is an oral ribonucleoside analogue that inhibits replication of SARS-CoV-2. The drug reduced the risk of hospitalization or death in patients with mild to moderate COVID-19 by ~50% in a phase 3 clinical trial. AT-527 is an oral nucleotide prodrug that reduced SARS-CoV-2 viral loads in patients hospitalized with COVID-19 in a phase 2 clinical trial. PF-07321332 is an oral SARS-CoV-2 protease inhibitor that is being tested in combination with low-dose ritonavir in a phase 2/3 clinical trial for prevention of COVID-19 infection.

Remdesivir and nirmatrelvir boosted with ritonavir are approved by the FDA and recommended as first-line therapy for COVID-19 by the NIH guidelines; molnupiravir is approved under an emergency use authorization by the FDA and is considered second-line therapy. At the time of this writing, Pemgarda, a monoclonal antibody to SARS-CoV-2, can be given under emergency use authorization to prevent COVID-19 in immunocompromised persons age 12 and older.

■ INVESTIGATIONAL AGENTS FOR RESPIRATORY VIRUS INFECTIONS

Favipiravir (T705) inhibits viral RNA polymerases and is active against influenza and other RNA viruses. It is approved for treatment of emerging influenza viruses in Japan. Presatovir is an RSV fusion inhibitor that was ineffective in two trials of RSV disease. DAS181 (Fludase) is a sialidase that cleaves sialic acid, a receptor for influenza A and B and parainfluenza viruses; it did not improve the clinical outcomes of patients with influenza, but in case reports transplant recipients with parainfluenza have improved clinically with the drug. Laninamivir octanoate inhibits the neuraminidase of influenza A and B viruses and is approved for treating influenza in Japan. RSV604 interacts with the RSV nucleocapsid and is undergoing phase 2 studies in transplant recipients.

ANTIVIRAL DRUGS FOR HUMAN PAPILLOMAVIRUS AND POXVIRUS INFECTIONS

Interferon α (IFN-α) inhibits replication of many RNA and DNA viruses in vitro. IFN-α is approved by the FDA for intralesional treatment of external anogenital warts caused by human papillomavirus (HPV). It is effective in resolving lesions in ~50% of cases, with a recurrence rate of ~25%.

Imiquimod is a toll-like receptor 7 agonist that induces production of IFN-α and other cytokines. It is approved as a topical cream for treatment of external genital and perianal warts caused by HPV in persons ≥12 years of age. This drug is effective in resolving lesions in ~40% of cases.

Tecovirimat is approved by the FDA for treatment of smallpox and inhibits replication of mpox and vaccinia viruses. Resistance to tecovirimat developed in a person treated with the drug for progressive vaccinia and has been reported in persons with mpox.

INVESTIGATIONAL ANTIVIRAL DRUGS FOR PICORNAVIRUS

Pocapavir inhibits picornaviruses by inhibiting virus uncoating and is being developed to reduce poliovirus shedding; resistance to the drug develops rapidly.

ANTIVIRAL DRUGS FOR HEPATITIS B VIRUS INFECTION

Eight drugs representing two classes are approved for the treatment of chronic HBV infection in the United States. One class, the nucleos(t)ide analogues, act as chain terminators of nascently replicating DNA thereby competitively inhibiting HBV reverse transcriptase; the other class, exogenous IFNs, mimic and augment the role of endogenous interferons (Table 196-3). The goal of therapy for chronic hepatitis B is to prevent progression to cirrhosis, liver failure, and hepatocellular carcinoma. This can be achieved through long-term inhibition of viral replication with reduction in hepatic inflammation, the driver of liver

TABLE 196-3 Antiviral Drugs for Chronic Hepatitis B Treatment in Adults

DRUG	ROUTE AND DOSE	DEVELOPMENT OF RESISTANCE	COMMON SIDE EFFECTS[a]	TREATMENT MONITORING	COMMENTS
Interferons					
Pegylated α2a	SC injection; 180 μg/week for 48 weeks	Not described in long-term studies.	Side effects are common and include fevers, chills, myalgia, fatigue, neurotoxicity, and leukopenia. Autoantibodies can develop, particularly antithyroid antibodies.	Complete blood counts should be performed biweekly for the first month and then monthly, renal and liver function testing monthly, thyroid function testing every 3 months.	Recommended as first-line therapy. Best treatment response seen among patients with HBV genotype A and B infections. Contraindicated in clinically significant portal hypertension and pregnancy.
Pegylated α2b	SC injection; 1.5 μg/kg per week for 48 weeks				
Nucelos(t)ide Analogues					
Lamivudine	Oral; 100 mg daily	30% after 1 year; 70% after 5 years	Malaise or fatigue, GI symptoms (nausea/vomiting, abdominal pain, diarrhea), headache, upper respiratory tract infection	Renal and liver function testing every 3–6 months Assessment of lactic acid level HBV DNA and serologic testing every 3–6 months	Monotherapy recommended if duration of therapy is to be <1 year, as in prophylaxis against HBV reactivation with immunosuppression or chemotherapy.
Adefovir	Oral; 10 mg daily	20–29% after 5 years Adefovir is usually active against lamivudine-resistant HBV strains.	Headache, asthenia, GI symptoms (abdominal pain, nausea)	Renal and liver function testing every 6 months Assessment of lactic acid level HBV DNA and serologic testing every 3–6 months	—
Telbivudine	Oral; 600 mg daily	11–25% after 2 years Cross-resistance is common between lamivudine- and telbivudine-resistant HBV strains.	Headache, fatigue, GI symptoms (abdominal pain)	Measurement of creatine kinase level if there is concern about myopathy Renal and liver function testing every 3–6 months Assessment of lactic acid level HBV DNA and serologic testing every 3–6 months	—
Entecavir	Oral; 0.5–1 mg daily	1–2% after 5 years in nucleos(t)ide-naïve patients; 50% after 5 years in lamivudine-resistant patients	Headache, fatigue, elevated alanine aminotransferase level	Renal and liver function testing every 3–6 months Assessment of lactic acid level HBV DNA and serologic testing every 3–6 months	Recommended as first-line therapy. Dose of 0.5 mg daily in treatment-naïve patients, 1 mg daily in treatment-experienced patients. Dose adjusted in renal dysfunction.
Tenofovir disoproxil	Oral; 300 mg daily	No resistance after up to 10 years of treatment	Headache, fatigue, nasopharyngitis, upper respiratory tract infection, nausea	Renal and liver function testing every 3–6 months Phosphorus assessment in patients with chronic kidney disease Assessment of lactic acid level HBV DNA and serologic testing every 3–6 months	Recommended as first-line therapy. Dosing frequency—but not dose—reduced in chronic kidney disease. May be used during pregnancy; possible risk of low birth weight.
Tenofovir alafenamide	Oral; 25 mg daily	No resistance after up to 3 years of treatment	Headache, fatigue, nasopharyngitis, upper respiratory tract infection	Renal and liver function testing every 3–6 months Phosphorus assessment in patients with chronic kidney disease Assessment of lactic acid level HBV DNA and serologic testing every 3–6 months	Recommended as first-line therapy. May be used during pregnancy; possible risk of low birth weight.
Emtricitabine	Oral; 200 mg daily	Not defined	Headache, GI symptoms (nausea, diarrhea, abdominal pain), fatigue, depression, insomnia, abnormal dreams, rash, asthenia, increased cough, rhinitis	Renal and liver function testing every 3–6 months Assessment of lactic acid level HBV DNA and serologic testing every 3–6 months	While not approved for treatment of chronic HBV infection, used interchangeably with lamivudine. Dosing frequency adjusted in chronic kidney disease.

[a] For emtricitabine, side effects were assessed only in combination with antiretroviral therapy.

Abbreviation: GI, gastrointestinal.

fibrosis. Virologic responses (defined by suppression of HBV replication), biochemical responses (improvement or normalization of liver function values), and histologic responses (reduction in inflammation and fibrosis on liver biopsy) are often achievable with current treatments. However, loss of hepatitis B e antigen (HBeAg) (an intermediate treatment endpoint), viral clearance with loss of hepatitis B surface antigen (HBsAg), and immune control (defined by a hepatitis B surface antibody [HBsAb] level of >10 IU/mL) are uncommon with current therapies.

Treatment with a nucleos(t)ide analogue is considered first-line therapy for chronic HBV infection because of its antiviral potency, favorable side-effect profile, and ease of administration. All drugs in this class are given by mouth once daily. While all nucleos(t)ide analogues carry a black box warning for lactic acidosis and severe hepatomegaly, these adverse events were observed in patients taking older nucleoside analogues (such as stavudine and didanosine for the treatment of HIV) and have not occurred in clinical trials of newer nucleos(t)ides for chronic HBV infection. Once initiated, nucleos(t)ide therapy must be continued for a long duration because of the risk of virus rebound and subsequent hepatitis flare if treatment is stopped. This can occur in up to 40–50% of patients and rarely may lead to hepatic decompensation. Viral rebound occurs because nucleos(t)ides analogues do not target the covalently closed circular DNA—an episomal form of viral DNA. Comparative studies of nucleos(t)ide analogues have demonstrated that newer drugs (entecavir, tenofovir, disoproxil, and tenofovir alafenamide) are associated with lower rates of viral resistance than older agents (lamivudine, telbivudine, and adefovir), but, if viral replication is effectively suppressed, histologic and biochemical improvement will occur in ~60–75% of patients without significant differences between antiviral agents or combinations. However, rates of HBsAg clearance remain extremely low (<1–5%).

Pegylated IFN-α2a also is considered a first-line therapy for chronic hepatitis B infection. Pegylated IFN-α2a has certain advantages over nucleos(t)ide analogues—including a finite dosing period of 48 weeks, absence of viral resistance, and higher rates of serologic response—but it has lower rates of biochemical and virologic responses (<40% for both). Downsides to pegylated IFN-α2a include its poor tolerability because of numerous side effects; it is contraindicated in patients with clinically significant portal hypertension and pregnancy.

Response rates are higher when pegylated IFN-α2a is combined with nucleos(t)ide therapy in treatment-naïve patients: overall rates of HBsAg loss after 48 weeks of combination therapy with pegylated IFN-α2a and tenofovir disoproxil fumarate (TDF) were low but significantly higher than when either was given alone: 9.1% versus 0% with TDF alone ($p<.001$) and 2.8% with IFN alone ($p<.005$). Combination therapy is not recommended because long-term nucleos(t)ide analogue monotherapy can achieve similar rates of HBsAg loss as 1 year of combination therapy and greater viral suppression. There is minimum benefit to adding pegylated IFN-α2a to ongoing nucleos(t)ide therapy.

The choice of a class of agent is dependent on the presence of comorbid conditions that prevent the use of one agent over another and patient preference.

■ LAMIVUDINE

Lamivudine is an oral cytidine analogue that competitively inhibits the viral reverse transcriptase activity of both HIV and HBV, preventing viral replication. Lamivudine was the first oral agent approved for therapy of chronic hepatitis B. Its long-term use was limited by high rates of viral resistance, approaching 30% among patients treated for 1 year and ~70% after 5 years of therapy. Its use in chronic hepatitis B has been superseded by agents with better resistance profiles.

While not approved for the treatment of chronic HBV infection, emtricitabine is a cytosine analogue similar in structure, activity, and resistance to lamivudine. Used alone, it offers no advantage over lamivudine, but in combination with tenofovir (both TDF and tenofovir alafenamide fumarate [TAF]) it is used as part of an antiretroviral regimen to treat patients with HIV/HBV co-infection requiring lifelong antiviral therapy and off-label in selected cases of established nucleoside resistance.

■ ADEFOVIR

Adefovir dipivoxil is the oral prodrug of adefovir—a monophosphate nucleotide analogue of adenosine. This drug is active against HBV, HIV, some herpesviruses (HSV and CMV), and poxviruses. Adefovir is effective for the management of HBV in treatment-naïve patients and those infected with lamivudine-resistant HBV. Viral resistance to adefovir is slower to emerge than resistance to lamivudine but still develops in 20–30% of patients after 5 years of treatment. Adefovir has been replaced with nucleos(t)ide analogues with higher barriers to resistance as first-line treatment for chronic hepatitis B.

■ TELBIVUDINE

Telbivudine, a β-L enantiomer of thymidine, was approved by the FDA in 2006 for the treatment of chronic HBV infection. It has little or no activity against HIV replication. Telbivudine is generally well tolerated and effective against HBV replication, but the risk of viral resistance (25% in HBeAg-positive and 11% in HBeAg-negative patients after 2 years of use), myopathy, peripheral neuropathy, and fatigue limited its use. Telbivudine was withdrawn from the U.S. market primarily for economic reasons.

■ ENTECAVIR

Entecavir is a cyclopentyl guanosine analogue that, once triphosphorylated, blocks HBV polymerase in multiple ways, inhibiting priming and reverse transcription of the HBV negative strand and positive-strand synthesis. Entecavir effectively inhibits HBV replication, with resulting biochemical and histologic improvement. This drug is active against some lamivudine-resistant HBV strains, but only at concentrations 20- to 30-fold higher than those obtained with the standard 0.5-mg dose; thus, a higher dose (1 mg daily) of entecavir is recommended for patients with previous lamivudine exposure. Entecavir resistance leading to viral rebound and clinical hepatitis is uncommon among previously untreated patients but may occur in up to 50% of patients with prior lamivudine resistance after 5 years of entecavir treatment. Entecavir-resistant strains retain susceptibility to tenofovir and occasionally adefovir. Entecavir is generally well tolerated and highly bioavailable but should be taken on an empty stomach because food interferes with its absorption. The drug is renally cleared, and dosing should be adjusted for a CrCl of <50 mL/min.

■ TENOFOVIR

Tenofovir is a nucleotide analogue of adenosine monophosphate with activity against both retroviruses and hepadnaviruses. Two prodrug forms, TDF and TAF, are approved by the FDA for the treatment of both HIV infection and HBV infection. Tenofovir potently inhibits HBV replication. Rates of viral suppression are similar between TDF and TAF at 3 years. After 10 years of continuous use, 96–98% of patients receiving TDF achieve complete viral suppression. TAF is associated with higher biochemical response compared with TDF. Clinical resistance to tenofovir has not been observed with up to 8 years of therapy. Both formulations are renally eliminated, and renal toxicity—including acute renal failure, Fanconi syndrome, and diabetes insipidus—has been reported. The risk is higher with TDF compared to TAF. In clinical trials of TAF, there have been no reported cases of Fanconi syndrome or proximal renal tubulopathy. Small declines in bone mineral density (~2.3% at 5 years with TDF and <1% at 3 years with TAF) have been observed. Routine monitoring of renal function during therapy with both agents is indicated, and dose frequency should be reduced in patients with GFR <50 mL/min.

■ INTERFERONS

IFNs have a broad spectrum of antiviral activity in addition to modulating the immune system. Recombinant α, β, γ, and λ IFNs have been evaluated in a variety of viral infections. Standard IFN- α2b was the first drug approved for treatment of chronic hepatitis B, but it has been largely replaced by pegylated IFN-α2a. Pegylation of IFN, through linkage of IFN to polyethylene glycol, results in slower absorption, decreased clearance, and more sustained serum IFN concentrations, thereby permitting a more convenient once-weekly dosing schedule. Consequently, pegylated IFN has supplanted standard IFN. IFNs are associated with numerous adverse effects including fever, myalgia, fatigue, somnolence, depression, confusion, leukopenia, and development of autoantibodies, including antithyroid antibodies, that limit its tolerability and patient acceptance.

Pegylated IFN-α2a is approved by the FDA for therapy in patients with chronic hepatitis B and C. Pegylated IFN-α2b is no longer available in the United States.

The administration of pegylated IFN-α2a for 48 weeks in patients with HBeAg-positive infection resulted in the loss of markers for HBV replication (e.g., HBeAg and HBV DNA in 29–36% and 8–14% of cases, respectively; 2–7% of patients also cleared HBsAg). In most patients who lose HBeAg and HBV DNA, serum aminotransferases return to normal levels, and the viral and biochemical responses are maintained in the long term. Predictors of a favorable response to pegylated IFN-α2a therapy include low pretherapy levels of HBV DNA, high pretherapy serum levels of alanine aminotransferase (ALT), a short duration of chronic HBV infection, HBV genotypes A and B, and active liver inflammation on biopsy. Poor responses are seen in patients with HBeAg-negative and immunosuppressed patients, including those infected with HIV.

ANTIVIRAL DRUGS FOR HEPATITIS C INFECTION

The goal of HCV treatment is long-term suppression of viral replication or a sustained virologic response (SVR). SVR is achieved when levels of HCV RNA in the serum remain undetectable 12 weeks after the end of treatment. SVR is considered synonymous with cure, as it is associated with durable suppression of HCV replication, lower all-cause and liver-related mortality, and a reduced risk of hepatocellular carcinoma. These benefits have been confirmed in patients with and without advanced liver disease and cirrhosis who received IFN-based and IFN-sparing, combination direct-acting antiviral drugs (DAAs).

Several targeted therapies with DAAs are effective against HCV (Table 196-4). Three classes of DAAs that target the NS5B RNA-dependent RNA polymerase, the NS3/4 protease, and NS5A, a zinc-binding phosphoprotein that is integral for HCV RNA replication, form the basis of curative regimens for chronic HCV infection. A combination of two or three DAAs is now the standard of care for the treatment of chronic HCV infection, regardless of genotype or fibrosis stage. Two pangenotypic regimens, glecaprevir/pibrentasvir and sofosbuvir/velpatasvir administered for 8 to 12 weeks, respectively, are the most widely used regimens with SVR rates that exceed 95% for all HCV genotypes.

Two pangenotypic regimens with high SVR rates are approved specifically for re-treatment of chronic HCV infection after initial treatment failure: glecaprevir/pibrentasvir and sofosbuvir/velpatasvir/voxilaprevir. In the setting of unfavorable resistance-associated variants (RAVs) or cirrhosis, re-treatment efficacy can frequently be improved by extension of the treatment course or the addition of ribavirin. Review of the online joint American Association for the Study of Liver Diseases/Infectious Diseases Society of America's HCV Guidelines is useful for selecting the appropriate DAA regimen. In addition, for all DAA-based treatments, checking for drug–drug interactions before the initiation of therapy is recommended.

Most regimens are well tolerated, but all DAAs carry a black-box warning about reactivation of HBV—mostly among HBsAg-positive persons and to a lesser extent patients with isolated anti-HBc following HCV suppression. In some cases, fulminant hepatitis, hepatic flare, and death have occurred in patients with untreated HBV infection who underwent treatment for chronic HCV infection. These risks are rare and can be safely managed with routine monitoring; treatment of HCV should not be deferred because of HBV co-infection.

NS5B POLYMERASE AND NS5A-CONTAINING REGIMENS

Sofosbuvir Sofosbuvir is the prodrug of a uridine inhibitor of the HCV NS5B RNA-dependent RNA polymerase. The active uridine nucleoside triphosphate results in termination of viral RNA replication. Sofosbuvir is approved by the FDA for the treatment of HCV

TABLE 196-4 Antiviral Drugs for Hepatitis C Treatment in Adults[a]

DRUG FORMULATION	ROUTE, DOSE, DURATION	MECHANISM(S) OF ACTION	SPECTRUM OF ACTIVITY	COMMON SIDE EFFECTS	COMMENTS
Sofosbuvir	Oral; 400 mg daily; duration varies (12–24 weeks)	Nucleoside analogue	Genotypes 1–6	Headache, fatigue	Should be combined with at least one other DAA from a different class.
Sofosbuvir/ledipasvir	Oral; 400 mg/90 mg daily; 8, 12, or 24 weeks	Nucleoside analogue/NS5A inhibitor	Genotypes 1, 4, 5, and 6	Headache, fatigue	Avoid coadministration with antacid medications.
Sofosbuvir/velpatasvir	Oral; 400 mg/100 mg daily; 12 weeks	Nucleoside analogue/NS5A inhibitor	Genotypes 1–6	Headache, fatigue	Avoid coadministration with antacid medications.
Sofosbuvir/velpatasvir/voxilaprevir	Oral; 400 mg/100 mg/100 mg once daily; 12 weeks	Nucleoside analogue/NS5A inhibitor/protease inhibitor	Genotypes 1–6	Headache, fatigue, diarrhea, nausea	Approved for re-treatment of patients with previous DAA experience. Avoid coadministration with antacid medications.
Elbasvir/grazoprevir	Oral; 50 mg/100 mg once daily; 12 or 16 weeks	NS5A inhibitor/protease inhibitor	Genotypes 1 and 4	Fatigue, anemia, headache, nausea	Pretreatment testing for resistance-associated variants recommended in patients infected with genotype 1a. Monitor hepatic function panel at 8 weeks and again at 12 weeks if patient is receiving 16 weeks of treatment.
Glecaprevir/pibrentasvir	Oral; 3 100-mg tablets/40 mg once daily; 8, 12, or 16 weeks	NS5A inhibitor/protease inhibitor	Genotypes 1–6	Headache, fatigue	—
Daclatasvir	Oral; 60-mg tablet once daily; 12 weeks Dose reduced to 30 mg once daily when taken with a strong CYP3A inhibitor Dose increased to 90 mg once daily when taken with moderate CYP3A inducers	NS5A inhibitor	Genotypes 1 and 3	Headache, fatigue	Use recommended only along with sofosbuvir—with or without ribavirin—for genotype 1 or 3 infection; no longer considered a first- or second-line regimen.
Ribavirin	Oral; 3–6 200-mg capsules once daily or in divided doses, based on weight, history of cardiovascular disease, and renal function	Nucleoside analogue, also unknown mechanisms	Unknown, used for all genotypes	Anemia, nausea, teratogenic in pregnancy	Used only as combined therapy with DAAs or interferon. Complete blood counts should be monitored after 2 weeks of treatment and as clinically indicated thereafter. Dose may be adjusted based on anemia and renal function.

[a] While these drugs are approved by the FDA for chronic but not acute HCV, they have been recommended for acute HCV by both the Infectious Diseases Society of America and the American Association for the Study of Liver Diseases.

Abbreviation: DAA, directly acting antiviral agent.

genotypes 1–4 and is active against genotypes 1–6. Resistance to sofosbuvir is conferred by an S282T substitution in the NS5B protein, but clinically significant resistance to sofosbuvir treatment has rarely been encountered and virologic breakthrough during sofosbuvir treatment is exceedingly rare. Sofosbuvir is approved for use with other DAAs as part of three fixed-dose combination regimens: as two-drug regimens with the NS5A protein inhibitors ledipasvir and velpatasvir, and as a three-drug regimen with velpatasvir and the protease inhibitor voxilaprevir. Both sofosbuvir and its active metabolite are renally cleared, and while the FDA has approved this drug only for patients with an estimated GFR of ≥30 mL/min, several studies have demonstrated its safety and efficacy in end-stage renal disease and for patients undergoing dialysis. Sofosbuvir has not been associated with significant toxicity or drug interactions with one notable exception: sofosbuvir potentiates amiodarone and may cause severe bradycardia, especially if coadministered with amiodarone and a beta blocker.

Sofosbuvir/Ledipasvir Ledipasvir is an NS5A protein inhibitor that is available only in combination with sofosbuvir. The fixed-dose combination of ledipasvir and sofosbuvir is effective against genotypes 1, 4, 5, and 6 with SVR rates of 95–100%. The standard duration of treatment is 12 weeks for genotypes 1 (all subgenotypes), 4, 5, and 6; however, treatment duration may be reduced to 8 weeks in treatment-naïve, genotype 1–infected noncirrhotic patients with baseline HCV RNA levels below 6 million copies/mL. Treatment should be extended to 24 weeks or ribavirin should be added in patients who have decompensated cirrhosis or previous DAA exposure. Ledipasvir is excreted via the biliary route, and no adjustment is needed for mild or moderate renal impairment. Several studies have shown that sofosbuvir/ledipasvir is safe in end-stage renal disease, but it remains FDA approved only for patients with a CrCl of >30 mL/min. No dose reduction is required for decompensated cirrhosis (Child–Turcotte–Pugh class B or C). Ledipasvir absorption is improved with food intake and is inhibited by antacids or proton pump inhibitors. Ledipasvir is an inhibitor of P-glycoprotein and may increase levels of tenofovir; renal function should be monitored in patients receiving both medications, although clinically significant interactions are unlikely during the relatively short period of treatment. Ledipasvir is generally well tolerated, and clinical trials have shown only a small increase in side effects, including headache and fatigue, over those occurring with placebo.

Sofosbuvir/Velpatasvir While chemically similar to ledipasvir, velpatasvir has an expanded spectrum of activity and exhibits improved efficacy over ledipasvir against HCV genotypes 2 and 3. Velpatasvir is available only in combination with sofosbuvir for the treatment of naïve patients with genotype 1–6 infection and all stages of fibrosis, including decompensated cirrhosis. SVR rates for patients without cirrhosis were 96–100% for all HCV genotypes. In contrast to sofosbuvir/ledipasvir treatment, shortening of the duration of sofosbuvir/velpatasvir therapy in these patients is not required. Similar to ledipasvir, velpatasvir should be taken with food, and coadministration with antacids or proton pump inhibitors should be avoided. Velpatasvir is in general well tolerated, and reported side effects are minimal.

Sofosbuvir/Velpatasvir/Voxilaprevir Available in a triple-drug combination with sofosbuvir and velpatasvir, voxilaprevir is a NS3/NS4A protease inhibitor that is active against HCV genotypes 1–6. The fixed-dose combination for 12 weeks is recommended for the re-treatment of patients with genotype 1–6 infection in whom SVR has not been attained after previous combination DAA treatment and for treatment-naïve genotype 3–infected patients with cirrhosis and the NS5A resistance-associated variant Y93H. Patients with genotype 3 infection who have failed an NS5A protein inhibitor–regimen have lower SVR rates to re-treatment with sofosbuvir/velpatasvir/voxilaprevir for 12 weeks; thus, it is recommended either to add ribavirin or, if ribavirin cannot be tolerated, to extend the duration of therapy to 24 weeks. Voxilaprevir is not recommended for patients with decompensated cirrhosis (see "Protease Inhibitors and Protease Inhibitor–Containing Regimens," below) or those with significant renal impairment and a CrCl of <30 mL/min. Voxilaprevir, like other protease inhibitors, is metabolized by the CYP3A system, and the effect of voxilaprevir may be reduced in the presence of other CYP inducers.

Sofosbuvir/Daclatasvir The combination of sofosbuvir with daclatasvir—the only NS5A protein inhibitor available individually rather than coformulated with other DAAs—is approved for the treatment of HCV genotypes 1 and 3. Daclatasvir binds the N terminus of the NS5A protein, both inhibiting viral RNA replication and blocking virion assembly. It is given in combination with sofosbuvir for 12 weeks and is safe for the treatment of patients with decompensated cirrhosis. Daclatasvir is a substrate of CYP3A, and the dose should be reduced if daclatasvir is given with a strong CYP3A inhibitor and increased if it is given with moderate CYP3A4 inducers. Daclatasvir absorption is not affected by food, and daclatasvir is highly protein bound. The dose does not need to be adjusted for renal impairment, and side effects are uncommon.

PROTEASE INHIBITOR–CONTAINING REGIMENS

Protease inhibitors are specifically designed to inhibit the HCV NS3/4A serine protease by mimicking the HCV polypeptide and, when bound by the viral protease, form a covalent bond with the catalytic NS3 serine residues, blocking further activity and preventing proteolytic cleavage of the HCV polyprotein into NS4A, NS4B, NS5A, and NS5B proteins. As a class, the protease inhibitors are hepatically metabolized and therefore should not be administered to patients with decompensated (Child–Turcotte–Pugh class B or C) cirrhosis. For patients receiving protease inhibitors, the current recommendation is that liver function tests should be monitored monthly.

Glecaprevir/Pibrentasvir Glecaprevir is a pangenotypic NS3/NS4A protease inhibitor that is coformulated with pibrentasvir, a pangenotypic NS5A protein inhibitor. Each medication individually has a high genetic barrier to resistance and is active against HCV genotypes 1–6. In patients infected with genotypes other than genotype 3, baseline resistance has no influence on glecaprevir treatment efficacy, and NS3/NS4A baseline polymorphisms have not been noted to correlate with virologic failure. Treatment duration varies with fibrosis and treatment experience: an 8-week course of therapy is recommended for treatment-naïve patients who are infected with any genotype and have any degree of fibrosis up to compensated cirrhosis, including patients with genotype 3 infection. This results in SVR rates of 95–100% across all HCV genotypes. Treatment-experienced cirrhotic patients should receive 12 weeks of treatment, and patients with prior NS5A protein inhibitor exposure with or without compensated cirrhosis should receive 16 weeks of therapy. The combination of glecaprevir/pibrentasvir should be taken with food. Clearance is via biliary excretion; therefore, no dose adjustment is required in end-stage renal disease. Because of the protease component, the combination of glecaprevir/pibrentasvir is not appropriate for patients with decompensated cirrhosis. Glecaprevir and pibrentasvir are only weak CYP3A inducers, but they inhibit the P glycoprotein, breast cancer resistance protein (BCRP), and organic anion transporter P1 (OATP1) drug transporters. When taken with other drugs that are substrates for these transporters, concentrations of both drugs may be increased. The combination regimen is generally well tolerated; mild headache, fatigue, diarrhea, and nausea have been reported.

Elbasvir/Grazoprevir The coformulation of elbasvir, an NS5A replication complex inhibitor, and grazoprevir, an NS3/NS4A protease inhibitor, is active against HCV genotypes 1 and 4. However, its efficacy in the treatment of HCV genotype 1a is reduced in the presence of baseline RAVs in the NS5A protein at positions M28, Q30, L31, and Y93; thus, in patients infected with genotype 1a, baseline resistance testing should be performed and, if the result is positive, ribavirin should be added and therapy should be extended to improve response rates. Susceptibility to grazoprevir is reduced with NS5A protein D168 substitutions, but few resistant isolates have been noted in cases of virologic failure; thus, testing for these substitutions before therapy is not recommended. Treatment duration is 12 weeks (genotype 1b or genotype 1a without baseline RAVs) or 16 weeks (in combination with ribavirin in patients with baseline NS5A protein polymorphisms and in genotype 4–infected patients with previous IFN exposure). Absorption of grazoprevir and elbasvir is unaffected by food, and the dose does not need to be adjusted in patients with chronic kidney disease or those who are undergoing dialysis. Elbasvir, like grazoprevir, is a substrate of the CYP3A system; coadministration with moderate or strong CYP3A inducers or with strong inhibitors is not recommended. Both components are well tolerated, and few side effects have been reported. The use of this drug combination, as with all those containing protease inhibitors, is contraindicated in decompensated cirrhosis.

INTERFERONS

Several IFN preparations have been studied and approved as therapeutic options for chronic HCV infection. Approved regimens combined IFN/pegylated IFN with ribavirin, a nonspecific nucleoside analogue with the antiviral effects discussed below. The approval of direct acting antiviral agents in 2014 led to revised guidance, and IFN therapy is no longer recommended for the treatment of hepatitis C.

RIBAVIRIN

Ribavirin, a synthetic oral triazole guanosine analogue, weakly inhibits both DNA and RNA polymerases, but its primary mechanism in HCV treatment is not well understood. It may promote infidelity of RNA viral replication, giving rise to unfit or less fit viral mutations, and also appears to stimulate IFN-response genes and modulate adaptive immune responses. The role of ribavirin in HCV therapy has changed over time. Ribavirin played an integral role in HCV treatment during the IFN era to prevent virologic relapse and, combined with sofosbuvir, was required as part of IFN-sparing regimens before other DAAs were available. However, adverse drug effects associated with higher doses (in heavier patients)—including hemolytic anemia, which is increased with renal failure—were frequently treatment-limiting. Other side effects include rash, myalgia, and fatigue. Ribavirin is teratogenic, and its use in women with child-bearing potential is therefore limited.

With the advent of several combination DAA-only, IFN-sparing regimens, there are often multiple ribavirin-free options for treatment. However, there are still several indications for ribavirin augmentation of combination DAA-based therapy. Most importantly, ribavirin improves the SVR rate by an average of 5% in treatment-naïve and treatment-experienced patients with genotype 1 infection, particularly that due to subgenotype 1a. The addition of ribavirin to treatment with paritaprevir/ritonavir/ombitasvir plus dasabuvir is recommended for patients with genotype 1a or 4 infection as well as for patients infected with genotype 1a who are receiving elbasvir/grazoprevir with baseline NS5A protein RAVs to overcome reduced susceptibility to elbasvir. Ribavirin is frequently included in regimens for re-treatment of genotype 1–infected, therapy-experienced patients with cirrhosis in order to preserve SVR rates while shortening re-treatment duration. SVR rates at 12 weeks were comparable in treatment-experienced cirrhotic patients receiving 24 weeks of ledipasvir/sofosbuvir and those receiving 12 weeks of ledipasvir/sofosbuvir plus ribavirin. Ribavirin also improves outcomes in treatment-experienced patients with genotype 3 infection—an ongoing therapeutic challenge even in the setting of current pangenotypic regimens. Ribavirin improves treatment response in other clinical settings as well, specifically in patients with decompensated cirrhosis for whom treatment protease inhibitors cannot be used and in patients with genotype 2 infection in resource-limited settings where ribavirin is more affordable than fixed-dose combination DAA regimens. Because of its broad antiviral effects, ribavirin is not known to select for any particular RAVs.

Absorption of ribavirin is improved by administration with food, and the drug is excreted renally. Lowering the dose of the drug may reduce toxicity. While determining red blood cell counts and hemoglobin levels after 2 weeks of therapy is recommended to monitor for hemolytic anemia, ribavirin can be administered safely to most patients for the relatively short period of DAA-based therapy. In patients with renal insufficiency and those with end-stage renal disease who are undergoing dialysis, the dose must be adjusted and the patient closely monitored for anemia.

In a recent large-scale study, ribavirin was effective in the treatment of chronic infection with hepatitis E virus, which can cause chronic inflammatory hepatitis in immunosuppressed patients, particularly solid-organ transplant recipients.

ANTIVIRAL DRUGS FOR HEPATITIS D INFECTION

INTERFERONS

At high doses, IFN-α and pegylated IFN-α are active against hepatitis D virus infection. In off-label use for hepatitis D, SVR was achieved in 25–35% of patients treated with IFN-α and 17–43% of patients treated with pegylated IFN-α for 48 weeks. Virologic and biochemical relapse occur frequently after stopping IFN. Extending the duration is associated with maintenance of clinical response and HBsAg loss in a few cases.

ENTRY INHIBITORS

Bulevirtide is a synthetic lipopeptide that mimics a region within pre-S1 of the large HBsAg and irreversibly binds to the sodium taurocholate cotransporting polypeptide, the hepatocyte entry receptor for both HDV and HBV. Bulevirtide indirectly reduces HDV replication by blocking viral entry and new rounds of infection. Bulevirtide is approved by the European Medicines Agency (EMA) at a dose of 2 mg subcutaneously once daily for use in patients with compensated chronic HDV infection. In off-label use, bulevirtide was shown to be effective in patients with decompensated liver disease due to chronic hepatitis D infection. In a clinical trial, bulevirtide given 2 or 10 mg subcutaneously once daily was able to suppress HDV viremia by at least 100-fold and normalize alanine aminotransferase (ALT) levels in 45% and 48% of patients, respectively; undetectable viremia was achieved in 12.2% and 20% of patients, respectively. Combined bulevirtide and pegylated interferon was superior to bulevirtide alone to reduce HDV RNA to an undetectable level. No emergence of viral resistance has been observed to date. Asymptomatic elevation in bile acids and injection-site reactions are the most common adverse reactions. Bulevirtide is not yet approved in the United States.

Acknowledgment

The authors gratefully acknowledge the contributions of Dr. Eleanor Wilson to this chapter in the previous edition.

FURTHER READING

Acosta E et al: Advances in the development of therapeutics for cytomegalovirus infections. J Infect Dis 221:S32, 2020.

American Association for the Study of Liver Diseases/Infectious Diseases Society of America: Recommendations for testing, managing, and treating hepatitis C. Available at *http://www.hcvguidelines.org*. Accessed October 28, 2023.

Asselah T, Rizzetto M: Hepatitis D virus infection. N Engl J Med 389:58, 2023.

Asselah T et al: Bulevirtide combined with pegylated interferon for chronic hepatitis D. N Engl J Med 391:133,2024.

Avery R et al: Maribavir for refractory cytomegalovirus infections with or without resistance post-transplant: Results from a phase 3 randomized clinical trial. Clin Infect Dis 75:690, 2022.

Chou R et al: Screening for hepatitis C virus infection in adolescents and adults: Updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 323:976, 2020.

Drysdale SB et al: Nirsevimab for prevention of hospitalizations due to RSV in infants. N Engl J Med 389:2425, 2023.

Gnann JW Jr, Whitley RJ: Genital herpes. N Engl J Med 375:666, 2016.

Hanula R et al: Evaluation of oseltamivir used to prevent hospitalization in outpatients with influenza: A systematic review and meta-analysis. JAMA Intern Med 184:18, 2024.

Ikematsu H et al: Baloxavir marboxil for prophylaxis against influenza in household contacts. N Engl J Med 383:309, 2020.

Ison MG et al: Early treatment with baloxavir marboxil in high-risk adolescent and adult outpatients with uncomplicated influenza

(CAPSTONE-2): A randomised, placebo-controlled, phase 3 trial. Lancet Infect Dis 20:1204, 2020.

Saullo JL, Miller RA: Cytomegalovirus therapy: Role of letermovir in prophylaxis and treatment in transplant recipients. Annu Rev Med 74:89, 2023.

Tang LE et al: Chronic hepatitis B infection: A review. JAMA 319:1802, 2018.

Uyeki T et al: Clinical practice guidelines by the Infectious Diseases Society of America: 2018 update on diagnosis, treatment, chemoprophylaxis, and institutional outbreak management of seasonal influenza. Clin Infect Dis 68:e1, 2019.

Venkatesan S et al: Neuraminidase inhibitors and hospital length of stay: A meta-analysis of individual participant data to determine treatment effectiveness among patients hospitalized with nonfatal 2009 pandemic influenza A (H1N1) virus infection. J Infect Dis 221:356, 2020.

Vyas A et al: Real-world outcomes associated with letermovir use for cytomegalovirus primary prophylaxis in allogeneic hematopoietic cell transplant recipients: A systematic review and meta-analysis of observational studies. Open Forum Infect Dis 10:ofac687, 2022.

Yardeni D et al: Current best practice in hepatitis B management and understanding long-term prospects for cure. Gastroenterology 164:42, 2023.

Section 12 Infections Due to DNA Viruses

197 Herpes Simplex Virus Infections

Lawrence Corey

DEFINITION

Herpes simplex viruses (HSV-1, HSV-2; *Herpesvirus hominis*) produce a variety of infections involving mucocutaneous surfaces, the peripheral nervous system (PNS), the central nervous system (CNS), and—on occasion—visceral organs. Prompt recognition and treatment reduce the morbidity and mortality rates associated with HSV infections.

ETIOLOGIC AGENT

The genome of HSV is a 152-kb linear, double-stranded DNA molecule (molecular weight, ~100 × 10^6) that encodes >90 transcription units with 84 identified proteins. The genomic structures of the two HSV subtypes are similar. The overall genomic sequence homology between HSV-1 and HSV-2 is ~50%, whereas the proteome homology is >80%. The homologous sequences are distributed over the entire genome map, and most of the polypeptides specified by one viral type are antigenically related to polypeptides of the other viral type. Many type-specific regions unique to HSV-1 and HSV-2 proteins do exist, and a number of them appear to be important in host immunity. These type-specific regions have been used to develop serologic assays that distinguish between the two viral subtypes. The most commonly used protein is glycoprotein G (US-4), which differs markedly in size and antigenic sites between HSV-1 and HSV-2. Either restriction endonuclease analysis or sequencing of viral DNA can be used to distinguish between the two subtypes and among strains of each subtype. Recombinant viruses (HSV-1/HSV-2) do circulate in nature. The variability of nucleotide sequences from clinical strains of HSV-1 and HSV-2 is such that HSV isolates obtained from two individuals can be differentiated by restriction enzyme patterns or genomic sequences. Epidemiologically related sources, such as sexual partners, mother–infant pairs, or persons involved in a common-source outbreak, can be inferred from such patterns. Deep sequencing of sequential isolates suggests that more than one variant of HSV-1 or HSV-2 can be found in a single individual and minor mutational changes do occur within anatomic sites and over time.

The viral genome is packaged in a regular icosahedral protein shell (capsid) composed of 162 capsomeres **(Chap. 195)**. The outer covering of the virus is a lipid-containing membrane (envelope) acquired as the DNA-containing capsid buds through the inner nuclear membrane of the host cell. Between the capsid and lipid bilayer of the envelope is the tegument. Viral replication has both nuclear and cytoplasmic phases. Only four of the 12 glycosylated envelope proteins appear to be essential for cell entry: glycoprotein D (gD), gH, gL, and gB. gD binds to cellular co-receptors that belong to the heparin sulfate or tumor necrosis factor receptor family of proteins, the immunoglobulin superfamily (nectin family), triggering a conformational change that alters activation of the gH-gL heterodimer complex that then activates gB and the fusogen glycoprotein gC. The ubiquity of these receptors contributes to the wide host range of herpesviruses. HSV replication is highly regulated. After fusion and entry, the nucleocapsid enters the cytoplasm and several viral proteins are released from the virion. Some of these viral proteins shut off host protein synthesis (by increasing cellular RNA degradation), whereas others "turn on" the transcription of immediate early genes of HSV replication. These immediate early gene products, designated *α genes*, are required for synthesis of the subsequent polypeptide group: the β polypeptides, many of which are regulatory proteins and enzymes required for DNA replication. Most current antiviral drugs interfere with β proteins, such as viral thymidine kinase (TK) and DNA polymerase. The third (γ) class of HSV genes encodes viral structural and tegument proteins and mostly requires viral DNA replication for expression. New antiviral drugs directed at viral assembly and release are under development.

After viral genome replication and structural protein synthesis, nucleocapsids are assembled in the cell's nucleus. Specific viral proteases clip the end of the DNA into procapsid. In the nucleus, the nucleocapsid binds through the inner nuclear membrane to genetic vessels that fuse with the outer membrane and moves the capsid into the cytoplasm. In some cells, viral replication in the nucleus forms two types of inclusion bodies: type A basophilic Feulgen-positive bodies that contain viral DNA and eosinophilic inclusion bodies that are devoid of viral nucleic acid or protein and represent a "scar" of viral infection. The cytoplasmic capsids move along microtubules to the Golgi network where a second round of envelopment occurs. The capsids acquire their lipid envelope and most of the tegument. Cellular machinery transports the infectious virus out of the cell.

Viral genomes are maintained by some neuronal cells in a repressed state called *latency*. Latency, which is associated with transcription of only a limited number of virus-encoded RNAs, accounts for the presence of viral DNA and RNA in neural tissue at times when infectious virus cannot be isolated. Maintenance and growth of neural cells from latently infected ganglia in tissue culture result in production of infectious virions (*explantation*) and in subsequent permissive infection of susceptible cells (*co-cultivation*). Activation of the viral genome may then occur, resulting in *reactivation*—the normal pattern of regulated viral gene expression and replication and HSV release. The release of virions from the neuron follows a complex process of anterograde transport down the length of neuronal axons. In experimental animals, ultraviolet light, systemic and local immunosuppression, and trauma to the skin or ganglia are associated with reactivation.

A noncoding region of the viral genome initially felt to be three noncoding regions and now felt to be a more diverse set of noncoding RNAs and micro-RNAs (miRNAs) collectively referred to as the latency-associated transcripts (LATs) are found in the nuclei of latently infected neurons, and deletion mutants of the LAT region exhibit reduced efficiency in their later reactivation. HSV DNA copy number is highly variable between neurons, with no direct correlation between HSV DNA copy numbers and LAT positivity. About 10% of ganglionic neurons contain viral DNA and only about 1% of these neurons express LATs. Substitution of HSV-1 LATs for HSV-2 LATs induces an HSV-1 reactivation pattern, suggesting this region of the genome apparently maintains—rather than establishes—latency. Viral miRNA appears to silence expression of the key neurovirulence factor infected-cell protein 34.5 (ICP34.5) and to bind in an antisense configuration to the immediate-early protein ICP0 messenger RNA to prevent expression, which is vital to HSV reactivation. While certain viral transcripts are known to be necessary for reactivation from latency, the molecular mechanisms of HSV latency are not fully understood and strategies to interrupt or maintain latency in neurons are incompletely understood.

While latency is the predominant state of virus on a per-neuron basis, the high frequency of oral and genital tract reactivation for HSV-1 and HSV-2 suggests that the viruses are rarely quiescent within the entire biomass of ganglionic tissue. The virus appears to be in a dynamic state—"mostly suppressed"—but with continual individual cells showing various degrees of viral transcriptional activity, and only a few of these infected neurons giving rise to actual reactivation. There is increasing recognition that HSV infection of the autonomic ganglia plays an important role in both initial and reactivation infections. In fact, deaths of animals from HSV-2 infection appear to be related to autonomic dysfunction of the bowel. Both HSV-1 and HSV-2 are shed subclinically. Most persons infected with HSV-2 and HSV-1 have frequent subclinical bursts of reactivation lasting 2–6 h, and the host tissue-based immune system can contain viral reactivation in the tissue before the development of clinical reactivation.

PATHOGENESIS

Exposure to HSV at mucosal surfaces or abraded skin sites permits entry of the virus into cells of the epidermis and dermis and initiation of viral replication therein. HSV infections are usually acquired subclinically. Whether clinical or subclinical, HSV acquisition is associated with sufficient viral replication to permit infection of sensory and/or autonomic nerve endings. On entry into the neuronal cell, the virus—or, more likely, the nucleocapsid—is transported intra-axonally to the nerve cell bodies in ganglia. Viral particles tether onto cellular proteins that motor along microtubules from axon tips (neurite endings) to neuronal cell bodies. In humans, the transit interval of spread to the ganglia after virus inoculation into peripheral tissue is unknown. During the initial phase of infection, viral replication occurs in ganglia and contiguous neural tissue. Virus then spreads to other mucocutaneous surfaces through centrifugal migration of infectious virions via peripheral nerves. This mode of spread helps explain the large surface area involved, the high frequency of new lesions distant from the initial crop of vesicles that is characteristic in patients with primary genital or oral–labial HSV infection, and the ability to recover virus from neural tissue distant from neurons innervating the inoculation site. Contiguous spread of locally inoculated virus also may take place and allow further mucosal extension of disease. Recent studies have demonstrated HSV viremia—another mechanism for extension of infection throughout the body—in ~30–40% of persons with primary HSV-2 infection; latent infection with both viral subtypes in both sensory and autonomic ganglia has been demonstrated. For HSV-1 infection, trigeminal ganglia are most commonly infected, although extension to the inferior and superior cervical ganglia also occurs. With genital infection, sacral nerve root ganglia (S2–S5) are most commonly affected. Autonomic ganglia, pelvic nerves, and vaginal nerve roots are commonly infected.

After resolution of primary disease, infectious HSV can no longer be cultured from the ganglia; however, neuronal infection, as defined by the presence of viral DNA, persists in ganglionic cells in the anatomic regions of the initial infection. The mechanism of reactivation from latency is unknown, although increasingly evidence of limited viral genes or miRNAs is identified in latently infected neurons. Evidence exists for viral antigen and activated host T cells at the ganglia and periphery, and immune responses in ganglia as well as peripheral tissue appear to influence the frequency and severity of HSV reactivation. HSV-specific T cells have been recovered from peripheral nerve root ganglia. Many of these resident CD8+ T cells are juxtaposed with latently HSV-1-infected neurons in the trigeminal ganglia and can block reactivation with both interferon (IFN) γ release and granzyme B–mediated degradation of the immediate-early protein ICP4. In addition, there appears to be a latent viral load in the ganglia that correlates positively with the number of neurons infected and the rate of reactivation but inversely with the number of T cells present. It is not known whether reactivating stimuli transiently suppress these immune cells, independently upregulate transcription of lytic genes, or both. Moreover, host containment in the mucosa has been demonstrated. Once virus reaches the dermal–epidermal junction, there are three possible outcomes: (1) rapid host containment of infection near the site of reactivation; (2) spread of small amounts of virus into the epidermis, with a micro-ulceration associated with low-titer subclinical shedding; and (3) widespread replication and necrosis of epithelial cells and subsequent clinical recurrence (the latter defined clinically by a skin blister and ulceration). Histologically, herpetic lesions involve a thin-walled vesicle or ulceration in the basal region, multinucleated cells that may include intranuclear inclusions, necrosis, and an acute inflammatory response. Re-epithelialization occurs once viral replication is restricted, almost always in the absence of a scar.

Analysis of the DNA from sequential isolates of HSV or from isolates from multiple infected ganglia in any one individual has revealed similar, if not identical, restriction endonuclease or DNA sequence patterns in most persons. As more sensitive genomic technologies are developed, evidence of multiple strains of the same subtype is increasingly being reported. For example, infection of individual neurons with multiple strains of drug-susceptible and drug-resistant virus in severely immunosuppressed patients indicates that ganglia can be reseeded during chronic infection. Because exposure to mucosal shedding is relatively common during a person's lifetime, current data suggest that exogenous infection with different strains of the same subtype does occur. The role strain variation plays in the varied reactivation pattern of disease is unknown.

IMMUNITY

Host responses influence the acquisition of HSV disease, the severity of infection, resistance to the development of latency, the maintenance of latency, and the frequency of recurrences. Both antibody-mediated and cell-mediated reactions are clinically important. Immunocompromised patients with defects in cell-mediated immunity experience more severe and more extensive HSV infections than those with deficits in humoral immunity, such as agammaglobulinemia. Experimental ablation of lymphocytes indicates that T cells play a major role in preventing lethal disseminated disease, although antibodies help reduce titers of virus in neural tissue. Some clinical manifestations of HSV appear to be related to the host immune response (e.g., stromal opacities associated with recurrent herpetic keratitis). The surface viral glycoproteins have been shown to be targets of antibodies that mediate neutralization and immune-mediated cytolysis (antibody-dependent cell-mediated cytotoxicity [ADCC]). Monoclonal antibodies to HSV viral glycoproteins have, in experimental infections, conferred protection against subsequent neurologic disease or ganglionic latency, and reduced subsequent reactivation in animals model experiments. Human studies of monoclonal antibodies are underway. Multiple cell populations, including neutrophils, macrophages, and a variety of T lymphocytes, play a role in host defenses against HSV infections, as do lymphokines generated by T lymphocytes. In animals, passive transfer of primed lymphocytes confers protection from subsequent HSV challenge. Maximal protection usually requires the activation of multiple T-cell subpopulations, including cytotoxic T cells and T cells responsible for delayed hypersensitivity. The latter may confer protection by the antigen-stimulated release of lymphokines (e.g., IFNs), which in turn have a direct antiviral effect and both activate and enhance a variety of specific and nonspecific effector cells. Cellular and humoral immune responses to HSV have been detected both in human ganglia and in mucosal tissue at the site of reactivation. The HSV virion contains a variety of genes that are directed at the inhibition of host responses. These include gene *ICP47*, which can bind to the cellular transporter-activating protein TAP-1 and reduce the ability of this protein to bind HSV peptides to human leukocyte antigen class I, thereby reducing recognition of viral proteins by cytotoxic T cells of the host. This effect can be overcome by the addition of IFN-γ, but this reversal requires 24–48 h; thus, the virus has time to replicate and invade other host cells. Entry of infectious HSV-1 and HSV-2 inhibits several signaling pathways of both CD4+ and CD8+ T cells, leading to their functional impairment in killing and influencing the spectrum of their cytokine secretion.

HSV-specific CD8+ T-cell responses appear to be an important component in viral clearance from lesions. Immunosuppressed patients with frequent and prolonged HSV lesions have fewer functional CD8+ T cells directed at HSV. HSV-specific CD8+ T cells have been shown to persist in the genital skin at the dermal–epidermal junction contiguous to nerve endings for months after lesion resolution. Even during clinical quiescence, these CD8+ T cells make both antiviral and cytotoxic proteins indicative of immune surveillance. These resident memory CD8+ T cells appear to be "first responders" capable of controlling viral reactivation at the site of viral release into the dermis. The communication with surrounding keratinocytes using cytokine release of interferon gamma initiates antiviral resistance mechanisms to HSV epithelial cell infection. Innate immunity also influences HSV infection. Severe HSV-1 infections occur in individuals with natural killer (NK) cell defects. Severe familial HSV-1 infections are associated with TLR3 polymorphisms, and the TLR signaling protein UNC93B and TLR2 polymorphisms influence disease reactivation. This rapid "on and off" interplay between the virus and the host helps explain the variability in clinical disease severity between episodes in any single individual. Differences of 30–60 min in host responses can result in 100- to 1000-fold differences in viral levels and can determine whether an episode of disease is subclinical or clinical.

There is a strong association between the magnitude of the CD8+ T-lymphocyte response and the clearance of virus from genital lesions. The location, effectiveness, and longevity of the CD8+ T lymphocytes (and other influencers of immune effector functions such as natural killer or CD4+ T-cell responses) may be important in the expression of disease and the likelihood of transmission over time.

■ EPIDEMIOLOGY

Seroepidemiologic studies have documented HSV infections worldwide. The global prevalence of HSV-1 is estimated at 66% of the population (3.7 billion people), while 13.2% or 492 million people aged 15–49 were estimated to live with HSV-2. The past 15 years have shown that the prevalence of HSV-2 is even higher in the developing than in the developed world. In sub-Saharan Africa, HSV-2 seroprevalence among pregnant women may approach 60%, and annual acquisition rates among teenage girls may verge on 20%. The global incidence has been estimated at ~23.9 million infections per year, with 491.5 million infected persons worldwide. As in the developed world, the rate of HSV-2 coital acquisition as well as the serologic prevalence are higher among women than among men. Most of this HSV-2 acquisition is preceded by acquisition of HSV-1; the frequency of genital HSV-1 in middle- and low-income countries is low at present.

Infection with HSV-1 is acquired more frequently and earlier in life than infection with HSV-2. From 70 to 90% of adults have antibodies to HSV-1 by the fifth decade of life. In populations of low socioeconomic status, most persons acquire HSV-1 infection before the third decade of life. Antibodies to HSV-2 are not detected routinely until puberty. Antibody prevalence rates correlate with past sexual activity and vary greatly among different population groups. There is evidence that the prevalence of HSV-2 has decreased slightly over the past decade or so in the United States. Serosurveys indicate that 15–20% of the U.S. population has antibodies to HSV-2. In most routine obstetric and family planning clinics, 15–30% of women have HSV-2 antibodies, although only 10% of those who are seropositive for HSV-2 report a history of genital lesions. As many as 50% of heterosexual adults attending sexually transmitted disease clinics have antibodies to HSV-2. A wide variety of serologic surveys has catalogued the widespread epidemic of HSV-2 in Central America, South America, and Africa. In Africa, HSV-2 seroprevalence has ranged from 40 to 70% in obstetric and other sexually experienced populations. Antibody prevalence rates average ~5–10% higher among women than among men.

Many studies continue to show that both incident and—more importantly—prevalent HSV-2 infection enhances the acquisition rate of HIV-1. More specifically, HSV-2 infection is associated on a population basis with a two- to fourfold increase in HIV-1 acquisition. This association has been amply demonstrated in heterosexual men and women in both the developed and developing worlds. Mathematical models suggest that ~33–50% of HIV-1 infections may be attributable to HSV-2 both in men who have sex with men (MSM) and in heterosexual women in sub-Saharan Africa. Epidemiologically, regions of the world with high HSV-2 prevalence and selected populations within such regions have a higher population-based incidence of HIV-1.

HSV-2 facilitates the spread of HIV into low-risk populations; prevalent HSV-2 appears to increase the risk of HIV infection by seven- to ninefold on a per-coital basis. In addition, HSV-2 is more frequently reactivated in and transmitted by persons co-infected with HIV-1 than in persons not co-infected. Thus, most areas of the world with a high HIV-1 prevalence also have a high HSV-2 prevalence. The shedding of HIV-1 virions from herpetic lesions in the genital region facilitates the spread of HIV through sexual contact. HSV-2 reactivation is associated with a localized persistent inflammatory response consisting of high concentrations of CCR5-enriched CD4+ T cells as well as inflammatory dendritic cells in the submucosa of the genital skin. These cells can support HIV infection and replication and thus are likely to account for the increased risk of HIV acquisition among persons with genital herpes. Unfortunately, antiviral therapy does not reduce this subclinical postreactivation inflammation, probably because of the inability of current antiviral agents to prevent the release of small amounts of HSV antigen into the genital mucosa.

Several studies suggest that many cases of "asymptomatic" genital HSV-2 infection are, in fact, simply unrecognized or confined to anatomic regions of the genital tract that are not easily visualized. Asymptomatic seropositive persons shed virus on mucosal surfaces almost as frequently as do those with symptomatic disease. This large reservoir of unidentified carriers of HSV-2 and the frequent asymptomatic reactivation of the virus from the genital tract have fostered the continued spread of genital herpes throughout the world.

HSV infections occur throughout the year. Transmission can result from contact with persons who have active ulcerative lesions or with persons who have no clinical manifestations of infection but who are shedding HSV from mucocutaneous surfaces. HSV reactivation on genital skin and mucosal surfaces is common. Most HSV-1 and HSV-2 episodes last 2–6 h; thus, replication of the virus and clearance by the host are rapid. Even with once-daily sampling, HSV DNA can be detected on 20% of days by polymerase chain reaction (PCR). Corresponding figures for HSV-1 in oral secretions are similar. Rates of shedding are highest during the initial years after acquisition, with viral shedding occurring on as many as 30–50% of days during this period. Immunosuppressed patients shed HSV from mucosal sites at an even higher frequency (20–80% of days). These high rates of mucocutaneous reactivation suggest that exposure to HSV from sexual or other close contact (kissing, sharing of glasses or silverware) is common and help explain the continuing spread and high seroprevalence of HSV infections worldwide. Reactivation rates vary widely among individuals. Among people living with HIV, a low CD4+ T-cell count and a high HIV-1 load are associated with increased rates of HSV reactivation. Daily antiviral chemotherapy for HSV-2 infection can reduce shedding rates but does not eliminate shedding, as measured by PCR or culture.

■ CLINICAL SPECTRUM

HSV has been isolated from nearly all visceral and mucocutaneous sites. The clinical manifestations and course of HSV infection depend on the anatomic site involved, the age and immune status of the host, and the antigenic type of the virus. Primary HSV infections (i.e., first infections with either HSV-1 or HSV-2 in which the host lacks HSV antibodies in acute-phase serum) are frequently accompanied by systemic signs and symptoms. Compared with recurrent episodes, primary infections, which involve both mucosal and extramucosal sites, are characterized by a longer duration of symptoms and virus isolation from lesions. The incubation period ranges from 1 to 26 days (median, 6–8 days). Both viral subtypes can cause genital and oral–facial infections, and the infections caused by the two subtypes are clinically indistinguishable. However, the frequency of reactivation of infection is influenced by anatomic site and virus type. Genital HSV-2 infection is twice as likely to reactivate and recurs 8–10 times more frequently than genital HSV-1 infection. Conversely, oral–labial HSV-1 infection recurs more frequently than oral–labial HSV-2 infection. Asymptomatic shedding rates follow the same pattern.

Oral–Facial Infections Gingivostomatitis and pharyngitis are the most common clinical manifestations of first-episode HSV-1 infection, whereas recurrent herpes labialis is the most common clinical manifestation of reactivation HSV-1 infection. HSV pharyngitis and gingivostomatitis usually result from primary infection and are most common among children and young adults. Clinical symptoms and signs, which include fever, malaise, myalgias, inability to eat, irritability, and cervical adenopathy, may last 3–14 days. Lesions may involve the hard and soft palate, gingiva, tongue, lip, and facial area. HSV-1 or HSV-2 infection of the pharynx usually results in exudative or ulcerative lesions of the posterior pharynx and/or tonsillar pillars. Lesions of the tongue, buccal mucosa, or gingiva may occur later in the course in one-third of cases. Fever lasting 2–7 days and cervical adenopathy are common. It can be difficult to differentiate HSV pharyngitis clinically from bacterial pharyngitis, *Mycoplasma pneumoniae* infections, and pharyngeal ulcerations of noninfectious etiologies (e.g., Stevens-Johnson syndrome). No substantial evidence suggests that reactivation of oral–labial HSV infection is associated with symptomatic recurrent pharyngitis.

Reactivation of HSV from the trigeminal ganglia may be associated with asymptomatic virus excretion in the saliva, development of intraoral mucosal ulcerations, or herpetic ulcerations on the vermilion border of the lip or external facial skin. About 50–70% of seropositive patients undergoing trigeminal nerve-root decompression and 10–15% of those undergoing dental extraction develop oral–labial HSV infection a median of 3 days after these procedures. Clinical differentiation of intraoral mucosal ulcerations due to HSV from aphthous, traumatic, or drug-induced ulcerations is difficult.

In immunosuppressed patients, HSV infection may extend into mucosal and deep cutaneous layers. Friability, necrosis, bleeding, severe pain, and inability to eat or drink may result. The lesions of HSV mucositis are clinically similar to mucosal lesions caused by cytotoxic drug therapy, trauma, or fungal or bacterial infections, and co-infections are common. Persistent ulcerative HSV infections are among the most common infections in patients with AIDS. HSV and *Candida* infections often occur concurrently. Systemic antiviral therapy speeds the rate of healing and relieves the pain of mucosal HSV infections in immunosuppressed patients. The frequency of HSV reactivation during the early phases of transplantation or induction chemotherapy is high (50–90%), and prophylactic systemic antiviral agents such as intravenous (IV) acyclovir and penciclovir or the oral congeners of these drugs are used to reduce reactivation rates. Patients with atopic eczema may also develop severe oral–facial HSV infections (*eczema herpeticum*), which may rapidly involve extensive areas of skin and occasionally disseminate to visceral organs. Extensive eczema herpeticum has resolved promptly with the administration of IV acyclovir. Erythema multiforme may also be associated with HSV infections **(see Figs. 59-9 and A1-24)**; some evidence suggests that HSV infection is the precipitating event in ~75% of cases of cutaneous erythema multiforme. HSV antigen has been demonstrated both in circulatory immune complexes and in skin lesion biopsy samples from these cases. Patients with severe HSV-associated erythema multiforme are candidates for chronic suppressive oral antiviral therapy.

HSV-1 and varicella-zoster virus (VZV) have been implicated in the etiology of Bell's palsy (flaccid paralysis of the mandibular portion of the facial nerve). Some but not all trials have documented quicker resolution of facial paralysis with the prompt initiation of antiviral therapy, with or without glucocorticoids. However, other trials have shown little benefit. There are advantages to the use of both antiviral drugs and glucocorticoids for moderate to severe Bell's palsy. Some experts feel glucocorticoids alone are preferred for mild disease.

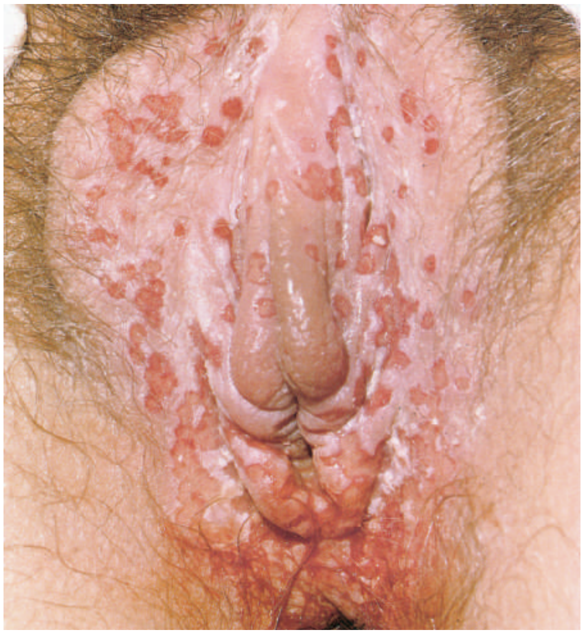

FIGURE 197-1 Genital herpes: primary vulvar infection, with multiple, extremely painful, punched-out, confluent, shallow ulcers on the edematous vulva and perineum. Micturition is often very painful. Associated inguinal lymphadenopathy is common. *(Reprinted with permission from K Wolff et al: Fitzpatrick's Color Atlas & Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005.)*

Genital Infections First-episode primary genital herpes is characterized by fever, headache, malaise, and myalgias. Pain, itching, dysuria, vaginal and urethral discharge, and tender inguinal lymphadenopathy are the predominant local symptoms. Widely spaced bilateral lesions of the external genitalia are characteristic **(Fig. 197-1)**. Lesions may be present in varying stages, including vesicles, pustules, or painful erythematous ulcers. The cervix and urethra are involved in >80% of women with first-episode infections. First episodes of genital herpes in patients who have had prior HSV-1 infection are occasionally associated with systemic symptoms: prior HSV-1 infection is associated with faster healing than true primary genital herpes. Detection of HSV DNA in serum has been found in ~30% of cases of true primary genital herpes. The clinical courses of acute first-episode genital herpes are similar for HSV-1 and HSV-2 infection. However, the recurrence rates of genital disease differ with the viral subtype: the 12-month recurrence rates among patients with first-episode HSV-2 and HSV-1 infections are ~90% and ~55%, respectively (median number of recurrences, 4 and <1, respectively). Recurrence rates for genital HSV-2 infections vary greatly among individuals and over time within the same individual. HSV has been isolated from the urethra and urine of men and women without external genital lesions. A clear mucoid discharge and dysuria are characteristics of symptomatic HSV urethritis. HSV has been isolated from the urethra of 5% of women with the dysuria–frequency syndrome. Occasionally, HSV genital tract disease is manifested by endometritis and salpingitis in women and by prostatitis in men. About 15% of cases of HSV-2 acquisition are associated with nonlesional clinical syndromes, such as aseptic meningitis, cervicitis, or urethritis. **A more complete discussion of the differential diagnosis of genital herpes is presented in Chap. 141**.

Both HSV-1 and HSV-2 can cause symptomatic or asymptomatic rectal and perianal infections. HSV proctitis is usually associated with rectal intercourse. However, subclinical perianal shedding of HSV is detected in women and men who report no rectal intercourse. This phenomenon is due to the establishment of latency in the sacral dermatome or sacral autonomic ganglia from prior genital tract infection, with subsequent reactivation in epithelial cells in the perianal region. Such reactivations are often subclinical. Symptoms of HSV proctitis include anorectal pain, anorectal discharge, tenesmus, and constipation. Sigmoidoscopy reveals ulcerative lesions of the distal 10 cm of the rectal mucosa. Rectal biopsies show mucosal ulceration, necrosis, polymorphonuclear and lymphocytic infiltration of the lamina propria, and (in occasional cases) multinucleated intranuclear inclusion-bearing cells. Perianal herpetic lesions are also found in immunosuppressed patients receiving cytotoxic therapy. Extensive perianal herpetic lesions and/or HSV proctitis is common among patients with HIV infection.

The recent outbreak of Mpox infections globally has made the differentiation of HSV anal rectal infection from Mpox infection of clinical and therapeutic importance. PCR-based assays clearly distinguish between the two entities. Acyclovir should be used for HSV-2 infection and tecovirimat for monkeypox infection.

Herpetic Whitlow Herpetic whitlow—HSV infection of the finger—may occur as a complication of primary oral or genital herpes by inoculation of virus through a break in the epidermal surface or by direct introduction of virus into the hand through occupational or some other type of exposure; either viral type may be isolated from the lesion. Clinical signs and symptoms include abrupt-onset edema, erythema, and localized tenderness of the infected finger. Vesicular or pustular lesions of the fingertip that are indistinguishable from lesions of pyogenic bacterial infection are seen. Fever, lymphadenitis, and epitrochlear and axillary lymphadenopathy are common. The infection may recur. Prompt diagnosis (to avoid unnecessary and potentially exacerbating surgical therapy and/or transmission) is essential. Antiviral therapy is usually recommended (see below).

Herpes Gladiatorum HSV may infect almost any area of skin. Mucocutaneous HSV infections of the thorax, ears, face, and hands have been described among wrestlers. Transmission of these infections is facilitated by trauma to the skin sustained during wrestling. Outbreaks of HSV among competitive wrestlers have illustrated the importance of prompt diagnosis and therapy to contain the spread of this infection.

Eye Infections HSV infection of the eye is the most common cause of corneal blindness in the United States. HSV keratitis presents as an acute onset of pain, blurred vision, chemosis, conjunctivitis, and characteristic dendritic lesions of the cornea. Use of topical glucocorticoids may exacerbate symptoms and lead to involvement of deep structures of the eye. Debridement, topical antiviral treatment, and/or IFN therapy hasten healing. However, recurrences are common, and the deeper structures of the eye may sustain immunopathologic injury. Stromal keratitis due to HSV appears to be related to T-cell–dependent destruction of deep corneal tissue. An HSV-1 epitope that is autoreactive with T cell–targeting corneal antigens has been postulated to be a factor in this infection. Chorioretinitis, usually a manifestation of disseminated HSV infection, may occur in neonates or in patients with HIV infection. HSV and VZV can cause acute necrotizing retinitis as an uncommon but severe manifestation. While VZV infection is the most common cause of acute retinal necropsy, both HSV-1 and HSV-2 may also be associated with this syndrome. Emergent ophthalmology consultation is recommended as residual blindness may occur; both systemic and intravitical antiviral therapy is recommended.

Central and Peripheral Nervous System Infections HSV accounts for 10–20% of all cases of sporadic viral encephalitis in the United States. The estimated incidence is ~2.3 cases per 1 million persons per year. Cases are distributed throughout the year, and the age distribution appears to be biphasic, with peaks at 5–30 and >50 years of age. HSV-1 causes >95% of cases.

The pathogenesis of HSV encephalitis varies. In children and young adults, primary HSV infection may result in encephalitis; presumably, exogenously acquired virus enters the CNS by neurotropic spread from the periphery via the olfactory bulb. However, most adults with HSV encephalitis have clinical or serologic evidence of mucocutaneous HSV-1 infection before the onset of CNS symptoms. In ~25% of the cases examined, the HSV-1 strains from the oropharynx and brain tissue of the same patient differ; thus, some cases may result from reinfection with another strain of HSV-1 that reaches the CNS. Two theories have been proposed to explain the development of actively replicating HSV in localized areas of the CNS in persons whose ganglionic and CNS isolates are similar. Reactivation of latent HSV-1 infection in trigeminal or autonomic nerve roots may be associated with extension of virus into the CNS via nerves innervating the middle cranial fossa. HSV DNA has been demonstrated by DNA hybridization in brain tissue obtained at autopsy—even from healthy adults. Thus, reactivation of long-standing latent CNS infection may be another mechanism for the development of HSV encephalitis. Recent studies have identified genetic polymorphisms among families with a high frequency of HSV encephalitis. Peripheral-blood mononuclear cells, fibroblasts, and neurons from these patients (predominantly children) appear to secrete reduced levels of IFN in response to HSV. Genetic mutations in *TLR3* documented in patients with HSV encephalitis suggest that some cases of sporadic HSV encephalitis may be related to host genetic determinants.

The clinical hallmark of HSV encephalitis has been the acute onset of fever and focal neurologic symptoms and signs, especially in the temporal lobe **(Fig. 197-2)**; gadolinium enhancing lesions are seen in the temporal lobe by magnetic resonance imaging (MRI). Clinical differentiation of HSV encephalitis from other viral encephalitides, focal infections, or noninfectious processes is difficult. Elevated cerebrospinal fluid (CSF) protein levels, leukocytosis (predominantly lymphocytes), and red blood cell counts due to hemorrhagic necrosis are common in HSV encephalitis. While brain biopsy has been the gold standard for defining HSV encephalitis, a highly sensitive and specific PCR for detection of HSV DNA in CSF has largely replaced biopsy for defining HSV CNS infection. Although titers of antibody to HSV in CSF and serum increase in most cases of HSV encephalitis, they rarely do so earlier than 10 days into the illness and, therefore, although useful in retrospect, generally are not helpful in establishing an early clinical diagnosis. In rare cases, demonstration of HSV antigen, HSV DNA, or HSV replication in brain tissue obtained by biopsy is highly sensitive; examination of such tissue also provides the opportunity to identify alternative, potentially treatable causes of encephalitis. Antiviral therapy with acyclovir reduces the rate of death from HSV encephalitis. Most authorities recommend the administration of IV

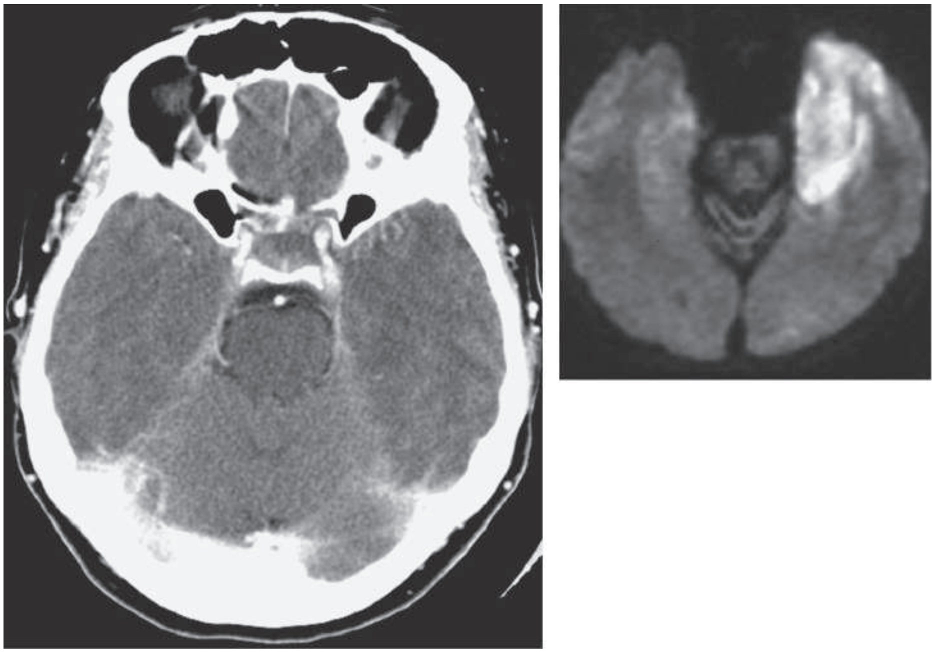

FIGURE 197-2 Computed tomography and diffusion-weighted magnetic resonance imaging scans of the brain of a patient with left-temporal-lobe herpes simplex virus encephalitis.

acyclovir to patients with presumed HSV encephalitis until the diagnosis is confirmed or an alternative diagnosis is made. All confirmed cases should be treated with IV acyclovir (30 mg/kg per day in three divided doses for 14–21 days). After the completion of therapy, the clinical recurrence of encephalitis requiring more treatment has been reported. For this reason, some authorities prefer to treat initially for 21 days, and many continue therapy until HSV DNA has been eliminated from the CSF. Even with therapy, neurologic sequelae are common, especially among persons >50 years of age.

HSV DNA has been detected in CSF from 3 to 15% of persons presenting to the hospital with aseptic meningitis. HSV meningitis, which is usually seen in association with primary genital HSV infection, is an acute, self-limited disease manifested by headache, fever, and mild photophobia and lasting 2–7 days. Lymphocytic pleocytosis in the CSF is characteristic. Neurologic sequelae of HSV meningitis are rare. HSV is the most commonly identified cause of recurrent lymphocytic meningitis (*Mollaret's meningitis*). Demonstration of HSV antibodies in CSF or persistence of HSV DNA in CSF can establish the diagnosis. For persons with frequent recurrences of HSV meningitis, daily antiviral therapy has reduced the frequency of recurrent episodes of symptomatic meningitis.

Autonomic nervous system dysfunction, especially of the sacral region, has been reported in association with both HSV and VZV infections. Numbness, tingling of the buttocks or perineal areas, urinary retention, constipation, CSF pleocytosis, and (in males) impotence may occur. Symptoms appear to resolve slowly over days or weeks. Occasionally, hypoesthesia and/or weakness of the lower extremities persists for many months. Transitory hypoesthesia of the area of skin innervated by the trigeminal nerve and vestibular system dysfunction (as measured by electronystagmography) are the predominant signs of disease. Rarely, transverse myelitis, manifested by a rapidly progressive symmetric paralysis of the lower extremities or Guillain-Barré syndrome, follows HSV infection. Similarly, PNS involvement (Bell's palsy) or cranial polyneuritis may be related to reactivation of HSV-1 infection.

There is increasing experimental evidence suggesting an association between herpesvirus pathogens, specifically HSV-1, and the development of sporadic Alzheimer's disease (AD). HSV-1 DNA is detected in brain tissue of patients with AD, and epidemiologically, HSV-1 antibodies are a significant risk factor for later AD onset. A wide variety of models of AD indicate that HSV-1 infection can induce neuronal death, tau phosphorylation, and intracellular expression of isoforms of amyloid precursor protein cleavage products that produce multicellular-like plaque structures associated with AD. There are no cogent data to indicate antiviral therapy would be of benefit to anyone with AD. An adequately powered prospective study of prolonged antiviral therapy has not been conducted.

HSV Lymphadenitis HSV-1 or HSV-2 lymphadenitis has been reported, with increasing frequency especially among patients with chronic lymphocytic leukemia. This condition represents a diagnostic dilemma often confused with lymphomatous transformation and occurs absent of oral or genital lesions. One histologic hint between HSV and lymphoblastic transformation is the presence of necrosis in biopsy tissue in HSV lymphadenitis. Isolation of HSV or demonstration of HSV DNA in tissue confers the diagnosis. Concomitant HSV viremia may also be present. Intravenous acyclovir has been associated with resolution of infection and clinical improvement.

Visceral Infections HSV infection of visceral organs usually results from viremia, and multiple-organ involvement is common. Occasionally, however, the clinical manifestations of HSV infection involve only the esophagus, lung, or liver. HSV esophagitis may result from direct extension of oral–pharyngeal HSV infection into the esophagus or may occur de novo by reactivation and spread of HSV to the esophageal mucosa via the vagus nerve. The predominant symptoms of HSV esophagitis are odynophagia, dysphagia, substernal pain, and weight loss. Multiple oval ulcerations appear on an erythematous base with or without a patchy white pseudomembrane. The distal esophagus is most commonly involved. With extensive disease, diffuse friability may spread to the entire esophagus. Neither endoscopic nor barium examination can reliably differentiate HSV esophagitis from *Candida* esophagitis or from esophageal ulcerations due to thermal injury, radiation, or corrosives. Endoscopically obtained secretions—for cytologic examination and culture or DNA detection by PCR—provide the most useful material for diagnosis. Systemic antiviral therapy usually reduces the severity and duration of symptoms and heals esophageal ulcerations.

HSV pneumonitis is uncommon except in severely immunosuppressed patients and may result from extension of herpetic tracheobronchitis into lung parenchyma. Focal necrotizing pneumonitis usually ensues. Hematogenous dissemination of virus from sites of oral or genital mucocutaneous disease may also occur, producing bilateral interstitial pneumonitis. Bacterial, fungal, and parasitic pathogens are commonly present in HSV pneumonitis. The mortality rate from untreated HSV pneumonia in immunosuppressed patients is high (>80%). HSV has also been isolated from the lower respiratory tract of persons with acute respiratory distress syndrome and prolonged intubation. Most authorities believe that the presence of HSV in tracheal aspirates in such settings is due to reactivation of HSV in the tracheal region and localized tracheitis in persons with long-term intubation. Such patients should be evaluated for extension of HSV infection into the lung parenchyma. While retrospective reviews of HSV tracheitis in intensive care unit patients suggest benefit from antiviral therapy, well-powered controlled trials assessing the role of antiviral agents used against HSV in ventilation-associated morbidity and mortality have not been conducted. The role of lower respiratory tract HSV infection in overall rates of morbidity and mortality associated with these conditions is unclear. HSV is an uncommon cause of hepatitis in immunocompetent patients. HSV infection of the liver is associated with fever, abrupt elevations of bilirubin and serum aminotransferase levels, and leukopenia (<4000 white blood cells/μL). Disseminated intravascular coagulation may also develop.

Other reported complications of HSV infection include monarticular arthritis, adrenal necrosis, idiopathic thrombocytopenia, and glomerulonephritis. Disseminated HSV infection in immunocompetent patients is rare. In immunocompromised patients, burn patients, or malnourished individuals, HSV occasionally disseminates to other visceral organs, such as the adrenal glands, pancreas, small and large intestines, and bone marrow. Dissemination, even recurrent dissemination, is being increasingly recognized in persons with chronic lymphocytic leukemia. Rarely, primary HSV infection in pregnancy disseminates and may be associated with the death of both mother and fetus. This uncommon event is usually related to the acquisition of primary infection in the third trimester. Disseminated HSV infection is best detected by the presence of HSV DNA in plasma or blood.

Neonatal HSV Infections Of all HSV-infected populations, neonates (infants <6 weeks) have the highest frequency of visceral and/or CNS infection. Without therapy, the overall rate of death from neonatal herpes is 65%; <10% of neonates with CNS infection develop normally. Although skin lesions are the most commonly recognized features of disease, many infants do not develop lesions at all or do so only well into the course of disease. Neonatal infection is usually acquired perinatally from contact with infected genital secretions at delivery. Congenitally infected infants have been reported. Of neonatal HSV infections, 30–50% are due to HSV-1 and 50–70% to HSV-2. A recent review of U.S. Medicaid databases indicates a frequency of ~1000 cases of neonatal HSV yearly, a rate of between 1 per 2000 and 1 per 3000 live births. The risk of developing neonatal HSV infection is 10 times higher for an infant born to a mother who has recently acquired HSV than for other infants. Neonatal HSV-1 infections may also be acquired through postnatal contact with immediate family members who have symptomatic or asymptomatic oral–labial HSV-1 infection or through nosocomial transmission within the hospital. All neonates with presumed herpes should be treated with IV acyclovir and then placed on maintenance oral antiviral therapy for the first 6–12 months of life. Antiviral chemotherapy with high-dose IV acyclovir (60 mg/kg per day) has reduced the mortality rate from neonatal herpes to ~15%.

However, rates of morbidity, especially among infants with HSV-2 infection involving the CNS, are still very high.

HSV in Pregnancy In the United States, 21% of all pregnant women and 51% of non-Hispanic black pregnant women are seropositive for HSV-2. However, the risk of mother-to-child transmission of HSV in the perinatal period is highest when the infection is acquired near the time of labor—that is, in previously HSV-seronegative women. The clinical manifestations of recurrent genital herpes—including the frequency of subclinical versus clinical infection, the duration of lesions, pain, and constitutional symptoms—are similar in pregnant and nonpregnant women. Recurrences increase in frequency over the course of pregnancy. However, when women are seropositive for HSV-2 at the outset of pregnancy, no effect on neonatal outcomes (including birth weight and gestational age) is seen. First-episode infections in pregnancy have more severe consequences for mother and infant. Maternal visceral dissemination during the third trimester occasionally occurs, as does premature birth or intrauterine growth retardation. The acquisition of primary disease in pregnancy, whether related to HSV-1 or HSV-2, carries the risk of transplacental transmission of virus to the neonate and can result in spontaneous abortion, although this outcome is relatively uncommon. For newly acquired genital HSV infection during pregnancy, most authorities recommend treatment with acyclovir (400 mg three times daily) or valacyclovir (500–1000 mg twice daily) for 7–10 days. However, the impact of this intervention on transmission is unknown. The high HSV-2 prevalence rate in pregnancy and the low incidence of neonatal disease in these women (1 case per 6000–20,000 live births) indicate that only a few infants are at risk of acquiring HSV. Therefore, cesarean section is not warranted for all women with recurrent genital disease. Because intrapartum transmission of infection accounts for the majority of cases, abdominal delivery need be considered only for women who are shedding HSV at delivery. Several studies have shown no correlation between recurrence of viral shedding before delivery and viral shedding at term. Hence, weekly virologic monitoring and amniocentesis are not recommended.

The frequency of transmission from mother to infant is markedly higher among women who acquire HSV near term (30–50%) than among those in whom HSV-2 infection is reactivated at delivery (<1%). Although maternal antibody to HSV-2 is protective, antibody to HSV-1 offers little or no protection against neonatal HSV-2 infection. Primary genital infection with HSV-1 leads to a particularly high risk of transmission during pregnancy and accounts for an increasing proportion of neonatal HSV cases. Moreover, during reactivation, HSV-1 appears more transmissible to the neonate than HSV-2. Only 2% of women who are seropositive for HSV-2 have HSV-2 isolated from cervical secretions at delivery, and only 1% of infants exposed in this manner develop infection, presumably because of the protective effects of maternally transferred antibodies and perhaps lower viral titers during reactivation. Despite the low frequency of transmission of HSV in this setting, 30–50% of infants with neonatal HSV are born to mothers with established genital herpes.

Isolation of HSV by cervicovaginal swab at the time of delivery is the greatest risk factor for intrapartum HSV transmission (relative risk = 346); however, culture-negative, PCR-positive cases of intrapartum transmission are well described. New acquisition of HSV (odds ratio [OR] = 49), isolation of HSV-1 versus HSV-2 (OR = 35), cervical versus vulvar HSV detection (OR = 15), use of fetal scalp electrodes (OR = 3.5), and young maternal age confer further risk of transmission, whereas cesarean delivery is protective (OR = 0.14). Physical examination poorly predicts the absence of shedding, and PCR far exceeds culture in terms of sensitivity and speed. Therefore, PCR detection at the onset of labor should be used to aid clinical decision-making for women with HSV-2 antibody. Because cesarean section appears to be an effective means of reducing maternal–fetal transmission, patients with recurrent genital herpes should be encouraged to come to the hospital early at the time of delivery for careful examination of the external genitalia and cervix as well as collection of a swab sample for viral isolation. Women who have no evidence of lesions can have a vaginal delivery. The presence of active lesions on the cervix or external genitalia is an indication for cesarean delivery.

If first-episode exposure has occurred (e.g., if HSV serologies show that the mother is seronegative or if the mother is HSV-1-seropositive and the isolate at delivery is found to be HSV-2), many authorities would initiate antiviral therapy for the infant with IV acyclovir. At a minimum, samples for viral cultures and PCR should be obtained from the throat, nasopharynx, eyes, and rectum of these infants immediately and at 5- to 10-day intervals. Lethargy, skin lesions, or fever should be evaluated promptly. All infants from whom HSV is isolated 24 h after delivery should be treated with IV acyclovir at recommended doses.

■ DIAGNOSIS

Both clinical and laboratory criteria are useful for diagnosing HSV infections. A clinical diagnosis can be made accurately when characteristic multiple vesicular lesions on an erythematous base are present. However, herpetic ulcerations may resemble skin ulcerations of other etiologies and considerable overlap including coinfection between HSV and Mpox infections exist. Mucosal HSV infections may also present as urethritis or pharyngitis without cutaneous lesions. Thus, laboratory studies to confirm the diagnosis and to guide therapy are recommended. While staining of scrapings from the base of the lesions with Wright's, Giemsa's (Tzanck preparation), or Papanicolaou's stain to detect giant cells or intranuclear inclusions of *Herpesvirus* infection is a well-described procedure, few clinicians are skilled in this technique, the sensitivity of staining is low (<30% for mucosal swabs), and these cytologic methods do not differentiate between HSV and VZV infections.

HSV infection is best confirmed in the laboratory by detection of virus, viral antigen, or viral DNA in scrapings from lesions. HSV DNA detection by PCR is the most sensitive laboratory technique for detecting mucosal or visceral HSV infections and is the recommended test for laboratory confirmation of a diagnosis. HSV causes a discernible cytopathic effect in a variety of cell culture systems, and this effect can be identified within 48–96 h after inoculation. Spin-amplified culture with subsequent staining for HSV antigen has shortened the time needed to identify HSV to <24 h. Culture is indicated when antiviral sensitivity testing is required. The sensitivity of all detection methods depends on the stage of the lesions (with higher sensitivity for vesicular than for ulcerative lesions), on whether the patient has a first or a recurrent episode of the disease (with higher sensitivity in first than in recurrent episodes), and on whether the sample is from an immunosuppressed or an immunocompetent patient (with more antigen or DNA in immunosuppressed patients). Laboratory confirmation permits subtyping of the virus; information on subtype may be useful epidemiologically and may help to predict the frequency of reactivation after first-episode oral–labial or genital HSV infection.

Both type-specific and type-common antibodies to HSV develop during the first several weeks after infection and persist indefinitely. Serologic assays with whole-virus antigen preparations, such as complement fixation, neutralization, indirect immunofluorescence, passive hemagglutination, radioimmunoassay, and enzyme-linked immunosorbent assay, can differentiate uninfected (seronegative) persons from those with past HSV-1 or HSV-2 infection, but they do not reliably distinguish between the two viral subtypes or help establish an active clinical diagnosis of HSV-2. Serologic assays that identify antibodies to the type-specific glycoprotein G of the two viral subtypes (G1 and G2) are available commercially and can be used to distinguish between the human antibody responses to HSV-1 and HSV-2. Point-of-care assays that provide results from capillary blood or serum during a clinic visit are available. False-positive and false-negative assays do occur, and there is a need for commercially available assays to distinguish prevalent HSV-2 infection, especially on the backgroiund of earlier acquired HSV-1. A western blot assay that can detect several HSV type-specific proteins can also be used and is, at present, the best confirmatory assay. The presence of type-specific HSV-2 antibody implies past HSV-2 infection—i.e., latent infection, likely subclinical reactivation, increased risk of HSV-1 acquisition, and potential transmission to others.

Acute- and convalescent-phase serum samples can be useful in demonstrating seroconversion during primary HSV-1 or HSV-2 infection. However, few available tests report titers, and increases in index values do not reflect first episodes in all patients. Serologic assays based on type-specific proteins should be used to identify asymptomatic carriers of HSV-1 or HSV-2. No reliable IgM method for defining acute HSV infection is available.

Several studies have shown that persons with previously unrecognized HSV-2 infection can be taught to identify symptomatic reactivations. Individuals seropositive for HSV-2 should be told about the high frequency of subclinical reactivation on mucosal surfaces that are not visible to the eye (e.g., cervix, urethra, perianal skin) or in microscopic ulcerations that may not be clinically symptomatic. Transmission of infection during such episodes is well established. HSV-2-seropositive persons should be educated about the high likelihood of subclinical shedding and the role that condoms (male or female) may play in reducing transmission. Antiviral therapy with valacyclovir (500 mg once daily) has been shown to reduce but not eliminate the transmission of HSV-2 between sexual partners.

TREATMENT

Herpes Simplex Virus Infections

Many aspects of mucocutaneous and visceral HSV infections are amenable to antiviral chemotherapy. For mucocutaneous infections, acyclovir and its congeners famciclovir and valacyclovir have been the mainstays of therapy. Several antiviral agents are available for topical use in HSV eye infections: idoxuridine, trifluorothymidine, topical vidarabine, and cidofovir and, more recently, topical pritelivir. For HSV encephalitis and neonatal herpes, IV acyclovir is the treatment of choice.

Most licensed antiviral agents for use against HSV inhibit the viral DNA polymerase. One class of drugs, typified by the drug acyclovir, is made up of substrates for the HSV enzyme TK. Acyclovir, ganciclovir, famciclovir, and valacyclovir are all selectively phosphorylated to the monophosphate form in virus-infected cells. Cellular enzymes convert the monophosphate form of the drug to the triphosphate, which is then incorporated into the viral DNA chain. Acyclovir is the agent most frequently used for the treatment of HSV infections and is available in IV, oral, and topical formulations. Valacyclovir, the valyl ester of acyclovir, offers greater bioavailability than acyclovir and thus can be administered less frequently. Famciclovir, the oral formulation of penciclovir, is clinically effective in the treatment of a variety of HSV-1 and HSV-2 infections. Ganciclovir is active against both HSV-1 and HSV-2; however, it is more toxic than acyclovir, valacyclovir, and famciclovir and generally is not recommended for the treatment of HSV infections. Anecdotal case reports suggest that ganciclovir may also be less effective than acyclovir for the treatment of HSV infections. All three recommended compounds—acyclovir, valacyclovir, and famciclovir—have proved effective in shortening the duration of symptoms and lesions of mucocutaneous HSV infections in both immunocompromised and immunocompetent patients **(Table 197-1)**. IV and oral formulations prevent reactivation of HSV in seropositive immunocompromised patients during induction chemotherapy or in the period immediately after bone marrow or solid organ transplantation. Chronic daily suppressive therapy reduces the frequency of reactivation disease among patients with frequent genital or oral–labial herpes. Only valacyclovir has been subjected to clinical trials that demonstrated reduced transmission of HSV-2 infection between sexual partners. IV acyclovir (30 mg/kg per day, given as a 10-mg/kg infusion over 1 h at 8-h intervals) is effective in reducing rates of death and morbidity from HSV encephalitis. Early initiation of therapy is a critical factor in outcome. The major side effect associated with IV acyclovir is transient renal insufficiency, usually due to crystallization of the compound in the renal parenchyma. This adverse reaction can be avoided if the medication is given slowly over 1 h and the patient is well hydrated. Because CSF levels of acyclovir average only 30–50% of plasma levels, the dosage of acyclovir used for treatment of CNS infection (30 mg/kg per day) is double that used for treatment of mucocutaneous or visceral disease (15 mg/kg per day). Even higher doses of IV acyclovir are used for neonatal HSV infection (60 mg/kg per day in three divided doses). Antiviral drugs neither eradicate latent infection nor affect the risk, frequency, or severity of subclinical or clinical recurrence after the drug is discontinued.

Increasingly, shorter courses of therapy are being used for recurrent mucocutaneous infection with HSV-1 or HSV-2 in immunocompetent patients. One-day courses of famciclovir and valacyclovir are clinically effective, more convenient, and generally less costly than longer courses of therapy (Table 197-1). These short-course regimens should be reserved for immunocompetent hosts.

SUPPRESSION OF MUCOCUTANEOUS HERPES

Recognition of the high frequency of subclinical reactivation provides a well-accepted rationale for the use of daily antiviral therapy to suppress reactivations of HSV, especially in persons with frequent clinical reactivations (e.g., those with recently acquired genital HSV infection). Immunosuppressed persons, including those with HIV infection, may also benefit from daily antiviral therapy. Daily acyclovir and valacyclovir reduce the frequency of HSV reactivations among HIV-positive persons. Regimens used include acyclovir (400–800 mg twice daily), famciclovir (500 mg twice daily), and valacyclovir (500 mg twice daily); valacyclovir at a dose of 4 g/d was associated with thrombotic thrombocytopenic purpura in one study of HIV-infected persons. Daily acyclovir therapy is associated with a modest reduction in the titer of HIV RNA in plasma (0.5-$\log_{10}$ reduction) and in the genital mucosa (0.33-$\log_{10}$ reduction). Episodes of subclinical reactivation still occur with chronic acyclovir or valaciclovir therapy.

REDUCED HSV TRANSMISSION TO SEXUAL PARTNERS

Once-daily valacyclovir (500 mg) has been shown to reduce transmission of HSV-2 between sexual partners. Transmission rates are higher from males to females and among persons with frequent HSV-2 reactivation. Serologic screening can be used to identify at-risk couples. Daily valacyclovir appears to be more effective at reducing subclinical shedding than daily famciclovir.

ACYCLOVIR RESISTANCE

Clinically relevant acyclovir-resistant strains of HSV do occur. Most of these strains have an altered substrate specificity for phosphorylating acyclovir. Thus, cross-resistance to famciclovir and valacyclovir is usually found. Occasionally, an isolate with altered TK specificity arises and is sensitive to famciclovir but not to acyclovir. In some patients infected with TK-deficient virus, higher doses of acyclovir are associated with clearing of lesions. In others, clinical disease progresses despite high-dose therapy. The majority of clinically significant acyclovir resistance has been seen in immunocompromised patients. HSV-2 isolates appear to be more resistant than HSV-1 strains. The frequency of acyclovir resistance is not systematically monitored; the lack of appreciable change in the frequency of acyclovir resistance in the general population past 45 years probably reflects the reduced transmission of TK-deficient mutants. Isolation of HSV from lesions persisting despite adequate dosages and blood levels of acyclovir should raise the suspicion of acyclovir resistance. Clinical management of acyclovir resistance can be challenging. Therapy with the antiviral drug foscarnet (40–80 mg/kg IV every 8 h until clinical resolution) is the only currently well demonstrated approach **(Chap. 196)**. Because of its toxicity and cost, this drug is usually reserved for patients with extensive mucocutaneous infections. Cidofovir is a nucleotide analogue and exists as a phosphonate or monophosphate form. Most TK-deficient strains of HSV are sensitive to cidofovir. Cidofovir ointment speeds healing of acyclovir-resistant lesions, but the topical drug itself can cause mucocutaneous ulcerations. No well-controlled trials of systemic cidofovir have been reported. Occasional cases may respond to topical imiquimod. Anecdotal case reports of successful

TABLE 197-1 Antiviral Chemotherapy for Herpes Simplex Virus (HSV) Infection

I. Mucocutaneous HSV infections

A. *Infections in immunosuppressed patients*

1. *Acute symptomatic first or recurrent episodes:* IV acyclovir (5 mg/kg q8h) or oral acyclovir (400 mg qid), famciclovir (500 mg bid or tid), or valacyclovir (500 mg bid) is effective. Treatment duration may vary from 7 to 21 days. IV therapy may be given for 2–10 days until clinical improvement and followed by oral therapy.

2. *Suppression of reactivation disease (genital or oral–labial):* IV acyclovir (5 mg/kg q8h) or oral valacyclovir (500 mg bid) or acyclovir (400–800 mg 3–5 times per day) prevents recurrences during the 30-day period immediately after transplantation. Longer-term HSV suppression is often used for persons with continued immunosuppression. In bone marrow and renal transplant recipients, oral valacyclovir (2 g/d) is also effective in reducing cytomegalovirus infection. Oral valacyclovir at a dose of 4 g/d has been associated with thrombotic thrombocytopenic purpura after extended use in HIV-positive persons. In HIV-infected persons, oral acyclovir (400–800 mg bid), valacyclovir (500 mg bid), or famciclovir (500 mg bid) is effective in reducing clinical and subclinical reactivations of HSV-1 and HSV-2.

B. *Infections in immunocompetent patients*

1. *Genital herpes*

a. *First episodes:* Oral acyclovir (200 mg 5 times per day or 400 mg tid), valacyclovir (1 g bid), or famciclovir (250 mg tid) for 7–14 days is effective. IV acyclovir (5 mg/kg q8h for 5 days) is given for severe disease or neurologic complications such as aseptic meningitis.

b. *Symptomatic recurrent genital herpes:* Short-course (1- to 3-day) regimens are preferred because of low cost, likelihood of adherence, and convenience. Oral acyclovir (800 mg tid for 2 days), valacyclovir (500 mg bid for 3 days), valacyclovir (1 g orally once a day for 3 days), or famciclovir (750 or 1000 mg bid for 1 day, a 1500-mg single dose, or 500 mg stat followed by 250 mg q12h for 2 days) effectively shortens lesion duration. Other options include oral acyclovir (200 mg 5 times per day), valacyclovir (500 mg bid), and famciclovir (125 mg bid for 5 days).

c. *Suppression of recurrent genital herpes:* Oral acyclovir (400–800 mg bid) or valacyclovir (500 mg daily) is given. Patients with >9 episodes per year should take oral valacyclovir (1 g daily or 500 mg bid) or famciclovir (250 mg bid or 500 mg bid).

2. *Oral–labial HSV infections*

a. *First episode:* Oral acyclovir is given (200 mg 5 times per day or 400 mg tid); an oral acyclovir suspension can be used (600 mg/m^2 qid). Oral famciclovir (250 mg bid) or valacyclovir (1 g bid) has been used clinically. The duration of therapy is 5–10 days.

b. *Recurrent episodes:* If initiated at the onset of the prodrome, single-dose or 1-day therapy effectively reduces pain and speeds healing. Regimens include oral famciclovir (a 1500-mg single dose or 750 mg bid for 1 day) or valacyclovir (a 2-g single dose or 2 g bid for 1 day). Self-initiated therapy with 6-times-daily topical penciclovir cream effectively speeds healing of oral–labial HSV infection. Topical acyclovir cream has also been shown to speed healing or single dose oral amenamevir 1200 mg.

c. *Suppression of reactivation of oral–labial HSV:* If started before exposure and continued for the duration of exposure (usually 5–10 days), oral acyclovir (400 mg bid) prevents reactivation of recurrent oral–labial HSV infection associated with severe sun exposure.

3. *Surgical prophylaxis of oral or genital HSV infection:* Several surgical procedures, such as laser skin resurfacing, trigeminal nerve-root decompression, and lumbar disk surgery, have been associated with HSV reactivation. IV acyclovir (3–5 mg/kg q8h) or oral acyclovir (800 mg bid), valacyclovir (500 mg bid), or famciclovir (250 mg bid) effectively reduces reactivation. Therapy should be initiated 48 h before surgery and continued for 3–7 days.

4. *Herpetic whitlow:* Oral acyclovir (200 mg) is given 5 times daily (alternative: 400 mg tid) for 7–10 days.

5. *HSV proctitis:* Oral acyclovir (400 mg 5 times per day) is useful in shortening the course of infection. In immunosuppressed patients or in patients with severe infection, IV acyclovir (5 mg/kg q8h) may be useful.

6. *Herpetic eye infections:* In acute keratitis, topical trifluorothymidine, vidarabine, idoxuridine, acyclovir, penciclovir, and interferon are all beneficial. Debridement may be required. Topical steroids may worsen disease.

II. Central nervous system HSV infections

A. *HSV encephalitis:* IV acyclovir (10 mg/kg q8h; 30 mg/kg per day) is given for 10 days or until HSV DNA is no longer detected in cerebrospinal fluid.

B. *HSV aseptic meningitis:* No studies of systemic antiviral chemotherapy exist. If therapy is to be given, IV acyclovir (15–30 mg/kg per day) should be used.

C. *Autonomic radiculopathy:* No studies are available. Most authorities recommend initial therapy with IV acyclovir (5–10 mg/kg q8h) followed by oral therapy for 21 days.

III. Neonatal HSV infections: IV acyclovir (60 mg/kg per day, divided into 3 doses) is given. The recommended duration of IV treatment is 21 days. Monitoring for relapse should be undertaken. Continued suppression with oral acyclovir suspension should be given for 3–4 months.

IV. Visceral HSV infections

A. *HSV esophagitis:* IV acyclovir (15 mg/kg per day) is given. In some patients with milder forms of immunosuppression, oral therapy with valacyclovir or famciclovir is effective.

B. *HSV pneumonitis and lymphadenitis:* No controlled studies exist. IV acyclovir (15 mg/kg per day) should be considered.

V. Disseminated HSV infections: No controlled studies exist. IV acyclovir (5 mg/kg q8h) should be tried. Adjustments for renal insufficiency may be needed. No definite evidence indicates that therapy will decrease the risk of death.

VI. Erythema multiforme associated with HSV: Anecdotal observations suggest that oral acyclovir (400 mg bid or tid) or valacyclovir (500 mg bid) will suppress erythema multiforme.

VII. Infections due to acyclovir-resistant HSV: IV foscarnet (40 mg/kg IV q8h) should be given until lesions heal. The optimal duration of therapy and the usefulness of its continuation to suppress lesions are unclear. The helicase-primase inhibitor pritelivir, 400 mg oral loading dose on day 1 followed by 100 mg once daily for 21–28 days, has reported effectiveness. Some patients may benefit from cutaneous application of trifluorothymidine or 1% cidofovir gel, both of which must be compounded at a pharmacy. These preparations should be applied once daily for 5–7 days. Topical imiquimod can be considered. IV cidofovir (5 mg/kg weekly) may be considered.

VIII. Acyclovir and pregnancy: No adverse effects to the fetus or newborn have been attributable to acyclovir. Acyclovir can be used in all stages of pregnancy and among women who are breastfeeding (the drug can be found in breast milk). Suppressive acyclovir treatment in late pregnancy (acyclovir 400 mg orally tid or valacyclovir 500 mg orally bid from ~34 weeks until delivery) reduces the frequency of cesarean delivery among women with recurrent genital herpes. Such treatment may not protect against transmission to neonates.

treatment of acyclovir resistant HSV with oral pritelivir have been reported in immunocompetent persons. Drugs from a new class that inhibit HSV-specific helicase/primase activity, amenamevir and pritelivir, are under clinical investigation and appear to offer a better toxicity profile for the treatment of acyclovir-resistant strains of HSV. Amenamevir is licensed for varicella-zoster infections in Japan, and a recent study has shown promise in recurrent oral–labial HSV.

ACYCLOVIR EFFICACY IN THE DEVELOPING WORLD

Initial studies of acyclovir-like drugs were performed solely in the developed world. While acyclovir, valacyclovir, and famciclovir are effective in the developing world, their clinical and virologic benefits, especially in reducing the frequency of genital lesions among patients in Africa, seem reduced from those in European and U.S. populations. The mechanism of this phenomenon is uncertain. Acyclovir therapy does not reduce the rate of HIV acquisition; however, HIV load among MSM in the United States decreased by 1.3 log$_{10}$ in contrast to 0.9 log$_{10}$ among Peruvian MSM and 0.5 log$_{10}$ among African women. Curiously, the anti-HIV drug tenofovir reduces HSV-2 acquisition among women in Africa, although it has no demonstrable clinical benefit or antiviral effects among persons with established HSV-2 infection in studies in the United States. The reasons for these disparate results are unclear.

PREVENTION

Efforts to control HSV disease on a population basis through suppressive antiviral therapy and/or educational programs have been limited. Barrier forms of contraception (especially condoms) decrease the likelihood of transmission of HSV infection, particularly during periods of asymptomatic viral excretion. When lesions are present, HSV infection may be transmitted by skin-to-skin contact despite the use of a condom. Nevertheless, the available data suggest that consistent condom use is an effective means of reducing the risk of genital HSV-2 transmission. Chronic daily antiviral therapy with valacyclovir can also be partially effective in reducing acquisition of HSV-2, especially among susceptible women. There are no comparative efficacy studies of valacyclovir versus condom use. Most authorities suggest both approaches. The need for a vaccine to prevent acquisition of HSV infection is great, especially in light of the role HSV-2 plays in enhancing the acquisition and transmission of HIV-1.

A substantial portion of neonatal HSV cases could be prevented by reducing the acquisition of HSV by women in the third trimester of pregnancy. Neonatal HSV infection can result from either the acquisition of maternal infection near term or the reactivation of infection at delivery in the already-infected mother. Women without known genital herpes should be counseled to abstain from vaginal intercourse during the third trimester with partners known to have or suspected of having genital herpes. Some authorities have recommended that antiviral therapy with acyclovir or valacyclovir be given to HSV-2-infected women in late pregnancy as a means of reducing reactivation of HSV-2 at term. Data are not available to support the efficacy of this approach, and the high treatment-to-prevention ratio makes this a difficult if not dubious public health strategy, even though it can reduce the frequency of HSV-associated cesarean delivery.

FURTHER READING

Centers for Disease Control and Prevention: 2021 Sexually transmitted diseases treatment guidelines. Available at *https://www.cdc.gov/std/treatment-guidelines/herpes.htm*.

James C et al: Herpes simplex virus: Global infection prevalence and incidence estimates, 2016. Bull World Health Organ 98:315, 2020.

Kawashima M et al: A phase 3, randomized, double-blind, placebo-controlled study evaluating a single, patient-initiated dose of amenamevir for recurrent herpes labialis. J Dermatol 50:311, 2022.

Looker KJ et al: Global and regional estimates of the contribution of herpes simplex virus type 2 infection to HIV incidence: A population attributable fraction analysis using published epidemiological data. Lancet Infect Dis 20:240, 2020.

Mahant S et al: Neonatal herpes simplex virus infection among Medicaid-enrolled children: 2009–2015. Pediatrics 143:e20183233, 2019.

Serris A et al: Pritelivir for recurrent acyclovir-resistant herpes simplex virus 2 infections in immunocompromised patients. J Antimicrob Chemother 77:2303, 2022.

198

Varicella-Zoster Virus Infections

Richard J. Whitley

DEFINITION

Varicella-zoster virus (VZV) causes two distinct clinical syndromes: varicella (chickenpox) and herpes zoster (shingles). Chickenpox, a ubiquitous and extremely contagious infection, is usually a benign illness of childhood characterized by an exanthematous vesicular rash. With reactivation of latent VZV (which is most common after the sixth decade of life), herpes zoster presents as a dermatomal vesicular rash and is usually associated with severe pain.

HISTORY AND ETIOLOGY

Early in the twentieth century, similarities in the histopathologic features of skin lesions resulting from varicella and herpes zoster were described. Viral isolates from patients with each of these diseases produced similar pathology in tissue culture—specifically, the appearance of eosinophilic intranuclear inclusions and multinucleated giant cells. These results suggested that the viruses were biologically similar. Restriction endonuclease analyses of viral DNA from a patient with chickenpox who subsequently developed herpes zoster verified the molecular identity of the two viruses responsible for these different clinical presentations.

VZV is a member of the family Herpesviridae, sharing with other members such structural characteristics as a lipid envelope surrounding a nucleocapsid with icosahedral symmetry, a total diameter of ~180–200 nm, and centrally located double-stranded DNA that is ~125,000 bp in length.

PATHOGENESIS AND PATHOLOGY

Primary Infection Transmission occurs readily by the respiratory route; the subsequent localized replication of the virus at an undefined site (presumably the nasopharynx) leads to seeding of the lymphatic/reticuloendothelial system and ultimately to the development of viremia. Viremia in patients with chickenpox manifests as a diffuse and scattered skin lesions that can be confirmed by the recovery of VZV from the blood (rarely) or routinely by the detection of viral DNA in either blood or lesions by polymerase chain reaction (PCR). Vesicles involve the corium or dermis, with degenerative changes characterized by ballooning, the presence of multinucleated giant cells, and eosinophilic intranuclear inclusions. Infection may involve localized blood vessels of the skin, resulting in necrosis and epidermal hemorrhage. With the evolution of disease, the vesicular fluid becomes cloudy because of the recruitment of polymorphonuclear leukocytes and the presence of degenerated cells and fibrin. Ultimately, the vesicles either rupture and release their fluid (which includes infectious virus) or are gradually reabsorbed.

Recurrent Infection The mechanism of reactivation of VZV that results in herpes zoster is unknown. The virus infects dorsal root ganglia during chickenpox, where it remains latent until reactivated. Histopathologic examination of representative dorsal root ganglia during active herpes zoster demonstrates hemorrhage, edema, and lymphocytic infiltration. Latent virus has been detected in sensory (dorsal, cranial, and enteric) ganglia.

Active replication of VZV in other organs, such as the lung or the brain, can occur during either chickenpox or herpes zoster but is uncommon in the immunocompetent host. Pulmonary involvement is characterized by interstitial pneumonitis, multinucleated giant cell formation, intranuclear inclusions, and pulmonary hemorrhage. Central nervous system (CNS) infection leads to histopathologic evidence of perivascular cuffing similar to that encountered in measles and other viral encephalitides. Focal hemorrhagic necrosis of the brain, characteristic of herpes simplex virus (HSV) encephalitis, develops infrequently in VZV infection.

EPIDEMIOLOGY AND CLINICAL MANIFESTATIONS

Chickenpox Humans are the only known reservoir for VZV. Chickenpox is highly contagious, with an attack rate of at least 90% among susceptible (seronegative) individuals. Persons of both sexes and all races are infected equally. The virus is endemic in the population at large; however, it becomes epidemic among susceptible individuals during seasonal peaks—namely, late winter and early spring in the temperate zone. For the most part, our knowledge of the disease's natural history and incidence predates the licensure of the chickenpox vaccine in 1995. Historically, children 5–9 years old were most commonly affected, accounting for 50% of all cases. Most other cases involved children 1–4 and 10–14 years old. Approximately 10% of the population of the United States over the age of 15 was susceptible to infection. VZV vaccination during the second year of life and followed by a preschool booster has dramatically changed the epidemiology of infection, causing a significant decrease in the annualized incidence of chickenpox, as noted below.

The incubation period of chickenpox ranges from 10 to 21 days but is usually 14–17 days. Secondary attack rates in susceptible siblings within a household are 70–90%. Patients are infectious ~48 h before the onset of the vesicular rash, during the period of vesicle formation (which generally lasts 4–5 days), and until all vesicles are crusted.

Clinically, chickenpox presents as a rash, low-grade fever, and malaise, although a few patients develop a prodrome 1–2 days before onset of the exanthem. In the immunocompetent patient, chickenpox is usually a benign illness associated with lassitude and with body temperatures of 37.8°–39.4°C (100°–103°F) of 3–5 days' duration. The skin lesions—the hallmark of the infection—include maculopapules, vesicles, and scabs in various stages of evolution **(Fig. 198-1; see also Fig. A1-30)**. These lesions, which evolve from maculopapules to vesicles over hours to days, appear on the trunk and face and rapidly spread to involve other areas of the body. Most are small and have an erythematous base with a diameter of 5–10 mm. Successive crops appear over a 2- to 4-day period. Lesions can also be found on the mucosa of the pharynx and/or the vagina. Their severity varies from one person to another. Some individuals have very few lesions, while others have as many as 2000. Younger children tend to have fewer vesicles than older individuals. Within families, secondary and tertiary cases are associated with a larger number of vesicles than the first family case. Immunocompromised patients—both children and adults, particularly those with leukemia—have lesions (often with a hemorrhagic base) that are more numerous and take longer to heal than those of immunocompetent individuals. Immunocompromised individuals are also at greater risk for visceral complications, which occur in 30–50% of cases and are fatal 15% of the time in the absence of antiviral therapy.

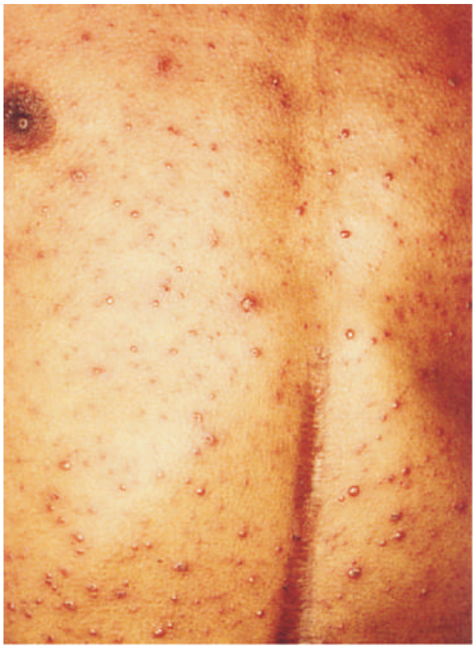

FIGURE 198-1 Varicella lesions at various stages of evolution: vesicles on an erythematous base, umbilical vesicles, and crusts.

The most common infectious complication of varicella is secondary bacterial superinfection of the skin, which is usually caused by *Streptococcus pyogenes* or *Staphylococcus aureus,* including strains that are methicillin-resistant. Skin infection results from excoriation of lesions after scratching. Gram stain of skin lesions may help clarify the etiology of unusually erythematous and pustulated lesions.

The most common extracutaneous site of involvement in children is the CNS. The syndrome of acute cerebellar ataxia and meningeal inflammation generally appears ~21 days after onset of the rash and rarely develops in the pre-eruptive phase. The cerebrospinal fluid (CSF) contains lymphocytes and elevated levels of protein. CNS involvement is a benign complication of VZV infection in immunocompetent children and generally does not require hospitalization. Aseptic meningitis, encephalitis, transverse myelitis, and Guillain-Barré syndrome can also occur. Encephalitis is reported in 0.1–0.2% of children with chickenpox. Reye syndrome, an historical complication for the most part, can occur in children concomitantly treated with aspirin; therefore, aspirin is no longer utilized. Other than supportive care, no specific therapy (e.g., although acyclovir is administered by some physicians) has proved efficacious for patients with CNS involvement.

Varicella pneumonia, the most serious complication following chickenpox, develops more often in adults (up to 20% of cases) than in children and is particularly severe in pregnant women. Pneumonia due to VZV usually has its onset 3–5 days into the illness and is associated with tachypnea, cough, dyspnea, and fever. Cyanosis, hypoxemia, pleuritic chest pain, and hemoptysis are frequently noted. Roentgenographic evidence of disease consists of nodular infiltrates and interstitial pneumonitis. Resolution of pneumonitis parallels improvement of the skin rash; however, patients may have persistent fever and compromised pulmonary function for weeks.

Other complications of chickenpox include myocarditis, corneal lesions, nephritis, arthritis, bleeding diatheses, acute glomerulonephritis, and hepatitis. Hepatic involvement, distinct from Reye syndrome and usually asymptomatic, is common in chickenpox and is generally characterized by elevated levels of liver enzymes, particularly aspartate and alanine aminotransferases.

Perinatal varicella is associated with mortality rates as high as 30% when maternal disease develops within 5 days before delivery or within 48 h thereafter. Illness in this setting is unusually severe because the newborn does not receive protective transplacental antibodies and has an immature immune system. *Congenital varicella*, with clinical manifestations of limb hypoplasia, cicatricial skin lesions, and microcephaly at birth, is extremely uncommon.

Herpes Zoster Herpes zoster (shingles) is a sporadic disease that results from reactivation of latent VZV from dorsal root ganglia. Most patients with shingles have no history of recent exposure to other individuals with VZV infection. Herpes zoster occurs at all ages, but its incidence is highest (5–10 cases per 1000 persons) among individuals in the sixth decade of life and beyond. Data suggest that at least 1.2 million cases occur annually in the United States. Vaccination has, likely, decreased the incidence of shingles. Recurrent herpes zoster is exceedingly rare except in immunocompromised hosts, especially those with AIDS.

Herpes zoster is characterized by a unilateral vesicular dermatomal eruption, often associated with severe pain. The dermatomes from T3 to L3 are most frequently involved. If the ophthalmic branch of the trigeminal nerve is involved, *zoster ophthalmicus* results. The factors responsible for the reactivation of VZV are not known. In children, reactivation is usually benign; in adults, it can be debilitating because of pain. The onset of disease is heralded by pain within the dermatome, which may precede lesions by 48–72 h; an erythematous maculopapular rash evolves rapidly into vesicular lesions **(Fig. 198-2)**. In the normal host, these lesions may remain few in number and continue to

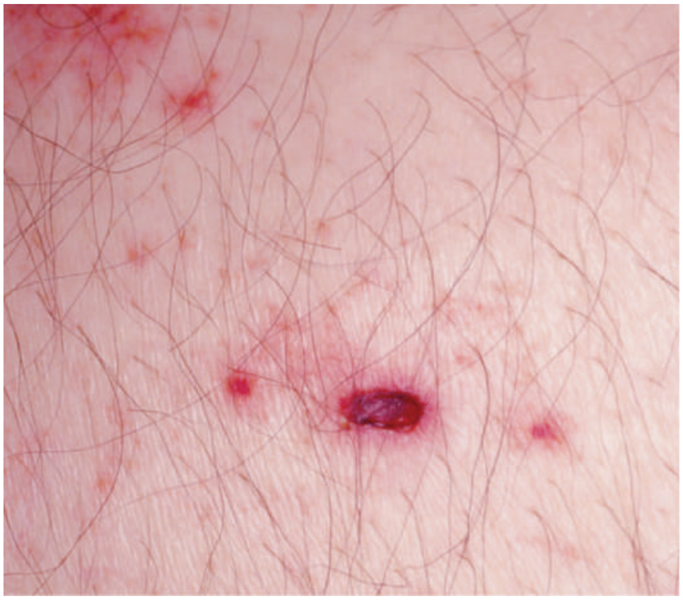

FIGURE 198-2 Close-up of lesions of disseminated zoster. Note lesions at different stages of evolution, including pustules and crusting. *(Photo courtesy of Lindsey Baden; with permission.)*

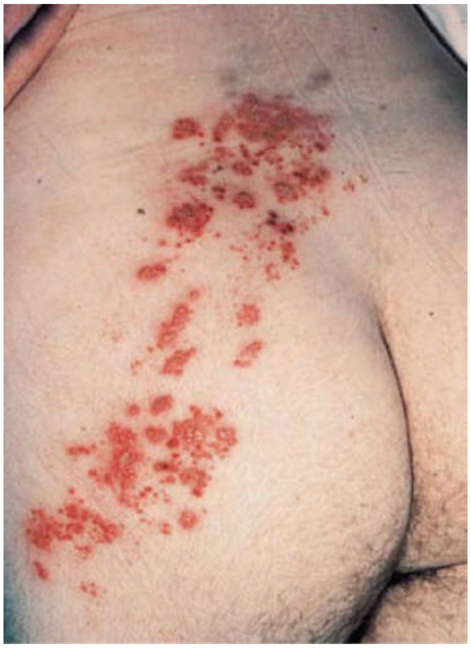

FIGURE 198-3 Herpes zoster in an HIV-infected patient is seen as hemorrhagic vesicles and pustules on an erythematous base grouped in a dermatomal distribution.

form for only 3–5 days. The total duration of disease is generally 7–10 days; however, it may take as long as 2–4 weeks for the skin to return to normal. Patients with herpes zoster can transmit infection to seronegative individuals, resulting in chickenpox. In a few patients, characteristic localization of pain to a dermatome with serologic evidence of herpes zoster has been reported in the absence of skin lesions, an entity known as *zoster sine herpetica*. When branches of the trigeminal nerve are involved, lesions may appear on the face, particularly the tip of the nose (Hutchinson sign), in the mouth, in the eye, or on the tongue. *Zoster ophthalmicus* is usually a debilitating condition that can result in blindness in the absence of antiviral therapy. In *Ramsay Hunt syndrome*, pain and vesicles appear in the external auditory canal, and patients lose their sense of taste in the anterior two-thirds of the tongue while developing ipsilateral facial palsy. The geniculate ganglion of the sensory branch of the facial nerve is involved.

In both normal and immunocompromised hosts, the most debilitating complication of herpes zoster is pain associated with acute neuritis and postherpetic neuralgia. Postherpetic neuralgia is uncommon in young individuals; however, at least 50% of patients over age 50 report some degree of pain in the involved dermatome for months after the resolution of cutaneous disease. Changes in sensation in the dermatome, resulting in either hypo- or hyperesthesia, are common.

CNS involvement may follow localized herpes zoster. Many patients without signs of meningeal irritation have CSF pleocytosis and moderately elevated levels of CSF protein. Symptomatic meningoencephalitis is characterized by headache, fever, photophobia, meningitis, and vomiting. A rare manifestation of CNS involvement is granulomatous angiitis with contralateral hemiplegia, which can be diagnosed by cerebral arteriography. Other neurologic manifestations include transverse myelitis with or without motor paralysis.

Like chickenpox, herpes zoster is more severe in immunocompromised than immunocompetent individuals. Lesions continue to form for >1 week, and scabbing is not complete in most cases until 3 weeks into the illness. Patients with Hodgkin disease and non-Hodgkin lymphoma are at greatest risk for progressive herpes zoster. Cutaneous dissemination **(Fig. 198-3)** develops in ~40% of these patients in the absence of therapy. Among patients with cutaneous dissemination, the risk of pneumonitis, meningoencephalitis, hepatitis, and other serious complications is increased by 5–10%. However, even in immunocompromised patients, disseminated zoster is rarely fatal.

Recipients of hematopoietic stem cell transplants are at particularly high risk of VZV infection. Of all cases of posttransplantation VZV infection, 30% occur within 1 year (50% of these within 9 months); 45% of the patients involved have cutaneous or visceral dissemination. The mortality rate in this situation is 10%. Postherpetic neuralgia, scarring, and bacterial superinfection are especially common in VZV infections occurring within 9 months of transplantation. Among infected patients, concomitant graft-versus-host disease increases the chance of dissemination and/or death.

DIFFERENTIAL DIAGNOSIS

The diagnosis of chickenpox is usually based on clinical presentation and is not difficult. The characteristic rash and a history of recent exposure should lead to a prompt diagnosis. Other viral infections that can mimic chickenpox include disseminated HSV infection in patients with atopic dermatitis and the disseminated vesiculopapular lesions sometimes associated with coxsackievirus infection, echovirus infection, or atypical measles. However, these rashes are more commonly morbilliform with a hemorrhagic component rather than vesicular or vesiculopustular. Rickettsialpox **(Chap. 192)** is sometimes confused with chickenpox; however, rickettsialpox can be distinguished easily by detection of the "herald spot" at the site of the mite bite and the development of a more pronounced headache. Serologic testing is also useful in differentiating rickettsialpox from varicella and can confirm susceptibility in adults unsure of their chickenpox history. Mpox should be considered in travelers returning from endemic areas **(Chap. 201)** and has been a recent concern in HIV-infected individuals, leading to vaccination campaigns. Smallpox concern has increased because of the threat of bioterrorism **(Chap. S4)**. The lesions of smallpox are larger than those of chickenpox and are all at the same stage of evolution at any given time.

Unilateral vesicular lesions in a dermatomal pattern should lead rapidly to the diagnosis of herpes zoster, although the occurrence of shingles without a rash has been reported. Both HSV and coxsackievirus infections can cause dermatomal vesicular lesions. Supportive diagnostic virology and fluorescent staining of skin scrapings with monoclonal antibodies are helpful in ensuring the proper diagnosis. In the prodromal stage of herpes zoster, the diagnosis can be exceedingly difficult and may be made only after lesions have appeared or by retrospective serologic assessment.

LABORATORY FINDINGS

Unequivocal confirmation of the diagnosis is possible only through the isolation of VZV in susceptible tissue-culture cell lines, the demonstration of either seroconversion or a fourfold or greater rise in antibody titer between acute-phase and convalescent-phase serum specimens, or the detection of VZV DNA by PCR. Specimens for detection of VZV DNA by PCR include lesions, blood, and saliva. A rapid impression can be obtained by a Tzanck smear, with scraping of the base of the lesions in an attempt to demonstrate multinucleated giant cells; however, the sensitivity of this method is low (~60%). PCR detection of viral DNA in vesicular fluid is available in many diagnostic laboratories and is the diagnostic method of choice. Direct immunofluorescent staining of cells from the lesion base or detection of viral antigens by other assays (such as the immunoperoxidase assay) is also useful. The most frequently employed serologic tools for assessing host response are the immunofluorescent detection of antibodies to VZV membrane antigens, the fluorescent antibody to membrane antigen (FAMA) test, immune adherence hemagglutination, and enzyme-linked immunosorbent assay (ELISA). The FAMA test and the ELISA appear to be most sensitive.

TREATMENT

Varicella-Zoster Virus Infections

Medical management of chickenpox in the immunologically normal host is directed toward the prevention of avoidable complications. Obviously, good hygiene includes daily bathing and soaks. Secondary bacterial infection of the skin can be avoided by meticulous skin care, particularly with close cropping of fingernails. Pruritus can be decreased with topical dressings or the administration of antipruritic drugs. Tepid water baths and wet compresses are better than drying lotions for the relief of itching. Administration of aspirin to children with chickenpox should be avoided because of its association with the development of Reye syndrome. Acyclovir (800 mg by mouth five times daily), valacyclovir (1 g three times daily), or famciclovir (250 mg three times daily) for 5–7 days is recommended for adolescents and adults with chickenpox of ≤24 h in duration. (Valacyclovir is licensed for use in children and adolescents. Famciclovir is recommended but not licensed for varicella.) Likewise, acyclovir therapy may be of benefit to children <12 years of age if initiated early in the disease (<24 h) at a dose of 20 mg/kg every 6 h. The advantages (i.e., pharmacokinetics) of the second-generation agents valacyclovir and famciclovir are described in **Chap. 196**.

Aluminum acetate soaks for the management of herpes zoster can be both soothing and cleansing. Patients with herpes zoster benefit from oral antiviral therapy, as evidenced by accelerated healing of lesions and resolution of zoster-associated pain with acyclovir, valacyclovir, or famciclovir. Acyclovir is administered at a dosage of 800 mg five times daily for 7–10 days. However, valacyclovir and famciclovir have superior pharmacokinetics and pharmacodynamics and should be used preferentially. Famciclovir, the prodrug of penciclovir, is at least as effective as acyclovir and perhaps more so; the dose is 500 mg by mouth three times daily for 7 days. Valacyclovir, the prodrug of acyclovir, accelerates healing and resolution of zoster-associated pain more promptly than acyclovir. The dose is 1 g by mouth three times daily for 5–7 days. Compared with acyclovir, both famciclovir and valacyclovir offer the advantage of less frequent administration. All three of these drugs are now available as generic products.

In severely immunocompromised hosts (e.g., transplant recipients, patients with cancer, particularly lymphoproliferative malignancies), both chickenpox and herpes zoster (including disseminated disease) should be treated, at least at the outset, with IV acyclovir, which reduces the occurrence of visceral complications but has little effect on healing of skin lesions or pain. The dose is 10 mg/kg every 8 h for 7 days. For low-risk immunocompromised hosts, oral therapy with valacyclovir or famciclovir appears beneficial. If medically feasible, it is desirable to decrease immunosuppressive treatment concomitant with the administration of IV acyclovir.

Patients with varicella pneumonia typically require ventilatory support. Persons with zoster ophthalmicus should be referred immediately to an ophthalmologist. Therapy for this condition consists of the administration of analgesics for severe pain and the use of atropine. Acyclovir, valacyclovir, and famciclovir all accelerate healing. Decisions regarding the use of corticosteroids should be made by the ophthalmologist.

The management of acute neuritis and/or postherpetic neuralgia can be particularly difficult. In addition to the judicious use of analgesics ranging from nonnarcotics to narcotic derivatives, drugs such as gabapentin, pregabalin, amitriptyline hydrochloride, lidocaine (patches), and fluphenazine hydrochloride are reportedly beneficial for pain relief. In one study, glucocorticoid therapy administered early in the course of localized herpes zoster significantly accelerated such quality-of-life improvements as a return to usual activity and termination of analgesic medications. The dose of prednisone administered orally was 60 mg/d on days 1–7, 30 mg/d on days 8–14, and 15 mg/d on days 15–21. This regimen is appropriate only for relatively healthy elderly persons with moderate or severe pain at presentation. Patients with osteoporosis, diabetes mellitus, glycosuria, or hypertension may not be appropriate candidates. Glucocorticoids should not be used without concomitant antiviral therapy.

■ PREVENTION

Three methods are used for the prevention of VZV infections. First, a live attenuated varicella vaccine (Oka) is recommended for all children >1 year of age (up to 12 years of age) who have not had chickenpox and for adults known to be seronegative for VZV. Two doses are recommended for all children: the first at 12–15 months of age and the second at ~4–6 years of age. VZV-seronegative persons >13 years of age should receive two doses of vaccine at least 1 month apart. The vaccine is both safe and efficacious. Breakthrough cases are mild and may result in spread of the vaccine virus to susceptible contacts. The universal vaccination of children has resulted in a decreased incidence of chickenpox in sentinel communities. Furthermore, the inactivated Oka vaccine, when administered to hematopoietic stem cell recipients, decreases the incidence of herpes zoster.

In individuals >50 years of age, the shingles vaccine of choice is Shingrix. It is a subunit vaccine (HZ/su) that consists of VZV glycoprotein E and the AS01$_B$ adjuvant. A randomized, placebo-controlled study administered two doses of vaccine or placebo 1 month apart to 15,411 participants aged 50 years or older. Overall vaccine efficacy for the prevention of herpes zoster virus was 97.2% (95% confidence interval, 93.7–99.0%; *p* <.001). Injection-site and systemic reactions were more frequent in vaccine recipients, but the proportions of participants who had serious adverse events were similar in the vaccine and placebo groups. The Advisory Committee on Immunization Practices recommends that persons in this age group be offered this vaccine in order to reduce the frequency of shingles and the severity of postherpetic neuralgia. Of note, vaccine immunity wanes over time, and reassessment of current recommendations or the use of a promising inactivated vaccine in development will be required.

A second approach is to administer varicella-zoster immune globulin (VZIG) to individuals who are susceptible, are at high risk for developing complications of varicella, and have had a significant exposure. This product should be given within 96 h (preferably within 72 h) of the exposure but may be administered up to 10 days with some efficacy. Indications for administration of VZIG appear in **Table 198-1**, which has been adapted from the American Academy of Pediatrics Red Book.

Lastly, antiviral therapy can be given as prophylaxis to individuals at high risk who are ineligible for vaccination or who are beyond the 96-h window after direct contact. While the initial studies have used acyclovir, similar benefit can be anticipated with either valacyclovir or famciclovir. Therapy is instituted 7 days after intense exposure. At this time, the host is midway into the incubation period. This approach significantly decreases disease severity, if not totally preventing disease.

TABLE 198-1 Recommendations for VZIG Administration

Exposure Criteria
1. Significant exposure to a person with chickenpox or zoster
a. Household: residence in the same household
b. Playmate: face-to-face indoor play
c. Hospital
Varicella: same 2- to 4-bed room or adjacent beds in a large ward, face-to-face contact with an infectious staff member or patient, visit by a person deemed contagious
Zoster: intimate contact (e.g., touching or hugging) with a person deemed contagious
d. Newborn infant: onset of varicella in the mother ≤5 days before delivery or ≤48 h after delivery; VZIG not indicated if the mother has zoster
2. Patient should receive VZIG as soon as possible but not >96 h after exposure.
Candidates (Provided They Have Significant Exposure) Include
1. Immunocompromised susceptible children without a history of varicella or varicella immunization
2. Susceptible pregnant women
3. Newborn infants whose mother had onset of chickenpox within 5 days before or within 48 h after delivery
4. Hospitalized premature infant (≥28 weeks of gestation) whose mother lacks a reliable history of chickenpox or serologic evidence of protection against varicella
5. Hospitalized premature infant (<28 weeks of gestation or ≤1000-g birth weight), regardless of maternal history of varicella or VZV serologic status

Abbreviation: VZIG, varicella-zoster immune globulin.

Source: Table is adapted from the American Academy of Pediatrics *Red Book.*

FURTHER READING

Arvin A: Aging, immunity, and the varicella-zoster virus. N Engl J Med 352:2266, 2005.

Arvin A, Abendroth A: Varicella-zoster virus, in *Fields Virology: DNA Viruses*. Vol 2. 7th ed, Howley P et al (eds). Philadelphia, PA, Wolters Kluwer, 2021, pp 445-488.

Cohen JI: A new vaccine to prevent herpes zoster. N Engl J Med 372:2149, 2015.

Gershon AA et al: Varicella zoster virus infection. Nat Rev Dis Primers 1:15016, 2015.

Gnann JW, Whitley RJ: Herpes zoster. N Engl J Med 347:340, 2002.

Kimberlin DW et al (eds): *Redbook: 2021-2024 Report of the Committee on Infectious Diseases*, 32nd ed. Itasca, IL, American Academy of Pediatrics, 2021.

Lai H et al: Efficacy of an adjuvanted herpes zoster subunit vaccine in older adults. N Engl J Med 372:2087, 2015.

Marin M et al: Monitoring varicella vaccine impact on varicella incidence in the United States: Surveillance challenges and changing epidemiology, 1995-2019. J Infect Dis 226:S392, 2022.

Morrison VA et al: Long-term persistence of zoster vaccine efficacy. Clin Infect Dis 60:900, 2015.

Shaw J, Gershon AA: Varicella virus vaccination in the United States. Viral Immunol 31:96, 2018.

Whitley RJ, Arvin A: Chickenpox and herpes zoster (varicella-zoster virus), in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 10th ed. Blaser MJ, Cohen JI, Holland SM (eds). Philadelphia, PA, Elsevier Press. In press.

Wutzler P et al: Varicella vaccination: The global experience. Expert Rev Vaccines 16:833, 2017.

199

Epstein-Barr Virus Infections, Including Infectious Mononucleosis

Jeffrey I. Cohen

DEFINITION

Epstein-Barr virus (EBV) is the cause of heterophile-positive infectious mononucleosis (IM), which is characterized by fever, sore throat, lymphadenopathy, and atypical lymphocytosis. EBV is also associated with several tumors, including nasopharyngeal and gastric carcinoma, Burkitt's lymphoma, Hodgkin's lymphoma, T-cell lymphoma, and (in patients with immunodeficiencies) B-cell lymphoma and smooth muscle tumors. Large epidemiology studies show a strong association of EBV with multiple sclerosis. The virus is a member of the family *Herpesviridae*. The two types of EBV that are widely prevalent in nature are not distinguishable by conventional serologic tests.

EPIDEMIOLOGY

EBV infections occur worldwide. These infections are most common in early childhood, with a second peak during late adolescence. By adulthood, >90% of individuals have been infected and have antibodies to the virus. IM is usually a disease of young adults. In lower socioeconomic groups and in areas of the world with deficient standards of hygiene (e.g., developing regions), EBV tends to infect children at an early age, and IM is uncommon. In areas with higher standards of hygiene, infection with EBV is often delayed until adulthood, and IM is more prevalent.

EBV is spread by contact with oral secretions. The virus is frequently transmitted from asymptomatic adults to infants and among young adults by transfer of saliva during kissing. Transmission by less intimate contact is rare. EBV has been transmitted by blood transfusion and by bone marrow transplantation. More than 90% of asymptomatic seropositive individuals shed the virus in oropharyngeal secretions. Shedding is increased in immunocompromised patients and those with IM.

PATHOGENESIS

EBV is transmitted by salivary secretions. The virus infects the epithelium of the oropharynx and the salivary glands and is shed from these cells. While B cells may become infected after contact with epithelial cells, studies suggest that lymphocytes in the tonsillar crypts can be infected directly. The virus then spreads through the bloodstream. The proliferation and expansion of EBV-infected B cells along with reactive T cells during IM result in enlargement of lymphoid tissue. Polyclonal activation of B cells leads to the production of antibodies to host-cell and viral proteins. During the acute phase of IM, up to 1 in every 100 B cells in the peripheral blood is infected by EBV; after recovery, 1–50 in every 1 million B cells is infected. During IM, there is an inverted CD4+/CD8+ T-cell ratio. The percentage of CD4+ T cells decreases, while there are large clonal expansions of CD8+ T cells; up to 40% of CD8+ T cells are directed against EBV antigens during acute infection. Memory B cells, not epithelial cells, are the reservoir for EBV in the body. When patients are treated with acyclovir, shedding of EBV from the oropharynx stops but the virus persists in B cells.

The EBV receptor (CD21) on the surface of B cells is also the receptor for the C3d component of complement. Another EBV receptor (CD35) on B cells binds to CD21. Human leukocyte antigen class II serves as a co-receptor for EBV entry into B cells. EBV infection of epithelial cells occurs by virus binding to ephrin A2 and results in viral replication and production of virions. When B cells are infected by EBV in vitro, they become transformed and can proliferate indefinitely. During latent infection of B cells, the EBV nuclear antigens (EBNAs), latent membrane proteins (LMPs), multiple microRNAs, and small EBV RNAs (EBERs) are expressed in vitro. EBV-transformed B cells secrete immunoglobulin; only a small fraction of these cells produce virus.

Cellular immunity is more important than humoral immunity in controlling EBV infection. In the initial phase of infection, suppressor T cells, natural killer (NK) cells, and nonspecific cytotoxic T cells are important in controlling the proliferation of EBV-infected B cells. Levels of markers of T-cell activation and serum interferon γ are elevated. Later in infection, human leukocyte antigen–restricted cytotoxic T cells that recognize EBNAs and LMPs and destroy EBV-infected cells are generated.

If T-cell immunity is compromised, EBV-infected B cells may begin to proliferate. When EBV is associated with lymphoma in immunocompetent persons, virus-induced proliferation is but one step in a multistep process of neoplastic transformation. In many EBV-containing tumors, LMP-1 mimics members of the tumor necrosis factor receptor family (e.g., CD40), transmitting growth-proliferating signals.

CLINICAL MANIFESTATIONS

Signs and Symptoms Most EBV infections in infants and young children either are asymptomatic or present as mild pharyngitis with or without tonsillitis. In contrast, ~75% of infections in adolescents present as IM. IM in the elderly often presents with nonspecific symptoms, including prolonged fever, fatigue, myalgia, and malaise. In contrast, pharyngitis, lymphadenopathy, splenomegaly, and atypical lymphocytes are relatively rare in elderly patients.

The incubation period for IM in young adults is ~4–6 weeks. A prodrome of fatigue, malaise, and myalgia may last for 1–2 weeks before the onset of fever, sore throat, and lymphadenopathy. Fever is usually low-grade and is most common in the first 2 weeks of the illness; however, it may persist for >1 month. Common signs and symptoms are listed along with their frequencies in **Table 199-1**. Lymphadenopathy and pharyngitis are most prominent during the first 2 weeks of the illness, while splenomegaly is more prominent during the second and third weeks. Lymphadenopathy most often affects the posterior cervical nodes but may be generalized. Enlarged lymph nodes are frequently tender and symmetric but are not fixed in place. Pharyngitis, often the most prominent sign, can be accompanied by enlargement of the tonsils with an exudate resembling that of streptococcal pharyngitis. A morbilliform or papular rash, usually on the arms or trunk, develops in ~5% of cases **(Fig. 199-1)**. Earlier studies reported that many patients treated with penicillin derivatives develop a macular rash; penicillin-associated rashes are not predictive of future adverse reactions to penicillins. More recent studies suggest that EBV-associated rashes may occur with similar frequency in those exposed to penicillin derivatives and those not taking these drugs. Erythema nodosum **(Fig. A1-39)** and erythema multiforme **(Fig. A1-24)** also have been described **(Chap. 61)**. The severity of the disease correlates with the levels of CD8+ T cells and EBV DNA in the blood. Most patients have symptoms for 2–4 weeks, but nearly 10% have fatigue that persists for ≥6 months.

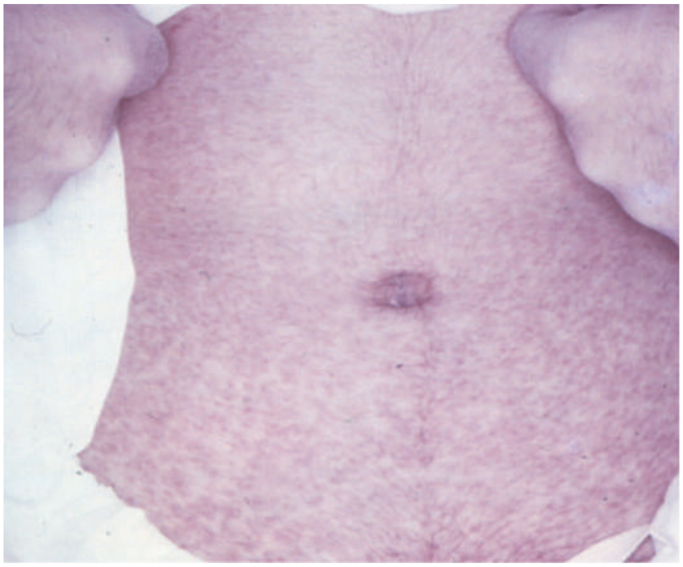

FIGURE 199-1 Rash in a patient with infectious mononucleosis due to Epstein-Barr virus. *(Courtesy of Maria Turner, MD; with permission.)*

TABLE 199-1 Signs and Symptoms of Infectious Mononucleosis

MANIFESTATION	MEDIAN PERCENTAGE OF PATIENTS (RANGE)
Symptoms	
Sore throat	75 (50–87)
Malaise	47 (42–76)
Headache	38 (22–67)
Abdominal pain, nausea, or vomiting	17 (5–25)
Chills	10 (9–11)
Signs	
Lymphadenopathy	95 (83–100)
Fever	93 (60–100)
Pharyngitis or tonsillitis	82 (68–90)
Splenomegaly	51 (43–64)
Hepatomegaly	11 (6–15)
Rash	10 (0–25)
Periorbital edema	13 (2–34)
Palatal enanthem	7 (3–13)
Jaundice	5 (2–10)

Laboratory Findings The white blood cell count is usually elevated and peaks at 10,000–20,000/μL during the second or third week of illness. Lymphocytosis is usually demonstrable, with >10% atypical lymphocytes. The latter cells are enlarged lymphocytes that have abundant cytoplasm, vacuoles, and indentations of the cell membrane **(Fig. 199-2)**. CD8+ T cells predominate among the atypical lymphocytes. Low-grade neutropenia and thrombocytopenia are common during the first month of illness. Liver function is abnormal in >90% of cases. Serum levels of aminotransferases and alkaline phosphatase are usually mildly elevated. The serum concentration of bilirubin is elevated in ~40% of cases.

Complications Most cases of IM are self-limited. Deaths are very rare and are most often due to central nervous system (CNS) complications, splenic rupture, upper-airway obstruction, or bacterial superinfection.

When CNS complications develop, they usually do so during the first 2 weeks of EBV infection; in some patients, especially children,

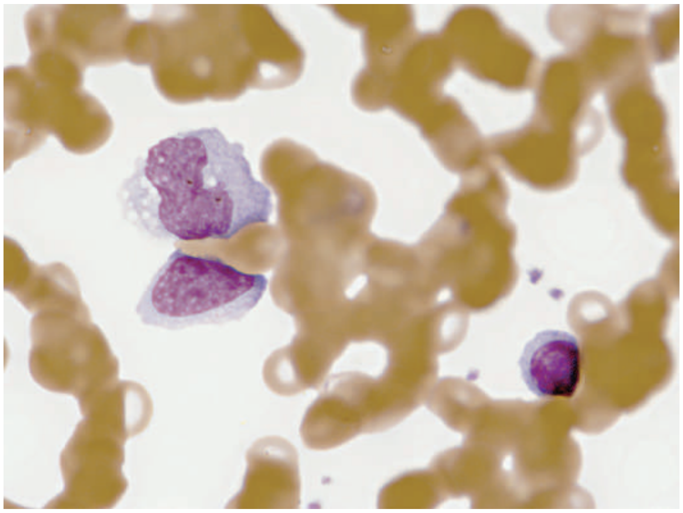

FIGURE 199-2 Atypical lymphocytes from a patient with infectious mononucleosis due to Epstein-Barr virus.

they are the only clinical manifestations of acute infection. Heterophile antibodies and atypical lymphocytes may be absent. Meningitis and encephalitis are the most common neurologic abnormalities, and patients may present with headache, meningismus, or cerebellar ataxia. Acute hemiplegia and psychosis also have been described. The cerebrospinal fluid contains mainly lymphocytes, with occasional atypical lymphocytes. Most cases resolve without neurologic sequelae. Acute EBV infection has also been associated with cranial nerve palsies (especially those involving cranial nerve VII), Guillain-Barré syndrome, acute transverse myelitis, and peripheral neuritis.

Autoimmune hemolytic anemia occurs in ~2% of cases during the first 2 weeks. In most cases, the anemia is Coombs-positive, with cold agglutinins directed against the red blood cell i antigen. Most patients with hemolysis have mild anemia that lasts for 1–2 months, but some patients have severe disease with hemoglobinuria and jaundice. Nonspecific antibody responses may also include rheumatoid factor, antinuclear antibodies, anti–smooth muscle antibodies, antiplatelet antibodies, and cryoglobulins. IM has been associated with red-cell aplasia, severe granulocytopenia, thrombocytopenia, pancytopenia, and hemophagocytic lymphohistiocytosis. The spleen ruptures in <0.5% of cases. Splenic rupture is more common among male than female patients and may manifest as abdominal pain, referred shoulder pain, or hemodynamic compromise.

Hypertrophy of lymphoid tissue in the tonsils or adenoids can result in upper-airway obstruction, as can inflammation and edema of the epiglottis, pharynx, or uvula. About 10% of patients with IM develop streptococcal pharyngitis after their initial sore throat resolves.

Other rare complications associated with acute EBV infection include hepatitis (which can be fulminant), myocarditis or pericarditis, pneumonia with pleural effusion, interstitial nephritis, genital ulcerations, and vasculitis.

EBV-Associated Diseases Other Than IM EBV-associated lymphoproliferative disease has been described in patients with congenital or acquired immunodeficiency, including those with severe combined immunodeficiency, patients with AIDS, and recipients of bone marrow or organ transplants who are receiving immunosuppressive drugs (especially cyclosporine). Proliferating EBV-infected B cells infiltrate lymph nodes and multiple organs, and patients present with fever and lymphadenopathy or gastrointestinal symptoms. Pathologic studies show B-cell hyperplasia or poly- or monoclonal lymphoma.

X-linked lymphoproliferative disease is a recessive disorder of young boys who have a normal response to childhood infections but develop fatal lymphoproliferative disorders after infection with EBV. The protein associated with most cases of this syndrome (SAP, encoded by *SH2D1A*) binds to a protein that mediates interactions of B and T cells. Most patients with this syndrome die of acute IM. Others develop hypogammaglobulinemia, malignant B-cell lymphomas, aplastic anemia, or agranulocytosis. Disease resembling X-linked lymphoproliferative disease, but with more prominent hemophagocytosis, has also been associated with mutations in *BIRC4*. Mutations in *ITK*, *MAGT1*, *CORO1A*, *TNFRSF9*, *IL27RA*, *CD70*, or *CD27* are associated with inability to control EBV and lymphoma. Mutations in other genes, such as *GATA2*, *PIK3CD*, *CTPS1*, *RLTPR*, *RSGRP1*, *TNFRSF9*, and several genes associated with severe combined immunodeficiency, also can predispose to severe or fatal EBV disease as well as other infections. Moreover, IM has proved fatal to some patients with no obvious preexisting immune abnormality.

Oral hairy leukoplakia **(Fig. 199-3)** is an early manifestation of infection with HIV in adults **(Chap. 208)**. Most patients present with raised, white corrugated lesions on the tongue (and occasionally on the buccal mucosa) that contain EBV DNA. Children infected with HIV can develop lymphoid interstitial pneumonitis; EBV DNA is often found in lung tissue from these patients.

Patients with chronic fatigue syndrome may have titers of antibody to EBV that are elevated but are not significantly different from those in healthy EBV-seropositive adults. These patients do not have elevated levels of EBV DNA in the blood. While some patients have malaise and fatigue that persist for weeks or months after IM, persistent EBV infection is not a cause of chronic fatigue syndrome. EBV reactivation (largely based on viral serology) has been associated with post-acute COVID-19 syndrome (PACS), but it is unclear that EBV is a cause of the symptoms of PACS. Chronic active EBV infection is very rare and is distinct from chronic fatigue syndrome. The affected patients have an illness lasting >3 months, with elevated levels of EBV DNA in the blood (in T or NK cells); high titers of antibody to EBV; and evidence of organ involvement, including hepatosplenomegaly, lymphadenopathy, and hepatitis, pneumonitis, uveitis, or neurologic disease. Some have somatic mutations in *DD3X* and other tumor driver genes.

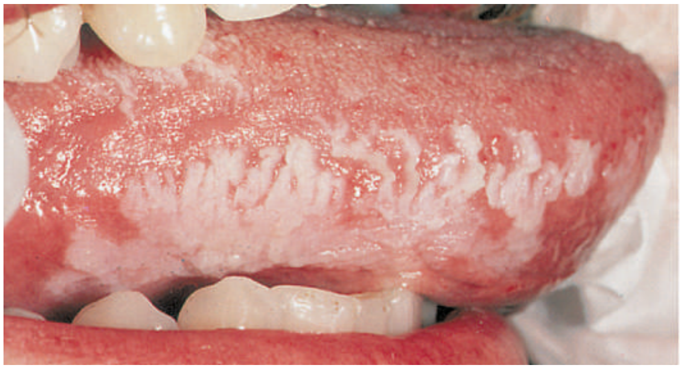

FIGURE 199-3 Oral hairy leukoplakia often presents as white plaques on the lateral surface of the tongue and is associated with Epstein-Barr virus infection.

EBV is associated with several malignancies. About 15% of cases of Burkitt's lymphoma in the United States and ~90% of those in Africa are associated with EBV **(Chap. 113)**. African patients with Burkitt's lymphoma have high levels of antibody to EBV, and their tumor tissue usually contains viral DNA. Malaria in African patients may impair cellular immunity to EBV and induce polyclonal B-cell activation with an expansion of EBV-infected B cells. In addition, malaria may target B cells and result in expansion of germinal centers, with consequently increased activity of activation-induced cytidine deaminase, which can mutate DNA. These changes may enhance the proliferation of B cells with elevated EBV DNA in the bloodstream, thereby increasing the likelihood of a *c-myc* translocation—the hallmark of Burkitt's lymphoma. EBV-containing Burkitt's lymphoma also occurs in patients with AIDS.

Anaplastic nasopharyngeal carcinoma is common in southern China and is uniformly associated with EBV; the affected tissues contain viral DNA and antigens. Patients with nasopharyngeal carcinoma often have elevated titers of antibody to EBV **(Chap. 82)**. Antibody to an EBV protein, BNLF1, in serum is a useful screening marker. Measurement of EBV DNA in plasma is useful for early detection of nasopharyngeal carcinoma. High levels of EBV plasma DNA before treatment or detectable levels of EBV DNA after radiation therapy correlate with lower rates of overall survival and relapse-free survival among patients with nasopharyngeal carcinoma.

Worldwide, the most common EBV-associated malignancy is gastric carcinoma. About 9% of these tumors are EBV-positive including >90% of gastric lymphoepithelioma-like carcinomas **(Chap. 85)**.

EBV has been associated with Hodgkin's lymphoma, especially the mixed-cellularity type **(Chap. 114)**. Patients with Hodgkin's lymphoma often have elevated titers of antibody to EBV. In about half of cases in the United States, viral DNA and antigens are found in Reed-Sternberg cells. The risk of EBV-positive Hodgkin's lymphoma is significantly increased in young adults for several years after EBV-seropositive IM. About 50% of non-Hodgkin's lymphomas in patients with AIDS are EBV-positive.

EBV is present in B cells of lesions from patients with lymphomatoid granulomatosis. In some cases, EBV DNA has been detected in tumors from immunocompetent patients with angiocentric nasal NK/T-cell lymphoma, aggressive NK leukemia/lymphoma, T-cell lymphoma, and CNS lymphoma. Studies have demonstrated viral DNA in leiomyosarcomas from AIDS patients and in smooth-muscle tumors from organ transplant recipients. Virtually all CNS lymphomas in AIDS patients are associated

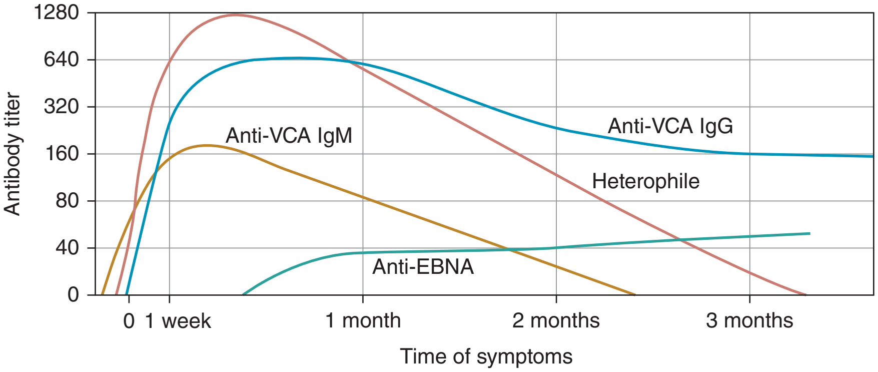

FIGURE 199-4 Pattern of Epstein-Barr virus (EBV) serology during acute infection. EBNA, Epstein-Barr nuclear antigen; VCA, viral capsid antigen. *(Reproduced with permission from JI Cohen, in NS Young et al [eds]: Clinical Hematology. Philadelphia, Mosby, 2006.)*

with EBV. EBV has been associated with multiple sclerosis; an epidemiology study of over a million military personnel found that the risk of multiple sclerosis was 32-fold higher after primary EBV infection, but not after other virus infections. In addition, a history of IM and higher levels of antibodies to EBNA before the onset of disease is more common in persons with multiple sclerosis than in the general population; additional research on the role of EBV in multiple sclerosis is needed.

DIAGNOSIS

Serologic Testing (Fig. 199-4) The heterophile test is used for the diagnosis of IM in children and adults. In the test for this antibody, human serum is absorbed with guinea pig kidney, and the heterophile titer is defined as the greatest serum dilution that agglutinates sheep, horse, or cow erythrocytes. The heterophile antibody does not interact with EBV proteins. A titer of ≥40 is diagnostic of acute EBV infection in a patient who has symptoms compatible with IM and atypical lymphocytes. Tests for heterophile antibodies are positive in 40% of patients with IM during the first week of illness and in 80–90% during the third week. Therefore, repeated testing may be necessary, especially if the initial test is performed early. Tests usually remain positive for 3 months after the onset of illness, but heterophile antibodies can persist for up to 1 year. These antibodies usually are not detectable in children <5 years of age, in the elderly, or in patients presenting with symptoms not typical of IM. The commercially available monospot test for heterophile antibodies is somewhat more sensitive than the classic heterophile test. The monospot test is ~75% sensitive and ~90% specific compared with EBV-specific serologies (see below). False-positive monospot results are more common among persons with connective tissue disease, lymphoma, viral hepatitis, and malaria.

EBV-specific antibody testing is used for patients with suspected acute EBV infection who lack heterophile antibodies and for patients with atypical infections. Titers of IgM and IgG antibodies to viral capsid antigen (VCA) are elevated in the serum of >90% of patients at the onset of disease. IgM antibody to VCA is most useful for the diagnosis of acute IM because it is present at elevated titers only during the first 2–3 months of the disease; in contrast, IgG antibody to VCA usually is not useful for diagnosis of IM but often is used to assess past exposure to EBV because it persists for life. Seroconversion to EBNA positivity also is useful for the diagnosis of acute infection with EBV. Antibodies to EBNA become detectable relatively late (3–6 weeks after the onset of symptoms) in nearly all cases of acute EBV infection and persist for the lifetime of the patient. These antibodies may be lacking in immunodeficient patients and in those with chronic active EBV disease.

Titers of other antibodies also may be elevated in IM; however, these elevations are less useful for diagnosis. Antibodies to early antigens are detectable 3–4 weeks after the onset of symptoms in patients with IM. About 70% of individuals with IM have antibodies to early antigen diffuse (EA-D) during the illness; the presence of EA-D antibodies is especially likely in patients with relatively severe disease. These antibodies usually persist for only 3–6 months. Levels of EA-D antibodies are elevated in patients with nasopharyngeal carcinoma or chronic active EBV infection. Antibodies to early antigen restricted (EA-R) are often found at elevated titers in patients with African Burkitt's lymphoma or chronic active EBV infection; however, they are not useful for diagnosis. IgA antibodies to EBV antigens have proved useful for the identification of patients with nasopharyngeal carcinoma and of persons at high risk for the disease.

Other Studies Detection of EBV DNA, RNA, or proteins has been valuable in demonstrating the association of the virus with various malignancies. The polymerase chain reaction has been used to detect EBV DNA in the cerebrospinal fluid of some AIDS patients with CNS lymphomas and to monitor the amount of EBV DNA in the blood of patients with lymphoproliferative disease. Detection of high levels of EBV DNA in blood for a few days to several weeks after the onset of IM may be useful if serologic studies yield equivocal results. Culture of EBV from throat washings or blood is not helpful in the diagnosis of acute infection, since EBV persists in the oropharynx and in B cells for the lifetime of the infected individual.

Differential Diagnosis Whereas ~90% of cases of IM are due to EBV, 5–10% of cases are due to cytomegalovirus (CMV) (Chap. 200). CMV is the most common cause of heterophile-negative mononucleosis; less common causes of IM and differences from IM due to EBV are shown in Table 199-2.

TABLE 199-2 Differential Diagnosis of Infectious Mononucleosis

	SIGN OR SYMPTOM				
ETIOLOGY	FEVER	ADENOPATHY	SORE THROAT	ATYPICAL LYMPHOCYTES	DIFFERENCES FROM EBV MONONUCLEOSIS
EBV infection	+	+	+	+	—
CMV infection	+	±	±	+	Older age at presentation, longer duration of fever
HIV infection	+	+	+	±	Diffuse rash, oral/genital ulcers, aseptic meningitis
Toxoplasmosis	+	+	±	±	Less splenomegaly; exposure to cats or raw meat
HHV-6 infection	+	+	+	+	Older age at presentation
Streptococcal pharyngitis	+	+	+	–	No splenomegaly, less fatigue
Viral hepatitis	+	±	–	±	Higher aminotransferase levels
Rubella	+	+	±	±	Maculopapular rash, no splenomegaly
Lymphoma	+	+	+	+	Fixed, nontender lymph nodes
Drugs[a]	+	+	–	±	Occurs at any age

[a] Most commonly phenytoin, carbamazepine, sulfonamides, or minocycline.

Abbreviations: CMV, cytomegalovirus; EBV, Epstein-Barr virus; HHV, human herpesvirus.

TREATMENT

EBV-Associated Disease

Therapy for IM consists of supportive measures, with rest and analgesia. Excessive physical activity during the first month should be avoided to reduce the possibility of splenic rupture, which often necessitates splenectomy. Glucocorticoid therapy is not indicated for uncomplicated IM and in fact may predispose to bacterial superinfection. Prednisone (40–60 mg/d for 2–3 days, with subsequent tapering of the dose over 1–2 weeks) has been used for the prevention of airway obstruction in patients with severe tonsillar hypertrophy, for autoimmune hemolytic anemia, for hemophagocytic lymphohistiocytosis, and for severe thrombocytopenia. Glucocorticoids have also been administered to rare patients with severe malaise and fever and to patients with severe CNS or cardiac disease.

Acyclovir has had no significant clinical impact on IM in controlled trials. In one study, the combination of acyclovir and prednisolone had no significant effect on the duration of symptoms of IM.

Acyclovir, at a dosage of 400–800 mg five times daily, has been effective for the treatment of oral hairy leukoplakia (despite common relapses). Posttransplantation EBV lymphoproliferative disease **(Chap. 148)** generally does not respond to antiviral therapy. When possible, therapy should be directed toward reduction of immunosuppression. Antibody to CD20 (rituximab) has been effective in some cases. Infusions of donor lymphocytes are often effective for stem cell transplant recipients, although graft-versus-host disease can occur. Infusions of HLA-matched EBV-specific cytotoxic T cells have been used to prevent EBV lymphoproliferative disease in high-risk settings as well as to treat the disease. Interferon α administration, cytotoxic chemotherapy, and radiation therapy (especially for CNS lesions) also have been used. Infusion of autologous EBV-specific cytotoxic T lymphocytes has shown promise in small studies of patients with nasopharyngeal carcinoma and Hodgkin's lymphoma. Treatment of several cases of X-linked lymphoproliferative disease with antibody to CD20 resulted in a successful outcome of what otherwise would probably have been fatal acute EBV infection.

PREVENTION

The isolation of patients with IM is unnecessary. A vaccine directed against the major EBV glycoprotein reduced the frequency of IM but did not affect the rate of asymptomatic infection in a phase 2 trial. Additional vaccines are in clinical trials.

FURTHER READING

Bjornevik K et al: Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. Science 375:296, 2022.

Chan KCA et al: Analysis of plasma Epstein-Barr virus DNA to screen for nasopharyngeal cancer. N Engl J Med 377:513, 2017.

Cohen JI et al: Epstein-Barr virus NK and T cell lymphoproliferative disease: Report of a 2018 international meeting. Leuk Lymphoma 61:808, 2020.

Dierickx D, Habermann TM: Post-transplantation lymphoproliferative disorders in adults. N Engl J Med 378:549, 2018.

Li T et al: Anti-Epstein-Barr virus BNLF2b for mass screening for nasopharyngeal cancer. N Engl J Med 389:808, 2023.

Mahaden JM et al: Tabelecleucel for allogeneic haematopoietic stem-cell or solid organ transplant recipients with Epstein-Barr virus-positive post-transplant lymphoproliferative disease after failure of rituximab or rituximab and chemotherapy (ALLELE): A phase 3, multicentre, open-label trial. Lancet Oncol 25:376, 2024.

Murray PG, Young LS: An etiological role for the Epstein-Barr virus in the pathogenesis of classical Hodgkin lymphoma. Blood 134:591, 2019.

Tangye SG, Latour S: Primary immunodeficiencies reveal the molecular requirements for effective host defense against EBV infection. Blood 135:644, 2020.

200

Cytomegalovirus and Human Herpesvirus Types 6, 7, and 8

Camille Nelson Kotton, Martin S. Hirsch

CYTOMEGALOVIRUS

DEFINITION

Cytomegalovirus (CMV), which was initially isolated from patients with congenital cytomegalic inclusion disease, is now recognized as an important pathogen in all age groups. In addition to inducing severe birth defects, CMV causes a wide spectrum of disorders in older children and adults, ranging from an asymptomatic subclinical infection to a mononucleosis syndrome in healthy individuals to disseminated disease in immunocompromised patients. Human CMV is one of several related species-specific viruses that cause similar diseases in various animals. All are associated with the production of characteristic enlarged cells—hence the name *cytomegalovirus*.

CMV, a β-herpesvirus, has double-stranded DNA, four species of mRNA, a protein capsid, and a lipoprotein envelope. Like other herpesviruses, CMV demonstrates icosahedral symmetry, replicates in the cell nucleus, and can cause either a lytic and productive or a latent infection. CMV can be distinguished from other herpesviruses by certain biologic properties, such as host range and type of cytopathology. Viral replication is associated with the production of large intranuclear inclusions and smaller cytoplasmic inclusions. CMV appears to replicate in a variety of cell types in vivo; in tissue culture it grows preferentially in fibroblasts. Although there is little clinical evidence that CMV is oncogenic in vivo, it does transform fibroblasts in rare instances, and genomic transforming fragments have been identified.

EPIDEMIOLOGY

CMV has a worldwide distribution. In many low- and middle-income regions, nearly all adults are seropositive for CMV, whereas only half of adults in the United States and Canada are seropositive. Immunocompromised CMV seropositive adults are more likely to undergo reactivation disease rather than primary infection. Data generated in specific regions should be considered in the context of local seropositivity rates, when appropriate.

Based on a recent systematic review, the overall prevalence of congenital CMV is estimated to be 0.67%, with a threefold higher prevalence in low- and middle-income countries (1.42%) compared with high-income countries (0.48%). Communal living and poor personal hygiene facilitate spread. Perinatal and early childhood infections are common. CMV may be present in breast milk, saliva, feces, and urine. Transmission can occur among young children in day-care centers and has been traced from infected toddler to pregnant mother to developing fetus. When an infected child introduces CMV into a household, 50% of nonimmune family members seroconvert within 6 months.

CMV is not readily spread by casual contact but rather requires repeated or prolonged intimate exposure for transmission. In late adolescence and young adulthood, CMV is often transmitted sexually, and asymptomatic carriage in semen or cervical secretions is common. Transfusion of CMV-seropositive blood products containing viable leukocytes may transmit CMV, with a frequency of 0.14–10% per unit transfused, although use of leukocyte-reduced or CMV-seronegative blood significantly decreases the risk of CMV transmission.

Once infected, an individual generally carries CMV for life, similar to other herpes viruses. The infection usually remains silent. CMV reactivation syndromes develop more frequently, however, when T lymphocyte–mediated immunity is compromised—for example, after organ transplantation, with lymphoid neoplasms and certain acquired immunodeficiencies (in particular, HIV infection; **Chap. 208**), or during critical illness in intensive care units. Most primary CMV infections in organ transplant recipients **(Chap. 148)** result from transmission via the graft or blood products. In CMV-seropositive transplant recipients, infection results from reactivation of latent virus in the recipients or from infection by a new strain from the donor. CMV infection may be associated with diseases as diverse as coronary artery stenosis and malignant gliomas, although these associations require further validation.

PATHOGENESIS

Congenital CMV infection can result from either primary or reactivation infection of the mother. However, clinical disease in the fetus or newborn is related largely to primary maternal infection **(Table 200-1)**. The major factors determining the severity of congenital infection are unclear, although a deficient capacity to produce precipitating antibodies and to mount T cell responses to CMV is associated with relatively severe disease.

Primary infection with CMV in late childhood or adulthood is often associated with a vigorous T lymphocyte response that may contribute to the development of a mononucleosis syndrome similar to that observed with Epstein-Barr virus **(Chap. 199)**. The hallmark of such infection is the appearance of atypical lymphocytes in the peripheral blood; these cells are predominantly activated CD8+ T lymphocytes. Polyclonal activation of B cells by CMV contributes to the development of rheumatoid factors and other autoantibodies during mononucleosis.

Once acquired, CMV persists indefinitely in host tissues. The sites of persistent infection may include multiple cell types and various organs. Transmission via blood transfusion or organ transplantation is due primarily to silent infection in these tissues. If the host's T cell responses become compromised by disease or by iatrogenic immunosuppression, latent virus can reactivate to cause a variety of syndromes. Chronic antigenic stimulation in the presence of immunosuppression (for example, after organ transplantation) appears to be an ideal setting for CMV activation and CMV disease. Certain particularly potent suppressants of T cell immunity (e.g., antithymocyte globulin, alemtuzumab) are associated with a high rate of clinical CMV syndromes. CMV may itself contribute to further T lymphocyte hyporesponsiveness, which often precedes superinfection with other opportunistic pathogens such as bacteria, molds, other viruses, and *Pneumocystis*.

PATHOLOGY

Cytomegalic cells in vivo (presumed to be infected epithelial cells) are two to four times larger than surrounding cells and often contain an 8- to 10-μm intranuclear inclusion that is eccentrically placed and is surrounded by a clear halo, producing an "owl's eye" appearance. Smaller granular cytoplasmic inclusions are demonstrated occasionally. Cytomegalic cells are found in a wide variety of organs, including the salivary gland, lung, liver, kidney, intestine, pancreas, adrenal gland, and central nervous system.

The cellular inflammatory response to infection consists of plasma cells, lymphocytes, and monocyte-macrophages. Granulomatous reactions occasionally develop, particularly in the liver. Immunopathologic reactions may contribute to CMV disease. Immune complexes have been detected in infected infants, sometimes in association with CMV-related glomerulopathies. Immune-complex glomerulopathy has also been observed in some CMV-infected patients after renal transplantation.

CLINICAL MANIFESTATIONS

Congenital CMV Infection Fetal infections range from subclinical to severe and disseminated. CMV seroconversion rates during pregnancy range from 1% to 7%. Of infants born to mothers with primary CMV infections during pregnancy, 5–20% will develop clinical manifestations, with a mortality rate of ~5%. Petechiae, hepatosplenomegaly, and jaundice are the most common presenting features (60–80% of cases). They can have "blueberry muffin"–like hemorrhagic purpuric eruptions, which when biopsied show histopathology with dermal erythropoiesis. Infections during the first trimester are associated with up to 40–50% of infected neonates developing sensorineural complications. Microcephaly with or without cerebral calcifications, intrauterine growth retardation, and prematurity are reported in 30–50% of cases. Inguinal hernias and chorioretinitis are less common. Laboratory abnormalities include elevated alanine aminotransferase levels in serum, thrombocytopenia, conjugated hyperbilirubinemia, hemolysis, and elevated protein levels in cerebrospinal fluid. The prognosis for severely infected infants is poor, and few survivors escape intellectual or hearing difficulties later in childhood. The differential diagnosis of cytomegalic inclusion disease in infants includes syphilis, toxoplasmosis, bacterial sepsis, and infection with a variety of viruses, including rubella, Zika, or herpes simplex virus.

Most congenital CMV infections are clinically inapparent at birth. Of asymptomatically infected infants, 7–11% develop sensorineural hearing loss over a 5-year period.

Perinatal CMV Infection The newborn may acquire CMV at delivery by passage through an infected birth canal or by postnatal contact with infected breast milk or other maternal secretions. Of infants who are breast-fed for >1 month by seropositive mothers, 40–60% become infected. Iatrogenic transmission can result from blood transfusion; use of leukocyte-reduced or CMV-seronegative blood products for transfusion into low-birth-weight seronegative infants or seronegative pregnant women decreases risk.

The great majority of infants infected at or after delivery remain asymptomatic. However, protracted interstitial pneumonitis has been associated with perinatally acquired CMV infection, particularly in premature infants, and occasionally has been accompanied by infection with *Chlamydia trachomatis*, *Pneumocystis*, or *Ureaplasma urealyticum*. Poor weight gain, adenopathy, rash, hepatitis, anemia, and atypical lymphocytosis may also be found, and CMV excretion often persists for months or years.

CMV Mononucleosis The most common clinical manifestation of CMV infection in immunocompetent hosts beyond the neonatal period is a heterophile antibody–negative mononucleosis syndrome, which may develop spontaneously or follow transfusion of

TABLE 200-1 Cytomegalovirus (CMV) Disease in the Immunocompromised Host

POPULATION	RISK FACTORS	PRINCIPAL SYNDROME(S)	TREATMENT	PREVENTION
Fetus/neonate	Primary maternal infection/ early pregnancy, reactivation infection	Cytomegalic inclusion disease	Ganciclovir followed by valganciclovir for symptomatic neonates	Avoidance of exposure; education of pregnant women about risks
Organ transplant recipient	Seropositivity of donor and/or recipient; potent immunosuppressive regimen; treatment of rejection	Febrile leukopenia (CMV syndrome); gastrointestinal disease; pneumonia; early infection may be asymptomatic	Ganciclovir or valganciclovir, ± CMV immunoglobulin for severe or resistant/refractory disease	Prophylaxis with ganciclovir, valganciclovir, or letermovir, or preemptive therapy
Hematopoietic stem cell transplant recipient	Graft-vs-host disease; older age of recipient; seropositive recipient; viremia	Pneumonia; gastrointestinal disease	Maribavir, foscarnet, or cidoforvir for resistant/ refractory infection	Prophylaxis with letermovir, ganciclovir, or valganciclovir, or preemptive therapy
Person with HIV	<50 CD4+ T cells/μL; CMV seropositivity	Retinitis; gastrointestinal disease; neurologic disease	Ganciclovir, valganciclovir, foscarnet, or cidofovir	Highly active antiretroviral therapy; prophylaxis with valganciclovir

leukocyte-containing blood products. Although the syndrome occurs at all ages, it most often involves sexually active young adults. With incubation periods of 20–60 days, the illness generally lasts for 2–6 weeks. Prolonged high fevers, sometimes with chills, profound fatigue, and malaise, characterize this disorder. Myalgias, headache, and splenomegaly are common, but in CMV mononucleosis (as opposed to Epstein-Barr virus mononucleosis), exudative pharyngitis and cervical lymphadenopathy are rare. Occasional patients develop rubelliform rashes, often after exposure to ampicillin or certain other antibiotics. Less common are interstitial or segmental pneumonia, myocarditis, pleuritis, arthritis, splanchnic vein thrombosis, and encephalitis. In rare cases, Guillain-Barré syndrome complicates CMV mononucleosis. The characteristic laboratory abnormality of CMV mononucleosis is relative lymphocytosis in peripheral blood, with >10% atypical lymphocytes. Total leukocyte counts may be low, normal, or markedly elevated. Although significant jaundice is uncommon, serum aminotransferase and alkaline phosphatase levels are often moderately elevated. Heterophile antibodies are absent; however, transient immunologic abnormalities are common and may include the presence of cryoglobulins, rheumatoid factors, cold agglutinins, and antinuclear antibodies. Hemolytic anemia, thrombocytopenia, and granulocytopenia complicate recovery in rare instances.

Most patients recover without sequelae, although postviral asthenia may persist for months. The excretion of CMV in urine, genital secretions, and/or saliva often continues for months or years. Rarely, CMV infection is fatal in immunocompetent hosts; survivors can have recurrent episodes of fever and malaise, sometimes associated with autonomic nervous system dysfunction (e.g., attacks of sweating or flushing).

CMV Infection in the Immunocompromised Host (Table 200-1) CMV is the most common viral pathogen complicating organ transplantation **(Chap. 148)**. In recipients of kidney, heart, lung, liver, pancreas, and vascularized composite (hand, face, other) transplants, CMV infection may result in a variety of clinical manifestations, including fever and leukopenia, hepatitis, colitis, pneumonitis, esophagitis, gastritis, and retinitis. CMV disease is an independent risk factor for both graft loss and death. Without prophylaxis, the period of maximal risk is between 1 and 4 months after transplantation. Disease likelihood and viral replication levels generally are greater after primary infection than after reactivation. Molecular studies indicate that seropositive organ transplant recipients are susceptible to infection with donor-derived, genotypically variant CMV strains. Reactivation infection, although common, is less likely than primary infection to be clinically significant. The overall risk of clinical disease is related to various factors, such as serologic mismatch (donor seropositive, recipient seronegative), degree of immunosuppression, use of antilymphocyte antibodies, lack of anti-CMV prophylaxis, and co-infection with other pathogens. The transplanted organ is particularly vulnerable as a target for CMV infection; thus, there is a tendency for CMV hepatitis to follow liver transplantation and for CMV pneumonitis to follow lung transplantation.

CMV viremia occurs in roughly one-third of hematopoietic stem cell transplant (HSCT) recipients; the risk of severe disease may be reduced by prophylaxis or preemptive therapy with antiviral drugs. The risk is greatest in the first 100 days after transplantation, and identified risk factors include certain types of immunosuppressive therapy, an allogeneic (rather than an autologous) graft, acute graft-versus-host disease, older age, and recipient seropositivity prior to transplant.

CMV is an important pathogen in persons with HIV (PWH) **(Chap. 208)**, in whom it may cause retinitis or disseminated disease, particularly when peripheral-blood CD4+ T cell counts fall below 50/μL. In PWH and CD4+ T cells under 200/uL, high blood CMV viral loads are a predictor of severe disease and death. As treatment for underlying HIV infection has improved, the incidence of serious CMV infections (e.g., retinitis) has decreased. During the first few weeks after starting highly active antiretroviral therapy, however, acute flare-ups of CMV retinitis may occur secondary to an immune reconstitution inflammatory syndrome (IRIS).

Syndromes produced by CMV in immunocompromised hosts (“CMV syndrome”) often begin with fatigue, fever, malaise, anorexia, night sweats, and arthralgias or myalgias. Liver function abnormalities, leukopenia, thrombocytopenia, and atypical lymphocytosis may be observed during these episodes. Without treatment, CMV infection may progress to more severe end-organ disease. The development of tachypnea, hypoxemia, and nonproductive cough signals respiratory involvement. Radiologic examination of the lung often shows bilateral interstitial or reticulonodular infiltrates that begin in the periphery of the lower lobes and spread centrally and superiorly; localized segmental, nodular, or alveolar patterns are less common. The differential diagnosis includes *Pneumocystis* infection; other viral, bacterial, or fungal infections; pulmonary hemorrhage; and injury secondary to irradiation or to treatment with cytotoxic drugs.

Gastrointestinal CMV involvement may be localized or extensive and almost exclusively affects immunocompromised hosts. Colitis is the most common clinical manifestation in organ transplant recipients. Ulcers of the esophagus, stomach, small intestine, or colon may result in bleeding or perforation. Clinicians should be aware that blood tests such as CMV antigenemia and viral load testing may yield negative results in the setting of intestinal disease. CMV infection may lead to exacerbations of underlying ulcerative colitis. Hepatitis occurs frequently, particularly after liver transplantation. Acalculous cholecystitis and adrenalitis also have been described.

CMV rarely causes meningoencephalitis in otherwise healthy individuals. Two forms of CMV encephalitis are seen in PWH. One resembles HIV encephalitis and often presents as progressive dementia; the other is a ventriculoencephalitis characterized by cranial-nerve deficits, nystagmus, disorientation, lethargy, and ventriculomegaly. In immunocompromised patients, CMV can also cause subacute progressive polyradiculopathy, which is often reversible if recognized and treated promptly.

CMV retinitis is an important cause of blindness in immunocompromised patients, particularly patients with advanced AIDS **(Chap. 208)**. Early lesions consist of small, opaque, white areas of granular retinal necrosis that spread in a centrifugal manner and are later accompanied by hemorrhages, vessel sheathing, and retinal edema **(Fig. 200-1)**. CMV retinopathy must be distinguished from that due to other conditions, including toxoplasmosis, candidiasis, and herpes simplex virus infection.

Fatal CMV infections are often associated with persistent viremia and the involvement of multiple organ systems. Progressive pulmonary infiltrates, pancytopenia, hyperamylasemia, and hypotension are characteristic features that are frequently found in conjunction with a terminal bacterial, fungal, or protozoan superinfection. Extensive adrenal necrosis with CMV inclusions is often documented at autopsy, as is CMV involvement of many other organs.

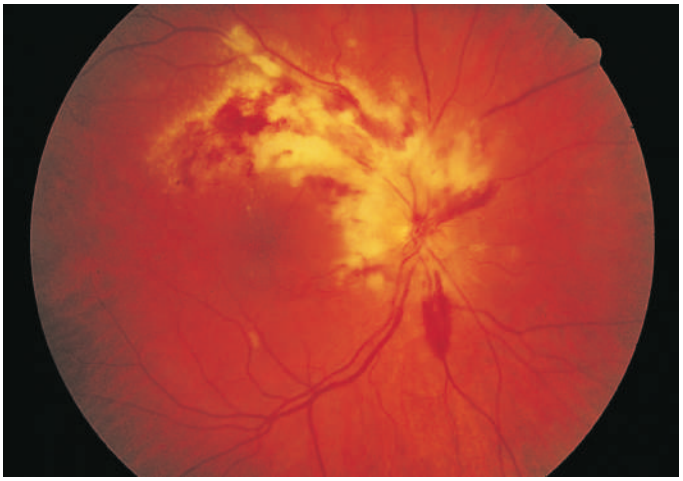

FIGURE 200-1 Cytomegalovirus infection in a patient with AIDS may appear as an arcuate zone of retinitis with hemorrhages and optic disk swelling. Often CMV is confined to the retinal periphery, beyond view of the direct ophthalmoscope.

DIAGNOSIS

CMV infection usually cannot be diagnosed reliably on clinical grounds alone. Isolation of CMV or detection of its antigens or DNA in appropriate clinical specimens is the preferred approach. The most common method of detection is quantitative nucleic acid testing (QNAT) for CMV by polymerase chain reaction (PCR) technology, for which blood or other specimens can be used; some centers use a CMV antigenemia test, an immunofluorescence assay that detects CMV antigens (pp65) in peripheral-blood leukocytes. CMV DNA in cerebrospinal fluid is useful in the diagnosis of CMV encephalitis or polyradiculopathy.

Virus excretion and/or viremia is readily detected by culture of appropriate specimens on human fibroblast monolayers. If CMV titers are high, as is common in congenital disseminated infection and in AIDS, characteristic cytopathic effects may be detected within a few days. However, in some situations (e.g., CMV mononucleosis), viral titers are low, and cytopathic effects may take several weeks to appear. Many laboratories expedite diagnosis with an overnight tissue-culture method (shell vial assay) involving centrifugation and an immunocytochemical detection technique employing monoclonal antibodies to an immediate-early CMV antigen. Isolation of CMV from urine, stool, or saliva does not, by itself, constitute proof of acute infection, since excretion from these sites may continue for months or years after illness. Detection of viremia by QNAT or antigenemia testing is a better predictor of acute infection.

A variety of serologic assays detect antibody to CMV. An increased level of IgG antibody to CMV may not be detectable for up to 4 weeks after primary infection. Detection of CMV-specific IgM is sometimes useful in the diagnosis of recent or active infection; however, circulating rheumatoid factors may result in occasional false-positive IgM tests. Serology is more helpful when used to predict risk of CMV infection and disease in transplant recipients and is not recommended to diagnose acute disease in immunocompromised patients.

PREVENTION

Prevention of CMV infection and disease in organ transplant and HSCT recipients is usually based on one of two methods: universal prophylaxis or preemptive therapy. With universal prophylaxis, antiviral drugs are used for a defined period, often 3 or 6 months. One clinical trial demonstrated that, in CMV-seronegative kidney transplant recipients with seropositive donors, prophylaxis with valganciclovir was more effective at prevention when given for 200 days rather than 100 days. With preemptive therapy, patients are monitored weekly for CMV viremia, and antiviral treatment is initiated once viremia is detected. Because of the bone marrow–suppressive effects of universal prophylaxis with ganciclovir/valganciclovir, preemptive therapy has been more commonly employed in HSCT recipients; letermovir provides excellent prophylaxis in higher-risk patients. For PWH, CMV end-organ disease is best prevented by using antiretroviral therapy sufficient to maintain CD4+ T-cell counts above 100/μL; primary prophylaxis with ganciclovir or valganciclovir is not recommended.

Several additional measures are useful for the prevention of CMV transmission to CMV-naïve, high-risk patients. The use of CMV-seronegative or leukocyte-depleted blood significantly decreases the rate of transfusion-associated transmission. In a placebo-controlled trial, a CMV glycoprotein B vaccine reduced infection rates among 464 CMV-seronegative women; this outcome raises the possibility that a vaccine will reduce rates of congenital infection, but further studies must validate this approach. A conditionally replication-defective CMV, termed V160, was recently shown in a phase 2 clinical trial to be well tolerated and immunogenic although three doses of vaccine did not reduce the incidence of primary CMV infection in CMV-seronegative women compared with placebo. The Triplex vaccine is in clinical trials in stem cell and liver transplant recipients; this vaccine has a recombinant viral vector with genes expressing three immunodominant proteins: UL83 (pp65), UL123 (IE1), and UL122 (IE2). An mRNA CMV vaccine, mRNA-1647, is in clinical trials; it contains six mRNAs that encode the membrane-bound pentamer complex and the full-length membrane-bound glycoprotein B. CMV immune globulin (intravenous immunoglobulin enriched from CMV-positive donors) has been studied in a variety of clinical situations (primary CMV infection in pregnancy, HSCT, solid organ transplantation), with conflicting results, and is used much less often in the era of multiple effective antiviral agents. A recent trial in pregnant women using CMV immunoglobulin before 24 weeks' gestation did not result in a lower incidence of congenital CMV infection or perinatal death compared with placebo.

Prophylactic acyclovir or valacyclovir at high doses may reduce rates of CMV infection and disease in renal transplant recipients; neither drug is effective in the treatment of active CMV disease. Valacyclovir may also reduce the risk of maternal-fetal CMV transmission in pregnant women with primary CMV infection.

TREATMENT

Cytomegalovirus Infection

Ganciclovir is a guanosine derivative that has considerably more activity against CMV than its congener acyclovir. After intracellular conversion by a viral phosphotransferase encoded by CMV gene region UL97, ganciclovir triphosphate is a selective inhibitor of CMV DNA polymerase. Several clinical studies have indicated response rates of 70–90% among people with HIV who are given ganciclovir for the treatment of CMV retinitis or colitis. In severe infections (e.g., CMV pneumonia in HSCT recipients), ganciclovir is sometimes combined with CMV immune globulin. Prophylactic or suppressive ganciclovir may be useful in high-risk HSCT or organ transplant recipients (e.g., those who are CMV-seropositive before transplantation). In many PWH, those with persistently low CD4+ T cell counts, and those with CMV disease, clinical and virologic relapses occur promptly if treatment with ganciclovir is discontinued. Therefore, prolonged prevention regimens may be recommended for such patients. Resistance to ganciclovir is more common among patients exposed to antivirals for >3 months and is usually related to mutations in the CMV UL97 gene (or, less commonly, the UL54 gene, or UL56 after letermovir exposure, or UL27 after maribavir exposure). The advent of CMV genotyping for resistance mutations has made it possible to rapidly obtain information regarding optimal treatment approaches against clinically resistant CMV.

Valganciclovir is an orally bioavailable prodrug that is rapidly metabolized to ganciclovir in intestinal tissues and the liver. Approximately 60–70% of an oral dose of valganciclovir is absorbed. An oral valganciclovir dose of 900 mg results in ganciclovir blood levels similar to those obtained with an IV ganciclovir dose of 5 mg/kg. Valganciclovir appears to be as effective as IV ganciclovir for both CMV induction (treatment) and prevention regimens, also offering the advantage of oral dosing.

Ganciclovir or valganciclovir therapy for CMV disease consists of a 14- to 21-day induction course (5 mg/kg IV twice daily for ganciclovir or 900 mg PO twice daily for valganciclovir), sometimes followed by suppressive therapy (e.g., valganciclovir, 900 mg/d). Peripheral-blood neutropenia develops in roughly one-quarter of treated patients but may be ameliorated by granulocyte colony-stimulating factor. Guidelines recommend checking weekly bloodwork for CMV NAT or antigenemia to ascertain virologic response. Whether to use suppressive therapy should depend on the overall level of immunocompromise and the risk of recurrent disease. Discontinuation of preventative regimens should be considered in PWH who, while receiving antiretroviral therapy, have a sustained (3- to 6-month) increase in CD4+ T cell counts to >100/μL. Compared with shorter (6-week) courses, prolonged (6-month) courses of valganciclovir had beneficial effects on hearing and developmental outcomes in infants with congenital CMV infection. For treatment of CMV retinitis, some clinicians use intravitreal injections of ganciclovir or foscarnet (see below) plus oral valganciclovir or intravenous ganciclovir, although no clinical trials have compared these approaches.

Foscarnet (sodium phosphonoformate) inhibits CMV DNA polymerase. Because this agent does not require phosphorylation to be active, it is also effective against most ganciclovir-resistant isolates. Foscarnet is less well tolerated than ganciclovir and causes considerable toxicity, including renal dysfunction, hypomagnesemia, hypokalemia, hypocalcemia, genital ulcers, dysuria, nausea, and paresthesia. Moreover, foscarnet administration requires the use of an infusion pump and close clinical monitoring. With aggressive hydration and dose adjustments for renal dysfunction, the toxicity of foscarnet can be reduced. The use of foscarnet should be avoided when a saline load cannot be tolerated (e.g., in cardiac insufficiency). The approved induction regimen is 60 mg/kg every 8 h for 2 weeks, although 90 mg/kg every 12 h is equally effective and no more toxic. Maintenance infusions should deliver 90–120 mg/kg once daily. No oral preparation is available. Foscarnet-resistant virus may emerge during extended therapy. This drug is used more frequently after HSCT than in other situations to avoid the myelosuppressive effects of ganciclovir; in general, foscarnet has been the first choice for infections with ganciclovir-resistant CMV. *Maribavir* is now being used for resistant/refractory CMV infections, as it was shown in a phase 3 trial to be more effective and have less toxicity compared with alternative therapies. *Letermovir* has efficacy for prophylaxis after HSCT and organ transplantation but induces rapid development of resistance when used during active infection; some experts recommend letermovir for prevention after treatment of resistant/refractory CMV.

Cidofovir is a nucleotide analogue with a long intracellular half-life that allows intermittent IV administration. Induction regimens of 5 mg/kg weekly for 2 weeks are followed by secondary prevention regimens of 3–5 mg/kg every 2 weeks. Cidofovir can cause severe nephrotoxicity through dose-dependent proximal tubular cell injury; however, this adverse effect can be tempered somewhat by saline hydration and probenecid. Cidofovir is used primarily for ganciclovir-resistant virus but is being replaced by maribavir; it may be useful for prevention after treatment of resistant/refractory CMV, as it could be given every 2 weeks.

HUMAN HERPESVIRUS (HHV) TYPES 6, 7, AND 8

HHV-6 AND HHV-7

HHV-6 and -7 seropositivity rates are generally high throughout the world. HHV-6 was first isolated in 1986 from peripheral-blood leukocytes of six persons with various lymphoproliferative disorders. There are two genetically distinct variants (HHV-6A and HHV-6B). HHV-6 appears to be transmitted by saliva and possibly by genital secretions. HHV 6 DNA integrates into chromosomal telomeres in several cell types, including germ cells, in about 1% of people. As a result, offspring may carry a copy of the viral genome in every nucleated cell (known as inherited chromosomally integrated or iciHHV-6). The full clinical implications of iciHIV-6 remain unclear; clinicians should be aware that they will always be HHV-6 NAT positive. Pregnant women with iciHHV-6 may be at higher risk for pre-eclampsia or spontaneous abortion.

Infection with HHV-6 frequently occurs during infancy as maternal antibody wanes. The peak age of acquisition is 9–21 months; by 24 months, seropositivity rates approach 80%. Older siblings appear to serve as a common source of transmission. In addition, congenital infection may occur, and ~1% of newborns are infected with HHV-6; transplacental infection with HHV-6 accounts for about 14%, with the others related to infection from inherited chromosomally integrated HHV-6. Congenital infection is generally asymptomatic, although subtle neurologic defects have been described. Most postnatally infected children develop symptoms (fever, fussiness, and diarrhea). A minority develop exanthem subitum (roseola infantum; **see Fig. A1-5**), a common illness characterized by fever with subsequent rash. In addition, ~10–20% of febrile seizures without rash during infancy are caused by HHV-6. After initial infection, HHV-6 persists in peripheral-blood mononuclear cells as well as in the central nervous system, salivary glands, and female genital tract.

In older age groups, HHV-6 has been associated with mononucleosis syndromes; in immunocompromised hosts, encephalitis, pneumonitis, syncytial giant-cell hepatitis, and disseminated disease are seen. In transplant recipients, HHV-6 infection may also be associated with graft dysfunction. Acute HHV-6-associated limbic encephalitis has been reported in hematopoietic stem cell transplant recipients and is characterized by memory loss, confusion, seizures, hyponatremia, and abnormal electroencephalographic and MRI results. High plasma loads of HHV-6 DNA in HSCT recipients are associated with allelic-mismatched donors, use of glucocorticoids, delayed monocyte and platelet engraftment, development of limbic encephalitis, and increased all-cause mortality rates. Mesial temporal lobe epilepsy has been associated with HHV-6 infections, and, like many other viruses, HHV-6 (as well as EBV, another herpes virus) has been implicated in the pathogenesis of multiple sclerosis, although further study is needed to distinguish between association and etiology.

HHV-7 was isolated in 1990 from T lymphocytes from the peripheral blood of a healthy 26-year-old man. The virus is frequently acquired during childhood, albeit at a later age than HHV-6. HHV-7 is commonly present in saliva, which is presumed to be the principal source of infection; breast milk and cervical secretions also may carry the virus. Viremia can be associated with either primary or reactivation infection. The most common clinical manifestations of childhood HHV-7 infections are fever and seizures. Some children present with respiratory or gastrointestinal signs and symptoms. An association has been made between HHV-7 and pityriasis rosea, but evidence is insufficient to indicate a causal relationship.

Clustering of HHV-6, HHV-7, and CMV infections in transplant recipients can make it difficult to sort out the roles of the various agents in individual clinical syndromes. HHV-6 and HHV-7 appear to be susceptible to ganciclovir and foscarnet, although definitive evidence of clinical response is lacking. Use of the narrower-spectrum, CMV-specific agent letermovir for CMV prevention after HSCT was shown not to result in higher rates of HHV-6.

HHV-8

Unique herpesvirus-like DNA sequences were reported during 1994 and 1995 in tissues derived from Kaposi's sarcoma (KS) and body cavity–based lymphoma occurring in people with HIV. The virus from which these sequences were derived is designated HHV-8 or Kaposi's sarcoma–associated herpesvirus (KSHV). HHV-8, which infects B lymphocytes, macrophages, and both endothelial and epithelial cells, appears to be causally related not only to KS and a subgroup of AIDS-related B cell body cavity–based lymphomas (primary effusion lymphomas) but also to multicentric Castleman disease, a lymphoproliferative disorder of B cells. The association of HHV-8 with several other diseases has been reported but not confirmed.

HHV-8 seropositivity occurs worldwide, with areas of high endemicity influencing rates of disease. Unlike other herpesvirus infections, HHV-8 infection is much more common in some geographic areas (e.g., central and southern Africa) than in others (North America, Asia, northern Europe). Pockets of increased HHV-8 infections have been observed among men who have sex with men (MSM) with HIV in the southern United States; risk factors for dual infections with HIV and HHV-8 in this population include oral/anal or oral/penile sex and methamphetamine use. In high-prevalence areas, infection occurs in childhood, and seropositivity is associated with families having numerous children who share eating and drinking utensils; HHV-8 may be transmitted in saliva. In low-prevalence areas, infections typically occur in adults, probably via sexual transmission. Concurrent epidemics of HIV-1 and HHV-8 infections among certain populations (e.g., MSM) in the late 1970s and early 1980s appear to have resulted in the frequent association of AIDS and KS. Transmission of HHV-8 may also be associated with organ transplantation, injection drug use, and blood transfusion; however, transmission via organ transplantation or blood transfusion in the United States appears to be quite rare.

Primary HHV-8 infection in immunocompetent children may manifest as fever and maculopapular rash. Among individuals with intact immunity, chronic asymptomatic infection is the rule, and neoplastic disorders generally develop only after subsequent immunocompromise. Immunocompromised persons with primary infection may present with fever, splenomegaly, lymphoid hyperplasia, pancytopenia, or rapid-onset KS. Quantitative analysis of HHV-8 DNA suggests a predominance of latently infected cells in KS lesions and frequent lytic replication in multicentric Castleman disease. The KS-associated herpesvirus inflammatory cytokine syndrome (KICS)—consisting of fever, lymphadenopathy, hepatosplenomegaly, cytopenias, and high levels of HHV-8, human and viral interleukin 6, and human interleukin 10—has been described in some HIV-infected patients and is associated with a high mortality rate.

Effective antiretroviral therapy for HIV-infected individuals has led to a marked reduction in rates of KS among persons dually infected with HHV-8 and HIV in resource-rich areas. HHV-8 itself is susceptible in vitro to ganciclovir, foscarnet, and cidofovir. A small, randomized, double-blind, placebo-controlled, crossover trial suggested that oral valganciclovir administered once daily reduced HHV-8 replication. However, clinical benefits of valganciclovir or other drugs in HHV-8 infection have not yet been demonstrated. Sirolimus inhibits the progression of dermal KS in kidney transplant recipients while providing effective immunosuppression. Rituximab alone or in combination with chemotherapy can lead to a survival of >90% at 5 years in HHV-8–associated multicentric Castleman disease.

FURTHER READING

Cytomegalovirus

Avery RK et al: Maribavir for refractory cytomegalovirus infections with or without resistance post-transplant: Results from a Phase 3 randomized clinical trial. Clin Infect Dis 75:690, 2022.

Chatzakis C et al: The effect of valacyclovir on secondary prevention of congenital cytomegalovirus infection, following primary maternal infection acquired periconceptionally or in the first trimester of pregnancy. An individual data meta-analysis. Am J Obstet Gynecol 230:109, 2024.

Das R et al: Safety, efficacy, and immunogenicity of a replication-deficient human cytomegalovirus vaccine, V160, in cytomegalovirus-seronegative women: a double-blind, randomized, placebo-controlled, phase 2b trial. Lancet Infect Dis 23:1383, 2023.

Drutman SB et al: Fatal cytomegalovirus infection in an adult with inherited NOS2 deficiency. N Engl J Med 382:437, 2020.

Fang M et al: High cytomegalovirus viral load is associated with 182-day all-cause mortality in hospitalized people with human immunodeficiency virus. Clin Infect Dis 76:1266, 2023.

Hughes BL et al: A trial of hyperimmune globulin to prevent congenital cytomegalovirus infection. N Engl J Med 385:436, 2021.

Kotton CN et al: The Third International Consensus Guidelines on the management of cytomegalovirus in solid-organ transplantation. Transplantation 102:900, 2018.

Limaye AP et al: Letermovir vs valganciclovir for prophylaxis of cytomegalovirus in high-risk kidney transplant recipients: A randomized clinical trial. JAMA 330:33, 2023.

Ssentongo P et al: Congenital cytomegalovirus infection burden and epidemiologic risk factors in countries with universal screening: A systematic review and meta-analysis. JAMA Netw Open 4:e2120736, 2021.

Human Herpesvirus (HHV) Types 6, 7, and 8

Gabrielli L et al: Inherited chromosomally integrated human herpesvirus 6: Laboratory and clinical features. Microrganisms 11:548, 2023.

Gaccioli F et al: Fetal inheritance of chromosomally integrated human herpesvirus 6 predisposes the mother to pre-eclampsia. Nat Microbiol 5:901, 2020.

Kampouri E et al: Human herpesvirus-6 reactivation and disease after allogeneic haematopoietic cell transplantation in the era of letermovir for cytomegalovirus prophylaxis. Clin Microbiol Infect 29:1450.e1, 2023.

Knights SM et al: High seroprevalence of Kaposi sarcoma–associated herpesvirus in men with HIV in the southern United States. Open Forum Infect Dis 10:ofad160, 2023.

Miura H et al: Inherited chromosomally integrated human herpesvirus 6 is a risk factor for spontaneous abortion. J Infect Dis 223:1717, 2021.

Pellett Madan R et al: Human herpesvirus 6, 7, and 8 in solid organ transplantation: Guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. Clin Transplant 33:e13518, 2019.

201

Monkeypox, Molluscum Contagiosum, and Other Poxvirus Infections

Faisal Syed Minhaj, Christina L. Hutson

POXVIRUSES

DEFINITION AND ETIOLOGY

Poxviruses (*Poxviridae*) are a family of double-stranded DNA viruses whose genomic structure is generally conserved across subfamilies, genera, and species. The central portion of the genome, which can range up to 200 kb, encodes the open reading frames (ORFs) required for replication or packaging of virions. The left and right ends of the genome encode genes with predicted functions in immune evasion, host interaction, or unknown roles. The complement of ORFs across different genera is largely responsible for differences in disease manifestations and/or virus host range. Four genera of poxviruses (*Orthopoxvirus*, *Parapoxvirus*, *Yatapoxvirus*, and *Molluscipoxvirus*) include species that can infect humans. Additionally, a currently unclassified poxvirus has been reported to cause human illness. Table 201-1 identifies these viruses, the majority of which are zoonotic, and lists some of their epidemiologic characteristics.

EPIDEMIOLOGY

Most poxviruses that infect humans are spread through direct contact; notable exceptions are some species of *Orthopoxvirus* (i.e., variola and monkeypox viruses [MPXV]). Variola virus (the virus that causes smallpox) is transmitted primarily by close contact and via respiratory secretions. In what seems to have been a rare circumstance near the end of global efforts to eradicate smallpox, it was reported that variola virus appeared to transmit via aerosol in a German hospital in Meschede. The degree to which potential aerosols had a role in smallpox transmission remains debated.

There are two geographically, genetically, and clinically distinct clades of MPXV. Clade I (formally Congo Basin clade) is endemic to Central Africa (Democratic Republic of the Congo, Republic of the Congo, Central African Republic, Cameroon, Gabon). Clade II (formally West African clade) is endemic to West Africa (Nigeria, Liberia, Sierra Leone, Cameroon, Cote d'Ivoire). Clade I MPXV causes a higher proportion of individuals with severe disease than Clade II, which is further subdivided into clade IIa, found endemically in West Africa, and clade IIb, which spread worldwide in 2022, with low levels of virus circulation continuing to date. In endemic areas, MPXV infections historically clustered around rural villages and within households with significant wildlife exposure. Spread to humans is primarily through direct contact (e.g., handling during food preparation) with infected animals leading to percutaneous or permucosal exposure. Human-to-human spread historically was primarily via close skin-to-skin contact;

TABLE 201-1 Poxviruses Causing Infection in Humans

GENUS, SPECIES	GEOGRAPHY	ZOONOTIC CHARACTERISTICS
Orthopoxvirus		
Variola (smallpox)	Eradicated, formerly worldwide	Solely a human pathogen
Monkeypox	Historically endemic to West and Central Africa, but spread worldwide in 2022	Squirrel species, Gambian rats, and dormice implicated as potential reservoir species; other species effective in transmitting disease to humans (pet North American prairie dogs, non-human primates); can be acquired during hunting/preparation of African wildlife for nutritional protein source
Cowpox	Europe	Rodents as reservoir; outbreaks associated with rodent pet trade; cats also effective transmitters of illness; previously, dairy cow teat lesions linked to human cutaneous lesions
Vaccinia and vaccinia-like viruses (e.g., buffalopox, Cantagalo, Araçatuba)	Europe, India, and South America	Rodents suspected as a potential reservoir; localized lesions on cattle or other ruminants (e.g., water buffalo for buffalopox) responsible for most human infections
Borealpox (formally known as Alaskapox)	United States (Alaska)	Northern red-backed voles, squirrels, and shrews suspected as potential reservoir species; potential transmission to humans suspected from contact of above species or pets (e.g., cats, dogs) who were infected
Akhmeta	Georgia (country)	Woodmice (*Apodemus* spp.); cows can be infected and possibly transmit to humans.
Molluscipoxvirus		
Molluscum contagiosum	Worldwide	Thought to be solely a human pathogen; closely related viruses described in other mammals
Parapoxvirus		
Orf	Worldwide	Handling of infected sheep and goats primarily responsible for transmission to humans
Pseudocowpox	Worldwide	Handling of infected dairy cattle
Bovine papular stomatitis	Worldwide	Handling of infected beef cattle
Deerpox	U.S. deer herds	Handling of infected deer
Sealpox	Seal/pinniped colonies worldwide	Handling of infected pinnipeds
Yatapoxvirus		
Tanapox	Africa	Possible nonhuman primate reservoir; potential arthropod mediator
Unclassified poxvirus		
NY-014[a]	United States (New York State)	Unknown

[a] Possibly an orthopoxvirus.

however, spread via respiratory secretions also was believed to occur. After decades without detection of MPXV infections, Nigeria experienced a rapid increase in mpox cases in 2017. As routine smallpox vaccination (which provides cross-protection for mpox) ended in 1980, it is unclear what role waning orthopoxvirus immunity played in its reemergence. These cases in Nigeria were unusual in that they occurred in urban centers in persons without known animal contact. In 2021, eight travel-associated mpox cases in persons who flew from Nigeria were identified in four countries (United States, United Kingdom, Israel, and Singapore), with limited subsequent human-to-human transmission. In 2022, clade IIb mpox spread globally, with transmission occurring primarily via close, intimate (e.g., sexual) contact and predominantly affecting gay, bisexual, and other men who have sex with men (MSM). MSM, particularly Black and Hispanic MSM in the United States, remain disproportionately affected by mpox to date. Additionally, up to 50% of those diagnosed with mpox in this outbreak are also living with HIV.

Other orthopoxviruses (apart from variola virus) (Table 201-1) are thought to spread only via direct contact or percutaneous/permucosal exposures to infected animals (or humans). The orthopoxvirus infections caused by cowpox and the vaccinia-like viruses are typically acquired initially through contact with an infected animal. Human-to-human transmission can also occur via contact with the lesion(s) of the infected human. In Europe, human cowpox infections have been associated with the pet rat trade, and vaccinia-like viruses (e.g., Belo Horizonte, Cantagalo, Araçatuba) are reported in handlers of dairy cattle in South America. Similarly, buffalopox has been reported in inhabitants of the Indian subcontinent exposed to infectious lesions on water buffalo. In the United States, vaccinia, the virus historically known as the substrate for smallpox vaccine, has caused infections in laboratory workers studying the virus. With all orthopoxvirus infections, the illness is considered infectious from symptom onset until all lesions have crusted, the crust(s) have separated from the skin, and a fresh layer of healthy skin has formed underneath.

The most common poxvirus encountered in practice is molluscum contagiosum virus (MCV), which is a molluscipoxvirus. MCV likely spreads through direct contact with and percutaneous exposure to another infected human. Like variola virus (an orthopoxvirus), MCV is considered to be a pathogen of humans only. Infections are commonly seen in pediatrics where transmission occurs through play activities. In adults, the disease can manifest similarly, but genital involvement also is noted as transmission often occurs through sexual exposure.

The epidemiology of tanapox (a yatapoxvirus) is poorly understood. Simian reservoirs and the potential for an arthropod vector are hypothesized. Rare cases of tanapox have only been seen in the United States from travelers returning from West or Central Africa. Human infections with parapoxviruses occur through direct contact with and percutaneous exposure to lesions developing at the site of contact with an infected animal.

PATHOGENESIS

The pathogenesis of orthopoxvirus infections is thought to involve systemic spread of disease from the site of virus inoculation to local lymph nodes, lymphoreticular tissue seeding, and finally the development of symptomatic (febrile) viremia and viral skin tropism. Disease severity is affected by the degree to which the innate immune and interferon responses control the initial stages of infection. During illness, patients may experience lymphadenopathy, fever, pain, and malaise. In immunocompromised persons more severe systemic manifestations are seen. Cases exemplifying this were seen throughout the global mpox outbreak in 2022. Individuals with intact immune systems develop lesions (often at the exposure site) about 1 week after exposure; these lesions progress through specified stages over the next 7–14 days, followed by scabbing and complete resolution by 14–21 days after rash onset. Lesion scabs contain viable virus, and it is only once all lesions scab over, the scabs separate from the skin, and newly formed intact skin forms that the infectious period ends. In contrast, persons with severe immunocompromised states (e.g., advanced HIV [i.e., CD4 T cell count <200 cells/mm^3], organ transplantation) can develop severe mpox. In these instances, the spread or growth of MPXV goes unchecked, and systemic spread of disease results in wide dissemination of the rash, with large confluent lesions persisting for months and additional organ involvement.

Other poxvirus infections—with the possible exception of yatapoxvirus infection, in which disease pathogenesis is poorly understood—likely involve only local growth of the virus at the site of inoculation or reinoculation. In some immunocompromised hosts, the lesions caused by *Parapoxvirus* infections can become quite large; such lesions are referred to as "giant orf."

CLINICAL MANIFESTATIONS

Systemic poxvirus infections (i.e., *Variola Virus*, *Monkeypox Virus* [*Orthopoxvirus* genome] and *Tanapox* [*Yatapoxvirus* genome])

Following exposure, both smallpox and mpox have a similar incubation period of up to 17 days. Classically, smallpox and mpox present as follows: the first clinical sign is fever, which is followed by rash onset days later. Other prodromal symptoms include malaise, sore throat, and headache. When lymphadenopathy was present during this stage and throughout illness, it was a key differentiating factor between smallpox (absent) and mpox (present) in endemic countries. The rash evolves through classic macular, papular, vesicular, and pustular phases (the last with central umbilication), with each stage lasting 2–3 days and lesions in the same anatomic location typically in the same stage of development. Diffuse, well-circumscribed, centrifugally distributed (i.e., lesions more prominent on the face and extremities, including palms and soles, than on the trunk) lesions are classic. Lesions are typically painful as they emerge and subsequently become pruritic during later stages. However, during the emergence of clade IIb mpox in 2022, several clinical manifestations differed from this classic presentation.

During the clade IIb outbreak, patients often presented with rash as their first sign or symptom of disease or concurrently or in the absence of prodromal symptoms; lesions were frequently localized, primarily in the anogenital region **(Fig. 201-1)**. The lesions themselves were smaller, fewer in number (often <10), and often in different stages at the same anatomic site. There does not appear to be an increase in disease severity in those without severe immunocompromise (i.e., HIV with CD4 T cell count >200 cells/mL). However, a minority of patients have developed severe manifestations of mpox, many of whom who are severely immunocompromised (e.g., advanced HIV, organ transplantation, comparable severe immunocompromise).

Severe Mpox Rare severe manifestations of mpox include ocular infection, neurologic complications, myopericarditis, and uncontrolled viral replication in severely immunocompromised patients. Mpox-related ocular disease can occur via autoinoculation (i.e., touching a lesion then the eye) or local spread from a nearby lesion. Symptoms include eye pain, redness, vision changes or loss, or periorbital swelling. Ocular disease can manifest as blepharitis, conjunctivitis, conjunctival lesions, keratitis, and vision loss. Neurologic complications include rare reports of encephalitis and myelitis. Patients may have severe headache, neck pain, altered mentation, or focal deficits. Myopericarditis has been reported in some patients with mpox, including complaints of shortness of breath or palpitations with elevations in cardiac biomarkers and electrocardiographic changes. Mucosal complications can affect alimentation, urination, or defecation due to painful or obstructive lesions. These can lead to strictures, edema, and severe lymphadenopathy. Complications from uncontrolled viral replication commonly occur in severely immunocompromised persons such as those with advanced HIV or organ transplantation. These patients often develop numerous large, coalescing, or necrotic lesions and can have other organ involvement including the gastrointestinal tract, liver, lungs, and brain, which can manifest as organ dysfunction. These patients may have disease lasting many months in which immune system optimization is crucial to recovery. Most deaths have occurred in the United States, predominantly among people with advanced HIV with CD4 <50 cells/mm^3.

Tanapox Patients infected with tanapox virus initially present with a very high fever, are often thought to have malaria (given the endemic location), and later develop 1–10 nodular lesions. These nodules are often in anatomic areas not typically covered by clothing. Lesions are seldom filled with fluid and more often contain necrotic tissue. The lesion size peaks around 2 weeks after initial formation, and lesions typically disappear spontaneously within 6 weeks.

Other Orthopoxvirus Infections Other orthopoxvirus infections are more localized in their presentation, with lesions likely developing directly at the site of contact with the virus. Akhmeta, Borealpox (formally known as Alaskapox), vaccinia, vaccinia-like, and cowpox virus infections are typically associated with a localized rash or lesion evolving through classical papular, vesicular, and pustular phases. In immunocompromised patients, presentation of these orthopoxvirus infections can be protracted or disseminated and, rarely, lead to death.

Other Poxvirus Infections Individuals infected with other poxviruses that cause localized disease (parapoxviruses and MCV) seldom report a febrile phase and instead experience slow and gradual development of a lesion or lesions. The lesion of molluscum contagiosum has a classic pearly appearance that sometimes umbilicates as it matures **(Fig. 201-2)**. There is little inflammation that surrounds the painless lesions, which persist for months but gradually regress after 6–12 months. Patients with immunocompromised status can have severe and prolonged disease; in patients with uncontrolled HIV and advanced HIV, immune reconstitution is usually sufficient to clear the virus. The rash lesions of parapoxvirus infections begin as erythematous papules, develop into a "target" lesion, and then become nodular and papilloma-like. "Giant" parapoxvirus infections have been reported in immunocompromised individuals.

DIFFERENTIAL DIAGNOSIS

A patient usually presents to the clinician with nodular or vesiculopustular lesions. Important elements of the history include travel, occupation (with risk varying dependent on the poxvirus; greater risk is in laboratory workers working with poxviruses, farmers, hunters, and sex workers), animal exposures, lesion evolution, sexual history, and symptom timing with respect to rash onset. Additionally, given the appearance and location, mpox lesions in the anogenital area may resemble common sexually transmitted infections (STIs) such as gonorrhea, chlamydia, or syphilis; up to 10% of patients with mpox may be coinfected with an STI. Other differential diagnoses in poxvirus infections include varicella, yaws, papillomavirus infection, herpes simplex virus, and (particularly in parapoxvirus infections) cutaneous anthrax. While the characteristic lesions of poxvirus infection coupled with an indicative exposure history are helpful in narrowing the differential, laboratory testing is needed to confirm the diagnosis.

Currently, the most common laboratory tool for diagnosis of poxvirus infection involves nucleic acid (i.e., molecular) testing. Nucleic acid–based diagnostics include polymerase chain reaction (PCR) and sequencing to fully characterize the isolate in some cases. This technology has led to the identification of a number of new poxviruses that can cause human infection, including Akhmeta, Borealpox, and NY-014. Orthopoxvirus molecular testing options drastically expanded during the global 2022 mpox outbreak that was caused by clade IIb MPXV. Prior to the outbreak, only select PCR assays (developed by the Centers for Disease Control and Prevention [CDC] for smallpox preparedness) were available within a subset of the Public Health Laboratories (PHL) within the Laboratory Response Network. During the 2022 outbreak, CDC and the U.S. Food and Drug Administration (FDA) collaborated to increase testing availability in commercial laboratories within the United States. PCR assays specific to non-variola orthopoxvirus (NVO), orthopoxvirus generic, generic mpox, and mpox clade II are now readily available for testing of most orthopoxviruses. Other countries experienced similar growth in orthopoxvirus testing options. Many laboratories have also introduced multiplex testing options that include both NVO- and MPXV-specific targets. Some laboratories (U.S. Government and PHL) offer mpox clade-specific testing. The orthopoxviruses also grow well in most standard clinical laboratory tissue cultures. The parapoxviruses are difficult to isolate via culture (primary cells are best), and MCV cannot be cultured. Electron microscopy identifies the characteristic large, brick-shaped virus particles on negative stain if orthopoxvirus, yatapoxvirus, or MCV is present. Parapoxviruses have an ovoid structure with crisscross spicules on negative-stain electron microscopy. MCV has a classic appearance, with Henderson-Patterson bodies, on pathologic analysis of a biopsy sample. Serologic assays can demonstrate orthopoxvirus reactivity, but most are unable to distinguish between orthopoxvirus species because of their broad antigenic similarity. Efforts in serology tests during the mpox outbreak are ongoing in order to differentiate infection from vaccination.

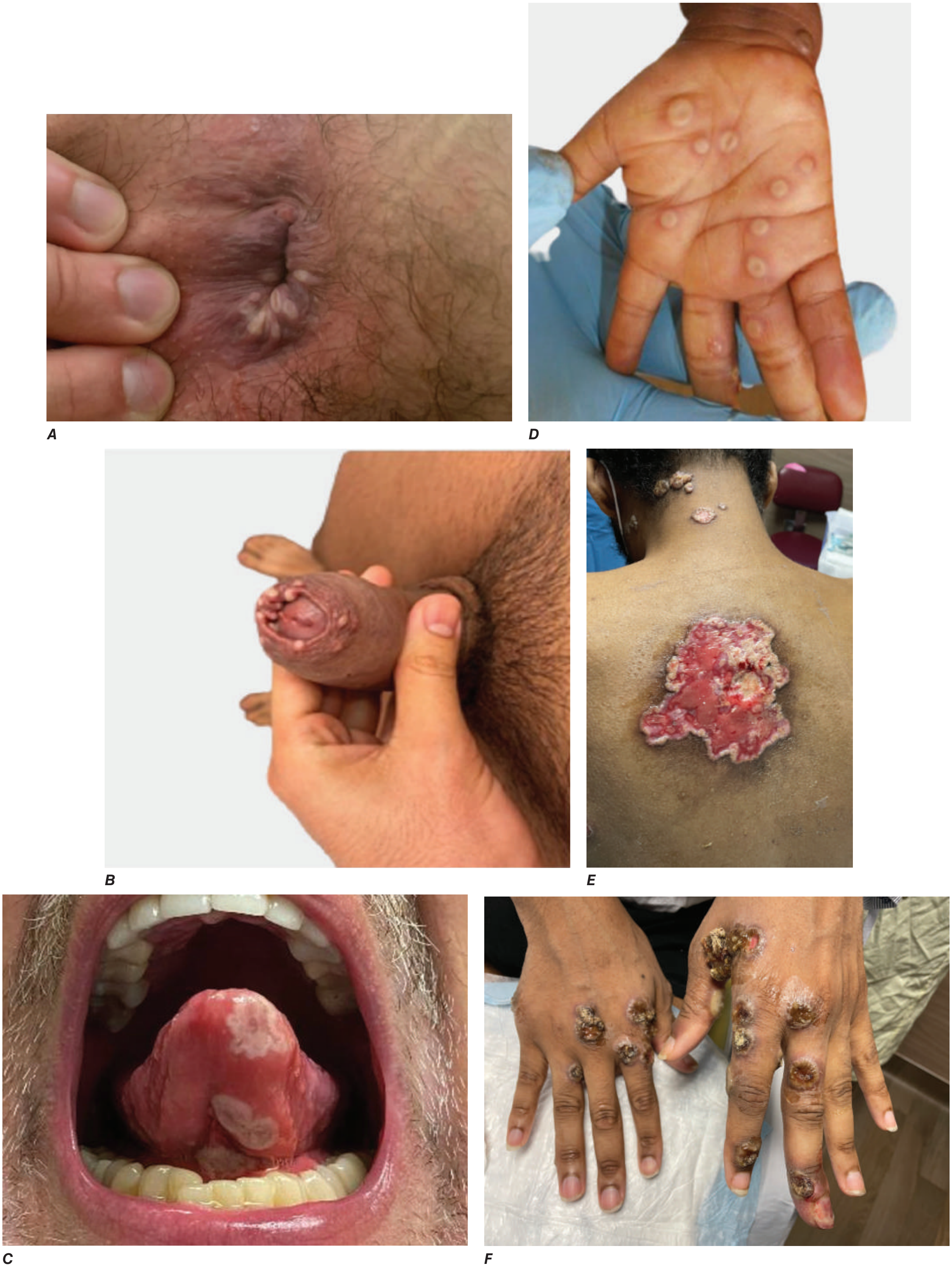

FIGURE 201-1 Mpox lesions. ***A–D.*** Standard lesions seen in mpox. Notice the well-circumscribed nature of the lesions even on mucosal surfaces; many of the lesions display the central umbilication unique to poxviruses, which often develop in later stages of illness. ***E and F.*** Severe mpox manifestations in a patient with advanced HIV. Notice the large coalescing lesion on the back, where the large lesion border is composed of individual lesions; additionally, large healing lesions are seen on the neck and hands. *(Source: CDC.)*

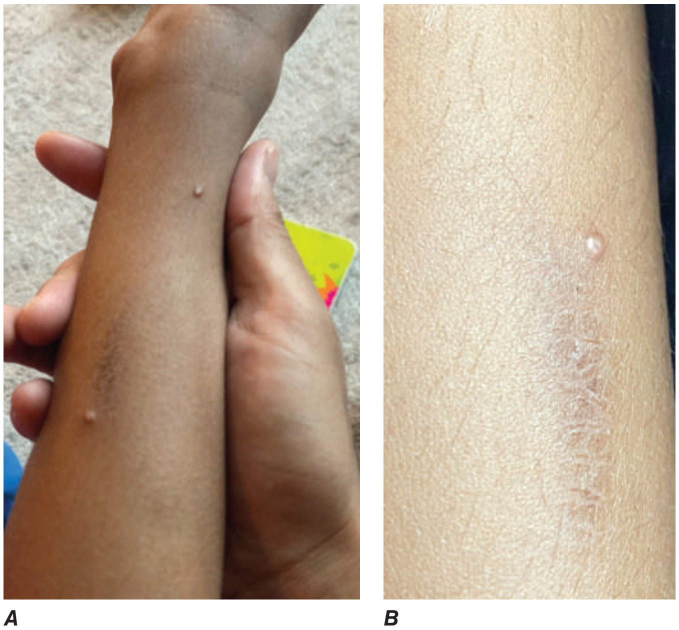

FIGURE 201-2 Molluscum contagiosum lesions. Notice the classic pearly appearance of molluscum contagiosum in ***A*** and ***B***. ***B*** also displays central umbilication. *(Source: CDC.)*

TREATMENT

Poxvirus

Treatment of poxvirus infection is largely supportive. Typical supportive care goals include prevention of autoinoculation to secondary sites and bacterial superinfection, pain control, and scar minimization. Disease-specific therapies are generally reserved for severe illness. Recently, as part of smallpox preparedness efforts, two antiviral agents active against the orthopoxviruses have been approved by FDA for the treatment of smallpox. As these agents did not require comparative human trials, their role in human orthopoxvirus infections continues to be investigated. Both antiviral agents are virustatic; therefore an immune response is crucial to recovery for any orthopoxvirus infection. For mpox specifically, given the high proportion of individuals coinfected with HIV, a key aspect of management is initiation, continuation, or reinitiation of antiretroviral therapy. The degree of HIV control is directly correlated with hospitalization, severe mpox manifestations, and mortality. As in other etiologies of immunosuppression, immune system optimization is critical for recovery. Immunosuppressive agents should be avoided or held if possible during illness.

Tecovirimat is an inhibitor of a viral egress protein that prevents cell-to-cell dissemination of mature virions; it is virustatic, so concomitant optimized immune function is essential to favorable clinical outcomes. It was used as an investigational drug in isolated cases of vaccinia, cowpox, and borealpox, and used extensively during the global mpox outbreak. It has a favorable safety profile and is the drug of choice for severe orthopoxviral infections; however, its effectiveness for treatment of human orthopoxvirus infections has not been systematically evaluated. Tecovirimat is dosed 600 mg orally or 200 mg intravenously every 12 h (with higher or more frequent dosing depending on patient weight) typically for 2 weeks. A fatty meal is necessary to optimize enteral bioavailability. For patients with severe immunocompromise and severe manifestations of mpox, the benefits of a prolonged treatment course (i.e., beyond the standard 14-day duration) may outweigh the harms (tecovirimat resistance). Importantly, a single-point amino acid change in the viral target of the drug can confer resistance to tecovirimat; therefore it should be reserved for severe orthopoxvirus infections given its primary purpose as therapy during a smallpox incident. Additionally, genotypic tecovirimat resistance does not consistently correlate with phenotypic resistance. To complicate tecovirimat resistance further, there have been different ranges of resistant virus observed, and it is unclear how these results should be interpreted for clinical treatment purposes. It is also unknown if resistance observed within one lesion specimen correlates with viral populations throughout the infected individual, and it has been observed that resistance develops differently (with different viral mutations detected) in swabs from different parts of the body. Therefore, it is reasonable to continue therapy even when resistance is suspected or detected in a specimen.

The other FDA-approved therapy for smallpox is brincidofovir, a prodrug of cidofovir. Either cidofovir or brincidofovir can be given along with tecovirimat for severe poxviral infections. Brincidofovir is only available orally and dosed at 200 mg weekly for two doses. Cidofovir can be used if intravenous therapy is required at 5 mg/kg weekly with concurrent probenecid. Limited animal data suggest that brincidofovir may be synergistic with tecovirimat in orthopoxvirus disease. Therefore, combination therapy can be considered in severe infections. Cidofovir and brincidofovir have higher barriers to resistance, therefore, it is less likely to occur. Topical cidofovir has been used in orthopoxvirus and MCV infections with mixed results.

Vaccinia immune globulin (VIG) is licensed for the treatment of adverse reactions to live, replicating smallpox (vaccinia virus) vaccine. The standard dose is 6000 U/kg intravenously; dosing can be repeated, and doses of up to 9000 U/kg can be used. Given antigenic similarities across the orthopoxvirus genus, VIG was used extensively for severe mpox with unclear efficacy. It is used primarily for people who may not be able to mount a sufficiently robust immune response to clear virus in patients with severe immunocompromise. Monoclonal antibodies similar to VIG are currently being studied for variola virus.

Trifluridine is active against ocular orthopoxviral infections and can be administered for treatment or ocular prophylaxis for periorbital lesions. Treatment dosing is instillation of one drop into the affected eye(s) every 2 h while awake for the first 2 weeks and then four times daily for an additional 2 weeks.

Treatment for MCV infection is on a case-by-case basis. Immunomodulatory therapies such as imiquimod have been used. If quicker resolution is desired, curettage and topical liquid nitrogen is available along with the FDA-approved agent cantharidin, which is applied by a clinician to lesions every 21 days for up to four doses.

■ PROGNOSIS

In immunocompetent hosts most poxvirus infections are self-limited, resolving in weeks or, in the case of molluscum contagiosum, months. The exceptions are the generalized orthopoxvirus infections caused by MPXV and variola virus, whose case–fatality rates are greater. In unvaccinated individuals smallpox carries a mortality of up to 30%, and MPXV clade I, IIa, and IIb mortality rates are 1.4–10%, 1%, and <1%, respectively. Immunocompromised hosts may have more severe orthopoxvirus and parapoxvirus infections (e.g., severe mpox, progressive vaccinia, eczema vaccinatum, severe borealpox) leading to higher mortality, or they may have atypical presentations (e.g., giant orf). In patients with advanced HIV, effective antiretroviral therapy is essential to favorable clinical outcomes. MCV infections can be diffuse in immunocompromised persons. Immune reconstitution inflammatory syndrome (IRIS) has been associated with recrudescence of MCV infections. Poxvirus reinfections are seldom reported.

■ PREVENTION

ACAM2000 is a live, replicating vaccinia virus vaccine administered as a single dose via a percutaneous needle. JYNNEOS is a live, nonreplicating, modified vaccinia Ankara-Bavarian Nordic (MVA-BN) vaccine administered via two subcutaneous doses 28 days apart. Awareness of occupational risks and adherence to appropriate barrier precautions effectively prevent most poxvirus infections. Pre-exposure vaccination is recommended for specific persons at risk of occupational exposure (e.g., research laboratorians) to orthopoxviruses via either ACAM2000 or JYNNEOS. Booster doses for these persons at sustained risk for occupational exposure to orthopoxviruses should be administered either 2 years, 3 years, or 10 years after primary vaccination depending on the vaccine administered and reason for vaccination. JYNNEOS is also recommended for persons at risk of mpox during mpox outbreaks and on the routine immunization schedule for persons with specific risk factors defined by CDC's Advisory Committee on Immunization Practices (ACIP). There is currently no recommendation for booster doses for these persons at risk or for persons to whom JYNNEOS was recommended during the mpox outbreak that started in 2022. Effectiveness of ACAM2000 is inferred from use of similar live, replicating vaccines during the smallpox eradication era when administration of a qualified vaccine 3–5 years earlier was viewed as 100% protective against variola virus. During mpox surveillance efforts in Democratic Republic of the Congo in the 1980s, smallpox vaccination 3–19 years earlier was 85% protective against disease among household contacts of people with mpox. The duration of efficacy is unclear. JYNNEOS was widely used during the global mpox outbreak, and effectiveness against mpox ranged from 36% to 75% for one-dose vaccination and 66% to 89% for two-dose vaccination. The duration of immunity following JYNNEOS vaccination compared with live, replicating vaccinia virus vaccines is unclear.

Acknowledgment

Inger K. Damon contributed to this chapter in the last edition and some material from that chapter has been retained here.

FURTHER READING

Chen X et al: Molluscum contagiosum virus infection. Lancet Infect Dis 13:877, 2013.

Rao AK et al: Interim clinical treatment considerations for severe manifestations of mpox — United States, February. MMWR Morb Mortal Wkly Rep 72:232, 2023.

Thornhill JP et al: Monkeypox virus infection in humans across 16 countries — April–June 2022. N Engl J Med 387:679, 2022.

202

Parvovirus Infections

Maria Söderlund-Venermo

Parvoviruses, members of the large family *Parvoviridae*, are small (diameter, ~22 nm), nonenveloped, icosahedral viruses with a linear single-stranded DNA genome of ~5000 nucleotides. The family includes viruses infecting many different animal hosts, from mammals to insects. Five main groups of parvoviruses infect humans: parvovirus B19 (B19V), adeno-associated viruses (AAVs), parvovirus 4 (parv4), human bocaviruses (HBoVs), and human protoparvoviruses (bufavirus and cutavirus).

PARVOVIRUS B19

DEFINITION

B19V belongs to the genus *Erythroparvovirus,* so named due to its narrow tropism of erythrocyte precursors in the bone marrow. B19V is divided into three genotypes (1, 2, and 3), with similar antigenic, pathogenic, and biological properties.

EPIDEMIOLOGY

B19V exclusively infects humans, and infection is common in virtually all parts of the world. Genotype 1 is currently predominant worldwide, whereas genotype 2 nowadays rarely causes active infections but remains persistent in tissues of older individuals. Genotype 3 is the most diverse and appears to be more common in the western parts of Africa.

Outbreaks of B19V infection, causing childhood rash (erythema infectiosum), are most common in schools and day-care centers and occur as epidemics a few years apart, in temperate climates, mostly in winter and spring. Within households, schools, and day-care centers, the infection rates approach 50%. The risk of infection increases in proportion to the number of children. Transmission occurs primarily via the respiratory route and occurs before the onset of rash or arthralgia. By the age of 15 years, ~50% of children have detectable IgG antibody to B19V; this seroprevalence may rise to 80% among the elderly. Especially in patients with a hemolytic disorder or compromised immune system, the viral load of B19V in blood can be extremely high (up to 10^{14} particles/mL), which increases the risk of transmission to hospital staff and family. Transmission can also occur via transfusion, particularly of pooled blood products. However, plasma pools are nowadays screened for B19V DNA, and high-titer pools are discarded. B19V is quite resistant to both heat and solvent-detergent inactivation.

PATHOGENESIS

B19V replicates in erythroid progenitors. This specificity may be due in part to a limited tissue distribution of the yet unknown primary B19V receptor that is recognized by the N-terminal unique region of B19-virus protein 1 (VP1u). Another important receptor, the blood group P antigen (globoside), recognized by the common VP region, VP2, is needed at a later intracellular step. Individuals who lack this P antigen are naturally resistant to B19V infection. Infection leads to high-titer viremia, with 10^{4-12} virus particles/mL detectable in the blood at the apex **(Fig. 202-1)**, and virus-induced cytotoxicity results in cessation of red cell production. The viral load will, however, quickly drop, leaving very low-level B19V DNA in the blood for months and even years after the acute infection. B19V DNAemia in nonacute infections has, however, also been shown to be due to nonencapsidated naked DNA being released from injured tissues. In immunocompetent individuals with normal hemopoiesis, the arrest of erythropoiesis is transient, with only a minimal drop in hemoglobin levels, which resolves as the immune response is mounted. However, in individuals with increased erythropoiesis (especially with hemolytic anemia), the cessation of red cell production can induce a transient crisis with severe anemia (Fig. 202-1). Similarly, if an individual (or a fetus) does not induce neutralizing antibodies to halt the lytic infection, erythroid production is compromised, and chronic anemia develops (Fig. 202-1).

In immunocompetent individuals, the immune-mediated phase of illness, which begins 2–3 weeks after acute infection as the IgM response peaks, manifests as the rash of erythema infectiosum or fifth disease alone or together with arthralgia and/or frank arthritis (see "Clinical Manifestations"). If immunocompromised patients with chronic B19V-induced anemia are given immunoglobulins, they may also present with a rash, which is due to antigen-antibody complexes in skin.

Even if B19V requires erythroid precursor cells for its replication, it can also enter nonpermissive cells, such as B cells, monocytes, and endothelial cells, by antibody-dependent enhancement (ADE), and remain presumably dormant for life in multiple tissues, such as the heart, liver, kidneys, synovia, brain, and even bones. This persistent presence of B19V DNA in our tissues does not generally seem to complicate normal health but may nevertheless be responsible for some disease presentations in predisposed individuals, as has been suggested in myocarditis, for example.

CLINICAL MANIFESTATIONS

Erythema Infectiosum Most B19V infections are asymptomatic or exhibit only a mild nonspecific illness. The main manifestation of symptomatic B19V infection is erythema infectiosum, also known as *fifth disease* or *slapped-cheek disease* **(Figs. 202-2 and A1-1*A*)**. Infection may begin with a minor febrile prodrome ~7–10 days after exposure, but it is often absent, and the classic facial rash develops

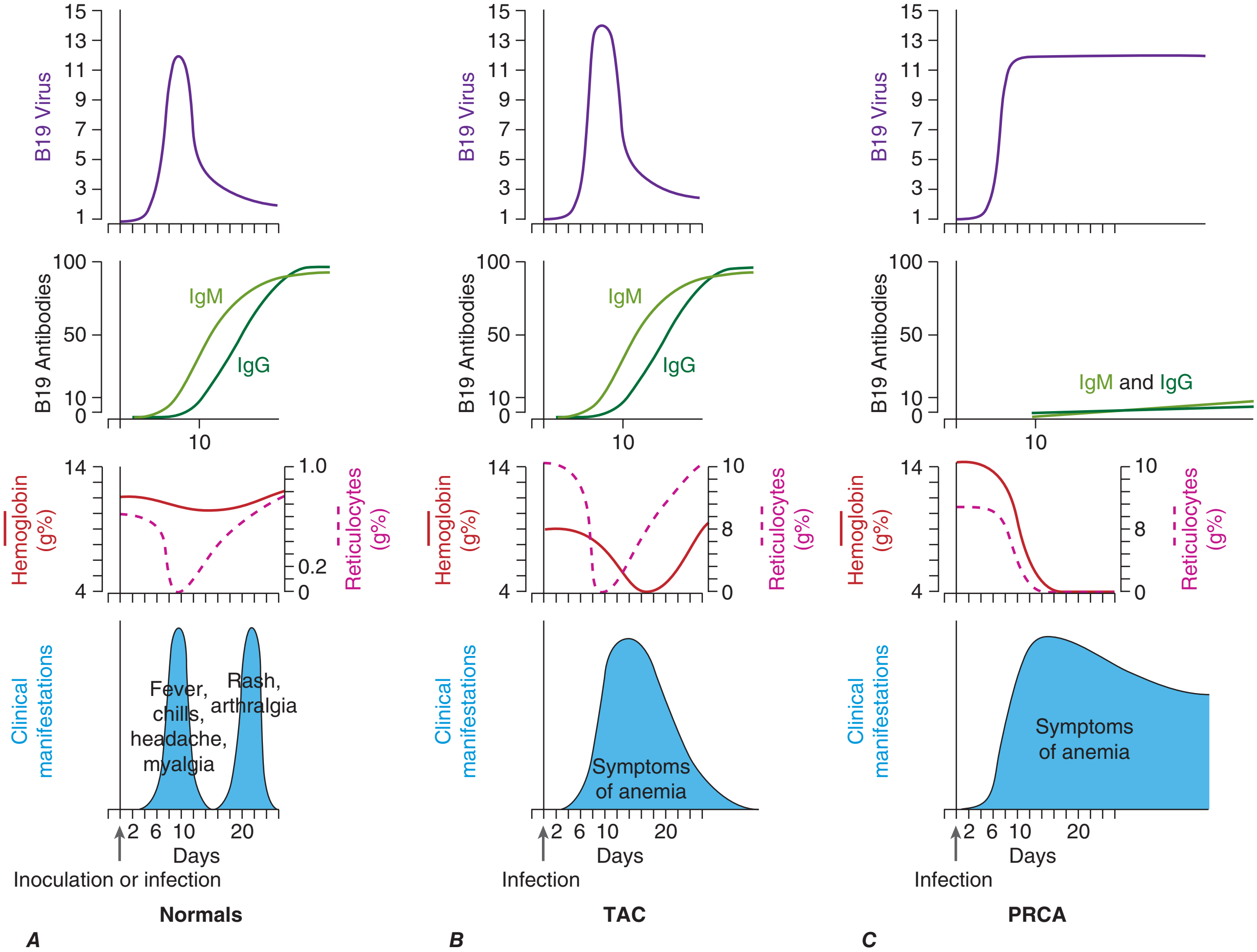

FIGURE 202-1 Schematic of the time course of parvovirus B19 infection in (***A***) normals (erythema infectiosum), (***B***) transient aplastic crisis (TAC), and (***C***) chronic anemia/pure red cell aplasia (PRCA). *(From The New England Journal of Medicine, Parvovirus B19, NS Young, KE Brown: 350:586. Copyright @2004 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

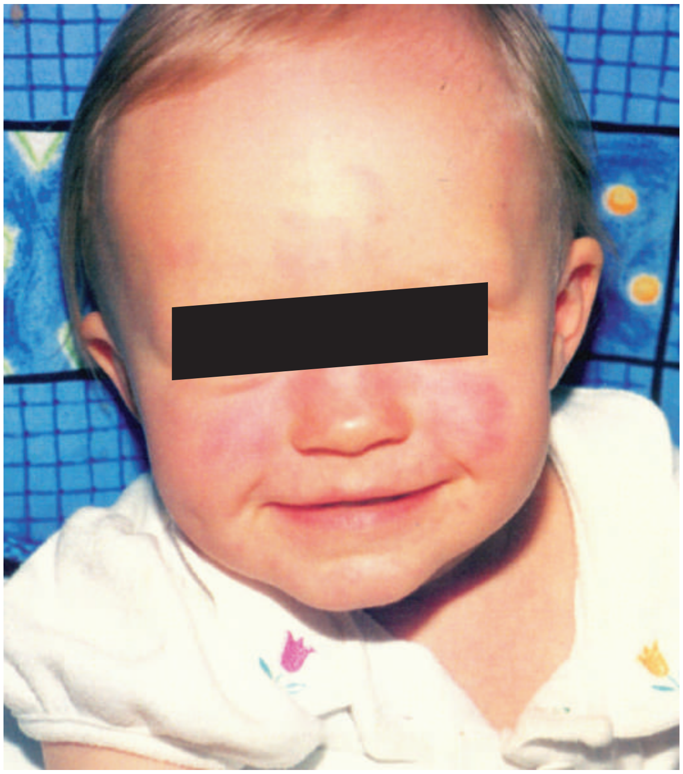
[redacted]

FIGURE 202-2 Young child with erythema infectiosum, or fifth disease, showing typical "slapped-cheek" appearance.

suddenly several days later. After 2–3 days, the erythematous maculopapular rash may spread to the trunk and extremities in a lacy reticular pattern. However, its pattern, intensity, and distribution vary, and B19V-induced rash is difficult to clinically distinguish from other viral exanthems, so a laboratory test should be used when a definite diagnosis is necessary, such as in pregnant women. Typically, the rash may recur for weeks when exercising or sunbathing, but the child is no longer infectious and can go to school. Adults typically do not exhibit the "slapped-cheek" appearance but present with arthralgia, with or without a macular rash. In children, arthritis and encephalitis are rare complications.

Polyarthropathy Syndrome Although uncommon among children, arthropathy occurs in ~50% of adults and is more common among women than among men. The distribution of the affected joints is often symmetrical, with arthralgia affecting the small joints of the hands and occasionally the ankles, knees, and wrists. Resolution usually occurs within a few weeks, but recurring symptoms can continue for months. The illness may mimic rheumatoid arthritis, and rheumatoid factor can often be detected in serum. However, the mere presence of B19V DNA in synovia is not enough to prove a causative relation, since healthy individuals also may exhibit viral DNA in their synovia.

Transient Aplastic Crisis Asymptomatic transient reticulocytopenia occurs in most individuals with B19V infection. However, in patients who depend on continual rapid production of red cells, infection can cause a transient aplastic crisis (TAC). B19V is the predominant cause of TAC in individuals with hemolytic disorders, hemoglobinopathies, red cell enzymopathies, and autoimmune hemolytic anemias. Patients present with severe to life-threatening anemia and a low reticulocyte count, and bone marrow examination reveals characteristic giant pronormoblasts and an absence of erythroid precursors. However, reticulocytopenia in sickle-cell patients with acute worsening of anemia is diagnostic without bone marrow examination. Patients are often febrile and very ill, often including other complications. As its name indicates, the illness is transient, and anemia resolves with the cessation of cytopathic infection in the erythroid progenitors, and lifelong immunity follows.

Pure Red Cell Aplasia (PRCA)/Chronic Anemia Chronic B19V infection has been reported in a wide range of immunocompromised patients who are unable to mount a neutralizing immune response, including those with certain congenital immunodeficiencies, AIDS **(Chap. 208)**, lymphoproliferative disorders (especially acute lymphocytic leukemia), and transplantation **(Chap. 148)**. PRCA patients have persistent anemia with reticulocytopenia, absent or low levels of B19V IgG, extremely high titers of B19V DNA in serum, and typically scattered giant pronormoblasts on bone marrow examination. Nonerythroid hematologic lineages are rarely affected, but transient neutropenia, lymphopenia, and thrombocytopenia (including idiopathic thrombocytopenic purpura) have been observed. B19V occasionally causes a hemophagocytic syndrome. The suspicion of B19V infection in such cases is often difficult due to the lack of normal B19V-related symptoms like rash or arthralgia, which are immune mediated. Diagnosis is, however, important due to the existence of effective therapy in form of repeated immunoglobulin administrations. Co-infection with *Plasmodium* and B19V has been suggested to play a role in the development of severe anemia in young children with malaria. B19V-infected immunocompetent individuals seldom show PRCA or chronic anemia.

Hydrops Fetalis B19V infection during pregnancy can lead to hydrops fetalis and/or fetal loss, due to either miscarriage (before 22 weeks of gestation) or fetal death (after 22 weeks of gestation). B19V probably causes 10–20% of all cases of nonimmune hydrops, which is characterized by gross edema and severe anemia. The risk of transplacental fetal infection is ~30%, and the excess risk of fetal loss (when the mother is infected before gestational week 20) is ~9%, but very low thereafter. Although B19V does not appear to be teratogenic, rare cases of eye damage, central nervous system (CNS) abnormalities, and congenital anemia have been reported. B19V infection may not cause any symptoms in the pregnant mother, so exposed seronegative mothers should undergo tests for B19V infection, and if found positive, they should be monitored regularly throughout pregnancy. Most fetal infections resolve themselves, but sometimes intrauterine red cell transfusions are needed.

Unusual Manifestations B19V infection may rarely cause hepatitis, vasculitis, myocarditis, glomerulosclerosis, or meningoencephalitis. A variety of other cardiac manifestations, CNS diseases, and autoimmune diseases have also been reported in conjunction with B19V infection. However, B19V DNA can be detected by polymerase chain reaction (PCR) for life in many tissues; therefore, this finding is of no known clinical significance, but its interpretation may cause confusion regarding B19V disease association.

DIAGNOSIS

Diagnosis of B19V infection in immunocompetent individuals is generally based on detection of B19V antibodies **(Table 202-1)**. IgM can be detected by indirect enzyme immunoassay (EIA) at the time of the rash in erythema infectiosum and by the third day of TAC in patients with hematologic disorders, and may remain detectable for ~3 months or longer. B19V IgG is detectable by the seventh day of illness and persists throughout life, whereby IgG positivity marks immunity. However, serum samples taken 2 weeks apart that show seroconversion or a fourfold or greater increase in IgG titer are considered diagnostic for acute infection. Modern serology can further measure the *quality* of IgG; as the immune response matures with time, the initially low avidity of IgG gradually increases within 6 months and can be measured with a denaturing EIA. Another way of timing the B19V infection is by comparing the IgG responses toward linear versus conformational B19V VP2 epitopes using epitope-type-specific (ETS) EIA. Both avidity and ETS EIAs differentiate between acute and past infection and thus increase the specificity of the diagnosis. Detection of B19V DNA in serum (or amniotic fluid) by PCR provides further help, especially in pregnancy, TAC, or chronic anemia. In acute infection at the height of viremia, $>10^{12}$ B19V DNA IU/mL of serum can be detected; nevertheless, the viral load falls rapidly within a few days but can remain detectable by PCR for months or even years after acute infection, even in healthy individuals, necessitating a quantitative (q)PCR. Of note, in tissue material, PCR alone should not be used to establish a B19V etiology because viral DNA remains in healthy bodies for decades.

TREATMENT

Parvovirus B19 Infection

No antiviral drugs against B19V are available for patient use, and treatment of B19V infection often targets symptoms only. However, cidofovir, and its lipid conjugate brincidofovir, as well as hydroxyurea, seem to inhibit B19V replication in vitro. TAC caused by B19V infection frequently necessitates treatment with repeated blood transfusions. In patients receiving chemotherapy, temporary cessation of treatment may result in an immune response and resolution. If this approach is unsuccessful or not applicable, commercial immunoglobulin can cure or ameliorate chronic B19V infection in immunosuppressed or otherwise immunocompromised patients. Generally, the intravenous IgG (IVIG) dose is 400 mg/kg daily for 5–10 days and the patient should be monitored for relapses. Administration of IVIG is not beneficial for the immune-mediated erythema infectiosum or B19V-associated arthropathies, which generally are self-limited. Intrauterine blood transfusion can prevent fetal loss in some cases of fetal hydrops; however, the risks need to be evaluated.

PREVENTION

No vaccine has been approved for the prevention of B19V infection, although vaccines based on B19V virus–like particles expressed in insect cells are known to be highly immunogenic. Phase 1 trials of a putative vaccine were discontinued because of adverse side effects, but others are under development.

TABLE 202-1 Diseases Associated with Human Parvovirus B19 Infection and Methods of Diagnosis

DISEASE	HOSTS	IgM	IgG	PCR	QUANTITATIVE PCR
Fifth disease	Healthy children	Positive	Positive	Positive	$>10^4$ IU/mL
Polyarthropathy syndrome	Healthy adults (more often women)	Positive within 3 months of onset	Positive	Positive	$>10^4$ IU/mL
Transient aplastic crisis	Patients with increased erythropoiesis	Negative/positive	Negative/positive	Positive	Often $>10^{12}$ IU/mL, but rapidly decreases
Persistent anemia/pure red cell aplasia	Immunodeficient or immunosuppressed patients	Negative/weakly positive	Negative/weakly positive	Positive	Often $>10^{12}$ IU/mL, but should be $>10^6$ in the absence of treatment
Hydrops fetalis/ congenital anemia	Fetuses (of mothers infected <20 weeks)	Negative/positive	Positive	Positive amniotic fluid or tissue	n/a

Abbreviations: IU, international units (1 IU equals ~1 genome); n/a, not applicable; PCR, polymerase chain reaction.

HUMAN BOCAPARVOVIRUSES

DEFINITION

Human bocavirus 1 (HBoV1) of the genus *Bocaparvovirus*, discovered in 2005, causes both upper and lower respiratory tract infections in young children and affects almost all children before the age of 5 years. Three related bocaviruses, HBoV2, HBoV3, and HBoV4, are mainly found in fecal samples and are thus considered enteric, with disputed roles in gastroenteritis.

EPIDEMIOLOGY

HBoVs are very common worldwide; the seroprevalences in Europe, already at 6 years of age, are 80% for HBoV1, 50% for HBoV2, 10% for HBoV3, and 0–1% for HBoV4, the latter being slightly more prevalent in China.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

HBoVs cause systemic infections with short viremia and induction of antibodies, which can be cross-reacting between the four HBoVs. HBoV1 causes mild to severe, rarely even life-threatening, respiratory tract infections (RTIs) in 0.5- to 5-year-old children (mean 2 years); it is the second to fourth most common finding (2–20%) in nasopharyngeal secretions of pediatric RTI patients. HBoV1 DNA, in decreasing levels, often remains detectable by sensitive PCRs in the airways for weeks or months after acute RTI, which is problematic, since RTI is commonly diagnosed by qualitative PCRs in respiratory secretions. This may lead to clinically incorrect interpretations of etiology, co-detections with other respiratory pathogens, and doubts of the role of HBoV1 in pathogenesis. However, clinical disease due to HBoV1 is associated with evidence of acute primary infection based on other diagnostic methods: sole infection, serology (IgG seroconversion or fourfold or greater increase in titer, IgM positivity, and/or low IgG avidity), viremia, and presence in respiratory secretions of spliced viral mRNA, high-load HBoV1 DNA ($>10^4$ genome copies/mL), or antigens. A combination of these methods should be applied in cases of severe or uncommon disease manifestations or for epidemiologic studies. HBoV1 causes both upper and lower RTIs, such as common cold, bronchiolitis, pneumonia, and exacerbations of asthma, with symptoms similar to those of other viral RTIs, with fever, cough, and wheezing commonly reported. In addition, diarrhea and acute otitis media often complicate HBoV1 infection. The role of the enteric HBoVs in childhood gastroenteritis remains to be established, but they have, like HBoV1, also been shown to cause encephalitis. There are no specific treatments or vaccines for HBoV infections.

OTHER HUMAN PARVOVIRUSES

BUFA- AND CUTAVIRUSES

Bufavirus and cutavirus of the genus *Protoparvovirus* were recently discovered in diarrheal stools by metagenomics. The global seroprevalence of bufavirus varies extensively from 3 to 85%, whereas that of cutavirus remains <6%. Although bufavirus DNA is found in 0.2–4% of diarrhea stools, the viral loads are low, and its pathogenic role is disputed. Interestingly, skin-persistent cutavirus has been shown to be associated with cutaneous T-cell lymphoma and its precursor, parapsoriasis en plaques, the latter with up to 67% genoprevalence in skin biopsies. Whether this association is causal or consequential remains to be confirmed.

PARVOVIRUS 4

Parvovirus 4 (parv4), in the genus *Tetraparvovirus,* was initially discovered in a patient with an acute viral syndrome and has since been found in pooled plasma donations and blood or tissues of injection drug users and hemophiliacs, suggesting parenteral spread. Even though parv4 in general is considered apathogenic, occasional studies have detected its DNA in patients with rash and respiratory, gastrointestinal, or central nervous system infections.

ADENO-ASSOCIATED VIRUSES

Adeno-associated viruses (AAV, with many serotypes), of the genus *Dependoparvovirus*, depend on helper functions from other viruses for their replication. Because of their ability to remain latent and their apathogenic nature, they have been developed as successful vectors in gene therapy. Nevertheless, in 2022, AAV2 was surprisingly associated with acute childhood hepatitis, but a causative role was not confirmed.

Acknowledgment

I thank the previous author, Kevin E. Brown, who wrote the prior edition's chapter. Some material from that chapter has been retained here.

FURTHER READING

Christensen A et al: Human bocaviruses and paediatric infections. Lancet Child Adolesc Health 3:418, 2019.

Crabol Y et al: Intravenous immunoglobulin therapy for pure red cell aplasia related to human parvovirus B19 infection: A retrospective study of 10 patients and review of the literature. Clin Infect Dis 56:968, 2013.

Maple PA et al: Identification of past and recent parvovirus B19 infection in immunocompetent individuals by quantitative PCR and enzyme immunoassays: A dual-laboratory study. J Clin Microbiol 52:947, 2014.

Matthews PC et al: Human parvovirus 4 'PARV4' remains elusive despite a decade of study F1000 Res 6:82, 2017.

Phan T et al: Cutavirus: A newly discovered parvovirus on the rise. Infect Gen Evol 80:104175, 2020.

Qiu J et al: Human parvoviruses. Clin Microbiol Rev 30:43, 2017.

Söderlund-Venermo M: Emerging human parvoviruses: The rocky road to fame. Ann Rev Virol 6:71, 2019.

Xiuong Y et al: The risk of maternal parvovirus B19 infection during pregnancy on fetal loss and fetal hydrops: A systematic review and meta-analysis. J Clin Virol 114:12, 2019.

203

Human Papillomavirus Infections

Darron R. Brown, Aaron C. Ermel

Interest in human papillomavirus (HPV) infection began in earnest in the 1980s after Harold zur Hausen postulated that infection with these viruses was associated with cervical cancer. It is now recognized that HPV infection of the human genital tract is extremely common and causes clinical conditions ranging from asymptomatic infection to genital warts (condylomata acuminata); dysplastic lesions and invasive cancers of the anus, penis, vulva, vagina, and cervix; and a subset of oropharyngeal cancers. This chapter describes the epidemiology of HPV as a virus and a pathogen, the natural history of HPV infections and associated cancers, strategies to prevent infection and HPV-associated disease, and treatment modalities for some conditions caused by HPV.

PATHOGENESIS

Overview HPV is an icosahedral, nonenveloped, 8000-base-pair, double-stranded DNA virus with a diameter of 55 nm. Like the genomes of other papillomaviruses, HPV's genome consists of an early (E) gene region, a late (L) gene region, and a noncoding region, which contains regulatory elements. The E1, E2, E5, E6, and E7 proteins are expressed early in the growth cycle and are necessary for viral replication and cellular transformation. The E6 and E7 proteins are responsible for malignant transformation, targeting the human cell-cycle regulatory molecules p53 and Rb (retinoblastoma protein) for degradation, respectively. Translation of the L1 and L2 transcripts and splicing of an E1^E4 transcript occur later. The L1 gene encodes the

54-kDa major capsid protein that makes up the majority of the virus shell; the 77-kDa L2 minor protein contributes a smaller percentage of the capsid mass.

More than 200 HPV types have been identified and are numerically designated on the basis of a unique L1 gene sequence. Approximately 40 HPV types are regularly identified in the anogenital tract; these types are subdivided into high-risk and low-risk categories depending on the associated risk of cervical cancer. For example, HPV types 6 and 11 cause genital warts and ~10% of low-grade cervical lesions and are thus designated low risk. HPV types 16 and 18 cause dysplastic lesions and a high percentage of invasive cancers of the cervix and are therefore considered high risk.

HPV is a tissue-tropic virus and targets basal keratinocytes of specific anatomic tissues after microtrauma allows exposure of these cells to the virus. HPV 1, for example, causes plantar warts but does not infect genital epithelium. The HPV replication cycle is completed as keratinocytes undergo differentiation. Virions are assembled in the nuclei of differentiated keratinocytes and can be visualized by electron microscopy. Infection is transmitted by contact with virus contained in these desquamated keratinocytes (or with free virus) from an infected individual.

The Immune Response to HPV Infection Unlike many viral infections, HPV infection has no viremic phase. This lack of viremia may account for the incomplete antibody response to HPV infection. Natural HPV infection of the genital tract gives rise to a detectable serum antibody response in 60–70% of individuals. Significant, although incomplete, protection against type-specific reinfection is associated with the presence of neutralizing antibodies. Serum antibodies likely reach the cervical epithelium and secretions by transudation and exudation. Therefore, protection against infection relates to the amount of neutralizing antibody at the site of infection and lasts as long as sufficient levels of neutralizing antibodies are present.

A cell-mediated immune response plays an important role in controlling progression of HPV infection. Histologic examination of lesions in individuals who experience regression of genital warts demonstrates infiltration by T cells and macrophages. CD4+ T cell regulation is particularly important in controlling HPV infections, as evidenced by the higher rates of infection and disease in immunosuppressed individuals, particularly those who are infected with HIV. Specific T cell responses may be measured against HPV proteins, the most important of which appear to be the E2 and E6 proteins. In women with HPV16 cervical infection, a strong T cell response to HPV16–derived E2 protein is associated with a lack of progression of cervical disease. However, measurable changes occur in the innate and adaptive immune systems of patients with HPV-associated cancers. There is suppression of the antigen-presentation process as well as suppression of antitumor activity. The end result is a reduction of HPV-specific antitumor immune responses and an increase in immunosuppressive cellular responses.

THE NATURAL HISTORY OF HPV-ASSOCIATED MALIGNANCY

HPV is transmitted by vaginal or anal intercourse, by oral sex, and probably by touching a partner's genitalia. In cross-sectional and longitudinal studies, ~50% of young women demonstrate evidence of HPV infection, with peaks during the teens and early twenties, within a few months after first coital experience. The number of lifetime sexual partners correlates with the likelihood of HPV infection and the subsequent risk of HPV-associated malignancy. HPV infection may occur in a monogamous person if that person's partner is infected.

Most HPV infections become undetectable after 6 to 9 months, a phenomenon known as "clearance." However, with prolonged follow-up and frequent sampling, the same HPV types may again be detected months or even years later. It is still debated whether such episodic detection indicates viral latency followed by reactivation or represents reinfection with an identical HPV type. Most evidence indicates that reactivation of latent virus is the cause of episodic detection of a specific HPV type.

While HPV is the causative agent of several cancers, most attention has focused on cervical cancer, which is the second most common cancer in women worldwide. More than 600,000 women are diagnosed and 300,000 die from invasive cervical cancer annually. More than 85% of all cervical cancer cases, as well as deaths, occur in women living in low-income countries, especially countries in sub-Saharan Africa, Asia, and South and Central America.

Evidence collected over 25 years shows that HPV causes nearly 100% of cervical cancers. Persistent HPV infection is the most significant risk factor for cervical cancer; relative risks range from 10 to 20 and exceed 100 in case–control studies and prospective studies, respectively. The time from HPV infection to cervical cancer may exceed 20 years. Cervical cancer peaks in the fifth and sixth decades of life for women living in developed countries and a decade or more earlier for women living in resource-poor countries. Persistent carriers of oncogenic HPV types are at greatest risk for high-grade cervical dysplasia and cancer.

Why HPV infections in some women but not others eventually lead to malignancy is not clear. Although oncogenic HPV infection is necessary for the development of cervical malignancy, only ~3–5% of infected women will ever develop this cancer, even in the absence of cytologic screening. Biomarkers that can predict which women will develop cervical cancer are not available, or incompletely characterized. Immunosuppression in general plays a significant role in redetection/reactivation of HPV infections, while other factors, such as smoking, hormonal changes, chlamydial infection, and nutritional deficits, have an impact on viral persistence and cancer.

The International Agency for Research on Cancer (IARC) has concluded that HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59 are carcinogenic in the uterine cervix. HPV type 16 is particularly virulent and causes at least 50% of cervical cancers. Worldwide, HPV types 16 and 18 cause at least 70% of cervical squamous cell carcinomas and 85% of cervical adenocarcinomas. Oncogenic types other than 16 or 18 cause the remaining 30% of cervical cancers. HPV types 16 and 18 also cause nearly 90% of anal cancers worldwide.

In addition to cervical and anal cancer, other HPV-associated cancers include vulvar and vaginal cancer (caused by HPV in 50–70% of cases), penile cancer (caused by HPV in 50% of cases), and at least 65% of oropharyngeal squamous cell carcinomas (OPSCCs). Over the past two decades, an epidemic of OPSCC related to oncogenic HPV infection, primarily HPV type 16, has developed. Rates of OPSCC in the United States have been increasing in men from a low of 0.27 case per 100,000 in 1973 to 0.57 case per 100,000 per year in 2004; rates in women have remained relatively stable at ~0.17 per 100,000 per year. The greatest increase in the incidence of OPSCC is among white men 40–50 years of age. Nearly 14,000 new cases were diagnosed in the United States in 2013. OPSCCs of the base of the tongue and tonsil cancer have increased annually by rates of 1.3 and 0.6%, respectively. Few data are available from developing countries about OPSCC.

THE EFFECTS OF HIV ON HPV-ASSOCIATED DISEASE

HIV infection accelerates the natural history of HPV infections. HIV-infected individuals are more likely than other individuals to develop genital warts, and their lesions are more recalcitrant to treatment. HIV infection has been consistently associated with precancerous cervical lesions, including low-grade cervical intraepithelial lesions (CIN) and CIN 3, the immediate precursor to cervical cancer. Women with HIV/AIDS have significantly higher rates of cervical cancer as well as subsets of some vulvar, vaginal, and oropharyngeal tumors **(Chap. 75)** than women in the general population. Studies indicate a direct relationship between low CD4+ T lymphocyte count and the risk of cervical cancer. Some studies show a reduced likelihood of HPV infection and precancerous lesions of the cervix in HIV-infected women given antiretroviral therapy (ART). However, the incidence of cervical cancer in HIV-infected women has not changed significantly since ART was introduced, possibly because of preexisting oncogenic HPV infections that occurred before ART was initiated.

The burden of HPV-associated cancers is expected to increase in HIV-infected patients, given the prolonged life expectancies provided with ART. For women living in developing countries where cervical cancer screening is not widely available, this trend will have significant consequences. Thus, elucidating the interactions of HIV infection and cervical cancer with cofactors such as diet, other sexually transmitted infections, and environmental exposures is an important focus of research that impacts women living in low- and middle-income countries.

Similar to that of cervical cancer, the incidence of anal cancer is strongly influenced by HIV infection. HIV-infected men who have sex with men (MSM) and HIV-infected women have much higher rates of anal cancer than HIV-uninfected populations. Specifically, the incidence among HIV-infected MSM has been found to be as high as 130 cases per 100,000 as opposed to 5 cases per 100,000 among HIV-negative MSM. The advent of ART has not impacted the incidence of anal cancer and high-grade anal intraepithelial neoplasia in the HIV-infected patient population.

More information regarding screening, prevention, and treatment in the HIV-infected population can be found at the Department of Health and Human Services website (*https://clinicalinfo.hiv.gov/en/guidelines*).

CLINICAL MANIFESTATIONS OF HPV INFECTION

HPV infects the male urethra, penis, and scrotum and the female vulva, vagina, and cervix. Perianal, anal, and oropharyngeal infections occur in both genders. Genital warts are caused primarily by HPV type 6 or 11 and appear as soft sessile growths with a surface that is either smooth or rough with multiple finger-like projections. Penile genital warts are usually 2–5 mm in diameter and often occur in groups. A second type of penile lesion, the keratotic plaque, is slightly raised above normal epithelium and has a rough, often pigmented surface. **Figs. 203-1 to 203-3** show vulvar and vaginal, penile, and perianal warts, respectively.

Vulvar warts are soft, whitish papules that are either sessile or have multiple fine, finger-like projections. These lesions are most often located in the introitus and labia. In nonmucosal areas, vulvar lesions are similar in appearance to those in men: dry and keratotic. Vulvar lesions can appear as smooth, sometimes pigmented papules that may coalesce. Vaginal lesions appear as multiple areas of elongated papillae. Biopsy of vulvar or vaginal lesions may reveal malignancy; differentiation based on clinical exam is not always reliable.

Subclinical cervical HPV infections are common, and the cervix may appear normal on examination. Cervical lesions often appear as papillary proliferations near the transformation zone. Irregular vascular loops are present beneath the surface epithelium. Patients who develop cervical cancer from HPV infection may present with a variety of symptoms. Early carcinomas appear eroded and bleed easily. More advanced carcinomas present as ulcerated lesions or as an exophytic cervical mass. Some cervical carcinomas are in the cervical canal and may be difficult to see. Bleeding, symptoms of a mass lesion in late stages, and metastatic disease that may manifest as bowel or bladder obstruction due to direct extension of the tumor also have been described.

Patients with squamous cell cancer of the anus **(Chap. 86)** have more variable presentations. The most common presentations include rectal bleeding and pain or a mass sensation. Twenty percent of patients who are diagnosed with anal cancer may not present with any specific symptoms at the time of diagnosis, and the lesion is found fortuitously.

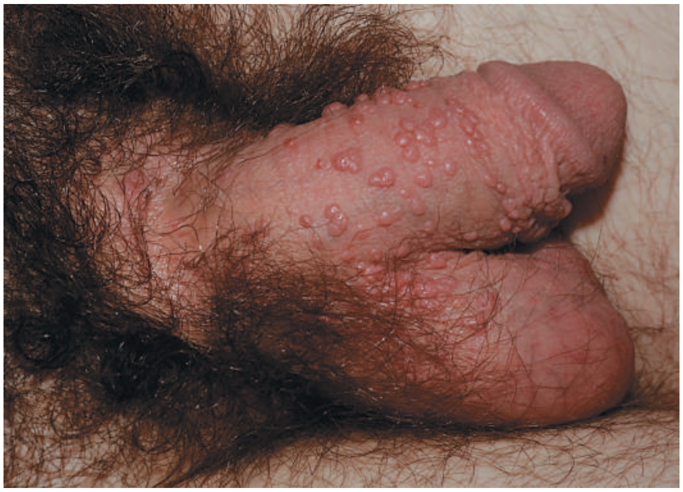

FIGURE 203-2 Penile genital warts caused by human papillomavirus. *(Reproduced with permission from K Wolff et al: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 8th ed. New York: McGraw-Hill, 2013.)*

PREVENTION OF HPV INFECTION AND DISEASE

Behaviors That Can Reduce Exposure to HPV HPV infections are transmitted through direct contact with infected genital skin

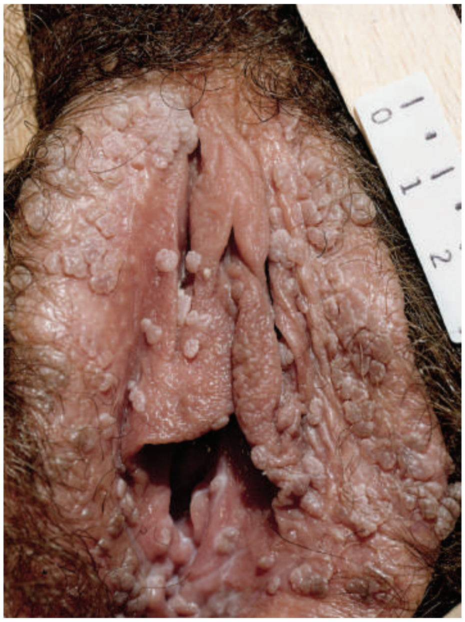

FIGURE 203-1 Warts of the vulva and vagina caused by human papillomavirus. *(Reproduced with permission from K Wolff et al: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 8th ed. New York: McGraw-Hill, 2013.)*

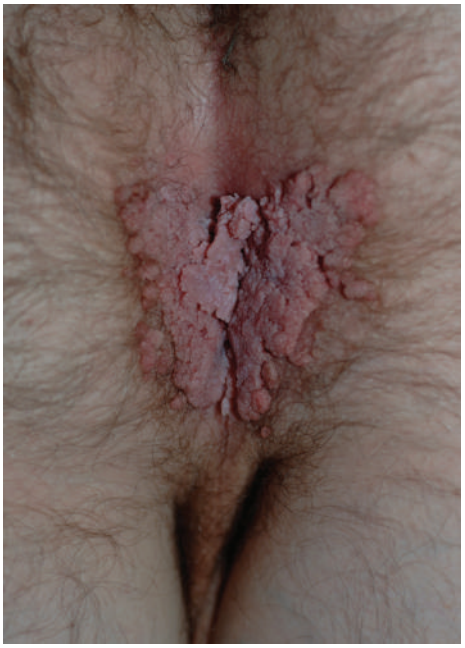

FIGURE 203-3 Perianal warts caused by human papillomavirus. *(Reproduced with permission from K Wolff et al: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 8th ed. New York: McGraw-Hill, 2013.)*

or mucosal surfaces and secretions. Abstinence may possibly reduce HPV infections: for both men and women, numerous studies indicate that HPV infection and HPV-associated diseases correlate with the number of lifetime sexual partners, and people with no history of sexual intercourse have a lower detection rate of HPV. Fewer studies look at nonpenetrative sex on the risk of HPV infection and disease, but several studies indicate that HPV can be spread by *any* sexual intimacy, including touching, oral sex, or use of sex toys. It is therefore possible that individuals who have not partaken in sexual intercourse can become infected.

Use of latex condoms reduces the risk of HPV infection and HPV-associated disease, such as genital warts and cervical precancers. Correct and consistent condom use has also been associated with regression of CIN in women and regression of HPV-associated penile lesions in men. As a preventive measure, condom use should be considered partially effective at best and not a substitute for cervical cancer screening or vaccination against HPV.

HPV Vaccines The development of HPV vaccines effective in preventing infection and HPV-associated disease represents a major development in the past decade. The vaccines use virus-like particles (VLPs) that consist of the HPV L1 major capsid protein. The L1 protein self-assembles into VLPs when expressed in eukaryotic cells (i.e., yeast or insect cells). These VLPs contain the same epitopes as actual HPV virions. However, they do not contain genetic material and therefore cannot transmit infection. The immunogenicity of the HPV vaccines relies on development of conformational neutralizing antibodies directed toward epitopes displayed on viral capsids.

Several large vaccine trials have been completed and demonstrate the high degree of safety and efficacy of HPV vaccines. There have been three HPV vaccines developed, tested, and U.S. Food and Drug Administration (FDA) approved, as described below.

BIVALENT VACCINE (CERVARIX) The bivalent HPV vaccine contains L1 VLPs of HPV types 16 and 18 and is marketed under the name of Cervarix (GlaxoSmithKline). This vaccine was tested in 18,644 women 15–25 years of age residing in the United States, South America, Europe, and Asia. It is administered by intramuscular injection three times (months 0, 1, and 6). The primary endpoints of the study included vaccine efficacy against persistent infections with HPV types 16 and 18. Investigators also assessed vaccine efficacy against CIN grade 2 or higher due to HPV 16 and 18 in women who had no evidence of HPV 16 or 18 infection at baseline. Vaccine efficacy related to HPV 16 or HPV 18 was 94.9% (95% confidence interval [CI], 87.7–98.4%) against CIN 2 or worse; 91.7% (95% CI, 66.6–99.1%) against CIN 3 or worse; and 100% (95% CI, −8.6–100%) against adenocarcinoma in situ (AIS). Adverse events associated with the bivalent vaccine were evaluated in phase 3 trials in a subset of 3077 women who received vaccine and 3080 women who received hepatitis A vaccine. Injection-site adverse events (pain, redness, and swelling) and systemic adverse events (fatigue, headache, and myalgia) were reported more frequently in the HPV vaccine group than in the control group. Serious adverse events, new-onset chronic disease, or medically significant conditions occurred in the same proportion (3.5%) of HPV vaccine recipients and control vaccine recipients. The bivalent HPV vaccine is approved in the United States for prevention of cervical cancer, CIN2 or worse, AIS, and CIN 1 caused by HPV types 16 and 18. This vaccine is approved for females 9–25 years of age. Cervarix is not currently marketed in the United States.

QUADRIVALENT VACCINE (GARDASIL) The quadrivalent L1 VLP vaccine (HPV types 6, 11, 16, and 18) is marketed under the name Gardasil (Merck). It is administered intramuscularly three times (months 0, 2, and 6). A combined efficacy analysis based on data from four randomized double-blind clinical studies including >20,000 participants was performed; results demonstrated that vaccine efficacy against external genital warts was 98.9% (95% CI, 93.7–100%). Vaccine efficacy was 95.2% (95% CI, 87.2–98.7%) against CIN; 100% (95% CI, 92.9–100%) against type 16- or 18-related CIN 2/3 or AIS; and 100% (95% CI, 55.5–100.0%) against type 16- or 18-related vulvar intraepithelial neoplasia grades 2 and 3 (VIN 2/3) and against vaginal intraepithelial neoplasia grades 2 and 3 (VaIN 2/3).

Safety data on the quadrivalent HPV vaccine are available from at least seven clinical trials, including nearly 12,000 women 9–26 years of age who received the vaccine and ~10,000 women who received aluminum-containing or saline placebo. A larger proportion of young women reported injection-site adverse events in the vaccine groups than in the placebo groups. Systemic adverse events were reported by similar proportions of vaccine and placebo recipients and were described as mild or moderate for most participants. The types of serious adverse events reported were similar for the two groups. Ten persons who received the quadrivalent vaccine and seven persons who received placebo died during the course of the trials; no deaths were considered to be vaccine related.

During the course of studies on the quadrivalent HPV vaccine, surveillance data for development of new medical conditions were collected for up to 4 years after vaccination. No statistically significant differences in the incidence of any medical conditions between vaccine and placebo recipients were demonstrated; this result indicated a very high safety profile for the vaccine. A recent safety review by the FDA and the Centers for Disease Control and Prevention (CDC) examined events related to Gardasil that had been reported to the Vaccine Adverse Events Reporting System (VAERS). The adverse events were consistent with what was seen in previous safety studies of the vaccine. Of note, rates of syncope and venous thrombotic events were higher with Gardasil than those usually observed with other vaccines.

The quadrivalent HPV vaccine is approved for (1) vaccination of females ages 9–26 years of age to prevent genital warts and cervical cancer caused by HPV types 6, 11, 16, and 18; (2) vaccination of the same population to prevent precancerous or dysplastic lesions, including cervical AIS, CIN 2/3, VIN 2/3, VaIN 2/3, and CIN 1; (3) vaccination of males 9–26 years of age to prevent genital warts caused by HPV types 6 and 11; and (4) vaccination of patients ages 9–26 years to prevent anal cancer and associated precancerous lesions due to HPV types 6, 11, 16, and 18. While the duration of protection has not been established, no evidence of waning protection has been found after a three-dose series of the quadrivalent HPV vaccine, even after 10 years of follow-up from clinical trials. The quadrivalent HPV vaccine is no longer available in the United States but is still available in many other countries, although production is not likely to continue in the future.

NINE-VALENT VACCINE (GARDASIL-9) In 2014, the FDA approved a new nine-valent L1 VLP vaccine. The nine-valent vaccine is marketed under the name Gardasil-9 (Merck). It is administered intramuscularly two times (months 0 and 6) for males and females from age 9 to the fifteenth birthday, then three times (months 0, 2, and 6) for males and females from age 15 through age 45. Three doses of HPV vaccine are recommended for persons starting the vaccination series on or after the fifteenth birthday and for persons with certain immunocompromising conditions, including HIV/AIDS.

The nine-valent vaccine targets HPV types 6, 11, 16, and 18 (the types also targeted by the quadrivalent HPV vaccine) as well as five additional oncogenic HPV types (31, 33, 45, 52, and 58). HPV types 16 and 18 together cause up to 80% of all cervical cancers worldwide, and worldwide data show that HPV types 31, 33, 35, 45, 52, and 58 are the next most frequently detected types in invasive cervical cancers. Mathematical models estimate that the level of protection conferred by the nine-valent HPV vaccine against all HPV-associated squamous cell cancers worldwide could be raised to at least 90%.

In clinical studies of females 16–26 years of age, the nine-valent HPV vaccine generated a noninferior antibody response to HPV types 6, 11, 16, and 18 compared with the quadrivalent HPV vaccine. Bridging immunologic studies in male and female vaccine recipients 9–15 years of age and in males 16–26 years of age indicated that the lower bound of the 95% CIs of the geometric mean titer ratio and seroconversion rates met criteria for noninferiority for all HPV types represented in the vaccine. In female recipients 16–26 years of age, vaccine efficacy against the combined endpoint of high-grade cervical, vulvar, or vaginal disease caused by any of the five additional oncogenic HPV types was 96.7% (95% CI, 80.9–99.8%). Like the other available HPV vaccines, the nine-valent HPV vaccine is safe and extremely well tolerated. The nine-valent HPV vaccine is approved for 9- to 45-year-old males and females and has an FDA indication for prevention of cervical, vaginal, vulvar, and anal cancer and genital warts due to vaccine types.

CROSS-PROTECTION OF HPV VACCINES Women who receive any of the available HPV vaccines produce neutralizing antibodies to virus types that are closely related to type 16 or 18. Analyses of data from clinical trials suggest that the HPV vaccines may offer limited cross-protection against nonvaccine virus types. Over short periods, the bivalent vaccine appears more efficacious against HPV types 31, 33, and 45 than the quadrivalent vaccine, but differences in study design make direct comparisons difficult if not impossible. In addition, in the bivalent vaccine trials, vaccine efficacy against persistent infections with HPV types 31 and 45 waned over time, whereas efficacy against persistent infection with HPV type 16 or 18 remained stable. These results suggest that cross-protection is likely to be shorter lived than efficacy against infection and disease caused by vaccine types.

SINGLE-DOSE VACCINATION As a cost saving measure, the World Health Organization has recommended that 9- to 15-year-olds may be vaccinated with a single dose of an HPV vaccine. This strategy appears to yield similar antibody titers to a two-dose regimen, but the level and duration of protection against infection and disease are not fully understood. A single-dose strategy has not been recommended by the FDA or other regulatory agencies in the United States.

RECOMMENDATIONS FOR HPV VACCINATION The most recent guidelines for HPV vaccination from the Advisory Committee on Immunization Practices (ACIP) are summarized below and provided in detail at *https://www.cdc.gov/vaccines/hcp/acip-recs/vacc-specific/hpv.html*.

No prevaccination testing of any kind is recommended to establish whether the HPV vaccine should be administered to an individual. The HPV vaccine should be administered, if possible, before exposure to HPV through sexual activity because the vaccines are preventative against specific HPV types and have no effect on preexisting, type-specific HPV infections. Either the bivalent (where available) or nine-valent HPV vaccines may be used. An individual can begin a vaccine series with one HPV vaccine and then complete the series with another. For those who have completed a vaccination series with the bivalent or quadrivalent vaccine, an additional full series (two or three doses, depending on age as indicated above) of vaccination with the nine-valent vaccine may be given, but there are no data to determine the effectiveness of this approach.

For children, adolescents, and adults (male and female) 9–26 years of age, the ACIP recommends HPV vaccination at age 11 or 12 years, although vaccination can be initiated at 9 years of age per FDA approval as above. “Catch-up” HPV vaccination is recommended for men and women through 26 years of age who are not adequately vaccinated.

For adults (male and female) 27–45 years of age, catch-up HPV vaccination is not routinely recommended. Instead, the ACIP now recommends “shared clinical decision-making” (see below) regarding HPV vaccination for adults in this age range who are not adequately vaccinated. HPV vaccines are not licensed for use in adults older than 45 years of age. For women, cervical cancer screening should continue according to age-specific guidelines regardless of having received an HPV vaccine (see cervical cancer screening section below).

SHARED CLINICAL DECISION-MAKING FOR ADULTS (MALES AND FEMALES) 27–45 YEARS OF AGE A discussion with adults 27–45 years of age should occur prior to routine recommendation of the HPV vaccine. HPV infection occurs soon after first sexual activity in most people, and vaccine effectiveness is therefore lower in older individuals due to prior infections. HPV exposure usually decreases among older age groups. Although HPV vaccination is safe for adults 27–45 years of age, the benefit to the population is likely to be minimal. However, some men and women who are not vaccinated may be at risk for acquisition of new HPV infections and could therefore benefit from HPV vaccination.

In considering HPV vaccination of adults 27–45 years of age, some key points emphasized by the ACIP that should be discussed include the following:

- HPV is a common sexually transmitted infection, and most HPV infections are asymptomatic and do not lead to clinical disease.
- Most sexually active adults have been exposed to HPV, although not necessarily all of the HPV types targeted by vaccines.
- Some adults are at risk for acquiring new HPV infections through sexual activity. For example, having a new sex partner is a risk factor for acquiring a new HPV infection.
- Persons in a long-term, mutually monogamous sexual partnership are unlikely to acquire a new HPV infection.
- Antibody testing cannot determine whether a person is immune or susceptible to a specific HPV type.
- HPV vaccines are very effective in persons who have not been exposed to vaccine-type HPV before vaccination.
- Vaccine effectiveness is likely to be lower among persons with multiple lifetime sex partners because these individuals have probably had previous infections with vaccine-type HPV.
- HPV vaccines are prophylactic (i.e., they prevent new HPV infections). They have no utility in preventing established HPV infection from progressing to clinical disease, and they do not have a role in treatment of HPV-associated disease.

RECOMMENDATIONS FOR HPV VACCINATION IN PEOPLE LIVING WITH HIV (PLWH) Guidelines for HPV vaccination of PLWH are summarized below and can be found in detail at *https://hivinfo.nih.gov*.

HPV vaccines are safe in PLWH. Administration of HPV vaccines generates high levels of antibody against HPV types represented in vaccine, although antibody levels are generally lower than in those who are HIV-uninfected. In addition, immune responses appear stronger among PLWH who have the highest CD4 counts and the lowest HIV viral loads. Studies also indicate that HPV vaccination induces an anamnestic response in PLWH. Regarding efficacy in protecting against HPV-associated disease, one randomized, double-blind, clinical trial evaluated the efficacy of the quadrivalent HPV vaccine in adults with HIV infection older than 27 years in prevention of new anal HPV infections or improvement in high-grade dysplastic anal lesions. The trial did not show efficacy, but many study participants had HPV infection detected at baseline, prior to vaccination.

HPV vaccination is recommended (three doses) for girls and boys with HIV infection 11–26 years of age. Because some individuals with HIV infection (similar to HIV-uninfected individuals) have had many sex partners prior to vaccination, HPV vaccination may be less beneficial in these patients than in those with few or no lifetime sex partners. Current data do not strongly support vaccination for those PLWH older than 26 years. The public health benefit for HPV vaccination of PLWH in this age range is likely to be minimal. However, although many PLWH ages 27–45 years will not fully benefit from the vaccine, there may be situations that suggest the possibility of vaccine benefit, and the same shared clinical decision-making (described above) between the provider and patient is recommended.

SCREENING FOR HPV-ASSOCIATED CANCER

Once HPV infection occurs, prevention of HPV-associated disease relies on screening. At present, screening for cervical cancer is widely accepted as cost-effective in preventing cervical cancer. Anal screening is accepted for screening in high-risk groups, though no national guidelines exist for screening intervals or ages for initiation and cessation of screening. In resource-rich countries, the primary method of cervical cancer screening is cytology via Pap smear. The American Society of Colposcopy and Cervical Pathology (ASCCP) guidelines recommend initiation of cervical cancer screening at age 21, no matter the age of sexual debut. Women 21–29 years old should have a Pap smear every 3 years if their initial and subsequent Pap smears are normal. Although adolescent and young women often test HPV DNA positive, they are at very low risk of cervical cancer. Because the presence of HPV DNA does not correlate with the presence of high-grade squamous intraepithelial neoplasia, co-testing (testing for HPV DNA at the time of Pap smear) is not recommended for women in this age group.

As a method of determining the need for colposcopy, HPV DNA co-testing is recommended for women 25–29 years of age in whom cytology detects abnormal squamous cells of undetermined significance (ASCUS). Women 30–65 should have a Pap smear every 3 years if testing for HPV DNA is not performed. The screening interval for women in this age group can be extended to every 5 years if HPV DNA co-testing is performed and results are negative. HPV testing is not recommended for partners of women with HPV or for screening of conditions other than cervical cancer.

The role of HPV DNA testing as a primary screen for cervical cancer is changing. In the United States, there are three commercially available assays (cobas® HPV Test [Roche Diagnostics], the BD Onclarity HPV Assay [Becton, Dickinson and Company], and Alinity m HR HPV Assay (Abbott Molecular Diagnostics)) that are FDA approved for primary screening using HPV DNA testing. However, more assays may gain approval for usage as the feasibility and evidence for their use in various populations globally come to light. These tests can be used to detect HPV DNA in specimens obtained from the cervix without cervical cytology for women ≥25 years of age. A positive result for HPV type 16 or 18 has a high enough positive predictive value in the general population that these women should have colposcopy performed. If high-risk HPV types other than HPV 16 or HPV 18 are detected, then cytology can be obtained. The complete set of algorithms for appropriate age-specific screening guidelines, HPV DNA testing, and the management of abnormal Pap smears are available through the ASCCP at *http://asccp.org/guidelines*.

For women ≤30 years of age who are infected with HIV, cervical cytology is the preferred method of cervical cancer screening and HPV DNA co-testing is not recommended. Cervical cancer screening should begin within 1 year of diagnosis of HIV infection, regardless of the mode of HIV transmission. If the first Pap smear is normal, then subsequent Pap smears should be performed annually until three negative tests are obtained. Cytology can then be obtained every 3 years. For women ≥30 years old, Pap testing is performed in the same manner as for younger women. However, HPV DNA co-testing can be used in women of this age group. If cytology and HPV DNA co-testing are negative, the next exam can be performed in 3 years. Positive HPV DNA co-test results are treated in the same manner as in HIV-uninfected women.

Women residing in developing countries with a lack of access to cervical screening programs have a higher rate of cervical cancer and a poorer cancer-specific survival. Approximately 75% of women living in developed countries have been screened in the past 5 years, as opposed to ~5% of women living in developing countries. Economic and logistic obstacles likely impede routine cervical cancer screening for these populations. Many poor countries rely on an alternative method—visual inspection with acetic acid (VIA)—for cervical cancer screening. While some studies show a reduction in cervical cancer mortality in communities where VIA is widely utilized, other studies do not. In addition, the low specificity of VIA is problematic. As newer methods that use detection of oncogenic HPV DNA become available, even resource-limited countries may be able to replace VIA with such methods and achieve a reduction in cervical cancers as a result.

There are no anal cancer screening guidelines endorsed by organizations such as the United States Preventative Services Task Force. The International Anal Neoplasia Society has published consensus guidelines regarding screening for anal cancer in patients at higher risk such as those with HIV. Current HIV treatment guidelines suggest that there may be a benefit to screening, but an effect on the associated morbidity and mortality of anal squamous cell cancer has not been consistently demonstrated. Although the most effective method of screening for anal cancer has not been determined, data published from a randomized controlled treatment trial noted a significant reduction in progression of high-grade precancerous lesions of the anus to anal cancer in those who received treatment. Further studies on optimal screening strategies for anal cancer are ongoing.

The incidence of HPV-associated head and neck cancers in the United States has overtaken the incidence of cervical cancer as of 2020, but there are no established guidelines for screening for HPV-associated head and neck cancers. However, HPV vaccination is likely to be effective for both anal and head and neck cancers associated with HPV.

TREATMENT

HPV-Associated Disease

A variety of treatment modalities are available for various HPV infections, but none has been proven to eliminate HPV from tissue adjacent to the destroyed and infected tissue. Treatment efficacies are limited by frequent recurrences, presumably due to reinfection from an infected partner, reactivation of latent virus, or autoinoculation from nearby infected cells. The goals of treatment include prevention of viral transmission, eradication of premalignant lesions, and reduction of symptoms.

Therapies are generally successful in eliminating visible lesions and grossly diseased tissue. Different therapies are indicated for genital warts, vaginal and cervical disease, and perianal and anal disease.

THERAPEUTIC OPTIONS

Imiquimod Imiquimod (5 or 3.75% cream) is a patient-applied topical immunomodulatory agent thought to activate immune cells by binding to a Toll-like receptor that leads to an inflammatory response. Imiquimod 5% cream is applied to genital warts at bedtime three times per week for up to 16 weeks. Warts are cleared in ~56% of patients, more often in women than in men; recurrence rates approach 13%. Local inflammatory side effects are common. Rates of clearance of genital warts are not as high with the 3.75% formulation as with the 5% preparation, but the duration of treatment is shorter (daily application required for a maximum of 8 weeks) and fewer local and systemic adverse reactions occur. Imiquimod should not be used to treat vaginal, cervical, or anal lesions. The safety of imiquimod during pregnancy has not been established.

Interferon Recombinant interferon α is used for intralesional treatment of genital warts, including perianal lesions. The recommended dosage is 1.0 × 10^6 IU of interferon into each lesion three times weekly for 3 weeks. Interferon therapy causes clearance of infected cells by immune-boosting effects. Adverse events include headache, nausea, vomiting, fatigue, and myalgia. Interferon therapy is costly and should be reserved for severe cases that do not respond to less expensive treatments. Interferon should not be used to treat vaginal, cervical, or anal lesions.

Cryotherapy Cryotherapy (liquid nitrogen treatment) for HPV-associated lesions causes cellular death. Genital warts usually disappear after two or three weekly sessions but often recur. Cryotherapy, which is nontoxic and is not associated with significant adverse reactions, can also be used for diseased cervical tissue. Local pain occurs frequently.

Surgical Methods Exophytic lesions can be surgically removed after intradermal injection of 1% lidocaine. This treatment is well tolerated but can cause scarring and requires hemostasis. Genital warts can also be destroyed by electrocautery, in which no additional hemostasis is required.

Laser Therapy Laser treatment affords destruction of exophytic lesions and other HPV-infected tissue while preserving normal tissue. Local anesthetics are generally adequate. Efficacy for genital lesions is at least equal to that of other therapies (60–90%), with low recurrence rates (5–10%). Complications include local pain, vaginal discharge, periurethral swelling, and penile or vulvar swelling. Laser therapy has also been used successfully for cervical dysplasia and anal disease caused by HPV.

Therapeutic Vaccines The innate and adaptive immune systems are altered in patients with HPV-associated cancers. Antitumor

TABLE 203-1 Recommended Treatments for Genital Warts Caused by Human Papillomavirus[a]

TREATMENT	IMIQUIMOD	CRYOTHERAPY	INTERFERON	SURGICAL REMOVAL	LASER
Effectiveness	Good	Good	Good	Excellent	Excellent
Recurrence	Frequent	Frequent	Frequent	Frequent	Frequent
Adverse effects	Frequent, mild to moderate	Mild, well tolerated	Frequent, moderately severe	Mild, well tolerated	Mild to moderate, well tolerated
Availability	Fair	Good	Fair	Good	Fair
Cost	Expensive	Inexpensive	Very expensive	Moderately expensive	Very expensive

[a]Imiquimod can be self-administered. All other treatments must be administered by a clinician.

immune responses are blunted by specific viral mechanisms. Numerous therapeutic vaccines that are being developed are designed to enhance the cell-mediated response to the HPV E6 and E7 oncoproteins, which are expressed in HPV-associated cancers. Such vaccines would enhance the ability to treat HPV-associated cancers, conditions that are very difficult to treat with current modalities. However, while progress has been made, no HPV vaccine is currently available for treatment of HPV infection or HPV-associated disease.

Other Therapies Both trichloroacetic acid and bichloroacetic acid are caustic agents that destroy warts by coagulation of proteins. Neither of these agents is recommended for treatment. Sinecatechins (15% ointment) and podophyllotoxin (0.05% solution or gel and 0.15% cream) are occasionally used for external genital warts, but other modalities listed above are as or more effective and are better tolerated.

RECOMMENDATIONS FOR TREATMENT

Table 203-1 lists available treatments for genital warts. An optimal therapy for HPV-related genital tract disease that combines high efficacy, low toxicity, low cost, and low recurrence is not available. For genital warts of the penis or vulva, cryotherapy is the safest, least expensive, and most effective modality. However, all available modalities for treatment of genital warts carry high rates of recurrence. Guidelines for the treatment of anogenital warts can be found on the CDC website (*https://www.cdc.gov/std/treatment-guidelines/anogenital-warts.htm*).

Women with vaginal lesions should be referred to a gynecologist experienced in colposcopy and treatment of these lesions. Treatment of cervical disease involves careful inspection, biopsy, and histopathologic grading to determine the severity and extent of disease. Women with evidence of HPV-associated cervical disease should be referred to a gynecologist familiar with HPV and experienced in colposcopy. Optimal follow-up of these patients includes colposcopic examination of the cervix and vagina on a yearly basis. Guidelines from the American College of Obstetricians and Gynecologists are available for the treatment of cervical dysplasia and cancer.

For anal or perianal lesions, cryotherapy or surgical removal is safest and most effective. Anoscopy and/or sigmoidoscopy should be performed in patients with perianal lesions, and suspicious lesions should be biopsied to rule out malignancy.

COUNSELING PATIENTS REGARDING HPV DISEASE

Most sexually active adults will be infected with HPV during their lives. The only way to avoid acquiring an HPV infection is to abstain from sexual activity, including intimate touching and oral sex. Practicing safe sex (partner reduction, use of condoms) may help reduce HPV transmission. Most HPV infections will be controlled by the immune system and cause no symptoms or disease. Some infections lead to genital warts and cervical precancers. Genital warts can be treated for cosmetic reasons and to prevent spread of infection to others. Even after resolution of genital warts, latent HPV may persist in normal-appearing skin or mucosa and thus theoretically may be transmitted to uninfected partners. Precancerous cervical lesions should be treated to prevent progression to cancer.

FURTHER READING

Akhatova A et al: Prophylactic human papillomavirus vaccination: From the origin to the current state. Vaccines (Basel) 10:1912, 2022.

Clifford GM et al: Carcinogenicity of human papillomavirus (HPV) types in HIV-positive women: A meta-analysis from HPV infection to cervical cancer. Clin Infect Dis 64:1228, 2017.

Garland SM et al: Impact and effectiveness of the quadrivalent human papillomavirus vaccine: A systematic review of 10 years of real-world experience. Clin Infect Dis 63:519, 2016.

Gavinski K, DiNardo D: Cervical cancer screening. Med Clin North Am 107:259, 2023.

Gelbard MK, Munger K: Human papillomaviruses: Knowns, mysteries, and unchartered territories. J Med Virol 95:e29191, 2023.

Giuliano AR et al: Efficacy of quadrivalent HPV vaccine against HPV infection and disease in males. N Engl J Med 364:401, 2011.

Gravitt PE, Winer RL: Natural history of HPV infection across the lifespan: Role of viral latency. Viruses 9:265, 2017.

Palefsky J et al: Treatment of anal high-grade squamous intraepithelial lesions to prevent anal cancer. N Engl J Med 386:2273, 2022.

Rosenblum HG et al: Declines in prevalence of human papillomavirus vaccine-type infection among females after introduction of vaccine — United States, 2003–2018. MMWR Morb Mortal Wkly Rep 70:415, 2021.

Schiffman M et al: Carcinogenic human papillomavirus infection. Nat Rev Dis Primers 2:16086, 2016.

Section 13 Infections Due to DNA and RNA Respiratory Viruses

204

Common Viral Respiratory Infections, Other Than COVID-19

James E. Crowe, Jr.

The most common and frequent infections in humans are respiratory virus infections. Influenza viruses and coronaviruses have been the agents responsible for the largest infectious disease pandemics. These viruses are easily transmitted by contact, droplets, and fomites. Furthermore, transmission can occur before the appearance of symptoms. These viruses are also associated with a large reproductive number (the number of secondary infections generated from one infected individual to others). Some classical respiratory viruses (e.g., rhinoviruses) enter the body through the respiratory tract, replicating and causing disease only in cells of the respiratory epithelium. Other, more systemic viruses (e.g., measles virus and severe acute respiratory syndrome coronavirus [SARS-CoV]) spread via the bloodstream and cause systemic disease;

however, they also may enter through and cause disease in the respiratory tract. Although infections with systemic viruses often induce lifelong immunity against disease, respiratory viruses that do not cause high-grade viremia usually can reinfect the same host many times throughout life. Reinfection with the same virus is common because of incomplete or waning immunity after natural infection. Hundreds of different viruses cause infection of the respiratory tract, and within each virus type, there can be a nearly unlimited diversity of field strains that vary antigenically, geographically, and over time (e.g., antigenically drifting influenza viruses or coronaviruses). Specific antiviral treatment options are limited, and only a few licensed vaccines are available. **For further discussion of common respiratory virus infections, see Chap. 37 and syndrome-specific chapters. For further discussion of SARS-CoV-2 virus infections, see Chap. 205.**

Common viral respiratory infections can be categorized in several ways, including by site of anatomic involvement, disease syndrome, or etiologic agent.

ANATOMIC SITES IN THE HUMAN RESPIRATORY TRACT

The type of respiratory disease that develops during virus infection is dictated to a large degree by the cell types and tissue organization in the respiratory tract. The vocal cords mark the transition between the upper and lower respiratory tracts. The upper respiratory tract is a complex anatomic system with interconnected structures, including the sinuses, middle-ear spaces, Eustachian tubes, conjunctiva, nasopharynx, oropharynx, and larynx. The tonsils and the adenoids are large collections of lymphoid tissue in the pharynx that participate in immunity but also are susceptible to infections. The lower respiratory tract structures include the trachea, bronchi, bronchioles, alveolar spaces, and lung tissue, including epithelial cells and blood vessels. The epithelial cell types that line the respiratory tract are varied in morphology and function, and their susceptibility to different virus infections varies. The principal types of cells in the major airways are ciliated or nonciliated epithelial cells, goblet cells, and Clara cells. Smooth-muscle cells form major tissue structures around the epithelial structures of the large airways of the lower respiratory tract down to the level of the bronchioles, and these cells are reactive to intrinsic and extrinsic signals, including viral infection or exposure to allergens or pollutants. The pathologic process of wheezing is driven by smooth-muscle contraction and obstruction of airways caused by mucus accumulation and epithelial sloughing in the lumen. Reactive airways causing wheezing are most often due to constriction of lumen size at the level of the bronchioles (which have the narrowest lumen diameter of the airways). The lung does not have smooth-muscle or ciliated cells but instead possesses pneumocytes of types I and II. Pneumonia **(Chap. 131)** is an infection of the pneumocytes in the lung tissue and the alveolar spaces. The alveolar spaces also contain cells of the monocyte lineage, such as macrophages, which patrol the air spaces.

DISEASE SYNDROMES

Since different respiratory viruses tend to have a predilection for replication in differing cells or regions of the respiratory tract, it is possible for the well-trained clinician with epidemiologic information to understand the most likely associations of viruses with clinical syndromes. The clinical diagnoses for virus infections in the upper respiratory tract are rhinitis or the common cold, sinusitis, otitis media, conjunctivitis, pharyngitis, tonsillitis, and laryngitis. Some upper respiratory tract infections affect more than one upper respiratory tract anatomic site during a single infection, such as the classical pattern of pharyngoconjunctival fever during adenovirus infection. Lower respiratory tract syndromes also can be associated easily with anatomic region, including tracheitis, bronchitis, bronchiolitis, pneumonia, and exacerbations of reactive airway disease or asthma. Bronchiolitis is a disease condition characterized by trapping of air in the lungs with difficulty in expiration (i.e., wheezing); it is caused by inflammation or infection of the bronchioles, the smallest and most highly resistant airways. Again, mixed syndromes occur, such as laryngotracheitis, usually termed croup. Croup, a disease condition characterized by difficulty in inspiration associated with a barky cough, is caused by inflammation or infection of the larynx, trachea, and bronchi. When respiratory symptoms occur in the context of a respiratory viral illness with significant systemic signs, infection with particular agents can be suspected (e.g., influenza, measles, SARS-CoV, SARS-CoV-2, or hantavirus pulmonary syndrome [HPS]), with exposure history taken into account.

ETIOLOGIC AGENTS

■ RESPIRATORY VIRUSES CAUSING DISEASE IN IMMUNOCOMPETENT HOSTS

Children have more frequent respiratory virus infections than adults; thus, it was natural that many early discoveries about the viral causes of respiratory infections came from pediatric studies. The principal causes of acute viral respiratory infections were determined in large epidemiologic studies in the 1960s and 1970s, when cell culture of infectious agents became available. More recently, studies of viral epidemiology have been conducted in adults, especially in special populations such as the elderly, nursing home residents, and immunocompromised individuals. Rapid antigen detection tests (based on immunoassays for detection of viral proteins) became available for respiratory syncytial virus (RSV) and influenza virus in the 1980s. With the availability of sensitive and specific molecular tests, such as reverse transcription combined with the polymerase chain reaction (RT-PCR), studies in the past several decades have greatly increased the extent to which we understand the causes of viral respiratory infections. Multiplex panels of RT-PCR tests capable of detecting a dozen or more viruses are commonly available for clinical testing of respiratory secretions. Nested multiplex PCR assays performed in two stages provide sensitive tests that have been especially helpful in studies of infection in adults, who often shed much lower concentrations of virus in secretions than do children. Typically, influenza viruses, RSV, and human metapneumovirus (hMPV) are the most common causes of serious lower respiratory tract disease in otherwise healthy subjects; parainfluenza viruses (PIVs) and adenoviruses also cause substantial disease. Rhinoviruses (the most common cause of the common cold syndrome) have been increasingly associated with lower respiratory tract syndromes. Rhinovirus infection is so common, even in asymptomatic individuals, that it has been hard to establish clear figures for the role of rhinovirus in lower respiratory disease. COVID-19 and the associated public health measures deployed in 2020–2021 altered the epidemiology of respiratory viruses such that conventional viruses were greatly reduced in incidence, exhibited altered seasonality, or even disappeared (influenza type B Yamagata lineage viruses) in the years immediately following. Generally, about two-thirds of cases of respiratory illness in a research setting can be associated with a specific viral agent. Besides the viruses mentioned above (and discussed below), several additional viruses identified with molecular tools have been associated with respiratory illness. Still, our diagnostic tools remain suboptimal since a specific infectious agent is not identified in approximately one-third of clinical respiratory illnesses in large surveillance studies. It is likely that in most of these cases pathogens are not detected because of the very low titers of virus in patient samples at the time of clinical presentation, which may occur after the period of peak virus shedding. It is also possible that novel agents are yet to be identified. As emerging tools for metagenomic and "virome" studies (with sequencing of all nucleic acids in a sample) are applied in these settings in coming years, new agents and new associations with disease will probably be discovered.

■ RESPIRATORY VIRUSES CAUSING DISEASE IN IMMUNOCOMPROMISED HOSTS

Special populations of patients are susceptible not only to the conventional respiratory viruses discussed above but also to agents causing symptoms during reactivation of latent viruses or new infections with opportunistic agents. Most prominently, reactivating latent viruses, such as herpes simplex virus (HSV) and cytomegalovirus (CMV) and adenoviruses, cause disease in immunocompromised humans. Patients at most risk are those with hematopoietic stem cell or solid organ transplantation, leukopenia caused by chemotherapy, or advanced

HIV-AIDS. In immunosuppressed patients with pneumonia, CMV is the virus recovered most frequently during deep respiratory tract diagnostic procedures such as bronchoalveolar lavage. These patients also are highly susceptible to more frequent and more severe disease caused by common respiratory viruses, including RSV, hMPV, PIVs, influenza viruses, rhinoviruses, and adenoviruses. Conventional acute respiratory viruses can cause chronic and sometimes fatal infections in these populations. Nosocomial transmission of respiratory viruses occurs in hematopoietic stem cell transplantation units, and the frequency of transmission can be high, with entire units affected.

SPECIFIC VIRAL CAUSES OF RESPIRATORY DISEASE

Orthomyxoviridae: Influenza Viruses (See also Chap. 206)

Influenza virus infection and influenza syndrome usually are associated with fever, myalgias, fatigue, sore throat, headache, and cough. Influenza causes severe and even fatal pneumonia, particularly in elderly patients, nursing home residents, immunocompromised persons, and very young children. Influenza pneumonia has an unusually high rate of complication by bacterial superinfection, with staphylococcal and streptococcal bacterial pneumonia occurring in as many as 10% of cases in some clinical series.

Influenza is a single-stranded, segmented, negative-sense, RNA genome virus of the family *Orthomyxoviridae*. There are four (sero) types of influenza viruses: A, B, C and D. Influenza A and C viruses infect multiple species, whereas influenza B virus infects humans almost exclusively. Type D viruses primarily infect cattle and are not known to infect humans. Type A viruses appear to be the most virulent for humans and most commonly cause severe disease manifestations, although type B viruses cause substantial morbidity. Based on antibody response, influenza A viruses can be subdivided into 18 different hemagglutinin (H) surface protein subtypes and 11 neuraminidase (N) surface protein subtypes. The subtypes that have caused major pandemics in humans are H1N1, which caused the 1918 pandemic; H2N2, which caused the 1957 pandemic; H3N2, which caused the 1968 pandemic; and H1N1pdm2009, which caused the 2009 pandemic. Currently, two type A subtypes (H1N1 and H3N2) and one type B lineage (Victoria) cause annual seasonal epidemics.

Major pandemics caused by new influenza viruses are always possible. Many highly pathogenic influenza viruses circulate in aquatic birds. Occasionally, avian viruses infect humans directly after close contact with infected wild birds or poultry. Co-housing of pigs (which have both avian and human influenza virus receptors) with poultry may increase the risk of reassortment of human and animal or bird viruses; reassortment can make the zoonotic viruses more fit for replication in humans. Several outbreaks of avian influenza have occurred in limited numbers of humans to date, and there is the risk of a worldwide pandemic with avian influenza viruses if a strain acquires the potential to spread efficiently from human to human. H5N1 influenza virus infection of humans, predominantly by direct chicken-to-human transmission, occurred during an epizootic in Hong Kong's poultry population in 1997. The disease affected many types of wild and domestic birds and caused a high rate of systemic disease and death in infected humans. This virus, carried in the gastrointestinal tract of wild birds, has spread throughout Asia and beyond and continues to evolve antigenically. An H5N1 virus was detected widely in dairy cattle in the United States in 2024 and inferred to be a highly pathogenic avian influenza (HPAI) virus. Avian H7N7 and H7N9 viruses also have caused zoonotic outbreaks. A significant outbreak of H7N9 virus infection began in China in March 2013, with high mortality, and there have been six outbreaks to date, the largest in 2016–2017 with 766 human infections. H7N9 is considered to have high potential to cause a future pandemic. H1N2 virus is endemic in pigs and affects humans with close contact. An H3N2 variant virus that differs antigenically from seasonal human viruses is endemic in pigs and occasionally infects children who have close contact with pigs in the United States. H3N8 is a subtype of equine influenza viruses that can infect humans who are in close contact with pigs. Rare human cases caused by H6, H9, and H10 subtype viruses have been reported. Type B influenza viruses co-circulate in humans during seasonal epidemics. Type B viruses mutate less frequently than type A viruses. The slower evolution of type B viruses is probably linked to the fact that they are almost exclusively human pathogens. There is only one B type of influenza, but these viruses began to diverge into two antigenically distinguishable lineages in the 1970s. The two virus lineages were named after the initial designated representative strains—B/Victoria/2/87 and B/Yamagata/16/88—and can be distinguished by serologic or genotyping laboratory tests. The evolution of B viruses over time spurred the inclusion of two type B virus antigens in seasonal influenza vaccines, expanding some multivalent vaccines from trivalent (H1N1, H3N2, B) to a quadrivalent format. During the COVID-19 pandemic, the diversity of influenza in humans has been reduced, as strains in lineage B/Yamagata and one clade of H3N2 known as 3c3.A were not detected. In 2023, the World Health Organization recommended future vaccines to be trivalent with only B/Victoria for the type B component.

Pneumoviridae/Paramyxoviridae (the formal species names of family *Pneumoviridae* were updated in 2023; Table 204-1)

• **RESPIRATORY SYNCYTIAL VIRUS** Human RSV (hRSV) (species name *Human orthopneumovirus hominis*) is a single-stranded, negative-sense, nonsegmented, RNA genome virus of the genus *Respirovirus* (formerly *Pneumovirus*) in the family *Paramyxoviridae*. Infection is ubiquitous, affecting most humans in the first several years of life and causing reinfections throughout life. RSV is among the most transmissible viruses of humans. Disease epidemics occur yearly, typically between October or November and March in temperate regions. RSV is one of the most common viral causes of severe lower respiratory tract illness in the elderly and in children; it is among the most important causes of hospitalization of elderly and infant patients throughout the world. There is only one serotype of RSV, but antigenic variability does occur in circulating field strains. In immune serum reciprocal cross-neutralization studies, the two antigenic subgroups, A and B, appear to be ~25% antigenically related; this relatedness may partially explain the susceptibility of humans to reinfection, which is very common and can be caused by viruses of the same subgroup or even the same strain. However, reinfection in otherwise healthy adults usually is associated with mild disease confined to the upper respiratory tract. Severe lower respiratory tract disease is common in the elderly, especially in frail institutionalized elderly populations. Immunocompromised patients of any age also are at risk of severe or prolonged disease, especially recipients of hematopoietic stem cell transplants. Wheezing is common with primary infection in children (bronchiolitis), and there is a strong association of RSV infection early in life and subsequent asthma, although it is unclear whether severe childhood RSV causes asthma or is the first manifestation of reactive airway disease. RSV causes exacerbations of asthma and is associated with acute exacerbations of chronic obstructive pulmonary disease (COPD), also referred to as acute exacerbations of chronic bronchitis (AECB).

HUMAN METAPNEUMOVIRUS hMPV (species name *Metapneumovirus hominis*) was discovered only in 2001 but probably has always been present in human populations. Infection occurs first in early childhood, and reinfections are common throughout life. This virus is similar in many respects to RSV, but it now is classified in the family *Pneumoviridae* and is a member of the genus *Metapneumovirus*. It causes both upper and lower respiratory disease. It appears to be

TABLE 204-1 Family *Pneumoviridae,* Human Pathogens with Current Species Names, the International Committee on Taxonomy of Viruses: 2023 Release

GENUS	CURRENT SPECIES NAME	FORMER SPECIES NAME(S)
Metapneumovirus	*Metapneumovirus hominis*	*Human metapneumovirus (hMPV)*
Orthopneumovirus	*Human orthopneumovirus hominis*	Human respiratory syncytial virus (hRSV) or *Human orthopneumovirus*

somewhat less virulent than RSV, causing about half as much severe lower respiratory tract disease, probably because it does not possess the nonstructural genes that RSV expresses in infected cells to abrogate the effect of host innate immune effectors like interferons. The clinical features of lower respiratory tract infections caused by hMPV are like those of such infections caused by other paramyxoviruses, most often including cough, coryza, and wheezing. Like RSV, hMPV plays an important role in exacerbations of asthma or COPD and causes pneumonia or wheezing in frail and institutionalized elderly individuals and immunocompromised patients.

***Paramyxoviridae* (the formal species names of family *Paramyxoviridae* were updated in 2023; Table 204-2) • HUMAN PARAINFLUENZA VIRUSES** The human PIVs (hPIV) are a group of four distinct serotypes (designated 1–4) of single-stranded, negative-sense RNA viruses belonging to the family *Paramyxoviridae*. hPIV3 (species name *Respirovirus pneumoniae*) most commonly causes severe disease, and repeated infection is common throughout life, although secondary infections often are mild or asymptomatic. Primary infections in children manifest as laryngotracheitis (croup), while subsequent infections typically are limited to the upper respiratory tract. hPIVs are detected with sensitive RT-PCR tests or, more classically, by cell culture with immunofluorescent microscopy or hemadsorption in reference laboratories.

MEASLES VIRUS **(See also Chap. 211)** Measles virus (species name *Morbillivirus hominis*) is also a paramyxovirus but of the genus *Morbillivirus*. This virus causes a systemic infection known as rubeola but also can manifest with respiratory symptoms. Measles virus probably is the most contagious respiratory virus infection of humans: it is transmitted efficiently not only by direct contact with infected persons or fomites (like other respiratory viruses) but also by small-particle aerosols. Measles virus infection is preventable by vaccination, but the pathogen is so transmissible that cases are inevitable—even in the United States—whenever vaccination rates fall below 90–95% in a population. The virus causes systemic illness, sometimes including severe pneumonia, when primary infection occurs in an unvaccinated adult or an immunocompromised person of any age. Therefore, vigilance in maintaining high vaccination rates is critical. With primary infection, the illness in children is typically milder; however, mortality rates in lower-resource countries are high, especially among persons with underlying risk factors, including malnutrition.

Symptoms of measles include ≥3 days of high fever and a classical set of upper and lower respiratory tract symptoms sometimes termed “the 3 Cs”: cough, coryza, and conjunctivitis. Unlike most respiratory viruses, measles virus circulates in the bloodstream and thus causes disseminated infection with systemic manifestations. Usually, a characteristic diffuse maculopapular rash appears within days of fever onset. Koplik’s spots **(see Fig. A1-2)**—typical mucosal lesions in the mouth that appear briefly—are considered diagnostic of measles infection in the setting of the typical rash and fever.

TABLE 204-2 Family *Paramyxoviridae* Human Pathogens with Current Species Names, the International Committee on Taxonomy of Viruses: 2023 Release

GENUS	CURRENT SPECIES NAME	FORMER SPECIES NAME(S)
Respirovirus	*Respirovirus laryngotracheitidis*	*Human respirovirus 1* or human parainfluenza virus type 1 (hPIV1)
	Respirovirus pneumoniae	Human parainfluenza virus type 3 (hPIV3) or *Human respirovirus 3*
Orthorubulavirus	*Orthorubulavirus parotitidis*	Mumps virus or *Mumps orthorubulavirus*
	Orthorubulavirus laryngotracheitidis	Human parainfluenza type 2 (hPIV2) or *Human orthorubulavirus 2*
	Orthorubulavirus hominis	Human parainfluenza type 4a (hPIV4a) or *Human orthorubulavirus 4*
		Human parainfluenza type 4b (hPIV4b) or *Human orthorubulavirus 4*
	Orthorubulavirus mammalis	Parainfluenza type 5 (PIV5) or *Mammalian orthorubulavirus 5*

Picornaviridae A wide variety of picornaviruses cause respiratory disease, including nonpolio enteroviruses, rhinoviruses, and parechoviruses **(Chap. 210)**. The designations of these viruses can be confusing: the *Enterovirus*, rhinovirus, and *Parechovirus* species names were changed (with the approval of the International Committee on Taxonomy of Viruses) to remove references to host species names (such as the formerly used terms human, simian, etc.), and they are frequently updated. These changes are summarized in **Table 204-3**. The genus *Enterovirus* consists of 15 species, including the enteroviruses and rhinoviruses affecting humans. The genus *Parechovirus* contains six species, one of which—Parechovirus A—encompasses 19 types: human parechoviruses (HPeVs) 1 and 2 are common human pathogens. These viruses exhibit seasonal patterns that differ from those of most other acute respiratory viruses. Rhinovirus infections occur year-round. *Enterovirus* infections occur most commonly in the summer months in temperate areas.

RHINOVIRUSES Rhinoviruses have single-stranded, positive-sense RNA genomes. Rhinoviruses A through C represent species in the *Enterovirus* genus of the family *Picornaviridae*. Rhinoviruses are the most common viral infective agents in humans and the most frequent cause of the common cold. Field isolates of rhinovirus are exceptionally diverse; they can be classified by serotyping into >100 serotypes or, alternatively, by genotyping into many genotypes that cause cold symptoms. The viral particles are icosahedral in structure and are nonenveloped. Rhinoviruses are responsible for at least half of all cases of the common cold. Rhinovirus-induced common colds may be complicated in children by otitis media and in adults by sinusitis. Most adults, in fact, have radiographic evidence of sinusitis during the common cold, which resolves without therapy. Therefore, the primary disease is probably best termed *rhinosinusitis*. Rhinovirus infection is associated with exacerbations of reactive airway disease in children and asthma

TABLE 204-3 Enterovirus, Rhinovirus, and Parechovirus Species Names, the International Committee on Taxonomy of Viruses: 2023 Release

GENUS	CURRENT SPECIES NAME	FORMER SPECIES NAME(S)
Enterovirus (15 species)	***Enterovirus alphacoxsackie***: consists of 25 serotypes, including coxsackieviruses and some nonpolio enteroviruses that cause respiratory disease	Human enterovirus A; Enterovirus A
	Enterovirus betacoxsackie: consists of 63 serotypes, including some coxsackieviruses, echoviruses, and nonpolio enteroviruses	Human enterovirus B; Enterovirus B
	Enterovirus coxsackiepol: consists of 23 serotypes, including the polioviruses	Human enterovirus C; Enterovirus C
	Enterovirus deconjuncti: consists of multiple serotypes and includes enterovirus D68	Human enterovirus D; Enterovirus D
	Enterovirus alpharhino	Human rhinovirus A; Rhinovirus A
	Enterovirus betarhino	Human rhinovirus B; Rhinovirus B
	Enterovirus cerhino	Human rhinovirus C; Rhinovirus C
Parechovirus (6 species)	***Parechovirus ahumpari***: consists of 19 types (1–19); human parechoviruses (HPeVs) 1 and 2 are common human pathogens	Parechovirus A; HPeV-1 and HPeV-2 were formerly classified in the genus *Enterovirus* as echoviruses 22 and 23, respectively

in adults. It is not clear whether rhinovirus is restricted to the upper respiratory tract and only indirectly induces inflammatory responses that affect the lower respiratory tract or whether the viruses spread to the lower respiratory tract. In the past, it was thought that these viruses did not often replicate or cause disease in the lower respiratory tract. However, recent studies have discerned strong epidemiologic associations of rhinoviruses with wheezing and asthma exacerbations, including episodes severe enough to require hospitalization. Rhinovirus C (species *Enterovirus cerhino*) has been associated with more severe disease syndromes, such as pneumonia or exacerbation of COPD. Rhinoviruses likely can infect the lower airways to some degree, inducing a local inflammatory response. Another possibility is that significant local infection of the upper respiratory tract may induce regional elaboration of mediators that causes lower airway disease. The association of rhinovirus infection with lower respiratory tract illness is difficult to study because diagnosis by cell culture is not sensitive. RT-PCR diagnostic tests are difficult to interpret because they are often positive for prolonged periods and even asymptomatic individuals may have a positive test. Comprehensive serologic studies to confirm infection are difficult because of the large number of serotypes. Nevertheless, most experts believe rhinoviruses are a common cause of serious lower respiratory tract illness.

ENTEROVIRUSES Nonpolio enteroviruses are common and distributed worldwide. Although infection often is asymptomatic, these viruses cause outbreaks of clinical respiratory disease, sometimes with fatal consequences. The species *Enterovirus alphacoxsackie* (formerly Enterovirus A) consists of 25 serotypes, including coxsackieviruses and some nonpolio enteroviruses that cause respiratory disease. Coxsackieviruses cause oral lesions and often are associated in children with hand-foot-and-mouth disease. The pharyngitis associated with this infection characteristically manifests with herpangina, a clinical syndrome of ulcers or small vesicles on the palate that often involves the tonsillar fossa and is associated with fever, difficulty swallowing, and throat pain. Outbreaks commonly occur in young children during the summer. Enterovirus A71 also causes large outbreaks of hand-foot-and-mouth disease, especially in Asia, sometimes leading to neurologic complications and even death. The species *Enterovirus betacoxsackie* (formerly Enterovirus B) consists of >90 serotypes, including the echoviruses (*echo* being an acronym for *e*nteric *c*ytopathic *h*uman *o*rphan, which may be an archaic notion since most echoviruses are associated with human diseases, most commonly in children). Echoviruses can be isolated from many children with upper respiratory tract infections during the summer months. Echovirus 11 has been associated with laryngotracheitis or croup. Epidemiologic studies also have associated echoviruses with epidemic pleurodynia, an acute illness characterized by sharp chest pain and fever. The species *Enterovirus coxsackiepol* (formerly Enterovirus C) consists of 23 serotypes, including the polioviruses. The species *Enterovirus deconjuncti* (formerly Enterovirus D) consists of five serotypes, including enterovirus D68, which has been associated with wheezing and a polio-like syndrome in children marked by acute flaccid myelitis.

PARECHOVIRUSES The genus *Parechovirus* comprises six species, one of which is *Parechovirus ahumpari* (formerly Parechovirus A), which can affect humans. The most common member of the genus *Parechovirus*, human parechovirus 1 (HPeV-1), is a frequent human pathogen. The genus also includes the closely related human parechovirus 2 (HPeV-2). HPeVs usually cause mild respiratory or gastrointestinal illness. Most infections occur in young children. The seroprevalence of HPeV-1 and HPeV-2 is high among adults.

Adenoviridae Viruses of the family *Adenoviridae* infect both humans and animals. As their designation indicates, adenoviruses were first isolated in human lymphoid tissues from surgically removed adenoids. In fact, some serotypes establish persistent asymptomatic infections in tonsil and adenoid tissues, and virus shedding can occur for months or years. These double-stranded DNA viruses are <100 nm in diameter and have nonenveloped icosahedral morphology. The large double-stranded DNA genome is linear and nonsegmented. The seven major human adenovirus species (designated A through G) fall into over 50 immunologically distinct serotypes. Human respiratory tract infections are caused mainly by the B and C species. Adenovirus infections can occur throughout the year. Many serotypes cause sporadic outbreaks, while others appear to be endemic in particular locations. Respiratory illnesses include mild disease such as the common cold and lower respiratory tract illnesses including croup, bronchiolitis, and pneumonia. Conjunctivitis is associated with infection by the B and D species. A particular constellation of symptoms referred to as pharyngoconjunctival fever is frequently associated with acute adenovirus infection. In contrast, gastroenteritis has been associated most frequently with virus serotypes 40 and 41 of species F. Immunocompromised patients are highly susceptible to severe disease during infection with respiratory adenoviruses. The syndrome of acute respiratory disease (ARD), especially common in stressful or crowded living conditions, was first recognized among military recruits during World War II and has continued to be a problem when vaccination has been suspended temporarily because of lapses in vaccine supply. ARD is most often associated with adenovirus types 4 and 7. Adenovirus vaccine containing live adenovirus types 4 and 7 taken orally as two tablets, which prevents most illness caused by these two virus types, is only available for U.S. military personnel 17–50 years of age. It is recommended by the Department of Defense for military recruits entering basic training or other military personnel at high risk for adenovirus infection.

Coronaviridae SARS-CoV-2 emerged in an outbreak in Wuhan, China, that spread worldwide, causing the severe pandemic of COVID-19. **SARS-CoV-2 is discussed separately in Chap. 205.**

Other members of the genus *Coronavirus* also contribute to respiratory illness, including severe disease. Dozens of coronaviruses affect animals. In the twentieth century, only two representative strains of human coronaviruses were known to cause disease: 229E (HCoV-229E) and OC43 (HCoV-OC43). An outbreak of infection with SARS-associated coronavirus (SARS-CoV) first showed that animal coronaviruses have the potential to cross from other species to humans, with devastating effects. The one major SARS-CoV epidemic to date (2002–2003) encompassed >8000 cases, with mortality rates approaching 10%. SARS-CoV causes a systemic illness with a respiratory route of entry. In contrast to most other viral pneumonias, SARS lacks upper respiratory symptoms, although cough and dyspnea occur in most patients. Typically, patients present with a nonspecific illness manifesting as fever, myalgia, malaise, and chills or rigors; watery diarrhea may occur as well. Investigators have reported the identification of a fourth human coronavirus, HCoV-NL63. Evidence is emerging that this new group 1 coronavirus is a common respiratory pathogen of humans, causing both upper and lower respiratory tract illness. HCoV-HKU1 was first described in January 2005 after its detection in a patient with pneumonia. Several cases of respiratory illness have been associated with this virus, but its infrequent identification suggests that this group 2 coronavirus has caused a low incidence of illness to date. The Middle East respiratory syndrome coronavirus (MERS-CoV), first isolated in 2012, causes severe disease in humans, with ~35% mortality and >2500 cases reported to date. MERS-CoV is a zoonotic virus (transmitted between animals and people). The virus likely emerged from bats in the Middle East, although studies have shown that humans are infected through direct or indirect contact with an intermediate host—infected dromedary camels.

Herpesviridae Several herpesviruses cause upper respiratory infections, especially infection of the oral cavity. Herpes simplex pharyngitis is associated with characteristic clinical findings, such as acute ulcerative stomatitis and ulcerative pharyngitis. HSV types 1 and 2—human herpesvirus (HHV) 1 (species *Simplexvirus humanalpha1*) and HHV2 (species *Simplexvirus humanalpha2*), respectively—both cause oral lesions **(Chap. 197)**, although >90% of oral infections are caused by HSV-1. Primary oral disease can be severe, especially in young children, who sometimes are admitted for rehydration therapy because of poor oral intake. A significant proportion of individuals suffer recurrences of symptomatic disease consisting of vesicles on the lips. Epstein-Barr virus (EBV) mononucleosis syndrome (Chap. 199) is often marked by acute or subacute exudative pharyngitis; in some cases, tonsillar swelling in EBV pharyngitis is so severe that airway occlusion appears imminent. Most of the viruses in the family *Herpesviridae*—including CMV (Chap. 200); EBV; varicella-zoster virus (VZV; Chap. 198); and HHV-6, -7, and -8 (Chap. 200)—can cause severe disease in immunocompromised patients, especially hematopoietic stem cell transplant recipients.

***Parvoviridae:* Human Bocavirus** Human bocavirus (HBoV) was identified in 2005 in respiratory samples from children with lower respiratory tract disease. Sequence analysis of the genome revealed that the virus is a member of the genus *Bocaparvovirus* (subfamily *Parvovirinae,* family *Parvoviridae*). This virus has been identified as the sole agent in a limited number of respiratory samples from individuals with respiratory tract disease, especially hospitalized young children, but the virus is also commonly found by RT-PCR tests in respiratory samples from healthy subjects.

***Retroviridae:* HIV** Pharyngitis occurs with primary HIV infection and may be associated with mucosal erosions and lymphadenopathy (Chap. 208).

***Papovaviridae:* Polyomaviruses** Polyomaviruses are small, double-stranded, DNA-genome, nonenveloped icosahedral viruses that may be oncogenic. Two major polyomaviruses, JC and BK viruses, are known to infect humans. Of adults in the United States, ≥80% are seropositive for these viruses. JC virus can infect the respiratory system, kidneys, or brain. BK virus infection causes a mild respiratory infection or pneumonia and can involve the kidneys of immunosuppressed transplant recipients.

EPIDEMIOLOGY

AGE

Age (along with the associated factor of prior exposure history) is a major determinant of risk for symptomatic disease during respiratory virus infection. Primary infection with most of the acute respiratory viruses often is more severe than secondary infection. Indeed, reinfection with most of these viruses occurs throughout life, but primary infection is much more likely to be associated with severe lower respiratory tract disease, while secondary infection typically is asymptomatic or associated with upper respiratory tract symptoms only. As these infections are ubiquitous, most primary infections (and thus many of the severe cases) occur during the first few years of life. Later, exposure to young children (in populations such as parents of young children and daycare workers) is a risk factor for frequent reinfection. Despite a lifetime of previous exposures, the risk of severe disease increases with age in the elderly, probably because of immune senescence and general medical decline.

SEASON

Infections with most of the conventional respiratory viruses (e.g., influenza virus, RSV, and hMPV) occur in winter. Typically, there is one dominant virus sweeping through a local community at any one time, a pattern that suggests some population-level interference with transmission. However, outbreaks can be closely spaced, and co-circulation of different viruses or antigenically diverse strains of one virus does occur. In the United States, some regional differences in seasonality have been noted; for example, RSV often appears in Florida and other southeastern states first. Seasons are, of course, reversed in the Northern and Southern hemispheres, so that winter epidemics occur roughly from November to March in the United States but from April to August in Australia; therefore, "winter" epidemics are almost always occurring somewhere in the world. Seasonal variances differ in the tropics, where acute respiratory viral infections are more common in the rainy season. The pandemic starting in 2020 caused by SARS-CoV-2 disrupted the seasonality of the common respiratory viruses for several years. SARS-CoV-2 outbreak patterns do not follow traditional seasonal patterns like other respiratory viruses, with peaks caused by antigenic variants at any time of the year thus far.

RISK FACTORS FOR DISEASE

Infection with these viruses is nearly universal, but disease expression varies among individuals infected with identical viruses. Therefore, investigators have sought to identify risk factors for severe disease. Most single risk factors identified have a moderate effect on the incidence of severe disease, but an accumulation of factors is associated with high risk. Underlying lung disease is a major factor, especially diseases associated with the need for chronic oxygen supplementation. COPD is one of the most profound risk factors. Other severe underlying medical conditions, especially cardiovascular disease, also enhance risk. Smoking (or exposure to wood smoke), low socioeconomic status, and male gender all contribute to minor increases in the risk of lower respiratory tract illness. Obesity causes a chronic state with features of inflammation that are associated with impaired immunity, reduced response to vaccination, and higher susceptibility to severe disease. Close exposure to infected people is a major factor. For instance, living in close quarters (e.g., housing for military trainees, college dormitories, or nursing homes) puts groups of individuals at risk for rapid outbreaks. A breakdown in isolation and hand-washing compliance procedures can lead to cycles of nosocomial transmission of infection in hospital inpatient wards and intensive care units. In assessments of severe lower respiratory tract illness, a history of travel to an area with unusual agents should be considered carefully (e.g., exposure to avian influenza outbreaks in Asia, exposure to MERS-CoV in the Middle East). In 2020–2021, the dominance of the SARS-CoV-2 outbreak and the associated health measures deployed reduced the incidence of conventional respiratory viruses.

TRANSMISSION

Most respiratory viruses are transmitted by two principal modes: fomites or large-particle aerosols of respiratory droplets spread directly from person to person by coughing or sneezing. Fomite transmission occurs indirectly when infected respiratory droplets are deposited on the hands or on inanimate objects and surfaces, with subsequent transfer of secretions to a susceptible person's nose or conjunctiva. Most respiratory viruses do not spread by small-particle aerosols across rooms or down halls, although measles virus and VZV do spread in this manner. Therefore, contact and droplet precautions are sufficient to prevent transmission in most settings; hand washing is especially critical in health care settings during the winter. Intensive studies of the SARS-CoV-2 pandemic are ongoing (see previous sections on COVID-19), but many experts agree that exposure to large-particle droplets likely is one of the major ways that SARS-CoV-2 spreads.

APPROACH TO THE PATIENT

Common Viral Respiratory Infections

The principal interventions that make a difference in the care of patients with acute respiratory virus infections are supportive, and these factors should be managed meticulously. Hypoxia is managed with supplemental oxygen and respiratory failure with mechanical ventilation. Because the tachypnea and fever that often accompany pneumonia and wheezing frequently result in dehydration, fluid management is important. The astute clinician can narrow the etiologic possibilities based on epidemiologic knowledge; information about viruses circulating in the community (widely available from local reference laboratories, county and state health departments, and the U.S. Centers for Disease Control and Prevention [CDC]); and the patient's exposure history, age, and immunologic status, including vaccination status. Proper use of rapid diagnostic tests is important. When diagnostic tests are applied only to samples from individuals at high risk of exposure to an infectious agent in the appropriate season, the positive predictive value of the test is increased. A central medical decision is whether to use a specific antibacterial or antiviral agent to treat a respiratory infection.

Antibiotics do not improve the outcome of uncomplicated respiratory virus infections in otherwise healthy subjects. Some viral infections, especially influenza, can be complicated by secondary bacterial infection. There are only a limited number of licensed antiviral drugs, which should be used when a specific viral etiology is determined. Antiviral treatment generally is effective only when administered early in the course of illness.

CLINICAL MANIFESTATIONS

The common cold is characterized by nasal congestion, sneezing, rhinorrhea, cough, and sore throat. Laryngitis is accompanied by hoarseness or dysphonia. Acute bronchitis is characterized by a dry or productive cough of <3 weeks' duration (most prevalent in winter) in the absence of signs and symptoms of pneumonia and of evidence of pneumonia on chest radiography and is primarily caused by viruses. Bacteria play a more prominent role in chronic bronchitis. Bronchiolitis is an acute illness with wheezing and evidence of upper respiratory infection, primarily seen in the winter in infants and young children. The typical clinical manifestations of acute pneumonia include cough, sputum production, dyspnea, and chest pain. More systemic signs and symptoms also occur in pneumonia, including fever, fatigue, sweats, headache, myalgia, and occasionally nausea, abdominal pain, and diarrhea.

DIAGNOSIS

The clinical diagnosis of a respiratory syndrome and the anatomic location of infection are based on history, physical examination, and radiography. A specific viral etiology can be determined by specific diagnostic tests. The gold standard for diagnosing a respiratory viral infection is virus isolation, performed by inoculation of cell cultures with fresh secretions and use of multiple cell types in a reference laboratory staffed by experienced technologists. Direct or indirect fluorescent antibody detection can be used to visualize virus-infected cells in nasal secretions. Rapid antigen-based diagnostic tests are used to detect influenza virus or RSV proteins in nasopharyngeal secretions. The most sensitive tests typically are RT-PCR molecular diagnostic tests that amplify and detect the presence of viral genomic RNA or DNA in respiratory secretions. Multiplex panels assaying a sample for a dozen or more common respiratory viruses are available. These tests must be used and interpreted carefully because of their extreme sensitivity. If care is not taken, it is relatively easy to contaminate a PCR test in the laboratory with small amounts of DNA from a previous reaction. In addition, because a viral genome can sometimes persist in nasal secretions for weeks after an infection resolves, a positive test may indicate a recently resolved rather than a currently acute infection. Despite these limitations, PCR tests generally are considered the most sensitive and specific tests available. Chest radiographs should be obtained for all patients with suspected pneumonia.

TREATMENT

Common Viral Respiratory Infections

INFLUENZA (SEE ALSO CHAP. 206)

Several drugs are licensed in the United States for the treatment or prophylaxis of influenza. Neuraminidase inhibitors act on both influenza A and B viruses by serving as transition-state analogues of the viral neuraminidase that is needed to release newly budded virion progeny from the surface of infected cells. The cell surface normally is coated heavily with the viral receptor sialic acid. Oseltamivir is administered orally twice daily and is effective for the prevention or treatment of uncomplicated influenza in otherwise healthy adults. Observational studies indicate that oseltamivir also may be beneficial during serious illness. The drug is generally well tolerated, with primarily gastrointestinal toxicity. Zanamivir, a powder that is administered through oral inhalation, exhibits effectiveness like that of oseltamivir. Moreover, zanamivir is active against some influenza virus strains that are resistant to oseltamivir. Inhalation of zanamivir powder may cause bronchospasm in patients with COPD or asthma. Peramivir is a neuraminidase inhibitor that acts as a transition-state analogue inhibitor of the influenza neuraminidase enzyme that is administered intravenously as a single 600-mg dose. It is efficacious in acute, uncomplicated influenza and was approved by the U.S. Food and Drug Administration (FDA) in 2014 for treatment of individuals who cannot take oral or inhaled medications. Laninamivir was approved in Japan for prophylaxis (2013) or treatment (2010) of influenza; it is under investigation in the United States. It is a polymeric zanamivir conjugate that is delivered by oral inhalation, and it exhibits greater potency and longer retention times than conventional zanamivir. Baloxavir marboxil is a relatively new class of drug for influenza. It is a prodrug whose metabolism releases the active agent baloxavir acid that inhibits influenza virus cap-dependent endonuclease activity in infected cells. This activity is used by the virus for a process in which the first 10–20 residues of a host cell RNA are removed and used as the 5′ cap and primer to initiate the synthesis of the influenza mRNA (a process sometimes termed "cap snatching"). Baloxavir marboxil was approved by the FDA in 2018 for treatment of acute uncomplicated flu within 2 days of illness onset in otherwise healthy people 12 years and older or those at high risk of developing flu-related complications. In 2020, the FDA approved an updated indication to include postexposure prevention of influenza for people ≥12 years old after contact with an infected person. The adamantanes amantadine and rimantadine were used in the past for the treatment of influenza A infection. These drugs interfere with the ion channel activity caused by the M2 protein of influenza A viruses, which is needed for viral particle uncoating after endocytosis. Widespread resistance occurred in many currently circulating influenza A viruses.

RSV INFECTION

Ribavirin is a nucleoside antimetabolite prodrug whose activation by kinases in the cell results in a 5′-triphosphate nucleotide form that inhibits RNA replication. The drug was licensed in an aerosol formula in the United States in 1986 for treatment of children with severe RSV-induced lower respiratory tract infection. The efficacy of aerosolized ribavirin therapy remains uncertain despite several clinical trials. Most centers use it infrequently, if ever, in otherwise healthy infants with severe RSV disease. Intravenous ribavirin has been used for adenovirus, hantavirus, measles virus, PIV, and influenza virus infections, although a good risk/benefit profile has not been clearly established for any of these uses.

OTHER VIRAL TARGETS

Pleconaril, an oral drug with good bioavailability for treatment of infections caused by picornaviruses, has been tested for treatment of rhinovirus infection. This drug acts by binding to a hydrophobic pocket in the VP1 protein and stabilizing the protein capsid, preventing release of viral RNA into the cell. Pleconaril reduces mucous secretions and other symptoms and is being further examined for this indication. Acyclovir and related compounds are guanine-analogue antiviral drugs used in the treatment of herpesvirus infections. HSV stomatitis in immunocompromised patients is treated with famciclovir or valacyclovir, and immunocompetent patients with severe oral disease compromising oral intake are sometimes treated with these agents. These compounds have also been used prophylactically to prevent the recurrence of outbreaks, with mixed results. Intravenous acyclovir is effective against HSV or VZV pneumonia in immunocompromised patients. Systemic therapy of CMV infection in immunocompromised patients has been studied with numerous small-molecule inhibitor drugs, including ganciclovir, valganciclovir, foscarnet, and cidofovir. The clinical utility of these drugs in the immunocompetent host is uncertain. The nucleotide analogue cidofovir also has activity against many other viruses, including adenoviruses. Intravenous cidofovir has been effective in the management of severe adenoviral infection in immunocompromised patients but may cause serious nephrotoxicity.

COMPLICATIONS: CO-INFECTIONS

Co-infections with two or more viruses can occur because of the overlap in the winter season of these viruses in temperate areas. In general, in careful studies using cell culture techniques for virus isolation, two or more viruses were isolated from respiratory secretions of otherwise healthy adults with acute respiratory illness in ~5–10% of cases. There is little evidence that more severe disease occurs during co-infections. The incidence of positive results in two molecular diagnostic tests (generally RT-PCR for these RNA viruses) is expected to be higher than that of culture because, as discussed above, molecular tests can remain positive for an extended period after shedding of infectious virus has ended.

PREVENTION

VACCINES

Numerous vaccines against influenza viruses have been licensed. In the United States, trivalent and quadrivalent inactivated intramuscular vaccines (covering H3N2, H1N1, and one or two B antigens) and a live attenuated trivalent vaccine for intranasal administration are available (although components of the live attenuated vaccine were only ~3% effective during the 2013–2016 seasons and that vaccine was not available during the 2016–2018 seasons). Vaccines are effective when the vaccine strains chosen for inclusion are highly related antigenically to the epidemic strain, but occasional antigenic mismatches cause negligible efficacy of a vaccine component. Antigenic drift caused by point mutations in the hemagglutinin (HA) and neuraminidase (NA) molecules leads to antigenic divergence, requiring the production of new vaccines each year. The segmented influenza genome allows reassortment of two viruses during co-infection of one individual or animal; sometimes the consequence is a major antigenic shift resulting in a pandemic. On average, pandemics occur every 20–30 years. There is current concern about the potential for an H5N1 or H7N9 pandemic, and experimental vaccines are being tested for these and other avian influenza viruses.

Vaccines were developed for adenovirus serotypes 4 and 7 and were approved for prevention of epidemic respiratory illness among military recruits. Essentially, these vaccines consisted of unmodified viruses given by the enteric route in capsules instead of by the respiratory route—the natural route of infection leading to disease. Inoculation by the altered route resulted in an immunizing asymptomatic infection. Most U.S. military recruits are vaccinated against adenovirus, and epidemic disease recurs in the absence of vaccination.

Two vaccines for RSV are now available (Arexy [GlaxoSmithKline] and Abrysvo [Pfizer]) based on recombinant subunit protein formulations of the virus surface fusion (F) protein. Both can be used to protect older adults. The CDC recommends a single dose of either vaccine for adults aged 60 and older who decide with their health care provider that RSV vaccination is right for them. Arexy also can be used in individuals 50 through 59 years of age who are at increased risk for lower respiratory tract disease caused by RSV. Abrysvo can be used to protect young infants through maternal immunization, leading to passive transfer of enhanced levels of serum antibodies from mother to fetus; a single dose is recommended for pregnant women between 32 and 36 weeks of pregnancy.

There are no licensed vaccines against rhinoviruses; as there is little or no cross-protection between serotypes, it will be challenging to develop a vaccine covering >100 serotypes. Efforts to develop seasonal coronavirus vaccines are in the preclinical stage. **SARS-CoV-2 vaccines are discussed in Chap. 205.**

PASSIVE PROTECTION WITH IMMUNOTHERAPY

Palivizumab, a humanized mouse monoclonal antibody to the F protein of RSV, was licensed for prevention of RSV hospitalization in high-risk infants, in half or more of whom it was effective. Experimental treatment of both immunocompetent and immunocompromised RSV-infected individuals with antibody was reported, but the efficacy of this approach has not been established. A next-generation RSV neutralizing antibody with higher potency and an extended half-life of ~3 months (nirsevimab) was approved in 2023 for all infants younger than 8 months of age born during RSV season or entering their first RSV season and some young children who are at increased risk for severe RSV disease and entering their second RSV season.

ISOLATION PROCEDURES, PERSONAL PROTECTIVE EQUIPMENT, AND HAND WASHING

Most respiratory viruses are spread by direct contact—i.e., body-surface to body-surface contact and physical transfer of microorganisms between a susceptible person and an infected person. Poor hand hygiene is probably the most common cause of contact transmission of viruses, which occurs often in family, school, and workplace settings. Transmission between health care workers and patients also takes place when hand-washing compliance is low. Fomites (objects or substances capable of carrying infectious organisms), including instruments, stethoscopes, and other objects in medical environments, can contribute to transmission. Small-particle-mediated airborne transmission can occur but is probably not the dominant mode of transmission for most respiratory viruses. Particle size affects the epidemiology of airborne pathogens. The composition and size distribution of the generated particles affect the duration of suspension of the infectious agents in the air, the distance across which they can be transported, the interval during which the virus remains infectious, and the site of deposition in the airway of a susceptible host. Direct exposure to large-particle aerosols (e.g., exposure at close range—up to 3 ft—to a cough or sneeze) causes some transmission. Particles of small size can remain suspended in the air for long periods; for instance, particles of ~1 μm can remain suspended for hours. However, in general, only a few respiratory viruses are thought to be transmitted by small-particle aerosols. Protection from transmission in health care environments can be achieved by proper implementation of and adherence to established procedures for the appropriate level of precaution.

Standard and Contact Precautions Standard precautions, the basic level of infection control that is always used in the care of all patients, reduces the risk of transmission of viruses from respiratory tract secretions and mucous membranes. Contact precautions, the second level, require a single room for the patient when possible and the use of additional personal protective equipment, including the wearing of clean, nonsterile gloves when touching a patient or coming into contact with secretions. Fluid-resistant nonsterile gowns are used to protect skin and clothing during activities in which contact with secretions is anticipated, and providers should wear each gown for the care of only one patient. A face mask is used when there is potential for direct contact with respiratory secretions. Eye protection (goggles or face shields) is worn in anticipation of potential splashing of respiratory secretions. Good hand hygiene should always follow any patient contact, including washing for 20 s with soap and warm water or cleaning with an alcohol-based hand rub. Providers should attempt to avoid the contamination of clothing and the transfer of microorganisms to other patients, surfaces, or environments.

Droplet Precautions Large-particle droplets are generated during sneezing and coughing and during the performance of some medical procedures, such as airway suctioning in critical care units or bronchoscopy. Such droplets may contain viruses, but their range is usually limited to about 3 ft. Transmission of large-particle droplets occurs when they are deposited on the nasal mucosa or conjunctivae. To prevent transmission in these settings, providers should implement droplet precautions. They should wear a face mask, such as a surgical mask, for close contact (within 3–6 ft of the patient). Patients also should wear a face mask when exiting the examination room and should avoid coming into close contact with other patients.

Airborne Precautions Airborne transmission occurs through the dissemination of airborne droplet nuclei (particles of ≤5 μm) or evaporated droplets containing viruses that can remain suspended in the air for long periods. Certain viruses that are carried by the airborne route can be inhaled by a susceptible host in the same room or over a long distance from the source patient, depending on environmental factors such as temperature and ventilation. Viruses transmitted by this route are SARS-CoV, measles virus, and VZV. Patients with these infections should be managed with personal respiratory protection and special ventilation and air handling. Providers should wear an N95 respirator selected with fit-testing, which must be repeated annually. Powered air-purifying respirators (PAPRs) are used in some cases. The patient should be housed in an airborne-infection isolation room—a negative-pressure room that has a minimum of six air exchanges per hour and exhausts through high-efficiency particulate air (HEPA) filtration or directly to the outside.

GLOBAL CONSIDERATIONS

HENDRA AND NIPAH VIRUSES

These emerging paramyxoviruses, which are grouped in their own new genus (*Henipavirus*), may not be respiratory pathogens in a conventional sense, but they probably infect humans by the respiratory route. Nipah virus is a newly recognized zoonotic virus, named after the location in Malaysia where it was first identified in 1999, and has caused significant outbreaks in Bangladesh and India. It has caused disease in humans who have had contact with infectious animals. Hendra virus (formerly called equine morbillivirus) is another closely related zoonotic paramyxovirus and was first isolated in Australia in 1994. The viruses have caused only a few localized outbreaks, but their wide host range and ability to cause high mortality raise concerns for the future. The natural host of these viruses is thought to be a certain species of fruit bat present in Australia and the Pacific. Pigs may be an intermediate host for transmission to humans in Nipah infection and horses in Hendra infection. Although the mode of transmission from animals to humans is not defined, inoculation of infected materials onto the respiratory tract probably plays a role. The clinical presentation usually appears to be an influenza-like syndrome that progresses to encephalitis, includes respiratory illness, and causes death in about half of identified cases.

BUNYAVIRIDAE: HANTAVIRUS

Intermittent outbreaks of hantavirus infection occur in South America and cause a severe lung infection: HPS. In addition, >800 cases of HPS have been reported in the United States, caused by the Sin Nombre hantavirus. The disease was first recognized during an outbreak in 1993. About one-third of recognized cases end in death. The Four Corners outbreak (at the intersection of the northwestern corner of New Mexico, the northeastern corner of Arizona, the southeastern corner of Utah, and the southwestern corner of Colorado) is well known; however, cases now have been reported in at least 32 states. Outbreaks of HPS also have occurred in South America, especially in Chile, caused by the related Andes hantavirus. Patients with HPS usually present with an influenza-like illness, including fever. Findings on physical examination are nonspecific, often consisting only of fever and elevated respiratory and heart rates. In addition to respiratory symptoms, abdominal pain is common. Diagnosis is often delayed until illness becomes severe, at which point intubation and mechanical ventilation may be required for respiratory support.

SUMMARY

Viruses are the leading causes of acute lower respiratory tract infection in most populations. Influenza virus and RSV are the most common pathogens; hMPV, PIV3, and rhinoviruses account for most other acute viral respiratory infections. Infection in otherwise healthy adults generally leads to partial immunity to these pathogens, with protection against severe lower respiratory disease. However, reinfection, with upper respiratory tract illness, is common throughout life. Special populations such as immunocompromised patients, institutionalized frail elderly patients, and patients with COPD are at highest risk for severe disease.

FURTHER READING

Beard KR et al: Treatment of influenza with neuraminidase inhibitors. Curr Opin Infect Dis 31:51, 2018.

Falsey AR et al: Bacterial complications of respiratory tract viral illness: A comprehensive evaluation. J Infect Dis 208:432, 2013.

Fry AM et al: Seasonal trends of human parainfluenza viral infections: United States, 1990–2004. Clin Infect Dis 43:1016, 2006.

Hammitt LL et al: Nirsevimab for prevention of RSV in healthy late-preterm and term infants. N Engl J Med 386:837, 2022.

Liu J-W et al: Comparison of antiviral agents for seasonal influenza outcomes in healthy adults and children: A systematic review and network meta-analysis. JAMA Netw Open 4:e2119151, 2021.

Monto AS, Cavallaro JJ: The Tecumseh study of respiratory illness. II. Patterns of occurrence of infection with respiratory pathogens, 1965–1969. Am J Epidemiol 94:280, 1971.

Papi A et al: Respiratory syncytial virus prefusion F protein vaccine in older adults. N Engl J Med 388:595, 2023.

Walsh EE et al: Efficacy and safety of a bivalent RSV prefusion F vaccine in older adults. N Engl J Med 388:1465, 2023.

Williams JV et al: Human metapneumovirus infection plays an etiologic role in acute asthma exacerbations requiring hospitalization in adults. J Infect Dis 192:1149, 2005.

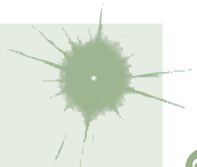

205

SARS-CoV-2 and COVID-19

James E. Crowe, Jr.

CORONAVIRUS DISEASE 2019 (COVID-19)

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2; ICTV name *Betacoronavirus pandemicum*) emerged in 2019 in an outbreak in Wuhan, China, that spread worldwide and caused a severe pandemic. SARS-CoV-2 is the cause of a respiratory disease called COVID-19. The virus is a member of the *Betacoronavirus* genus that not only includes the highly pathogenic viruses severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1; ICTV name *Betacoronavirus pandemicum*) (which caused a smaller epidemic in 2002–2003) and Middle East respiratory syndrome coronavirus (MERS-CoV or HCoV-EMC; ICTV name *Betacoronavirus cameli*) (which caused small epidemics in 2012, 2015, and 2018 with continuing sporadic cases), but also contains the common cold viruses human coronavirus OC43 (HCoV-OC43 or betacoronavirus 1; ICTV name *Betacoronavirus gravedinis*) and human coronavirus HKU1 (HCoV-HKU1; ICTV name *Betacoronavirus hongkonense)*. These are enveloped, positive-sense RNA viruses encoded by a viral RNA genome that is quite large, a single linear RNA segment of nearly 30,000 nucleotides that encodes four structural proteins, designated the S (spike), E (envelope), M (membrane), and N (nucleocapsid) proteins, and a large polyprotein that is cleaved into 16 nonstructural proteins in infected cells. The trimeric S protein is primed by the transmembrane protease serine 2 (TMPRSS2) to facilitate entry of SARS-CoV-2. SARS-CoV-2 S protein is a type 1 fusion machine that also mediates attachment using a receptor-binding domain (RBD) that binds to the human angiotensin-converting enzyme 2 (hACE2) protein receptor.

EPIDEMIOLOGY

The origins of the virus and mode of exposure to the first human cases remain unresolved, but the virus likely existed prior to the outbreak in some form in a bat reservoir. Infection was first detected in humans in 2019 in Wuhan, China; it rapidly spread by human-to-human transmission through all provinces of China and then worldwide. The World Health Organization (WHO) designated SARS-CoV-2 a Public Health Emergency of International Concern on January 30, 2020, and declared the outbreak a pandemic on March 11, 2020, leading to several years of global disruption. On May 5, 2023, the WHO declared an end to the global COVID-19 emergency. As of September 2024, the virus had caused >700 million cases and >7 million deaths worldwide, although most countries have now stopped accurate reporting of cases. The basic reproduction number (R_0) (the expected number of cases generated directly by one case in a population in which all individuals are susceptible to infection) of SARS-CoV-2 has been estimated to peak around 6, which is substantially higher than that of seasonal influenza (typically 1–2). Densely populated settings such as prisons, cruise ships, nursing homes, airplanes, and large indoor gatherings facilitate high transmission efficiency. Transmission in outdoor settings is less common. Healthcare workers and those working in dentistry have a high potential for exposure. Certain individuals may have contributed to extraordinarily high transmission events (so-called "superspreaders"). Transmission does occur in school settings, although schools have not been considered a primary driver of population transmission and children have lower rates of severe disease than adults. Spread of SARS-CoV-2 is primarily via respiratory droplets transmitted between persons in proximity when the droplets make direct contact with mucous membranes. Airborne transmission by small particle from person-to-person may occur, but airborne transmission over long distances is unlikely. Fomite transmission by contact with contaminated surfaces occurs; therefore, hand washing in environments of exposure has been advised. Large-scale frequent surface decontamination efforts have been deployed in public spaces, but the effect of these cleanings on reducing transmission is uncertain.

For replication, coronaviruses use RNA-dependent RNA polymerases that are error prone, allowing viral variants to occur frequently. Most variant residues have no effect or are deleterious to the virus. However, some variations affect transmission, disease severity, antigenicity (and thus efficacy and effectiveness of immunity induced by prior vaccination or infection), the effectiveness of therapies, the sensitivity and specificity of diagnostic tools, and the effectiveness of social and public health measures. Rapid evolution of variants has been a distinguishing feature of the COVID-19 pandemic. The WHO has established a Technical Advisory Group on SARS-CoV-2 Virus Evolution to follow and report variant evolution. The group has developed terminology concerning variants under monitoring (VUMs), variants of interest (VOIs), and variants of concern (VOCs) to draw attention to virus evolution. In 2024, the WHO launched a WHO Coronavirus Network (CoViNet) to facilitate early and accurate detection of coronaviruses and variant tracking. The WHO Technical Advisory Group on COVID-19 Vaccine Composition assesses the analyses of the likely effects of emerging VOCs on the performance of COVID-19 vaccines.

Advanced age is the principal risk factor for severe illness from COVID-19 (marked by need for hospitalization, intensive care, and mechanical ventilation). More than 95% of COVID-19 deaths occur in persons older than 45 years, and >80% of deaths occur in those older than 65 years. Preexisting social and health disparities put some groups at increased risk of illness or death from COVID-19, including persons with disabilities and many racial/ethnic minority groups. Male sex is associated with higher risk of severe disease (odds ratio, ~1.8). Most individuals who die have preexisting comorbidities. The risk of severe COVID-19 illness increases markedly with elevated body mass index (BMI). Overweight condition (BMI >25 but <30), obesity (BMI ≥30 but <40), and severe obesity (BMI ≥40) are risk factors for progressively increased severe COVID-19. Substance use, such as alcohol, opioid, or cocaine use disorder, and current or former smoking both increase risk. Pregnant women are more likely to suffer more severe illness. Most other medical conditions increase the risk of severe illness, but conditions that especially increase risk are as follows: (1) chronic lung diseases, including COPD, moderate to severe asthma, cystic fibrosis, and pulmonary hypertension interstitial lung disease; (2) cancer or cancer treatments, including hematologic malignancies, solid organ transplant, and stem cell transplant; (3) immunodeficiency, including primary immunodeficiency caused by inherited genetic defects or secondary or acquired immunodeficiency caused by prolonged use of corticosteroids, other immunosuppressive drugs, or HIV type 1 (HIV-1) infection; (4) hemoglobin blood disorders, including thalassemia or sickle cell disease; (5) cerebrovascular disease, such as stroke; (6) cognitive impairment or other neurologic conditions; (7) heart conditions, including arterial hypertension, heart failure, coronary artery disease, and cardiomyopathies; (8) obstructive sleep apnea; (9) chronic inflammatory, autoimmune, and rheumatic diseases; (10) type 1 or type 2 diabetes mellitus; (11) chronic liver disease, especially cirrhosis; and (12) genetic conditions, especially Down syndrome. A multisystem inflammatory syndrome in children (MIS-C) or pediatric multisystem inflammatory syndrome (PMIS) has been associated with COVID-19, comprising a persistent fever, involvement of multiple organ systems (including gastrointestinal, dermatologic, cardiac, renal, hematologic, and neurologic), and elevated circulating inflammatory markers. The highest-risk individuals for MIS-C in the United States are Black and Latino children age 3–12 years. A similar syndrome in adults (MIS-A) may occur rarely.

PREVENTATIVE MEASURES

Early in the epidemic, public health methods for prevention were limited mostly to nonpharmaceutical interventions (NPIs), including social distancing (staying at least 6 ft from other persons in public to avoid infection), social isolation (staying away from other persons when infected), quarantine (staying at home for 14 days after potential exposure), limiting travel, and working from home. When a local epidemic persists, prior to entering health care settings, patients often are screened for clinical signs or symptoms common in COVID-19, especially fever, respiratory symptoms (cough, dyspnea, sore throat), myalgias, and anosmia/hyposmia. Universal masking has been required in many health care settings during epidemic conditions, although effectiveness of masking in community or health settings has been difficult to assess. Most mask studies have been observational and confounded by lack of consistency in concurrent interventions, study of exposure to variants with differing R_0 values, differences in mask type (N95, KN95, FFP2, paper medical or surgical, or cloth masks), poor compliance, and other factors. Most experts conclude that masking, use of personal protective equipment, and hand washing are highly appropriate in health care settings with infected patients and may have some benefit in community settings. Most organizations and governments responsible for public health decisions have discontinued mandatory policies in favor of softer recommendations.

CLINICAL MANIFESTATIONS

The disease course varies widely, including asymptomatic infection, mild disease, moderate disease, or severe disease requiring hospitalization, oxygen therapy, intensive care, and mechanical ventilation. A substantial proportion of patients (possibly a third of those infected) are asymptomatic, but those individuals can transmit the virus to others. Most individuals with symptomatic infection have mild disease (no pneumonia). Severe disease, typically requiring hospitalization and involving pneumonia and associated manifestations (dyspnea, radiographic involvement of more than half of the lung, and/or hypoxia with oxygen saturation ≤94%), is common. Critical disease with manifestations of respiratory failure requiring mechanical ventilation, multiorgan failure, or shock occurs and requires intensive care.

DIAGNOSIS OF COVID-19

Testing for COVID-19 is recommended during periods of respiratory symptoms, or 5 days or later after exposure without active symptoms. COVID-19 tests usually use a sample taken from the nasopharynx or sometimes a saliva sample. There are two major types of approved diagnostic tests: antigen tests (for proteins of SARS-CoV-2) and molecular tests (for genetic material of SARS-CoV-2). Antigen tests are often used as screening tests and sometimes called rapid or at-home tests. Approved protein-detection antigen test devices are reliable and accurate, but they are less accurate than molecular tests, especially when testing exposed persons without symptoms. COVID-19 diagnostic tests are widely available from health care professionals, hospitals and clinics, and some pharmacies or other testing sites in the community. In the United States, at-home COVID-19 tests approved by the U.S. Food and Drug Administration (FDA) are available for free or for purchase. Molecular tests are nucleic acid amplification tests (NAATs), most often performed using reverse transcription polymerase chain reaction (RT-PCR) to convert the viral RNA genome to DNA copies (cDNA) and amplify the copy number for enhanced sensitivity. PCR tests are not rapid and cannot be performed at home, so they are typically performed in professional health care settings and processed in a reference laboratory using specialized equipment. Nasopharyngeal swabs are used mostly commonly, while saliva testing also has been implemented, especially in large-scale population screening efforts.

Other more general laboratory tests have been used during medical management of severe or critical illness. Most immune profiling tests reveal widespread abnormalities consistent with systemic disease including lymphopenia and thrombocytopenia; elevated inflammatory markers, such as interleukin 6 (IL-6), tumor necrosis factor α, ferritin, and C-reactive protein; elevated liver enzymes and lactate dehydrogenase; elevated markers of acute kidney injury; elevated D-dimer and prothrombin time; and elevated troponin and creatine phosphokinase. Research-grade tests show that beneficial components of the adaptive immune response, including antibodies and T cells, also arise during the first 1–2 weeks after exposure. Chest radiographs may exhibit abnormal findings such as consolidation and ground-glass opacities that are distributed bilaterally, especially in the lower lung regions, but may also be normal despite respiratory compromise. Chest computed tomography (CT) has features (ground-glass opacifications with or without mixed consolidation, pleural thickening, interlobular septal thickening, and air bronchograms) that can be systematically interpreted as typical, indeterminate, or atypical for COVID-19. Chest CT may be more sensitive than radiographs, but CT should be used principally for medical management of respiratory disease, not as a primary diagnostic tool for COVID-19. Lung ultrasound also has been used to image the lungs to detect some COVID-19 abnormalities.

■ CLINICAL COURSE

The onset of disease manifests typically within 4–5 days after exposure and nearly always within 14 days. Symptoms include cough, fever, myalgia, headache, dyspnea, sore throat, and gastrointestinal symptoms of nausea, vomiting, or diarrhea. Sudden onset of dysgeusia and anosmia (loss of taste and smell) occurs in a substantial number of cases, which often resolves in weeks to months. Diverse dermatologic findings occur in patients with COVID-19. General decline of health status, including onset or worsening of dementia, can occur in older individuals, especially those with cognitive impairment. Mental health consequences of the acute disease, isolation measures, and medical management regimens are common, including fatigue, depression, general and social anxiety, sleep disturbances, cognitive deficits, post-traumatic symptoms, and substance abuse disorder. Long COVID is a chronic condition that occurs after SARS-CoV-2 infection and is present for at least 3 months with a wide range of symptoms or conditions that may improve, worsen, or be ongoing. The symptoms are varied and can include neuropsychiatric disorders and pain syndromes in addition to respiratory and metabolic changes. People with long COVID may experience many symptoms related to alterations in mental health and decreased brain function and experience increased suicidal ideation, which may increase suicide risk.

■ COMPLICATIONS

Severe complications of infection can occur. The major complication in patients with severe disease is acute respiratory distress syndrome requiring oxygen therapy and mechanical ventilation. Disseminated intravascular coagulation is a complication in severe disease. Thromboembolic complications are common in severe disease, mostly occurring as venous thromboembolism, including pulmonary embolism or deep vein thrombosis. Events stemming from arterial thrombosis, including acute stroke or ischemia of the limbs, are reported. Cardiac complications manifest as heart failure, myocardial injury, or arrhythmias and cardiovascular syndromes, especially shock. Acute kidney injury requiring dialysis can occur. Encephalopathy occurs in critically ill patients, and delirium in the intensive care unit setting reduces overall survival. Other neurologic complications including seizures, ataxia, or motor or sensory deficits have been reported. Those with COVID-19 disease and laboratory markers of excessive inflammatory response can exhibit a pattern of persistent fever and multiorgan disease with high risk of fatal outcome. An excessive proinflammatory host response to SARS-CoV-2 infection likely contributes directly to pulmonary pathology and severity of COVID-19. Manifestations typically mediated by autoantibodies have been reported. Disease is usually caused by direct viral pathogenesis in tissues or the associated immune response, but secondary bacterial or fungal infections do occur, usually as bacteremia or respiratory infections.

■ GENERAL MEDICAL MANAGEMENT

Medical management of COVID-19 is focused on severe respiratory illness and systemic disease manifestations. As bacterial infection is an uncommon complication of COVID-19, antibiotics are not generally indicated, but when the diagnosis is uncertain, empiric antibiotic regimens for community-acquired or health care–associated pneumonia should be considered. Since there is such a substantial risk of thromboembolic complications, many experts recommend pharmacologic prophylaxis of venous thromboembolism for all hospitalized patients with COVID-19. Nonsteroidal anti-inflammatory drugs (NSAIDs) are often used as antipyretic agents, but questions have been raised about a possible association between NSAID use and worse outcomes with COVID-19; when possible, the preferred antipyretic agent is acetaminophen. Immunosuppressed individuals are at higher risk of severe illness or death; therefore, on a case-by-case basis, providers should decide whether to continue immunomodulatory agents such as steroids or other immunosuppressive drugs that were indicated for preexisting conditions prior to onset of COVID-19. Generally, experts agree that the best course usually is to continue common preexisting medications of aspirin, statins, and angiotensin-converting enzyme inhibitors or angiotensin receptor blockers.

The time to recovery from COVID-19 is affected by the severity of disease, the individual's preexisting comorbidities, and age. Generally, symptomatic infection is an acute syndrome that resolves in 2 weeks in ~80% of persons, especially following mild or moderate disease. Individuals with severe disease often require longer for recovery, on the order of several months. However, a subset of individuals with infection progress to a recurring or persisting pattern of symptoms, most commonly including fatigue, cognitive deficits, cough, dyspnea, or chest pain. Those with severe acute pulmonary or cardiac injury may have persisting respiratory or cardiac impairment. Diverse long-term adverse mental health consequences of infection are common, and the public health measures used to manage the pandemic also have led to social isolation with adverse mental health consequences.

■ SPECIFIC TREATMENTS

The approach to specific treatment of COVID-19 of varying levels of severity continues to evolve, especially because emerging SARS-CoV-2 variants exhibit differing properties of transmission efficiency and capacity to cause disease. Most options fall into the categories of small-molecule viral inhibitors, polyclonal or monoclonal antibodies, and immunomodulators. The recommended use of these medical countermeasures can be grouped under categories of outpatient management of mild disease or hospital management of severe disease. We recommend consulting up-to-date recommendations from groups authorized to provide expert or governmental guidelines, including the National Institutes of Health (NIH) COVID-19 Treatment Guidelines Panel in the United States or the WHO for international settings (see "Further Reading," below). Many but not all of the guidelines from such groups are harmonized. During the early years of the pandemic, numerous countermeasures were *authorized* (i.e., not necessarily fully *approved*) under the Emergency Use Authorization (EUA) authority that allows the FDA to facilitate the availability and use of medical countermeasures prior to full licensure when the secretary of the Department of Health and Human Services declares that an EUA is appropriate for a public health emergency. However, the WHO ended Public Health Emergency of International Concern status for COVID on May 5, 2023, and now proper drug use once again typically requires FDA approval.

Antiviral drugs (including small-molecule inhibitors and polyclonal or monoclonal antibodies) have the most potential to alter the clinical course early in infection, since they may reduce the peak titer of virus, a parameter that is likely correlated with severity of disease. Later in the clinical course, anti-inflammatory medications may be of more benefit since the pathogenesis of disease is driven increasingly more over time by tissue inflammation and systemic inflammatory responses than by direct viral cytopathic effect.

OUTPATIENT MANAGEMENT

Most patients with COVID-19 can be managed as outpatients. Individuals who are infected but have mild disease can be treated with supportive care only. The decision to use antiviral drugs is driven principally by medical risk factors, but it should also consider available resources, the clinical situation, and the patient's social circumstances. Shared clinical decision-making should be used regarding COVID-19–specific therapy to accommodate the individual's values and preferences in addition to medical risk factors. Risk-benefit ratios are difficult to assess. Reducing viral load has the potential to reduce acute disease and complications, but COVID medications also have the potential to cause adverse effects or drug–drug interactions, and they carry the risk of causing a rebound COVID-19 phenomenon that then may require prolongation of isolation. COVID-19–specific therapy is not recommended for individuals who have asymptomatic SARS-CoV-2 infection. Symptomatic patients who are most likely to benefit from specific antiviral treatment are individuals who are older, unvaccinated, with multiple medical comorbidities, and/or with immunocompromised conditions. Others who lack these risks might benefit to some degree, but since their risk of hospitalization or death is low the impetus to risk side effects is much less compelling. For symptomatic adults who are at increased risk, many experts recommend treating immunocompetent unvaccinated adults >50 years of age, all adults ≥65 years of age regardless of vaccination status, immunocompetent adults of any age who have multiple medical risk factors for progression to severe disease, and immunocompromised adults of any age. As of late 2024, the principal antiviral drug for use in outpatients is nirmatrelvir-ritonavir, which should be administered as soon as possible and within 5 days after symptom onset. The typical dose is nirmatrelvir 300 mg and ritonavir 100 mg orally twice daily for 5 days, but the dose should be adjusted for reduced renal function and should be avoided in patients with severe renal dysfunction. Drug interactions are possible, so prior to prescribing, providers should review all medications and consider potential drug interactions in consultation with a pharmacist or using online tools. In a minority of cases, viral shedding occurs again after an initial improvement, with or without mild symptoms, requiring an extension of the period of isolation. Patients on treatment should be counseled to return for reevaluation if the patients experience worsening or new-onset mental status changes such as confusion or worsening respiratory symptoms such as dyspnea. Patients should be counseled on an appropriate period of self-isolation to achieve infection control, per current recommendations of the U.S. Centers for Disease Control and Prevention (CDC).

If nirmatrelvir-ritonavir is not available or suitable, alternatives that are more difficult to administer are available. Remdesivir is approved, but administration requires three IV doses over 3 days. The processing in vivo of this broad-spectrum antiviral prodrug causes intracellular delivery of an active chemical that serves as a ribonucleotide analogue inhibitor of SARS-CoV-2 RNA polymerase. Molnupiravir is an antiviral small-molecule drug that is metabolized into a ribonucleoside analogue that inhibits viral reproduction by promoting widespread mutations in the replication of viral RNA by RNA-directed RNA polymerase. In December 2021, the FDA granted an EUA to molnupiravir for use in certain populations where other treatments are not feasible, but with a black box warning about its mutagenic properties that limit its use in patients with pregnancy potential. Virus-neutralizing antibodies delivered as monoclonal human antibodies or antibody combinations or as constituents of high-titer convalescent plasma also have been authorized for postexposure prophylaxis or treatment of outpatient COVID-19, but new variants arise and reduce their effectiveness and often lead to withdrawal of authorization.

Many other therapies have been used without specific authorization or approval and are of uncertain benefit. Antibiotics are not recommended for routine treatment of COVID-19, as bacterial complications are unusual. Dexamethasone, prednisone, and other corticosteroids have been used in the setting of COVID-19 infection accompanied by acute exacerbation of chronic obstructive pulmonary disease or asthma.

INPATIENT MANAGEMENT

Indications for Hospitalization COVID-19 patients with symptoms suggestive of more severe disease—especially moderate to severe dyspnea, cyanosis, change in mental status or chest pain, reduced urine output or anuria, with or without oxygen saturation ≤94%—should be evaluated in an emergency department. The NIH COVID-19 Treatment Guidelines Panel recommends hospitalization for patients with oxygen saturation of <94% on room air, respiratory rate of >30 breaths/min, $\text{PaO}_2/\text{FiO}_2$ <300 mmHg, or lung infiltrates on chest radiograph occupying >50% of lung fields. The decision to hospitalize varies regionally and due to differences in available resources and social risk factors.

Management in Hospital Upon hospitalization, the initial assessment should seek to define medical comorbidities or immunocompromising factors and to identify dysfunction in critical organ systems. Fever should be controlled, preferably with acetaminophen. The use of NSAIDs has not been aligned with poor COVID-19 outcomes, but most experts suggest avoiding their use or limiting them to lower effective doses. If the patient uses common chronic medications, those generally should be continued, including aspirin (unless there is a clear bleeding risk), statins, and angiotensin-converting enzyme inhibitors or angiotensin receptor blockers. Hospitalized patients should receive therapeutic-dose anticoagulation to prevent venous thromboembolism. For patients without any risk factors, additional care is primarily supportive. Patients with risk factors for severe disease typically are treated with remdesivir. Oxygen should be used to support oxygen saturation. Patients receiving low-flow supplemental oxygen are usually treated with remdesivir (for antiviral effect) and low-dose dexamethasone (for anti-inflammatory treatment). Patients treated with high-dose glucocorticoids should be monitored for common adverse effects, especially hyperglycemia and increased risk of co-infection. If the oxygen requirement escalates during the first 4 days of hospitalization despite those treatments, with severe disease and elevated inflammatory markers on laboratory testing, the Janus kinase inhibitor baricitinib or the monoclonal IL-6–blocking antibody tocilizumab can be used. Janus kinase inhibitors typically are used for treatment of rheumatoid arthritis because of their known immunomodulatory effects, which probably also improve inflammation during COVID-19, but baricitinib may also mediate some direct antiviral effects by interfering with viral entry into cells. Alternatively, tocilizumab can be used for immunomodulatory effect. This biologic drug is a humanized monoclonal antibody against the IL-6 receptor that blocks the effect of the cytokine IL-6, which plays an important role in pathogenesis. These immunomodulatory therapies are also usually used for patients admitted to an intensive care unit to receive high-flow supplemental oxygen or noninvasive ventilation, mechanical ventilation, or extracorporeal membrane oxygenation. Abatacept or infliximab are alternate immunomodulatory agents that can be used if the first-line immunomodulators are not available.

ANTIBODY-BASED THERAPIES

Passive immunization with SARS-CoV-2 antibodies to achieve antiviral immunity or therapeutic effect has been tested using convalescent plasma (blood plasma taken from persons who have recovered from COVID-19) or human monoclonal antibodies (mAbs). The typical overall composite titer of SARS-CoV-2–neutralizing antibodies in convalescent plasma following a single primary infection is moderately low, limiting its effectiveness and reproducibility. Human mAbs are recombinant proteins made in the laboratory based on the genes encoding an antibody obtained typically from a single SARS-CoV-2–specific B cell isolated from the peripheral blood of a convalescent individual. Numerous mAb products obtained EUA for use in outpatients with laboratory-confirmed mild to moderate SARS-CoV-2 infection who were at high risk for progressing to severe disease and/or hospitalization, although these EUAs for therapy are now revoked because new variants have escaped their activity.

TREATMENT OF COMPLICATIONS

Severe and complicated COVID-19 is a complex disease that can affect the respiratory and cardiovascular systems and may involve all other major organs. Disease severity and mortality are often determined by complications due to the systemic effects of a hyperinflammatory and hypercoagulable state. Acute respiratory distress syndrome with severe hypoxia and viral sepsis are common, with acute cardiac injury, arrhythmias, thromboembolic events, acute kidney injury, cytokine storm, oxidative stress, and thrombosis and shock. Each of these organ system diseases must be managed in the context of ongoing viral infection. Bacterial superinfection of COVID-19 probably occurs, but the incidence is uncertain. There are insufficient data to recommend empiric broad-spectrum antimicrobial therapy in the absence of another indication, although some experts routinely administer broad-spectrum antibiotics as empiric therapy for bacterial pneumonia to all patients with COVID-19 and moderate or severe hypoxemia. Ideally, providers initiating empiric therapy should attempt to deescalate or stop antibiotics if there is no ongoing evidence of bacterial infection. Several EUAs have been issued for medical management of complications during COVID-19, including replacement solutions for continuous renal replacement therapy and drugs for sedation via continuous infusion in intensive care. Anticoagulation in the face of COVID-19–associated thromboembolic events is an especially complex situation and requires expert consultation.

ANTIBODY-BASED PREVENTION

Antibodies of the IgG isotype typically have an average serum half-life of about 21 days, and thus passive transfer of immune serum or antibodies confers protection of modest duration. During COVID-19, however, proof of principle for the use of a long-acting antibody (LAAB) combination for longer-term prevention of disease was established with the emergency use authorization of the two-antibody combination drug tixagevimab co-packaged with cilgavimab. These antibodies were engineered with variant residues in the Fc portion of the antibody that altered the kinetics of binding and retention to the Fc receptor neonatal (FcRn) that regulates antibody half-life in vivo. This first-in-human authorization of an LAAB enabled protection of individuals who could not be effectively vaccinated, such as the immunocompromised. The FDA terminated the EUA in January 2023 due to the widespread emergence of escape variants in circulation. However, this program established a paradigm for vaccine surrogate use of LAABs in high-risk individuals.

In 2024, the FDA granted an EUA for a second LAAB, pemivibart, which is currently limited to use when the combined national frequency of variants with substantially reduced susceptibility to pemivibart is ≤90%. Pemivibart requires IV administration because of the large (4.5 g) dose for pre-exposure prophylaxis of COVID-19 in adults and adolescents (≥12 years of age weighing ≥40 kg) who have moderate to severe immune compromise due to certain medical conditions or receipt of certain immunosuppressive medications or treatments and are unlikely to mount an adequate immune response to COVID-19 vaccination.

PREVENTION BY IMMUNOPROPHYLAXIS

Four COVID vaccines—two based on mRNA (Pfizer-BioNTech and Moderna), one on a subunit protein approach (Novavax), and one on viral vectored technology (Janssen)—have been authorized or approved in the United States. The EUA for the viral vector vaccine based on a human adenovirus that was modified to contain the gene for making the SARS-CoV-2 spike protein was withdrawn in June 2023 due to association with thrombosis with thrombocytopenia syndrome (TTS). Currently, the mRNA vaccines are available to everyone 6 months and older, and the subunit protein vaccine is for anyone 12 years and older. The principal goal of vaccination is preventing severe COVID-19 by reducing critical illness or death, hospitalization, or medically attended illness in emergency departments and urgent care visits for COVID-19. Prevention of all symptomatic SARS-CoV-2 infections is beneficial and often measured as a secondary endpoint in clinical trials. Preliminary data also suggest that vaccinated persons with COVID are less likely than unvaccinated persons to suffer long COVID syndrome.

COVID vaccine designs were developed in record time in 2020 because of years of prior research on candidate vaccines for related coronaviruses, including SARS-CoV-1 and MERS-CoV. The viral surface spike (S) protein that is the target of protective neutralizing antibodies was engineered from the native "metastable" membrane-bound form to a stabilized "prefusion" conformation of soluble protein. This protein then was encoded by recombinant mRNA formulated in lipid nanoparticle delivery components. Phase 3 clinical data for the original mRNA vaccines delivered in a two-dose primary series showed 95% efficacy for preventing symptomatic COVID, a stunning outcome for this rapid development program. The durability of immunity proved relatively short in subsequent real-world effectiveness studies in adults, which showed that the protection from the mRNA waned over time. In addition, antigenic variant viruses continue to emerge on a regular basis, further eroding the level of vaccine-induced immunity. In response, manufacturers and regulators have supported a program of progressively updating mRNA vaccines to boost levels of immunity and broaden coverage to recognize newer variants. The original mRNA vaccines from both Pfizer and Moderna protected against the original SARS-CoV-2 virus. The mRNAs encoding spike protein antigens have been replaced three times since, with vaccines targeting different variants of the Omicron virus strain. In 2022, vaccines encoding the S protein from both the original strain and the Omicron variants designated BA.4 and BA.5 ("bivalent" vaccines) were deployed. In 2022, the XBB subvariant of Omicron emerged by recombination of two BA.2 sublineage viruses, leading to the replacement of vaccines in 2023 with a monovalent vaccine based on XBB. In 2024, the mRNA vaccines were updated to protect against KP.2, a further 2024 sublineage known as a FLiRT variant because it contains a phenylalanine (F) to leucine (L) mutation and an arginine (R) to threonine (T) mutation in the S protein. Even as these updated vaccines were rapidly released, new variants continued to emerge.

The protein vaccine contains a recombinant form of S protein instead of mRNA encoding S. The gene encoding the S protein is introduced into an insect virus vector (baculovirus) and expressed in insect cells, and then purified and formulated as a nonreplicating nanoparticle that displays multiple copies of S. The protein nanoparticle construct is then formulated with a saponin-based adjuvant named Matrix-M to increase immunogenicity. The protein vaccine demonstrated 100% effectiveness against moderate and severe disease in phase 3 trial results published in 2021. The vaccine was updated for 2024–2025 with antigen based on a JN.1 variant (that emerged prior to the KP.2 variants on which 2024 mRNA vaccines are based), although the manufacturer reported preclinical studies in which the JN.1 nanoparticle vaccine induced cross-reactive antibodies to the later KP.2 and KP.3 variants.

FURTHER READING

Centers for Disease Control and Prevention: COVID-19. Available at *https://www.cdc.gov/covid/*. Accessed September 10, 2024.

Centers for Disease Control and Prevention: Infection control guidance: SARS-CoV-2. Available at *https://www.cdc.gov/covid/hcp/infection-control/*. Accessed September 10, 2024.

Infectious Diseases Society of America: IDSA guidelines on the treatment and management of patients with COVID-19. Available at *https://www.idsociety.org/practice-guideline/covid-19-guideline-treatment-and-management/*. Accessed September 10, 2024.

National Institutes of Health: Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. Available at *https://www.ncbi.nlm.nih.gov/books/NBK570371/*. Accessed September 10, 2024.

World Health Organization: Therapeutics and COVID-19: Living guideline. Available at *https://www.who.int/publications/i/item/WHO-2019-nCoV-therapeutics-2023.2*. Accessed December 21, 2024.

206 Influenza

Kathleen M. Neuzil

DEFINITION

The term influenza represents both a clinically defined respiratory illness accompanied by systemic symptoms of fever, malaise, and myalgia and the name of the orthomyxoviruses that cause this syndrome. Although this term is sometimes used more generally to denote any viral respiratory illness, many features distinguish influenza from these other illnesses, most particularly its systemic symptoms, its propensity to cause sharply peaked winter epidemics in temperate climates, and its capacity to spread rapidly among close contacts. The morbidity and mortality associated with influenza epidemics are documented closely in the United States by the Centers for Disease Control and Prevention (CDC), which records clinical cases of influenza-like illness, cases of virologically documented influenza, and excess deaths due to pneumonia and influenza combined.

ETIOLOGIC AGENTS

Three influenza viruses occur in humans: A, B, and C. These viruses are irregularly circular in shape, measure 80–120 nm in diameter, and have a lipid envelope and prominent spikes that are formed by the two surface glycoproteins, hemagglutinin (H) and neuraminidase (N) **(Fig. 206-1)**. The hemagglutinin functions as the viral attachment protein, binding to sialic acid receptors on the cells that line the superficial epithelium of the respiratory tract. The neuraminidase cleaves the virus from the cell membrane to facilitate its release from the cell and prevents self-aggregation of viruses. Influenza A viruses have eight single-strand negative-sense RNA segments in their genomes that encode hemagglutinin and neuraminidase as well as internal genes, including polymerase, matrix, nucleoprotein, and nonstructural genes. The segmented nature of the genome allows gene reassortment; an analogy for reassortment is the shuffling of a deck of cards. Reassortment takes place when a single cell is infected with two different strains of influenza.

Among the influenza viruses, the A viruses are of paramount importance for several reasons: (1) the plasticity of their genomes, which enables them to react to the prevailing immunity in the community by modifying their immunogenic epitopes, particularly on the hemagglutinin surface protein (*antigenic drift*); (2) the segmentation of their genomes, which allows genes coding both surface and internal proteins to be reassorted between influenza A variants (*antigenic shift*); and (3) their extensive mammalian and avian reservoirs, in which multiple variants with distinct hemagglutinin and neuraminidase genes lie in wait. As a result of all of these factors, influenza A virus has the ability, particularly after an antigenic shift, to cause a worldwide epidemic (*pandemic*). The most severe influenza A pandemic in modern history took place in 1918; ~50 million deaths were attributed to the culpable influenza A H1N1 virus in the years surrounding 1918.

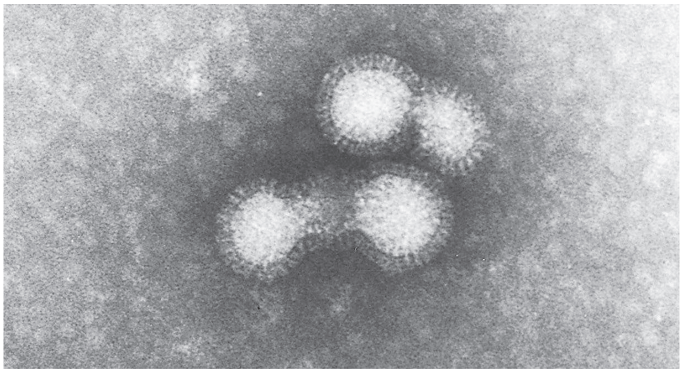

FIGURE 206-1 An electron micrograph of influenza A virus (×40,000). *(From YZ Cohen, R Dolin: Influenza, in Harrison's Principles of Internal Medicine, 19th ed. DL Kasper et al [eds]. New York: McGraw-Hill, 2015, p 1209.)*

The influenza A viruses are further classified by their surface glycoproteins (H and N), the geographic location of their isolation, their sequential number among isolated viruses, and their year of isolation. Thus, for the 2024–2025 season, U.S.-licensed egg-based influenza vaccines will contain hemagglutinin derived from an influenza A/Victoria/4897/2022 (H1N1) pdm09-like virus, an influenza A/Thailand/8/2022 (H3N2)-like virus, and an influenza B/Austria/1359417/2021 (B/Victoria lineage)-like virus. U.S. cell culture–based inactivated and recombinant influenza vaccines will contain HA derived from an influenza A/Wisconsin/67/2022 (H1N1) pdm09-like virus, an influenza A/Massachusetts/18/2022 (H3N2)-like virus, and an influenza B/Austria/1359417/2021 (B/Victoria lineage)-like virus.

EPIDEMIOLOGY

Influenza virus causes outbreaks during the cooler months of the year and thus has a mirror-image season in the antipodes compared with that in the Northern Hemisphere. The circulation of strains in the Southern Hemisphere has some predictive value for vaccine composition in the Northern Hemisphere, and vice versa. This information is important as the degree of antigenic drift is one determinate of vaccine efficacy. Vaccine composition typically must change in at least one component yearly in anticipation of the predicted circulating strains.

A typical outbreak begins in early winter and lasts 4–5 weeks in a given community, although its impact on the country as a whole will be of considerably longer duration. When influenza activity exceeds a predetermined baseline, an influenza outbreak is classified as an *epidemic*. Influenza's impact is reflected in increased school and work absenteeism, increased visits to emergency departments and primary care physicians, and increased hospitalizations, particularly of older persons and individuals with underlying cardiopulmonary disease. The impact often is most easily recognized in the pediatric population, whose school absenteeism quickly peaks.

Influenza's global spread and causative strain(s) in a given year are well documented by the surveillance networks of the World Health Organization (WHO) and the CDC. The severity of an epidemic depends on the transmissibility and virulence of the viral strain, the susceptibility of the population, the adaptation of the virus to its human host, and the degree of antigenic match to the recommended vaccine. None of these parameters is totally predictable for influenza A.

Influenza is largely spread by small- and large-particle droplets; however, emerging data support a role for aerosol transmission. Transmission is likely modulated by temperature and humidity. Spread is facilitated by the coughing and sneezing that accompany the illness. Within families, the illness is often introduced by a preschool or school-aged child. In the United States, influenza virus circulation in the first quarter of 2020 declined sharply within 2 weeks of the COVID-19 emergency declaration and widespread implementation of community mitigation measures and travel restrictions. The decline occurred in other Northern Hemisphere countries and the tropics, and in 2020, Southern Hemisphere temperate climates had virtually no influenza circulation. Influenza activity remained at low levels during the 2020–2021 Northern Hemisphere season, and increased in seasons thereafter. While changes in health care–seeking behavior and testing priorities during the pandemic may have contributed, such declines in influenza detection were noted even in areas with continued or increased testing, implicating community mitigation measures as the most likely reason.

Influenza A Viruses When a major shift in the hemagglutinin and/or the neuraminidase occurs, with introduction of a new serotype from an animal or avian reservoir, an influenza A strain has the potential to cause a pandemic. In modern influenza history, such shifts occurred in 1918 (H1N1), 1957 (H2N2), 1968 (H3N2), 1977 (H1N1), and 2009 (H1N1pdm) **(Table 206-1)**. On the basis of analysis of serum antibody profiles in the elderly, epidemics that took place in the 1890s have been attributed to H3N2 and H2N2 viruses. Epidemics typical of influenza have been documented throughout recorded history.

TABLE 206-1 Emergence of Antigenic Subtypes of Influenza A Virus Associated with Pandemic or Epidemic Disease

YEARS	SUBTYPE	EXTENT OF OUTBREAK
1889–1890	H2N8[a]	Severe pandemic
1900–1903	H3N8[a]	Moderate epidemic
1918–1919	H1N1[b] (formerly HswN1)	Severe pandemic
1933–1935	H1N1[b] (formerly H0N1)	Mild epidemic
1946–1947	H1N1	Mild epidemic
1957–1958	H2N2	Severe pandemic
1968–1969	H3N2	Moderate pandemic
1977–1978[c]	H1N1	Mild pandemic
2009–2010[d]	H1N1	Pandemic

[a]As determined by retrospective serologic survey of individuals alive during those years ("seroarchaeology"). [b]Hemagglutinins formerly designated as Hsw and H0 are now classified as variants of H1. [c]From this time until 2016–2017, viruses of the H1N1 and H3N2 subtypes circulated in alternating years or concurrently. [d]A novel influenza A/H1N1 virus emerged to cause this pandemic.

Source: Adapted from YZ Cohen, R Dolin. Influenza. In: Kasper DL, et al, eds. *Harrison's Principles of Internal Medicine*, 19th ed. New York, McGraw-Hill, 2015, p. 1209.

In some epidemics, a younger age group proves especially susceptible. This is the case with current H1N1 epidemics, where individuals born before 1968 had likely been exposed to related viral strains and thus were relatively protected against the current strain. The 1918 epidemic was striking in this regard: the most severely infected individuals were infants and previously healthy young adults—the latter being a group not typically found to have high influenza mortality **(Fig. 206-2)**. The 1918 epidemic increased all-cause mortality and led to more deaths than all military losses in World War I. Despite the attention paid to the risk and impact of pandemic disease, it is generally appreciated that—with the exception of 1918—cumulatively more illness occurs during yearly epidemics combined than in pandemics. All of the annual influenza A epidemics in the past 50 years have been caused by H1N1 and/or H3N2 strains. H2N2 strains circulated between 1957 and 1968, and H1N1 strains circulated prior to that, including in 1918.

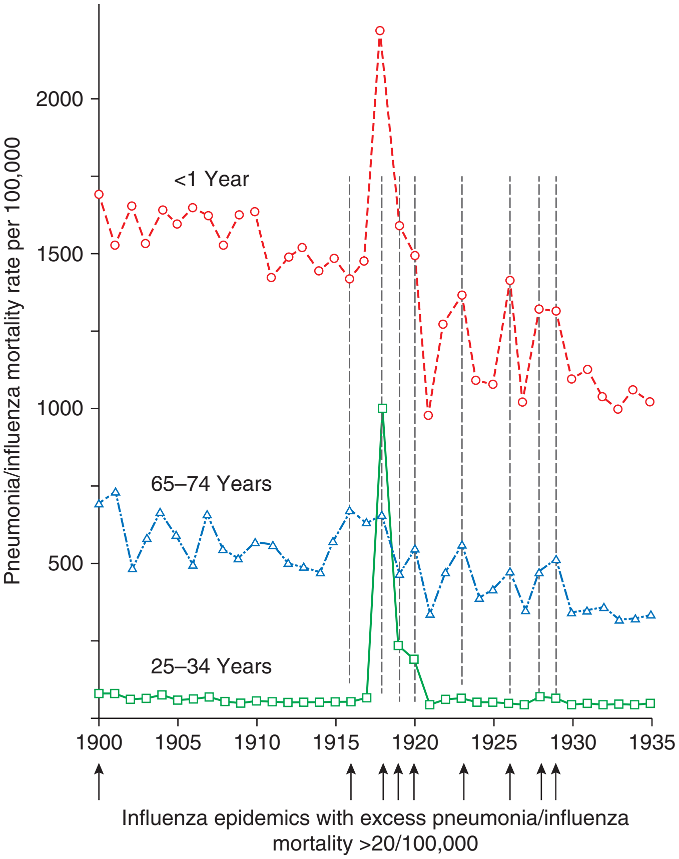

FIGURE 206-2 Excess pneumonia/influenza deaths in 1900–1953, demonstrating the dramatic peaks of deaths among young infants and young adults (25–34 years of age) in 1918. *(Data are from public health records collated by PF Wright.)*

Avian Influenza Viruses Wild birds are considered the natural hosts for influenza viruses, and potential pandemic viruses continue to emerge with higher-numbered hemagglutinins (e.g., H5, H6, H7, H8, H9) reflecting some of the 18 distinct H subtypes in avian reservoirs. Bird migration contributes to rapid global spread. Highly pathogenic influenza H5 and H7, in particular, have been associated with widespread outbreaks in poultry, human infections through contact with infected birds, and limited human-to-human transmission. Most human infections have occurred in individuals who have had direct contact with domesticated birds or who have visited live-bird markets.

Some avian strains—notably H5 strains—are highly pathogenic in humans, as was the 1918 strain. The reasons for the high pathogenicity of certain strains are not entirely clear. After the sequencing of the 1918 virus recovered from the lungs of bodies buried in the Arctic permafrost, the virus was genetically reconstructed under carefully controlled isolation conditions. In animal studies of this viable 1918 virus, both the hemagglutinin and the ribonucleoprotein contributed to high levels of replication accompanied by an abnormally enhanced innate immune response characterized by proinflammatory cytokines. Perhaps this "cytokine storm" is the best explanation for the enhanced illness in young, immunologically vigorous individuals during the 1918 pandemic. Sequencing demonstrated that the 1918 virus was of avian origin. Although the 1918 virus was first identified in military camps in the United States, its impact cannot be attributed to the disruption of war—the illness was well documented in countries such as Iceland that were not directly involved in World War I.

The same concerns about a "cytokine storm" have been raised about the H5N1 viruses that first emerged in Hong Kong in 1996. These viruses exhibited high pathogenicity in individuals who had direct contact with domestic fowl, with mortality rates close to 50%, but also displayed poor human-to-human transmissibility. Pathogenicity appears to be a function not just of the viruses' surface proteins, but also of an optimal gene constellation including all eight segmented influenza genes. However, unlike the 1918 strain, the H5N1 viruses have, to date, caused only sporadic disease, as have other limited clusters of a highly pathogenic H7N9 virus.

Sporadic H5 avian infections have been reported in many mammals, including sea lions, sea elephants, foxes, goats, and zoo animals. In 2024, a multistate outbreak of highly pathogenic H5 avian influenza in dairy cows was first reported, with rare human infections from animal exposures. It is unclear why higher-numbered avian hemagglutinin strains have not acquired the degree of transmissibility necessary to cause pandemic disease.

Swine Influenza Viruses Swine play an important role in interspecies transmission of influenza. It is postulated that epithelial cells in the swine respiratory tract may play a specific role as a "mixing vessel," allowing the reassortment of genes from avian and human sources and thereby permitting the transmission of avian viruses to humans. The nature of the sialic acid receptors for influenza virus hemagglutinin partially accounts for host preference. Humans have largely α-2,6-galactose receptors, while birds have α-2,3-galactose receptors. Swine have both types of receptors on their respiratory epithelial cells—hence their postulated role in facilitating reassortment and host adaptation of avian strains to growth in humans. The swine origin 2009 H1N1pdm strain was a reassorted virus with gene segment origins from avian, human, and swine hosts.

Influenza B and C Viruses The influenza B viruses are more genetically stable than the influenza A viruses and are mainly associated with human infection. Two lineages of influenza B have circulated for the past 40 years (B/Yamagata-like and B/Victoria-like viruses), and it has proven difficult to predict which strain will be dominant in a given year. Co-circulation of both B lineages—Victoria and Yamagata—began in 2011. This led to the incorporation of representatives of both influenza A lineages plus influenza A/H1N1 and H3N2

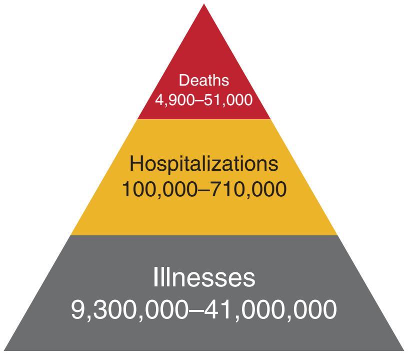

FIGURE 206-3 Pyramid of impact of influenza illness. Estimated range of annual burden of influenza in the United States from 2010–2023. *(From https://www.cdc.gov/flu-burden/php/about/index.html?CDC_AAref_Val=https://www.cdc.gov/flu/about/burden/index.)*

viruses into quadrivalent vaccines, first marketed in the United States in 2013. However, since March 2020, B/Yamagata influenza viruses have not circulated in the population. Thus, influenza vaccines in the United States will revert to a trivalent vaccine for the 2024–2025 season, containing only a B/Victoria-like virus.

Influenza C viruses cause intermittent mild disease. The clinical information about this virus is limited because of the small number of isolated viruses compared to influenza A or B viruses.

Influenza-Associated Morbidity and Mortality Influenza virus infects people of all ages and causes mild to severe illness, and even death in some cases. The impact of influenza is highly variable from year to year and can be depicted as a pyramid of illnesses, medical visits, hospitalizations, and deaths **(Fig. 206-3)**. Infection rates are highest among children, with complications and hospitalizations from seasonal influenza being greatest among certain high-risk groups during most epidemics. These groups are assigned the highest priority for vaccination and other preventive and therapeutic measures. Their caregivers and close contacts are also prioritized targets of interventions **(Table 206-2)**.

Mortality attributable to influenza, reported as excess over the anticipated sine-wave curve of pneumonia and influenza deaths during the year, varied between 4900 and 51,000 deaths annually from 2010 to 2023. The dramatic effect of the COVID-19 pandemic on excess pneumonia and influenza mortality data is evident from the comparison of 2020 data with data from the prior three seasons **(Fig. 206-4)**. Due to this outsized effect of COVID-19, the method for calculating mortality due to influenza changed starting with the 2023–2024 season and is now restricted to percentage of deaths with influenza listed on the death certificate. In contrast to mortality surveillance in adults, influenza-associated pediatric mortality is based on laboratory confirmation. Upon normalization of influenza circulation after the COVID-19 pandemic, 184 children died in 2022–2023 and 138 children died in 2023–2024 from laboratory-confirmed influenza. These numbers are undoubtedly underestimates, since all children are not tested for influenza, and even among those who are tested, tests are less sensitive later in the illness.

PATHOGENESIS AND IMMUNITY

At a cellular level, influenza virus binds to sialic acid receptors and enters the epithelial cell through receptor-mediated endocytosis. The virus then enters an endosome, where acidification promotes proteolytic cleavage of the hemagglutinin, exposing a fusion domain. The influenza hemagglutinin undergoes a marked structural reorganization in this cleavage step. Hemagglutinin cleavage may be one of the factors that restrict viral growth to epithelial cells, as a unique protease in the respiratory milieu is required for this cleavage to occur. The fusion domain allows the viral RNA to enter the cytoplasm. The nucleoprotein is transported into the nucleus of the cell, where transcription to a positive-sense RNA and replication take place. Viral proteins then assemble on the apical surface of the infected cell and, after incorporation of cellular membrane, bud from the membrane back into the mucosal milieu.

TABLE 206-2 High-Risk Groups Who Should Be Assigned a High Priority for Influenza Immunization and Treatment[a]

High-Risk Group
Children 6–59 months of age
Adults ≥50 years of age
Persons with chronic pulmonary (including asthma), cardiovascular (except isolated hypertension), renal, hepatic, neurologic, hematologic, or metabolic disorders (including diabetes mellitus)
Persons who are immunocompromised (any cause, including medications or HIV infection)
Women who are or plan to be pregnant during the influenza season
Children and adolescents (6 months through 18 years of age) who are receiving aspirin- or salicylate-containing medications and who might be at risk for Reye syndrome
Residents of nursing homes and other long-term-care facilities
American Indians/Alaska Natives
Persons who are extremely obese (body mass index ≥40)
Contacts and Caregivers
Caregivers and contacts of those at risk: health care personnel in inpatient and outpatient care settings who have the potential for exposure to patients or to infectious materials, medical emergency-response workers, autopsy personnel, employees of nursing home and long-term-care facilities who have contact with patients or residents, and students and trainees in these professions who have contact with patients
Household contacts and caregivers of children ≤59 months (i.e., <5 years) of age (particularly contacts of infants <6 months old) and adults ≥50 years of age
Household contacts (including children) and caregivers of persons who are in a high-risk group

[a]No hierarchy is implied by order of listing.

Source: Centers for Disease Control and Prevention 2023–2024 summary of recommendations for influenza vaccine (*https://www.cdc.gov/mmwr/volumes/72/rr/rr7202a1.htm*).

Influenza infection is initiated in the upper respiratory tract via aerosolized virus. The cells infected with influenza virus are primarily the ciliated cells of the respiratory tract. Denudation of the superficial epithelium probably accounts for much of the symptomatology and may predispose to secondary bacterial infections. The onset of symptoms follows an incubation period that, for a viral illness, is very short: 48–72 h. The infection spreads to the lungs but, even there, remains confined to the epithelial layer. Influenza virus is associated with systemic symptoms of fever, malaise, and myalgia. These manifestations are presumed to be mediated by cytokines, and excess cytokine production has been implicated in the acute toxicity of H5N1 and other highly pathogenic influenza viruses.

The immune response to influenza virus occurs at the systemic and mucosal levels and involves both T and B cells. The B-cell responses are directed primarily toward antigenic epitopes on the two surface glycoproteins—i.e., hemagglutinin and neuraminidase. At a structural level, the four recognized epitopes on the hemagglutinin are largely confined to the globular head of the protein, which collectively constitute the targets for hemagglutination inhibition (HAI) antibodies. HAI and neutralizing antibodies are highly correlated; HAI antibody levels are used as a measure of susceptibility to clinical infection and thus as a measure of vaccine-induced protection. In a child or an adult without prior vaccination or with the emergence of a distinctly new strain, serum HAI antibody is a surrogate for protection. However, in individuals with both vaccine-induced and natural immunity, the protective efficacy of a vaccine based on serum HAI antibody is more difficult to predict.

There is considerable research interest in the induction and protective role of broadly neutralizing antibodies that recognize less antigenically variable regions on the stalk of the hemagglutinin. The results of these studies have led to investments toward research and development of a universal influenza vaccine, although no such vaccines are yet available in clinical practice.

The role of T-cell immunity, which primarily recognizes internal protein epitopes, remains unclear in humans. However, T-cell

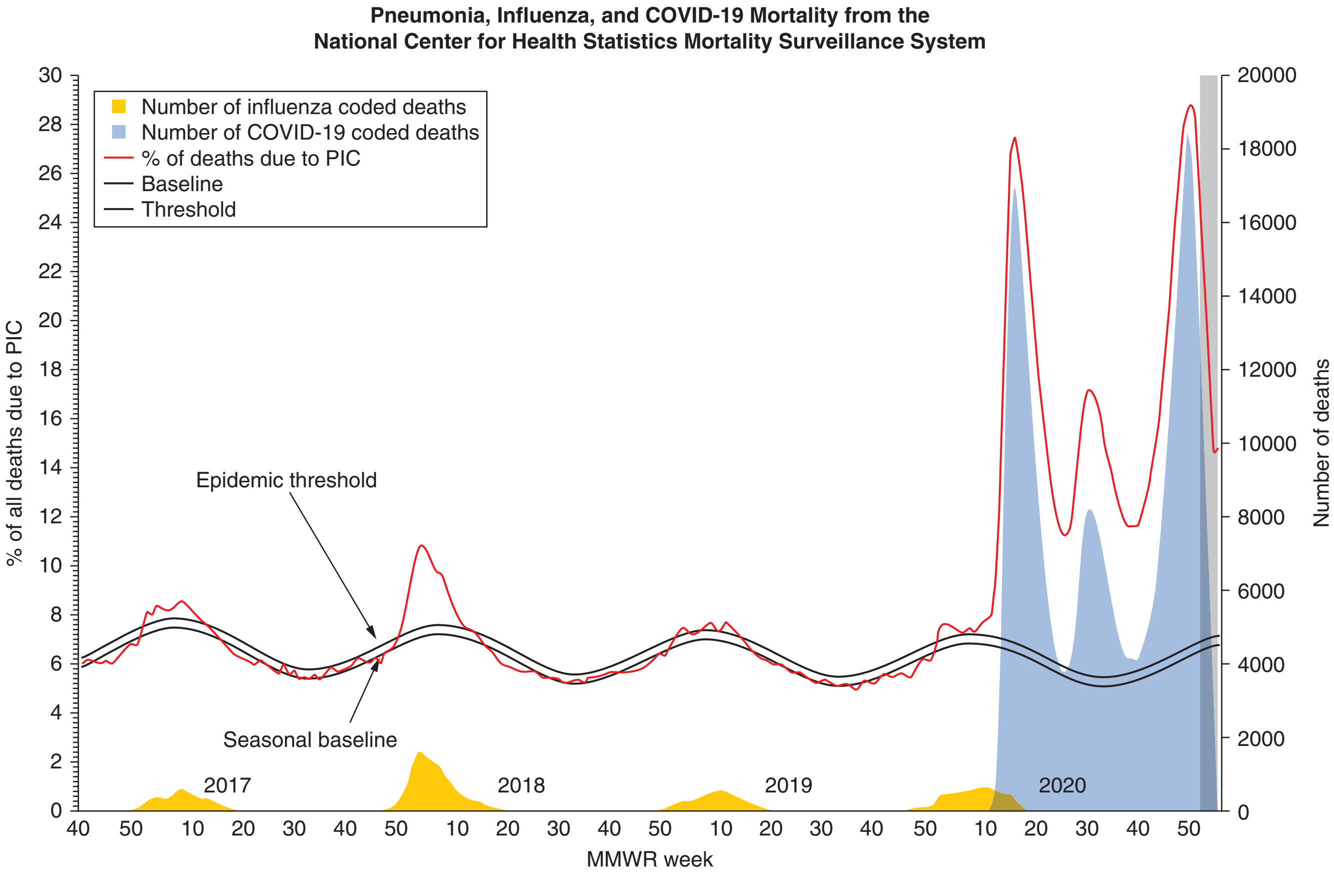

FIGURE 206-4 Pneumonia, influenza, and COVID-19 mortality. MMWR, *Morbidity and Mortality Weekly Report*; PIC, pneumonia, influenza, COVID-19. Data through the week ending January 23, 2021, as of January 28, 2021. *(From https://www.cdc.gov/fluview/?CDC_AAref_Val=https://www.cdc.gov/flu/weekly/index)*

immunity is thought to play a role in clearance of an influenza infection that quite reproducibly develops 8–10 days after exposure. A role for T cells in protection against acquisition of infection has also been proposed.

CLINICAL MANIFESTATIONS

Attack rates of clinical influenza vary considerably from year to year. With the advent of molecular diagnostic tests, prospective studies with regular sampling demonstrate that asymptomatic or minimally symptomatic cases of influenza are more common than previously recognized. When symptomatic, influenza is primarily a respiratory illness causing cough, sore throat, and rhinorrhea or nasal congestion. The illness has a sudden onset and is epidemiologically linked to close contact with persons who have similar symptoms and often to community-wide respiratory illness. What distinguishes influenza from most other respiratory viral illnesses is the degree of accompanying fever, chills, fatigue, myalgia, and malaise. SARS-CoV-2 is the exceptional respiratory virus that also has a remarkable systemic component **(Chaps. 204 and 205)**. Symptoms of influenza typically begin within 48–72 h of exposure.

Respiratory symptoms, particularly recurrent cough, persist well beyond the 2–5 days of systemic symptoms. There is a postinfectious delay in return to normal levels of activity. Pulmonary function is persistently decreased after acute influenza. Persons with a regular exercise routine (e.g., runners) note a decrease from their prior level of performance that typically lasts for a month or more. In the elderly, the respiratory presentation may be less prominent, but there is often a decline in baseline activity and a loss of appetite.

On physical examination, the patient with influenza appears ill, with sweating, coughing, nonpurulent conjunctivitis, and diffuse pharyngeal erythema. With lower respiratory involvement, pulmonary examination typically reveals nonlocalizing scattered rales, rhonchi, and wheezes. When present, localized pulmonary findings suggest relatively complicated pneumonia with a bacterial component. Muscle pain may be elicited by pressure, particularly in the calves and thighs. There are rare gastrointestinal findings. No rash is associated with influenza.

■ COMPLICATIONS

Most persons who become ill with influenza virus infection recover without serious complications or sequelae. Complications of influenza occur most commonly in persons ≥65 years of age, young children, persons of all ages with underlying cardiopulmonary disease and immunosuppression, and women who are in the second or third trimester of pregnancy.

Respiratory Complications Pneumonia characterized by progressive air hunger, localized pulmonary findings on physical examination, and radiographic findings of diffuse infiltrates or consolidation is the most common complication of influenza. Pneumonia in influenza can be primary influenza viral pneumonia, secondary bacterial pneumonia, or mixed viral and bacterial pneumonia. Primary viral pneumonia is characterized by increasing dyspnea, persistent fever, and—in more severe cases—cyanosis. Primary influenza pneumonia was typical in the 1918 pandemic and occurs with H5N1 virus, as initially described in Hong Kong in 1997. Pathologically, a marked inflammatory reaction in the alveolar septa is characterized by infiltration of monocytes, lymphocytes, and macrophages, with variable numbers of neutrophils. Destruction and hemorrhage are seen in the respiratory epithelium. Large amounts of virus can be recovered from the lungs.

In secondary bacterial pneumonia or mixed viral and bacterial pneumonia, illness may be biphasic, with evidence of recovery from the primary influenza illness followed by recrudescence of fever and pulmonary symptoms. Localizing findings may be detected on pulmonary examination and/or x-ray. The development of secondary bacterial infection is not surprising, as influenza de-epithelializes the airways and destroys ciliary function, allowing bacterial contamination. Another proposed mechanism for bacterial/viral enhancement is the production by *Staphylococcus* and *Pseudomonas* of proteases that enhance cleavage of the influenza hemagglutinin and thereby facilitate viral replication. The risk of secondary bacterial disease is greatest in elderly patients and those with chronic obstructive pulmonary disease (COPD).

Some influenza strains cause laryngotracheobronchitis, bronchiolitis, or croup in children. Otitis media—a common accompaniment to influenza in children—may also be due to a combination of influenza virus and bacteria.

Extrapulmonary Complications Although influenza is believed to spread only rarely beyond the respiratory epithelial cells, where unique endogenous proteases facilitate hemagglutinin cleavage and productive infection, this disease causes not only prominent systemic complaints but also a variety of extrapulmonary manifestations. The most common extrapulmonary manifestation of influenza is myositis, which is seen more often in influenza B and is characterized by severe muscle pain, elevated creatinine phosphokinase levels, and myoglobinuria that can lead to renal failure. The muscles are extremely tender to touch. Myo-/pericarditis is seen less frequently. However, a consistent epidemiologic link exists between influenza epidemics and excess cardiovascular hospitalizations.

Neurologic involvement, while rare, does occur following influenza infection. Influenza-associated encephalopathy or encephalitis is characterized by rapid progression within a few days of influenza infection. Transverse myelitis and parkinsonian symptoms have been reported. Postinfectious acute demyelinating encephalomyelitis can follow influenza as well as other viral infections. Neurologic manifestations are more frequent in children as compared to adults. Children most commonly present with febrile seizures, increased seizure frequency among those with seizure disorders, or self-limited encephalopathy. More serious manifestations of meningitis, encephalitis, and focal brain lesions may occur, particularly in children with preexisting neurologic conditions.

Guillain-Barré syndrome can develop after influenza and was reported after a widespread influenza vaccination effort in the fall of 1976 that was undertaken in anticipation of a swine influenza epidemic (which never materialized). Until aspirin was recognized as a cofactor in its precipitation, Reye syndrome, an acute hepatic decompensation, was seen commonly in children and adolescents with influenza, particularly those infected with influenza B virus. Subsequently, the use of aspirin for fever control and symptom relief in children with viral infections was strongly discouraged, and Reye syndrome has virtually disappeared from clinical practice.

LABORATORY FINDINGS AND DIAGNOSIS

There is a strong argument for establishing a microbiologic diagnosis from both an individual-patient and a public-health perspective. This information is particularly valuable early in the season, when the extent of influenza and the precise circulating strain(s) are uncertain; in the management of high-risk or hospitalized patients; in settings such as long-term-care facilities and hospitals, where the institution of specific infection-control measures is appropriate; and in any patient with influenza-like illness if the test results will influence clinical management.

Influenza virus is most easily recovered from nasal or pharyngeal specimens. A number of rapid influenza diagnostic tests (RIDTs) are available. They work by detecting viral antigens and can provide results within 10–15 min. Rapid molecular assays (i.e., nucleic acid amplification tests [NAATs]) detect viral genetic material. Several NAATs are authorized for home use, including tests that detect and differentiate between SARS-CoV-2, influenza A, and influenza B in self-collected anterior nasal swab samples. In addition to RIDTs and rapid molecular assays, several influenza diagnostic tests are available in specialized hospital and public health laboratories, including reverse transcription polymerase chain reaction (RT-PCR) and viral culture. Many nucleic acid–based tests are multiplex and target a panel of common respiratory pathogens— influenza, respiratory syncytial virus, parainfluenzavirus, and coronaviruses including SARS-CoV-2—an advantage in the ill hospitalized patient and during outbreaks of other respiratory pathogens. Clinicians should not use viral culture for initial or primary diagnosis of influenza because results will not be available in a timely manner to inform clinical management, but viral culture can confirm the strain and allow for antiviral sensitivity testing.

Serologic confirmation of infection is also possible but requires paired serum samples, with the convalescent-phase sample obtained 2 weeks after infection. Other laboratory tests are of limited value. Mild leukopenia is seen in influenza, and a white blood cell count >15,000/μL suggests a secondary bacterial component in influenza pneumonia.

DIFFERENTIAL DIAGNOSIS

Influenza may be diagnosed clinically based on an acute presentation of a febrile respiratory illness during high periods of influenza circulation. However, less common presentations of influenza and cases occurring outside of peak influenza season are frequently misdiagnosed on the basis of symptoms alone. Influenza symptoms and signs may overlap with symptoms of other respiratory viruses. Respiratory syncytial virus often co-circulates with influenza virus; it particularly affects the youngest children, causing bronchiolitis, but it can also infect the elderly, leading to an influenza-like nonspecific respiratory illness and a decline in mobility, nutrition, and pulmonary function, with resultant hospitalization.

Persons with COVID-19 have a wide range of symptoms reported, ranging from mild to severe illness. Many of these symptoms—fever, chills, cough, shortness of breath, fatigue, muscle aches, headaches, congestion or runny nose—overlap with the symptoms of influenza. While new loss of taste (ageusia) or smell (anosmia) may distinguish COVID-19 from influenza, they are reported in the minority of infected persons. When SARS-CoV-2 and influenza viruses are co-circulating, clinicians should consider both viruses, as well as co-infection, in patients with acute respiratory illness symptoms. The similar clinical presentations reiterate the importance of testing in order to inform treatment decisions.

IMMUNIZATION

Vaccination is the best approach to prevent influenza. The vaccines currently available in the United States are increasing in number and diversity **(Table 206-3)**. These vaccines fall into two broad categories: parenterally administered inactivated influenza vaccines and intranasally administered live-attenuated influenza vaccines. Current vaccines are further classified based on production substrate (eggs, cell), antigen dose and valence (trivalent or quadrivalent), and the presence or absence of adjuvants. Current inactivated influenza vaccines are designed with the common goal to induce immunity to the

TABLE 206-3 Categories of Vaccines Licensed for Prevention of Seasonal Influenza, United States

		NONREPLICATING VACCINES			
	LIVE ATTENUATED	STANDARD INACTIVATED	HIGH-DOSE INACTIVATED	RECOMBINANT	ADJUVANTED INACTIVATED
Route	Intranasal	Intramuscular	Intramuscular	Intramuscular	Intramuscular
Approved ages	2–49 years	≥6 months	≥65 years	≥18 years	≥65 years
HA[a]	15	15	60	45	15
Substrate	Eggs	Eggs/cell culture	Eggs	Cell culture	Eggs
Number of strains	3	3	3	3	3

[a] Hemagglutinin content in micrograms per strain.

hemagglutinin surface glycoprotein of the influenza virus. No effort is made to standardize the neuraminidase content.

As the viral surface hemagglutinin undergoes frequent antigenic drift, the seasonal influenza vaccine is reformulated as often as twice annually to match the strains projected to circulate in the following influenza season. The decision about vaccine composition must be made ~10 months before the seasonal peak in influenza virus circulation; this decision is made by committees at the WHO. Subsequently, the U.S. Food and Drug Administration (FDA), which has regulatory authority over vaccines in the United States, convenes an advisory committee that considers the recommendations of WHO, reviews and discusses similar data, and makes a final decision regarding vaccine virus composition of influenza vaccines licensed and marketed in the United States. This timing can result in a mismatch of vaccine composition with the viral strains that are actually prevalent in the upcoming season. Influenza vaccine is unique in being given seasonally in the months immediately preceding an outbreak in temperate climates. In the United States, vaccine is typically available starting in August or September.

The performance of current influenza vaccines varies by year, vaccine formulation, and the underlying age, health condition, and prior virus and vaccine exposure of the recipient. Unfortunately, the relative contribution of each of these factors has not been well-elucidated, given the many variables involved and the complex interplay of infection and host response. Depending on the degree to which vaccine strains match circulating strains, seasonal influenza vaccines will confer more or less protection, as antibody against influenza is for the most part strain specific. A meta-analysis of randomized controlled trials of influenza vaccine efficacy over 12 influenza seasons showed inactivated influenza vaccines had a pooled efficacy of 59% (95% confidence interval, 51–67%) among those aged 18–65 years. Since 2004–2005, the CDC has estimated the effectiveness of seasonal influenza vaccine to prevent laboratory-confirmed influenza associated with medically attended respiratory illness. During that period, effectiveness ranged from ~40 to 60% across all age groups during seasons when most circulating influenza vaccines are antigenically similar to the recommended influenza vaccine components; effectiveness was lower in years with strain mismatch. Importantly, studies support that influenza vaccine mitigates disease severity. For example, observational studies in children support that influenza vaccination reduces intensive care unit hospitalizations and deaths by an estimated 74 and 65%, respectively.

Newer technologies have been developed to overcome some of the limitations of current vaccines. The first fully recombinant vaccine was approved by the FDA in 2017. Both recombinant and cell-based vaccines may overcome the egg adaptation of vaccine strains that may contribute to diminished vaccine effectiveness. Oil-in-water adjuvanted vaccines and high-dose vaccines elicit greater immune responses than traditional inactivated influenza vaccines and are approved in the United States for persons ≥65 years of age. In most head-to-head comparisons, high-dose vaccines have shown superior effectiveness to standard dose. While evidence is more limited, select comparisons of recombinant and adjuvanted vaccines with standard vaccines likewise show improved effectiveness.

In head-to-head comparisons in pediatric populations in the 1990s, a live, attenuated, intranasally administered vaccine (LAIV) exhibited an efficacy exceeding that of injected inactivated vaccines. LAIV is a desirable option in children given the ease of intranasal administration and theoretical advantage of stimulating mucosal immunity by the topical route. However, in the 2014–2016 influenza seasons, LAIV had lower replicative fitness and no demonstrable efficacy assignable to the vaccine's H1N1 component. Consequently, advisory committees in the United States and elsewhere suspended the recommendations for use of LAIV until manufacturing improvements allowed reinstatement of recommendations for its use in 2018. Since that time LAIV has performed comparably to inactivated influenza vaccines in annual effectiveness assessments.

Inactivated influenza vaccines have been licensed for >60 years and have a robust safety and tolerability profile. While local reactions are most common following inactivated influenza vaccines, rare adverse events may occur. These include Guillain-Barré syndrome, identified in 1976 and less frequently during other years; oculorespiratory syndrome, first recognized in 2000; and febrile seizures first reported in young children in Australia in 2010. Adjuvanted vaccines in general cause more local pain and erythema than unadjuvanted vaccines. LAIVs have been associated with excess wheezing and hospitalizations in children younger than 2 years and thus are not licensed for use in this age group.

The recommendations for use, the approved age range of each product, the route of administration, and the anticipated side effects are updated annually by the CDC (*https://www.cdc.gov/acip-recs/hcp/vaccine-specific/flu.html.*) In the United States, routine annual influenza vaccination is recommended for all persons 6 months of age and older. For persons 65 years of age and over, higher-dose or adjuvanted vaccines are preferred. For other age groups, no preferential recommendation is made for one influenza vaccine product over another. Two doses of vaccine should be given to children <9 years of age who have not received at least two lifetime doses of influenza vaccine prior to the start of the season. All other children should receive one dose. Groups at special risk of experiencing or transmitting influenza and for whom influenza immunization is a particularly high priority are listed in Table 206-2.

History of severe allergic reaction (e.g., anaphylaxis) to any vaccine component is a contraindication to influenza vaccines. A history of Guillain-Barré syndrome within 6 weeks of a previous dose of influenza vaccine is considered a precaution for the use of all influenza vaccines. Egg allergy alone necessitates no additional safety measures for influenza vaccination beyond those recommended for any vaccine recipient, regardless of severity of previous reaction to egg.

TREATMENT

Influenza

Antiviral therapy for influenza has been limited by the paucity of available drugs, the short duration of symptoms in uncomplicated influenza, and the changing patterns of drug resistance in influenza viral strains. In the past, influenza A infection could be treated with the M-2 channel blockers amantadine and rimantadine. Widespread resistance has currently relegated these compounds to historical interest only.

Neuraminidase inhibitors have been the mainstay for treatment of influenza A and B viruses for many years. As their name implies, these drugs inhibit the influenza neuraminidase and thus limit the egress of influenza virus from an infected cell. They are most effective in patients whose influenza illness is recognized early and confirmed by rapid diagnostic testing or on the basis of clinical and epidemiologic evidence. In experimental trials, these drugs hasten the resolution of symptoms if given within 48 h of infection. There are indications for their use both prophylactically—either throughout the season or, when a case is recognized in a close contact, in the short term—and therapeutically. The anticipated effect of early administration is the resolution of symptoms 1–2 days sooner than without treatment. The use of neuraminidase inhibitors is recommended for complicated influenza infections in hospitalized patients in the absence of formal proof of efficacy and when diagnosis may have been delayed. All the available neuraminidase inhibitors carry a risk of development of resistance, particularly with prolonged administration (e.g., to an immunodeficient individual with persistent recovery of influenza virus). Resistance to neuraminidase inhibitors is not widespread among currently circulating influenza A or B strains, but its development has been demonstrated in the laboratory, and clinical resistance could influence the utility of these drugs.

The defined risk groups who can benefit from neuraminidase inhibitors include children <2 years of age, adults >65 years of age, patients with chronic conditions, immunosuppressed individuals, pregnant women, women who have delivered infants ≤2 weeks previously, patients <19 years old who are receiving long-term aspirin treatment, Native Americans (including Alaska Natives), morbidly obese individuals, and residents of nursing homes or chronic-care facilities. This list resembles that of candidates whose vaccination is a high priority (Table 206-2). Use of neuraminidase inhibitors should be considered in selected high-risk cases despite a history of vaccination.

The available neuraminidase inhibitors are oral oseltamivir, nasal-spray zanamivir, and intravenous peramivir. Oseltamivir, which is most widely used, is an orally absorbed drug that is converted to its active component, oseltamivir carboxylate, in the liver. Gastrointestinal symptoms, especially nausea, may accompany the administration of oseltamivir. Because zanamivir is not orally bioavailable, it is given as an inhaled dry powder dispersed through a Diskhaler device.

The usual duration of therapy with either oral oseltamivir or intranasal zanamivir is 5 days, with twice-a-day dosing. Oseltamivir is preferred for treatment of pregnant women and is approved for treatment at any age, beginning at 14 days of life in infants. Poor oral intake or absorption is a contraindication to the use of oseltamivir, although this drug can also be given by oro/nasal tube. Asthma and COPD are relative contraindications to the use of intranasal zanamivir; this agent is approved for treatment in persons 7 years and older. For hospitalized patients with suspected or confirmed influenza, initiation of antiviral treatment with oral or enterically administered oseltamivir is recommended as soon as possible. For patients who cannot tolerate or absorb oral or enterically administered oseltamivir, the use of a single infusion of intravenous peramivir should be considered. Peramivir is licensed for individuals ≥2 years of age. The most current recommendations and details on influenza antiviral drug use and release are available through the CDC (*https://www.cdc.gov/flu/professionals/antivirals/summary-clinicians.htm*).

In 2018, a first-in-class compound, baloxavir marboxil (XOFLUZA), was approved by the FDA for persons 12 years and older for prophylaxis or treatment of uncomplicated influenza within 2 days of onset of illness. Baloxavir inhibits cap-dependent endonuclease, has activity against influenza A and B, and is a single-dose formulation. In clinical studies, if given within 48 h of symptoms, baloxavir decreased symptom duration, viral shedding, and antibiotic use in healthy individuals with uncomplicated influenza. However, development of resistance is a concern, with 2–10% of the trial participants who received baloxavir showing viral escape with reduced drug susceptibility. CDC does not recommend use of baloxavir in pregnant women, breastfeeding mothers, outpatients with complicated or progressive illness, severely immunosuppressed people, or hospitalized patients because of lack of information on use of baloxavir for these groups.

Other critical aspects of treatment include maintenance of fluid and electrolyte balance, oxygen supplementation, fever control with nonsteroidal anti-inflammatory drugs, and treatment of suspected secondary bacterial complications with antibiotics. Appropriate respiratory isolation of patients should be practiced in accordance with local hospital guidelines.

FURTHER READING

Barry JM: *The Great Influenza: The Story of the Deadliest Pandemic in History*. New York, Penguin Books, 2005.

Chung JR: Effects of influenza vaccination in the United States during the 2018–2019 influenza season. Clin Infect Dis 71:e368, 2020.

Erbelding EJ: A universal influenza vaccine: The strategic plan for the National Institute of Allergy and Infectious Diseases. J Infect Dis 218:347, 2018.

Fineberg HV: Pandemic preparedness and response—lessons from the H1N1 influenza of 2009. N Engl J Med 370:1335, 2014.

Kash JC, Taubenberger JK: The role of viral, host, and secondary bacterial factors in influenza pathogenesis. Am J Pathol 185:1528, 2015.

Osterholm MT et al: Efficacy and effectiveness of influenza vaccines: A systemic review and meta-analysis. Lancet Infect Dis 12:36, 2012.

Treanor JJ: Influenza vaccination. N Engl J Med 375:1261, 2016.

Uyeki TM et al: Influenza. Lancet 400:693, 2022.

Watanabe T et al: 1918 influenza virus hemagglutinin (HA) and the viral RNA polymerase complex enhance viral pathogenicity, but only HA induces aberrant host responses in mice. J Virol 87:5239, 2013.

Wright PF et al: Correlates of immunity to influenza as determined by challenge of children with live, attenuated influenza vaccine. Open Forum Infect Dis 3:108, 2016.

Section 14 Infections Due to Human Immunodeficiency Virus and Other Human Retroviruses

207

The Human Retroviruses

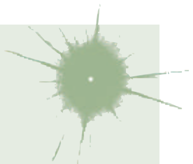

Dan L. Longo, Anthony S. Fauci

The retroviruses, which make up a large family (Retroviridae), infect mainly vertebrates. These viruses have a unique replication cycle whereby their genetic information is encoded by RNA rather than DNA. Retroviruses contain an RNA-dependent DNA polymerase (a reverse transcriptase) that directs the synthesis of a DNA form of the viral genome after infection of a host cell. The designation *retrovirus* denotes that information in the form of RNA is transcribed into DNA in the host cell—a sequence that overturned a central dogma of molecular biology: that information passes unidirectionally from DNA to RNA to protein. The observation that RNA was the source of genetic information in the causative agents of certain animal tumors led to a number of paradigm-shifting biologic insights regarding not only the direction of genetic information passage but also the viral etiology of certain cancers and the concept of oncogenes as normal host genes scavenged and altered by a viral vector.

The family Retroviridae includes seven subfamilies **(Table 207-1)**. Members of two of the families infect humans with pathologic consequences: the deltaretroviruses, of which human T-cell lymphotropic virus (HTLV) type 1 is the most important in humans; and the lentiviruses, of which HIV is the most important in humans.

The wide variety of interactions of a retrovirus with its host range from completely benign events (e.g., silent carriage of endogenous retroviral sequences in the germline genome of many animal species) to rapidly fatal infections (e.g., exogenous infection with an oncogenic virus such as Rous sarcoma virus in chickens). The ability of

TABLE 207-1 Classification of Retroviruses: The Family Retroviridae

GENUS	EXAMPLE(S)	FEATURE
Alpharetrovirus	Rous sarcoma virus	Contains *src* oncogene
Betaretrovirus	Mouse mammary tumor virus	Exogenous or endogenous
Gammaretrovirus	Abelson murine leukemia virus	Contains *abl* oncogene
Deltaretrovirus	HTLV-1	Causes T-cell lymphoma and neurologic disease
Epsilonretrovirus	Walleye dermal sarcoma virus	Not known to be pathogenic in humans
Lentivirus	HIV-1, HIV-2	Causes AIDS
Spumavirus	Simian foamy virus	Not known to be pathogenic in humans

retroviruses to acquire and alter the structure and function of host cell genetic sequences has revolutionized our understanding of molecular carcinogenesis. The viruses can insert into the germline genome of the host cell and behave as a transposable or movable genetic element. They can activate or inactivate genes near the site of integration into the genome. They can rapidly alter their own genome by recombination and mutation under selective environmental stimuli.

Most human viral diseases occur as a consequence of tissue destruction either directly by the virus itself or indirectly by the host's response to the virus. Although these mechanisms are operative in retroviral infections, retroviruses have additional mechanisms of inducing disease, including the malignant transformation of an infected cell and the induction of an immunodeficiency state through selective destruction or dysfunction of immune-competent cells that renders the host susceptible to opportunistic diseases (infections and neoplasms; **Chap. 208**).

In addition to their role as acquired causes of infections, retroviruses and lentiviruses are employed as tools to alter gene expression in target cells. Gene therapy applications are expanding, and in addition to the benefits they bestow, a new set of viral-mediated adverse events have emerged.

STRUCTURE AND LIFE CYCLE

All retroviruses are similar in structure, genome organization, and mode of replication. Retroviruses are 70–130 nm in diameter and have a lipid-containing envelope surrounding an icosahedral capsid with a dense inner core. The core contains two identical copies of the single-strand RNA genome. The RNA molecules are 8–10 kb long and are complexed with reverse transcriptase and tRNA. Other viral proteins, such as integrase, are also components of the virion particle. The RNA has features usually found in mRNA: a cap site at the 5′ end of the molecule, which is important in the initiation of mRNA translation, and a polyadenylation site at the 3′ end, which influences mRNA turnover (i.e., messages with shorter polyA tails turn over faster than messages with longer polyA tails). However, the retroviral RNA is not translated; instead, it is transcribed into DNA. The DNA form of the retroviral genome is called a *provirus*.

The replication cycle of retroviruses proceeds in two phases **(Fig. 207-1)**. In the first phase, the virus enters the cytoplasm after binding to one or more specific cell-surface receptors; the viral RNA and reverse transcriptase synthesize a double-strand DNA version of the RNA template; and the provirus moves into the nucleus and integrates into the host cell genome. This proviral integration is permanent. Although some animal retroviruses integrate into a single specific site of the host genome in every infected cell, the human retroviruses integrate randomly. This first phase of replication depends entirely on gene products in the virus. The second phase includes the synthesis and processing of viral genomes, mRNAs, and proteins using host cell machinery, often under the influence of viral gene products. Virions are assembled and released from the cell by budding from the membrane; host cell membrane proteins are frequently incorporated into the envelope of the virus. Proviral integration occurs during the S-phase of the cell cycle; thus, in general, nondividing cells are resistant to retroviral infection. Only the lentiviruses are able to infect nondividing cells. Once a host cell is infected, it is infected for the life of the cell.

Retroviral genomes include both coding and noncoding sequences **(Fig. 207-2)**. In general, noncoding sequences are important recognition signals for DNA or RNA synthesis or processing events and are located in the 5′ and 3′ terminal regions of the genome. All retroviral genomes are terminally redundant, containing identical sequences called *long terminal repeats* (LTRs). The ends of the retroviral RNA genome differ slightly in sequence from the integrated retroviral DNA. In the latter, the LTR sequences are repeated in both the 5′ and the 3′ terminus of the virus. The LTRs contain sequences involved in initiating the expression of the viral proteins, the integration of the provirus, and the polyadenylation of viral RNAs. The primer binding site, which is critical for the initiation of reverse transcription, and the viral packaging sequences are located outside the LTR sequences. The coding regions include the *gag* (group-specific antigen, core protein), *pol* (RNA-dependent DNA polymerase), and *env* (envelope) genes. The *gag* gene encodes a precursor polyprotein that is cleaved to form three to five capsid proteins; a fraction of the Gag precursor proteins also contain a protease responsible for cleaving the Gag and Pol polyproteins. A Gag-Pol polyprotein gives rise to the protease that is responsible for cleaving the Gag-Pol polyprotein. The *pol* gene encodes three proteins: the reverse transcriptase, the integrase, and the protease. The reverse transcriptase copies the viral RNA into the double-strand DNA provirus, which inserts itself into the host cell DNA via the action of integrase. The protease cleaves the Gag-Pol polyprotein into smaller protein products. The *env* gene encodes the envelope glycoproteins: one protein that binds to specific surface receptors and determines what cell types can be infected and a smaller transmembrane protein that anchors the complex to the envelope. **Fig. 207-3** shows how the retroviral gene products make up the virus structure.

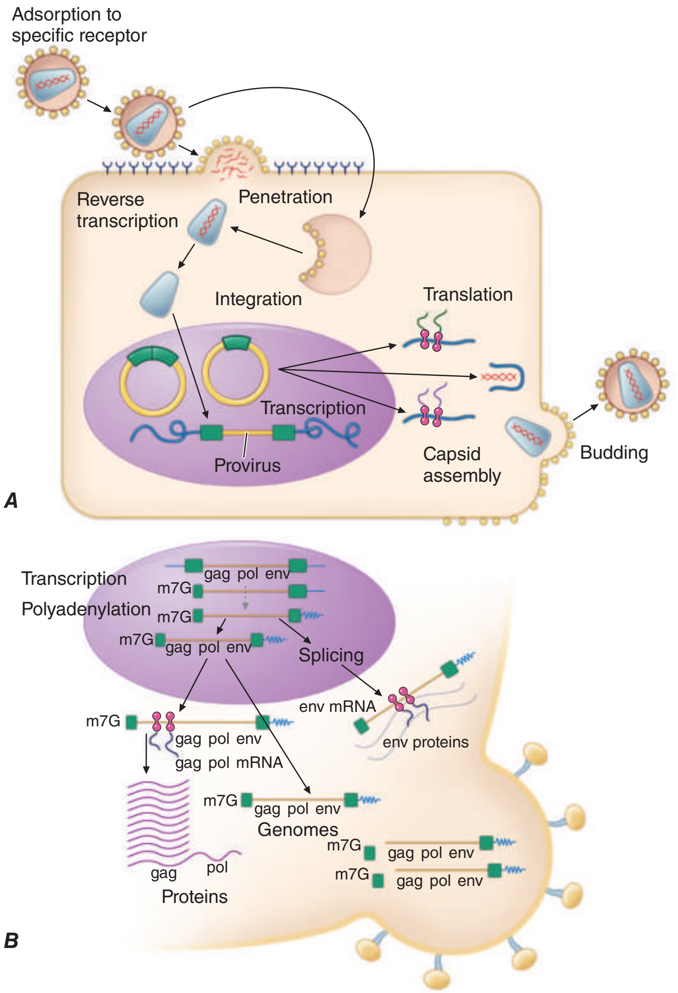

FIGURE 207-1 The life cycle of retroviruses. ***A.*** Overview of virus replication. The retrovirus enters a target cell by binding to a specific cell-surface receptor; once the virus is internalized, its RNA is released from the nucleocapsid and is reverse-transcribed into proviral DNA. The provirus is inserted into the genome and then transcribed into RNA; the RNA is translated; and virions assemble and are extruded from the cell membrane by budding. ***B.*** Overview of retroviral gene expression. The provirus is transcribed, capped, and polyadenylated. Viral RNA molecules then have one of three fates: they are exported to the cytoplasm, where they are packaged as the viral RNA in infectious viral particles; they are spliced to form the message for the envelope polyprotein; or they are translated into Gag and Pol proteins. Most of the messages for the Pol protein fail to initiate Pol translation because of a stop codon before its initiation; however, in a fraction of the messages, the stop codon is missed, and the Pol proteins are translated. *(Reproduced with permission from JM Coffin, in BN Fields, DM Knipe [eds]: Fields Virology. New York, Raven, 1990.)*

HTLVs have a region between *env* and the 3′ LTR that encodes several proteins and transcripts in overlapping reading frames (Fig. 207-2). Tax is a 40-kDa protein that does not bind to DNA but induces the expression of host cell transcription factors that alter host cell gene expression

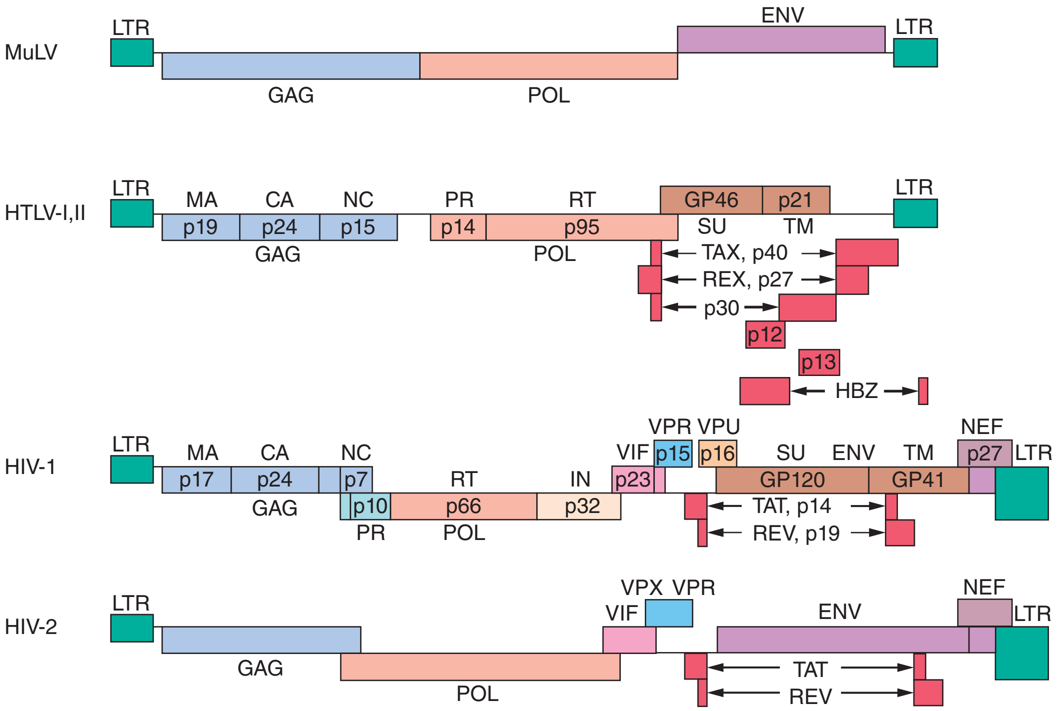

FIGURE 207-2 Genomic structure of retroviruses. The murine leukemia virus MuLV has the typical three structural genes: *gag*, *pol*, and *env*. The *gag* region gives rise to three proteins: matrix (MA), capsid (CA), and nucleic acid–binding (NC) proteins. The *pol* region encodes both a protease (PR) responsible for cleaving the viral polyproteins and a reverse transcriptase (RT). In addition, HIV *pol* encodes an integrase (IN). The *env* region encodes a surface protein (SU) and a small transmembrane protein (TM). The human retroviruses have additional gene products translated in each of the three possible reading frames. HTLV-1 and HTLV-2 have *tax* and *rex* genes with exons on either side of the *env* gene. HIV-1 and HIV-2 have six accessory gene products: *tat, rev, vif, nef, vpr*, and either *vpu* (in HIV-1) or *vpx* (in HIV-2). The genes for these proteins are located mainly between the *pol* and *env* genes. GP, glycoprotein; HBZ, HTLV-1 basic leucine zipper domain–containing protein; LTR, long terminal repeat.

and is capable of inducing cell transformation under certain circumstances. Rex is a 27-kDa protein that regulates the expression of viral mRNAs. Other transcripts from this region (p12, p13, and p30) tend to restrict expression of viral genes and diminish the immunogenicity of infected cells. The protein of *HBZ*, a product of the complementary proviral DNA strand, interacts with many cellular transcription factors and signaling proteins. It stimulates proliferation of infected cells and is the only viral product universally expressed in HTLV-1-infected tumor cells. These proteins are produced from messages that are similar but that are spliced differently from overlapping but distinct exons.

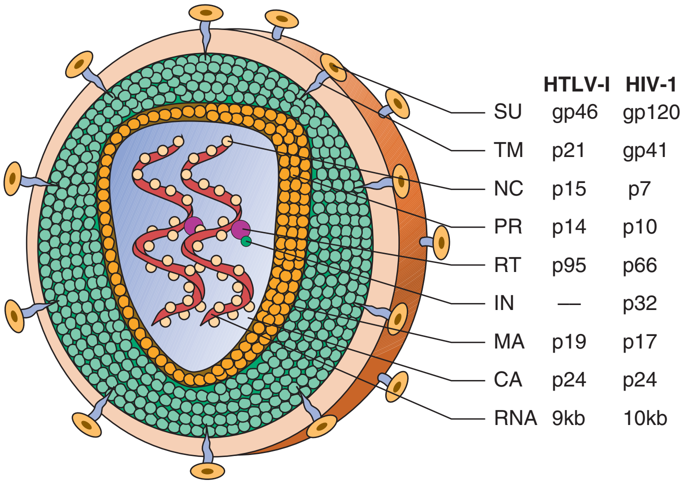

FIGURE 207-3 Schematic structure of human retroviruses. The surface glycoprotein (SU) is responsible for binding to receptors of host cells. The transmembrane protein (TM) anchors SU to the virus. NC is a nucleic acid–binding protein found in association with the viral RNA. A protease (PR) cleaves the polyproteins encoded by the *gag, pol*, and *env* genes into their functional components. RT is reverse transcriptase, and IN is an integrase present in some retroviruses (e.g., HIV-1) that facilitates insertion of the provirus into the host genome. The matrix protein (MA) is a Gag protein closely associated with the lipid of the envelope. The capsid protein (CA) forms the major internal structure of the virus, the core shell.

The lentiviruses in general, and HIV-1 and -2 in particular, contain a larger genome than other pathogenic retroviruses. They contain an untranslated region between *pol* and *env* that encodes portions of several proteins, varying with the reading frame into which the mRNA is spliced. Tat is a 14-kDa protein that augments the expression of virus from the LTR. The Rev protein of HIV-1, similar to the Rex protein of HTLV, regulates RNA splicing and/or RNA transport. The Nef protein downregulates CD4, the cellular receptor for HIV; alters host T-cell–activation pathways; and enhances viral infectivity. The Vif protein is necessary for the proper assembly of the HIV nucleoprotein core in many types of cells; without Vif, proviral DNA is not efficiently produced in these infected cells. In addition, the Vif protein targets APOBEC (apolipoprotein B mRNA-editing enzyme catalytic polypeptide, a cytidine deaminase that mutates the viral sequence) for proteasomal degradation, thus blocking its virus-suppressing effect. Vpr, Vpu (HIV-1 only), and Vpx (HIV-2 only) are viral proteins encoded by translation of the same message in different reading frames. As noted above, oncogenic retroviruses depend on cell proliferation for their replication; lentiviruses can infect nondividing cells, largely through effects mediated by Vpr. Vpr facilitates transport of the provirus into the nucleus and can induce other cellular changes, such as G_2 growth arrest and differentiation of some target cells. Vpx is structurally related to Vpr, but its functions are not fully defined. Vpu promotes the degradation of CD4 in the endoplasmic reticulum and stimulates the release of virions from infected cells.

Retroviruses can be either exogenously acquired (by infection with an infected cell or a free virion capable of replication) or transmitted in the germline as endogenous virus. Endogenous retroviruses are often replication defective. The human genome contains endogenous retroviral sequences, but there are no known replication-competent endogenous retroviruses in humans.

In general, viruses that contain only the *gag, pol*, and *env* genes either are not pathogenic or take a long time to induce disease; these observations indicate the importance of the other regulatory genes in viral disease pathogenesis. The pathogenesis of neoplastic transformation by retroviruses relies on the chance integration of the provirus at a spot in the genome resulting in the expression of a cellular gene (protooncogene) that becomes transforming by virtue of its unregulated expression. For example, avian leukosis virus causes B-cell leukemia by inducing the expression of *myc*. Some retroviruses possess captured and altered cellular genes near their integration site, and these viral oncogenes can transform the infected host cell. Viruses that have oncogenes often have lost a portion of their genome that is required for replication. Such viruses need helper viruses to reproduce, a feature that may explain why these acute transforming retroviruses are rare in nature. All human retroviruses identified to date are exogenous and are not acutely transforming (i.e., they lack a transforming oncogene).

These remarkable properties of retroviruses have led to experimental efforts to use them as vectors to insert specific genes into particular cell types, a process known as *gene therapy* or *gene transfer*. The process could be used to repair a genetic defect or to introduce a new property that could be used therapeutically; for example, a gene (e.g., thymidine kinase) that would make a tumor cell susceptible to killing by a drug (e.g., ganciclovir) could be inserted. One source of concern about the use of retroviral vectors in humans is that replication-competent viruses might rescue endogenous retroviral replication, with unpredictable results. This concern is not merely hypothetical: the detection of proteins encoded by endogenous retroviral sequences on the surface of cancer cells implies that the genetic events leading to the cancer were able to activate the synthesis of these usually silent genes.

Lentiviruses are being widely examined as gene therapy tools to correct genetic defects (e.g., thalassemia, sickle cell disease, hemophilia) and to generate effective cancer therapies (e.g., chimeric antigen receptor T cells, natural killer [NK] cells, and macrophages). The success of these therapies has been variable. Gene therapy interventions are often not permanent as expression of the introduced gene wanes over time. The random insertion of genetic material into multiple sites of the genome in some cells has resulted in insertional mutagenesis and the generation of novel malignancies. While gene therapies are overall only rarely associated with second malignancies, the expectation is that the use of directed insertion of genetic material into a specific site in the genome (e.g., with CRISPR technology) will reduce the risk.

HUMAN T-CELL LYMPHOTROPIC VIRUS

HTLV-1, a delta retrovirus, was isolated in 1980 from a T-cell lymphoma cell line from a patient originally thought to have cutaneous T-cell lymphoma. Later it became clear that the patient had a distinct form of lymphoma (originally reported in Japan) called *adult T-cell leukemia/lymphoma* (ATL). Serologic data have determined that HTLV-1 is the cause of at least two important diseases: ATL and tropical spastic paraparesis, also called *HTLV-1-associated myelopathy* (HAM). HTLV-1 may also play a role in infective dermatitis, arthritis, uveitis, and Sjögren's syndrome.

Two years after the isolation of HTLV-1, HTLV-2 was isolated from a patient with an unusual form of hairy cell leukemia that affected T cells. Epidemiologic studies of HTLV-2 failed to reveal a consistent disease association. Similarly, HTLV-3 and HTLV-4 have been identified but have no known disease association.

BIOLOGY AND MOLECULAR BIOLOGY

Because the biology of HTLV-1 and that of HTLV-2 are similar, the following discussion will focus on HTLV-1.

Human glucose transporter protein 1 (GLUT-1) functions as a receptor for HTLV-1, probably acting together with neuropilin-1 (NRP1) and heparan sulfate proteoglycans. The heparan sulfate proteoglycans do not appear to be involved with HTLV-2 cell entry. Generally, only T cells are productively infected, but infection of B cells and other cell types is occasionally detected. The most common outcome of HTLV-1 infection is latent carriage of randomly integrated provirus in CD4+ T cells. HTLV-1 does not contain an oncogene and does not insert into a unique site in the genome. Indeed, most infected cells express no viral gene products. The only viral gene product that is routinely expressed in tumor cells transformed by HTLV-1 in vivo is *hbz*. The *tax* gene is thought to be critical to the transformation process but is not expressed in the tumor cells of many ATL patients, possibly because of the immunogenicity of *tax*-expressing cells. Cells transformed in vitro, by contrast, actively transcribe HTLV-1 RNA and produce infectious virions. Most HTLV-1-transformed cell lines are the result of the infection of a normal host T cell in vitro. It is difficult to establish cell lines derived from authentic ATL cells.

Although *tax* does not itself bind to DNA, it is located in the nucleus and induces the expression of a wide range of host cell gene products, including transcription factors (especially c-rel/nuclear factor κB [NF-κB], ets-1 and -2, and members of the fos/jun family), cytokines (e.g., interleukin [IL] 2, granulocyte-macrophage colony-stimulating factor, and tumor necrosis factor), membrane proteins and receptors (major histocompatibility [MHC] molecules and IL-2 receptor α), and chromatin remodeling complexes. The genes activated by *tax* are generally controlled by transcription factors of the c-rel/NF-κB and cyclic AMP response element binding (CREB) protein families. It is unclear how this induction of host gene expression leads to neoplastic transformation; *tax* can interfere with G_1 and mitotic cell-cycle checkpoints, block apoptosis, inhibit DNA repair, and promote antigen-independent T-cell proliferation. Induction of a cytokine-autocrine loop has been proposed; however, IL-2 is not the crucial cytokine. The involvement of IL-4, IL-7, and IL-15 has been proposed.

In light of the irregular expression of *tax* in ATL cells, it has been suggested that *tax* is important in the early phases of transformation but is not essential for the maintenance of the transformed state. The maintenance role is thought to be due to *hbz* expression. As is clear from the epidemiology of HTLV-1 infection, transformation of an infected cell is a rare event and may depend on heterogeneous second, third, or fourth genetic hits. No consistent chromosomal abnormalities have been described in ATL; however, aneuploidy is common, and individual cases with p53 mutations and translocations involving the T-cell receptor genes on chromosome 14 have been reported. *Tax* may repress certain DNA repair enzymes, permitting the accumulation of genetic damage that would normally be repaired. However, the molecular pathogenesis of HTLV-1-induced neoplasia is not fully understood.

FEATURES OF HTLV-1 INFECTION

Epidemiology HTLV-1 infection is transmitted in at least three ways: from mother to child, especially via breast milk; through sexual activity, more commonly from men to women; and through the blood—via contaminated transfusions or contaminated needles. The virus is most commonly transmitted perinatally. Compared with HIV, which can be transmitted in cell-free form, HTLV-1 is less infectious, and its transmission usually requires cell-to-cell contact.

HTLV-1 is endemic in southwestern Japan and Okinawa, where >1 million persons are infected. Antibodies to HTLV-1 are present in the serum of up to 35% of Okinawans, 10% of residents of the Japanese island of Kyushu, and <1% of persons in nonendemic regions of Japan. Despite this high prevalence of infection, only ~500 cases of ATL are diagnosed in this area each year. Clusters of infection have been noted in other areas of eastern Asia, such as Taiwan; in the Caribbean basin, including northeastern South America; in northwestern South America; in central and southern Africa; in Italy, Israel, Iran, and Papua New Guinea; in the Arctic; and in the southeastern part of the United States **(Fig. 207-4)**. An estimated 10–20 million persons have HTLV-1 infection worldwide.

Progressive spastic or ataxic myelopathy developing in an individual who is HTLV-1 positive (i.e., who has serum antibodies to HTLV-1) may be due to direct infection of the nervous system with the virus, but destruction of the pyramidal tracts appears to involve HTLV-1-infected CD4+ T cells; a similar disorder may result from infection with HIV or HTLV-2. In rare instances, patients with HAM are seronegative but have detectable antibody to HTLV-1 in cerebrospinal fluid (CSF).

The cumulative lifetime risk of developing ATL is 3% among HTLV-1-infected patients, with a threefold greater risk among men than among women; a similar cumulative risk is projected for HAM (4%), but with women more commonly affected than men. The distribution of these two diseases overlaps the distribution of HTLV-1, with >95%

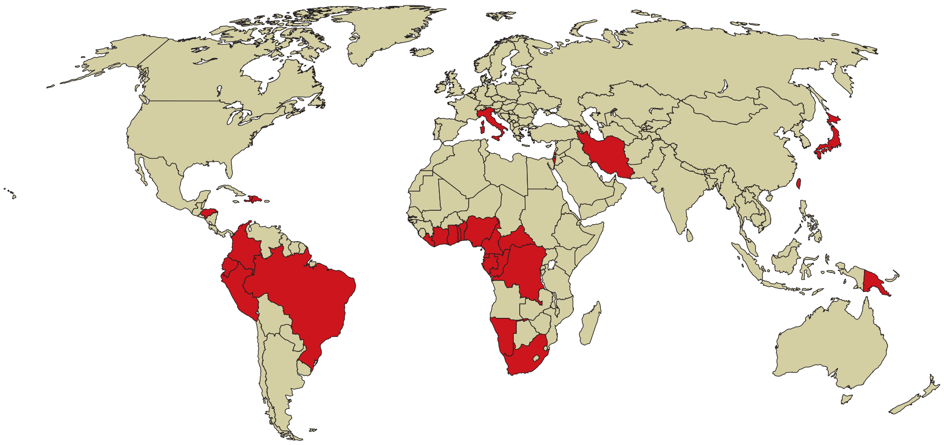

FIGURE 207-4 Global distribution of HTLV-1 infection. Countries with a prevalence of HTLV-1 infection of 1–5% are shaded darkly. Note that the distribution of infected patients is not uniform in endemic countries. For example, the people of southwestern Japan and northeastern Brazil are more commonly affected than those in other regions of those countries.

of affected patients showing serologic evidence of HTLV-1 infection. The latency period between infection and the emergence of disease is 20–30 years for ATL. For HAM, the median latency period is ~3.3 years (range, 4 months to 30 years). The development of ATL is rare among persons infected by blood products; however, ~20% of patients with HAM acquire HTLV-1 from contaminated blood. ATL is more common among perinatally infected individuals, whereas HAM is more common among persons infected via sexual transmission.

Associated Diseases • **ATL** Four clinical types of HTLV-1-induced neoplasia have been described: acute, lymphomatous, chronic, and smoldering. All of these tumors are monoclonal proliferations of CD4+ postthymic T cells with clonal proviral integrations and clonal T-cell receptor gene rearrangements.

ACUTE ATL About 60% of patients who develop malignancy have classic acute ATL, which is characterized by a short clinical prodrome (~2 weeks between the first symptoms and the diagnosis) and an aggressive natural history (median survival period, 6 months). The clinical picture is dominated by rapidly progressive skin lesions, pulmonary involvement, hypercalcemia, and lymphocytosis with cells containing lobulated or "flower-shaped" nuclei **(see Fig. 113-7)**. The malignant cells have monoclonal proviral integrations and express CD4, CD3, and CD25 (low-affinity IL-2 receptors) on their surface. Serum levels of CD25 can be used as a tumor marker. Anemia and thrombocytopenia are rare. The skin lesions may be difficult to distinguish from those in mycosis fungoides. Lytic bone lesions, which are common, do not contain tumor cells but rather are composed of osteolytic cells, usually without osteoblastic activity. Despite the leukemic picture, bone marrow involvement is patchy in most cases.

The hypercalcemia of ATL is multifactorial; the tumor cells produce osteoclast-activating factors (tumor necrosis factor α, IL-1, lymphotoxin) and can also produce a parathyroid hormone–like molecule. Affected patients have an underlying immunodeficiency that makes them susceptible to opportunistic infections similar to those seen in patients with AIDS **(Chap. 208)**. The pathogenesis of the immunodeficiency is unclear. Pulmonary infiltrates in ATL patients reflect leukemic infiltration half the time and opportunistic infections with organisms such as *Pneumocystis* and other fungi the other half. Gastrointestinal symptoms are nearly always related to opportunistic infection. *Strongyloides stercoralis* is a gastrointestinal parasite that has a pattern of endemic distribution similar to that of HTLV-1. HTLV-1-infected persons also infected with this parasite may develop ATL more often or more rapidly than those without *Strongyloides* infections. Serum concentrations of lactate dehydrogenase and alkaline phosphatase are often elevated in ATL. About 10% of patients have leptomeningeal involvement leading to weakness, altered mental status, paresthesia, and/or headache. Unlike other forms of central nervous system (CNS) lymphoma, ATL may be accompanied by normal CSF protein levels. The diagnosis depends on finding ATL cells in the CSF **(Chap. 113)**.

LYMPHOMATOUS ATL The lymphomatous type of ATL occurs in ~20% of patients and is similar to the acute form in its natural history and clinical course, except that circulating abnormal cells are rare and lymphadenopathy is evident. The histology of the lymphoma is variable but does not influence the natural history. In general, the diagnosis is suspected on the basis of the patient's birthplace (see "Epidemiology," above) and the presence of skin lesions and hypercalcemia. The diagnosis is confirmed by the detection of antibodies to HTLV-1 in serum.

CHRONIC ATL Patients with the chronic form of ATL generally have normal serum levels of calcium and lactate dehydrogenase and no involvement of the CNS, bone, or gastrointestinal tract. The median duration of survival for these patients is 2 years. In some cases, chronic ATL progresses to the acute form of the disease.

SMOLDERING ATL Fewer than 5% of patients have the smoldering form of ATL. In this form, the malignant cells have monoclonal proviral integration; <5% of peripheral-blood cells exhibit typical morphologic abnormalities; hypercalcemia, adenopathy, and hepatosplenomegaly do not develop; the CNS, the bones, and the gastrointestinal tract are not involved; and skin lesions and pulmonary lesions may be present. The median survival period for this small subset of patients appears to be ≥5 years.

HAM (TROPICAL SPASTIC PARAPARESIS) In contrast to ATL, in which there is a slight predominance of male patients, HAM affects female patients disproportionately. HAM resembles multiple sclerosis in certain ways **(Chap. 455)**. The onset is insidious. Symptoms include weakness or stiffness in one or both legs, back pain, and urinary incontinence. Sensory changes are usually mild, but peripheral neuropathy may develop. The disease generally takes the form of slowly progressive and unremitting thoracic myelopathy; one-third of patients are bedridden within 10 years of diagnosis, and one-half are unable to walk unassisted by this point. Patients display spastic paraparesis or paraplegia with hyperreflexia, ankle clonus, and extensor plantar responses. Cognitive function is usually spared; cranial nerve abnormalities are unusual.

Magnetic resonance imaging (MRI) reveals lesions in both the white matter and the paraventricular regions of the brain as well as in the spinal cord. Pathologic examination of the spinal cord shows symmetric degeneration of the lateral columns, including the corticospinal tracts; some cases involve the posterior columns as well. The spinal meninges and cord parenchyma contain an inflammatory infiltrate that includes CD8+ T cells with myelin destruction.

HTLV-1 is not usually found in cells of the CNS but may be detected in a small population of lymphocytes present in the CSF. In general, HTLV-1 replication is greater in HAM than in ATL, and patients with HAM have a stronger immune response to the virus. Antibodies to HTLV-1 are present in the serum and appear to be produced in the CSF of HAM patients, where titers are often higher than in the serum. The pathophysiology of HAM may involve the induction of autoimmune destruction of neural cells by T cells with specificity for viral components such as Tax or Env proteins. One theory is that susceptibility to HAM may be related to the presence of human leukocyte antigen (HLA) alleles capable of presenting viral antigens in a fashion that leads to autoimmunity. Insufficient data are available to confirm an HLA association. However, antibodies in the sera of HAM patients have been shown to bind a neuron-specific antigen (heteronuclear ribonuclear protein A1 [hnRNP A1]) and to interfere with neurotransmission in vitro.

It is unclear what factors influence whether HTLV-1 infection will cause disease and, if it does, whether it will induce a neoplasm (ATL) or an autoimmune disorder (HAM). Differences in viral strains, the susceptibility of particular MHC haplotypes, the route of HTLV-1 infection, the viral load, and the nature of the HTLV-1-related immune response are putative factors, but few definitive data are available.

OTHER PUTATIVE HTLV-1-RELATED DISEASES Even in the absence of the full clinical picture of HAM, bladder dysfunction is common in HTLV-1-infected women. In areas where HTLV-1 is endemic, diverse inflammatory and autoimmune diseases have been attributed to the virus, including uveitis, dermatitis, pneumonitis, rheumatoid arthritis, and polymyositis. However, a causal relationship between HTLV-1 and these illnesses has not been established.

Prevention Women in endemic areas should not breast-feed their children, and blood donors should be screened for serum antibodies to HTLV-1. As in the prevention of HIV infection, the practice of safe sex and the avoidance of needle sharing are important.

TREATMENT

HTLV-1 Infection

For the small number of patients who develop HTLV-1–related disease, therapies are not curative. In patients with the acute and lymphomatous types of ATL, the disease progresses rapidly. Hypercalcemia is generally controlled by glucocorticoid administration and cytotoxic therapy directed against the neoplasm. The tumor is highly responsive to combination chemotherapy that is used against other forms of lymphoma; however, patients are susceptible to overwhelming bacterial and opportunistic infections, and ATL relapses within 4–10 months after remission in most cases (Fig. 207-5). The combination of interferon α and zidovudine may extend survival. Because viral replication is not clearly associated with ATL progression, zidovudine is probably effective through its cytotoxic

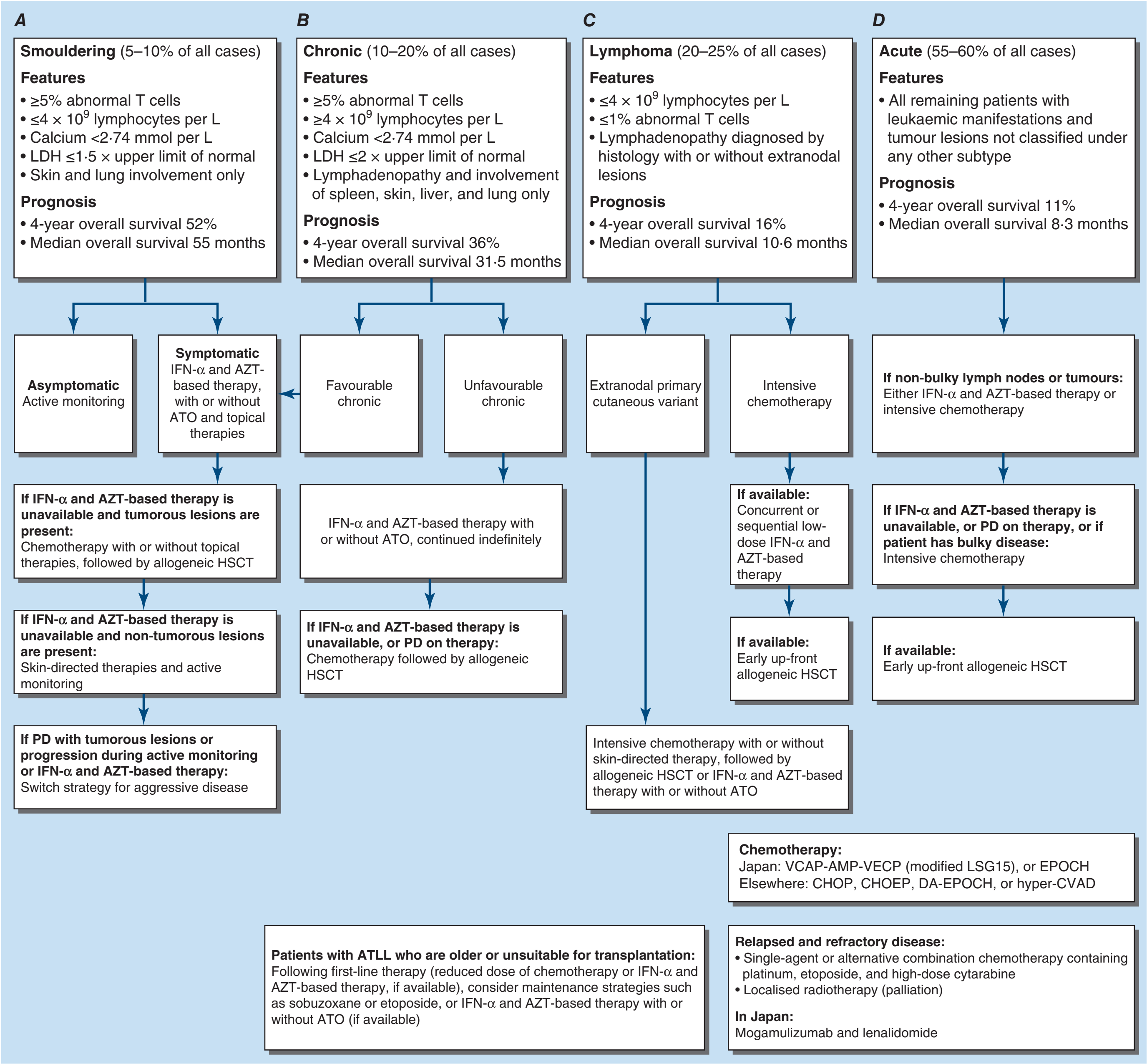
A
Smouldering (5–10% of all cases)
Features
• ≥5% abnormal T cells
• ≤4 × 10^9 lymphocytes per L
• Calcium <2·74 mmol per L
• LDH ≤1·5 × upper limit of normal
• Skin and lung involvement only
Prognosis
• 4-year overall survival 52%
• Median overall survival 55 months
B
Chronic (10–20% of all cases)
Features
• ≥5% abnormal T cells
• ≥4 × 10^9 lymphocytes per L
• Calcium <2·74 mmol per L
• LDH ≤2 × upper limit of normal
• Lymphadenopathy and involvement of spleen, skin, liver, and lung only
Prognosis
• 4-year overall survival 36%
• Median overall survival 31·5 months
C
Lymphoma (20–25% of all cases)
Features
• ≤4 × 10^9 lymphocytes per L
• ≤1% abnormal T cells
• Lymphadenopathy diagnosed by histology with or without extranodal lesions
Prognosis
• 4-year overall survival 16%
• Median overall survival 10·6 months
D
Acute (55–60% of all cases)
Features
• All remaining patients with leukaemic manifestations and tumour lesions not classified under any other subtype
Prognosis
• 4-year overall survival 11%
• Median overall survival 8·3 months
Asymptomatic
Active monitoring
Symptomatic
IFN-α and AZT-based therapy, with or without ATO and topical therapies
Favourable chronic
Unfavourable chronic
Extranodal primary cutaneous variant
Intensive chemotherapy
If non-bulky lymph nodes or tumours: Either IFN-α and AZT-based therapy or intensive chemotherapy
If IFN-α and AZT-based therapy is unavailable and tumorous lesions are present: Chemotherapy with or without topical therapies, followed by allogeneic HSCT
IFN-α and AZT-based therapy with or without ATO, continued indefinitely
If available: Concurrent or sequential low-dose IFN-α and AZT-based therapy
If IFN-α and AZT-based therapy is unavailable, or PD on therapy, or if patient has bulky disease: Intensive chemotherapy
If IFN-α and AZT-based therapy is unavailable and non-tumorous lesions are present: Skin-directed therapies and active monitoring
If IFN-α and AZT-based therapy is unavailable, or PD on therapy: Chemotherapy followed by allogeneic HSCT
If available: Early up-front allogeneic HSCT
If available: Early up-front allogeneic HSCT
If PD with tumorous lesions or progression during active monitoring or IFN-α and AZT-based therapy: Switch strategy for aggressive disease
Intensive chemotherapy with or without skin-directed therapy, followed by allogeneic HSCT or IFN-α and AZT-based therapy with or without ATO
Chemotherapy:
Japan: VCAP-AMP-VECP (modified LSG15), or EPOCH
Elsewhere: CHOP, CHOEP, DA-EPOCH, or hyper-CVAD
Patients with ATLL who are older or unsuitable for transplantation: Following first-line therapy (reduced dose of chemotherapy or IFN-α and AZT-based therapy, if available), consider maintenance strategies such as sobuzoxane or etoposide, or IFN-α and AZT-based therapy with or without ATO (if available)
Relapsed and refractory disease:
• Single-agent or alternative combination chemotherapy containing platinum, etoposide, and high-dose cytarabine
• Localised radiotherapy (palliation)
In Japan:
Mogamulizumab and lenalidomide

FIGURE 207-5 An approach to HTLV-1-associated adult T-cell leukemia/lymphoma. ATLL, adult T-cell leukemia or lymphoma; ATO, arsenic trioxide; AZT, azidothymidine; CHOEP, cyclophosphamide, vincristine, doxorubicin, etoposide, and prednisolone; CHOP, cyclophosphamide, vincristine, doxorubicin, and prednisolone; DA-EPOCH, dose-adjusted etoposide, prednisolone, vincristine, and cyclophosphamide; HSCT, hematopoietic stem cell transplantation; Hyper-CVAD, cyclophosphamide, vincristine, doxorubicin, and dexamethasone with alternating high-dose methotrexate and cytarabine; LDH, lactate dehydrogenase; PD, progressive disease; VCAP-AMP-VECP, vincristine, cyclophosphamide, doxorubicin, and prednisolone; doxorubicin, ranimustine, and prednisolone; vindesine, etoposide, carboplatin, and prednisolone. *(Reproduced with permission from JS O'Donnell et al: Integrated molecular and immunological features of human T-lymphotropic virus type 1 infection and disease progression to adult T-cell leukaemia or lymphoma. Lancet Haematology 10:e539, 2023.)*

effects (as a chain-terminating thymidine analogue) rather than its antiviral effects. Selected series have reported high rates of response and a 40% rate of 5-year survival; however, this level of response has not been universal. LSG15, a multidrug chemotherapy program developed in Japan, induces complete responses in about one-third of patients, about half of whom survive for >2 years; however, the median survival time is about 13 months. High-dose therapy with bone marrow transplantation has been widely tested in Japan. Median survival has not been influenced by this treatment; however, up to 25% of patients survive free of disease for 4 years. Lenalidomide has been reported to have a 42% response rate in patients with relapsed ATL, extending median survival to 20 months despite a short 4-month progression-free survival period. Mogamulizumab, an antibody to CCR4 (a receptor for a number of chemokines, including RANTES and TARC), improved response rates when added to chemotherapy. An experimental approach using an yttrium-90-labeled or toxin-conjugated antibody to the IL-2 receptor appears promising but is not widely available. Patients with the chronic or smoldering form of ATL may be managed with an expectant approach: treat any infections, and watch and wait for signs of progression to acute disease.

Patients with HAM may obtain some benefit from the use of glucocorticoids to reduce inflammation during acute episodes of myelopathy. Antiretroviral regimens have not been effective. In one study, danazol (200 mg three times daily) produced significant neurologic improvement in five of six treated patients, with resolution of urinary incontinence in two cases, decreased spasticity in three, and restoration of the ability to walk after confinement to a wheelchair in two. Antibody to IL-15 receptor β chain has been tested with some promising clinical effects in small numbers of patients. Physical therapy and rehabilitation are important components of management.

FEATURES OF HTLV-2 INFECTION

Epidemiology HTLV-2 is endemic in certain Native American tribes and in Africa. It is generally considered to be a New World virus that was brought from Asia to the Americas 10,000–40,000 years ago during the migration of infected populations across the Bering land bridge. The mode of transmission of HTLV-2 is probably the same as that of HTLV-1 (see above). HTLV-2 may be less readily transmitted sexually than HTLV-1.

Studies of large cohorts of injection drug users with serologic assays that reliably distinguish HTLV-1 from HTLV-2 indicated that the vast majority of HTLV-positive cohort members were infected with HTLV-2. The seroprevalence of HTLV in a cohort of 7841 injection drug users from drug treatment centers in Baltimore, Chicago, Los Angeles, New Jersey (Asbury Park and Trenton), New York City (Brooklyn and Harlem), Philadelphia, and San Antonio was 20.9%, with >97% of cases due to HTLV-2. The seroprevalence of HTLV-2 was higher in the Southwest and the Midwest than in the Northeast. In contrast, the seroprevalence of HIV-1 was higher in the Northeast than in the Southwest or the Midwest. Approximately 3% of the cohort members were infected with both HTLV-2 and HIV-1. The seroprevalence of HTLV-2 increased linearly with age. Women were significantly more likely than men to be infected with HTLV-2; the virus is thought to be more efficiently transmitted from male to female than from female to male.

Associated Diseases Although HTLV-2 was isolated from a patient with a T-cell variant of hairy cell leukemia, this virus has not been consistently associated with a particular disease and in fact has been thought of as "a virus searching for a disease." However, evidence is accumulating that HTLV-2 may play a role in certain neurologic, hematologic, and dermatologic diseases. These data require confirmation, particularly in light of the previous confusion regarding the relative prevalences of HTLV-1 and HTLV-2 among injection drug users.

Prevention Avoidance of needle sharing, adherence to safe-sex practices, screening of blood (by assays for HTLV-1, which also detect HTLV-2), and avoidance of breast-feeding by infected women are important principles in the prevention of spread of HTLV-2.

HUMAN IMMUNODEFICIENCY VIRUS

HIV-1 and HIV-2 are members of the lentivirus subfamily of Retroviridae and are the only lentiviruses known to infect humans. The lentiviruses are slower-acting than viruses that cause acute infection (e.g., influenza virus) but not than other retroviruses. The features of acute primary infection with HIV resemble those of more classic acute infections. The characteristic chronicity of HIV disease is consistent with the designation *lentivirus*. **For a detailed discussion of HIV, see Chap. 208.**

FURTHER READING

Forlani G et al: HTLV-1 infection and pathogenesis: New insights from cellular and animal models. Int J Mol Sci 22:8001, 2021.

Katsuya H et al: Treatment and survival among 1594 patients with ATL. Blood 126:2570, 2015.

Letafati A et al: Therapeutic approaches for HTLV-1-associated adult T-cell leukemia/lymphoma: A comprehensive review. Med Oncol 40:295, 2023.

Ma G et al: Multifaceted functions and roles of HBZ in HTLV-1 pathogenesis. Retrovirology 13:16, 2016.

Moir S et al: Pathogenic mechanisms of HIV disease. Annu Rev Pathol 6:223, 2011.

Ohmoto A, Fuji S: Non-cancerous complications in HTLV-1 carriers. Expert Rev Anti Infect Ther 22:307, 2024.

Yamauchi J et al: An update on human T-cell leukemia virus type I (HTLV-1)-associated myelopathy/tropical spastic paraparesis (HAM/TSP) focusing on clinical and laboratory biomarkers. Pharmacol Ther 218:107669, 2021.

208

Human Immunodeficiency Virus Disease: AIDS and Related Disorders

Anthony S. Fauci, Gregory K. Folkers,
H. Clifford Lane

The Acquired Immune Deficiency Syndrome (AIDS) was first recognized in the United States in the summer of 1981, when the U.S. Centers for Disease Control and Prevention (CDC) reported the unexplained occurrence of *Pneumocystis jirovecii* (formerly *P. carinii*) pneumonia in five previously healthy homosexual men in Los Angeles and of Kaposi's sarcoma (KS) with or without *P. jirovecii* pneumonia and other opportunistic infections in 26 previously healthy homosexual men in New York, San Francisco, and Los Angeles. The disease was soon recognized in male and female injection drug users; in hemophiliacs and blood transfusion recipients; among female sexual partners of men with AIDS; and among infants born to mothers with AIDS. In 1983, what became known as human immunodeficiency virus (HIV) was isolated from a patient with lymphadenopathy, and by 1984 HIV was demonstrated clearly to be the causative agent of AIDS. In 1985, a sensitive enzyme-linked immunosorbent assay (ELISA) was developed to detect antibodies to HIV. This led to an appreciation of the scope and evolution of the HIV epidemic at first in the United States and other developed nations and ultimately among developing nations throughout the world (see "HIV Infection and AIDS Worldwide," below). The staggering worldwide evolution of the HIV pandemic has been matched by an explosion of information in the areas of HIV virology, pathogenesis (both immunologic and virologic), treatment of

HIV infection, treatment and prophylaxis of the opportunistic diseases associated with HIV infection, and prevention of HIV infection. The information flow related to HIV disease is enormous and continues to expand, and it has become almost impossible for the health care generalist to stay abreast of the literature. The purpose of this chapter is to present the most current information available on the scope of the pandemic, as well as the pathogenesis, treatment, and prevention of HIV disease. Above all, the aim is to provide a solid scientific basis and practical clinical guidelines for a state-of-the-art approach to the care of persons with HIV.

■ DEFINITION

The current CDC classification system for HIV infection and AIDS categorizes patients based on clinical conditions associated with HIV infection together with the level of the CD4+ T lymphocyte count. A confirmed HIV case can be classified in one of three clinical stages (A, B, or C) and one of three CD4+ T lymphocyte categories (1, 2, or 3). Advanced HIV disease (AIDS) is classified as stage C, requiring the diagnosis of one or more specific opportunistic illnesses **(Table 208-1)**, and as stage 3 in anyone over 6 years of age if the CD4+ T lymphocyte count is below 200 cells/μL **(Table 208-2)**.

The definition and staging criteria of AIDS are complex and comprehensive and were established for surveillance purposes rather than for the practical care of patients. Thus, the clinician should not focus on whether the patient fulfills the strict definition of AIDS but should view HIV disease as a spectrum ranging from primary infection, with or without the acute syndrome, to the relatively asymptomatic stage, to advanced stages associated with opportunistic diseases (see "Pathophysiology and Pathogenesis," below).

TABLE 208-1 CDC Stage 3 (AIDS)-Defining Opportunistic Illnesses in HIV Infection

Bacterial infections, multiple or recurrent[a]
Candidiasis of bronchi, trachea, or lungs
Candidiasis of esophagus
Cervical cancer, invasive[b]
Coccidioidomycosis, disseminated or extrapulmonary
Cryptococcosis, extrapulmonary
Cryptosporidiosis, chronic intestinal (>1 month's duration)
Cytomegalovirus disease (other than liver, spleen, or nodes), onset at age >1 month
Cytomegalovirus retinitis (with loss of vision)
Encephalopathy attributed to HIV
Herpes simplex: chronic ulcers (>1 month's duration) or bronchitis, pneumonitis, or esophagitis (onset at age >1 month)
Histoplasmosis, disseminated or extrapulmonary
Isosporiasis, chronic intestinal (>1 month's duration)
Kaposi's sarcoma
Lymphoma, Burkitt's (or equivalent term)
Lymphoma, immunoblastic (or equivalent term)
Lymphoma, primary, of brain
Mycobacterium avium complex or *Mycobacterium kansasii*, disseminated or extrapulmonary
Mycobacterium tuberculosis of any site, pulmonary,[b] disseminated, or extrapulmonary
Mycobacterium, other species or unidentified species, disseminated or extrapulmonary
Pneumocystis jirovecii (previously known as *Pneumocystis carinii*) pneumonia
Pneumonia, recurrent[b]
Progressive multifocal leukoencephalopathy
Salmonella septicemia, recurrent
Toxoplasmosis of brain, onset at age >1 month
Wasting syndrome attributed to HIV

[a] Only among children age <6 years. [b] Only among adults, adolescents, and children age ≥6 years.

Source: MMWR 63(RR-03), April 11, 2014.

TABLE 208-2 CDC HIV Infection Stages 1–3 Based on Age-Specific CD4+ T Lymphocyte Count or CD4+ T Lymphocyte Percentage of Total Lymphocytes[a]

	AGE ON DATE OF CD4 T+ LYMPHOCYTE TEST					
	<1 YEAR		1–5 YEARS		6 YEARS THROUGH ADULT	
STAGE[a]	CELLS/μL	%	CELLS/μL	%	CELLS/μL	%
1	≥1500	≥34	≥1000	≥30	≥500	≥26
2	750–1499	26–33	500–999	22–29	200–499	14–25
3	<750	<26	<500	<22	<200	<14

[a] The stage is based primarily on the CD4+ T lymphocyte count; the CD4+ T lymphocyte count takes precedence over the CD4+ T lymphocyte percentage, and the percentage is considered only if the count is missing.

Source: MMWR 63(RR-03), April 11, 2014.

ETIOLOGIC AGENT

HIV is the etiologic agent of AIDS; it belongs to the family of human retroviruses (Retroviridae) and the subfamily of lentiviruses **(Chap. 207)**. Nononcogenic lentiviruses cause disease in other animal species, including sheep, horses, goats, cattle, cats, and monkeys. The four retroviruses known to cause human disease belong to two distinct groups: the human T lymphotropic viruses (HTLV)-1 and HTLV-2, which are transforming retroviruses; and the human immunodeficiency viruses, HIV-1 and HIV-2, which cause cytopathic effects either directly or indirectly **(Chap. 207)**. The most common cause of HIV disease throughout the world, and certainly in the United States, is HIV-1, which comprises several subtypes with different geographic distributions (see "Molecular Heterogeneity of HIV-1," below). HIV-2 was first identified in 1986 in West African patients and was originally confined to West Africa. However, cases traced to West Africa or to sexual contacts with West Africans have been identified throughout the world. HIV-1 strains are broken down into one major group (M) and three minor groups (N, O, P), and HIV-2 viruses are divided into groups A through H. Each group likely derives from a separate transfer to humans from a nonhuman primate reservoir. HIV-1 viruses likely came from chimpanzees and/or gorillas, and HIV-2 from sooty mangabeys. The AIDS pandemic is primarily caused by the HIV-1 M group viruses. Although HIV-1 group O and HIV-2 viruses have been found in numerous countries, including those in the developed world, they have caused much more localized epidemics. Reported infections with group N and group P viruses are rare and confined almost entirely to residents of Cameroon or travelers from Cameroon. The taxonomic relationship between primate lentiviruses is shown in **Fig. 208-1**.

■ MORPHOLOGY OF HIV

Electron microscopy shows that the HIV virion is an icosahedral structure **(Fig. 208-2)** containing numerous external spikes formed by the two major envelope proteins that exist as a trimeric heterodimer, the external gp120 and the transmembrane gp41. The virion buds from the surface of the infected cell (Fig. 208-2*A*) and incorporates a variety of host cellular proteins into its lipid bilayer. The structure of HIV-1 is schematically diagrammed in Fig. 208-2*B*.

■ REPLICATION CYCLE OF HIV

HIV is an RNA virus whose hallmark is the reverse transcription of its genomic RNA to proviral DNA by the enzyme *reverse transcriptase*. The replication cycle of HIV begins with the high-affinity binding via surface-exposed residues within the gp120 protein to its receptor on the host cell surface, the CD4 molecule **(Fig. 208-3)**. The CD4 molecule is a 55-kDa protein found predominantly on a subset of T lymphocytes that are responsible for helper function in the immune system **(Chap. 360)**. Once it binds to CD4, the gp120 protein undergoes a conformational change that facilitates binding to one of two major co-receptors. The two major co-receptors for HIV-1 are CCR5 and CXCR4. Both receptors belong to the family of seven-transmembrane-domain G protein–coupled cellular receptors, and the use of one or the other or both receptors by the virus for entry into the cell is an important determinant of the cellular tropism of the virus. Cell-to-cell

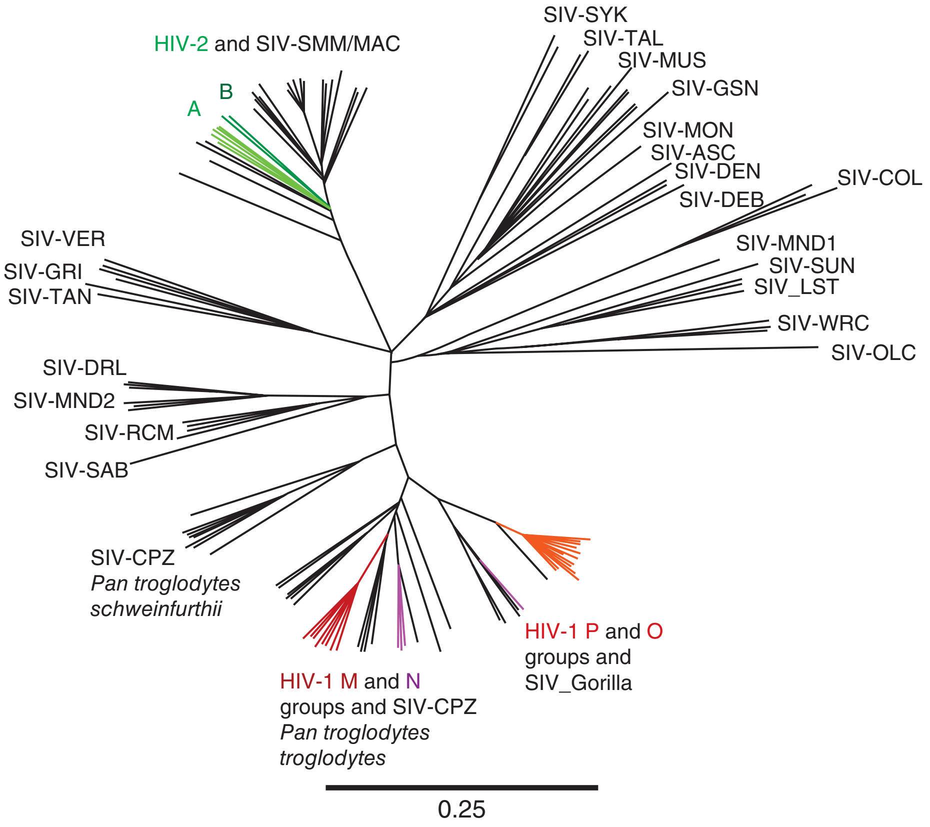

FIGURE 208-1 A phylogenetic tree based on the nearly complete genomes (*gag* through *nef* genes) of primate immunodeficiency viruses. The scale (0.25) indicates a 25% phylogenetically corrected genetic distance at the nucleotide level. Clades in color represent viruses (HIV-1, HIV-2) identified in humans after relatively recent transfers from chimpanzee, gorilla, and sooty mangabey reservoirs. *(Prepared by Brian Foley, PhD, of the HIV Sequence Database, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory; additional information at www.hiv.lanl.gov/content/sequence/HelpDocs/subtypes.html.)*

spread is also facilitated by accessory molecules such as the C-type lectin receptor *DC-SIGN* expressed on certain dendritic cells (DCs) that bind to the HIV gp120 envelope protein, allowing virus captured on DCs to spread to CD4+ T cells. Following binding of the envelope protein to the CD4 molecule associated with the above-mentioned conformational change in the viral envelope gp120, which facilitates binding to a co-receptor, fusion with the host cell membrane occurs as the newly exposed gp41 molecule penetrates the plasma membrane of the target cell and then folds upon itself to bring the virion and target cell together **(Fig. 208-4)**. Following fusion, uncoating of the capsid protein shell is initiated—a step that facilitates reverse transcription and leads to formation of the preintegration complex, composed of viral RNA, enzymes, and accessory proteins and surrounded by capsid and matrix proteins (Fig. 208-3). All these post-fusion viral components constitute the HIV replication complex, including the outer capsid shell, which plays an integral role in supporting reverse transcription of viral RNA. As the preintegration complex traverses the cytoplasm to reach the nucleus, the viral reverse transcriptase enzyme catalyzes the reverse transcription of the genomic RNA into DNA, resulting in the formation of double-stranded HIV proviral DNA. At several steps of the replication cycle, the virus is vulnerable to various cellular factors that can block the progression of infection. The cytoplasmic tripartite motif-containing protein 5α (TRIM5α) is a host restriction factor that interacts with retroviral capsids, causing their premature disassembly and induction of innate immune responses. The apolipoprotein B mRNA editing enzyme (catalytic polypeptide-like 3 [APOBEC3]) family of cellular proteins also inhibits progression of virus infection after virus has entered the cell and prior to entering the nucleus. APOBEC3 proteins, which are incorporated into virions and released into the cytoplasm of a newly infected cell, bind to the single minus-strand DNA intermediate and deaminate viral cytidine, causing hypermutation of retroviral genomes. HIV has evolved a powerful strategy to protect itself from APOBEC. The viral protein Vif targets APOBEC3 for proteasomal degradation. SAMHD1 is another post-entry host factor that prevents reverse transcription by depleting pools of deoxynucleotides (dNTPs). The type I interferon (IFN)-induced myxovirus resistance protein 2 (MX2) is another restriction factor associated with innate immunity that inhibits HIV-1 nuclear entry.

Following reverse transcription, the proviral DNA accesses the nuclear pore and is transferred from the cytoplasm to the nucleus, where it is integrated into a host cell chromosome through the action of another virally encoded enzyme, *integrase* (Fig. 208-3). HIV proviral DNA preferentially integrates into the host genomic DNA in regions of active transcription. This provirus may remain transcriptionally inactive (latent) or may manifest varying levels of gene expression, up to active transcription and production of virus.

Cellular activation plays an important role in the replication cycle of HIV and is critical to the pathogenesis of HIV disease (see “Pathophysiology and Pathogenesis,” below). Following initial binding, fusion, and internalization of the nucleic acid contents of virions into the target cell, incompletely reverse-transcribed DNA intermediates are labile in quiescent cells and do not integrate efficiently into the host cell genome unless cellular activation occurs shortly after infection. Furthermore, some degree of activation of the host cell is required for the initiation of transcription of the integrated proviral DNA into either genomic RNA or mRNA. This latter process may not necessarily be associated with the detectable expression of the classic cell-surface markers of activation. This is reflected in the fact that cell-associated HIV RNA transcribed from competent or defective proviruses can be detected in infected resting CD4+ T cells. In this regard, activation of HIV expression from the latent state depends on the interaction of various cellular and viral factors. Following transcription, HIV mRNA is translated into proteins that undergo modification through glycosylation, myristoylation, phosphorylation, and cleavage. The viral particle is formed by the assembly of HIV proteins, enzymes, and genomic RNA at the plasma membrane of the cells. Budding of the progeny virion through the lipid bilayer of the host cell membrane is the point at which the core acquires its external envelope and where the host restriction factor tetherin can inhibit the release of budding particles. Tetherin is an IFN-induced type II transmembrane protein that interferes with virion detachment. The HIV accessory protein vpu counteracts this effect through direct interactions with tetherin. During or soon after budding, the virally encoded protease catalyzes the cleavage of the gag-pol precursor to yield the mature virion. Progression through the virus replication cycle is profoundly influenced by a variety of viral regulatory gene products. Likewise, each point in the replication cycle of HIV is a real or potential target for therapeutic intervention. Thus far, the reverse transcriptase, protease, and integrase enzymes as well as the process of virus–target cell binding and fusion have proved to be susceptible to pharmacologic disruption.

■ HIV GENOME

Figure 208-5 illustrates schematically the arrangement of the HIV genome. Like other retroviruses, HIV-1 has genes that encode the structural proteins of the virus: *gag* encodes the proteins that form the core of the virion (including p24 antigen); *pol* encodes the enzymes responsible for protease processing of viral proteins, reverse

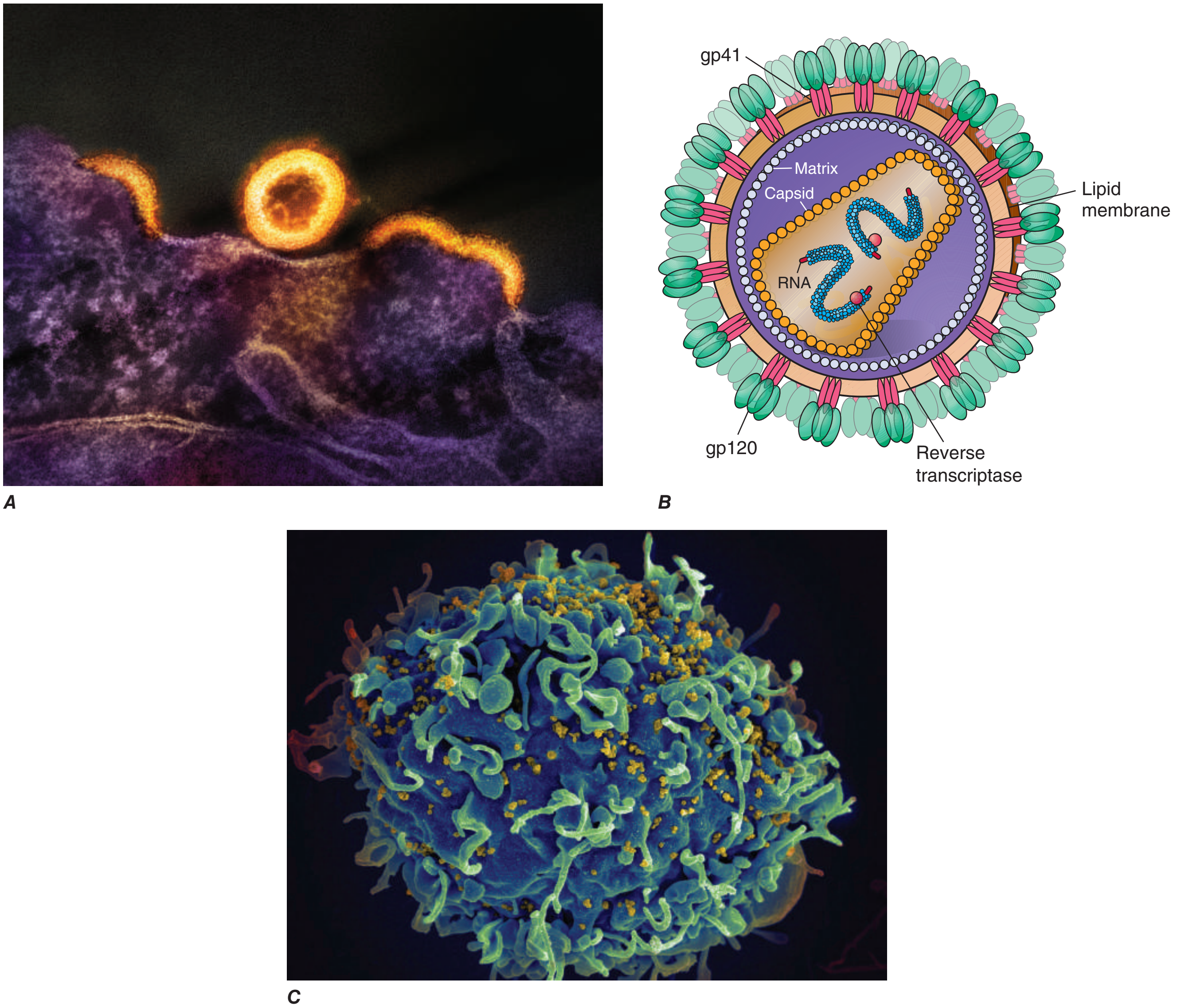

FIGURE 208-2 ***A***. Transmission electron micrograph of HIV-1 virus particles (*colorized yellow/gold*) replicating from an HIV-infected H9 T cell (*purple*). Budding virus particles that have not yet separated from the cell appear as semicircles. A separated, spherical immature particle is seen at the center of the image. *(Image captured at the NIAID Integrated Research Facility [IRF] in Fort Detrick, Maryland. Courtesy of NIAID.)* ***B.*** Structure of HIV-1, including the gp120 envelope, gp41 transmembrane components of the envelope, genomic RNA, enzyme reverse transcriptase, p18(17) inner membrane (matrix), and p24 core protein (capsid). *(Courtesy of George V. Kelvin. Adapted from RC Gallo: Sci Am 256:46, 1987.)* ***C.*** Scanning electron micrograph of a human H9 T cell (blue/green) infected with HIV-1 virus particles (yellow). *(Image captured at NIAID Rocky Mountain Laboratories in Hamilton, Montana. Courtesy of NIAID.)*

transcription, and integration; and *env* encodes the envelope glycoproteins. However, HIV-1 is more complex than other retroviruses, particularly those of the nonprimate group, in that it also contains at least six other regulatory genes (*tat*, *rev*, *nef*, *vif*, *vpr*, and *vpu*), which code for proteins involved in the modification of the host cell to enhance virus growth and the regulation of viral gene expression. Several of these proteins are thought to play a role in the pathogenesis of HIV disease; their various functions are listed in Fig. 208-5. Flanking these genes are long terminal repeats (LTRs), which contain regulatory elements involved in gene expression (Fig. 208-5). The major difference between the genomes of HIV-1 and HIV-2 is the fact that HIV-2 lacks the *vpu* gene and has a *vpx* gene not contained in HIV-1.

MOLECULAR HETEROGENEITY OF HIV-1

Molecular analyses of HIV isolates reveal varying levels of sequence diversity over all regions of the viral genome. For example, the degree of difference in the coding sequences of the viral envelope protein ranges from a few percent (very close, among isolates from the same infected individual) to more than 50% (extreme diversity, between isolates from the different groups of HIV-1: M, N, O, and P). The changes tend to cluster in hypervariable regions. HIV can evolve by several means, including simple base substitution, insertions and deletions, recombination, and gain and loss of glycosylation sites. HIV sequence diversity arises directly from the limited fidelity of the reverse transcriptase, i.e., a tendency toward copying errors. The balance of immune pressure and functional constraints on proteins influences the regional level of variation within proteins. For example, envelope, which is exposed on the surface of the virion and is under immune selective pressure from both antibodies and cytolytic T lymphocytes, is extremely variable, with clusters of mutations in hypervariable domains. In contrast, reverse transcriptase, with important enzymatic functions, is relatively conserved in the absence of drug-induced selective pressure, particularly around the active site. The extraordinary variability of HIV-1 contrasts markedly with the relative stability of HTLV-1 and 2.

The four groups (M, N, O, and P) of HIV-1 are the result of four separate chimpanzee-to-human (or possibly gorilla-to-human for groups O and P) transfers. Group M (major), which is responsible for most infections in the world, has diversified into subtypes and intersubtype recombinant forms, due to “sub-epidemics” within humans after one of those transfers.

Among primate lentiviruses, HIV-1 is most closely related to viruses isolated from chimpanzees and gorillas (Fig. 208-1). The chimpanzee subspecies *Pan troglodytes troglodytes* has been established to be a

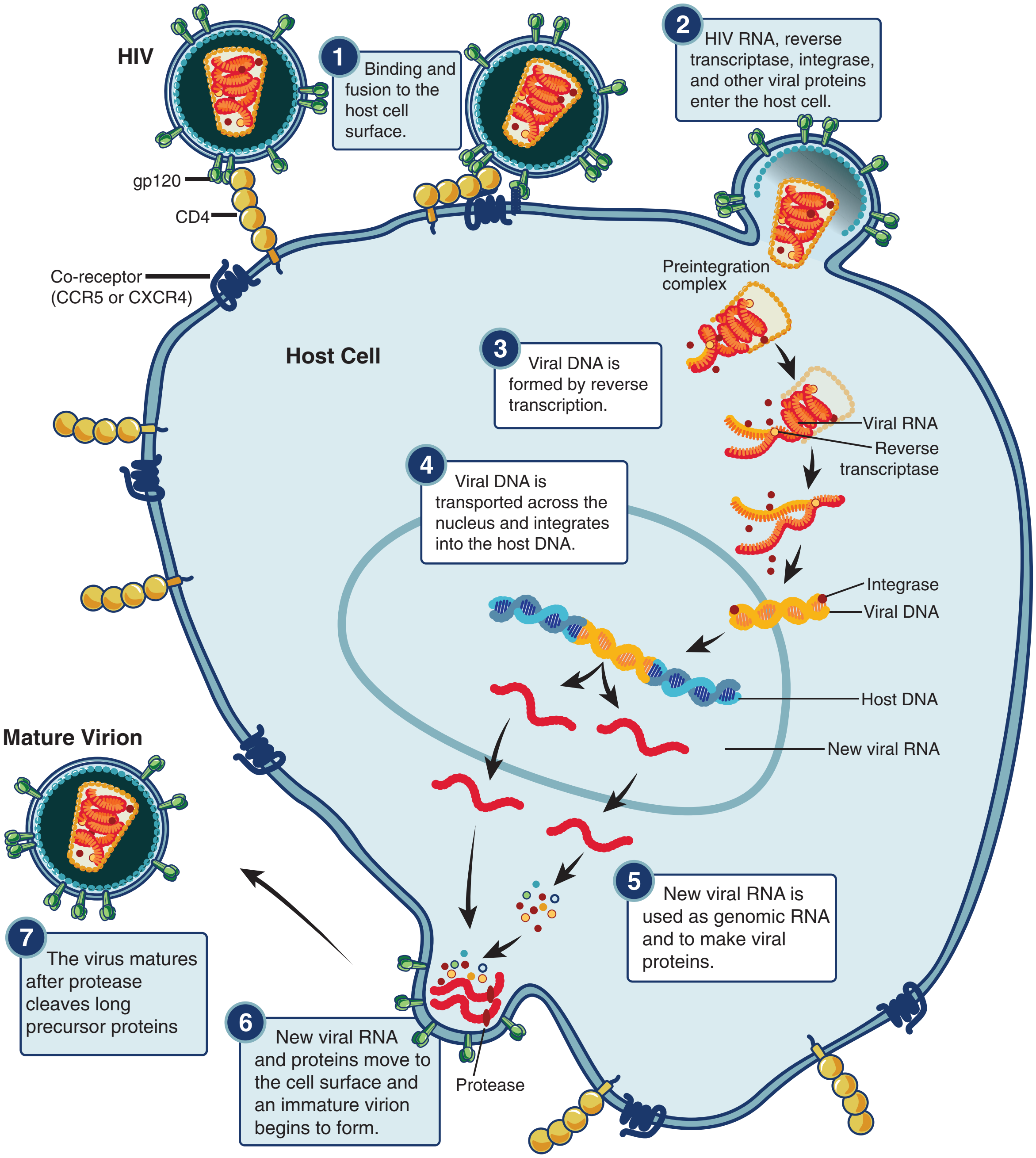
HIV
gp120
CD4
Co-receptor (CCR5 or CXCR4)
1 Binding and fusion to the host cell surface.
2 HIV RNA, reverse transcriptase, integrase, and other viral proteins enter the host cell.
Preintegration complex
Host Cell
3 Viral DNA is formed by reverse transcription.
Viral RNA
Reverse transcriptase
4 Viral DNA is transported across the nucleus and integrates into the host DNA.
Integrase
Viral DNA
Host DNA
New viral RNA
Mature Virion
5 New viral RNA is used as genomic RNA and to make viral proteins.
7 The virus matures after protease cleaves long precursor proteins
6 New viral RNA and proteins move to the cell surface and an immature virion begins to form.
Protease

FIGURE 208-3 The replication cycle of HIV. See text for description. *(From the National Institute of Allergy and Infectious Diseases.)*

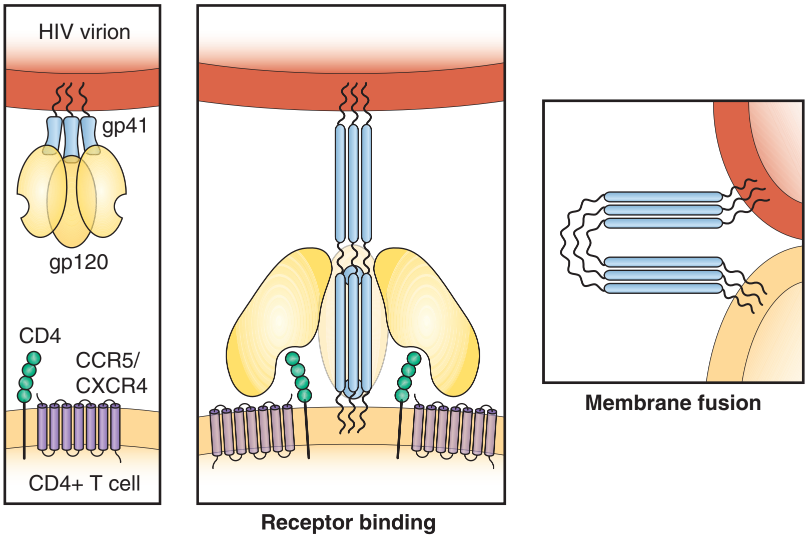
HIV virion
gp41
gp120
CD4
CCR5/ CXCR4
CD4+ T cell
Receptor binding
Membrane fusion

FIGURE 208-4 Binding and fusion of HIV-1 with its target cell. HIV-1 binds to its target cell via the CD4 molecule, leading to a conformational change in the gp120 molecule that allows it to bind to the co-receptor CCR5 (for R5-using viruses). The virus then firmly attaches to the host cell membrane in a coiled-spring fashion via the newly exposed gp41 molecule. Virus–cell fusion occurs as the transitional intermediate of gp41 undergoes further changes to form a hairpin structure that draws the two membranes into close proximity (see text for details). *(Adapted from D Montefiori, JP Moore: HIV vaccines. Magic of the occult? Science 283:336, 1999.)*

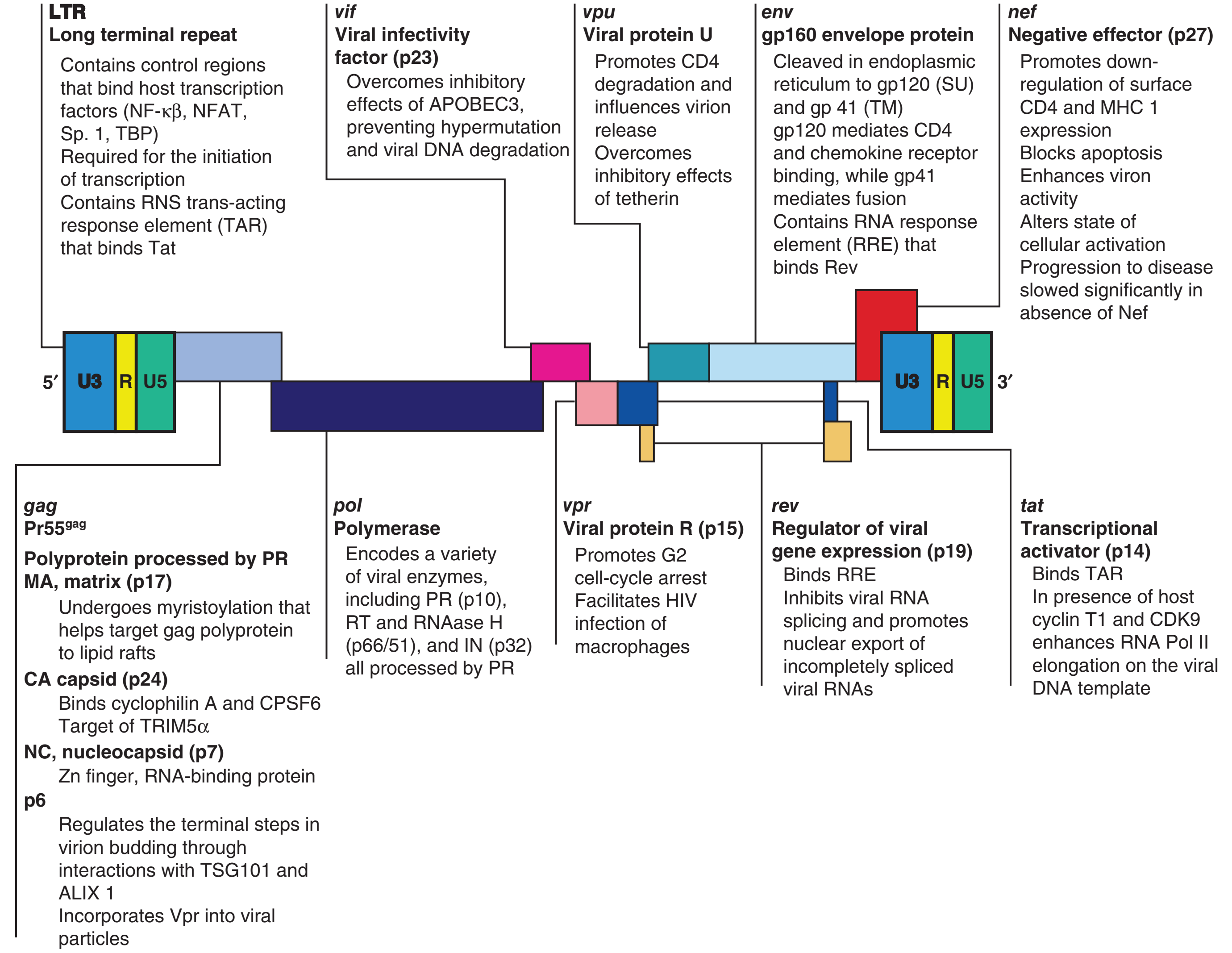
LTR
Long terminal repeat
Contains control regions that bind host transcription factors (NF-κβ, NFAT, Sp. 1, TBP)
Required for the initiation of transcription
Contains RNS trans-acting response element (TAR) that binds Tat
vif
Viral infectivity factor (p23)
Overcomes inhibitory effects of APOBEC3, preventing hypermutation and viral DNA degradation
vpu
Viral protein U
Promotes CD4 degradation and influences virion release
Overcomes inhibitory effects of tetherin
env
gp160 envelope protein
Cleaved in endoplasmic reticulum to gp120 (SU) and gp 41 (TM)
gp120 mediates CD4 and chemokine receptor binding, while gp41 mediates fusion
Contains RNA response element (RRE) that binds Rev
nef
Negative effector (p27)
Promotes down-regulation of surface CD4 and MHC 1 expression
Blocks apoptosis
Enhances viron activity
Alters state of cellular activation
Progression to disease slowed significantly in absence of Nef
5′
U3
R
U5
U3
R
U5
3′
gag
Pr55gag
Polyprotein processed by PR
MA, matrix (p17)
Undergoes myristoylation that helps target gag polyprotein to lipid rafts
CA capsid (p24)
Binds cyclophilin A and CPSF6
Target of TRIM5α
NC, nucleocapsid (p7)
Zn finger, RNA-binding protein
p6
Regulates the terminal steps in virion budding through interactions with TSG101 and ALIX 1
Incorporates Vpr into viral particles
pol
Polymerase
Encodes a variety of viral enzymes, including PR (p10), RT and RNAase H (p66/51), and IN (p32) all processed by PR
vpr
Viral protein R (p15)
Promotes G2 cell-cycle arrest
Facilitates HIV infection of macrophages
rev
Regulator of viral gene expression (p19)
Binds RRE
Inhibits viral RNA splicing and promotes nuclear export of incompletely spliced viral RNAs
tat
Transcriptional activator (p14)
Binds TAR
In presence of host cyclin T1 and CDK9 enhances RNA Pol II elongation on the viral DNA template

FIGURE 208-5 Organization of the genome of the HIV provirus together with a summary description of its 9 genes encoding 15 proteins. *(Reproduced with permission from WC Greene et al: Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. Nat Med 8:673, 2002.)*

natural reservoir of the HIV-1 M and N groups. The rare viruses of the HIV-1 O and P groups are most closely related to viruses found in Cameroonian gorillas. The M group comprises ten subtypes, or *clades*, designated A, B, C, D, F, G, H, J, K, and L, as well as more than 150 known circulating recombinant forms (CRFs) and numerous unique recombinant forms. Inter-subtype recombinants are generated by infection of an individual with two subtypes that then recombine and create a virus with a selective advantage. These CRFs range from highly prevalent forms—such as CRF01_AE, common in southeast Asia, and CRF02_AG in west and central Africa—to numerous CRFs that are relatively rare, either because they are of a more recent origin (newly recombined) or because they have not broken out into a major population. The subtypes and CRFs constitute the major lineages of the M group of HIV-1. HIV-1 M group subtype C dominates the global pandemic, and although there is speculation that it is more transmissible than other subtypes, solid data on variations in transmissibility between subtypes are lacking. Human population densities, access to prevention and treatment, prevalence of genital ulcers, iatrogenic transmissions, and other confounding host factors are all possible reasons why one subtype has spread more than another.

Figure 208-6 schematically diagrams the worldwide distribution of HIV-1 subtypes by region. Nine strains account for most new HIV infections globally: HIV-1 subtypes A, B, C, D, F, and G and three of the CRFs, CRF01_AE, CRF02_AG, and CRF07_BC. Subtype C viruses (of the M group) are by far the most common form, likely accounting for ~50% of infections worldwide. In sub-Saharan Africa, home to approximately two-thirds of all people living with HIV/AIDS, most infections are caused by subtype C, with smaller proportions of infections caused by subtype A, subtype D, CRF02_AG, and other subtypes and recombinants. In South Africa, the country with the largest number of prevalent infections (7.7 million in 2023), 98% of all HIV-1 isolates sequenced are of subtype C. In Asia, HIV-1 isolates of the CRF01_AE lineage and subtypes B and C predominate. CRF01_AE accounts for most infections in south and southeast Asia, while ~97% of infections in India, home to an estimated 2.5 million people with HIV, are of subtype C (see "HIV Infection and AIDS Worldwide," below). Subtype B viruses are overwhelmingly predominant in the United States, Canada, certain countries in South America, western Europe, and Australia. It is thought that, purely by chance, subtype B was seeded into the United States and Europe in the late 1970s, thereby establishing an overwhelming founder effect. Many countries have co-circulating viral subtypes that are giving rise to new CRFs. Sequence analyses of HIV-1 isolates from infected individuals indicate that recombination among viruses of different clades likely occurs when an individual is infected with viruses of more than one subtype, particularly in geographic areas where subtypes overlap, and more often in sub-epidemics driven by injection drug use than in those driven by sexual transmission.

The extraordinary diversity of HIV, reflected by the presence of multiple subtypes, circulating recombinant forms, and continuous viral evolution, has implications for possible differential rates of transmission, rates of disease progression, and the development of resistance to antiretroviral drugs. This diversity may also prove to be a formidable obstacle to HIV vaccine development, as a broadly useful vaccine would need to induce protective responses against a wide range of viral strains.

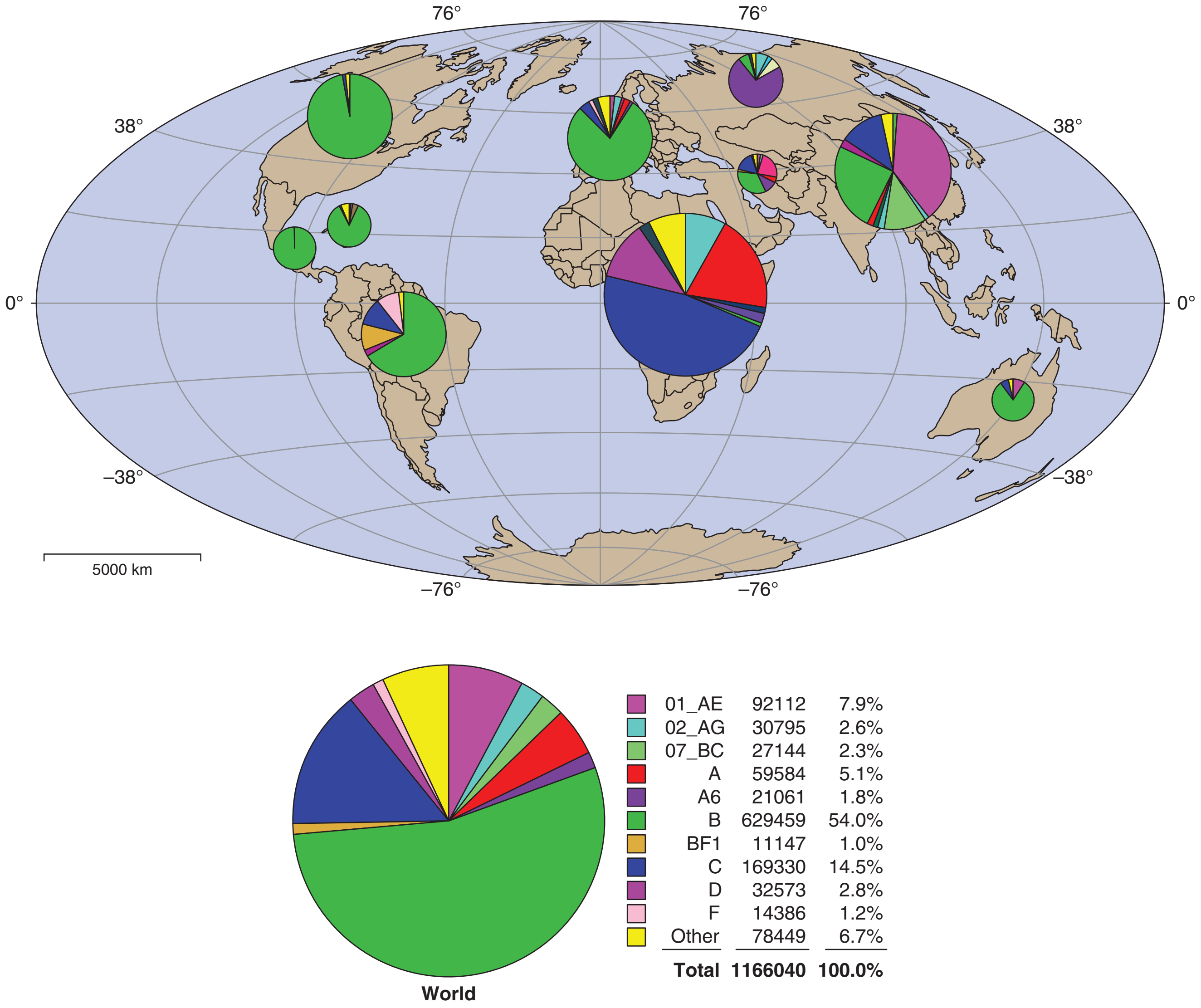

FIGURE 208-6 Global geographic distribution of HIV-1 subtypes and recombinant forms. Distributions derived from relative frequency of subtypes among >1.16 million HIV genomic sequences in the Los Alamos National Laboratory HIV Sequence Database. *(Additional information available at www.hiv.lanl.gov/components/sequence/HIV/geo/geo.comp.)*

TRANSMISSION

HIV is transmitted primarily by sexual contact (both heterosexual and male to male); by blood and blood products; and by infected mothers to infants intrapartum, perinatally, or via breast milk. After four decades of experience and observations, there is no evidence that HIV is transmitted by any other modality. Table 208-3 lists the estimated risk of HIV transmission for various types of exposures.

SEXUAL TRANSMISSION

HIV infection is predominantly a sexually transmitted infection (STI) worldwide. By far the most common mode of infection, particularly in developing countries, is heterosexual transmission, although in many western countries male-to-male sexual transmission dominates. Although a wide variety of factors including viral load and the presence of ulcerative genital diseases influence the efficiency of heterosexual transmission of HIV, such transmission is generally inefficient. A systemic review found a low per-act risk of heterosexual transmission in the absence of antiretrovirals: 0.04% for female-to-male transmission and 0.08% for male-to-female transmission during vaginal intercourse in the absence of antiretroviral therapy or condom use (Table 208-3). As discussed below, this risk approaches zero in settings where the infected partner is on effective antiretroviral therapy or when the uninfected partner is on a program of pre-exposure or postexposure prophylaxis (PrEP or PEP).

HIV has been demonstrated in seminal fluid both within infected mononuclear cells and in cell-free material. The virus appears to

TABLE 208-3 Estimated Per-Act Probability of Acquiring HIV from an Infected Source, By Exposure Act

TYPE OF EXPOSURE	RISK PER 10,000 EXPOSURES
Parenteral	
Blood transfusion	9250
Needle-sharing during injection drug use	63
Percutaneous (needle-stick)	23
Sexual	
Receptive anal intercourse	138
Insertive anal intercourse	11
Receptive penile-vaginal intercourse	8
Insertive penile-vaginal intercourse	4
Receptive oral intercourse	Low
Insertive oral intercourse	Low
Other[a]	
Biting	Negligible
Spitting	Negligible
Throwing body fluids (including semen or saliva)	Negligible
Sharing sex toys	Negligible

[a] HIV transmission through these exposure routes is technically possible but unlikely and not well documented.

Source: CDC, *www.cdc.gov/hivpartners/php/riskandprevention/.*

concentrate in the seminal fluid, particularly in situations where there are increased numbers of lymphocytes and monocytes in the fluid, as seen in genital inflammatory states such as urethritis and epididymitis, conditions closely associated with other STIs. The virus has also been demonstrated in cervical smears and vaginal fluid. There is an elevated risk of HIV transmission associated with unprotected receptive anal intercourse (URAI) among both men and women compared to the risk associated with unprotected receptive vaginal intercourse. Although data are limited, the per-act risk for HIV transmission via URAI has been estimated to be ~1.4% (Table 208-3). The risk of HIV acquisition associated with URAI is higher than that seen in penile–vaginal intercourse probably because only a thin, fragile rectal mucosal membrane separates the deposited semen from potentially susceptible cells in and beneath the mucosa, and microtrauma of the mucosal membrane has been associated with anal intercourse. Anal douching and sexual practices that traumatize the rectal mucosa also increase the likelihood of infection. It is likely that anal intercourse provides at least two modalities of infection: (1) direct inoculation into blood in cases of traumatic tears in the mucosa; and (2) infection of susceptible target cells, such as Langerhans cells, in the mucosal layer in the absence of trauma. Insertive anal intercourse also confers an increased risk of HIV acquisition compared with insertive vaginal intercourse in the receptive partner since the vaginal mucosa is several layers thicker than the rectal mucosa and less likely to be traumatized during intercourse. Nonetheless, the virus can be transmitted to either partner through vaginal intercourse. As noted in Table 208-3, male-to-female HIV transmission is more efficient than female-to-male transmission. The differences in reported transmission rates between men and women may be due in part to the prolonged exposure of the vaginal and cervical mucosa to infected seminal fluid; the endometrium also can be exposed to virus when semen enters through the cervical os. By comparison, the penis and urethral orifice of the uninfected male partner are only exposed relatively briefly to infected vaginal fluid.

Among various cofactors examined in studies of heterosexual HIV transmission, the presence of other STIs has been strongly associated with HIV transmission. In this regard, there is a close association between genital ulcerations and transmission, owing to both susceptibility to infection and infectivity. Infections with microorganisms such as *Treponema pallidum* **(Chap. 187)**, *Haemophilus ducreyi* **(Chap. 162)**, and herpes simplex virus (HSV; **Chap. 197**) are important causes of genital ulcerations linked to transmission of HIV. In addition, pathogens responsible for non-ulcerative inflammatory STIs such as those caused by *Chlamydia trachomatis* **(Chap. 194)**, *Neisseria gonorrhoeae* **(Chap. 161)**, and *Trichomonas vaginalis* **(Chap. 236)** also are associated with an increased risk of transmission of HIV infection. Bacterial vaginosis, an infection related to sexual behavior, but not strictly an STI, also may be linked to an increased risk of transmission of HIV infection. Multiple studies have suggested that treating STIs and genital tract syndromes may help decrease transmission of HIV. This effect is most prominent in populations in which the prevalence of HIV infection is relatively low. It is noteworthy that this principle may not apply to the treatment of HSV infections since it has been shown that even following anti-HSV therapy with resulting healing of HSV-related genital ulcers, HIV acquisition is not reduced. Biopsy studies revealed that the likely explanation is that HIV receptor–positive inflammatory cells persisted in the genital tissue despite the healing of ulcers, and so HIV-susceptible targets remained at the site.

The quantity of HIV-1 in plasma (viral load) is a primary determinant of the risk of HIV-1 transmission. In a cohort of heterosexual couples in Uganda discordant for HIV infection and not receiving antiretroviral therapy, the mean serum HIV RNA level was significantly higher among individuals with HIV whose partners seroconverted than among those whose partners did not seroconvert. In fact, transmission was rare when the infected partner had a plasma level of <1700 copies of HIV RNA per milliliter, even when genital ulcer disease was present **(Fig. 208-7)**. The rate of HIV transmission per coital act was highest during the early stage of HIV infection when plasma HIV RNA levels were high and in advanced disease with high viral set points.

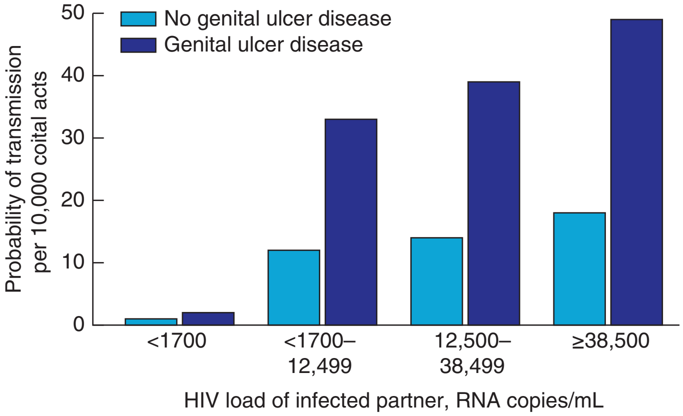

FIGURE 208-7 Probability of HIV transmission per coital act among monogamous, heterosexual, HIV-serodiscordant couples in Uganda. *(From RH Gray et al: Lancet 357:1149, 2001.)*

Antiretroviral therapy dramatically reduces plasma viremia in most people with HIV (see "Antiretroviral Therapy" and "HIV Prevention," below) and is associated with a dramatic reduction in risk of transmission, an approach widely referred to as *treatment as prevention* or TasP. Multiple studies have demonstrated that if the viral load of a person with HIV is reduced by antiretroviral therapy to <20 copies/mL as measured by conventional commercial assays, there is essentially no chance of sexual transmission to the person's sexual partner. This is true for heterosexuals as well as men who have sex with men, leading to the commonly used description of this phenomenon as "undetectable equals untransmittable" or "U = U."

Multiple studies including large, randomized, controlled trials clearly have indicated that male circumcision is associated with a lower risk of acquisition of HIV infection for heterosexual men. Studies also suggest that circumcision is protective against HIV acquisition for men who have sex with men reporting mainly or only insertive sex. The benefit of circumcision may be due to increased susceptibility of uncircumcised men to ulcerative STIs, as well as to other factors such as microtrauma to the foreskin and glans penis. In addition, the highly vascularized inner layer of foreskin tissue contains a high density of Langerhans cells as well as increased numbers of CD4+ T cells, macrophages, and other cellular targets for HIV. Finally, the moist environment under the foreskin may promote the presence or persistence of microbial flora that, via inflammatory changes, may lead to even higher concentrations of target cells for HIV in the foreskin. In addition, randomized clinical trials have demonstrated that male circumcision also reduces HSV type 2, human papillomavirus virus (HPV), and genital ulcer disease in men as well as HPV, genital ulcer disease, bacterial vaginosis, and *Trichomonas vaginalis* infections among female partners of circumcised men. Thus, there may be an added indirect benefit of diminution of risk for HIV acquisition to the female sexual partners of circumcised men.

In some studies, the use of oral contraceptives was associated with an increase in incidence of HIV infection over and above that which might be expected by not using a condom for birth control. This phenomenon may be due to drug-induced changes in the cervical mucosa, rendering it more vulnerable to penetration by the virus. Adolescent girls might also be more susceptible to infection upon exposure due to the properties of an immature genital tract with increased cervical ectopy or exposed columnar epithelium.

Oral sex is a much less efficient mode of transmission of HIV than is anal intercourse or vaginal intercourse (Table 208-3). Multiple studies have reported that the incidence of transmission of infection by oral sex among couples discordant for HIV is extremely low. However, there have been well-documented reports of HIV transmission that likely resulted from fellatio or cunnilingus. Therefore, the assumption that oral sex is completely safe is not warranted.

The association of alcohol consumption and illicit drug use with unsafe sexual behavior, both homosexual and heterosexual, leads to an increased risk of sexual transmission of HIV. Methamphetamine and other so-called "club drugs" such as 3,4-methylenedioxymethamphetamine (MDMA; also known as "ecstasy"), ketamine, gamma-hydroxybutyrate (GHB), and inhaled nitrites (known as "poppers"), sometimes taken in conjunction with PDE-5 inhibitors such as sildenafil (Viagra), tadalafil (Cialis), or vardenafil (Levitra), have been associated with risky sexual practices and increased risk of HIV infection, particularly among men who have sex with men.

■ TRANSMISSION THROUGH INJECTION DRUG USE

HIV can be transmitted to injection drug users (IDUs) who are exposed to HIV while sharing injection paraphernalia such as needles, syringes, the water in which drugs are mixed, or the cotton through which drugs are filtered. Parenteral transmission of HIV during injection drug use does not require IV puncture; subcutaneous ("skin popping") or intramuscular ("muscling") injections can transmit HIV as well, even though these behaviors are sometimes erroneously perceived as low risk. Among IDUs, the risk of HIV infection increases with the duration of injection drug use; the frequency of needle sharing; the number of partners with whom paraphernalia are shared; comorbid psychiatric conditions such as antisocial personality disorder; the use of cocaine in injectable form or smoked as "crack"; and the use of injection drugs in a geographic location with a high prevalence of HIV infection. As noted in Table 208-3, the per-act risk of transmission from injection drug use with a contaminated needle has been estimated to be approximately 0.6%.

■ TRANSMISSION BY TRANSFUSED BLOOD AND BLOOD PRODUCTS

HIV can be transmitted to individuals who receive HIV-contaminated blood transfusions, blood products, or transplanted tissue. The vast majority of HIV infections acquired via contaminated blood transfusions, blood components, or transplanted tissue in resource-rich countries occurred prior to the spring of 1985, when mandatory testing of donated blood for HIV-1 was initiated. It is estimated that >90% of individuals exposed to HIV-contaminated blood products become infected (Table 208-3). Transfusions of whole blood, packed red blood cells, platelets, leukocytes, and plasma are all capable of transmitting HIV infection. In contrast, hyperimmune gamma globulin, hepatitis B immune globulin, plasma-derived hepatitis B vaccine, and Rh$_o$ immune globulin have not been associated with transmission of HIV infection. The procedures involved in processing these products either inactivate or remove the virus.

Currently, in the United States and in most developed countries, the following measures have made the risk of transmission of HIV infection by transfused blood or blood products extremely small: the screening of blood donations for antibodies to HIV-1 and HIV-2 and determination of the presence of HIV nucleic acid usually in minipools of several specimens; the careful selection of potential blood donors with health history questionnaires to exclude individuals with risk behaviors; and opportunities for self-deferral and the screening out of HIV-negative individuals with serologic testing for infections that have shared risk factors with HIV, such as hepatitis B and C and syphilis. The chance of infection of a hemophiliac via clotting factor concentrates has essentially been eliminated because of standard screening of blood together with the added layer of safety resulting from heat treatment of the concentrates. It is currently estimated that the risk of infection with HIV in the United States via transfused screened blood is approximately 1 in 2 million units. Since nearly 21 million blood components are transfused in the United States each year, eliminating the risk of transfusion-related HIV transmission likely will not be possible. Transmission of HIV (both HIV-1 and HIV-2) by blood or blood products is still an ongoing threat in certain developing countries where routine screening of blood is not universally practiced. Furthermore, there have been reports in certain countries of sporadic breakdowns in routinely available screening procedures in which contaminated blood or plasma was transfused, resulting in small clusters of patients becoming infected.

■ OCCUPATIONAL TRANSMISSION OF HIV: HEALTH CARE WORKERS, LABORATORY WORKERS, AND THE HEALTH CARE SETTING

There is a small but definite occupational risk of HIV transmission to health care workers and laboratory personnel and potentially others who work with HIV-containing materials, particularly when sharp objects are used. More than 300,000 health care workers are stuck with needles or other sharp medical instruments in the United States each year. The global number of HIV infections among health care workers attributable to sharps injuries has been estimated to be 1000 cases (range, 200–5000) per year. In the United States, a total of 58 documented cases of occupational HIV transmission to health care workers, and 150 possible transmissions, have been reported by the CDC. Since 1999, only one confirmed case (a laboratory technician sustaining a needle puncture while working with a live HIV culture in 2008) has been reported.

Exposures that place a health care worker at potential risk of HIV infection are percutaneous injuries (e.g., a needle stick or cut with a sharp object) or contact of mucous membrane or nonintact skin (e.g., exposed skin that is chapped, abraded, or afflicted with dermatitis) with blood, tissue, or other potentially infectious body fluids. Large, multi-institutional studies have indicated that the risk of HIV transmission following skin puncture from a needle or a sharp object that was contaminated with blood from a person with documented HIV infection is ~0.23%, and after a mucous membrane exposure it is ~0.09% (see "HIV and the Health Care Worker," below) if the injured and/or exposed person is not treated within 24 hours with antiretroviral drugs. The risk of hepatitis B virus (HBV) infection following a similar type of exposure is ~6–30% in nonimmune individuals. If a susceptible worker is exposed to HBV, postexposure prophylaxis with hepatitis B immune globulin and initiation of HBV vaccine is >90% effective in preventing HBV infection. The risk of HCV infection following percutaneous injury is ~1.8% **(Chap. 350)**.

Rare HIV transmission after nonintact skin exposure has been documented. The average risk for transmission by this route has not been precisely determined; however, it is estimated to be less than the risk for mucous membrane exposure. Transmission of HIV through intact skin has not been documented. All health care workers experiencing a puncture wound or mucous membrane exposures involving blood from a patient with documented HIV infection should be treated prophylactically with combination antiretroviral therapy (ART). This practice, referred to as *postexposure prophylaxis* or PEP, has dramatically reduced the occurrence of puncture-related transmissions of HIV to health care workers.

In addition to blood and visibly bloody body fluids, semen and vaginal secretions also are considered potentially infectious; however, they have not been implicated in occupational transmission from patients to health care workers. The following fluids also are considered potentially infectious: cerebrospinal fluid, synovial fluid, pleural fluid, peritoneal fluid, pericardial fluid, and amniotic fluid. The risk for transmission after exposure to fluids or tissues other than HIV-infected blood has not been quantified, but it is probably considerably lower than the risk after blood exposures. Feces, nasal secretions, saliva, sputum, sweat, tears, urine, and vomitus are not considered potentially infectious for HIV unless they are visibly bloody. Rare cases of HIV transmission via human bites have been reported, but not in the setting of occupational exposure.

An increased risk for HIV infection following percutaneous exposures to HIV-infected blood is associated with exposures involving a relatively large quantity of blood, as in the case of a device visibly contaminated with the patient's blood, a procedure that involves a hollow-bore needle placed directly in a vein or artery, or a deep injury. Factors that might be associated with mucocutaneous transmission of HIV include exposure to an unusually large volume of blood and prolonged contact. In addition, the risk increases for exposures to blood from untreated patients with high levels of HIV in the blood. Since the beginning of the HIV epidemic, there have been rare instances where transmission of infection from a health care worker to patients seemed highly probable. Despite this small number of documented cases, the risk of HIV transmission involving infected health care workers to patients is extremely low in developed countries—in fact, too low to be measured accurately. In this regard, several retrospective epidemiologic studies have been performed tracing thousands of patients of dentists, physicians, surgeons, obstetricians, and gynecologists with HIV, and no cases of HIV transmission that could be linked to the health care providers were identified other than the already identified documented cases.

Breaches in infection control and the reuse of contaminated syringes, failure to properly sterilize surgical instruments, and/or hemodialysis equipment also have resulted rarely in the transmission of HIV from patient to patient in hospitals, nursing homes, and outpatient settings. Finally, these very rare occurrences of transmission of HIV as well as HBV and HCV to and from health care workers in the workplace underscore the importance of the use of universal precautions when caring for all patients (see below and **Chap. 147**).

■ MOTHER-TO-CHILD TRANSMISSION OF HIV

HIV infection can be transmitted from an infected mother to her fetus during pregnancy, during delivery, or by breast-feeding. This remains a persistent form of transmission of HIV infection in certain developing countries. Virologic analyses of aborted fetuses indicate that HIV can be transmitted to the fetus during the first or second trimesters of pregnancy. However, maternal transmission occurs most commonly during birth. Two studies performed in Rwanda and the Democratic Republic of Congo (then called Zaire) indicated that among all mother-to-child transmissions of HIV, the relative proportions were 23–30% before birth, 50–65% during birth, and 12–20% via breast-feeding.

In the absence of antiretroviral therapy for the mother during pregnancy, labor, and delivery, and for the infant prophylactically following birth, the overall probability of transmission of HIV from mother to infant/fetus ranges from 15% to 25% in industrialized countries and from 25% to 35% in developing countries. These differences may relate to the adequacy of prenatal care as well as to the stage of HIV disease and the general health of the mother during pregnancy. Higher rates of transmission have been reported to be associated with many factors—the best documented of which is the presence of high maternal levels of plasma viremia, with the risk increasing linearly with the level of maternal plasma viremia. It is very unlikely that mother-to-child transmission will occur if the mother's level of plasma viremia is <1000 copies of HIV RNA/mL of blood and extremely unlikely if the level is <50 copies/mL. Increased mother-to-child transmission is also correlated with closer human leukocyte antigen (HLA) match between mother and child. A prolonged interval between membrane rupture and delivery is another well-documented risk factor for transmission of HIV. Other conditions that are potential risk factors, but that have not been consistently demonstrated, are the presence of chorioamnionitis at delivery; STIs during pregnancy; illicit drug use during pregnancy; cigarette smoking; preterm delivery; and obstetric procedures such as amniocentesis, amnioscopy, fetal scalp electrodes, and episiotomy. Today, the rate of mother-to-child transmission has fallen to <1% in pregnant women who are receiving ART for their HIV infection. Such treatment, combined with cesarean section delivery, has rendered mother-to-child transmission of HIV an extremely unusual event in the United States and other developed nations. In this regard, both the United States Public Health Service and the World Health Organization guidelines recommend that all pregnant women with HIV receive lifelong ART for the health of the mother (regardless of plasma HIV RNA copy number or CD4+ T cell count) as well as to prevent perinatal transmission.

Breast-feeding is an important modality of transmission of HIV infection in certain developing countries, particularly where untreated mothers continue to breast-feed for prolonged periods. The risk factors for mother-to-child transmission of HIV via breast-feeding by an untreated mother include detectable levels of HIV in breast milk, the presence of mastitis, low maternal CD4+ T cell counts, and maternal vitamin A deficiency. The risk of HIV infection via breast-feeding is highest in the early months of breast-feeding. In addition, exclusive breast-feeding has been reported to carry a lower risk of HIV transmission than mixed feeding. In developed countries, breast-feeding of babies by a mother with HIV is contraindicated since alternative forms of adequate nutrition, i.e., formulas, are readily available. In developing countries, where breast-feeding may be essential for the overall health of the infant, the continuation of ART in the infected mother during the period of breastfeeding markedly diminishes the risk of transmission of HIV to the infant. In fact, treatment of a pregnant woman with ART should be provided for the benefit of the woman as much as for the prevention of mother-to-child transmission and should be continued beyond the pregnancy, for life.

■ TRANSMISSION OF HIV BY OTHER BODY FLUIDS

Although HIV can be isolated typically in low titers from saliva of a small proportion of infected individuals, there is no convincing evidence that saliva can transmit HIV infection, either through kissing or through other exposures, such as occupationally to health care workers. Saliva contains endogenous antiviral factors; among these factors, HIV-specific immunoglobulins of IgA, IgG, and IgM isotypes are detected readily in salivary secretions of infected individuals. It has been suggested that large glycoproteins such as mucins and thrombospondin 1 sequester HIV into aggregates for clearance by the host. In addition, multiple soluble salivary factors inhibit HIV to various degrees in vitro, probably by targeting host cell receptors rather than the virus itself. Perhaps the best studied of these, secretory leukocyte protease inhibitor (SLPI), blocks HIV infection in several cell culture systems, and it is found in saliva at levels that approximate those required for inhibition of HIV in vitro. In this regard, higher salivary levels of SLPI in breast-fed infants were associated with a decreased risk of HIV transmission through breast milk. It has also been suggested that submandibular saliva reduces HIV infectivity by stripping gp120 from the surface of virions, and that saliva-mediated disruption and lysis of HIV-infected cells occurs because of the hypotonicity of oral secretions. Transmission of HIV by a human bite can occur but is a rare event. Although virus can be detected, if not isolated, from virtually any body fluid, there is no evidence that HIV transmission can occur as a result of exposure to tears, sweat, or urine. However, there have been isolated cases of transmission of HIV infection by body fluids that may or may not have been contaminated with blood. Most of these situations occurred in the setting of a close relative providing intensive nursing care for a person with HIV without observing universal precautions, underscoring the importance of adhering to such precautions in the handling of body fluids and wastes from people with HIV.

EPIDEMIOLOGY

■ HIV INFECTION AND AIDS WORLDWIDE

HIV infection/AIDS is a global pandemic, with cases reported from virtually every country. In 2023, an estimated 39.9 million individuals were living with HIV infection, according to the Joint United Nations Programme on HIV/AIDS (UNAIDS). An estimated 95% of people living with HIV/AIDS reside in low- and middle-income countries; ~53% are female, and 1.4 million are children <15 years. The regional distribution of these cases is illustrated in **Fig. 208-8**. The estimated number

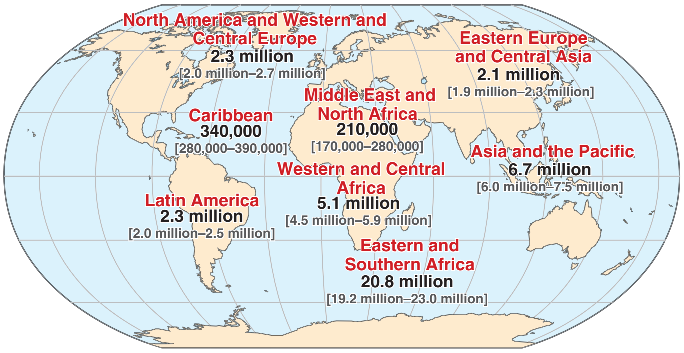
North America and Western and Central Europe
2.3 million
[2.0 million–2.7 million]
Eastern Europe and Central Asia
2.1 million
[1.9 million–2.3 million]
Middle East and North Africa
210,000
[170,000–280,000]
Caribbean
340,000
[280,000–390,000]
Asia and the Pacific
6.7 million
[6.0 million–7.5 million]
Western and Central Africa
5.1 million
[4.5 million–5.9 million]
Latin America
2.3 million
[2.0 million–2.5 million]
Eastern and Southern Africa
20.8 million
[19.2 million–23.0 million]

FIGURE 208-8 Estimated number of adults and children living with HIV infection as of December 2023. Total: 39.9 million (36.1 million–44.6 million). *(From Joint United Nations Programme on HIV/AIDS [UNAIDS].)*

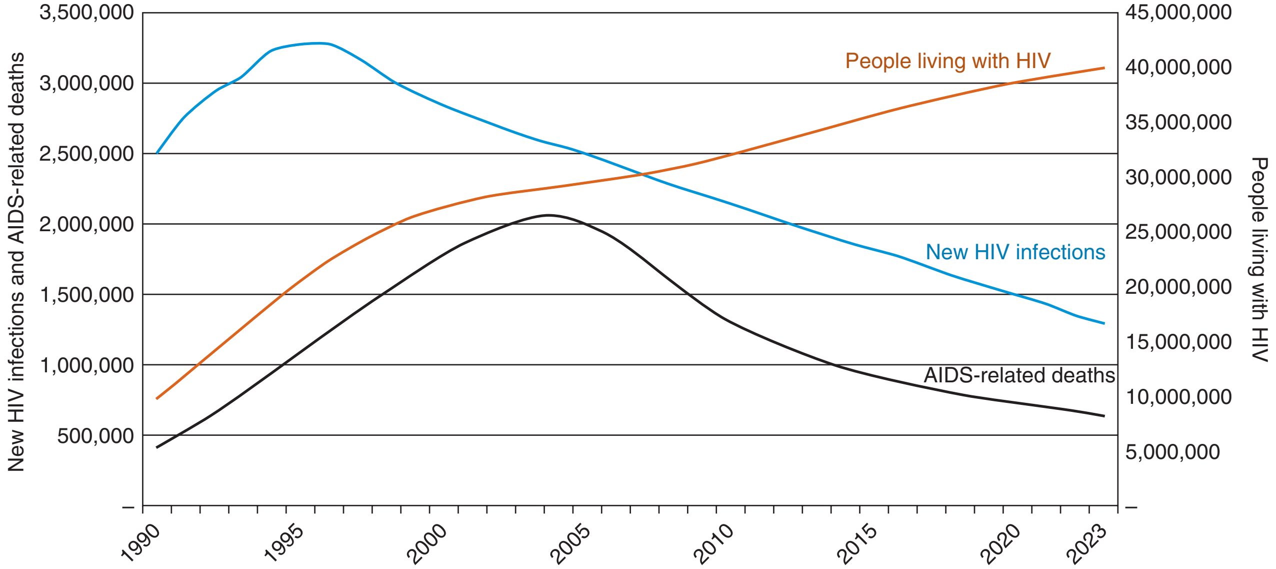

FIGURE 208-9 Global estimates of new HIV cases, AIDS-related deaths, and HIV prevalence, 1990–2023. *(From UNAIDS.)*

of people living with HIV—i.e., the global prevalence—has increased more than fourfold since 1990, reflecting the combined effects of continued high rates of new HIV infections and the life-prolonging impact of antiretroviral therapy **(Fig. 208-9)**. The overall global prevalence of HIV infection among persons 15–49 years of age is ~0.7%, with rates varying widely by country and region as illustrated in **Fig. 208-10**. The WHO Eastern and Southern Africa region remains most severely affected, with more than 1 in 20 adults (5.7%) living with HIV and accounting for more than half of the people with HIV worldwide. An estimated 88.4 million people have become infected with HIV since the start of the pandemic.

In 2023, an estimated 1.3 million new cases of HIV infection occurred worldwide, including 120,000 among children <15 years; about one-third of new infections were among people age 15–24 years. The regional distribution of new HIV cases in 2023 is shown in **Fig. 208-11**. Globally, members of certain high-risk populations are disproportionately affected by HIV infection, often because of factors such as marginalization, discrimination, and in some cases criminalization. Sex workers, people who inject drugs, transgender people, prisoners, gay men and other men who have sex with men, the clients of sex workers, and the sexual partners of these key populations accounted for 55% of all new HIV infections in 2022 **(Fig. 208-12)**.

New HIV infections globally have fallen by 60% since their peak in 1995 (Fig. 208-9). In 2023, 1.3 million people were newly infected with HIV, compared with 3.3 million people in 1995. Reductions in global HIV incidence likely reflect progress with HIV prevention efforts and the increased provision to HIV-infected people of antiretroviral therapy, which makes them much less likely to transmit the virus to sexual partners. Between 2010 and 2023, new HIV infections declined by 39%, from 2.1 million to 1.3 million. During the same period a ~62%

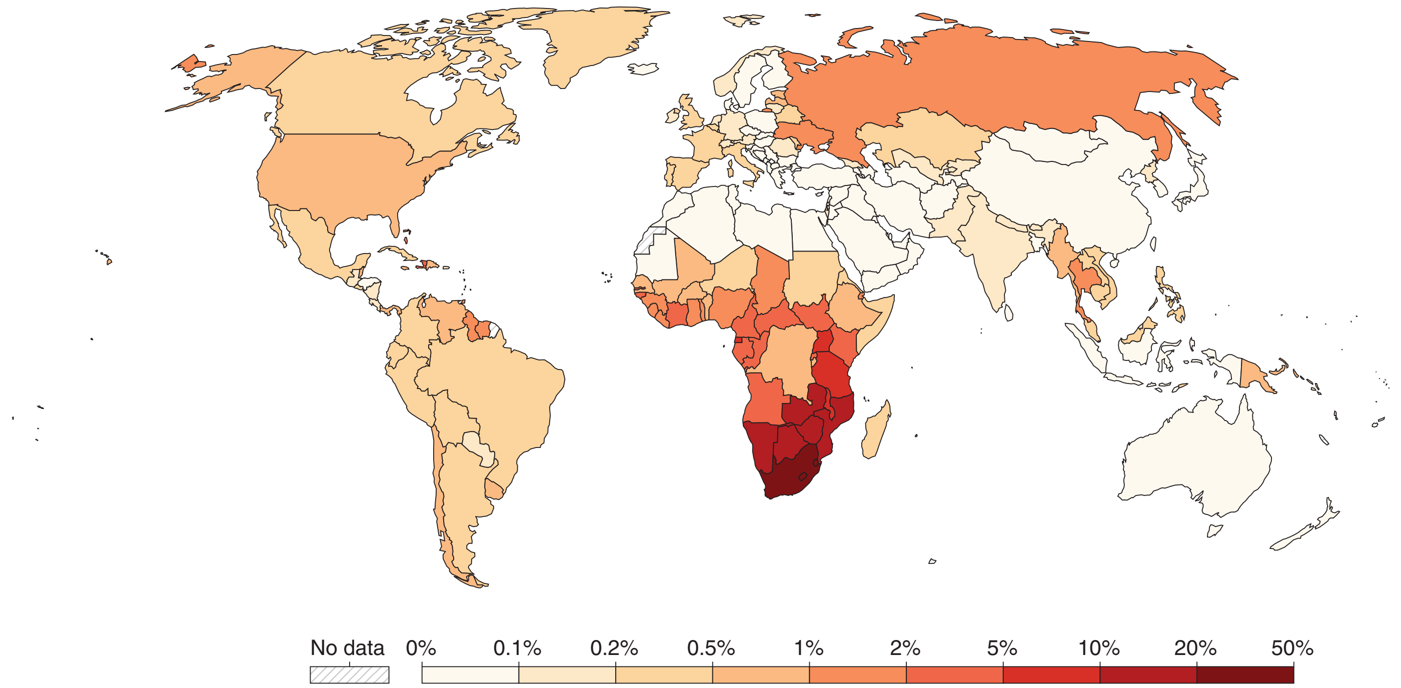

FIGURE 208-10 Adult HIV prevalence rates by country, 2021. Data are estimates for adults age 15–49 years. *(Reproduced from IHME, Global Burden of Disease with major processing by Our World in Data; 2024.)*

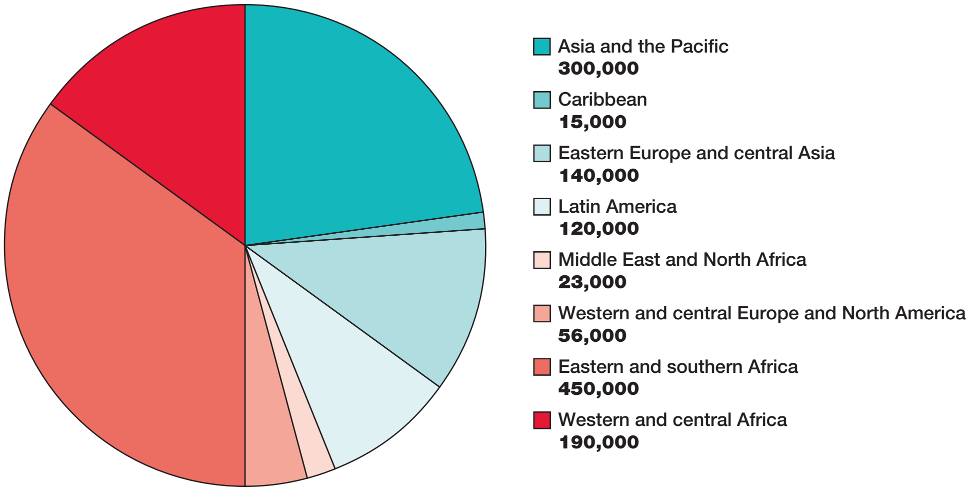

FIGURE 208-11 Distribution of new HIV infections, by region, 2023. *(Reproduced with permission from UNAIDS.)*

reduction in HIV infections among children <15 years was observed, from 300,000 in 2010 to 120,000 in 2023. This progress is due largely to the increasing availability of antiretroviral medications to prevent the transmission of HIV from mother to infant. An estimated 30.7 million people with HIV globally, 77% of all people living with HIV, were accessing antiretroviral therapy as of December 2023, up from 7.7 million people in 2010. Among pregnant women with HIV, 84% had access to antiretroviral medicines to prevent transmission of HIV to their child in 2023.

In 2023, global AIDS deaths totaled 630,000 (including 76,000 children <15 years), a 65% decrease since 2000 that coincides with a rapid expansion of access to antiretroviral therapy **(Fig. 208-13)**. Since the beginning of the HIV pandemic, an estimated 42.3 million persons globally have died of an AIDS-related illness.

The HIV epidemic has occurred in "waves" in different regions of the world, each wave having somewhat different characteristics depending on the demographics of the country and region in question and the timing of the introduction of HIV into the population. Although the AIDS epidemic was first recognized in the United States and shortly thereafter in Western Europe, it very likely began in sub-Saharan Africa (see above), a region particularly devastated by the epidemic.

The 21 countries of the **Eastern and Southern Africa** region are home to about 7% of the world's population but had 20.8 million people living with HIV in 2023, >50% of the global total (Fig. 208-8). HIV prevalence among adults age 15–49 years across the region is 5.7%. Of 21 countries in the region, 17 have generalized epidemics, that is, their national prevalence is >1%. In 6 countries in the region, >10% of the adult population age 15–49 has HIV infection (Fig. 208-10). South Africa has the highest number of people living with HIV in the world (7.7 million); Eswatini (formerly known as Swaziland) has the highest adult HIV prevalence globally (25.1%). Heterosexual exposure is the primary mode of HIV transmission in most countries in this region, as is the case throughout sub-Saharan Africa. Women and girls age 15 years and older account for ~63% of all HIV infections in the region. Key populations, notably sex workers and clients of sex workers, accounted for 23% of new infections in the region in 2022.

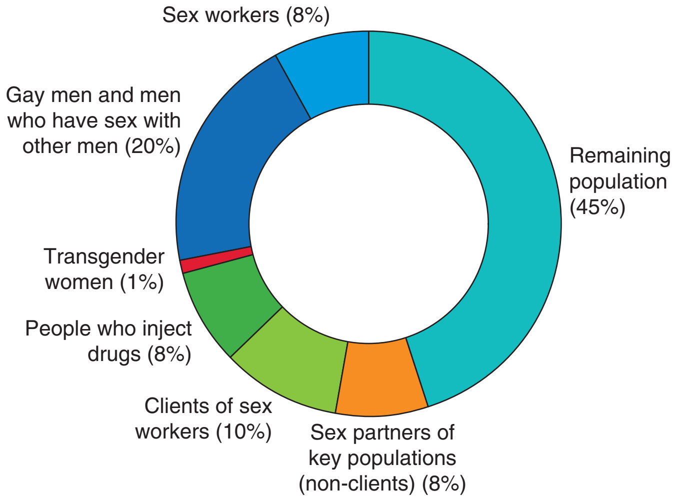

FIGURE 208-12 Global distribution of new HIV infections by population. Data for 2022. *(From UNAIDS.)*

Major progress in the HIV response has been made in the Eastern and Southern Africa region in recent years. The annual number of people acquiring HIV in the region fell by 59% between 2010 and 2023, from 1.1 million to 450,000. AIDS-related deaths decreased by 57% in that period, from 600,000 in 2010 to 260,000 in 2023. As of 2023, 93% of people living with HIV in the region knew their infections status; 83% of those living with HIV (17.4 million people) were on antiretroviral therapy, and 78% had suppressed viral loads.

The 25 countries of the **Western and Central Africa** region are home to 5.1 million people living with HIV, of whom 3.1 million are women age ≥15 years and 380,000 are children. HIV prevalence in most countries is relatively low compared with Eastern and Southern Africa. HIV prevalence among adults across the region overall stands at 1.2%. About 40% of new infections in the region in 2023 occurred in Nigeria, a large country with an HIV seroprevalence rate of 1.3%. As in Eastern and Southern Africa, heterosexual transmission accounts for most HIV transmission in West and Central Africa, with sex workers and their clients accounting for about one-fifth of new infections.

The **Middle East and North Africa** region has one of the lowest HIV prevalence rates in the world (<0.1%), although annual new infections more than doubled between 2010 and 2023, from 11,000 to 23,000. In 2023, an estimated 230,000 people were living with HIV in the 19 countries in the region. Cases are largely concentrated among IDUs, men who have sex with men, and sex workers and their clients.

In **Asia and the Pacific**, an estimated 6.7 million people were living with HIV at the end of 2023, with an overall adult prevalence rate of 0.2%. With 300,000 new infections in 2023, the region accounted for one-quarter of all new infections globally that year. The region saw a 13% decrease in new infections between 2010 and 2023, with reductions in Thailand and Vietnam offset by increases in Afghanistan, Bangladesh, Fiji, the Lao People's Democratic Republic, Papua New Guinea, and the Philippines. In the same time frame, AIDS-related deaths in the region fell by 51% to 150,000. Among countries in this region, only Thailand has an adult seroprevalence rate that reaches 1%. However, the populations of many countries are so large that even low infection and seroprevalence rates result in large numbers of people living with HIV. In this regard, three populous countries—China, India, and Indonesia—account for around three-quarters of all people living with HIV in the region. Key populations (Fig. 208-11) and their

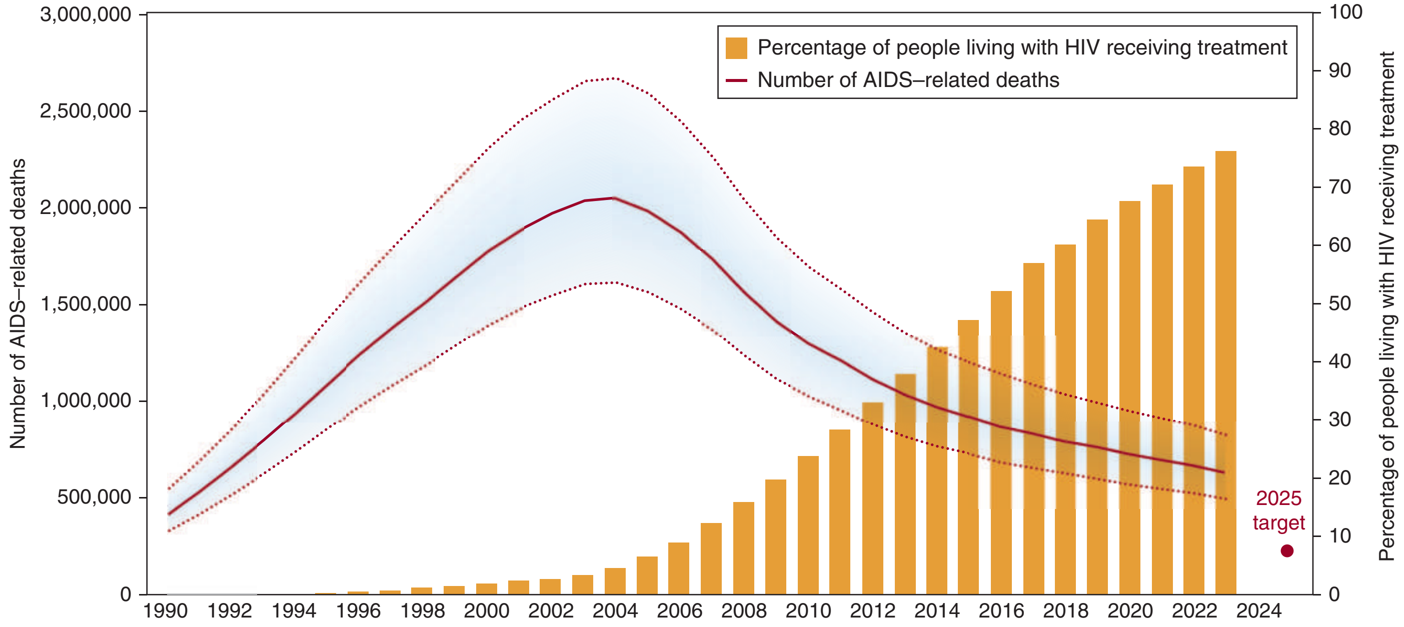

FIGURE 208-13 Global antiretroviral therapy coverage and number of AIDS-related deaths, 1990-2023. *(Data from The Urgency of Now: AIDS at a Crossroads. Geneva: Joint United Nations Programme on HIV/AIDS; 2024.)*

partners accounted for ~80% of new HIV infections in the region in 2022. Rising numbers of new infections among gay men and other men who have sex with men in this region are a major concern.

The HIV epidemic continues to expand in **Eastern Europe and Central Asia**, with a 20% increase in annual new HIV infections and 34% increase in AIDS deaths between 2010 and 2023. There are about 2.1 million people living with HIV in the region, where the epidemic has been driven by injection drug use but has an increasing proportion of new HIV infections transmitted sexually. Members of key populations and their sexual partners account for ~94% of new infections in the region. The Russian Federation, Ukraine, Kazakhstan, and Uzbekistan reported 92% of all registered cases in the region in 2023.

Approximately 2.3 million people were living with HIV/AIDS in **Latin America** at the end of 2023. The annual number of new infections in the region increased by 9% between 2010 and 2023 to 120,000. About two-thirds of new infections were among key populations and their sex partners, with increasing rates among men who have sex with men, sex workers, and transgender women. AIDS-related deaths declined by 28% to 30,000 from 2010 to 2023. Brazil is home to the largest number of HIV-infected persons (1,00,000) in the region.

In the **Caribbean**, an estimated 340,000 people were living with HIV in 2023, with 15,000 new infections that year. New infections have declined by 22% in the region since 2010. About 90% of new infections in the region in 2023 were in Cuba, the Dominican Republic, Haiti, and Jamaica. People from key populations and their sex partners accounted for almost half of new infections in the region in 2023.

Approximately 2.3 million people were living with HIV/AIDS in **North America and Western and Central Europe** in 2023. While modes of transmission vary greatly by country, HIV in the region disproportionately affects men who have sex with men. New infections in the region fell by 24% between 2010 and 2023 to 56,000, while AIDS-related deaths fell 34%, to 13,000. Key populations accounted for ~83% of new infections in the region in 2022.

HIV INFECTION IN THE UNITED STATES

As of 2022, an estimated 1.2 million individuals in the United States were living with HIV infection, ~13% of whom were unaware of their infection. Among people with diagnosed HIV in 2022, about 76% have received some HIV care, and about 65% have achieved viral suppression (see below).

Nearly 40% of people living with HIV in the United States are Black/African American and 26% are Hispanic/Latino; ~60% are men who have sex with men, according to CDC estimates. The HIV prevalence rate among all individuals age 13 years or older in the United States is 0.4%. Approximately 1.4% of Black/African-American adults are living with HIV in the United States, more than any other racial/ethnic group.

The estimated annual number of new HIV infections in the United States has fallen by more than two-thirds since its height in the late 1980s of about 130,000 per year. CDC data indicate further progress in recent years, as overall incidence fell 12% from 2018 to 2022, from 36,200 new infections to 31,800. The estimated distribution of new infections in 2022, broken down by transmission category, is shown in **Fig. 208-14**.

In the United States, the burden of HIV infection is not evenly distributed across states and regions. In most areas of the country, HIV is concentrated in urban areas. In the southern United States, larger percentages of diagnoses are in smaller metropolitan and nonmetropolitan areas. HIV has disproportionately affected minority populations in the United States in both urban and rural areas. Among new HIV infections from 2018 to 2022 in the United States, 37% were among Blacks/African Americans, a group that constitutes only 12% of the U.S. population. Hispanics/Latinos, 18% of the U.S. population, accounted for 33% of new HIV infections. The rate of new HIV infections in 2022 by race/ethnicity in the United States is shown in **Fig. 208-15**.

Perinatal HIV transmission, from an HIV-infected mother to her baby, has declined significantly in the United States, largely due to the implementation of guidelines for the universal counseling and voluntary HIV testing of pregnant women and the use of antiretroviral

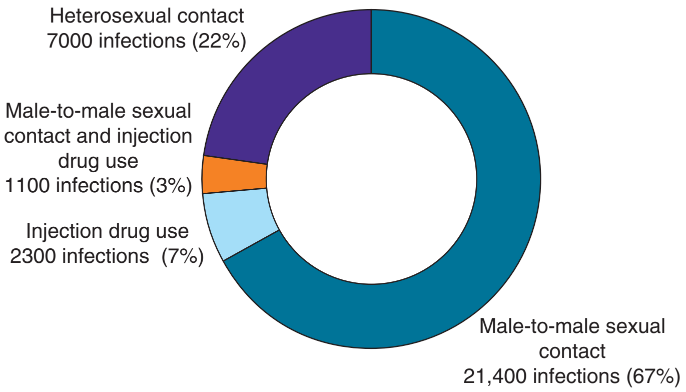

FIGURE 208-14 Estimated distribution of new HIV infections in the United States among people age ≥13 years, by transmission category, 2022. Total: 31,800. *(From CDC: HIV Surveillance Supplemental Report 29 [No. 1], 2024.)*

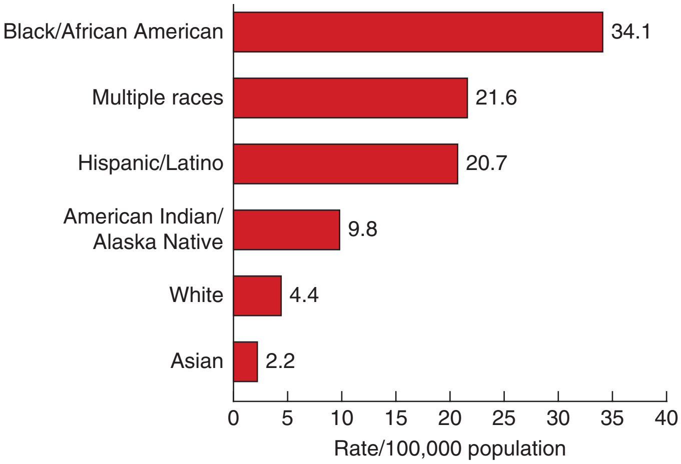

FIGURE 208-15 Estimated rate of new HIV infections in the United States among people age ≥13 years, by race/ethnicity (per 100,000 population), 2022. *(From CDC: HIV Surveillance Supplemental Report 29 [No. 1], 2024.)*

therapy for pregnant women and newborn infants to prevent infection. In 2022, 62 children were newly diagnosed with HIV infection in the United States, down from a peak of ~1750 in 1991.

The rate of HIV-related deaths in the United States rose steadily through the 1980s, peaked around 1994, and fell rapidly through 1997 **(Fig. 208-16)**. Since then, the rate of death due to HIV disease continues to decline, decreasing from 2.6 deaths per 100,000 in 2010 to 1.3 deaths per 100,000 in 2022. This trend is likely due to several factors, including improved prophylaxis and treatment of opportunistic infections, growing experience among the health professions in caring for HIV-infected individuals, improved access to health care, and a decrease in new infections. However, the most influential factor clearly has been the increased use—and continued refinement—of antiretroviral therapy (ART), generally administered in a combination of two, three, or four agents. Despite much progress, HIV disease continues to remain among the leading causes of death among all persons age 25–44 years in the United States, ranking 13th in 2022. HIV disease as an underlying cause of death is particularly high among Black/African-American persons age 25–44 years, ranking as the 8th leading cause among Black males and the 12th leading cause among Black females in 2022.

PATHOPHYSIOLOGY AND PATHOGENESIS

The hallmark of HIV disease is a profound immunodeficiency resulting primarily from a progressive quantitative and qualitative deficiency of the subset of T lymphocytes referred to as *helper T cells* occurring in a setting of aberrant immune activation creating a state of immunosuppression. The *helper* subset of T cells is defined phenotypically by the presence on its surface of the CD4 molecule **(Chap. 360)**, which serves as the primary cellular receptor for HIV. A co-receptor also must be present together with CD4 for efficient binding, fusion, and entry of HIV-1 into its target cells (Figs. 208-3 and 208-4). HIV-1 uses two major co-receptors, CCR5 and CXCR4, for fusion and entry; these co-receptors are also the primary receptors for certain chemoattractant cytokines termed *chemokines* and belong to the seven-transmembrane-domain G protein–coupled family of receptors. Multiple mechanisms responsible for cellular depletion and/or immune dysfunction of CD4+ T cells have been demonstrated in vitro. These include direct infection and destruction of these cells by HIV, as well as indirect effects such as immune clearance of infected cells; cell death associated with aberrant immune activation and inflammation, including caspase 1–mediated pyroptosis prompted by tissue CD4+ T cells undergoing abortive/nonproductive HIV infection; and immune exhaustion due to persistent cellular activation with resulting cellular dysfunction. Patients with CD4+ T cell levels below certain thresholds are at high risk of developing a variety of opportunistic diseases, particularly the infections and neoplasms that are AIDS-defining illnesses. Some features of AIDS, such as Kaposi's sarcoma and certain neurologic abnormalities, cannot be explained completely by the immunodeficiency caused by HIV infection, since these complications may occur prior to the development of severe immunologic impairment.

The combination of viral pathogenic and immunopathogenic events that occur during the course of HIV disease from the moment of initial (primary) infection through the development of advanced-stage disease is complex and varied. It is important to appreciate that the pathogenic mechanisms of HIV disease are multifactorial and multiphasic and are different at different stages of the disease. Therefore, it is essential to consider the typical clinical course of an untreated individual with HIV to better appreciate these pathogenic events **(Fig. 208-17)**.

■ EARLY EVENTS IN HIV INFECTION: PRIMARY INFECTION AND INITIAL DISSEMINATION OF VIRUS

Using rectal or vaginal mucosal transmission in nonhuman primates as a model, the earliest events (within hours) that occur following exposure of HIV to the mucosal surface determine whether an infection will be established or aborted as well as the subsequent course of events following infection. Although the mucosal barrier is relatively effective in limiting access of HIV to susceptible targets in the submucosal tissue, the virus can cross the barrier by transport on Langerhans cells, an epidermal type of DC, just beneath the surface or through microscopic rents in the mucosa. Significant disruptions in the mucosal barrier as seen in ulcerative genital disease facilitate viral entry and increase the

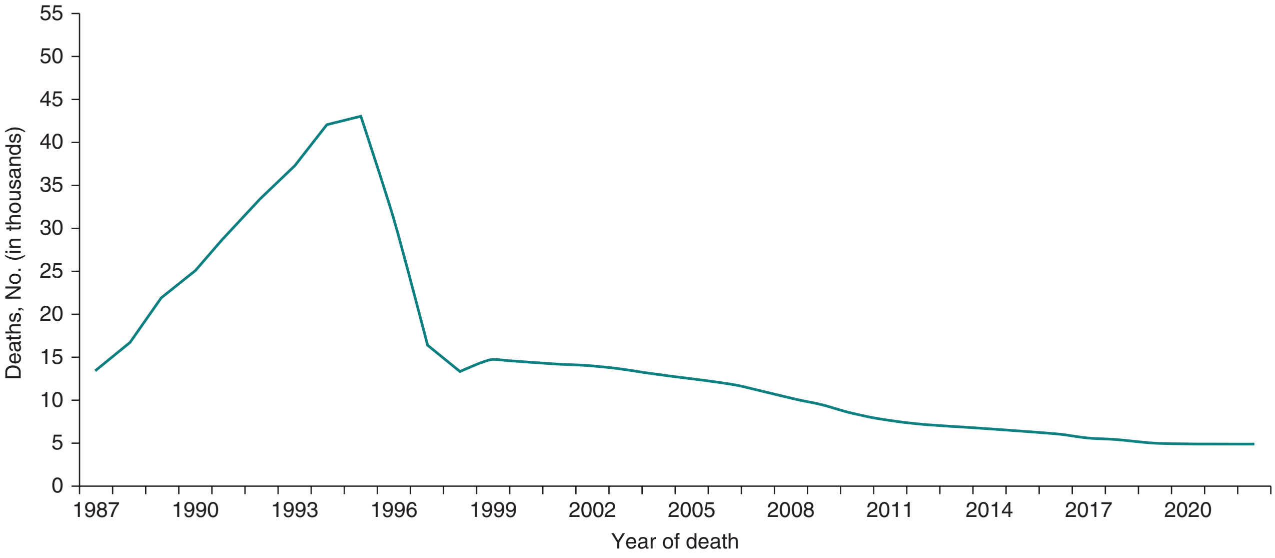

FIGURE 208-16 Trends in annual age-adjusted rates of death in the general population with HIV infection as the underlying cause, United States, 1987–2022. *(From CDC.)*

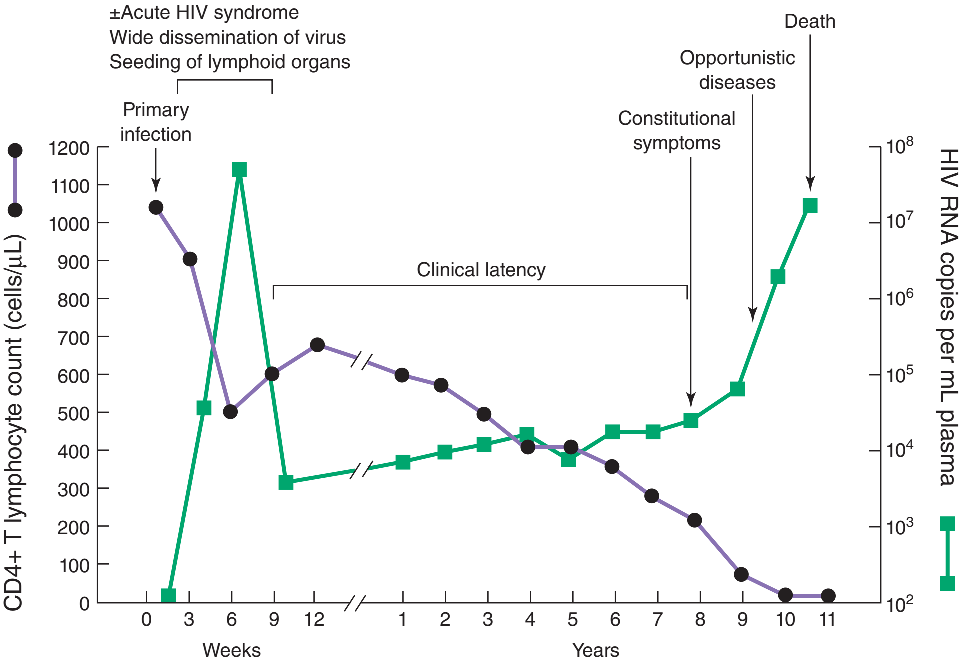

FIGURE 208-17 Typical course of an untreated HIV-infected individual. See text for detailed description. *(From G Pantaleo, C Graziosi, AS Fauci: The Immunopathogenesis of Human Immunodeficiency Virus Infection. N Engl J Med 328:327, 1993. Copyright © 1993 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

efficiency of infection. Viruses then seek susceptible targets, which are primarily CD4+ T cells that are spatially dispersed in the mucosa. This spatial dispersion of targets provides a significant obstacle to the establishment of infection. Such obstacles account for the low efficiency of sexual transmission of HIV (see "Sexual Transmission," above). Both "partially" resting CD4+ T cells and activated CD4+ T cells serve as early amplifiers of infection. Resting CD4+ T cells are more abundant; however, activated CD4+ T cells support productive infection and thus generate larger amounts of virus. For infection to become established, the basic reproductive rate (R_0) must become equal to or greater than 1, i.e., each infected cell would infect at least one other cell. Once infection is established, the virus replicates in lymphoid cells in the mucosa, the submucosa, and to some extent the lymphoreticular tissues that drain the gut or genital tissues. For a variable period up to several days, the virus is typically not detected in the plasma. This period is called the "eclipse" phase of infection. As more virus is produced within several days to weeks, it is disseminated, first to the draining lymph nodes and then to other lymphoid compartments where it has easy access to dense concentrations of CD4+ T cell targets, allowing for a burst of high-level plasma viremia that is readily detectable by currently available assays measuring viral RNA **(Fig. 208-18)**. The gut-associated lymphoid tissue (GALT) is one target of HIV infection and a location where CD4+ T cells (primarily memory cells) are infected and depleted, both by direct viral effects and by activation-associated apoptosis. Once virus replication reaches this threshold and virus is widely disseminated, infection is firmly established throughout the lymphoid tissues of the body and persists for the life of the individual. It is important to point out that the efficiency of initial infection of susceptible cells may vary somewhat with the route of infection. Virus that enters directly into the bloodstream via infected blood or blood products (i.e., transfusions, use of contaminated needles for injection drugs, sharp-object injuries, maternal-to-fetal transmission either intrapartum or perinatally, or sexual intercourse where there is enough trauma to cause bleeding) is likely first cleared from the circulation to the spleen and other lymphoid organs, where primary focal infections begin, followed by wider dissemination throughout other lymphoid tissues as described above.

It has been demonstrated that sexual transmission of HIV is the result of a single infectious event and that a viral genetic bottleneck exists for transmission with selective transmission of certain viruses. In this regard, certain characteristics of the HIV envelope glycoprotein have a major influence on transmission, at least in subtype A and C viruses. Transmitting viruses, often referred to as "founder viruses," are typically a small fraction of the circulating viremia of the transmitting partner and are less-diverged viruses with signature sequences including shorter V1–V2 loop sequences and fewer predicted N-linked glycosylation sites relative to the major circulating variants. These

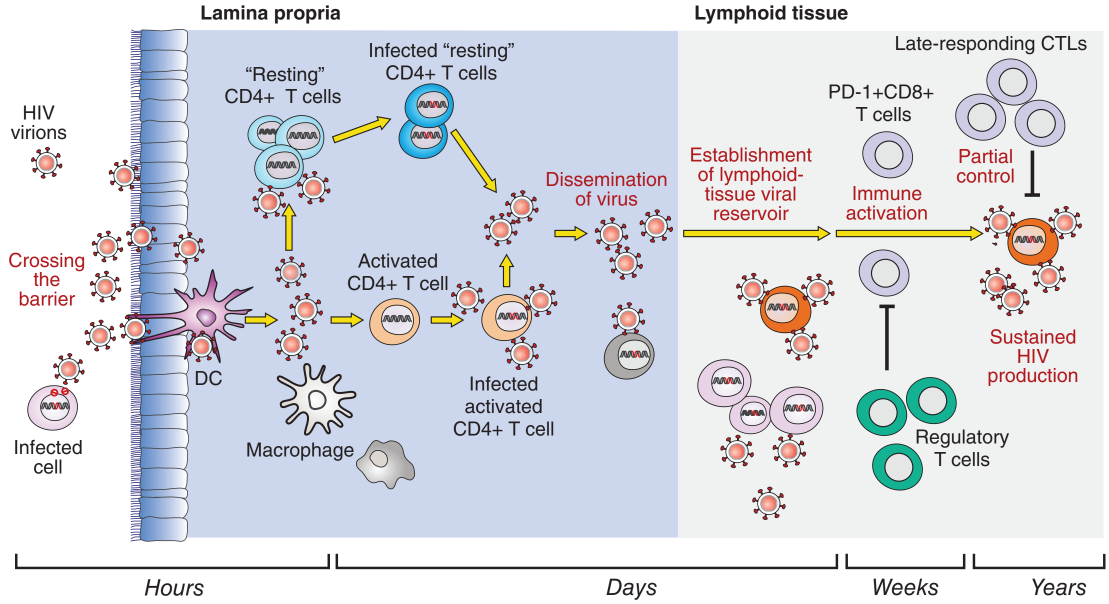

FIGURE 208-18 Summary of early events in HIV infection. See text for detailed description. CTLs, cytolytic T lymphocytes; HIV, human immunodeficiency virus. *(Adapted from AT Haase: Nat Rev Immunol 5:783, 2005.)*

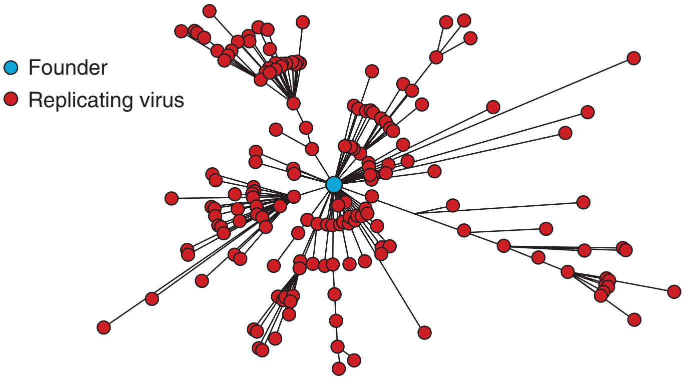

FIGURE 208-19 As HIV diverges from founder to chronically replicating virus, it accumulates N-linked glycosylation sites. See text for detailed description. *(Adapted from CA Derdeyn et al: Science 303:2019, 2004; B Chohan et al: J Virol 79:6528, 2005; and BF Keele et al: Proc Natl Acad Sci USA 105:7552, 2008.)*

viruses are almost exclusively R5 strains and are usually sensitive to neutralizing antibody. Once replication proceeds in the newly infected partner, the founder virus diverges and accumulates glycosylation sites, becoming progressively more resistant to neutralization **(Fig. 208-19)**.

The acute burst of viremia and wide dissemination of virus in primary HIV infection may be associated with an *acute HIV syndrome*, which occurs to varying degrees in ~50% of individuals within 2 to 4 weeks of initial infection (see below). This syndrome is usually associated with high levels of plasma viremia reflected in millions of copies of HIV RNA per milliliter of plasma that can last for several weeks. Acute mononucleosis-like symptoms are well correlated with the presence of high levels of plasma viremia. Virtually all patients develop some degree of plasma viremia during primary infection, which contributes to virus dissemination throughout the lymphoid tissue, even though they may remain asymptomatic or not recall experiencing symptoms. The peak level of plasma viremia in primary HIV infection does not necessarily determine the rate of disease progression; however, the set point of the level of steady-state plasma viremia after ~1 year correlates with the rate of disease progression in the untreated patient and with immunologic and virologic aberrancies that may persist in the treated patient. The strikingly high levels of viremia observed in many patients during acute HIV infection is felt to be associated with a higher likelihood of transmission of the virus to others by a variety of routes including sexual transmission, shared needles and syringes, and mother-to-child transmission intrapartum, perinatally, or via breast milk.

■ ESTABLISHMENT OF CHRONIC INFECTION

Persistence of Virus Replication HIV infection is unique among human viral infections. Despite the robust cellular and humoral immune responses that are mounted following primary infection (see "Immune Response to HIV," below), once infection has been established the virus succeeds in escaping complete immune-mediated clearance, paradoxically seems to thrive on immune activation, and is never eliminated completely from the body. Rather, a chronic infection develops and persists with varying degrees of continual virus replication in the untreated patient for a median of ~10 years before the patient becomes clinically ill (see "Advanced HIV Disease," below). It is this establishment of a chronic, persistent infection that is the hallmark of HIV disease. Throughout the often-protracted course of chronic infection, virus replication can invariably be detected in untreated patients by widely available molecular assays that measure copies of virion-associated HIV RNA in plasma (copies per milliliter). Levels of virus vary greatly in most untreated patients, usually ranging from fewer than 50 to greater than a million copies of HIV RNA per milliliter of plasma. Studies using highly sensitive molecular techniques have demonstrated that even in treated patients in whom plasma viremia is suppressed to below detection limits of commercial assays (20–50 copies of HIV RNA per milliliter depending on assay kit manufacturer) by ART, there is a continual low level of virion production in most infected patients. In other human viral infections (with some exceptions) if the host survives, the virus is completely cleared from the body and a state of immunity against subsequent infection develops. HIV infection very rarely kills the host during primary infection. Certain viruses, such as HSV **(Chap. 197)**, are not completely cleared from the body after infection, but instead enter a latent state; in these cases, clinical latency is accompanied by microbiologic latency. This is not the case with HIV infection as described above. Chronicity associated with persistent virus replication can also be seen in certain cases of HBV and HCV infections **(Chap. 352)**; however, in these infections the immune system is not a target of the virus.

Escape of HIV from Effective Immune System Control Inherent to the establishment of chronicity of HIV infection is the ability of the virus to evade adequate control and elimination by the cellular and humoral immune responses. There are several mechanisms whereby the virus accomplishes this evasion. Paramount among these is the establishment of a sustained level of replication associated with the generation of viral diversity via mutation and recombination. The selection of mutants that escape control by CD8+ cytolytic T lymphocytes (CTLs) is critical to the propagation and progression of HIV infection. The high rate of virus replication associated with inevitable mutations also contributes to the inability of antibody to neutralize and/or clear the autologous virus. Furthermore, for reasons that remain unclear, the humoral immune system does not readily produce classic neutralizing antibodies against the HIV envelope and does so only after years of persistent virus replication and after the infection is firmly established (see below). Extensive analyses of sequential HIV isolates and host responses have demonstrated that viral escape from B cell and CD8+ T cell responses occurs early after infection and allows the virus to continually evade effective immune responses. Virus-specific CD8+ CTLs expand greatly during primary HIV infection, and they likely represent the high-affinity responses that would be expected to be most efficient in eliminating virus-infected cells; however, with very rare exceptions, viral control is incomplete as viral replication persists at relatively high levels in the majority of individuals. In addition to viral escape from CTLs through high rates of mutation, it is thought that the initially strong immune response becomes qualitatively dysfunctional owing to the overwhelming immune activation associated with persistent viral replication, leading to immune "exhaustion" that affects both arms of adaptive immunity. Several studies have indicated that exhaustion of HIV-specific CD8+ T cells during prolonged immune activation is associated with upregulation of several inhibitory receptors, such as the programmed death (PD) 1 molecule (of the B7-CD28 family of molecules), T cell immunoreceptor with Ig and ITIM domains (TIGIT), T cell immunoglobulin and mucin domain–containing molecule 3 (Tim-3), and lymphocyte activating gene 3 (Lag-3), collectively referred to as *immune-checkpoint receptors*. Upregulation of these surface proteins restricts polyreactivity and proliferative capacity, functional attributes of CD8+ T cells that are essential for effective killing of pathogens. Another mechanism contributing to the evasion by HIV of immune system control is the downregulation of HLA class I molecules on the surface of HIV-infected cells by the viral proteins Nef, Tat, and Vpu, resulting in the lack of ability of CD8+ CTLs to recognize and kill infected target cells. Although this downregulation of HLA class I molecules would seem to favor elimination of HIV-infected cells by natural killer (NK) cells, this latter mechanism does not remove HIV-infected cells effectively (see below). Another potential means of escape of HIV-infected cells from elimination by CD8+ CTLs is the sequestration of infected cells in immunologically privileged sites such as the central nervous system (CNS), as well as the low frequency of virus-specific CD8+ CTLs in areas of lymphoid tissues, namely germinal centers, where HIV actively replicates.

The principal targets of neutralizing antibodies against HIV are the envelope proteins gp120 and gp41. HIV employs at least three mechanisms to evade neutralizing antibody responses: hypervariability in the primary sequence of the envelope, extensive glycosylation of the envelope, and conformational masking of neutralizing epitopes.

Several studies that have followed the evolution of the humoral immune response to HIV from the earliest points after primary infection indicate that the virus continually mutates to escape the emerging antibody response such that the sequential antibodies that are induced do not neutralize the currently autologous virus. *Broadly neutralizing antibodies* capable of neutralizing a wide range of primary HIV isolates in vitro occur in only about 20% of people with HIV, and, when they do occur, 2–3 years of infection with continual virus replication are generally required to drive the affinity maturation of the antibodies. Unfortunately, by the time these broadly neutralizing antibodies are formed, they are ineffective in containing the virus currently replicating in the patient. Persistent viremia also results in exhaustion of B cells like the exhaustion reported for CD8+ T cells, adding to the defects in the humoral response to HIV.

CD4+ T cell help is essential for the integrity of both humoral and cell-mediated antigen-specific immune responses. HIV preferentially infects activated CD4+ T cells including HIV-specific CD4+ T cells, and so this loss of viral-specific helper T cell responses has profoundly negative consequences for the immunologic control of HIV replication. Furthermore, this loss occurs early in the course of infection, and animal studies indicate that 40–70% of all memory CD4+ T cells in the GALT are eliminated during acute infection. During chronic HIV viremia, CD4+ T cells also exhibit evidence of exhaustion, reflected in upregulation of the cytotoxic T lymphocyte–associated antigen 4 (CTLA-4), also a member of the B7-CD28 family.

Finally, the escape of HIV from immune-mediated elimination during primary infection allows the formation of a pool of latently infected CD4+ T cells, referred to as the *viral reservoir*, which may not be recognized or completely eliminated by virus-specific CTLs or by ART (see below). Thus, despite a potent immune response and the marked downregulation of virus replication following primary HIV infection, HIV succeeds in establishing a state of chronic infection with a variable degree of persistent virus replication. During this period most patients make the clinical transition from acute primary infection to variable periods of clinical latency or smoldering disease activity (see below).

The HIV Reservoir: Obstacles to the Eradication of Virus

A pool of latently infected, resting CD4+ T cells that serves as at least one component of the persistent reservoir of virus exists in virtually all people with HIV, including those who are receiving ART. Such cells carry an integrated form of HIV DNA in the genome of the host and can remain in this state until an activation signal drives the expression of HIV transcripts. Only a small fraction of the latently infected cells in the viral reservoir contains replication-competent virus, with the overwhelming majority of cells containing defective proviruses incapable of a full replication cycle. However, upon activation of the reservoir variable degrees of sustained virus replication invariably occur. This form of latency is to be distinguished from preintegration latency, in which HIV enters a resting CD4+ T cell and, in the absence of an activation signal, reverse transcription of the HIV genome occurs to a certain extent but the resulting proviral DNA fails to integrate into the host genome. This period of preintegration latency may last hours to days, and if no activation signal is delivered to the cell, the proviral DNA loses its capacity to initiate a productive infection. If these cells do become activated prior to decay of the preintegration complex, reverse transcription proceeds to completion and the virus continues along its replication cycle (see above and Fig. 208-20). The pool of cells that are in the postintegration state of latency is established early during primary HIV infection. Despite the suppression of plasma viremia to <20–50 copies per milliliter by potent regimens of ART administered over several years, this pool of latently infected cells persists and can give rise to replication-competent virus upon cellular activation ex vivo. Modeling studies built on projections of decay curves have estimated that in such a setting of prolonged viral suppression, it would require many years to the entire life of the host for the pool of latently infected cells to be eliminated. This has not been documented to occur spontaneously in any patients very likely because the latent viral reservoir is long-lived and is continually replenished by the low levels of persistent virus replication that may remain below the limits of detection of current assays (see below) as well as by the expansion by proliferation of the pool of latently infected cells (Fig. 208-20), even in patients who for the most part are treated successfully. Reservoirs of HIV-infected cells, latent or otherwise, can exist in multiple compartments including the lymphoid tissue, peripheral blood, and CNS (likely in cells of the monocyte/macrophage lineage) as well as in other unidentified locations. Over the past several years attempts have been made to eliminate HIV in the latent viral reservoir using agents that activate resting CD4+ T cells and/or reinitiate viral expression without systemic activation during the course of ART; however, such attempts, referred to as "shock and kill," have thus far been unsuccessful. This persistent reservoir of infected cells remains a major obstacle to the goal of eradication of virus from infected individuals and hence a classic "cure," despite the favorable clinical outcomes that have resulted from ART. Consequently, intense efforts are being directed toward investigating the feasibility of achieving ART-free HIV remission through passive transfer of long-acting broadly neutralizing antibodies and therapeutic agents that could enhance the host immune responses against the virus.

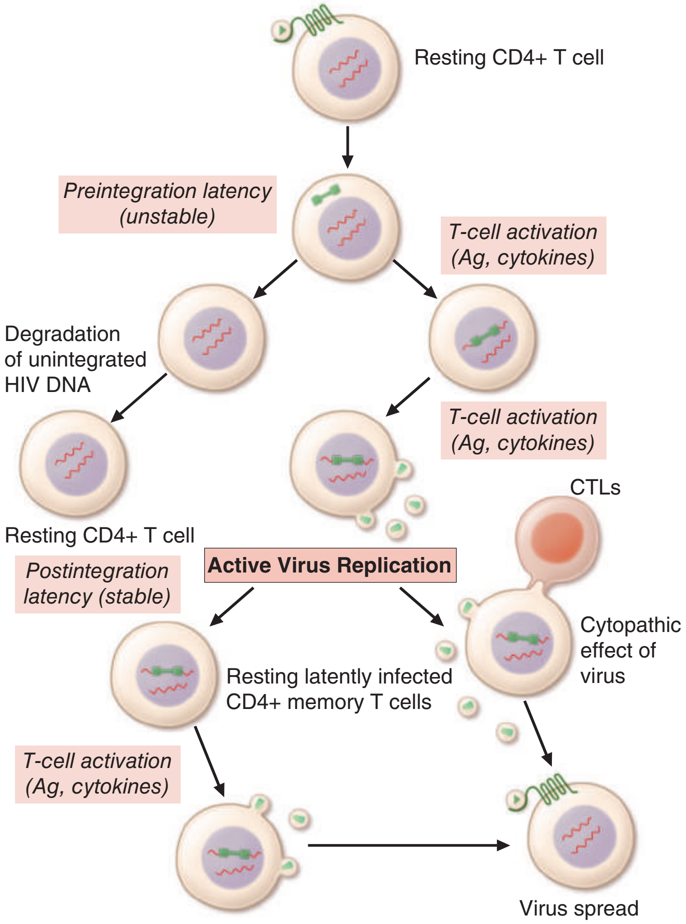
Resting CD4+ T cell
Preintegration latency (unstable)
T-cell activation (Ag, cytokines)
Degradation of unintegrated HIV DNA
T-cell activation (Ag, cytokines)
CTLs
Resting CD4+ T cell
Postintegration latency (stable)
Active Virus Replication
Cytopathic effect of virus
Resting latently infected CD4+ memory T cells
T-cell activation (Ag, cytokines)
Virus spread

FIGURE 208-20 Generation of latently infected, resting CD4+ T cells in people with HIV. See text for details. Ag, antigen; CTLs, cytolytic T lymphocytes. *(Courtesy of TW Chun.)*

Viral Dynamics The dynamics of viral production and turnover have been quantified using mathematical modeling in the setting of the administration of reverse transcriptase and protease inhibitors to people with HIV in clinical studies. Treatment with these drugs resulted in a precipitous decline in the level of plasma viremia, which typically fell by well over 90% within 2 weeks. It was determined on the basis of modeling the kinetics of viral decline and the emergence of resistant mutants during therapy that 93–99% of the circulating virus originated from recently infected, rapidly turning over CD4+ T cells and that ~1–7% of circulating virus originated from longer-lived cells, likely monocytes/macrophages. A negligible amount of circulating virus originated from the pool of latently infected cells (Fig. 208-21).

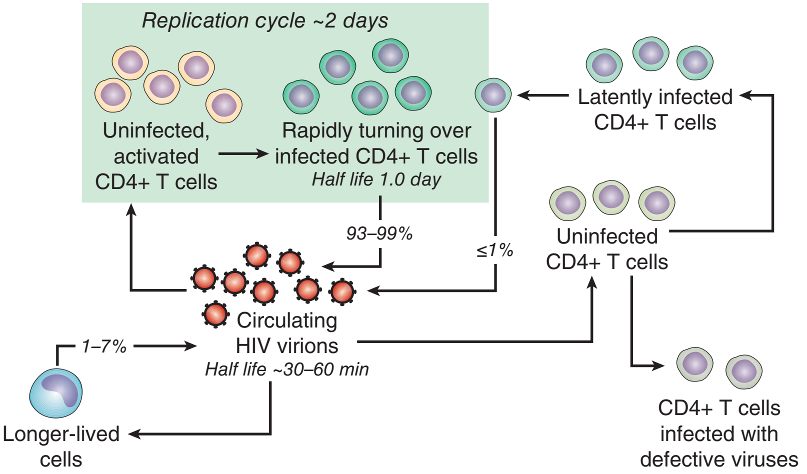
Replication cycle ~2 days
Uninfected, activated CD4+ T cells
Rapidly turning over infected CD4+ T cells
Half life 1.0 day
Latently infected CD4+ T cells
93–99%
≤1%
Uninfected CD4+ T cells
Circulating HIV virions
Half life ~30–60 min
1–7%
Longer-lived cells
CD4+ T cells infected with defective viruses

FIGURE 208-21 Dynamics of HIV infection in vivo. See text for detailed description. *(Adapted from Perelson AS et al: HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time. Science 271:1582, 1996.)*

It was also determined that the half-life of a circulating virion was ~30–60 min and that of productively infected cells was 1 day. Given the relatively steady level of plasma viremia and of infected cells, it appears that extremely large amounts of virus (~10^{10}–10^{11} virions) are produced and cleared from the circulation each day. In addition, data suggest that the minimal duration of the HIV-1 replication cycle in vivo is ~2 days. Other studies have demonstrated that the decrease in plasma viremia that results from treatment with ART correlates closely with a decrease in virus replication in lymph nodes, further confirming that lymphoid tissue is the main site of HIV replication and the main source of plasma viremia.

The level of steady-state viremia, called the viral *set point*, at ~1 year following acquisition of HIV infection has important prognostic implications for the progression of HIV disease in the untreated patient. It has been demonstrated that, as a group, untreated people with HIV who have a low set point at 6 months to 1 year following infection progress to AIDS much more slowly than do individuals whose set point is very high at that time **(Fig. 208-22)**.

Clinical Latency versus Microbiologic Latency With the exception of certain long-term nonprogressors and "elite controllers" of HIV replication, the level of CD4+ T cells in the blood inevitably decreases progressively in viremic people with HIV in the absence of ART. The decline in CD4+ T cells may be gradual or abrupt, the latter usually reflecting a significant spike in the level of plasma viremia. Most patients are relatively asymptomatic while this progressive decline is taking place (see below) and are often described as being in a state of *clinical latency*. However, this term is misleading; it does not mean disease latency, since progression, although slow in many cases and often without symptoms, is generally relentless as evidenced by readily detectable plasma viremia, during this period. Furthermore, clinical latency should not be confused with microbiologic latency since varying levels of virus replication inevitably occur during this period of clinical latency. Even in those rare patients, such as elite controllers, who have <50 copies of HIV RNA per milliliter in the absence of therapy, there is virtually always some degree of low-level ongoing virus replication.

■ ADVANCED HIV DISEASE

In untreated patients or in patients in whom therapy has not adequately controlled virus replication, after a variable period, usually measured in years, the CD4+ T cell count falls below a critical level (<200/μL) and the patient becomes highly susceptible to opportunistic disease (Fig. 208-17). For this reason, the CDC case definition of stage 3 (AIDS) includes all people with HIV >5 years of age with CD4+ T cell counts below this level (Table 208-2). Patients may experience constitutional signs and symptoms or may develop an opportunistic disease abruptly without any prior symptoms. The depletion of CD4+ T cells continues to be progressive and unrelenting in this phase. It is not uncommon for CD4+ T cell counts in the untreated patient to drop to as low as 10/μL or even to zero. In countries where ART as well as prophylaxis and treatment for opportunistic infections are readily accessible, survival is increased dramatically even in those patients with advanced HIV disease. In contrast, untreated patients who progress to this severest form of immunodeficiency usually succumb to opportunistic infections or neoplasms (see below).

■ LONG-TERM SURVIVORS, LONG-TERM NONPROGRESSORS, AND ELITE CONTROLLERS

It is important to distinguish between the terms *long-term survivor* and *long-term nonprogressor*. Long-term nonprogressors are by definition long-term survivors; however, the reverse is not always true. Predictions from one study that antedated the availability of effective ART estimated that ~13% of homosexual/bisexual men who were infected at an early age may remain free of clinical AIDS for >20 years. Many of these individuals may have gradually progressed in their degree of immune deficiency; however, they certainly survived for a considerable period. With the advent of effective ART, the survival of people with HIV has dramatically increased. Early in the AIDS pandemic, prior to the availability of antiretroviral therapy, if a patient presented with a life-threatening opportunistic infection, the median survival was 26 weeks from the time of presentation. Currently, a 20-year-old individual with HIV who is appropriately treated with ART can expect to live at least 50 years according to mathematical model projections. In the face of ART, long-term survival is now commonplace. Definitions of long-term nonprogressors have varied considerably over the years, and so such individuals constitute a heterogeneous group. Long-term nonprogressors were first described in the 1990s. Originally, individuals were considered to be long-term nonprogressors if they had been infected with HIV for a long period (≥10 years), their CD4+ T cell counts were in the normal range, their plasma viremia remained relatively low (undetectable to several thousand copies of HIV RNA/mL plasma), and they remained clinically stable over years without receiving ART. Approximately 5–15% of people

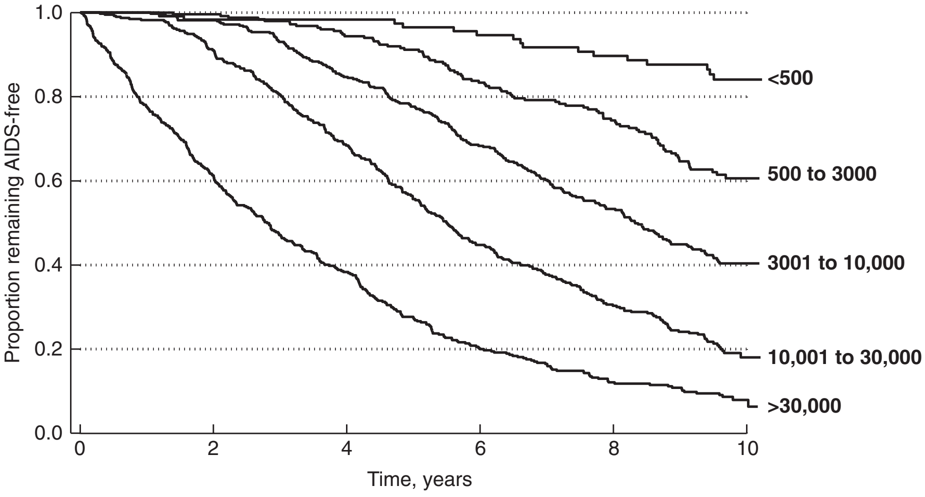

FIGURE 208-22 Relationship between levels of virus and rates of disease progression. Kaplan-Meier curves showing proportion of 1604 untreated patients remaining AIDS-free over 10 years, stratified by baseline HIV-1 RNA categories (copies per milliliter). *(From Multicenter AIDS Cohort Study; JW Mellors, A Muñoz, JV Giorgi, JB Margolick, CJ Tassoni, P Gupta, LA Kingsley, JA Todd, AJ Saah, R Detels, JP Phair, CR Rinaldo, Jr.)*

with HIV fell into this broader nonprogressor category. However, this group was rather heterogeneous and over time a significant proportion of these individuals progressed and ultimately required antiretroviral therapy. From this broader group, a much smaller subgroup of "elite" controllers was identified, and they constituted a fraction of 1% of people with HIV. These elite controllers, by definition, have extremely low levels of plasma viremia that is often undetectable by standard assays and normal CD4+ T cell counts. It is noteworthy that their HIV-specific immune response, especially HIV-specific CD8+ CTLs that can clear infected CD4+ T cells, is robust and distinctly superior to those of progressors with HIV infection. In this group of elite controllers, certain HLA class I haplotypes are overrepresented, particularly HLA-B57-01 and HLA-B27-05. Outside of the subgroup of elite controllers, multiple other genetic factors have been shown to be involved to a greater or lesser degree in the control of virus replication and thus in the rate of HIV disease progression (see "Genetic Factors in HIV-1 and AIDS Pathogenesis," below).

LYMPHOID ORGANS AND HIV PATHOGENESIS

Regardless of the portal of entry of HIV, lymphoid tissues are the major anatomic sites for the establishment and propagation of HIV infection. Despite the use of measurements of plasma viremia to determine the level of disease activity, virus replication occurs mainly in lymphoid tissue and not in blood; indeed, the level of plasma viremia directly reflects virus production in lymphoid tissue.

Some patients experience progressive generalized lymphadenopathy early in the course of the infection; others experience varying degrees of transient lymphadenopathy. Lymphadenopathy reflects the cellular activation and immune response to the virus in the lymphoid tissue, which is generally characterized by follicular or germinal center hyperplasia. Lymphoid tissue involvement is a common denominator of virtually all patients with HIV infection, even those without easily detectable lymphadenopathy.

Examinations of lymph tissue and peripheral blood in patients and monkeys during various stages of HIV and simian immunodeficiency virus (SIV) infection, respectively, have led to substantial insight into the pathogenesis of HIV disease. In most of the original human studies, peripheral lymph nodes were the predominant sources for analyses into changes in lymphoid tissues associated with HIV and SIV infection, whereas more recent studies have expanded to include the GALT, where the earliest burst of virus replication occurs associated with marked depletion of CD4+ T cells. A variety of techniques, including sensitive molecular and imaging approaches to visualize virus and cells in location or suspension, have been employed to describe events associated with HIV disease. During acute HIV infection resulting from mucosal transmission, virus replication progressively amplifies from scattered lymphoid cells in the lamina propria of the gut to draining lymph nodes, leading to high levels of plasma viremia. The GALT plays a major role in the amplification of virus replication, and virus is disseminated from replication in the GALT to peripheral lymphoid tissues. A profound degree of cellular activation occurs within lymphoid tissues (see below) and is reflected in follicular or germinal center hyperplasia. At this time copious amounts of extracellular virions (both infectious and defective) are trapped on the processes of the follicular dendritic cells (FDCs) that form the stromal cell network in the light zones of lymph node germinal centers. Virions that have bound complement components on their surfaces attach to the surface of FDCs via interactions with complement receptors and likely via Fc receptors that bind to antibodies that are attached to the virions. The use of in situ hybridization techniques, including those that allow detection of viral RNA in the context of tissue architecture, has revealed that HIV is primarily expressed in CD4+ T cells of the paracortical area and, to a lesser extent, in specialized CD4+ T cells (see below) in light zones of germinal centers **(Fig. 208-23)**. The persistence of trapped virus on the surface of FDC likely reflects both a long-lived viral reservoir and virus that is replaced by continual expression in nearby CD4+ T cells. The trapped virus, either as whole virion or shed envelope, also serves as a continual activator of CD4+ T cells, thus driving further virus replication.

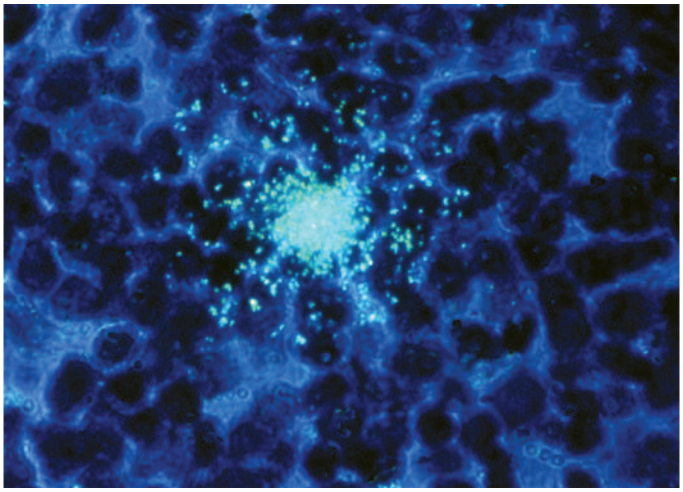

FIGURE 208-23 HIV in the lymph node of an HIV-infected individual. An individual cell infected with HIV shown expressing HIV RNA by in situ hybridization using a radiolabeled molecular probe. Original ×500. *(Reproduced with permission from G Pantaleo et al: HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. Nature 362:355, 1993.)*

During the early stages of HIV disease, the architecture of lymphoid tissues is generally preserved and may even be hyperplastic owing to an increased presence of B cells and specialized CD4+ T cells called follicular helper CD4+ T cells (TF_H) in prominent germinal centers. Extracellular virions can be seen by electron microscopy attached to FDC processes. The trapping of antigen is a physiologically normal function for the FDCs, which present antigen to B cells and secrete factors such as CXCL13 that retain B and TF_H cells in the light zones of germinal centers. These FDC functions, along with stimulatory factors produced by TF_H cells, contribute to the generation of B cell memory. However, in the case of HIV, persistent cellular activation, resulting in a shift to secretion of proinflammatory cytokines such as interleukin (IL) 1β, tumor necrosis factor (TNF) α, IFN-γ, and IL-6, can induce viral replication (see below) and diminish the effectiveness of the immune response against the virus. In addition, the CD4+ TF_H cells that are recruited into the germinal center to provide help to B cells in the generation of an HIV-specific immune response are highly susceptible to infection and may be an important component of the HIV reservoir. Thus, in HIV infection, a normal physiologic function of the immune system, i.e., the generation of an HIV-specific immune response that contributes to the clearance of virus, can also have deleterious consequences.

As HIV disease progresses, the architecture of lymphoid tissues begins to disrupt. Confocal microscopy reveals destruction of the fibroblastic reticular cell (FRC) and FDC networks in the T cell zone and B cell follicles/germinal centers, respectively. The mechanisms of destruction are not completely understood, but they are thought to be associated with collagen deposition causing fibrosis and a shift in the expression of certain cytokines, namely decreases in IL-7 and lymphotoxin α, which are critical to the maintenance of lymphoid tissues and their lymphocyte constituents, and increased levels of transforming growth factor (TGF) β. As the disease progresses to an advanced stage, there is complete disruption of the architecture of the lymphoid tissues, accompanied by dissolution of the FRC and FDC networks. At this point, the lymph nodes are "burnt out." This destruction of lymphoid tissue compounds the immunodeficiency of HIV disease and contributes both to the inability to control HIV replication and to the inability to mount adequate immune responses against opportunistic pathogens and vaccination. The events from primary infection to the ultimate destruction of the immune system are illustrated in **Fig. 208-24**. In nonhuman primate studies and some human studies that have examined GALT following SIV or HIV infection, the basal level of cellular activation combined with virus-mediated activation leads to the rapid infection and elimination of an estimated 50–90% of CD4+ T cells in the gut.

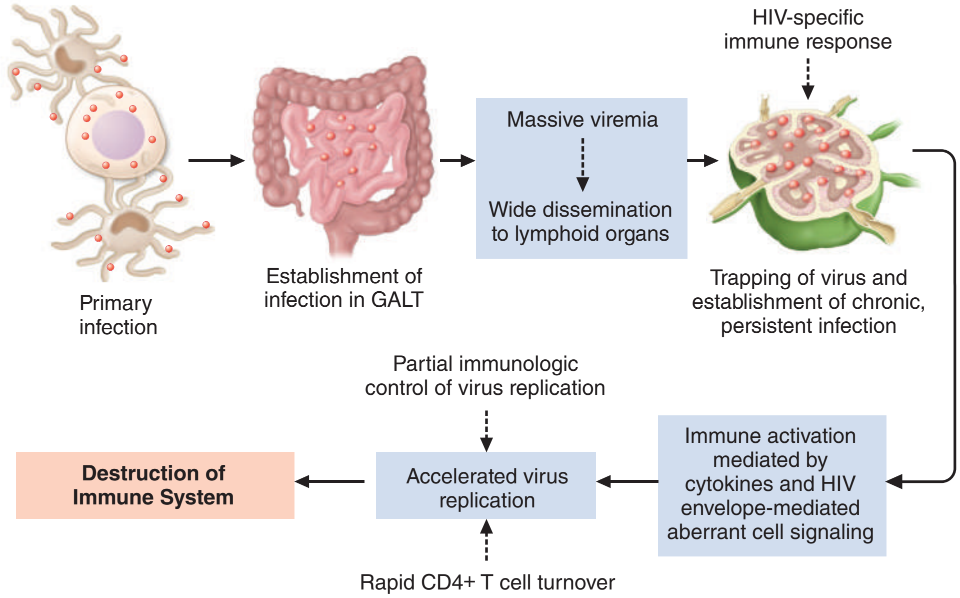
HIV-specific immune response
Massive viremia
Wide dissemination to lymphoid organs
Primary infection
Establishment of infection in GALT
Trapping of virus and establishment of chronic, persistent infection
Partial immunologic control of virus replication
Destruction of Immune System
Accelerated virus replication
Immune activation mediated by cytokines and HIV envelope-mediated aberrant cell signaling
Rapid CD4+ T cell turnover

FIGURE 208-24 Events that transpire from primary HIV infection through the establishment of chronic persistent infection to the ultimate destruction of the immune system. See text for details. GALT, gut-associated lymphoid tissue.

THE ROLE OF IMMUNE ACTIVATION AND INFLAMMATION IN HIV PATHOGENESIS

Activation of the immune system and variable degrees of inflammation are essential components of any appropriate immune response to a foreign antigen. However, immune activation and inflammation, which are aberrant in certain individuals with HIV, play a critical role in the pathogenesis of HIV disease as well as other chronic conditions associated with HIV infection. Immune activation and inflammation in individuals with HIV contribute substantially to (1) the replication of HIV, (2) the induction of immune dysfunction, and (3) the increased incidence of chronic conditions such as premature cardiovascular disease **(Table 208-4)**.

INDUCTION OF HIV REPLICATION BY ABERRANT IMMUNE ACTIVATION

The immune system is normally in a state of homeostasis, awaiting perturbation by foreign antigenic stimuli. Once the immune response deals with and clears the antigen, the system returns to relative quiescence **(Chap. 360)**. This is generally not the case in HIV infection where, in the untreated patient, virus replication is invariably persistent with very few exceptions and as a result immune activation is persistent. HIV replicates most efficiently in activated CD4+ T cells; in HIV infection, chronic activation provides the cell substrates necessary for persistent virus replication throughout the course of HIV disease, particularly in the untreated patient. Even in certain patients receiving ART whose levels of plasma viremia are suppressed to <20 copies per milliliter, there are low but detectable degrees of virus replication that drives low-level persistent immune activation. In addition, immune activation may result from RNA transcription of the integrated DNA of defective proviruses. From a virologic standpoint, although quiescent CD4+ T cells can be infected with HIV, albeit inefficiently, reverse transcription, integration, and virus spread are much more efficient in activated cells. Furthermore, cellular activation induces expression of virus in cells latently infected with HIV. In essence, immune activation and inflammation provide the engine that drives HIV replication. In addition to endogenous factors such as cytokines, multiple exogenous factors such as other microbes that induce cellular activation can enhance HIV replication and thus may play a role in HIV pathogenesis.

TABLE 208-4 Conditions Associated with Persistent Immune Activation and Inflammation in Patients with HIV Infection

Accelerated aging syndrome
Bone fragility
Cancers
Cardiovascular disease
Diabetes
Kidney disease
Liver disease
Neurocognitive dysfunction

Co-infection with a range of viruses, such as HSV types 1 and 2, cytomegalovirus (CMV), human herpesvirus (HHV) 6, Epstein-Barr virus (EBV), HBV, HCV, adenovirus, and HTLV-1 have been shown to upregulate HIV expression. In addition, infestation with nematodes has been shown to be associated with a heightened state of immune activation that facilitates HIV replication; in certain studies, deworming of the infected host has resulted in a decrease in plasma viremia. Two diseases of great global health significance, malaria and tuberculosis (TB), have been shown to increase HIV viral load in dually infected individuals. Globally, *Mycobacterium tuberculosis* is the most common opportunistic infection in people with HIV **(Chap. 183)**. In addition to the fact that individuals with HIV are more likely to develop active TB after exposure and to reactivate latent TB, it has been demonstrated that active TB can accelerate the course of HIV infection. It has also been shown that levels of plasma viremia are greatly elevated in individuals with HIV who have active TB and who are not receiving ART, compared with pre-TB levels and levels of viremia after successful treatment of the active TB. The situation is similar in the interaction between HIV and malaria parasites **(Chap. 231)**. Acute infection with *Plasmodium falciparum* of individuals with HIV increases viral load, and the increased viral load is reversed by effective treatment of malaria.

MICROBIAL TRANSLOCATION AND PERSISTENT IMMUNE ACTIVATION

One proposed mechanism of persistent immune activation involves the disruption of the mucosal barrier in the gut due to HIV replication in submucosal lymphoid tissue. As a result of this disruption, there is an increase in the products of bacteria, particularly lipopolysaccharide (LPS), that translocate from the bowel lumen through the damaged mucosa to the circulation, leading to persistent systemic immune activation and inflammation. This effect can persist even after the HIV viral load is brought to <20 copies/mL by ART. Other related factors that are thought to contribute to the pathogenesis of HIV include depletion in the GALT of IL-17–producing T cells, which are responsible for defense against extracellular bacteria and fungi, as well as alterations in gut microbiota and the metabolic pathways involved.

PERSISTENT IMMUNE ACTIVATION AND INFLAMMATION INDUCE IMMUNE DYSFUNCTION The immune activated state in HIV infection is reflected by hyperactivation of B cells leading to hypergammaglobulinemia; increased lymphocyte turnover; activation of monocytes; expression of activation markers and immune checkpoint receptors on CD4+ and CD8+ T cells; increased activation-associated cellular apoptosis and pyroptosis; lymph node hyperplasia, particularly during the chronic phase prior to disease progression; increased secretion of proinflammatory cytokines, particularly IL-6 and type I interferons; elevated levels of high-sensitivity C-reactive protein (CRP), CXC chemokine ligand 10 (CXCL10), D-dimer, neopterin, β_2-microglobulin, soluble (s) CD14, sTNFR, sCD27, sCD163, and sCD40L; and autoimmune phenomena (see "Autoimmune Phenomena," below). Even in the absence of direct infection of a target cell, HIV envelope proteins can interact with cellular receptors (CD4 molecules and chemokine receptors) to deliver potent activation signals resulting in calcium flux, the phosphorylation of certain proteins involved in signal transduction, co-localization of cytoplasmic proteins including those involved in cell trafficking, immune dysfunction, and, under certain circumstances, apoptosis and pyroptosis. From an immunologic standpoint, chronic exposure of the immune system to a particular antigen over an extended period may ultimately lead to an inability to sustain an adequate immune response to the antigen in question. In many chronic viral infections, including HIV infection, persistent viremia is associated with "functional exhaustion" of virus-specific T cells, decreasing their capacity to proliferate and perform effector functions. It has been demonstrated that this phenomenon of immune exhaustion may be mediated, at least in part, by the upregulation of inhibitory receptors on HIV-specific T cells, such as PD-1, LAG-3, and Tim-3 that are shared by both CD4+ and CD8+ T cells, as well as CTLA-4 on CD4+ and 2B4 and CD160 on CD8+ T cells. Furthermore, the ability of the immune system to respond to a broad spectrum of non-HIV antigens may be compromised if immunocompetent bystander cells are maintained in a state of chronic activation.

The deleterious effects of chronic immune activation on the progression of HIV disease are well established. As in most conditions of persistent antigen exposure, the host maintains activation of antigen (HIV)-specific responses but must also prevent excessive activation and potential immune-mediated damage to tissues. Certain studies suggest that normal immunoregulatory mechanisms that act to keep hyperimmune activation in check, particularly CD4+, FoxP3+, and CD25+ regulatory T cells (T-regs), may be dysfunctional or depleted in the context of advanced HIV disease. One possibility is a role for the inhibitory receptor LAG-3 (see below), which is overexpressed on exhausted T cells and shown to inhibit the proliferation of T-regs.

Apoptosis Apoptosis is a form of programmed cell death that is a normal mechanism for the elimination of effete cells in organogenesis as well as in the cellular proliferation that occurs during a normal immune response **(Chap. 360)**. Apoptosis can occur by intrinsic or extrinsic pathways, the latter of which is largely dependent on cellular activation, and in this regard the aberrant cellular activation associated with HIV disease is correlated with a heightened state of apoptosis. HIV can trigger activation-induced cell death through the upregulation of the death receptors, such as Fas/CD95, TNFR1, or TNF-related apoptosis-inducing ligand (TRAIL) receptors 1 and 2. Their corresponding ligands FasL, TNF, and TRAIL also are upregulated in HIV disease. HIV-induced stress and alterations in homeostasis also can trigger intrinsic apoptosis due to the downregulation of antiapoptotic proteins such as Bcl-2. Other mechanisms of HIV-induced cell death have been described, including autophagy, necrosis, necroptosis, and pyroptosis. The phenomenon of pyroptosis, an inflammatory form of cell death involving the upregulation of the proinflammatory enzyme caspase 1 and release of the proinflammatory cytokines IL-1β and IL-18, has been linked to a bystander effect of HIV replication on depletion of CD4+ T cells (see "Pathophysiology and Pathogenesis," above). The process of pyroptosis generates multimeric complexes called inflammasomes, which can also be activated by LPS. Certain viral gene products have been associated with enhanced susceptibility to apoptosis; these include Env, Tat, and Vpr. In contrast, Nef has been shown to possess antiapoptotic properties. The intensity of apoptosis correlates with the general state of activation of the immune system and not with the stage of disease or with viral burden. Multiple studies, including those examining lymphoid tissue, have demonstrated that the rate of apoptosis is elevated in HIV infection and that apoptosis is seen in "bystander" cells such as CD8+ T cells and B cells as well as in uninfected CD4+ T cells. It is likely that this bystander apoptosis of immunocompetent cells related to immune activation contributes to the general immunologic abnormalities in HIV disease.

MEDICAL CONDITIONS ASSOCIATED WITH PERSISTENT IMMUNE ACTIVATION AND INFLAMMATION IN HIV DISEASE It has become clear, as the survival of people with HIV has increased, that multiple, previously unrecognized medical complications are associated with HIV disease—and that these complications relate to chronic immune activation and inflammation (Table 208-4). These complications can appear even after patients have experienced years of ART-induced adequate control of viral replication (plasma viremia <50 copies per milliliter of plasma) for several years. Other chronic conditions that have been reported include bone fragility, certain cancers, diabetes, kidney and liver disease, and neurocognitive dysfunction, thus presenting an overall picture of accelerated aging.

Autoimmune Phenomena Autoimmune phenomena are commonly observed in people with HIV, and they reflect, at least in part, chronic immune activation and the dysregulation of B and T cells. Although these phenomena usually occur in the absence of autoimmune disease, a wide spectrum of clinical manifestations that may be associated with autoimmunity have been described (see "Immunologic and Rheumatologic Diseases," below). Autoimmune phenomena include antibodies against autoantigens expressed on intact lymphocytes and other cells, or against proteins released from dying cells. Antiplatelet and antierythrocyte antibodies have some clinical relevance in that they may contribute to thrombocytopenia and autoimmune hemolytic anemia, respectively, in HIV disease (see below). Antibodies to nuclear and cytoplasmic components of cells have been reported, as have antibodies to cardiolipin and phospholipids, as well as surface receptors, including CD4, and serum proteins. However, these manifestations are relatively low in the era of ART. Molecular mimicry, either from opportunistic pathogens or from HIV itself, also is a trigger or cofactor in autoimmunity. Antibodies against the HIV envelope proteins, especially gp41, often cross-react with host proteins; the best-known examples are antibodies directed against the membrane-proximal external region (MPER) of gp41 that also react with phospholipids and cardiolipin. The phenomenon of polyreactive HIV-specific antibodies may be beneficial to the host (see "Immune Response to HIV," below).

The increased occurrence and/or exacerbation of certain autoimmune diseases have been reported in HIV infection; these diseases include psoriasis, idiopathic thrombocytopenic purpura, autoimmune hemolytic anemia, Graves' disease, antiphospholipid syndrome, and primary biliary cirrhosis. Most of these manifestations were described prior to the advent of ART and have decreased in frequency since its widespread use. However, with increasing availability of ART, an *immune reconstitution inflammatory syndrome* (IRIS) has been increasingly observed in infected individuals, particularly those with low CD4+ T cell counts (see below). IRIS is an autoimmune-like phenomenon characterized by a paradoxical deterioration of clinical condition, which is usually compartmentalized to a particular organ system in individuals in whom ART has recently been initiated. It is associated with a decrease in viral load and at least partial recovery of immune competence, which is usually associated with increases in CD4+ T cell counts. The immunopathogenesis of this syndrome is felt to be related to an increase in immune response against the presence of residual antigens that are usually microbial and is most commonly seen with underlying mycobacterial (*Mycobacterium tuberculosis* [TB] or *avium* complex [MAC]), fungal (cryptococcal), and viral (CMV, HHV) infections. This syndrome is discussed in more detail below.

■ CYTOKINES AND OTHER SOLUBLE FACTORS IN HIV PATHOGENESIS

The immune system is homeostatically regulated by a complex network of immunoregulatory cytokines, which are pleiotropic and redundant and operate in an autocrine and paracrine manner. They are expressed continuously, even during periods of apparent quiescence of the immune system. On perturbation of the immune system by antigenic challenge, the expression of cytokines increases to varying degrees **(Chap. 360)**. Cytokines that are important components of this immunoregulatory network are thought to play major roles in HIV disease, during both the early and chronic phases of infection. A potent proinflammatory "cytokine storm" is induced during the acute phase of HIV infection, likely a response of inflammatory cells to virus replicating at very high levels. Cytokines and chemokines that are induced during this early phase include the type I interferon IFN-α, IL-15, and CXCL10, followed by IL-6, IL-12, and TNF-α, and a delayed peak of the anti-inflammatory cytokine IL-10. Soluble factors of innate immunity also are induced shortly after infection, including neopterin and β-microglobulin. Several of these early-expressed cytokines and factors are not downregulated following the early phase of HIV infection, as seen in other self-resolving viral infections, and persist during the chronic phase of infection and contribute to maintaining high levels of immune activation. Among the cytokines and factors associated with early innate immune responses, they are intended to contain viral replication, although paradoxically most are potent inducers of HIV expression/replication because of their ability to induce immune activation that leads to enhanced viral production and an increase in readily available target cells for HIV (activated CD4+ T cells). The induction of IFN-α, one of the first cytokines induced during primary HIV infection and an important element of innate immune sensing, is thought to play a particularly important role in HIV pathogenesis by inducing a large number of IFN-associated genes that activate the immune system, alter the homeostasis of CD4+ T cells, and influence the virus variants that are selected during the HIV transmission bottleneck. Other cytokines that are elevated during the chronic phase of HIV infection and linked to immune activation include IFN-γ, the *CC-chemokine* RANTES (CCL5), macrophage inflammatory protein (MIP) 1β (CCL4), and IL-18.

Multiple cytokines and soluble factors have been associated with HIV pathogenesis at various stages of disease, in various tissues or organs, and in the regulation of HIV replication. Plasma levels of IP-10 are predictive of disease progression, whereas the proinflammatory cytokine IL-6, marker of monocyte/macrophage activation soluble CD14 (sCD14), and coagulation marker D-dimer are associated with increased risk of all-cause mortality in people with HIV. In particular, IL-6, sCD14, and D-dimer are associated with increased risk of cardiovascular disease and other causes of death, even in individuals receiving ART. IL-18 has also been shown to play a role in the development of the HIV-associated lipodystrophy syndrome. Elevated levels of TNF-α and IL-6 have been demonstrated in plasma and cerebrospinal fluid (CSF), and increased expression of TNF-α, IL-1β, IFN-γ, and IL-6 has been demonstrated in the lymph nodes of people with HIV prior to disease progression and a shift to TGF-β in advanced disease (see "Lymphoid Organs and HIV Pathogenesis," above). RANTES (CCL5), MIP-1α (CCL3), and MIP-1β (CCL4) **(Chap. 360)** inhibit infection by and spread of R5 HIV-1 strains, while *stromal cell–derived factor* (SDF) 1 inhibits infection by and spread of X4 strains. The mechanisms whereby the CC-chemokines RANTES (CCL5), MIP-1α (CCL3), and MIP-1β (CCL4) inhibit infection of R5 strains of HIV, or SDF-1 blocks X4 strains of HIV, involve blocking of the binding of the virus to its co-receptors, the CC-chemokine receptor CCR5 and the CXC-chemokine receptor CXCR4, respectively. Other soluble factors that have not yet been fully characterized, such as soluble CD8 antiviral factor (CAF), also have been shown to suppress HIV replication, independent of co-receptor usage.

LYMPHOCYTE TURNOVER IN HIV INFECTION

The immune systems of patients with HIV infection are characterized by a profound increase in lymphocyte turnover that is immediately reduced with effective ART. Studies utilizing in vivo or in vitro labeling of lymphocytes in the S-phase of the cell cycle have demonstrated a tight correlation between the degree of lymphocyte turnover and plasma viremia. This increase in turnover is seen in CD4+ and CD8+ T lymphocytes as well as B lymphocytes and can be observed in peripheral blood and lymphoid tissue. Mathematical models derived from these data suggest that one can view the lymphoid pool as consisting of dynamically distinct subpopulations of cells that are differentially affected by HIV infection. A major consequence of HIV infection appears to be a shift in cells from a more quiescent pool to a pool with a higher turnover rate. It is likely that a consequence of a higher rate of turnover is a higher rate of cell death. It has been suggested that the more rapid decline in CD4+ compared with CD8+ T cells may be linked to alterations in inflammatory and homeostatic cytokines that cause increased activation-induced death without replenishment of CD4+ T cells. (See **Table 208-5** for additional mechanisms of depletion.)

TABLE 208-5 Proposed Mechanisms of CD4+ T Cell Dysfunction and Depletion

DIRECT MECHANISMS	INDIRECT MECHANISMS
Loss of plasma membrane integrity due to viral budding	Aberrant intracellular signaling events
Accumulation of unintegrated viral DNA Activation of DNA-dependent protein kinase during viral integration into host genome	Autoimmunity
Interference with cellular RNA processing	Innocent bystander killing of viral antigen–coated cells
Intracellular gp120–CD4 autofusion events	Apoptosis, pyroptosis (caspase 1–associated inflammation), autophagy
Syncytia formation	Inhibition of lymphopoiesis from reduced survival cytokines and lymphoid tissue integrity Activation-induced cell death Elimination of HIV-infected cells by virus-specific immune responses

THE ROLE OF VIRAL RECEPTORS AND CO-RECEPTORS IN HIV PATHOGENESIS

CCR5 AND CXCR4 As mentioned above, HIV-1 utilizes two major co-receptors along with CD4 to bind to, fuse with, and enter target cells; these co-receptors are CCR5 and CXCR4, which are also receptors for certain endogenous chemokines. Strains of HIV that utilize CCR5 as a co-receptor are referred to as *R5 viruses*. Strains of HIV that utilize CXCR4 are referred to as *X4 viruses*. Many virus strains are *dual tropic* in that they utilize both CCR5 and CXCR4; these are referred to as *R5X4 viruses*.

The natural chemokine ligands for the major HIV co-receptors can readily block entry of HIV. For example, the CC-chemokines RANTES (CCL5), MIP-1α (CCL3), and MIP-1β (CCL4), which are the natural ligands for CCR5, block entry of R5 viruses, whereas SDF-1, the natural ligand for CXCR4, blocks entry of X4 viruses. The mechanism of inhibition of viral entry is a steric inhibition of binding that is not dependent on signal transduction **(Fig. 208-25)**.

The transmitting virus is almost invariably an R5 virus that predominates during the early stages of HIV disease, although in the era of deep sequencing, more X4 variants have been detected in early disease than were previously reported. In the absence of ART or in therapeutic failures, there is a transition to a predominantly X4 virus in approximately half of individuals infected with subtype B virus. The transition is often preceded by dual R5X4 strains, and detection of X4 variants is associated with a relatively rapid decline in CD4+ T cell counts, increased HIV plasma viremia, and progression of disease. However, the other half of infected individuals progress in their disease while maintaining predominance of an R5 virus, and individuals infected with non–subtype B clades more rarely switch from CCR5 tropism to CXCR4 tropism than do those infected with subtype B. The reason for this difference is unclear.

The basis for the tropism of different envelope glycoproteins for either CCR5 or CXCR4 relates to the ability of the HIV envelope, including the third variable region (V3 loop) of gp120, to interact with these co-receptors. In this regard, binding of gp120 to CD4 induces a conformational change in gp120 that increases its affinity for the relevant co-receptor. Finally, R5 viruses are more efficient in infecting monocytes/macrophages and microglial cells of the brain (see "Neuropathogenesis in HIV Disease," below).

THE INTEGRIN *α4β7* The integrin *α4β7* is an accessory receptor for HIV. It is not essential for the binding and infection of a CD4+ T cell by HIV; however, it likely plays an important role in the transmission of HIV at mucosal surfaces such as the genital tract and gut and contributes somewhat to the pathogenesis of HIV disease. The integrin *α4β7*, which is the gut homing receptor for peripheral T cells, binds

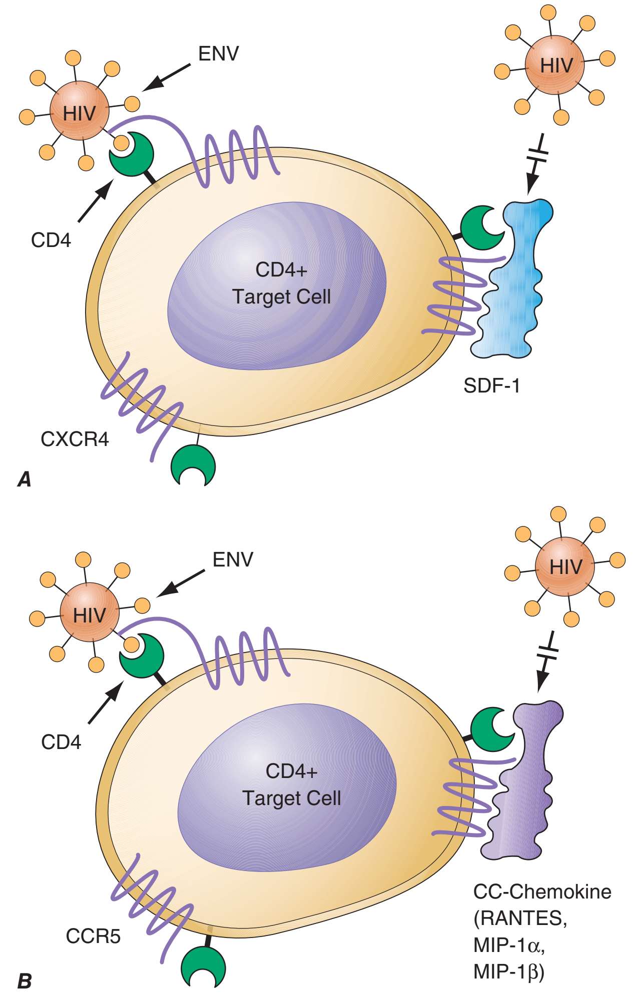

FIGURE 208-25 Model for the role of co-receptors CXCR4 and CCR5 in the efficient binding and entry of X4 (***A***) and R5 (***B***) strains of HIV-1, respectively, into CD4+ target cells. Blocking of this initial event in the virus life cycle can be accomplished by inhibition of binding to the co-receptor by the normal ligand for the receptor in question. The ligand for CXCR4 is stromal cell–derived factor (SDF-1); the ligands for CCR5 are RANTES, MIP-1α, and MIP-1β.

in its activated form to a specific tripeptide in the V2 loop of gp120, resulting in rapid activation of leukocyte function–associated antigen 1 (LFA-1), the central integrin in the establishment of virologic synapses, which facilitate efficient cell-to-cell spread of HIV. It has been demonstrated that $\alpha4\beta7^{high}$ CD4+ T cells are more susceptible to productive infection than are $\alpha4\beta7^{low-neg}$ CD4+ T cells because this cellular subset is enriched with metabolically active CD4+ T cells that are $CCR5^{high}$. These cells are present in the mucosal surfaces of the gut and genital tract. Importantly, it has been demonstrated that the virus that is transmitted during sexual exposure binds much more efficiently to α4β7 than does the virus that diversifies from the transmitting virus over time by mutation, particularly involving the accumulation of glycogens on the surface of the HIV envelope (see "Early Events in HIV Infection: Primary Infection and Initial Dissemination of Virus," above).

CELLULAR TARGETS OF HIV

CD4+ T lymphocytes and to a lesser extent CD4+ cells of the myeloid lineage are the principal targets of HIV and are the only cells that can be productively infected with HIV. Circulating DCs have been reported to express low levels of CD4, although high expression of the restriction factor SAMHD1 in myeloid (mDC) and plasmacytoid (pDC) DCs limits HIV replication in these cells by depleting intracellular pools of dNTPs and directly degrading viral RNA. Epidermal Langerhans cells express CD4 and have been infected by HIV in vivo, although they too restrict replication by high expression of the host restriction factor langerin. As has been shown in vivo for DCs, FDCs, and B cells, Langerhans cells are more likely to bind and transfer virus to activated CD4+ T cells than to be productively infected themselves.

Of potential clinical relevance is the demonstration that thymic precursor cells, which were assumed to be negative for CD3, CD4, and CD8 molecules, express low levels of CD4 and can be infected with HIV in vitro. In addition, human thymic epithelial cells transplanted into an immunodeficient mouse can be infected with HIV by direct inoculation of virus into the thymus. Since these cells may play a role in the normal regeneration of CD4+ T cells, it is possible that their infection and depletion contribute, at least in part, to the impaired ability of the CD4+ T cell pool to completely reconstitute itself in certain infected individuals in whom ART has suppressed plasma viremia to below the level of detection (see below). In addition, CD34+ monocyte precursor cells have been shown to be infected in vivo in patients with advanced HIV disease. It is likely that these cells express low levels of CD4, and therefore it is not essential to invoke CD4-independent mechanisms to explain the infection. The clinical relevance of this finding is unclear.

QUALITATIVE AND QUANTITATIVE ABNORMALITIES OF MONONUCLEAR CELLS

CD4+ T Cells The primary immunopathogenic lesion in HIV infection involves CD4+ T cells, and the range of CD4+ T cell abnormalities in advanced HIV infection is broad. The defects are both quantitative and qualitative and ultimately impact virtually every limb of the immune system, indicating the critical dependence of the integrity of the immune system on the inducer/helper function of CD4+ T cells. In advanced HIV disease, most of the observed immune defects can ultimately be explained by the quantitative depletion of CD4+ T cells. However, T cell dysfunction can be demonstrated in patients early in the course of infection, even when the CD4+ T cell count is in the low–normal range. The degree and spectrum of dysfunctions increase as the disease progresses, reflecting the range of CD4+ T cell functional heterogeneity, especially in lymphoid tissues. One of the first sites of intense HIV replication is in the GALT where CD4+ T_H17 cells reside; they are important for host defense against extracellular pathogens in the intestinal mucosa and help maintain the integrity of the gut epithelium. In HIV infection, they are depleted by direct and indirect effects of viral replication and cause loss of gut homeostasis and integrity, as well as a shift toward a T_H1 phenotype. Studies have shown that even after many years of ART, normalization of the CD4+ T cells in the GALT remains incomplete. In lymph nodes, HIV perturbs another important subset of the CD4+ helper T lineage, namely TF_H cells (see "Lymphoid Organs and HIV Pathogenesis," above). TF_H cells, which are derived either directly from naïve CD4+ T cells or from other T_H precursors, migrate into B cell follicles during germinal center reactions and provide help to antigen-specific B cells through cell–cell interactions and secretion of cytokines to which B cells respond, the most important of which is IL-21. In addition, it has been shown that people with HIV who have broadly neutralizing antibodies have higher frequencies of memory TF_H CD4+ T cells. As with T_H17 cells, TF_H cells are highly susceptible to HIV infection. However, in contrast to T_H17 and most other CD4+ T cell subsets, the number of TF_H cells is increased in lymph nodes of people with HIV, especially those who are viremic. It is unclear whether this increase is helpful to responding B cells, although the likely outcome is that the increase in numbers is detrimental to the quality of the humoral immune response against HIV (see "Immune Response to HIV," below). In addition, defects of central memory cells are a critical component of HIV immunopathogenesis. The progressive loss of antigen-specific CD4+ T cells has important implications for the control of HIV infection. In this regard, there is a correlation between the maintenance of HIV-specific CD4+ T cell proliferative responses and improved control of infection. Essentially every T cell function has been reported to be abnormal at some stage of HIV infection. Loss of polyfunctional HIV-specific CD4+ T cells, especially those that produce IL-2, occurs early in disease, whereas IFN-producing CD4+ T cells are maintained longer and do not correlate with control of HIV viremia. Other abnormalities include impaired expression of IL-2 receptors, defective IL-2 production, reduced expression of the IL-7 receptor (CD127), and a decreased proportion of CD4+ T cells that express CD28, a major co-stimulatory molecule necessary for the normal activation of T cells, which is also depleted due to aging. Cells lacking expression of CD28 do not respond normally to activation signals and may express markers of terminal activation including HLA-DR, CD38, and CD45RO. As mentioned above ("The Role of Immune Activation and Inflammation in HIV Pathogenesis"), a subset of CD4+ T cells referred to as *T regulatory cells*, or T-regs, may be involved in damping aberrant immune activation that propagates HIV replication. The presence of these T-reg cells correlates with lower viral loads and higher CD4+/CD8+ T cell ratios. A loss of this T-reg capability with advanced disease may be detrimental to the control of virus replication.

It is difficult to explain completely the profound immunodeficiency noted in people with HIV solely based on direct infection and quantitative depletion of CD4+ T cells. This is particularly apparent during the early stages of HIV disease, when CD4+ T cell numbers may be only marginally decreased. In this regard, it is likely that CD4+ T cell dysfunction results from a combination of depletion of cells due to direct infection of the cell and multiple virus-related but indirect effects on the cell (Table 208-5). Several of these effects have been demonstrated ex vivo and/or by the analysis of cells isolated from the peripheral blood. Soluble viral proteins, particularly gp120, can bind with high affinity to the CD4 molecules on uninfected T cells and monocytes; in addition, virus and/or viral proteins can bind to DCs or FDCs. HIV-specific antibodies can recognize these bound molecules and potentially collaborate in the elimination of the cells by ADCC. HIV envelope glycoproteins gp120 and gp160 manifest high-affinity binding to the CD4 molecule as well as to various chemokine receptors. Intracellular signals transduced by gp120 through both CD4 and CCR5/CXCR4 have been associated with a number of immunopathogenic processes including anergy, apoptosis, and abnormalities of cell trafficking. The molecular mechanisms responsible for these abnormalities include dysregulation of the T cell receptor–phosphoinositide pathway, p56lck activation, phosphorylation of focal adhesion kinase, activation of the MAP kinase and ras signaling pathways, and downregulation of the co-stimulatory molecules CD40 ligand and CD80.

The inexorable decline in CD4+ T cell counts that occurs in most untreated people with HIV may result in part from the inability of the immune system to regenerate over an extended period of time the rapidly turning over CD4+ T cell pool efficiently enough to compensate for both HIV-mediated and naturally occurring attrition of cells. In this regard, the degree and duration of decline of CD4+ T cells at the time of initiation of therapy is an important predictor of the restoration of these cells. A person who maintains a very low CD4+ T cell count for a considerable period before the initiation of ART almost invariably has an incomplete reconstitution of such cells. At least two major mechanisms may contribute to the failure of the CD4+ T cell pool to reconstitute itself adequately over the course of HIV infection. The first is the destruction of lymphoid precursor cells, including thymic and bone marrow progenitor cells; the other is the gradual disruption of the lymphoid tissue architecture and microenvironment, which is essential for efficient regeneration of immunocompetent cells. Finally, during the advanced stages of CD4+ T lymphopenia, there are increased serum levels of the homeostatic cytokine IL-7. It was initially felt that this elevation was a homeostatic response to the lymphopenia; however, recent findings suggest that the increase in serum IL-7 was a result of reduced utilization of the cytokine related to the loss of cells expressing the IL-7 receptor, CD127, which serves as a normal physiologic regulator of IL-7 production.

CD8+ T Cells A relative CD8+ T lymphocytosis is generally associated with high levels of HIV plasma viremia and likely reflects an immune response to the virus as well as dysregulated homeostasis associated with generalized immune activation. During the late stages of HIV infection, there may be a significant reduction in the numbers of CD8+ T cells despite the presence of high levels of viremia. HIV-specific CD8+ CTLs have been demonstrated in people with HIV early in the course of disease, and their emergence often coincides with a decrease in plasma viremia—an observation that is a factor in the proposal that virus-specific CTLs can control HIV disease for a finite period of time in a certain percentage of infected individuals. However, emergence of HIV escape mutants that ultimately evade these HIV-specific CD8+ T cells has been described in most people with HIV who are not receiving ART. In addition, as the disease progresses, the functional capability of these cells gradually decreases, at least in part due to the persistent nature of HIV infection that causes functional exhaustion via the upregulation of inhibitory receptors such as PD-1, TIGIT, LAG-3, and TIM-3 on HIV-specific CD8+ T cells (see "The Role of Immune Activation and Inflammation in HIV Pathogenesis," above). As chronic immune activation persists, there are also systemic effects on CD8+ T cells, such that as a population they assume an abnormal phenotype characterized by expression of activation markers such as co-expression of HLA-DR and CD38 with an absence of expression of the IL-2 receptor (CD25) and a reduced expression of the IL-7 receptor (CD127). In addition, CD8+ T cells lacking CD28 expression are increased in HIV disease, reflecting a skewed expansion of a less differentiated CD8+ T cell subset. This skewing of subsets is also associated with diminished polyfunctionality, a qualitative difference that distinguishes elite controllers from progressors. Elite controllers can also be distinguished from progressors by the maintenance in the former of a high proliferative capacity of their HIV-specific CD8+ T cells coupled to increases in perforin expression and elimination of infected targets, characteristics that are markedly diminished in advanced HIV disease. It has been reported that the phenotype of CD8+ T cells in people with HIV may be of prognostic significance. Those individuals whose CD8+ T cells developed a phenotype of HLA-DR+/CD38– following seroconversion had stabilization of their CD4+ T cell counts, whereas those whose CD8+ T cells developed a phenotype of HLA-DR+/CD38+ had a more aggressive course and a poorer prognosis. In addition to the defects in HIV-specific CD8+ CTLs, functional defects in other MHC-restricted CTLs, such as those directed against influenza and CMV, have been demonstrated. CD8+ T cells secrete a variety of soluble factors that inhibit HIV replication, including the CC-chemokines RANTES (CCL5), MIP-1α (CCL3), and MIP-1β (CCL4) and potentially several yet-unidentified factors. The presence of high levels of HIV viremia in vivo as well as exposure of CD8+ T cells in vitro to HIV envelope, both of which are associated with aberrant immune activation, have been shown to be associated with a variety of cellular functional abnormalities. Furthermore, since the integrity of CD8+ T cell function depends in part on adequate inductive signals from CD4+ T cells, the defect in CD8+ CTLs is likely compounded by the quantitative loss and qualitative dysfunction of CD4+ T cells.

B Cells The predominant defect in B cells from people with HIV is one of aberrant cellular activation, which is reflected by increased propensity to terminal differentiation and immunoglobulin secretion, as well as increased expression of markers of activation and exhaustion. As a result of activation and differentiation in vivo and induction of inhibitory pathways of regulation, B cells from HIV viremic patients manifest a decreased capacity to undergo cell signaling and mount a proliferative response ex vivo. B cells from people with HIV manifest enhanced spontaneous secretion of immunoglobulins in vitro, a process that reflects their highly differentiated state in vivo. There is also an increased incidence of EBV-related B cell lymphomas in people with HIV that are likely due to combined effects of defective T cell immune surveillance and increased B cell turnover that increases the risk of oncogenesis. Untransformed B cells cannot be infected with HIV, although HIV or its products can activate B cells directly. B cells from patients with high levels of viremia bind virions to their surface via the CD21 complement receptor. It is likely that in vivo activation of B cells by replication-competent or defective virus as well as viral products during the viremic state accounts at least in part for their activated phenotype. B cell subpopulations from people with HIV undergo a number of changes over the course of HIV disease, including the attrition of resting memory B cells and replacement with several aberrant memory and differentiated B cell subpopulations that collectively express reduced levels of CD21 and either increased expression of activation markers or inhibitory receptors associated with functional exhaustion. The more activated and differentiated B cells are also responsible for increased secretion of immunoglobulins and increased susceptibility to Fas-mediated apoptosis. In more advanced disease, there is also the appearance of immature B cells associated with CD4+ T cell lymphopenia. Despite increased frequencies of germinal center B cells and CD4+ TF_H cells, both of which are required for effective humoral immunity, cognate B cell–CD4+ T cell interactions in lymphoid tissues are perturbed in people with HIV, especially those with persistent viremia. In vivo, the aberrant activated state of B cells manifests itself by hypergammaglobulinemia and by the presence of circulating immune complexes that bind the B cells and restrict their capacity to respond to further stimulation. People with HIV respond poorly to primary and secondary immunizations with protein and polysaccharide antigens. Using immunization with influenza vaccine, it has been demonstrated that there is a memory B cell defect in people with HIV, particularly those with high levels of HIV viremia. There is also evidence that responses to HIV and non-HIV antigens in infected individuals, especially those who remain viremic, are enriched in abnormal subsets of B cells that either are highly prone to apoptosis or show signs of functional exhaustion. Taken together, these B cell defects are likely responsible at least in part for the inadequate humoral response to HIV as well as to decreased response to vaccinations and the increase in certain bacterial infections seen in advanced HIV disease in adults. In addition, they likely contribute to the inadequacy of host defenses against bacterial infections that play a role in the increased morbidity and mortality of children with HIV. The absolute number of circulating B cells also may be depressed in HIV infection; this phenomenon likely reflects increased activation-induced apoptosis as well as a redistribution of cells out of the circulation and into the lymphoid tissue—phenomena that are associated with ongoing viral replication.

Monocytes/Macrophages Circulating monocytes are generally normal in number in people with HIV; however, there is evidence of increased activation within this lineage. The increased level of sCD14 and other biomarkers (see above) reported in people with HIV is an indirect marker of monocyte activation in vivo. Levels of sCD14 can remain elevated in individuals whose plasma viremia has been suppressed by ART for several years, an indicator of the residual immune activation and inflammation observed in HIV infection and effects on the monocyte/macrophage lineage. Multiple other abnormalities of circulating monocytes have been reported in people with HIV, many of which may be related directly or indirectly to aberrant in vivo immune activation. In this regard, increased levels of lipopolysaccharide (LPS) are found in the sera of people with HIV due, at least in part, to translocation across the gut mucosal barrier (see above). LPS is a highly inflammatory bacterial product that preferentially binds to macrophages through CD14 and Toll-like receptors, resulting in cellular activation. In the peripheral blood, expansion of monocytes that express the intermediate and nonclassical marker CD16 and markers of activation (HLA-DR) and stimulation (CD40 and CD86) has been described, especially in viremic individuals. Activated monocytes are also responsible for secretion of inflammatory cytokines and chemokines observed in HIV infection, including CXCL10, IL-1β, and IL-6. Monocytes express the CD4 molecule and several co-receptors for HIV on their surface, and thus are potential targets of HIV infection. However, in vivo infection of circulating monocytes is difficult to demonstrate, although infection of tissue macrophages and macrophage-lineage cells in the brain (infiltrating macrophages or resident microglial cells) and lung (pulmonary alveolar macrophages) can be demonstrated easily. Tissue macrophages are an important source of HIV during the inflammatory response associated with opportunistic infections and can serve as persistent reservoirs of HIV infection, thus representing an obstacle to the eradication of HIV by antiretroviral drugs.

Dendritic and Langerhans Cells DCs and Langerhans cells are not productively infected with HIV, likely in part due to their expression of host restriction factors, including APOBEC3G and SAMHD1 (see above). However, they are thought to play an important role in the initiation of HIV infection by virtue of the ability of HIV to bind to cell-surface C-type lectin receptors, particularly DC-SIGN (see above) and langerin. However, while langerin provides a host barrier for replication by trafficking HIV to acidic compartments for degradation, DC-SIGN retains HIV in early endosomal compartments. This allows efficient presentation of intact virus to CD4+ T cell targets that become infected; complexes of infected CD4+ T cells and DCs provide an optimal microenvironment for virus replication. Furthermore, pDCs secrete large amounts of IFN-α in response to viral infections and as such play an important role in innate sensing of HIV during early phase of infection. The numbers of circulating pDCs and mDCs are decreased in HIV infection through mechanisms that remain unclear, although several studies have shown increased lymphoid tissue recruitment of DCs associated with lymphoid hyperplasia and inflammation. The mDCs are also involved in the initiation of adaptive immunity in draining lymph nodes by presenting antigen to T cells and B cells, as well as by secreting cytokines such as IL-12, IL-15, and IL-18 that activate other immune cells, although these functions are perturbed in HIV infection.

Natural Killer Cells and Innate Lymphoid Cells NK cells represent the prototypical member of innate lymphoid cells (ILCs) that collectively provide tissue homeostasis and immunosurveillance against virus-infected cells, certain tumor cells, and allogeneic cells **(Chap. 360)**. There are no convincing data that HIV productively infects NK cells in vivo; however, functional abnormalities in NK cells have been observed throughout the course of HIV disease, and the severity of these abnormalities increases as disease progresses. NK cells are part of the innate immune system and act by direct killing of infected cells and secretion of antiviral cytokines and chemokines. In early HIV infection there is an increase in the activation of NK cells, and the capacity to secrete IFN-γ is maintained, although they manifest reduced cytotoxic function as a result of altered maturation. During chronic HIV infection, both NK cell cytotoxicity and cytokine secretion become impaired. Given that HIV infection of target cells downregulates HLA-A and B, but not HLA-C and D molecules, this may explain in part the relative inability of NK cells to kill HIV-infected target cells. However, the NK cell impairments, especially in patients with high levels of virus replication, are associated with an expansion of an "anergic" CD56–/CD16+ NK cell subset. This abnormal subset of NK cells manifests an increased expression of inhibitory NK cell receptors (iNKRs) and a substantial decrease in expression of natural cytotoxicity receptors (NCRs) and shows a markedly impaired lytic activity. The overrepresentation of this abnormal subset of NK cells may explain in part the observed defects in NK cell function in people with HIV and likely begins to occur during primary infection. The relative expression of iNKRs and NCRs—as well as their ligands, which include HLA class I molecules—has an impact on the antiviral functions associated with NK cells, including direct killing and ADCC. Polymorphisms in iNKR and NCR alleles have been linked to HIV-1 disease outcomes, and there are indications that the early control of HIV may be mediated by cytotoxic NK cell-mediated responses. NK cells may also serve as sources of HIV-inhibitory soluble factors, including CC-chemokines such as MIP-1α (CCL3), MIP-1β (CCL4), and RANTES (CCL5). Finally, both inflammatory cytokines and alterations in the GALT of HIV infected individuals disrupt NK cells and other ILCs.

GENETIC FACTORS IN HIV-1 AND AIDS PATHOGENESIS

Candidate gene approaches and genome-wide association studies (GWAS) have identified polymorphisms in host genes that contribute to inter-individual variation in (1) the risk of acquiring HIV, (2) the steady-state levels of HIV that are established soon after infection (virologic set point), (3) the rate at which untreated HIV infection progresses to AIDS defined by a CD4+ T cell count that is lower than 200 cells/mm^3 and/or development of AIDS-defining illnesses, (4) the level of immune reconstitution (e.g., CD4+ cell counts) achieved and risk of non-AIDS-associated diseases after initiation of virally suppressive antiretroviral therapy (ART), and (5) adverse reactions to antiretroviral agents. The key polymorphisms that influence these five outcomes are summarized in **Table 208-6**, and their identification has

TABLE 208-6 Host Genetic Factors Influencing HIV/AIDS Pathogenesis and Therapy Responses

GENE[a]	GENETIC VARIATION	MECHANISMS[b]	GENETIC ASSOCIATIONS[c]
Genes in MHC Locus			
HLA-B	*B*27* and *B*57*	Altered presentation of specific HIV antigens	Slower progression to AIDS; lower viral load
	*B*35*	Restriction of specific HIV peptide presentation	Faster progression to AIDS; higher viral load
	HLA-Bw4	Providing ligands for activating KIR	Slower progression to AIDS
	*B*57:01*	Altered presentation of specific HIV antigens (as above); possible abacavir-specific activation of cytokine-producing CD8+ T cells in carriers of this allele	Slower progression to AIDS; higher risk of abacavir-associated hypersensitivity
	HLA-B? 21M allele	Enhanced HLA-E expression levels correlated with higher HLA-A expression and inhibition of NKG2A-expressing cells	Higher viral load, reduced CD4+ counts, and accelerated disease progression
	*B*57:03* bearing the rs2523608-A allele	Altered presentation of specific HIV antigens	Variant overexpressed in HIV-1 controllers of African descent
HLA class I allele	Homozygosity of *HLA*-class I alleles	Reduced repertoire for epitope recognition	Faster progression to AIDS; increased risk of mother-to-child transmission
	Shared donor–recipient *HLA* alleles	Preadaptation of HIV strains	Faster progression to AIDS
	Rare *HLA* alleles	Limited adaptation of HIV strains; less frequent escape mutants	Protection against HIV infection
HLA class II allele	*HLA-DRB1* alleles	Influence on protein specificity of CD4+ T cell responses to HIV Gag and Nef proteins	*HLA-DRB1*15:02*—lower viral load *HLA-DRB1*03:01*—higher viral load
HLA extended haplotype	A1-B8-DR3-DQ2 (AH 8.1)	Increased proinflammatory responses; higher TNF-α production	Faster progression to AIDS
HLA-C	rs9264942-C allele (35 kb upstream of *HLA-C*) in linkage with rs67384697-Del	Increased expression of HLA-C by reducing binding of miRNA-148a	Decreased viral load set point
	rs5010528-G (1 kb upstream of HLA-C)	Unknown	Higher risk of developing nevirapine-associated hypersensitivity
HCP5	rs2395029-G	Linkage disequilibrium with *HLA-B*57:01*	Lower viral load and slower progression to AIDS
MICA	Noncoding SNV near *MICA*, rs4418214-T	May affect HLA class I peptide presentation—linkage with protective *HLA-B* alleles	Enriched in HIV-1 controllers
PSORS1C3	rs3131018-A	May affect HLA class I peptide presentation	Enriched in HIV-1 controllers
ZNRD1	rs9261174-C	Possible interference in processing of HIV transcripts; influence *ZNRD1* expression	Slower disease progression to AIDS
Chemokine Receptors			
CCR5	rs333: 32-bp deletion in the ORF (Δ32) found in persons of European descent	Truncated CCR5 protein; reduced co-receptor activity of R5 HIV strain	*Δ32/Δ32*: CCR5-null state associated with resistance to acquiring HIV infection
			Δ32/wild type: slower progression to AIDS; better CD4+ T cell recovery during ART
	Promoter SNVs, haplotypes (*HHA* to *HHG*2*)	Altered CCR5 expression, e.g., HHE haplotype correlates with high CCR5 expression	HHE/HHE: increased HIV susceptibility and faster progression to AIDS
	rs1015164 G→A (34 kb downstream from *CCR5* and close to *CCRL2*)	Increased expression of the lncRNA *RP11-24-11.2*, which corresponds to an antisense transcript that overlaps CCR5 (*CCR5AS*); results in increased CCR5 expression	rs1015164A allele associated with higher viral load
CCR2	rs1799864: SNV in ORF (64 V→I)	Linkage with polymorphisms in *CCR5* promoter	64I-bearing haplotype associated with delayed progression to AIDS
CCRL2	rs3204849: SNV in ORF (167 Y→F)	SNV in linkage with *CCR5* haplotype	167F associated with accelerated progression to AIDS and PCP
CXCR6	rs2234358: G→T in the 3′ UTR	Trafficking of effector T cells and activation of NK T cells; minor HIV co-receptor	Prevalence of rs2234358-T lower in long-term nonprogressors and viremic controllers of African descent
CX3CR1	SNVs in ORF: rs3732379 (249 V→I) and rs3732378 (280 T→M)	Alleles bearing 249I and 280M reduce receptor expression and binding of fractalkine, the CX3CR1 ligand	249I and 280M associated with faster AIDS progression in persons of European descent
DARC	rs2814778: Promoter SNV (–46T→C) found in persons of African descent	–46C/C associated with absent *DARC* expression (Duffy null), low neutrophil counts, and altered circulating chemokine levels as well as HIV binding to RBCs and trans-infection of HIV-1	–46C/C: increased risk of acquiring HIV but slower HIV disease progression; Duffy null–associated low neutrophil trait associated with increased HIV risk
Chemokines			
CCL3L, CCL4L	Gene copy number of *CCL3L* and *CCL4L*	High numbers of *CCL3L* and *CCL4L* gene-containing segmental duplications correlated with high *CCL3L* and *CCL4L* levels	Gene copy number lower than population median associated with increased HIV-AIDS susceptibility and lower CD4+ T cell recovery during ART
CCL5	Promoter SNVs	Altered gene expression	Influenced HIV-AIDS susceptibility
CCL2	rs1024611: Promoter SNV (–2578 T→G)	–2578G allele: increased *CCL2* expression and monocyte recruitment	–2578G/G associated with increased risk of developing HIV-1–associated dementia and faster AIDS onset

(*Continued*)

TABLE 208-6 Host Genetic Factors Influencing HIV/AIDS Pathogenesis and Therapy Responses (Continued)

GENE[a]	GENETIC VARIATION	MECHANISMS[b]	GENETIC ASSOCIATIONS[c]
CXCL12	rs7919208: Promoter SNV (G→A)	rs7919208A creates a new transcription factor binding site associated with increased *CXCL12* expression	rs7919208A associated with higher susceptibility to HIV-related non-Hodgkin's lymphoma
Cytokines			
IL-6	rs1800795: Promoter SNV (–174 G→C)	–174G/G associated with increased IL-6 and CRP levels	–174G/G associated with high risk of KS development and variable recovery of CD4+ T cells during ART
IL-7RA	rs6897932: Coding SNV (244 T→I)	244 I/I associated with increased signal transduction and proliferation in response to IL-7	244 I/I associated with faster CD4+ T cell recovery during ART
IL-10	rs1800872: Promoter SNV (–592 C→A)	–592A associated with decreased IL-10 levels	–592A associated with increased HIV infection risk and AIDS progression rate
Drug-Metabolizing Enzyme Gene			
CYP2B6	Multiple variants (e.g., rs3745274 [516 G→ T], i.e., *CYP2B6*6*)	*CYP2B6* variants influence enzyme activity	516T/T associated with higher risk of adverse reactions to efavirenz
Innate Immunity Genes			
MBL	Alleles defined by 3 coding SNVs	Low plasma concentration and structural variation of MBL protein	Slow progression to AIDS with heterozygosity for coding SNVs
	X allele (promoter SNV –221)	Decreased levels of MBL protein	Faster progression to AIDS with X/X genotype
APOBEC3G	rs8177832: ORF SNV (186 H→R)	Reduced anti–HIV-1 activity	186R associated with rapid AIDS progression in persons of African descent
APOBEC3F	Haplotype tagged by rs2076101 in ORF (231 I→V)	231V variant may influence Vif-mediated *APOBEC3F* degradation	231V associated with lower viral load, slower progression to AIDS and PCP
TLR7	rs179008: ORF SNV (32A→T) on Chr. X	Lower TLR7 mRNA translation efficiency and impaired TLR7-dependent IFN-α production	rs179008-T associated with lower viral load and cell-associated HIV-1 DNA in women
PARD3B	rs11884476 near exon 20: C→G	Direct interaction with HIV signaling through SMAD family of proteins	rs11884476-G associated with slower progression to AIDS
IFNL4	rs368234815: Frameshift mutation (TT→ΔG)	Polymorphism in *IFNL4* gene in linkage with a *IFNL3* variant; this haplotype influences *IFNL3* levels	rs368234815-ΔG associated with higher prevalence of AIDS-defining illnesses and potentially increased HIV-1 infection risk
	rs8099917: T→G	Unknown	rs8099917-G associated with higher susceptibility to KS
Others			
ApoE	E4 allele defined by two coding SNVs	ApoE is an HIV-1–inducible inhibitor of HIV-1 replication and infectivity in macrophages	E4/E4 associated with rapid AIDS progression and HIV-associated dementia
ApoL1/MYH9	Several risk haplotypes, including G1	Overexpression of the *ApoL1* kidney risk variants may increase kidney cell death	Increased risk for HIV-associated nephropathy
RYR3	rs2229116: ORF SNV (A →G)	Unknown; potential impact on calcium signaling and homeostasis	rs2229116-G associated with subclinical atherosclerosis during ART
PROX1	rs17762192-G 36 kb upstream of *PROX1*	Unknown; presumably due to its impact on *PROX1* expression, which is a negative regulator of IFN-γ	rs17762192-G associated with reduced rate of HIV disease progression
Gene–Gene Interaction			
KIR+HLA	*KIR3DS1* interaction with *HLA-Bw4-80Ile*	Altered NK cell activity required to eliminate HIV-infected cells	*KIR3DS1/HLA-Bw4-80Ile* associated with delayed AIDS onset
	KIR2DL3 interaction with *HLA-C1*	Reduction of inhibitory KIR likely resulting in increased immune activation, impaired killing of latently infected cells, and a higher proviral burden	*HLA-C1*$^{+}$ *KIR2DL*$^{+}$ associated with better immune recovery during ART
	KIR3DL1 I47V interaction with *HLA-B*57:01*	Variation in an immune NK cell receptor that binds *B*57:01*, modifying the protective effect of *B*57:01*	Increasing copy numbers of 47V associated with lower viral load in persons carrying *HLA-B*57*
LILRB2+HLA	*LILRB2* interaction with HLA class I	Regulation of dendritic cells by *LILRB2-HLA* engagement	Control of HIV-1
CCL3L1+ CCR5	Low *CCL3L1* gene copies + detrimental *CCR5* genotypes	Low *CCL3L1* and high *CCR5* expression	Increased HIV/AIDS susceptibility and reduced immune reconstitution during ART

[a] Representative genes and polymorphisms and [b] possible mechanisms are listed. [c] Some of the associations are population specific and may display cohort-specific effects. Most of the associations were derived from persons of European descent.

Abbreviations: APOBEC, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like; ApoE, apolipoprotein E; ApoL1, apolipoprotein L1; ART, antiretroviral therapy; CCL, CC ligand; CCL3L, CCL3-like; CCR5, CC chemokine receptor 5; CCR5AS, CCR5 antisense RNA; CCRL2, CC chemokine receptor–like 2; CRP, C-reactive protein; CYP2B6, cytochrome P450 family 2 subfamily B member 6; CXCL12, chemokine (C-X-C motif) ligand 12; CXCR6, chemokine (C-X-C motif) receptor 6; CX3CR1, chemokine (C-X3-C motif) receptor 1; DARC, Duffy antigen receptor for chemokines; Del, deletion; HCP5, HLA class I histocompatibility antigen protein P5; HHE, human haplogroup E; HLA, human leukocyte antigen; IFN, interferon; IFNL3, interferon λ3 gene; IFNL4, interferon λ4 gene; IL, interleukin; IL-7RA, interleukin 7 receptor α; KIR, killer cell immunoglobulin-like receptors; KS, Kaposi's sarcoma; LILRB2, leukocyte immunoglobulin-like receptor B2; MBL, mannose-binding lectin; MHC, major histocompatibility complex; MICA, MHC class I polypeptide-related sequence A; MYH9, myosin heavy chain 9; NK, natural killer; ORF, open reading frame; PARD3B, par-3 family cell polarity regulator beta; PCP, *Pneumocystis jirovecii* pneumonia; PROX1, prospero homeobox 1; PSORS1C3, psoriasis susceptibility 1 candidate 3; RYR3, ryanodine receptor 3; SMAD, mothers against decapentaplegic homolog; SNV, single nucleotide variant; rs#, SNV identification number; TLR7, Toll-like receptor 7; TNF-α, tumor necrosis factor α; UTR, untranslated region; VL, viral load; ZNRD1, zinc ribbon domain containing 1; +, present; –, absent.

Sources: SK Ahuja, W He. Reviews for additional information; P An et al: Trends Genet 26:119, 2010; J Fellay: Antivir Ther 14:731, 2009; RA Kaslow et al: J Infect Dis 191:S68, 2005; D van Manen et al: Retrovirology 9:70, 2012; MP Martin et al: Immunol Rev 254:245, 2013; S Limou et al: Front Immunol 4:118, 2013; PJ McLaren et al: Curr Opin HIV AIDS 10:110, 2015; PJ McLaren et al: Proc Natl Acad Sci USA 112:14658, 2015; PJ McLaren, M Carrington: Nat Immunol 16:577, 2015; P An et al: PLoS Genet 12:e1005921, 2016; F Pereyra et al: Science 330:1551, 2010; I Bartha et al: PLoS Comput Biol 13:e1005339, 2017; S Kulkarni et al: Nat Immunol 20:824, 2019; S Le Clerc et al: Front Genet 10:799 2019; V Kalidasan et al: Front Microbiol 11:46 2020; SN Gingras et al: Hum Genet 139:865 2020.

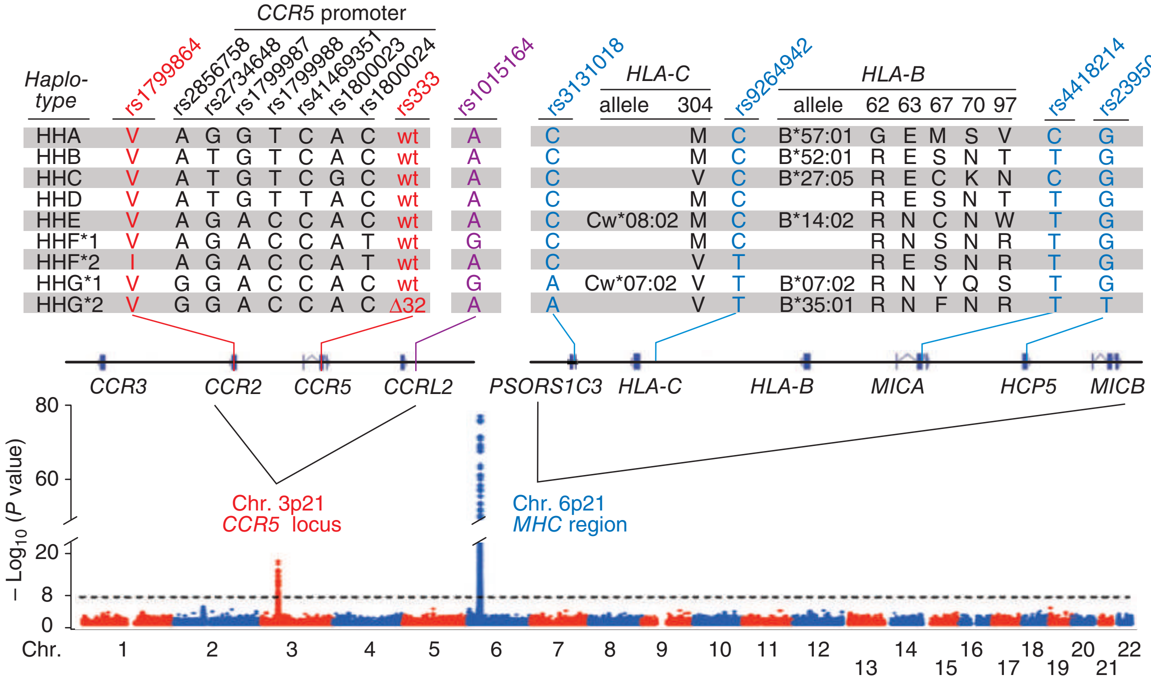

Haplo-type				*CCR5* promoter								*HLA-C*			*HLA-B*							
	rs1799864	rs2856758	rs2734648	rs1799987	rs1799988	rs41469351	rs1800023	rs1800024	rs333	rs1015164	rs3131018	allele	304	rs9264942	allele	62	63	67	70	97	rs4418214	rs2395029
HHA	V	A	G	G	T	C	A	C	wt	A	C		M	C	B*57:01	G	E	M	S	V	C	G
HHB	V	A	T	G	T	C	A	C	wt	A	C		M	C	B*52:01	R	E	S	N	T	T	G
HHC	V	A	T	G	T	C	G	C	wt	A	C		V	C	B*27:05	R	E	C	K	N	C	G
HHD	V	A	T	G	T	T	A	C	wt	A	C		M	C		R	E	S	N	T	T	G
HHE	V	A	G	A	C	C	A	C	wt	A	C	Cw*08:02	M	C	B*14:02	R	N	C	N	W	T	G
HHF*1	V	A	G	A	C	C	A	T	wt	G	C		M	C		R	N	S	N	R	T	G
HHF*2	I	A	G	A	C	C	A	T	wt	A	C		V	T		R	E	S	N	R	T	G
HHG*1	V	G	G	A	C	C	A	C	wt	G	A	Cw*07:02	V	T	B*07:02	R	N	Y	Q	S	T	G
HHG*2	V	G	G	A	C	C	A	C	Δ32	A	A		V	T	B*35:01	R	N	F	N	R	T	T

FIGURE 208-26 Schema depicting haplotypes within two regions that contribute significantly to HIV-AIDS susceptibility. *Top:* Haplotypes (left, *CCR5*; right, HLA alleles). *Bottom:* GWAS Manhattan plots schematized. Chr, chromosome. *Horizontal dotted line:* genome level significance threshold.

greatly advanced our understanding of the genes that influence HIV-AIDS pathogenesis and ART-associated immune reconstitution. Of particular interest are polymorphisms in two chromosomal regions, as they are associated with consistent effects on HIV acquisition, virologic set point, and/or rates of HIV disease progression: the region in chromosome 3 that includes the gene that encodes the HIV co-receptor *CC-chemokine receptor 5* (*CCR5*) and the *major histocompatibility locus* (*MHC*) in chromosome 6 **(Fig. 208-26)**.

GENETICS OF CCR5: FROM BENCH TO BEDSIDE While the discovery of CCR5 as a major co-receptor for cell entry of HIV-1 was established by in vitro studies, genetic association studies established its seminal role in HIV pathogenesis. Initial in vitro studies revealed that a 32-bp deletion (*Δ32*) in the coding region of *CCR5* contributes to resistance to CCR5 using R5 strains of HIV. The *CCR5 Δ32* allele encodes a truncated protein that is not expressed on the cell surface. Congruently, genotype-phenotype association studies in large cohorts demonstrated that individuals homozygous for the *CCR5 Δ32* allele (*Δ32/Δ32*) lack CCR5 surface expression and are highly resistant to acquiring HIV infection; heterozygosity for the *CCR5 Δ32* allele is associated with a lower risk of acquiring HIV.

The distribution of the *CCR5 Δ32* allele is population specific. Approximately 1% of individuals of European ancestry are homozygous for the *CCR5 Δ32* allele. Depending on the geographic region in Europe, up to 18% of individuals are heterozygous for the *CCR5 Δ32* allele. The *CCR5 Δ32* allele is rare in other populations. The evolutionary pressure that resulted in the emergence of the *CCR5 Δ32* allele in the European population remains unknown and has been speculated to be secondary to an ancestral pandemic, such as the plague.

Subsequent studies identified single nucleotide variants (SNVs) in the promoter (regulatory) region of *CCR5* that influence gene expression levels. Alleles bearing specific cassettes of linked polymorphisms (haplotypes) were identified and designated as human haplogroups A to G*2 (*HHA* to *HHG*2*) (Fig. 208-26). The *CCR5 Δ32* polymorphism is found on the *HHG*2* haplotype. *CCR5* haplotypes *A–D* versus *E–G*2* differ by bearing GT versus AC at polymorphic sites rs1799987 and rs1799988 (Fig. 208-26). *CCR5-HHA* haplotype represents the ancestral haplotype (found in chimpanzees) and is associated with lower *CCR5* gene expression, whereas the *CCR5-HHE* haplotype is associated with higher CCR5 expression. Methylation of DNA is a common epigenetic signaling mechanism that cells use to lock genes in the "off" position, and polymorphisms in *CCR5* haplotypes may mediate their effects by influencing DNA methylation levels in the *CCR5* locus. The *CCR5-HHE* and *CCR5-HHA* haplotypes are more sensitive and resistant, respectively, to T cell activation–induced demethylation of the *CCR5* locus.

In worldwide populations, *HHE* and *HHC* are prevalent haplotypes, whereas the ancestral *HHA* haplotype is more common in persons of African ancestry. The associations of *CCR5* haplotypes with HIV acquisition and/or HIV disease course are largely consistent with their effects on *CCR5* gene expression. For example, homozygosity for the *CCR5-HHE* haplotype is associated with an increased risk of acquiring HIV, progressing rapidly to AIDS, and reduced immune recovery during ART. The *HHA* haplotype is associated with slower disease progression in African populations and has been speculated to be a basis for why chimpanzees (who all carry the ancestral *CCR5-HHA* haplotype) naturally infected with SIV resist disease progression. The pairing of the *HHC* and *CCR5 Δ32*-bearing *HHG*2* haplotypes (*HHC/HHG*2* genotype) is associated with a lower risk of acquiring HIV infection and slower rate of HIV disease progression, whereas the pairing of the *HHE* haplotype with the *HHG*2* haplotype is associated with the opposite effects. The *CCR2*-64I–bearing *HHF*2* haplotype is associated with a slower HIV disease course.

Consistent with these genetic associations, polymorphisms in genes encoding ligands for CCR5 have also been associated with variable HIV susceptibility and disease progression rates. Examples include copy number variations of *CCL3L1* and SNVs in *CCL5*. The sum of these studies established a pivotal role of CCR5 and its ligands in HIV-AIDS pathogenesis and, potentially, immune recovery.

The discovery that the *CCR5 Δ32/Δ32* genotype is associated with strong resistance to HIV infection, and that uninfected persons bearing this genotype did not appear to have impaired immunity, led to the development of two kinds of novel therapies. First, it spurred the development of a new class of therapies approved by the U.S. Food and Drug Administration: entry inhibitors (e.g., maraviroc) that block the interaction of CCR5 with the HIV envelope. Second, it led to the evaluation of novel experimental cellular therapies. A person with HIV and acute myelogenous leukemia was given an allogeneic stem cell transplantation from an HLA-compatible person whose cells lacked expression of CCR5 due to the *Δ32/Δ32* genotype. There was no evidence of HIV-1 infection for 13 years in the patient who underwent the transplant; the patient eventually died due to recurrence of leukemia.

This observation provided a proof of concept for an HIV cure and led to the development of additional novel cellular therapies involving autologous transplantation of CD4+ T cells in which the *CCR5* gene is inactivated ex vivo using new gene editing procedures. Similar cellular strategies have had mixed success, mainly due to the latent viral reservoir in various tissues.

DISCOVERY OF HLA CLASS I ALLELES THAT ASSOCIATE WITH VIROLOGIC CONTROL OF HIV INFECTION There is a strong association between variations within the *HLA-B* gene with protective (e.g., _HLA-B*57_ and _HLA-B*27_ alleles) or detrimental (e.g., _HLA-B*35_ allele) outcomes during HIV infection. Carriage of the _HLA-B*57_ and/or _HLA-B*27_ alleles is associated with slower disease progression. The beneficial effects of these alleles may relate in part to their associations with a lower virologic set point as well as to higher cell-mediated immunity in people with HIV. The protective effect of the _HLA-B*57_ and _HLA-B*27_ alleles on the HIV disease course is underscored by the finding that the prevalence of these alleles is higher among persons with long-term nonprogression and persons who control HIV replication spontaneously (elite controllers). In contrast, the _HLA-B*35_ allele has been associated with faster progression to AIDS and higher viral load. The prevalence of the *HLA-B* alleles differs between populations. _HLA-B*57:01_ in Europeans and _HLA-B*57:03_ in persons of African descent are the protective alleles. In some populations (e.g., Japanese) where the _HLA-B*57/HLA-B*27_ alleles are absent, _HLA-B*51_ is associated with a protective phenotype.

Possession of the protective *HLA-B* alleles is associated with broader and stronger CD8+ T cell responses to HIV epitopes. The mechanisms underlying the differential effects of the *HLA-B* alleles on the course of HIV disease may relate to differences in the ability of antigen-presenting cells to present immunodominant HIV epitopes to T helper or cytotoxic T lymphocytes in the context of MHC-encoded molecules. This may result in differential immune responses that influence viral replication. In this regard, the *HLA-B* alleles that impact the course of HIV disease differ in their amino acid residues in the HLA-B peptide-binding groove; this difference may play a critical role in virologic control.

The *HLA-B–21M* allele does not influence *HLA-B* gene expression; however, it is in linkage with *HLA-B* haplotypes that are associated with higher *HLA-A* and *HLA-E* expression. Higher *HLA-A* levels associate with poorer control of HIV as well as higher viral load, reduced CD4+ T cell counts, and accelerated progression to AIDS. HLA-E is the ligand for natural killer (NK) cell NKG2A, an inhibitory receptor. Engagement of NKG2A with HLA-E inhibits NK cells that would normally be potent eliminators of virally infected cells. Thus, targeting NKG2A might provide a therapeutic avenue for HIV treatment.

Investigators have also examined the influence of extended *HLA* haplotypes (linked alleles) on the course of HIV disease. The extended *HLA* ancestral haplotype (AH) 8.1 is defined by the presence of *HLA-A1*, *HLA-B8*, and *HLA-DR3* alleles. AH 8.1 is the most common ancestral haplotype in persons of European descent (present in 10%) and is associated with multiple autoimmune diseases in HIV-seronegative persons. These associations of AH 8.1 are thought to be due to a genetically determined hyperresponsiveness characterized by high TNF-α production and lack of complement C4A. Strong epidemiologic data indicate that carriage of AH 8.1 in HIV-seropositive persons is associated with a rapid decline in the number of CD4+ T cells and faster progression to AIDS development. Gene–gene interactions between HLA alleles and other genes (e.g., killer cell immunoglobulin-like receptors) also may influence HIV disease progression rates.

POLYMORPHISMS IDENTIFIED BY GWAS THAT ASSOCIATE WITH VIROLOGIC CONTROL AND DISEASE PROGRESSION GWAS have not identified additional genetic variations that associate with the risk of HIV-1 acquisition, presumably due to a paucity of well-characterized risk cohorts in which level of exposure has been quantified. By contrast, large-scale GWAS have identified SNVs, especially in the MHC, that influence HIV viral load, including in a large group of individuals termed *HIV controllers (including elite controllers)* who spontaneously (without ART) control viral replication. GWAS in people with HIV of European ancestry identified four SNVs in genes in the HLA class I loci that associated with virologic control. These SNVs are within or in the vicinity of *PSORS1C3*, *HLA-C*, *MICA*, and *HCP5* genes (Fig. 208-26). As noted in this figure, the individual effects of these alleles are difficult to discern because of linkage disequilibrium. The protective effects of the SNVs in *HCP5* and *MICA* may relate to their linkage with known protective *HLA-B* alleles. The protective *HCP5* allele is in linkage disequilibrium with the _HLA-B*57:01_ allele, and the protective *MICA* allele tags with the _HLA-B*57:01_ and _HLA-B*27:05_ alleles. The protective *HLA-C* SNV is associated with higher HLA-C expression, which has been associated with viral control and better HIV outcomes. This protective SNV (rs9264942; T→C) resides 35 kb upstream of the *HLA-C* gene and is in strong linkage disequilibrium with a 3′-UTR indel263 SNV (rs67384697; G→deletion), generating the T-G or C-deletion haplotypes **(Fig. 208-27)**. miR-148a binds to the 3′-UTR region encompassing the rs67384697 SNV and silences *HLA-C* expression. Binding of miR-148a to the 3′-UTR is disrupted on the mRNA transcribed from the C-deletion haplotype; this disruption associates with less silencing of the mRNA and therefore higher HLA-C cell surface expression, which associates with better HIV disease outcomes (Fig. 208-27). Conversely, binding of miR-148a to the 3′-UTR is intact on the mRNA transcribed from the T-G haplotype; this binding associates with silencing of the mRNA and therefore lower HLA-C cell surface expression associates with worse HIV disease outcomes (Fig. 208-27). GWAS in persons of African descent have identified an SNV (rs2523608) that tags the _HLA-B*57:03_ allele that is known to associate with HIV-1 control and a slower disease course. Together, these GWAS data underscore the importance of variations in HLA class I loci in control of viral replication.

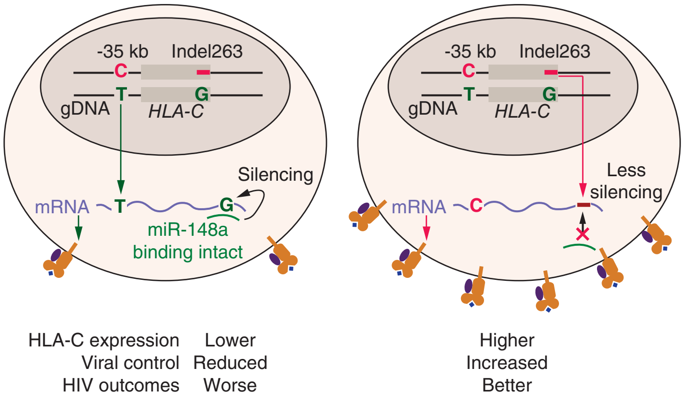

FIGURE 208-27 Linkage disequilibrium between two variants in the HLA-C locus and their influence on binding of miR-148a to the 3′-untranslated region (UTR). Altered binding of miR-148a associates with HLA-C protein expression levels and, in turn, viral control and HIV disease outcomes. Effects associated with T-G (left) and C-del (right) haplotypes are depicted. The C-deletion haplotype prevents binding of miR-148a to 3′-UTR of HLA-C (less silencing). Kb, kilobase.

A recent GWAS suggested that an SNV (rs1015164G→A) approximately 34 kb downstream of the *CCR5* loci is associated with a higher viral load set point (Fig. 208-26) and lower CD4+ T cell counts in therapy-naïve HIV-seropositive persons. rs1015164 maps to a lncRNA gene in proximity to the *CCRL2* gene (Fig. 208-26). The lncRNA is transcribed from the antisense strand of *CCR5* and was therefore named *CCR5AS*. The rs1015164A allele is associated with higher expression of *CCR5AS* in CD4+ T cells, which in turn was associated with increased levels of *CCR5* mRNA. Although the detrimental effect of the rs1015164A allele was suggested to be independent of the detrimental effects of the abovementioned *CCR5-HHE* haplotype, further investigation is warranted as the rs1015164A allele and *CCR5-HHE* haplotype are in a high degree of linkage disequilibrium.

Most GWAS studies have been performed in European populations, limiting generalizability to other populations. Additionally, GWAS are generally not suitable for identifying rare variants (<1% prevalence). Therefore, next-generation sequencing (NGS) approaches were suggested to identify these rare variants. However, a recent NGS study suggests that exonic variants with large effect sizes are unlikely to have a major contribution to host control of HIV infection. Mathematical modeling revealed that variations in host genes may explain about 10% of the observed variability in HIV viral load, whereas viral genetic diversity may explain 29% of the variability.

GENETIC ASSOCIATIONS WITH SPECIFIC AIDS AND NON-AIDS CONDITIONS

Carotid artery disease Many of the non-AIDS events in HIV-seropositive individuals resemble those attributable to immune senescence and those found in the HIV-seronegative aging population. A functional SNV in the ryanodine receptor 3 (*RYR3*) gene was found to be associated with an increased risk of common carotid intima–media thickness (cIMT), which is a surrogate for subclinical atherosclerosis. Functional studies on *RYR3* and its isoforms demonstrate a major role of these receptors in modulating endothelial function and atherogenesis via calcium-signaling pathways, providing a biologically plausible mechanism by which the SNV in *RYR3* may associate with increased cIMT risk.

Kidney disease HIV-1–associated nephropathy (HIVAN) is a form of focal sclerosing glomerulonephritis caused by direct infection of kidney epithelial cells with HIV. HIVAN is more common in persons of African descent. There is evidence that polymorphisms in the *MYH9* gene and in the neighboring *APOL1* gene are a strong determinant of susceptibility to HIVAN in persons of African descent. The effect of carrying two *APOL1* risk alleles explains nearly 35% of HIVAN. Overexpression of the *APOL1* kidney risk variants may associate with increased kidney cell death.

HIV-associated neurocognitive disorder HIV-associated neurocognitive disorder (HAND) comprises a spectrum of neurocognitive deficits due to HIV infection. Variations in the apolipoprotein E (*ApoE*) gene have strong associations with Alzheimer's disease in the HIV-seronegative population. In HIV-seropositive persons, possession of the *E4/E4* genotype has been associated with dementia, peripheral neuropathy, and impairment in cognition as well as immediate and delayed verbal memory. Macrophage recruitment and activation play a central role in the development of many of the HAND syndromes. Variations in chemokines that play an influential role in macrophage activation and recruitment, namely *CCL2* (*MCP-1*) and *CCL3* (*MIP-1*α), have been shown to influence the risk of developing HAND. Variations in mitochondrial genes also have been associated with a risk of AIDS and HAND. A GWAS identified a polymorphism in chromosome 14 in the T cell receptor α locus that may influence neurocognitive outcomes.

HIV-1–associated pneumocystis pneumonia Human Apobec3 cytidine deaminases are intrinsic resistance factors to HIV-1. However, HIV-1 encodes a viral infectivity factor (Vif) that degrades *APOBEC3* proteins. Association studies suggest a role of genetic variations in the APOBEC3 family in HIV disease. A common haplotype derived from 6 SNVs in the *APOBEC3F* gene and tagged by a codon-changing variant is associated with a significantly lower viral load set point, slower rate of progression to AIDS, and delayed development of *Pneumocystis jirovecii* pneumonia (PCP). In addition, a coding SNV in the *CCRL2* gene is associated with accelerated progression to AIDS and rapid development of PCP.

HIV-related non-hodgkin lymphoma (NHL) The relative risk of developing NHL in HIV-seropositive persons is highly elevated compared with the general population. NHL represents approximately 34% of all identified cancers in HIV-seropositive persons. A recent GWAS identified a promoter SNV in the *CXCL12* gene that was associated with higher susceptibility to develop HIV-related NHL. The effect of this SNV is likely causal as it creates new transcription factor binding sites, impacting *CXCL12* expression.

ASSOCIATIONS WITH ART-RELATED ADVERSE EVENTS Abacavir, an effective antiretroviral agent, is associated with significant risk of hypersensitivity reactions (2–9% of cases). Interestingly, while the *HLA-B*57:01* allele is associated with a slower HIV disease course, possession of this allele is associated with a higher risk of abacavir-associated hypersensitivity, possibly due to the abacavir-specific activation of cytokine-producing CD8+ T cells only in *HLA-B*57:01* carriers. Pharmacogenetic screening for the *HLA-B*57:01* allele is recommended before initiation of abacavir treatment.

The antiretroviral agent nevirapine is associated with hypersensitivity reactions in 6–10% of patients, including Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN). rs5010528G, a strong proxy for *HLA-C*04:01* carriage, was associated with high risk of SJS and TEN during nevirapine treatment. In addition, efavirenz was among the first antiretroviral agents to be co-formulated into single-pill regimens for mass rollout globally. Several genetic variants in the drug-metabolizing enzyme CYP2B6 have been associated with high efavirenz plasma concentrations and increased risk of adverse neuropsychiatric effects. For example, homozygosity for one such variant, rs3745274 T/T, increases the risk of adverse reactions to efavirenz up to fivefold, and this risk genotype is much more common in Africans (13.7%) than Europeans (5.6%).

NEUROPATHOGENESIS IN HIV DISEASE

While there has been a remarkable decrease in the incidence of the severe forms of HIV encephalopathy among those with access to treatment in the era of effective ART, people with HIV can still experience milder forms of neurocognitive impairment despite adequate ART. Factors that contribute to the neurocognitive decline include lack of complete control of HIV replication in the brain; production of HIV proteins that may be neurotoxic; low CD4+ T cell nadir; chronic immune activation; comorbidities such as drug abuse, microvascular disease, older age, and diabetes; and the potential for neurotoxicity of certain antiretroviral drugs. HIV has been demonstrated in the brain and CSF of infected individuals with and without neuropsychiatric abnormalities. As opposed to lymphoid tissues, there are no resident lymphocytes in the brain. The main cell types that are infected in the brain in vivo are the perivascular macrophages and the microglial cells, which can sometimes form syncytia resulting in multinucleated giant cells; low-level viral replication is also seen in perivascular astrocytes. It has been proposed that monocytes that have already been infected in the blood can migrate into the brain, where they then reside as macrophages, or macrophages can be directly infected while residing within the brain. The precise mechanisms whereby HIV enters the brain are unclear; however, they are thought to relate, at least in part, to the ability of virus-infected and immune-activated macrophages to induce adhesion molecules such as E-selectin and vascular cell adhesion molecule 1 (VCAM-1) on brain endothelium. Other studies have demonstrated that HIV gp120 enhances the expression of intercellular adhesion molecule 1 (ICAM-1) in glial cells and HIV Tat protein can disrupt the tight junctions of the brain endothelial cells to facilitate entry of HIV-infected cells into the CNS. Virus isolates from the brain are preferentially R5 strains as opposed to X4 strains; in this regard, people with HIV who are heterozygous for *CCR5-Δ32* appear to be relatively protected against the development of HIV encephalopathy. Once HIV enters the brain due to pressures of the local environment, it evolves to develop distinct sequences in the env, tat, and LTR genes. These unique sequences have been associated with neurocognitive dysfunction; however, it is unclear if they are causal (see below).

People with HIV may manifest white matter lesions as well as neuronal loss. The white matter lesions are due to axonal injury and a disruption of the blood-brain barrier and not due to demyelination. Given the absence of evidence of HIV infection of neurons, HIV-mediated effects on neurons are thought to involve indirect pathways whereby viral proteins, particularly gp120 and Tat, trigger the release of endogenous neurotoxins from macrophages and to a lesser extent from astrocytes. In addition, it has been demonstrated that both HIV-1 Nef and Tat can induce chemotaxis of leukocytes, including monocytes, into the CNS. Neurotoxins can be released from monocytes as a consequence of infection and/or immune activation. Monocyte-derived neurotoxic factors have been reported to kill neurons via a variety of mechanisms including activation of the *N*-methyl-D-aspartate (NMDA) receptors and induction of oxidative stress. In addition, HIV gp120 shed by virus-infected monocytes could cause neurotoxicity by antagonizing the function of vasoactive intestinal peptide (VIP), by elevating intracellular calcium levels, and by decreasing neurotrophic factor levels in the cerebral cortex. A variety of monocyte-derived cytokines can contribute directly or indirectly to the neurotoxic effects in HIV infection; these include TNF-α, IL-1, IL-6, TGF-β, IFN-γ, platelet-activating factor, and endothelin. Furthermore, among the CC-chemokines, elevated levels of monocyte chemotactic protein 1 (MCP-1 or CCL-2) in the brain and CSF have been shown to correlate best with the presence and degree of HIV encephalopathy in ART-naïve patients. In addition, infection and/or activation of monocyte-lineage cells can result in increased production of eicosanoids, quinolinic acid, nitric oxide, excitatory amino acids such as L-cysteine and glutamate, arachidonic acid, platelet-activating factor, free radicals, TNF-α, and TGF-β, which may contribute to neurotoxicity. Astrocytes may play diverse roles in HIV neuropathogenesis. Reactive gliosis or astrocytosis has been demonstrated in the brains of people with HIV, and TNF-α and IL-6 have been shown to induce astrocyte proliferation. In addition, astrocyte-derived IL-6 can induce HIV expression in infected cells in vitro. Furthermore, it has been suggested that astrocytes may downregulate macrophage-produced neurotoxins. Evidence of neuronal injury can be demonstrated by measuring neurofilament levels in CSF. Treatment with ART leads to improvement in neuropsychiatric manifestations and a decrease in these cytokine levels in CSF, suggesting that they are driven by the virus or by its products. However, even in patients on long-term ART, there may be evidence of persistently activated lymphocytes in the CSF. It is unclear if these lymphocytes may contribute to neuronal injury in the brain or are critical for controlling the CNS viral reservoir. However, some individuals may develop a subacute encephalitis due to an IRIS reaction (see below). This often occurs weeks or a few months after initiation of ART in individuals with low CD4+ T cell counts. It is thought that the recovery of CD4+ T cells causes a lymphocyte response to the CNS HIV reservoir. The contribution of host genetic factors to development of neuropsychiatric manifestations of HIV infection has not been well studied. However, evidence supports the role of several genetic factors including the E4 allele for apoE in an increased risk of HIV-associated neurocognitive disorders and peripheral neuropathy.

It has also been suggested that the CNS may serve as a relatively sequestered site for a reservoir of latently infected cells that might be a barrier for the eradication of virus by ART (see "The HIV Reservoir: Obstacles to the Eradication of Virus," above).

■ PATHOGENESIS OF KAPOSI'S SARCOMA

There are at least four distinct epidemiologic forms of KS: (1) the classic form that occurs in older men of predominantly Mediterranean or eastern European Jewish backgrounds with no recognized contributing factors; (2) the equatorial African form that occurs in all ages, also without any recognized precipitating factors; (3) the form associated with organ transplantation and its attendant iatrogenic immunosuppressed state; and (4) the form associated with HIV-1 infection. In the latter two forms, KS is an opportunistic disease; in people with HIV, unlike typical opportunistic infections, its occurrence is not strictly related to the level of depression of CD4+ T cell counts. The pathogenesis of KS is complex; fundamentally, it is an angioproliferative disease that is not a true neoplastic sarcoma, at least not in its early stages. It is a manifestation of excessive proliferation of spindle cells that are believed to be of vascular origin and have features in common with endothelial and smooth-muscle cells. In HIV disease the development of KS is dependent on the interplay of a variety of factors including HIV-1 itself, human herpes virus 8 (HHV-8), immune activation, and cytokine secretion. Numerous epidemiologic and virologic studies have clearly linked HHV-8, which is also referred to as *Kaposi's sarcoma–associated herpesvirus* (KSHV), to KS not only in people with HIV but also in individuals with the other forms of KS. HHV-8 is a γ-herpesvirus related to EBV and *herpesvirus saimiri*. It encodes a homologue to human IL-6 and, in addition to KS, has been implicated in the pathogenesis of body cavity lymphoma, multiple myeloma, and monoclonal gammopathy of undetermined significance. Sequences of HHV-8 are found universally in the lesions of KS, and patients with KS are virtually all seropositive for HHV-8. HHV-8 DNA sequences can be found in the B cells of 30–50% of patients with KS and 7% of patients with AIDS without clinically apparent KS.

Between 1% and 2% of eligible blood donors are positive for antibodies to HHV-8, while the prevalence of HHV-8 seropositivity in men with HIV is 30–35%. The prevalence of HHV-8 seropositivity in women with HIV is ~4%. This finding is reflective of the lower incidence of KS in women. It has been debated whether HHV-8 is the transforming agent in KS; the bulk of the cells in the tumor lesions of KS are not neoplastic cells. However, it has been demonstrated that endothelial cells can be transformed in vitro by HHV-8. In this regard, HHV-8 possesses genes, including homologues of the IL-8 receptor, Bcl-2, and cyclin D, which can potentially transform the host cell. Despite the complexity of the pathogenic events associated with the development of KS in people with HIV, HHV-8 is the etiologic agent of this disease. The initiation and/or propagation of KS requires an activated state and is mediated, at least in part, by cytokines. Multiple factors, including TNF-α, IL-1β, IL-6, granulocyte-macrophage colony-stimulating factor (GM-CSF), basic fibroblast growth factor, and oncostatin M, function in an autocrine and paracrine manner to sustain the growth and chemotaxis of the KS spindle cells. In this regard, KSHV-derived IL-6 has been demonstrated to induce proliferation of lymphoma cells and to inhibit the cytostatic effects of IFN-α on KSHV-infected lymphoma cells.

IMMUNE RESPONSE TO HIV

As detailed above and below, following the initial burst of viremia during primary infection, people with HIV mount robust immune responses that in most cases substantially curtail the levels of plasma viremia and likely contribute to delaying the ultimate development of clinically apparent disease for a median of 10 years in untreated individuals. This immune response contains elements of both humoral and cell-mediated immunity involving both adaptive and innate immune responses **(Table 208-7; Fig. 208-28)**. It is directed against multiple antigenic determinants of the HIV virion as well as against viral proteins expressed on the surface of infected cells. Ironically, those CD4+ T cells with T cell receptors specific for HIV are theoretically those CD4+ T cells most likely to be activated—and thus to serve as early targets for productive HIV infection and the cell death or dysfunction associated with infection. Thus, an early consequence of HIV infection is interference with and decrease of the helper T cell population needed to generate an effective immune response.

Although a great deal of investigation has been directed toward delineating and better understanding the components of this immune response, it remains unclear which immunologic effector mechanisms are most important in delaying progression of infection and which, if any, play a role in the pathogenesis of HIV disease. This lack of knowledge has also hampered the ability to develop an effective vaccine for HIV disease.

TABLE 208-7 Elements of the Immune Response to HIV

Humoral immunity
Neutralizing antibodies
Type specific
Group specific
Broadly neutralizing
Antibodies participating in antibody-dependent cellular cytotoxicity (ADCC)
Protective
Pathogenic (bystander killing)
Enhancing antibodies
Complement
Cell-mediated immunity
Helper CD4+ T lymphocytes
Class I MHC–restricted cytotoxic CD8+ T lymphocytes
CD8+ T cell–mediated inhibition (noncytolytic)
ADCC
Natural killer cells

Abbreviation: MHC, major histocompatibility complex.

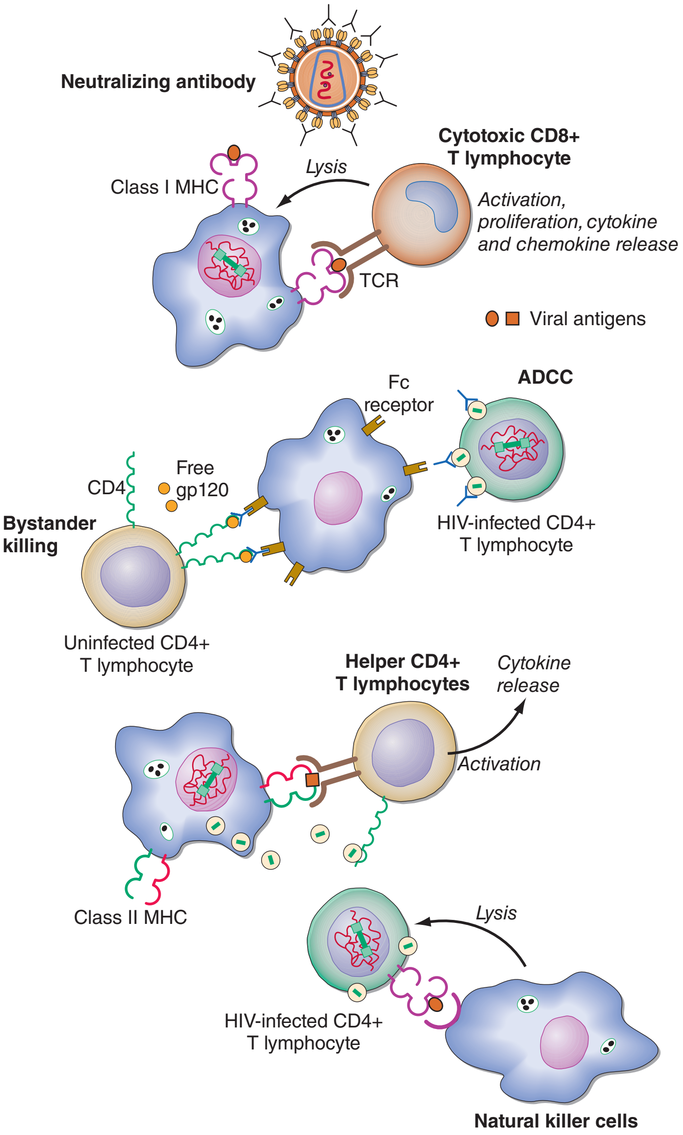
Neutralizing antibody
Cytotoxic CD8+ T lymphocyte
Lysis
Class I MHC
Activation, proliferation, cytokine and chemokine release
TCR
● ■ Viral antigens
ADCC
Fc receptor
Free gp120
CD4
Bystander killing
HIV-infected CD4+ T lymphocyte
Uninfected CD4+ T lymphocyte
Helper CD4+ T lymphocytes
Cytokine release
Activation
Class II MHC
Lysis
HIV-infected CD4+ T lymphocyte
Natural killer cells

FIGURE 208-28 Schematic representation of the different immunologic effector mechanisms thought to be active in the setting of HIV infection. Detailed descriptions are given in the text. ADCC, antibody-dependent cellular cytotoxicity; MHC, major histocompatibility complex; TCR, T cell receptor.

HUMORAL IMMUNE RESPONSE

Antibodies to HIV usually appear within 3–6 weeks and almost invariably within 12 weeks of primary infection **(Fig. 208-29)**; rare exceptions are in individuals who have defects in the ability to produce HIV-specific antibodies. Along with plasma levels of HIV RNA and p24 antigen, detection of these antibodies forms the basis of most diagnostic screening tests for HIV infection. The appearance of HIV-binding antibodies detected by ELISA and Western blot assays occurs prior to the appearance of neutralizing antibodies; the latter generally appear following the initial decreases in plasma viremia and are more closely related to the appearance of HIV-specific CD8+ T lymphocytes. The first antibodies detected are those directed against the immunodominant region of the envelope gp41, followed by the appearance of antibodies to the structural or gag protein p24 and the gag precursor p55. Antibodies to p24 gag are followed by the appearance of antibodies to the outer envelope glycoprotein (gp120), the gag protein p17, and the products of the *pol* gene (p31 and p66). In addition, one may see antibodies to the low-molecular-weight regulatory proteins encoded by the HIV genes *vpr*, *vpu*, *vif*, *rev*, *tat*, and *nef*. On rare occasions, levels of HIV-specific antibodies may decline during treatment of acute HIV infection; otherwise they remain persistently positive.

While antibodies to multiple antigens of HIV are produced, the precise functional significance of these different antibodies is unclear. The only viral proteins that elicit neutralizing antibodies are the envelope proteins gp120 and gp41. Antibodies directed toward the envelope proteins of HIV have been characterized both as being protective and as possibly contributing to the pathogenesis of HIV disease. Among the protective antibodies are those that function to neutralize HIV directly and prevent the spread of infection to additional cells, as well as those that participate in ADCC. The first neutralizing antibodies are directed against the autologous infecting virus and appear after approximately 12 to 24 weeks of infection. Due to its high rate of mutation the virus is usually able to quickly escape these (and subsequent) neutralizing antibodies. One important mechanism of immune escape is the addition of N-linked glycosylation sites, forming a glycan shield that interferes with envelope recognition by these initial antibodies.

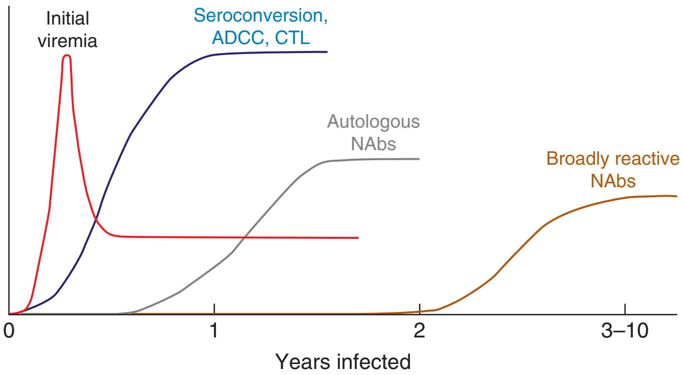

FIGURE 208-29 Relationship between initial HIV viremia and the development of antibodies to HIV. Within 3 to 6 weeks of initial HIV infection, nonneutralizing antibodies to HIV appear. These antibodies are capable of mediating antibody-dependent cellular cytotoxicity (ADCC). The decline in plasma viremia generally correlates with the appearance of cytotoxic T lymphocytes (CTL). After approximately 3 months, autologous neutralizing antibodies (NAbs) capable of neutralizing prior circulating strains of HIV appear. After 2 or more years, broadly reactive NAbs appear. *(Reproduced with permission from of Annual Reviews, Inc. from The Role of Antibodies in HIV Vaccines, JR Mascola and DC Monteori. 28:413, 2010.)*

A number of broad and potent HIV-neutralizing envelope-specific antibodies have been isolated from people with HIV in studies designed to better understand the host response to HIV infection. Approximately 20% of patients develop antibodies capable of neutralizing highly diverse strains. These usually appear 2 or more years following infection in the face of continual viremia. These studies have revealed at least five major sites within the HIV envelope trimer that are able to elicit broadly neutralizing antibodies. These sites include antibodies directed toward the CD4 binding site (CD4bs) of gp120, those binding glycan-dependent epitopes in the V1/V2 region of gp120, those near the base of the V3 region of gp120, those binding to the gp120/gp41 bridge, and those binding to the membrane-proximal region of gp41 **(Fig. 208-30)**. Several of these antibodies contain unique features including high levels of somatic hypermutation, selective germ-line gene usage (especially for CD4bs antibodies), and long

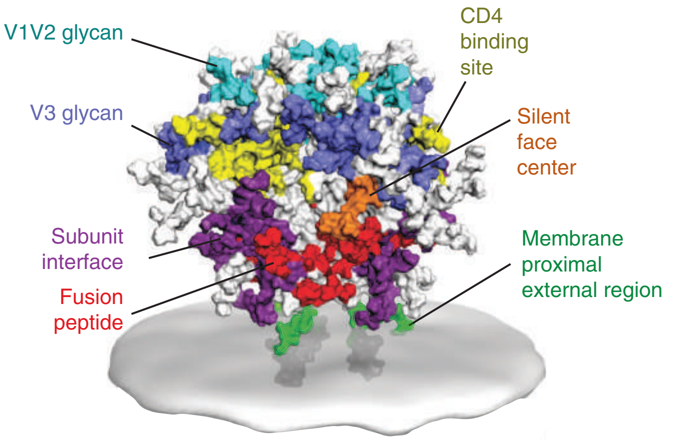

FIGURE 208-30 Known targets of broadly neutralizing antibodies against HIV-1. *(Courtesy of J Stuckey, GY Chuang.)*

heavy chain complementary determining regions (especially CDRH3). Of note, while these antibodies are broadly neutralizing in vitro, their precise in vivo significance is unclear and the patients from whom they were derived demonstrate evidence of ongoing viral replication unless treated with ART.

The other major class of protective antibodies are those that participate in ADCC, a form of cell-mediated immunity **(Chap. 360)** in which NK cells that bear Fc receptors bind specific anti-HIV antibodies that then juxtapose the NK cells to and destroy cells expressing HIV antigens. The levels of anti-envelope antibodies capable of mediating ADCC are highest in the earlier stages of HIV infection. Antibodies to both gp120 and gp41 have been shown to participate in ADCC-mediated killing of HIV-infected cells. In vitro, IL-2 can augment ADCC-mediated killing.

In addition to playing a role in host defense, HIV-specific antibodies have also been implicated in disease pathogenesis. Antibodies directed to gp41, when present in low titer, have been shown in vitro to be capable of facilitating infection of cells through an Fc receptor–mediated mechanism known as *antibody enhancement*. Thus, the same regions of the envelope protein of HIV that give rise to antibodies capable of mediating ADCC can also elicit the production of antibodies that can facilitate infection of cells in vitro. In addition, it has been postulated that anti-gp120 antibodies that participate in the ADCC killing of HIV-infected cells might also kill uninfected CD4+ T cells if the uninfected cells had bound free gp120, a phenomenon referred to as *bystander killing*.

One of the most primitive components of the humoral immune system is the complement system **(Chap. 360)**. This element of innate immunity consists of ~30 proteins that are found circulating in blood or associated with cell membranes. While HIV alone is capable of directly activating the complement cascade, the resulting lysis is weak due to the presence of host cell regulatory proteins captured in the virion envelope during budding. It is possible that complement-opsonized HIV virions have increased infectivity in a manner analogous to antibody-mediated enhancement.

■ CELLULAR IMMUNE RESPONSE

T cell–mediated immunity plays a major role in host defense against most viral infections **(Chap. 360)** and is thought to be an important component of the host immune response to HIV. T cell immunity can be divided into two major categories: that mediated by *helper/inducer CD4+ T cells* and that mediated by *cytotoxic/immunoregulatory CD8+ T cells*.

HIV-specific CD4+ T cells can be detected in the majority of people with HIV via flow cytometry to measure intracellular cytokine production in response to MHC class II tetramers pulsed with HIV peptides or through lymphocyte proliferation assays utilizing HIV antigens such as p24. These cells likely play a critical role in the orchestration of the immune response to HIV by providing help to HIV-specific B cells and CD8+ T cells. They may also be capable of directly killing HIV-infected cells. HIV-specific CD4+ T cells may be preferential targets of HIV infection by HIV-infected antigen-presenting cells during the generation of an immune response to HIV (Fig. 208-28). However, they also are likely to undergo clonal expansions in response to HIV antigens and thus survive as a population of cells despite the virus. No clear correlations exist between levels of HIV-specific CD4+ T lymphocytes and plasma HIV RNA levels; however, in the setting of high viral loads, CD4+ T cell responses to HIV antigens appear to shift from one of proliferation and IL-2 production to one of IFN-γ production. Thus, while a reverse correlation exists between the level of p24-specific proliferation and levels of plasma HIV viremia, the nature of the causal relationship between these parameters is unclear.

MHC class I–restricted, HIV-specific CD8+ T cells have been identified in the peripheral blood of patients with HIV-1 infection. These cells include CTLs that produce perforins and granzyme, and T cells that can be induced by HIV antigens to express an array of cytokines such as IFN-γ, IL-2, MIP-1β, and TNF-α. Multiple HIV antigens, including Gag, Env, Pol, Tat, Rev, and Nef, can elicit CD8+ T cell responses. CTLs have been identified in the peripheral blood of patients within weeks of HIV infection and prior to the appearance of plasma virus. The selective pressure they exert on the evolution of the population of circulating viruses reflects their potential role in control of HIV infection. These CD8+ T lymphocytes, through their HIV-specific antigen receptors, bind to and cause the lytic destruction of target cells bearing autologous MHC class I molecules presenting HIV antigens. Two types of CTL activity can be demonstrated in the peripheral blood or lymph node mononuclear cells of people with HIV. The first type directly lyses appropriate target cells in culture without prior in vitro stimulation (*spontaneous CTL activity*). The other type of CTL activity reflects the *precursor frequency of CTLs* (CTLp); this type of CTL activity can be demonstrated by stimulation of CD8+ T cells in vitro with a mitogen such as phytohemagglutinin or anti-CD3 antibody to expand the population prior to detection.

In addition to CTLs, CD8+ T cells capable of being induced by HIV antigens to express cytokines such as IFN-γ also appear in the setting of HIV-1 infection. It is not clear whether these are the same or different effector pools compared with those cells mediating cytotoxicity; in addition, the relative roles of each in host defense against HIV are not fully understood. It does appear that these CD8+ T cells are driven to in vivo expansion by HIV antigen. There is a direct correlation between levels of CD8+ T cells capable of producing IFN-γ in response to HIV antigens and plasma levels of HIV-1 RNA. Thus, while these cells are clearly induced by HIV-1 infection, in most instances they are not able to effectively control infection. One exception may be a subset of patients who control viral replication in the absence of antiretroviral drugs and are referred to as *elite nonprogressors* (see "Long-Term Survivors, Long-Term Nonprogressors, and Elite Controllers," above). The peripheral blood of these patients contains a population of CD8+ T cells that undergo substantial in vitro proliferation in response to HIV antigens and express perforin and granzyme.

At least three other forms of cell-mediated immunity to HIV have been described: non-cytolytic CD8+ T cell–mediated suppression of HIV replication, ADCC, and NK cell activity. *Non-cytolytic CD8+ T cell–mediated suppression of HIV replication* refers to the ability of CD8+ T cells from a patient with HIV to inhibit the replication of HIV in tissue culture without killing infected targets. There is no requirement for HLA compatibility between the CD8+ T cells and the HIV-infected cells. This effector mechanism is thus nonspecific and appears to be mediated by soluble factor(s) including the CC-chemokines RANTES (CCL5), MIP-1α (CCL3), and MIP-1β (CCL4). These CC-chemokines are potent suppressors of HIV replication and operate at least in part via blockade of the HIV co-receptor (*CCR5*) for R5 (macrophage-tropic) strains of HIV-1 (see above). *ADCC*, as described above in relation to humoral immunity, involves the killing of HIV-expressing cells by NK cells armed with specific antibodies directed against HIV antigens. Finally, *NK cells* alone have been shown to be capable of killing HIV-infected target cells in tissue culture. This primitive cytotoxic mechanism of host defense is directed toward nonspecific surveillance for neoplastic transformation and viral infection through recognition of altered class I MHC molecules.

DIAGNOSIS AND LABORATORY MONITORING OF HIV INFECTION

The establishment of HIV as the causative agent of AIDS and related syndromes early in 1984 was followed by the rapid development of sensitive screening tests for HIV infection. By March 1985, blood donors in the United States were routinely screened for antibodies to HIV. In 1996, blood banks in the United States added the p24 antigen capture assay to the screening process to help identify the rare, infected individuals who were donating blood in the time (up to 3 months) between infection and the development of antibodies. In 2002, the ability to detect early infection with HIV was further enhanced by the licensure of nucleic acid testing (NAT) as a routine part of blood donor screening. These refinements decreased the interval between infection and detection (window period) from 22 days for antibody testing to 16 days with p24 antigen testing and subsequently to 12 days with NAT. The development of sensitive assays for monitoring levels of plasma viremia ushered in a new era of being able to monitor the progression of HIV disease more closely. Utilization of these tests, coupled with the measurement of levels of CD4+ T lymphocytes in peripheral blood, is an important component of the management of persons with HIV infection.

DIAGNOSIS OF HIV INFECTION

The CDC has recommended that screening for HIV infection be performed as a matter of routine health care. The diagnosis of HIV infection depends on the demonstration of antibodies to HIV and/or the direct detection of HIV or one of its components. As noted above, antibodies to HIV generally appear in the circulation 3–12 weeks following infection. In addition to laboratory-based screening tests, several home tests are available.

The standard blood screening tests for HIV infection are based on the detection of antibodies to HIV and/or the p24 antigen (see below) of HIV. A common laboratory-based platform is the ELISA, also referred to as an *enzyme immunoassay* (EIA). This solid-phase assay is an extremely good screening test with a sensitivity of >99.5%. Most diagnostic laboratories use commercial kits that contain antigens from both HIV-1 and HIV-2 and thus can detect antibodies to either. These kits use both natural and recombinant antigens and are continuously updated to increase their sensitivity to newly discovered species, such as group O viruses (Fig. 208-1). The fourth-generation EIA tests combine detection of antibodies to HIV-1 or HIV-2 with detection of the p24 antigen of HIV. EIA tests are generally scored as positive (highly reactive), negative (nonreactive), or indeterminate (partially reactive). While the EIA is an extremely sensitive test, it is not optimal with regard to specificity. This is particularly true in studies of low-risk individuals, such as volunteer blood donors. In this latter population, as few as 10% of EIA-positive individuals are subsequently confirmed to have HIV infection. Among the factors associated with false-positive EIA tests are antibodies to class II antigens (such as may be seen following pregnancy, blood transfusion, or transplantation), autoantibodies, hepatic disease, recent influenza vaccination, acute viral infections, and administration of an HIV vaccine. For these reasons, anyone suspected of having HIV infection based on a positive or inconclusive fourth-generation EIA result should have the result confirmed with a more specific assay such as an HIV-1– or HIV-2–specific antibody immunoassay or a plasma HIV RNA level. One can estimate whether an individual has a recent infection with HIV-1 by comparing the results on a standard EIA test that will score positive for all infected individuals with the results on an assay modified to be less sensitive ("detuned assay") that will score positive for individuals with established HIV infection and negative for individuals with recent infection. In rare instances, an individual with HIV treated early in the course of infection may revert to a negative EIA. This does *not* indicate clearing of infection; rather, it signifies levels of ongoing exposure to virus or viral proteins insufficient to maintain a measurable antibody response. When these individuals have discontinued therapy, viruses and antibodies have reappeared.

CDC recommendations indicate that a positive fourth-generation assay confirmed by a second HIV-1– or HIV-2–specific immunoassay or a plasma HIV RNA level is adequate for diagnosis. The Western blot, which had previously been used for a confirmatory test, is no longer used for this purpose.

A guideline for the use of these serologic tests in attempting to make a diagnosis of HIV infection is depicted in **Fig. 208-31**. In patients in whom HIV infection is suspected, the appropriate initial test is a fourth-generation HIV-1/2 antigen antibody immunoassay. If the result is negative, unless there is strong reason to suspect early HIV infection (as in a patient exposed within the previous 3 months), the diagnosis is ruled out and retesting should be performed only as clinically indicated. If the immunoassay is indeterminate or positive, the test should be repeated. If the repeat is negative on two occasions, one can assume that the initial positive reading was due to a technical error in the performance of the assay and that the patient is negative. If the repeat is indeterminate or positive, one should proceed to an HIV-1/HIV-2 antibody differentiation immunoassay such as the Bio-Rad Geenius. If testing is positive for HIV-1 and/or HIV-2 one may make a diagnosis of HIV-1 and/or HIV-2 infection. If the HIV-1/HIV-2 antibody testing is negative or indeterminate one should proceed to HIV-1 RNA testing with a nucleic acid test (NAT; see below). If the NAT is positive, in the presence of a negative antibody test, one can make a diagnosis of acute HIV-1 infection. If the NAT test is negative for HIV-1 one should consider additional NAT testing for HIV-2. One can conclude a false-positive fourth-generation test in the setting of repeated negative or indeterminate HIV-2/HIV-2 antibody tests in the setting of negative NAT tests.

In addition to these standard laboratory-based assays for detecting antibodies to HIV, a series of point-of-care tests can provide results in 1 to 60 minutes. While the sensitivity and specificity of these tests are generally quite high, it is generally recommended that any positive results be confirmed with standard laboratory testing. Currently in the United States there is one FDA-approved rapid test kit available for use at home (OraQuick). It provides results in approximately 20 minutes. There are several other tests for which the sample is obtained at home and mailed to the lab. A positive result with any of these tests should be followed with confirmatory laboratory testing by a health care professional.

A variety of laboratory tests are available for the direct detection of HIV or its components **(Table 208-8)**. These tests may be of considerable help in making a diagnosis of HIV infection when the antibody determination assays are indeterminate. In addition, the tests detecting levels of HIV RNA can be used to determine prognosis and to assess the response to antiretroviral therapies. The simplest, least expensive, and most rarely used of the direct detection tests is the *p24 antigen capture assay*. This is an EIA-type assay in which the solid phase consists of antibodies to the p24 antigen of HIV. It detects the viral protein p24 in the blood of people with HIV where it exists either as free antigen or complexed to anti-p24 antibodies. It is currently part of the fourth-generation HIV-1/2 antigen antibody immunoassay test recommended for initial screening. Overall, ~30% of individuals with untreated HIV infection have detectable levels of free p24 antigen. This increases to ~50% when samples are treated with a weak acid to dissociate antigen–antibody complexes. Throughout the course of HIV infection, an equilibrium exists between p24 antigen and anti-p24 antibodies. During the first few weeks of infection, before an immune response develops, there is a brisk rise in p24 antigen levels. After the development of anti-p24 antibodies, these levels decline. Late in the course of infection, when

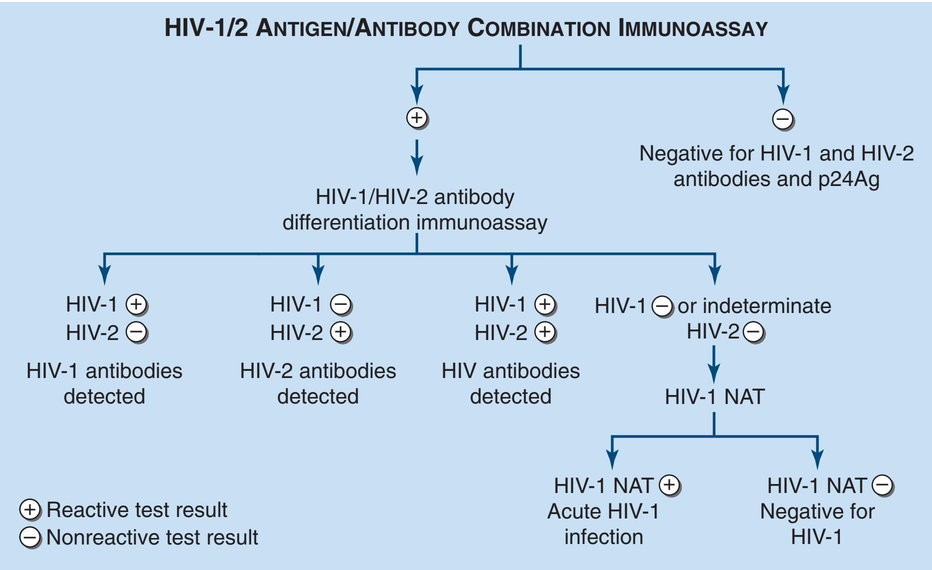
HIV-1/2 Antigen/Antibody Combination Immunoassay
⊕
⊖
Negative for HIV-1 and HIV-2 antibodies and p24Ag
HIV-1/HIV-2 antibody differentiation immunoassay
HIV-1 ⊕
HIV-2 ⊖
HIV-1 antibodies detected
HIV-1 ⊖
HIV-2 ⊕
HIV-2 antibodies detected
HIV-1 ⊕
HIV-2 ⊕
HIV antibodies detected
HIV-1 ⊖ or indeterminate
HIV-2 ⊖
HIV-1 NAT
HIV-1 NAT ⊕
Acute HIV-1 infection
HIV-1 NAT ⊖
Negative for HIV-1
⊕ Reactive test result
⊖ Nonreactive test result

FIGURE 208-31 CDC algorithm for making a diagnosis of HIV infection using tests for antibody, antigen, and RNA. NAT, nucleic acid test. *(Adapted from stacks.cdc.gov/view/cdc/23446.)*

TABLE 208-8 Characteristics of Tests for Direct Detection of HIV

TEST	TECHNIQUE	SENSITIVITY[a]	COST/TEST[b]
Immune complex–dissociated p24 antigen capture assay	Measurement of levels of HIV-1 core (p24) protein in an EIA-based format following dissociation of antigen–antibody complexes by weak acid treatment	Positive in 50% of patients; detects down to 15 pg/mL of p24 protein	\$1–2
HIV RNA by PCR	Target amplification of HIV-1 RNA via reverse transcription followed by PCR	Reliable to 20 copies/mL of HIV RNA	\$75–150
HIV RNA by bDNA	Measurement of levels of particle-associated HIV RNA in a nucleic acid capture assay employing signal amplification	Reliable to 50 copies/mL of HIV RNA	\$75–150
HIV RNA by TMA	Target amplification of HIV-1 RNA via reverse transcription followed by T7 RNA polymerase	Reliable to 100 copies/mL of HIV RNA	\$225
HIV RNA by NASBA	Isothermal nucleic acid amplification with internal controls	Reliable to 80 copies/mL of HIV RNA	\$75–150

[a]Sensitivity figures refer to those approved by the U.S. Food and Drug Administration. [b]Prices may be lower in large-volume settings.

Abbreviations: bDNA, branched DNA; cDNA, complementary DNA; EIA, enzyme immunoassay; NASBA, nucleic acid sequence–based amplification; PCR, polymerase chain reaction; TMA, transcription-mediated amplification.

circulating levels of virus are high, p24 antigen levels also increase, particularly when detected by techniques involving dissociation of antigen-antibody complexes. The p24 antigen capture assay has its greatest use as a screening test for HIV infection in patients suspected of having the acute HIV syndrome (see below), as high levels of p24 antigen are present prior to the development of antibodies. Its use as a stand-alone test for routine blood donor screening for HIV infection has been replaced by the use of "fourth-generation" assays that combine antigen and antibody testing or NAT.

The ability to measure and monitor levels of HIV RNA in the plasma of patients with HIV infection has been of extraordinary value in furthering our understanding of the pathogenesis of HIV infection, in monitoring the response to ART, and in providing a diagnostic tool in settings where measurements of anti-HIV antibodies may be misleading, such as in acute infection and neonatal infection. In addition to the commercially available tests for measuring HIV RNA, *DNA PCR* assays also are employed by research laboratories for making a diagnosis of HIV infection by amplifying HIV proviral DNA from peripheral blood mononuclear cells. The commercially available RNA detection tests have a sensitivity of 20–50 copies of HIV RNA per milliliter of plasma. Research laboratory–based RNA assays can detect as few as one HIV RNA copy per milliliter, while the DNA PCR tests can detect proviral DNA at a frequency of one copy per 10,000–100,000 cells. Thus, these tests are extremely sensitive. One frequent consequence of a high degree of sensitivity is some loss of specificity, and false-positive results have been reported with each of these techniques. For this reason, a positive EIA with a positive HIV RNA assay can be considered the "gold standard" for a diagnosis of HIV infection, and the interpretation of other test results must be done with this in mind.

In the RT-PCR technique, following DNAse treatment, a cDNA copy is made of the HIV RNA species present in plasma typically using tRNALys3 as a primary primer. This is the same host cell primer used during the reverse transcription stage of the viral life cycle **(Fig. 208-32)**. Because HIV is an RNA virus, this will result in the production of DNA copies of the HIV genome in amounts proportional to the amount of HIV RNA present in plasma. This cDNA is then amplified and characterized using standard PCR techniques. In addition to being diagnostic and prognostic tools, RT-PCR and DNA-PCR also are useful for amplifying defined areas of the HIV genome for sequence analysis and have become an important technique for studies of sequence diversity and microbial resistance to antiretroviral agents. In patients with a positive or indeterminate EIA test and an indeterminate Western blot, and in patients in whom serologic testing may be unreliable (such as patients with hypogammaglobulinemia or advanced HIV disease), these tests for quantitating HIV RNA in plasma or detecting proviral DNA in peripheral blood mononuclear cells are valuable tools for making a diagnosis of HIV infection; however, they should be used for diagnosis only when standard serologic testing has failed to provide a definitive result.

■ LABORATORY MONITORING OF PATIENTS WITH HIV INFECTION

The integration of clinical and laboratory data is essential to optimal management of patients with HIV infection. The close relationship between clinical manifestations of HIV infection and CD4+ T cell count has made measurement of CD4+ T cell numbers a routine part of the initial evaluation of people with HIV. The discovery of HIV as the cause of AIDS led to the development of sensitive tests that allow one to monitor the levels of HIV in the blood. Determinations of peripheral blood CD4+ T cell counts and measurements of the plasma levels of HIV RNA provide a powerful set of tools for determining prognosis and monitoring response to therapy.

CD4+ T Cell Counts The CD4+ T cell count is the laboratory test generally accepted as the best indicator of the immediate state of immunologic competence of the patient with HIV infection. This measurement has been shown to correlate very well with the level of immunologic competence. Patients with CD4+ T cell counts <200/μL are at high risk of disease from *P. jirovecii*, while patients with CD4+ T cell counts <50/μL are also at high risk of disease from CMV, mycobacteria of the *M. avium* complex (MAC), and/or *T. gondii* (Fig. 208-32). Once the CD4+ T cell count is <200/μL, patients should be placed on a regimen for *P. jirovecii* prophylaxis. Once the count is <50/μL, primary prophylaxis for MAC infection is indicated unless the patient is immediately started on ART. As with any laboratory measurement, one may wish to obtain two determinations prior to any significant changes in patient management based on CD4+ T cell count alone. Patients with HIV infection should have CD4+ T cell measurements performed at the time of diagnosis and every 3–6 months for the first 2 years thereafter. More frequent measurements should be made if a declining trend is noted. For patients who have been on ART for at least 2 years with HIV RNA levels persistently <50 copies/mL and CD4 counts 300–500/μL, monitoring may be decreased to every year. For those with CD4 counts >500/μL, the monitoring of the CD4 count is felt by many to be optional. There are a handful of clinical situations in which the CD4+ T cell count may be misleading. Patients with HTLV-1/HIV co-infection may have elevated CD4+ T cell counts that do not accurately reflect their degree of immune competence. In patients with hypersplenism or those who have undergone splenectomy, and in patients receiving medications that suppress the bone marrow such as IFN-α, the CD4+ T cell percentage may be a more reliable indication of immune function than the CD4+ T cell count. A CD4+ T cell percentage of 15 is comparable to a CD4+ T cell count of 200/μL.

HIV RNA Determinations Facilitated by highly sensitive techniques for the precise quantitation of small amounts of nucleic acids, the measurement of serum or plasma levels of HIV RNA has become an essential component in the monitoring of patients with HIV infection. As discussed in "Diagnosis of HIV Infection," above, the most used technique is the RT-PCR assay. This assay generates data in the form of number of copies of HIV RNA per milliliter of serum or plasma, and commercial assays can reliably detect as few as 20 copies of HIV RNA per milliliter of plasma. Research-based assays can detect down to one copy per milliliter. While it is common practice to describe levels of HIV RNA below these cut-offs as "undetectable," this is a term that should be avoided as it is imprecise and leaves the false impression that the level of virus is 0. By utilizing more sensitive,

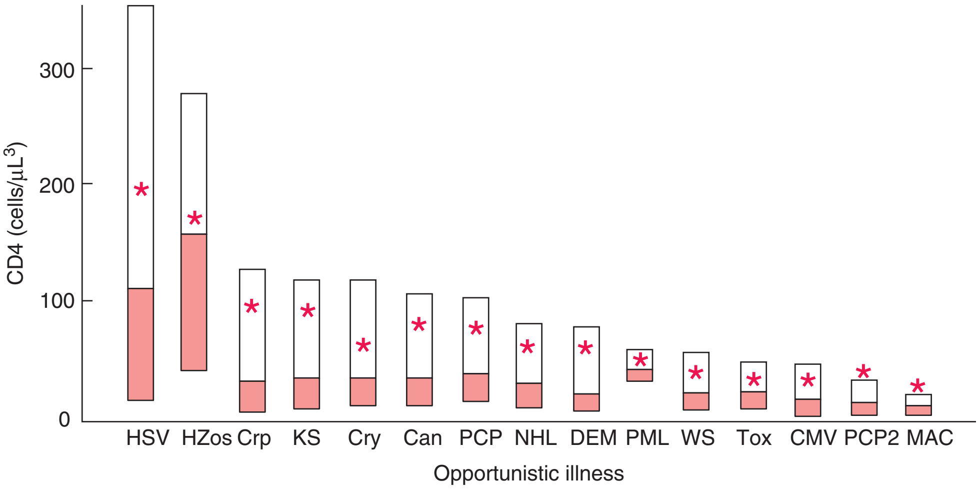

FIGURE 208-32 Relationship between CD4+ T cell counts and the development of opportunistic diseases. Boxplot of the median (*line inside the box*), first quartile (*bottom of the box*), third quartile (*top of the box*), and mean (*asterisk*) CD4+ lymphocyte count at the time of the development of opportunistic disease. Can, candidal esophagitis; CMV, cytomegalovirus infection; Crp, cryptosporidiosis; Cry, cryptococcal meningitis; DEM, AIDS dementia complex; HSV, herpes simplex virus infection; HZos, herpes zoster; KS, Kaposi's sarcoma; MAC, *Mycobacterium avium* complex bacteremia; NHL, non-Hodgkin's lymphoma; PCP, primary *Pneumocystis jirovecii* pneumonia; PCP2, secondary *P. jirovecii* pneumonia; PML, progressive multifocal leukoencephalopathy; Tox, *Toxoplasma gondii* encephalitis; WS, wasting syndrome. *(From Annals of Internal Medicine, RD Moore, RE Chaisson: Natural History of Opportunistic Disease in an HIV-Infected Urban Clinical Cohort. 124(7):633-642, 1996. Copyright © 1996 American College of Physicians. All Rights Reserved. Reprinted with the permission of American College of Physicians, Inc.)*

nested PCR techniques and by studying tissue levels of virus as well as plasma levels, HIV RNA can be detected in virtually every patient with HIV infection. There are notable exceptions to this that involve patients who underwent cytoreductive therapy followed by transplant from *CCR5Δ32* donors.

Measurements of changes in HIV RNA levels over time have been of great value in delineating the relationship between levels of virus and rates of disease progression (Fig. 208-22), the rates of viral turnover, the relationship between immune system activation and viral replication, and the time to development of drug resistance. HIV RNA measurements are greatly influenced by the state of activation of the immune system and may fluctuate greatly in the setting of secondary infections or immunization. For these reasons, decisions based on HIV RNA levels should never be made on a single determination. Measurements of plasma HIV RNA levels should be made at the time of HIV diagnosis. At the time of diagnosis, ART should be initiated, and HIV RNA levels monitored approximately every 4 weeks until the effectiveness of the therapeutic regimen is determined by the development of a new steady-state level of HIV RNA. In most instances of effective antiretroviral therapy, the plasma level of HIV RNA will drop to <50 copies/mL within 6 months of the initiation of treatment. During therapy, levels of HIV RNA should be monitored every 3–6 months to evaluate the continuing effectiveness of therapy.

HIV Resistance Testing The availability of multiple antiretroviral drugs as treatment options has generated a great deal of interest in the potential for measuring the sensitivity of an individual's HIV viral quasispecies to different antiretroviral agents. HIV resistance testing can be done through either genotypic or phenotypic measurements. In the genotypic assays, sequence analyses of the HIV genomes obtained from patients are compared with sequences of viruses with known antiretroviral resistance profiles. In the phenotypic assays, the in vivo growth of patient-derived viral isolates or genetically constructed pseudoviruses is compared with the growth of reference strains of the virus in the presence or absence of different antiretroviral drugs. These tests are quite good at identifying those antiretroviral agents that have been utilized in the past and suggesting agents that may be of future value in a given patient. Resistance testing is recommended at the time of initial diagnosis and, if therapy is not initiated at that time, at the time of initiation of ART. Drug resistance testing is also indicated in the setting of virologic failure and should be performed while the patient is still on the failing regimen because of the propensity for the pool of HIV quasispecies to rapidly revert to wild-type in the absence of the selective pressures of ART. In the hands of experts, resistance testing enhances the short-term ability to decrease viral load by ~0.5 log compared with changing drugs merely based on drug history. In addition to the use of resistance testing to help in the selection of new drugs in patients with virologic failure, it may also be of value in selecting an initial regimen for treatment of therapy-naïve individuals. This is particularly true in geographic areas with a high level of background resistance. The patient needs to have an HIV-1 RNA level above 500–1000 copies/mL for an accurate resistance determination. Resistance assays lose their consistency at lower levels of plasma viremia.

Co-Receptor Tropism Assays Following the licensure of maraviroc as the first CCR5 antagonist for the treatment of HIV infection (see below), it became necessary to be able to determine whether a patient's virus was likely to respond to this treatment. Patients tend to have CCR5-tropic virus early in the course of infection, with a trend toward CXCR4 viruses later in disease. The antiretroviral agent maraviroc is effective only against CCR5-tropic viruses. Because the genotypic determinants of cellular tropism are poorly defined, a phenotypic assay is necessary to determine this property of HIV. The Trofile assay (Monogram Biosciences) is available to make this determination. This assay clones the envelope regions of the patient's virus into an indicator virus that is then used to infect target cells expressing either CCR5 or CXCR4 as their co-receptor. The assay takes weeks to perform and is expensive. Another, less costly option is to obtain a genotypic assay of the V3 region of HIV-1 and then employ a computer algorithm to predict viral tropism from the sequence. While this approach is less expensive than the classic phenotypic assay, there are fewer data to validate its predictive value.

Other Tests A variety of other laboratory tests have been studied as potential markers of HIV disease activity. Among these are quantitative culture of replication-competent HIV from plasma, peripheral blood mononuclear cells, or resting memory CD4+ T cells; circulating levels of β_2-microglobulin, soluble IL-2 receptor, IgA, acid-labile endogenous IFN, or TNF-α; and the presence or absence of activation markers such as CD38, HLA-DR, and PD-1 on CD4+ or CD8+ T cells. Nonspecific serologic markers of inflammation and/or coagulation such as IL-6, D-dimer, and sCD14 have been shown to have a high correlation with all-cause mortality **(Table 208-9)**. While these measurements have value as markers of disease activity and help to increase our understanding of the pathogenesis of HIV disease, they do not currently play a major role in the monitoring of patients with HIV infection.

TABLE 208-9 Association between High-Sensitivity CRP, Il-6, and D-Dimer with All-Cause Mortality in Patients with HIV Infection

	UNADJUSTED		ADJUSTED	
MARKER	ODDS RATIO (FOURTH/FIRST)	*P*	ODDS RATIO (FOURTH/FIRST)	*P*
Hs-CRP	2.0	.05	2.8	.03
IL-6	8.3	<.0001	11.8	<.0001
D-dimer	12.4	<.0001	26.5	<.0001

Abbreviations: Hs-CRP, high-sensitivity C-reactive protein; IL-6, interleukin 6.

Source: From LH Kuller et al: PLoS Med 5:e203, 2008.

CLINICAL MANIFESTATIONS

The clinical consequences of HIV infection encompass a spectrum ranging from an acute syndrome associated with primary infection to a prolonged asymptomatic state to advanced disease. It is best to regard HIV disease as beginning at the time of primary infection and progressing through various stages. As mentioned above, active virus replication and progressive immunologic impairment occur throughout the course of HIV infection in most patients. Except for the rare, true, "elite" virus controllers or long-term nonprogressors (see "Long-Term Survivors, Long-Term Nonprogressors, and Elite Controllers," above), HIV disease in untreated patients inexorably progresses even during the clinically latent stage. Since the mid-1990s, ART has had a major impact on preventing and reversing the progression of disease over extended periods of time in a substantial proportion of adequately treated patients. Today, a person diagnosed with HIV infection and treated with ART has a close to normal life expectancy.

■ ACUTE HIV INFECTION

It is estimated that 50–70% of individuals with HIV infection experience an acute clinical syndrome ~3–6 weeks after primary infection **(Fig. 208-33)**. Varying degrees of clinical severity have been reported, and although it has been suggested that symptomatic seroconversion leading to the seeking of medical attention indicates an increased risk for an accelerated course of disease, there does not appear to be a correlation between the level of the initial burst of viremia in acute HIV infection and the subsequent course of disease. The typical clinical findings in the acute HIV syndrome are listed in **Table 208-10**; they occur along with a burst of plasma viremia. It has been reported that several symptoms of the acute HIV syndrome (fever, skin rash, pharyngitis, and myalgia) occur less frequently in those infected by injection drug use compared with those infected by sexual contact. The syndrome is typical of an acute viral syndrome and has been likened to acute infectious mononucleosis. Symptoms usually persist for one to several weeks and gradually subside as an immune response to HIV develops and the levels of plasma viremia decrease. Opportunistic infections have been reported during this stage of infection, reflecting the immunodeficiency that results from reduced numbers of CD4+ T cells and likely also from the dysfunction of CD4+ T cells owing to viral protein and endogenous cytokine-induced perturbations of cells (Table 208-5) associated with the extremely high levels of plasma viremia. The Fiebig staging system has been used to describe the different stages of acute HIV infection, ranging from stage 1 (HIV RNA positive alone) to stage VI (HIV RNA and full Western blot positive). A number of immunologic abnormalities accompany the acute HIV syndrome, including multiphasic perturbations of the numbers of circulating lymphocyte subsets. The numbers of total lymphocytes and T cell subsets (CD4+ and CD8+) are initially reduced. An inversion of the CD4+/CD8+ T cell ratio occurs later because of a rise in the number of CD8+ T cells. In fact, there may be a selective and transient expansion of CD8+ T cell subsets, as determined by T cell receptor analysis (see above). The total circulating CD8+ T cell count may remain elevated or return to normal; however, CD4+ T cell levels usually remain somewhat depressed, although there may be a slight rebound toward normal. Lymphadenopathy occurs in ~70% of individuals with primary HIV infection. Most patients recover spontaneously from this syndrome, and many are left with only a mildly depressed CD4+ T cell count that remains stable for a variable period before beginning its progressive decline; in some individuals, the CD4+ T cell count returns to the normal range. Approximately 10% of patients manifest a fulminant course of immunologic and clinical deterioration after primary infection, even after the disappearance of initial symptoms. In most patients, primary infection with or without the acute syndrome is followed by a prolonged period of clinical latency or smoldering low disease activity.

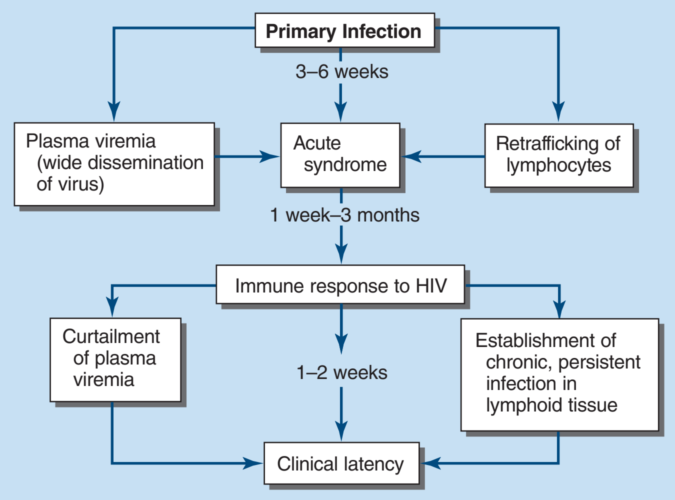
Primary Infection
3–6 weeks
Plasma viremia (wide dissemination of virus)
Acute syndrome
Retrafficking of lymphocytes
1 week–3 months
Immune response to HIV
Curtailment of plasma viremia
1–2 weeks
Establishment of chronic, persistent infection in lymphoid tissue
Clinical latency

FIGURE 208-33 The acute HIV syndrome. See text for detailed description. *(From G Pantaleo, C Graziosi, AS Fauci: The Immunopathogenesis of Human Immunodeficiency Virus Infection. N Engl J Med 328:327, 1993. Copyright © 1993 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

TABLE 208-10 Clinical Findings in the Acute HIV Syndrome

General	Neurologic
Fever	Meningitis
Pharyngitis	Encephalitis
Lymphadenopathy	Peripheral neuropathy
Headache/retroorbital pain	Myelopathy
Arthralgias/myalgias	Dermatologic
Lethargy/malaise	Erythematous maculopapular rash
Anorexia/weight loss	Mucocutaneous ulceration
Nausea/vomiting/diarrhea	

Source: Reproduced with permission from B Tindall, DA Cooper: Primary HIV infection: Host responses and intervention strategies. AIDS 5:1, 1991.

■ THE ASYMPTOMATIC STAGE—CLINICAL LATENCY

Although the length of time from initial infection to the development of clinical disease varies greatly, the median time for untreated patients is ~10 years. As emphasized above, HIV disease with active virus replication is ongoing and progressive during this asymptomatic period. The rate of disease progression is directly correlated with HIV RNA levels. Patients with high levels of HIV RNA in plasma progress to symptomatic disease faster than patients with low levels of HIV RNA (Fig. 208-22). Some patients referred to as *long-term nonprogressors* show little if any decline in CD4+ T cell counts over extended periods of time. These patients generally have extremely low levels of HIV RNA; a subset, referred to as *elite nonprogressors*, exhibits HIV RNA levels <50 copies/mL. Certain other patients remain entirely asymptomatic even though their CD4+ T cell counts show a steady progressive decline to extremely low levels. In these patients, the appearance of an opportunistic disease may be the first manifestation of HIV infection. During the asymptomatic period of HIV infection, the average rate of CD4+ T cell decline is ~50/μL per year in an untreated patient. When the CD4+ T cell count falls to <200/μL, the resulting state of immunodeficiency is severe enough to place the patient at high risk for opportunistic infections and neoplasms and, hence, for clinically apparent disease.

■ SYMPTOMATIC DISEASE

Symptoms of HIV disease can appear at any time during the course of HIV infection. Generally, the spectrum of illnesses that one observes changes as the CD4+ T cell count declines. The more severe and life-threatening complications of HIV infection occur in patients with CD4+ T cell counts <200/μL. A diagnosis of AIDS is made in any individual age 6 years and older with HIV infection and a CD4+ T cell count <200/μL (stage 3, Table 208-2) and in anyone with HIV infection who develops one of the HIV-associated diseases considered to be indicative of a severe defect in cell-mediated immunity (Table 208-1). While the causative agents of the secondary infections are characteristically opportunistic organisms such as *P. jirovecii*, atypical mycobacteria, CMV, and other organisms that do not ordinarily cause disease in the absence of a compromised immune system, they also include several common bacterial and mycobacterial pathogens. Following the widespread use of ART and implementation of guidelines for the prevention of opportunistic infections **(Table 208-11)**, the incidence of these

TABLE 208-11 NIH/CDC/IDSA 2024 Guidelines for the Prevention of Opportunistic Infections in Persons Infected with HIV

PATHOGEN	INDICATIONS	FIRST CHOICE(S)	ALTERNATIVES
Pneumocystis jirovecii	CD4+ T cell count <100–200/μL if HIV RNA level above detection limits *or* CD4+ T cell count <100 /μL regardless of HIV RNA level OR Oropharyngeal candidiasis *or* Prior bout of PCP May stop prophylaxis if CD4+ T cell count >200/μL for ≥3 months	Trimethoprim-sulfamethoxazole (TMP-SMX), 1 DS tablet qd PO *or* TMP-SMX, 1 SS tablet qd PO	Dapsone 50 mg bid PO or 100 mg/d PO *or* Dapsone 50 mg/d PO + Pyrimethamine 50 mg/week PO + Leucovorin 25 mg/week PO *or* (Dapsone 200 mg PO + Pyrimethamine 75 mg PO + Leucovorin 25 mg weekly PO) *or* Aerosolized pentamidine, 300 mg via Respirgard II nebulizer every month *or* Atovaquone 1500 mg/d PO *or* TMP-SMX 1 DS tablet 3?/week PO
Mycobacterium tuberculosis		Important to double check drug–drug interactions with ART regimen	
Isoniazid sensitive	Skin test >5 mm *or* Positive IFN-γ release assay *or* Prior positive test without treatment *or* Close contact with case of active pulmonary TB Same with high probability of exposure to drug-resistant TB	Isoniazid 15 mg/kg PO once weekly (900 mg maximum) plus rifapentine (600 mg for 25.1-32 kg; 750 mg for 32.1-49.9 kg; 900 mg ≥ 50.0 kg Pyridoxine 50 mg PO weekly for 12 weeks *or* Isoniazid 300 mg PO daily + rifampin 600 mg PO daily + pyridoxine 25–50 mg PO daily for 3 months	Isoniazid 300 mg PO daily + pyridoxine 25–50 mg PO daily for 6–9 months Rifampin 600 mg PO daily for 4 months Isoniazid 300 mg PO daily + rifapentine (<35 kg: 300 mg; 35–45 kg: 450 mg; >45 kg: 600 mg) PO daily + pyridoxine 25–50 mg PO daily for 4 weeks (only for persons on an efavirenz-based ART)
Drug resistant	Consult local public health authorities		
Mycobacterium avium complex	CD4+ T cell count <50/μL unless ART immediately initiated—or if not able to achieve viral suppression *or* Prior documented disseminated disease May stop prophylaxis once fully suppressive ART in place	Azithromycin 1200 mg weekly PO or 600 mg twice weekly PO *or* Clarithromycin 500 mg twice daily PO	Rifabutin (dose adjusted based on cART regimen)
Toxoplasma gondii	T. gondii IgG antibody positive and CD4+ T cell count <100/μL *or* Prior toxoplasmic encephalitis and CD4+ T cell count <200/μL May stop prophylaxis if CD4+ T cell count >200/μL and HIV RNA suppressed for ≥3 months	TMP-SMX 1 DS tablet PO daily *or* Sulfadiazine 2000–4000 mg in 2–4 divided doses daily PO + Pyrimethamine 25–50 mg/d PO + Leucovorin 10–25 mg/d PO	TMP-SMX 1 DS 3 times weekly PO *or* TMP-SMX, 1 SS PO daily *or* Dapsone 50 mg/d PO + Pyrimethamine 50 mg weekly PO + Leucovorin 25 mg weekly PO *or* (Dapsone 200 mg PO + Pyrimethamine 75 mg PO + Leucovorin 25 mg PO) weekly *or* Atovaquone 1500 mg PO daily ± (Pyrimethamine 75 mg PO + Leucovorin 10 mg PO) daily *or* (Clindamycin 600 mg q8h PO + Pyrimethamine 25 mg/d PO + Leucovorin 10mg/d PO) daily
Varicella zoster virus	Significant exposure to chickenpox or shingles in a patient with no history of immunization or prior exposure to either	Varicella zoster immune globulin, IM, within 10 d of exposure (800-843-7477)	Acyclovir 800 mg PO 5 times a day for 5–7 days *or* Valacyclovir 1 g PO tid for 5–7 days

(*Continued*)

TABLE 208-11 NIH/CDC/IDSA 2024 Guidelines for the Prevention of Opportunistic Infections in Persons Infected with HIV (Continued)

PATHOGEN	INDICATIONS	FIRST CHOICE(S)	ALTERNATIVES
Cryptococcus neoformans	Prior documented disease—primary prophylaxis not recommended in United States May stop secondary prophylaxis 1 year after initiation of therapy if CD4+ T cell count >100/μL, no evidence of active fungal infection, and HIV RNA levels suppressed for >3 months	Fluconazole 200 mg/d PO	Itraconazole 200 mg/d PO
Histoplasma capsulatum	Prior documented disease or CD4+ T cell count <150 μL and high risk (endemic area or occupational exposure) May stop prophylaxis after 1 year if CD4+ T cell count >150/μL and patient on effective ART for ≥6 months	Itraconazole 200 mg bid PO	Fluconazole 400 mg/d PO
Coccidioides immitis	Prior documented disease *or* positive serology and CD4+ T cell count <250/μL if from a disease endemic area. (For this indication prophylaxis can be stopped if CD4+ T cell count ≥250 for 6 months.)	Fluconazole 400 mg PO daily	
Talaromyces (formerly *Penicillium*) *marneffei*	Prior documented disease Patients with CD4+ T cell counts <100 who live or stay in northern Thailand, Southern China, or Vietnam May stop secondary prophylaxis in patients on ART with CD4+ T cell count >100/μL for ≥6 months	Itraconazole 200 mg/d PO	Fluconazole 400 mg PO once weekly
Salmonella species	Prior recurrent bacteremia—primary prophylaxis not generally recommended	Based on sensitivity of initial infection	
Bartonella	Prior infection May stop after 3–4 months of therapy and with a CD4+ T cell count >200/μL for >3 months and effective ART for at least 6 months	Doxycycline 200 mg/d PO *or* Azithromycin 1200 mg weekly PO *or* Clarithromycin 500 mg bid PO	
Cytomegalovirus	Prior end-organ disease—primary prophylaxis not recommended May stop secondary prophylaxis if CD4+ T cell count >100/μL for 6 months and no evidence of active CMV disease Restart if CD4+ T cells <100/μL	Valganciclovir 900 mg once daily PO	
Immunizations Generally Recommended			
Hepatitis B virus	All susceptible (anti-HBc– and anti-HBs–negative) patients	Hepatitis B vaccine: 3 doses	
Hepatitis A virus	All susceptible (anti-HAV–negative) patients	Hepatitis A vaccine: 2 doses	
Influenza virus	All patients annually	Inactivated trivalent influenza virus vaccine 1 dose yearly	
COVID-19	Annually or per updated CDC recommendations (*www.cdc.gov/covid/vaccines/immunocompromised-people.html*)		
Streptococcus pneumoniae	All patients, preferably before CD4+ T cell count ≤200/μL Reimmunize persons initially immunized at a CD4+ T cell count <100/μL whose CD4+ T cell count then increases to >200/μL	Pneumococcal conjugated vaccine (15 or 20) 0.5 mL IM; if 15-valent is used, follow in 8 weeks or more by pneumococcal polysaccharide vaccine (23)	
Human papillomavirus	All patients 13–26 years of age	HPV vaccine; 3 doses	
Mpox	All patients with HIV who have potential for mpox exposure	Mpox vaccine (MVA-BN); 2 doses	
Meningococcal disease	All patients with HIV infection >18 years	Two doses of quadrivalent meningococcal conjugate vaccine	
Zoster	Age <18 years	Two-dose series of recombinant zoster vaccine (Shingrix) IM 2–6 months apart	

(*Continued*)

TABLE 208-11 NIH/CDC/IDSA 2024 Guidelines for the Prevention of Opportunistic Infections in Persons Infected with HIV (Continued)

PATHOGEN	INDICATIONS	FIRST CHOICE(S)	ALTERNATIVES
Recommended for Prevention of Severe or Frequent Recurrences			
Herpes simplex	Frequent/severe recurrences	Valacyclovir 500 mg twice daily PO *or* Acyclovir 400 mg twice daily PO *or* Famciclovir 500 mg twice daily PO	
Candida (mucocutaneous)	Frequent/severe recurrences—primary prophylaxis not recommended	Fluconazole 100–200 mg daily PO	Posaconazole oral suspension 400 mg daily PO *or* Itraconazole 200 mg oral suspension daily *or* Posaconazole tablet 300 mg PO daily

Abbreviations: ART, antiretroviral therapy; bid, twice daily; cART, combination antiretroviral therapy; DS, double-strength; IFN-γ, interferon γ; IM, intramuscular; PCP, *Pneumocystis jirovecii* pneumonia; PO, by mouth; qd, daily; SS, single-strength; TB, tuberculosis; tid, three times a day.

secondary infections has decreased dramatically **(Fig. 208-34)**. Overall, the clinical spectrum of HIV disease is constantly changing as patients live longer and new and better approaches to treatment and prophylaxis are developed. In addition to the classic, original AIDS-defining illnesses, patients with HIV infection also have an increase in several serious non-AIDS illnesses, including non-AIDS-related cancers and cardiovascular, renal, and hepatic disease. Non-AIDS events (Table 208-4) now dominate the disease burden for patients with HIV infection successfully treated with ART. In developed countries, AIDS-related illnesses are responsible for only ~25% of deaths in patients with HIV infection. A similar percentage of deaths are due to non-AIDS-defining malignancies. Cardiovascular disease and liver disease each account for approximately 15% of deaths with 3% of deaths due to suicide The physician providing care to a patient with HIV infection must be well versed in general internal medicine as well as HIV-related opportunistic diseases. In general, it should be stressed that a key element of treatment of symptomatic complications of HIV disease, whether they are primary or secondary, is achieving good control of HIV replication with the use of ART and instituting primary and secondary prophylaxis for opportunistic infections as indicated.

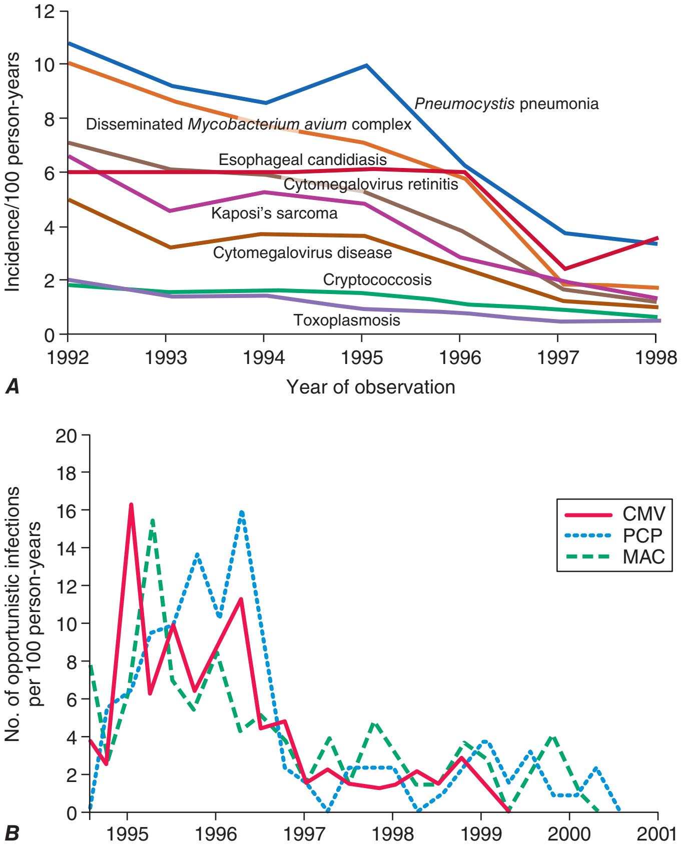

FIGURE 208-34 ***A.*** Decrease in the incidence of opportunistic infections and Kaposi's sarcoma in people with HIV with CD4+ T cell counts <100/μL from 1992 through 1998. *(Reproduced with permission from JE Kaplan et al: Epidemiology of human immunodeficiency virus-associated opportunistic infections in the United States in the era of highly active antiretroviral therapy. Clin Infect Dis 30:S5-14, 2000.)* ***B.*** Quarterly incidence rates of cytomegalovirus (CMV), *Pneumocystis jirovecii* pneumonia (PCP), and *Mycobacterium avium* complex (MAC) from 1995 to 2001. *(Reproduced with permission from Palella FJ Jr et al; HIV Outpatient Study Investigators. Durability and predictors of success of highly active antiretroviral therapy for ambulatory patients with HIV. AIDS 16:1617, 2002.)*

Diseases of the Respiratory System Acute bronchitis and sinusitis are prevalent during all stages of HIV infection. The most severe cases tend to occur in patients with lower CD4+ T cell counts. Sinusitis presents as fever, nasal congestion, and headache. The diagnosis is made by CT or MRI. The maxillary sinuses are most commonly involved; however, disease is also frequently seen in the ethmoid, sphenoid, and frontal sinuses. While some patients may improve without antibiotic therapy, radiographic improvement is quicker and more pronounced in patients who have received antimicrobial therapy. It is postulated that this high incidence of sinusitis results from an increased frequency of infection with encapsulated organisms such as *H. influenzae* and *S. pneumoniae*. In patients with low CD4+ T cell counts one may see mucormycosis infections of the sinuses. In contrast to the course of this infection in other patient populations, mucormycosis of the sinuses in patients with HIV infection may progress more slowly. In this setting aggressive, frequent local debridement in addition to local and systemic amphotericin B may result in effective treatment.

Pulmonary disease is one of the most frequent complications of HIV infection. The most common manifestation of pulmonary disease is pneumonia. Three of the 10 most common AIDS-defining illnesses are recurrent bacterial pneumonia, tuberculosis, and pneumonia due to the unicellular fungus *P. jirovecii*. Other major causes of pulmonary infiltrates include other mycobacterial infections, other fungal infections, nonspecific interstitial pneumonitis, KS, and lymphoma.

Bacterial pneumonia is seen with an increased frequency in patients with HIV infection, with 0.8–2.0 cases per 100 person-years. Patients with HIV infection are particularly prone to infections with encapsulated organisms. *S. pneumoniae* **(Chap. 153)** and *H. influenzae* **(Chap. 162)** are responsible for most cases of bacterial pneumonia in patients with AIDS. This may be a consequence of altered B cell function and/or defects in neutrophil function that may be secondary to HIV disease (see above). Pneumonias due to *S. aureus* **(Chap. 152)** and *P. aeruginosa* **(Chap. 170)** also are reported to occur with an increased frequency in patients with HIV infection. *S. pneumoniae* (pneumococcal) infection may be the earliest serious infection to occur in patients with

HIV disease. This can present as pneumonia, sinusitis, and/or bacteremia. Patients with untreated HIV infection have a 6-fold increase in the incidence of pneumococcal pneumonia and a 100-fold increase in the incidence of pneumococcal bacteremia. Pneumococcal disease may be seen in patients with relatively intact immune systems. In one study, the baseline CD4+ T cell count at the time of a first episode of pneumococcal pneumonia was ~300/μL. Of interest is the fact that the inflammatory response to pneumococcal infection appears proportional to the CD4+ T cell count. Due to this high risk of pneumococcal disease, immunization with the conjugated pneumococcal vaccine followed by booster immunization with the 23-valent pneumococcal polysaccharide vaccine is one of the generally recommended prophylactic measures for patients with HIV infection. This is likely most effective if given while the CD4+ T cell count is >200/μL and, if given to patients with lower CD4+ T cell counts, should be repeated once the count has been above 200 for 6 months. Although clear guidelines do not exist, it also makes sense to repeat immunization every 5 years. The incidence of bacterial pneumonia is cut in half when patients quit smoking.

Pneumocystis pneumonia (PCP; **Chap. 227**) is caused by the fungus *P. jirovecii* and was once the hallmark of AIDS. It has dramatically declined in incidence following the development of effective prophylactic regimens and the widespread use of ART. It is, however, still the single most common cause of pneumonia in patients with HIV infection in the United States and can be identified as a likely etiologic agent in 25% of cases of pneumonia in patients with HIV infection, with an incidence of about 1 case per 100 person-years. Approximately 30% of cases of HIV-associated PCP occur in patients who are unaware of their HIV status. The risk of PCP is greatest among those who have experienced a previous bout of PCP and those who have CD4+ T cell counts of <200/μL. Overall, 79% of patients with PCP have CD4+ T cell counts <100/μL and 95% of patients have CD4+ T cell counts <200/μL. Recurrent fever, night sweats, thrush, and unexplained weight loss also are associated with an increased incidence of PCP. For these reasons, it is strongly recommended that all patients with CD4+ T cell counts <200/μL (or a CD4 percentage <15) receive some form of PCP prophylaxis. The incidence of PCP is approaching zero in patients with known HIV infection receiving appropriate ART and prophylaxis. In the United States, primary PCP is now occurring at a median CD4+ T cell count of 36/μL, while secondary PCP is occurring at a median CD4+ T cell count of 10/μL.

Patients with PCP generally present with fever and a cough that is usually nonproductive or productive of only scant amounts of white sputum. They may complain of a characteristic retrosternal chest pain that is worse on inspiration and is described as sharp or burning. HIV-associated PCP may have an indolent course characterized by weeks of vague symptoms and should be included in the differential diagnosis of fever, pulmonary complaints, or unexplained weight loss in any patient with HIV infection and <200 CD4+ T cells/μL. The most common finding on chest x-ray is either a normal film, if the disease is suspected early, or a faint bilateral interstitial infiltrate. The classic finding of a dense perihilar infiltrate is unusual in patients with AIDS. In patients with PCP who have been receiving aerosolized pentamidine for prophylaxis, one may see an x-ray picture of upper lobe cavitary disease, reminiscent of TB. Other less common findings on chest x-ray include lobar infiltrates and pleural effusions. Thin-section CT may demonstrate a patchy ground-glass appearance. Routine laboratory evaluation is usually of little help in the differential diagnosis of PCP. A mild leukocytosis is common, although this may not be obvious in patients with prior neutropenia. Elevation of lactate dehydrogenase is common. Arterial blood-gases may indicate hypoxemia with a decline in $\mathrm{Pa}_{\mathrm{o}_2}$ and an increase in the arterial-alveolar (a–A) gradient. Arterial blood-gas measurements not only aid in making the diagnosis of PCP but also provide important information for staging the severity of the disease and directing treatment (see below). A definitive diagnosis of PCP requires demonstration of the organism in samples obtained from induced sputum, bronchoalveolar lavage, transbronchial biopsy, or open-lung biopsy. PCR has been used to detect specific DNA sequences for *P. jirovecii* in clinical specimens where histologic examinations have failed to make a diagnosis.

In addition to pneumonia, other clinical problems have been reported in people with HIV as a result of infection with *P. jirovecii*. Otic involvement may be seen as a primary infection, presenting as a polypoid mass involving the external auditory canal. In patients receiving aerosolized pentamidine for prophylaxis against PCP, one may see a variety of extrapulmonary manifestations of *P. jirovecii*. These include ophthalmic lesions of the choroid, a necrotizing vasculitis that resembles Buerger disease, bone marrow hypoplasia, and intestinal obstruction. Other organs that have been involved include lymph nodes, spleen, liver, kidney, pancreas, pericardium, heart, thyroid, and adrenals. Organ infection may be associated with cystic lesions that may appear calcified on CT or ultrasound.

The standard treatment for PCP or disseminated pneumocystosis is trimethoprim-sulfamethoxazole (TMP-SMX). A high (20–85%) incidence of side effects, particularly skin rash and bone marrow suppression, is seen with TMP-SMX in patients with HIV infection. Alternative treatments for mild to moderate PCP include dapsone/trimethoprim, clindamycin/primaquine, and atovaquone. IV pentamidine is the treatment of choice for severe disease in the patient unable to tolerate TMP-SMX. For patients with a $\mathrm{Pa}_{\mathrm{o}_2}$ <70 mmHg or with an a–A gradient >35 mmHg, adjunct glucocorticoid therapy should be used in addition to specific antimicrobials. Overall, treatment should be continued for 21 days and followed by secondary prophylaxis. Prophylaxis for PCP is indicated for any individual with HIV who has experienced a prior bout of PCP, any patient with a CD4+ T cell count of <200/μL or a CD4 percentage <15, any patient with unexplained fever for >2 weeks, and any patient with a recent history of oropharyngeal candidiasis. The preferred regimen for prophylaxis is TMP-SMX, one double-strength tablet daily. This regimen also provides protection against toxoplasmosis and some bacterial respiratory pathogens. For patients who cannot tolerate TMP-SMX, alternatives for prophylaxis include dapsone plus pyrimethamine plus leucovorin, aerosolized pentamidine administered by the Respirgard II nebulizer, and atovaquone. Primary or secondary prophylaxis for PCP can be discontinued in those patients treated with ART who maintain good suppression of HIV (<50 copies/mL) and CD4+ T cell counts >200/μL for at least 3 months.

M. tuberculosis, once thought to be on its way to extinction in the United States, experienced a resurgence associated with the HIV epidemic **(Chap. 183)**. Worldwide, approximately one-third of all AIDS-related deaths are associated with TB, and TB is the primary cause of death for 10–15% of patients with HIV infection. In the United States ~5% of untreated AIDS patients have active TB. Patients with HIV infection are more likely to have active TB by a factor of 100 when compared with an HIV-negative population. For an asymptomatic HIV-negative person with a positive purified protein derivative (PPD) skin test, the risk of reactivation TB is around 1% per year. For the patient with untreated HIV infection, a positive PPD skin test, and no signs or symptoms of TB, the rate of reactivation TB is 7–10% per year. Untreated TB can accelerate the course of HIV infection. Levels of plasma HIV RNA increase in the setting of active TB and decline in the setting of successful TB treatment. Active TB is most common in patients 25–44 years of age, in African Americans and Hispanics, in patients in New York City and Miami, and in patients in developing countries. In these demographic groups, 20–70% of the new cases of active TB are in patients with HIV infection. The epidemic of TB embedded in the epidemic of HIV infection probably represents the greatest health risk to the general public and the health care profession associated with the HIV epidemic. In contrast to infection with atypical mycobacteria such as MAC, active TB often develops relatively early in the course of HIV infection and may be an early clinical sign of HIV disease. In one study, the median CD4+ T cell count at presentation of TB was 326/μL.

The clinical manifestations of TB in people with HIV are quite varied and generally show different patterns as a function of the CD4+ T cell count. In patients with relatively high CD4+ T cell counts, the typical pattern of pulmonary reactivation occurs: patients present with fever, cough, dyspnea on exertion, weight loss, night sweats, and a chest x-ray revealing cavitary apical disease of the upper lobes. In patients with lower CD4+ T cell counts, disseminated disease is more common.

In these patients the chest x-ray may reveal diffuse or lower-lobe bilateral reticulonodular infiltrates consistent with miliary spread, pleural effusions, and hilar and/or mediastinal adenopathy. Infection may be present in bone, brain, meninges, GI tract, lymph nodes (particularly cervical lymph nodes), and viscera. Some patients with advanced HIV infection and active TB may have no symptoms of illness, and thus screening for TB should be part of the initial evaluation of every patient with HIV infection. Approximately 60–80% of people with HIV and TB have pulmonary disease, and 30–40% have extrapulmonary disease. Respiratory isolation and a negative-pressure room should be used for patients in whom a diagnosis of pulmonary TB is being considered. This approach is critical to limit nosocomial and community spread of infection. Culture of the organism from an involved site provides a definitive diagnosis. Blood cultures are positive in 15% of patients. This figure is higher in patients with lower CD4+ T cell counts. In the setting of fulminant disease, one cannot rely on the accuracy of a negative PPD skin test to rule out a diagnosis of TB. In addition, IFN-γ release assays may be difficult to interpret due to high backgrounds as a consequence of HIV-associated immune activation. TB is one of the conditions associated with HIV infection for which cure is possible with appropriate therapy. Therapy for TB is generally the same in the patient with HIV as in the HIV-negative patient **(Chap. 183)**. Due to the possibility of multidrug-resistant or extensively drug-resistant TB, drug susceptibility testing should be performed to guide therapy. Due to pharmacokinetic interactions, adjusted doses of rifabutin and/or changes in ART are required when treating TB in the setting of HIV infection. Treatment is most effective in programs that involve directly observed therapy. Initiation of ART and/or anti-TB therapy may be associated with clinical deterioration due to immune reconstitution inflammatory syndrome (IRIS) reactions. These are most common in patients initiating both treatments at the same time, may occur as early as 1 week after initiation of ART therapy, and are seen more frequently in patients with advanced HIV disease. For these reasons it is recommended that initiation of ART be delayed in antiretroviral-naïve patients with CD4 counts ≥50 cells/μL until 2–8 weeks following the initiation of treatment for TB. For patients with CD4 counts <50 cells/μL the benefits of more immediate ART outweigh the risks of IRIS, and ART should be started as soon as possible in those patients. Effective prevention of active TB can be a reality if the health care professional is aggressive in looking for evidence of latent or active TB by making sure that all patients with HIV infection receive a PPD skin test or evaluation with an IFN-γ release assay. Anergy testing is not of value in this setting. Since these tests rely on the host mounting an immune response to *M. tuberculosis*, patients with CD4+ T cell counts <200 cells/μL should be retested if their CD4+ T cell counts rise to persistently above 200. Patients at risk of continued exposure to TB should be tested annually. People with HIV with a skin-test reaction of >5 mm, those with a positive IFN-γ release assay, or those who are close household contacts of persons with active TB should receive treatment with 12 weeks of once-weekly isoniazid and rifapentin.

Atypical mycobacterial infections also are seen with an increased frequency in patients with HIV infection. Infections with at least 12 different mycobacteria have been reported, including *M. bovis* and representatives of all four Runyon groups. The most common atypical mycobacterial infection is with *M. avium* or *M. intracellulare* species—the *Mycobacterium avium* complex (MAC). Infections with MAC are seen mainly in patients in the United States and are rare in Africa. It has been suggested that prior infection with *M. tuberculosis* decreases the risk of MAC infection. MAC infections probably arise from organisms that are ubiquitous in the environment, including both soil and water. There is little evidence for person-to-person transmission of MAC infection. The presumed portals of entry are the respiratory and GI tracts. MAC infection is a late complication of HIV infection, occurring predominantly in patients with CD4+ T cell counts of <50/μL. The average CD4+ T cell count at the time of diagnosis is 10/μL. The most common presentation is disseminated disease with fever, weight loss, and night sweats. At least 85% of patients with MAC infection are mycobacteremic, and large numbers of organisms can often be demonstrated on bone marrow biopsy. The chest x-ray is abnormal in ~25% of patients, with the most common pattern being that of a bilateral, lower-lobe infiltrate suggestive of miliary spread. Alveolar or nodular infiltrates and hilar and/or mediastinal adenopathy also can occur. Other clinical findings include endobronchial lesions, abdominal pain, diarrhea, and lymphadenopathy. Anemia and elevated liver alkaline phosphatase are common. The diagnosis is made by the culture of blood or involved tissue. The finding of two consecutive sputum samples positive for MAC is highly suggestive of pulmonary infection. Cultures may take 2 weeks to turn positive. Therapy consists of a macrolide, usually clarithromycin, with ethambutol. Some physicians elect to add a third drug from among rifabutin, ciprofloxacin, or amikacin in patients with extensive disease. Therapy is continued until resolution of clinical signs and symptoms, negative cultures, and CD4+ T cell counts >100/μL for 3–6 months in the setting of ART. Primary prophylaxis for MAC is indicated in patients with HIV infection and CD4+ T cell counts <50/μL not immediately starting ART (Table 208-11). This may be discontinued in patients in whom ART induces a sustained suppression of viral replication regardless of the change in CD4+ T cell count.

Rhodococcus equi is a gram-positive, pleomorphic, acid-fast, non-spore-forming bacillus that can cause pulmonary and/or disseminated infection in patients with advanced HIV infection. Fever and cough are the most common presenting signs. Radiographically one may see cavitary lesions and consolidation. Blood cultures are often positive. Treatment is based on antimicrobial sensitivity testing.

Fungal infections of the lung, in addition to PCP, can be seen in patients with AIDS. Patients with pulmonary cryptococcal disease present with fever, cough, dyspnea, and, in some cases, hemoptysis. A focal or diffuse interstitial infiltrate is seen on chest x-ray in >90% of patients. In addition, one may see lobar disease, cavitary disease, pleural effusions, and hilar or mediastinal adenopathy. More than half of patients are fungemic, and 90% of patients have concomitant CNS infection. *Coccidioides immitis* is a mold that is endemic in the southwest United States. It can cause a reactivation pulmonary syndrome in patients with HIV infection. Most patients with this condition will have CD4+ T cell counts <250/μL. Patients present with fever, weight loss, cough, and extensive, diffuse reticulonodular infiltrates on chest x-ray. One may also see nodules, cavities, pleural effusions, and hilar adenopathy. While serologic testing is of value in the immunocompetent host, serologies are negative in 25% of patients with HIV and coccidioidal infection. Invasive aspergillosis is not an AIDS-defining illness and is generally not seen in patients with AIDS in the absence of neutropenia or administration of glucocorticoids. When it does occur, *Aspergillus* infection may have an unusual presentation in the respiratory tract of patients with AIDS, where it gives the appearance of a pseudomembranous tracheobronchitis. Primary pulmonary infection of the lung may be seen with *histoplasmosis*. The most common pulmonary manifestation of histoplasmosis, however, is in the setting of disseminated disease, presumably due to reactivation. In this setting respiratory symptoms are usually minimal, with cough and dyspnea occurring in 10–30% of patients. The chest x-ray is abnormal in ~50% of patients, showing either a diffuse interstitial infiltrate or diffuse small nodules, and the urine will often be positive for *Histoplasma* antigen.

Two forms of *idiopathic interstitial pneumonia* have been identified in patients with HIV infection: lymphoid interstitial pneumonitis (LIP) and nonspecific interstitial pneumonitis (NIP). LIP, a common finding in children, is seen in about 1% of adult patients with untreated HIV infection. This disorder is characterized by a benign infiltrate of the lung and is thought to be part of the polyclonal activation of lymphocytes seen in the context of HIV and EBV infections. Transbronchial biopsy is diagnostic in 50% of the cases, with an open-lung biopsy required for diagnosis in the remainder of cases. This condition is generally self-limited, and no specific treatment is necessary. Severe cases have been managed with brief courses of glucocorticoids. Although rarely a clinical problem since the use of ART, evidence of NIP may be seen in up to half of all patients with untreated HIV infection. Histologically, interstitial infiltrates of lymphocytes and plasma cells in a perivascular and peribronchial distribution are present. When symptomatic, patients present with fever and nonproductive cough occasionally accompanied by mild chest discomfort. Chest x-ray is usually normal or may reveal a faint interstitial pattern. Like LIP, NIP is a self-limited process for which no therapy is indicated other than appropriate management of the underlying HIV infection. HIV-related pulmonary arterial hypertension (HIV-PAH) is seen in ~0.5% of people with HIV. Patients may present with an array of symptoms including shortness of breath, fatigue, syncope, chest pain, and signs of right-sided heart failure. Chest x-ray reveals dilated pulmonary vessels and right-sided cardiomegaly with right ventricular hypertrophy seen on electrocardiogram. ART does not appear to be of clear benefit, and the prognosis is quite poor with a median survival in the range of 2 years.

Neoplastic diseases of the lung including KS and lymphoma are discussed below in the section on neoplastic diseases.

Diseases of the Cardiovascular System Heart disease is a relatively common postmortem finding in people with HIV (25–75% in autopsy series). The most common form of heart disease is coronary heart disease. In one large series the overall rate of myocardial infarction (MI) was 3.5/1000 patient-years, 28% of these events were fatal, and MI was responsible for 7% of all deaths in the cohort. In patients with HIV infection, cardiovascular disease may be associated with classic risk factors such as smoking, a direct consequence of HIV infection and the accompanying immune activation, or a complication of ART. In general, patients with HIV infection have higher levels of triglycerides, lower levels of high-density lipoprotein cholesterol, and a higher prevalence of smoking than cohorts of individuals without HIV infection. The finding that the rate of cardiovascular disease events was lower in patients on antiretroviral therapy than in those randomized to undergo a treatment interruption identified a clear association between HIV replication and risk of cardiovascular disease. In one study, a baseline CD4+ T cell count of <500/μL was found to be an independent risk factor for cardiovascular disease comparable in magnitude to that attributable to smoking. While the precise pathogenesis of this association remains unclear, it is likely related to the immune activation and increased propensity for coagulation seen because of HIV replication. Exposure to HIV protease inhibitors and certain reverse transcriptase inhibitors has been associated with increases in total cholesterol and/or risk of MI. Any increases in the risk of death from MI resulting from the use of certain antiretrovirals must be balanced against the marked increases in overall survival brought about by these drugs.

Another form of heart disease associated with HIV infection is a dilated cardiomyopathy associated with congestive heart failure (CHF) referred to as *HIV-associated cardiomyopathy*. This generally occurs as a late complication of HIV infection and, histologically, displays elements of myocarditis. For this reason, some have advocated it be treated with IV immunoglobulin (IVIg). HIV can be directly demonstrated in cardiac tissue in this setting, and there is debate over whether HIV plays a direct role in this condition. Patients present with typical findings of CHF including edema and shortness of breath. Patients with HIV infection may also develop cardiomyopathy as side effects of IFN-α or nucleoside analogue therapy. These are reversible once therapy is stopped. KS, cryptococcosis, Chagas' disease, and toxoplasmosis can involve the myocardium, leading to cardiomyopathy. In one series, most patients with HIV infection and a treatable myocarditis were found to have myocarditis associated with toxoplasmosis. Most of these patients also had evidence of CNS toxoplasmosis. Thus, MRI or double-dose contrast CT scan of the brain should be included in the workup of any patient with advanced HIV infection and cardiomyopathy.

A variety of other cardiovascular problems are found in patients with HIV infection. Pericardial effusions may be seen in the setting of advanced HIV infection. Predisposing factors include TB, CHF, mycobacterial infection, cryptococcal infection, pulmonary infection, lymphoma, and KS. While pericarditis is quite rare, in one series 5% of patients with HIV disease had pericardial effusions considered to be moderate or severe. Tamponade and death have occurred in association with pericardial KS, presumably owing to acute hemorrhage. Nonbacterial thrombotic endocarditis has been reported and should be considered in patients with unexplained embolic phenomena. IV pentamidine, when given rapidly, can result in hypotension as a consequence of cardiovascular collapse.

Diseases of the Oropharynx and Gastrointestinal System Oropharyngeal and GI diseases are common features of HIV infection. They are most frequently due to secondary infections. In addition, oral and GI lesions may occur with KS and lymphoma.

Oral lesions, including *thrush*, *hairy leukoplakia*, and *aphthous ulcers* **(Fig. 208-35)**, are particularly common in patients with untreated HIV infection. Thrush, due to *Candida* infection, and oral hairy leukoplakia, associated with EBV, are usually indicative of fairly advanced immunologic decline; they generally occur in patients with CD4+ T cell counts of <300/μL. In one study, 59% of patients with oral candidiasis went on to develop AIDS in the next year. Thrush appears as a white, cheesy exudate, often on an erythematous mucosa in the posterior oropharynx. While most commonly seen on the soft palate, early lesions are often found along the gingival vestibule. The diagnosis is made by direct examination of a scraping for pseudohyphal elements. Culturing is of no diagnostic value, as patients with HIV infection may have a positive throat culture for *Candida* in the absence of thrush. Oral hairy leukoplakia presents as white, frondlike lesions, generally along the lateral borders of the tongue and sometimes on the adjacent buccal mucosa (Fig. 208-35). Despite its name, oral hairy leukoplakia is not considered a premalignant condition. Lesions are associated with florid replication of EBV. While usually more disconcerting as a sign of HIV-associated immunodeficiency than a clinical problem in need of treatment, severe cases of oral hairy leukoplakia have been reported to respond to topical podophyllin or systemic therapy with anti-herpesvirus agents. Aphthous ulcers of the posterior oropharynx also are seen with regularity in patients with untreated HIV infection (Fig. 208-35). These lesions are of unknown etiology and can be quite painful and interfere with swallowing. Topical anesthetics provide immediate symptomatic relief of short duration. The fact that thalidomide is an effective treatment for this condition suggests that the pathogenesis may involve the action of tissue-destructive cytokines. Palatal, glossal, or gingival ulcers may also result from cryptococcal disease or histoplasmosis.

Esophagitis **(Fig. 208-36)** may present with odynophagia and retrosternal pain. Upper endoscopy is generally required to make an accurate diagnosis. Esophagitis may be due to *Candida*, CMV, or HSV. While CMV tends to be associated with a single large ulcer, HSV infection is more often associated with multiple small ulcers. The esophagus may also be the site of KS and lymphoma. Like the oral mucosa, the esophageal mucosa may have large, painful ulcers of unclear etiology that may respond to thalidomide. While achlorhydria is a common problem in patients with HIV infection, other gastric problems are generally rare. Among the neoplastic conditions involving the stomach are KS and lymphoma.

Infections of the small and large intestine leading to diarrhea, abdominal pain, and occasionally fever are among the most significant GI problems in people with HIV. They include infections with bacteria, protozoa, and viruses.

Bacteria may be responsible for infections of the GI tract in patients with HIV infection. Infections with enteric pathogens such as *Salmonella*, *Shigella*, and *Campylobacter* are more common in men who have sex with men and are often more severe and more apt to relapse in patients with HIV infection. Patients with untreated HIV have approximately a 20-fold increased risk of infection with *S. typhimurium*. They may present with a variety of nonspecific symptoms including fever, anorexia, fatigue, and malaise of several weeks' duration. Diarrhea is common but may be absent. Diagnosis is made by culture of blood and stool. Long-term therapy with ciprofloxacin is the recommended treatment. HIV-infected patients also have an increased incidence of *S. typhi* infection in areas of the world where typhoid is a problem. *Shigella* spp., particularly *S. flexneri*, can cause severe intestinal disease in people with HIV. Up to 50% of patients with GI disease will develop bacteremia. *Campylobacter* infections occur with an increased frequency in patients with HIV infection. While *C. jejuni* is the strain most frequently isolated, infections with many other strains have been

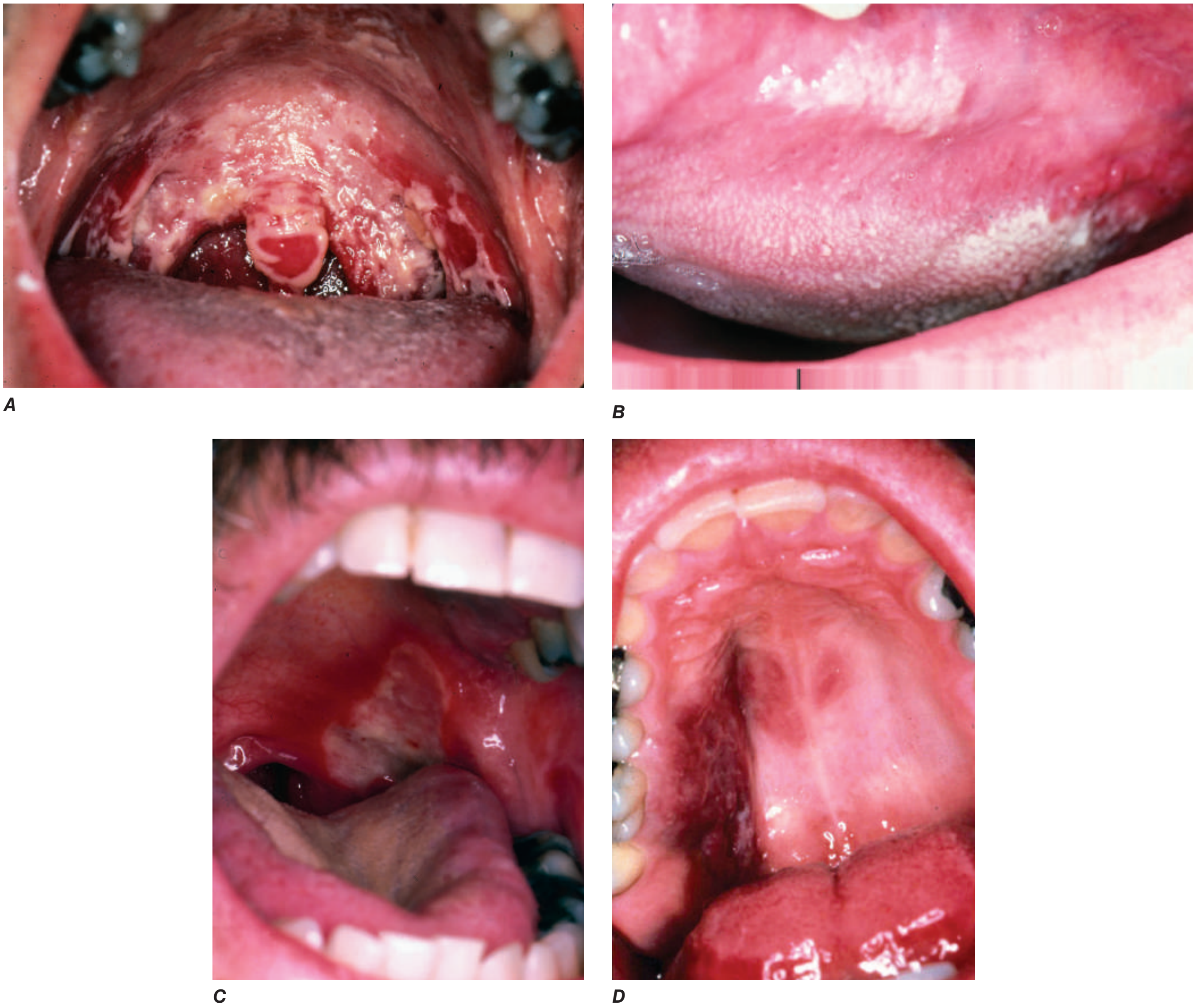

FIGURE 208-35 Various oral lesions in people with HIV. ***A.*** Thrush. ***B.*** Hairy leukoplakia. ***C.*** Aphthous ulcer. ***D.*** Kaposi's sarcoma.

reported. Patients usually present with crampy abdominal pain, fever, and bloody diarrhea. Infection may also present as proctitis. Stool examination reveals the presence of fecal leukocytes. Systemic infection can occur, with up to 10% of infected patients exhibiting bacteremia. Most strains are sensitive to erythromycin. Abdominal pain and diarrhea may be seen with MAC infection, and patients with HIV infection may have persistent diarrhea due to enteroaggregative *E. coli*.

Fungal infections including histoplasmosis, coccidioidomycosis, and penicilliosis have all been identified as a cause of fever and diarrhea in patients with HIV infection. Peritonitis has been seen with *C. immitis*.

Cryptosporidia, microsporidia, and *Isospora belli* **(Chap. 236)** are the most common opportunistic protozoa that infect the GI tract and cause diarrhea in patients with HIV. Cryptosporidial infection may present in a variety of ways, ranging from a self-limited or intermittent diarrheal illness in patients in the early stages of HIV infection to a severe, life-threatening diarrhea in severely immunodeficient individuals. In patients with untreated HIV infection and CD4+ T cell counts of <300/μL, the incidence of cryptosporidiosis is ~1% per year. In 75% of cases the diarrhea is accompanied by crampy abdominal pain, and 25% of patients have nausea and/or vomiting. Cryptosporidia may also cause biliary tract disease in the patient with HIV, leading to cholecystitis with or without accompanying cholangitis and pancreatitis secondary to papillary stenosis. The diagnosis of cryptosporidial diarrhea is made by stool examination or biopsy of the small intestine. The

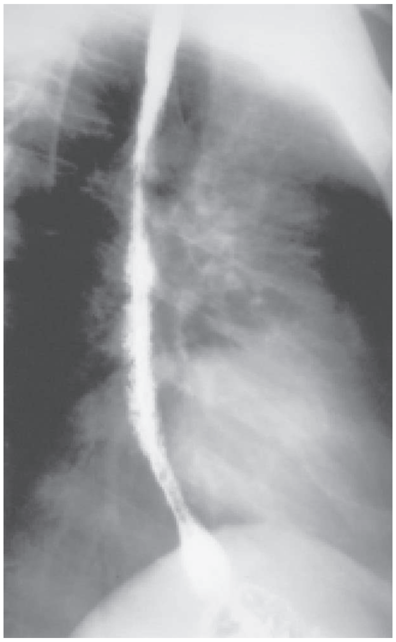

FIGURE 208-36 Barium swallow of a patient with *Candida* esophagitis. The flow of barium along the mucosal surface is grossly irregular.

diarrhea is noninflammatory, and the characteristic finding is the presence of oocysts that stain with acid-fast dyes. Therapy is predominantly supportive and marked improvements have been reported in the setting of effective ART. Treatment with up to 2000 mg/d of nitazoxanide (NTZ) is associated with improvement in symptoms or a decrease in shedding of organisms in about half of patients. Its overall role in the management of this condition remains unclear. Patients can minimize their risk of developing cryptosporidiosis by avoiding contact with human and animal feces, by not drinking untreated water from lakes or rivers, and by not eating raw shellfish.

Microsporidia are small, unicellular, obligate intracellular parasites that reside in the cytoplasm of enteric cells **(Chap. 236)**. The main species causing disease in humans is *Enterocytozoon bieneusi*. The clinical manifestations are similar to those described for cryptosporidia and include abdominal pain, malabsorption, diarrhea, and cholangitis. The small size of the organism may make it difficult to detect; however, with the use of chromotrope-based stains, organisms can be identified in stool samples by light microscopy. Definitive diagnosis generally depends on electron-microscopic examination of a stool specimen, intestinal aspirate, or intestinal biopsy specimen. In contrast to cryptosporidia, microsporidia have been noted in a variety of extraintestinal locations, including the eye, brain, sinuses, muscle, and liver, and they have been associated with conjunctivitis and hepatitis. The most effective way to deal with microsporidia in a patient with HIV infection is to restore the immune system by treating the HIV infection with ART. Albendazole, 400 mg bid, has been reported to be of benefit in some patients.

I. belli is a coccidian parasite **(Chap. 236)** most commonly found as a cause of diarrhea in patients from tropical and subtropical regions. Its cysts appear in the stool as large, acid-fast structures that can be differentiated from those of cryptosporidia based on size, shape, and number of sporocysts. The clinical syndromes of *Isospora* infection are identical to those caused by cryptosporidia. The important distinction is that infection with *Isospora* is generally easy to treat with TMP-SMX. While relapses are common, a thrice-weekly regimen of TMP-SMX appears adequate to prevent recurrence.

CMV colitis **(Chap. 200)** was once seen as a consequence of advanced immunodeficiency in 5–10% of patients with AIDS. It is much less common with the advent of ART. CMV colitis presents as diarrhea, abdominal pain, weight loss, and anorexia. The diarrhea is usually nonbloody, and the diagnosis is achieved through endoscopy and biopsy. Multiple mucosal ulcerations are seen at endoscopy, and biopsies reveal characteristic intranuclear and cytoplasmic inclusion bodies. Secondary bacteremias may result as a consequence of thinning of the bowel wall. Treatment is with either valganciclovir/ganciclovir or foscarnet for 3–6 weeks. Relapses are common, and maintenance therapy is typically necessary in patients whose HIV infection is poorly controlled. Patients with CMV disease of the GI tract should be carefully monitored for evidence of CMV retinitis.

In addition to disease caused by specific secondary infections, patients with HIV infection may also experience a chronic diarrheal syndrome for which no etiologic agent other than HIV can be identified. This entity is referred to as *AIDS enteropathy* or *HIV enteropathy*. It is most likely a direct result of HIV infection in the GI tract and improves with ART. Histologic examination of the small bowel in these patients reveals low-grade mucosal atrophy with a decrease in mitotic figures, suggesting a hyporegenerative state. Patients often have decreased or absent small-bowel lactase and malabsorption with accompanying weight loss.

The initial evaluation of a patient with HIV infection and diarrhea should include a set of stool examinations, including culture, examination for ova and parasites, and examination for *Clostridium difficile* toxin. Approximately 50% of the time this workup will demonstrate infection with pathogenic bacteria, mycobacteria, or protozoa. If the initial stool examinations are negative, additional evaluation, including upper and/or lower endoscopy with biopsy, will yield a diagnosis of microsporidial or mycobacterial infection of the small intestine ~30% of the time. In patients for whom this diagnostic evaluation is nonrevealing, a presumptive diagnosis of HIV enteropathy can be made if the diarrhea has persisted for >1 month. An algorithm for the evaluation of diarrhea in patients with HIV infection is given in **Fig. 208-37**.

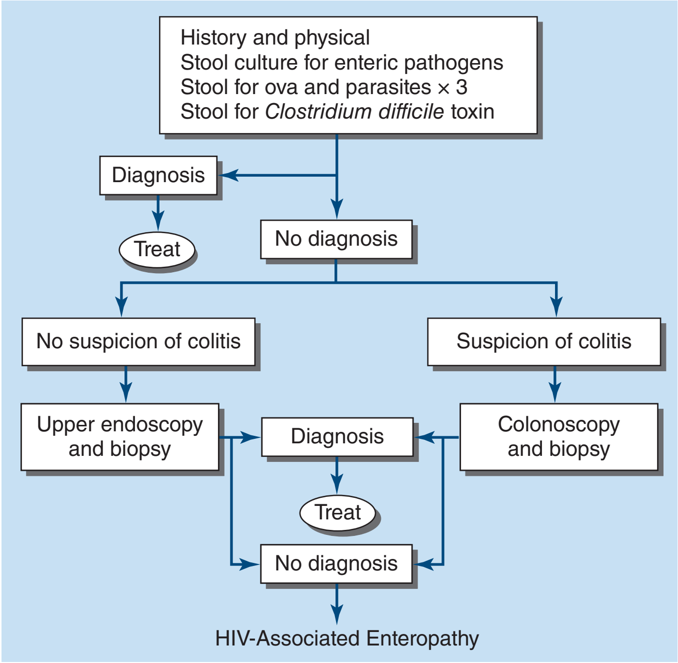
History and physical
Stool culture for enteric pathogens
Stool for ova and parasites × 3
Stool for *Clostridium difficile* toxin
Diagnosis
Treat
No diagnosis
No suspicion of colitis
Suspicion of colitis
Upper endoscopy and biopsy
Diagnosis
Colonoscopy and biopsy
Treat
No diagnosis
HIV-Associated Enteropathy

FIGURE 208-37 Algorithm for the evaluation of diarrhea in a patient with HIV infection. HIV-associated enteropathy is a diagnosis of exclusion and can be made only after other, generally treatable, forms of diarrheal illness have been ruled out.

Rectal lesions are common in patients with HIV, particularly the perirectal ulcers and erosions due to the reactivation of HSV **(Fig. 208-38)**. These lesions may appear quite atypical, as denuded skin without vesicles. They typically respond well to treatment with valacyclovir, famciclovir, or foscarnet. Other rectal lesions encountered in patients with HIV infection include condylomata acuminata, KS, and intraepithelial neoplasia (see below).

Hepatobiliary Diseases Diseases of the hepatobiliary system are a major problem in patients with HIV infection. It has been estimated that approximately 15% of the deaths of patients with HIV infection

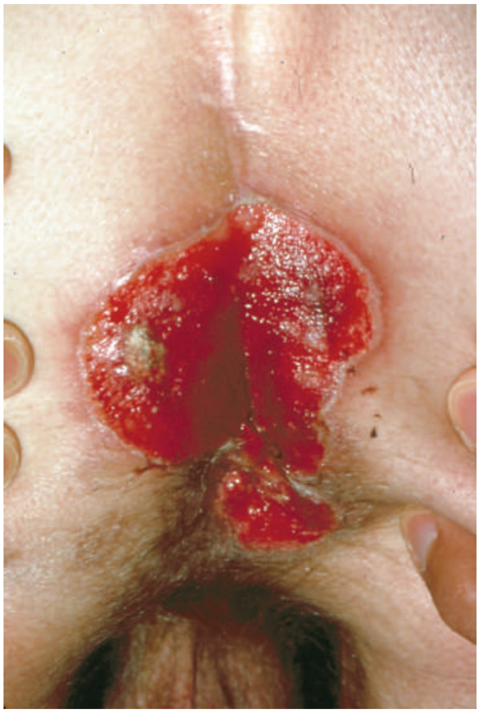

FIGURE 208-38 Severe, erosive perirectal herpes simplex in a patient with AIDS.

are related to liver disease. While this is predominantly a reflection of the problems encountered in the setting of co-infection with hepatitis B or C, it is also a reflection of the hepatic injury, ranging from hepatic steatosis to hypersensitivity reactions to immune reconstitution, that can be seen in the context of ART.

The prevalence of co-infection with HIV and hepatitis viruses varies by geographic region. In the United States, ~90% of people with HIV have evidence of infection with HBV; 6–14% have chronic HBV infection; 5–50% of patients are co-infected with HCV; and co-infections with hepatitis D, E, and/or G viruses are common. Among IV drug users with HIV infection, rates of HCV infection range from 70 to 95%. HIV infection has a significant impact on the course of hepatitis virus infection. It is associated with approximately a 3-fold increase in the development of persistent hepatitis B surface antigenemia. Patients infected with both HBV and HIV have decreased evidence of inflammatory liver disease. The presumption that this is due to the immunosuppressive effects of HIV infection is supported by the observations that this situation can be reversed, and one may see the development of more severe hepatitis following the initiation of effective ART. In studies of the impact of HIV on HBV infection, 4- to 10-fold increases in liver-related mortality rates have been noted in patients with HIV and active HBV infection compared to rates in patients with either infection alone. There is, however, only a slight increase in overall mortality rate in people with HIV who are also hepatitis B surface antigen (HBsAg)–positive. IFN-α is less successful as treatment for HBV in patients with HIV co-infection. Lamivudine, emtricitabine, adefovir/tenofovir/entecavir, and telbivudine alone or in combination are useful in the treatment of hepatitis B in patients with HIV infection. It is important to remember that all the above-mentioned drugs also have activity against HIV and should not be used alone in patients with HIV infection, to avoid the emergence of quasispecies of HIV resistant to these drugs. For this reason, the treatment of hepatitis B infection in a patient with HIV infection should always be done in the setting of ART, and alterations in ART need to take into account that the current regimen is also treating HBV. HCV infection is more severe in the patient with HIV infection; it does not appear to affect overall mortality rates in people with HIV when other variables such as age, baseline CD4+ T cell count, and use of ART are taken into account. In the setting of HIV and HCV co-infection, levels of HCV are approximately 10-fold higher than in the HIV-negative patient with HCV infection. There is a 50% higher overall mortality rate with a five-fold increased risk of death due to liver disease in patients chronically infected with both HCV and HIV. Use of directly acting agents for the treatment of HCV leads to cure rates approaching 100%, even in patients with HIV co-infection. Successful treatment of HCV in patients with HIV decreases mortality. Hepatitis A virus infection is not seen with an increased frequency in patients with HIV infection. It is recommended that all patients with HIV infection who have not experienced natural infection be immunized with hepatitis A and/or hepatitis B vaccines. Infection with hepatitis G virus, also known as GB virus C, is seen in ~50% of patients with HIV infection. For reasons that are currently unclear, there are data to suggest that patients with HIV infection co-infected with this virus have a decreased rate of progression to AIDS.

A variety of other infections also may involve the liver. Granulomatous hepatitis may be seen as a consequence of mycobacterial or fungal infections, particularly MAC infection. Hepatic masses may be seen in the context of TB, peliosis hepatis, or fungal infection. Among the fungal opportunistic infections, *C. immitis* and *Histoplasma capsulatum* **(Chap. 218)** are those most likely to involve the liver. Biliary tract disease in the form of papillary stenosis or sclerosing cholangitis has been reported in the context of cryptosporidiosis, CMV infection, and KS. When no diagnosis can be made, the term *AIDS cholangiopathy* is used. Hemophagocytic lymphohistiocytosis of the liver has been seen in the setting of Hodgkin's disease and may occur prior to diagnosis of the underlying neoplasm.

Many of the drugs used to treat HIV infection are metabolized by the liver and can cause liver injury. Fatal hepatic reactions have been reported with a wide array of antiretrovirals including nucleoside analogues, nonnucleoside analogues, and protease inhibitors. Nucleoside analogues work by inhibiting DNA synthesis. This can result in toxicity to mitochondria, which can lead to disturbances in oxidative metabolism. This may manifest as hepatic steatosis and, in severe cases, lactic acidosis and fulminant liver failure. It is important to be aware of this condition and to watch for it in patients with HIV infection receiving nucleoside analogues. It is reversible if diagnosed early and the offending agent(s) discontinued. Nevirapine has been associated with at times fatal fulminant and cholestatic hepatitis, hepatic necrosis, and hepatic failure. Indinavir may cause mild to moderate elevations in serum bilirubin in 10–15% of patients in a syndrome similar to Gilbert's syndrome. A similar pattern of hepatic injury may be seen with atazanavir. In the patient receiving ART with an unexplained increase in hepatic transaminases, strong consideration should be given to drug toxicity.

Pancreatic injury is most commonly a consequence of drug toxicity, notably that secondary to pentamidine or dideoxynucleosides. While up to half of patients in some series have biochemical evidence of pancreatic injury, <5% of patients show any clinical evidence of pancreatitis that is not linked to a drug toxicity.

Diseases of the Kidney and Genitourinary Tract Diseases of the kidney or genitourinary tract may be a direct consequence of HIV infection, due to an opportunistic infection or neoplasm, or related to drug toxicity. Overall, microalbuminuria is seen in ~20% of untreated patients with HIV; significant proteinuria is seen in closer to 2%. The presence of microalbuminuria has been associated with an increase in all-cause mortality. *HIV-associated nephropathy* (HIVAN) was first described in IDUs and was initially thought to be IDU nephropathy in patients with HIV infection; it is now recognized as a true direct complication of HIV infection. Although most patients with this condition have CD4+ T cell counts <200/μL, HIV-associated nephropathy can be an early manifestation of HIV infection and is also seen in children. More than 90% of reported cases have been in African-American or Hispanic individuals; the disease is not only more prevalent in these populations but also more severe and at one point was the third leading cause of end-stage renal failure among African Americans age 20–64 years in the United States. Proteinuria is the hallmark of this disorder. Edema and hypertension are rare. Ultrasound examination reveals enlarged, hyperechogenic kidneys. A definitive diagnosis is obtained through renal biopsy. Histologically, focal segmental glomerulosclerosis is present in 80%, and mesangial proliferation in 10–15% of cases. Prior to effective antiretroviral therapy, this disease was characterized by relatively rapid progression to end-stage renal disease. Patients with HIV-associated nephropathy should be treated for their HIV infection. Treatment with angiotensin-converting enzyme (ACE) inhibitors and/or prednisone, 60 mg/d, have been reported to be of benefit in some cases. The incidence of this disease in patients receiving adequate ART has not been well defined; however, the impression is that it has decreased in frequency and severity. It is the leading cause of end-stage renal disease in patients with HIV infection.

Among the drugs commonly associated with renal damage in patients with HIV disease are pentamidine, amphotericin, adefovir, cidofovir, tenofovir, and foscarnet. Switching from tenofovir disoproxil fumarate (TDF) to tenofovir alafenamide (TAF) may lead to a decrease in renal injury from TDF. TMP-SMX may compete for tubular secretion with creatinine and cause an increase in the serum creatinine level. The pharmacokinetic booster cobicistat, a component of several fixed-drug ART formulations, inhibits renal tubular secretion of creatinine leading to increased serum creatinine levels without a true decline in glomerular filtration rate. Sulfadiazine may crystallize in the kidney and result in an easily reversible form of renal shutdown, while indinavir or atazanavir may form renal calculi. Adequate hydration is the mainstay of treatment and prevention for these latter two conditions.

Genitourinary tract infections are seen with a high frequency in patients with HIV infection; they present with skin lesions, dysuria, hematuria, and/or pyuria and are managed in the same fashion as in patients without HIV infection. Infections with HSV are covered below ("Dermatologic Diseases"). Infections with *T. pallidum*, the etiologic agent of *syphilis*, play an important role in the HIV epidemic. In HIV-negative individuals, genital syphilitic ulcers as well as the ulcers of chancroid are major predisposing factors for heterosexual transmission of HIV infection. While most people with HIV and syphilis have a typical presentation of the latter, a variety of formerly rare clinical problems may be encountered in the setting of dual infection. Among them are *lues maligna*, an ulcerating lesion of the skin due to a necrotizing vasculitis and associated with unexplained fever; nephrotic syndrome; and neurosyphilis. The most common presentation of syphilis in the patient with HIV is that of *condylomata lata*, a form of secondary syphilis. Neurosyphilis may be asymptomatic or may present as acute meningitis, neuroretinitis, deafness, or stroke. The rate of neurosyphilis may be as high as 1% in patients with HIV infection, and one should consider a lumbar puncture to look for neurosyphilis in all patients with HIV infection and secondary syphilis. As a consequence of the immunologic abnormalities seen in the setting of HIV infection, diagnosis of syphilis through standard serologic testing may be challenging. On the one hand, a significant number of patients have false-positive Venereal Disease Research Laboratory (VDRL) tests due to polyclonal B cell activation. On the other hand, the development of a new positive VDRL may be delayed in patients with new infections, and the anti–fluorescent treponemal antibody (anti-FTA) test may be negative in the setting of immunodeficiency. Thus, dark-field examination of appropriate specimens should be performed in any patient in whom syphilis is suspected, even if the patient has a negative VDRL. Similarly, any patient with a positive serum VDRL test, neurologic findings, and an abnormal spinal fluid examination should be considered to have neurosyphilis and treated accordingly, regardless of the CSF VDRL result. In any setting, patients treated for syphilis need to be carefully monitored to ensure adequate therapy. Approximately one-third of patients with HIV infection will experience a Jarisch-Herxheimer reaction upon initiation of therapy for syphilis.

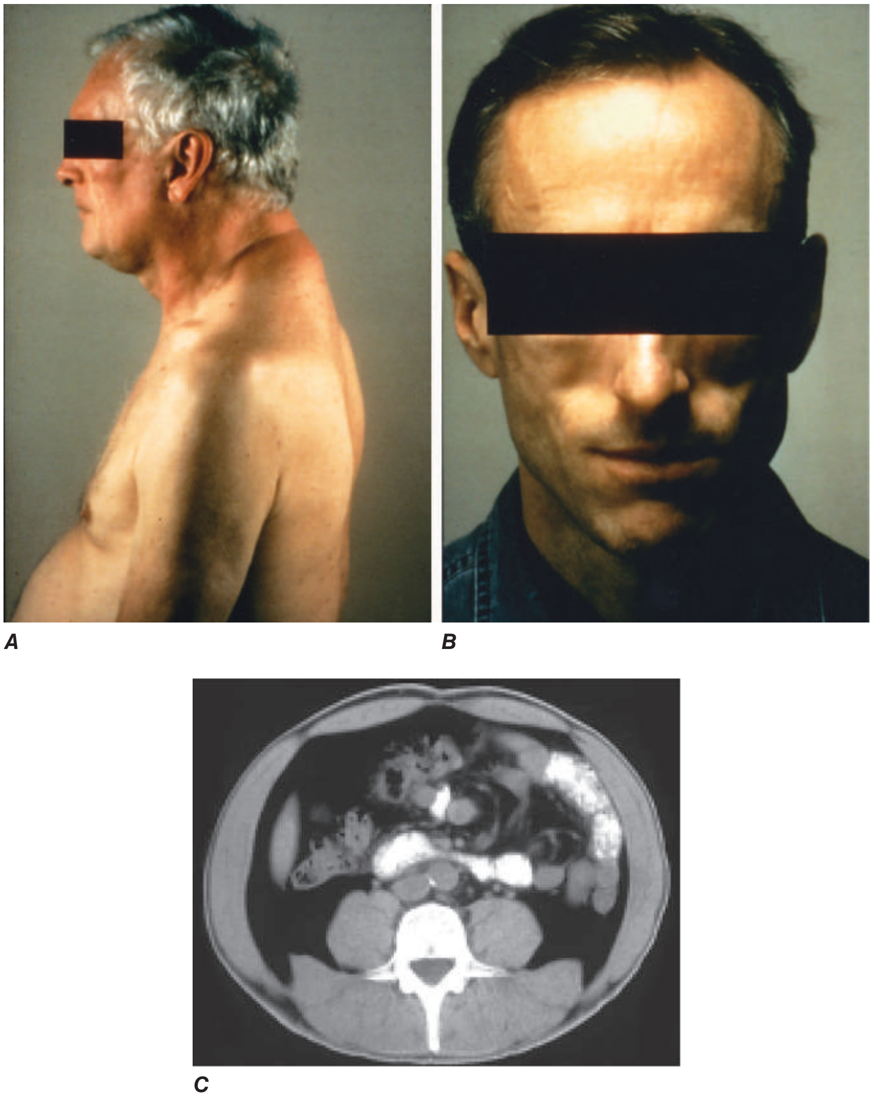

FIGURE 208-39 Characteristics of lipodystrophy. ***A.*** Truncal obesity and buffalo hump. ***B.*** Facial wasting. ***C.*** Accumulation of intraabdominal fat on CT scan.

Vulvovaginal candidiasis is a common problem in women with HIV infection. Symptoms include pruritus, discomfort, dyspareunia, and dysuria. Vulvar infection may present as a morbilliform rash that may extend to the thighs. Vaginal infection is usually associated with a white discharge, and plaques may be seen along an erythematous vaginal wall. Diagnosis is made by microscopic examination of the discharge for pseudohyphal elements in a 10% potassium hydroxide solution. Mild disease can be treated with topical therapy. More serious disease can be treated with fluconazole. Other causes of vaginitis include *Trichomonas* and mixed bacteria.

Diseases of the Endocrine System and Metabolic Disorders

A variety of endocrine and metabolic disorders are seen in the context of HIV infection. These may be a direct consequence of HIV infection, secondary to opportunistic infections or neoplasms, or related to medication side effects. Between 33% and 75% of patients with HIV infection receiving thymidine analogues or protease inhibitors as a component of ART develop a syndrome often referred to as *lipodystrophy*, consisting of elevations in plasma triglycerides, total cholesterol, and apolipoprotein B, as well as hyperinsulinemia and hyperglycemia. Many of the patients have been noted to have a characteristic set of body habitus changes associated with fat redistribution, consisting of truncal obesity coupled with peripheral wasting **(Fig. 208-39)**. Truncal obesity is apparent as an increase in abdominal girth related to increases in mesenteric fat, a dorsocervical fat pad ("buffalo hump") reminiscent of patients with Cushing's syndrome, and enlargement of the breasts. The peripheral wasting, or lipoatrophy, is particularly noticeable in the face and buttocks and by the prominence of the veins in the legs. These changes may develop at any time ranging from ~6 weeks to several years following the initiation of ART. Approximately 20% of the patients with HIV-associated lipodystrophy meet the criteria for the *metabolic syndrome* as defined by the International Diabetes Federation or the U.S. National Cholesterol Education Program (NCEP) Adult Treatment Panel III. The lipodystrophy syndrome has been reported in association with regimens containing a variety of different drugs, and while initially reported in the setting of protease inhibitor therapy, it appears that similar changes can also be induced by protease-sparing regimens. It has been suggested that the lipoatrophy changes are particularly severe in patients receiving the thymidine analogues stavudine and zidovudine. Current treatment guidelines avoid these drugs and recommend drugs with fewer of these side effects. NCEP guidelines should be followed in the management of these lipid abnormalities **(Chap. 419)**, with a further recommendation that moderate-intensity statin therapy be given to all individuals with HIV infection between the ages of 40 and 75 years unless their 10-year risk of cardiovascular disease is <5%. In all patients, consideration should be given to changing the components of ART with avoidance of thymidine analogues (azidothymidine and stavudine) and offending protease inhibitors. Due to concerns regarding drug interactions, the most utilized lipid-lowering agents in this setting are gemfibrozil and atorvastatin. Lactic acidosis is associated with certain ART medications. This is most often seen with the nucleoside analogue reverse transcriptase inhibitors and can be fatal.

Patients with advanced HIV disease may develop hyponatremia due to the syndrome of inappropriate antidiuretic hormone (vasopressin) secretion (SIADH) because of increased free-water intake and decreased free-water excretion. SIADH is usually seen in conjunction with pulmonary or CNS disease. Low serum sodium may also be due to adrenal insufficiency; a concomitant high serum potassium should alert one to this possibility. Hyperkalemia may be secondary to adrenal insufficiency, HIV nephropathy; or medications, particularly trimethoprim and pentamidine. Hypokalemia may be seen in the setting of tenofovir or amphotericin therapy. Adrenal gland disease may be due to mycobacterial infections, CMV disease, cryptococcal disease, histoplasmosis, or ketoconazole toxicity. Iatrogenic Cushing's syndrome with suppression of the hypothalamic-pituitary-adrenal axis may be seen with the use of local glucocorticoids (injected or inhaled) in patients receiving ritonavir or cobicistat. This is due to inhibition of the hepatic enzyme CYP3A4 by ritonavir leading to prolongation of the glucocorticoid half-life.

Thyroid function may be altered in 10–15% of patients with HIV infection. Both hypo- and hyperthyroidism may be seen. The predominant abnormality is subclinical hypothyroidism. In the setting of ART, up to 10% of patients have been noted to have elevated thyroid-stimulating hormone levels, suggesting that this may be a manifestation of immune reconstitution. Immune-reconstitution Graves' disease may occur as a late (9–48 months) complication of ART. In advanced HIV disease, infection of the thyroid gland may occur with opportunistic pathogens, including *P. jirovecii*, CMV, mycobacteria, *Toxoplasma gondii*, and *Cryptococcus neoformans*. These infections are generally associated with a nontender, diffuse enlargement of the thyroid gland. Thyroid function is usually normal. Diagnosis is made by fine-needle aspirate or open biopsy.

Depending on the severity of disease, HIV infection is associated with *hypogonadism* in 20–50% of men and is lowest in the setting of ART. While this is generally a complication of underlying illness, testicular dysfunction may also be a side effect of ganciclovir therapy. In some surveys, up to two-thirds of patients report decreased libido and one-third complain of erectile dysfunction. Androgen-replacement therapy should be considered in patients with symptomatic hypogonadism. HIV infection does not seem to have a significant effect on the menstrual cycle outside the setting of advanced disease.

Immunologic and Rheumatologic Diseases Immunologic and rheumatologic disorders are common in patients with HIV infection and range from excessive immediate-type hypersensitivity reactions **(Chap. 363)** to an increase in the incidence of reactive arthritis **(Chap. 386)** to conditions characterized by a diffuse infiltrative lymphocytosis. The occurrence of these phenomena is an apparent paradox in the setting of the profound immunodeficiency and immunosuppression that characterizes HIV infection and reflects the complex nature of the immune system and its regulatory mechanisms.

Drug allergies are the most significant allergic reactions occurring in patients with HIV and appear to become more common as the disease progresses. They occur in up to 65% of patients who receive therapy with TMP-SMX for PCP. In general, these drug reactions are characterized by erythematous, morbilliform eruptions that are pruritic, tend to coalesce, and are often associated with fever. Nonetheless, ~33% of patients can be maintained on the offending therapy, and thus these reactions are not an immediate indication to stop the drug. Anaphylaxis is extremely rare in patients with HIV infection, and patients who have a cutaneous reaction during a single course of therapy can still be considered candidates for future treatment or prophylaxis with the same agent. The one exception to this is the nucleoside analogue abacavir, where fatal hypersensitivity reactions have been reported with rechallenge. This hypersensitivity is strongly associated with the HLA-B5701 haplotype, and a hypersensitivity reaction to abacavir is an absolute contraindication to future therapy. For other agents, including TMP-SMX, desensitization regimens are moderately successful. While the mechanisms underlying these allergic-type reactions remain unknown, patients with HIV infection have been noted to have elevated IgE levels that increase as the CD4+ T cell count declines. The numerous examples of patients with multiple drug reactions suggest that a common pathway is involved.

HIV infection shares many similarities with a variety of autoimmune diseases, including a substantial polyclonal B cell activation that is associated with a high incidence of antiphospholipid antibodies, such as anticardiolipin antibodies, VDRL antibodies, and lupus-like anticoagulants. In addition, people with HIV have an increased incidence of antinuclear antibodies. Despite these serologic findings, there is no evidence that people with HIV have an increase in two of the more common autoimmune diseases, i.e., systemic lupus erythematosus and rheumatoid arthritis. In fact, it has been observed that these diseases may be somewhat ameliorated by the concomitant presence of HIV infection, suggesting that an intact CD4+ T cell limb of the immune response plays an integral role in the pathogenesis of these conditions. Similarly, there are anecdotal reports of patients with common variable immunodeficiency **(Chap. 362)**, characterized by hypogammaglobulinemia, who have had a normalization of Ig levels following the development of HIV infection, suggesting a possible role for overactive CD4+ T cell immunity in certain forms of that syndrome. The one autoimmune disease that may occur with an increased frequency in patients with HIV infection is a variant of primary Sjögren's syndrome **(Chap. 373)** in which patients with HIV infection develop a syndrome consisting of parotid gland enlargement, dry eyes, and dry mouth. This condition is associated with lymphocytic infiltrates of the salivary gland and lung. One also can see peripheral neuropathy, polymyositis, renal tubular acidosis, and hepatitis. In contrast to Sjögren's syndrome, in which the lymphocytic infiltrates are composed predominantly of CD4+ T cells, in patients with HIV infection the infiltrates are composed predominantly of CD8+ T cells. In addition, while patients with Sjögren's syndrome are mainly women who have autoantibodies to Ro and La and who frequently have HLA-DR3 or B8 MHC haplotypes, people with HIV who have this syndrome are usually African-American men who do not have anti-Ro or anti-La and who most often have the HLA-DR5 haplotype. This syndrome appears to be less common with the increased use of effective ART. The term *diffuse infiltrative lymphocytosis syndrome* (DILS) is used to describe this entity and to distinguish it from Sjögren's syndrome.

Approximately one-third of people with HIV experience arthralgias; furthermore, 5–10% are diagnosed as having some form of reactive arthritis, such as Reiter's syndrome or psoriatic arthritis as well as undifferentiated spondyloarthropathy **(Chap. 374)**. These syndromes occur with increasing frequency as the competency of the immune system declines. This association may be related to an increase in the number of infections with organisms that may trigger a reactive arthritis with progressive immunodeficiency or to a loss of important regulatory T cells. Reactive arthritides in people with HIV generally respond well to standard treatment; however, therapy with methotrexate has been associated with an increase in the incidence of opportunistic infections and should be used with caution and only in severe cases.

People with HIV also experience a variety of joint problems without obvious cause that are referred to generically as *HIV-* or *AIDS-associated arthropathy*. This syndrome is characterized by subacute oligoarticular arthritis developing over a period of 1–6 weeks and lasting 6 weeks to 6 months. It generally involves the large joints, predominantly the knees and ankles, and is nonerosive with only a mild inflammatory response. X-rays are nonrevealing. Nonsteroidal anti-inflammatory drugs are only marginally helpful; however, relief has been noted with the use of intraarticular glucocorticoids. A second form of arthritis also thought to be secondary to HIV infection is called *painful articular syndrome*. This condition, reported as occurring in up to 10% of AIDS patients, presents as an acute, severe, sharp pain in the affected joint. It affects primarily the knees, elbows, and shoulders; lasts 2–24 h; and may be severe enough to require narcotic analgesics. The cause of this arthropathy is unclear; however, it is thought to result from a direct effect of HIV on the joint. This condition is reminiscent of the fact that other lentiviruses, in particular the caprine arthritis-encephalitis virus, are capable of directly causing arthritis.

A variety of other immunologic or rheumatologic diseases have been reported in people with HIV, either de novo or in association with opportunistic infections or drugs. Using the criteria of widespread musculoskeletal pain of at least 3 months' duration and the presence of at least 11 of 18 possible tender points by digital palpation, 11% of a cohort of people with HIV containing 55% IDUs were diagnosed as having *fibromyalgia* **(Chap. 385)**. While the incidence of frank arthritis was less in this population than in other studied populations that consisted predominantly of men who have sex with men, these data support the concept that there are musculoskeletal problems that occur as a direct result of HIV infection. CNS angiitis and polymyositis also have been reported in people with HIV. Septic arthritis is surprisingly rare, especially given the increased incidence of staphylococcal bacteremias seen in this population. When septic arthritis has been reported, it has usually been due to *Staphylococcus aureus*, systemic fungal infection with *C. neoformans*, *Sporothrix schenckii*, or *H. capsulatum* or to systemic mycobacterial infection with *M. tuberculosis*, *M. haemophilum*, *M. avium*, or *M. kansasii*.

Patients with HIV infection treated with ART have been found to have an increased incidence of osteonecrosis or avascular necrosis of the hip and shoulders. In a study of asymptomatic patients, 4.4% were found to have evidence of osteonecrosis on MRI. While precise cause-and-effect relationships have been difficult to establish, this complication has been associated with the use of lipid-lowering agents, systemic glucocorticoids, and testosterone; bodybuilding exercise; alcohol consumption; and the presence of anticardiolipin antibodies. Osteoporosis has been reported in 7% of women with HIV infection, with 41% of women demonstrating some degree of osteopenia. Several studies have documented decreases in bone mineral density of 2–6% in the first 2 years following the initiation of ART. This may be particularly apparent with tenofovir-containing regimens.

Immune Reconstitution Inflammatory Syndrome (IRIS) Following the initiation of effective ART, a paradoxical worsening of pre-existing, untreated, or partially treated opportunistic infections may be noted. One may also see exacerbations of pre-existing autoimmune conditions or the development of new autoimmune conditions following the initiation of antiretrovirals **(Table 208-12)**. IRIS related to a known pre-existing infection or neoplasm is referred to as *paradoxical IRIS*, while IRIS associated with a previously undiagnosed condition is referred to as *unmasking IRIS*. The term *immune reconstitution disease* (IRD) is sometimes used to distinguish IRIS manifestations related to opportunistic diseases from IRIS manifestations related to autoimmune diseases. IRD is particularly common in patients with underlying untreated mycobacterial or fungal infections. Some form of IRIS is seen in 10–30% of patients following the initiation of ART, depending on the clinical setting, and is most common in patients starting therapy with CD4+ T cell counts <50 cells/μL who have a precipitous drop in HIV RNA levels following the initiation of ART. Signs and symptoms may appear anywhere from 2 weeks to 2 years after the initiation of ART and can include localized lymphadenitis, prolonged fever, pulmonary infiltrates, hepatitis, increased intracranial pressure, uveitis, sarcoidosis, and Graves' disease. The clinical course can be protracted, and severe cases can be fatal. The underlying mechanism appears to be related to a phenomenon similar to type IV hypersensitivity reactions and reflects the immediate improvements in immune function that occur as levels of HIV RNA drop and the immunosuppressive effects of HIV infection are controlled. In severe cases, the use of immunosuppressive drugs such as glucocorticoids may be required to blunt the inflammatory component of these reactions while specific antimicrobial therapy takes effect.

TABLE 208-12 Characteristics of Immune Reconstitution Inflammatory Syndrome (IRIS)

Paradoxical worsening of an existing clinical condition or abrupt appearance of a new clinical finding (unmasking) is seen following the initiation of antiretroviral therapy
Occurs weeks to months following the initiation of antiretroviral therapy
Is most common in patients starting therapy with a CD4+ T cell count <50/μL who experience a precipitous drop in viral load
Is frequently seen in the setting of tuberculosis, particularly when ART is started soon after initiation of anti-TB therapy
Can be fatal

Abbreviations: ART, antiretroviral therapy; TB, tuberculosis.

TABLE 208-13 Causes of Bone Marrow Suppression in Patients with HIV Infection

DISEASES	MEDICATIONS
HIV infection	Zidovudine
Mycobacterial infections	Dapsone
Fungal infections	Trimethoprim-sulfamethoxazole
B19 parvovirus infection	Pyrimethamine
Lymphoma	5-flucytosine
	Ganciclovir
	Interferon α
	Trimetrexate
	Foscarnet

Diseases of the Hematopoietic System Disorders of the hematopoietic system including lymphadenopathy, anemia, leukopenia, and/or thrombocytopenia are common throughout the course of HIV infection and may be the direct result of HIV, manifestations of secondary infections and neoplasms, or side effects of therapy **(Table 208-13)**. Direct histologic examination and culture of lymph node or bone marrow tissue are often diagnostic. A significant percentage of bone marrow aspirates from patients with HIV infection have been reported to contain lymphoid aggregates, the precise significance of which is unknown. Initiation of ART will lead to reversal of most hematologic complications that are the direct result of HIV infection.

Some patients, otherwise asymptomatic, may develop *persistent generalized lymphadenopathy* as an early clinical manifestation of HIV infection. This condition is defined as the presence of enlarged lymph nodes (>1 cm) in two or more extrainguinal sites for >3 months without an obvious cause. The lymphadenopathy in this setting is due to marked follicular hyperplasia in the node in response to HIV infection. The nodes are generally discrete and freely movable. This feature of HIV disease may be seen at any point in the spectrum of immune dysfunction and is not associated with an increased likelihood of developing AIDS. Paradoxically, a loss in lymphadenopathy or a decrease in lymph node size outside the setting of ART may be a prognostic marker of disease progression. In patients with CD4+ T cell counts >200/μL, the differential diagnosis of lymphadenopathy includes TB, KS, Castleman's disease, and lymphoma. In patients with more advanced disease, lymphadenopathy may also be due to atypical mycobacterial infection, toxoplasmosis, systemic fungal infection, or bacillary angiomatosis. While indicated in patients with CD4+ T cell counts <200/μL, lymph node biopsy is not indicated in patients with early-stage disease unless there are signs and symptoms of systemic illness, such as fever and weight loss, or unless the nodes begin to enlarge, become fixed, or coalesce. Monoclonal gammopathy of unknown significance (MGUS) **(Chap. 116)**, defined as the presence of a serum monoclonal IgG, IgA, or IgM in the absence of a clear cause, has been reported in 3% of patients with HIV infection. The overall clinical significance of this finding in patients with HIV infection is unclear, although it has been associated with other viral infections, non-Hodgkin's lymphoma, and plasma cell malignancy.

Anemia is the most common hematologic abnormality in patients with HIV and, in the absence of a specific treatable cause, is independently associated with a poor prognosis. While generally mild, anemia can be quite severe and require chronic blood transfusions. Among the specific reversible causes of anemia in the setting of HIV infection are drug toxicity, systemic fungal and mycobacterial infections, nutritional deficiencies, and parvovirus B19 infections. The antiretroviral zidovudine may block erythroid maturation prior to its effects on other marrow elements. A characteristic feature of zidovudine therapy is an elevated mean corpuscular volume (MCV). Another drug used in patients with HIV infection that has a selective effect on the erythroid series is dapsone. This drug can cause a serious hemolytic anemia in patients who are deficient in glucose-6-phosphate dehydrogenase and can create a functional anemia in others through induction of methemoglobinemia. Folate levels are usually normal in people with HIV; however, vitamin B_{12} levels may be depressed as a consequence of achlorhydria or malabsorption. True autoimmune hemolytic anemia is rare, although ~20% of patients with HIV infection may have a positive direct antiglobulin test as a consequence of polyclonal B cell activation. Infection with parvovirus B19 also may cause anemia. It is important to recognize this possibility given the fact that it responds well to treatment with IVIg. Erythropoietin levels in patients with HIV infection and anemia are generally lower than expected given the degree of anemia. Treatment with erythropoietin may result in an increase in hemoglobin levels. An exception to this is a subset of patients with zidovudine-associated anemia in whom erythropoietin levels may be quite high.

During the course of HIV infection, neutropenia may be seen in approximately half of patients. In most instances it is mild; however, it can be severe and can put patients at risk of spontaneous bacterial infections. This is most frequently seen in patients with severely advanced HIV disease and in patients receiving potentially myelosuppressive therapies. In the setting of neutropenia, diseases not commonly seen in patients with HIV, such as aspergillosis or mucormycosis, may occur. Both granulocyte colony-stimulating factor (G-CSF) and GM-CSF increase neutrophil counts in patients with HIV infection regardless of the cause of the neutropenia. Earlier concerns about the potential of these agents to also increase levels of HIV were not confirmed in controlled clinical trials.

Thrombocytopenia may be an early consequence of HIV infection. Approximately 3% of patients with untreated HIV infection and CD4+ T cell counts ≥400/μL have platelet counts <150,000/μL. For untreated patients with CD4+ T cell counts <400/μL, this incidence increases to 10%. Thrombocytopenia is more common in patients with hepatitis C co-infection, cirrhosis, and/or ongoing high-level HIV replication. Thrombocytopenia is rarely a serious clinical problem in patients with HIV infection and generally responds well to successful ART. Clinically, it resembles the thrombocytopenia seen in patients with idiopathic thrombocytopenic purpura **(Chap. 120)**. Immune complexes containing anti-gp120 antibodies and anti-anti-gp120 antibodies have been noted in the circulation and on the surface of platelets in patients with HIV infection. Patients with HIV infection have also been noted to have a platelet-specific antibody directed toward a 25-kDa component of the surface of the platelet. Other data suggest that the thrombocytopenia in patients with HIV infection may be due to a direct effect of HIV on megakaryocytes. Whatever the cause, it is very clear that the most effective medical approach to this problem has been the use of ART. For patients with platelet counts <20,000/μL, a more aggressive approach combining IVIg or anti-Rh Ig for an immediate response and ART for a more lasting response is appropriate. Rituximab has been used with some success in otherwise refractory cases. Splenectomy is a rarely needed option and is reserved for patients refractory to medical management. Because of the risk of serious infection with encapsulated organisms, all patients with HIV infection about to undergo splenectomy should be immunized with vaccines to prevent disease from *S. pneumoniae*, *N. meningitidis*, and *H. influenzae* type b. It should be noted that, in addition to causing an increase in the platelet count, removal of the spleen will result in an increase in the peripheral blood lymphocyte count, making CD4+ T cell counts unreliable markers of immunocompetence. In this setting, the clinician should rely on the CD4+ T cell percentage for making diagnostic decisions with respect to the likelihood of opportunistic infections. A CD4+ T cell percentage of 15 is approximately equivalent to a CD4+ T cell count of 200/μL. In patients with early HIV infection, thrombocytopenia has also been reported as a consequence of classic thrombotic thrombocytopenic purpura **(Chap. 120)**. This clinical syndrome, consisting of fever, thrombocytopenia, hemolytic anemia, and neurologic and renal dysfunction, is a rare complication of early HIV infection. As in other settings, the appropriate management is the use of salicylates and plasma exchange. Other causes of thrombocytopenia include lymphoma, mycobacterial infections, and fungal infections.

The incidence of venous thromboembolic disease such as deep-vein thrombosis or pulmonary embolus is approximately 1% per year in patients with HIV infection. This is approximately 10 times higher than that seen in an age-matched population. Factors associated with an increased risk of clinical thrombosis include age >45, history of an opportunistic infection, lower CD4 count, and estrogen use. Abnormalities of the coagulation cascade, including decreased protein S activity, increases in factor VIII, anticardiolipin antibodies, PAR-1 expression on T cells, or lupus-like anticoagulant, have been reported in more than 50% of patients with HIV infection. The clinical significance of this increased propensity toward thromboembolic disease is likely reflected in the observation that elevations in D-dimer are strongly associated with all-cause mortality in patients with HIV infection (Table 208-9).

Dermatologic Diseases Dermatologic problems occur in >90% of patients with HIV infection. From the macular, roseola-like rash seen with the acute seroconversion syndrome to extensive end-stage KS, cutaneous manifestations of HIV disease can be seen throughout the course of HIV infection. Among the more common nonneoplastic problems are seborrheic dermatitis, folliculitis, and opportunistic infections. Extrapulmonary pneumocystosis may cause a necrotizing vasculitis. Neoplastic conditions are covered in a separate section below.

Seborrheic dermatitis occurs in 3% of the general population and in up to 50% of patients with HIV infection. Seborrheic dermatitis increases in prevalence and severity as the CD4+ T cell count declines. In patients with HIV, seborrheic dermatitis may be aggravated by concomitant infection with *Pityrosporum*, a yeastlike fungus; use of topical antifungal agents has been recommended in cases refractory to standard topical treatment.

Folliculitis is among the most prevalent dermatologic disorders in patients with HIV infection and is seen in ~20% of patients. It is more common in patients with CD4+ T cell counts <200 cells/μL. *Pruritic papular eruption* is one of the most common pruritic rashes in patients with HIV infection. It appears as multiple papules on the face, trunk, and extensor surfaces and may improve with ART. *Eosinophilic pustular folliculitis* is a rare form of folliculitis that is seen with increased frequency in patients with HIV infection. It presents as multiple, urticarial perifollicular papules that may coalesce into plaque-like lesions. Skin biopsy reveals an eosinophilic infiltrate of the hair follicle, which in certain cases has been associated with the presence of a mite. Patients typically have an elevated serum IgE level and may respond to treatment with topical anthelmintics. Pruritus is a common symptom in patients with HIV infection and can lead to *prurigo nodularis*. Patients with HIV infection have also been reported to develop a severe form of *Norwegian scabies* with hyperkeratotic psoriasiform lesions.

Both *psoriasis* and *ichthyosis*, although they are not reported to be increased in frequency, may be particularly severe when they occur in patients with HIV infection. Pre-existing psoriasis may become guttate in appearance and more refractory to treatment in the setting of HIV infection.

Reactivation herpes zoster (*shingles*) is seen in 10–20% of patients with HIV infection. This reactivation syndrome of varicella-zoster virus indicates a modest decline in immune function and may be the first indication of clinical immunodeficiency. In one series, patients who developed shingles did so an average of 5 years after HIV infection. In a cohort of patients with HIV infection and localized zoster, the subsequent rate of the development of AIDS was 1% per month. In that study, AIDS was more likely to develop if the outbreak of zoster was associated with severe pain, extensive skin involvement, or involvement of cranial or cervical dermatomes. The clinical manifestations of reactivation zoster in patients with HIV, although indicative of immunologic compromise, are not as severe as those seen in other immunodeficient conditions. Thus, while lesions may extend over several dermatomes, involve the spinal cord, and/or be associated with frank cutaneous dissemination, visceral involvement has not been reported. In contrast to patients without a known underlying immunodeficiency state, patients with HIV infection tend to have recurrences of shingles with a relapse rate of ~20%. Valacyclovir, famciclovir, or acyclovir is the treatment of choice. IV acyclovir can be used in severe cases, and foscarnet may be of value in patients with acyclovir-resistant virus.

Infection with *herpes simplex virus* in people with HIV is associated with recurrent orolabial, genital, and perianal lesions as part of recurrent reactivation syndromes **(Chap. 197)**. As HIV disease progresses and the CD4+ T cell count declines, these infections become more frequent and severe. Lesions often appear as beefy red, are exquisitely painful, and tend to occur high in the gluteal cleft (Fig. 208-38). Perirectal HSV may be associated with proctitis and anal fissures. HSV should be high in the differential diagnosis of any patient with HIV who has a poorly healing, painful perirectal lesion. In addition to recurrent mucosal ulcers, recurrent HSV infection in the form of *herpetic whitlow* can be a problem in patients with HIV infection, presenting with painful vesicles or extensive cutaneous erosion. Valacyclovir, famciclovir, or acyclovir is the treatment of choice in these settings. It is noteworthy that even subclinical reactivation of herpes simplex may be associated with increases in plasma HIV RNA levels.

Diffuse skin eruptions due to *Molluscum contagiosum* may be seen in patients with advanced HIV infection. These flesh-colored, umbilicated lesions resemble those of *Talaromyces (formerly Penicillium) marnefei* or *Cryptococcosis.* They tend to regress with effective ART and can also be treated with local therapy. Similarly, *condyloma acuminatum* lesions may be more severe and more widely distributed in patients with low CD4+ T cell counts. Imiquimod cream may be helpful in some cases. Atypical mycobacterial infections may present as erythematous cutaneous nodules, as may fungal infections, *Bartonella*, *Acanthamoeba*, and KS. Cutaneous infections with *Aspergillus* have been noted at the site of IV catheter placement.

The skin of patients with HIV infection is often a target organ for drug reactions **(Chap. 63)**. Although most skin reactions are mild and not necessarily an indication to discontinue therapy, some patients may have particularly severe cutaneous reactions to drugs, including erythroderma, *Stevens-Johnson syndrome*, and toxic epidermal necrolysis. This is particularly true for sulfa drugs, nonnucleoside reverse transcriptase inhibitors, abacavir, amprenavir, darunavir, fosamprenavir, and tipranavir. Similarly, patients with HIV infection are often quite photosensitive and burn easily following exposure to sunlight or as a side effect of radiation therapy **(Chap. 64)**.

HIV infection and its treatment may be accompanied by cosmetic changes of the skin that are not of great clinical importance but may be troubling to patients. Yellowing of the nails and straightening of the hair, particularly in African-American patients, have been reported as a consequence of HIV infection. Zidovudine therapy has been associated with elongation of the eyelashes and the development of a bluish discoloration to the nails, again more common in African-American patients. Therapy with clofazimine may cause a yellow-orange discoloration of the skin and urine.

Neurologic Diseases Clinical disease of the nervous system accounts for a significant degree of morbidity in a high percentage of patients with HIV infection **(Table 208-14)**. The neurologic problems that occur in people with HIV may be either primary to the pathogenic processes of HIV infection or secondary to opportunistic infections or neoplasms. Among the more frequent opportunistic diseases that involve the CNS are toxoplasmosis, cryptococcosis, progressive multifocal leukoencephalopathy, and primary CNS lymphoma. Other less common problems include mycobacterial infections; syphilis; and infection with CMV, herpes zoster, HTLV-1, *Trypanosoma cruzi*, or *Acanthamoeba*. Overall, secondary diseases of the CNS have been reported to occur in approximately one-third of patients with AIDS. These data antedate the widespread use of ART, and this frequency is considerably lower in patients with suppressed viral replication. Primary processes related to HIV infection of the nervous system are reminiscent of those seen with other lentiviruses, such as the maedi-visna virus of sheep.

Neurologic problems directly attributable to HIV occur throughout the course of infection and may be inflammatory, demyelinating, or degenerative in nature. The term *HIV-associated neurocognitive disorders* (HAND) is used to describe a spectrum of disorders that range from asymptomatic neurocognitive impairment (ANI) to minor neurocognitive disorder (MND) to clinically severe dementia. The most severe form, *HIV-associated dementia* (HAD), also referred to as the *AIDS dementia complex*, or *HIV encephalopathy*, is considered an AIDS-defining illness. Many HIV-infected patients have some neurologic problem during the course of their disease. Even in the setting of suppressive ART, approximately 50% of people with HIV can be shown to have mild to moderate neurocognitive impairment using sensitive neuropsychiatric testing. As noted in the section on pathogenesis, damage to the CNS may be a direct result of viral infection of the CNS macrophages or glial cells or may be secondary to the release of neurotoxins and potentially toxic cytokines such as IL-1β, TNF-α, IL-6, and TGF-β. It has been reported that people with HIV who have the E4 allele for apoE are at increased risk for AIDS encephalopathy and peripheral neuropathy. Virtually all patients with HIV infection have some degree of nervous system involvement with the virus. This is evidenced by the fact that CSF findings are abnormal in ~90% of untreated patients, even during the asymptomatic phase of HIV infection. CSF abnormalities include pleocytosis (50–65% of patients), detection of viral RNA (~75%), elevated CSF protein (35%), and evidence of intrathecal synthesis of anti-HIV antibodies (90%). It is important to point out that evidence of infection of the CNS with HIV does not imply impairment of cognitive function. The neurologic function of a person with HIV should be considered normal unless clinical signs and symptoms suggest otherwise.

TABLE 208-14 Neurologic Diseases in Patients with HIV Infection

HIV-1 infection	Opportunistic infections
Aseptic meningitis	Toxoplasmosis
HIV-associated neurocognitive disorders (HAND), including HIV encephalopathy/AIDS dementia complex	Cryptococcosis
Myelopathy	Progressive multifocal leukoencephalopathy
Vacuolar myelopathy	Cytomegalovirus
Pure sensory ataxia	Syphilis
Paresthesia/dysesthesia	*Mycobacterium tuberculosis*
Peripheral neuropathy	HTLV-1 infection
Acute inflammatory demyelinating polyneuropathy (Guillain-Barré syndrome)	Amebiasis
Chronic inflammatory demyelinating polyneuropathy (CIDP)	Neoplasms
Mononeuritis multiplex	Primary CNS lymphoma
Distal symmetric polyneuropathy	Kaposi's sarcoma
Myopathy	

Aseptic meningitis may occur at any time in the course of HIV infection; however, it is rare following the development of AIDS. This suggests that clinical aseptic meningitis in the context of HIV infection is an immune-mediated disease. In the setting of acute primary infection, patients may experience a syndrome of headache, photophobia, and meningismus. Rarely, an acute encephalopathy due to encephalitis may occur. Cranial nerve involvement may be seen, predominantly cranial nerve VII but occasionally V and/or VIII. CSF findings include a lymphocytic pleocytosis, elevated protein level, and normal glucose level. This syndrome, which cannot be clinically differentiated from other viral meningitides **(Chap. 143)**, usually resolves spontaneously within 2–4 weeks; however, in some patients, signs and symptoms may become chronic.

Fungal meningitis is the leading infectious cause of meningitis in patients with AIDS **(Chaps. 143 and 144)**. While the vast majority of these are due to *C. neoformans*, up to 12% may be due to *C. gattii*. Cryptococcal meningitis is the initial AIDS-defining illness in ~2% of patients and generally occurs in patients with CD4+ T cell counts <100/μL. Cryptococcal meningitis is particularly common in untreated patients with AIDS in Africa, occurring in ~5% of patients. Most patients present with a picture of subacute meningoencephalitis with fever, nausea, vomiting, altered mental status, headache, and meningeal signs. The incidence of seizures and focal neurologic deficits is low. The CSF profile may be normal or may show only modest elevations in WBC or protein levels and decreases in glucose. The opening pressure in the CSF is usually elevated. In addition to meningitis, patients may develop cryptococcomas and cranial nerve involvement. Approximately one-third of patients also have pulmonary disease. Uncommon manifestations of cryptococcal infection include skin lesions that resemble *molluscum contagiosum*, lymphadenopathy, palatal and glossal ulcers, arthritis, gastroenteritis, myocarditis, and prostatitis. The prostate gland may serve as a reservoir for smoldering cryptococcal infection. The diagnosis of cryptococcal meningitis is made by identification of organisms in spinal fluid with India ink examination or by the detection of cryptococcal antigen. Blood cultures for fungus are often positive. A biopsy may be needed to make a diagnosis of CNS cryptococcoma and to distinguish inadequately treated infection from immune reconstitution syndrome. Initial treatment is with IV amphotericin B 0.7 mg/kg daily, or liposomal amphotericin 3–4 mg/kg daily, with flucytosine 25 mg/kg qid for at least 2 weeks if possible. Decreases in renal function in association with amphotericin can lead to increases in flucytosine levels and subsequent bone marrow suppression. Therapy continues with amphotericin alone until the CSF culture turns negative followed by fluconazole 800 mg/d PO for 8 weeks, and then fluconazole 200 mg/d until the CD4+ T cell count has increased to >200 cells/μL for 6 months in response to ART. Repeated lumbar puncture may be required to manage increased intracranial pressure. Symptoms may recur with initiation of ART as an immune reconstitution syndrome (see above). For this reason, it is recommended that patients receive 4–6 weeks of antifungal therapy prior to initiation of ART. Other fungi that may cause meningitis in patients with HIV infection are *C. immitis* and *H. capsulatum*. Meningoencephalitis has also been reported due to *Acanthamoeba* or *Naegleria*.

HIV-associated dementia consists of a constellation of signs and symptoms of CNS disease. While this is generally a late complication of HIV infection that progresses slowly over months, it can be seen in patients with CD4+ T cell counts >350 cells/μL. A major feature of this entity is the development of dementia, defined as a decline in cognitive ability from a previous level. It may present as impaired ability to concentrate, increased forgetfulness, difficulty reading, or increased difficulty performing complex tasks. Initially these symptoms may be indistinguishable from findings of situational depression or fatigue. In contrast to "cortical" dementia (such as Alzheimer's disease), aphasia, apraxia, and agnosia are uncommon, leading some investigators to classify HIV encephalopathy as a "subcortical dementia" characterized by defects in short-term memory and executive function (see below). In addition to dementia, patients with HIV encephalopathy may also have motor and behavioral abnormalities. Among the motor problems are unsteady gait, poor balance, tremor, and difficulty with rapid alternating movements. Increased tone and deep tendon reflexes may be found in patients with spinal cord involvement. Late stages may be complicated by bowel and/or bladder incontinence. Behavioral problems include apathy, irritability, and lack of initiative, with progression to a vegetative state in some instances. Some patients develop a state of agitation or mild mania. These changes usually occur without significant changes in level of alertness. This contrasts with the finding of somnolence in patients with dementia due to toxic/metabolic encephalopathies.

HIV-associated dementia is the initial AIDS-defining illness in ~3% of patients with HIV infection and thus only rarely precedes clinical evidence of immunodeficiency. Clinically significant encephalopathy eventually develops in ~25% of untreated patients with AIDS. As immunologic function declines, the risk and severity of HIV-associated dementia increases. Autopsy series suggest that 80–90% of patients with HIV infection have histologic evidence of CNS involvement. Several classification schemes have been developed for grading HIV encephalopathy; a commonly used clinical staging system is outlined in **Table 208-15**.

The precise cause of HIV-associated dementia remains unclear, although the condition is thought to be a result of a combination of direct effects of HIV on the CNS and associated immune activation. HIV has been found in the brains of patients with HIV encephalopathy by Southern blot, in situ hybridization, PCR, and electron microscopy. Multinucleated giant cells, macrophages, and microglial cells appear to be the main cell types harboring virus in the CNS. Histologically, the major changes are seen in the subcortical areas of the brain and include pallor and gliosis, multinucleated giant cell encephalitis, and vacuolar myelopathy. Less commonly, diffuse or focal spongiform changes occur in the white matter. Areas of the brain involved in motor function, language, and judgment are most severely affected.

TABLE 208-15 Clinical Staging of HAND According to Frascati Criteria

STAGE	NEUROCOGNITIVE STATUS[a]	FUNCTIONAL STATUS[b]
Asymptomatic	1 SD below mean in 2 cognitive domains	No impairments in activities of daily living
Mild neurocognitive disorder	1 SD below mean in 2 cognitive domains	Impairments in activities of daily living
HIV-associated dementia	2 SD below mean in 2 cognitive domains	Notable impairments in activities of daily living

[a]Neurocognitive testing should include assessment of at least 5 domains, including attention/information processing, language, abstraction/executive function, complex perceptual motor skills, memory (including learning and recall), simple motor skills, or sensory perceptual skills. Appropriate norms must be available to establish the number of domains in which performance is below 1 SD. [b]Functional status is typically assessed by self-reporting but might be corroborated by a collateral source. No agreed measures exist for HIV-associated neurocognitive disorder criteria. Note that, for diagnosis of HIV-associated neurocognitive disorder, other causes of dementia must be ruled out and potential confounding effects of substance use or psychiatric illness should be considered.

Source: Adapted from A Antinori et al: Neurology 69:1789, 2007.

There are no specific criteria for a diagnosis of HIV-associated dementia, and this syndrome must be differentiated from other diseases that affect the CNS of patients with HIV (Table 208-14). The diagnosis of dementia depends on demonstrating a decline in cognitive function. This can be accomplished objectively with the use of a Mini-Mental State Examination (MMSE) in patients for whom prior scores are available. For this reason, it is advisable for all patients with a diagnosis of HIV infection to have a baseline MMSE. However, changes in MMSE scores may be absent in patients with mild HIV encephalopathy. Imaging studies of the CNS, by either MRI or CT, often demonstrate evidence of cerebral atrophy **(Fig. 208-40)**. MRI may also reveal small areas of increased density on T2-weighted images. Lumbar puncture is an important element of the evaluation of patients with

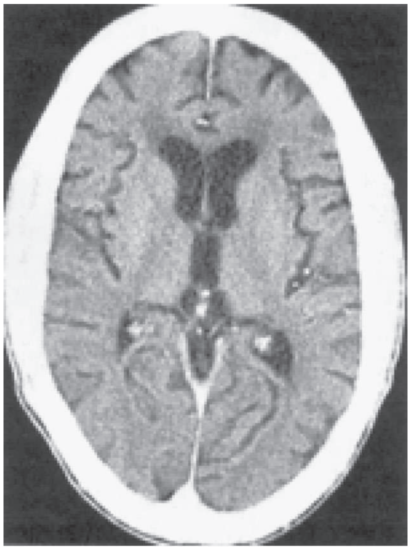

FIGURE 208-40 AIDS dementia complex. Postcontrast CT scan through the lateral ventricles of a 47-year-old man with AIDS, altered mental status, and dementia. The lateral and third ventricles and the cerebral sulci are abnormally prominent. Mild white matter hypodensity is seen adjacent to the frontal horns of the lateral ventricles.

HIV infection and neurologic abnormalities. It is generally most helpful in ruling out or making a diagnosis of opportunistic infections. In HIV encephalopathy, patients may have the nonspecific findings of an increase in CSF cells and protein level. While HIV RNA can often be detected in the spinal fluid and HIV can be cultured from the CSF, this finding is not specific for HIV encephalopathy. There appears to be no correlation between the presence of HIV in the CSF and the presence of HIV encephalopathy. Elevated levels of monocyte chemoattractant protein-1 (MCP-1), β_2-microglobulin, neopterin, and quinolinic acid (a metabolite of tryptophan reported to cause CNS injury) have been noted in the CSF of patients with HIV encephalopathy. These findings suggest that these factors as well as inflammatory cytokines may be involved in the pathogenesis of this syndrome.

Combination antiretroviral therapy is of benefit in patients with HIV-associated dementia. Improvement in neuropsychiatric test scores has been noted for both adult and pediatric patients treated with antiretrovirals. The rapid improvement in cognitive function noted with the initiation of ART suggests that at least some component of this problem is quickly reversible, again supporting at least a partial role of soluble mediators in the pathogenesis. It should also be noted that these patients have an increased sensitivity to the side effects of neuroleptic drugs. The use of these drugs for symptomatic treatment is associated with an increased risk of extrapyramidal side effects; therefore, patients with HIV encephalopathy who receive these agents must be monitored carefully. It is felt by many physicians that the decrease in the prevalence of severe cases of HAND brought about by ART has resulted in an increase in the prevalence of milder forms of this disorder.

Seizures may be a consequence of opportunistic infections, neoplasms, or HIV encephalopathy **(Table 208-16)**. The seizure threshold is often lower than normal in patients with advanced HIV infection due in part to the frequent presence of electrolyte abnormalities. Seizures are seen in 15–40% of patients with cerebral toxoplasmosis, 15–35% of patients with primary CNS lymphoma, 8% of patients with cryptococcal meningitis, and 7–50% of patients with HIV encephalopathy. Seizures may also be seen in patients with CNS tuberculosis, aseptic meningitis, and progressive multifocal leukoencephalopathy. Seizures may be the presenting clinical symptom of HIV disease. In one study of 100 patients with HIV infection presenting with a first seizure, cerebral mass lesions were the most common cause, responsible for 32 of the 100 new-onset seizures. Of these 32 cases, 28 were due to toxoplasmosis and 4 to lymphoma. HIV encephalopathy accounted for an additional 24 new-onset seizures. Cryptococcal meningitis was the third most common diagnosis, responsible for 13 of the 100 seizures. In 23 cases, no cause could be found, and it is possible that these cases represent a subcategory of HIV encephalopathy. Of these 23 cases, 16 (70%) had 2 or more seizures, suggesting that anticonvulsant therapy is indicated in all patients with HIV infection and seizures unless a rapidly correctable cause is found. Due to a variety of drug–drug interactions between antiseizure medications and antiretrovirals, drug levels need to be monitored carefully.

Patients with HIV infection may present with *focal neurologic deficits* from a variety of causes. The most common causes are toxoplasmosis, progressive multifocal leukoencephalopathy, and CNS lymphoma. Other causes include cryptococcal infections (discussed above; also **Chap. 221**), stroke, and reactivation of Chagas' disease.

TABLE 208-16 Causes of Seizures in Patients with HIV Infection

DISEASE	OVERALL CONTRIBUTION TO FIRST SEIZURE, %	FRACTION OF PATIENTS WHO HAVE SEIZURES, %
HIV encephalopathy	24–47	7–50
Cerebral toxoplasmosis	28	15–40
Cryptococcal meningitis	13	8
Primary central nervous system lymphoma	4	15–30
Progressive multifocal leukoencephalopathy	1	20

Source: From DM Holtzman et al: Am J Med 87:173, 1989.

Toxoplasmosis **(Chap. 235)** has been one of the most common causes of secondary CNS infections in patients with AIDS, but its incidence is decreasing in the era of ART. It is most common in patients from the Caribbean and from France, where the seroprevalence of *T. gondii* is around 50%. This figure is closer to 15% in the United States. Toxoplasmosis is generally a late complication of HIV infection and usually occurs in patients with CD4+ T cell counts <200/μL. Cerebral toxoplasmosis is thought to represent a reactivation of latent tissue cysts. It is 10 times more common in patients with antibodies to the organism than in patients who are seronegative. Patients diagnosed with HIV infection should be screened for IgG antibodies to *T. gondii* during the time of their initial workup. Those who are seronegative should be counseled about ways to minimize the risk of primary infection including avoiding the consumption of undercooked meat and careful hand washing after contact with soil or changing the cat litter box. The most common clinical presentation of cerebral toxoplasmosis in patients with HIV infection is fever, headache, and focal neurologic deficits. Patients may present with seizure, hemiparesis, or aphasia as a manifestation of these focal deficits or with a picture more influenced by the accompanying cerebral edema and characterized by confusion, dementia, and lethargy, which can progress to coma. The diagnosis is usually suspected on the basis of MRI findings of multiple lesions in multiple locations, although in some cases only a single lesion is seen. Pathologically, these lesions generally exhibit inflammation and central necrosis and, as a result, demonstrate ring enhancement on contrast MRI **(Fig. 208-41)** or, if MRI is unavailable or contraindicated, on double-dose contrast CT. There is usually evidence of surrounding edema. In addition to toxoplasmosis, the differential diagnosis of single or multiple enhancing mass lesions in the patient with HIV includes primary CNS lymphoma and, less commonly, TB or fungal or bacterial abscesses. The definitive diagnostic procedure is brain biopsy. However, given the morbidity rate that can accompany this procedure, it is usually reserved for the patient who has failed 2–4 weeks of empiric therapy for toxoplasmosis. If the patient is seronegative for *T. gondii*, the likelihood that a mass lesion is due to toxoplasmosis is <10%. In that setting, one may choose to be more aggressive and perform a brain biopsy sooner. Standard treatment is sulfadiazine and pyrimethamine with leucovorin as needed for a minimum of 4–6 weeks. Alternative therapeutic regimens include clindamycin in combination with pyrimethamine; atovaquone plus pyrimethamine; and azithromycin plus pyrimethamine plus rifabutin. Relapses are common, and it is recommended that patients with a history of prior toxoplasmic encephalitis receive maintenance therapy with sulfadiazine, pyrimethamine, and leucovorin as long as their CD4+ T cell counts remain <200 cells/μL. Patients with CD4+ T cell counts <100/μL and IgG antibody to *Toxoplasma* should receive primary prophylaxis for toxoplasmosis.

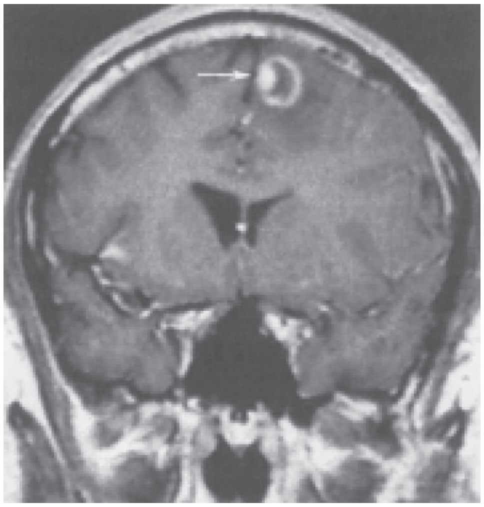

FIGURE 208-41 Central nervous system toxoplasmosis. A coronal postcontrast T1-weighted MRI scan demonstrates a peripheral enhancing lesion in the left frontal lobe, associated with an eccentric nodular area of enhancement (*arrow*); this so-called eccentric target sign is typical of toxoplasmosis.

Fortunately, the same daily regimen of a single double-strength tablet of TMP-SMX used for *P. jirovecii* prophylaxis provides adequate primary protection against toxoplasmosis. Secondary prophylaxis/maintenance therapy for toxoplasmosis may be discontinued in the setting of effective ART and increases in CD4+ T cell counts to >200/μL for 6 months.

JC virus, a human polyomavirus that is the etiologic agent of *progressive multifocal leukoencephalopathy* (PML), is an important opportunistic pathogen in patients with AIDS **(Chap. 142)**. While ~80% of the general adult population has antibodies to JC virus, indicative of prior infection, <10% of healthy adults show any evidence of ongoing viral replication. PML is the only known clinical manifestation of JC virus infection. It is a late manifestation of AIDS and is seen in ~1–4% of patients with AIDS. The lesions of PML begin as small foci of demyelination in subcortical white matter that eventually coalesce. The cerebral hemispheres, cerebellum, and brainstem may all be involved. Patients typically have a protracted course with multifocal neurologic deficits, with or without changes in mental status. Approximately 20% of patients experience seizures. Ataxia, hemiparesis, visual field defects, aphasia, and sensory defects may occur. Headache, fever, nausea, and vomiting are rarely seen. Their presence should suggest another diagnosis. MRI typically reveals multiple, nonenhancing white matter lesions that may coalesce and have a predilection for the occipital and parietal lobes. The lesions show signal hyperintensity on T2-weighted images and diminished signal on T1-weighted images. The measurement of JC virus DNA levels in CSF has a diagnostic sensitivity of 76% and a specificity of close to 100%. Prior to the availability of ART, most patients with PML died within 3–6 months of the onset of symptoms. Paradoxical worsening of PML has been seen with initiation of ART as an immune reconstitution syndrome. There is no specific treatment for PML; however, a median survival of 2 years and survival of >15 years have been reported in patients with PML treated with ART for their HIV disease. Despite having a significant impact on survival, only ~50% of patients with HIV infection and PML show neurologic improvement with ART. Studies with other antiviral agents such as cidofovir have failed to show clear benefit. Factors influencing a favorable prognosis for PML in the setting of HIV infection include a CD4+ T cell count >100/μL at baseline and the ability to maintain an HIV viral load of <500 copies/mL. Baseline HIV-1 viral load does not have independent predictive value of survival. PML is one of the few opportunistic infections that continues to occur with some frequency despite the widespread use of ART.

Reactivation American trypanosomiasis may present as acute meningoencephalitis with focal neurologic signs, fever, headache, vomiting, and seizures. Accompanying cardiac disease in the form of arrhythmias or heart failure should increase the index of suspicion. The presence of antibodies to *T. cruzi* supports the diagnosis. In South America, reactivation of *Chagas' disease* is considered to be an AIDS-defining condition and may be the initial AIDS-defining condition. Most cases occur in patients with CD4+ T cell counts <200 cells/μL. Lesions appear radiographically as single or multiple hypodense areas, typically with ring enhancement and edema. They are found predominantly in the subcortical areas, a feature that differentiates them from the deeper lesions of toxoplasmosis. *T. cruzi* amastigotes, or trypanosomes, can be identified from biopsy specimens or CSF. Other CSF findings include elevated protein and a mild (<100 cells/μL) lymphocytic pleocytosis. Organisms can also be identified by direct examination of the blood. Treatment consists of benzimidazole (2.5 mg/kg bid) or nifurtimox (2 mg/kg qid) for at least 60 days, followed by maintenance therapy for the duration of immunodeficiency with either drug at a dose of 5 mg/kg three times a week. As is the case with cerebral toxoplasmosis, successful therapy with antiretrovirals may allow discontinuation of maintenance therapy for Chagas' disease.

Stroke may occur in patients with HIV infection. In contrast to the other causes of focal neurologic deficits in patients with HIV infection, the symptoms of a stroke are sudden in onset. Patients with HIV infection have an increased prevalence of many classic risk factors associated with stroke, including smoking and diabetes. It has been reported that HIV infection itself can lead to an increase in carotid artery stiffness. The relative increase in risk for stroke as a consequence of HIV infection is more pronounced in women and in individuals between the ages of 18 and 29. Among the secondary infectious diseases in patients with HIV infection that may be associated with stroke are vasculitis due to cerebral varicella zoster or neurosyphilis and septic embolism in association with fungal infection. Other elements of the differential diagnosis of stroke in the patient with HIV infection include atherosclerotic cerebral vascular disease, thrombotic thrombocytopenic purpura, and cocaine or amphetamine use.

Primary CNS lymphoma is discussed below in the section on neoplastic diseases.

Spinal cord disease, or myelopathy, is present in ~20% of patients with AIDS, often as part of HIV-associated neurocognitive disorder. In fact, 90% of the patients with HIV-associated myelopathy have some evidence of dementia, suggesting that similar pathologic processes may be responsible for both conditions. Three main types of spinal cord disease are seen in patients with AIDS. The first of these is a vacuolar myelopathy, as mentioned above. This condition is pathologically similar to subacute combined degeneration of the cord, such as that occurring with pernicious anemia. Although vitamin B_{12} deficiency can be seen in patients with AIDS as a primary complication of HIV infection, it does not appear to be responsible for most cases of myelopathy seen in patients with HIV infection. However, it should be included in the differential diagnosis. Vacuolar myelopathy is characterized by a subacute onset and often presents with gait disturbances, predominantly ataxia and spasticity; it may progress to include bladder and bowel dysfunction. Physical findings include evidence of increased deep tendon reflexes and extensor plantar responses. The second form of spinal cord disease involves the dorsal columns and presents as a pure sensory ataxia. The third form is also sensory in nature and presents with paresthesias and dysesthesias of the lower extremities. In contrast to the cognitive problems seen in patients with HIV encephalopathy, these spinal cord syndromes do not respond well to antiretroviral drugs, and therapy is mainly supportive.

One important disease of the spinal cord that also involves the peripheral nerves is a *myelopathy* and *polyradiculopathy* seen in association with CMV infection. This entity is generally seen late in the course of HIV infection and is fulminant in onset, with lower extremity and sacral paresthesias, difficulty in walking, areflexia, ascending sensory loss, and urinary retention. The clinical course is rapidly progressive over a period of weeks. CSF examination reveals a predominantly neutrophilic pleocytosis, and CMV DNA can be detected by CSF PCR. Intravenous therapy with ganciclovir or foscarnet can lead to rapid improvement, and prompt initiation of therapy is important in minimizing the degree of permanent neurologic damage. Combination therapy with both drugs should be considered in patients who have been previously treated for CMV disease. Other diseases involving the spinal cord in patients with HIV infection include HTLV-1-associated myelopathy (HAM) **(Chap. 207)**, neurosyphilis **(Chap. 187)**, infection with herpes simplex **(Chap. 197)** or varicella-zoster **(Chap. 198)**, TB **(Chap. 183)**, and lymphoma **(Chap. 113)**.

Peripheral neuropathies are common in patients with HIV infection. They occur at all stages of illness and take a variety of forms. Early in the course of HIV infection, an acute inflammatory demyelinating polyneuropathy resembling Guillain-Barré syndrome may occur **(Chap. 458)**. In other patients, a progressive or relapsing-remitting inflammatory neuropathy resembling chronic inflammatory demyelinating polyneuropathy (CIDP) has been noted. Patients commonly present with progressive weakness, areflexia, and minimal sensory changes. CSF examination often reveals a mononuclear pleocytosis, and peripheral nerve biopsy demonstrates a perivascular infiltrate suggesting an autoimmune etiology. Plasma exchange or IVIg has been tried with variable success. Because of the immunosuppressive effects of glucocorticoids, they should be reserved for severe cases of CIDP refractory to other measures. Another autoimmune peripheral neuropathy seen in patients with AIDS is mononeuritis multiplex **(Chap. 457)** due to a necrotizing arteritis of peripheral nerves. The most common peripheral neuropathy in patients with HIV infection is a *distal sensory polyneuropathy* (DSPN) also referred to as painful sensory neuropathy

(HIV-SN), predominantly sensory neuropathy, or distal symmetric peripheral neuropathy. This condition may be a direct consequence of HIV infection or a side effect of ART with dideoxynucleosides. It is more common in taller individuals, older individuals, and those with lower CD4 counts. Two-thirds of patients with AIDS may be shown by electrophysiologic studies to have some evidence of peripheral nerve disease. Presenting symptoms are usually painful burning sensations in the feet and lower extremities. Findings on examination include a stocking-type sensory loss to pinprick, temperature, and touch sensation and a loss of ankle reflexes. Motor changes are mild and are usually limited to weakness of the intrinsic foot muscles. Response of this condition to antiretrovirals has been variable, perhaps because antiretrovirals are responsible for the problem in some instances. When due to dideoxynucleoside therapy, patients with lower extremity peripheral neuropathy may complain of a sensation that they are walking on ice. Other entities in the differential diagnosis of peripheral neuropathy include diabetes mellitus, vitamin B_{12} deficiency, and side effects from metronidazole or dapsone. For distal symmetric polyneuropathy that fails to resolve following the discontinuation of dideoxynucleosides, therapy is symptomatic; gabapentin, carbamazepine, tricyclics, or analgesics may be effective for dysesthesias. Treatment-naïve patients may respond to ART.

Myopathy may complicate the course of HIV infection; causes include HIV infection itself, zidovudine, and the generalized wasting syndrome (discussed below). HIV-associated myopathy may range in severity from an asymptomatic elevation in creatine kinase levels to a subacute syndrome characterized by proximal muscle weakness and myalgias. Quite pronounced elevations in creatine kinase may occur in asymptomatic patients, particularly after exercise. The clinical significance of this as an isolated laboratory finding is unclear. A variety of both inflammatory and noninflammatory pathologic processes have been noted in patients with more severe myopathy, including myofiber necrosis with inflammatory cells, nemaline rod bodies, cytoplasmic bodies, and mitochondrial abnormalities. Profound muscle wasting, often with muscle pain, may be seen after prolonged zidovudine therapy. This toxic side effect of the drug is dose-dependent and is related to its ability to interfere with the function of mitochondrial polymerases. It is reversible following discontinuation of the drug. Red ragged fibers are a histologic hallmark of zidovudine-induced myopathy.

Ophthalmologic Diseases Ophthalmologic problems occur in ~50% of patients with advanced HIV infection. The most common abnormal findings on funduscopic examination are cotton-wool spots. These are hard white spots that appear on the surface of the retina and often have an irregular edge. They represent areas of retinal ischemia secondary to microvascular disease. At times they are associated with small areas of hemorrhage and thus can be difficult to distinguish from CMV retinitis. In contrast to CMV retinitis, however, these lesions are not associated with visual loss and tend to remain stable or improve over time.

One of the most devastating consequences of HIV infection is CMV retinitis. Patients at high risk of CMV retinitis (CD4+ T cell count <100/μL) should undergo an ophthalmologic examination every 3–6 months. The majority of cases of CMV retinitis occur in patients with a CD4+ T cell count <50/μL. Prior to the availability of ART, this CMV reactivation syndrome was seen in 25–30% of patients with AIDS. In the ART era this has dropped to close to 2%. CMV retinitis usually presents as a painless, progressive loss of vision. Patients may also complain of blurred vision, "floaters," and scintillations. The disease is usually bilateral, although typically it affects one eye more than the other. The diagnosis is made on clinical grounds by an experienced ophthalmologist. The characteristic retinal appearance is that of perivascular hemorrhage and exudate. In situations where the diagnosis is in doubt due to an atypical presentation or an unexpected lack of response to therapy, vitreous or aqueous humor sampling with molecular diagnostic techniques may be of value. CMV infection of the retina results in a necrotic inflammatory process, and the visual loss that develops is irreversible. CMV retinitis may be complicated by rhegmatogenous retinal detachment as a consequence of retinal atrophy in areas of prior inflammation. Therapy for CMV retinitis consists of oral valganciclovir, IV ganciclovir, or IV foscarnet, with cidofovir as an alternative. Combination therapy with ganciclovir and foscarnet has been shown to be slightly more effective than either ganciclovir or foscarnet alone in the patient with relapsed CMV retinitis. A 3-week induction course is followed by maintenance therapy with oral valganciclovir. If CMV disease is limited to the eye, intravitreal injections of ganciclovir or foscarnet may be considered. Intravitreal injections of cidofovir are generally avoided due to the increased risk of uveitis and hypotony. Maintenance therapy is continued until the CD4+ T cell count remains >100 μL for >6 months. The majority of patients with HIV infection and CMV disease develop some degree of uveitis with the initiation of ART. The etiology of this is unknown; however, it has been suggested that this may be due to the generation of an enhanced immune response to CMV as an IRIS (see above). In some instances, this has required the use of topical glucocorticoids.

Both HSV and varicella zoster virus can cause a rapidly progressing, bilateral, necrotizing retinitis referred to as the *acute retinal necrosis syndrome*, or *progressive outer retinal necrosis* (PORN). This syndrome, in contrast to CMV retinitis, is associated with pain, keratitis, and iritis. It is often associated with orolabial HSV or trigeminal zoster. Ophthalmologic examination reveals widespread pale gray peripheral lesions. This condition is often complicated by retinal detachment. It is important to recognize and treat this condition with IV ganciclovir or IV acyclovir (if definitely due to HSV) as quickly as possible to minimize the loss of vision.

Several other secondary infections may cause ocular problems in patients with HIV. *P. jirovecii* can cause a lesion of the choroid that may be detected as an incidental finding on ophthalmologic examination. These lesions are typically bilateral, are from half to twice the disc diameter in size, and appear as slightly elevated yellow-white plaques. They are usually asymptomatic and may be confused with cotton-wool spots. Chorioretinitis due to toxoplasmosis can be seen alone or, more commonly, in association with CNS toxoplasmosis. KS may involve the eyelid or conjunctiva, while lymphoma may involve the retina. Syphilis may lead to a uveitis that is highly associated with the presence of neurosyphilis.

Additional Disseminated Infections and Wasting Syndrome Infections with species of the small, gram-negative, *Rickettsia*-like organism *Bartonella* **(Chap. 177)** are seen with increased frequency in patients with HIV infection. While it is not considered an AIDS-defining illness by the CDC, many experts view infection with *Bartonella* as indicative of a severe defect in cell-mediated immunity. It is usually seen in patients with CD4+ T cell counts <100/μL and is a significant cause of unexplained fever in patients with advanced HIV infection. Among the clinical manifestations of *Bartonella* infection are bacillary angiomatosis, cat-scratch disease, and trench fever. *Bacillary angiomatosis* is usually due to infection with *B. henselae* and is linked to exposure to flea-infested cats. It is characterized by a vascular proliferation that leads to a variety of skin lesions that have been confused with the skin lesions of KS. In contrast to the lesions of KS, the lesions of bacillary angiomatosis generally blanch, are painful, and typically occur in the setting of systemic symptoms. Infection can extend to the lymph nodes, liver (peliosis hepatis), spleen, bone, heart, CNS, respiratory tract, and GI tract. *Cat-scratch disease* is also due to infection with *B. henselae* and generally begins with a papule at the site of inoculation. This is followed several weeks later by the development of regional adenopathy and malaise. Infection with *B. quintana* is transmitted by lice and has been associated with case reports of trench fever, endocarditis, adenopathy, and bacillary angiomatosis. The organism is quite difficult to culture, and diagnosis often relies on identifying the organism in biopsy specimens using the Warthin-Starry or similar stains, PCR, and/or seroconversion. Treatment is with either doxycycline or erythromycin for at least 3 months.

Histoplasmosis is an opportunistic infection that is seen most frequently in patients in the Mississippi and Ohio River valleys, Puerto Rico, the Dominican Republic, and South America. These are all areas in which infection with *H. capsulatum* is endemic **(Chap. 218)**.

Because of this limited geographic distribution, histoplasmosis is only seen in approximately 0.5% of AIDS cases in the United States. Histoplasmosis is generally a late manifestation of HIV infection; however, it may be the initial AIDS-defining condition. In one study, the median CD4+ T cell count for patients with histoplasmosis and AIDS was 33/μL. While disease due to *H. capsulatum* may present as a primary infection of the lung, disseminated disease, presumably due to reactivation, is the most common presentation in patients with HIV. Patients usually present with a 4- to 8-week history of fever and weight loss. Hepatosplenomegaly and lymphadenopathy are each seen in about 25% of patients. CNS disease, either meningitis or a mass lesion, is seen in 15% of patients. Bone marrow involvement is common, with thrombocytopenia, neutropenia, and anemia occurring in 33% of patients. Approximately 7% of patients have mucocutaneous lesions consisting of a maculopapular rash and skin or oral ulcers. Respiratory symptoms are usually mild, with chest x-ray showing a diffuse infiltrate or diffuse small nodules in ~50% of cases. The gastrointestinal tract may be involved. Diagnosis is made by silver staining of tissue, by culturing the organisms from blood, bone marrow, or tissue, or by detecting antigen in blood or urine. Treatment is typically with liposomal amphotericin B followed by maintenance therapy with oral itraconazole until the serum *Histoplasma* antigen is <2 units, the patient has been on antiretrovirals for at least 6 months, and the CD4 count is >150 cells/μL. In the setting of mild infection, it may be appropriate to initiate therapy with itraconazole alone.

Following the spread of HIV infection to southeast Asia, disseminated infection with the fungus *Talaromyces (formerly Penicillium) marneffei* was recognized as a complication of HIV infection and is considered an AIDS-defining condition in those parts of the world where it occurs. *T. marneffei* is the third most common AIDS-defining illness in Thailand, following TB and cryptococcosis. It is more frequently diagnosed in the rainy than the dry season. Clinical features include fever, generalized lymphadenopathy, hepatosplenomegaly, anemia, thrombocytopenia, and papular skin lesions with central umbilication resembling the lesions of *Molluscum contagiosum*. Treatment is with amphotericin B followed by itraconazole until the CD4+ T cell count is >100 cells/μL for at least 6 months.

Visceral leishmaniasis **(Chap. 233)** is recognized with increasing frequency in patients with HIV infection who live in or travel to areas endemic for this protozoal infection transmitted by sandflies. The clinical presentation is one of hepatosplenomegaly, fever, and hematologic abnormalities. Lymphadenopathy and other constitutional symptoms may be present. A chronic, relapsing course is seen in two-thirds of co-infected patients. Organisms can be detected by PCR and, with special techniques, isolated from cultures of bone marrow aspirates. Histologic stains are often diagnostic but may be negative. Antibody titers are of little help. Patients with HIV infection usually respond well initially to standard therapy with amphotericin B or pentavalent antimony compounds. Eradication of the organism is difficult, however, and relapses are common.

Patients with HIV infection are at a slightly increased risk of infection with malaria and of clinical malaria. This is particularly true for patients from nonendemic areas who are at risk for primary infection and in patients with lower CD4+ T cell counts. HIV-positive individuals with CD4+ T cell counts <300 cells/μL have a poorer response to malaria treatment than others. Co-infection with malaria is associated with a modest increase in HIV viral load. The risk of malaria may be decreased with TMP-SMX prophylaxis.

Generalized wasting is an AIDS-defining condition; it is defined as involuntary weight loss of >10% associated with intermittent or constant fever and chronic diarrhea or fatigue lasting >30 days in the absence of a defined cause other than HIV infection. Prior to the widespread use of ART it was the initial AIDS-defining condition in ~10% of patients with AIDS in the United States. Generalized wasting is rarely seen today with the earlier initiation of antiretrovirals. A constant feature of this syndrome is severe muscle wasting with scattered myofiber degeneration and occasional evidence of myositis. Glucocorticoids may be of some benefit; however, this approach must be carefully weighed against the risk of compounding the immunodeficiency of HIV infection. Androgenic steroids, growth hormone, and total parenteral nutrition have been used as therapeutic interventions with variable success.

Neoplastic Diseases The neoplastic diseases considered to be AIDS-defining conditions are Kaposi's sarcoma, non-Hodgkin's lymphoma, and invasive cervical carcinoma. In addition, there is also an increase in the incidence of a variety of non-AIDS-defining malignancies including Hodgkin's disease; multiple myeloma; leukemia; melanoma; and cervical, brain, testicular, oral, lung, gastric, liver, renal, and anal cancers. Since the introduction of potent ART, there has been a marked reduction in the incidence of KS (Fig. 208-34). The non-AIDS-defining malignancies now account for more morbidity and mortality in patients with HIV infection than the AIDS-defining malignancies and are responsible for approximately 10% of the deaths in patients with HIV infection. Rates of non-Hodgkin's lymphoma have declined; however, this decline has not been as dramatic as the decline in rates of KS. In contrast, ART has had little effect on human papillomavirus (HPV)-associated malignancies. As patients with HIV infection live longer, a wider array of cancers is seen in this population. While some may only reflect known risk factors (e.g., smoking, alcohol consumption, co-infection with other viruses such as hepatitis B) that are increased in patients with HIV infection, some may be a direct consequence of HIV and are clearly increased in patients with lower CD4+ T cell counts.

Kaposi's sarcoma is a multicentric neoplasm presenting as multiple vascular nodules in the skin, mucous membranes, and viscera. The clinical course of KS ranges from indolent, with only minor skin or lymph node involvement, to fulminant, with extensive cutaneous and visceral involvement. In the initial period of the AIDS epidemic, KS was a prominent clinical feature of the first cases of AIDS, occurring in 79% of the patients diagnosed in 1981. By 1989 it was seen in only 25% of cases, by 1992 the number had decreased to 9%, and by 1997 the number was <1%. HHV-8 (KSHV) has been strongly implicated as a viral cofactor in the pathogenesis of KS.

Clinically, KS has varied presentations and may be seen at any stage of HIV infection, even in the presence of a normal CD4+ T cell count. The initial lesion may be a small, raised, reddish-purple nodule on the skin **(Fig. 208-42)**, a discoloration on the oral mucosa (Fig. 208-35*D*),

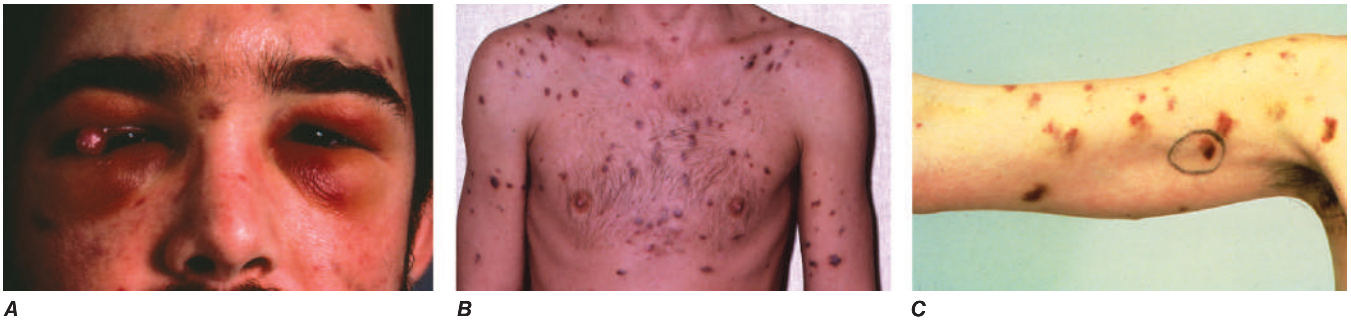

FIGURE 208-42 Kaposi's sarcoma in three patients with AIDS demonstrating (***A***) periorbital edema and bruising; (***B***) classic truncal distribution of lesions; and (***C***) upper extremity lesions.

or a swollen lymph node. Lesions often appear in sun-exposed areas, particularly the tip of the nose, and have a propensity to occur in areas of trauma (Koebner phenomenon). Because of the vascular nature of the tumors and the presence of extravasated red blood cells in the lesions, their colors range from reddish to purple to brown and often take the appearance of a bruise, with yellowish discoloration and tattooing. Lesions range in size from a few millimeters to several centimeters in diameter and may be either discrete or confluent. KS lesions most commonly appear as raised macules; however, they can also be papular, particularly in patients with higher CD4+ T cell counts. Confluent lesions may give rise to surrounding lymphedema and may be disfiguring when they involve the face and disabling when they involve the lower extremities or the surfaces of joints. Apart from skin, the lymph nodes, GI tract, and lung are the organ systems most commonly affected by KS. Lesions have been reported in virtually every organ, including the heart and the CNS. In contrast to most malignancies, in which lymph node involvement implies metastatic spread and a poor prognosis, lymph node involvement may be seen very early in KS and is of no special clinical significance. In fact, some patients may present with disease limited to the lymph nodes. These are generally patients with relatively intact immune function and thus the patients with the best prognosis. Pulmonary involvement with KS generally presents with shortness of breath. Some 80% of patients with pulmonary KS also have cutaneous lesions. The chest x-ray characteristically shows bilateral lower lobe infiltrates that obscure the margins of the mediastinum and diaphragm **(Fig. 208-43)**. Pleural effusions are seen in 70% of cases of pulmonary KS, a fact that is often helpful in the differential diagnosis. GI involvement is seen in 50% of patients with KS and usually takes one of two forms: (1) mucosal involvement, which may lead to bleeding that can be severe; these patients sometimes also develop symptoms of GI obstruction if lesions become large; and (2) biliary tract involvement. KS lesions may infiltrate the gallbladder and biliary tree, leading to a clinical picture of obstructive jaundice similar to that seen with sclerosing cholangitis. Several staging systems have been proposed for KS. One in common use was developed by the National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group; it distinguishes patients on the basis of tumor extent, immunologic function, and presence or absence of systemic symptoms **(Table 208-17)**.

A diagnosis of KS is based on biopsy of a suspicious lesion. Histologically one sees a proliferation of spindle cells and endothelial cells, extravasation of red blood cells, hemosiderin-laden macrophages, and, in early cases, an inflammatory cell infiltrate. Included in the differential diagnosis are lymphoma (particularly for oral lesions), bacillary angiomatosis, and cutaneous mycobacterial infections.

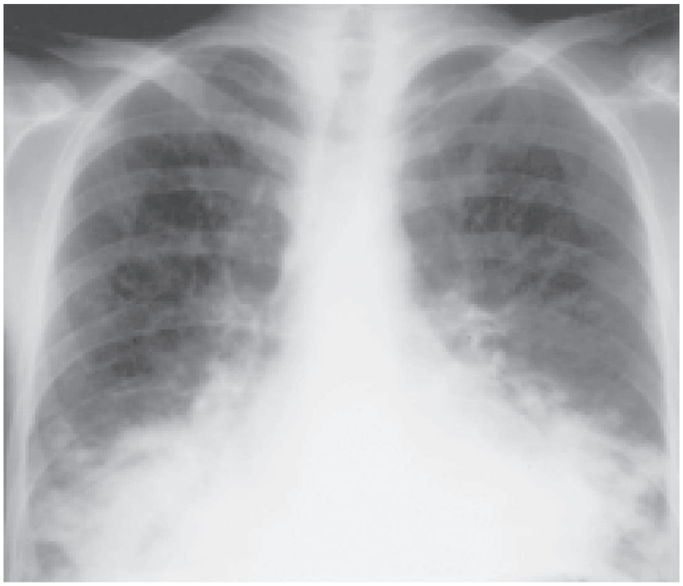

FIGURE 208-43 Chest x-ray of a patient with AIDS and pulmonary Kaposi's sarcoma. The characteristic findings include dense bilateral lower lobe infiltrates obscuring the heart borders and pleural effusions.

TABLE 208-17 National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group TIS Staging System for Kaposi's Sarcoma

PARAMETER	GOOD RISK (STAGE 0): ALL OF THE FOLLOWING	POOR RISK (STAGE 1): ANY OF THE FOLLOWING
Tumor (T)	Confined to skin and/or lymph nodes and/or minimal oral disease	Tumor-associated edema or ulceration Extensive oral lesions GI lesions Nonnodal visceral lesions
Immune system (I)	CD4+ T cell count ≥200/μL	CD4+ T cell count <200/μL
Systemic illness (S)	No B symptoms[a]	B symptoms[a] present
	Karnofsky performance status ≥70	Karnofsky performance status <70
	No history of opportunistic infection, neurologic disease, lymphoma, or thrush	History of opportunistic infection, neurologic disease, lymphoma, or thrush

[a] Defined as unexplained fever, night sweats, >10% involuntary weight loss, or diarrhea persisting for more than 2 weeks.

Management of KS **(Table 208-18)** should be conducted in consultation with an expert since definitive treatment guidelines do not exist. In the majority of cases, effective ART will go a long way in achieving control. ART has been associated with the spontaneous regression of KS lesions. Paradoxically, it has also been associated with the initial appearance of KS as a form of IRIS. For patients in whom tumor persists or is compromising vital functions or in whom control of HIV replication is not possible, a variety of options exist. In some cases, lesions remain quite indolent, and many of these patients can be managed with no specific treatment. Fewer than 10% of AIDS patients with KS die as a consequence of their malignancy, and death from secondary infections is considerably more common. Thus, whenever possible one should avoid treatment regimens that may further suppress the immune system and increase susceptibility to opportunistic infections. Treatment is indicated under two main circumstances. The first is when a single lesion or a limited number of easily accessible lesions are causing significant discomfort or cosmetic problems, such as with prominent facial lesions, lesions overlying a joint, or lesions in the oropharynx that interfere with swallowing or breathing. Under these circumstances, treatment with localized radiation, intralesional vinblastine, topical 9-*cis*-retinoic acid, or cryotherapy may be helpful. It should be noted that patients with HIV infection are particularly sensitive to the side effects of radiation therapy. This is especially true with respect to the development of radiation-induced mucositis; doses of radiation directed at mucosal surfaces, particularly in the head and

TABLE 208-18 Management of AIDS-Associated Kaposi's Sarcoma

- Observation and optimization of antiretroviral therapy
- Single or limited number of lesions
  - Radiation
  - Intralesional vinblastine
  - Cryotherapy
- Extensive disease; inadequate response to ART
  - Initial therapy
    - Interferon α (if CD4+ T cells >150/μL)
    - Liposomal daunorubicin
  - Subsequent therapy
    - Liposomal doxorubicin
    - Paclitaxel
    - Pomalidomide
    - Combination chemotherapy with low-dose doxorubicin, bleomycin, and vinblastine (ABV)
    - Targeted radiation

neck region, should be adjusted accordingly. The second indication for KS-directed treatment is for patients with a large number of lesions or in patients with visceral involvement. In these patients, systemic therapy, either IFN-α or chemotherapy, should be considered. The single most important determinant of response appears to be the CD4+ T cell count. This relationship between response rate and baseline CD4+ T cell count is particularly true for IFN-α. The response rate to IFN-α for patients with CD4+ T cell counts >600/μL is ~80%, while the response rate for patients with counts <150/μL is <10%. In contrast to the other systemic therapies, IFN-α provides an added advantage of having antiretroviral activity; thus, it may be the appropriate first choice for single-agent systemic therapy for early patients with disseminated disease. A variety of chemotherapeutic agents also have been shown to have activity against KS. Five of them—liposomal daunorubicin, liposomal doxorubicin, vinblastine, paclitaxel, and the thalidomide analogue pomalidomide—have been approved by the FDA for this indication. Liposomal daunorubicin and pomalidomide are approved as first-line therapy for patients with advanced KS despite ART. They have fewer side effects than conventional chemotherapy. In contrast, liposomal doxorubicin and paclitaxel are approved only for KS patients who have failed standard chemotherapy. Response rates vary from 23% to 88%, appear to be comparable to what had been achieved earlier with combination chemotherapy regimens, and are greatly influenced by CD4+ T cell count. Vinblastine is most commonly used as an intralesional injection or as part of a combination regimen.

Lymphomas **(Chaps. 113 and 114)** occur with an increased frequency in patients with congenital or acquired T cell immunodeficiencies. AIDS is no exception; at least 6% of all patients with AIDS develop lymphoma at some time during the course of their illness. This is a 10- to 20-fold increase in incidence compared with the general population. In contrast to the situation with KS, primary CNS lymphoma, and most opportunistic infections, the incidence of AIDS-associated systemic lymphomas has not experienced a dramatic decrease as a consequence of the widespread use of effective ART. Lymphoma occurs in all risk groups, with the highest incidence in patients with hemophilia and the lowest incidence in patients from the Caribbean or Africa with heterosexually acquired infection. Lymphoma is a late manifestation of HIV infection, generally occurring in patients with CD4+ T cell counts <200/μL. As HIV disease progresses, the risk of lymphoma increases. The attack rate for lymphoma increases exponentially with increasing duration of HIV infection and decreasing level of immunologic function. At 3 years following a diagnosis of HIV infection, the risk of lymphoma is 0.8% per year; by 8 years after infection, it is 2.6% per year. As individuals with HIV infection live longer as a consequence of improved ART and better treatment and prophylaxis of opportunistic infections, it is anticipated that the incidence of lymphomas may increase.

Three main categories of lymphoma are seen in patients with HIV infection: grade III or IV immunoblastic lymphoma, Burkitt's lymphoma, and primary CNS lymphoma. Approximately 90% of these lymphomas are B cell in phenotype; more than half contain EBV DNA. Some are associated with KSHV. These tumors may be either monoclonal or oligoclonal in nature and are probably in some way related to the pronounced polyclonal B cell activation seen in patients with HIV infection.

Immunoblastic lymphomas account for ~60% of the cases of lymphoma in patients with AIDS. The majority of these are diffuse large B cell lymphomas (DLBCL). They are generally high grade and would have been classified as diffuse histiocytic lymphomas in earlier classification schemes. This tumor is more common in older patients, increasing in incidence from 0% in people with HIV <1 year old to >3% in those >50 years of age. Two variants of immunoblastic lymphoma that are seen primarily in patients with HIV are primary effusion lymphoma (PEL) and its solid variant, plasmacytic lymphoma of the oral cavity. PEL, also referred to as body cavity lymphoma, presents with lymphomatous pleural, pericardial, and/or peritoneal effusions in the absence of discrete nodal or extranodal masses. The tumor cells do not express surface markers for B cells or T cells and are felt to represent a preplasmacytic stage of differentiation. While both KSHV and EBV DNA sequences have been found in the genomes of the malignant cells from patients with body cavity lymphoma, KSHV is felt to be the driving force behind the oncogenesis (see above).

Small noncleaved cell lymphoma (*Burkitt's lymphoma*) accounts for ~20% of the cases of lymphoma in patients with AIDS. It is most frequent in patients 10–19 years old and usually demonstrates characteristic c-*myc* translocations from chromosome 8 to chromosome 14 or 22. Burkitt's lymphoma is not commonly seen in the setting of immunodeficiency other than HIV-associated immunodeficiency, and the incidence of this particular tumor is more than 1000-fold higher in the setting of HIV infection than in the general population. In contrast to African Burkitt's lymphoma, where 97% of the cases contain EBV genome, only 50% of HIV-associated Burkitt's lymphomas are EBV-positive.

Primary CNS lymphoma accounts for ~20% of the cases of lymphoma in patients with HIV infection. In contrast to HIV-associated Burkitt's lymphoma, primary CNS lymphomas are usually positive for EBV. In one study, the incidence of Epstein-Barr positivity was 100%. This malignancy does not have a predilection for any particular age group. The median CD4+ T cell count at the time of diagnosis is ~50/μL. Thus, CNS lymphoma generally presents at a later stage of HIV infection than does systemic lymphoma. This may explain, at least in part, the poorer prognosis for this subset of patients.

The clinical presentation of lymphoma in patients with HIV infection is quite varied, ranging from focal seizures to rapidly growing mass lesions in the oral mucosa **(Fig. 208-44)** to persistent unexplained fever. At least 80% of patients present with extranodal disease, and a similar percentage have B-type symptoms of fever, night sweats, and/or weight loss. Virtually any site in the body may be involved. The most common extranodal site is the CNS, involved in approximately one-third of all patients with lymphoma. Approximately 60% of these cases are primary CNS lymphoma. Primary CNS lymphoma generally presents with focal neurologic deficits, including cranial nerve findings, headaches, and/or seizures. MRI or CT generally reveals a limited number (one to three) of 3- to 5-cm lesions **(Fig. 208-45)**. The lesions often show ring enhancement on contrast administration and may occur in any location. Contrast enhancement is usually less pronounced than that seen with toxoplasmosis. Lesions of CNS lymphoma are most commonly seen deep in the white matter. The main diseases in the differential diagnosis are cerebral toxoplasmosis and cerebral Chagas' disease. In addition to the 20% of lymphomas in people with HIV that are primary CNS lymphomas, CNS disease is also seen in patients with HIV and systemic lymphoma. Approximately 20% of patients with systemic lymphoma have CNS disease in the form of leptomeningeal involvement. This fact underscores the importance of lumbar puncture in the staging evaluation of patients with systemic lymphoma.

Systemic lymphoma is seen at earlier stages of HIV infection than primary CNS lymphoma. In one series the mean CD4+ T cell count was 226/μL. In addition to lymph node involvement, systemic

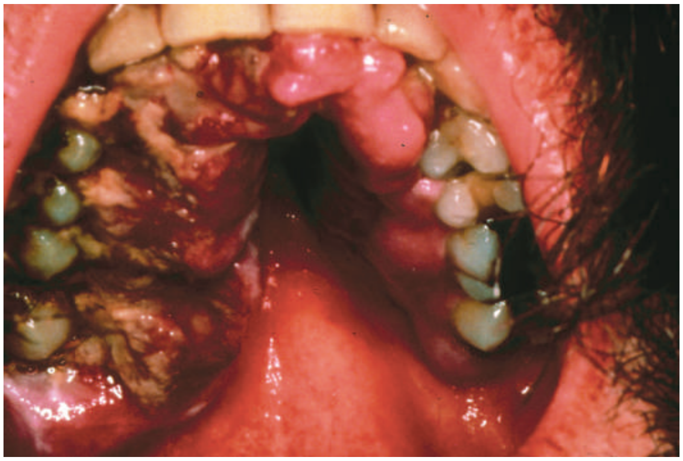

FIGURE 208-44 Immunoblastic lymphoma involving the hard palate of a patient with AIDS.

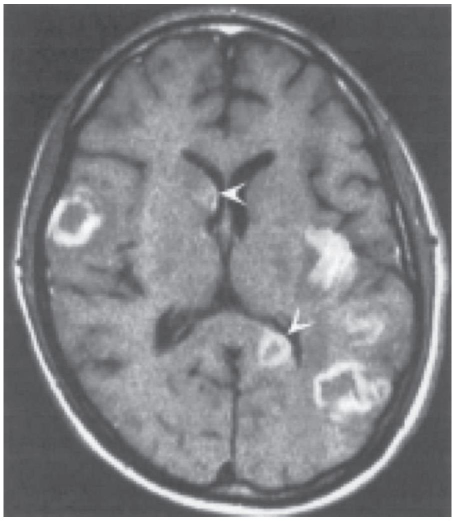

FIGURE 208-45 Central nervous system lymphoma. Postcontrast T1-weighted MRI scan in a patient with AIDS, altered mental status, and hemiparesis. Multiple enhancing lesions, some ring-enhancing, are present. The left sylvian lesion shows gyral and subcortical enhancement, and the lesions in the caudate and splenium (*arrowheads*) show enhancement of adjacent ependymal surfaces.

lymphoma may commonly involve the GI tract, bone marrow, liver, and lung. GI tract involvement is seen in ~25% of patients. Any site in the GI tract may be involved, and patients may complain of difficulty swallowing or abdominal pain. The diagnosis is usually suspected on the basis of CT or MRI of the abdomen. Bone marrow involvement is seen in ~20% of patients and may lead to pancytopenia. Liver and lung involvement are each seen in ~10% of patients. Pulmonary disease may present as a mass lesion, multiple nodules, or an interstitial infiltrate.

Both conventional and unconventional approaches have been employed in an attempt to treat HIV-related lymphomas. Systemic lymphoma is generally treated by the oncologist with combination chemotherapy. Earlier disappointing figures are being replaced with more optimistic results for the treatment of systemic lymphoma following the availability of more effective ART and the use of rituximab in CD20+ tumors. While there is some controversy regarding the use of antiretrovirals during chemotherapy, there is no question that their use overall in patients with HIV lymphoma has improved survival. Concerns regarding synergistic bone marrow toxicities with chemotherapy and ART are mitigated with the use of ART regimens that avoid myelosuppressive antiretrovirals. As in most situations in patients with HIV disease, those with higher CD4+ T cell counts tend to fare better although not as well as patients with lymphoma without HIV. Response rates as high as 72% with a median survival of 33 months and disease-free intervals up to 9 years have been reported. Treatment of primary CNS lymphoma remains a significant challenge. Treatment is complicated by the fact that this illness usually occurs in patients with advanced HIV disease. Palliative measures such as radiation therapy provide some relief. The prognosis remains poor in this group, with a 2-year survival of 20–30%.

Multicentric Castleman's disease (MCD) is a KSHV-associated lymphoproliferative disorder that is seen with an increased frequency in patients with HIV infection. While the incidence of Kaposi's sarcoma has decreased, the incidence of MCD has increased in the setting of ART. While not a true malignancy, MCD shares many features with lymphoma including generalized lymphadenopathy, hepatosplenomegaly, and systemic symptoms of fever, fatigue, and weight loss. Pulmonary symptoms may be seen in ~50% of patients. KS is present in 75–82% of cases. Lymph node biopsies reveal a predominance of interfollicular plasma cells and/or germinal centers with vascularization and an "onion skin" (hyaline vascular) appearance. Prior to the availability of ART, patients with HIV and multicentric Castleman's disease had a 15-fold increased risk of developing non-Hodgkin's lymphoma compared with patients with HIV in general. Treatment typically involves chemotherapy. Rituximab may be of benefit, but it has been associated with worsening of coexisting KS. Similarly, the use of corticosteroids may increase the risk of developing KS. The median survival of patients with treated multicentric Castleman's disease pre-ART was initially reported as 14 months. This has increased to a 2-year survival of more than 90% in the era of ART.

Evidence of infection with *human papillomavirus* (HPV), associated with *intraepithelial dysplasia of the cervix* or *anus*, is approximately twice as common in people with HIV as in the general population and can lead to intraepithelial neoplasia and eventually invasive cancer. In a series of studies, men with HIV were examined for evidence of anal dysplasia, and Papanicolaou (Pap) smears were found to be abnormal in 20–80%. These changes tend to persist and are generally not affected by ART, raising the possibility of a subsequent transition to a more malignant condition. While the incidence of an abnormal Pap smear of the cervix is ~5% in otherwise healthy women, the incidence of abnormal cervical smears in women with HIV infection is 30–60%, and *invasive cervical cancer* is included as an AIDS-defining condition. While only small increases in the absolute numbers of cervical or anal cancers have been seen as a consequence of HIV infection, the relative risk of these conditions when one compares men and women with HIV to men and women without HIV infection is on the order of 10- to 100-fold and decreased with ART. Given the high rates of dysplasia and relative risks for cervical and anal cancer, a comprehensive gynecologic and rectal examination, including Pap smear, is indicated at the initial evaluation and 6 months later for all patients with HIV infection. If these examinations are negative at both time points, the patient should be followed with yearly evaluations. If an initial or repeat Pap smear shows evidence of severe inflammation with reactive squamous changes, the next Pap smear should be performed at 3 months. If, at any time, a Pap smear shows evidence of squamous intraepithelial lesions, colposcopic examination with biopsies as indicated should be performed. The 2-year survival rate for women with HIV and invasive cervical cancer is 64% compared with 79% for women without HIV. In addition to rectal and cervical lesions, HPV can also lead to head and neck cancers. In one study of men who have sex with men, 25% were found to have oral HPV; high-risk HPV genotypes were three times more common in men with HIV. The most common HPV genotypes in the general population and the genotypes upon which current HPV vaccines are based are 6, 11, 16, and 18. In the population of people with HIV, other genotypes such as 58 and 53 also are prominent. This raises a concern about the level of effectiveness of the current HPV vaccines for patients with HIV. Despite this, it is recommended that patients with HIV infection be vaccinated against HPV.

IDIOPATHIC CD4+ T LYMPHOCYTOPENIA

A syndrome was recognized in 1992 characterized by an absolute CD4+ T cell count of <300/μL or <20% of total T cells on a minimum of two occasions at least 6 weeks apart; no evidence of HIV-1, HIV-2, HTLV-1, or HTLV-2 on testing; and the absence of any defined immunodeficiency or therapy associated with decreased levels of CD4+ T cells. By mid-1993, ~100 patients had been described. After extensive multicenter investigations, a series of reports were published in early 1993, which together allowed a number of conclusions. Idiopathic CD4+ lymphocytopenia (ICL) is a very rare syndrome, as determined by studies of blood donors and cohorts of HIV-seronegative men who have sex with men. Cases were clearly identified as early as 1983. The definition of ICL based on CD4+ T cell counts coincided with the ready availability of testing for CD4+ T cells in patients suspected of being immunodeficient. However, as a result of immune deficiency, certain patients with ICL develop some of the opportunistic diseases (particularly cryptococcosis, nontuberculous mycobacterial infections, and HPV disease) seen in patients with HIV. In one study, HPV-associated disease was seen in 29% of patients, cryptococcosis in 24%, molluscum contagiosum in 95%, and nontuberculous mycobacterial disease in 5%. Approximately 10–30% of patients may exhibit an autoimmune disease; this is more common in those with higher CD4 counts. The syndrome is demographically, clinically, and immunologically unlike HIV infection and AIDS. Fewer than half of the reported ICL patients had risk factors for HIV infection, and there were wide geographic and age distributions. The fact that a significant proportion of initially diagnosed patients did have risk factors probably reflects a selection bias, in that physicians who take care of patients with HIV were more likely to monitor CD4+ T cells. Approximately half of the patients are women, compared with approximately one-fifth among people with HIV in the United States. Many patients with ICL remained clinically stable, and their condition may not deteriorate progressively as is common with seriously immunodeficient patients with HIV. Approximately 15% of patients with ICL experience spontaneous reversal of the CD4+ T lymphocytopenia. Immunologic abnormalities in ICL are somewhat different from those of HIV infection. ICL patients often have increases in CD4+ T cell activation with decreases in CD8+ T cells and B cells. Furthermore, immunoglobulin levels are either normal or, more commonly, decreased in patients with ICL, compared with the usual hypergammaglobulinemia of people with HIV. Virologic studies of these patients have revealed no evidence of HIV-1, HIV-2, HTLV-1, or HTLV-2 or of any other mononuclear cell–tropic virus. Furthermore, there has been no epidemiologic evidence to suggest that a transmissible microbe was involved. The cases of ICL have been widely dispersed, with no clustering. Close contacts and sexual partners who were studied were clinically well and were serologically, immunologically, and virologically negative for HIV. ICL is a heterogeneous syndrome, and it is highly likely that there is no common cause; however, there may be common causes among subgroups of patients that are currently unrecognized.

Patients who present with laboratory data consistent with ICL should be worked up for underlying diseases that could be responsible for the immune deficiency. If no underlying cause is detected, no specific therapy should be initiated. However, if opportunistic diseases occur, they should be treated appropriately (see above). Depending on the level of the CD4+ T cell count, patients should receive prophylaxis for the commonly encountered opportunistic infections.

TREATMENT

AIDS and Related Disorders

GENERAL PRINCIPLES OF PATIENT MANAGEMENT

The CDC guidelines call for the testing for HIV infection to be a part of routine medical care. It is recommended that the patient be informed of the intention to test, as is the case with other routine laboratory determinations, and be given the opportunity to "opt out." Such an approach is critical to the goal of identifying as many infected individuals as possible since 13% of the 1.2 million individuals in the United States with HIV are not aware of their status. In the setting of routine testing, although it is difficult, pretest counseling is an important part of the process. No matter how well prepared a patient is for adversity, the discovery of a diagnosis of HIV infection is a devastating event. Thus, physicians should be sensitive to this fact and, where possible, utilize pretest counseling to at least partially prepare the patient should the results demonstrate the presence of HIV infection. Following a diagnosis of HIV infection, the health care provider should be prepared to immediately activate support systems for the newly diagnosed patient and initiate ART. These supports should include individuals who can spend time talking to the newly diagnosed person and ensuring that he or she is emotionally stable and ready to begin therapy. Most communities have HIV support centers that can be of great help in these difficult situations.

The treatment of patients with HIV infection requires not only a comprehensive knowledge of the possible disease processes that may occur and up-to-date knowledge of and experience with ART, but also the ability to deal with the problems of a chronic, potentially life-threatening illness. A comprehensive knowledge of internal medicine is required to deal with the changing spectrum of illnesses associated with HIV infection, many of which are similar to a state of accelerated aging. The appropriate use of potent ART and other treatment and prophylactic interventions is of critical importance in providing each patient with the best opportunity to live a long and healthy life with HIV infection. In contrast to the earlier days of this epidemic, a diagnosis of HIV infection needs no longer be equated with having an inevitably fatal disease. In addition to medical interventions, the health care provider has a responsibility to provide each patient with appropriate counseling and education concerning their disease as part of a comprehensive care plan. Patients must be educated about the potential transmissibility of their infection and about the fact that while health care providers may refer to levels of the virus as "undetectable," this is only a reflection of the sensitivity of the assay being used to measure the virus, rather than a comment on the presence or absence of the virus. It is important for patients to be aware that the virus is still present in virtually all patients who have ever been diagnosed with HIV infection and capable of being transmitted in the absence of effective ART. Thus, there must be frank discussions concerning sexual practices and the sharing of syringes and other paraphernalia used in illicit drug use. The treating physician not only must be aware of the latest medications available for patients with HIV infection but also must educate patients concerning the natural history of their illness, listen to their concerns, and be sensitive to their fears. As with other diseases, therapeutic decisions should be made in consultation with the patient, when possible, and with the patient's proxy if the patient is incapable of making decisions. In this regard, it is recommended that all patients with HIV infection, and in particular those with CD4+ T cell counts <200/μL, designate a trusted individual with durable power of attorney to make medical decisions on their behalf, if necessary.

Following a diagnosis of HIV infection, several examinations and laboratory studies should be performed to help determine the extent of disease and provide baseline standards for future reference **(Table 208-19)**. In addition to routine chemistry, fasting lipid profile, aspartate aminotransferase, alanine aminotransferase, total and direct bilirubin, fasting glucose and hematology screening panels, Pap smear, urinalysis, and chest x-ray, one should also obtain a CD4+ T cell count, a plasma HIV RNA level, an HIV resistance test, a rapid plasma reagin (RPR) or Venereal Disease Research Laboratory (VDRL) test, an anti-*Toxoplasma* antibody titer, and serologies for hepatitis A, B, and C. A PPD test or IFN-γ release assay should be done and an MMSE performed and recorded. A pregnancy test should be done for any women for whom the drug efavirenz is being considered, and HLA-B5701 testing should be done in all patients in whom the drug abacavir is being considered.

TABLE 208-19 Initial Evaluation of the Patient with HIV Infection
History and physical examination
Routine chemistry and hematology
AST, ALT, alkaline phosphatase, direct and indirect bilirubin
Lipid profile and fasting glucose
CD4+ T lymphocyte count
Plasma HIV RNA level
HIV resistance testing
HLA-B5701 screening
RPR or VDRL test
Anti-*Toxoplasma* antibody titer
Urinalysis
PPD skin test or IFN-γ release assay
Mini-Mental Status Examination
Serologies for hepatitis A, hepatitis B, and hepatitis C
Immunization with pneumococcal polysaccharide; influenza; HPV as indicated
Immunization with hepatitis A and hepatitis B if seronegative
Counseling regarding natural history and transmission
Help contacting others who might be infected

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; PPD, purified protein derivative; RPR, rapid plasma reagin; VDRL, Venereal Disease Research Laboratory.

TABLE 208-20 HIV Resources Available on the World Wide Web

CLINICALINFO.HIV.GOV	US-approved HIV treatment guidelines
CLINICALTRIALS.GOV	Information on clinical trials
HIV.GOV	Epidemiological data and prevention information

Patients should be immunized with pneumococcal polysaccharide, with annual influenza and COVID-19 shots, and, if seronegative for these viruses, with HPV, hepatitis A, and hepatitis B vaccines. The status of hepatitis C infection should be determined. In addition, patients should be counseled with regard to sexual practices and needle sharing, and counseling should be offered to people whom the patient knows, or suspects, also may be infected. Once these baseline activities are performed, short- and long-term medical management strategies should be developed based on the most recent information available and modified as new information becomes available. The field of HIV medicine is changing rapidly, and it is difficult to remain fully up to date. Fortunately, there are a series of excellent sites on the Internet that are frequently updated, and they provide the most recent information on a variety of topics, including consensus panel reports on treatment **(Table 208-20)**.

ANTIRETROVIRAL THERAPY

Combination antiretroviral therapy (ART), also referred to as highly active antiretroviral therapy (HAART), is the cornerstone of management of patients with HIV infection and should be initiated as soon as possible following a diagnosis of HIV infection. One exception to immediate initiation of ART is in the setting of cryptococcal meningitis or TB where several weeks of specific antimicrobial therapy prior to initiation of ART may decrease the risk of severe IRIS. Following the initiation of widespread use of ART in the United States in 1995–1996, marked declines were noted in the incidence of most AIDS-defining conditions (Fig. 208-34). Suppression of HIV replication is an important component in prolonging and improving the quality of life for the patient as well as minimizing the risk of transmission of HIV to others. Adequate suppression of HIV replication requires strict adherence to prescribed regimens of antiretroviral drugs. This has been facilitated by the coformulations of antiretrovirals and the development of once-daily, monthly regimens, and every-six-month regimens with the expectation that even longer-lasting drugs are on the horizon. Among the decisions that need to be made in the context of prescribing ART are selection of the best initial regimen, determining when a given regimen should be changed, and deciding what regimen should be selected when a change is made. The care provider and patient must come to a mutually agreeable plan based on the best available data. In an effort to facilitate this process, the U.S. Department of Health and Human Services makes available on the Internet (*clinicalinfo.hiv.gov/en/guidelines*) a series of periodically updated guidelines, including *"Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adults and Adolescents*" and "*Guidelines for the Prevention and Treatment of Opportunistic Infections in Persons Infected with Human Immunodeficiency Virus*." At present, an extensive clinical trials network, involving both clinical investigators and patient advocates, is in place attempting to develop improved approaches to therapy. Consortia comprising representatives of academia, industry, independent foundations, and the federal government engage in drug development, including a wide-ranging series of clinical trials. As a result, new therapies and new therapeutic strategies are continually emerging. New drugs are often available through expanded-access programs prior to official licensure. Given the complexity of this field, decisions regarding ART are best made in consultation with experts.

Currently available drugs for the treatment of HIV infection as part of a combination regimen fall into five categories: those that inhibit the viral reverse transcriptase enzyme (nucleoside and nucleotide reverse transcriptase inhibitors; nonnucleoside reverse transcriptase inhibitors), those that inhibit the viral protease enzyme (protease inhibitors), those that inhibit the viral integrase enzyme (integrase strand transfer inhibitors), those that interfere with viral entry (fusion inhibitors; CCR5 antagonists; CD4 antagonists) and those that interfere with the viral capsid **(Table 208-21; Fig. 208-46)**. A typical regimen will include two nucleoside/nucleotide reverse transcriptase inhibitors (usually a tenofovir-based drug or abacavir + 3TC or FTC) plus a nonnucleoside reverse transcriptase inhibitor, an integrase inhibitor, or a protease inhibitor boosted with a pharmacokinetic enhancer (ritonavir or cobicistat). More recent studies have also supported the two-drug regimen of dolutegravir plus 3TC for initial therapy in hepatitis B–negative patients with baseline HIV RNA levels <500,000 copies/mL. Numerous fixed-drug formulations combining two or more of these antiretroviral drugs have been licensed **(Table 208-22)**. Prior to initiation of therapy and at any time a change in therapy due to treatment failure is being considered, drug resistance testing should be performed to help guide the selection of drugs to be used in combination. A summary of known resistance mutations for antiretroviral drugs is shown in **Fig. 208-47**.

While most patients with HIV infection will be infected with HIV-1, some patients, especially those with an epidemiologic link to West Africa, may be infected with HIV-2. While the principles of treatment are the same as those for persons infected with HIV-1, it is important to note that the nonnucleoside reverse transcriptase inhibitors, enfuvirtide, and fostemsavir are not active against HIV-2 and should not be used as part of ART regimens in HIV-2–infected individuals.

The FDA-approved reverse transcriptase inhibitors include the *nucleoside analogues* zidovudine, didanosine, zalcitabine, stavudine, lamivudine, abacavir, and emtricitabine; the *nucleotide analogues* tenofovir disoproxil and tenofovir alafenamide; and the *nonnucleoside reverse transcriptase inhibitors* nevirapine, delavirdine, efavirenz, etravirine, rilpivirine, long-acting rilpivirine, and doravirine (Table 208-21). These represent the first class of drugs licensed for the treatment of HIV infection. They are indicated for this use as part of combination regimens. It should be stressed that none of these drugs should be used as monotherapy for HIV infection due to the relative ease with which drug resistance may develop under such circumstances. Thus, when lamivudine, emtricitabine, or tenofovir is used to treat hepatitis B infection in the setting of HIV infection, one should ensure that the patient is also on additional antiretroviral medication. Similarly, when any of these three medications are discontinued, one needs to be vigilant for a flare of hepatitis B in coinfected patients. The reverse transcriptase inhibitors block the HIV replication cycle at the point of RNA-dependent DNA synthesis, the reverse transcription step. While the nonnucleoside reverse transcriptase inhibitors are quite selective for the HIV-1 reverse transcriptase, the nucleoside and nucleotide analogues inhibit a variety of DNA polymerases in addition to those of the HIV-1 reverse transcriptase. For this reason, serious side effects are more varied with the nucleoside analogues and include mitochondrial damage that can lead to hepatic steatosis and lactic acidosis as well as peripheral neuropathy and pancreatitis. The use of either of the thymidine analogues zidovudine and stavudine has been associated with a syndrome of hyperlipidemia, glucose intolerance/insulin resistance, and fat redistribution often referred to as *lipodystrophy syndrome* (discussed in "Diseases of the Endocrine System and Metabolic Disorders," above). For these reasons, the older drugs in this class, zidovudine, didanosine, zalcitabine, and stavudine, are no longer recommend for use in the United States due to their side-effect profiles. The nucleoside and nucleotide transcriptase inhibitors preferred for use in combination regimens according to the DHHS Panel on the use of antiretroviral drugs are lamivudine, emtricitabine, abacavir, tenofovir disoproxil, and tenofovir alafenamide. Given its renal toxicity, tenofovir disoproxil should be limited to use in patients with creatinine clearance (CrCl) >70 while tenofovir alafenamide should generally be limited to use in patients with CrCl >30. The preferred nonnucleoside reverse transcriptase inhibitors are efavirenz, rilpivirine, and doravirine. Of note, rilpivirine is approved for treatment only in ART-naïve patients with HIV RNA levels <100,000 copies/mL and is contraindicated in patients taking proton pump inhibitors.

TABLE 208-21 Antiretroviral Drugs Licensed in the United States for the Treatment of HIV Infection

DRUG[a]	INDICATION	DOSE IN COMBINATION	SUPPORTING DATA	TOXICITY
Nucleoside or Nucleotide Reverse Transcriptase Inhibitors				
Lamivudine (Epivir, 2′3′-dideoxy-3′-thiacytidine, 3TC)	In combination with other antiretroviral agents for the treatment of HIV infection	150 mg bid 300 mg qd	In combination with zidovudine superior to zidovudine alone with respect to changes in CD4+ T cell counts in 495 patients who were zidovudine-naïve and 477 patients who were zidovudine-experienced; overall CD4+ T cell counts for the zidovudine group were at baseline by 24 weeks, while in the group treated with zidovudine plus lamivudine they were 10–50 cells/μL above baseline; 54% decrease in progression to AIDS/death compared with zidovudine alone	Flare of hepatitis in HBV-co-infected patients who discontinue drug
Emtricitabine (FTC, Emtriva)	In combination with other antiretroviral agents for the treatment of HIV infection	200 mg qd	Comparable to lamivudine in combination with stavudine and nevirapine/efavirenz	Hepatotoxicity in HBV-co-infected patients who discontinue drug, skin discoloration
Abacavir (Ziagen)	For treatment of HIV infection in combination with other antiretroviral agents	300 mg bid	Abacavir + zidovudine + 3TC equivalent to indinavir + zidovudine + 3TC with regard to viral load suppression (~60% in each group with <400 HIV RNA copies/mL plasma) and CD4+ T cell increase (~100/μL in each group) at 24 weeks	Hypersensitivity reaction In HLA-B5701+ individuals (can be fatal); fever, rash, nausea, vomiting, malaise or fatigue, and loss of appetite
Tenofovir disoproxil fumarate (Viread)	For use in combination with other antiretroviral agents when treatment is indicated	300 mg qd	Reduction of ~0.6 log in HIV-1 RNA levels when added to background regimen in treatment-experienced patients	Renal, osteomalacia, flare of hepatitis in HBV-co-infected patients who discontinue drug
Tenofovir alafenamide (Vemlidy)	In combination with emtricitabine and other antiretroviral agents for treatment of HIV-1 infection	25 mg qd	92% of patients treated in combination with emtricitabine, elvitegravir, and cobicistat had HIV-1 RNA levels <50 copies/mL	Nausea, less renal toxicity than tenofovir disoproxil fumarate
Nonnucleoside Reverse Transcriptase Inhibitors				
Nevirapine (Viramune)	In combination with other antiretroviral agents for treatment of progressive HIV infection	200 mg/d × 14 days then 200 mg bid *or* 400 mg extended release qd	Increase in CD4+ T cell count, decrease in HIV RNA when used in combination with nucleosides	Skin rash, hepatotoxicity
Efavirenz (Sustiva)	For treatment of HIV infection in combination with other antiretroviral agents	600 mg qhs	Efavirenz + zidovudine + 3TC comparable to indinavir + zidovudine + 3TC with regard to viral load suppression (a higher percentage of the efavirenz group achieved viral load <50 copies/mL, but the discontinuation rate in the indinavir group was unexpectedly high, accounting for most treatment "failures"); CD4 cell increase (~140/μL in each group) at 24 weeks	Rash, dysphoria, elevated liver function tests, drowsiness, abnormal dreams, depression, lipid abnormalities, potentially teratogenic
Etravirine (Intelence)	In combination with other antiretroviral agents in treatment-experienced patients whose HIV is resistant to nonnucleoside reverse transcriptase inhibitors and other antiretroviral medications	200 mg bid	Higher rates of HIV RNA suppression to <50 copies/mL (56% vs 39%); greater increases in CD4+ T cell count (89 vs 64 cells) compared to placebo when given in combination with an optimized background regimen	Rash, nausea, hypersensitivity reactions
Rilpivirine (Edurant)	In combination with other drugs in previously untreated patients when treatment is indicated	25 mg qd	Noninferior to efavirenz with respect to suppression at week 48 in 1368 treatment-naïve individuals, except in patients with pretherapy HIV RNA levels >100,000 where it was inferior	Nausea, dizziness, somnolence, vertigo, less CNS toxicity and rash than efavirenz
Protease Inhibitors				
Ritonavir (Norvir)	In combination with other antiretroviral agents for treatment of HIV infection when treatment is warranted	600 mg bid (also used in lower doses as pharmacokinetic booster)	Reduction in the cumulative incidence of clinical progression or death from 34% to 17% in patients with CD4+ T cell count <100/μL treated for a median of 6 months	Nausea, abdominal pain, hyperglycemia, fat redistribution, lipid abnormalities, may alter levels of many other drugs, paresthesias, hepatitis
Atazanavir (Reyataz)	For treatment of HIV infection in combination with other antiretroviral agents	400 mg qd or 300 mg qd + ritonavir 100 mg qd when given with efavirenz	Comparable to efavirenz when given in combination with zidovudine + 3TC in a study of 810 treatment-naïve patients; comparable to nelfinavir when given in combination with stavudine + 3TC in a study of 467 treatment-naïve patients	Hyperbilirubinemia, PR prolongation, nausea, vomiting, hyperglycemia, fat maldistribution, rash transaminase elevations, renal stones

(*Continued*)

TABLE 208-21 Antiretroviral Drugs Licensed in the United States for the Treatment of HIV Infection (Continued)

DRUG[a]	INDICATION	DOSE IN COMBINATION	SUPPORTING DATA	TOXICITY
Darunavir (Prezista)	In combination with 100 mg ritonavir for combination therapy in treatment-experienced adults	600 mg + 100 mg ritonavir bid with food	At 24 weeks, patients with prior extensive exposure to antiretrovirals treated with a new combination including darunavir showed a –1.89-log change in HIV RNA levels and a 92-cell increase in CD4+ T cells compared with –0.48 log and 17 cells in the control arm	Diarrhea, nausea, headache, skin rash, hepatotoxicity, hyperlipidemia, hyperglycemia
Entry Inhibitors				
Enfuvirtide (Fuzeon)	In combination with other agents in treatment-experienced patients with evidence of HIV-1 replication despite ongoing antiretroviral therapy	90 mg SC bid	In treatment of experienced patients, superior to placebo when added to new optimized background (37% vs 16% with <400 HIV RNA copies/mL at 24 weeks; + 71 vs + 35 CD4+ T cells at 24 weeks)	Local injection reactions, hypersensitivity reactions, increased rate of bacterial pneumonia
Maraviroc (Selzentry)	In combination with other antiretroviral agents in adults infected with only CCR5-tropic HIV-1	150–600 mg bid depending on concomitant medications (see text)	At 24 weeks, among 635 patients with CCR5-tropic virus and HIV-1 RNA >5000 copies/mL despite at least 6 months of prior therapy with at least 1 agent from 3 of the 4 antiretroviral drug classes, 61% of patients randomized to maraviroc achieved HIV RNA levels <400 copies/mL compared with 28% of patients randomized to placebo	Hepatotoxicity, nasopharyngitis, fever, cough, rash, abdominal pain, dizziness, musculoskeletal symptoms
Ibalizumab (Trogarzo)	In combination with other antiretroviral agents in patients with multidrug-resistant HIV-1	Single loading dose of 2000 mg followed by a maintenance dose of 800 mg every 2 weeks	At 25 weeks, 50% of patients with multidrug-resistant HIV-1 with HIV-1 RNA >1000 copies/mL treated with an optimized background of 1 active drug and ibalizumab achieved HIV RNA levels <200 copies/mL	Rash, diarrhea, nausea
Fostemsavir (Rukobia)	In combination with other antiretroviral agents in patients with multi-drug resistant HIV-1	600 mg bid	At 96 weeks, 64% of patients with multi-drug resistant HIV-1 treated with an optimized background and fostemsavir achieved an HIV-1 RNA level <200 copies/mL	QTc prolongation at higher doses, elevation in liver enzymes in those with hepatitis B or C
Integrase Inhibitors				
Raltegravir (Isentress)	In combination with other antiretroviral agents	400 mg bid	At 24 weeks, among 436 patients with 3-class drug resistance, 76% of patients randomized to receive raltegravir achieved HIV RNA levels <400 copies/mL compared with 41% of patients randomized to receive placebo	Nausea, headache, diarrhea, CPK elevation, muscle weakness, rhabdomyolysis
Elvitegravir (Available only in combination with cobicistat, tenofovir, and emtricitabine [Stribild])	Fixed-dose combination	1 tablet daily	Noninferior to raltegravir or atazanavir/ritonavir in treatment-experienced patients	Diarrhea, nausea, upper respiratory infections, headache
Dolutegravir (Tivicay)	In combination with other antiretroviral agents	50 mg daily for treatment-naïve patients 50 mg twice daily for treatment-experienced patients or those also receiving efavirenz or rifampin	Noninferior to raltegravir, superior to efavirenz or darunavir/ritonavir	Insomnia, headache, hypersensitivity reactions, hepatotoxicity
Bictegravir (Available only in combination with tenofovir alafenamide and emtricitabine [Biktarvy])	For treatment of HIV infection in adults	50 mg bictegravir/ 25 mg tenofovir alafenamide/200 mg emtricitabine qd	Noninferior to dolutegravir/tenofovir/ emtricitabine and noninferior to dolutegravir/ abacavir/lamivudine	Nausea, diarrhea, headache
Cabotegravir (Vocabria)	In combination with rilpivirine for treatment of HIV infection in adults	Oral lead-in of 30 mg + 25 mg rilpivirine for 1 month; followed by an initial injection of 600 mg (3 mL) IM + 900 mg (3 mL) rilpivirine IM; followed by monthly injections of 400 mg (2 mL) IM + 600 mg (2 mL) rilpivirine IM	Noninferior to abacavir/dolutegravir/ lamivudine or dolutegravir + 2 nucleoside/tide reverse transcriptase inhibitors Noninferior to nonnucleoside reverse transcriptase inhibitor + 2 nucleoside/ tide reverse transcriptase inhibitors or a protease inhibitor + 2 nucleoside/tide reverse transcriptase inhibitors or an integrase inhibitor and 2 nucleoside/tide reverse transcriptase inhibitors	Injection-site reactions

(*Continued*)

TABLE 208-21 Antiretroviral Drugs Licensed in the United States for the Treatment of HIV Infection (Continued)

DRUG[a]	INDICATION	DOSE IN COMBINATION	SUPPORTING DATA	TOXICITY
Lenacapavir (Sunlenca)	In combination with other agents for treatment of multidrug-resistant HIV infection in adults	Induction with either (927 mg SC + 600 mg orally day 1; 600 mg orally day 2) or (600 mg orally days 1 and 2; 300 mg orally day 8; 927 mg SC day 15) followed by 927 mg SC every 6 months	At 52 weeks, 83% of patients with multi-drug resistant HIV-1 treated with an optimized background and lenacapavir achieved an HIV-1 level <50 copies/mL	Injection site reactions, nausea

[a]Initial trade names are provided. Generic forms may be available.

Abbreviations: ARC, AIDS-related complex; bid, twice daily; CPK, creatine phosphokinase; HBV, hepatitis B virus; IM, intramuscular; qd, once daily; qhs, once daily at bedtime; SC, subcutaneous.

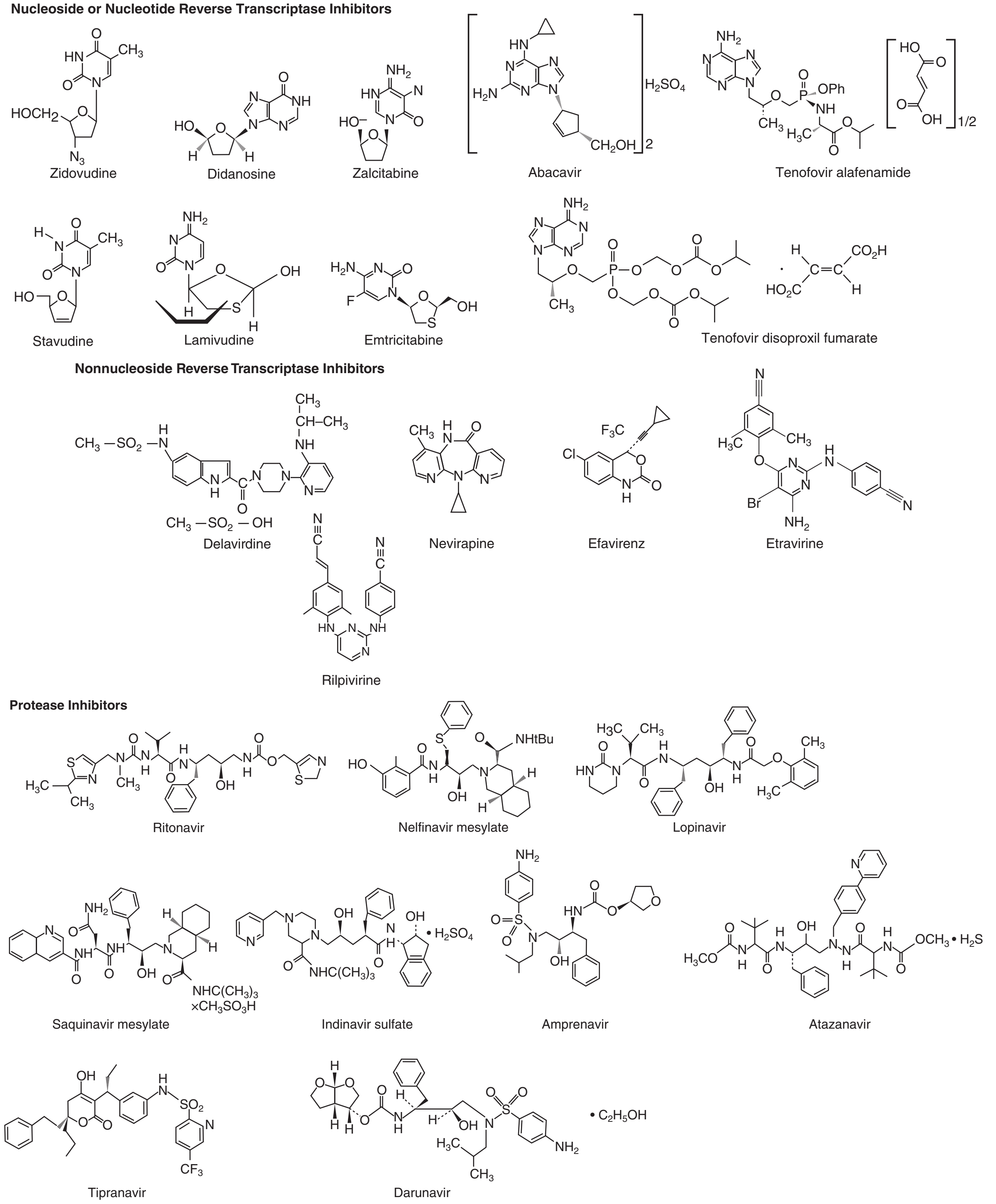
Nucleoside or Nucleotide Reverse Transcriptase Inhibitors
Zidovudine
Didanosine
Zalcitabine
Abacavir
Tenofovir alafenamide
Stavudine
Lamivudine
Emtricitabine
Tenofovir disoproxil fumarate
Nonnucleoside Reverse Transcriptase Inhibitors
Delavirdine
Nevirapine
Efavirenz
Etravirine
Rilpivirine
Protease Inhibitors
Ritonavir
Nelfinavir mesylate
Lopinavir
Saquinavir mesylate
Indinavir sulfate
Amprenavir
Atazanavir
Tipranavir
Darunavir

FIGURE 208-46 Molecular structures of antiretroviral agents.

Entry Inhibitors

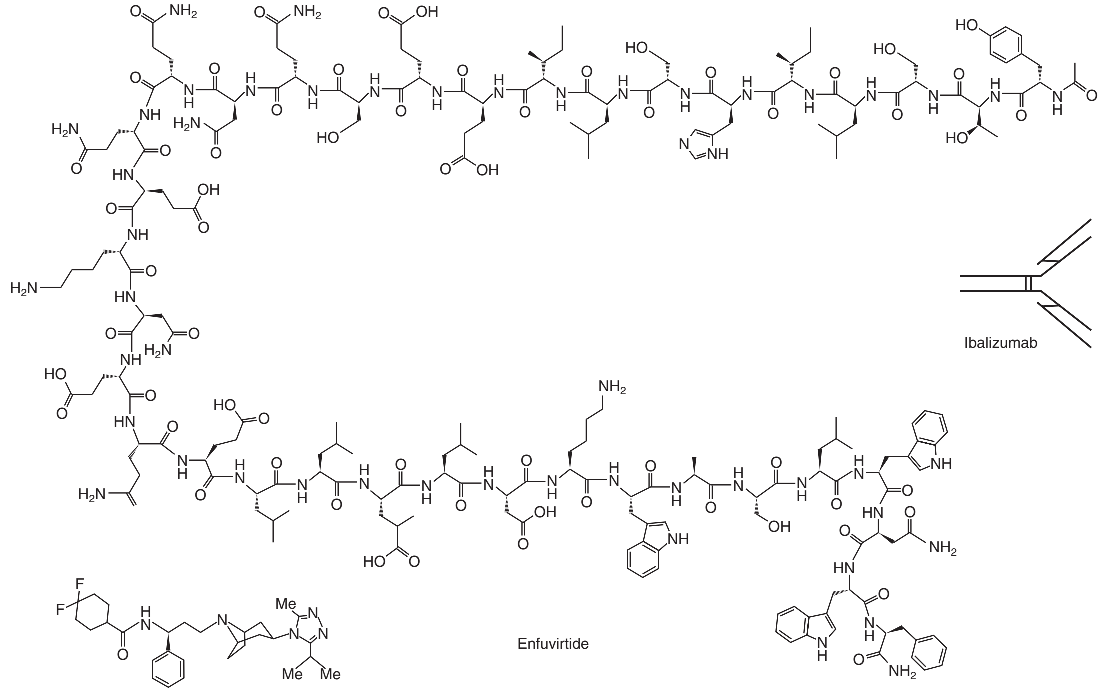

Ibalizumab

Enfuvirtide

Maraviroc

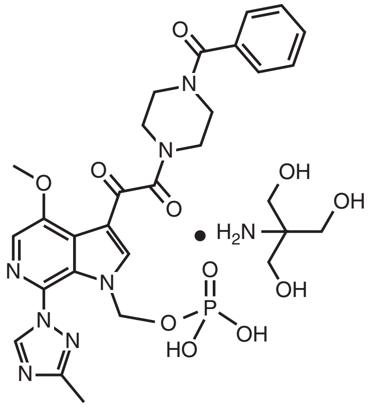

Fostemsavir

Integrase Inhibitors

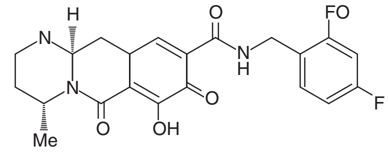

Elvitegravir

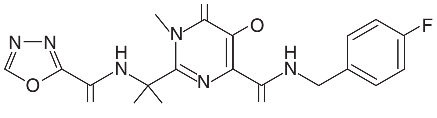

Raltegravir

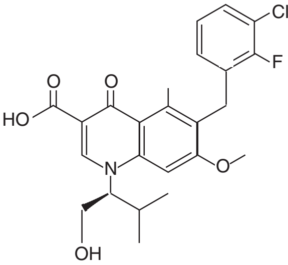

Dolutegravir

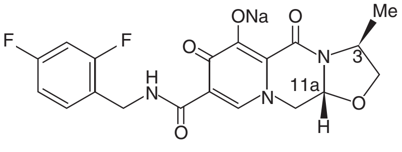

Cabotegravir

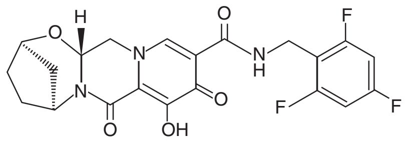

Bictegravir

Capsid Inhibitor

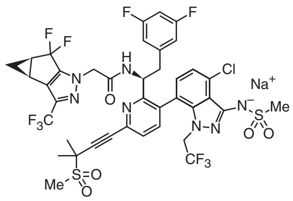

Lenacapavir

FIGURE 208-46 (Continued)

TABLE 208-22 Combination Formulations of Antiretroviral Drugs

NAME	COMBINATION
ABC/3TC (generic)	Abacavir 600 mg/lamivudine 300 mg
EFV/TDF/FTC (generic)[a]	Efavirenz 600 mg/tenofovir disoproxil fumarate 300 mg/ emtricitabine 200 mg
Biktarvy[a]	Bictegravir 50 mg/tenofovir alafenamide 25 mg/emtricitabine 200 mg
Cabenuva[a]	Cabotegravir + rilpivirine (long-acting injection)
Cimduo	Tenofovir disoproxil fumarate 300 mg/lamivudine 300 mg
Complera[a]	Rilpivirine 25 mg/tenofovir disoproxil fumarate 300 mg/ emtricitabine 200 mg
Delstrigo[a]	Doravirine 100 mg/tenofovir disoproxil fumarate 300 mg/ lamivudine 300 mg
Descovy	Tenofovir alafenamide 25 mg/emtricitabine 200 mg
Dovato[a]	Dolutegravir 50 mg/lamivudine 300 mg
Genvoya[a]	Elvitegravir 150 mg/cobicistat 150 mg/tenofovir alafenamide 10 mg/emtricitabine 200 mg
Juluca[a]	Dolutegravir 50 mg/rilpivirine 25 mg
Odefsey[a]	Rilpivirine 25 mg/tenofovir alafenamide 25 mg/emtricitabine 200 mg
Stribild[a]	Elvitegravir 150 mg/cobicistat 150 mg/tenofovir disoproxil fumarate 300 mg/emtricitabine 200 mg
Symfi[a]	Efavirenz 600 mg/tenofovir disoproxil fumarate 300 mg/lamivudine 300 mg
Symfi Lo[a]	Efavirenz 400 mg/tenofovir disoproxil fumarate 300 mg/lamivudine 300 mg
Symtuza[a]	Darunavir 800 mg/cobicistat 150 mg/tenofovir alafenamide 10 mg/ emtricitabine 200 mg
Triumeq[a]	Dolutegravir 50 mg/abacavir 600 mg/lamivudine 300 mg
Truvada	Tenofovir disoproxil fumarate 300 mg/emtricitabine 200 mg

[a] Complete, once-daily, single-tablet regimens.

Source: Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adults and Adolescents, USPHS.

The HIV-1 *protease inhibitors* (saquinavir, indinavir, ritonavir, nelfinavir, amprenavir, fosamprenavir, lopinavir/ritonavir, atazanavir, atazanavir/cobicistat, tipranavir, darunavir, and darunavir/cobicistat) are an important part of the therapeutic armamentarium of antiretrovirals. While possessing antiviral properties of its own, ritonavir is typically used as a pharmacokinetic enhancer due to its high affinity for several isoforms of cytochrome P450 (3A4, 2D6) leading to large increases in the plasma concentrations of co-administered drugs metabolized by these pathways. As in the case of reverse transcriptase inhibitors, resistance to protease inhibitors can develop rapidly in the setting of monotherapy, and thus these agents should be used only as part of combination therapeutic regimens. Based on superior efficacy and side-effect profile, ritonavir-boosted darunavir in combination with emtricitabine and tenofovir (disoproxil or alafenamide) is the preferred protease inhibitor strategy according to the DHHS Panel on the use of antiretroviral drugs.

Integrase strand transfer inhibitors act by blocking the action of the HIV integrase enzyme and thus preventing integration of the HIV provirus into the host cell genome. They are among the most potent and safest of the antiretroviral drugs and frequently part of initial combination regimens. The five licensed integrase inhibitors are raltegravir, cabotegravir, elvitegravir, dolutegravir, and bictegravir. Cabotegravir is an integrase inhibitor that is given in combination with rilpivirine as a monthly injection. Prior to initiation of the monthly injections, patients should initially be treated with oral preparations of the two drugs to be sure they are well tolerated. Elvitegravir is always given in combination with cobicistat, which acts to boost the concentrations of elvitegravir. Cobicistat also inhibits tubular secretion of creatinine, resulting in increases in serum creatinine, and is not recommended for patients with estimated creatinine clearances <70 mL/min. Bictegravir is available only in combination with tenofovir alafenamide and emtricitabine. When used as part of initial ART, integrase inhibitor–containing regimens have been associated with greater weight gain than nonnucleoside reverse transcriptase inhibitor– or protease inhibitor–containing regimens.

Entry inhibitors act by interfering with the binding of HIV to its receptor or co-receptor or by interfering with the process of fusion (see above). The first drug in this class to be licensed was the fusion inhibitor enfuvirtide, or T-20, followed by the CCR5 antagonist maraviroc. The anti-CD4 monoclonal antibody ibalizumab was licensed in 2018, and the small molecule fostemsavir in 2020. Given that maraviroc is effective only against CCR5-tropic viruses, a co-receptor tropism assay should be performed when use of this agent is being considered.

The *capsid inhibitor* lenacapavir acts by interfering with viral replication by binding to the interface between p24 capsid subunits in hexamers. This then interferes with multiple steps in the viral lifecyle including nuclear uptake of proviral DNA, virus assembly and release, and capsid core formation. It is licensed for treatment of individuals who are heavily treatment-experienced with multidrug-resistant virus that is not suppressed on the current regimen. It is given by subcutaneous injection every 6 months after an oral or oral plus subcutaneous loading regimen.

PRINCIPLES OF THERAPY

The principles of therapy for HIV infection have been articulated by a panel sponsored by the U.S. Department of Health and Human Services as a working group of the NIH Office of AIDS Research Advisory Council. These principles are summarized in **Table 208-23**. As noted in these guidelines, ART of HIV infection does not lead to eradication or cure of HIV. The possible exceptions are a limited number of individuals with HIV infection and cancer who received allogeneic stem cell transplants from donors who were homozygous for the CCR5Δ32 mutation (see above) and thus resistant to HIV infection.

Treatment decisions must consider the fact that one is dealing with a chronic infection that requires life-long therapy. Patients initiating antiretroviral therapy must be willing to commit to ongoing treatment and understand the importance of adherence to their prescribed regimen. The importance of adherence is illustrated by the observation that treatment interruption is associated with rapid increases in HIV RNA levels, rapid declines in CD4+ T cell counts, and an increased risk of clinical progression. While it seems reasonable to assume that the complications associated with ART could be minimized by intermittent treatment regimens designed to minimize exposure to the drugs in question, all efforts to do so have paradoxically been associated with an increase in serious adverse events in the patients randomized to intermittent therapy, demonstrating that some "non-AIDS-defining" serious adverse events such as heart attack and stroke are linked to HIV replication. Thus, unless contraindicated for reasons of toxicity, patients started on ART should remain on ART.

At present, the U.S. Department of Health and Human Services Guidelines panel recommends that everyone with HIV infection be treated with ART and that therapy be initiated a soon as possible after diagnosis with some exceptions noted below. ART has been associated with a decrease in disease progression in patients at all stages of HIV infection and leads to a decrease in the risk of transmission of infection. In addition, one may wish to administer a 6-week course of therapy to uninfected individuals immediately following a high-risk exposure to HIV. The combination of tenofovir and emtricitabine is also licensed for pre-exposure prophylaxis in individuals at high risk of HIV infection, as is an injectable, long-acting formulation of cabotegravir that may be even more effective. For patients diagnosed with an opportunistic infection and HIV infection at the same time and a CD4+ count ≥50 cells/μL, one may consider a 2- to 4-week delay in the initiation of antiretroviral therapy during which time treatment is focused on the opportunistic infection. This delay may decrease the severity of any subsequent immune reconstitution inflammatory syndrome by lowering the antigenic burden of the opportunistic infection. This is particularly true for patients with TB or cryptococcal meningitis. For patients with advanced HIV infection (CD4+ <50 cells/μL), however, ART should be initiated as soon as possible.

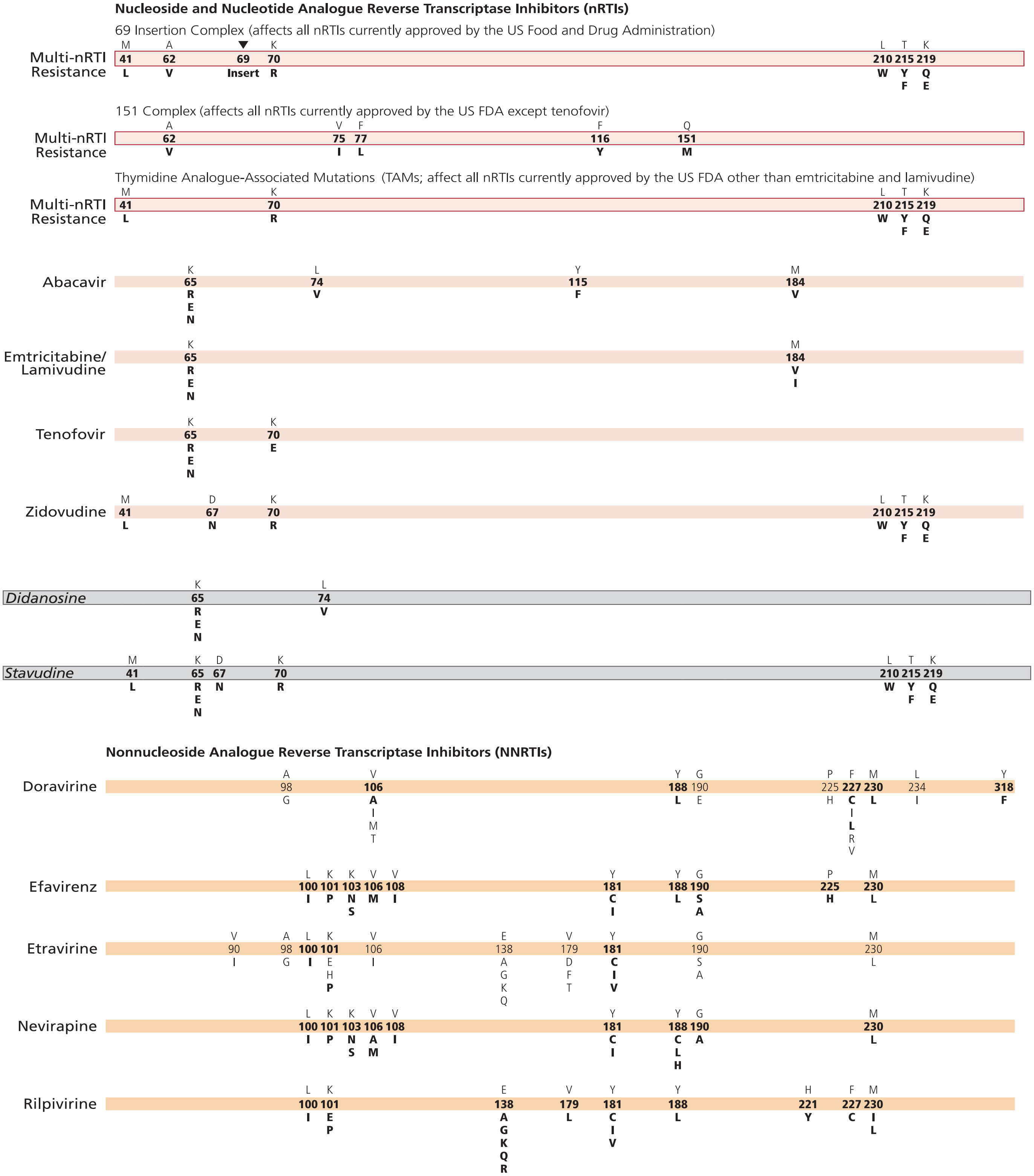
Nucleoside and Nucleotide Analogue Reverse Transcriptase Inhibitors (nRTIs)
69 Insertion Complex (affects all nRTIs currently approved by the US Food and Drug Administration)
Multi-nRTI Resistance
M 41 L
A 62 V
▼ 69 Insert
K 70 R
L 210 W
T 215 Y F
K 219 Q E
151 Complex (affects all nRTIs currently approved by the US FDA except tenofovir)
Multi-nRTI Resistance
A 62 V
V 75 I
F 77 L
F 116 Y
Q 151 M
Thymidine Analogue-Associated Mutations (TAMs; affect all nRTIs currently approved by the US FDA other than emtricitabine and lamivudine)
Multi-nRTI Resistance
M 41 L
K 70 R
L 210 W
T 215 Y F
K 219 Q E
Abacavir
K 65 R E N
L 74 V
Y 115 F
M 184 V
Emtricitabine/ Lamivudine
K 65 R E N
M 184 V I
Tenofovir
K 65 R E N
K 70 E
Zidovudine
M 41 L
D 67 N
K 70 R
L 210 W
T 215 Y F
K 219 Q E
Didanosine
K 65 R E N
L 74 V
Stavudine
M 41 L
K 65 R E N
D 67 N
K 70 R
L 210 W
T 215 Y F
K 219 Q E
Nonnucleoside Analogue Reverse Transcriptase Inhibitors (NNRTIs)
Doravirine
A 98 G
V 106 A I M T
Y 188 L
G 190 E
P 225 H
F 227 C I L R V
M 230 L
L 234 I
Y 318 F
Efavirenz
L 100 I
K 101 P
K 103 N S
V 106 M
V 108 I
Y 181 C I
Y 188 L
G 190 S A
P 225 H
M 230 L
Etravirine
V 90 I
A 98 G
L 100 I
K 101 E H P
V 106 I
E 138 A G K Q
V 179 D F T
Y 181 C I V
G 190 S A
M 230 L
Nevirapine
L 100 I
K 101 P
K 103 N S
V 106 A M
V 108 I
Y 181 C I
Y 188 C L H
G 190 A
M 230 L
Rilpivirine
L 100 I
K 101 E P
E 138 A G K Q R
V 179 L
Y 181 C I V
Y 188 L
H 221 Y
F 227 C
M 230 I L

FIGURE 208-47 Amino acid substitutions conferring resistance to antiretroviral drugs. For each amino acid residue, the letter above the bar indicates the amino acid associated with wild-type virus and the letter(s) below indicate the substitution(s) that confer viral resistance. The number shows the position of the mutation in the protein. Mutations selected by protease inhibitors in Gag cleavage sites are not listed. HR1, first heptad repeat; NAMs, nRTI-associated mutations; NNRTI, nonnucleoside reverse transcriptase inhibitor; nRTI, nucleoside reverse transcriptase inhibitor; PI, protease inhibitor. Amino acid abbreviations: A, alanine; C, cysteine; D, aspartate; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine. *(Reproduced with permission from AM Wensing et al: 2022 Update of the drug resistance mutations in HIV-1. 13, 2022. Reproduced with permission from IAS–USA.)*

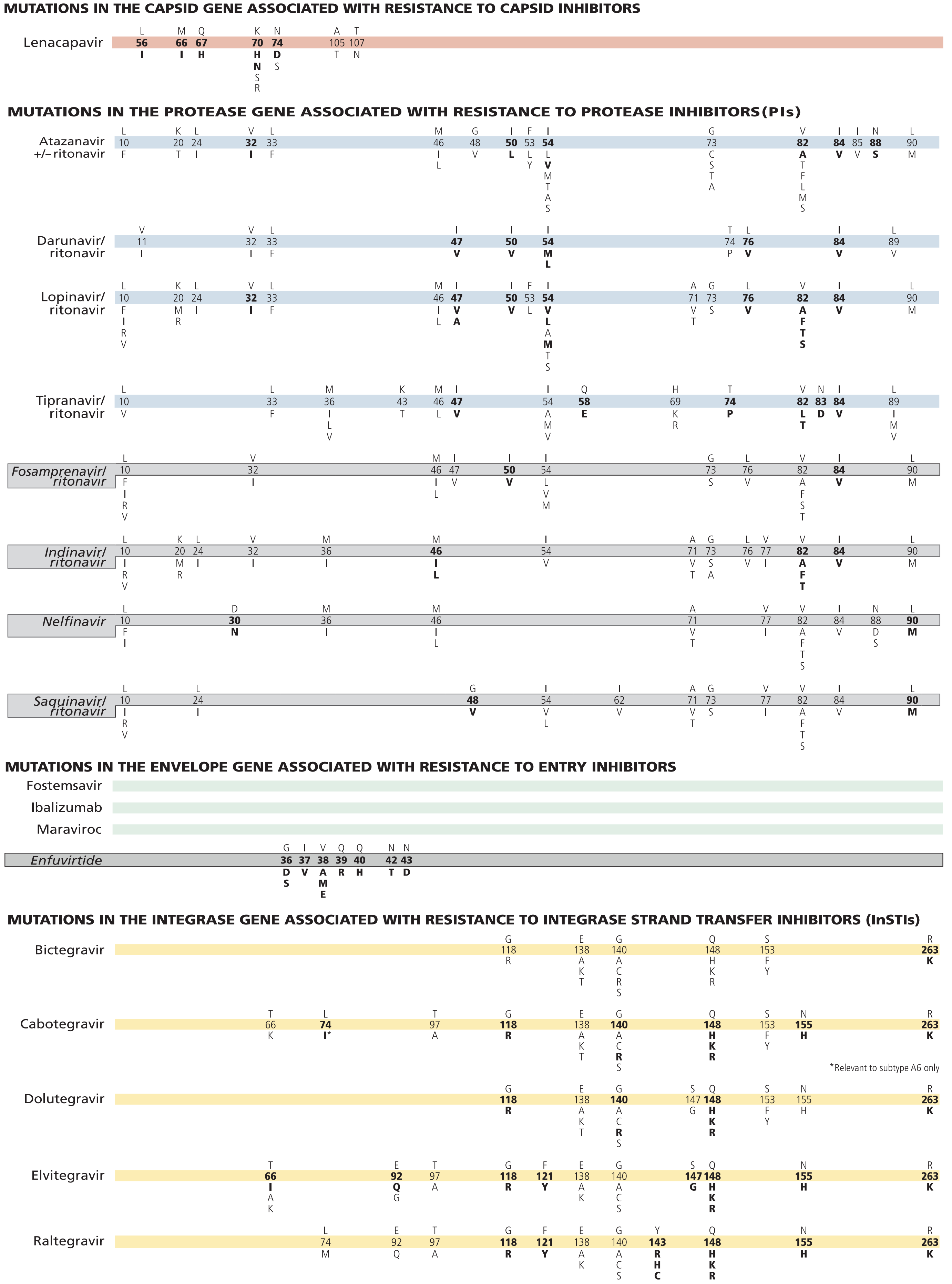
MUTATIONS IN THE CAPSID GENE ASSOCIATED WITH RESISTANCE TO CAPSID INHIBITORS
Lenacapavir
L56 I; M66 I; Q67 H; K70 H N S R; N74 D S; A105 T; T107 N
MUTATIONS IN THE PROTEASE GENE ASSOCIATED WITH RESISTANCE TO PROTEASE INHIBITORS(PIs)
Atazanavir +/– ritonavir
L10 F; K20 T; L24 I; V32 I; L33 F; M46 I L; G48 V; I50 L; F53 L Y; I54 L V M T A S; G73 C S T A; V82 A T F L M S; I84 V; I85 V; N88 S; L90 M
Darunavir/ ritonavir
V11 I; V32 I; L33 F; I47 V; I50 V; I54 M L; T74 P; L76 V; I84 V; L89 V
Lopinavir/ ritonavir
L10 F I R V; K20 M R; L24 I; V32 I; L33 F; M46 I L; I47 V A; I50 V; F53 L; I54 V L A M T S; A71 V T; G73 S; L76 V; V82 A F T S; I84 V; L90 M
Tipranavir/ ritonavir
L10 V; L33 F; M36 I L V; K43 T; M46 L; I47 V; I54 A M V; Q58 E; H69 K R; T74 P; V82 L T; N83 D; I84 V; L89 I M V
Fosamprenavir/ ritonavir
L10 F I R V; V32 I; M46 I L; I47 V; I50 V; I54 L V M; G73 S; L76 V; V82 A F S T; I84 V; L90 M
Indinavir/ ritonavir
L10 I R V; K20 M R; L24 I; V32 I; M36 I; M46 I L; I54 V; A71 V T; G73 S A; L76 V; V77 I; V82 A F T; I84 V; L90 M
Nelfinavir
L10 F I; D30 N; M36 I; M46 I L; A71 V T; V77 I; V82 A F T S; I84 V; N88 D S; L90 M
Saquinavir/ ritonavir
L10 I R V; L24 I; G48 V; I54 V L; I62 V; A71 V T; G73 S; V77 I; V82 A F T S; I84 V; L90 M
MUTATIONS IN THE ENVELOPE GENE ASSOCIATED WITH RESISTANCE TO ENTRY INHIBITORS
Fostemsavir
Ibalizumab
Maraviroc
Enfuvirtide
G36 D S; I37 V; V38 A M E; Q39 R; Q40 H; N42 T; N43 D
MUTATIONS IN THE INTEGRASE GENE ASSOCIATED WITH RESISTANCE TO INTEGRASE STRAND TRANSFER INHIBITORS (InSTIs)
Bictegravir
G118 R; E138 A K T; G140 A C R S; Q148 H K R; S153 F Y; R263 K
Cabotegravir
T66 K; L74 I*; T97 A; G118 R; E138 A K T; G140 A C R S; Q148 H K R; S153 F Y; N155 H; R263 K
*Relevant to subtype A6 only
Dolutegravir
G118 R; E138 A K T; G140 A C R S; S147 G; Q148 H K R; S153 F Y; N155 H; R263 K
Elvitegravir
T66 I A K; E92 Q G; T97 A; G118 R; F121 Y; E138 A K; G140 A C S; S147 G; Q148 H K R; N155 H; R263 K
Raltegravir
L74 M; E92 Q; T97 A; G118 R; F121 Y; E138 A K; G140 A C S; Y143 R H C; Q148 H K R; N155 H; R263 K

FIGURE 208-47 (Continued)

TABLE 208-23 Principles of Therapy of HIV Infection

1. Ongoing HIV replication leads to immune system damage, progression to AIDS, and systemic immune activation.
2. Plasma HIV RNA levels indicate the magnitude of HIV replication and the rate of CD4+ T cell destruction. CD4+ T cell counts indicate the current level of competence of the immune system.
3. Maximal suppression of viral replication is a goal of therapy; the greater the suppression the less likely the appearance of drug-resistant quasispecies.
4. The most effective therapeutic strategies involve the simultaneous initiation of combinations of effective anti-HIV drugs with which the patient has not been previously treated and that are not cross-resistant with antiretroviral agents that the patient has already received.
5. The antiretroviral drugs used in combination regimens should be used according to optimum schedules and dosages.
6. The number of available drugs is limited. Any decisions on antiretroviral therapy have a long-term impact on future options for the patient.
7. Women should receive optimal antiretroviral therapy regardless of pregnancy status.
8. The same principles apply to children and adults, but the treatment of HIV-infected children involves unique pharmacologic, virologic, and immunologic considerations.
9. Compliance is an important part of ensuring maximal effect from a given regimen. The simpler the regimen, the easier it is for the patient to be compliant.

Source: Modified from *Principles of Therapy of HIV Infection*, USPHS, and the Henry J. Kaiser Family Foundation.

Once the decision has been made to initiate therapy, the health care provider must decide which drugs to use as the first regimen. The decision regarding choice of drugs not only will affect the immediate response to therapy but also will have implications regarding options for future therapeutic regimens. The initial regimen is usually the most effective insofar as the virus has yet to be under any selective pressure to develop significant drug resistance. HIV is capable of rapidly developing resistance to any single agent, and therapy must be given as a multidrug combination. Given that patients can be infected with viruses that harbor drug resistance mutations, it is recommended that a viral genotype be done at the time of initiation of therapy to ensure that appropriate antiretroviral agents have been selected. The combination regimens currently recommended for initial therapy in most treatment-naïve patients are listed in **Table 208-24**. It is currently debated whether treatment-naïve individuals with <50 copies/mL of HIV RNA benefit from ART. While these individuals are at low risk of disease progression in the short term, they do have evidence of persistent immune activation that may have long-term consequences.

Following the initiation of therapy, one should expect a rapid, at least 1-log (10-fold) reduction in plasma HIV RNA levels within 1–2 months and then a slower decline in plasma HIV RNA levels to <50 copies/mL within 6 months. During this same time there should be a rise in the CD4+ T cell count of 100–150/cells μL that is also particularly brisk during the first month of therapy. Subsequently, one should anticipate a CD4+ T cell count increase of 50–100 cells/year until numbers approach normal. Many clinicians feel that failure to achieve these endpoints is an indication for a change in therapy. Other reasons for a change in therapy include a persistently declining CD4+ T cell count, a consistent increase in HIV RNA levels to >200 copies/mL, clinical deterioration, or drug toxicity **(Table 208-25)**. As in the case of initiating therapy, changing therapy may have a lasting impact on future therapeutic options. When changing therapy because of treatment failure (clinical progression or worsening laboratory parameters), it is important to attempt to provide a regimen with at least two new active drugs. This decision can be guided by resistance testing (see below). In the patient in whom a change is made for reasons of drug toxicity, a simple replacement of one drug is reasonable. It should be stressed that in attempting to sort out a drug toxicity it may be advisable to hold all therapy for a period of time to distinguish between drug toxicity and disease progression. Drug toxicity will usually begin to show signs of reversal within 1–2 weeks. Prior to changing a treatment regimen because of drug failure, it is important to ensure that the patient has been adherent to the prescribed regimen. As in the case of initial therapy, the simpler the new therapeutic regimen, the easier it is for the patient to be compliant. Plasma HIV RNA levels should be monitored within 2–4 weeks after initiation of ART or following a change in regimen, every 4–8 weeks until HIV RNA levels are suppressed to <200 copies/mL, and then every 3–6 months during therapy.

TABLE 208-24 Initial Combination Regimens Recommended for Most Treatment-Naïve Patients Regardless of HIV RNA Level or CD4 Count

For people who do not have a history of using Carbotegravir-LA as PrEP:
- Bictegravir + tenofovir alafenamide + emtricitabine (Biktarvy)
- Dolutegravir + tenofovir[a] + emtricitabine[b]
- Dolutegravir + lamivudine (Dovato) only for those with HIV RNA ≤ 500,000 copies/ml

For people who have a history of Carbotegravir-LA as PrEP, INSTI genotype resistance testing should be performed; if starting prior to resistance testing results:
- Darunavir/cobicistat or darunavir/ritonavir + tenofovir[a] + emtricitabine[b]

[a]Tenofovir alafenamide and tenofovir disoproxil fumarate are two forms of tenofovir approved by FDA. Tenofovir alafenamide has fewer bone and renal toxicities, while tenofovir disoproxil fumarate is associated with lower lipid levels. [b]Lamivudine may substitute for emtricitabine and vice versa.

Source: Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adults and Adolescents, USPHS.

TABLE 208-25 Indications for Changing Antiretroviral Therapy in Patients with HIV Infection[a]

Less than a 1-log drop in plasma HIV RNA by 4 weeks following the initiation of therapy

A reproducible significant increase (defined as threefold or greater) from the nadir of plasma HIV RNA level not attributable to intercurrent infection, vaccination, or test methodology

Persistently declining CD4+ T cell numbers

Clinical deterioration

Side effects

[a]Generally speaking, a change should involve the initiation of at least two drugs felt to be effective in the given patient. The exception to this is when a change is being made to manage toxicity, in which case a single substitution is reasonable.

Source: Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adults and Adolescents, USPHS.

In order to determine an optimal therapeutic regimen for initial therapy or for a patient on a failing regimen, one may attempt to measure antiretroviral drug susceptibility through genotyping or phenotyping of HIV quasispecies and to determine adequacy of dosing through measurement of drug levels. Genotyping may be done through cDNA sequencing. Phenotypic assays typically measure the enzymatic activity of viral enzymes in the presence or absence of different concentrations of different drugs and have also been used to determine co-receptor tropism. These assays will generally detect quasispecies present at a frequency of ≥10%. Next-generation sequencing may allow detection of quasispecies at frequencies down to 1%. It is generally recommended that resistance testing be used in confirming initial therapy choices in settings where the risk of transmission of resistant virus is high (such as the United States and Europe) and in determining new regimens for patients experiencing virologic failure while on therapy. Resistance testing may be of particular value in distinguishing drug-resistant virus from poor patient compliance. Due to the rapid rate at which drug-resistant viruses revert to wild-type, it is recommended that resistance testing performed in the setting of drug failure be conducted while the patient is still on the failing regimen. Measurement of plasma drug levels can also be used to tailor individual treatment. The inhibitory quotient, defined as the trough blood level/IC_{50} of the patient's virus, is used by some to determine the adequacy of dosing of a given treatment regimen. Despite the best of efforts there will still be patients with ongoing high levels of HIV replication while receiving the best available therapy. These patients will receive benefit from remaining on antiretroviral therapy even though it is not fully suppressive.

In addition to the licensed medications discussed above, a large number of experimental agents are being evaluated as possible therapies for HIV infection. Therapeutic strategies are being developed to interfere with virtually every step of the replication cycle of the virus (Fig. 208-3) and in an attempt to eliminate the reservoir of infected cells to "cure" HIV infection. In addition to directly acting antiviral drugs, other strategies, generically referred to as "immune-based therapies," are being developed as a complement to antiviral therapy. Among the antiviral agents in early clinical trials are additional nucleoside and nucleotide analogues, protease inhibitors, fusion inhibitors, receptor and co-receptor antagonists, and integrase inhibitors—as well as new antiviral strategies including long-acting injectables, antisense nucleic acids, and maturation inhibitors. Among the immune-based therapies being evaluated are monoclonal antibodies, bone marrow transplantation, adoptive transfer of lymphocytes genetically modified to resist infection or enhance HIV-specific immunity, active immunotherapy with inactivated HIV or its components, IL-7, and IL-15.

HIV AND THE HEALTH CARE WORKER

Health care workers, especially those who deal with large numbers of patients with HIV, have a small but definite risk of becoming infected with HIV as a result of professional activities (see "Occupational Transmission of HIV: Health Care Workers, Laboratory Workers, and the Health Care Setting," above).

In the United States, 58 health care workers for whom case investigations have been completed have had documented seroconversions to HIV following occupational exposures. Only one of these has occurred since 1999. Approximately 85% of the exposures resulting in infection have been due to percutaneous (puncture/cut injury) exposures to HIV-infected blood. In addition, at least 150 possible cases of occupationally acquired HIV infection have been reported among health care personnel in the United States. The number of these workers who actually acquired their infection through occupational exposures is not known. Taken together, data from several large studies suggest that the risk of HIV infection following a percutaneous exposure to HIV-contaminated blood in an individual who does not receive postexposure prophylaxis (PEP) is ~0.23%, and after a mucous membrane exposure, ~0.09%. Although episodes of HIV transmission after nonintact skin exposure have been documented, the average risk for transmission by this route has not been precisely quantified but is estimated to be less than the risk for mucous membrane exposures. The risk for transmission after exposure to body fluids or tissues other than HIV-infected blood also has not been quantified but is probably considerably lower than for blood exposures. A seroprevalence survey of 3420 orthopedic surgeons, 75% of whom practiced in an area with a relatively high prevalence of HIV infection and 39% of whom reported percutaneous exposure to patient blood, usually through an accident involving a suture needle, failed to reveal any cases of possible occupational infection, suggesting that the risk of infection with a suture needle may be considerably less than that with a blood-drawing (hollow-bore) needle.

Most cases of health care worker seroconversion have occurred as a result of needle-stick injuries. When one considers the circumstances that result in needle-stick injuries, it is immediately obvious that adhering to the standard guidelines for dealing with sharp objects would result in a significant decrease in this type of accident. In one study, 27% of needle-stick injuries resulted from improper disposal of the needle (more than half of these were due to recapping the needle), 23% occurred during attempts to start an IV line, 22% occurred during blood drawing, 16% were associated with an IM or SC injection, and 12% were associated with giving an IV infusion.

Occupational exposures to HIV should be considered as a medical emergency to ensure timely postexposure management and administration of PEP. A delay of even sending 72 h in the initiation of PEP may be the difference preventing and not preventing infection. Recommendations regarding PEP must take into account that a variety of circumstances determine the risk of transmission of HIV following occupational exposure. In this regard, several factors have been associated with an increased risk for occupational transmission of HIV infection, including deep injury, the presence of visible blood on the instrument causing the exposure, injury with a device that had been placed in the vein or artery of the source patient, and advanced HIV disease in the source patient. Other important considerations when considering PEP in the health care worker include known or suspected pregnancy or breast-feeding, the possibility of exposure to drug-resistant virus, and the toxicities of different PEP regimens. Regardless of the decision to use PEP, the wound should be cleansed immediately and antiseptic applied. If a decision is made to offer PEP, U.S. Public Health Service guidelines recommend that PEP regimens contain three (or more) antiretroviral drugs administered for a 4-week duration for all occupational exposures to HIV. Detailed guidelines are available from the *Updated U.S. Public Health Service Guidelines for the Management of Occupational Exposures to HIV and Recommendations for Postexposure Prophylaxis* (CDC, 2018). The report emphasizes the importance of adherence to PEP when it is indicated, and close follow-up of exposed workers should be provided including counseling, baseline and follow-up HIV testing, and monitoring for drug toxicity. Follow-up appointments should begin within 72 h of an HIV exposure and may be concluded 4 months after exposure. For consultation on the treatment of occupational exposures to HIV and other bloodborne pathogens, the clinician managing the exposed patient can call the National Clinicians' Post-Exposure Prophylaxis Hotline (PEPline) at 888-448-4911. This service is available 24 hours a day at no charge. (Additional information is available at *www.nccc.ucsf.edu*.) PEPline support may be especially useful in challenging situations, such as when drug-resistant HIV strains are suspected or if the health care worker is pregnant.

Health care workers can minimize their risk of occupational HIV infection by following the CDC guidelines of June 2015, which include adherence to universal precautions and assuming that blood and other body fluids from all patients are potentially infectious. Therefore, the following infection control precautions should be adhered to at all times: (1) routinely use barriers (such as gloves and/or goggles) when anticipating contact with blood or body fluids; (2) immediately wash hands and other skin surfaces after contact with blood or body fluids; and (3) carefully handle and dispose of sharp instruments during and after use. For further information contact the CDC at 800-CDC-INFO (232-4636) or see *www.cdc.gov/cdc-info/*. The risk of HBV infection following a needle-stick injury from a hepatitis antigen–positive patient is much higher than the risk of HIV infection (see "Transmission," above). There are multiple examples of needle-stick injuries where the patient was positive for both HBV and HIV and the health care worker became infected only with HBV. For these reasons, it is advisable, given the high prevalence of HBV infection in people with HIV, that all health care workers dealing with patients with HIV be immunized with the HBV vaccine.

TB is another infection common to patients with HIV that can be transmitted to the health care worker. For this reason, all health care workers should know their PPD status, have it checked yearly, and, where appropriate, receive 6 months of isoniazid treatment if their skin test converts to positive. In addition, all patients in whom a diagnosis of active pulmonary TB is being entertained should be placed immediately in respiratory isolation, pending results of the diagnostic evaluation. The emergence of drug-resistant organisms, including extensively drug-resistant TB strains, has made TB an increasingly important problem for health care workers. This is particularly true for the health care worker with pre-existing HIV infection.

HIV PREVENTION

Many proven interventions, usually applied in combination, have a role in preventing the transmission of HIV **(Fig. 208-48)**. Education, counseling, and behavior modification are the cornerstones of any HIV prevention strategy. A major problem in the United States and elsewhere is that many infections are passed on by those who do not know that they are infected. Of the ~1.2 million persons in the United States

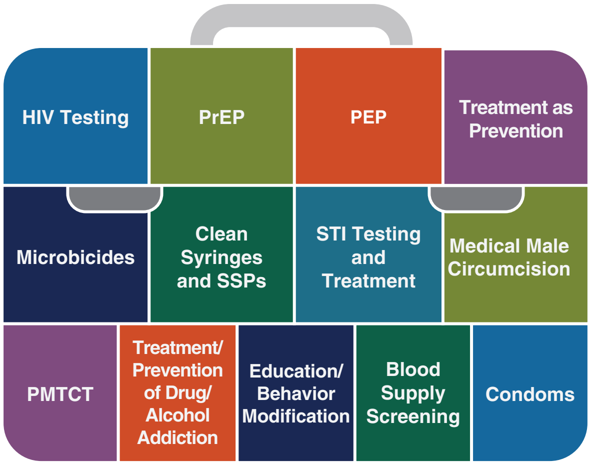
HIV Testing
PrEP
PEP
Treatment as Prevention
Microbicides
Clean Syringes and SSPs
STI Testing and Treatment
Medical Male Circumcision
PMTCT
Treatment/ Prevention of Drug/ Alcohol Addiction
Education/ Behavior Modification
Blood Supply Screening
Condoms

FIGURE 208-48 The HIV prevention "toolkit." See text for detailed description. PMTCT, prevention of mother-to-child transmission of HIV; PEP, postexposure prophylaxis with antiretroviral drugs; PrEP, pre-exposure prophylaxis with antiretroviral drugs; SSPs, syringe services programs. *(From: The White House. 2021. National HIV/AIDS Strategy for the United States 2022–2025. Washington, DC.)*

with HIV, it is estimated that ~13% do not know their HIV status and that a substantial proportion of all new infections are transmitted by those people. In this regard, the CDC has recommended HIV testing as part of routine medical care and that all individuals between the ages of 13 and 64 years be tested at least once. These individuals should be informed of the testing and be tested without the need for written informed consent. Each individual can "opt out" of testing; however, testing should otherwise be routinely administered. Individuals who are practicing high-risk behavior should be tested more often and should use pre-exposure prophylaxis (PrEP) (see below). Partners engaged in monogamous sexual relationships who wish to be assured of safety should both be tested for HIV antibody. If both are negative, it must be understood that any divergence from monogamy puts both partners at risk; open discussion of the importance of honesty in such relationships should be encouraged.

When the HIV status of either partner is not known, or when one partner is positive, there are a number of options. *Use of condoms* can markedly decrease the chance of HIV transmission. It should be remembered that condoms are not 100% effective in preventing transmission of HIV infection, and there is a ~10% failure rate of condoms used for contraceptive purposes. Most condom failures result from breakage or improper usage, such as not wearing the condom for the entire period of intercourse. Latex condoms are preferable since viruses have been shown to leak through natural skin condoms. Petroleum-based gels should never be used for lubrication of the condom, since they increase the likelihood of condom rupture.

Microbicides composed of gels or rings containing antiretroviral drugs have been shown to be variably efficacious in preventing acquisition of HIV infection in women engaging in vaginal intercourse. The considerable degree of variability in efficacy relates to the generally poor adherence of participants to the use of the intervention. One product, a vaginal ring that releases the antiretroviral drug dapivirine from the ring into the vagina slowly over 28 days, has been recommended by WHO as an additional prevention choice for women at substantial risk of HIV infection as part of combination prevention approaches.

Large, prospective clinical trials have clearly demonstrated that ART for people with HIV has an important role in HIV prevention. The initial results of the HPTN 052 clinical trial published in 2011 demonstrated a 96% reduction in HIV transmission risk among heterosexual HIV-discordant couples where the partner with HIV started ART immediately versus delayed ART initiation. The final results of HPTN 052, published in 2016, reported no HIV transmissions within these couples when the partner with HIV had a suppressed viral load (defined as having a viral load of <400 copies of HIV RNA per milliliter). Three subsequent studies reported similar results, with no genetically linked infections while the partner with HIV was virally suppressed even though couples were engaging in sex without a condom and not using PrEP. These three studies included >500 HIV-discordant heterosexual couples and >1100 HIV-discordant couples of men who have sex with men. Combined, these couples engaged in >125,000 sex acts without a condom or PrEP over more than 2600 couple-years of observation. Collectively, the studies demonstrated that if the viral load of the infected partner is decreased to below detectable levels by antiretroviral therapy, sexual transmission to the uninfected partner does not occur. This is true for heterosexuals and men who have sex with men, leading, as noted above, to the commonly used phrase "undetectable equals untransmittable" or U = U.

Pre-exposure prophylaxis (PrEP) with antiretroviral medication also is highly effective in preventing HIV acquisition by at-risk uninfected men who have sex with men and heterosexual men and women. Accumulated data indicate that high adherence to a PrEP regimen of emtricitabine + tenofovir disoproxil fumarate, taken as 1 pill per day or on demand (immediately before and following a sexual encounter), is 99% effective in preventing HIV acquisition if subjects adhere strictly to the regimen. Subsequent studies indicated similar, if not better, efficacy with cabotegravir injections given every 2 months as a maintenance regimen. A recent phase 3 clinical trial conducted in South Africa and Uganda indicated that the twice-yearly injectable HIV capsid inhibitor lenacapavir demonstrated even greater efficacy in preventing HIV infection in cisgender women. Additional phase 3 studies of lenacapavir for PrEP are being conducted in cisgender men, transgender women, transgender men, and gender-nonbinary individuals. More limited data demonstrate the utility of PrEP for people who inject drugs. CDC estimates that approximately 1.2 million people in the United States are at "substantial" risk for HIV infection and should be counseled about PrEP.

Adult male circumcision, which has been shown to result in a 50–65% reduction in HIV acquisition in the circumcised subject, is currently being pursued, particularly in developing nations, as a component of HIV prevention (see above). The most effective way to prevent transmission of HIV infection among IDUs is to stop the use of injectable drugs. Unfortunately, that is extremely difficult to accomplish unless the individual enters a treatment program. For those who will not or cannot participate in a drug treatment program and who will continue to inject drugs, the avoidance of sharing of needles and other paraphernalia ("works") is the next best way to avoid transmission of infection. However, the cultural and social factors that contribute to the sharing of paraphernalia are complex and difficult to overcome. Under these circumstances, paraphernalia should be cleaned after each usage with a virucidal solution, such as sodium hypochlorite (undiluted household bleach). *Needle exchange programs* have been highly successful in decreasing HIV transmission among injection drug users without increasing the use of injection drugs. As noted, above, oral PrEP also is effective in preventing acquisition of HIV infections among IDUs. It is important for IDUs to be tested for HIV infection and counseled to avoid transmission to their sexual partners. Prevention of transmission through blood or blood products and prevention of mother-to-child transmission are discussed in "Transmission," above.

■ HIV VACCINES

There is currently no safe and effective vaccine approved for the prevention of HIV infection. Successful vaccines for other diseases are predicated on the assumptions that the body can mount an adequate immune response to the microbe or virus in question during natural infection and that the vaccine will mimic the natural response to infection. Even with serious diseases, such as smallpox, poliomyelitis, measles, and influenza among others, the body usually clears the infectious agent and provides protection, which is usually lifelong against future exposure to the same pathogen. Unfortunately, this is not the case with HIV infection since the natural immune response to HIV infection is unable to clear the virus from the body and cases of superinfection are not rare.

Some of the factors that contribute to the problematic nature of developing a preventive HIV vaccine are (1) the high mutability of the virus; (2) the fact that the infection can be transmitted by cell-free or cell-associated virus; (3) the fact that the HIV provirus integrates itself into the genome of the target cell and may remain in a latent form unexposed to the immune system; (4) the likely need for the development of effective mucosal immunity; and, importantly, (5) the difficulty that the immune system has in readily mounting broadly neutralizing antibodies in response to natural infection with HIV (see below).

Early attempts to develop a vaccine with the envelope protein gp120 aimed at inducing neutralizing antibodies in humans were unsuccessful; the elicited antisera failed to neutralize primary isolates of HIV. In this regard, two phase 3 trials were undertaken in the United States and Thailand using soluble gp120, and the vaccines failed to protect human volunteers from HIV infection. In addition, two separate vaccine trials aimed at eliciting CD8+ T cell responses to prevent infection and, if unsuccessful in preventing infection, to control postinfection viremia, also failed at both goals. In 2009, a vaccine using a poxvirus vector prime expressing various viral proteins followed by an envelope protein boost was assessed in a 16,000-person clinical trial (RV144) conducted in Thailand among predominantly low-HIV-prevalence heterosexuals. The vaccine provided the first positive, albeit very modest, signal ever reported in an HIV vaccine trial, showing 31% protection against acquisition of infection. Such a result is certainly not sufficient justification for clinical use of the vaccine; however, it served as an important first step in the direction of the development of a safe and effective vaccine against HIV infection.

Follow-up studies of RV144 indicate that nonneutralizing or weakly neutralizing antibody responses against certain constant epitopes in the otherwise highly variable V1–V2 region of the HIV envelope may be associated with the modest degree of protection observed in that clinical trial. Three additional similar studies were undertaken in high-HIV-prevalence countries in sub-Saharan Africa as well as in the Americas and certain European countries in attempts to improve on the results of RV144 by a variety of approaches, including increasing the number of vaccine boosts with envelope protein, the use of mosaic antigens, and the addition of adjuvant. Unfortunately, all three of these phase 3 studies of candidate vaccines failed to show efficacy. Another study was terminated early due to lack of efficacy.

An area of HIV vaccine research that is currently being actively pursued is the attempt to induce broadly neutralizing antibodies by developing as immunogens for vaccination certain epitopes on the HIV envelope that are the targets of naturally occurring broadly neutralizing antibodies during HIV infection (Fig. 208-30). It is curious that only about 20% of people with HIV develop broadly neutralizing antibodies in response to natural infection and they do so only after 2–3 years of ongoing infection. By the time these antibodies appear, they can neutralize a broad range of primary HIV isolates, but they appear to be ineffective against the autologous virus in the infected subject. Upon close examination, these broadly neutralizing antibodies manifest a high degree of somatic mutations that accumulated over time and are responsible for their affinity maturation and broadly neutralizing capacity. The goal of current efforts is to develop the conformationally correct HIV envelope epitopes that, when used as immunogens, would direct the immune response of an uninfected individual to the production of broadly neutralizing antibodies over a reasonable time frame by sequential immunizations. It remains to be seen whether this approach will be feasible.

■ FURTHER READING

Bekker LG et al: HIV infection. Nat Rev Dis Primers 9:42, 2023.

Beyrer C et al: Is HIV epidemic control by 2030 realistic? Lancet HIV 7:e489, 2024.

Centers for Disease Control and Prevention (CDC): Clinical Guidance for PrEP. Available at *www.cdc.gov/hivnexus/hcp/prep/*.

Centers for Disease Control and Prevention (CDC): Clinical Guidance for PEP. Available at *www.cdc.gov/hivnexus/hcp/pep/*.

Centers for Disease Control and Prevention (CDC): Clinical Care of HIV. Available at *www.cdc.gov/hivnexus/hcp/clinical-care/*.

Cohn LB et al: Biology of the HIV-1 latent reservoir and implications for cure strategies. Cell Host Microbe 27:519, 2020.

Collins DR et al: CD8+ T cells in HIV control, cure and prevention. Nat Rev Immunol 20:471, 2020.

Eisinger RW et al: HIV viral load and transmissibility of HIV infection: Undetectable equals untransmittable. JAMA 321:451, 2019.

Fauci AS, Lane HC: Four decades of HIV/AIDS—much accomplished, much to do. N Engl J Med 383:1, 2020.

Grosso TM et al: HIV and aging, biological mechanisms, and therapies: What do we know? AIDS Rev 25:79, 2022.

Haynes BF et al: Strategies for HIV-1 vaccines that induce broadly neutralizing antibodies. Nat Rev Immunol 23:142, 2023.

Horberg M et al: Primary care guidance for persons with human immunodeficiency virus: 2024 update by the HIV Medicine Association of the Infectious Diseases Society of America. Clin Infect Dis, 2024. [Epub ahead of print.]

Kazer SW: Evolution and diversity of immune responses during acute HIV Infection. Immunity 53:908, 2020.

Landovitz RJ et al: Prevention, treatment and cure of HIV infection. Nat Rev Microbiol 21:657, 2023.

Lisco A et al: Reappraisal of idiopathic CD4 lymphocytopenia at 30 years. N Engl J Med 388:1681, 2023.

Moir S, Fauci AS: B-cell responses to HIV infection. Immunol Rev 275:33, 2017.

Nkolola JP, Barouch DH: Prophylactic HIV-1 vaccine trials: Past, present, and future. Lancet HIV 11:e117, 2024.

Panel on Opportunistic Infections in Adults and Adolescents with HIV: Guidelines for the Prevention and Treatment of Opportunistic Infections in Adults and Adolescents with HIV. Available at *clinicalinfo.hiv.gov/en/guidelines/hiv-clinical-guidelines-adult-and-adolescent-opportunistic-infections/whats-new*.

Saez-Cirion A, Sereti I: Immunometabolism and HIV-1 pathogenesis: Food for thought. Nat Rev Immunol 21:5, 2021.

UN Joint Programme On HIV/AIDS (UNAIDS): 2024 global AIDS report — The Urgency of Now: AIDS at a Crossroads. Available at *www.unaids.org/en/resources/documents/2024/global-aids-update-2024*.

U.S. Department of Health and Human Services Panel on Antiretroviral Guidelines for Adults and Adolescents: Guidelines for the use of antiretroviral agents in adults and adolescents living with HIV. Available at *clinicalinfo.hiv.gov/en/guidelines/hiv-clinical-guidelines-adult-and-adolescent-arv/whats-new*.

Venter WDF et al: The long wait for long-acting HIV prevention and treatment formulations. Lancet HIV 11:e711, 2024.

Section 15 Infections Due to RNA Viruses

209

Viral Gastroenteritis

Umesh D. Parashar, Roger I. Glass

Acute infectious gastroenteritis is a common illness that affects persons of all ages worldwide. It is a leading cause of death among children in developing countries, accounting for an estimated 0.5 million deaths each year, and is responsible for up to 6–8% of all hospitalizations among children in industrialized countries, including the United States. Elderly persons, especially those with debilitating health conditions, also are at risk of severe complications and death from acute gastroenteritis. Among healthy young adults, acute gastroenteritis is rarely fatal but incurs substantial medical and social costs, including those of time lost from work.

Several enteric viruses have been recognized as important etiologic agents of acute infectious gastroenteritis **(Table 209-1, Fig. 209-1)**.

TABLE 209-1 Viral Causes of Gastroenteritis among Humans

VIRUS	FAMILY	GENOME	PRIMARY AGE GROUP AT RISK	CLINICAL SEVERITY	DETECTION ASSAYS
Group A rotavirus	Reoviridae	Double-strand segmented RNA	Children <5 years	+ + +	EIA (commercial), RT-PCR, EM, PAGE
Norovirus	Caliciviridae	Positive-sense single-strand RNA	All ages	+ +	RT-PCR, EM, EIA (commercial)
Sapovirus	Caliciviridae	Positive-sense single-strand RNA	Children <5 years	+	RT-PCR, EM
Astrovirus	Astroviridae	Positive-sense single-strand RNA	Children <5 years	+	EIA, RT-PCR, EM
Adenovirus (mainly types 40 and 41)	Adenoviridae	Double-strand DNA	Children <5 years	+/+ +	EIA (commercial), PCR, EM

Abbreviations: EIA, enzyme immunoassay; EM, electron microscopy; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; RT-PCR, reverse-transcription PCR.

Although most viral gastroenteritis is caused by RNA viruses, the DNA viruses that are occasionally involved (e.g., adenovirus types 40 and 41) are included in this chapter. Illness caused by these viruses is characterized by the acute onset of vomiting and/or diarrhea, which may be accompanied by fever, nausea, abdominal cramps, anorexia, and malaise. As shown in Table 209-2, several features can help distinguish gastroenteritis caused by viruses from that caused by bacterial agents. However, the distinction based on clinical and epidemiologic parameters alone is often difficult, and laboratory tests are required to confirm the diagnosis.

HUMAN CALICIVIRUSES

Etiologic Agent The Norwalk virus is the prototype strain of a group of small (27–40 nm), nonenveloped, round, icosahedral viruses with relatively amorphous surface features on visualization by electron microscopy. Molecular cloning and characterization have demonstrated that the viruses have a single, positive-strand RNA genome ~7.5 kb in length and possess a single virion-associated protein—similar to that of typical caliciviruses—with a molecular mass of 60 kDa. On the basis of these molecular characteristics, these viruses are presently classified into two genera belonging to the family Caliciviridae, the *noroviruses* and the *sapoviruses*, which are further classified into genogroups and genotypes. Of the 10 recognized norovirus genogroups in humans and animals, 35 different genotypes belonging to 5 genogroups (GI, GII, GIV, GVIII, and GIX) are known to infect humans.

Epidemiology Infections with human caliciviruses are common worldwide, and most adults have antibodies to these viruses. Antibody is acquired at an earlier age in developing countries—a pattern consistent with the presumed fecal–oral mode of transmission. Infections occur year-round, although, in temperate climates, a distinct increase has been noted in cold-weather months. Noroviruses may be the most common infectious agents of mild gastroenteritis in the community and affect all age groups, whereas sapoviruses primarily cause gastroenteritis in children. Noroviruses also cause traveler's diarrhea, and outbreaks have occurred among military personnel deployed to various parts of the world. The limited data available indicate that norovirus may be the second most common viral agent (after rotavirus) among young children and the most common agent among older children and adults. In the United States and some other developed countries, with the decline in severe rotavirus disease following implementation of a rotavirus vaccination program, norovirus has become the leading cause of medically attended gastroenteritis in young children. Noroviruses are also recognized as the major cause of epidemics of gastroenteritis worldwide. In the United States, ~50% of all reported outbreaks of gastroenteritis are caused by noroviruses.

Norovirus is transmitted predominantly by the fecal–oral route but is also present in vomitus. Because an inoculum with very few viruses can be infectious, transmission can occur by aerosolization, by contact with contaminated fomites, and by person-to-person contact. Viral shedding and infectivity are greatest during the acute illness, but challenge studies with Norwalk virus in volunteers indicate that viral antigen may be shed by asymptomatically infected persons and also by symptomatic persons before the onset of symptoms and for several weeks after the resolution of illness. Viral shedding can be prolonged in immunocompromised individuals.

Pathogenesis The exact sites and cellular receptors for attachment of viral particles have not been determined. Data suggest that carbohydrates that are similar to human histo-blood group antigens (HBGA) and are present on the gastroduodenal epithelium of individuals with the secretor phenotype may serve as ligands for the attachment of norovirus. Additional studies must more fully elucidate norovirus–carbohydrate interactions, including strain-specific variations. After the infection of volunteers, reversible lesions are noted in the upper jejunum, with broadening and blunting of the villi, shortening of the microvilli, vacuolization of the lining epithelium, crypt hyperplasia,

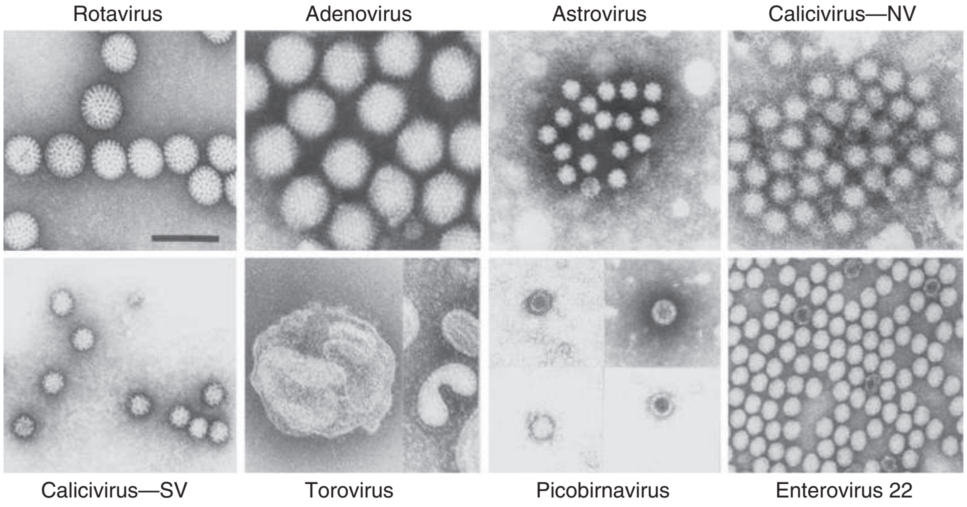

FIGURE 209-1 Viral agents of gastroenteritis. NV, norovirus; SV, sapovirus.

TABLE 209-2 Characteristics of Gastroenteritis Caused by Viral and Bacterial Agents

FEATURE	VIRAL GASTROENTERITIS	BACTERIAL GASTROENTERITIS
Setting	Incidence similar in developing and developed countries	More common in settings with poor hygiene and sanitation
Infectious dose	Low (10–100 viral particles) for most agents	High ($>10^5$ bacteria) for *Escherichia coli*, *Salmonella*, *Vibrio*; medium (10^2–10^5 bacteria) for *Campylobacter jejuni*; low (10–100 bacteria) for *Shigella*
Seasonality	In temperate climates, winter seasonality for most agents; year-round occurrence in tropical areas	More common in summer or rainy months, particularly in developing countries with a high disease burden
Incubation period	1–3 days for most agents; can be shorter for norovirus	1–7 days for common agents (e.g., *Campylobacter*, *E. coli*, *Shigella*, *Salmonella*); a few hours for bacteria producing preformed toxins (e.g., *Staphylococcus aureus*, *Bacillus cereus*)
Reservoir	Primarily humans	Depending on bacterial species, human (e.g., *Shigella*, *Salmonella*), animal (e.g., *Campylobacter*, *Salmonella*, *E. coli*), and water (e.g., *Vibrio*) reservoirs exist
Fever	Common with rotavirus and norovirus; uncommon with other agents	Common with agents causing inflammatory diarrhea (e.g., *Salmonella*, *Shigella*)
Vomiting	Prominent and can be the only presenting feature, especially in children	Common with bacteria producing preformed toxins; less prominent in diarrhea due to other agents
Diarrhea	Common; nonbloody in almost all cases	Prominent and occasionally bloody with agents causing inflammatory diarrhea
Duration	1–3 days for norovirus and sapovirus; 2–8 days for other viruses	1–2 days for bacteria producing preformed toxins; 2–8 days for most other bacteria
Diagnosis	This is often a diagnosis of exclusion in clinical practice. Commercial enzyme immunoassays are available for detection of rotavirus and adenovirus, but identification of other agents is limited to research and public health laboratories.	Fecal examination for leukocytes and blood is helpful in differential diagnosis. Culture of stool specimens, sometimes on special media, can identify several pathogens. Molecular techniques are useful epidemiologic tools but are not routinely used in most laboratories.
Treatment	Supportive therapy to maintain adequate hydration and nutrition should be given. Antibiotics and antimotility agents are contraindicated.	Supportive hydration therapy is adequate for most patients. Antibiotics are recommended for patients with dysentery caused by *Shigella* or diarrhea caused by *Vibrio cholerae* and for some patients with *Clostridium difficile* colitis.

and infiltration of the lamina propria by polymorphonuclear neutrophils and lymphocytes. The lesions persist for at least 4 days after the resolution of symptoms and are associated with malabsorption of carbohydrates and fats and a decreased level of brush-border enzymes. Adenylate cyclase activity is not altered. No histopathologic changes are seen in the stomach or colon, but gastric motor function is delayed, and this alteration is believed to contribute to the nausea and vomiting that are typical of this illness.

Clinical Manifestations Gastroenteritis caused by human caliciviruses has a sudden onset following an average incubation period of 24 h (range, 12–72 h). The illness generally lasts 12–60 h and is characterized by one or more of the following symptoms: nausea, vomiting, abdominal cramps, and diarrhea. Vomiting is more prevalent among children, whereas a greater proportion of adults develop diarrhea. Constitutional symptoms are common, including headache, fever, chills, and myalgias. The stools are characteristically loose and watery, without blood, mucus, or leukocytes. White cell counts are generally normal; rarely, leukocytosis with relative lymphopenia may be observed. Death is a rare outcome and usually results from severe dehydration in vulnerable persons (e.g., elderly patients with debilitating health conditions).

Immunity Approximately 50% of persons challenged with Norwalk virus become ill and acquire short-term immunity against the infecting strain. In early human volunteer studies, immunity to Norwalk virus appeared to correlate inversely with level of antibody; i.e., persons with higher levels of preexisting antibody to Norwalk virus were more susceptible to illness on rechallenge. This paradoxical observation was later explained by data indicating that some individuals have a genetic predisposition to illness, with specific HBGA phenotypes influencing susceptibility to norovirus infection. Contemporary data show that functional antibodies that block norovirus binding to HBGAs correlate with protective immunity in human volunteer challenge and vaccination studies. Furthermore, initial studies have demonstrated that norovirus grown in vitro in the newly developed human intestinal enteroid (HIE) cell-based system can be neutralized by sera containing blocking antibodies.

Diagnosis Cloning and sequencing of the genomes of Norwalk and several other human caliciviruses have allowed the development of assays based on polymerase chain reaction (PCR) for detection of virus in stool and vomitus. PCR-based detection assays for norovirus combined with those for multiple enteric pathogens are commercially available and are increasingly used in clinical settings. Virus-like particles (VLPs) produced by expression of capsid proteins in a recombinant baculovirus vector have been used to develop enzyme immunoassays (EIAs) for detection of virus in stool or a serologic response to a specific viral antigen. Commercial EIA kits for detection in stool have limited sensitivity and usefulness in clinical practice and are of greatest utility in outbreaks, in which many specimens are tested and only a few need be positive to identify norovirus as the cause.

TREATMENT

Infections with Norwalk and Related Human Caliciviruses

The disease is self-limited, and oral rehydration therapy is generally adequate. If severe dehydration develops, IV fluid therapy is indicated. No specific antiviral therapy is available.

Prevention Epidemic prevention relies on situation-specific measures, such as control of contamination of food and water, exclusion of ill food handlers, and reduction of person-to-person spread through good personal hygiene and disinfection of contaminated fomites. The role of immunoprophylaxis is not clear, given the lack of long-term immunity from natural disease, but efforts to develop norovirus vaccines are ongoing. Vaccines based on VLPs are being tested in human volunteers. In a proof-of-concept trial, the efficacy of a monovalent GI.1 VLP vaccine was 47% among volunteers who received the vaccine intranasally and were then challenged with a homologous strain. In a second trial, norovirus disease severity was reduced in volunteers who received a bivalent G1.1/GII.4 VLP vaccine intramuscularly (with the GII.4 component including a consensus sequence from three different GII.4 strains) and were subsequently challenged with a GII.4 norovirus strain. Data from the first field efficacy study of this bivalent vaccine conducted in ~4700 healthy U.S. Navy recruits given 1 intramuscular injection of the bivalent vaccine were recently reported. While the primary endpoint of protection against homotypic infection could not be evaluated because only 6 total moderate/severe cases due to GI.1 or GII.4 norovirus strains occurred during the trial, the vaccine efficacy was 61.8% (95.01% confidence interval, 20.8–81.6%) for moderate/severe norovirus acute gastroenteritis due to any type. These initial data are encouraging; however, key issues to be further studied include the duration of protection and the level of heterotypic protection against antigenically distinct strains, particularly given the continuing and rapid natural evolution leading to the emergence of novel norovirus strains.

ROTAVIRUS

Etiologic Agent Rotaviruses are members of the family Reoviridae. The viral genome consists of 11 segments of double-strand RNA enclosed in a triple-layered, nonenveloped, icosahedral capsid 75 nm in diameter. Viral protein 6 (VP6), the major structural protein, is the target of commercial immunoassays and determines the group specificity of rotaviruses. Seven major groups of rotavirus (A through G) exist; human illness is caused primarily by group A and, to a much lesser extent, by groups B and C. Two outer-capsid proteins, VP7 (G-protein) and VP4 (P-protein), determine serotype specificity, induce neutralizing antibodies, and form the basis for binary classification of rotaviruses (G and P types). The segmented genome of rotavirus allows genetic reassortment (i.e., exchange of genome segments between viruses) during co-infection—a property that plays a role in viral evolution and that has been utilized in the development of reassortant animal/human rotavirus–based vaccines.

Epidemiology Worldwide, nearly all children are infected with rotavirus by 3–5 years of age. Neonatal infections are common but are often asymptomatic or mild, presumably because of protection by maternal antibody or breast milk. Compared with rotavirus disease in industrialized countries, disease in developing countries occurs at a younger age, is less seasonal, is more frequently caused by uncommon or multiple rotavirus strains, and is more often fatal. Moreover, because of suboptimal access to hydration therapy, rotavirus is a leading cause of diarrheal death among children in the developing world, with the highest mortality rates among children in sub-Saharan Africa and southern Asia **(Fig. 209-2)**.

First infections after 3 months of age are likely to be symptomatic, and the incidence of disease peaks among children 4–23 months of age. Reinfections are common, but the severity of disease decreases with each repeat infection. Therefore, severe rotavirus infections are less common among older children and adults than among younger individuals. Nevertheless, rotavirus can cause illness in parents and caretakers of children with rotavirus diarrhea, immunocompromised persons, travelers, and elderly individuals and should be considered in the differential diagnosis of gastroenteritis among adults.

In tropical settings, rotavirus disease occurs year-round, with less pronounced seasonal peaks than in temperate settings, where rotavirus disease occurs predominantly during the cooler fall and winter months. Before the introduction of rotavirus vaccine in the United States, the rotavirus season each year began in the Southwest during the autumn and early winter (October through December) and migrated across the continent, peaking in the Northeast during late winter and spring (March through May). The reasons for this characteristic pattern are not clear but may be correlated with state-specific differences in birth rates, which could influence the rate of accumulation of susceptible infants after each rotavirus season. After the implementation of routine vaccination of U.S. infants against rotavirus in 2006, the characteristic prevaccine geotemporal pattern of U.S. rotavirus was dramatically altered, and these changes were accompanied by substantial declines in rotavirus detections by a national network of sentinel laboratories.

During episodes of rotavirus-associated diarrhea, virus is shed in large quantities in stool (10^7–10^{12}/g). Viral shedding detectable by EIA usually subsides within 1 week but may persist for >30 days in immunocompromised individuals; it may be detected for longer periods by sensitive molecular assays, such as PCR. The virus is transmitted predominantly through the fecal–oral route. Spread through respiratory secretions, person-to-person contact, or contaminated environmental surfaces has been postulated to explain the rapid acquisition of antibody in the first 3 years of life, regardless of sanitary conditions.

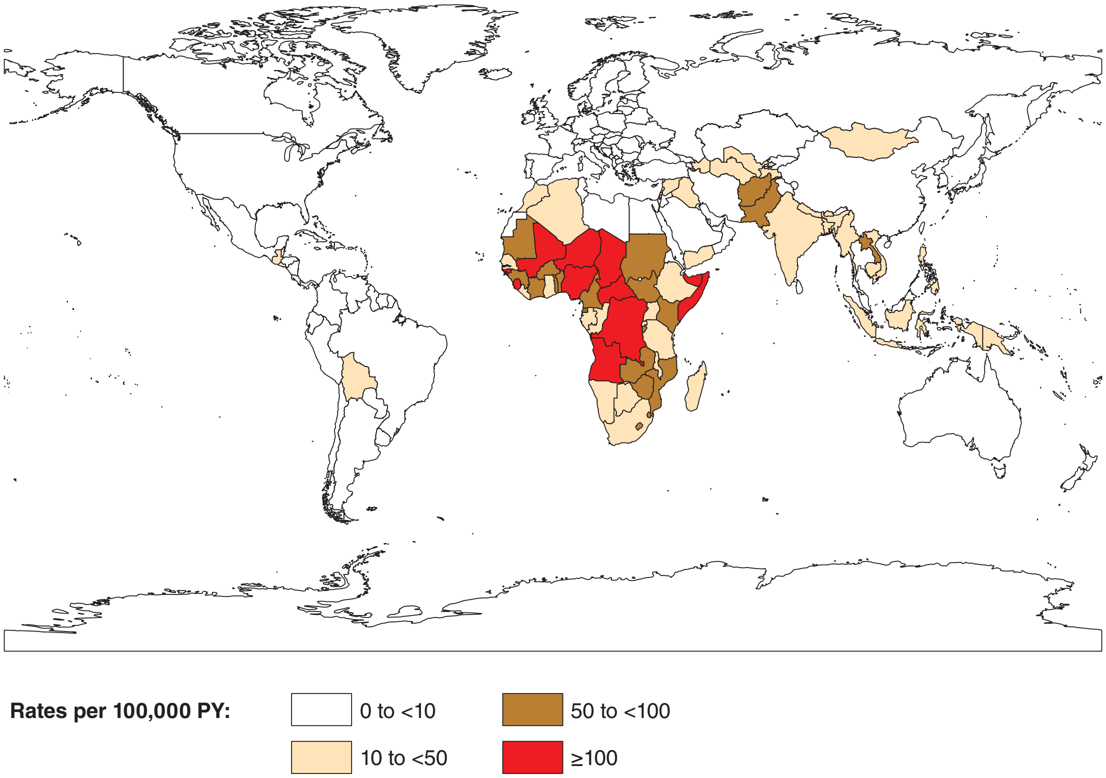

FIGURE 209-2 Rotavirus mortality rates by country, per 100,000 children <5 years of age. *(From JE Tate et al: Global, regional, and national estimates of rotavirus mortality in children <5 years of age, 2000–2013. Clin Infect Dis 62:S96, 2016.)*

At least 10 different G serotypes of group A rotavirus have been identified in humans, but only 5 types (G1 through G4 and G9) are common. While human rotavirus strains that possess a high degree of genetic homology with animal strains have been identified, animal-to-human transmission appears to be uncommon.

Group B rotaviruses have been associated with several large epidemics of severe gastroenteritis among adults in China since 1982 and have also been identified in India. Group C rotaviruses have been associated with a small proportion of pediatric gastroenteritis cases in several countries worldwide.

Pathogenesis Rotaviruses infect and ultimately destroy mature enterocytes in the villous epithelium of the proximal small intestine. The loss of absorptive villous epithelium, coupled with the proliferation of secretory crypt cells, results in secretory diarrhea. Brush-border enzymes characteristic of differentiated cells are reduced, and this change leads to the accumulation of unmetabolized disaccharides and consequent osmotic diarrhea. Studies in mice indicate that a nonstructural rotavirus protein, NSP4, functions as an enterotoxin and contributes to secretory diarrhea by altering epithelial cell function and permeability. In addition, rotavirus may evoke fluid secretion through activation of the enteric nervous system in the intestinal wall. Rotavirus antigenemia and viremia are common among children with acute rotavirus infection, although the antigen and RNA levels in serum are substantially lower than those in stool.

Clinical Manifestations The clinical spectrum of rotavirus infection ranges from subclinical infection to severe gastroenteritis leading to life-threatening dehydration. After an incubation period of 1–3 days, the illness has an abrupt onset, with vomiting frequently preceding the onset of diarrhea. Up to one-third of patients may have a temperature of >39°C. The stools are characteristically loose and watery and only infrequently contain red or white cells. Gastrointestinal symptoms generally resolve in 3–7 days.

Respiratory and neurologic features in children with rotavirus infection have been reported, but causal associations have not been proven. Moreover, rotavirus infection has been associated with a variety of other clinical conditions (e.g., sudden infant death syndrome, necrotizing enterocolitis, intussusception, Kawasaki disease, and type 1 diabetes), but no causal relationship has been confirmed with any of these syndromes.

Rotavirus does not appear to be a major opportunistic pathogen in children with HIV infection. In severely immunodeficient children, rotavirus can cause protracted diarrhea with prolonged viral excretion and, in rare instances, can disseminate systemically. Persons who are immunosuppressed for bone marrow transplantation also are at risk for severe or even fatal rotavirus disease.

Immunity Protection against rotavirus disease is correlated with the presence of virus-specific secretory IgA antibodies in the intestine and, to some extent, the serum. Because virus-specific IgA production at the intestinal surface is short-lived, complete protection against disease is only temporary. However, each infection and subsequent reinfection confers progressively greater immunity; thus, severe disease is most common among young children with first or second infections. Immunologic memory is believed to be important in the attenuation of disease severity upon reinfection.

Diagnosis Illness caused by rotavirus is difficult to distinguish clinically from that caused by other enteric viruses. Because large quantities of virus are shed in feces, the diagnosis can usually be confirmed by a wide variety of commercially available EIAs or by techniques for detecting viral RNA, such as gel electrophoresis, probe hybridization, or PCR.

TREATMENT

Rotavirus Infections

Rotavirus gastroenteritis can lead to severe dehydration; appropriate treatment should be instituted early. Standard oral rehydration therapy is successful for most children who can take fluids by mouth, but IV fluid replacement may be required for patients who are severely dehydrated or are unable to tolerate oral therapy because of frequent vomiting. The therapeutic roles of probiotics, bismuth subsalicylate, enkephalinase inhibitors, and nitazoxanide have been evaluated in clinical studies but are not clearly defined. Antibiotics and antimotility agents should be avoided. In immunocompromised children with chronic symptomatic rotavirus disease, orally administered immunoglobulins or colostrum may result in the resolution of symptoms, but the best choices regarding agents and their doses have not been well studied, and treatment decisions are often empirical.

Prevention Efforts to develop rotavirus vaccines were pursued because it was apparent—given the similar rates in less developed and industrialized nations—that improvements in hygiene and sanitation were unlikely to reduce disease incidence. The first rotavirus vaccine licensed in the United States in 1998 was withdrawn from the market within 1 year because it was linked with a low incidence of intussusception, a form of bowel obstruction.

In 2006, promising safety and efficacy (85–98% against severe rotavirus disease) data for two new rotavirus vaccines—RotaTeq (Merck, United States) and Rotarix (GlaxoSmithKline, Belgium)—were reported from large clinical trials conducted in North America, Europe, and Latin America. Both vaccines are now recommended for routine immunization of all U.S. infants, and their use has led to a >70–80% decline in rotavirus hospitalizations and emergency department visits at hospitals across the United States. Somewhat unexpectedly, rotavirus vaccination of young infants has also resulted in the added benefit of declines in rotavirus disease among children who miss vaccination and even among older children and adults who are not eligible for vaccination in some settings. The reason is likely to be a reduction in community transmission of rotavirus because of vaccination—i.e., herd protection. In April 2009, the World Health Organization (WHO) recommended the use of rotavirus vaccines in all countries worldwide. As of December 2003, over 120 countries, including several low-income countries in Africa and Asia, have incorporated rotavirus vaccine into their national childhood immunization programs (Fig. 209-3). Large declines in severe morbidity and mortality from childhood diarrhea have been documented in many countries. Postmarketing surveillance has identified a low risk of intussusception in some high- and middle-income countries; however, the benefits of vaccination exceed the risks, and no changes in vaccine administration policy have been implemented. An intussusception risk has not been identified in several postmarketing evaluations in developing countries to date.

The different epidemiology of rotavirus disease and the greater prevalence of co-infection with other enteric pathogens, of comorbidities, and of malnutrition in developing countries may adversely affect the performance of oral rotavirus vaccines, as is the case with oral vaccines against poliomyelitis, cholera, and typhoid in these regions. Therefore, evaluation of the efficacy of rotavirus vaccines in resource-poor settings of Africa and Asia was specifically recommended, and these trials have now been completed. As anticipated, the efficacy of rotavirus vaccines was moderate (50–65%) in these settings when compared with that in industrialized countries. Despite modest efficacy, routine use of rotavirus vaccines in low-income African countries with a heavy disease burden has yielded substantial public health benefits.

Several manufacturers in emerging markets, including India, China, Vietnam, Indonesia, and Brazil, are developing candidate rotavirus vaccines. Beginning in 2016, two Indian-made rotavirus vaccines—Rotavac (Bharat Biotech, India) and Rotasiil (Serum Institute, India)—were implemented in India's routine childhood immunization program, which has since expanded to all Indian states with a birth cohort of >25 million. In trials conducted in low-income countries, the efficacy of Rotavac and Rotasiil ranged from 36 to 66%, similar to the efficacy of multinational vaccines in these settings. In 2018, these two vaccines were prequalified by WHO, allowing their procurement with funding support from Gavi, the Vaccine Alliance, in low-income countries outside India.

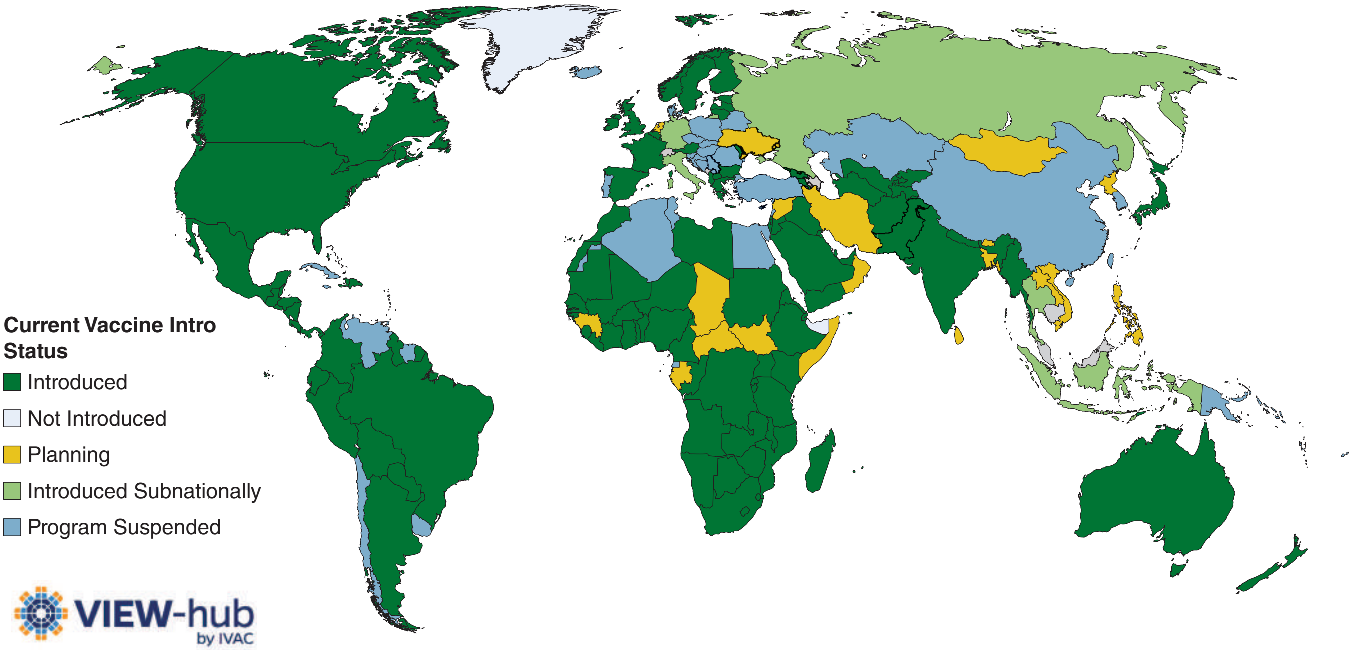

January 8, 2024 © The International Vaccine Access Center (IVAC)

FIGURE 209-3 Countries that have implemented national rotavirus vaccination programs, January 8, 2024. *(Source: View-Hub, http://www.view-hub.org/viz/. © The International Vaccine Access Center [IVAC].)*

OTHER VIRAL AGENTS OF GASTROENTERITIS

Enteric *adenoviruses* of serotypes 40 and 41 belonging to subgroup F are 70- to 80-nm viruses with double-strand DNA that cause ~2–12% of all diarrhea episodes in young children. Unlike adenoviruses that cause respiratory illness, enteric adenoviruses are difficult to cultivate in cell lines, but they can be detected with commercially available EIAs. Adenovirus types 31 and 42–49 have been linked to diarrhea in HIV-infected and other immunocompromised persons.

Astroviruses are 28- to 30-nm viruses with a characteristic icosahedral structure and a positive-sense, single-strand RNA. At least seven serotypes have been identified, of which serotype 1 is most common. Astroviruses are primarily pediatric pathogens, causing ~2–10% of cases of mild to moderate gastroenteritis in children. The availability of simple immunoassays to detect virus in fecal specimens and of molecular methods to confirm and characterize strains will permit more comprehensive assessment of the etiologic role of these agents.

Toroviruses are 100- to 140-nm, enveloped, positive-strand RNA viruses that are recognized as causes of gastroenteritis in horses (Berne virus) and cattle (Breda virus). Their role as a cause of diarrhea in humans is still unclear, but studies from Canada have demonstrated associations between torovirus excretion and both nosocomial gastroenteritis and necrotizing enterocolitis in neonates. These associations require further evaluation.

Picobirnaviruses are small, bisegmented, double-strand RNA viruses that cause gastroenteritis in a variety of animals. Their role as primary causes of gastroenteritis in humans remains unclear, but several studies have found an association between picobirnaviruses and gastroenteritis in HIV-infected adults.

Several other viruses (e.g., enteroviruses, reoviruses, pestiviruses, Aichivirus, and parvovirus B) have been identified in the feces of patients with diarrhea, but their etiologic role in gastroenteritis has not been proven. Diarrhea has also been noted as a manifestation of infection with recently recognized viruses that primarily cause severe respiratory illness: the SARS-CoV viruses, the influenza A/H5N1 virus, and the current pandemic strain of influenza A/H1N1 virus.

FURTHER READING

Armah G et al: Vaccine value profile for norovirus. Vaccine 41 Suppl 2:S134, 2023.

Clark A et al: Estimating the global impact of rotavirus vaccines on child mortality. Int J Infect Dis 137:90, 2023.

Cohen A et al: Aetiology and incidence of diarrhoea requiring hospitalisation in children under 5 years of age in 28 low-income and middle-income countries: findings from the Global Pediatric Diarrhea Surveillance. BMJ Global Health 7:e009548, 2022.

Flynn T et al: Viral gastroenteritis. Lancet 403:862, 2024.

Tate J et al: Global, regional, and national estimates of rotavirus mortality in children <5 years of age, 2000–2013. Clin Infect Dis 62:S96, 2016.

210

Enterovirus, Parechovirus, and Reovirus Infections

Jeffrey I. Cohen

ENTEROVIRUSES

CLASSIFICATION AND CHARACTERIZATION

Enteroviruses, members of the family Picornaviridae, are so designated because of their ability to multiply in the gastrointestinal tract. Despite their name, these viruses are not a prominent cause of gastroenteritis. Enteroviruses encompass more than 115 human serotypes: 3 serotypes of poliovirus, 23 serotypes of coxsackievirus A, 6 serotypes of coxsackievirus B, 29 serotypes of echovirus, enteroviruses 68–71, and multiple new enteroviruses (beginning with enterovirus 73) that have been identified by molecular techniques. Human enteroviruses have been reclassified into four species designated A–D. Echoviruses 22 and 23 have been reclassified as parechoviruses 1 and 2 on the basis of low nucleotide homology and differences in viral proteins. Enterovirus and parechovirus surveillance conducted in the United States by the Centers for Disease Control and Prevention (CDC) in 2022 showed that the most common enteroviruses and parechoviruses were enterovirus D68 (41.1% of cases) and human parechovirus 2 (20.7%), followed by coxsackievirus A6, A9, B3, and B5, each of which accounted for 4–5% of all isolates.

Human enteroviruses contain a single-stranded RNA genome surrounded by an icosahedral capsid comprising four viral proteins. These viruses have no lipid envelope and are stable in acidic environments, including the stomach. They are susceptible to chlorine-containing cleansers but resistant to inactivation by standard disinfectants (e.g., alcohol, detergents) and can persist for days at room temperature.

PATHOGENESIS AND IMMUNITY

Much of what is known about the pathogenesis of enteroviruses has been derived from studies of poliovirus infection. After ingestion, poliovirus is thought to infect epithelial cells in the mucosa of the gastrointestinal tract and then to spread to and replicate in the submucosal lymphoid tissue of the tonsils and Peyer's patches. The virus next spreads to the regional lymph nodes, a viremic phase ensues, and the virus replicates in organs of the reticuloendothelial system. In some cases, a second episode of viremia occurs and the virus replicates further in various tissues, sometimes causing symptomatic disease.

It is uncertain whether poliovirus reaches the central nervous system (CNS) during viremia or whether it also spreads via peripheral nerves. Since viremia precedes the onset of neurologic disease in humans, it has been assumed that the virus enters the CNS via the bloodstream. The poliovirus receptor is a member of the immunoglobulin superfamily. Poliovirus infection is limited to primates, largely because their cells express the viral receptor. Studies demonstrating the poliovirus receptor in the end-plate region of muscle at the neuromuscular junction suggest that, if the virus enters the muscle during viremia, it could travel across the neuromuscular junction up the axon to the anterior horn cells. Studies of monkeys and of transgenic mice expressing the poliovirus receptor show that, after IM injection, poliovirus does not reach the spinal cord if the sciatic nerve is cut. Taken together, these findings suggest that poliovirus can spread directly from muscle to the CNS by neural pathways.

Poliovirus can usually be cultured from the blood 3–5 days after infection, before the development of neutralizing antibodies. While viral replication at secondary sites begins to slow 1 week after infection, it continues in the gastrointestinal tract. Poliovirus is shed from the oropharynx for up to 3 weeks after infection and from the gastrointestinal tract for as long as 12 weeks; hypogammaglobulinemic patients can shed poliovirus for >20 years. During replication in the gastrointestinal tract, attenuated oral poliovirus can mutate, reverting to a more neurovirulent phenotype within a few days; however, additional mutations are probably required for full neurovirulence. One patient with hypogammaglobulinemia who had been infected 12 years earlier and was receiving IV immune globulin suddenly developed quadriplegia and respiratory muscle paralysis and died; analysis showed that the virus had reverted to a more wild-type sequence.

Humoral and secretory immunity in the gastrointestinal tract is important for the control of enterovirus infections. Enteroviruses induce specific IgM, which usually persists for <6 months, and specific IgG, which persists for life. Capsid protein VP1 is the predominant target of neutralizing antibody, which generally confers lifelong protection against subsequent disease caused by the same serotype but does not prevent infection or virus shedding. Enteroviruses also induce cellular immunity of uncertain significance. Patients with impaired cellular immunity are not known to develop unusually severe disease when infected with enteroviruses. In contrast, the severe infections in patients with agammaglobulinemia emphasize the importance of humoral immunity in controlling enterovirus infections. Disseminated enterovirus infections have occurred in hematopoietic cell transplant recipients. IgA antibodies are instrumental in reducing poliovirus replication in and shedding from the gastrointestinal tract. Breast milk contains IgA specific for enteroviruses and can protect humans from infection.

EPIDEMIOLOGY

Enteroviruses have a worldwide distribution. More than 50% of nonpoliovirus enterovirus infections and >90% of poliovirus infections are subclinical. When symptoms do develop, they are usually nonspecific and occur in conjunction with fever; only a minority of infections are associated with specific clinical syndromes. The incubation period for most enterovirus infections ranges from 2 to 14 days but usually is <1 week.

Enterovirus infection is more common in socioeconomically disadvantaged areas, especially in those where conditions are crowded and in tropical areas where hygiene is poor. Infection is most common among infants and young children; serious illness develops most often during the first few days of life and in older children and adults. In developing countries, where children are infected at an early age, poliovirus infection has less often been associated with paralysis; in countries with better hygiene, older children and adults are more likely to be seronegative, become infected, and develop paralysis. Passively acquired maternal antibody reduces the risk of symptomatic infection in neonates. Young children are the most frequent shedders of enteroviruses and are usually the index cases in family outbreaks. In temperate climates, enterovirus infections occur most often in the summer and fall; no seasonal pattern is apparent in the tropics.

Most enteroviruses are transmitted primarily by the fecal–oral or oral–oral route. Patients are most infectious shortly before and after the onset of symptomatic disease, when virus is present in the stool and throat. The ingestion of virus-contaminated food or water also can cause disease. Certain enteroviruses (such as enterovirus 70, which causes acute hemorrhagic conjunctivitis) can be transmitted by direct inoculation from the fingers to the eye. Airborne transmission is important for some viruses that cause respiratory tract disease, such as coxsackievirus A21. Enteroviruses can be transmitted across the placenta from mother to fetus, causing severe disease in the newborn. The transmission of enteroviruses through blood transfusions or insect bites has not been documented. Nosocomial spread of coxsackievirus and echovirus has taken place in hospital nurseries. Outbreaks of enteroviruses correlate with levels of preexisting immunity to specific serotypes and birth rates.

CLINICAL FEATURES

Poliovirus Infection Most infections with poliovirus are asymptomatic. After an incubation period of 3–6 days, ~5% of patients present with a minor illness (abortive poliomyelitis) manifested by fever, malaise, sore throat, anorexia, myalgias, and headache. This condition usually resolves in 3 days. About 1% of patients present with aseptic meningitis (nonparalytic poliomyelitis). Examination of cerebrospinal fluid (CSF) reveals lymphocytic pleocytosis, a normal glucose level, and a normal or slightly elevated protein level; CSF polymorphonuclear leukocytes may be present early. In some patients, especially children, malaise and fever precede the onset of aseptic meningitis.

PARALYTIC POLIOMYELITIS The least common presentation is that of paralytic disease. After one or several days, signs of aseptic meningitis are followed by severe back, neck, and muscle pain and by the rapid or gradual development of motor weakness. In some cases, the disease appears to be biphasic, with aseptic meningitis followed first by apparent recovery but then (1–2 days later) by the return of fever and the development of paralysis; this form is more common among children than among adults. Weakness is generally asymmetric, is proximal more than distal, and may involve the legs (most commonly); the arms; or the abdominal, thoracic, or bulbar muscles. Paralysis develops during the febrile phase of the illness and usually does not progress after defervescence. Urinary retention also may occur. Examination reveals weakness, fasciculations, decreased muscle tone, and reduced or absent reflexes in affected areas. Transient hyperreflexia sometimes precedes the loss of reflexes. Patients frequently report sensory symptoms, but objective sensory testing usually yields normal results. Bulbar paralysis may lead to dysphagia, difficulty in handling secretions, or dysphonia. Respiratory insufficiency due to aspiration, involvement of the respiratory center in the medulla, or paralysis of the phrenic or intercostal nerves may develop, and severe medullary involvement may lead to circulatory collapse. Most patients with paralysis recover some function weeks to months after infection. About two-thirds of patients have residual neurologic sequelae.

Paralytic disease is more common among older individuals, pregnant women, and persons exercising strenuously or undergoing trauma at the time of CNS symptoms. Tonsillectomy predisposes to bulbar poliomyelitis, and IM injections increase the risk of paralysis in the involved limb(s).

VACCINE-ASSOCIATED POLIOMYELITIS The risk of developing poliomyelitis after oral vaccination is estimated at 1 case per 2.5 million doses. The risk is ~2000 times higher among immunodeficient persons, especially persons with hypo- or agammaglobulinemia. Before 1997, an average of eight cases of vaccine-associated poliomyelitis occurred—in both vaccinees and their contacts—in the United States each year. With the change in recommendations first to a sequential regimen of inactivated poliovirus vaccine (IPV) and oral poliovirus vaccine (OPV) in 1997 and then to an all-IPV regimen in 2000, the number of cases of vaccine-associated polio declined. From 1997 to 1999, six such cases were reported in the United States followed by one case in 2022.

POSTPOLIO SYNDROME The *postpolio syndrome* presents as new onset of weakness, fatigue, fasciculations, and pain with additional atrophy of the muscle group involved during the initial paralytic disease 20–40 years earlier. The syndrome is more common among women and with increasing time after acute disease. The onset is usually insidious, and weakness occasionally extends to muscles that were not involved during the initial illness. The prognosis is generally good; progression to further weakness is usually slow, with plateau periods of 1–10 years. The postpolio syndrome is thought to be due to progressive dysfunction and loss of motor neurons that compensated for the neurons lost during the original infection and not to persistent or reactivated poliovirus infection.

Other Enteroviruses An estimated 5–10 million cases of symptomatic disease due to enteroviruses other than poliovirus occur in the United States each year. Among neonates, enteroviruses are the most common cause of aseptic meningitis and nonspecific febrile illnesses. Certain clinical syndromes are more likely to be caused by certain serotypes (Table 210-1).

NONSPECIFIC FEBRILE ILLNESS (SUMMER GRIPPE) The most common clinical manifestation of enterovirus infection is a nonspecific febrile illness. After an incubation period of 3–6 days, patients present with an acute onset of fever, malaise, and headache. Occasional cases are associated with upper respiratory symptoms, and some cases include nausea and vomiting. Symptoms often last for 3–4 days, and most cases resolve in a week. While infections with other respiratory viruses occur more often from late fall to early spring, febrile illness due to enteroviruses frequently occurs in the summer and early fall.

TABLE 210-1 Manifestations Commonly Associated with Enterovirus Serotypes

	SEROTYPE(S) OF INDICATED VIRUS	
MANIFESTATION	COXSACKIEVIRUS	ECHOVIRUS (E) AND ENTEROVIRUS (Ent)
Acute hemorrhagic conjunctivitis	A24	E70
Aseptic meningitis	A2, 4, 7, 9, 10; B1–5	E4, 6, 7, 9, 11, 13, 16, 18, 19, 30, 33; Ent70, 71
Encephalitis	A9; B1–5	E3, 4, 6, 7, 9, 11, 18, 25, 30; Ent71
Exanthem	A4, 5, 6, 9, 10, 16; B1, 3–5	E4–7, 9, 11, 16–19, 25, 30; Ent71
Generalized disease of the newborn	B1–5	E4–7, 9, 11, 14, 16, 18, 19
Hand-foot-and-mouth disease	A5–7, 9, 10, 16; B1, 2, 5	Ent71
Herpangina	A1–10, 16, 22; B1–5	E6, 9, 11, 16, 17, 25, 30; Ent71
Myocarditis, pericarditis	A4, 9, 16; B1–5	E6, 9, 11, 22
Paralysis	A4, 7, 9; B1–5	E2–4, 6, 7, 9, 11, 18, 30; EntD68, 70, 71
Pleurodynia	A1, 2, 4, 6, 9, 10, 16; B1–6	E1–3, 6, 7, 9, 11, 12, 14, 16, 19, 24, 25, 30
Pneumonia	A9, 16; B1–5	E6, 7, 9, 11, 12, 19, 20, 30; EntD68, 71

GENERALIZED DISEASE OF THE NEWBORN Most serious enterovirus infections in infants develop during the first week of life, although severe disease can occur up to 3 months of age. Neonates often present with an illness resembling bacterial sepsis, with fever, irritability, and lethargy. Laboratory abnormalities include leukocytosis with a left shift, thrombocytopenia, elevated values in liver function tests, and CSF pleocytosis. The illness can be complicated by myocarditis and hypotension, fulminant hepatitis and disseminated intravascular coagulation, meningitis or meningoencephalitis, or pneumonia. It may be difficult to distinguish neonatal enterovirus infection from bacterial sepsis, although a history of a recent virus-like illness in the mother provides a clue.

ASEPTIC MENINGITIS AND ENCEPHALITIS In children and young adults, enteroviruses are the cause of up to 90% of cases of aseptic meningitis in which an etiologic agent can be identified. Patients with aseptic meningitis typically present with an acute onset of fever, chills, headache, photophobia, and pain on eye movement. Nausea and vomiting also are common. Examination reveals meningismus without localizing neurologic signs; drowsiness or irritability also may be apparent. In some cases, a febrile illness may remit and return several days later in conjunction with signs of meningitis. Other systemic manifestations may provide clues to an enteroviral cause, including diarrhea, myalgias, rash, pleurodynia, myocarditis, and herpangina. Examination of the CSF invariably reveals pleocytosis; the CSF cell count shows a shift from neutrophil to lymphocyte predominance within 1 day of presentation, and the total cell count does not exceed 1000/μL. The CSF glucose level is usually normal (in contrast to the low CSF glucose level in mumps), with a normal or slightly elevated protein concentration. Partially treated bacterial meningitis may be particularly difficult to exclude in some instances. Enteroviral meningitis is more common in summer and fall in temperate climates, while viral meningitis of other etiologies is more common in winter and spring. Symptoms ordinarily resolve within a week, although CSF abnormalities can persist for several weeks. Enteroviral meningitis is often more severe in adults than in children. Neurologic sequelae are rare, and most patients have an excellent prognosis.

Enteroviral encephalitis is much less common than enteroviral aseptic meningitis. Occasional highly inflammatory cases of enteroviral meningitis may be complicated by a mild form of encephalitis that is recognized on the basis of progressive lethargy, disorientation, and sometimes seizures. Less commonly, severe primary encephalitis may develop. An estimated 10–35% of cases of viral encephalitis are due to enteroviruses. Immunocompetent patients generally have a good prognosis.

Patients with hypogammaglobulinemia, agammaglobulinemia, or severe combined immunodeficiency may develop chronic meningitis or encephalitis; about half of these patients have a dermatomyositis-like syndrome, with peripheral edema, rash, and myositis. They may also have chronic hepatitis. Patients may develop neurologic disease while receiving immunoglobulin replacement therapy. Echoviruses (especially echovirus 11) are the most common pathogens in this situation.

Paralytic disease due to enteroviruses other than poliovirus occurs sporadically and is usually less severe than poliomyelitis. Most cases are due to enterovirus 70 or 71 or to coxsackievirus A7 or A9. Guillain-Barré syndrome is also associated with enterovirus infection. While earlier studies suggested a link between enteroviruses and chronic fatigue syndrome, most recent studies have not demonstrated such an association.

ACUTE FLACCID MYELITIS Patients with acute flaccid myelitis present with fever or respiratory symptoms and progress within hours to a few days to flaccid paralysis in one or more limbs. The disease is much more frequent in children. Less commonly, the disease can affect cranial nerves and respiratory or bulbar muscles. Like polio and some other enteroviruses, the disease affects the anterior horn cells in the spinal cord; gray matter changes can be seen on MRI of the spinal cord. The CSF shows a lymphocytic pleocytosis and often a mildly elevated protein. Cases of acute flaccid myelitis have occurred in late summer or early fall since 2012. Several studies have shown antibodies to enteroviruses in the CSF; antibodies to enterovirus D68 are most frequently detected. While enterovirus D68 has been detected in respiratory, stool, and nasopharyngeal samples from patients with acute flaccid myelitis, the virus has been rarely detected in the CSF. Treatment is supportive, and most patients have persistent neurologic deficits.

PLEURODYNIA (BORNHOLM DISEASE) Patients with pleurodynia present with an acute onset of fever and spasms of pleuritic chest or upper abdominal pain. Chest pain is more common in adults, and abdominal pain is more common in children. Paroxysms of severe, knifelike pain usually last 15–30 min and are associated with diaphoresis and tachypnea. Fever peaks within an hour after the onset of paroxysms and subsides when pain resolves. The involved muscles are tender to palpation, and a pleural rub may be detected. The white blood cell count and chest x-ray results are usually normal. Most cases are due to coxsackievirus B and occur during epidemics. Symptoms resolve in a few days, and recurrences are rare. Treatment includes the administration of nonsteroidal anti-inflammatory agents or the application of heat to the affected muscles.

MYOCARDITIS AND PERICARDITIS Enteroviruses are estimated to cause up to one-third of cases of acute myocarditis. Coxsackievirus B and its RNA have been detected in pericardial fluid and myocardial tissue in some cases of acute myocarditis and pericarditis. Most cases of enteroviral myocarditis or pericarditis occur in newborns, adolescents, or young adults. More than two-thirds of patients are male. Patients often present with an upper respiratory tract infection that is followed by fever, chest pain, dyspnea, arrhythmias, and occasionally heart failure. A pericardial friction rub is documented in half of cases, and the electrocardiogram shows ST-segment elevations or ST- and T-wave abnormalities. Serum levels of myocardial enzymes are often elevated. Neonates commonly have severe disease, while older children and adults recover completely. Up to 10% of cases progress to chronic dilated cardiomyopathy. Chronic constrictive pericarditis also may be a sequela.

EXANTHEMS Enterovirus infection is the leading cause of exanthems in children in the summer and fall. While exanthems are associated with many enteroviruses, certain types have been linked to specific syndromes. Echoviruses 9 and 16 have frequently been associated with exanthem and fever. Rashes may be discrete or confluent, beginning on the face and spreading to the trunk and extremities. Echovirus 9 is the most common cause of a rubelliform (discrete) rash. Unlike the rash of rubella, the enteroviral rash occurs in the summer and is not associated with lymphadenopathy. Roseola-like rashes develop after defervescence, with macules and papules on the face and trunk. The Boston exanthem, caused by echovirus 16, is a roseola-like rash. A variety of other rashes have been associated with enteroviruses, including erythema multiforme **(see Fig. A1-24*A*)** and vesicular, urticarial, petechial, bullous, or purpuric lesions. Enanthems also occur, including lesions that resemble the Koplik's spots seen with measles **(see Fig. A1-2)**.

HAND-FOOT-AND-MOUTH DISEASE (FIG. 210-1) After an incubation period of 4–6 days, patients with hand-foot-and-mouth disease present with fever, anorexia, and malaise; these manifestations are followed by the development of sore throat and vesicles **(see Fig. A1-22)** on the buccal mucosa and often on the tongue and then by the appearance of tender vesicular lesions on the dorsum of the hands, sometimes with involvement of the palms. The vesicles may form bullae and quickly ulcerate. About one-third of patients also have lesions on the palate, uvula, or tonsillar pillars, and one-third have a rash on the feet (including the soles) or on the buttocks. Generalized rashes also have been reported. The disease is highly infectious, with attack rates of close to 100% among young children. The lesions usually resolve in 1 week. Most cases are due to coxsackievirus A16 or enterovirus 71.

An epidemic of enterovirus 71 infection in Taiwan in 1998 resulted in thousands of cases of hand-foot-and-mouth disease or herpangina (see below). Severe complications included CNS disease, myocarditis, and pulmonary hemorrhage. About 90% of those who died were children ≤5 years old, and death was associated with pulmonary edema or pulmonary hemorrhage. CNS disease included aseptic meningitis, flaccid paralysis (similar to that seen in poliomyelitis), and rhombencephalitis with myoclonus and tremor or ataxia. The mean age of patients with CNS complications was 2.5 years, and MRI in cases with encephalitis usually showed brainstem lesions. Follow-up of children at 6 months showed persistent dysphagia, cranial nerve palsies, hypoventilation, limb weakness, and atrophy; at 3 years, persistent neurologic sequelae were documented, with delayed development and impaired cognitive function.

Yearly epidemics of enterovirus 71 infection have occurred in China since 2008, with thousands of cases and hundreds of deaths each year. Infections have been associated with fever, rash, brainstem encephalitis with myoclonic jerks, and limb trembling; some cases have progressed to seizures and coma. Lung findings include pulmonary edema and hemorrhage. While the level of creatine kinase MB is sometimes elevated, myocardial necrosis generally is not found.

Cyclic epidemics occur every 2–3 years in other Asian countries. However, the virus circulates at lower rates in the United States, Europe, and Africa. In the United States, hand-foot-and-mouth disease is most commonly associated with coxsackievirus A16. Between November 2011 and February 2012, outbreaks of hand-foot-and-mouth disease

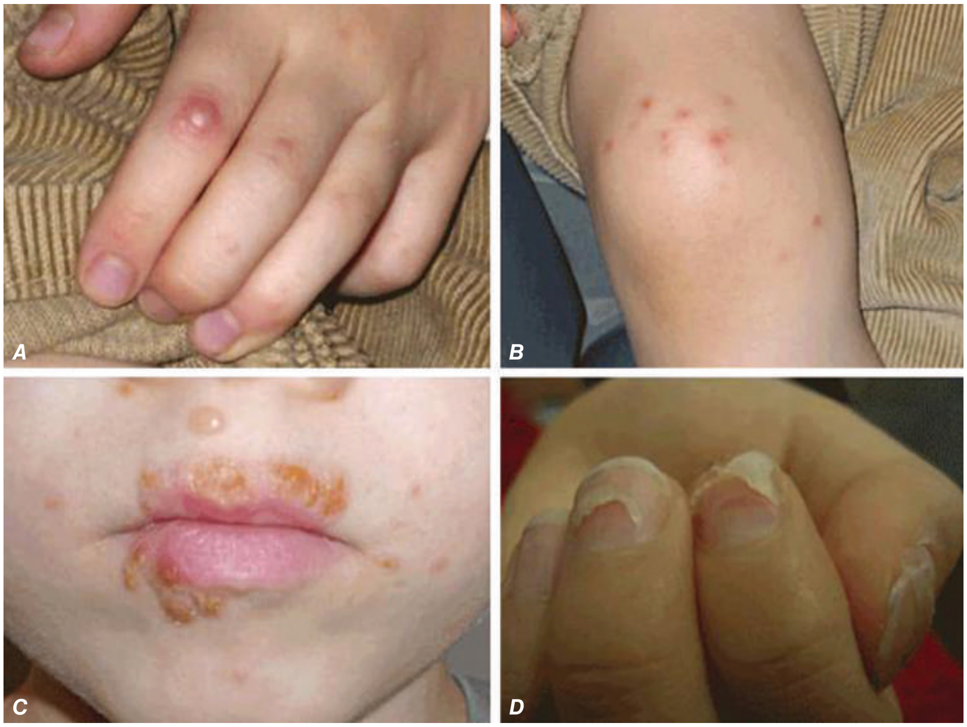

FIGURE 210-1 Vesicular eruptions of the hand (*A*), knee (*B*), and mouth (*C*) of a 6-year-old boy with coxsackievirus A6 infection. Several of his fingernails were shed 2 months later (*D*). *(Images reprinted courtesy of Centers for Disease Control and Prevention/Emerging Infectious Diseases.)*

due to coxsackievirus A6 occurred in several U.S. states, and 19% of the affected persons were hospitalized.

HERPANGINA Herpangina is usually caused by coxsackievirus A and presents as acute-onset fever, sore throat, odynophagia, and grayish-white papulovesicular lesions on an erythematous base that ulcerate. The lesions can persist for weeks; are present on the soft palate, anterior pillars of the tonsils, and uvula; and are concentrated in the posterior portion of the mouth. In contrast to herpes stomatitis, enteroviral herpangina is not associated with gingivitis. Acute lymphonodular pharyngitis associated with coxsackievirus A10 presents as white or yellow nodules surrounded by erythema in the posterior oropharynx. The lesions do not ulcerate.

ACUTE HEMORRHAGIC CONJUNCTIVITIS Patients with acute hemorrhagic conjunctivitis present with an acute onset of severe eye pain, blurred vision, photophobia, and watery discharge from the eye. Examination reveals edema, chemosis, and subconjunctival hemorrhage and often shows punctate keratitis and conjunctival follicles as well **(Fig. 210-2)**. Preauricular adenopathy is often found. Epidemics and nosocomial spread have been associated with enterovirus 70 and coxsackievirus A24. Outbreaks have been due to coxsackievirus A24 in China and India (2010), Japan (2011), and Thailand (2014). Systemic symptoms, including headache and fever, develop in 20% of cases, and recovery is usually complete in 10 days. The sudden onset and short duration of the illness help to distinguish acute hemorrhagic conjunctivitis from other ocular infections, such as those due to adenovirus and *Chlamydia trachomatis*. Paralysis has been associated with some cases of acute hemorrhagic conjunctivitis due to enterovirus 70 during epidemics.

OTHER MANIFESTATIONS Enteroviruses are an infrequent cause of childhood pneumonia and the common cold. From mid-August 2014 to January 2015, enterovirus D68 infection was confirmed in more than 1000 persons with mild to severe respiratory illnesses in 49 U.S. states. Nearly all reported cases were in children, many of whom had asthma. Severe respiratory illness due to enterovirus D68 continues to occur; case numbers were lower in 2020, likely due to mitigation efforts for COVID-19, but rose in 2022. A prospective study of >300 children showed that prolonged shedding of enteroviruses in the stool was associated with development of islet cell autoantibodies and type 1 diabetes. Coxsackievirus B has been isolated at autopsy from the pancreas of a few children presenting with type 1 diabetes mellitus; however, most attempts to isolate the virus have been unsuccessful. Other diseases that have been associated with enterovirus infection include parotitis, bronchitis, bronchiolitis, croup, infectious lymphocytosis, polymyositis, acute arthritis, and acute nephritis.

DIAGNOSIS

Isolation of enterovirus in cell culture had been the traditional diagnostic procedure; PCR is used now more often. Cultures of stool, nasopharyngeal, or throat samples from patients with enterovirus diseases do not prove that the virus is directly associated with disease because these sites are frequently colonized for weeks in patients with subclinical infections. Isolation of virus from the throat is more likely to be associated with disease than is isolation from the stool since virus is shed for shorter periods from the throat. Cultures of CSF, serum, fluid from body cavities, or tissues are positive less frequently, but a positive result is indicative of disease caused by enterovirus. Cultures are more likely to be positive earlier than later in the course of infection. Cultures may be negative because of the presence of neutralizing antibody, lack of susceptibility of the cells used, or inappropriate handling of the specimen. Coxsackievirus A may require inoculation into special cell-culture lines or into suckling mice.

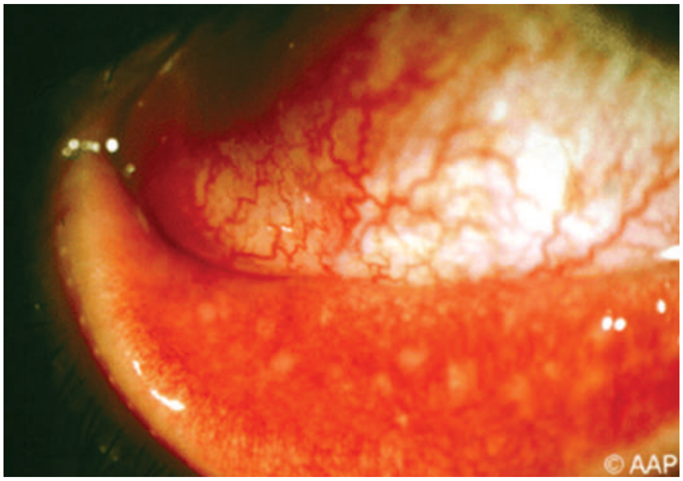

FIGURE 210-2 Acute hemorrhagic conjunctivitis due to enterovirus 70. *(Image reprinted courtesy of Jerri Ann Jenista, MD.)*

Identification of the enterovirus serotype is useful primarily for epidemiologic studies and, with a few exceptions, has little clinical utility. It is important to identify serious infections with enterovirus during epidemics and to distinguish the vaccine strain of poliovirus from the other enteroviruses in the throat or in the feces. Stool and throat samples for culture as well as acute- and convalescent-phase serum specimens should be obtained from all patients with suspected poliomyelitis. In the absence of a positive CSF culture, a positive culture of stool obtained within the first 2 weeks after the onset of symptoms is most often used to confirm the diagnosis of poliomyelitis. If poliovirus infection is suspected, two or more fecal and throat swab samples should be obtained at least 1 day apart and cultured for enterovirus as soon as possible. If poliovirus is isolated, it should be sent to the CDC for identification as either wild-type or vaccine virus.

Reverse-transcription polymerase chain reaction (PCR) has been used to amplify viral nucleic acid from CSF, serum, urine, stool, conjunctiva, throat swabs, and tissues. A pan-enterovirus PCR assay can detect all human enteroviruses. With the proper controls, PCR of the CSF is highly sensitive (70–100%) and specific (>80%) and is more rapid than culture. PCR of the CSF is less likely to be positive when patients present ≥3 days after the onset of meningitis or with enterovirus 71 infection; in these cases, PCR of throat or rectal swabs—although less specific than PCR of CSF—should be considered.

PCR of serum is also highly sensitive and specific in the diagnosis of disseminated disease. PCR may be particularly helpful for the diagnosis and follow-up of enterovirus disease in immunodeficient patients receiving immunoglobulin therapy, whose viral cultures may be negative. Antigen detection is less sensitive than PCR.

Serologic diagnosis of enterovirus infection is limited by the large number of serotypes and the lack of a common antigen. Demonstration of seroconversion may be useful in rare cases for confirmation of culture results, but serologic testing is usually limited to epidemiologic studies. Serum should be collected and frozen soon after the onset of disease and again ~4 weeks later. Measurement of neutralizing titers is the most accurate method for antibody determination; measurement of complement-fixation titers is usually less sensitive. Titers of virus-specific IgM are elevated in both acute and chronic infection.

TREATMENT

Enterovirus Infections

Most enterovirus infections are mild and resolve spontaneously; however, intensive supportive care may be needed for cardiac, hepatic, or CNS disease. IV, intrathecal, or intraventricular immunoglobulin has been used with apparent success in some cases for the treatment of chronic enterovirus meningoencephalitis and dermatomyositis in patients with hypogammaglobulinemia or agammaglobulinemia. The disease may stabilize or resolve during therapy; however, some patients decline inexorably despite therapy. IV immunoglobulin often prevents severe enterovirus disease in these patients. IV administration of immunoglobulin with high titers of antibody to the infecting virus has been used in some cases of life-threatening infection in neonates, who may not have maternally acquired antibody. In one trial involving neonates with enterovirus infections, immunoglobulin containing very high titers of antibody to the infecting virus reduced rates of viremia; however, the study was too small to show a substantial clinical benefit. The level of enteroviral antibodies varies with the immunoglobulin preparation. While a phase 2 trial of pleconaril for neonatal enterovirus sepsis showed that the time to serum PCR negativity was reduced and the survival rate increased in newborns who had confirmed enterovirus infections and were treated with the drug, the differences did not reach significance and the drug is not available on a compassionate-use basis. Pocapavir and vapendavir also are being tested for enterovirus infections; resistance developed rapidly to OPV in a clinical trial of pocapavir. Glucocorticoids are contraindicated.

Good hand-washing practices and the use of gowns and gloves are important in limiting nosocomial transmission of enteroviruses during epidemics. Enteric precautions are indicated for 7 days after the onset of enterovirus infections. Inactivated enterovirus 71 vaccines have been licensed in China.

PREVENTION AND ERADICATION OF POLIOVIRUS

(See also Chap. 129) After a peak of 57,879 cases of poliomyelitis in the United States in 1952, the introduction of IPV in 1955 and of OPV in 1961 ultimately eradicated disease due to wild-type poliovirus in the Western Hemisphere. Such disease has not been documented in the United States since 1979, when cases occurred among religious groups who had declined immunization. In the Western Hemisphere, paralysis due to wild-type poliovirus was last documented in 1991. Paralysis due to vaccine-derived poliovirus (VDPV) was reported in 2022 in New York (see below).

In 1988, when ~350,000 cases of polio occurred in 125 countries, the World Health Organization adopted a resolution to eradicate poliomyelitis by the year 2000. Wild-type poliovirus type 2 and wild-type poliovirus type 3 were declared eradicated in 2015 and 2019, respectively. The Americas were certified free of indigenous wild-type poliovirus transmission in 1994, the Western Pacific Region in 2000, the European Region in 2002, and Southeast Asia in 2014. After a nadir of 496 cases in 2001, 21 countries that had previously been free of polio reported cases imported from 6 polio-endemic countries in 2002–2005. By 2006, polio transmission had been reduced in most of these 21 countries. In 2017, there were 22 cases of wild-type polio, the lowest ever reported for 1 year—all of these cases were from Pakistan and Afghanistan. In 2021 wild-type polio remerged in Africa. After another peak of 176 cases in 2019, the number of cases of wild-type polio had fallen to 12 in 2023, all from 2 countries and all due to polio type 1 **(Table 210-2)**. Polio is a source of concern for unimmunized or partially immunized travelers. Clearly, global eradication of polio is necessary to eliminate the risk of importation of wild-type virus.

TABLE 210-2 Laboratory-Confirmed Cases of Poliomyelitis in 2023

COUNTRY	WILD-TYPE POLIO	VACCINE-DERIVED POLIO
Pakistan	6	0
Afghanistan	6	0
Democratic Republic of the Congo	0	223
Nigeria	0	87
Chad	0	55
Guinea	0	47
Madagascar	0	24
Mali	0	15
Central African Republic	0	14
Others	0	61[a]
Total	12	526

[a] Others with <10 cases; Kenya, Somalia, Yemen 8 cases; Cote d'Ivoire, Indonesia 6 cases; Mozambique 5 cases; Benin, Burkina Fossa, Niger, South Sudan 3 cases; United Republic of Tanzania 2 cases; Burundi, Ethiopia, Israel, Mauritania, Zambia, Zimbabwe 1 case each.

Outbreaks are thought to have been facilitated by suboptimal rates of vaccination, isolated pockets of unvaccinated children, poor sanitation and crowding, improper vaccine-storage conditions, and a reduced level of response to one of the serotypes in the vaccine. While the global eradication campaign has markedly reduced the number of cases of endemic polio, doubts have been raised as to whether eradication is a realistic goal, given the large number of asymptomatic infections and the political instability in developing countries.

Use of OPV, especially in areas with low vaccination rates, has been associated with vaccine-derived polio due to mutations that result in restoration of viral fitness and neurovirulence during prolonged replication in individuals or person-to-person transmission. Vaccine-derived polio was recognized in Egypt in 1983–1993, and hundreds of cases have been reported in many countries, including 385 cases in Nigeria in 2005–2012. Epidemics have been rapidly terminated after intensive vaccination with OPV. In 2005, a case of vaccine-derived polio occurred in an unvaccinated U.S. woman returning from a visit to Central and South America. In the same year, an unvaccinated immunocompromised infant in Minnesota was found to be shedding VDPV; further investigation identified 4 of 22 infants in the same community who were shedding the virus. All 5 infants were asymptomatic. These outbreaks emphasize the need for maintaining high levels of vaccine coverage and continued surveillance for circulating virus. In 2016, only 5 cases were reported; however, this number increased with a peak of 1113 cases in 2020 and subsequently declined to 526 cases of vaccine-derived polio in 2023 from 24 countries; 96% of these cases were from Africa and 3% were from the Eastern Mediterranean Region (Table 210-2). This decline in VDPV type 2 is associated with the use of a safer, novel type 2 OPV (see below). A case of vaccine-derived polio type 2 occurred in an unvaccinated adult in New York in 2022 and was genetically linked to wastewater collected in the area about 1 month before and after the case; the patient had not traveled internationally during the incubation period. VDPV due to OPV2 has been detected in wastewater from other countries in 2022, including Canada, Israel, and the United Kingdom, where virus transmission has been eliminated. From 2018 to March 2020, 92% of cases of vaccine-derived polio were due to type 2 virus. Cessation of vaccination with type 2 OPV is believed to be responsible for this increase in polio type 2. IPV is used in most industrialized countries and OPV in most developing countries, including those in which polio still is or recently was endemic. While IM injections of other vaccines (live or attenuated) can be given concurrently with OPV, unnecessary IM injections should be avoided during the first month after OPV vaccination because they increase the risk of vaccine-associated paralysis. Since 1988, an enhanced-potency inactivated poliovirus vaccine has been available in the United States.

After several doses of OPV alone, the seropositivity rate for individual poliovirus serotypes may still be suboptimal for children in developing countries; one or more supplemental doses of IPV can increase the rate of seropositivity for these serotypes. Against a given serotype, monovalent OPV containing only that serotype is more immunogenic than trivalent vaccine because of a lack of interference from other serotypes. Given the eradication of wild-type poliovirus type 2 and the establishment of OPV type 2 as the primary cause of vaccine-derived polio, bivalent OPV (types 1 and 3), which had been shown to be superior to trivalent OPV in inducing antibodies to types 1 and 3, replaced trivalent OPV vaccine in April 2016. However, outbreaks of vaccine-derived polio due to polio type 2 have required vaccination with monovalent OPV type 2. A novel type 2 oral poliovirus vaccine (nOPV2) was engineered to be impaired in reversion to neurovirulence. After ~700 million doses of nOPV2 were administered in response to circulating VDPV type 2, from 2021 to 2023, 61 cases of paralysis were associated with nOPV2, with all the cases in Africa. The rate of emergence of paralysis with nOPV2 is about 10-fold lower than with monovalent OPV2. Another approach to reduce VDPV is to vaccinate with IPV followed by OPV. Addition of at least one dose of trivalent IPV after immunization with bivalent OPV will also reduce the risk of vaccine-derived polio associated with type 2 virus and enhance immunity to poliovirus types 1 and 3. Accordingly, in 2016, ~90% of countries included trivalent IPV in their immunization schedules.

TABLE 210-3 Recommendations for Poliovirus Vaccination of Adults

1. Most adults in the United States have little risk for exposure to polioviruses, and most are immune as a result of vaccination during childhood. Vaccination with IPV is recommended for those at greater risk for exposure to polioviruses than the general population:
    a. travelers to areas or countries where polio is epidemic or endemic;
    b. laboratory workers who handle specimens that might contain polioviruses;
    c. health care workers or other caregivers who have close contact with patients who might be excreting wild-type polioviruses; and
    d. adults who are identified by public health authorities as being part of a group or population at increased risk because of an outbreak.
2. Adults who are unvaccinated or whose vaccination status is unknown and who are at increased risk should receive three doses of IPV. Two doses of IPV should be administered at intervals of 4–8 weeks; a third dose should be administered 6–12 months after the second.
3. Adults who have had a primary series of polio vaccine and who are at increased risk should receive another dose of IPV. Currently, data do not indicate a need for more than a single lifetime booster dose with IPV for adults.

Abbreviation: IPV, inactivated poliovirus vaccine.

Source: From *https://www.cdc.gov/vaccines/vpd/polio/hcp/recommendations.html.*

As the frequency of wild-type polio declines and reports of polio associated with circulating VDPV increase, the World Health Organization is investigating whether IPV can be produced from OPV strains that require less biocontainment, ultimately replacing OPV.

OPV and IPV induce antibodies that persist for at least 5 years. Both vaccines induce IgG and IgA antibodies. Compared with recipients of IPV, recipients of OPV shed less virus and less frequently develop reinfection with wild-type virus after exposure to poliovirus. Although IPV is safe and efficacious, OPV offers the advantages of ease of administration, lower cost, and induction of intestinal immunity resulting in a reduction in the risk of community transmission of wild-type virus. Because of progress toward global eradication of polio and the continued occurrence of cases of vaccine-associated polio, an all-IPV regimen was recommended in 2000 for childhood poliovirus vaccination in the United States, with vaccine administration at 2, 4, and 6–18 months and 4–6 years of age. The risk of vaccine-associated polio should be discussed before OPV is administered. Recommendations for vaccination of adults are listed in **Table 210-3**.

There are concerns about discontinuing vaccination in the event that endemic spread of poliovirus is eliminated. Among the reasons for these concerns are that poliovirus is shed from some immunocompromised persons for >25 years, that vaccine-derived poliovirus can circulate and cause disease, and that wild-type poliovirus is present in research laboratories and vaccine manufacturing facilities. Antivirals and monoclonal antibodies are in development to reduce or terminate shedding of poliovirus by long-term virus excretors. Pocapavir was shown to reduce shedding of OPV type 1 in a clinical trial, but rapid development of resistance with virus transmission, despite reduced shedding, indicates that combination therapy with other antivirals and/or monoclonal antibodies will be needed.

PARECHOVIRUSES

Human parechoviruses (HPeVs), like enteroviruses, are members of the family Picornaviridae. The 19 serotypes of HPeV commonly cause infections in early childhood. Infections with HPeV type 1 (HPeV-1) occur throughout the year, while other parechovirus infections occur more commonly in summer and fall. Infections with HPeVs present similarly to those due to enteroviruses and may cause generalized disease of the newborn, aseptic meningitis, encephalitis, seizures, paralysis, exanthems, respiratory tract disease, rash, hepatitis, and gastroenteritis. While HPeV-1 is the most common serotype and generally causes mild disease, deaths of infants in the United States have been associated with HPeV-1, HPeV-3, and HPeV-6. HPeVs can be isolated from the same sites as enteroviruses, including the nasopharynx, stool, and respiratory tract secretions. PCR using pan-enterovirus primers does not detect HPeVs, and while PCR assays are performed by the CDC and research laboratories, many commercial laboratories do not perform the test.

REOVIRUSES

Reoviruses are double-stranded RNA viruses encompassing three serotypes. Serologic studies indicate that most humans are infected with reoviruses during childhood. Most infections either are asymptomatic or cause mild upper respiratory tract symptoms. Reovirus is considered a rare cause of mild gastroenteritis or meningitis in infants and children. Speculation regarding an association of reovirus type 3 with idiopathic neonatal hepatitis and extrahepatic biliary atresia is based on an elevated prevalence of antibody to reovirus in some affected patients and the detection of viral RNA by PCR in hepatobiliary tissues in some studies. Orthoreoviruses have been associated with human disease—e.g., Melaka and Kampar viruses with fever and acute respiratory disease in Malaysia and Nelson Bay virus with acute respiratory disease in a traveler from Bali.

FURTHER READING

Lee SE et al: Progress toward poliomyelitis eradication – Worldwide, January 2021–March 2023. MMWR Morb Mortal Wkly Rep 72:517, 2023.

Link-Gelles R et al: Public health response to a case of paralytic poliomyelitis in an unvaccinated person and detection of poliovirus in wastewater - New York, June–August 2022. Morb Mortal Wkly Rep 19;1065, 2022.

Ma KC et al: Increase in acute respiratory illnesses among children and adolescents associated with rhinoviruses and enteroviruses, including enteroviris D68–United States, July–September 2022. MMWR Morb Mortal Wkly Rep 71:1265, 2022.

Pallansch MA: Circulating poliovirus in New York—new instance of an old problem. N Engl J Med 387:1725, 2022.

Tomatis Souverbielle C et al: Update on nonpolio enterovirus and parechovirus infections in neonates and young infants. Curr Opin Pediatr 35:380, 2023.

211

Measles (Rubeola)

Alex C. Kong, William J. Moss

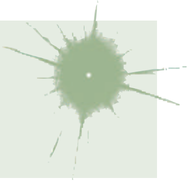

DEFINITION

Measles is a highly contagious viral disease characterized by a prodromal illness of fever, cough, coryza, and conjunctivitis followed by the appearance of a generalized maculopapular rash. Before the widespread use of measles vaccines, it was estimated that measles caused >2 million deaths worldwide each year.

GLOBAL CONSIDERATIONS

Remarkable progress has been made in reducing global measles incidence and mortality rates through measles vaccination. In the Americas, intensive vaccination and surveillance efforts—based in part on the successful Pan American Health Organization strategy of periodic nationwide measles vaccination campaigns (supplementary immunization activities [SIAs])—and high levels of routine measles vaccine coverage interrupted endemic transmission of measles virus. The World Health Organization's (WHO's) Region of the Americas was declared to have eliminated measles in September 2016—the first region in the world to do so. However, endemic measles virus transmission was reestablished, and the region lost its measles elimination status. As such, no WHO region has achieved and sustained measles elimination status, highlighting the importance of maintaining high measles vaccination coverage.

In the United States, high-level coverage with two doses of measles vaccine eliminated endemic measles virus transmission in 2000. However, imported cases and low measles vaccine coverage in some communities threaten this goal. The 1274 measles cases reported in the United States in 2019 represent the highest count since 1992. Historically low levels of measles cases were reported early after the COVID-19 pandemic began but were not sustained, with resurgence of measles globally and in the United States.

Progress also has been made in reducing measles incidence and mortality rates in sub-Saharan Africa and Asia because of increasing routine measles vaccine coverage and provision of a second dose of measles vaccine through mass measles vaccination campaigns and routine childhood immunization programs. From 2000 to 2022, the estimated annual number of global measles deaths per year decreased 82%, from 772,854 (95% confidence interval [CI]: 580,969–1,064,580) to 136,216 (95% CI: 97,058–190,234). Measles vaccination prevented an estimated 57.2 million deaths over this period. Despite this progress, the fact that >100,000 children die each year from a preventable disease such as measles attests to the need for greater resources and efforts to identify and reach unvaccinated children.

The COVID-19 pandemic caused severe disruptions to immunization activities, further threatening progress toward measles control and elimination. Almost 40 million children were estimated to have missed a dose of measles vaccine in 2021, including 25 million children who missed their first dose. Large-scale measles outbreaks occurred in 22 countries in 2021 and 37 countries in 2022. Over the same period, the estimated number of measles cases and deaths increased by 18% and 43%, respectively.

The Measles and Rubella Partnership (MRP)—formerly the Measles and Rubella Initiative (MRI)—is working to improve immunization coverage and address setbacks caused by the pandemic. Since its inception in 2001, MRI has played an important role in reducing global measles incidence and mortality rates, providing governments and communities in 88 countries with technical and financial support for routine immunization activities, mass vaccination campaigns, and disease surveillance systems. In 2023, MRI was rebranded as MRP, and the partnership that had historically been led by the American Red Cross, the United Nations Foundation, UNICEF, and the U.S. Centers for Disease Control and Prevention (CDC) expanded to formally include longtime partners Gavi, the Vaccine Alliance and the Bill and Melinda Gates Foundation as core partners.

■ ETIOLOGY

Measles virus is a spherical, nonsegmented, single-stranded, negative-sense RNA virus and a member of the *Morbillivirus* genus in the family Paramyxoviridae. Measles was originally a zoonotic infection, arising from animal-to-human transmission of an ancestral morbillivirus thousands of years ago, when human populations attained sufficient size to sustain virus transmission. Although RNA viruses typically have high mutation rates, measles virus is an antigenically monotypic virus, i.e., the surface proteins responsible for inducing protective immunity retained their antigenic structure across time and distance because of their key role in binding cellular receptors. The public health significance of this stability is that measles vaccines developed decades ago from a single strain of measles virus remain protective worldwide. Both wild-type and attenuated measles viruses are inactivated by ultraviolet light and heat, necessitating a cold chain for vaccine transport and storage.

■ EPIDEMIOLOGY

Measles virus is one of the most highly contagious directly transmitted pathogens. Outbreaks can occur in populations in which <10% of persons are susceptible. Chains of transmission are common among household contacts, school-age children, and health care workers. There are no latent or persistent measles virus infections that result in a prolonged infectious period, nor are there animal reservoirs for the virus. Thus, measles virus can be maintained in human populations only by an unbroken chain of acute infections, which requires a continuous supply of susceptible individuals. Newborns become susceptible to measles virus infection when passively acquired maternal antibodies are lost, generally before 6–9 months of age. When not immunized, these infants account for the bulk of new susceptible individuals that sustain measles virus transmission.

Endemic measles has a typical temporal pattern characterized by yearly seasonal epidemics superimposed on longer epidemic cycles of 2–5 years or more. In temperate climates, annual measles outbreaks typically occur in the late winter and early spring. These annual outbreaks are probably attributable to social networks facilitating transmission (e.g., congregation of children at school) and environmental factors favoring the viability and transmission of measles virus. Measles cases continue to occur during interepidemic periods in large populations but at low incidence. The longer epidemic cycles occurring every several years result from the accumulation of susceptible persons over successive birth cohorts and the subsequent decline in the number of susceptibles following an outbreak.

Secondary attack rates among susceptible household and institutional contacts generally exceed 90%. The average age at which measles occurs depends on rates of contact with infected persons, protective maternal antibody decline, and vaccine coverage. In densely populated urban settings with low-level vaccination coverage, measles is a disease of infants and young children. The cumulative incidence can reach 50% by 1 year of age, with a significant proportion of children acquiring measles before 9 months—the age at which the first of two routine vaccine doses are administered in many countries, in line with the schedule recommended by the WHO's Expanded Programme on Immunization. As measles vaccine coverage increases or population density decreases, the age distribution shifts toward older children. In such situations, measles cases predominate in school-age children. Infants and young children, although susceptible if not protected by maternal antibodies or vaccination, are not exposed to measles virus at a rate sufficient to cause a heavy disease burden in this age group. As vaccination coverage increases further, the age distribution of cases may be shifted into adolescence and adulthood. This distribution is seen in measles outbreaks in the United States and necessitates targeted measles vaccination programs for these older age groups. Some countries have a bimodal distribution, with measles cases predominantly in young infants and adults.

Persons with measles are infectious for several days before and after the onset of rash, when levels of measles virus in blood and body fluids are highest and when cough, coryza, and sneezing that facilitate virus spread are most severe. The contagiousness of measles before the onset of recognizable disease hinders the effectiveness of isolation measures.

Medical settings are well-recognized sites of measles virus transmission. Children may present to health care facilities during the prodrome, when the diagnosis is not obvious, although the child is infectious and is likely to infect susceptible contacts. Susceptible health care workers can acquire measles from infected children and transmit measles virus to others. Nosocomial transmission can be reduced by maintenance of a high index of clinical suspicion particularly during outbreaks, use of appropriate isolation precautions when measles is suspected, administration of measles vaccine to susceptible children and health care workers, and documentation of health care workers' immunity to measles (i.e., proof of receipt of two doses of measles vaccine or detection of IgG antibodies to measles virus).

As efforts at measles control are increasingly successful, public perceptions of the risk of measles diminish and may be replaced by concerns about possible adverse events associated with measles vaccine. Consequently, measles outbreaks have occurred because of opposition to vaccination on religious or philosophical grounds or unfounded fears of serious adverse events (see "Active Immunization," below, and Chap. 3).

■ PATHOGENESIS

Measles virus is transmitted primarily by respiratory droplets over short distances and, less commonly, by small-particle aerosols that remain suspended in the air for long periods. Airborne transmission appears to be important in certain settings, including schools, physicians' offices, hospitals, and enclosed public places. The virus can be transmitted by direct contact with infected secretions but does not survive for long on fomites.

The incubation period for measles is ~10 days to fever onset and 14 days to rash onset. This period may be shorter in infants and longer (up to 3 weeks) in adults. Infection is initiated when measles virus is deposited in the respiratory tract, oropharynx, or conjunctivae **(Fig. 211-1*A*)**. During the first 2–4 days after infection, measles virus proliferates locally in the respiratory mucosa, primarily in dendritic cells and lymphocytes, and spreads to draining lymph nodes. Virus then enters the bloodstream by budding from infected lymphocytes, producing the viremia that disseminates infection throughout the body. Replication of measles virus in the target organs, together with the host's immune response, are responsible for the signs and symptoms of measles that occur 8–12 days after infection and mark the end of the incubation period **(Fig. 211-1*B*)**.

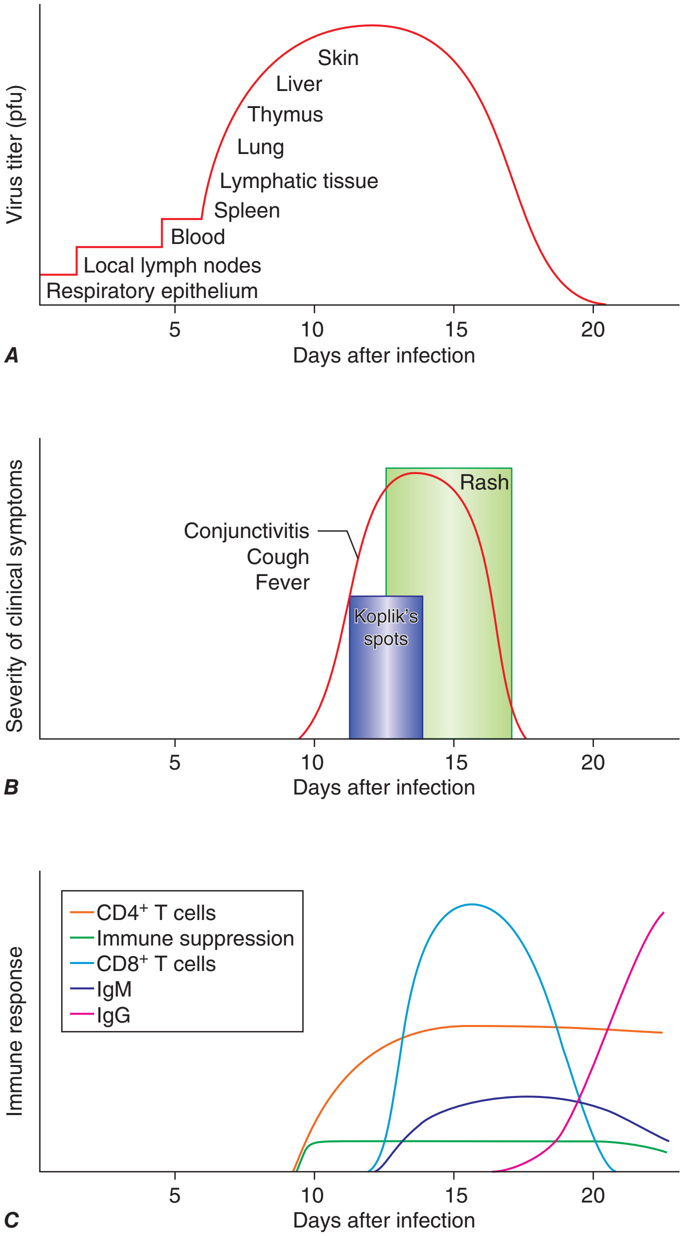

FIGURE 211-1 Measles virus infection: pathogenesis, clinical features, and immune responses. ***A.*** Spread of measles virus, from initial infection of the respiratory tract through dissemination to the skin. ***B.*** Appearance of clinical signs and symptoms, including Koplik's spots and rash. ***C.*** Antibody and T-cell responses to measles virus. The signs and symptoms of measles arise coincident with the host immune response. *(Reproduced with permission from WJ Moss and DE Griffin: Global measles elimination. Nat Rev Microbiology 4:900, 2006.)*

IMMUNE RESPONSES

Host immune responses to measles virus are essential for viral clearance, clinical recovery, and the establishment of long-term protective immunity **(Fig. 211-1*C*)**. Early nonspecific (innate) immune responses during the prodromal phase include activation of natural killer cells and increased production of antiviral proteins. The adaptive immune responses consist of measles virus–specific antibody and cellular responses. The protective efficacy of antibodies to measles virus is illustrated by the immunity conferred to infants from passively acquired maternal antibodies and the protection of exposed, susceptible individuals after administration of anti–measles virus immunoglobulin. The first measles virus–specific antibodies produced after infection are of the IgM subtype, with a subsequent switch to predominantly IgG1 and IgG3 isotypes. The IgM antibody response is typically absent following reexposure or revaccination and serves as a marker of primary infection.

The importance of cellular immunity to measles virus is demonstrated by the ability of children with agammaglobulinemia (congenital inability to produce antibodies) to recover fully from measles and the contrasting picture for children with severe defects in T lymphocyte function who often develop severe or fatal disease **(Chap. 362)**. The initial predominant T_H1 response (characterized by interferon-γ) is essential for viral clearance, and the later T_H2 response (characterized by interleukin-4) promotes the development of measles virus–specific antibodies that are critical for protection against reinfection.

The duration of protective immunity following wild-type measles virus infection is generally thought to be lifelong. Immunologic memory to measles virus includes both continued production of measles virus–specific antibodies by long-lived plasma cells and memory B cells as well as circulation of measles virus–specific CD4+ and CD8+ T lymphocytes.

However, the intense immune responses induced by measles virus infection are paradoxically associated with depressed responses to unrelated (non–measles virus) antigens. This state of immunosuppression persists for at least several weeks to months beyond resolution of the acute illness, enhances susceptibility to secondary infections with bacteria and viruses that cause pneumonia and diarrhea, and is thus responsible for a substantial proportion of measles-related morbidity and deaths. Delayed-type hypersensitivity responses to recall antigens, such as tuberculin, are suppressed, and cellular and humoral responses to new antigens are impaired. Reactivation of latent tuberculosis and remission of autoimmune diseases after measles have been described and are attributed to this period of immune suppression. Importantly, measles results in reductions in the magnitude and diversity of antibodies against previously encountered viral and bacterial antigens, impairing immunologic memory. This mechanism may explain why child morbidity and mortality due to other infectious diseases may be increased for >2 years after measles.

APPROACH TO THE PATIENT

Measles

Clinicians should consider measles in persons presenting with fever and generalized erythematous rash, particularly when measles virus is known to be circulating or the patient has a history of travel to endemic areas. Appropriate precautions must be taken to prevent nosocomial transmission. The diagnosis requires laboratory confirmation except during large outbreaks in which an epidemiologic link to a confirmed case can be established. Care is largely supportive and consists of the administration of vitamin A and antibiotics for secondary bacterial infections (see "Treatment," below). Complications of measles, including bacterial infections and encephalitis, may occur after acute illness and require careful monitoring, particularly in immunocompromised persons.

CLINICAL MANIFESTATIONS

In most persons, the signs and symptoms of measles are highly characteristic (Fig. 211-1*B*). Fever and malaise beginning ~10 days after exposure are followed by cough, coryza, and conjunctivitis. These signs and symptoms increase in severity over 4 days. Koplik's spots **(see Fig. A1-2)** develop on the buccal mucosa ~2 days before the rash appears. Koplik's spots are pathognomonic of measles and consist of bluish white dots ~1 mm in diameter surrounded by erythema. The lesions appear first on the buccal mucosa opposite the lower molars but rapidly increase in number and may involve the entire buccal mucosa. They fade with the onset of rash.

The characteristic rash of measles **(see Fig. A1-3)** begins 2 weeks after infection, when the clinical manifestations are most severe, and signal the host's immune response to the replicating virus. Headache, abdominal pain, vomiting, diarrhea, and myalgia may be present. The rash of measles begins as erythematous macules behind the ears and on the neck and hairline. The rash progresses to involve the face, trunk, and arms, with involvement of the legs and feet by the end of the second day. Areas of confluent rash appear on the trunk and extremities, and petechiae may be present. The rash fades slowly in the same order of progression as it appeared, usually beginning on the third or fourth day after onset. Resolution of the rash may be followed by desquamation, particularly in undernourished children.

Because the characteristic rash of measles is a consequence of the cellular immune response, it may not develop in persons with impaired cellular immunity. These persons have a high case–fatality rate and frequently develop giant cell pneumonitis caused by measles virus.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of measles includes other causes of fever, morbilliform rash, and conjunctivitis, including rubella, Kawasaki disease, infectious mononucleosis, roseola, scarlet fever, Rocky Mountain spotted fever, enterovirus or adenovirus infection, and drug sensitivity. Rubella is a milder illness without cough and with distinctive posterior auricular or suboccipital lymphadenopathy. The rash of roseola (exanthem subitum) **(see Fig. A1-5)** appears after fever has subsided. The atypical lymphocytosis in infectious mononucleosis contrasts with the leukopenia commonly observed in children with measles.

DIAGNOSIS

Measles is readily diagnosed on clinical grounds by clinicians familiar with the disease, particularly during outbreaks. Koplik's spots are especially helpful because they appear early and are pathognomonic. Clinical diagnosis is more difficult (1) during the prodromal illness; (2) when the rash is attenuated by passively acquired antibodies or prior immunization; (3) when the rash is absent or delayed in immunocompromised children or severely undernourished children with impaired cellular immunity; and (4) in regions where the incidence of measles is low and other pathogens are responsible for most illnesses with fever and rash. The CDC case definition for measles requires (1) a generalized maculopapular rash of at least 3 days' duration; (2) fever of at least 38.3°C (101°F); and (3) cough, coryza, or conjunctivitis.

Serology is the most common method of laboratory diagnosis. The detection of measles virus–specific IgM in a single specimen of serum or oral fluid is considered diagnostic of acute infection, as is a fourfold or greater increase in measles virus–specific IgG antibody levels between acute- and convalescent-phase serum specimens. Primary infection in the immunocompetent host results in antibodies that are often detectable within 1–3 days of rash onset and reach peak levels in 2–4 weeks. However, measles virus–specific IgM antibodies may not be detectable until 4–5 days or more after rash onset, resulting in false-negative test results if the specimen is obtained too early, and usually fall to undetectable levels within 4–8 weeks of rash onset.

Several methods for measurement of antibodies to measles virus are available. Neutralization tests are sensitive and specific, and the results are highly correlated with protective immunity. However, these tests require propagation of measles virus in cell culture and thus are expensive and laborious. Commercially available measles IgM enzyme immunoassays are most frequently used. Measles can also be diagnosed by isolation of the virus in cell culture from respiratory secretions, nasopharyngeal or conjunctival swabs, blood, or urine. Direct detection of giant cells in respiratory secretions, urine, or tissue obtained by biopsy provides another method of diagnosis.

For detection of measles virus RNA by reverse-transcription polymerase chain reaction, primers targeted to highly conserved regions of measles virus genes are used. Extremely sensitive and specific, this assay may also permit identification and characterization of measles virus genotypes for molecular epidemiologic studies and can distinguish wild-type from vaccine virus strains.

TREATMENT

Measles

There is no specific antiviral therapy for measles. Treatment consists of general supportive measures such as hydration and administration of antipyretic agents. Because secondary bacterial infections are a major cause of morbidity and death attributable to measles, effective case management involves prompt antibiotic treatment for patients who have clinical evidence of bacterial infection, including pneumonia and otitis media.

Vitamin A (available in oral and parenteral formulations) is effective for the treatment of measles and can markedly reduce rates of morbidity and mortality. The WHO recommends administration of once-daily oral doses of 200,000 IU of vitamin A for 2 consecutive days to all children with measles who are ≥12 months of age. Lower doses are recommended for younger children: 100,000 IU per day for children 6–11 months of age and 50,000 IU per day for children <6 months old. A third dose is recommended 2–6 weeks later for children with evidence of vitamin A deficiency. While such deficiency is not a widely recognized problem in the United States, many American children with measles do, in fact, have low serum levels of vitamin A, and these children experience increased measles-associated morbidity.

Anecdotal reports have described the recovery of previously healthy pregnant and immunocompromised patients with measles pneumonia and of immunocompromised patients with measles encephalitis after treatment with aerosolized and IV ribavirin. However, the clinical benefits of ribavirin in measles have not been conclusively demonstrated in clinical trials.

COMPLICATIONS

Most complications of measles involve the respiratory tract and include the effects of measles virus itself and secondary bacterial infections. Giant cell pneumonitis due to replication of measles virus in the lungs can develop in immunocompromised persons. Acute laryngotracheobronchitis (croup) can occur during measles and may result in airway obstruction, particularly in young children. Many children with measles develop diarrhea, which contributes to and can exacerbate existing undernutrition.

Most complications of measles result from secondary bacterial infections of the respiratory tract that are attributable to a state of immune suppression after acute measles. Otitis media and bronchopneumonia are most common. Recurrence of fever or failure of fever to subside with the rash suggests secondary bacterial infection.

Severe complications of measles involve the central nervous system (CNS). Post-measles encephalomyelitis complicates ~1 in 1000 cases, affecting mainly older children and adults. Encephalomyelitis occurs within 2 weeks of rash onset and is characterized by fever, seizures, and a variety of neurologic abnormalities. The finding of periventricular demyelination, the induction of immune responses to myelin basic protein, and the absence of measles virus in the brain suggest that post-measles encephalomyelitis is an autoimmune disorder triggered by measles virus infection. Rarer CNS complications that occur months to years after acute infection are measles inclusion body encephalitis (MIBE) and subacute sclerosing panencephalitis (SSPE). In contrast to post-measles encephalomyelitis, MIBE and SSPE are caused by persistent measles virus infection. MIBE is a rare but fatal complication that affects individuals with defective cellular immunity and typically occurs months after infection. SSPE is a slowly progressive disease characterized by seizures and progressive deterioration of cognitive and motor functions, with death occurring 5–15 years after measles virus infection. SSPE most often develops in persons infected with measles virus at <2 years of age.

PROGNOSIS

Most persons with measles recover and develop long-term protective immunity to reinfection. Measles case–fatality proportions vary with the average age of infection, the nutritional and immunologic status of the population, measles vaccine coverage, and access to health care. Among previously vaccinated persons who do become infected, disease is less severe and mortality rates are significantly lower. In most developed countries, the case–fatality rate is 0.01–0.1%, but in endemic areas of sub-Saharan Africa, the measles case–fatality rate may be 5–10% or even higher. Measles is a major cause of childhood deaths in refugee camps and in internally displaced populations, where case–fatality rates have been as high as 20–30%.

PREVENTION

Passive Immunization Human immunoglobulin given shortly after exposure can attenuate the clinical course of measles. In immunocompetent persons, administration of immunoglobulin within 72 h of exposure usually prevents measles virus infection and almost always prevents clinical measles. Administered up to 6 days after exposure, immunoglobulin will still prevent or modify the disease. Prophylaxis with immunoglobulin is recommended for susceptible household and nosocomial contacts who are at risk of developing severe measles, particularly children <1 year of age, immunocompromised persons (including immunocompromised persons living with HIV who were previously immunized with live attenuated measles vaccine), and pregnant women. Except for premature infants, children <6 months of age usually will be partially or completely protected by passively acquired maternal antibody. Infants born to women with vaccine-induced measles immunity become susceptible to measles at a younger age than infants born to women with acquired immunity from natural infection. If measles is diagnosed in a household member, all unimmunized children in the household should receive immunoglobulin. The recommended dose is 0.5 mL/kg given intramuscularly with a maximum total dose of 15 mL. Immunocompromised and pregnant persons should receive 400 mg/kg intravenously. IV immunoglobulin contains antibodies to measles virus, and the usual dose of 100–400 mg/kg generally provides adequate prophylaxis for measles exposures occurring as long as 3 weeks or more after IV immunoglobulin administration.

Active Immunization The first live attenuated measles vaccine was developed by passage of the Edmonston strain in chick embryo fibroblasts to produce the Edmonston B virus, which was licensed in 1963 in the United States but was reactogenic. Further passage of Edmonston B virus produced the more attenuated and less reactogenic Schwarz vaccine. The Moraten ("more attenuated Enders") strain, which was licensed in 1968 and is used in the United States, is genetically identical to the Schwarz strain. The Edmonston-Zagreb vaccine, also derived from the Edmonston B strain, is widely used in many countries and was passaged in human diploid cells.

Lyophilized measles vaccines are relatively stable, but reconstituted vaccine rapidly loses potency. Live attenuated measles vaccines are inactivated by light and heat and lose about half their potency at 20°C and almost all their potency at 37°C within 1 h after reconstitution. Therefore, a cold chain must be maintained before and after reconstitution. Antibodies first appear 12–15 days after vaccination, and titers peak at 1–3 months. Measles vaccines are often combined with other live attenuated virus vaccines, such as those for mumps and rubella (MMR) and for mumps, rubella, and varicella (MMRV).

The recommended age of first vaccination varies from 6 to 15 months and represents a balance between the optimal age for seroconversion and the probability of acquiring measles before that age. The proportions of children who develop protective levels of antibody after the first measles vaccination approximate 85% at 9 months of age and 95% at 12 months. Common childhood illnesses concomitant with vaccination may reduce the level of immune response, but such illnesses are not valid reasons to withhold vaccination. Measles vaccines have been well tolerated and immunogenic in children and adults living with HIV, although antibody levels may wane more rapidly. Because of the potential severity of wild-type measles virus infection in children living with HIV, routine measles vaccination is recommended except for those who are severely immunocompromised. Measles vaccination is contraindicated in individuals with other severe deficiencies of cellular immunity because of the possibility of disease due to progressive pulmonary or CNS infection with the vaccine virus.

The duration of vaccine-induced immunity is at least several decades, if not longer. Rates of secondary vaccine failure 10–15 years after immunization have been estimated at ~5% but are likely lower when vaccination takes place after 12 months of age. Decreasing antibody concentrations do not necessarily imply a complete loss of protective immunity as a secondary immune response usually develops after reexposure to measles virus, with a rapid rise in antibody titers in the absence of overt clinical disease.

Standard doses of currently licensed measles vaccines are safe for immunocompetent children and adults. Fever up to 39.4°C (103°F) occurs in ~5–15% of seronegative vaccine recipients, and ~5% of vaccine recipients develop a transient rash. Mild transient thrombocytopenia has been reported, with an incidence of 1 case per ~40,000 MMR recipients.

Since the publication of a report in 1998 falsely hypothesizing that MMR vaccine may cause a syndrome of autism and intestinal inflammation, much public attention has focused on this purported association. The events that followed publication of this report led to diminished vaccine coverage in the United Kingdom and provide important lessons in the misinterpretation of epidemiologic evidence and the communication of scientific results to the public. The publication that incited the concern was a case series describing 12 children with a regressive developmental disorder and chronic enterocolitis; 9 of these children had autism. In 8 of the 12 cases, the parents associated onset of the developmental delay with MMR vaccination. This simple temporal association was misinterpreted and misrepresented as a possible causal relationship, first by the lead author of the study and then by elements of the media and the public. Subsequently, many comprehensive reviews and additional epidemiologic studies refuted evidence of a causal relationship between MMR vaccination and autism, and the offending publication was retracted.

PROSPECTS FOR MEASLES ERADICATION

Progress in global measles control has renewed discussion of measles eradication. In contrast to poliovirus eradication, the eradication of measles virus will not entail challenges posed by prolonged shedding of potentially virulent vaccine viruses and asymptomatic reservoirs. However, in comparison with smallpox eradication, higher levels of population immunity will be necessary to interrupt measles virus transmission, more highly skilled health care workers will be required to administer measles vaccines, and containment through case detection and ring vaccination will be more difficult for measles virus because of infectivity before rash onset.

New tools, such as microarray patches (MAPs) to deliver measles vaccine, could facilitate mass vaccination campaigns and vaccination of hard-to-reach children such as those residing in remote rural areas. In May 2023, Micron Biomedical reported positive results in a phase 1/2 clinical trial in The Gambia comparing the results of immunization by measles- and rubella-containing (MR) vaccine administered by MAP versus subcutaneous injection. This trial was the first of its kind to use microarray technology in children and found similar rates of seroconversion in MR vaccine–naïve children (92.9–100% for MAP and 89.7–100% for subcutaneous) and seroprotection in all age groups (93.2–100% for MAP and 89.8–100% for subcutaneous). Further studies and the overcoming of manufacturing and regulatory hurdles will be needed before MAPs become available.

Despite enormous progress, measles remains a leading vaccine-preventable cause of childhood mortality worldwide and continues to cause outbreaks in communities with low vaccination coverage rates. As the world looks to rebuild immunization services disrupted by the COVID-19 pandemic and improve these services through the ambitious MRP agenda, measles outbreaks will continue to remind us of the challenges to be overcome.

Acknowledgment

The authors acknowledge the substantial contributions of Kaitlin Rainwater-Lovett, who coauthored prior editions of this chapter.

FURTHER READING

De Swart RL, Moss WJ: The immunological basis for immunization series: Module 7: Measles. Update 2020. Geneva: World Health Organization, 2020.

Griffin DE: Measles immunity and immunosuppression. Curr Opin Virol 46:9, 2020.

Hübschen JM et al: Measles. Lancet 399:678, 2022.

Mina MJ et al: Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens. Science 366:599, 2019.

Minta AA et al: Progress toward measles elimination—Worldwide, 2000–2022. MMWR Morb Mortal Wkly Rep 72:1262, 2023.

Moss WJ: Measles. Lancet 380:2490, 2017.

Moss WJ et al: Feasibility assessment of measles and rubella eradication. Vaccine 39:3544, 2021.

Phadke VK et al: Vaccine refusal and measles outbreaks in the US. JAMA 324:1344, 2020.

Strebel PM, Orenstein WA: Measles. N Engl J Med 381:349, 2019.

World Health Organization: Measles vaccines: WHO position paper—April 2017. Wkly Epidemiol Rec 92:205, 2017.

212

Rubella (German Measles)

Alan C. Ou, Ludmila M. Perelygina,
Laura A. Zimmerman, Susan E. Reef

Rubella was historically viewed as a variant of measles or scarlet fever. After an epidemic of rubella in Australia in the early 1940s, the ophthalmologist Norman Gregg noticed the occurrence of congenital cataracts among infants whose mothers had reported rubella during early pregnancy, and congenital rubella syndrome (CRS; see "Clinical Manifestations," below) was first described. Not until 1962 was a separate viral agent for rubella isolated.

ETIOLOGY

Rubella virus is a member of the Matonaviridae family the genus *Rubivirus*. This single-strand RNA enveloped virus measures 40–80 nm in diameter. Its nucleocapsid consists of ~10-kb positive-sense RNA genome surrounded by a protein shell composed of a core protein and a single-layer lipoprotein envelope with spike-like projections containing two glycoproteins, E1 and E2. There is only one antigenic type of rubella virus, and humans are its only known reservoir.

PATHOGENESIS AND PATHOLOGY

Although the pathogenesis of postnatal (acquired) rubella has been well documented, data on pathology are limited because of the mildness of the disease. Rubella virus is spread from person to person via respiratory droplets. Primary implantation and replication in the nasopharynx are followed by spread to the lymph nodes. Subsequent viremia occurs, which in pregnant women often results in infection of the placenta. Placental virus replication may lead to infection of fetal organs. The pathology of CRS in the infected fetus is well defined, with almost all organs found to be infected; however, the pathogenesis of CRS is only poorly delineated. In tissue, infections with rubella virus have diverse effects, ranging from no obvious impact to cell destruction. The hallmark of fetal infection is a chronic infection with persistence throughout fetal development in utero and for up to 1 year after birth.

Individuals with acquired rubella may shed virus from 7 days before rash onset to ~5–7 days thereafter. Both clinical and subclinical infections are considered contagious. Infants with CRS may shed large quantities of virus from bodily secretions, particularly from the throat and in the urine, up to 1 year of age. Outbreaks of rubella, including some in nosocomial settings, have originated with index cases of CRS. Thus, only individuals immune to rubella virus should have contact with infants who have CRS or who are congenitally infected with rubella virus but are not showing signs of CRS.

EPIDEMIOLOGY

The largest recent rubella epidemic in the United States took place in 1964–1965, when an estimated 12.5 million cases occurred, resulting in ~20,000 cases of CRS. Since the introduction of the routine rubella vaccination program in the United States in 1969, the number of rubella cases reported each year has dropped by >99%; the rate of vaccination coverage with rubella-containing vaccine (RCV) has been >90% among children 19–35 months old since 1996. In the United States, a goal for the elimination of rubella and CRS by 2000 was set in 1989. Interruption of endemic transmission of rubella virus was achieved by 2001. In 2004, a panel of experts agreed unanimously that rubella was no longer an endemic disease in the United States. The criteria used to document lack of endemic transmission included low disease incidence, high nationwide rubella antibody seroprevalence, outbreaks that were few and contained (i.e., small numbers of cases), and lack of endemic virus transmission (as assessed by genetic sequencing). Although interruption of endemic transmission has been sustained since 2001, rubella virus importations continue to occur, and cases continue to develop among susceptible persons. During 2010–2022, 66 cases of rubella were reported; 71% of these cases were in persons 20–49 years old—an age group that includes women of childbearing age. During this period, 13 cases of CRS were reported, all from foreign-born mothers. Therefore, health care providers should remain vigilant, considering the possibility of rubella virus infection in adults (especially those emigrating or returning from countries without rubella control programs) and recognizing the potential for CRS among their infants.

The Global Measles and Rubella Strategic Framework 2021–2030 envisions a "world free from measles and rubella" with the goal to "achieve and sustain the regional measles and rubella elimination goals." By 2023, five of the six World Health Organization (WHO) regions had rubella elimination goals (see "Prevention" below). Although rubella and CRS are no longer endemic in the WHO Region of the Americas, they remain important public health problems globally. The number of rubella cases reported worldwide in 2000 was ~700,000; this figure declined to 17,407 in 2022. However, the number of rubella cases may be underestimated because cases are often mild, patients may not seek care, cases may not be recognized or may not be reported, and, in some countries, cases are identified through measles surveillance systems that are not specific for rubella. Despite a continued increase in the number of countries with rubella vaccination programs, 25% of the world's children remained unvaccinated against rubella in 2022. In 2010, it was estimated that 105,000 cases of CRS occurred annually globally.

CLINICAL FEATURES

Acquired Rubella Acquired rubella commonly presents with a generalized maculopapular rash that usually lasts up to 3 days **(Fig. 212-1)**, although as many as 50% of cases may be subclinical or without rash. When the rash occurs, it is usually mild and may be difficult to detect in persons with darker skin. In younger children, rash is usually the first sign of illness. However, in older children and adults, a 1- to 5-day prodrome often precedes the rash and may include low-grade fever, malaise, and upper respiratory symptoms. The incubation period is 14 days (range, 12–23 days).

Lymphadenopathy, particularly occipital and postauricular, may be noted during the second week after exposure. Although acquired rubella is usually thought of as a benign disease, arthralgia and arthritis are common in infected adults, particularly women. Thrombocytopenia and encephalitis are less common complications.

Congenital Rubella Syndrome The most serious consequence of rubella virus infection can develop when a woman becomes infected

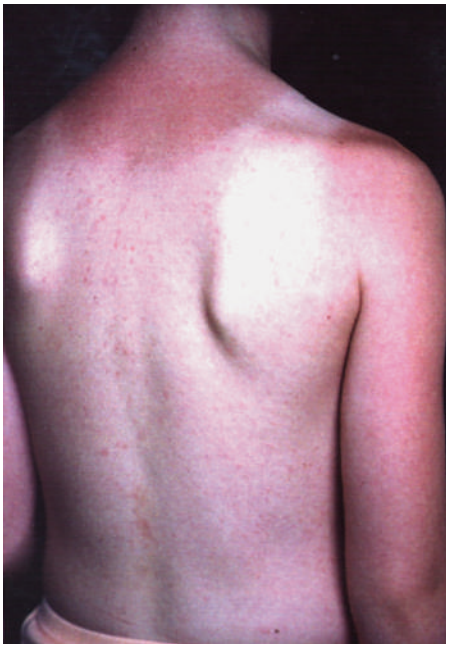

FIGURE 212-1 Mild maculopapular rash of rubella in a child.

during pregnancy, particularly during the first trimester. The resulting complications may include miscarriage, fetal death, premature delivery, or live birth with congenital defects. Infants infected with rubella virus in utero may have myriad physical defects (Table 212-1), which most commonly relate to the eyes, ears, and heart. This constellation of severe birth defects is known as CRS. In addition to permanent manifestations, there are a host of transient physical manifestations, including thrombocytopenia with purpura/petechiae (e.g., dermal erythropoiesis, "blueberry muffin syndrome"). Some infants may be born with congenital rubella virus infection but have no apparent signs or symptoms of CRS and are referred to as "infants with congenital rubella virus infection only."

DIAGNOSIS

Acquired Rubella Clinical diagnosis of acquired rubella is difficult because of the mimicry of many illnesses with rashes, the varied clinical presentations, and the high rates of subclinical and mild disease. Illnesses that may be similar to rubella in presentation include scarlet fever, roseola, toxoplasmosis, fifth disease, measles, Zika, and illnesses with suboccipital and postauricular lymphadenopathy. Thus, laboratory documentation of rubella virus infection is considered the only reliable way to confirm acute disease.

Laboratory assessment of rubella virus infection is conducted by serologic and virus detection methods. For acquired rubella, serologic diagnosis is most common and depends on the demonstration of IgM antibodies in an acute-phase serum specimen or a fourfold rise in IgG antibody titer between acute- and convalescent-phase specimens. To detect a rise in IgG antibody titer indicative of acute disease, the acute-phase serum specimen should be collected within 7–10 days after onset of illness and the convalescent-phase specimen ~14–21 days after the first specimen. The enzyme-linked immunosorbent assay IgM capture technique is considered most accurate for serologic diagnosis, but the indirect IgM assays also are acceptable. After rubella virus infection, IgM antibody may be detectable for up to 6 weeks. In case of a negative result for IgM in specimens taken earlier than day 5 after rash onset, serologic testing should be repeated.

TABLE 212-1 Common Transient and Permanent Manifestations in Infants with Congenital Rubella Syndrome

TRANSIENT MANIFESTATIONS	PERMANENT MANIFESTATIONS
Hepatosplenomegaly	Hearing impairment/deafness
Interstitial pneumonitis	Congenital heart defects (patent ductus arteriosus, pulmonary arterial stenosis)
Thrombocytopenia with purpura/petechiae (e.g., dermal erythropoiesis or "blueberry muffin syndrome")	Eye defects (cataracts, cloudy cornea, microphthalmos, pigmentary retinopathy, congenital glaucoma)
Hemolytic anemia	Microcephaly
Bony radiolucencies	Central nervous system sequelae (mental and motor delay, autism)
Intrauterine growth retardation	
Adenopathy	
Meningoencephalitis	

Although uncommon, reinfection with rubella virus is possible, and IgM antibodies may be present. In this instance, IgG avidity testing is used in conjunction with IgG testing to distinguish primary rubella infection from reinfection. The detection of low-avidity antibodies in a patient's serum indicates recent infection. The presence of mature (high-avidity) IgG antibodies most likely indicates an infection occurring at least 2 months previously. Avidity testing may be particularly useful in diagnosing rubella in pregnant women and assessing the risk of CRS.

Rubella virus is typically detected in the nasopharynx during the prodromal period and for as long as 2 weeks after rash onset. However, viral specimens (nasopharyngeal swabs are preferred but throat swabs or urine are also acceptable) should be collected as soon after symptom onset as possible, preferably 1–3 days after onset, but no later than 7 days after onset. Rubella is usually diagnosed by viral RNA detection in a reverse-transcriptase polymerase chain reaction (RT-PCR) assay; rubella virus isolation can also be used to diagnose rubella.

Congenital Rubella Syndrome The classic triad of CRS—clinical manifestations of cataracts, hearing impairment, and heart defects—is seen in ~10% of infants with CRS. Infants may present with different combinations of defects depending on when infection occurs during gestation. Hearing impairment is the most common single defect of CRS. However, as with acquired rubella, laboratory diagnosis of congenital infection is highly recommended, particularly because most features of the clinical presentation are nonspecific and may be associated with other intrauterine infections. Early diagnosis of CRS allows the prompt implementation of infection control measures and facilitates appropriate medical intervention for specific disabilities.

Diagnostic tests used to confirm CRS include serologic assays and virus detection. In an infant with congenital infection, serum IgM antibodies are normally present for up to 6 months but may be detectable for up to 1 year after birth. In some instances, IgM may not be detectable until 1 month of age; thus, infants who have symptoms consistent with CRS but who test negative shortly after birth should be retested at 1 month. A rubella serum IgG titer persisting beyond the time expected after passive transfer of maternal IgG antibody (i.e., a rubella titer that does not decline at the expected rate of a twofold dilution per month) is another serologic criterion used to confirm CRS.

In congenital infection, rubella virus is detected most commonly from nasopharyngeal and throat swabs and urine. Infants with congenital rubella may excrete virus for up to 1 year, but specimens for RT-PCR are most likely to be positive if obtained within the first 6 months after birth. Rubella virus in infants with CRS can also be detected by virus culture.

Rubella Diagnosis in Pregnant Women In the United States, screening for rubella IgG antibodies is recommended as part of routine prenatal care. Pregnant women with a positive IgG antibody serologic test are considered immune. Susceptible pregnant women should be vaccinated postpartum.

A susceptible pregnant woman exposed to rubella virus should be tested for IgM antibodies and, if positive, confirmed by testing for low-avidity IgG antibodies to determine whether she was infected during pregnancy. Pregnant women with evidence of acute infection must be clinically monitored, and gestational age at the time of maternal infection must be determined to assess the possibility of risk to the fetus. Among women infected with rubella virus during the first 10 weeks of gestation, the risk of delivering an infant with CRS is 90%. The risk of birth defects declines with infection later in gestation, and fetal defects are rarely associated with maternal rubella after the 16th week of gestation, although sensorineural hearing deficit may occur with infection as late as 20 weeks. Because of the potential for false-positive results, rubella IgM antibody testing is not recommended for pregnant women with no history of illness or contact with a rubella-like illness.

TREATMENT

Rubella

No specific therapy is available for rubella virus infection. Symptom-based treatment for various manifestations, such as fever and arthralgia, is appropriate. Immunoglobulin does not prevent rubella virus infection after exposure and therefore is not recommended as routine postexposure prophylaxis. Although immunoglobulin may modify or suppress symptoms, it can create an unwarranted sense of security: infants with congenital rubella have been born to women who received immunoglobulin shortly after exposure. Administration of immunoglobulin should be considered only if a pregnant woman who has been exposed to a person with rubella will not consider termination of the pregnancy under any circumstances. In such cases, IM administration of 20 mL of immunoglobulin within 72 h of rubella exposure may reduce—but does not eliminate—the risk of rubella.

PREVENTION

After the isolation of rubella virus in the early 1960s and the occurrence of a devastating rubella pandemic in 1964–1965, a vaccine for rubella was developed and licensed in 1969. The majority of rubella-containing vaccines (RCVs) used worldwide are combined measles and rubella (MR) or measles, mumps, and rubella (MMR) formulations. A tetravalent measles, mumps, rubella, and varicella (MMRV) vaccine is available but is not widely used. Available rubella-containing vaccines are live attenuated vaccine virus.

The public health burden of rubella virus infection is measured primarily through the occurrence of CRS cases among women who were infected during pregnancy. The 1964–1965 rubella epidemic in the United States resulted in >30,000 infections during pregnancy. CRS occurred in ~20,000 infants born alive, including >11,000 infants who were deaf, >3500 infants who were blind, and almost 2000 infants with intellectual disability. The medical cost of this epidemic exceeded \$1.5 billion. It has been estimated that the lifetime medical costs for children with CRS range from \$11,255 in low-income countries to \$934,000 in high-income countries.

In most countries, there are few data to document the epidemiology of CRS, but clusters of CRS cases have been reported in developing countries. Before the introduction of routine immunization against rubella in the United States, the incidence of CRS was 0.1–0.2 case per 1000 live births during endemic periods and 1–4 cases per 1000 live births during epidemic periods. Where rubella virus is circulating and women of childbearing age are susceptible, CRS cases will continue to occur.

The most effective method of preventing acquired rubella and CRS is through vaccination with an RCV. One dose induces seroconversion in ≥95% of persons ≥1 year of age. Immunity is considered long-term and is probably lifelong. The most commonly used vaccine globally is the RA27/3 virus strain. The recommendation for routine rubella vaccination schedules in the United States is a first dose of MMR vaccine at 12–15 months of age and a second dose at 4–6 years. Other persons recommended to receive a dose of a rubella-containing vaccine include adolescents and adults without documented evidence of immunity, individuals in congregate settings (e.g., college students, military personnel, childcare and health care workers), international travelers, and susceptible women before and after pregnancy.

Because of the theoretical risk of transmission of live attenuated rubella vaccine virus to the developing fetus, women known to be pregnant should not receive RCV. In addition, pregnancy should be avoided for 28 days after receipt of RCV. In follow-up studies of ~3000 unknowingly pregnant women who received rubella vaccine, no infant was born with CRS. Receipt of RCV during pregnancy is not ordinarily a reason to consider termination of the pregnancy.

In 2022, 175 (90%) of the 194 member countries of the WHO recommended inclusion of RCV in the routine childhood vaccination schedule **(Fig. 212-2)**. Goals for the elimination of rubella and CRS

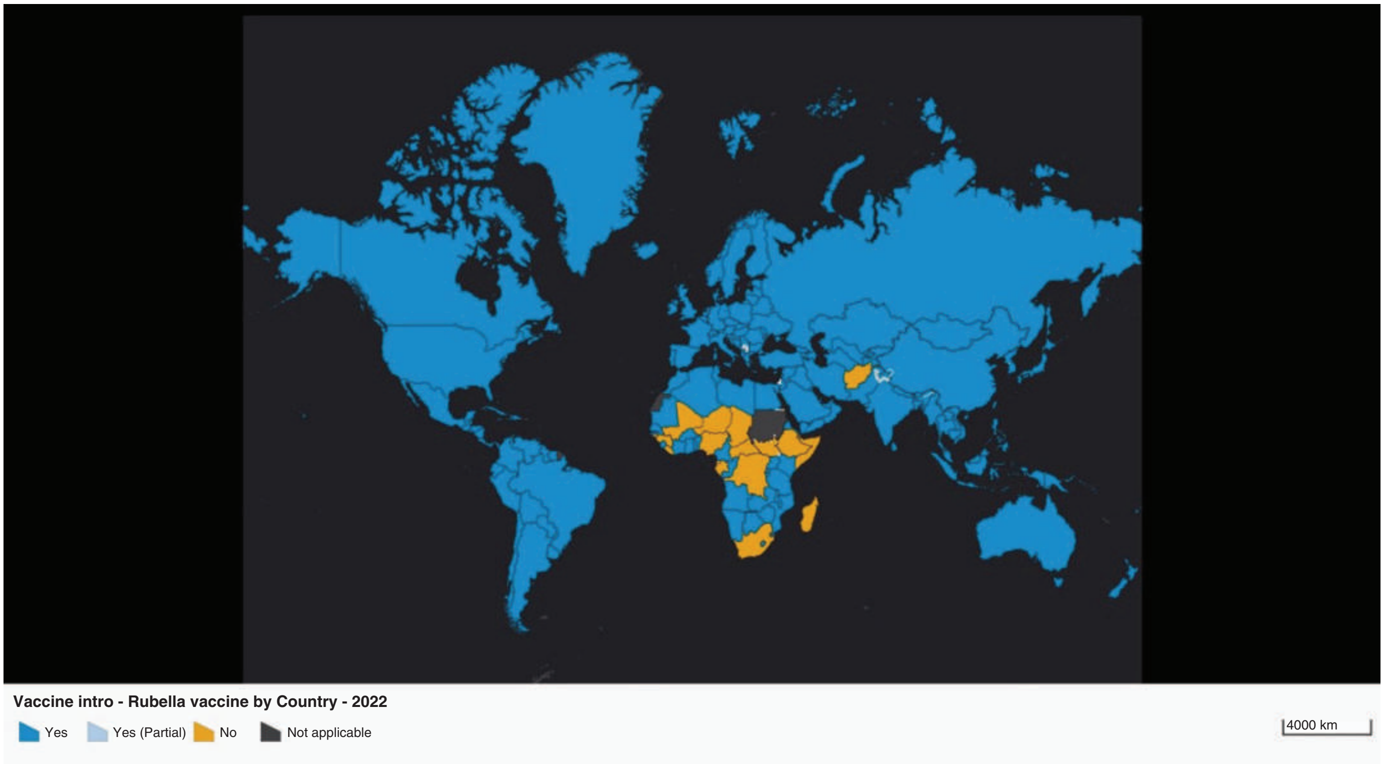

FIGURE 212-2 Countries using rubella vaccine in national childhood immunization schedules, 2022. Disclaimer—The boundaries and names shown and the designations used on this map do not imply the expression or any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city, or area nor of its authorities, or concerning the delimitation of its frontiers or boundaries. *Dotted* or *dashed lines* on maps represent approximate border lines for which there may not be full agreement. *(Reproduced with permission from the World Health Organization; 2024.)*

have been established in the WHO American, European, Southeast Asian, African, and Western Pacific regions. The Eastern Mediterranean region has not yet set such goals. Unfortunately, the COVID-19 pandemic, which began in late 2019, led to extensive disruptions of routine vaccination services in many countries. Thus, it is essential that all children are up to date with rubella vaccination, especially those who missed vaccination during the pandemic. To protect against rubella vaccine throughout the life course, additional strategies to immunize adolescents and adults will be needed to ensure adults of childbearing age are protected from the risk of having an infant with CRS.

FURTHER READING

Centers for Disease Control and Prevention: Control and prevention of rubella: Evaluation and management of suspected outbreaks, rubella in pregnant women, and surveillance for congenital rubella syndrome. MMWR Morb Mortal Wkly Rep 50:1, 2001.

Centers for Disease Control and Prevention: Notice to readers: Revised ACIP recommendation for avoiding pregnancy after receiving a rubella-containing vaccine. MMWR Morb Mortal Wkly Rep 50:1117, 2001.

Centers for Disease Control and Prevention: *Manual for the Surveillance of Vaccine-Preventable Diseases*, SW Roush et al (eds). Atlanta, Centers for Disease Control and Prevention, Last Reviewed: March 6, 2020, Chapters 14–15 (Rubella & Congenital Rubella Syndrome). Available at *https://www.cdc.gov/vaccines/pubs/surv-manual/chapters.html*. Accessed March 5, 2024.

Centers for Disease Control and Prevention: Rubella, in *Epidemiology and Prevention of Vaccine Preventable Diseases*, 14th ed. E Hall et al (eds). Washington, DC, Public Health Foundation, August 2021, Chapter 20. Available at *https://www.cdc.gov/vaccines/pubs/pinkbook/front-matter.html*. Accessed March 5, 2024.

Ou AC et al: Progress toward rubella and congenital rubella syndrome control and elimination–Worldwide, 2012–2022. MMWR Morb Mortal Wkly Rep 73:162, 2024.

Reef SE, Plotkin SA: Rubella vaccine, in *Vaccines*, SA Plotkin, WA Orenstein (eds). Philadelphia, Saunders, 2024, pp 1025–1056.

Thompson K, Odahowski C: The costs and valuation of health impacts of measles and rubella risk management policies. Risk Anal 36:1357, 2016.

Vynnycky E et al: Estimates of the global burden of congenital rubella syndrome, 1996–2019. Int J Infect Dis 137:149, 2023.

World Health Organization: Rubella, module 11, in *The Immunological Basis for Immunization Series*. Geneva, WHO, 2008. Available at *https://www.who.int/publications/i/item/9789241596848*. Accessed August 10, 2023.

World Health Organization: Rubella vaccines: WHO position paper. Wkly Epidemiol Rec 95:301, 2020. Available at *https://www.who.int/publications/i/item/WHO-WER9527*. Accessed August 10, 2023.

World Health Organization: Global Vaccine Action Plan 2011–2020. Geneva, WHO, 2013. Available at *http://www.who.int/immunization/global_vaccine_action_plan/GVAP_doc_2011_2020/en/*. Accessed August 10, 2023.

World Health Organization: Measles and rubella strategic framework 2021–2030. Geneva: WHO; 2020. License: CC BY-NC-SA 3.0 IGO.

World Health Organization: Immunization Agenda 2030: A global strategy to leave no one behind. Available at *https://www.who.int/teams/immunization-vaccines-and-biologicals/strategies/ia2030*. Accessed March 5, 2024.

213

Mumps

Jessica Leung, Nina B. Masters

Mumps is an acute, self-limited, systemic viral illness typically characterized by parotitis or other salivary gland swelling. Although mumps was once considered a universal childhood disease in the United States, routine mumps vaccination—with a one-dose policy implemented in 1977 and a two-dose policy in 1989—led to a >99% reduction in cases by the early 2000s. However, since 2006, there has been an increase in mumps cases in the United States, the majority among fully vaccinated persons. Mumps should be suspected in all patients with parotitis or mumps complications (see "Clinical Manifestations"), regardless of age, vaccination status, or travel history.

ETIOLOGIC AGENT

Mumps is an acute viral illness caused by a paramyxovirus from the *Rubulavirus* genus in the *Paramyxoviridae* family. This single-stranded, negative-sense, enveloped RNA virus is ~15.3 kb in size and encodes several minor proteins and seven major proteins. There is only one mumps virus serotype. One of the seven major encoded proteins, the small hydrophobic (SH) protein exhibits hypervariability among strains; thus, the SH gene nucleotide sequence (316 nucleotides) is used to genotype the virus for molecular epidemiologic purposes. Mumps virus is rapidly inactivated by formalin, ether, chloroform, heat, and ultraviolet light.

The 12 known genotypes of mumps virus are designated by the letters A to N (except E and M). In the United States, >98% of mumps virus specimens genotyped from 2015 through 2017 were genotype G. Most mumps vaccines licensed globally are composed of virus strains from genotype A, B, or N. The mumps virus strain (Jeryl Lynn) used in vaccines in the United States is genotype A.

EPIDEMIOLOGY

Mumps occurs worldwide and is endemic in many countries. In the absence of routine vaccination, the annual incidence of mumps is 100–1000 cases per 100,000 population, with epidemic peaks every 2–5 years. From 1999 to 2018, on average, >500,000 mumps cases were reported to the World Health Organization annually, with fewer reported cases from 2019 to 2021 (between 160,000 and 270,000). However, the global mumps incidence is challenging to estimate, as few countries routinely collect the pertinent data. As of 2021, mumps vaccine was introduced in 123 WHO member states. Mumps incidence has been reduced by 97–99% in countries with a routine two-dose measles, mumps, and rubella (MMR) vaccination schedule and by 87–88% in those with a one-dose vaccination program. However, since the mid-2000s, large mumps outbreaks have been reported among populations with high two-dose MMR coverage in countries with routine mumps immunization programs. Most outbreaks have occurred in settings with intense or frequent close contact, such as universities, close-knit communities, and correctional facilities, and most of these cases have occurred in fully vaccinated persons. Despite these outbreaks, mumps incidence is still much higher in countries that do not have routine mumps vaccination.

In the United States, prior to licensure of a vaccine for mumps in 1967, >100,000 mumps cases occurred annually. After the implementation of a one-dose mumps vaccination policy in 1977 and a subsequent two-dose policy in 1989, reported mumps cases declined to an annual average of ~300 by the early 2000s. However, since 2006, there has been an increase in mumps cases reported in the United States, with several peak years **(Fig. 213-1)**. During the highest peak in recent cases, from January 2016 through June 2017, 150 mumps outbreaks and 9200 outbreak-associated cases were reported in a range of settings and groups, including schools, universities, athletic teams and facilities, church groups, workplaces, and large parties and events. While a majority of cases occurred in fully vaccinated young adults in association with large university outbreaks, about one-third of cases have

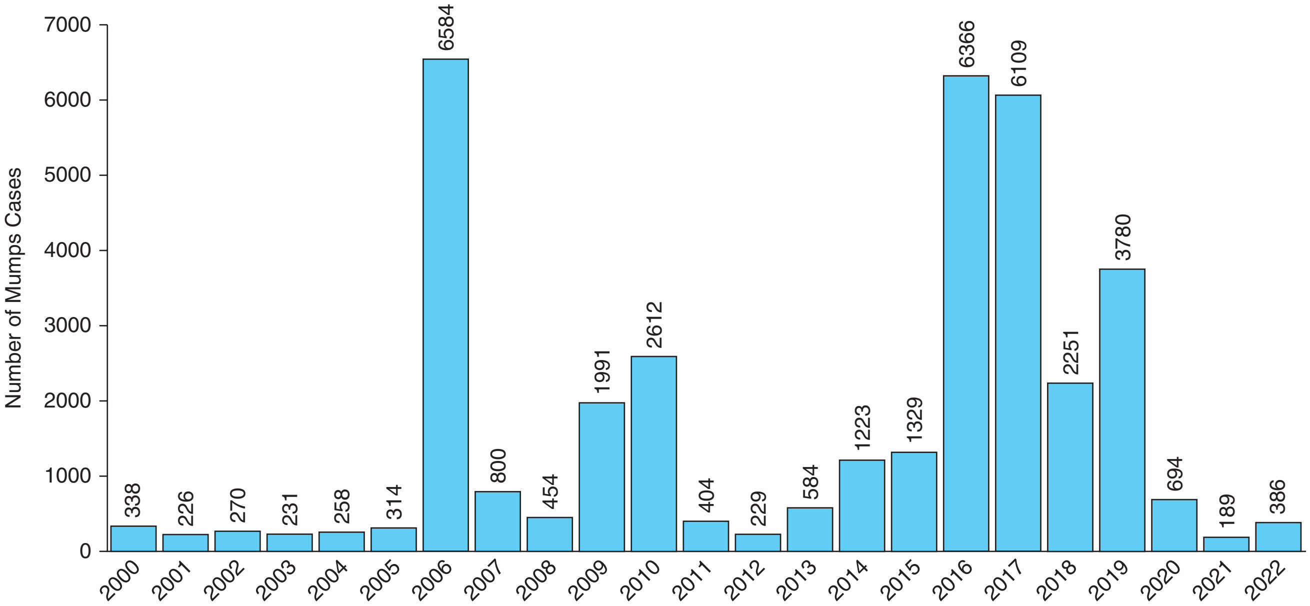

FIGURE 213-1 Reported mumps cases: United States, 2000–2022. *(Source: National Notifiable Diseases Surveillance System (NNDSS), Notifiable Infectious Disease Data Tables. Atlanta, GA, CDC Division of Health Informatics and Surveillance, 2022. Available at https://www.cdc.gov/nndss/data-statistics/infectious-tables/index.html.)*

affected children or adolescents, most of whom were vaccinated. Since 2020, there has been a large reduction in reported mumps cases likely due to social distancing and other COVID-19 prevention measures, with 150–700 reported annually. As of 2023, mumps is endemic in the United States, and there are no elimination goals for the disease.

Multiple factors are likely involved in being at risk for mumps infection among vaccinated persons. Following vaccination, these factors include (1) failure to develop an immune response, (2) the development of a low-level immune response that is insufficient for protection, (3) a decrease in immunity over time (waning immunity) after initial development of a vaccine-induced immune response, (4) lower levels of vaccine-induced antibodies to the circulating wild-type virus strains than to the vaccine virus strain, and (5) a lower frequency of subclinical immunologic boosting due to lack of exposure to wild-type virus during periods of low disease incidence.

PATHOGENESIS

Humans are the only known natural reservoir for mumps virus, which is transmitted through direct contact with respiratory droplets or saliva of an infected person. The average incubation period is 16–18 days, with a range of 12–25 days. A person is most infectious from 2 days before until 5 days after onset of parotitis or other salivary gland swelling. However, mumps virus has been detected in saliva as early as 7 days before onset and as late as 9 days after onset of these manifestations. Mumps virus has been isolated from urine and seminal fluid up to 14 days after onset of parotitis, although no studies have assessed transmissibility of the virus through these fluids.

Primary mumps virus replication likely occurs in the nasal mucosa or upper respiratory mucosal epithelium. Given the range of symptoms, it is assumed that, after infection of the respiratory mucosa, the virus spreads to regional lymph nodes. Mononuclear cells and cells within regional lymph nodes can become infected; such infection facilitates the development of viremia, which usually lasts 3–5 days. Viremia can result in a range of acute inflammatory reactions, most commonly in the salivary glands (resulting in parotitis) and the testes (resulting in orchitis). Other sites of virus dissemination include the kidneys (reflected in the frequency of viruria), the central nervous system (CNS), the pancreas, the heart, the ovaries, the mammary glands, the perilymphatic fluid within the cochlea, and (during pregnancy) the fetus.

Little is known about the pathology of mumps since the disease is rarely fatal. Affected salivary glands contain perivascular and interstitial mononuclear-cell infiltrates and exhibit hemorrhage with prominent edema. Serum and urine amylase levels may be elevated as a result of inflammation and tissue damage in the parotid gland. Necrosis of acinar and epithelial duct cells is evident in the salivary glands and in the germinal epithelium of the seminiferous tubules of the testes. The virus probably enters cerebrospinal fluid (CSF) through the choroid plexus or via transiting mononuclear cells during plasma viremia. Although relevant data are limited, in many cases, mumps encephalitis appears to be a para- or postinfectious process (as suggested by perivenous demyelination and perivascular mononuclear-cell inflammation) rather than the result of a direct cytotoxic effect caused by viral invasion of the CNS. However, although rare, primary mumps encephalitis does occur, as shown by mumps virus isolation from brain tissue. Infection of the perilymphatic fluid likely develops via retrograde penetration by the virus from the cervical lymph nodes following viremia, but infection could also occur via the CSF in cases of mumps CNS infection, given that the perilymph communicates with the CSF. Virus in the perilymph can result in infection of the cochlea and damage to the organ of Corti and the tectorial membrane, leading to transient or permanent deafness. Evidence of placental and intrauterine spread has been found in both early and late gestation. Virus frequently disseminates to the kidneys, but kidney involvement in mumps is almost always benign.

CLINICAL MANIFESTATIONS

While typically presenting with parotitis or other salivary gland swelling, mumps infection can be asymptomatic or present as nonspecific respiratory symptoms, though serious complications such as sensorineural hearing loss can occur. Fully vaccinated persons can contract mumps, but vaccinated persons are at a lower risk for mumps and mumps complications. Mumps infection is asymptomatic in ~20% of unvaccinated patients; the proportion asymptomatic among vaccinated persons is unknown.

Parotitis can be preceded by several days by a prodrome of low-grade fever, malaise, myalgia, headache, and anorexia. Parotitis typically lasts for 5 days (range, 3–7 days); most cases resolve within 10 days. Parotitis is generally bilateral and may not occur synchronously on both sides; unilateral involvement occurs in about one-third of cases. Swelling of the parotid gland is accompanied by tenderness and obliteration of the space between the earlobe and the angle of the mandible **(Figs. 213-2 and 213-3)**. The patient frequently reports an earache and jaw pain and finds it difficult to eat, swallow, or talk. The orifice of the parotid duct is commonly red and swollen. The submaxillary and sublingual glands are involved less often than the parotid gland and are rarely involved alone. In ~6% of mumps cases, obstruction of lymphatic drainage secondary to bilateral salivary gland swelling may lead to presternal

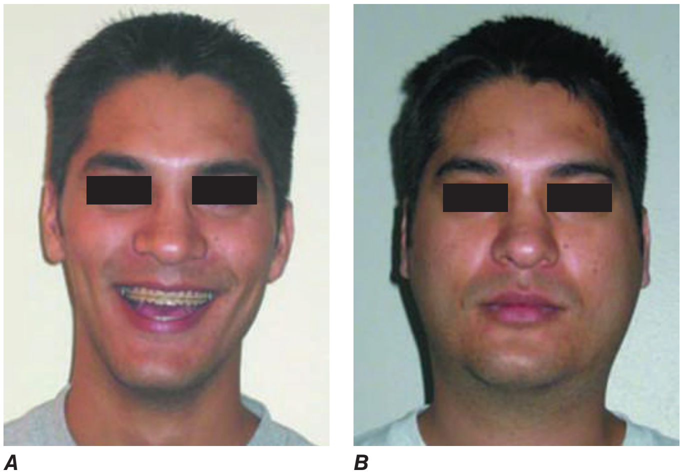

FIGURE 213-2 The same person before mumps acquisition (*A*) and on day 3 of acute bilateral parotitis (*B*). *(Courtesy of patient C.M. From JD Shanley: The resurgence of mumps in young adults and adolescents. Cleve Clin J Med 74:42, 2007. Reprinted with permission. Copyright © 2007 Cleveland Clinic Foundation. All rights reserved.)*

pitting edema, associated often with submandibular adenitis and rarely with the more life-threatening supraglottic edema.

The most frequent complications of mumps include orchitis, oophoritis, mastitis, pancreatitis, hearing loss, meningitis, and encephalitis. Complications can occur in the absence of parotitis and are more common among adults than among children and among males than among females, likely due to rates of orchitis.

Orchitis (testicular inflammation), usually accompanied by fever, is the most common complication, developing in up to 30% of unvaccinated and 6% of vaccinated postpubertal males. This complication is rare in children. Orchitis typically occurs during the first week of parotitis but can develop up to 6 weeks after parotitis. Both testes are involved in ~10–30% of cases. The testis is painful and tender and can be enlarged to several times its normal size. Pain and swelling may last for 1 week, while tenderness may last for several weeks. Testicular atrophy develops in ~30–50% of affected testicles. The development of anti-sperm antibodies, reduced testosterone production, and impaired sperm mobility through oligospermia, azoospermia, or asthenospermia may lead to temporary sterility or subfertility. However, no studies have assessed the risk of permanent infertility in men with mumps orchitis.

Approximately 7% of unvaccinated and ≤1% of vaccinated postpubertal women develop oophoritis, which may be associated with lower abdominal pain and vomiting. The rate of mastitis in mumps has been estimated to be as high as 30% among unvaccinated postpubertal women and as low as ≤1% among vaccinated postpubertal women. Pancreatitis occurs in ~4% of unvaccinated and <1% of vaccinated mumps patients. Mumps pancreatitis, which may present as abdominal pain, is difficult to diagnose because an elevated serum amylase level can be associated with either parotitis or pancreatitis. However, serum lipase is elevated in pancreatitis and the presence of both elevated serum amylase and lipase can help determine if pancreatitis is present in addition to parotitis. Hearing loss associated with mumps infection can occur in up to 4% of unvaccinated and <1% of vaccinated mumps patients. Mumps-related hearing loss is usually sudden in onset, unilateral, and transient and may be associated with vestibular symptoms. Bilateral and permanent hearing loss are rare.

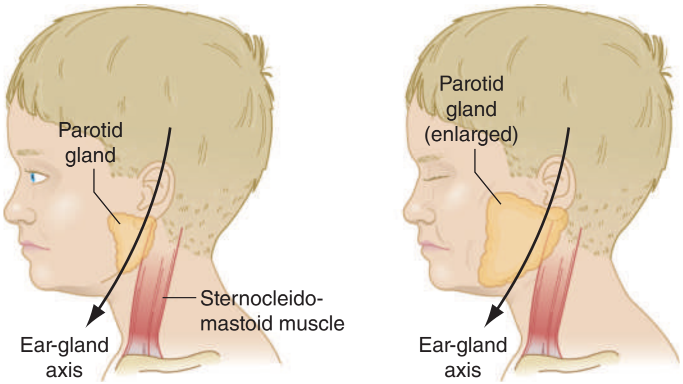

FIGURE 213-3 Schematic drawings of a normal parotid gland (left) and a parotid gland infected with mumps virus (right). An enlarged cervical lymph node is usually posterior to the imaginary line. *(Reproduced with permission from A Gershon et al: Krugman's Infectious Diseases of Children, 11th ed. Philadelphia, Elsevier, 2004.)*

Mumps virus is highly neurotropic, with subclinical CNS involvement occurring in up to 55% of patients as manifested by CSF pleocytosis. However, symptomatic CNS infection is less common. Aseptic meningitis occurs in ≤1% of vaccinated patients and up to 10% of unvaccinated patients and is a self-limited manifestation without significant risk of death or long-term sequelae. Symptoms of aseptic meningitis, including stiff neck, headache, and drowsiness, typically appear ~5 days after parotitis. Encephalitis develops in ≤1% of patients, who present with high fever, marked changes in the level of consciousness, seizures, and focal neurologic symptoms. Electroencephalographic abnormalities may be seen. Permanent sequelae are sometimes identified in survivors, and adult infections more commonly have poor outcomes than pediatric infections. The mortality rate associated with mumps encephalitis is ~1.5%. Other CNS problems occasionally associated with mumps include cerebellar ataxia, facial palsy, transverse myelitis, hydrocephalus, Guillain-Barré syndrome, flaccid paralysis, and behavioral changes.

Although rare and self-limited, myocarditis and endocardial fibroelastosis may represent severe complications of mumps infection; however, mumps-associated electrocardiographic abnormalities have been reported in up to 15% of cases. Other unusual complications include thyroiditis, nephritis, arthritis, hepatic disease, keratouveitis, and thrombocytopenic purpura. Abnormal renal function is common, but severe, life-threatening nephritis is rare.

Mumps infection in pregnant women is generally benign and is not more severe than in women who are not pregnant. Evidence suggesting an association between maternal mumps infection and an increased rate of spontaneous abortion or intrauterine fetal death is inconclusive.

Both mumps reinfection after natural infection and recurrent infection (in which parotid gland swelling resolves and then, weeks to months later, develops on the same or the other side) can occur. In the past, mumps reinfection was thought to be rare, but more recent data have suggested that it may be more common than previously thought.

Death due to mumps is exceedingly rare.

DIFFERENTIAL DIAGNOSIS

Mumps is the only cause of epidemic parotitis, although an increase in parotitis cases may also result from increased influenza activity—specifically, infection with influenza A virus subtype H3N2. Other infectious causes of parotitis include parainfluenza virus types 1–3, Epstein-Barr virus, human herpesviruses 6A and 6B, herpes simplex viruses types 1 and 2, coxsackievirus A, adenovirus, parvovirus B19, echovirus, lymphocytic choriomeningitis virus, and HIV. Laboratory testing for sporadic parotitis cases caused by these infectious pathogens can help rule out mumps.

Parotitis can also develop in the setting of sarcoidosis, Sjögren's syndrome, Mikulicz's syndrome, Parinaud's oculoglandular syndrome, uremia, diabetes mellitus, laundry starch ingestion, malnutrition, cirrhosis, and some drug treatments. Unilateral parotitis can be caused by ductal obstruction, cysts, and tumors. In the absence of parotitis or other salivary gland enlargement, symptoms of other visceral-organ and/or CNS involvement may predominate, and a laboratory diagnosis is required. Other entities should be considered when manifestations consistent with mumps appear in organs other than the parotid. For example, testicular torsion may produce a painful scrotal mass resembling that seen in mumps orchitis. Orchitis can also be caused by bacterial infections in the prostate and urinary tract, sexually transmitted diseases such as chlamydia and gonorrhea, and other viral infections such as those with coxsackievirus, varicella, echovirus, and cytomegalovirus. Oophoritis can also be caused by sexually transmitted diseases such as chlamydia and gonorrhea. A number of viruses (e.g., enteroviruses) can cause aseptic meningitis that is clinically indistinguishable from that due to mumps virus.

LABORATORY DIAGNOSIS

If mumps is suspected, infection is confirmed by virologic methods, but serologic testing can aid in diagnosis. Especially in vaccinated patients, a negative virologic or serologic result in a person with clinical signs of mumps does not rule out mumps infection.

Virologic methods for confirming mumps include reverse transcription polymerase chain reaction (RT-PCR) and viral culture. RT-PCR is preferred because of its sensitivity, specificity, and timeliness. Mumps virus and viral RNA can be detected in blood, saliva, urine, and CSF. Buccal swabs provide the best specimens for virus detection. The parotid gland should be massaged for 30 s prior to collection of the buccal swab sample. As maximal viral shedding occurs within 5 days after symptom onset, specimens for mumps virologic testing ideally should be collected as close to parotitis onset as possible. The diagnostic yield of urine specimens increases over time up to 10 days after parotitis onset, but buccal specimens are more likely than urine specimens to result in virus detection at any time point.

Serologic methods that can aid in the diagnosis of mumps include detection of mumps-specific IgM antibodies or a fourfold rise between acute- and convalescent-phase IgG antibodies. In unvaccinated persons, IgM antibody is usually detectable within 5 days after onset, reaches a maximal level a week after onset, and remains elevated for weeks or months. Failure to detect mumps IgM in vaccinated patients is very common, as the IgM response is often undetectable, transient, or delayed in these individuals. Collection of specimens >3 days after onset may improve IgM detection. Additionally, IgM can yield false-positive results due to serologic cross-reactions. Use of IgG testing is generally not recommended, as IgG titers in vaccinated or previously infected patients may already be elevated at the time of acute-phase specimen collection, such that a fourfold rise is not detected in the convalescent-phase specimen.

TREATMENT

Mumps

Mumps is generally a benign, self-resolving illness. Therapy for parotitis and other clinical manifestations is symptom based and supportive. The administration of analgesics and the application of warm or cold compresses to the parotid area may be helpful. Testicular pain may be minimized by the local application of cold compresses and gentle support for the scrotum. Anesthetic blocks also may be used. Neither the administration of glucocorticoids nor incision of the tunica albuginea is of proven value in severe orchitis. Mumps immune globulin is not recommended for postexposure prophylaxis or treatment.

PREVENTION

Vaccination is the best prevention measure against mumps. Mumps vaccine is commonly included as part of the combination measles–mumps–rubella (MMR) vaccine or the combination measles–mumps–rubella–varicella (MMRV) vaccine. All mumps vaccines currently on the market are live attenuated virus vaccines. Strains used in mumps vaccines have included Jeryl Lynn, RIT 4385, Urabe Am9, Rubini, Leningrad-3 and Leningrad-Zagreb. Urabe- and Rubini-containing vaccines are no longer available. The Jeryl Lynn and RIT 4385 strains are the only strains used in mumps vaccines in the United States since 1967.

In the United States, children are recommended to receive the first MMR dose at 12–15 months of age and the second dose at 4–6 years. MMR vaccine is licensed for use in persons age ≥12 months; MMRV vaccine is licensed for use in persons age 12 months through 12 years. Due to the potential increased risk of febrile seizures, it is recommended that MMR vaccine and varicella vaccine be administered for the first dose in children 12–15 months of age. Adequate vaccination against mumps is defined as two doses of MMR for school-aged children (i.e., grades K–12) and for adults at high risk (i.e., health care workers, international travelers, and students at post–high school educational institutions) and one dose for preschool-aged children and adults not at high risk. During an outbreak, a second dose should be considered for children age 1–4 years and adults who have received one dose. In 2017, after an increase in cases among persons with two MMR doses and a study demonstrating added benefit of a third MMR vaccine dose for individual protection, a third dose was recommended for use during outbreaks, specifically for groups whom public health authorities identify as at increased risk of acquiring mumps; public health authorities will inform providers of these groups at increased risk. As the duration of protection provided by a third dose of MMR vaccine is unknown and may be short term (<1 year), there is no current recommendation for a routine third dose.

The effectiveness of Jeryl Lynn containing MMR vaccine in preventing mumps was 72% after one dose (RR 0.24, 95% CI 0.08 to 0.76; 6 cohort studies; 9915 children; moderate certainty evidence), 86% after two doses (RR 0.12, 95% CI 0.04 to 0.35; 5 cohort studies; 7792 children; moderate certainty evidence). The effectiveness of the mumps component is lower than that of the measles component (two-dose effectiveness of 97%) and the rubella component (one-dose effectiveness of 97%). Incremental vaccine effectiveness of a third MMR dose—compared with two doses—during outbreaks is estimated at 78% (range, 61–88%).

In general, most recipients of mumps vaccine will seroconvert after vaccination and will have detectable antibodies to mumps virus; however, antibody levels start to decline soon after vaccination. Vaccine-induced neutralizing antibodies to wild-type strains may be lower in titer and may decline more rapidly than antibodies to the vaccine strain (Jeryl Lynn). However, most young adults given two vaccine doses in childhood appear to retain memory B cells.

Mumps vaccines are generally very safe. Urabe and Leningrad-Zagreb mumps strain vaccines have been associated with a slightly increased risk of aseptic meningitis, but there is no evidence of this risk for Jeryl Lynn and RIT 4385 mumps strain vaccines. There is a twofold greater risk of febrile seizures among children 12–23 months of age after receipt of the first dose of MMRV vaccine than after the first dose of MMR vaccine, with or without simultaneous varicella vaccination; this risk has not been found among vaccinated children 4–6 years of age.

There is no known immune correlate of protection for mumps; a positive IgG titer indicates only that a person has been exposed to mumps virus through either vaccination or natural infection and does not predict protection against infection. Therefore, all close contacts of a mumps patient should be advised to self-monitor for mumps symptoms for 25 days after their last exposure. Further, IgG titers should not be used to infer immunity in close contacts as it may indicate acute infection rather than immunity. MMR vaccine has not been shown to prevent illness or alter clinical severity in persons already infected with mumps virus and is not recommended as postexposure prophylaxis for immediate close contacts of mumps patients.

Acknowledgment

The authors acknowledge and thank Drs. Mariel Marlow and Stephen Rubin, authors of prior editions of this chapter.

FURTHER READING

Di Pietrantonj C et al: Vaccines for measles, mumps, rubella, and varicella in children. Cochrane Database Syst Rev 4:CD004407, 2020.

Krow-Lucal E et al: Measles, mumps, rubella vaccine (PRIORIX): Recommendations of the Advisory Committee on Immunization Practices—United States, 2022. MMWR Morb Mortal Wkly Rep 71:1465, 2022.

Marin M et al: Recommendation of the Advisory Committee on Immunization Practices for use of a third dose of mumps virus–containing vaccine in persons at increased risk for mumps during an outbreak. MMWR Morb Mortal Wkly Rep 67:33, 2018.

Masarani M et al: Mumps orchitis. J R Soc Med 99:573, 2006.

McClean HQ et al: Prevention of measles, rubella, congenital rubella syndrome, and mumps, 2013: Summary recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 62:1, 2013.

Rasheed MAU et al: Decreased humoral immunity to mumps in young adults immunized with MMR vaccine in childhood. Proc Natl Acad Sci USA 116:19071, 2019.

Rota JS et al: Comparison of the sensitivity of laboratory diagnostic methods from a well-characterized outbreak of mumps in New York City in 2009. Clin Vaccine Immunol 20:391, 2013.

Rubin S et al: Molecular biology, pathogenesis and pathology of mumps virus. J Pathol 235:242, 2015.

Shepersky L et al: Mumps in vaccinated children and adolescents: 2007–2019. Pediatrics 148:e2021051873, 2021.

Tappe J et al: Characteristics of reported mumps cases in the United States: Vaccine 2018-2023. Vaccine 2024. Available at *https://pubmed.ncbi.nlm.nih.gov/39019661/*.

World Health Organization: WHO immunological basis for immunization series. Module 16: Mumps update 2020. Available at *https://apps.who.int/iris/bitstream/handle/10665/338004/9789240017504-eng.pdf*.

214 Rabies and Other Rhabdovirus Infections

Alan C. Jackson

RABIES

Rabies is a rapidly progressive, acute infectious disease of the central nervous system (CNS) in humans and animals that is caused by infection with rabies virus. The infection is normally transmitted from animal vectors via a bite exposure. Rabies has encephalitic and paralytic forms that progress to death.

ETIOLOGIC AGENT

Rabies virus is a member of the family *Rhabdoviridae*. Two genera in this family, *Lyssavirus* and *Vesiculovirus*, contain species that cause human disease. Rabies virus is a lyssavirus that infects a broad range of mammals and causes serious neurologic disease when transmitted to humans. This single-strand RNA virus has a nonsegmented, negative-sense (antisense) genome that consists of 11,932 nucleotides and encodes five proteins: nucleocapsid protein, phosphoprotein, matrix protein, glycoprotein, and a large polymerase protein. Rabies virus variants, which can be characterized by distinctive nucleotide sequences, are associated with specific animal reservoirs. Six other non–rabies virus species in the *Lyssavirus* genus have been reported to cause a clinical picture similar to rabies. Vesicular stomatitis virus, a vesiculovirus, causes vesiculation and ulceration in cattle, horses, and other animals and causes a self-limited, mild, systemic illness in humans (see "Other Rhabdoviruses," below).

EPIDEMIOLOGY

Rabies is a zoonotic infection that occurs in a variety of mammals throughout the world except in Antarctica and on some islands. Rabies virus is usually transmitted to humans by the bite of an infected animal. Worldwide, endemic canine rabies is estimated to cause 59,000 human deaths annually. Most of these deaths occur in Asia and Africa, with rural populations and children disproportionally affected. Thus, in many resource-poor and resource-limited countries, canine rabies continues to be a threat to humans. However, in Latin America, rabies control efforts in dogs have been quite successful in recent years. Endemic canine rabies has been eliminated from the United States and most other resource-rich countries. Rabies is endemic in wildlife species, and a variety of animal reservoirs have been identified in different countries of the world **(Fig. 214-1)**. Surveillance data from 2021 identified 3579 confirmed animal cases of rabies in the United States and Puerto Rico. Only 9.6% of these cases were in domestic animals, including 222 cases in cats, 50 in dogs, and 42 in cattle. In North American wildlife reservoirs, including bats, raccoons, skunks, and foxes, the infection is endemic, with involvement of one or more rabies virus variants in each reservoir species **(Fig. 214-2)**. "Spillover" of rabies to other wildlife species and to domestic animals occurs. Bat rabies virus variants are present in every state except Hawaii and are responsible for most indigenously acquired human rabies cases in the United States. Raccoon rabies is endemic along the entire eastern coast of the United States. Skunk rabies is present in the midwestern states, with another focus in California. Rabies in foxes occurs in New Mexico, Arizona, and Alaska. In the United States, there were five human rabies deaths in 2021 and none in 2019, 2020, and 2022.

In Canada and Europe, epizootics of rabies in red foxes have been well controlled with the use of baits containing rabies vaccine. A similar approach, along with additional measures, is used in Canada to control incursions of raccoon rabies from the United States.

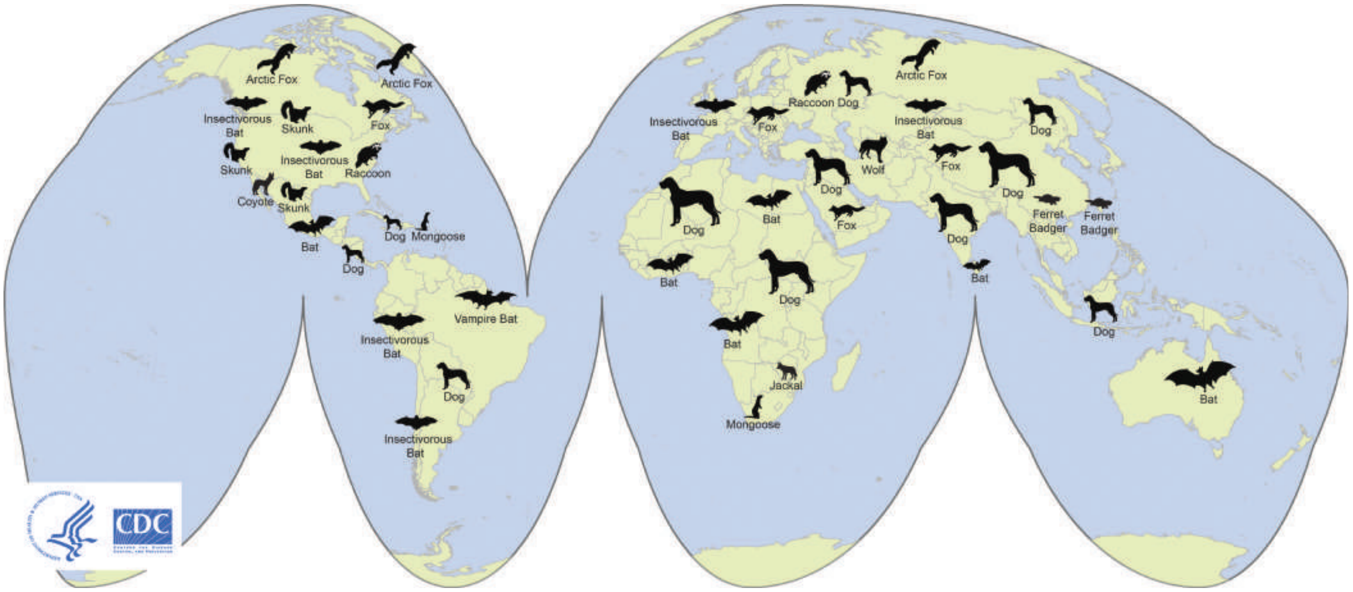

FIGURE 214-1 Distribution of global rabies vectors. *(Courtesy of the Centers for Disease Control and Prevention.)*

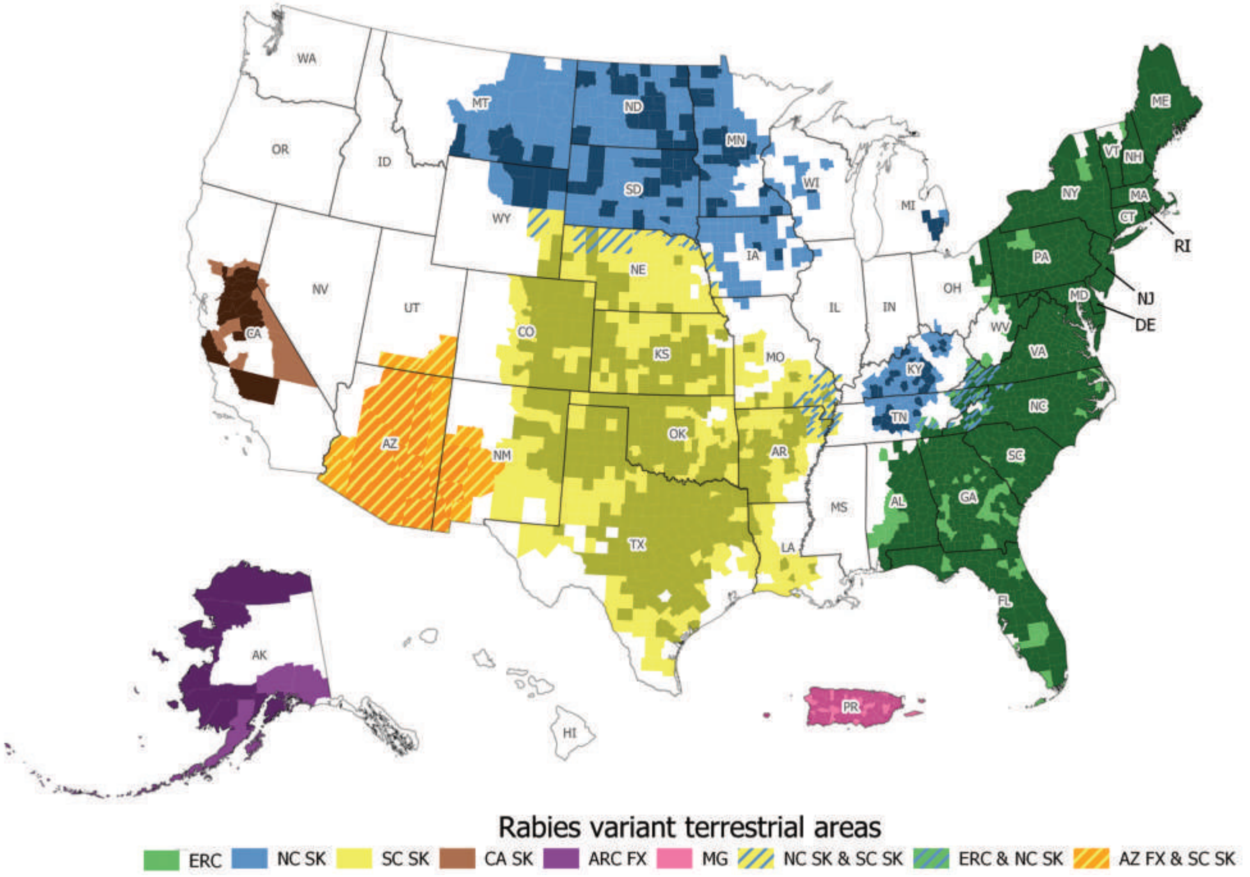

FIGURE 214-2 Distribution of the major rabies virus variants among wild terrestrial reservoirs in the United States and Puerto Rico, 2017–2021. *Darker shading* indicates counties with confirmed animal rabies cases in the past 5 years; *lighter shading* represents counties bordering enzootic counties without animal rabies cases that did not satisfy criteria for adequate surveillance. Small nonenzootic areas with no rabies cases reported in the past 15 years are shaded if they are in the vicinity of known enzootic counties and do not satisfy criteria for adequate surveillance. ARC FX rabies virus variant, Arctic fox RVV; AZ FX, Arizona fox RVV; CA SK, California skunk RVV; ERC, Eastern raccoon RVV; MG, Dog-mongoose RVV; NC SK, North central skunk RVV; SC SK, South central skunk RVV. *(Reproduced with permission from X Ma et al: Rabies surveillance in the United States during 2021. J Am Vet Med Assoc 261:1045, 2023.)*

Rabies virus variants isolated from humans or other mammalian species can be identified by reverse-transcription polymerase chain reaction (RT-PCR) amplification and sequencing or by characterization with monoclonal antibodies. These techniques are helpful in human cases with no known history of exposure. Worldwide, most human rabies (~99%) is transmitted from dogs in countries with endemic canine rabies and dog-to-dog transmission, and human cases can be imported by travelers returning from these regions. In North America, indigenously acquired human disease is usually associated with transmission from bats; there may be no known history of bat bite or other bat exposure in these cases. Most human cases are due to a bat rabies virus variant associated with silver-haired and tricolored bats and, less commonly, from Brazilian (Mexican) free-tail bats. These are small bats whose bite may not be recognized, and the virus has adapted for replication at skin temperature and in cell types that are present in the skin.

Transmission from nonbite exposures is relatively uncommon. Aerosols generated in the laboratory or in caves containing millions of Brazilian free-tail bats have rarely caused human rabies. Transmission has resulted from corneal transplantation and also from solid-organ transplantation and a vascular conduit (for a liver transplant) from undiagnosed donors with rabies in Texas, Florida, Germany, Kuwait, Syria, and China. Human-to-human transmission is extremely rare, although hypothetical concern about transmission to health care workers has prompted the implementation of barrier techniques to prevent exposures from patients with rabies.

PATHOGENESIS

The incubation period of rabies (defined as the interval between exposure and the onset of clinical disease) is usually 20–90 days, but in rare cases is either as short as a few days or >1 year. During most of the incubation period, rabies virus is thought to be present at or close to the site of inoculation **(Fig. 214-3)**. In muscles, the virus is known to bind to nicotinic acetylcholine receptors on postsynaptic membranes at neuromuscular junctions, but the exact details of viral entry into the skin and subcutaneous tissues have not yet been clarified. Rabies virus spreads centripetally along peripheral nerves toward the spinal cord or brainstem via retrograde fast axonal transport (rate, up to ~250 mm/d), with delays at intervals of ~12 h at each synapse. Once the virus enters the CNS, it rapidly disseminates to other regions of the CNS via fast axonal transport along neuroanatomic connections. Neurons are prominently infected in rabies; infection of astrocytes is unusual. After CNS infection becomes established, there is centrifugal spread along sensory and autonomic nerves to other tissues, including the salivary glands, heart, adrenal glands, and skin. Rabies virus replicates in acinar cells of the salivary glands and is secreted in the saliva of rabid animals that serve as vectors of the disease. There is no well-documented evidence for hematogenous spread of rabies virus.

Pathologic studies show mild inflammatory changes in the CNS in rabies, with mononuclear inflammatory infiltration in the leptomeninges, perivascular regions, and parenchyma, including microglial nodules called *Babes nodules*. Degenerative neuronal changes usually are not prominent, and there is little evidence of neuronal death; neuronophagia is observed occasionally. The pathologic changes are surprisingly mild in light of the clinical severity and fatal outcome of the disease. The most characteristic pathologic finding in rabies is the *Negri body* **(Fig. 214-4)**. Negri bodies are eosinophilic cytoplasmic inclusions in brain neurons that are composed of rabies virus proteins and viral RNA. These inclusions occur in a minority of infected neurons, are commonly observed in Purkinje cells of the cerebellum

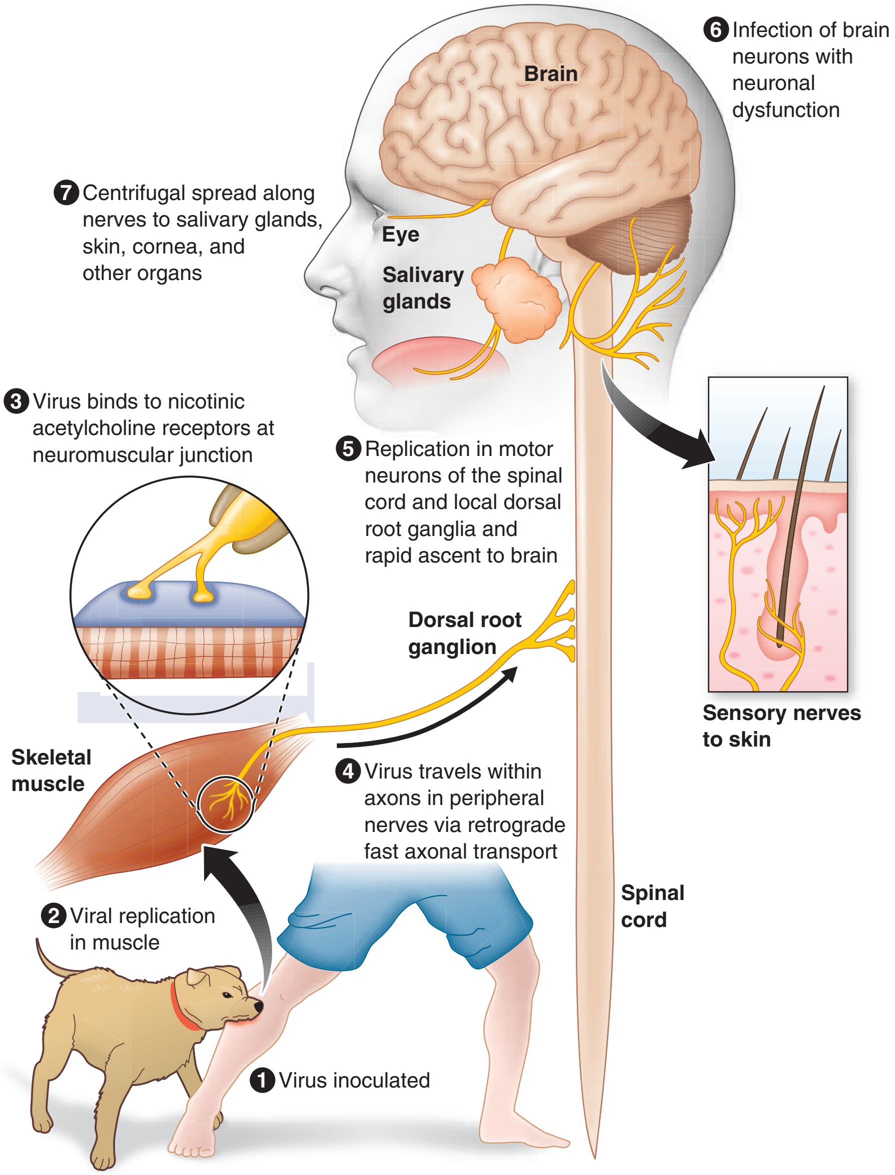
⑥ Infection of brain neurons with neuronal dysfunction
Brain
⑦ Centrifugal spread along nerves to salivary glands, skin, cornea, and other organs
Eye
Salivary glands
③ Virus binds to nicotinic acetylcholine receptors at neuromuscular junction
⑤ Replication in motor neurons of the spinal cord and local dorsal root ganglia and rapid ascent to brain
Dorsal root ganglion
Sensory nerves to skin
Skeletal muscle
④ Virus travels within axons in peripheral nerves via retrograde fast axonal transport
Spinal cord
② Viral replication in muscle
① Virus inoculated

FIGURE 214-3 Schematic representation of events in rabies pathogenesis following peripheral inoculation of rabies virus by an animal bite. *(Reproduced with permission from AC Jackson: Pathogenesis, in Rabies: Scientific Basis of the Disease and Its Management, 3rd ed. Oxford, UK, Elsevier Academic Press, 2013.)*

and in pyramidal neurons of the hippocampus, and are less frequently seen in cortical and brainstem neurons. Negri bodies are not observed in all cases of rabies. The lack of prominent degenerative neuronal changes has led to the concept that neuronal dysfunction—rather than neuronal death—is responsible for clinical disease in rabies. Experimental studies indicate that oxidative stress due to mitochondrial dysfunction plays an important role. The basis for behavioral changes, including the aggressive behavior of rabid animals, is not well understood but may be related to infection of serotonergic neurons in the brainstem.

■ CLINICAL MANIFESTATIONS

For rabies prevention, the emphasis must be on postexposure prophylaxis (PEP) initiated after a recognized exposure and before any symptoms or signs develop. Rabies should usually be suspected on the basis of the clinical presentation with or without a history of an exposure. The disease generally presents as atypical encephalitis with relative preservation of consciousness. Rabies may be difficult to recognize late in the clinical course when progression to coma has occurred. A minority of patients (~20%) present with acute flaccid paralysis. There are prodromal, acute neurologic, and comatose phases that usually progress to death despite aggressive therapy **(Table 214-1)**.

Prodromal Features The clinical features of rabies begin with nonspecific prodromal manifestations, including fever, malaise, headache, nausea, and vomiting. Anxiety or agitation also may occur. The earliest specific neurologic symptoms of rabies include paresthesias, pain, or pruritus near the site of the exposure, one or more of which occur in 50–80% of patients and strongly suggest rabies. The wound has usually healed by this point, and these symptoms probably reflect infection with associated inflammatory changes in local dorsal root or cranial sensory ganglia.

Encephalitic Rabies Two acute neurologic forms of rabies are seen in humans: the encephalitic (furious) form in 80% and the paralytic form in 20%. Some of the manifestations of encephalitic rabies, including fever, confusion, hallucinations, combativeness, and seizures, may be seen in other viral encephalitides as well. Autonomic dysfunction is common in rabies and may result in hypersalivation, gooseflesh, cardiac arrhythmia, and priapism. In encephalitic rabies, episodes of hyperexcitability are typically followed by periods of complete lucidity that become shorter as the disease progresses. Rabies encephalitis is distinguished by early brainstem involvement, which results in the classic features of *hydrophobia* (involuntary, painful contraction of the diaphragm and accessory respiratory, laryngeal, and pharyngeal muscles in response to swallowing liquids) **(Fig. 214-5)** and *aerophobia* (the same features caused by stimulation from a draft of air). These symptoms are probably due to dysfunction of infected brainstem neurons that normally inhibit inspiratory neurons near the nucleus ambiguus, resulting in exaggerated defense reflexes that protect the respiratory tract. The combination of hypersalivation and pharyngeal dysfunction is responsible for the classic appearance of "foaming at the mouth." Brainstem dysfunction progresses rapidly, and coma—followed within days by death—is the rule unless the course is prolonged by supportive measures. With such measures, late complications can include cardiac and/or respiratory failure, disturbances of water balance (syndrome of inappropriate antidiuretic hormone secretion or diabetes insipidus), noncardiogenic pulmonary edema, and gastrointestinal hemorrhage. Cardiac arrhythmias may be due to dysfunction affecting vital centers in the brainstem or autonomic pathways or to myocarditis. Multiple-organ failure is common in patients treated aggressively in critical care units.

Paralytic Rabies About 20% of patients have paralytic rabies in which muscle weakness predominates and cardinal features of encephalitic rabies (hyperexcitability, hydrophobia, and aerophobia) are lacking. There is early and prominent flaccid muscle weakness, often beginning in the bitten extremity and spreading to produce quadriparesis and facial weakness. Sphincter involvement is common, sensory involvement is usually mild, and these cases are commonly misdiagnosed as Guillain-Barré syndrome. Patients with paralytic rabies generally survive a few days longer than those with encephalitic rabies, but multiple-organ failure nevertheless ensues.

■ LABORATORY INVESTIGATIONS

Most routine laboratory tests in rabies yield normal results or show nonspecific abnormalities. Complete blood counts are usually normal. Examination of cerebrospinal fluid (CSF) often reveals mild

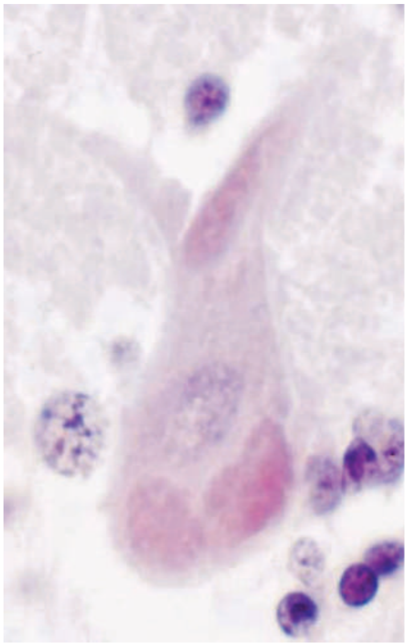

FIGURE 214-4 Three large Negri bodies in the cytoplasm of a cerebellar Purkinje cell from an 8-year-old boy who died of rabies after being bitten by a rabid dog in Mexico. *(Reproduced with permission from AC Jackson, E Lopez-Corella. N Engl J Med 335:568, 1996; © Massachusetts Medical Society.)*

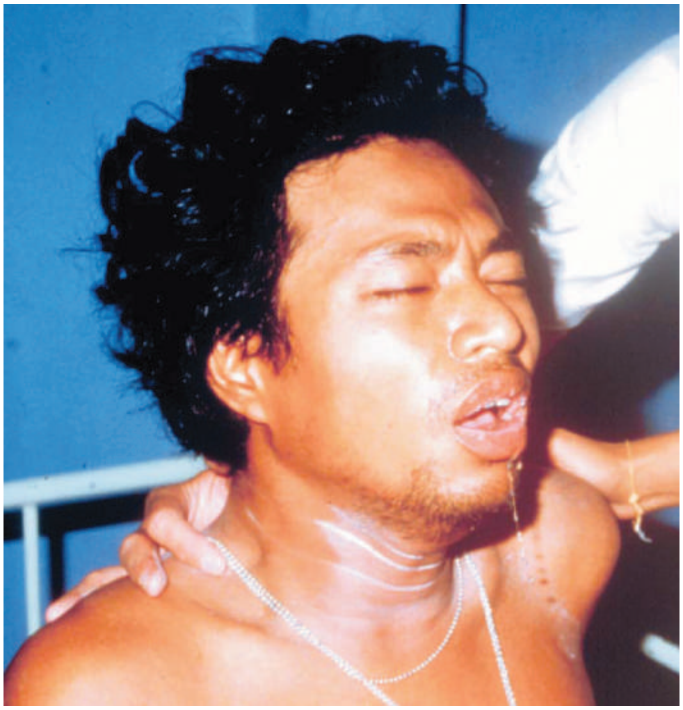

FIGURE 214-5 Hydrophobic spasm of inspiratory muscles associated with terror in a patient with encephalitic (furious) rabies who is attempting to swallow water. *(Copyright DA Warrell, Oxford, UK; with permission.)*

mononuclear-cell pleocytosis with a mildly elevated protein level. Severe pleocytosis (>1000 white cells/μL) is unusual and should prompt a search for an alternative diagnosis. Imaging is usually performed to exclude other diagnostic possibilities. Computed tomography (CT) head scans are usually normal in rabies. Magnetic resonance imaging (MRI) brain scans may show signal abnormalities in the brainstem or other gray-matter areas, but these findings are variable and nonspecific. Electroencephalograms typically show only nonspecific abnormalities. Of course, important tests in suspected cases of rabies include those that may identify an alternative, potentially treatable diagnosis (see "Differential Diagnosis," below).

DIAGNOSIS

In North America, a diagnosis of rabies often is not considered until relatively late in the clinical course, even with a typical clinical presentation. This diagnosis should be considered in patients presenting with acute atypical encephalitis or acute flaccid paralysis, including those in whom Guillain-Barré syndrome is suspected. The absence of an animal-bite history is common in North America, particularly due to unrecognized bat exposures. The lack of hydrophobia is not unusual in rabies. Once rabies is suspected, rabies-specific laboratory tests should be performed to confirm the diagnosis. Diagnostically useful specimens include serum, CSF, fresh saliva, skin biopsy samples from the neck, and brain tissue (rarely obtained before death). Because skin biopsy relies on the demonstration of rabies virus antigen and/or RNA in cutaneous nerves at the base of hair follicles, samples are usually taken from hairy skin at the nape of the neck. Corneal impression smears are of low diagnostic yield and are generally not performed. Negative antemortem rabies-specific laboratory tests never exclude a diagnosis of rabies, and tests may need to be repeated after an interval for diagnostic confirmation.

TABLE 214-1 Clinical Stages of Rabies

STAGE	TYPICAL DURATION	SYMPTOMS AND SIGNS
Incubation period	20–90 days	None
Prodrome	2–10 days	Fever, malaise, anorexia, nausea, vomiting; paresthesias, pain, or pruritus at the wound site
Acute Neurologic Disease		
Encephalitic (80%)	2–7 days	Anxiety, agitation, hyperactivity, bizarre behavior, hallucinations, autonomic dysfunction, hydrophobia
Paralytic (20%)	2–10 days	Flaccid paralysis in limb(s) progressing to quadriparesis with facial paralysis
Coma, death[a]	0–14 days	

[a] Recovery is rare.

Source: Adapted from MAW Hattwick: Rabies virus, in *Principles and Practice of Infectious Diseases,* GL Mandell et al (eds). New York, Wiley, 1979.

Rabies Virus–Specific Antibodies In a previously unimmunized patient, serum neutralizing antibodies to rabies virus are diagnostic. However, because rabies virus infects immunologically privileged neuronal tissues, serum antibodies may not develop until late in the disease. Antibodies may be detected within a few days after the onset of symptoms, but some patients die without detectable antibodies. The presence of rabies virus–specific neutralizing antibodies in the CSF suggests rabies encephalitis, regardless of immunization status. A diagnosis of rabies is questionable in patients who recover from their illness without developing serum neutralizing antibodies to rabies virus.

RT-PCR Amplification Detection of rabies virus RNA by RT-PCR is highly sensitive and specific. This technique can detect virus in fresh saliva samples, skin biopsy specimens, CSF (less sensitive), and brain tissues. In addition, RT-PCR with genetic sequencing can distinguish among rabies virus variants, permitting identification of the probable source of an infection.

Direct Fluorescent Antibody Testing Direct fluorescent antibody (DFA) testing with rabies virus antibodies conjugated to fluorescent dyes is highly sensitive and specific for the detection of rabies virus antigen in tissues; the test can be performed quickly and applied to skin biopsy and brain tissue samples. In skin biopsy samples, rabies virus antigen may be detected in cutaneous nerves at the base of hair follicles.

DIFFERENTIAL DIAGNOSIS

The diagnosis of rabies may be difficult without a history of animal exposure, and no exposure to an animal (e.g., a bat) may be recalled. The presentation of rabies is usually quite different from that of acute viral encephalitis due to most other causes, including herpes simplex encephalitis and arboviral (e.g., West Nile) encephalitis. Early neurologic symptoms may occur at the site of the bite, and there may be early features of brainstem involvement with preservation of consciousness. Anti–*N*-methyl-D-aspartate receptor (anti-NMDA) encephalitis occurs in young patients (especially females) and is characterized by behavioral changes, autonomic instability, hypoventilation, and seizures. Many other antibodies are also associated with autoimmune encephalitis. Postinfectious (immune-mediated) encephalomyelitis may follow influenza, measles, mumps, and other infections; it may also occur as a sequela of immunization with rabies vaccines derived from neural tissues, which are now infrequently used and only in resource-limited and resource-poor countries. Rabies may present with unusual neuropsychiatric symptoms and may be misdiagnosed as a psychiatric disorder. Rabies hysteria (now classified as a somatic symptom disorder) may occur as a psychological response to the fear of rabies and is often characterized by a shorter incubation period than rabies, aggressive behavior, inability to communicate, and a long course with recovery.

As previously mentioned, paralytic rabies may mimic Guillain-Barré syndrome. In these cases, fever, bladder dysfunction, a normal sensory examination, and CSF pleocytosis favor a diagnosis of rabies. Conversely, Guillain-Barré syndrome may occur as a complication of rabies vaccination with a neural tissue–derived product (e.g., suckling mouse brain vaccine) and may be mistaken for paralytic rabies (i.e., vaccine failure).

TREATMENT

Rabies

There is no established treatment for rabies. Aggressive management with supportive care in critical care units has resulted in the survival of at least 33 patients with rabies. Many of these survivors have recently been reported from India. There have been many recent treatment failures (>60) with the combination of therapeutic (induced) coma, antiviral drugs, and sometimes ketamine—measures that were used in a healthy survivor in whom neutralizing antibodies to rabies virus were detected at presentation. No antiviral therapy has demonstrated any efficacy in the treatment of rabies. Expert opinion is recommended before a course of experimental therapy is embarked upon. A palliative approach may be appropriate for many patients who are not considered candidates for aggressive management.

PROGNOSIS

Rabies is an almost uniformly fatal disease but is nearly always preventable after recognized exposures with appropriate postexposure therapy during the early incubation period (see below). All but 2 of 33 documented survivors of rabies received one or more doses of rabies vaccine before disease onset. The survivors who had not received vaccine had neutralizing antibodies to rabies virus in CSF at clinical presentation. Most patients with rabies die within several days of the onset of illness, despite aggressive care in a critical care unit.

PREVENTION

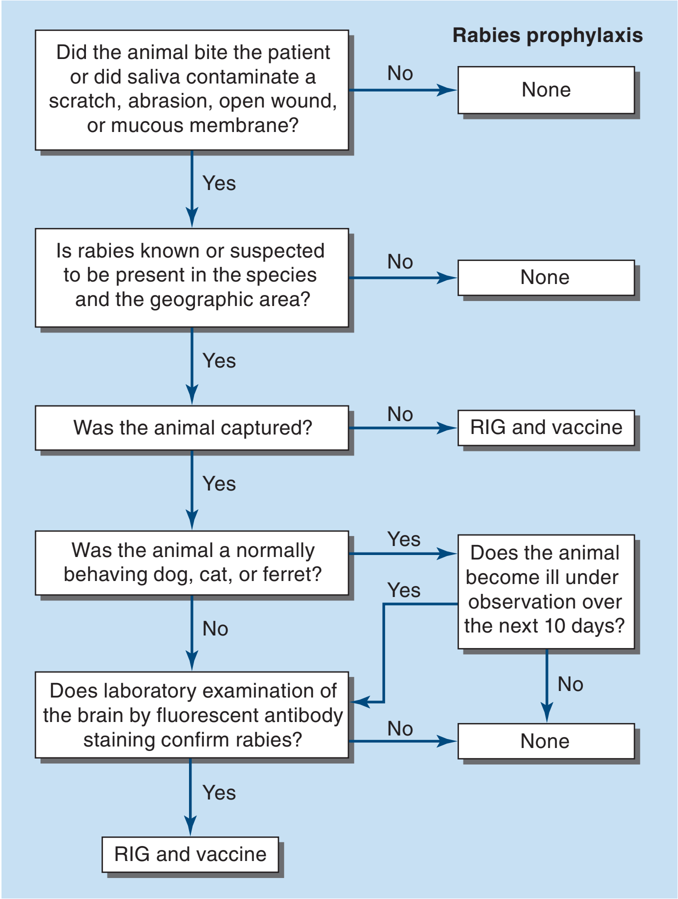
Rabies prophylaxis
Did the animal bite the patient or did saliva contaminate a scratch, abrasion, open wound, or mucous membrane?
No
None
Yes
Is rabies known or suspected to be present in the species and the geographic area?
No
None
Yes
Was the animal captured?
No
RIG and vaccine
Yes
Was the animal a normally behaving dog, cat, or ferret?
Yes
Does the animal become ill under observation over the next 10 days?
Yes
No
No
Does laboratory examination of the brain by fluorescent antibody staining confirm rabies?
No
None
Yes
RIG and vaccine

FIGURE 214-6 Algorithm for rabies postexposure prophylaxis. RIG, rabies immune globulin. *(From L Corey, in Harrison's Principles of Internal Medicine, 15th ed. E Braunwald et al [eds]: New York, McGraw-Hill, 2001.)*

Postexposure Prophylaxis Since there is no effective therapy for rabies, it is extremely important to prevent the disease after an animal exposure. Figure 214-6 shows the steps involved in making decisions about PEP. Based on the exposure history and local epidemiologic information, the physician must decide whether initiation of PEP is warranted. Healthy dogs, cats, or ferrets may be confined and observed for 10 days. PEP is not necessary if the animal remains healthy. If the animal develops signs of rabies during the observation period, it should be euthanized immediately; the head should be transported to the laboratory under refrigeration, rabies virus should be sought by DFA testing, and viral isolation should be attempted by cell culture and/or mouse inoculation. Any animal other than a dog, cat, or ferret should be euthanized immediately and the head submitted for laboratory examination. In high-risk exposures and in areas where canine rabies is endemic, rabies prophylaxis should be initiated without waiting for laboratory results. If the laboratory results prove to be negative, it may safely be concluded that the animal's saliva did not contain rabies virus, and immunization should be discontinued. If an animal escapes after an exposure, it must be considered rabid, and PEP must be initiated unless information from public health officials indicates otherwise (i.e., there is no endemic rabies in the area). Although controversial, the use of PEP may be warranted when a person (e.g., a small child or a sleeping adult) has been present in the same space as a bat and an unrecognized bite cannot be reliably excluded.

PEP includes local wound care and both active and passive immunization. It is important that current recommendations are followed very closely because minor deviations can lead to failure of prophylactic measures. Local wound care is essential and may greatly decrease the risk of rabies virus infection. Wound care should not be delayed, even if the initiation of immunization is postponed pending the results of the 10-day observation period. All bite wounds and scratches should be washed thoroughly with soap and water. Devitalized tissues should be debrided, tetanus prophylaxis given, and antibiotic treatment initiated whenever indicated.

Previously unvaccinated persons (but not those who have previously been immunized) should be passively immunized with rabies immune globulin (RIG). If RIG is not immediately available, it should be administered no later than 7 days after the first vaccine dose. After day 7, endogenous antibodies are being produced, and passive immunization may actually be counterproductive. If anatomically feasible, the entire dose of RIG (20 IU/kg) should be infiltrated at the site of the bite, and any RIG remaining after infiltration of the bite site should be administered IM at a distant site. Recent recommendations by the World Health Organization indicate that under certain circumstances the remainder of the dose does not need to be administered after local infiltration of the wound(s). With multiple or large wounds, the RIG preparation may need to be diluted in order to obtain a sufficient volume for adequate infiltration of all wound sites. If the exposure involves a mucous membrane, the entire dose should be administered IM. Rabies vaccine and RIG should never be administered at the same site or with the same syringe. Commercially available RIG in the United States is purified from the serum of hyperimmunized human donors. These human RIG preparations are much better tolerated than are the equine-derived preparations still in use in some countries (see below). Serious adverse effects of human RIG are uncommon. Local pain and low-grade fever may occur.

Two purified inactivated rabies vaccines are available for rabies PEP in the United States. They are highly immunogenic and remarkably safe compared with earlier vaccines. Four 1-mL doses of rabies vaccine should be given IM in the deltoid area. (The anterolateral aspect of the thigh also is acceptable in children.) Gluteal injections, which may not always reach muscle, should not be given and have been associated with rare vaccine failures. Ideally, the first dose should be given as soon as possible after exposure; failing that, it should be given without further delay. The three additional doses should be given on days 3, 7, and 14; a fifth dose on day 28 is no longer recommended except in immunocompromised patients. Pregnancy is not a contraindication for immunization. Glucocorticoids and other immunosuppressive medications may interfere with the development of active immunity and should not be administered during PEP unless they are essential. Routine measurement of serum neutralizing antibody titers is not required, but titers should be measured 2–4 weeks after immunization in immunocompromised persons. Local reactions (pain, erythema, edema, and pruritus) and mild systemic reactions (fever, myalgias, headache, and nausea) are common; anti-inflammatory and antipyretic medications may be used, but immunization should not be discontinued. Systemic allergic reactions are uncommon, but anaphylaxis does occur rarely and can be treated with epinephrine and antihistamines. The risk of rabies development should be carefully considered before the decision is made to discontinue vaccination because of an adverse reaction.

Most of the burden of rabies PEP is borne by persons with the fewest resources. In addition to the rabies vaccines discussed above, vaccines grown in either primary cell lines (hamster or dog kidney) or continuous cell lines (Vero cells) are satisfactory and are available in many countries outside the United States. Less expensive vaccines derived from neural tissues are still used in a diminishing number of developing countries; however, these vaccines are associated with serious neuroparalytic complications, including postinfectious encephalomyelitis and Guillain-Barré syndrome. The use of these vaccines should be discontinued as soon as possible, and progress has been made in this regard. Worldwide, more than 10 million individuals receive postexposure rabies vaccine each year.

If human RIG is unavailable, purified equine RIG can be used in the same manner at a dose of 40 IU/kg. The incidence of anaphylactic reactions and serum sickness has been low with recent equine RIG products.

Preexposure Rabies Vaccination Preexposure rabies prophylaxis should be considered for people with an occupational or recreational risk of rabies exposures and also for certain travelers to rabies-endemic areas. The primary schedule consists of two doses of rabies vaccine given on days 0 and 7. Serum neutralizing antibody tests help determine the need for subsequent booster doses. When a previously immunized individual is exposed to rabies, two booster doses of vaccine should be administered on days 0 and 3. Wound care remains essential. As stated above, RIG should not be administered to previously vaccinated persons.

OTHER RHABDOVIRUSES

OTHER LYSSAVIRUSES

A growing number of lyssaviruses other than rabies virus have been discovered to infect bat populations in Europe, Africa, Asia, and Australia. Six of these viruses have produced a very small number of cases of a human disease indistinguishable from rabies, including European bat lyssaviruses 1 and 2, Australian bat lyssavirus, Irkut virus, and Duvenhage virus. Mokola virus, a lyssavirus that has been isolated from shrews with an unknown reservoir species in Africa, may also produce human disease indistinguishable from rabies.

VESICULAR STOMATITIS VIRUS

Vesicular stomatitis is a viral disease of cattle, horses, pigs, and some wild mammals. Vesicular stomatitis virus is a member of the genus *Vesiculovirus* in the family *Rhabdoviridae*. Outbreaks of vesicular stomatitis in horses and cattle occur sporadically in the southwestern United States. The animal infection is associated with severe vesiculation and ulceration of oral tissues, teats, and feet and may be clinically indistinguishable from the more dangerous foot-and-mouth disease. Epidemics are usually seasonal, typically beginning in the late spring, and are probably due to arthropod vectors. Direct animal-to-animal spread can also occur, although the virus cannot penetrate intact skin. Transmission to humans usually results from direct contact with infected animals (particularly cattle) and occasionally follows laboratory exposure. In human disease, early conjunctivitis is followed by an acute influenza-like illness with fever, chills, nausea, vomiting, headache, retrobulbar pain, myalgias, substernal pain, malaise, pharyngitis, and lymphadenitis. Small vesicular lesions may be present on the buccal mucosa or on the fingers. Encephalitis is very rare. The illness usually lasts 3–6 days, with complete recovery. Subclinical infections are common. A serologic diagnosis can be made on the basis of a rise in titer of complement-fixing or neutralizing antibodies. Therapy is symptom-based.

FURTHER READING

Fooks AR et al: Current status of rabies and prospects for elimination. Lancet 384:1389, 2014.

Fooks AR, Jackson AC (eds): *Rabies: Scientific Basis of the Disease and Its Management*, 4th ed. London, Elsevier Academic Press, 2020.

Jackson AC: Treatment of rabies. In: Post TW, ed. UpToDate. Waltham, Massachusetts: Wolters Kluwer, 2023. *www.uptodate.com*.

Letchworth GJ et al: Vesicular stomatitis. Vet J 157:239, 1999.

Manning SE et al: Human rabies prevention—United States, 2008: Recommendations of the Advisory Committee on Immunization Practices. MMWR Recomm Rep 57:1, 2008.

World Health Organization: *WHO Expert Consultation on Rabies: Third Report* (WHO Technical Report Series No. 1012). Geneva, World Health Organization, 2018. Available at *https://iris.who.int/bitstream/handle/10665/272364/9789241210218-eng.pdf*. Accessed September 20, 2024.

215

Arthropod-Borne and Rodent-Borne Virus Infections

Jens H. Kuhn, Ian Crozier

This chapter summarizes the major features of selected arthropod-borne and rodent-borne viruses and associated infections and/or disease. Numerous viruses of this category are spread in nature among animals without ever infecting humans. Other viruses incidentally infect humans, with few causing disease. Some viruses are regularly introduced into the human population by arthropods (specifically, insects and ticks) or by chronically infected rodents.

These zoonotic viruses are taxonomically diverse and therefore differ fundamentally in virion morphology, replication strategy, genomic organization, and genome sequence. Although a virus's classification in a taxon is enlightening regarding natural maintenance strategies, sensitivity to antiviral agents, and aspects of pathogenesis, it does not necessarily predict which clinical symptoms and signs (if any) the viral infection will cause in humans. Zoonotic viruses are evolving, and "new" zoonotic viruses are regularly discovered. The epizootiology and epidemiology of zoonotic viruses continue to change because of environmental alterations affecting vectors, reservoirs, and hosts, including wildlife, livestock, and humans. Zoonotic viruses are most numerous in the tropics but are also found in temperate and frigid climates. The distribution and seasonal activity of zoonotic viruses may vary in a manner likely to depend largely on ecologic conditions (e.g., rainfall and temperature), which can affect the density of virus vectors and reservoirs, likelihood of exposure, and development of infection.

Arthropod-borne viruses (arboviruses) infect their vectors after ingestion of blood meals from viremic, usually nonhuman vertebrates; some infections may be the result of saliva-activated transmission. Arthropod vectors develop chronic systemic infection as the viruses penetrate their guts and spread throughout their bodies to their salivary glands; such virus dissemination, referred to as *extrinsic incubation*, typically lasts 1–3 weeks in mosquitoes. If the salivary glands become involved, the arthropod vectors are competent to continue the chain of transmission by infecting vertebrates during subsequent blood meals. Alternatively, virus maintenance in arthropod vectors may be achieved through *transovarial transmission* to progeny. Generally, the arthropods are unharmed by the infections and natural vertebrate partners have only transient viremia without overt disease.

Rodent-borne viruses are maintained in nature by transmission among rodents, which become chronically infected. A high degree of rodent–virus specificity may be observed, and overt disease in the reservoir hosts is rare.

ETIOLOGY

Arthropod-borne and rodent-borne zoonotic viruses belong mostly to the classes *Alsuviricetes* (family *Togaviridae*), *Bunyaviricetes* (families *Arenaviridae*, *Hantaviridae, Nairoviridae, Peribunyaviridae,* and *Phenuiviridae*), *Flasuviricetes* (family *Flaviviridae*), *Insthoviricetes* (family *Orthomyxoviridae*), *Monjiviricetes* (families *Bornaviridae* and *Rhabdoviridae*), and *Resentoviricetes* (families *Sedoreoviridae* and *Spinareoviridae*) **(Table 215-1)**. An exception is Syr-Darya Valley fever virus, an ixodid tick-borne cardiovirus (*Pisoniviricetes*: *Picornaviridae*) that causes febrile disease and has been found in Central Asia.

■ *ALSUVIRICETES: TOGAVIRIDAE*

Members of the family *Togaviridae* have linear, positive-sense RNA genomes (≈9.7–11.8 kb) and form enveloped icosahedral virions (≈60–70 nm in diameter) that bud from the plasma membrane of the infected cell. The togavirids discussed in this chapter are all members of the genus *Alphavirus* and are transmitted to vertebrates by mosquitoes.

■ *BUNYAVIRICETES: ARENAVIRIDAE*

Members of the family *Arenaviridae* that infect humans are all assigned to the genus *Mammarenavirus*. Mammarenaviruses form spherical, oval, or pleomorphic enveloped and spiked virions (≈50–300 nm in diameter) that bud from the plasma membrane of the infected cell. The particles contain two genomic single-stranded RNAs (small [S], ≈3.5 kb, and large [L], ≈7.5 kb), encoding structural proteins in an ambisense orientation. Most mammarenaviruses persist by chronically infecting rodents. Human mammarenaviruses are maintained by muroid rodents that are often continuously viremic and commonly transmit viruses vertically and horizontally. One mammarenavirus associated with human infections is maintained by shrews. Strikingly, each mammarenavirus is predominantly adapted to one particular type of rodent. Humans usually become infected through inhalation of or direct contact with infected rodent excreta or secreta (e.g., from aerosolized whole rodents in harvesting machines or aerosolized dried urine or feces when sweeping floors in barns or houses) and direct contact with rodents in traps. Person-to-person transmission of mammarenaviruses occurs but is uncommon.

■ *BUNYAVIRICETES: HANTAVIRIDAE, NAIROVIRIDAE, PERIBUNYAVIRIDAE,* AND *PHENUIVIRIDAE*

Members of these families that infect humans form spherical to pleomorphic enveloped virions containing three genomic RNAs (S, ≈1–2 kb; medium [M], 3.6–5.3 kb; and L, 6.4–12.3 kb) of negative-sense (hantavirids, nairovirids, and peribunyavirids) or negative-sense or ambisense polarities (phenuivirids). These bunyaviricetes mature into particles ≈80–120 nm in diameter in the Golgi complexes of infected cells and exit these cells by exocytosis.

Hantavirids that infect humans are classified in the genus *Orthohantavirus* and are maintained by muroid rodents that chronically shed virions. As with mammarenaviruses, individual orthohantaviruses are usually specifically adapted to a particular type of rodent. However, orthohantaviruses do not cause chronic viremia in their rodent hosts and are transmitted only horizontally from rodent to rodent. Also like mammarenaviruses, orthohantaviruses infect humans primarily through inhalation of or direct contact with rodent excreta or secreta, and person-to-person transmission occurs but is not common (with the notable exception of Andes virus).

Nairovirids that infect humans are classified in the genus *Orthonairovirus*. Orthonairoviruses are predominantly maintained by ixodid ticks, which transmit these viruses vertically (transovarially and transstadially) to progeny and horizontally through contact with viremic vertebrate hosts and subsequent transmission to healthy vertebrates. Humans are usually infected via a tick bite or during handling of infected vertebrates.

Peribunyavirids of one genus (*Orthobunyavirus*) infect humans. Orthobunyaviruses are primarily mosquito-borne (and, in rare circumstances, transmitted by midges or sandflies) and have viremic vertebrate intermediate hosts. Many orthobunyaviruses are transmitted vertically (transovarially) to their mosquito hosts. Numerous orthobunyaviruses have been associated with human infection and disease.

Phenuivirids are transmitted vertically (transovarially) in their arthropod vectors and horizontally through viremic vertebrate hosts. Human phenuivirids are found in four genera: *Bandavirus*, *Phlebovirus*, *Tanzavirus*, and *Uukuvirus*. Bandaviruses and uukuviruses are transmitted by ticks, whereas human phleboviruses are transmitted by sandflies (the ecology of tanzaviruses is unclear).

■ *FLASUVIRICETES: FLAVIVIRIDAE*

The family *Flaviviridae* currently includes only one genus (*Orthoflavivirus*) that comprises arthropod-borne human viruses. Orthoflaviviruses have single-stranded positive-sense RNA genomes (≈11 kb) and form spherical enveloped particles or virions (40–60 nm in diameter). The orthoflaviviruses discussed in this chapter belong to two phylogenetically and antigenically distinct groups that are transmitted among vertebrates by mosquitoes (generally, *Aedes* and *Culex* spp.) and, typically, ixodid ticks, respectively. Vectors are usually infected when they feed on viremic hosts. As in the case of most other viruses discussed in this chapter, humans are accidental hosts, typically infected by arthropod bites. Arthropods maintain orthoflavivirus infections horizontally, although vertical (transovarial) transmission has been documented. Under some circumstances, orthoflaviviruses can be transmitted by aerosol or via contaminated food products; in particular, raw milk can transmit tick-borne encephalitis virus.

■ *INSTHOVIRICETES: ORTHOMYXOVIRIDAE*

The family *Orthomyxoviridae* includes two genera of medically relevant arthropod-borne viruses: *Quaranjavirus* and *Thogotovirus*. Quaranjaviruses are transmitted among birds by argasid ticks, whereas thogotoviruses have a predilection for mammalian host reservoirs and can be transmitted by argasid and ixodid ticks and mosquitoes.

■ *MONJIVIRICETES: BORNAVIRIDAE AND RHABDOVIRIDAE*

Bornavirids and rhabdovirids that infect humans have linear nonsegmented, negative-sense RNA genomes (bornavirids, ≈9 kb; rhabdovirids, ≈11–15 kb) and form spherical (bornavirids, 90–130 nm in diameter) and bullet-shaped to pleomorphic enveloped particles (rhabdovirids, 100–430 nm long and 45–100 nm wide). The only known rodent-borne bornavirid that infects humans is a member of genus *Orthobornavirus* and is transmitted by squirrels. The rhabdovirid genus *Tibrovirus* contains several uncharacterized human pathogens

TABLE 215-1 Zoonotic Arthropod- and Rodent-Borne Viruses That Infect Humans

VIRUS TAXON	VIRUS (ABBREVIATION)	MAJOR NONHUMAN HOST(S)[a]	VECTOR(S)	SYNDROME[b]
Alphavirus (family *Togaviridae*)	Barmah Forest virus (BFV)	Cattle, horses, marsupials?	Biting midges (*Culicoides marksi*), mosquitoes (*Aedes camptorhynchus, Ae. normanensis, Ae. notoscriptus, Ae. vigilax, Culex annulirostris*)	A&R
	Chikungunya virus (CHIKV)	Bats, nonhuman primates	Mosquitoes (predominantly *Aedes aegypti* and *Ae. Albopictus*)	A&R[c]
	Eastern equine encephalitis virus (EEEV)	Freshwater swamp passeriform birds, but also opportunistic amphibians, other birds (emu, gallinaceous poultry, pheasants), reptiles, and mammals (goats, horses, pigs, rodents)	Mosquitoes (*Aedes*, *Coquillettidia*, *Culex* spp., *Anopheles quadrimaculatus*, *Culiseta melanura*, *Mansonia perturbans*, *Psorophora* spp.)	E
	Everglades virus (EVEV)	Cotton deermice (*Peromyscus gossypinus*), hispid cotton rats (*Sigmodon hispidus*)	Mosquitoes (*Culex cedecei*)	F&M, E
	Madariaga virus (MADV)	Likely bats, marsupials, reptiles, rodents	Mosquitoes (*Culex*, *Culiseta* spp.)	F&M, E
	Mayaro virus (MAYV)	Nonhuman primates, possums, rodents; possibly caimans, horses, sheep	Mosquitoes (predominantly *Haemagogus* spp., but also *Aedes*, *Culex*, *Mansonia*, *Psorophora*, *Sabethes* spp.)	A&R
	Mucambo virus (MUCV)	Nonhuman primates, rodents	Mosquitoes (*Culex*, *Ochlerotatus* spp.)	F&M, E
	O'nyong-nyong virus[d] (ONNV)	Unknown	Mosquitoes (in particular *Anopheles gambiae*, *A. funestus*, *Mansonia* spp.)	A&R
	Ross River virus (RRV)	Marsupials, rodents	Mosquitoes (*Aedes normanensis*, *Ae. vigilax*, *Culex annulirostris*)	A&R
	Semliki Forest virus (SFV)	Birds, rodents	Mosquitoes (*Aedes*, *Culex* spp.)	A&R
	Sindbis virus[e] (SINV)	Typically birds, but also frogs and rats	Typically mosquitoes (*Culex*, *Culiseta* spp.), but tick isolation has been reported	A&R
	Tonate virus (TONV)	Birds, Suriname crested oropendolas (*Psarocolius decumanus*)	Mosquitoes (*Anopheles, Coquillettidia*, *Culex*, *Mansonia*, *Uranotaenia*, *Wyeomyia* spp.), sandflies (*Lutzomyia* spp.)	F&M, E
	Una virus (UNAV)	Birds, horses, nonhuman primates, rodents	Mosquitoes (*Aedes*, *Anopheles*, *Coquillettidia*, *Culex*, *Ochlerotatus*, *Psorophora* spp.)	F&M
	Venezuelan equine encephalitis virus (VEEV)	Equids, rodents	Mosquitoes (*Aedes*, *Culex* spp., *Psorophora confinnis*)	F&M, E
	Western equine encephalitis virus (WEEV)	Equids, lagomorphs, passeriform birds, pheasants	Mosquitoes (*Aedes* spp., *Culex tarsalis*, *Culiseta* spp.)	E
Bandavirus (family *Phenuiviridae*)	Bhanja virus[f] (BHAV)	Cattle, four-toed hedgehog (*Atelerix albiventris*), goats, sheep, striped ground squirrels (*Xerus erythropus*)	Ixodid ticks (*Amblyomma*, *Dermacentor*, *Haemaphysalis*, *Hyalomma*, *Rhipicephalus* spp.)	F&M, E
	Heartland virus (HRTV)	Cattle, deer, elk, goats, raccoons, sheep?	Ixodid ticks (*Amblyomma americanum*)	F&M
	Severe fever with thrombocytopenia syndrome virus[g] (SFTSV)	Camels, cats, cattle, chickens, dogs, goats, hedgehogs, pigs, rodents, sheep?	Ixodid ticks (predominantly *Haemaphysalis longicornis*, but also *Amblyomma testudinarium*, *H. concinna*, *H. flava*, *Ixodes nipponensis*, *Rhipicephalus microplus*)	F&M, VHF
Coltivirus (family *Spinareoviridae*)	Colorado tick fever virus (CTFV)	Predominantly golden-mantled ground squirrels (*Spermophilus lateralis*), but also bushy-tailed woodrats (*Neotoma cinerea*), Columbian ground squirrels (*Spermophilus columbianus*), eastern deermice (*Peromyscus maniculatus*), least chipmunks (*Tamias minimus*), North American porcupines (*Erethizon dorsata*), Richardson ground squirrel (*Spermophilus richardsonii*), Uinta chipmunks (*Tamias umbrinus*), yellow-pine chipmunks (*Tamias amoenus*) among lesser important mammal hosts	Ixodid ticks (*Dermacentor andersoni*, possibly *D. occidentalis*)	F&M, E
	Eyach virus (EYAV)	Lagomorphs, rodents	Ixodid ticks (*Ixodes ricinus*, *I. ventalloi*)	F&M, E
	Salmon River virus (SRV)	Unknown	Ixodid ticks (*Ixodes* spp.)	F&M, E
Mammarenavirus (family *Arenaviridae*)	Chapare virus (CHAPV)	Unidentified cricetid rodents	None	VHF
	Flexal virus (FLEV)	Unidentified cricetid rodents	None	(F&M)
	Guanarito virus (GTOV)	Predominantly short-tailed zygodonts (*Zygodontomys brevicauda*) but also Alston's cotton rats (*Sigmodon alstoni*)	None	VHF

(*Continued*)

TABLE 215-1 Zoonotic Arthropod- and Rodent-Borne Viruses That Infect Humans (Continued)

VIRUS TAXON	VIRUS (ABBREVIATION)	MAJOR NONHUMAN HOST(S)[a]	VECTOR(S)	SYNDROME[b]
	Junín virus (JUNV)	Predominantly drylands lauchas (*Calomys musculinus*) but also Azara's akodonts (*Akodon azarae*) and little lauchas (*Calomys laucha*)	None	VHF
	Lassa virus (LASV)	Predominantly Natal mastomys (*Mastomys natalensis*), but also reddish-white mastomys (*M. erythroleucus*), African hylomuscus (*Hylomyscus pamfi*), and Baoule's mice (*Mus baoulei*)	None	F&M, VHF
	Lymphocytic choriomeningitis virus (LCMV)	Predominantly house mice (*Mus musculus*), but also murid long-tailed field mice (*Apodemus sylvaticus*), soft-furred mice (*Praomys* spp.), and golden hamsters (*Mesocricetus auratus*)	None	F&M, E, (VHF)
	Lujo virus (LUJV)	Unknown	None	VHF
	Machupo virus (MACV)	Big lauchas (*Calomys callosus*)	None	VHF
	Sabiá virus (SBAV)	Unidentified cricetid rodents	None	VHF
	Whitewater Arroyo virus (WWAV)[l]	White-throated woodrats (*Neotoma albigula*)	None	(E)
Orbivirus (family *Sedoreoviridae*)	Kemerovo virus (KEMV)	Birds, rodents	Ixodid ticks (*Ixodes persulcatus*)	F&M, E
	Lebombo virus (LEBV)	Unknown	Mosquitoes (*Aedes*, *Mansonia* spp.)	F&M
	Orungo virus (ORUV)	Camels, cattle, goats, nonhuman primates, sheep	Mosquitoes (*Aedes*, *Anopheles*, *Culex* spp.)	F&M, E
	Tribeč virus (TRBV)[m]	Bank voles (*Myodes glareolus*), birds, common pine voles (*Microtus subterraneus*), goats, hares	Ixodid ticks (*Ixodes persulcatus*, *I. ricinus*)	F&M
Orthobornavirus (family *Bornaviridae*)	Variegated squirrel bornavirus 1 (VSBV1)	Finlayson's squirrel (*Callosciurus finlaysonii*), Prevost's squirrels (*Callosciurus prevostii*), Swinhoe's striped squirrel (*Tamiops swinhoei*), variegated squirrels (*Sciurus variegatoides*)	None	E
Orthobunyavirus (family *Peribunyaviridae*)	Apeú virus (APEUV)	Bare-tailed woolly opossums (*Caluromys philander*) and other opossums; rodents; black howlers (*Alouatta caraya*), tufted capuchins (*Cebus apella*)	Mosquitoes (*Aedes*, *Culex* spp.)	F&M
	Bangui virus (BGIV)	Unknown	Unknown	F&M
	Batai virus (BATV)[n]	Birds, camels, cattle, goats, rodents, sheep	Mosquitoes (*Aedes abnormalis*, *A. curtipes*, *Anopheles barbirostris*, *A. maculipennis*, *A. phillippinensis*, *Culex gelidus*, other spp.)	F&M
	Bunyamwera virus (BUNV)	Birds, cows, goats, horses, sheep	Mosquitoes (*Aedes* spp.)	F&M
	Bwamba virus (BWAV)	Unknown	Mosquitoes (*Aedes*, *Anopheles*, *Mansonia* spp.)	F&M
	Cache Valley virus (CVV)	Cattle, deer, dogs, foxes, groundhogs, horses, nonhuman primates, pigs, rabbits, raccoons, rodents	Mosquitoes (*Aedes*, *Anopheles*, *Coquillettidia*, *Culiseta*, *Psorophora* spp.)	F&M, E
	California encephalitis virus (CEV)	Lagomorphs, rodents	Mosquitoes (*Aedes*, *Culex*, *Culiseta*, *Psorophora* spp.)	F&M, E
	Caraparú virus (CARV)	Rodents, tufted capuchins (*C. apella*)	Mosquitoes (*Culex* spp.)	F&M
	Catú virus (CATUV)	Bats, capuchins (*Cebus* spp.), opossums, rodents	Mosquitoes (*Culex* spp.)	F&M
	Cristoli virus (/)	Unknown	Mosquitoes?	E
	Fort Sherman virus (FSV)	Cattle, goats, horses, sheep?	Mosquitoes?	F&M
	Gan Gan virus (GGV)	Unknown	Mosquitoes (*Aedes*, *Culex* spp.)	A&R
	Germiston virus (GERV)	Rodents	Mosquitoes (*Culex* spp.)	F&M
	Guamá virus (GMAV)	Bats, capuchins (*Cebus* spp.), howlers (*Alouatta* spp.), marsupials, rodents	Mosquitoes (*Aedes*, *Culex*, *Limatus*, *Mansonia*, *Psorophora*, *Trichoprosopon* spp.)	F&M
	Guaroa virus (GROV)	Unknown	Mosquitoes (*Anopheles* spp.)	F&M
	Ilesha virus (ILEV)	Unknown	Mosquitoes (*Anopheles gambiae*)	F&M, (VHF)
	Inkoo virus (INKV)	Cattle, foxes, hares, moose, reindeer, rodents	Mosquitoes (*Aedes* spp.)	F&M, E

(*Continued*)

TABLE 215-1 Zoonotic Arthropod- and Rodent-Borne Viruses That Infect Humans (Continued)

VIRUS TAXON	VIRUS (ABBREVIATION)	MAJOR NONHUMAN HOST(S)[a]	VECTOR(S)	SYNDROME[b]
	Iquitos virus (IQTV)	Unknown	Unknown	F&M
	Itaquí virus (ITQV)	Capuchins (*Cebus* spp.), opossums, rodents	Mosquitoes (*Culex* spp.)	F&M
	Jamestown Canyon virus (JCV)	Predominantly white-tailed deer (*Odocoileus virginianus*), but also bison, elk, and moose	Mosquitoes (*Aedes*, Anopheles, Coquillettidia, *Culiseta*, *Ochlerotatus* spp.)	F&M, E
	Keystone virus (KEYV)	Rabbits, squirrels	Mosquitoes (*Ochlerotatus atlanticus*)	E
	La Crosse virus (LACV)	Chipmunks, squirrels	Mosquitoes (*Ochlerotatus triseriatus*)	F&M, E
	Lumbo virus (LUMV)	Unknown	Mosquitoes (*Aedes pembaensis*)	F&M, E
	Madrid virus (MADV)	Capuchins (*Cebus* spp.), opossums, rodents	Mosquitoes (*Culex* spp.)	F&M
	Maguari virus (MAGV)	Birds, cattle, horses, sheep, water buffalo	Mosquitoes (*Aedes*, *Anopheles*, *Culex*, *Psorophora*, *Wyeomyia* spp.)	F&M
	Marituba virus (MTBV)	Capuchins (*Cebus* spp.), opossums, rodents	Mosquitoes (*Culex* spp.)	F&M
	Murutucú virus (MURV)	Capuchins (*Cebus* spp.), opossums, pale-throated sloths (*Bradypus tridactylus*), rodents	Mosquitoes (*Coquillettidia*, *Culex* spp.)	F&M
	Nepuyo virus (NEPV)	Bats (*Artibeus* spp.), rodents	Mosquitoes (*Culex* spp.)	F&M
	Ngari virus (NRIV)	Cattle, goats, sheep	Mosquitoes (*Aedes*, *Anopheles*, *Culex* spp.)	F&M, VHF
	Nyando virus (NDV)	Unknown	Mosquitoes (*Aedes*, *Anopheles* spp.), sandflies (*Lutzomyia* spp.)	F&M
	Oriboca virus (ORIV)	Capuchins (*Cebus* spp.), opossums, rodents	Mosquitoes (*Aedes*, *Culex*, *Mansonia*, *Psorophora* spp.)	F&M
	Oropouche virus (OROV)	Marmosets (*Callithrix* spp.), pale-throated sloths (*B. tridactylus*)	Biting midges (*Culicoides paraensis*), mosquitoes (*Coquillettidia venezuelensis*, *Culex quinquefasciatus*, *Mansonia* spp., *Ochlerotatus serratus*)	F&M, E
	Ossa virus (OSSAV)	Rodents	Mosquitoes (*Culex* spp.)	F&M
	Pongola virus (PGAV)	Cattle, donkeys, goats, sheep	Mosquitoes (*Aedes*, *Anopheles*, *Mansonia* spp.)	F&M
	Restan virus (RESV)	Unknown	Mosquitoes (*Culex* spp.)	F&M
	Shokwe virus (SHOV)	Rodents	Mosquitoes (*Aedes*, *Anopheles*, *Mansonia* spp.)	F&M
	Shuni virus (SHUV)	Horses, livestock	Mosquitoes (*Culex theileri*, *Culicoides* spp.)	E
	Snowshoe hare virus (SSHV)	Collared lemmings, snowshoe hares, squirrels, voles	Mosquitoes (*Aedes*, *Culiseta*, *Ochlerotatus*, *Simulium* spp.)	F&M, E
	Tacaiuma virus (TCMV)	Nonhuman primates	Mosquitoes (*Anopheles*, *Haemagogus* spp.)	F&M
	Ťahyňa virus (TAHV)	Boars, cattle, deer, dogs, eulipotyphla, foxes, hares, horses, pigs, rodents	Mosquitoes (*Aedes*, *Anopheles*, *Culex*, *Culiseta* spp.)	F&M, E
	Tataguine virus (TATV)	Unknown	Mosquitoes (*Anopheles* spp.)	F&M
	Trubanaman virus (TRUV)	Unknown	Mosquitoes (*Anopheles*, *Culex* spp.)	(A&R)
	Wyeomyia virus (WYOV)	Unknown	Mosquitoes (*Wyeomyia* spp.)	F&M
	Xingu virus (XINV)	Unknown	Unknown	F&M
	Zungarococha virus (ZUNV)	Unknown	Unknown	F&M
Orthoflavivirus (family *Flaviviridae*)	Alkhurma hemorrhagic fever virus (AHFV)[j]	Livestock?	Argasid ticks (*Ornithodoros savignyi*), ixodid ticks (*Hyalomma dromedarii*, *H. rufipes*)	VHF
	Apoi virus (APOIV)	Field mice, voles	Unknown	E
	Banzi virus (BANV)	Rodents?	Mosquitoes (*Culex rubinotus*)	F&M
	Bussuquara virus (BSQV)	Cotton rats, monkeys, spiny-rats	Mosquitoes (*Culex* spp.)	F&M
	Cacipacoré virus (CPCV)	Birds, horses, nonhuman primates, water buffalo	Mosquitoes (*Aedes*, *Anopheles*, *Culex* spp.)	F&M
	Dakar bat virus (DBV)	Bats	Unknown	F&M
	Dengue viruses 1–4 (DENV 1–4)	Nonhuman primates	Mosquitoes (predominantly *Aedes aegypti*, *A. albopictus*)	F&M, VHF
	Edge Hill virus (EHV)	Bandicoots, dogs, wallabies	Mosquitoes (*Aedes vigilax*, *Culex annulirostris*)	F&M
	Ilhéus virus (ILHV)	Birds, coatis, nonhuman primates, rodents, reptiles, sloths, water buffalo	Mosquitoes (*Aedes*, *Culex*, *Coquillettidia*, *Haemagogus*, *Ochlerotatus*, *Psorophora*, *Sabethes*, *Trichoprosopon* spp.)	E

(*Continued*)

TABLE 215-1 Zoonotic Arthropod- and Rodent-Borne Viruses That Infect Humans (Continued)

VIRUS TAXON	VIRUS (ABBREVIATION)	MAJOR NONHUMAN HOST(S)[a]	VECTOR(S)	SYNDROME[b]
	Japanese encephalitis virus (JEV)	Ardeid wading birds (in particular herons), horses, pigs	Mosquitoes (*Culex* spp., in particular *C. tritaeniorhynchus*)	E
	Karshi virus (KSIV)	Great gerbils (*Rhombomys opimus*)	Argasid ticks (*Ornithodoros capensis*), ixodid ticks (*Hyalomma asiaticum*)	F&M, E
	Kédougou virus (KEDV)	Unknown	Mosquitoes (*Culex* spp.)	F&M
	Kokobera virus (KOKV)	Macropods, horses	Mosquitoes (*Culex* spp.)	A&R
	Koutango virus (KOUV)	Gerbils	Mosquitoes (*Aedes* spp.)	F&M
	Kyasanur Forest disease virus (KFDV)[k]	Indomalayan vandeleurias (*Vandeleuria oleracea*), roof rats (*Rattus rattus*)	Ixodid ticks (predominantly *Haemaphysalis spinigera*)	VHF
	Modoc virus (MODV)	North American deermice (*Peromyscus maniculatus*)	Unknown	E
	Murray Valley encephalitis virus[h] (MVEV)	Birds (egrets, herons)	Mosquitoes (predominantly *C. annulirostris*)	E
	Ntaya virus (NTAV)	Birds, domestic animals	Mosquitoes (*Culex* spp.)	F&M
	Omsk hemorrhagic fever virus (OHFV)	Migratory birds, rodents	Ixodid ticks (predominantly *Dermacentor* spp.)	VHF
	Powassan virus (POWV)	Red squirrels (*Tamiasciurus hudsonicus*), white-footed deermice (*Peromyscus leucopus*), woodchucks (*Marmota monax*), other small mammals	Ixodid ticks (in particular *Ixodes cookei*, other *Ixodes* spp., *Dermacentor* spp.)	E
	Rio Bravo virus (RBV)	Bats	Unknown	F&M
	Rocio virus (ROCV)	Rufous-collared sparrows (*Zonotrichia capensis*), various migratory birds, horses, water buffalo (*Bubalus bubalis*)	Mosquitoes (*Aedes*, *Culex*, *Psorophora* spp.)	E
	Sepik virus (SEPV)	Unknown	Mosquitoes (*Mansonia septempunctata*)	F&M
	Spondweni virus (SPOV)	Cattle, sheep	Mosquitoes (*Aedes*, *Culex*, *Mansonia* spp.)	F&M
	St. Louis encephalitis virus (SLEV)	Columbiform and passeriform birds (finches, sparrows)	Mosquitoes (predominantly *Culex* spp., in particular *C. nigripalpus*, *C. pipiens*, *C. quinquefasciatus*, *C. tarsalis*)	E
	Tick-borne encephalitis virus (TBEV)	Passeriform birds, deer, eulipotyphla, goats, grouse, small mammals, rodents, sheep	Ixodid ticks (*Ixodes gibbosus*, *I. persulcatus*, *I. ricinus*; sporadically *Dermacentor*, *Haemaphysalis*, *Hyalomma* spp.)	F&M, E, (VHF)
	Usutu virus (USUV)	Accipitriform, columbiform, passeriform, and strigiform birds	Mosquitoes (*Culex* spp., in particular *C. pipiens*, but also *Aedes* spp, *Anopheles* spp., *Ochlerotatus* spp.)	(E)
	Wesselsbron virus (WESSV)	Cattle, goats, rodents, sheep	Mosquitoes (*Aedes* spp.)	F&M
	West Nile virus (WNV)[i]	Passeriform birds (blackbirds, crows, finches, sparrows), small mammals, horses	Mosquitoes (*Culex* spp., in particular *C. pipiens*, *C. quinquefasciatus*, *C. restuans*, *C. tarsalis*)	E
	Yellow fever virus (YFV)	Nonhuman primates (*Alouatta*, *Ateles*, *Cebus*, *Cercopithecus*, *Colobus* spp.)	Mosquitoes (*Aedes* spp., in particular *Ae. aegypti*)	VHF
	Zika virus (ZIKV)	Nonhuman primates (*Macaca*, *Pongo* spp.)	Mosquitoes (*Aedes* spp.)	F&M, A&R
Orthohantavirus (family *Hantaviridae*)	Amur virus (AMRV)	Korean field mice (*Apodemus peninsulae*)	None	VHF
	Anajatuba virus (ANJV)	Fornes' colilargos (*Oligoryzomys fornesi*)	None	P
	Andes virus (ANDV)	Long-tailed colilargos (*Oligoryzomys longicaudatus*)	None	P
	Araraquara virus (ARAV)	Hairy-tailed akodonts (*Necromys lasiurus*)	None	P
	Araucária virus (ARAUV)	Black-footed colilargos (*Oligoryzomys nigripes*)	None	P
	Bayou virus (BAYV)	Marsh rice rats (*Oryzomys palustris*)	None	P
	Bermejo virus (BMJV)	Chacoan colilargos (*Oligoryzomys chacoensis*)	None	P
	Black Creek Canal virus (BCCV)	Hispid cotton rats (*Sigmodon hispidus*)	None	P
	Blue River virus (BRV)	White-footed deermice (*Peromyscus leucopus*)	None	P
	Caño Delgadito virus (CADV)	Alston's cotton rats (*Sigmodon alstoni*)	None	P

(*Continued*)

TABLE 215-1 Zoonotic Arthropod- and Rodent-Borne Viruses That Infect Humans (Continued)

VIRUS TAXON	VIRUS (ABBREVIATION)	MAJOR NONHUMAN HOST(S)[a]	VECTOR(S)	SYNDROME[b]
	Castelo dos Sonhos virus (CASV)	Brazilian colilargos (*Oligoryzomys eliurus*)	None	P
	Catacamas virus (CATV)	Coues' oryzomys (*Oryzomys couesi*)	None	P
	Choclo virus (CHOV)	Fulvous colilargos (*Oligoryzomys fulvescens*)	None	F&M, P
	Dobrava virus (DOBV)	Caucasus field mice (*Apodemus ponticus*), striped field mice (*A. agrarius*), yellow-necked field mice (*A. flavicollis*)	None	VHF
	El Moro Canyon virus (ELMCV)	Western harvest mice (*Reithrodontomys megalotis*)	None	VHF
	Gōu virus (GOUV)	Brown rats (*Rattus norvegicus*), roof rats (*R. rattus*), Oriental house rats (*R. tanezumi*)	None	VHF
	Hantaan virus (HTNV)	Striped field mice (*Apodemus agrarius*)	None	VHF
	Juquitiba virus (JUQV)	Black-footed colilargos (*Oligoryzomys nigripes*)	None	P
	Kurkino virus (KURV)	Striped field mice (*Apodemus agrarius*)	None	VHF
	Laguna Negra virus (LANV)	Little lauchas (*Calomys laucha*)	None	P
	Lechiguanas virus (LECV)	Flavescent colilargos (*Oligoryzomys flavescens*)	None	P
	Maciel virus (MCLV)	Dark-furred akodonts (*Necromys obscurus*)	None	P
	Maripa virus (MARV)	Unknown	None	P
	Monongahela virus (MGLV)	Eastern deermice (*Peromyscus maniculatus*)	None	P
	Muju virus (MUJV)	Korean red-backed voles (*Myodes regulus*)	None	VHF
	New York virus (NYV)	White-footed deermice (*Peromyscus leucopus*)	None	P
	Orán virus (ORNV)	Long-tailed colilargos (*Oligoryzomys longicaudatus*)	None	P
	Paranoá virus (PARV)	Unknown	None	P
	Pergamino virus (PRGV)	Azara's akodonts (*Akodon azarae*)	None	P
	Puumala virus (PUUV)	Bank voles (*Myodes glareolus*)	None	P, VHF
	Rio Mamoré virus (RIOMV)	Common bristly mice (*Neacomys spinosus*)	None	P
	Saaremaa virus (SAAV)	Striped field mice (*Apodemus agrarius*)	None	VHF
	Seoul virus (SEOV)	Brown rats (*Rattus norvegicus*), roof rats (*R. rattus*)	None	VHF
	Sin Nombre virus (SNV)	Western deermice (*Peromyscus sonoriensis*)	None	P
	Sochi virus (SOCV)	Caucasus field mice (*Apodemus ponticus*)	None	VHF
	Tula virus (TULV)	Common voles (*Microtus arvalis*), East European voles (*M. levis*), field voles (*M. agrestis*)	None	(P), VHF
	Tunari virus (TUNV)	Unknown	None	P
Orthonairovirus (family *Nairoviridae*)	Aigai virus (AIGV)	Cattle, goats, tortoises	Ixodid ticks (*Hyalomma* spp., *Rhipicephalus* spp.)	VHF
	Avalon virus (AVAV)	European herring gulls (*Larus argentatus*)	Ixodid ticks (*Ixodes uriae*)	(Polyradiculoneuritis?)
	Crimean-Congo hemorrhagic fever virus (CCHFV)	Cattle, dogs, goats, hares, hedgehogs, mice, ostriches, sheep	Predominantly ixodid ticks (*Hyalomma* spp.)	VHF
	Dugbe virus (DUGV)	Northern giant pouched rats (*Cricetomys gambianus*), Zébu cattle (*Bos primigenius*)	Biting midges (*Culicoides* spp.), ixodid ticks (*Amblyomma*, *Hyalomma*, *Rhipicephalus* spp.)	F&M
	Erve virus (ERVEV)	Greater white-toothed shrews (*Crocidura russula*)	Unknown	(Thunderclap headache?)
	Issyk-kul virus (ISKV)	Bats, birds	Biting midges (*Culicoides schultzei*), horseflies (*Tabanus agrestis*), mosquitoes (*Aedes caspius*, *Anopheles hyrcanus*), argasid ticks (*Argas vespertilionis*, *A. pusillus*), ixodid ticks (*Ixodes vespertilionis*)	F&M

(*Continued*)

TABLE 215-1 Zoonotic Arthropod- and Rodent-Borne Viruses That Infect Humans (Continued)

VIRUS TAXON	VIRUS (ABBREVIATION)	MAJOR NONHUMAN HOST(S)[a]	VECTOR(S)	SYNDROME[b]
	Nairobi sheep disease virus[o] (NSDV)	Sheep	Ixodid ticks (*Amblyomma*, *Haemaphysalis*, *Rhipicephalus* spp.), mosquitoes (*Culex* spp.)	F&M
	Sönglǐng virus (SGLV)	Great gerbils (*Rhombomys opimus*)	Ixodid ticks (*Ixodes crenulatus, Ixodes persulcatus, Haemaphysalis concinna*, and *Haemaphysalis longicornis*)	F&M
	Tǎchéng tick virus 1 (TcTV1)	Sheep, tortoises	Ixodid ticks (*Dermacentor marginatus*, *Hyalomma aegyptium*)	F&M
	Tamdy virus (TAMV)	Gerbils, other mammals (including Bactrian camels), birds	Ixodid ticks (*Dermacentor*, *Hyalomma* spp.)	F&M
	Wetland virus (WELV)	Horses, pigs, sheep, North China zokors	Ixodid ticks (*Dermacentor*, *Ixodes*, *Haemaphysalis* spp.)	F&M
	Yezo virus (YEZV)	Unknown	Ixodid ticks (*Ixodes persulcatus*)	F&M
Phlebovirus (family *Phenuiviridae*)	Adria virus (ADRV)	Unknown	Sandflies	E
	Alenquer virus (ALEV)	Unknown	Unknown	F&M
	Candirú virus (CDUV)	Unknown	Unknown	F&M
	Chagres virus (CHGV)	Unknown	Sandflies (*Lutzomyia* spp.)	F&M
	Chios virus (/)	Unknown	Unknown	E
	Coclé virus (CCLV)	Unknown	Sandflies	F&M
	Echarate virus (ECHV)	Unknown	Unknown	F&M
	Granada virus (GRV = GRAV)	Unknown	Sandflies	F&M
	Maldonado virus (MLOV)	Unknown	Unknown	F&M
	Morumbi virus (MRBV = MRMBV)	Unknown	Unknown	F&M
	Punta Toro virus (PTV)	Unknown	Sandflies (*Lutzomyia* spp.)	F&M
	Rift Valley fever virus (RVFV)	Cattle, sheep	Mosquitoes (*Aedes*, *Anopheles*, *Coquillettidia*, *Culex*, *Eretmapodites*, *Mansonia* spp.)	F&M, E, VHF
	Sandfly fever Cyprus virus (SFCV)	Unknown	Unknown	F&M
	Sandfly fever Ethiopia virus (SFEV)	Unknown	Sandflies	F&M
	Sandfly fever Naples virus (SFNV)	Unknown	Sandflies (*Phlebotomus papatasi, P. perfiliewi, P. perniciosus*)	F&M
	Sandfly fever Sicilian virus (SFSV)	Eulipotyphla, least weasels (*Mustela nivalis*), rodents	Sandflies (predominantly *Phlebotomus papatasi*)	F&M
	Sandfly fever Turkey virus (SFTV)	Unknown	Sandflies (*Phlebotomus* spp.)	F&M
	Serra Norte virus (SRNV)	Unknown	Unknown	F&M
	Toscana virus (TOSV)	Migratory birds?	Sandflies (*Phlebotomus papatasi, P. perfiliewi, Sergentomyia* spp.)	F&M, E
Quaranjavirus (family *Orthomyxoviridae*)	Quaranfil virus (QRFV)	Birds	Argasid ticks (*Argas arboreus*)	F&M
Seadornavirus (family *Sedoreoviridae*)	Banna virus (BAV)	Cattle, pigs	Biting midges (*Culicoides* spp.), mosquitoes (*Aedes, Anopheles, Culiseta* spp.), ticks	E
Tanzavirus (family *Phenuviridae*)	Dar es Salaam virus (DeSV)	Unknown	Unknown	F&M
Tibrovirus (family *Rhabdoviridae*)	Bas-Congo virus (BASV)	Artiodactyls?	Biting midges?	(VHF)
	Ekpoma virus 1 (EKV1)	Artiodactyls?	Biting midges?	F&M
	Ekpoma virus 2 (EKV2)	Artiodactyls?	Biting midges?	F&M
	Mundri virus (MUNV)	Artiodactyls?	Biting midges?	E (New onset nodding syndrome)
Thogotovirus (family *Orthomyxoviridae*)	Bourbon virus (BRBV)	Raccoons (*Procyon lotor*), white-tailed deer (*Odocoileus virginianus*)?	Ixodid ticks (*Amblyomma americanum*)	F&M
	Dhori virus (DHOV)[p]	Bats, camels, horses	Mosquitoes (*Aedes, Anopheles, Culex* spp.), argasid ticks (*Ornithodoros* spp.), ixodid ticks (*Dermacentor, Hyalomma* spp.)	F&M, E
	Thogoto virus (THOV)	Camels, cattle	Ixodid ticks (*Amblyomma, Hyalomma, Rhipicephalus* spp.)	F&M, E

(*Continued*)

TABLE 215-1 Zoonotic Arthropod- and Rodent-Borne Viruses That Infect Humans (Continued)

VIRUS TAXON	VIRUS (ABBREVIATION)	MAJOR NONHUMAN HOST(S)[a]	VECTOR(S)	SYNDROME[b]
Uukuvirus (family *Phenuiviridae*)	Tăchéng tick virus 2 (TcTV2)	Unknown	Ixodid ticks (*Dermacentor marginatus, Dermacentor nuttalli, Dermacentor silvarum, Hyalomma asiaticum*)	(E)
	Uukuniemi virus (UUKV)	Birds, cattle, rodents	Ixodid ticks (*Ixodes* spp.)	F&M
Vesiculovirus (family *Rhabdoviridae*)	Chandipura virus (CHPV)	Hedgehogs	Mosquitoes (*Aedes aegypti*), sandflies (*Phlebotomus, Sergentomyia* spp.)	F&M, E
	Isfahan virus (ISFV)	Great gerbils (*Rhombomys opimus*)	Sandflies (*Phlebotomus papatasi*)	F&M
	Piry virus (PIRYV)	Gray four-eyed opossums (*Philander opossum*)	Mosquitoes (*Aedes, Culex, Toxorhynchites* spp.)	F&M
	Vesicular stomatitis Indiana virus (VSIV)	Cattle, horses, pigs	Sandflies (*Lutzomyia* spp.)	F&M
	Vesicular stomatitis New Jersey virus (VSNJV)	Cattle, horses, pigs	Biting midges (*Culicoides* spp.), chloropid flies, mosquitoes (*Culex, Mansonia* spp.), muscoid flies (*Musca* spp.), simuliid flies	F&M

[a] Mammalian names as listed in Wilson & Reeder's *Mammal Species of the World*, 3rd edition (*https://www.departments.bucknell.edu/biology/resources/msw3/*).
[b] Abbreviations refer to the syndromes most associated with the viruses: A&R, arthritis and rash; E, encephalitis; F&M, fever and myalgia; P, pulmonary; VHF, viral hemorrhagic fever. Abbreviations are placed in parentheses when cases are controversial. [c] In the older literature, chikungunya virus often is also listed as a causative agent of VHF. However, later studies revealed that, in most cases, people with "chikungunya hemorrhagic fever" were co-infected with one or more dengue viruses, an observation suggesting that the VHF was severe dengue. [d] Also known as Igbo-Ora virus. [e] Also known as Ockelbo virus (OCKV), Pogosta virus, and Karelian fever virus (KFV). [f] Also known as Palma virus (PALV). [g] Alternatives used in the literature are Huáiyángshān virus (HYSV) and Hénán fever virus (HNFV). [h] Also known as Alfuy virus (ALFV). [i] Also includes Kunjin virus (KUNV). [j] Also spelled Alkhumra hemorrhagic fever virus (AHFV) and known as Alkhurma/Alkhumra virus (ALKV). [k] Also known as Nánjiànyí(n) virus. [l] Whitewater Arroyo virus is often listed as a causative agent of VHF in the literature, but convincing data associating this virus with VHF have not been published. [m] Also known as Brezová virus, Cvilín virus, Kharagysh virus, Koliba virus, or Lipovník virus. [n] Also known as Čalovo virus (CVOV) or Chittoor virus (CHITV). [o] Also known as Ganjam virus (GV). [p] Also known as Astra virus and Batken virus (BKNV).

that cause fever and myalgia (F&M) syndrome and possibly viral hemorrhagic fever, whereas the genus *Vesiculovirus* includes confirmed human arthropod-borne viruses, all of which are transmitted by insects (biting midges, mosquitoes, and sandflies). **The general properties of rhabdovirids are discussed in more detail in Chap. 214.**

■ *RESENTOVIRICETES: SEDOREOVIRIDAE* AND *SPINAREOVIRIDAE*

The families *Sedoreoviridae* and *Spinarevoviridae* are established for viruses with linear, multisegmented, double-stranded RNA genomes (≈16–29 kb in total). These viruses produce particles that have icosahedral symmetry and are 60–80 nm in diameter. In contrast to all other virions discussed in this chapter, resentovirions are not enveloped and thus are insensitive to detergent inactivation. Human arthropod-borne viruses are found within the genera *Coltivirus* (family *Spinareoviridae*), *Orbivirus*, and *Seadornavirus* (both in family *Sedoreoviridae*). Arthropod-borne coltiviruses possess 12 genome segments and are transmitted by numerous tick types transstadially but not transovarially. Therefore, maintenance of the transmission cycle involves viremic mammalian hosts infected by tick bites. Arthropod-borne orbiviruses have 10 genome segments and are transmitted by mosquitoes or ixodid ticks, whereas relevant seadornaviruses have 12 genome segments and are transmitted exclusively by mosquitoes.

EPIDEMIOLOGY

Arthropod-borne and rodent-borne viruses are generally confined to the areas inhabited by their reservoir hosts and/or vectors. Consequently, a patient's geographic origin or travel history can provide important clues in the differential diagnosis. **Table 215-2** lists the approximate geographic distribution of most arthropod-borne and rodent-borne infections. Many of these diseases can be acquired in rural or urban settings. However, diseases commonly associated with urban outbreaks include yellow fever (YF), dengue without/with warning signs (previously called dengue fever), severe dengue (previously called dengue hemorrhagic fever and dengue shock syndrome), chikungunya virus disease, hemorrhagic fever with renal syndrome (HFRS) caused by Seoul virus, sandfly fever caused by sandfly fever Naples and Sicilian viruses, and Oropouche virus disease.

DIAGNOSIS

In patients with suspected viral infection, a history of mosquito bite(s) has little diagnostic significance, but a history of tick bite(s) is more useful. Patients infected with mammarenaviruses or orthohantaviruses sometimes report exposure to rodents. Although clinical signs and epidemiologic clues may enable presumptive etiologic diagnosis in epidemic settings, laboratory diagnosis is required in all cases. For most arthropod-borne and rodent-borne virus diseases, acute-phase serum samples (collected within 3 or 4 days of onset) have yielded isolates. Paired serum samples have been used to show rising antibody titers. Intensive efforts to develop rapid tests for viral hemorrhagic fevers (VHFs) have resulted in reliable antigen-detection enzyme-linked immunosorbent assays (ELISAs), IgM-capture ELISAs, and multiplex polymerase chain reaction (PCR) assays. These tests can provide a diagnosis based on a single serum sample within a few hours and are particularly useful for assessment of patients with severe disease. More sensitive reverse transcriptase PCR (RT-PCR) assays may yield diagnoses based on samples without detectable antigen and may also provide useful truly quantifiable genetic information about the etiologic agent.

At diagnosis, patients with encephalitides generally are no longer viremic or antigenemic and usually do not have virions in cerebrospinal fluid (CSF). In this situation, either serologic detection of IgM or RT-PCR detection of viral nucleic acid may enable diagnosis. Increasingly, IgM-capture ELISAs are used for the simultaneous testing of serum and CSF. IgG ELISA or classic serology is useful in the evaluation of past exposure to viruses, many of which circulate in areas with minimal medical infrastructures and sometimes cause only mild or subclinical infections (and are not recognized).

CLINICAL DISEASE SYNDROMES

There is a wide spectrum of possible human responses to infection with arthropod-borne or rodent-borne viruses, and knowledge of the outcome of most of these infections is limited. People infected with these viruses may not develop or recognize symptoms or signs of illness. If viral disease is recognized, it can usually be grouped into one of five broad syndromic categories: F&M, arthritis and rash (A&R), encephalitis, pulmonary disease, or VHF **(Table 215-3)**. Although a useful clinical heuristic, it should be acknowledged that these categories often overlap in the complex spectra of disease caused by arthropod-borne and rodent-borne viruses. Furthermore, illness caused by many of these viruses is often best known by the most severe disease phenotypes, which typically do not result in the most common disease manifestations. For example, infections with West Nile virus (WNV) and Venezuelan equine encephalitis virus (VEEV) are discussed in this chapter as encephalitides, but during epidemics, many patients present with much milder F&M. Similarly, Rift Valley fever virus (RVFV)

TABLE 215-2 Geographic Distribution of Zoonotic Arthropod-Borne or Rodent-Borne Viral Diseases

	TYPE OF DISEASE[b]					
AREA[a]	***BUNYAVIRICETES*** **(*Arenaviridae, Hantaviridae, Nairoviridae, Peribunyaviridae, Phenuiviridae*)**	***FLASUVIRICETES*** **(*Flaviviridae*)**	***INSTHOVIRICETES*** **(*Orthomyxoviridae*)**	***RESENTOVIRICETES*** **(*Sedoreoviridae, Spinareoviridae*)**	***MONJIVIRICETES*** **(*Bornaviridae, Rhabdoviridae*)**	***ALSUVIRICETES*** **(*Togaviridae*)**
Africa	Aigai, Bangui, Batai, Bhanja, Bunyamwera, and Bwamba virus infections, Crimean-Congo hemorrhagic fever, Dar es Salaam, Dugbe, Germiston, Ilesha virus infections, Lassa fever, Lujo virus, Nairobi sheep disease virus infection, Ngari, Nyando, and Pongola virus infections, Rift Valley fever, sandfly fever, Shokwe, Shuni, Tataguine virus infections	Alkhurma hemorrhagic fever, Banzi virus infection, Dakar bat, dengue without/with warning signs/severe dengue, Kédougou, Koutango, Ntaya, Spondweni, Usutu, Wesselsbron, West Nile virus infections, yellow fever, Zika virus disease	Dhori, Quaranfil, Thogoto virus infections	Lebombo, Orungo, Tribeč virus infections	(Bas-Congo virus infection,) Ekpoma virus 1, Ekpoma virus 2, Mundri virus infections	Chikungunya virus disease, o'nyong-nyong fever, Semliki Forest, Sindbis virus infections
Central Asia	Bhanja, Issyk-kul virus infections, Crimean-Congo hemorrhagic fever, sandfly fever, Ťahyňa, Tamdy virus infections	Far Eastern tick-borne encephalitis, Karshi, Powassan, West Nile virus infections	Dhori virus infections	—	Isfahan virus infection	Sindbis virus infection
Eastern Asia	Crimean-Congo hemorrhagic fever, hemorrhagic fever with renal syndrome, sandfly fever, severe fever with thrombocytopenia syndrome, Tăchéng tick virus 1 (and 2) and Ťahyňa, Tamdy, Sōnglǐng, wetland, and Yezo virus infections	Apoi virus infection, dengue without/with warning signs/severe dengue, Far Eastern tick-borne encephalitis, Japanese encephalitis, Karshi virus infection, Kyasanur Forest disease	—	Banna virus infection	—	—
Southern Asia	Batai, Bhanja virus infections, Crimean-Congo hemorrhagic fever, hemorrhagic fever with renal syndrome, Nairobi sheep disease virus infection, sandfly fever	Dengue without/with warning signs/severe dengue, Japanese encephalitis, Kyasanur Forest disease, West Nile virus infection, Zika virus disease	Dhori, Quaranfil, Thogoto virus infections	—	Chandipura, Isfahan virus infections, variegated squirrel bornavirus 1?	Chikungunya virus disease
South-Eastern Asia	Batai virus infection, hemorrhagic fever with renal syndrome	Dengue without/with warning signs/severe dengue, Japanese encephalitis, West Nile virus infection, Zika virus disease	—	—	—	Chikungunya virus disease
Western Asia	Aigai, Batai, Bhanja virus infections, Crimean-Congo hemorrhagic fever, hemorrhagic fever with renal syndrome, sandfly fever, Tăchéng tick virus 1, Tamdy virus infections	Alkhurma hemorrhagic fever, Central European tick-borne encephalitis, dengue without/with warning signs/severe dengue, West Nile virus infection	Dhori, Quaranfil virus infections	—	—	Chikungunya virus disease
Latin/Central America and the Caribbean	Alenquer, Apeú virus infections, Argentinian hemorrhagic fever, Bolivian hemorrhagic fever, "Brazilian hemorrhagic fever", Bunyamwera, Candirú, Caraparú, Catú, Chagres, Chapare, Coclé, Echarate, (Flexal,) Fort Sherman, Guamá, Guaroa virus infections, hantavirus pulmonary syndrome, Iquitos, Itaquí, Juquitiba virus infections, lymphocytic choriomeningitis, Madrid, Maguari, Maldonado, Marituba, Mayaro, Morumbi, Murutucú, Nepuyo, Oriboca virus infections, Oropouche virus disease, Ossa, Punta Toro, Restan, Serra Norte, Tacaiuma, Trinidad virus infections, Venezuelan hemorrhagic fever, Wyeomyia, Xingu, Zungarococha virus infections	Bussuquara, Cacipacoré virus infections, dengue without/with warning signs/severe dengue, Ilhéus, Rio Bravo virus infections, Rocio viral encephalitis, Spondweni, St Louis encephalitis, yellow fever, Zika virus disease	—	—	Piry fever, vesicular stomatitis fever	Chikungunya virus disease, Madariaga, Mayaro, Mucambo, Tonate, Una virus infections, Venezuelan equine encephalitis

(*Continued*)

TABLE 215-2 Geographic Distribution of Zoonotic Arthropod-Borne or Rodent-Borne Viral Diseases (Continued)

AREA[a]	TYPE OF DISEASE[b]					
	BUNYAVIRICETES ***(Arenaviridae, Hantaviridae, Nairoviridae, Peribunyaviridae, Phenuiviridae)***	***FLASUVIRICETES*** ***(Flaviviridae)***	***INSTHOVIRICETES*** ***(Orthomyxoviridae)***	***RESENTOVIRICETES*** ***(Sedoreoviridae, Spinareoviridae)***	***MONJIVIRICETES*** ***(Bornaviridae, Rhabdoviridae)***	***ALSUVIRICETES*** ***(Togaviridae)***
Northern America	(Avalon,) Cache Valley virus infections, California encephalitis, hantavirus pulmonary syndrome, Heartland and Keystone virus infections, lymphocytic choriomeningitis, Nepuyo, snowshoe hare, (Whitewater Arroyo virus) infections	Dengue without/with warning signs/severe dengue, Modoc virus infection, Powassan virus disease, Rio Bravo virus infection, St Louis encephalitis, West Nile virus infection, Zika virus disease	Bourbon virus infection	Colorado tick fever, Salmon River virus infection	Vesicular stomatitis fever	Eastern equine encephalitis, Everglades virus infection, western equine encephalitis
Europe	Adria, Aigai, (Avalon,) Batai, Bhanja, Cristoli virus infections, California encephalitis, Crimean-Congo hemorrhagic fever, (Erve virus infection), hemorrhagic fever with renal syndrome, Inkoo, Issyk-kul virus infections, lymphocytic choriomeningitis, sandfly fever, snowshoe hare, Ťahyňa, Uukuniemi virus infections	Central European tick-borne encephalitis, dengue without/with warning signs/severe dengue, Ntaya virus infection, Omsk hemorrhagic fever, Powassan, Usutu, West Nile virus infections	Dhori, Thogoto virus infections	Eyach, Kemerovo, Tribeč virus infections	—	Chikungunya virus disease, Sindbis virus infection
Oceania	Batai, Gan Gan, (Trubanaman) virus infections	Dengue without/ with warning signs/ severe dengue, Edge Hill virus infection, Japanese encephalitis, Kokobera virus infection, Murray Valley encephalitis, Sepik virus, West Nile virus infections, Zika virus disease	—	—	—	Barmah Forest virus infection, Ross River disease, Sindbis virus infection

[a] Geographic names here and throughout the chapter are as recommended by the UN geoscheme (*https://unstats.un.org/unsd/methodology/m49/*). [b] Disease names according to the World Health Organization's International Classification of Diseases 11th revision (ICD-11; *https://icd.who.int/browse11/l-m/en*). Quotation marks indicate common usage in the absence of ICD-11 recognition. Diseases not acknowledged by the ICD-11 are designated as "virus infection(s)."

is best known as a cause of VHF, but the prevalence of milder F&M syndrome is far higher after infection, with encephalitis and ocular disease also occurring occasionally. Lymphocytic choriomeningitis virus (LCMV) is classified in this chapter as a cause of F&M syndrome because it is the most common disease manifestation. Even when central nervous system (CNS) disease evolves during infection with this virus, neurologic manifestations are usually mild and preceded by F&M. However, this virus may also cause fetal microcephaly. Overlap between syndromic categories is further complicated by evolving nomenclature around their classification. For example, infection with any dengue virus (DENV-1–4) is considered as a cause of F&M syndrome because this disease presentation, historically called "dengue fever," is by far the most common manifestation. However, severe manifestations of DENV infection have a complicated pathogenesis: the historical classification of disease as "dengue hemorrhagic fever" included a subset of patients with "dengue shock syndrome," which is of tremendous consequence for pediatric populations in particular areas of the world. Further complicating this overlap, a relatively recent World Health Organization (WHO) revision of disease classification recommended a less descriptive but more pragmatic use of "dengue without warning signs," "dengue with warning signs," and "severe dengue" to describe the same spectrum and enhance clinical management and case reporting.

Unfortunately, most of the known arthropod-borne and rodent-borne viral diseases have not been studied in detail with modern medical approaches. Thus, available data may be incomplete, at relatively low resolution, or biased toward severe disease. Data on geographic distribution are often difficult to interpret. Frequently, the literature is not clear as to whether the data pertain to the distribution of a particular virus or to the areas where human disease has been observed. In addition, the designations for viruses and viral diseases have changed multiple times over decades. In this chapter, virus and taxon names are in line with the latest reports of the International Committee on Taxonomy of Viruses, and disease names are in accordance with the WHO's International Classification of Diseases 11th revision (ICD-11). When needed for clarity or historical reference, alternative nomenclature is used. Considering this syndromic approach, it should be noted that the variable clinical manifestations of particular viruses may be captured over a number of sections.

■ FEVER AND MYALGIA (F&M)

F&M is by far the most common clinical syndrome and has the most favorable outcomes after arthropod-borne and rodent-borne infections. However, it is also a common prodrome en route to more invasive syndromes associated with these infections. In logical sequence, therefore, F&M will be discussed initially but should not be considered in isolation. Indeed, clinicians evaluating patients with F&M should pay careful attention for symptoms and signs that suggest the A&R, neuroinvasive, pulmonary, or VHF syndromes that are subsequently discussed in this chapter. Although many of the viruses listed in Table 215-1 probably cause F&M, only some of these viruses have prominent associations with the syndrome that are considered biomedically important.

F&M syndrome typically begins with the abrupt onset of fever, chills, intense myalgia, and malaise. Patients may also report joint pain, but true arthritis is not found. Anorexia is characteristic and may be accompanied by nausea or vomiting. Headache is common and may be severe, with photophobia and retroorbital pain. Physical findings are minimal and usually confined to conjunctival injection, pharyngeal erythema/exudate, muscle tenderness, abdominal tenderness,

TABLE 215-3 Clinical Syndromes Caused by Zoonotic Arthropod-Borne or Rodent-Borne Viruses

SYNDROME	VIRUS
Fever and myalgia (F&M)	*Arenaviridae*: (Flexal,) Lassa and lymphocytic choriomeningitis viruses *Flaviviridae*: Bussuquara, Banzi, Cacipacoré, Dakar bat, dengue 1–4, Edge Hill, Karshi, Kédougou, Koutango, Ntaya virus, Rio Bravo, Sepik, Spondweni, tick-borne encephalitis, Wesselsbron, and Zika viruses *Hantaviridae*: Choclo virus *Nairoviridae*: Dugbe, Issyk-kul, Nairobi sheep disease, Sōnglǐng viruses, Tǎchéng tick virus 1, Tamdy, wetland, Yezo virus *Orthomyxoviridae*: Bourbon, Dhori, and Thogoto viruses *Peribunyaviridae*: Apeú, Bangui, Batai, Bunyamwera, Bwamba, Cache Valley, California encephalitis, Caraparú, Catú, Fort Sherman, Germiston, Guamá, Guaroa, Ilesha, Inkoo, Iquitos, Itaquí, Jamestown Canyon, La Crosse, Lumbo, Madrid, Maguari, Marituba, Nepuyo, Ngari, Nyando, Oriboca, Oropouche, Ossa, Pongola, Restan, Shokwe, snowshoe hare, Tacaiuma, Ťahyňa, Tataguine, Wyeomyia, Xingu, and Zungarococha viruses *Phenuiviridae*: Alenquer, Bhanja, Candirú, Chagres, Dar es Salaam, Echarate, Heartland, Maldonado, Morumbi, Punta Toro, Rift Valley fever, sandfly fever Cyprus, sandfly fever Ethiopia, sandfly fever Naples, sandfly fever Sicilian, sandfly fever Turkey, Serra Norte, severe fever with thrombocytopenia syndrome, Toscana, and Uukuniemi viruses *Sedoreoviridae*: Kemerovo, Lebombo, Orungo, and Tribeč viruses *Spinareoviridae*: Colorado tick fever, Eyach, and Salmon River viruses *Rhabdoviridae*: Ekpoma virus 1, Ekpoma virus 2, Chandipura, Isfahan, Piry, vesicular stomatitis Indiana, and vesicular stomatitis New Jersey viruses *Togaviridae*: Everglades, Madariaga, Mucambo, Tonate, Una, and Venezuelan equine encephalitis viruses
Arthritis and rash (A&R)	*Flaviviridae*: Kokobera and Zika viruses *Peribunyaviridae*: Gan Gan virus, (Trubanaman virus) *Togaviridae*: Barmah Forest, chikungunya, Mayaro, o'nyong-nyong, Ross River, Semliki Forest, and Sindbis viruses
Encephalitis	*Arenaviridae*: lymphocytic choriomeningitis virus, (Whitewater Arroyo virus) *Flaviviridae*: Apoi, Ilhéus, Japanese encephalitis, Karshi, Modoc, Murray Valley encephalitis, Powassan, Rocio, St. Louis encephalitis, tick-borne encephalitis, Usutu, and West Nile viruses *Orthomyxoviridae*: Dhori and Thogoto viruses *Bornaviridae*: variegated squirrel bornavirus 1 *Peribunyaviridae:* Cache Valley, California encephalitis, Cristoli, Inkoo, Jamestown Canyon, Keystone, La Crosse, Lumbo, Oropouche, snowshoe hare, Shuni, and Ťahyňa viruses *Phenuiviridae*: Adria, Bhanja, Chios, Rift Valley fever, (Tǎchéng tick virus 2), and Toscana viruses *Sedoreoviridae*: Banna, Kemerovo, and Orungo viruses *Spinareoviridae*: Colorado tick fever, Eyach, and Salmon River viruses *Rhabdoviridae*: Chandipura, Mundri viruses *Togaviridae*: eastern equine encephalitis, Everglades, Madariaga, Mucambo, Tonate, Venezuelan equine encephalitis, and western equine encephalitis viruses
Pulmonary disease	*Hantaviridae*: Anajatuba, Andes, Araucária, bayou, Bermejo, Black Creek Canal, Blue River, Caño Delgadito, Castelo dos Sonhos, Catacamas, Choclo, Juquitiba, Laguna Negra, Lechiguanas, Maciel, Monongahela, New York, Orán, Paranoá, Pergamino, Puumala, Rio Mamoré, Sin Nombre, (Tula,) and Tunari viruses
Viral hemorrhagic fever (VHF)	*Arenaviridae*: Chapare, Guanarito, Junín, Lassa, Lujo, (lymphocytic choriomeningitis,) Machupo, and Sabiá viruses *Hantaviridae*: Amur, Dobrava, El Moro Canyon, gōu, Hantaan, Kurkino, Muju, Puumala, Saaremaa, Seoul, Sochi, and Tula viruses *Nairoviridae*: Crimean-Congo hemorrhagic fever virus *Peribunyaviridae*: (Ilesha virus,) Ngari viruses *Phenuiviridae*: Rift Valley fever and severe fever with thrombocytopenia syndrome viruses *Flaviviridae*: Alkhurma hemorrhagic fever, dengue 1–4, Kyasanur Forest disease, Omsk hemorrhagic fever, (tick-borne encephalitis,) and yellow fever viruses *Rhabdoviridae*: (Bas-Congo virus)

Viruses are placed in parentheses if cases were controversial.

and possibly the presence of a nonpruritic maculopapular rash that may have a petechial component. The spectrum of disease varies from subclinical to temporarily incapacitating. When present, the evolution and duration of symptoms/signs are quite variable (generally 2–5 days) and may be biphasic after some viral infections. Less-common findings include epistaxis (not necessarily indicating a bleeding diathesis) and signs of aseptic meningitis. Even in the presence of headache, photophobia, and meningismus, challenges in obtaining and examining the CSF in remote areas makes diagnosis difficult. Although upper and lower respiratory symptoms/signs and radiographic evidence of pulmonary infiltrates are noted in some patients, the agents causing this syndrome are not primary respiratory pathogens.

F&M syndrome is the most nonspecific of the disease syndromes caused by arboviral infections. The early stages of other syndromes discussed in this chapter begin similarly and are encompassed in a broad differential diagnosis that also includes community-acquired parasitic infections (e.g., malaria), bacterial infections (e.g., anicteric leptospirosis and rickettsial diseases), and other viral infections. F&M syndrome is often described as "influenza-like," but the usual absence of cough and coryza makes influenza an unlikely confounder except at the earliest stages. Treatment is supportive, but acetylsalicylic acid is generally avoided because of the potential for exacerbated bleeding or Reye syndrome. Complete recovery is the expected outcome for symptomatic patients, although prolonged asthenia and nonspecific symptoms have been described, particularly after infection with LCMV or DENV-1–4. Nonetheless, it must be reemphasized that this nonspecific syndrome may be the prodrome for arthritic, neuroinvasive, pulmonary, or VHF syndromes with less favorable outcomes.

Efforts to prevent viral infections causing this syndrome are best targeted to vector control, which, however, may be expensive or impossible. Destruction of mosquito breeding sites is generally the most economically and environmentally sound approach. Emerging containment technologies include the release of genetically modified mosquitoes and the spread of *Wolbachia* bacteria to limit mosquito multiplication rates. Depending on the vector and its habits, other possible strategies include the use of window screens on dwellings or other barriers (e.g., permethrin-impregnated bed nets), judicious application of arthropod repellents (e.g., *N*,*N*,-diethyltoluamide [DEET]) to the skin, use of long-sleeved (ideally permethrin-impregnated) clothing, and avoidance of vector habitats, particularly at peak feeding times.

Bunyaviricetes F&M syndrome is caused by numerous bunyaviricetes, many of which cause isolated individual infections and usually do not cause epidemics. These viruses include arenavirids (mammarenaviruses), hantavirids (orthohantaviruses), nairovirids (orthonairoviruses), peribunyavirids (orthobunyaviruses), and phenuivirids (bandaviruses and phleboviruses).

ARENAVIRIDS Infection with LCMV is the only human mammarenavirus infection resulting predominantly in F&M syndrome. LCMV is transmitted to humans from the common house mouse (*Mus musculus*) by aerosols of excreta or secreta. The virus is maintained in the mouse mainly by vertical transmission from infected dams. Infected mice remain viremic and shed virus for life, with high concentrations of virus in all tissues. Mouse-acquired infections of colonies of pet hamsters also can serve as a link for human acquisition. Patients may have a history of residence in rodent-infested housing or other exposure to rodents. An antibody prevalence of ≈5–10% has been reported among urban adults from Argentina, Germany, and the United States. In addition, laboratory infections among scientists and animal caretakers can occur because the virus is widely used in immunology laboratories to study T-lymphocyte function and can silently infect cell cultures and passaged tumor lines. Transmission has occurred via organ transplantation.

Lymphocytic choriomeningitis (LCM) differs from the general F&M syndrome in a characteristic gradual onset of the otherwise typical illness. Orchitis, transient alopecia, arthritis, pharyngitis, cough, and maculopapular rash are associated with LCM. An estimated one fourth (or less) of patients experience an initial febrile phase of 3–6 days. After a brief remission, many develop recurrent fever accompanied by meningeal (severe headache, photophobia, nausea and vomiting, and neck stiffness) and/or less common encephalitic signs (altered mental status or level of consciousness, sensorimotor deficits) as part of a ≈1-week CNS phase. Patients with neuroinvasive disease, including patients with clear-cut signs of encephalitis, almost always recover fully. Rarely, disease is complicated by transient hydrocephalus (that may require shunting) and myelitis. During the initial febrile phase, leukopenia and thrombocytopenia are common, and virus can usually be isolated from blood. During the CNS phase, the virus may be found in the CSF, and antibodies are detected in the blood. The pathogenesis of LCM is thought to resemble manifestations resulting from direct intracranial inoculation of the virus into adult mice. The onset of the immune response leads to T-cell-mediated immunopathologic meningitis. During the meningeal phase, CSF monocyte counts range from the hundreds to the low thousands per microliter, and, unusually for a viral meningitis, hypoglycorrhachia is found in one-third of patients.

IgM-capture ELISA, immunochemistry, and RT-PCR are used in the diagnosis of lymphocytic choriomeningitis. IgM-capture ELISA of serum and CSF usually yields positive results; RT-PCR assays have been developed for CSF detection. Particular diagnostic challenges arise in immunosuppressed patients with fulminant infections transmitted by recent organ transplantation: in the absence of typical immune responses, molecular diagnosis via RT-PCR or immunohistochemistry may be required. Infection should be suspected in acutely ill febrile patients with marked leukopenia and thrombocytopenia and meningitis syndromes. In patients with aseptic meningitis, a diagnosis of LCM is suggested by the following: a well-marked febrile prodrome, adult age, occurrence in the autumn, low CSF glucose levels, or CSF monocyte counts of >1,000/μL. Throughout a pregnancy, maternal-to-fetal transmission may occur, leading to fetal death (in the first trimester) or to consequent congenital hydrocephalus, microcephaly, and/or chorioretinitis (in the second or third trimesters). Because a mild maternal infection may be unrecognized or unrecalled, specific antibodies to the virus should be sought in both mother and fetus under suspicious circumstances, particularly in neonatal hydrocephalus testing negative for toxoplasmosis, rubella, cytomegalovirus, herpes simplex, and HIV-1/HIV-2 (TORCH) pathogens.

PERIBUNYAVIRIDS Apeú, Caraparú, Itaquí, Madrid, Marituba, Murutucú, Nepuyo, Oriboca, Ossa, Restan, and Zungarococha viruses are among the most common causes of arboviral infection in humans entering South American jungles. These viruses cause acute febrile disease and are transmitted by mosquitoes in neotropical forests. Oropouche virus is transmitted in urban settings in Central and South America by biting midges (*Culicoides paraensis*), which often breed to high density in cacao husks and other vegetable detritus found in towns and cities. Explosive epidemics involving thousands of patients with Oropouche virus disease have been reported in Brazil and Peru, with smaller outbreaks in Panama and French Guiana and most recent outbreaks in Cuba and Haiti. Direct human-to-human transmission does not occur. Sloths and nonhuman primates may play a role as vertebrate hosts during maintenance sylvatic cycles. After a 3- to 10-day incubation period, most patients develop typical F&M syndrome with a biphasic evolution (recurrence of symptoms in the second week of illness) and, uncommonly, hemorrhagic manifestations. However, a small percentage of patients develop severe illness with neurologic (meningoencephalitis, Guillain-Barré) or hemorrhagic syndromes that have been associated with fatal outcome. Recently, infection of pregnant women has been associated with spontaneous miscarriage, intrauterine fetal death, and congenital anomalies. Oropouche virus has been detected in umbilical cord blood and fetal tissues in these clinical settings. Serologic detection (IgM or a rise in antibody titer on paired samples) or PCR confirms the diagnosis of acute F&M syndrome; IgM and viral nucleic acid have been detected in the CSF in rare cases of neuroinvasive disease. Postacute sequelae have not been described, though some patients have prolonged asthenia after the acute syndrome. Specific therapeutics or vaccines are not available.

Iquitos virus, a recently discovered reassortant and close relative of Oropouche virus, causes disease that is easily mistaken for Oropouche virus disease; its overall epidemiologic significance remains to be determined.

PHENUIVIRIDS The genus *Phlebovirus* includes numerous viruses that may cause human infection. Sandfly fever Cyprus virus, sandfly fever Ethiopia virus, sandfly fever Sicilian virus, and sandfly fever Turkey virus (and the encephalitis-causing Chios virus) are genetically and antigenically very closely related. In contrast, sandfly fever Naples virus (SFNV) is only genetically and antigenically distantly related to these viruses. SFNV has not been detected in sandflies, humans, or nonhuman vertebrates since the 1980s and therefore may be extinct, but closely related viruses, such as Granada virus and Toscana virus, continue to circulate. Toscana virus is thus far the only *Phlebovirus* transmitted by sandflies that is known to cause central and peripheral nervous system disease, such as encephalitis, meningitis, or polymyeloradiculopathy. *Phlebotomus* sandflies transmit the virus, probably by biting small mammals and humans. Female sandflies may be infected orally during blood meals and may transmit the virus to progeny when they lay their eggs. This prominent transovarial transmission confounds virus control.

Sandfly fever is found in the circum-Mediterranean area, extending to the east through the Balkans into parts of China as well as into Western Asia. Sandflies are found in both rural and urban settings and are known for their short flight ranges and small sizes; the latter enables them to penetrate standard mosquito screens and netting. Epidemics have been described in the wake of natural disasters and wars. After World War II, extensive spraying in parts of Europe to control malaria greatly reduced sandfly populations and transmission of SFNV; the incidence of sandfly fever continues to be low.

A common pattern of disease in endemic areas consists of high attack rates among visitors (including military personnel) but little or no disease in the local population, having developed protective immunity after childhood infection. Toscana virus infection is common during the summer among rural residents and vacationers, particularly in

Italy, Spain, and Portugal; several cases have been identified in travelers returning to Germany and Scandinavia. Clinical presentation is typical for F&M syndrome, with onset 3–6 days after a sandfly bite (a papule at the bite site may be observed). Notably, outbreaks of aseptic meningitis (virus can be isolated from the CSF) or meningoencephalitis have been associated with Toscana virus infection.

Coclé and Punta Toro viruses are not directly related to sandfly fever viruses but, like them, are transmitted by sandflies and cause a sandfly-fever-like disease. These two viruses are maintained in Latin American and Caribbean tropical forests, respectively, where the vectors rest on tree buttresses. Epidemics have not been reported, but antibody prevalence among inhabitants of villages in endemic areas indicates a cumulative lifetime exposure rate of >50% in the case of Punta Toro virus.

Heartland virus is notable as an emerging virus in the eastern, midwestern, and southern United States (>60 cases since its first description in 2012), where it causes often-fatal infection among males >50 years of age with comorbidities.

Flavivirids The most clinically significant orthoflaviviruses that cause F&M syndrome are DENV-1–4 and Zika virus (ZIKV). In fact, dengue without or with warning signs ("dengue," historically called "dengue fever"—to be distinguished from severe dengue) is probably the most prevalent arthropod-borne viral disease worldwide, with ≈400 million infections occurring per year, of which ≈100 million (25%) cause clinical illness. Dengue is endemic in >100 countries worldwide, including those in Africa, the Americas, the eastern Mediterranean, Southeastern Asia, and the western Pacific. More than half of the world's population is considered at risk, although Asia bears 70% of the global burden, with alarming increases over the past decade including, for example, >400,000 cases in 2019 in the Philippines. Year-round transmission of DENV-1–4 occurs between latitudes 25°N and 25°S, but seasonal forays of the viruses into the United States and Europe have been documented. The principal vectors for all four viruses are YF mosquitoes (*Aedes aegypti*). Through increasing spread of mosquitoes throughout the tropics and subtropics and international travel by infected humans, large areas of the world have become vulnerable to the introduction of DENV-1–4. Thus, dengue and severe dengue (see "Viral Hemorrhagic Fever,") are becoming increasingly common. For instance, conditions favorable to DENV-1–4 transmission via YF mosquitoes exist in Hawaii and the southern United States. The range of a lesser vector of DENV-1–4, the Asian tiger mosquito (*Aedes albopictus*), now extends from Asia to the continental United States, the Indian Ocean, parts of Europe, and Hawaii. Also anthropophilic, YF mosquitoes typically breed near human habitation, using relatively fresh water in locations such as jars, vases, discarded containers, coconut husks, and old tires. These mosquitoes usually bite during the day. Bursts of dengue and severe dengue cases are to be expected in the southern United States, particularly along the Mexican border, where containers of water may be infested with YF mosquitoes. Air-conditioned buildings with screened vents may inhibit transmission of many arboviruses, including DENV-1–4.

Most primary DENV infections are subclinical. After an incubation period of ≈4–7 days, symptomatic patients present with three evolving phases: febrile, critical, and recovery. Although most patients presenting with F&M syndrome do not go through a critical phase, early recognition of the critical phase consistent with severe dengue must be considered in all patients. In most patients, dengue begins with the typical sudden onset of high-grade fever, frontal headache, retroorbital pain, back pain, and severe myalgia. These symptoms and signs gave rise to the colloquial designation of dengue as "break-bone fever." A transient macular rash is often present at illness onset, and conjunctival redness, pharyngeal erythema, lymphadenopathy, and hepatomegaly may be noted on physical examination. The illness may last a week and include additional symptoms and signs (anorexia, nausea or vomiting, and marked cutaneous hypersensitivity). Near the time of defervescence (days 3–5), a maculopapular rash begins on the trunk and spreads to the extremities and the face. Epistaxis and scattered petechiae are often noted in uncomplicated dengue (without/with warning signs), and preexisting gastrointestinal lesions may bleed during the acute illness. A positive tourniquet test—i.e., the detection of 10 or more new petechiae in one square inch of the upper arm after a 5-min blood pressure cuff inflation to midway between systolic and diastolic pressure—may demonstrate microvascular damage, but this finding is more likely to be associated with severe dengue. After defervescence and a brief remission, a subset of patients has recurrence of fever and other signs in a "saddle-back" pattern. Regardless, most patients recover from acute illness by 7–10 days, though convalescence can be prolonged.

Laboratory findings of dengue (without/with warning signs) include leukopenia, thrombocytopenia, and, in many cases, modest elevations of serum aspartate aminotransferase (AST) activity without hepatic synthetic dysfunction. The diagnosis is made by antigen-detection ELISA or RT-PCR during the acute phase or by IgM ELISA or paired serology during recovery. Virus is readily isolated from blood in the acute phase if mosquito inoculation or mosquito cell culture is used. Further clinical management and vaccine prevention are discussed in reference to severe dengue.

ZIKV is an emerging pathogen that is transmitted to nonhuman primates and humans by *Aedes* mosquitoes. The virus was discovered 1947 in a sentinel rhesus monkey (*Macaca mulatta*) and *Aedes africanus* mosquitoes in the Zika Forest in what was then the British Protectorate of Uganda. Human ZIKV infection was first documented during a YF outbreak in 1954 in Nigeria. Later, ZIKV infections were recognized in Southeastern Asia and Southern Asia. Prior to 2007, only 14 clinically identified cases of Zika virus disease had been reported. In recent years, the number of reported ZIKV infections has increased steadily and rapidly, with large but generally mild disease outbreaks on Yap Island and Micronesia (2007) and in Cambodia (2010), the Philippines (2012), and French Polynesia (2013–2014). ZIKV disease in the Americas was first reported on Easter Island, Chile (2014), and in Brazil (2015). By the end of 2015, an estimated 440,000 to 1.3 million cases had occurred in Brazil. At the end of May 2017, ZIKV infections had been recorded on five continents in 85 countries, including Mexico and the United States. Beginning in 2018, the global activity of ZIKV declined rather rapidly for unknown reasons.

Phylogenetic analysis of all available African ZIKV isolates revealed two geographically overlapping clades (Eastern and Western Africa). A descendant Asian lineage, represented by viruses collected from mosquitoes trapped in homes in Malaysia, was first reported in 1969. All ZIKV isolates causing human cases outside of Africa trace back to this Asian lineage.

Human infections are usually asymptomatic or benign and self-resolving and are most commonly misdiagnosed as dengue without/with warning signs or influenza. Typically, ZIKV disease is characterized by low-grade fever, an itchy maculopapular rash, arthralgia/myalgia, nonpurulent conjunctivitis, headache, and malaise. Other clinical manifestations include involvement of the gastrointestinal (nausea, vomiting, abdominal pain, and/or diarrhea), genitourinary (hematospermia), and, less commonly, cardiac (myocarditis and/or pericarditis), ocular (uveitis), and auditory (hearing loss) systems. The most concerning complications of ZIKV infection are neurologic syndromes, such as, Guillain-Barré syndrome and congenital fetal microcephaly, which may be associated either with fetal death or serious developmental delay in newborn infants. Other neurologic complications include encephalitis, meningoencephalitis, transverse myelitis, demyelinating polyneuropathies, cerebrovascular ischemia, retinopathies, and neurologic birth defects. Although most human ZIKV infections are acquired after bites by infected female mosquitoes, transmission may also occur perinatally or via breast-feeding, sexual contact with an infected person, transfusion of blood products, or organ transplantation. Infectious ZIKV or ZIKV RNA has been documented in the semen of male patients for up to 69 and 188 days after illness onset, respectively; sexual transmission associated with persistence has been documented up to 41 days after illness onset. Antiviral treatments (curative or preventive) and licensed vaccines against ZIKV are not yet available. Of note, some evidence suggests that prior exposure to ZIKV infection may increase the risk of severe disease syndrome and poor outcomes from DENV infection.

Orthomyxovirids Bourbon virus is emerging as a rare cause of F&M in humans in at least 14 states of the United States (≈60 cases since discovery in 2015). Onset is typical, but severe disease progression occurs among older individuals with comorbidities, including multisystem organ (kidney, respiratory, and/or hemodynamic) failure that has been fatal.

Sedoreovirids Several orbiviruses (Lebombo, Kemerovo, Orungo, and Tribeč viruses) can cause F&M in humans. Lebombo and Orungo viruses are transmitted by mosquitoes, whereas Kemerovo and Tribeč viruses are transmitted by ticks.

Spinareovirids Several coltiviruses (Colorado tick fever, Eyach, and Salmon River viruses) can cause F&M in humans. All are transmitted by ticks. The most significant spinareovirid arthropod-borne disease is Colorado tick fever. Several hundred patients with this disease are reported annually in the United States and Canada. The infection is acquired between March and November through the bite of Rocky Mountain wood ticks (*Dermacentor andersoni*) in mountainous western regions at altitudes of 1200–3000 m. Small mammals serve as amplifying hosts. After a tick exposure (reported by almost all patients) and a mean incubation period of 1–14 days, the most common presentation is F&M, often with headache; rash develops in a minority of patients. Meningoencephalitis is not uncommon, especially in children, and disseminated intravascular coagulation (DIC) with hemorrhagic manifestations, pericarditis, myocarditis, orchitis, and pulmonary disease have also been reported. Leukopenia and thrombocytopenia are noted. The disease usually lasts 7–10 days and is often biphasic. Historically, the most important differential diagnostic considerations have been Rocky Mountain spotted fever (although Colorado tick fever is much more common in Colorado) and tularemia. Notably, seropositivity may be delayed for 10–14 days, limiting the utility of typical serologic testing early in the disease course; molecular RT-PCR testing may be used in this setting. Also, Colorado tick fever virus replicates for several weeks in erythropoietic cells and can be found in erythrocytes. This feature, detected in erythroid smears stained by immunofluorescence, can be diagnostically helpful and is important during screening of blood donors.

■ ARTHRITIS AND RASH (A&R)

Arthritides are common clinical presentations (or manifestations) of several viral diseases, such as hepatitis B, hepatitis C, parvovirus B19 infection, and rubella, and occasionally accompany infection due to adenovirids, enteroviruses, herpesvirids, mumps virus, or HIV-1/HIV-2. Arthropod-borne alphaviruses are also common causes of arthritides—usually acute febrile syndromes with joint involvement often accompanied by a maculopapular rash. Rheumatic involvement includes arthralgia with or without typical inflammatory signs, including periarticular redness, swelling (less commonly, joint effusions), and immobility. Most alphavirus infections are less severe and have fewer articular manifestations in children than in adults. In temperate climates, these ailments are summer diseases. The most significant alphavirus causes of arthritides are chikungunya virus disease, Ross River disease, Barmah Forest virus infection, and Sindbis virus infection. ZIKV infections may also be associated with joint manifestations. Less significant but historically notable are viruses that caused isolated cases or epidemics. A large (>2 million cases), albeit isolated, epidemic was caused by o'nyong-nyong virus from 1959 to 1962 (o'nyong-nyong fever). Mayaro virus, Semliki Forest virus, and Una virus caused isolated cases or limited and infrequent outbreaks (30 to several hundred cases per year). Symptoms and signs of infections with these viruses often are similar to those observed with chikungunya virus disease. Two orthobunyaviruses—Gan Gan virus and Trubanaman virus—and the orthoflavivirus Kokobera virus have been associated with single cases of polyarthritic disease. Except for a vaccine to prevent chikungunya virus disease, no specific therapies or licensed vaccines exist.

Chikungunya Virus Disease Historically, chikungunya virus (CHIKV) was considered endemic in rural areas of Africa, with intermittent outbreaks occurring in towns and cities of both Africa and Asia. In 2004, a large epidemic began in the Indian Ocean region (specifically on the islands of Réunion and Mauritius) and was most likely spread by travelers. YF mosquitoes and Asian tiger mosquitoes are the major CHIKV vectors. All regions with established populations of these two vectors have now documented local mosquito-borne transmission, and the virus has now been identified across Africa, Asia, Europe, and the Americas, including in the continental United States, where suitable vector mosquitoes are present in southern states and local transmission has been documented. From December 2013 to June 2023, >3.6 million cases were reported from 50 countries or territories in the Americas.

The disease is more common in adults. After an incubation period of 2–10 days, the abrupt onset of fever (often severe, with a saddleback pattern) and severe arthralgia are accompanied by constitutional symptoms and signs, including chills, abdominal pain, anorexia, conjunctival injection, headache, nausea, and photophobia. Migratory polyarthritis mainly affects the small joints of the ankles, feet, hands, and wrists, but the larger joints may be involved. Rash may appear at the outset or several days into the illness; its development often coincides with defervescence around day 2 or 3 of clinical illness. The rash is most intense on the trunk and limbs and may desquamate. In hospitalized patients with severe acute disease, uncommon but reported manifestations include cardiorespiratory, neurologic, renal, neurologic, and ocular complications; deaths have been reported, typically in the context of large outbreaks. Young children develop less prominent symptoms and signs and are therefore less frequently hospitalized. Children also often develop a bullous rather than a maculopapular/petechial rash. Although pregnant women are not at risk of more severe disease, there is a high risk of maternal–fetal transmission, particularly during the intrapartum period, that, in some cases, has led to fetal death. Recovery may require weeks, and a significant portion of middle-aged to older patients develop chronic arthritis or arthralgia syndromes (typically involving the same joints) that may be disabling. The risk of chronic musculoskeletal syndromes is increased in patients with preexisting osteoarthritis or severe acute disease who test positive for the human leukocyte antigen B27 subtype (HLA-B27) and may be associated with cryoglobulins. Although petechiae and epistaxis are occasionally seen, CHIKV should not be considered as a cause of VHF. Laboratory abnormalities may include transient lymphopenia and mild thrombocytopenia, as well as elevated activities of AST and concentrations of C-reactive protein (CRP). Treatment of chikungunya virus disease relies on acetaminophen and nonsteroidal anti-inflammatory drugs (NSAIDs). For patients with refractory arthritis, glucocorticoids or disease-modifying antirheumatic drugs (DMARDs) may be considered. A live-attenuated vaccine to prevent chikungunya virus disease has been approved by the U.S. Food and Drug Administration (FDA) and is undergoing postmarketing evaluation. Other vaccine and therapeutic candidates in preclinical development may benefit from the designation of CHIKV as the representative pathogen for arthritic alphaviruses.

Ross River Disease and Barmah Forest Virus Infection Ross River virus (RRV) and Barmah Forest virus (BFV) cause diseases that are clinically indistinguishable (hence the previously common disease designation of "epidemic polyarthritis" for both infections). RRV has caused epidemics in Australia, Papua New Guinea, and the South Pacific since the beginning of the 20th century. In 1979 and 1980, the virus swept through the Pacific Islands, causing >500,000 infections. From 1991 to 2011, the virus caused 92,559 infections in rural and suburban areas of Australia. From 2014 to 2015, >10,000 cases were recorded in Australia. RRV is predominantly transmitted by *Aedes normanensis*, *Aedes vigilax*, and *Culex annulirostris* mosquitoes. Marsupials (e.g., kangaroos, wallabies, and koalas) and rodents are probably the main vertebrate hosts, but infection also occurs in horses and other livestock. BFV infections have been increasingly documented since the early 1990s. For instance, from 1991 to 2011, 21,815 cases were recorded in Australia, and new data indicate that the disease also occurs in Papua New Guinea. BFV is transmitted by both *Aedes* and *Culex* mosquitoes and has been isolated from biting midges. The vertebrate hosts remain to be determined, but serologic studies implicate horses and possums.

Of the human RRV and BFV infections surveyed, 55–75% were asymptomatic; however, clinical illness can be debilitating when it occurs. After a typical incubation period of 3–9 days, patients develop the sudden onset of symmetric joint pain predominantly involving the peripheral extremities. Generally, a nonitchy, diffuse, maculopapular rash (more common in BFV infection) develops coincidentally or follows shortly; but, in some patients, rash can precede joint pain by several days. Constitutional symptoms (e.g., low-grade fever, asthenia, headache, myalgia, and nausea) are not prominent or are absent in many cases. Most patients are incapacitated for considerable periods (6 months or more) by joint involvement that interferes with grasping, sleeping, and walking. Ankle, interphalangeal, knee, metacarpophalangeal, and wrist joints are most often involved, although elbows, shoulders, and toes may also be affected. Periarticular swelling and tenosynovitis are common, and one third of patients have true arthritis (more common in Ross River disease). Myalgia and nuchal stiffness may accompany joint pain. Only half of all patients with arthritis can resume normal activities within 4 weeks, and 10% continue to limit their activities after 3 months. Occasionally, patients are symptomatic for >1 year but without progressive arthropathy.

In the diagnosis of either infection, clinical laboratory values are normal or variable. Tests for rheumatoid factor and antinuclear antibodies are negative, and the erythrocyte sedimentation rate is acutely elevated. Joint fluid contains 1000–60,000 monocytes per microliter, and viral antigen can usually be detected in macrophages. Serologic demonstration of rising antibody titers is the cornerstone of diagnosis, though it should be noted that virus-specific IgM antibodies may persist for month to years. Isolation of the virus from blood after mosquito inoculation or growth of the virus in cell culture is possible early in the illness. Because of the great economic impact of annual epidemics in Australia, an inactivated RRV vaccine has been under advanced development; phase 3 trials were completed in 2015 with promising results, but the candidate vaccine has not yet been developed for the market. Analgesics or NSAIDS are used for symptomatic treatment.

Sindbis Virus Infection Sindbis virus is typically transmitted to birds primarily by infected *Culex* and *Culiseta* mosquitoes that also vector human transmission over a range that includes Africa, Asia, Europe, and Australia. Infections with northern European or southern African variants are particularly likely in rural environments. Although infections may be subclinical, rash and arthralgia typically develop after an incubation period of <1 week. Constitutional clinical signs are not marked, and fever is modest or absent. The rash, which lasts ≈1 week, begins on the trunk, spreads to the extremities, and evolves from macules to papules that often vesiculate. The arthritis is polyarticular, migratory, and incapacitating, with resolution of the acute phase in a few days. The ankles, elbows, knees, phalangeal joints, wrists, and—to a much lesser extent—proximal and axial joints are involved. Persistence of joint pain and occasionally arthritis may continue for months or even years despite lack of deformities.

■ ENCEPHALITIS

The major encephalitis viruses are found in the families *Flaviviridae, Peribunyaviridae, Rhabdoviridae,* and *Togaviridae*. However, individual agents of other families, including Dhori virus and Thogoto virus (*Orthomyxoviridae*) and Banna virus (*Sedoreoviridae*), have caused isolated cases of encephalitis. Arboviral encephalitides are seasonal diseases, commonly occurring in the warmer months. Their incidence varies markedly with time and place, depending on ecologic factors that determine vector activity and human exposure. The causative viruses differ substantially in terms of the ratio of clinical to subclinical infections, case-fatality rate (CFR), and postacute sequelae. Humans are not considered important amplifiers of these viruses.

All the viral encephalitides discussed in this section share a similar pathogenesis. An infected arthropod ingests blood from a human and, through this contact, initiates infection. Initial viremia, thought to originate from the lymphoid system, leads to multifocal entry into the CNS, presumably via infection of olfactory neuroepithelium and subsequent passage through the cribriform plate via infected macrophages or infection of brain capillary endothelial cells. During the viremic phase, there may be little or no recognizable disease except in tick-borne orthoflavivirus encephalitides, which have clearly delineated phases of fever and systemic illness.

CNS dysfunction or damage and associated clinical symptoms/signs arise partly from direct neuronal infection and subsequent damage and partly from edema, inflammation, and other indirect effects of the host response. The usual pathologic features of arboviral encephalitides are focal necroses of neurons, inflammatory glial nodules, and perivascular lymphoid cuffing. Involved areas display the "luxury perfusion" phenomenon, with normal or increased total blood flow and low oxygen extraction. The typical patient presents with a prodrome of nonspecific constitutional symptoms and signs, including fever, abdominal pain, sore throat, and respiratory signs. Headache, meningeal signs, photophobia, and vomiting quickly follow. The severity of human infection widely varies from an absence of symptoms/signs to febrile headache, aseptic meningitis, and full-blown encephalitis. The proportions and severity of these manifestations vary with the infecting virus and with known (e.g., age) and mostly unknown host immunogenetic determinants. Encephalitic involvement of deeper brain structures may be signaled by lethargy, somnolence, and cognitive deficit detected by mental status examination. More severely affected patients are obviously disoriented and may become comatose. Tremors, loss of abdominal reflexes, cranial nerve palsies, hemiparesis, monoparesis, difficulty swallowing, limb-girdle syndrome, and frontal lobe signs are common. Spinal and motor neuron diseases have been documented after WNV and Japanese encephalitis virus infections. Seizures and focal signs may be evident early or may appear during the disease. Some patients present with an abrupt onset of fever, convulsions, and other signs of CNS involvement. Encephalitis usually lasts a few days to several weeks and may be fatal, or recovery may be slow (with weeks or months before the return of maximal recoverable function) or incomplete (with persisting long-term deficits). Difficulty concentrating, fatigability, tremors, and personality changes are common during recovery.

The diagnosis of arboviral encephalitides depends on careful history-taking (travel/exposures), physical examination (in particular, of a febrile patient with signs of CNS dysfunction), and virus-specific or serology-specific laboratory testing. Clinicians should (1) consider empirical acyclovir treatment for herpesvirus meningoencephalitis and antibiotic treatment for bacterial meningitis until test results are received; (2) exclude intoxination and metabolic or oncologic causes, including paraneoplastic syndromes, hyperammonemia, liver failure, and anti-*N*-methyl-D-aspartate (NMDA) receptor encephalitis; and (3) rule out a brain abscess or a cerebrovascular event. Leptospirosis, neurosyphilis, Lyme disease, cat-scratch disease, opportunistic infections in immunocompromised patients, and more recently described viral encephalitides (e.g., Nipah virus infection), among others, should be considered in the differential diagnosis if epidemiologically relevant. CSF examination usually shows a modest increase in white blood cell (WBC) counts—in the tens or hundreds or perhaps a few thousand. In early phases of disease, a significant proportion of CSF WBCs may be polymorphonuclear leukocytes (PMNs), but monocytes later predominate. CSF glucose concentrations are generally normal. There are exceptions to this pattern of findings. In eastern equine encephalitis, for example, PMNs may predominate during the first 72 h of disease, and hypoglycorrhachia may be detected. In lymphocytic choriomeningitis, lymphocyte counts may be in the thousands, and glucose concentrations may be diminished. A humoral immune response is usually detectable at or near the onset of disease. Both serum (sampled in the acute or convalescent phase) and CSF should be examined for IgM antibodies, and plaque-reduction neutralization testing and/or RT-PCR should be conducted to detect viruses. Viruses generally cannot be isolated from blood or CSF, although Japanese encephalitis virus has been recovered from CSF of patients with severe disease. RT-PCR analysis of CSF may yield positive results but cannot exclude diagnoses. Metagenomic/high-throughput sequencing of CSF may support diagnosis of arboviral encephalitides. Viral antigen is present in brain tissue, although its distribution may be focal. Electroencephalography usually shows diffuse abnormalities and thus is not directly helpful.

Experience with medical imaging is still evolving but is generally nonspecific, as most patients do not present with pathognomonic abnormalities; however, imaging may be useful to rule out other suspected causes of disease. Both computed tomography (CT) and magnetic resonance imaging (MRI) scans may show normal results except for evidence of preexisting conditions or occasional diffuse edema. Notably, images acquired early may be normal and imaging abnormalities may only be detected later as the disease evolves. Although not pathognomonic, characteristic imaging abnormalities for some arboviral encephalitides have been reported, such as with eastern equine encephalitis (focal abnormalities) and severe Japanese encephalitis (hemorrhagic bilateral thalamic abnormalities).

Supportive care for severely ill patients may require management of elevated intracranial pressure, the syndrome of inappropriate secretion of antidiuretic hormone, respiratory failure, or seizures. Specific therapies for these viral encephalitides are not available. The only practical preventive measures are vector management and personal protective measures against the arthropod transmitting the virus. Direct human-to-human transmission has not been confirmed. For Japanese or Central European/Far Eastern tick-borne encephalitides, vaccination should be considered in specific circumstances (see relevant sections that follow).

Flavivirids The most significant orthoflavivirus encephalitides are Central European/Far Eastern tick-borne encephalitides, Japanese encephalitis, St. Louis encephalitis, and WNV infection. Murray Valley encephalitis and Rocio virus infection resemble Japanese encephalitis but are documented only occasionally in Australia and Brazil. Powassan virus has caused ≈400 cases of often-severe disease (CFR ≈8%), frequently occurring among children in eastern Canada and the United States. Usutu virus causes sporadic human infections (e.g., 105 cases from 2012 to 2021 in Europe), but such infections may be underdiagnosed.

CENTRAL EUROPEAN/FAR EASTERN TICK-BORNE ENCEPHALITIDES Tick-borne encephalitis virus (TBEV) is currently subdivided into four groups: the western/European subtype (previously called Central European encephalitis virus), the (Ural-)Siberian subtype (previously called Russian spring–summer encephalitis virus), the Far Eastern subtype, and the louping ill subtype (previously called louping ill virus, or, in Japan, Negishi virus). Small mammals, grouse, deer, and sheep are the vertebrate amplifiers for these viruses, which are transmitted by ticks. The risk of infection varies by geographic area and can be highly localized. Human infections usually follow outdoor activities during which tick bites occur or consumption of raw (unpasteurized) milk or cheese from infected goats or, less commonly, infected cows or sheep. Milk seems to represent the main transmission route for louping ill subtype viruses, which cause disease very rarely. Several thousand infections with TBEV are recorded each year among people of all ages. Tick-borne encephalitis occurs between April and October, with a peak in June and July.

Western/European viruses classically caused bimodal disease. After an incubation period of 7–14 days, the illness begins with an influenza-like febrile prodrome (fever, arthralgia/myalgia, headaches, and nausea and/or vomiting) that lasts for 2–4 days and is thought to correlate with viremia. A subsequent remission for several days is followed by the recurrence of fever and the onset of neurologic signs. The neuroinvasive phase (7–10 days before onset of improvement) varies from mild aseptic meningitis (more common among younger patients) to severe (meningo)encephalitis with coma, seizures, tremors, and motor signs. Spinal and medullary involvement can lead to typical limb-girdle paralysis and respiratory paralysis. Most patients with western/European virus infections recover (CFR ≈1–2%), and only a minority of patients have significant residual deficits. However, patients with (Ural-)Siberian virus infections have worse outcomes (CFR ≈7–8%) and are more likely to experience prolonged infections and develop chronic disability.

Infections with Far Eastern viruses generally run a more abrupt and severe course. The encephalitic syndrome caused by these viruses sometimes begins without a remission from the fever-myalgia phase and has more severe manifestations than the western/European syndrome. CFR is high (≈20–40%), and major sequelae—most notably, lower motor neuron paralyses of the proximal muscles of the extremities, trunk, and neck—are common, developing in approximately half of patients. Thrombocytopenia may occur during the initial febrile illness, resembling the early hemorrhagic phase of some other tick-borne orthoflavivirus infections, such as Kyasanur Forest disease. In the early stage of the illness, virus may be detected by PCR or isolated from the blood; however, after the onset of CNS manifestations, virus cannot typically be detected in or isolated from the CSF, and diagnosis requires detection of IgM antibodies in serum and/or CSF.

Diagnosis of Central European/Far Eastern tick-borne encephalitides primarily relies on serology and detection of viral genomes by RT-PCR. There is no specific therapy for infection. However, effective alum-adjuvanted, formalin-inactivated virus vaccines (FSME-IMMUN and Encepur) are produced in Austria, Germany, and Russia in chicken embryo cells. Two doses of the Austrian vaccine separated by an interval of 1–3 months appear to be effective in the field, though antibody responses are similar when vaccine doses are given 2 weeks apart. Because rare cases of postvaccination Guillain-Barré syndrome have been reported, vaccination should be reserved for people likely to experience rural exposure in an endemic area during the season of transmission. Cross-neutralization for the western/European and Far Eastern variants has been established, but there are no published field studies on cross-protection among formalin-inactivated vaccines.

Because up to 4% of ticks in endemic areas may be infected, the use of immunoglobulin prophylaxis of Central European/Far Eastern tick-borne encephalitides has increased in some regions. Prompt administration of high-titer-specific immunoglobulin is routine in some areas (e.g., Russia) but has been discontinued in many European countries because of concerns for antibody-mediated enhancement of infections and/or disease.

JAPANESE ENCEPHALITIS Japanese encephalitis is the most significant viral encephalitis in Asia. Each year, ≈68,000 cases and ≈13,600–20,400 deaths are reported. Japanese encephalitis virus is found throughout Asia—including in the Russian Far East, Japan, China, India, Pakistan, and Southeastern Asia—and causes occasional epidemics on western Pacific Islands. The virus has previously been considered endemic in Australia's Torres Strait Islands; from 2021 to 2022, 45 cases causing seven deaths were identified across a broad region of eastern and southern Australia. The virus is particularly common in areas where irrigated rice fields attract the natural avian vertebrate hosts and provide abundant breeding sites for *Culex tritaeniorhynchus* mosquitoes, which transmit the virus to humans. Amplification by pigs, which subsequently abort pregnancies, and horses, which develop encephalitis, may be significant, as well. Vaccination of domestic pigs and horses may reduce the transmission of the virus. After an incubation period of 5–15 days, clinical signs of Japanese encephalitis range from nonspecific febrile illness (nausea, vomiting, diarrhea, and/or cough) to aseptic meningitis, meningoencephalitis, acute flaccid paralysis, and severe encephalitis. Common findings are cerebellar signs, cranial nerve palsies, and cognitive and speech impairments. A Parkinsonian presentation and seizures are typical in severe cases. Patients may present with neuropsychiatric manifestations, including abnormal behavior and acute psychosis. MRI detection of thalamic abnormalities is specific but insensitive. CFR in hospitalized patients is high (≈20–30%) and long-term neurologic dysfunction and disability are common in survivors. However, most long-term residents of endemic areas eventually seroconvert after natural environmental exposures and develop protective immunity. Effective vaccines are available and indicated for naïve individuals who are relocating or traveling frequently or over extended periods to endemic rural areas. Usually, two intramuscular doses of the vaccine are given 28 days apart, with the second dose administered at least 1 week prior to travel.

ST. LOUIS ENCEPHALITIS St. Louis encephalitis virus is transmitted among birds by mosquitoes. This virus causes a low-level endemic infection among rural residents of the central and western United States, where *Culex tarsalis* mosquitoes serve as vectors. Urban mosquitoes (*Culex pipiens* and *Culex quinquefasciatus*) have been responsible for epidemics, resulting in hundreds or even thousands of cases in cities of the central and eastern United States. Most cases occur in June through October in the United States, but sporadic cases of the disease have been noted throughout the year in Latin/Central America and the Caribbean. The urban mosquitoes breed in accumulations of stagnant water and sewage with high organic content. The elimination of open sewers and trash-filled drainage systems is expensive and may not be possible. However, installation of screens on windows and vents of houses and implementation of personal protective measures may be effective in preventing infection. Mosquitoes are most active outdoors at dusk; bites can be avoided by modifying activities and using repellents.

Most St. Louis encephalitis virus infections are subclinical; susceptibility to severe disease increases with age. Infections that result in aseptic meningitis or mild encephalitis are concentrated among children and young adults, whereas severe and/or fatal cases primarily affect the elderly. Infection rates are similar in all age groups; a pathophysiologic explanation for higher risk of severe disease in older individuals is unclear. After an incubation period of 4–21 days, patients typically present with a nonspecific prodrome (fever, malaise, myalgia, and/or headache), followed by rapid-onset CNS manifestations with neurologic abnormalities that commonly include nuchal rigidity, hypotonia, hyperreflexia, myoclonus, and tremors. Severe cases can include cranial nerve palsies, hemiparesis, and seizures. Of interest, during and after the prodrome, patients often report dysuria and may have viral antigenuria and pyuria. Overall CFR is ≈7% but may be increased up to ≈20% among patients >60 years of age. Recovery is slow, and emotional lability, difficulty concentrating, memory issues, asthenia, and tremors are commonly prolonged in older convalescent patients. The CSF of patients with St. Louis encephalitis usually contains tens to hundreds of WBCs, with a lymphocytic predominance and a left shift. The CSF glucose concentration is normal in these patients.

WEST NILE VIRUS INFECTION WNV is now the primary cause of arboviral encephalitis in the United States. From 1999 to 2022, 28,684 cases of neuroinvasive disease (e.g., meningitis, encephalitis, and acute flaccid paralysis), with 2641 deaths, and 26,891 cases of nonneuroinvasive infection, with 135 deaths, were reported. WNV was initially described as being transmitted among wild birds by *Culex* mosquitoes in Africa, Asia, and southern Europe. In addition, the virus has been implicated in severe and fatal hepatic necrosis in Africa. WNV was introduced into New York City via diseased birds in 1999 and subsequently spread to other areas of the northeastern United States, causing die-offs among crows, exotic zoo birds, and other birds. The virus has continued to spread and is now found in all states (of the United States) as well as in Canada, Mexico, South America, and the Caribbean islands. *C. pipiens* mosquitoes remain the major vectors in the northeastern United States, but mosquitoes of several other *Culex* species and Asian tiger mosquitoes are also involved. Jays compete with crows and other corvids as amplifiers elsewhere in the United States.

The majority of WNV infections are subclinical. After an incubation period of 3–14 days, 20% of those infected develop typical F&M syndrome (without CNS involvement); <1% develop aseptic meningitis and severe encephalitis, particularly among the elderly. The febrile syndrome associated with WNV infection is notable for the frequent—rather than occasional—appearance of a maculopapular rash concentrated on the trunk (especially in children) and the development of lymphadenopathy. Back pain, fatigue, headache, myalgia, retroorbital pain, sore throat, nausea and vomiting, and arthralgia (but not arthritis) are common symptoms. Most patients fully recover, though asthenia may persist for several weeks. Patients who develop neuroinvasive disease typically have meningitis, encephalitis, or acute flaccid paralysis syndromes, often with some overlap. The risk of encephalitis, neurologic sequelae, and death is increased in male, elderly, diabetic, and hypertensive patients and in patients with previous CNS disease, but the pathophysiologic determinants of such a wide spectrum of disease phenotype and severity have otherwise been unexplained. Host genetic risk factors for neuroinvasive disease and death may include chemokine receptor CCR5 deficiency. Recent identification of a high prevalence (≈35%) of preexisting autoantibodies against type I interferons in patients with encephalitis (vs patients with subclinical WNV infection, in whom the prevalence approximated the general population) suggests this predisposition may underlie 40% of cases of WNV encephalitis. In addition to the more severe motor and cognitive sequelae, milder neurologic findings may include cranial nerve palsies, tremor, slight abnormalities in motor skills, and loss of executive functions. Given the burden of disease, intense clinical interest, and widely available diagnostics, more unusual features are increasingly described (including chorioretinitis, myocarditis, myositis and rhabdomyolysis, orchitis, flaccid paralysis with histologic lesions resembling poliomyelitis, and initial presentation with fever and focal neurologic deficits in the absence of diffuse encephalitis). Immunosuppressed patients may have fulminant courses or may develop persistent CNS infection. Postacute neurologic sequelae are common, especially in patients with severe neuroinvasive disease syndromes, and thus far have persisted in some patients for >5 years. Virus transmission through both transplantation and blood transfusion has necessitated screening of blood and organ donors by nucleic acid–based tests. A low but non-zero risk of maternal-to-fetal/neonatal transmission has been reported. Diagnosis rests upon detection of IgM antibodies in serum or CSF. In patients with potential prior exposures to sero-cross-reactive flaviviruses, plaque-reduction neutralization testing for specific neutralizing antibodies or molecular testing (PCR) may aid in specific diagnosis. Treatment is supportive only, and ventilatory support may be required for severe neuroinvasive disease. Although an equine vaccine is available, prevention of WNV infection in humans relies on avoidance of mosquito bites, vector control, and safe handling of potentially infected carcasses.

Peribunyavirids • **CALIFORNIA ENCEPHALITIS** California encephalitis virus has been implicated in only a very few cases of encephalitis (California encephalitis *sensu stricto*), whereas its close relative, La Crosse virus (LACV), is the major cause of this disease (historically ≈80–100 cases per year in the United States, although prevalence has been lower in recent years). La Crosse encephalitis is most reported in the upper midwestern United States but is also found in other areas of the central and southeastern parts of the country, such as West Virginia, Tennessee, North Carolina, and Georgia. Inkoo, Jamestown Canyon, Lumbo, snowshoe hare, and Ťahyňa viruses are close relatives of LACV that also cause human disease (California encephalitis *sensu lato*, including La Crosse encephalitis). Transovarial infection is a strong component of transmission of these viruses in *Aedes* and *Ochlerotatus* mosquitoes. The vector of LACV is the *Ochlerotatus triseriatus* mosquito. These mosquitoes are infected by transovarial transmission, venereal transmission, and feeding on viremic chipmunks and other mammals. *O. triseriatus* mosquitoes breed in standing water in locations such as tree holes and abandoned tires; they bite during daylight hours. Risk factors for human cases include recreation in forested areas, residence at a forest's edge, and nearby standing water. Intensive environmental modification based on these findings has reduced the incidence of disease in a highly endemic area in the midwestern United States.

Most humans are infected from July through September. Asian tiger mosquitoes efficiently transmit LACV to mice and have transmitted it transovarially in the laboratory. This aggressive anthropophilic mosquito has the capacity to urbanize, and its possible transmission of virus to humans is of concern. The prevalence of antibody to LACV in humans is 20% or higher in endemic areas, indicating that infection is common but most often asymptomatic or subclinical.

Neuroinvasive disease after LACV infection varies from aseptic meningitis accompanied by confusion to severe and occasionally fatal encephalitis (CFR <0.5%). CNS disease has been recognized primarily in children <15 years of age. After an incubation period of ≈3–7 days, a usual nonspecific febrile prodrome is followed by the sudden onset of CNS manifestations, including headache and lethargy, often with nausea, vomiting, convulsions (in half of patients), and coma (in one-third of patients). Focal seizures, hemiparesis, tremor, aphasia, chorea, Babinski signs, and other evidence of significant neurologic dysfunction are common during acute neuroinvasive disease. Most patients recover completely, but ≈10% develop postacute sequelae, including recurrent seizures, focal neurologic deficits, and cognitive and behavioral problems that may affect learning abilities.

The WBC count is commonly elevated and left-shifted in patients with LACV infection, sometimes reaching 20,000/μL. CSF leukocyte counts are typically 30–500/μL, usually with a monocyte predominance (although 25–90% of cells are PMNs in some patients). The blood protein concentration is normal or slightly increased, and the glucose concentration is normal. Specific virologic diagnosis based on IgM-capture assays of serum and CSF is efficient. The only human anatomic site from which virus has been isolated is the brain.

Treatment is supportive over a 1- to 2-week acute phase and focused on the prevention and management of status epilepticus, cerebral edema, and the syndrome of inappropriate secretion of antidiuretic hormone. A phase 2B clinical trial of IV ribavirin in children with LACV infection was discontinued during dose escalation because of adverse effects.

Jamestown Canyon virus has been implicated in several cases of encephalitis in adults (≈10–40 cases per year since 2011), usually with a significant respiratory illness at onset. Human infection with this virus has been documented in Northern America (in the United States in at least 10 states and in Canada), where the vector mosquito (*Aedes stimulans*) feeds on its main host, the white-tailed deer (*Odocoileus virginianus*). Ťahyňa virus can be found in Africa, China, Central Europe, and Russia. The virus is a prominent cause of undifferentiated febrile disease but can also cause pharyngitis, pulmonary syndromes, aseptic meningitis, or meningoencephalitis.

Rhabdovirids • **CHANDIPURA VIRUS INFECTION** Chandipura virus is an emerging and increasingly significant human virus in India, where it is transmitted among hedgehogs by mosquitoes and sandflies. In humans, the disease begins as a rapid-onset influenza-like illness, with fever, headache, abdominal pain, nausea, and vomiting. These manifestations are followed by infection-related or autoimmune-mediated encephalitis that may cause altered mental status and seizures. Notably, neither the virus nor abnormal cells are detected in CSF; the brain has not been examined histopathologically, but limited cerebral imaging data query whether CNS disease is more likely related to cerebrovascular events (vasospasm/vasculitis) than to infectious encephalitis. Several hundred cases of Chandipura virus infection are recorded in India every year in children, with a CFR ≈45–80%. Infections with other arthropod-borne rhabdovirids (Isfahan, Piry, vesicular stomatitis Indiana, and vesicular stomatitis New Jersey viruses) may imitate the early febrile stage of Chandipura virus infection.

Togavirids • **EASTERN EQUINE ENCEPHALITIS** This disease is encountered primarily in swampy foci along the east coast and, more recently and specifically, the southeastern coast of the United States, with a few inland foci as far removed from the coastlines as Michigan, Wisconsin, Arkansas, and Montana (≈4–40 cases/year). Humans are mostly infected from June through October. During this period, *Culiseta* mosquitoes infect birds and virus spills over into other vectors, such as *Aedes sollicitans* or *Aedes vexans* mosquitoes, which are more likely to feed on mammals. Especially with spread to the southeastern coast, there is concern over the amplifying role of introduced Asian tiger mosquitoes, which have been found to be infected with eastern equine encephalitis virus and are an effective experimental vector in the laboratory. Horses are a common target for the virus. Contact with unvaccinated horses may be associated with human disease, but horses probably do not play a significant role in amplification of the virus.

Most infections are subclinical. After an incubation period of ≈5–10 days, symptomatic individuals develop a nonspecific febrile prodrome, after which a small proportion (2% of adults, 6% of children) develop sudden and rapidly progressive neuroinvasive disease with (meningo)encephalitis syndromes associated with focal neurologic deficits, seizures, profoundly altered mental status, and coma. Neuroinvasive disease is highly fatal (CFR ≥30–50%) and >50% survivors have residual postacute sequelae. Acute PMN CSF pleocytosis, often occurring during the first 1–3 days of disease, is another indication of severity, and it parallels peripheral leukocytosis with a left shift. Extensive necrotic lesions and PMN infiltrates are found at postmortem examination of the brain. Specific treatment is not available, although the therapeutic role of IV immunoglobulin (IVIG) remains under query. A formalin-inactivated vaccine has been used to protect laboratory workers but is not generally available or applicable.

VENEZUELAN EQUINE ENCEPHALITIS VEEV is separated into epizootic viruses (subtypes IA/B and IC) and enzootic viruses (subtypes ID, IE, and IF). Closely related enzootic viruses are Everglades virus, Mucambo virus, and Tonate virus. Enzootic viruses are found primarily in humid tropical-forest habitats and are maintained between culicoid mosquitoes and rodents. These viruses cause acute febrile syndromes in humans but are not pathogenic for horses and do not cause epizootics. Everglades virus has caused sporadic cases of encephalitis in humans in Florida. Extrapolation from the rate of genetic change suggests that Everglades virus may have been introduced into Florida <200 years ago. Everglades virus is most closely related to the ID-subtype viruses that appear to have given evolutionary rise to the epizootic variants active in South America.

Epizootic viruses have an unknown natural cycle but periodically cause extensive epizootics/epidemics in equids and humans in the Americas. These epizootics/epidemics are the result of high-level viremia in horses and mules, which transmit the infection to several types of mosquitoes, which in turn infect humans. Humans also have high-level viremia, but their role in virus transmission is unclear. Relatively restricted epizootics of Venezuelan equine encephalitis (VEE) occurred repeatedly in South America in <10-year intervals from the 1930s until 1969, when a massive epizootic, including tens of thousands of equine and human infections, spread throughout Central America and Mexico, reaching southern Texas in 1971. Genetic sequencing suggested that the virus from that outbreak originated from residual "un-inactivated" IA/B-subtype virus in veterinary vaccines. The outbreak was terminated in Texas with a live-attenuated vaccine (TC-83), originally developed for human use by the U.S. Army; the epizootic virus was then used for further production of inactivated veterinary vaccines. No further major outbreaks occurred until 1995 and 1996, when large epizootics of VEE occurred in Colombia/Venezuela and Mexico, respectively. Of the >85,000 clinical cases, 4% (more children than adults) had neurologic symptoms/signs, resulting in 300 deaths. The viruses involved in these epizootics, as well as previously epizootic IC viruses, are close phylogenetic relatives of known enzootic ID viruses. This finding suggests that active virus evolution and selection of epizootic viruses are underway in South America.

During epizootics, extensive human infection is typical, with clinical disease occurring in 10–60% of infected individuals. Most infections result in notable acute febrile syndromes, whereas relatively few (5–15%) result in neurologic disease. A low rate of CNS invasion is supported by the absence of encephalitis after the many infections (from exposure to aerosols) that have occurred in the laboratory setting. Small-molecule and monoclonal antibody-based antiviral therapeutics are in preclinical development but not yet approved. These (and vaccine; see below) efforts may benefit from the designation of VEEV as a representative pathogen for the encephalitic alphaviruses.

The prevention of epizootic VEE depends on vaccination of horses with the attenuated TC-83 vaccine or with an inactivated vaccine prepared from that variant. Enzootic viruses are genetically and antigenically different from epizootic viruses, and protection against the former with vaccines prepared from the latter is relatively ineffective. Humans can be protected by immunization with similar vaccines prepared from Everglades virus, Mucambo virus, and VEEV, but the use of the vaccines is restricted to laboratory personnel because of reactogenicity, possible fetal pathogenicity, and limited availability.

WESTERN EQUINE ENCEPHALITIS The primary maintenance cycle of western equine encephalitis virus in the western United States and Canada involves *Aedes*, *C. tarsalis*, and *Culiseta* mosquitoes and birds (principally sparrows and finches). Equids and humans become infected, and both suffer encephalitis without amplifying the virus. St. Louis encephalitis virus is transmitted via a similar cycle in the same regions harboring western equine encephalitis virus; disease caused by the former occurs about a month earlier than that caused by the latter (July through October). Large epidemics of western equine encephalitis occurred in the western and central United States and Canada from the 1930s through the 1950s but have been uncommon since then. From 1964 through 2010, only 640 cases were reported in the United States. This decline in incidence may reflect, in part, the integrated approach to mosquito management employed in irrigation projects and, in part, the increasing use of agricultural pesticides. The decreased incidence of western equine encephalitis almost certainly reflects the increased tendency for humans to be indoors behind closed or screened windows at dusk—the peak biting period of the major vector.

Most infections are subclinical or mild. After an incubation period of ≈5–10 days, most symptomatic patients develop a nonspecific febrile prodrome with spontaneous recovery. The small minority of patients who develop neuroinvasive disease present with a typical viral meningoencephalitis syndrome (headache, vomiting, neck and back stiffness, and/or altered mental status). The frequency and morbidity of severe neurologic disease are increased among infants and very young children; in addition, CFR is high in this and in very elderly populations (3–7% overall). One-third of individuals who have convulsions during the acute illness have subsequent seizure activity. Infants <1 year of age—particularly those in the first months of life—are at serious risk of persistent motor and cognitive deficits. Of those 5–9 years of age, twice as many males as females develop clinical encephalitis; the contribution of biological versus behavioral factors (increased outdoor exposure to the vector) to incidence differences is uncertain. Specific treatment is not available. A formalin-inactivated vaccine has been used to protect laboratory workers but is not generally available.

■ PULMONARY DISEASE

Hantavirus pulmonary syndrome (HPS) was first described in 1993, but retrospective identification of cases by immunohistochemistry (1978) and serology (1959) supports the idea that HPS is a recently discovered disease rather than a truly new one. The causative agents are orthohantaviruses of a distinct phylogenetic lineage that is associated with the cricetid rodent subfamily Sigmodontinae. Sin Nombre virus, which chronically infects western deermice (*Peromyscus sonoriensis*), is the most important agent of HPS in the United States. Several other related viruses (Anajatuba, Andes, Araraquara, Araucária, bayou, Bermejo, Black Creek Canal, Blue River, Caño Delgadito, Castelo dos Sonhos, Catacamas, Choclo, Juquitiba, Laguna Negra, Lechiguanas, Maciel, Maripa, Monongahela, New York, Orán, Paranoá, Pergamino, Rio Mamoré, and Tunari viruses) cause the disease in Northern and South America. Andes virus is unusual in that it has been implicated in human-to-human transmission. HPS particularly affects rural residents living in dwellings permeable to rodent entry or workers in occupations that risk rodent exposure. Each type of rodent has particular habits; in the case of deermice, these behaviors include living in and around human habitation.

After a typical incubation period of 2–3 weeks, HPS begins with a nonspecific prodrome of ≈3–4 days (range 1–11 days) comprising fever, malaise, myalgia, and—in many cases—gastrointestinal disturbances (e.g., abdominal pain, nausea, and vomiting). Dizziness is common, and vertigo is occasional. Severe prodromal symptoms/signs may bring some patients to medical attention, but most cases are recognized as the pulmonary phase begins, with typical signs of slightly lowered blood pressure, tachycardia, tachypnea, mild hypoxemia, thrombocytopenia, and early radiographic signs of pulmonary edema. Physical findings in the chest are often surprisingly scant. The conjunctival and cutaneous signs of vascular involvement seen in orthohantavirus VHFs (see "Viral Hemorrhagic Fever," below) are uncommon but may be more frequent with HPS outside Northern America. Early recognition and triage of patients at this stage are critical because, shortly after onset, hypotension and noncardiogenic edema progress rapidly to hypoxemic respiratory failure.

The differential diagnosis of HPS includes undifferentiated sepsis syndromes and severe acute respiratory infections and should be informed by local epidemiology. Considerations include severe atypical pneumonia (*Legionella*, *Mycoplasma*, and Q fever), influenza, rickettsial disease, meningococcemia, septicemic plague, tularemia, leptospirosis, dengue, and YF. Fever with severe abdominal pain and tenderness merits consideration of the causes of a "surgical" abdomen and pyelonephritis. A specific diagnosis is best made when antigen-specific IgM, present in most symptomatic patients, is detected in acute-phase serum. Sin Nombre virus antigen testing detects antibodies to the related HPS-causing orthohantaviruses. Occasionally, heterotypic viruses will cross-react (only to the IgG ELISA), but the very low seroprevalence of these viruses in normal populations queries this laboratory finding. Orthohantaviral RNA is cleared quickly, although RT-PCR may be useful to detect virus in tissue biopsies, thrombus, or at autopsy. RT-PCR and DNA sequencing may be needed in specifically identifying the infecting virus in areas outside the home range of deermice, in atypical cases, and for molecular epidemiology.

During the prodrome, differentiating HPS from other causes is difficult, but diagnostically helpful features emerge at or soon after initial presentation. A new dry cough, usually absent at disease onset, may signal the pulmonary phase. Early in disease, interstitial edema may be evident radiographically; with progression, bilateral alveolar infiltrates in a central pattern and with a normal-sized heart signal the development of noncardiogenic edema. Pleural effusions are often seen. Coincident thrombocytopenia, circulating atypical lymphocytes (lymphoblasts), and a left shift (often with marked leukocytosis) are considered almost diagnostic in experienced centers. Hemoconcentration, hypoalbuminemia, and proteinuria also aid in diagnosis. Although thrombocytopenia and partial thromboplastin time (PTT) prolongation are usually found, clinical evidence of coagulopathy or laboratory indications of DIC are found in only a minority of severely ill patients. Patients with severe illness also have acidosis and elevated serum lactate concentrations. Oliguria and mildly increased values in serum creatinine concentrations are common, but patients with severe HPS often have markedly decreased glomerular filtration. Some American orthohantaviruses other than Sin Nombre virus (e.g., Andes virus) have been associated with more severe renal dysfunction, but few such cases have been studied.

After recognition, early triage-based monitoring and management of hemodynamic, respiratory, and renal dysfunction during the first few hours are critical to improved outcomes. During this period, hypotension and hemoconcentration suggest intravascular volume depletion; however, aggressive IV fluid therapy is not advised in the setting of widespread capillary leak, frequently depressed cardiac output, and oliguria. Mild hypoxemia may be managed by oxygen administration alone, but mechanical ventilatory support of respiratory failure and shock is often needed. Pulmonary artery catheterization enables pulmonary capillary wedge pressure–guided inotropic and vasopressor support and cautious intravascular volume replacement. The most critically ill patients with a low cardiac index or refractory hypoxemia despite this support may benefit from extracorporeal membrane oxygenation, ideally before the onset of shock. CFR is high (≈30–40%), even with appropriate management, but most patients surviving the first 48 h of hospitalization are extubated and discharged within a few days with no apparent long-term sequelae. Although ribavirin inhibits orthohantaviruses in vitro and has been effectively used in patients with HFRS due to Hantaan virus, an underpowered randomized clinical trial of ribavirin for HPS failed to show efficacy. Preclinical development of novel small-molecule– and monoclonal antibody-based therapeutics for HPS is ongoing.

■ VIRAL HEMORRHAGIC FEVER

VHF syndrome encompasses a small number of acute viral infections that cause severe disease and multisystem organ dysfunction sharing a mutual pathophysiologic disruption to vascular function, stability, and integrity. In practice, viral infections classified as VHFs vary widely across a spectrum of clinical phenotypes and severity; indeed, hemorrhagic manifestations are typically not the most prominent nor severe features of disease. Nonetheless, the VHF designation largely captures a useful common theme: hemorrhagic manifestations are a sign of widespread vascular dysfunction or damage that plays a central role in local, organ-specific, and systemic expression of disease. Direct or indirect damage to the microvasculature leads to increased permeability and (particularly when platelet function is decreased) to actual disruption and local hemorrhage most evident in the skin and mucous membranes. Cutaneous flushing and conjunctival suffusion are examples of common observable abnormalities due to dysfunction of local circulation. Similarly, affected microvasculature may be fragile to trauma and external forces, such as the application of a blood pressure cuff causing a positive tourniquet sign. Although overt hemorrhage occurs infrequently, subtle local hemorrhagic signs (e.g., bleeding of the gums or at IV catheter sites) are often present on careful examination. Circulatory dysfunction may also contribute to specific organ dysfunction or damage that may be particularly prominent in some VHF cases; for instance, the kidneys are primary targets in HFRS, and the liver is a primary target in YF. Regarding severe systemic manifestations, hypotension and shock are not usually a direct result of life-threatening blood loss but rather an indication of generalized circulatory dysfunction contributing to hemodynamic instability. Despite this common theme, the pathogenesis of VHF is poorly understood and varies among the viruses regularly implicated in the syndrome. In some viral infections, direct damage to the vascular system or even to parenchymal cells of target organs is an important factor; in other viral infections, soluble mediators are thought to play a major role in the development of hemorrhage or hemodynamic derangement.

Symptom initiation in the acute phase in most VHF syndromes is associated with ongoing virus replication and viremia, typically with nonspecific fever and arthralgia/myalgia, usually of abrupt onset (mammarenavirus infections are exceptions, as they often develop gradually). Within a few days, patients present for medical attention with severe fatigue and prostration that is often accompanied by anorexia, severe headache, sore throat, chest pain, gastrointestinal symptoms (abdominal pain, nausea, vomiting, and/or diarrhea), dizziness, hyperesthesia, and photophobia. Initial examination often reveals only an acutely ill febrile patient with asthenia, conjunctival suffusion, tenderness to palpation of muscles or abdomen, and borderline hypotension or postural hypotension (perhaps with tachycardia). Petechiae (often best visualized in the axillae), flushing of the head and thorax, periorbital edema, and proteinuria are common. AST activities are usually elevated at presentation or within a day or two thereafter. Hemoconcentration from vascular leakage is usually evident and is most marked in HFRS and in severe dengue. Seriously ill patients develop more severe clinical signs, including initial organ dysfunction typical of the causative virus; continued high viral loads coupled with ineffective and dysregulated immune responses lead to progressive multiorgan dysfunction characterized by shock, acute kidney injury, CNS injury (encephalopathy, coma, seizures), severe hemorrhagic manifestations, and death.

One of the major diagnostic clues is travel to an endemic or active outbreak area within the incubation period for a given syndrome. Except in infections with Seoul virus, DENV-1–4, and yellow fever virus (YFV), which have urban hosts/vectors, rural travel exposures are more commonly associated with a VHF. In addition, several travel-associated diseases—falciparum malaria, shigellosis, typhoid fever, leptospirosis, relapsing fever, and rickettsial diseases—should also be considered in the differential diagnosis, as they are treatable despite being potentially fatal.

Early recognition of VHF is crucial for timely initiation of appropriate supportive clinical care and virus-specific therapy (if available). Key components of clinical care include prompt isolation and hospitalization even in absence of clinical signs; rapid triage and ongoing monitoring of vital signs; the prevention and management of dehydration, hypovolemia, and hemodynamic instability (with closely monitored IV fluids and vasopressors); the prevention and management of acute kidney injury and electrolyte abnormalities; empiric or specific treatment of common concurrent or secondary bacterial or parasitic infections; replacement of blood products (packed red blood cells, clotting factors, and platelets as indicated); and the usual precautionary measures used in the treatment of patients with hemorrhagic diatheses. DIC should be treated only with clear laboratory evidence and if laboratory monitoring of therapy is feasible; there is no proven benefit of such therapy. Of note, VHF patients are commonly hypovolemic from insensible fluid and gastrointestinal losses and require intravascular fluids to maintain blood pressure and glomerular filtration; however, VHF patients may also have decreased cardiac output, capillary leak syndromes, and oliguria that require judicious fluid management and inotropic/vasopressor support. Specific therapy is available for several of the VHFs. Strict barrier nursing and other precautions against infection of medical staff and visitors are indicated when VHFs are encountered, except when the illness is due to DENV-1–4, orthohantaviruses, RVFV, or YFV.

Novel VHF-causing agents are still being discovered. Some of these viruses have recently emerged and continue to cause disease (e.g., severe fever with thrombocytopenia syndrome virus [SFTSV]). Others have caused isolated VHF-like cases without recurrence, including viruses for whom Koch's postulates are yet to be fulfilled (e.g., Bas-Congo virus).

Bunyaviricetes The most significant VHF-causing bunyaviricetes are arenavirids (Junín, Lassa, and Machupo viruses), hantavirids, nairovirids (Crimean-Congo hemorrhagic fever virus), and phenuivirids (RVFV and SFTSV). Other bunyaviricetes—e.g., the Garissa variant of Ngari virus and Ilesha virus (both orthobunyaviruses) or Chapare, Guanarito, Lujo, and Sabiá viruses (all mammarenaviruses)—have caused sporadic VHF outbreaks.

ARGENTINIAN AND BOLIVIAN HEMORRHAGIC FEVERS These severe diseases (with CFR reaching >50% in individual outbreaks) are caused by Junín virus and Machupo virus, respectively. Clinical presentations of both diseases are similar, but epidemiology differs because of the distribution and behavior of the distinct rodent reservoirs. Argentinian hemorrhagic fever has been recorded only in rural areas of Argentina, whereas Bolivian hemorrhagic fever seems to be confined to rural Bolivia. Unusually for arboviruses, infection with the causative agents almost always results in disease, and all ages and both sexes are affected. Person-to-person or nosocomial transmission is rare but has occurred. The transmission of Junín virus from male survivors to their partners suggests the need to avoid intimate contact in early convalescence (and counseling in this regard for all patients with mammarenavirus hemorrhagic fevers). In contrast to LF (see below), thrombocytopenia—often marked—is the rule, mucosal hemorrhage is common, and CNS dysfunction (e.g., marked confusion, tremors of the upper extremities and tongue, and cerebellar signs) is much more common in disease caused by Junín virus and Machupo virus (20–30%). Some cases follow a predominantly neurologic course, with a poor prognosis. Convalescence is protracted and frequently associated with sequelae such as hearing loss.

Typical clinical laboratory findings (thrombocytopenia, leukopenia, and proteinuria) support the diagnosis. Argentinian and Bolivian hemorrhagic fevers are readily treated with convalescent plasma given within the first 8 days of illness; however, the use of convalescent plasma may be associated with neurologic sequelae. In the absence of passive antibody therapy, (off-label) IV ribavirin in the dose recommended for LF is likely to be effective in all South American VHFs caused by mammarenaviruses. A safe, effective, live-attenuated vaccine exists for Argentinian hemorrhagic fever. After vaccination of >250,000 high-risk people in the endemic area, the incidence of this VHF decreased markedly. In experimental animals, this vaccine is cross-protective against Bolivian hemorrhagic fever. Other mammarenaviruses have caused isolated or sporadic VHF in South America, including Guanarito virus (Venezuelan hemorrhagic fever) and Chapare and Sabiá viruses. Multivalent vaccine approaches against all five VHF-causing mammarenaviruses in this region are in preclinical development.

LASSA FEVER Lassa virus (LASV) is known to cause endemic and epidemic disease in Nigeria, Sierra Leone, Guinea, and Liberia, although it is probably more widely distributed in Western Africa. In Western Africa alone, probably tens to hundreds of thousands of LASV infections occur annually, and ≈20% of these develop LF, which can be a prominent cause of febrile disease. For example, in one hospital in Sierra Leone, laboratory-confirmed LF is consistently responsible for one-fifth of admissions to the medical wards. LASV can be transmitted by close person-to-person contact. Nosocomial spread has occurred but is uncommon if appropriate infection prevention and control practices are followed. All ages and both sexes are affected; the incidence of disease is highest in the dry season, but transmission occurs year-round.

The majority of LASV infections are subclinical or mild. After a highly variable incubation period of 1–3 weeks, symptomatic patients have the gradual onset of a nonspecific febrile prodrome (among the VHF agents, only mammarenaviruses are typically associated with a gradual onset of illness). Progression in the second week of illness includes sore throat, cough, retrosternal chest and back pain, and gastrointestinal symptoms (abdominal pain, vomiting, and/or diarrhea). A maculopapular rash may be noted in light-skinned patients. Further disease progression includes conjunctival infection, head and neck edema, and respiratory (pneumonia and/or pleural effusions) and cardiac (pericarditis and/or pericardial effusions) manifestations. In late stages, multisystem organ dysfunction occurs with acute kidney injury, CNS manifestations (encephalopathy, coma, and/or seizures), shock, and hemorrhage (seen in only ≈15–30% of LF patients overall). Laboratory abnormalities may include leukocytosis, thrombocytopenia, elevated serum creatinine levels, hyperkalemia, hypoalbuminemia, elevated AST and ALT activities, and proteinuria. Viral load, age, pregnancy status, clinical signs (CNS features, face and neck swelling, hypotension, hemorrhage, and/or jaundice), and laboratory abnormalities (high serum creatinine levels, AST activity, leukocytosis, and/or thrombocytopenia) have been associated with fatal outcome. Pregnancy is associated with higher case fatality, especially during the last trimester, when fetal death is 90%. Interruption of pregnancy via medical abortion may increase survival rates, but data on LF and pregnancy outcomes remain sparse. In survivors, a high rate (≈30%) of sensorineural deafness (onset during the acute illness or in convalescence, usually bilateral) occurs; other postacute sequelae of LF are poorly defined but include neurologic and mental health issues. The virus is detected in the urine during convalescence and has been detected in seminal fluid early in recovery. Reinfection may occur but has not been associated with severe disease.

Aggressive supportive and critical care, including renal replacement therapy when available, is required to improve outcomes. Regarding LASV-specific therapy, observational studies in the 1980s informed the current practice of treating patients with (off-label) ribavirin (IV route preferred). Historically, ribavirin appeared to be partially effective in reducing fatality from that documented among retrospective controls. However, possible side effects, such as reversible anemia (which usually does not require transfusion), dependent hemolytic anemia, and bone marrow suppression, need to be kept in mind. When used, ribavirin should be given by slow IV infusion in a dose of 32 mg/kg; this dose should be followed by 16 mg/kg every 6 h for 4 days and then by 8 mg/kg every 8 h for 6 days. Small-molecule and monoclonal antibody therapeutics are in advanced preclinical development, and research networks in Western Africa are preparing for randomized clinical trials that will be critical to inform best specific therapeutic approaches. Inactivated LASV vaccines failed in preclinical studies, but several promising vaccine platforms are under experimental evaluation.

HEMORRHAGIC FEVER WITH RENAL SYNDROME HFRS is the most significant VHF today, with >100,000 cases of severe disease in Asia annually and thousands of mild infections in Europe. The disease is widely distributed in Eurasia. The major causative viruses are Puumala virus (Europe), Dobrava virus (the Balkans), and Hantaan virus (Eastern Asia). Amur, gōu, Kurkino, Muju, Saaremaa, Sochi, and Tula viruses also cause HFRS but much less frequently and in more geographically confined areas that are determined by the distribution of reservoir hosts. Seoul virus is an exception; because it is associated with brown rats (*Rattus norvegicus*), which migrate on ships, the virus has a worldwide distribution. Despite the wide distribution of Seoul virus, only mild or moderate HFRS occurs in Asia; human disease has been difficult to identify in many areas of the world, although an outbreak in North America has been described. Most cases of HFRS occur in rural residents or vacationers; again, the exception is Seoul virus infection, which may be acquired in an urban, rural, or laboratory setting. Classic Hantaan virus infection in Korea and in rural China is most common in the spring and fall and is related to rodent density and agricultural practices. Human infection is acquired primarily through aerosols of rodent urine, although virus is also present in rodent saliva and feces. Patients with HFRS are not infectious.

Most patients with hantavirus infections do not develop severe disease. Classic features of HFRS include fever, hypotension, hemorrhage, and acute kidney injury, with severe disease expressed in four identifiable stages after a typical 1- to 2-week incubation period:

1. The *febrile* stage lasts 3 or 4 days and is identified by the abrupt onset of fever, headache, severe myalgia, thirst, anorexia, and often nausea and vomiting. Photophobia, retroorbital pain, and pain on ocular movement are common, and the vision may become blurred with ciliary body inflammation. Flushing over the face, the anterior neck, and back is characteristic, as are pharyngeal injection, periorbital edema, and conjunctival suffusion. Petechiae often develop in areas of pressure, the conjunctivae, and the axillae. Back pain and tenderness to percussion at the costovertebral angle reflect massive retroperitoneal edema. Thrombocytopenia and mild to moderate DIC is present, and hemorrhagic manifestations may occur. Other early laboratory findings include proteinuria and an active urinary sediment.
2. During the *hypotensive* stage (lasting from a few hours to 48 h), the blood pressure falls (sometimes with shock) with a compensatory tachycardia (vs the relative bradycardia typical of the febrile phase). Kinin activation and capillary leak are marked, leading to hemoconcentration. Leukocytosis with a left shift develops, and thrombocytopenia persists. Atypical lymphocytes—which, in fact, are activated $CD8^+$ and, to a lesser extent, $CD4^+$ T cells—circulate. Renal perfusion is compromised from local and systemic circulatory changes resulting in necrosis of tubules, particularly at the corticomedullary junction, and oliguria. Consequent inability to concentrate the urine develops, and the urine's specific gravity falls to 1.010. Proteinuria is marked. Diffuse hemorrhage may appear, including hematemesis, hemoptysis, and melena.
3. During the *oliguric* stage, hemorrhagic tendencies continue, probably in large part because of uremic bleeding defects. Oliguria persists for 3–10 days before the return of renal function marks the onset of the polyuric stage.
4. During the *polyuric* stage (diuresis with hyposthenuria), the inability to concentrate the urine may lead to volume depletion and electrolyte abnormalities.

Mild cases of HFRS may be much less stereotypical, presenting only with fever, gastrointestinal abnormalities, and transient oliguria followed by hyposthenuria. Infections with Puumala virus, the most common cause of HFRS in Europe ("nephropathia epidemica"), result in a similar general but much-attenuated presentation. Bleeding manifestations are found in only 10% of patients, hypotension rather than shock is usually documented, and oliguria is present in only about half of patients. The dominant features may be fever, abdominal pain, proteinuria, mild oliguria, and sometimes blurred vision or glaucoma, followed by polyuria and hyposthenuria in recovery. CFR is <1%.

HFRS should be suspected in patients with exposure in an endemic (typically rural) area. Prompt recognition of the disease permits rapid hospitalization and expectant management of shock and renal failure. Clinical laboratory parameters include leukocytosis (which may be leukemoid and with a left shift), thrombocytopenia, elevated serum creatinine level, proteinuria, and hematuria. HFRS is readily diagnosed by an IgM-capture ELISA that is positive at admission or within 24–48 h thereafter. The isolation of orthohantaviruses is difficult, but RT-PCR of renal biopsy, thrombus (collected early in the clinical course), or postmortem tissues may yield molecular diagnosis and should be used if definitive identification of the infecting virus is required.

Mainstays of therapy are management of shock with vasopressor support and modest crystalloid infusion, IV human serum albumin administration, timely renal replacement therapy to prevent volume overload that may result in pulmonary edema, and control of hypertension to prevent intracranial hemorrhage. Use of IV ribavirin has reduced CFR and morbidity in severe cases, provided treatment is begun within the first 4 days of illness. CFR approaches 15% but should be lower than 5% with adequate supportive and critical care. Generally, surviving patients make a full recovery, even of glomerular function, and postacute sequelae have not been identified.

CRIMEAN-CONGO HEMORRHAGIC FEVER (CCHF) CCHFV causes severe VHF over a wide endemic area (Africa, Asia, and Europe), determined primarily by the distribution of ixodid *Hyalomma* ticks. Incidence has been increasing in known endemic areas (particularly in Western Asia), and recent autochthonous infections in nonendemic areas, such as Spain, suggest that the at-risk geography is expanding. Because of the propensity of CCHFV-transmitting ticks to feed on domestic livestock and some wild mammals, veterinary serosurveys are the most effective mechanism for the monitoring of virus circulation. Human infections are acquired via tick bites, during the crushing of infected ticks, and contact with blood or body fluids of infected animals or humans (including nosocomial and vertical transmission). Domestic animals do not become ill but do develop viremia. Thus, risk of acquiring CCHFV occurs during sheep shearing, animal slaughter, or contact with infected hides or carcasses from recently slaughtered infected animals. Nosocomial outbreaks are common and are usually related to extensive blood exposure or needlesticks.

The majority of infections are subclinical. After an incubation period that varies by exposure to a tick bite (1–3 days) or blood or body fluids (3–7 days, perhaps shorter after nosocomial exposures to highly viremic patients), patients develop a sudden-onset nonspecific febrile prodrome typical to VHF. Severe illness is also typical, with the possible exceptions of a profound and dysregulated inflammatory component, more severe hemorrhagic manifestations (including extended ecchymoses and/or retinal and pulmonary hemorrhage), and more severe liver injury, resulting in jaundice in some patients. Recovery (for most patients; sequelae are not identified) or rapid progression to fatal disease occurs in the second week of illness with multiorgan (renal and respiratory) dysfunction, shock, and profound coagulopathy leading to massive bleeding. Clinical laboratory findings indicate coagulopathy as well as markers of liver (elevations in activities of AST/ALT and bilirubin concentrations) and muscle (elevations in creatine phosphokinase levels) injury. Predictors of fatal outcome include clinical features (hematemesis, melena, and/or altered mental status), laboratory biomarkers (viral load, thrombocytopenia, elevated AST/ALT levels, and/or obvious DIC), and delayed CCHFV-specific IgM and IgG responses. The mainstay of treatment is supportive and critical care, including extracorporeal support of organ dysfunction. The pathophysiology argues for combined virus-specific and disease-modifying immunomodulatory approaches. The benefit of IV ribavirin, suggested in observational studies that were likely confounded, has not been borne out in randomized trials. Small-molecule antivirals (favipiravir and molnupiravir), monoclonal antibodies, and the use of convalescent plasma are in preclinical evaluation, but clinical studies are needed. The routine use of steroids and/or IVIG is not currently recommended. No human or veterinary vaccines are recommended.

RIFT VALLEY FEVER The natural range of RVFV was previously confined to sub-Saharan Africa, with circulation of the virus markedly enhanced by substantial rainfall. The El Niño Southern Oscillation phenomenon of 1997 facilitated subsequent spread of Rift Valley fever (RVF) to the Arabian Peninsula, with epidemic disease in 2000. The virus has also been found in Madagascar and Egypt, where it caused major epidemics in 1977–1979, 1993, and thereafter. RVFV is maintained by transovarial transmission in floodwater *Aedes* mosquitoes and presumably also via a vertebrate amplifier. Increased transmission during particularly heavy rains leads to epizootics characterized by high-level viremia in cattle, goats, or sheep. Numerous types of mosquitoes feed on these animals and become infected, thereby increasing the possibility of human infections. Remote sensing via satellite can detect the ecologic changes associated with high rainfall that predict the likelihood of RVFV transmission. High-resolution satellites can also detect the special depressions in floodwaters from which the mosquitoes emerge. The virus can be transmitted to humans by mosquitoes but most commonly by contact with blood or aerosols from domestic animals. Therefore, transmission risk is high during birthing, and both abortuses and placentas need to be handled with caution. Risk is also high during animal slaughter but decreases thereafter as anaerobic glycolysis in postmortem tissues results in an acidic environment that rapidly inactivates bunyaviricetes in carcasses. Neither person-to-person nor nosocomial transmission of RVFV has been documented.

RVFV is unusual in that it causes several clinical syndromes. Most infections are either subclinical or produce a typical F&M syndrome after an incubation period of 2–6 days. A much smaller proportion (<1%) of patients develop typical VHF notable for prominent liver involvement (jaundice) and an extremely rapid course with high case fatality (50%, within 3–6 days), often with renal failure and DIC/hemorrhagic manifestations. Further, <1% of patients develop a late onset (1–4 weeks) meningoencephalitis syndrome that is not usually fatal but does lead to severe neurologic sequelae. Perhaps 10% of otherwise mild infections develop a late-onset (1–3 weeks) retinal vasculitis causing blurred or loss of vision. Funduscopic examination reveals edema, hemorrhages, and infarction of the retina as well as optic nerve degeneration. In a small proportion of patients (<1 in 200), retinal vasculitis is followed by viral encephalitis. Of interest, the late onset of both retinal disease and meningoencephalitis—after serum neutralizing antibody has developed—suggests (so far, undefined) immunopathology.

No proven therapy exists for RVF. Epidemic disease is best prevented by vaccination of livestock. The ability of this virus to propagate after introduction into Egypt suggests that other potentially receptive areas, including the United States, should develop response plans. RVF, like VEE, is likely to be controlled only with adequate stocks of an effective live-attenuated vaccine, but global stocks are unavailable. A formalin-inactivated vaccine confers immunity in humans; however, quantities are limited, and three injections are required. This vaccine is recommended for potentially exposed laboratory workers and for veterinarians working in sub-Saharan Africa. A new live-attenuated vaccine, MP-12, is being tested in humans (phase 2 trials have been completed). The vaccine is safe and licensed for use in sheep and cattle. A modified replication-incompetent chimpanzee adenovirus (ChAdOx1) vectored vaccine has recently been shown to be safe and immunogenic in phase 1 trials. In addition, several vaccines are being developed specifically for use in animals.

SEVERE FEVER WITH THROMBOCYTOPENIA SYNDROME This tick-borne disease is caused by SFTSV. The primary tick vector, the Asian long-horned tick (*Haemaphysalis longicornis*), is endemic in areas of southeast Asia and Oceania, and geographic distribution has been increasing, likely via migratory birds (in Asia) or by importation of pets or livestock (in the United States, where the vector is now widespread and considered invasive). Numerous human infections have been reported during the past few years from China, and cases have also been detected in Japan, the Republic of Korea, Vietnam, and Thailand. Seroprevalence in endemic areas of China approaches 5%. The clinical presentation ranges from mild nonspecific febrile syndromes to severe VHF with high CFR (>12%) particularly in older individuals. After a 1- to 2-week incubation period, patients develop a nonspecific febrile prodrome that may include abdominal pain, vomiting, and diarrhea. Common laboratory abnormalities include prominent thrombocytopenia, leukopenia (often with atypical lymphocytes), markers of hepatocyte injury, and elevated ferritin levels. Most patients recover spontaneously; however, in the second week of illness, some patients develop severe disease with multiorgan dysfunction that can include acute renal injury, myocarditis, meningoencephalitis, DIC with hemorrhage, and hemophagocytic lymphohistiocytosis. Diagnosis requires RT-PCR, as serologic tests are not available. No vaccines or approved therapeutics are available (though ribavirin and favipiravir have been used in patients) and treatment relies on supportive care.

Flasuviricetes The most significant orthoflaviviruses that cause VHF are the mosquito-borne DENV-1–4 and YFV. These viruses are widely distributed and cause tens to hundreds of thousands of infections each year. Alkhurma hemorrhagic fever virus (>600 cases since discovery in 1995), Kyasanur Forest disease virus (≈10,000 cases over 60 years), and Omsk hemorrhagic fever virus (isolated infections every year with intermittent larger outbreaks) are geographically very restricted but prevalent tick-borne orthoflaviviruses that cause VHF, sometimes with subsequent viral encephalitis. Tick-borne encephalitis virus has caused VHF in a few patients. There is currently no therapy for infection with these VHFs, but an inactivated vaccine has been used in India to prevent Kyasanur Forest disease (and there is a vaccine to prevent tick-borne encephalitis; see above).

SEVERE DENGUE Although most individuals infected with DENV-1–4 have either subclinical infection or F&M syndrome (the febrile phase), some of these patients enter a critical (or plasma leakage) phase—often as fever declines—and develop the criteria for severe dengue: a definite or presumptive diagnosis of dengue infection plus either (1) plasma leakage severe enough to cause shock or respiratory distress, or (2) severe bleeding, or (3) severe organ dysfunction. Early recognition and action long before this critical period are crucial to initiate appropriate supportive care. Indeed, all patients with a presumptive diagnosis of dengue (of any severity) should be initially assessed for designated warning signs (dengue with warning signs, directing inpatient management) as well as for the criteria for severe dengue (i.e., a provisional diagnosis of severe dengue should be made clinically).

The complex determinants of risk for this progression to severe dengue include host and viral factors but center most notably around the potential for immune-mediated enhancement of disease. Several weeks after convalescence from infection with DENV-1–4, the transient protection conferred by that infection against reinfection with a heterotypic DENV usually wanes. Heterotypic reinfection may result in classic dengue without/with warning signs or, less commonly, in severe dengue. In the past 20 years, YF mosquitoes have progressively reinvaded Latin America and other areas, and frequent travel by infected individuals has introduced multiple variants of DENV-1–4 from many geographic areas. Thus, the pattern of hyperendemic transmission of multiple DENV serotypes established in the Americas and the Caribbean has led to the emergence of severe dengue as a major problem. Among the millions of DENV-1–4 infections, ≈500,000 cases of severe dengue occur annually, with a CFR of ≈2.5%. The induction of vascular permeability and shock depends on multiple factors, such as the presence or absence of enhancing and nonneutralizing antibodies, age (susceptibility to severe dengue drops considerably after 12 years of age), sex (females are more often affected than males), race (white people are more often affected than black people), nutritional status, and the timing and sequence of infections (e.g., DENV-1 infection followed by DENV-2 infection seems to be more severe than DENV-4 infection followed by DENV-2 infection). The presence of neutralizing antibodies is associated with decreased viremia on subsequent infection. In addition, considerable heterogeneity exists among variants in each DENV population. For instance, Southeastern Asian DENV-2 variants have more potential to cause severe dengue than do other variants. Recent evidence points to a key role for the DENV NS1 protein in the vascular leak phenomenon associated with severe dengue.

In milder cases of severe dengue, restlessness, lethargy, thrombocytopenia (<100,000/μL), and hemoconcentration are detected 2–5 days after the onset of typical dengue, usually at the time of defervescence. The maculopapular rash that often develops in dengue (without/with warning signs) may also appear in severe dengue. However, severe dengue is most notoriously identified as the consequence of a vascular leak syndrome leading to intravascular volume depletion, hypoalbuminemia, serosal effusions (pleural, ascitic), and, in severe cases, circulatory collapse (i.e., shock, typically lasting 2–3 days), often with an accompanying narrowed pulse pressure, hepatomegaly, and cyanosis. Bleeding tendencies (evidenced by a positive tourniquet test and petechiae) or overt bleeding in the absence of underlying causes (e.g., preexisting gastrointestinal lesions) may be detected but are less common in children. Organ involvement may include mild hepatic injury, CNS abnormalities (e.g., altered mental status, seizures), cardiac abnormalities (e.g., arrhythmias), renal disturbances (e.g., acute kidney injury), and ocular dysfunction.

A virologic diagnosis of severe dengue can be made by the usual means (nucleic acid amplification or antigen detection) in the first 5 days of infection, after which diagnosis relies on serologic testing. Combination testing—point-of-care rapid tests for NS1 antigen and IgM antibody assays—is increasingly used in the clinical setting. The early detection of IgG (starting at 4 days after illness onset, rapidly rising) suggests a secondary rather than primary infection, in which IgG is detected later and at much lower titers. However, serologic cross-reactivity after infection with (or vaccination for) antigenically similar orthoflaviviruses (e.g., YFV, ZIKV, and Japanese encephalitis virus) confounds diagnosis.

As above, patients with severe dengue should be hospitalized and rapidly triaged (and then closely monitored) for hemodynamic instability and bleeding. In general, appropriate IV fluids should be administered as early as possible, preferably before shock develops. Based on clinical parameters (vital signs, urine output, hematocrit, bleeding, and signs of volume overload), algorithmic approaches have been developed to standardize appropriate fluid management (crystalloid vs colloid vs blood rate) and monitoring of patients with and without shock and/or bleeding. CFR varies greatly depending on early recognition and quality of treatment. However, most patients with severe dengue respond well to supportive therapy, and the overall CFR at experienced clinical centers in the tropics is probably as low as 1%. When patients enter the convalescent phase, significant fluid reabsorption (of extravascular fluid and effusions) occurs as the plasma leak resolves, blood pressure improves, and diuresis occurs. Notably, a patchy erythematous pruritic rash may develop and desquamate during recovery. Significant postacute sequelae include prolonged fatigue and depression.

The key to control of both dengue (without/with warning signs) and severe dengue is the control of YF mosquitoes, which also reduces the risk of urban YF and CHIKV circulation. Control efforts have been handicapped by the presence of nondegradable tires and long-lived plastic containers in trash repositories—creating perfect mosquito breeding grounds upon filling with water during rainfall—and by insecticide resistance. Urban poverty and an inability of the public health community to mobilize the human population to respond to the need to eliminate mosquito breeding sites are also factors in lack of mosquito control. New approaches that may be considered in the future of vector control include the release of *Aedes* mosquitoes infected with *Wolbachia* or carrying dominant lethal genetic mutations that will be passed on to offspring. A tetravalent live-attenuated dengue vaccine based on the attenuated YFV 17D platform (CYD-TDV, or Dengvaxia) was licensed in 2015 and registered in 20 countries for individuals 9–45 years of age. However, retrospective analysis of phase 3 trials in Latin America and Asia suggested protection from severe dengue only in previously seropositive individuals; indeed, the risk of severe dengue was actually increased in seronegative vaccine recipients over that in nonvaccinated seronegative individuals, suggesting that a "first serologic hit" from the vaccine predisposes naïve recipients to more severe primary dengue infection. Strategic revision to avoid vaccine-enhanced disease pivoted to include prevaccination serologic screening aimed to avoid vaccination of seropositive individuals; however, the requirement for prevaccine testing and decision-making is challenging. Recently, live-attenuated candidate vaccines based on modified recombinant DENVs have been or are being evaluated, with these learnings in mind. Qdenga (TAK-3), given in two doses 3 month apart, has shown overall efficacy against infection and hospitalization in seropositive and seronegative children, although not against DENV-3 (in fact, a very small signal of potentially increased risk of hospitalization after DENV-3 infection was observed), and the trial did not include DENV-4 infections. This vaccine has been approved for use in several European countries. Recommendations globally are currently for introduction in settings with high transmission intensity for maximal public health impact while reducing potential risk in seronegative individuals. Further postmarket evaluation is ongoing. More recently, a phase 3 trial of Butantan-DV, a single-dose tetravalent vaccine, was shown to prevent symptomatic DENV-1 and DENV-2 infection; in the absence of DENV-3 and DENV-4 infections during the trial, efficacy against these serotypes is unknown. TVOO3, a single-dose vaccine designed to balance DENV serotype-specific components, is currently undergoing phase 3 clinical evaluation; similar concerns about safety are being addressed.

YELLOW FEVER YFV had caused major epidemics in Africa and Europe before its transmission by YF mosquitoes was discovered in 1900. Urban YF became established in the Americas because of colonization with YF mosquitoes—originally an African mosquito. Subsequently, different types of mosquitoes and nonhuman primates were found to maintain YFV in Africa and in Central and South American rain forests. Transmission to humans is incidental, occurring via bites from mosquitoes that have fed on viremic monkeys. After the identification of *Ae. aegypti* as the vector of YF, containment strategies were aimed at increased mosquito control. Today, urban YFV transmission occurs only in some African cities, but the threat exists in the cities of South America, where reinfestation by YF mosquitoes has taken place, and DENV-1–4 transmission by these mosquitoes is common. Despite the existence of a highly effective and safe vaccine, several hundred jungle cases of YF occur annually in South America, and 84,000–170,000 severe jungle and urban cases (resulting in 29,000–60,000 fatal outcomes) occurred in Africa in 2013 alone. In 2016, a large urban outbreak (Luanda, Angola) spilled over to generate local transmission in large cities in neighboring countries (e.g., Kinshasa, Democratic Republic of the Congo) as well as travel-related cases in China; the signal of a global threat that included exportation to Asia stimulated ongoing efforts to identify and vaccinate highest-risk populations in 40 targeted countries in Africa and South America, to reactively vaccinate people in outbreak settings, and to increase measures to prevent exportation.

YF is a typical VHF notable for prominent hepatic necrosis. After an incubation period of 3–6 days, patients present with a nonspecific febrile illness (fatigue, myalgia, backache, headaches, photophobia, anorexia, nausea or vomiting) associated with viremia typically lasting 3–4 days. After defervescence, 10–15% of patients develop recrudescent fever and "intoxication" characterized by severe dysfunction of the liver and other organs. Hepatic failure leads to the characteristic jaundice, bleeding (gastrointestinal tract, nasopharyngeal mucosa), abdominal pain with nausea and vomiting, and hyperammonemic encephalopathy; acute kidney injury leads to oliguria, azotemia, and marked albuminuria; and myocardial injury and encephalitis have been described. Abnormalities in liver function tests range from modest elevations of hepatic aminotransferase activities in mild cases to severe liver injury, hyperbilirubinemia, and the synthetic dysfunction of acute hepatic failure. After early leukopenia, leukocytosis occurs as disease progresses, and coagulation abnormalities are common. Treatment is supportive only. Although most infections are subclinical, 50% of patients who enter the toxic phase die in the next 7–10 days.

Urban YF can be prevented by the control of YF mosquitoes. The continuing sylvatic cycles require vaccination of all visitors to areas of potential transmission with live-attenuated variant 17D vaccine virus, which cannot be transmitted by mosquitoes. With few exceptions, reactions to the vaccine are minimal; immunity is provided within 10 days and lasts for at least 25–35 years. Recent meta-analytic review of 39 clinical studies suggests seroprotection is high (>90%) and perhaps lifelong after one vaccine dose in adults from nonendemic areas, raising questions about the utility of further boosting in this population. An egg allergy mandates caution in vaccine administration. Although fetal harm has not been documented, pregnant women should be immunized only if they are at risk of exposure to YFV. Because vaccination has been associated with several cases of encephalitis in children <6 months of age, it is contraindicated in this age group and not recommended for infants 6–8 months of age unless the risk of exposure is very high. Rare, serious, multisystemic adverse reactions (occasionally fatal), including vaccine-associated "viscerotropic" YF, have been reported, particularly affecting the elderly and those with congenital (mutations affecting interferon α [IFNα] signaling) or acquired (autoantibodies to IFNα, thymoma) impairment affecting viral control. The risk-to-benefit ratio should be weighed before vaccine administration to individuals ≥60 years of age. Nevertheless, the number of deaths of unvaccinated travelers with YF exceeds the number of deaths from vaccination, and a liberal vaccination policy for travelers to involved areas should be pursued. Timely information on changes in YF distribution and YF vaccine requirements can be obtained from the U.S. Centers for Disease Control and Prevention (*https://wwwnc.cdc.gov/travel/yellowbook/2024/preparing/yellow-fever-vaccine-malaria-prevention-by-country*).

ACKNOWLEDGMENT

The authors gratefully acknowledge the major contributions of Clarence J. Peters and additional contributions by Rémie N. Charrel to this chapter in previous editions and thank Anya Crane (IRF-Frederick) for editing the manuscript.

FURTHER READING

Centers for Disease Control and Prevention: Arbovirus catalog. Available at *https://wwwn.cdc.gov/arbocat/*. Accessed February 4, 2024.

Howley PM, Knipe DM (eds): *Fields Virology. Volume 1: Emerging Viruses*, 7th ed. Philadelphia, Wolters Kluwer/Lippincott Williams & Wilkins, 2020.

International Committee on Taxonomy of Viruses (ICTV): Virus taxonomy: The ICTV report on virus classification and taxon nomenclature. Available at *https://ictv.global/report*. Accessed February 4, 2024.

Lvov DK et al: *Zoonotic Viruses of Northern Eurasia: Taxonomy and Ecology*. London, Elsevier/Academic Press, 2015.

Singh SK, Ruzek D (eds): *Viral Hemorrhagic Fevers*. Boca Raton, FL, CRC Press, 2013.

Vasilakis N, Gubler DJ (eds): *Arboviruses: Molecular Biology, Evolution and Control*. Haverhill, UK, Caister Academic Press, 2016.

WEBSITE

International Committee on Taxonomy of Viruses (ICTV). *https://ictv.global/*. Accessed February 4, 2024.

216

Filovirids: Orthoebolavirus and Orthomarburgvirus Infections

Jens H. Kuhn, Ian Crozier

Several viruses in the family *Filoviridae* cause severe infections in humans that are often fatal. Introduction of filovirids into human populations is an extremely rare event that most likely occurs by direct or indirect contact with reservoir hosts (known and unknown) or by contact with sick or deceased filovirid-infected mammals. Filovirids are highly infectious but not exceptionally contagious. Human-to-human transmission occurs through direct person-to-person contact or exposure to infected bodily fluids or tissues; there is no evidence of aerosol or respiratory droplet transmission in natural outbreak settings. Infections manifest initially with a nonspecific influenza-like febrile illness that rapidly progresses to commonly include gastrointestinal manifestations and, in severe illness, coagulopathy, multiple-organ dysfunction syndrome, shock, and death. Although the prevalence and source remain controversial, serologic footprints of subclinical acute filovirid infections have been identified since the first descriptions of filovirid disease outbreaks. Filovirid disease survivors may be persistently infected in immune-privileged tissue compartments, commonly in the male reproductive tract, but also in the central nervous system (CNS), and in intraocular tissues and fluids; notably, filoviral persistence has been rarely associated with the re-ignition of outbreaks. Historically, the prevention of filovirid infections has consisted primarily of tried-and-true epidemiologic approaches (e.g., isolation of cases, contact tracing, effective infection prevention and control, and safe burial practices). Treatment of disease traditionally consisted only of limited supportive clinical care (often constrained by in-field capacity); indeed, filovirid-specific vaccines or therapeutic agents had not been rigorously evaluated in humans prior to the 2013–2016/2021 outbreaks of Ebola virus disease (EVD) that occurred in Western Africa. Building on the knowledge gained in Western Africa and during the 2018–2020 EVD outbreak in the Democratic Republic of the Congo, prevention and treatment strategies now include the widespread deployment of an effective vaccine specific to Ebola virus (EBOV); the use of effective therapeutics based on virus-specific monoclonal antibodies (mAbs), which were identified in a first-of-its-kind randomized controlled trial; and the delivery of improved supportive care. Although these advances have essentially become new standards for prevention and treatment of EVD, the same cannot yet be said for other filovirid diseases.

Filovirids are categorized as World Health Organization (WHO) Risk Group 4 pathogens. Consequently, all work with material potentially containing replicating filovirids should be conducted in maximum containment (biosafety level 4) laboratories, or the viruses should be inactivated prior to activities in biosafety level 2/3 laboratories. These viruses must be handled by experienced personnel wearing appropriate personal protective equipment (PPE; see "Control and Prevention," below) and following rigorous standard operating procedures. In addition, when filovirid disease outbreaks are suspected or confirmed, the appropriate national authorities and WHO reference laboratories should be contacted immediately.

ETIOLOGY

The family *Filoviridae* includes nine official genera **(Figs. 216-1 and 216-2)**. Human pathogens are found in two of these genera, *Orthoebolavirus* and *Orthomarburgvirus*. Collectively, these pathogens cause "filovirid" ("filovirus") disease (FVD; International Classification of Diseases, Eleventh Revision [ICD-11], code 1D60). FVD is subdivided into "Ebola disease" (EBOD; ICD-11, code 1D60.0), caused by four of six classified orthoebolaviruses (Bundibugyo virus [BDBV], EBOV, Sudan virus [SUDV], and Taï Forest virus [TAFV]), and "Marburg disease" (MARD; ICD-11, code 1D60.1), caused by the two orthomarburgviruses, Marburg virus and Ravn virus. Accordingly, specific diseases are optimally named with the virus-specific cause such that, for example, "EVD" is distinguished from "Sudan virus disease (SVD)."

Mammalian filovirids have linear, nonsegmented, negative-sense RNA genomes that are ≈19 kb in length. These genomes contain seven genes that encode seven structural proteins: a nucleoprotein (NP), a polymerase cofactor (VP35), a matrix protein (VP40), a glycoprotein ($GP_{1,2}$), a transcriptional activator (VP30), a ribonucleoprotein complex-associated protein (VP24), and a large (L) protein that contains an RNA-directed RNA polymerase domain. Orthoebolaviruses, but not orthomarburgviruses, additionally encode three nonstructural proteins of unknown functions (sGP, ssGP, and Δ-peptide). Filovirions are unique among human virus particles in that they are predominantly pleomorphic filaments but also assume torus-like or 6-like shapes (width ≈91–98 nm; average length <1 μm). These enveloped virions contain helical ribonucleoprotein capsids and are covered with $GP_{1,2}$ spikes **(Fig. 216-3)**.

- Realm *Riboviria*
  - Kingdom *Orthornavirae*
    - Phylum *Negarnaviricota*
      - Subphylum *Haploviricotina*
        - Class *Monjiviricetes*
          - Order *Mononegavirales*
            - Family *Filoviridae*
              - Genus *Cuevavirus*
                - Species *Cuevavirus lloviuense*
                  - Virus: Lloviu virus (LLOV)
              - Genus *Dianlovirus*
                - Species "*Dianlovirus dehongense*" (proposed)
                  - Virus: Déhóng virus (DEHV)
                - Species *Dianlovirus menglaense*
                  - Virus: Měnglà virus (MLAV)
              - Genus *Loebevirus*
                - Species *Loebevirus percae*
                  - Virus: Lötschberg virus (LTBV)
              - Genus *Oblavirus*
                - Species *Oblavirus percae*
                  - Virus: Oberland virus (OBLV)
              - Genus *Orthoebolavirus*
                - Species *Orthoebolavirus bombaliense*
                  - Virus: Bombali virus (BOMV)
                - Species *Orthoebolavirus bundibugyoense*
                  - Virus: Bundibugyo virus (BDBV)
                - Species *Orthoebolavirus restonense*
                  - Virus: Reston virus (RESTV)
                - Species *Orthoebolavirus sudanense*
                  - Virus: Sudan virus (SUDV)
                - Species *Orthoebolavirus taiense*
                  - Virus: Taï Forest virus (TAFV)
                - Species *Orthoebolavirus zairense*
                  - Virus: Ebola virus (EBOV)
              - Genus *Orthomarburgvirus*
                - Species *Orthomarburgvirus marburgense*
                  - Virus 1: Marburg virus (MARV)
                  - Virus 2: Ravn virus (RAVV)
              - Genus *Striavirus*
                - Species *Striavirus antennarii*
                  - Virus: Xīlǎng virus (XILV)
              - Genus *Tapjovirus*
                - Species *Tapjovirus bothropis*
                  - Virus: Tapajós virus (TAPV)
              - Genus *Thamnovirus*
                - Species *Thamnovirus kanderense*
                  - Virus: Kander virus (KNDV)
                - Species *Thamnovirus percae*
                  - Virus: Fiwi virus (FIWIV)
                - Species *Thamnovirus thamnaconi*
                  - Virus: Huángjiāo virus (HUJV)

Filovirids that are known to infect humans are depicted in the same nonblack colors as in figures.

FIGURE 216-1 Filovirid taxonomy (2024).

EPIDEMIOLOGY

The majority of recorded FVD outbreaks, including the 2013–2016 EVD outbreak, can be traced back to single index patients who transmitted the infection to others. Although small outbreaks may not have been recorded historically, the epidemiology of EVD transmission chains suggests that only ≈50 natural host-to-human spillover events have occurred since the discovery of filovirids in 1967. However, outbreaks associated with reignition of human-to-human transmission from a persistently infected survivor of a previous outbreak have more recently been described. Outbreak frequency, size, and overall

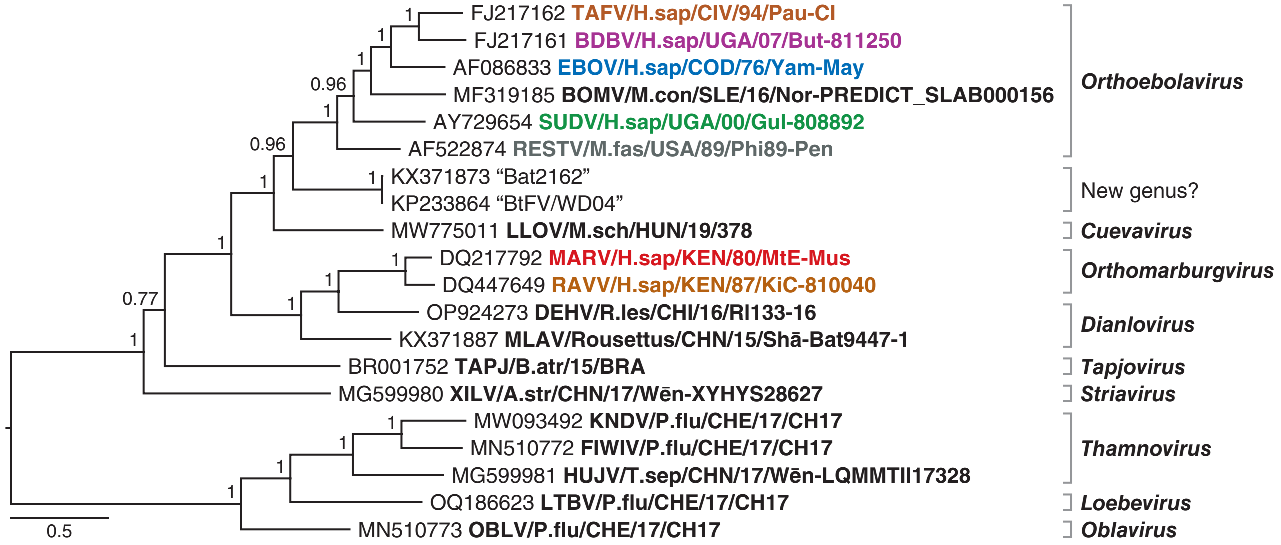
FJ217162 TAFV/H.sap/CIV/94/Pau-CI
FJ217161 BDBV/H.sap/UGA/07/But-811250
AF086833 EBOV/H.sap/COD/76/Yam-May
MF319185 BOMV/M.con/SLE/16/Nor-PREDICT_SLAB000156
AY729654 SUDV/H.sap/UGA/00/Gul-808892
AF522874 RESTV/M.fas/USA/89/Phi89-Pen
KX371873 "Bat2162"
KP233864 "BtFV/WD04"
MW775011 LLOV/M.sch/HUN/19/378
DQ217792 MARV/H.sap/KEN/80/MtE-Mus
DQ447649 RAVV/H.sap/KEN/87/KiC-810040
OP924273 DEHV/R.les/CHI/16/RI133-16
KX371887 MLAV/Rousettus/CHN/15/Shā-Bat9447-1
BR001752 TAPJ/B.atr/15/BRA
MG599980 XILV/A.str/CHN/17/Wēn-XYHYS28627
MW093492 KNDV/P.flu/CHE/17/CH17
MN510772 FIWIV/P.flu/CHE/17/CH17
MG599981 HUJV/T.sep/CHN/17/Wēn-LQMMTII17328
OQ186623 LTBV/P.flu/CHE/17/CH17
MN510773 OBLV/P.flu/CHE/17/CH17
Orthoebolavirus
New genus?
Cuevavirus
Orthomarburgvirus
Dianlovirus
Tapjovirus
Striavirus
Thamnovirus
Loebevirus
Oblavirus
0.96
0.96
0.77
0.5

FIGURE 216-2 Filovirid phylogeny/evolution. Midpoint-rooted maximum-likelihood tree inferred by using filovirid large gene (*L*) sequences. Bootstrap values are shown at each node. The scale bar indicates nucleotide substitutions per site. Tips of branches are labeled with GenBank accession numbers followed by filovirid isolate designation. Filovirids that are known to infect humans are depicted in the same nonblack colors as in other figures. BDBV, Bundibugyo virus; BOMV, Bombali virus; DEHV, Déhóng virus; EBOV, Ebola virus; FIWIV, Fiwi virus; HUJV, Huángjiāo virus; KNDV, Kander virus; LLOV, Lloviu virus; LTBV, Lötschberg virus; MARV, Marburg virus; MLAV, Měnglà virus; OBLV, Oberland virus; RAVV, Ravn virus; RESTV, Reston virus; SUDV, Sudan virus; TAFV, Taï Forest virus; TAPV, Tapajós virus; XILV, Xīlǎng virus. *(Adapted and expanded from JH Kuhn et al:* Filoviridae*, in Fields Virology, Vol 1, 7th ed, PM Howley et al (eds). Philadelphia, Wolters Kluwer/Lippincott Williams & Wilkins, 2020, pp 449–503. Analysis courtesy of Nicholas Di Paola, PhD, USAMRIID, Fort Detrick, MD, USA. Figure courtesy of Jiro Wada, NIH/NIAID/DCR/IRF-Frederick, Fort Detrick, MD, USA.)*

case-fatality rate are likely the result of complex interactions of the specific filovirid, the reservoir hosts (known and unknown), the susceptible human population (e.g., varying with age, unknown genetic determinants of susceptibility and disease severity, and risk behavior), and the geographic setting (e.g., local public health capacity, socioeconomic conditions, and cultural practices).

As of November 13, 2024, 35,608 cases of human filovirid disease and 15,886 deaths had been recorded **(Fig. 216-4)**. These numbers emphasize both the high case-fatality rate (number of deaths per number of sick people; 44.6%) and the overall low mortality (reflecting the impact on the healthy population) related to filovirid diseases. Of these totals, 28,652 cases and 11,325 deaths occurred during the 2013–2016 EVD outbreak in Western Africa; this was the largest of all recorded FVD outbreaks. Natural FVD outbreaks had not been considered a global threat until regional and then global spread during this outbreak challenged that tenet. Filovirids that are pathogenic in humans appear to be exclusively endemic to equatorial (Western, Middle, and Eastern) Africa **(Fig. 216-5)**, although this distribution may change if natural or artificial environmental alterations lead to filovirid host migration and increased contacts between nonhuman hosts and humans.

Outbreaks have been contained when high-risk activities (e.g., ritual washing as part of burial practices) have been limited or been made safer with appropriate infection prevention and control. Of particular importance is accessibility to health care centers with staff trained and equipped (e.g., with PPE) for adequate infection prevention and control, to limit both community spread and nosocomial amplification. The incidence of FVD may have increased over the past two decades (Fig. 216-4 and Fig. 216-5), but debate continues as to whether this increase is due to increased filovirid activity, more frequent human interaction with filovirid hosts, or improvement in surveillance capabilities.

FVD outbreaks are associated with distinct meteorologic and geographic conditions and are probably associated with distinct hosts or reservoirs. The four orthoebolaviruses that cause disease in humans appear to be endemic in humid rainforests. EVD outbreaks in particular have often been associated with hunting in forests or contact with bushmeat (i.e., meat from apes, other nonhuman primates, duikers, or bush pigs). Ecologic studies suggest that EBOV may play a role in extensive and often-fatal epizootics among wild ape populations. However, only one orthoebolavirus, TAFV, has been isolated from nonhuman primates in the wild. Orthomarburgviruses, on the other hand, seem to infect hosts inhabiting arid woodlands, and associated outbreaks have almost always been epidemiologically linked to individuals visiting or working in natural or engineered caves or mines. The cave-dwelling Egyptian rousette (*Rousettus aegyptiacus* (Geoffroy, 1810)), a pteropodid (fruit) bat, serves as a natural and subclinically infected reservoir for both Marburg virus and Ravn virus. Although bats are suspected hosts for orthoebolaviruses as

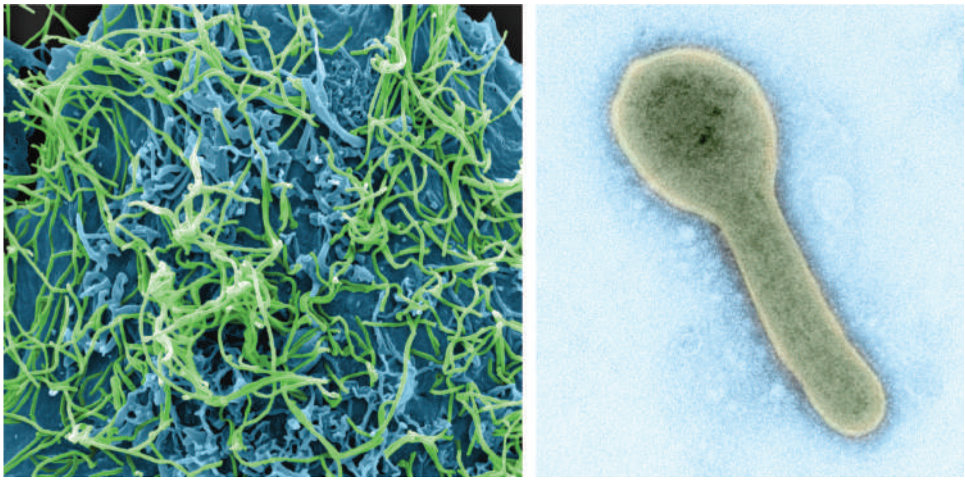

FIGURE 216-3 Ultrastructure of filovirions. ***Left:*** Colorized scanning electron micrograph of Ebola virus particles (*green*) attached to the surface of an infected grivet (*Chlorocebus aethiops* (Linnaeus, 1758)) Vero E6 producer cell (*blue*). ***Right:*** Colorized transmission electron micrograph of a Marburg virus particle collected from purified Vero E6 producer cell supernatant. *(Figure courtesy of John G. Bernbaum and Jiro Wada, NIH/NIAID/DCR/IRF-Frederick, Fort Detrick, MD, USA.)*

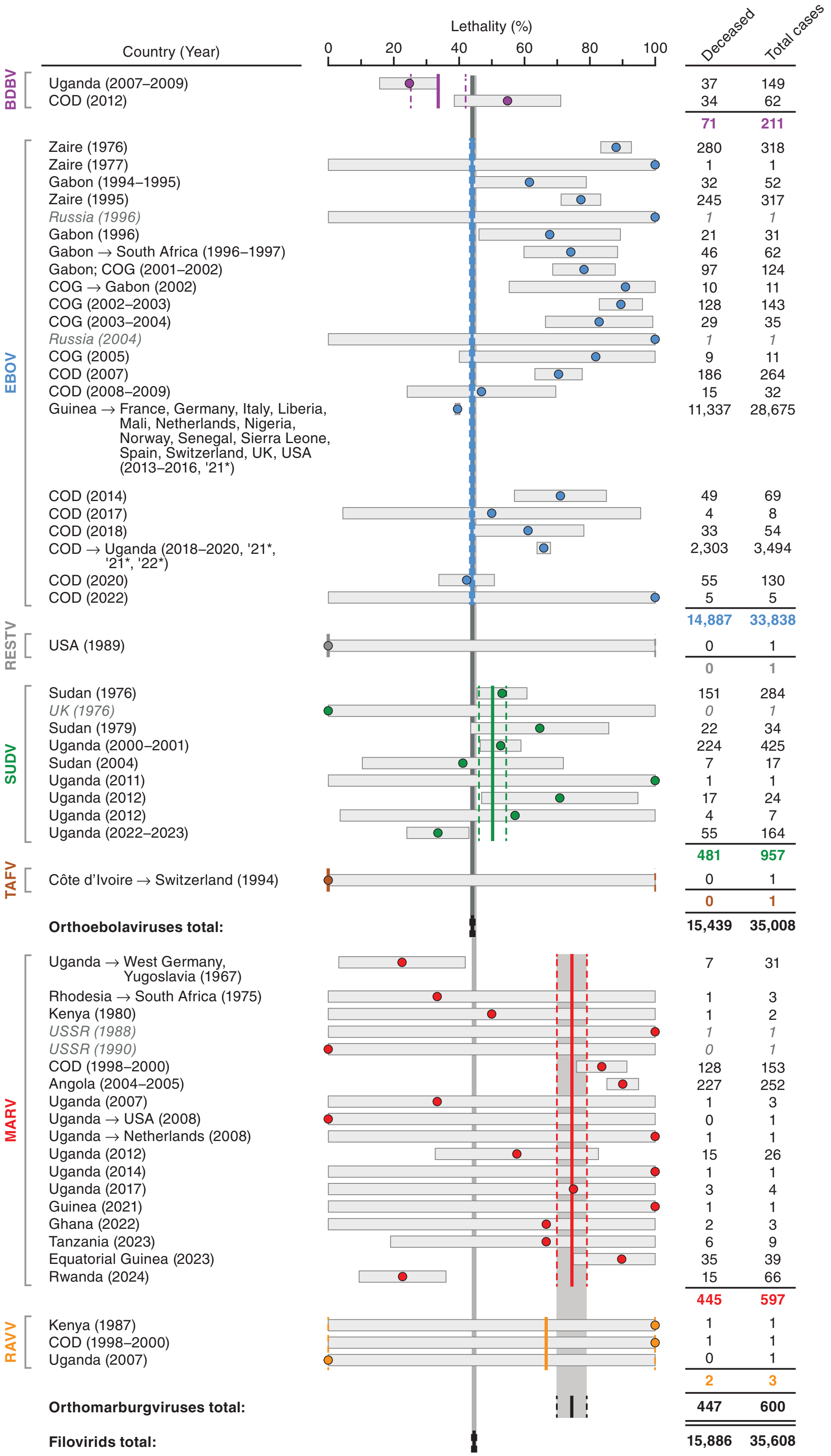

FIGURE 216-4 Characteristics of outbreaks of human filovirid disease. Seven of 17 classified filovirids have caused disease in humans. ***Left column:*** Outbreaks considered to be caused by independent zoonotic spillover are listed by virus in chronological order in the left column. Outbreaks considered to be related to reignition from a persistently infected survivor of a prior outbreak are included with the initial outbreak (denoted by asterisk). Laboratory infections are in *gray italicized* text. International case exportations are indicated with *arrows*. ***Right column:*** Numbers of fatal cases and total cases are summarized. ***Middle column:*** The case-fatality rate (*colored dots*) for each outbreak is plotted on a 0–100% scale along with 99% confidence intervals (*gray horizontal bars*). The overall case-fatality rate for disease caused by a particular virus is delineated by *vertical colored lines*, with *vertical colored dashed lines* indicating the corresponding 99% confidence intervals. The overall case-fatality rates for all *Orthoebolavirus*, all *Orthomarburgvirus*, and all filovirid disease outbreaks are shown by (underlaid) *vertical gray bars*. BDBV, Bundibugyo virus; COD, Democratic Republic of the Congo (formerly Zaire); COG, Republic of the Congo; EBOV, Ebola virus; MARV, Marburg virus; RAVV, Ravn virus; RESTV, Reston virus; SUDV, Sudan virus; TAFV, Taï Forest virus; UK, United Kingdom; USSR, Union of Soviet Socialist Republics (today Russia). *(Adapted and expanded from JH Kuhn et al: Evaluation of perceived threat differences posed by filovirus variants. Biosecur Bioterror 9:361, 2011. Figure courtesy of Jiro Wada, NIH/NIAID/DCR/IRF-Frederick, Fort Detrick, MD, USA.)*

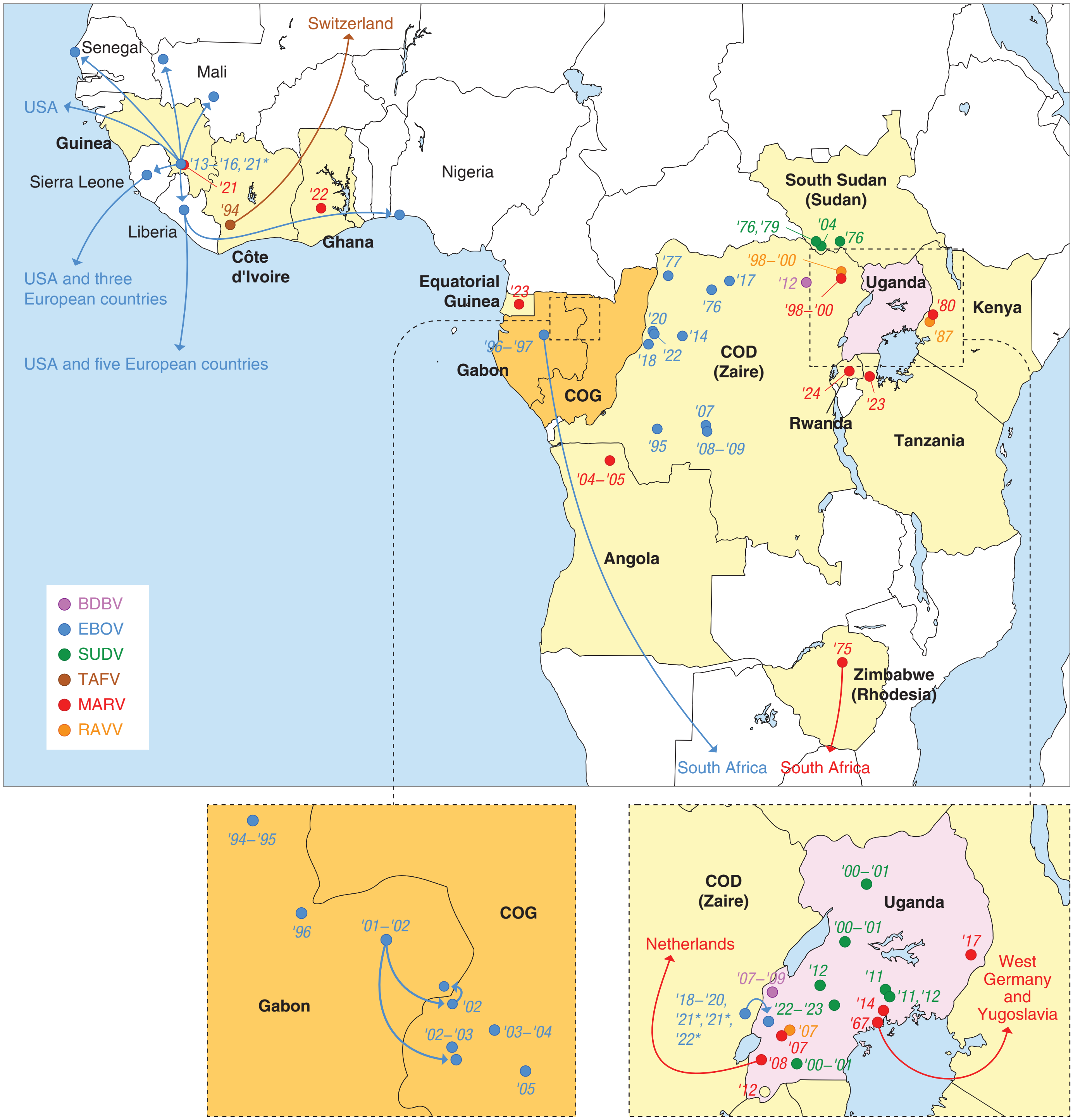
Switzerland
Senegal
Mali
USA
Guinea
'13–'16, '21*
'21
'94
'22
Sierra Leone
Nigeria
Liberia
Côte d'Ivoire
Ghana
USA and three European countries
USA and five European countries
Equatorial Guinea
'23
'96–'97
Gabon
COG
South Sudan (Sudan)
'76, '79
'04
'76
'77
'17
'98–'00
'12
'98–'00
Uganda
'76
'80
Kenya
'87
'20
'14
'18
'22
COD (Zaire)
'24
'23
Rwanda
'07
'95
'08–'09
Tanzania
'04–'05
Angola
'75
Zimbabwe (Rhodesia)
South Africa
South Africa
BDBV
EBOV
SUDV
TAFV
MARV
RAVV
'94–'95
'96
'01–'02
COG
'02
Gabon
'02–'03
'03–'04
'05
COD (Zaire)
'00–'01
Uganda
'00–'01
Netherlands
'17
West Germany and Yugoslavia
'07–'09
'12
'11
'18–'20, '21*, '21*, '22*
'22–'23
'11, '12
'14
'07
'67
'07
'08
'00–'01
'12

FIGURE 216-5 Geographic distribution of human filovirid disease outbreaks and years of occurrence. *Arrows* indicate international case exportations. BDBV, Bundibugyo virus; COD, Democratic Republic of the Congo (formerly Zaire); COG, Republic of the Congo; EBOV, Ebola virus; MARV, Marburg virus; RAVV, Ravn virus; SUDV, Sudan virus; TAFV, Taï Forest virus. *(Figure courtesy of Jiro Wada, NIH/NIAID/DCR/IRF-Frederick, Fort Detrick, MD, USA.)*

well, definitive proof is lacking. To date, the nonpathogenic Bombali virus is the only orthoebolavirus for which coding-complete genome nucleic acid has been detected in bats. EBOV and Reston virus have been loosely connected to frugivorous and insectivorous bats by means of antibody or genome fragment detection, whereas the hosts of BDBV, SUDV, and TAFV are enigmatic.

PATHOGENESIS

Human infections typically occur through direct exposure of skin lesions or mucosal surfaces to contaminated body fluids or material or by parenteral inoculation (e.g., via accidental needlesticks or reuse of needles in poorly equipped hospitals). Numerous studies, both in vitro and in vivo (in several animal models of human disease), have illuminated aspects of FVD pathogenesis **(Fig. 216-6)**. The $GP_{1,2}$ spikes on the surface of filovirions determine their cell and tissue tropism by engaging cell-surface molecules and the intracellular filovirid receptor NPC intracellular cholesterol transporter 1 (NPC1).

One of the hallmarks of filovirid pathogenesis is a pronounced modulation and dysregulation of immune responses. The first targets of filovirions are local macrophages, monocytes, and dendritic cells. Several structural proteins of filovirions (i.e., VP35, VP40, and/or VP24) then suppress intrinsic and innate immune responses by, for example, inhibiting the type I interferon antiviral pathways. This immunomodulation ultimately enables a productive filovirid infection, resulting in very high viral titers (>10^6 PFU/mL of serum in humans) with dissemination to most tissues. In tissues, filovirions infect additional

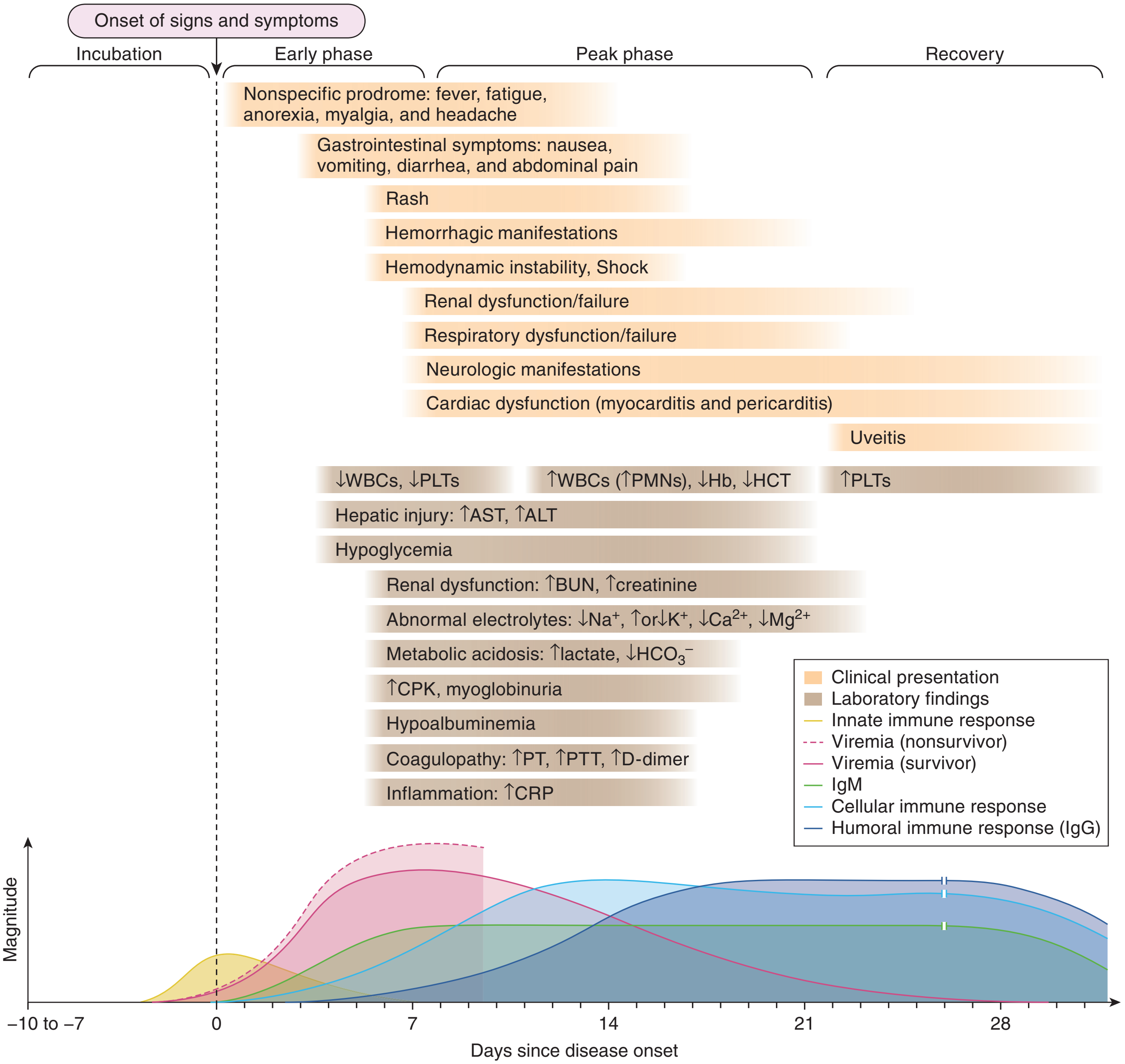

FIGURE 216-6 Ebola virus disease course. ALT, alanine aminotransferase; AST, aspartate aminotransferase; BUN, blood urea nitrogen; CPK, creatine phosphokinase; CRP, C-reactive protein; Hb, hemoglobin; HCT, hematocrit; PLTs, platelets; PMNs, polymorphonuclear leukocytes; PT, prothrombin time; PTT, partial thromboplastin time; WBCs, white blood cells. *(Adapted and expanded from JH Kuhn et al:* Filoviridae, *in Fields Virology, 7th ed, Vol. 1. Howley PM et al (eds). Philadelphia, Wolters Kluwer/Lippincott Williams & Wilkins, 2020, pp 449–503. Figure courtesy of Jiro Wada, NIH/NIAID/DCR/IRF-Frederick, Fort Detrick, MD, USA.)*

phagocytic cells, including other macrophages (alveolar, peritoneal, and pleural macrophages; Kupffer cells in the liver; and microglia in the CNS), epithelial cells (e.g., adrenal cortical cells and hepatocytes), stromal cells (fibroblasts), and endothelial cells. Infection is cytolytic in some—but not all—infected cells; e.g., hepatocyte necrosis likely contributes to elevated aminotransferase activities, and hepatic synthetic dysfunction contributes to coagulopathy. Infection leads to the secretion of soluble signaling molecules (varying with the cell type) that likely contribute to forward dysregulation of immune responses and ultimately to multiorgan dysfunction syndromes. For instance, infected macrophages secrete proinflammatory cytokines, leading to further recruitment of monocyte-derived macrophages to the site of infection. In contrast, infected dendritic cells are not activated to secrete cytokines, and the expression of major histocompatibility class II antigens is partially suppressed, with consequently deficient antigen presentation. Immunosuppression also occurs in part by massive lymphoid depletion in lymph nodes, spleen, and thymus in the absence of effective humoral and cell-mediated immune responses, especially in severe or fatal disease. Results from animal studies suggest that depletion is a direct consequence of considerable lymphocyte death; this explanation would also account for the severe lymphopenia that develops in patients. In addition to potential florid filovirid dissemination, another consequence of lymphocyte depletion may be susceptibility to secondary bacterial and fungal infections.

Other pathogenic hallmarks of filovirid infections include coagulopathy and endothelial dysfunction. Along with hepatic synthetic dysfunction, disseminated intravascular coagulation may contribute though this remains controversial. Thrombocytopenia, increased concentrations of tissue factor, consumption of clotting factors, increased concentrations of fibrin degradation products (D-dimers), and declining concentrations of protein C are typical features of severe disease. Consequently, fibrin deposition and microthrombotic small-vessel occlusion and necrotic/hypoxic infarction may occur in some tissues, particularly in the gonads and, less often, in the kidneys and spleen. In addition, petechiae, ecchymoses, extensive visceral effusions, and other hemorrhagic signs are observed in internal organs, mucous membranes, and skin. However, actual severe blood loss is a rare event (although it frequently occurs during or after childbirth). Most likely, aberrance in cytokines or other factors, such as nitric oxide, and direct infection and activation of endothelial cells are responsible for upregulated permeability of blood-vessel endothelia. This upregulation leads to fluid redistribution and interstitial tissue edema and hypovolemic or septic shock are common features of severe disease.

Despite this long list of pathogenetic hallmarks, increasing evidence from clinical settings suggests that effective filovirid-specific adaptive immune responses do develop, coinciding with control and clearance of viremia and subsequent clinical improvement in surviving patients. However, depending on the severity of illness (including organ dysfunction and late complications), clinical illness may be protracted and recovery incomplete.

CLINICAL MANIFESTATIONS

Diseases caused by orthoebolaviruses and orthomarburgviruses present very similar clinical phenotypes that cannot be distinguished at the bedside and, for all practical purposes, may be considered the same disease; this approach may change as higher-resolution characterization of human FVD accrues. The prevalence and character of clinical signs do not differ significantly in disease caused by disparate filovirids (except for the possibly apathogenic Reston virus), although, with the exception of the 2013–2016 EVD outbreak, the numbers of deeply characterized clinical observations are very low. The incubation period is 2–25 days (most commonly 7–10 days), after which infected people develop a nonspecific influenza-like syndrome characterized by sudden onset of fever and chills, severe headaches, cough, myalgia, pharyngitis, arthralgia of the larger joints, development of a maculopapular rash, and other symptoms/signs. A subsequent second phase (≈5–7 days after disease onset and thereafter) involves the gastrointestinal (nausea and vomiting and/or diarrhea, sometimes with abdominal pain), respiratory (chest pain, cough, and/or dyspnea), cardiovascular (hypotension and/or edema), and central nervous (confusion, headache, and/or coma) systems. Common hemorrhagic manifestations include subconjunctival injection, petechial rash, gingival bleeding, and bleeding at injection sites; epistaxis, hematemesis, hematuria, and melena occur but are less common. Patients usually succumb to acute disease 4–14 days after infection, often with severe multiorgan dysfunction that commonly includes shock and acute renal or respiratory failure.

Typical laboratory findings are leukopenia (with cell counts as low as 1000/μL) with a left shift prior to leukocytosis, thrombocytopenia (with counts as low as 50,000/μL), increased activities of liver enzymes (aspartate aminotransferase > alanine aminotransferase, γ-glutamyltransferase), increased creatinine and urea concentrations with proteinuria, electrolyte derangement (hypokalemia or hyperkalemia, hyponatremia, and/or hypocalcemia), hypoglycemia, hypoalbuminemia, prolonged prothrombin and partial thromboplastin times, and elevated creatine phosphokinase activities. Nonspecific markers of systemic inflammation (e.g., C-reactive protein concentrations) may be markedly elevated in severely ill patients.

DIAGNOSIS

Filovirid disease cannot be diagnosed based on clinical presentation alone. Numerous diseases common in equatorial Africa need to be considered in the differential diagnosis of a febrile patient. Almost all of these occur at a much higher incidence than FVD and are much more likely diagnoses in non-outbreak settings; however, during outbreaks and in peri-outbreak periods, timely and accurate laboratory diagnosis to rule in or rule out filovirid infection is critical. The most important infectious disease mimics of FVD are falciparum malaria and typhoid fever; also important are enterohemorrhagic *Escherichia coli* enteritis, gram-negative septicemia (including shigellosis), meningococcal septicemia, rickettsial infections, fulminant viral hepatitis, leptospirosis, measles, and other high-consequence viral infections (in particular, Lassa fever and yellow fever). Rarer noninfectious possibilities, including venomous snakebites, warfarin intoxication, and the many causes of acquired or inherited coagulopathy, also must be considered in the bleeding patient. An exposure history—including exposure to caves or mines; direct contact with bats, nonhuman primates, or "bushmeat"; direct contact with severely ill local residents; or admission to rural hospitals with patient-to-patient or patient-to-health-care-worker clusters of illness—should raise the index of suspicion.

If FVD is suspected based on epidemiologic and/or clinical manifestations, infectious disease specialists and the proper public health authorities should be notified immediately. Laboratory diagnosis of FVD is relatively straightforward but ideally requires the capacity for maximum containment, which usually is not available in filovirid-endemic countries. Increasingly, laboratory diagnosis is performed with patient samples inactivated in mobile field "glove boxes" by on-site personnel trained in the safe use of diagnostic assays adapted for field use in lower-containment settings. Consequently, diagnostic samples should be collected and processed with great caution and with the use of appropriate PPE and strict barrier techniques. With adherence to established biosafety precautionary measures, samples should be sent in suitable transport media to national or international reference laboratories. Acute-phase blood/serum is the preferred diagnostic specimen because it usually contains high titers of filovirions.

The current assay of choice for the diagnosis of filovirid infection is reverse-transcription polymerase chain reaction (RT-PCR) targeting one or more filovirid genes; a typical detection limit is 1000–5000 PFU/mL of serum, depending on the assay. When available, safe, rapid-turnaround, and standardized in-field PCR-based approaches (e.g., the Cepheid GeneXpert platform) are now standardly deployed during outbreaks. Antigen-capture enzyme-linked immunosorbent assays (ELISAs) or immunochromatographic assays for the detection of filovirion proteins are in development and may be useful as point-of-care rapid diagnostic tests (RDTs) in the future; sensitivity limitations of EBOV-specific antigen-based RDTs restrict current use to diagnosis of postmortem samples only. Direct immunoglobulin M (IgM) capture, direct IgG capture, or IgM-capture ELISAs (including multiplex approaches) are used for the detection of filovirion-targeting antibodies from patients in the later stages of disease (i.e., those who have been able to mount a detectable antibody response), including survivors. All of these assays can be conducted on samples treated with guanidinium isothiocyanate (for RT-PCR), cobalt-60 irradiation (for ELISA), or other effective measures (heat or irradiation treatment) that render filovirids noninfectious. Virus isolation in cell culture and plaque assays for quantification or diagnostic confirmation are relatively simple but must be performed in maximum-containment laboratories. If available, electron microscopic examination of inactivated samples or cultures for the unique filamentous structures of filovirions can further support diagnosis (Fig. 216-3). Formalin-fixed skin biopsy samples and possibly skin swabs can be useful for safe postmortem diagnoses. In-field (or near in-field) rapid genome sequencing was first deployed to inform classic epidemiology in Western Africa during the 2013–2016 EVD outbreak and has become a mainstay of outbreak control and response, even in endemic settings.

TREATMENT

Filovirid Disease

Clinical management of patients with suspected or confirmed filovirid disease should be conducted by health care workers who have been well-trained in the complex care of FVD patients and who are using appropriate PPE in care environments with appropriate infection prevention and control measures (see "Control and Prevention," below). Historically, management of FVD had been entirely supportive (and even that limited by resource constraints) because the efficacy and safety of specific antiviral countermeasures had not been rigorously studied outside of animal models of disease. The 2013–2016 EVD outbreak in Western Africa highlighted the need to conduct rigorous, feasible, and ethical clinical research during outbreaks. Building on challenges and lessons learned in that setting, the first-of-its-kind Pamoja Tulinde

Maisha (PALM) randomized clinical trial was conducted during the 2018–2020/2021 EVD outbreak in the Democratic Republic of the Congo and identified two therapeutics based on mAbs specific to EBOV to improve survival rates. mAb114 (ansuvimab-zykl) and REG-EB3 (atoltivimab, maftivimab, and odesivimab-ebgn) were subsequently approved for the treatment of EVD in PCR-confirmed adults (including pregnant women) and children. Both are administered via carefully monitored single-dose infusions initiated as soon as possible after diagnosis. In addition, efforts during this outbreak demonstrated the will and capacity to deliver more advanced supportive and critical care accompanying specific therapeutics in the African outbreak setting. Although evidence is lacking, consensus treatment strategies include those generally recommended for severe septicemia/sepsis/shock **(Chap. 315)** and should be applied with an emphasis on standard approaches—i.e., monitoring and response to respiratory dysfunction (e.g., oxygen), circulatory dysfunction (e.g., intravascular fluid repletion and vasopressor support), and CNS dysfunction (e.g., ruling out of reversible causes, notably hypoglycemia)—as well as the detection and management of acute kidney injury, hemorrhage, electrolyte derangements, and nutritional status and the prevention and treatment of secondary or co-infections. Pain management and administration of antipyretics, antiemetics, and antidiarrheal agents may be considered. Crucial strategies to improve outcomes in the most severely ill FVD patients include preventing organ dysfunction and providing safe and effective temporary organ support (e.g., mechanical ventilation and renal replacement) to expand the window for administration of medical countermeasures and development of effective endogenous immune responses.

COMPLICATIONS

Even in patients who have initial virologic or clinical improvement, complications in the second or third week of illness may include secondary infections; persistent renal dysfunction and electrolyte abnormalities; respiratory insufficiency; neurologic compromise (e.g., EBOV-related meningoencephalitis, cerebrovascular events, and seizures); cardiac dysfunction (e.g., myocarditis and pericarditis); and venous thrombosis. Pregnancy and labor are historically associated with severe complications and poor outcomes in filovirid disease due to clotting factor consumption, fetal loss, and/or severe blood loss during birth; in the era of EBOV-specific therapeutics and optimized supportive care, maternal and fetal outcomes are improved.

A number of sequelae have been self-reported or historically described in survivors of FVD, including prolonged and sometimes incapacitating arthralgia and myalgia, asthenia, alopecia, visual problems (including uveitis), hearing loss, memory loss and neurocognitive dysfunction, mental health conditions (anxiety, depression, and posttraumatic stress disorder), and reproductive problems. Well-controlled observational studies of EVD survivors were first conducted in the aftermath of the 2013–2016 Western Africa EVD outbreak. Compared with their close-contact controls, Liberian EVD survivors had an increased incidence of headache, arthralgia and myalgia, memory loss, fatigue, and urinary frequency as well as abnormal results in abdominal, chest, neurologic, musculoskeletal, and ocular exams. Of individual and public health significance is the potential for persistence of filovirids in immune-privileged tissue compartments (and their associated fluids) in FVD survivors, most commonly in the semen of male survivors (with the rare but documented potential for sexual transmission) and rarely in the CNS (causing recrudescent meningoencephalitis), the eye (causing recrudescent uveitis), and the placenta (causing transmission or placental insufficiency). True relapses that resemble the course of primary FVD are extremely rare but have been described. The risk of the remote reignition of outbreaks from persistently infected FVD survivors is apparently rare but obviously consequential. Notably, these events have only been detected after the two largest EVD outbreaks, suggesting a likely limitation, with the caveat that available samples and genomic tools to determine outbreak relatedness have historically not been available. EVD outbreaks in the Democratic Republic of the Congo in 2021 were genomically linked to the previous 2018–2020 outbreak. After a 5- to 7-year interim, an EVD outbreak in Guinea in 2021 was genomically linked to the 2013–2016 Western African outbreak; this event remains enigmatic as corroborating epidemiologic links have not been found.

PROGNOSIS

Among the most severe of acute viral diseases in humans, FVD generally has a poor prognosis, although with much greater heterogeneity than was historically assumed; i.e., the 70–90% case-fatality rate ascribed for many decades to EVD has required revision. With an incomplete evidence base, the outcome probably depends on factors that include the particular filovirid causing the infection (Fig. 216-4), host factors (age, immune status, coinfections, and unknown host genetic factors), virus exposure route and dose, viral load, the presence and severity of organ dysfunction, and—critically—the availability of filovirid-specific countermeasures and requisite supportive care. Despite the receipt of effective EBOV-specific therapeutics, outcomes in severe EVD patients with high viral loads and complex organ dysfunction remain suboptimal and will require future research to evaluate and advance virus-targeted (e.g., combination therapeutic), disease-targeted (e.g., immunomodulatory), and optimized supportive care strategies. After resolution of acute disease, the long-term health and survival outcomes for FVD survivors are unknown. It is uncertain how increased access to filovirid-specific therapeutics and life-saving support will impact filovirid persistence in survivors; short- and long-term surveillance will be necessary to avoid individual consequences (e.g., relapse and/or recrudescent inflammatory syndromes) and public health consequences (e.g., reignition of outbreak transmission chains in the peri-outbreak or post-outbreak period).

Although remarkable progress in understanding and improving outcomes in EVD has occurred in recent years, the same cannot yet be said for disease caused by filovirids other than EBOV, either in acute disease or in convalescence.

CONTROL AND PREVENTION

Prevention of viral exposure in nature is challenged by an incomplete understanding of filovirid ecology. To prevent orthomarburgvirus infection, people entering or living in areas where Egyptian rousettes can be found should avoid direct or indirect contact with these animals. Prevention in nature is more difficult in the case of pathogenic orthoebolaviruses, largely because ecologic reservoirs have not yet definitively been identified. EVD outbreaks have been associated with hunting or consumption of nonhuman primates more than exposure to bats. Indeed, the mechanism of introduction of orthoebolaviruses into nonhuman primate populations, if it occurs at all, is unclear. (Only one pathogenic orthoebolavirus, TAFV, has unequivocally been detected in wild nonhuman primates.) Therefore, to prevent orthoebolavirus infection, the only evidenced advice that can be offered to travelers and locals is to avoid contact with "bushmeat," nonhuman primates, and bats. In any setting, the local involvement of medical anthropologists and careful community engagement is strongly advised to ensure preventive risk communication is not received as threatening or patronizing.

Biomedical prevention strategies have historically been limited to tried-and-true pillars of outbreak control, centering on the identification and isolation of cases, contact tracing, ensuring that health care workers and other response personnel have appropriate training and capacity in infection prevention and control, and preventing high-risk transmission events. Measures aimed at preventing and controlling infection, including relatively simple barrier nursing techniques, vigilant use of appropriate PPE, quarantine, and contact tracing, usually effectively terminate or at least contain FVD outbreaks. Isolation of infected people and their contacts and avoidance of direct person-to-person contact without appropriate PPE usually prevent further spread, as the virions are not transmitted through droplets or aerosols under natural conditions. Typical protective gear sufficient to prevent filovirid infections consists of disposable gloves, gowns, and shoe covers and a face shield and/or goggles. If available, N-95 or N-100 respirators may be used to further limit infection risk. Positive-air-pressure respirators should be considered for high-risk medical procedures, such as intubation or suctioning. Medical equipment used in the care of an infected patient, such as gloves or syringes, should never be reused. Because filovirions are enveloped, disinfecting with detergents (e.g., 1% sodium deoxycholate, diethyl ether, or phenolic compounds) is relatively straightforward. Bleach solutions are recommended at 1:100 for surface disinfection and 1:10 for application to excreta or corpses. Whenever possible, potentially contaminated materials should be autoclaved, irradiated, or destroyed.

Emerging from research conducted during the 2013–2016 EVD outbreak in Western Africa, a vaccine based on a recombinant vesicular stomatitis Indiana virus expressing EBOV $GP_{1,2}$ (rVSV-ZEBOV/Ervebo) was the first filovirid vaccine approved for use in the United States and the European Union (EU). It is now widely deployed in both a reactive-ring vaccination strategy, targeting close contacts and their contacts in EVD outbreak settings, and for the preexposure vaccination of health care workers in at-risk regions. More recently, a heterologous dose vaccine candidate incorporating EBOV $GP_{1,2}$ into an adenovirus vector (Ad26.ZEBOV-GP/Zabdeno) followed by a vaccinia virus vector incorporating multiple filovirid antigens (MVA-BN-filo/Mvabea) has been shown to be safe and immunogenic in humans. Though evaluation of efficacy in a clinical trial has not been possible, immunobridging data gained during nonhuman primate experimentation led to regulatory authorization in the EU under "exceptional circumstances"; the two-dose requirement likely limits current use to proactive preexposure prevention in "peri-outbreak" settings rather than reactive "in-outbreak" reactive strategies. Development and evaluation of this and other vaccine candidates continue toward complementary preventive approaches for non-outbreak or peri-outbreak settings, with emphasis on the durability of immune responses and increases in preventive breadth toward other filovirids.

Even in the absence of high-level evidence, expert consensus informs the targeted use of EBOV-specific vaccine or postexposure prophylaxis (PEP) to prevent infection or disease in health care workers considered to have had a high-risk EBOV exposure (e.g., after needlestick injury). Evidence is needed to inform the use of PEP in high-risk contacts in the field outbreak setting. For male survivors, abstinence from sexual activity with a partner for at least 12 months after disappearance of clinical signs is recommended, unless testing proves semen to be free of filovirid RNA. (The use of condoms is generally recommended for all sexual activities.) Reproductive tract and CNS tissues, including ocular tissues and fluids from survivors, should be handled with appropriate precautions until demonstrated to be filovirid-RNA free. The role of filovirid-specific therapeutics in the prevention or treatment of filoviral persistence is unclear.

■ FURTHER READING

Cnops L et al: Essentials of filoviral load quantification. Lancet Infect Dis 16:e134, 2016.

Crozier I et al: The evolution of medical countermeasures for Ebola virus disease: Lessons learned and next steps. Vaccines (Basel) 10:1213, 2022.

Dudas G et al: Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature 544:309, 2017.

Hoenen T et al: Therapeutic strategies to target the Ebola virus life cycle. Nat Rev Microbiol 17:593, 2019.

Jacob ST et al: Ebola virus disease. Nat Rev Dis Primers 6:13, 2020.

Kuhn JH et al: *Filoviridae*, in *Fields Virology*, Vol 1, 7th ed, PM Howley et al (eds). Philadelphia, Wolters Kluwer/Lippincott Williams & Wilkins, 2020, pp 449–503.

Matz KM et al: Ebola vaccine trials: Progress in vaccine safety and immunogenicity. Expert Rev Vaccines 18:1229, 2019.

Mulangu S et al: A randomized, controlled trial of Ebola virus disease therapeutics. N Engl J Med 381:2293, 2019.

Regules JA et al: A recombinant vesicular stomatitis virus Ebola vaccine. N Engl J Med 376:330, 2017.

Section 16 Fungal Infections

217 Pathogenesis, Diagnosis, and Treatment of Fungal Infections

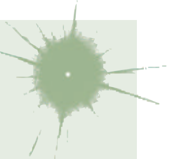

Michail S. Lionakis, John E. Edwards, Jr.

DEFINITION AND ETIOLOGY

In recent decades, human fungal infections have dramatically increased worldwide as a result of the AIDS pandemic, the widespread use of antibacterial agents, and the introduction of cytotoxic agents and precision medicine biologics for the treatment of autoimmune and neoplastic diseases and for use in patients undergoing solid organ transplantation or hematopoietic stem cell transplantation. Moreover, of great concern has been the recent rise in fungal infections caused by drug-resistant species, such as azole- and/or echinocandin-resistant *Candida glabrata* and *Candida auris* and azole-resistant *Aspergillus fumigatus*. Among the ~5 million fungal species, only a few cause human infections **(Table 217-1)**.

Fungal infections are classified as *mucocutaneous* and *deep organ* infections on the basis of anatomic location and as *endemic* and *opportunistic* infections on the basis of epidemiology. Mucocutaneous infections can cause serious morbidity but are rarely fatal. Deep organ infections cause severe illness and often carry a high mortality rate. The endemic mycoses are caused by fungi that are not part of the normal human microbiota but are environmentally acquired. The opportunistic mycoses are caused by fungi (*Candida, Aspergillus*) that often are components of the human microbiota and whose ubiquity in nature renders them easily acquired by immunosuppressed hosts (Table 217-1). Opportunistic fungi cause serious infections when impaired host immune responses allow the organisms to transition from commensals to invasive pathogens. Endemic fungi typically cause self-limited disease in immunocompetent hosts but severe illness in immunosuppressed patients.

Fungi are morphologically classified as *yeast*, *mold*, and *dimorphic*. Yeasts are seen as round single cells or budding organisms. Molds grow as filamentous forms called *hyphae* both at room temperature and in tissue. *Aspergillus*, Mucorales, and dermatophytes that infect skin and nails are mold fungi. Variations exist within this classification system. For instance, when *Candida* infects tissue, both yeasts and filamentous forms (pseudohyphae) may be present (except in the cases of *C. glabrata* and *C. auris*, which form only yeasts in tissue); in contrast, *Cryptococcus* exists only in yeast form. *Dimorphic* is the term used to describe fungi that have two forms; they grow as yeasts or large spherical structures in tissue but as filamentous forms at room temperature in the environment (Table 217-1).

Patients acquire deep organ infection by molds and endemic dimorphic fungi via inhalation. Skin dermatophytes are primarily environmentally acquired, but human-to-human transmission may also occur. The commensal *Candida* invades deep tissues from sites of mucocutaneous colonization, usually in the gastrointestinal tract or the skin in the case of *C. auris*.

In this chapter, we outline general principles of immunology, diagnosis, and treatment related to the most common human fungal infections.

■ PATHOGENESIS

In the past decade, our understanding of fungal recognition pathways and of tissue-specific innate and adaptive antifungal host defense mechanisms has markedly expanded. A major breakthrough has been the discovery and functional characterization of the C-type lectin receptor/spleen tyrosine kinase/caspase recruitment domain–containing protein 9 (CLR/SYK/CARD9) signaling pathway, which mediates fungal polysaccharide recognition and orchestrates proinflammatory mediator production, leukocyte recruitment, inflammasome

TABLE 217-1 Major Fungal Infections, Associated At-Risk Patient Populations, and Diagnostic Tests

INFECTION (MOST COMMON FUNGAL GENERA AND SPECIES)	CLINICAL SYNDROME(S)	RISK FACTOR(S)	DIAGNOSTIC TEST(S)
Mold (Filamentous) Fungi			
Aspergillosis (*Aspergillus fumigatus, A. terreus, A. flavus, A. niger, A. nidulans*[a])	Pneumonia or disseminated infection ABPA Keratitis	Neutropenia, glucocorticoids, HSCT, post-influenza or COVID-19, BTK inhibition Atopic individuals Direct inoculation	*Culture* of BAL fluid: low sensitivity, nonspecific (colonization, contamination) *Histologic examination* of tissue[b]: acute-angle septate hyphae *Biomarkers:* GM (BAL > serum); serum BDG (nonspecific)
Mucormycosis (*Rhizopus, Rhizomucor, Mucor, Cunninghamella,* and *Lichtheimia* spp.)	Sinopulmonary infection Rhinocerebral infection Necrotizing skin infection	Neutropenia, HSCT Diabetic ketoacidosis Direct inoculation (e.g., tornado victims)	*Culture* of BAL fluid or sinus tissue: very low sensitivity *Histologic examination* of tissue: ribbon-like aseptate hyphae *Biomarkers:* Negative
Fusariosis (*Fusarium solani, F. oxysporum*)	Pneumonia or disseminated infection Keratitis	Neutropenia Direct inoculation	*Culture* of tissue or blood: one of the few molds recovered from blood *Histologic examination* of tissue: acute-angle septate hyphae *Biomarkers:* GM can be positive; BDG (nonspecific)
Scedosporiosis (*Scedosporium apiospermum*)	Pneumonia or disseminated infection	Neutropenia, glucocorticoids, HSCT	*Culture* of BAL: low sensitivity, nonspecific (colonization, contamination) *Histologic examination* of tissue: acute-angle septate hyphae *Biomarkers:* BDG can be positive
Phaeohyphomycosis (*Cladophialophora, Alternaria, Phialophora, Rhinocladiella, Exophiala,* and *Exserohilum* spp.)	Sinopulmonary, CNS, or disseminated infection Skin infection Allergic sinusitis	HSCT, neutropenia, glucocorticoids, healthy individuals (for CNS), TNF-α inhibition Direct inoculation Atopic individuals	*Culture* of ordinarily sterile site *Histologic examination* of tissue: cell walls may appear dark brown or golden on H&E; Fontana-Masson may stain fungal melanin
Dermatophytosis (*Trichophyton, Microsporum,* and *Epidermophyton* spp.)	Skin and nail infections	Healthy individuals	*Culture* or *microscopic examination* of scrapings or clippings: chains of arthrospores (diagnostic)
Eumycetoma (*Madurella mycetomatis*)	Skin and subcutaneous infections	Healthy individuals	*Culture* and *macroscopic and histologic examination* of grains harvested from biopsy or aspiration
Yeast Fungi			
Mucosal candidiasis[c] (*Candida albicans, C. glabrata*)	Oropharyngeal or esophageal candidiasis Vulvovaginal candidiasis	AIDS, glucocorticoids Antibiotic use	*Culture* of mucosal surfaces *Histologic examination* of esophageal tissue or wet preparation (10% KOH) of vaginal discharge: yeast and/or pseudohyphae
Invasive candidiasis[c] (*C. albicans, C. glabrata, C. parapsilosis, C. tropicalis, C. auris*)	Candidemia Disseminated infection (spleen, liver, kidney, eye, heart, CNS)	Critical illness (ICU) Neutropenia, glucocorticoids	*Culture* of blood: low sensitivity *Histologic examination* of tissue: yeast and/or pseudohyphae *Biomarkers/other tests:* BDG (nonspecific); T2 magnetic resonance in whole blood
Cryptococcosis (*Cryptococcus neoformans, C. gattii*)	Pneumonia Osteomyelitis Meningoencephalitis	AIDS, glucocorticoids Sarcoidosis AIDS, AAbs to IFN-γ or GM-CSF, BTK or JAK inhibition	*Culture* of CSF, BAL fluid, blood *Microscopic examination* of tissue or CSF: encapsulated yeast (GMS, India ink, mucicarmine stain) *Biomarkers: Cryptococcus* Ag (serum, CSF) is sensitive and specific
Trichosporonosis[d] (*Trichosporon asahii, T. mucoides, T. asteroides*)	Superficial skin infection (white piedra) Disseminated infection (skin, eye)	Healthy individuals Neutropenia, glucocorticoids, HSCT, SOT	*Culture* of tissue or blood *Histologic examination* of tissue: yeasts, hyphae, and arthroconidia *Biomarkers:* BDG can be positive
Endemic Dimorphic Fungi			
Histoplasmosis (*Histoplasma capsulatum, H. duboisii* [in Africa])	Self-limited pneumonia Disseminated infection (liver, bone, bone marrow) Fibrosing mediastinitis	Healthy individuals AIDS, SOT, glucocorticoids, AAbs to IFN-γ, JAK or TNF-α inhibition	*Culture* of blood or tissue: low sensitivity; weeks needed for growth *Histologic examination* of tissue: yeast with narrow-based budding *Other tests: Histoplasma* Ag (urine > serum > BAL); BDG can be positive; serology (CF) can be useful in non-AIDS patients
Blastomycosis (*Blastomyces dermatitidis, B. gilchristii*)	Pneumonia Disseminated infection (skin, bone, mucosal surfaces, genitourinary tract)	Healthy individuals AIDS, glucocorticoids, TNF-α inhibition	*Culture* of BAL or tissue: low sensitivity; weeks needed for growth *Histologic examination* of tissue: yeast with broad-based budding *Other tests:* serology (CF, ID) has low sensitivity; *Blastomyces* Ag test cross-reacts with other endemic fungi; GM can be positive
Coccidioidomycosis (*Coccidioides immitis, C. posadasii*)	Self-limited pneumonia Disseminated infection (CNS, bone)	Healthy individuals AIDS, glucocorticoids, TNF-α inhibition	*Culture* is diagnostic[e] *Histologic examination:* spherules *Other tests:* serology (CF, ID); *Coccidioides* Ag test can be useful in CNS infection; BDG can be positive
Paracoccidioidomycosis (*Paracoccidioides brasiliensis, P. lutzii*)	Pneumonia Disseminated infection (skin, bone, mucosal surfaces)	Healthy individuals AIDS, glucocorticoids	*Culture* of tissue: active disease; several weeks needed for growth *Histologic examination* of KOH preparations or tissue: yeast with budding in steering-wheel pattern *Other tests:* serology (ID, CF); *Paracoccidioides* Ag test

(*Continued*)

TABLE 217-1 Major Fungal Infections, Associated At-Risk Patient Populations, and Diagnostic Tests (Continued)

INFECTION (MOST COMMON FUNGAL GENERA AND SPECIES)	CLINICAL SYNDROME(S)	RISK FACTOR(S)	DIAGNOSTIC TEST(S)
Sporotrichosis (*Sporothrix schenckii*)	Lymphocutaneous infection (ascending lymphangitis) Disseminated infection	Direct inoculation AIDS, glucocorticoids	*Culture* of tissue (diagnostic) *Histologic examination*: cigar-shaped yeast, often with surrounding asteroid body
Talaromycosis (*Talaromyces marneffei*)	Pneumonia Disseminated infection (skin, bone, mucosal surfaces)	Healthy individuals AIDS, glucocorticoids, AAbs to IFN-γ	*Culture* of tissue (diagnostic) *Histologic examination* of tissue: yeasts with transverse septa *Biomarkers:* GM is often positive
Adiaspiromycosis (*Emmonsia crescens, E. parva*)	Pneumonia	Occupational dust exposure	*Culture:* nonculturable *Histologic examination*: thick-walled adiaspore within granuloma
Emergomycosis (*Emergomyces africanus, E. pasteurianus*)	Disseminated infection (lungs, skin)	AIDS, SOT	*Culture* of infected tissue *Histologic examination* of tissue: yeast with narrow-based budding *Biomarkers: Histoplasma* Ag can be positive
Chromoblastomycosis (*Fonsecaea pedrosoi, F. monophora*)	Skin and subcutaneous tissue infections	Healthy individuals	*Culture* of infected tissue *Histologic examination* of scrapings (KOH) or tissue (GMS): sclerotic bodies (pathognomonic)
Other Fungi			
Pneumocystosis[f] (*Pneumocystis jirovecii*)	Pneumonia Disseminated infection (eye, CNS, skin, gastrointestinal tract)	AIDS, glucocorticoids, BTK inhibition AIDS	*Culture:* nonculturable *Histologic examination* (gold standard): special (GMS, Diff-Quik) or immunofluorescence stains *Biomarkers/other tests:* BDG (nonspecific); BAL fluid PCR (sensitive; can be positive in colonized individuals)

[a] *A. nidulans* is seen almost exclusively in chronic granulomatous disease. [b] GMS or PAS stains. [c] Some *Candida* species form pseudohyphae. [d] *Trichosporon* species are yeast-like fungi that also generate septate hyphae and arthroconidia. [e] *Coccidioides* is a laboratory hazard. It is important to notify the microbiology laboratory if this infection is suspected. [f] *Pneumocystis* is present in cyst and trophozoite forms.

Abbreviations: AAbs, autoantibodies; ABPA, allergic bronchopulmonary aspergillosis; Ag, antigen; BAL, bronchoalveolar lavage; BDG, β-D-glucan; BTK, Bruton's tyrosine kinase; CF, complement fixation; CNS, central nervous system; CSF, cerebrospinal fluid; GM, galactomannan; GM-CSF, granulocyte-macrophage colony-stimulating factor; GMS, Gomori methenamine silver; H&E, hematoxylin and eosin; HSCT, hematopoietic stem cell transplantation; ICU, intensive care unit; ID, immunodiffusion; IFN-γ, interferon γ; JAK, Janus kinase; KOH, potassium hydroxide; PAS, periodic acid–Schiff; PCR, polymerase chain reaction; SOT, solid organ transplantation; TNF-α, tumor necrosis factor α.

activation, and Th17 cell differentiation upon fungal invasion. Human inherited CARD9 deficiency causes severe mucocutaneous and invasive fungal disease and is the only known primary immunodeficiency to feature fungus-specific infection susceptibility without a predisposition to other infections, autoimmunity, allergy, or cancer. Notably, CARD9-deficient patients develop infections by certain fungi in certain tissues, including (1) chronic mucocutaneous candidiasis linked to defective interleukin (IL) 17 responses; (2) infections of the central nervous system (CNS) caused by *Candida* (but also by *Aspergillus* and phaeohyphomycetes) linked to impaired microglial-neutrophilic responses; and (3) deep dermatophytosis. Thus, the clinical use of SYK inhibitors for autoimmunity and cancer may cause opportunistic fungal disease. Human inherited deficiency of Toll-like receptor (TLR) signaling does not lead to spontaneous fungal disease, yet polymorphisms in TLR pathway molecules may increase the risk of fungal disease in critically ill or immunosuppressed persons, and TLR stimulation may boost protective CLR immunity, as has been shown with the TLR7 agonist imiquimod in chromoblastomycosis.

The development of clinically relevant animal models of mycoses and the phenotypic characterization of fungal infections that develop in patients with primary immunodeficiencies and in recipients of immune pathway-targeting biologics have led to the delineation of fungus-, cell-, and tissue-specific requirements for antifungal host defense **(Fig. 217-1)**.

At the mucosal interface, IL-17-producing lymphoid cells play a critical role in protection by driving epithelial cell production of antimicrobial peptides that restrict mucosal *Candida* invasion. Indeed, AIDS patients are at risk for mucosal—but not invasive—candidiasis. Concordantly, inherited deficiency of IL-17 signaling caused by mutations in *IL17F*, *IL17RA*, *IL17RC*, or *TRAF3IP2* (encoding the IL-17 receptor adaptor ACT1) or pharmacologic inhibition of IL-17 signaling by biologics that target IL-12p40, IL-23p19, IL-17A, IL-17A/IL-17F, or IL-17RA cause mucosal—but not invasive—candidiasis. Other conditions that underlie a predisposition to chronic mucocutaneous candidiasis include primary immunodeficiencies due to mutations in *STAT3*, *STAT1*, *DOCK8*, *JNK1*, *IRF8*, *RORC*, and *CARD9*, all of which impair Th17 cells, as well as thymoma and autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy (APECED), which feature autoantibodies to IL-17A, IL-17F, and IL-22. In APECED, exacerbated T cell–derived interferon γ (IFN-γ)/STAT1 responses disrupt the integrity of the oral epithelial barrier, thereby promoting mucosal fungal invasion and infection; remission of candidiasis can be achieved with JAK inhibition in APECED and STAT1 gain of function (GOF). Of note, vaginal candidiasis (unlike oropharyngeal and esophageal candidiasis) develops in the setting of antibiotic treatment, not AIDS or IL-17–targeted biologics; this observation underscores the role of the microbiota in fungal control at the vaginal—but not the oral—mucosa.

On the other hand, neutrophils—but not lymphocytes—are critical for control of invasive infections caused by *Aspergillus* (and other inhaled molds) and *Candida* (Fig. 217-1). Indeed, patients with chemotherapy-induced neutropenia and patients undergoing allogeneic hematopoietic stem cell transplantation are at risk for invasive aspergillosis and candidiasis. Both oxidative and nonoxidative burst–dependent effector mechanisms are operational within neutrophils for fungal killing. Inherited deficiency in neutrophil superoxide generation due to mutations in the six subunits of the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex causes chronic granulomatous disease, a prototypic primary immunodeficiency that carries a lifetime risk for invasive aspergillosis of ~40%; infrequently (i.e., in <5% of cases, mostly infants), chronic granulomatous disease predisposes to invasive candidiasis. The unexpected development of invasive mold infections in recipients of Bruton's tyrosine kinase (BTK) inhibitors has recently uncovered the critical role of BTK in promoting myeloid phagocyte-dependent antifungal effector functions.

Host defenses against fungi that reside within macrophages, such as *Cryptococcus*, *Pneumocystis*, and endemic dimorphic fungi, depend on the interplay of IFN-γ–producing lymphoid cells and IL-12–producing macrophages that enable intramacrophagic fungal killing (Fig. 217-1).

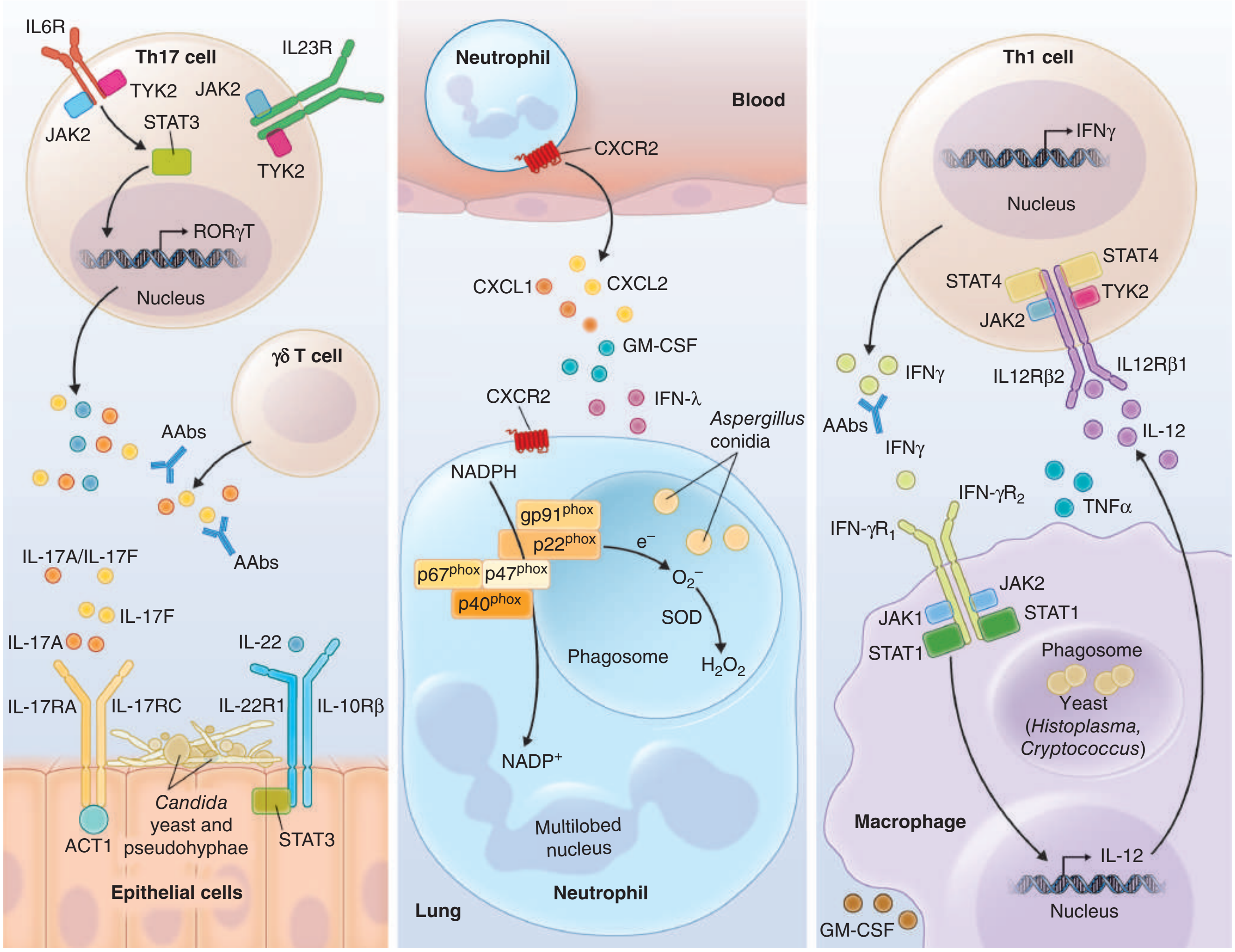
IL6R
Th17 cell
IL23R
TYK2
JAK2
JAK2
STAT3
TYK2
RORγT
Nucleus
γδ T cell
AAbs
AAbs
IL-17A/IL-17F
IL-17F
IL-17A
IL-22
IL-17RA
IL-17RC
IL-22R1
IL-10Rβ
Candida yeast and pseudohyphae
ACT1
STAT3
Epithelial cells
Neutrophil
Blood
CXCR2
CXCL1
CXCL2
GM-CSF
CXCR2
IFN-λ
Aspergillus conidia
NADPH
gp91phox
p22phox
e−
p67phox
p47phox
p40phox
O2−
SOD
Phagosome
H2O2
NADP+
Multilobed nucleus
Lung
Neutrophil
Th1 cell
IFNγ
Nucleus
STAT4
STAT4
TYK2
JAK2
IFNγ
IL12Rβ2
IL12Rβ1
AAbs
IFNγ
IL-12
IFN-γR2
TNFα
IFN-γR1
JAK2
JAK1
STAT1
STAT1
Phagosome
Yeast (*Histoplasma*, *Cryptococcus*)
Macrophage
IL-12
Nucleus
GM-CSF

FIGURE 217-1 Host defense against fungi. ***Left:*** Production of IL-17A, IL-17F, and IL-22 by Th17 cells, Tc17 cells, γδ T cells, and innate lymphoid cells confers protection from mucosal *Candida* invasion. STAT3 promotes Th17 differentiation via RORγt induction. IL-17A and IL-17F bind to IL-17RA and IL-17RC on epithelial cells and signal via ACT1 to produce antimicrobial peptides that inhibit fungal growth. IL-22 binds to its receptor on epithelial cells and activates STAT3 to mediate epithelial proliferation and repair. ***Middle:*** Activation of $CXCR2^+$ neutrophils recruited from blood in the *Aspergillus*-infected lung enables assembly of the six subunits of NADPH oxidase and superoxide generation that promotes fungal killing. Production of reactive oxygen species by neutrophils is facilitated by recruited monocyte-derived and plasmacytoid dendritic cells via type I and type III IFNs and GM-CSF. ***Right:*** The interaction of Th1 cells with macrophages is protective against intramacrophagic endemic dimorphic fungi, *Pneumocystis*, and *Cryptococcus*. Upon fungal uptake, macrophages produce IL-12 that binds to its receptor on T cells and activates STAT4, with consequent release of IFN-γ. IFN-γ binds to its receptor on macrophages and activates STAT1, thereby enabling fungal killing. TNF-α and GM-CSF are also critical for macrophage activation. AAbs, autoantibodies; IL, interleukin; IFN, interferon; JAK, Janus kinase; GM-CSF, granulocyte-macrophage colony-stimulating factor; NADPH, nicotinamide adenine dinucleotide phosphate; RORγt; RAR-related orphan receptor γ; SOD, superoxide dismutase; STAT, signal transducer and activator of transcription; TNF, tumor necrosis factor; TYK2, tyrosine kinase 2.

Indeed, AIDS patients and those receiving glucocorticoids, which affect lymphocytes and macrophages both quantitatively and qualitatively, are at risk for severe infections by these fungi. Accordingly, inherited impairment of the IL-12/IFN-γ signaling axis caused by mutations in *IL12RB1*, *IFNGR1*, *IFNGR2*, *STAT1*, *IRF8*, or *GATA2* underlies susceptibility to severe infection by intramacrophagic fungi (and other intramacrophagic pathogens, such as mycobacteria and salmonellae). In addition, the IFN-γ–targeting monoclonal antibody emapalumab, JAK inhibitors that block IFN-γ–dependent cellular responses, and autoantibodies to IFN-γ predispose to infection with intramacrophagic fungi, as do biologics targeting tumor necrosis factor α (TNF-α) and autoantibodies to granulocyte-macrophage colony-stimulating factor (GM-CSF). The latter predisposing factors reveal the central role of these two Th1-associated cytokines—TNF-α and GM-CSF—in macrophage activation.

Taken together, these observations show that the cellular and molecular factors that drive protective antifungal immune responses vary greatly with the anatomic site of the infection, the offending fungus, and the patient population (Table 217-1). The growing body of data on human immunologic responses to fungi promises to inform precision medicine strategies for risk assessment, prophylaxis, immunotherapy, and vaccination of vulnerable patients.

DIAGNOSIS

The diagnostic modalities used for various fungal infections are outlined in Table 217-1 and are detailed in the chapters on specific mycoses that follow in this section. Definitive diagnosis of a fungal infection requires histopathologic identification of the fungus invading tissue with parallel culture of the fungus from the specimen. Certain fungi have distinctive morphologic features that facilitate diagnosis (Table 217-1). The stains most often used to identify fungi are periodic acid–Schiff (PAS) and Gomori methenamine silver (GMS). *Candida*, unlike other fungi, is visible on Gram-stained tissue smears. Hematoxylin and eosin stains define accompanying histologic features of fungal disease (granuloma formation, angioinvasion, necrosis) but are insufficient to reliably identify fungi in tissue. A positive India ink stain of cerebrospinal fluid (CSF) is diagnostic for cryptococcosis. Most laboratories use calcofluor white staining coupled with fluorescence microscopy to identify fungi in fluid specimens. A positive fungal culture of blood or tissue may signify either patient colonization or lab contamination instead of true infection, with the most likely scenario depending on the fungus and the anatomic site. In blood, *Candida* can be detected with any of the widely used automated blood culture systems, but the lysis-centrifugation technique increases the sensitivity of blood cultures for both *Candida* and other less common fungi (e.g., *Histoplasma*). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) is now used extensively for detection and speciation of fungi recovered from culture.

The several available fungal-antigen and serologic tests vary in sensitivity and specificity. The most reliable of these tests are the antibody to *Coccidioides*, *Histoplasma* antigen, and cryptococcal polysaccharide antigen. Serologic tests are also available for other endemic dimorphic fungi (Table 217-1). The galactomannan test—especially in the bronchoalveolar lavage fluid—is useful for the diagnosis of aspergillosis; however, false-negative results are common, particularly in patients receiving antifungal prophylaxis, and false-positive results may occur with other fungal infections. The β-glucan test has a high negative predictive value for invasive candidiasis but lacks specificity. T2 magnetic resonance is now approved by the U.S. Food and Drug Administration (FDA) for detection of *Candida* in blood. Several polymerase chain reaction and nucleic acid hybridization assays exist for fungal detection but are not standardized and are not widely used in the clinic.

ANTIFUNGAL DRUGS

This section provides a brief overview of available agents for the treatment of fungal infections. Drug regimens and schedules are detailed in the chapters on specific mycoses that follow in this section. Since fungal organisms, like human cells, are eukaryotic, the identification of drugs that selectively kill or inhibit fungi but that are not toxic to human cells poses challenges. Indeed, far fewer antifungal than antibacterial agents have been introduced into clinical medicine.

Early initiation of appropriate antifungal therapy is a critical determinant of favorable outcome, as has been shown for candidemia, aspergillosis, and mucormycosis. In addition, source control of the infection is important—e.g., with removal of the central venous catheter in candidemia, drainage of abdominal abscesses in intraabdominal candidiasis, and surgical debridement of sinus tissue in mucormycosis. Moreover, an essential factor in a favorable prognosis in patients with opportunistic mycoses is the achievement of immune reconstitution—e.g., with neutrophil recovery, tapering of glucocorticoids or other immunosuppressive drugs, or initiation of combination antiretroviral therapy in AIDS.

AMPHOTERICIN B

The advent of amphotericin B (AmB) in the 1950s revolutionized the treatment of deep-seated mycoses. Before the availability of AmB, cryptococcal meningitis and other disseminated fungal infections were nearly always fatal. AmB remains the broadest-spectrum antifungal agent. Its fungicidal mechanism of action involves forming extramembranous sponge-like aggregates that extract fungal ergosterol from lipid bilayers. AmB remains the preferred antifungal agent for the treatment of mucormycosis and fusariosis and for induction therapy for cryptococcal meningitis and disseminated infections caused by endemic dimorphic fungi. However, AmB has several limitations, including lack of a licensed oral formulation and significant toxicity from the intravenous preparations, primarily renal and infusion-related (fever, chills, thrombosis). The introduction of lipid AmB formulations has ameliorated these toxicities, and the lipid formulations have largely replaced the original deoxycholate formulation in resource-rich settings. In developing countries, AmB deoxycholate is still widely used because of the high cost of the lipid formulations. The two lipid formulations commonly used in the clinic are liposomal AmB and AmB lipid complex, which exhibit comparable efficacy, toxicity, and tissue penetration profiles.

AZOLES

Azoles offer important advantages over AmB, such as the availability of oral and IV formulations and a lack of renal toxicity. The mechanism of action of azoles involves inhibition of lanosterol 14α-demethylase and ergosterol synthesis in the fungal cell membrane, with a consequent accumulation of toxic sterol intermediates and growth arrest. Unlike AmB, azoles are considered fungistatic.

Fluconazole Fluconazole plays an important role in the treatment of several fungal infections. Its major advantages are the availability of oral and IV formulations, a long half-life, penetration into most body fluids (ocular fluid, CSF, urine), and minimal toxicity. This drug rarely causes liver toxicity; high doses may result in alopecia, dry mouth, and a metallic taste. Notably, the administration of even low doses of fluconazole to pregnant women for the treatment of vaginal candidiasis was recently linked to miscarriage and stillbirth. Fluconazole has no activity against molds and most endemic dimorphic fungi and is less active than the newer azoles against *C. glabrata* and *C. krusei*.

Fluconazole is the preferred agent for the treatment of coccidioidal meningitis, although relapses may occur despite therapy. Fluconazole is also used as consolidation and maintenance therapy for cryptococcal meningitis and for the treatment of mucosal candidiasis. It is used for treating candidemia in patients who are not critically ill or immunosuppressed; in these patients, fluconazole was found to be as efficacious as AmB. Because of increasing rates of azole-resistant *Candida* strains, echinocandin treatment is preferred, which is then replaced by fluconazole once a susceptible *Candida* species is recovered. Fluconazole is effective as prophylaxis in recipients of high-risk liver and allogeneic bone marrow transplants, although many centers now use posaconazole in neutropenic patients, given its added activity against molds. Fluconazole prophylaxis in leukemic patients, in AIDS patients with low CD4+ T-cell counts, and in patients on surgical intensive care units is controversial.

Itraconazole Itraconazole is available in oral (capsule, suspension) and IV formulations and has broader antifungal activity—i.e., against molds and endemic dimorphic fungi—than fluconazole. Itraconazole is the drug of choice for mild to moderate histoplasmosis and blastomycosis and has also been used to treat chronic coccidioidomycosis, phaeohyphomycosis, sporotrichosis, and mucocutaneous mycoses such as oropharyngeal candidiasis, tinea versicolor, tinea capitis, and onychomycosis. Although it is approved by the FDA for use in febrile neutropenic patients, most centers now use newer azoles in such patients. Disadvantages of itraconazole include its poor CSF penetration, the use of cyclodextrin in its oral suspension and IV formulation, and its variable level of absorption in the capsule form, which requires monitoring of blood levels in patients receiving capsules for disseminated mycoses. A new formulation of itraconazole, called SUBA-itraconazole (for "super-bioavailability"), which results in improved absorption and less variable plasma levels, has recently been approved by the FDA. Itraconazole is a potent CYP3A4 inhibitor; this characteristic leads to significant drug interactions. The drug causes hepatotoxicity and cardiac toxicity that may manifest as congestive heart failure.

Voriconazole Voriconazole is also available in oral and IV formulations, has far broader antifungal activity than fluconazole (including *C. glabrata*, *C. krusei*, *Aspergillus*, *Scedosporium*, and endemic dimorphic fungi—but not Mucorales), and penetrates into most body fluids (ocular fluid, CSF). It is the preferred agent for the treatment of aspergillosis and also has been used to treat scedosporiosis and as step-down (but not primary) therapy for coccidioidomycosis, blastomycosis, and histoplasmosis. Voriconazole is considerably more expensive than fluconazole, and as with itraconazole, its use is associated with numerous interactions with drugs typically used in patients at risk for fungal infections. Hepatotoxicity, visual disturbances, and skin rashes (including photosensitivity) are relatively common, and long-term use requires skin cancer surveillance. A unique toxicity of voriconazole among azoles is fluorosis-associated periostitis. It is crucial to monitor drug levels because (1) voriconazole is metabolized in the liver by CYP2C9, CYP3A4, and CYP2C19; and (2) human genetic variation in CYP2C19 activity exists and can lead to significant interpatient variability in drug levels. Dosages should be reduced in patients with hepatic, but not renal, failure; however, because the IV formulation is prepared in cyclodextrin, it should be given with caution to patients with severe renal failure.

Posaconazole Posaconazole has broader activity than voriconazole, including activity against Mucorales. Both oral (suspension, tablet) and IV formulations are available. Posaconazole is approved by the FDA for antifungal prophylaxis in neutropenic leukemic patients and allogeneic hematopoietic stem cell transplant recipients as well as for treatment of oropharyngeal candidiasis, including infections refractory to fluconazole or itraconazole. Posaconazole has been reported to be effective salvage therapy for aspergillosis, mucormycosis, fusariosis, cryptococcosis, histoplasmosis, and coccidioidomycosis, although controlled clinical trials are lacking. The tablet formulation is not hampered by the suboptimal absorption that occurs with the suspension; the tablet also results in higher and more reliable blood levels of the drug. Posaconazole is less hepatotoxic than voriconazole and does not cause the skin, visual, or bone toxicity that occurs with voriconazole. However, the use of posaconazole is linked to significant P450-related drug interactions and can lead to pseudohyperaldosteronism, which may manifest with hypokalemia, edema, and/or hypertension.

Isavuconazole Isavuconazole is available in oral and IV formulations and has broad antifungal activity similar to that of posaconazole. Isavuconazole is approved by the FDA for treatment of aspergillosis (on the basis of a randomized controlled trial that found it noninferior to voriconazole) and mucormycosis (on the basis of an open-label, noncomparative trial of 37 patients). Isavuconazole appears to be less hepatotoxic than voriconazole; it does not cause skin, bone, or visual toxicity, and it causes fewer P450-associated drug interactions than voriconazole.

Oteseconazole The newest azole—a tetrazole—oteseconazole, is approved by the FDA for the treatment of recurrent vulvovaginal candidiasis. It has potent activity against *Candida* (including azole-resistant strains) species and endemic dimorphic fungi. Its long half-life allows for once-weekly dosing following induction. Oteseconazole does not appear to cause hepatotoxicity or P450-associated drug interactions, but it is contraindicated in women of reproductive potential as it may cause fetal ocular toxicity.

■ ECHINOCANDINS

The echinocandins include the FDA-approved drugs caspofungin, anidulafungin, micafungin, and rezafungin, which are available solely as an IV formulation and inhibit β-1,3-glucan synthase, an enzyme that is crucial for fungal cell-wall synthesis but is not a constituent of human cells. Rezafungin, a new generation echinocandin, has a prolonged half-life that allows for weekly dosing with front-loading. The different echinocandins have comparable efficacy, toxicity, and tissue penetration profiles; are fungicidal for *Candida* and fungistatic for *Aspergillus*; and have no activity against other molds, *Cryptococcus*, or endemic dimorphic fungi. Their most common use to date is in candidal infections. These drugs offer three major advantages: minimal toxicity, minimal drug interactions, and activity against all *Candida* species. The minimal inhibitory concentrations (MICs) of echinocandins are higher against *Candida parapsilosis* than against other *Candida* species, but the higher MICs do not translate into less clinical efficacy against this species.

In controlled trials, caspofungin was as efficacious as AmB against candidemia and invasive candidiasis and as efficacious as fluconazole against candidal esophagitis. Caspofungin has also been efficacious as salvage therapy for aspergillosis. Anidulafungin is approved by the FDA as therapy for candidemia in nonneutropenic patients and for *Candida* esophagitis, abdominal infection, and peritonitis. In controlled trials, anidulafungin was noninferior and possibly superior to fluconazole against candidemia and invasive candidiasis and was as efficacious as fluconazole against candidal esophagitis. Micafungin is approved by the FDA for the treatment of candidal esophagitis and candidemia and for antifungal prophylaxis in hematopoietic stem cell transplantation. Moreover, micafungin yielded favorable results when used for the treatment of invasive aspergillosis and candidiasis in open-label trials. Rezafungin was recently approved by the FDA for the treatment of candidemia and invasive candidiasis.

■ IBREXAFUNGERP

Ibrexafungerp is an oral antifungal agent that belongs to a new class of glucan synthase inhibitors called triterpenoids and has potent activity against *Candida* species. Ibrexafungerp has been approved by the FDA for the treatment of vulvovaginal candidiasis and recurrent vulvovaginal candidiasis as it maintains its antifungal efficacy at low pH that is often observed in the vaginal mucosa. Its use is contraindicated in pregnant and lactating women.

■ FLUCYTOSINE (5-FLUOROCYTOSINE)

Flucytosine use has diminished as newer antifungal drugs have been developed. Its mechanism of action involves intrafungal conversion to 5-fluorouracil, which inhibits fungal DNA synthesis. The use of flucytosine in combination with AmB as induction therapy for cryptococcal meningitis is based on the drugs' synergistic interaction and favorable flucytosine CSF penetration that promotes a rapid decline of the cryptococcal burden in the CSF. Flucytosine is also used in combination with AmB for the treatment of candidal meningitis and endocarditis, although comparative trials with AmB monotherapy are lacking. Flucytosine monotherapy is not recommended as it is associated with the development of resistance. Flucytosine can cause bone marrow suppression and liver toxicity, which are intensified when the drug is used with AmB.

■ TERBINAFINE

Terbinafine inhibits squalene epoxidase and ergosterol synthesis, is used for onychomycosis and ringworm infection, and is as effective as itraconazole in both conditions. Although active against other fungi, terbinafine penetrates poorly into tissues beyond the skin and nails and therefore is not preferred for systemic mycoses. Terbinafine carries a risk for hepatotoxicity.

■ TOPICAL ANTIFUNGAL AGENTS

A detailed discussion of topical agents for mucocutaneous mycoses is beyond the scope of this chapter; the reader is referred to **Chap. 225** and the dermatology literature. Azoles such as clotrimazole, miconazole, and ketoconazole are often used topically to treat common cutaneous mycoses as well as oropharyngeal and vaginal candidiasis. In vaginal candidiasis, oral fluconazole given once has the advantage of not requiring repeated intravaginal application. The polyenes nystatin and AmB have also been used topically for oropharyngeal and vaginal candidiasis. Agents from other classes that are used to treat these conditions include ciclopirox, haloprogin, terbinafine, naftifine, tolnaftate, and undecylenic acid.

■ FURTHER READING

Bennett JE: Introduction to mycoses, in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 9th ed, JE Bennett et al (eds). Philadelphia, Elsevier Saunders, 2020, pp 3082–3086.

Lionakis MS et al: Immune responses to fungal pathogens and therapeutic prospects. Nat Rev Immunol 23:433, 2023.

Pappas PG et al: Clinical mycology today: A synopsis of the mycoses study group education and research consortium (MSGERC) second biennial meeting, September 27–30, 2018, Big Sky, Montana, a proposed global research agenda. Med Mycol 58:569, 2020.

218

Histoplasmosis

Chadi A. Hage, L. Joseph Wheat

■ ETIOLOGY

Histoplasma capsulatum, a thermal dimorphic fungus, is the etiologic agent of histoplasmosis. In most endemic areas in North America, *H. capsulatum* var. *capsulatum* is the causative agent. In Central and South America, histoplasmosis is common and is caused by genetically different clades of *H. capsulatum* var. *capsulatum*. In Africa, *H. capsulatum* var. *duboisii* is also found. Yeasts of var. *duboisii* are larger than those of var. *capsulatum*.

Mycelia—the naturally infectious form of *Histoplasma*—have a characteristic appearance, with microconidial and macroconidial forms **(Fig. 218-1)**. Microconidia are oval and are small enough

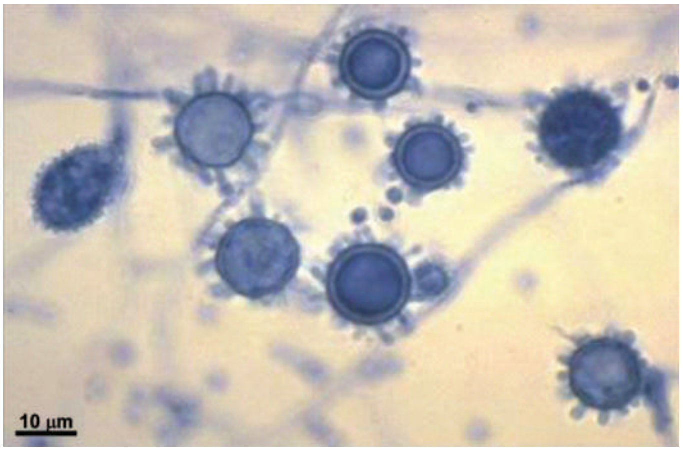

FIGURE 218-1 Spiked spherical conidia of *H. capsulatum* (lacto-phenol cotton blue stain) grown in the laboratory at room temperature.

(2–4 μm) to reach the terminal bronchioles and alveoli. Shortly after infecting the host, mycelia transform into the yeasts that are found inside macrophages and other phagocytes. The yeast forms are characteristically small (2–5 μm), with occasional narrow budding **(Fig. 218-2)**. In the laboratory, mycelia are best grown at room temperature, whereas yeasts are grown at 37°C on enriched media.

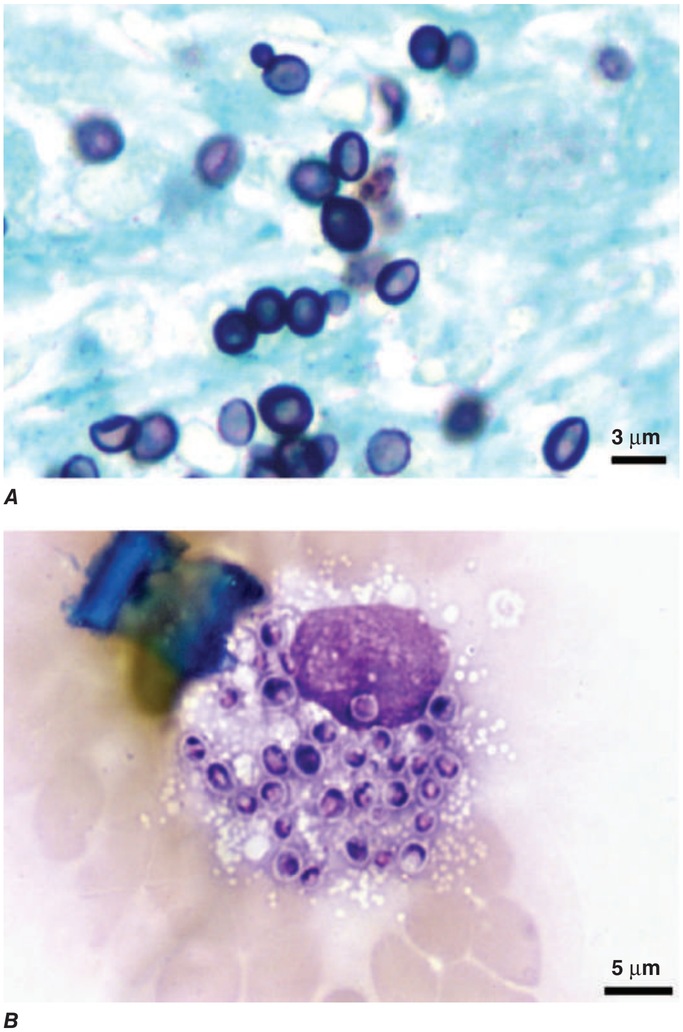

FIGURE 218-2 ***A.*** Small (2–5 μm) narrow budding yeasts of *H. capsulatum* from bronchoalveolar lavage fluid (Grocott's methenamine silver stain). ***B.*** Intracellular yeasts of *H. capsulatum* within an alveolar macrophage from a patient with AIDS and disseminated histoplasmosis (Giemsa stain).

EPIDEMIOLOGY

Histoplasmosis is the most prevalent endemic mycosis in North America. Although this fungal disease has been reported throughout the world, its endemicity is particularly notable in the Ohio and Mississippi river valleys of North America and in certain parts of Mexico, Central and South America (Brazil), Africa, and Asia. Histoplasmosis is increasingly reported outside of the traditionally known endemic areas. The geographic distribution of histoplasmosis is related to the humid and acidic nature of the soil in the endemic areas. Soil enriched with bird or bat droppings promotes the growth and sporulation of *Histoplasma.* Disruption of soil containing the organism leads to aerosolization of the microconidia and exposure of humans nearby. Activities associated with high-level exposure and high infection attack rate include spelunking, excavation, cleaning of chicken coops, demolition and remodeling of old buildings, and cutting of dead trees. Most cases seen outside of highly endemic areas represent imported disease, e.g., in Europe, histoplasmosis is diagnosed fairly often, mostly in emigrants from or travelers to endemic areas on other continents. The epidemiology of histoplasmosis is changing as a result of global climate changes and with the continued expansion of at-risk populations and the acceleration of intercontinental and international travel that brings this infection to areas of the world that are not known to be endemic. The population at risk for histoplasmosis continues to grow as a result of increasing numbers of patients receiving immunosuppressive therapies for autoimmune disorders, cancers, and organ transplants.

PATHOGENESIS AND PATHOLOGY

Infection follows inhalation of microconidia (Fig. 218-1). Once they reach the alveolar spaces, microconidia are rapidly recognized and engulfed by alveolar macrophages, where they transform into yeasts (Fig. 218-2), a process that is integral to the pathogenesis of histoplasmosis and is dependent on the availability of calcium and iron inside the phagocytes. The yeasts are capable of evading the immune system and multiplying inside resting macrophages. Neutrophils and then lymphocytes are attracted to the site of infection. Before the development of cellular immunity, yeasts use the phagosomes as a vehicle for translocation to local draining lymph nodes, whence they spread hematogenously throughout the reticuloendothelial system. Effective cellular immunity develops ~2 weeks after infection. T cells produce interferon-γ to assist the macrophages in killing the organism and controlling the progression of disease. Interleukin 12 and tumor necrosis factor α (TNF-α) play an essential role in cellular immunity to *H. capsulatum*. In the immunocompetent host, macrophages, lymphocytes, and epithelial cells eventually organize and form granulomas that contain the organisms. These granulomas typically fibrose and calcify; calcified lung nodules, mediastinal lymph nodes, and hepatosplenic calcifications are frequently found in healthy individuals from endemic areas. In immunocompetent hosts, infection with *H. capsulatum* confers protective immunity to reinfection. In patients with impaired cellular immunity, the infection is not properly contained and can disseminate throughout the reticuloendothelial system. Progressive disseminated histoplasmosis (PDH) can involve multiple organs, most commonly the lungs, bone marrow, spleen, liver **(Fig. 218-3)**, adrenal glands, and mucocutaneous membranes. Unlike latent tuberculosis, inactive histoplasmosis does not reactivate. In patients with mildly impaired immune systems, active infection may smolder and eventually worsen with further decline in immunity.

Structural lung disease (e.g., emphysema) impairs the clearance of pulmonary histoplasmosis leading to the development of chronic pulmonary disease. This chronic process is characterized by progressive inflammation, tissue necrosis, and fibrosis mimicking cavitary tuberculosis.

CLINICAL MANIFESTATIONS

The clinical spectrum of histoplasmosis ranges from asymptomatic infection to life-threatening illness. The attack rate and the extent and severity of the disease depend on the intensity of exposure, the immune status of the exposed individual, and the underlying lung architecture of the host.

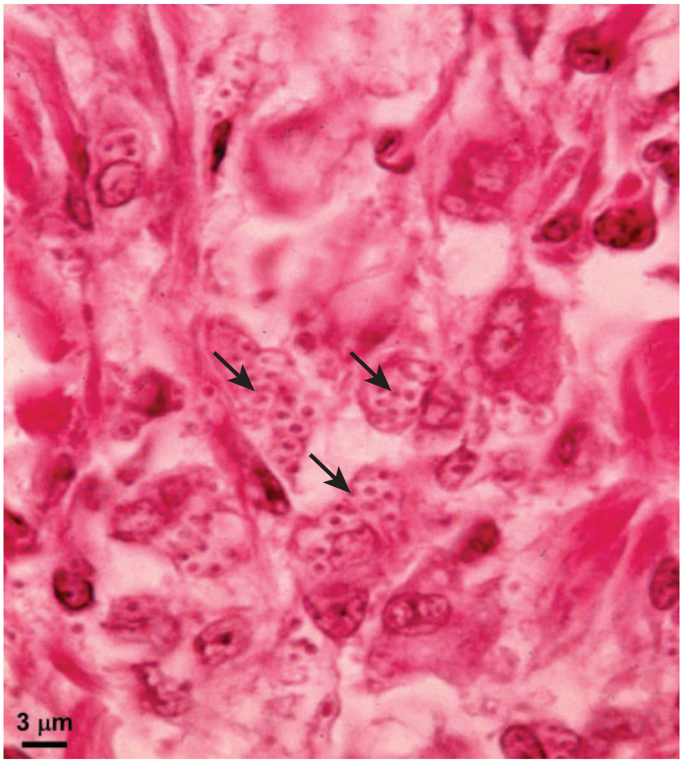

FIGURE 218-3 Intracellular yeasts (*arrows*) of *H. capsulatum* in a liver biopsy specimen (hematoxylin and eosin stain) from a patient who developed progressive disseminated histoplasmosis while receiving anti–tumor necrosis factor therapy for rheumatoid arthritis.

In immunocompetent individuals with low-level exposure, most *Histoplasma* infections are either asymptomatic or mild and self-limited. Of adults residing in endemic areas, up to 75% have immunologic and/or radiographic evidence of previous infection without clinical manifestations. Asymptomatic lung nodules representing controlled histoplasmosis are frequently found on chest computed tomography (CT) scans obtained during screening for lung cancer in smokers from endemic areas. When symptoms of acute histoplasmosis develop, they usually appear 1–4 weeks after exposure. Heavy exposure leads to a flulike illness with fever, chills, sweats, headache, myalgia, anorexia, dry cough, dyspnea, and chest pain. Chest radiographs usually show signs of pneumonitis with prominent hilar or mediastinal adenopathy. Pulmonary infiltrates may be focal with light exposure or diffuse with heavy exposure. Rheumatologic symptoms of arthralgia or arthritis, often associated with erythema nodosum, occur in 5–10% of patients with acute histoplasmosis. Pericarditis may also develop. These manifestations represent inflammatory responses to the acute pulmonary infection rather than extrapulmonary spread. Affected hilar or mediastinal lymph nodes may undergo necrosis and coalesce to form large mediastinal masses that can cause compression of great vessels, proximal airways, and the esophagus. These necrotic lymph nodes may also rupture and create fistulas between mediastinal structures (e.g., bronchoesophageal fistulae).

PDH is typically seen in immunocompromised individuals, who account for ~70% of cases. Common risk factors include AIDS (CD4+ T-cell count, <200/µL), extremes of age, the administration of immunosuppressive medications to prevent or treat rejection following transplantation (e.g., prednisone, mycophenolate, calcineurin inhibitors), and the use of methotrexate, anti-TNF-α agents, and other biologic response modifiers for autoimmune disorders. PDH may also occur in healthy individuals, some of whom may have rare undiagnosed genetic immunodeficiencies of the relevant immune pathways (IFN-γ and TNF-α). Workup for these conditions should be considered in healthy subjects with PDH.

The clinical spectrum of PDH ranges from an acute, rapidly fatal course—with diffuse interstitial or reticulonodular lung infiltrates causing respiratory failure, shock, coagulopathy, and multiorgan failure—to a subacute or chronic course with a focal organ distribution. Common manifestations include fever, weight loss, hepatosplenomegaly, and thrombocytopenia. Other findings may include meningitis or focal brain lesions, ulcerations of the oral mucosa, gastrointestinal ulcerations and bleeding, and adrenal insufficiency. Prompt recognition of this devastating illness is of paramount importance in patients with severe manifestations or with underlying immunosuppression, especially those due to AIDS **(Chap. 208)** or anti-TNF therapy.

Chronic cavitary histoplasmosis is seen in smokers who have structural lung disease (e.g., bullous emphysema). This chronic illness is characterized by productive cough, dyspnea, low-grade fever, night sweats, and weight loss. Chest radiographs usually show upper-lobe infiltrates, cavitation, and pleural thickening—findings resembling those of tuberculosis. Without treatment, the course is slowly progressive.

Fibrosing mediastinitis is an uncommon but serious complication of histoplasmosis. In certain patients, acute infection is followed for unknown reasons by progressive fibrosis around the hilar and mediastinal lymph nodes, encasing mediastinal structures with potentially devastating consequences. Major manifestations include superior vena cava syndrome, obstruction of pulmonary vessels, and airway obstruction. Patients may experience recurrent pneumonia, hemoptysis, or respiratory failure. Fibrosing mediastinitis is fatal in up to one-third of cases.

In healed histoplasmosis, calcified mediastinal nodes or lung parenchymal nodules may erode through the walls of the airways and cause hemoptysis and expectoration of calcified material. This condition is called *broncholithiasis*.

The clinical features and management of histoplasmosis caused by the genetically different clades in Central and South America are similar to those of the disease in North America. African histoplasmosis caused by var. *duboisii* is clinically distinct and is characterized by frequent skin and bone involvement.

■ DIAGNOSIS

Recommendations for the diagnosis and treatment of histoplasmosis are summarized in **Table 218-1**. Once suspected, the diagnosis of histoplasmosis is usually straightforward as many diagnostic tools are now available in the United States. This is not the case in resource-limited endemic regions of Central America, South America, and Africa, where the diagnosis is often delayed, with consequently poor outcomes.

Fungal culture remains the gold standard diagnostic test for histoplasmosis. However, culture results may not be known for up to 1 month, and cultures are often negative in less severe cases. Cultures are positive in ~75% of patients with PDH and chronic pulmonary histoplasmosis. Cultures of bronchoalveolar lavage (BAL) fluid are positive in about half of patients with acute pulmonary histoplasmosis causing diffuse infiltrates and hypoxemia. In PDH, the culture yield is highest for BAL fluid, bone marrow aspirate, and blood. Cultures of sputum or bronchial washings are usually positive in chronic pulmonary histoplasmosis. Cultures are typically negative, however, in other forms of histoplasmosis.

Fungal stains of cytopathology or biopsy materials showing structures resembling *Histoplasma* yeasts are helpful in the diagnosis of PDH, yielding positive results in about half of cases. Yeasts can be seen in BAL fluid (Fig. 218-2) from patients with diffuse pulmonary infiltrates, in bone marrow biopsy samples, and in biopsy specimens of other involved organs (e.g., liver, adrenal glands). Occasionally, yeasts are seen within circulating phagocytes on blood smears from patients with severe PDH. However, staining artifacts and other fungal elements sometimes stain positively and may be misidentified as *Histoplasma* yeasts. Culture and pathology are no longer performed in most patients because diagnosis is more often established by antigen detection and/or serology, more rapidly and without subjecting the patient to invasive procedures.

The detection of *Histoplasma* antigen in body fluids is extremely useful in the diagnosis of PDH and acute diffuse pulmonary histoplasmosis.

TABLE 218-1 Recommendations for the Diagnosis and Treatment of Histoplasmosis

TYPE OF HISTOPLASMOSIS	DIAGNOSTIC TESTS	TREATMENT RECOMMENDATIONS	COMMENTS
Acute pulmonary, mild to moderate with no improvement by the time of diagnosis	*Histoplasma* antigen (BAL fluid, serum, urine) Cytopathology and fungal culture of BAL fluid *Histoplasma* serology (ID and CF), (EIA): IgG and IgM	Itraconazole (200 mg bid) for 6–12 weeks. Monitor renal and hepatic function.	Patients with mild cases usually recover without therapy, but itraconazole should be considered if the patient's condition is not already improving by the time the diagnosis is established.
Acute pulmonary, severe illness, respiratory failure (ARDS)	*Histoplasma* antigen (BAL fluid, serum, urine) Cytopathology and fungal culture of BAL fluid *Histoplasma* serology (ID and CF), (EIA): IgG and IgM	Lipid AmB (3–5 mg/kg per day) ± glucocorticoids for 1–2 weeks; then itraconazole (200 mg bid) for 6–12 weeks. Monitor renal and hepatic function.	Adjust itraconazole dose to achieve blood levels of 2–5 μg/mL.
Chronic/cavitary pulmonary	*Histoplasma* serology (ID and CF), (EIA): IgG and IgM Fungal culture of sputum or BAL fluid	Itraconazole (200 mg bid) Adjust dose to achieve blood levels of 2–5 μg/mL for at least 12 months. Monitor hepatic function.	Continue treatment until radiographic findings show no further improvement. Monitor for relapse after treatment is stopped.
Progressive disseminated	*Histoplasma* antigen (BAL fluid, serum, urine) *Histoplasma* serology (ID and CF), (EIA): IgG and IgM Fungal culture of blood or bone marrow aspirate Cytopathology on biopsy of affected organ	Lipid AmB (3–5 mg/kg per day) for 1–2 weeks; then itraconazole (200 mg bid); adjust dose to achieve blood levels of 2–5 μg/mL for at least 12 months. Monitor renal and hepatic function.	Liposomal AmB is preferred, but the AmB lipid complex may be used because of cost. Chronic antifungal maintenance therapy may be necessary if the degree of immunosuppression cannot be substantially reduced. Monitor antigen levels in urine and serum during treatment to ensure proper response.
Central nervous system	*Histoplasma* antigen CSF *Histoplasma* serology (ID and CF), (EIA): IgG and IgM Fungal culture of CSF	Liposomal AmB (5 mg/kg per day) for 4–6 weeks; then itraconazole (200 mg bid) Adjust dose to achieve blood levels of 2–5 μg/mL for at least 12 months. Monitor renal and hepatic function.	A longer course of lipid AmB is recommended because of the high risk of relapse. Itraconazole should be continued until CSF or MRI abnormalities clear.

Abbreviations: AmB, amphotericin B; BAL, bronchoalveolar lavage; CF, complement fixation; CSF, cerebrospinal fluid; EIA, enzyme immunoassay; ID, immunodiffusion; MRI, magnetic resonance imaging.

The sensitivity of this method is >95% in patients with PDH and >80% in patients with severe acute pulmonary histoplasmosis resulting from heavy exposure, if both urine and serum are tested. Antigen levels correlate with severity of illness in PDH and can be used to follow disease progression, as levels predictably decrease with effective therapy. Increasing antigen levels also predict relapse. *Histoplasma* antigen can be detected in cerebrospinal fluid from patients with *Histoplasma* meningitis and in BAL fluid from those with pulmonary histoplasmosis. Lateral flow antigen detection allows the diagnosis of histoplasmosis at the bedside as a point-of-care testing method, which might offer access to rapid testing in resource-limited areas of the world where traditional antigen testing might not be available. Cross-reactivity occurs with African histoplasmosis, blastomycosis, coccidioidomycosis, paracoccidioidomycosis, talaromycosis, and rarely aspergillosis.

Serologic tests, including immunodiffusion (ID), complement fixation (CF), and IgG and IgM enzyme immunoassay (EIA), are useful for the diagnosis of histoplasmosis, especially in immunocompetent patients. One month may be required for the detection of antibodies after the onset of infection by ID or CF, but antibodies may be detected earlier by more sensitive methods (EIA). IgM appears first then declines, and IgG appears later and increases during the infection. EIA for IgG and IgM antibodies provides a more accurate method for monitoring changes and antibody levels. Serologic tests are especially useful for the diagnosis of chronic pulmonary histoplasmosis. Limitations of ID and CF, however, include insensitivity early in the course of infection and reduced sensitivity in immunosuppressed patients, especially those receiving immunosuppression for organ transplantation. Also, antibodies may persist for several years after infection. Positive results from past infection may lead to a misdiagnosis of active histoplasmosis in a patient with another disease process.

TREATMENT

Histoplasmosis

Treatment is indicated for all patients with PDH or chronic pulmonary histoplasmosis as well as for most symptomatic patients with acute pulmonary histoplasmosis who have not improved by the time the diagnosis is established, especially in those with diffuse infiltrates and difficulty breathing. In most other cases of pulmonary histoplasmosis, treatment is not recommended, especially if the immune system of the host is intact and the degree of exposure is not heavy. The symptoms usually are mild, subacute, and not progressive, and the illness resolves without therapy. Treatment should be considered if the symptoms are not improving within a month.

The preferred treatments for histoplasmosis (Table 218-1) include the lipid formulations of amphotericin B in severe cases and itraconazole in others. Liposomal amphotericin B is more effective and better tolerated than the deoxycholate formulation and is more effective in patients with AIDS and PDH. The deoxycholate formulation of amphotericin B is an alternative to a lipid formulation for patients at low risk for nephrotoxicity and if liposomal amphotericin B is not available. Posaconazole and isavuconazole are alternatives for patients who cannot take itraconazole. *Histoplasma* may develop resistance to fluconazole and voriconazole, and they are not the preferred alternative to itraconazole, especially in immunocompromised patients.

In severe cases requiring hospitalization, a lipid formulation of amphotericin B is used first for 2 weeks, followed by itraconazole. In patients with meningitis, a lipid formulation of amphotericin B should be given for 4–6 weeks before switching to itraconazole. In immunosuppressed patients, the degree of immunosuppression should be reduced if possible, although immune reconstitution inflammatory syndrome (IRIS) may ensue. Antiretroviral treatment improves the outcome of PDH in patients with AIDS and is recommended; however, whether antiretroviral treatment should be delayed to avoid IRIS is unknown.

Blood levels of itraconazole should be monitored to ensure adequate drug exposure, with target concentrations of the parent drug and its hydroxy metabolites measuring 2–5 μg/mL. Drug interactions should be carefully assessed; itraconazole not only is cleared by cytochrome P450 metabolism but also inhibits cytochrome P450. This profile causes interactions with many other medications used routinely in organ transplant recipients.

The duration of treatment for acute pulmonary histoplasmosis is 6–12 weeks, while that for PDH and chronic pulmonary histoplasmosis is at least 1 year. Antigen levels in urine and serum should be monitored during and for at least 1 year after therapy for PDH. Stable or rising antigen levels suggest treatment failure or relapse and should raise concerns regarding proper intake of itraconazole (capsule formulation with food), adherence to treatment, drug absorption, itraconazole serum concentrations, and drug interactions.

Lifelong itraconazole maintenance therapy is recommended for patients with persistently suppressed immunity but not for those with immune recovery—e.g., patients with AIDS who complete at least 1 year of itraconazole and show no signs of active infection including *Histoplasma* antigen levels <2 ng/mL, CD4+ T-cell count recovery to at least 150/μL (preferably >250/μL), and HIV suppression with viral load <50 copies/mL on antiretroviral therapy. Similarly, maintenance therapy may not be necessary in other immunocompromised patients if the clinical findings have cleared, antigen levels are <2 ng/mL, and immunosuppression is substantially reduced.

Fibrosing mediastinitis, which represents a chronic fibrotic reaction to past mediastinal histoplasmosis rather than an active infection, does not respond to antifungal therapy. Often patients with mediastinal granuloma have chronic or progressive courses and receive treatment with itraconazole and corticosteroids to reduce disease progression.

FURTHER READING

Abdallah W et al: Diagnosis of histoplasmosis using the MVista *Histoplasma* galactomannan antigen qualitative lateral flow-based immunoassay: A multicenter study. Open Forum Infect Dis 8:ofab454, 2021.

Azar MM et al: Clinical perspectives in the diagnosis and management of histoplasmosis. Clin Chest Med 38:403, 2017.

Azar MM et al: Current concepts in the epidemiology, diagnosis, and management of histoplasmosis. Semin Respir Crit Care Med 41:13, 2020.

Bahr NC et al: Histoplasmosis infections worldwide: Thinking outside of the Ohio River valley. Curr Trop Med Rep 2:70, 2015.

Hage CA et al: A multicenter evaluation of tests for diagnosis of histoplasmosis. Clin Infect Dis 53:448, 2011.

219

Coccidioidomycosis

Neil M. Ampel

DEFINITION AND ETIOLOGY

Coccidioidomycosis, commonly known as Valley fever (see "Epidemiology," below), is caused by dimorphic soil-dwelling fungi of the genus *Coccidioides*. Genetic analysis has demonstrated the existence of at least two species, *C. immitis* and *C. posadasii*. These species are indistinguishable with regard to the clinical disease they cause and their appearance on routine laboratory media, although *C. posadasii* may grow more quickly at higher temperatures. Thus, the organisms will be referred to simply as *Coccidioides* for the remainder of this chapter.

EPIDEMIOLOGY

Coccidioidomycosis is confined to the Western Hemisphere between the latitudes of 40°N and 40°S. In the United States, areas of high endemicity include the southern San Joaquin Valley of California (hence the sobriquet "Valley fever") and the south-central region of Arizona. However, infection may be acquired in other areas of the southwestern United States, including the other southern counties in California, southern Nevada, southwestern Utah, southern New Mexico, and western Texas. Cases acquired well outside the recognized endemic areas, including in eastern Washington state and in northeastern Utah, have also been described, and a recent climate model suggests that the endemic range may expand further east and north over time. Outside the United States, coccidioidomycosis is endemic to northern Mexico as well as to localized regions of Central America. In South America, there are endemic foci in Colombia, Venezuela, northeastern Brazil, Paraguay, Bolivia, and north-central Argentina.

The fungus is known to inhabit the native soils within the endemic regions. Because of the difficulty in isolating *Coccidioides* from environmental sites, the precise characteristics of potentially infectious soil are not known, but alluvial alkaline deposits in regions of relative aridity with moderate temperature ranges are most closely associated with endemicity. In the United States, several outbreaks of coccidioidomycosis have been associated with soil from archeologic excavations of Amerindian sites both within and outside of the recognized endemic region. When isolated from the soil, *Coccidioides* is found no more than 2–20 cm below the surface, nor is it usually isolated from cultivated soil. In addition to its saphrophytic soil phase, *Coccidioides* appears to have an endozoan life cycle, likely infecting burrowing rodents. This would suggest that the organisms are concentrated in these rodents and their burrows rather than extant in the soil.

In endemic areas, most cases of human coccidioidomycosis occur without obvious soil or dust exposure. Climatic factors may increase the risk of infection. Periods of aridity following rainy seasons are associated with increases in the number of symptomatic infections. The number of reported cases of coccidioidomycosis has been steadily increasing over the past two decades. In 2019, a total of 20,003 cases were reported. The majority of these were either from Arizona or California, in approximately equal numbers, with Nevada, New Mexico, and Utah reporting a small fraction of cases. The factors associated with these increases have not been elucidated but likely include an influx of older individuals without prior coccidioidal infection into endemic areas, construction activity, increased reporting, and changing climate.

PATHOGENESIS, PATHOLOGY, AND IMMUNE RESPONSE

The life cycle of *Coccidioides* is depicted in **Fig. 219-1**. On agar media and in the environment, *Coccidioides* organisms exist as filamentous molds or mycelia. Individual filaments (*hyphae*) elongate and branch, some growing upward. Portions of these aerial hyphae thicken and septate, and alternating cells then degenerate by autolysis, leaving barrel-shaped viable spores called *arthroconidia*. Measuring approximately 2 μm by 5 μm, individual arthroconidia can easily dislodge from the hypha and become airborne. These may persist in the air for prolonged periods. When arthroconidia are inhaled by a susceptible host, their small size allows them to evade initial mechanical mucosal defenses and reach deep into the bronchial tree, where infection is initiated.

Once within a susceptible host, the individual arthroconidia enlarge, become rounded, and develop internal septations. The resulting structures, called *spherules*, may attain sizes up to 200 μm and are unique to *Coccidioides*. The septations encompass uninuclear elements called *endospores*. Spherules may rupture and release packets of endospores that can themselves develop into new spherules, thus propagating infection locally. If returned to artificial media or the soil, the fungus reverts to its mycelial stage.

Clinical observations and data from animal studies strongly support the critical role of a robust cellular immune response in the host's control of coccidioidomycosis. Necrotizing granulomas containing spherules are typically identified in patients with resolved pulmonary infection. In disseminated disease in which infection is not controlled, granulomas are generally poorly formed or do not develop at all, and a polymorphonuclear leukocyte response is frequently seen. In patients who are asymptomatic or in whom the initial pulmonary infection has resolved, delayed-type hypersensitivity to coccidioidal skin antigens has been routinely documented and T-cell activation has been observed in peripheral blood stimulated with coccidioidal antigens. In persons who have developed protective cellular immunity, spherules exist in the lungs for prolonged periods, possibly lifelong, in a latent

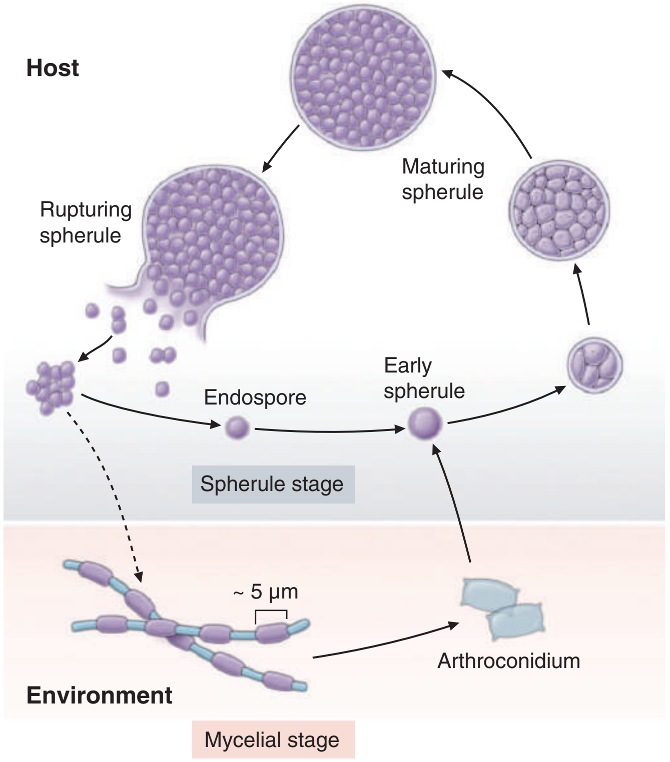

FIGURE 219-1 Life cycle of *Coccidioides*, including the mycelial phase in the environment and the spherule phase in the host.

state. This results in the persistence of protective immunity but also places those infected at risk of recurrent active disease if that immunity should wane.

CLINICAL AND LABORATORY MANIFESTATIONS

After infection, 60% of individuals are completely asymptomatic. The other 40% have symptoms that are related primarily to pulmonary infection, including fever, cough, and pleuritic chest pain. Symptoms generally occur from several days to up to 3 weeks after inhalation of arthroconidia. The risk of symptomatic illness increases with age.

In addition to these local symptoms indicative of infection, there are several manifestations of primary pulmonary coccidioidomycosis that are due to an immunologic response. Most prominent among these are cutaneous reactions. A diffuse, erythematous maculopapular rash, known as toxic erythema, has been noted in some cases. In addition, erythema nodosum **(see Fig. A1-39)**—typically over the lower extremities—and erythema multiforme **(see Fig. A1-24)**—usually in a necklace distribution—may occur. Lesions consistent with Sweet syndrome have also been reported **(Chap. 21)**. Cutaneous manifestations are especially common in women. Symmetrical arthralgias ("desert rheumatism"), often involving the ankles, knees, and hips, may also occur with or without cutaneous manifestations.

Primary pulmonary coccidioidomycosis is often misdiagnosed as a community-acquired bacterial pneumonia. However, the diagnosis of primary pulmonary coccidioidomycosis is strongly suggested by the findings of rash or symmetrical arthralgias in a patient with an appropriate exposure history. The finding of any of the following is also strongly suggestive of coccidioidomycosis: a history of night sweats, marked fatigue, peripheral-blood eosinophilia, failure to improve with antibacterial therapy, and an upper lobe infiltrate or hilar or mediastinal lymphadenopathy on chest imaging.

In most patients, primary pulmonary coccidioidomycosis resolves without sequelae over several weeks. However, several pneumonic complications may arise. Pulmonary nodules are residua of the primary pneumonia. Generally single, frequently located in the upper lobes, and usually ≤4 cm in diameter, nodules are often discovered on routine chest radiograph in asymptomatic patients. Calcification is uncommon. Coccidioidal pulmonary nodules can be difficult to distinguish radiographically from pulmonary malignancies. Like malignancies, coccidioidal nodules often enhance on positron emission tomography. However, unlike malignancies, routine computed tomography (CT) imaging often demonstrates multiple nodules and there may be microsatellite lesions, scattered smaller nodules surrounding the larger one. These findings are not specific, and biopsy may be required to distinguish between these two entities.

Pulmonary cavities occur when a nodule extrudes its contents into the bronchial tree, resulting in a thin-walled shell. Frequently asymptomatic, these cavities may on occasion be associated with persistent cough, hemoptysis, and pleuritic chest pain. They may also become secondarily infected with bacterial oral flora, environmental fungi such as *Aspergillus* species, or even with *Coccidioides* growing from the cavity wall. Rarely, a cavity may rupture into the pleural space, causing pyopneumothorax. In such cases, patients present with acute dyspnea, and the chest radiograph reveals a collapsed lung with a pleural air-fluid level. Chronic or persistent pulmonary coccidioidomycosis manifests with prolonged fever, cough, and weight loss and is radiographically associated with pulmonary scarring, fibrosis, and cavities. It is most common in patients who already have chronic lung disease due to other etiologies.

In some cases, primary pneumonia presents as a diffuse reticulonodular pulmonary process in association with dyspnea and fever. This manifestation of pulmonary coccidioidomycosis may occur in settings of intense environmental exposure or in those with profoundly suppressed cellular immunity, such as persons with untreated HIV-1 infection and markedly depressed peripheral blood CD4 cell count. In the latter case, this picture is associated with unrestrained fungal growth and is frequently associated with fungemia.

Clinical dissemination occurs in approximately 1% of infected individuals and is defined as finding the fungus outside the thoracic cavity. Dissemination is more likely to occur in male patients, particularly those of African and perhaps Filipino ancestry, and in persons with depressed cellular immunity, including patients with HIV-1 infection and peripheral blood CD4 counts of <250/µL, those receiving chronic glucocorticoid therapy, those with allogeneic solid-organ transplants, and those being treated with tumor necrosis factor-α antagonists or other biological response modifiers. Women who acquire new coccidioidal infection during the second or third trimester of pregnancy or postpartum also are at significant risk for disseminated disease. Common sites for dissemination include the skin, bones, joints, soft tissues, and meninges. Dissemination may follow symptomatic or asymptomatic pulmonary infection and may involve only one site or multiple anatomic foci. When it occurs, clinical dissemination is usually evident within the first 6 months after primary pulmonary infection and is generally associated with a lack of cellular immunity to *Coccidioides*.

Of the disseminated syndromes, coccidioidal meningitis is the most dire and is uniformly fatal if untreated. Patients usually present with a persistent dull headache, often accompanied by lethargy, and confusion. Nuchal rigidity, if present, is not severe. Examination of cerebrospinal fluid (CSF) demonstrates lymphocytic pleocytosis with profound hypoglycorrhachia and elevated protein levels. CSF eosinophilia is occasionally observed. The diagnosis is usually established by finding coccidioidal complement-fixing antibody (see below) in the CSF in association with the CSF inflammatory pattern described above. The fungus is isolated from the CSF in fewer than one-third of cases. Magnetic resonance imaging with gadolinium frequently demonstrates enhancement in the basilar meninges. With or without appropriate therapy, patients may develop hydrocephalus, usually communicating, which presents clinically as a marked decline in mental status, often with gait disturbances. An elevated opening CSF pressure on lumbar puncture and dilated ventricles on brain imaging are hallmarks of this condition.

DIAGNOSIS

Serology is the most common method to establish the diagnosis of coccidioidomycosis. Several techniques are available, including the traditional tube-precipitin (TP) and complement-fixation (CF) assays, immunodiffusion TP and CF (IDTP and IDCF), and enzyme immunoassay (EIA) to detect IgM and IgG antibodies. TP and IgM antibodies are found in serum soon after infection and persist for weeks to months. They are not useful for gauging severity of disease. The CF and IgG antibodies occur later in the course of the disease and persist longer than TP and IgM antibodies. Rising CF titers are a reflection of fungal growth and are associated with clinical progression, and the presence of CF antibody in CSF is indicative of coccidioidal meningitis. Antibodies disappear over time in persons whose clinical illness resolves and the presence of coccidioidal antibodies in the serum does not indicate immunity or control of coccidioidomycosis.

Because of its commercial availability, relatively rapid turnaround time, and higher sensitivity than other tests, EIAs for IgM and IgG are recommended for the initial diagnosis of coccidioidomycosis. Both tests, but particularly the IgM EIA, are occasionally falsely positive, and results should be interpreted cautiously when used for screening purposes in asymptomatic individuals. If either the IgM or IgG EIA serology is positive, immunodiffusion tests (IDTP and IDCF) or traditional TP and CF assays should be performed for confirmation. If the IDCF or traditional CF assays are qualitatively positive, a CF titer should be requested. The optical density obtained from the EIA IgG does not correlate with the serologic titer of either the IDCF or traditional CF assay and should not be used as a surrogate for a titer. The CF titer is an important prognostic tool. High titers, for example, those ≥1:32, suggest unrestrained fungal growth, predict disease activity, and suggest the possibility of extrathoracic dissemination.

Coccidioides grows within 3–7 days at 37°C on a variety of artificial media, including blood agar. Therefore, it is always useful to obtain samples of sputum or other respiratory fluids and tissue for culture in suspected cases of coccidioidomycosis. The clinical laboratory should always be alerted to the possibility of this diagnosis, since *Coccidioides* poses a significant laboratory hazard if it is inadvertently inhaled. The organism can also be identified directly. While treatment of samples with potassium hydroxide is rarely fruitful, examination of sputum or other respiratory fluids after Papanicolaou, Gomori methenamine silver, or calcofluor white staining may reveal spherules in a significant proportion of patients with pulmonary coccidioidomycosis. For fixed tissues (e.g., those obtained from biopsy specimens), spherules with surrounding inflammation can be demonstrated with hematoxylin-eosin or Gomori methenamine silver staining.

A commercially available test for coccidioidal antigenuria and antigenemia has been developed and appears to be useful in immunosuppressed patients with severe or disseminated disease. It is also useful when the CSF is assayed in cases of suspected coccidioidal meningitis. False-positive results may occur in cases of histoplasmosis or blastomycosis. Some laboratories offer genomic detection by polymerase chain reaction; this assay does not appear to be more sensitive than culture but can be more rapid.

TREATMENT

Coccidioidomycosis

Currently, two classes of antifungal agents are useful for the treatment of coccidioidomycosis **(Table 219-1)**. While once prescribed routinely, amphotericin B in all its formulations is now reserved for only the most severe cases of dissemination and for intrathecal or intraventricular administration in patients with coccidioidal meningitis in whom other antifungal therapy has failed. The original formulation of amphotericin B, which is dispersed with deoxycholate, is usually administered intravenously in doses of 0.7–1.0 mg/kg either daily or three times per week. The newer lipid-based formulations are associated with less renal toxicity, but there are no studies indicating whether they lead to better improvement than the deoxycholate formulation in coccidioidomycosis. The lipid dispersions are administered intravenously at doses of 3–5 mg/kg daily or three times per week.

Triazole antifungals are currently the principal drugs used to treat most cases of coccidioidomycosis. Clinical trials have demonstrated the usefulness of both fluconazole and itraconazole as initial agents. Fluconazole has been the triazole of choice for the treatment of coccidioidomycosis in the past because of availability, cost, predictable oral absorption, and perceived lack of severe adverse events. However, there are no studies demonstrating its superiority over other triazole antifungals, and evidence indicates that itraconazole is more effective against bone and joint disease and may be more effective for other forms of coccidioidomycosis. For both drugs, a minimal oral adult dosage of 400 mg/d should be used. The maximum dose of itraconazole is 200 mg three times daily, but higher doses of fluconazole may be given and 800 mg is frequently prescribed for coccidioidal meningitis. The newer triazole antifungals, voriconazole and posaconazole, are useful for all types of clinical disease, including meningitis, and should be considered in cases where fluconazole or itraconazole therapy has failed. To date, isavuconazole has been used in limited circumstances in coccidioidomycosis but also appears to be effective, including for meningitis. High-dose triazole therapy may be teratogenic during the first trimester of pregnancy, and triazole therapy should be avoided in pregnant women during this period.

TABLE 219-1 Clinical Presentations of Coccidioidomycosis, Their Frequency, and Recommended Initial Therapy for the Immunocompetent Host

CLINICAL PRESENTATION	FREQUENCY, %	RECOMMENDED THERAPY
Asymptomatic infection	60	None
Primary pneumonia (focal)	40	In most cases, none[a]
Diffuse pneumonia	<1	Amphotericin B followed by prolonged oral triazole therapy
Pulmonary sequelae	5	
Nodule		None
Cavity		In most cases, none[b]
Chronic pneumonia		Prolonged triazole therapy
Disseminated disease	≤1	
Skin, bone, joint, soft tissue disease		Prolonged triazole therapy[c]
Meningitis		Lifelong triazole therapy[d]

[a]Treatment is indicated for hosts with depressed cellular immunity as well as for those with prolonged symptoms and signs of increased severity, including night sweats for >3 weeks, weight loss of >10%, a complement-fixation titer of >1:16, and extensive pulmonary involvement on chest radiography. [b]Treatment (usually with the oral triazoles fluconazole and itraconazole) is recommended for persistent symptoms. [c]In severe cases, some clinicians would use amphotericin B as initial therapy. [d]Intraventricular or intrathecal amphotericin B is recommended in cases of triazole failure. Hydrocephalus may occur, requiring a cerebrospinal fluid shunt.

Note: See text for dosages and durations.

There are several new antifungal agents that are promising for the treatment of coccidioidomycosis in the future. These include olorofim, oteseconazole, ibrexafungerp, fosmanogepix, and oral liquid nanocrystal amphotericin B. At this time, only olorofim has been studied in human coccidioidomycosis, and none are currently approved for use in coccidioidomycosis. However, olorofim has been studied in human coccidioidomycosis and has received U.S. Food and Drug Administration breakthrough therapy designation for difficult-to-treat Coccidioides meningitis.

Most patients with focal primary pulmonary coccidioidomycosis do not require antifungal therapy. Patients for whom antifungal therapy should be considered include those with underlying cellular immunodeficiencies and those with prolonged symptoms and signs of extensive disease. Specific criteria include symptoms persisting for ≥2 months, night sweats occurring for >3 weeks, weight loss of >10%, a serum CF antibody titer of >1:16, and extensive pulmonary involvement apparent on chest radiography. When antifungal therapy is used, either fluconazole or itraconazole at 400 mg daily anywhere from 6 weeks to 6 months is considered appropriate.

Diffuse pulmonary coccidioidomycosis represents a special situation. Because most patients with this form of disease are profoundly hypoxemic and critically ill, many clinicians favor beginning therapy with an amphotericin B formulation combined with an oral triazole antifungal. The triazole antifungal therapy is continued alone once clinical improvement occurs and should be continued for 6 months to 1 year.

The nodules that may occur after primary pulmonary coccidioidomycosis do not require treatment. As noted above, these nodules are not easily distinguished from pulmonary malignancies by means of radiographic imaging. Close clinical follow-up and biopsy may be required to separate these two entities. Most pulmonary cavities do not require therapy. Antifungal treatment should be considered in patients with persistent cough, pleuritic chest pain, and hemoptysis. Occasionally, pulmonary coccidioidal cavities become secondarily infected (see above). This development is often manifested by an air-fluid level within the cavity. Bacterial oral flora or *Aspergillus* species are commonly involved, and therapy directed at these organisms should be considered. Surgical removal of the cavity may be required in cases of persistent productive cough and hemoptysis or in those cases of a persistently growing cavity. In addition, cavities >4 cm in diameter are unlikely to resolve spontaneously, and surgical extirpation should be considered. Surgery is always required in cases of pyopneumothorax. For chronic pulmonary coccidioidomycosis, prolonged antifungal therapy—lasting for at least 1 year—is usually required, with monitoring of symptoms, radiographic changes, sputum cultures, and serologic titers.

Most cases of disseminated coccidioidomycosis require prolonged antifungal therapy. The duration of treatment is based on clinical improvement in conjunction with a significant decline in serum CF antibody titer. Such therapy routinely is continued for at least several years. Relapse occurs in 15–30% of individuals once therapy is discontinued and it is important to continue to monitor such patients on a regular basis (e.g., every 3–4 months) after antifungal therapy is discontinued.

Coccidioidal meningitis poses a special challenge. While most patients with this form of disease respond to treatment with oral triazoles, 80% experience relapse when therapy is stopped. Therefore, lifelong therapy is recommended. In cases of triazole failure, intrathecal or intraventricular amphotericin B may be used. Installation requires considerable expertise and should be undertaken only by an experienced health care provider. Shunting of CSF in addition to appropriate antifungal therapy is required in cases of meningitis complicated by hydrocephalus. It is prudent to obtain expert consultation in all cases of coccidioidal meningitis.

PREVENTION

There are no proven methods to reduce the risk of acquiring coccidioidomycosis among residents of an endemic region, but avoidance of inhalation of uncultivated soil or dust is a reasonable measure. For individuals with suppressed cellular immunity, the risk of developing symptomatic coccidioidomycosis is greater than that in the general population. Among those about to undergo allogeneic solid-organ transplantation, antifungal therapy is appropriate prior to transplantation when there is evidence of active or recent coccidioidomycosis. Several transplant centers in the endemic region provide universal antifungal prophylaxis for 6 months to 1 year after solid-organ and allogeneic stem cell transplantation, and lifelong universal prophylaxis has been advocated after lung transplantation to prevent the development of coccidioidomycosis after transplantation.

Cases of donor-transmitted coccidioidomycosis have been reported. Donors who are living or have lived in a coccidioidal endemic region should be screened serologically for coccidioidomycosis before transplantation and organ donation deferred if there is evidence of active infection.

Data on the use of antifungal agents for prophylaxis in other situations are limited. The administration of prophylactic antifungals is not recommended for HIV-1-infected persons who live in an endemic region. Most experts would administer a triazole antifungal to patients with a history of active coccidioidomycosis or a positive coccidioidal serology in whom therapy with tumor necrosis factor-α antagonists or other biological response modifiers is being considered.

There are recent efforts to develop a vaccine for coccidioidomycosis, and a live avirulent product has demonstrated promising results in a canine model. Future studies will determine if this is a viable strategy for preventing or ameliorating coccidioidal infection in humans.

FURTHER READING

Galgiani JN et al: 2016 Infectious Diseases Society of America (IDSA) clinical practice guideline for the treatment of coccidioidomycosis. Clin Infect Dis 63:e112, 2016.

Gorris ME et al: Expansion of coccidioidomycosis endemic regions in the United States in response to climate change. Geohealth 3:308, 2019.

Shubitz LF et al: Δcps1 vaccine protects dogs against experimentally induced coccidioidomycosis. Vaccine 39:6894, 2021.

Taylor JW, Barker BM: The endozoan, small-mammal reservoir hypothesis and the life cycle of *Coccidioides* species. Med Mycol 57:S16, 2019.

Troung CN et al: Universal lifelong fungal prophylaxis and risk of coccidioidomycosis in lung transplant recipients living in an endemic area. Clin Infect Dis 74:1966, 2021.

220

Blastomycosis

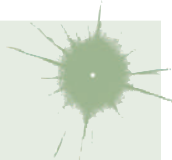

Gregory M. Gauthier, Bruce S. Klein

DEFINITION

Blastomycosis is a pyogranulomatous disease that follows the inhalation of *Blastomyces* conidia or hyphal fragments. Typically, *Blastomyces* causes pulmonary infection; however, a subset of patients will have disseminated disease that involves organs such as the skin, bone, brain, or genitourinary system. Blastomycosis is considered a primary fungal infection because it affects persons with either intact or impaired immune systems. A delay in diagnosis is common because blastomycosis mimics other diseases such as bacterial pneumonia, tuberculosis, and malignancy. Diagnosis involves culture- and nonculture-based tests. Amphotericin B formulations and triazoles are the drugs of choice for treatment.

ETIOLOGY

Blastomyces is a species complex comprising *B. dermatitidis*, *B. gilchristii*, *B. helicus*, *B. percursus*, *B. emzantsi*, *B. silverae*, and *B. parvus*. *B. silverae* and *B. parvus* are not known to commonly infect humans. *Blastomyces* species exhibit thermal dimorphism, which involves the ability to convert between hyphal and yeast morphologies in response to temperature. In the soil (22–25°C), *Blastomyces* grows as septate hyphae that produce infectious conidia. Among the *Blastomyces* species, *B. helicus* is unique because its hyphae grow in a coiled pattern and it does not sporulate under in vitro growth conditions. In organs and tissues (37°C), *Blastomyces* grows as a pathogenic yeast **(Fig. 220-1)** that elicits pyogranulomatous inflammation. The yeast form of all *Blastomyces* species grows as broad-based budding yeast cells, with subtle differences in size among the different species (4–29 μm).

EPIDEMIOLOGY

Although the majority of *Blastomyces* infections occur in North America, blastomycosis is a systemic fungal infection of global importance, with infections also occurring in Africa and Asia. In the United States, the traditional geographic range for *Blastomyces* includes the Mississippi

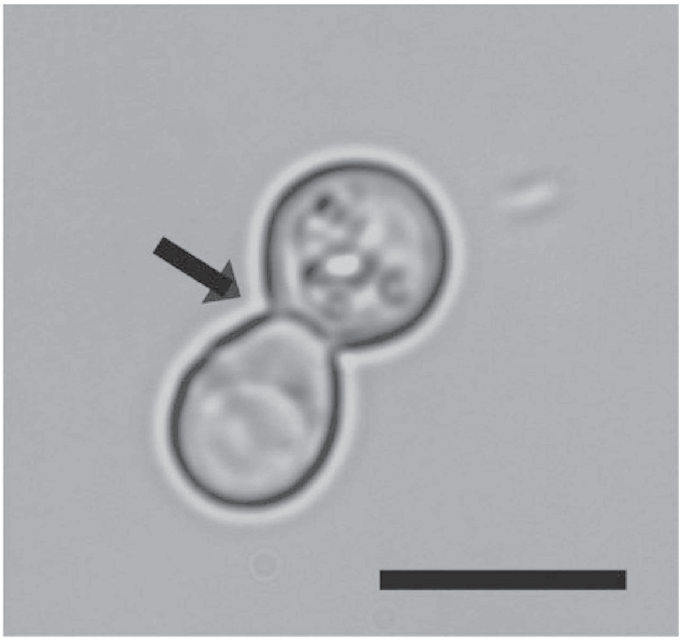

FIGURE 220-1 ***Blastomyces*** **yeast at 37°C**, with broad-based budding between mother and daughter cells (*arrow*). Bar = 10 μm. *(Gregory M. Gauthier, MD, MS.)*

and Ohio River basins, the St. Lawrence River basin, states bordering the Great Lakes, and southeastern states. In Canada, the traditional geographic range includes the provinces of Saskatchewan, Manitoba, Ontario, and Quebec. In North America, *B. dermatitidis* is located throughout the traditional geographic range. *B. gilchristii* is geographically restricted to Minnesota, Wisconsin, Canada, and areas along the St. Lawrence River. *B. dermatitidis* and *B. gilchristii* are thought to have diverged 1.9 million years ago during the Pleistocene epoch, with *B. gilchristii* restricted to formerly glaciated areas. *B. dermatitidis* is found in glaciated and nonglaciated areas. In the environment, *B. dermatitidis* and *B. gilchristii* are not uniformly distributed; rather, they grow in ecologic niches often referred to as *microfoci*, which are characterized by acidic, sandy soils that are found near water and that contain decaying organic matter such as vegetation or wood. In upstate New York State, blastomycosis is an emerging pathogen in the Capitol District and upper Susquehanna River Subbasin with *B. dermatitidis* (88.4%) more common than *B. gilchristii* (11.6%). In Canada, blastomycosis is an emerging pathogen in the province of Saskatchewan. *B. helicus* infections have been reported in the western United States (California, Montana, Idaho, Colorado, Nebraska, Texas) and Canada (Saskatchewan, Alberta); their ecologic niche has yet to be defined. The geographic range and ecologic niche for *B. parvus* and *B. silverae* are unknown.

Outside of North America, blastomycosis has been reported in Africa (>100 cases), India (<10 cases), and Israel. On the basis of morphologic analysis, nearly all clinical isolates of *Blastomyces* in Africa were originally thought to be *B. dermatitidis*. However, molecular phylogenetic analysis of human clinical isolates has demonstrated that multiple *Blastomyces* species exist in Africa, including *B. dermatitidis*, *B. gilchristii*, *B. percursus*, and *B. emzantsi*. A combination of internal transcribed spacer (ITS) sequencing, multilocus sequence typing (MLST), and whole genome sequencing was used to identify a new species, *B. emzantsi*, and to differentiate *B. percursus* from other *Blastomyces* species. MLST has identified a *B. dermatitidis* isolate from Rwanda and *B. gilchristii* from Zimbabwe and South Africa. Analysis of 20 isolates from South Africa collected over a 40-year period identified them as either *B. emzantsi* or *B. percursus*. The geographic distribution and ecologic niche of the four *Blastomyces* species in Africa are unknown. In India, there have been fewer than 10 autochthonous cases of blastomycosis, with the majority identified by morphologic analysis. One autochthonous case (caused by *B. percursus*) with molecular confirmation has been reported from Israel.

Epidemiologic information about blastomycosis derives primarily from passive laboratory surveillance, retrospective studies, and outbreak investigations. The lack of sensitive skin testing and serologic testing, along with difficulty in isolating *Blastomyces* from the environment by culture or molecular methods, has limited an in-depth epidemiologic understanding of blastomycosis. In North America, blastomycosis is reportable in five U.S. states (Minnesota, Wisconsin, Michigan, Arkansas, and Louisiana) and two Canadian provinces (Manitoba, Ontario). The annual incidence of blastomycosis in the traditional endemic area ranges from 0.11 to 2.17 cases/100,000 persons. In older persons (Medicare beneficiaries, 1999–2008), the nationwide annual incidence of blastomycosis was 0.7/100,000, with the highest rates in the midwestern and southern regions of the United States. Analysis of Healthcare Cost and Utilization Project (HCUP) data estimated that 11,776 persons were hospitalized for blastomycosis in the United States from 2010 through 2020, with the majority of patients from midwestern (58.8%) and southern (31.4%) states. In certain places, such as Vilas County, Wisconsin, and Kenora, Ontario, blastomycosis is hyperendemic, with annual incidence rates ranging from 40 to 117 cases/100,000 persons. Incidence data likely underestimate the true burden of infection because they are limited to persons with clinically apparent infection. Patients with asymptomatic or subclinical infections are undercounted.

Most *Blastomyces* infections are sporadic and can occur in either rural or urban areas. There have been at least 20 outbreaks of blastomycosis in the United States since the mid-1950s. Wisconsin, Minnesota, and North Carolina have had multiple outbreaks. The majority of outbreaks have been in rural areas, but several have occurred in urban settings. Activities associated with outbreaks include construction (of homes, cabins, factories, and roads), excavation of dirt, participation in water sports (canoeing, tubing on a river, and fishing), and exposure to a community compost pile or to beaver dams. *Blastomyces* infection is typically acquired from disturbed soil, which liberates infectious particles that are then inhaled into the lungs.

An investigation of a blastomycosis outbreak in Marathon County, Wisconsin (2009–2010), found that 45% of patients were of Hmong ethnicity. A retrospective study from the Marshfield Clinic in Wisconsin (1999–2014) found that 14.4% of patients with blastomycosis were of Asian ethnicity—a figure higher than was anticipated given that <2.5% of the population within the catchment area is Asian, including a large Hmong population. These findings suggest that persons of Hmong ethnicity have an increased risk of acquiring blastomycosis. A combination of whole genome sequencing and immunologic analyses indicated that polymorphisms in the interleukin 6 (IL-6) gene in the Hmong population result in decreased IL-6 production, which in turn impairs development of IL-17-producing $CD4^+$ T lymphocytes. IL-17 is a critical cytokine for recruitment and activation of innate immune cells such as neutrophils and macrophages active against *Blastomyces*. Thus, alterations in IL-6 production may be responsible for the increased risk of blastomycosis in the Hmong population. Although data are limited, persons of Hmong ethnicity do not appear to be at increased risk for disseminated blastomycosis. Increased incidence rates of blastomycosis have also been reported in indigenous people of Canada and the United States. Compared with Caucasians, Asian and indigenous persons with blastomycosis tend to have fewer underlying medical conditions and to be younger.

PATHOGENESIS

A defining feature of the *Blastomyces* species complex is the ability to respond to shifts in temperature by switching between hyphal and yeast forms. In the soil, *Blastomyces* grows as mold cells with hyphae that produce conidia. Hyphal growth promotes environmental survival, genetic diversity through mating, and production of infectious conidia that facilitate transmission of *Blastomyces* from the environment to mammals, including humans. At 37°C (the core temperature of mammals), *Blastomyces* hyphae and conidia convert into pathogenic yeast that upregulate yeast phase–specific virulence factors and downregulate host immune defenses, thereby facilitating infection. Virulence traits that *Blastomyces* shares with *Histoplasma*, *Coccidioides*, *Sporothrix*, and *Paracoccidioides* are thermotolerance at 37°C, intracellular survival, and capacity to cause infection in persons with either healthy or impaired immune defenses. Although *Emergomyces* and *Talaromyces marneffei* (formerly *Penicillium marneffei*) exhibit thermal dimorphism, growth as yeast at 37°C, and intracellular survival, these dimorphic fungi tend to cause infection primarily in immunocompromised persons.

The morphologic switch from hyphae to yeast at 37°C is driven chiefly by temperature and is coupled with the uptake of exogenous cysteine. Cysteine uptake is required to complete the transition to the yeast form because it helps restart mitochondrial respiration, which ceases during the morphologic switch. Over the past two decades, knowledge about the genetic mechanisms that promote the temperature-dependent transition between hyphae and yeast has substantially increased. The discovery of dimorphism-regulating kinase 1 (*DRK1*), which encodes a group III hybrid histidine kinase that is part of the high-osmolarity glycerol (HOG) signaling pathway, provided genetic proof that that the transition to yeast is essential for virulence of the thermally dimorphic fungi. Disruption of *DRK1* by gene deletion or RNA interference resulted in *Blastomyces* cells that grew as hyphae at 37°C instead of yeast. Although viable at 37°C, these cells had altered cell-wall composition, failed to upregulate the *Blastomyces* adhesin 1 (*BAD1*, formerly WI-1) virulence factor, and were avirulent in a mouse model of lethal pulmonary infection. Subsequent studies of *Histoplasma* and *Talaromyces* demonstrated that the function of *DRK1* is conserved with regard to thermal dimorphism and virulence.

The temperature-dependent transition in the other direction—from yeast to hyphae—is regulated in part by a GATA-transcription factor encoded by siderophore biosynthesis repressor in *Blastomyces* (*SREB*), which influences neutral lipid metabolism. In addition, sensing of chitin by *NGT1* and *NGT2* *N*-acetylglucosamine transporters accelerates the conversion to hyphae following a drop in temperature from 37°C to 22°C. These two mechanisms are conserved in *Histoplasma capsulatum*.

As a primary fungal pathogen, *Blastomyces* is one of the few fungi that can infect immunocompetent persons. In its yeast form, *Blastomyces* evades and modulates immune defenses. Following disruption of soil, conidia that are aerosolized and inhaled into the lungs are phagocytosed by pulmonary macrophages, in which a subset of the conidia germinate as yeast and replicate during the early phases of infection. *Blastomyces* is also capable of replicating outside of macrophages. Upon conversion to the yeast phase, an essential virulence factor encoded by *BAD1* is upregulated. *BAD1* encodes a multifunctional 120-KDa cell-surface protein that facilitates yeast adherence to lung epithelial cells via interaction with heparin sulfate, attachment to host immune cells by binding to CR3 and CD14 complement receptors, and downregulation of tumor necrosis factor alpha (TNF-α) in macrophages and neutrophils. In addition, the BAD1 protein impairs activation of CD4+ T lymphocytes, thereby decreasing the production of IL-17 and interferon gamma (IFN-γ). In vivo transcriptional profiling of *B. dermatitidis* yeast during pulmonary infection demonstrated that *BAD1* is the most highly upregulated gene. Deletion of *BAD1* renders *B. dermatitidis* avirulent in a murine model of pulmonary infection. Thus, BAD-1 is essential for virulence in *B. dermatitidis* and likely in *B. gilchristii* as well. In contrast, *BAD1* is absent from the sequenced genomes of *B. helicus*, *B. parvus*, *B. silverae*, *B. percursus*, and *B. emzantsi*.

Additional factors that contribute to the virulence of *Blastomyces* yeast include relative resistance to oxidative stress, upregulation of catalase and superoxide dismutase during infection, active uptake of zinc by a *PRA1*-encoded zincophore and transmembrane transporter (*ZRT1*), and cleavage of granulocyte-macrophage colony-stimulating factor by dipeptidyl peptidase IVA, which blocks activation of innate immune cells (macrophages, neutrophils) and their recruitment to the lung.

APPROACH TO THE PATIENT

Blastomycosis

On the basis of outbreak investigations, it is estimated that 50% of persons exposed to *Blastomyces* develop symptomatic infection after a 3-week to 3-month incubation period. The relatively long incubation period means that patients can be diagnosed with blastomycosis throughout the year. Blastomycosis has been referred to as the "the great pretender" because it can mimic infectious and noninfectious diseases. Blastomycotic pneumonia clinically and radiographically resembles community-acquired bacterial pneumonia, viral pneumonia, tuberculosis, and lung cancer. Patients often receive two or three courses of antibiotics before pulmonary blastomycosis is diagnosed. Without fungal stain and culture, cutaneous lesions can mimic skin cancer, sarcoidosis, and pyoderma gangrenosum. Rarely, blastomycosis can mimic laryngeal cancer. The most important aspect of the approach to a patient with a compatible illness is the consideration of *Blastomyces* as an etiologic agent in the differential diagnosis. This awareness facilitates early diagnosis and treatment, enhancing the potential for improved clinical outcomes. Clinical clues to blastomycosis, especially in persons who reside in or visit endemic regions, include pneumonia that does not improve with antibiotic treatment, pneumonia with extrapulmonary manifestations (e.g., skin lesions, osteomyelitis, central nervous system [CNS] involvement), and skin ulcers that do not respond to standard therapy. Blastomycosis should also be considered in persons from (or visited) an endemic area who have unexplained respiratory failure or acute respiratory distress syndrome (ARDS). In addition, a detailed exposure history can elevate blastomycosis in the differential diagnosis; approximately 50–60% of patients will have environmental risk factors for blastomycosis. This history should also include inquiries about a pet or family member with blastomycosis; these factors have been reported in 7.7–10% and 4–9% of patients, respectively.

CLINICAL MANIFESTATIONS

Pulmonary Blastomycosis Pulmonary manifestations occur in 69–93% of patients with symptomatic blastomycosis and are the most common clinical feature of infection. Signs and symptoms can include fever, chills, productive or nonproductive cough, shortness of breath, hemoptysis, malaise, and decreased appetite. Pulmonary blastomycosis also can manifest as asymptomatic infection, a brief influenza-like illness, acute pneumonia, chronic pneumonia, or ARDS. Radiographic findings in the lungs include lobar consolidation, a mass lesion, interstitial infiltrates, nodule(s), a miliary pattern, cavitary disease, and diffuse involvement of multiple lobes. Hilar adenopathy, pleural effusion, and empyema are uncommon. No distinctive features differentiate blastomycosis from other pulmonary diseases. Diabetes, receipt of a solid organ transplant, immunosuppression, and multilobar pneumonia are risk factors for severe pulmonary blastomycosis. Approximately 4–15% of patients with pulmonary blastomycosis develop ARDS, which is characterized by a fulminant course and high mortality rates ranging from 40 to 89% in most studies. The mortality rate in ARDS is increased when the diagnosis is delayed.

Disseminated Blastomycosis Disseminated blastomycosis occurs in 15–48% of patients and has the potential to involve nearly any organ in the body. The most common site of dissemination is the skin, in which the infection can manifest as papules, ulcers, verrucous lesions, or abscesses. The second most common site is bone, with consequent osteomyelitis characterized by bone pain, soft tissue swelling, soft tissue abscess, and sinus tract formation. Typically, a single bone is involved; however, multifocal osteomyelitis can occur. The most common sites for osteomyelitis include the spine, long bones, and ribs. Dissemination to the CNS (e.g., manifesting as meningitis, an abscess, or a mass lesion), the larynx, or the genitourinary system (e.g., to the prostate or epididymis) occurs in fewer than 10%; the majority of the affected patients have concomitant involvement of other organs, such as the lung or the skin.

Factors that influence dissemination include the infecting *Blastomyces* species, the duration of pulmonary symptoms, and concomitant AIDS. Multiple studies from Wisconsin, a state in which *B. dermatitidis* and *B. gilchristii* are endemic, have demonstrated that *B. dermatitidis* is more likely to cause disseminated infection (31.4–47.8% of cases), whereas *B. gilchristii* tends to remain localized to the lung (90.7–92.2%). Surprisingly, immunosuppression has only a minimal influence on dissemination, an observation suggesting that *Blastomyces* virulence factors have a greater impact than host immune defenses. The frequency of disseminated blastomycosis among solid organ transplant recipients, persons receiving cancer chemotherapy, and patients undergoing pharmacologic immunosuppression is similar to that among patients with intact immune systems. Although patients treated with TNF-α antagonists are considered at risk for blastomycosis, the clinical manifestations and frequency of disseminated disease are unknown in this group because of a paucity of published data. Persons with AIDS and CD4$^+$ T lymphocyte counts of <100/μL are an exception: they are at increased risk for CNS dissemination. Blastomycosis in pregnancy is uncommon, is typically diagnosed in the second or third trimester (91%), and most frequently manifests as pneumonia (74%) or disseminated infection (48%). Transmission to the neonate by either the transplacental route or aspiration of infected vaginal secretions is rare. Persons infected with *B. helicus* can have localized pulmonary infection or disseminated disease; they are typically immunosuppressed (e.g., as a result of solid organ transplantation, chemotherapy, HIV infection, or lupus) and have a high mortality rate (71.4% in seven patients). In contrast to *B. dermatitidis* and *B. gilchristii*, *B. helicus* commonly causes fungemia. Infections with *B. percursus* and *B. emzantsi* are often of long duration (persisting for 4 weeks to 5 years) and can involve the lungs or become disseminated (skin, bone, brain).

DIAGNOSIS

Timely diagnosis of blastomycosis requires a high degree of clinical suspicion because its clinical and radiographic presentations mimic more common etiologies, such as community-acquired pneumonia, malignancy, and tuberculosis. Laboratory findings such as leukocytosis, mild anemia, increased C-reactive protein level, and elevated erythrocyte sedimentation rate are nonspecific. Once suspected, the diagnosis of blastomycosis is straightforward and involves microscopic examination of stained specimens, fungal culture, and antigen testing. The poor sensitivity of complement fixation (9%) and immunodiffusion (28%) renders serologic testing diagnostically dispensable. However, a recently developed serologic test designed to detect antibodies to BAD1 has a sensitivity of 87% and a specificity of 94–99%.

A presumptive diagnosis of blastomycosis can be made by staining of clinical specimens and looking for broad-based budding yeast with a doubly refractile cell wall. Along with the broad-based budding pattern, yeast size (4–29 μm) allows *Blastomyces* to be distinguished from other fungi. An exception is *B. helicus*, which has the potential to be confused with *Histoplasma* because of its small-sized yeast. Respiratory tract specimens such as sputum, tracheal aspirate, and bronchoalveolar (BAL) fluid can be stained with calcofluor, 10% potassium hydroxide, or Papanicolaou stain. Purulent drainage can be stained in a similar manner. The sensitivity of staining of respiratory samples ranges from 50 to 90%. Tissue samples for histopathology should be stained with Gomori methenamine silver or periodic acid–Schiff stain and assessed for pyogranulomatous inflammation and broad-based budding yeast. Traditional stains, such as Gram's stain or hematoxylin and eosin, do not permit optimal visualization of *Blastomyces* yeast.

Growth of *Blastomyces* in cultures of respiratory tissue or body fluid samples provides a definitive diagnosis of blastomycosis but typically requires 5–28 days of incubation. Special media such as Sabouraud dextrose, potato dextrose, and brain–heart infusion are required because *Blastomyces* does not grow well on standard bacteriologic media. Most clinical microbiology laboratories incubate fungal cultures at 25–30°C, a temperature that results in hyphal growth of *Blastomyces*. Unfortunately, *Blastomyces* hyphae are not morphologically distinct enough to confirm diagnosis. Thus, fungal identification and diagnosis are commonly confirmed via chemiluminescent DNA probe or, less commonly, via conversion to yeast upon incubation at 37°C. Diagnosis can also be confirmed by polymerase chain reaction. Neither the chemiluminescent DNA probe nor morphologic analysis of yeast by light microscopy differentiates among the different species of *Blastomyces*. Moreover, some species, such as *B. emzantsi*, are difficult to convert to yeast at 37°C. The species of *Blastomyces* is not typically determined in clinical labs because DNA sequencing is required.

An antigen test that detects a conserved galactomannan component in the *Blastomyces* cell wall has supplanted serologic testing. This test can be performed on urine, blood, BAL fluid, and cerebrospinal fluid. The sensitivity of the antigen test is 85–93% for urine and 57–82% for serum. Infection burden appears to influence test sensitivity, with a lower burden of infection resulting in reduced sensitivity. The antigen test can detect *B. dermatitidis*, *B. gilchristii*, and *B. helicus*; however, its utility for detection of other *Blastomyces* species is unknown. Cross-reactions in the antigen test occur during infection with other dimorphic fungi, including *H. capsulatum* (96%), *Paracoccidioides* species (100%), and *T. marneffei* (70%). Among these, only *H. capsulatum* is found in the same endemic region as *Blastomyces*. Rare cross-reactions can occur with *Aspergillus* and *Cryptococcus* infections. Antigen levels in urine and blood decline with successful treatment, and their measurement can be used to monitor the response to antifungal therapy.

TREATMENT

Blastomycosis

Guidelines for the treatment of blastomycosis have been published by the Infectious Diseases Society of America (2008), the American Thoracic Society (2011), the American Society of Transplantation (2019), and European Confederation of Medical Mycology (2021). Although there are isolated reports of self-limited pulmonary blastomycosis, there are no criteria to determine which patients will experience a resolution of infection. Thus, treatment is recommended for all patients with blastomycosis in order to prevent progressive infection, respiratory failure, and disseminated disease. Antifungal selection is influenced by immune status, CNS involvement, pregnancy, medical comorbidities (e.g., congestive heart failure, prolonged QT interval), and drug–drug interactions. Antifungal drugs active against *Blastomyces* include amphotericin B (AmB) formulations and triazoles. The minimal amount of beta-(1,3)-glucan in the *Blastomyces* yeast cell wall renders echinocandins ineffective, and they should not be used to treat blastomycosis. Hematologic, hepatic, and renal function should be assessed prior to initiation of antifungal therapy, and possible drug–drug interactions should be evaluated. In addition, patients should be educated about proper administration of triazole antifungals. For example, itraconazole capsules require an acidic gastric environment for optimal absorption and should be taken with food and an acidic beverage to improve bioavailability; they cannot be used by persons taking antacids, H2 antagonists, or proton pump inhibitors. In contrast, itraconazole solution can be given to patients receiving gastric acid–lowering therapies and should be taken without food.

Treatment for blastomycosis is summarized in **Table 220-1**. For immunocompetent patients with pulmonary or disseminated blastomycosis of mild or moderate severity (e.g., treatable in the outpatient setting), itraconazole therapy for 6 months is recommended. For severe blastomycosis (e.g., that requiring hospitalization), induction therapy with lipid AmB for 7–14 days (or until clinical improvement), followed by itraconazole treatment for 6–12 months, is recommended. Although not well studied, combination antifungal therapy with lipid AmB and itraconazole (or another azole) can be considered for patients with severe pulmonary blastomycosis. In patients with ARDS, prednisone can be considered; however, the benefits of steroids are unclear. Osteomyelitis due to blastomycosis requires at least 12 months of antifungal therapy, and some patients may require surgical debridement. For blastomycosis involving the CNS, lipid AmB is administered for 4–6 weeks and is followed by treatment with voriconazole, itraconazole, or fluconazole for at least 12 months. Although fluconazole has excellent CNS penetration, its minimum inhibitory concentration (MIC) against *B. dermatitidis* and *B. gilchristii* is higher than that of either itraconazole or voriconazole. Emerging data suggest that isavuconazonium sulfate can be used for treatment of CNS blastomycosis.

Immunosuppressed patients should be treated with 7–14 days of lipid AmB followed by 12 months of itraconazole. For patients

TABLE 220-1 Treatment of Blastomycosis

PATIENT POPULATION	SEVERITY OF INFECTION	SITE OF INFECTION	THERAPY
Immunocompetent	Mild to moderate[a]	Lung	Itraconazole for 6–12 months[b]
		Disseminated	Itraconazole for 6–12 months[b] (≥12 months for osteomyelitis)
	Severe[c]	CNS	Lipid AmB (5 mg/kg daily[d,e] for 4–6 weeks) followed by voriconazole (200–400 mg bid), itraconazole[b] or fluconazole (800 mg daily) for at least 12 months of treatment
		Lung	Lipid AmB (3–5 mg/kg daily[e,f] for 7–14 days) followed by itraconazole[b] for 6–12 months
		Disseminated	Lipid AmB (3–5 mg/kg daily[e,f] for 7–14 days) followed by itraconazole[b] for 12 months of treatment (≥12 months for osteomyelitis)
Immunocompromised	Any severity	CNS	Lipid AmB (5 mg/kg daily[d,e] for 4–6 weeks) followed by voriconazole (200–400 mg bid), itraconazole[b] or fluconazole (800 mg daily) for at least 1 year of treatment[g]
		Lung or disseminated	Lipid AmB (3–5 mg/kg daily[e,f] for 7–14 days) followed by itraconazole[b] for 12 months[g]
Pregnant[h]	Any severity	Any site	Lipid AmB (3–5 mg/kg daily[e,f] for 6–8 weeks), with avoidance of azole antifungals

[a] Mild to moderate infections can typically be managed in the outpatient setting. [b] A loading dose of 200 mg PO tid for 3 days followed by 200 mg PO daily or bid, with dosing based on serum itraconazole levels. The goal for levels of total itraconazole (i.e., itraconazole plus hydroxyitraconazole) is 1–5 μg/mL. Liquid itraconazole has greater bioavailability than the capsule formulation. Liquid itraconazole and oral capsules are administered differently (see text for details). Serum itraconazole levels should be measured after steady state has been reached (2 weeks). Because of the drug's long half-life, blood for serum itraconazole determinations can be drawn regardless of the time of administration. In contrast, serum drug levels for voriconazole, posaconazole, and isavuconazole should be measured before a dose is administered when steady state has been reached (~1 week). [c] Severe blastomycosis requires hospitalization on a medical ward, an intermediate care unit, or an intensive care unit. [d] Lipid amphotericin B (AmB) is the preferred formulation because it has the best central nervous system (CNS) penetration among AmB formulations. For patients with CNS blastomycosis that results in neurologic dysfunction, surgical intervention should be considered. [e] For patients with CNS blastomycosis, severe pulmonary blastomycosis, or severe disseminated blastomycosis, combination therapy with lipid AmB plus a triazole antifungal can be considered; however, this combination has not been formally studied. For patients with acute respiratory distress syndrome, adjunctive steroid therapy with prednisone (40–60 mg daily for 1–2 weeks) can be considered; however, the benefit of steroid administration is not well defined. [f] If lipid AmB is not available, then AmB deoxycholate (0.7–1. 0 mg/kg daily) can be substituted; however, this formulation is associated with higher rates of nephrotoxicity and infusion reactions than lipid AmB. [g] Consider lifelong suppression with itraconazole (200 mg daily) if immunosuppression cannot be reversed. This decision should be made on a case-by-case basis; not all immunosuppressed patients require lifelong suppressive therapy. In addition, lifelong antifungal suppression can be considered in patients who experience relapse after appropriate therapy. [h] All women of childbearing age should undergo pregnancy testing before initiation of therapy.

requiring irreversible immunosuppression, indefinite suppressive azole therapy may be needed; however, in light of the heterogeneity of this patient population, a decision about suppressive therapy should be made on a case-by-case basis. The majority of solid organ transplant recipients do not require lifelong suppression because rates of relapse are low when treatment guidelines are followed.

For pregnant women, lipid AmB treatment for 6–8 weeks is recommended because, unlike the triazole antifungals, lipid AmB is not teratogenic. Fluconazole can cause craniofacial, skeletal, and cardiac defects in the developing fetus (Antley-Bixler-like syndrome); voriconazole and posaconazole also can result in skeletal abnormalities. Itraconazole increases the risk of spontaneous abortion. Before starting antifungal therapy, women of childbearing age with blastomycosis should have a pregnancy test.

Voriconazole, posaconazole, and isavuconazonium sulfate have potent activity against *B. dermatitidis* and *B. gilchristii* and can be considered as alternatives for persons who cannot tolerate itraconazole. These agents, along with itraconazole and AmB, also exhibit good activity against newly identified species of *Blastomyces*, such as *B. helicus*, *B. percursus*, and *B. emzantsi*. Fluconazole MICs against *B. percursus* and *B. emzantsi* are higher than those of other triazoles. Moreover, fluconazole appears to have poor activity against *B. helicus*, *B. parvus*, and *B. silverae*.

PROGNOSIS

Mortality rates for blastomycosis range from 5 to 13%; most deaths are associated with respiratory failure due to ARDS. Analysis of 11,776 hospital admissions for blastomycosis from 2010 through 2020 (HCUP data) demonstrated a 7.9% mortality rate, with increased mortality in patients who were older or who had underlying chronic obstructive pulmonary disease or concomitant malignancy (solid or hematologic). The vast majority of patients who recover from pulmonary blastomycosis do not experience long-term loss of pulmonary function. Cutaneous blastomycosis typically results in scarring.

PREVENTION

Prevention of blastomycosis is challenging because most infections are sporadic and unpredictably acquired from the environment. However, substantial progress has been made in understanding vaccine-mediated immunity conferred by a live, attenuated vaccine strain that is deficient in BAD1. When injected subcutaneously into mice, the *B. dermatitidis* BAD1-null vaccine strain induces sterilizing immunity by activating T_H17 lymphocytes to protect against lethal pulmonary challenge. Major antigenic components of the vaccine identified to date include calnexin and *Blastomyces* endoglucanase-2, which is conserved in other pathogenic fungi, including *Histoplasma capsulatum*, *Coccidioides* species, *Aspergillus* species, *Fonsecaea pedrosoi*, and *Pseudogymnoascus destructans*. Neither the BAD1-null attenuated vaccine nor recombinant antigen-based vaccines are commercially available.

FURTHER READING

Benedict K et al: Blastomycosis-associated hospitalizations, United States 2010-2020. J Fungi 9:867, 2023.

Limper AH et al: An official American Thoracic Society statement: Treatment of fungal infections in adult pulmonary and critical care patients. Am J Respir Crit Care Med 183:96, 2011.

Smith D et al: Clinical testing guidance for coccidioidomycosis, histoplasmosis, and blastomycosis in patients with community-acquired pneumonia for primary and urgent care providers. Clin Infect Dis 78:1559, 2024.

221

Cryptococcosis

Arturo Casadevall, Shmuel Shoham

DEFINITION AND ETIOLOGY

Cryptococcus, a genus of yeast-like fungi, is the etiologic agent of cryptococcosis. In 2022, the World Health Organization (WHO) declared *Cryptococcus neoformans* as a critical priority pathogen. Until recently, cryptococcal strains were separated into two species. However, genome sequencing studies have now revealed tremendous diversity among isolates previously assigned to each species, leading to the proposal that each of the prior species classifications includes numerous new species. Hence, *C. neoformans* and *C. gattii* are now considered as species complexes. However, for clinical purposes, these species complexes cause indistinguishable disease referred to as cryptococcosis. Consequently, this chapter will continue to use the nomenclature *C. neoformans* and *C. gattii* with the understanding that these terms refer to species complexes.

EPIDEMIOLOGY

Cryptococcosis was first described in the 1890s but remained rare until the mid-twentieth century, when advances in diagnosis and increases in the number of immunosuppressed individuals markedly raised its reported prevalence. Although serologic evidence of cryptococcal *infection* is common among immunocompetent individuals, cryptococcal *disease* (cryptococcosis) is rare in the absence of impaired immunity. Conditions associated with high risk for *C. neoformans* infection include hematologic malignancies, receipt of solid-organ transplants, advanced liver disease, diabetes mellitus, illnesses necessitating glucocorticoid therapy, and advanced HIV infection with CD4+ T lymphocyte counts of <200/μL. In contrast, *C. gattii*–related disease is not generally associated with specific immune deficits and often occurs in immunocompetent individuals but is associated with autoantibodies to granulocyte-macrophage colony-stimulating factor.

Cryptococcal infection is acquired from the environment. *C. neoformans* and *C. gattii* species complexes inhabit different ecologic niches. *C. neoformans* is frequently found in soils contaminated with avian excreta and can be recovered from shaded and humid soils contaminated with pigeon droppings. In contrast, *C. gattii* is not found in bird feces. Instead, it inhabits a variety of arboreal species, including eucalyptus trees. *C. neoformans* strains are found throughout the world; however, var. *grubii* (serotype A) strains are far more common than var. *neoformans* (serotype D) strains among both clinical and environmental isolates. *C. gattii* was thought to be largely limited to tropical regions until an outbreak of cryptococcosis caused by a new serotype B strain began in Vancouver in 1999. This outbreak has extended into the United States, and *C. gattii* infections are being encountered increasingly in several states in the Pacific Northwest.

The global burden of cryptococcosis was estimated in 2012 at ~1 million cases, with >600,000 deaths annually, although the prevalence of this disease has declined since then with the greater availability of antiretroviral therapy (ART) for HIV. Since the onset of the HIV pandemic in the early 1980s, the overwhelming majority of cryptococcosis cases have occurred in patients with AIDS **(Chap. 208)**. To comprehend the impact of HIV infection on the epidemiology of cryptococcosis, it is instructive to note that in the early 1990s there were >1000 cases of cryptococcal meningitis each year in New York City—a figure far exceeding that for all cases of bacterial meningitis. With the advent of effective ART, the incidence of AIDS-related cryptococcosis has been sharply reduced among treated individuals. Therefore, most cases of cryptococcosis now occur in resource-limited regions of the world. The disease remains distressingly common in regions where ART is not readily available (e.g., parts of Africa and Asia); in these regions, up to one-third of patients with AIDS have cryptococcosis. Among HIV-infected persons, those with a decreased percentage of memory B cells expressing IgM may be at greater risk for cryptococcosis. In wealthier countries, most cryptococcosis cases are in people without HIV.

PATHOGENESIS

Cryptococcal infection is almost always acquired by inhalation of aerosolized particles. The exact nature of the infectious particles is not known; the two leading candidate forms are small desiccated yeast cells and basidiospores. Occasionally, infection is acquired via direct inoculation of the skin. Little is known about the pathogenesis of initial infection. Serologic studies have shown that cryptococcal infection is acquired in childhood, but it is not known whether the initial infection is symptomatic. Given that cryptococcal infection is common while disease is rare, the consensus is that pulmonary defense mechanisms in immunologically intact individuals are highly effective at containing this fungus. It is not clear whether initial infection leads to a state of immunity or whether most individuals are subject throughout life to frequent and recurrent infections that resolve without clinical disease. However, some human cryptococcal infections lead to a state of latency in which viable organisms are harbored for prolonged periods, possibly in granulomas. Thus, the inhalation of cryptococcal cells and/or spores can be followed by either clearance or establishment of the latent state. The consequences of prolonged harboring of cryptococcal cells in the lung are not known, but evidence from animal studies indicates that the organisms' prolonged presence could alter the immunologic milieu in the lung and predispose to allergic airway disease.

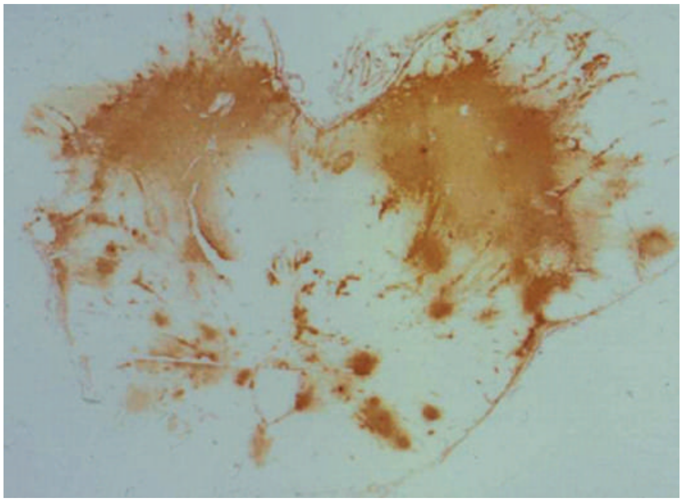

FIGURE 221-1 Cryptococcal antigen in human brain tissue, as revealed by immunohistochemical staining. Brown areas show polysaccharide deposits in the midbrain of a patient who died of cryptococcal meningitis. *(Reproduced with permission from SC Lee et al: Immunohistochemical localization of capsular polysaccharide antigen in the central nervous system cells in cryptococcal meningoencephalitis. Am J Pathol 148:1267, 1996.)*

Cryptococcosis usually presents clinically as pulmonary disease and/or chronic meningoencephalitis. The mechanisms by which the fungus undergoes extrapulmonary dissemination and enters the central nervous system (CNS) remain poorly understood. Current evidence suggests that both direct fungal-cell migration across the endothelium and fungal-cell carriage inside macrophages as "Trojan horse" invaders can occur. *Cryptococcus* species have well-defined virulence factors that include the expression of the polysaccharide capsule, the ability to make melanin, and the elaboration of enzymes (e.g., phospholipase and urease) that enhance the survival of fungal cells in tissue. Among these virulence factors, the capsule and melanin production have been most extensively studied. The cryptococcal capsule is antiphagocytic, and the capsular polysaccharide has been associated with numerous deleterious effects on host immune function. Cryptococcal infections can elicit little or no tissue inflammatory response. The immune dysfunction seen in cryptococcosis has been attributed to the release of copious amounts of capsular polysaccharide into tissues, where it probably interferes with local immune responses **(Fig. 221-1)**. In clinical practice, the capsular polysaccharide is the antigen that is measured as a diagnostic marker of cryptococcal infection.

APPROACH TO THE PATIENT

Cryptococcosis

Cryptococcosis should be included in the differential diagnosis when any patient presents with findings suggestive of chronic pulmonary or CNS infection. Concern about cryptococcosis is heightened by a history of headache and neurologic symptoms in a patient with an underlying immunosuppressive disorder such as advanced HIV infection, malignancy, immunosuppressive use, or solid-organ transplantation. Evaluation of cerebrospinal fluid (CSF) is critical for diagnosis of CNS disease and should include measurements of CSF pressure, protein, glucose, cell count, Gram stain, cultures, and a cryptococcal antigen assay.

CLINICAL MANIFESTATIONS

The clinical manifestations reflect the site of infection. The spectrum of disease consists predominantly of meningoencephalitis and pneumonia, but skin and soft tissue infections also occur and cryptococcosis can affect any tissue or organ. Symptoms may arise from fungal tissue invasion and from an overactive immune response. In CNS disease, symptoms may reflect elevated intracranial pressure (ICP). CNS involvement usually presents as signs and symptoms of chronic meningitis, such as headache, fever, lethargy, sensory deficits, memory deficits, cranial nerve paresis, vision deficits, and meningismus. Cryptococcal meningitis differs from bacterial meningitis in that many patients present with symptoms of several weeks in duration. In addition, classic characteristics of meningeal irritation, such as meningismus, may be absent. Indolent cases can present as subacute dementia or depression. Meningeal cryptococcosis can lead to sudden catastrophic vision loss.

Pulmonary cryptococcosis usually presents as cough, increased sputum production, and chest pain. Radiographic findings include nodules, infiltrates, and masses. *C. gattii* pneumonia can present with granulomatous masses known as *cryptococcomas*. Fever develops in a minority of cases. Pulmonary cryptococcosis can follow an indolent course, and many cases probably do not come to clinical attention. In fact, many cases are discovered incidentally during the workup of an abnormal chest radiograph obtained for other purposes. Pulmonary cryptococcosis can be associated with antecedent diseases such as malignancy, diabetes, and tuberculosis.

Skin lesions are common in patients with disseminated cryptococcosis and can be highly variable, including papules, plaques, purpura, vesicles, tumor-like lesions, and rashes. The spectrum of cryptococcosis in HIV-infected patients is so varied and has changed so much since the advent of ART that a distinction between HIV-related and HIV-unrelated cryptococcosis is no longer pertinent. In highly immunosuppressed patients, the lesions of cutaneous cryptococcosis often resemble those of molluscum contagiosum **(Fig. 221-2; Chap. 201)**.

DIAGNOSIS

A diagnosis of cryptococcosis requires the demonstration of cryptococcal cells or antigen in normally sterile tissues. Visualization of encapsulated fungal cells in CSF mixed with India ink remains a useful rapid diagnostic technique. Cryptococcal cells can also be visualized on Gram stain. However, these staining techniques may yield negative results in patients with a low fungal burden or be misread as positive by inexperienced operators. Cultures of CSF and blood that are positive for cryptococcal cells are diagnostic for cryptococcosis. CSF examination usually reveals evidence of chronic meningitis with mononuclear cell pleocytosis and increased protein levels. A particularly useful test is cryptococcal antigen (CRAg) detection in CSF and blood. The assay is based on serologic detection of cryptococcal polysaccharide and is both sensitive and specific. A major advance in recent years was the introduction of rapid point-of-care CRAgs that provide results in minutes. A positive CRAg test provides strong presumptive evidence for cryptococcosis; however, because the result is often negative in pulmonary cryptococcosis, the test is less useful for diagnosing pulmonary disease and is of only limited usefulness in monitoring the response to therapy. Polymerase chain reaction is available and effective for the diagnosis of cryptococcosis, but physicians should seek confirmation of positive or negative tests with independent methods such as CRAg and culture.

In areas of Africa where there is a high prevalence of HIV infection, routine screening of blood for CRAg in HIV-infected patients with low CD4+ T lymphocyte counts may identify individuals at high risk of cryptococcal disease who are candidates for antifungal therapy. CRAg screening has shown that a significant proportion of HIV-infected patients hospitalized with pneumonia in Thailand harbor cryptococcal infection. Inexpensive point-of-care CRAg tests in the form of lateral flow assays provide rapid and accurate information and are of great diagnostic benefit in resource-limited regions.

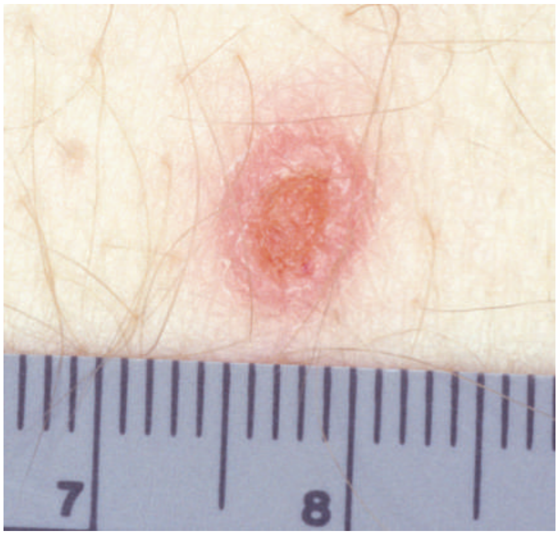

FIGURE 221-2 Disseminated fungal infection. A liver transplant recipient developed six cutaneous lesions similar to the one shown. Biopsy and serum antigen testing demonstrated *Cryptococcus*. Important features of the lesion include a benign-appearing fleshy papule with central umbilication resembling molluscum contagiosum. *(Photo courtesy of Dr. Lindsey Baden; with permission.)*

TREATMENT

Cryptococcosis

When the infection is nonsevere and limited to the lungs or other non-CNS sites, treatment is typically with fluconazole. While pulmonary cryptococcosis in an immunocompetent host sometimes resolves without therapy, given the propensity of *Cryptococcus* species to disseminate from the lung, the inability to gauge the host's immune status precisely, and the availability of low-toxicity therapy in the form of fluconazole, the current recommendation for nonsevere pulmonary cryptococcosis in an immunocompetent individual is fluconazole 400 mg daily for 6–12 months. Fluconazole can cause drug interactions, QT interval prolongation, and liver dysfunction (especially at higher doses), and the dose should be adjusted for renal function. In general, extrapulmonary cryptococcosis without CNS involvement requires less intensive therapy, with the caveat that morbidity and death are associated with meningeal involvement. Thus, the decision to categorize cryptococcosis as "extrapulmonary without CNS involvement" should be made only after evaluation of the CSF reveals no evidence of cryptococcal infection.

When cryptococcosis is severe or involves the CNS, treatment involves induction, consolidation, and maintenance phases. The induction phase ideally includes a combination of intravenous amphotericin B (AmB) and oral flucytosine (5-FC). If feasible, lipid formulations of AmB should be used because those have less toxicity than deoxycholate AmB. The recommended daily doses are liposomal AmB 3–5 mg/kg, AmB lipid complex 5 mg/kg, and deoxycholate AmB 0.7–1 mg/kg. The dose of 5-FC is 25 mg/kg four times daily. 5-FC can cause bone marrow suppression, and the dose should be adjusted for renal function. The duration of induction therapy is a minimum of 2 weeks but may require extension to 4–6 weeks when there is poor initial response to therapy, neurologic complications, or brain cryptococcomas and in people without HIV. Cryptococcal cells with negative cultures may be seen on CSF staining, and elevated cryptococcal antigen may be present in CSF and blood, long after clearance of cultures. These do not indicate treatment failure.

For people with HIV, WHO recommends several alternative approaches to induction. These include a single dose of liposomal AmB 10 mg/kg given along with 14 days of 5-FC at 25 mg/kg four times daily *and* fluconazole 1200 mg daily. If liposomal AmB is not available, a 7-day course of AmB deoxycholate 1 mg/kg per day can be used in its place. If AmB is not available, an all-oral 14-day induction regimen of fluconazole 1200 mg daily and 5-FC 25 mg/kg four times daily can be used. The consolidation phase is typically with fluconazole 400–800 mg daily for 8 weeks. This is followed by fluconazole maintenance at 200–400 mg daily for a year or longer.

The goal of maintenance is to prevent recurrent disease, and its duration depends on the immune status of the patient. Individuals who remain deeply immunocompromised, such as those with HIV whose CD4+ T lymphocyte counts remain <200/μL or have not attained control of HIV with ART may need to remain on fluconazole indefinitely. Newer triazoles like voriconazole, posaconazole, and isavuconazole are highly active against cryptococcal strains and appear to be clinically effective, but clinical experience with these agents is limited. An oral formulation (cochleated) of AmB is in development. The echinocandins are not effective against *Cryptococcus* species. Addition of a short course of interferon γ to antifungal therapy in patients with HIV infection increases clearance of cryptococci from the CSF. Antifungal drug resistance has not been a major problem with cryptococcal strains, but there are increasing reports of drug-resistant strains, including some emerging during the prolonged therapy needed for cryptococcosis. Hence, cryptococcosis that is refractory to antifungal therapy should prompt an investigation into the susceptibility of the clinical isolates in question.

Cryptococcal meningoencephalitis may be associated with increased ICP, which can damage the brain and cranial nerves. CSF pressure should be measured and increased ICP diligently managed in such patients. The mechanism leading to increased ICP is often excess CSF fluid. Neurologic symptoms, including headache, blurred vision, cranial nerve abnormalities, and altered mental status, are typical clues to increased ICP. Management requires reduction of pressure by repeated therapeutic lumbar punctures and the placement of shunts. Neither mannitol nor acetazolamide is effective. Glucocorticoids are not helpful unless the elevated ICP is from a brain lesion with associated mass effect and edema.

Several clinical syndromes in cryptococcosis are driven by an overactive immune response. In people with HIV disease who are treated with ART, an immune reconstitution inflammatory syndrome (IRIS) occurs when immunity rebounds in the setting of treated cryptococcosis (or an undiagnosed asymptomatic infection). Similar syndromes can occur in transplant recipients whose immunosuppressive regimens have been reduced to help control the infection and even in non-immunocompromised people with infection that had appeared to be resolving. The immune response triggers an inflammatory reaction that can be difficult to distinguish from a relapsing infection. Symptoms can include fevers, headache, lymphadenopathy, and pulmonary, CNS, and cutaneous manifestations. Administration of prophylactic dexamethasone in HIV-associated cryptococcosis is not recommended because it has been associated with reduced fungal clearance and increased mortality. The approach to patients that have already developed IRIS and the related immune-driven syndromes must attempt to balance resurgent immunity against immune-mediated damage. Management is individualized and can involve the use of glucocorticoids to reduce inflammation. Apart from the difficulties in distinguishing these inflammatory syndromes from cryptococcal relapse, their management is complex because the cause is often triggered by the desirable outcome of improving immunity, which is important in controlling cryptococcal infection and preventing relapses.

A major consideration for clinicians treating symptomatic AIDS-related cryptococcosis is when to begin ART, which can trigger rebounding immunity. Current recommendations are to start ART 4–6 weeks after initiating antifungal therapy. Screening with serum CRAg is recommended for asymptomatic people with HIV and CD4+ T lymphocyte counts of <100–200/μL. Positive tests should be followed up with careful evaluation for active disease including CSF analysis. For those with antigenemia but no evidence of active disease, a preemptive fluconazole regimen is recommended with 14 days of induction dosing at 800 mg daily and up to 1200 mg daily when fungal burden is high (e.g., CRAg titers exceed 1:80) followed by consolidation and maintenance dosing. Initiation of ART should start 2 weeks after initiation of antifungal therapy in such circumstances.

PROGNOSIS AND COMPLICATIONS

Even with antifungal therapy, cryptococcosis is associated with high rates of morbidity and death. For most patients with cryptococcosis, the most important prognostic factors are the extent and the duration of the underlying immunologic deficits that predisposed them to develop the disease. Cryptococcosis is often curable with antifungal therapy in individuals with no apparent immunologic dysfunction, but in patients with irreversible immunosuppression, the best that can be hoped for is that antifungal therapy will induce remission, which can then be maintained with lifelong suppressive therapy. Before the advent of ART, the median overall survival period for AIDS patients with cryptococcosis was <1 year. Cryptococcosis in patients with underlying neoplastic disease has a particularly poor prognosis. For CNS cryptococcosis, poor prognostic markers are a CSF assay positive for yeast cells on initial India ink examination (evidence of a heavy fungal burden), high CSF pressure, low CSF glucose levels, low CSF pleocytosis (<2/μL), recovery of yeast cells from extraneural sites, absence of antibody to capsular polysaccharide, a CSF or serum cryptococcal antigen level of ≥1:32, and concomitant glucocorticoid therapy or hematologic malignancy. A response to treatment does not guarantee cure since relapse of cryptococcosis is common even among patients with relatively intact immune systems and immune reconstitution syndromes can occur in patients who had been improving with antifungal therapy. Complications of CNS cryptococcosis include cranial nerve deficits, vision and hearing loss, and cognitive impairment.

PREVENTION

No vaccine is available for cryptococcosis. Primary prophylaxis with fluconazole 100–200 mg daily is an approach in high-risk HIV patients (e.g., CD4+ T lymphocyte count <100/μL) and can be used in places where cryptococcal antigen testing is not readily available. Since ART raises the CD4+ T lymphocyte count, it constitutes an immunologic form of prophylaxis.

FURTHER READING

Alanio A: Dormancy in *Cryptococcus neoformans*: 60 years of accumulating evidence. J Clin Invest 130:3353, 2020.

Boyer-Chammard T et al: Recent advances in managing HIV-associated cryptococcal meningitis. F1000Res 8:F1000 Faculty Rev-743, 2019.

Kwon-Chung KJ et al: The case for adopting the "species complex" nomenclature for the etiologic agents of cryptococcosis. mSphere 2:e00357, 2017.

Robertson EJ et al: *Cryptococcus neoformans* ex vivo capsule size is associated with intracranial pressure and host immune response in HIV-associated cryptococcal meningitis. J Infect Dis 209:74, 2014.

Ssebambulidde K et al: Treatment recommendations for non-HIV associated cryptococcal meningoencephalitis including management of post-infectious inflammatory response syndrome. Front Neurol 13:994396, 2022.

Tugume L et al: Cryptococcal meningitis. Nat Rev Dis Primers 9:62, 2023.

222

Candidiasis

Michail S. Lionakis, Shakti Singh,
Ashraf S. Ibrahim, John E. Edwards, Jr.

The genus *Candida* encompasses >150 species, only a few of which cause disease in humans. With rare exceptions (although the exceptions are increasing in number), the human pathogens are *C. albicans*, *C. guilliermondii* (recently revised to *Meyerozyma guilliermondii*), *C. krusei* (recently revised to *Pichia kudriavzevii*), *C. parapsilosis*, *C. tropicalis*, *C. lusitaniae* (recently revised to *Clavispora lusitaniae*), *C. dubliniensis*, *C. glabrata* (recently revised to *Nakaseomyces glabratus*), and the emerging, multidrug-resistant *C. auris*, which has been responsible for several outbreaks in health care facilities in recent years. Ubiquitous in nature, they inhabit the gastrointestinal tract (including the mouth and oropharynx), the female genital tract, and the skin in the majority of healthy persons. Although cases of candidiasis have been described since antiquity in debilitated patients, the advent of *Candida* species as common human pathogens dates to the introduction of modern therapeutic approaches that suppress normal host defense mechanisms. Of those advances, the most important are the use of antibacterial agents that alter the normal human microbiota and allow nonbacterial species to become more prevalent in the commensal flora, the use of indwelling intravenous catheters, and the use of cytotoxic, immunosuppressive treatments for malignant and autoimmune disorders. With the introduction of antifungal agents, the causes of *Candida* infections shifted from an almost complete dominance of *C. albicans* to the common involvement of *C. glabrata* and the other species listed above. The non-*albicans* species now account for approximately half of all cases of candidemia and hematogenously disseminated candidiasis. Recognition of this change is clinically important since the various species differ in susceptibility and are increasingly resistant to the newer antifungal agents.

Candida is a small, thin-walled, ovoid yeast that measures 4–6 μm in diameter and reproduces by budding. Organisms of this genus occur in three forms in tissue: blastospores, pseudohyphae, and hyphae. *Candida* grows readily on simple media; lysis centrifugation enhances its recovery from blood. Species are identified by biochemical testing (currently with automated devices) or on special agar (e.g., CHROMagar).

EPIDEMIOLOGY

Candida are present in humans as commensals, in animals, in foods, and on inanimate objects. In developed countries, where contemporary medical therapeutics are commonly used, *Candida* species are now among the most common nosocomial pathogens. In the United States, these species are among the four most common pathogens isolated from the blood of hospitalized patients. In fact, in a recent point-prevalence study in the United States, *Candida* species were the most common organisms infecting the bloodstream of hospitalized patients. In regions where advanced medical care is not readily available, mucocutaneous *Candida* infections, such as thrush, are more common than deep-organ infections, which rarely occur. However, the incidence of deep-organ candidiasis has been increasing steadily as advances in health care—such as therapy with broad-spectrum antibiotics, more aggressive treatment of cancer, and the use of immunosuppression for sustaining organ transplants—have been implemented. In aggregate, the global incidence of infections due to *Candida* species has risen over the past few decades.

C. auris is an emerging species of *Candida* that has spread rapidly in recent years to >50 countries and is a major public health concern. It was designated an urgent threat in the 2019 Centers for Disease Control and Prevention (CDC) Antimicrobial Resistance Threat Report, and it was included within the critical priority group in the 2022 World Health Organization (WHO) fungal priority pathogens list. This concern stems from its occurrence in health care facilities, its ability to adhere to and persist long term on inanimate objects (in hospitals) and the human skin despite decolonization efforts, its association with substantial mortality, its propensity for misidentification as other *Candida* species, the incomplete understanding of its environmental reservoirs, and its multidrug resistance to the current antifungal therapeutic armamentarium, with some *C. auris* strains being resistant to all antifungal drug classes currently available for treatment. *C. auris* (*auris* meaning *ear* in Latin) was first identified in 2009 from the ear drainage of a patient with an ear infection in Japan. However, subsequent retrospective analysis of *Candida* strain collections identified the earliest known *C. auris* strain to date back to 1996 in South Korea. Notably, whole genome sequencing analysis of *C. auris* strains from South Asia, East Asia, South America, South Africa, and Iran found that although strains within each geographic region are closely related to each other, they are distinct compared to strains from other geographic regions. These findings indicate that *C. auris* emerged independently in multiple geographic locations around the same time; the epidemiologic reasons for this emergence remain poorly understood but may relate to the increasing use of antifungal drugs and climate change.

The presence of a central venous catheter and/or other invasive medical devices and recent residence in nursing homes are major risk factors for *C. auris* colonization and infection. Screening of selected patients who are in a hospital or nursing home where *C. auris* has been cultured and are at risk for dissemination from a colonization site may help in implementing effective infection control measures. Hand hygiene with an alcohol-based hand sanitizer is recommended when hands are not visibly soiled, in which case washing with soap and water is preferred. Identifying the source of contamination, if possible, and using an Environmental Protection Agency (EPA)-registered hospital-grade disinfectant effective against *Clostridioides difficile* spores are desirable. If a patient develops an invasive or bloodstream infection, it is recommended that the health care facility informs the CDC, or a similar agency in other countries, and adheres to recommendations for infection control, including isolation of patients (contact or enhanced barrier precautions), use of proper personal protective coverings, enforcement of hospital environment hygiene, and communicating with other health care facilities if the patient is being transferred.

PATHOGENESIS

In the most severe form of *Candida* infection, the organisms disseminate hematogenously and form microabscesses and small macroabscesses in major organs. Although the exact mechanism is not known, *Candida* probably enters the bloodstream from mucosal surfaces after growing to large numbers as a consequence of bacterial suppression by antibacterial drugs and breaches in the integrity of the mucosal barrier; alternatively, in some instances, the organism may enter the bloodstream from the skin via central venous catheters. A change from the blastospore stage to the pseudohyphal and hyphal stages is generally considered integral to *Candida*'s penetration into tissue. However, *C. glabrata* and *C. auris* can cause life-threatening infection, even though they do not transform into pseudohyphae or hyphae. Adherence to both epithelial and endothelial cells is thought to be the first step in invasion and infection; several adhesins have been identified as well as a mucosal toxin, candidalysin. Biofilm formation also is considered important in pathogenesis. Numerous reviews of cases of hematogenously disseminated candidiasis have identified the predisposing factors or conditions associated with disease **(Table 222-1)**.

Several genes that are involved in the pathogenesis of other *Candida* species—such as those responsible for biofilm formation, proteinases, lipases, phospholipases, hydrolases, adhesins, secreted aspartyl proteases, and transporters involved in azole resistance—are also present in *C. auris*. Unlike other *Candida* species, several *C. auris* strains exhibit aggregate-forming properties in vivo, which may enable immune evasion. In addition, *C. auris* shows a unique tolerance to high temperature and saline concentrations and can grow optimally at up to 42°C and in a 10% saline concentration, making it possible to exist and persist in harsh environments. Furthermore, *C. auris* has significant affinity for

TABLE 222-1 Well-Recognized Factors and Conditions Predisposing to Hematogenously Disseminated Candidiasis

- Antibacterial agents
- Indwelling intravenous catheters
- Hyperalimentation fluids
- Indwelling urinary catheters
- Parenteral glucocorticoids
- Severe burns
- CARD9 deficiency (central nervous system)
- Abdominal and thoracic surgery
- Cytotoxic chemotherapy
- Immunosuppressive agents for organ transplantation
- Respirators
- Myeloperoxidase deficiency
- Neutropenia
- Low birth weight (neonates)
- Diabetes

abiotic surfaces such as plastic materials and medical devices, as well as human skin and nasal and ear cavities, which may account for its persistent colonization capabilities. The *C. auris*–specific adhesin surface colonization factor 1 (SCF1) was recently identified as a key fungal factor governing biofilm formation, colonization of skin and medical devices, and virulence during invasive infection.

Innate immunity is the most important defense mechanism against hematogenously disseminated candidiasis, and the neutrophil is the most potent component of this defense. Macrophages also play an important host defense role. On the other hand, interleukin (IL) 17–producing innate and adaptive lymphoid cells contribute significantly to defense against mucocutaneous candidiasis as evidenced by several monogenic disorders of IL-17 receptor signaling that manifest with *chronic mucocutaneous candidiasis* (CMC) (see "Clinical Manifestations," below). Although many immunocompetent individuals have antibodies to *Candida*, the role of these antibodies in defense against the organism is not clear. Multiple genetic polymorphisms in host immune-related genes that predispose to both disseminated and focal candidiasis have been identified and may contribute to patient susceptibility.

CLINICAL MANIFESTATIONS

Mucocutaneous Candidiasis *Thrush* is characterized by white, adherent, painless, discrete or confluent patches in the mouth, on the tongue, or in the esophagus, occasionally with fissuring at the corners of the mouth. This form of disease caused by *Candida* can also occur at points of contact with dentures (called "denture sore mouth"). Organisms are identifiable in gram-stained scrapings from lesions. The occurrence of thrush in a young, otherwise healthy-appearing person should prompt an investigation for underlying HIV infection. More commonly, thrush is seen as a nonspecific manifestation of severe debilitating illness. Vulvovaginal candidiasis is accompanied by pruritus, pain, and vaginal discharge, which is usually thin but may contain whitish "curds" in severe cases. In contrast to oral thrush, HIV is not considered a major risk factor for vulvovaginal candidiasis. Instead, many women who receive antibiotics, particularly β-lactams, may develop vulvovaginal candidiasis. A subset of patients with recurrent vulvovaginitis may have a deficiency in the surface expression of Dectin-1, encoded by *CLEC7A*, a major recognition factor for β-glucan on the surface of *Candida* and/or in the downstream adaptor molecule CARD9, which ultimately increases the propensity for recurrent mucocutaneous (including vaginal) infections.

Other *Candida* skin infections include *paronychia*, a painful swelling at the nail–skin interface; *onychomycosis*, a fungal nail infection rarely caused by this genus; *intertrigo*, an erythematous irritation with redness and pustules in the skin folds; *balanitis*, an erythematous-pustular infection of the glans penis; *erosio interdigitalis blastomycetica*, an infection between the digits of the hands or toes; *folliculitis*, with pustules developing most frequently in the area of the beard; *perianal candidiasis*, a pruritic, erythematous, pustular infection surrounding the anus; *mastitis*; and *diaper rash*, a common erythematous, pustular perineal infection in infants. *Generalized disseminated cutaneous candidiasis*, another form of infection that occurs primarily in infants, is characterized by widespread eruptions over the trunk, thorax, and extremities. The diagnostic macronodular lesions of hematogenously disseminated candidiasis **(Fig. 222-1)** indicate a high probability of dissemination to multiple organs as well as the skin. While the lesions are seen predominantly in immunocompromised patients treated with cytotoxic drugs, they may also develop in patients without neutropenia.

CMC is a heterogeneous infection of the hair, nails, skin, and mucous membranes that persists despite intermittent antifungal therapy. The onset of disease usually comes in infancy or within the first two decades of life, but in rare cases, it occurs in later life. The condition may be mild and limited to a specific area of the skin or nails, or it may take a severely disfiguring form (*Candida* granuloma) characterized by exophytic outgrowths on the skin. CMC is usually associated with specific immunologic dysfunction; most frequently reported is a failure of lymphocytes to secrete or respond to type-17 cytokines following stimulation by *Candida* antigens in vitro. A subset of the affected patients has mutations in the IL-17 receptors IL-17RA or IL-17RC, its adaptor molecule ACT1 (TRAF3IP2), or, more often, in STAT1.

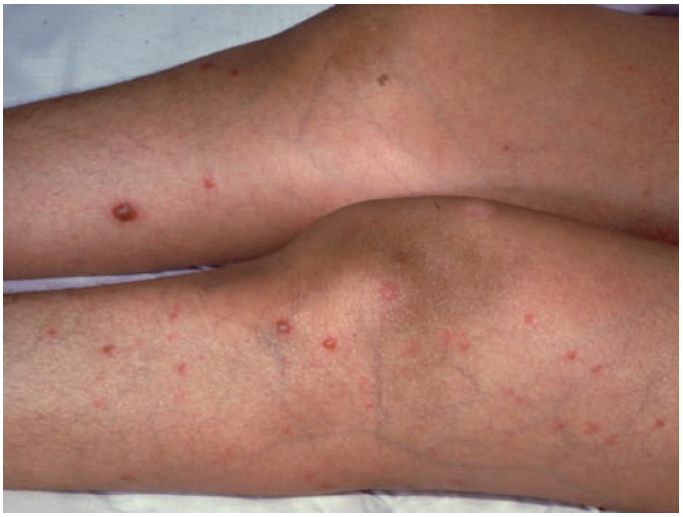

FIGURE 222-1 Macronodular skin lesions associated with hematogenously disseminated candidiasis. *Candida* organisms are usually but not always visible on histopathologic examination. The fungi grow when a portion of the biopsied specimen is cultured. Therefore, for optimal identification, both histopathology and culture should be performed. *(Image courtesy of Dr. Noah Craft and the Victor Newcomer collection at UCLA, archived by Logical Images, Inc.; with permission.)*

Approximately half of patients with CMC have associated endocrine abnormalities either in the setting of gain-of-function mutations in *STAT1* or in the context of *a*utoimmune *p*oly*e*ndocrinopathy–*c*andidiasis–*e*ctodermal *d*ystrophy (APECED) syndrome. This syndrome is due to mutations in the autoimmune regulator (*AIRE*) gene and is most prevalent among Finns, Iranian Jews, and Sardinians. Conditions that usually follow the onset of the disease include hypoparathyroidism, adrenal insufficiency, autoimmune thyroiditis, autoimmune hepatitis, autoimmune pneumonitis, alopecia, pernicious anemia, intestinal malabsorption, and primary hypogonadism. In addition, dental enamel dysplasia, vitiligo, nail dystrophy, asplenia, and calcification of the brain and tympanic membranes may occur. Patients with CMC rarely develop hematogenously disseminated candidiasis, reflecting their intact neutrophil function.

Deeply Invasive Candidiasis Deeply invasive *Candida* infections may or may not be due to hematogenous seeding. Deep esophageal infection may result from penetration by organisms from superficial esophageal erosions; joint or deep-wound infection from contiguous spread of organisms from the skin; kidney infection from catheter-initiated ascending spread of organisms through the urinary tract; infection of intraabdominal organs and the peritoneum from perforation of the gastrointestinal tract; and gallbladder infection from retrograde migration of organisms from the gastrointestinal tract into the biliary drainage system.

However, more commonly, deeply invasive candidiasis results from hematogenous seeding of various organs as a complication of candidemia. Once the organism gains access to the intravascular compartment (either from the gastrointestinal tract or, less often, from the skin through the site of an indwelling intravascular catheter), it may spread hematogenously to a variety of deep organs. The brain, chorioretina **(Fig. 222-2)**, heart, and kidneys are most commonly infected and the liver and spleen are less commonly affected in nonneutropenic hosts (but most often involved in neutropenic patients). In fact, nearly any organ can become involved, including the endocrine glands, pancreas, heart valves (native or prosthetic), skeletal muscle, joints (native or prosthetic), bones, and meninges. *Candida* organisms can also spread hematogenously to the skin and cause classic macronodular lesions (Fig. 222-1). Frequently, painful muscular involvement is evident beneath the area of affected skin. Chorioretinal involvement and skin involvement are highly significant since both findings are associated with a very high probability of abscess formation in multiple

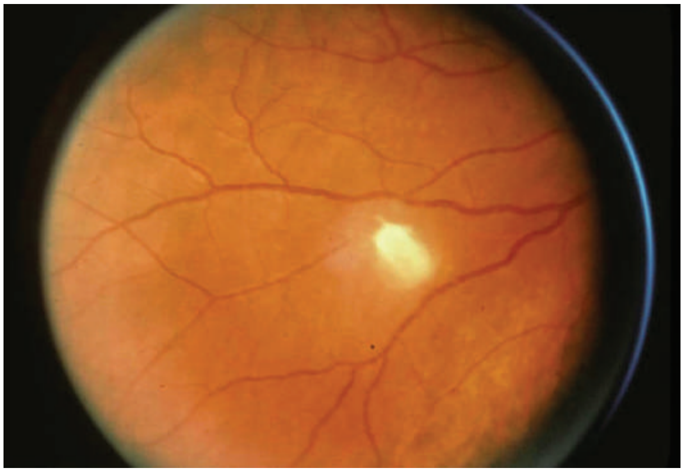

FIGURE 222-2 Hematogenous *Candida* endophthalmitis. A classic off-white lesion projecting from the chorioretina into the vitreous causes the surrounding haze. The lesion is composed primarily of inflammatory cells rather than organisms. Lesions of this type may progress to cause extensive vitreal inflammation and eventual loss of the eye. Partial vitrectomy, combined with IV and possibly intravitreal antifungal therapy, may be helpful in controlling the lesions. *(Image courtesy of Dr. Gary Holland; with permission.)*

deep organs as a result of generalized hematogenous seeding. Ocular involvement (Fig. 222-2) may require specific treatment (e.g., partial vitrectomy or intraocular injection of antifungal agents) to prevent permanent blindness. An ocular examination is indicated for patients with candidemia, whether or not they have ocular manifestations. *C. auris* invasive infections are similar to those of other *Candida* species and are most frequently associated with recent surgical procedures, immunosuppression, invasive devices such as catheters or various support or drainage tubes, and extended hospital stays. In the majority of invasive infections, *C. auris* has been isolated from the blood, but invasion of the kidney or spleen, and its recovery from cerebrospinal, bile, peritoneal, and pleural fluids demonstrate its invasiveness and dissemination potential. *C. auris*–associated candidemia can be life-threatening, with a crude mortality rate of 30–60%.

DIAGNOSIS

The diagnosis of *Candida* infection is established by visualization of pseudohyphae or hyphae on wet mount (saline and 10% KOH), tissue Gram stain, periodic acid–Schiff stain, or methenamine silver stain in the presence of inflammation. Absence of organisms on hematoxylin-eosin staining does not reliably exclude *Candida* infection. The most challenging aspect of diagnosis is determining which patients with *Candida* isolates have hematogenously disseminated candidiasis. For instance, recovery of *Candida* from sputum, urine, or peritoneal catheters may indicate mere colonization rather than deep-seated infection, and *Candida* isolation from the blood of patients with indwelling intravascular catheters may reflect inconsequential seeding of the blood from or growth of the organisms on the catheter. Despite extensive research into both antigen and antibody detection systems, there is currently no widely available and validated diagnostic test to distinguish patients with inconsequential seeding of the blood from those whose positive blood cultures represent hematogenous dissemination to multiple organs. Many studies have examined the utility of the β-glucan test; at present, its greatest utility is its negative predictive value (~90%). Meanwhile, the presence of ocular or macronodular skin lesions is highly suggestive of widespread infection of multiple deep organs. Despite extensive diagnostic tests for hematogenous dissemination, such as polymerase chain reaction and T2 technology, no test is fully validated or widely available at present. Matrix-assisted laser desorption–ionization–time-of-flight mass spectrometry (MALDI-TOF MS) is now being used extensively for detection and speciation and is useful for the correct diagnosis of *C. auris*.

C. auris can be misdiagnosed in the microbiology laboratory, often leading to inappropriate treatment and delay in the implementation of appropriate infection control measures. Preliminary testing by culturing the fungus and examination of colony morphology may help in the initial identification, but this must be confirmed with more advanced diagnostic methods. For example, features such as budding yeast morphology, absence of hyphal growth or germ tubes, and growth at 40–42°C (unlike other *Candida* species) on CHROMagar that may appear white, pink, red, or purple should raise suspicion for *C. auris* **(Fig. 222-3)**.

Several advanced molecular techniques accurately identify *C. auris* strains and therefore are being used for the follow-up testing and confirmation of the specimens that failed to be identified by traditional methods. MALDI-TOF equipment with upgraded libraries, such as Bruker Biotyper MALDI-TOF (CA System library version claim 4 or research use only [RUO] libraries versions 2014 [5627] and more recent), and using the bioMérieux VITEK MALDI-TOF MS (IVD v3.2 or RUO libraries with Saramis Ver 4.14 database and Saccharomycetaceae update), are the most common methods of *C. auris* identification. Other supplemental MALDI-TOF databases, such as MicrobeNet, which include additional *C. auris* strains from the four phylogenetic clades (i.e., South Asian, East Asian, South American, and South African) also can be used for the identification of *C. auris* strains. Sequencing of the D1–D2 region of the 28s rDNA or the internal transcribed region (ITS) of rDNA can also correctly identify *C. auris*. Recently, an automated, qualitative nucleic acid multiplex in vitro diagnostic test by GenMark called ePlex Blood Culture Identification Fungal Pathogen (BCID-FP) Panel was approved by the U.S. Food and Drug Administration for *C. auris* testing. Also, several polymerase chain reaction–based detection methods have been reported to identify

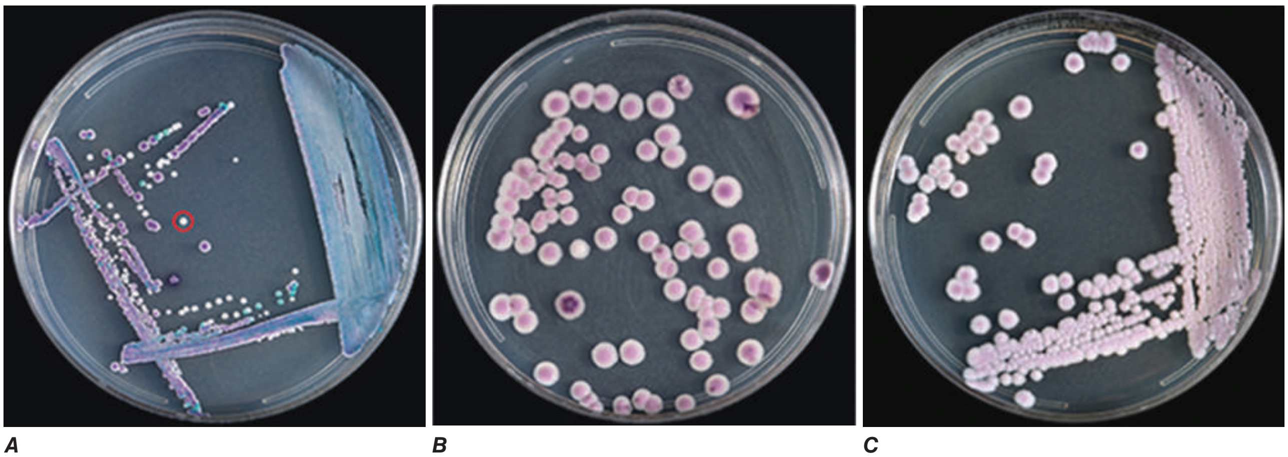

FIGURE 222-3 *C. auris* colony morphology and color on CHROMagar plates. *A.* *Candida* mixed culture: culture of *C. glabrata* (purple), *C. tropicalis* (navy blue), and *C. auris* (white, circled in red). ***B.*** *C. auris* showing multiple colony morphologies. ***C.*** *C. auris* after Salt SAB Dulcitol Broth enrichment. *(From CDC: Identification of Candida auris. Available at: https://www.cdc.gov/fungal/candida-auris/identification.html.)*

TABLE 222-2 Typical Decision-Making Steps in the Diagnosis of *C. auris*

NO.	METHOD	DATABASE/SOFTWARE	INITIAL FINDING		CONFIRMATION
1.	Bruker Biotyper MALDI-TOF	RUO libraries	*C. auris*		*C. auris*
		CA System library	*C. auris*		*C. auris*
2.	bioMérieux VITEK MS MALDI-TOF	RUO library	*C. auris*		*C. auris*
		IVD library (v3.2)	*C. auris*		*C. auris*
		Older IVD libraries	*C. haemulonii*		*C. auris* possible: Needs further workup
			C. lusitaniae		*C. auris* possible: Needs further workup
			No identification		*C. auris* possible: Needs further workup
3.	VITEK 2 YST	Software version 8.01	*C. auris*		*C. auris* confirmed
			C. haemulonii		*C. auris* possible: Needs further workup
			C. duobushaemulonii		*C. auris* possible: Needs further workup
			Candida spp. not identified		*C. auris* possible: Needs further workup
		Older versions	*C. haemulonii*		*C. auris* possible: Needs further workup
			C. duobushaemulonii		*C. auris* possible: Needs further workup
			Candida spp. not identified		*C. auris* possible: Needs further workup
4.	API 20C		*Rhodotorula glutinis,* if characteristic red color absent		*C. auris* possible: Needs further workup
			C. sake		*C. auris* possible: Needs further workup
			Candida spp. not identified		*C. auris* possible: Needs further workup
5.	API ID 32C		*C. intermedia*		*C. auris* possible: Needs further workup
			C. sake		*C. auris* possible: Needs further workup
			Saccharomyces kluyveri		*C. auris* possible: Needs further workup
6.	BD Phoenix		*C. catenulata*		*C. auris* possible: Needs further workup
			C. haemulonii		*C. auris* possible: Needs further workup
			Candida spp. not identified		*C. auris* possible: Needs further workup
7.	MicroScan		*C. lusitaniae*	No hyphal growth present	Can rule out *C. lusitaniae*, *C. guilliermondii*, and *C. parapsilosis.* *C. auris* possible: Needs further workup
			C. guilliermondii		
			C. parapsilosis		
			C. lusitaniae	Hyphal growth present	Possibly *C. lusitaniae*, *C. guilliermondii*, *C. parapsilosis*, or *C. auris*: Needs further workup
			C. guilliermondii		
			C. parapsilosis		
			C. famata		*C. auris* possible: Needs further workup
			Candida spp. not identified		*C. auris* possible: Needs further workup
8.	RapID Yeast Plus		*C. parapsilosis* → Test on corneal agar	No hyphal growth present	Can rule out *C. parapsilosis. C. auris* possible: Needs further workup
				Hyphal growth present	Possibly *C. parapsilosis or* *C. auris*: Needs further workup
			Candida spp. not identified		*C. auris* possible: Needs further workup
9.	GenMark ePlex BCID-FP Panel		*C. auris*		*C. auris* confirmed

Abbreviations: IVD, in vitro diagnostic; RUO, research use only.

Source: Adapted from CDC: Identification of *Candida auris*. Available at: *https://www.cdc.gov/fungal/candida-auris/pdf/Testing-algorithm_by-Method_508.pdf.*

C. auris in various specimens. **Table 222-2** outlines the typical decision-making steps in the diagnosis of *C. auris* by using different methods. A suspicious *C. auris* specimen is usually sent to a regional reference laboratory for further testing and confirmation of *C. auris*.

TREATMENT

Candida Infections

MUCOCUTANEOUS *CANDIDA* INFECTION

The treatment of mucocutaneous candidiasis is summarized in **Table 222-3**.

CANDIDEMIA AND SUSPECTED HEMATOGENOUSLY DISSEMINATED CANDIDIASIS

All patients with candidemia are treated with a systemic antifungal agent. A certain percentage of patients, including many of those who have candidemia associated with an indwelling intravascular catheter, probably have "benign" candidemia rather than deep-organ seeding. However, because there is no reliable way to distinguish benign candidemia from deep-organ infection, and because antifungal drugs less toxic than amphotericin B are available, antifungal treatment for candidemia—with or without clinical evidence of deep-organ involvement—has become the standard of practice. In addition, if an indwelling intravascular catheter is

TABLE 222-3 Treatment of Mucocutaneous Candidal Infections

DISEASE	PREFERRED TREATMENT	ALTERNATIVES
Cutaneous	Topical azole	Topical nystatin
Vulvovaginal	Oral fluconazole (150 mg) or ibrexafungerp (300 mg twice daily for 1 day) or azole cream or suppository	Nystatin suppository
Oral (thrush)	Fluconazole tablets (100–200 mg/d)	Clotrimazole trashes, nystatin
Esophageal	Fluconazole tablets (100–200 mg/d) or itraconazole solution (200 mg/d)	Caspofungin, micafungin, or amphotericin B

TABLE 222-4 Agents for the Treatment of Disseminated Candidiasis

AGENT	ROUTE OF ADMINISTRATION	DOSE[a]	COMMENT
Amphotericin B deoxycholate	IV only	0.5–1.0 mg/kg daily	Mostly replaced by lipid formulations
Amphotericin B lipid formulations			Not approved as primary therapy by the U.S. Food and Drug Administration, but used commonly because they are less toxic than amphotericin B deoxycholate
Liposomal (AmBiSome, Abelcet)	IV only	3.0–5.0 mg/kg daily	
Lipid complex (ABLC)	IV only	3.0–5.0 mg/kg daily	
Colloidal dispersion (ABCD)	IV only	3.0–5.0 mg/kg daily	Associated with frequent infusion reactions
Azoles[b]			
Posaconazole	IV and oral	300 mg/d (IV) 200 mg tid (oral)	Approved for prophylaxis
Fluconazole	IV and oral	400 mg/d	Most commonly used
Voriconazole	IV and oral	400 mg/d	Multiple drug interactions, visual hallucinations, fluorosis, phototoxicity Approved for candidemia in nonneutropenic patients
Echinocandins			Broad spectrum against *Candida* species; approved for disseminated candidiasis; less toxic than amphotericin B formulations
Caspofungin	IV only	50 mg/d	
Anidulafungin	IV only	100 mg/d	
Micafungin	IV only	100 mg/d	
Rezafungin	IV only	400 mg loading dose, 200 mg once weekly thereafter	

[a]For loading doses and adjustments in renal failure, see Pappas PG et al: Clinical practice guidelines for the management of candidiasis: 2016 update by the Infectious Diseases Society of America. Clin Infect Dis 62:e1, 2016. The recommended duration of therapy is 2 weeks beyond the last positive blood culture and the resolution of signs and symptoms of infection. [b]Although ketoconazole is approved for the treatment of disseminated candidiasis, it has been replaced by the newer agents listed in this table. Posaconazole has been approved for prophylaxis in neutropenic patients and for oropharyngeal candidiasis.

present, it is best to remove or replace the device whenever feasible. Moreover, an infectious disease consultation is recommended as it has been associated with improved patient outcomes.

The drugs used for the treatment of candidemia and suspected disseminated candidiasis are listed in **Table 222-4**. Various lipid formulations of amphotericin B, four echinocandins, the azoles fluconazole and voriconazole, and in some instances, the newer triazole posaconazole are used; no agent within a given class has been clearly identified as superior to the others. Most institutions choose an agent from each class on the basis of their own specific microbial epidemiology, strategies to minimize toxicities, and cost considerations. An echinocandin is the first choice of treatment. The U.S. Food and Drug Administration (FDA) recently approved the new-generation echinocandin rezafungin for the treatment of candidemia and invasive candidiasis in adult patients; rezafungin has a prolonged half-life, which allows for once-weekly 200-mg dosing following a 400-mg front-loading dose. Echinocandin treatment continues until sensitivities or speciation is determined. In stable patients, many centers then switch to fluconazole if a sensitive strain is identified and there is no evidence of hematogenous dissemination. For hemodynamically unstable or neutropenic patients, initial treatment with echinocandins is warranted, and once the clinical response has been assessed and the pathogen specifically identified, the regimen can be altered according to the sensitivities. At present, the vast majority of *C. albicans* isolates are sensitive to fluconazole. Isolates of *C. glabrata* and *C. krusei* are less sensitive to fluconazole and more sensitive to polyenes and echinocandins. *C. parapsilosis* is less sensitive to echinocandins in vitro; however, this lesser sensitivity is considered clinically insignificant. Posaconazole has been approved for prophylaxis, including against *Candida*, in neutropenic patients. Itraconazole is rarely used for *Candida* nowadays, and isavuconazole is not recommended for this indication.

Antifungal drug resistance is one of the hallmarks of *C. auris* infections. Some *C. auris* strains have multidrug resistance with elevated minimal inhibitory concentrations (MICs) to all three major antifungal classes—azoles, echinocandins, and polyenes—resulting in limited treatment options. A CDC study reported antifungal resistance in *C. auris* strains obtained from 54 patients in India, Pakistan, South Africa, and Venezuela: 93% were resistant to fluconazole, 35% to amphotericin B, and 7% to echinocandins; 41% of the tested strains were resistant to two antifungal classes, and, alarmingly, 4% of the tested strains were resistant to all three classes of antifungal drugs. A 2023 CDC report indicated that rates of *C. auris* echinocandin resistance tripled in 2021 compared to the prior 2 years. Almost all *C. auris* strains that have been identified have elevated MICs for fluconazole with variable susceptibilities to other triazoles **(Table 222-5)**, associated with mutations in *ERG11*-encoded lanosterol demethylase and/or overexpression of drug transporters/efflux pumps.

Due to the high rates of azole resistance among *C. auris* strains, the use of echinocandins is recommended as first-line therapy for *C. auris* infection. By contrast, the CDC discourages the use of antifungal drugs for the treatment of colonization of *C. auris* in the absence of invasive or bloodstream infection. A history of patient travel or residence in a health care or nursing facility with a known

TABLE 222-5 Typical MICs of Available Antifungal Drugs for *C. auris*

DRUG	TENTATIVE RESISTANCE BREAKPOINTS[a]	MIC RANGE, µg/mL		
		MIC	MIC_{50}	MIC_{90}
Amphotericin B	≥2	0.06–8	0.5–1	2–4
Fluconazole	≥32	0.12–≥64	≥64	≥64
Itraconazole	N/A	0.032–2	0.06–0.5	0.25–1
Voriconazole	N/A	0.032–16	0.5–2	2–8
Posaconazole	N/A	0.015–16	0.016–0.5	0.125–2
Isavuconazole	N/A	0.015–4	0.125–0.25	0.5–2
Caspofungin	≥2	0.03–16	0.25–1	1–2
Anidulafungin	≥4	0.015–16	0.125–0.5	0.5–1
Micafungin	≥4	0.015–8	0.125–0.25	0.25–2

[a]Tentative resistance breakpoints per Centers for Disease Control and Prevention (CDC) (*www.cdc.gov/fungal/candida-auris/c-auris-antifungal.html*).

Abbreviations: MIC, minimum inhibitory concentration; N/A, not available.

Source: Adapted from CDC: Antifungal susceptibility testing and interpretation. Available at: *www.cdc.gov/fungal/candida-auris/c-auris-antifungal.html*.

TABLE 222-6 List of CDC-Recommended Echinocandin Doses for the Treatment of *C. auris* Infections

DRUG	ADULTS	CHILDREN (>2 MONTHS)	INFANTS (<2 MONTHS)
Caspofungin	Loading dose 70 mg IV, then 50 mg IV daily	Loading dose 70 mg/m^2 per day IV, then 50 mg/m^2 per day IV	25 mg/m^2 per day IV
Anidulafungin	Loading dose 200 mg IV, then 100 mg IV daily	Not approved for use in children	Not approved for use in children
Micafungin	100 mg IV daily	2 mg/kg per day IV with option to increase to 4 mg/kg per day IV in children at least 40 kg	10 mg/kg per day IV

Abbreviation: CDC, Centers for Disease Control and Prevention.

Source: Adapted from CDC: Treatment and management of infections and colonization. Available at: *www.cdc.gov/fungal/candida-auris/c-auris-treatment.html*.

outbreak of *C. auris* infection, as well as drug susceptibility data of identified strains, act as a guide for the effective choice of treatment of invasive and bloodstream infections. *C. auris* is known to develop antibiotic resistance during treatment. Therefore, the emergence of antifungal resistance should be closely monitored with follow-up cultures and repeat susceptibility testing. Antibiotic stewardship should be implemented to ameliorate the risk of development of drug resistance. Patients may remain colonized with *C. auris* during or after the successful treatment of invasive *C. auris* infection. Therefore, infection control measures should be implemented throughout patient care. Table 222-6 outlines CDC-recommended echinocandin doses for the initial antifungal treatment for *C. auris* infections.

In cases of echinocandin resistance, liposomal amphotericin B (5 mg/kg per day) can be considered. For neonates and infants (<2 months old), amphotericin B deoxycholate (1 mg/kg per day) treatment can be initiated. If this fails, liposomal amphotericin B (5 mg/kg per day) can be given. In very severe cases, if all treatment options fail, echinocandins per CDC recommendations can be given (Table 222-6). Other considerations for *C. auris* infection management can be referenced from the 2016 Infectious Diseases Society of America (IDSA) Clinical Practice Guideline for the Management of Candidiasis.

Some generalizations exist regarding the management of specific *Candida* infections. Recovery of *Candida* from sputum is almost never indicative of underlying pulmonary candidiasis and does not by itself warrant antifungal treatment. Similarly, *Candida* in the urine of a patient with an indwelling bladder catheter may represent colonization only, rather than bladder or kidney infection. However, the threshold for systemic treatment is lower in general in severely ill patients in this category since it is impossible to distinguish colonization from lower or upper urinary tract infection. If the isolate is *C. albicans*, most clinicians use oral fluconazole rather than a bladder washout with amphotericin B, which was more commonly used in the past. Although echinocandins are poorly excreted into the urine, they may be an option, especially for non-*albicans* isolates. The doses and duration are the same as for disseminated candidiasis. The significance of the recovery of *Candida* from abdominal drains in postoperative patients is unclear, but again, the threshold for treatment is generally low because most of the affected patients have been subjected to risk factors predisposing them to disseminated candidiasis. In addition, there has been a considerable increase in the recognition and diagnosis of intraabdominal candidiasis.

Removal of the infected valve and long-term antifungal administration constitute appropriate treatment for *Candida* endocarditis. Although definitive studies are not available, patients usually are treated for weeks with a systemic antifungal agent (Table 222-4) and then given chronic suppressive therapy for months or years (sometimes indefinitely) with an oral azole (usually fluconazole at 400–800 mg/d).

Hematogenous *Candida* endophthalmitis is a special problem requiring ophthalmologic consultation. When lesions are expanding or are threatening the macula, an IV polyene combined with flucytosine (25 mg/kg four times daily) has been the regimen of choice, although comparative studies with other regimens have not yet been reported. As more data on the newer triazoles (e.g., voriconazole) and the echinocandins become available, new strategies involving these agents are developing, although it is important to note that echinocandins exhibit low penetration in ocular tissue. Of paramount importance is the decision to perform a partial vitrectomy. This procedure debulks the infection and can preserve sight, which may otherwise be lost due to vitreal scarring. All patients with candidemia should undergo ophthalmologic examination because of the relatively high frequency of this ocular complication (up to 15–20% in some case series). This examination can detect a developing eye lesion early in its course; in addition, identification of a lesion signifies a probability of ~90% of deep-organ abscesses and may prompt prolongation of therapy for candidemia beyond the recommended 2 weeks after the last positive blood culture. Although the basis for the consensus is a very small data set, the recommended treatment for *Candida* meningitis is a polyene (Table 222-4) plus flucytosine (25 mg/kg four times daily). Development of *Candida* meningoencephalitis in an otherwise immunocompetent individual should raise suspicion for deficiency in the C-type lectin receptor adaptor molecule CARD9 and should prompt genetic testing to rule out this monogenic disorder. Successful treatment of *Candida*-infected prosthetic material (e.g., an artificial joint) nearly always requires removal of the infected material followed by long-term administration of an antifungal agent selected on the basis of the isolate's sensitivity and the logistics of administration.

PROPHYLAXIS

The use of antifungal agents to prevent *Candida* infections has been controversial, but some general principles have emerged. Most centers administer prophylactic fluconazole (400 mg/d) to recipients of allogeneic hematopoietic stem cell transplantation. High-risk liver transplant recipients are also given fluconazole prophylaxis in most centers. The use of prophylaxis for neutropenic patients has varied considerably from center to center; many centers that elect to give prophylaxis to this population use either fluconazole (200–400 mg/d) or a lipid formulation of amphotericin B (AmBisome, 1–2 mg/d). Caspofungin (50 mg/d) also has been used. Posaconazole (200 mg three times daily) has been approved by the FDA for prophylaxis in neutropenic patients; it is gaining in popularity and may replace fluconazole in settings when mold activity is desired.

Prophylaxis is sometimes given to surgical patients at very high risk for candidiasis. The widespread use of prophylaxis for nearly all patients in general surgical or medical intensive care units is not—and should not be—a common practice for three reasons: (1) the incidence of disseminated candidiasis is relatively low, (2) the cost–benefit ratio is suboptimal, and (3) increased resistance with widespread prophylaxis is a valid concern.

Prophylaxis for oropharyngeal or esophageal candidiasis in HIV-infected patients is not recommended unless there are frequent recurrences.

FURTHER READING

Lionakis MS, Edwards JE jr: *Candida* species, in *Mandell, Douglas, and Bennett's Principles of Infectious Diseases*, 10th ed. Blaser MJ et al (eds). Philadelphia, Elsevier, 2025.

Pappas PG et al: Invasive candidiasis. Nat Rev Dis Primers 62:e1, 2018.

Pechacek J, Lionakis MS: Host defense mechanisms against *Candida auris*. Expert Rev Anti Infect Ther 21:1087, 2023.

Proctor DM et al: Integrated genomic, epidemiologic investigation of *Candida auris* skin colonization in a skilled nursing facility. Nat Med 27:1401, 2021.

Santana DJ et al: A *Candida auris*-specific adhesin, Scf1, governs surface association, colonization, and virulence. Science 381:1461, 2023.

Tsai SV et al: Burden of candidemia in the United States, 2017. Clin Infect Dis 71:e449, 2020.

223 Aspergillosis

David W. Denning

Aspergillosis is the collective term used to describe all disease entities caused by any one of ~50 pathogenic and allergenic species of *Aspergillus*. Only those species that grow at 37°C can cause invasive infection, although some species without this ability can cause allergic syndromes. Each common pathogenic species is actually a complex of many species (many of them cryptic) but is referred to as a single species here for simplicity. *A. fumigatus* is responsible for most cases of invasive aspergillosis, >95% of cases of chronic aspergillosis, and most allergic syndromes. *A. flavus* is more prevalent in some hospitals and causes a higher proportion of cases of sinus infections, cutaneous infections, and keratitis than *A. fumigatus*. *A. niger* can cause invasive infection but more commonly colonizes the respiratory tract and causes external otitis. *A. terreus* causes only invasive disease, usually with a poor prognosis. *A. nidulans* occasionally causes invasive infection, primarily in patients with chronic granulomatous disease (CGD).

EPIDEMIOLOGY AND ECOLOGY

Aspergillus has a worldwide distribution, most commonly growing in decomposing plant materials (i.e., compost) and in bedding. This hyaline (nonpigmented), septate, branching mold produces vast numbers of conidia (spores) on stalks above the surface of mycelial growth. *Aspergilli* are found in indoor and outdoor air, on surfaces, and in water from surface reservoirs. Daily exposures vary from a few to many millions of conidia; high numbers of conidia are encountered in hay barns and other very dusty environments. The required size of the infecting inoculum is uncertain; however, only intense exposures (e.g., during construction work, handling of moldy bark or hay, or composting) are demonstrated to cause disease—acute community-acquired pulmonary aspergillosis—in healthy immunocompetent individuals or exacerbations of COPD. Allergic syndromes are exacerbated by continuous antigenic exposure arising from sinus or airway colonization or from nail infection. High-efficiency particulate air (HEPA) filtration is often protective against infection; thus, HEPA filters should be installed and monitored for efficiency in operating rooms, burn units and in areas of the hospital that house high-risk patients.

The incubation period of invasive aspergillosis after exposure is highly variable, extending in documented cases from 2 to 90 days. Thus, community acquisition of an infecting strain frequently manifests as invasive infection during hospitalization, although nosocomial acquisition is also common. Outbreaks usually are directly related to a contaminated air source or construction in the hospital.

Global aspergillosis incidence and prevalence have been estimated (Table 223-1). The frequency of different manifestations of aspergillosis varies considerably with geographic location; most notably, chronic granulomatous sinusitis is rare outside the Middle East and India. Fungal (mycotic) keratitis is particularly common in Southeast Asia but occurs globally. Chronic pulmonary aspergillosis follows pulmonary tuberculosis in ~6–13% of treated cases and also mimics pulmonary tuberculosis as smear-negative or "clinically diagnosed" tuberculosis. *Aspergillus* onychomycosis, usually of the toenail, has been reported in as low as <1% and as high as 35% of cases of onychomycosis and is more common in diabetes.

TABLE 223-1 Disease Frequency and Diagnostic Sensitivity for Different Manifestations of Aspergillosis

	TYPE OF DISEASE		
PARAMETER	**INVASIVE**	**CHRONIC**	**ALLERGIC**
Incidence/100,000[a]	27.6	23.6	?[b]
Prevalence/100,000[a]	—	55.4	286[c]
Global burden[a]	~2,116,000	~6,141,000	~11,690,000
Mortality rate without treatment[a]	~100%	~50%	<1%
Respiratory Diagnostic Sensitivity[d]			
Culture[e]	√	√-√√[e]	√-√√[e]
Microscopy	√	√	√
Antigen	√√√	√√	?
Real-time PCR	√√	√√	√√
Blood Diagnostic Sensitivity[d]			
Culture	x	x	x
Antigen	√√	√	√
β-D-Glucan	√√	√	?
Real-time PCR	√√	x	x
IgG antibody[f]	√√[f]	√√√	√√√
IgE antibody	x	√√	√√√√

[a] Denning DW: Global incidence and mortality of severe fungal disease. Lancet Infect Dis 24:e428, 2024. [b] Allergic fungal disease can develop at any age, usually in adulthood; the annual frequency with which it occurs is not known. [c] Allergic bronchopulmonary aspergillosis and severe asthma with fungal sensitization. [d] Key for sensitivity: 1 check = limited (as the text indicates, 10–30% for culture); 2 checks = higher; 3 checks = >80%; and 4 checks = ~95%. [e] Multiple specimens, fungal media, and high-volume fungal culture increase sensitivity. [f] High cross-reactivity between *A. fumigatus* and other *Aspergillus* spp. [g] IgG is a useful assay in subacute invasive and nonneutropenic patients but may not be positive on initial testing.

Abbreviation: PCR, polymerase chain reaction.

RISK FACTORS AND PATHOGENESIS

The primary risk factors for invasive aspergillosis are profound neutropenia, glucocorticoid use, and underlying respiratory disease; risk increases with longer duration of these conditions. Higher doses of glucocorticoids increase the risk of both acquisition of invasive aspergillosis and death from the infection. Neutrophil and/or phagocyte dysfunction also is an important risk factor, as evidenced by aspergillosis in CGD, advanced HIV infection, and relapsed leukemia. Invasive aspergillosis is increasingly recognized (if actively sought) in medical intensive care units (2–5%), those with severe influenza (8–25%) and severe COVID-19 (~10%), and patients admitted to hospital with chronic obstructive pulmonary disease (COPD; 1.3–3.9%). Temporary abrogation of protective responses from glucocorticoid use or compensatory anti-inflammatory response syndrome is a significant risk factor. Many patients have some evidence of prior pulmonary disease—typically, a history of pneumonia or COPD. Many new immunomodulating agents, such as infliximab and ibrutinib, increase the risk of invasive aspergillosis, as does severe liver disease.

Patients with chronic pulmonary aspergillosis have a wide spectrum of underlying pulmonary disease, including current and prior tuberculosis, sarcoidosis, prior pneumothorax, or COPD. These patients are apparently immunocompetent, but natural killer and/or interleukin 12 or gamma interferon production defects are common. Their inflammatory immune (T_H1-like) response is suboptimal, and persistent inflammation is typical. Glucocorticoids accelerate disease progression.

Allergic bronchopulmonary aspergillosis (ABPA) usually complicates asthma and cystic fibrosis. Many genetic associations indicate a strong basis for the development of a T_H2-like and "allergic" response to *A. fumigatus*. Remarkably, high-dose glucocorticoid treatment for exacerbations of ABPA almost never leads to invasive aspergillosis. Fungal, and especially *Aspergillus*, sensitization is especially common in those with poorly controlled asthma. Environmental exposure and *Aspergillus* sensitization drive some COPD exacerbations. Most patients with *Aspergillus* bronchitis have bronchiectasis, with or without cystic fibrosis, including heterozygous *CFTR* mutations.

Different genetic traits are associated with invasive, chronic, and allergic aspergillosis; the majority of people probably are not at risk for aspergillosis. Multiple gene variants appear to be necessary for susceptibility to each form of aspergillosis.

CLINICAL FEATURES AND APPROACH TO THE PATIENT (Table 223-2)

Invasive Pulmonary Aspergillosis Both the frequency of invasive disease and the pace of its progression increase with greater

TABLE 223-2 Major Manifestations of Aspergillosis

ORGAN	TYPE OF DISEASE			
	INVASIVE (ACUTE AND SUBACUTE)	CHRONIC	SAPROPHYTIC	ALLERGIC
Lung	Angioinvasive (in neutropenia), nonangioinvasive, granulomatous	Chronic cavitary, chronic fibrosing, bronchitis, *Aspergillus* nodule	Aspergilloma (single), airway colonization	Allergic bronchopulmonary, severe asthma with fungal sensitization, hypersensitivity pneumonitis
Sinus	Acute invasive	Chronic invasive, chronic granulomatous	Maxillary fungal ball	Allergic fungal sinusitis, eosinophilic fungal rhinosinusitis
Brain	Abscess, hemorrhagic infarction, meningitis	Granulomatous, meningitis	None	None
Skin	Acute disseminated, locally invasive (trauma, burns, IV access)	External otitis, onychomycosis	None	None
Heart	Endocarditis (native or prosthetic), pericarditis	None	None	None
Eye	Keratitis, endophthalmitis	None	None	None described

degrees of immunocompromise. Invasive aspergillosis is arbitrarily classified as acute and subacute, with courses of ≤1 month and 1–3 months, respectively. More than 80% of cases of invasive aspergillosis involve the lungs, and most are community acquired. The most common clinical features are no symptoms at all, fever, cough (sometimes productive), nondescript chest discomfort, trivial hemoptysis, and shortness of breath. Although the fever often responds to glucocorticoids, the disease progresses. In ventilated patients, screening for *Aspergillus* antigen on tracheobronchial lavage fluid is necessary for diagnosis as radiology is not distinctive. The keys to early diagnosis in at-risk patients are a high index of suspicion, screening for circulating antigen (in leukemia), and urgent computed tomography (CT) of the thorax. Invasive aspergillosis is one of the most common diagnostic errors revealed at autopsy.

Invasive Sinusitis The sinuses are involved in 5–10% of cases of invasive aspergillosis, especially affecting patients with leukemia and recipients of hematopoietic stem cell transplants. In addition to fever, the most common features are nasal or facial discomfort, blocked nose, and nasal discharge (sometimes bloody). Endoscopic examination of the nose reveals pale, dusky or necrotic-looking tissue in any location. CT or magnetic resonance imaging (MRI) of the sinuses is essential but does not distinguish invasive *Aspergillus* sinusitis from preexisting allergic or bacterial sinusitis early in the disease process.

Tracheobronchitis Occasionally, only the airways are infected by *Aspergillus*. The resulting manifestations seen on bronchoscopy range from acute or chronic bronchitis to ulcerative or pseudomembranous tracheobronchitis. These entities are particularly common among lung transplant recipients and patients on artificial ventilation. Obstruction with mucous plugs may occur and is called obstructing bronchial aspergillosis in immunocompromised patients and mucous impaction in other patients, such as those with ABPA.

Aspergillus Bronchitis Recurrent chest infections that only partially improve with antibiotic treatment and are associated with significant breathlessness or coughing up of thick sputum plugs are typical features of *Aspergillus* bronchitis. Patients are not significantly immunocompromised and usually have bronchiectasis or cystic fibrosis. Occasional patients present with respiratory failure because of airway obstruction with mucus. Concurrent bacterial bronchitis is common. The diagnosis rests on recurrent detection of *Aspergillus* in the airway by microscopy, culture, or polymerase chain reaction (PCR). *Aspergillus* IgG is usually detectable.

Disseminated Aspergillosis In the most severely immunocompromised patients, *Aspergillus* disseminates from the lungs to multiple organs—most often to the brain but also to the skin, thyroid, bone, kidney, liver, gastrointestinal tract, eye (endophthalmitis), and heart valve. Aside from cutaneous lesions, the most common features are gradual clinical deterioration over 1–3 days, with low-grade fever and features of mild sepsis, and nonspecific abnormalities in laboratory tests. In most cases, at least one localization becomes apparent before death. Blood cultures are almost always negative.

Cerebral Aspergillosis Hematogenous dissemination to the brain is a devastating complication of invasive aspergillosis. Single or multiple lesions may develop. In acute disease, hemorrhagic infarction is most typical, and cerebral abscess is common. Rarer manifestations include meningitis, mycotic aneurysm, and cerebral granuloma (mimicking a brain tumor). Local spread from cranial sinuses also occurs. Postoperative infection develops rarely and is exacerbated by glucocorticoids, which are often given after neurosurgery. The presentation can be either acute or subacute, with mood changes, focal signs, seizures, and decline in mental status. MRI is the most useful immediate investigation; unenhanced CT of the brain is usually nonspecific, and contrast is often contraindicated because of poor renal function. Cerebral aspergillosis is disproportionately common in those on ibrutinib.

Endocarditis Most cases of *Aspergillus* endocarditis are prosthetic-valve infections resulting from contamination during surgery. Native-valve disease is reported, especially as a feature of disseminated infection and in persons using illicit IV drugs. Culture-negative endocarditis with large vegetations is the most common presentation; embolectomy occasionally reveals the diagnosis.

Cutaneous Aspergillosis Dissemination of *Aspergillus* occasionally results in cutaneous features, usually an erythematous or purplish nontender area that progresses to a necrotic eschar. Direct invasion of the skin occurs in neutropenic patients at the site of IV catheter insertion and in burn patients. Surgical, burn, and trauma wounds may become infected with *Aspergillus* (especially *A. flavus*).

Chronic Pulmonary Aspergillosis The hallmark of chronic cavitary pulmonary aspergillosis (Fig. 223-1) is one or more pulmonary cavities expanding over a period of months or years in association with pulmonary symptoms and systemic manifestations such as fatigue and weight loss. Often mistaken initially for tuberculosis, >90% of chronic cavitary pulmonary aspergillosis cases occur in patients with prior pulmonary disease (e.g., current or past tuberculosis or nontuberculous mycobacterial infection, sarcoidosis, rheumatoid lung disease, pneumothorax, bullae) or lung surgery. The onset is insidious, and systemic features (weight loss, fatigue) may be more prominent than pulmonary symptoms. On CT scan, an irregular internal cavity surface and thickened cavity walls are typical and indicative of disease activity. Irregular material, fluid level, and a well-formed fungal ball are seen in some cavities. Multiple cavities are more common than a single cavity, and most cavities are in the upper lobes. Pleural thickening and pericavitary infiltrates are typical and most obvious if a positron emission tomography scan has been done as part of the workup. Chronic cavitary pulmonary aspergillosis is usually caused by *A. fumigatus*, but *A. niger* has been implicated, particularly in diabetic patients, as have other species. IgG antibodies to *Aspergillus* are detectable in ~90% of patients with chronic cavitary pulmonary aspergillosis. Some patients have concurrent infections with mycobacteria and/or other bacterial pathogens. The most significant complication is life-threatening hemoptysis, which may be the presenting manifestation. If untreated, chronic cavitary pulmonary aspergillosis typically

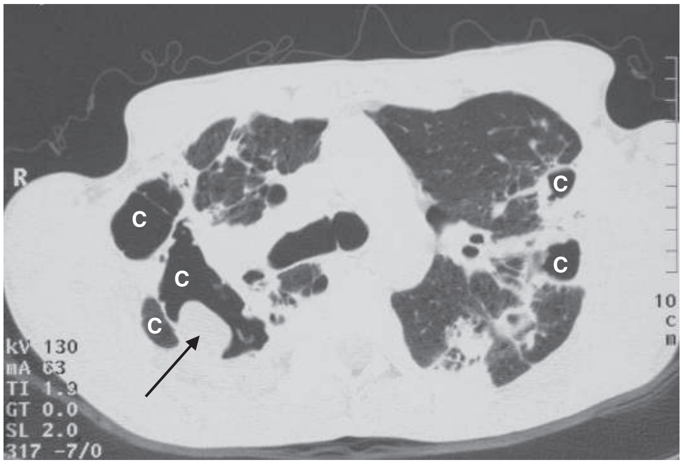

FIGURE 223-1 Computed tomography (CT) scan image of the chest in a patient with long-standing bilateral chronic cavitary pulmonary aspergillosis. This patient had a history of several bilateral pneumothoraces and had required bilateral pleurodesis in 1990. CT then demonstrated multiple bullae, and sputum cultures grew *A. fumigatus*. The patient had initially weakly and later strongly positive serum IgG *Aspergillus* antibody tests. This scan (2003) shows a mixture of thick- and thin-walled cavities in both lungs (each marked with *C*), with a probable fungal ball (*black arrow*) protruding into the large cavity on the patient's right side (*R*). There is also considerable pleural thickening bilaterally.

progresses (sometimes relatively rapidly) to unilateral or upper-lobe fibrosis. This end-stage entity is termed *chronic fibrosing pulmonary aspergillosis* (destroyed lung).

Aspergilloma Aspergilloma (fungal ball) is a late manifestation of chronic cavitary pulmonary aspergillosis, but some patients are asymptomatic. The inside of a pulmonary cavity allows fungal growth that peels off, forming the layers of the fungal ball. Signs and symptoms associated with single (simple) aspergillomas are minor, including cough (sometimes productive), hemoptysis, wheezing, and mild fatigue. More significant signs and symptoms are associated with chronic cavitary pulmonary aspergillosis and should be treated as such. About 10% of fungal balls resolve spontaneously (by being coughed up), but the cavity may still be infected and the patient symptomatic.

***Aspergillus* Nodule** A recently recognized form of chronic pulmonary aspergillosis is the *Aspergillus* nodule, which may resemble early lung carcinoma and may cavitate. Nodules may be single or multiple and 5–50 mm in diameter. Larger mass lesions are rarely seen. Nodules are usually avid on positron emission tomography. IgG antibodies to *Aspergillus* are detectable in ~65% of patients with an *Aspergillus* nodule.

Chronic *Aspergillus* Sinusitis Three entities are subsumed under this broad designation: fungal ball of the sinus, chronic invasive sinusitis, and chronic granulomatous sinusitis. Fungal ball of the sinus is limited to the maxillary sinus (except in rare cases involving the sphenoid sinus) in which the sinus cavity is filled with a fungal ball. Maxillary disease is associated with prior upper-jaw root canal work and chronic (bacterial) sinusitis. About 90% of CT scans show focal hyperattenuation related to concretions; on MRI scans, the T2-weighted signal is decreased, whereas it is increased in bacterial sinusitis. Removal of the fungal ball is curative. No tissue invasion is demonstrable histologically or radiologically.

In contrast, chronic invasive sinusitis is a slowly destructive process that most commonly affects the ethmoid and sphenoid sinuses. Patients are usually but not always immunocompromised to some degree (e.g., as a result of diabetes or HIV infection). Imaging of the cranial sinuses shows opacification of one or more sinuses, local bone destruction, and invasion of local structures. The differential diagnosis is wide, including other infections. Apart from a history of chronic nasal discharge and blockage, loss of the sense of smell, and persistent headache, the usual presenting features are related to local involvement of critical structures. The orbital apex syndrome (blindness and proptosis) is characteristic. Facial swelling, cavernous sinus thrombosis, carotid artery occlusion, and pituitary fossa and brain and skull-base invasion are complications.

Chronic granulomatous sinusitis due to *Aspergillus* is most commonly seen in the Middle East and India and is often caused by *A. flavus*. It typically presents late, with facial swelling and unilateral proptosis. The prominent granulomatous reaction histologically distinguishes this disease from chronic invasive sinusitis, in which tissue necrosis with a low-grade mixed-cell infiltrate is typical. IgG antibodies to *A. flavus* are usually detectable.

Allergic Bronchopulmonary Aspergillosis In almost all cases, ABPA represents a hypersensitivity reaction to *A. fumigatus*; other causes are other aspergilli and other fungi. ABPA occurs in ~2.5% of patients with asthma who are referred to secondary care, although it may be less common in the United States and more common in those from the Indian subcontinent. In cystic fibrosis, up to 15% of teenagers are affected. Episodes of bronchial obstruction with mucous plugs leading to coughing fits, "pneumonia," consolidation, and breathlessness are typical. Many patients report coughing up thick sputum casts, often brown in color. Eosinophilia commonly develops before systemic glucocorticoids are given. The cardinal diagnostic test is detection of *Aspergillus*-specific IgE (or a positive skin-prick test in response to *A. fumigatus* extract) together with an elevated serum level of total IgE (usually >1000 IU/mL). The presence of hyperattenuated mucus in airways is highly specific. Bronchiectasis is characteristic, and a few patients develop chronic cavitary pulmonary aspergillosis.

Severe Asthma with Fungal Sensitization (SAFS) Many adults with severe asthma do not fulfill the criteria for ABPA and yet are allergic to fungi. Although *A. fumigatus* is a common allergen, numerous other fungi (e.g., *Cladosporium* and *Alternaria* species) are implicated by skin-prick testing and/or specific IgE testing. Serum total IgE concentrations are <1000 IU/mL, and bronchial-wall thickening is common. ABPA and SAFS are collectively referred to as *fungal asthma*.

Allergic Fungal Rhinosinusitis Like the lungs, the sinuses manifest allergic responses to *Aspergillus* and other fungi. The affected patients present with chronic (i.e., perennial) sinusitis that is relatively unresponsive to antibiotics. Many of these patients have nasal polyps, and all have congested nasal mucosae and sinuses full of mucoid material. The histologic hallmarks of allergic fungal sinusitis are local eosinophilia and Charcot-Leyden crystals. Removal of abnormal mucus and polyps, with local and occasionally systemic administration of glucocorticoids, usually leads to resolution. Persistent or recurrent signs and symptoms may require more extensive surgery (ethmoidectomy) and sometimes oral antifungal therapy. Recurrence is common, often after another bacterial or viral infection.

Superficial Aspergillosis *Aspergillus* can cause keratitis, onychomycosis, and otitis externa. The first can be difficult to diagnose early enough to save the patient's sight. Natamycin (5%) eye drops are the optimal therapy for fungal keratitis, often with surgery. Otitis externa usually resolves with debridement and local application of antifungal agents.

■ DIAGNOSIS

Several techniques are required to establish the diagnosis of any form of aspergillosis with confidence (Table 223-1).

Acute Invasive Aspergillosis Patients with acute invasive aspergillosis have a relatively heavy load of fungus in the affected organ; thus, antigen detection, PCR, microscopy, culture, and/or histopathology usually confirm the diagnosis. However, the pace of progression leaves only a narrow window for making the diagnosis without losing the patient, and some invasive procedures are not possible because of coagulopathy, respiratory compromise, and other factors. Many cases of invasive aspergillosis are missed clinically and are diagnosed only at autopsy. Histologic examination of affected tissue reveals either infarction, with invasion of blood vessels by many fungal hyphae, or acute necrosis, with limited inflammation and fewer hyphae. *Aspergillus* hyphae are hyaline, narrow, and septate, with branching at 45° in infected tissue. Hyphae can be seen in cytology or microscopy preparations, which therefore provide a rapid means of presumptive diagnosis.

One *Aspergillus* antigen test relies on detection of galactomannan release from *Aspergillus* organisms during growth, the other a novel protein antigen. Respiratory sample antigen detection is more sensitive than serum and is critical in the intensive care unit patient in whom radiology is nonspecific. Positive serum antigen results usually precede clinical or radiologic features by several days. The sensitivity of antigen detection is reduced by antifungal therapy.

A positive culture supports the diagnosis, given that multiple other (rarer) fungi can mimic *Aspergillus* species histologically, but only 10–30% of patients with invasive aspergillosis have a positive culture. Bacterial agar is less sensitive than fungal media for culture; thus, if physicians do not request fungal culture, the diagnosis may be missed. High-volume fungal cultures enhance yield. A positive culture may represent noninvasive forms of aspergillosis or airway colonization. Both antigen detection and real-time PCR are faster and much more sensitive than culture of respiratory samples and blood.

Definitive confirmation of a diagnosis of invasive aspergillosis requires (1) a positive culture of a sample taken directly from an ordinarily sterile site (e.g., a brain abscess) or (2) positive results of both histologic testing and culture (or molecular confirmation of *Aspergillus* spp.) of a sample taken from an affected organ (e.g., sinuses or skin). Most diagnoses of invasive aspergillosis are inferred from fewer data, including the presence of the *halo sign* on a thoracic CT scan—a localized ground-glass appearance representing hemorrhagic infarction surrounds a nodule or consolidation. Halo signs are present for ~7 days early in the course of infection in neutropenic patients and are a good prognostic feature, reflecting an early diagnosis. Nodules with halo signs are a feature of COVID-19 itself; invasive aspergillosis diagnosis requires supportive evidence. Other characteristic radiologic features of invasive pulmonary aspergillosis include nodules and pleural-based infarction or cavitation, but nonspecific consolidation is common (Fig. 223-2).

Chronic Aspergillosis For chronic aspergillosis, *Aspergillus* antibody testing combined with characteristic imaging is sufficient for the diagnosis. Biopsy of *Aspergillus* nodules reveals hyphae surrounded by cells of chronic inflammation and sometimes granulomas. Antibody titers fall slowly with successful therapy. Cultures are infrequently positive but are important in checking for azole resistance. PCR of sputum is often strongly positive. Some patients with chronic pulmonary aspergillosis often have elevated titers of total and *Aspergillus*-specific IgE.

ABPA, SAFS, and Allergic *Aspergillus* Sinusitis ABPA and SAFS are diagnosed serologically with elevated specific and total serum IgE levels or with skin-prick tests. Allergic *Aspergillus* sinusitis is usually diagnosed histologically, accompanied by *Aspergillus* IgE antibody.

TREATMENT

Aspergillosis

Antifungal drugs active against *Aspergillus* include voriconazole, itraconazole, posaconazole, isavuconazole, caspofungin, micafungin, and amphotericin B (AmB). Possible interactions with other drugs must be considered before azoles are prescribed. In addition, plasma itraconazole and voriconazole concentrations vary substantially from one patient to another, and many authorities recommend monitoring levels to ensure that drug concentrations are adequate but not excessive. Initial IV administration is preferred for acute invasive aspergillosis and oral administration for all other diseases that require antifungal therapy. Current recommendations are shown in Table 223-3.

Voriconazole, isavuconazole, and posaconazole are the preferred agents for invasive aspergillosis; caspofungin, micafungin, and lipid-associated AmB are second-line agents. AmB is not active against *A. terreus* or *A. nidulans*; multi-azole resistance in *A. fumigatus* is present in 3–20% of isolates but is increasing, especially in Southeast Asia; and *A. niger* is resistant to isavuconazole. An infectious disease consultation is advised for patients with invasive disease, given the complexity of management. Immune reconstitution can complicate recovery. The duration of therapy for invasive aspergillosis varies from ~3 months to several years, depending on the patient's immune status and response to therapy. Relapse occurs if the response is suboptimal and immune reconstitution is not complete.

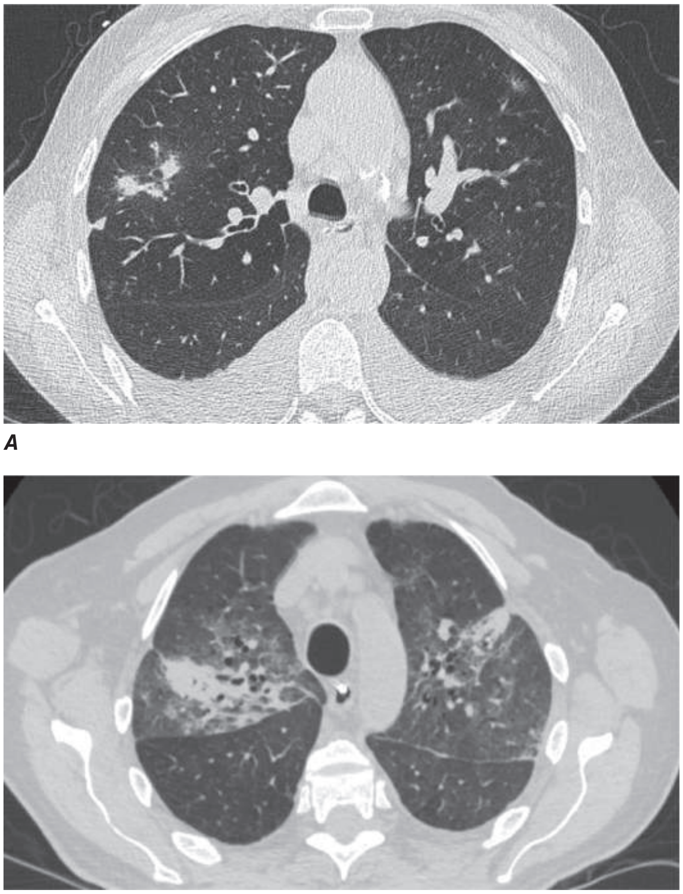

FIGURE 223-2 Markedly different appearances of invasive aspergillosis on computed tomography scan of the thorax. ***A.*** Patient with myelodysplasia and moderate neutropenia showing small right-sided nodules with minimal surrounding ground glass and a separate area of ground glass only on the left laterally. ***B.*** Patient with multiple myeloma undergoing intensive chemotherapy with corticosteroids showing bilateral areas of consolidation and some nonspecific atelectasis with probable ground glass surrounding the right-sided lesion. The anterior component of the left-sided lesion is demarcated by the fissure.

Voriconazole is currently the preferred oral agent for chronic aspergillosis with itraconazole or posaconazole as substitutes when failure, emergence of resistance, or adverse events occur. Because chronic cavitary pulmonary aspergillosis responds slowly, therapy for at least 12 months is necessary, and disease control may require years of treatment, whereas the duration of treatment for other forms of chronic and allergic aspergillosis requires case-by-case evaluation. Glucocorticoids should be avoided in chronic cavitary pulmonary aspergillosis unless covered by adequate antifungal therapy; mortality is increased by 240% with corticosteroids. Acute exacerbations of ABPA respond well to voriconazole, itraconazole, or a short course of glucocorticoids—long-term azole therapy usually helps minimize corticosteroid exposure and maintain remission. Antifungal response in *Aspergillus* bronchitis is gratifying, but relapse after 4 months of therapy is common.

Resistance in *A. fumigatus* to one or more azoles, although uncommon, is increasingly found globally. Resistance may be derived from azole fungicide use for crops. Resistance arising from multiple mechanisms may develop during long-term treatment, and a positive culture during antifungal therapy is an indication for susceptibility testing.

Surgical treatment is important in several forms of aspergillosis, including fungal ball of the sinus and single aspergillomas, in which

TABLE 223-3 Treatment of Aspergillosis[a]

INDICATION	PRIMARY TREATMENT	PRECAUTIONS	SECONDARY TREATMENT	COMMENTS
Invasive disease[b]	Voriconazole, isavuconazole, posaconazole	Drug interactions (especially with voriconazole)[c]	AmB, caspofungin, micafungin	As primary therapy, azoles have a 20% higher response rate than AmB. Therapeutic drug monitoring is recommended for voriconazole and isavuconazole in ICU patients.
Prophylaxis	Posaconazole tablet, itraconazole solution SUBA-itraconazole	Vincristine, cyclophosphamide interaction	Micafungin, aerosolized AmB	Some centers monitor plasma levels of itraconazole and posaconazole.
Single aspergilloma	Surgical resection	Multicavity disease: poor outcome of surgery, medical therapy preferable	Itraconazole, voriconazole, intracavity AmB	Single large cavities with an aspergilloma are best resected. Relapse reduced by pre- and perioperative antifungal therapy.
Chronic pulmonary disease[b]	Voriconazole, itraconazole	Poor absorption of itraconazole capsules with proton pump inhibitors or H_2 blockers	Posaconazole, IV AmB, IV micafungin	Resistance may emerge during treatment, especially if plasma drug levels are low. Therapeutic azole drug monitoring recommended.
ABPA/SAFS ("fungal asthma")	Itraconazole	Some glucocorticoid interactions, including with inhaled formulations	Voriconazole, posaconazole, inhaled AmB	Long-term therapy is helpful in most cases. No evidence indicates whether therapy modifies progression to bronchiectasis/fibrosis.

[a]For information on duration of therapy and drug resistance in certain *Aspergillus* species, see text. [b]An infectious disease consultation is appropriate for these patients. [c]Online drug-interaction resource: *www.aspergillus.org.uk/content/antifungal-drug-interactions*.

Note: After loading doses, the oral dose is usually 200 mg bid for voriconazole and itraconazole, 65 mg bid for SUBA-itraconazole, 300 mg qd for posaconazole tablets, and 200 mg qd for isavuconazole. The IV dose of voriconazole for adults is 6 mg/kg twice at 12-h intervals (loading doses) followed by 4 mg/kg q12h; a larger dose is required for children and teenagers; a lower dose may be safer for persons >70 years of age. Plasma monitoring is helpful in optimizing the dosage. The IV dose of isavuconazole is 200 mg tid for 2 days (loading dose) followed by 200 mg qd. Caspofungin is given as a single loading dose of 70 mg and then at 50 mg/d; some authorities use 70 mg/d for patients weighing >80 kg, and lower doses are required with hepatic dysfunction. Micafungin is given as 50 mg/d for prophylaxis and as at least 150 mg/d for treatment; this drug has not yet been approved by the U.S. Food and Drug Administration (FDA) for this indication. AmB deoxycholate is given at a daily dose of 1 mg/kg if tolerated. Several strategies are available for minimizing renal dysfunction. Lipid-associated AmB is given at 3 mg/kg (AmBisome) or 5 mg/kg (Abelcet). Different regimens are available for aerosolized AmB, but none is FDA approved. Other considerations that may alter dose selection or route include age; concomitant medications; renal, hepatic, or intestinal dysfunction; and drug tolerability.

Abbreviations: ABPA, allergic bronchopulmonary aspergillosis; AmB, amphotericin B; SAFS, severe asthma with fungal sensitization ICU, intensive care unit.

surgery is curative; invasive aspergillosis involving bone, heart valve, sinuses, and proximal areas of the lung (to avoid catastrophic hemoptysis); brain abscess; keratitis; and endophthalmitis. In allergic fungal sinusitis, removal of abnormal mucus and polyps, with local and occasionally systemic glucocorticoid treatment, usually leads to resolution. Persistent or recurrent signs and symptoms may require more extensive surgery (ethmoidectomy) and possibly antifungal therapy. Surgery is problematic in chronic cavitary pulmonary aspergillosis, often resulting in serious complications. Bronchial artery embolization is preferred for problematic hemoptysis.

PROPHYLAXIS

In situations in which moderate or high risk is predicted (e.g., after induction therapy for acute myeloid leukemia), the need for antifungal prophylaxis for superficial and systemic candidiasis and for invasive aspergillosis is generally accepted. Fluconazole is commonly used in these situations but has no activity against *Aspergillus* species. Itraconazole solution or super bioavailability (SUBA)-itraconazole capsules provide enough bioavailability for modest efficacy, the latter with fewer adverse events. Posaconazole tablets are more effective in reducing infection rates and the need for empirical antifungal therapy. Some data support the use of IV micafungin in those with azole contraindications. No prophylactic regimen is completely successful.

OUTCOME

Invasive aspergillosis is curable if immune reconstitution occurs, whereas allergic and chronic forms are not. The mortality rate for invasive aspergillosis is 30–70% if the infection is treated but is 100% if the diagnosis is missed. Cerebral aspergillosis, *Aspergillus* endocarditis, and bilateral extensive invasive pulmonary aspergillosis have very poor outcomes, as does invasive infection in persons with late-stage AIDS, relapsed uncontrolled leukemia or liver failure.

For those with chronic cavitary pulmonary aspergillosis, ~70% deteriorate clinically without antifngal therapy. The mortality rate for chronic cavitary pulmonary aspergillosis is 15–20% in the first year, ~40% over 5 years, and 50–60% over 10 years if the patient is actively treated with antifungal agents. Antifungal therapy fails in ~30% of recipients and still more often if azole resistance is present.

Both ABPA and SAFS patients respond to glucocorticoids and antifungal therapy; ~60% respond to itraconazole and ~80% to voriconazole and posaconazole (if tolerated). Inhaled amB is effective in and tolerated by ~15% of patients. Long-term antifungal therapy often allows oral glucocorticoids to be stopped and inhaled glucocorticoid dose. Relapse after discontinuation is common.

FURTHER READING

Denning DW, Pfavayi LT: Poorly controlled asthma: Easy wins and future prospects for addressing fungal allergy. Allergol Int 72:493, 2023.

Kang N et al: Clinical characteristics and treatment outcomes of pathologically confirmed *Aspergillus* nodules. J Clin Med 9:2185, 2020.

Li Z, Denning DW: The impact of corticosteroids on the outcome of fungal disease: A systematic review and meta-analysis. Curr Fungal Infect Rep 17:54, 2023.

Sehgal IS et al: Efficacy of 12-months oral itraconazole versus 6-months oral itraconazole to prevent relapses of chronic pulmonary aspergillosis: An open-label, randomised controlled trial in India. Lancet Infect Dis 22:1052, 2022.

Ullman AJ et al: Diagnosis and management of *Aspergillus* diseases: Executive summary of the 2017 ESCMID-ECMM-ERS guideline. Clin Microbiol Infect 24:e1ee38, 2018.

224

Mucormycosis

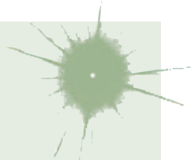

Brad Spellberg, Ashraf S. Ibrahim

Mucormycosis represents a group of life-threatening infections caused by fungi of the order Mucorales of the subphylum Mucoromycotina. Mucormycosis is highly invasive and relentlessly progressive, resulting in higher rates of morbidity and mortality than many other infections. The mortality rates from mucormycosis have declined in recent years as a result of early initiation of more effective antifungal therapies. However, mortality remains high overall, often driven by progression of the underlying predisposing condition.

TABLE 224-1 Taxonomy of Fungi Causing Mucormycosis (Subphylum Mucoromycotina, Order Mucorales)

FAMILY	GENUS (SPECIES LISTED FOR SOME)
Mucoraceae	*Rhizopus oryzae*
	Rhizopus delemar
	Rhizopus microsporus
	Rhizomucor
	Mucor
	Actinomucor
Lichtheimiaceae	*Lichtheimia* (formerly *Mycocladus*, formerly *Absidia*)
Cunninghamellaceae	*Cunninghamella*
Thamnidiaceae	*Cokeromyces*
Mortierellaceae	*Mortierella*
Saksenaceae	*Saksenaea*
	Apophysomyces
Syncephalastraceae	*Syncephalastrum*

ETIOLOGY

The fungal order Mucorales consists of seven families that are known to cause mucormycosis (Table 224-1). *Rhizopus oryzae* and *R. delemar* (both in the family Mucoraceae) are by far the most common causes of mucormycosis in the Western Hemisphere. Less frequently isolated species of the Mucoraceae that cause a similar spectrum of infections include *R. microsporus*, *Rhizomucor pusillus*, *Lichtheimia corymbifera* (formerly *Absidia corymbifera*), *Apophysomyces elegans*, and *Mucor* species. Increasing numbers of cases of mucormycosis due to infection by mold in the family Cunninghamellaceae have also been reported, particularly in highly immunocompromised patients. Other Mucorales can be the major cause of disease in certain geographic areas (e.g., *A. elegans* in India and *M. irregularis* in China), or in outbreaks following natural disasters (e.g., *A. trapeziformis* outbreak following Joplin tornado). Only rare case reports have demonstrated the ability of fungi in the remaining families of the Mucorales to cause mucormycosis.

PATHOGENESIS

The Mucorales are ubiquitous environmental fungi to which humans are constantly exposed. These fungi cause infection primarily in patients with uncontrolled diabetes, defects in phagocytic function (e.g., neutropenia or exposure to high doses of corticosteroids), and/or elevated levels of free iron, which supports fungal growth in serum and tissues. In the past, iron-overloaded patients with end-stage renal failure who were treated with deferoxamine had a high risk of developing rapidly fatal disseminated mucormycosis; deferoxamine is an iron chelator for the human host, but it serves as a fungal xenosiderophore, directly delivering iron to the Mucorales. Furthermore, patients with diabetic ketoacidosis (DKA) are at high risk of developing rhinocerebral mucormycosis. The acidosis causes dissociation of iron from sequestering proteins, resulting in enhanced fungal survival and virulence. The ketoacid β-hydroxybutyrate also increases expression of host and fungal receptors that result in fungal adherence and penetration into tissues.

Nevertheless, the majority of diabetic patients who present with mucormycosis are not acidotic, and, even absent acidosis, hyperglycemia directly contributes to the risk of mucormycosis by at least four likely mechanisms: (1) hyperglycation of iron-sequestering proteins, disrupting normal iron sequestration; (2) upregulation of a mammalian cell receptor (GRP78) that binds to Mucorales, enabling tissue penetration (due to both a direct effect of hyperglycemia and increasing levels of free iron); (3) induction of poorly characterized defects in phagocytic function; and (4) enhanced expression of CotH, a Mucorales-specific protein that mediates host cell invasion by binding to GRP78 (due to hyperglycemia and the resulting free iron).

EPIDEMIOLOGY

Mucormycosis typically occurs in patients with diabetes mellitus, solid-organ or hematopoietic stem cell transplantation (HSCT), prolonged neutropenia or corticosteroid use, or malignancy. During the pandemic, mucormycosis occurred commonly as a co-infection in patients with COVID-19 severe enough to warrant hospitalization. Such infections were heavily described in low- and middle-income countries (LMICs) and in India in particular. However, the incidence of mucormycosis was higher in India than most other countries even before the pandemic. Hence, it is unclear if the virus itself somehow predisposed to fungal superinfection or if the routine use of high-dose corticosteroid therapy for COVID-19, possibly combined with high rates of diabetes mellitus in affected populations, was the primary driver of such infections.

As mentioned, the majority of diabetic patients are not acidotic on presentation with mucormycosis. Furthermore, patients often have no previously recognized history of diabetes mellitus when they present with mucormycosis. In these instances, presentation with mucormycosis may result in the first clinical recognition of hyperglycemia, which often has been unmasked by recent glucocorticoid use. Thus, a high index of suspicion of mucormycosis must be maintained, even in the absence of a known history of diabetes, if hyperglycemia is present and patients present with sinusitis, particularly with possible orbital extension.

In patients undergoing HSCT, mucormycosis develops at least as commonly during nonneutropenic as during neutropenic periods, probably because of glucocorticoid treatment of graft-versus-host disease. Mucormycosis can occur as isolated cutaneous or subcutaneous infection in immunologically normal individuals after traumatic implantation of soil or vegetation (e.g., due to natural disasters, motor vehicle accidents, or severe injuries in war zones) or in nosocomial settings via direct access through intravenous catheters, subcutaneous injections, or maceration of the skin by a moist dressing.

Patients receiving antifungal prophylaxis with either itraconazole or voriconazole may be at increased risk of mucormycosis. These patients typically present with disseminated mucormycosis, the most lethal form of disease. Breakthrough mucormycosis also has been described in patients receiving posaconazole, isavuconazole, or echinocandin prophylaxis.

CLINICAL MANIFESTATIONS

Mucormycosis presentations are often as one of five well-defined clinical syndromes: rhino-orbital-cerebral, pulmonary, cutaneous, gastrointestinal, and disseminated disease. However, infection of any body site can occur. Patients with specific defects in host defense tend to develop specific syndromes. For example, patients with diabetes mellitus and/or DKA typically develop the rhino-orbital-cerebral form and much more rarely develop pulmonary or disseminated disease. In contrast, pulmonary mucormycosis occurs most commonly in leukemic patients who are receiving chemotherapy and in patients undergoing HSCT.

Rhino-Orbital-Cerebral Disease Rhino-orbital-cerebral mucormycosis continues to be the most common form of the disease worldwide. Most cases occur in patients with diabetes, although such cases are also described in the transplantation setting, often along with glucocorticoid-induced diabetes mellitus. The initial symptoms of rhino-orbital-cerebral mucormycosis are nonspecific and include eye or facial pain and facial numbness followed by the onset of conjunctival suffusion and swelling, and blurry vision. In contrast to the acute, bright red, periocular skin manifestations typical of acute bacterial orbital cellulitis, periorbital skin in patients with rhino-orbital-cerebral mucormycosis may take on a more dusky, subacute appearance. Fever may be absent in up to half of cases. White blood cell counts are typically elevated as long as the patient has functioning bone marrow. If untreated, infection usually spreads from the ethmoid sinus to the orbit, resulting in compromise of extraocular muscle function and proptosis, typically with chemosis. From the orbit, the fungus can spread contiguously or hematogenously to the frontal lobe of the brain and/or via venous drainage to the cavernous sinus. Onset of signs and symptoms in the contralateral eye, with resulting bilateral proptosis, chemosis, vision loss, and ophthalmoplegia, is ominous, suggesting the development of cavernous sinus thrombosis.

Upon visual inspection, infected tissue often has a normal appearance during the earliest stages of fungal spread, which can make diagnosis difficult; blind biopsies of normal-appearing sinus tissue are warranted when suspicion for mucormycosis is high. Tissue then progresses through an erythematous phase, with or without edema, before the onset of a violaceous appearance and finally the development of a black necrotic eschar. Infection can sometimes extend from the sinuses into the mouth and produce painful necrotic ulcerations of the hard palate, but this is a late finding that suggests extensive, well-established infection.

One common misperception about mucormycosis is that it is always rapidly progressive. In fact, the rate of progression is extremely variable and is possibly dependent on the immune status of the patient, the infectious inoculum, and the causative Mucorales species, some of which are more virulent and/or have faster growth rates than others. Patients may go from initial symptoms to death in days; alternatively, it can take months or even a year or more for lethal progression to occur.

Pulmonary Disease Pulmonary mucormycosis is the second most common manifestation. Symptoms include dyspnea, cough, and chest pain; fever is often but not invariably present. Angioinvasion results in necrosis, cavitation, and/or hemoptysis. Lobar consolidation, isolated masses, nodular disease, cavities, or wedge-shaped infarcts may be seen on chest radiography. Chest computed tomography (CT) is the best method for determining the extent of pulmonary mucormycosis and may demonstrate evidence of infection before it is seen on chest x-ray. In the setting of cancer, where mucormycosis may be difficult to differentiate from aspergillosis, the presence of ≥10 pulmonary nodules, pleural effusion, or concomitant sinusitis make mucormycosis more likely. It is important to distinguish mucormycosis from aspergillosis because treatments for these infections may differ. Indeed, voriconazole—the first-line treatment for aspergillosis—exacerbates mucormycosis in mouse and fly models of infection. Isavuconazole was noninferior to voriconazole for the treatment of aspergillosis in a randomized controlled trial and also has activity against Mucorales. Hence if there is doubt about whether infection is caused by a septated mold (e.g., *Aspergillus*) or a Mucorales, inclusion of isavuconazole in a treatment regimen is a reasonable consideration. Consideration must also be given to the possibility of dual infection with both a septated mold and Mucorales; dual infection is not infrequently encountered in highly compromised patients.

Cutaneous Disease Cutaneous mucormycosis may result from external implantation of the fungus or from hematogenous dissemination. External implantation–related infection has been described in the setting of soil exposure from trauma (e.g., in a motor vehicle accident, a natural disaster, or combat-related injuries), penetrating injury with plant material (e.g., a thorn), injections of medications (e.g., insulin), catheter insertion, contamination of surgical dressings, and use of tape to secure endotracheal tubes. Cutaneous disease can be highly invasive, penetrating into muscle, fascia, and even bone. Necrotizing fasciitis caused by mucormycosis carries a mortality rate approaching 80%. Necrotic cutaneous lesions in the setting of hematogenous dissemination also are associated with an extremely high mortality rate. However, with prompt, aggressive surgical debridement, isolated, nonnecrotizing cutaneous mucormycosis has a favorable prognosis and a low mortality rate.

Gastrointestinal Disease In the past, gastrointestinal mucormycosis occurred primarily in premature neonates in association with disseminated disease and necrotizing enterocolitis. However, there has been a marked increase in case reports describing adults with neutropenia, glucocorticoid use, or other immunocompromising conditions. In addition, gastrointestinal disease has been reported as a nosocomial process following administration of medications mixed with contaminated wooden applicator sticks. Nonspecific abdominal pain and distention associated with nausea and vomiting are the most common symptoms. Gastrointestinal bleeding is common, and fungating masses may be seen in the stomach at endoscopy. The disease may progress to visceral perforation, with extremely high mortality rates.

Disseminated and Miscellaneous Forms of Disease Hematogenously disseminated mucormycosis may originate from any primary site of infection. The most common site of dissemination is the brain, but metastatic lesions may also be found in any other organ. Mortality rates for widely disseminated mucormycosis exceed 90%; however, these high rates are likely to be due in part to the underlying predisposing condition leading to the infection and the inability to surgically remove the infected foci.

Mucormycosis may affect any body site, including bones, mediastinum, trachea, kidneys, peritoneum (in association with dialysis), scalp (causing a kerion), and even isolated infection of teeth.

DIAGNOSIS

A high index of suspicion is required for diagnosis of mucormycosis. Unfortunately, autopsy series have shown that up to half of cases are diagnosed only postmortem. Because the Mucorales are environmental isolates, definitive diagnosis requires a positive culture from a sterile site (e.g., a needle aspirate, a tissue biopsy specimen, or pleural fluid) or histopathologic evidence of invasive mucormycosis. A probable diagnosis of mucormycosis can be established by culture from a nonsterile site (e.g., sputum or bronchoalveolar lavage) or the detection of Mucorales on the surface of histopathology samples (without visualization of evidence of invasion) when a patient has appropriate risk factors as well as clinical and radiographic evidence of disease. In such cases, given the urgency of administering therapy early, the patient should be treated while confirmation of the diagnosis is awaited.

Biopsy with histopathologic examination remains the most sensitive and specific modality for definitive diagnosis **(Fig. 224-1)**. Biopsy reveals characteristic wide (≥6- to 30-μm), thick-walled, ribbon-like, aseptate hyphal elements that branch at right angles. Other fungi, including *Aspergillus, Fusarium*, and *Scedosporium* species, have septa, are thinner, and branch at acute angles. However, artificial septa may result from folding of tissue during processing (which may also alter the appearance of the angle of branching), which can make Mucorales appear to have septa. Thus, the width and the ribbon-like form of the fungus are likely the most reliable features distinguishing mucormycosis from other pathogenic molds. The Mucorales are visualized most effectively with periodic acid–Schiff or hematoxylin and eosin; in contrast to many other fungi, methenamine silver may not result in optimal staining. While histopathology can identify the Mucorales, species can be identified only by culture. Several studies showed that polymerase chain reaction (PCR) of Mucorales-specific targets is useful in diagnosing mucormycosis. However, the U.S. Food and Drug Administration (FDA) has not approved any of these PCR-based assays for this purpose.

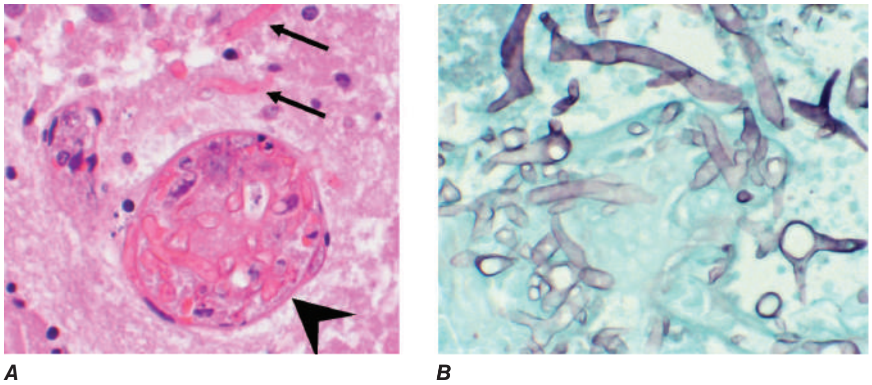

FIGURE 224-1 Histopathology sections of *Rhizopus delemar* in infected brain. *A.* Broad, ribbon-like, nonseptate hyphae in the parenchyma (*arrows*) and a thrombosed blood vessel with extensive intravascular hyphae (*arrowhead*) (hematoxylin and eosin). ***B.*** Extensive, broad, ribbon-like hyphae invading the parenchyma (Gomori methenamine silver).

Unfortunately, cultures are positive in fewer than half of cases of mucormycosis. Nevertheless, the Mucorales are not fastidious organisms and tend to grow quickly (i.e., within 48–96 h) on culture media. The likely explanation for the low sensitivity of culture is that the Mucorales form long filamentous structures that are killed by tissue homogenization—the standard method for preparing tissue cultures in the clinical microbiology laboratory. Therefore, the laboratory should be advised when a diagnosis of mucormycosis is suspected, and the tissue should be cut into sections and placed in the center of culture dishes rather than homogenized. Because there is also substantial variability among isolates in optimal growth temperature, growth at both room temperature and 37°C is advisable.

Imaging techniques often yield subtle findings that underestimate the extent of disease. For example, the most common finding on CT or magnetic resonance imaging (MRI) of the head or sinuses of a patient with rhino-orbital mucormycosis is sinusitis that is indistinguishable from bacterial sinusitis. While sinusitis is almost always seen on CT scans in patients with the rhino-orbital-cerebral disease, erosion through the sinus bones and into the orbit is rarely seen on CT even when it is clinically present. MRI is more sensitive (~80%) for detecting orbital and central nervous system (CNS) disease than is CT. High-risk patients should always undergo endoscopy and/or surgical exploration, with biopsy of the areas of suspected infection. If mucormycosis is suspected, initial empirical therapy with a polyene antifungal agent should be initiated while the diagnosis is being confirmed.

DIFFERENTIAL DIAGNOSIS

Other mold infections, including aspergillosis, scedosporiosis, fusariosis, and infections caused by the dematiaceous fungi (brown-pigmented soil organisms), can cause clinical syndromes identical to mucormycosis. Histopathologic examination usually allows distinction of the Mucorales from these other organisms, and a positive culture permits definitive species identification. As stated above, it is important to distinguish the Mucorales from these other fungi, as the preferred antifungal treatments differ (i.e., polyenes for the Mucorales vs expanded-spectrum triazoles for most septate molds). The entomophthoromycoses caused by *Basidiobolus* and *Conidiobolus* also can cause identical clinical syndromes. These fungi cannot be readily distinguished from the Mucorales by histopathology but can be reliably distinguished by culture. Fortunately, entomophthoromycoses are uncommon in developed countries and can be treated with polyenes; it is not urgent to distinguish them from mucormycosis.

In a patient with sinusitis and proptosis, orbital cellulitis and cavernous sinus thrombosis caused by bacterial pathogens (most commonly *Staphylococcus aureus*, but also streptococcal and gram-negative species) must be excluded. *Klebsiella rhinoscleromatis* is a rare cause of an indolent facial rhinoscleroma syndrome that may appear similar to mucormycosis. Finally, the Tolosa-Hunt syndrome is characterized by painful ophthalmoplegia, ptosis, headache, and cavernous sinus inflammation; biopsies and clinical follow-up may be needed to distinguish the Tolosa-Hunt syndrome from mucormycosis; there is a lack of progression of the former entity.

TREATMENT

Mucormycosis

GENERAL PRINCIPLES

Optimizing the chances for successful treatment of mucormycosis requires four steps: (1) early initiation of therapy; (2) surgical debridement, when possible; (3) rapid reversal of underlying predisposing risk factors, if possible; and (4) proceeding to treat underlying malignancy, if present, without waiting to complete antifungal therapy first.

Early initiation of antifungal therapy requires maintaining a high index of suspicion for at-risk patients. Multiple studies have found that earlier initiation of polyene-based therapy improves survival of patients with mucormycosis. Because the disease can present subtly at first and because confirmation of the diagnosis can take days, therapy often must be started empirically before the diagnosis is established. When there is a reasonable suspicion of mucormycosis, clinicians should not hesitate to initiate therapy with a lipid polyene as soon as possible, since the toxicity of lipid polyenes (unlike that of amphotericin B [AmB] deoxycholate) is rarely substantial after one or two doses.

Blood vessel thrombosis and resulting tissue necrosis during mucormycosis can result in poor penetration of antifungal agents to the site of infection. Therefore, debridement of all necrotic tissues can help eradicate the disease. Surgery has been found (by logistic regression and in multiple case series) to be an independent variable for favorable outcome in patients with mucormycosis. However, these data are confounded by the fact that sicker patients are often unable to tolerate surgical procedures. Thus, a moderated approach in which tissue is debrided when and to the extent it is safe to do so is advisable. Limited data from a retrospective study support the use of intraoperative frozen sections to delineate the margins of infected tissues, with sparing of tissues lacking evidence of infection.

Rapidly reversing hyperglycemia, acidosis, or iron overload, and lowering corticosteroid doses are important to achieving cure. Indeed, a recent study confirmed that resolution of acidosis in mice with DKA via the administration of sodium bicarbonate (used in the mice in lieu of insulin) improved survival. Administration of glucocorticoids predisposes animals to death from mucormycosis in experimental models. Similarly, iron administration to patients with active mucormycosis should be avoided, as iron exacerbates infection in experimental models. Blood transfusion typically results in some liberation of free iron due to hemolysis, so a conservative approach to red blood cell transfusions is advisable.

One of the most common errors made in management of mucormycosis is the belief that mucormycosis must be eradicated before an underlying malignancy can be treated. This belief can result in halting or delaying treatment for the underlying disease (e.g., chemotherapy or transplantation) until the mucormycosis is cured. Three fallacies belie this concern. First, mucormycosis will not be definitively eradicated until near-normal immunity is restored; the antifungals provide a holding action and are unlikely to be curative until the underlying disease is treated. Second, modern antifungals can halt progression of mucormycosis temporarily, enabling aggressive chemotherapy or transplantation to be administered to cure the underlying disease. Finally, the primary driver of death in such patients is typically progression of the underlying disease due to failure to treat it appropriately.

Initially, some consideration can be given to moderating the level of aggressiveness of the chemotherapy and resulting duration and depth of neutropenia. The aggressiveness of immune suppression and antifungal therapy can then be adjusted during the course of treatment in response to changes in clinical status. Nevertheless, chemotherapy should be given sufficiently aggressively to attempt cure of the underlying disease. These patients are extremely complex, and multidisciplinary, team-based care is advisable.

ANTIFUNGAL THERAPY

Primary therapy for mucormycosis should be based on a polyene antifungal agent **(Table 224-2)**, except perhaps in mild localized infection (e.g., isolated suprafascial cutaneous infection) that has been mostly excised surgically in an immunocompetent patient. Lipid formulations of AmB are significantly less nephrotoxic than AmB deoxycholate, can be administered at higher doses, and are probably more effective for this purpose. Liposomal amphotericin B (LAmB) may be preferred to amphotericin B lipid complex (ABLC) for management of brain infection based on retrospective survival data and superior brain penetration; there is no clear efficacy advantage of either agent for infections outside the brain, although LAmB may be less nephrotoxic than ABLC (but is also considerably more expensive in many settings).

Starting dosages of 1 mg/kg per day for AmB deoxycholate and 5 mg/kg per day for LAmB and ABLC are commonly given to adults

TABLE 224-2 Antifungal Options for the Treatment of Mucormycosis[a]

DRUG	RECOMMENDED DOSAGE	ADVANTAGES AND SUPPORTING STUDIES	DISADVANTAGES
First-Line Antifungal Therapy			
AmB deoxycholate	1.0–1.5 mg/kg once per day	• >5 decades of clinical experience • Inexpensive • FDA approved for treatment of mucormycosis	• Highly toxic • Poor CNS penetration
LAmB	5–10 mg/kg once per day	• Less nephrotoxic than AmB deoxycholate • Better CNS penetration than AmB deoxycholate or ABLC • Better outcomes than with AmB deoxycholate in murine models and a retrospective clinical review	• Expensive
ABLC	5 mg/kg once per day	• Less nephrotoxic than AmB deoxycholate • Murine and retrospective clinical data suggest benefit of combination therapy with echinocandins	• Expensive • Possibly less efficacious than LAmB for CNS infection
Second-Line/Salvage Option			
Isavuconazole	200 mg of isavuconazole (372 mg of isavuconazonium sulfate), load q8h × 6 followed by once-daily dosing	• Efficacy similar to that of LAmB in mouse models • FDA approved for treatment of mucormycosis • May be a rational empirical option when septate mold vs mucormycosis is not yet established	• Much less clinical experience • Clinical study supporting approval was small and historically controlled
Posaconazole	200 mg four times per day	• In vitro activity against the Mucorales, with lower MICs than isavuconazole • Retrospective data for salvage therapy in mucormycosis	• Substantially lower blood levels than isavuconazole • No data on initial therapy for mucormycosis, and no evidence for posaconazole in combination with other therapy • Experience limited, potential use for salvage therapy
Combination Therapy[b]			
Echinocandin plus lipid polyene	Standard echinocandin doses	• Favorable toxicity profile • Synergistic in murine disseminated mucormycosis • Retrospective clinical data suggest superior outcomes for rhino-orbital-cerebral mucormycosis.	• Limited clinical data on combination therapy
Lipid polyene plus azole (posaconazole or isavuconazole)	Standard doses	• Favorable toxicity profile	• Limited efficacy data, with no available evidence of superiority vs monotherapy
Triple therapy (lipid polyene plus echinocandin plus azole)	Standard doses	• Maximal aggressiveness	• Limited efficacy data, with no available evidence for superiority vs monotherapy or dual therapy

[a]Primary therapy should generally include a polyene. Non-polyene-based regimens may be appropriate for patients who refuse polyene therapy or for relatively immunocompetent patients with mild disease (e.g., isolated suprafascial cutaneous infection) that can be surgically eradicated. [b]Prospective randomized trials are necessary to confirm the suggested benefit (from animal and small retrospective human studies) of combination therapy for mucormycosis. Dose escalation of any echinocandin is not recommended because of a paradoxical loss of benefit of combination therapy at echinocandin doses of ≥3 mg/kg per day.

Abbreviations: ABLC, AmB lipid complex; AmB, amphotericin B; CNS, central nervous system; FDA, U.S. Food and Drug Administration; LAmB, liposomal AmB; MIC, minimal inhibitory concentration.

Source: Reproduced with permission from B Spellberg et al: Recent advances in the management of mucormycosis: From bench to bedside. Clin Infect Dis 48:1743, 2008.

and children to treat mucormycosis. Dose escalation of LAmB to 7.5 or 10 mg/kg per day for CNS mucormycosis may be considered in light of the limited penetration of polyenes into the brain. A recent observational study failed to find superior outcomes of LAmB dosed at 10 mg/kg per day. However, very few patients with brain involvement were included, and the analysis was not stratified by site of infection. Therefore, 10 mg/kg per day may be reasonable to consider for treating aggressive brain infection but is unlikely of benefit, and will increase toxicity, for other patients. Because of autoinduction of metabolism, which results in paradoxically lower drug levels, there is no advantage to escalating the LAmB dose above 10 mg/kg per day, and doses of 5 mg/kg per day are probably adequate for nonbrain infections. ABLC dose escalation above 5 mg/kg per day is not advisable given the lack of relevant data and the drug's potential toxicity.

In multiple studies, various combinations of lipid polyenes (both ABLC and LAmB) plus echinocandins (e.g., caspofungin, micafungin, and anidulafungin) improved survival rates among mice with disseminated mucormycosis (including CNS disease). Furthermore, combination lipid polyene–echinocandin therapy was associated with significantly better outcomes than polyene monotherapy in a retrospective clinical study involving patients with rhino-orbital-cerebral mucormycosis (including CNS disease). The effect of echinocandins appears to be to downmodulate the virulence of the fungus and reduce tissue necrosis and destruction from fungal invasion. On the basis of such data, some experts prefer combination lipid polyene–echinocandin therapy as a first-line option. However, at least one retrospective study did not find an advantage of any combination regimens (including polyene-azole, polyene-echinocandin, or others) in patients who primarily had malignancy as the underlying disease. Ultimately, definitive randomized controlled trials are needed to establish whether the combination is superior in efficacy to monotherapy for mucormycosis. When used, echinocandins should be administered at standard, FDA-approved doses, since dose escalation has resulted in paradoxical loss of efficacy in preclinical models.

In contrast to deferoxamine, the iron chelator deferasirox is fungicidal against clinical isolates of the Mucorales. In mice with DKA and disseminated mucormycosis, combination deferasirox–LAmB therapy resulted in synergistic improvement of survival rates and reduced the fungal burden in the brain. Unfortunately, a small, randomized, double-blind, phase 2 safety clinical trial of adjunctive therapy with deferasirox (plus LAmB) documented excess mortality among patients treated with deferasirox. Of note, the study population included primarily patients with active malignancy, and few patients in the study had diabetes mellitus as their only risk factor. Deferasirox is therefore contraindicated as therapy in patients with active malignancy, but its role in patients who have diabetes mellitus without malignancy (the setting in which its preclinical efficacy was optimal) remains uncertain.

Posaconazole and isavuconazole are the only FDA-approved azoles with reliable in vitro activity against the Mucorales. However, there are limited data regarding the efficacy of posaconazole monotherapy for mucormycosis, and in contrast to polyene-echinocandin therapy, no data support the use of combination posaconazole-polyene regimens. Although the minimal inhibitory concentrations (MICs) of isavuconazole against the Mucorales are four- to eightfold higher than those of posaconazole, blood levels may be higher with standard isavuconazole dosing than with posaconazole. Isavuconazole is FDA-approved for the treatment of mucormycosis based on a small, historically controlled study. Given this limited dataset, many experts continue to think that lipid polyenes are first-line options and that isavuconazole, like posaconazole, is best reserved for oral transitional therapy in patients whose condition has substantially improved on polyene-based therapy, or for salvage therapy in patients who are intolerant of polyene-based regimens or whose infection is refractory to these regimens.

No studies in mice or humans have ever shown benefit of adding posaconazole to another antifungal agent for combination therapy of mucormycosis. In contrast, in mice, a combination of isavuconazole plus polyene regimens did prolong survival time and lowered tissue fungal burden compared to monotherapy. However, this combination has not been studied compared to monotherapy in people. Some experts use triple therapy with a polyene, echinocandin, and either posaconazole or isavuconazole for patients who have extensive disease or whose disease has progressed on prior therapy. Empirical, dual lipid polyene–azole therapy is a rational choice in a patient with likely invasive mold infections when septate molds and mucormycosis are both in the differential diagnosis and the etiologic agent has not yet been confirmed. Alternatively, initial therapy with isavuconazole monotherapy may be reasonable for a brief period of time in a stable patient if mucormycosis is felt to be possible but less likely than a septated mold infection.

The roles of recombinant cytokines and neutrophil transfusions in the primary treatment of mucormycosis are not clear, although it is intuitive that earlier recovery of neutrophil counts should improve survival rates. Limited data from uncontrolled case series have described the use of hyperbaric oxygen in centers with the appropriate technical expertise and facilities; its efficacy remains undefined. As mentioned previously, one study in mice with DKA found that administration of sodium bicarbonate improved survival from mucormycosis; however, because insulin was not administered to the mice, it is unclear whether the therapeutic effect is clinically relevant.

In general, antifungal therapy for mucormycosis should be continued until resolution of clinical signs and symptoms of infection and resolution of underlying immunosuppression. However, after a week or two of daily therapy in a patient who is clinically improving, it is reasonable to consider switching to thrice-weekly lipid polyene doses—with ultimate weaning down to twice-weekly doses—for maintenance therapy. For patients with mucormycosis who are receiving immunosuppressive medications, secondary antifungal prophylaxis is typically continued for as long as the immunosuppressive regimen is administered. Transitioning to azoles for chronic suppression is a reasonable alternative to continuing polyene therapy in this setting, with re-initiation of polyenes during periods of deep neutropenia.

One common source of error in the long-term management of mucormycosis is follow-up radiology. Analysis of data from the DEFEAT Mucor study indicated that early radiographic progression (within the first 2 weeks) did not predict long-term mortality. Changing the therapeutic plan based on early radiographic changes can result in therapeutic errors. For example, it is common for CNS Mucorales to cavitate in the brain parenchyma over time. This does not necessarily reflect therapeutic failure, but rather may reflect increased immune reactivity to the fungus, particularly in patients recovering from neutropenia or with removal of immune suppression. Thus, it may not be advisable to obtain serial radiographic studies in the short-term, and if such studies are obtained, caution should be used in reacting to their results. Greater emphasis should be placed on clinical response, particularly within the first 2–4 weeks after initiation of therapy.

PROGNOSIS

Over the past two decades, the prognosis of mucormycosis has substantially improved with aggressive antifungal therapy. Even CNS infection is often successfully treated. As mentioned, the key driver of outcome may be control of the patient's predisposing condition.

FURTHER READING

Alqarihi A et al: Mucormycosis in 2023: An update on pathogenesis and management. Front Cell Infect Microbiol 13:1254919, 2023.

Baldin C, Ibrahim AS: Molecular mechanisms of mucormycosis: The bitter and the sweet. PLoS Pathog 13:e1006408, 2017.

Cornely O et al: Global guideline for the diagnosis and management of mucormycosis: An initiative of the ECMM in cooperation with ESCMID/EFISG. Lancet Infect Dis 19:e405, 2019.

Danion F et al: What is new in pulmonary mucormycosis? J Fungi (Basel) 9:307, 2023.

Pettrikos G et al: Epidemiology of mucormycosis in Europe. Clin Microbiol Infect 20:67, 2014.

225

Superficial Fungal Infections

Joseph Pechacek, Carol A. Kauffman, Michail S. Lionakis

Fungal infections of the skin and skin structures are caused by molds and yeasts that do not invade deeper tissues but rather cause disease merely by inhabiting the superficial layers of skin, hair follicles, and nails. These agents cause the most common human fungal infections but only rarely cause serious infections.

MALASSEZIA INFECTIONS

Etiologic Agents, Epidemiology, and Pathogenesis *Malassezia* species, primarily *M. furfur* and *M. pachydermatis*, are lipophilic yeasts that generally cause only minor skin infections but, on occasion, can cause invasive infection. *Malassezia* species are part of the indigenous human microbiota found in the stratum corneum of the back, chest, scalp, and face—areas rich in sebaceous glands. The organisms do not invade below the stratum corneum and generally elicit little inflammation. Cutaneous interleukin (IL)-17 signaling controls *Malassezia* species in the skin, but excessive, *Malassezia*-induced IL-17 responses have been associated with exacerbating atopic inflammation.

Clinical Manifestations *Malassezia* species cause tinea versicolor (also called *pityriasis versicolor*), folliculitis, and seborrheic dermatitis. Tinea versicolor presents as flat round scaly patches of hypo- or hyperpigmented skin on the neck, chest, or upper arms. The lesions are usually asymptomatic but can be pruritic. They can be mistaken for vitiligo, but the latter is not scaly. Folliculitis occurs on the back and chest and mimics bacterial folliculitis. Seborrheic dermatitis manifests as erythematous pruritic scaly lesions in the eyebrows, moustache, nasolabial folds, and scalp (dandruff). Seborrheic dermatitis can be severe in patients with AIDS. Fungemia and disseminated infection occur rarely with *Malassezia* species, and they almost always occur in premature neonates receiving parenteral lipid nutrition preparations.

Diagnosis *Malassezia* infections are diagnosed clinically in most cases. If scrapings are collected on a microscope slide on which a drop of potassium hydroxide has been placed, a mixture of budding, bottle-shaped yeasts and short septate hyphae is seen. To culture *Malassezia* from patients with suspected disseminated infection, sterile olive oil must be added to the medium.

Treatment and Prognosis Topical creams and lotions, including selenium sulfide shampoo, ketoconazole shampoo or cream, and terbinafine cream, are effective in treating *Malassezia* infections and are usually given for 2 weeks. Mild topical steroid creams are sometimes used to treat seborrheic dermatitis. For extensive disease, oral itraconazole or fluconazole (200 mg/day) can be used for 5–7 days. The rare cases of fungemia caused by *Malassezia* species are treated with amphotericin B (AmB) or an azole, such as voriconazole, prompt removal of central catheters, and discontinuation of parenteral lipid infusions. *Malassezia* skin infections are benign and self-limited, although recurrences are common. The outcome of systemic infection depends on the host's underlying conditions, but most infected neonates do well.

■ MUCOCUTANEOUS *CANDIDA* INFECTIONS

Candidiasis is discussed in Chap. 222, and this section briefly describes mucocutaneous manifestations of candidiasis.

Etiologic Agents, Epidemiology, and Pathogenesis The most common species responsible for mucocutaneous candidiasis—e.g., oral thrush, esophageal candidiasis, vulvovaginal candidiasis (VVC)—is *Candida albicans*. *Candida* species are commensal organisms residing primarily in the oral mucosa, gastrointestinal tract, and genitourinary tract with the skin being colonized in certain settings, such as after antibiotic use in the intensive care unit or by certain species such as *C. parapsilosis* or *C. auris*. *C. auris* is an emerging multidrug-resistant species with a propensity for long-term colonization of the human skin, health care environments, and medical devices and for causing nosocomial outbreaks. *C. auris* grows avidly in human sweat and colonizes deep skin layers including hair follicles. IL-17–producing lymphoid cells at mucocutaneous barrier sites such as T_H17 cells, γδ T cells, and innate lymphoid cells mediate control against *Candida*. Risk factors for oral thrush include inherited and pharmacologic defects of IL-17 signaling, AIDS, and conditions with excessive mucosal interferon γ responses such as autoimmune polyendocrine syndrome type 1. Antibiotic use, diabetes, and pregnancy are major risk factors for VVC. Skin candidiasis is uncommon and most often seen in inborn errors of immunity that cause chronic mucocutaneous candidiasis, primarily *STAT1* gain-of-function.

Diagnosis Diagnosis of oral thrush is generally clinical with white mucosal plaques that are removed with gentle scraping. Culture reveals the responsible *Candida* species. Biopsy of affected tissue may be needed to differentiate thrush from mimics such as leukoplakia. Biopsy of affected oral or esophageal tissue shows yeasts, pseudohyphae, and hyphae invading into the epithelium; observing only yeast forms superficial to epithelial tissue usually represents solely colonization. VVC is diagnosed with culture and microscopy of vaginal secretions collected via swabs. Skin and nail candidiasis is diagnosed by skin biopsy or nail clippings with culture and microscopy.

Treatment and Prognosis Mild oral thrush may be treated with topical agents such as nystatin suspension or clotrimazole troches, but moderate and severe disease or esophageal involvement should be treated with systemic, primarily azole therapy. VVC may be treated with topical or oral azoles. Newer options for VVC include the triterpenoid antifungal ibrexafungerp, which inhibits β-glucan, and oteseconazole, a long-acting tetrazole that cannot be used in patients of reproductive potential owing to potential fetal ocular toxicity.

■ DERMATOPHYTE INFECTIONS

Etiologic Agents, Epidemiology, and Pathogenesis The molds that cause skin infections in humans include the genera *Trichophyton*, *Microsporum*, and *Epidermophyton*. These organisms, which are not components of the normal skin microbiota, can live within the keratinized structures of the skin—hence the term *dermatophytes*. Dermatophytes occur worldwide, and infections with these organisms are extremely common with an estimated ~1 billion people affected globally. Some organisms cause disease only in humans and can be transmitted by person-to-person contact and by fomites, such as hairbrushes or wet floors, that have been contaminated by infected individuals. Several species cause infections in cats and dogs and can readily be transmitted from these animals to humans, and others are spread from contact with soil. The characteristic ring shape of cutaneous lesions is the result of the organisms' outward growth in a centrifugal pattern in the stratum corneum. Fungal nail invasion usually occurs through the lateral or superficial nail plates and then spreads throughout the nail; when hair shafts are invaded, the organisms can be found either within the shaft or surrounding it. Symptoms are caused by the inflammatory reaction elicited by fungal antigens and not by tissue invasion. Dermatophyte infections are restricted by IL-17 responses and occur more often in males; progesterone can inhibit dermatophyte growth.

Clinical Manifestations Dermatophyte infection of the skin is often called *ringworm*. This term is confusing because worms are not involved. *Tinea*, the Latin word for *worm*, describes the serpentine nature of the skin lesions. Tinea is a less confusing term and can be used with the name of the body part affected—e.g., tinea capitis (head), tinea pedis (feet), tinea corporis (body), tinea cruris (crotch), and tinea unguium (nails, more often termed *onychomycosis*).

Tinea capitis occurs most commonly in 3- to 7-year-old children. Children with tinea capitis usually present with well-demarcated scaly patches in which hair shafts are broken off right above the skin; alopecia can result. Tinea corporis manifests as well-demarcated, annular, pruritic, scaly lesions that undergo central clearing. Usually, one or several small lesions are present; however, in some patients, tinea corporis can involve much of the trunk. The rash should be differentiated from contact dermatitis, eczema, and psoriasis. Tinea cruris is seen almost exclusively in men. The perineal rash is erythematous and pustular, has a discrete scaly border, is without satellite lesions, and is usually pruritic. The rash must be differentiated from intertriginous candidiasis, erythrasma, and psoriasis.

Tinea pedis also is also more common among men. It usually starts in the web spaces of the toes; peeling, maceration, and pruritus are followed by development of a scaly pruritic rash along the lateral and plantar surfaces of the feet. Hyperkeratosis of the soles of the feet often ensues. Tinea pedis has been implicated in lower-extremity cellulitis, as streptococci and staphylococci can gain entrance to the tissues through fissures between the toes. Onychomycosis affects toenails more often than fingernails and is most common among persons who have tinea pedis. The nail becomes thickened and discolored and may crumble; onycholysis almost always occurs. Onychomycosis is more common in older adults and in persons with vascular disease, diabetes mellitus, and nail trauma. Fungal infection must be differentiated from psoriasis, which can mimic onychomycosis but usually has associated skin lesions.

Diagnosis Many dermatophyte infections are diagnosed by their clinical appearance. If the diagnosis is in doubt, scrapings should be taken from the edge of a lesion with a scalpel blade, transferred to a slide to which a drop of potassium hydroxide is added, and examined under a microscope for the presence of hyphae. Cultures are indicated if an outbreak is suspected or the patient does not respond to therapy.

Treatment and Prognosis Dermatophyte infections usually respond to topical therapy with azoles or terbinafine. Lotions or sprays

TABLE 225-1 Suggested Oral Treatment for Extensive Tinea Infections and Onychomycosis

ANTIFUNGAL AGENT	SUGGESTED DOSAGE	COMMENTS
Extensive Tinea Infection		
Terbinafine	250 mg/day for 1–2 weeks	Adverse reactions minimal with short treatment period
Itraconazole[a]	200 mg/day for 1–2 weeks	Adverse reactions minimal with short treatment period except for drug interactions
Onychomycosis		
Terbinafine	250 mg/day for 3 months	Slightly superior to itraconazole; monitor for hepatotoxicity
Itraconazole[a]	200 mg/day for 3 months *or* 200 twice daily for 1 week each month for 3 months	Drug interactions frequent; monitor for hepatotoxicity; rarely causes hypokalemia, hypertension, edema; use with caution in patients with congestive heart failure Other triazoles such as posaconazole may be considered as alternative therapy

[a] Itraconazole capsules require food and gastric acid for absorption, whereas itraconazole solution is taken on an empty stomach. The newer SUBA-itraconazole formulation exhibits improved oral bioavailability and reduced interpatient variability.

are easier than creams to apply to large or hairy areas. Particularly for tinea cruris, the affected area should be kept as dry as possible. When patients have extensive skin lesions, oral itraconazole or terbinafine can hasten resolution **(Table 225-1)**. Terbinafine interacts with fewer drugs than itraconazole and is generally the first-line systemic agent.

Onychomycosis generally does not respond to topical therapy, although efinaconazole topical solution applied to the affected nail for as long as a year has been beneficial in several trials. Itraconazole and terbinafine both accumulate in the nail plate and can treat onychomycosis (Table 225-1). The major decision to be made regarding therapy is whether the extent of nail involvement justifies the use of systemic antifungal agents that may have adverse effects, may interact with other drugs, and are costly. Relapses of tinea cruris and tinea pedis are common and should be treated as early as possible with topical creams to avoid development of more extensive disease. Relapses of onychomycosis follow treatment in 25–30% of cases.

SPOROTRICHOSIS

Etiologic Agent, Epidemiology, and Pathogenesis *Sporothrix schenckii* complex comprises six closely related organisms; *S. schenckii* and *S. brasiliensis* are the species that cause most human infections. *S. globosa* has grown in importance in Asia. *Sporothrix* species are relatively intolerant to heat and are found worldwide in sphagnum moss, decaying vegetation, and soil. Sporotrichosis most commonly affects persons who participate in outdoor activities such as landscaping, gardening, and tree farming. Infected animals—such as dogs, armadillos, and most commonly cats—can transmit *S. schenckii* to humans. A large ongoing outbreak of sporotrichosis in Brazil caused by *S. brasiliensis* has been traced to cats, which are highly susceptible to this infection. Sporotrichosis is primarily a localized infection of skin and subcutaneous tissues that follows traumatic conidial inoculation. Osteoarticular sporotrichosis is uncommon, occurring most often in middle-aged men who abuse alcohol, and pulmonary sporotrichosis occurs almost exclusively in persons with chronic obstructive pulmonary disease following fungal inhalation. Dissemination occurs almost entirely in markedly immunocompromised patients, especially those with AIDS.

Clinical Manifestations and Differential Diagnosis Days or weeks after inoculation, a papule develops at the site and then usually ulcerates but is not very painful. Similar lesions develop sequentially along the lymphatic channels proximal to the original lesion **(Fig. 225-1)**. Some patients develop a fixed cutaneous lesion that can be verrucous or ulcerative and that remains localized without lymphatic extension. The differential diagnosis of lymphocutaneous sporotrichosis includes nocardiosis (especially by *Nocardia brasiliensis*), tularemia, nontuberculous mycobacterial infection (especially by *Mycobacterium marinum*), and leishmaniasis. Osteoarticular sporotrichosis is usually monoarticular and can present as chronic synovitis or septic arthritis. Pulmonary sporotrichosis presents with cavitary or occasionally multifocal noncavitary pneumonia and must be differentiated from tuberculosis and other fungal pneumonias. Numerous ulcerated skin lesions, with or without spread to visceral organs (including the central nervous system [CNS]), are characteristic of disseminated sporotrichosis.

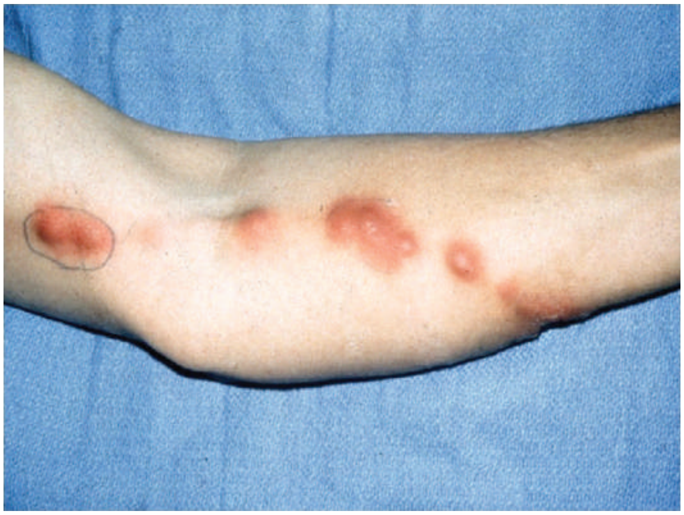

FIGURE 225-1 Several nodular lesions in a lymphangitic spread pattern that developed on the patient's arm after traumatic inoculation. Cultures from the biopsied lesion (*circle*) yielded growth of *Sporothrix schenckii.*

Diagnosis *S. schenckii* usually grows readily as a mold on Sabouraud's agar when material from a cutaneous lesion biopsy is incubated at room temperature. Histopathologic examination of biopsy material shows a mixed granulomatous and pyogenic reaction, and tiny oval or cigar-shaped yeasts sometimes can be seen with special stains.

Treatment and Prognosis Guidelines for the management of the various forms of sporotrichosis have been published by the Infectious Diseases Society of America **(Table 225-2)**. Itraconazole is the drug of choice for cutaneous and lymphocutaneous sporotrichosis. Fluconazole is less effective, voriconazole is ineffective, and posaconazole has been used successfully in a small number of patients. Saturated solution of potassium iodide (SSKI) continues to be used for lymphocutaneous infection because of its low cost; however, SSKI is poorly

TABLE 225-2 Suggested Treatment for Sporotrichosis

DISEASE	FIRST-LINE THERAPY	ALTERNATIVES/COMMENTS
Cutaneous, lymphocutaneous	Itraconazole, 200 mg/day until 2–4 weeks after lesions resolve	SSKI, increasing doses[a] Terbinafine, 500 mg twice daily
Pulmonary, osteoarticular	Itraconazole, 200 mg twice daily for 12 months	Lipid AmB[b] for severe pulmonary disease until stable; then itraconazole
Disseminated, central nervous system	Lipid AmB[b] for 4–6 weeks	Itraconazole, 200 mg twice daily after AmB for 12 months Patients with AIDS: itraconazole maintenance, 200 mg/day until CD4+ T cell count is >200/μL for ≥12 months

[a] The starting dosage is 5–10 drops three times daily in water or juice. The dosage is increased weekly by 10 drops per dose, as tolerated, up to 40–50 drops three times daily. [b] The dose of lipid AmB is 3–5 mg/kg daily; the higher dose should be used when the central nervous system is involved.

Abbreviations: AmB, amphotericin B; SSKI, saturated solution of potassium iodide.

tolerated because of adverse reactions, including metallic taste, salivary gland swelling, rash, and fever. High-dose terbinafine may be effective for lymphocutaneous infection. Treatment for lymphocutaneous sporotrichosis is continued for 2–4 weeks after all lesions have resolved, usually for a total of 3–6 months. The success rate for treatment of lymphocutaneous sporotrichosis is 90–100%.

Pulmonary and osteoarticular forms of sporotrichosis are treated with itraconazole for at least 1 year. Severe pulmonary infection and disseminated sporotrichosis, including that involving the CNS, should be treated initially with amphotericin B (AmB), with a switch to itraconazole after improvement has been noted. Lifelong suppressive therapy with itraconazole is often required for patients with AIDS. These forms of sporotrichosis respond poorly to antifungal therapy.

EUMYCETOMA AND CHROMOBLASTOMYCOSIS

Etiologic Agents, Epidemiology, and Pathogenesis Dematiaceous or brown-black fungi, the common soil organisms that cause phaeohyphomycoses, contain melanin, which causes the hyphae and conidia to be darkly pigmented. The term *phaeohyphomycosis* is used to describe any infection with a pigmented mold. This definition encompasses two cutaneous syndromes discussed herein—eumycetoma and chromoblastomycosis—and all other types of invasive infections caused by these organisms, which are discussed in **Chap. 226**. It is important to note that eumycetomas can be caused by hyaline molds and by brown-black molds and that only ~50% of all mycetomas are due to fungi. Actinomycetes cause the remainder **(Chap. 180)**.

The most common cause of eumycetoma is *Madurella* species. *Fonsecaea*, *Phialophora*, and *Cladophialophora* species are responsible for most cases of chromoblastomycosis and exhibit a pathogenic muriform phase. These infections are most often seen in resource-limited rural settings and affect mostly adult men. Inoculation is via minor, often unnoticed, trauma with soil or plant material contamination. Although exposure to these pathogens is common, the development of clinical disease is infrequent. Eumycetoma is found predominantly in tropical and subtropical areas with most cases reported from Sudan, Mexico, and India. Neutrophils mediate early inflammation around the hyphae-containing grains, and macrophages form granulomas at the site of infection.

Chromoblastomycosis is also prevalent in tropical and subtropical regions, with most cases reported from Latin America, Madagascar, and China. A broader geographic distribution is seen with certain species such as *Cladophialophora carrionii*, which thrives in semiarid climates and is a chief cause of chromoblastomycosis in Australia. Following traumatic inoculation, transformation of the fungus into thick-walled, multiseptated, muriform cells ensues. These cells can evade phagocytosis, resulting in chronic disease development.

Clinical Manifestations and Differential Diagnosis Eumycetoma—commonly called Madura foot—is a chronic cutaneous and subcutaneous infection that usually occurs on the lower extremities and manifests with swelling, sinus tract development, and appearance of grains that are actually fungal colonies discharged from the sinus tract. As the infection progresses, adjacent fascia and bony structures may become involved. The disease is indolent and disfiguring, progressing slowly over years. Complications include fractures of infected bone and bacterial superinfection.

Chromoblastomycosis is an indolent subcutaneous infection characterized by nodular, verrucous, or plaque-like painless lesions that occur predominantly on the lower extremities and grow slowly over months to years. There is hardly ever extension to adjacent structures, as is seen with eumycetoma. Long-term consequences include bacterial superinfection, chronic lymphedema, and (rarely) the development of squamous cell carcinoma.

Diagnosis A tentative clinical diagnosis of mycetoma can be made when a patient presents with a lesion characterized by swelling, sinus tracts, and grains. Histopathologic examination and culture are necessary to confirm that the etiologic agent is a mold and not an actinomycete. In chromoblastomycosis, the diagnosis rests on the histologic demonstration of muriform cells or "sclerotic bodies" in tissues; culture establishes the causative pigmented mold. In culture-negative cases, PCR-based detection of the 28S ribosomal subunit or the rRNA internal transcribed spacer (ITS) may assist with identification.

Treatment and Prognosis Treatment of eumycetoma and chromoblastomycosis involves both surgical excision of lesions and antifungal therapy. Surgical removal of lesions is most effective if performed before extensive spread has occurred and is strongly recommended as a mainstay of treatment for both syndromes. In chromoblastomycosis, cryosurgery, heat therapy, and laser therapy have been used with variable success. Optimal antifungal therapy may vary depending on the causative fungal agent. Eumycetoma has been treated with itraconazole, voriconazole, posaconazole, and less commonly terbinafine with variable success rates. Long-term antifungal therapy is required, and cost or tolerability of these agents may limit their use. Itraconazole, terbinafine, and flucytosine have been used to treat chromoblastomycosis, again with variable success and with cost or tolerability limiting long-term treatment. Muriform cells may respond to antifungal therapy differently from hyphal forms, so caution is needed in interpreting in vitro susceptibility results. Species identification may be useful in determining optimal therapy, as different dematiaceous molds can differ in their susceptibilities to azoles, AmB, and flucytosine.

FURTHER READING

Bonifaz A, Tirado-Sanchez A: Cutaneous disseminated and extracutaneous sporotrichosis: Current status of a complex disease. J Fungi 3:6, 2017.

Hay R: Dermatophytosis (ringworm) and other superficial mycoses, in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 10th ed. Blaser MJ et al (eds). Philadelphia, Elsevier, 2025.

Hospenthal D: Uncommon fungi and related species, in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 10th ed. Blaser MJ et al (eds). Philadelphia, Elsevier, 2025.

226

Uncommon Disseminated Fungal Infections

Joseph Pechacek, Carol A. Kauffman, Michail S. Lionakis

ENDEMIC MYCOSES (DIMORPHIC FUNGI)

Dimorphic fungi—so named for their ability to convert between mold and yeast forms (or spherules or adiaspores in the case of *Coccidioides* and *Emmonsia crescens*, respectively) in a temperature-dependent manner—are found in discrete environmental niches as molds that produce conidia, which are the infectious form. In infected tissues and at temperatures of >35°C, they are found in the yeast form. The most common invasive endemic mycoses in the United States—histoplasmosis, coccidioidomycosis, and blastomycosis—are discussed in **Chaps. 218–220**, respectively, and sporotrichosis is discussed in **Chap. 225**.

PARACOCCIDIOIDOMYCOSIS

Etiologic Agent, Epidemiology, and Pathogenesis The species constituting the *Paracoccidioides brasiliensis* complex and the less frequently reported *Paracoccidioides lutzii* are thermally dimorphic fungi found in humid areas of Central and South America. The highest incidence of *Paracoccidioides* infections is seen in Brazil, Colombia, Venezuela, and Ecuador although infections have been identified as far north as Mexico. Most cases are found in an ecological niche characterized by high rainfall and humidity and the presence of tropical or subtropical forests. A striking male-to-female ratio ranges from 14:1 to as high as 70:1 in various reports; this observation is largely accounted for by chronic disease, as the acute form of the infection lacks this stark male predominance. Most patients are middle-aged or elderly men from rural areas. Paracoccidioidomycosis develops after the inhalation of aerosolized conidia encountered in the environment into the lungs. In this early stage, neutrophils and macrophages play an important role in controlling the infection and preventing dissemination. Granulomas may be found in the lung and at sites of extrapulmonary dissemination in infected patients. Disease rarely develops at the time of the initial infection but appears years later, presumably after reactivation of a latent infection. Inborn errors of immunity affecting the interleukin (IL)-12/interferon (IFN)-γ signaling axis predispose to severe, disseminated paracoccidioidomycosis.

Clinical Manifestations The majority of patients exposed to *Paracoccidioides* develop asymptomatic pulmonary disease, which is contained and does not become clinically apparent. When clinical disease develops, two major syndromes are associated with paracoccidioidomycosis: the acute or juvenile form and the chronic or adult form. The acute form is uncommon, develops typically in persons <30 years old, and manifests primarily as disseminated infection of the reticuloendothelial system. The interval from initial fungal exposure to developing the acute form of paracoccidioidomycosis varies from as early as 45 days to as late as several months. Immunocompromised individuals also can develop this type of rapidly progressive disease. Peripheral eosinophilia is often seen in this form of the disease. The chronic form of paracoccidioidomycosis accounts for ~90% of cases and predominantly affects older men. The primary manifestations are progressive pulmonary disease, primarily in the lower lobes, with fibrosis and ulcerative and nodular mucocutaneous lesions that occur primarily in mucous membranes of the upper respiratory tract and that must be differentiated from leishmaniasis **(Chap. 233)** and squamous cell carcinoma **(Chap. 82)**. Metastatic foci of paracoccidioidomycosis may also occur via hematogenous or lymphatic spread, most often involving the central nervous system (CNS), skin, or adrenal glands, which can lead to the development of adrenal insufficiency.

Diagnosis The diagnosis is established by growth of the mold form of *Paracoccidioides* in culture at room temperature. A presumptive diagnosis can be made by detection of the distinctive thick-walled yeast in tissue biopsy specimens, which appears as a central yeast with multiple daughter cells attached circumferentially via narrow-necked buds and is classically described as a mariner's wheel. Serologic testing also can aid in diagnosis, although it should be noted that certain available tests may not detect *P. lutzii*.

Treatment and Prognosis Itraconazole is the treatment of choice for paracoccidioidomycosis **(Table 226-1)** owing to its longer-standing experience in this disease, although voriconazole and posaconazole also are active in vitro and have been effective in small patient cohorts. Sulfonamides have been used for years and are the least costly agents; however, the response is slower and the relapse rate higher. Seriously ill patients, such as those with respiratory failure, should be initially treated with amphotericin B (AmB) with subsequent azole therapy following clinical improvement. Patients with paracoccidioidomycosis have an excellent response to therapy, regardless of the causative species, but pulmonary fibrosis can be progressive in those with chronic disease.

TABLE 226-1 Suggested Treatment for Uncommon Disseminated Fungal Infections

DISEASE	FIRST-LINE THERAPY	ALTERNATIVES/COMMENTS
Paracoccidioidomycosis		
Chronic (adult form)	Itraconazole[a], 100–200 mg/day for 6–12 months	Voriconazole[a], 200 mg twice daily for 6–12 months Posaconazole[a], 300 mg/day for 6–12 months TMP-SMX, 160/800 mg twice daily for 12–36 months
Acute (juvenile form)	Lipid AmB[b] until improvement	Itraconazole[a], 200 mg twice daily after AmB for 12 months Voriconazole or posaconazole at doses noted above may be used
Talaromycosis		
Mild or moderate	Itraconazole[a], 200 mg twice daily for 12 weeks	Voriconazole[a], 200 mg twice daily Posaconazole[a] 300 mg/day
Severe	Lipid AmB[b] until improvement	Itraconazole[a], 200 mg twice daily after AmB for 12 weeks Voriconazole or posaconazole may be considered as an alternative
Maintenance therapy (AIDS)	Itraconazole, 200 mg/day until CD4+ T cell count is >100/μL for ≥6 months	
Emergomycosis (Emmonsiosis)		
Mild or moderate	Itraconazole[a] 200 mg twice daily for 12 weeks	Voriconazole[a], 200 mg twice daily Posaconazole[a] 300 mg/day
Severe	Lipid AmB[b] until improvement	Itraconazole[a], 200 mg twice daily after AmB for 12 weeks Voriconazole or posaconazole may be considered as an alternative
Phaeohyphomycosis	Voriconazole[a], 200 mg twice daily Itraconazole[a], 200 mg twice daily Posaconazole[a], 300 mg/day	Lipid AmB may be effective against some mold species
Fusariosis	Voriconazole[a], 200–300 mg twice daily Lipid AmB, 5 mg/kg/day Posaconazole[a], 300 mg/day	Lipid AmB plus voriconazole or posaconazole is used by some physicians for initial therapy. Investigational olorofim or fosmanogepix may be considered if available.
Scedosporiosis and lomentosporiosis	Voriconazole[a], 200–300 mg twice daily Posaconazole[a], 300 mg/day	Not susceptible to AmB. *Lomentospora prolificans* is resistant to almost all commercially available antifungal drugs. Investigational olorofim or fosmanogepix may be considered if available.
Trichosporonosis	Voriconazole[a], 200–300 mg twice daily	Posaconazole[a], 300 mg/day

[a] Dosing should be adjusted for appropriate drug trough levels. [b] The dose of lipid AmB is 3–5 mg/kg daily; the higher dose should be used when the central nervous system is involved.

Abbreviations: AmB, amphotericin B; TMP-SMX, trimethoprim-sulfamethoxazole.

■ TALAROMYCOSIS

Etiologic Agent, Epidemiology, and Pathogenesis *Talaromyces marneffei* (formerly *Penicillium marneffei*) is a thermally dimorphic fungus that is endemic in the soil in certain areas of Vietnam, Thailand, and other southeastern Asian countries. The epidemiology of talaromycosis (formerly *penicilliosis*) is linked to bamboo rats that are infected with the fungus but rarely manifest disease. The disease occurs most often among persons living in rural areas in which the rats are found, but there is no evidence for transmission of the infection directly from rats to humans. Accidental laboratory inoculation has also been reported. Infection is rare in immunocompetent hosts, and most cases are reported in persons who have advanced AIDS. Other immune deficiencies associated with a higher risk of developing talaromycosis include corticosteroid use, solid organ transplantation, idiopathic CD4 lymphocytopenia, inborn errors of immunity affecting the IL-12/IFN-γ signaling axis, and neutralizing autoantibodies against IFN-γ. Infection is presumed to result from the inhalation of conidia from the environment though a definitive environmental reservoir that has yet to be identified. The organism converts to the yeast phase in the lungs and then spreads hematogenously throughout the reticuloendothelial system. IFN-γ–primed monocytes/macrophages and neutrophils both play a role in control of *Talaromyces* via fungal phagocytosis and killing.

Clinical Manifestations The clinical manifestations of talaromycosis mimic those of disseminated histoplasmosis and include fever, fatigue, weight loss, dyspnea, lymphadenopathy, hepatosplenomegaly, and mucocutaneous lesions, which appear as papules that often umbilicate and resemble molluscum contagiosum **(Chap. 201)**. Patients with mild talaromycosis may present with isolated skin lesions.

Diagnosis Talaromycosis is diagnosed by culture of *T. marneffei* from blood or from biopsy samples of skin, bone marrow, or lymph node. The organism usually grows within 1 week as a mold producing a distinctive red pigment that diffuses into the agar. Histopathologic examination of tissues, blood smears, or material from skin lesions shows oval or elliptical yeast-like organisms that reproduce via bipolar fission (i.e., appearing as two elliptical yeasts separated by a central septation) and can quickly establish a presumptive diagnosis. Intramacrophage yeast-like organisms can also be seen on tissue or lymph node sections or bone marrow aspirates. *T. marneffei*–specific antigen and polymerase chain reaction (PCR)-based assays are under investigation.

Treatment and Prognosis Disseminated talaromycosis is usually fatal if not treated. With treatment, the mortality rate is ~10%. For mild or moderate infection, itraconazole is the drug of choice. Voriconazole also can be used, and posaconazole has good in vitro activity and has been successfully used in few case reports. Severe infection should be treated with AmB until clinical improvement occurs; then therapy can be changed to an effective triazole (Table 226-1). For patients with AIDS, suppressive therapy with itraconazole is recommended until the CD4+ T cell count has been >100 cells/μL for at least 6 months. Primary prophylaxis may be considered in certain patients with AIDS in endemic areas who are unable to receive effective antiretroviral therapy.

■ EMERGOMYCOSIS (DISSEMINATED EMMONSIOSIS)

Etiologic Agents, Epidemiology, and Pathogenesis As the name implies, *Emergomyces* (several members of which are formerly of the genus *Emmonsia*) is a genus of emerging thermally dimorphic fungi consisting of seven species, six of which have been shown to cause human disease. Among them, *E. pasteurianus* is the most geographically widespread species, and *E. africanus* has been the most studied given its relatively high incidence of infection in South Africa. Infection by *Emergomyces* species have now been reported in Europe, Asia (primarily in China and India), Africa (Uganda and South Africa), and North America (United States and Canada). Although *Emergomyces* has yet to be cultured from the environment and no animal reservoir has been identified thus far, infection is thought to begin, as with other dimorphic fungi, through inhalation of conidia released by the mold form of the pathogen in the environment. The organism then undergoes temperature-dependent transformation into the invasive yeast form within the lungs and disseminates hematogenously to cause systemic disease in immunocompromised patients. Clinically apparent disease appears largely restricted to patients with impaired cellular immunity, including those with AIDS, hematologic malignancy, or solid organ transplantation, and recipients of immunosuppressive therapy such as corticosteroids.

Clinical Manifestations Pulmonary disease can manifest with parenchymal lesions and/or endobronchial lesions. Although isolated lung involvement has been described, the vast majority of patients with pulmonary disease have concomitant cutaneous involvement, which most frequently manifests with disseminated ulcers, papules, or nodules. Dissemination to the liver, bone marrow, and lymph nodes may also occur, whereas dissemination to the CNS is rare.

Diagnosis Diagnosis is made by culturing *Emergomyces* species from blood or tissue or by showing tissue invasion upon histologic examination. *Emergomyces* may grow from aerobic blood culture bottles within 7 days; the use of fungal isolator culture bottles may improve sensitivity. The yeast form appears on histopathology as intracellular and extracellular round or ovoid yeasts with narrow-based budding that may be misidentified as *Histoplasma capsulatum*; therefore, definitive identification is made on mold phase in culture. It is worth noting that the commercially available *Histoplasma* antigen test has a high degree of cross-reactivity with *Emergomyces* species.

Treatment and Prognosis Given the lack of clinical trials to guide optimal treatment of emergomycosis, treatment recommendations are currently based on expert opinion and largely mirror treatment recommendations for disseminated histoplasmosis. Immunosuppressed patients with emergomycosis should initiate treatment with AmB until clinical improvement occurs followed by therapy with a mold-active azole. Itraconazole is the azole most often used, although in vitro susceptibility testing results of voriconazole and posaconazole have been favorable. In case reports, itraconazole monotherapy has been used successfully for mild disease. Prolonged secondary prophylaxis may be considered in patients without a reversible cause of immunodeficiency.

■ ADIASPIROMYCOSIS

Adiaspiromycosis is a rare infection caused predominantly by the mold *Emmonsia crescens*, which is found in soil. Occupational dust exposure is a major risk factor. Human infections have been broadly reported including from the Americas, Europe, and Asia. The lungs are by far the most frequently affected organ, although cases of ocular and appendiceal disease have been described. The disease is defined by the presence of adiaspores in affected tissue, which results in a marked inflammatory response. After exposure to conidia via inhalation, ingestion, or mucosal contact with dust, the host mounts an immune response leading to the development of noncaseating granulomas. Once within host tissue and driven by thermal pressure, the conidia enlarge to thick-walled adiaspores measuring up to 500 μm in diameter that are incapable of budding or replication. Clinical sequelae, including respiratory failure, are caused by the host's granulomatous inflammatory response in infected tissue. As *E. crescens* has not been cultured from human tissue with adiaspiromycosis, diagnosis depends upon histopathologic identification of adiaspores with surrounding granulomas, often with the assistance of molecular methods. The disease is often self-limiting. However, in severe disease, corticosteroids, with or without antifungal therapy, have been used successfully in several reports.

MOLD INFECTIONS

Molds produce conidia or spores, which cause infection via inhalation or traumatic skin inoculation. In infected tissues, filamentous structures termed *hyphae* form. The most common invasive mold infections—aspergillosis and mucormycosis—are discussed in **Chaps. 223** and **224**, respectively. Dematiaceous or brown-black molds cause phaeohyphomycoses that can be either localized or disseminated infections, the latter mostly in immunocompromised patients. Three genera of hyaline (nonpigmented) molds, *Fusarium*, *Scedosporium*, and *Lomentospora*, have become prominent pathogens among immunocompromised patients. Infections with these fungi mimic aspergillosis in their clinical manifestations and their histopathologic appearance in tissues.

PHAEOHYPHOMYCOSES

Dematiaceous or brown-black fungi cause phaeohyphomycoses. The term *phaeohyphomycosis* is used to describe any infection with a melanin-producing fungus with filamentous growth. Melanin is a virulence factor for all pigmented molds. Most dematiaceous fungi cause localized subcutaneous infections after direct inoculation. Disseminated infections and serious focal visceral infections can occur in immunocompetent individuals but are much more often seen in immunocompromised patients. Two specific cutaneous phaeohyphomycoses—eumycetoma and chromoblastomycosis—are discussed in **Chap. 225**.

Etiologic Agents, Epidemiology, and Pathogenesis A large number of genera and species of pigmented molds can cause human infection. Most are found in the soil or on plants, and some cause economically important plant diseases. *Alternaria*, *Exophiala*, *Curvularia*, *Cladophialophora*, *Bipolaris*, and *Wangiella* species are among the more common pigmented molds reported to cause human infection. Infections with dematiaceous molds are acquired by traumatic inoculation into the eye or through the skin, by inhalation, or by injection of contaminated medication. In 2012, *Exserohilum* species caused a large outbreak in the United States of severe—and in some patients fatal—infections of the CNS following the injection of methylprednisolone contaminated with this fungus. Subcutaneous infections follow traumatic inoculation and are frequently caused by *Curvularia*, *Exophiala*, and *Phialophora*, among other genera. Several organisms, specifically *Cladophialophora bantiana* and *Rhinocladiella mackenziei*, are neurotropic and likely to cause CNS infection. When a patient is immunocompromised or when a pigmented mold is injected directly into a deep structure, these organisms become opportunistic pathogens, invading blood vessels and mimicking better-known opportunistic infections, such as aspergillosis.

Disseminated phaeohyphomycosis may occur in patients with solid organ or hematopoietic stem cell transplantation or after corticosteroid use. It has also been described in patients with autosomal recessive deficiency in CARD9, an intracellular adaptor protein that relays fungus-sensing signals from multiple cell-surface C-type lectin receptors (CLRs). A recent report indicated that ~70% of putatively immunocompetent patients who developed disseminated and brain phaeohyphomycosis carried deleterious variants in *CLEC7A*, the gene that encodes for the CARD9-coupled CLR Dectin-1, which recognizes fungal cell wall β-glucan. Dectin-1 and CARD9 protect against this infection by mediating the production of proinflammatory cytokines in response to fungal infection, promoting macrophage fungal killing.

Clinical Manifestations Dematiaceous molds are the most common cause of allergic fungal sinusitis and a less common cause of invasive fungal sinusitis. Keratitis occurs with traumatic corneal inoculation. Even in many immunocompromised patients, inoculation through the skin generally produces only localized nodular lesions at the entry site. Other immunocompromised patients develop pneumonia, brain abscess, or disseminated infection. Patients with brain abscess often have no other signs of dissemination at the time of diagnosis and typically present with an indolent course. Certain species, such as *Cladophialophora bantiana*, exhibit a propensity for intraventricular spread within the CNS. In the *Exserohilum* outbreak mentioned above, epidural injections of fungus-contaminated glucocorticoids led to meningitis, basilar stroke, epidural abscess and phlegmon, vertebral osteomyelitis, and arachnoiditis.

Diagnosis The specific diagnosis of infection with a pigmented mold is established by growth of the organism in culture. Histologically, observation of melanized hyphal forms appearing golden-brown on biopsy specimens stained with hematoxylin and eosin supports the diagnosis of phaeohyphomycosis. Specialized melanin stains, such as the Fontana-Masson stain, highlight melanized fungal structures in infected tissues and provide important diagnostic support **(Fig. 226-1)**. PCR assays are increasingly used in the diagnosis of infection due to dematiaceous molds but are available only through fungal reference laboratories.

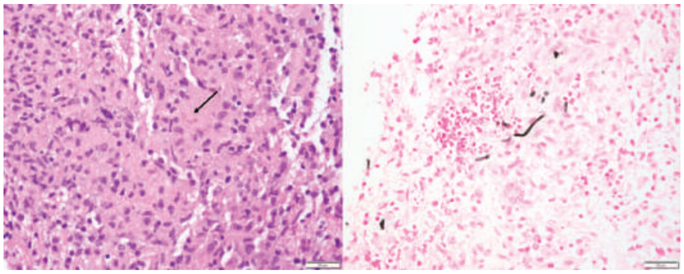

FIGURE 226-1 Histologic stains revealing pigmented mold from brain biopsy of a man with multifocal brain lesions. Cultures grew *Cladophialophora bantiana*. Images show tissue stained with hematoxylin and eosin (*left panel*) showing golden-brown pigmented hyphae (*arrow*) within multinucleated giant cells and associated neutrophilic infiltrate and Fontana-Masson stain (*right panel*) highlighting the melanin found in the cell wall of this organism. Scale bars, 20 μm. *(Images taken by Dr. Stefania Pittaluga, NCI.)*

Treatment and Prognosis Surgical debridement is vital for the treatment of phaeohyphomycosis in addition to antifungal treatment, which may need to be prolonged, especially in patients with CNS involvement. The choice of antifungal agent to treat disseminated and focal visceral infections with brown-black molds is based on the location and extent of the infection, in vitro test results, and clinical experience with the specific infecting organism. AmB is not effective against many of these organisms but has been used successfully against some species (Table 226-1). Itraconazole, voriconazole, or posaconazole can be used in the treatment of localized infections. Voriconazole has been the preferred azole when infections involve the CNS because it reaches adequate concentrations at that anatomic site; recent reports support posaconazole as an alternative azole. For certain pigmented molds, such as *C. bantiana*, flucytosine may have a role in addition to AmB or an azole. Disseminated and focal visceral infections, especially those involving the CNS, are associated with high mortality rates.

FUSARIOSIS

Etiologic Agent, Epidemiology, and Pathogenesis *Fusarium* species, which are found worldwide in soil and on plants, have emerged as major opportunistic pathogens in markedly immunocompromised patients. Strains of the *F. solani* complex are most commonly responsible for human infections, followed by *F. oxysporum*, *F. proliferatum*, and other species. *Fusarium* are molds with distinctive banana-shaped macroconidia. While *Fusarium* keratitis and onychomycosis can occur in immunocompetent patients, disseminated disease develops in immunodeficient patients. An outbreak of severe *Fusarium* keratitis among soft contact lens wearers was traced back to a particular brand of contact lens solution and individual contact lens cases that had been contaminated with this mold. In 2022 and 2023, two separate outbreaks of *F. solani* meningitis were reported in previously healthy young women who had received epidural anesthesia for plastic surgery procedures at two private clinics in Mexico. The case-fatality rate was ~50%.

Disseminated infection is reported most often in patients who have a hematologic malignancy, are neutropenic, have received hematopoietic stem cell or solid organ transplantation or corticosteroids, or have severe burn wounds. Neutrophils are the most critical cells for protection against fusariosis; hence the correlation with profound and prolonged neutropenia and susceptibility to disseminated infection. Neutrophil-derived reactive oxygen species are partly responsible for effective host defense against *Fusarium*. Patients with chronic granulomatous disease (CGD), who carry deleterious mutations in the NADPH oxidase complex that impair the phagocyte oxidative burst, are at risk for fusariosis, albeit with far lower frequency compared with aspergillosis. Several studies have shown that the prognosis of fusariosis in neutropenic patients is significantly improved in those who recovered their neutrophil counts relative to those who remained persistently neutropenic.

Exposure to *Fusarium* can be through inhalation of conidia or direct inoculation via trauma, but disseminated disease can also occur secondary to pre-existing local disease, such as sinusitis or onychomycosis. For example, the presence of *Fusarium* onychomycosis preceding allogeneic hematopoietic stem cell transplantation has been associated with a higher risk of developing disseminated fusariosis post-transplantation.

Clinical Manifestations In immunocompetent persons, *Fusarium* species cause localized infections of various organs. These organisms are a common cause of fungal keratitis, which can extend into the anterior chamber of the eye, cause loss of vision, and require corneal transplantation. Onychomycosis due to *Fusarium* species, while a potential source of discomfort in immunocompetent patients, is a source of subsequent hematogenous dissemination and should be aggressively sought and treated in neutropenic patients. In profoundly immunocompromised patients, fusariosis is angioinvasive, and clinical manifestations mimic those of aspergillosis. Pulmonary infection is characterized by multiple nodular lesions that are best identified on CT scan. Fusariosis typically presents with persistent fever in neutropenic patients despite the administration of broad-spectrum antibiotics. Disseminated fusariosis differs from disseminated aspergillosis in that skin lesions are extremely common with fusariosis and may develop shortly after or even concurrently with the development of fever; the lesions are nodular or necrotic, are usually painful, and may appear over time in different locations **(Fig. 226-2)**.

Diagnosis The diagnostic approach usually includes both documentation of the growth of *Fusarium* species from involved tissue and demonstration of invasion by histopathologic techniques that show acute angle septate hyphae in tissues. The organism is difficult to differentiate from *Aspergillus* species in tissues; therefore, identification with culture is imperative. An extremely helpful diagnostic clue is growth in blood cultures, which are positive in as many as 50% of patients with disseminated fusariosis. A positive blood culture may lead to rapid identification; subsequent rapid growth on agar reveals the characteristic banana-shaped macroconidia of *Fusarium*. Further identification of the species and antimicrobial susceptibility testing, generally in a reference laboratory, may aid in choosing the most appropriate therapy.

Treatment and Prognosis *Fusarium* species are resistant to many antifungal agents. A lipid formulation of AmB, voriconazole, or posaconazole is recommended. Many physicians use both a lipid formulation of AmB and either voriconazole or posaconazole because susceptibility information is not standardized and is not always predictive of clinical response, bearing in mind that the prognosis is largely dictated by the degree of immune restoration. *F. solani* complex is frequently resistant to all azoles, while other species may have lower minimum inhibitory concentrations (MICs) to voriconazole and posaconazole. AmB is the most active antifungal drug against *Fusarium*; however, the MICs are relatively high for these organisms compared with other molds (Table 226-1). Mortality rates for disseminated fusariosis can be as high as 85%, with survival dependent on reversal of neutropenia. With the improved antifungal therapy now available, mortality rates have fallen to ~50%. However, if neutropenia persists, the mortality rate approaches 100%.

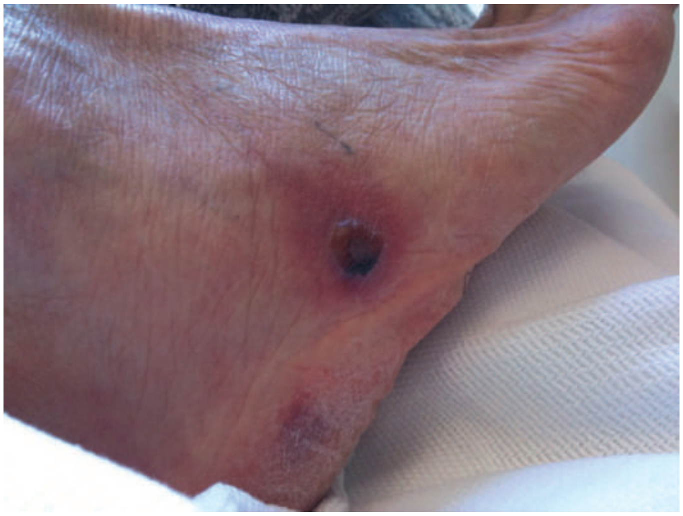

FIGURE 226-2 Painful necrotic foot lesion that developed over a week in a woman with acute leukemia who had been neutropenic for 2 months. *Fusarium* species grew from a punch biopsy. *(Courtesy of Dr. Nessrine Ktaich.)*

Two newer investigational antifungal agents show promise for the treatment of fusariosis. Fosmanogepix, which inhibits fungal Gwt1 and glycosylphosphatidylinositol–mannoprotein synthesis and anchoring, exhibits in vitro activity against *F. solani*, including against strains that are resistant to all available antifungals; it has been reported to successfully treat a few patients with meningeal and disseminated fusariosis. Olorofim, which inhibits fungal dihydroorotate dehydrogenase and pyrimidine synthesis, has shown favorable, but strain-specific, activity against *F. solani* complex in vitro.

SCEDOSPORIOSIS AND LOMENTOSPORIOSIS

Etiologic Agent, Epidemiology, and Pathogenesis *Scedosporium apiospermum* complex, which is composed of several related species, is reported more often as a cause of human infection than *Lomentospora prolificans*, formerly *Scedosporium prolificans*, but both are major emerging opportunistic pathogens in immunocompromised hosts and can cause pneumonia, disseminated infection, and brain abscess. Organisms of the *S. apiospermum* complex are found worldwide in temperate climates in tidal flats, swamps, ponds, manure, and soil. *L. prolificans* is also found in soil but is more geographically restricted with a greater representation of reported infections from Australia and Spain. Infection occurs predominantly through inhalation of conidia, but direct traumatic inoculation through the skin or the eye can also occur. *S. apiospermum* infections of the lungs and CNS have been described in putatively immunocompetent patients following near-drowning accidents, likely owing to the associated heavy burden of fungal exposure from contaminated water.

Clinical Manifestations Among immunocompetent persons, *Scedosporium* and *Lomentospora* species are a prominent cause of eumycetoma **(Chap. 225)**. Keratitis resulting from accidental corneal inoculation is a sight-threatening infection. Traumatic inoculation during accidents may also lead to deep-seated infections, such as osteoarticular infections including septic arthritis, in immunocompetent patients. In patients who have hematologic malignancies, especially those with acute leukemia and prolonged neutropenia, recipients of solid organ or hematopoietic stem cell transplants, CGD patients, and patients receiving corticosteroids, these organisms are angioinvasive and can cause pneumonia and extrapulmonary dissemination, with a propensity for CNS spread. Pulmonary infection mimics that of aspergillosis both clinically and in appearance on CT scan: nodules, cavities, and lobar infiltrates are common radiographic findings. Contiguous spread from the lung to adjacent anatomic sites, such as the vertebrae, has rarely been described. Disseminated infection often involves the heart, brain, and many other organs. Skin lesions are not as common or as painful as those in patients with fusariosis.

Diagnosis Diagnosis depends on the growth of *Scedosporium* or *Lomentospora* species from infected tissue and the histologic demonstration of acute-angle septate hyphae invading tissues. Culture evidence is essential because these molds are difficult to morphologically differentiate from *Aspergillus* in tissues. Demonstration of tissue invasion is essential because these ubiquitous environmental molds can be mere contaminants or colonizers in the respiratory tract, especially in persons with bronchiectasis. As with *Fusarium*, *L. prolificans* can grow in blood cultures, but *S. apiospermum* usually does not.

Treatment and Prognosis *Scedosporium* and *Lomentospora* species are intrinsically resistant to AmB, echinocandins, and some azoles. Voriconazole has been the agent of choice for *S. apiospermum*, and posaconazole has been increasingly used for this infection (Table 226-1). *L. prolificans* is resistant in vitro to almost every commercially available antifungal agent; the addition of agents such as terbinafine to a voriconazole regimen has been attempted because in vitro data suggest possible synergy between these two agents against some strains of *L. prolificans*. Mortality rates for invasive *S. apiospermum* infection are ~50%, but those for invasive *L. prolificans* infection remain as high as 85–100%. The novel antifungal agents fosmanogepix and olorofim have shown in vitro activity against both *S. apiospermum* and *L. prolificans* and have demonstrated efficacy in preclinical mouse models of these infections. Ongoing clinical trials for these agents will help determine their clinical efficacy in infected patients.

YEAST INFECTIONS

In addition to the far more common human pathogenic yeasts *Candida* and *Cryptococcus*, which are discussed in **Chaps. 221** and **222** the yeast-like fungus *Trichosporon* has emerged as a significant opportunistic pathogen among immunocompromised patients.

TRICHOSPORONOSIS

Etiologic Agent, Epidemiology, and Pathogenesis The genus *Trichosporon* encompasses many species, some of which cause localized infection of hair and nails; white piedra is a superficial infection of the hair and scalp caused by *T. ovoides* that is characterized by the appearance of white nodules along the hair shaft. The most common species responsible for invasive infection is *Trichosporon asahii,* although other species also can cause disseminated disease. *Trichosporon* species grow as yeast-like colonies in vitro; in vivo, however, hyphae, pseudohyphae, and arthroconidia can be seen in addition to yeast forms. These yeasts are commonly found in soil, sewage, and water and in rare instances can colonize the human skin and the gastrointestinal and respiratory tracts. Most infections follow inhalation or entry via central venous catheters. The presence of foreign material is a risk factor for trichosporonosis, as both catheter-associated infections and prosthetic cardiac valve infections have been described. Systemic infection occurs almost exclusively in immunocompromised hosts, especially those who have hematologic malignancies, are neutropenic, have received a solid organ or hematopoietic stem cell transplantation, or are receiving corticosteroids. Severe neutropenia is a major risk factor for disseminated disease. Myeloid phagocytes and their oxidative burst are critical for protection against *Trichosporon*, as illustrated by the development of invasive trichosporonosis in patients with CARD9 deficiency or CGD.

Clinical Manifestations Disseminated trichosporonosis resembles invasive candidiasis, and fungemia is often the initial manifestation of infection typically presenting with fever. Pneumonia, skin lesions, and sepsis are common, although many patients can present without clinically apparent lung infection. The skin lesions in disseminated disease begin as papules or nodules surrounded by erythema and progress to central necrosis. Renal involvement is common in disseminated disease and may cause funguria and hematuria. A rare chronic form of infection mimics hepatosplenic candidiasis, now known as chronic disseminated candidiasis.

Diagnosis The diagnosis of systemic *Trichosporon* infection is established by growth of the organism from infected tissues or from blood. Histopathologic examination of a skin lesion showing a mixture of yeast forms, arthroconidia, pseudohyphae, and hyphae can lead to an early presumptive diagnosis of trichosporonosis. The presence of arthroconidia in particular can help differentiate trichosporonosis from candidemia. The serum cryptococcal antigen latex agglutination test may be positive in patients with disseminated trichosporonosis because *T. asahii* and *Cryptococcus neoformans* share polysaccharide antigens.

Treatment and Prognosis Rates of response to AmB have been disappointing, and many *Trichosporon* isolates are resistant to AmB in vitro. Voriconazole has been the antifungal agent of choice (Table 226-1). The mortality rates for disseminated *Trichosporon* infection have been as high as 70% but are decreasing with the use of voriconazole; however, neutrophil recovery is vital in overcoming this infection, and mortality rates with persistent neutropenia remain very high despite antifungal therapy.

FURTHER READING

De Almeida Junior JN, Hennequin C: Invasive *Trichosporon* infections: A systematic review on a re-emerging fungal pathogen. Front Microbiol 7:1629, 2016.

Hospenthal D: Uncommon fungi and related species in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 10th ed, Blaser MJ et al (eds). Philadelphia, Elsevier, 2025.

Neoh CF et al: Scedosporiosis and lomentosporiosis: Modern perspectives on these difficult-to-treat rare mold infections. Clin Microbiol Rev 37:e0000423, 2024.

Nucci M et al: Fusariosis. Semin Respir Crit Care Med 36:706, 2015.

Revankar SG et al: A Mycoses Study Group international prospective study of phaeohyphomycosis: An analysis of 99 proven/probable cases. Open Forum Infect Dis 4:ofx200, 2017.

227

Pneumocystis Infections

Alison Morris, Henry Masur

DEFINITION AND DESCRIPTION

Pneumocystis is an opportunistic pathogen that is an important cause of pneumonia in immunocompromised hosts, particularly those with HIV infection **(Chap. 208)**, organ transplants, or hematologic malignancies and those receiving high-dose glucocorticoids or certain immunosuppressive monoclonal antibodies. *Pneumocystis* was discovered in rodents in 1909 and was initially believed to be a protozoan. Because *Pneumocystis* cannot be cultured, understanding its biology has been limited, but molecular techniques have demonstrated that the organism is actually a fungus. Formerly known as *Pneumocystis carinii*, the species specific to humans has been renamed *Pneumocystis jirovecii*; the species that is specific to rats is *Pneumocystis carinii*; there are multiple variants of *Pneumocystis* that are specific to other animals.

EPIDEMIOLOGY

Pneumocystis jirovecii pneumonia (PCP) came to medical attention in the early 1950s when pathologists in Czechoslovakia recognized *Pneumocystis* in the alveolar exudates of infants involved in nursery outbreaks of interstitial pneumonia, outbreaks that had been described in Europe since the 1920s. Among adults, PCP was rarely recognized until the populations of immunosuppressed adults increased due to the development of immunosuppressive therapies for solid-organ transplantation, bone marrow transplantation, cancer, and autoimmune disorders, and until the development of better pulmonary diagnostic techniques such as bronchoscopy. In 1981, PCP was first reported in men who had sex with men and in intravenous (IV) drug users who had no obvious cause of immunosuppression. These cases were subsequently recognized as the first cases of what came to be known as the acquired immunodeficiency syndrome (AIDS) **(Chap. 208)**.

The incidence of PCP increased dramatically as the AIDS epidemic grew: without PCP chemoprophylaxis or antiretroviral therapy (ART), 80–90% of patients with HIV/AIDS in North America and Western Europe ultimately developed one or more episodes of PCP. While its incidence declined with the introduction of anti-*Pneumocystis* prophylaxis and durably effective ART, PCP has continued to be an important cause of AIDS-associated morbidity in the United States and Western Europe, particularly in individuals who do not know they are infected with HIV until they are profoundly immunosuppressed and in people living with HIV (PLWH) with CD4+ T lymphocyte counts of <200/μL who are not receiving ART or PCP prophylaxis.

PCP also develops in HIV-uninfected patients who are immunocompromised secondary to congenital immunodeficiencies, hematologic or malignant neoplasms, stem cell or solid-organ transplantation, and treatment with immunosuppressive medications. The incidence of PCP depends on the degree and duration of immunosuppression. PCP is increasingly reported among individuals receiving tumor necrosis factor α inhibitors and certain (but not all) immunosuppressive monoclonal antibodies for autoimmune, rheumatologic, or neoplastic diseases. In many health care systems, PCP occurs more often due to non-HIV-related immunosuppression than due to HIV infection.

While clinical disease due to *Pneumocystis* in immunocompetent hosts has not been clearly documented, studies have shown that *Pneumocystis* organisms can cause subclinical infection among children and adults who are not immunocompromised and can be associated with pulmonary pathology. The relevance of these organisms to acute or chronic clinical syndromes, such as chronic obstructive pulmonary disease (COPD), in immunocompetent patients is being investigated.

In some developing countries, the incidence of PCP among PLWH has been reported to be lower than that in more industrialized countries. This lower incidence may be due to competing mortality from infectious diseases such as tuberculosis and bacterial pneumonia, which typically occur before patients become immunosuppressed enough to develop PCP. Geographic variations in *Pneumocystis* exposure and underdiagnosis attributable to lack of diagnostic resources also may explain the apparent lower frequency of PCP in some countries.

PATHOGENESIS AND PATHOLOGY

Life Cycle and Transmission The life cycle of *Pneumocystis* likely involves both sexual and asexual reproduction. The organism exists as a trophic form, a cyst, and a precyst. Studies in rodents show that immunocompetent animals can serve as reservoirs for respiratory transmission of *P. carinii* (the infecting species in rats) to immunocompetent and immunosuppressed rats. Human *Pneumocystis* is thought to be transmitted by a respiratory route as well. *P. jirovecii*, like all pneumocystis species, is host-specific. Thus, humans are not infected, for example, by *P. carinii* (rodents) or *P. oryctolagi* (rabbits), but are only infected by *P. jirovecii,* which is from other humans.

Serologic and molecular studies have demonstrated that most humans are exposed to *P. jirovecii* and infected early in life. It was historically thought that *Pneumocystis* pneumonia usually developed from reactivation of latent infection. However, molecular evidence makes it clear that children and adults can develop PCP from primary infection or reinfection. The source of infection is thought to be either healthy or immunosuppressed individuals who themselves experienced recent infection or reinfection, or immunosuppressed persons with clinical PCP. Nosocomial outbreaks among immunosuppressed persons occur in inpatient and outpatient settings. The utility of droplet or airborne isolation for preventing transmission from patients with PCP to other immunosuppressed individuals has been debated; no clear evidence exists, although it seems prudent to isolate patients with active PCP from other immunosuppressed patients using at least droplet precautions.

Role of Immunity Defects in cellular and/or humoral immunity predispose to development of PCP. Such defects may be congenital, or they may be acquired as a result of HIV infection or treatment with certain immunosuppressive drugs such as glucocorticoids, fludarabine, temozolomide, temsirolimus, rituximab, or alemtuzumab. CD4+ T cells are critical in host defense against *Pneumocystis*. Among PLWH, the incidence of PCP is inversely related to the CD4+ T-cell count: at least 80% of cases occur at counts of <200/μL, and most cases actually develop at counts of <100/μL. HIV viral load is another factor that predisposes patients to PCP. Clinicians must recognize that PCP can occur at CD4+ T-cell counts >200/μL in persons with HIV infection, but such occurrences are uncommon, especially with CD4+ T-cell counts substantially higher than 200/μL.

CD4+ T-cell counts are less useful in predicting the risk of PCP in patients who are immunosuppressed for reasons other than HIV infection. In these populations, a CD4+ T-cell count <200/μL is a sensitive indicator of susceptibility to PCP. However, a CD4+ T-cell count >200/μL in many populations does not imply protection.

Lung Pathology *Pneumocystis* has a unique tropism for the lung. Organisms are presumably inhaled into the alveolar space after being exhaled by another human. Clinically apparent pneumonia occurs only if an individual is immunocompromised. *Pneumocystis* proliferates in the lung, provoking a mononuclear cell response. The alveoli become filled with proteinaceous material, and alveolar damage results in increased alveolar-capillary injury and surfactant abnormalities. Stained lung sections typically show foamy, vacuolated alveolar exudates composed largely of viable and nonviable organisms **(Fig. 227-1*A*)**. Interstitial edema and fibrosis may develop, and organisms can be seen in the alveolar space with silver or other stains. The organisms can also be seen when tissue is subjected to colorimetric or immunofluorescent staining **(Fig. 227-1*B*–*D*)**.

CLINICAL FEATURES

Clinical Presentation PCP presents as acute or subacute pneumonia that may initially be characterized by a vague sense of dyspnea alone, but that subsequently manifests as fever and nonproductive cough with progressive shortness of breath. Patients may ultimately progress to respiratory failure and death. Extrapulmonary manifestations of PCP are rare, but can include involvement of almost any organ, most notably the lymph nodes, spleen, and liver.

Physical Examination, Oxygen Saturation, and Imaging The physical examination findings in PCP are nonspecific. Patients have decreased oxygen saturation—at rest or with exertion—that, without treatment, progresses to severe hypoxemia. Patients may initially have a normal chest examination and no adventitious sounds, but later develop diffuse rales and signs of consolidation.

Laboratory Findings The results of routine laboratory tests are nonspecific in PCP. Serum levels of lactate dehydrogenase (LDH) are often elevated as a result of pulmonary damage; however, a normal LDH level does not rule out PCP, nor is an elevated LDH value specific for PCP. The peripheral white blood cell count may be elevated in relation to the patient's baseline values, but the increase is usually modest and the baseline may have been below usual normal limits due to HIV infection. Hepatic and renal function are typically normal.

Radiographic Findings Although the initial chest radiograph may be normal when patients have mild symptoms, the classic radiographic appearance of symptomatic PCP consists of diffuse bilateral interstitial infiltrates that are perihilar and symmetric **(Fig. 227-2*A*)**—another finding that is not specific for PCP. The interstitial infiltrates can progress to alveolar filling **(Fig. 227-2*B*)**. High-resolution chest computed tomography (CT) shows diffuse ground-glass opacities in virtually all patients with PCP, often before a routine chest radiograph becomes abnormal **(Fig. 227-2*C*)**. A normal chest CT essentially rules out the diagnosis of PCP. Pneumatoceles and pneumothoraces are characteristic chest radiographic findings, especially in patients with HIV infection **(Fig. 227-2*D*)**. A wide variety of atypical radiographic findings have been described, including asymmetric patterns, upper-lobe infiltrates, mediastinal adenopathy, nodules, cavities, and effusions.

DIAGNOSIS

The optimal sample for a specific microbiologic diagnostic examination depends on how ill the patient is and what resources are available. Before the 1990s, diagnoses of PCP were usually established by open lung biopsy; later, transbronchial lung biopsy was employed. Hematoxylin and eosin (H and E) staining of pulmonary tissue

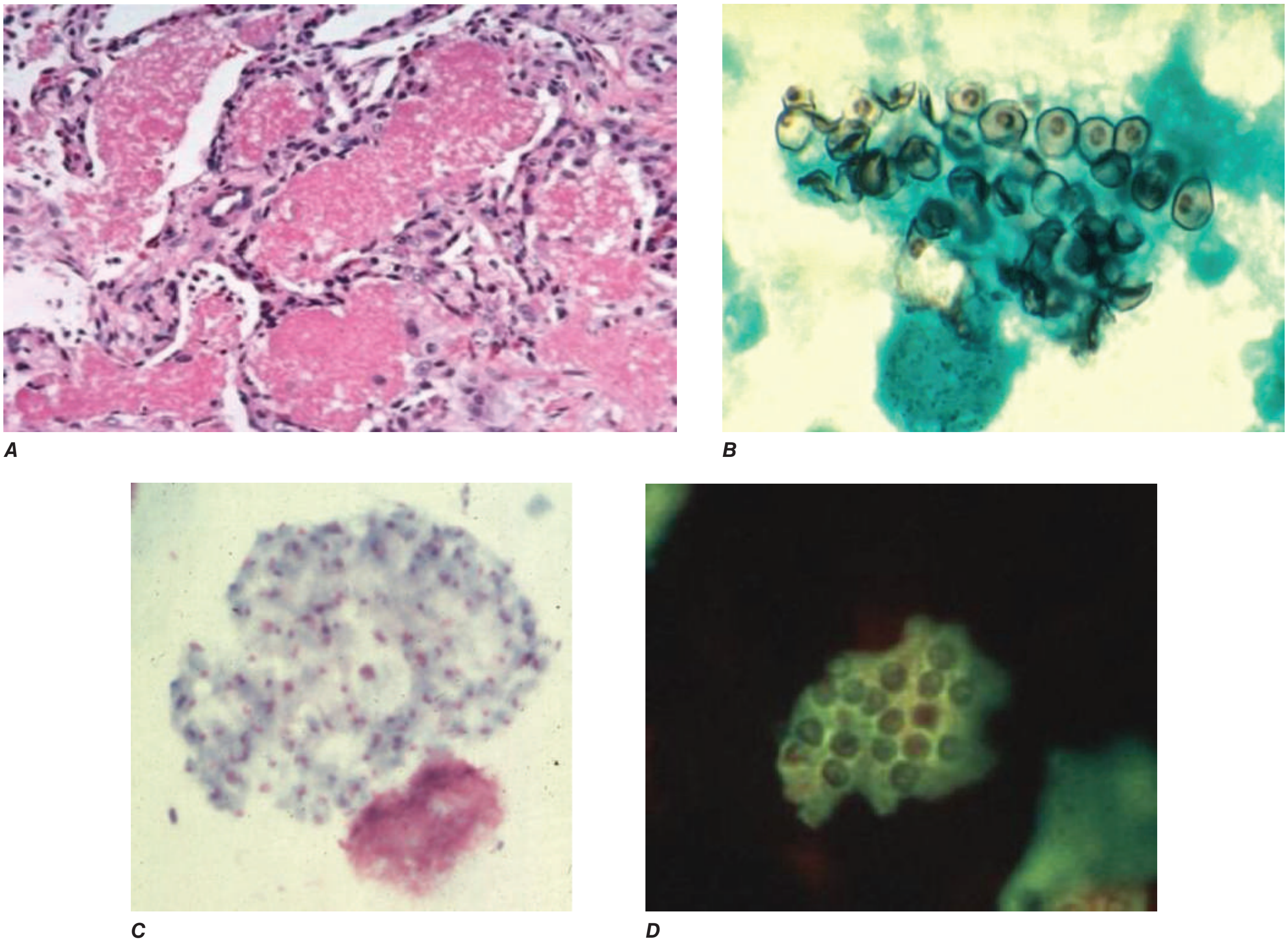

FIGURE 227-1 Direct microscopy of *Pneumocystis* pneumonia. ***A.*** Transbronchial lung biopsy stained with hematoxylin and eosin shows eosinophilic alveolar filling. ***B.*** Methenamine silver–stained bronchoalveolar lavage (BAL) fluid. ***C.*** Giemsa-stained BAL fluid. ***D.*** Immunofluorescent stain of BAL fluid.

demonstrates a foamy alveolar filling and a mononuclear interstitial infiltrate (Fig. 227-1*A*). This appearance is pathognomonic for PCP even though the organisms cannot be specifically identified with this H and E stain. Since human pneumocystis has never been successfully cultured, the diagnosis is typically established in lung tissue or pulmonary secretions by staining of the cyst and or trophozoite—e.g., with methenamine silver (Fig. 227-1*B*), toluidine blue O, or Giemsa (Fig. 227-1*C*)—or by staining with a specific immunofluorescent antibody (Fig. 227-1*D*).

Since the 1990s, bronchoalveolar lavage (BAL) has become the most common method for obtaining a sample of respiratory secretions in which to detect *Pneumocystis* organisms. The demonstration of organisms in BAL fluid is almost 100% sensitive and specific for PCP. The organisms are identified in pulmonary secretions with the specific stains indicated above for lung biopsy. While expectorated sputum or throat swabs have very low sensitivity, an induced sputum sample obtained and interpreted by an experienced provider can be highly sensitive and specific; however, the sensitivity is dependent on both the characteristics of the patient, the quality of the sputum sample, and the experience of the center conducting the test and is widely variable (55–90%).

Many laboratories offer polymerase chain reaction (PCR) testing of respiratory specimens for *Pneumocystis* in preference to direct microscopy of stained respiratory secretions. However, PCR tests are so sensitive that it is difficult to distinguish patients with colonization (i.e., those whose acute lung disease is due to some other process but who have low levels of *Pneumocystis* DNA in their lungs) from those with acute pneumonia due to *Pneumocystis.* The PCR cycle number is probably helpful, i.e., the lower the cycle number, the more DNA was detected. However, PCR tests for PCP are not standardized and specimen collection can vary considerably resulting in variable quality of the specimens assessed. Therefore, there is no specific PCR threshold that can be deemed to be especially convincing. Such PCR tests on appropriate samples may be more useful for ruling out a diagnosis of PCP if they are negative than for definitively attributing the disease to *Pneumocystis* if they are positive.

There has been considerable interest in serologic tests, such as assays for (1→3)-β-D-glucan, a component of the fungal cell wall. Serum (1→3)-β-D-glucan levels are frequently elevated in patients with PCP. However, serum or BAL (1→3)-β-D-glucan levels are not perfectly sensitive or highly specific for PCP. There are increasing numbers of studies of serum PCR tests for *Pneumocystis*, but such serum tests are not yet useful for establishing the presence or absence of PCP.

■ COURSE AND PROGNOSIS

Untreated, PCP is invariably fatal. Patients with HIV infection often have an indolent course that may present early as mild exercise intolerance or chest tightness without fever or cough and a normal or nearly normal posterior–anterior chest radiograph. However, this process progresses over days, weeks, or even a few months to fever, cough, diffuse alveolar infiltrates, and profound hypoxemia. Some patients with HIV infection and most patients with other types of immunosuppression have more acute disease that progresses over a few days to respiratory failure. Rare patients also develop distributive shock due to PCP. A few unusual patients present with extrapulmonary manifestations in the skin or soft tissue, retina, brain, liver, kidney, or spleen. Extrapulmonary disease is nonspecific in presentation and can be diagnosed only by histology. When there is extrapulmonary clinical disease in a

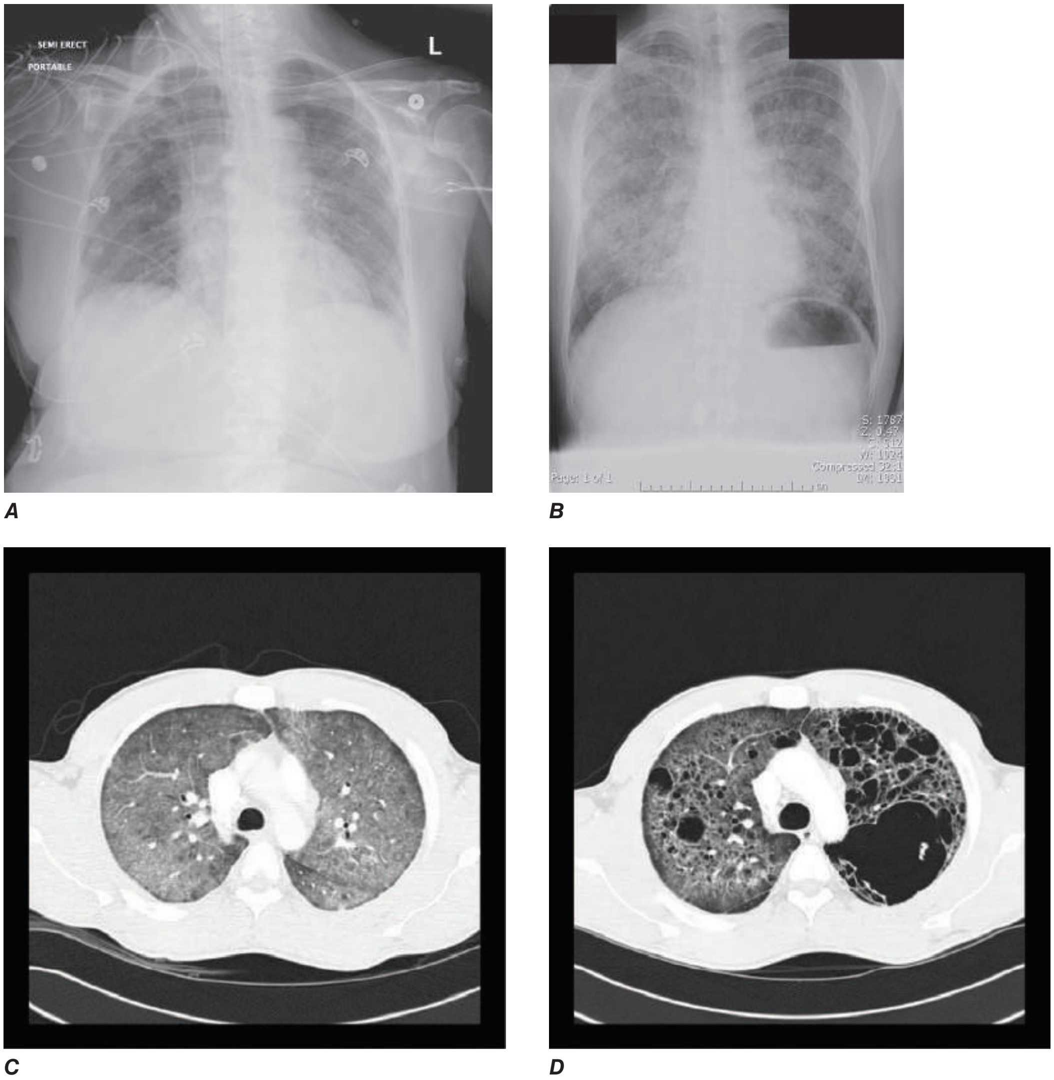

FIGURE 227-2 Radiographs in *Pneumocystis* pneumonia. ***A.*** Posterior–anterior chest radiograph showing symmetric interstitial infiltrates. ***B.*** Posterior–anterior chest radiograph showing symmetric alveolar infiltrates (*courtesy of Alison Morris*). ***C.*** Computed tomography (CT) image demonstrating symmetric interstitial infiltrates and ground-glass opacities. ***D.*** CT image showing symmetric interstitial infiltrates, ground-glass opacities, and pneumatoceles.

patient with PCP, the priority is to determine what other concurrent infectious or neoplastic process might be present, given the rarity of extrapulmonary pneumocystosis.

Factors that influence mortality risk of PCP include the patient's age and degree of immunosuppression as well as the presence of preexisting lung disease, the need for mechanical ventilation, and the development of a pneumothorax. With advances in supportive critical care, the prognosis for patients with PCP who require intubation and respiratory support has improved and now depends to a large extent on comorbidities and the prognosis of the underlying disease. Some patients do not respond to therapy for 4–8 days, and thus supportive care for a minimum of 10 days is a reasonable consideration if such support is compatible with the patient's wishes and the prognosis of comorbidities. In fact, if patients of any level of severity are treated with specific therapy but without corticosteroids, they often deteriorate during the first few days, presumably due to enhanced inflammation induced by dying organisms. Patients whose condition continues to deteriorate after 3 or 4 days or have not improved after 7–10 days should be reevaluated to determine whether other infectious processes are present (either having been missed on initial evaluation or having developed during treatment), whether initial anti-*Pneumocystis* treatment has failed, or whether noninfectious processes (e.g., congestive heart failure, pulmonary emboli, pulmonary hypertension, drug toxicity, or a neoplastic process) are causing pulmonary dysfunction.

TREATMENT

P. jirovecii Pneumonia

The treatment of choice for PCP is trimethoprim-sulfamethoxazole (TMP-SMX), given either IV or PO for 14 days to non-HIV-infected patients with mild disease and for 21 days to all other patients **(Table 227-1)**. TMP-SMX, which interferes with the organism's folate metabolism, is at least as effective as alternative agents and is better tolerated. TMP-SMX can cause leukopenia, hepatitis, rash, fever, elevation of potassium and creatinine, and anaphylactic and anaphylactoid reactions. Patients with HIV infection have an unusually high incidence of hypersensitivity to TMP-SMX. Monitoring of serum drug levels is useful if renal function or toxicities are issues in order to enhance the likelihood that therapy will be effective and toxicity will be avoided. Maintenance of a 2-h postdose serum sulfamethoxazole level of 100–150 μg/mL has been associated with a successful outcome. Resistance to TMP-SMX cannot be measured by organism growth inhibition in the laboratory because human *Pneumocystis* cannot be successfully cultured. However, mutations in the target gene for sulfamethoxazole that confer in vitro sulfa resistance when found in other organisms have been recognized in *Pneumocystis*. The clinical relevance of these mutations for the response to therapy is uncertain. Sulfadiazine plus pyrimethamine, an oral regimen more often used for treatment of toxoplasmosis, also is highly effective.

TABLE 227-1 Treatment of *Pneumocystis* Pneumonia[a]

DRUG(S)	DOSE, ROUTE	ADVERSE EFFECTS
First-Choice Agent		
TMP-SMX	TMP (5 mg/kg) plus SMX (25 mg/kg) q6–8h PO or IV (i.e., 2 double-strength tablets tid or qid)	Fever, rash, cytopenias, hepatitis, hyperkalemia
Alternative Agents		
Atovaquone	750 mg bid PO	Rash, fever, hepatitis
Clindamycin ***plus*** Primaquine	300–450 mg q6h PO or 600 mg q6–8h IV 15–30 mg qd PO	Hemolysis (G6PD deficiency), methemoglobinemia, neutropenia, rash
Pentamidine	3–4 mg/kg qd IV	Hypotension, azotemia, cardiac arrhythmias (torsades des pointes), pancreatitis, dysglycemias, hypocalcemia, neutropenia, hepatitis
Adjunctive Agent		
Prednisone or methylprednisolone	40 mg bid × 5 d, 40 mg qd × 5 d, 20 mg qd × 11 d; PO or IV	Peptic ulcer disease, hyperglycemia, mood alteration, hypertension

[a]Treatment can be administered for 14 days to non-HIV-infected patients with mild disease and for 21 days to all other patients.

Abbreviations: G6PD, glucose-6-phosphate dehydrogenase; TMP-SMX, trimethoprim-sulfamethoxazole.

Either intravenous pentamidine or the combination of clindamycin plus primaquine is are options for patients who cannot tolerate TMP-SMX and for patients in whom treatment with TMP-SMX appears to be failing. Pentamidine must be administered IV over at least 60 min to avoid potentially lethal hypotension. Adverse effects can be severe and irreversible and include renal dysfunction, dysglycemia (life-threatening hypoglycemia that can occur days or weeks after initial infusion and be followed by hyperglycemia), neutropenia, and torsades de pointes. Clindamycin plus primaquine is effective, but primaquine can be given only by the oral route—a disadvantage for patients who cannot ingest or absorb oral drugs and primaquine can cause hemolysis in patients with glucose-6-phosphate dehydrogenase deficiency and can cause methemoglobinemia. Oral atovaquone is also a reasonable option for patients with mild disease who have no impediments to absorbing an oral drug that requires a high-fat meal for optimal absorption. There is some evidence for activity of echinocandins against the cyst form (but not the trophozoite form) of *Pneumocystis*, but the role for echinocandins as part of combination therapy is currently uncertain, and these drugs should never be used as single drug therapy for PCP.

A major advance in therapy for PCP was the recognition that glucocorticoids could improve survival rates among PLWH with moderate to severe disease (initial room air Po_2 <70 mmHg or alveolar–arterial oxygen gradient ≥35 mmHg). Glucocorticoids appear to reduce the pulmonary inflammation that occurs after specific therapy is started and organisms begin to die, eliciting inflammation. Therapy with glucocorticoids should be the standard of care for patients with HIV infection and moderate to severe PCP. Therapy with glucocorticosteroids is also probably effective for patients with other immunodeficiencies who have moderate to severe PCP. This treatment should be started for moderate or severe disease when therapy for PCP is initiated, even if the diagnosis is suspected but has not yet been confirmed. If PLWH or HIV-uninfected patients are receiving high-dose glucocorticoids when they develop PCP, there are theoretical advantages to decreasing the steroid dose to improve immune function, but there is no convincing evidence on which to base any specific strategy.

No definitive trials have identified the best therapeutic algorithm for patients in whom TMP-SMX treatment for PCP is failing. If no other treatable infectious or noninfectious processes are detected and pulmonary dysfunction appears to be due to PCP alone, many authorities would switch from TMP-SMX to either IV pentamidine or IV clindamycin plus oral primaquine. Clindamycin-primaquine is certainly less toxic than IV pentamidine. Some authorities would add the second drug or drug combination to TMP-SMX rather than switching regimens. If patients are not already receiving them, glucocorticoids should be added to the regimen; the dosage and regimen are usually chosen empirically and depend on what glucocorticoid regimen (if any) the patient was receiving when PCP therapy was begun.

For PLWH who present with PCP before the initiation of ART, ART should be started within the first 2 weeks of therapy for PCP in most situations. Immune reconstitution inflammatory syndrome (IRIS) can occur, and the decision to initiate ART thus requires considerable clinical judgement that factors in the severity of pneumonia, the response of PCP to therapy, and concurrent medical conditions.

■ PREVENTION

The most effective method for preventing PCP is to eliminate the cause of immunosuppression by withdrawing immunosuppressive therapy or treating the underlying cause (e.g., HIV infection). Patients who are susceptible to PCP benefit from chemoprophylaxis during the period of susceptibility. For patients with HIV infection, CD4+ T-cell counts are a reliable marker of susceptibility, and counts <200/μL are an indication to start or continue prophylaxis **(Table 227-2)**.

For patients who are immunosuppressed as a result of factors other than HIV infection, CD4+ T-cell counts <200/μL are a plausible but not absolute marker of susceptibility. However, counts >200/μL are not a reliable marker of protection from PCP. For these patients who are immunosuppressed due to causes other than HIV infection, the period of susceptibility is usually estimated on the basis of experience with the underlying disease and the specific immunosuppressive regimen. Cessation of prophylaxis has been associated with clusters of cases in certain patient populations, such as solid-organ transplant recipients, where the period of susceptibility is not well-defined. Patients receiving a prolonged course of high-dose glucocorticoids

TABLE 227-2 Prophylaxis of *Pneumocystis* Pneumonia

DRUG(S)	DOSE, ROUTE	COMMENTS
First-Choice Agent		
TMP-SMX	1 tablet (double- or single-strength) qd PO	Incidence of hypersensitivity is high. Rechallenge for non-life-threatening hypersensitivity; consider dose-escalation protocol.
Alternative Agents		
Dapsone	50 mg bid or 100 mg qd PO	Hemolysis is associated with G6PD deficiency.
Dapsone ***plus*** Pyrimethamine ***plus*** Leucovorin	50 mg qd PO 50 mg weekly PO 25 mg weekly PO	Leucovorin ameliorates cytopenias due to pyrimethamine.
Dapsone ***plus*** Pyrimethamine ***plus*** Leucovorin	200 mg weekly PO 75 mg weekly PO 25 mg weekly PO	Leucovorin ameliorates cytopenias due to pyrimethamine.
Pentamidine	300 mg monthly via Respirgard II nebulizer	Aerosol may cause bronchospasm. Pentamidine is probably less effective than TMP-SMX or dapsone regimens.
Atovaquone	1500 mg qd PO	Requires fatty meal for optimal absorption.

Abbreviations: G6PD, glucose-6-phosphate dehydrogenase; TMP-SMX, trimethoprim-sulfamethoxazole.

appear to be particularly susceptible to PCP. The glucocorticoid exposure threshold that warrants chemoprophylaxis is ill-defined, but such preventive therapy should be strongly considered for any patient who is receiving more than the equivalent of 20 mg of prednisone daily for 30 days or who is receiving glucocorticoids in conjunction with other immunosuppressive agents. Clinical experience also suggests that chemoprophylaxis is useful for patients receiving certain immunosuppressive agents (e.g., tumor necrosis factor inhibitors, antithymocyte globulin, rituximab, and alemtuzumab). The duration of such chemoprophylaxis is empirically estimated based on prior clinical experience and immunologic factors that would plausibly relate to immunity, such as CD4+ T-cell counts, recognizing that such estimates are not precise.

TMP-SMX is the most effective prophylactic drug; few patients experience a PCP breakthrough when they are reliably taking a recommended TMP-SMX chemoprophylactic regimen. Several TMP-SMX regimens have been used successfully. Regimens of one single-strength or double-strength tablet daily are the regimens with which there is the most experience, but one double-strength tablet two or three times weekly also has been recommended for various PLWH and non-HIV-infected populations of patients.

For patients who cannot tolerate TMP-SMX (usually because of hypersensitivity or bone marrow suppression), alternative drugs include daily dapsone, weekly dapsone-pyrimethamine, atovaquone, and monthly aerosol pentamidine. Patients who develop hypersensitivity to TMP-SMX can sometimes tolerate the drug if a gradual dose-escalation protocol is used. Atovaquone is effective and well tolerated; however, this drug is available only as an oral preparation, and gastrointestinal absorption is unpredictable in patients with abnormal gastrointestinal motility or function. Aerosolized pentamidine is effective, but it is not as effective as TMP-SMX and may not provide protection in areas of the lung that are not well-ventilated. Dapsone cross-reacts with sulfonamides in a substantial fraction of patients and is rarely useful in patients with a history of life-threatening reactions to TMP-SMX.

FURTHER READING

Buchacz K et al: Incidence of AIDS-defining opportunistic infections in a multicohort analysis of HIV-infected persons in the United States and Canada, 2000–2010. J Infect Dis 214:862, 2016.

Del Corpo O et al: Diagnostic accuracy of serum (1-3)-β-D-glucan for *Pneumocystis jirovecii* pneumonia: A systematic review and meta-analysis. Clin Microbiol Infect 26:1137, 2020.

Lécuyer R et al: Characteristics and prognosis factors of *Pneumocystis jirovecii* pneumonia according to underlying disease. Chest 165:1319, 2024.

Le Gal S et al: *Pneumocystis* infection outbreaks in organ transplantation units in France: A nation-wide survey. Clin Infect Dis 70:2216, 2020.

Ma L et al: Genome analysis of three *Pneumocystis* species reveals adaptation mechanisms to life exclusively in mammalian hosts. Nat Commun 7:10740, 2016.

Panel on Opportunistic Infections in HIV-Infected Adults and Adolescents: Guidelines for the prevention and treatment of opportunistic infections in HIV-infected adults and adolescents: Recommendations from the Centers for Disease Control and Prevention, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America. Available at *https://clinicalinfo.hiv.gov/en/guidelines/hiv-clinical-guidelines-adult-and-adolescent-opportunistic-infections/pneumocystis-0?view=full*. Accessed December 21, 2024.

Wills NK et al: The prevalence of laboratory-confirmed *Pneumocystis jirovecii* in HIV-infected adults in Africa: A systematic review and meta-analysis. Med Mycol 59:802, 2021.

Zolopa A et al: Early antiretroviral therapy reduces AIDS progression/death in individuals with acute opportunistic infections: A multicenter randomized strategy trial. PLoS One 4:e5575, 2009.

Section 17 Protozoal and Helminthic Infections: General Considerations

228

Introduction to Parasitic Infections

Sharon L. Reed, Charles E. Davis

The word *parasite* comes originally from the Greek *parasitos* (*para*, alongside of; and *sitos*, food), meaning someone who eats at another's table or lives at another's expense. Although the same is true of many bacteria and viruses, the designation *parasite* is reserved, by convention, for helminths and protozoa. These organisms are larger and more complex than bacteria, with a eukaryotic cell structure similar to that of human host cells. Historically, this similarity has made it difficult to find effective antiparasitic agents that do not cause unacceptable toxicity to human cells. Fortunately, intensive research and modern techniques have now provided suitable agents for safe and effective treatment of most parasitic infections. **See Chap. S12 for details on diagnostic procedures and Chap. 229 for details on treatment.**

Internal parasites of human beings are divided into two types: helminths (worms) and protozoa. *Helminths* are multicellular organisms that can often be seen with the naked eye **(Chap. 237)**. There are two phyla: Platyhelminthes (flat worms) and Nemathelminthes (roundworms). Both phyla include some genera that mature in the gastrointestinal tract and others that migrate through the tissue after ingestion or skin penetration. **Tables S12-1 and S12-2** present the helminthic genera, their definitive and intermediate hosts, their geographic distributions, and the parasitic stages in the human body.

The key to understanding which helminths use humans as definitive hosts is to remember that helminth ova develop into larvae, and larval stages develop into adults. Humans serve as the definitive host when they ingest helminth larvae, which develop into adults in the intestine and usually cause mild disease, often without any symptoms. (The exception is ingestion of the late-stage larvae of the somatic or tissue flukes, as shown in **Table S12-1**.) In contrast, if humans ingest helminth ova and serve as the intermediate host, the ova develop into larvae, which penetrate the intestine, migrate through the tissue, and invade organs where they mature into adults. Intermediate hosts with parasitic invasion of organs may experience severe disease.

Protozoa are microscopic single-celled organisms. Among the many differences between helminths and protozoans, the most important is the ability of protozoa (like bacteria) to multiply within the human body and cause overwhelming infections. A major mechanism promoting unrestrained growth is evasion of the host immune response either by antigenic variation (*Trypanosoma brucei*) or by survival inside host cells (e.g., *Plasmodium*, *Babesia*, *Cryptosporidium*, *Leishmania*, and *Toxoplasma*). In contrast, almost all helminths require stages in other hosts to complete their life cycles and multiply. As a result, except for *Strongyloides* and *Capillaria*, which can complete their life cycle in humans, increases in the burden of infection with helminths require repeated exogenous reinfections. Thus, permanent residents of endemic countries, who are exposed repeatedly, may have heavy severe infections, while most travelers with one or two exposures are unlikely to experience the full spectrum of chronic helminthic infections.

In contrast to helminthic infections, naïve patients with their first protozoal infection usually are the most severely affected because partial immunity often limits the number of parasites during recurrent infections. Protozoan replication to large numbers in the host also promotes the development of drug-resistant forms, especially in malaria **(Chap. 229)**. Because protozoa belong to many different phyla, it is easier to understand the pathogenesis and management of protozoal infections when they are classified by the site of infection (intestinal protozoans, free-living amebae, and blood and tissue protozoans)

(Table S12-3). Immunocompromised hosts are at risk of disseminated infection with several protozoa, including *Leishmania*, *Toxoplasma*, *Cryptosporidium*, *and Trypanosoma cruzi*, which are AIDS-defining illnesses. Among the helminths, *Strongyloides* can disseminate in immunocompromised individuals.

HELMINTHIC INFECTIONS

The Platyhelminthes (flatworms) are categorized as tapeworms (cestodes) and flukes (trematodes). Tapeworms are composed of a head or scolex bearing the holdfast organs and segments, which become gravid as they mature. Some tapeworms can reach lengths of many yards; the longest tapeworms develop in the intestine, where they rarely cause serious disease. In contrast, flukes are small leaf-shaped organisms whose size is not a measure of disease severity.

FLATWORMS

Cestodes Tapeworms cause either intestinal or somatic infection, depending on the species. *Intestinal* infections occur when the human host ingests larvae in the tissue of the intermediate host, whereas *somatic* infections occur when humans accidentally ingest ova excreted from the wild or domesticated definitive animal host.

INTESTINAL TAPEWORMS As shown in **Table S12-1**, humans acquire most intestinal tapeworms by eating the insufficiently cooked flesh of the intermediate host. Thus, *Taenia saginata* is commonly called the beef tapeworm, *Taenia solium* the pork tapeworm, and *Diphyllobothrium latum* the fish tapeworm. *Hymenolepis nana* is capable of completing its life cycle in the human intestine and is acquired by ingestion of infected grain beetles or of ova from infected humans or mice. None of these parasites causes significant damage, and infection is usually asymptomatic. There are two occasional exceptions. When people ingest *T. solium* ova from their own intestine or from another infected individual, it can cause somatic infection. *D. latum* avidly absorbs vitamin B_{12} in the intestine and can cause pernicious anemia in 1–2% of infected Scandinavians with a genetic predisposition.

SOMATIC TAPEWORMS There are three major causes of somatic tapeworm infections. Two species of *Echinococcus* cause echinococcosis. *E. granulosus* is acquired by accidental ingestion of ova from dogs infected when fed the infected tissues of sheep or other animals by sheepherders or hunters. *E. multilocularis* is transmitted primarily in sub-Arctic areas when humans ingest ova from foxes, dogs, or cats that have been infected through consumption of the tissues of infected rodents. Both species cause hydatid cysts when the eggs hatch into larvae, penetrate the intestine, and migrate into the liver or lung. Ingested *T. solium* ova cause somatic disease (cysticercosis) when the larvae penetrate the intestine, migrate into tissue, and form cysts (*cysterci*), usually in the muscles or central nervous system (CNS).

Trematodes Flukes also cause both intestinal and somatic infections **(Chap. 241 and Table S12-1)**. Most fluke infections are localized to Asia, Africa, Southeast Asia, or the Pacific islands. Infection with intestinal flukes is usually asymptomatic, although heavy infections sometimes cause abdominal discomfort and mucous diarrhea. Liver flukes and lung flukes cause somatic infections when humans ingest a larval form from an intermediate host. Adults develop in the intestine, migrate into adjacent tissues, and cause disease. The major liver flukes (*Clonorchis sinensis*, *Opisthorchis* spp., and *Fasciola hepatica*) are causes of recurrent bacterial cholangitis (due to obstruction) or portal hypertension and cirrhosis. Only *F. hepatica* can be acquired worldwide; it is especially common in sheep-raising areas, where the animals ingest water plants (e.g., watercress). The lung flukes (*Paragonimus* spp.) occur globally except in Europe; most lesions occur as pulmonary cysts, although occasional lesions develop in the CNS or the abdominal cavity.

The blood flukes cause schistosomiasis, one of the most common and serious parasitic infections **(Chap. 241 and Table S12-1)**. The major species are *Schistosoma mansoni*, *S. haematobium*, and *S. japonicum*. All are transmitted to humans when free-swimming larvae exit an infected snail in freshwater and penetrate the skin. Swimmer's itch sometimes follows skin penetration but is usually of short duration. The larvae then wander in the skin until they find a blood vessel and migrate to the target organ. *S. mansoni* and *S. japonicum* migrate to the mesentery vessels and eventually make their way to the liver, while *S. haematobium* targets the veins around the ureter and bladder. Extensive egg deposition by *S. mansoni* and *S. japonicum* and the immune reactions to the ova cause granuloma formation and, with many repeated exposures, portal vein obstruction and cirrhosis. The same process in the ureters and bladders during infection with *S. haematobium* eventually interferes with urine flow and leads to repeated urinary tract infections and kidney damage.

ROUNDWORMS

Nematodes Roundworms are nonsegmented bisexual organisms. The species that infect humans include intestinal and tissue groups. Humans may also acquire certain nonhuman mammalian roundworms that either can be limited to the skin or can migrate to tissues and cause serious disease (the larva migrans syndromes).

INTESTINAL ROUNDWORMS The major intestinal roundworms are *Ascaris lumbricoides*, *Necator americanus* (New World hookworms), *Ancylostoma duodenale* (Old World hookworms), *Trichuris trichiura* (whipworms), *Enterobius vermicularis* (pinworms), and *Strongyloides stercoralis*. Taken together, infections caused by intestinal roundworms are the most common infections in the world. *Ascaris*, hookworms, and *Trichuris* infect about 900 million individuals, and at least 100–370 million have strongyloidiasis. These infections are most common in resource-poor developing countries, especially where people defecate outside and/or human feces is used as fertilizer ("night soil"). Infection is transmitted either by ingestion of ova (*A. lumbricoides*, *T. trichiura*, and *E. vermicularis*) or by active penetration of the skin by larvae (hookworms and *S. stercoralis*) **(Table S12-2)**.

Intestinal roundworms cause serious health problems in residents of endemic regions with poor sanitation, but travelers are at low risk of developing significant disease from most of these parasites. Intestinal blockage and malnutrition from heavy *Ascaris* infections and anemia from heavy hookworm infections are now restricted to areas of heavy endemicity. Except in the case of *Strongyloides* and *Capillaria*, which can reproduce in the body, multiple exposures over time are necessary for the development of severe disease. *Strongyloides* infection persists over decades and can disseminate when the immune system is compromised. Although *Capillaria* remains localized to the intestine, infections can become so heavy that protein-losing enteropathy and malnutrition cause serious disease.

The life cycles of *Ascaris* and the hookworms involve migration through the heart and lungs before development into adults in the intestine. In particular, *Ascaris* occasionally causes eosinophilic pneumonia (Loeffler's syndrome) during heavy infections. Pinworms are the most common causes of intestinal roundworm infection persisting in the United States and other developed countries. The anal and perineal itching caused by pinworm migration out of the anus and subsequent egg deposition is well known to families throughout the world.

TISSUE ROUNDWORMS The major diseases caused by tissue roundworms are filariasis, angiostrongyliasis, gnathostomiasis, and trichinellosis. By far, the most important globally is filariasis; the thread-like filarial worms infect an estimated 80 million individuals in tropical and subtropical areas of the world. Four filarial species cause three distinct diseases: lymphatic filariasis (*Wuchereria bancrofti* and *Brugia malayi*), river blindness (*Onchocercus volvulus*), and loiasis (*Loa loa*, the African eye worm). Humans, the major reservoir, acquire these infections from bites of infected arthropods **(Table S12-2)**. The larvae develop into adults, which remain static in tissue: the lymphatics for lymphatic filariasis and subcutaneous tissue for *O. volvulus* and *L. loa*. After adults mate, next-stage larvae are produced, and their migration causes additional damage.

Repeated bouts of migrating larvae and blocking of the lymphatics by adults are necessary to establish the syndrome of lymphatic filariasis; thus, it is unusual for the short-term traveler (<3 months' residence in an endemic region) to develop significant disease. In river blindness, the larvae produced by adult *O. volvulus* migrate through the skin and eye, causing skin damage and eventual blindness. Loiasis is a milder disease restricted to central and western Africa. Although both the adults and the larvae of *L. loa* migrate through the skin and eye, many infected individuals are asymptomatic, and the infection is often diagnosed only when an adult worm migrates across the subconjunctival tissue and is visible to the patient and the physician. Red lumps in the skin from heavy cutaneous migration are called *Calabar swellings*.

The other four major roundworm tissue infections are acquired by ingestion of larvae in undercooked food. The sources for trichinellosis are swine and other large mammals; for gnathostomiasis, freshwater fish and chicken; for ancylostomiasis, snails, fish, prawns, and crabs; and for Guinea worm, infected water fleas. Guinea worm infection (dracunculiasis, caused by *Dracunculus medinensis*) has been almost eradicated. *Trichinella spiralis* larvae penetrate the intestine and migrate widely, with a preference for skeletal tissue; the release of eosinophils and IgE causes muscle soreness and may cause palpebral swelling and other manifestations of generalized allergic reactions. *Angiostrongylus cantonensis* is the most common parasitic cause of eosinophilic meningitis. Ingested larvae penetrate the intestine and migrate to the brain and meninges, where they quickly die and attract massive numbers of eosinophils. Although complications can occur, most individuals recover spontaneously. *Gnathostoma spinigerum* larvae also penetrate the intestine and migrate, showing a preference for the skin, eyes, and meninges. Mechanical damage from the migration and inflammation produced by the resultant immune reaction can cause boil-like lesions on the skin, painful eye damage, and eosinophilic meningitis. Although eosinophilic meningitis caused by *G. spinigerum* is less common than that caused by *A. cantonensis*, it is often more severe and can result in paralysis or brain hemorrhage.

PROTOZOAL INFECTIONS

INTESTINAL PROTOZOA

Entamoeba histolytica is the one intestinal protozoan that causes invasive disease. This disease consists of dysentery or bloody diarrhea that must be differentiated from that due to bacteria such as *Salmonella*, *Campylobacter*, and *Shigella*. Although amebiasis usually has a slower onset with lower fever than these bacterial infections, *E. histolytica* can disseminate from the bloodstream to cause distant abscesses, particularly of the liver. The diagnosis cannot be made by identification of the characteristic cyst or trophozoites **(Chap. 230)**, as they are identical to those of the noninvasive *E. dispar*, which is more common globally.

Cryptosporidium and *Giardia* are the most common water-borne protozoal infections. *Cryptosporidium* can cause major outbreaks because it is highly infectious and resistant to high levels of chlorine **(Chap. 236)**. Without immune reconstitution, immunosuppressed patients, particularly those with AIDS, can develop severe, even fatal watery diarrhea. Infections caused by the remaining intestinal protozoans—*Giardia*, *Isospora*, *Cyclospora*, and microsporidia **(Chap. 236)**—have a much more indolent course, with intermittent diarrhea. Microsporidia, unique intracellular protozoa that form infectious spores, may cause limited gastrointestinal infection in immunocompetent hosts, but patients with AIDS can develop chronic diarrhea and wasting or disseminated infection to the biliary or respiratory tract.

FREE-LIVING AMOEBAS

The free-living amoebas *Acanthamoeba* and *Naegleria* are found worldwide in freshwater and brackish water **(Chap. 230 and Table S12-3)**. Organisms of these two genera cause very different syndromes. In immunocompromised individuals, *Acanthamoeba* may cause invasive infection, with brain masses and skin lesions. However, all humans are susceptible to *Acanthamoeba* keratitis after trauma to the eye and exposure to contaminated water. In contrast, naeglerial meningitis, acquired in warm lakes or hot springs, causes sudden pyogenic and usually fatal meningitis. *Balamuthia*, reported only from the Americas, causes indolent meningoencephalitis, with both cerebrospinal fluid pleocytosis and a space-occupying lesion, in immunocompetent patients. Despite the availability of miltefosine, which is active in vitro against *Naegleria*, infection of the CNS is almost universally fatal.

BLOOD AND TISSUE PROTOZOANS

Plasmodium* and *Babesia Malaria, caused by six species of *Plasmodium*, carries higher mortality rates than any other parasitic infection **(Chap. 231)**. All species are transmitted in tropical and subtropical areas by female *Anopheles* mosquitoes. *Plasmodium falciparum* is most common in sub-Saharan Africa, where it causes more than 80% of malaria infections and 90% of malarial deaths. Infection with *P. falciparum* may be particularly severe because the organism can invade any erythrocyte, reaches very high parasite loads, damages organs by adhering to vascular epithelium, and is the most likely *Plasmodium* species to be resistant to antimalarial drugs. *Plasmodium vivax*, the dominant cause of malaria outside sub-Saharan Africa, reaches lower levels of parasitemia and exhibits less drug resistance because it invades only reticulocytes with Duffy antigen. Many Africans, especially in the western part of the continent, lack the Duffy blood group; consequently, *Plasmodium ovale*, another cause of milder malaria, can compete successfully with *P. vivax*. Both *P. vivax* and *P. ovale* produce persistent liver forms, which must be treated with primaquine **(Chap. 229)**. Because malaria can cause a variety of symptoms ranging from acute fever to coma, this diagnosis must be considered in any traveler or immigrant from a malarial area. Recently, locally acquired *P. vivax* malaria has been detected in Florida and Texas and *P. falciparum* in Maryland; thus, malaria should be included in the differential diagnoses of unexplained fever even in patients who have never left the United States. *Babesia* also infects erythrocytes and may cause a nonspecific febrile illness or, in asplenic patients, severe infection. This parasite is carried by ixodid ticks and is geographically limited to the northeastern and midwestern United States, with only sporadic cases in Europe and other temperate areas.

Trypanosomes The three species of trypanosomes all have flagellated bloodstream forms, but they cause very different diseases. *T. cruzi*, the cause of Chagas disease, is transmitted in South and Central America in the feces of blood-sucking reduviid bugs **(Chap. 234)**. After initial parasitemia, patients are often asymptomatic for years while the parasite multiplies intracellularly in muscle and ganglion cells. Although only a minority of patients go on to develop organ damage (megaesophagus and cardiomyopathy), all infected patients can spread the disease through transfusions, mother-to-child transmission, and organ transplants.

African trypanosomiasis is limited to sub-Saharan Africa, where it is transmitted by the bite of a tsetse fly. A history of a tsetse bite and the presence of a painful chancre are strong diagnostic clues **(Chap. 234)**. Although the parasites causing this disease in western Africa (*Trypanosoma brucei gambiense*) and eastern Africa (*T. brucei rhodesiense*) look identical, they are genetically and clinically distinct. *T. b. gambiense* causes low-level parasitemia with cyclical fevers over months or years before CNS invasion, whereas *T. b. rhodesiense* causes high-level parasitemia, invades the CNS early on, and can lead to death within weeks of onset.

Leishmania Leishmaniasis is caused by more than 20 species of obligate intracellular protozoa transmitted by sandflies, which are present in almost 100 countries in tropical and temperate zones **(Chap. 233)**. A wide spectrum of clinical symptoms result, ranging from self-healing, painless skin ulcers to mucocutaneous disease with destruction of the nose and palate to disseminated visceral leishmaniasis with hepatic and splenic involvement. The resulting disease depends on the infecting strain and the host immune response. Visceral leishmaniasis can present as an acute febrile illness, with the later development of hepatosplenomegaly, and is an AIDS-defining illness in HIV-infected patients. More than 90% of cases of visceral leishmaniasis occur in India, Bangladesh, Ethiopia, Sudan, and Brazil.

Toxoplasma *Toxoplasma gondii* is an obligate intracellular parasite that is found worldwide. Infection follows ingestion of oocysts in food or water contaminated by cat feces, ingestion of tissue cysts in undercooked meat, or transplacental transmission. After gastrointestinal invasion, tachyzoites can invade any nucleated cell and cause lifelong infection in most patients **(Chap. 235)**. Clinical manifestations depend on the host's age and immune status at the time of infection. Congenital toxoplasmosis results from primary maternal infection; outcomes are most severe early in pregnancy and include visual, hearing, and cognitive impairments. Babies infected later in pregnancy may appear normal but can develop chorioretinitis decades later. Primary infection in immunocompetent hosts may be asymptomatic, may present as an infectious mononucleosis–like syndrome, or may manifest as chorioretinitis during outbreaks. During immunosuppression by AIDS or organ transplantation, reactivation of latent cerebral infection can be fatal unless diagnosed and treated early.

APPROACH TO THE PATIENT

Parasitic Infection

A thorough history and physical examination are the keys to diagnosis of any disease and particularly of parasitic infections. Because many of the more serious parasitic infections are uncommon in the United States, a travel history, particularly to developing nations, is a critical component. The longer the stay in an area endemic for significant parasitic infections, the greater the risk, even for healthy travelers. In addition, other factors increase the chance of acquiring these infections. Notably, immunocompromise greatly increases the likelihood of developing some of the more serious parasitic infections. Even healthy travelers with adventure itineraries, extensive travel to rural areas, or involvement in war zones or refugee camps are at increased risk. Immigrants from developing countries may seek care for symptoms or signs associated with parasitic infections.

Information on the patient's immunization history and adherence to appropriate malarial chemoprophylaxis is critical. The recent approval of the first parasitic vaccines against *P. falciparum* is very exciting, but it will be targeted initially only for children in high prevalence areas. Typhoid fever is much less likely to be the cause of prolonged fever in an immunized individual. Similarly, hepatitis A or B is unlikely to be the cause of jaundice and fever in fully immunized patients. In this era of increasing drug resistance, even adherence to appropriate malarial chemoprophylaxis does not guarantee that fever is not malarial. Nevertheless, most travelers who acquire malaria have taken inadequate or no prophylaxis. Although these considerations do not prove that the symptoms are caused by parasites, they narrow the differential diagnosis.

There are many other important aspects of the history, including when symptoms began. Was the individual still in the endemic area at the time, or did the symptoms commence after return to the United States? If they started during travel, was any treatment received? Malaria must be the first consideration in a febrile patient returning from an endemic area. If the patient was well upon return from travel, the timing of symptom onset is a critical point. For example, if the chief manifestation is fever that began >10–14 days after departure from the endemic region, many tropical diseases can be ruled out, including dengue fever, chikungunya fever, and Zika virus infection. On the other hand, fever beginning several months or later after return makes malaria a likely diagnosis. Travelers' diarrhea, the most common complaint of travelers, is usually caused by bacteria or viruses and resolves in a short time with or without treatment. Travelers' diarrhea that persists for weeks is much more likely to be parasitic in origin.

Most patients who consult physicians after international travel either have troublesome symptoms or have been referred for symptoms or signs whose source was unclear to a referring caregiver. After a careful travel history including the individual's symptoms and the exact geographic zones visited, a thorough physical examination must be conducted. The symptoms, signs, and physical findings should help to establish possible diagnoses. Table 228-1 breaks down the symptoms of major parasitic infections by organ system and geographic distribution, with comments on clinical and epidemiologic associations.

TABLE 228-1 Parasitic Infections, by Organ System and Signs/Symptoms[a]

ORGAN SYSTEM, MAJOR SIGN(S)/SYMPTOM(S)	PARASITE(S)	GEOGRAPHIC DISTRIBUTION	COMMENTS
Skin			
Serpentine rash	Hookworm	Worldwide	Can cause anemia in heavy infections
	Strongyloides	Moist tropics and subtropics	Disseminated infection in immunocompromise
	Toxocara (animal roundworm)	Tropical and temperate zones	Cutaneous or visceral larva migrans
Itchy skin rash	*Onchocerca*	Mexico, Central/South America, Africa	Larvae detectable in skin snips and nodules
Painless ulcers	*Leishmania*	Tropics and subtropics	Amastigotes detectable in biopsies; may cause destructive mucocutaneous infection; AIDS-defining infection
Skin nodules	*Onchocerca*	Mexico, South America, Africa	Large nodules of adult worms
	Loa loa (African eye worm)	Western and central Africa	Migratory nodules
	Gnathostoma	Southeast Asia and China	Migratory nodules with eosinophilia
Painful nodules, especially involving feet	*Dracunculus* (Guinea worm)	Africa	Nearly eradicated
Central Nervous System			
Somnolence, seizures, coma	*Plasmodium falciparum*	Subtropics and tropics	Cerebral malaria, especially in children
	Trypanosoma brucei rhodesiense	Sub-Saharan eastern Africa	Painful chancre from tsetse fly bite; death in weeks to months
Space-occupying lesions, seizures	*Acanthamoeba*	Worldwide	Immunocompromised individuals
	Balamuthia	Americas	Indolent meningoencephalitis with brain mass
	Toxoplasma	Worldwide	Reactivation disease in immunocompromise; ring-enhancing lesions; AIDS-defining infection
	Taenia solium	Mexico, Central/South America, Africa	Cysticercosis; variable sized or calcified larval cysts on CT
	Schistosoma japonicum	Far East	Aberrant eggs can form brain or spinal cord masses.
	Schistosoma mansoni	Africa, Central/South America	Aberrant eggs can form brain or spinal cord masses.

(*Continued*)

TABLE 228-1 Parasitic Infections, by Organ System and Signs/Symptoms[a] (Continued)

ORGAN SYSTEM, MAJOR SIGN(S)/SYMPTOM(S)	PARASITE(S)	GEOGRAPHIC DISTRIBUTION	COMMENTS
Pyogenic meningitis	*Naegleria*	Worldwide	Motile trophozoites in fresh cerebrospinal fluid; pyogenic; rapid death
Eosinophilic meningitis	*Angiostrongylus* (rat lung worm)	Southeast Asia, Pacific, Caribbean	Most common cause globally of eosinophilic meningitis; spontaneous resolution
	Gnathostoma	Southeast Asia and China	Migratory nodules
Eyes			
Painful corneal ulcers	*Acanthamoeba*	Worldwide	Freshwater and brackish water; corneal trauma; long-wear contact lenses
Corneal opacification	*Onchocerca*	Mexico, Central/South America, Africa	Immune response to microfilaria in cornea
Congenital or adult visual loss	*Toxoplasma*	Worldwide	Primary infection in pregnancy and normal hosts; reactivation infection in immunocompromised
Retinal mass	*Toxocara*	Worldwide	Ocular larva migrans
Visible roundworm in eye	*Onchocerca*	Mexico, Central/South America, Africa	Worms may cross eye during migration.
	L. loa	Western and central Africa	Worms may cross eye during migration.
Pain, possible vision loss	*Gnathostoma*	Southeast Asia and China	Migratory skin nodules, eosinophilia
Lungs			
Pulmonary nodule/abscess	*Paragonimus*	Far East, Africa, Americas	Ectopic migration to abdomen or central nervous system
Cough, transient infiltrates, eosinophilia	Migrating helminths	Worldwide	Loeffler's syndrome from migrating *Ascaris*, hookworm, *Strongyloides*
Heart			
Pulmonary edema	*P. falciparum* (complication)	Tropics and subtropics	End-organ damage from severe malaria
Cardiomegaly, arrhythmias	*Trypanosoma cruzi*	Mexico, Central/South America	Late amastigote infection of myocardium; AIDS-defining infection
Gastrointestinal Tract			
Hepatosplenomegaly	Malaria (multiple episodes)	Tropics and subtropics	Splenomegaly with anemia and recurrent fever are hallmarks of malaria.
	S. mansoni	Africa, Central/South America	Portal obstruction with cirrhosis and late varices
	Leishmania donovani complex	Tropics and subtropics	Visceral leishmaniasis; AIDS-defining infection
Hepatomegaly	*Entamoeba histolytica*	Tropics	Acute with fever, right-upper-quadrant pain; or chronic with enlarged liver; hypoechoic abscess(es) on ultrasound or CT
	Echinococcus	Sheep-raising areas	Characteristic cysts of liver > lung
	Fasciola	Sheep-raising areas	Eosinophilia
Cholangitis	*Clonorchis*	China, Southeast Asia	Recurrent cholangitis and late cholangiocarcinoma
	Microsporidia	Worldwide	AIDS
	Cryptosporidium	Worldwide	AIDS-defining infection
Bloody diarrhea	*E. histolytica*	Tropics	Less fever than in diarrhea of bacterial etiology
	S. mansoni	Africa, Central/South America	Only in heavy, acute infection with fever and eosinophilia
	S. japonicum	Far East	Only in heavy, acute infection
Watery diarrhea	*Cryptosporidium*	Worldwide	Severe in immunocompromised patients
	Giardia	Worldwide	Foul-smelling stool with steatorrhea
	Isospora belli	Worldwide	Fever, abdominal pain, chronic diarrhea
	Microsporidia	Worldwide	Chronic diarrhea with AIDS
	Capillaria	Southeast Asia, Egypt	Malabsorption, wasting
Passage of large roundworm (>6 cm)	*Ascaris*	Worldwide	Patients may confuse the roundworm with an earthworm.
Small roundworms visible around anus	Pinworm	Worldwide	Anal itching; eggs rarely detected by ova and parasite (O&P) exam
	Trichuris	Worldwide	Rectal prolapse with heavy infection in children
Passage of tapeworm segments	*T. solium* or *Taenia saginata*	Worldwide	Usual reason for seeking medical care
	Diphyllobothrium latum	Worldwide	Pernicious anemia in genetically predisposed Scandinavians
Genitourinary System			
Itchy discharge	*Trichomonas vaginalis*	Worldwide	Common sexually transmitted disease of both sexes
Hematuria	*Schistosoma haematobium*	Africa	Hematuria with negative cultures, urinary tract infections, and late bladder cancer

(*Continued*)

TABLE 228-1 Parasitic Infections, by Organ System and Signs/Symptoms[a] (Continued)

ORGAN SYSTEM, MAJOR SIGN(S)/SYMPTOM(S)	PARASITE(S)	GEOGRAPHIC DISTRIBUTION	COMMENTS
Muscular System			
Myalgias, myositis	*Trichinella*	Worldwide	Palpebral swelling; high-level eosinophilia
Bloodstream			
Fever without localizing symptoms	*Plasmodium*	Tropics and subtropics	Consider in any patient from a malarious area.
	Babesia	New England, United States	Geographically limited; worse with splenectomy
	T. brucei rhodesiense, *T. brucei gambiense*	Sub-Saharan Africa	Limited to tsetse fly range; painful chancre; adenopathy and cyclical fevers; early (*rhodesiense*) or late (*gambiense*) central nervous system involvement
	Filariae	Asia, India	Periodic fever with eosinophilia, adenolymphangitis, chronic lymphangitis
	L. donovani complex	Tropics and subtropics	Hepatosplenomegaly, fever, wasting; AIDS-defining infection

[a] See also text and Tables S12-1, S12-2, and S12-3 for vectors and routes of transmission.

FURTHER READING

Blackburn D et al: Outbreak of locally acquired mosquito-transmitted malaria—Florida and Texas, May–July 2023. MMWR 72:973, 2023.

Conrad MD et al: Evolution of partial resistance to artemisinins in malaria parasites in Uganda. N Engl J Med 389:722, 2023.

Diaz AV et al: Reaching the World Health Organization elimination targets for schistosomiasis: The importance of a One Health perspective. Philos Trans R Soc Lond B Biol Sci 378:20220274, 2023.

Loukas A et al: The yin and yang of human soil-transmitted helminths. Int J Parasitol 51:1243, 2021.

Rubin EJ: Making the worm turn. N Engl J Med 388:1908, 2023.

229

Agents Used to Treat Parasitic Infections

Thomas A. Moore

Parasitic infections continue to afflict more than half of the world's population and impose a substantial health burden, particularly in underdeveloped nations, where they are most prevalent. The reach of some parasitic diseases, including malaria, has expanded over the past few decades due to factors such as deforestation, population shifts, global warming, and other climatic events. Although there have been significant advances in vaccine development and vector control, chemotherapy remains the single most effective means of controlling parasitic infections. Efforts to combat the spread of some diseases are hindered by the development and spread of drug resistance, the limited introduction of new antiparasitic agents, the proliferation of counterfeit medications, profiteering, and, most recently, the widespread and unsupported use of antiparasitic agents to treat COVID-19, all of which have dramatically increased the cost of these once-affordable agents. However, there are good reasons to be optimistic. Ambitious global initiatives aimed at controlling or eliminating threats such as AIDS, tuberculosis, and malaria continue to demonstrate success. The ongoing efforts of multinational partnerships to address the substantial burden imposed by neglected tropical diseases have generated new effective antiparasitic agents. In addition, agents approved for other uses are being re-evaluated for antiparasitic efficacy, and some have been subsequently repurposed.

This chapter deals exclusively with the agents used to treat infections due to parasites. Specific treatment recommendations for the parasitic diseases of humans are listed in subsequent chapters. Many of the agents discussed herein are approved by the U.S. Food and Drug Administration (FDA) but are considered investigational for the treatment of certain infections. Drugs marked in the text with an asterisk (*) are available through the Centers for Disease Control and Prevention (CDC) Drug Service (telephone: 404-639-3670; email: drugservice@cdc.gov). Drugs marked with a dagger (†) are available only through their manufacturers; contact information for these manufacturers may be available from the CDC.

Table 229-1 presents a brief overview of each agent (including some drugs that are covered in other chapters), along with major toxicities, spectrum of activity, and safety for use during pregnancy and lactation.

Albendazole Like all benzimidazoles, albendazole acts by selectively binding to free β-tubulin in nematodes, inhibiting the polymerization of tubulin and the microtubule-dependent uptake of glucose. Irreversible damage occurs in gastrointestinal (GI) cells of the nematodes, resulting in starvation, death, and expulsion by the host. This fundamental disruption of cellular metabolism offers treatment for a wide range of parasitic diseases.

Albendazole is poorly absorbed from the GI tract, a feature that is advantageous for the treatment of intestinal helminths but not for that of tissue helminth infections (e.g., hydatid disease and neurocysticercosis), which requires a sufficient amount of active drug to reach the site of infection. Administration with a high-fat meal (~40 g) increases the drug's absorption by up to fivefold. The metabolite albendazole sulfoxide is responsible for the drug's therapeutic effect outside the gut lumen. Albendazole sulfoxide crosses the blood-brain barrier, reaching a level significantly higher than that achieved in plasma. The high concentrations of albendazole sulfoxide attained in cerebrospinal fluid (CSF) may explain the efficacy of albendazole in the treatment of neurocysticercosis.

Albendazole is extensively metabolized in the liver, but there are few data regarding the drug's use in patients with hepatic disease. Single-dose albendazole therapy in humans is largely without side effects (overall frequency, ≤1%). More prolonged courses (e.g., as administered for cystic and alveolar echinococcal disease) have been associated with liver function abnormalities and bone marrow toxicity. Thus, when prolonged use is anticipated, the drug should be administered in treatment cycles of 28 days interrupted by 14-day intervals off therapy. Prolonged therapy with full-dose albendazole (800 mg/d) should be approached cautiously in patients also receiving drugs with known effects on the cytochrome P450 system.

Amodiaquine Amodiaquine has been widely used in the treatment of malaria for >60 years. Like chloroquine (the other major 4-aminoquinoline), amodiaquine is now of limited use because of the spread of resistance. Amodiaquine interferes with hemozoin formation through complexation with heme. It is rapidly absorbed and acts as a prodrug after oral administration; the principal plasma metabolite, monodesethylamodiaquine, is the predominant antimalarial agent. Amodiaquine

TABLE 229-1 Overview of Agents Used for the Treatment of Parasitic Infections

DRUGS BY CLASS	PARASITIC INFECTION(S)	ADVERSE EFFECTS	MAJOR DRUG–DRUG INTERACTIONS	PREGNANCY CLASS[a]	BREAST MILK
4-Aminoquinolines					
Amodiaquine	Malaria[b]	Agranulocytosis, hepatotoxicity	No information	Not assigned	Yes[c]
Chloroquine	Malaria[b]	*Occasional:* pruritus, nausea, vomiting, headache, hair depigmentation, exfoliative dermatitis, reversible corneal opacity *Rare:* irreversible retinal injury, nail discoloration, blood dyscrasias	Antacids and kaolin: reduced absorption of chloroquine Ampicillin: bioavailability reduced by chloroquine Cimetidine: increased serum levels of chloroquine Cyclosporine: serum levels increased by chloroquine	Not assigned[d]	Yes[c]
Piperaquine	Malaria[b]	*Occasional:* GI disturbances	None reported	Not assigned	Yes
8-Aminoquinolines					
Primaquine	Malaria[b]	*Frequent:* hemolysis in patients with G6PD deficiency *Occasional:* methemoglobinemia, GI disturbances *Rare:* CNS symptoms	Quinacrine: potentiated toxicity of primaquine	Contraindicated	Yes
Tafenoquine	Malaria[b]	*Frequent:* hemolysis in patients with G6PD deficiency, mild GI upset *Occasional:* methemoglobinemia, headache	No information	Not assigned	Yes
Aminoalcohols					
Halofantrine	Malaria[b]	*Frequent:* abdominal pain, diarrhea *Occasional:* ECG disturbances (dose-related prolongation of QTc and PR interval), nausea, pruritus; contraindicated in persons who have cardiac disease or who have taken mefloquine in the preceding 3 weeks	Concomitant use of agents that prolong QTc interval contraindicated	C	No information
Lumefantrine	Malaria[b]	*Occasional:* nausea, vomiting, diarrhea, abdominal pain, anorexia, headache, dizziness	Plasma levels increased by darunavir and nevirapine, decreased by etravirine	Not assigned	No information
Aminoglycosides					
Paromomycin	Amebiasis,[b] infection with *Dientamoeba fragilis*, giardiasis, cryptosporidiosis, leishmaniasis	*Frequent:* GI disturbances (oral dosing only) *Occasional:* nephrotoxicity, ototoxicity, vestibular toxicity (parenteral dosing only)	No major interactions	Oral: B Parenteral: not assigned[d]	No information
Amphotericin B	Leishmaniasis,[e] amebic meningoencephalitis	*Frequent:* fever, chills, hypokalemia, hypomagnesemia, nephrotoxicity *Occasional:* vomiting, dyspnea, hypotension	Antineoplastic agents: renal toxicity, bronchospasm, hypotension Glucocorticoids, ACTH, digitalis: hypokalemia Zidovudine: increased myelo- and nephrotoxicity	B	No information
Amphotericin B deoxycholate					
Amphotec (InterMune)					
Amphotericin B lipid complex, ABLC (Abelcet)					
Amphotericin B, liposomal (AmBisome)					
Antimonials	Leishmaniasis	*Frequent:* arthralgias/myalgias, pancreatitis, ECG changes (QT prolongation, T-wave flattening or inversion)			
Meglumine antimoniate[h]			Antiarrhythmics and tricyclic antidepressants: increased risk of cardiotoxicity	Not assigned	No information
Artemisinin and derivatives	Malaria[g]	*Occasional:* neurotoxicity (ataxia, convulsions), nausea, vomiting, anorexia, contact dermatitis			
Arteether			No information	Not assigned	Yes[c]
Artemether			Artemether levels decreased by darunavir, etravirine, and nevirapine	C	Yes[c]
Artesunate[h]			Mefloquine: levels decreased and clearance accelerated by artesunate	C	Yes[c]
Dihydroartemisinin			Mefloquine: increased absorption	Not assigned	Yes[c]

(*Continued*)

TABLE 229-1 Overview of Agents Used for the Treatment of Parasitic Infections (Continued)

DRUGS BY CLASS	PARASITIC INFECTION(S)	ADVERSE EFFECTS	MAJOR DRUG–DRUG INTERACTIONS	PREGNANCY CLASS[a]	BREAST MILK
Atovaquone	Malaria,[b] babesiosis	*Frequent:* nausea, vomiting *Occasional:* abdominal pain, headache	Plasma levels decreased by rifampin, tetracycline, atazanavir, efavirenz, lopinavir/ritonavir; bioavailability decreased by metoclopramide	C	No information
Azoles Fluconazole Itraconazole Ketoconazole	Leishmaniasis	*Serious:* hepatotoxicity *Rare:* exfoliative skin disorders, anaphylaxis	Warfarin, oral hypoglycemics, phenytoin, cyclosporine, theophylline, digoxin, dofetilide, quinidine, carbamazepine, rifabutin, busulfan, docetaxel, vinca alkaloids, pimozide, alprazolam, diazepam, midazolam, triazolam, verapamil, atorvastatin, cerivastatin, lovastatin, simvastatin, tacrolimus, sirolimus, indinavir, ritonavir, saquinavir, alfentanil, buspirone, methylprednisolone, trimetrexate: plasma levels increased by azoles Carbamazepine, phenobarbital, phenytoin, isoniazid, rifabutin, rifampin, antacids, H_2-receptor antagonists, proton pump inhibitors, nevirapine: decreased plasma levels of azoles Clarithromycin, erythromycin, indinavir, ritonavir: increased plasma levels of azoles	C	Yes
Benzimidazoles					
Albendazole	Ascariasis, capillariasis, clonorchiasis, cutaneous larva migrans, cysticercosis,[b] echinococcosis,[b] enterobiasis, eosinophilic enterocolitis, gnathostomiasis, hookworm, lymphatic filariasis, microsporidiosis, strongyloidiasis, trichinellosis, trichostrongyliasis, trichuriasis, visceral larva migrans	*Occasional:* nausea, vomiting, abdominal pain, headache, reversible alopecia, elevated aminotransferases *Rare:* leukopenia, rash	Dexamethasone, praziquantel: plasma level of albendazole sulfoxide increased by ~50%	C	Yes[c]
Mebendazole	Ascariasis,[b] capillariasis, eosinophilic enterocolitis, enterobiasis,[b] hookworm,[b] trichinellosis, trichostrongyliasis, trichuriasis,[b] visceral larva migrans	*Occasional:* diarrhea, abdominal pain, elevated aminotransferases *Rare:* agranulocytosis, thrombocytopenia, alopecia	Cimetidine: inhibited mebendazole metabolism	C	No information
Thiabendazole	Strongyloidiasis,[b] cutaneous larva migrans,[b] visceral larva migrans[b]	*Frequent:* anorexia, nausea, vomiting, diarrhea, headache, dizziness, asparagus-like urine odor *Occasional:* drowsiness, giddiness, crystalluria, elevated aminotransferases, psychosis *Rare:* hepatitis, seizures, angioneurotic edema, Stevens-Johnson syndrome, tinnitus	Theophylline: serum levels increased by thiabendazole	C	No information
Triclabendazole	Fascioliasis, paragonimiasis	*Occasional:* abdominal cramps, diarrhea, biliary colic, transient headache	No information	Not assigned	Yes
Benznidazole	Chagas disease	*Frequent:* rash, pruritus, nausea, leukopenia, paresthesias	No major interactions	Not assigned	No information
Clindamycin	Babesiosis, malaria, toxoplasmosis	*Occasional:* pseudomembranous colitis, abdominal pain, diarrhea, nausea/vomiting *Rare:* pruritus, skin rashes	No major interactions	B	Yes[c]
Diloxanide furoate	Amebiasis	*Frequent:* flatulence *Occasional:* nausea, vomiting, diarrhea *Rare:* pruritus	None reported	Contraindicated	No information

(*Continued*)

TABLE 229-1 Overview of Agents Used for the Treatment of Parasitic Infections (Continued)

DRUGS BY CLASS	PARASITIC INFECTION(S)	ADVERSE EFFECTS	MAJOR DRUG–DRUG INTERACTIONS	PREGNANCY CLASS[a]	BREAST MILK
Eflornithine[f] (difluoromethylornithine, DFMO)	Trypanosomiasis	*Frequent:* pancytopenia *Occasional:* diarrhea, seizures *Rare:* transient hearing loss	No major interactions	Contraindicated	No information
Emetine and dehydroemetine	Amebiasis, fascioliasis	*Severe:* cardiotoxicity *Frequent:* pain at injection site *Occasional:* dizziness, headache, GI symptoms	None reported	X	No information
Folate antagonists					
Dihydrofolate reductase inhibitors					
Pyrimethamine	Malaria,[b] isosporiasis, toxoplasmosis[b]	*Occasional:* folate deficiency *Rare:* rash, seizures, severe skin reactions (toxic epidermal necrolysis, erythema multiforme, Stevens-Johnson syndrome)	Sulfonamides, proguanil, zidovudine: increased risk of bone marrow suppression when used concomitantly	C	Yes
Proguanil and chlorproguanil	Malaria	*Occasional:* urticaria *Rare:* hematuria, GI disturbances	Atazanavir, efavirenz, lopinavir/ritonavir: plasma levels of proguanil decreased	C	Yes
Trimethoprim	Cyclosporiasis, isosporiasis	Hyperkalemia, GI upset, mild stomatitis	Methotrexate: reduced clearance Warfarin: effect prolonged Phenytoin: hepatic metabolism increased	C	Yes
Dihydropteroate synthetase inhibitors: sulfonamides Sulfadiazine Sulfamethoxazole Sulfadoxine	Malaria,[b] toxoplasmosis[b]	*Frequent:* GI disturbances, allergic skin reactions, crystalluria *Rare:* severe skin reactions (toxic epidermal necrolysis, erythema multiforme, Stevens-Johnson syndrome), agranulocytosis, aplastic anemia, hypersensitivity of the respiratory tract, hepatitis, interstitial nephritis, hypoglycemia, aseptic meningitis	Thiazide diuretics: increased risk of thrombocytopenia in elderly patients Warfarin: effect prolonged by sulfonamides Methotrexate: levels increased by sulfonamides Phenytoin: metabolism impaired by sulfonamides Sulfonylureas: effect prolonged by sulfonamides	B	Yes
Dihydropteroate synthetase inhibitors: sulfones					
Dapsone	Leishmaniasis, malaria, toxoplasmosis	*Frequent:* rash, anorexia *Occasional:* hemolysis, methemoglobinemia, neuropathy, allergic dermatitis, anorexia, nausea, vomiting, tachycardia, headache, insomnia, psychosis, hepatitis *Rare:* agranulocytosis	Rifampin: lowered plasma levels of dapsone	C	Yes
Fumagillin	Microsporidiosis	*Rare:* neutropenia, thrombocytopenia	None reported	No information	No information
Furazolidone	Giardiasis	*Frequent:* nausea/vomiting, brown urine *Occasional:* rectal itching, headache *Rare:* hemolytic anemia, disulfiram-like reactions, MAO inhibitor interactions	Risk of hypertensive crisis when administered for >5 days with MAO inhibitors	C	No information
Iodoquinol	Amebiasis,[b] balantidiasis, *D. fragilis* infection	*Occasional:* headache, rash, pruritus, thyrotoxicosis, nausea, vomiting, abdominal pain, diarrhea *Rare:* optic neuritis, peripheral neuropathy, seizures, encephalopathy	No major interactions	C	No information
Lactones					
Ivermectin	Ascariasis, cutaneous larva migrans, gnathostomiasis, loiasis, lymphatic filariasis, onchocerciasis,[b] scabies, strongyloidiasis,[b] trichuriasis	*Occasional:* fever, pruritus, headache, myalgias *Rare:* hypotension	No major interactions	C	Yes[c]
Moxidectin	Onchocerciasis	*Occasional:* fever, pruritus, headache, myalgias *Rare:* orthostatic hypotension, elevated transaminases	No major interactions	C	Yes[c]

(*Continued*)

TABLE 229-1 Overview of Agents Used for the Treatment of Parasitic Infections (Continued)

DRUGS BY CLASS	PARASITIC INFECTION(S)	ADVERSE EFFECTS	MAJOR DRUG–DRUG INTERACTIONS	PREGNANCY CLASS[a]	BREAST MILK
Macrolides					
Azithromycin	Babesiosis	*Occasional:* nausea, vomiting, diarrhea, abdominal pain *Rare:* angioedema, cholestatic jaundice	Cyclosporine and digoxin: levels increased by azithromycin Nelfinavir: increased levels of azithromycin	B	Yes
Spiramycin[h]	Toxoplasmosis	*Occasional:* GI disturbances, transient skin eruptions *Rare:* thrombocytopenia, QT prolongation in an infant, cholestatic hepatitis	No major interactions	Not assigned[d]	Yes[c]
Mefloquine	Malaria[b]	*Frequent:* lightheadedness, nausea, headache *Occasional:* confusion; nightmares; insomnia; visual disturbance; transient and clinically silent ECG abnormalities, including sinus bradycardia, sinus arrhythmia, first-degree AV block, prolongation of QTc interval, and abnormal T waves *Rare:* psychosis, convulsions, hypotension	Administration of halofantrine <3 weeks after mefloquine use may produce fatal QTc prolongation. Mefloquine may lower plasma levels of anticonvulsants. Levels are decreased and clearance is accelerated by artesunate. Mefloquine decreases plasma levels of ritonavir and possibly other protease inhibitors.	C	Yes
Melarsoprol[f]	Trypanosomiasis	*Frequent:* myocardial injury, encephalopathy, peripheral neuropathy, hypertension *Occasional:* G6PD-induced hemolysis, erythema nodosum leprosum *Rare:* hypotension	No major interactions	Not assigned	No information
Metrifonate	Schistosomiasis	*Frequent:* abdominal pain, nausea, vomiting, diarrhea, headache, vertigo, bronchospasm *Rare:* cholinergic symptoms	No major interactions	B	No
Miltefosine	Leishmaniasis,[b] primary amebic meningoencephalitis	*Frequent:* mild and transient (1–2 days) GI disturbances within first 2 weeks of therapy (resolve after treatment completion); motion sickness *Occasional:* reversible elevations of creatinine and aminotransferases	No major interactions	Not assigned	No information
Niclosamide	Intestinal cestode infections[b]	*Occasional:* nausea, vomiting, dizziness, pruritus	No major interactions	B	No information
Nifurtimox	Chagas disease	*Frequent:* nausea, vomiting, abdominal pain, insomnia, paresthesias, weakness, tremors *Rare:* seizures (all reversible and dose-related)	No major interactions	Not assigned	No information
Nitazoxanide	Cryptosporidiosis,[b] giardiasis[b]	*Occasional:* abdominal pain, diarrhea *Rare:* vomiting, headache	Increases plasma levels of highly protein-bound drugs (e.g., phenytoin, warfarin)	B	No information
Nitroimidazoles					
Metronidazole	Amebiasis,[b] balantidiasis, dracunculiasis, giardiasis [b], trichomoniasis,[b] *D. fragilis* infection	*Frequent:* nausea, headache, anorexia, metallic aftertaste *Occasional:* vomiting, insomnia, vertigo, paresthesias, disulfiram-like effects *Rare:* seizures, peripheral neuropathy	Warfarin: effect enhanced by metronidazole Disulfiram: psychotic reaction Phenobarbital, phenytoin: accelerate elimination of metronidazole Lithium: serum levels elevated by metronidazole Cimetidine: prolonged half-life of metronidazole Oral solutions of antiretrovirals containing alcohol: disulfiram effect due to alcohol	B	Yes
Tinidazole	Amebiasis,[b] giardiasis, trichomoniasis	*Occasional:* nausea, vomiting, metallic taste	*See* metronidazole	C	Yes

(*Continued*)

TABLE 229-1 Overview of Agents Used for the Treatment of Parasitic Infections (Continued)

DRUGS BY CLASS	PARASITIC INFECTION(S)	ADVERSE EFFECTS	MAJOR DRUG–DRUG INTERACTIONS	PREGNANCY CLASS[a]	BREAST MILK
Oxamniquine	Schistosomiasis	*Occasional:* dizziness, drowsiness, headache, orange urine, elevated aminotransferases *Rare:* seizures	No major interactions	C	No information
Pentamidine isethionate	Leishmaniasis, trypanosomiasis	*Frequent:* hypotension, hypoglycemia, pancreatitis, sterile abscesses at IM injection sites, GI disturbances, reversible renal failure *Occasional:* hepatotoxicity, cardiotoxicity, delirium *Rare:* anaphylaxis	No major interactions	C	No information
Piperazine and derivatives					
Piperazine	Ascariasis, enterobiasis	*Occasional:* nausea, vomiting, diarrhea, abdominal pain, headache *Rare:* neurotoxicity, seizures	None reported	C	No information
Diethylcarbamazine[f]	Lymphatic filariasis, loiasis, tropical pulmonary eosinophilia	*Frequent:* dose-related nausea, vomiting *Rare:* fever, chills, arthralgias, headache	None reported	Not assigned[d]	No information
Praziquantel	Clonorchiasis,[b] cysticercosis, diphyllobothriasis, hymenolepiasis, taeniasis, opisthorchiasis, intestinal trematodes, paragonimiasis, schistosomiasis[b]	*Frequent:* abdominal pain, diarrhea, dizziness, headache, malaise *Occasional:* fever, nausea *Rare:* pruritus, singultus	No major interactions	B	Yes
Pyrantel pamoate	Ascariasis, eosinophilic enterocolitis, enterobiasis,[b] hookworm, trichostrongyliasis	*Occasional:* GI disturbances, headache, dizziness, elevated aminotransferases	No major interactions	C	No information
Pyronaridine	Malaria	*Occasional:* headache, nausea	None reported to date	B	Yes
Quinacrine[h]	Giardiasis[b]	*Frequent:* headache, nausea, vomiting, bitter taste *Occasional:* yellow-orange discoloration of skin, sclerae, urine; begins after 1 week of treatment and lasts up to 4 months after drug discontinuation *Rare:* psychosis, exfoliative dermatitis, retinopathy, G6PD-induced hemolysis, exacerbation of psoriasis, disulfiram-like effects	Primaquine: toxicity potentiated by quinacrine	C	No information
Quinine and quinidine	Malaria, babesiosis	*Frequent:* cinchonism (tinnitus, high-tone deafness, headache, dysphoria, nausea, vomiting, abdominal pain, visual disturbances, postural hypotension), hyperinsulinemia resulting in life-threatening hypoglycemia *Occasional:* deafness, hemolytic anemia, arrhythmias, hypotension due to rapid IV infusion	Carbonic anhydrase inhibitors, thiazide diuretics: reduced renal elimination of quinidine Amiodarone, cimetidine: increased quinidine levels Nifedipine: decreased quinidine levels; quinidine slows metabolism of nifedipine Phenobarbital, phenytoin, rifampin: accelerated hepatic elimination of quinidine Verapamil: reduced hepatic clearance of quinidine Diltiazem: decreased clearance of quinidine	X	Yes[c]
Quinolones					
Ciprofloxacin	Cyclosporiasis, isosporiasis	*Occasional:* nausea, diarrhea, vomiting, abdominal pain/discomfort, headache, restlessness, rash *Rare:* myalgias/arthralgias, tendon rupture, CNS symptoms (nervousness, agitation, insomnia, anxiety, nightmares, or paranoia); convulsions	Probenecid: increased serum levels of ciprofloxacin Theophylline, warfarin: serum levels increased by ciprofloxacin	C	Yes

(*Continued*)

TABLE 229-1 Overview of Agents Used for the Treatment of Parasitic Infections (Continued)

DRUGS BY CLASS	PARASITIC INFECTION(S)	ADVERSE EFFECTS	MAJOR DRUG–DRUG INTERACTIONS	PREGNANCY CLASS[a]	BREAST MILK
Suramin[f]	Trypanosomiasis	*Frequent:* immediate: fever, urticaria, nausea, vomiting, hypotension; delayed (up to 24 h): exfoliative dermatitis, stomatitis, paresthesias, photophobia, renal dysfunction *Occasional:* nephrotoxicity, adrenal toxicity, optic atrophy, anaphylaxis	No major interactions	Not assigned	No information
Tetracyclines	Balantidiasis, *D. fragilis* infection, malaria; lymphatic filariasis (doxycycline)	*Frequent:* GI disturbances *Occasional:* photosensitivity dermatitis *Rare:* exfoliative dermatitis, esophagitis, hepatotoxicity	Warfarin: effect prolonged by tetracyclines	D	Yes

[a]Based on U.S. Food and Drug Administration (FDA) pregnancy categories of A–D, X. [b]Approved by the FDA for this indication. [c]Not believed to be harmful. [d]Use in pregnancy is recommended by international organizations outside the United States. [e]Only AmBisome has been approved by the FDA for this indication. [f]Available through the CDC. [g]Only artemether (in combination with lumefantrine) and artesunate have been approved by the FDA for this indication. [h]Available through the manufacturer.

Abbreviations: ACTH, adrenocorticotropic hormone; AV, atrioventricular; CNS, central nervous system; ECG, electrocardiogram; G6PD, glucose 6-phosphate dehydrogenase; GI, gastrointestinal; MAO, monoamine oxidase.

and its metabolites are all excreted in urine, but there are no recommendations concerning dosage adjustment in patients with impaired renal function. Agranulocytosis and hepatotoxicity can develop with repeated use; therefore, this drug should not be used for prophylaxis. Indeed, because of its adverse effects and widespread resistance, amodiaquine is no longer in use in Europe or the United States, and it was dropped from malaria control programs as a single agent by the WHO in 1990; however, it remains effective in some areas when combined with other antimalarial drugs (e.g., artesunate, sulfadoxine-pyrimethamine).

Amphotericin B See Table 229-1 and Chap. 217.

Antimonials† Despite associated adverse reactions and the need for prolonged parenteral treatment, the pentavalent antimonial compounds (designated Sb^{v}) have remained the first-line therapy for all forms of leishmaniasis throughout the world, primarily because they are affordable and effective and have survived the test of time. Pentavalent antimonials are active only after bioreduction to the trivalent Sb(III) form, which inhibits trypanothione reductase, a critical enzyme involved in the oxidative stress management of *Leishmania* species. The fact that *Leishmania* species use trypanothione rather than glutathione (which is used by mammalian cells) may explain the parasite-specific activity of antimonials. The drugs are taken up by the reticuloendothelial system, and their activity against *Leishmania* species may be enhanced by this localization.

Resistance is a major problem in some areas. Although low-level unresponsiveness to Sb^{v} was identified in India in the 1970s, incremental increases in both the recommended daily dosage (to 20 mg/kg) and the duration of treatment (to 28 days) satisfactorily compensated for the growing resistance until around 1990. Since then, the capacity of Sb^{v} to induce long-term cure in patients with kala-azar who live in eastern India has steadily eroded. Co-infection with HIV impairs the treatment response.

Pentavalent antimonials are available in aqueous solution and are administered parenterally. Antimony appears to have two elimination phases. When the drug is administered IV, the mean half-life of the first phase is <2 h; the mean half-life of the terminal elimination phase is nearly 36 h. This slower phase may be due to conversion of pentavalent antimony to a trivalent form that is the likely cause of the side effects often seen with prolonged therapy.

In 2020, the global manufacturer of sodium stibogluconate notified the CDC that the product would be discontinued due to the inability to source the necessary raw materials. As a result, CDC ended its distribution program following expiration of the last lot in use. However, meglumine antimoniate is available from the manufacturer through the FDA (301-796-1400).

Artemisinin Derivatives* Artesunate, artemether, artemotil, and the parent compound artemisinin are sesquiterpene lactones derived from the wormwood plant *Artemisia annua*. These agents are at least 10-fold more potent in vivo than other antimalarial drugs and presently show no cross-resistance with known antimalarial drugs; thus, they have become first-line agents for the treatment of severe falciparum malaria. The artemisinin compounds are rapidly effective against the asexual blood forms of *Plasmodium* species but are not active against intrahepatic forms. With the exception of artesunate, artemisinin and its derivatives are highly lipid soluble and readily cross both host and parasite cell membranes. One factor that explains the drugs' highly selective toxicity against malaria is that parasitized erythrocytes concentrate artemisinin and its derivatives to concentrations 100-fold higher than those in uninfected erythrocytes. The antimalarial effect of these agents results primarily from the active metabolite dihydroartemisinin; in the presence of heme or molecular iron, the endoperoxide moiety of dihydroartemisinin decomposes, generating free radicals and other metabolites that damage parasite proteins. The compounds are available for oral, rectal, IV, or IM administration, depending on the derivative. Following FDA approval in May 2020, artesunate became commercially available in the United States.

Artemisinin and its derivatives are cleared rapidly from the circulation. Their short half-lives limit their value for prophylaxis and monotherapy. Side effects appear to be minor, although sinus bradycardia and transient first-degree heart block have been reported. Although seen in animal models, embryotoxicity and neurotoxicity have not been identified in humans despite active investigation. These agents should be used only in combination with another, longer-acting agent (e.g., artesunate-mefloquine, dihydroartemisinin-piperaquine). A combined formulation of artemether and lumefantrine is widely available for the treatment of acute uncomplicated falciparum malaria acquired in areas where *Plasmodium falciparum* is resistant to chloroquine and antifolates.

Atovaquone Atovaquone is a hydroxynaphthoquinone that exerts broad-spectrum antiprotozoal activity via selective inhibition of parasite mitochondrial electron transport. This agent exhibits potent activity against toxoplasmosis and babesiosis when used with pyrimethamine and azithromycin, respectively. Atovaquone possesses a novel mode of action against *Plasmodium* species, inhibiting the electron transport system at the level of the cytochrome bc1 complex. The drug is active against both the erythrocytic and the exoerythrocytic stages of *Plasmodium* species; however, because it does not eradicate hypnozoites from the liver, patients with *P. vivax* or *P. ovale* infections must be given radical prophylaxis.

Malarone is a fixed-dose combination of atovaquone and proguanil used for malaria prophylaxis as well as for the treatment of acute, uncomplicated *P. falciparum* malaria. Malarone has been shown to be effective in regions with multidrug-resistant *P. falciparum*. Resistance to atovaquone develops rapidly via mutations in the parasite's mitochondrial cytochrome b complex. However, the mutations result in sterility of female parasites; thus, atovaquone-resistant parasites cannot be transmitted to another person. This situation may explain why clinical resistance has yet to be reported.

The bioavailability of atovaquone varies considerably. Absorption after a single oral dose is slow, increases two- to threefold with a fatty meal, and is dose-limited above 750 mg. The elimination half-life is increased in patients with moderate hepatic impairment. Because of the potential for drug accumulation, the use of atovaquone is generally contraindicated in persons with a creatinine clearance rate <30 mL/min. No dosage adjustments are needed in patients with mild to moderate renal impairment.

Azithromycin See Table 229-1 and Chap. 149.

Azoles See Table 229-1 and Chap. 217.

Benznidazole Introduced in 1971, this oral nitroimidazole derivative is used to treat Chagas disease, with cure rates of 80–90% recorded in acute infections. Benznidazole is believed to exert its trypanocidal effects by generating oxygen radicals to which the parasites are more sensitive than mammalian cells because of a relative deficiency in antioxidant enzymes. Benznidazole also appears to alter the balance between pro- and anti-inflammatory mediators by downregulating the synthesis of nitrite, interleukin (IL) 6, and IL-10 in macrophages. Benznidazole is highly lipophilic and readily absorbed. The drug is extensively metabolized; only 5% of the dose is excreted unchanged in the urine. Benznidazole is well tolerated; adverse effects are rare and usually manifest as GI upset or pruritic rash. Following FDA approval in 2018, this drug is now commercially available in the United States.

Chloroquine This 4-aminoquinoline has marked, rapid schizonticidal and gametocidal activity against blood forms of *P. ovale* and *P. malariae* and against susceptible strains of *P. vivax* and *P. falciparum*. It is not active against intrahepatic forms (*P. vivax* and *P. ovale*). Parasitized erythrocytes accumulate chloroquine in significantly greater concentrations than do normal erythrocytes. Chloroquine, a weak base, concentrates in the food vacuoles of intraerythrocytic parasites because of a relative pH gradient between the extracellular space and the acidic food vacuole. Once it enters the acidic food vacuole, chloroquine is rapidly converted to a membrane-impermeable protonated form and is trapped. Continued accumulation of chloroquine in the parasite's acidic food vacuoles results in drug levels that are 600-fold higher at this site than in plasma. The high accumulation of chloroquine results in an increase in pH within the food vacuole to a level above that required for the acid proteases' optimal activity, inhibiting parasite heme polymerase; as a result, the parasite is effectively killed with its own metabolic waste. Compared with susceptible strains, chloroquine-resistant plasmodia transport chloroquine out of intraparasitic compartments more rapidly and maintain lower chloroquine concentrations in their acid vesicles. Hydroxychloroquine, a congener of chloroquine, is equivalent to chloroquine in its antimalarial efficacy but is preferred to chloroquine for the treatment of autoimmune disorders because it produces less ocular toxicity when used in high doses.

Chloroquine is well absorbed. However, because it exhibits extensive tissue binding, a loading dose is required to yield effective plasma concentrations. A therapeutic drug level in plasma is reached 2–3 h after oral administration (the preferred route). Chloroquine can be administered IV, but excessively rapid parenteral administration can result in seizures and death from cardiovascular collapse. The mean half-life of chloroquine is 4 days, but the rate of excretion decreases as plasma levels decline, making once-weekly administration possible for prophylaxis in areas with sensitive strains. About one-half of the parent drug is excreted in urine, but the dose should not be reduced for persons with acute malaria and renal insufficiency.

Ciprofloxacin See Table 229-1 and Chap. 149.

Clindamycin See Table 229-1 and Chap. 149.

Dapsone See Table 229-1 and Chap. 188.

Dehydroemetine Emetine is an alkaloid derived from ipecac; dehydroemetine is synthetically derived from emetine and is considered less toxic. Both agents are active against *Entamoeba histolytica* and appear to work by blocking peptide elongation and thus inhibiting protein synthesis. Emetine is rapidly absorbed after parenteral administration, rapidly distributed throughout the body, and slowly excreted in the urine in unchanged form. Both agents are contraindicated in patients with renal disease.

Diethylcarbamazine* A derivative of the antihelminthic agent piperazine with a long history of successful use, diethylcarbamazine (DEC) remains the treatment of choice for lymphatic filariasis and loiasis and has also been used for visceral larva migrans. Although piperazine itself has no antifilarial activity, the piperazine ring of DEC is essential for the drug's activity. DEC's mechanism of action remains to be fully defined. Proposed mechanisms include immobilization due to inhibition of parasite cholinergic muscle receptors, disruption of microtubule formation, and alteration of helminthic surface membranes resulting in enhanced killing by the host's immune system. DEC enhances adherence properties of eosinophils. The development of resistance under drug pressure (i.e., a progressive decrease in efficacy when the drug is used widely in human populations) has not been observed, although DEC has variable effects when administered to persons with filariasis. Monthly administration provides effective prophylaxis against both bancroftian filariasis and loiasis.

DEC is well absorbed after oral administration, with peak plasma concentrations reached within 1–2 h. No parenteral form is available. The drug is eliminated largely by renal excretion, with <5% found in feces. If more than one dose is to be administered to an individual with renal dysfunction, the dose should be reduced commensurate with the reduction in creatinine clearance rate. Alkalinization of the urine prevents renal excretion and increases the half-life of DEC. Use in patients with onchocerciasis can precipitate a Mazzotti reaction, with pruritus, fever, and arthralgias. Like other piperazines, DEC is active against *Ascaris* species. Patients co-infected with this nematode may expel live worms after treatment.

Diloxanide Furoate Diloxanide furoate, a substituted acetanilide, is a luminally active agent used to eradicate the cysts of *E. histolytica*. After ingestion, diloxanide furoate is hydrolyzed by enzymes in the lumen or mucosa of the intestine, releasing furoic acid and the ester diloxanide; the latter acts directly as an amebicide.

Diloxanide furoate is given alone to asymptomatic cyst passers. For patients with active amebic infections, diloxanide is generally administered in combination with a 5-nitroimidazole such as metronidazole or tinidazole. Diloxanide furoate is rapidly absorbed after oral administration. When coadministered with a 5-nitroimidazole, diloxanide levels peak within 1 h and disappear within 6 h. About 90% of an oral dose is excreted in the urine within 48 h, chiefly as the glucuronide metabolite. Diloxanide furoate is contraindicated in pregnant and breast-feeding women and in children <2 years of age.

Eflornithine* Eflornithine (difluoromethylornithine, or DFMO) is a fluorinated analogue of the amino acid ornithine. Although originally designed as an antineoplastic agent, eflornithine has proven effective against some trypanosomatids.

Eflornithine has specific activity against all stages of infection with *Trypanosoma brucei gambiense*; however, it is inactive against *T. b. rhodesiense*. The drug acts as an irreversible suicide inhibitor of ornithine decarboxylase, the first enzyme in the biosynthesis of the polyamines putrescine and spermidine. Polyamines are essential for the synthesis of trypanothione, an enzyme required for the maintenance of intracellular thiols in the correct redox state and for the removal of reactive oxygen metabolites. However, polyamines are also essential for cell division in eukaryotes, and ornithine decarboxylase is similar in trypanosomes and mammals. The selective antiparasitic activity of eflornithine is partly explained by the structure of the trypanosomal enzyme, which lacks a 36-amino-acid C-terminal sequence found on mammalian ornithine decarboxylase. This difference results in a lower turnover of ornithine decarboxylase and a more rapid decrease of polyamines in trypanosomes than in the mammalian host. The diminished effectiveness of eflornithine against *T. b. rhodesiense* appears to be due to the parasite's ability to replace the inhibited enzyme more rapidly than *T. b. gambiense*.

Eflornithine is less toxic but more costly than conventional therapy. It can be administered IV or PO. The dose should be reduced in renal failure. Eflornithine readily crosses the blood-brain barrier; CSF levels are highest in persons with the most severe central nervous system (CNS) involvement.

Flubendazole This agent, a methylcarbamate benzimidazole, is highly active against a broad spectrum of gut nematodes and filaria. Its antihelminthic effect is similar to other benzimidazoles; however, it is also an effective inducer of reactive oxygen species and disrupts glucose transport and metabolism. It has limited solubility in water, so bioavailability is very low, but absorption increases if taken after meals, and is further enhanced when reconstituted as a cyclodextrin-based formulation. It is approved for human use in Europe but not in the Unites States.

Fosmidomycin Originally developed in the 1970s as an antibiotic, this *Streptomyces*-derived agent is an aminopropylphosphonic acid that exhibits antimalarial activity through inhibition of 1-deoxy-D-xylulose 5-phosphate reductoisomerase, an essential enzyme on the non-mevalonate pathway of isoprenoid biosynthesis, a pathway that is common in most eukaryotes but absent in humans. For >15 years, fosmidomycin has been evaluated alone and in combination with other antimalarials, and it has demonstrated both safety and efficacy. However, the drug is highly charged—resulting in a short plasma half-life (3.5 h)—causing high rates of recrudescent malaria, particularly in children. It is not currently available, but clinical studies are ongoing.

Fumagillin† Originally discovered as an antiangiogenic compound derived from the fungus *Aspergillus fumigatus*, fumagillin is a water-insoluble antibiotic that is active against microsporidia and is used topically to treat ocular infections due to *Encephalitozoon* species. When given systemically, fumagillin was effective but caused thrombocytopenia in all recipients in the second week of treatment; this side effect was readily reversed when administration of the drug was stopped. Fumagillin acts by binding to methionine aminopeptidase 2, thus inhibiting microsporidial replication by irreversibly blocking the active site.

Furazolidone This nitrofuran derivative is an effective alternative agent for the treatment of giardiasis and also exhibits activity against *Isospora belli*. Because it is the only agent active against *Giardia* that is available in liquid form, it is most often used to treat young children. Furazolidone undergoes reductive activation in *Giardia lamblia* trophozoites—an event that, unlike the reductive activation of metronidazole, involves an NADH oxidase. The killing effect correlates with the toxicity of reduced products, which damage important cellular components, including DNA. Although furazolidone had been thought to be largely unabsorbed when administered orally, the occurrence of systemic adverse reactions indicates that this is not the case. More than 65% of the drug dose can be recovered from the urine as colored metabolites. Omeprazole reduces the oral bioavailability of furazolidone.

Furazolidone is a monoamine oxidase (MAO) inhibitor; thus, caution should be used in its concomitant administration with other drugs (especially indirectly acting sympathomimetic amines) and in the consumption of food and drink containing tyramine during treatment. However, hypertensive crises have not been reported in patients receiving furazolidone, and it has been suggested that—because furazolidone inhibits MAOs gradually over several days—the risks are small if treatment is limited to a 5-day course. Because hemolytic anemia can occur in patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency and glutathione instability, furazolidone treatment is contraindicated in mothers who are breast-feeding and in neonates.

Halofantrine This 9-phenanthrenemethanol is one of three classes of arylaminoalcohols first identified as potential antimalarial agents by the World War II Malaria Chemotherapy Program. Its activity is believed to be similar to that of chloroquine, although it is an oral alternative for the treatment of malaria due to chloroquine-resistant *P. falciparum*.

Halofantrine is thought to share one or more mechanisms with the 4-aminoquinolines, forming a complex with ferriprotoporphyrin IX and interfering with the degradation of hemoglobin. It has been shown to bind to plasmepsin, a hemoglobin-degrading enzyme unique to plasmodia.

Halofantrine exhibits erratic bioavailability, but its absorption is significantly enhanced when it is taken with a fatty meal. The elimination half-life of halofantrine is 1–2 days; it is excreted mainly in feces. Halofantrine is metabolized into *N*-debutyl-halofantrine by the cytochrome P450 enzyme CYP3A4. Grapefruit juice should be avoided during treatment because it increases both halofantrine's bioavailability and halofantrine-induced QT interval prolongation by inhibiting CYP3A4 at the enterocyte level. Halofantrine should not be given simultaneously with or <3 weeks after mefloquine because of the potential occurrence of a fatal prolongation of the QTc interval on electrocardiography.

Iodoquinol Iodoquinol (diiodohydroxyquin), a hydroxyquinoline, is an effective luminal agent for the treatment of amebiasis, balantidiasis, and infection with *Dientamoeba fragilis*. Its mechanism of action is unknown. It is poorly absorbed. Because the drug contains 64% organically bound iodine, it should be used with caution in patients with thyroid disease. Iodine dermatitis occurs occasionally during iodoquinol treatment. Protein-bound serum iodine levels may be increased during treatment and can interfere with certain tests of thyroid function. These effects may persist for as long as 6 months after discontinuation of therapy. Iodoquinol is contraindicated in patients with liver disease. Most serious are the reactions related to prolonged high-dose therapy (optic neuritis, peripheral neuropathy), which should not occur if the recommended dosage regimens are followed.

Ivermectin Ivermectin (22,23-dihydroavermectin) is a derivative of the macrocyclic lactone avermectin produced by the soil-dwelling actinomycete *Streptomyces avermitilis*. Ivermectin is active at low doses against a wide range of helminths and ectoparasites. It is the drug of choice for the treatment of onchocerciasis, strongyloidiasis, cutaneous larva migrans, and scabies. Ivermectin is highly active against microfilariae of the lymphatic filariases but has no macrofilaricidal activity. When ivermectin is used in combination with other agents such as DEC or albendazole for treatment of lymphatic filariasis, synergistic activity is seen. Although active against the intestinal helminths *Ascaris lumbricoides* and *Enterobius vermicularis*, ivermectin is only variably effective in trichuriasis and is ineffective against hookworms. Widespread use of ivermectin for treatment of intestinal nematode infections in sheep and goats has led to the emergence of drug resistance in veterinary practice; this development may portend problems in human medical use.

Data suggest that ivermectin acts by opening the neuromuscular membrane-associated, glutamate-dependent chloride channels. The influx of chloride ions results in hyperpolarization and muscle paralysis—particularly of the nematode pharynx, with consequent blockage of the oral ingestion of nutrients. As these chloride channels are present only in invertebrates, paralysis is seen only in the parasite.

Ivermectin is available for administration to humans only as an oral formulation. The drug is highly protein bound; it is almost completely excreted in feces. Both food and beer increase the bioavailability of ivermectin significantly. Ivermectin is distributed widely throughout the body; animal studies indicate that it accumulates at the highest concentration in adipose tissue and liver, with little accumulation in the brain. Few data exist to guide therapy in hosts with conditions that may influence drug pharmacokinetics.

Ivermectin is generally administered as a single dose of 150–200 μg/kg. In the absence of parasitic infection, the adverse effects of ivermectin in therapeutic doses are minimal. Adverse effects in patients with filarial infections include fever, myalgia, malaise, lightheadedness, and (occasionally) postural hypotension. The severity of such side effects is related to the intensity of parasite infection, with more symptoms in individuals with a heavy parasite burden. In onchocerciasis, skin edema, pruritus, and mild eye irritation may also occur. The adverse effects are generally self-limiting and only occasionally require symptom-based treatment with antipyretics or antihistamines. More severe complications of ivermectin therapy for onchocerciasis include encephalopathy in patients heavily infected with *Loa loa*.

Lumefantrine Lumefantrine (benflumetol), a fluorene arylaminoalcohol derivative synthesized in the 1970s by the Chinese Academy of Military Medical Sciences (Beijing), has marked blood schizonticidal activity against a wide range of plasmodia. This agent conforms structurally and in mode of action to other arylaminoalcohols (quinine, mefloquine, and halofantrine). Lumefantrine exerts its antimalarial effect as a consequence of its interaction with heme, a degradation product of hemoglobin metabolism. Although its antimalarial activity is slower than that of the artemisinin-based drugs, the recrudescence rate with the recommended lumefantrine regimen is lower. The pharmacokinetic properties of lumefantrine are reminiscent of those of halofantrine, with variable oral bioavailability, considerable augmentation of oral bioavailability by concomitant fat intake, and a terminal elimination half-life of ~4–5 days in patients with malaria.

Artemether and lumefantrine have synergistic activity, and the combined formulation of artemether and lumefantrine is effective for the treatment of falciparum malaria in areas where *P. falciparum* is resistant to chloroquine and antifolates.

Mebendazole This benzimidazole is a broad-spectrum antiparasitic agent widely used to treat intestinal helminthiases. Its mechanism of action is similar to that of albendazole; however, it is a more potent inhibitor of parasite malic dehydrogenase and exhibits a more specific and selective effect against intestinal nematodes than the other benzimidazoles.

Mebendazole is available only in oral form but is poorly absorbed from the GI tract; only 5–10% of a standard dose is measurable in plasma. The proportion absorbed from the GI tract is extensively metabolized in the liver. Metabolites appear in the urine and bile; impaired liver or biliary function results in higher plasma mebendazole levels in treated patients. No dose reduction is warranted in patients with renal function impairment. Because mebendazole is poorly absorbed, its incidence of side effects is low. Transient abdominal pain and diarrhea sometimes occur, usually in persons with massive parasite burdens.

Mefloquine Mefloquine is used for prophylaxis of chloroquine-resistant malaria; high doses can be used for treatment. Despite the development of drug-resistant strains of *P. falciparum* in parts of Africa and Southeast Asia, mefloquine remains an effective drug throughout most of the world. Cross-resistance of mefloquine with halofantrine and with quinine has been documented in limited areas. Like quinine and chloroquine, this quinoline is active only against the asexual erythrocytic stages of malarial parasites. Unlike quinine, however, mefloquine has a relatively poor affinity for DNA and, as a result, does not inhibit the synthesis of parasitic nucleic acids and proteins. Although both mefloquine and chloroquine inhibit hemozoin formation and heme degradation, mefloquine differs in that it forms a complex with heme that may be toxic to the parasite.

Mefloquine HCl is poorly water soluble and intensely irritating when given parenterally; thus, it is available only in tablet form. Its absorption is adversely affected by vomiting and diarrhea but is significantly enhanced when the drug is administered with or after food. About 98% of the drug binds to protein. Mefloquine is excreted mainly in the bile and feces; therefore, no dose adjustment is needed in persons with renal insufficiency. The drug and its main metabolite are not appreciably removed by hemodialysis. No dosage adjustments are indicated for the achievement of plasma concentrations in dialysis patients. Pharmacokinetic differences have been detected among various ethnic populations; however, these distinctions are of minor importance compared with host immune status and parasite susceptibility. In patients with impaired liver function, the elimination of mefloquine may be prolonged, leading to higher plasma levels.

Mefloquine should be used with caution by individuals participating in activities requiring alertness and fine-motor coordination because dizziness, vertigo, or tinnitus can develop and persist. If the drug is to be administered for a prolonged period, periodic evaluations are recommended, including liver function tests and ophthalmic examinations. Sleep abnormalities (insomnia, abnormal dreams) have occasionally been reported. Psychosis and seizures occur rarely; mefloquine should not be prescribed to patients with neuropsychiatric conditions. The development of acute anxiety, depression, restlessness, or confusion may be considered prodromal to a more serious event, and the drug should be discontinued.

Concomitant use of quinine, quinidine, or drugs producing β-adrenergic blockade may cause significant electrocardiographic abnormalities or cardiac arrest. Halofantrine must not be given simultaneously with or <3 weeks after mefloquine because a potentially fatal prolongation of the QTc interval on electrocardiography may occur. No data exist on mefloquine use after halofantrine use. Administration of mefloquine with quinine or chloroquine may increase the risk of convulsions. Mefloquine may lower plasma levels of anticonvulsants. Caution should be exercised with concomitant antiretroviral therapy, because mefloquine has been shown to exert variable effects on ritonavir pharmacokinetics that are not explained by hepatic CYP3A4 activity or ritonavir protein binding. Vaccinations with attenuated live bacteria should be completed at least 3 days before the first dose of mefloquine.

Women of childbearing age who are traveling to areas where malaria is endemic should be warned against becoming pregnant and encouraged to practice contraception during malaria prophylaxis with mefloquine and for up to 3 months thereafter. However, in the case of unplanned pregnancy, use of mefloquine is not considered an indication for pregnancy termination. Analysis of prospectively monitored cases demonstrates a prevalence of birth defects and fetal loss comparable to background rates.

Melarsoprol* Melarsoprol has been used since 1949 for the treatment of human African trypanosomiasis. This trivalent arsenical compound is indicated for the treatment of African trypanosomiasis with neurologic involvement and for the treatment of early disease that is resistant to suramin or pentamidine. Melarsoprol, like other drugs containing heavy metals, interacts with thiol groups of several different proteins; however, its antiparasitic effects appear to be more specific. Trypanothione reductase is a key enzyme involved in the oxidative stress management of both *Trypanosoma* and *Leishmania* species, helping to maintain an intracellular reducing environment by reduction of disulfide trypanothione to its dithiol derivative dihydrotrypanothione. Melarsoprol sequesters dihydrotrypanothione, depriving the parasite of its main sulfhydryl antioxidant, and inhibits trypanothione reductase, depriving the parasite of the essential enzyme system that is responsible for keeping trypanothione reduced. These effects are synergistic. The selectivity of arsenical action against trypanosomes is due at least in part to the greater melarsoprol affinity of reduced trypanothione than of other monothiols (e.g., cysteine) on which the mammalian host depends for maintenance of high thiol levels. Melarsoprol enters the parasite via an adenosine transporter; drug-resistant strains lack this transport system.

Melarsoprol is always administered IV. A small but therapeutically significant amount of the drug enters the CSF. The compound is excreted rapidly, with ~80% of the arsenic found in feces.

Melarsoprol is highly toxic. The most serious adverse reaction is reactive encephalopathy, which affects 6% of treated individuals and usually develops within 4 days of the start of therapy, with an average case–fatality rate of 50%. Glucocorticoids are administered with melarsoprol to prevent this development. Because melarsoprol is intensely irritating, care must be taken to avoid infiltration of the drug.

Metrifonate Metrifonate has selective activity against *Schistosoma haematobium*. This organophosphorous compound is a prodrug that is converted nonenzymatically to dichlorvos (2,2-dichlorovinyl dimethylphosphate, DDVP), a highly active chemical that irreversibly inhibits the acetylcholinesterase enzyme. Schistosomal cholinesterase is more susceptible to dichlorvos than is the corresponding human enzyme. The exact mechanism of action of metrifonate is uncertain, but the drug is believed to inhibit tegumental acetylcholine receptors that mediate glucose transport.

Metrifonate is administered in a series of three doses at 2-week intervals. After a single oral dose, metrifonate produces a 95% decrease in plasma cholinesterase activity within 6 h, with a fairly rapid return to normal. However, 2.5 months are required for erythrocyte cholinesterase levels to return to normal. Treated persons should not be exposed to neuromuscular blocking agents or organophosphate insecticides for at least 48 h after treatment.

Metronidazole and Other Nitroimidazoles See Table 229-1 and Chap. 149.

Miltefosine In the early 1990s, miltefosine (hexadecylphosphocholine), originally developed as an antineoplastic agent, was discovered to have significant antiproliferative activity against *Leishmania* species, *T. cruzi*, and *T. brucei* parasites in vitro and in experimental animal models. Miltefosine is the first oral drug that has proved to be highly effective and comparable to amphotericin B against visceral leishmaniasis in India, where antimonial-resistant cases are prevalent. Miltefosine is also effective in previously untreated visceral infections. Cure rates in cutaneous leishmaniasis are comparable to those obtained with antimony. Miltefosine is also effective against the free-living ameba *Naegleria fowleri*.

The activity of miltefosine is attributed to interaction with cell signal transduction pathways and inhibition of phospholipid and sterol biosynthesis. Resistance to miltefosine has not been observed clinically. The drug is readily absorbed from the GI tract, is widely distributed, and accumulates in several tissues. The efficacy of a 28-day treatment course in Indian visceral leishmaniasis is equivalent to that of amphotericin B therapy; however, it appears that a shortened course of 21 days may be equally efficacious.

General recommendations for the use of miltefosine are limited by the exclusion of specific groups from the published clinical trials: persons <12 or >65 years of age, persons with the most advanced disease, breast-feeding women, HIV-infected patients, and individuals with significant renal or hepatic insufficiency.

Moxidectin Like ivermectin, moxidectin is a macrocyclic lactone that is an effective antihelminthic. In 2018, the FDA approved its use for the treatment of onchocerciasis. The primary mode of action of moxidectin is believed to be like that of ivermectin; however, there are likely different binding sites, as suggested by the identification of ivermectin-resistant helminths that are susceptible to moxidectin. The drug is well tolerated, with most adverse effects attributed to death of microfilariae. Some adverse effects occurred more commonly compared with ivermectin, including orthostatic hypotension (5 vs 2%) and elevated transaminases (1 vs 0.6%). In clinical trials, no clinically significant differences in the pharmacokinetics were observed with age, gender, weight, or renal impairment. The effect of hepatic dysfunction is unknown.

Niclosamide† Niclosamide is active against a wide variety of adult tapeworms but not against tissue cestodes. The drug uncouples oxidative phosphorylation in parasite mitochondria, thereby blocking the uptake of glucose by the intestinal tapeworm and resulting in the parasite's death. Niclosamide rapidly causes spastic paralysis of intestinal cestodes in vitro. Its use is limited by its side effects, the necessarily long duration of therapy, the recommended use of purgatives, and—most important—limited availability (i.e., on a named-patient basis from the manufacturer).

Niclosamide is poorly absorbed. Tablets are given on an empty stomach in the morning after a liquid meal the night before, and this dose is followed by another 1 h later. For treatment of hymenolepiasis, the drug is administered for 7 days. A second course is often prescribed. The scolex and proximal segments of the tapeworms are killed on contact with niclosamide and may be digested in the gut. However, disintegration of the adult tapeworm results in the release of viable ova, which theoretically can result in autoinfection. Although fears of the development of cysticercosis in patients with *Taenia solium* infections have proved unfounded, it is still recommended that a brisk purgative be given 2 h after the first dose.

Nifurtimox This nitrofuran compound is an inexpensive and effective oral agent for the treatment of acute Chagas disease. Trypanosomes lack catalase and have very low levels of peroxidase; as a result, they are very vulnerable to by-products of oxygen reduction. When nifurtimox is reduced in the trypanosome, a nitro anion radical is formed and undergoes autooxidation, resulting in the generation of the superoxide anion O_2^-, hydrogen peroxide (H_2O_2), hydroperoxyl radical (HO_2), and other highly reactive and cytotoxic molecules. Despite the abundance of catalases, peroxidases, and superoxide dismutases that neutralize these destructive radicals in mammalian cells, nifurtimox has a poor therapeutic index. Prolonged use is required, but the course may have to be interrupted because of drug toxicity, which develops in 40–70% of recipients. Nifurtimox is well absorbed and undergoes rapid and extensive biotransformation; <0.5% of the original drug is excreted in urine. In 2020, the FDA approved this agent for the treatment of Chagas disease in children; it is now commercially available.

Nitazoxanide Nitazoxanide is a 5-nitrothiazole compound used for the treatment of cryptosporidiosis and giardiasis; it is active against other intestinal protozoa as well. The drug is approved for use in children 1–11 years of age.

The antiprotozoal activity of nitazoxanide is believed to be due to interference with the pyruvate-ferredoxin oxidoreductase (PFOR) enzyme–dependent electron transfer reaction that is essential to anaerobic energy metabolism. Studies have shown that the PFOR enzyme from *G. lamblia* directly reduces nitazoxanide by transfer of electrons in the absence of ferredoxin. The DNA-derived PFOR protein sequence of *Cryptosporidium parvum* appears to be similar to that of *G. lamblia*. Interference with the PFOR enzyme–dependent electron transfer reaction may not be the only pathway by which nitazoxanide exerts antiprotozoal activity.

After oral administration, nitazoxanide is rapidly hydrolyzed to an active metabolite, tizoxanide (desacetyl-nitazoxanide). Tizoxanide then undergoes conjugation, primarily by glucuronidation. It is recommended that nitazoxanide be taken with food; however, no studies have been conducted to determine whether the pharmacokinetics of tizoxanide and tizoxanide glucuronide differ in fasted versus fed subjects. Tizoxanide is excreted in urine, bile, and feces, and tizoxanide glucuronide is excreted in urine and bile. The pharmacokinetics of nitazoxanide in patients with impaired hepatic and/or renal function have not been studied. Tizoxanide is highly bound to plasma protein (>99.9%). Therefore, caution should be used when administering this agent concurrently with other highly plasma protein–bound drugs that have narrow therapeutic indices, as competition for binding sites may occur.

Oxamniquine This tetrahydroquinoline derivative is an effective alternative agent for the treatment of *S. mansoni*, although susceptibility to this drug exhibits regional variation. Oxamniquine exhibits anticholinergic properties, but its primary mode of action seems to rely on ATP-dependent enzymatic drug activation generating an intermediate that alkylates essential macromolecules, including DNA. In treated adult schistosomes, oxamniquine produces marked tegumental alterations like those seen with praziquantel but that develop less rapidly, becoming evident 4–8 days after treatment.

Oxamniquine is administered orally as a single dose and is well absorbed. Food retards absorption and reduces bioavailability. About 70% of an administered dose is excreted in urine as a mixture of pharmacologically inactive metabolites. Patients should be warned that their urine might have an intense orange-red color. Side effects are uncommon and usually mild, although hallucinations and seizures have been reported.

Paromomycin (Aminosidine) First isolated in 1956, this aminoglycoside is an effective oral agent for the treatment of infections due to intestinal protozoa. Parenteral paromomycin appears to be effective against visceral leishmaniasis in India.

Paromomycin inhibits protozoan protein synthesis by binding to the 30S ribosomal RNA in the aminoacyl-tRNA site, causing misreading of mRNA codons. Paromomycin is less active against *G. lamblia* than standard agents; however, like other aminoglycosides, paromomycin is poorly absorbed from the intestinal lumen, and the high levels of drug in the gut compensate for this relatively weak activity. If absorbed or administered systemically, paromomycin can cause ototoxicity and nephrotoxicity. However, systemic absorption is very limited, and toxicity should not be a concern in persons with normal kidneys. Topical formulations are not generally available.

Pentamidine Isethionate This diamidine is an effective alternative agent for some forms of leishmaniasis and trypanosomiasis. It is available for parenteral and aerosolized administration. Although its mechanism of action remains undefined, it is known to exert a wide range of effects, including interaction with trypanosomal kinetoplast DNA; interference with polyamine synthesis by a decrease in the activity of ornithine decarboxylase; and inhibition of RNA polymerase, topoisomerase, ribosomal function, and the synthesis of nucleic acids and proteins.

Pentamidine isethionate is well absorbed, highly tissue bound, and excreted slowly over several weeks, with an elimination half-life of 12 days. No steady-state plasma concentration is attained in persons given daily injections; the result is extensive accumulation of pentamidine in tissues, primarily the liver, kidney, adrenal gland, and spleen. Pentamidine does not penetrate well into the CNS. Pulmonary concentrations of pentamidine are increased when the drug is delivered in aerosolized form, but not when it is delivered systemically.

Rapid (<1-h) infusion of intravenous pentamidine often results in hypotension. Because electrolyte disturbances and mild to moderate nephrotoxicity occur commonly, pentamidine should be used with caution with other nephrotoxic agents. Pancreatitis and QT prolongation may also occur; cumulative damage to pancreatic islet cells may result in drug-induced diabetes mellitus. Similarly, hypoglycemia can develop, although much less commonly when pentamidine is given by the inhaled route.

Piperaquine This bisquinoline was synthesized in the 1960s and used widely for malaria control in China until resistance emerged. The development of artemisinin-based combination therapy led to its evaluation as a partner drug, and it is now combined with dihydroartemisinin. Piperaquine is highly lipophilic and has a prolonged half-life (~20 days), thus providing a period of posttreatment prophylaxis. The drug's mechanisms of action and resistance have not been well studied but are presumed to be similar to the other 4-aminoquinolines.

Piperazine The antihelminthic activity of piperazine is confined to ascariasis and enterobiasis. Piperazine acts as an agonist at extrasynaptic γ-aminobutyric acid (GABA) receptors, causing an influx of chloride ions in the nematode somatic musculature. Although the initial result is hyperpolarization of the muscle fibers, the ultimate effect is flaccid paralysis, leading to the expulsion of live worms. Patients should be warned, as this occurrence can be unsettling.

Praziquantel This heterocyclic pyrazinoisoquinoline derivative is highly active against a broad spectrum of trematodes and cestodes. It is the mainstay of treatment for schistosomiasis and is a critical part of community-based control programs.

All of the effects of praziquantel can be attributed either directly or indirectly to an alteration of intracellular calcium concentrations. Although the exact mechanism of action remains unclear, the major mechanism is disruption of the parasite tegument, causing tetanic contractures with loss of adherence to host tissues and, ultimately, disintegration or expulsion. Praziquantel induces changes in the antigenicity of the parasite by causing the exposure of concealed antigens. Praziquantel also produces alterations in schistosomal glucose metabolism, including decreases in glucose uptake, lactate release, glycogen content, and ATP levels.

Praziquantel exerts its parasitic effects directly and does not need to be metabolized to be effective. It is well absorbed but undergoes extensive first-pass hepatic clearance. Levels of the drug are increased when it is taken with food, particularly carbohydrates, or with cimetidine. Serum levels are reduced by glucocorticoids, chloroquine, carbamazepine, and phenytoin. Praziquantel is completely metabolized in humans, with 80% of the dose recovered as metabolites in urine within 4 days. It is not known to what extent praziquantel crosses the placenta, but retrospective studies suggest that it is safe in pregnancy.

Patients with schistosomiasis who have heavy parasite burdens may develop abdominal discomfort, nausea, headache, dizziness, and drowsiness. Symptoms begin 30 min after ingestion, may require spasmolytics for relief, and usually disappear spontaneously after a few hours.

Primaquine Phosphate Primaquine, an 8-aminoquinoline, has a broad spectrum of activity against all stages of plasmodial development in humans but has been used most effectively for eradication of the hepatic stage of these parasites. Primaquine must be metabolized by the host to be effective. It is, in fact, rapidly metabolized; only a small fraction of the dose of the parent drug is excreted unchanged. Although the parasiticidal activity of the three oxidative metabolites remains unclear, they are believed to affect both pyrimidine synthesis and the mitochondrial electron transport chain. The metabolites appear to have significantly less antimalarial activity than primaquine; however, their hemolytic activity is greater than that of the parent drug.

Primaquine causes marked hypotension after parenteral administration and therefore is given only by the oral route. It is rapidly and almost completely absorbed from the GI tract. Patients should be tested for G6PD deficiency before they receive primaquine. The drug may induce the oxidation of hemoglobin into methemoglobin, regardless of the G6PD status of the patient. Primaquine is otherwise well tolerated.

Proguanil (Chloroguanide) Proguanil inhibits plasmodial dihydrofolate reductase and is used with atovaquone for oral treatment of uncomplicated malaria or with chloroquine for malaria prophylaxis in parts of Africa without widespread chloroquine-resistant *P. falciparum*.

Proguanil exerts its effect primarily by means of the metabolite cycloguanil, whose inhibition of dihydrofolate reductase in the parasite disrupts deoxythymidylate synthesis, thus interfering with a key pathway involved in the biosynthesis of pyrimidines required for nucleic acid replication. There are no clinical data indicating that folate supplementation diminishes drug efficacy; women of childbearing age for whom atovaquone/proguanil is prescribed should continue taking folate supplements to prevent neural tube birth defects.

Proguanil is extensively absorbed regardless of food intake. The drug is 75% protein bound. The main routes of elimination are hepatic biotransformation and renal excretion; 40–60% of the proguanil dose is excreted by the kidneys. Drug levels are increased and elimination is impaired in patients with hepatic insufficiency.

Pyrantel Pamoate Pyrantel is a tetrahydropyrimidine formulated as pamoate. This safe, well-tolerated, inexpensive drug is used to treat a variety of intestinal nematode infections but is ineffective in trichuriasis. Pyrantel pamoate is usually effective in a single dose. Its target is the nicotinic acetylcholine receptor on the surface of nematode somatic muscle. Pyrantel depolarizes the neuromuscular junction of the nematode, resulting in its irreversible paralysis and allowing the natural expulsion of the worm.

Pyrantel pamoate is poorly absorbed from the intestine; >85% of the dose is passed unaltered in feces. The absorbed portion is metabolized and excreted in urine. Piperazine is antagonistic to pyrantel pamoate and should not be used concomitantly.

Pyrantel pamoate has minimal toxicity at the oral doses used to treat intestinal helminthic infection. It is not recommended for pregnant women or for children <12 months old.

Pyrimethamine When combined with short-acting sulfonamides, this diaminopyrimidine is effective in malaria, toxoplasmosis, and isosporiasis. Unlike mammalian cells, the parasites that cause these infections cannot use preformed pyrimidines obtained through salvage pathways but rather rely completely on de novo pyrimidine synthesis, for which folate derivatives are essential cofactors. The efficacy of pyrimethamine is increasingly limited by the development of resistant strains of *P. falciparum* and *P. vivax*. Single amino acid substitutions to parasite dihydrofolate reductase confer resistance to pyrimethamine by decreasing the enzyme's binding affinity for the drug.

Pyrimethamine is well absorbed; the drug is 87% bound to human plasma proteins. In healthy volunteers, drug concentrations remain at therapeutic levels for up to 2 weeks; drug levels are lower in patients with malaria.

At the usual dosage, pyrimethamine alone causes little toxicity except for occasional skin rashes and, more rarely, blood dyscrasias. Bone marrow suppression sometimes occurs at the higher doses used for toxoplasmosis; at these doses, the drug should be administered with folinic acid.

Pyronaridine This potent antimalarial is a benzonaphthyridine derivative first synthesized by Chinese researchers in 1970. Like chloroquine, pyronaridine targets hematin formation, inhibiting the production of β-hematin by forming complexes with it, with consequent enhancement of hematin-induced hemolysis. However, this drug is more potent than chloroquine: for complete lysis, pyronaridine is required at only 1/100th of the concentration needed with chloroquine. It also inhibits glutathione-dependent heme degradation. Despite its similar mode of action, pyronaridine remains effective against chloroquine-resistant strains. When combined with artesunate, it is effective for the treatment of acute, uncomplicated infection caused by *P. falciparum* or *P. vivax* in areas of low transmission with evidence of artemisinin resistance.

Pyronaridine is readily absorbed, widely distributed throughout the body, metabolized by the liver, and excreted in urine and stool. Its use is contraindicated in patients with severe liver or kidney impairment. Pyronaridine inhibits both CYP2D6 and P-glycoprotein in vitro, and these effects may have clinical relevance for patients taking medications for cardiac disease (e.g., metoprolol and digoxin).

Quinacrine† Despite being one of the first antimalarials, the antiprotozoal mechanism of quinacrine has not been fully elucidated. It is generally considered to intercalate into DNA and thereby inhibit replication and transcription. The drug also inhibits adenosine uptake, ATP incorporation into RNA, and NADH oxidase—the same enzyme that activates furazolidone. The differing relative quinacrine uptake rate between human cells and *G. lamblia* may explain the selective toxicity of the drug. Resistance correlates with decreased drug uptake.

Quinacrine is rapidly absorbed from the intestinal tract and is widely distributed in body tissues. Alcohol is best avoided because of a disulfiram-like effect. Although its production was discontinued in 1992, quinacrine can be obtained commercially from compounding pharmacies.

Quinine and Quinidine When combined with another agent, the cinchona alkaloid quinine is effective for the oral treatment of both uncomplicated, chloroquine-resistant malaria and babesiosis. Quinine acts rapidly against the asexual blood stages of all forms of the human malaria parasites. For severe malaria, only quinidine (the dextroisomer of quinine) is available in the United States. Quinine concentrates in the acidic food vacuoles of *Plasmodium* species. The drug inhibits the nonenzymatic polymerization of the highly reactive, toxic heme molecule into the nontoxic polymer pigment hemozoin.

Quinine is readily absorbed when given orally. In patients with malaria, the elimination half-life of quinine increases according to the severity of the infection. However, toxicity is avoided by an increase in the concentration of plasma glycoproteins. The cinchona alkaloids are extensively metabolized, particularly by CYP3A4; only 20% of the dose is excreted unchanged in urine. The drug's metabolites are also excreted in urine and may be responsible for toxicity in patients with renal failure. Renal excretion of quinine is decreased when cimetidine is taken and increased when the urine is acidic. The drug readily crosses the placenta.

Quinidine is both more potent as an antimalarial and more toxic than quinine. Its use requires cardiac monitoring. Dose reduction is necessary in persons with severe renal impairment.

Spiramycin† This macrolide is used to treat acute toxoplasmosis in pregnancy and congenital toxoplasmosis. While the mechanism of action is similar to that of other macrolides, the efficacy of spiramycin in toxoplasmosis appears to stem from its rapid and extensive intracellular penetration, which results in macrophage drug concentrations 10–20 times greater than serum concentrations.

Spiramycin is rapidly and widely distributed throughout the body and reaches concentrations in the placenta up to five times those in serum. This agent is excreted mainly in bile. Indeed, in humans, the urinary excretion of active compounds represents only 20% of the administered dose.

Serious reactions to spiramycin are rare. Of the available macrolides, spiramycin appears to have the lowest risk of drug interactions. Complications of treatment are rare but, in neonates, can include life-threatening ventricular arrhythmias that disappear with drug discontinuation.

Spiramycin is not formally approved for use in the United States, but it is accessible through a compassionate use program for toxoplasmosis in pregnancy through the FDA (301-796-1400).

Sulfonamides See Table 229-1 and Chap. 149.

Suramin* This derivative of urea is the drug of choice for the early stage of African trypanosomiasis. The drug is polyanionic and acts by forming stable complexes with proteins, thus inhibiting multiple enzymes essential to parasite energy metabolism. Suramin appears to inhibit all trypanosome glycolytic enzymes more effectively than it inhibits the corresponding host enzymes.

Suramin is parenterally administered. It binds to plasma proteins and persists at low levels for several weeks after infusion. Its metabolism is negligible. This drug does not penetrate the CNS.

Tafenoquine Tafenoquine is an 8-aminoquinoline with causal prophylactic activity. Its prolonged half-life (2–3 weeks) allows longer dosing intervals when the drug is used for prophylaxis. Tafenoquine has been well tolerated in clinical trials. When tafenoquine is taken with food, its absorption is increased by 50% and the most commonly reported adverse event—mild GI upset—is diminished. Like primaquine, tafenoquine is a potent oxidizing agent, causing hemolysis in patients with G6PD deficiency as well as methemoglobinemia. It has been commercially available since FDA approval in 2018.

Tetracyclines See Table 229-1 and Chap. 149.

Thiabendazole Discovered in 1961, thiabendazole remains one of the most potent of the numerous benzimidazole derivatives. However, its use has declined significantly because of a higher frequency of adverse effects than is seen with other, equally effective agents.

Thiabendazole is active against most intestinal nematodes that infect humans. Although the exact mechanism of its antihelminthic activity has not been fully elucidated, it is likely to be similar to that of other benzimidazole drugs: namely, inhibition of polymerization of parasite β-tubulin. The drug also inhibits the helminth-specific enzyme fumarate reductase. In animals, thiabendazole has anti-inflammatory, antipyretic, and analgesic effects, which may explain its usefulness in dracunculiasis and trichinellosis. Thiabendazole also suppresses egg and/or larval production by some nematodes and may inhibit the subsequent development of eggs or larvae passed in feces. Despite the emergence and global spread of thiabendazole-resistant trichostrongyliasis among sheep, there have been no reports of drug resistance in humans.

Thiabendazole is available in tablet form and as an oral suspension. The drug is rapidly absorbed from the GI tract but can also be absorbed through the skin. Thiabendazole should be taken after meals. This agent is extensively metabolized in the liver before ultimately being excreted; most of the dose is excreted within the first 24 h. The usual dose of thiabendazole is determined by the patient's weight, but some treatment regimens are parasite specific. No adjustments are recommended in patients with renal or hepatic failure; only cautious use is advised.

Coadministration of thiabendazole to patients taking theophylline can result in an increase in theophylline levels by >50%. Therefore, serum levels of theophylline should be monitored closely in this situation.

Tinidazole This nitroimidazole is effective for the treatment of amebiasis, giardiasis, and trichomoniasis. Like metronidazole, tinidazole must undergo reductive activation by the parasite's metabolic system before it can act on protozoal targets. Tinidazole inhibits the synthesis of new DNA in the parasite and causes degradation of existing DNA. The reduced free-radical derivatives alkylate DNA, with consequent cytotoxic damage to the parasite. This damage appears to be produced by short-lived reduction intermediates, resulting in helix destabilization and strain breakage of DNA. The mechanism of action and side effects of tinidazole are similar to those of metronidazole, but adverse events appear to be less frequent and severe with tinidazole. In addition, the significantly longer half-life of tinidazole (>12 h) offers potential cure with a single dose.

Tribendimidine Tribendimidine, a diamidine derivative of aminophenylamidine amidantel, is a cholinergic agonist that is selective for the nicotinic acetylcholine receptors of nematode muscle. This novel agent has a broad spectrum of activity against a wide variety of helminths and is highly effective against food-borne trematodes, with a similar cure rate to praziquantel. Clinical trials have demonstrated efficacy of a single dose alone or in combination with other helminthics against soil-transmitted helminth infections. The drug is an L-type nicotinic acetylcholine receptor agonist and exhibits the same method of action as levamisole and pyrantel; therefore, it may not be effective in regions where resistance to these agents is widespread. It is not available in the United States.

Triclabendazole In contrast to other benzimidazoles, the antihelminthic activity of triclabendazole is highly specific for *Fasciola* and *Paragonimus* species, with little activity against nematodes, cestodes, and other trematodes. Triclabendazole is effective against all stages of *Fasciola* species. The active sulfoxide metabolite of triclabendazole binds to fluke tubulin by assuming a unique nonplanar configuration and disrupts microtubule-based processes. Resistance to triclabendazole in veterinary use has been reported in Australia and Europe; however, no resistance has been documented in humans.

Triclabendazole is rapidly absorbed after oral ingestion; administration with food enhances its absorption and shortens the elimination half-life of the active metabolite. Both the sulfoxide and the sulfone metabolites are highly protein bound (>99%). Treatment with triclabendazole is typically given in one or two doses. No clinical data are available regarding dose adjustment in renal or hepatic insufficiency; however, given the short course of therapy and extensive hepatic metabolism of triclabendazole, dose adjustment is unlikely to be necessary. No information exists on drug interactions.

Trimethoprim-Sulfamethoxazole See Table 229-1 and Chap. 149.

FURTHER READING

Dziduch K et al: The current directions of searching for antiparasitic drugs. Molecules 27:1534, 2022.

Fehintola FA et al: Drug interactions in the treatment and chemoprophylaxis of malaria in HIV infected individuals in sub-Saharan Africa. Curr Drug Metab 12:51, 2011.

Keiser J, Häberli C: Evaluation of commercially available anthelminthics in laboratory models of human intestinal nematode infections. ACS Infect Dis 7:1177, 2021.

Keiser J et al: Antiparasitic drugs for paediatrics: Systematic review, formulations, pharmacokinetics, safety, efficacy, and implications for control. Parasitology 138:1620, 2011.

Kelesidis T, Falagas ME: Substandard/counterfeit antimicrobial drugs. Clin Microbiol Rev 28:443, 2015.

Pink R et al: Opportunities and challenges in antiparasitic drug discovery. Nat Rev Drug Discov 4:727, 2005.

Section 18 Protozoal Infections

230

Amebiasis and Infection with Free-Living Amebae

Rosa M. Andrade, Sharon L. Reed

AMEBIASIS

DEFINITION

Amebiasis is an infection caused by *Entamoeba histolytica*, an intestinal protozoan. Its spectrum of clinical syndromes ranges from asymptomatic colonization (90% of cases) to invasive amebiasis, which accounts for 10% of affected individuals. Invasive amebiasis frequently presents as intestinal colitis (dysentery or diarrhea) or as extraintestinal amebiasis, in which abscesses of the liver are more commonly found than involvement of the lungs or brain.

LIFE CYCLE AND TRANSMISSION

E. histolytica is acquired by ingestion of viable cysts from fecally contaminated water, food, or hands **(Fig. 230-1)**. Food-borne exposure is the most prevalent form of transmission. It occurs when food handlers are shedding cysts or food is being grown with feces-contaminated soil, fertilizer, or water. Less common means of transmission include oral and anal sexual practices and, in rare instances, direct rectal inoculation through colonic irrigation devices. Motile trophozoites are released from cysts in the small intestine and, in most patients, remain as harmless commensals in the large bowel. After encystation, infectious cysts are shed in the stool and can survive for several weeks in a moist environment. In some patients, the trophozoites invade either the bowel mucosa, causing symptomatic colitis, or the bloodstream, causing distant abscesses of the liver, lungs, or brain. The trophozoites may not encyst in patients with active dysentery, and motile hematophagous trophozoites are frequently present in fresh stools. Trophozoites are rapidly killed by exposure to air or stomach acid and therefore cannot transmit infection.

EPIDEMIOLOGY

E. histolytica infection typically affects underdeveloped tropical regions with poor sanitation systems and hygiene, occurring often in children <5 years of age. This infection is widespread in the Indian subcontinent and Africa, parts of East Asia (Thailand), and Central and South America (Mexico and Colombia). According to the Global Burden of Disease 2016 study, amebiasis accounts for 26,748 all-age deaths, including 4567 children <5 years old.

In contrast, returning travelers, recent immigrants, men who have sex with men (MSM), military personnel, and inmates of institutions are the main groups at risk for amebiasis in developed countries. The GeoSentinel Surveillance Network, which encompasses information from tropical medicine clinics on six continents, showed that, among long-term travelers (trip duration, >6 months), diarrhea due to *E. histolytica* was among the most common diagnoses. In countries like Japan and Spain, amebiasis is emerging as a sexually transmitted disease, first reported in HIV-positive MSM. However, in Japan, *E. histolytica* infections, symptomatic or asymptomatic, seem to be steadily spreading among non-HIV men and women.

Worldwide, *E. histolytica* is the second most common cause of death related to parasitic infection (after malaria). Invasive colitis and liver abscesses are tenfold more common among men than among women; this difference has been attributed to a disparity in complement-mediated killing and effects of testosterone on the secretion of interferon γ. The wide spectrum of clinical disease caused by *Entamoeba* is due in part to the differences between the two major infecting species,

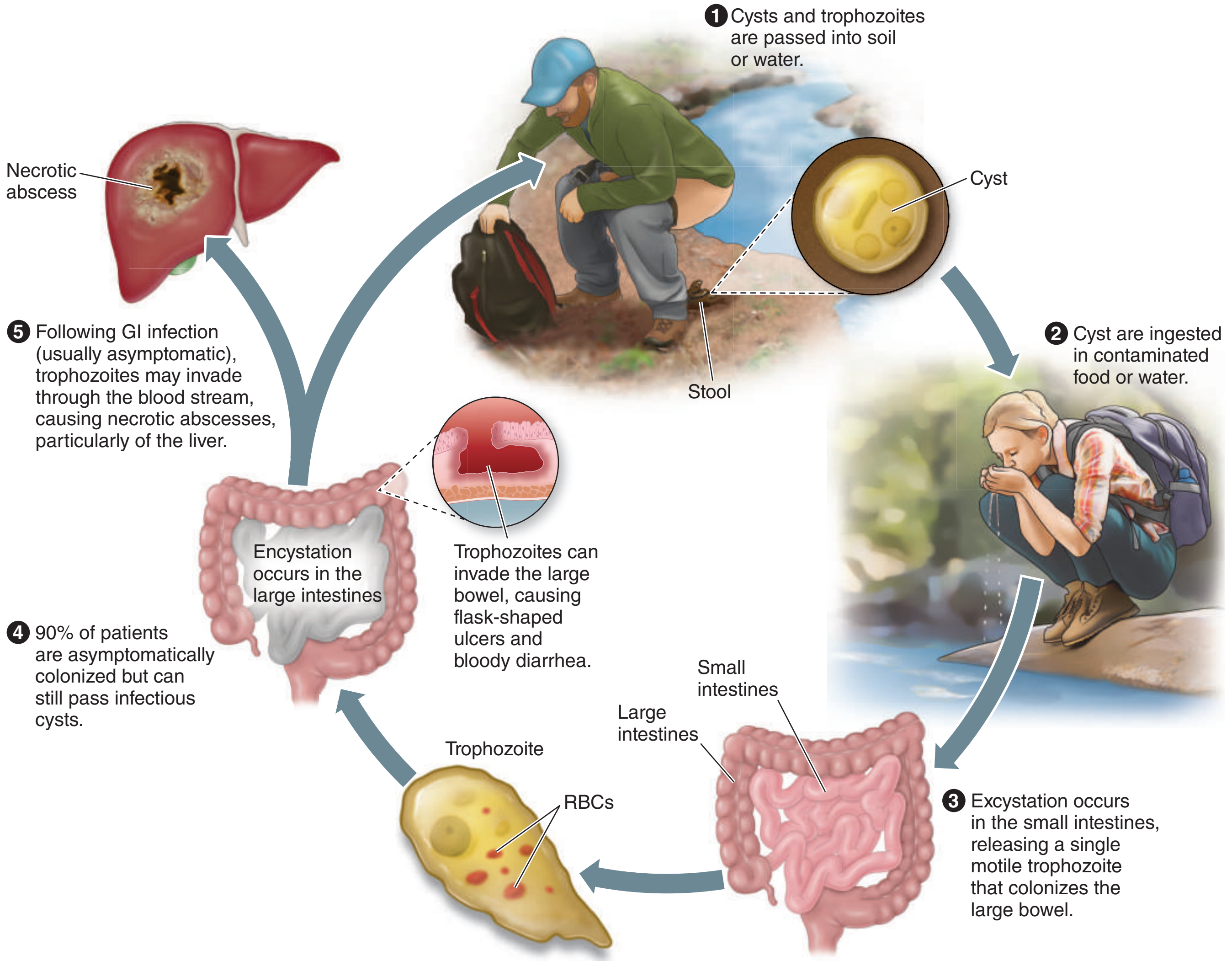

1 Cysts and trophozoites are passed into soil or water.
Cyst
Stool
2 Cyst are ingested in contaminated food or water.
Small intestines
Large intestines
3 Excystation occurs in the small intestines, releasing a single motile trophozoite that colonizes the large bowel.
Trophozoite
RBCs
Encystation occurs in the large intestines
Trophozoites can invade the large bowel, causing flask-shaped ulcers and bloody diarrhea.
4 90% of patients are asymptomatically colonized but can still pass infectious cysts.
5 Following GI infection (usually asymptomatic), trophozoites may invade through the blood stream, causing necrotic abscesses, particularly of the liver.
Necrotic abscess

FIGURE 230-1 Life cycle of *Entamoeba histolytica*. GI, gastrointestinal; RBCs, red blood cells.

E. histolytica and *E. dispar. E. histolytica* has unique surface antigens, is genetically distinct, and possesses virulence properties that distinguish it from the morphologically identical *E. dispar*.

Most asymptomatic carriers, including MSM and patients with AIDS, harbor *E. dispar* and have self-limited infections. In this respect, *E. dispar* is dissimilar to other enteric pathogens such as *Cryptosporidium* and *Cystoisospora belli*, which can cause self-limited illnesses in immunocompetent hosts but devastating diarrhea in patients with AIDS. These observations indicate that *E. dispar* is incapable of causing invasive disease. Through genomic sequencing, new species of *Entamoeba* have been identified: *E. moshkovskii* and *E. bangladeshi*. These new species are microscopically indistinguishable from *E. histolytica*. Although *E. moshkovskii* causes diarrhea, weight loss, and colitis in mice, a prospective evaluation of children from the Mirpur community of Dhaka, Bangladesh, found that most children who had diarrheal diseases associated with *E. moshkovskii* were simultaneously infected with at least one other enteric pathogen. *E. bangladeshi nov.* sp., *Bangladesh* was first reported in 2012 in this Bangladeshi community; however, it has been isolated in South African subjects of all ages in recent years. Additional clinical and epidemiologic studies are needed to discern the true role of *E. bangladeshi* in the human host.

PATHOGENESIS AND PATHOLOGY

Both trophozoites and cysts are found in the intestinal lumen, but only trophozoites of *E. histolytica* invade tissue. The trophozoite is 20–60 μm in diameter and contains vacuoles and a nucleus with a characteristic central nucleolus. Trophozoites attach to colonic mucus and epithelial cells by Gal/GalNAc adherence lectin and release glycosidases and proteases that cause degradation of mucus polymers. Extracellular cysteine proteinases degrade collagen, elastin, IgA, IgG, and the anaphylatoxins C3a and C5a. After disruption of the mucous layer, trophozoites damage the mucosa by contact-dependent and contact-independent cytotoxicity. The contact-dependent cytotoxicity is attributable to induction of apoptotic cell death; trogocytosis-mediated cell death (ingestion of fragments of living cells); and lysis of inflammatory cells (neutrophils, monocytes, and lymphocytes), colonic cells, and hepatic cells through release of phospholipase A and pore-forming peptides. Contact-independent cytotoxicity follows production of inflammatory mediators, such as prostaglandin E2, by trophozoites, ultimately leading to increased ion permeability of intercellular tight junctions.

E. histolytica trophozoites are constantly exposed to reactive oxygen and nitrogen species arising from their own metabolism and from the host during tissue invasion. The ability to resist reactive oxygen species or reactive nitrogen species such as nitric oxide or *S*-nitrosothiols (e.g., *S*-nitrosoglutathione [GSNO] and *S*-nitrosocysteine [CySNO]) is also a virulence factor. Since *E. histolytica* lacks glutathione and glutathione reductase, it relies on its thioredoxin–thioredoxin reductase system to prevent, regulate, and repair the damage caused by oxidative stress. This antioxidant system is versatile: it has the ability to reduce reactive nitrogen species and use an alternative electron donor, such as nicotinamide adenine dinucleotide. In addition, trophozoites preincubated with *E. coli* 055 are more resistant to hydrogen peroxide–mediated killing by using bacterially produced malate dehydrogenase and oxaloacetate for protection from oxidative stress. Metronidazole, the current standard of therapy for amebiasis, seems to exert its antiparasitic effect through inhibition of this antioxidant system.

Phagocytosis is a virulence factor that leads to a defective proliferation of *E. histolytica* if inhibited. Trophozoites use membrane-associated carbohydrate-binding proteins to phagocytose intestinal bacteria, especially gram-negative Enterobacteriaceae, for their nutrients. Interactions with commensal bacteria, such as *Escherichia coli*, can attenuate the virulence of *E. histolytica* by decreasing the expression of Gal/GalNAc lectin. In contrast, ingestion of enteropathogenic bacteria, such as enteropathogenic *E. coli* and *Shigella dysenteriae*, increases expression of the Gal/GalNAc lectin and enhances *E. histolytica* cysteine protease activity.

The interactions between *E. histolytica* and the gut microbiome can determine the severity of disease. During the first 2 years of life, the gut immune system and the microbiome mature rapidly. In one study, ~80% of children from the Bangladeshi community of Dhaka were found to be infected with *E. histolytica* by 2 years of age. Fecal anti–Gal/GalNAc lectin IgA was associated with protection from reinfection, while a high parasite burden in the first year of life was associated with the expansion of *Prevotella copri* in their gut microbiome and presence of diarrhea. In adults, the diversity of the gut microbiome can determine the pathogenesis of disease. As shown in a Japanese cohort through next-generation sequencing analysis, the gut microbiome in asymptomatic *E. histolytica* carriers was more homogeneous, with high abundance of *Ruminococcaceae*, *Coriobacteriaceae*, and *Clostridiaceae*, while these bacteria were absent or in low abundance in patients with invasive disease (symptomatic). These observations suggest the protective role of these bacteria.

Antimicrobial peptides, such as cathelicidins, are an important component of innate immunity and are induced by *E. histolytica* upon intestinal invasion in a mouse model. In this model, cecal cathelicidin-related antimicrobial peptide mRNA increased by >4-fold at 3 days and >100-fold at 7 days. However, *E. histolytica* remained resistant to cathelicidin-mediated killing, probably because the antimicrobial peptide was digested by amebic cysteine proteinases. Other in vitro studies have shown that α- and β-defensins may damage *E. histolytica* membrane integrity.

IgA plays a critical role in acquired immunity to *E. histolytica*. A study in Bangladeshi schoolchildren revealed that an intestinal IgA response to Gal/GalNAc reduced the risk of new *E. histolytica* infection by 64%. Serum IgG antibody is not protective; titers correlate with the duration of illness rather than with the severity of disease. Indeed, Bangladeshi children with a serum IgG response were more likely than those without such a response to develop new *E. histolytica* infection. In infants from this same Bangladeshi community, passive immunity conferred by maternal parasite-specific IgA via breastfeeding resulted in a 39% reduced risk of infection and a 64% reduced risk of diarrheal disease from *E. histolytica* during the first year of life. However, this protection appeared to be species-specific, with little or no protection conferred from infections with other species such as *E. dispar* or *E. bangladeshi*.

This Bangladeshi cohort has furthered our understanding of the genetic susceptibility factors associated with *E. histolytica* disease. Heterozygosity of the major histocompatibility complex (MHC) class II allele DQB1*0601 was found to protect against amebic intestinal disease, which supports the role of antigen processing and CD4+ T cells in resistance to amebiasis. Adipocyte leptin receptors (LEPRs) are expressed on intestinal epithelial cells, prevent apoptosis, promote tissue repair, and may decrease neutrophil infiltration. In this cohort, a single amino acid substitution (Q223R) in LEPRs nearly quadrupled the risk for amebic intestinal disease in children and increased the risk for amebic liver abscesses in adults. Similarly, variations in the locus of cAMP-responsive element modulator/cullin 2 (CREM/CUL2) may increase the risk for diarrhea in children who acquired *E. histolytica* within their first year of life. Interestingly, both genetic variations, Q223R and CREM, are overrepresented in this geographical region.

The earliest intestinal lesions are microulcerations of the mucosa of the cecum, sigmoid colon, or rectum that release erythrocytes, inflammatory cells, and epithelial cells. A colonoscopy reveals small ulcers with heaped-up margins and normal intervening mucosa **(Fig. 230-2*A*)**. Submucosal extension of ulcerations under viable-appearing surface mucosa causes the classic "flask-shaped" ulcer containing trophozoites at the margins of dead and viable tissues. Although neutrophilic infiltrates may accompany early lesions in animals, human intestinal infection is marked by a paucity of inflammatory cells, probably in part because of the killing of neutrophils by trophozoites **(Fig. 230-2*B*)**. Treated ulcers characteristically heal with little or no scarring. Occasionally, however, full-thickness necrosis and perforation occur.

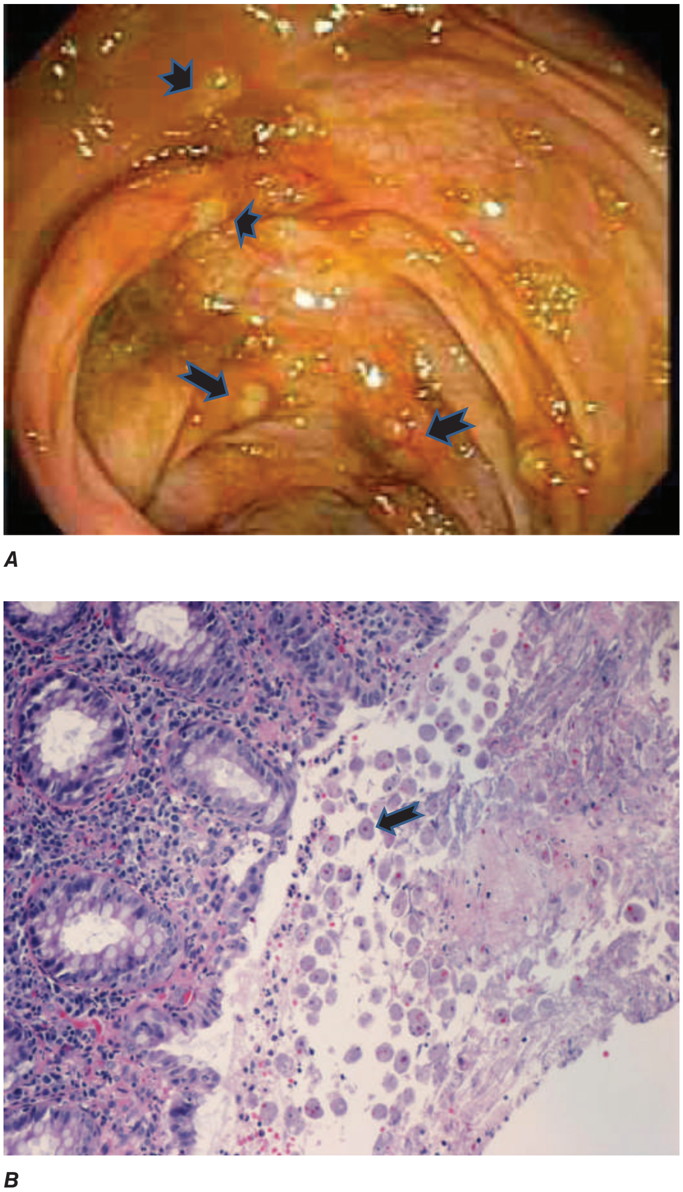

FIGURE 230-2 Endoscopic and histopathologic features of intestinal amebiasis. ***A.*** Appearance of ulcers on colonoscopy (*arrows*). ***B.*** Inflammatory infiltrate and *Entamoeba histolytica* trophozoites (*arrows*) in invasive amebic colitis (hematoxylin and eosin). *(Courtesy of the Department of Pathology and Gastroenterology, San Diego VA Medical Center.)*

Rarely, intestinal infection results in the formation of a mass lesion, or *ameboma*, in the bowel lumen. The overlying mucosa is usually thin and ulcerated, while other layers of the wall are thickened, edematous, and hemorrhagic; this condition results in exuberant formation of granulation tissue with little fibrous-tissue response.

Amebic liver abscesses are age- and gender-dependent. Men 30–60 years of age are most commonly infected at a rate 10–12 times higher than women in the same age group. Studies in animal models have demonstrated that testosterone may increase susceptibility to amebic liver abscess by modulating the secretion of interferon γ by natural killer T cells, which are activated through *E. histolytica* lipopeptidophosphoglycan present on the surface of ameba trophozoites. Liver abscesses are always preceded by intestinal colonization, which may be asymptomatic. Blood vessels may be compromised early by wall lysis and thrombus formation. Trophozoites invade veins to reach the liver through the portal venous system. *E. histolytica* is resistant to complement-mediated lysis—a property critical to survival in the bloodstream. Inoculation of amebae into the portal system of hamsters results in an acute cellular infiltrate consisting predominantly of neutrophils. Later, the neutrophils are lysed by contact with amebae, and the release of neutrophil toxins may contribute to necrosis of hepatocytes. The liver parenchyma is replaced by necrotic material that is surrounded by a thin rim of congested liver tissue. Although the necrotic contents of a liver abscess are classically described as "anchovy paste," the fluid is variable in color; it is composed of bacteriologically sterile granular debris with few or no cells. Amebae, if seen, tend to be found near the capsule of the abscess.

CLINICAL SYNDROMES

Intestinal Amebiasis The most common type of amebic infection is asymptomatic cyst passage (90% of patients). Even in highly endemic areas, most patients harbor *E. dispar*.

Symptomatic amebic colitis develops 2–6 weeks after the ingestion of infectious cysts. A gradual onset of lower abdominal pain and mild diarrhea is followed by malaise, weight loss, and diffuse lower abdominal or back pain. Cecal involvement may mimic acute appendicitis. Patients with full-blown dysentery may pass 10–12 stools per day. The stools contain little fecal material and consist mainly of blood and mucus. In contrast to those with bacterial diarrhea, fewer than 40% of patients with amebic dysentery are febrile. Virtually all patients have heme-positive stools.

More fulminant intestinal infection, with severe abdominal pain, high fever, and profuse diarrhea, is rare and occurs predominantly in children. Patients may develop toxic megacolon, in which there is severe bowel dilation with intramural air. Patients receiving glucocorticoids are at risk for severe amebiasis. The association between severe amebiasis complications and glucocorticoid therapy emphasizes the importance of excluding amebiasis when inflammatory bowel disease is suspected. An occasional patient presents with only an asymptomatic or tender abdominal mass caused by an ameboma, which is easily confused with cancer on barium studies. A positive serologic test or biopsy can prevent unnecessary surgery in this setting.

Environmental enteropathy ("impoverished gut"; blunted small-intestinal villi with lamina propria inflammation) is observed in tropical developing areas with endemic enteric infections, such as amebiasis. It is associated with functional gastrointestinal impairment causing malnutrition and stunted growth in children within the first 2 years of life. Bangladeshi children with symptomatic *E. histolytica* infections were 2.9 times more likely to be malnourished and 4.7 times more likely to be short for their age than were children without symptomatic infections. These factors affect their cognitive development and may be linked to loss of productivity in adulthood.

Amebic Liver Abscess Extraintestinal infection by *E. histolytica* most often involves the liver. Of travelers who develop an amebic liver abscess after leaving an endemic area, 95% do so within 5 months. Young patients with an amebic liver abscess are more likely than older patients to present in the acute phase with prominent symptoms of <10 days' duration. Most patients are febrile and have right-upper-quadrant pain, which may be dull or pleuritic in nature and may radiate to the shoulder. Point tenderness over the liver and right-sided pleural effusion are common. Jaundice is rare. Although the initial site of infection is the colon, fewer than one-third of patients with an amebic abscess have active diarrhea. Older patients from endemic areas are more likely to have a subacute course lasting 6 months, with weight loss and hepatomegaly. About one-third of patients with chronic presentations are febrile. Thus, the clinical diagnosis of an amebic liver abscess may be difficult to establish because the symptoms and signs are often nonspecific. Since 10–15% of patients present only with fever, amebic liver abscess must be considered in the differential diagnosis of fever of unknown origin **(Chap. 22)**.

Complications of Amebic Liver Abscess Pleuropulmonary involvement, which is reported in 20–30% of patients, is the most frequent complication of amebic liver abscess. Manifestations include sterile effusions, contiguous spread from the liver, and rupture into the pleural space. Sterile effusions and contiguous spread usually resolve with medical therapy, but frank rupture into the pleural space requires drainage. A hepatobronchial fistula may cause cough productive of large amounts of necrotic material that may contain amebae. This dramatic complication carries a good prognosis. Abscesses that rupture into the peritoneum may present as an indolent leak or an acute abdomen and require both percutaneous catheter drainage and medical therapy. Rupture into the pericardium, usually from abscesses of the left lobe of the liver, carries the gravest prognosis; it can occur during medical therapy and requires surgical drainage.

Involvement of Other Extraintestinal Sites The genitourinary tract may become involved by direct extension of amebiasis from the colon or by hematogenous spread of the infection. Painful genital ulcers, characterized by a punched-out appearance and profuse discharge, may develop secondary to extension from either the intestine or the liver. Both of these conditions respond well to medical therapy. Cerebral involvement has been reported in fewer than 0.1% of patients in large clinical series. Symptoms and prognosis depend on the size and location of the lesion.

DIAGNOSTIC TESTS

Laboratory Diagnosis Stool examinations, serologic tests, and noninvasive imaging of the liver remain the most important procedures in the diagnosis of amebiasis in most parts of the world. Fecal findings suggestive of amebic colitis include a positive test for heme, a paucity of neutrophils, and amebic cysts or trophozoites. The definitive diagnosis of amebic colitis is made by the demonstration of hematophagous trophozoites of *E. histolytica*. Because trophozoites are killed rapidly by water, drying, or barium, it is important to examine at least three fresh stool specimens. Examination of a combination of wet mounts, iodine-stained concentrates, and trichrome-stained preparations of fresh stool and concentrates for cysts or trophozoites confirms the diagnosis in 75–95% of cases. Cultures of amebae are more sensitive but are not routinely available. If stool examinations are negative, sigmoidoscopy with biopsy of the edge of ulcers may increase the yield, but this procedure is dangerous during fulminant colitis because of the risk of perforation. Trophozoites in a biopsy specimen from a colonic mass confirm the diagnosis of ameboma, but trophozoites are rare in liver aspirates because they are found in the abscess capsule and not in the readily aspirated necrotic center. Accurate diagnosis requires experience, since the trophozoites may be confused with neutrophils and the cysts must be differentiated morphologically from those of *Entamoeba hartmanni*, *Entamoeba coli*, and *Endolimax nana*, which do not cause clinical disease and do not warrant therapy. Unfortunately, the cysts of *E. histolytica* cannot be distinguished microscopically from those of *E. dispar*, *E. moshkovskii*, or *E. bangladeshi*. Therefore, the microscopic diagnosis of *E. histolytica* can be made only by the detection of *Entamoeba* trophozoites that have ingested erythrocytes.

Serology is an important addition to the methods used for parasitologic diagnosis of invasive amebiasis. Enzyme-linked immunosorbent assays are positive in >90% of cases with colitis, ameboma, or liver abscess. Positive results in conjunction with the appropriate clinical syndrome suggest active disease because serologic findings usually revert to negative within 6–12 months. Even in highly endemic areas such as South Africa, fewer than 10% of asymptomatic individuals have a positive amebic serology. The interpretation of the indirect hemagglutination test is difficult because titers may remain positive for as long as 10 years.

Early in infection, up to 10% of patients with acute amebic liver abscess may have negative serologic findings, which usually become positive within a week. In contrast to carriers of *E. dispar*, most asymptomatic carriers of *E. histolytica* develop antibodies. Thus, serologic tests are helpful in assessing the risk of invasive amebiasis in asymptomatic, cyst-passing individuals in nonendemic areas. Serologic tests should also be performed in patients with ulcerative colitis before the institution of glucocorticoid therapy to prevent the development of severe colitis or toxic megacolon owing to unsuspected amebiasis.

More sensitive, specific, and rapid tests in stool are now in use in most developed countries. Enzyme immunoassays using monoclonal antibodies for the Gal/GalNAc lectin allow specific detection of *E. histolytica* in 96-well plates or a single cassette for rapid detection. Sensitivity is 87% and specificity >90% in stool and can also be used in liver abscess aspirates. The number of commercially available PCR tests has significantly expanded. Most are multiplex for GI protozoa alone or include major bacterial and viral pathogens, as well. The sensitivity and specificity approaches 100%, and results can be available in as little as an hour. Their cost precludes use in most developing countries.

Other amplification tests such as a loop-mediated isothermal amplification (LAMP) assay may be a potential alternative for direct detection of *E. histolytica* DNA in pus samples from amebic liver abscesses. LAMP is a relatively simple, rapid, and low-cost method of DNA amplification that could be a better alternative for diagnosis in developing countries.

Routine hematology and chemistry tests usually are not very helpful in the diagnosis of invasive amebiasis. About three-fourths of patients with an amebic liver abscess have leukocytosis (>10,000 cells/μL); this condition is particularly likely if symptoms are acute or complications have developed. Invasive amebiasis does not elicit eosinophilia. Anemia, if present, is usually multifactorial. Even with large liver abscesses, liver enzyme levels are normal or minimally elevated. The alkaline phosphatase level is most often elevated and may remain so for months. Aminotransferase elevations suggest acute disease or a complication.

Radiographic Studies Radiographic barium studies are potentially dangerous in acute amebic colitis. Amebomas are usually identified first by a barium enema, but biopsy is necessary for differentiation from carcinoma.

Radiographic techniques such as ultrasonography, CT, and MRI are all useful for detection of the round or oval hypoechoic cyst. More than 80% of patients who have had symptoms for >10 days have a single abscess of the right lobe of the liver **(Fig. 230-3)**. Approximately 50% of patients who have had symptoms for <10 days have multiple abscesses. Findings associated with complications include large abscesses (>10 cm) in the superior part of the right lobe, which may rupture into the pleural space; multiple lesions, which must be differentiated from pyogenic abscesses; and lesions of the left lobe, which may rupture into the pericardium. Because abscesses resolve slowly and may increase in size despite a clinical response to therapy, frequent follow-up ultrasonography may prove confusing. Complete resolution of a liver abscess within 6 months can be anticipated in two-thirds of patients, but 10% may have persistent abnormalities for a year.

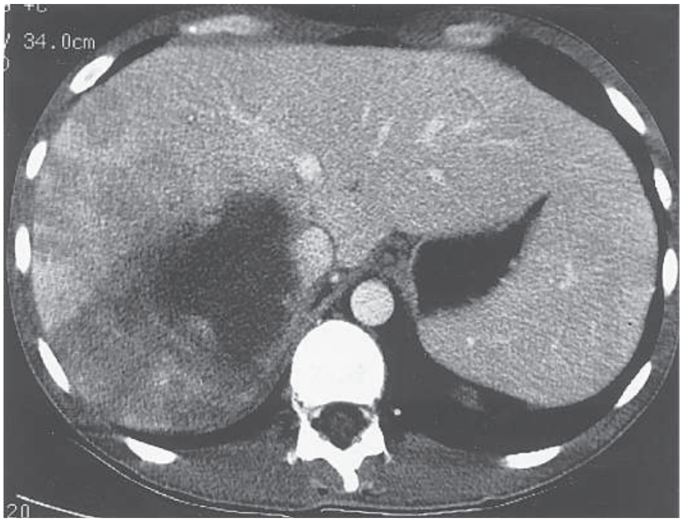

FIGURE 230-3 Abdominal CT scan of a large amebic abscess of the right lobe of the liver. *(Courtesy of the Department of Radiology, UCSD Medical Center, San Diego; with permission.)*

Differential Diagnosis The differential diagnosis of intestinal amebiasis includes bacterial diarrheas **(Chap. 138)** caused by *Campylobacter* **(Chap. 169)**; enteroinvasive *Escherichia coli* **(Chap. 166)**; and species of *Shigella* **(Chap. 172)**, *Salmonella* **(Chap. 171)**, and *Vibrio* **(Chap. 173)**. Because the typical patient with amebic colitis has less prominent fever than in these other conditions as well as heme-positive stools with few neutrophils, correct diagnosis requires bacterial cultures, microscopic examination of stools, amebic serologic testing, and stool specific antigens or PCR. As has been mentioned, amebiasis must be ruled out in any patient thought to have inflammatory bowel disease.

Because of the variety of presenting signs and symptoms, amebic liver abscess can easily be confused with pulmonary or gallbladder disease or with any febrile illness with few localizing signs, such as malaria **(Chap. 231)** or typhoid fever **(Chap. 171)**. The diagnosis should be considered in members of high-risk groups who have recently traveled outside the United States **(Chap. 130)** and in inmates of institutions. Once radiographic studies have identified an abscess in the liver, the most important differential diagnosis is between amebic and pyogenic abscess. Patients with pyogenic abscess typically are older and have a history of underlying bowel disease or recent surgery. Amebic serology is helpful, but aspiration of the abscess, with Gram's staining and culture of the material, may be required for differentiation of the two diseases.

TREATMENT

Amebiasis

INTESTINAL DISEASE (TABLE 230-1)

The drugs used to treat amebiasis can be classified according to their primary site of action. Luminal amebicides are poorly absorbed and reach high concentrations in the bowel, but their activity is limited to cysts and trophozoites close to the mucosa. Only two luminal drugs are available in the United States: iodoquinol and paromomycin. Indications for the use of luminal agents include eradication of cysts in patients with colitis or a liver abscess and treatment of asymptomatic carriers. The majority of asymptomatic individuals who pass cysts are colonized with *E. dispar*, which does not warrant specific therapy. However, it is prudent to treat asymptomatic individuals who pass cysts unless *E. dispar* colonization can be definitively demonstrated by specific antigen-detection tests.

Tissue amebicides reach high concentrations in the blood and tissue after oral or parenteral administration. The development of nitroimidazole compounds, especially metronidazole, was a major advance in the treatment of invasive amebiasis. Patients with amebic colitis should be treated with IV or oral metronidazole. Side effects include nausea, vomiting, abdominal discomfort, and a disulfiram-like reaction. Another, longer-acting imidazole compound, tinidazole, is likewise effective and is available in the United States. All patients should also receive a full course of therapy with

TABLE 230-1 Drug Therapy for Amebiasis

INDICATION	THERAPY
Asymptomatic carriage	Luminal agent: iodoquinol (650-mg tablets), 650 mg tid for 20 days; *or* paromomycin (250-mg tablets), 500 mg tid for 10 days
Acute colitis	Metronidazole (250- or 500-mg tablets), 750 mg PO or IV tid for 5–10 days; *or* tinidazole, 2 g/d PO for 3 days ***plus*** Luminal agent as above
Amebic liver abscess	Metronidazole, 750 mg PO or IV for 5–10 days; *or* tinidazole, 2 g PO once; *or* ornidazole,[a] 2 g PO once ***plus*** Luminal agent as above

[a] Not available in the United States.

a luminal agent, since metronidazole does not eradicate cysts. Resistance to metronidazole has been selected in the laboratory but has not been found in clinical isolates. Relapses are not uncommon and probably represent reinfection or failure to eradicate amebae from the bowel because of an inadequate dosage or duration of therapy.

AMEBIC LIVER ABSCESS

Metronidazole is the drug of choice for amebic liver abscess. Longer-acting nitroimidazoles (tinidazole and ornidazole) have been effective as single-dose therapy in developing countries. With early diagnosis and therapy, mortality rates from uncomplicated amebic liver abscess are <1%. There is no evidence that combined therapy with two drugs is more effective than the single-drug regimen. Studies of South Africans with liver abscesses demonstrated that 72% of patients without intestinal symptoms had bowel infection with *E. histolytica*; thus, all treatment regimens should include a luminal agent to eradicate cysts and prevent further transmission. Amebic liver abscess recurs rarely.

More than 90% of patients respond dramatically to metronidazole therapy with decreases in both pain and fever within 72 h. Indications for aspiration of liver abscesses are (1) the need to rule out a pyogenic abscess, particularly in patients with multiple lesions; (2) the lack of a clinical response in 3–5 days; (3) the threat of imminent rupture; and (4) the need to prevent rupture of left-lobe abscesses into the pericardium. There is no evidence that aspiration, even of large abscesses (up to 10 cm), accelerates healing. Percutaneous drainage may be successful even if the liver abscess has already ruptured. Surgery should be reserved for instances of bowel perforation and rupture into the pericardium.

PREVENTION

Amebic infection is spread by ingestion of food or water contaminated with cysts. Since an asymptomatic carrier may excrete up to 15 million cysts per day, prevention of infection requires adequate sanitation and eradication of cyst carriage. In high-risk areas, infection can be minimized by the avoidance of unpeeled fruits and vegetables and the use of bottled water. Because cysts are resistant to readily attainable levels of chlorine, disinfection by iodination (tetraglycine hydroperiodide) is recommended. There is no effective prophylaxis.

INFECTION WITH FREE-LIVING AMEBAE

EPIDEMIOLOGY

There are multiple genera of free-living amebae, but the major human pathogens are *Acanthamoeba*, *Naegleria*, and *Balamuthia*. All of these parasites can cause serious central nervous system (CNS) infections, which are almost always fatal. *Acanthamoeba* and *Naegleria* are distributed throughout the world and have been isolated from a wide variety of fresh and brackish water, including water from taps, lakes, hot springs, swimming pools, heating and air-conditioning units, and hospital water networks, and even from the nasal passages of healthy children. Encystation may protect these protozoa from desiccation and food deprivation. The persistence of *Legionella pneumophila* in water supplies is attributable in part to chronic infection of free-living amebae, particularly *Acanthamoeba*. In vitro studies have suggested that several pathogens that can resist phagosome-mediated killing may be able to survive within water systems in free-living amebae. These pathogens include *Pseudomonas aeruginosa*, nontuberculous *Mycobacteria* (both slow-growing species—e.g., those in the *Mycobacterium avium* complex, *M. kansasii*, and *M. gordonae*—and rapid-growing species—e.g., *M. chelonae and M. abscessus*), and viruses such as adenoviruses and echoviruses.

In contrast, the environmental niche of free-living amebae of the genus *Balamuthia* appears to be soil. A soil sample from a flowerpot was linked to a fatal infection in a child. Cases have been reported from all continents except Africa, but the majority of cases are from warm, dry areas of the southwestern United States and Latin America.

With better recognition of these pathogens, additional risk factors have been identified. Since 2010, five cases of *Naegleria fowleri* infection have been reported in northern U.S. states and have been associated with exposure to piped water, which represents a new ecologic niche. Since 2009, three clusters of *Balamuthia mandrillaris* infections have been associated with organ transplantation. *Acanthamoeba* species have caused large outbreaks of microbial keratitis associated with contact lens wear.

NAEGLERIA INFECTIONS

Primary amebic meningoencephalitis (PAM) is a fulminant CNS infection caused by the free-living ameba *Naegleria fowleri*, which thrives in warm freshwater of lakes and rivers. In the United States, 157 cases of PAM were reported from 1962 through 2022 for a total of 381 cases worldwide. Historically in southern states and more frequently during the summer, the number of infections reported annually has remained stable (0–8). However, cases in northern states have been on the rise. In 2010, a PAM case was reported for the first time from Minnesota; this case was followed by additional cases from Minnesota, Indiana, and Kansas in 2011 and 2012. With climate change, other areas may be at risk because of higher temperatures. It commonly affects healthy, male individuals (75%) with a median age of 14 years (range, 1 month–85 years). Most patients (91%) were exposed to recreational freshwater from lakes, reservoirs, rivers, streams, or ditches. The remaining cases were due to tap-water exposure through nasal irrigation with a neti pot, playing on a backyard waterslide, and swimming in a poorly maintained pool.

Despite evidence of extra-CNS dissemination of *N. fowleri* in PAM patients post-mortem, transplantation of their solid organs have not been associated with PAM in transplant recipients.

The risk of PAM in transplant recipients needs to be carefully considered in individual cases.

PAM occurs after *N. fowleri* trophozoites enter the CNS through the olfactory neuroepithelium following inhalation of water or dust contaminated with trophozoites or cysts. Trophozoites then migrate to the brain, avoiding the protective blood-brain barrier and invoking a neutrophilic response. After an incubation period of 2–15 days, severe headache, high fever, nausea, vomiting, and meningismus develop. Photophobia and palsies of the third, fourth, and sixth cranial nerves are common. Rapid progression to seizures and coma may follow. The prognosis is uniformly poor: most patients die within a week.

The diagnosis of *Naegleria* infection should be considered in any patient who has purulent meningitis without evidence of bacteria on Gram's staining, antigen detection assay, and culture. Other laboratory findings resemble those for fulminant bacterial meningitis, with elevated intracranial pressure, high white blood cell counts (up to 20,000/μL), and elevated protein concentrations and low glucose levels in cerebrospinal fluid (CSF). Diagnosis depends on the detection of motile trophozoites in wet mounts of fresh spinal fluid. Antibodies to *Naegleria* species have been detected in healthy adults; thus, serologic testing is not useful in the diagnosis of acute infection. Diagnostic PCR and histochemical staining of biopsies are available through the CDC.

A number of antimicrobial agents have in vitro activity against *N. fowleri*, but the prognosis remains poor. The few survivors have been treated with different combinations of amphotericin B, fluconazole, azithromycin, rifampin, and dexamethasone. The new antiparasitic agent miltefosine—an alkylphosphocholine compound used to treat breast cancer and visceral leishmaniasis—is active in vitro against *Naegleria*, *Acanthamoeba*, and *Balamuthia* and is now commercially available (*impavido.com*). All seven survivors among reported worldwide cases (381) included miltefosine as well as the five drugs above. Early diagnosis, prompt combination therapy including miltefosine, and aggressive management of neurologic complications, including therapeutic hypothermia, are important factors in better outcomes. A clinician whose patient may have PAM should contact the CDC Emergency Operations Center at (770) 488-7100 for assistance in diagnosis by PCR and treatment recommendations (which should include miltefosine).

ACANTHAMOEBA INFECTIONS

Granulomatous Amebic Encephalitis Infection with at least eight different *Acanthamoeba* species follows a more indolent course than *Naegleria* infection and typically occurs in chronically ill or debilitated patients. Risk factors include lymphoproliferative disorders, chemotherapy, glucocorticoid therapy, lupus erythematosus, post-transplant, and AIDS. Infection usually reaches the CNS hematogenously from a primary focus in the sinuses, skin, or lungs. In the CNS, the onset is insidious, and the syndrome often mimics a space-occupying lesion. Altered mental status, headache, and stiff neck may be accompanied by focal findings such as cranial nerve palsies, ataxia, and hemiparesis. Cutaneous ulcers or hard nodules containing amebae are frequently detected in AIDS patients with disseminated *Acanthamoeba* infection.

Examination of the CSF for trophozoites may be diagnostically helpful, but lumbar puncture may be contraindicated because of increased intracerebral pressure. CT frequently reveals cortical and subcortical lesions of decreased density consistent with embolic infarcts. In other patients, multiple enhancing lesions with edema may mimic the CT appearance of toxoplasmosis **(Chap. 235)**. Demonstration of the trophozoites and cysts of *Acanthamoeba* on wet mounts or in biopsy specimens establishes the diagnosis. Culture on nonnutrient agar plates seeded with *Escherichia coli* also may be helpful. A real-time PCR assay to detect *Acanthamoeba*, *Naegleria*, and *Balamuthia* as well as a fluorescein-labeled antiserum for the detection of protozoa in biopsy specimens are available from the CDC. Granulomatous amebic encephalitis in patients with AIDS or transplant recipients may have an accelerated course (with survival for only 3–40 days) because of the difficulty these individuals have in forming granulomas. Various antimicrobial agents have been used to treat *Acanthamoeba* infection, but miltefosine should be included in combination therapy.

Keratitis The incidence of keratitis caused by *Acanthamoeba* has increased in the past 20 years, in part as a result of improved diagnosis. Earlier infections were associated with trauma to the eye and exposure to contaminated water. At present, most infections are linked to extended-wear contact lenses, and rare cases are associated with laser-assisted in situ keratomileusis (LASIK). Risk factors include the use of homemade saline, the wearing of lenses while swimming, and inadequate disinfection. Since contact lenses presumably cause microscopic trauma, early corneal findings may be nonspecific. The first symptoms usually include tearing and the painful sensation of a foreign body. Once infection is established, progression is rapid. The characteristic clinical sign is an annular, paracentral corneal ring representing a corneal abscess. Deeper corneal invasion and loss of vision may follow.

The differential diagnosis includes bacterial, mycobacterial, and herpetic infection. The irregular polygonal cysts of *Acanthamoeba* **(Fig. 230-4)** may be identified in corneal scrapings or biopsy material, and trophozoites can be grown on special media. Cysts are resistant to available drugs, and the results of medical therapy have been disappointing. Some reports have suggested partial responses to polyhexamethylene biguanide (0.2%) or propamidine isethionate (0.1%) eyedrops. Severe infections usually require keratoplasty.

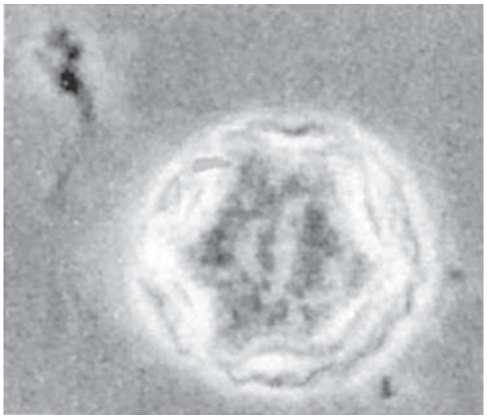

FIGURE 230-4 Double-walled cyst of *Acanthamoeba castellani*, as seen by phase-contrast microscopy. *(From DJ Krogstad et al, in A Balows et al [eds]: Manual of Clinical Microbiology, 5th ed. Washington, DC, American Society for Microbiology, 1991.)*

BALAMUTHIA INFECTIONS

Balamuthia mandrillaris is a free-living ameba that was first identified in 1986 as the cause of a fatal infection in a mandrill baboon at the Wild Animal Park in San Diego, California. Since then more than 100 cases have been reported in the United States and 200 worldwide. The parasite has been isolated from soil and dust and is probably widespread in the environment. It is an important etiologic agent of granulomatous amebic encephalitis, cutaneous lesions, and sinus infections in humans. The potential risk factors for granulomatous amebic encephalitis identified by the California Encephalitis Project include young age, immunocompromising conditions, and Hispanic ethnicity. The infection likely starts with percutaneous or mucous membrane exposure and then spreads hematogenously to the brain and other organs—a pattern that explains the risk for transmission through organ transplantation. In 2009–2012, three clusters of organ transplant–transmitted *B. mandrillaris* infections were detected by recognition of severe unexpected illness in multiple recipients from the same donor after an incubation period of 17–24 days.

Frequently, *Balamuthia* affects immunocompetent individuals, in whom the course is typically subacute, with focal neurologic signs, fever, seizures, and headaches leading to death within 1 week to several months after onset. Skin lesions may occur on the face, trunk, or extremities. In addition to dust inhalation, inoculation of trophozoites or cysts from stagnant water may occur through open wounds or mucous membranes. Diagnosis relies on examination of CSF, which reveals mononuclear or neutrophilic pleocytosis, elevated protein levels, and normal to low glucose concentrations. Amebae are rarely isolated from CSF and require culture in the presence of cell monolayers. Multiple hypodense lesions are usually detected with imaging studies **(Fig. 230-5)**. The differential diagnosis includes tuberculomas **(Chap. 183)** and neurocysticercosis **(Chap. 242)**. Testing for *Balamuthia* is part of the multiplex PCR test from the CDC. Fluorescent antibodies for serology and immunohistochemistry also are available.

The nine surviving patients in the United States have been treated with a variety of drugs, including pentamidine, flucytosine,

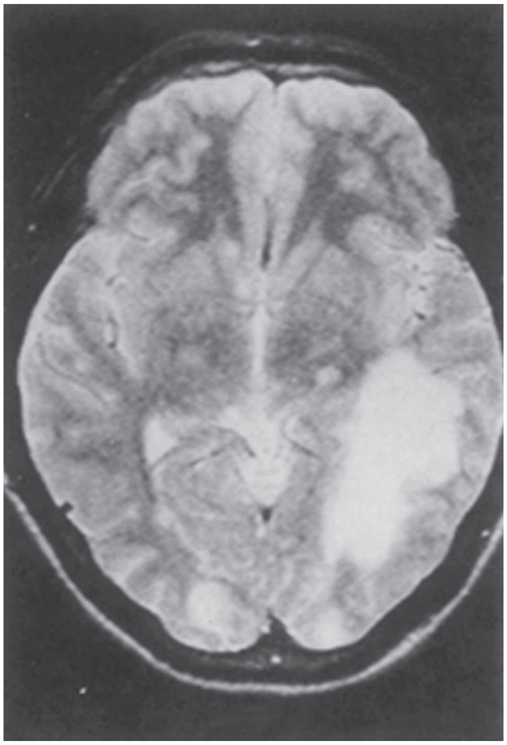

FIGURE 230-5 Brain MRI of amebic meningoencephalitis due to *Balamuthia mandrillaris*. A large lesion in the parieto-occipital lobe and other smaller lesions are seen. *(Courtesy of the Department of Radiology, UCSD Medical Center, San Diego.)*

sulfadiazine, and macrolides. The CDC recommends that miltefosine now be included, as for treatment of other free-living amebae. Clinicians should contact the CDC Emergency Operations Center at (770)-488-7100 for assistance in diagnosis and treatment.

FURTHER READING

Amebiasis

Shirley DAT et al: A review of the global burden, new diagnostics, and current therapeutics for amebiasis. Open Forum Infect Dis 5:ofy161, 2018.

Uddin MJ et al: Host protective mechanisms to intestinal amebiasis. Trends Parasitol 37:165, 2021.

Yanagawa Y et al: Clinical features and gut microbiome of asymptomatic *Entamoeba histolytica* infection. Clin Infect Dis 73:e3163, 2021.

Yanagawa Y et al: Diversity of plasticity of virulent characteristics of *Entamoeba histolytica.* Trop Med and Infect Dis 8:255, 2023.

Free-Living Amebae

Cope JR et al: The epidemiology and clinical features of *Balamuthia mandrillaris* disease in the United States, 1974–2016. Clin Infect Dis 68:1815, 2019.

Gharpure R et al: Epidemiology and clinical characteristics of primary amebic meningoencephalitis caused by *Naegleria fowleri*: A global review. Clin Infect Dis 73:e19, 2021.

Haston JC et al: The epidemiology and clinical features of non-keratitis *Acanthamoeba* infections in the United States, 1956–2020. Open Forum Infect Dis 10:ofac682, 2023.

231

Malaria

Elizabeth A. Ashley, Nicholas J. White

> *Humanity has but three great enemies: Fever, famine, and war; of these by far the greatest, by far the most terrible, is fever.*
>
> *—William Osler, 1896*

Malaria is a protozoan disease transmitted by the bite of infected female *Anopheles* mosquitoes and is the most important of the parasitic diseases of humans. In 2022, there were an estimated 249 million cases in 85 malaria-endemic countries and 608,000 deaths (i.e., ~1660 deaths each day). Two countries had an estimated 43% of these deaths: Nigeria (31%) and Democratic Republic of the Congo (12%). Malaria was eliminated from the United States, Canada, Europe, and Russia >50 years ago, but its prevalence rose in many parts of the tropics between 1970 and 2000. In response to this rise, there was substantial investment aimed at increasing access to accurate diagnosis, effective treatments, and insecticide-treated bed nets. Between 2000 and 2015, malaria mortality rates decreased dramatically as a result of highly effective control programs in several countries; since then, however, progress has reversed and estimated global case numbers, mainly in sub-Saharan Africa, have risen steadily. Meanwhile an increasing number of countries that had low malaria transmission are now targeting malaria elimination. This ambitious goal is threatened by increasing resistance to antimalarial drugs and insecticides.

Malaria remains today, as it has been for centuries, a heavy burden on tropical communities, a threat to nonendemic countries, and a danger to travelers.

ETIOLOGY AND PATHOGENESIS

Six species of the genus *Plasmodium* cause nearly all malarial infections in humans. These are *P. falciparum*, *P. vivax*, two morphologically identical sympatric species of *P. ovale* (*curtisi* and *wallikeri*), *P. malariae*, and—in Southeast Asia—the monkey malaria parasite *P. knowlesi* **(Table 231-1)**. Occasionally humans are also infected with the monkey parasites *P. simium* (South America) and *P. cynomolgi* (Southeast Asia). While almost all deaths are caused by falciparum malaria, *P. knowlesi* and occasionally *P. vivax* can also cause severe illness. Human infection begins when a female anopheline mosquito inoculates plasmodial *sporozoites* from its salivary glands during a blood meal **(Fig. 231-1)**. These microscopic motile forms of the malaria parasite are carried rapidly via the bloodstream to the liver, where they invade hepatic parenchymal cells and begin a period of asexual reproduction. By this amplification process (known as *intrahepatic* or *preerythrocytic schizogony*), a single sporozoite may produce from 10,000 to >30,000 daughter merozoites. These few swollen infected liver cells containing the hepatic schizonts eventually burst, discharging motile *merozoites* into the bloodstream. The merozoites then invade red blood cells (RBCs) to become *trophozoites* and, in nonimmune subjects, multiply 6- to 20-fold every 48 h (*P. knowlesi*, 24 h; *P. malariae*, 72 h). When the parasites reach densities of ~50/μL of blood (~100 million parasites in total in the blood of an adult), the symptomatic stage of the infection begins. In *P. vivax* and *P. ovale* infections, a proportion of the intrahepatic forms do not divide immediately but remain inert for a period ranging from 2 weeks to ≥1 year. These dormant forms, or *hypnozoites*, are the cause of the relapses that characterize infection with these species.

Attachment of merozoites to erythrocytes is mediated via a complex interaction with several different binding ligands and specific erythrocyte surface receptors. *P. falciparum* merozoites bind via

TABLE 231-1 Characteristics of *Plasmodium* Species Infecting Humans

	FINDING FOR INDICATED SPECIES				
CHARACTERISTIC	***P. FALCIPARUM***	***P. VIVAX***	***P. OVALE*[a]**	***P. MALARIAE***	***P. KNOWLESI***
Duration of intrahepatic phase (days)	5.5	8	9	15	5.5
Number of merozoites released per infected hepatocyte	30,000	10,000	15,000	15,000	20,000
Approximate duration of erythrocytic cycle (h)	48	48	50	72	24
Red cell preference	Younger cells (but can invade cells of all ages)	Reticulocytes and cells up to 2 weeks old	Reticulocytes	Older cells	Younger cells
Morphology	Usually only ring forms; banana-shaped gametocytes	Irregularly shaped large rings and trophozoites; enlarged erythrocytes; Schüffner's dots	Infected erythrocytes, enlarged and oval with tufted ends; Schüffner's[b] dots	Band or rectangular forms of trophozoites common	Resembles *P. falciparum* (early trophozoites) or *P. malariae* (later trophozoites, including band forms)
Pigment color	Black	Yellow-brown	Dark brown	Brown-black	Dark brown
Ability to cause relapses	No	Yes	Yes	No	No

[a]Genomic studies have revealed *P. ovale* to be two sympatric species: *P. ovale curtisi* and *P. ovale wallikeri*, which are morphologically very similar but may have different incubation periods and latencies. [b]Also known as James's dots.

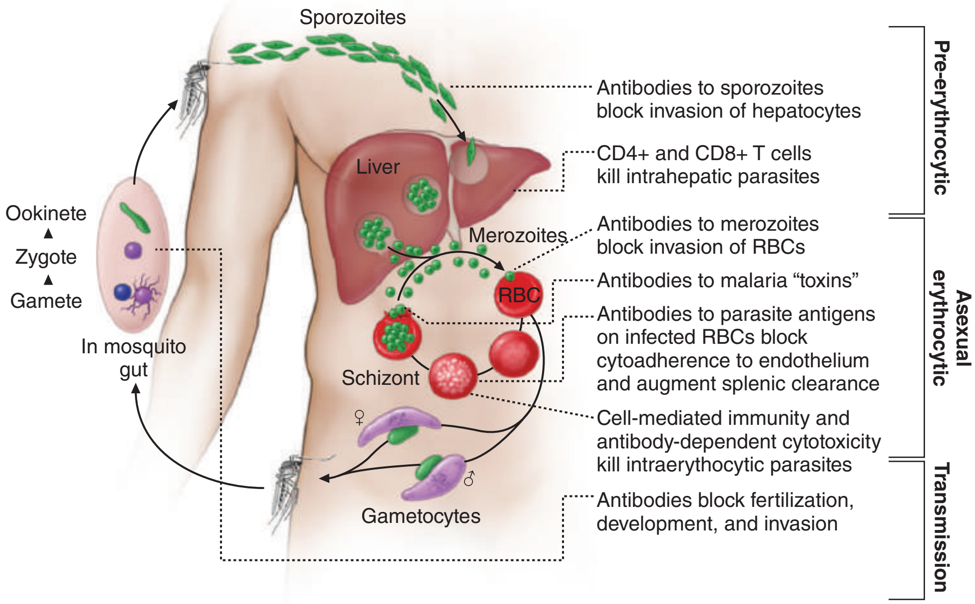
Sporozoites
Antibodies to sporozoites block invasion of hepatocytes
Liver
CD4+ and CD8+ T cells kill intrahepatic parasites
Pre-erythrocytic
Ookinete
Zygote
Gamete
In mosquito gut
Merozoites
Antibodies to merozoites block invasion of RBCs
RBC
Antibodies to malaria "toxins"
Schizont
Antibodies to parasite antigens on infected RBCs block cytoadherence to endothelium and augment splenic clearance
Asexual erythrocytic
♀
Cell-mediated immunity and antibody-dependent cytotoxicity kill intraerythocytic parasites
♂
Gametocytes
Antibodies block fertilization, development, and invasion
Transmission

FIGURE 231-1 The malaria transmission cycle from mosquito to human and targets of immunity. In *Plasmodium vivax* and *P. ovale* infections some liver stage parasites remain dormant ("hypnozoites") and awake weeks or months later to cause relapses. RBC, red blood cell.

erythrocyte- binding antigen 175 to glycophorin A and via EBA-140 to glycophorin C. The other glycophorins (B and D) also contribute. The merozoite reticulocyte-binding protein homologue 5 (PfRh5) plays a critical role binding to red cell basigin (CD147, EMMPRIN).

P. vivax binds to receptors on developing erythrocytes. The Duffy blood-group antigen Fy^a or Fy^b plays an important role in invasion of *P. vivax* and *P. knowlesi*. Most West Africans and people with origins in that region have the Duffy-negative FyFy phenotype and are generally resistant to *P. vivax* malaria.

During the first few hours of intraerythrocytic development, the small "ring forms" of the different malaria species appear similar under light microscopy. As the trophozoites enlarge, species-specific characteristics become evident, malaria pigment (hemozoin) becomes visible, and the parasite assumes an irregular or ameboid shape. By the end of the intraerythrocytic life cycle, the parasite has consumed two-thirds of the RBC's hemoglobin and has grown to occupy most of the cell. It is now called a *schizont*. Multiple nuclear divisions have taken place (*schizogony* or *merogony*). The infected RBC then ruptures to release 6–30 daughter merozoites, each potentially capable of invading a new RBC and repeating the cycle. The disease in human beings is caused by the direct effects of the asexual parasite—RBC invasion and destruction—and by the host's reaction. Some of the blood-stage parasites develop into morphologically distinct, longer-lived sexual forms (*gametocytes*) that can transmit malaria. In falciparum malaria, a delay of several asexual cycles precedes this switch to gametocytogenesis. Female gametocytes typically outnumber males by 4:1. Immature (stage I to IV) gametocytes are sequestered preferentially in the bone marrow and are also found in the spleen, brain, and intestine.

After being ingested in the blood meal of a biting female anopheline mosquito, the male gametocyte exflagellates and divides rapidly into eight motile male gametes. These fuse with female gametocytes, undergoing two rounds of sexual division (meiosis) to form a zygote in the insect's midgut. This zygote matures into an ookinete, which penetrates and encysts in the mosquito's gut wall. The resulting oocyst expands by asexual division until it bursts to liberate myriad motile sporozoites, which then migrate in the hemolymph to the salivary gland of the mosquito to await inoculation into another human at the next feed, thus completing the parasite's life cycle.

EPIDEMIOLOGY

Malaria occurs throughout most of the tropical regions of the world **(Fig. 231-2)**. *P. falciparum* predominates in Africa, New Guinea, and Hispaniola (i.e., the Dominican Republic and Haiti); *P. vivax* is more

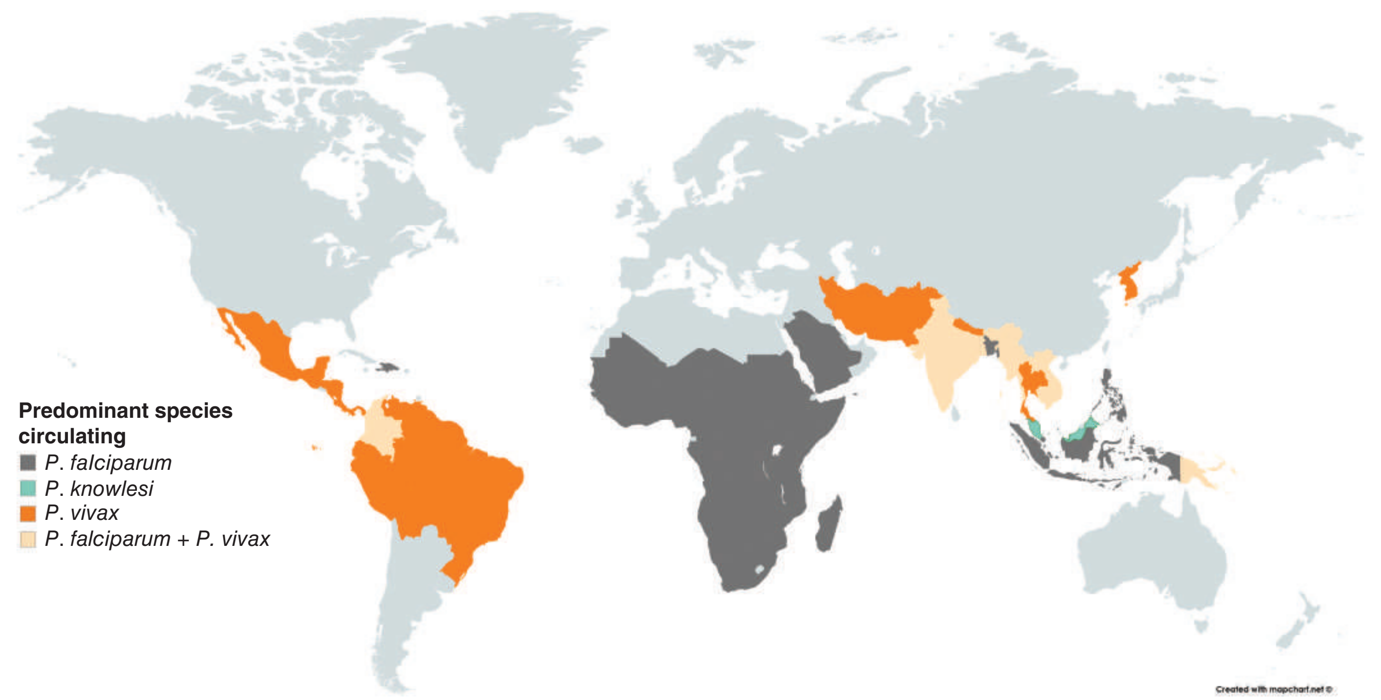
Predominant species circulating
■ P. falciparum
■ P. knowlesi
■ P. vivax
■ P. falciparum + P. vivax
Created with mapchart.net ©

FIGURE 231-2 Malaria-endemic countries showing predominant *Plasmodium* species. *Plasmodium vivax* is common in the Horn of Africa and in Mauritania but relatively unusual elsewhere in the continent.

common in Central and South America and most of Asia. The prevalence of these two species is approximately equal in Oceania. *P. malariae* is found in most endemic areas, especially throughout sub-Saharan Africa, but is much less common. *P. ovale* is relatively unusual outside of Africa. In endemic areas, submicroscopic asymptomatic infections (detectable by polymerase chain reaction [PCR]) with all human malaria parasites are much more common than microscopy-detected infections. *P. knowlesi* causes human infections commonly on the island of Borneo and, to a lesser extent, elsewhere in Southeast Asia, where the main hosts, long-tailed and pig-tailed macaques, are found.

The epidemiology of malaria is complex and may vary considerably even within relatively small geographic areas. Endemicity traditionally has been defined in terms of rates of microscopy-detected parasitemia or palpable spleens in children 2–9 years of age and has been classified as hypoendemic (<10%), mesoendemic (11–50%), hyperendemic (51–75%), and holoendemic (>75%). In holo- and hyperendemic areas (e.g., certain regions of tropical Africa or coastal New Guinea) where there is intense *P. falciparum* transmission, people may sustain one or more infectious mosquito bites per week and are infected repeatedly throughout their lives. In such settings, malaria morbidity and mortality are substantial during early childhood. Immunity against disease is hard won in these areas following repeated symptomatic infections in childhood, but, if the child survives, infections become increasingly asymptomatic. These asymptomatic older children and adults are a major source of malaria transmission. As control measures progress and urbanization expands, environmental conditions become less conducive to malaria transmission, and all age groups may lose protective immunity and become susceptible to illness. Constant, frequent, year-round infection is termed *stable transmission*. In areas where transmission is low, erratic, or focal, full protective immunity is not acquired, and symptomatic disease may occur at all ages. This situation usually exists in hypoendemic areas and is termed *unstable transmission*. Even in stable transmission areas, there is often an increased incidence of symptomatic malaria during the rainy season coinciding with increased mosquito breeding and transmission. Malaria can behave like an epidemic disease in some areas, particularly those with unstable malaria, such as northern India (the Punjab region), the Horn of Africa, Rwanda, Burundi, southern Africa, and Madagascar. Epidemics may occur when changes in environmental, economic, or social conditions (e.g., heavy rains following drought or migration—usually of refugees or workers—from a nonmalarious region to an area of high transmission) are compounded by relaxation or breakdown in malaria control and prevention services caused by underinvestment, war, or civil disorder. Epidemics often result in high mortality rates among all age groups. The principal determinants of the epidemiology of malaria are the number (density), the human-biting habits, and the longevity of the anopheline mosquito vectors. More than 100 of the >400 anopheline species can transmit malaria, but the ~40 species that do so commonly vary considerably in their efficiency as malaria vectors. More specifically, the transmission of malaria is directly proportional to the density of the vector, the square of the number of human bites per day per mosquito, and the tenth power of the probability of the mosquito's surviving for 1 day. Mosquito longevity is particularly important as a determinant of malaria transmissibility because the parasite's development within the mosquito—from gametocyte ingestion to subsequent inoculation (*sporogony*)—lasts 8–30 days, depending on ambient temperature. In order to transmit malaria, the vector mosquito must therefore survive for >7 days. Sporogony is not completed at cooler temperatures—i.e., <16°C (<60.8°F) for *P. vivax* and <21°C (<69.8°F) for *P. falciparum*; thus, transmission does not occur below these temperatures or at high altitudes. Global warming is resulting in malaria transmission at higher altitudes than previously. The most effective mosquito vectors of malaria are those, such as the *Anopheles gambiae* species complex in Africa, that are long-lived, occur in high densities in tropical climates, breed readily, and bite humans in preference to other animals. The entomologic inoculation rate (i.e., the number of sporozoite-positive mosquito bites per person per year) is the most common measure of malaria transmission and varies from <1 in some parts of Latin America and Southeast Asia to >300 in parts of tropical Africa.

PATHOPHYSIOLOGY

ERYTHROCYTE CHANGES

After invading an erythrocyte, the growing malarial parasite progressively consumes and degrades intracellular proteins, principally hemoglobin. The potentially toxic heme is detoxified by lipid-mediated crystallization to biologically inert hemozoin (malaria pigment). The parasite also alters the RBC membrane by changing its transport properties, exposing cryptic surface antigens, and inserting new parasite-derived proteins. The RBC becomes more irregular in shape, more antigenic, and less deformable.

In *P. falciparum* infections, membrane protuberances appear on the erythrocyte's surface 12–15 h after cell invasion. These "knobs" extrude a high-molecular-weight, antigenically variant, strain-specific erythrocyte membrane adhesive protein (PfEMP1) that mediates attachment to receptors on venular and capillary endothelium (*cytoadherence*). Several vascular receptors have been identified; intercellular adhesion molecule 1 and endothelial protein C receptor are important in the brain, VAR2CSA (which binds chondroitin sulfate A) predominates in the placenta, and CD36 binds parasitized RBCs in most other organs. Erythrocytes containing more mature parasites stick inside and eventually block capillaries and venules. Infected RBCs may also adhere to uninfected RBCs (to form rosettes) and to other parasitized erythrocytes (agglutination). The processes of cytoadherence, rosetting, and agglutination are central to the pathogenesis of falciparum malaria. They result in the sequestration of infected RBCs in vital organs (particularly the brain), where they interfere with microcirculatory flow and metabolism. Sequestered *P. falciparum* parasites continue to develop out of reach of the principal host defense mechanism: splenic processing and filtration. As a consequence, only the younger ring forms of the asexual *P. falciparum* parasites circulate in the peripheral blood, and the level of peripheral parasitemia variably underestimates the true number of parasites within the body. In severe malaria, uninfected erythrocytes also become less deformable, which compromises their passage through the partially obstructed capillaries and venules and shortens their survival.

In the other human malarias, significant sequestration does not occur, and all stages of the parasite's development are evident on peripheral-blood smears. *P. vivax* and *P. ovale* show a marked predilection for young RBCs and *P. malariae* for old cells; these species produce a level of parasitemia that seldom exceeds 2%. In contrast, *P. falciparum* can invade erythrocytes of all ages and may be associated with very high parasite densities. Dangerously high parasite densities may also occur in *P. knowlesi* infections, with rapid increases as a result of the shorter (24-h) asexual life cycle.

HOST RESPONSE

Initially, the host responds to malaria infection by activating nonspecific defense mechanisms. Splenic immunologic and filtrative clearance functions are augmented, and the removal of both parasitized and uninfected erythrocytes is accelerated. The spleen also removes damaged ring-form parasites (a process known as "pitting") from within the red cell and returns the once-infected cells back to the circulation, where their survival is shortened. The parasitized cells escaping splenic removal are destroyed when the schizont ruptures. The material released induces monocyte/macrophage activation and the release of proinflammatory cytokines, which cause fever and other pathologic effects. Temperatures of ≥40°C (≥104°F) damage mature parasites; in untreated infections, the effect of such temperatures is to further synchronize the parasitic cycle, with eventual production of the regular fever spikes and rigors that originally characterized the different malarias. These regular fever patterns (*quotidian*, daily; *tertian*, every 2 days; *quartan*, every 3 days) are seldom seen today as patients receive prompt and effective antimalarial treatment.

The geographic distributions of the thalassemias, sickle cell disease, hemoglobins C and E, hereditary ovalocytosis, and glucose-6-phosphate dehydrogenase (G6PD) deficiency closely resemble that of falciparum malaria before the introduction of control measures. This similarity suggests that these genetic disorders confer protection against death from falciparum malaria. HbA/S heterozygotes (sickle cell trait) have a sixfold reduction in the risk of dying from severe falciparum malaria and are correspondingly protected from the bacterial infections that complicate malaria. Hemoglobin S–containing RBCs impair parasite growth at low oxygen tensions, and *P. falciparum*–infected RBCs containing hemoglobin S or C exhibit reduced cytoadherence because of reduced surface presentation of the adhesin PfEMP1. Parasite multiplication in HbA/E heterozygotes is reduced at high parasite densities. In Melanesia, children with α-thalassemia have more frequent malaria (both vivax and falciparum) in the early years of life, which appears to protect them against severe disease. In Melanesian ovalocytosis and in Africans with the Dantu blood group, rigid erythrocytes resist merozoite invasion. G6PD deficiency provides some protection against severe *P. falciparum* infections but has a much stronger protective effect against *P. vivax* infections.

Nonspecific host defense mechanisms stop the infection's expansion, and the subsequent strain-specific immune response then controls the infection. Eventually, exposure to sufficient strains confers protection from high-level parasitemia and disease but not from infection (*premunition*). Asymptomatic parasitemia is very common among adults and older children living in regions with stable and intense malaria transmission (i.e., holo- or hyperendemic areas) and also in parts of low-transmission areas. Parasitemia in asymptomatic infections fluctuates in density but often averages ~5000/mL—which is just below the level of microscopy detection but sufficient to generate transmissible densities of gametocytes. Immunity is mainly specific for both the species and the strain of infecting malarial parasite. Both humoral immunity and cellular immunity are necessary for protection, but the mechanisms of each are incompletely understood (Fig. 231-1). Immune individuals have a polyclonal increase in serum levels of IgM, IgG, and IgA, although much of this antibody is unrelated to protection. Antibodies to a variety of parasite antigens presumably act in concert to limit in vivo replication of the parasite. In *P. falciparum* infections, the variant surface adhesin PfEMP1 is the most important of these antigens. Passive transfer of maternal antibody contributes to the partial protection of infants from severe malaria in the first months of life. This complex immunity to disease declines when a person lives outside an endemic area for several months or longer.

Several factors retard the development of cellular immunity to malaria. These factors include the absence of major histocompatibility antigens on the surface of infected RBCs, which precludes direct T-cell recognition; malaria antigen–specific immune unresponsiveness; and the enormous strain diversity of malarial parasites, along with the ability of the parasites to express variant immunodominant antigens on the erythrocyte surface that change during the course of infection. Parasites may persist in the blood for months or years (or, in the case of *P. malariae*, for decades) if treatment is not given. The complexity of the immune response in malaria, the sophistication of the parasites' evasion mechanisms, and the lack of a good in vitro correlate with clinical immunity have all slowed progress toward effective vaccines.

CLINICAL FEATURES

Malaria is a common cause of fever in tropical countries. Clinical diagnosis is notoriously unreliable. The first symptoms of malaria are nonspecific; the lack of a sense of well-being, headache, fatigue, abdominal discomfort, and muscle aches followed by fever are all similar to the symptoms of a minor viral illness. In some instances, a prominence of headache, chest pain, abdominal pain, cough, arthralgia, myalgia, or diarrhea may suggest another diagnosis. Although headache may be severe in malaria, the neck stiffness and photophobia of meningitis do not occur. While myalgia may be prominent, it is not usually as severe as in dengue fever, and the muscles are not tender as in leptospirosis or typhus. Nausea, vomiting, and orthostatic hypotension are common. The classic malarial paroxysms, in which fever spikes, chills, and rigors occur at regular intervals, are unusual and at presentation suggest infection (often relapse) with *P. vivax* or *P. ovale*. The temperature of nonimmune individuals and children often rises above 40°C (104°F), with tachycardia and sometimes delirium. Childhood febrile convulsions may occur with any of the malarias, but generalized seizures are associated specifically with falciparum malaria and may herald the development of encephalopathy (cerebral malaria). Many clinical abnormalities have been described in acute malaria, but most patients with uncomplicated infections have few abnormal physical findings other than fever, malaise, mild anemia, and (in some cases) a palpable spleen. Anemia is common among young children living in areas with stable transmission (e.g., much of West Africa) and increases in prevalence where resistance has compromised the efficacy of antimalarial drugs. Frequent vivax malaria relapse is an important cause of anemia in young children in some areas (e.g., on the island of New Guinea). In nonimmune individuals with acute malaria, the spleen takes several days to become palpable, but splenic enlargement is found in a high proportion of otherwise healthy individuals in malaria-endemic areas and reflects repeated infections. Slight enlargement of the liver is also common, particularly among young children. Jaundice may develop in patients with otherwise uncomplicated malaria and usually resolves over 1–3 weeks. Malaria is not associated with a rash. Petechial hemorrhages in the skin or mucous membranes—features of viral hemorrhagic fevers and leptospirosis—develop only very rarely in severe falciparum malaria, although thrombocytopenia is usual.

SEVERE FALCIPARUM MALARIA

Appropriately and promptly treated, uncomplicated falciparum malaria (i.e., where the patient can sit or stand unaided and can swallow medicines and food) carries a mortality rate of <0.1%. However, once vital-organ dysfunction occurs or the total proportion of erythrocytes infected increases to >2% (a level corresponding to >10^{12} parasites in an adult), mortality risk rises steeply, depending on the immunity of the host. The major manifestations of severe falciparum malaria are shown in **Table 231-2**, and features indicating a poor prognosis are listed in **Table 231-3**.

Cerebral Malaria Coma is a characteristic and ominous feature of falciparum malaria and, even with treatment, has been associated with death rates of ~20% among adults and 15% among children. Any obtundation, delirium, or abnormal behavior in falciparum malaria should be taken very seriously. The onset of coma may be gradual or sudden following a convulsion.

Cerebral malaria manifests as a diffuse symmetric encephalopathy; focal neurologic signs are unusual. Although some passive resistance to head flexion may be detected, signs of meningeal irritation are absent. The eyes may be divergent, and bruxism and a pout reflex are common, but other primitive reflexes are usually absent. The corneal reflexes are preserved, except in deep coma. Muscle tone may be either increased or decreased. The tendon reflexes are variable, and the plantar reflexes may be flexor or extensor; the abdominal and cremasteric reflexes are absent. Flexor or extensor posturing may be seen. On routine funduscopy, ~15% of patients have retinal hemorrhages; with pupillary dilation and indirect ophthalmoscopy, this figure increases to 30–40%. Papilledema (8% among children, rare among adults) and cotton wool spots (<5%) also occur. More specific funduscopic abnormalities **(Fig. 231-3)** include discrete spots of retinal opacification (30–60%) and decolorization of a retinal vessel or segment of vessel (occasional cases). Convulsions, which are usually generalized and often repeated, occur in ~10% of adults and up to 50% of children with cerebral malaria. More covert seizure activity is common, particularly among children, and may manifest as repetitive tonic–clonic eye movements or even hypersalivation. Whereas adults rarely (<3% of cases) suffer obvious neurologic sequelae, ~10% of children surviving cerebral malaria—especially those with hypoglycemia, severe anemia, repeated seizures, and deep coma—have residual neurologic deficits when they regain consciousness; hemiplegia, cerebral palsy, cortical blindness, deafness, and impaired cognition may all occur. The majority of these deficits improve markedly or resolve completely within 6 months. However, the prevalence of some other deficits increases over time; ~10% of children surviving cerebral malaria have a persistent language deficit. There may also be deficits in learning, planning and executive functions, attention, memory, and nonverbal functioning. The incidence of epilepsy is increased and life expectancy is decreased among these children.

TABLE 231-2 Manifestations of Severe Falciparum Malaria

SIGNS	MANIFESTATIONS
Major	
Unarousable coma/ cerebral malaria	Failure to localize or respond appropriately to noxious stimuli; coma persisting for >30 min after generalized convulsion; a Glasgow Coma Score <11, or in young children a Blantyre Coma Score of <3
Acidemia/acidosis	Arterial pH of <7.25, base deficit >8 meq/L, or plasma bicarbonate level of <15 mmol/L; venous lactate level of ≥5 mmol/L; manifests as labored deep breathing, often termed "respiratory distress"
Severe normochromic, normocytic anemia	Hematocrit of <15% or hemoglobin level of <50 g/L (<5 g/dL) with parasite density of >10,000/μL[a]
Renal failure	Serum or plasma creatinine level of >265 μmol/L (>3 mg/dL) or blood urea level of >20 μmol/L[b]
Pulmonary edema/ adult respiratory distress syndrome	Noncardiogenic pulmonary edema, often aggravated by overhydration; radiologically confirmed or oxygen saturation <92% on room air with a respiratory rate >30/min, often with chest wall indrawing and crepitations on auscultation
Hypoglycemia	Plasma glucose level of <2.2 mmol/L (<40 mg/dL)
Hypotension/shock	Systolic blood pressure of <70 mmHg in children 1–5 years or <80 mmHg in adults; with evidence of impaired perfusion or capillary refill >2 s
Bleeding/ disseminated intravascular coagulation	Significant bleeding and hemorrhage from the gums, nose, and gastrointestinal tract and/or evidence of disseminated intravascular coagulation
Convulsions	More than two generalized seizures in 24 h; signs of continued seizure activity, sometimes subtle (e.g., tonic-clonic eye movements without limb or face movement)
Other	
Hemoglobinuria[c]	Macroscopic black, brown, or red urine; not associated with effects of oxidant drugs and red blood cell enzyme defects (such as G6PD deficiency)
Extreme weakness	Prostration; inability to sit unaided[d]
Hyperparasitemia	Parasitemia level of >5% in nonimmune patients (>10% in any patient)
Jaundice	Serum bilirubin level of >50 mmol/L (>3 mg/dL) if combined with a parasite density of >100,000/μL or other evidence of vital-organ dysfunction

[a]This is nonspecific and may include patients with chronic anemia; a parasitemia threshold of 100,000/μL is more specific for acute malarial anemia. [b]These are criteria for adults. Lower values reflect severe malaria in children. [c]Hemoglobinuria may also occur in uncomplicated malaria and in patients with G6PD deficiency, particularly if they take oxidant drugs such as primaquine. [d]In children who are normally able to sit.

Abbreviation: G6PD, glucose-6-phosphate dehydrogenase.

TABLE 231-3 Features Indicating a Poor Prognosis in Severe Falciparum Malaria

Clinical

- Marked agitation
- Hyperventilation (respiratory distress)
- Low core temperature (<36.5°C; <97.7°F)
- Bleeding
- Deep coma
- Repeated convulsions
- Anuria
- Shock

Laboratory

- Biochemistry
  - Hypoglycemia (<2.2 mmol/L)
  - Hyperlactatemia (≥5 mmol/L)
  - Acidemia (arterial pH <7.25, base deficit >8 meq/L, or serum HCO_3 <15 mmol/L)
  - Elevated serum creatinine (>265 μmol/L)
  - Elevated total bilirubin (>50 μmol/L)
  - Elevated liver enzymes (AST/ALT 3 times upper limit of normal)
  - Elevated muscle enzymes (CPK ↑, myoglobin ↑)
  - Elevated urate (>600 μmol/L)
- Hematology
  - Leukocytosis (>12,000/μL)
  - Severe anemia (PCV <10%)
  - Coagulopathy
  - Low platelet count (<50,000/μL)
  - Prolonged prothrombin time (>3 s)
  - Prolonged partial thromboplastin time
  - Decreased fibrinogen (<200 mg/dL)
- Parasitology
  - Hyperparasitemia
  - Increased mortality at >100,000/μL
  - High mortality at >500,000/μL
  - >20% of parasites identified as pigment-containing trophozoites and schizonts
  - >5% of neutrophils contain visible malaria pigment

Note: Increased risk of concomitant bacteremia in adults if >20% parasitemia.

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; CPK, creatine phosphokinase; PCV, packed cell volume.

Hypoglycemia Hypoglycemia, an important and common complication of severe malaria, is associated with a poor prognosis and is particularly problematic in children and pregnant women. Hypoglycemia in malaria results from both a failure of hepatic gluconeogenesis and an increase in the consumption of glucose by the host and, to a much lesser extent, the malaria parasites. Hypoglycemia may be compounded by quinine, a powerful stimulant of pancreatic insulin secretion, which is still sometimes used in endemic areas for the treatment of both severe and uncomplicated falciparum malaria. In severe disease, the clinical diagnosis of hypoglycemia is difficult: the usual physical signs (sweating, gooseflesh, tachycardia) are absent, and the neurologic impairment caused by hypoglycemia cannot be distinguished from that caused by malaria.

Acidosis Acidosis, resulting from accumulation of organic acids, is an important cause of death from severe malaria, which in adults is often compounded by renal impairment. Hypovolemia is not a major contributor to acidosis. Lactic acidosis is caused by the combination of anaerobic glycolysis in tissues where sequestered parasites interfere with microcirculatory flow, lactate production by the parasites, and a failure of hepatic and renal lactate clearance. Hyperlactatemia commonly coexists with hypoglycemia. Acids of gut origin are important contributors to acidosis. In children, ketoacidosis may also contribute. Hydroxyphenyllactic acid, α-hydroxybutyric acid, and β-hydroxybutyric acid concentrations are elevated. Acidotic breathing, sometimes called "respiratory distress," is a sign of poor prognosis. It is followed often by circulatory failure refractory to volume expansion or inotropic drug treatment and ultimately by respiratory arrest. Plasma concentrations of bicarbonate or lactate are the best biochemical prognosticators in severe malaria.

Noncardiogenic Pulmonary Edema Adults with severe falciparum malaria may develop noncardiogenic pulmonary edema (adult respiratory distress syndrome) even after several days of antimalarial therapy. The mortality rate is >80%. The pathogenesis is unclear. Pulmonary edema can be readily precipitated by overly vigorous administration of IV fluid. Noncardiogenic pulmonary edema can also develop in otherwise uncomplicated vivax malaria, where recovery is usual.

Renal Impairment Acute kidney injury is common in severe falciparum malaria. Oliguric renal failure requiring temporary renal replacement therapy is an important manifestation in adults but is unusual among children. The pathogenesis of renal failure is unclear but may be related to erythrocyte sequestration and agglutination interfering with renal microcirculatory flow and metabolism. Clinically and pathologically, this syndrome manifests as acute tubular

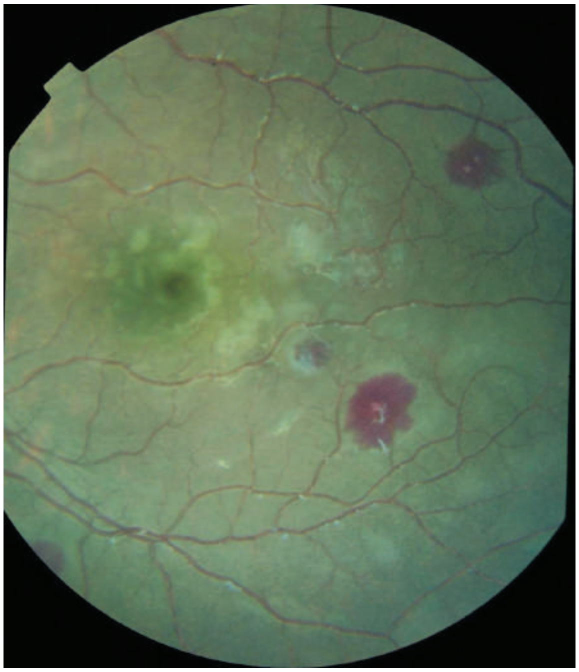

FIGURE 231-3 The eye in cerebral malaria: perimacular whitening and pale-centered retinal hemorrhages. *(Courtesy of N. Beare, T. Taylor, S. Harding, S. Lewallen, and M. Molyneux; with permission.)*

necrosis. Acute renal failure may occur simultaneously with other vital-organ dysfunction (in which case the mortality risk is high) or may progress as other disease manifestations resolve. In survivors, urine flow resumes in a median of 4 days, and serum creatinine levels return to normal in a mean of 17 days **(Chap. 321)**. Early dialysis or hemofiltration considerably improves the chances of survival, particularly in acute hypercatabolic renal failure.

Hematologic Abnormalities Anemia results from accelerated RBC removal by the spleen, obligatory RBC destruction at parasite schizogony, and ineffective erythropoiesis. In severe malaria, the deformability of both infected and uninfected RBCs is reduced. The degree of reduced deformability correlates with prognosis and with the development of anemia. Splenic clearance of all RBCs is increased. In nonimmune individuals and in areas with unstable transmission, anemia can develop rapidly and transfusion is often required. A hemoglobin of ≤3 g/dL on presentation is associated with increased mortality. Acute hemolytic anemia with massive hemoglobinuria (“blackwater fever”) may occur. Hemoglobinuria may contribute to renal injury. Some patients with blackwater fever have G6PD deficiency, but in the majority of cases, it is unclear why massive hemolysis has occurred. In nonimmune patients, sudden hemolysis may follow many days after artesunate treatment of hyperparasitemia, usually as a result of relatively synchronous loss of once-parasitized “pitted” RBCs. As a consequence of repeated malarial infections, children in high-transmission areas are usually anemic and often develop severe anemia. This results from both shortened survival of uninfected RBCs and marked dyserythropoiesis. Anemia is a common consequence of antimalarial drug resistance, which results in repeated or continued infection.

Slight coagulation abnormalities are common in falciparum malaria, and mild thrombocytopenia is usual (a normal platelet count should question the diagnosis of malaria). Fewer than 5% of patients with severe malaria have significant bleeding with evidence of disseminated intravascular coagulation. Hematemesis from stress ulceration or acute gastric erosions may occur rarely.

Liver Dysfunction Mild hemolytic jaundice is common in malaria. Severe jaundice is associated with *P. falciparum* infections; is more common among adults than among children; and results from hemolysis, hepatocyte injury, and cholestasis. Liver failure does not occur. When accompanied by other vital-organ dysfunction (often renal impairment), liver dysfunction carries a poor prognosis. Hepatic dysfunction contributes to hypoglycemia, lactic acidosis, and impaired drug metabolism. Occasional patients with falciparum malaria may develop deep jaundice (with hemolytic, hepatic, and cholestatic components) without evidence of other vital-organ dysfunction, in which case the prognosis is good.

TABLE 231-4 Relative Incidence of Severe Complications of Falciparum Malaria

COMPLICATION	NONPREGNANT ADULTS	PREGNANT WOMEN	CHILDREN
Anemia	+	++	+++
Convulsions	+	+	+++
Hypoglycemia	+	+++	+++
Jaundice	+++	+++	+
Renal failure	+++	+++	+
Pulmonary edema	++	+++	+

Note: –, rare; +, infrequent; ++, frequent; +++, very frequent.

Other Complications HIV/AIDS and malnutrition predispose to more severe malaria in nonimmune individuals. Malaria anemia is worsened by concurrent infections with intestinal helminths, hookworm in particular. Approximately 6% of children diagnosed with severe malaria have concomitant bacteremia. In adults, the proportion is lower (<1%), except in those with very high parasite counts (>20% parasitemia). In areas of moderate and high malaria transmission, differentiating severe malaria from sepsis with incidental parasitemia in childhood is very difficult, and severe malaria is overdiagnosed. In endemic areas, *Salmonella* spp. bacteremia has been associated specifically with *P. falciparum* infections. Chest infections and catheter-associated urinary tract infections are common among patients who are unconscious for >3 days. Aspiration pneumonia may follow generalized convulsions. The frequencies of complications of severe falciparum malaria are summarized in **Table 231-4**.

MALARIA IN PREGNANCY

Malaria in early pregnancy causes fetal loss. In areas of high malaria transmission, falciparum malaria in primi- and secundigravid women is associated with low birth weight (average reduction, ~170 g) and consequently increased infant mortality rates. In general, infected mothers in areas of stable transmission remain asymptomatic despite intense accumulation of parasitized erythrocytes in the placental microcirculation. Maternal HIV infection predisposes pregnant women to more frequent and higher-density malaria infections, predisposes their newborns to congenital malarial infection, and exacerbates the reduction in birth weight associated with malaria.

In areas with unstable transmission of malaria, pregnant women are prone to severe infections and are particularly likely to develop high *P. falciparum* parasitemias complicated by anemia, hypoglycemia, and acute pulmonary edema. Fetal distress, premature labor, and stillbirth or low birth weight are common results. Fetal death is common in severe malaria. Congenital malaria occurs in <5% of newborns of infected mothers; its frequency and the level of parasitemia are related directly to the timing of maternal infection and the parasite density in maternal blood and in the placenta. *P. vivax* malaria in pregnancy is also associated with a reduction in birth weight (average, 110 g), but in contrast to falciparum malaria, this effect is more pronounced in multigravid than in primigravid women. About 300,000 women die in childbirth yearly, with most deaths occurring in low-income countries; maternal death from hemorrhage at childbirth is correlated with malaria-induced anemia.

MALARIA IN CHILDREN

Most of the estimated >600,000 deaths from falciparum malaria each year are in young African children. Convulsions, coma, hypoglycemia, metabolic acidosis, and severe anemia are relatively common among children with severe malaria, whereas deep jaundice, oliguric acute kidney injury, and acute pulmonary edema are unusual. Severely anemic children may present with labored deep breathing, which previously has been attributed incorrectly to "anemic congestive cardiac failure" but is usually caused by metabolic acidosis. In general, children tolerate antimalarial drugs well and respond rapidly to treatment. Younger children, and in particular malnourished children, require higher body weight–adjusted antimalarial drug doses than older children or adults.

TRANSFUSION MALARIA

Malaria can be transmitted by blood transfusion, needlestick injury, or organ transplantation. The incubation period is often short because there is no preerythrocytic stage of development, and thus there are no relapses of *P. vivax* and *P. ovale* infections. The clinical features and management of these cases are the same as for naturally acquired infections, although primaquine is not needed for radical cure of vivax or ovale malaria as there are no liver stages.

CHRONIC COMPLICATIONS OF MALARIA

HYPERREACTIVE MALARIAL SPLENOMEGALY

Chronic or repeated malarial infections produce hypergammaglobulinemia; normochromic, normocytic anemia; and, in certain situations, splenomegaly. Some residents of malaria-endemic areas in tropical countries exhibit an abnormal immunologic response to repeated infections that is characterized by massive splenomegaly, hepatomegaly, marked elevations in serum IgM and malarial antibody titers, hepatic sinusoidal lymphocytosis, and (in Africa) peripheral B-cell lymphocytosis. This syndrome has been associated with the production of cytotoxic IgM antibodies to CD8+ T lymphocytes, antibodies to CD5+ T lymphocytes, and an increase in the ratio of CD4+ to CD8+ T cells. These events may lead to uninhibited B-cell production of IgM and the formation of cryoglobulins (IgM aggregates and immune complexes). This immunologic process stimulates lymphoid hyperplasia and clearance activity and eventually produces splenomegaly. Patients with hyperreactive malarial splenomegaly present with an abdominal mass or a dragging sensation in the abdomen and occasional sharp abdominal pains suggesting perisplenitis. There is usually anemia and some degree of pancytopenia (hypersplenism). In some cases, malaria parasites cannot be found in peripheral-blood smears by microscopy. Respiratory and skin infections are common and many patients die of overwhelming sepsis. Persons with hyperreactive malarial splenomegaly living in endemic areas should receive antimalarial chemoprophylaxis; the results are usually good. In nonendemic areas, antimalarial treatment is advised. Some cases have been mistaken for hematologic malignancy. However, in other cases refractory to therapy, clonal lymphoproliferation may develop, and this can evolve into a malignant lymphoproliferative disorder.

QUARTAN MALARIAL NEPHROPATHY

Chronic or repeated infections with *P. malariae* (and possibly with other malarial species) may cause soluble immune complex injury to the renal glomeruli, resulting in the nephrotic syndrome. Other unidentified factors must contribute to this process since only a very small proportion of infected patients develop renal disease. The histologic appearance is that of focal or segmental glomerulonephritis with splitting of the capillary basement membrane. Subendothelial dense deposits are seen on electron microscopy, and immunofluorescence reveals deposits of complement and immunoglobulins and *P. malariae* antigens are often visible. A coarse-granular pattern of basement membrane immunofluorescent deposits (predominantly IgG3) with selective proteinuria carries a better prognosis than a fine-granular, predominantly IgG2 pattern with nonselective proteinuria. Quartan nephropathy is mainly a disease of children and is rarely reported nowadays. It usually responds poorly to treatment with either antimalarial agents or glucocorticoids and cytotoxic drugs.

BURKITT'S LYMPHOMA AND EPSTEIN-BARR VIRUS INFECTION

It is possible that malaria-related immune dysregulation provokes infection with lymphoma viruses. Childhood Burkitt's lymphoma is strongly associated with Epstein-Barr virus (EBV) and with high transmission of *P. falciparum*. Chronic *P. falciparum* malaria drives large numbers of EBV-infected cells through the lymph node germinal centers and deregulates activation-induced cytidine deaminase, resulting in DNA damage, c-myc translocations, and sometimes lymphoma.

DIAGNOSIS OF MALARIA

When a patient in or from a malarious area presents with fever, thick and thin blood smears should be prepared and *examined immediately* to confirm the diagnosis and identify the species of infecting parasite **(Figs. 231-4 through 231-10)**. In general, if the blood smear is negative when examined by an experienced microscopist, the patient does not have malaria. If reliable microscopy is not available, a rapid test should be performed. Malaria is not a clinical diagnosis.

DEMONSTRATION OF THE PARASITE

The definitive diagnosis of malaria rests on the demonstration of asexual forms of the parasite in stained peripheral-blood smears. Of the Romanowsky stains, Giemsa at pH 7.2 is preferred; Field's, Wright's, or Leishman's stain can also be used. Staining of parasites with the fluorescent dye acridine orange allows more rapid diagnosis of malaria (but not speciation of the infection) in patients with low-level parasitemia.

Both thin (Figs. 231-4, 231-5, 231-10) and thick (Figs. 231-6, 231-7, 231-8, and 231-9) blood smears should be examined. The thin blood smear should be air-dried, fixed in anhydrous methanol, then stained; the RBCs in the tail of the film should then be examined under oil immersion (×1000 magnification). The density of parasitemia is expressed as the number of parasitized erythrocytes per 1000 RBCs. The thick blood film should be of uneven thickness. The smear should be dried thoroughly and stained without fixing. As many layers of

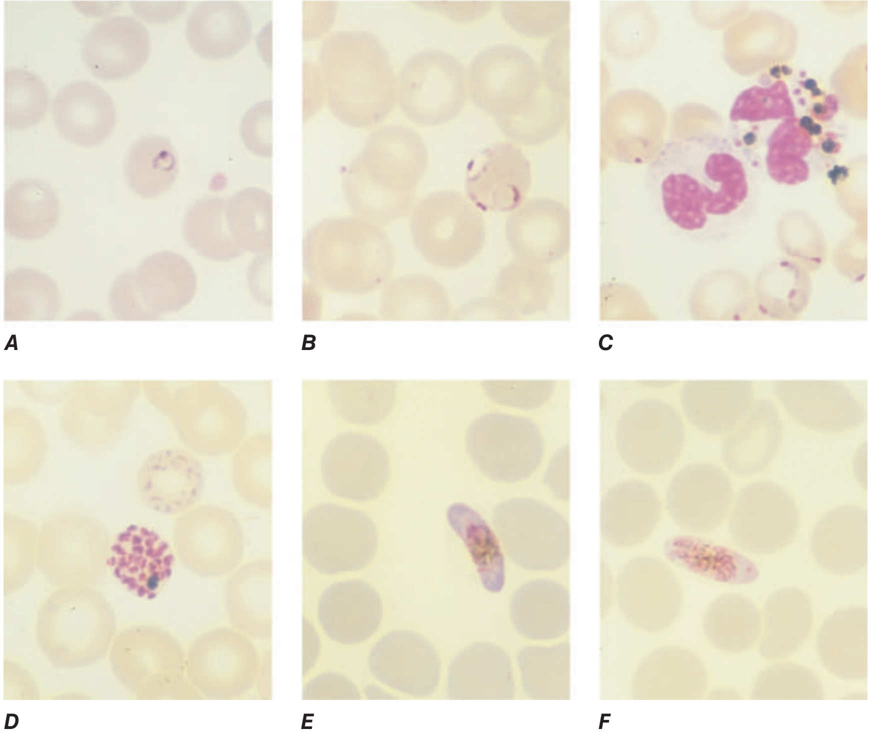

FIGURE 231-4 Thin blood films of *Plasmodium falciparum*. ***A.*** Young trophozoite. ***B.*** Old trophozoite. ***C.*** Trophozoites in erythrocytes and pigment in polymorphonuclear cells. ***D.*** Mature schizont. ***E.*** Female gametocyte. ***F.*** Male gametocyte. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

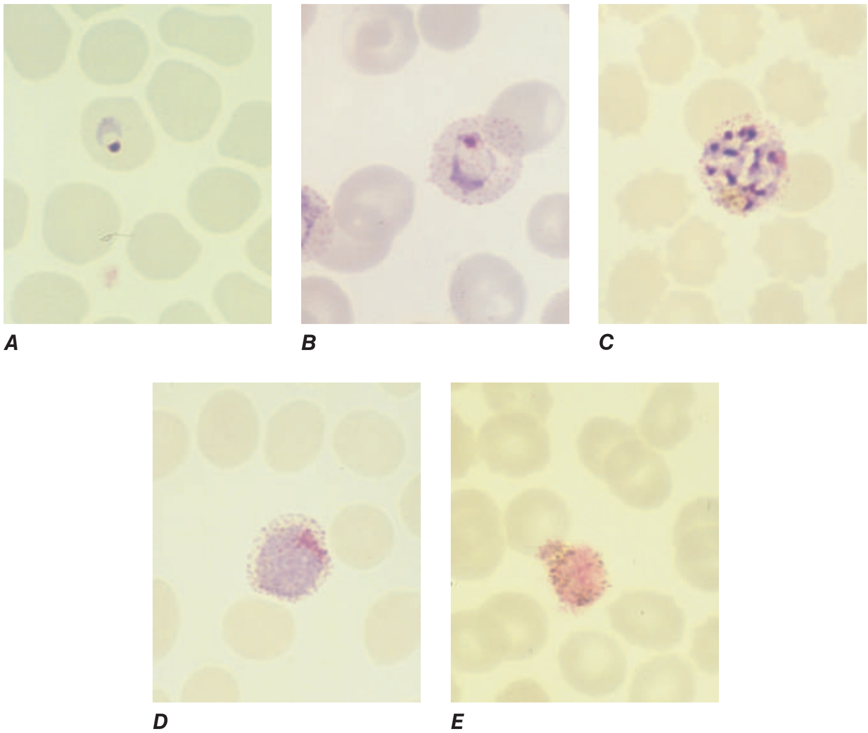

FIGURE 231-5 Thin blood films of *Plasmodium vivax*. *A.* Young trophozoite. ***B.*** Old trophozoite. ***C.*** Mature schizont. ***D.*** Female gametocyte. ***E.*** Male gametocyte. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

erythrocytes overlie one another and are lysed during the staining procedure, the thick film has the advantage of concentrating the parasites (by 40- to 100-fold compared with a thin blood film) and thus increasing diagnostic sensitivity. Both parasites and white blood cells (WBCs) are counted, and the number of parasites per unit volume is calculated from the total leukocyte count. Alternatively, a WBC count of 8000/μL is assumed. This figure is converted to the number of parasitized erythrocytes per microliter. A minimum of 200 WBCs should be counted under oil immersion. Interpretation of blood smears, particularly thick films, requires some experience because artifacts are common. Before a thick smear is judged to be negative, 100–200 fields should be examined. In high-transmission areas, the presence of up to 10,000 parasites/μL of blood may be tolerated without symptoms or signs in partially immune individuals. Thus, in these areas, the detection of low-density malaria parasitemia is sensitive but has low specificity in identifying malaria as the cause of illness. Because the prevalence of asymptomatic parasitemia is often high, low-density parasitemia is a common incidental finding in other conditions causing fever.

■ RAPID DIAGNOSTIC TESTS

Rapid, simple, sensitive, and specific antibody-based diagnostic stick or card tests that detect *P. falciparum*–specific histidine-rich protein 2 (PfHRP2), lactate dehydrogenase, or occasionally aldolase antigens in finger-prick blood samples have become the main method of malaria diagnosis in endemic areas **(Table 231-5)**. Some of these rapid diagnostic tests (RDTs) carry a second antibody (either pan-malaria or *P. vivax*–specific) and so distinguish falciparum malaria from the less dangerous malarias. PfHRP2-based RDTs may remain positive for several weeks after acute infection. This prolonged positivity is a disadvantage in high-transmission areas where infections are frequent but helps in the diagnosis of severe malaria in patients who have taken antimalarial drugs and cleared peripheral parasitemia but who still have a strongly positive PfHRP2 test. A major disadvantage of RDTs is that they do not quantify parasitemia. Widespread use of PfHRP2 RDTs has put strong selection pressure on *P. falciparum* populations in some areas, leading to an increased prevalence of mutant parasites with deletion of *PfHRP2/3* genes that are not detected by the current generation of PfHRP2-based tests. This is a particular problem in the horn of Africa but increasingly reported elsewhere.

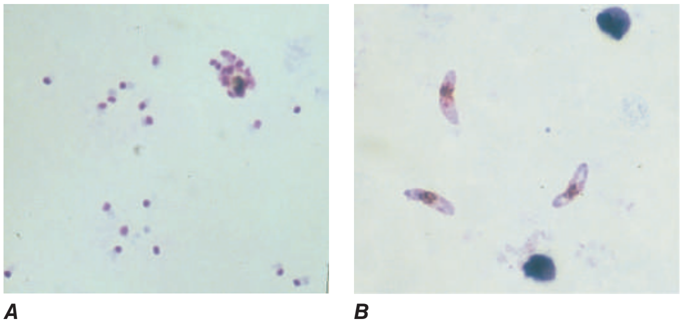

FIGURE 231-6 Thick blood films of *Plasmodium falciparum*. *A.* Trophozoites. ***B.*** Gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

The relationship between parasite density and prognosis is complex and variable; in general, patients with >10^5 parasites/μL are at increased risk of dying, but nonimmune patients may die with much lower counts, and partially immune persons may tolerate parasite densities many times higher with only minor symptoms. In severe malaria, a poor prognosis is indicated by a predominance of more mature *P. falciparum* parasites (i.e., >20% of parasites with visible pigment) in the peripheral-blood film or by the presence of phagocytosed malarial pigment in >5% of neutrophils (an indicator of recent schizogony). In *P. falciparum* infections, gametocytemia peaks 1 week after the peak of asexual parasite densities. Because the mature gametocytes of *P. falciparum* (unlike those of other plasmodia) are not affected by most antimalarial drugs, their persistence does not mean there is drug resistance or a need to re-treat if a full course of appropriate treatment has been given. Phagocytosed malarial pigment seen inside peripheral-blood monocytes may provide a clue to recent infection if malaria parasites are not detectable. After parasite clearance, this intraphagocytic malarial pigment is often evident for several days in peripheral-blood films, and for longer in bone marrow aspirates or smears of fluid expressed after intradermal puncture.

Molecular diagnosis by PCR amplification of parasite nucleic acid is more sensitive than microscopy or rapid diagnostic tests for detecting malaria parasites and defining malarial species. PCR is used in reference centers but should not be used for primary diagnosis in endemic

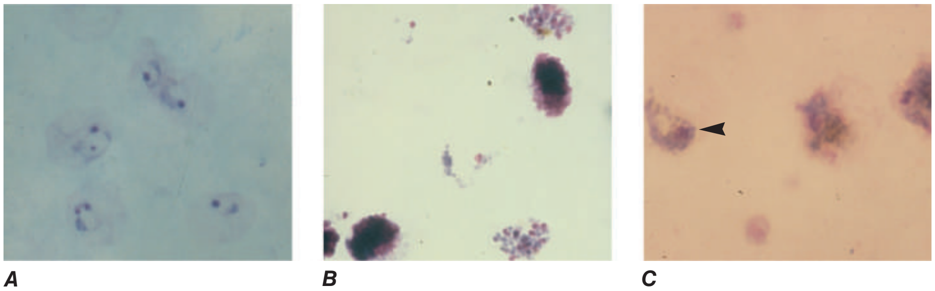

FIGURE 231-7 Thick blood films of *Plasmodium vivax*. *A.* Trophozoites. ***B.*** Schizonts. ***C.*** Gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

areas as it is too sensitive. In epidemiologic surveys, ultrasensitive PCR detection (1000 times more sensitive than microscopy) has proved very useful in identifying asymptomatic infections as control and eradication programs drive parasite prevalences down to very low levels and in identifying residual "hot spots" of transmission. Serologic diagnosis with either indirect fluorescent antibody or enzyme-linked immunosorbent assays is useful for screening of prospective blood donors and may prove useful as a measure of transmission intensity in future epidemiologic studies. Serology has no place in the diagnosis of acute illness.

■ LABORATORY FINDINGS IN ACUTE MALARIA

Normochromic, normocytic anemia is usual. The leukocyte count is generally normal, although it may be raised in very severe infections. There is slight monocytosis, lymphopenia, and eosinopenia, with reactive lymphocytosis and eosinophilia in the weeks after acute infection. The platelet count is usually reduced to ~10^5/μL (a normal platelet count may point to another diagnosis). The erythrocyte sedimentation rate, plasma viscosity, and levels of C-reactive protein and other acute-phase proteins are elevated. Severe infections may be accompanied by prolonged prothrombin and partial thromboplastin times and by more severe thrombocytopenia. Antithrombin III levels are reduced even in mild infection. In uncomplicated malaria, plasma concentrations of electrolytes, blood urea nitrogen (BUN), and creatinine are usually normal. Findings in severe malaria may include metabolic acidosis, with low plasma concentrations of glucose, sodium, bicarbonate, phosphate, and albumin, together with elevations in lactate, BUN, creatinine (adjusted for age and body mass), urate, muscle and liver enzymes, and conjugated and unconjugated bilirubin. Hypergammaglobulinemia is usual in immune and semi-immune subjects living in malaria-endemic areas. Urinalysis generally gives normal results. In adults and children with cerebral malaria, the mean cerebrospinal fluid (CSF) opening pressure at lumbar puncture is ~160 mm H_2O; usually the CSF content is normal or there is a slight elevation of total protein level (<1.0 g/L [<100 mg/dL]) and cell count (<20/μL).

TREATMENT

Malaria

Patients with severe malaria and those unable to take oral drugs should receive parenteral antimalarial therapy immediately **(Table 231-6)**. Antimalarial drug susceptibility testing can be performed but yields results too slowly to influence the choice of treatment. Molecular markers of resistance to various antimalarial drugs have been identified such as *dhfr* mutations (antifols), *Pfcrt* mutations (chloroquine), *Pfkelch13* mutations (artemisinin), *Pfpm1* or *Pfpm2* gene amplification (piperaquine) and *Pfmdr1*gene amplification (mefloquine), which are used in population surveillance studies.

The World Health Organization (WHO) recommends artemisinin-based combination therapy (ACT) as first-line treatment for uncomplicated *P. falciparum* malaria in malaria-endemic areas. An ACT comprises an artemisinin derivative (e.g., artesunate, artemether, or dihydroartemisinin), combined with a single partner drug from another antimalarial drug class in a fixed-dose coformulation. ACTs are also the recommended first-line treatment for *P. knowlesi* infections, and either chloroquine or an ACT is recommended for the other malarias. The choice of ACT depends on the likely sensitivity of the infecting parasites to the partner drug. ACTs may be unavailable in temperate countries, where treatment recommendations are limited to the registered available drugs. Despite increasing evidence of chloroquine resistance in *P. vivax* (from parts of Indonesia, Oceania, eastern and southern Asia, and Central and South America), chloroquine remains an effective treatment for *P. vivax* malaria in many areas and for *P. ovale* and *P. malariae* infections everywhere.

Artemisinin resistance in *P. falciparum* emerged in Southeast Asia in the late 2000s, where it was followed by piperaquine and mefloquine resistance in some areas. Significant artemisinin resistance is now prevalent throughout the Greater Mekong Subregion and has now emerged and spread in East Africa. As new antimalarials are still years away, it has been suggested that current ACTs should now combine two slowly eliminated partner drugs to provide mutual protection against resistance (triple ACTs). Falsified or substandard antimalarial drugs are sold in many Asian and African countries and may be the cause of treatment failures. Characteristics of antimalarial drugs are shown in **Table 231-7**.

SEVERE MALARIA

In large randomized controlled clinical trials, parenteral artesunate, a water-soluble artemisinin derivative, reduced severe falciparum malaria mortality rates by 35% in Asian adults and children and by 22.5% in African children compared with quinine treatment. Artesunate therefore is now the drug of choice for all patients with

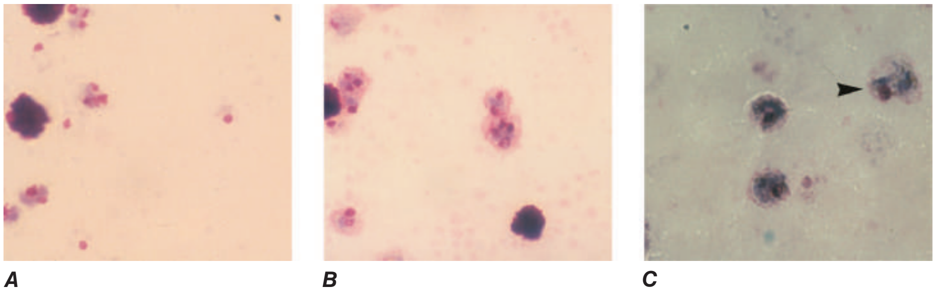

FIGURE 231-8 Thick blood films of *Plasmodium ovale*. *A.* Trophozoites. ***B.*** Schizonts. ***C.*** Gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

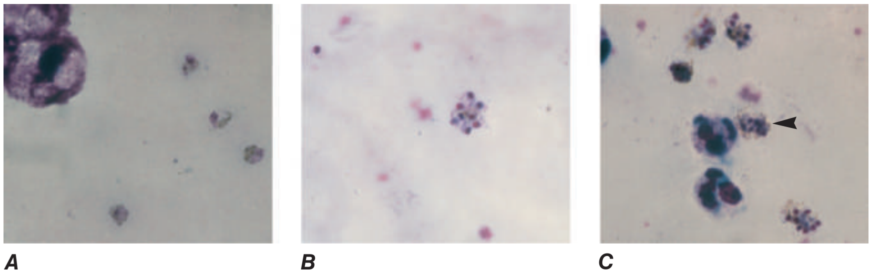

FIGURE 231-9 Thick blood films of *Plasmodium malariae*. ***A.*** Trophozoites. ***B.*** Schizonts. ***C.*** Gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

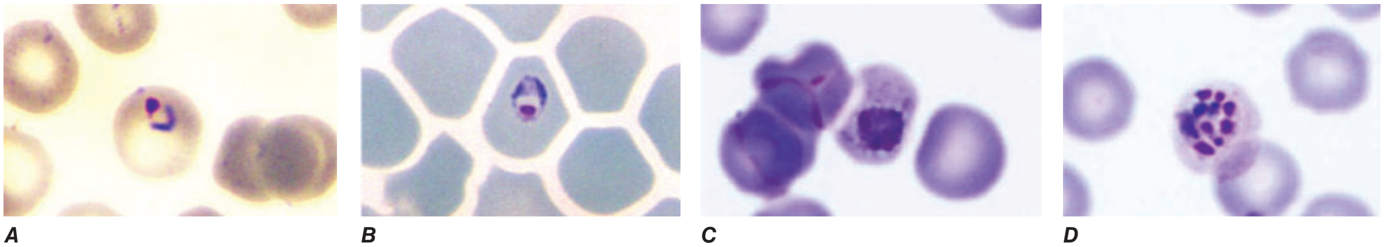

FIGURE 231-10 Thin blood films of *Plasmodium knowlesi*. ***A***. Young ring stages ***B.*** Older ring stages ***C.*** Mature trophozoites ***D.*** Schizonts (***B*** and ***C*** may look very similar to the "band forms" of *P. malariae*). *(From Kesinee Chotivanich, Cell and Tissue Culture Resources Unit laboratory and MORU, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand.)*

TABLE 231-5 Standard Methods for the Diagnosis of Malaria[a]

METHOD	PROCEDURE	ADVANTAGES	DISADVANTAGES
Thick blood film[b]	Blood should be uneven in thickness but thin enough that the hands of a watch can be read through part of the spot. Stain dried, unfixed blood spot with Giemsa, Field's, or another Romanowsky stain. Count number of asexual parasites per 200 WBCs (or per 500 WBCs at low densities). Count and report gametocytes separately.[c]	Sensitive (0.001% parasitemia); species specific; inexpensive	Requires experience (artifacts may be misinterpreted as low-level parasitemia); underestimates true count
Thin blood film[d]	Stain fixed smear with Giemsa, Field's, or another Romanowsky stain. Count number of RBCs containing asexual parasites per 1000 RBCs. In severe malaria, assess stage of parasite development and count neutrophils containing malaria pigment.[e] Count and report gametocytes separately.[c]	Rapid; species specific; inexpensive; in severe malaria, provides prognostic information[e]	Insensitive (<0.05% parasitemia); uneven distribution of *P. vivax*, as enlarged infected red cells concentrate at leading edge
PfHRP2 dipstick or card test	A drop of blood is placed on the stick or card, which is then immersed in washing solutions. Monoclonal antibody capture of parasitic antigens reads out as a colored band.	Robust and relatively inexpensive; rapid; sensitivity similar to or slightly lower than that of thick films (~0.001% parasitemia)	Detects only *Plasmodium falciparum*; remains positive for weeks after high-density infections[f]; does not quantitate *P. falciparum* parasitemia; evasion of detection by certain strains due to polymorphisms/deletions in *HRP2/3* genes
Plasmodium LDH dipstick or card test	A drop of blood is placed on the stick or card, which is then immersed in washing solutions. Monoclonal antibody capture of parasitic antigens reads out as two colored bands. One band is genus specific (all malarias) or *P. vivax* specific, and the other band is specific for *P. falciparum.*	Rapid; sensitivity similar to or slightly lower than that of thick films for *P. falciparum* (~0.001% parasitemia)	May miss low-level parasitemia with *P. vivax*, *P. ovale*, and *P. malariae* and may not speciate these organisms; does not quantitate *P. falciparum* parasitemia; lower sensitivity for detection of *P. knowlesi*, which may be misidentified as *P. falciparum*
Microtube concentration methods with acridine orange staining	Blood is collected in a specialized tube containing acridine orange, anticoagulant, and a float. After centrifugation, which concentrates the parasitized cells around the float, fluorescence microscopy is performed.	Sensitivity similar or superior to that of thick films (~0.001% parasitemia); ideal for processing large numbers of samples rapidly	Does not speciate or quantitate; requires fluorescence microscopy

[a] Malaria cannot be diagnosed clinically with accuracy, but treatment should be started on clinical grounds if laboratory confirmation is likely to be delayed. In areas of the world where malaria is endemic and transmission rates are high, low-level asymptomatic parasitemia is common in otherwise healthy people. Thus, malaria may not be the cause of a fever in a parasitemic patient, although in this context, the presence of >10,000 parasites/μL (~0.2% parasitemia) does indicate that malaria is likely to be the cause. Antibody and polymerase chain reaction (PCR) tests have no role in the diagnosis of malaria except that PCR is increasingly used for genotyping and speciation in mixed infections and for detection of low-level parasitemia in asymptomatic residents of endemic areas. [b] Asexual parasites/200 WBCs × 40 = parasite count/μL (assumes a WBC count of 8000/μL). See Figs. 231-6 through 231-10. [c] *P. falciparum* gametocytemia may persist for days or weeks after clearance of asexual parasites. Gametocytemia without asexual parasitemia does not indicate active infection. [d] Parasitized RBCs (/1000) × hematocrit × 125.6 = parasite count/μL. See Figs. 231-4, 231-5 and 231-10. [e] The presence of >100,000 parasites/μL (~2% parasitemia) is associated with an increased risk of severe malaria, but some patients have severe malaria with lower counts. At any level of parasitemia, the finding that >50% of parasites are tiny rings (cytoplasm thickness less than half of nucleus width) carries a relatively good prognosis. In a severely ill patient, the presence of visible pigment in >20% of parasites or of phagocytosed pigment in >5% of polymorphonuclear leukocytes (indicating massive recent schizogony) carries a worse prognosis. [f] Persistence of PfHRP2 is a disadvantage in high-transmission settings, where many asymptomatic people have positive tests, but can be used to diagnostic advantage in low-transmission settings when a sick patient has previously received unknown treatment (which, in endemic areas, often consists of antimalarial drugs). In this situation, a positive PfHRP2 test indicates that the illness is falciparum malaria, even if the blood smear is negative.

Abbreviations: LDH, lactate dehydrogenase; PfHRP2, *P. falciparum* histidine-rich protein 2; RBCs, red blood cells; WBCs, white blood cells.

TABLE 231-6 Regimens for the Treatment of Malaria[a]

TYPE OF DISEASE OR TREATMENT	REGIMEN(S)
Uncomplicated Malaria	
Known chloroquine-sensitive strains of *Plasmodium vivax*, *P. malariae*, *P. ovale*, *P. falciparum*[b]	Chloroquine (10 mg of base/kg stat followed by 5 mg/kg at 12, 24, and 36 h *or* by 10 mg/kg at 24 h and 5 mg/kg at 48 h) *or* Amodiaquine (10–12 mg of base/kg qd for 3 days)
Radical treatment for *P. vivax or P. ovale* infection (prevention of relapse)	In addition to chloroquine or amodiaquine or ACT, primaquine (0.5 mg of base/kg qd in Southeast Asia and Oceania [total dose 7 mg/kg] and 0.25 mg/kg elsewhere [total dose 3.5 mg/kg]) should be given for 14 days to prevent relapse.[c] In mild G6PD deficiency, 0.75 mg of base/kg should be given once weekly for 8 weeks. Primaquine should not be given in severe G6PD deficiency. Single dose tafenoquine (adult dose 300 mg) is being introduced for radical cure in some areas alongside quantitative G6PD testing. Can be given only If G6PD levels are normal.
P. falciparum malaria	Artesunate[d,e] (4 mg/kg qd for 3 days) *plus* sulfadoxine (25 mg/kg)/pyrimethamine (1.25 mg/kg) as a single dose
	or
	Artesunate[d] (4 mg/kg qd for 3 days) *plus* amodiaquine (10 mg of base/kg qd for 3 days)[d,e]
	or Artemether-lumefantrine[d] (1.5/9 mg/kg bid for 3 days with food)
	or
	Artesunate[d] (4 mg/kg qd for 3 days) *plus* mefloquine (24–25 mg of base/kg—either 8 mg/kg qd for 3 days or 15 mg/kg on day 2 and then 10 mg/kg on day 3)[f]
	or
	DHA-piperaquine[d] (target dose: 4/24 mg/kg qd for 3 days in children weighing <25 kg and 4/18 mg/kg qd for 3 days in persons weighing ≥25 kg) *or* Artesunate-pyronaridine[d] (4/12 mg/kg qd for 3 days)
Second-line treatment/treatment of imported malaria (if other ACT not available)	Artesunate[e] (2 mg/kg qd for 7 days) *or* quinine (10 mg of salt/kg tid for 7 days) *plus 1 of the following 3:* Tetracycline[f] (4 mg/kg qid for 7 days) Doxycycline[f] (3 mg/kg qd for 7 days) Clindamycin (10 mg/kg bid for 7 days) *or* Atovaquone-proguanil (20/8 mg/kg qd for 3 days with food)
Severe Falciparum Malaria[g,h,i]	
	Artesunate[e] (2.4 mg/kg stat IV followed by 2.4 mg/kg at 12 and 24 h and then daily if necessary; for children weighing <20 kg, give 3 mg/kg per dose)
	or, if unavailable,
	Artemether[e] (3.2 mg/kg stat IM followed by 1.6 mg/kg qd)
	or, if unavailable,
	Quinine dihydrochloride (20 mg of salt/kg[j] infused over 4 h, followed by 10 mg of salt/kg infused over 2–8 h q8h[k])

[a] In endemic areas where malaria transmission is low, except in pregnant women and infants, a single dose of primaquine (0.25 mg of base/kg) should be added as a gametocytocide to all falciparum malaria treatments to prevent transmission. This is safe, even in G6PD deficiency. [b] Very few areas now have chloroquine-sensitive *P. falciparum* malaria. [c] Recent large studies indicate that these total doses can be condensed into 7-day primaquine regimens. [d] In areas where the partner drug to artesunate is known to be effective. Fixed-dose co-formulated combinations are available. The World Health Organization recommends artemisinin combination regimens as first-line therapy for falciparum malaria in all tropical countries and advocates use of fixed-dose combinations. [e] Artemisinin derivatives are not readily available in some temperate countries. [f] Tetracycline and doxycycline should not be given to pregnant women or to children <8 years of age. [g] Oral treatment should be substituted as soon as the patient recovers sufficiently to take fluids by mouth. [h] Artesunate is the drug of choice when available. The data from large studies in Southeast Asia showed a 35% lower mortality rate than with quinine, and very large studies in Africa showed a 22.5% reduction in mortality rate compared with quinine. The doses of artesunate in children weighing <20 kg should be 3 mg/kg. [i] Patients with severe malaria acquired in an area where artemisinin resistance has been confirmed should be given parenteral artesunate and quinine together at standard treatment doses. [j] A loading dose should not be given if therapeutic doses of quinine have definitely been administered in the previous 24 h. [k] Infusions can be given in 0.9% saline and 5–10% dextrose in water. Infusion rates for quinine should be carefully controlled.

Abbreviations: ACT, artemisinin combination therapy; DHA, dihydroartemisinin; G6PD, glucose-6-phosphate dehydrogenase.

severe malaria everywhere. Artesunate is given by IV injection but is also absorbed rapidly following IM injection. Artemether and the closely related drug artemotil (arteether) are oil-based formulations given by IM injection; they are erratically absorbed and do not confer the same survival benefit as artesunate. A rectal formulation of artesunate has been developed as a community-based prereferral treatment for patients in the rural tropics who cannot take oral medications. Prereferral administration of rectal artesunate has been shown to decrease mortality rates among severely ill children without access to immediate parenteral treatment. IV artesunate is approved by the U.S. Food and Drug Administration for the treatment of severe malaria and can be obtained from major drug distributors in the United States. Although parenteral quinine is steadily being replaced by parenteral artesunate in endemic areas, it may still have a small role in the treatment of presumed artemisinin-resistant severe falciparum malaria, where a precautionary approach is adopted and both artesunate and quinine are given together in full doses.

Severe falciparum malaria constitutes a medical emergency requiring intensive nursing care and careful management. Frequent evaluation of the patient's condition is essential. Adjunctive treatments such as high-dose glucocorticoids, urea, heparin, dextran, desferrioxamine, antibody to tumor necrosis factor α, high-dose phenobarbital (20 mg/kg), mannitol, or large-volume fluid or albumin boluses have proved either ineffective or harmful in clinical trials and should not be used. In acute renal failure or severe metabolic acidosis, hemofiltration or hemodialysis should be started as early as possible.

In severe malaria, parenteral antimalarial treatment should be started immediately, and in endemic areas, because of the difficulty in distinguishing sepsis from severe malaria, broad-spectrum antibiotics should also be started in children. Artesunate, given

TABLE 231-7 Properties of Antimalarial Drugs

DRUG(S)[a]	PHARMACOKINETIC PROPERTIES	ANTIMALARIAL ACTIVITY	MINOR TOXICITY	MAJOR TOXICITY
Quinine	Good oral and IM absorption (quinine); *Cl* and V_d reduced, but plasma protein binding (principally to α1 acid glycoprotein) increased (90%) in malaria; quinine $t_{1/2}$: 16 h in malaria, 11 h in healthy persons	Acts mainly on trophozoite blood stage; kills gametocytes of *P. vivax, P. ovale,* and *P. malariae* (but not *P. falciparum*); no action on liver stages	*Common:* cinchonism (tinnitus, high-tone hearing loss, nausea, vomiting, dysphoria, postural hypotension); ECG QT interval prolongation (usually by <10%). *Rare:* diarrhea, visual disturbance, rashes. *Note:* very bitter taste	*Common:* hypoglycemia. *Rare:* hypotension, blindness, deafness, cardiac arrhythmias, thrombocytopenia, hemolysis, hemolytic-uremic syndrome, vasculitis, cholestatic hepatitis, neuromuscular paralysis.
Chloroquine	Good oral absorption, very rapid IM and SC absorption; complex pharmacokinetics; enormous *Cl* and V_d (unaffected by malaria); blood concentration profile determined by distribution processes in malaria; $t_{1/2}$: 1–2 months. Active desethyl metabolite about 25% of parent drug concentrations	As for quinine, but acts slightly earlier in asexual cycle	*Common:* nausea, dysphoria, pruritus in dark-skinned patients, postural hypotension, ECG QT prolongation. *Rare:* accommodation difficulties, keratopathy, hypoglycemia, rash. *Note:* bitter taste but usually well tolerated	*Acute:* hypotensive shock (parenteral), cardiac arrhythmias, neuropsychiatric reactions. *Chronic:* retinopathy (cumulative dose, >100 g), skeletal and cardiac myopathy
Piperaquine	Adequate oral absorption, may be enhanced by fats; similar pharmacokinetics to chloroquine; $t_{1/2}$: 21–28 days	As for chloroquine; resistance has emerged in Southeast Asia	Occasional epigastric pain, diarrhea, ECG QT prolongation	None identified
Amodiaquine	Good oral absorption; largely converted to active metabolite desethylamodiaquine; $t_{1/2}$: 4–5 days	As for chloroquine, but more active against chloroquine-resistant *P. falciparum*	Nausea (tastes better than chloroquine), dysphoria, headache, bradycardia, ECG QT prolongation	Agranulocytosis; hepatitis, mainly with prophylactic use; should not be used with efavirenz
Primaquine	Complete oral absorption; active metabolite produced mainly via CYP2D6; $t_{1/2}$: 5–7 h	Radical cure; eradicates hepatic forms of *P. vivax* and *P. ovale;* kills *P. falciparum* gametocytes; kills developing liver stages of all species	Nausea, vomiting, diarrhea, abdominal pain, hemolysis, methemoglobinemia	Serious hemolytic anemia in severe G6PD deficiency; hemoglobinuria
Mefloquine	Adequate oral absorption; no parenteral preparation; $t_{1/2}$: 14–20 days (shorter in malaria)	As for quinine; resistance has emerged in Southeast Asia	Nausea, giddiness, dysphoria, fuzzy thinking, sleeplessness, nightmares, sense of dissociation	Neuropsychiatric reactions, convulsions, encephalopathy
Lumefantrine	Highly variable absorption related to fat intake; $t_{1/2}$: 3–4 days	As for quinine	None identified	None identified
Artemisinin and derivatives (artemether, artesunate)	Good oral absorption; good absorption of IM artesunate but slow and variable absorption of IM artemether; artesunate and artemether biotransformed to active metabolite dihydroartemisinin; all drugs eliminated very rapidly; $t_{1/2}$: <1 h	Broader stage specificity and more rapid than other drugs; no action on liver stages; kills all but fully mature gametocytes of *P. falciparum*	Reduction in reticulocyte count (but not anemia); neutropenia at high doses; in some cases, delayed anemia after treatment of severe malaria with hyperparasitemia	Anaphylaxis, urticaria, fever
Pyrimethamine	Good oral absorption, variable IM absorption; $t_{1/2}$: 4 days	For blood stages, acts mainly on mature forms; causal prophylactic	Well tolerated	Megaloblastic anemia, pancytopenia, pulmonary infiltration
Proguanil[b] (chloroguanide)	Good oral absorption; biotransformed to active metabolite cycloguanil; $t_{1/2}$: 16 h; biotransformation reduced by oral contraceptive use and in pregnancy	Causal prophylactic; not used alone for treatment	Well tolerated; mouth ulcers and rare alopecia	Megaloblastic anemia in renal failure
Atovaquone[b]	Highly variable absorption related to fat intake; $t_{1/2}$: 30–70 h	Acts mainly on trophozoite blood stage	None identified	None identified
Tetracycline, doxycycline[c]	Excellent absorption; $t_{1/2}$: 8 h for tetracycline, 18 h for doxycycline	Weak antimalarial activity; should not be used alone for treatment	Gastrointestinal intolerance, deposition in growing bones and teeth (tetracycline), photosensitivity, moniliasis, benign intracranial hypertension	Renal failure in patients with impaired renal function (tetracycline)
Pyronaridine	Rapid variable absorption, large V_d; $t_{1/2}$: 12–14 days	Acts mainly on trophozoite blood stage; kills gametocytes of *P. vivax, P. ovale,* and *P. malariae* (but not *P. falciparum*); no action on liver stages	Gastrointestinal intolerance, anemia, transient elevation of aminotransferases, hypoglycemia, headache	None identified
Arterolane	$t_{1/2}$: 3 h	Broad stage specificity; no action on liver stages; kills all but fully mature gametocytes of *P. falciparum*	Gastrointestinal intolerance, transient elevation of aminotransferases	None identified

[a] Several antimalarial drugs are formulated as different salts (e.g., phosphate, sulfate, hydrochloride) and are therefore prescribed as base equivalents. For example, chloroquine phosphate 250 salt contains 155 mg base equivalent. It is very important to check when prescribing that the correct dose is being given. [b] Atovaquone and proguanil are prescribed as a fixed-dose combination. This and proguanil alone should not be given if the estimated glomerular filtration rate is <30 mL/min. [c] Tetracycline and doxycycline should not be given to pregnant women or to children <8 years of age.

Abbreviations: Cl, systemic clearance; ECG, electrocardiogram; G6PD, glucose-6-phosphate dehydrogenase; V_d, total apparent volume of distribution.

by either IV or IM injection, is simple to administer, very safe, and rapidly effective. It does not require dose adjustments in liver dysfunction or renal failure. It should be used in pregnant women with severe malaria. If artesunate is unavailable and artemether or quinine is used, an initial loading dose must be given so that therapeutic concentrations are reached as soon as possible. Quinine causes dangerous hypotension if injected rapidly and so must be administered carefully by rate-controlled infusion only. If this approach is not possible, quinine may be given by deep IM injections into the anterior thigh. The optimal therapeutic range for quinine in severe malaria is not known with certainty, but total plasma concentrations of 8–15 mg/L for quinine are effective and do not cause serious toxicity. If the patient remains seriously ill or in acute renal failure for >2 days, maintenance doses of quinine should be reduced by 30–50% to prevent toxic accumulation of the drug. The initial dose should never be reduced. Convulsions should be treated promptly with IV (or rectal) benzodiazepines. The role of prophylactic anticonvulsants in children is uncertain. If respiratory support is not available, a full loading dose of phenobarbital (20 mg/kg) to prevent convulsions should *not* be given as it may cause respiratory arrest. Levetiracetam is the preferred anticonvulsant to control seizures.

When the patient is unconscious, the blood glucose level should be measured every 6 h for at least 24 h. All patients should receive a continuous infusion of dextrose, and blood concentrations ideally should be maintained above 4 mmol/L. Hypoglycemia (<2.2 mmol/L or 40 mg/dL) usually occurs within the first 24 h and should be treated immediately with bolus glucose. The parasite count and hematocrit should be measured every 6–12 h. Anemia develops rapidly. There is uncertainty as to the optimal thresholds for transfusion as there is some evidence that moderate anemia may be beneficial in a patient with severe malaria and vital organ dysfunction. It is recommended that if the hematocrit falls to <20%, whole blood (preferably fresh) or packed cells should be transfused slowly, with careful attention to circulatory status. In areas with higher malaria transmission, where blood for transfusion is in short supply, a threshold of 15% is widely used. Renal function should be checked at least daily. Children presenting with very severe anemia (hemoglobin <4 g/dL) and acidotic breathing require immediate blood transfusion. Management of fluid balance is difficult in severe malaria, particularly in adults, because of the thin dividing line between overhydration (leading to pulmonary edema) and underhydration (contributing to renal impairment). Fluid balance management is *different* from that in sepsis: fluid boluses are potentially dangerous in severe malaria. Nasogastric feeding should be delayed in nonintubated patients (for 60 h in adults and 36 h in children) to reduce the risk of aspiration pneumonia. As soon as the patient can take fluids, oral therapy should be substituted for parenteral treatment and a full 3-day course of ACT given. Mefloquine should be avoided as follow-on treatment for severe malaria because of the increased risk of post-malaria neurologic syndrome.

In areas of high transmission of both *P. falciparum* and *P. vivax* (the island of New Guinea), severe and potentially life-threatening anemia is common among children, and both species contribute. Elsewhere, severe vivax malaria may occur but is uncommon. Many patients have had comorbidities contributing to vital-organ dysfunction. Noncardiogenic pulmonary edema may occur.

P. knowlesi can cause severe disease associated with high parasite densities. Acute kidney injury, respiratory distress, and shock have all been described, but cerebral malaria does not occur. Treatment for severe vivax and knowlesi malaria should follow the recommendations given for falciparum malaria.

UNCOMPLICATED MALARIA

P. falciparum and *P. knowlesi* infections should be treated with an artemisinin-based combination because of their propensity for high parasite densities and severe disease. Infections with sensitive strains of *P. vivax*, *P. malariae*, and *P. ovale* should be treated either with an ACT or oral chloroquine (total dose, 25 mg of base/kg). The ACT regimens now recommended are safe and effective in adults, children, and pregnant women (all trimesters). The WHO recommends that artemether-lumefantrine should be given preferentially in the first trimester of pregnancy, if available, since there is the most experience with this drug. The rapidly eliminated artemisinin component is an artemisinin derivative (artesunate, artemether, or dihydroartemisinin) given for 3 days, and the partner drug is usually a more slowly eliminated antimalarial to which *P. falciparum* in the area is sensitive. Six ACT regimens are currently recommended by the WHO: artemether-lumefantrine, artesunate-mefloquine, dihydroartemisinin-piperaquine, artesunate-sulfadoxine-pyrimethamine, artesunate-amodiaquine, and artesunate-pyronaridine. In areas of low malaria transmission or to contain resistance spread, a single dose of primaquine (0.25 mg/kg) should be added to ACT as a *P. falciparum* gametocytocide to reduce the transmissibility of the infection. This low dose of primaquine is safe even in G6PD deficiency. Pregnant women should not be given primaquine. Atovaquone-proguanil is highly effective everywhere, although it is seldom used in endemic areas because of its high cost and the propensity for rapid emergence of resistance. Clinical recovery is slower after atovaquone-proguanil treatment than after ACT. Of great concern is the spread of artemisinin-resistant *P. falciparum* in Southeast Asia and East Africa. Infections with these resistant parasites are cleared slowly from the blood, with parasite clearance half-lives >5 h and clearance times typically exceeding 3 days. Cure rates with ACT have fallen to unacceptably low levels in some areas. Triple antimalarial combinations are under evaluation with promising results to date.

The 3-day ACT regimens are all well tolerated, although mefloquine is associated with increased rates of vomiting and dizziness. As second-line treatment for recrudescence following first-line therapy, a different ACT regimen may be given.

Patients should be monitored for vomiting for 1 h after the administration of any oral antimalarial drug. If there is vomiting, the dose should be repeated. Symptom-based treatment, with acetaminophen (paracetamol) administration, lowers fever and thereby reduces the patient's propensity to vomit these drugs. Minor central nervous system (CNS) reactions (nausea, dizziness, sleep disturbances) are common. The incidence of serious adverse neuropsychiatric reactions to mefloquine treatment is ~1 in 1000 in Asia but may be as high as 1 in 200 among African and white ethnic groups. Amodiaquine may also cause adverse CNS effects. All the antimalarial quinolines (chloroquine, piperaquine amodiaquine, mefloquine, and quinine) exacerbate the orthostatic hypotension associated with malaria, and all are tolerated better by children than by adults. Chloroquine, amodiaquine, and piperaquine all prolong ventricular repolarization (QT prolongation) but, at currently recommended doses, are not proarrhythmic. Pregnant women, young children, patients unable to tolerate oral therapy, and nonimmune individuals (e.g., travelers) with suspected malaria should be evaluated carefully and hospitalization considered. If there is any doubt as to the identity of the infecting malarial species, treatment for falciparum malaria should be given. A negative blood smear read by an experienced microscopist makes malaria very unlikely but does not rule it out completely; thick blood films should be checked again 1 and 2 days later to exclude the diagnosis. Nonimmune patients receiving treatment for malaria should have daily parasite counts performed to confirm reduction until the thick films are negative for asexual parasite stages. If the asexual parasitemia has not cleared by 7 days (and adherence is assured), second-line treatment should be administered.

To eradicate persistent liver stages and prevent relapse (radical treatment), primaquine (0.5 mg of base/kg in East Asia and Oceania and 0.25 mg/kg elsewhere) should be given once daily for 14 days to patients with *P. vivax* or *P. ovale* infection after laboratory tests for G6PD deficiency have proved negative. The same total dose may be given over 7 days. If the patient has a mild variant of G6PD deficiency, primaquine can be given in a dose of 0.75 mg of base/kg (maximum, 45 mg) once weekly for 8 weeks. Pregnant women with vivax or ovale malaria should not be given primaquine but should receive suppressive prophylaxis with chloroquine (5 mg of base/kg per week) until 1 month after delivery, after which radical treatment can be given. The slowly eliminated 8-aminoquinoline tafenoquine has been registered in some countries (currently recommended adult dose 300 mg, which may be too low; dose-optimization studies are ongoing). This allows radical cure to be given in a single dose. The consequent risk of protracted hemolysis in G6PD deficiency, including in female heterozygotes who may test as normal with current G6PD screens (which detect <30–40% of normal enzyme activity), requires that all patients should have a quantitative test of G6PD activity before receiving tafenoquine. Only those with >70% of normal activity should receive the drug.

MANAGEMENT OF COMPLICATIONS OF MALARIA

Acute Renal Failure If plasma levels of BUN or age- and weight-adjusted creatinine rise despite adequate rehydration, fluid administration should be restricted to prevent volume overload. As in other forms of hypercatabolic acute kidney injury, renal replacement therapy is best performed early **(Chap. 321)**. Hemofiltration and hemodialysis are more effective than peritoneal dialysis and are associated with lower mortality risk. Some patients with renal impairment pass small volumes of urine sufficient to allow control of fluid balance; these cases can be managed conservatively if other indications for dialysis do not arise. Renal function usually improves within days, but full recovery may take weeks.

Acute Pulmonary Edema (Acute Respiratory Distress Syndrome) This syndrome is caused by increased pulmonary capillary permeability. Patients should be positioned with the head of the bed at a 45° elevation and should be given oxygen and IV diuretics. Positive-pressure ventilation should be started early if the immediate measures fail **(Chap. 316)**. Rarely, patients may require extracorporeal membrane oxygenation.

Hypoglycemia An initial slow injection of 20% dextrose (2 mL/kg over 10 min) should be followed by an infusion of 10% dextrose (0.10 g/kg per hour). The blood glucose level should be checked regularly thereafter as recurrent hypoglycemia is common, particularly among patients receiving quinine. In severely ill patients, hypoglycemia commonly occurs together with metabolic (lactic) acidosis and carries a poor prognosis.

Sepsis In malaria-endemic areas where a high proportion of children are parasitemic, it is usually impossible to distinguish severe malaria from bacterial sepsis with confidence. These children should be treated with both antimalarials and broad-spectrum antibiotics with activity against nontyphoidal *Salmonella* species from the outset. Ideally all patients with severe malaria should have blood cultures taken, and empirical antibiotics should also be given to adults with >20% parasitemia. Antibiotics should be considered for severely ill patients of any age who are not responding to antimalarial treatment or who deteriorate unexpectedly.

Other Complications Patients who develop spontaneous bleeding should be given fresh blood and IV vitamin K. Convulsions should be treated with IV or rectal benzodiazepines and, if necessary, respiratory support. Aspiration pneumonia should be suspected in any unconscious patient with convulsions, particularly with persistent hyperventilation; IV antimicrobial agents and oxygen should be administered, and pulmonary hygiene should be undertaken.

GLOBAL CONSIDERATIONS

The goal of global eradication of malaria remains a challenge. After substantial progress from 2000 to 2015, the global burden of malaria has been steadily increasing. Malaria eradication will require strong leadership, increased national commitment, and substantial international support. The two main tools used to control malaria are insecticide-treated bed nets (ITNs), previously shown to reduce all-cause mortality in African children by 20%, and the ACTs. Pyrethroid resistance and artemisinin resistance in Africa pose major threats to global malaria control. Newer ITNs combine pyrethroids with chlorfenapyr (a pyrrole insecticide) or pyriproxyfen (an insect growth regulator). New drugs to treat malaria are in development but still years away. Two very similar pre-erythrocytic vaccines (RTS,S/AS01 and RTS/matrix M) have been recommended by WHO for deployment. These vaccines provide relatively short-duration protection and are given to young children. Other challenges to malaria eradication include the widespread distribution of *Anopheles* breeding sites, the enormous number of infected persons, behavioral changes (to avoid ITN contact) in anopheline mosquito vectors, and inadequacies in human and material resources, infrastructure, and control programs. Eliminating vivax malaria is further hindered by the lack of a simple, safe, radical curative regimen. The animal reservoir of *P. knowlesi* presents another challenge to eradication of this species.

MALARIA PREVENTION

Malaria may be contained by judicious use of insecticides to kill the mosquito vector, rapid diagnosis, patient management, and—where effective and feasible—administration of intermittent preventive treatments, seasonal malaria chemoprevention, or chemoprophylaxis to high-risk groups such as pregnant women and young children. Indoor residual spraying with insecticides is practiced as part of an integrated vector control program in endemic countries, and a combination neonicotinoid-pyrethroid has been developed in an attempt to delay the emergence of insecticide resistance. Focal elimination of *P. falciparum* can be accelerated safely by mass treatment with slowly eliminated antimalarials such as dihydroartemisinin-piperaquine. In addition to the two malaria vaccines now available, an irradiated live sporozoite vaccine is in late-stage development, and research on many other vaccine candidates is ongoing.

■ PERSONAL PROTECTION AGAINST MALARIA

Simple measures to reduce the frequency of bites by infected mosquitoes in malarious areas are very important. These measures include the avoidance of exposure to mosquitoes at their peak feeding times (usually dusk to dawn) and the use of insect repellents containing 10–35% DEET (or, if DEET is unacceptable, 7% picaridin), suitable clothing, and ITNs or other insecticide-impregnated materials. Widespread use of bed nets treated with residual pyrethroids reduces the incidence of malaria in areas where vectors bite indoors at night.

■ CHEMOPROPHYLAXIS

(**Table 231-8**; *https://wwwnc.cdc.gov/travel/yellowbook/2020/travel-related-infectious-diseases/malaria*) Recommendations for malaria prophylaxis depend on knowledge of local patterns of drug sensitivity in *Plasmodium* species and the likelihood of acquiring malarial infection. Drugs effective against resistant *P. falciparum* should be used (atovaquone-proguanil, doxycycline, or mefloquine). Chemoprophylaxis is never entirely reliable, and malaria should always be considered in the differential diagnosis of fever in patients who have traveled to endemic areas, even if they are taking prophylactic antimalarial drugs.

Chemoprevention in Pregnancy Pregnant women planning to visit malarious areas should be warned about the potential risks and advised to avoid all nonessential travel. All pregnant women who live in endemic areas should be encouraged to attend regular antenatal clinics. Mefloquine is the only drug advised for pregnant women traveling to areas with drug-resistant malaria; this drug is generally considered safe in the second and third trimesters of pregnancy; the data on first-trimester exposure, although limited, are reassuring. Chloroquine and proguanil are regarded as safe, but there are now very few regions where these drugs can be recommended for protection. The safety of other prophylactic antimalarial agents in pregnancy has not been established. In endemic areas, intermittent preventive treatment in pregnancy (IPTp) involves giving treatment doses of sulfadoxine-pyrimethamine (SP) at each antenatal visit (maximum, once monthly) in the second and third trimesters of pregnancy. Women with HIV infection who are taking trimethoprim-sulfamethoxazole as prophylaxis should not

TABLE 231-8 Drugs Used in the Prophylaxis of Malaria[a]

DRUG	USAGE	ADULT DOSE	PEDIATRIC DOSE	COMMENTS
Atovaquone-proguanil (Malarone)	Prophylaxis in areas with chloroquine- or mefloquine-resistant *Plasmodium falciparum*	1 adult tablet PO[b]	5–8 kg: ½ pediatric tablet[c] daily ≥8–10 kg: ¾ pediatric tablet daily ≥10–20 kg: 1 pediatric tablet daily ≥20–30 kg: 2 pediatric tablets daily ≥30–40 kg: 3 pediatric tablets daily ≥40 kg: 1 adult tablet daily	Begin 1–2 days before travel to malarious areas. Take daily at the same time each day while in the malarious areas and for 7 days after leaving such areas. Atovaquone-proguanil is contraindicated in persons with severe renal impairment (creatinine clearance rate, <30 mL/min). In the absence of data, it is not recommended for children weighing <5 kg, pregnant women, or women breast-feeding infants weighing <5 kg. Atovaquone-proguanil should be taken with food or a milky drink.
Chloroquine phosphate (Aralen and generic)	Prophylaxis only in the very few areas with chloroquine-sensitive *P. falciparum*[c] or areas with *P. vivax* only	300 mg of base (500 mg of salt) PO once weekly	5 mg of base/kg (8.3 mg of salt/kg) PO once weekly, up to maximum adult dose of 300 mg of base	Begin 1–2 weeks before travel to malarious areas. Take weekly on the same day of the week while in the malarious areas and for 4 weeks after leaving such areas. Chloroquine may exacerbate psoriasis.
Doxycycline (many brand names and generic)	Prophylaxis in areas with chloroquine- or mefloquine-resistant *P. falciparum*[d]	100 mg PO qd (except in pregnant women; see Comments)	≥8 years of age: 2 mg/kg PO qd, up to adult dose	Begin 1–2 days before travel to malarious areas. Take daily at the same time each day while in the malarious areas and for 4 weeks after leaving such areas. Doxycycline is contraindicated in children aged <8 years and in pregnant women.
Hydroxychloroquine sulfate (Plaquenil)	An alternative to chloroquine for primary prophylaxis only in the very few areas with chloroquine-sensitive *P. falciparum*[d] or areas with *P. vivax* only	310 mg of base (400 mg of salt) PO once weekly	5 mg of base/kg (6.5 mg of salt/kg) PO once weekly, up to maximum adult dose of 310 mg of base	Begin 1–2 weeks before travel to malarious areas. Take weekly on the same day of the week while in the malarious areas and for 4 weeks after leaving such areas. Hydroxychloroquine may exacerbate psoriasis.
Mefloquine (Lariam and generic)	Prophylaxis in areas with chloroquine-resistant *P. falciparum*[d]	228 mg of base (250 mg of salt) PO once weekly	≤9 kg: 4.6 mg of base/kg (5 mg of salt/kg) PO once weekly >9–19 kg: ¼ tablet[e] once weekly >19–30 kg: ½ tablet once weekly >30–45 kg: ¾ tablet once weekly >45 kg: 1 tablet once weekly	Begin 1–2 weeks before travel to malarious areas. Take weekly on the same day of the week while in the malarious areas and for 4 weeks after leaving such areas. Mefloquine is contraindicated in persons allergic to this drug or related compounds (e.g., quinine and quinidine) and in persons with active or recent depression, generalized anxiety disorder, psychosis, schizophrenia, other major psychiatric disorders, or seizures. Use with caution in persons with psychiatric disturbances or a history of depression. Mefloquine is not recommended for persons with cardiac conduction abnormalities.
Primaquine	For prevention of malaria in areas with mainly *P. vivax*	30 mg of base (52.6 mg of salt) PO qd	0.5 mg of base/kg (0.8 mg of salt/kg) PO qd, up to adult dose; should be taken with food	Begin 1–2 days before travel to malarious areas. Take daily at the same time each day while in the malarious areas and for 7 days after leaving such areas. Primaquine is contraindicated in persons with G6PD deficiency. It is also contraindicated during pregnancy.
Primaquine	Terminal prophylaxis to decrease risk of relapses of *P. vivax* and *P. ovale*	30 mg of base (52.6 mg of salt) PO qd for 14 days after departure from the malarious area	0.5 mg of base/kg (0.8 mg of salt/kg), up to adult dose, PO qd for 14 days after departure from the malarious area	This therapy is indicated for persons who have had prolonged exposure to *P. vivax* and/or *P. ovale*. It is contraindicated in persons with G6PD deficiency as well as during pregnancy.

[a] Several antimalarial drugs are formulated as different salts (e.g., phosphate, sulfate, hydrochloride) and are therefore prescribed as base equivalents. For example, chloroquine phosphate 250 salt contains 155 mg base equivalent. It is very important to check when prescribing that the correct dose is being given. [b] An adult tablet contains 250 mg of atovaquone and 100 mg of proguanil hydrochloride. [c] A pediatric tablet contains 62.5 mg of atovaquone and 25 mg of proguanil hydrochloride. [d] Very few areas now have chloroquine-sensitive falciparum malaria. [e] One tablet contains 228 mg of base (250 mg of salt).

Abbreviation: G6PD, glucose-6-phosphate dehydrogenase.

Source: Centers for Disease Control and Prevention, *https://www.cdc.gov/malaria/travelers/index.html*.

be given concomitant SP. Dihydroartemisinin-piperaquine is being evaluated as an alternative. Children born to nonimmune mothers in malaria-endemic areas (usually expatriates moving to these areas) should receive prophylaxis from birth.

Chemoprevention for Children in Endemic Areas Seasonal malaria chemoprevention (SMC) involves giving monthly treatment doses of amodiaquine and SP to children (between 3 months and 5 years of age) in endemic areas of Africa. Originally designed for the intense 3- to 4-month rainy season malaria transmission period across the Sahel region of Africa, it is now introduced in other parts of Africa where malaria is less seasonal and more drug resistant, although its effectiveness there is not known. Hundreds of millions of doses are now dispensed annually. Perennial malaria chemoprevention (PMC) is a loosely described intervention involving giving treatment doses, usually of SP, to young children when they attend health services for immunizations or whenever convenient for control programs. It has replaced intermittent preventive treatment for infants. Children should not receive both SMC and PMC.

Chemoprevention in Travelers Travelers to a malaria-endemic region should start taking antimalarial drugs 2 days to 2 weeks before departure so that any untoward reactions can be detected before travel and therapeutic antimalarial blood concentrations will be present if and when any infections develop (Table 231-8). Antimalarial prophylaxis should continue for 4 weeks after the traveler has left the endemic area, except if atovaquone-proguanil or primaquine has been taken; these drugs have significant activities against the liver stage of the infection (causal prophylaxis) and can be discontinued 1 week after departure from the endemic area. If suspected malaria develops while a traveler is abroad, obtaining a reliable diagnosis and antimalarial treatment locally is a top priority. Presumptive self-treatment for malaria with atovaquone-proguanil (for 3 consecutive days) or one of the artemisinin-based combinations can be considered under special circumstances; medical advice on self-treatment should be sought before departure for malaria-endemic areas and as soon as possible after illness begins. Every effort should be made to confirm the diagnosis.

Atovaquone-proguanil (3.75/1.5 mg/kg or 250/100 mg, daily adult dose) is a fixed-combination, once-daily prophylactic agent that is very well tolerated by adults and children. This combination is effective against all types of malaria, including multidrug-resistant falciparum malaria. Atovaquone-proguanil is best taken with food or a milky drink to optimize absorption. It is not recommended if the estimated glomerular filtration rate is <30 mL/min. There are insufficient data on safety for prophylaxis in pregnancy.

Mefloquine (250 mg of salt weekly, adult dose) has been widely used for malarial prophylaxis because it is usually effective against multidrug-resistant falciparum malaria and is reasonably well tolerated. Mefloquine has been associated with rare episodes of psychosis and seizures at prophylactic doses; these reactions are more frequent at the higher doses used for treatment. More common side effects with prophylactic mefloquine include mild nausea, dizziness, fuzzy thinking, disturbed sleep patterns, vivid dreams, dysphoria, and malaise. Mefloquine is contraindicated for use by travelers with known hypersensitivity and by persons with active or recent depression, anxiety disorder, psychosis, schizophrenia, another major psychiatric disorder, or seizures; it is not recommended for persons with cardiac conduction abnormalities, although the evidence that it is cardiotoxic is very weak. Daily administration of doxycycline (100 mg daily, adult dose) is an effective alternative to atovaquone-proguanil or mefloquine. Doxycycline is generally well tolerated but may cause vulvovaginal thrush, diarrhea, and photosensitivity and is not recommended for prophylaxis in children <8 years old or pregnant women, although evidence that it is harmful in children is lacking.

Chloroquine can no longer be relied upon to prevent *P. falciparum* infections but is still used to prevent and treat malaria due to the other human *Plasmodium* species and for *P. falciparum* malaria in Central American countries west and north of the Panama Canal and in Caribbean countries. Chloroquine-resistant *P. vivax* has been reported from parts of eastern Asia, Oceania, and Central and South America. High-level resistance in *P. vivax* is prevalent in Oceania and Indonesia. Chloroquine is generally well tolerated, although some patients cannot take it because of malaise, headache, visual symptoms (due to reversible keratopathy), gastrointestinal intolerance, alopecia, or pruritus. Chloroquine is considered safe in pregnancy. With chronic administration for >5 years, a characteristic dose-related retinopathy may develop, but this condition is rare at the doses used for antimalarial prophylaxis. Idiosyncratic or allergic reactions are also rare. Skeletal and/or cardiac myopathy is a potential problem with protracted prophylactic use, although it is more likely to occur at the high doses used in the treatment of rheumatoid arthritis. Neuropsychiatric reactions and skin rashes are unusual. Amodiaquine should not be used for weekly prophylaxis because continuous weekly use is associated with a high risk of agranulocytosis (~1 person in 2000) and hepatotoxicity (~1 person in 16,000). Chloroquine, amodiaquine, and piperaquine all cause moderate electrocardiograph QT prolongation but have not been associated with ventricular arrhythmias at therapeutic doses.

Primaquine (0.5 mg of base/kg or a daily adult dose of 30 mg taken with food), an 8-aminoquinoline compound, has proved safe and effective in the prevention of drug-resistant falciparum and vivax malaria in adults. Primaquine can be considered for adults (with the exception of pregnant women) who are intolerant to other recommended drugs. Abdominal pain can be prevented by taking primaquine with food. G6PD deficiency must be excluded before primaquine is prescribed. In the past, the dihydrofolate reductase inhibitors pyrimethamine and proguanil (chloroguanide) were administered widely, but the rapid selection of resistance in both *P. falciparum* and *P. vivax* has limited their use. Whereas antimalarial quinolines such as chloroquine (a 4-aminoquinoline) act only on the erythrocyte stage of parasitic development, the dihydrofolate reductase inhibitors (as well as atovaquone and primaquine) also inhibit preerythrocytic growth in the liver (causal prophylaxis) and development in the mosquito (sporontocidal activity). Proguanil is safe and well tolerated, although mouth ulceration occurs in ~8% of persons using this drug; it is considered safe for antimalarial prophylaxis in pregnancy. Prophylactic use of the combination of SP or amodiaquine is not recommended for weekly administration because of an unacceptable incidence of severe toxicity, principally exfoliative dermatitis and other skin rashes (SP) (incidence, 1 in 7000; fatal reactions, 1 in 18,000), agranulocytosis (amodiaquine), hepatitis, and pulmonary eosinophilia (both drugs).

Because of the increasing spread and intensity of antimalarial drug resistance, the Centers for Disease Control and Prevention (CDC) recommends that travelers and their providers consider their destination, type of travel, and current medications and health risks when choosing antimalarial chemoprophylaxis. There is an increasingly appreciated problem of falsified and substandard antimalarial drugs (and other medicines) on the shelves of pharmacies in Southeast Asia and sub-Saharan Africa; hence, travelers should purchase their preventive drugs from a reputable source before going to a malarious country. Consultation for the evaluation of prophylaxis failures or treatment of malaria can be obtained from state and local health departments and the CDC Malaria Hotline (770) 488-7788 or (855) 856-4713 [toll free]) or the CDC Emergency Operations Center (770-488-7100 [after hours]).

ACKNOWLEDGMENT

The authors gratefully acknowledge the substantial contributions of Joel G. Breman, MD, to this chapter in the previous editions.

FURTHER READING

World Health Organization: WHO guidelines for malaria. October 16, 2023. Available at: *https://www.who.int/teams/global-malaria-programme/guidelines-for-malaria*. Accessed March 1, 2024.

232 Babesiosis

Edouard Vannier, Jeffrey A. Gelfand

Babesiosis is an emerging infectious disease caused by protozoan parasites of the genus *Babesia* that invade and eventually lyse red blood cells (RBCs). Most cases occur in the United States during the summer months and are caused by *Babesia microti*, a species typically found in small rodents and transmitted by the deer tick, *Ixodes scapularis*. Symptoms are those of a flu-like illness. For most patients, a single standard course of atovaquone plus azithromycin is sufficient to achieve cure. Highly immunocompromised patients and asplenic individuals are at risk of persistent infection and should be treated for a longer duration. Adjunct red cell exchange can be useful for severe cases. In the absence of vaccine and chemoprophylaxis, persons at risk of severe babesiosis should minimize their exposure to ticks and, if possible, avoid endemic areas.

ETIOLOGY, EPIDEMIOLOGY, AND MODES OF TRANSMISSION

A few *Babesia* species have been implicated as etiologic agents of human babesiosis. These species use wild or domesticated mammals as reservoir hosts and are maintained in their enzootic cycle by ticks. Humans are incidental, dead-end hosts. Most cases are reported from across the Northern Hemisphere; the predominant etiologic agent varies by continent.

United States • **GEOGRAPHIC DISTRIBUTION** Most cases (>95%) are caused by *B. microti* and are reported from the Northeast (Massachusetts, Rhode Island, Connecticut, Vermont, New Hampshire, Maine, New York, New Jersey) and the upper Midwest (Minnesota, Wisconsin) **(Fig. 232-1)**. Other *Babesia* species seldom cause disease. Symptomatic infection with *Babesia duncani* and *B. duncani*–type organisms has been reported from Washington State, Oregon, and California. Symptomatic infection with *Babesia divergens*–like organisms has been documented in the central states of Arkansas, Missouri, Michigan, and Kentucky and also in Washington State and Pennsylvania.

INCIDENCE Cases are reported weekly by state health departments to the Centers for Disease Control and Prevention (CDC) via the National Notifiable Diseases Surveillance System. In 2023, more than 3200 cases were reported from 33 of the 40 states in which babesiosis is notifiable. In 2011, when babesiosis became a nationally notifiable disease, 1126 cases were reported to the CDC. The increase in incidence over time is best explained by a greater density of ticks in highly endemic areas and by the northward expansion of these ticks. Although *B. microti* and *Borrelia burgdorferi* (the agent of Lyme disease; **Chap. 191**) are transmitted by the same tick species, the deer tick *I. scapularis*, the geographic expansion of babesiosis has lagged behind that of Lyme disease. This delay likely reflects the poor ecologic fitness of *B. microti* compared with that of *B. burgdorferi* and is consistent with the observation that *B. burgdorferi* helps maintain *B. microti* in their shared enzootic cycle. Other factors contributing to the rise in the incidence of babesiosis include greater exposure to ticks due to forest fragmentation in suburban areas and increased leisure activities in grassy or wooded areas.

MODES OF TRANSMISSION • **Tick Bite** *B. microti* is acquired primarily during the blood meal of an *I. scapularis* tick. Less than one-half (~45%) of patients recall a tick bite within the 8 weeks prior to symptom onset. Both nymphs and adult ticks can transmit *B. microti*; tick larvae are not infected because *B. microti* is not transmitted transovarially. Most cases (~90%) present from June through August because nymphs—the primary vector—are active from late spring to early summer. Patients who present in the fall may have acquired *B. microti* from an adult female tick. *B. duncani* and *B. divergens*–like organisms are thought to be transmitted by *Dermacentor albipictus* and *Ixodes dentatus* ticks, respectively.

Blood Transfusion More than 300 cases of transfusion-transmitted babesiosis (TTB) caused by *B. microti* have been reported. Most cases involve packed RBCs; a few have been attributed to frozen-deglycerolized RBCs and whole blood–derived platelets contaminated with RBCs but none to apheresis platelets. At the time of transfusion, the age of the refrigerated RBC units has ranged from 4 to 42 days, indicating that *B. microti* remains viable throughout the RBC unit's shelf life. Acellular blood components (plasma, cryoprecipitate) have not been implicated. TTB has occurred year-round because *B. microti* can persist for more than a year in untreated asymptomatic carriers. Given the seasonality of tick-borne babesiosis, two-thirds of the donations implicated in TTB cases have been obtained from June through October. Like that of tick-borne babesiosis, the incidence of TTB sharply increased during the first two decades of the millennium. In May 2020, the American Red Cross (ARC) began to test all blood donations collected in the 15 jurisdictions that account for 99% of tick-borne babesiosis cases and 95% of TTB cases. During the following 13 months, no suspected TTB cases were reported to the ARC. In contrast, from 2010 to 2017, *B. microti*–positive donations were linked to 81 TTB cases (~10 cases per year). Screening of the blood supply relies on the amplification of parasite 18S rRNA gene transcripts. In addition to *B. microti*, the Procleix Babesia

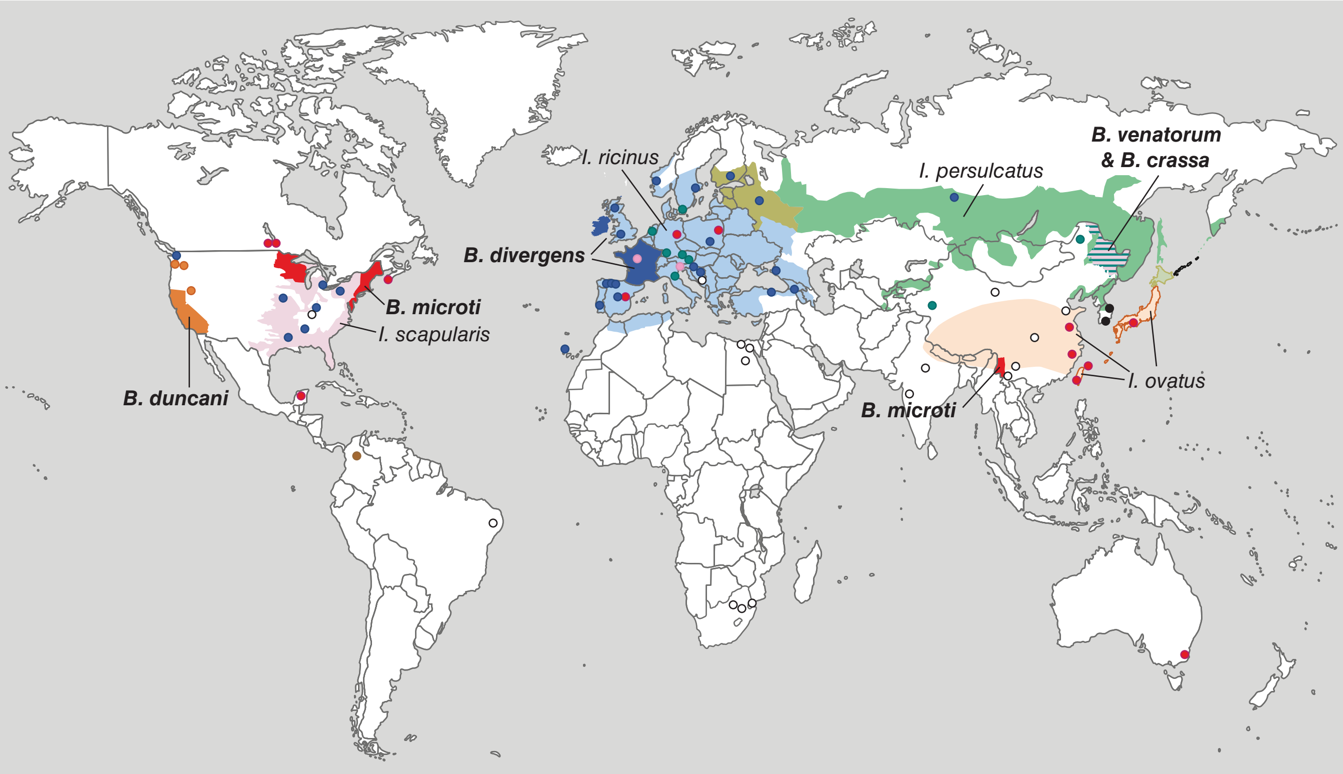

FIGURE 232-1 Geographic distribution of human babesiosis and associated tick vectors. Dark colors indicate areas where human babesiosis is endemic or sporadic (defined by ≥5 cases). Light colors indicate areas where tick vectors are present but human babesiosis is rare (<5 cases), undocumented, or absent. *Circles* depict single cases except in six locations (Colombia, Mexico, Montenegro, Poland, and the provinces of Gansu and Shandong in China) where all patients were diagnosed at one hospital or identified via survey in one location. Colors distinguish the etiologic agents: *red* for *Babesia microti*, *orange* for *B. duncani*, *blue* for *B. divergens* and *B. divergens–like*, *green* for *B. venatorum*, *pink* for *B. crassa–like*, *brown* for *B. bovis* and *B. bigemina*, and *black* for *B. motasi–like*. *White circles* depict cases caused by uncharacterized *Babesia* isolates or by isolates for which the full sequence of the 18S rRNA gene was not available or ruled out all of the *Babesia* species listed above. Asymptomatic infections and cases of travel-associated babesiosis are omitted.

Assay detects *B. duncani* and *B. divergens*, which have been implicated in a handful of TTB cases. The assay is exquisitely sensitive; the 95% limit of detection is 3 parasites per mL of blood for *B. microti* and *B. duncani*, and 2 parasites per mL of blood for *B. divergens*.

Vertical Transmission Passage of *B. microti* across the placenta has been documented but is rare. Symptoms typically develop during the third to the sixth week of life and often consist of fever accompanied by pallor and lethargy. Parasitemia at presentation has ranged from 2% to 5%. Both the neonate and the mother typically are seropositive for *B. microti* IgG antibody. Most cases of neonatal babesiosis, however, are acquired through blood transfusion or tick bite. Infants (<1 year of age) account for <1% of the annual number of cases of babesiosis reported to the CDC.

Solid Organ Transplantation This unusual mode of transmission has been highlighted in a single case report. Two patients received a diagnosis of babesiosis 8 weeks after transplantation of a kidney allograft obtained from a single donor who had received multiple transfusions shortly before his death. Corneas from the deceased donor were transplanted, but neither recipient was infected with *B. microti*; this outcome suggests that RBCs that had remained in the vasculature or fluids of the donated kidneys were the source of *B. microti*.

Risk Factors Most patients (~80%) who present with symptoms of babesiosis are ≥50 years. Those who are admitted to a hospital are a decade older (median age, 68 years) than those who are not (median age, 59 years). Aside from tick exposure and advanced age, major risk factors for severe babesiosis include asplenia and immunosuppression. Asplenia can be congenital, functional (e.g., due to celiac disease or hemoglobinopathies such as sickle cell disease and thalassemia), or acquired (due to splenectomy). Immunosuppression often is iatrogenic and associated with conditions such as autoimmune disorders, chronic inflammatory disorders, malignancies, or transplantation. Immunosuppression can be inherent in comorbidities such as X-linked agammaglobulinemia, common variable hypogammaglobulinemia, and HIV/AIDS. Risk factors for TTB include conditions that require transfusion of blood components, particularly RBCs.

Outside the United States Travel-associated babesiosis may become more common if, as anticipated, *Babesia* species continue to emerge worldwide (Fig. 232-1).

EUROPE More than 40 cases have been attributed to *B. divergens* since the index case was reported from Croatia in 1957. The infection is rarely diagnosed in immunocompetent individuals; most patients lack a spleen or have functional hyposplenism. Most cases occur in France and Ireland, particularly in regions with cattle farms. *B. divergens*, a parasite of cattle, is transmitted by the castor bean tick *Ixodes ricinus*. Other *Babesia* species rarely cause disease. *B. divergens*–like parasites have been implicated in 1 case on the Canary Islands and in another in western France. Five cases caused by *Babesia venatorum* have been reported from Italy, Austria, Germany, and Sweden. Two cases of infection with *Babesia crassa*–like organisms have been reported from Slovenia and western France. All 9 patients had been splenectomized. In Germany, a normosplenic patient with acute myeloid leukemia presumably acquired *B. microti* during the transfusion of whole blood–derived platelets. In eastern Poland, mild babesiosis caused by *B. microti* has occurred in normosplenic patients. *I. ricinus* is the vector for *B. microti* and *B. venatorum* and the presumed vector for *B. divergens*–like parasites. *Haemaphysalis concinna* is the likely vector for *B. crassa*–like organisms. *B. microti* is found in small rodents, *B. venatorum* in roe deer, and *B. crassa* in sheep. Rabbits are suspected reservoirs for *B. divergens*–like parasites.

ASIA *B. microti* was recognized as a human pathogen in Taiwan in the late 1990s. In the past decade, babesiosis has gained the status of emerging infectious disease in mainland China. In the northeastern province of Heilongjiang, *B. venatorum* and *B. crassa*–like organisms have caused mild disease in immunocompetent individuals, whereas antibodies specific for *B. microti* have been detected in blood donors. The taiga tick *Ixodes persulcatus* is a competent vector for *B. microti* and the presumed vector for *B. venatorum*. In this province, *B. crassa*–like organisms are found in *I. persulcatus* and *H. concinna* ticks. A case of *B. venatorum* infection was documented in a child residing in the northwestern Xinjiang Autonomous Region. Several cases of *B. microti* infection have been reported from southern China, mostly from Yunnan; two of the patients were co-infected with *Plasmodium* species, one with *P. falciparum* and the other with *P. vivax*.

The single case of babesiosis reported from Japan was acquired through blood transfusion and was caused by a *B. microti* organism that defines the Kobe lineage. Two cases have been reported from South Korea; both occurred in splenectomized individuals and were caused by *Babesia motasi*–like organisms, which are parasites of sheep and goats. In one case, *Haemaphysalis longicornis* was the presumed vector.

REMAINDER OF THE WORLD Three cases of babesiosis caused by *B. microti* have been reported from Canada: two from southern Manitoba and one from southern Nova Scotia. Asymptomatic *B. microti* infection has been detected in a few blood donors from southwestern Ontario. Babesiosis is a reportable disease in Manitoba and Quebec. In Yucatan, Mexico, four cases of febrile illness have been attributed to *B. microti*. Evidence that *Babesia bovis* and *Babesia bigemina* may cause human illness has come from Uraba, a region of Colombia where cattle ranching is important and malaria is endemic. In western India, the death of a young, splenectomized farmer was caused by a novel *Babesia* species. A fatal case of *B. microti* infection has been reported from New South Wales, Australia; the causative agent was likely imported.

CLINICAL MANIFESTATIONS

United States • ***B. MICROTI* INFECTION** Symptoms typically appear 1–4 weeks after the bite of an infected tick but 3–7 weeks (median, 37 days; range, 11–176 days) after transfusion of contaminated blood components. Patients experience a gradual onset of fatigue with or without malaise that is followed within days by fever and one or more of the following: chills, night sweats, headache, myalgia, and anorexia. Fever is persistent or intermittent and has reached 40.9°C (105.6°F). Less common symptoms include arthralgia, nausea, dry cough, neck stiffness, emotional lability, and sore throat. Diarrhea, vomiting, abdominal pain, and joint swelling are rare. Dark urine and jaundice raise the suspicion of severe hemolytic anemia and may be accompanied by shortness of breath.

On physical examination, fever is the salient feature. The skin may be pale or yellowish. A focal red rash, if present, denotes the site of the tick bite; an erythema migrans rash **(Fig. A1-8)** signifies concurrent Lyme disease **(Chap. 191)** or southern tick-associated rash illness (STARI) **(Chap. 190)**. Ecchymoses and petechiae are rare. Examination of the eyes may be remarkable for scleral icterus, which is consistent with severe hemolysis. Retinopathy with splinter hemorrhages and retinal infarcts are rare. Tenderness of the abdominal upper left quadrant suggests splenomegaly, which may be accompanied by hepatomegaly. Left-sided abdominal pain raises suspicion of splenic infarction or splenic rupture. Unexplained hypotension accompanied by tachycardia reinforces suspicion of splenic rupture. Splenic infarction and subcapsular hematoma can occur in the absence of splenic rupture. Splenic infarction and splenic rupture are confirmed by CT. Splenic infarcts appear as wedge-shaped hypodense lesions. A hyperdense fluid surrounding the spleen is consistent with hemoperitoneum caused by splenic rupture.

Severe babesiosis requires hospital admission. The median length of hospital stay is 4 days (interquartile range, 3–7). Severe babesiosis can be accompanied by one or several complications. The leading complication is acute renal failure (20%). The second most common complication is acute respiratory failure (7%). Less common complications include supraventricular arrhythmia, heart failure, disseminated intravascular coagulation, and shock. At least one transfusion of red blood cells is given to one of five patients hospitalized for babesiosis. Patients diagnosed >7 days after symptom onset are predisposed to severe complicated babesiosis, which has been defined as an illness requiring admission to an intensive care unit (ICU); an illness complicated by acute respiratory distress syndrome, heart failure or shock; or an illness requiring dialysis, intubation or RBC exchange. Among clinical features, diarrhea and nausea or vomiting are strong predictors of severe babesiosis as just defined. Asplenia and autoimmune disorders predispose to severe disease, but underlying cardiac conditions do not. Concurrent Lyme disease does not decrease the risk of severe disease but has been associated with a lower risk for hospitalization; the implication is that a diagnosis of babesiosis should be considered in any patient diagnosed with Lyme disease.

Despite therapy, babesiosis can be fatal. Prior to the turn of the millennium, when clindamycin plus quinine was the regimen of choice for treatment of babesiosis, fatality rates ranging from 5% to 9% were reported. Since atovaquone plus azithromycin has become the first-line therapy, a fatal outcome is rare. Of the 7612 cases of babesiosis reported to the CDC in 2011–2015, 46 (0.6%) ended in death. Death is more frequent among those ≥65 years of age. A review of claims for 10,305 Medicare recipients who were diagnosed with babesiosis between 2006 and 2013 revealed that 1% had died within 30 days. Among those admitted to a hospital, the fatality rate was 3%. Fatality is high in immunocompromised patients (~20%) and those with splenic rupture or splenic infarction (6%).

OTHER BABESIA INFECTIONS The eight documented cases of *B. duncani* infection reported in the United States were moderate to severe; one patient died. Symptoms were similar to those evoked by *B. microti*. All seven patients infected with *B. divergens*–like organisms had been splenectomized and experienced a severe illness that required hospitalization; three died.

Global Considerations Most cases of *B. divergens* infection in Europe have occurred in individuals who lack a spleen or have functional hyposplenism. Symptoms develop suddenly and consist of fever (>41°C [>105.8°F]), headache, lumbar or abdominal pain, and dark urine. Jaundice and hemoglobinuria are common. Without immediate therapy, including immediate red cell exchange, patients often develop renal failure with or without acute respiratory distress or failure. In spleen-intact individuals, *B. divergens* infection typically evokes a mild illness. All 5 patients infected with *B. venatorum* in Europe had been splenectomized; their illness ranged from mild to severe, and none died. The 32 cases of *B. venatorum* infection reported from northeastern China occurred in spleen-intact residents. Symptoms were similar to those evoked by *B. microti*, although chills were rare. Seven patients were hospitalized for intermittent fever as high as 40°C. Only 4 patients were treated with clindamycin (without quinine); all 32 patients recovered. Cases of *B. crassa*–like infection reported from northeastern China also occurred in spleen-intact residents. Fever, fatigue, and myalgia were less common than among patients infected with *B. microti*, but headache was as common and nausea or vomiting more common. No patient was admitted to a hospital. Only 3 of the 31 patients were given clindamycin (without quinine); none died. In Europe, 2 cases of *B. crassa*–like infection occurred in splenectomized individuals and were severe. Neither ended in death, although 1 patient required intensive care. Cases of *B. motasi*–like infection in South Korea were severe; 1 patient died but the other recovered following clindamycin monotherapy.

■ PATHOGENESIS

Anemia RBC debris generated by hemolysis may accumulate in the kidney vasculature and cause renal failure. Free hemoglobin is rapidly complexed by haptoglobin. Once haptoglobin is depleted, the heme group is oxidized, released from hemoglobin, and complexed by hemopexin. Excessive hemolysis results in excess free hemoglobin, which scavenges and consumes nitric oxide, leading to thrombus formation and vascular inflammation. RBC lysis in small capillaries of the spleen may trigger localized necrosis, leading to splenic infarction. Exposed to oxidative stress, RBCs become poorly deformable and are filtered out by splenic macrophages as they attempt to pass through the red pulp. Erythrophagocytosis, along with the mounting of an immune response, contributes to splenomegaly and splenic rupture. Massive phagocytosis of RBCs, platelets, and white blood cells has led to hemophagocytic lymphohistiocytosis, a fatal condition. Persistent anemia, despite resolution of infection, has been attributed to autoantibodies that tag RBCs for clearance and may activate the complement system.

Inflammation Fever, chills, and sweats likely result from the systemic inflammatory response triggered by RBC lysis. Excessive inflammation promotes end-organ pathology, leading to renal and pulmonary compromise. The spleen is the immunodominant organ in babesiosis. This feature, along with the clearance of parasitized RBCs, explains why asplenia is a major risk factor for severe disease. Protective immunity involves CD4$^+$ T cells, particularly T$_h$1 cells, as revealed by high-grade parasitemia in mice depleted of CD4$^+$ T cells or treated with an interferon γ–neutralizing antibody. The importance of CD4$^+$ T cells is corroborated by the severity of babesiosis in patients with AIDS and in allograft recipients. In immunocompetent mice, Tfh cell expansion is accompanied with germinal center formation and antibody secretion. Although nearly every immunocompetent patient tests positive for *B. microti* antibody at diagnosis, a role for antibodies is uncertain. Persistent, relapsing babesiosis often occurs in patients treated with rituximab for a cancer or an autoimmune disorder; this observation suggests that B cells, and presumably antibodies, are critical for parasite clearance in some individuals.

■ DIAGNOSIS

A diagnosis of babesiosis should be considered for patients who experience symptoms compatible with babesiosis and may have been exposed to ticks in an endemic area, particularly from late spring to early fall, or were transfused with blood components, particularly packed RBCs, in the past 6 months. Given that *I. scapularis* ticks can be coinfected with *B. microti* and *B. burgdorferi*, babesiosis should be considered in any Lyme disease patient for whom symptoms worsen or do not abate within days or weeks of initiation of appropriate antibiotic therapy. Conversely, because one-half of patients diagnosed with babesiosis are infected with *B. burgdorferi*, a diagnosis of babesiosis should prompt a diagnostic evaluation for Lyme disease. Other tick-borne pathogens, although rarely implicated in cases of coinfection with *B. microti*, may be considered; these include *Anaplasma phagocytophilum* **(Chap. 192)** and *Borrelia miyamotoi* **(Chap. 190)**.

Routine Laboratory Testing The complete blood count often is remarkable for anemia. An elevated reticulocyte count signifies stress-induced erythropoiesis. Low levels of haptoglobin or elevated levels of lactate dehydrogenase are consistent with hemolysis. Severe anemia often is preceded by severe thrombocytopenia. The white blood cell (WBC) count is reduced, unchanged, or elevated. Elevated levels of alkaline phosphatase, aspartate aminotransferase, and alanine aminotransferase signify hepatocyte injury. Elevated total bilirubin levels result from hemolysis but may also denote hepatic compromise. Elevated levels of blood urea nitrogen and serum creatinine indicate renal compromise. Urinalysis may reveal excess urobilinogen, hemoglobinuria, and/or proteinuria. Given that babesiosis is an imitator of HELLP (hemolysis, elevated liver enzymes, and low platelet count) syndrome, a diagnosis of babesiosis should be considered for pregnant women who are at risk of tick exposure and have laboratory abnormalities that define this syndrome.

There is no consensus on the use of a particular laboratory parameter as predictor of severe babesiosis. In a study of severe disease as defined above (see "Clinical Manifestations"), a total bilirubin level of >1.9 mg/dL was highly predictive of severe disease, whereas WBC counts of $<5 \times 10^3/\mu L$ were associated with a better prognosis. Parameters associated with severe disease also included WBC counts of $>10 \times 10^3/\mu L$ and creatinine levels of >1.2 mg/dL. An earlier study identified alkaline phosphatase levels of >125 IU/L and WBC counts of $>5 \times 10^3/\mu L$ as strong predictors of severe disease, in this case defined as a hospital stay of >2 weeks, an ICU stay of >2 days, or death.

Specific Testing A definitive diagnosis of babesiosis is made by microscopic examination of Giemsa-stained thin blood smears

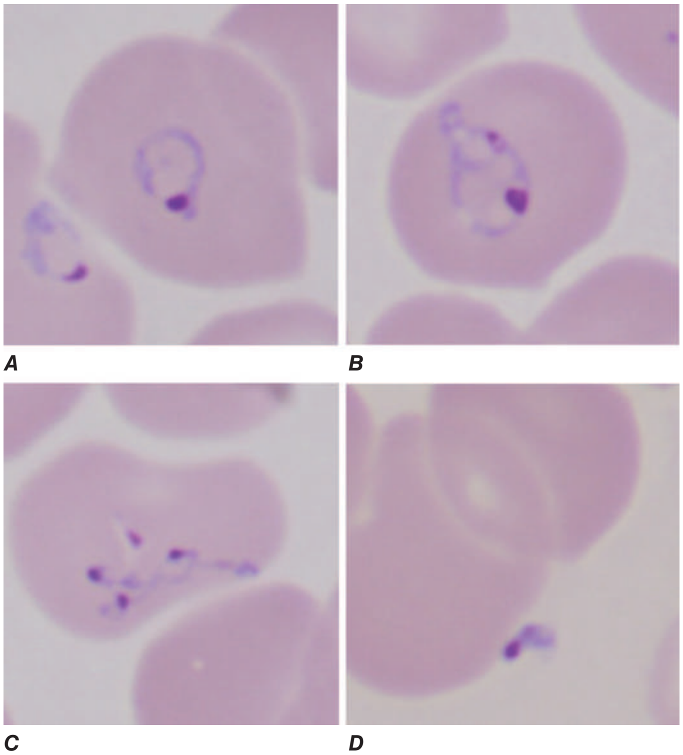

FIGURE 232-2 Giemsa-stained thin blood films showing *Babesia microti* parasites. *B. microti* is an obligate parasite of erythrocytes. Trophozoites may appear as ring forms (***A***) or as ameboid forms (***B***). Merozoites can be arranged in tetrads that are pathognomonic (***C***). Extracellular parasites can be noted (***D***), particularly when parasitemia is high. *(Reproduced with permission from E Vannier, PJ Krause: Human babesiosis. N Engl J Med 366:2397, 2012.)*

(Fig. 232-2) or amplification of *Babesia* DNA in blood. *Babesia* trophozoites appear round, oval, or ameboid. The ring form is most common and lacks the central brownish (hemozoin) deposit typical of *Plasmodium falciparum* late-stage trophozoites **(see Fig. A2-1)**. For travelers who have returned from *P. falciparum*–endemic areas and reside in a *Babesia*-endemic area, a negative result in the BinaxNOW malaria test readily rules out falciparum malaria when microscopy cannot. The presence of extracellular merozoites, particularly when parasitemia is high, and the absence of gametocytes and schizonts also distinguish babesiosis from malaria. Merozoites are arranged in pairs and occasionally in tetrads (the “Maltese cross”). Tetrads are pathognomonic of babesiosis and can be seen in human erythrocytes infected with *B. microti*, *B. duncani*, *B. venatorum*, or *B. divergens*–like organisms. Parasitemia typically ranges from 0.1 to 10% in immunocompetent patients but has reached 30–40% in immunocompromised patients. Parasitemia of >4% is a risk factor for severe, complicated disease. Splenic complications, however, often occur in patients with low-grade parasitemia (median, 1%; range, 0.1–30%).

If parasites cannot be identified by microscopy and babesiosis is still suspected, amplification of *Babesia* DNA is recommended. Real-time polymerase chain reaction (PCR) assays, which amplify the parasite 18S rRNA gene, detect as few as 1–10 parasites/μL of blood. Use of a fluorescent probe allows for speciation of the causative agent. Real-time PCR is recommended for the monitoring of low-grade infection in patients at risk of relapsing babesiosis.

A single positive serologic result is not sufficient to establish a diagnosis of babesiosis because antibodies can persist for >1 year after the illness has resolved and the parasite has been cleared. An indirect fluorescent antibody test is most commonly used. For *B. microti*, IgM titers of ≥1:20 and IgG titers of ≥1:64 are considered positive. IgG titers of ≥1:1024 suggest active or recent infection. Antibodies to *B. microti* do not react with *B. duncani* or *B. divergens* antigen. Sera from patients infected with *B. venatorum* or *B. crassa*–like organisms react with *B. divergens* antigen.

TREATMENT

Babesiosis

MILD TO MODERATE *B. MICROTI* ILLNESS

Mild to moderate babesiosis caused by *B. microti* is treated with atovaquone plus azithromycin administered orally for 7–10 days. Dosages for adults and children are provided in **Table 232-1**. Symptoms usually abate within 48 h after initiation of therapy and resolve within 1–2 weeks. If symptoms persist despite therapy for babesiosis, initiating doxycycline while testing for *B. burgdorferi* or other tick-borne pathogens such as *A. phagocytophilum* and *B. miyamotoi* is essential. Fatigue may persist for weeks to months but does not warrant, on its own, that treatment be extended or resumed. Parasite DNA can be detected for as long as 3 months, but follow-up PCR testing is not recommended because relapse of infection in immunocompetent individuals is unlikely.

SEVERE *B. MICROTI* ILLNESS

First-Line Antimicrobial Therapy The preferred regimen for the treatment of severe babesiosis caused by *B. microti* is oral atovaquone plus IV azithromycin (Table 232-1). Use of this combination is supported by a retrospective study of 40 patients who were admitted for severe babesiosis, including 11 who were admitted to the ICU. All but 1 of the 40 patients improved following initiation of treatment with atovaquone plus azithromycin, and were discharged. Clindamycin plus quinine, the first regimen to ever bring cure to a babesiosis patient, is an alternative choice. Use of quinine is impeded by a risk for QTc prolongation and cinchonism, including tinnitus; these adverse events often require premature discontinuation. A prospective, nonblind, randomized clinical trial established that atovaquone plus azithromycin is as effective as clindamycin plus quinine in clearing *B. microti* parasites and resolving symptoms of non-life-threatening babesiosis. Adverse events were reported by 15% of patients treated with atovaquone plus azithromycin but 72% of patients treated with clindamycin plus quinine. No trial has compared the 2 regimens for treatment of severe babesiosis.

Intravenous azithromycin should be initiated at 500 mg/d, along with atovaquone. Laboratory parameters should be monitored daily until symptoms abate and parasitemia is <4%. Thereafter, azithromycin can be administered orally. If the patient has an intact spleen and is not immunocompromised, the dosage can be reduced to 250 mg/d. The regimen is administered for 7–10 days, but the duration should be extended if symptoms persist. If the patient is asplenic or immunocompromised, azithromycin should be maintained at 500 mg/d. Given the risk for prolonged or relapsing babesiosis in such patients, the regimen should be administered until symptoms have resolved and parasites are no longer seen on blood smear for at least 2 weeks.

Adjunct Exchange Transfusion RBC exchange (RCE) is recommended when parasitemia is high and the clinical status of the patient deteriorates in the context of complications such as severe hemolysis or critical organ dysfunction, particularly renal or pulmonary compromise. The primary purpose of RCE is to rapidly reduce parasite burden; RCE also corrects anemia. Plasma exchange seldom is used in severe, complicated babesiosis although it removes circulating inflammatory mediators and byproducts of hemolysis such as free hemoglobin, free heme, and unconjugated bilirubin. Therapeutic apheresis is performed in close consultation with transfusion medicine services.

The criteria for RCE are not strictly defined. It is common practice to initiate RCE when parasitemia is >10%, although no study has been conducted to identify the threshold above which RCE provides greatest benefit. A recent case series illustrates the limitation of using parasitemia as the sole criterion to initiate RCE in the hope of preventing death. In this study of 19 patients, pre-RCE parasitemia (mean, 12.9%; 95% confidence interval [CI], 9.4–16.4) was a predictor of post-RCE length of hospital stay but not of mortality;

TABLE 232-1 Treatment of Human Babesiosis

ADULTS	CHILDREN
Mild to Moderate *B. microti* Infection[a]	
Atovaquone (750 mg q12h PO) *plus* Azithromycin (500 mg/d PO on day 1 followed by 250 mg/d PO on subsequent days)	Atovaquone (20 mg/kg q12h PO; maximum, 750 mg/dose) *plus* Azithromycin (10 mg/kg qd PO on day 1 [maximum, 500 mg], 5 mg/kg qd PO thereafter [maximum, 250 mg])
Severe *B. microti* Infection[b,c]	
Preferred[d] Atovaquone (750 mg q12h PO) *plus* Azithromycin (500 mg qd IV followed by 250–500 mg qd PO)	**Preferred** Atovaquone (20 mg/kg q12h PO; maximum, 750 mg/dose) *plus* Azithromycin (10 mg/kg qd IV followed by 10 mg/kg qd PO [maximum, 500 mg])
Alternative[e,f] Clindamycin (600 mg q6h IV followed by 600 mg q8h PO) *plus* Quinine (650 mg q6–8h PO) Consider exchange transfusion	**Alternative** Clindamycin (7–10 mg/kg q6–8h IV followed by 7–10 mg/kg q6–8h PO [maximum, 600 mg/dose]) *plus* Quinine (8 mg/kg q8h PO; maximum, 650 mg/dose) Consider exchange transfusion
***B. divergens* Infection**[g]	
Immediate complete exchange transfusion *plus* Clindamycin (600 mg q6–8h IV) *plus* Quinine (650 mg q8h PO)	Immediate complete exchange transfusion *plus* Clindamycin (7–10 mg/kg q6–8h IV; maximum, 600 mg/dose) *plus* Quinine (8 mg/kg q8h PO; maximum, 650 mg/dose)

[a] Treat for 7–10 days. [b] Treat for 7–10 days, but extend duration if symptoms persist. [c] For severely immunocompromised patients, antimicrobial therapy should be given for at least 6 consecutive weeks, including 2 final weeks during which parasites are no longer detected on blood smear. [d] If the risk of QTc prolongation or allergy associated with use of azithromycin is a concern, clindamycin can be substituted for azithromycin. For severely immunocompromised patients, IV clindamycin can be added to atovaquone plus azithromycin at initiation of treatment. [e] Clindamycin plus quinine is no longer the preferred regimen because quinine often is discontinued due to QTc prolongation or other side effects, including tinnitus. This regimen can be considered for cases that respond poorly to atovaquone plus azithromycin. [f] Other alternative regimens have been used successfully, as documented in a limited number of case reports. If quinine toxicity is a concern, atovaquone can be substituted for quinine. For cases that respond poorly to atovaquone plus azithromycin, atovaquone-proguanil can be added to the two-drug regimen or can be substituted for atovaquone. [g] A few cases of *B. divergens* infection in Europe have been treated successfully with atovaquone plus azithromycin or atovaquone-proguanil plus azithromycin.

this finding is consistent with the concept that end-organ dysfunction is a correlate of host inflammation rather than parasite burden. Post-RCE parasitemia (mean, 3.4%; 95% CI, 1.9–4.9) was associated neither with post-RCE length of hospital stay nor with mortality; this finding advocates against the use of repeat RCE. There was a trend toward an association between creatinine levels at admission and mortality, pointing to the greater benefit of RCE in patients with renal compromise. RCE has been withheld when the patient is stable and parasitemia has begun to resolve following initiation of antimicrobial therapy. In a study that compared 9 such patients with 19 patients who underwent RCE and received antimicrobial therapy, the two groups differed neither by the parasitemia at admission (mean, 11.0% vs 11.6%) nor by the time to parasitemia <1% (day 7 vs day 8).

Highly Immunocompromised Patients Patients are at risk for persistent, relapsing babesiosis when the immune response to *Babesia* is severely impaired. Major risk factors include asplenia, HIV/AIDS, and immunosuppressive regimens given for transplantation, an autoimmune disorder, or malignancy. Many of these regimens include rituximab or another B cell–depleting antibody. Antimicrobial therapy should be administered for at least 6 consecutive weeks, including 2 weeks beyond the first definitive negative PCR test for *B. microti*. Given the duration of treatment, atovaquone plus azithromycin is the preferred regimen. Azithromycin should be administered intravenously and initiated at 500 mg/d. Laboratory parameters should be monitored daily until symptoms abate and parasitemia is <4%. Thereafter, azithromycin can be administered orally, but the dosage should be maintained at 500 mg/d because lower dosages may promote antimicrobial resistance. Once the patient is no longer critically ill, routine laboratory testing can be performed every 2 to 3 days. When parasites are no longer observed on blood smear, real-time PCR should be used to monitor the infection. After discontinuation of therapy, close follow-up is recommended. If symptoms recur, blood smears and/or real-time PCR should be ordered.

Antimicrobial Resistance Parasitologic or clinical relapse has been documented in highly immunocompromised patients, particularly when antimicrobial therapy is interrupted or administered at a reduced dosage. Some patients who relapse while or after being treated with atovaquone plus azithromycin have been managed with clindamycin plus quinine (Table 232-1, footnote e). When quinine toxicity is a concern, clindamycin has been added to atovaquone plus azithromycin. An informed approach is to identify mutations that have arisen in the parasite genome and may explain antimicrobial resistance. Atovaquone targets the parasite cytochrome b (Cytb). When resistance to atovaquone is predicted to be partial, atovaquone-proguanil can be substituted for atovaquone (Table 232-1, footnote f). The rationale is that, as with *P. falciparum*, proguanil lowers the concentration of atovaquone required to collapse the mitochondrial membrane potential. Resistance to azithromycin has been attributed to missense mutations in the ribosomal protein subunit L4 gene (*RLP4*) and to a mutation in domain V of the 23S rRNA gene, which carries the peptidyl transferase activity of the parasite ribosome. Domain V is also the target of clindamycin. When resistance to azithromycin is explained by a mutation in *RPL4*, clindamycin can be substituted for azithromycin. A mutation in domain V that prevents binding of azithromycin and clindamycin precludes substitution of one for the other. Such mutation was recently identified in a patient who relapsed while being treated with a three-drug regimen that included azithromycin and clindamycin. Following initiation of tafenoquine at the loading dose of 600 mg (given over 3 consecutive days), parasitemia rapidly declined. Tafenoquine was administered along with atovaquone and proguanil. Maintenance therapy, which consisted of a 300-mg weekly dose of tafenoquine, was discontinued 5 weeks after the last positive *B. microti* PCR test. Tafenoquine, which acts by inducing reactive oxygen species, is well suited to substitute for drugs that target a particular gene product. Tafenoquine should not be prescribed to patients with glucose-6-phosphate dehydrogenase deficiency because their RBCs have reduced antioxidant capacity and would not sustain the oxidative stress caused by tafenoquine.

Splenic Rupture Splenic rupture typically occurs in young, healthy patients with low-grade parasitemia. If the patient is hemodynamically unstable, emergent splenectomy should be performed. If the patient is hemodynamically stable but bleeding persists, splenic arterial embolization should be considered. In the absence of hemoperitoneum, splenic rupture should be managed without surgery but with careful hemodynamic monitoring. Removal of the spleen leaves patients at risk for relapsing babesiosis or severe disease caused by other microorganisms.

OTHER *BABESIA* INFECTIONS

B. duncani and *B. divergens*–like infections typically have been treated with IV clindamycin (600 mg three or four times daily or 1200 mg twice daily) plus oral quinine (600–650 mg three times daily) for 7–10 days. RBC exchange is initiated for patients with severe hemolytic anemia and/or end-organ failure. In a recent case of severe *B. divergens*-like infection, a RBC exchange was performed and cure was achieved following a 2-week course of atovaquone plus azithromycin.

GLOBAL CONSIDERATIONS

In Europe, *B. divergens* infection is considered a medical emergency. The recommended approach is immediate, complete RCE combined with administration of clindamycin plus oral quinine (Table 232-1). Some cases have been cured with RCE and clindamycin monotherapy. Anemia may persist for >1 month and require blood transfusion. A severe case of *B. divergens* infection resolved during therapy with atovaquone plus azithromycin. A relapse in a spleen-intact individual was treated with atovaquone-proguanil plus azithromycin. The first-line therapy for *B. venatorum* infection in Europe has been IV or oral clindamycin plus quinine. In a patient intolerant to quinine, infection was cured after administration of atovaquone plus azithromycin. A pediatric case of mild *B. venatorum* infection in China was successfully treated by a standard course of atovaquone plus azithromycin.

PREVENTION

Given the lack of vaccine and chemoprophylaxis, individuals who reside in endemic areas, especially those at risk of severe babesiosis, should wear protective clothing, apply tick repellents to the skin and permethrin to clothing, and limit outdoor activities where ticks abound from May through October. The skin should be thoroughly examined after outdoor activities and ticks carefully removed with tweezers. As babesiosis continues to expand into new areas and because climate change has begun to reshape this expansion, physicians should be increasingly aware of this once neglected disease.

FURTHER READING

Krause PJ et al: Clinical practice guidelines by the Infectious Diseases Society of America (IDSA): 2020 guideline on diagnosis and management of babesiosis. Clin Infect Dis 72:185, 2021.

Rogers R et al: Broad antimicrobial resistance in a case of relapsing babesiosis successfully treated with tafenoquine. Clin Infect Dis 76:741, 2023.

Swanson M et al: Trends in reported babesiosis cases — United States, 2011–2019. MMWR Morb Mortal Wkly Rep 72:273, 2023.

Tannous T et al: Red cell exchange as adjunctive therapy for babesiosis: Is it really effective? Transfus Med Rev 35:16, 2021.

Tonnetti L et al: *Babesia* blood testing: The first-year experience. Transfusion 62:135, 2022.

233

Leishmaniasis

Shyam Sundar

Encompassing a complex group of disorders, leishmaniasis is caused by unicellular eukaryotic obligatory intracellular protozoa of the genus *Leishmania* and primarily affects the host's reticuloendothelial system. *Leishmania* species produce widely varying clinical syndromes ranging from self-healing cutaneous ulcers to fatal visceral disease. These syndromes fall into three broad categories: visceral leishmaniasis (VL), cutaneous leishmaniasis (CL), and mucosal leishmaniasis (ML).

ETIOLOGY AND LIFE CYCLE

Leishmaniasis is caused by ~20 species of the genus *Leishmania* in the order Kinetoplastida and the family Trypanosomatidae **(Table 233-1)**. Several clinically important species are of the subspecies *Viannia*. The organisms are transmitted by phlebotomine sandflies of the genus *Phlebotomus* in the "Old World" (Asia, Africa, and Europe) and the genus *Lutzomyia* in the "New World" (the Americas). Transmission may be anthroponotic (i.e., the vector transmits the infection from infected humans to healthy humans) or zoonotic (i.e., the vector transmits the infection from an animal reservoir to humans). Human-to-human transmission via shared infected needles has been documented in IV drug users in the Mediterranean region. In utero transmission to the fetus occurs rarely.

Leishmania organisms occur in two forms: extracellular, flagellate promastigotes (length, 10–20 μm) in the sandfly vector and intracellular, nonflagellate amastigotes (length, 2–4 μm; **Fig. 233-1** in vertebrate hosts, including humans. Promastigotes are introduced through the proboscis of the female sandfly into the skin of the vertebrate host. Neutrophils predominate among the host cells that first encounter and take up promastigotes at the site of parasite delivery. The infected neutrophils may undergo apoptosis and release viable parasites that are taken up by macrophages, or the apoptotic cells may themselves be taken up by macrophages and dendritic cells. The parasites multiply as amastigotes inside macrophages, causing cell rupture with subsequent invasion of other macrophages. While feeding on infected hosts, sandflies pick up amastigotes, which transform into the flagellate form in the flies' posterior midgut and multiply by binary fission; the promastigotes then migrate to the anterior midgut and can infect a new host when flies take another blood meal.

EPIDEMIOLOGY

Leishmaniasis occurs in 99 countries—most of them developing—in tropical and temperate regions **(Fig. 233-2)**. More than 1 billion people live in areas endemic for leishmaniasis and are at risk of infection. An estimated 30,000 new cases of VL and more than 1 million new cases of CL occur annually. CL is common in South America, Africa, and Asia, whereas VL commonly occurs in East Africa, Brazil, and the Indian subcontinent. ML is limited to South America. The distribution of *Leishmania* is limited by the distribution of sandfly vectors.

VISCERAL LEISHMANIASIS

VL (also known as *kala-azar*, a Hindi term meaning "black fever") is caused by the *Leishmania donovani* complex, which includes *L. donovani* and *Leishmania infantum*; these species are responsible for anthroponotic and zoonotic transmission, respectively. East Africa now has the highest incidence of VL, followed by Brazil and the Indian subcontinent. In the Indian subcontinent (India, Nepal, and Bangladesh), where a VL elimination program has been implemented, VL incidence has markedly declined. In these three countries, the World Health Organization (WHO) recently announced the elimination of VL as a public health problem in Bangladesh and a steep decline in the incidence in Nepal, and India reported a 98.7% decline. Zoonotic VL is reported from all countries in the Middle East, Pakistan, and other countries from western Asia to China. Endemic foci also exist in the independent states of the former Soviet Union, mainly Georgia and Azerbaijan. In the Horn of Africa, Sudan, South Sudan, Ethiopia, Kenya, Uganda, and Somalia report VL. In Sudan and South Sudan, large outbreaks are thought to be anthroponotic, although zoonotic transmission also occurs. VL is rare in West and sub-Saharan Africa.

Mediterranean VL, long an established endemic disease due to *L. infantum*, has a large canine reservoir and was seen primarily in infants before the advent of HIV infection. In Mediterranean Europe, 70% of adult VL cases are associated with HIV co-infection. The combination is deadly because of the combined impact of the two infections on the immune system. IV drug users are at particular risk. Other forms of immunosuppression (e.g., that associated with organ transplantation) also predispose to VL. In the Americas, disease caused by *L. infantum* is endemic from Mexico to Argentina, but 90% of cases in the New World are reported from northeastern Brazil. After the introduction of highly active antiretroviral therapy, the incidence of HIV–VL co-infection declined significantly in Europe; however, ~30 and 5% of VL patients are co-infected with HIV in Ethiopia and India, respectively.

Immunopathogenesis The majority of individuals infected by *L. donovani* or *L. infantum* mount a successful immune response and control the infection, never developing symptomatic disease. Forty-eight hours after intradermal injection of killed promastigotes, these individuals exhibit delayed-type hypersensitivity (DTH) to leishmanial antigens in the leishmanin skin test (also called the Montenegro

TABLE 233-1 Geographic Distribution and Characteristic Epidemiology of Leishmaniases

ORGANISM, ENDEMIC REGION	CLINICAL SYNDROME	SPECIES	VECTOR	RESERVOIR	TRANSMISSION	SETTING
***Leishmania donovani* Complex**						
South Asia	VL, PKDL	*L. donovani*	*Phlebotomus argentipes*	Humans	Anthroponotic	Rural, domestic
Sudan, South Sudan, Somalia, Ethiopia, Kenya, Uganda	VL, PKDL	*L. donovani*	*P. orientalis,* *P. martini*	Humans, rodents in Sudan, canines	Anthroponotic, occasionally zoonotic	Majority peridomestic, occasionally sylvatic
Mediterranean basin, Middle East, Central Asia, China	VL, CL	*L. infantum*	*P. perniciosus,* *P. ariasi*	Dogs, foxes, jackals	Zoonotic	Domestic, peridomestic
Middle East, Saudi Arabia, Yemen	VL	*L. donovani*	*P. perniciosus,* *P. ariasi*	Dogs, foxes, jackals	Zoonotic	Domestic, peridomestic
Central and South America	VL, CL	*L. infantum*[a]	*Lutzomyia longipalpis*	Foxes, dogs, opossums	Zoonotic	Domestic, peridomestic, periurban
Azerbaijan, Armenia, Georgia, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan	VL	*L. infantum*	*P. turanicus*	Humans, dogs, foxes	Anthroponotic, zoonotic	Domestic
L. tropica						
Western India to Turkey, parts of North and East Africa	CL, leishmaniasis recidivans	*L. tropica*	*P. sergenti*	Humans	Anthroponotic	Urban domestic, peridomestic
L. major						
Western and Central Asia, North and sub-Saharan Africa	CL	*L. major*	*P. papatasi,* *P. duboscqi*	Nile rats, rodents	Zoonotic	Sylvatic, peridomestic
Kazakhstan, Turkmenistan, Uzbekistan	CL	*L. major*	*P. papatasi,* *P. duboscqi*	Gerbils	Zoonotic	Rural
L. aethiopica						
Ethiopia, Uganda, Kenya	CL, DCL	*L. aethiopica*	*P. longipes,* *P. pedifer*	Hyraxes	Zoonotic	Sylvatic, peridomestic
Subspecies *Viannia*						
Peru, Ecuador	CL, ML	*L. (V.) peruviana*	*Lutzomyia verrucarum,* *L. peruensis*	Wild rodents	Zoonotic	Andean Valleys
Guyana, Surinam, French Guyana, Ecuador, Brazil, Colombia, Bolivia	CL, ML	*L. (V.) guyanensis*	*L. umbratilis*	Sloths, arboreal anteaters, opossums	Zoonotic	Tropical forest
Central America, Ecuador, Colombia	CL, ML	*L. (V.) panamensis*	*L. trapidoi*	Sloths	Zoonotic	Tropical forest and deforested areas
South and Central America	CL, ML	*L. (V.) braziliensis*	*Lutzomyia* spp., *L. umbratilis,* *Psychodopygus wellcomei*	Forest rodents, peridomestic animals	Zoonotic	Tropical forest and deforested areas
***L. mexicana* Complex**						
Central America and northern parts of South America	CL, ML, DCL	*L. amazonensis*	*L. flaviscutellata*	Forest rodents	Zoonotic	Tropical forest and deforested areas
	CL, ML, DCL	*L. mexicana*	*L. olmeca*	Variety of forest rodents and marsupials	Zoonotic	Tropical forest and deforested areas
	CL, DCL	*L. pifanoi*	*L. olmeca*	Variety of forest rodents and marsupials	Zoonotic	Tropical forest and deforested areas

[a] *L. infantum* is designated *L. chagasi* in the New World.

Abbreviations: CL, cutaneous leishmaniasis; DCL, diffuse cutaneous leishmaniasis; ML, mucosal leishmaniasis; PKDL, post–kala-azar dermal leishmaniasis; VL, visceral leishmaniasis.

skin test). Results in mouse models indicate that the development of acquired resistance to leishmanial infection is controlled by the production of interleukin (IL) 12 by antigen-presenting cells and the subsequent secretion of interferon (IFN) γ, tumor necrosis factor (TNF) α, and other proinflammatory cytokines by the T helper 1 (T_H1) subset of T lymphocytes. The immune response in patients developing active VL is complex; in addition to increased production of multiple proinflammatory cytokines and chemokines, patients with active disease have markedly elevated levels of IL-10 in serum as well as enhanced IL-10 mRNA expression in lesional tissues. A direct role for IL-10 in the pathology of VL in humans is supported by studies demonstrating that IL-10 blockade can enhance antigen-specific IFN-γ responses in whole blood from VL patients and IL-10 neutralization promotes parasitic killing. The main disease-promoting activity of IL-10 in VL may be to condition host macrophages for enhanced survival and growth of the parasite. IL-10 can render macrophages unresponsive to activation signals and inhibit killing of amastigotes by downregulating the production of TNF-α and nitric oxide. Multiple antigen-presentation functions of dendritic cells and macrophages are also suppressed by IL-10. Patients with such suppression do not have positive leishmanin skin tests, nor do their peripheral-blood mononuclear cells respond to leishmanial antigens in vitro. Organs of the reticuloendothelial system

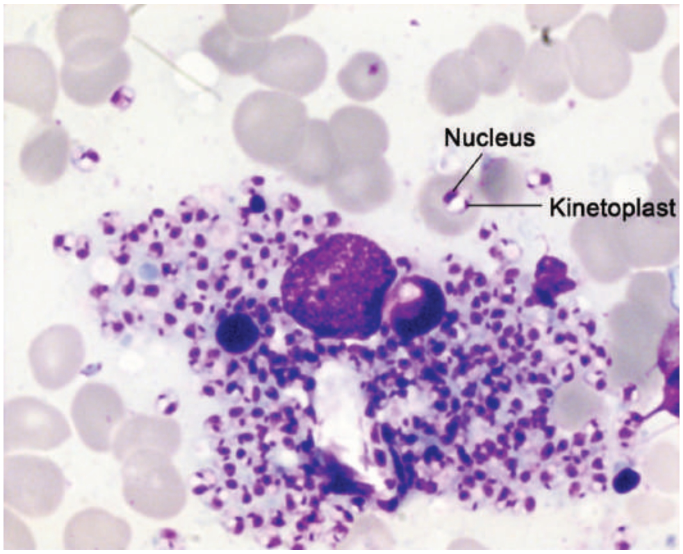

FIGURE 233-1 A macrophage with numerous intracellular amastigotes (2–4 μm) in a Giemsa-stained splenic smear from a patient with visceral leishmaniasis. Each amastigote contains a nucleus and a characteristic kinetoplast consisting of multiple copies of mitochondrial DNA. A few extracellular parasites are also visible.

are predominantly affected, with remarkable enlargement of the spleen, liver, and lymph nodes in some regions. The tonsils and intestinal submucosa are also heavily infiltrated with parasites. Bone marrow dysfunction results in pancytopenia.

Clinical Features On the Indian subcontinent and in the Horn of Africa, persons of all ages are affected by VL. In endemic areas of the Americas and the Mediterranean basin, immunocompetent infants and small children as well as immunodeficient adults are affected especially often. The incubation period is 2–3 months but may be up to 1 year or more. The most common presentation of VL is an abrupt onset of moderate- to high-grade fever associated with rigor and chills. Fever may continue for several weeks with decreasing intensity, and the patient may become afebrile for a short period before experiencing another bout of fever. The spleen may be palpable by the second week of illness and, depending on the duration of illness, may become hugely enlarged **(Fig. 233-3)**. Hepatomegaly (usually moderate) soon follows. Lymphadenopathy is common in most endemic regions of the world except the Indian subcontinent, where it is rare. Patients lose weight and feel weak, and the skin gradually develops dark discoloration due to hyperpigmentation that is most easily seen in brown-skinned individuals. In advanced illness, hypoalbuminemia may manifest as pedal edema and ascites. Anemia appears early and may become severe enough to cause congestive heart failure. Epistaxis, retinal hemorrhages, and gastrointestinal bleeding are associated with thrombocytopenia. Secondary infections such as pneumonia, tuberculosis, bacillary or amebic dysentery, and gastroenteritis are common. Herpes zoster, chickenpox, boils in the skin, and scabies may also occur. Untreated, the disease is fatal in most patients, including 100% of those with HIV co-infection.

Leukopenia and anemia occur early and are followed by thrombocytopenia. There is a marked polyclonal increase in serum immunoglobulins. Serum levels of hepatic aminotransferases are raised in a significant proportion of patients, and serum bilirubin levels are elevated occasionally. Renal dysfunction is uncommon.

Laboratory Diagnosis Demonstration of amastigotes in smears of tissue aspirates remains the gold standard for the diagnosis of VL (Fig. 233-1). The sensitivity of splenic smears is >95%, whereas smears of bone marrow (60–85%) and lymph node aspirates (50%) are less sensitive. Culture of tissue aspirates increases sensitivity. Splenic aspiration is invasive and may be dangerous in untrained hands. To circumvent these invasive procedures, several serologic techniques are currently used to detect antibodies to *Leishmania*. An enzyme-linked immunosorbent assay (ELISA) and the indirect immunofluorescent antibody test (IFAT) are used in sophisticated laboratories.

In the field, however, a rapid immunochromatographic test based on the detection of antibodies to a recombinant antigen (rK39) consisting of 39 amino acids conserved in the kinesin region of *L. infantum* is used worldwide. The test requires only a drop of fingerprick blood or serum, and the result can be read within 15 min. Except in East Africa (where both its sensitivity and specificity are lower), the sensitivity of the rK39 rapid diagnostic test (RDT) in immunocompetent individuals is ~98% and its specificity is ~90%. RDTs— based on a new synthetic polyprotein, rK28—performed better in Sudan, but are not available commercially. Since these antibody detection tests remain positive for years after cure, they cannot be used for measurement of cure or detection of relapse. Qualitative detection of leishmanial nucleic acid by polymerase chain reaction (PCR) and quantitative detection by

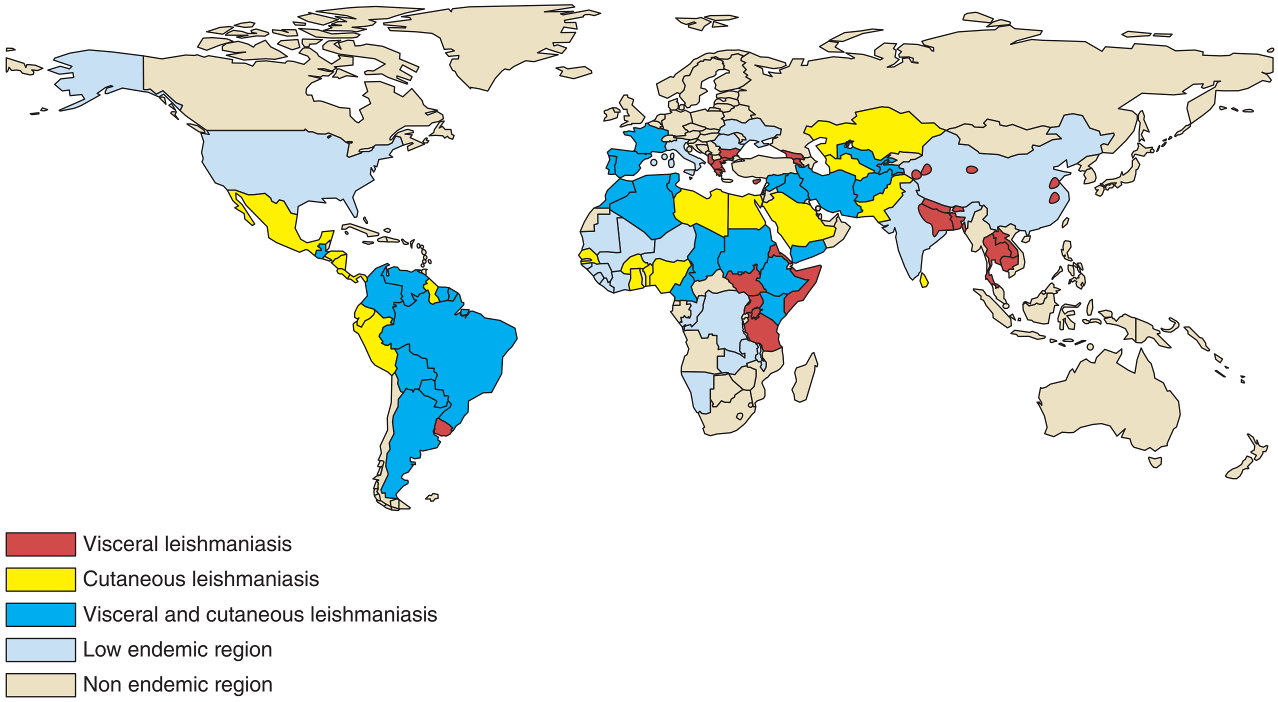

FIGURE 233-2 Worldwide distribution of human leishmaniasis.

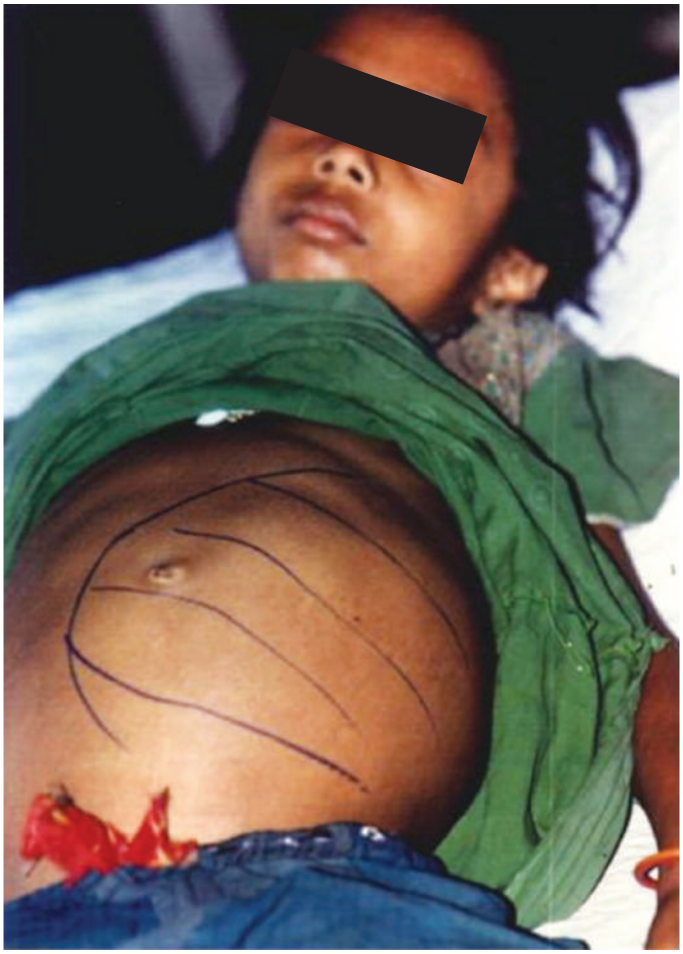

FIGURE 233-3 A patient with visceral leishmaniasis has a hugely enlarged spleen visible through the surface of the abdomen. Splenomegaly is the most important feature of visceral leishmaniasis.

real-time PCR are highly sensitive and have the added advantage of species identification. However, because the capacity to perform these tests is confined to specialized laboratories, they are yet to be used for routine diagnosis of VL in endemic areas. Using blood samples, loop-mediated isothermal amplification (LAMP) of nucleic acids performs better compared with microscopy. A simple visual readout with minimum laboratory equipment requirements makes it attractive for point-of-care diagnostics for VL and post–kala-azar dermal leishmaniasis (PKDL), and it is likely to be used increasingly in the future.

Differential Diagnosis VL is easily mistaken for malaria. Other febrile illnesses that may mimic VL include typhoid fever, tuberculosis, brucellosis, schistosomiasis, and histoplasmosis. Splenomegaly due to portal hypertension, chronic myeloid leukemia, tropical splenomegaly syndrome, and (in Africa) schistosomiasis may also be confused with VL. Fever with neutropenia or pancytopenia in patients from an endemic region strongly suggests a diagnosis of VL; hypergammaglobulinemia in patients with long-standing illness strengthens the diagnosis. In nonendemic countries, a careful travel history is essential when any patient presents with fever.

TREATMENT

Visceral Leishmaniasis

GENERAL CONSIDERATIONS

Severe anemia should be corrected by blood transfusion, and other comorbid conditions should be managed promptly. Treatment of VL is complex because the optimal drug, dosage, and duration vary with the endemic region. Despite completing recommended treatment, some patients experience relapse (most often within 6–12 months), and prolonged follow-up is recommended. A pentavalent antimonial is the drug of choice in most endemic regions of the world, but there is widespread resistance to antimony in the Indian state of Bihar, where either amphotericin B (AmB)—deoxycholate or liposomal—or miltefosine is preferred. Dose requirements for AmB are lower in India than in the Americas, Africa, or the Mediterranean region. In Mediterranean countries, where cost is seldom an issue, liposomal AmB (LAmB) is the drug of choice. In immunocompetent patients, relapses are uncommon with AmB in its deoxycholate and lipid formulations. Antileishmanial therapy has recently evolved as new drugs and delivery systems have become available and resistance to antimonial compounds has emerged.

Except for AmB (deoxycholate and lipid formulations), antileishmanial drugs are available in the United States only from the Centers for Disease Control and Prevention Drug Service (telephone: 404-639-3670; email: drugservice@cdc.gov).

PENTAVALENT ANTIMONIAL COMPOUNDS

Two pentavalent antimonial (Sb^V) preparations are available: sodium stibogluconate (100 mg of Sb^V/mL) and meglumine antimoniate (85 mg of Sb^V/mL). The daily dose is 20 mg/kg by IV infusion or IM injection, and therapy continues for 28–30 days. Cure rates exceed 90% in Africa, the Americas, and most of the Old World but are <50% in Bihar, India, as a result of resistance. Adverse reactions to Sb^V treatment are common and include arthralgia, myalgia, and elevated serum levels of aminotransferases. Electrocardiographic changes are common. Concave ST-segment elevation is not significant, but prolongation of QT_c to >0.5 s may herald ventricular arrhythmia and sudden death. Chemical pancreatitis is common but usually does not require discontinuation of treatment; severe clinical pancreatitis occurs in immunosuppressed patients.

AMPHOTERICIN B

AmB (and its liposomal formulations) is currently used as a first-line drug in the Indian Subcontinent. Because of a better safety profile, LAmB is now the preferred drug of choice in other parts of the world, as well. Conventional AmB deoxycholate is administered in doses of 0.75–1.0 mg/kg on alternate days for a total of 15 infusions. Fever with chills is an almost universal adverse reaction to AmB infusions. Nausea and vomiting are also common, as is thrombophlebitis in the infused veins. Acute toxicities can be minimized by administration of antihistamines like chlorpheniramine and antipyretic agents like acetaminophen before each infusion. AmB can cause renal dysfunction and hypokalemia and, in rare instances, elicits hypersensitivity reactions, bone marrow suppression, and myocarditis, all of which can be fatal.

Several lipid formulations of AmB, developed to replace the deoxycholate formulation, are preferentially taken up by reticuloendothelial tissues. Because very little free drug is available to cause toxicity, a large amount of drug can be delivered over a short period. LAmB has been used extensively to treat VL in all parts of the world. With a terminal half-life of ~150 h, LAmB can be detected in the liver and spleen of animals for several weeks after a single dose. In addition to oral miltefosine (see below), this is the only drug approved by the U.S. Food and Drug Administration (FDA) for the treatment of VL; the regimen is 3 mg/kg daily on days 1–5, 14, and 21 (total dose, 21 mg/kg). However, the total-dose requirement for different regions of the world varies widely. In Asia, it is 10–15 mg/kg; in Africa, ~18 mg/kg; and in Mediterranean/American regions, ≥20 mg/kg. The daily dose is flexible (1–10 mg/kg). In a study in India, a single dose of 10 mg/kg cured infection in 96% of patients. This single-dose regimen is used in the elimination program in the Indian subcontinent. Adverse effects of LAmB are usually mild and include infusion reactions, backache, and occasional reversible nephrotoxicity.

PAROMOMYCIN

Paromomycin (aminosidine) is an aminocyclitol-aminoglycoside antibiotic with antileishmanial activity. Its mechanism of action against *Leishmania* is yet to be established. Paromomycin is approved in India for the treatment of VL at an IM dose of 11 mg of base/kg daily for 21 days; this regimen produces a cure rate of 94.6%. However, the optimal dose has not been established in other endemic regions. Paromomycin is a relatively safe drug, but some patients develop hepatotoxicity, reversible ototoxicity, and (in rare instances) nephrotoxicity and tetany. Paromomycin, in combination with Sb^V, is used in sub-Saharan Africa.

MILTEFOSINE

Miltefosine, an alkylphosphocholine, is the first oral compound approved for the treatment of leishmaniasis in several endemic countries including the United States. This drug has a long half-life (150–200 h); its mechanism of action is not clearly understood. The recommended therapeutic regimens consist of 28 consecutive days of treatment in doses of 50 mg for patients weighing <25 kg, a twice-daily dose of 50 mg for patients weighing ≥25 kg, and 2.5 mg/kg for children 2–11 years of age. In children, allometric dosing has been suggested for adequate exposure to the drug. These regimens have resulted in a cure rate of ~94% in India. However, recent studies from the Indian subcontinent indicate a decline in the cure rate. Doses in other regions remain to be established. Because of its long half-life, miltefosine is prone to induce resistance in *Leishmania*. Its adverse effects include mild to moderate vomiting and diarrhea in 40 and 20% of patients, respectively; these reactions usually subside spontaneously after a few days. Rare instances of severe allergic dermatitis, hepatotoxicity, and nephrotoxicity have been reported. Because miltefosine is expensive and is associated with significant adverse events, it is best administered as directly observed therapy after meals to minimize the gastrointestinal adverse events, and to ensure completion of treatment and to minimize the risk of resistance induction. Because miltefosine is teratogenic in rats, its use is contraindicated during pregnancy and (unless contraceptive measures are strictly adhered to for at least 5 months after treatment) in women of childbearing age.

MULTIDRUG THERAPY

Multidrug therapy for leishmaniasis is likely to be preferred in the future. Its potential advantages in VL include (1) better compliance and lower costs associated with shorter treatment courses and decreased hospitalization, (2) less toxicity due to lower drug doses and/or shorter duration of treatment, and (3) a reduced likelihood that resistance to either agent will develop. In a study from India, one dose of LAmB (5 mg/kg) followed by miltefosine for 7 days, or paromomycin for 10 days, or both miltefosine and paromomycin simultaneously for 10 days (in their usual daily doses) produced a cure rate of >97% (all three combinations). In Africa, a combination of Sb^V and paromomycin given for 17 days was as effective and safe as Sb^V given for 30 days and is the preferred first-line treatment of VL. Recently, in a phase 3 trial from East Africa, this prevailing 17-day regimen was compared with combinations of paromomycin and miltefosine for 14 days as doses described above. The cure rates were similar, and this new combination has been proposed as an alternative to the standard 17-day regimen.

Prognosis of Treated VL Patients Recovery from VL is quick. Within a week after the start of treatment, defervescence, regression of splenomegaly, weight gain, and recovery of hematologic parameters are evident. With effective treatment, no parasites are recovered from tissue aspirates at the posttreatment evaluation. Continued clinical improvement over 12 months is suggestive of cure. A small percentage of patients (with the exact figure depending on the regimen used) relapse but respond well to retreatment with AmB or its lipid formulations.

VL in the Immunocompromised Host HIV/VL co-infection has been reported from 45 countries. Where both infections are endemic, VL behaves as an opportunistic infection in HIV-1-infected patients. HIV infection can increase the risk of VL development by several-fold in endemic areas. Co-infected patients usually show the classic signs of VL, but they may present with atypical features due to loss of immunity and involvement of unusual anatomic locations, e.g., infiltration of the skin, oral mucosa, gastrointestinal tract, lungs, and other organs. Serodiagnostic tests may be negative in up to 50% of patients. Parasites can be recovered from unusual sites such as bronchoalveolar lavage fluid and buffy coat. LAmB is the drug of choice for HIV/VL co-infection—both for primary treatment and for treatment of relapses. A total dose of 40 mg/kg, administered as 4 mg/kg on days 1–5, 10, 17, 24, 31, and 38, is considered optimal for the Americas and Mediterranean countries and is approved by the FDA, but most patients experience a relapse within 1 year. Recently, there has been a significant change in the WHO recommendations for treatment of HIV–VL co-infection in the Indian subcontinent and East Africa. LAmB infusions (total dose 30 mg/kg; 5 mg/kg on days 1, 3, 5, 7, 9, and 11) co-administered with oral miltefosine (100 mg/d for 14 days for the Indian subcontinent and 28 days for East Africa) should now be used to treat HIV–VL co-infection. AmB deoxycholate can also be used where LAmB is not accessible. Reconstitution of patients' immunity by antiretroviral therapy has led to a dramatic decline in the incidence of co-infection in the Mediterranean basin. In contrast, HIV–VL co-infection is on the rise in African and Asian countries. Ethiopia is worst affected: up to 30% of VL patients are also infected with HIV. Because restoration of the CD4+ T-cell count to >200/μL does decrease the frequency of relapse, antiretroviral therapy (in addition to antileishmanial therapy) is a cornerstone of the management of HIV–VL co-infection. Secondary prophylaxis with pentamidine or lipid AmB has been shown to delay relapses, but no regimen has been established as optimal.

Post–Kala-Azar Dermal Leishmaniasis On the Indian subcontinent and in Sudan and other East African countries, 2–50% of patients develop skin lesions concurrent with or after the cure of VL. Most common are hypopigmented macules, papules, and/or nodules or diffuse infiltration of the skin and sometimes of the oral mucosa. The African and Indian diseases differ in several respects; important features of PKDL in these two regions are listed in **Table 233-2**, and disease in an Indian patient is depicted in **Fig. 233-4**.

In PKDL, parasites are scanty in hypopigmented macules but may be seen and cultured more easily from nodular lesions. Cellular infiltrates are heavier in nodules than in macules. Lymphocytes are the dominant cells; next most common are histiocytes and plasma cells. In about half of cases, epithelioid cells—scattered individually or forming compact granulomas—are seen. The diagnosis is based on history, clinical findings supported by demonstration of parasites in the slit skin smears, or by demonstrating parasitic DNA in PCR. Sensitivity of skin smears is quite low, and positive serology does not help as it remains positive for several years after the cure of VL. Indian PKDL was treated with prolonged courses (up to 120 days) of pentavalent antimonials. This prolonged course was toxic and frequently led to noncompliance.

TABLE 233-2 Clinical, Epidemiologic, and Therapeutic Features of Post–Kala-Azar Dermal Leishmaniasis: East Africa and the Indian Subcontinent

FEATURE	EAST AFRICA	INDIAN SUBCONTINENT
Most affected country	Sudan and South Sudan	Bangladesh
Incidence among patients with VL	~50%	5–15%
Interval between VL and PKDL	During VL to 6 months	6 months to 3 years
Age distribution	Mainly children	Any age
History of prior VL	Yes	Not necessarily
Rashes of PKDL in presence of active VL	Yes	No
Treatment	Sodium stibogluconate for 2–3 months	Miltefosine for 12 weeks
Natural course	Spontaneous cure in majority of patients	Spontaneous cure rarely

Abbreviations: PKDL, post–kala-azar dermal leishmaniasis; VL, visceral leishmaniasis.

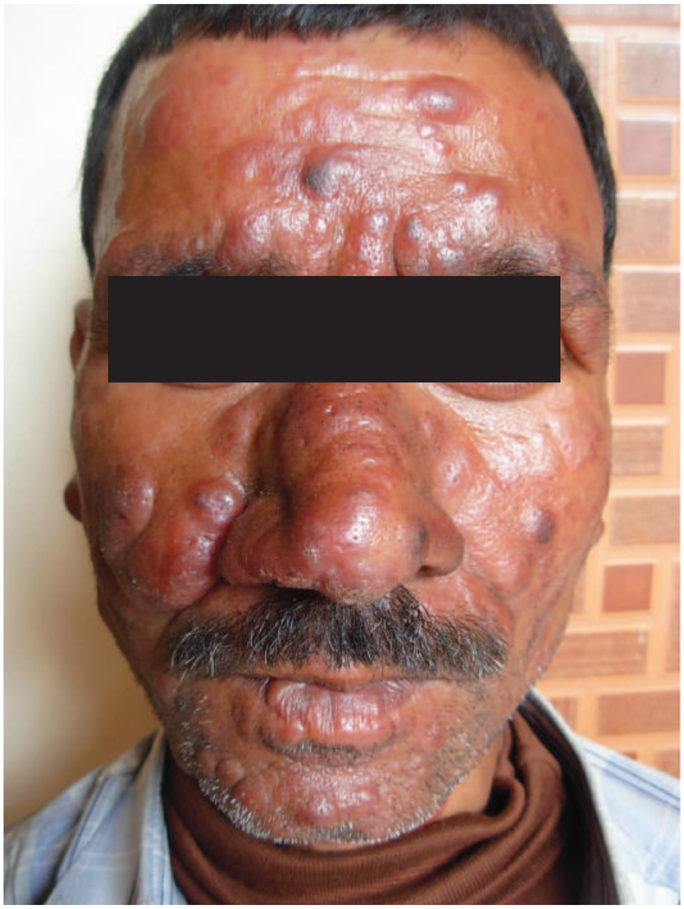
[redacted]

FIGURE 233-4 Post–kala-azar dermal leishmaniasis in an Indian patient. Note nodules of varying size involving the entire face. The face is erythematous, and the surface of some of the large nodules is discolored.

The alternative—three to four 20-day courses of AmB spread over several months—is expensive and unacceptable for most patients. Except for cosmetic reasons, these patients do not have any physical limitation, and thus motivation for such long and arduous treatment is very low. This leads to either no or incomplete treatment. In the Indian subcontinent, the currently recommended regimen is oral miltefosine for 12 weeks, in the usual daily doses. Though initially cure rates were high, a decline in its efficacy is now being reported in some studies. Furthermore, there are reports of ocular toxicity with this regimen in 3.7% of patients. The efficacy of either LAmB alone (20 mg/kg) or in combination with miltefosine (for 3 weeks) in PKDL has been recently tested in the Indian subcontinent; the efficacy was greater than 80% with both regimens. In East Africa, a majority of patients experience spontaneous healing. In those with persistent lesions, the response to 60 days of treatment with a pentavalent antimonial is good.

■ CUTANEOUS LEISHMANIASIS

CL can be broadly divided into Old World and New World forms. Old World CL caused by *Leishmania tropica* is anthroponotic and is confined to urban or suburban areas throughout its range. Zoonotic CL is most commonly due to *Leishmania major*, which naturally parasitizes several species of desert rodents that act as reservoirs over wide areas of the Middle East, Africa, and central Asia. Local outbreaks of human disease are common. Major outbreaks currently affect Afghanistan, Syria, Iraq, Lebanon, and Turkey in association with refugees and population movement. CL is increasingly seen in tourists and military personnel on mission in CL-endemic regions of countries and as a co-infection in HIV-infected patients. *Leishmania aethiopica* is restricted to the highlands of Ethiopia, Kenya, and Uganda, where it is a natural parasite of hyraxes. New World CL is mainly zoonotic and is most often caused by *Leishmania mexicana, Leishmania (Viannia) panamensis*, and *Leishmania amazonensis*. A wide range of forest animals act as reservoirs, and human infections with these species are predominantly rural. As a result of extensive urbanization and deforestation, *Leishmania (Viannia) braziliensis* has adapted to peridomestic and urban animals, and CL due to this organism is increasingly becoming an urban disease. In the United States, a few cases of CL have been acquired indigenously in Texas.

Immunopathogenesis As in VL, the proinflammatory (T_H1) response in CL may result in either asymptomatic or subclinical infection. However, in some individuals, the immune response causes ulcerative skin lesions, the majority of which heal spontaneously, leaving a scar. Healing is usually followed by immunity to reinfection with that species of parasite.

Clinical Features A few days or weeks after the bite of a sandfly, a papule develops and grows into a nodule that ulcerates over weeks or months. The base of the ulcer, which is usually painless, consists of necrotic tissue and crusted serum, but secondary bacterial infection sometime occurs. The margins of the ulcer are raised and indurated. Lesions may be single or multiple and vary in size from 0.5 to >3 cm **(Fig. 233-5)**. Lymphatic spread and lymph gland involvement may be palpable and may precede the appearance of the skin lesion. There may be satellite lesions, especially in *L. major* and *L. tropica* infections. The lesions usually heal spontaneously after 2–15 months. Lesions due to *L. major* and *L. mexicana* tend to heal rapidly, whereas those due to *L. tropica* and parasites of subspecies *Viannia* heal more slowly. In CL caused by *L. tropica*, new lesions—usually scaly, erythematous papules and nodules—develop in the center or periphery of a healed sore, a condition known as *leishmaniasis recidivans*. Lesions of *L. mexicana* and *Leishmania (Viannia) peruviana* closely resemble those seen in the Old World; however, lesions on the pinna of the ear are common, chronic, and destructive in the former infections. *L. mexicana* is responsible for chiclero's ulcer, the so-called self-healing sore of Mexico. CL lesions on exposed body parts (e.g., the face and hands), permanent scar formation, and social stigmatization may cause anxiety and depression and may affect the quality of life of CL patients.

Differential Diagnosis A typical history (an insect bite followed by the events leading to ulceration) in a resident of or a returning traveler from an endemic zone, strongly suggests CL. Cutaneous tuberculosis, fungal infections, leprosy, sarcoidosis, and malignant ulcers are sometime mistaken for CL.

Laboratory Diagnosis Demonstration of amastigotes in material obtained from a lesion remains the diagnostic gold standard. Microscopic examination of slit skin smears, aspirates, or biopsies of the lesion is used for detection of parasites. Culture of smear or biopsy material may yield *Leishmania*. PCR or LAMP is more sensitive than

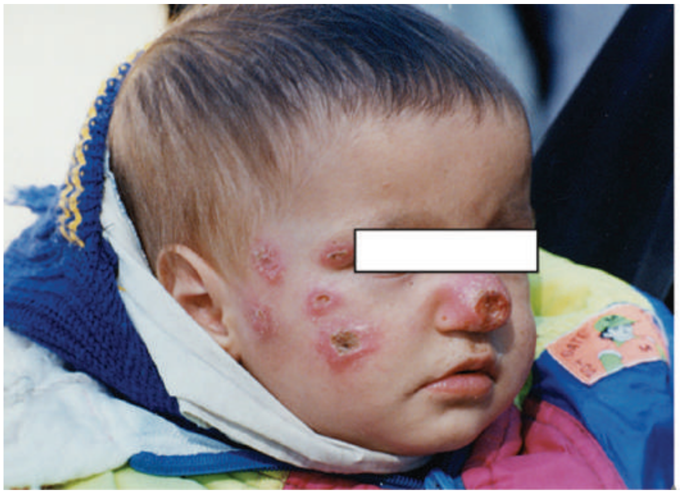
[redacted]

FIGURE 233-5 Cutaneous leishmaniasis in a Bolivian child. There are multiple ulcers resulting from several sandfly bites. The edges of the ulcers are raised. *(Courtesy of P. Desjeux, Retired Medical Officer, World Health Organization, Geneva, Switzerland.)*

microscopy and culture and allows identification of *Leishmania* to the species level. Quantitative PCR may be used to monitor treatment responses. This information is important in decisions about therapy because responses to treatment can vary with the species.

TREATMENT

Cutaneous Leishmaniasis

Although lesions heal spontaneously in the majority of cases, their spread or persistence indicates that treatment may be needed. One or a few small lesions due to "self-healing species" can be treated with topical agents. Systemic treatment is required for lesions over the face, hands, or joints; multiple lesions; large ulcers; lymphatic spread; New World CL with the potential for development of ML; and CL in HIV-co-infected patients.

A pentavalent antimonial is the first-line drug for all forms of CL and is used in a dose of 20 mg/kg for 20 days. The exceptions to this rule are CL caused by *Leishmania (Viannia) guyanensis*, for which pentamidine isethionate is the drug of choice (two injections of 4 mg of salt/kg separated by a 48-h interval), and CL due to *L. aethiopica*, which responds to paromomycin (16 mg/kg daily) but not to antimonials. Relapses usually respond to a second course of treatment. In Peru, topical imiquimod (5–7.5%) plus parenteral antimonials have been shown to cure CL more rapidly than antimonials alone. Azoles and triazoles have been used with mixed responses in both Old and New World CL but have not been adequately assessed for this indication in clinical trials. In *L. major* infection, oral fluconazole (200 mg/d for 6 weeks) resulted in a higher rate of cure than placebo (79% vs 34%) and also cured infection faster. Adverse effects include gastrointestinal symptoms and hepatotoxicity. Ketoconazole (600 mg/d for 28 days) is 76–90% effective in CL due to *L. (V.) panamensis* and *L. mexicana* in Panama and Guatemala.

Orally administered miltefosine makes it very attractive and is a major advancement in the treatment of CL. Miltefosine has been used in CL in doses of 2.5 mg/kg for 28 days. This agent is effective against *L. major* infections. Recent Pan American Health Organization guidelines recommend intralesional pentavalent antimonials in patients with localized CL caused by *L. braziliensis* and *L. amazonensis*. There is a strong recommendation for oral miltefosine treatment for New World CL where *L. panamensis, L. mexicana, L. guyanensis*, and *L. braziliensis* are the species involved. Overall, there has been a tilt toward the use of miltefosine in CL and ML instead of antimonials, both in the New and Old Worlds. In Brazil, miltefosine cured 71% of patients with *L. (V.) guyanensis* infection. Other drugs, such as dapsone, allopurinol, rifampin, azithromycin, and pentoxifylline, have been used either alone or in combinations, but most of the relevant studies have had design limitations that preclude meaningful conclusions.

Small lesions (≤3 cm in diameter) may conveniently be treated weekly until cure with an intralesional injection of a pentavalent antimonial at a dose adequate to blanch the lesion (0.2–2.0 mL). An ointment containing 15% paromomycin sulfate, either alone or with 0.5% gentamicin or 12% methylbenzonium chloride, cured 70–82% of lesions due to *L. major* in 20 days and may be suitable for lesions caused by other species. Heat therapy with an FDA-approved radiofrequency generator and cryotherapy with liquid nitrogen have also been used successfully.

Diffuse Cutaneous Leishmaniasis (DCL) DCL is a rare form of leishmaniasis caused by *L. amazonensis* and *L. mexicana* in South and Central America and by *L. aethiopica* in Ethiopia and Kenya. DCL is characterized by the lack of a cell-mediated immune response to the parasite, the uncontrolled multiplication of which thus continues unabated. The DTH response does not develop, and lymphocytes do not respond to leishmanial antigens in vitro. DCL patients have a polarized immune response with high levels of immunosuppressive cytokines, including IL-10, transforming growth factor (TGF) β, and IL-4, and low concentrations of IFN-γ. Profound immunosuppression leads to widespread cutaneous disease. Lesions may initially be confined to the face or a limb but spread over months or years to other areas of the skin. They may be symmetrically or asymmetrically distributed and include papules, nodules, plaques, and areas of diffuse infiltration. These lesions do not ulcerate. The overlying skin is usually erythematous in pale-skinned patients. The lesions are teeming with parasites, which are therefore easy to recover. DCL does not heal spontaneously and is difficult to treat. If relapse and drug resistance are to be prevented, treatment should be continued for some time after lesions have healed and parasites can no longer be isolated. In the New World, repeated 20-day courses of pentavalent antimonials are given, with an intervening drug-free period of 10 days. Miltefosine has been used for several months with a good initial response. Combinations should be tried. In Ethiopia, a combination of paromomycin (14 mg/kg per day) and sodium stibogluconate (10 mg/kg per day) is effective.

■ MUCOSAL LEISHMANIASIS

The subgenus *Viannia* is widespread from the Amazon basin to Paraguay and Costa Rica and is responsible for deep sores and for ML (Table 233-1). In *L. (V.) braziliensis* infections, cutaneous lesions may be simultaneously accompanied by mucosal spread of the disease or followed by spread years later. ML is typically caused by *L. (V.) braziliensis* and rarely by *L. amazonensis*, *L. (V.) guyanensis*, and *L. (V.) panamensis*. Young men with chronic lesions of CL are at particular risk. Overall, ~3% of infected persons develop ML. Not every patient with ML has a history of prior CL. ML is almost entirely confined to the Americas. In rare cases, ML may also be caused by Old World species like *L. major, L. infantum* (*L. chagasi*), or *L. donovani*.

Immunopathogenesis and Clinical Features The immune response is polarized toward a T_H1 response, with marked increases of IFN-γ and TNF-α and varying levels of T_H2 cytokines (IL-10 and TGF-β). Patients have a stronger DTH response with ML than with CL, and their peripheral-blood mononuclear cells respond strongly to leishmanial antigens. The parasite spreads via the lymphatics or the bloodstream to mucosal tissues of the upper respiratory tract. Intense inflammation leads to destruction, and severe disability ensues. Lesions in or around the nose or mouth (espundia; **Fig. 233-6**) are the typical presentation of ML. Patients usually provide a history of self-healed CL preceding ML by 1–5 years. Typically, ML presents as nasal stuffiness and bleeding followed by destruction of nasal cartilage, perforation of the nasal septum, and collapse of the nasal bridge. Subsequent involvement of the pharynx and larynx leads to difficulty in swallowing and phonation. The lips, cheeks, and soft palate may also be affected. Secondary bacterial infection is common, and aspiration pneumonia may be fatal. Despite the high degree of T_H1 immunity and the strong DTH response, ML does not heal spontaneously.

Laboratory Diagnosis Tissue biopsy is essential for identification of parasites, but the rate of detection is poor unless PCR techniques are used. The strongly positive DTH response fails to distinguish between past and present infection.

TREATMENT

Mucosal Leishmaniasis

The regimen of choice is a pentavalent antimonial agent administered at a dose of 20 mg of Sb^V/kg for 30 days. Patients with ML require long-term follow-up with repeated oropharyngeal and nasal examination. With failure of therapy or relapse, patients may receive another course of an antimonial but then become unresponsive, presumably because of resistance in the parasite. In this situation, AmB should be used. An AmB deoxycholate dose totaling 25–45 mg/kg is appropriate. There are no controlled trials of LAmB, but administration of 2–3 mg/kg for 20 days is considered adequate. Miltefosine (2.5 mg/kg for 28 days) cured 71% of ML patients in Bolivia. The more extensive the disease, the worse is the prognosis; thus, prompt, effective treatment and regular follow-up are essential.

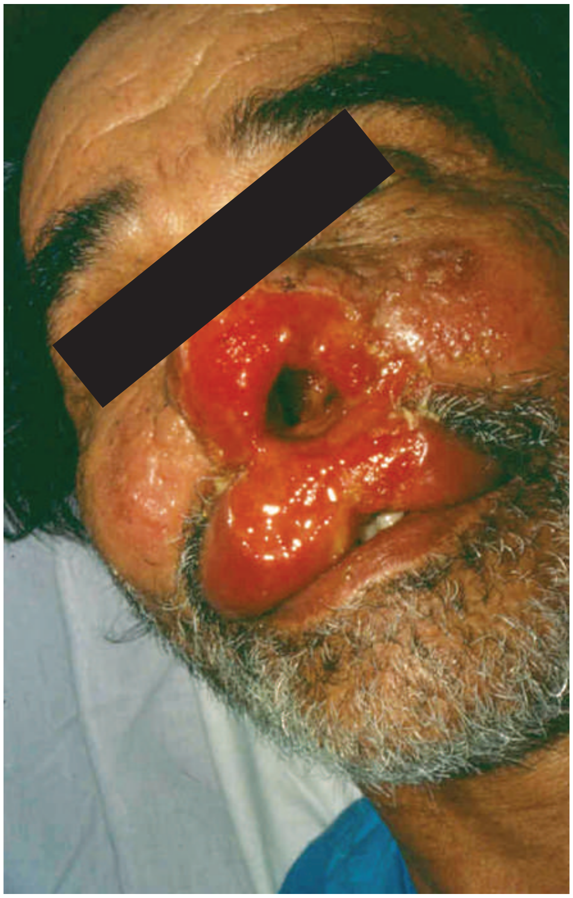
[redacted]

FIGURE 233-6 Mucosal leishmaniasis in a Brazilian patient. There is extensive inflammation around the nose and mouth, destruction of the nasal mucosa, ulceration of the upper lip and nose, and destruction of the nasal septum. *(Courtesy of R. Dietz, Universidade Federal do Espírito Santo, Vitória, Brazil.)*

PREVENTION OF LEISHMANIASIS

No vaccine is available for humans for any form of leishmaniasis, although several candidates are in early phases of development. Inoculation with live *L. major* ("leishmanization") is practiced in Iran; 80% of recipients were protected, according to one report. Anthroponotic leishmaniasis is controlled by case finding, treatment, and vector control with insecticide-impregnated bed nets and curtains and residual insecticide spraying. Control of zoonotic leishmaniasis is more difficult. Use of insecticide-impregnated collars for dogs, treatment of infected domestic dogs, and culling of street dogs are measures that have been used with uncertain efficacy to prevent transmission of *L. infantum*. In Brazil, canine vaccines have been found to promote a decrease in the human and canine incidence of zoonotic VL. Two vaccines, Leishmune and Leish-Tec, are licensed in Brazil; Leishmune provides significant protection to vaccinated dogs. CaniLeish and LetiFend are the two licensed canine vaccines approved for use in Europe. Personal prophylaxis with bed nets and repellants may reduce the risk of CL infections in the New World.

FURTHER READING

Dorlo TP et al: Optimal dosing of miltefosine in children and adults with visceral leishmaniasis. Antimicrob Agents Chemother 56:3864, 2012.

Mann S et al: A review of leishmaniasis: Current knowledge and future directions. Curr Trop Med Rep 8:121, 2021.

Pan American Health Organization: Synthesis of evidence and recommendations: Guideline for the treatment of leishmaniasis in the region of the Americas 47:e43, 2023.

Singh OP, Sundar S: Visceral leishmaniasis elimination in India: Progress and the road ahead. Expert Rev Anti Infect Ther 20:1381, 2022.

World Health Organization: WHO guideline for the treatment of visceral leishmaniasis in HIV co-infected patients in East Africa and South-East Asia. Geneva: World Health Organization, 2022. Available at *https://iris.who.int/bitstream/handle/10665/354703/9789240048294-eng.pdf?sequence=1*. Accessed September 15, 2023.

234 Chagas Disease and African Trypanosomiasis

François Chappuis, Yves Jackson

Myriads of protozoan parasites of the genus *Trypanosoma* infect plants and animals worldwide. Among these, three are of clinical significance for humans: *T. cruzi* causes Chagas disease, and *T. brucei gambiense* and *T. brucei rhodesiense* cause human African trypanosomiasis (HAT), which is also known as "sleeping sickness." Despite obvious differences in their geographic distribution, parasitic life cycle, clinical presentation, treatment, and outcome, these vector-borne diseases are archetypal examples of neglected tropical diseases. More broadly, these infectious diseases affect neglected populations of the lowest socioeconomic class who have limited access to care and who live either in remote rural areas of low- or middle-income tropical/subtropical countries or in urban areas of both endemic and nonendemic countries. Most drugs to treat these conditions are several decades old, their availability is limited, and their efficacy and/or safety is suboptimal.

Other trypanosome species (e.g., *T. congolense* and *T. evansi*) predominantly cause nonhuman zoonoses and only occasionally cause illness in humans.

CHAGAS DISEASE (AMERICAN TRYPANOSOMIASIS)

DEFINITION

First described in 1909 by Carlos Chagas, Chagas disease (American trypanosomiasis) is a zoonosis caused by the flagellated protozoan *T. cruzi*. After a frequently asymptomatic acute phase, 30–40% of patients develop potentially life-threatening chronic cardiomyopathy and/or digestive tract dysfunction over decades. Acute reactivation may occur in immunocompromised patients. Chagas disease imposes an important human and social burden in Latin America and has recently spread outside its natural boundaries to become a global public health problem. The vast majority of affected individuals are unaware of being infected and do not have access to appropriate clinical management and counseling.

TRANSMISSION

Vectorial Transmission *T. cruzi* infection is primarily a zoonosis transmitted to a range of wild and domestic mammals by blood-sucking triatomine bugs. Sylvatic, peridomiciliary, and intradomiciliary vectorial cycles sometimes overlap. Over a large geographic area in the Americas (from northern Argentina to the southern United States), most human infections are intradomiciliary, arising from a triatomine bite during nighttime sleep. Feces released by triatomines during a blood meal contain the infective metacyclic form of *T. cruzi* that enters the human body through cutaneous breaks, mucosae, or conjunctivae. Despite some laboratory research showing the potential for transmission by bedbugs, there is no evidence that bedbugs actually transmit *T. cruzi* to humans.

Nonvectorial Transmission Other modes of transmission can cause infection in both endemic and nonendemic regions. *T. cruzi* can be transmitted congenitally from mother to newborn, by transfusion of blood products, by tissue or organ transplantation, or by ingestion of contaminated food or drink. Congenital infection occurs in 1–10% of newborns of infected mothers. The risk of infection from contaminated blood products is low (1.7% overall, 13% for platelet recipients, and close to 0 for recipients of red blood cells and plasma). Transmission by infected organ and tissue transplants mostly affects heart, liver, and kidney recipients. Oral transmission is increasingly reported after ingestion of contaminated food (berries) or drinks (fruit or sugar cane juice) and occasionally causes outbreaks.

EPIDEMIOLOGY

An estimated 6–7 million people are infected by *T. cruzi*, including >1 million individuals with chronic cardiomyopathy. However, the true global burden of Chagas disease is in fact uncertain. The highest numbers of infected individuals reside in Argentina, Brazil, and Mexico; the prevalence is highest in Bolivia (6.1%), Argentina (3.6%), and Paraguay (2.1%). In highly endemic regions of these countries, the prevalence may exceed 40%. Formerly restricted to poor rural populations, the distribution of cases—and, to some extent, *T. cruzi* transmission—has progressively extended to cities in the context of rapid urbanization and rural migration. A recent history of migration from a rural area is the main risk factor in urban settings.

Overall, the prevalence and incidence of Chagas disease have sharply declined in recent decades because of improved housing and socioeconomic conditions as well as public health interventions, including regional vector-control initiatives, implementation of systematic screening of blood products, and improved detection of congenital transmission. Several countries have been declared free of domiciliary transmission as a result of sustained residual insecticide-spraying campaigns. This progress is threatened by adaptation of the vector to the periurban environment, its resurgence in areas where spraying has been discontinued, the development of resistance to pyrethroid insecticides, and the persistence of peridomiciliary transmission. A growing number of localized outbreaks are being reported in previously stable areas, with the Amazon basin particularly at risk.

Chagas disease distribution has expanded to nonendemic countries in the context of increased global travel, with cases reported more frequently in North America, Western Europe, Australia, and Japan. The United States harbors up to 300,000 cases, mostly among immigrants from Central America. In addition, sporadic vector-borne infections occur in the southern states. Western Europe has an estimated 68,000–123,000 cases, and Japan and Australia a few thousand cases. Despite the implementation of blood bank screening and of some dedicated medical programs, only a small proportion of cases have been identified and properly managed to date. A low level of awareness among health care professionals and difficulties experienced by some groups in accessing care appear to be major drivers. At-risk migrant communities are frequently subject to factors that render them socially, legally, or economically vulnerable. Moreover, the cultural perception of Chagas as a disease embedded in poverty can create a social stigma that complicates its management at the community level. In contrast to immigrants, international tourists visiting endemic countries are at very low risk of being infected, whether by reduviid bug bites or by other routes, and reports of Chagas disease in travelers are rare.

PATHOLOGY

Several *T. cruzi* strains have been identified. These strains have partially overlapping transmission cycles and geographic distributions, but no definitive evidence supports an association of certain strains with specific clinical manifestations or with variation in disease severity. The rarity of digestive tract involvement north of the Amazon basin suggests that specific parasitic and host genetic factors may influence the disease course. The pathogenesis of Chagas disease results from the complex interactions between the pathogen and the host immune response. Many questions about the relative importance of these interactions, including the role of autoimmune mechanisms, remain unanswered. After local penetration of trypomastigotes, parasites rapidly enter the bloodstream and disseminate through the body, infecting a wide range of nucleated cells in which they differentiate into amastigotes **(Fig. 234-1)**. The innate immune response triggered by parasite mucins and DNA leads to a predominantly T helper 1 response. The production of various proinflammatory cytokines and the activation of CD8+ T lymphocytes reduce parasitemia to a subpatent level within 4–8 weeks, a point marking the end of the acute phase.

Immune evasion mechanisms allow persistent low-intensity proliferation of amastigotes and their release into the bloodstream, with subsequent infection of potentially all types of nucleated cells—notably cardiac, skeletal, and smooth-muscle cells. Mechanisms that have been postulated to determine the pathogenic evolution toward cardiomyopathy include the parasites' persistence and the host's inability to downregulate the initial immune response, resulting in cell-mediated damage and an imbalance of T helper 1 and 2 responses with excessive production of proinflammatory cytokines. Secondary mechanisms, such as microcirculation abnormalities and dysautonomia, also may influence the progression of tissue damage. Genome-wide association studies suggest that genetic variation may contribute to cardiomyopathy development.

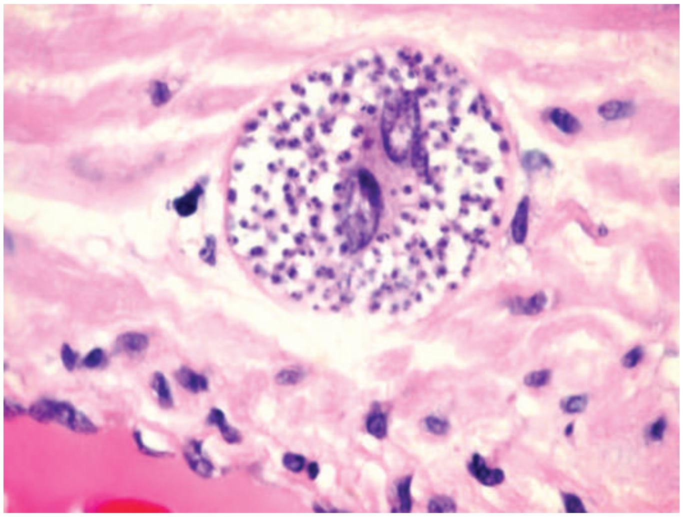

FIGURE 234-1 A cluster of *Trypanosoma cruzi* amastigotes with an inflammatory infiltrate in the placenta of a congenitally infected newborn infant.

In the myocardium, chronic inflammation results in cellular destruction and the development of fibrosis leading to a segmental loss of contractility and dilation of the chambers, with the associated risk of left ventricle apical aneurism. Focal hypoperfusion and tissue damage are sources of ventricular arrhythmias, while scarring lesions mostly affect the conduction system. Autonomic cell destruction leads to vagal and sympathetic denervation whose exact clinical significance remains to be clarified.

T. cruzi appears to have a direct toxic effect on digestive tract intramural autonomic ganglion cells. Over time, the loss of neural cells affects muscular tone, leading to motility disorders and ultimately to organ dilation (megaviscera syndrome). The esophagus and colon are primarily affected, but lesions may occur along the whole digestive tract. Inadequate relaxation of the lower esophageal sphincter causes symptoms of achalasia, whereas damage to the colon ultimately mimics Hirschsprung's disease, with severe constipation and the risk of volvulus and toxic dilation.

Factors reducing the cellular immune response, such as HIV infection, posttransplantation immunosuppressive therapies, or hematologic malignancies, may increase intracellular replication of amastigotes, with increased parasitemia (reactivation). Lesions develop predominantly in the central nervous system (CNS), the heart, and the skin. Among HIV-positive patients, the risk of reactivation is ~20% in the absence of antiretroviral therapies and occurs when the CD4+ T cell count falls to <100/μL. Clinically manifest *T. cruzi* reactivation is an AIDS-defining opportunistic infection.

CLINICAL MANIFESTATIONS

The clinical manifestations of *T. cruzi* infection vary greatly among individuals. The infection course is divided into two phases that are associated with different clinical features, duration, and prognosis **(Table 234-1)**. The acute phase remains undetected and undiagnosed in most individuals. While 5–10% of these early infections spontaneously resolve without treatment, *T. cruzi* persists for life in most individuals (the chronic phase); 60–70% of these individuals never develop apparent tissue damage (the indeterminate form), but the remaining 30–40% progress toward detectable organ damages of variable severity over decades (the determinate form). These chronic complications include cardiac (20–30%), digestive (5–20%), or mixed (5–10%) disorders. There

TABLE 234-1 Characteristics of the Stages of *Trypanosoma cruzi* Infection

PHASE OR SETTING	CONTEXT	ONSET OF FIRST SYMPTOMS	CLINICAL MANIFESTATIONS	DURATION	PROGNOSIS
Acute (congenital)	~5% risk of maternal transmission to newborn	At birth or weeks after delivery	>90% asymptomatic; rare lymphadenopathy, hepatosplenomegaly, jaundice, respiratory distress, growth retardation	2–8 weeks	Favorable when infant is born alive; unknown rate of in utero or neonatal death
Acute	Vector-borne transmission; oral transmission (ingestion of contaminated food/drinks); blood product transfusion; tissue/organ transplantation	1–2 weeks after vectorial transmission; may be sooner (days) after oral transmission or later (months) after transfusion/transplantation	>90% asymptomatic or mild febrile illness; local swelling at inoculation site (eyelid [Romaña sign] or skin [chagoma]); polyadenopathy; splenomegaly; myocarditis, hepatitis, and encephalitis more frequent with oral transmission	4–8 weeks	Mortality: 0.1–5% with oral transmission or myocarditis/encephalitis
Chronic (indeterminate form)	Balanced immune response after acute phase subsides	No symptoms	Normal clinical examination and ECG result	Lifelong or until determinate phase	No attributable mortality
Chronic (determinate form)	Predominant inflammatory response (in cardiomyopathy only)	Years to decades after initial infection	Dyspnea, chest pain, palpitation, syncope, sudden death, stroke, dysphagia, regurgitation, constipation, fecaloma, volvulus, peripheral neuropathy	Chronic	5-year mortality: 2–63%, depending on extent of cardiac damage; most important causes of death: cardiac failure and sudden death, followed by stroke
Acute (reactivation)	Severe immunosuppression	Variable	Myocarditis, erythema nodosum, panniculitis, *Toxoplasma*-like focal brain lesion, meningoencephalitis	Variable	Mortality depends on rapidity of diagnosis and treatment and on underlying conditions

Abbreviation: ECG, electrocardiography.

is no predictor of evolution toward clinical manifestations during the chronic phase. In patients with cardiomyopathy, bundle branch blocks are usually the first signs and may cause no symptoms for years until more severe conduction system disease, arrhythmias, and left ventricular dysfunction occur. Advanced cardiac damage entails a worse prognosis than other cardiomyopathies—notably, ischemic heart disease.

APPROACH TO THE PATIENT

Chagas Disease (American Trypanosomiasis)

More than 90% of infections go undiagnosed, and cases are frequently identified at a late stage once chronic complications develop. The vast majority of *T. cruzi*–infected individuals are asymptomatic (i.e., in the indeterminate form of the chronic phase). An awareness of potential Chagas disease is important for general practitioners as well as for physicians from various specialties, including gastroenterologists, cardiologists, neurologists, obstetricians, pediatricians, and infectious disease specialists. Outside endemic areas, screening for Chagas disease should be proposed when any Latin American individual has evocative symptoms and signs, including abnormalities on electrocardiography (ECG) or increased risk of (1) *T. cruzi* infection (Chagas disease in the mother or other family members; origins in a highly endemic country or area; history of unscreened blood transfusion in Latin America); (2) transmission to others (e.g., via pregnancy or blood or organ donation); or (3) reactivation (current or pending immunosuppression). Screening of the relatives of an index case may identify additional cases.

DIAGNOSIS AND STAGING

Diagnostic Confirmation Diagnostic strategies depend on the clinical phase (Table 234-2). Detection of circulating parasites by microscopy of the blood after concentration (e.g., by the Strout method, microhematocrit) or by nucleic acid–based assay (polymerase chain reaction [PCR]) is the best diagnostic approach when the parasitemia level is high—i.e., during the acute phases, including reactivation. Once parasitemia becomes undetectable by microscopy (a point marking the end of the acute phase), diagnosis relies on immunologic tests that detect anti–*T. cruzi* IgG. The most common techniques include a conventional or recombinant enzyme-linked immunosorbent assay (ELISA) and immunofluorescence assays. Two positive serologic tests using different techniques and targeting different antigens confirm the diagnosis of Chagas disease during the chronic phase. In the presence of discordant serologic results, a third serologic test is warranted. Some of the immunochromatographic rapid diagnostic tests on the market have sufficient sensitivity and specificity to be used as first-line screening tests where laboratory facilities are not easily accessible. If the rapid diagnostic test result is positive, at least one conventional serologic assay is necessary to confirm infection.

Diagnosis of congenital infection relies on examination of cord and/or peripheral blood by microscopy or PCR during the first days or weeks of life. A test conducted after 4 weeks of age is most accurate: PCR earlier in life may be falsely positive, likely because of the passage of *T. cruzi* DNA fragments from the mother to the child. If results are negative, serologic tests should be performed at 9 months of age, once maternal antibodies have been cleared.

During the chronic phase, the limited sensitivity (50–80%) of conventional PCR restricts its usefulness for primary diagnosis; however, PCR can document therapeutic failure if it yields positive results after the completion of treatment. In the United States, the Centers for Disease Control and Prevention (CDC) provides reference laboratory testing (see contact information in the treatment section).

TABLE 234-2 Diagnostic Procedures of Choice for Clinical Stages of *T. cruzi* Infection

STAGE	TECHNIQUE OF CHOICE	SAMPLE	DIAGNOSTIC CRITERIA
Acute	Microscopy after concentration, PCR	Peripheral blood, cerebrospinal or other body fluids	Positivity in one test
Acute (early congenital during first 9 months of life)	Microscopy after concentration, PCR	Cord or peripheral blood	Positivity in one test
Chronic (indeterminate and determinate forms)	Serology	Peripheral blood	Positivity in two tests with different techniques and antigens
Reactivation	Microscopy after concentration, PCR	Peripheral blood, cerebrospinal or other body fluids	Positivity with evidence of increasing parasitemia on serial samples or extremely high parasite load

Abbreviation: PCR, polymerase chain reaction.

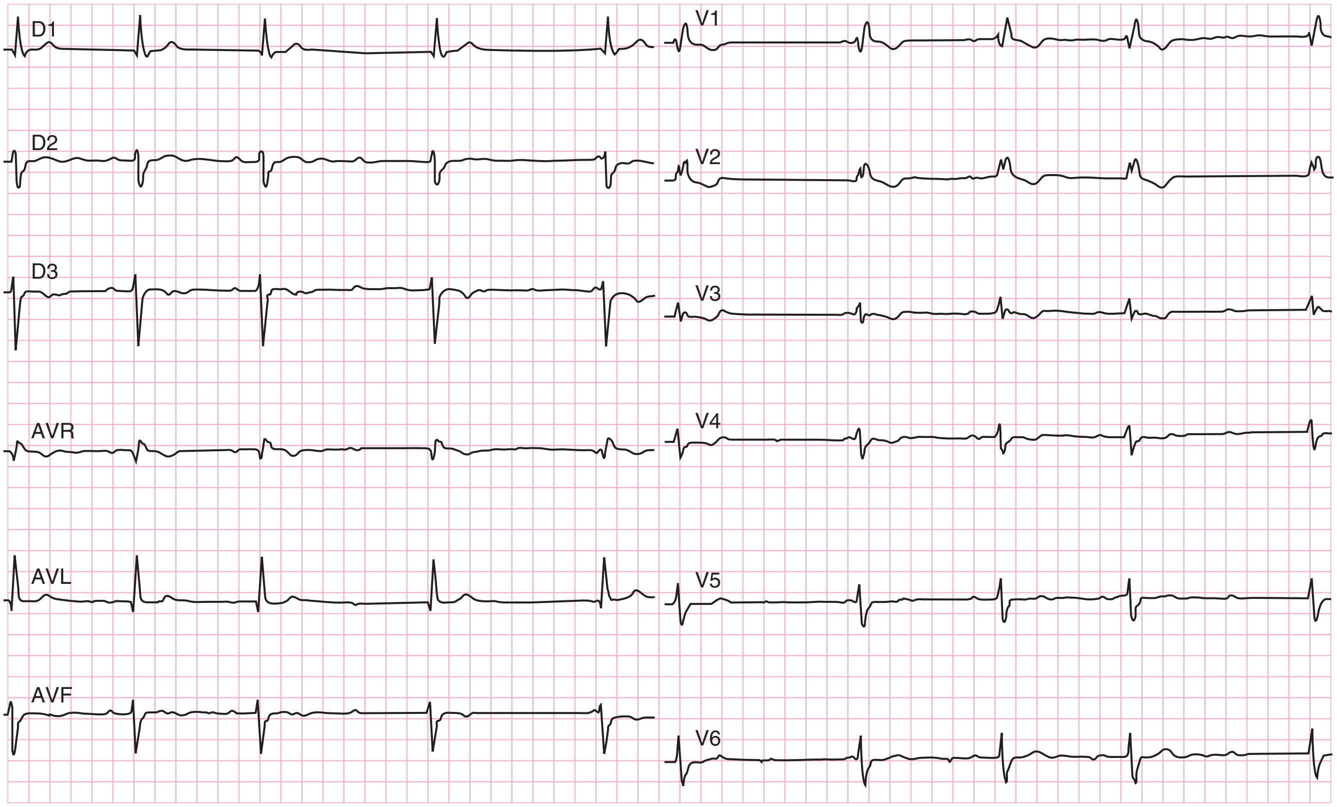

FIGURE 234-2 Electrocardiogram of a 43-year-old patient shows bradycardia with high-grade atrioventricular blocks.

Disease Staging Once *T. cruzi* infection is confirmed, clinicians should assess the presence of complications and concomitant factors that may influence the course of the disease. The initial evaluation includes a thorough cardiac, neurologic, and digestive history and a clinical examination. Twelve-lead ECG with a 30-s strip is a good screening test for Chagas-associated cardiomyopathy. The most frequently found abnormalities are right bundle branch block, left anterior fascicular block, ventricular premature beats, repolarization disorders, Q waves, and low QRS voltage **(Fig. 234-2)**. An abnormal ECG result or the presence of suggestive cardiac symptoms warrants further investigation. Echocardiography and the 24-h Holter test are the preferred methods for assessment of chamber dilation, apical aneurysm, ventricular dysfunction, and arrhythmias. Depending on the findings, the workup can be supplemented by MRI or electrophysiologic studies. Gastroenterologic investigations are performed in patients with suggestive symptoms, such as dysphagia and severe constipation. Barium esophagography and enema are first-line diagnostic procedures, which can be supplemented by esophageal manometry. Megacolon is diagnosed when the sigmoid or descending colon diameter is ≥6.5 cm.

Comorbidities, including other cardiovascular risk factors, immunosuppressive conditions, and other chronic infections (e.g., with *Strongyloides stercoralis* or HIV) should be investigated.

TREATMENT

Chagas Disease (American Trypanosomiasis)

ETIOLOGIC TREATMENT

Only two drugs, benznidazole and nifurtimox **(Table 234-3)**, have shown persistent efficacy against *T. cruzi* infection when administered for ≥30 days. While these drugs have been used since the early 1970s, many questions remain about their mode of action and efficacy at the different stages of infection. The treatment goal depends on the clinical stage; the overall objectives are to cure patients who have recent infection or reactivation, to reduce morbidity, and to prevent transmission at later stages. Treatment is most effective during the acute (including congenital) phase and the early chronic phase (i.e., in patients <18 years of age), with a 60–100% cure rate. The efficacy of treatment during the indeterminate form of the chronic phase in patients >18 years old is not known; however, treatment may protect against the development of cardiac damage later in life and eliminate the risk of vertical transmission when given before conception. In adults with chronic cardiomyopathy, benznidazole has no impact on disease progression and mortality risk. Neither benznidazole nor nifurtimox is effective against

TABLE 234-3 Chagas Treatment Regimens and Adverse Reactions to Benznidazole and Nifurtimox

DRUG	REGIMEN	DURATION	ADVERSE EVENTS IN ADULTS (FREQUENCY)	PREMATURE DISCONTINUATION (RATE)
Benznidazole	Age <12 years: 5–7.5 mg/kg per day in 2 doses Age >12 years: 5 mg/kg per day in 2 doses	30–60 days	Allergic dermatitis (29–50%), anorexia and weight loss (5–40%), paresthesia (0–30%), peripheral neuropathy (0–30%), nausea and vomiting (0–5%), leukopenia and thrombocytopenia (<1%)	7–20%
Nifurtimox	Age <10 years: 15–20 mg/kg per day in 3 or 4 doses Age 11–16 years: 12.5–15 mg/kg per day in 3 or 4 doses Age >16 years: 8–10 mg/kg per day in 3 or 4 doses	60–90 days	Anorexia and weight loss (50–81%), nausea and vomiting (15–50%), abdominal discomfort (12–40%), headaches (13–70%), dizziness and vertigo (12–33%), anxiety and depression (10–49%), insomnia (10–54%), myalgia (13–30%), peripheral neuropathy (2–5%), memory loss (6–14%), leukopenia (<1%)	6–44%

Source: From C Bern: Chagas' Disease. N Engl J Med 373:456, 2015. Copyright © 2015 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

digestive complications. Treatment is contraindicated during pregnancy and in advanced renal or hepatic failure. Preferred regimens and drug tolerance vary with age. Adverse events are more frequent among adults, who are therefore at increased risk of premature treatment discontinuation (Table 234-3). As benznidazole seems better tolerated than nifurtimox in adults, it is the recommended first-line drug in this age range. Close (e.g., weekly) clinical and biologic monitoring is necessary during treatment. While treatment is usually prescribed for 60 days, the optimal duration remains a matter of debate, with a growing interest in shorter courses.

Treatment should be undertaken for all children, women of child-bearing age, patients in the acute phase, and patients with reactivation. Given the uncertainties about the impact of treatment, the decision to treat patients >18 years old who have the indeterminate form of the chronic phase should be made on an individual basis after discussing the pros and cons with the patient. A negative pregnancy test is mandatory before initiating treatment, as the recommended drugs have not been proven to be safe in pregnancy. The efficacy of second-line treatment (e.g., nifurtimox after failure with benznidazole) has not been evaluated to date.

The limited efficacy of current regimens and the understanding that living parasites are a driver of immunopathologic processes have fueled interest in novel therapeutic approaches. These include the addition of immunomodulatory interventions to antiparasitic treatment and the use of combinations of antiparasitic drugs. Information on drugs can be obtained through the CDC (Parasitic Diseases Public Inquiries line [404-718-4745] or chagas@cdc.gov), or the CDC Emergency Operations Center (770-488-7100). The U.S. Food and Drug Administration has approved the use of benznidazole for treatment of children 2–12 years and nifurtimox for those 0–17 years of age, and in older patients based on clinical decision.

NONETIOLOGIC TREATMENT

The management of Chagas cardiomyopathy generally follows the management guidelines for heart failure, conduction disturbances, or ventricular arrhythmia of other etiologies. Given the high risk of sudden death, early initiation of treatment with amiodarone or implantation of a cardioverter defibrillator should be considered in the presence of pathologic electrophysiologic abnormalities. Anticoagulation is recommended for primary and secondary prevention of cardioembolic events in the presence of an intramural thrombus or apical aneurysm. Strict control of other cardiovascular risk factors is warranted. Chagas cardiomyopathy is a prominent indication for heart transplantation in Latin America; some evidence indicates that the results are better than in cardiomyopathy of other etiologies. Posttransplantation immunosuppression requires close monitoring, given the high risk of reactivation.

Treatment of digestive dysmotility includes dietary counseling and meals rich in fiber and hydration, with smaller portions eaten more frequently. Drugs releasing the lower esophageal sphincter (e.g., nifedipine or isosorbide dinitrate before meals), pneumatic balloon dilatation, or laparoscopic myotomy improve upper gastrointestinal symptoms in the early stage. Use of botulinum toxin is effective but requires repeated injections. Laxatives and enemas alleviate chronic constipation in most patients. Surgery is indicated in patients with distressing symptoms that are refractory to medical treatment.

CLINICAL FOLLOW-UP

Defining the optimal cure after treatment remains very challenging and is a crucial topic of research. While the search for biomarkers (including through proteomics) to identify early indicators of treatment response holds some promise, serologic follow-up remains the cornerstone of posttreatment monitoring in the acute phase. In the chronic phase, there is no assay of proven value for documentation of response. The time needed for negative seroconversion after treatment indeed depends on the duration of infection. The interval is short (usually months, sometimes up to 2 years) when infection is treated during the acute (including congenital) phase. In contrast, decades are required in adults infected during childhood. A positive result in a posttreatment PCR indicates treatment failure, but a negative result cannot be interpreted because of the low sensitivity of PCR during the chronic phase. The status of patients with negative PCR results but persistent positive serology is therefore uncertain, but these patients should be considered potentially infective as long as serologic tests continue to yield positive results. All patients, treated or not, should be regularly monitored. The basic yearly assessment includes history-taking for detection of new symptoms, clinical examination, and 12-lead ECG.

PREVENTION

In the absence of a vaccine, preventive measures—primary (prevention of *T. cruzi* transmission), secondary (avoidance of complications), and tertiary (reduction of morbidity and mortality)—are necessary. Screening of blood donations is being progressively implemented in endemic areas and in countries to which high-risk groups are immigrating, and screening should be extended to organ donation. When sustained over prolonged periods, vector control is an effective and cost-effective strategy to curb intradomiciliary transmission. Insecticide-impregnated bed nets (as used for malaria) provide individual protection against reduviid bug bites. Screening of child-bearing-age and pregnant Latin American migrant women has been highly cost-effective in Spain, although the cost per case detected varies with the prevalence of infection in the targeted population. Early identification of cases through passive and active screening of the population at risk, along with provision of treatment, may reduce the risk of complications and secondary transmission, particularly congenital transmission. Finally, identification and treatment of cardiac complications and prevention of cardioembolic events at an early stage positively influence the disease course.

GLOBAL CONSIDERATIONS

With its geographic expansion, Chagas disease has become a global health issue, predominantly affecting vulnerable people on four continents. Yet, as with other neglected tropical diseases, progress against Chagas is limited by a lack of research and development and a lack of financial and political commitment. For example, the production and registration of existing drugs, and access to them, are still problematic in many countries, including the United States. Difficulties in research on and development of new drugs are compounded by the lack of financial incentives. The future of Chagas disease is likely to be influenced by global phenomena. Climatic changes, population aging, increasing prevalence of noncommunicable comorbidities (e.g., diabetes, hypertension) in low- and middle-income countries, and increasing use of immunosuppressive drugs are likely to impact the epidemiology, clinical course, and burden of Chagas disease. To tackle these challenges, clinical, public health, and policy interventions need to be scaled up and improved in areas of high or hidden prevalence (e.g., in the Chaco Region of Argentina, Bolivia, and Paraguay and in Mexico, Western Europe, and the United States, respectively).

HUMAN AFRICAN TRYPANOSOMIASIS (SLEEPING SICKNESS)

DEFINITION

HAT is a life-threatening illness caused by infection with extracellular protozoan parasites that are transmitted by tsetse flies in sub-Saharan Africa. *T. b. gambiense* and *T. b. rhodesiense* are the two pathogenic subspecies affecting humans; their epidemiologic and clinical features largely differ.

EPIDEMIOLOGY

The geographic range of HAT is restricted to sub-Saharan Africa in line with the distribution of its vector, the tsetse fly (*Glossina* species; **Fig. 234-3**). HAT due to *T. b. gambiense* is endemic in 24 countries of western and central Africa. Between 1999 and 2020, the number of reported cases fell by 98% (from 27,862 to 565) as a result of successful control measures based on systematic screening of populations at

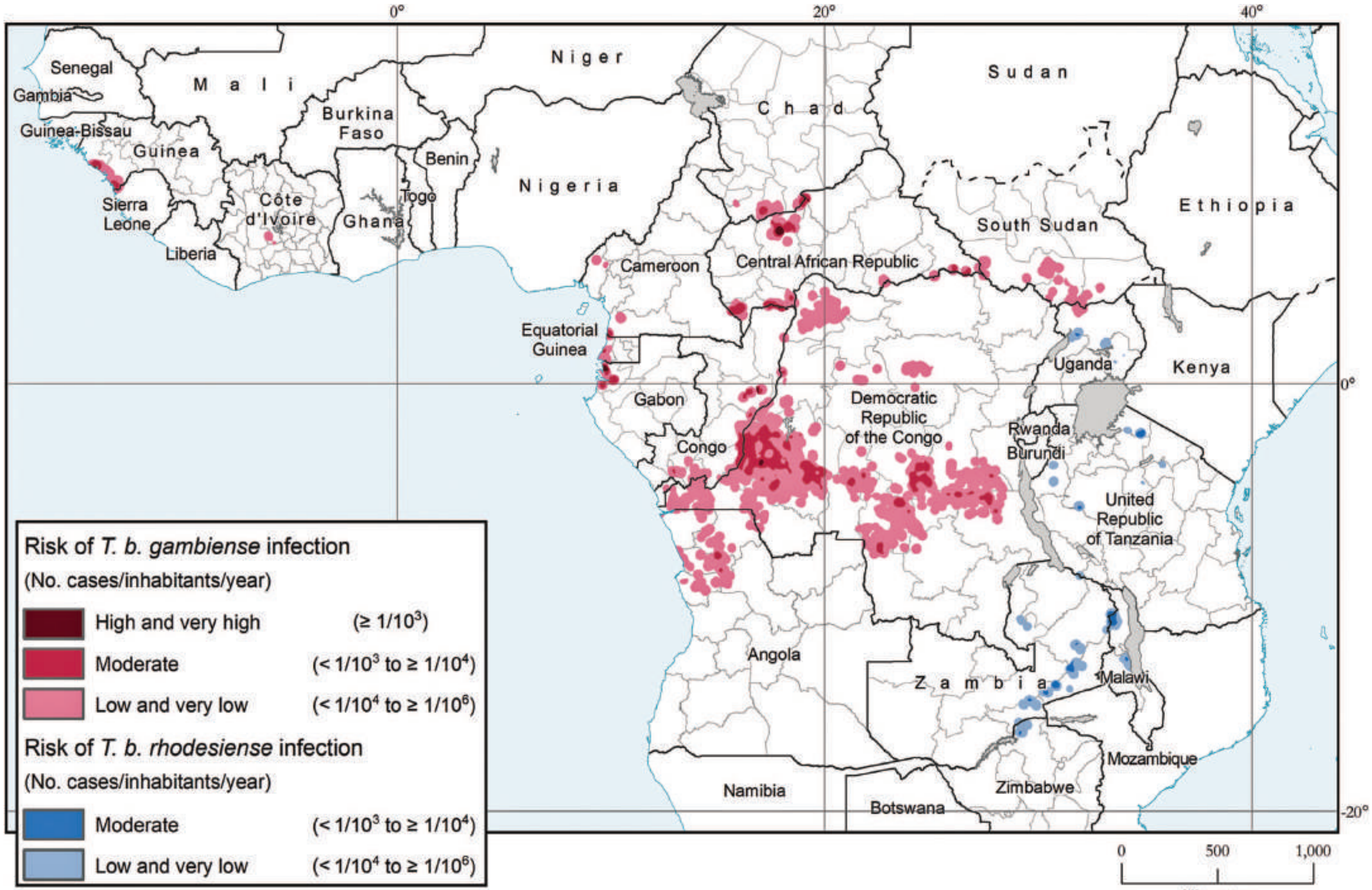
0°
20°
40°
0°
-20°
Senegal
Gambia
Guinea-Bissau
Guinea
Sierra Leone
Liberia
Côte d'Ivoire
Ghana
Togo
Benin
Burkina Faso
Mali
Niger
Nigeria
Chad
Sudan
South Sudan
Ethiopia
Cameroon
Central African Republic
Equatorial Guinea
Gabon
Congo
Democratic Republic of the Congo
Uganda
Kenya
Rwanda
Burundi
United Republic of Tanzania
Angola
Zambia
Malawi
Mozambique
Zimbabwe
Namibia
Botswana
Risk of T. b. gambiense infection
(No. cases/inhabitants/year)
High and very high (≥ 1/10³)
Moderate (< 1/10³ to ≥ 1/10⁴)
Low and very low (< 1/10⁴ to ≥ 1/10⁶)
Risk of T. b. rhodesiense infection
(No. cases/inhabitants/year)
Moderate (< 1/10³ to ≥ 1/10⁴)
Low and very low (< 1/10⁴ to ≥ 1/10⁶)
0
500
1,000
kilometres

FIGURE 234-3 Areas at risk for human African trypanosomiasis, 2016–2020. *(Reproduced from Franco JR et al: The elimination of human African trypanosomiasis: Achievements in relation to WHO road map targets for 2020. PLoS Negl Trop Dis 2022; 16(1): e0010047, Figure 4.)*

risk, diagnostic confirmation, and treatment of infected individuals. During the same period, the number of reported cases of HAT due to *T. b. rhodesiense* fell by 84% (from 619 to 98) in the 10 disease-endemic countries of eastern and southeastern Africa. However, the ratio of reported to unreported cases remains uncertain for disease caused by both species. In 2020, most cases of *T. b. gambiense* HAT were reported by the Democratic Republic of the Congo (DRC; 70%), whereas Malawi reported most of the cases caused by *T. b. rhodesiense* (91%). The geographic distributions of *T. b. gambiense* and *T. b. rhodesiense* do not overlap, but the two species are present in distinct regions of Uganda. A roadmap for HAT elimination as a public health problem by 2020 was mapped out by the World Health Organization (WHO) with two primary indicators: the number of cases reported annually (target: <2000; reached since 2018) and the area at risk reporting ≥1 case/10,000 people/year (target: reduction of 90% by 2016–2020 compared with the 2000–2004 baseline; 83% reduction reached in 2020). The next goal set by WHO is the global elimination of transmission by 2030.

Humans are the predominant reservoir of *T. b. gambiense*. Rare cases of vertical (in utero) or transfusional transmission have been reported, but almost all patients are infected by the bite of tsetse flies during their daily activities along or near rivers, where the flies live and reproduce. In contrast, *T. b. rhodesiense* causes zoonosis in a variety of wild and domesticated animals (e.g., antelopes and cattle, respectively), which act as reservoirs. Humans are infected by *T. b. rhodesiense* via tsetse bites in woodland savannah. Honey gatherers, game park rangers, poachers, and firewood collectors are particularly at risk. Imported cases of HAT are occasionally diagnosed among African immigrants and other travelers. While long-term travelers (>30 days) are at increased risk of *T. b. gambiense* HAT, most imported cases of *T. b. rhodesiense* HAT are seen in short-term travelers, typically following visits to game parks.

■ PATHOLOGY AND PATHOGENESIS

T. b. rhodesiense and *T. b. gambiense*, unlike other trypanosome species, can infect humans because they resist lytic factors in human serum—namely, apolipoprotein L-1 (*APOL1*). Human *APOL1* variants are prevalent in individuals of African ancestry, conferring protection against livestock trypanosome species, but at the cost of increasing the likelihood of chronic kidney disease. The serum resistance–associated protein is responsible for resistance in *T. b. rhodesiense*, whereas other mechanisms, notably involving the *T. b. gambiense*–specific glycoprotein (TgsGP) gene, are used by *T. b. gambiense*.

Trypanosomes are transmitted to humans by the tsetse bite, proliferate, and induce a local inflammatory reaction that is sometimes clinically apparent as a chancre. Trypanosomes then disseminate into the hematolymphatic system, with lymph nodes becoming enlarged after infiltration by mononuclear cells and lymphocytes. The degree of enlargement of the liver and spleen is usually mild to moderate, with infiltration by mononuclear cells as a prominent feature. Trypanosomes multiply in the blood, but their presence and density vary. This variation is mainly due to a cyclic immune-evasion process, whereby the parasite population can be decimated by the host's immune response until the reemergence of offspring parasites that express a different variant surface glycoprotein to which the immune system is temporarily blind. Each trypanosome genome encodes a repertoire of ~1000 variant surface glycoproteins between which the parasites can switch genetically. Trypanosomes also multiply in extravascular tissues during the first stage of illness. The skin, skeletal muscles, serous membranes (peritoneum, pleurae, and pericardium), and heart can be involved, with interstitial infiltration of mononuclear cells and vasculitis evident on microscopic examination. Myocarditis and pericarditis with myocardial degeneration and interstitial hemorrhage are common features of *T. b. rhodesiense* infection.

The CNS is invaded weeks to months (*T. b. rhodesiense*) or months to years (*T. b. gambiense*) after initial infection. This invasion corresponds to the second stage of HAT, which is defined by the presence of trypanosomes or mononuclear cells in the cerebrospinal fluid (CSF). The white matter is predominantly affected, with perivascular proliferation of astrocytes, microglial cells, and Mott's (morular) cells that contain IgM in intracellular vacuoles. The location of white-matter lesions in the brain correlates with the main neurologic clinical features. The cerebral cortex and neurons are spared until the terminal stages of illness. Because reversible inflammatory lesions predominate over the irreversible destruction of tissue, neuropsychiatric symptoms and signs resolve partially or completely during or after treatment of second-stage HAT.

APPROACH TO THE PATIENT

Human African Trypanosomiasis

HAT is usually lethal in the absence of treatment. Therefore, early diagnosis is crucial; physicians should include HAT in the differential diagnosis of several clinical syndromes when a patient has traveled or lived in at-risk sub-Saharan African countries, and obtaining a thorough recent and remote travel history from the patient is a prerequisite for diagnosis. In particular, HAT due to *T. b. gambiense* should be suspected in patients with persistent and intermittent fever or headaches, progressive neuropsychiatric disorders, and biologic signs of systemic inflammation, even if the last exposure occurred several years previously. HAT due to *T. b. rhodesiense* should be suspected in patients with an acute febrile illness and a recent exposure to tsetse flies in an eastern African country, especially if diagnostic tests for malaria are negative.

■ CLINICAL MANIFESTATIONS

The clinical presentations of *T. b. gambiense* and *T. b. rhodesiense* HAT usually differ. *T. b. gambiense* HAT is a slowly evolving illness with a long incubation period (months to years) and a prolonged disease course. In contrast, *T. b. rhodesiense* HAT is an acute febrile illness with a short (<3-week) incubation period and a shorter (weeks to months) disease course. There are exceptions to this classic pattern. Acute forms of *T. b. gambiense* HAT have been reported, especially among travelers, and chronic forms of *T. b. rhodesiense* HAT occur in the southern range of its geographic distribution (e.g., Zambia and Malawi). Trypano-resistance (i.e., self-resolving first-stage infections) and trypano-tolerance (i.e., the long-term persistence of parasites [e.g., in the skin] without clinical features of disease) have been reported for *T. b. gambiense*. Concomitant HIV co-infection does not seem to predispose individuals to an increased risk of HAT, and the impact of the virus on the clinical presentation of HAT is not known.

T. b. gambiense The occurrence of trypanosomal chancre is reported in a sizeable proportion of travelers, but very rarely in patients living in endemic areas, where the nonpurulent, painful, and itchy nodule can easily be confused with a lesion caused by the bite of another arthropod. The chancre spontaneously disappears in 1–3 weeks.

SYSTEMIC FEATURES After an asymptomatic incubation period that usually lasts for weeks or months but occasionally lasts for years, patients may present with irregular and remittent fever, sometimes accompanied by fatigue, malaise, and myalgia. Fever is more frequent among travelers than among natives, but the absence of fever in no way rules out the disease. Circinate or serpiginous rashes, commonly called trypanids, can occur on the trunk and on proximal parts of the extremities. Trypanids are almost impossible to detect on dark skin and have been reported only in Caucasians. Pruritus is a common but nonspecific symptom that affects up to half of patients during the second stage. Painless edema of the face and extremities occasionally occurs during the first phase.

Enlarged lymph nodes—a classic sign of HAT—are detected in 38–85% of patients at both disease stages. Cervical palpation is essential in patients with suspected HAT. The lateroposterior cervical group (Winterbottom sign) and the supraclavicular group are most commonly affected. Lymph nodes are movable, soft initially, harder later, and painless. A variable proportion of patients present with mild to moderate hepatomegaly and splenomegaly. Signs of myocarditis and pericarditis are occasionally detected by ECG and echocardiography but are usually clinically silent. Symptoms of HAT may mimic hypothyroidism or adrenal insufficiency, but thyroid and adrenal function tests yield normal results. Loss of libido, impotence, and amenorrhea, with decreased levels of testosterone and estradiol, are common in second-stage patients and are most likely caused by dysfunction of the hypothalamic–pituitary axis.

NEUROPSYCHIATRIC FEATURES Most patients with second-stage illness have no or only mild specific neuropsychiatric symptoms and signs, which, when they develop, tend to do so late in the disease course. In contrast, some nonspecific features, such as headaches and mood and behavioral changes, are present in both disease stages but become more permanent and severe during the second stage. As mentioned earlier, HAT is commonly called "sleeping sickness" because of various sleep disturbances (daytime somnolence, nocturnal insomnia) that are more pronounced late in the second stage. Dysregulation of the daily sleep/wake cycle and fragmentation of sleeping patterns are characteristic. Depending on the area of the brain affected, various neurologic syndromes also can develop, including disorders that are pyramidal-related (e.g., motor weakness, rare instances of hemiplegia), extrapyramidal-related (e.g., rigidity, paratonia), and cerebellar-related (e.g., ataxia, abnormal gait). Fine tremor, resting myoclonus, and abnormal (athetoid or choreic) movements also have been reported. Mental disorder is a key feature of HAT and can easily be misdiagnosed as primary psychiatric illness. Common presentations are antisocial or aggressive behavior, mood disorders (e.g., irritability, indifference), apathy or hyperactivity, and depression or psychosis (e.g., delirium, hallucinations). In the final stage of illness, decreased consciousness, dementia, and sometimes epilepsy are present, leading to coma, bed sores, aspiration pneumonia, or other bacterial infections and ultimately to death.

T. b. rhodesiense The clinical presentation of *T. b. rhodesiense* HAT can be similar to that of *T. b. gambiense* HAT in areas (e.g., Zambia, Malawi) that characteristically harbor specific parasite genotypes and host factors. The typical acute form with an incubation period of <3 weeks generally occurs in the northern range of the disease's distribution (e.g., Tanzania, Uganda) and in travelers. The initial trypanosomal chancre is clinically similar to that seen in *T. b. gambiense* HAT but is more common, especially among travelers.

SYSTEMIC FEATURES Fever can be high and occurs in both first- and second-stage patients, often in association with headaches and with diffuse myalgia and arthralgia. Pruritus and edema of the face and legs can be present. Lymphadenopathies have been reported in variable proportions in both disease stages and predominately affect the submandibular, axillary, and inguinal regions. Mild to moderate hepatomegaly and splenomegaly are documented in a minority of patients. Myocarditis and pericarditis appear to influence clinical course and outcome, even though clinical features of cardiac failure or arrhythmia have not been prominent findings in large case series. In contrast, conduction abnormalities, with various degrees of atrioventricular block, have been reported in travelers. Sepsis-like features, with disseminated intravascular coagulation and multiple-organ failure, can occur in the terminal stage.

NEUROPSYCHIATRIC FEATURES Neuropsychiatric symptoms and signs in *T. b. rhodesiense* HAT are reported with varying frequency but overall are similar to those described above for *T. b. gambiense* HAT. The notable exception in *T. b. rhodesiense* disease is a more rapid evolution toward coma and death.

■ DIAGNOSIS

The clinical and biologic features of *T. b. gambiense* and *T. b. rhodesiense* HAT—anemia, thrombocytopenia, elevated levels of C-reactive protein and IgM—are not sufficiently specific, and current drug regimens are not sufficiently practical to allow the initiation of treatment solely on the basis of suspicion. Diagnostic confirmation is therefore required in all patients.

T. b. gambiense The diagnosis of *T. b. gambiense* HAT is based on a three-step approach: screening, diagnostic confirmation, and staging.

SCREENING Immunologic (serologic) methods constitute the preferred screening tool. The card agglutination test for trypanosomiasis (CATT) has been used in most endemic areas for several decades. The test reagent contains stained, freeze-dried trypanosomes of selected variable-antigen types. If specific antibodies are present in the patient's blood or serum, agglutination can be seen with the naked eye. The sensitivity of the CATT on undiluted blood or serum is 69–100% (>90% in most studies), with some regional variation; its specificity is 84–99%. The CATT and associated equipment (e.g., a rotator) are manufactured and distributed by the Institute of Tropical Medicine in Antwerp, Belgium, but are not widely available outside endemic areas. In recent years, lateral flow tests have been developed and commercialized, first based on whole parasites and later on recombinant antigens. Their diagnostic performance is comparable to that of the CATT. Other serologic test formats (ELISA, immunofluorescence, indirect hemagglutination) are available in some reference laboratories in both endemic and nonendemic countries.

DIAGNOSTIC CONFIRMATION The microscopic observation of trypanosomes in the lymph, blood, skin, or CSF confirms the diagnosis. Direct observation of motile trypanosomes on a wet preparation of lymph obtained by cervical lymph node puncture is simple and cheap but has limited sensitivity (50–65% in most studies). Trypanosomes can be found in the blood but often occur at low densities. Therefore, stained thin and thick blood smears have very low sensitivity. Sensitivity is improved (to 40–60% in most studies) with the microhematocrit centrifugation technique, which is based on microscopic examination of the buffy coat after centrifugation of four to six microhematocrit tubes. The most sensitive method (~90%) is the miniature anion-exchange centrifugation technique, which is based on the visualization of trypanosomes in eluate after the passage of a large volume (500 μL) of blood through an anion-exchange column and subsequent centrifugation. Trypanosomes can also be visualized on microscopic examination of skin biopsies, even when parasites have not been visualized in the blood or lymph node.

STAGING Staging is based on the examination of CSF obtained by lumbar puncture. Second-stage HAT is defined by the presence in CSF of a raised leukocyte count (>5/μL) and/or of trypanosomes. The latter can be detected in the cell-counting chamber or, preferably, after centrifugation of the CSF. Staging is no longer an obligatory step in settings where fexinidazole is used as first-line treatment for both first- and second-stage HAT patients, except for young children (<6 years or weighing <20 kg) and for patients with neuropsychiatric symptoms and signs consistent with severe HAT, i.e., mental confusion, abnormal behavior, logorrhea, anxiety, ataxia, tremor, motor weakness, speech impairment, abnormal gait or movements, or seizures (see "Treatment," below).

Several molecular methods based on PCR or loop-mediated isothermal amplification have been developed, mostly based on the detection of multiple-copy DNA targets of the Trypanozoon group (to which *T. brucei* belongs) or the single-copy TgsGP gene of *T. b. gambiense*. None of these methods has been fully validated for diagnostic purposes, and a positive result of their application to blood should be interpreted as suspected rather than confirmed HAT. Molecular methods applied to CSF (to detect biomarkers) have not proved more accurate than classic methods for staging and have yielded false-positive results in a substantial proportion of cases.

T. b. rhodesiense The diagnosis of *T. b. rhodesiense* HAT is usually simpler because parasites are more numerous in body fluids. They can occasionally be visualized in a chancre aspirate. In light of the lack of available serologic tests and the high sensitivity of parasite detection methods in blood, wet mounts, thin/thick smears **(Fig. 234-4)**, and the microhematocrit or other concentration techniques are used for both screening and confirmation. Because the modalities of treatment of *T. b. rhodesiense* are stage dependent, staging remains an obligatory step, and the definition and methods used are the same as for *T. b. gambiense* HAT.

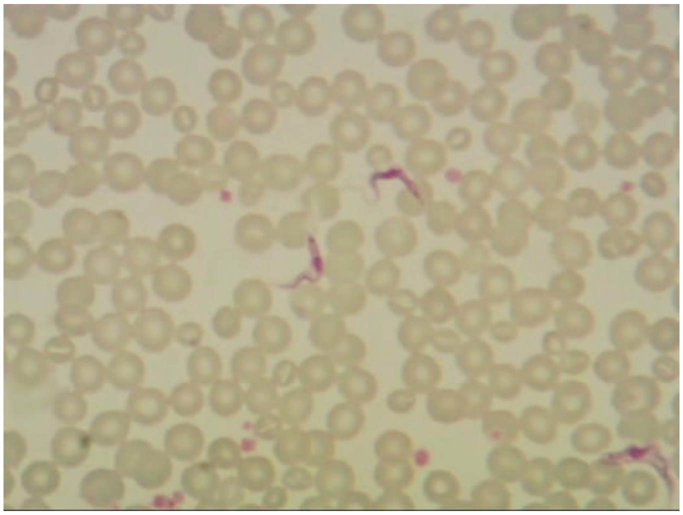

FIGURE 234-4 ***Trypanosoma brucei rhodesiense* in blood** (thin smear, Giemsa stain). *(Credit to the DPDx team, U.S. Centers for Disease Control and Prevention, Atlanta.)*

TREATMENT

Human African Trypanosomiasis

The management of HAT is based on general supportive therapy (e.g., rehydration, pain management), treatment of concomitant infections (e.g., malaria, pneumonia), and antiparasitic treatment. The modalities of antitrypanosomal treatment depend on the *Trypanosoma* species, the stage of illness, and the presence of contraindications **(Table 234-4)**.

T. B. GAMBIENSE

Fexinidazole, a nitroimidazole compound, is the first effective oral treatment against HAT. It is administered with food for 10 days, divided into a 4-day loading phase and a 6-day maintenance phase. It is highly effective (>95% cure rate) in patients with first-stage and nonsevere second-stage HAT, the latter being defined as <100 leukocytes/μL in the CSF. Fexinidazole is associated with a lower cure rate (87%) in patients with severe second-stage (≥100 leukocytes/μL in the CSF) HAT. The most relevant adverse reactions reported in clinical trials are vomiting, headache, and neuropsychiatric disorders (e.g., insomnia, anxiety, agitation). Fexinidazole is contraindicated in patients with hepatic insufficiency or at increased risk of QT interval prolongation. In the absence of safety and efficacy data, it remains contraindicated in small children (<6 years and/or weighing <20 kg).

Pentamidine isethionate is highly effective (>95%) against first-stage *T. b. gambiense* HAT and is an excellent alternative to fexinidazole when the latter is contraindicated or not available. It is generally well tolerated and can therefore be administered in peripheral health care centers in endemic countries **(Fig. 234-5)**. Hypotension after injection is common but generally mild. Hypoglycemia or hyperglycemia occasionally occurs, but permanent diabetes is very rare. Severe adverse events, such as acute pancreatitis and anaphylaxis, occur extremely rarely.

Nifurtimox–eflornithine combination therapy is very effective (>95% cure rate) and safe in patients with second-stage HAT, including patients with severe (≥100 leukocytes/μL in the CSF) illness. Common adverse reactions include gastrointestinal disturbances (nausea, vomiting, abdominal pain), headache, anorexia, and reversible bone marrow toxicity (anemia, leukopenia). Convulsions and psychosis are reported in <5% of patients.

Acoziborole, administered orally as a single-dose treatment (three tablets), cured >95% of 208 patients age >15 years with first- and second-stage HAT. Provided that its efficacy and safety is confirmed in a higher number of individuals, acoziborole could become the preferred treatment for *T. b. gambiense* HAT in the future.

TABLE 234-4 Treatment of Human African Trypanosomiasis (HAT)

	FIRST-LINE TREATMENT		
DISEASE AND STAGE	**DRUG(S) AND ROUTE**	**DOSE AND DURATION**	**ALTERNATIVE TREATMENT**
***T. b. gambiense* HAT**			
First stage	Fexinidazole PO	≥35 kg: 1800 mg for 4 days, followed by 1200 mg for 6 days 20–34 kg: 1200 mg for 4 days, followed by 600 mg for 6 days[a]	Pentamidine isethionate IM or IV[b]: 4 mg/kg per day for 7 days
Nonsevere second stage (6–99 leukocytes/μL in the cerebrospinal fluid [CSF])	Fexinidazole PO	≥35 kg: 1800 mg for 4 days, followed by 1200 mg for 6 days 20–34 kg: 1200 mg for 4 days, followed by 600 mg for 6 days[a]	Eflornithine: 200 mg/kg bid for 7 days *plus* Nifurtimox: 5 mg/kg tid for 10 days
Severe second stage (≥100 leukocytes/μL in the CSF)	Eflornithine IV + nifurtimox PO	Eflornithine: 200 mg/kg bid for 7 days Nifurtimox: 5 mg/kg tid for 10 days	Fexinidazole: ≥35 kg: 1800 mg for 4 days, followed by 1200 mg for 6 days 20–34 kg: 1200 mg for 4 days, followed by 600 mg for 6 days[a]
***T. b. rhodesiense* HAT**			
First stage	Suramin IV	4–5 mg/kg on day 1 followed by 5 weekly injections of 20 mg/kg (e.g., days 3, 10, 17, 24, 31)[c]	Pentamidine isethionate IM or IV[b]: 4 mg/kg per day for 7 days
Second stage	Melarsoprol IV	2.2 mg/kg per day for 10 days	—

[a]Fexinidazole should not be administered in children <6 years and weighing <20 kg. [b]For IV administration, slow infusion (60–120 min) should be used. [c]The maximal dose is 1 g per injection; the drug should be diluted in distilled water.

Sources: Control and surveillance of human African trypanosomiasis: Report of a WHO Expert Committee. WHO Technical Report Series 984, 2013; WHO interim guidelines for the treatment of gambiense human African trypanosomiasis. August 2019; *https://www.who.int/publications/i/item/9789241550567*.

T. B. RHODESIENSE

Suramin has been used for >90 years and remains the first-line treatment for first-stage *T. b. rhodesiense* HAT. Common adverse events are pyrexia and nephrotoxicity, which is usually mild and reversible but necessitates surveillance of albuminuria and renal function before each dose.

Because eflornithine is ineffective against *T. b. rhodesiense*, melarsoprol, an arsenic-based derivative, remains in use for second-stage *T. b. rhodesiense* HAT. Reactive encephalopathy is a life-threatening adverse event that occurs in 5–18% of patients, with an associated mortality rate of 10–70%. The efficacy of concomitant high-dose prednisolone to prevent reactive encephalopathy in patients with *T. b. rhodesiense* HAT is not known. Other severe but less frequent adverse reactions to melarsoprol include exfoliative dermatitis, bloody diarrhea, peripheral neuropathy, renal dysfunction, and liver toxicity. Phlebitis is common, as is soft tissue necrosis if the drug is accidentally given paravenously.

The 10-day fexinidazole oral treatment regimen (see above for *T. b. gambiense* HAT) was recently studied as an alternative to suramin and melarsoprol in Malawi and Uganda. Fexinidazole was administered in 45 patients with first-stage (n = 10) and second-stage (n = 35) patients and cured 43 (96%).

FIGURE 234-5 Intramuscular injection of pentamidine by a nurse in a village health center, Province Orientale, Democratic Republic of the Congo.

PROGNOSIS

Provided that treatment guidelines are properly followed, >95% of patients with first-stage and second-stage *T. b. gambiense* HAT are definitively cured with fexinidazole, pentamidine, and nifurtimox–eflornithine combination therapy. The overall case–fatality rate is <1% except in very advanced cases. Because relapses can occur long after completion of treatment, follow-up visits are advised every 6 months for at least 2 years. If clinical features of HAT are present, both blood and CSF examinations are indicated. Patients with second-stage *T. b. rhodesiense* HAT are at a 5–10% risk of dying during or after melarsoprol treatment, but relapses are very rare.

GLOBAL CONSIDERATIONS

The elimination of sleeping sickness as a public health problem has been achieved, thanks to increased control activities run by national control programs and nongovernmental medical organizations, improved funding, and the end of several civil wars (e.g., in Angola) in the past 20 years. Funding for research, development, and implementation of improved diagnostic (e.g., rapid diagnostic tests), therapeutic (e.g., oral drugs), and vector control tools remains crucial to sustain recent achievements and to move on to the next objective, i.e., the global elimination of transmission by 2030.

FURTHER READING

Bern C et al: Chagas disease in the United States: A public health approach. Clin Microbiol Rev 33:e00023-19, 2019.

Büscher P et al: Human African trypanosomiasis. Lancet 390:2397, 2017.

de Sousa AS et al: Chagas disease. Lancet 403:203, 2024.

Lindner AK et al: New WHO guidelines for treatment of gambiense human African trypanosomiasis including fexinidazole: Substantial changes for clinical practice. Lancet Infect Dis 20:e38, 2020.

Urech K et al: Sleeping sickness in travelers—Do they really sleep? PLoS Negl Trop Dis 5:e1358, 2011.

235

Toxoplasma Infections

Kami Kim

DEFINITION

Toxoplasmosis is caused by infection with the obligate intracellular parasite *Toxoplasma gondii*. Acute infection acquired after birth is typically asymptomatic, but some immunocompetent individuals can present with systemic or ocular disease. Acute infection is thought to result in the lifelong chronic persistence of cysts in the host's tissues. The classic presentation of toxoplasmosis is encephalitis in immunocompromised individuals (especially HIV-positive individuals or transplant recipients) in whom latent infection has reactivated. Among the clinical manifestations of the disease are lymphadenopathy, encephalitis, myocarditis, pneumonitis, and retinitis. Congenital toxoplasmosis is an infection of newborns that results from the transplacental passage of parasites from an infected mother to the fetus. These infants may be asymptomatic at birth, but many children later manifest signs and symptoms, including chorioretinitis, strabismus, epilepsy, and psychomotor retardation. Toxoplasmosis can also present as acute disease (typically chorioretinitis) associated with food- or waterborne sources.

ETIOLOGY

T. gondii is an intracellular coccidian that infects both birds and mammals. Up to a third of the world's population is thought to be infected latently with this organism. There are two distinct stages in the life cycle that are transmissible to humans **(Fig. 235-1)**. Tissue cysts that contain bradyzoites are transmitted in undercooked meat. After an intermediate host (e.g., a human, mouse, sheep, pig) ingests the cyst, it is rapidly digested by the acidic-pH gastric secretions. Sporulated oocysts that contain sporozoites are products of the sexual cycle in feline intestines and acquired by ingestion of food or water contaminated with infected cat feces. Bradyzoites or sporozoites are released, enter the intestinal epithelium, and transform into rapidly dividing tachyzoites. The tachyzoites can infect and replicate in all mammalian cells except red blood cells. The parasite actively penetrates the cell and forms a parasitophorous vacuole. Parasite replication continues within the vacuole. After the parasites reach a critical mass, intracellular signaling within the host and the parasite results in parasite egress from the vacuole. The host cell is destroyed, and the released tachyzoites infect adjoining cells. Parasites can disseminate throughout the body as free tachyzoites or within phagocytic cells in the bloodstream or via lymphatics. Tachyzoites actively invade host cells and can cross epithelial and endothelial barriers.

The tachyzoite replication cycle within an infected organ causes cytopathology and clinical symptoms. Most tachyzoites are eliminated by the host's humoral and cell-mediated immune responses. Tissue cysts containing bradyzoites develop 7–10 days after systemic tachyzoite infection. These tissue cysts occur in various host organs but persist principally within the central nervous system (CNS) and muscle. The development of this chronic stage completes the asexual portion of the life cycle. Active infection in the immunocompromised host is usually due to the spontaneous release of encysted parasites that undergo rapid transformation into tachyzoites within the CNS that cannot be contained by the immune system.

The *sexual* stage in the life cycle takes place in the cat (the definitive host) and is defined by the formation of oocysts within the feline host intestine. This enteroepithelial cycle begins with the ingestion of the bradyzoite tissue cysts and, after several intermediate stages, culminates in the production of gametes. Gamete fusion produces a zygote, which envelops itself in a rigid wall and is secreted in the feces as an unsporulated oocyst. After 2–3 days of exposure to air at ambient temperature, the noninfectious oocyst sporulates to produce eight sporozoite progeny. The sporulated oocyst can be ingested by an intermediate host. It is in the intermediate host that *T. gondii* completes its life cycle.

Sporulated oocysts are environmentally hardy and very infectious; they are thought to be sources of waterborne outbreaks such as those reported in Victoria (British Columbia, Canada) and in South America. In the Northern Hemisphere, *T. gondii* strains are predominantly of three genotypes. Strains found in South and Central America are more virulent than those from the Northern Hemisphere, are frequently of the type I virulent genotype or atypical genotypes, and are more likely to be associated with symptomatic disease, usually ocular posterior uveitis. Ocular toxoplasmosis should be considered in a person from Central or South America with ocular symptoms and retinal abnormalities. Severe disease, including sepsis, fever of unknown origin, and pneumonia, has been reported and should be considered in a patient with travel history to South or Central America. There have been reports of outbreaks in North America among individuals who have ingested undercooked game, especially venison, and these strains appear more virulent as the attack rate is often high. Prevalence of *T. gondii* in Africa suggests *T. gondii* infection is common and that some strains may also be quite virulent.

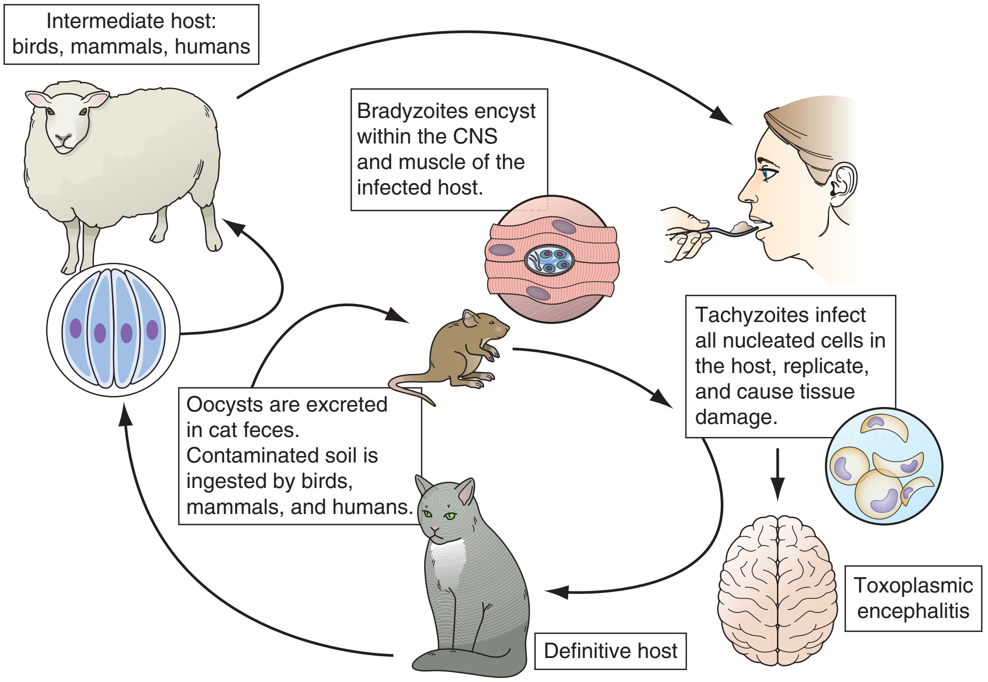
Intermediate host:
birds, mammals, humans
Bradyzoites encyst within the CNS and muscle of the infected host.
Tachyzoites infect all nucleated cells in the host, replicate, and cause tissue damage.
Oocysts are excreted in cat feces. Contaminated soil is ingested by birds, mammals, and humans.
Toxoplasmic encephalitis
Definitive host

FIGURE 235-1 Life cycle of *Toxoplasma gondii*. The cat is the definitive host in which the sexual phase of the cycle is completed. Oocysts shed in cat feces can infect a wide range of animals, including birds, rodents, grazing domestic animals, and humans. The bradyzoites found in the muscle of food animals may infect humans who eat insufficiently cooked meat products, particularly lamb and pork. Although human disease can take many forms, congenital infection and encephalitis from reactivation of latent infection in the brains of immunosuppressed persons are the most important manifestations. CNS, central nervous system. *(Courtesy of Dominique Buzoni-Gatel, Institut Pasteur, Paris.)*

EPIDEMIOLOGY

T. gondii infects a wide range of mammals and birds. Its seroprevalence depends on the locale and the age of the population. Generally, hot arid climatic conditions are associated with a low prevalence of infection. In the United States and most European countries, the seroprevalence increases with age and exposure. In the United States, seroprevalence has steadily decreased, with 11% of individuals >6 years old having serologic evidence of *Toxoplasma* exposure in a 2011–2014 survey, with foreign-born Americans having a higher rate of seroprevalence. In most other regions of the world, the seroprevalence is higher, with a seroprevalence as high as 78% reported in Brazil. Because of increased awareness of foodborne infections, the prevalence of seropositivity has decreased worldwide over the past two decades, although it remains very high in Central and South America.

TRANSMISSION

Oral Transmission Most cases of human *Toxoplasma* infection are thought to be acquired by the oral route. Transmission can be attributable to ingestion of sporulated oocysts from contaminated soil, food, or water. During acute feline infection, a cat may excrete as many as 100 million oocysts per day. These sporozoite-containing oocysts are highly infectious and may remain viable for many years in soil or water. Humans infected during an oocyst-transmitted infection develop stage-specific antibodies to the oocyst/sporozoite.

Children and adults also acquire infection from tissue cysts containing bradyzoites. Undercooking or insufficient freezing of meat is an important source of infection in the developed world. Toxoplasmosis has been associated with eating raw or undercooked food including ground beef, lamb, or venison or drinking unpasteurized goat milk. More recent epidemiologic studies have associated acute infections with ingestion of untreated water or shellfish (oysters, mussels, and clams).

Transmission via Blood or Organs In addition to being transmitted orally, *T. gondii* can be transmitted directly from a seropositive donor to a seronegative recipient in a transplanted heart, heart–lung, kidney, liver, or pancreas. Viable parasites can be cultured from refrigerated anticoagulated blood, which may be a source of infection in individuals receiving blood transfusions. *T. gondii* reactivation has been reported in bone marrow, hematopoietic stem cell, and liver transplant recipients as well as in individuals with AIDS. Finally, laboratory personnel can be infected after contact with contaminated needles or glassware or with infected tissue.

Transplacental Transmission On average, about one-third of all women who acquire infection with *T. gondii* during pregnancy transmit the parasite to the fetus; the remainder give birth to normal, uninfected babies. Of the various factors that influence fetal outcome, gestational age at the time of infection is the most critical (see below). Recrudescent maternal infection is rarely the source of congenital disease, although rare cases of transmission by immunocompromised women (e.g., those infected with HIV or those receiving high-dose glucocorticoids) have been reported. Thus, women who are seropositive before pregnancy usually are protected against acute infection and do not give birth to congenitally infected neonates.

There is essentially no risk for congenital infection if the mother becomes infected ≥6 months before conception. If infection is acquired <6 months before conception, the likelihood of transplacental infection increases as the interval between infection and conception decreases. Women with documented acute toxoplasmosis should be counseled to use appropriate measures to prevent pregnancy for 6 months after infection. In pregnancy, if the mother becomes infected during the first trimester, the incidence of transplacental infection is lowest (~15%), but the disease in the neonate is most severe. If maternal infection occurs during the third trimester, the incidence of transplacental infection is greatest (65%), but the infant is usually asymptomatic at birth. Infected infants who are normal at birth may have a higher incidence of learning disabilities and chronic neurologic sequelae than uninfected children. Only a small proportion (20%) of women infected with *T. gondii* develop clinical signs of infection. Often the diagnosis is first appreciated when routine postconception serologic tests show evidence of specific antibody. Chronic toxoplasmosis has not been thought to affect pregnancy, but there have been recent studies suggesting a greater incidence of adverse pregnancy outcomes without evidence of reactivated toxoplasmosis. The lingering effects of chronic infection are controversial (see below) and an area of ongoing investigation.

PATHOGENESIS

Upon the host's ingestion of either tissue cysts containing bradyzoites or oocysts containing sporozoites, the parasites are released from the cysts by the digestive process. Bradyzoites are resistant to the effect of pepsin and invade the host's gastrointestinal tract. Within enterocytes (or other gut-associated cells), the parasites undergo morphologic transformation, giving rise to invasive tachyzoites. From the gastrointestinal tract, parasites disseminate to a variety of organs, particularly lymphatic tissue, skeletal muscle, myocardium, retina, placenta, and the CNS. At these sites, the parasite infects host cells, replicates, and invades the adjoining cells. In this fashion, the hallmarks of the infection develop: cell death and focal necrosis surrounded by an acute inflammatory response.

In the immunocompetent host, both the humoral and the cellular immune responses control infection; parasite virulence and tissue tropism may be strain specific. Tachyzoites are sequestered by a variety of immune mechanisms, including induction of parasiticidal antibody, activation of macrophages with radical intermediates, production of interferon γ (IFN-γ), and stimulation of CD8+ cytotoxic T lymphocytes. These antigen-specific lymphocytes are capable of killing both extracellular parasites and target cells infected with parasites. As tachyzoites are cleared from the acutely infected host, tissue cysts containing bradyzoites begin to appear, usually within the CNS, skeletal muscle, and the retina. *Toxoplasma* secretes signaling molecules into infected host cells, and these molecules modulate host gene expression, host metabolism, and host immune response. It was initially thought that cysts with bradyzoites are not eliminated by the immune system, but more recent studies in the murine model indicate that both CD8+ T cells and alternatively activated macrophages are able to kill cysts in vivo. The ability to eliminate cysts may depend on the genetic background of the infected host as well as the parasite genotype that initially infected the host.

Immunocompromised or fetal hosts lack the immune factors necessary to control the spread of tachyzoite infection. This altered immune state allows the persistence of tachyzoites and gives rise to progressive focal destruction in affected organs (i.e., necrotizing encephalitis, pneumonia, and myocarditis).

It is thought that all infected individuals have persistent infection with cysts containing bradyzoites with lifelong infection remaining subclinical. Although bradyzoites are in a slow metabolic phase, bradyzoites can replicate, and cysts do rupture within the CNS. These subclinical cycles of cyst ruptures followed by development of new bradyzoite-containing cysts are the probable source of recrudescent infection in immunocompromised individuals and the most likely stimulus for the persistence of antibody titers in the immunocompetent host.

The persistence of toxoplasmosis has been hypothesized to be a contributing factor to a variety of neuropsychiatric conditions, including schizophrenia and bipolar disease, but the contribution of chronic toxoplasmosis to human disease remains controversial. In rodents, chronic *T. gondii* infection has significant effects on behavior, increasing predation. Epidemiologic studies such as the National Health and Nutrition Examination Survey (NHANES) study show a correlation between *T. gondii* seropositivity and a number of chronic diseases, including diabetes and cognitive dysfunction.

PATHOLOGY

Cell death and focal necrosis due to replicating tachyzoites induce an intense mononuclear inflammatory response in any tissue or cell type infected. Tachyzoites rarely can be visualized by routine histopathologic staining of these inflammatory lesions. However, immunofluorescent staining with parasitic antigen–specific antibodies can reveal the organism. In contrast to the inflammatory process caused by tachyzoites, bradyzoite-containing cysts cause inflammation only at the early stages of development. Once the cysts reach maturity, the inflammatory process is blunted, and the cysts remain relatively immunologically quiescent within the brain matrix until they rupture.

Lymph Nodes During acute infection, lymph node biopsy demonstrates characteristic findings, including follicular hyperplasia and irregular clusters of tissue macrophages with eosinophilic cytoplasm. Granulomas rarely are evident in these specimens. Although tachyzoites are not usually visible, parasites can be demonstrated by subinoculation of infected tissue into mice, with resultant disease, or by PCR. PCR amplification of DNA fragments of *Toxoplasma* genes is effective and sensitive in establishing lymph node infection by tachyzoites.

Eyes In the eye, infiltrates of monocytes, lymphocytes, and plasma cells may produce uni- or multifocal lesions. Granulomatous lesions and retinochoroiditis can be observed in the posterior chamber after acute necrotizing retinitis. Other ocular complications include iridocyclitis, cataracts, and glaucoma. *T. gondii* is the most common cause of posterior uveitis in immunocompetent individuals and ocular toxoplasmosis is a common clinical presentation in outbreaks.

Central Nervous System During CNS involvement, both focal and diffuse meningoencephalitis can be documented, with evidence of necrosis and microglial nodules. Necrotizing encephalitis in patients without AIDS is characterized by small diffuse lesions with perivascular cuffing in contiguous areas. In the AIDS population, polymorphonuclear leukocytes may be present in addition to monocytes, lymphocytes, and plasma cells. Cysts containing bradyzoites frequently are found contiguous with the necrotic tissue border. As a consequence of antiretroviral therapy (ART) for AIDS, the incidence of toxoplasmosis has decreased in the developed world. The incidence of toxoplasmosis in underresourced settings is not known due to lack of diagnostic infrastructure but is likely to be higher than in the United States, particularly in regions of the world with populations with large numbers of untreated patients living with HIV.

Lungs and Heart Among patients with AIDS who die of toxoplasmosis, 40–70% have involvement of the lungs and heart. Interstitial pneumonitis can develop in neonates and immunocompromised patients and rarely in immunocompetent individuals. Thickened and edematous alveolar septa infiltrated with mononuclear and plasma cells are apparent. This inflammation may extend to the endothelial walls. Tachyzoites and bradyzoite-containing cysts have been observed within the alveolar membrane. Superimposed bronchopneumonia can be caused by other microbial agents. Cysts and aggregates of parasites in cardiac muscle tissue are evident in patients with AIDS who die of toxoplasmosis. Focal necrosis surrounded by inflammatory cells is associated with hyaline necrosis and disrupted myocardial cells. Pericarditis is associated with toxoplasmosis in some patients.

Gastrointestinal Tract Rare cases of human gastrointestinal tract infection with *T. gondii* have presented as ulcerations in the mucosa. Acute infection in certain strains of inbred mice (C57BL/6) results in lethal ileitis within 7–9 days. This inflammatory bowel disease has been recognized in several other mammalian species, including pigs and nonhuman primates.

Other Sites Pathologic changes during disseminated infection are similar to those described for the lymph nodes, eyes, and CNS. In patients with AIDS, the skeletal muscle, pancreas, stomach, and kidneys can be involved, with necrosis, invasion by inflammatory cells, and (rarely) tachyzoites detectable by routine staining. Large necrotic lesions may cause direct tissue destruction. In addition, secondary effects from acute infection of these various organs, including pancreatitis, myositis, and glomerulonephritis, have been reported.

■ HOST IMMUNE RESPONSE

Acute *Toxoplasma* infection evokes a cascade of protective immune responses in the immunocompetent host. *Toxoplasma* enters the host at the gut mucosal level and evokes a mucosal immune response that includes the production of antigen-specific secretory IgA. Titers of serum IgA antibody directed at the tachyzoite surface antigen p30/SAG1 are a useful marker for congenital and acute toxoplasmosis.

Within the host, *T. gondii* rapidly induces detectable levels of both IgM and IgG serum antibodies. Monoclonal gammopathy of the IgG class can occur in congenitally infected infants. IgM levels may be increased in newborns with congenital infection. The polyclonal IgG antibodies evoked by infection are parasiticidal in vitro in the presence of serum complement and are the basis for the Sabin-Feldman dye test. However, cell-mediated immunity is the major protective response evoked by the parasite during host infection. Macrophages are activated after phagocytosis of antibody-opsonized parasites. If the parasite is not phagocytosed and enters the macrophage, monocytes, or dendritic cells by active penetration, these "Trojan horses" represent a mechanism for transport and dissemination to distant organs. *Toxoplasma* stimulates a robust interleukin (IL) 12 response by human dendritic cells.

The CD4+ and CD8+ T-cell responses are antigen-specific and further stimulate the production of a variety of important lymphokines that expand the T-cell and natural killer cell repertoire. *T. gondii* is a potent inducer of a T_H1 phenotype, with IL-12 and IFN-γ playing an essential role in the control of the parasites' growth in the host. Regulation of the inflammatory response is at least partially under the control of a T_H2 response that includes the production of IL-4 and IL-10 in seropositive individuals. Human T-cell clones of both the CD4+ and the CD8+ phenotypes are cytolytic against parasite-infected macrophages. These T-cell clones produce cytokines that are "microbistatic." IL-18, IL-7, and IL-15 upregulate the production of IFN-γ and may be important during acute and chronic infection. The effect of IFN-γ may be paradoxical, with stimulation of a host downregulatory response as well.

Although *T. gondii* infection is believed to be recrudescent in patients with AIDS or other immunocompromised states, antibody titers are not useful in establishing reactivation or in following the activity of infection. An absence of positive serologies suggests an alternative diagnosis, although AIDS patients may have borderline positive or low serologies and transplant patients treated with immunosuppressive agents including those specific for B cells may become seronegative. T cells from AIDS patients with reactivation of toxoplasmosis fail to secrete both IFN-γ and IL-2. This alteration in the production of these critical immune cytokines contributes to the persistence of infection. *Toxoplasma* infection develops late in the course of AIDS (CD4+ count <100/μL), when the loss of T cell–dependent protective mechanisms, particularly CD8+ T cells, becomes most pronounced.

■ CLINICAL MANIFESTATIONS

In persons whose immune systems are intact, acute toxoplasmosis is usually asymptomatic and self-limited. This condition can go unrecognized in 80–90% of adults and children with acquired infection. The asymptomatic nature of this infection makes diagnosis difficult in mothers infected during pregnancy. In contrast, the wide range of clinical manifestations in congenitally infected children includes severe neurologic complications such as hydrocephalus, microcephaly, intellectual disability, and chorioretinitis. If prenatal infection is severe, multiorgan failure and subsequent intrauterine fetal death can occur. In children and adults, chronic infection can persist throughout life, with little consequence to the immunocompetent host.

Toxoplasmosis in Immunocompetent Patients The most common manifestation of acute toxoplasmosis is cervical lymphadenopathy. The nodes may be single or multiple, are usually nontender, are discrete, and vary in firmness. Lymphadenopathy also may be found in suboccipital, supraclavicular, inguinal, and mediastinal areas. Generalized lymphadenopathy occurs in 20–30% of symptomatic patients. Between 20 and 40% of patients with lymphadenopathy also have headache, malaise, fatigue, and fever (usually with a temperature of <40°C [<104°F]). A smaller proportion of symptomatic individuals have myalgia, sore throat, abdominal pain, maculopapular rash, meningoencephalitis, and confusion.

Rare complications associated with infection in the normal immune host include pneumonia, myocarditis, encephalopathy, pericarditis, and polymyositis. These manifestations are often associated with more virulent parasitic genotypes/strains. Signs and symptoms associated with acute infection usually resolve within several weeks, although the lymphadenopathy may persist for some months. In one epidemic, toxoplasmosis was diagnosed correctly in only 3 of the 25 patients who consulted physicians. If toxoplasmosis is considered in the differential diagnosis, routine laboratory and serologic screening should precede node biopsy.

In North America and Europe, there are three predominant genotypes, but strains are more genetically diverse in Central and South America. Genotypes of *T. gondii* prevalent in South America are more virulent than those typically seen in North America or Europe. These genotypes may be associated with acute or recurrent ocular disease in immunocompetent individuals and have also been associated with pneumonitis and a fulminant sepsis picture in immunologically normal individuals. Thus, a detailed history, particularly regarding travel and countries of residence, is critical for establishing a diagnosis. Individuals from South or Central America may have frequent recurrences of ocular toxoplasmosis that may require treatment and suppressive therapy.

The results of routine laboratory studies are usually unremarkable except for minimal lymphocytosis, an elevated erythrocyte sedimentation rate, and a nominal increase in serum aminotransferase levels. Evaluation of cerebrospinal fluid (CSF) in cases with evidence of encephalopathy or meningoencephalitis shows an elevation of intracranial pressure, mononuclear pleocytosis (10–50 cells/mL), a slight increase in protein concentration, and (occasionally) an increase in the gamma globulin level. PCR amplification of the *Toxoplasma* DNA target sequence in CSF is specific for active toxoplasmosis, but not sensitive. PCR of ocular fluid or bronchoalveolar lavage may also be positive in those with ocular or pulmonary toxoplasmosis, respectively. The CSF of chronically infected individuals is normal.

Infection of Immunocompromised Patients Patients with AIDS, transplant patients, and those receiving immunosuppressive therapy for lymphoproliferative disorders are at greatest risk for developing acute toxoplasmosis. Toxoplasmosis has also been reported after treatment with antibodies to tumor necrosis factor. The infection may be due either to reactivation of latent infection or to acquisition of parasites from exogenous sources such as blood or transplanted organs. In individuals with AIDS, >95% of cases of *Toxoplasma* encephalitis (TE) are believed to be due to recrudescent infection. In most of these cases, encephalitis develops when the CD4+ T-cell count falls below 100/μL. In immunocompromised hosts, the disease may be rapidly fatal if untreated. Thus, accurate diagnosis and initiation of appropriate therapy are necessary to prevent fulminant infection.

Toxoplasmosis is a principal opportunistic infection of the CNS in persons with AIDS. Individuals with AIDS who are seropositive for *T. gondii* are at high risk for encephalitis. Before the advent of highly effective ART, about one-third of the 15–40% of adult AIDS patients in the United States who were latently infected with *T. gondii* developed TE. TE may still be a presenting infection in individuals who are unaware of their positive HIV status.

Individuals may be at relatively high risk for reactivation after allogeneic hematopoietic stem cell transplantation (HSCT), particularly if complicated by graft-versus-host reaction. Weekly PCR screening of blood from patients seropositive prior to HSCT is recommended, although not all centers routinely monitor HSCT patients for toxoplasmosis. Screening *Toxoplasma* serologies (donor and recipient) before transplantation identifies patients potentially at risk for reactivated toxoplasmosis. Serologies should be performed prior to initiation of immunosuppressive agents.

The signs and symptoms of acute toxoplasmosis in immunocompromised patients principally involve the CNS **(Fig. 235-2)**. More than 50% of patients with clinical manifestations have intracerebral involvement. Clinical findings at presentation range from nonfocal to focal dysfunction. CNS findings include encephalopathy, meningoencephalitis, and mass lesions. Patients may present with altered mental status (75%), fever (10–72%), seizures (33%), headaches (56%), and focal neurologic findings (60%), including motor deficits, cranial nerve palsies, movement disorders, dysmetria, visual-field loss, and aphasia. Patients who present with evidence of diffuse cortical dysfunction develop evidence of focal neurologic disease as infection progresses. This altered condition is due not only to the necrotizing encephalitis caused by direct invasion by the parasite but also to secondary effects, including vasculitis, edema, and hemorrhage. The onset of infection can range from an insidious process over several weeks to an acute presentation with fulminant focal deficits, including hemiparesis, hemiplegia, visual-field defects, localized headache, and focal seizures.

Although lesions can occur anywhere in the CNS, the areas most often involved appear to be the brainstem, basal ganglia, pituitary gland, and corticomedullary junction. Brainstem involvement gives rise to a variety of neurologic dysfunctions, including cranial nerve palsy, dysmetria, and ataxia. With basal ganglion infection, patients may develop hydrocephalus, choreiform movements, and choreoathetosis. *Toxoplasma* usually causes encephalitis, and meningeal involvement is uncommon. CSF findings may be unremarkable or may include a modest increase in cell count and in protein—but not glucose—concentration.

Cerebral toxoplasmosis must be differentiated from other opportunistic infections or tumors in the CNS of AIDS patients. The differential diagnosis includes herpes simplex encephalitis, cryptococcal meningitis, progressive multifocal leukoencephalopathy, and primary CNS lymphoma. Involvement of the pituitary gland can give rise to panhypopituitarism and hyponatremia from inappropriate secretion of vasopressin (antidiuretic hormone). HIV-associated neurocognitive disorder (HAND) may present as cognitive impairment, attention loss, and altered memory. Brain biopsy in patients who have been treated for TE but who continue to exhibit neurologic dysfunction often fails to identify organisms.

Autopsies of *Toxoplasma*-infected patients have demonstrated the involvement of multiple organs, including the lungs, gastrointestinal tract, pancreas, skin, eyes, heart, and liver. *Toxoplasma* pneumonia can be confused with *Pneumocystis* pneumonia. Respiratory involvement usually presents as dyspnea, fever, and a nonproductive cough and may rapidly progress to acute respiratory failure with hemoptysis, metabolic acidosis, hypotension, and (occasionally) disseminated intravascular coagulation. Histopathologic studies demonstrate necrosis and a mixed cellular infiltrate. The presence of organisms is a helpful diagnostic indicator, but organisms can also be found in healthy tissue. Infection of the heart is usually asymptomatic but can be associated with cardiac tamponade or biventricular failure. Infections of the gastrointestinal tract and the liver have been documented.

Congenital Toxoplasmosis Between 400 and 4000 infants born each year in the United States are affected by congenital toxoplasmosis. Acute infection in mothers acquiring *T. gondii* during pregnancy is usually asymptomatic; most women are diagnosed via prenatal serologic screening. Infection of the placenta leads to hematogenous

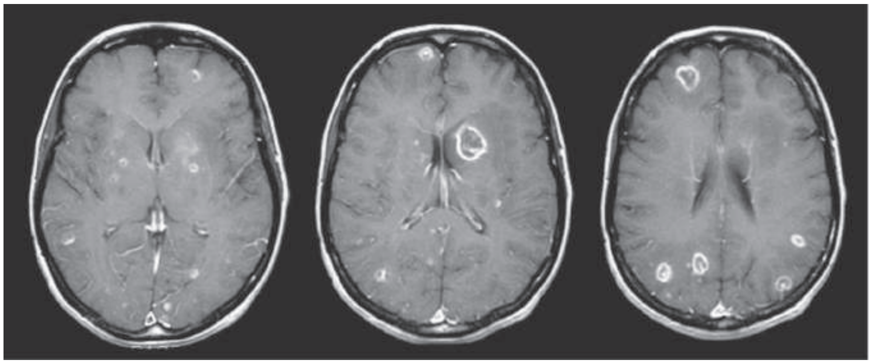

FIGURE 235-2 Toxoplasmic encephalitis in a 36-year-old patient with AIDS. The multiple lesions are demonstrated by magnetic resonance imaging scanning (T1-weighted with gadolinium enhancement). *(Courtesy of Clifford Eskey, Dartmouth Hitchcock Medical Center, Hanover, NH; with permission.)*

infection of the fetus. As gestation proceeds, the proportion of fetuses that become infected increases, but the clinical severity of the infection declines. Although infected children may initially be asymptomatic, the persistence of *T. gondii* can result in reactivation and clinical disease—most frequently chorioretinitis—decades later. Factors associated with relatively severe disabilities include delays in diagnosis and in initiation of therapy, neonatal hypoxia and hypoglycemia, profound visual impairment (see "Ocular Infection," below), uncorrected hydrocephalus, and increased intracranial pressure. If treated appropriately, upward of 70% of children have normal developmental, neurologic, and ophthalmologic findings at follow-up evaluations. Treatment for 1 year with pyrimethamine, a sulfonamide, and folinic acid is tolerated with minimal toxicity (see "Treatment," below).

Ocular Infection Infection with *T. gondii* is estimated to cause 35% of all cases of chorioretinitis in the United States and Europe. It was formerly thought that the majority of cases of ocular disease were due to congenital infection. Ocular toxoplasmosis in immunocompetent individuals occurs more commonly than was previously appreciated and has been associated with outbreaks traced to oocyst contamination of soil or water in Victoria (British Columbia) and in South America. Outbreaks have also been reported in North America after ingestion of undercooked venison.

A variety of ocular manifestations are documented, including blurred vision, scotoma, photophobia, and eye pain. Macular involvement occurs, with loss of central vision, and nystagmus is secondary to poor fixation. Involvement of the extraocular muscles may lead to disorders of convergence and to strabismus. Ophthalmologic examination should be undertaken in newborns with suspected congenital infection. As the inflammation resolves, vision improves, but episodic flare-ups of chorioretinitis, which progressively destroy retinal tissue and lead to glaucoma, are common. The ophthalmologic examination reveals yellow-white, cotton-like patches with indistinct margins of hyperemia. As the lesions age, white plaques with distinct borders and black spots within the retinal pigment become more apparent. Lesions usually are located near the posterior pole of the retina; they may be single but are more commonly multiple. Congenital lesions may be unilateral or bilateral and show evidence of massive chorioretinal degeneration with extensive fibrosis. Surrounding these areas of involvement are a normal retina and vasculature. In patients with AIDS, retinal lesions are often large, with diffuse retinal necrosis, and include both free tachyzoites and cysts containing bradyzoites. Toxoplasmic chorioretinitis may be a prodrome to the development of encephalitis.

DIAGNOSIS

Tissue and Body Fluids The differential diagnosis of acute toxoplasmosis can be made by appropriate culture, serologic testing, and PCR **(Table 235-1)**. PCR is the mainstay for detection of organisms in tissue or biological fluids, but a negative PCR does not rule out toxoplasmosis. Isolation or PCR of *T. gondii* from the patient's body fluids (blood, CSF, or bronchoalveolar lavage) reflects acute infection, whereas isolation from biopsied tissue is an indication only of the presence of tissue cysts and should not be misinterpreted as evidence of acute toxoplasmosis. Persistent parasitemia in patients with latent, asymptomatic infection is rare. Histologic examination of lymph nodes may suggest the characteristic changes described above. Demonstration of tachyzoites in lymph nodes establishes the diagnosis of acute toxoplasmosis. Histologic demonstration of cysts containing bradyzoites confirms prior infection with *T. gondii* but may represent latent rather than acute infection.

Serology Serologic testing has become the routine method of diagnosis. Diagnosis of acute infection with *T. gondii* can be established by detection of the simultaneous presence of IgG and IgM antibodies to *Toxoplasma* in serum. The presence of circulating IgA favors the diagnosis of an acute infection. The Sabin-Feldman dye test, the indirect fluorescent antibody test, and the enzyme-linked immunosorbent assay (ELISA) all measure circulating IgG antibody to *Toxoplasma*. Positive IgG titers (>1:10) can be detected as early as 2–3 weeks after infection. These titers usually peak at 6–8 weeks and decline slowly to a new baseline level that persists for life. Antibody avidity increases with time and can be useful in difficult cases during pregnancy for establishing when infection may have occurred. The serum IgM titer should be measured in concert with the IgG titer to better establish the time of infection; either the double-sandwich IgM-ELISA or the IgM-immunosorbent assay (IgM-ISAGA) should be used. Both assays are specific and sensitive, with fewer false-positive results than other commercial tests. The double-sandwich IgA-ELISA is more sensitive than the IgM-ELISA for detecting congenital infection in the fetus and newborn. Although a negative IgM result with a positive IgG titer indicates distant infection, IgM can persist for >1 year and should not necessarily be considered a reflection of acute disease. If acute toxoplasmosis is suspected, a more extensive panel of serologic tests can be performed. In the United States, testing is available at the Remington Laboratory for Specialty Diagnostics (*https://www.sutterhealth.org/services/lab-pathology/toxoplasma-serology-laboratory*).

TABLE 235-1 Differential Laboratory Diagnosis of Toxoplasmosis

CLINICAL SETTING	ALTERNATIVE DIAGNOSIS	DISTINGUISHING CHARACTERISTICS
Mononucleosis syndrome	Epstein-Barr virus infection	Serology/PCR
	Cytomegalovirus infection	PCR/viral load/serology
	HIV infection	Serology/antigen/viral load
	Bartonella infection (cat-scratch disease)	Biopsy (PCR or culture)/serology
	Lymphoma	Biopsy
Congenital infection	Cytomegalovirus infection	PCR
	Herpes simplex virus infection	PCR
	Rubella virus infection	Serology
	Syphilis	Serology
	Listeriosis	Bacterial culture
Chorioretinitis in immunocompetent individual	Tuberculosis	Bacterial culture/PCR
	Syphilis	Serology
	Histoplasmosis	Serology/culture/antigen
Chorioretinitis in AIDS patient	Cytomegalovirus infection	Characteristic exam
	Syphilis	Serology
	Herpes simplex virus infection	PCR
	Varicella-zoster virus infection	PCR
	Fungal infection	PCR/culture
CNS lesions in AIDS patient	Lymphoma or metastatic tumor	Tissue biopsy
	Brain abscess	Culture/biopsy
	Progressive multifocal leukoencephalopathy	PCR for JC virus
	Fungal infection	Antigen/PCR/biopsy/culture
	Mycobacterial infection	PCR/biopsy/culture

Abbreviations: CNS, central nervous system; PCR, polymerase chain reaction.

Source: Reproduced with permission from JD Schwartzman: Toxoplasmosis, in *Principles and Practice of Clinical Parasitology.* Hoboken, Wiley; 2001.

Molecular Diagnostics Molecular approaches can directly detect *T. gondii* in biologic samples independent of the serologic response. Results obtained with PCR have high specificity and clinical utility in the diagnosis of toxoplasmosis but may only be available in specialty laboratories. While very specific, depending on the body fluid type tested, the sensitivity of PCR of body fluids may be low, so diagnostic algorithms typically incorporate serologic testing of blood or body fluids. Real-time PCR, if available, can provide quantitative results. Molecular epidemiologic studies with polymorphic markers have been useful in correlating clinical signs and symptoms of disease with different *T. gondii* genotypes that may vary in virulence. Next-generation sequencing approaches (metagenomics) of blood or body fluids are often useful, particularly in immunocompromised individuals.

The Immunocompetent Adult or Child For the patient who presents with lymphadenopathy only, a positive IgM titer is an indication of acute infection—and an indication for therapy, if clinically warranted (see "Treatment," below). The serum IgM titer should be determined again in 3 weeks. An elevation in the IgG titer without an increase in the IgM titer suggests that infection is present but is not acute. If there is a borderline increase in either IgG or IgM, the titers should be reassessed in 3–4 weeks.

The Immunocompromised Host A presumptive clinical diagnosis of TE in patients with AIDS is based on clinical presentation, history of exposure (as evidenced by positive serology), and radiologic evaluation. To detect latent infection with *T. gondii*, HIV-infected persons should be tested for IgG antibody to *Toxoplasma* soon after HIV infection is diagnosed. When these criteria are used, the predictive value is as high as 80%. More than 97% of patients with AIDS and toxoplasmosis have IgG antibody to *T. gondii* in serum. IgM serum antibody usually is not detectable. Although IgG titers do not correlate with active infection, serologic evidence of infection almost always precedes the development of TE. It is therefore important to determine the *Toxoplasma* antibody status of all patients infected with HIV. Antibody titers may range from negative to 1:1024 in patients with AIDS and TE. Fewer than 3% of patients have no demonstrable antibody to *Toxoplasma* at diagnosis of TE.

Patients with TE have focal or multifocal abnormalities demonstrable by computed tomography (CT) or magnetic resonance imaging (MRI). Neuroradiologic evaluation should include double-dose contrast CT of the head. By this test, single and frequently multiple contrast-enhancing lesions (<2 cm) may be identified. MRI usually demonstrates multiple lesions located in both hemispheres, with the basal ganglia and corticomedullary junction most commonly involved; MRI provides a more sensitive evaluation of the efficacy of therapy than does CT (Fig. 235-2). These findings are not pathognomonic of *Toxoplasma* infection, because 40% of CNS lymphomas are multifocal and 50% are ring-enhancing. For both MRI and CT scans, the rate of false-negative results is ~10%. The finding of a single lesion on an MRI scan increases the likelihood of primary CNS lymphoma (in which solitary lesions are four times more likely than in TE) and strengthens the argument for the performance of a brain biopsy. A therapeutic trial of anti-*Toxoplasma* medications is frequently used to assess the diagnosis. Treatment of presumptive TE with pyrimethamine plus sulfadiazine or clindamycin results in quantifiable clinical improvement in >50% of patients by day 3. Leucovorin is administered to prevent bone marrow toxicity. By day 7, >90% of treated patients show evidence of improvement. In contrast, if patients fail to respond or have lymphoma, clinical signs and symptoms worsen by day 7. Patients in this category require brain biopsy with or without a change in therapy. This procedure can now be performed by a stereotactic CT-guided method that reduces the potential for complications. Brain biopsy for *T. gondii* identifies organisms in 50–75% of cases. PCR amplification of CSF may also confirm toxoplasmosis or suggest alternative diagnoses (Table 235-1), such as progressive multifocal leukoencephalopathy (JC virus positive) or primary CNS lymphoma (Epstein-Barr virus positive).

CT and MRI with contrast are currently the standard diagnostic imaging tests for TE. As in other conditions, the radiologic response may lag behind the clinical response. Resolution of lesions may take from 3 weeks to 6 months. Some patients show clinical improvement despite worsening radiographic findings.

Congenital Infection The issue of concern when a pregnant woman has evidence of recent *T. gondii* infection is whether the fetus is infected. PCR analysis of the amniotic fluid has replaced fetal blood sampling. Serologic diagnosis is based on the persistence of IgG antibody or a positive IgM titer after the first week of life (a time frame that excludes placental leak). The IgG determination should be repeated every 2 months. An increase in IgM beyond the first week of life is indicative of acute infection. Up to 25% of infected newborns may be seronegative and have normal routine physical examinations. Thus, assessment of the eye and the brain, with ophthalmologic testing, CSF evaluation, and radiologic studies, is important in establishing the diagnosis.

Ocular Toxoplasmosis The serum antibody titer may not correlate with the presence of active lesions in the fundus, particularly in cases of congenital toxoplasmosis. In general, a positive IgG titer (measured in undiluted serum if necessary) in conjunction with typical lesions establishes the diagnosis. If lesions are atypical and the serum antibody titer is in the low-positive range, the diagnosis is presumptive. The parasitic antigen–specific polyclonal IgG assay as well as parasite–specific PCR may facilitate the diagnosis. PCR of ocular samples has better yield than PCR of blood, but negative PCR does not rule out the diagnosis. Diagnosis may also be established by ocular fluid Western blot or comparison of ocular fluid antibody with blood antibody (Goldmann-Witmer coefficient). The clinical diagnosis of ocular toxoplasmosis can be supported in 60–90% of cases by laboratory tests, depending on the time of anterior chamber puncture and the panel of antibody analyses used.

TREATMENT

Toxoplasmosis

CONGENITAL INFECTION

Congenitally infected neonates are treated with daily oral pyrimethamine (1 mg/kg) and sulfadiazine (100 mg/kg) with folinic acid for 1 year. Depending on the signs and symptoms, prednisone (1 mg/kg per day) may be used for congenital infection. Some U.S. states and some countries routinely screen pregnant women (France, Austria) and/or newborns (Denmark, Massachusetts). Management and treatment regimens vary with the country and the treatment center. Most experts use spiramycin to treat pregnant women who have acute toxoplasmosis early in pregnancy and use pyrimethamine/sulfadiazine/folinic acid to treat women who seroconvert after 18 weeks of pregnancy or in cases of documented fetal infection. Studies suggest that treatment during pregnancy decreases the severity of infection. Many women who are infected in the first trimester elect termination of pregnancy. Those who do not terminate pregnancy are offered prenatal antibiotic therapy to reduce the frequency and severity of *Toxoplasma* infection in the infant. The optimal duration of treatment for a child with asymptomatic congenital toxoplasmosis is not clear, although most clinicians in the United States would treat the child for 1 year in light of cohort investigations conducted by the National Collaborative Chicago-Based, Congenital Toxoplasmosis Study.

INFECTION IN IMMUNOCOMPETENT PATIENTS

Immunologically competent adults and older children who have only lymphadenopathy do not require specific therapy unless they have persistent, severe symptoms. Patients with ocular toxoplasmosis are usually treated for 6 weeks with pyrimethamine plus either sulfadiazine or clindamycin and sometimes with prednisone. Trimethoprim-sulfamethoxazole (TMP-SMX) can also be given if pyrimethamine cannot be obtained (5 mg/kg bid based on TMP). Treatment should be supervised by an ophthalmologist familiar with *Toxoplasma* disease. Ocular disease can be self-limited without treatment, but therapy is typically considered for lesions that are severe or close to the fovea or optic disc. Prolonged treatment with TMP-SMX prevents recurrences of ocular toxoplasmosis while on treatment and is often considered in individuals with frequent flares in a 1- to 2-year period. Flares are more common in individuals who have acquired infection in South America. Whether treatment improves long-term visual outcomes is unclear. Other clinical presentations of toxoplasmosis in immunocompetent individuals are treated 2–4 weeks with duration and decision to treat based upon response and severity of clinical symptoms.

INFECTION IN IMMUNOCOMPROMISED PATIENTS

Clinical Treatment Immunocompromised patients, such as patients with AIDS and/or transplant recipients, should be treated for acute toxoplasmosis as toxoplasmosis is rapidly fatal if untreated. The recommended treatment is pyrimethamine (200 mg load, 50 mg/d if <60 kg, 75 mg/d if >60 kg) plus sulfadiazine (1000 mg qid for <60 kg; 1500 mg qid >60 kg) plus leucovorin (10–25 mg/d) to reduce hematologic toxicity for a minimum of 6 weeks. In cases of sulfa intolerance, clindamycin (600 mg qid) can be substituted. TMP-SMX (5 mg/kg of TMP, 25 mg/kg SMX bid) appears to be an effective alternative for treatment of TE in resource-poor settings where the preferred combination of pyrimethamine plus sulfadiazine is not available. Pyrimethamine is very expensive in the United States, so many clinicians prescribe TMP-SMX if pyrimethamine cannot be obtained. Most experts continue to prefer pyrimethamine plus sulfadiazine because of better synergy between the drugs and fewer clinical failures relative to TMP-SMX.

Primary Prophylaxis in AIDS The incidence of TE has declined as the survival of patients with HIV infection has increased through the use of ART. The incidence of TE in underresourced settings is unknown because serologic testing and imaging are not available. AIDS patients who are seropositive for *T. gondii* and who have a CD4+ T lymphocyte count of <100/μL should receive prophylaxis against TE.

The daily dose of TMP-SMX (one double-strength tablet qd) recommended for prophylaxis of *Pneumocystis jirovecii* pneumonia (PJP; formerly *Pneumocystis carinii* or PCP) is effective against TE. If patients cannot tolerate TMP-SMX, the recommended alternative is dapsone-pyrimethamine, which likewise is effective against PJP. Atovaquone with or without pyrimethamine also can be considered. Prophylactic monotherapy with dapsone, pyrimethamine, azithromycin, clarithromycin, or aerosolized pentamidine is probably insufficient.

Discontinuing Primary Prophylaxis Prophylaxis against TE can be discontinued in HIV-positive patients who have responded to ART and whose CD4+ T lymphocyte count has been >200/μL for 3 months. Although patients with CD4+ T lymphocyte counts of <100/μL are at greatest risk for developing TE, the risk that this condition will develop when the count has increased to 100–200/μL has not been established. Thus, prophylaxis should be discontinued when the count has increased to >200/μL. Prophylaxis should be recommenced if the CD4+ T lymphocyte count again decreases to <100–200/μL.

Individuals who have completed initial therapy for TE should receive treatment indefinitely until immune reconstitution, with a CD4+ T-cell count of >200/μL, is achieved with ART. Combination therapy with pyrimethamine plus sulfadiazine plus leucovorin is effective. An alternative to sulfadiazine in this regimen is clindamycin or TMP-SMX.

Discontinuing Secondary Prophylaxis (Long-Term Maintenance Therapy) Patients receiving secondary prophylaxis for TE are at low risk for recurrence when they have completed initial therapy for TE, remain asymptomatic, and have evidence of restored immune function. Individuals with HIV infection should have a CD4+ T lymphocyte count of >200/μL for at least 6 months after combined ART (cART). A repeat MRI brain scan is recommended. Secondary prophylaxis should be reintroduced if the CD4+ T lymphocyte count decreases to <200/μL.

Prophylaxis in Allogeneic HSCT or Solid Organ Transplant The incidence of toxoplasmosis is lower in seropositive allogeneic HSCT given TMP-SMX prophylaxis, so prophylaxis should be given after engraftment for at least 6 months with weekly PCR monitoring performed after transplant. TMP-SMX regimens will prevent both toxoplasmosis and *P. jirovecii*.

PREVENTION

Seronegative immunocompromised or pregnant persons should be counseled regarding sources of *Toxoplasma* infection. The chances of primary infection with *Toxoplasma* can be reduced by not eating undercooked meat and by avoiding oocyst-contaminated material (i.e., a cat's litter box). Specifically, lamb, beef, pork, and venison should be cooked to an internal temperature of 63°C (145°F) measured in the thickest portion of the cut and rested for 3 min. Ground meat should be cooked to 71°C (145°F), whereas poultry should be cooked to 74°C (165°F). Hands should be washed thoroughly after work in the garden, and all fruits and vegetables should be washed. Freezing meat to −20°C (−4°F) also kills cysts. Ingestion of raw shellfish is a risk factor for toxoplasmosis, given that the filter-feeding mechanism of clams and mussels concentrates oocysts.

If the patient owns a cat, the litter box should be cleaned or changed daily, preferably by an HIV-negative, nonpregnant person; alternatively, patients should wash their hands thoroughly after changing the litter box. Litter boxes should be changed daily if possible, as freshly excreted oocysts will not have sporulated and will not be infectious. Patients should be encouraged to keep their cats inside and not to adopt or handle stray cats. Cats should be fed only canned or dried commercial food or well-cooked table food, not raw or undercooked meats. Patients need not be advised to part with their cats or to have their cats tested for toxoplasmosis. Ideally, blood intended for transfusion into *Toxoplasma*-seronegative immunocompromised individuals should be screened for antibody to *T. gondii*. Although such serologic screening is not routinely performed, seronegative women should be screened for evidence of infection several times during pregnancy if they are exposed to environmental conditions that put them at risk for infection with *T. gondii*. HIV-positive individuals should adhere closely to these preventive measures.

ACKNOWLEDGMENT

The author would like to acknowledge Dr. Lloyd Kasper for his numerous contributions to our understanding of the pathogenesis of toxoplasmosis and his essential role in preparation of this chapter for prior editions.

FURTHER READING

Aerts R et al: Guidelines for the management of *Toxoplasma gondii* infection and disease in patients with haematological malignancies and after haematopoietic stem-cell transplantation: Guidelines from the 9th European Conference on Infections in Leukaemia, 2022. Lancet Infect Dis 24:e291, 2024.

Cifuentes-Gonzalez C et al: Risk factors for recurrences and visual impairment in patients with ocular toxoplasmosis: A systematic review and meta-analysis. PLoS One A18:e0283845, 2023.

Jones JL et al: *Toxoplasma gondii* infection in the United States, 2011–2014. Am J Trop Med Hyg 98:551 2018.

Schumacher AC et al: Toxoplasmosis outbreak associated with *Toxoplasma gondii*-contaminated venison–high attack rate, unusual clinical presentation, and atypical genotype. Clin Infect Dis 72:1557, 2021.

Song G et al: *Toxoplasma gondii* seropositivity and cognitive functioning in older adults: An analysis of cross-sectional data of the National Health and Nutrition Examination Survey 2011–2014. BMJ Open 14:e071513, 2024.

236

Protozoal Intestinal Infections and Trichomoniasis

Peter F. Weller, Edward T. Ryan

PROTOZOAL INFECTIONS

GIARDIASIS

Giardia duodenalis (also known as *G. lamblia* or *G. intestinalis*) is a cosmopolitan protozoal parasite that inhabits the small intestines of humans and other mammals. Giardiasis is one of the most common parasitic diseases in both developed and developing countries worldwide, causing both endemic and epidemic intestinal disease and diarrhea. In resource-constrained areas with limited access to adequate sanitation and safer water, prevalence rates of giardiasis may be 20–40%, especially among young children.

Life Cycle and Epidemiology (Fig. 236-1) Infection follows the ingestion of environmentally hardy cysts, which excyst in the small intestine, releasing flagellated trophozoites (Fig. 236-2) that multiply by binary fission. *Giardia* remains a pathogen of the proximal small bowel and does not disseminate hematogenously. Trophozoites remain free in the lumen or attach to the mucosal epithelium by means of a ventral sucking disk. As a trophozoite encounters altered conditions, it forms a morphologically distinct cyst, which is the stage of the parasite usually found in the feces. Trophozoites may be present and even predominate in loose or watery stools, but it is the resistant cyst that survives outside the body and is responsible for transmission. Cysts do not tolerate heating or desiccation, but they do remain viable for months in cold fresh water. The number of cysts excreted varies widely but can approach 10^7 per gram of stool.

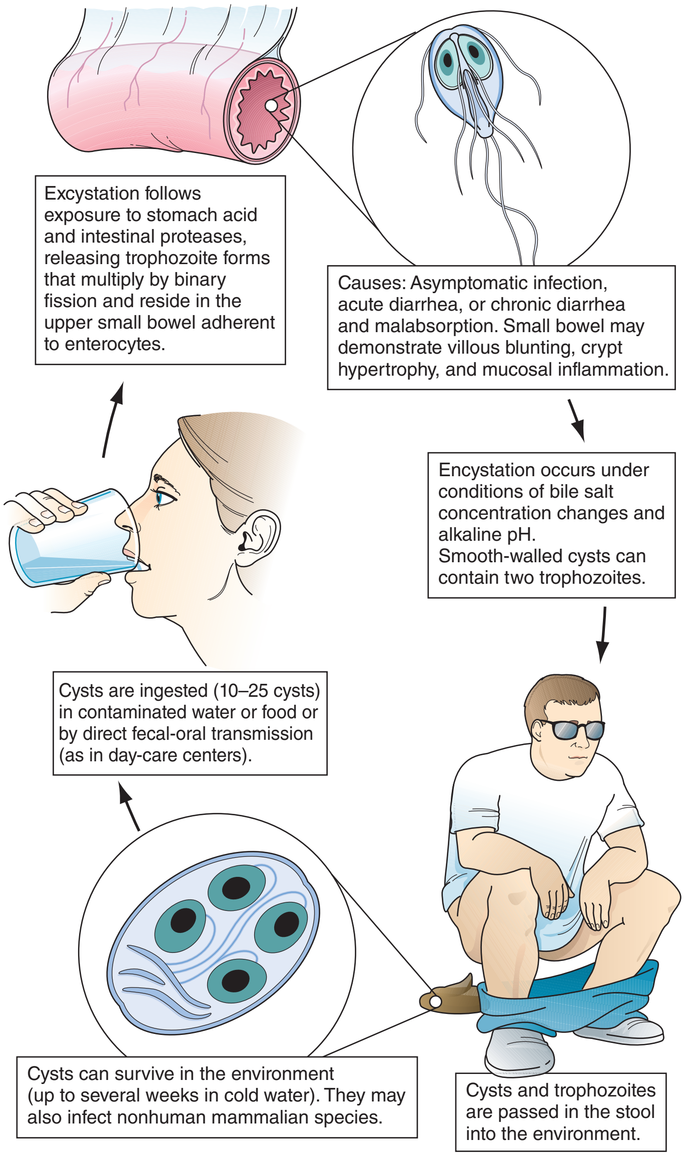
Excystation follows exposure to stomach acid and intestinal proteases, releasing trophozoite forms that multiply by binary fission and reside in the upper small bowel adherent to enterocytes.
Causes: Asymptomatic infection, acute diarrhea, or chronic diarrhea and malabsorption. Small bowel may demonstrate villous blunting, crypt hypertrophy, and mucosal inflammation.
Encystation occurs under conditions of bile salt concentration changes and alkaline pH. Smooth-walled cysts can contain two trophozoites.
Cysts are ingested (10–25 cysts) in contaminated water or food or by direct fecal-oral transmission (as in day-care centers).
Cysts can survive in the environment (up to several weeks in cold water). They may also infect nonhuman mammalian species.
Cysts and trophozoites are passed in the stool into the environment.

FIGURE 236-1 Life cycle of *Giardia*. *(Reproduced with permission from RL Guerrant et al [eds]: Tropical Infectious Diseases: Principles, Pathogens and Practice, 2nd ed. Elsevier, 2006.)*

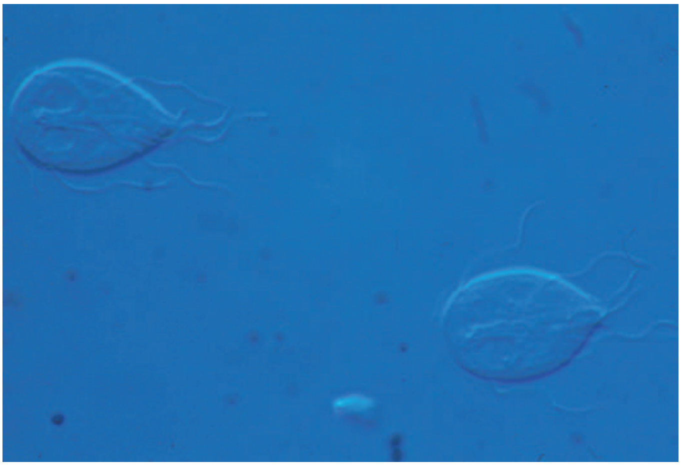

FIGURE 236-2 Flagellated, binucleate *Giardia* trophozoites.

Ingestion of as few as 10 cysts is sufficient to cause infection in humans. Because cysts are infectious when excreted, person-to-person transmission occurs where fecal hygiene is poor. Giardiasis is especially prevalent in day-care centers; person-to-person spread also takes place in other institutional settings with poor fecal hygiene and during anal–oral contact. If food is contaminated with *Giardia* cysts after cooking or preparation, foodborne transmission can occur. Waterborne transmission accounts for episodic infections (e.g., in campers and travelers) and for major epidemics in metropolitan areas. Surface water, ranging from mountain streams to large municipal reservoirs, can become contaminated with fecally derived *Giardia* cysts. The efficacy of water as a means of transmission is enhanced by the small infectious inoculum of *Giardia*, the prolonged survival of cysts in cold water, and the resistance of cysts to killing by routine chlorination methods that are adequate for controlling bacteria. Viable cysts can be eradicated from water by either boiling or filtration.

In the United States, *Giardia* (like *Cryptosporidium*; see below) can cause waterborne epidemics of gastroenteritis. *Giardia* is common in developing countries, and infections may be acquired by travelers.

There are several recognized genotypes or assemblages of *G. duodenalis*. Human infections are due to assemblages A and B, whereas other assemblages are more common in other animals, including cats and dogs. Like beavers from reservoirs implicated in epidemics, dogs and cats have been found to be infected with assemblages A and B; this finding suggests both that these animals may have been infected from human sources and that they might be sources of further human infections.

Giardiasis, like cryptosporidiosis, creates a significant economic burden because of the costs incurred in the installation of water filtration systems required to prevent waterborne epidemics, in the management of epidemics that involve large communities, and in the evaluation and treatment of endemic infections.

Pathophysiology The reasons that some, but not all, infected patients develop clinical manifestations and the mechanisms by which *Giardia* causes alterations in small-bowel function are largely unknown. Although trophozoites adhere to the epithelium, they are not invasive but may elicit apoptosis of enterocytes, epithelial barrier dysfunction, and epithelial cell malabsorption and secretion.

TABLE 236-1 Diagnosis of Intestinal Protozoal Infections

PARASITE	STOOL O+P	FECAL ACID-FAST STAIN	FECAL ANTIGEN IMMUNOASSAYS	FECAL NAATS	OTHER
Giardia	+		+	+	DFA
Cryptosporidium	±	+	+	+	DFA
Cystoisospora	±	+		+	
Cyclospora	±	+		+	
Dientamoeba	±		+	+	
Balantidium	+				
Microsporidia	–			+	Special fecal stains, tissue biopsies

Abbreviations: DFA, direct immunofluorescence assay; NAATs, nucleic acid amplification tests; O+P, conventional ova and parasites.

Consequent lactose intolerance and, in a minority of infected adults and children, significant malabsorption are clinical signs of the loss of epithelial brush-border enzyme activities. In most infections, the morphology of the bowel is unaltered; however, in chronically infected, symptomatic patients, the histopathologic findings (including flattened villi) and the clinical manifestations at times resemble those of tropical sprue and gluten-sensitive enteropathy. The pathogenesis of diarrhea in giardiasis is not known.

The natural history of *Giardia* infection varies markedly. Infections may be asymptomatic, transient, recurrent, or chronic. *G. duodenalis* parasites vary genotypically, and such variations might contribute to different courses of infection. Parasite as well as host factors may be important in determining the course of infection and disease. Both cellular and humoral responses develop in human infections, but their precise roles in disease pathogenesis and/or control of infection are unknown. Because patients with hypogammaglobulinemia suffer from prolonged, severe infections that are poorly responsive to treatment, humoral immune responses appear to be important. The greater susceptibilities of the young than of the old and of newly exposed persons than of chronically exposed populations suggest that at least partial protective immunity may develop.

Clinical Manifestations Disease manifestations of giardiasis range from asymptomatic carriage to fulminant diarrhea and malabsorption. Most infected persons are asymptomatic, but in epidemics, the proportion of symptomatic cases may be higher. Symptoms may develop suddenly or gradually. In persons with acute giardiasis, symptoms develop after an incubation period that lasts at least 5–6 days and usually 1–3 weeks. Prominent early symptoms include diarrhea, abdominal pain, bloating, belching, flatus, nausea, and vomiting. Although diarrhea is common, upper intestinal manifestations such as nausea, vomiting, bloating, and abdominal pain may predominate. The duration of acute giardiasis is usually >1 week, although diarrhea often subsides. Individuals with chronic giardiasis may present with or without having experienced an antecedent acute symptomatic episode. Diarrhea is not necessarily prominent, but increased flatus, loose stools, sulfurous belching, and (in some instances) weight loss occur. Symptoms may be continual or episodic and may persist for years. Some persons who have relatively mild symptoms for long periods recognize the extent of their discomfort only in retrospect. Fever, the presence of blood and/or mucus in the stools, and other signs and symptoms of colitis are uncommon and suggest a different diagnosis or a concomitant illness. Symptoms tend to be intermittent yet recurring and gradually debilitating, in contrast with the acute disabling symptoms associated with many enteric bacterial infections. Because of the less severe illness early on and the propensity for chronic infections, patients may seek medical advice late in the course of the illness; however, disease can be severe, resulting in malabsorption, weight loss, growth retardation in children, and dehydration. A number of extraintestinal manifestations have been described, such as urticaria, anterior uveitis, and arthritis; whether these are caused by giardiasis or concomitant processes is unclear.

Giardiasis can be severe in patients with hypogammaglobulinemia and can complicate other preexisting intestinal diseases, such as that occurring in cystic fibrosis. In patients with AIDS, *Giardia* can cause enteric illness that is refractory to treatment.

Diagnosis (Table 236-1) Giardiasis is diagnosed by detection of parasite antigens in the feces, by identification of cysts in the feces or of trophozoites in the feces or small intestines, or by nucleic acid amplification tests (NAATs). Cysts are oval and measure 8–12 μm × 7–10 μm. Mature cysts characteristically contain four nuclei. Trophozoites are pear-shaped, dorsally convex, flattened parasites with two nuclei and four pairs of flagella (Fig. 236-2). The diagnosis is sometimes difficult to establish. Direct examination of fresh or properly preserved stools as well as concentration methods can be used. Because cyst excretion is variable and may be undetectable at times, repeated examination of stool, sampling of duodenal fluid, and biopsy of the small intestine may be required to detect the parasite. Tests for parasitic antigens in stool are often more sensitive than or comparable to microscopic examination, are easier to perform, have a more rapid turn-around time, are often cost-comparable, and are often the first-line test in many institutions. NAATs are also highly sensitive.

TREATMENT

Giardiasis

Cure rates with metronidazole (250–500 mg thrice daily for 5 days) are usually >90%. Tinidazole (2 g once by mouth) may be more effective than metronidazole. Nitazoxanide (500 mg twice daily for 3 days) is an alternative agent for treatment of giardiasis. Paromomycin, an oral aminoglycoside that is not well absorbed, can be given to symptomatic pregnant patients, although information is limited on how effectively this agent eradicates infection. Albendazole and mebendazole also can be used as alternative agents, although their efficacy is lower than that of single-dose tinidazole. Quinacrine, secnidazole, ornidazole, and furazolidone are other alternative agents that have demonstrated some efficacy.

Almost all patients respond to therapy and are cured, although some with chronic giardiasis experience delayed resolution of symptoms after eradication of *Giardia*. For many of the latter patients, residual symptoms probably reflect delayed regeneration of intestinal brush-border enzymes. Continued infection should be documented by stool examinations before treatment is repeated. Patients who remain infected after repeated treatments should be evaluated for reinfection through family members, close personal contacts, and environmental sources as well as for hypogammaglobulinemia. In cases refractory to multiple treatment courses, prolonged therapy with metronidazole (750 mg thrice daily for 21 days) or therapy with varied combinations of multiple agents has been successful.

Prevention Giardiasis can be prevented by consumption of uncontaminated food and water and by personal hygiene during the provision of care for infected children. Boiling or filtering potentially contaminated water prevents infection.

■ CRYPTOSPORIDIOSIS

The coccidian parasite *Cryptosporidium* causes diarrheal disease that is self-limited in immunocompetent human hosts but can be severe in persons with AIDS or other forms of immunodeficiency. Two species of *Cryptosporidium*, *C. hominis* and *C. parvum*, cause most human infections.

Life Cycle and Epidemiology *Cryptosporidium* species are widely distributed in the world. In resource-constrained areas, cryptosporidiosis is the second most common cause of moderate to severe diarrhea during the first two years of life (following rotavirus). Cryptosporidiosis is acquired by the consumption of oocysts (50% infectious dose: ~10–100 *C. hominis/parvum* oocysts in nonimmune individuals), which excyst to liberate sporozoites that in turn enter and infect intestinal epithelial cells. The parasite's further development involves both asexual and sexual cycles, which produce forms capable of infecting other epithelial cells and of generating oocysts that are passed in the feces. *Cryptosporidium* species infect a number of animals, and *C. parvum* can spread from infected animals to humans. Since oocysts are immediately infectious when passed in feces, person-to-person transmission takes place in day-care centers and among household contacts and medical providers. Waterborne transmission (especially that of *C. hominis*) accounts for infections in travelers and for common-source epidemics. Oocysts are quite hardy and resist killing by routine chlorination. Both drinking water and recreational water (e.g., pools, waterslides) have been increasingly recognized as sources of infection.

Pathophysiology Although intestinal epithelial cells harbor cryptosporidia in an intracellular vacuole, the means by which secretory diarrhea is elicited remain uncertain. No characteristic pathologic changes are found by biopsy. The distribution of infection can be spotty within the principal site of infection, the small bowel. Cryptosporidia are found in the pharynx, stomach, and large bowel of some patients and at times in the respiratory tract. Especially in patients with AIDS, involvement of the biliary tract can cause papillary stenosis, sclerosing cholangitis, or cholecystitis.

Clinical Manifestations Asymptomatic infections can occur in both immunocompetent and immunocompromised hosts. In immunocompetent persons, symptoms develop after an incubation period of ~1 week and consist principally of watery nonbloody diarrhea, sometimes in conjunction with abdominal pain, nausea, anorexia, fever, and/or weight loss. In these hosts, the illness usually subsides after 1–2 weeks. In contrast, in immunocompromised hosts (especially those with AIDS and CD4+ T-cell counts <100/μL), diarrhea can be chronic, persistent, and remarkably profuse, causing clinically significant fluid and electrolyte depletion. Stool volumes may range from 1 to 25 L/d. Weight loss, wasting, and abdominal pain may be severe. Biliary tract involvement can manifest as mid-epigastric or right-upper-quadrant pain.

Diagnosis (Table 236-1) Evaluation starts with fecal examination for small oocysts, which are smaller (4–5 μm in diameter) than the fecal stages of most other parasites. Because conventional stool examination for ova and parasites (O+P) does not detect *Cryptosporidium*, specific testing must be requested. Detection is enhanced by evaluation of stools (obtained on multiple days) by several techniques, including modified acid-fast and direct immunofluorescent stains and enzyme immunoassays. NAATs also are useful. Cryptosporidia can also be identified by light and electron microscopy at the apical surfaces of intestinal epithelium from biopsy specimens of the small bowel and, less frequently, the large bowel.

TREATMENT

Cryptosporidiosis

Nitazoxanide, approved by the U.S. Food and Drug Administration (FDA) for the treatment of cryptosporidiosis, is available in tablet form for adults (500 mg twice daily for 3 days) and as an elixir for children. This agent has not been effective for the treatment of immunosuppressed patients or HIV-infected patients, in whom improved immune status due to antiretroviral therapy can lead to amelioration of cryptosporidiosis. Otherwise, treatment includes supportive care with replacement of fluids and electrolytes and administration of antidiarrheal agents. Biliary tract obstruction may require papillotomy or T-tube placement. Prevention requires minimizing exposure to infectious oocysts in human or animal feces. Use of submicron water filters may minimize acquisition of infection from drinking water.

CYSTOISOSPORIASIS

The coccidian parasite *Cystoisospora belli* (formerly referred to as *Isospora belli*) causes human intestinal disease. Infection is acquired by the consumption of oocysts, after which the parasite invades intestinal epithelial cells and undergoes both sexual and asexual cycles of development. Oocysts excreted in stool are not immediately infectious but must undergo further maturation.

Although *C. belli* infects many animals, little is known about the epidemiology or prevalence of this parasite in humans. It is most common in tropical and subtropical countries. Acute infections can begin abruptly with fever, abdominal pain, and watery nonbloody diarrhea and can last for weeks or months. In patients who have AIDS or are immunocompromised for other reasons, infections often are not self-limited but rather resemble cryptosporidiosis, with chronic, profuse watery diarrhea. Eosinophilia, which is not found in other enteric protozoan infections, may be detectable in immunocompetent hosts. The diagnosis (Table 236-1) is usually made by detection of the large (~25 μm) oocysts in stool by modified acid-fast staining. Oocyst excretion may be low-level and intermittent; if repeated stool examinations are unrevealing, sampling of duodenal contents by aspiration or small-bowel biopsy (often with electron microscopic examination) may be necessary. NAATs are effective newer diagnostic tools.

TREATMENT

Cystoisosporiasis

Trimethoprim-sulfamethoxazole (TMP-SMX, 160/800 mg two times daily for 7–10 days; and, for HIV-infected patients, then continuing three times daily for 3–4 weeks) is effective. For patients intolerant of sulfonamides, pyrimethamine (50–75 mg/d) or ciprofloxacin 500 mg two times a day can be used. Relapses can occur in persons with AIDS and necessitate maintenance therapy with TMP-SMX (160/800 mg three times per week).

CYCLOSPORIASIS

Cyclospora cayetanensis, a cause of diarrheal illness, is globally distributed: illness due to *C. cayetanensis* has been reported in the United States, Asia, Africa, Latin America, and Europe. The epidemiology of this parasite has not yet been fully defined, but waterborne transmission and foodborne transmission (e.g., by basil, sweet peas, and imported raspberries) have been recognized. The full spectrum of illness attributable to *Cyclospora* has not been delineated. Some infected patients may be without symptoms, but many have diarrhea, flulike symptoms, and flatulence and belching. The illness can be self-limited, can wax and wane, or, in many cases, can involve prolonged diarrhea, anorexia, and upper gastrointestinal symptoms, with sustained fatigue and weight loss in some instances. Diarrheal illness may persist for >1 month. *Cyclospora* can cause enteric illness in patients infected with HIV.

The parasite is detectable in epithelial cells of small-bowel biopsy samples and elicits secretory diarrhea by unknown means. The absence of fecal blood and leukocytes indicates that disease due to *Cyclospora* is not caused by destruction of the small-bowel mucosa. The diagnosis (Table 236-1) can be made by detection of spherical 8- to 10-μm oocysts in the stool, although routine stool O+P examinations are not sufficient. Specific fecal examinations must be requested to detect the oocysts, which are variably acid-fast and are fluorescent when viewed with ultraviolet light microscopy. NAATs are sensitive. Cyclosporiasis should be considered in the differential diagnosis of prolonged diarrhea, with or without a history of travel by the patient to other countries.

TREATMENT

Cyclosporiasis

Cyclosporiasis is treated with TMP-SMX (160/800 mg twice daily for 7–10 days). HIV-infected patients may require longer treatment courses, may experience relapses after treatment, and thus may require longer-term suppressive maintenance therapy. Nitazoxanide and ciprofloxacin may be alternative agents in patients unable to receive TMP-SMX.

MICROSPORIDIOSIS

Microsporidia are obligate intracellular spore-forming protozoa that infect many animals and cause disease in humans, especially as opportunistic pathogens in AIDS. Microsporidia are members of a distinct phylum, Microspora, which contains hundreds of genera and thousands of species. The various microsporidia are differentiated by their developmental life cycles, ultrastructural features, and molecular taxonomy based on ribosomal RNA. The complex life cycles of the organisms result in the production of infectious spores **(Fig. 236-3)**. Currently, at least 17 species of microsporidia are recognized as causes of human disease, most commonly of the genera *Enterocytozoon* and *Encephalitozoon*. Although some microsporidia are probably prevalent causes of self-limited or asymptomatic infections in immunocompetent patients, little is known about how microsporidiosis is acquired.

Microsporidiosis is most common among patients with AIDS, less common among patients with other types of immunocompromise, and rare among immunocompetent hosts. In patients with AIDS, intestinal infections with *Enterocytozoon bieneusi* and *Encephalitozoon intestinalis* are recognized to contribute to chronic diarrhea and wasting; these infections have been found in 10–40% of patients with chronic diarrhea. Both organisms have been found in the biliary tracts of patients with cholecystitis. *E. intestinalis* may also disseminate to cause fever, diarrhea, sinusitis, cholangitis, and bronchiolitis. In patients with AIDS, *Encephalitozoon hellem* has caused superficial keratoconjunctivitis as well as sinusitis, respiratory tract disease, and disseminated

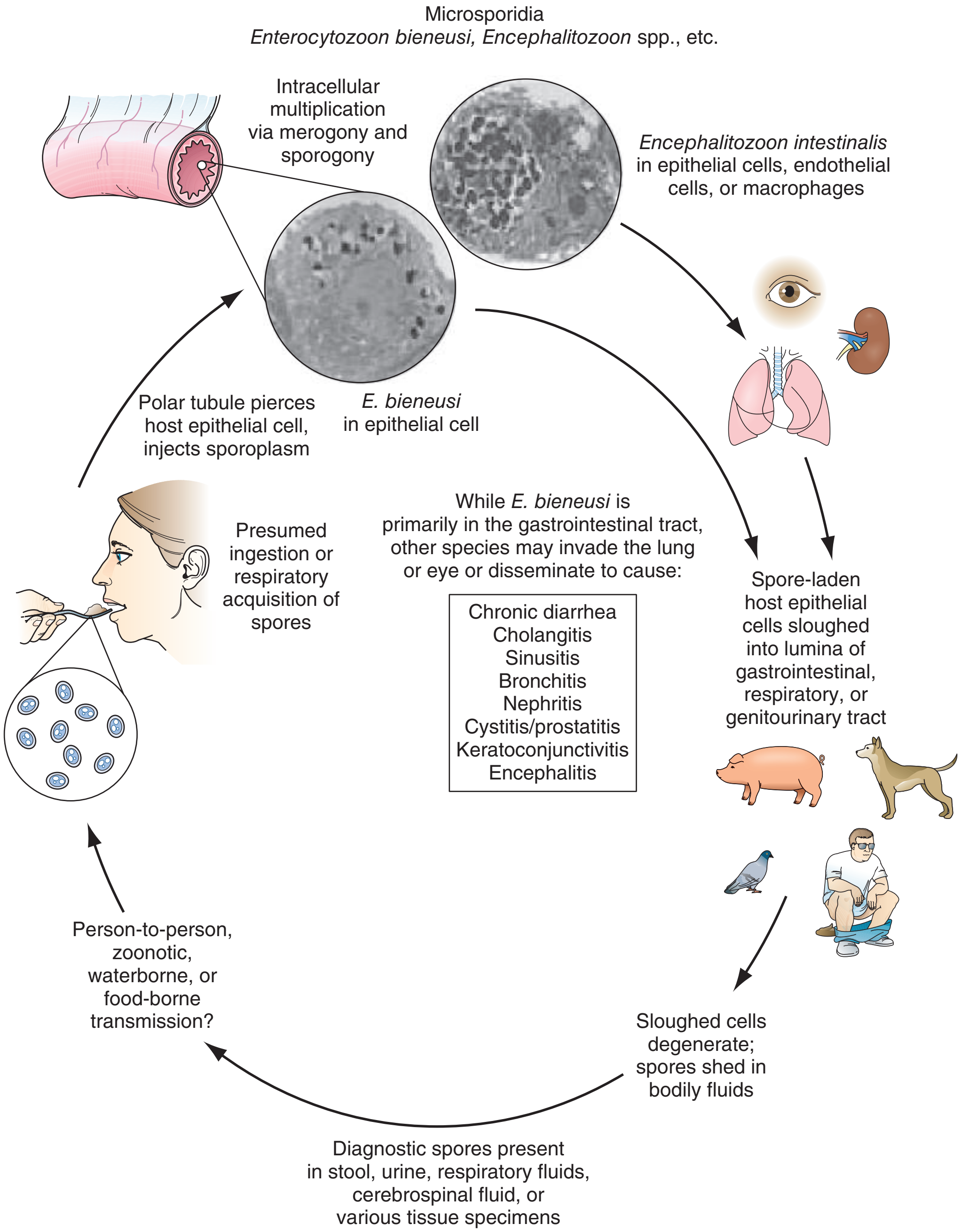
Microsporidia
Enterocytozoon bieneusi, Encephalitozoon spp., etc.
Intracellular multiplication via merogony and sporogony
Encephalitozoon intestinalis in epithelial cells, endothelial cells, or macrophages
Polar tubule pierces host epithelial cell, injects sporoplasm
E. bieneusi in epithelial cell
Presumed ingestion or respiratory acquisition of spores
While *E. bieneusi* is primarily in the gastrointestinal tract, other species may invade the lung or eye or disseminate to cause:
Chronic diarrhea
Cholangitis
Sinusitis
Bronchitis
Nephritis
Cystitis/prostatitis
Keratoconjunctivitis
Encephalitis
Spore-laden host epithelial cells sloughed into lumina of gastrointestinal, respiratory, or genitourinary tract
Person-to-person, zoonotic, waterborne, or food-borne transmission?
Sloughed cells degenerate; spores shed in bodily fluids
Diagnostic spores present in stool, urine, respiratory fluids, cerebrospinal fluid, or various tissue specimens

FIGURE 236-3 Life cycle of microsporidia. *(Reproduced with permission from RL Guerrant et al [eds]: Tropical Infectious Diseases: Principles, Pathogens and Practice, 2nd ed. Elsevier 2006.)*

infection. Myositis due to *Pleistophora* has been documented. *Nosema*, *Vittaforma*, and *Microsporidium* have caused stromal keratitis associated with trauma in immunocompetent patients.

Microsporidia are small gram-positive organisms with mature spores measuring 0.5–2 μm × 1–4 μm. Diagnosis of microsporidial infections in tissue often requires electron microscopy, although intracellular spores can be visualized by light microscopy with hematoxylin and eosin, Giemsa, or tissue Gram's stain. Light microscopy with a modified trichrome stain or fluorescent assays can be used on stool, tissue, or urine. NAATs are useful for diagnosis and speciation. Definitive therapies for microsporidial infections remain to be established. For superficial keratoconjunctivitis due to *E. hellem*, *E. cuniculi*, *E. intestinalis*, and *E. bieneusi*, topical therapy with fumagillin suspension has shown promise **(Chap. 229)**. For enteric infections with *E. intestinalis* in HIV-infected patients, therapy with albendazole may be efficacious **(Chap. 229)**.

OTHER INTESTINAL PROTOZOA

Balantidiasis *Balantidium coli* is a large ciliated protozoal parasite that can produce a spectrum of large-intestinal disease analogous to amebiasis. The parasite is widely distributed in the world. Since it infects pigs, cases in humans are more common where pigs are raised. Infective cysts can be transmitted from person to person and through water, but many cases are due to the ingestion of cysts derived from porcine feces in association with slaughtering, with use of pig feces for fertilizer, or with contamination of water supplies by pig feces.

Ingested cysts liberate trophozoites, which reside and replicate in the large bowel. Many patients remain asymptomatic, but some have persisting intermittent diarrhea, and a few develop more fulminant dysentery. In symptomatic individuals, the pathology in the bowel—both gross and microscopic—is similar to that seen in amebiasis, with varying degrees of mucosal invasion, focal necrosis, and ulceration. Balantidiasis, unlike amebiasis, only rarely spreads hematogenously to other organs. The diagnosis is made by detection of the parasite in stool or sampled colonic tissue. Tetracycline (500 mg four times daily for 10 days) is an effective therapeutic agent. Metronidazole is an alternative agent.

Blastocystosis *Blastocystis hominis* is an organism of uncertain pathogenicity. Some patients who pass *B. hominis* in their stools are asymptomatic, whereas others have diarrhea and associated intestinal symptoms. Diligent evaluation reveals other potential bacterial, viral, or protozoal causes of diarrhea in some but not all patients with symptoms. Because the pathogenicity of *B. hominis* is uncertain and because therapy for *Blastocystis* infection is neither specific nor uniformly effective, patients with prominent intestinal symptoms should be fully evaluated for other infectious causes of diarrhea. If diarrheal symptoms associated with *Blastocystis* are prominent, metronidazole (500–750 mg thrice daily for 5–10 days) or tinidazole (2 g once) can be tried. Alternative agents include paromomycin, TMP-SMX, and nitazoxanide.

Dientamoebiasis *Dientamoeba fragilis* is unique among intestinal protozoa in that it has a trophozoite stage but not a cyst stage. How trophozoites survive to transmit infection is not known. When symptoms develop in patients with *D. fragilis* infection, they are generally mild and include intermittent diarrhea, abdominal pain, and anorexia. The diagnosis is made by the detection of trophozoites in stool; the lability of these forms accounts for the greater yield when fecal samples are preserved immediately after collection. NAATs are more sensitive than fecal microscopy. Paromomycin (25–35 mg/kg per day in three doses for 7 days) or metronidazole (500–750 mg three times daily for 10 days) is appropriate for treatment. Tetracycline, doxycycline, tinidazole, secnidazole, and ornidazole are alternative agents.

Sarcocystosis Humans can be the definitive or dead-end intermediate host for *Sarcocystis* spp. sarcocystoses are zoonotic infections. Humans can act as definitive hosts (the host in which a parasite completes its sexual life cycle) following consumption of undercooked beef or pork that contain the protozoa *Sarcocystis hominis* or *S. suihominis*, respectively. In this setting, humans can develop intestinal symptoms including nausea, vomiting, and diarrhea that is often self-resolving within a few days. Human intestinal infection can result in transient shedding of infectious oocysts that are consumed by cattle or pigs in the environment, and in these hosts the protozoa encyst in muscle tissue, allowing completion of the life cycle when these tissues are consumed by humans. Separately, humans can develop a transient myositis with peripheral eosinophilia following ingestion of oocysts of *S. nesbitti* shed by snakes. Intestinal sarcocystosis can be detected by stool examination, although oocysts are not usually detectable during the diarrheal phase. Intestinal sarcocystosis is self-limited and treated supportively.

TRICHOMONIASIS

Various species of trichomonads can be found in the mouth (in association with periodontitis) and occasionally in the gastrointestinal tract. *Trichomonas vaginalis*—one of the most prevalent protozoal parasites in the United States—is a pathogen of the genitourinary tract and a major cause of symptomatic vaginitis **(Chap. 141)**.

Life Cycle and Epidemiology *T. vaginalis* is a pear-shaped, actively motile organism that measures about 10 × 7 μm, replicates by binary fission, and inhabits the lower genital tract of females and the urethra and prostate of males. There are 2–3 million infections each year in the United States. While the organism can survive for a few hours in moist environments and could be acquired by direct contact, person-to-person venereal transmission accounts for virtually all cases of trichomoniasis. Its prevalence is greatest among persons with multiple sexual partners and among those with other sexually transmitted diseases **(Chap. 141)**.

Clinical Manifestations Many men infected with *T. vaginalis* are asymptomatic, although some develop urethritis and a few have epididymitis or prostatitis. In contrast, infection in women, which has an incubation period of 5–28 days, is more often symptomatic and manifests with malodorous vaginal discharge (often yellow), vulvar erythema and itching, dysuria or urinary frequency (in 30–50% of patients), and dyspareunia. These manifestations, however, do not clearly distinguish trichomoniasis from other types of infectious vaginitis.

Diagnosis Detection of motile trichomonads by microscopic examination of wet mounts of vaginal or prostatic secretions has been the conventional means of diagnosis. Although this approach provides an immediate diagnosis, its sensitivity for the detection of *T. vaginalis* is only ~50–60% in routine evaluations of vaginal secretions. Direct immunofluorescent antibody staining is more sensitive (70–90%) than wet-mount examinations. *T. vaginalis* can be recovered from the urethra of both males and females and is detectable in males after prostatic massage. NAATs are highly sensitive and specific for urine and for endocervical and vaginal swabs from women and often considered first-line diagnostics.

TREATMENT

Trichomoniasis

Metronidazole (either a single 2-g dose or 500-mg doses twice daily for 7 days), tinidazole (a single 2-g dose) or secnidazole (a single 2-g dose) is effective. All sexual partners must be treated concurrently to prevent reinfection, especially from asymptomatic males. In males with persistent symptomatic urethritis after therapy for nongonococcal urethritis, metronidazole therapy should be considered for possible trichomoniasis. Alternatives to metronidazole for treatment during pregnancy are not readily available. Reinfection often accounts for apparent treatment failures, but strains of *T. vaginalis* exhibiting high-level resistance to metronidazole have been encountered. Treatment of these resistant infections with higher oral doses, parenteral doses, or concurrent oral and vaginal doses of metronidazole or with tinidazole has been successful.

FURTHER READING

Bourli P et al: Waterborne transmission of protozoan parasites: A review of worldwide outbreaks–an update 2017–2022. J Water Health 21:1421, 2023.

Buret AG et al: Update on *Giardia*: Highlights from the Seventh International *Giardia* and *Cryptosporidim* Conference. Parasite 27:49, 2020.

Carter BL et al: Health sequelae of human cryptosporidiosis in industrialized countries: A systematic review. Parasit Vectors 13:443, 2020.

Coffey CM et al: Evolving epidemiology of reported giardiasis cases in the United States, 1995–2016. Clin Infect Dis 72:764, 2021.

Fitri LE et al: Diagnostic methods of common intestinal protozoa: Current and future immunological and molecular methods. Trop Med Infect Dis 7:253, 2022.

Hemphill A et al: Comparative pathobiology of the intestinal protozoan parasites *Giardia lamblia*, *Entamoeba histolytica* and *Cryptosporidium parvum*. Pathogens 8:116, 2019.

Mathison BA, Pritt BS: Cyclosporiasis—updates on clinical presentation, pathology, clinical diagnosis, and treatment. Microorganisms 9:1863, 2021.

Ramanan P, Pritt BS: Extraintestinal microsporidiosis. J Clin Microbiol 52:3839, 2014.

Van Den Broucke S et al: Microscopic detection of intestinal *Sarcocystis* infection diagnosed in international travelers at the Institute of Tropical Medicine, Antwerp, Belgium, from 2001 to 2020. Am J Trop Med Hyg 109:327, 2023.

Van German TO, Muzny CA: Recent advances in the epidemiology, diagnosis, and management of *Trichomonas vaginalis* infection. F1000Res 8:1666, 2019.

Van Gerwen OT et al: Trichomoniasis. Infect Dis Clin North Am 37: 245, 2023.

Van Gestel RSFE et al: A clinical guideline on *Dientamoeba fragilis* infections. Parasitology 146:1131, 2018.

Widmer G et al: Update on *Cryptosporidium* spp: Highlights from the Seventh International *Giardia* and *Cryptosporidim* Conference. Parasite 27:14, 2020.

Section 19 Helminthic Infections

237 Introduction to Helminthic Infections

Peter F. Weller, Edward T. Ryan

The word *helminth* is derived from the Greek *helmins* ("parasitic worm"). Helminthic worms are highly prevalent and, depending on the species, may exist as free-living organisms or as parasites of plant or animal hosts. The parasitic helminths have co-evolved with specific mammalian and other host species. Accordingly, most helminthic infections are restricted to nonhuman hosts, and only rarely do these zoonotic helminths accidentally cause human infections.

Helminthic parasites of humans belong to two phyla: Nemathelminthes, which includes nematodes (roundworms), and Platyhelminthes, which includes cestodes (tapeworms) and trematodes (flukes). Helminthic parasites of humans reside within the human body and hence are the cause of true infections. In contrast, parasites of other genera that reside only on mucocutaneous surfaces of humans (e.g., the parasites causing myiasis and scabies) are considered to represent infestations rather than infections.

Helminthic parasites differ substantially from protozoan parasites in several respects. First, protozoan parasites are unicellular organisms, whereas helminthic parasites are multicellular worms that possess differentiated organ systems. Second, helminthic parasites have complex life cycles that require sequential stages of development outside the human host. Thus, most helminths do not complete their replication within the human host; rather, they develop to a certain stage within the mammalian host and, as part of their obligatory life cycle, must mature further outside that host. During the "extra-human" stages of their life cycle, helminths exist either as free-living organisms or as parasites within another host species and thereafter mature into new developmental stages capable of infecting humans. Thus, with only two exceptions (*Strongyloides stercoralis* and *Capillaria philippinensis*, which are capable of internal human reinfections), increases in the number of adult helminths (i.e., the "worm burden") within the human host require repeated exogenous reinfections. In the case of protozoan parasites, a brief, even singular exposure (e.g., a single mosquito bite transmitting malaria) may lead rapidly to intense parasite loads and overwhelming infections; in contrast, for all but the two helminths noted above, increases in worm burden require multiple and usually ongoing exposures to infectious forms, such as ingestion of eggs of intestinal helminths or waterborne exposures to infectious cercariae of *Schistosoma mansoni*. This requirement is germane both to the consideration of helminthic infections in individuals and to ongoing global efforts to interrupt and/or minimize the acquisition of helminthic infections by humans.

Third, helminthic infections have a predilection toward stimulation of host immune responses that elicit eosinophilia within human tissues and blood. The many protozoan infections characteristically do not elicit eosinophilia in infected humans, with only three exceptions (two intestinal protozoan parasites, *Cystoisospora belli* and *Dientamoeba fragilis*, and tissue-borne *Sarcocystis* species). The magnitude of helminth-elicited eosinophilia tends to correlate with the extent of tissue invasion by larvae or adult helminths. For example, in several helminthic infections, including acute schistosomiasis (Katayama syndrome), paragonimiasis, and hookworm and *Ascaris* infections, eosinophilia is most pronounced during the early phases of infection, when migrations of infecting larvae and progression of subsequent developmental stages through the tissues are greatest. In established infections, local eosinophilia is often present around helminths in tissues, but blood eosinophilia may be intermittent, mild, or absent. In helminthic infections in which parasites are well contained within tissues (e.g., echinococcal cysts) or confined within the lumen of the intestinal tract (e.g., adult *Ascaris* or tapeworms), eosinophilia is usually absent.

NEMATODES

Nematodes are nonsegmented roundworms. Species of nematodes are remarkably diverse and abundant in nature. Among the many thousands of nematode species, few are parasites of humans. Most nematodes are free-living, and these species have variably evolved to survive in diverse ecologic niches, including saltwater, freshwater, or soil. The well-studied organism *Caenorhabditis elegans* is a free-living nematode. Nematodes can be either beneficial or deleterious parasites of plants. Parasitic nematodes have co-evolved with specific mammalian hosts and have no capacity to live their full life cycles in other hosts. Uncommonly, humans are exposed to infectious stages of nonhuman nematode parasites, and the resultant zoonotic nematode infections can elicit inflammatory and immune responses as larval forms migrate and die in the unsuitable human host. Examples include pulmonary coin lesions due to mosquito-transmitted infections with the dog heartworm *Dirofilaria immitis*; eosinophilic meningoencephalitis due to ingested eggs of the raccoon ascarid *Baylisascaris procyonis*; and eosinophilic meningitis due to ingestion of larvae of the rat lungworm *Angiostrongylus cantonensis*.

Nematode parasites of humans include worms that reside in the intestinal tract or localize in extraintestinal vascular or tissue sites. Roundworms are sexually dimorphic, with separate male and female forms that copulate to produce offspring (except for *S. stercoralis*, whose adult females are hermaphroditic in the human intestinal tract). Depending on the species, fertilized females release either larvae or eggs containing larvae. Nematodes have five developmental stages: an adult stage and four sequential larval stages. These parasites characteristically are surrounded by a durable outer cuticular layer. Nematodes have a nervous system; a muscular system, including muscle cells under the cuticle; and a developed intestinal tract, including an oral cavity and an elongated gut that ends in an anal pore. Adults may range in size from minute to >1 meter in length (with *Dracunculus medinensis*, for example, at the long end of this spectrum).

Humans acquire infections with nematode parasites by various routes, depending on the parasitic species. Ingestion of eggs passed in human feces is a major global health problem with many of the intestinal helminths (e.g., *Ascaris lumbricoides*). In other species, infecting larvae penetrate skin exposed to fecally contaminated soil (e.g., *S. stercoralis*, hookworms) or traverse the skin after the bite of infected insect vectors (e.g., filariae). Some nematode infections are acquired by consumption of specific animal-derived foods (e.g., trichinellosis from raw or undercooked pork or wild carnivorous mammals). As noted above, only two nematodes, *S. stercoralis* and *C. philippinensis*, can internally reinfect humans; thus, for all other nematodes, any increases in worm burden must be due to continued exogenous reinfections.

■ CESTODES

Tapeworms are the cestode parasites of humans. Adult tapeworms are elongated, segmented, hermaphroditic flatworms that reside in the intestinal lumen or, in their larval forms, may live in extraintestinal tissues. Tapeworms include a head (*scolex*) and a number of attached segments (*proglottids*). The worms attach to the intestinal tract via their scolices, which may possess suckers, hooks, or grooves. The scolex is the site of formation of new proglottids. Tapeworms do not have a functional gut tract; rather, each tapeworm segment passively and actively obtains nutrients through its specialized surface tegument. Mature proglottids possess both male and female sex organs, but insemination usually occurs between adjacent proglottids. Fertilized proglottids release eggs that are passed in the feces. When ingested by an intermediate host, an egg releases an oncosphere that penetrates the gut and develops further in tissues as a cysticercus. Humans acquire infection by ingesting animal tissues that contain cysticerci, and the resultant tapeworms develop and reside in the proximal small bowel (e.g., *Taenia solium*, *T. saginata*). Alternatively, if humans ingest eggs of these cestodes that have been passed in human or animal feces, oncospheres develop and can cause space-occupying extraintestinal cystic lesions in tissues; examples include cysticercosis due to *T. solium* and hydatid disease due to species of *Echinococcus*.

■ TREMATODES

Trematodes of medical importance include blood flukes, intestinal flukes, and tissue flukes. Adult flukes are often leaf-shaped flatworms. Oral and/or ventral suckers help adult flukes maintain their positions in situ. Flukes have an oral cavity but no distal anal pore. Nutrients are obtained both through their integument and by ingestion into the blind intestinal tract. Flukes are hermaphroditic except for blood flukes (schistosomes), which are sexually dimorphic. Eggs are passed in human feces (*Fasciola*, *Fasciolopsis*, *Clonorchis*, *Schistosoma japonicum*, *S. mansoni*), urine (*Schistosoma haematobium*), or sputum and feces (*Paragonimus*). Expelled eggs release miracidia—usually in water—that infect specific snail species. Within snails, parasites multiply and cercariae are released. Depending on the species, cercariae can penetrate the skin (schistosomes) or can develop into metacercariae that can be ingested with plants (e.g., watercress for *Fasciola*) or with fish (*Clonorchis*) or crabs (*Paragonimus*).

■ CONCLUSION

Many of the so-called neglected tropical diseases are due to helminthic infections. The health impacts of many helminthic infections are varied and are based on the frequent need for repeated exposures to increase the worm burdens in infected humans. In global regions where exposures to specific helminths occur even in childhood (e.g., fecally derived intestinal nematodes, mosquito-transmitted filariae, or waterborne snail-transmitted schistosomes), the morbidities in infected individuals can include nutritional, developmental, cognitive, and functional impairments. Ongoing global mass-treatment programs are currently aimed at diminishing the local intensity of infection and prevalences of specific helminths and their consequent impacts on the health of local populations.

238

Trichinellosis and Other Tissue Nematode Infections

Peter F. Weller, Edward T. Ryan

Nematodes are elongated, symmetric roundworms. Parasitic nematodes of medical significance may be broadly classified as either predominantly intestinal or tissue nematodes. The intestinal nematodes are covered in **Chap. 239**; filarial tissue-dwelling nematodes and *Dracunculus medinensis* infection are both covered in **Chap. 240**. This chapter covers the tissue nematodes that cause trichinellosis, visceral and ocular larva migrans, cutaneous larva migrans, cerebral angiostrongyliasis, and gnathostomiasis. All of these zoonotic infections result from incidental exposure to infectious nematodes. The clinical symptoms of these infections are due largely to invasive larval stages that (except in the case of *Trichinella*) do not reach maturity in humans.

■ TRICHINELLOSIS

Trichinellosis develops after the ingestion of meat containing cysts of *Trichinella* (e.g., pork or other meat from a carnivore/omnivore). Although most infections are mild and asymptomatic, heavy infections can cause severe enteritis, periorbital edema, myositis, and (infrequently) death.

Life Cycle and Epidemiology At least nine species of *Trichinella* and 13 genotypes are recognized as causes of infection in humans. Two species are distributed worldwide: *T. spiralis*, which is found in a great variety of carnivorous and omnivorous animals, and *T. pseudospiralis*, which is found in mammals and birds. *T. nativa* is present in Arctic and subarctic regions and infects bears, foxes, and walruses; *T. nelsoni* is found in equatorial eastern Africa, where it is common among felid predators and scavengers such as hyenas and bush pigs; and *T. britovi* is found in Europe, western Africa, and western Asia among carnivores but not among domestic swine. *T. murrelli* is present in wild animals in North American and Japan. *T. papuae* is found in Papua New Guinea, Thailand, Taiwan, and Cambodia in domestic and feral pigs and in saltwater crocodiles and turtles. *T. zimbabwensis* is present in crocodiles in Tanzania. *T. patagoniensis* is found in cougars in South America.

After human consumption of trichinous meat, encysted larvae are liberated by digestive acid and proteases **(Fig. 238-1)**. The larvae invade the small-bowel mucosa and mature into adult worms. After ~1 week, female worms release newborn larvae that migrate via the circulation to striated muscle. The larvae of all species except *T. pseudospiralis*, *T. papuae*, and *T. zimbabwensis* then encyst by inducing a radical transformation in the muscle cell architecture. Host immune responses may help to expel intestinal adult worms but have few deleterious effects on muscle-dwelling larvae.

Human trichinellosis classically has been caused by the ingestion of infected pork products and thus can occur in almost any location where the meat of domestic or wild swine is eaten. Increasingly, human trichinellosis has also been acquired from the meat of other animals, including dogs (in parts of Asia and Africa), horses (in Italy and France), and bears and walruses (in northern regions). Although cattle and horses (being herbivores) are not natural hosts of *Trichinella*, beef and horse meat have been implicated in outbreaks when contaminated or adulterated with trichinous pork or perhaps following ingestion of fodder contaminated with meat parts from naturally infected animals. Approximately 10–20 cases of trichinellosis are reported annually in the United States, down from 400–500 cases per year in the 1940s, largely reflecting the impact of control programs targeting education, pig-raising, and the processing and freezing of pork. Consumption of wild game (especially bear and walrus meat) now accounts for the majority of cases reported each year in North America. Most mild cases probably remain undiagnosed.

Pathogenesis and Clinical Features Clinical symptoms of trichinellosis arise from the successive phases of parasite enteric invasion,

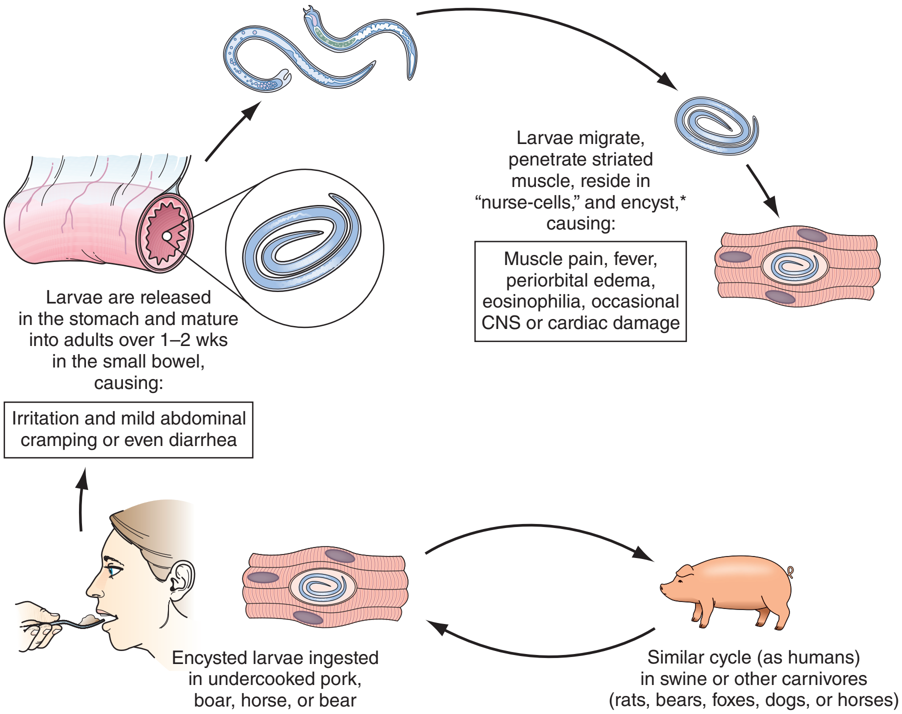
Larvae migrate,
penetrate striated
muscle, reside in
"nurse-cells," and encyst,*
causing:
Muscle pain, fever,
periorbital edema,
eosinophilia, occasional
CNS or cardiac damage
Larvae are released
in the stomach and mature
into adults over 1–2 wks
in the small bowel,
causing:
Irritation and mild abdominal
cramping or even diarrhea
Encysted larvae ingested
in undercooked pork,
boar, horse, or bear
Similar cycle (as humans)
in swine or other carnivores
(rats, bears, foxes, dogs, or horses)

**T. papuae, T. zimbabwensis,* and *T. pseudospiralis* do not encyst.

FIGURE 238-1 Life cycle of *Trichinella spiralis* (cosmopolitan); *nelsoni* (equatorial Africa); *britovi* (Europe, western Africa, western Asia); *nativa* (Arctic); *murrelli* (North America); *papuae* (Papua New Guinea); *zimbabwensis* (Tanzania); and *pseudospiralis* (cosmopolitan). CNS, central nervous system. *(Reproduced with permission from RL Guerrant et al [eds]: Tropical Infectious Diseases: Principles, Pathogens and Practice, 2nd ed. Elsevier, 2006.)*

larval migration, and muscle encystment (Fig. 238-1). Most light infections (those with <10 larvae per gram of muscle) are asymptomatic, whereas heavy infections (which can involve >50 larvae per gram of muscle) can be life-threatening. An initial enteric phase due to release of ingested muscle larvae may elicit diarrhea, abdominal pain, constipation, and nausea during the first weeks after infection.

Symptoms due to larval migration and muscle invasion begin to appear in the second week after infection. The migrating *Trichinella* larvae provoke a marked local and systemic hypersensitivity reaction, with fever and eosinophilia. Periorbital and facial edema is common, as are hemorrhages in the subconjunctivae, retina, and nail beds ("splinter" hemorrhages). A maculopapular rash, headache, cough, dyspnea, or dysphagia sometimes develops. Myocarditis with tachyarrhythmias or heart failure—and, less commonly, encephalitis or pneumonitis—may develop and accounts for most deaths of patients with trichinellosis.

Upon onset of larval encystment in muscle 2–3 weeks after infection, symptoms of myositis with myalgias, muscle edema, and weakness develop, usually overlapping with the inflammatory reactions to migrating larvae. The most commonly involved muscle groups include the extraocular muscles; the biceps; and the muscles of the jaw, neck, lower back, and diaphragm. Peaking ~3 weeks after infection, symptoms subside only gradually during a prolonged convalescence. Uncommon infections with *T. pseudospiralis*, whose larvae do not encapsulate in muscles, elicit a prolonged polymyositis-like illness.

Laboratory Findings and Diagnosis Blood eosinophilia develops in >90% of patients with symptomatic trichinellosis and may peak at a level of >50% 2–4 weeks after infection. Serum levels of muscle enzymes, including creatine phosphokinase, are elevated in most symptomatic patients. Patients should be questioned thoroughly about their consumption of pork or wild animal meat and about illness in other individuals who ate the same meat. A presumptive clinical diagnosis can be based on fevers, eosinophilia, periorbital edema, and myalgias after a suspect meal. A rise in the titer of parasite-specific antibody, which usually does not occur until after the third week of infection, confirms the diagnosis. Alternatively, a definitive diagnosis requires surgical biopsy of at least 1 g of involved muscle; the yields are highest near tendon insertions in symptomatic muscles. The fresh muscle tissue should be compressed between glass slides and examined microscopically **(Fig. 238-2)** because larvae may be missed by examination of routine histopathologic sections alone.

TREATMENT

Trichinellosis

Most lightly infected patients recover uneventfully with bed rest, antipyretics, and analgesics. Glucocorticoids like prednisone **(Table 238-1)** are beneficial for severe myositis and myocarditis. Mebendazole and albendazole are active against enteric stages of the parasite, but their efficacy against encysting/encysted larvae has not been conclusively demonstrated.

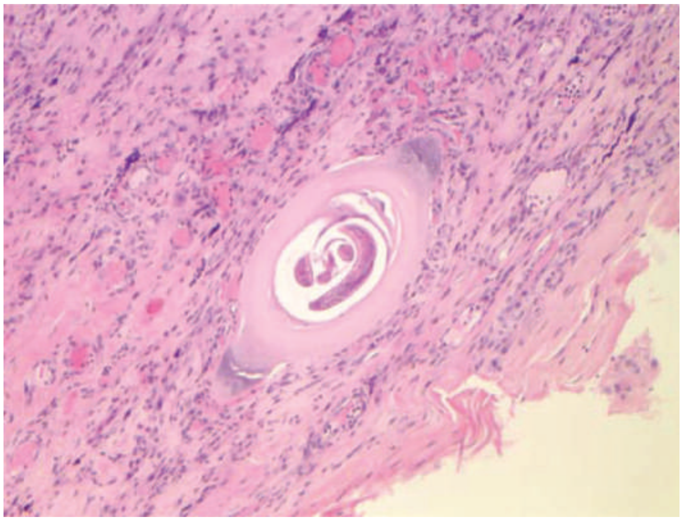

FIGURE 238-2 *Trichinella* larva encysted in a characteristic hyalinized capsule in striated muscle tissue. *(CDC/Wadsworth Center, New York State Department of Health.)*

TABLE 238-1 Therapy for Tissue Nematode Infections

INFECTION	SEVERITY	TREATMENT
Trichinellosis	Mild	Supportive
	Moderate	Albendazole (400 mg bid × 10–14 days) *or*
		Mebendazole (200–400 mg tid × 3 days, then 500 mg tid × 10 days)
	Severe	Add glucocorticoids (e.g., prednisone, 1 mg/kg qd × 5 days)
Visceral larva migrans	Mild to moderate	Supportive
	Severe	Glucocorticoids (as above)
	Ocular	Not fully defined; albendazole (800 mg bid for adults, 400 mg bid for children) with glucocorticoids × 5–20 days has been effective
Cutaneous larva migrans		Ivermectin (single dose, 200 μg/kg) *or*
		Albendazole (200 mg bid × 3 days)
Angiostrongyliasis	Mild to moderate	Supportive Not fully defined: albendazole (15 mg/kg per day in two divided doses × 14 days) with glucocorticoids (as above)
	Severe	Not fully defined: albendazole (15 mg/kg per day in two divided doses × 14 days) with glucocorticoids (as above)
Gnathostomiasis		Ivermectin (200 μg/kg per day × 2 days)[a] *or*
		Albendazole (400 mg bid × 21 days)[a]

[a]These agents may have efficacy for cutaneous gnathostomiasis; their use in neurologic and ocular gnathostomiasis is less certain and may be detrimental. Steroids are often given during neurologic and ocular gnathostomiasis.

Prevention Larvae are usually killed by cooking pork until it is no longer pink or by freezing it at −15°C for 3 weeks. However, Arctic *T. nativa* larvae in walrus or bear meat are relatively resistant and may remain viable despite freezing.

VISCERAL AND OCULAR LARVA MIGRANS

Visceral larva migrans is a syndrome caused by nematodes that are normally parasitic for nonhuman host species. In humans, these nematode larvae do not develop into adult worms but instead migrate through host tissues and elicit eosinophilic inflammation. The most common form of visceral larva migrans is toxocariasis due to larvae of the canine ascarid *Toxocara canis*; the syndrome is due less commonly to the feline ascarid *T. cati* and even less commonly to the pig ascarid *Ascaris suum*. Rare cases with eosinophilic meningoencephalitis have been caused by the raccoon ascarid *Baylisascaris procyonis*.

Life Cycle and Epidemiology The canine roundworm *T. canis* is distributed among dogs worldwide. Ingestion of infective eggs by dogs is followed by liberation of *Toxocara* larvae, which penetrate the gut wall and migrate intravascularly into canine tissues, where most remain in a developmentally arrested state. During pregnancy, some larvae resume migration in bitches and infect puppies prenatally (through transplacental transmission) or after birth (through suckling). Thus, in lactating bitches and puppies, larvae return to the intestinal tract and develop into adult worms, which produce eggs that are released in the feces. Eggs must undergo embryonation over several weeks to become infectious. Humans acquire toxocariasis mainly by ingesting soil/sand—such as found in public parks and playgrounds—contaminated by dog or cat feces that contains infective *T. canis* eggs.

Pathogenesis and Clinical Features Clinical disease most commonly afflicts preschool children. After humans ingest *Toxocara* eggs, the larvae hatch and penetrate the intestinal mucosa, from which they are carried by the circulation to a wide variety of organs and tissues. The larvae invade the liver, lungs, central nervous system (CNS), and other sites, provoking intense local eosinophilic granulomatous responses. The degree of clinical illness depends on larval number and tissue distribution, reinfection, and host immune responses. Most light infections are asymptomatic and may be evidenced only by blood eosinophilia. Characteristic symptoms of visceral larva migrans include fever, malaise, anorexia and weight loss, cough, wheezing, and rashes. Hepatosplenomegaly is common. These features may be accompanied by extraordinary peripheral eosinophilia at levels that may approach 90%. Uncommonly, seizures or behavioral disorders develop. Rare deaths are due to severe neurologic, pneumonic, or myocardial involvement.

The ocular form of the larva migrans syndrome occurs when *Toxocara* larvae invade the eye. An eosinophilic granulomatous mass, most commonly in the posterior pole of the retina, develops around the entrapped larva. The retinal lesion can mimic retinoblastoma in appearance, and mistaken diagnosis of the latter condition can lead to unnecessary enucleation. The spectrum of eye involvement also includes endophthalmitis, uveitis, and chorioretinitis. Unilateral visual disturbances, strabismus, and eye pain are the most common presenting symptoms. In contrast to visceral larva migrans, ocular toxocariasis usually develops in older children or young adults with no history of pica; these patients seldom have eosinophilia or visceral manifestations.

Diagnosis In addition to eosinophilia, leukocytosis and hypergammaglobulinemia may be evident in visceral larva migrans; these attributes are often absent in ocular larva migrans. Transient pulmonary infiltrates are apparent on chest x-rays of about one-half of patients with symptoms of pneumonitis. The clinical diagnosis can be confirmed by an enzyme-linked immunosorbent assay for toxocaral antibodies, although this assay may be negative in ocular larva migrans. Stool examination for parasite eggs is worthless in toxocariasis, since the larvae do not develop into egg-producing adults in humans.

TREATMENT

Visceral and Ocular Larva Migrans

The vast majority of *Toxocara* infections are self-limited and resolve without specific therapy. In patients with severe myocardial, CNS, or pulmonary involvement, glucocorticoids may be employed to reduce inflammatory complications. Available anthelmintic drugs, including mebendazole and albendazole, have not been shown conclusively to alter the course of larva migrans, although most individuals with moderate to severe visceral larva migrans receive albendazole with concomitant steroids (Table 238-1). Treatment of ocular disease is not fully defined, but the administration of steroids (to decrease ocular inflammation) and albendazole has been effective. Control measures include prohibiting dog excreta in public parks and playgrounds, deworming dogs, and preventing pica in children.

CUTANEOUS LARVA MIGRANS

Cutaneous larva migrans ("creeping eruption") is a serpiginous skin eruption caused by burrowing larvae of animal hookworms, usually the dog and cat hookworm *Ancylostoma braziliense*. The larvae hatch from eggs passed in dog and cat feces and mature in the soil. Humans become infected after skin contact with soil in areas frequented by dogs and cats. Cutaneous larva migrans is prevalent among children and travelers in regions with warm humid climates.

After larvae penetrate the skin, erythematous lesions form along the tortuous tracks of their migration along the dermal-epidermal junction; the larvae advance several centimeters in a day. The intensely pruritic lesions may occur anywhere on the body and can be numerous if the patient has lain on the ground. Vesicles and bullae may form. The animal hookworm larvae do not mature in humans and, without treatment, will die after an interval ranging from weeks to a couple of months, with resolution of skin lesions. The diagnosis is made on clinical grounds. Skin biopsies only rarely detect diagnostic larvae. Symptoms can be alleviated by ivermectin or albendazole (Table 238-1).

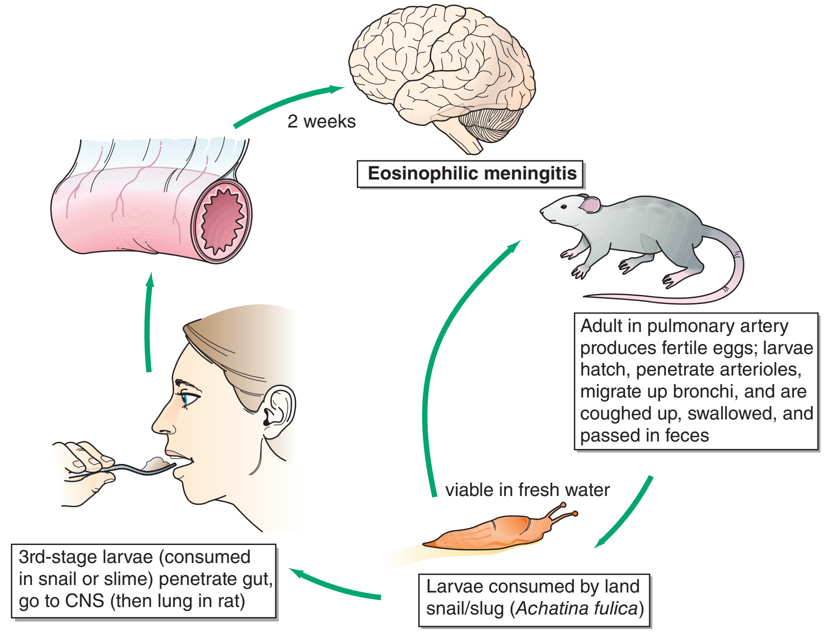
2 weeks
Eosinophilic meningitis
Adult in pulmonary artery produces fertile eggs; larvae hatch, penetrate arterioles, migrate up bronchi, and are coughed up, swallowed, and passed in feces
viable in fresh water
Larvae consumed by land snail/slug (*Achatina fulica*)
3rd-stage larvae (consumed in snail or slime) penetrate gut, go to CNS (then lung in rat)

FIGURE 238-3 Life cycle of *Angiostrongylus cantonensis* (rat lung worm) found in Southeast Asia and the Pacific Basin as well as on Caribbean islands, in countries of Central and South America, and in the southern United States. CNS, central nervous system. *(Reproduced with permission from RL Guerrant et al [eds]: Tropical Infectious Diseases: Principles, Pathogens and Practice, 2nd ed. Elsevier, 2006.)*

ANGIOSTRONGYLIASIS

Angiostrongylus cantonensis, the rat lungworm, is the most common cause of human eosinophilic meningitis (Fig. 238-3).

Life Cycle and Epidemiology This infection occurs principally in Southeast Asia and the Pacific Basin (including Hawaii) but has spread to other areas of the world, including the Caribbean islands, countries in Central and South America, and the southern United States. *A. cantonensis* larvae produced by adult worms in the rat lung migrate to the gastrointestinal tract and are expelled with the feces. They develop into infective larvae in land snails and slugs. Humans acquire the infection by ingesting raw infected mollusks; vegetables contaminated by mollusk slime; or crabs, freshwater shrimp, and certain marine fish that have themselves eaten infected mollusks. The larvae then migrate to the brain.

Pathogenesis and Clinical Features The parasites eventually die in the CNS, but not before initiating pathologic consequences that, in heavy infections, can result in permanent neurologic sequelae or death. Migrating larvae cause marked local eosinophilic inflammation and hemorrhage, with subsequent necrosis and granuloma formation around dying worms. Clinical symptoms develop 2–35 days after the ingestion of larvae, typically at about 2 weeks. Patients usually present with an insidious or abrupt excruciating frontal, occipital, or bitemporal headache. Neck stiffness, nausea and vomiting, and paresthesias are also common. Fever, cranial and extraocular nerve palsies, seizures, paralysis, and lethargy are uncommon.

Laboratory Findings Examination of cerebrospinal fluid (CSF) is mandatory in suspected cases and usually reveals an elevated opening pressure, a white blood cell count of 150–2000/μL, and an eosinophilic pleocytosis of >20%. The protein concentration is usually elevated and the glucose level normal. The larvae of *A. cantonensis* are only rarely seen in CSF. Peripheral-blood eosinophilia may be mild. The diagnosis is generally based on the clinical presentation of eosinophilic meningitis together with a compatible epidemiologic history. Serologic tests are sometimes available in endemic regions, although they have not been standardized and may be negative at onset of clinical symptoms. A polymerase chain reaction assay that detects *A. cantonensis* DNA in CSF or other tissues is available, including through the U.S. Centers for Disease Control and Prevention (CDC) and the Hawaii Department of Public Health in the United States.

TREATMENT

Angiostrongyliasis

Many individuals with eosinophilic meningitis from angiostrongyliasis improve with supportive therapy alone. Steroids are often given to decrease inflammation. Specific chemotherapy is of uncertain benefit and should never be given without steroids to prevent exacerbation of inflammatory brain lesions; when given, albendazole is usually employed. (Table 238-1). Repeated lumbar punctures with removal of CSF can assist with symptom relief. In most patients, cerebral angiostrongyliasis has a self-limited course, and recovery is complete. The infection may be prevented by adequately cooking snails, crabs, and prawns and inspecting vegetables for mollusk infestation. Other parasitic or fungal causes of eosinophilic meningitis in endemic areas may include gnathostomiasis (see below), paragonimiasis (Chap. 241), schistosomiasis (Chap. 241), neurocysticercosis (Chap. 242), and coccidioidomycosis (Chap. 219).

A separate species of *Angiostrongylus*, *A. costaricensis*, can inhabit the mesenteric arteries of humans, causing an eosinophilic inflammatory mass in abdominal viscera. *A. costaricensis* is a zoonosis (rat–mollusk/slug life cycle) found in the New World, especially Latin America, and occurs following ingestion of slugs or raw produce contaminated by slug slime. Diagnosis is most commonly made at the time of surgical excision, which is often curative.

GNATHOSTOMIASIS

Infection of human tissues with larvae of *Gnathostoma spinigerum* can cause eosinophilic meningoencephalitis, migratory cutaneous swellings, or invasive masses of the eye and visceral organs.

Life Cycle and Epidemiology Human gnathostomiasis occurs in many countries and is notably endemic in Southeast Asia and parts of China and Japan. In nature, the mature adult worms parasitize the gastrointestinal tract of dogs and cats. First-stage larvae hatch from eggs passed into water and are ingested by *Cyclops* species (water fleas). Infective third-stage larvae develop in the flesh of many animal species (including fish, frogs, eels, snakes, chickens, and ducks) that have eaten either infected *Cyclops* or another infected second intermediate host. Humans typically acquire the infection by eating raw or undercooked fish or poultry. Raw fish dishes, such as *som fak* in Thailand and *sashimi* in Japan, account for many cases of human gnathostomiasis. Some cases in Thailand result from the local practice of applying frog or snake flesh as a poultice.

Pathogenesis and Clinical Features Clinical symptoms are due to the aberrant migration of a single larva into cutaneous, visceral, neural, or ocular tissues. After invasion, larval migration may cause local inflammation, with pain, cough, or hematuria accompanied by fever and eosinophilia. Painful, itchy, migratory swellings may develop in the skin, particularly in the distal extremities or periorbital area. Cutaneous swellings usually last ~1 week but often recur intermittently over many years. Larval invasion of the eye can provoke a sight-threatening inflammatory response. Invasion of the CNS results in eosinophilic meningitis with myeloencephalitis, a serious complication due to ascending larval migration along a large nerve tract. Patients characteristically present with agonizing radicular pain and paresthesias in the trunk or a limb, which are followed shortly by paraplegia. Cerebral involvement, with focal hemorrhages and tissue destruction, is often fatal.

Diagnosis and Treatment Cutaneous migratory swellings with marked peripheral eosinophilia, supported by an appropriate geographic and dietary history, generally constitute an adequate basis for a clinical diagnosis of gnathostomiasis. However, patients may present with ocular or cerebrospinal involvement without antecedent cutaneous swellings. In the latter case, eosinophilic pleocytosis is demonstrable (usually along with hemorrhagic or xanthochromic CSF), but worms are almost never recovered from CSF. Computed tomography or magnetic resonance imaging of the brain during neuronal gnathostomiasis often demonstrates cerebral hemorrhage from the destructive migration of the parasite. Surgical removal of the parasite from subcutaneous or ocular tissue, though rarely feasible, is both diagnostic and therapeutic. Albendazole or ivermectin may be helpful, especially for cutaneous gnathostomiasis (Table 238-1). At present, cerebrospinal involvement is managed with supportive measures and generally with a course of glucocorticoids; the role of albendazole or ivermectin is uncertain and could be detrimental. Gnathostomiasis can be prevented by adequate cooking of fish and poultry in endemic areas.

FURTHER READING

Auer H, Walochnik J: Toxocariasis and the clinical spectrum. Adv Parasitol 109:111, 2020.

Centers for Disease Control and Prevention: Surveillance for trichinellosis—United States, 2015, Annual Summary. Atlanta, GA: U.S. Department of Health and Human Services, CDC, 2017.

Clinical Subcommittee of the Hawaii Governor's Joint Task Force on Rat Lungworm Disease. Preliminary guidelines for the diagnosis and treatment of human neuroangiostrongyliasis (rat lungworm disease) in Hawaii. *https://health.hawaii.gov/docd/files/2018/08/RLWD_Preliminary_Clinical_Guidelines_FINAL_082918.pdf.* Accessed December 1, 2023.

Diaz JH et al: The disease ecology, epidemiology, clinical manifestations, and management of trichinellosis linked to consumption of wild animal meat. Wilderness Environ Med 31:235, 2020.

Lupi O et al: Mucocutaneous manifestations of helminth infections. Nematodes. J Am Acad Dermatol 73:929, 2015.

Martins YC et al: Central nervous system manifestations of *Angiostrongylus cantonensis* infection. Acta Trop 141:46, 2015.

Rostami A et al: Meat sources of infection for outbreaks of human trichinellosis. Food Microbiol 64:65, 2017.

Sitcar AD et al: Raccoon roundworm infection associated with central nervous system disease and ocular disease—six states, 2013–2015. Morbid Mortal Wkly Rep 65:930, 2016.

239 Intestinal Nematode Infections

David J. Diemert, Thomas B. Nutman

More than one billion individuals worldwide are infected with one or more species of intestinal nematode. Table 239-1 summarizes biologic and common clinical features of infections due to the major intestinal parasitic nematodes. These parasites are most common in regions with inadequate sanitation and disposal of fecal waste, particularly in resource-limited countries in the tropics and subtropics, although they have also been seen with increasing frequency among immigrants and refugees to developed countries. Although intestinal nematode infections are not usually fatal, they contribute to malnutrition, impaired physical and cognitive development, and diminished work capacity. It is interesting that these helminth infections may protect some individuals from allergic diseases. Humans may on occasion be infected with nematode parasites that ordinarily infect animals; these zoonotic infections produce diseases such as trichostrongyliasis, anisakiasis, capillariasis, and abdominal angiostrongyliasis.

Intestinal nematodes are roundworms that range in length from 1 mm to many centimeters when mature (Table 239-1). Their life cycles are complex and highly varied; some species, including *Strongyloides stercoralis* and *Enterobius vermicularis*, can be transmitted directly from person to person, while others, such as *Ascaris lumbricoides*, *Trichuris trichiura,* and the hookworms, require a soil phase for development. Because most helminth parasites cannot fully complete their life cycle within the same human host, heavy burdens of adult worms require repeated exposure to the parasite in its infectious stage, whether larva or egg. Hence, clinical disease, as opposed to asymptomatic (or subclinical) infection, generally develops only with repeated exposure and is typically related to infection intensity. In children with marginal nutritional status, intestinal helminth infections may impair growth and development. Eosinophilia and elevated serum IgE antibody levels are features of many helminth infections, the latter particularly when the helminth life cycle involves tissue migration such as *A. lumbricoides, S. stercoralis,* or the hookworms. Significant protective immunity to intestinal nematodes appears not to develop in humans.

ASCARIASIS

A. lumbricoides is the largest intestinal nematode parasite of humans, reaching up to 40 cm in length. Most infected individuals have low worm burdens and are asymptomatic. Clinical disease arises from larval migration in the lungs or the effects of adult worms in the intestines.

Life Cycle Adult worms live in the lumen of the small intestine. Mature female *Ascaris* worms are extraordinarily fecund, each producing over 200,000 eggs a day that are expelled within feces. Ascarid eggs, which are remarkably resistant to environmental stresses, become infective after several weeks of maturation in warm, moist soil and can remain infective for years. After infective eggs are swallowed, gastric acid dissolves their protective outer layer, releasing larvae into the intestine that invade the mucosa, migrate through the circulation to the lungs, penetrate into the alveoli, ascend the bronchial tree, and return—through swallowing—to the small intestine, where they develop into adult worms. The time between initial infection and egg detection in the feces is typically between 2 and 3 months. Adult worms live for 1–2 years.

Epidemiology *Ascaris* is widely distributed in tropical and subtropical regions as well as in other humid areas in more temperate regions of the world. Transmission typically occurs through fecally contaminated soil and is due either to inadequate treatment of sewage or to the use of human feces as fertilizer. With their propensity for poorer oral hygiene, younger children are most often affected. Infection outside endemic areas, though uncommon, can occur when eggs on transported produce are ingested.

TABLE 239-1 Major Human Intestinal Parasitic Nematodes

	PARASITIC NEMATODE				
FEATURE	*ASCARIS LUMBRICOIDES* (ROUNDWORM)	*NECATOR AMERICANUS, ANCYLOSTOMA DUODENALE, ANCYLOSTOMA CEYLANICUM* (HOOKWORM)	*STRONGYLOIDES STERCORALIS*	*TRICHURIS TRICHIURA* (WHIPWORM)	*ENTEROBIUS VERMICULARIS* (PINWORM)
Global prevalence in humans (millions)	730	440	100	465	200
Endemic areas	Hot, humid regions	Hot, humid regions	Hot and warm, humid regions	Hot, humid regions	Worldwide
Infective stage	Egg	Filariform larva	Filariform larva	Egg	Egg
Route of infection	Oral	Percutaneous	Percutaneous or autoinfective	Oral	Oral
Gastrointestinal location of worms	Small intestine	Jejunal mucosa	Small intestinal, mucosa	Cecum, colonic mucosa	Cecum, appendix
Adult worm size	15–40 cm	7–13 mm	1–2 mm	30–50 mm	2–13 mm
Pulmonary passage of larvae	Yes	Yes	Yes	No	No
Incubation period[a] (days)	60–75	40–100	25–30	70–90	35–45
Longevity	1 year	*N. americanus*: 2–5 years *A. duodenale:* 6–8 years *A. ceylanicum*: 6–8 years[b]	Decades (due to autoinfection)	5 years	2 months
Fecundity (eggs/day/worm)	240,000	*N. americanus:* 9000–10,000 *A. duodenale:* 10,000–28,000 *A. ceylanicum*: 5,000–15,000	5000–10,000	3000–7000	2000–10,000
Principal symptoms	Gastrointestinal symptoms; rarely, biliary obstruction or, in heavy infections, gastrointestinal obstruction	Iron-deficiency anemia in moderate and heavy infections	Gastrointestinal symptoms; malabsorption or sepsis in hyperinfection	Gastrointestinal symptoms, rectal prolapse, or anemia in heavy infection	Perianal pruritus
Diagnosis	Eggs in stool	Eggs in fresh stool, larvae in old stool	Larvae in stool or duodenal aspirate; sputum in hyperinfection; serology	Eggs in stool	Eggs from perianal skin on cellulose acetate tape
Treatment	Mebendazole Albendazole Ivermectin Moxidectin Pyrantel pamoate	Mebendazole Albendazole	Ivermectin Albendazole	Mebendazole Albendazole Ivermectin	Mebendazole Albendazole Pyrantel pamoate

[a]Time from infection to egg production by mature female worm. [b]Assumed but no evidence base in humans.

Clinical Features During the lung phase of larval migration, ~9–12 days after egg ingestion, patients may develop an irritating nonproductive cough and burning substernal discomfort that is aggravated by coughing or deep inspiration. Dyspnea and blood-tinged sputum are less common. Fever can occur. Eosinophilia develops during this symptomatic phase and subsides slowly over weeks. Chest imaging may reveal rounded infiltrates a few millimeters to several centimeters in size that are hallmarks of eosinophilic pneumonitis (Löffler's syndrome). These infiltrates may be transient and intermittent, clearing after several weeks. Where there is seasonal transmission of the parasite such as on the Arabian peninsula, seasonal pneumonitis with eosinophilia may develop in previously infected and sensitized hosts.

In light infections, adult worms in the small intestine usually cause no symptoms. In heavier infections, a large bolus of entangled worms can cause pain and small-bowel obstruction, sometimes complicated by perforation, intussusception, or volvulus. Obstruction is more common in children due to their having intestinal lumens of narrow diameter. Single worms may cause disease when they migrate into aberrant sites. A large worm can enter and occlude the biliary tree, causing biliary colic, cholecystitis, cholangitis, pancreatitis, or (rarely) intrahepatic abscesses. Migration of an adult worm up the esophagus can provoke coughing or vomiting up the worm. In highly endemic areas, intestinal and biliary ascariasis can rival acute appendicitis as a cause of surgical acute abdomen.

Laboratory Findings Most cases of ascariasis can be diagnosed by microscopic detection of characteristic *Ascaris* eggs (65 × 45 μm) in fecal samples, although increasingly, polymerase chain reaction (PCR) of worm DNA extracted from stool is being used in research and some clinical settings. Occasionally, patients present after passing an adult worm—identifiable by its large size and smooth cream-colored surface—in the stool or, much less commonly, through the mouth or nose. During the early transpulmonary migratory phase, when eosinophilic pneumonitis occurs, larvae can be found in sputum or gastric aspirates before diagnostic eggs appear in the stool. The eosinophilia that is prominent during this early stage usually decreases to minimal levels in established infection. Adult worms may be visualized, occasionally serendipitously, on contrast radiographic studies of the gastrointestinal tract. A plain abdominal film may reveal masses of worms in gas-filled loops of bowel in patients with intestinal obstruction. Pancreaticobiliary worms can be detected by ultrasound and endoscopic retrograde cholangiopancreatography; the latter method has been used to extract biliary *Ascaris* worms.

TREATMENT

Ascariasis

Ascariasis should always be treated to prevent potentially serious complications. Albendazole (400 mg once), mebendazole (100 mg twice daily for 3 days or 500 mg once), pyrantel pamoate (11 mg/kg once with a maximum dose not to exceed 1 g), ivermectin (150–200 μg/kg once), and moxidectin (8 mg once) are effective. These medications are contraindicated in pregnancy, however. Mild diarrhea and abdominal pain are uncommon side effects of these agents.

Partial intestinal obstruction should be managed with nasogastric suction, IV fluid administration, and instillation of piperazine through the nasogastric tube, although complete obstruction and its severe complications require immediate surgical intervention.

HOOKWORM

Three species (*Necator americanus, Ancylostoma duodenale*, and *Ancylostoma ceylanicum*) are responsible for most human hookworm infections. Most infected individuals are asymptomatic. Hookworm disease develops from a combination of factors—heavy worm burden, prolonged duration of infection, and inadequate iron intake—and results in iron-deficiency anemia and, on occasion, hypoproteinemia.

Life Cycle Adult hookworms, which are ~1 cm long, use buccal teeth (*Ancylostoma*) or cutting plates (*Necator*) to attach to the small-bowel mucosa where they secrete enzymes that enable them to invade submucosal tissues and ingest blood and villous tissue. Female adult hookworms produce thousands of eggs daily. The eggs are deposited with feces in soil, where rhabditiform larvae hatch and develop over a 1-week period into infectious filariform larvae. Infective larvae penetrate the skin and reach the lungs by way of the bloodstream. There they invade alveoli and ascend the airways to the larynx before being swallowed and reaching the small intestine. The prepatent period from skin penetration to appearance of eggs in the feces is ~6–8 weeks, but it may be longer with *Ancylostoma* spp. Less commonly, larvae of *Ancylostoma* spp., if swallowed orally, can survive and develop directly into adult worms that parasitize the intestinal mucosa. Adult hookworms may survive over a decade but usually live ~6–8 years for *A. duodenale* and 2–5 years for *N. americanus*.

Epidemiology Over 400 million people are infected with hookworms worldwide. Whereas *N. americanus* is the predominant species throughout the tropics and subtropics, *A. duodenale* is more geographically restricted to areas of North Africa and northern Asia. *A. ceylanicum* is much less common than the other two species and is mainly found in Southeast Asia. Age prevalence studies have shown a constant increase in hookworm prevalence throughout childhood and into adulthood. Older children have the greatest intensity of hookworm infection; however, in rural areas where fields are fertilized with human feces, older working adults also may be heavily infected.

Clinical Features Most hookworm infections are of light intensity and therefore are clinically asymptomatic. Infective larvae may provoke a pruritic maculopapular rash ("ground itch") at the site of skin penetration as well as serpiginous tracks of subcutaneous migration (similar to those of cutaneous larva migrans; **Chap. 238**) in previously sensitized hosts. Larvae migrating through the lungs occasionally cause mild transient pneumonitis, but this condition develops less frequently in hookworm infection than in ascariasis. In the early intestinal phase, infected persons may develop epigastric pain, abdominal bloating, and other abdominal symptoms accompanied by eosinophilia. The major consequence of chronic hookworm infection is iron deficiency. Symptoms are minimal if iron stores are adequate, but undernourished individuals develop symptoms of progressive iron-deficiency anemia and hypoproteinemia, including fatigue, weakness, and shortness of breath.

Laboratory Findings The diagnosis is established by the finding of characteristic 40 × 60 μm oval hookworm eggs in the feces. Stool concentration techniques may be required to detect light infections. Eggs of the three species are indistinguishable by light microscopy, whereas PCR provides significant improvement in species-specific diagnosis, although this remains a research tool that is not available in commercial laboratories. In a stool sample that is not fresh, hookworm eggs may have hatched to release rhabditiform larvae, which must be differentiated from those of *S. stercoralis*. Hypochromic microcytic anemia, occasionally with eosinophilia or hypoalbuminemia, is characteristic of hookworm disease.

TREATMENT
Hookworm Infection

Hookworm infection can be treated with several safe and effective anthelmintic drugs, including albendazole (400 mg once daily for 3 days) and mebendazole (500 mg once daily or 100 mg twice daily for 3 days). Mild iron-deficiency anemia can often be treated with oral iron supplementation alone. Severe hookworm disease with protein loss and malabsorption necessitates nutritional support and oral iron replacement along with deworming. There is significant concern that the benzimidazoles (mebendazole and albendazole) are becoming much less effective against human hookworms.

Ancylostoma caninum* and *Ancylostoma braziliense *A. caninum*, the canine hookworm, has been identified as a cause of human eosinophilic enteritis, especially in northeastern Australia. In this rare zoonotic infection, adult hookworms attach to the small intestine (where they may be visualized by endoscopy) and elicit abdominal pain and intense local eosinophilia. Treatment with mebendazole (100 mg twice daily for 3 days) or albendazole (400 mg once) or endoscopic removal is effective. Both of these animal hookworm species can cause cutaneous larva migrans ("creeping eruption"; **Chap. 238**).

STRONGYLOIDIASIS

S. stercoralis is distinguished by its ability—unique among helminths (except for *Capillaria*; see below)—to complete its life cycle and replicate within a human host. This capacity permits ongoing cycles of autoinfection as infective larvae are internally produced that can develop into egg-producing adult worms without exiting the host. Infection with *S. stercoralis* can thus persist for decades without further exposure to exogenous infective larvae. In immunocompromised hosts, large numbers of invasive *Strongyloides* larvae can disseminate widely and can be fatal.

Life Cycle In addition to a parasitic life cycle in the human host, *Strongyloides* can undergo a free-living cycle of development in the soil **(Fig. 239-1)**. This adaptability facilitates the parasite's survival in the absence of mammalian hosts. Rhabditiform larvae passed in feces can transform into infectious filariform larvae either directly or after a free-living phase of development. Humans acquire *S. stercoralis* when filariform larvae in fecally contaminated soil penetrate the skin or mucous membranes. The larvae then travel through the bloodstream to the lungs, where they break into the alveolar spaces, ascend the bronchial tree, are swallowed, and thereby reach the small intestine. There the larvae mature into adult worms that embed in the mucosa of the proximal small bowel. The minute (1- to 2-mm long) parasitic adult female worms reproduce by parthenogenesis; adult males do not exist in the human host. Eggs hatch in the intestinal mucosa, releasing rhabditiform larvae that migrate to the lumen and pass with the feces into soil. Alternatively, rhabditiform larvae in the bowel can develop directly into filariform larvae that penetrate the colonic wall or perianal skin and enter the circulation to repeat the migration that establishes ongoing internal reinfection.

Epidemiology *S. stercoralis* is spottily distributed in tropical areas and other warm, humid regions and is common in Southeast Asia, sub-Saharan Africa, and Brazil. In the United States, the parasite is endemic in parts of the Southeast and is found in immigrants, refugees, travelers, and military personnel who have lived in endemic areas.

Clinical Features In uncomplicated strongyloidiasis, many patients are asymptomatic or have mild cutaneous and/or abdominal symptoms. Recurrent urticaria, often involving the buttocks, is the most common cutaneous manifestation. Larvae migrating in the skin can elicit a pathognomonic serpiginous eruption known as *larva currens*. This pruritic, raised, erythematous lesion advances as rapidly as 10 cm/h along the course of larval migration. Adult parasites burrow into the duodenojejunal mucosa and can cause abdominal (usually midepigastric) pain, which resembles peptic ulcer disease except that

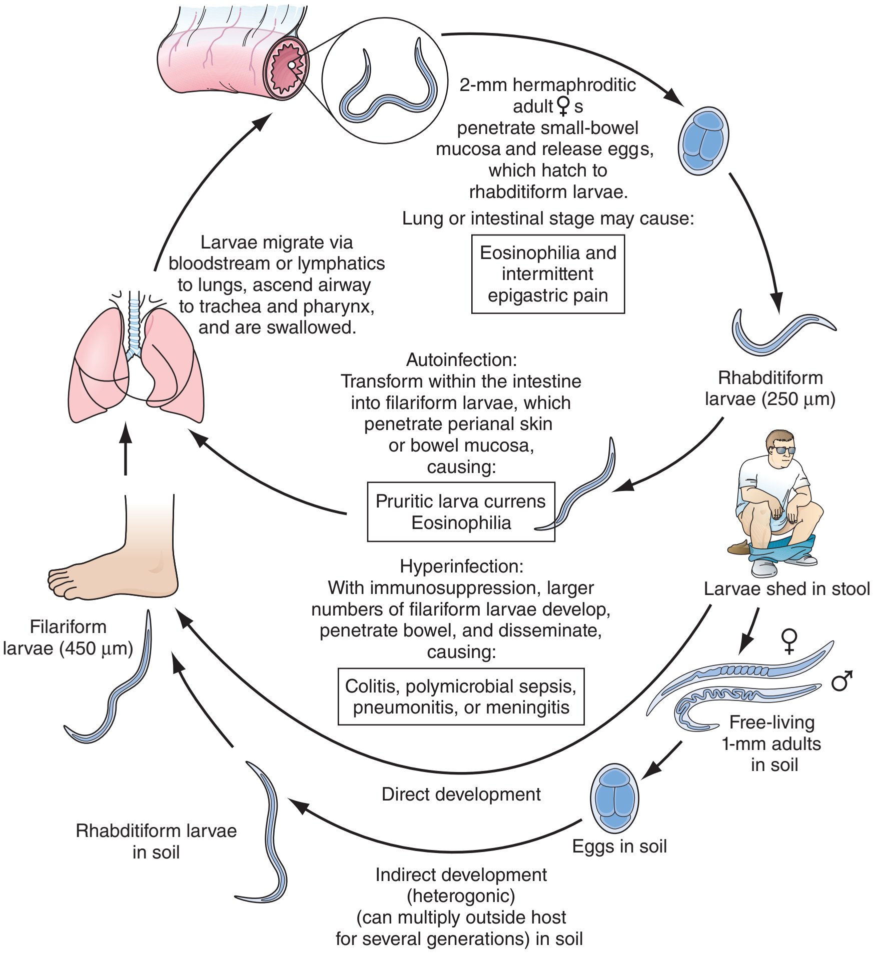
2-mm hermaphroditic adult ♀s penetrate small-bowel mucosa and release eggs, which hatch to rhabditiform larvae.
Lung or intestinal stage may cause:
Eosinophilia and intermittent epigastric pain
Rhabditiform larvae (250 μm)
Larvae shed in stool
♀
♂
Free-living 1-mm adults in soil
Eggs in soil
Direct development
Indirect development (heterogonic) (can multiply outside host for several generations) in soil
Rhabditiform larvae in soil
Filariform larvae (450 μm)
Larvae migrate via bloodstream or lymphatics to lungs, ascend airway to trachea and pharynx, and are swallowed.
Autoinfection: Transform within the intestine into filariform larvae, which penetrate perianal skin or bowel mucosa, causing:
Pruritic larva currens Eosinophilia
Hyperinfection: With immunosuppression, larger numbers of filariform larvae develop, penetrate bowel, and disseminate, causing:
Colitis, polymicrobial sepsis, pneumonitis, or meningitis

FIGURE 239-1 Life cycle of *Strongyloides stercoralis*. *(Reproduced with permission from RL Guerrant et al [eds]: Tropical Infectious Diseases: Principles, Pathogens and Practice, 2nd ed. Elsevier, 2006.)*

it is aggravated by food ingestion. Nausea, watery diarrhea, abdominal bloating, occult gastrointestinal bleeding, and weight loss can occur. Small-bowel obstruction may develop with early, heavy infection. Pulmonary symptoms are rare in uncomplicated strongyloidiasis. Eosinophilia is common, with levels fluctuating over time.

The ongoing autoinfection cycle of *S. stercoralis* is normally constrained by the host's immune system. Impairment of host cell-mediated immunity, especially with high-dose glucocorticoid therapy and less commonly with other immunosuppressive medications, or associated with infection with human T-cell lymphotropic virus type 1 (HTLV-1), may lead to hyperinfection and the generation of large numbers of filariform larvae. Colitis, enteritis, or malabsorption may develop. Hyperinfection can also lead to disseminated strongyloidiasis, characterized by larvae invading not only gastrointestinal tissues and the lungs but also the central nervous system, skin, peritoneum, liver, and kidneys. Moreover, bacteremia may develop because of the spread of enteric flora through disrupted mucosal barriers. Gram-negative sepsis, pneumonia, or meningitis may complicate or dominate the clinical course. Eosinophilia is often absent in severely infected patients. Disseminated strongyloidiasis, particularly in patients with unsuspected infection who are taking glucocorticoids, has a high mortality rate.

Diagnosis In uncomplicated strongyloidiasis, visualizing rhabditiform larvae by light microscopic examination of a fecal sample is diagnostic. Rhabditiform larvae are ~250 μm long, with a short buccal cavity that distinguishes them from hookworm larvae. In uncomplicated infections, few larvae are passed and single stool examinations detect only about one-third of cases. Serial examinations and the use of the agar plate detection method improve the sensitivity of stool diagnosis. Again, PCR has begun to be used more widely and provides increased diagnostic sensitivity. In hyperinfection, a single microscopy-based stool examination is usually positive given the large number of larvae being produced. *Strongyloides* larvae may also be found by sampling of the duodenojejunal contents by aspiration or biopsy. An enzyme-linked immunosorbent assay for serum antibodies to antigens of *Strongyloides* is a sensitive method for diagnosing uncomplicated infections. Such serologic testing should be performed for patients with prior potential exposure, especially those with eosinophilia and/or those who may be starting glucocorticoid therapy or undergoing solid organ transplantation. In disseminated strongyloidiasis, filariform larvae are easily detected in stool as well as in samples obtained from sites of potential larval migration, including sputum, bronchoalveolar lavage fluid, or surgical drainage fluid.

TREATMENT

Strongyloidiasis

Even in the asymptomatic state, strongyloidiasis must be treated because of the potential for subsequent hyperinfection and fatal dissemination. Ivermectin (200 μg/kg daily for 2 days) is more effective than albendazole (400 mg daily for 3 days). Decreases in eosinophil count and antibody titer indicate a response to treatment. For disseminated strongyloidiasis, treatment with ivermectin should be extended for at least 14 days after fecal examinations have become negative. In potentially immunocompromised hosts, the course of ivermectin should be repeated 2 weeks after initial treatment. Ivermectin has been successfully given per rectum in those unable to take ivermectin orally and can be given parenterally (using veterinary preparations) through single-patient expanded access protocols.

TRICHURIASIS

Most infections with *Trichuris trichiura* are asymptomatic, but heavy infections may cause gastrointestinal symptoms. Like the other soil-transmitted helminths, whipworm is distributed globally in the tropics and subtropics and is most common among poor children from resource-limited regions of the world.

Life Cycle Adult *Trichuris* worms reside in the colon and cecum, their thin anterior portions embedded in the superficial mucosa. Thousands of eggs laid daily by adult female worms pass with the feces and mature in the soil. After ingestion, infective eggs hatch in the duodenum, releasing larvae that mature before migrating to the large bowel. The entire cycle takes ~3 months, and adult worms may live for several years.

Clinical Features Most individuals infected with *T. trichiura* have no symptoms or eosinophilia. Heavy infections may result in anemia, abdominal pain, anorexia, and bloody or mucoid diarrhea resembling inflammatory bowel disease. Rectal prolapse can result from massive infections in children due to protracted tenesmus. Moderately heavy *T. trichiura* burdens also contribute to impaired physical growth.

Diagnosis and Treatment The characteristic 50 × 20 μm lemon-shaped *T. trichiura* eggs are readily detected on microscopic examination of stool. Adult worms, which are 3–5 cm long, are occasionally seen on proctoscopy. PCR is being used increasingly in settings where it is available. Albendazole (400 mg daily for 3 days) or mebendazole (100 mg twice daily for 3 days) is safe and moderately effective for treatment. The addition of ivermectin (200 μg/kg daily for 3 doses) to either benzimidazole drug increases cure rates significantly.

ENTEROBIASIS (PINWORM)

E. vermicularis is more common in temperate climates than in the tropics. In the United States, ~40 million persons are infected with pinworms, with most cases occurring in children.

Life Cycle and Epidemiology Adult *Enterobius* worms are ~5–10 mm long and dwell in the cecum. Gravid female worms migrate nocturnally out of the anus to the perianal region where they release up to 10,000 immature eggs each. The eggs become infective within hours and are transmitted by hand-to-mouth passage. From ingested eggs, larvae hatch and mature into adults. This life cycle takes ~1 month, and adult worms survive for ~2 months. Autoinfection results from perianal scratching and transport of infective eggs on the fingers or under the nails to the mouth. Retroinfection can also occur by larvae released from embryonated eggs on the perianal skin migrating back into the rectum. Because of the ease of person-to-person spread, pinworm infections are common among family members.

Clinical Features Most pinworm infections are asymptomatic. Perianal pruritus is the cardinal symptom. The itching, which is often worse at night as a result of the nocturnal emergence of female worms, may lead to excoriation and bacterial superinfection. Heavy infections have been associated with abdominal pain, weight loss, and rarely acute appendicitis. Uncommonly, pinworms may migrate into the female genital tract, causing vulvovaginitis and pelvic or peritoneal granulomas. Eosinophilia is uncommon.

Diagnosis Since pinworm eggs are not released in feces, the diagnosis cannot be made by conventional fecal ova and parasite microscopy. Instead, eggs are detected by the application of clear cellulose acetate tape to the perianal region in the morning. After the tape is transferred to a slide, microscopic examination can detect pinworm eggs, which are oval, measure 55 × 25 μm, and are flattened along one side.

TREATMENT

Enterobiasis

Infected children and adults should be treated with mebendazole (100 mg once), albendazole (400 mg once), or pyrantel pamoate (11 mg/kg once), with the same treatment repeated after 2 weeks because the drugs do not kill eggs or early-stage larvae. Treatment of household members is recommended to eliminate asymptomatic reservoirs of potential reinfection. Linens and clothing should be washed in hot water followed by a hot dryer to kill deposited eggs.

TRICHOSTRONGYLIASIS

Trichostrongylus species, which are normally parasites of herbivorous animals, occasionally infect humans, particularly in Asia, Africa, and Australia. Humans acquire the infection by accidentally ingesting *Trichostrongylus* larvae on contaminated leafy vegetables. The larvae do not migrate in humans but mature directly into adult worms in the small bowel. The pathology caused by these worms is similar to hookworm, although they ingest far less blood. Most infected persons are asymptomatic, but heavy infections may give rise to mild anemia and eosinophilia. In stool examinations, *Trichostrongylus* eggs resemble hookworm eggs but are larger (85 × 115 μm). Treatment consists of mebendazole, albendazole, or pyrantel pamoate **(Chap. 229)**.

ANISAKIASIS

Anisakiasis is a gastrointestinal infection caused by the accidental ingestion in uncooked saltwater fish of nematode larvae belonging to the family Anisakidae. The incidence of anisakiasis in the United States has increased as a result of the growing popularity of raw fish dishes. Most cases occur in Japan, the Netherlands, and Chile, where raw fish—sashimi, pickled green herring, and ceviche, respectively—are national culinary staples. Anisakid nematodes parasitize large sea mammals such as whales, dolphins, and seals. As part of a complex parasitic life cycle involving marine food chains, infectious larvae migrate to the musculature of a variety of fish species that serve as intermediate hosts. Both *Anisakis simplex* and *Pseudoterranova decipiens* have been implicated in human anisakiasis, but an identical gastric syndrome may be caused by the red larvae of eustrongylid parasites of fish-eating birds.

When humans consume infected raw or undercooked fish, live larvae may be coughed up within 48 h. Alternatively, larvae may immediately penetrate the mucosa of the stomach. Within hours, violent upper abdominal pain accompanied by nausea and occasionally vomiting ensues, mimicking an acute abdomen. The diagnosis can be established by direct visualization on upper endoscopy, outlining of the worm by contrast radiographic studies, or histopathologic examination of extracted tissue. Extraction of the burrowing larvae during endoscopy is curative. In addition, larvae may pass to the small bowel, where they penetrate the mucosa and provoke a vigorous eosinophilic granulomatous response. Symptoms may appear 1–2 weeks after the infective meal, with intermittent abdominal pain, diarrhea, nausea, and fever resembling the manifestations of Crohn's disease. Ingestion of *Anisakis*-derived proteins through consumption of fish meat containing *Anisakis* parasites can elicit allergic gastrointestinal and even anaphylactic responses.

Anisakid eggs are not found in the stool since the larvae do not mature in humans. Serologic tests have been developed but are not widely available.

Anisakid larvae in saltwater fish are killed by cooking to 60°C, freezing at –20°C for 3 days, or commercial blast freezing, but usually not by salting, marinating, or cold smoking. No medical treatment is available; surgical or endoscopic removal should be undertaken.

CAPILLARIASIS

Intestinal capillariasis is caused by ingestion of raw fish infected with *Capillaria philippinensis*. Subsequent autoinfection can lead to a severe wasting syndrome. The disease occurs in the Philippines and Thailand and, on occasion, elsewhere in Asia. The natural life cycle of *C. philippinensis* involves fish from fresh and brackish water. When humans eat infected raw fish, the larvae mature in the intestine into adult worms, which produce invasive larvae that cause intestinal inflammation and villus loss and can develop into new adult worms without leaving the human host. Capillariasis has an insidious onset with nonspecific abdominal pain and watery diarrhea. If untreated, repeated rounds of autoinfection can lead to protein-losing enteropathy, severe malabsorption, and ultimately death from cachexia, cardiac failure, or superinfection. The diagnosis is made by identifying characteristic peanut-shaped (20 × 40 μm) eggs on microscopic examination of stool. Severely ill patients require hospitalization and supportive therapy in addition to prolonged anthelmintic treatment with albendazole (400 mg twice daily for 10 days) or mebendazole (500 mg daily for 20 days; **Chap. 229**).

ABDOMINAL ANGIOSTRONGYLIASIS

Abdominal angiostrongyliasis is primarily a disease of children living in Latin America. The zoonotic parasite *Angiostrongylus costaricensis* causes eosinophilic ileocolitis after the ingestion of contaminated vegetation. *A. costaricensis* normally parasitizes the cotton rat and other rodents, with slugs and snails serving as intermediate hosts. Humans become infected by accidentally ingesting infective larvae in mollusk slime deposited on fruits and vegetables; children are at highest risk. The larvae penetrate the gut wall and migrate to mesenteric arterioles, where they develop into adult worms. Eggs deposited in the gut wall provoke an intense eosinophilic granulomatous reaction, and adult worms may cause mesenteric arteritis, thrombosis, or frank bowel infarction. Symptoms may mimic those of appendicitis, including abdominal pain and tenderness, fever, vomiting, and a palpable mass in the right iliac fossa. Leukocytosis and eosinophilia are prominent. CT with contrast typically shows inflamed bowel, often with concomitant obstruction, but a definitive diagnosis is usually made histologically after partial bowel resection. Pathologic examination reveals a thickened bowel wall with eosinophilic granulomas surrounding the *Angiostrongylus* eggs. In nonsurgical cases, the diagnosis rests solely on clinical grounds because larvae and eggs cannot be detected in the stool and serologic tests are not available. Medical therapy for abdominal angiostrongyliasis is of uncertain efficacy. Careful observation and surgical resection for severe symptoms are the mainstays of treatment.

Acknowledgment

The authors wish to acknowledge and thank Peter F. Weller, MD, author of prior editions of this chapter.

FURTHER READING

Costa IN et al: Diagnosis of human strongyloidiasis: Application in clinical practice. Acta Trop 223:106081, 2021.

Else KJ et al: Whipworm and roundworm infections. Nat Rev Dis Primers 6:44, 2020.

Holland C et al: Global prevalence of Ascaris infection in humans (2010–2021): A systematic review and meta-analysis. Infect Dis Poverty 11:113, 2022.

Krolewiecki A et al: Strongyloidiasis: A neglected tropical disease. Infect Dis Clin North Am 33:135, 2019.

Loukas A et al: Hookworm infection. Nat Rev Dis Primers 2:16088, 2016.

Montressor A et al: The global progress of soil-transmitted helminthiases control in 2020 and World Health Organization targets for 2030. PLoS Negl Trop Dis 14:e0008505, 2020.

Moser W et al: Efficacy of recommended drugs against soil transmitted helminths: Systematic review and network meta-analysis. BMJ 358:j4307, 2017.

O'Connell EM et al: *Ancylostoma ceylanicum* hookworm in Myanmar refugees, Thailand, 2012-2015. Emerg Infect Dis 24:1472, 2018.

Shamsi S et al: A critical review of anisakidosis cases occurring globally. Parasitol Res 122:1733, 2023.

240

Filarial and Related Infections

Philip J. Budge, Thomas B. Nutman

Filarial worms are nematodes that dwell in the subcutaneous tissues and the lymphatics. Eight filarial species infect humans **(Table 240-1)**; of these, four—*Wuchereria bancrofti*, *Brugia malayi*, *Onchocerca volvulus*, and *Loa loa*—are responsible for most symptomatic filarial infections. Filarial parasites, which infect an estimated 170 million persons worldwide, are transmitted by specific species of mosquitoes or other arthropods and have a complex life cycle, including insect-borne infective larval stages and adult worms that reside in either lymphatic or subcutaneous tissues of humans. The offspring of adults are microfilariae, which, depending on their species, are 200–250 μm long and 5–7 μm wide, may or may not be enveloped in a loose sheath, and either circulate in the blood or migrate through the skin (Table 240-1). To complete the life cycle, microfilariae are ingested by the arthropod vector and develop over 1–2 weeks into new infective larvae. Adult worms live for many years, whereas microfilariae survive for 3–36 months. The intracellular bacterial endosymbiont *Wolbachia* has been found in all stages of *Brugia*, *Wuchereria*, *Mansonella*, and *Onchocerca* species and has become a target for antifilarial chemotherapy.

Usually, infection is established only with repeated, prolonged exposures to infective larvae. Since the clinical manifestations of filarial diseases develop relatively slowly, these infections should be considered chronic, with possible long-term debilitating effects. The nature, severity, and timing of clinical manifestations in patients with filarial infections who are native to endemic areas and have lifelong exposure may differ significantly from those who are travelers or who have recently moved to these areas. Characteristically, filarial disease is more acute and intense in newly exposed individuals than in natives of endemic areas.

LYMPHATIC FILARIASIS

Lymphatic filariasis is caused by *W. bancrofti*, *B. malayi*, or *Brugia timori*. The threadlike adult parasites reside in afferent lymphatics or lymph nodes, where they may remain viable for more than two decades.

EPIDEMIOLOGY

W. bancrofti, the most widely distributed filarial parasite of humans, affects an estimated 51 million people and is found throughout the tropics and subtropics, including Asia and the Pacific Islands, Africa, areas of South America, and the Caribbean basin. Humans are the only definitive host for the parasite. Generally, the subperiodic form is found only in the Pacific Islands; elsewhere, *W. bancrofti* is nocturnally periodic. Nocturnally periodic forms of microfilariae are scarce in peripheral blood by day and increase at night, whereas subperiodic forms are present in peripheral blood at all times and reach maximal levels in the afternoon. Natural vectors for *W. bancrofti* are *Culex* mosquitoes in urban settings and *Anopheles* or *Aedes* mosquitoes in rural areas.

Brugian filariasis due to *B. malayi* occurs primarily in eastern India, Indonesia, Malaysia, and the Philippines. *B. malayi* also has two forms distinguished by the periodicity of microfilaremia. The more common nocturnal form is transmitted in areas of coastal rice fields, while the subperiodic form is found in forests. *B. malayi* naturally infects cats as well as humans. The distribution of *B. timori* is limited to the islands of southeastern Indonesia.

PATHOLOGY

The principal pathologic changes of lymphatic filariasis result from inflammatory damage to the lymphatics, which is typically caused by adult worms and not by microfilariae. Adult worms live in afferent lymphatics or sinuses of lymph nodes and cause lymphatic dilation and thickening of the vessel walls. The infiltration of plasma cells,

TABLE 240-1 Characteristics of the Filariae

ORGANISM	PERIODICITY	DISTRIBUTION	VECTOR	LOCATION OF ADULT	MICROFILARIAL LOCATION	SHEATH
Wuchereria bancrofti	Nocturnal	Cosmopolitan areas worldwide, including South America, Africa, southern Asia, Papua New Guinea, China, Indonesia	*Culex, Anopheles* (mosquitoes)	Lymphatic tissue	Blood	+
	Subperiodic	Eastern Pacific	*Aedes* (mosquitoes)	Lymphatic tissue	Blood	+
Brugia malayi	Nocturnal	Southeast Asia, Indonesia, India	*Mansonia, Anopheles* (mosquitoes)	Lymphatic tissue	Blood	+
	Subperiodic	Indonesia, Southeast Asia	*Coquillettidia, Mansonia* (mosquitoes)	Lymphatic tissue	Blood	+
Brugia timori	Nocturnal	Indonesia	*Anopheles* (mosquitoes)	Lymphatic tissue	Blood	+
Loa loa	Diurnal	West and Central Africa	*Chrysops* (deerflies)	Subcutaneous tissue	Blood	+
Onchocerca volvulus	None	South and Central America, Africa	*Simulium* (blackflies)	Subcutaneous tissue	Skin, eye	–
Mansonella ozzardi	None	South and Central America	*Culicoides* (midges)	Undetermined site	Blood	–
	None	Caribbean	*Simulium* (blackflies)	Undetermined site	Blood	–
Mansonella perstans	None	South and Central America, Africa	*Culicoides* (midges)	Body cavities, mesentery, perirenal tissue	Blood	–
Mansonella streptocerca	None	West and Central Africa	*Culicoides* (midges)	Subcutaneous tissue	Skin	–

eosinophils, and macrophages in and around the infected vessels, along with endothelial and connective tissue proliferation, leads to tortuosity of the lymphatics and damaged or incompetent lymph valves. Lymphedema and chronic stasis changes with hard or brawny edema develop in the overlying skin. These consequences of filarial infection are due both to the direct effects of the worms and to the host's inflammatory response to the parasite. Inflammatory responses are believed to cause the granulomatous and proliferative processes that precede total lymphatic obstruction. It is thought that the lymphatic vessel remains patent as long as the worm remains viable, and that the death of the worm leads to enhanced granulomatous reactions and fibrosis. Lymphatic obstruction results, and despite collateralization, lymphatic function is compromised.

CLINICAL FEATURES

The most common presentations of the lymphatic filariases are asymptomatic (or subclinical) microfilaremia, hydrocele **(Fig. 240-1)**, acute adenolymphangitis (ADL), and chronic lymphatic disease. In areas where *W. bancrofti* or *B. malayi* is endemic, most infected individuals have few overt clinical manifestations of filarial infection despite the presence of circulating microfilariae in the peripheral blood. Although they may be clinically asymptomatic, virtually all persons with *W. bancrofti* or *B. malayi* microfilaremia have some degree of subclinical disease that includes microscopic hematuria and/or proteinuria, dilated (and tortuous) lymphatics (visualized by imaging), and—in men with *W. bancrofti* infection—scrotal lymphangiectasia (detectable by ultrasound). Despite these findings, most individuals appear to remain clinically asymptomatic for years; in relatively few does the infection progress to either acute or chronic disease.

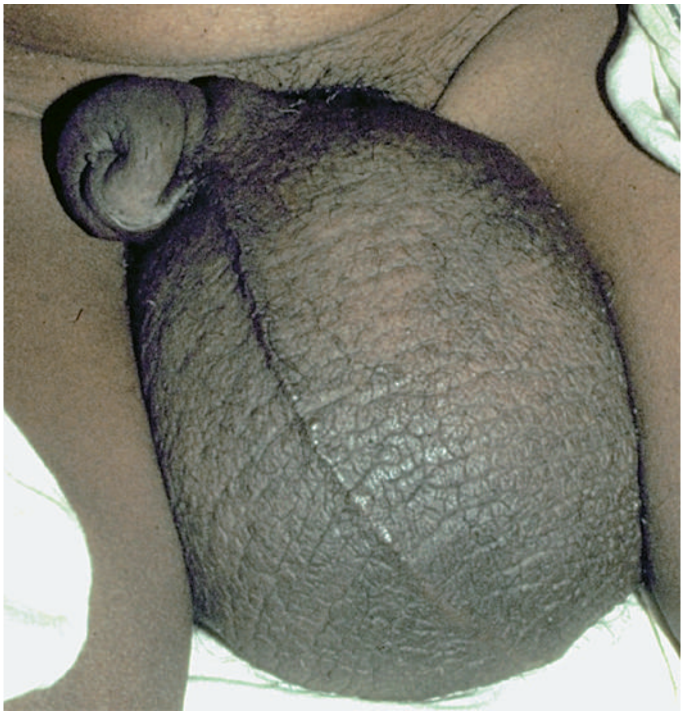

FIGURE 240-1 Hydrocele associated with *Wuchereria bancrofti* infection.

ADL is characterized by high fever, lymphatic inflammation (lymphangitis and lymphadenitis), and transient local edema. The lymphangitis is retrograde, extending peripherally from the lymph node draining the area where the adult parasites reside. Regional lymph nodes are often enlarged, and the entire lymphatic channel can become indurated and inflamed. Concomitant local thrombophlebitis can occur as well. In brugian filariasis, a single local abscess may form along the involved lymphatic tract and subsequently rupture to the surface. The lymphadenitis and lymphangitis can involve both the upper and lower extremities in both bancroftian and brugian filariasis, but involvement of the genital lymphatics occurs almost exclusively with *W. bancrofti* infection. This genital involvement can be manifested by funiculitis, epididymitis, and scrotal pain and tenderness. In endemic areas, another type of acute disease—dermatolymphangioadenitis (DLA)—is recognized as a syndrome that includes high fever, chills, myalgias, and headache. Edematous inflammatory plaques clearly demarcated from normal skin are seen. Vesicles, ulcers, and hyperpigmentation also may be noted. There is often a history of trauma, burns, irradiation, insect bites, punctiform lesions, or chemical injury. Entry lesions, especially in the interdigital area, are common. DLA is often diagnosed as cellulitis.

If lymphatic damage progresses, transient lymphedema can develop into lymphatic obstruction and the permanent changes associated with elephantiasis **(Fig. 240-2)**. Brawny edema follows early pitting edema, the subcutaneous tissues thicken, and hyperkeratosis occurs. Fissuring of the skin develops, as do hyperplastic changes. Superinfection of these poorly vascularized tissues becomes a problem. In bancroftian filariasis, in which genital involvement is common, hydroceles may develop (Fig. 240-1); in advanced stages, this condition may evolve into scrotal lymphedema and scrotal elephantiasis. Furthermore, if there is obstruction of the retroperitoneal lymphatics, increased renal lymphatic pressure leads to rupture of the renal lymphatics and the development of chyluria, which is usually intermittent and most prominent in the morning.

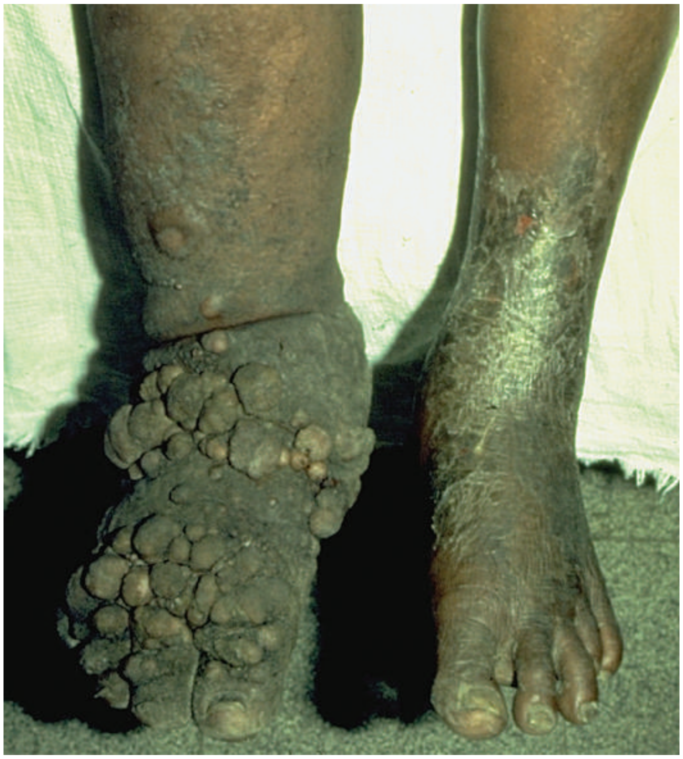

FIGURE 240-2 Elephantiasis of the lower extremity associated with *Wuchereria bancrofti* infection.

The clinical manifestations of filarial infections in travelers or transmigrants who have recently entered an endemic region are distinctive. Given a sufficient number of bites by infected vectors, usually over a 3- to 6-month period, recently exposed patients can develop acute lymphatic or scrotal inflammation with or without urticaria and localized angioedema. Lymphadenitis of epitrochlear, axillary, femoral, or inguinal lymph nodes is often followed by evolving retrograde lymphangitis. Acute attacks are short-lived and are not usually accompanied by fever. With prolonged exposure to infected mosquitoes, these attacks, if untreated, become more severe and lead to permanent lymphatic inflammation and obstruction.

DIAGNOSIS

A definitive diagnosis can be made only by detection of the parasites and hence can be difficult. Adult worms localized in lymphatic vessels or nodes are largely inaccessible. Microfilariae can be found in blood, in hydrocele fluid, or (occasionally) in other body fluids. Such fluids can be examined microscopically, either directly or—for greater sensitivity—after concentration of the parasites by the passage of fluid through a polycarbonate cylindrical-pore filter (pore size, 3 μm) or by the centrifugation of fluid fixed in 2% formalin (Knott's concentration technique). The timing of blood collection is critical and should be based on the periodicity of the microfilariae in the endemic region involved. Many infected individuals do not have microfilaremia, and definitive diagnosis in such cases can be difficult. Assays for circulating antigens of *W. bancrofti* permit the diagnosis of microfilaremic and cryptic (amicrofilaremic) infection. Two tests are commercially available: an enzyme-linked immunosorbent assay (ELISA) and a rapid-format lateral flow assay. Both assays have sensitivities of 93–100% and specificities approaching 100%, although false positives can occur in individuals infected with *Loa loa* (see "Loiasis" below). There are currently no tests for circulating antigens in brugian filariasis.

Polymerase chain reaction (PCR)–based assays for DNA of *W. bancrofti* and *B. malayi* in blood have been developed. A number of studies indicate that the sensitivity of this diagnostic method is equivalent to or greater than that of parasitologic methods.

In cases of suspected lymphatic filariasis, examination of the scrotum, the lymph nodes, or (in female patients) the breast by means of high-frequency ultrasound in conjunction with Doppler techniques may result in the identification of motile adult worms within dilated lymphatics **(Video 240-1)**. Worms may be visualized in the lymphatics of the spermatic cord in up to 80% of men infected with *W. bancrofti*. Live adult worms have a distinctive pattern of movement within the lymphatic vessels (termed the *filarial dance sign*). Radionuclide lymphoscintigraphic imaging of the limbs reliably demonstrates widespread lymphatic abnormalities in both subclinical microfilaremic persons and those with clinical manifestations of lymphatic pathology. Although of potential utility in the delineation of anatomic changes associated with infection, lymphoscintigraphy is unlikely to assume primacy in the diagnostic evaluation of individuals with suspected infection; it is principally a research tool, although it has been used more widely for assessment of lymphedema of any cause. Eosinophilia and elevated serum concentrations of IgE and antifilarial antibody support the diagnosis of lymphatic filariasis. There is, however, extensive cross-reactivity between filarial antigens and antigens of other helminths. Of note, *W. bancrofti*– and *B. malayi*–specific antigens have been identified and are now available for use in rapid diagnostic tests with specificities of >98%. However, seropositivity cannot be equated with active infection: residents of endemic areas can become sensitized to filarial antigens through exposure to infective mosquitoes without having patent filarial infections.

The ADL associated with lymphatic filariasis must be distinguished from thrombophlebitis, infection, and trauma. Retrograde evolution is a characteristic feature that helps distinguish filarial lymphangitis from ascending bacterial lymphangitis. Chronic filarial lymphedema must also be distinguished from the lymphedema of malignancy, postoperative scarring, trauma, chronic edematous states, and congenital lymphatic system abnormalities.

TREATMENT

Lymphatic Filariasis

With newer definitions of clinical syndromes in lymphatic filariasis and new tools to assess clinical status (e.g., ultrasound, lymphoscintigraphy, circulating filarial antigen assays, PCR), approaches to treatment based on infection status can be considered.

Orally administered diethylcarbamazine (DEC; 6 mg/kg daily for 12 days), which has both macro- and microfilaricidal properties, remains the drug of choice for the treatment of active lymphatic filariasis (defined by microfilaremia, antigen positivity, or adult worms on ultrasound), although albendazole (400 mg twice daily by mouth for 21 days) also has demonstrated macrofilaricidal efficacy. A 4- to 6-week course of oral doxycycline (targeting the intracellular *Wolbachia*) also has significant macrofilaricidal activity, as does DEC/albendazole used daily for 7 days. The addition of DEC to a 3-week course of doxycycline is also efficacious in lymphatic filariasis.

Regimens that combine single doses of albendazole (400 mg) with either DEC (6 mg/kg) or ivermectin (200 μg/kg) all have a sustained microfilaricidal effect and are the mainstay of programs for the eradication of lymphatic filariasis in Africa (albendazole/ivermectin) and elsewhere (albendazole/DEC) (see "Prevention and Control," below). More recently, a regimen using single doses of the three major antifilarial drugs (albendazole/DEC/ivermectin) has been shown to sustain microfilarial clearance out to at least 2 years.

As already mentioned, a growing body of evidence indicates that, although they may be asymptomatic, virtually all persons with *W. bancrofti* or *B. malayi* microfilaremia have some degree of subclinical disease (hematuria, proteinuria, abnormalities on lymphoscintigraphy). Thus, early treatment of asymptomatic persons who have microfilaremia is recommended to prevent further lymphatic damage. For ADL, supportive treatment (including the administration of antipyretics and analgesics) is recommended, as is antibiotic therapy if secondary bacterial infection is likely. Similarly, because lymphatic disease is associated with the presence of adult worms, treatment with DEC is recommended for microfilaria-negative carriers of adult worms.

In persons with chronic manifestations of lymphatic filariasis, treatment regimens that emphasize hygiene, prevention of secondary bacterial infections, and physiotherapy have gained wide acceptance for morbidity control. These regimens are similar to those recommended for lymphedema of most nonfilarial causes and are known by a variety of names, including *complex decongestive physiotherapy* and *complex lymphedema therapy*. Hydroceles (Fig. 240-1) can be managed surgically. With chronic manifestations of lymphatic filariasis, drug treatment should be reserved for individuals who have evidence of active infection; however, in some settings, a 6-week course of doxycycline has been shown to provide improvement in filarial lymphedema irrespective of disease activity.

Side effects of DEC treatment in infected individuals may include fever, chills, arthralgias, headaches, nausea, and vomiting. Both the development and the severity of these reactions are directly related to the number of microfilariae circulating in the bloodstream. The adverse reactions may represent either an acute hypersensitivity reaction to the antigens being released by dead and dying parasites or an inflammatory reaction induced by the intracellular *Wolbachia* endosymbionts freed from their intracellular niche.

Ivermectin has a side effect profile similar to that of DEC when used in lymphatic filariasis. In patients infected with *L. loa* who have high levels of microfilaremia, DEC—like ivermectin (see "Loiasis," below)—can elicit severe encephalopathic complications. When used in single-dose regimens for the treatment of lymphatic filariasis, albendazole is associated with relatively few side effects.

PREVENTION AND CONTROL

To protect themselves against filarial infection, individuals must avoid contact with infected mosquitoes by using personal protective measures, including bed nets, particularly those impregnated with insecticides such as permethrin. Mass drug administration (MDA) is the current approach to elimination of lymphatic filariasis as a public health problem. The underlying tenet of this approach is that mass annual distribution of antifilarial chemotherapy—albendazole with either DEC (for all areas except those where onchocerciasis is co-endemic; see section on onchocerciasis treatment, below) or ivermectin or with both ivermectin and DEC (triple-drug therapy)—will profoundly suppress microfilaremia. If the suppression is sustained, then transmission can be interrupted.

Created by the World Health Organization in 1997, the Global Programme to Eliminate Lymphatic Filariasis is based on mass administration of single annual doses of DEC plus albendazole in non-African regions and of albendazole plus ivermectin in Africa. Available information from late 2022 indicated that >935 million persons in 68 countries had thus far participated. Not only has lymphatic filariasis been eliminated in some defined areas, but collateral benefits—avoidance of disability and treatment of intestinal helminths and other conditions (e.g., scabies and louse infestation)—also have been noted. The strategy of the global program is being refined, and attempts are being made to integrate this effort with other mass-treatment strategies (e.g., deworming programs, malaria control, and trachoma control) in an integrated control strategy.

TROPICAL PULMONARY EOSINOPHILIA

Tropical pulmonary eosinophilia (TPE) is a distinct syndrome that develops in some individuals infected with the lymphatic-dwelling filarial species. Most cases have been reported from India, Pakistan, Sri Lanka, Brazil, Guyana, and Southeast Asia; the decreasing incidence of TPE in the past decade probably reflects global MDA efforts.

CLINICAL FEATURES

The main features include a history of residence in filaria-endemic regions, paroxysmal cough and wheezing (usually nocturnal and probably related to the nocturnal periodicity of microfilariae), weight loss, low-grade fever, lymphadenopathy, and pronounced blood eosinophilia (>3000 eosinophils/μL). Chest x-rays or computed tomography (CT) scans may be normal but generally show increased bronchovascular markings. Diffuse miliary lesions or mottled opacities may be present in the middle and lower lung fields. Tests of pulmonary function show restrictive abnormalities in most cases and obstructive defects in half. Characteristically, total serum IgE levels (4–40 KIU/mL) and antifilarial antibody levels are markedly elevated.

PATHOLOGY

In TPE, microfilariae and parasite antigens are rapidly cleared from the bloodstream by the lungs. The clinical symptoms result from allergic and inflammatory reactions elicited by the cleared parasites. In some patients, trapping of microfilariae in other reticuloendothelial organs can cause hepatomegaly, splenomegaly, or lymphadenopathy. A prominent, eosinophil-enriched, intra-alveolar infiltrate is common, and with it comes the release of cytotoxic proinflammatory eosinophil granule proteins that may mediate some of the pathology seen in TPE. In the absence of successful treatment, interstitial fibrosis can lead to progressive pulmonary damage.

DIFFERENTIAL DIAGNOSIS

TPE must be distinguished from asthma, Löffler's syndrome, allergic bronchopulmonary aspergillosis, allergic granulomatosis with polyangiitis (eosinophilic granulomatosis with polyangiitis or Churg-Strauss syndrome), other systemic vasculitides (most notably, periarteritis nodosa), chronic eosinophilic pneumonia, and the hypereosinophilic syndromes (HESs).

TREATMENT

Tropical Pulmonary Eosinophilia

DEC is used at a daily dosage of 4–6 mg/kg for 14 days. Symptoms usually resolve within 3–7 days after the initiation of therapy. Relapse, which occurs in ~12–25% of cases (sometimes after an interval of several years), requires re-treatment.

ONCHOCERCIASIS

EPIDEMIOLOGY

Onchocerciasis ("river blindness") is caused by the filarial nematode *O. volvulus*, which infects an estimated 37 million individuals in 31 countries worldwide. The majority of individuals infected with *O. volvulus* live in the equatorial region of Africa extending from the Atlantic coast to the Red Sea. In the Americas, the only remaining countries with isolated foci are Venezuela and Brazil. The infection is also found in Yemen.

ETIOLOGY

Infection in humans begins with the deposition of infective larvae on the skin by the bite of an infected blackfly. The larvae develop into adults, which are typically found in subcutaneous nodules. About 7 months to 3 years after infection, the gravid female releases microfilariae that migrate out of the nodule and throughout the tissues, concentrating in the dermis. Infection is transmitted to other persons when a female fly ingests microfilariae from the host's skin and these microfilariae then develop into infective larvae. Adult *O. volvulus* females and males are ~40–60 cm and ~3–6 cm in length, respectively. The life span of adults can be as long as 18 years, with an average of ~9 years. Because the blackfly vector breeds along free-flowing rivers and streams (particularly in rapids) and generally restricts its flight to an area within several kilometers of these breeding sites, both biting and disease transmission are most intense in these locations.

PATHOLOGY

Onchocerciasis primarily affects the skin, eyes, and lymph nodes. In contrast to the pathology in lymphatic filariasis, the damage in onchocerciasis is elicited by microfilariae and not by adult parasites. In the skin, there are mild but chronic inflammatory changes that can result in loss of elastic fibers, atrophy, and fibrosis. The subcutaneous nodules (*onchocercomata*) consist primarily of fibrous tissues surrounding the adult worm, often with a peripheral ring of inflammatory cells surrounded by an endothelial layer (characterized as lymphatic in origin). In the eye, neovascularization and corneal scarring lead to corneal opacities and blindness. Inflammation in the anterior and posterior chambers frequently results in anterior uveitis, chorioretinitis, and optic atrophy. Although punctate opacities are due to an inflammatory reaction surrounding dead or dying microfilariae, the pathogenesis of most manifestations of onchocerciasis is still unclear.

■ CLINICAL FEATURES

Skin Pruritus and rash are the most common manifestations of onchocerciasis. The pruritus can be incapacitating; the rash is typically a papular eruption **(Fig. 240-3)** that is generalized rather than localized to a particular region of the body. Long-term infection results in exaggerated and premature wrinkling of the skin, loss of elastic fibers, and epidermal atrophy that can lead to loose, redundant skin and hypo- or hyperpigmentation. Localized eczematoid dermatitis can cause hyperkeratosis, scaling, and pigmentary changes. In an immunologically hyperreactive form of onchodermatitis (commonly termed *sowdah* or *localized onchodermatitis*), the affected skin darkens as a consequence of the profound inflammation that occurs as microfilariae in the skin are cleared.

Onchocercomata These subcutaneous nodules, which can be palpable and/or visible, contain the adult worm. They are most common over the coccyx and sacrum, the trochanter of the femur, the lateral anterior crest, and other bony prominences. Nodules vary in size and characteristically are firm and not tender. It has been estimated that, for every palpable nodule, there are four deeper nonpalpable ones.

Ocular Tissue Visual impairment is the most serious complication of onchocerciasis and usually affects only those persons with moderate or heavy infections. Lesions may develop in all parts of the eye. The most common early finding is conjunctivitis with photophobia. Punctate keratitis—acute inflammatory reactions surrounding dying microfilariae and manifested as “snowflake” opacities—is common among younger patients and resolves without apparent complications. Sclerosing keratitis occurs in 1–5% of infected persons and is the leading cause of onchocercal blindness. Anterior uveitis and iridocyclitis develop in ~5% of infected persons. Characteristic chorioretinal lesions develop as a result of atrophy and hyperpigmentation of the retinal pigment epithelium. Constriction of the visual fields and overt optic atrophy may occur.

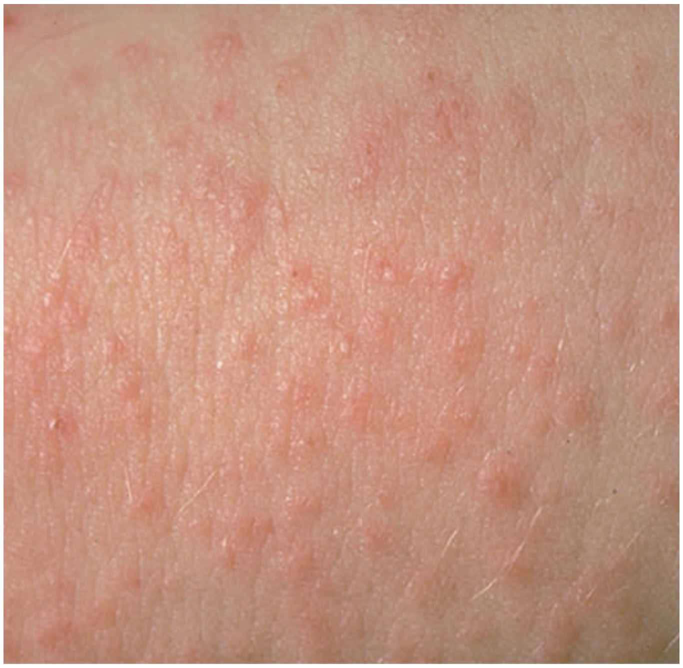

FIGURE 240-3 Papular eruption as a consequence of onchocerciasis.

Lymph Nodes Mild to moderate lymphadenopathy is common, particularly in the inguinal and femoral areas, where the enlarged nodes may hang down in response to gravity (“hanging groin”), sometimes predisposing to inguinal and femoral hernias.

Other Manifestations Some heavily infected individuals develop cachexia with loss of adipose tissue and muscle mass. A form of dwarfism, Nakalanga dwarfism, has been attributed to pituitary involvement in this infection. An association between onchocerciasis and epilepsy (including an epidemic form termed nodding syndrome) has gained attention recently. Among adults who become blind, there is a three- to fourfold increase in mortality rate.

■ DIAGNOSIS

Definitive diagnosis depends on the detection of an adult worm in an excised nodule or, more commonly, of microfilariae in a skin snip. Skin snips are obtained with a corneal-scleral punch or by lifting of the skin with the tip of a needle and excision of a small (1- to 3-mm) piece with a sterile scalpel blade. Both methods collect a blood-free skin biopsy sample extending to just below the epidermis. The biopsy tissue can be incubated in tissue culture medium or in saline on a glass slide or flat-bottomed microtiter plate. After incubation for 2–4 h (or occasionally overnight in light infections), microfilariae emergent from the skin can be seen by low-power microscopy or can be detected by PCR.

Eosinophilia and elevated serum IgE levels are common, but, because these features are seen in many parasitic infections, are not diagnostic in themselves. Immunoassays to detect antibodies to *Onchocerca*-specific antigens are being used both in specialized laboratories and at the point of contact in rapid-diagnostic formats.

TREATMENT

Onchocerciasis

The main goals of therapy are to prevent the development of irreversible lesions and to alleviate symptoms. Chemotherapy is the mainstay of management. Ivermectin, a semisynthetic macrocyclic lactone active against microfilariae, is the first-line agent for the treatment of onchocerciasis. It is given orally in a single dose of 150 μg/kg, either yearly or semiannually. More frequent ivermectin administration (every 3 months) has been suggested to ameliorate pruritus and skin disease.

After treatment, most individuals have few or no reactions. Pruritus, cutaneous edema, and/or maculopapular rash occur in ~1–10% of treated individuals. In areas of Africa co-endemic for *O. volvulus* and *L. loa*, however, ivermectin is contraindicated (as it is for pregnant or breast-feeding women) because of severe posttreatment encephalopathy, especially in patients who are heavily microfilaremic for *L. loa* (>30,000 microfilariae/mL). Although ivermectin treatment results in a marked drop in microfilarial density, its effect can be short-lived (<3 months in some cases). Thus, it is occasionally necessary to give ivermectin more frequently for persistent symptoms. Another macrocyclic lactone, moxidectin, has a more prolonged microfilaricidal effect but is not yet commercially available.

A 6-week course of doxycycline is macrofilaristatic, rendering female adult worms sterile for long periods.

■ PREVENTION

Vector control has been beneficial in highly endemic areas in which breeding sites are vulnerable to insecticide spraying, but most areas endemic for onchocerciasis are not suited to this type of control. Community-based administration of ivermectin every 6–12 months is being used to interrupt transmission in endemic areas. This measure, in conjunction with vector control, has already helped eliminate the

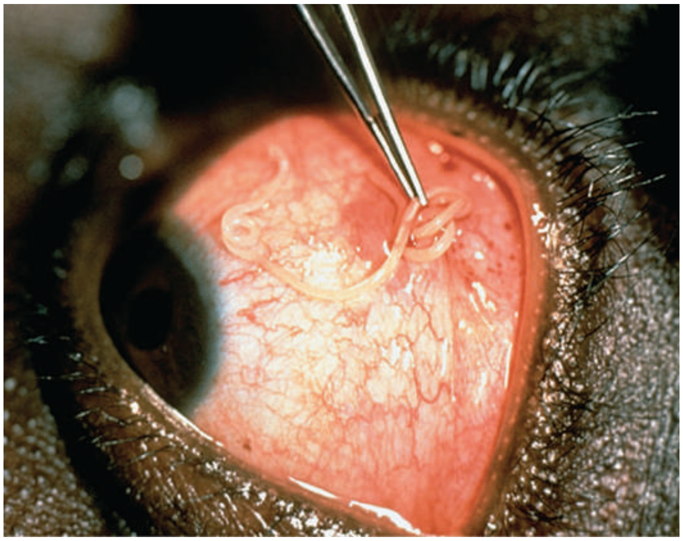

FIGURE 240-4 Adult *Loa loa* worm being surgically removed after its subconjunctival migration.

infection in most of Latin America and has reduced the prevalence of disease in many endemic foci in Africa. No drug has proved useful for prophylaxis of *O. volvulus* infection.

LOIASIS

ETIOLOGY AND EPIDEMIOLOGY

Loiasis is caused by *L. loa* (the African eye worm), which is present in the rainforests of West and Central Africa. Adult parasites (females, 50–70 mm long and 0.5 mm wide; males, 25–35 mm long and 0.25 mm wide) live in subcutaneous tissues. Microfilariae circulate in the blood with a diurnal periodicity that peaks between 10:00 A.M. and 2:00 P.M.

CLINICAL FEATURES

Manifestations of loiasis in natives of endemic areas may differ from those in temporary residents or visitors. Among the indigenous population, loiasis is often an asymptomatic infection with microfilaremia. Infection may be recognized only after subconjunctival migration of an adult worm **(Fig. 240-4)** or may be manifested by episodic *Calabar swellings*—evanescent localized areas of angioedema and erythema developing on the extremities and less frequently at other sites. Nephropathy, encephalopathy, and cardiomyopathy can occur but are infrequently diagnosed. Although historically considered relatively benign in endemic populations, recent studies show loiasis is associated with chronic morbidity and increased mortality in endemic populations. In patients who are not residents of endemic areas, allergic symptoms predominate, episodes of Calabar swelling tend to be more frequent, microfilaremia is less common, and eosinophilia and increased levels of antifilarial antibodies are characteristic.

PATHOLOGY

The pathogenesis of the manifestations of loiasis is poorly understood. Calabar swellings are thought to result from a hypersensitivity reaction to adult worm antigens.

DIAGNOSIS

Definitive diagnosis of loiasis requires the detection of microfilariae in the peripheral blood or the isolation of the adult worm from the eye (Fig. 240-4) or from a subcutaneous biopsy specimen collected from a site of swelling developing after treatment. PCR-based assays for the detection of *L. loa* DNA in blood are available in specialized laboratories and are highly sensitive and specific, as are some newer recombinant antigen–based serologic techniques. In practice, the diagnosis must often be based on a characteristic history and clinical presentation, blood eosinophilia, and elevated levels of antifilarial antibodies, particularly in travelers to an endemic region, who are often amicrofilaremic.

TREATMENT

Loiasis

DEC (8–10 mg/kg per day administered orally for 21 days) is effective against both the adult and the microfilarial forms of *L. loa*, but multiple courses are frequently necessary before loiasis resolves completely. In cases of heavy microfilaremia, allergic or other inflammatory reactions can take place during treatment, including central nervous system involvement with coma and encephalitis. Heavy infections can be treated initially with apheresis to remove the microfilariae and with glucocorticoids (40–60 mg of prednisone per day) followed by low doses of DEC (0.5 mg/kg per day). If antifilarial treatment has no adverse effects, the prednisone dose can be tapered rapidly and the dose of DEC gradually increased to 8–10 mg/kg per day.

Albendazole or ivermectin is effective in reducing microfilarial loads, although neither is approved for this purpose by the U.S. Food and Drug Administration. Moreover, ivermectin is contraindicated in patients with >30,000 microfilariae/mL because this drug has been associated with severe adverse events (including encephalopathy and death) in heavily infected patients with loiasis in West and Central Africa. DEC (300 mg weekly) is an effective prophylactic regimen for loiasis.

STREPTOCERCIASIS

Mansonella streptocerca, found mainly in the tropical forest belt of Africa from Ghana to the Democratic Republic of the Congo, is transmitted by biting midges. The major clinical manifestations involve the skin and include pruritus, papular rashes, and pigmentation changes. Many infected individuals have inguinal adenopathy, although most are asymptomatic. The diagnosis is made by detection of the characteristic microfilariae in skin snips. Ivermectin at a single dose of 150 μg/kg leads to sustained suppression of microfilariae in the skin and is probably the treatment of choice for streptocerciasis.

MANSONELLA PERSTANS INFECTION

M. perstans, distributed across the center of Africa and in northeastern South America, is transmitted by midges. Adult worms reside in serous cavities—pericardial, pleural, and peritoneal—as well as in the mesentery and the perirenal and retroperitoneal tissues. Microfilariae circulate in the blood without periodicity. The clinical and pathologic features of the infection are poorly defined. Most patients appear to be asymptomatic, but manifestations may include transient angioedema and pruritus of the arms, face, or other parts of the body (analogous to the Calabar swellings of loiasis); fever; headache; arthralgias; and right-upper-quadrant pain. Occasionally, pericarditis and hepatitis occur. The diagnosis is based on the demonstration of microfilariae in blood or serosal effusions. Perstans filariasis is often associated with peripheral-blood eosinophilia and antifilarial antibody elevations.

With the identification of a *Wolbachia* endosymbiont in *M. perstans*, doxycycline (200 mg twice a day) for 6 weeks has been established as the first effective treatment for this infection.

MANSONELLA OZZARDI INFECTION

The distribution of *M. ozzardi* is restricted to Central and South America and certain Caribbean islands. Adult worms are rarely recovered from humans. Microfilariae circulate in the blood without periodicity. Although this organism has often been considered nonpathogenic, headache, articular pain, fever, pulmonary symptoms, adenopathy, hepatomegaly, pruritus, and eosinophilia have been ascribed to *M. ozzardi* infection. The diagnosis is made by detection of microfilariae in peripheral blood. Ivermectin is effective in treating this infection.

ZOONOTIC FILARIAL INFECTIONS

Dirofilariae that affect primarily dogs, cats, and raccoons occasionally infect humans incidentally, as do *Brugia* and *Onchocerca* parasites that affect small mammals. Because humans are an abnormal host, the parasites never develop fully. Pulmonary dirofilarial infection caused by the canine heartworm *Dirofilaria immitis* generally presents in humans as a solitary pulmonary nodule. Chest pain, hemoptysis, and cough are uncommon. Infections with *Dirofilaria repens* (from dogs) or *Dirofilaria tenuis* (from raccoons) can cause local subcutaneous nodules in humans. Zoonotic *Brugia* infection can produce isolated lymph node enlargement, whereas zoonotic *Onchocerca* species (particularly *O. lupi*) can cause subconjunctival masses. Eosinophilia levels and antifilarial antibody titers are not commonly elevated. Excisional biopsy is both diagnostic and curative. These infections usually do not respond to antifilarial chemotherapy.

DRACUNCULIASIS (GUINEA WORM INFECTION)

ETIOLOGY AND EPIDEMIOLOGY

The incidence of dracunculiasis, caused by *Dracunculus medinensis*, has declined dramatically because of global eradication efforts. However, between 2017 and 2020, there were increases in the number of human cases, followed by a decline in 2021 and 2022. At the end of 2022, there were a total of 13 human cases of Guinea worm disease across four African countries, with 6 cases in Chad, 5 cases in South Sudan, and 1 each in Ethiopia and Central African Republic. Low-level persistence of infection in animal reservoirs (dogs and cats) is a major challenge to ongoing eradication efforts.

Humans acquire *D. medinensis* when they ingest water containing infective larvae derived from *Cyclops*, a crustacean that is the intermediate host. Larvae penetrate the stomach or intestinal wall, mate, and mature. The adult male probably dies; the female worm develops over a year and migrates to subcutaneous tissues, usually in the lower extremity. As the thin female worm, ranging in length from 30 cm to 1 m, approaches the skin, a blister forms that, over days, breaks down and forms an ulcer. When the blister opens, large numbers of motile, rhabditiform larvae can be released into stagnant water; ingestion by *Cyclops* completes the life cycle.

CLINICAL FEATURES

Few or no clinical manifestations of dracunculiasis are evident until just before the blister forms, when there is an onset of fever and generalized allergic symptoms, including periorbital edema, wheezing, and urticaria. The emergence of the worm is associated with local pain and swelling. When the blister ruptures (usually as a result of immersion in water) and the adult worm releases larva-rich fluid, symptoms are relieved. The shallow ulcer surrounding the emerging adult worm heals over weeks to months. Such ulcers, however, can become secondarily infected, the result being cellulitis, local inflammation, abscess formation, or (uncommonly) tetanus. Occasionally, the adult worm does not emerge but becomes encapsulated and calcified.

DIAGNOSIS

The diagnosis is based on the findings developing with the emergence of the adult worm, as described above.

TREATMENT

Dracunculiasis

Gradual extraction of the worm by winding of a few centimeters on a stick each day remains the common and effective practice. Worms may be excised surgically. No drug is effective in treating dracunculiasis.

PREVENTION

Prevention, which remains the only real control measure, depends on the provision of safe drinking water.

Acknowledgment

The authors wish to acknowledge and thank Peter F. Weller, MD, author of prior editions of this chapter.

FURTHER READING

Hopkins DR et al: Progress toward global eradication of dracunculiasis—January 2022–June 2023. Morb Mortal Wkly Rep 72:1230, 2023.

King CL et al: Single-dose triple-drug therapy for *Wuchereria bancrofti*—5-year follow-up. N Engl J Med 382:1956, 2020.

Opoku NO et al: Single dose moxidectin versus ivermectin for *Onchocerca volvulus* infection in Ghana, Liberia, and the Democratic Republic of the Congo: A randomised, controlled, double-blind phase 3 trial. Lancet 392:1207, 2018.

Ramharter M et al: The African eye worm: Current understanding of the epidemiology, clinical disease, and treatment of loiasis. Lancet Infect Dis 24:e165, 2024.

Taylor MJ et al: Lymphatic filariasis and onchocerciasis. Lancet 376:1175, 2010.

VIDEO 240-1 Filarial dance sign. Ultrasound showing dilated lymphatic vessel (upper right) containing live, motile, adult worms.

241

Schistosomiasis and Other Trematode Infections

Birgitte Jyding Vennervald,
Anna-Sofie Stensgaard

Trematodes, or flatworms, are a group of helminths that belong to the phylum Platyhelminthes. The adult flatworms share some common characteristics, such as macroscopic size (from one to several centimeters); dorsoventrally flattened, bilaterally symmetric bodies; and two suckers—oral and ventral. Except for schistosomes, which have separate sexes, all human parasitic trematodes are hermaphroditic. Their life cycles involve a mammalian/human definitive host, in which sexual reproduction by adult worms takes place, and an intermediate host (snails), in which asexual multiplication occurs. Some species of trematodes have more than one intermediate host.

Humans are infected either by direct penetration of intact skin (schistosomiasis) or by ingestion of raw freshwater fish, crustaceans, or aquatic plants with metacercariae—the infective larval stage.

Significant trematode infections of humans may be divided according to the location of the adult worms: blood, liver (biliary tree), intestines, or lungs **(Table 241-1)**. Adult worms do not multiply within the mammalian host but can live for up to 30 years. Infections are often chronic.

Although it is relatively rare to encounter patients with trematode infections in the United States, many millions of people are infected worldwide. Both schistosomiasis and food-borne trematode infections are poverty-related chronic diseases with high morbidity and a significant public health impact. Various factors may increase the spread of the infections globally. Increasing temperatures may render new areas suitable for the intermediate host snails, and an increase in travel and migration may increase the number of patients with trematode infections—for example, in the United States.

APPROACH TO THE PATIENT

Trematode Infection

In the evaluation of a patient in whom trematode infection is suspected, certain questions are highly relevant and can assist in establishing a diagnosis: Where have you been? If you have traveled, when

TABLE 241-1 Major Human Trematode Infections

TREMATODE	TRANSMISSION ROUTE	GEOGRAPHIC DISTRIBUTION
Blood Flukes		
Intestinal schistosomiasis		
Schistosoma mansoni	Skin penetration by cercariae released from snails (*Biomphalaria* spp.)	Africa, Brazil, Venezuela, Surinam, the Caribbean (low risk)
Shistosoma japonicum	Skin penetration by cercariae released from snails (*Oncomelania* spp.)	China, Indonesia, Philippines
Schistosoma guineensis and *Schistosoma intercalatum*	Skin penetration by cercariae released from snails (*Bulinus* spp.)	Rainforest areas of Central Africa
Schistosoma mekongi	Skin penetration by cercariae released from snails (*Neotricula aperta*)	Several districts of Cambodia and Lao People's Democratic Republic (PDR)
Urogenital schistosomiasis		
Schistosoma haematobium	Skin penetration by cercariae released from snails (*Bulinus* spp.)	Africa, Middle East, Corsica (France)
Liver Flukes		
Clonorchis sinensis	Ingestion of metacercariae in freshwater fish	Asia, including Republic of Korea, China, Taiwan, Vietnam
Opisthorchis viverrini	Ingestion of metacercariae in freshwater fish	Northeast Thailand, Lao PDR, Cambodia, Vietnam
Opisthorchis felineus	Ingestion of metacercariae in freshwater fish	Former Soviet Union, Kazakhstan, Ukraine, Turkey
Fasciola hepatica	Ingestion of metacercariae on aquatic plants or in water	Worldwide
Fasciola gigantica	Ingestion of metacercariae on aquatic plants or in water	Africa, Asia
Intestinal Flukes		
Fasciolopsis buski	Ingestion of metacercariae on aquatic plants	Bangladesh, China, India, Indonesia, Lao PDR, Malaysia, Taiwan, Thailand, Vietnam
Echinostoma spp.	Ingestion of freshwater fish, frogs, mussels, snails	China, India, Indonesia, Japan, Malaysia, Russia, Republic of Korea, Philippines, Thailand
Heterophyes heterophyes, several other species	Ingestion of metacercariae in freshwater or brackish-water fish	Egypt, Greece, Islamic Republic of Iran, Italy, Japan, Republic of Korea, Sudan, Tunisia, Turkey
Lung Flukes		
Paragonimus westermani	Ingestion of metacercariae in crayfish or crabs	Tropical and subtropical areas of eastern and southern Asia and sub-Saharan Africa
Paragonimus kellicotti	Ingestion of metacercariae in crayfish or crabs	North America

did you return? What activities have you been involved in (trekking, swimming, whitewater rafting)? What have you been eating (local dishes while traveling; raw, poorly cooked, or pickled freshwater fish or crustaceans)? Definitive diagnosis is based on detection of parasite eggs in stool, urine, sputum, and sometimes tissue samples or on serologic tests. The presence of eosinophilia and a history of travel to endemic areas should raise suspicion of trematode infection. The U.S. Centers for Disease Control and Prevention (CDC) can provide guidance with respect to diagnosis and treatment.

SCHISTOSOMIASIS

Human schistosomiasis is caused by six species of the parasitic genus *Schistosoma*: *S. mansoni*, *S. japonicum*, *S. mekongi*, *S. intercalatum*, and the recently described *S. guineensis* cause intestinal disease, and *S. haematobium* causes urogenital disease (Table 241-1). The infection may cause considerable intestinal, hepatic, and genitourinary morbidity. Avian schistosomes may penetrate human skin but die in subcutaneous tissue, producing only cutaneous manifestations.

ETIOLOGY

Schistosoma infection is contracted through contact with freshwater bodies harboring infected intermediate-host snails **(Fig. 241-1)**. Cercariae, the infective larval stage released from the snail, penetrate intact human skin within a few minutes after attaching to the skin. After penetration, the cercariae transform to schistosomula, which then enter a small vein or lymphatic vessel, circulate in the bloodstream through the lung capillaries, and are pumped via the heart to all parts of the body to reach the portal vein. There, the worms mature into adult males or females, pair, and migrate to their final location in the mesenteric or pelvic venous plexus.

The interval from cercarial penetration to sexual maturation and egg production, termed the *prepatent period*, lasts 5–7 weeks (up to 12 weeks for *S. haematobium*). The female worm then begins to produce eggs, which are excreted via feces or, for *S. haematobium*, urine. Approximately 50% of eggs are retained in tissue, where they are responsible for organ-specific morbidity (see "Pathogenesis," below). When excreted eggs reach water, they hatch and release a free-swimming larval stage (*miracidium*), which, after penetrating a host snail, undergoes several rounds of asexual multiplication. After ~4–6 weeks, infective cercariae are shed from the infected snails into the water. One snail, infected by one miracidium, can shed thousands of cercariae per day for several months; thus, the transmission potential of schistosomes is enormous. Environmental temperature is an important determinant of the development rate and the number of cercariae produced and released by the snails.

The schistosome egg **(Fig. 241-2)** is the only stage of the parasites' life cycle that can be detected in humans, either in excreta or in tissue biopsies. The eggs are large and can easily be distinguished morphologically from other helminth eggs. *S. haematobium* eggs are ~140 μm long, with a terminal spine; *S. mansoni* eggs are ~150 μm long, with a lateral spine; and *S. japonicum* eggs are smaller, rounder, and ~90 mm long, with a small lateral spine or knob.

Adult schistosomes are ~1–2 cm long. The male worm is flat, and the body forms a groove or gynecophoric canal in which the mature adult female is held like a sausage in a hotdog roll. Females are longer, thinner, and rounded. The females produce hundreds (African species) to thousands (Asian species) of eggs per day. Each ovum contains a ciliated miracidium larva, which secretes proteolytic enzymes that help the eggs to migrate into the lumen of the bladder (*S. haematobium*) or the intestine (other species). The lifespan of an adult schistosome averages 3–5 years but can be as long as 30 years. Schistosome worms feed on red blood cells; the debris is regurgitated in the host's blood, where it can be detected as circulating antigens (see "Diagnosis," below).

Adult schistosomes persist in the bloodstream for years and have evolved strategies of evading attack using immune effector mechanisms. This immune evasion is a result of several processes, such as binding of host proteins to the schistosome surface, which renders the parasite invisible to the host immune system.

The genome of schistosomes is relatively large (~300 Mb). Whole-genome sequences are available for *S. mansoni*, *S. japonicum*, and *S. haematobium*.

EPIDEMIOLOGY

Because of the complex life cycle of schistosomes, with snails as an intermediate host and humans as the final host, transmission is dependent on freshwater habitats that are suitable for the snails, are areas of human activity, and have climatic conditions favoring the survival of the snails and the development of the parasites inside the snail host. These requirements are reflected in the global distribution of

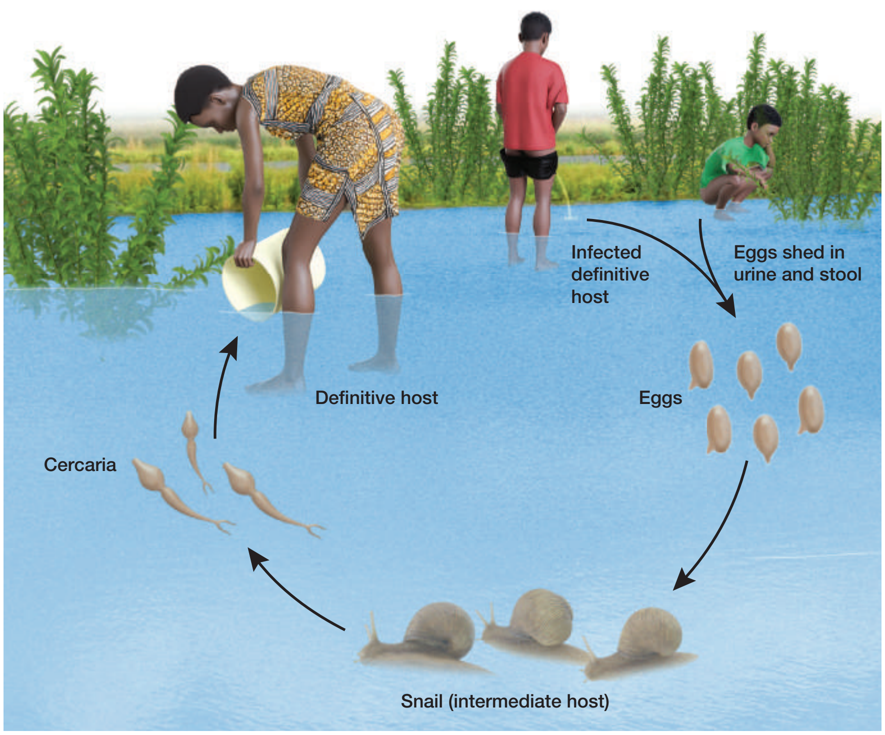

FIGURE 241-1 Life cycle of schistosomiasis.

schistosomiasis as well as in its microgeographic distribution within an endemic area. For *S. mansoni, S. haematobium*, *S. intercalatum*, and *S. guineensis*, humans are the most important definitive host. *S. japonicum* and *S. mekongi* are zoonotic parasites, with a wide range of definitive hosts such as pigs, water buffaloes, and various rodents.

It is estimated that at least 251.4 million people required preventive treatment in 2021, of whom more than 75.3 million people were reported to have been treated. Schistosomiasis transmission has been reported from 78 countries, of which 51 endemic countries have moderate to high transmission **(Fig. 241-3)**. More than 70% of infected people live in sub-Saharan Africa. Schistosomiasis is the most important of the neglected tropical diseases and is second only to malaria in public health impact. It is a poverty-related disease, and infection is prevalent in areas where adequate water supplies and sanitary facilities are lacking. In these areas, people come into contact with infested water through a variety of activities, including bathing, washing clothes, and collecting water for drinking or cooking. In some areas, adults have a high occupational risk of exposure; fishermen, canal cleaners, and workers in rice fields fall into this category. Among children, playing in water and swimming pose a risk. Large-scale irrigation and hydroelectric power operations can create suitable habitats for host snails and thus increase the risk of schistosomiasis transmission.

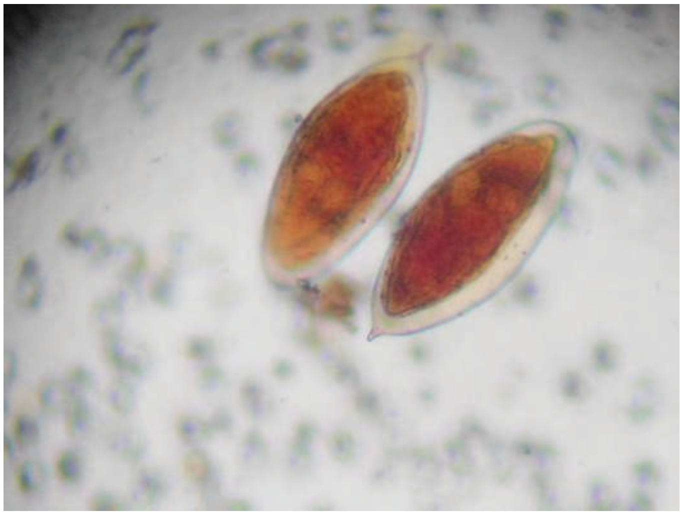

FIGURE 241-2 *Schistosoma haematobium* eggs.

In general, children living in endemic areas initially acquire infection at ~3–4 years of age—i.e., when they are old enough to walk and come into contact with infested water. However, infection does occur in much younger children. As children grow older, the prevalence and intensity of infection increase, peaking around puberty. A characteristic feature of schistosomiasis infection in human populations is a convex age–prevalence curve, with low prevalence in very young children, higher prevalence in older children with a peak at 10–15 years of age, and declining prevalence in adults. The same pattern is observed between age and intensity of infection and is attributable to various factors. Generally, children have more frequent, prolonged, and extensive water contact than adults through activities like playing and swimming. Furthermore, several studies have indicated that acquired immunity to schistosomiasis develops slowly over several years, so that adults are reinfected to a much lesser extent than children. These factors, combined with progressive spontaneous death of adult worms from infections acquired during childhood, lead to lower levels of infection in the adult population.

It is important to note that ongoing climate change, resulting in rising water temperatures and altered precipitation patterns, affects the distribution, reproduction, survival, and dispersal of the intermediate host snails, as well as the rate of parasite development within the snail. Therefore, climate change could considerably alter the distribution and abundance of the intermediate host snail and its schistosome parasites, resulting in notable shifts in the global distribution, disease dynamics, and transmission of schistosomiasis in the future.

PATHOGENESIS

Cercarial invasion may be associated with dermatitis arising from dermal and subdermal inflammatory reactions in response to dying cercariae that trigger innate immune responses. However, most manifestations of schistosomiasis—in the acute, established, and chronic phases of infection—are due to immunologic reactions to eggs retained in host tissues.

Around the time when oviposition commences, acute schistosomiasis (Katayama fever) may occur (see “Clinical Features,” below). Antigen excess from eggs results in the formation of soluble immune complexes, which may be deposited in several tissues and initiate a serum sickness–like illness. All evidence suggests that schistosome eggs, and not adult worms, induce the organ-specific morbidity caused by schistosome infections. Approximately half of the eggs are not excreted via feces or urine but are trapped in intestinal or hepatic tissue (*S. mansoni*, *S. japonicum*, and *S. mekongi*) or in the bladder and urogenital system (*S. haematobium*). The eggs induce a granulomatous host immune response composed primarily of lymphocytes, eosinophils, and alternatively activated macrophages. The lymphocytes produce various T_H2 cytokines such as interleukins 4, 5, and 13. Later, in the chronic phase of infection, regulatory cytokines are responsible for immunomodulation or downregulation of host responses to schistosome eggs and play an important role in reducing the size of granulomas.

When *S. mansoni* or *S. japonicum* eggs are swept into the small portal branches of the liver via the portal vein, they lodge in the presinusoidal periportal tissues. The formation of granulomas around the eggs can cause significant enlargement of the spleen and liver. High-intensity infections in children are often accompanied by hepatosplenomegaly that generally decreases over time, partly because the

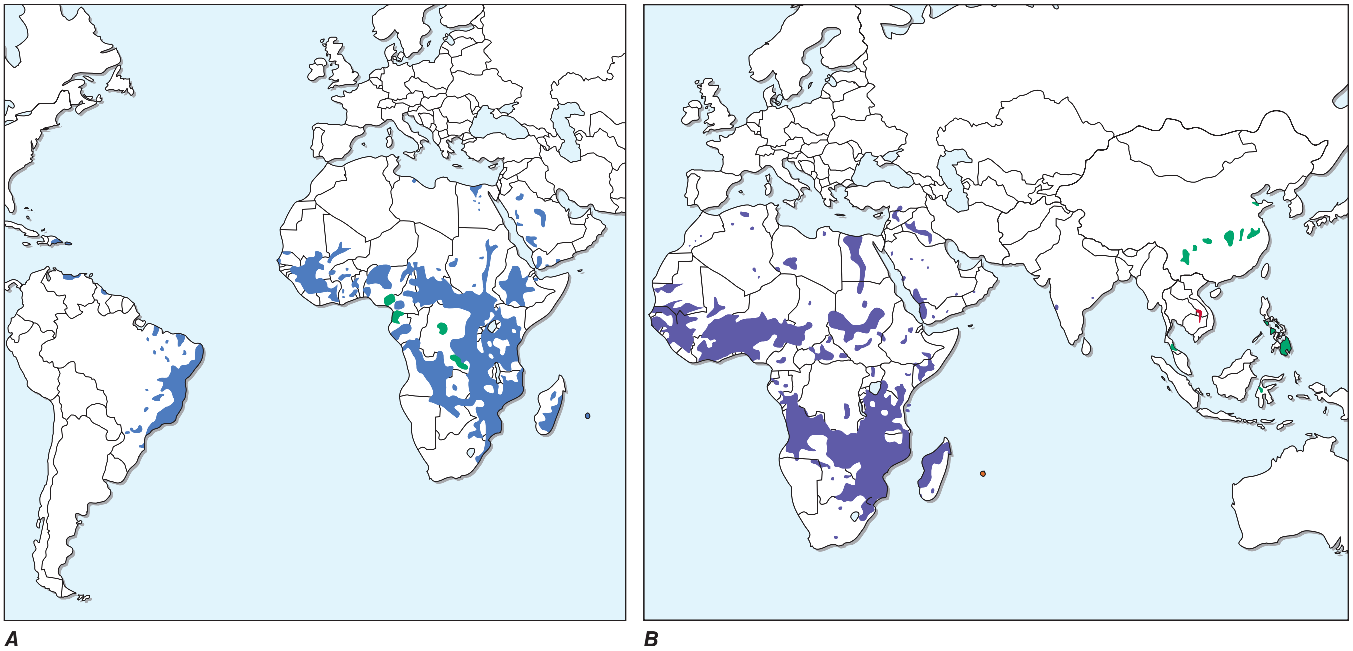

FIGURE 241-3 Global distribution of human schistosomiasis. ***A.*** *Schistosoma mansoni* infection (*dark blue*) is endemic in Africa, the Middle East, South America, and a few Caribbean countries. *S. intercalatum* infection (*green*) is endemic in sporadic foci in West and Central Africa. ***B.*** *Schistosoma haematobium* infection (*purple*) is endemic in Africa and the Middle East. The major endemic countries for *S. japonicum* infection (*green*) are China, the Philippines, and Indonesia. *Schistosoma mekongi* infection (*red*) is endemic in sporadic foci in Southeast Asia. *(Reprinted from CH King, AAF Mahmoud: Schistosomiasis and other trematode infections, in DL Kasper et al [eds], Harrison's Principles of Internal Medicine, 19th ed. New York, McGraw-Hill Education, 2015, pp 1423–1429.)*

number of eggs being deposited in the tissue gradually declines after the early teenage years as partial immunity to new infections develops and partly because of immunologic downregulation of the granulomatous response. However, in some infected individuals, egg-induced granulomatous responses lead to severe periportal fibrosis (*Symmers clay pipestem fibrosis*), with deposition of collagen around the portal vein, occlusion of the smaller portal branches, and severe, often irreversible, pathology. Occlusion of the portal branches may result in marked portal hypertension **(Fig. 241-4)**.

The signs and symptoms of *S. haematobium* infection relate to the worms' predilection for the veins of the urogenital plexus and result from deposition of eggs in the bladder, ureters, and genital organs. During established active infection, clusters of living eggs in the urogenital tissues can be found surrounded by intense inflammatory reactions and intense tissue eosinophilia. Movement of egg clusters into the lumen of the bladder is often followed by sloughing off of the epithelial surface, ulceration, and bleeding. Intense egg-induced tissue inflammation can result in bladder wall thickening and development of masses and pseudopolyps. Inflammation and granuloma formation around the ureteral ostia can lead to hydronephrosis.

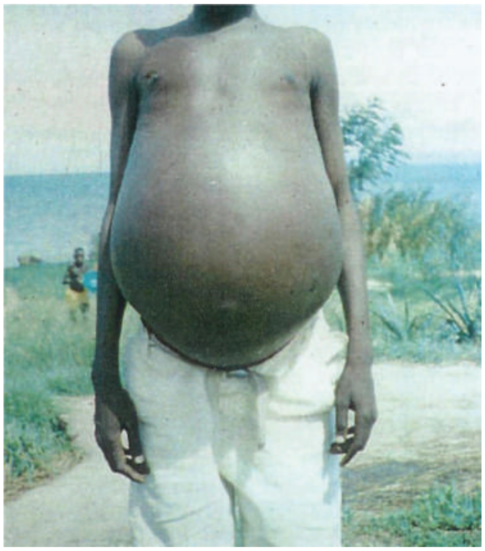

FIGURE 241-4 Portal hypertension with ascites. *(Photo courtesy of Birgitte J. Vennervald.)*

Generally, late chronic-stage infections are characterized by accumulation of dead calcified eggs in tissue. Characteristic cervical lesions are found in *S. haematobium* infections, including active-stage lesions with intense tissue inflammation around live eggs and chronic-stage sandy patches with clusters of calcified eggs.

CLINICAL FEATURES

In general, disease manifestations of schistosomiasis occur in three stages—acute, active, and chronic—according to the duration and intensity of infection.

Cercarial Dermatitis ("Swimmer's Itch") Cercarial penetration of the skin may result in a maculopapular rash called cercarial dermatitis or "swimmer's itch." Cercarial dermatitis can develop in people who have not previously been exposed to schistosomiasis (e.g., travelers), whereas it is rare among people living in endemic areas. A particularly severe form of cercarial dermatitis is commonly seen after exposure to cercariae from avian schistosomes. These cercariae cannot complete their development in humans and die in the skin, causing an inflammatory allergic reaction. This form of cercarial dermatitis can occur in people who have been in contact with water from lakes (e.g., in Europe or the United States) where various species of water birds, such as ducks, geese, and swans, are found. The rash may last for 1–2 weeks. This condition normally requires no treatment, but systemic antihistamines, topical antihistamines, or glucocorticoids can be used to reduce symptoms.

Acute Schistosomiasis (Katayama Fever) Symptomatic acute schistosomiasis, also known as Katayama fever or Katayama syndrome, is usually seen in travelers who have contracted the infection for the first time. The onset occurs between 2 weeks and 3 months after exposure to the parasite. The symptoms may appear suddenly and include fever, myalgia, general malaise and fatigue, headache, nonproductive cough, and intestinal symptoms such as abdominal tenderness or pain. Various combinations of these symptoms are often accompanied by eosinophilia and transient pulmonary infiltrates. Many patients recover spontaneously from acute schistosomiasis after 2–10 weeks, but the illness follows a more severe clinical course in some individuals, with weight loss, dyspnea, diarrhea, and hepatomegaly. Severe cerebral or spinal cord manifestations may occur, and even light infections may cause severe illness. The syndrome can, in rare cases, be fatal.

Differential diagnosis includes many other febrile infectious diseases with acute onset, including malaria, salmonellosis, and acute hepatitis. Fever and eosinophilia occur in trichinosis, tropical eosinophilia, invasive ankylostomiasis, strongyloidiasis, visceral larva migrans, and infections with *Opisthorchis* and *Clonorchis* species. Katayama fever is rare in people chronically exposed to infection in areas endemic for *S. mansoni* or *S. haematobium*.

Intestinal Schistosomiasis (*S. mansoni, S. japonicum*) In intestinal schistosomiasis, adult worms are located in the mesenteric veins, and disease manifestations are associated with parasite eggs passing through or becoming trapped in intestinal tissue. This event induces mucosal granulomatous inflammation with microulcerations, superficial bleeding, and sometimes pseudopolyposis. The symptoms tend to be more pronounced with a high intensity of infection and include intermittent abdominal pain, loss of appetite, and sometimes bloody diarrhea. The clinical manifestations of *S. intercalatum, S. guineensis*, and *S. mekongi* infection are generally milder.

Hepatosplenic Schistosomiasis Hepatosplenic schistosomiasis is caused by schistosome eggs trapped in liver tissue and occurs in *S. mansoni* and *S. japonicum* infections. There are two distinct clinical entities: early inflammatory hepatosplenomegaly and late hepatosplenic disease with periportal fibrosis.

Early inflammatory hepatosplenic schistosomiasis is the main entity seen in children and adolescents. The liver is enlarged, especially the left lobe, and is smooth and firm. The spleen is enlarged, often extending below the umbilicus, and is firm or hard. Generally, ultrasonography shows no hepatic fibrosis. This form of hepatosplenic schistosomiasis may be found in up to 80% of infected children. Its severity is closely associated with the intensity of infection and may also be associated with concomitant chronic exposure to malaria.

Late hepatosplenic schistosomiasis with periportal or Symmers fibrosis may develop in young and middle-aged adults with long-standing, high-level exposure to infection. Patients with periportal fibrosis may excrete very few or no eggs in feces. During the early stage, the liver is enlarged, especially the left lobe; it is smooth and firm or hard. The spleen is enlarged, often massively, and is firm or hard. The patient may report a left hypochondrial mass with discomfort and anorexia. Ultrasonography reveals typical periportal fibrosis and dilation of the portal vein. Other complications include delayed growth and puberty, especially in *S. japonicum* infections, and severe anemia. Severe hepatosplenic schistosomiasis may lead to portal hypertension, but hepatic function usually remains normal, even in cases with marked periportal fibrosis and portal hypertension.

Ascites, attributable both to portal hypertension and to hypoalbuminemia, may be seen, especially in *S. japonicum* infection. Patients with severe hepatosplenic disease and portal hypertension may develop esophageal varices detectable by endoscopy or ultrasound. These patients may experience repeated bouts of hematemesis, melena, or both. Hematemesis is the most severe complication of hepatosplenic schistosomiasis, and death may result from massive loss of blood.

Urogenital Schistosomiasis (*S. haematobium*) The signs and symptoms of *S. haematobium* infection relate to the worms' predilection for the veins of the urogenital tract. Two stages of infection are recognized. An active stage occurring mainly in children, adolescents, and younger adults is characterized by egg excretion in the urine, with proteinuria and macroscopic or microscopic hematuria and deposition of eggs in the urinary tract. A chronic stage in older individuals is characterized by sparse or no urinary egg excretion despite urogenital tract pathology.

A characteristic sign in the active stage is painless, terminal hematuria. Dysuria and suprapubic discomfort or pain are associated with active urogenital schistosomiasis and may persist throughout the course of active infection. Eggs deposited in the bladder mucosa may give rise to an intense inflammatory response of the bladder wall, which may cause ureteric obstruction and lead to hydroureter and hydronephrosis. These early inflammatory lesions, including obstructive uropathy, can be visualized by ultrasonography.

As the infection progresses, the inflammatory component decreases and fibrosis becomes more prominent. The symptoms at this stage are nocturia, urine retention, dribbling, and incontinence. Cystoscopy reveals "sandy patches" composed of large numbers of calcified eggs surrounded by fibrous tissue and an atrophic mucosal surface. The ureters are less commonly involved, but ureteral fibrosis can cause irreversible obstructive uropathy that can progress to uremia.

Egg deposition may cause granulomas and lesions in the genital organs, most commonly in the cervix and vagina in women and the seminal vessels in men. The results may include dyspareunia, abnormal vaginal discharge, contact bleeding, and lower back pain in women and perineal pain, painful ejaculation, and hematospermia in men. Genital symptoms like bloody discharge and genital itch are associated with *S. haematobium* infection in school-aged girls living in schistosomiasis-endemic areas. Symptoms such as hematospermia and perineal discomfort have been described in travelers, and eggs have been demonstrated in seminal fluid. An association between female genital schistosomiasis and HIV infection has been demonstrated, but the impact of genital schistosomiasis on HIV transmission needs further elucidation.

S. haematobium has been classified by the International Agency for Research on Cancer (IARC) as definitely carcinogenic to humans (i.e., a group 1 carcinogen). Chronic *S. haematobium* infection is associated with squamous cell carcinoma of the urinary bladder.

Other Manifestations Worms and eggs can sometimes be located in ectopic sites, causing site-specific manifestations and symptoms. Neuroschistosomiasis is one of the most severe clinical forms of schistosomiasis and is caused by the inflammatory response around eggs in the cerebral or spinal venous plexus. *S. mansoni* and *S. haematobium* worms can end up in the spinal venous plexus, where they may cause transverse myelitis—an acute complication sometimes seen in travelers returning home with schistosomiasis. *S. japonicum* is mainly associated with granulomatous lesions in the brain, causing epileptic seizures, encephalopathy with headache, visual impairment, motor deficit, and ataxia. Pulmonary schistosomiasis is caused by portacaval shunting of eggs into the lung capillaries, where they induce granulomas in the perialveolar area. The consequences may be fibrosis, pulmonary hypertension, and cor pulmonale.

DIAGNOSIS

Anamnestic information on recent travels to endemic areas and exposure to freshwater bodies through recreational or other activities is important in the diagnosis of schistosomiasis in travelers. Information about exact geographic locations can facilitate identification of the relevant species of *Schistosoma*. Eosinophilia is a common finding and is often associated with helminthic infections such as schistosomiasis.

Detection of schistosome eggs in stool or urine is indicative of active infection and is the standard diagnostic method. The diagnosis is often based on the detection of eggs in a fixed small amount of excreta—e.g., 50 mg of stool or filtration of 10 mL of urine. This method is widely used among populations in endemic areas and allows quantitation of the level of infection (eggs per gram of feces or per 10 mL of urine). However, levels of egg excretion in people from nonendemic areas may be very low, in which case a larger sample and concentration methods (e.g., formol-ether concentration) may be needed.

Eggs can also be detected in rectal biopsies (both *S. mansoni* and *S. haematobium*) and occasionally in Pap smears and semen samples (*S. haematobium*). Polymerase chain reaction (PCR)–based detection of parasite DNA in stool or urine is more sensitive than parasitologic methods and is increasingly used. *Schistosoma* DNA can be detected in cerebrospinal fluid samples for diagnosis of neuroschistosomiasis.

Serology, with detection of specific antibodies to schistosomes, is useful in travelers but less so in people from endemic areas where transmission is ongoing. The serologic assays employed at the CDC are a Falcon assay screening test/enzyme-linked immunosorbent assay (FAST-ELISA) using *S. mansoni* adult microsomal antigen and a confirmatory species-specific immunoblot assay performed in light of the patient's travel history.

Schistosome proteoglycans—circulating anodic and cathodic antigens (CAAs and CCAs)—regurgitated into the bloodstream by the feeding worms can be detected in serum and urine by ELISA or monoclonal antibody–based lateral flow assays. The presence of CAA or CCA is an indication of active infection, and levels of these antigens correlate well with the intensity of infection. However, detection of CAAs and CCAs is not currently suitable for diagnosis in travelers, who are likely to have low levels of infection and very few worms, but promising results have been obtained using an ultrasensitive lateral flow assay. A commercially available point-of-care assay (Rapid Medical Diagnostics, Pretoria, South Africa) that detects CCA in urine is now widely used for screening of infected communities in relation to mass drug administration programs.

TREATMENT

Schistosomiasis

The drug of choice for treatment of schistosomiasis is praziquantel. It is administered orally, is available as 600-mg tablets, and is effective against all schistosome species infecting humans. The drug is safe and well tolerated. Standard regimens are shown in **Table 241-2**. In patients who are not cured by initial treatment, the same dose can be repeated at weekly intervals for 2 weeks. Since praziquantel does not affect the young migrating stages of the schistosomes, it may be necessary to repeat the dose 6–12 weeks later, especially if eosinophilia or symptoms persist despite treatment.

As a general principle, all patients with acute schistosomiasis should be treated with praziquantel. Glucocorticoids can be added in Katayama fever to suppress the hypersensitivity reaction. However, treatment for acute schistosomiasis or Katayama fever must be adjusted appropriately for each case, and in the most severe cases, management in an acute-care setting is necessary.

Praziquantel is effective in cerebral *S. japonicum* infections, resulting in rapid dissipation of cerebral edema and resolution of cerebral masses. However, glucocorticoids and anticonvulsants are sometimes needed in neuroschistosomiasis.

The effect of antischistosomal treatment on disease manifestations depends on the stage and severity of the lesions. Early hepatosplenomegaly, mild or moderate fibrosis, and urinary bladder lesions seen during active infection resolve after chemotherapy. However, for late-stage manifestations (e.g., severe fibrosis with portal hypertension), praziquantel treatment is only one component of management, since the main complications are due to obstructive pathology. Management of portal hypertension and prevention of bleeding from esophageal varices should follow clinical guidelines for treatment of these conditions.

TABLE 241-2 Treatment of Schistosomiasis and Food-Borne Trematode Infections

INFECTION	DRUG OF CHOICE	ADULT DOSE[a]
Schistosoma mansoni, *S. haematobium,* *S. intercalatum, S. guineensis*	Praziquantel[b]	40 mg/kg PO in 2 divided doses for 1 day
S. japonicum, S. mekongi	Praziquantel	60 mg/kg PO in 3 divided doses for 1 day
Clonorchis sinensis, *Opisthorchis viverrini,* *Opisthorchis felineus*	Praziquantel	25 mg/kg PO tid for 2 consecutive days
Fasciola hepatica, Fasciola gigantica	Triclabendazole[c]	2 doses of 10 mg/kg PO given 12 h apart
Fasciolopsis buski	Praziquantel	75 mg/kg PO in 3 divided doses for 1 day
Echinostoma spp., *Heterophyes heterophyes,* several other species	Praziquantel	25 mg/kg PO tid
Paragonimus westermani, *Paragonimus kellicotti*	Praziquantel Triclabendazole[c]	25 mg/kg PO tid for 2 consecutive days 10 mg/kg PO once (or twice, 12–24 h apart)

[a] The pediatric dose is the same as the adult dose in all instances. [b] The safety of praziquantel in children <4 years old has not been established, although many children in this age group have been treated with praziquantel during mass drug-administration programs. [c] In February 2019, the U.S. Food and Drug Administration (FDA) approved triclabendazole for treatment of fascioliasis in patients at least 6 years of age.

PREVENTION AND CONTROL

Schistosomiasis is contracted through direct contact with infested freshwater. Travelers should be made aware of the risk of infection if they come into contact with freshwater sources in schistosomiasis-endemic areas. For people living in rural areas where schistosomiasis is endemic, it may be very difficult, if not impossible, to avoid water contact—for example, during occupational activities such as fishing and working in rice fields. Schistosomiasis is a poverty-related disease, and access to safe water and good sanitary facilities may rarely be available. Because *S. japonicum* is a zoonotic parasite, preventive measures should target not only the human population but also animals such as water buffalo, which act as reservoirs for infection.

Praziquantel treatment of infected people, often during mass drug-administration programs, is a cornerstone of the management and control of schistosomiasis. Regular treatment will reduce the level of schistosomiasis morbidity in affected populations. However, treatment should be combined with other relevant strategies, such as control of the intermediate host snails, improved water-quality and sanitation facilities, and health education. Schistosomiasis control measures should be integrated into local health programs.

There have been intensive efforts to develop vaccines, but none is yet available. Three vaccine candidates have been tested in clinical phase 1 or 2 trials. Only one candidate, *S. haematobium* 28GST, has been tested in a clinical phase 3 trial in populations living in an endemic area. The vaccine candidate was immunogenic and well tolerated by infected children, but a sufficient efficacy was not reached.

FOOD-BORNE TREMATODE INFECTIONS

Food-borne trematode infections are a group of zoonotic diseases caused by hepatic, intestinal, and pulmonary parasitic flukes. These infections are contracted by ingestion of infective parasites in undercooked aquatic food or water plants. Infections can cause severe liver and lung disease, and together, this may result in 2 million life-years lost to disability and death worldwide every year.

LIVER FLUKES

The most important liver flukes causing human infections are the related species *Opisthorchis viverrini* and *Opisthorchis felineus*, which cause opisthorchiasis; *Clonorchis sinensis*, which causes clonorchiasis; and *Fasciola hepatica* and *Fasciola gigantica*, which cause fascioliasis (Table 241-1).

Opisthorchiasis and Clonorchiasis *O. viverrini* is found mainly in northeastern Thailand, Laos, and Cambodia; *O. felineus* mainly in Europe and Asia, including the former Soviet Union; and *C. sinensis* in Asia, including Korea, China, Taiwan, Vietnam, Japan, and Asian regions of Russia. Parasite eggs excreted from infected humans or animals are ingested by a host snail (the first intermediate host), where they undergo several developmental stages. Cercariae are then released from the snail and penetrate freshwater fish (the second intermediate host), encysting as metacercariae in the muscles or under the scales. Humans become infected by eating raw or undercooked fish from endemic countries. After ingestion, the metacercariae excyst in gastric juices and migrate via the duodenum, the ampulla of Vater, and the extrahepatic biliary system to the intrahepatic bile ducts.

TABLE 241-3 Clinical Features of Food-Borne Trematode Infections

	SYMPTOMS OR SIGNS		
INFECTION	**EARLY OR ACUTE STAGE**	**ESTABLISHED OR CHRONIC STAGE**	**COMPLICATIONS**
Liver Flukes			
Clonorchis sinensis, Opisthorchis viverrini, Opisthorchis felineus	Often asymptomatic; sometimes hepatitis-like symptoms and high fever (especially with *O. felineus*)	Biliary colic, cholestatic jaundice, recurrent cholangitis and cholelithiasis; hepatomegaly, gallbladder enlargement, periductal fibrosis. Light infections are often asymptomatic and remain so for years.	Pancreatitis, cholangiocarcinoma[a]
Fasciola hepatica, Fasciola gigantica	Acute onset (1–4 weeks after infection) with high fever, weight loss, sometimes urticaria and liver tenderness	Biliary colic, cholestatic jaundice, recurrent cholangitis and cholelithiasis; thickening, enlargement, and fibrosis of biliary ducts; sometimes repeated relapses of acute symptoms	Pancreatitis. In rare cases: ectopic infections in the central nervous system, orbital area, gastrointestinal tract, lungs, and other organs. Rarely, fascioliasis can be fatal.
Intestinal Flukes			
Fasciolopsis buski, Echinostoma spp., *Heterophyes heterophyes*, several other species	Often asymptomatic; sometimes nonspecific gastrointestinal symptoms	Heavy infection may lead to ulceration of intestinal mucosa and malabsorption. Mild infections are often asymptomatic.	Malnutrition, anemia; rarely, ectopic infection in the central nervous system
Lung Flukes			
Paragonimus westermani, Paragonimus kellicotti	Often asymptomatic; sometimes insidious onset with anorexia and weight loss	Bronchitis-, asthma-, and tuberculosis-like symptoms and signs such as chronic cough, dyspnea, bloody ("rusty") sputum	Pulmonary cyst formation; ectopic infection in the central nervous system, eyes, skin, heart, abdominal and reproductive organs

[a] Carcinogenesis has not yet been established for *O. felineus.*

The clinical manifestations of infection with *Opisthorchis* species and *C. sinensis* are similar. Pathologic changes are typically seen in the bile ducts, liver, and gallbladder (Table 241-3). Tissue damage and intense inflammation are caused by mechanical and chemical irritation and immune responses to worms or worm products, and chronic inflammation may result in the development of cholangiocarcinoma. Both *O. viverrini* and *C. sinensis* are classified by the IARC as definitely carcinogenic (class 1). Acute and light infections are mostly asymptomatic, but hepatitis-like signs and symptoms, with high fever and chills, have been reported, especially in *O. felineus* infections. In general, only heavily infected people have symptoms and severe complications (Table 241-3).

The diagnosis of these infections is based on microscopic identification of parasite eggs in stool specimens. The eggs of *Opisthorchis* are indistinguishable from those of *Clonorchis*.

Fascioliasis Fascioliasis occurs in many areas of the world and usually is caused by *Fasciola hepatica*, a common liver fluke of sheep and cattle. *F. hepatica* is found in more than 50 countries on all continents except Antarctica; *F. gigantica* is less widespread. The areas with the highest known rates of human *Fasciola* infection are in the Andean highlands of Bolivia and Peru. In other areas where fascioliasis is found, human cases are sporadic.

Unlike the other liver flukes, *Fasciola* species have no second intermediate host, as their infectious metacercariae adhere directly to aquatic plants. Humans usually acquire infection by ingesting aquatic plants, such as watercress, that contain viable metacercariae or by drinking water with free metacercariae.

After metacercariae have excysted in the duodenum, *Fasciola* species migrate through the intestinal wall into the body cavity, penetrate the liver capsule, and move through the liver into the bile ducts. This migration route is different from that of other liver flukes and gives rise to symptoms during the acute migratory phase; the parasites may cause tissue destruction, focal bleeding, and inflammation. Some migrating flukes may deviate from their usual route to cause ectopic infections. In the established latent stage of infection, the parasites may cause bile duct inflammation, resulting in thickening and expansion of the ducts, fibrosis, and ultimately biliary obstruction (Table 241-3). Although some infected people are asymptomatic in the latent phase, others may experience repeated relapses of acute manifestations.

The most widely used diagnostic approach is direct detection of *Fasciola* eggs by microscopic examination of stool or of duodenal or biliary aspirates. Eggs generally cannot be detected until 3–4 months after exposure, whereas antibodies to the parasite may become detectable 2–4 weeks after exposure. More than one stool specimen may be needed for diagnosis, especially in light infections.

■ INTESTINAL FLUKES

More than 70 species of intestinal flukes can cause human infection. These parasites are found in different geographic areas, with a relatively high prevalence in Southeast Asia. Humans are infected by ingestion of infective metacercariae attached to aquatic plants (*Fasciolopsis buski*) or encysted in freshwater fish. Flukes mature in the human intestines, and eggs are passed with feces. Mechanical irritation of the intestinal wall and inflammation may lead to nonspecific gastrointestinal symptoms such as diarrhea, constipation, and abdominal pain. Most individuals infected with intestinal flukes are asymptomatic, but heavy infections can be severe, with intestinal mucosal ulcerations and malabsorption (Table 241-3). The diagnosis is established by detection of eggs in stool samples. However, eggs from various intestinal trematodes are often morphologically similar, and it is very difficult to distinguish among species. A cautionary note: *Fasciola* eggs can be difficult to distinguish on the basis of morphologic criteria from the eggs of the intestinal fluke *F. buski*. The distinction has implications for therapy: infection with *F. buski* is treated with praziquantel, which is not effective against fascioliasis (Table 241-2).

■ LUNG FLUKES

Paragonimiasis is a parasitic lung infection caused by lung flukes of the genus *Paragonimus*. It is a food-borne parasitic zoonosis, with most cases reported from Asia and attributable to consumption of raw or undercooked freshwater crustaceans. *Paragonimus westermani* and related species (e.g., *Paragonimus africanus*) are endemic in West Africa, Central and South America, and Asia. The United States has one indigenous species of lung fluke, *Paragonimus kellicotti.*

Paragonimus species require two intermediate hosts: first, a freshwater snail; and second, a freshwater crustacean, such as a freshwater crab. Humans are infected by consuming raw or undercooked infected crustaceans containing *Paragonimus* metacercariae. *Paragonimus* infects other carnivores such as cats, dogs, foxes, rodents, and pigs in addition to humans. After ingestion, metacercariae quickly penetrate the duodenum and traverse the peritoneal cavity, diaphragm, and parietal pleura to mature into hermaphroditic worm pairs in the pleural spaces or lungs within 6–10 weeks. Adults cross-fertilize in cystic cavities in the pleural spaces or lungs within another 4–16 weeks and release unembryonated eggs into bronchioles. The eggs are then coughed up in bloody ("rusty") sputum and either discharged in sputum or swallowed and later excreted in feces. Unembryonated eggs are passed from the mammalian host into freshwater ecosystems, where they infect intermediate host snails.

The symptoms and signs of paragonimiasis are fever, cough, hemoptysis, and peripheral eosinophilia. Some patients with paragonimiasis and low parasite burdens may remain relatively asymptomatic for prolonged periods or may have recurrent attacks of cough, sputum production, fever, and night sweats that mimic tuberculosis. Infective metacercariae may migrate to extrapulmonary sites such as the brain (cerebral paragonimiasis).

Pulmonary paragonimiasis is diagnosed by detection of parasite ova in sputum and/or feces. Serology can be helpful in egg-negative cases and in cerebral paragonimiasis. Anamnestic information about the consumption of raw or undercooked freshwater crabs by immigrants, expatriates, and returning travelers—and, in the United States, the consumption of raw or undercooked crayfish from freshwater river systems where *P. kellicotti* is endemic—is important in patients presenting with fever, cough, hemoptysis, pleural effusions, and peripheral eosinophilia.

TREATMENT

Food-Borne Trematode Infections

Praziquantel and triclabendazole are the two drugs of choice; Table 241-2 summarizes the dosages recommended for the various trematode infections. All confirmed cases of human paragonimiasis should be treated with praziquantel (Table 241-2) to avoid the complications of extrapulmonary disease. Surgical management may be needed for pulmonary or cerebral lesions.

■ CONTROL AND PREVENTION

Drugs are currently the main method of controlling the morbidity associated with food-borne trematode infections, but integrated programs (including improved sanitation; food inspections; and information, education, and communication campaigns) are important for sustainable disease control. Collaboration with other sectors (e.g., agricultural, environmental, and educational) is necessary to tackle highly complex situations in which human behavior, biological factors, and agricultural practices all play a role.

■ FURTHER READING

Andrade G et al: Decline in infection-related morbidities following drug-mediated reductions in the intensity of *Schistosoma* infection: A systematic review and meta-analysis. PLoS Negl Trop Dis 11:e0005372, 2017.

Cucchetto G et al: High-dose or multi-day praziquantel for imported schistosomiasis? A systematic review. J Travel Med 26:taz050, 2019.

De Leo GA et al: Schistosomiasis and climate change. BMJ 371:m4324, 2020.

Fried B, Abruzzi A: Food-borne trematode infections of humans in the United States of America. Parasitol Res 106:1263, 2010.

Fürst T et al: Global burden of human food-borne trematodiasis: A systematic review and meta-analysis. Lancet Infect Dis 12:210, 2012.

Jordan P et al (eds): *Human Schistosomiasis*. CAB International, Wallingford, 1993.

Keiser J, Utzinger J: Food-borne trematodiases. Clin Microbiol Rev 22:466, 2009.

Mcmanus DP et al: Schistosomiasis. Nat Rev Dis Primers 4:13, 2018.

Ross AG et al: Katayama syndrome. Lancet Infect Dis 7:218, 2007.

Sripa B et al: Update on pathogenesis of opisthorchiasis and cholangiocarcinoma. Adv Parasitol 102:97, 2018.

World Health Organization: *Female Genital Schistosomiasis: A Pocket Atlas for Clinical Health-Care Professionals*. Geneva, World Health Organization, 2015. Available at *http://brightresearch.org/wp-content/uploads/2016/05/FGS-pocket-atlas_eng.pdf*. WHO/HTM/NTD/2015.4, 2015. Accessed March 16, 2020.

242

Cestode Infections

A. Clinton White, Jr., Miguel M. Cabada

Cestodes, or tapeworms, are members of the flatworm phylum (Platyhelminthes) and comprise the subphylum (Cestoda). The adult worms are segmented worms found in the gastrointestinal tract of the definitive host. The larval forms are typically cystic and found in the tissues of the intermediate host. Human cestode infections include both tapeworm and larval-form infections. For tapeworm infections, humans are the definitive hosts, with adult tapeworms living in the gastrointestinal tract (*Taenia saginata*, *Dibothriocephalus latus*, and *Dipylidium caninum*). Humans are an intermediate or dead-end host with larval-stage parasites living in the tissues for *Echinococcus* spp., *Spirometra* spp., and *Taenia multiceps*. Humans may be either the intermediate hosts or the definitive hosts for *Taenia solium* and are the hosts of both stages of *Rodentolepis nana* (formerly *Hymenolepis nana*).

The tapeworm forms are typically elongated and ribbon-shaped and attach to the intestinal mucosa by means of sucking cups or hooks located on the scolex. A short, narrow neck is found at the base of the scolex from which proglottids (segments) form. As proglottids mature, they are displaced from the neck by the formation of new, less mature segments. The elongating chain of attached proglottids, called the strobila, constitutes the bulk of the tapeworm. The length of the strobila varies between species. Some tapeworms consist of more than 1000 proglottids and may be several meters long. Mature proglottids are hermaphroditic and produce eggs, which are intermittently released.

Most human tapeworms require at least one intermediate host to complete the life cycle. After ingestion of the eggs or proglottids by an intermediate host, the eggs are activated to release the invasive larvae (oncospheres). The oncosphere penetrates the intestinal mucosa and migrates to tissues and develops into an encysted form known as a cysticercus (single scolex), a coenurus (multiple scolices), or a hydatid (cyst with daughter cysts, each containing multiple protoscolices). After ingestion of the cystic forms by the definitive host, the scolex evaginates and develops into a tapeworm.

■ *TAENIASIS SAGINATA* AND *TAENIASIS ASIATICA*

The beef tapeworm *T. saginata* is found in all countries where raw or undercooked beef is eaten. It is most prevalent in subSaharan African and Middle Eastern countries. *Taenia asiatica* is closely related to *T. saginata* and is found in Asia, with pigs as intermediate hosts. Since the clinical manifestations and morphology of the two species are very similar, they are discussed together.

Etiology and Pathogenesis Humans serve as the definitive host for the adult stage of *T. saginata* and *T. asiatica*. The scolex attaches to the small intestines and the strobila with 1000–2000 proglottids can reach 8 m in length. The scolex of *T. asiatica* has an unarmed (with no hooks) rostellum and four suckers, whereas *T. saginata* has no rostellum but attaches via four prominent suckers. Each gravid segment has 15–30 uterine branches (in contrast to 8–12 for *T. solium*). The eggs of the three human-infecting *Taenia* species are indistinguishable morphologically and are 30–40 μm in diameter, contain the oncosphere with six hooklets, and have a thick brown striated shell. Eggs can live for months to years on vegetation until they are ingested by cattle or other herbivores (*T. saginata*) or pigs (*T. asiatica*). After ingestion, the oncosphere egresses the egg, invades into the intestinal wall, and is distributed throughout the body via the bloodstream. The invasive larva transforms into the cysticercus (cystic forms) in muscle or viscera of the intermediate host. When raw or undercooked meat containing the larvae is ingested, the cysticercus evaginates and forms a tapeworm in the human intestines. The adult worm matures over several months to produce eggs. The proglottids or eggs are then shed intermittently with stool.

Clinical Manifestations Patients may be asymptomatic or may note passing proglottids in their feces. Patients may note the sensation of something moving while the motile proglottids of *T. saginata* are passed. Patients occasionally note abdominal pain or discomfort, nausea, change in appetite, or weakness. Weight loss is unusual.

Diagnosis Diagnosis depends on detection of eggs or proglottids in the stool. Eggs may not be found by stool examination but are sometimes only present in the perianal area; thus, when suspected, the perianal region should be examined with use of a cellophane-tape swab (as in pinworm infection; **Chap. 239**). Antigen-detection assays are more sensitive than microscopic examination but are not commercially available. Distinguishing *T. saginata* or *T. asiatica* from *T. solium* requires examination of mature proglottids or the scolex or via molecular tests. Eosinophilia and elevated levels of serum IgE are usually absent.

TREATMENT

Taeniasis saginata and *Taeniasis asiatica*

A single dose of praziquantel (10 mg/kg) is highly effective. Niclosamide (adult dose, 2 g; 1 g for children weighing 11–34 kg) is also effective but is less available. Nitazoxanide can also be used.

Prevention Adequate cooking of beef and pork viscera or exposure to temperatures as low as 56°C for 5 min will destroy cysticerci. Refrigeration, salting for long periods, or freezing at −10°C for 9 days also kills the parasites in beef. General preventive measures include inspection of beef and pork viscera and proper disposal of human feces.

■ *TAENIASIS SOLIUM* AND CYSTICERCOSIS

T. solium, also known as the pork tapeworm, causes two forms of infection in humans: adult tapeworms in the intestine (taeniasis) or larval forms in the tissues (cysticercosis). Humans are the only hosts for the *T. solium* tapeworm; pigs are the usual intermediate hosts, although other animals may rarely harbor the cysticercus forms.

T. solium has a wide global distribution including all areas where pigs are raised with access to human feces. Cysticercosis is highly prevalent in Latin America, subSaharan Africa, China, India, and Southeast Asia. Cysticercosis also occurs in nonendemic nations due to the immigration of tapeworm carriers or cysticercosis-infected persons from endemic areas.

Etiology and Pathogenesis The tapeworms of *T. solium* reside primarily in the small intestines. The scolex attaches to the mucosa via sucking disks and the armed rostellum with two rows of hooklets. The mature tapeworm can reach a length of 3 m and have up to 1000 proglottids. The adult tapeworms are thought to live for just few years. Each mature proglottid may produce approximately 50,000 eggs. Proglottids are released from the terminal end of the tapeworm intermittently and excreted into the feces. The eggs are immediately infective for both humans and pigs. After ingestion of eggs by the intermediate host, the oncospheres are released, penetrate the intestinal wall, and are carried via the bloodstream to tissues. In the pig, larvae usually mature in striated muscle of the neck, tongue, and trunk. By 60–90 days, the invasive larvae transform into the encysted larval stage. In the pigs, cysticerci typically survive for months to years, until the pig is slaughtered. Humans develop intestinal tapeworm infections after ingestion of contaminated pork. Human cysticercosis follows ingestion of *T. solium* eggs. Transmission is associated with close contact with a tapeworm carrier. The eggs are sticky and may be found under the fingernails of tapeworm carriers. The tapeworm carrier can also infect themselves by ingestion of ova, likely via the fecal-oral route.

Clinical Manifestations Intestinal infections (taeniasis) with *T. solium* are often asymptomatic. Some patients note passage of proglottids in stool. The proglottids are typically off-white in color and 1–3 cm in length, 0.5–1 cm wide, and about 1 mm thick. The clinical manifestations of cysticercosis are variable. Cysticerci can be found anywhere in the body but are commonly detected in the central nervous system (CNS), skeletal muscle, subcutaneous tissue, or the eye. Involvement of the brain, spine, or cerebrospinal fluid (CSF) is termed neurocysticercosis. The clinical manifestations of human cysticercosis vary with the location of the cysticerci as well as with the extent of associated inflammatory responses or scarring. The most common clinical manifestations are neurologic symptoms. Headache is common with all CNS forms of disease. Seizures are the most frequent clinical manifestation and are associated with inflammation of the brain parenchyma surrounding the cysticercus. Seizures may be focal, focal with secondary generalization, or generalized. Another common clinical manifestation is with symptoms of hydrocephalus, which may result from CSF flow obstruction by cysticerci and/or accompanying inflammation or communicating hydrocephalus from arachnoiditis. The symptoms of increased intracranial pressure may include headache, nausea and vomiting, dizziness, and ataxia. Patients with hydrocephalus may present with altered mental status or papilledema with altered visual acuity. Some patients present with intermittent acute hydrocephalus (termed Bruns' syndrome) associated with change in position due to the cysticercus working as a ball valve. Cysticerci at the base of the brain or in the subarachnoid space may cause chronic meningitis or arachnoiditis, communicating hydrocephalus, mass lesions, hemorrhages, or strokes.

Diagnosis The diagnosis of tapeworm infection with *T. solium* is made by the demonstration of eggs or proglottids, as described for *T. saginata*. However, eggs and proglottids are only shed intermittently, limiting the sensitivity of direct testing. Antigen-capture enzyme-linked immunosorbent assay (ELISA), polymerase chain reaction (PCR), and serology for tapeworm stage-specific antigens are more sensitive but are only available as research techniques. The diagnosis of neurocysticercosis can be difficult since the symptoms are nonspecific and there is no readily available material for demonstration of the parasite. A group of international experts proposed revised diagnostic criteria **(Table 242-1)**. The diagnosis is only certain with definite demonstration of the parasite (absolute criteria). Definitive diagnosis is possible with histologic observation of the parasite in excised tissue, by funduscopic visualization of the parasite in the subretinal or vitreous spaces of the eye, or by neuroimaging demonstrating a cystic lesion containing a characteristic scolex **(Fig. 242-1)**. With high-resolution neuroimaging, the scolex can often be identified. In other cases, a clinical diagnosis is based on a combination of clinical presentation, radiographic studies, exposure or evidence demonstrating presence of the parasites by antigen-detection, quantitative PCR, or even next-generation sequencing in spinal fluid.

Neuroimaging is the primary major diagnostic method (Fig. 242-1). Demonstration of a cystic lesion with a mural nodule consistent with a scolex ("dot-in-hole") is diagnostic. Major findings include cystic lesions with or without enhancement (e.g., ring enhancement), one or more nodular calcifications (which may also have associated edema or enhancement), focal enhancing lesions, or cystic lesions in the subarachnoid space. Cysticerci in the brain parenchyma are usually 5–20 mm in diameter and round. Cystic lesions in the subarachnoid space or fissures may enlarge up to 6 cm in diameter and may be lobulated. The cyst wall for cysticerci in the subarachnoid space or ventricles is usually very thin, and the cyst fluid is often isodense with CSF. Thus, obstructive hydrocephalus or enhancement of the basilar meninges may be the only finding on computed tomography (CT) in extraparenchymal neurocysticercosis. However, since these findings are less specific, they are considered only minor criteria. Cysticerci in the ventricles or subarachnoid space are more readily identified by magnetic resonance imaging (MRI), especially fast imaging employing steady-state acquisition (FIESTA) or three-dimensional constructive interference in steady state (3D CISS). CT is more sensitive than MRI in identifying calcified lesions, whereas MRI is more sensitive than CT for identifying small cystic lesions, scolexes, and enhancement. Spontaneous resolution, resolution after albendazole therapy, or mobile cystic lesions within the ventricles are findings that can support the diagnosis of neurocysticercosis.

Exposure history significantly modifies the interpretation of neuroimaging studies. Detection of specific antibodies to or antigens of *T. solium* are major exposure criteria. Antibody tests using unfractionated

TABLE 242-1 Revised Diagnostic Criteria for Neurocysticercosis[a]

1. **Absolute criteria**
   - **a.** Histologic demonstration of the parasite from biopsy of a brain or spinal cord lesion
   - **b.** Visualization of subretinal cysticercus
   - **c.** Conclusive demonstration of a scolex with a cystic lesion on neuroimaging studies
2. **Neuroimaging criteria**
   - **a. Major neuroimaging criteria**
     - Cystic lesions without a discernible scolex, typical small enhancing lesions, multilobulated cystic lesions in the subarachnoid space, typical parenchymal brain calcifications
   - **b. Confirmative neuroimaging criteria**
     - Resolution of cystic lesions spontaneously or after cysticidal drug therapy
     - Migration of ventricular cysts documented on sequential neuroimaging studies
   - **c. Minor neuroimaging criteria**
     - Obstructive hydrocephalus or abnormal enhancement of basal leptomeninges
3. **Clinical/exposure criteria**
   - **a. Major clinical/exposure criteria**
     - Detection of specific anticysticercal antibodies (e.g., by enzyme-linked immunoelectrotransfer blot [EITB]) or cysticercal antigens by well-standardized immunodiagnostic tests
     - Cysticercosis outside the central nervous system
     - Evidence of a household contact with *T. solium* infection
   - **b. Minor clinical/exposure criteria**
     - Clinical manifestations suggestive of neurocysticercosis
     - Individuals coming from or living in an area where cysticercosis is endemic

[a] Diagnosis is confirmed by one absolute criterion, by two major criteria or one major and one confirmatory neuroimaging criteria plus any clinical/exposure criterion, or by one major neuroimaging criterion plus two clinical/exposure criteria (including at least one major clinical/exposure criterion), together with the exclusion of other pathologies producing similar neuroimaging findings. A probable diagnosis is supported by one major neuroimaging criterion plus any two clinical/exposure criteria or by one minor neuroimaging criterion plus at least one major clinical/exposure criterion.

Source: Reproduced with permission from OH Del Brutto et al: Revised diagnostic criteria for neurocysticercosis. J Neurol Sci 372:202, 2017.

antigens (e.g., ELISAs using crude parasite antigen) have high rates of false-positive and false-negative results and should be avoided. An immunoblot assay (enzyme-linked immunoelectrotransfer blot [EITB]) using lentil lectin–purified glycoproteins is >99% specific and sensitive in patients with multiple cysts. However, patients with single intracranial lesions or with calcifications may be seronegative. Serum samples are more sensitive than CSF using EITB. Each of the diagnostic antigens has been cloned, and assays using recombinant or synthetic antigens are in development. Assays using monoclonal antibodies to detect parasite antigen in the blood, CSF, or urine may also facilitate diagnosis and patient follow-up. Antigen-detection assays are currently available commercially in Europe but not in the United States. More recently, real-time PCR has been employed for diagnosis and follow-up of extraparenchymal disease.

Other major clinical/exposure criteria for neurocysticercosis include the presence of cysticerci outside the CNS (e.g., typical cigar-shaped calcifications in muscle) or exposure to a tapeworm carrier or a household member infected with *T. solium*. Minor clinical/exposure criteria include residence in an endemic village or clinical symptoms suggestive of neurocysticercosis (e.g., seizures or obstructive hydrocephalus).

Studies from India validated clinical criteria for diagnosis in selected cases. In patients from endemic areas who had single enhancing lesions presenting with seizures, a normal physical examination, and no evidence of systemic disease (e.g., no fever, adenopathy, or chest radiographic abnormalities), the presence on CT of round lesions 5–20 mm in diameter with no midline shift was almost always caused by neurocysticercosis. Definite or probable diagnosis can be made using the criteria and combinations of criteria listed in the footnote of Table 242-1. Patients often have CSF pleocytosis with a predominance of lymphocytes, neutrophils, or eosinophils. The protein level in CSF may be elevated; the glucose concentration is usually normal but can be markedly reduced.

TREATMENT

Taenia solium and Cysticercosis

Tapeworm infection by *T. solium* infection is treated with a single dose of praziquantel (10 mg/kg). However, praziquantel may occasionally trigger an inflammatory response in the CNS if concomitant cryptic cysticercosis is present. Niclosamide (2 g) is also effective but is not as widely available.

INITIAL MANAGEMENT OF NEUROCYSTICERCOSIS

Initial management of neurocysticercosis should focus on treatment of seizures or hydrocephalus. Seizures can be controlled with antiseizure medications. Seizure medications can usually be tapered after 6 months in patients with single enhancing lesions in whom imaging normalizes and in whom there are no breakthrough seizures. Subjects with multiple parenchymal lesions require more prolonged therapy. However, antiseizure medications can often be tapered off after 2 years if lesions resolve without development of calcifications and patients remain free of seizures. Patients with calcified lesions are at higher risk of recurrent seizures, especially if the lesions are associated with perilesional edema or enhancement.

MANAGING HYDROCEPHALUS

For patients with hydrocephalus, the reduction of intracranial pressure should be the priority of the initial therapy. Patients with cysticerci in the cerebral ventricles typically present with obstructive hydrocephalus, and the preferred approach is removal of the cysticercus via neurosurgery. Cysticerci in the lateral or third ventricles should be removed via neuroendoscopy. Antiparasitic drugs make the cysticerci more friable and should be avoided prior to surgery. The cysticerci in the fourth ventricle can be approached by microdissection using a posterior approach or, in some cases, via neuroendoscopy. When complete removal of the cysticercus is not possible, a diverting procedure, such as ventriculoperitoneal shunting, can be used to manage hydrocephalus. Historically, shunt failure was a major problem. The risk of shunt failure may be limited by administration of antiparasitic drugs and glucocorticoids.

ANTIPARASITIC DRUGS AND ANTI-INFLAMMATORY THERAPY

Antiparasitic drug treatment is never an emergency in neurocysticercosis and should wait until patients are stabilized with antiseizure and anti-inflammatory medications and exclusion of intraocular disease. Antiparasitic drugs should never be started in patients with elevated intracranial pressure. Antiparasitics do hasten resolution of neuroradiologic abnormalities in parenchymal neurocysticercosis. Clinical benefits consist mainly of decreasing the number of recurrent generalized seizures. In viable parenchymal cysticercosis, most authorities recommend antiparasitic drugs, especially albendazole (15 mg/kg per day for 8–28 days). A combination of albendazole and praziquantel (50 mg/kg per day) is more effective in patients with more than two cystic lesions. A longer course or combination therapy is needed in patients with multiple subarachnoid cysticerci. Both antiparasitic agents may exacerbate the inflammatory response around the dying parasite, thereby exacerbating seizures or hydrocephalus. Patients receiving these drugs should be carefully monitored. High-dose glucocorticoids should always be used during treatment (e.g., dexamethasone 0.1–0.4 mg/kg per day or prednisone 60 mg/d).

For patients with subarachnoid cysts or giant cysticerci, anti-inflammatory medications such as glucocorticoids are needed to reduce arachnoiditis and accompanying vasculitis. Most authorities recommend prolonged courses of antiparasitic drugs as well as shunting when hydrocephalus is present. Patients typically require prolonged anti-inflammatory treatment along with antiparasitics.

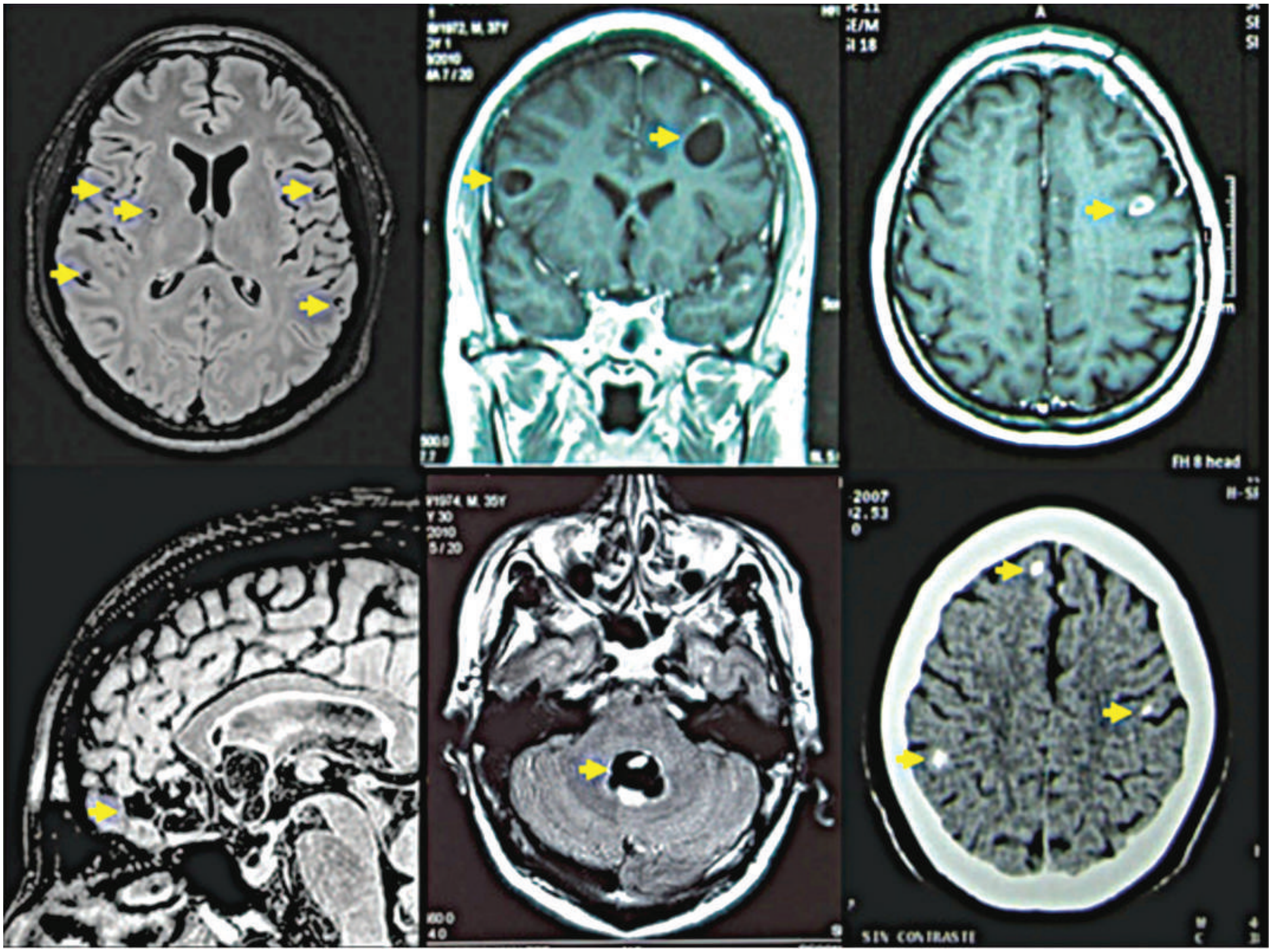

FIGURE 242-1 **Neurocysticercosis is caused by *Taenia solium*.** Neurologic infection can be classified based on the location and viability of the parasites. ***Upper left:*** Parenchymal viable cysts (FLAIR MRI sequence). ***Upper center:*** Parenchymal viable cysts (postcontrast T1 MRI sequence). ***Upper right:*** Single enhancing lesion (postcontrast T1 MRI sequence). ***Bottom left:*** Extensive basal subarachnoid neurocysticercosis in the anterior fossa (FLAIR MRI sequence). ***Bottom center:*** Viable cyst in the fourth ventricle (FLAIR MRI sequence). ***Bottom right:*** Intraparenchymal brain calcifications (noncontrasted CT scan). Lesions are marked with *arrowheads*. FLAIR, fluid-attenuated inversion recovery. *(Modified with permission from White AC Jr, Garcia HH. Updates on the management of neurocysticercosis. Curr Opin Infect Dis. 2018;31(5):377-382. Lippincott Williams & Wilkins.)*

Methotrexate and, in some cases, tumor necrosis factor inhibitors have been used as steroid-sparing agents in patients requiring prolonged therapy. In patients with diffuse cerebral edema and elevated intracranial pressure due to multiple inflamed parenchymal lesions, glucocorticoids are the mainstay of therapy, and antiparasitic drugs should be avoided. For ocular and spinal lesions, drug-induced inflammation may cause irreversible damage. Intraocular disease should be managed surgically. Recent data suggest that spinal disease is best managed using both medical and surgical therapy.

Prevention Prevention of *T. solium* tapeworm infection consists of precautions in handling pork, as described for *T. asiatica*, and thoroughly cooking or freezing pork to destroy the cysticerci. Pork inspections and condemnation of infected meat prevent transmission. The prevention of cysticercosis involves good personal hygiene including handwashing, effective disposal of feces, and treatment and prevention of human intestinal infections. Optimal eradication programs to eradicate *T. solium* in endemic areas include mass chemotherapy administered to human and porcine populations and vaccinations of pigs. A vaccine for porcine infection is licensed in India and a few other countries.

ECHINOCOCCOSIS

Echinococcosis (also known as hydatid disease) refers to infection by the larval stage of *Echinococcus* species (*E. granulosus* sensu lato, *E. multilocularis*, or *E. vogeli*). *E. granulosus* sensu lato parasites produce cystic hydatid disease or cystic echinococcosis, which is prevalent worldwide in most areas where livestock is raised in association with dogs. *E. granulosus* sensu lato is a complex of several distinct species with important genotypic and phenotypic differences. Human cystic echinococcosis is caused by *E. granulosus* sensu stricto (genotypes 1–3), *E. canadensis* (genotypes 6–8 and 10), and *E. ortleppi* (genotype 5). Some classify genotypes 6 and 7 as a separate species—*E. intermedius*. *E. granulosus* sensu lato parasites are found on all continents, with areas of high prevalence in western China, central Asia, the Middle East, the Mediterranean region, eastern Africa, and parts of South America. *E. multilocularis* causes multilocular or alveolar echinococcosis characterized by locally invasive lesions. Alveolar echinococcosis is prevalent in Alpine, sub-Arctic, and Arctic regions of the northern hemisphere, including western China, central Asia, central and northern Europe, and in isolated areas of North America with an expanding range of endemic areas. Neotropical echinococcosis (formerly termed polycystic hydatid disease) is caused by *E. vogeli* and *E. oligarthrus*, which are only found in limited foci in South America.

Echinococcal species require both intermediate and definitive hosts. The definitive hosts are usually canines (dogs, foxes, wolves) that harbor the small tapeworms in the intestine and shed eggs in stool. After the ingestion of eggs, oncospheres invade through the intestines into the circulatory system and form cysts in the intermediate hosts including sheep, cattle, goats, camels, pig, and horses for *E. granulosus* sensu lato and mice and other rodents for *E. multilocularis*. When a dog

(*E. granulosus*) or fox/wolf (*E. multilocularis*) ingests viscera containing cysts, the protoscolices in the cyst fluid develop into tapeworms in the intestine, completing the life cycle. Humans are a dead-end intermediate host and not part of the parasite's life cycle.

Etiology The adult tapeworms of *E. granulosus* sensu lato are small (5 mm long) and live for 5–20 months in the small intestine of dogs. Each tapeworm has only three proglottids: one immature, one mature, and one gravid. The latter are shed and release eggs that are morphologically similar to *Taenia* eggs. Heavy infections of dogs with many tapeworms are common in endemic areas. When humans ingest the eggs, the invasive oncospheres are released from eggs, penetrate the intestinal mucosa, enter the portal circulation, and are carried mostly to the liver and lungs. However, virtually any organ can be infected including kidneys, spleen, heart, bone, and brain. Larvae of *E. granulosus* sensu lato develop into fluid-filled unilocular hydatid cysts. The wall of the cystic lesion consists of an external membrane and an inner germinal layer, which are surrounded by the host's adventitial layer that may contain different patterns of inflammation and fibrosis. Daughter cysts and germinating cystic structures called brood capsules develop from the inner aspect of the germinal layer. New organisms, called protoscolices, develop in large numbers within the brood capsule. A protoscolex is an invaginated scolex with the capacity to form an adult tapeworm if ingested by a definitive host or form a new cystic lesion if released in the intermediate host's tissues. The cysts expand slowly over a period of years and may contain thousands of protoscolices.

E. multilocularis has a life cycle involving wild canines such as foxes or wolves, which are the main definitive hosts. Domestic dogs can also serve as definitive hosts of these tapeworms. Small rodents are the main intermediate hosts. Humans are dead-end intermediate hosts that develop alveolar echinococcosis. The larval stage of this parasite forms multilocular, small irregular cysts with proliferating and invasive capacity. The parasite larvae invade the host tissue by peripheral extension of processes from the germinal layer. These lesions do not contain brood capsules or protoscolices.

Clinical Manifestations The slowly enlarging echinococcal cysts generally remain asymptomatic for years until their expanding size elicits organ-specific symptoms. Spontaneous or traumatic rupture may lead to type I allergic reactions including anaphylaxis. The liver is involved in two-thirds of *E. granulosus* sensu lato infections and nearly all *E. multilocularis* infections. The lungs are involved in about 20% of *E. granulosus* sensu lato infections. The parasites are often discovered incidentally on a routine x-ray or ultrasound study prior to onset of symptoms.

Symptoms of hepatic cystic echinococcosis may include abdominal fullness or pain or a palpable mass in the right upper quadrant. Compression of a bile duct or leakage of cyst fluid into the biliary tree may present with symptoms mimicking cholelithiasis. Jaundice can result from biliary obstruction. Rupture of or leakage from a hydatid cyst may present more acutely with symptoms including fever, pruritus, urticaria, eosinophilia, or anaphylaxis. Cystic echinococcosis in the lungs may present with chronic cough, shortness of breath, chest pain, or hemoptysis. Rupture into the bronchial tree leads to sudden expectoration of the cyst fluid and membranes and rupture into the pleural cavity may produce pleuritic chest pain and hydatid empyema. Rupture of hydatid cysts, which can occur spontaneously, after trauma, or during surgery, may lead to release of protoscolices into the patient's tissues, each of which can form new cysts. Hydatid disease may also involve bone (invasion of the medullary cavity with bony erosion producing pathologic fractures), the CNS (space-occupying lesions), the heart (conduction defects, pericarditis), and the pelvis (pelvic mass).

E. multilocularis characteristically presents as a slow-growing hepatic mass, which typically presents decades after the initial infection. The lesions resemble tumors causing progressive destruction of the liver and extension into adjoining structures. Frequent symptoms include upper-quadrant and epigastric discomfort. Physical examination may reveal liver enlargement and obstructive jaundice. The lesions may infiltrate adjoining organs (e.g., diaphragm, kidneys, or lungs) or may metastasize to the spleen, lungs, or brain.

Diagnosis Imaging studies are the main diagnostic methods to detect and evaluate echinococcal cysts. Chest x-ray or CT can identify pulmonary cysts due to *E. granulosus* sensu lato, which appear as rounded masses of uniform density **(Fig. 242-2)**.

Ultrasound, CT, or MRI can be used to identify cystic echinococcosis lesions in solid organs of the abdomen, particularly in the liver. The ultrasound classification proposed by the World Health Organization Informal Working Group on Echinococcosis has diagnostic and management applications for liver cystic echinococcosis **(Figs.** 242-2 **and 242-3)**. MRI can also be used to classify lesions. MRI and CT may be more useful than ultrasound to evaluate the presence of cyst complications, such as communication with the biliary tree, that preclude some management options. Imaging of cystic echinococcosis shows well-defined cysts walls and, in some cases, internal trabeculation, dense cyst material, and/or calcifications (Fig. 242-3). In some cases, the protoscolices and brood capsules of *E. granulosus* complex may be visible within the cysts as fine particles termed hydatid sand. Identification of daughter cysts within the larger cyst is diagnostic of cystic echinococcosis disease. Eggshell or mural calcification on CT is also diagnostic of *E. granulosus* infection. In contrast, ultrasound or CT of alveolar hydatid cysts often reveals an indistinct solid mass. Some cases will display central necrosis or plaque-like calcifications.

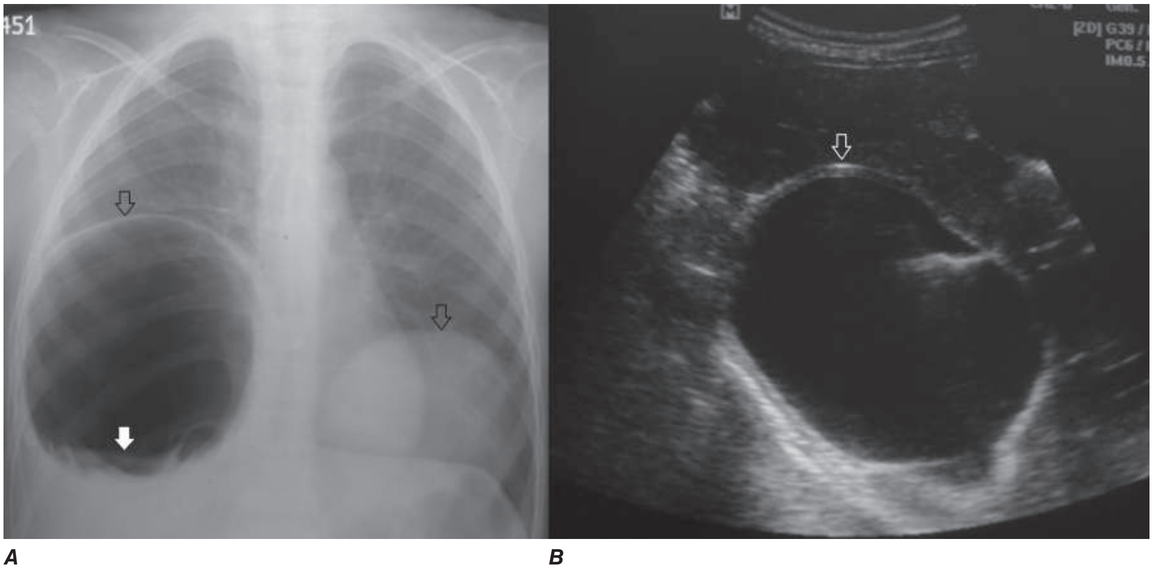

FIGURE 242-2. Imaging studies of cystic echinococcosis. ***A.*** Chest x-ray film of a patient with bilateral cysts. The hollow arrows show well-defined cyst walls with a complicated right chest cyst and an intact left chest cyst. ***B.*** Liver ultrasound of a patient with CE1 cysts. The hollow arrow shows a well-defined bilayer cyst wall.

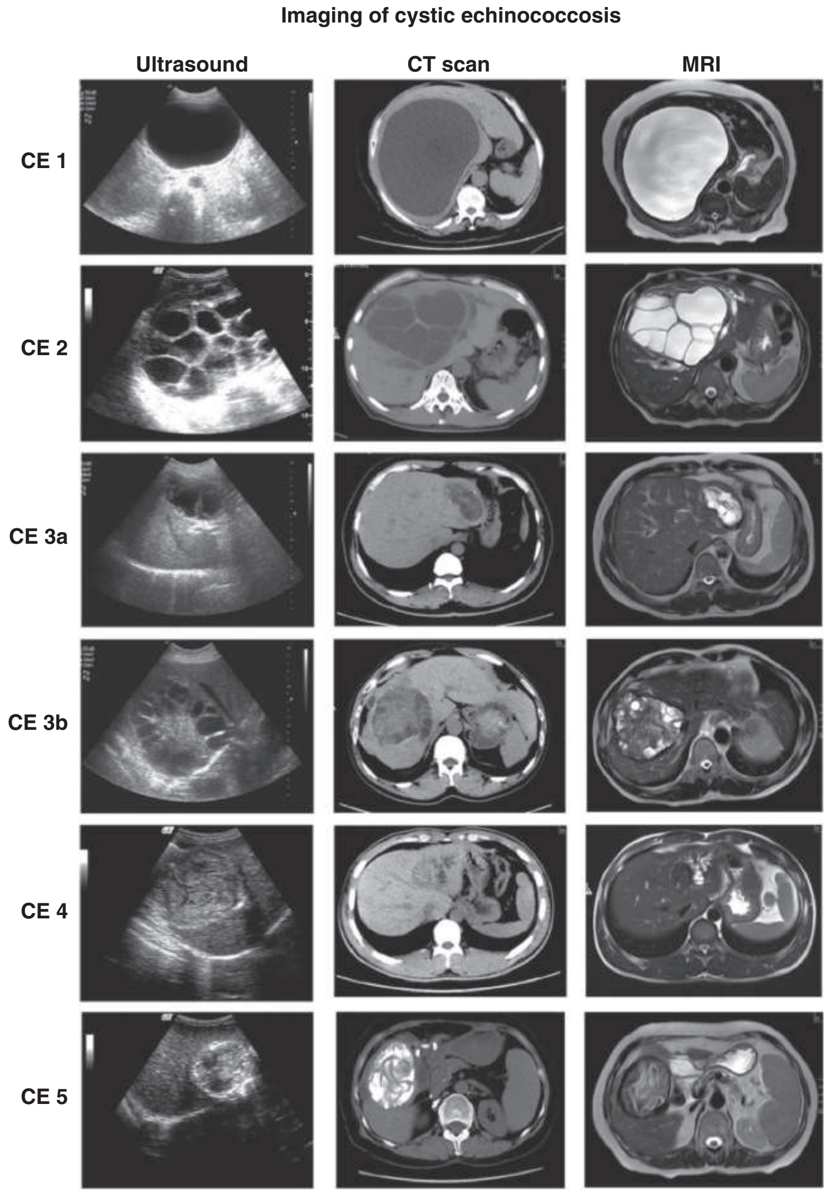

FIGURE 242-3 Management of cystic hydatid disease caused by *Echinococcus granulosus* should be based on viability of the parasite, which can be estimated from radiographic appearance. Staging is done by imaging studies including ultrasound, CT, or MRI and includes lesions classified as active, transitional, and inactive. *Active* cysts include types CL (with a cystic lesion and no visible cyst wall), CE1 (with a visible cyst wall and internal echoes [*snowflake sign*]), and CE2 (with a visible cyst wall and internal septation). *Transitional* cysts may have detached laminar membranes (CE3a) or may be partially collapsed (CE3b). *Inactive cysts* include types CE4 (a nonhomogeneous mass) and CE5 (a cyst with a thick calcified wall).

For cystic echinococcosis disease, a definitive diagnosis can also be made by the examination of aspirated fluids for protoscolices and/or hooklets. However, due to the potential risk of fluid leakage resulting in either dissemination of infection or, more rarely, anaphylactic reactions, aspiration should only be performed by experienced interventionists and following the precautions used for percutaneous treatment (see below). Serodiagnostic assays can be useful, but current serologic tests are insensitive and cannot be used to exclude the diagnosis of echinococcosis. ELISA and immunoblot assays for specific antibody are positive in ~90% of cases of hydatid liver disease. By contrast, the sensitivity is only ~50% for patients with cysts in the lungs.

TREATMENT

Echinococcosis

CYSTIC HYDATID DISEASE

Optimal treatment of cystic echinococcosis varies depending on the size, stage, location, and clinical manifestations of cysts. In the past, surgery was the main treatment method, but numerous studies have demonstrated that other treatment modalities may be just as effective and lead to less morbidity. Staging is recommended for cystic echinococcosis of the liver, which allows assessment of the cyst size and viability (Fig. 242-3). CL, CE1, and CE3a lesions <5 cm in diameter may respond well to chemotherapy with albendazole alone. Other small liver cysts (<5 cm) such as CE2 or CE3b are less responsive to medical treatment alone. Larger CE1 lesions and uncomplicated CE3a lesions in the liver can often be managed by PAIR (percutaneous aspiration, instillation of scolicidal agents, and reaspiration). Contraindications to PAIR include cysts communicating with the biliary tree, large cysts (>10 cm), or superficial cysts, which are more likely to rupture and spill protoscolices. Albendazole (15 mg/kg daily in two divided doses) should be initiated at least 2 days before the procedure and continued for at least 4 weeks afterward. Fine-needle ultrasound or CT-guided aspiration should enter the cyst through solid tissues to limit spillage of cyst fluid. Aspiration can confirm the diagnosis by microscopic demonstration of protoscolices or hooks. After aspiration, bilirubin should be measured in the cyst fluid using a dipstick, or contrast material should be injected to detect occult communications with the biliary tract. If no bile is found and no communication is visualized, instillation of scolicidal agents (usually hypertonic saline) and reaspiration are performed. PAIR, when performed by a skilled practitioner, results in cure and relapse rates equivalent to surgery, with less perioperative morbidity and shorter hospitalization. In some centers, CE2 lesions have been treated by modified catheter drainage, including puncture of each daughter cyst within the primary cyst.

Patients with treatment failure can often be treated again successfully with PAIR or additional courses of medical therapy. Response to treatment is best assessed by serial imaging studies, with attention to cyst size and consistency. Cysts may not demonstrate complete radiologic resolution even though no viable protoscolices are present. Those cysts classified as CE4 or CE5 are considered nonviable and require periodic reevaluations to assess for reactivation.

Surgery remains the treatment of choice for complicated cystic echinococcosis (e.g., cysts communicating with the biliary tract), for most thoracic and intracranial cysts, and when PAIR is not possible. Liver cysts should be removed via a pericystectomy, in which the entire cyst and the surrounding fibrous tissue are removed to prevent spillage and recurrence. Recent reports demonstrate that, in experienced hands, cysts can often be safely removed by laparoscopic or robotic surgery. The risks posed by leakage of fluid during surgery or PAIR include anaphylaxis and dissemination of protoscolices. Spillage can be minimized by careful dissection and using surgical draping soaked in hypertonic saline. Infusion of scolicidal agents is no longer recommended because of problems with hypernatremia, intoxication, or sclerosing cholangitis. Albendazole should be administered adjunctively, beginning several days to weeks before resection of the liver cyst and continuing for several weeks afterwards. Albendazole at 10-15 mg/kg divided in two daily doses (400 mg BID in adults) is recommended for treatment in association with clinical and laboratory monitoring.

ALVEOLAR HYDATID DISEASE

Surgical resection is required to attempt cure of *E. multilocularis* infection. Complete removal of the parasite continues to offer the best chance for cure. Patients who have undergone presumed curative resection should be treated with albendazole for at least 2 years after presumptively curative surgery. Positron emission tomography can be used to follow disease activity. Unfortunately, most cases are only diagnosed at a stage in which complete resection is impossible; in these cases, albendazole treatment should be continued indefinitely, with careful monitoring. In some cases of larger lesions, complete removal of the liver followed by liver transplantation has been used. Unfortunately, the immunosuppression required to prevent rejection of the transplanted liver also promotes proliferation of *E. multilocularis* and reinfection of the transplanted liver. Thus, indefinite treatment with albendazole is required.

Prevention Cystic echinococcosis can be prevented by administering praziquantel to infected dogs, by preventing dogs from having access to viscera from infected animals, or by vaccinating sheep. Limiting the number of stray dogs that are more likely to evade prevention measures may help decrease transmission. In Europe, *E. multilocularis* infection is associated with human settlement encroaching into forested areas, gardening in those areas, and collecting activities in forested areas. Thus, gloves and hand hygiene should be used in these situations. Praziquantel-impregnated bait can be used to treat tapeworms in wild canines.

■ *RODENTOLEPIS NANA* (PREVIOUSLY *HYMENOLEPIS NANA*)

The dwarf tapeworm, previously known as *H. nana*, is the most common human cestode infection. Recent molecular data have led to the reclassification of this organism to a different genus and new name *Rodentolepis nana*. *R. nana* is endemic worldwide, including temperate and tropical regions. Transmission is mostly person-to-person by the fecal-oral route.

Etiology and Pathogenesis *R. nana* is the only human cestode that does not require an intermediate host. Both the larval and adult stages coexist in the intestine of infected persons. The adult tapeworm is only ~2 cm long and lives in the human proximal ileum. The tiny proglottids are rarely seen in the stool. They release spherical eggs, 30–44 μm in diameter, containing the invasive larvae termed oncosphere, which has six hooklets. The eggs are immediately infective after leaving the host and survive for ≤10 days in the environment. After the eggs are ingested, the oncosphere is released, penetrates the intestinal villi, and develops into the cysticercoid larval form within the epithelium. After a few days, these larvae reenter the intestinal lumen, attach to the mucosa, and mature into adult tapeworms over 10–12 days. The life span of adult *R. nana* worms is typically ~4–10 weeks. However, infection is perpetuated by cycles of reinfection and autoinfection in which some eggs hatch in the intestinal lumen and form the cysticercoid larva without leaving the host.

Clinical Manifestations *R. nana* infection is most often asymptomatic. However, infection may be associated with diarrhea, abdominal pain, and weight loss, particularly in children with the highest burden of infection.

Diagnosis Infection is diagnosed by finding the characteristic *R. nana* eggs in microscopy of the stool.

TREATMENT

Rodentolepis nana Infection

The treatment of choice for *R. nana* is praziquantel (25 mg/kg once), which is active against both the adult worms and the cysticercoids in the intestinal wall. Nitazoxanide (500 mg twice a day for 3 days) has been used as an alternative treatment.

Prevention Since *R. nana* is acquired by the fecal-oral route, improved sanitation and personal hygiene can be used to eliminate disease. Hand washing in the household and school is important. Mass chemotherapy and improved hygiene have been used to control epidemics.

■ *HYMENOLEPIASIS DIMINUTA*

Hymenolepis diminuta is a cestode of rodents that occasionally causes infection in small children. Infection is acquired by ingesting uncooked cereal and other foods contaminated by fleas and other insects that serve as intermediate hosts for *H. diminuta*. Infection is diagnosed by detection of eggs in the stool. The treatment of choice is praziquantel (25 mg/kg once).

■ DIPHYLLOBOTHRIASIS (*DIBOTHRIOCEPHALUS/ ADENOCEPHALUS*)

The broad fish tapeworms comprised several species that were formerly classified under the genus *Diphyllobothrium*. Molecular and phylogenetic studies have now demonstrated several important differences between these parasites, which led to modification of the taxonomic classification of cestodes of the *Diphyllobothriidae* family infecting humans. *Dibothriocephalus latus* (formerly *Diphyllobothrium latum*), *Adenocephalus pacificus* (formerly *Diphyllobothrium pacificum*), and *Dibothriocephalus nihonkaiensis* (formerly *Diphyllobothrium nihonkaiensis*) are the most common species infecting humans. These parasites were initially identified in freshwater lakes, rivers, and deltas of the Northern Hemisphere and central Africa. However, they are also found in marine environments in the northern Pacific Ocean (*D. nihonkaisensis*) and the Pacific coast of South America (*A. pacificus*).

Etiology and Pathogenesis The adult tapeworm can reach a length of up to 25 m, making them the longest tapeworms of humans. The scolex attaches to the small intestinal mucosa by a modified sucker called *bothria*, which is located on the elongated scolex. The mature tapeworms have 3000–4000 proglottids, producing ~1 million eggs per day. *D. latus* eggs hatch in fresh water and release the free-swimming coracidium. The coracidia are ingested by small freshwater crustaceans

(*Cyclops* or *Diaptomus* species). Within the infected crustaceans, the procercoid larvae develop. When the infected crustaceans are ingested by fish, the procercoid larva migrate to the fish's flesh and transform into a sparganum or plerocercoid larva. Humans acquire the infection by ingesting infected raw or smoked fish. Within 3–5 weeks, the tapeworm matures into an adult in the human intestine. For *A. pacificus*, the definite hosts include seals, dogs, and humans, and the second intermediate hosts are marine fish.

Clinical Manifestations Most infections by these tapeworms are asymptomatic. Some patients note abdominal discomfort, diarrhea, vomiting, weakness, or weight loss. The proglottids may be passed in stool or found incidentally during endoscopy. Diphyllobothriidae can cause acute abdominal pain or intestinal obstruction in rare cases. *D. latus* tapeworms have avid receptors for vitamin B_{12} that can interfere with absorption and, in patients with other risk factors, produce vitamin B_{12} deficiency. In patients with B_{12} deficiency, pernicious anemia and neurologic sequelae may develop.

Diagnosis The diagnosis is made by the detection of the characteristic eggs in the stool. The eggs possess a single shell with an operculum at one end and a knob at the other, which may be confused with trematode eggs by the inexperienced technician. Mild to moderate eosinophilia may be detected. Examination of the tapeworms passed also provides a diagnosis, as proglottids have a characteristic uterus with a rosette-like shape.

TREATMENT

Diphyllobothriasis

Praziquantel (5–10 mg/kg once) is highly effective against all Diphyllobothriid species. Parenteral vitamin B_{12} may be given for B_{12} deficiency.

Prevention Heating fish to 54°C for 5 min or freezing at –18°C for 24 h kills the larval forms. Placing fish in brine with a high salt concentration for long periods can also kill the plerocercoid larvae.

DIPYLIDIASIS

Dipylidium caninum is a common tapeworm of dogs and cats. Dogs, cats, and occasionally humans become infected by swallowing fleas harboring the intermediate forms (cysticercoid larvae). Children are more often infected than adults. Infections are usually asymptomatic except for the passage of motile proglottids in stool and, less often, vague abdominal symptoms. The diagnosis is made by the detection of proglottids or the characteristic egg packets in the stool. Eggs within the packet resemble other *Taenia* eggs. The motile proglottids resemble flattened grains of rice and, in small children, may migrate out of the anus. Praziquantel is the treatment of choice. Prevention should mainly focus on antiparasitic and flea treatment of dogs and cats.

SPARGANOSIS

The plerocercoid larvae of Diphyllobothriid tapeworms of the genus *Spirometra* cause human sparganosis. Dogs, cats, and pigs are hosts of the adult tapeworm; crustaceans of the *Cyclops* species in fresh water are the first intermediate host; and snakes, frogs, or birds are the second intermediate hosts. Human infection commonly follows topical application of poultices with infected flesh from snakes, frogs, or birds used in traditional medicine. Infection can also be acquired by the ingestion of water containing infected *Cyclops* and raw or undercooked meat from infected snakes, birds, or some mammals. Infection commonly presents as a subcutaneous swelling that contains the parasite. Periorbital infections can present with swelling and intraocular infections may lead to blindness. Infections of the brain can present as a mass or slowly migrating lesions. Proliferative lesions may cause tissue infiltration and are poorly responsive to medical treatment. Surgical excision is used to treat localized sparganosis.

COENUROSIS

Coenurosis is a rare infection of humans by the larval stage (coenurus) of the dog tapeworms *Taenia multiceps* or *Taenia serialis*. The main clinical manifestations are space-occupying cystic lesions in various tissues. The commonly involved tissues include the CNS or subcutaneous tissue. Surgical excision is usually required for both definitive diagnosis and treatment. There are limited data on response to antiparasitic treatment.

Acknowledgment

The authors acknowledge and thank Peter F. Weller, MD, author of prior editions of this chapter.

FURTHER READING

Brunetti E et al: Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans. Acta Trop 114:1, 2010.

Del Brutto OH et al: Revised diagnostic criteria for neurocysticercosis. J Neurol Sci 372:202, 2017.

Kern P et al: The echinococcoses: Diagnosis, clinical management and burden of disease. Adv Parasitol 96:259, 2017.

Nash TE et al: Natural history of treated subarachnoid neurocysticercosis. Am J Trop Med Hyg 102:78, 2020.

Nguyen DC et al: The brief case: The boy who cried worm. J Clin Microbiol 61:e00553, 2013.

Wen H et al: Echinococcosis: Advances in the 21st century. Clin Microbiol Rev 32:e00075, 2019.

White AC Jr et al: Diagnosis and treatment of neurocysticercosis: 2017 clinical practice guidelines by the Infectious Diseases Society of America (IDSA) and the American Society of Tropical Medicine and Hygiene (ASTMH). Clin Infect Dis 66:1159, 2018.

This page intentionally left blank

Disorders of the Cardiovascular System PART 6

Section 1 Introduction to Cardiovascular Disorders

243

Approach to the Patient with Possible Cardiovascular Disease

Joseph Loscalzo

THE MAGNITUDE OF THE PROBLEM

Cardiovascular diseases comprise the most prevalent serious disorders in industrialized nations and are a rapidly increasing problem in developing nations **(Chap. 245)**. Age-adjusted death rates for coronary heart disease have declined by two-thirds in the past four decades in the United States, reflecting the identification and reduction of risk factors as well as improved treatments and interventions for the management of coronary artery disease, arrhythmias, and heart failure. Nonetheless, cardiovascular diseases remain the most common causes of death, responsible for nearly 700,000 deaths each year. Approximately one-third of these deaths are sudden. In addition, cardiovascular diseases are highly prevalent, diagnosed in nearly half of the adult population. The growing prevalence of obesity **(Chap. 414)**, type 2 diabetes mellitus **(Chap. 415)**, and metabolic syndrome **(Chap. 420)**, which are important risk factors for atherosclerosis, now threatens to reverse the progress that has been made in the age-adjusted reduction in the mortality rate of coronary heart disease. Globally, the impact of cardiovascular diseases is steadily mounting, with an estimated 19 million deaths worldwide.

For many years, cardiovascular disease was considered to be more common in men than in women. In fact, cardiovascular disease is the leading cause of all deaths among women and men **(Chap. 410)**. In addition, although the absolute number of deaths secondary to cardiovascular disease has declined over the past decades in men, this number has actually risen in women. Inflammation, obesity, type 2 diabetes mellitus, and the metabolic syndrome appear to play more prominent roles in the development of coronary atherosclerosis in women than in men. Coronary artery disease (CAD) is more frequently associated with dysfunction of the coronary microcirculation in women than in men. Exercise electrocardiography has a lower diagnostic accuracy in the prediction of epicardial obstruction in women than in men.

NATURAL HISTORY

Cardiovascular disorders often present acutely, as in a previously asymptomatic person who develops an acute myocardial infarction **(Chap. 286)**, or a previously asymptomatic patient with hypertrophic cardiomyopathy **(Chaps. 266–270)** or with a prolonged QT interval **(Chap. 259)** whose first clinical manifestation is syncope or even sudden death. However, the alert physician may recognize the patient at risk for these complications long before they occur and often can take measures to prevent them. For example, a patient with acute myocardial infarction will often have had risk factors for atherosclerosis for many years. Had these risk factors been recognized, their elimination or reduction might have delayed or even prevented the infarction. Similarly, a patient with hypertrophic cardiomyopathy may have had a heart murmur for years and a family history of this disorder. These findings could have led to an echocardiographic examination, recognition of the condition, and appropriate therapy long before the occurrence of a serious acute manifestation.

Patients with valvular heart disease or idiopathic dilated cardiomyopathy, by contrast, may have a prolonged course of gradually increasing dyspnea and other manifestations of chronic heart failure that is punctuated by episodes of acute deterioration only late in the course of the disease. Understanding the natural history of various cardiac disorders is essential for applying appropriate diagnostic and therapeutic measures to each stage of the condition, as well as for providing the patient and family with the likely prognosis.

CARDIAC SYMPTOMS

The symptoms caused by heart disease result most commonly from myocardial ischemia, disturbance of the contraction and/or relaxation of the myocardium, obstruction to blood flow, or an abnormal cardiac rhythm or rate. Ischemia, which is caused by an imbalance between the heart's oxygen supply and demand, is manifest most frequently as chest discomfort **(Chap. 15)**, whereas reduction of the pumping ability of the heart commonly leads to fatigue and elevated intravascular pressure upstream of the failing ventricle. The latter results in abnormal fluid accumulation, with peripheral edema **(Chap. 43)** or pulmonary congestion and dyspnea **(Chap. 39)**. Obstruction to blood flow, as occurs in valvular stenosis, can cause symptoms resembling those of myocardial failure **(Chap. 264)**. Cardiac arrhythmias often develop suddenly, and the resulting symptoms and signs—palpitations **(Chap. 45)**, dyspnea, hypotension, and syncope **(Chap. 23)**—generally occur abruptly and may disappear as rapidly as they develop.

Although dyspnea, chest discomfort, edema, and syncope are cardinal manifestations of cardiac disease, they occur in other conditions, as well. Thus, dyspnea is observed in disorders as diverse as pulmonary disease, marked obesity, and anxiety **(Chap. 39)**. Similarly, chest discomfort may result from a variety of noncardiac and cardiac causes other than myocardial ischemia **(Chap. 15)**. Edema, an important finding in untreated or inadequately treated heart failure, also may occur with primary renal disease and in hepatic cirrhosis **(Chap. 43)**. Syncope occurs not only with serious cardiac arrhythmias but in a number of neurologic conditions as well **(Chap. 23)**. Whether heart disease is responsible for these symptoms frequently can be determined by carrying out a careful clinical examination **(Chap. 246)**, supplemented by noninvasive testing using electrocardiography at rest and during exercise **(Chap. 247)**, echocardiography, and cardiopulmonary imaging **(Chap. 248)**.

Myocardial or coronary function that may be adequate at rest may be insufficient during exertion. Thus, dyspnea and/or chest discomfort that appear during activity are characteristic of patients with heart disease, whereas the opposite pattern, that is, the appearance of these symptoms at rest and their remission during exertion, is rarely observed in such patients. It is important, therefore, to question the patient carefully about the relation of symptoms to exertion.

Many patients with cardiovascular disease may be asymptomatic both at rest and during exertion but may present with an abnormal physical finding such as a heart murmur, elevated arterial pressure, or an abnormality of the electrocardiogram (ECG) or imaging test. It is important to assess the global risk of CAD in asymptomatic individuals, using a combination of clinical assessment and measurement of cholesterol and its fractions, as well as other biomarkers, such as C-reactive protein, in some patients. Since the first clinical manifestation of CAD may be catastrophic—sudden cardiac death, acute myocardial infarction, or stroke in previous asymptomatic persons—it is mandatory to identify those at high risk for such events and institute further testing and preventive measures.

DIAGNOSIS

As outlined by the New York Heart Association (NYHA), the elements of a complete cardiac diagnosis include the systematic consideration of the following:

1. *The underlying etiology*. Is the disease congenital, hypertensive, ischemic, or inflammatory in origin?
2. *The anatomic abnormalities*. Which chambers are involved? Are they hypertrophied, dilated, or both? Which valves are affected? Are they regurgitant and/or stenotic? Is there pericardial involvement? Has there been a myocardial infarction?

TABLE 243-1 New York Heart Association Functional Classification

Class I
No limitation of physical activity
No symptoms with ordinary exertion
Class II
Slight limitation of physical activity
Ordinary activity causes symptoms
Class III
Marked limitation of physical activity
Less than ordinary activity causes symptoms
Asymptomatic at rest
Class IV
Inability to carry out any physical activity without discomfort
Symptoms at rest

3. *The physiologic disturbances*. Is an arrhythmia present? Is there evidence of congestive heart failure or myocardial ischemia?
4. *Functional disability*. How strenuous is the physical activity required to elicit symptoms? The classification provided by the NYHA has been found to be useful in describing functional disability **(Table 243-1)**.

One example may serve to illustrate the importance of establishing a complete diagnosis. In a patient who presents with exertional chest discomfort, the identification of myocardial ischemia as the etiology is of great clinical importance. However, the simple recognition of ischemia is insufficient to formulate a therapeutic strategy or prognosis until the underlying anatomic abnormalities responsible for the myocardial ischemia, for example, coronary atherosclerosis or aortic stenosis, are identified and a judgment is made about whether other physiologic disturbances that cause an imbalance between myocardial oxygen supply and demand, such as severe anemia, thyrotoxicosis, or supraventricular tachycardia, play contributory roles. Finally, the severity of the disability should govern the extent and tempo of the workup and strongly influence the therapeutic strategy that is selected.

The establishment of a correct and complete cardiac diagnosis usually commences with the history and physical examination **(Chap. 246)**. Indeed, the clinical examination remains the basis for the diagnosis of a wide variety of disorders. The clinical examination may then be supplemented by five types of laboratory tests: (1) ECG **(Chap. 247)**; (2) noninvasive imaging examinations (chest x-ray, echocardiogram, radionuclide imaging, computed tomographic imaging, positron emission tomography, and magnetic resonance imaging) **(Chap. 248)**; (3) blood tests to assess risk (e.g., lipid determinations, C-reactive protein) or cardiac function (e.g., brain natriuretic peptide [BNP] **[Chap. 264]**); (4) occasionally, specialized invasive examinations (i.e., cardiac catheterization and coronary arteriography **[Chap. 249]**); and (5) genetic tests to identify monogenic cardiac diseases (e.g., hypertrophic cardiomyopathy **[Chaps. 266–270]**, Marfan's syndrome **[Chap. 425]**, and abnormalities of cardiac ion channels that lead to prolongation of the QT interval and an increase in the risk of sudden death **[Chap. 253]**). These genetic tests are becoming more widely available.

FAMILY HISTORY

In eliciting the history of a patient with known or suspected cardiovascular disease, particular attention should be directed to the family history. Familial clustering is common in many forms of heart disease. Mendelian transmission of single-gene defects may occur, as in hypertrophic cardiomyopathy **(Chaps. 266–270)**, Marfan's syndrome **(Chap. 425)**, and sudden death associated with a prolonged QT syndrome **(Chap. 259)**. Premature coronary disease and essential hypertension, type 2 diabetes mellitus, and hyperlipidemia (the most important risk factors for CAD) are usually polygenic disorders. Although familial transmission may be less obvious than in the monogenic disorders, it is helpful in assessing risk and prognosis in polygenic disorders, as well. Familial clustering of cardiovascular diseases not only may occur on a genetic basis but also may be related to familial dietary or behavior patterns, such as excessive ingestion of salt or calories and cigarette smoking.

ASSESSMENT OF FUNCTIONAL IMPAIRMENT

When an attempt is made to determine the severity of functional impairment in a patient with heart disease, it is helpful to ascertain the level of activity and the rate at which it is performed before symptoms develop. Thus, it is not sufficient to state that the patient complains of dyspnea. The breathlessness that occurs after running up two long flights of stairs denotes far less functional impairment than do similar symptoms that occur after taking a few steps on level ground. In addition, the degree of customary physical activity at work and during recreation should be considered. The development of two-flight dyspnea in a well-conditioned marathon runner may be far more significant than the development of one-flight dyspnea in a previously sedentary person. The history should include a detailed consideration of the patient's therapeutic regimen. For example, the persistence or development of edema, breathlessness, and other manifestations of heart failure in a patient who is receiving optimal doses of diuretics and other therapies for heart failure **(Chap. 264)** is far graver than are similar manifestations in the absence of treatment. In an effort to determine the progression of symptoms, and thus the severity of the underlying illness, it may be useful to ascertain what, if any, specific tasks the patient could have carried out 6 months or 1 year earlier that they cannot perform at present.

ELECTROCARDIOGRAM

(See also Chap. 247) Although an ECG usually should be recorded in patients with known or suspected heart disease, with the exception of the identification of arrhythmias, conduction abnormalities, ventricular hypertrophy, and acute myocardial infarction, it generally does not establish a specific diagnosis. The range of normal electrocardiographic findings is wide, and the tracing can be affected significantly by many noncardiac factors, such as age, body habitus, and serum electrolyte concentrations. In general, electrocardiographic changes should be interpreted in the context of other abnormal cardiovascular findings.

ASSESSMENT OF THE PATIENT WITH A HEART MURMUR

(Fig. 243-1) The cause of a heart murmur can often be readily elucidated from a systematic evaluation of its major attributes: timing, duration, intensity, quality, frequency, configuration, location, and radiation when considered in the light of the history, general physical examination, and other features of the cardiac examination, as described in **Chap. 246**.

The majority of heart murmurs are midsystolic and soft (grades I–II/VI). When such a murmur occurs in an asymptomatic child or young adult *without* other evidence of heart disease on clinical examination, it is usually benign and echocardiography generally is not required.

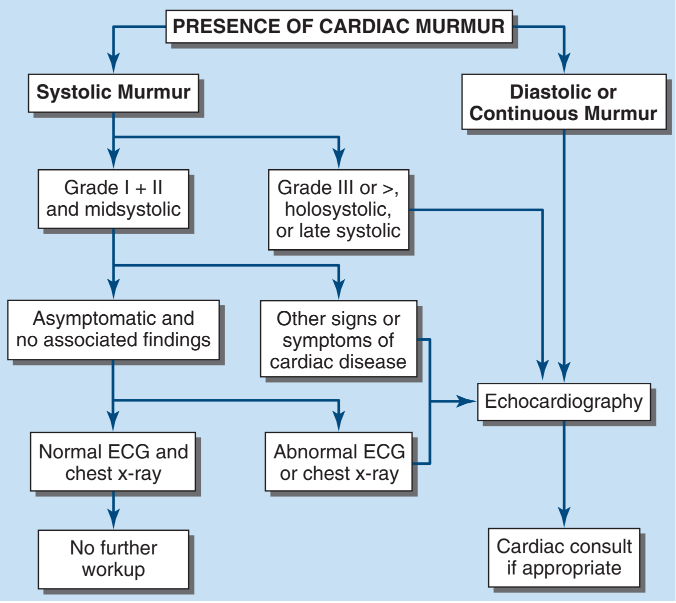
PRESENCE OF CARDIAC MURMUR
Systolic Murmur
Diastolic or Continuous Murmur
Grade I + II and midsystolic
Grade III or >, holosystolic, or late systolic
Asymptomatic and no associated findings
Other signs or symptoms of cardiac disease
Echocardiography
Normal ECG and chest x-ray
Abnormal ECG or chest x-ray
No further workup
Cardiac consult if appropriate

FIGURE 243-1 Approach to the evaluation of a heart murmur. ECG, electrocardiogram. *(Reproduced with permission from E Braunwald, L Goldman [eds]: Primary Cardiology, 2nd ed. Philadelphia, Saunders, 2003.)*

By contrast, two-dimensional and Doppler echocardiography **(Chap. 248)** are indicated in patients with loud systolic murmurs (grades ≥III/VI), especially those that are holosystolic or late systolic, and in most patients with diastolic or continuous murmurs.

PITFALLS IN CARDIOVASCULAR MEDICINE

Increasing subspecialization in internal medicine and the perfection of advanced diagnostic techniques in cardiology can lead to several undesirable consequences. Examples include the following:

1. Failure by the *noncardiologist* to recognize important cardiac manifestations of systemic illnesses. For example, the presence of mitral stenosis, patent foramen ovale, and/or transient atrial arrhythmia should be considered in a patient with stroke, or the presence of pulmonary hypertension and cor pulmonale should be considered in a patient with scleroderma or Raynaud's syndrome. A cardiovascular examination should be carried out to identify and estimate the severity of the cardiovascular involvement that accompanies many noncardiac disorders.
2. Failure by the *cardiologist* to recognize underlying systemic disorders in patients with heart disease. For example, hyperthyroidism should be considered in an elderly patient with atrial fibrillation and unexplained heart failure, and Lyme disease should be considered in a patient with unexplained (fluctuating) atrioventricular block. A cardiovascular abnormality may provide the clue critical to the recognition of some systemic disorders. For example, an unexplained pericardial effusion may provide an early clue to the diagnosis of tuberculosis or a neoplasm.
3. Overreliance on and overutilization of laboratory tests, particularly invasive techniques, for the evaluation of the cardiovascular system. Cardiac catheterization and coronary arteriography **(Chap. 249)** provide precise diagnostic information that may be crucial in developing a therapeutic plan in patients with known or suspected CAD. Although a great deal of attention has been directed to these examinations, it is important to recognize that they serve to *supplement*, not *supplant*, a careful examination carried out with clinical and noninvasive techniques. A coronary arteriogram should not be performed in lieu of a careful history in patients with chest pain suspected of having ischemic heart disease. Although coronary arteriography may establish whether the coronary arteries are obstructed and to what extent, the results of the procedure by themselves often do not provide a definitive answer to the question of whether a patient's symptom of chest discomfort is attributable to coronary atherosclerosis and whether or not revascularization is indicated.

Despite the value of invasive tests in certain circumstances, they entail some small risk to the patient, involve discomfort and substantial cost, and place a strain on medical facilities. Therefore, they should be carried out only if the results can be expected to modify the patient's management.

DISEASE PREVENTION AND MANAGEMENT

The prevention of heart disease, especially of CAD, is one of the most important tasks of primary health care givers as well as cardiologists. Prevention begins with risk assessment, followed by attention to lifestyle, such as achieving optimal weight, physical activity, and smoking cessation, and then aggressive treatment of all abnormal risk factors, such as hypertension, hyperlipidemia, and diabetes mellitus **(Chap. 415)**.

After a complete diagnosis has been established in patients with known heart disease, a number of management options are usually available. Several examples may be used to demonstrate some of the principles of cardiovascular therapeutics:

1. In the absence of evidence of heart disease, the patient should be clearly informed of this assessment and *not* be asked to return at intervals for repeated examinations. If there is no evidence of disease, such continued attention may lead to the patient's developing inappropriate concern about the possibility of heart disease.
2. If there is no evidence of cardiovascular disease but the patient has one or more risk factors for the development of ischemic heart disease **(Chap. 284)**, a plan for their reduction should be developed and the patient should be retested at intervals to assess compliance and efficacy in risk reduction.
3. Asymptomatic or mildly symptomatic patients with valvular heart disease that is anatomically severe should be evaluated periodically, every 6–12 months, by clinical and noninvasive examinations. Early signs of deterioration of ventricular function may signify the need for surgical treatment before the development of disabling symptoms, irreversible myocardial damage, and excessive surgical risk **(Chap. 272)**.
4. In patients with CAD **(Chap. 284)**, available practice guidelines should be considered in the decision on the form of treatment (medical, percutaneous coronary intervention, or surgical revascularization). Mechanical revascularization may be employed too frequently in the United States and too infrequently in Eastern Europe and developing nations. The mere presence of angina pectoris and/or the demonstration of critical coronary arterial narrowing at angiography should not reflexively evoke a decision to treat the patient by revascularization. Instead, these interventions should be limited to patients with CAD in whom revascularization has been shown to improve the natural history (e.g., acute coronary syndrome or multivessel CAD with left ventricular dysfunction).

FURTHER READING

Tsao CW et al: Heart disease and stroke statistics—2023 update: A report from the American Heart Association. Circulation 147:e93, 2023.

244

Basic Biology of the Cardiovascular System

Joseph Loscalzo, John F. Keaney, Jr., Calum A. MacRae

DEVELOPMENTAL BIOLOGY OF THE CARDIOVASCULAR SYSTEM

The heart forms early during embryogenesis **(Fig. 244-1)**, circulating blood, nutrients, molecular signals, and oxygen to the other developing organs while continuing to grow and undergo complex morphogenetic changes. Early cardiac progenitors arise within crescent-shaped fields of lateral splanchnic mesoderm under the influence of multiple cues and migrate to the midline to form the linear heart tube: a single layer of endocardium and a single layer of spontaneously beating cardiomyocytes.

The simple linear heart tube undergoes chamber specification and asymmetric looping, coordinated with longitudinal and concentric growth of different regions of the heart tube, to produce the presumptive atria and ventricles. Cells continue to migrate into the heart at both ends from additional heart fields in pharyngeal mesoderm as looping and growth occur. These cells exhibit distinctive gene expression (e.g., Islet-1) and distinctive physiology (e.g., calcium handling), contributing to discrete areas of the adult heart, including the right atrium and the right ventricle. These different embryonic origins of cells within the right and left ventricles correlate with distinctive single-cell RNA sequencing profiles decades later and help explain why some forms of congenital and adult cardiac diseases affect different regions of the heart.

After looping and chamber formation, a series of morphogenetic events divide the left side from the right side of the heart, separate the

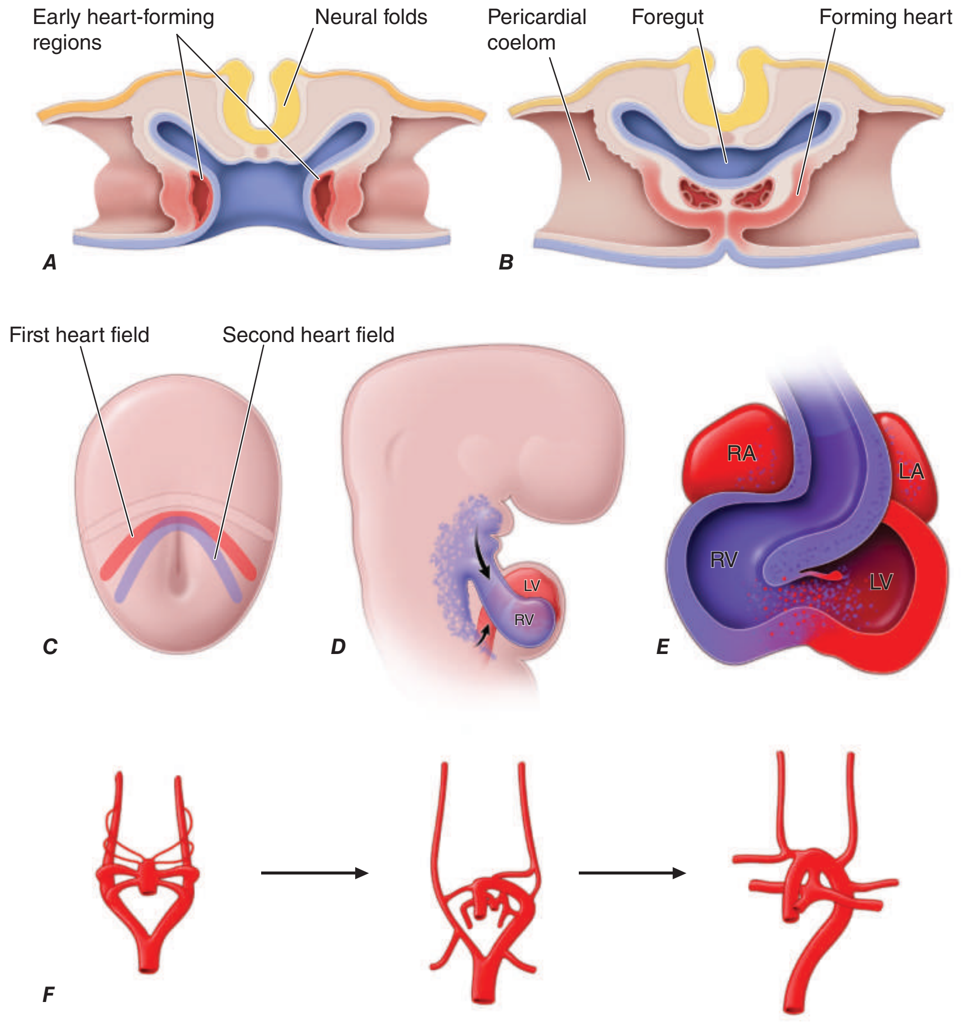

FIGURE 244-1 ***A.*** Schematic depiction of a transverse section through an early embryo depicts the bilateral regions where early heart tubes form. ***B.*** The bilateral heart tubes subsequently migrate to the midline and fuse to form the linear heart tube. ***C.*** At the early cardiac crescent stage of embryonic development, cardiac precursors include a primary heart field fated to form the linear heart tube and a second heart field fated to add myocardium to the inflow and outflow poles of the heart. ***D.*** Second heart field cells populate the pharyngeal region before subsequently migrating to the maturing heart. ***E.*** Large portions of the right ventricle and outflow tract and some cells within the atria derive from the second heart field. ***F.*** The aortic arch arteries form as symmetric sets of vessels that then remodel under the influence of the neural crest to form the asymmetric mature vasculature. LA, left atrium; LV, left ventricle; RA, right atrium; RV, right ventricle.

atria from the ventricles, and fashion the aorta and pulmonary artery from the truncus arteriosus. Cardiac valves form between the atria and the ventricles and between the ventricles and the outflow vessels. Early in development, myocardial cells secrete an extracellular matrix rich in hyaluronic acid, or "cardiac jelly," which accumulates within the endocardial cushions, precursors of the valves. Signals from overlying myocardial cells trigger migration, invasion, and phenotypic changes in underlying endocardial cells, an epithelial-mesenchymal transformation, that then invade and populate the endocardial cushion matrix. Mesenchymal cells then proliferate and form the mature valve leaflets.

The great vessels form as a series of symmetric bilateral aortic arch arteries that remodel asymmetrically to define the mature central vasculature. Migrating neural crest cells from the dorsal neural tube orchestrate this process and are necessary for aortic arch remodeling and septation of the truncus arteriosus. Smooth-muscle cells within the tunica media of the aortic arch, the ductus arteriosus, and the carotid arteries all derive from neural crest. By contrast, smooth muscle within the descending aorta arises from lateral plate mesoderm, and smooth muscle of the proximal outflow tract arises from the second heart field. Neural crest cells are sensitive to both vitamin A and folic acid, and congenital heart diseases involving abnormal remodeling of the aortic arch arteries are observed with maternal deficiencies of these vitamins. Shared embryonic origins of different cardiovascular cell types lead to syndromic associations between various congenital heart diseases and a range of extracardiac abnormalities.

Coronary artery formation requires the addition of yet another cell population to the embryonic heart. Epicardial cells arise in the proepicardial organ, a derivative of the septum transversum, which also contributes to the fibrous portion of the diaphragm and to the liver. Proepicardial cells contribute smooth muscle to the coronary arteries and are required for proper coronary patterning. Other cell types within the heart (e.g., fibroblasts) also can arise from the proepicardium.

The cardiac conduction system, which generates and propagates electrical impulses, differentiates from cardiomyocyte precursors. The conduction system is composed of slow-conducting (proximal) components, such as the sinoatrial (SA) and atrioventricular (AV) nodes, as well as fast-conducting (distal) components, including the His bundle, bundle branches, and Purkinje fibers. Precursors within the sinus venosus give rise to the SA node, whereas those within the AV canal mature into heterogeneous cell types that compose the AV node. Decremental conduction through the AV node delays electrical impulses between atria and ventricles, enabling sequential antegrade contraction. The AV node also reduces the transmission of higher impulse rates to the vulnerable ventricle, whereas the distal conduction system rapidly propagates each impulse throughout the ventricles. The conduction system is composed of complex and heterogeneous cell populations with distinct gap junction proteins and ion channels that define the particular local electrical properties. Developmental defects in the conduction system can lead to clinical electrophysiologic disorders, such as congenital heart block or pre-excitation (Wolff-Parkinson-White syndrome) **(Chap. 256)**.

■ ORIGIN OF VASCULAR CELLS

Smooth-muscle cells are of varied origin. Some upper-body arterial smooth-muscle cells derive from the neural crest, whereas lower-body arteries develop smooth-muscle cells from neighboring mesodermal structures. Embryonic endothelial progenitor cells are derived from mesoderm. In adults, resident vascular or bone marrow–derived endothelial progenitors may aid repair of damaged or aging arteries. Bone marrow clonality, increasingly prevalent in aging, may impart significant clonality to endothelial cell populations. Vascular stem cells resident in the vessel wall may give rise to some smooth-muscle cells in injured or atheromatous arteries.

THE BLOOD VESSEL

■ VASCULAR ULTRASTRUCTURE

Blood vessels participate in disease biology as well as physiologic function in virtually every organ system. The smallest blood vessels—capillaries—consist of a monolayer of endothelial cells on a basement membrane adjacent to a discontinuous layer of smooth-muscle-like cells known as *pericytes* **(Fig. 244-2A)**. Arteries typically have

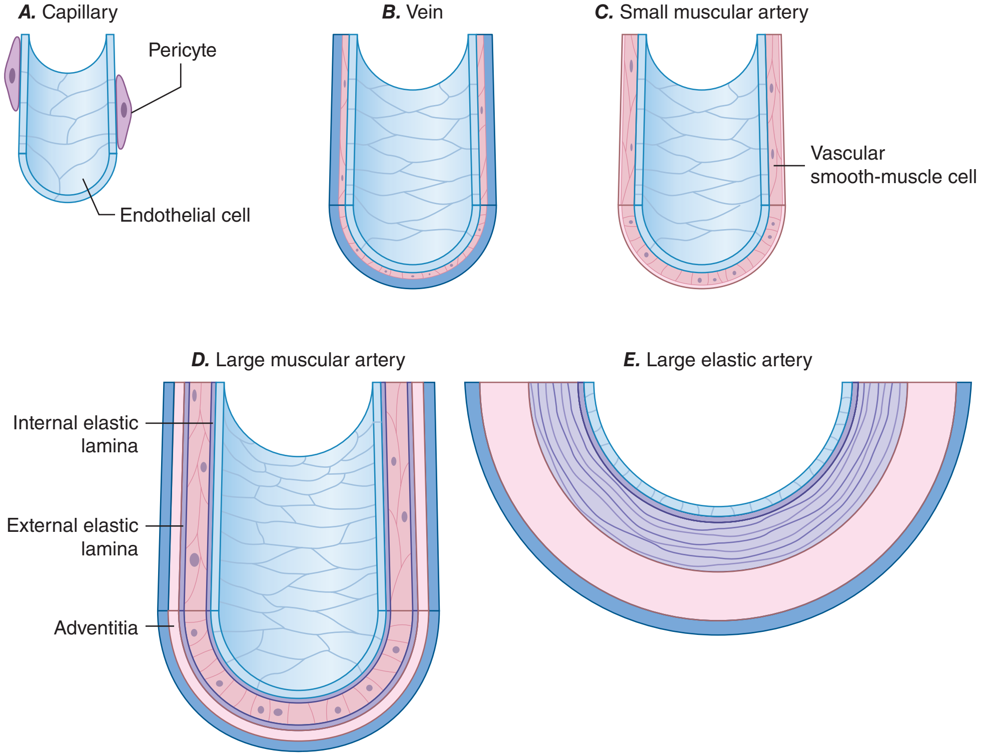

FIGURE 244-2 Schematics of the structures of various types of blood vessels. ***A.*** Capillaries consist of an endothelial tube in contact with a discontinuous population of pericytes. ***B.*** Veins typically have thin medias and thicker adventitias. ***C.*** A small muscular artery features a prominent tunica media. ***D.*** Larger muscular arteries have a prominent media with smooth-muscle cells embedded in a complex extracellular matrix. ***E.*** Larger elastic arteries have cylindrical layers of elastic tissue alternating with concentric rings of smooth-muscle cells as well as *vasa vasorum* to facilitate tissue blood supply.

a trilaminar structure **(Fig. 244-2*B–E*)**. The *intima* consists of a monolayer of endothelial cells continuous with those of the capillaries. The middle layer, or *tunica media*, consists of a syncytium of smooth-muscle cells that in veins are much sparser than in arteries (Fig. 244-2*B*). The outer layer, or *adventitia*, consists of extracellular matrix with fibroblasts, mast cells, and nerve terminals. Larger arteries require nourishment of the tunica media that is accomplished via their own vasculature, the *vasa vasorum* (Fig. 244-2*E*).

Arterioles are small muscular arteries (Fig. 244-2*C*) that regulate blood pressure and flow through arterial beds. Medium-size muscular arteries also contain prominent smooth-muscle layers (Fig. 244-2*D*) that participate in atherogenesis. Larger elastic arteries have a highly structured tunica media with concentric bands of smooth-muscle cells, interspersed with strata of elastin-rich extracellular matrix (Fig. 244-2*E*). Larger arteries form an internal elastic lamina between intima and media while an external elastic lamina partitions the media from surrounding adventitia.

VASCULAR CELL BIOLOGY

Endothelial Cell The endothelium forms the interface between tissues and the blood compartment, regulating the passage of molecules and cells. This function of endothelial cells as a selectively permeable barrier fails in vascular diseases, including atherosclerosis, hypertension, and renal disease, as well as in pulmonary edema, sepsis, and other situations exhibiting “capillary leak.”

The endothelium also participates in the local regulation of vascular tone and blood flow. Endogenous endothelium-derived substances, such as prostacyclin, endothelium-derived hyperpolarizing factor, nitric oxide (NO), and hydrogen peroxide (H_2O_2), provide tonic stimulation of endothelial homeostatic properties under physiologic conditions in vivo **(Table 244-1)**. Impaired production or excess catabolism of these substances can mediate dysfunctional properties of the endothelium. A major homeostatic influence on the endothelium is laminar blood flow, and the measurement of flow-mediated dilatation directly assesses endothelial vasodilator function in humans **(Fig. 244-3)**.

Endothelial cells also produce potent vasoconstrictor substances such as endothelin. Excessive production of reactive oxygen species, such as superoxide anion (O_2^-), by endothelial or smooth-muscle cells under pathologic conditions (e.g., excessive exposure to angiotensin II) can promote local oxidative stress and inactivate NO.

Endothelial cells also regulate cellular traffic through tissues. Normal endothelium exhibits limited interaction with circulating leukocytes, but bacterial products such as endotoxin or proinflammatory cytokines can induce endothelial cells to express an array of adhesion molecules that selectively bind various classes of leukocytes in different pathologic conditions. The adhesion molecules and chemokines generated during acute bacterial infection tend to recruit granulocytes, while in chronic inflammatory diseases such as tuberculosis or atherosclerosis, the adhesion molecules expressed favor monocyte recruitment. Endothelial cell injury participates in the pathophysiology of many immune-mediated diseases. For example, complement-mediated lysis of endothelial cells contributes to tissue injury. The foreign histocompatibility complex antigens on endothelial cells in solid-organ allografts can promote allograft arteriopathy, while immune-mediated endothelial injury also plays a role in thrombotic thrombocytopenic purpura or hemolytic-uremic syndrome.

The endothelium also regulates the balance between thrombosis and hemostasis through a highly tuned set of regulatory pathways. For example, inflammatory cytokines, bacterial endotoxin, or angiotensin II can activate endothelial cells to produce substantial quantities of plasminogen activator inhibitor 1 (PAI-1), the major inhibitor of fibrinolysis. Inflammatory stimuli also induce endothelial expression of the potent procoagulant tissue factor, a contributor to disseminated intravascular coagulation in sepsis; similar effects are observed in hyperglycemia. Thus, in pathologic circumstances, endothelial dysfunction tends to promote local thrombus accumulation rather than combat it.

Endothelial cells regulate the growth of subjacent smooth-muscle cells by elaborating heparan sulfate glycosaminoglycans that inhibit smooth-muscle proliferation. In the setting of vascular injury, endothelium-derived growth factors and chemoattractants (e.g., platelet-derived growth factor) induce the migration and proliferation of vascular smooth-muscle cells. Dysregulation of these growth-stimulatory molecules may promote smooth-muscle accumulation in atherosclerotic lesions.

TABLE 244-1 Endothelial Functions in Health and Disease

HOMEOSTATIC PROPERTIES	DYSFUNCTIONAL PROPERTIES
Optimize balance between vasodilation and vasoconstriction	Impaired dilation, vasoconstriction
Antithrombotic, profibrinolytic	Prothrombotic, antifibrinolytic
Anti-inflammatory	Proinflammatory
Antiproliferative	Proproliferative
Antioxidant	Prooxidant
Selective permeability	Impaired barrier function

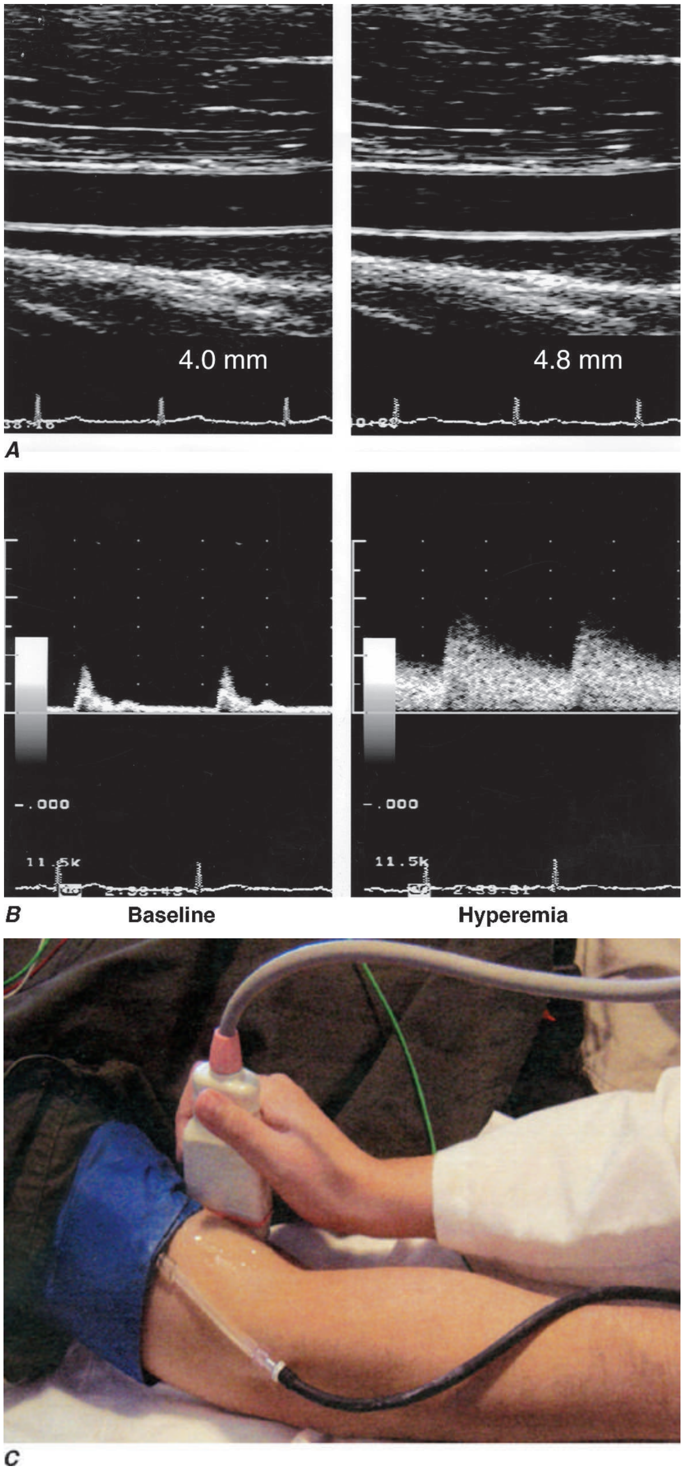

FIGURE 244-3 Assessment of endothelial function in vivo using blood pressure cuff occlusion and release. Upon deflation of the cuff, an ultrasound probe monitors changes in diameter (***A***) and blood flow (***B***) of the brachial artery (***C***). *(Reproduced with permission of J. Vita, MD.)*

Vascular Smooth-Muscle Cell Contraction and relaxation of vascular smooth-muscle cells in muscular arteries determine blood pressure, regional flow, and the afterload experienced by the left ventricle (see below). Venous tone regulates venous tree capacitance and influences ventricular preload. Smooth-muscle cells in the adult vessel seldom replicate in the absence of arterial injury or inflammatory activation, but proliferation and migration of arterial smooth-muscle cells contribute to arterial stenoses in atherosclerosis, arteriolar remodeling in hypertension, and the hyperplastic response of arteries to injury. In the pulmonary circulation, smooth-muscle migration and proliferation underlie the vascular pathology that occurs in sustained high-flow states such as left-to-right shunts in congenital heart disease with resulting pulmonary hypertension.

Smooth-muscle cells secrete the bulk of vascular extracellular matrix. Excessive production of collagen and glycosaminoglycans contributes to the remodeling, altered biomechanics, and physiology of arteries affected by hypertension or atherosclerosis. In larger elastic arteries, such as the aorta, the ability to store the kinetic energy of systole promotes tissue perfusion during diastole. Arterial stiffness, associated with aging or disease, increases left ventricular afterload and portends a poor outcome.

Like endothelial cells, vascular smooth-muscle cells not only respond to paracrine stimuli from other cells but can themselves serve as a source of such stimuli. For example, proinflammatory stimuli induce smooth-muscle cells to elaborate cytokines and other mediators that drive thrombosis and fibrinolysis as well as proliferation.

Vascular Smooth-Muscle Cell Contraction The principal mechanism for vascular smooth-muscle cell contraction is increased cytoplasmic calcium concentration due to transmembrane influx and triggered release from intracellular calcium stores **(Fig. 244-4)**. In vascular smooth-muscle cells, voltage-dependent L-type calcium channels open with membrane depolarization. Local influx of calcium, termed *calcium sparks*, can trigger release from intracellular stores, which results in more contraction and increased vessel tone (see below). Opposing currents balance the effects of individual ionic fluxes promoting homeostasis, which is tightly regulated by neural and metabolic influences.

Vasoconstricting agonists also increase intracellular [Ca^{2+}] by various mechanisms including receptor-dependent phospholipase C activation producing hydrolysis of phosphatidylinositol 4,5-bisphosphate to generate diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP_3). These membrane lipid derivatives, in turn, activate protein kinase C and increase intracellular [Ca^{2+}]. In addition, IP_3 binds specific sarcoplasmic reticulum (SR) receptors to increase calcium efflux from this storage pool into the cytoplasm.

Vascular smooth-muscle cell contraction depends on myosin light chain phosphorylation that reflects the balance between the activity of relevant kinases and phosphatases. Calcium activates myosin light chain kinase via calmodulin, augmenting myosin ATPase activity and enhancing contraction. Conversely, myosin light chain phosphatase reduces myosin ATPase activity and contractile force. Other kinase/phosphorylase combinations result in a complex regulatory network that refines vascular tone and links it to physiologic requirements.

Control of Vascular Smooth-Muscle Cell Tone The autonomic nervous system and endothelial cells modulate vascular smooth-muscle cells through similar convergent pathways. Autonomic neurons enter vessel media and modulate vascular smooth-muscle cell tone in response to baroreceptors and chemoreceptors within the aortic arch or carotid bodies and to thermoreceptors in the skin. Rapidly acting reflex arcs modulated by central inputs respond to multiple sensory inputs as well as emotional stimuli through three neuronal classes: *sympathetic*, whose principal neurotransmitters are epinephrine and norepinephrine; *parasympathetic*, whose principal neurotransmitter is acetylcholine; and *nonadrenergic/noncholinergic*, which include two subgroups—nitrergic, whose principal neurotransmitter is NO, and peptidergic, whose principal neurotransmitters are substance P, vasoactive intestinal peptide, calcitonin gene-related peptide, and the nonpeptide, adenosine triphosphate (ATP).

Each of these neurotransmitters acts through specific receptors on the vascular smooth-muscle cell to modulate intracellular Ca^{2+} and, consequently, contractile tone. Norepinephrine activates α-adrenergic receptors, and epinephrine activates both α and β receptors. In most blood vessels, norepinephrine activates postjunctional α_1 receptors in large arteries and α_2 receptors in small arteries and arterioles, leading to vasoconstriction. Most blood vessels express β_2-adrenergic receptors on their vascular smooth-muscle cells and respond to β agonists by cyclic AMP–dependent relaxation. Acetylcholine released from parasympathetic neurons may bind to muscarinic receptors on either vascular smooth-muscle cells, causing vasoconstriction, or on

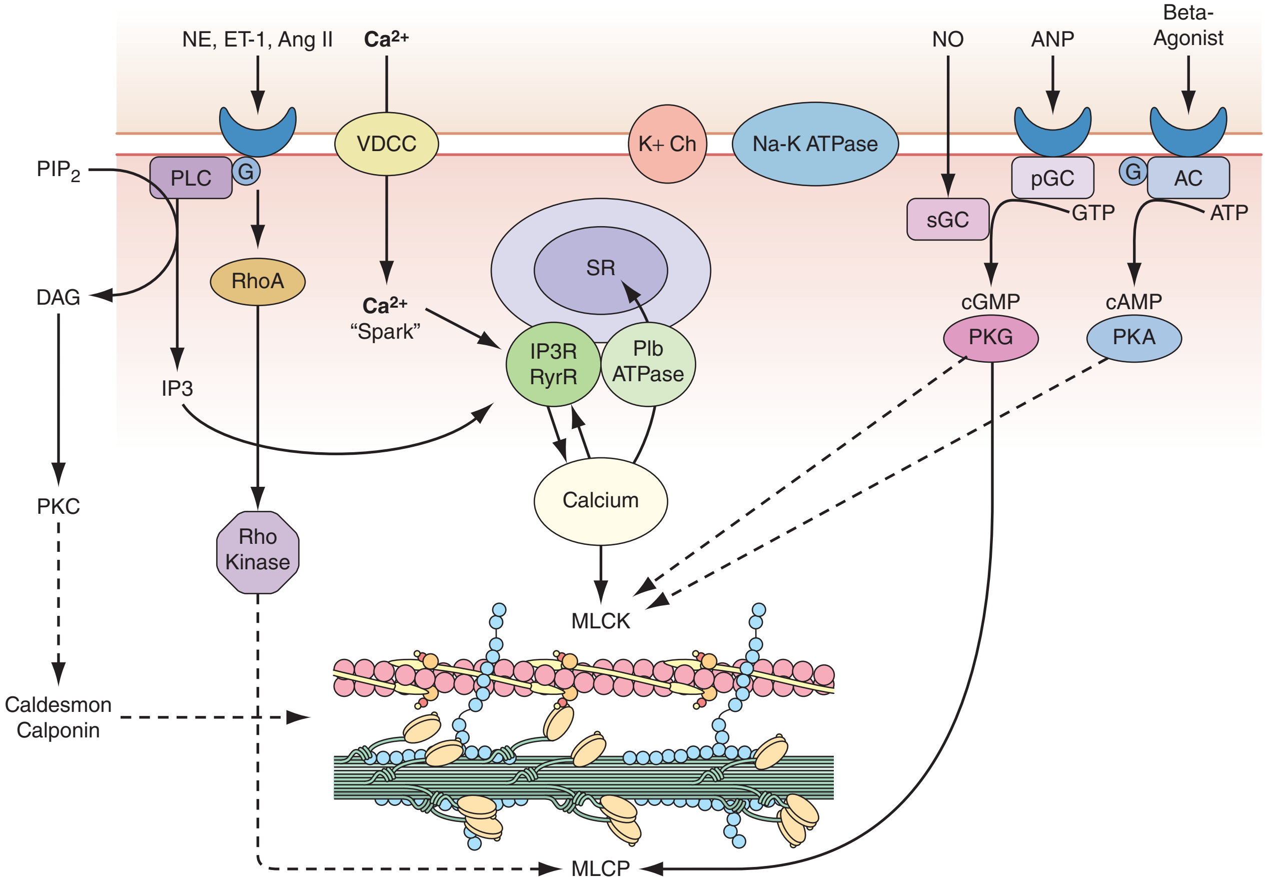

FIGURE 244-4 Regulation of vascular smooth-muscle cell calcium concentration and actomyosin ATPase-dependent contraction. AC, adenylyl cyclase; Ang II, angiotensin II; ANP, atrial natriuretic peptide; DAG, diacylglycerol; ET-1, endothelin-1; G, G protein; IP_3, inositol 1,4,5-trisphosphate; MLCK, myosin light chain kinase; MLCP, myosin light chain phosphatase; NE, norepinephrine; NO, nitric oxide; pGC, particular guanylyl cyclase; PIP_2, phosphatidylinositol 4,5-bisphosphate; PKA, protein kinase A; PKC, protein kinase C; PKG, protein kinase G; PLC, phospholipase C; sGC, soluble guanylyl cyclase; SR, sarcoplasmic reticulum; VDCC, voltage-dependent calcium channel. *Solid lines* depict stimulatory interaction, and *dashed lines* represent inhibition. *(Reproduced with permission from B Berk, in Vascular Medicine, 3rd ed. Philadelphia, Saunders, Elsevier; 2006.)*

endothelial cells, causing NO-dependent vasorelaxation. Nitrergic neurons release NO, which relaxes vascular smooth-muscle cell via the cyclic GMP–dependent and –independent mechanisms outlined, and other peptidergic inputs that regulate vascular tone. **For the detailed molecular physiology of the autonomic nervous system, see Chap. 451.**

The release of endothelial effectors of vascular smooth-muscle cell tone integrates the smooth-muscle response to mechanical (e.g., shear stress, cyclic strain) and biochemical stimuli (purinergic agonists, muscarinic agonists, peptidergic agonists). In addition to these local paracrine modulators, a complex system of circulating modulators ranging from norepinephrine to the natriuretic peptides also modify vascular smooth-muscle cell tone.

ARTERIOGENESIS AND ANGIOGENESIS

Recruitment and growth of blood vessels (arteriogenesis) and new capillaries (angiogenesis) can occur in response to conditions such as chronic hypoxemia and tissue ischemia. Growth factors, including vascular endothelial growth factor (VEGF) and fibroblast growth factor (FGF), can activate a signaling cascade that stimulates endothelial proliferation and tube formation, defined as *angiogenesis*. Guidance molecules, including members of the semaphorin family of secreted peptides, direct blood vessel patterning by attracting or repelling nascent endothelial tubes. The recruitment and expansion of preexisting collateral vascular networks in response to a blocked artery, an example of arteriogenesis, can result from selective activation of both growth factors and, perhaps, local or circulating endothelial progenitor cells. True vascular regeneration, or the development of a new blood vessel that includes all three cell layers, normally does not occur in adult mammals, but recent scientific advances might help obviate such limitations.

CELLULAR BASIS OF CARDIAC CONTRACTION

CARDIAC ULTRASTRUCTURE

Most of the ventricular mass is composed of cardiomyocytes, normally 60–140 μm in length and 17–25 μm in diameter **(Fig. 244-5A)**. Each cell contains multiple myofibrils that run the length of the cell and are composed of series of repeating sarcomeres. The cytoplasm between the myofibrils contains other cell constituents, including a single centrally located nucleus, mitochondria, and the intracellular membrane system, the SR.

The *sarcomere*, the structural and functional unit of contraction, lies between adjacent Z lines, which on transmission electron microscopy are seen as dark repeating bands. The distance between Z lines varies with the degree of contraction or stretch of the muscle and ranges between 1.6 and 2.2 μm. At the center of the sarcomere is a dark band of constant length (1.5 μm), the A band, which is flanked by two lighter bands, the I bands, which are of variable length. The sarcomere of heart muscle, like that of skeletal muscle, consists of interdigitating thick and thin myofilaments. Thicker filaments, composed principally of the protein myosin, traverse the A band; they are about 10 nm (100 Å) in diameter, with tapered ends. Thinner filaments, composed primarily of actin, course from the Z lines through the I band into the A band; they are ~5 nm (50 Å) in diameter and 1.0 μm in length. Thus, thick and thin filaments overlap only within the (dark) A band, whereas the (light) I band contains only thin filaments. On electron-microscopic examination, bridges extend between the thick and thin filaments within the A band; these are myosin heads (see below) bound to actin filaments.

THE CONTRACTILE PROCESS

The sliding filament model for muscle contraction rests on the central observation that both the thick and the thin filaments are constant in

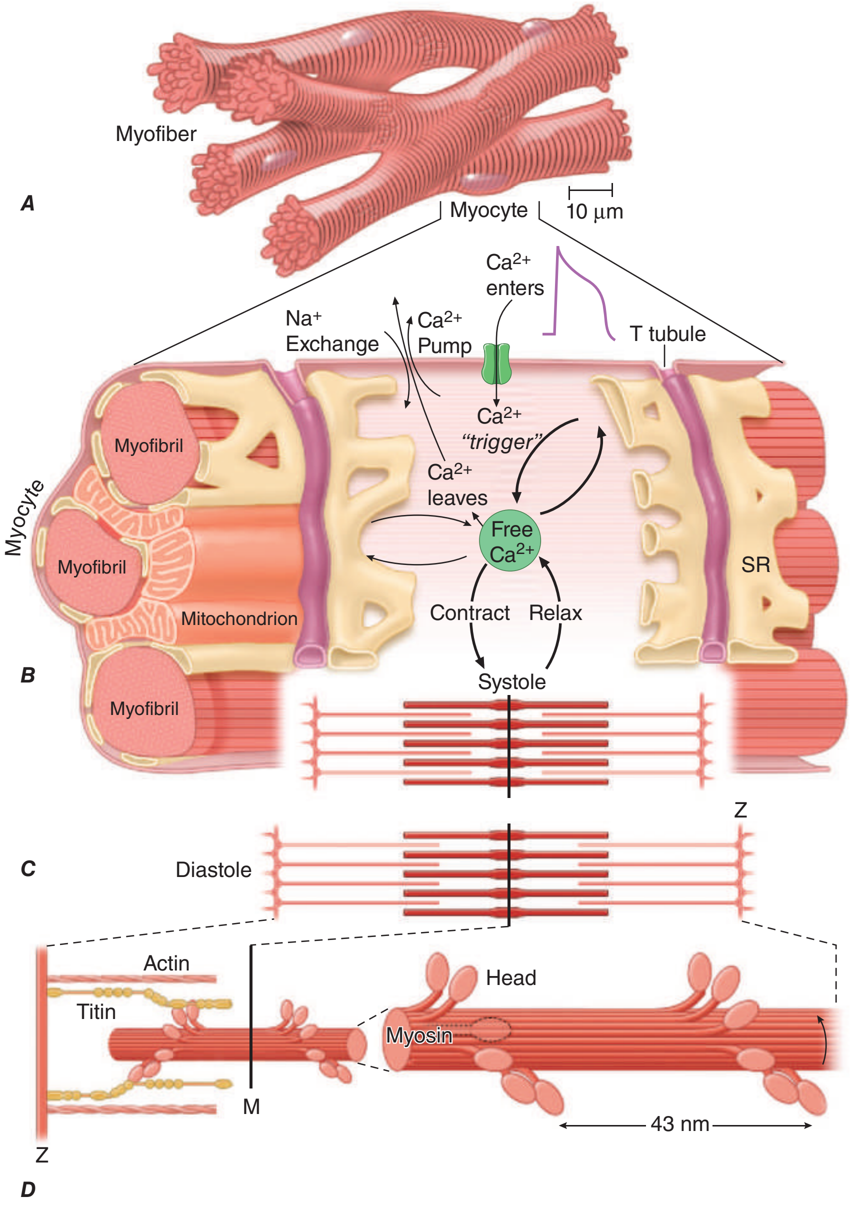

FIGURE 244-5 ***A*** shows the branching myocytes making up the cardiac myofibers. ***B*** illustrates the critical role played by the changing [Ca^{2+}] in the myocardial cytosol. Ca^{2+} ions are schematically shown as entering through the calcium channel that opens in response to the wave of depolarization that travels along the sarcolemma. These Ca^{2+} ions "trigger" the release of more calcium from the sarcoplasmic reticulum (SR) and thereby initiate a contraction-relaxation cycle. Eventually the small quantity of Ca^{2+} that has entered the cell leaves predominantly through an Na^+/Ca^{2+} exchanger, with a lesser role for the sarcolemmal Ca^{2+} pump. The varying actin-myosin overlap is shown for (***B***) systole, when [Ca^{2+}] is maximal, and (***C***) diastole, when [Ca^{2+}] is minimal. ***D.*** The myosin heads, attached to the thick filaments, interact with the thin actin filaments. *(Reproduced with permission from LH Opie: Heart Physiology: From Cell to Circulation, 4th ed. Philadelphia, Lippincott, Williams & Wilkins, 2004.)*

length during both contraction and relaxation. With activation, the actin filaments are propelled farther into the A band. In this process, the A band remains constant in length, whereas the I band shortens and the Z lines move toward one another.

The *myosin* molecule is a complex, asymmetric protein with a molecular mass of about 500,000 Da; it has a rod-like portion that is about 150 nm (1500 Å) in length with a globular portion (head) at its end. The globular portions of myosin form the bridges to actin and are the site of ATPase activity. In thick myofilaments, composed of ~300 longitudinally stacked myosin molecules, the rod-like segments of myosin assume an orderly, polarized orientation, with outwardly projecting globular heads interacting with actin to generate force and shorten **(Fig. 244-5*B*)**.

Actin has a molecular mass of about 47,000 Da. Thin filaments consist of a double helix of two chains of actin molecules wound about each other on a larger molecule, tropomyosin. A group of regulatory proteins—troponins C, I, and T—localize at regular intervals on this filament **(Fig. 244-6)**. In contrast to myosin, actin lacks intrinsic enzymatic activity but combines reversibly with myosin in the presence of ATP and Ca^{2+}. Calcium activates the myosin ATPase, which breaks down ATP to supply the energy for contraction (Fig. 244-6). The activity of myosin ATPase determines the rate of actomyosin cross-bridge formation and breakdown, and ultimately determines contraction velocity. In relaxed muscle, tropomyosin inhibits this interaction. *Titin* (Fig. 244-5*D*) an enormous, flexible, myofibrillar protein, connects myosin to the Z line; its elasticity contributes to the passive mechanical characteristics of the heart. Dystrophin, a cytoskeletal protein that binds to the dystroglycan complex at membrane adherens junctions, tethers the sarcomere to the cell membrane at these regions of tight coupling to adjacent myocytes. Mutations in multiple sarcomeric and cytoskeletal proteins cause different Mendelian disorders involving the heart and skeletal muscle and also sensitize individuals to toxic cardiomyopathies (e.g., due to alcohol or chemotherapy) and to those caused by other acquired stressors, such as inflammatory or peripartum cardiomyopathy.

During activation of the cardiac myocyte, Ca^{2+} binds the heterotrimer troponin C, resulting in regulatory conformational changes in tropomyosin and exposing actin cross-bridge interaction sites

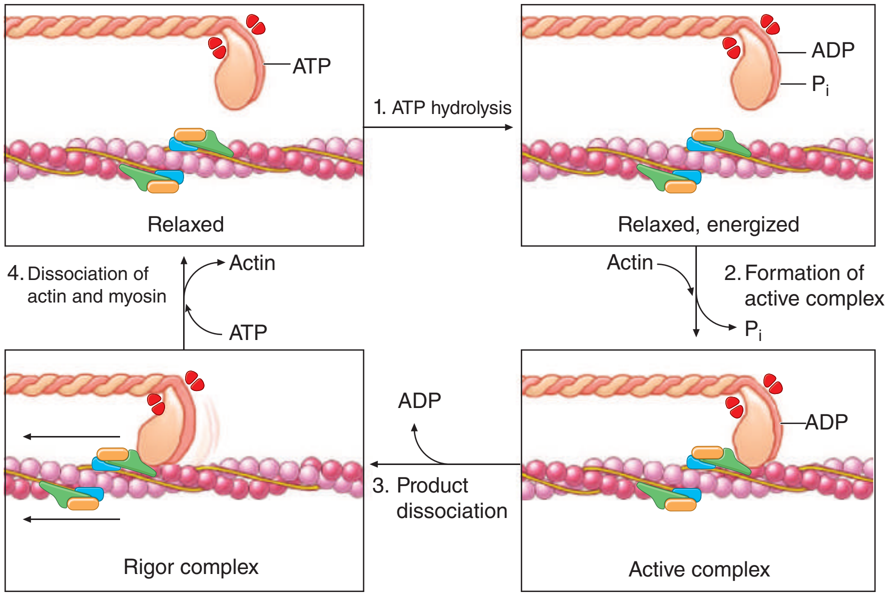

FIGURE 244-6 Four steps in cardiac muscle contraction and relaxation. In relaxed muscle (*upper left*), ATP bound to the myosin cross-bridge dissociates the thick and thin filaments. ***Step 1:*** Hydrolysis of myosin-bound ATP by the ATPase site on the myosin head transfers the chemical energy of the nucleotide to the activated cross-bridge (*upper right*). When cytosolic Ca^{2+} concentration is low, as in relaxed muscle, the reaction cannot proceed because tropomyosin and the troponin complex on the thin filament do not allow the active sites on actin to interact with the cross-bridges. Therefore, even though the cross-bridges are energized, they cannot interact with actin. ***Step 2:*** When Ca^{2+} binding to troponin C has exposed active sites on the thin filament, actin interacts with the myosin cross-bridges to form an active complex (*lower right*) in which the energy derived from ATP is retained in the actin-bound cross-bridge, whose orientation has not yet shifted. ***Step 3:*** The muscle contracts when ADP dissociates from the cross-bridge. This step leads to the formation of the low-energy rigor complex (*lower left*) in which the chemical energy derived from ATP hydrolysis has been expended to perform mechanical work (the "rowing" motion of the cross-bridge). ***Step 4:*** The muscle returns to its resting state, and the cycle ends when a new molecule of ATP binds to the rigor complex and dissociates the cross-bridge from the thin filament. This cycle continues until calcium is dissociated from troponin C in the thin filament, which causes the contractile proteins to return to the resting state with the cross-bridge in the energized state. ADP, adenosine diphosphate; ATP, adenosine triphosphate; ATPase, adenosine triphosphatase. *(Reproduced with permission from AM Katz: Heart failure: Cardiac function and dysfunction, in Atlas of Heart Diseases, 3rd ed, WS Colucci [ed]. Philadelphia, Current Medicine, 2002.)*

(Fig. 244-6). Repetitive interaction between myosin heads and actin filaments is termed *cross-bridge cycling* and results in sliding of the actin along the myosin filaments, with muscle shortening and/or the development of tension. The splitting of ATP then dissociates the myosin cross-bridge from actin. In the presence of ATP (Fig. 244-6), actin and myosin filaments bind and dissociate cyclically if sufficient Ca^{2+} is present; these processes cease when [Ca^{2+}] falls below a critical level, and the troponin-tropomyosin complex once more inhibits actin-myosin interactions **(Fig. 244-7)**.

Cytoplasmic [Ca^{2+}] is a principal determinant of the inotropic state of the heart. Most agents that stimulate myocardial contractility (positive inotropic stimuli), including digitalis glycosides and β-adrenergic agonists, increase cytoplasmic [Ca^{2+}], triggering cross-bridge cycling. Increased adrenergic neuronal activity stimulates myocardial contractility through norepinephrine release, activation of β-adrenergic receptors, and, via G$_s$-stimulated guanine nucleotide-binding proteins, activation of the adenylyl cyclase, which leads to the formation of the intracellular second messenger cyclic AMP from ATP (Fig. 244-7). Cyclic AMP in turn activates protein kinase A (PKA), which phosphorylates sarcolemmal Ca^{2+} channels, thereby enhancing the influx of Ca^{2+} into the myocyte.

The SR **(Fig. 244-8)**, a complex network of anastomosing intracellular channels, invests the myofibrils. The transverse tubules, or T system, closely related to the SR, both structurally and functionally, arise as sarcolemmal invaginations that extend into the myofibrillar bundles along the Z lines, i.e., the ends of the sarcomeres.

CARDIAC ACTIVATION

In the inactive state, the cardiomyocyte membrane is electrically polarized; i.e., the interior has a negative charge relative to the outside of the cell, with a transmembrane potential of −80 to −100 mV **(Chap. 250)**. The sarcolemma, which in the resting state is largely impermeable to Na$^+$, and a Na$^+$- and K$^+$-pump energized by ATP extrudes Na$^+$ from the cell and maintains the resting potential. In this resting state, intracellular [K$^+$] is relatively high and [Na$^+$] is far lower; conversely, extracellular [Na$^+$] is high and [K$^+$] is low. At the same time, extracellular [Ca^{2+}] greatly exceeds free intracellular [Ca^{2+}].

The action potential has four phases **(see Fig. 250-1*B*)**. Depolarizing current spreads across the cell membrane, penetrating deeply into the cell via the T tubular system. During the action potential plateau (phase 2), there is a slow inward current through sarcolemmal L-type Ca^{2+} channels (Fig. 244-8). The absolute quantity of Ca^{2+} traversing sarcolemmal and T tubular membranes is modest and insufficient to fully activate contraction. However, this initial Ca^{2+} current, through *Ca^{2+}-induced Ca^{2+} release,* triggers substantial Ca^{2+} release from the SR, inducing contraction.

Ca^{2+} is released from the SR through a Ca^{2+} release channel, a cardiac isoform of the ryanodine receptor (RyR2). Several regulatory proteins inhibit RyR2 and thus SR Ca^{2+} release. Inherited disorders or exogenous factors affecting the efficiency or stability of SR Ca^{2+} handling can impair contraction, leading to heart failure or to ventricular arrhythmias.

The Ca^{2+} released from the SR diffuses to interact with myofibrillar troponin C (Fig. 244-7), repressing this protein's inhibition of contraction, and so activating myofilaments to shorten. During repolarization, the activity of the SR Ca^{2+}-ATPase (SERCA$_{2A}$) leads to Ca^{2+} uptake against a concentration gradient into the SR where it complexes with another specialized protein, *calsequestrin*. The uptake of Ca^{2+} is ATP (energy)-dependent and lowers cytoplasmic [Ca^{2+}] to a level where actomyosin interaction is inhibited and myocardial relaxation occurs. There is also a sarcolemmal exchange of Ca^{2+} for Na$^+$ (Fig. 244-8), reducing cytoplasmic [Ca^{2+}]. Additional control of calcium compartmentalization results from cyclic AMP–dependent PKA phosphorylation of the SR protein *phospholamban*, permitting SERCA$_{2A \text{ activation}}$, increasing SR Ca^{2+} uptake, and so accelerating relaxation rates, and loading the SR with Ca^{2+} for subsequent cycles of release and contraction.

Thus, the combination of the cell membrane, transverse tubules, and SR, transmitting the action potential, releasing and then re-accumulating Ca^{2+}, controls the cyclic contraction and relaxation of heart muscle.

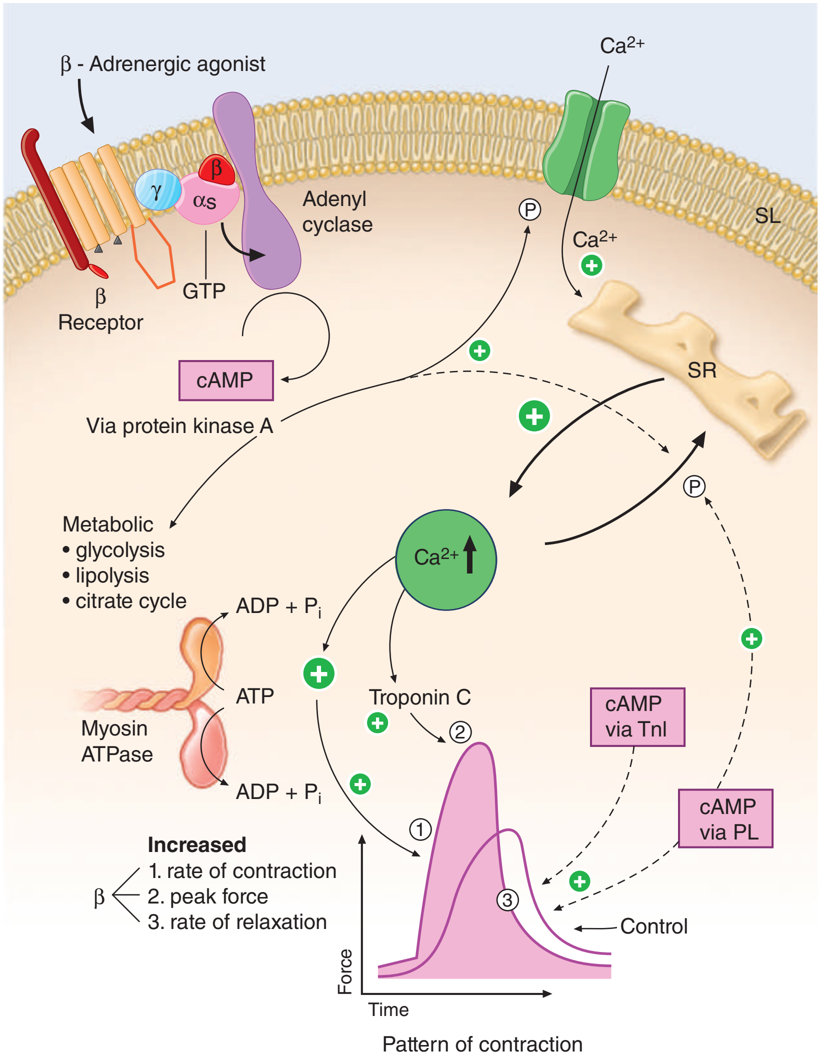

FIGURE 244-7 Signal systems involved in positive inotropic and lusitropic (enhanced relaxation) effects of α-adrenergic stimulation. When the β-adrenergic agonist interacts with the β receptor, a series of G protein–mediated changes leads to activation of adenylyl cyclase and the formation of cyclic adenosine monophosphate (cAMP). The latter acts via protein kinase A to stimulate metabolism (*left*) and phosphorylate the Ca^{2+} channel protein (*right*). The result is an enhanced opening probability of the Ca^{2+} channel, thereby increasing the inward movement of Ca^{2+} ions through the sarcolemma (SL) of the T tubule. These Ca^{2+} ions release more calcium from the sarcoplasmic reticulum (SR) to increase cytosolic Ca^{2+} and activate troponin C. Ca^{2+} ions also increase the rate of breakdown of adenosine triphosphate (ATP) to adenosine diphosphate (ADP) and inorganic phosphate (Pi). Enhanced myosin ATPase activity explains the increased rate of contraction, with increased activation of troponin C explaining increased peak force development. An increased rate of relaxation results from the ability of cAMP to activate as well the protein phospholamban, situated on the membrane of the SR, that controls the rate of uptake of calcium into the SR. The latter effect explains enhanced relaxation (lusitropic effect). P, phosphorylation; PL, phospholamban; TnI, troponin I. *(Reproduced with permission from LH Opie: Heart Physiology: From Cell to Circulation, 4th ed. Philadelphia, Lippincott, Williams & Wilkins, 2004.)*

Genetic or pharmacologic alterations of any component can disturb any of the functions of this finely tuned system.

CONTROL OF CARDIAC PERFORMANCE AND OUTPUT

The extent of shortening of heart muscle and, therefore, ventricular stroke volume in the intact heart depends on three major influences: (1) the length of the muscle at the onset of contraction, i.e., the preload; (2) the tension that the muscle must develop during contraction, i.e., the afterload; and (3) muscle contractility, i.e., the extent and velocity of shortening at any given preload and afterload. Table 244-2 lists the major determinants of preload, afterload, and contractility.

THE ROLE OF MUSCLE LENGTH (PRELOAD)

Preload determines sarcomere length at the onset of contraction. Contractile force is optimal at specific sarcomere lengths (~2.2 μm) where both myofilament Ca^{2+} sensitivity is maximal, and myofilament interactions and activation of contraction are most efficient. The relationship between initial muscle fiber length and the developed force is the basis of Starling's law of the heart, which states that, within limits, the ventricular contraction force depends on the end-diastolic length of the cardiac muscle; in vivo, end-diastolic length relates closely to the ventricular end-diastolic volume.

CARDIAC PERFORMANCE

Ventricular end-diastolic or "filling" pressure can serve as a surrogate for end-diastolic volume. In isolated heart and heart-lung preparations, stroke volume varies directly with the end-diastolic fiber length (preload) and inversely with the arterial resistance (afterload), and as the heart fails—i.e., as its contractility declines—it delivers a progressively smaller stroke volume from a normal or even elevated end-diastolic volume. The relation between ventricular end-diastolic pressure and the stroke work of the ventricle (the ventricular function curve) provides a working definition of cardiac contractility in the intact organism. An increase in contractility is accompanied by a shift of the ventricular function curve upward and to the left (greater stroke work at any level of ventricular end-diastolic pressure, or lower end-diastolic volume at any level of stroke work), whereas a shift downward and to the right characterizes reduction of contractility (Fig. 244-9).

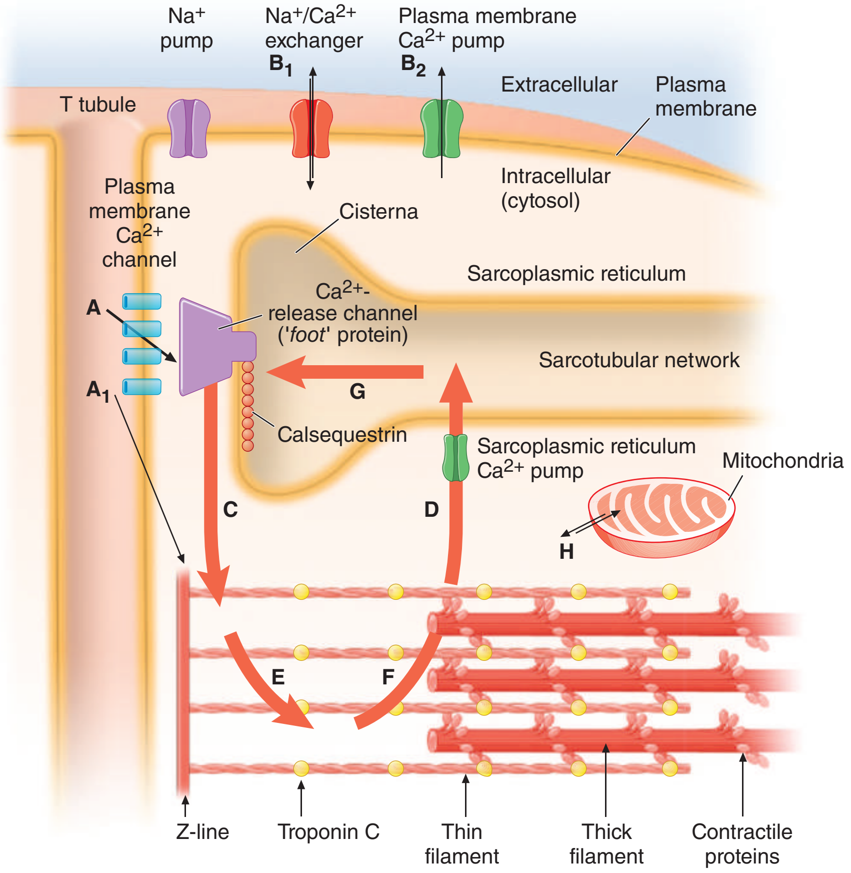

FIGURE 244-8 The Ca^{2+} fluxes and key structures involved in cardiac excitation-contraction coupling. The *arrows* denote the direction of Ca^{2+} fluxes. The *thickness* of each arrow indicates the magnitude of the calcium flux. Two Ca^{2+} cycles regulate excitation-contraction coupling and relaxation. The larger cycle is entirely intracellular and involves Ca^{2+} fluxes into and out of the sarcoplasmic reticulum, as well as Ca^{2+} binding to and release from troponin C. The smaller extracellular Ca^{2+} cycle occurs when this cation moves into and out of the cell. The action potential opens plasma membrane Ca^{2+} channels to allow passive entry of Ca^{2+} into the cell from the extracellular fluid (*arrow A*). Only a small portion of the Ca^{2+} that enters the cell directly activates the contractile proteins (*arrow A_1*). The extracellular cycle is completed when Ca^{2+} is actively transported back out to the extracellular fluid by way of two plasma membrane fluxes mediated by the sodium-calcium exchanger (*arrow B_1*) and the plasma membrane calcium pump (*arrow B_2*). In the intracellular Ca^{2+} cycle, passive Ca^{2+} release occurs through channels in the cisternae (*arrow C*) and initiates contraction; active Ca^{2+} uptake by the Ca^{2+} pump of the sarcotubular network (*arrow D*) relaxes the heart. Diffusion of Ca^{2+} within the sarcoplasmic reticulum (*arrow G*) returns this activator cation to the cisternae, where it is stored in a complex with calsequestrin and other calcium-binding proteins. Ca^{2+} released from the sarcoplasmic reticulum initiates systole when it binds to troponin *C* (*arrow E*). Lowering of cytosolic [Ca^{2+}] by the sarcoplasmic reticulum (SR) causes this ion to dissociate from troponin (*arrow F*) and relaxes the heart. Ca^{2+} also may move between mitochondria and cytoplasm (H). *(Reproduced with permission from AM Katz: Physiology of the Heart, 4th ed. Philadelphia, Lippincott, Williams & Wilkins, 2005.)*

VENTRICULAR AFTERLOAD

In the intact heart, as ex vivo, the extent and velocity of shortening of ventricular muscle fibers at any level of preload and of myocardial contractility relate inversely to the afterload, i.e., the instantaneous load opposing shortening. In the intact heart, the afterload may be defined as the tension developed in the ventricular wall during ejection. Afterload is determined by the aortic pressure as well as by the volume of the ventricular cavity and myocardial tissue characteristics including thickness. Laplace's law models the tension of the myocardial fiber as the product of intra-cavitary ventricular pressure and ventricular radius divided by wall thickness. Therefore, at any given aortic pressure, the afterload on a dilated left ventricle exceeds that on a normal-sized ventricle. Conversely, at the same aortic pressure and ventricular diastolic volume, the afterload on a hypertrophied ventricle is lower than that on a normal chamber. Aortic pressure in turn depends on the peripheral vascular resistance, the biomechanics of the arterial tree, and the volume of blood it contains at the onset of ejection.

Ventricular afterload finely regulates cardiovascular performance **(Fig. 244-10)**. As noted, elevations in both preload and contractility increase myocardial fiber shortening, whereas increases in afterload reduce it. The extent of myocardial fiber shortening and left ventricular size determine stroke volume. An increase in arterial pressure induced by vasoconstriction, for example, augments afterload, which opposes myocardial fiber shortening, reducing stroke volume.

When myocardial contractility is impaired and the ventricle dilates, afterload rises (Laplace's law) and limits cardiac output. Increased afterload also may result from neural and humoral stimuli that occur in response to a fall in cardiac output. This increased afterload may reduce cardiac output further, thereby increasing ventricular volume and initiating a vicious circle, especially in patients with ischemic heart disease and limited myocardial O_2 supply. Treatment with vasodilators has the opposite effect; when afterload falls, cardiac output rises **(Chaps. 266–270)**.

Under normal circumstances, the various influences acting on cardiac performance interact in a complex fashion to maintain cardiac output at a level responsive to the requirements of tissue metabolic demands (Fig. 244-10). Interference with a single mechanism may not influence the cardiac output due to homeostatic adjustments. For example, a moderate reduction of blood volume or the loss of the atrial contribution to ventricular contraction can be tolerated without a reduction in resting cardiac output. Under these circumstances, other factors, such as adrenergic neuronal impulses increasing cardiac contractility, heart rate, and venous tone, will serve as compensatory mechanisms and sustain cardiac output in a normal individual. Ultimately, understanding the complex interactions between so many different phasic variables requires rigorous models to predict relevant outcomes, and led to the early application of systems engineering principles in medicine.

TABLE 244-2 Determinants of Stroke Volume

- I. **Ventricular Preload**
  - A. Blood volume
  - B. Distribution of blood volume
    - 1. Body position
    - 2. Intrathoracic pressure
    - 3. Intrapericardial pressure
    - 4. Venous tone
    - 5. Pumping action of skeletal muscles
  - C. Atrial contraction
- II. **Ventricular Afterload**
  - A. Systemic vascular resistance
  - B. Elasticity of arterial tree
  - C. Arterial blood volume
  - D. Ventricular wall tension
    - 1. Ventricular radius
    - 2. Ventricular wall thickness
- III. **Myocardial Contractility**[a]
  - A. Intramyocardial $[Ca^{2+}]$ ↑↓
  - B. Cardiac adrenergic nerve activity ↑↓[b]
  - C. Circulating catecholamines ↑↓[b]
  - D. Cardiac rate ↑↓[b]
  - E. Exogenous inotropic agents ↑
  - F. Myocardial ischemia ↓
  - G. Myocardial cell death (necrosis, apoptosis, autophagy) ↓
  - H. Alterations of sarcomeric and cytoskeletal proteins ↓
    - 1. Genetic
    - 2. Hemodynamic overload
  - I. Myocardial fibrosis ↓
  - J. Chronic overexpression of neurohormones ↓
  - K. Ventricular remodeling ↓
  - L. Chronic and/or excessive myocardial hypertrophy ↓

[a] Arrows indicate directional effects of determinants of contractility. [b] Contractility rises initially but later becomes depressed.

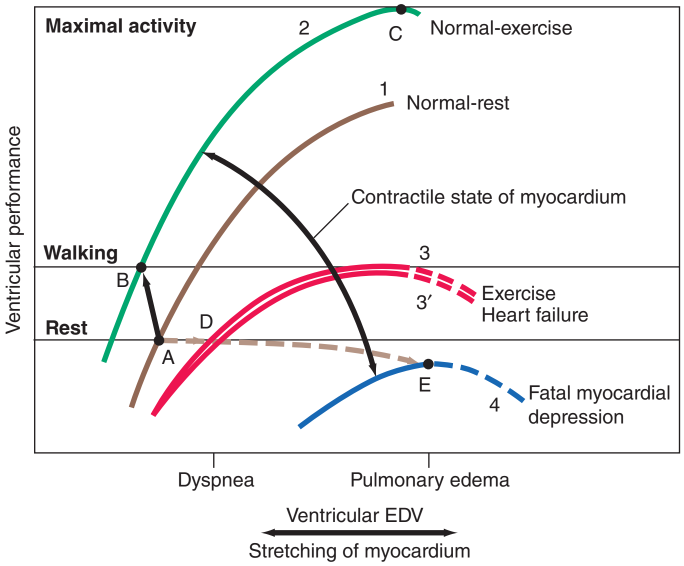

FIGURE 244-9 The interrelations among influences on ventricular end-diastolic volume (EDV) through stretching of the myocardium and the contractile state of the myocardium. Levels of ventricular EDV associated with filling pressures that result in dyspnea and pulmonary edema are shown on the abscissa. Levels of ventricular performance required when the subject is at rest, while walking, and during maximal activity are designated on the ordinate. The *broken lines* are the descending limbs of the ventricular-performance curves, which are rarely seen during life but show the level of ventricular performance if end-diastolic volume could be elevated to very high levels. For further explanation, see text. *(Reproduced with permission from WS Colucci, and E Braunwald: Pathophysiology of heart failure, in Braunwald's Heart Disease, 7th ed, Philadelphia: Elsevier, 2005.)*

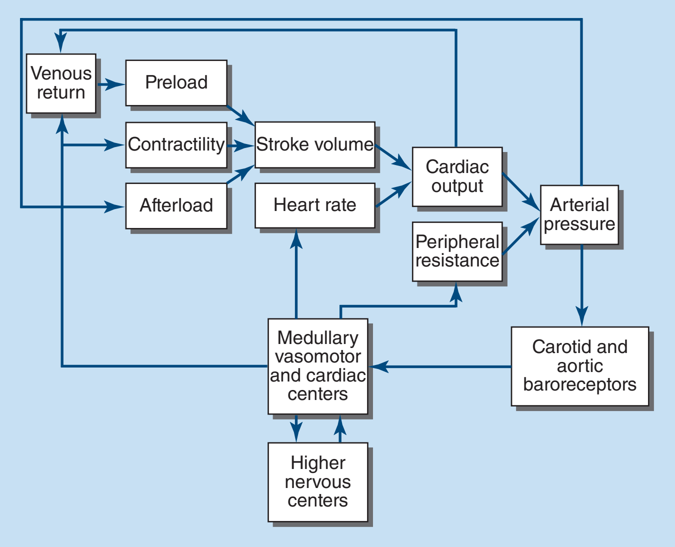

FIGURE 244-10 Interactions in the intact circulation of preload, contractility, and afterload in producing stroke volume. Stroke volume combined with heart rate determines cardiac output, which, when combined with peripheral vascular resistance, determines arterial pressure for tissue perfusion. The characteristics of the arterial system also contribute to afterload, an increase that reduces stroke volume. The interaction of these components with carotid and aortic arch baroreceptors provides a feedback mechanism to higher medullary and vasomotor cardiac centers and to higher levels in the central nervous system to effect a modulating influence on heart rate, peripheral vascular resistance, venous return, and contractility. *(Reproduced with permission from MR Starling: Physiology of myocardial contraction, in Atlas of Heart Failure: Cardiac Function and Dysfunction, 3rd ed, WS Colucci and E Braunwald [eds]. Philadelphia: Current Medicine; 2002.)*

EXERCISE

The integrated response to exercise illustrates typical interactions among the three determinants of stroke volume: preload, afterload, and contractility (Fig. 244-9). Hyperventilation, the pumping action of the exercising muscles, and venoconstriction during exercise all augment venous return and hence ventricular filling and preload (Table 244-2). Simultaneously, the increase in neuronal and humoral adrenergic stimulation of the myocardium and the tachycardia that occur during exercise combine to augment the myocardial contractility (Fig. 244-9, curves 1 and 2), together elevating stroke volume and stroke work, with little or no change in end-diastolic pressure and volume (Fig. 244-9, points A and B). Vasodilation occurs in the exercising muscles, thus limiting the increase in afterload that otherwise would occur as cardiac output rises to levels as high as five times greater than basal levels during maximal exercise. This vasodilation ultimately allows the achievement of elevated cardiac outputs during exercise at arterial pressures only moderately higher than the resting state.

ASSESSMENT OF CARDIAC FUNCTION

Several techniques can define impaired cardiac function in clinical practice. Cardiac output and stroke volume may decline in the presence of heart failure, but these variables are often within normal limits, especially at rest, even late in disease. A more sensitive index of cardiac function is the ejection fraction, i.e., the ratio of stroke volume to end-diastolic volume (normal value = 67 ± 8%), which is frequently depressed in systolic heart failure even when stroke volume is normal. Alternatively, abnormally elevated ventricular end-diastolic volume (normal value = 75 ± 20 mL/m^2) or end-systolic volume (normal value = 25 ± 7 mL/m^2) signifies left ventricular systolic impairment.

Noninvasive techniques, particularly echocardiography, radionuclide scintigraphy, and cardiac magnetic resonance imaging (MRI) **(Chap. 248)**, have great value in the clinical assessment of myocardial function. They provide measurements of end-diastolic and end-systolic volumes, ejection fraction, and systolic shortening rate, and they allow assessment of ventricular filling (see below) as well as regional contraction, relaxation, and tissue characterization. The latter measurements

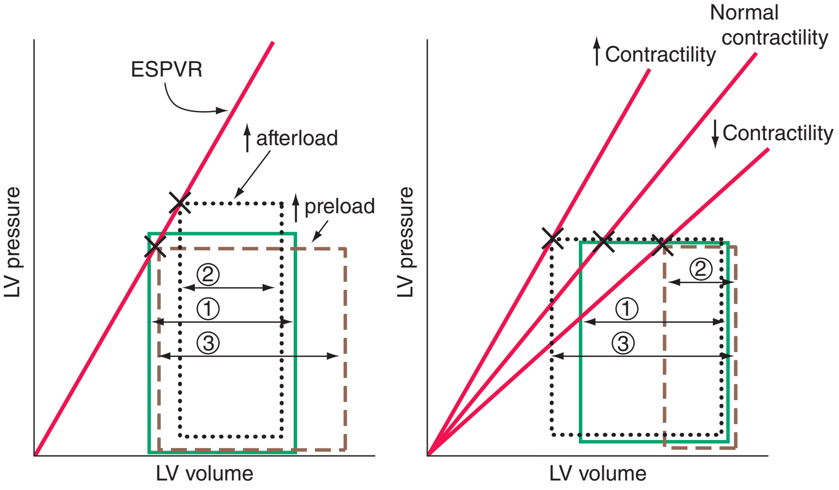

FIGURE 244-11 The responses of the left ventricle to increased afterload, increased preload, and increased and reduced contractility are shown in the pressure-volume plane. ***Left***. Effects of increases in preload and afterload on the pressure-volume loop. Because there has been no change in contractility, the end-systolic pressure-volume relationship (ESPVR) is unchanged. With an increase in afterload, stroke volume falls (1 → 2); with an increase in preload, stroke volume rises (1 → 3). ***Right***. With increased myocardial contractility and constant left ventricular end-diastolic volume, the ESPVR moves to the left of the normal line (lower end-systolic volume at any end-systolic pressure) and stroke volume rises (1 → 3). With reduced myocardial contractility, the ESPVR moves to the right; end-systolic volume is increased, and stroke volume falls (1 → 2).

have particular importance in ischemic heart disease, as myocardial infarction causes regional myocardial damage.

Strong dependence on ventricular loading conditions influences the precision of measurements of cardiac output, ejection fraction, and ventricular volumes as indices of cardiac function. Thus, a depressed ejection fraction and lowered cardiac output may occur in patients with normal ventricular function but reduced preload, as occurs in hypovolemia, or with increased afterload, as occurs in acutely elevated arterial pressure.

The end-systolic left ventricular pressure-volume relationship has particular value as an index of ventricular performance as it does not depend on preload and afterload **(Fig. 244-11)**. At any level of myocardial contractility, left ventricular end-systolic volume varies inversely with end-systolic pressure; as contractility declines, end-systolic volume (at any level of end-systolic pressure) rises. Invasive measurement of end-systolic left ventricular pressure-volume loops adds rigor to research studies of left ventricular function, but these techniques are less pragmatic than the more readily assessed indices obtained in routine clinical practice, such as ventricular volumes and ejection fraction. Integrated cardiopulmonary exercise testing with formal analysis of exhaled gases is now more broadly available and can estimate maximal oxygen delivery as an indirect metric of physiologic reserve. Longitudinal measurements of some aspects of cardiovascular physiology are increasingly feasible with implantable or wearable devices.

■ DIASTOLIC FUNCTION

Ventricular filling is influenced by several characteristics of the myocardium including: (1) the extent and speed of myocardial relaxation; and (2) the passive stiffness of the ventricular wall. The former is largely a function of the rate of uptake of Ca^{2+} by the SR that may be enhanced by adrenergic activation and reduced by ischemia due to limited ATP available for pumping Ca^{2+} into the SR (see above). For the latter, ventricular stiffness increases with hypertrophy, fibrosis, and conditions that infiltrate the ventricle, such as amyloid, or can result from an extrinsic constraint (e.g., pericardial constriction) **(Fig. 244-12)**.

Ventricular filling can be assessed by measuring flow velocity across the mitral valve using Doppler ultrasound. Normally, inflow velocity is more rapid in early diastole than during atrial systole. However, with mild to moderately impaired relaxation, the rate of early diastolic filling declines, as presystolic filling rates rise. With further stiffening, flow is "pseudo-normalized," as early ventricular filling becomes more rapid with rising left atrial pressure upstream of the left ventricle.

■ CARDIAC METABOLISM

The heart requires a continuous supply of energy (ATP) not only to drive mechanical contraction, but also to maintain ionic and biochemical homeostasis. The development of tension, the frequency of contraction, and myocardial contractility levels are the principal determinants of the heart's energy and oxygen requirements, representing ~15% of that of the entire organism.

The heart's ATP production requires the generation of acetyl coenzyme A that can be derived from (in descending order) free fatty acids (FFAs), glucose, lactate, amino acids, and ketone bodies. Myocardial FFAs derive from circulating FFAs, whereas the cardiomyocyte's glucose derives from plasma as well as from myocardial glycogen stores (via glycogenolysis). These two principal sources of acetyl coenzyme

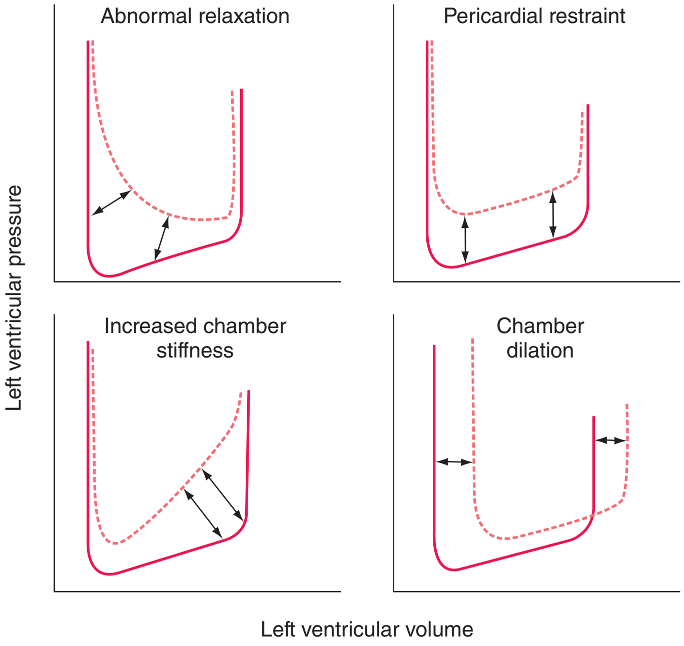

FIGURE 244-12 Mechanisms that cause diastolic dysfunction reflected in the pressure-volume relation. The bottom half of the pressure-volume loop is depicted. *Solid lines* represent normal subjects; *broken lines* represent patients with diastolic dysfunction. *(Reproduced with permission from JD Carroll et al: The differential effects of positive inotropic and vasodilator therapy on diastolic properties in patients with congestive cardiomyopathy. Circulation; 1986; 74: 815.)*

A are metabolized distinctly in cardiac muscle. Glucose is converted in the cytoplasm into pyruvate, which passes into mitochondria for conversion into acetyl-CoA that then undergoes oxidation. FFAs are converted to acyl-CoA in the cytoplasm and acetyl-CoA in the mitochondria. Acetyl-CoA enters the citric acid (Krebs) cycle to produce ATP by oxidative phosphorylation; ATP then enters the cytoplasm from the mitochondrial compartment. Intracellular adenosine diphosphate (ADP), resulting from ATP breakdown, enhances ATP production.

In the fasted, resting state, circulating FFAs furnish most of the heart's acetyl-CoA (~70%). In the fed state, with elevations of blood glucose and insulin, glucose oxidation increases and FFA oxidation subsides. Increased cardiac work, inotropic agents, hypoxia, and mild ischemia all enhance myocardial glucose uptake, production (glycogenolysis), and metabolism to pyruvate (glycolysis). Exercise raises circulating lactate levels and myocardial utilization of acetyl-CoA. By contrast, β-adrenergic stimulation raises the circulating levels and metabolism of FFAs in favor of glucose. Severe myocardial ischemia inhibits cytoplasmic pyruvate dehydrogenase, producing incomplete glucose metabolism to lactic acid (anaerobic glycolysis). Anaerobic glycolysis produces much less ATP per mole of glucose than does aerobic glucose metabolism. High concentrations of circulating FFAs, which can occur when adrenergic stimulation is superimposed on severe ischemia, reduce oxidative phosphorylation, and the myocardial content of ATP declines, impairing contraction. In addition, FFA breakdown products may exert toxic or arrhythmogenic effects on cardiac cell membranes.

Myocardial energy is stored as creatine phosphate (CP), which is in equilibrium with ATP, the immediate energy source. In states of reduced energy availability, the CP stores decline first. Cardiac hypertrophy, fibrosis, tachycardia, increased wall tension due to ventricular dilation, and increased intracytoplasmic $[Ca^{2+}]$ all contribute to increased myocardial energy needs. When coupled with reduced coronary flow reserve, as occurs with obstruction of coronary arteries or abnormalities of the coronary microcirculation, an imbalance in myocardial ATP production relative to demand may occur, and the resulting ischemia can worsen or cause heart failure.

CARDIAC CELLULAR INTERVENTIONS: REGENERATION AND REPAIR

Adult mammalian myocardial cells are fully differentiated and have little or no regenerative potential; however, there is evidence that the immature mammalian heart has some limited regenerative potential that rapidly becomes constrained with increasing maturity and workload. Considerable current effort is being devoted to evaluating the utility of various approaches to facilitate the transient release of these constraints on regeneration to enhance cardiac repair after injury. Recent work has also shown that by engineering chimeric antigen receptors on T lymphocytes (CAR-T), it may be possible to reverse fibrosis in the heart. The success of such approaches would offer the exciting possibility of reconstructing an infarcted or failing ventricle.

FURTHER READING

Bautch VL, Caron KM: Blood and lymphatic vessel formation. Cold Spring Harb Perspect Biol 7:a008268, 2015.

Dejana E et al: The molecular basis of endothelial cell plasticity. Nat Commun 8:14361, 2017.

Green DJ et al: Vascular adaptation to exercise in humans: Role of hemodynamic stimuli. Physiol Rev 97:495, 2017.

Libby P et al (eds): *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*, 11th ed. Philadelphia, Elsevier, 2022.

MacLeod KT: Recent advances in understanding cardiac contractility in health and disease. F1000Res 5(F1000 Faculty Rev):1770, 2016.

Page E et al (eds): *Handbook of Physiology: A Critical Comprehensive Presentation of Physiological Knowledge and Concepts. Section 2: The Cardiovascular System, Volume I: The Heart*. New York, Oxford University Press, 2002.

Rurik JG et al: CAR T cells produced in vivo to treat cardiac injury. Science 375:91, 2022.

Spinale FG: Assessment of cardiac function—Basic principles and approaches. Compr Physiol 5:1911, 2015.

Srivastava D: Making or breaking the heart: From lineage determination to morphogenesis. Cell 126:1037, 2006.

Taegtmeyer H et al: Cardiac metabolism in perspective. Comp Physiol 6:1675, 2016.

245

Epidemiology of Cardiovascular Disease

J. Michael Gaziano, Thomas A. Gaziano

Cardiovascular disease (CVD) is now the most common cause of death worldwide. Before 1900, infectious diseases and malnutrition were the most common causes, and CVD was responsible for <10% of all deaths. In 2022, CVD accounted for over 19 million deaths worldwide (33%), with the same rate now occurring in both high-income countries and low- and middle-income countries.

THE EPIDEMIOLOGIC TRANSITION

The global rise in CVD is the result of an unprecedented transformation in the causes of morbidity and mortality during the twentieth century. Known as the epidemiologic transition, this shift is driven by industrialization, urbanization, and associated lifestyle and demographic changes and is taking place in every part of the world among all races, ethnic groups, and cultures. The transition is divided into four basic stages: pestilence and famine, receding pandemics, degenerative and man-made diseases, and delayed degenerative diseases. A fifth stage, characterized by an epidemic of inactivity and obesity, is emerging in many countries that have moderate economic development **(Table 245-1)**.

The *age of pestilence and famine* is marked by malnutrition, infectious diseases, and high infant and child mortality that are offset by high fertility. Tuberculosis, dysentery, cholera, and influenza are often fatal, resulting in a mean life expectancy of about 30 years. CVD, which accounts for <10% of deaths, takes the form of rheumatic heart disease and cardiomyopathies due to infection and malnutrition. Despite advances in treatment of those with rheumatic heart disease, incidence and prevalence rates increased between 1990 and 2019.

Per capita income and life expectancy increase during the *age of receding pandemics* as the emergence of public health systems, cleaner water supplies, and improved nutrition combine to drive down deaths from infectious disease and malnutrition. Infant and childhood mortality also decline, but deaths due to CVD increase to between 10 and 35% of all deaths. Rheumatic valvular disease, hypertension, coronary heart disease (CHD), and stroke are the predominant forms of CVD. Almost 40% of the world's population is currently in this stage.

The *age of degenerative and man-made diseases* is distinguished by mortality from noncommunicable diseases—primarily CVD—surpassing mortality from malnutrition and infectious diseases. Caloric intake, particularly from animal fat, increases. CHD and stroke are prevalent, and between 35 and 65% of all deaths can be traced to CVD. Typically, the rate of CHD deaths exceeds that of stroke by a ratio of 2:1 to 3:1. During this period, average life expectancy surpasses the age of 50. Roughly 35% of the world's population falls into this category.

In the *age of delayed degenerative diseases*, CVD and cancer remain the major causes of morbidity and mortality, with CVD accounting for 40% of all deaths. However, age-adjusted CVD mortality declines, aided by preventive strategies (for example, smoking cessation programs and effective blood pressure control), acute hospital management, and technologic advances, such as the availability of bypass

TABLE 245-1 Five Stages of the Epidemiologic Transition

STAGE	DESCRIPTION	DEATHS RELATED TO CVD, %	PREDOMINANT CVD TYPE
Pestilence and famine	Predominance of malnutrition and infectious diseases as causes of death; high rates of infant and child mortality; low mean life expectancy	<10	Rheumatic heart disease, cardiomyopathies caused by infection and malnutrition
Receding pandemics	Improvements in nutrition and public health lead to decrease in rates of deaths related to malnutrition and infection; precipitous decline in infant and child mortality rates	10–35	Rheumatic valvular disease, hypertension, CHD, and stroke (predominantly hemorrhagic)
Degenerative and man-made diseases	Increased fat and caloric intake and decrease in physical activity lead to emergence of hypertension and atherosclerosis; with increase in life expectancy, mortality from chronic, noncommunicable diseases exceeds mortality from malnutrition and infectious disease	35–65	CHD and stroke (ischemic and hemorrhagic)
Delayed degenerative diseases	CVD and cancer are the major causes of morbidity and mortality; better treatment and prevention efforts help avoid deaths among those with disease and delay primary events; age-adjusted CVD morality declines; CVD affecting older and older individuals	40–50	CHD, stroke, and congestive heart failure
Inactivity and obesity	Overweight and obesity increase at alarming rate; diabetes and hypertension increase; decline in smoking rates levels off; a minority of the population meets physical activity recommendations	<40	CHD, stroke, and congestive heart failure, peripheral vascular disease

Abbreviations: CHD, coronary heart disease; CVD, cardiovascular disease.

Source: Data from AR Omran: The epidemiologic transition: A theory of the epidemiology of population change. Milbank Mem Fund Q 49:509, 1971; and SJ Olshansky, AB Ault: The fourth stage of the epidemiologic transition: The age of delayed degenerative diseases. Milbank Q 64:355, 1986.

surgery. CHD, stroke, and congestive heart failure are the primary forms of CVD. About 15% of the world's population is now in the age of delayed degenerative diseases or is exiting this age and moving into the fifth stage of the epidemiologic transition.

In the industrialized world, physical activity continues to decline while total caloric intake increases. The resulting epidemic of overweight and obesity may signal the start of the *age of inactivity and obesity*. Rates of type 2 diabetes mellitus, hypertension, and lipid abnormalities are on the rise, trends that are particularly evident in children. As these risk factor trends continue, age-adjusted CVD mortality rates that have fallen for decades during the fourth phase have started to increase in recent years during this fifth phase.

PATTERNS IN THE EPIDEMIOLOGIC TRANSITION

Unique regional features have modified aspects of the transition in various parts of the world. High-income countries experienced declines in CVD death rates by as much as 50–60% over the past 60 years, whereas CVD deaths increased by 15% over the past 20 years in the low- and middle-income range and the rate of change has been faster. However, given the large amount of available data, the United States serves as a useful reference point for comparisons. The age of pestilence and famine occurred before 1900, with a largely agrarian economy and population. Infectious diseases accounted for more deaths than any other cause. By the 1930s, the country proceeded through the age of receding pandemics. The establishment of public health infrastructures resulted in dramatic declines in infectious disease mortality rates. Lifestyle changes due to rapid urbanization resulted in a simultaneous increase in CVD mortality rates, reaching ~390 per 100,000. Between 1930 and 1965, the country entered the age of degenerative and man-made diseases. Infectious disease mortality rates fell to fewer than 50 per 100,000 per year, whereas CVD mortality rates reached peak levels with increasing urbanization and lifestyle changes in diet, physical activity, and tobacco consumption. The age of delayed degenerative diseases took place between 1965 and 2000. New therapeutic approaches, preventive measures, and exposure to public health campaigns promoting lifestyle modifications led to substantial declines in age-adjusted mortality rates and a steadily rising age at which a first CVD event occurs.

Currently, the United States is entering what appears to be a fifth phase. The decline in the age-adjusted CVD death rate of 3% per year through the 1970s and 1980s has tapered off in the 1990s to 2% through 2010. There was then limited reduction in the 2010–2019 period, and in 2020 and 2021, there was an increase in CVD mortality rates that had not been seen since the early 1960s. Competing trends appear to be at play. On the one hand, an increase in the prevalence of diabetes and obesity, a slowing in the rate of decline in smoking, and a leveling off in the rate of detection and worsening treatment for hypertension are in the negative column. On the other hand, cholesterol levels continue to decline in the face of increased statin use.

Many high-income countries (HICs)—which together account for 15% of the population—have proceeded through four stages of the epidemiologic transition in roughly the same pattern as the United States. CHD is the dominant form of CVD in these countries, with rates that tend to be two- to fivefold higher than stroke rates. However, variations exist. Whereas North America, Australia, and central northwestern European HICs experienced significant increases then rapid declines in CVD rates, southern and central European countries experienced a more gradual rise and fall in rates. More specifically, central European countries (i.e., Austria, Belgium, and Germany) declined at slower rates compared to their northern counterparts (i.e., Finland, Sweden, Denmark, and Norway). Countries such as Portugal, Spain, and Japan never reached the high mortality rates that the United States and other countries did, with CHD mortality rates at 200 per 100,000, or less. The countries of Western Europe also exhibit a clear north/south gradient in absolute rates of CVD, with rates highest in northern countries (i.e., Finland, Ireland, and Scotland) and lowest in Mediterranean countries (i.e., France, Spain, and Italy). Japan is unique among the HICs, most likely due to the unique dietary patterns of its population. Although stroke rates increased dramatically, CHD rates did not rise as sharply in Japan. However, Japanese dietary habits are undergoing substantial changes, reflected in an increase in cholesterol levels.

Patterns in low- and middle-income countries (LMICs; gross national income per capita $11,666) depend, in part, on cultural differences, secular trends, and responses at the country level, with regard to both public health and treatment infrastructure. Although communicable diseases continue to be a major cause of death, CVD has emerged as a significant health concern in LMICs. With 85% of the world's population, LMICs are driving the rates of change in the global burden of CVD **(Fig. 245-1)**. In most LMICs, an urban/rural gradient has emerged for CHD, stroke, and hypertension, with higher rates in urban centers.

However, although CVD rates are rapidly rising globally, vast differences exist among the regions and countries, and even within the countries themselves **(Fig. 245-2)**. The East Asia and Pacific regions appear to be fully in the third phases of the epidemiologic transition. CVD is a major cause of death in China, but like Japan, stroke causes more deaths than CHD in a ratio of about three to one. Vietnam and Cambodia, on the other hand, are just emerging from the pestilence and famine transition. The Middle East and North Africa regions also appear to be well into the third phase of the epidemiologic transition, with increasing life expectancy and CVD death rates just below those of HICs. In general, Latin America appears to be in the third phase of

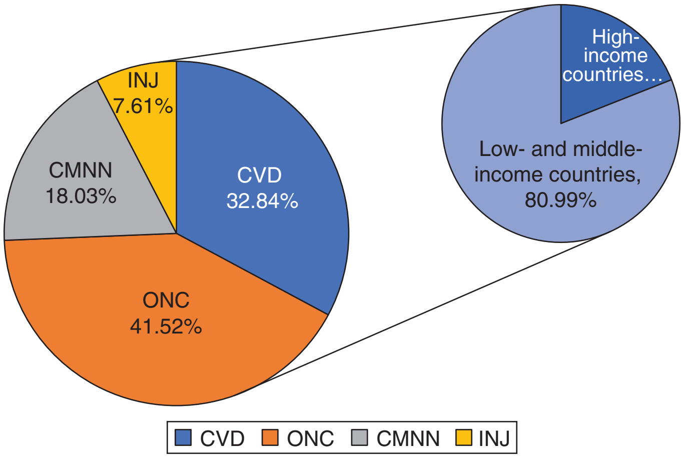

FIGURE 245-1 Global deaths by cause, 2019. CMNN, communicable, maternal, neonatal, and nutritional disorders; CVD, cardiovascular diseases; INJ, injuries; ONC, other noncommunicable diseases. *(Based on data from Global Burden of Disease Study 2017. Global Burden of Disease Study 2017 [GBD 2017] Results. Seattle, United States: Institute for Health Metrics and Evaluation [IHME], 2020.)*

the transition, although there is vast regional heterogeneity with some areas in the second phase of the transition and more in the fourth. The Eastern Europe and Central Asia regions, however, are firmly in the peak of the third phase, with the highest death rates due to CVD (~55%) in the world. Importantly, deaths due to CHD are not limited to the elderly in this region and have a significant effect on working-age populations. South Asia—and more specifically, India, which accounts for the greatest proportion of the region's population—is experiencing an alarming increase in heart disease. The transition appears to be in the Western style, with CHD as the dominant form of CVD. However, rheumatic heart disease continues to be a major cause of morbidity and mortality. As in South Asia, rheumatic heart disease is also an important cause of CVD morbidity and mortality in sub-Saharan Africa, which largely remains in the first phase of the epidemiologic transition.

Many factors contribute to this heterogeneity among LMICs. First, the regions are in various stages of the epidemiologic transition. Second, vast differences in lifestyle and behavioral risk factors exist. Third, racial and ethnic differences may lead to altered susceptibilities to various forms of CVD. In addition, it should be noted that for most countries in these regions, accurate country-wide data on cause-specific mortality are not complete.

GLOBAL TRENDS IN CARDIOVASCULAR DISEASE

Over the past 5 years, there have been changes in the trends of CVD that are reflective of both trends in demographics and management of disease, but also of the way deaths and diseases have been measured and estimated. In 2017, the Global Burden of Disease (GBD) Study updated its estimates with several important changes based on newly available data, refinement in the causes of death, and the introduction of new modeling techniques. The major changes include the addition of an independent estimation of population and fertility, the addition of over 127 country-years of vital registration and verbal autopsy data, revisions of some deaths from "misclassified" to dementia, Parkinson's disease and atrial fibrillation, and the addition of new diseases such as nonrheumatic calcific aortic and degenerative mitral valve disease. CVD accounts for 32% of deaths worldwide, a number expected to increase. In 2019, CHD accounted for 16.2% of all deaths globally and the largest portion (10%) of global years of life lost (YLLs) and disability-adjusted life-years (DALYs) (7.2%). Stroke moved from the third to the second largest cause of death (11.6% of all deaths) and remained the third largest contributor to global YLLs (7.5%) and DALYs (5.7%). Together, CHD and stroke accounted for more than a quarter of all deaths worldwide. The burden of stroke is of growing concern among LMICs. The impact of stroke on DALYs and mortality rates is more than three times greater in LMICs as compared to HICs.

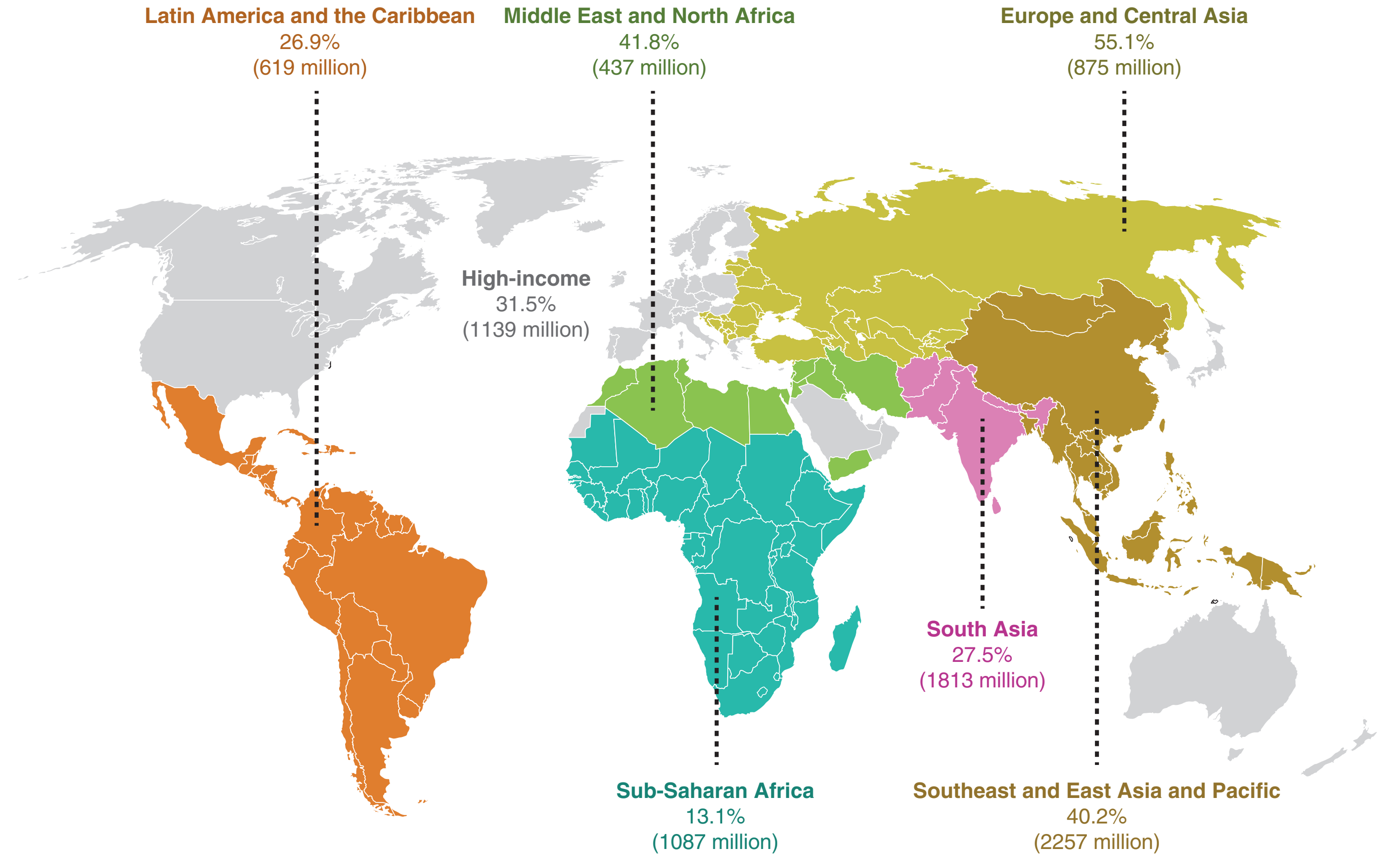

FIGURE 245-2 Cardiovascular disease deaths as a percentage of total deaths and total population in seven economic regions of the world defined by the World Bank. *(Based on data from Global Burden of Disease Study 2017. Global Burden of Disease Study 2019 [GBD 2019] Results. Seattle, United States: Institute for Health Metrics and Evaluation [IHME], 2023.)*

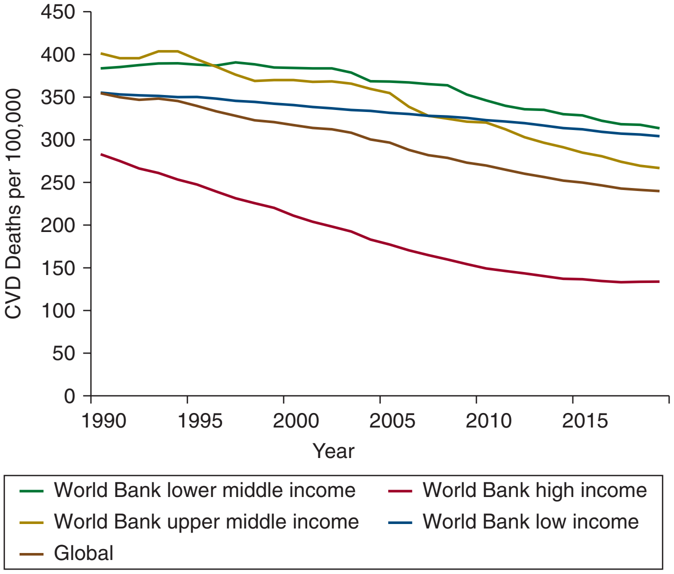

FIGURE 245-3 Age-standardized cardiovascular diseases (CVD) death rate per 100,000 from 1990 to 2019, by World Bank income. *(Based on data from Global Burden of Disease Study 2019. Global Burden of Disease Study 2019 [GBD 2019] Results. Seattle, United States: Institute for Health Metrics and Evaluation [IHME], 2023.)*

With 85% of the world's population, LMICs largely drive global CVD rates and trends. More than 15 million CVD deaths occurred in LMICs in 2019, compared to 3.5 million in HICs. Globally, there is evidence of significant delays in age of occurrence and/or improvements in case fatality rates; between 1990 and 2019, the number of CVD deaths increased by 54%, but age-adjusted death rates decreased by 32.4% in the same period. Age-standardized death rates, however, have declined faster in HICs than in middle-income and lower-income regions **(Fig. 245-3)**. Population growth has been greater in LMICs compared to HICs. As a result of slower rates of population growth in HICs, overall CVD deaths remained steady. However, in the LMICs, the population aging and growth outstripped gains in age-adjusted mortality reductions such that overall CVD deaths continued to climb over the past 25 years **(Fig. 245-4)**.

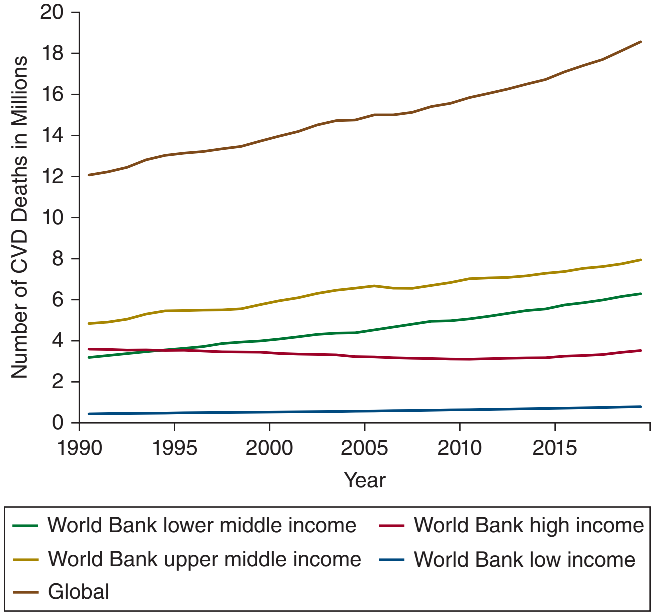

FIGURE 245-4 Number of cardiovascular diseases (CVD) deaths from 1990 to 2019, by World Bank income. *(Based on data from Global Burden of Disease Study 2019. Global Burden of Disease Study 2017 [GBD 2019] Results. Seattle, United States: Institute for Health Metrics and Evaluation [IHME], 2023.)*

Although HIC population growth will be fueled by immigration from LMICs, the populations of HICs will shrink as a proportion of the world's population. The modest decline in CVD death rates that began in the HICs in the latter third of the twentieth century will continue, but the rate of decline appears to be slowing. However, these countries are expected to see an increase in the prevalence of CVD, as well as the absolute number of deaths as the population ages.

Significant portions of the population living in LMICs have entered the third phase of the epidemiologic transition, and some are entering the fourth stage. Changing demographics play a significant role in future predictions for CVD throughout the world. For example, the population in Eastern Europe declined by 3% between 2010 and 2023, whereas it grew by 30% in South Asia. CVD rates will also have an economic impact. Even assuming no increase in CVD risk factors, most countries, but especially India and South Africa, will see a large number of people between 35 and 64 die of CVD over the next 30 years, as well as an increasing level of morbidity among middle-aged people related to heart disease and stroke.

RISK FACTORS

Global variation in CVD rates is related to temporal and regional variations in known risk factors and behaviors. Ecologic analyses of major CVD risk factors and mortality demonstrate high correlations between expected and observed mortality rates for the three main risk factors—smoking, serum cholesterol, and hypertension—and suggest that many regional variations are based on differences in conventional risk factors.

Behavioral Risk Factors • **TOBACCO** Over 1.3 billion people use tobacco worldwide. Tobacco use currently causes about 8.7 million deaths annually (15.4% of all deaths), 3.2 million of which are CVD related. The population of the HIC group smokes (21.6%) at almost double the rate of the low-income countries (11.2%), whereas the middle-income country group's smoking rate (19.5%) approximates the global average (22.3%). From 2007 to 2017, smoking rates decreased across low-, middle-, and high-income country groups, with relative reductions of 19%, 12%, and 20%, respectively. By 2030, the global average smoking rate is expected to decline from 19% to 16% (women, 4%; men, 28%); however, the number of tobacco users is expected to rise owing to population growth. Secondhand smoke is another well-established cause of CVD, responsible for 598,000 CVD deaths of nonsmokers in 2019. Although smoking bans have both immediate and long-term benefits, implementation varies greatly between countries.

DIET Total caloric intake per capita increases as countries develop. With regard to CVD, a key element of dietary change is an increase in intake of saturated animal fats and hydrogenated vegetable fats, which contain atherogenic *trans* fatty acids, along with a decrease in intake of plant-based foods and an increase in simple carbohydrates. Fat contributes <20% of calories in rural China and India, <30% in Japan, and over 35% in the United States. Caloric contributions from fat appear to be increasing in the United States but falling in the other HICs.

PHYSICAL INACTIVITY The increased mechanization that accompanies the economic transition leads to a shift from physically demanding, agriculture-based work to largely sedentary industry- and office-based work. Physical inactivity is responsible for 1.3 million global deaths in 2019. The global prevalence of physical inactivity has remained steady between 2001 and 2016 (28.5% to 27.5%). Mortality rates attributable to inactivity are highest in North Africa and the Middle East and in Central and Eastern Europe and lowest in sub-Saharan Africa.

METABOLIC RISK FACTORS

Examination of trends in metabolic risk factors provides insight into changes in the CVD burden globally. Here we describe four metabolic risk factors—lipid levels, hypertension, obesity, and diabetes mellitus—using data from the Global Burden of Disease, Injuries, and Risk Factors Study (GBD 2019). The GBD project identified and compiled mortality and morbidity data from 195 countries from 1980 to 2017.

Lipid Levels Worldwide, high low-density lipoprotein cholesterol levels are estimated to play a role in 41% of ischemic heart disease deaths and 9% of stroke deaths, amounting to 4 million deaths in 2019. Although mean population plasma cholesterol levels tend to rise as countries move through the epidemiologic transition, mean serum total cholesterol levels have decreased globally between 1980 and 2008 by 0.08 mmol/L per decade in men and 0.07 mmol/L per decade in women. Large declines occurred in Australasia, North America, and Western Europe (0.19–0.21 mmol/L). Countries in the East Asia and Pacific region experienced increases of >0.08 mmol/L in both men and women. More recent research including Mendelian studies suggests that lipoprotein(a) may act as an individual predictor of CVD risk beyond traditional total or low-density lipoprotein cholesterol through increased cellular lipid accumulation, endothelial dysfunction, and impacts on coagulation. It appears to be elevated in ~20% of the global population, although fewer data are available from LMICs. Nonrandomized data suggest higher rates among those of African descent with twice the levels of Caucasians, with East Asians and South Asians having intermediate levels. There are limited clinical trial data on clinical agents that target lipoprotein(a), although PCSK9 inhibitors lower it or other specific targets, so this remains an area of intense research.

Hypertension Elevated blood pressure is an early indicator of the epidemiologic transition. Observational studies show increased risk of CVD beginning with systolic blood pressures (SBPs) >110–115 mmHg. Between 1990 and 2015, the global prevalence of SBP ≥110–115 mmHg increased from 73,119 to 81,373 per 100,000, whereas the prevalence of SBP ≥140 mmHg rose from 17,307 to 20,526 per 100,000. In 2015, of the estimated 3.47 billion adults with SBP ≥110–115 mmHg, 874 million (25%) had SBP ≥140 mmHg. While SBP ≥140 mmHg accounts for only 25% of those with elevated blood pressure, it accounted for 73% (7.8 million) of deaths due to SBP of ≥110–115 mmHg in 2015. Worldwide, 53% of stroke deaths (3.44 million) and 53% of CHD deaths (4.86 million) are attributable to high blood pressure, accounting for 10.8 million deaths in 2019. From 1990 to 2015, the number of deaths related to SBP ≥140 mmHg increased in all LMIC groups but fell in HICs. Between 1980 and 2008, the age-standardized prevalence of uncontrolled hypertension decreased even as the number of people with uncontrolled hypertension increased due to population growth and aging. However, concerning trends of declining hypertension control rates were seen in the United States since 2016 with many middle-income countries now outperforming high-income nations. Rising mean population blood pressure also occurs as populations industrialize and move from rural to urban settings. For example, the prevalence of hypertension in urban India is 33.8%, but varies between 14.5 and 31.7% in rural regions. One major concern in LMICs is the high rate of undetected, and therefore untreated, hypertension. This may explain, at least in part, the higher stroke rates in these countries in relation to CHD rates during the early stages of the transition. The high rates of hypertension throughout Asia, especially undiagnosed hypertension, likely contribute to the high prevalence of hemorrhagic stroke in the region.

Obesity In 2025, an estimated 892 million adults will be obese. Global obesity prevalence is projected to be 16.1% among adults and is increasing throughout the world, particularly in developing countries where the trajectories are steeper than those experienced by the developed countries. High body mass index (BMI) contributed to 5.0 million deaths worldwide (7.1% deaths from any cause); CVD was the leading cause of these deaths (1.95 million) and also of associated DALYs (160 million) followed by diabetes (0.6 million deaths, 30.4 million DALYs). Women are more affected by obesity than men; from 1975 to 2014, global mean age-standardized BMI increased from 22.1 to 24.4 kg/m^2 in females and from 21.7 to 24.2 kg/m^2 in males, whereas the prevalence of obesity increased from 6.4% to 14.9% in females and 3.2% to 10.8% in males. The proportion of the world's adult women who are either overweight or obese rose from 29.8% to 38.0% between 1980 and 2013, while an increase from 28.8% to 36.9% was observed for men. Country and regional differences are observed. The highest prevalence of male obesity is in the United States, Southern and Central Latin America, Australasia, and Central and Western Europe. For females, the highest prevalence of obesity is in Southern and North Africa, the Middle East, Central and Southern Latin America, and the United States. The lowest prevalence for both males and females was observed in South and Southeast Asia and in East, Central, and West Africa.

Diabetes Mellitus As a consequence of, or in addition to, increasing BMI and decreasing levels of physical activity, worldwide rates of diabetes—predominantly type 2 diabetes—are on the rise. According to the most recent data from the GBD project, the prevalence of diabetes increased 129.7% for males and 120.9% for females between 1990 and 2017. An estimated 476 million people worldwide have diabetes, and the International Diabetes Foundation predicts this number will reach 693 million by 2045. Nearly 50% of people with diabetes are undiagnosed, and 80% live in LMICs. The Middle East and North Africa have the highest regional age-standardized prevalence (8.7% of the population) and incidence rates (400 per 100,000) of diabetes, whereas East Asia and the Pacific have the lowest (5.8%; 249 per 100,000). Future growth will also largely occur in the Middle East and Africa, along with other LMICs in South Asia and sub-Saharan Africa.

COVID-19 The impact of COVID-19 on CVD was significant. First, patients with acute conditions were hesitant to present to hospitals for fear of infection, resulting in delayed presentation for those with acute coronary syndrome (ACS) or stroke and missing critical early care. Additionally, even those who presented with ACS had worse outcomes sometimes related to testing delays. Furthermore, the presence of obesity and/or diabetes mellitus worsened outcomes. Finally, control rates of hypertension and dyslipidemia may have been exacerbated due to difficulty of access to providers.

GENETIC RISK FACTORS

A great deal of effort has recently been invested in understanding how genes affect cardiovascular health in populations. These efforts have focused on germline genetic variants that are related to specific CVDs as well as those that are associated with cardiovascular risk factors. In both cases, every year, the number of associated variants has increased meaningfully to the point that it appears that hundreds or even thousands of variants are associated with these conditions, each explaining a small amount of the population variability in disease and risk factors. Collections of variants have been combined in polygenomic risk scores, but these too explain only a small amount of the variability of the disease in the population. Much more data will emerge in the coming years about these associations, the mechanisms that explain these associations, the relationships of variants that are specific to certain tissues such as the heart or the brain, and the interactions between genetic and lifestyle factors in causing disease. Currently, most of the data are among those with European ancestry; however, large-scale efforts are underway to understand the relationships between genes and diseases and their risk factors around the world. The early data suggest nontrivial differences among various world populations. Beyond germline risk, there appears to be increased cardiovascular risk associated with age-related expansion of hematopoietic clones with somatic mutations, including loss-of-function alleles of certain genes. Individuals with these mutations without other hematologic abnormalities are defined as having clonal hematopoiesis of indeterminate potential (CHIP). Recent studies suggest those with CHIP have up to a twofold increased risk of developing CHD.

SUMMARY

Although CVD rates are declining in the HICs, they are increasing in many other regions of the world. The consequences of this preventable epidemic will be substantial on many levels, including individual mortality and morbidity, family suffering, and staggering economic costs.

Three complementary strategies can be used to lessen the impact. First, the overall burden of CVD risk factors can be lowered through population-wide public health measures, such as national campaigns against cigarette smoking, unhealthy diets, and physical inactivity. Second, it is important to identify higher risk subgroups of the population who stand to benefit the most from specific, low-cost prevention interventions, including screening for and treatment of hypertension and elevated cholesterol. Simple, low-cost interventions, such as the "polypill"—a regimen of aspirin, a statin, and an antihypertensive agent—now show trial data with reductions in events for both primary and secondary prevention and have been added to the WHO Essential Medicines list. Third, resources should be allocated to acute, as well as secondary, prevention interventions. For countries with limited resources, a critical first step in developing a comprehensive plan is better assessment of cause-specific mortality and morbidity, as well as the prevalence, of the major preventable risk factors.

In the meantime, the HICs must continue to bear the burden of research and development aimed at prevention and treatment, being mindful of the economic limitations of many countries. The concept of the epidemiologic transition provides insight into how to alter the course of the CVD epidemic. The efficient transfer of low-cost preventive and therapeutic strategies could alter the natural course of this epidemic and thereby reduce the excess global burden of preventable CVD.

■ FURTHER READING

Boutari C et al: A 2022 update on the epidemiology of obesity and a call to action: As its twin COVID-19 pandemic appears to be receding, the obesity and dysmetabolism pandemic continues to rage on. Metabolism 133:155217, 2022.

Clerkin KJ et al: COVID-19 and cardiovascular disease. Circulation 141:1648, 2020.

Gaziano T, Gaziano JM: Global burden of cardiovascular disease, in *Heart Disease: A Textbook of Cardiovascular Medicine*, 12th ed, Braunwald E (ed). Philadelphia, Elsevier, 2022.

Mensah G et al: Global burden of cardiovascular diseases and risks, 1990-2022. J Am Coll Cardiol 82:2350, 2023.

Tsao CW et al: Heart disease and stroke statistics–2022 update: A report from the American Heart Association. Circulation 145:e153, 2022.

Section 2 Diagnosis of Cardiovascular Disorders

246 Physical Examination of the Cardiovascular System

Patrick T. O'Gara, Joseph Loscalzo

The approach to a patient with known or suspected cardiovascular disease begins with the time-honored traditions of a directed history and a targeted physical examination. The scope of these activities depends on the clinical context, ranging from an elective ambulatory follow-up visit to a more urgent bedside encounter. There has been a gradual decline in physical examination skills over the past few decades at every level, from student to faculty specialist, a development of great concern to both clinicians and medical educators. Classic cardiac findings are recognized by only a minority of internal medicine and family practice residents. Despite popular perceptions, clinical performance does not improve predictably with experience; instead, the acquisition of new examination skills may become more difficult for a busy individual practitioner. Less time is now devoted to mentored cardiovascular examinations during the training of students and residents. One widely recognized outcome of these trends is the progressive utilization of noninvasive imaging studies to establish the presence and severity of cardiovascular disease even when the examination findings imply a low pretest probability of significant pathology. Proponents of the use of hand-held ultrasound devices to identify and characterize structural cardiac disease have called for its incorporation into educational curricula. Techniques to improve competency in bedside examination skills include repetition, patient-centered teaching conferences, visual display feedback of auscultatory events using Doppler echocardiographic imaging, and simulation-based training. The use of digital stethoscopes may enhance learning and is foundational to the application of computer- or artificial intelligence–assisted evaluation of auscultatory events.

The findings from the history and physical examination can help establish the presence, severity, and prognosis of several cardiovascular diseases. For example, observations regarding heart rate and blood pressure, signs of pulmonary congestion, and the presence of mitral regurgitation (MR) contribute importantly to bedside risk assessment in patients with acute coronary syndromes and can inform clinical decision-making before the results of cardiac biomarker testing are known. The prognosis of patients with heart failure can be predicted on the basis of the jugular venous pressure (JVP) and the presence or absence of a third heart sound (S_3). Accurate characterization of cardiac murmurs provides important insight into the natural history of many valvular and congenital heart lesions. Finally, the important role played by the physical examination in enhancing the clinician-patient relationship cannot be overstated.

■ THE GENERAL PHYSICAL EXAMINATION

The examination begins with an assessment of the general appearance of the patient, with notation of age, posture, demeanor, and overall health status. Is the patient in pain or resting quietly, dyspneic or diaphoretic? Does the patient choose to avoid certain body positions to reduce or eliminate pain, as might be the case with suspected acute pericarditis? Are there clues indicating that dyspnea may have a pulmonary cause, such as a barrel chest deformity with an increased anterior-posterior diameter, tachypnea, and pursed-lip breathing? Skin pallor, cyanosis, and jaundice can be appreciated readily and provide additional clues. The appearance of a chronically ill-appearing emaciated patient may suggest the presence of long-standing heart failure or another systemic disorder, such as a malignancy. Various genetic syndromes, often with cardiovascular involvement, can also be recognized easily, such as trisomy 21, Marfan syndrome, and Holt-Oram syndrome. Height and weight should be measured routinely, and both body mass index and body surface area should be calculated. Knowledge of the waist circumference and the waist-to-hip ratio can be used to predict long-term cardiovascular risk. Mental status, level of alertness, and mood should be assessed continuously during the interview and examination.

Skin Central cyanosis occurs with significant right-to-left shunting at the level of the heart or lungs, allowing deoxygenated blood to reach the systemic circulation. Peripheral cyanosis or acrocyanosis, in contrast, is usually related to reduced extremity blood flow due to small vessel constriction, as seen in patients with severe heart failure, shock, or peripheral vascular disease; it can be aggravated by the use of β-adrenergic blockers with unopposed α-mediated vasoconstriction. Differential cyanosis refers to isolated cyanosis affecting the lower but not the upper extremities in a patient with a large patent ductus arteriosus (PDA) and secondary pulmonary hypertension with right-to-left to shunting at the great vessel level. Telangiectasias on the lips, tongue, and mucous membranes, as part of the Osler-Weber-Rendu syndrome (hereditary hemorrhagic telangiectasia), resemble spider nevi and can be a source of right-to-left shunting when also present in the lung. Malar telangiectasias also are seen in patients with advanced mitral stenosis (MS) or scleroderma. An unusually tan or bronze discoloration of the skin may suggest hemochromatosis as the cause of the associated systolic heart failure. Jaundice, which may be visible first in the sclerae, has a broad differential diagnosis but, in the appropriate setting, can be consistent with advanced right heart failure and congestive hepatomegaly. Various hereditary lipid disorders sometimes are associated with subcutaneous xanthomas, particularly along the tendon sheaths or over the extensor surfaces of the extremities. Severe hypertriglyceridemia can be associated with eruptive xanthomatosis and lipemia retinalis. Palmar crease xanthomas are specific for type III hyperlipoproteinemia. Pseudoxanthoma elasticum, a disease associated with premature atherosclerosis, is manifested by a leathery, cobblestoned appearance of the skin in the axilla and neck creases and by angioid streaks on funduscopic examination. Extensive lentigenes have been described in a variety of development delay–cardiovascular syndromes, including Carney's syndrome, which includes multiple atrial myxomas. Cutaneous manifestations of sarcoidosis such as lupus pernio and erythema nodosum may suggest this disease as a cause of an associated dilated cardiomyopathy, especially with heart block, intraventricular conduction delay, or ventricular tachycardia.

Head and Neck Dentition and oral hygiene should be assessed in every patient both as a source of potential infection and as an index of general health. A high-arched palate is a feature of Marfan syndrome and other connective tissue disease syndromes. Bifid uvula has been described in patients with Loeys-Dietz syndrome, and orange tonsils are characteristic of Tangier disease. The ocular manifestations of hyperthyroidism have been well described. Many patients with congenital heart disease have associated hypertelorism, low-set ears, or micrognathia. Blue sclerae are a feature of osteogenesis imperfecta. An arcus senilis pattern lacks specificity as an index of coronary heart disease risk. The funduscopic examination is an often-underused method by which to assess the microvasculature, especially among patients with established atherosclerosis, hypertension, or diabetes mellitus. A mydriatic agent may be necessary for optimal visualization. A funduscopic examination should be performed routinely in the assessment of patients with suspected endocarditis and those with a history of acute visual change. Branch retinal artery occlusion or visualization of a Hollenhorst plaque can narrow the differential diagnosis rapidly in the appropriate setting. Relapsing polychondritis may manifest as an inflamed pinna or, in its later stages, as a saddle-nose deformity because of destruction of nasal cartilage; granulomatosis with polyangiitis can also lead to a saddle-nose deformity.

Chest Midline sternotomy, right anterior thoracotomy, left posterolateral thoracotomy, or infraclavicular scars should not be overlooked and may provide the first clue regarding an underlying cardiovascular disorder in patients unable to provide a relevant history. A prominent venous collateral pattern may suggest subclavian or vena caval obstruction. If the head and neck appear dusky and slightly cyanotic and the venous pressure is grossly elevated without visible pulsations, a diagnosis of superior vena cava syndrome should be entertained. Thoracic cage abnormalities have been well described among patients with connective tissue disease syndromes. They include pectus carinatum ("pigeon chest") and pectus excavatum ("funnel chest"). Obstructive lung disease is suggested by a barrel chest deformity, especially with tachypnea, pursed-lip breathing, and use of accessory muscles. The characteristically severe kyphosis and compensatory lumbar, pelvic, and knee flexion of ankylosing spondylitis should prompt careful auscultation for a murmur of aortic regurgitation (AR). Straight back syndrome refers to the loss of the normal kyphosis of the thoracic spine and has been described in patients with mitral valve prolapse (MVP) and its variants. In some patients with cyanotic congenital heart disease, the chest wall appears to be asymmetric, with anterior displacement of the left hemithorax. The respiratory rate and pattern should be noted during spontaneous breathing and sleep, with additional attention to depth, audible wheezing, and stridor. Lung examination can reveal adventitious sounds indicative of pulmonary edema, pneumonia, or pleuritis.

Abdomen In some patients with advanced obstructive lung disease, the point of maximal cardiac impulse may be in the epigastrium. The liver is frequently enlarged and tender in patients with chronic heart failure. Systolic pulsations over the liver signify severe tricuspid regurgitation (TR). Splenomegaly may be a feature of infective endocarditis, particularly when symptoms have persisted for weeks or months. Ascites is a nonspecific finding but may be present with advanced chronic right heart failure, constrictive pericarditis, hepatic cirrhosis, or an intraperitoneal malignancy. The finding of an elevated JVP implies a cardiovascular etiology. In nonobese patients, the aorta typically is palpated between the epigastrium and the umbilicus. The sensitivity of palpation for the detection of an abdominal aortic aneurysm (pulsatile and expansile mass) decreases as a function of body size. Because palpation alone is not sufficiently accurate to establish this diagnosis, a screening ultrasound examination is advised when appropriate. The presence of an arterial bruit over the abdomen suggests high-grade atherosclerotic disease, although precise localization is difficult.

Extremities The temperature and color of the extremities, the presence of clubbing, arachnodactyly, and pertinent nail findings can be surmised quickly during the examination. Clubbing implies the presence of central right-to-left shunting, although it has also been described in patients with endocarditis. Its appearance can range from cyanosis and softening of the root of the nail bed, to the classic loss of the normal angle between the base of the nail and the skin, to the skeletal and periosteal bony changes of hypertrophic osteoarthropathy, which is seen rarely in patients with advanced lung or liver disease. Patients with the Holt-Oram syndrome have an unopposable, "fingerized" thumb, whereas patients with Marfan syndrome may have arachnodactyly and a positive "wrist" (overlapping of the thumb and fifth finger around the wrist) or "thumb" (protrusion of the thumb beyond the ulnar aspect of the hand when the fingers are clenched over the thumb in a fist) sign. The Janeway lesions of endocarditis are nontender, slightly raised hemorrhagic lesions on the palms and soles, whereas Osler's nodes are tender, raised nodules on the pads of the fingers or toes. Splinter hemorrhages are classically identified as linear petechiae in the midposition of the nail bed and should be distinguished from the more common traumatic petechiae, which are seen closer to the distal edge.

Lower extremity or presacral edema in the setting of an elevated JVP defines volume overload and may be a feature of chronic heart failure or constrictive pericarditis. Lower extremity edema in the absence of jugular venous hypertension may be due to profound hypoalbuminemia as seen in nephrotic syndrome or liver failure. Other causes include lymphatic or venous obstruction or, more commonly, venous insufficiency, as would be further suggested by the appearance of varicosities, venous ulcers (typically medial in location), and brownish cutaneous discoloration from hemosiderin deposition (eburnation). Pitting edema can also be seen in patients who use dihydropyridine calcium channel blockers. A Homans' sign (posterior calf pain on active dorsiflexion of the foot against resistance) is neither specific nor sensitive for deep venous thrombosis. Muscular atrophy or the absence of hair along an extremity is consistent with severe arterial insufficiency or a primary neuromuscular disorder.

CARDIOVASCULAR EXAMINATION

Jugular Venous Pressure and Waveform The JVP is the single most important bedside measurement from which to estimate the volume status. The internal jugular vein is preferred because the external jugular vein is valved and not directly in line with the superior vena cava and right atrium. Nevertheless, the external jugular vein has been used to discriminate between high and low central venous pressure (CVP). Precise estimation of the central venous or right atrial pressure from bedside assessment of the jugular venous waveform can be difficult. Venous pressure traditionally has been measured as the vertical distance between the top of the jugular venous pulsation and the sternal inflection point (angle of Louis). A distance >4.5 cm at 30° elevation is considered abnormal. However, the actual distance between the mid-right atrium and the angle of Louis varies considerably as a function of both body size and the patient angle at which the assessment is made (30°, 45°, or 60°). The use of the sternal angle as a reference point leads to systematic underestimation of CVP, and this method should be used less for semiquantification than to distinguish a normal from an elevated CVP. The use of the clavicle may provide an easier reference

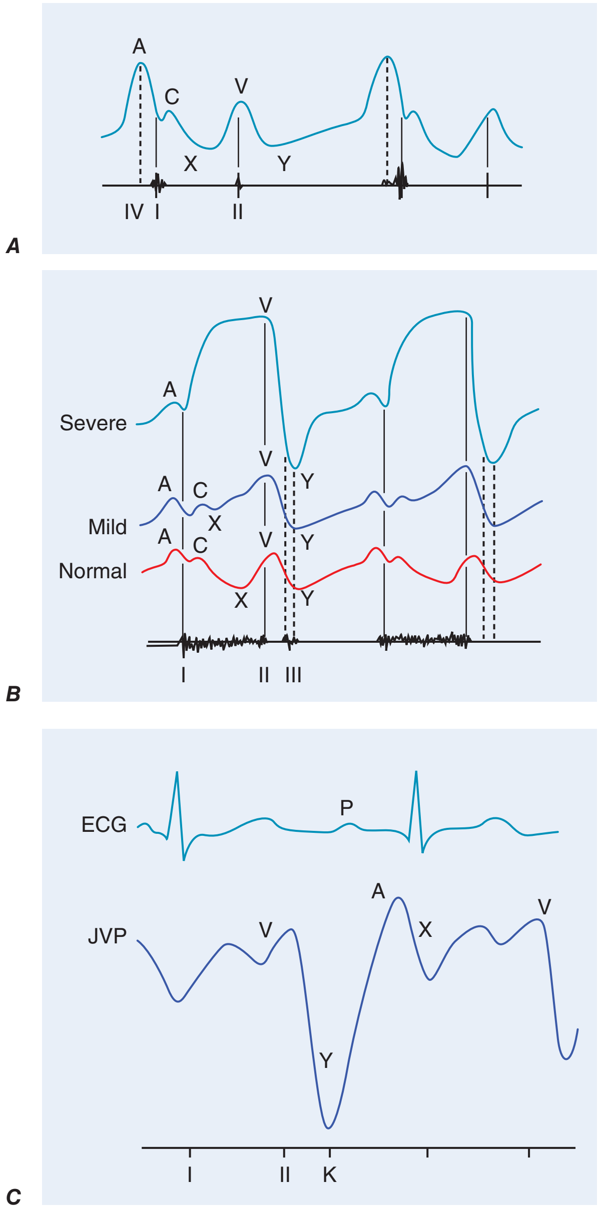

FIGURE 246-1 ***A.*** Jugular venous pulse wave tracing (*top*) with heart sounds (*bottom*) in a patient with reduced right ventricular compliance. The A wave represents right atrial presystolic contraction and occurs just after the electrocardiographic P wave and just before the first heart sound (I). In this example, the A wave is accentuated and larger than normal due to decreased right ventricular compliance, as also suggested by the right-sided S$_4$ (IV). The C wave may reflect the carotid pulsation in the neck and/or an early systolic increase in right atrial pressure as the right ventricle pushes the closed tricuspid valve into the right atrium. The *x* descent follows the A wave just as atrial pressure continues to fall. The V wave represents atrial filling during ventricular systole and peaks at the second heart sound (II). The *y* descent corresponds to the fall in right atrial pressure after tricuspid valve opening. ***B.*** Jugular venous wave forms in mild (*middle*) and severe (*top*) tricuspid regurgitation, compared with normal (*bottom*), with phonocardiographic representation of the corresponding heart sounds below. With increasing degrees of tricuspid regurgitation, the waveform becomes "ventricularized." ***C.*** Electrocardiogram (ECG) (*top*), jugular venous waveform (JVP) (*middle*), and heart sounds (*bottom*) in pericardial constriction. Note the prominent and rapid *y* descent, corresponding in timing to the pericardial knock (K). *(Reproduced with permission from J Abrams: Synopsis of Cardiac Physical Diagnosis, 2nd ed. Boston, Butterworth Heinemann, 2001.)*

for standardization. Venous pulsations above this level in the sitting position are clearly abnormal, as the distance between the clavicle and the right atrium is at least 10 cm. The patient should always be placed in the sitting position, with the legs dangling below the bedside, when an elevated pressure is suspected in the semisupine position. It should also be noted that bedside estimates of CVP are made in centimeters of water, and must be converted to millimeters of mercury to provide correlation with accepted hemodynamic norms (1.36 cmH$_2$O = 1.0 mmHg).

The venous waveform sometimes can be difficult to distinguish from the carotid pulse, especially during casual inspection. Nevertheless, the venous waveform has several characteristic features, and its individual components can be appreciated in most patients **(Fig. 246-1)**. The arterial pulsation is not easily obliterated with palpation; the venous waveform in patients with sinus rhythm is usually biphasic, while the carotid pulse is monophasic; and the jugular venous pulsation should change with changes in posture or inspiration (unless the venous pressure is quite elevated).

The venous waveform is divided into several distinct peaks. The *a* wave reflects right atrial presystolic contraction and occurs just after the electrocardiographic P wave, preceding the first heart sound (S$_1$). A prominent *a* wave is seen in patients with reduced right ventricular compliance; a cannon *a* wave occurs with atrioventricular (AV) dissociation and right atrial contraction against a closed tricuspid valve. In a patient with a wide complex tachycardia, the appreciation of cannon *a* waves in the jugular venous waveform identifies the rhythm as ventricular in origin. The *a* wave is not present with atrial fibrillation. The *x* descent defines the fall in right atrial pressure after tricuspid valve opening, and occurs after inscription of the *a* wave. The *c* wave, which occurs as the closed tricuspid valve is pushed into the right atrium during early ventricular systole, interrupts this *x* descent and is followed by a further descent. The *v* wave represents atrial filling (atrial diastole) and occurs during ventricular systole. The height of the *v* wave is determined by right atrial compliance as well as the volume of blood returning to the right atrium either antegrade from the cavae or retrograde through an incompetent tricuspid valve. In patients with TR, the *v* wave is accentuated and the subsequent fall in pressure (*y* descent) is rapid. With progressive degrees of TR, the *v* wave merges with the *c* wave, and the right atrial and jugular vein waveforms become "ventricularized." The *y* descent, which follows the peak of the *v* wave, can become prolonged or blunted with obstruction to right ventricular inflow, as may occur with tricuspid stenosis or pericardial tamponade. Normally, the venous pressure should fall by at least 3 mmHg with inspiration. Kussmaul's sign is defined by either a rise or a lack of fall of the JVP with inspiration and is classically associated with constrictive pericarditis, although it has been reported in patients with restrictive cardiomyopathy, massive pulmonary embolism, right ventricular infarction, and advanced left ventricular (LV) systolic heart failure. It is also a common, isolated finding in patients after cardiac surgery without other hemodynamic abnormalities.

Venous hypertension sometimes can be elicited by passive leg elevation or performance of the abdominojugular reflux maneuver. When these signs are positive, a volume-overloaded state with limited compliance of an overly distended or constricted venous system is present. Abdominojugular reflux is produced with firm and consistent pressure over the upper portion of the abdomen, preferably over the right upper quadrant, for >15 s. A positive response is defined by a sustained rise of >3 cm in the JVP during the application of firm abdominal pressure. The response should be assessed after 10 s of continuous pressure to allow for respiratory artifacts and tensing of the abdominal muscles to subside. Patients must be coached to refrain from breath holding or a Valsalva-like maneuver during the procedure. Performance of the abdominojugular reflux maneuver is useful in predicting a pulmonary artery wedge pressure >15 mmHg in patients with heart failure.

Although the JVP estimates right ventricular filling pressure, it has a predictable relationship with the pulmonary artery wedge pressure. In a large study of patients with advanced heart failure, the presence of a right atrial pressure >10 mmHg (as predicted on bedside examination) had a positive value of 88% for the prediction of a pulmonary artery wedge pressure of >22 mmHg. In addition, an elevated JVP has prognostic significance in patients with both symptomatic heart failure and asymptomatic LV systolic dysfunction. The presence of an elevated JVP is associated with a higher risk of subsequent hospitalization for heart failure, death from heart failure, or both.

Assessment of Blood Pressure Measurement of blood pressure usually is delegated to a medical assistant but should be repeated by the examining clinician. Accurate measurement depends on body position, arm size, time of measurement, place of measurement, type of device, device size, technique, and examiner. In general, physician-recorded blood pressures are higher than both nurse-recorded pressures and self-recorded pressures at home. Blood pressure is best measured in the seated position with the arm at the level of the heart and the feet on the floor with the back supported, using an appropriately sized cuff, after 5–10 min of relaxation. When it is measured in the supine position, the arm should be raised to bring it to the level of the mid-right atrium. The length and width of the blood pressure cuff bladder should be 80 and 40% of the arm's circumference, respectively. A common source of error in practice is to use an inappropriately small cuff, resulting in marked overestimation of true blood pressure, or an inappropriately large cuff, resulting in underestimation of true blood pressure. The cuff should be inflated to 30 mmHg above the expected systolic pressure and the pressure released at a rate of 2–3 mmHg/s. Systolic and diastolic pressures are defined by the first and fifth Korotkoff sounds, respectively. Very low (even 0 mmHg) diastolic blood pressures may be recorded in patients with chronic, severe AR or a large arteriovenous fistula because of enhanced diastolic "run-off." In these instances, both the phase IV and phase V Korotkoff sounds should be recorded. Blood pressure is best assessed at the brachial artery level, though it can be measured at the radial, popliteal, or pedal artery level. In general, systolic pressure increases and diastolic pressure decreases when measured in more distal arteries. Blood pressure should be measured in both arms, and the difference should be <10 mmHg. A blood pressure differential that exceeds this threshold may be associated with atherosclerotic or inflammatory subclavian artery disease, supravalvular aortic stenosis, aortic coarctation, or aortic dissection. Systolic leg pressures are usually as much as 20 mmHg higher than systolic arm pressures. Greater leg–arm pressure differences are seen in patients with chronic severe AR as well as patients with extensive and calcified lower extremity peripheral arterial disease. The ankle-brachial index (systolic pressure in the dorsalis pedis and/or posterior tibial artery divided by the higher of the two brachial artery pressures) is a powerful predictor of long-term cardiovascular mortality.

The blood pressure measured in an office or hospital setting may not accurately reflect the pressure in other venues. "White coat hypertension" (elevated clinic blood pressure and normal out of clinic blood pressure) is defined by at least three separate clinic-based measurements >130/80 mmHg and at least two non-clinic-based measurements <130/80 mmHg in the absence of any evidence of target organ damage. Individuals with white coat hypertension may not benefit from drug therapy, although they may be more likely to develop sustained hypertension over time. Masked hypertension (normal or low clinic blood pressure but elevated out of clinic blood pressure) should be suspected when normal or even low blood pressures are recorded in the office in patients with advanced atherosclerotic disease, especially when evidence of target organ damage is present or bruits are audible. Higher systolic blood pressures measured with a 24-h ambulatory blood pressure device are associated with a higher risk of cardiovascular disease and all-cause death independent of blood pressures measured in the outpatient setting.

Orthostatic hypotension is defined by a fall in systolic pressure >20 mmHg or in diastolic pressure >10 mmHg within 3 minutes of assumption of the upright posture from a supine position. There may also be a lack of a compensatory tachycardia, an abnormal response that suggests autonomic insufficiency, as may be seen in patients with diabetes mellitus or Parkinson's disease. Orthostatic hypotension is a common cause of postural lightheadedness/syncope and should be assessed routinely in patients for whom this diagnosis might pertain. It can be exacerbated by advanced age, dehydration, certain medications, food, deconditioning, and ambient temperature/humidity.

Arterial Pulse The carotid artery pulse occurs just after the ascending aortic pulse. The aortic pulse is best appreciated in the epigastrium, just above the level of the umbilicus. Peripheral arterial pulses that should be assessed routinely include the subclavian, brachial, radial, ulnar, femoral, popliteal, dorsalis pedis, and posterior tibial. In patients in whom the diagnosis of either temporal arteritis or polymyalgia rheumatica is suspected, the temporal arteries also should be examined. Although one of the two pedal pulses may not be palpable in up to 10% of normal subjects, the pair should be symmetric. The integrity of the arcuate system of the hand is assessed by Allen's test, which is performed routinely before instrumentation of the radial artery. The pulses should be examined for their symmetry, volume, timing, contour, amplitude, and duration. If necessary, simultaneous auscultation of the heart can help identify a delay in the arrival of an arterial pulse. Simultaneous palpation of the radial and femoral pulses may reveal a femoral delay in a patient with upper extremity hypertension and suspected aortic coarctation. The carotid upstrokes should never be examined simultaneously or before listening for a bruit. Light pressure should always be used to avoid precipitation of carotid hypersensitivity syndrome and syncope in a susceptible elderly individual. The arterial pulse usually becomes more rapid and spiking as a function of its distance from the heart, a phenomenon that reflects the muscular status of the more peripheral arteries and the summation of the incident and reflected waves. In general, the character and contour of the arterial pulse depend on the stroke volume, ejection velocity, vascular compliance, and systemic vascular resistance. The pulse examination can be misleading in patients with reduced cardiac output and in those with stiffened arteries from aging, chronic hypertension, or peripheral arterial disease.

The character of the pulse is best appreciated at the carotid level (Fig. 246-2). A weak and delayed pulse (*pulsus parvus et tardus*) defines severe aortic stenosis (AS). Some patients with AS may also have a slow, notched, or interrupted upstroke (anacrotic pulse) with a thrill

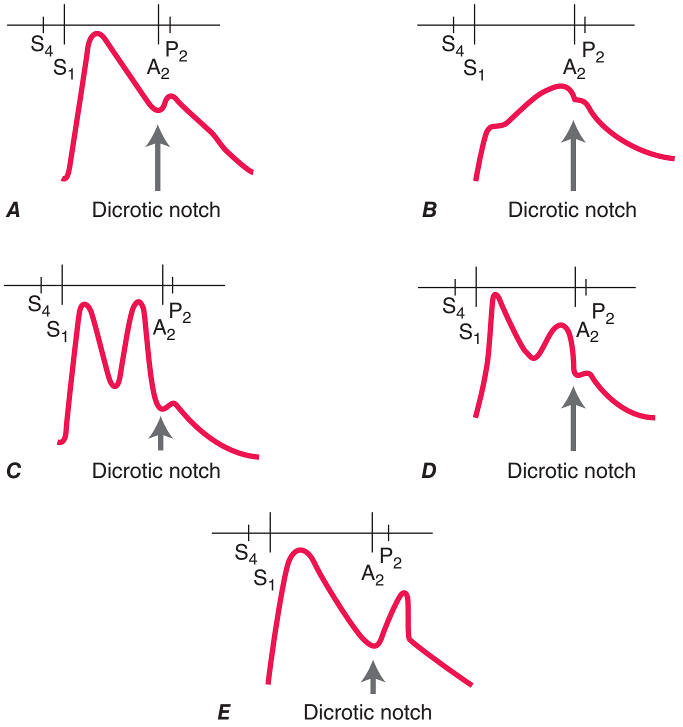

FIGURE 246-2 Schematic diagrams of the configurational changes in carotid pulse and their differential diagnoses. Heart sounds are also illustrated. ***A.*** Normal. S_4, fourth heart sound; S_1, first heart sound; A_2, aortic component of second heart sound; P_2, pulmonic component of second heart sound. ***B.*** Aortic stenosis. Anacrotic pulse with slow upstroke (tardus) to a reduced peak (parvus). ***C.*** Bisferiens pulse with two peaks in systole. This pulse is rarely appreciated in patients with severe aortic regurgitation. ***D.*** Bisferiens pulse in hypertrophic obstructive cardiomyopathy. There is a rapid upstroke to the first peak (percussion wave) and a slower rise to the second peak (tidal wave). ***E.*** Dicrotic pulse with peaks in systole and diastole. This waveform may be seen in patients with sepsis or during intraaortic balloon counterpulsation with inflation just after the dicrotic notch. *(Reproduced with permission from K Chatterjee, W Parmley [eds]: Cardiology: An Illustrated Text/Reference. Philadelphia, Gower Medical Publishers, 1991.)*

or shudder. With chronic severe AR, by contrast, the carotid upstroke has a sharp rise and rapid fall-off (Corrigan's or water-hammer pulse). Some patients with advanced AR may have a bifid or bisferiens pulse, in which two systolic peaks can be appreciated. A bifid pulse is also described in patients with hypertrophic obstructive cardiomyopathy (HOCM), with inscription of percussion and tidal waves. A bifid pulse is easily appreciated in patients on intraaortic balloon counterpulsation (IABP), in whom the second pulse is diastolic in timing.

Pulsus paradoxus refers to a fall in systolic pressure >10 mmHg with inspiration that is seen in patients with pericardial tamponade but also is described in those with massive pulmonary embolism, hemorrhagic shock, severe obstructive lung disease, and tension pneumothorax. Pulsus paradoxus is measured by noting the difference between the systolic pressure at which the Korotkoff sounds are first heard (during expiration) and the systolic pressure at which the Korotkoff sounds are heard with each heartbeat, independent of the respiratory phase. Between these two pressures, the Korotkoff sounds are heard only intermittently and during expiration. The cuff pressure must be decreased slowly to appreciate the finding. It can be difficult to measure pulsus paradoxus in patients with tachycardia, atrial fibrillation, or tachypnea. A pulsus paradoxus may be palpable at the brachial artery or femoral artery level when the pressure difference exceeds 15 mmHg. This inspiratory fall in systolic pressure is an exaggerated consequence of interventricular dependence.

Pulsus alternans, in contrast, is defined by beat-to-beat variability of pulse amplitude. It is present only when every other phase I Korotkoff sound is audible as the cuff pressure is lowered slowly, typically in a patient with a regular heart rhythm and independent of the respiratory cycle. Pulsus alternans is seen in patients with severe LV systolic dysfunction and is thought to be due to cyclic changes in intracellular calcium and action potential duration. When pulsus alternans is associated with electrocardiographic T-wave alternans, the risk for an arrhythmic event appears to be increased.

Ascending aortic aneurysms can rarely be appreciated as a pulsatile mass in the right parasternal area. Appreciation of a prominent abdominal aortic pulse should prompt noninvasive imaging with ultrasound or computed tomography for better characterization. Femoral and/or popliteal artery aneurysms should be sought in patients with abdominal aortic aneurysm disease.

The level of a claudication-producing arterial obstruction can often be identified on physical examination (Fig. 246-3). For example, in a patient with calf claudication, a decrease in pulse amplitude between the common femoral and popliteal arteries will localize the obstruction to the level of the superficial femoral artery, although inflow obstruction above the level of the common femoral artery may coexist. Auscultation for carotid, subclavian, abdominal aortic, and femoral artery bruits should be routine. However, the correlation between the presence of a bruit and the degree of vascular obstruction is poor. A cervical bruit is a weak indicator of the degree of carotid artery stenosis; the absence of a bruit does not exclude the presence of significant luminal obstruction. If a bruit extends into diastole or if a thrill is present, the obstruction is usually severe. Another cause of an arterial bruit is an arteriovenous fistula with enhanced flow.

The likelihood of significant lower extremity peripheral arterial disease increases with typical symptoms of claudication, cool skin, abnormalities on pulse examination, or the presence of a vascular bruit. Abnormal pulse oximetry (a >2% difference between finger and toe oxygen saturation) can be used to detect lower extremity peripheral

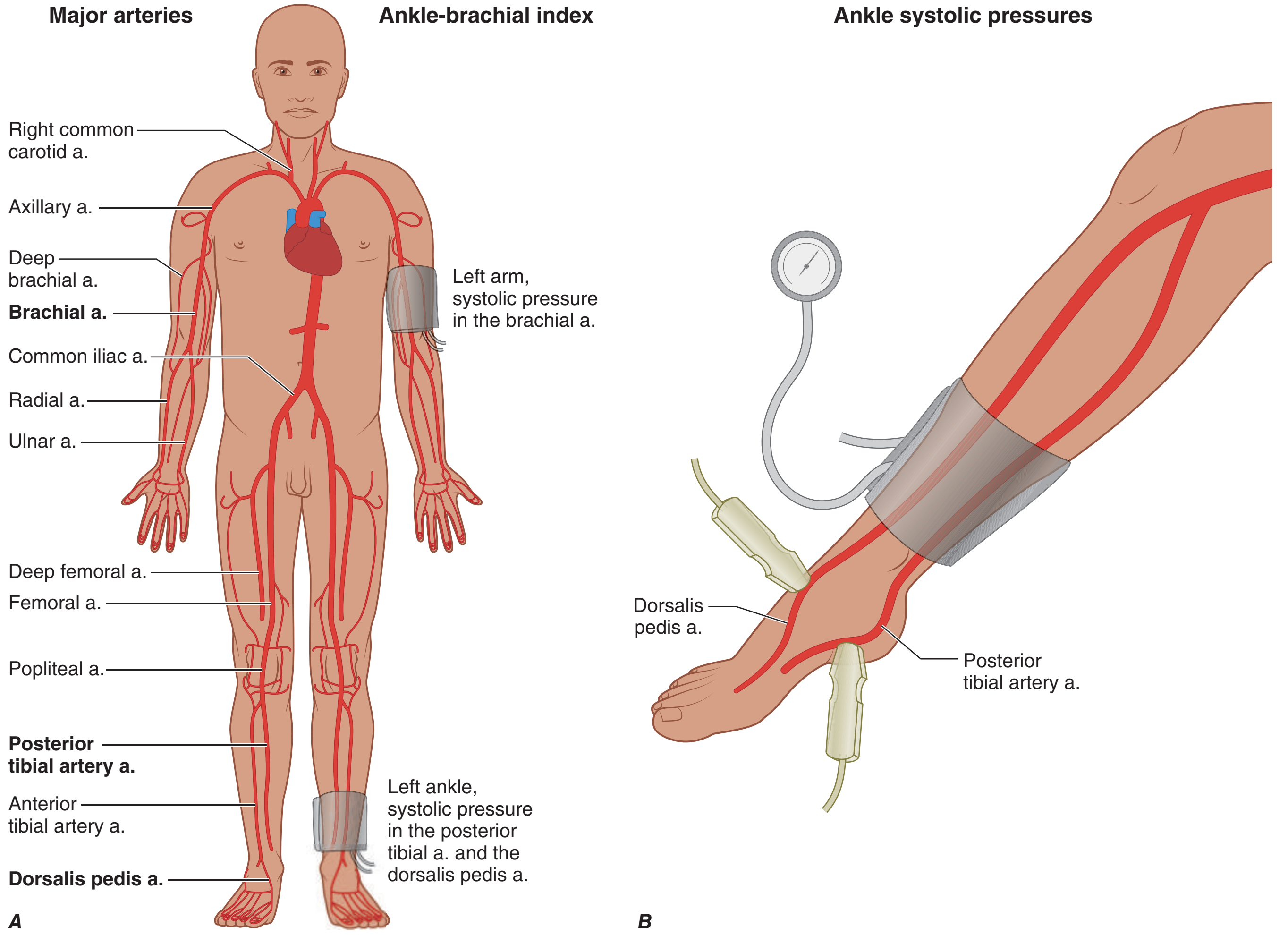

FIGURE 246-3 ***A.*** Anatomy of the major arteries of the leg. ***B.*** Measurement of the ankle systolic pressure. The ankle-brachial index (ABI) is calculated by dividing the lower of the two ankle pressures (i.e., the dorsalis pedis or posterior tibia) by the higher of the two arm pressures (i.e., left or right arm). Left and right ABIs should be recorded. *(Adapted from NA Khan et al: Does the clinical examination predict lower extremity peripheral arterial disease? JAMA 295:536, 2006.)*

arterial disease and is comparable in its performance characteristics to the ankle-brachial index.

Inspection and Palpation of the Heart The LV apex beat may be visible in the midclavicular line at the fifth intercostal space in thin-chested adults. Visible pulsations anywhere other than this expected location are abnormal. The left anterior chest wall may heave in patients with an enlarged or hyperdynamic left or right ventricle. As noted previously, a visible right upper parasternal pulsation may be suggestive of ascending aortic aneurysm disease. In thin, tall patients and patients with advanced obstructive lung disease and flattened diaphragms, the cardiac impulse may be visible in the epigastrium and should be distinguished from a pulsatile liver edge.

Palpation of the heart begins with the patient in the supine position at 30° and can be enhanced by placing the patient in the left lateral decubitus position. The normal LV impulse is <2 cm in diameter and moves quickly away from the fingers; it is better appreciated at end expiration, with the heart closer to the anterior chest wall. Characteristics such as size, amplitude, and rate of force development should be noted.

Enlargement of the LV cavity is manifested by a leftward and downward displacement of an enlarged apex beat. A sustained apex beat is a sign of pressure overload, such as that which may be present in patients with AS or chronic hypertension. A palpable presystolic impulse corresponds to the fourth heart sound (S_4) and is indicative of reduced LV compliance and the forceful contribution of atrial contraction to ventricular filling. A palpable third sound (S_3), which is indicative of a rapid early filling wave in patients with heart failure, may be present even when the gallop itself is not audible. A large LV aneurysm may sometimes be palpable as an ectopic impulse, discrete from the apex beat and often dyskinetic. HOCM may very rarely cause a triple cadence beat at the apex with contributions from a palpable S_4 and the two components of the bisferiens systolic pulse.

Right ventricular pressure or volume overload may create a sternal lift. Signs of either TR (*cv* waves in the jugular venous pulse) and/or pulmonary arterial hypertension (a loud single or palpable P_2) would be confirmatory. The right ventricle can enlarge to the extent that left-sided events are obscured. A zone of retraction between the right ventricular and LV impulses sometimes can be appreciated in patients with right ventricle pressure or volume overload when they are placed in the left lateral decubitus position. Systolic and diastolic thrills signify turbulent and high-velocity blood flow. Their locations help identify the origin of heart murmurs.

CARDIAC AUSCULTATION

Heart Sounds Ventricular systole is defined by the interval between the first (S_1) and second (S_2) heart sounds **(Fig. 246-4)**. The first heart sound (S_1) includes mitral and tricuspid valve closure. Normal splitting can be appreciated in young patients and those with right bundle branch block, in whom tricuspid valve closure is relatively delayed. The intensity of S_1 is determined by the distance over which the anterior leaflet of the mitral valve must travel to return to its annular plane, leaflet mobility, LV contractility, and the PR interval. S_1 is classically loud in the early phases of rheumatic MS and in patients with hyperkinetic circulatory states or short PR intervals. S_1 becomes softer in the later stages of MS when the leaflets are rigid and calcified, after exposure to β-adrenergic receptor blockers, with long PR intervals, and with LV contractile dysfunction. The intensity of heart sounds, however, can be reduced by any process that increases the distance between the stethoscope and the responsible cardiac event, including mechanical ventilation, obstructive lung disease, obesity, pneumothorax, and a pericardial effusion.

Aortic and pulmonic valve closure constitutes the second heart sound (S_2). With normal or physiologic splitting, the A_2–P_2 interval increases with inspiration and narrows during expiration. This physiologic interval will widen with right bundle branch block because of the further delay in pulmonic valve closure and in patients with severe MR because of the premature closure of the aortic valve. An unusually narrowly split or even a singular S_2 is a feature of pulmonary arterial hypertension. Fixed splitting of S_2, in which the A_2–P_2 interval is wide and does not change during the respiratory cycle, occurs in patients with a secundum atrial septal defect (ASD). Reversed or paradoxical splitting refers to a pathologic delay in aortic valve closure, such as that which occurs in patients with left bundle branch block, right ventricular pacing, severe AS, HOCM, and acute myocardial ischemia. With reversed or paradoxical splitting, the individual components of S_2 are audible at end expiration, and their interval narrows with inspiration, the opposite of what would be expected under normal physiologic conditions. P_2 is considered loud when its intensity exceeds that of A_2 at the base, when it can be palpated in the area of the proximal main pulmonary artery (second left interspace), or when both components of S_2 can be appreciated at the lower left sternal border or apex. The intensity of A_2 and P_2 decreases with AS and pulmonic stenosis (PS), respectively. In these conditions, a single S_2 may result.

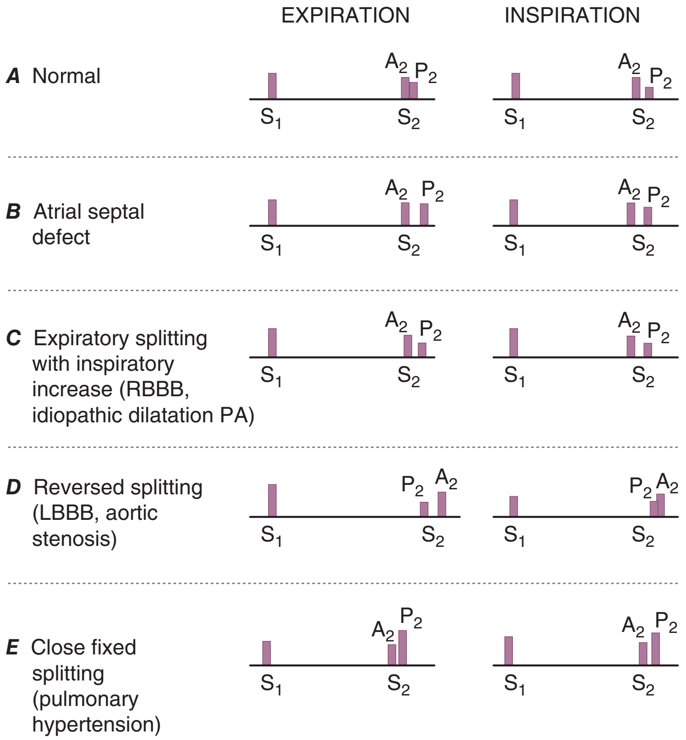
EXPIRATION
INSPIRATION
A Normal
A2 P2
S1
S2
B Atrial septal defect
C Expiratory splitting with inspiratory increase (RBBB, idiopathic dilatation PA)
D Reversed splitting (LBBB, aortic stenosis)
P2 A2
E Close fixed splitting (pulmonary hypertension)

FIGURE 246-4 Heart sounds. ***A.*** Normal. S_1, first heart sound; S_2, second heart sound; A_2, aortic component of the second heart sound; P_2, pulmonic component of the second heart sound. ***B.*** Atrial septal defect with fixed splitting of S_2. ***C.*** Physiologic but wide splitting of S_2 with right bundle branch block (RBBB). PA, pulmonary artery. ***D.*** Reversed or paradoxical splitting of S_2 with left bundle branch block (LBBB). ***E.*** Narrow splitting of S_2 with pulmonary hypertension. *(Reproduced with permission from NO Fowler: Diagnosis of Heart Disease, Springer Nature; 1991.)*

Systolic Sounds An ejection sound is a high-pitched early systolic sound that corresponds in timing to the upstroke of the carotid pulse. It usually is associated with congenital bicuspid aortic or pulmonic valve disease; however, ejection sounds are also sometimes audible in patients with isolated aortic or pulmonary root dilation and normal semilunar valves. The ejection sound that accompanies bicuspid aortic valve disease becomes softer and then inaudible as the valve calcifies and becomes more rigid. The ejection sound that accompanies PS moves closer to the first heart sound as the severity of the stenosis increases. In addition, the pulmonic ejection sound is the only right-sided acoustic event that *decreases* in intensity with inspiration. Ejection sounds are often heard more easily at the lower left sternal border than they are at the base. Nonejection sounds (clicks), which occur after the onset of the carotid upstroke, are related to MVP and may be single or multiple. The nonejection click may introduce a murmur. This click-murmur complex will move away from the first heart sound with maneuvers that increase ventricular preload, such as squatting. On standing, the click and murmur move closer to S_1.

Diastolic Sounds The high-pitched opening snap (OS) of MS occurs after a very short interval after the second heart sound. The A_2–OS interval is inversely proportional to the height of the left atrial–left ventricular diastolic pressure gradient. The intensity of both S_1 and the OS of MS decreases with progressive calcification and rigidity of the anterior mitral leaflets. The pericardial knock (PK) is also high-pitched and occurs slightly later than the OS, corresponding in timing to the abrupt cessation of ventricular expansion after tricuspid valve opening and to an exaggerated *y* descent seen in the jugular venous waveform in patients with constrictive pericarditis. A tumor plop is a lower-pitched sound that rarely can be heard in patients with atrial myxoma. It may be appreciated only in certain positions and arises from the diastolic prolapse of the tumor across the mitral valve.

The third heart sound (S_3) occurs during the rapid filling phase of ventricular diastole. It can be a normal finding in children, adolescents, and young adults; however, in older patients, it signifies heart failure. A left-sided S_3 is a low-pitched sound best heard over the LV apex. A right-sided S_3 is usually better heard over the lower left sternal border and becomes louder with inspiration. A left-sided S_3 in patients with chronic heart failure is predictive of cardiovascular morbidity and mortality. Interestingly, an S_3 is equally prevalent among heart failure patients with preserved and reduced LV ejection fraction.

The fourth heart sound (S_4) occurs during the atrial filling phase of ventricular diastole and indicates LV presystolic expansion. An S_4 is more common among patients who derive significant benefit from the atrial contribution to ventricular filling, such as those with chronic LV hypertrophy or active myocardial ischemia. An S_4 is not present with atrial fibrillation.

Cardiac Murmurs Heart murmurs result from audible vibrations that are caused by increased turbulence and are defined by their timing within the cardiac cycle. Not all murmurs are indicative of structural heart disease, and the accurate identification of a benign or functional systolic murmur often can obviate the need for additional testing in healthy subjects, especially children and adolescents. The duration, frequency, configuration, and intensity of a heart murmur are dictated by the magnitude, variability, and duration of the responsible pressure difference between two cardiac chambers, the two ventricles, or the ventricles and their respective great arteries. The intensity of a heart murmur is graded on a scale of 1 to 6; a thrill is present with murmurs of grade 4 or greater intensity. Other attributes of the murmur that aid in its accurate identification include its location, radiation, and response to bedside maneuvers. Although clinicians can detect and correctly identify heart murmurs with only fair reliability, a careful and complete bedside examination usually can identify individuals with valvular heart disease for whom transthoracic echocardiography and clinical follow-up are indicated and exclude subjects for whom no further evaluation is necessary.

Systolic murmurs can be early, mid, late, or holosystolic in timing **(Fig. 246-5)**. Acute severe MR results in a decrescendo early systolic murmur, the characteristics of which are related to the progressive attenuation of the LV to left atrial pressure gradient during systole because of the steep and rapid rise in left atrial pressure in this context. Severe MR associated with posterior leaflet prolapse or flail radiates anteriorly and to the base, where it can be confused with the murmur of AS. MR that is due to anterior leaflet involvement radiates posteriorly and to the axilla. With acute TR in patients with normal pulmonary artery pressures, an early systolic murmur that may increase in intensity with inspiration may be heard at the left lower sternal border, with regurgitant *cv* waves visible in the jugular venous pulse.

A midsystolic murmur begins after S_1 and ends before S_2; it is typically crescendo-decrescendo in configuration. AS is the most common cause of a midsystolic murmur in an adult. It is often difficult to estimate the severity of the valve lesion on the basis of the physical examination findings, especially in older hypertensive patients with stiffened carotid arteries or patients with low cardiac output in whom the intensity of the systolic heart murmur is misleadingly soft. Examination findings consistent with severe AS would include *parvus et tardus* carotid upstrokes, a late-peaking grade 3 or greater midsystolic murmur, a soft A_2, a sustained LV apical impulse, and an S_4. It is sometimes difficult to distinguish aortic sclerosis from more advanced degrees of valve stenosis. The former is defined by focal thickening and calcification of the aortic valve leaflets that is not severe enough to result in obstruction. These valve changes are associated with a Doppler jet velocity across the aortic valve of 2.5 m/s or less. Patients with aortic sclerosis can have grade 2 or 3 midsystolic murmurs identical in their acoustic characteristics to the murmurs heard in patients with more advanced degrees of AS. Other causes of a midsystolic heart murmur include pulmonic valve stenosis (with or without an ejection sound), HOCM, increased pulmonary blood flow in patients with a large ASD and left-to-right shunting, and several states associated with accelerated blood flow in the absence of structural heart disease, such as fever, thyrotoxicosis, pregnancy, anemia, and normal childhood/adolescence.

The murmur of HOCM has features of both obstruction to LV outflow and MR, as would be expected from knowledge of the pathophysiology of this condition. The systolic murmur of HOCM usually can be distinguished from other causes on the basis of its response to bedside maneuvers, including Valsalva, passive leg raising, and standing/squatting. In general, maneuvers that decrease LV preload (or increase LV contractility) will cause the murmur to intensify, whereas maneuvers that increase LV preload or afterload will cause a decrease in the intensity of the murmur. Accordingly, the systolic murmur of HOCM becomes louder during the strain phase of the Valsalva maneuver and after standing quickly from a squatting position. The murmur becomes softer with passive leg raising and when squatting. The murmur of AS is typically loudest in the second right interspace with radiation into the carotids, whereas the murmur of HOCM is best heard between the lower left sternal border and the apex. The murmur of PS is best heard in the second left interspace. The midsystolic murmur associated with enhanced pulmonic blood flow in the setting of a large ASD is usually loudest at the mid-left sternal border.

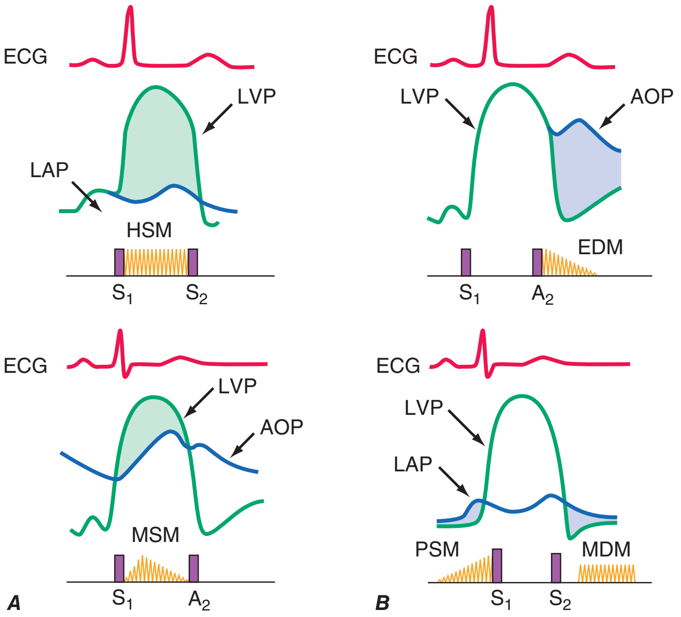

FIGURE 246-5 ***A. Top.*** Graphic representation of the systolic pressure difference (*green shaded area*) between left ventricle and left atrium with phonocardiographic recording of a holosystolic murmur (HSM) indicative of mitral regurgitation. ECG, electrocardiogram; LAP, left atrial pressure; LVP, left ventricular pressure; S_1, first heart sound; S_2, second heart sound. ***Bottom.*** Graphic representation of the systolic pressure gradient (*green shaded area*) between left ventricle and aorta in patient with aortic stenosis. A midsystolic murmur (MSM) with a crescendo-decrescendo configuration is recorded. AOP, aortic pressure. ***B. Top.*** Graphic representation of the diastolic pressure difference between the aorta and left ventricle (*blue shaded area*) in a patient with aortic regurgitation, resulting in a decrescendo, early diastolic murmur (EDM) beginning with A_2. ***Bottom.*** Graphic representation of the diastolic left atrial–left ventricular gradient (*blue areas*) in a patient with mitral stenosis with a mid-diastolic murmur (MDM) and late presystolic murmurs (PSM).

A late systolic murmur, heard best at the apex, indicates MVP. As previously noted, the murmur may or may not be introduced by one or more nonejection clicks. Differential radiation of the murmur, as previously described, may help identify the specific leaflet involved by the myxomatous process. The click-murmur complex behaves in a manner directionally similar to that demonstrated by the murmur of HOCM during the Valsalva and stand/squat maneuvers **(Fig. 246-6)**. The murmur of MVP can be identified by the accompanying nonejection click.

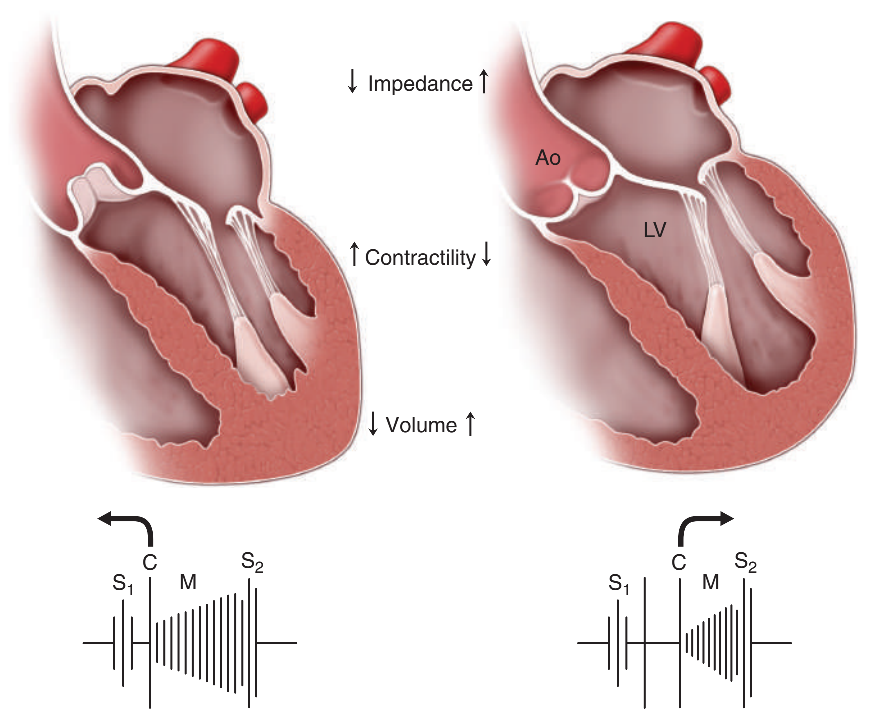

FIGURE 246-6 Behavior of the click (C) and murmur (M) of mitral valve prolapse with changes in loading (volume, impedance) and contractility. S_1, first heart sound; S_2, second heart sound. With standing (left side of figure), volume and impedance decrease, as a result of which the click and murmur move closer to S_1. With squatting (right), the click and murmur move away from S_1 due to the increases in left ventricular volume and impedance (afterload). Ao, aorta; LV, left ventricle. *(Adapted from RA O'Rourke, MH Crawford: Curr Prob Cardiol 1:9, 1976.)*

Holosystolic murmurs are plateau in configuration and reflect a continuous and wide pressure gradient between the left ventricle and left atrium with chronic MR, the left ventricle and right ventricle with a ventricular septal defect (VSD), and the right ventricle and right atrium with TR. In contrast to acute MR, in chronic MR, the left atrium is enlarged and its compliance is normal or increased to the extent that there is little if any further increase in left atrial pressure from any increase in regurgitant volume. The murmur of MR is best heard over the cardiac apex. The intensity of the murmur increases with maneuvers that increase LV afterload, such as sustained hand grip. The murmur of a VSD (without significant pulmonary hypertension) is holosystolic and loudest at the mid-left sternal border, where a thrill is usually present. The murmur of TR is loudest at the lower left sternal border, increases in intensity with inspiration (Carvallo's sign), and is accompanied by visible *cv* waves in the jugular venous wave form and, on occasion, by pulsatile hepatomegaly.

Diastolic Murmurs In contrast to some systolic murmurs, diastolic heart murmurs always signify structural heart disease (Fig. 246-5). The murmur associated with acute, severe AR is relatively soft and of short duration because of the rapid rise in LV diastolic pressure and the progressive diminution of the aortic-LV diastolic pressure gradient. In contrast, the murmur of chronic severe AR is classically heard as a decrescendo, blowing diastolic murmur along the left sternal border in patients with primary valve pathology and sometimes along the right sternal border in patients with primary aortic root pathology. When the jet of AR is directed posteriorly along the anterior leaflet of the mitral valve, the intensity of the murmur may be attenuated, and severity of the valve lesion underestimated. Detection of subtle AR murmurs can be enhanced by auscultation and end-expiration with the patient leaning forward. With chronic, severe AR, the pulse pressure is wide and the arterial pulses are bounding in character. Signs of significant diastolic run-off are often absent in the acute phase. The murmur of pulmonic regurgitation is also heard along the left sternal border. It is most commonly due to pulmonary hypertension and enlargement of the annulus of the pulmonic valve. S_2 is single and loud and may be palpable. There is a right ventricular/parasternal lift that is indicative of chronic right ventricular pressure overload. A less impressive murmur of PR is present after repair of tetralogy of Fallot or pulmonic valve atresia. In this postoperative setting, the murmur is softer and lower-pitched, and the severity of the accompanying pulmonic regurgitation can be underestimated significantly.

MS is the classic cause of a mid- to late-diastolic murmur, which is best heard over the apex in the left lateral decubitus position, is low-pitched or rumbling, and is introduced by an OS in the early stages of the rheumatic disease process. Presystolic accentuation refers to an increase in the intensity of the murmur just before the first heart sound and occurs in patients with sinus rhythm. It is absent in patients with atrial fibrillation. The auscultatory findings in patients with rheumatic tricuspid stenosis typically are obscured by left-sided events, although they are similar in nature to those described in patients with MS. "Functional" mitral or tricuspid stenosis refers to the generation of mid-diastolic murmurs that are created by increased and accelerated transvalvular diastolic flow, even in the absence of valvular obstruction, in the setting of severe MR, severe TR, or a large ASD with left-to-right shunting. The Austin Flint murmur of chronic severe AR is a low-pitched mid- to late apical diastolic murmur that sometimes can be confused with MS. The Austin Flint murmur typically decreases in intensity after exposure to vasodilators, whereas the murmur of MS may be accompanied by an OS and also may increase in intensity after vasodilators because of the associated increase in cardiac output. Unusual causes of a mid-diastolic murmur include atrial myxoma, complete heart block, and acute rheumatic mitral valvulitis.

Continuous Murmur A continuous murmur is predicated on a pressure gradient that persists between two cardiac chambers or blood vessels across systole and diastole. The murmurs typically begin in systole, envelop the second heart sound (S_2), and continue through some portion of diastole. They can often be difficult to distinguish from individual systolic and diastolic murmurs in patients with mixed valvular heart disease. The classic example of a continuous murmur is that associated with a PDA, which usually is heard in the second or third interspace at a slight distance from the sternal border. Other causes of a continuous murmur include a ruptured sinus of Valsalva aneurysm with creation of an aortic–right atrial or right ventricular fistula, a coronary or great vessel arteriovenous fistula, and an arteriovenous fistula constructed to provide dialysis access. There are two types of benign continuous murmurs. The cervical venous hum is heard in children or adolescents in the supraclavicular fossa. It can be obliterated with firm pressure applied to the diaphragm of the stethoscope, especially when the subject turns their head toward the examiner. The mammary soufflé of pregnancy relates to enhanced arterial blood flow through engorged breasts. The diastolic component of the murmur can be obliterated with firm pressure over the stethoscope.

TABLE 246-1 Effects of Physiologic Interventions on the Intensity of Heart Murmurs and Sounds

Respiration
Right-sided murmurs and sounds generally increase with inspiration, except for the PES. Left-sided murmurs and sounds are usually louder during expiration.
Valsalva Maneuver
Most murmurs decrease in length and intensity. Two exceptions are the systolic murmur of HOCM, which usually becomes much louder, and that of MVP, which becomes longer and often louder. After release of the Valsalva maneuver, right-sided murmurs tend to return to control intensity earlier than do left-sided murmurs.
After VPB or AF
Murmurs originating at normal or stenotic semilunar valves increase in the cardiac cycle after a VPB or in the cycle after a long cycle length in AF. By contrast, systolic murmurs due to AV valve regurgitation do not change or become shorter (MVP).
Positional Changes
With *standing*, most murmurs diminish, with two exceptions being the murmur of HOCM, which becomes louder, and that of MVP, which lengthens and often is intensified. With *squatting*, most murmurs become louder, but those of HOCM and MVP usually soften and may disappear. Passive leg raising usually produces the same results.
Exercise
Murmurs due to blood flow across normal or obstructed valves (e.g., PS, MS) become louder with both isotonic and submaximal isometric (hand grip) exercise. Murmurs of MR, VSD, and AR also increase with hand grip exercise. However, the murmur of HOCM often decreases with nearly maximum hand grip exercise. Left-sided S_4 and S_3 sounds are often accentuated by exercise, particularly when due to ischemic heart disease.

Abbreviations: AF, atrial fibrillation; AR, aortic regurgitation; HOCM, hypertrophic obstructive cardiomyopathy; MR, mitral regurgitation; MS, mitral stenosis; MVP, mitral valve prolapse; PES, pulmonic ejection sound; PR, pulmonic regurgitation; PS, pulmonic stenosis; TR, tricuspid regurgitation; TS, tricuspid stenosis; VPB, ventricular premature beat; VSD, ventricular septal defect.

Dynamic Auscultation Diagnostic accuracy can be enhanced by the performance of simple bedside maneuvers to identify heart murmurs and characterize their significance (Table 246-1). Except for the pulmonic ejection sound, right-sided events increase in intensity with inspiration and decrease with expiration; left-sided events behave oppositely (100% sensitivity, 88% specificity). As previously noted, the intensity of the murmurs associated with MR, VSD, and AR will increase in response to maneuvers that increase LV afterload, such as hand grip and vasopressors. The intensity of these murmurs will decrease after exposure to vasodilating agents. Squatting is associated with an abrupt increase in LV preload and afterload, whereas rapid standing results in a sudden decrease in preload. In patients with MVP, the click and murmur move away from the first heart sound with squatting because of the delay in onset of leaflet prolapse at higher ventricular volumes. With rapid standing, however, the click and murmur move closer to the first heart sound as prolapse occurs earlier in systole at a smaller chamber dimension. The murmur of HOCM behaves similarly, becoming softer and shorter with squatting (95% sensitivity, 85% specificity) and longer and louder on rapid standing (95% sensitivity, 84% specificity). A change in the intensity of a systolic murmur in the first beat after a premature beat or in the beat after a long cycle length in patients with atrial fibrillation suggests valvular AS rather than MR, particularly in an older patient in whom the murmur of the AS may be well transmitted to the apex (Gallavardin effect). Of note, however, the systolic murmur of HOCM also increases in intensity in the beat after a premature beat. This increase in intensity of any LV outflow murmur in the beat after a premature beat relates to the combined effects of enhanced LV filling (from the longer diastolic period) and postextrasystolic potentiation of LV contractile function. In either instance, forward flow will accelerate, causing an increase in the gradient across the LV outflow tract (dynamic or fixed) and a louder systolic murmur. In contrast, the intensity of the murmur of MR does not change in a postpremature beat, because there is relatively little change in the nearly constant LV to left atrial pressure gradient or further alteration in mitral valve flow. Bedside exercise can sometimes be performed to increase cardiac output and, secondarily, the intensity of both systolic and diastolic heart murmurs. Most left-sided heart murmurs decrease in intensity and duration during the strain phase of the Valsalva maneuver. The murmurs associated with MVP and HOCM are the two notable exceptions. The Valsalva maneuver also can be used to assess the integrity of the heart and vasculature in the setting of advanced heart failure.

Prosthetic Heart Valves The first clue that prosthetic valve dysfunction may contribute to recurrent symptoms is frequently a change in the quality of the heart sounds or the appearance of a new murmur. The heart sounds with a bioprosthetic valve resemble those generated by native valves. A mitral bioprosthesis usually may be associated with a grade 1 to 2 midsystolic murmur along the left sternal border (created by turbulence across the valve struts as they project into the LV outflow tract) as well as by a soft mid-diastolic murmur that occurs with normal LV filling. This diastolic murmur often can be heard only in the left lateral decubitus position and after exercise. A high-pitched or holosystolic apical murmur is indicative of pathologic MR due to a paravalvular leak and/or intra-annular bioprosthetic regurgitation from leaflet degeneration, for which diagnostic noninvasive imaging is indicated. Clinical deterioration can occur rapidly after the first expression of mitral bioprosthetic valve failure. A tissue valve in the aortic position is always associated with a grade 1 to 3 midsystolic murmur at the base or just below the suprasternal notch. A diastolic murmur of AR is abnormal in any circumstance. Mechanical valve dysfunction may first be suggested by a decrease in the intensity of either the opening or the closing sound. A high-pitched apical systolic murmur in patients with a mechanical mitral prosthesis and a diastolic decrescendo murmur in patients with a mechanical aortic prosthesis indicate paravalvular regurgitation. Patients with prosthetic valve thrombosis may present clinically with signs of shock, muffled heart sounds, and soft murmurs in both systole and diastole.

Pericardial Disease A pericardial friction rub is nearly 100% specific for the diagnosis of acute pericarditis, although the sensitivity of this finding is not nearly as high, because the rub may come and go over the course of an acute illness or be very difficult to elicit. The rub is heard as a leathery or scratchy three-component or two-component sound, although it may be monophasic. Classically, the three components are ventricular systole, rapid early diastolic filling, and late presystolic filling after atrial contraction in patients in sinus rhythm. It is necessary to listen to the heart in several positions. Additional clues to the presence of acute pericarditis may be present from the history and 12-lead electrocardiogram. The rub typically disappears as the volume of any pericardial effusion increases. Pericardial tamponade can be diagnosed with a sensitivity of 98%, a specificity of 83%, and a positive likelihood ratio of 5.9 (95% confidence interval 2.4–14) by a pulsus paradoxus that exceeds 12 mmHg in a patient with a large pericardial effusion.

The findings on physical examination are integrated with the symptoms previously elicited with a careful history to construct an appropriate differential diagnosis and proceed with indicated imaging and laboratory assessment. The physical examination is an irreplaceable component of clinician-patient interaction. Educational efforts to enhance competence may result in improved diagnostic efficiency. The adjunctive use of point-of-care ultrasound and artificial intelligence tools is expected to increase.

FURTHER READING

Chamsi-Pasha MA et al: Handheld echocardiography: Current state and future perspectives. Circulation 136:2178, 2017.

Drazner MH et al: Value of clinician assessment of hemodynamics in advanced heart failure: The ESCAPE trial. Circ Heart Fail 1:170, 2008.

Fanaroff AC et al: Does this patient with chest pain have acute coronary syndrome? The Rational Clinical Examination Systematic Review. JAMA 314:1955, 2015.

Fang LC et al: History, physical examination, and the virtual visit. An evidence-based approach, in *Braunwald's Heart Disease. A Textbook of Cardiovascular Medicine*, 12th ed, RO Bonow et al (eds). Philadelphia, Elsevier/Saunders, 2022, p. 123.

Jani V et al: The discerning ear: Cardiac auscultation in the era of artificial intelligence and telemedicine. Eur Heart J Digit Health 2:456, 2021.

247

Electrocardiography

Ary L. Goldberger

An *electrocardiogram* (ECG or EKG) is a graphical representation of electrical activity generated by the heart. The signals, detected by means of metallic electrodes attached to the extremities and chest wall, are amplified and recorded by the *electrocardiograph* device. ECG leads are configured to display the instantaneous *differences* in electrical potentials between specific sets of electrodes. The utility of the ECG derives from its immediate availability as a noninvasive, inexpensive, and highly versatile test. In addition to its use in detecting arrhythmias, conduction disturbances, and myocardial ischemia/infarction, the ECG may reveal findings related to life-threatening metabolic disturbances, drug toxicities, and increased susceptibility to sudden cardiac arrest **(see also Chaps. 317 and 420)**. The importance of electrocardiologic abnormalities in the diagnosis, prognosis, and management of muscular dystrophies and other hereditary neuromuscular diseases is discussed separately **(see Chap. 360)**.

ELECTROPHYSIOLOGIC BACKGROUND

Depolarization of the heart is the initiating event for cardiac contraction. The electrical currents that spread through the heart are produced by three components: cardiac pacemaker cells, specialized conduction tissue, and the heart muscle itself. The ECG records only the depolarization (stimulation) and repolarization (recovery) electrical activity generated by the "working" atrial and ventricular myocardium **(see also Chaps. 251 and 253)**.

The stimulus initiating the normal heartbeat originates in the *sinoatrial* (SA) *node* **(Fig. 247-1)**, which possesses spontaneous automaticity. Spread of the depolarization wave through the right and left atria induces contraction of these chambers. Next, the impulse stimulates specialized conduction tissues in the atrioventricular (AV) nodal and His-bundle areas; together, these two regions constitute the AV junction. The bundle of His bifurcates into two main divisions, the right and left bundle branches, which rapidly transmit depolarization wavefronts in a synchronous way to the right and left ventricular myocardium via the Purkinje fibers. The main left bundle fans out into left anterior and left posterior fascicular subdivisions. The depolarization wavefronts then spread through the ventricular wall, from endocardium to epicardium, triggering coordinated ventricular contraction. Since the cardiac depolarization and repolarization wavefronts have direction and magnitude, they can be represented by vectors.

BASIC ECG WAVEFORMS AND INTERVALS

The ECG waveforms are labeled alphabetically, beginning with the P wave, which represents atrial depolarization **(Fig. 247-2)**. The QRS complex represents ventricular depolarization; the ST-T-U complex (ST segment, T wave, and U wave) represents ventricular repolarization. The J point is the junction between the end of the QRS complex and the beginning of the ST segment. Atrial repolarization waveforms (ST-T_a) are usually too low in amplitude to be detected, but they may become apparent in acute pericarditis, atrial infarction, and AV heart block syndromes.

The QRS-T waveforms of the surface ECG correspond to sequential phases of simultaneously obtained ventricular *action potentials*, the intracellular recordings from single myocardial fibers **(Chap. 251)**. The rapid upstroke (phase 0) of the action potential corresponds to the onset of QRS. The plateau (phase 2) corresponds to the isoelectric ST segment, and active repolarization (phase 3) corresponds to the inscription of the T wave. Factors that decrease the slope of phase 0 by impairing the influx of Na^+ (e.g., hyperkalemia and drugs such as flecainide) tend to increase QRS duration. Factors that prolong phase 2 or 3 (e.g., amiodarone, hypocalcemia) increase the QT interval. In contrast, factors (e.g., hypercalcemia, digoxin) associated with shortening of ventricular repolarization duration abbreviate the QT. The hereditary short QT syndrome and its relationship to sudden cardiac arrest are discussed in **Chap. 262**.

The ECG is usually recorded on graph paper divided into a grid of 1-mm^2 boxes. When the recording (sweep) speed is 25 mm/s, the smallest (1 mm) horizontal divisions correspond to 40 ms (0.04 s), with heavier lines at intervals of 200 ms (0.20 s). Vertically, the ECG graph measures the amplitude of a specific wave or deflection (1 mV = 10 mm with standard calibration; the voltage criteria for hypertrophy mentioned below are given in millimeters). There are four major sets of ECG intervals: RR, PR, QRS, and QT/QT_c (Fig. 247-2). The instantaneous heart rate (beats per minute) can be computed from the interbeat (RR) interval by dividing the number of large (0.20 s) time units

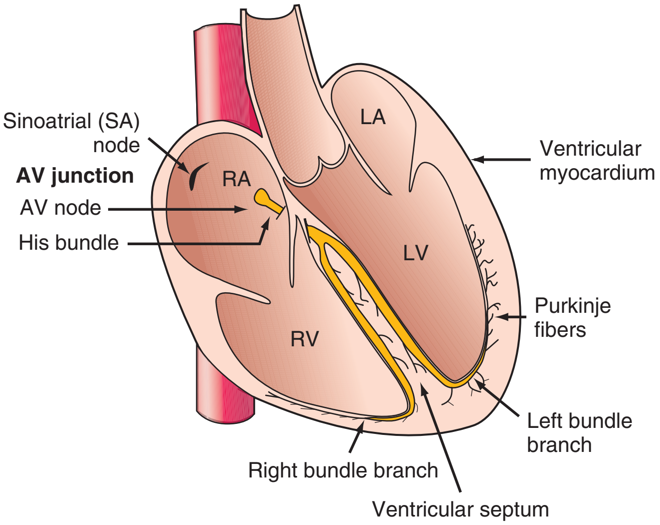

FIGURE 247-1 Schematic of the cardiac conduction system. AV, atrioventricular; LA, left atrium; LV, left ventricle; RA, right atrium; RV, right ventricle.

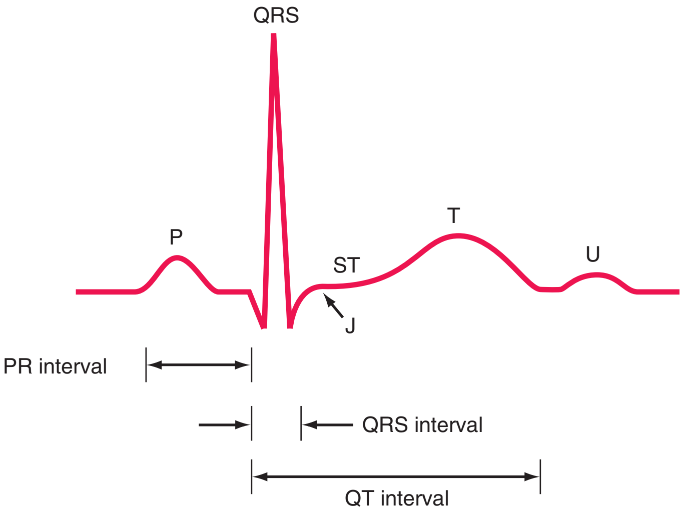

FIGURE 247-2 Basic ECG waveforms and intervals. Not shown is the RR interval, the time between consecutive QRS complexes.

between consecutive R waves into 300 or the number of small (40 ms) segments into 1500. The PR interval measures the time (normally 120–200 ms) between atrial and ventricular depolarization, which includes the physiologic delay imposed by stimulation of cells in the AV junction area. The QRS interval (normally 100–110 ms or less) reflects the duration of ventricular depolarization. The QT interval subtends both ventricular depolarization and (primarily) repolarization times and varies inversely with the heart rate. A variety of formulas has been proposed for computing a rate-corrected QT interval, termed QT_c, but without formal consensus. The classic "square root" formula ($QT_c = QT/\sqrt{RR}$, computed in second units) has been criticized for systematic errors at both lower and higher heart rates. One alternative is the Framingham formula (given here for units of milliseconds): $QT_c = QT + 0.154\,(1000 - RR)$. The following upper normal limits (based visually on the longest QT) have been proposed: QT_c of 460 ms in women and 450 ms in men. Lower limits are less well defined. Visual or electronic QT/QT_c measurements should be assessed in light of inherent limitations in their precise determination from standard ECGs waveforms.

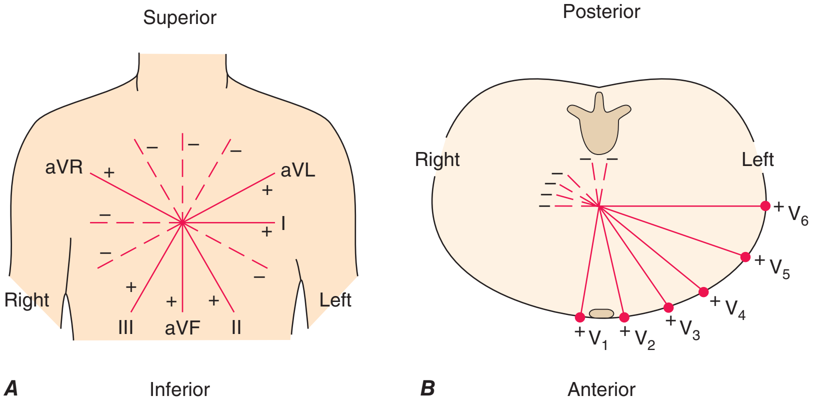

FIGURE 247-3 The six frontal plane (*A*) and six horizontal plane (*B*) leads provide a three-dimensional representation of cardiac electrical activity.

ECG LEADS

The 12 conventional ECG leads are divided into two groups: six limb (extremity) leads and six chest (precordial) leads. The limb leads record potentials transmitted onto the *frontal plane* (**Fig. 247-3*A***); the chest leads record potentials transmitted onto the *horizontal plane* (**Fig. 247-3*B***).

The orientation and polarity of the frontal plane leads are represented on a hexaxial diagram (**Fig. 247-4**). The six chest leads are obtained by exploring electrodes as shown in **Fig. 247-5**.

Each lead is analogous to a different video camera angle "looking" at the same events—atrial and ventricular depolarization and repolarization—from different spatial orientations. The 12-lead ECG can be supplemented with additional leads in special circumstances. For example, right precordial leads V_3R to V_6R are useful in detecting evidence of acute right ventricular ischemia/infarction. Bedside monitors and ambulatory ECGs (e.g., Holter monitors, event recorders, patch electrode and other medical wearable devices) usually employ only one or two modified leads. The standard ECG leads are configured such that a positive (upright) deflection is recorded in a lead if a wave of depolarization spreads toward the positive pole of that lead, and a negative deflection is recorded if the wave spreads toward the negative pole. If the *mean* orientation of the depolarization vector is at right angles to a particular lead axis, a biphasic (equally positive and negative) deflection will be inscribed.

GENESIS OF THE NORMAL ECG

P WAVE

The normal atrial depolarization vector is oriented downward and toward the subject's left, reflecting the spread of depolarization from the sinus node to the right and then the left atrial myocardium. Since this vector points toward the positive pole of lead II and toward the negative pole of lead aVR, the sinus-generated P wave will be positive in lead II and negative in aVR. By contrast, activation of the atria from an ectopic pacemaker in the lower part of either atrium or in the AV junction region may produce retrograde P waves (negative in II, positive in aVR). The normal P wave in lead V_1 may be biphasic with a positive component reflecting right atrial depolarization, followed by a small (<1 mm^2) negative component reflecting left atrial depolarization.

QRS COMPLEX

Normal ventricular depolarization proceeds as a rapid, continuous spread of activation wave fronts. This complex process can be divided

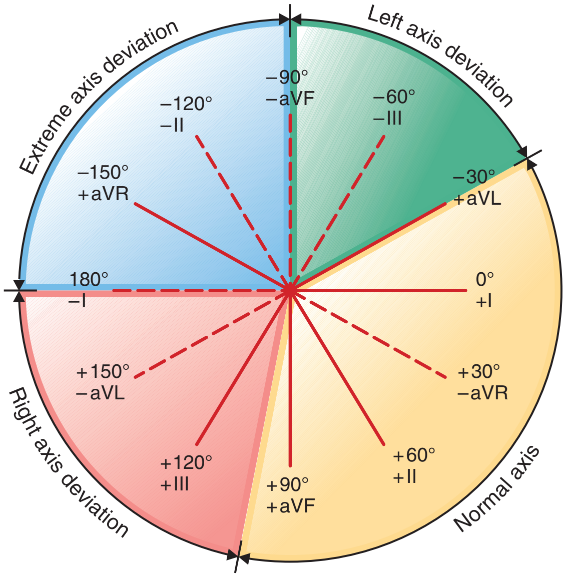

FIGURE 247-4 The frontal plane (limb or extremity) leads are represented on a hexaxial diagram. Each ECG lead has a specific spatial orientation and polarity. The positive pole of each lead axis (*solid line*) and the negative pole (*hatched line*) are designated by their angular position relative to the positive pole of lead I (0°). The mean electrical axis of the QRS complex is measured with respect to this display.

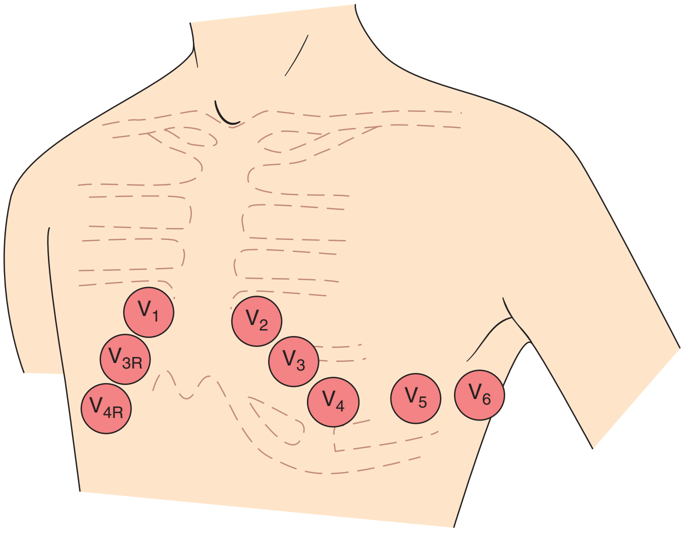

FIGURE 247-5 The horizontal plane (chest or precordial) leads are obtained with electrodes in the locations shown. Additional posterior leads are sometimes placed on the same horizontal plane as V_4 to facilitate detection of acute posterolateral infarction (V_7, midaxillary line; V_8, posterior axillary line; and V_9, posterior scapular line). Right chest leads (V_3R–V_6R) may enhance detection of right ventricular involvement in the context of inferior infarction.

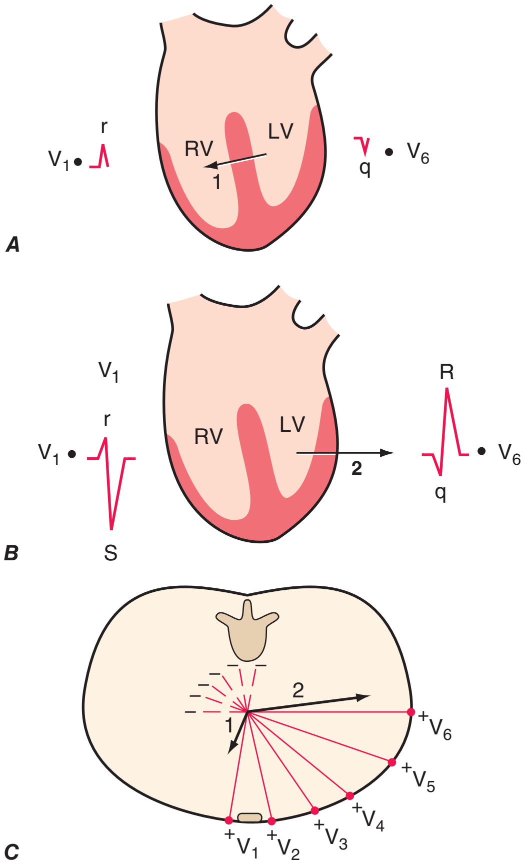

FIGURE 247-6 Ventricular depolarization can be divided into two major phases, each represented by a vector. ***A.*** The first phase (*arrow 1*) denotes depolarization of the ventricular septum, beginning on the left side and spreading to the right. This process is represented by a small "septal" r wave in lead V_1 and a small septal q wave in lead V_6. ***B.*** Simultaneous depolarization of the left and right ventricles (LV and RV) constitutes the second phase. Vector 2 is oriented to the left and posteriorly, reflecting the electrical predominance of the LV. ***C.*** Vectors (*arrows*) representing these two phases are shown in reference to the horizontal plane leads. *(Reproduced with permission from AL Goldberger et al: Goldberger's Clinical Electrocardiography: A Simplified Approach, 10th ed. Philadelphia, Elsevier, 2024.)*

into two major sequential phases, and each can be represented by a mean vector **(Fig. 247-6)**. The first and shortest phase is depolarization of the interventricular septum, proceeding from the left to the right and anteriorly (vector 1). The second and major phase results from the simultaneous depolarization of the right and left ventricles (vector 2). This phase is normally dominated by the more massive left ventricle, so that vector 2 points leftward and posteriorly. Therefore, a right precordial lead (V_1) will record this biphasic depolarization process with a small positive deflection (septal r wave) followed by a larger negative deflection (S wave). A left precordial lead, for example, V_6, will record the same sequence with a small negative deflection (septal q wave) followed by a relatively tall positive deflection (R wave). Intermediate leads show a relative increase in R-wave amplitude (normal R-wave progression) and a decrease in S-wave amplitude progressing across the chest from right to left. The lead where the R and S waves are of about equal amplitude is referred to as the *transition zone* (usually V_3 or V_4) **(Fig. 247-7)**.

The QRS pattern in the extremity leads may vary considerably from one normal subject to another depending on the *electrical axis* of the QRS, which describes the mean orientation of the QRS vector with reference to the six frontal plane leads. Normally, the QRS axis ranges from −30° to +100° (Fig. 247-4). An axis more negative than −30° is referred to as *left axis deviation*, and an axis more positive than +90 to +100° is referred to as *right axis deviation*. Left axis deviation may occur as a normal variant but is more commonly associated with left ventricular hypertrophy, a block in the anterior fascicle of the left bundle system (left anterior fascicular block or hemiblock), or inferior myocardial infarction. Right axis deviation also may occur as a normal variant (particularly in children and young adults), as a spurious finding due to reversal of the left and right arm electrodes, or in conditions such as right ventricular overload (acute or chronic), lateral infarction, dextrocardia, left pneumothorax, and left posterior fascicular block.

T WAVE AND U WAVE

Normally, the mean T-wave vector is oriented roughly concordant with the mean QRS vector (within about 45° in the frontal plane). Since depolarization and repolarization are electrically opposite processes, this normal QRS–T-wave vector concordance indicates that repolarization normally must proceed in the reverse direction from depolarization (i.e., from ventricular epicardium to endocardium). The normal U wave is a small, rounded deflection (≤1 mm) that follows the T wave and usually has the same polarity as the T wave. An abnormal increase in U-wave amplitude is most commonly due to hypokalemia or drugs (e.g., dofetilide, amiodarone, sotalol, quinidine). Very prominent U waves, as part of prolonged ventricular repolarization syndromes, are a marker of increased susceptibility to *torsades de pointes* **(Chap. 253)**.

MAJOR ECG ABNORMALITIES

CARDIAC ENLARGEMENT AND HYPERTROPHY

Right atrial overload (acute or chronic) may lead to an increase in P-wave amplitude (≥2.5 mm) **(Fig. 247-8)**, previously referred to as "P-pulmonale." Left atrial overload typically produces a biphasic P wave in V_1 with a broad negative component or a broad (≥120 ms), often with a notched P wave in one or more limb leads (Fig. 247-8). This pattern, historically referred to as "P-mitrale," may occur with *interatrial conduction delays* in the absence of actual atrial enlargement, leading to the more general designation of *left atrial abnormality*.

Right ventricular hypertrophy due to a sustained, severe pressure load (e.g., with pulmonic valve stenosis or certain pulmonary artery hypertension syndromes) is characterized by a relatively tall R wave in lead V_1 (R ≥ S wave), usually with right axis deviation **(Fig. 247-9)**; alternatively, there may be a qR pattern in V_1 or V_3R. ST depression and T-wave inversion in the right to mid-precordial leads are also

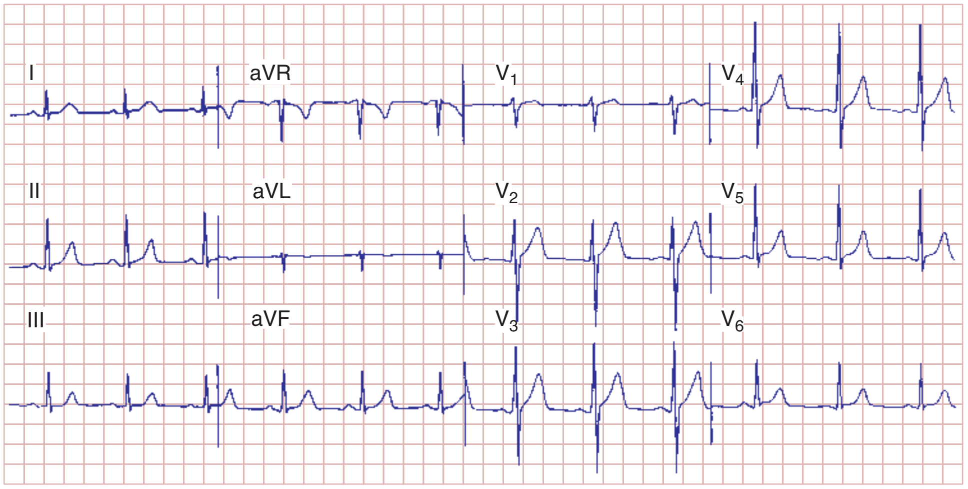

FIGURE 247-7 Normal electrocardiogram from a healthy male subject. Sinus rhythm is present with a heart rate of 75 beats per minute. PR interval is 160 ms; QRS interval (duration) is 80 ms; QT interval is 360 s; QT_c (Framingham formula) is about 390 ms; the mean QRS axis is about +70°. The precordial leads show normal R-wave progression with the transition zone (R wave ≈ S wave) in lead V_3.

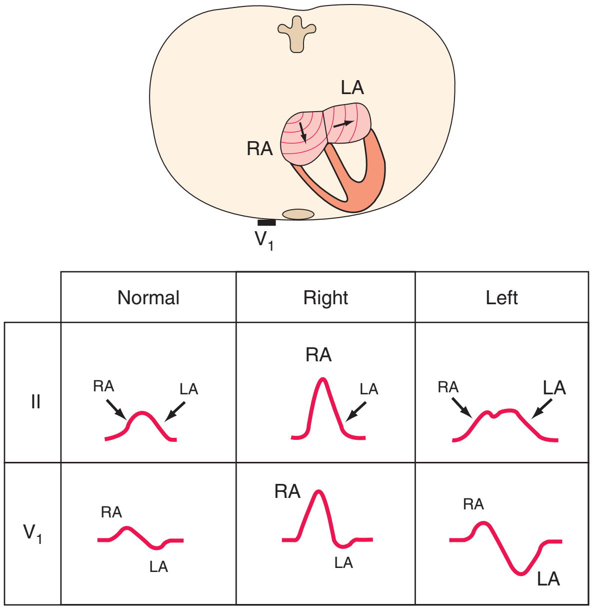

FIGURE 247-8 Right atrial (RA) overload may cause tall, peaked P waves in the limb or precordial leads. Left atrial (LA) abnormality may cause broad, often notched P waves in the limb leads and a biphasic P wave in lead V_1 with a prominent negative component representing delayed depolarization of the LA. *(Reproduced with permission from MK Park, WG Guntheroth: How to Read Pediatric ECGs, 4th ed. St. Louis, Mosby/Elsevier, 2006.)*

often present. This pattern, formerly called right ventricular "strain," is attributable to repolarization abnormalities in acutely or chronically overloaded muscle. Prominent S waves may occur in the left lateral precordial leads. Right ventricular hypertrophy due to ostium secundum atrial septal defects, with the accompanying right ventricular volume overload, is commonly associated with an incomplete or complete right bundle branch block pattern in concert with a rightward QRS axis.

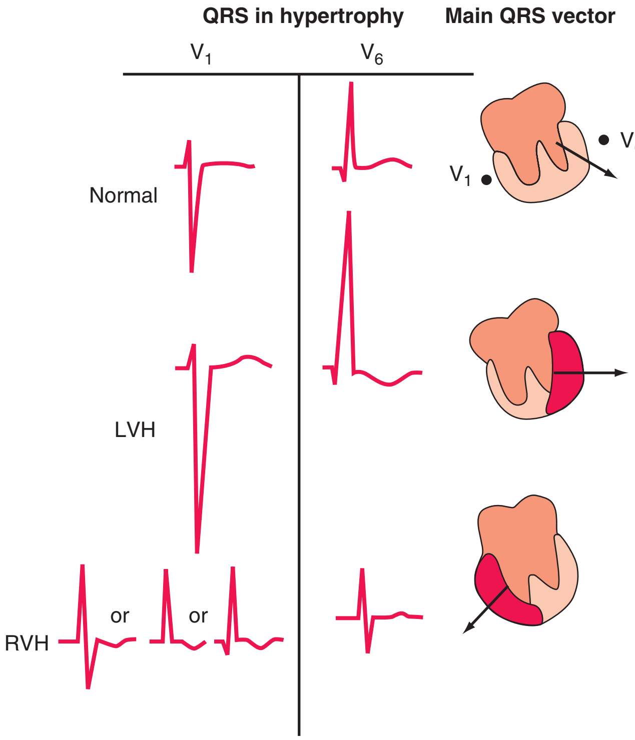

FIGURE 247-9 Left ventricular hypertrophy (LVH) increases the amplitude of electrical forces directed to the left and posteriorly. In addition, repolarization abnormalities may cause ST-segment depression and T-wave inversion in leads with a prominent R wave. Right ventricular hypertrophy (RVH) may shift the QRS vector to the right; this effect usually is associated with an R, RS, or qR complex in lead V_1. T-wave inversions may be present in right precordial leads.

Acute cor pulmonale due to pulmonary thromboembolism **(Chap. 290)** or acute respiratory distress syndromes (e.g., COVID-19) may be associated with a normal ECG or a variety of abnormalities. Sinus tachycardia is the most common arrhythmia, although other tachyarrhythmias, such as atrial fibrillation or flutter, may occur. The QRS axis may shift to the right, sometimes in concert with the so-called $S_1Q_3T_3$ pattern (prominence of the S wave in lead I and the Q wave in lead III, with T-wave inversion in lead III). Acute right ventricular dilation also may be associated with slow R-wave progression and ST-T abnormalities in V_1 to V_4 simulating acute anterior infarction. A right ventricular conduction disturbance may appear.

Chronic cor pulmonale due to obstructive lung disease **(Chap. 307)** usually does not produce the classic ECG patterns of right ventricular hypertrophy noted above. Instead of tall right precordial R waves, emphysema is more typically associated with diminished r waves in right to mid-precordial leads (slow R-wave progression) due in part to downward displacement of the diaphragm and the heart. Low-voltage complexes are commonly present, owing to hyperaeration.

Multiple voltage criteria for *left ventricular hypertrophy* (Fig. 247-9) have been proposed based on the presence of tall left precordial R waves and deep right precordial S waves (e.g., $SV_1 + [RV_5 \text{ or } RV_6] > 35$ mm). Repolarization abnormalities (ST depression with T-wave inversions, formerly called the left ventricular "strain" pattern) may appear in leads with prominent R waves. However, prominent precordial voltages occur as a common normal variant, especially in athletic or young individuals. Left ventricular hypertrophy may increase limb lead voltage with or without increased precordial voltage (e.g., $RaVL + SV_3 > 20$ mm in women and >28 mm in men). The presence of left atrial abnormality increases the likelihood of underlying left ventricular hypertrophy in cases with borderline voltage criteria. Left ventricular hypertrophy often progresses to incomplete or complete left bundle branch block. The sensitivities of conventional voltage criteria for left ventricular hypertrophy are low in middle age to older adults and may be decreased further in obese persons and smokers, as well as with right bundle branch block. ECG evidence for left ventricular hypertrophy is a major noninvasive marker of increased risk of cardiovascular morbidity and mortality rates, including sudden cardiac death. However, because of false-positive and false-negative diagnoses, the ECG is of limited utility in diagnosing atrial or ventricular enlargement. More definitive anatomic and functional information is provided by echocardiographic and cardiac magnetic resonance imaging studies **(Chaps. 248 and A9)**.

■ BUNDLE BRANCH BLOCKS AND RELATED PATTERNS

Intrinsic impairment of conduction in either the right or the left bundle system (intraventricular conduction disturbances) leads to prolongation of the QRS interval. With complete bundle branch blocks, the widest QRS interval is ≥120 ms in duration; with incomplete blocks, the QRS interval is between about 110 and 120 ms. The QRS vector usually is oriented in the direction of the myocardial region where depolarization is delayed **(Fig. 247-10)**. Thus, with right bundle branch block, the terminal QRS vector is oriented to the right and anteriorly (rSR′ in V_1 and qRS in V_6, typically). Left bundle branch block alters both early and later phases of ventricular depolarization. The major QRS vector is directed to the left and posteriorly. In addition, the normal early left-to-right pattern of septal activation is disrupted such that septal depolarization proceeds from right to left as well. As a result, left bundle branch block generates wide, predominantly negative (QS) complexes in lead V_1 and entirely positive (R) complexes in V_6. Waveform patterns identical to those of left bundle branch block, preceded by a sharp (sometimes very low amplitude) spike, are seen in most cases of electronic right ventricular pacing due to the relative delay in left ventricular activation.

Bundle branch block may occur in a variety of conditions. In subjects without structural heart disease, right bundle branch block is seen more commonly than left bundle branch block. Right bundle branch block also occurs with heart disease, both congenital (e.g., atrial septal defect) and acquired (e.g., valvular, ischemic). Left bundle branch block is often a marker of one of four underlying conditions associated with increased risk of cardiovascular morbidity and mortality rates: coronary

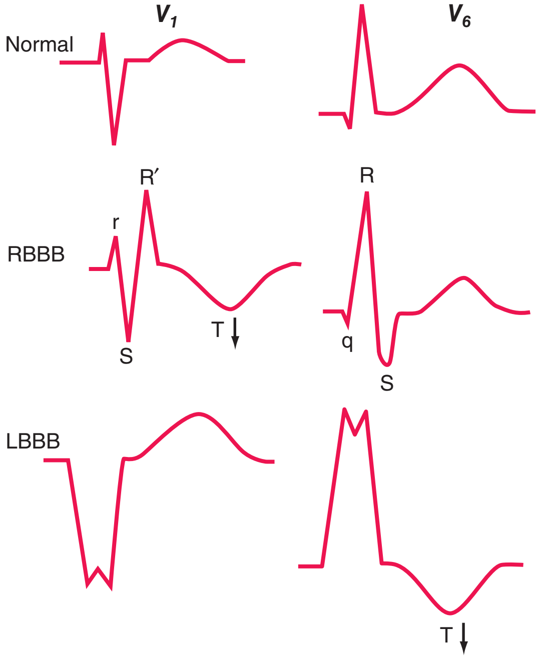

FIGURE 247-10 Comparison of typical QRS-T patterns in right bundle branch block (RBBB) and left bundle branch block (LBBB) with the normal pattern in leads V_1 and V_6. Note the secondary T-wave inversions (*arrows*) in leads with an rSR′ complex with RBBB and in leads with a wide R wave with LBBB.

heart disease (frequently with impaired left ventricular function), hypertensive heart disease, aortic valve disease (including after transcatheter aortic valve replacement), and cardiomyopathy. Bundle branch blocks may be chronic or intermittent. A bundle branch block may be rate-related, most commonly observed when the heart rate exceeds some critical value.

Bundle branch blocks and depolarization abnormalities secondary to artificial pacemakers not only affect ventricular depolarization (QRS) but also are characteristically associated with *secondary repolarization* (ST-T) abnormalities. With bundle branch blocks, the T wave is typically opposite in polarity to the last deflection of the QRS (Fig. 247-10). This discordance of the QRS–T-wave vectors is caused by the altered sequence of repolarization that occurs as a consequence of altered depolarization. In contrast, *primary repolarization* abnormalities are independent of QRS changes and are related instead to actual alterations in the electrical properties of the myocardial fibers themselves (e.g., in the resting membrane potential or action potential duration), not just to changes in the sequence of repolarization. Ischemia, electrolyte imbalance, and drugs such as digoxin all cause such primary ST–T-wave changes. Primary and secondary T-wave changes may coexist. For example, T-wave inversions in the right precordial leads with left bundle branch block or in the left precordial leads with right bundle branch block may be important markers of underlying ischemia or other abnormalities. A distinctive abnormality simulating right bundle branch block with ST-segment elevations in the right chest leads is seen with the Brugada pattern **(Chap. 262)**.

Partial blocks in the left bundle system (left anterior or posterior fascicular blocks; formerly called hemiblocks) generally do not prolong the QRS duration substantially. Instead, they are associated with shifts in the frontal plane QRS axis (leftward or rightward, respectively). Left anterior fascicular block (QRS axis more negative than –45°) is probably the most common cause of marked left axis deviation in adults. In contrast, left posterior fascicular block (QRS axis more rightward than +110–120°) is extremely rare as an isolated finding and requires exclusion of other factors causing right axis deviation. Intraventricular conduction delays also can be caused by factors extrinsic (toxic) to the conduction system that slow ventricular conduction, particularly hyperkalemia or drugs (e.g., class 1 antiarrhythmic agents, tricyclic antidepressants, phenothiazines). Prolongation of QRS duration does not necessarily indicate a conduction delay but may be due to *preexcitation* of the ventricles via a bypass tract, as in Wolff-Parkinson-White (WPW) patterns **(Chap. 256)** and related variants.

MYOCARDIAL ISCHEMIA AND INFARCTION

(See also Chap. 286) The ECG is central to the diagnosis of acute and chronic ischemic heart disease. Ischemia exerts complex time-dependent effects on the electrical properties of myocardial cells. Severe, acute ischemia lowers the resting membrane potential and shortens the duration of the action potential. Such changes cause a voltage gradient between normal and ischemic zones. As a consequence, current flows between those regions. These *currents of injury* are represented on the surface ECG by deviation of the ST segment **(Fig. 247-11)**. When the acute ischemia is *transmural*, the ST vector usually is shifted in the direction of the outer (epicardial) layers, producing ST elevations and sometimes, in the earliest stages of ischemia, tall, positive so-called hyperacute T waves over the ischemic zone. With ischemia confined primarily to the *subendocardium*, the ST vector typically shifts toward the subendocardium and ventricular cavity, so that overlying (e.g., anterior precordial) leads show ST-segment depression (with ST elevation in lead aVR). Multiple factors affect the amplitude of acute ischemic ST deviations. Profound ST elevation or depression in multiple leads usually indicates very severe ischemia. The division of acute myocardial infarction due to obstructive coronary artery disease into ST-segment elevation and non-ST elevation types is useful since the consistent efficacy of emergency (minutes to hours) reperfusion therapy is limited to the former group. Indications for acute reperfusion therapy in non-ST elevation myocardial infarction are a focus of ongoing investigation **(Chap. 285)**. Takotsubo syndrome, as well as other causes of myocardial infarction without atherosclerotic coronary disease, can simulate the patterns of acute or evolving ST-segment elevation or non-ST-segment elevation infarction **(Chap. 285)**.

The ECG leads are usually more helpful in localizing regions of ST elevation than non-ST elevation ischemia. For example, acute transmural anterior (including apical and lateral) wall ischemia is reflected by ST elevations or increased T-wave positivity in one or more of the precordial leads (V_1–V_6) and leads I and aVL. Inferior wall ischemia produces changes in leads II, III, and aVF. "Posterior" wall ischemia (almost always associated with lateral or inferior involvement) may be indirectly recognized by *reciprocal* ST depressions in leads V_1 to V_3 (thus constituting an ST elevation "equivalent" acute coronary syndrome). Acute right ventricular ischemia usually produces ST elevations in right-sided chest leads (Fig. 247-5). When ischemic ST elevations occur as the earliest sign of acute infarction, they typically are followed within a period ranging from hours to days by evolving T-wave inversions and often by Q waves occurring in the same lead distribution. Reversible transmural

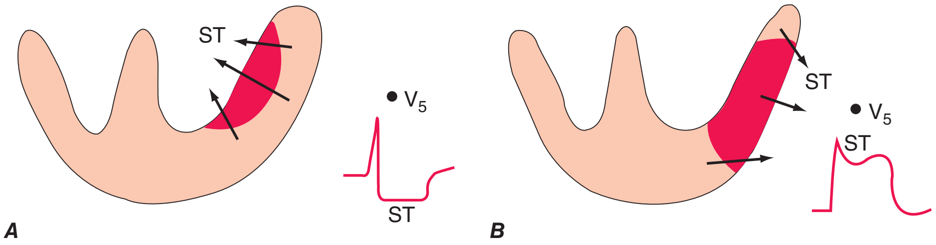

FIGURE 247-11 Acute ischemia causes a current of injury. *A.* With predominant subendocardial ischemia, the resultant ST vector will be directed toward the inner layer of the affected ventricle and the ventricular cavity. Overlying leads therefore will record ST depression. ***B.*** With ischemia involving the outer ventricular layer (transmural or epicardial injury), the ST vector will be directed outward. Overlying leads will record ST elevation.

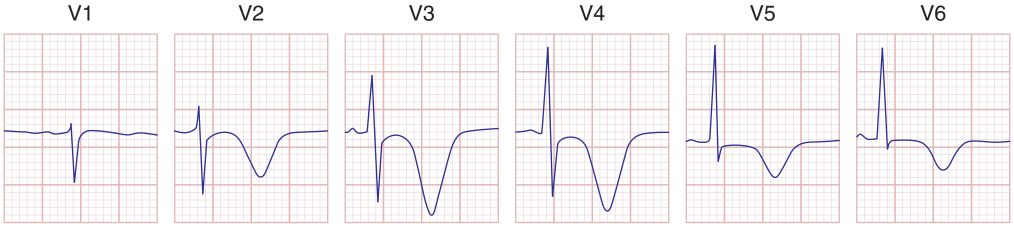

FIGURE 247-12 Severe anterior wall ischemia (with or without infarction) may cause prominent T-wave inversions in the precordial leads and in leads I and aVL. This pattern (sometimes referred to as the Wellens T wave sign) is usually associated with a high-grade stenosis of the left anterior descending coronary artery.

ischemia, for example, due to coronary vasospasm (Prinzmetal's angina) may cause transient ST-segment elevations without development of Q waves. Depending on the severity and duration of ischemia, ischemic ST elevations may resolve completely in minutes or be followed by T-wave inversions that persist for hours or even days. Patients with ischemic chest pain who present with deep T-wave inversions in multiple precordial leads (e.g., V_1–V_4, and sometimes I and aVL) with or without cardiac enzyme elevations typically have severe obstruction in the left anterior descending coronary artery **(Fig. 247-12)**.

With infarction, depolarization (QRS) changes often accompany repolarization (ST-T) abnormalities. Necrosis of sufficient myocardial tissue may lead to decreased R-wave amplitude or abnormal Q waves (even in the absence of transmural ischemia) in the anterior or inferior leads **(Fig. 247-13)**. Abnormal Q waves were once considered markers of transmural myocardial infarction, whereas subendocardial infarcts were thought not to produce Q waves. However, transmural infarcts may occur without Q waves, and subendocardial (nontransmural) infarcts may be associated with Q waves. Therefore, evolving or chronic infarcts are more appropriately classified as "Q-wave" or "non-Q-wave" **(Chap. A7)**. Loss of depolarization forces due to posterior or lateral infarction may cause reciprocal increases in R-wave amplitude in leads V_1 and V_2 without diagnostic Q waves in any of the conventional leads. (Additional leads V_7–V_9 may show acute changes.) In the weeks and months after infarction, these ECG changes may persist or begin to resolve. Complete normalization of the ECG after Q-wave infarction is uncommon but may occur, particularly with smaller infarcts. In contrast, ST-segment elevations that persist for several weeks or more after a Q-wave infarct usually correlate with a severe underlying wall motion disorder, although not necessarily a frank ventricular aneurysm.

The ECG has important limitations in both sensitivity and specificity in the diagnosis of acute and chronic ischemic heart disease. Although a single normal ECG does not exclude ischemia or even acute infarction, a normal ECG *throughout* the course of an acute infarct is distinctly uncommon. Prolonged chest pain without diagnostic ECG changes therefore should always prompt a careful search for other noncoronary causes of chest pain **(Chap. 15)**. Furthermore, the diagnostic changes of acute or evolving ischemia are often masked by the presence of left bundle branch block, electronic ventricular pacemaker patterns, and WPW preexcitation. However, clinicians may also overdiagnose ischemia or infarction based on the presence of ST-segment elevations or depressions; T-wave inversions; tall, positive T waves; or Q waves *not* related to ischemic heart disease (*pseudoinfarct* patterns). For example, ST-segment elevations simulating acute ischemia/infarction may occur with acute pericarditis or myocarditis, including COVID-19 infections, as a normal variant (including the typical "early repolarization" pattern), or in a variety of other conditions **(Table 247-1)**. Similarly, tall T waves do not invariably represent hyperacute ischemic changes but may also be caused by normal variants, hyperkalemia, or cerebrovascular injury, among other causes.

ST-segment elevations and tall, positive T waves are common findings in leads V_1 and V_2 in left bundle branch block or left ventricular hypertrophy in the absence of ischemia. The differential diagnosis of Q waves includes physiologic or positional variants, ventricular hypertrophy, acute or chronic noncoronary myocardial injury, hypertrophic cardiomyopathy, and ventricular conduction disorders. Ventricular hypertrophy, hypokalemia, drugs such as digoxin, and a variety of other factors may cause ST-segment depression mimicking subendocardial ischemia. Prominent T-wave inversion may occur with

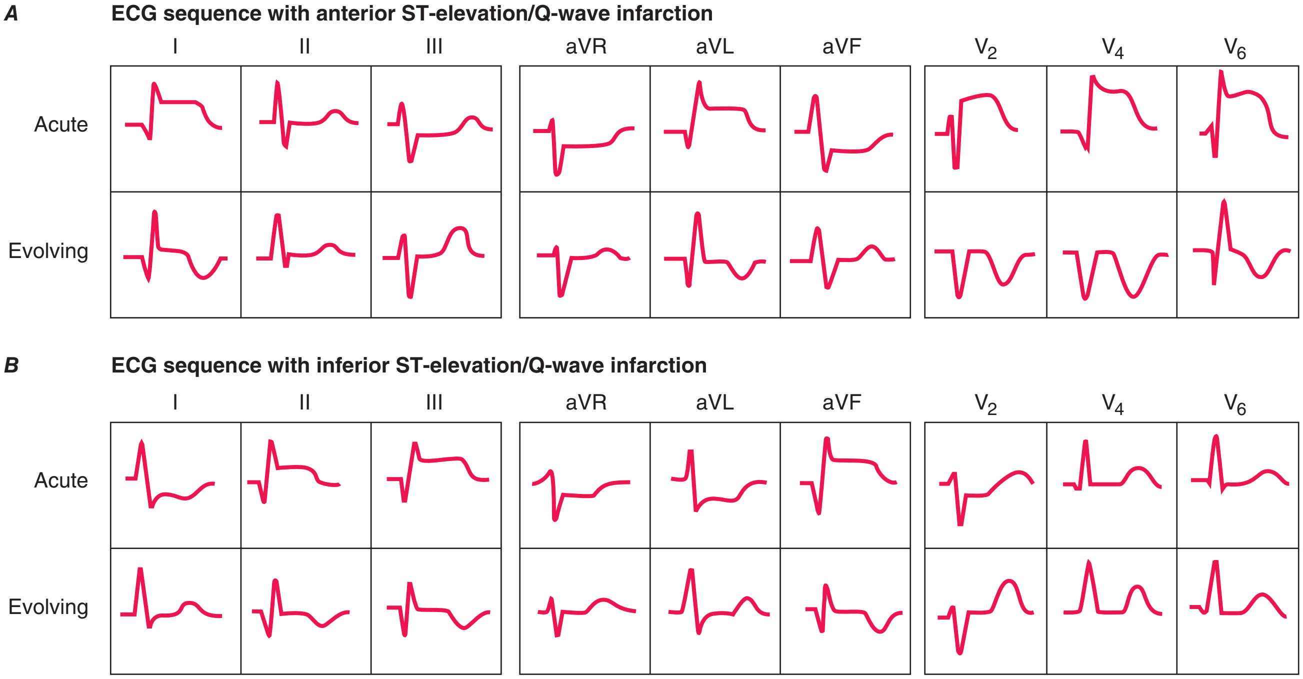
A
ECG sequence with anterior ST-elevation/Q-wave infarction
I
II
III
aVR
aVL
aVF
V2
V4
V6
Acute
Evolving
B
ECG sequence with inferior ST-elevation/Q-wave infarction
I
II
III
aVR
aVL
aVF
V2
V4
V6
Acute
Evolving

FIGURE 247-13 Sequence of depolarization and repolarization changes with acute and evolving anterior (*A*) and (*B*) inferior ST-elevation/Q-wave infarctions. With anterior infarcts, ST elevation in leads I and aVL and the precordial leads may be accompanied by reciprocal ST depressions in leads II, III, and aVF. Conversely, acute inferior (or posterolateral) infarcts may be associated with reciprocal ST depressions in leads V_1 to V_3. *(Reproduced with permission from AL Goldberger et al: Goldberger's Clinical Electrocardiography: A Simplified Approach, 10th ed. Philadelphia, Elsevier, 2024.)*

TABLE 247-1 Differential Diagnosis of ST-Segment Elevations

Myocardial ischemia/infarction
Noninfarction transmural ischemia (e.g., Prinzmetal's syndrome)
Acute myocardial infarction
Due to atherosclerotic coronary occlusion
Due to nonatherosclerotic causes (e.g., takotsubo syndrome, coronary dissection)
Post-myocardial infarction (left ventricular motion abnormality/aneurysm)
Acute pericarditis
Normal variants (including benign "early repolarization" patterns)
Left ventricular hypertrophy/left bundle branch block[a]
Other (rarer)
Acute pulmonary embolism[a]
Brugada patterns (right bundle branch block–like morphology with ST elevations in right precordial leads)
Class 1C antiarrhythmic drugs[a]
DC cardioversion (transient)
Hypercalcemia[a]
Hyperkalemia[a]
Hypothermia (J [Osborn] waves)
Nonischemic myocardial injury
Myocarditis syndromes (infectious and noninfectious)
Tumor invading left ventricle
Trauma to ventricles

[a] Usually localized to V_1–V_2 or V_3.

Source: Modified from AL Goldberger et al: *Goldberger's Clinical Electrocardiography: A Simplified Approach,* 10th ed. Elsevier, 2024.

ventricular hypertrophy, cardiomyopathies, myocarditis, and "stress cardiomyopathies" associated with takotsubo syndrome and cerebrovascular injury (particularly intracranial bleeds), among others causes. Diagnostic confusion may also occur when nonischemic T-wave inversions ("cardiac memory" effect) appear in normally conducted beats in patients with intermittent wide QRS complexes, most commonly due to ventricular pacing or to left bundle branch block.

METABOLIC FACTORS AND DRUG EFFECTS

A variety of metabolic abnormalities and pharmacologic agents alter the ECG and, in particular, cause changes in repolarization (ST-T-U) and sometimes QRS prolongation. Certain life-threatening electrolyte disturbances may be diagnosed initially and monitored from the ECG. *Hyperkalemia* produces a sequence of changes **(Fig. 247-14)**, usually beginning with narrowing and peaking (tenting) of the T waves. Further elevation of extracellular K^+ leads to AV conduction disturbances, diminution in P-wave amplitude, and widening of the QRS interval. Severe hyperkalemia eventually causes cardiac arrest with a slow sinusoidal type of mechanism ("sine-wave" pattern) followed by asystole. *Hypokalemia* **(Fig. 247-15)** prolongs ventricular repolarization, often with prominent U waves. Prolongation of the QT interval is also seen with drugs that increase the duration of the ventricular action potential: class 1A antiarrhythmic agents and related drugs (e.g., quinidine, procainamide, tricyclic antidepressants, phenothiazines) and class III agents (e.g., amiodarone [Fig. 247-15], dofetilide, sotalol, ibutilide). Systemic *hypothermia* (Fig. 247-15) also prolongs repolarization, usually with a distinctive convex elevation of the J point (Osborn wave) and bradycardia. Marked QT prolongation, sometimes with deep, wide T-wave inversions, may occur with intracranial bleeds, particularly subarachnoid hemorrhage ("CVA T-wave" pattern) (Fig. 247-15). *Hypocalcemia* typically prolongs the QT interval (ST portion), whereas *hypercalcemia* shortens it **(Fig. 247-16)**. Digitalis glycosides also shorten the QT interval, often with a characteristic "scooping" of the ST–T-wave complex (*digitalis effect*).

NONSPECIFIC ST-T CHANGES AND LOW QRS VOLTAGE

Many other factors are associated with ECG changes, particularly alterations in ventricular repolarization. T-wave flattening, minimal T-wave inversions, or slight ST-segment depression ("nonspecific ST–T-wave changes") may occur with a variety of electrolyte and acid-base disturbances, infectious or inflammatory processes, central nervous system disorders, endocrine abnormalities, many drugs, ischemia, hypoxia, and virtually any type of cardiopulmonary abnormality, in addition to physiologic changes (e.g., with posture or after meals). Low QRS voltage is arbitrarily defined as peak-to-trough QRS amplitudes of ≤5 mm in the six limb leads and/or ≤10 mm in the chest leads. Multiple factors may be responsible. Among the most serious include pericardial **(Fig. 247-17)** or pleural effusions, chronic obstructive pulmonary disease, cardiac amyloid, and anasarca.

ELECTRICAL ALTERNANS SYNDROMES

Electrical alternans—a beat-to-beat alternation in one or more components of the ECG signal—is a common type of nonlinear cardiovascular response to a variety of hemodynamic and electrophysiologic perturbations. Total electrical alternans (P-QRS-T) with sinus tachycardia is a relatively specific sign of pericardial effusion, usually with cardiac tamponade (Fig. 247-17). In contrast, pure repolarization

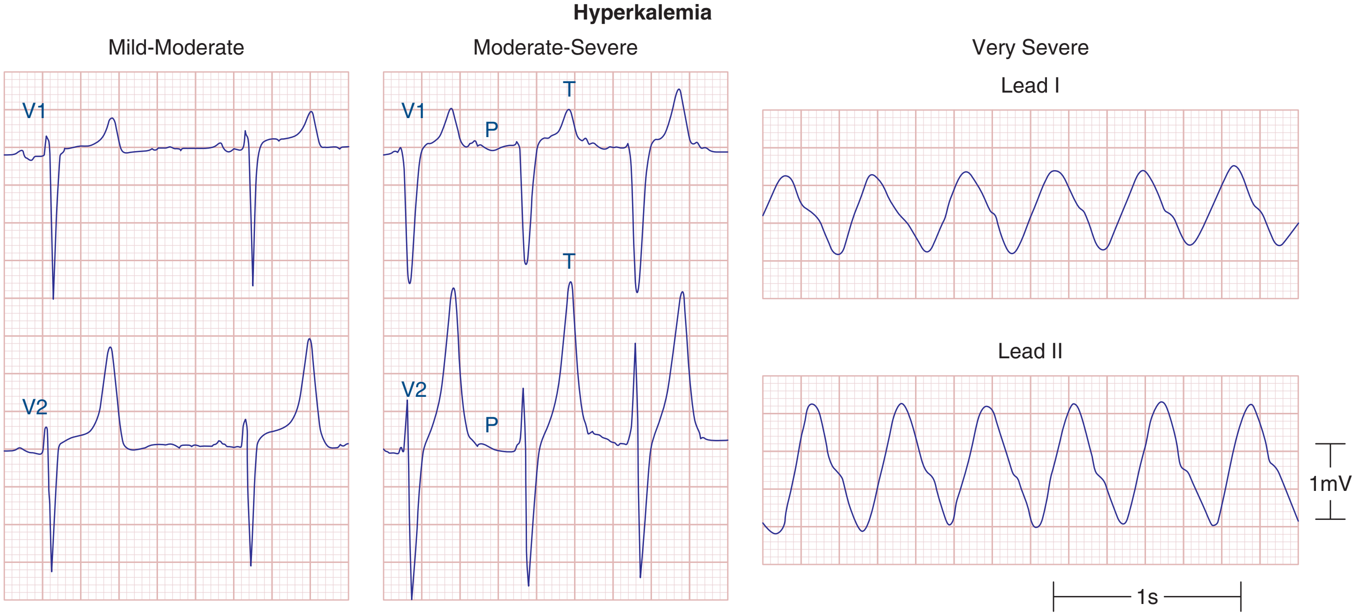

FIGURE 247-14 The earliest ECG change with hyperkalemia is usually peaking ("tenting") of the T waves. With further increases in the serum potassium concentration, the QRS complexes widen, the P waves decrease in amplitude and may disappear, and finally a sine-wave pattern leads to asystole unless emergency therapy is given. *(Reproduced with permission from AL Goldberger et al: Goldberger's Clinical Electrocardiography: A Simplified Approach, 10th ed. Philadelphia, Elsevier, 2024.)*

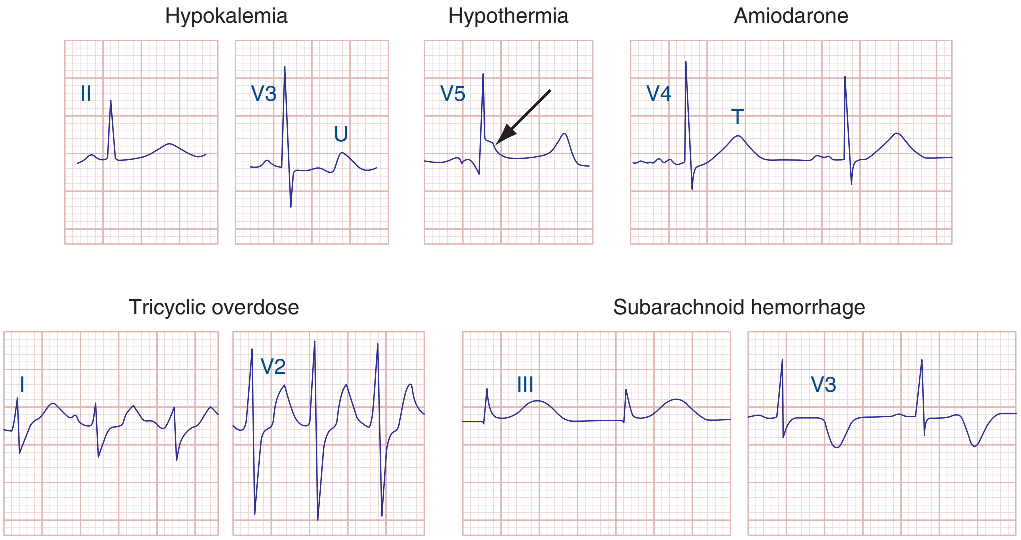
Hypokalemia
Hypothermia
Amiodarone
II
V3
U
V5
V4
T
Tricyclic overdose
Subarachnoid hemorrhage
I
V2
III
V3

FIGURE 247-15 A variety of metabolic derangements, drug effects, and other factors may prolong ventricular repolarization with QT prolongation or prominent U waves. Prominent repolarization prolongation, particularly if due to hypokalemia, inherited "channelopathies," or certain pharmacologic agents, indicates increased susceptibility to *torsades des pointes* ventricular tachycardia **(Chap. 261)**. Marked systemic hypothermia is associated with a distinctive convex "hump" at the J point (Osborn wave, *arrow*) attributed to altered transmural ventricular action potential characteristics. Note QRS and QT prolongation along with sinus tachycardia in the case of tricyclic antidepressant overdose.

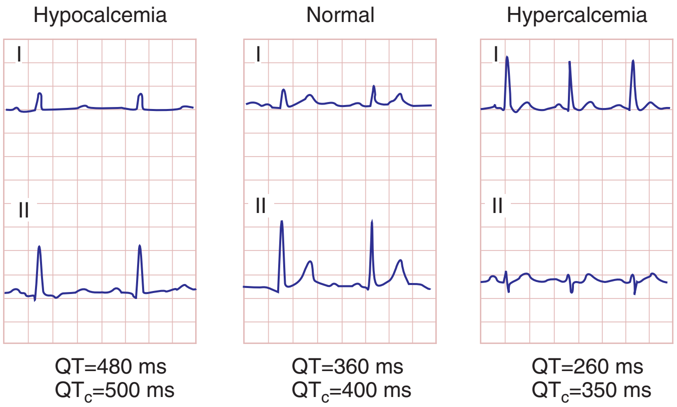
Hypocalcemia
Normal
Hypercalcemia
I
II
I
II
I
II
QT=480 ms
QT_c=500 ms
QT=360 ms
QT_c=400 ms
QT=260 ms
QT_c=350 ms

FIGURE 247-16 Prolongation of the Q-T interval (ST-segment portion) is typical of hypocalcemia. Hypercalcemia may cause abbreviation of the ST segment with relative or absolute shortening of the QT interval.

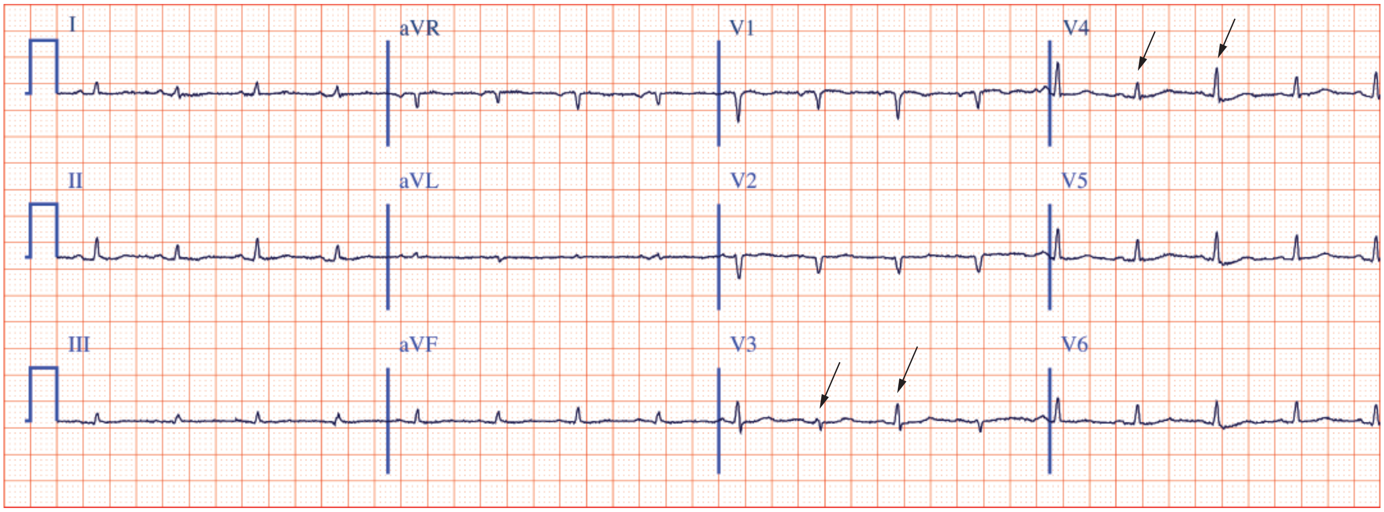
I
aVR
V1
V4
II
aVL
V2
V5
III
aVF
V3
V6

FIGURE 247-17 Classic triad of findings for pericardial effusion with cardiac tamponade: (1) sinus tachycardia; (2) low QRS voltages and (3) electrical alternans (best seen in leads V_3 and V_4 in this case; *arrows*). This triad is highly suggestive of pericardial effusion, usually with tamponade physiology, but is of limited sensitivity. *(Adapted from LA Nathanson et al: ECG Wave-Maven. ecg.bidmc.harvard.edu.)*

(ST-T or U wave) alternans is a sign of electrical instability and may precede life-threatening ventricular tachyarrhythmias.

■ CLINICAL INTERPRETATION OF THE ECG

Accurate analysis of ECGs requires thoroughness and care. The patient's age, biologic sex, and clinical status should always be considered. Many mistakes in ECG interpretation are errors of omission. Therefore, a systematic approach is essential. The following 14 parameters should be analyzed carefully in every ECG: (1) standardization (calibration) and technical features (including lead placement and artifacts); (2) rhythm; (3) heart rate; (4) PR interval/AV conduction; (5) QRS interval; (6) QT/QT_c intervals; (7) mean QRS electrical axis; (8) P waves; (9) QRS voltages; (10) precordial R-wave progression; (11) abnormal Q waves; (12) ST segments; (13) T waves; and (14) U waves. Comparison with any previous ECGs is invaluable.

■ COMPUTERIZED ELECTROCARDIOGRAPHY

Computerized systems are widely used for storage and immediate retrieval of vast numbers of ECG records. Fully automated computerized ECG analyses still have major limitations and, therefore, should not be accepted without careful clinician review.

■ FURTHER READING

Cook DA et al: Accuracy of physicians' electrocardiogram interpretations: A systematic review and meta-analysis. JAMA Intern Med 180:1461, 2020.

Goldberger AL et al: *Goldberger's Clinical Electrocardiography: A Simplified Approach*, 10th ed. Philadelphia, Elsevier, 2024.

Ismail H et al: Neuromuscular disorders and the role of the clinical electrophysiologist. J Am Coll Cardiol Clin Electrophysiol 3:1069, 2017.

Mirvis DM, Goldberger AL: Electrocardiography, in *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*, 12th ed, Libby P et al (eds). Philadelphia, Elsevier, 2021, pp. 141-174.

Nathanson LA et al: ECG Wave-Maven. Self-assessment program for students and physicians. Available at *ecg.bidmc.harvard.edu*. Accessed January 1, 2024.

Sharma S et al: International recommendations for electrocardiographic interpretation in athletes. Eur Heart J 39:1466, 2018.

248

Noninvasive Cardiac Imaging: Echocardiography, Nuclear Cardiology, and Magnetic Resonance/Computed Tomography Imaging

Marcelo F. DiCarli, Raymond Y. Kwong, Scott D. Solomon

The ability to image the heart and blood vessels noninvasively has been one of the greatest advances in cardiovascular medicine since the development of the electrocardiogram (ECG). Cardiac imaging complements history taking and physical examination, blood and laboratory testing, and exercise testing in the diagnosis and management of most diseases of the cardiovascular system. Modern cardiovascular imaging consists of echocardiography (cardiac ultrasound), nuclear scintigraphy including positron emission tomography (PET) imaging, magnetic resonance imaging (MRI), and computed tomography (CT). These studies, often used in conjunction with exercise or pharmacologic stress testing, can be used independently or can provide complementary information for patient diagnosis or prognosis. In this chapter, we review the principles of each of these modalities and the utility and relative benefits of each for the most common cardiovascular diseases.

PRINCIPLES OF MULTIMODALITY CARDIAC IMAGING

■ ECHOCARDIOGRAPHY

Echocardiography uses high-frequency sound waves (ultrasound) to penetrate the body, reflect from relevant structures, and generate an image. The basic physical principles of echocardiography are identical to other types of ultrasound imaging, although the hardware and software are optimized for evaluation of cardiac structure and function. Early echocardiography machines displayed "M-mode" echocardiograms in which a single ultrasound beam was displayed over time on a moving sheet of paper **(Fig. 248-1, *left panel*)**. Modern echocardiographic machinery uses phased array transducers that contain up to 512 elements and emit ultrasound in sequence. The reflected ultrasound is then sensed by the receiving elements, and images are generated by microprocessors that use information about the timing and magnitude of the reflected ultrasound to generate an image **(Fig. 248-1, *right panel*)**. This sequence happens repeatedly in "real time" to generate moving images with frame rates that are typically >30 frames per second but can exceed 100 frames per second. The gray scale of the image features indicates the intensity of the reflected ultrasound; fluid or blood appears black, and highly reflective structures, such as calcifications on cardiac valves or the pericardium, appear white. Tissues such as myocardium appear grayer, and tissues such as muscle display a unique speckle pattern.

The spatial resolution of ultrasound is dependent on the frequency of the ultrasound (inverse of wavelength). The higher the frequency of the ultrasound beam, the greater the spatial resolution and ability to discern small structures, but at the cost of penetration, which is decreased at higher frequencies. Thus, higher frequencies are used in pediatric imaging or transesophageal echocardiography where the transducer is typically much closer to the structures being interrogated.

Three-dimensional ultrasound transducers use a waffle-like matrix array transducer and receive a pyramidal data sector. Three-dimensional echocardiography is being increasingly used for assessment of congenital heart disease and valves, particularly with transesophageal echocardiography, although current image quality lags behind two-dimensional (2D) ultrasound **(Fig. 248-2)**.

In addition to the generation of 2D images that provide information about cardiac structure and function, echocardiography can be used to interrogate blood flow within the heart and blood vessels by using the Doppler principle to ascertain the velocity of blood flow. When ultrasound emitted from a transducer reflects off red blood cells that are moving toward the transducer, the reflected ultrasound will return at a slightly higher frequency than emitted; the opposite is true when flow moves away from the transducer. That frequency difference, termed the Doppler shift, is directly related to the velocity of the flow of the red blood cells. The velocity of blood flow between two chambers will be directly related to the pressure gradient between those chambers. A modified form of the Bernoulli equation,

$$p = 4v^2$$

where p = the pressure gradient and v = the velocity of blood flow in meters per second, can be used to calculate this pressure gradient in most clinical circumstances. This principle can be used to determine the pressure gradient between chambers and across valves and has become central to the quantitative assessment of valvular heart disease.

There are three types of Doppler ultrasound that are typically used in standard echocardiographic examinations: spectral Doppler, which consists of both pulsed wave Doppler and continuous wave Doppler, and color flow Doppler. Both types of spectral Doppler will display a waveform representing the velocity of blood flow, with time on the horizontal axis and velocity on the vertical axis. Pulsed wave Doppler

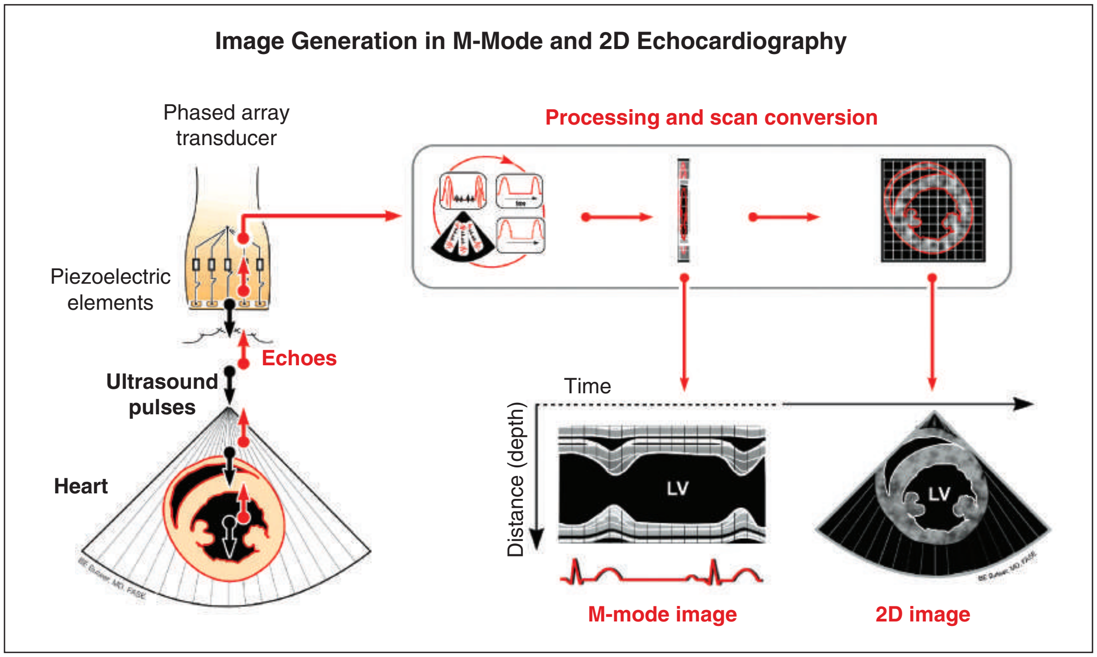
Image Generation in M-Mode and 2D Echocardiography
Phased array transducer
Processing and scan conversion
Piezoelectric elements
Echoes
Ultrasound pulses
Heart
Time
Distance (depth)
LV
LV
M-mode image
2D image

FIGURE 248-1 **Principle of image generation in two-dimensional (2D) echocardiography.** An electronically steerable phased-array transducer emits ultrasound from piezoelectric elements, and returning echoes are used to generate a 2D image (*right*) using a scan converter. Early echocardiography machines used a single ultrasound beam to generate an "M-mode" echocardiogram (see text), although modern equipment generates M-mode echocardiograms digitally from the 2D data. LV, left ventricle.

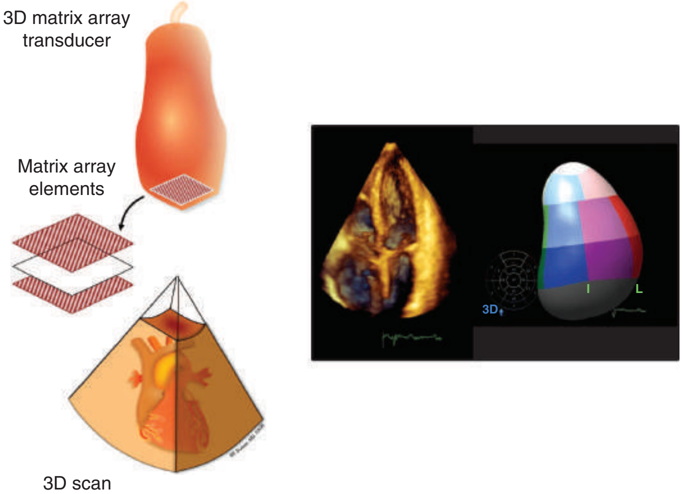
3D matrix array transducer
Matrix array elements
3D scan
3D

FIGURE 248-2 **Three-dimensional (3D) probe** and 3D image.

is used to interrogate relatively low velocity flow and has the ability to determine blood flow velocity at a particular location within the heart. Continuous wave Doppler is used to assess high velocity flow but can only identify the highest velocity in a particular direction and cannot interrogate the velocity at a specific depth location. Both techniques can only accurately assess velocities that are in the direction of the ultrasound scan lines, and velocities that are at an angle to the direction of the ultrasound beam will be underestimated. Color flow Doppler is a form of pulsed wave Doppler in which the velocity of blood flow is color encoded according to a scale and superimposed on a 2D grayscale image in real time, giving the appearance of real-time flow within the heart. The Doppler principle can also be used to assess the velocity of myocardial motion, which is a sensitive way to assess myocardial function **(Fig. 248-3)**. A standard full transthoracic echocardiographic examination consists of a series of 2D views made up of different imaging planes from various scanning locations and spectral and color flow Doppler assessment.

Transesophageal Echocardiography Transesophageal echocardiography is a form of echocardiography in which the transducer

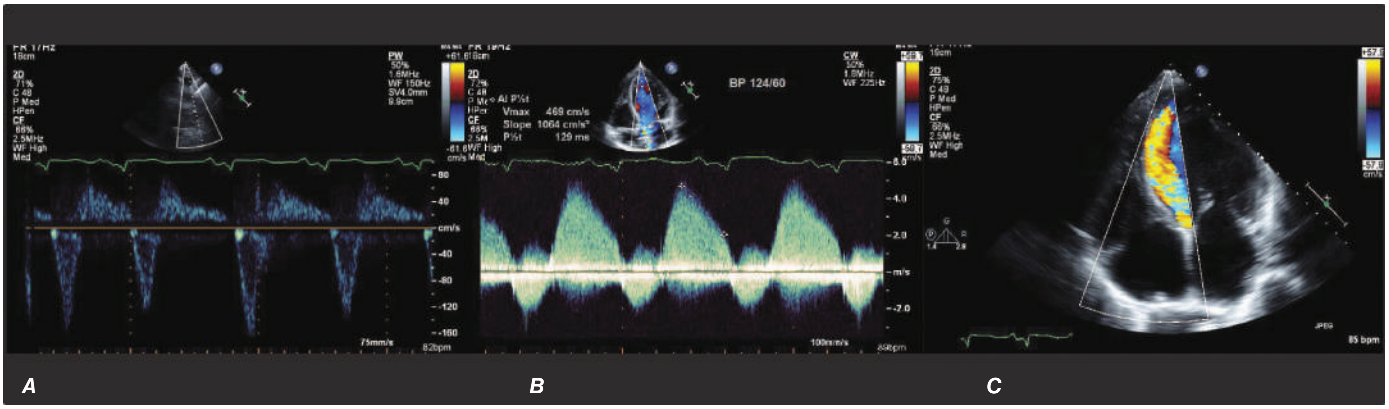
A
B
C
BP 124/60
Vmax 469 cm/s
Slope 1064 cm/s²
P½t 129 ms

FIGURE 248-3 **Three types of Doppler ultrasound.** ***A and B.*** Pulsed and continuous wave Doppler waveforms with time on horizontal axis and velocity of blood flow on vertical axis. ***C.*** Color flow Doppler, where velocities are encoded by colors according to scale on right side of screen and superimposed on a two-dimensional grayscale image.

FIGURE 248-4 Two examples of handheld ultrasound equipment: V-Scan (General Electric, *left*) and Sonosite (*right*).

is located on the tip of an endoscope that can be inserted into the esophagus. This procedure allows closer, less obstructed views of cardiac structures, without having to penetrate through chest wall, muscle, and ribs. Because less penetration is needed, a higher frequency probe can be used, and image quality and spatial resolution are generally higher than with standard transthoracic imaging, particularly for structures that are more posterior. Transesophageal echocardiography has become the test of choice for assessment of small lesions in the heart such as valvular vegetations, especially in the setting of a prosthetic valve disease, and intracardiac thrombi, including assessment of the left atrial appendage, which is difficult to visualize with standard transthoracic imaging, and for assessment of congenital abnormalities. Transesophageal echocardiography requires both topical and systemic anesthesia, generally conscious sedation, and carries additional risks such as potential damage to the esophagus, including the rare possibility of perforation, aspiration, and anesthesia-related complications. Patients generally need to give consent for transesophageal echocardiography and be monitored during and after the procedure. Transesophageal echocardiography can be carried out in intubated patients and is routinely used for intraoperative monitoring during cardiac surgery.

Stress Echocardiography Stress echocardiography is routinely used to assess cardiac function during exercise and can be used to identify myocardial ischemia or to assess valvular function under exercise conditions. Stress echocardiography is typically performed in conjunction with treadmill or bicycle exercise testing but can also be performed using pharmacologic stress most typically with an intravenous infusion of dobutamine (see section on stress imaging below).

Point-of-Care/Handheld Echocardiography Whereas typical echocardiographic equipment is large, bulky, and expensive, small handheld or point-of-care ultrasound (POCUS) equipment developed over the last decade now offers diagnostic quality imaging in a package small enough to be carried on rounds **(Fig. 248-4)**. These relatively inexpensive point-of-care devices are gaining full diagnostic capabilities and currently represent an excellent screening tool if used by an experienced operator. As these units become even smaller and less expensive, they are being increasingly used not just by cardiologists, but also by emergency medicine physicians, intensivists, anesthesiologists, and internists.

Nevertheless, echocardiography is a nontomographic modality. The image obtained is dependent on the skill of the operator who, holding a transducer, identifies standard views from which measurements can be made. Images that are obtained off-axis can result in incorrect measurements such as chamber volumes or ejection fraction. As point-of-care echocardiography becomes more commonplace, practitioners will need sufficient training to obtain and interpret images. Artificial intelligence has been incorporated into several devices to help guide the user to obtain optimal images and to help with interpretation (see below). Myocardial strain or deformation imaging has emerged as an alternative way to assess cardiac contractile performance. Global and/or regional myocardial strain can be assessed by either Doppler or, more commonly, 2D echocardiography. Global longitudinal strain is assessed from the apical view and calculated as the endocardial perimeter length in end-diastole minus the endocardial perimeter length in end-systole and divided by the endocardial perimeter length in end-diastole. It is a more robust measure of contractile function than volumetric-based ejection fraction and has been shown to be predictive of outcome in a variety of cardiac diseases, including heart failure and following myocardial infarction.

RADIONUCLIDE IMAGING

Radionuclide imaging techniques are commonly used for the evaluation of patients with known or suspected coronary artery disease (CAD), including for initial diagnosis and risk stratification as well as the assessment of myocardial viability. In addition, radionuclide imaging is commonly used in the evaluation of patients with suspected cardiac amyloidosis, myocardial and vascular inflammation, and infective endocarditis. These techniques use small amounts of radiopharmaceuticals **(Table 248-1)**, which are injected intravenously and trapped in the heart and/or vascular cells. Radioactivity within the heart and vasculature decays by emitting gamma rays. The interaction between these gamma rays and the detectors in specialized scanners (single-photon emission computed tomography [SPECT] and PET) creates a scintillation event or light output, which can be captured by digital recording equipment to form an image of the heart and vasculature. Like CT and MRI, radionuclide images also generate tomographic (three-dimensional) views of the heart and vasculature.

Radiopharmaceuticals Used in Clinical Imaging Table 248-1 summarizes the most commonly used radiopharmaceuticals in clinical SPECT and PET imaging.

Protocols for Stress Myocardial Perfusion Imaging Both exercise and pharmacologic stress can be used for myocardial

TABLE 248-1 Radiopharmaceuticals for Clinical Nuclear Cardiology

RADIOPHARMACEUTICAL	IMAGING TECHNIQUE	PHYSICAL HALF-LIFE	APPLICATION
Technetium-99m sestamibi	SPECT	6 h	Myocardial perfusion imaging
Technetium-99m tetrofosmin	SPECT	6 h	Myocardial perfusion imaging
Thalium-201	SPECT	72 h	Myocardial perfusion imaging
Iodine-123 metaiodobenzylguanidine *(MIBG)*	SPECT	13 h	Cardiac sympathetic innervation
Rubidium-82	PET	76 s	Myocardial perfusion imaging
^{13}N-ammonia	PET	10 min	Myocardial perfusion imaging
^{18}F-fluorodeoxyglucose	PET	110 min	Myocardial viability, infection and inflammation imaging
Technetium-99m pyrophosphate (PYP), technetium-99m hydroxymethylene diphosphonate (HMDP), technetium-99m 3,3-diphosphono-1,2-propanodicarboxylic acid (DPD)	SPECT	6 h	Cardiac amyloidosis

Abbreviations: PET, positron emission tomography; SPECT, single-photon emission computed tomography.

perfusion imaging. Exercise stress is generally preferred because it is physiologic and provides additional clinically important information (i.e., clinical and hemodynamic responses, ST-segment changes, exercise duration, and functional status). However, submaximal effort will lower the sensitivity of the test and should be avoided, especially if the test is requested for initial diagnosis of CAD. In patients who are unable to exercise or who exercise submaximally, pharmacologic stress offers an adequate alternative to exercise stress testing. Pharmacologic stress can be accomplished either with coronary vasodilators, such as adenosine, dipyridamole, or regadenoson, or β_1-receptor agonists, such as dobutamine. For patients unable to exercise, vasodilators are the most commonly used stress agents in combination with myocardial perfusion imaging. Dobutamine is a potent β_1-receptor agonist that increases myocardial oxygen demand by augmenting contractility, heart rate, and blood pressure such as exercise. It is generally used as an alternative to vasodilator stress in patients with chronic pulmonary disease, in whom vasodilators may be contraindicated. Dobutamine is also commonly used as a pharmacologic alternative to stress testing in stress echocardiography.

Myocardial Perfusion and Viability Imaging Protocols

Imaging protocols are tailored to the individual patient based on the clinical question, patient's risk, ability to exercise, body mass index, and other factors.

For SPECT imaging, technetium-99m (^{99m}Tc)-labeled tracers are the most used imaging agents because they are associated with the best image quality and the lowest radiation dose to the patient **(Fig. 248-5)**. Selection of the protocol (stress-only, single-day, or 2-day) depends on the patient and clinical question. After intravenous injection, myocardial uptake of ^{99m}Tc-labeled tracers is rapid (1–2 min). After uptake, these tracers become trapped intracellularly in mitochondria and show minimal change over time. This is why ^{99m}Tc tracers can be helpful in patients with chest pain of unclear etiology occurring at rest because patients can be injected while having chest pain and imaged sometime later after symptoms subside. Indeed, a normal myocardial perfusion study following a rest injection in a patient with active chest pain effectively excludes myocardial ischemia as the cause of chest pain (high negative predictive value). While used commonly in the past for perfusion imaging, thallium-201 protocols are now rarely used because they are associated with a higher radiation dose to the patient.

PET myocardial perfusion imaging is an alternative to SPECT and is associated with improved diagnostic accuracy and lower radiation dose to patients (Table 248-1). The ultra-short half-life of some PET radiopharmaceuticals in clinical use (e.g., rubidium-82) is the primary reason why imaging is generally combined with pharmacologic stress, as opposed to exercise. However, exercise is possible for relatively longer-lived radiotracers (e.g., ^{13}N-ammonia). PET imaging protocols are typically faster than SPECT, but more expensive. In comparison to SPECT, PET has improved spatial and contrast resolution and provides absolute measures of myocardial perfusion (in mL/min per gram of tissue), thereby providing the patients' regional and global coronary flow reserve. The latter helps improve diagnostic accuracy and risk stratification, especially in obese patients, women, and higher risk individuals (e.g., diabetes mellitus) **(Fig. 248-6)**. Contemporary PET and SPECT scanners are combined with a CT scanner (so-called *hybrid PET/CT* and *SPECT/CT*). CT is used primarily to guide patient positioning in the field of view and for correcting inhomogeneities in radiotracer distribution due to attenuation by soft tissues (so-called *attenuation correction*). However, it can also be used to obtain diagnostic data including coronary artery calcium (CAC) score and/or CT coronary angiography (discussed below).

For the evaluation of myocardial viability in patients with ischemic cardiomyopathy, myocardial perfusion imaging (with SPECT or PET) is usually combined with metabolic imaging (i.e., fluorodeoxyglucose [FDG] PET). In hospital settings lacking access to PET scanning, thallium-201 SPECT imaging is an excellent alternative. FDG-PET is also used in the evaluation of myocardial and vascular inflammation and in patients with suspected infective endocarditis.

Bone scintigraphy with SPECT is currently used for the evaluation of patients with suspected cardiac amyloidosis. As discussed below under applications in new-onset heart failure, bone-seeking ^{99m}Tc radiotracers (Table 248-1) are currently used to diagnose transthyretin cardiac amyloidosis with high accuracy.

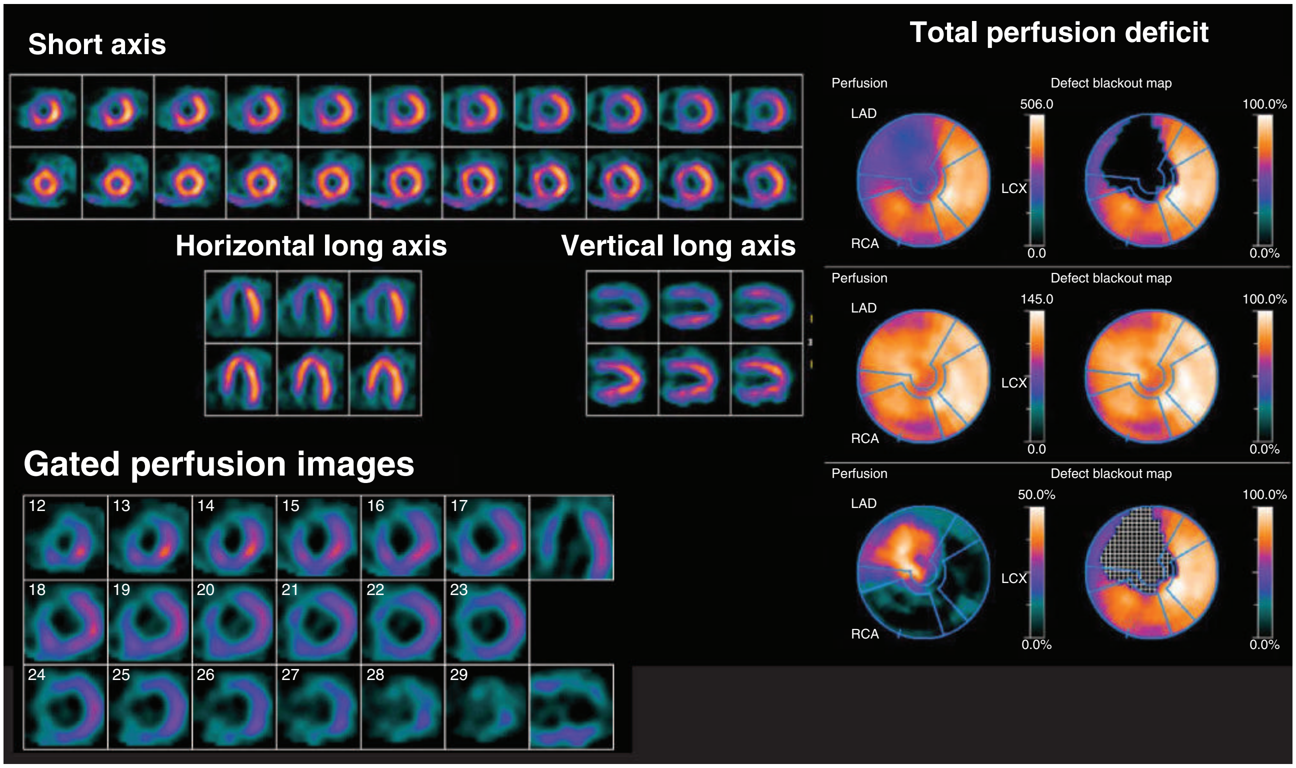

FIGURE 248-5 Tomographic stress (top of each pair) and rest myocardial perfusion images with technetium-99m sestamibi single-photon emission computed tomography imaging demonstrating a large perfusion defect throughout the anterior and anteroseptal walls. The *right panel* demonstrates the quantitative extent of the perfusion abnormality at stress (*top bull's-eye*) and at rest (*middle bull's-eye*), and the extent of defect reversibility (*lower bull's-eye*). The *lower left panel* demonstrates electrocardiogram-gated myocardial perfusion images from which one can determine the presence of regional wall motion abnormalities and calculate left ventricular volumes and ejection fraction.

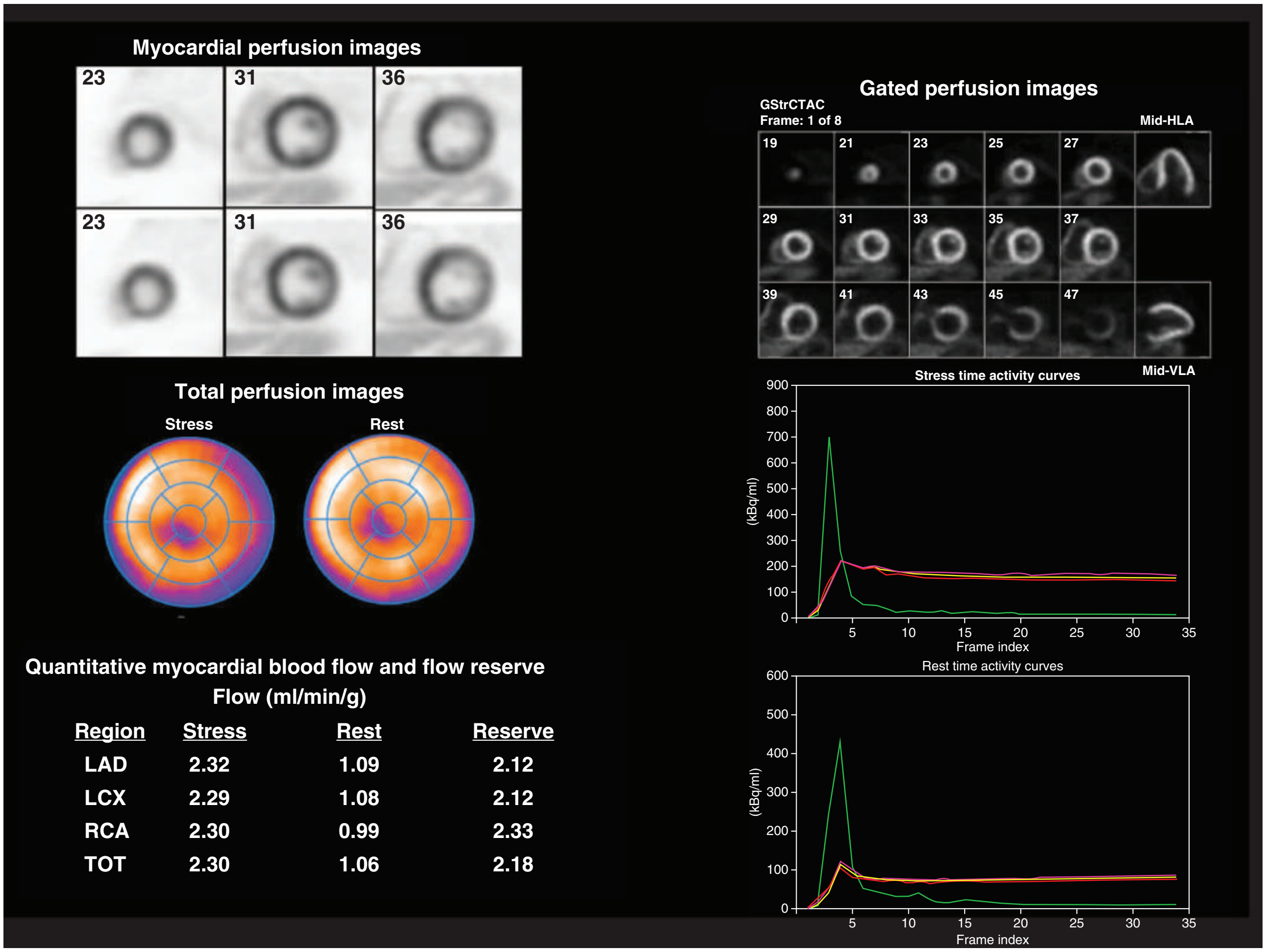

Region	Stress	Rest	Reserve
LAD	2.32	1.09	2.12
LCX	2.29	1.08	2.12
RCA	2.30	0.99	2.33
TOT	2.30	1.06	2.18

FIGURE 248-6 Multidimensional cardiac imaging protocol with positron emission tomography. The *left upper panel* demonstrates stress and rest short-axis images of the left and right ventricles demonstrating normal regional myocardial perfusion. The *middle panel* demonstrates the quantitative bull's-eye display to evaluate the extent and severity of perfusion defects. The *lower right panel* illustrates the time-activity curves for quantification of myocardial blood flow. The *right upper panel* demonstrates electrocardiogram-gated myocardial perfusion images from which one can determine the presence of regional wall motion abnormalities and calculate left ventricular volumes and ejection fraction. LAD, left anterior descending artery; LCX, left circumflex artery; RCA, right coronary artery; TOT, total left ventricle.

■ CARDIAC COMPUTED TOMOGRAPHY

CT acquires images by passing a thin x-ray beam through the body at many angles to generate cross-sectional images. The x-ray transmission measurements are collected by a detector array and digitized into pixels that form an image. The grayscale information in individual pixels is determined by the attenuation of the x-ray beam along its path by tissues of different densities, referenced to the value for water in units known as Hounsfield units. In the resulting CT images, bone appears bright white, air is black, and blood and muscle show varying shades of gray. However, due to the limited contrast between cardiac chambers and vascular structures, iodinated contrast agents are necessary for most cardiovascular indications. Cardiac CT produces tomographic images of the heart and surrounding structures. With modern CT scanners, a three-dimensional dataset of the heart can be acquired in 5–15 s with submillimeter spatial resolution.

CT Calcium Scoring CT calcium scoring is the simplest application of cardiac CT and does not require administration of iodinated contrast. The presence of coronary artery calcification has been associated with increased burden of atherosclerosis and cardiovascular mortality. Coronary calcium is then quantified (e.g., Agatston score) and categorized as minimal (0–10), mild (10–100), moderate (100–400), or severe (>400) **(Fig. 248-7)**. CAC scores are then normalized by age and gender and reported as percentile scores. Population-based studies in asymptomatic cohorts have reported high cardiac prognostic value of CT calcium score. With appropriate techniques, the radiation dose associated with CAC scanning is very low (~1–2 mSv).

CT Coronary Angiography Coronary CT angiography (CTA) is a clinically important alternative to stress testing in selected patients with suspected CAD. Imaging of the coronary arteries by CT is challenging because of their small luminal size and because of cardiac and respiratory motion. Respiratory motion can be reduced by breath-holding, and cardiac motion is best reduced by slowing the patient's heart rate, ideally to under 60 beats/min, using intravenous or oral beta blockade or other rate-lowering drugs. When performing a coronary CTA, image quality is further enhanced using sublingual nitroglycerin to enlarge the coronary lumen just prior to contrast injection. Imaging the whole-heart volume is synchronized to the administration of weight-based and appropriately timed intravenous iodinated contrast. Image acquisition is linked to the timing of the cardiac cycle through ECG triggering. The resulting images are then postprocessed using a three-dimensional workstation, which facilitates interpretation of the coronary anatomy and estimation of the severity of atherosclerosis (Fig. 248-7). A modified version of the enhanced protocol outlined for coronary CTA is used for the evaluation of patients with structural heart disease, especially for preprocedural planning of those undergoing transcatheter valve replacement.

■ CARDIAC MAGNETIC RESONANCE

Cardiac magnetic resonance (CMR) imaging is based on imaging of protons in hydrogen, which is an advantage, given the abundance of water in the human body. When the body is placed inside an MRI scanner, protons in different tissues, such as in simple fluid or complex macromolecules such as fat or protein, interact with the magnetic field

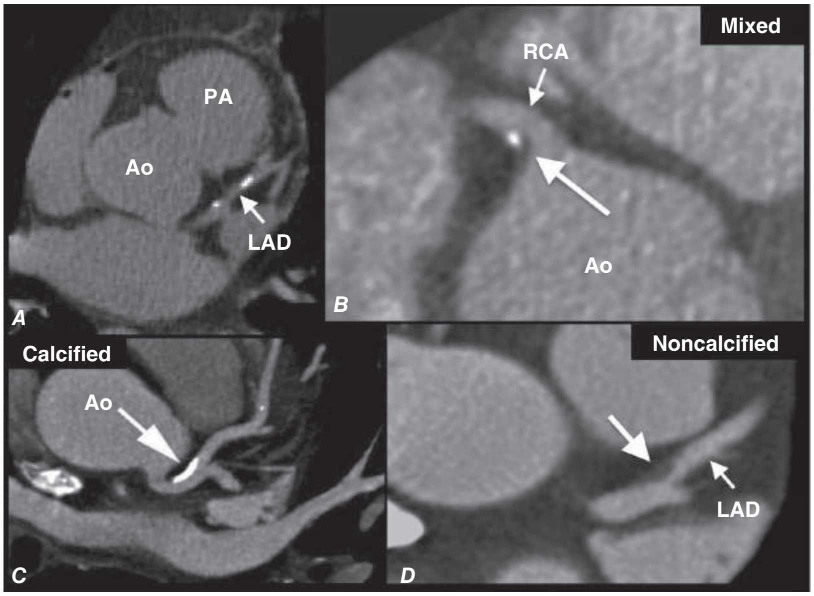

FIGURE 248-7 Examples of non-contrast- and contrast-enhanced coronary imaging with computed tomography (CT). ***A.*** Calcified coronary plaques in the distal left main and proximal left anterior descending coronary artery (LAD) in a noncontrast cardiac CT scan. Calcium deposits are dense and present as bright white structures on CT, even without contrast enhancement. ***B, C, and D.*** Different types of atherosclerotic plaques on contrast-enhanced CT scans. Importantly, noncalcified plaques are evident only on contrast-enhanced CT scans. AO, aorta; PA, pulmonary artery; RCA, right coronary artery.

at their unique frequencies. A set of orthogonal gradient coils in the scanner is designed to locate protons spatially so that radiofrequency (RF) energy pulses can be delivered to select imaging planes of interest. Once the RF pulses stop, the energy absorbed will be released, collected by the phased-array receiver coils placed on the patient's body surface, digitally recorded in a data matrix known as the K-space, then reconstructed into a magnetic resonance image. The large arrays of software methods of delivering RF pulses are known as pulse sequences, which aim to extract different types of cardiac structural or physiologic information. In CMR, T1-weighted pulse sequences are most common, and they assess cardiac structure and function, blood flow, and myocardial perfusion with pharmacologic stress. T2-weighted and T2*-weighted pulse sequences, on the other hand, evaluate myocardial edema and myocardial iron infiltration, respectively. In recent years, T1 and T2 mapping have become routinely used in experienced centers in quantifying myocardial tissue characteristics, with T1 mapping most commonly used to scale the spectrum of myocardial inflammation or fibrosis and T2 mapping for myocardial edema. Using a combination of these methods above, a CMR can accurately assess cardiac structures and function, myocardial infarction and viability, ischemia, and infiltration. Currently, the most common indications for a CMR include assessing the etiology of cardiomyopathy, myocardial ischemia from chest pain syndromes, myocardial viability for benefits of revascularization, pericardial diseases, and myocardial substrates of arrhythmias. Vector ECG-gating and repetitive patient breath-holding are used by convention to suppress cardiac and respiratory motions, respectively. However, with technical advent, rapid data collection algorithms and diaphragmatic position gating have eliminated the need for ECG gating and breath-holding in challenging situations. A list of common pulse sequences used in CMR is shown in Table 248-2.

ASSESSMENT OF CARDIAC STRUCTURE AND FUNCTION

Echocardiography, CMR, and cardiac CT are all capable of assessing cardiac structure and function, although echocardiography is generally considered the primary imaging method for these assessments. Radionuclide imaging can also be used to assess left ventricular regional and global systolic function. Echocardiography is most often used to assess the size of all four chambers and thickness of ventricular walls, which are affected by both cardiac and systemic diseases.

The structure of the left ventricle is generally assessed by determining its volume and mass. Left ventricular volumes can be easily estimated from 2D echocardiography by using methods incorporating geometric assumptions. The accuracy of these echocardiographic methods is reduced when foreshortening of the imaging plane leads to underestimation of volumes. Moreover, these methods require accurate delineation of the endocardial border. In this regard, high-resolution tomographic techniques such as CMR or cardiac CT are more accurate for volumetric assessment. Three-dimensional (3D) echocardiography does not require any geometric assumptions about the left ventricle for quantification of volumes and ejection fraction. However, 3D echocardiographic imaging requires substantial expertise and currently is not widely used in practice.

Left ventricular dilatation is common to several cardiac diseases. For example, regional dysfunction secondary to myocardial infarction can ultimately lead to progressive ventricular dilatation or remodeling. Although dilatation often begins in the region affected by the

TABLE 248-2 Clinical Cardiac Magnetic Resonance Pulse Sequences and Their Application

PULSE SEQUENCE	KEY IMAGING INTERESTS
Cardiac Morphology	
Still frame imaging (black or bright blood)	Cardiac structures
Cardiac Function	
Cine imaging	Left ventricular volume and function
Cine myocardial tagging	Left ventricular deformation (strain)
Blood Flow Imaging	
Velocity-encoded phase contrast	Cardiac and great vessel flow
Stress Testing	
Myocardial perfusion imaging	Regional myocardial blood flow
Cine imaging	Regional wall motion
Myocardial Tissue Characterization	
Late gadolinium enhancement	Myocardial infarction and infiltrative disease
T2-weighted imaging	Myocardial edema
Iron content imaging	Myocardial iron infiltration
Magnetic Resonance Angiography	
Aorta, peripheral and coronary arteries	Luminal stenosis and vessel wall remodeling

infarction, subsequent compensatory dilatation can occur in remote myocardial regions as well. The presence of regional wall motion abnormalities associated with ventricular thinning (reflecting scar) in a coronary distribution is strongly suggestive of an ischemic etiology. Regional wall motion can be accurately assessed by echocardiography, CMR, and cardiac CT. Direct assessment of infarcted myocardium is possible with both CMR (evident as areas of late gadolinium enhancement [LGE]) and radionuclide imaging (as assessed by regional perfusion or metabolic defects at rest). CMR can be particularly useful in determining etiology of cardiomegaly and ventricular dysfunction, with LGE in coronary distributions being nearly pathognomonic for infarction **(Video 248-1)**.

More global ventricular dilatation is seen in cardiomyopathy and dilatation due to valvular heart disease. Idiopathic, nonischemic cardiomyopathies will typically result in global ventricular dilatation and dysfunction, with thinning of the walls. Patients with substantial ventricular dyssynchrony due to conduction abnormalities will have a typical pattern of contraction (e.g., delay of contraction of the lateral wall with left bundle branch block). As discussed later in this chapter, regurgitant lesions of either the mitral or aortic valves can lead to substantial ventricular dilatation, and assessment of ventricular size is integral in the evaluation and timing of surgical correction. Because changes in ventricular size are used clinically to determine which patients should undergo valve surgery, accurate assessment of changes in ventricular size is essential. Although serial echocardiography can provide these data, serial assessment by CMR may be more accurate when appreciation of subtle changes over time is important.

Left ventricular wall thickness and mass are also important measures of cardiac and systemic disease. The left ventricle will hypertrophy under any condition in which its afterload is increased, including conditions that obstruct outflow, such as aortic stenosis, hypertrophic cardiomyopathy, and subaortic membranes; in postcardiac aortic obstruction seen in coarctation; or in systemic conditions characterized by increased afterload, such as hypertension. The pattern of ventricular hypertrophy can change depending on the etiology. Aortic stenosis and hypertension are typically characterized by concentric hypertrophy, in which the ventricular walls thicken "concentrically" and cavity size is usually small. In volume overload conditions such as mitral or aortic regurgitation, there may be minimal increase in ventricular wall thickness, but substantial ventricular dilatation leads to marked increases in left ventricular mass.

Ventricular wall thickness can be measured, and ventricular mass can be calculated by either echocardiography or CMR. Although radionuclide imaging and cardiac CT can also provide measures of left ventricular mass, they are not generally used for this purpose. Although measurement of wall thickness with echocardiography is relatively straightforward and accurate, determining left ventricular mass by echocardiography requires using one of several formulas that consider both wall thickness and ventricular cavity dimensions. Assessment of left ventricular mass volumetrically by CMR has the advantage of not requiring geometric assumptions and is thus more accurate than echocardiography.

■ ASSESSMENT OF LEFT VENTRICULAR SYSTOLIC FUNCTION

Assessment of ejection fraction, or the percentage of blood ejected with each beat, has been the primary method to assess systolic function and is generally calculated by subtracting end-systolic volume from end-diastolic volume and dividing by end-diastolic volume. All cardiac imaging modalities can provide direct measurements of left ventricular ejection fraction (LVEF). As discussed above, tomographic techniques (e.g., CMR, CT, and radionuclide imaging) are generally more accurate and reproducible than echocardiography because there are no geometric assumptions and these techniques are not dependent on operator skill, although echocardiography remains the most utilized method. A LVEF of 55% or greater is generally considered normal, and an LVEF of 50–55% is considered in the low-normal range, although these can vary widely, and normal ejection fraction tends to be higher in women than in men.

Newer methods to assess systolic function, such as myocardial strain or deformation imaging using speckle-tracking methods on echocardiography, myocardial tagging, and more recently, feature tracking on CMR, can provide a more sensitive approach to the detection of systolic dysfunction, in part because these measures are geometric assumption-independent. Additional assessments based on these novel methods include assessment of myocardial twist and torsion. Regional strain patterns can differ in various diseases. For example, in subtypes of cardiac amyloidosis, it is common to see a reduction in myocardial strain at the base of the heart with relative sparing of the apex. Strain imaging is being increasingly utilized in conditions such as valvular heart disease and early detection of cardiotoxicity following chemotherapy and/or radiation therapy. In addition to the estimation or calculation of ejection fraction, stroke volume can be assessed by any of the imaging methods by subtracting the end-systolic volume from the end-diastolic volume or by quantifying forward flows using echocardiographic Doppler methods or phase-contrast CMR imaging. They offer an estimated measure of cardiac output other than LVEF.

■ ASSESSMENT OF LEFT VENTRICULAR DIASTOLIC FUNCTION

Echocardiography remains the primary method for clinical assessment of diastolic function, in part because echocardiography has the highest temporal resolution of all the imaging techniques. Recent advances in Doppler tissue imaging allow for accurate assessment of the velocity of myocardial wall motion by assessing the excursion of the mitral annulus in diastole. Mitral annular relaxation velocity, or E′, is inversely related to the time constant of relaxation, tau, and has been shown to have prognostic significance in patients with heart failure. Dividing the standard mitral inflow maximal velocity, E, by the mitral annular relaxation velocity yields E/E′, which has been shown to correlate with left ventricular filling pressures. The utility of standard E and A wave ratios for assessment of diastolic function has been questioned. Mitral deceleration time can be a useful measure if very short (<150 ms), suggesting restrictive physiology and severe diastolic dysfunction. Left atrial volume is considered an integrator of diastolic function, and minimal volume may be more reflective of left ventricular filling pressures than maximal volume. Several grading methods for diastolic function have been proposed that take into account a number of diastolic parameters, including Doppler tissue-based relaxation velocities, pulmonary venous Doppler, and left atrial size **(Fig. 248-8)**. Diastolic function worsens with aging, and most diastolic parameters need to be adjusted for age.

■ ASSESSMENT OF RIGHT VENTRICULAR FUNCTION

Right ventricular size and function have been shown to be prognostically important in a variety of conditions, and can be assessed by echocardiography, CMR, CT, or radionuclide imaging methods. CMR is considered the most accurate noninvasive technique to evaluate the structure and ejection fraction of the right ventricle **(Video 248-2)**. Assessment of the right ventricle by echocardiography has generally been qualitative, owing in part to the unusual geometry of the right ventricle. However, several quantitative methods are available for assessment of right ventricular function, including fractional area change (FAC = [diastolic area − systolic area]/diastolic area), which has been shown to correlate with outcomes in heart failure and after myocardial infarction. Excursion of the tricuspid annulus (tricuspid annular plane systolic excursion [TAPSE]) is another method to assess right ventricular function, although it is mostly used in research settings.

Abnormalities of right ventricular size and function are generally secondary to either diseases that affect the right ventricle intrinsically or diseases in which the right ventricle responds to abnormalities elsewhere in the heart or pulmonary vasculature. Intrinsic diseases that affect the right ventricle include congenital abnormalities, including hypoplastic right ventricle and arrhythmogenic right ventricular dysplasia, and acquired conditions, such as right ventricular infarction and infiltrative diseases. Right ventricular dilatation can occur due to both chronic and acute processes. Long-standing pulmonary hypertension or pulmonary outflow tract obstruction leads to right ventricular hypertrophy and ultimately dilatation. An acute process that can cause profound right ventricular dilatation and dysfunction is acute pulmonary embolism. In the setting of acute occlusion of a pulmonary artery or branch, an acute rise in pulmonary vascular resistance causes a previously normal right ventricle to dilate and fail due to the increased afterload. In acute pulmonary embolism, right ventricular dilatation and dysfunction are signs of substantial hemodynamic compromise and are associated with a marked increased risk of death. In addition to right ventricular dilatation, acute pulmonary embolism is often associated with a specific pattern of regional right ventricular dysfunction, commonly referred to as the McConnell sign, characterized by preservation of right ventricular wall motion in the basal and apical regions and dyskinesis in the region of the mid right ventricular free wall. This abnormality is highly specific for acute pulmonary embolism and is likely secondary to acute increases in right ventricular load.

Any disease that causes increased pulmonary vascular resistance can lead to right ventricular dilatation and dysfunction. For example, long-standing chronic obstructive pulmonary disease increases pulmonary vascular resistance and results in cor pulmonale. Acute pneumonia can cause findings that are similar to acute pulmonary embolism, and right ventricular dysfunction has been a hallmark of severe COVID-19 disease due to either macro- or microthrombosis in the pulmonary vasculature. In patients with right ventricular dilatation without obvious pulmonary disease, intracardiac shunts should be considered. The increased flow through the pulmonary vasculature resulting from an atrial septal or ventricular septal defect can, over time, result in elevation in pulmonary vascular resistance with subsequent dilatation and hypertrophy of the right ventricle. Right ventricular dilatation and dysfunction can be seen in left-sided heart disease, and patients who develop right ventricular dilatation and dysfunction due to predominantly left-sided disease have worse outcomes.

In addition to assessment of left and right ventricular structure and function, assessment of the other cardiac chambers also provides important clues to intracardiac and systemic diseases. Enlargement of the left atrium is common in patients with hypertension and is also suggestive of increased left ventricular filling pressures; indeed, left atrial size is often termed the "hemoglobin A_{1c}" of diastolic function, because left atrial enlargement reflects long-standing increase in left-sided filling pressures. Right atrial dilatation and dilatation of the inferior vena cava are common in conditions in which central venous pressure is elevated.

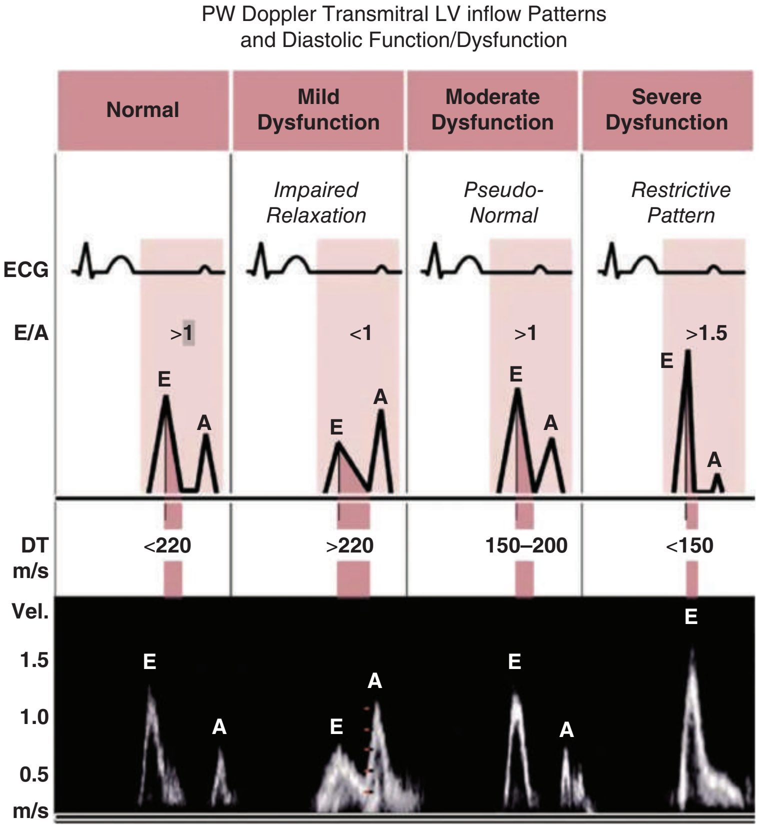

FIGURE 248-8 Mitral inflow Doppler waveforms in diastolic dysfunction. Left ventricle diastolic filling patterns (normal, impaired relaxation, pseudonormal, and restrictive filling). See text for details. ECG, electrocardiogram; LV, left ventricle; PW, pulsed wave. *(Modified from CY Ho, BE Bulwer: Echocardiographic assessment of diastolic function, in Essential Echocardiography. A Practical Handbook with DVD. CY Ho, BE Bulwer (eds). Totowa, NJ: Humana Press; 2007, p 124.)*

PATIENT SAFETY CONSIDERATIONS

RADIATION EXPOSURE

Both cardiac CT and radionuclide imaging expose patients to ionizing radiation. Several recent publications have raised concern regarding the potential harmful effects of ionizing radiation associated with cardiac imaging. The *effective dose* is a measure used to estimate the biologic effects of radiation and is expressed in millisieverts (mSv). However, measuring the radiation effective dose associated with diagnostic imaging is complex and imprecise and often results in varying estimates, even among experts. The effective dose from a typical myocardial perfusion SPECT scan ranges between ~4 and 11 mSv, depending on the protocol and type of scanner used. The effective dose from a typical myocardial perfusion PET scan is lower, ~2.0–4 mSv. Radiation exposure associated with cardiac CT is variable and, as with radionuclide imaging, also depends on the imaging protocol and scanner used. Although historic radiation doses with cardiac CT have been quite high, the introduction of newer technologies (e.g., x-ray tube modulation, prospective ECG gating) has resulted in a significant dose reduction. The current average radiation dose for a coronary CTA ranges from 3 to 15 mSv and, in selected cases, can be as low as 1 mSv. Imaging laboratories follow the ALARA (as low as reasonably achievable) principle when balancing the clinical need and imaging approach. By comparison, the average dose for invasive coronary angiography is ~7 mSv, whereas exposure to radiation from natural sources in the United States amounts to ~3 mSv annually.

The risk of a fatal malignancy from medical imaging–related radiation is difficult to estimate precisely but is likely small and difficult to discern from the background risk of natural malignancies. The small but potential radiation risks from imaging mandate an assessment of the risk-versus-benefit ratio in the individual patient. In this context, one must also consider the risks of missing important diagnostic information by not performing a test (which could potentially influence near-term management and outcomes) for a theoretical concern of a small long-term risk of malignancy. Before ordering any test, especially one associated with ionizing radiation, we must ensure the appropriateness of the study and that the potential benefits outweigh the risks. The likelihood that the study being considered will affect the patient's clinical management should be addressed before testing is performed. It is also important that "routine" follow-up scans in asymptomatic individuals be avoided.

CONTRAST AGENTS

Contrast agents are commonly used in cardiac CT, CMR, and echocardiography. Although their use significantly enhances the diagnostic information of each of these tests, there are also potential risks from the administration of contrast agents that should be considered.

The risk of adverse reactions from iodinated contrast agents used in cardiac CT is well established. The precise pathogenesis of contrast reactions following intravascular administration of iodinated contrast media is not known. The overall incidence of contrast reactions is 0.4–3% with nonionic formulations and higher for ionic formulations. Most contrast adverse reactions are mild and self-limiting. The risk of contrast-induced nephropathy (CIN) in patients with relatively normal renal function (estimated glomerular filtration rate [eGFR] >60 mL/min) is low. In most patients, CIN is self-limited, and renal function usually returns to baseline within 7–10 days, without progressing to chronic renal failure. However, this risk increases in patients with GFR <60 mL/min, especially older diabetic subjects. In such patients, appropriate screening and pre- and postscan hydration are necessary.

The use of gadolinium-based contrast agents (GBCAs) improves CMR imaging by enhancing myocardial tissue contrast or delineating cardiac or vascular structures. All GBCAs are chelated to render the compounds nontoxic and facilitate renal excretion. Mild reactions from GBCAs, such as skin itchiness or redness, occur in ~1% of patients, but severe or anaphylactic reactions are very rare at approximately one in 200,000 patients. Older generation (group I) linear-structured GBCAs have been associated with a rare but serious condition known as nephrogenic systemic fibrosis (NSF), which is an interstitial inflammatory reaction manifested as fibrosis of tissues or internal organs and even death. Risk factors to developing NSF include high-dose use of group I GBCA in presence of severe renal dysfunction (eGFR <30 mL/min per 1.73 m^2), need for hemodialysis, an eGFR <15 mL/min per 1.73 m^2, acute renal deterioration, and concurrent proinflammatory/systemic illnesses. Newer generation (group II) macrocyclic-structured GBCAs have an excellent safety profile across all patients with chronic kidney dysfunction and indeed are now the standard GBCA of choice in most MRI centers. The American College of Radiology considers using group II agents as safe in patients with severe renal dysfunction or on dialysis. For patients with a clinical indication for MRI, checking the serum eGFR level is no longer recommended before the use of group II GBCA. With the widespread use of group II GBCA, routine pretest screening, and weight-based dosing, a near-zero incidence of NSF has been reported in the past decade.

Contrast agents can also be used in echocardiography. Injected agitated saline is used routinely to assess cardiac shunts, because these "bubbles" are too large to traverse the pulmonary circulation. After saline injection, the presence of bubbles in the left side of the heart is indicative of shunt, although the location can sometimes be difficult to determine. The current U.S. Food and Drug Administration (FDA)-approved use of echocardiographic contrast agents is for opacification of left-sided chambers and to improve delineation of left ventricular endocardial border in patients with suboptimal echocardiograms. These agents are either albumin- or lipid-based microspheres filled with inert gases, typically perfluorocarbons. They are considered extremely safe, although they have, in extremely rare instances, been associated with allergic reactions and neurologic events.

SAFETY CONSIDERATIONS OF CMR IN PATIENTS WITH PACEMAKERS AND DEFIBRILLATORS

There are now numerous FDA-approved MRI-conditional internal cardiac defibrillators and pacemakers that are safe to undergo an MRI study up to 3T, under well-defined scanning conditions in experienced centers. For non–FDA-approved cardiac devices (legacy devices), collective evidence has also indicated that MRI studies can be safely performed at 1.5T under normal operational settings, in absence of fractured, epicardial, or abandoned leads, and when an experienced staff is available to interrogate the cardiac device before and after MRI study. The Centers for Medicare and Medicaid Services have approved and expanded coverage of MRI studies in patients with an implanted legacy device since April 2018.

PATIENT-CENTERED APPLICATIONS OF CARDIAC IMAGING

CORONARY ARTERY DISEASE

The basis for the diagnostic application of imaging tests in patients with known or suspected CAD should be viewed considering the pretest probability of disease as well as the specific characteristics of imaging tests (i.e., sensitivity and specificity). In symptomatic patients, the prevalence or pretest probability of CAD differs based on the type of symptom (typical angina, atypical angina, noncardiac chest pain), as well as on age, gender, and coronary risk factors. In an individual patient, the results of the initial test inform the posttest likelihood of CAD. In patients undergoing sequential testing (e.g., ECG treadmill testing followed by stress imaging), the posttest probability of disease after the first test becomes the pretest likelihood of disease for the second test. Regardless of the sequence, the expectation is that a test will provide sufficient information to confirm or exclude the diagnosis of CAD and that such information will allow accurate risk stratification to be able to guide management decisions.

Table 248-3 summarizes the relative diagnostic accuracies of cardiac imaging modalities for the diagnosis of CAD.

It is important to highlight that most studies included in meta-analyses of the diagnostic accuracy of cardiac imaging modalities for the diagnosis of CAD were retrospective, small, single-center studies, comprising predominantly male patients with a high prevalence of CAD (>50–60%). Multicenter studies assessing the performance of individual modalities or comparing different modalities have consistently resulted in more modest diagnostic accuracies, tracking more closely with how these tests perform in clinical practice.

Stress Echocardiography The hallmark of myocardial ischemia during stress echocardiography is the development of new regional wall motion abnormalities and reduced systolic wall thickening **(Video 248-3)**. Stress echocardiography can be performed in conjunction with exercise or dobutamine stress. Stress echocardiography is best at identifying inducible wall motion abnormalities in previously normally contracting segments. In a patient with wall motion abnormalities at rest, the specificity of stress echocardiography is reduced, and worsening regional function of a previously abnormal segment might reflect worsening contractile function in the setting of increased wall stress rather than new evidence of inducible ischemia.

The advantages of stress echocardiography over other stress imaging techniques include its relatively good diagnostic accuracy, widespread

TABLE 248-3 Comparative Diagnostic Accuracy of Cardiac Imaging Approaches to Coronary Artery Disease

IMAGING MODALITY	PUBLISHED DATA	SENSITIVITY	SPECIFICITY
Exercise echocardiography	15 studies (n = 1849 patients)	84%	82%
Dobutamine echocardiography	28 studies (n = 2246 patients)	80%	84%
SPECT MPI	113 studies (n = 11,212 patients)	88%	76%
Myocardial perfusion PET	9 studies (n = 650 patients)	93%	81%
CMR perfusion	37 studies (n = 2841 patients)	91%	81%
CMR wall motion	14 studies (n = 754 patients)	83%	86%
Coronary CTA	18 studies (n = 1286 patients)	99%	89%

Note: In these studies, the diagnosis of coronary artery disease was based on the presence of a >50% or >70% stenosis on invasive coronary angiography.

Abbreviations: CMR, cardiac magnetic resonance; CTA, computed tomography angiography; MPI, myocardial perfusion imaging; PET, positron emission tomography; SPECT, single-photon emission computed tomography.

availability, no use of ionizing radiation, and relatively low cost. Limitations of stress echocardiography include (1) the technical challenges associated with image acquisition at peak exercise because of exertional hyperpnea and cardiac excursion, (2) the fact that rapid recovery of wall motion abnormalities can be seen with mild ischemia (especially with one-vessel disease, which limits sensitivity), (3) difficulty detecting residual ischemia within an infarcted territory because of resting wall motion abnormality, (4) high operator dependence for acquisition of echocardiographic data and analysis of images, and (5) the fact that good-quality complete images viewing all myocardial segments occurs in only 85% of patients. Newer techniques including second harmonic imaging and the use of intravenous contrast agents improve image quality, but their effect on diagnostic accuracy has not been well documented.

As with nuclear perfusion imaging, stress echocardiography is often used for risk stratification in patients with suspected or known CAD. A negative stress echocardiogram is associated with an excellent prognosis, allowing identification of patients at low risk. Conversely, the risk of adverse events increases with the extent and severity of wall motion abnormalities on stress echocardiography.

Stress Radionuclide Imaging SPECT myocardial perfusion imaging is the most common form of stress imaging tests for CAD evaluation. The presence of a reversible myocardial perfusion defect is indicative of ischemia **(Fig. 248-9, *left panel*)**, whereas a fixed perfusion defect generally reflects prior myocardial infarction **(Fig. 248-9, *right panel*)**. As discussed above, PET has advantages compared to SPECT, but it is not widely available and is more expensive and, thus, considered an emerging technology in clinical practice.

Nuclear perfusion imaging is another robust approach to diagnose obstructive CAD, quantify the magnitude of inducible myocardial ischemia, assess the extent of tissue viability, and guide therapeutic management (i.e., selection of patients for revascularization). One of the most valuable clinical applications of radionuclide perfusion imaging is for risk stratification. It is well established that patients with a normal SPECT or PET study exhibit a low rate of major adverse cardiac events of <1% annually. Importantly, the risks of death and myocardial infarction increase linearly with increasing magnitude of perfusion abnormalities, reflecting the extent and severity of CAD.

Despite the widespread use and clinical acceptance of radionuclide imaging in CAD evaluation, a recognized limitation of this approach is that it often uncovers only coronary territories supplied by the most severe stenoses. Consequently, it is relatively insensitive to accurately delineate the extent of obstructive angiographic CAD, especially in the setting of multivessel disease. The use of quantitative myocardial blood flow and coronary flow reserve with PET can help mitigate this limitation. In patients with so-called "balanced" ischemia or diffuse CAD, measurements of coronary flow reserve uncover areas of myocardium at risk that would generally be missed by performing only relative assessments of myocardial perfusion **(Fig. 248-10)**. Conversely, a normal coronary flow reserve is associated with a very high negative predictive value for excluding high-risk angiographic CAD. These measurements of coronary flow reserve also contribute to risk stratification across the spectrum of ischemic changes, including patients with visually normal myocardial perfusion.

HYBRID CT AND NUCLEAR PERFUSION IMAGING Because many of the newer generation nuclear medicine scanners integrate CT and a gamma camera in the same acquisition gantry, it is now possible to acquire and quantify myocardial scar and ischemia and CAC scoring from a single dual-modality study (SPECT/CT or PET/CT) **(Fig. 248-11)**. The rationale for this integrated approach is predicated on the fact that the perfusion imaging approach is designed to uncover only obstructive atherosclerosis. Conversely, CAC scoring provides a quantitative measure of the anatomic extent of atherosclerosis. This provides an opportunity to improve the conventional models for risk assessment using nuclear imaging alone, especially in patients without known CAD.

Cardiac CT Voluminous plaques are more prone to calcification, and stenotic lesions frequently contain large amounts of calcium. Indeed, there is evidence that high CAC scores are generally predictive of a higher likelihood of obstructive CAD, and the available data support the concept of a threshold phenomenon governing this relationship (i.e., Agatston score >400). However, given the fact that CAC scores are not specific markers of obstructive CAD, one should be cautious in using this information as the basis for referral of patients to coronary angiography, especially in symptomatic patients with low-risk stress tests. Conversely, CAC scores <400, especially in symptomatic patients with intermediate-high likelihood of CAD, as in those with typical angina, may be less effective in excluding CAD, especially in young symptomatic men and women who may have primarily noncalcified atherosclerosis **(Fig. 248-12)**.

As discussed above, the improved temporal and spatial resolution of modern multidetector CT scanners offers a unique noninvasive approach to delineate the extent and severity of coronary

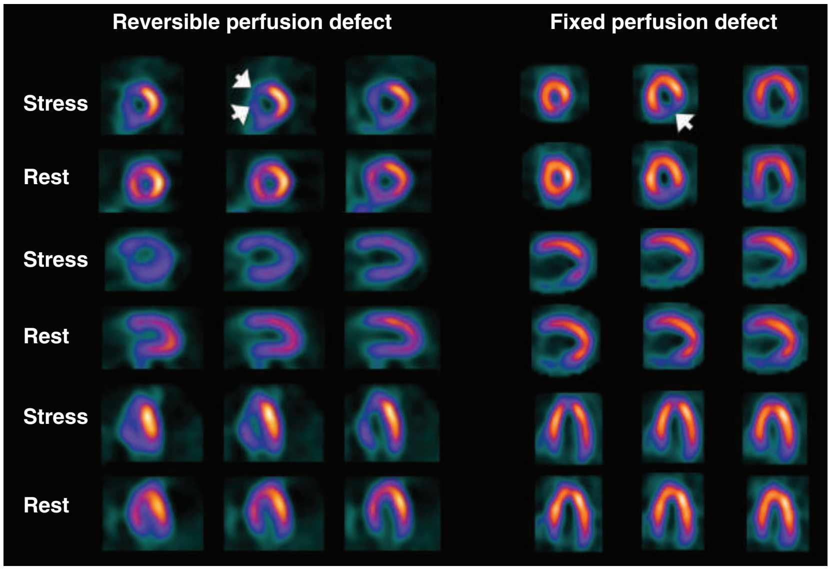

FIGURE 248-9 Selected technetium-99m sestamibi myocardial perfusion single-photon emission computed tomography images of two different patients demonstrating a reversible perfusion defect involving the anterior and septal left ventricular wall, reflecting ischemia in the left anterior descending coronary territory (*arrows in left panel*), and a fixed perfusion defect involving the inferior and inferolateral walls consistent with myocardial scar in the right coronary territory (*arrow in right panel*).

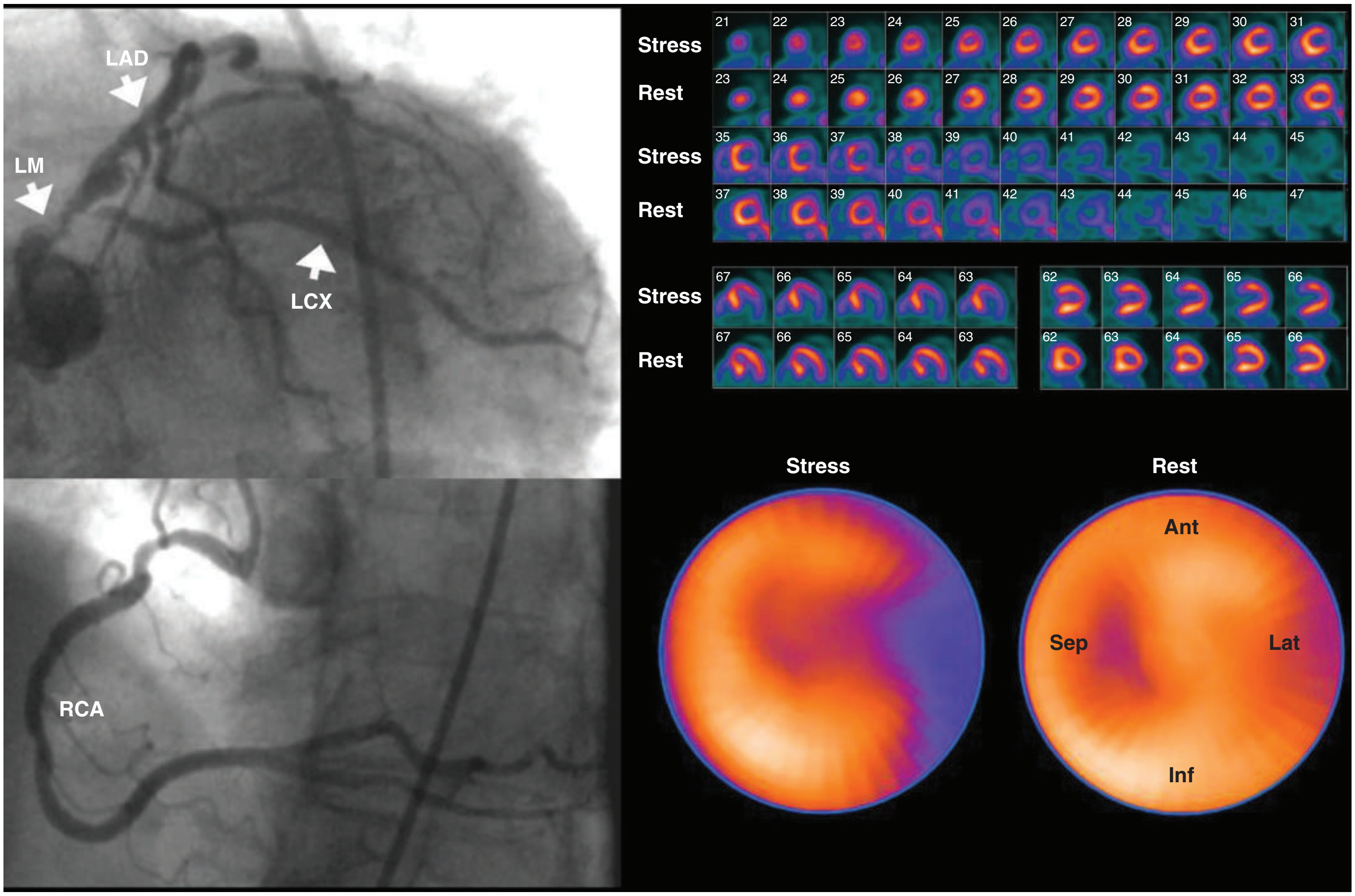
LAD
LM
LCX
RCA
Stress
Rest
Stress
Rest
Stress
Rest
Stress
Rest
Ant
Sep
Lat
Inf

FIGURE 248-10 Coronary angiographic (*left panel*) and rubidium-82 myocardial perfusion positron emission tomography images (*right panel*) in an 85-year-old female with diabetes presenting with chest pain. The coronary angiogram demonstrates significant stenoses of the left main and circumflex coronary arteries. However, the perfusion images demonstrate only a reversible lateral wall defect. Quantification of stress and rest myocardial blood flow demonstrated a significant, global reduction on coronary flow reserve (estimated at 1.2, normal value >2.0), reflecting extensive myocardium risk that was underestimated by the semiquantitative estimates of myocardial perfusion. LAD, left anterior descending artery; LCX, left circumflex artery; LM, left main artery; RCA, right coronary artery.

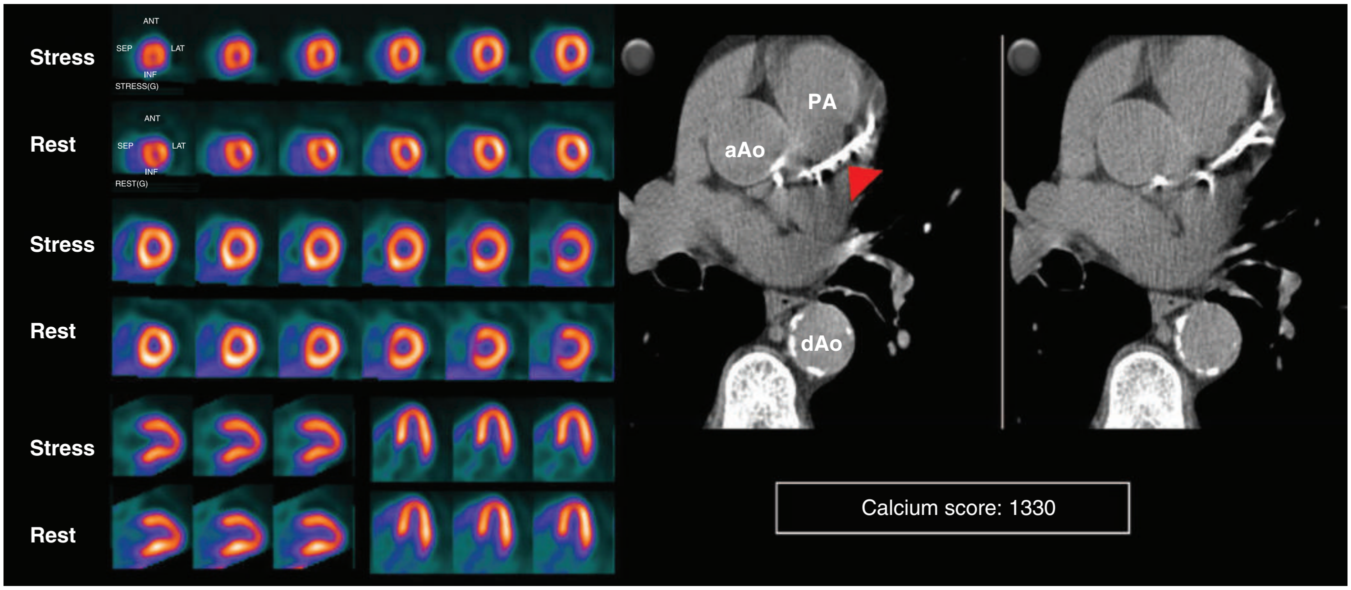
Stress
Rest
Stress
Rest
Stress
Rest
ANT
SEP
LAT
INF
STRESS(G)
REST(G)
PA
aAo
dAo
Calcium score: 1330

FIGURE 248-11 Stress and rest rubidium-82 myocardial perfusion positron emission tomography (PET) images (*left*) and noncontrast gated computed tomography (CT) images (*right*) delineating the extent and severity of coronary artery calcifications obtained with integrated PET/CT imaging. The images demonstrate extensive atherosclerosis (Agatston coronary calcium score = 1330) without flow-limiting disease based on the normal perfusion study. aAo, ascending aorta; dAo, descending aorta; PA, pulmonary artery.

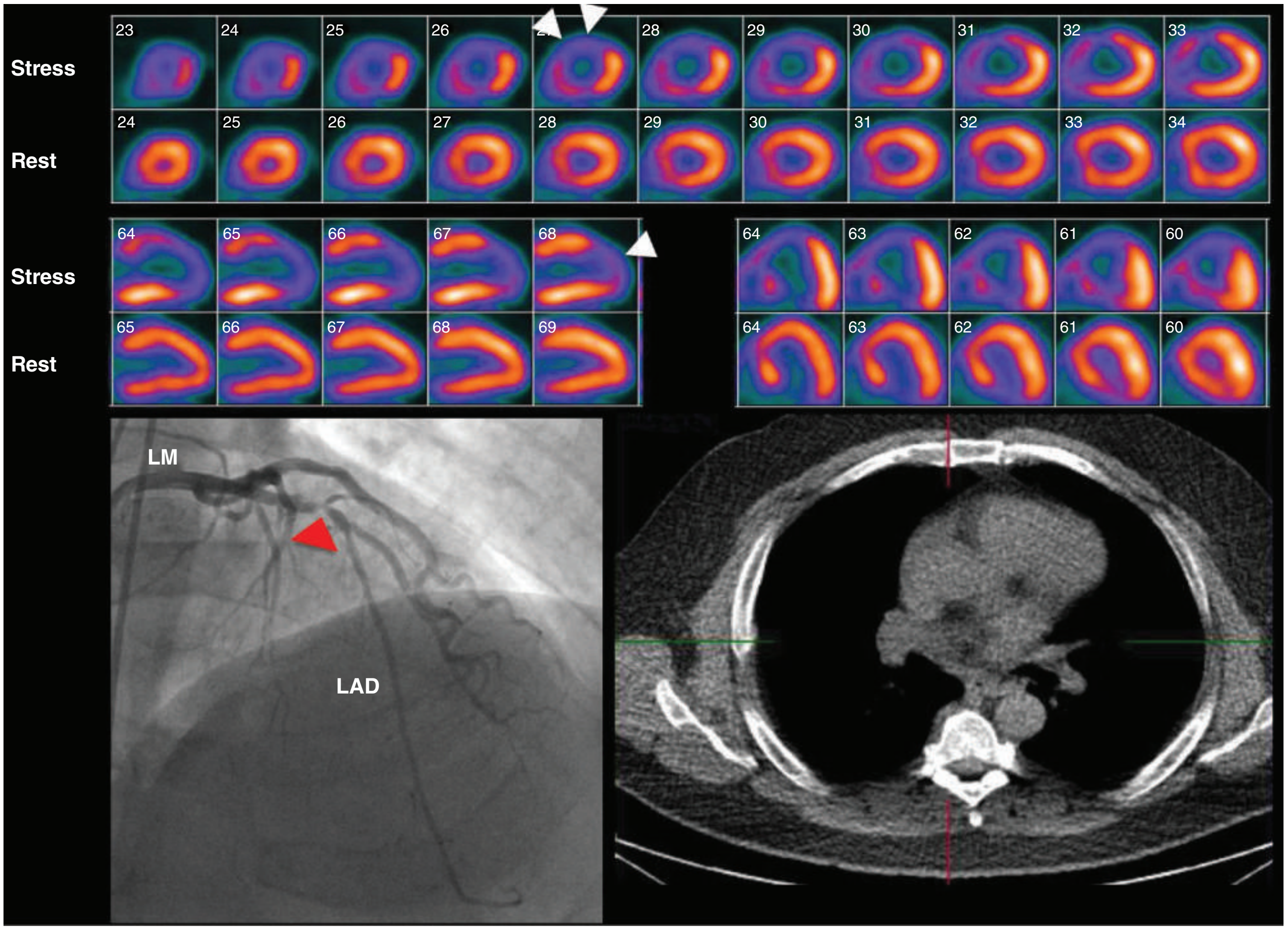

FIGURE 248-12 Stress and rest rubidium-82 myocardial perfusion positron emission tomography images (*top*), noncontrast gated computed tomography images (*lower right*), and selected coronary angiographic images obtained on a 59-year-old male patient with atypical angina. Despite the absence of significant coronary calcifications (Agatston calcium score = 0), the perfusion images demonstrated a dense and reversible perfusion defect involving the anterior and anteroseptal walls (*arrows*), reflecting significant obstructive disease in the left anterior descending coronary artery (LAD), confirmed on angiography. LM, left main artery.

atherosclerosis with coronary CTA. The extremely high sensitivity of this approach offers a very effective means for excluding the presence of CAD (high negative predictive value) (Table 248-3). In the setting of high coronary calcium scores (e.g., >400), however, specificity is reduced because the blooming artifact of calcium does not allow one to evaluate the vessel lumen accurately. Given the high negative predictive value of CTA, a normal scan result effectively excludes obstructive CAD and abolishes the need for further investigation. As discussed below, this may be quite useful in patients with low-intermediate clinical risk presenting to the emergency room for chest pain. However, the limited capability of this technique to determine which coronary plaques are flow limiting can make abnormal scan results more difficult to interpret, especially in terms of the possible need of revascularization. There are emerging data suggesting that by adding a stress myocardial perfusion CT evaluation (similar to stress perfusion CMR) **(Fig. 248-13, *top panel*)** or an estimated fractional flow reserve (so-called FFR_{CT}) **(Fig. 248-13, *lower panel*)**, one can define the hemodynamic significance of anatomic stenosis. FFR_{CT} is beginning to enter routine clinical practice. However, CT myocardial perfusion remains an emerging technology.

As with invasive coronary angiography, assessments of the extent of CAD by CTA can also provide useful prognostic information. A low 1-year cardiac event rate has been reported for patients without coronary atherosclerosis on CTA. For patients with obstructive CAD, the risk of adverse cardiac events increases proportionally with the extent of angiographically obstructive CAD. There is new evidence that even the presence of nonobstructive atherosclerosis increases the risk of adverse cardiac events.

Although CTA can be helpful in assessing patency of bypass grafts, the assessment of stents is somewhat more challenging because the limited spatial resolution of CT and stent diameter (<3 mm being associated with the highest number of partial lumen visualization and nondiagnostic scans) both contribute to limited clinical results.

CMR Imaging CMR evaluates for ischemia from CAD by assessing regional myocardial perfusion or regional wall motion at rest and during pharmacologic stress with an intravenous infusion of vasodilator agent or dobutamine. Myocardial perfusion is evaluated by injecting a GBCA bolus followed by imaging data acquisition as the contrast passes through the cardiac chambers and into the myocardium. Relative perfusion deficits are recognized as regions of low signal intensity (black) within the myocardium **(Video 248-4)**. Several minutes after GBCA injection, LGE imaging allows detection of bright areas of myocardial scar (white), which permits comparison of regions of hypoperfusion and infarction to quantify myocardial ischemia **(Fig. 248-14)**.

With better delineation of the endocardial borders, dobutamine CMR has better diagnostic accuracy than dobutamine echocardiography for detection of CAD, especially in patients with poor acoustic window (Table 248-3). High-dose dobutamine carries the risk of serious ventricular arrhythmias (~1%), but most cases can be prevented with proper monitoring of vital signs and regional cine function. The advantages of stress perfusion CMR over SPECT include its higher spatial resolution, which allows detection of subendocardial ischemia or infarction that may be missed by SPECT. As with other imaging modalities, stress CMR studies also provides robust risk stratification in practice. In a recent randomized controlled trial, a stress CMR-guided strategy was shown to improve the guidance toward the use of invasive investigation and coronary revascularization.

Selecting a Testing Strategy in Patients without Known CAD

As discussed above, there are many options for the evaluation of a patient with suspected CAD presenting with chest pain symptoms. The critical questions to be answered by a testing strategy include the

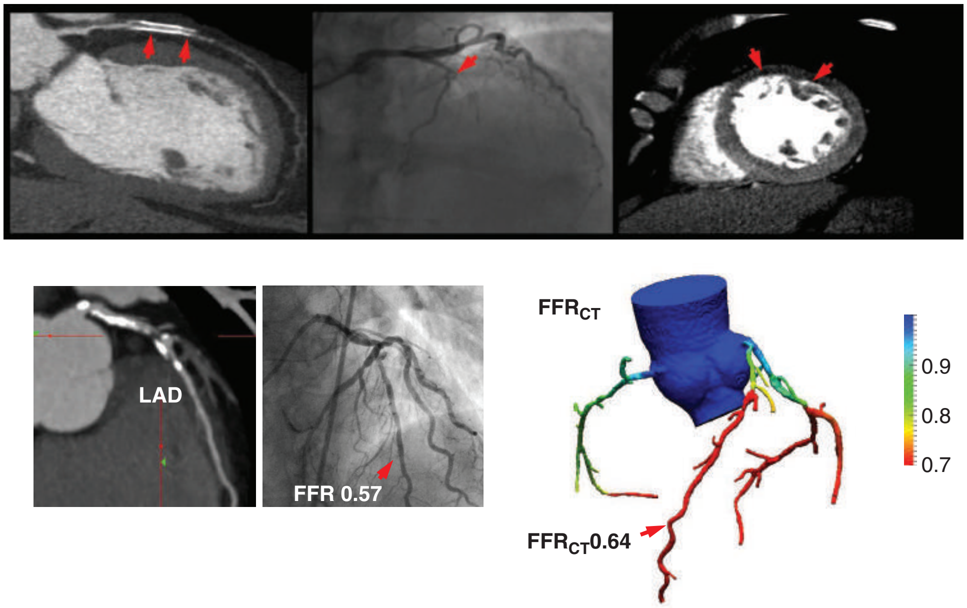

FIGURE 248-13 Examples of novel approaches to the assessment of flow-limiting coronary artery disease (CAD) with cardiac computed tomography (CT). In the *top panel*, representative views of a coronary CT angiogram (CTA; *left*), coronary angiogram (*middle*), and stress myocardial perfusion CT (*right*) images in a patient with CAD and prior stenting of the left anterior descending coronary artery (LAD) are presented. On the CTA, the stent (*arrows*) is totally occluded as evidenced by the loss of contrast enhancement distal to the stent. The coronary angiogram demonstrates a concordant total occlusion of the LAD. On the perfusion CT images, there is a black rim (*arrows*) involving the anterior and anterolateral walls, indicating the lack of contrast opacification during stress consistent with myocardial ischemia. *(Images courtesy of CORE 320 investigators.)* The *lower panel* illustrates an example of fractional flow reserve (FFR) estimates with coronary CTA (*left*) compared to the reference standard of invasive FFR. The FFR reflects the pressure differential between a coronary segment distal to a stenosis and the aorta. In normal coronary arteries, there is no gradient, and FFR is 1. An FFR <0.80 is consistent with a hemodynamically significant stenosis. *(Images courtesy of Dr. James Min, Cornell University, New York.)*

following: (1) Does the chest pain reflect obstructive CAD? (2) What are the short- and long-term risks? (3) Does the patient need to be considered for revascularization? Large-scale clinical trials have demonstrated the benefits of guideline-directed medical therapies where the majority of the patients with stable CAD are at low risk of serious cardiac events on guideline-directed medical therapy alone, and coronary revascularization is best reserved for patients with severe symptoms despite adequate medical therapy. However, imaging will continue to play a significant role in diagnosing the etiology of chest pain and risk assessment of individual patients.

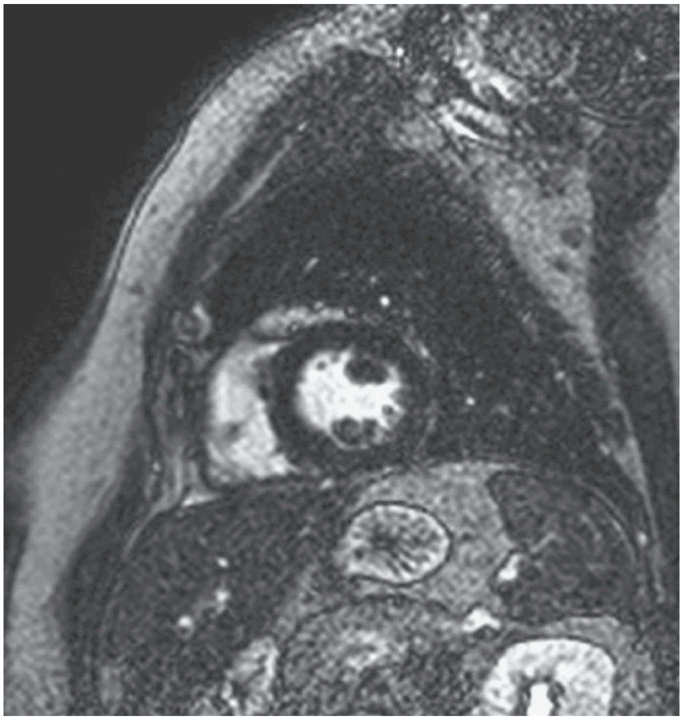

FIGURE 248-14 The image shows the late gadolinium enhancement image of a mid short-axis view. There is no evidence of infarction in the anterior wall, which would be seen as bright white areas, indicating that the stress perfusion defect primarily represents myocardial ischemia. This patient had a significant stenosis of the left anterior descending coronary artery.

For symptomatic patients without a prior history of CAD, the 2021 American College of Cardiology/American Heart Association chest pain guidelines assigned a class I recommendation for stress imaging testing and coronary CTA. Most common stress imaging strategies include stress echocardiography and radionuclide imaging. However, CMR and PET have both been shown to have higher test accuracy than SPECT. In patients with intermediate clinical risk, stress imaging with either SPECT or echocardiography has been shown to reclassify patients accurately who are initially classified as intermediate risk by exercise treadmill testing (ETT) as low or high risk. ETT as the initial testing strategy is still valuable for patients with a normal or nearly normal resting ECG who can exercise.

The use of exercise testing in women presents difficulties that are not seen in men, reflecting the differences in the lower prevalence of obstructive CAD in women and the different accuracy of exercise testing in men and women. Compared with men, the lower pretest probability of disease in women means that more test results are false positive. In some of these patients, a positive ETT may reflect true myocardial ischemia caused by microvascular coronary artery dysfunction (so-called *microvascular disease*). In addition, the inability of many women to exercise to maximum aerobic capacity, the greater prevalence of mitral valve prolapse and microvascular disease, and possibly other reasons may contribute to the differences with men as well. The difficulties of using exercise testing for diagnosing obstructive CAD in women have led to speculation that stress imaging may be preferred over standard stress testing. However, recent data from the WOMEN study suggest that in symptomatic, low-risk women who can exercise, standard ETT is a very effective initial diagnostic strategy as compared to stress radionuclide imaging. Indeed, the 2-year outcomes were similar in both diagnostic strategies, and the ETT-first approach resulted in 48% lower costs compared to exercise radionuclide imaging.

An imaging strategy is the recommended first step for patients who are unable to exercise to an adequate workload and/or those with abnormal resting ECGs (e.g., left ventricular hypertrophy with strain, left bundle branch block). Importantly, the most recent documents regarding appropriate use of imaging also considered that an imaging strategy may be an appropriate first step in patients with intermediate-high likelihood of CAD (e.g., diabetics, renal impairment) due to increased overall sensitivity for diagnosis of CAD and improved risk stratification. In considering an imaging strategy, the evidence supporting the role of ischemia assessment versus anatomy must be considered. From the discussion above, for patients with atypical chest pain and a low-pretest risk of CAD, a normal coronary CTA is helpful because it effectively excludes the presence of obstructive CAD and the need for further testing, defines a low clinical risk, and makes management decisions regarding referral to coronary angiography straightforward. However, in patients with an intermediate or higher pretest risk, coronary CTA is less effective in excluding obstructive CAD because of its limited accuracy to define stenosis severity and predict ischemia; however, abnormal CTA results are more problematic to interpret and to use as the basis for defining the potential need of invasive coronary angiography and revascularization. In such patients, a follow-up stress test is usually required to determine the possible need of revascularization **(Fig. 248-15)**.

The justification of stress imaging in testing strategies has hinged on the identification of which patients may benefit from a revascularization strategy by means of noninvasive estimates of jeopardized myocardium rather than angiography-derived anatomic stenoses. However, recent evidence from the ISCHEMIA trial suggests that optimal medical therapy provides comparable prognostic benefit to an early invasive strategy in patients with moderate or severe myocardial ischemia. Coronary revascularization is reserved for patients with very extensive evidence of ischemia by stress imaging and/or inadequate symptom control by optimal medical therapy. While the available data suggest similar diagnostic accuracy for SPECT and echocardiography but higher accuracy for PET and CMR, the choice of strategy depends on availability and local expertise.

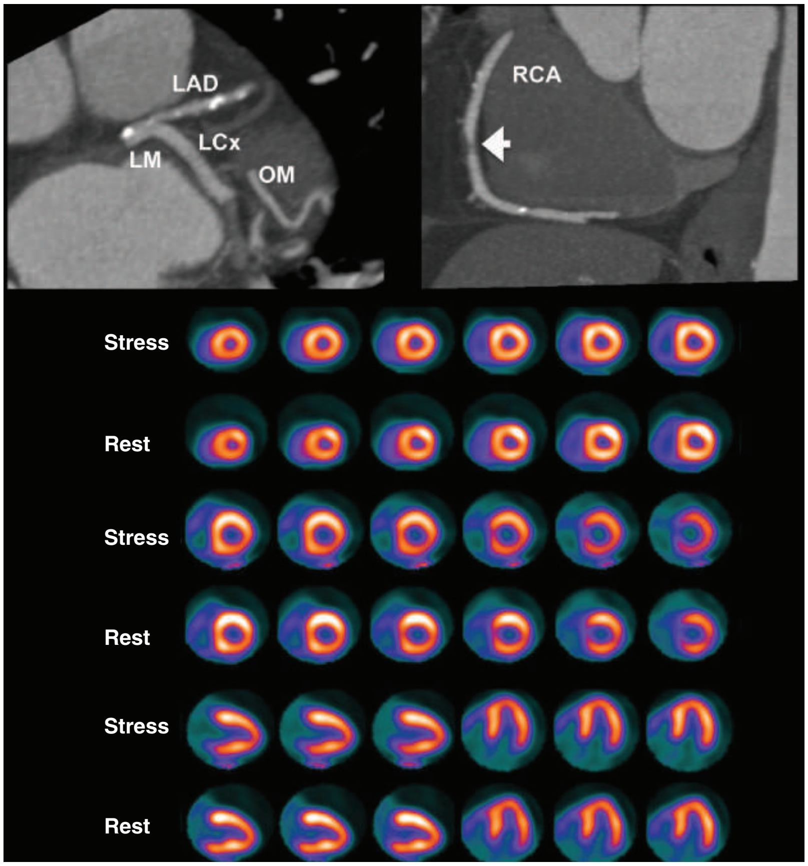

FIGURE 248-15 Selected views from coronary computed tomography angiographic (CTA) images (*top panel*) and stress and rest rubidium-82 myocardial perfusion positron emission tomography images (*lower panel*) obtained on a 64-year-old male patient with atypical angina. The CTA images demonstrate dense focal calcifications in the left main (LM) and left anterior descending (LAD) coronary arteries and a significant noncalcified plaque in the mid right coronary artery (RCA; *arrow*). The myocardial perfusion images demonstrated no evidence of flow-limiting stenosis. LCx, left circumflex artery; OM, obtuse marginal branch.

Selecting a Testing Strategy in Patients with Known CAD Use and selection of testing strategies in symptomatic patients with established CAD (i.e., prior angiography, prior myocardial infarction, prior revascularization) differ from those in patients without prior CAD. Although standard ETT may help distinguish cardiac from noncardiac chest pain, exercise ECG has several limitations following myocardial infarction and revascularization (especially coronary artery bypass grafting). These patients frequently have rest ECG abnormalities. In addition, there is a clinical need to document both the magnitude and localization of ischemia to be able to direct therapy, especially the potential need for targeted revascularization. Consequently, imaging tests are preferred for evaluating patients with known CAD.

There are also important differences in the effectiveness of imaging tests in these patients. As discussed above, coronary CTA is limited in patients with prior revascularization. While CTA provides excellent visualization of the bypass grafts, the native circulation tends to get heavily calcified and is generally not a good target for imaging with CTA. Likewise, blooming artifacts from metallic stents also limit the application of coronary CTA in patients with prior percutaneous coronary intervention. If an anatomic strategy is indicated, direct referral to invasive angiography is preferred.

Stress imaging approaches are especially useful and preferred in symptomatic patients with established CAD. As in patients without prior CAD, normal imaging studies in symptomatic patients with established CAD also identify a low-risk cohort. In those with abnormal stress imaging studies, the degree of abnormality relates to posttest risk. In addition, stress imaging approaches can localize and quantify the magnitude of ischemia, thereby assisting in planning targeted revascularization procedures. As in patients without prior CAD, the choice of stress imaging strategy depends on availability and local expertise.

Testing Strategy Considerations in Patients Presenting with Chest Pain to the Emergency Department Although acute chest pain is a frequent reason for patient visits to the emergency department (ED), only a small minority of those presentations represent an acute coronary syndrome (ACS). Strategies used in the evaluation of these patients include novel cardiac biomarkers (e.g., serum troponins), conventional stress testing (ETT), and noninvasive cardiac imaging. It is generally accepted that the primary goal of this evaluation is exclusion of ACS and other serious conditions rather than detection of CAD.

The routine evaluation of acute chest pain in most centers in the United States includes admission to a chest pain unit to rule out ACS with the use of serial ECGs and cardiac biomarkers. In selected patients, stress testing with or without imaging may be used for further risk stratification. Stress echocardiography and radionuclide imaging are among the most frequently used imaging approaches in these patients. Multiparametric CMR imaging has also been used successfully in patients with acute chest pain **(Video 248-5)**. Due to its ability to probe multiple aspects of myocardial physiology, cardiac anatomy, and tissue characterization with LGE imaging, CMR is useful in diagnosing conditions that mimic ACS (e.g., acute myocarditis, takotsubo cardiomyopathy, pericarditis) **(Fig. 248-16)**.

As discussed above, coronary CTA is a rapid and accurate imaging technique to exclude the presence of CAD and is well suited for the evaluation of patients with acute chest pain **(Fig. 248-17)**. Four randomized clinical trials have demonstrated the feasibility, safety, and accuracy of coronary CTA in the ED as compared to usual care (which typically includes stress imaging). Patients in these trials had a very low clinical

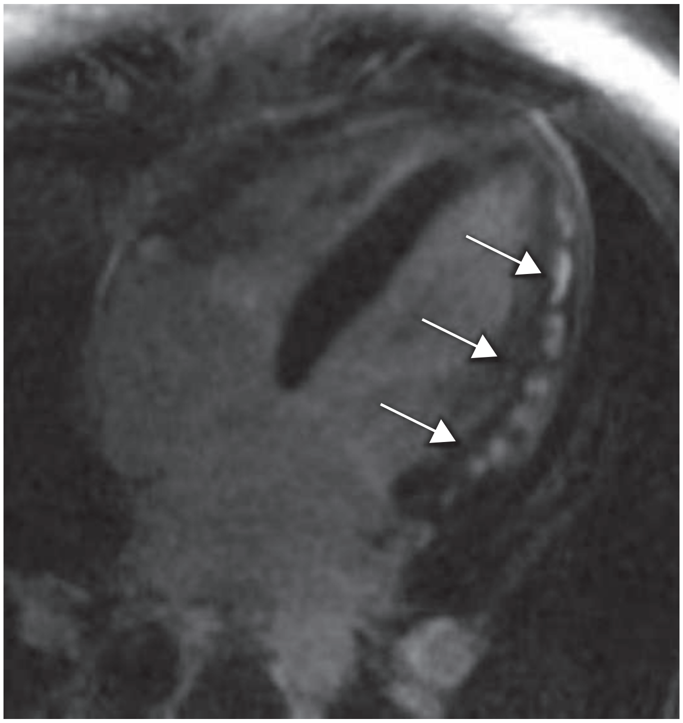

FIGURE 248-16 A four-chamber long-axis late gadolinium enhancement (LGE) image of a patient with acute myocarditis. Note that the LGE primarily involved the epicardial aspect of the myocardium (*arrows*), sparing the endocardium, which is a feature that distinguishes myocarditis from myocardial infarction, which affects the endocardium. Also note the multiple foci of LGE in this case affecting the lateral wall of the left ventricle. Viral myocarditis often presents with this pattern.

risk. Overall, there were no deaths and very few myocardial infarctions without differences between the groups. Likewise, there were no differences in postdischarge ED visits or rehospitalizations. These studies showed decreased length of stay with coronary CTA, and most but not all reported cost savings. An observation from a recent meta-analysis was that, compared to usual care, more patients assigned to coronary CTA underwent cardiac catheterization (6.3 vs 8.4%, respectively) and revascularization (2.6 vs 4.6%, respectively). The relative increased frequency in the referral to cardiac catheterization and revascularization after coronary CTA compared to stress imaging testing strategies has also been observed in patients with stable chest pain syndromes.

Taken together, the available data clearly suggest that not all patients presenting with acute chest pain require specialized imaging testing. Patients with very low clinical risk and negative biomarkers (especially high-sensitivity troponin assays) can be safely triaged. The use of imaging tests in patients with low-intermediate risk should be carefully considered, especially given the trade-offs discussed above.

VALVULAR HEART DISEASE

Abnormalities of any of the four valvular structures in the heart can lead to significant cardiac dysfunction, heart failure, or even death. Echocardiography, CMR, and cardiac CT can be used for the evaluation of valvular heart disease, although echocardiography is generally considered the first imaging test for the assessment of valvular heart disease. In addition, echocardiography is the most cost-effective screening method for valvular heart disease. In some cases, CMR can complement echocardiography when echocardiographic acoustic window is inadequate, by quantifying blood flow data more precisely or providing complimentary assessment of adjacent vascular structures relevant to the valvular condition.

Echocardiography can be used to assess both regurgitant and stenotic lesions of any of the cardiac valves. Typical indications for echocardiography to assess valvular heart disease include cardiac murmurs identified on physical examination, symptoms of breathlessness that may represent valvular heart disease, syncope or presyncope, and preoperative exams in patients undergoing bypass surgery. A standard echocardiographic examination should include qualitative and quantitative assessment of all valves regardless of indication and should serve as an adequate screening test for significant valvular disease.

Assessment of Aortic Stenosis Aortic stenosis, one of the most common forms of valvular heart disease, most often occurs because of gradual progression of valvular calcification in both normal and congenitally abnormal valves. Assessment of aortic stenosis is most commonly performed with echocardiography, although techniques for quantitative assessment of aortic stenosis with CMR have been developed and increasingly used over the past decade. Echocardiographic assessment generally begins with visual inspection of the valve. This allows for assessment of valvular morphology, whether it is tricuspid, bicuspid, or some variant; degree of leaflet calcification; and leaflet excursion.

The normal aortic valve consists of three leaflets or cusps: the right coronary, the left coronary, and the noncoronary cusps. Abnormalities of cusp development are some of the most common congenital heart anomalies, the most common of which is bicuspid aortic valve, with two opening leaflets rather than three **(Fig. 248-18)**. The aortic valve can be visualized on echocardiography, although sometimes it can be difficult to distinguish true bicuspid aortic valve from variants, including the presence of a vestigial commissure (raphe). Bicuspid aortic valve, one of the most common congenital anomalies, predisposes to both aortic stenosis and aortic insufficiency.

The degree of aortic stenosis is assessed by estimating both the pressure gradient across the valve and the valve area. Patients with moderate aortic stenosis or higher generally have peak instantaneous velocities of 3.0 m/s and higher, and often higher than 4.0 m/s, corresponding to pressure gradients of 36 and 64 mmHg, respectively. Because pressure gradients across the aortic valve can be underestimated in patients with severe left ventricular dysfunction, estimation of

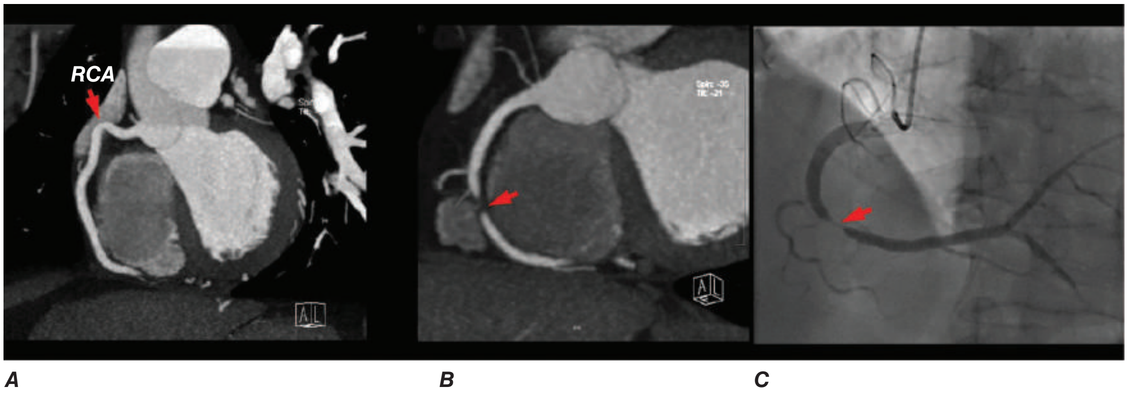

FIGURE 248-17 Representative coronary computed tomography angiographic (CTA) images of two patients presenting to the emergency department with chest pain and negative biomarkers. The patient in ***A*** had angiographically normal coronary arteries; the panel shows a representative view of the right coronary artery (RCA). ***B*** and ***C*** show a corresponding significant stenosis in the mid portion of the RCA on both the CTA (***B***) and invasive angiographic view (***C***). *(Images used with permission from Dr. Quynh Truong, Massachusetts General Hospital, Boston, MA.)*

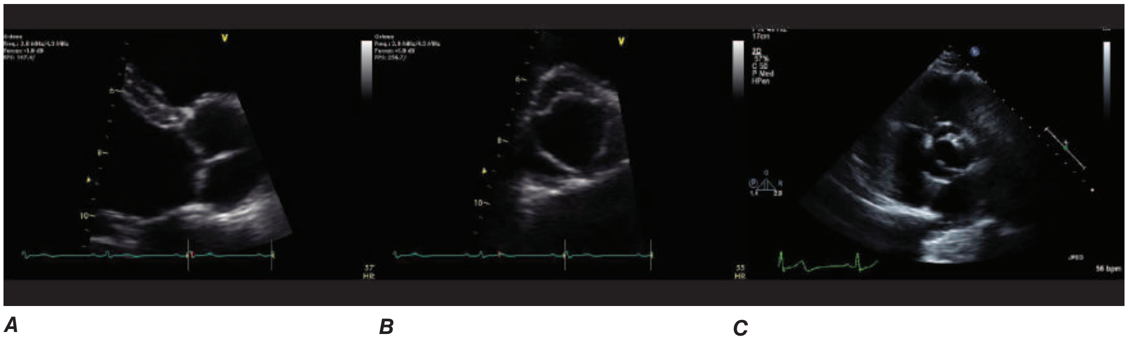

FIGURE 248-18 Normal aortic valve in the parasternal long-axis view (***A***) and short-axis view (***B***), and bicuspid aortic valve showing typical 10 o'clock to 4 o'clock leaflet orientation (***C***).

valve area by the continuity principle is the most accurate technique for assessing the severity of the stenosis. However, evaluation of the patient with so-called low-flow or low-gradient aortic stenosis can be challenging and sometimes require provocative testing such as dobutamine echocardiography. In these cases, it is important to distinguish whether the valve is indeed capable of opening further or simply behaving like a stenotic valve because of the low-pressure gradient.

Aortic valve areas <1.0 cm^2 are generally considered severe, and valve areas <0.6 cm^2 are considered critical. Because patients with good left ventricular function can often tolerate severe aortic stenosis for a considerable period of time, valve areas or gradients alone should not be used to determine whether an individual patient should undergo aortic valve surgery, as this remains a clinical decision.

Some patients with apparent aortic stenosis have subvalvular or even supravalvular obstruction. Hypertrophic cardiomyopathy represents the classic form of subvalvular aortic stenosis, but this is usually easily distinguished from aortic stenosis on echocardiography as the valve leaflets can be seen opening during systole. Subaortic membranes can behave very similarly to leaflet aortic stenosis, and the membranes themselves can be very thin and difficult to visualize, although the presence of a murmur, a gradient across the valve with aortic leaflets that appear to open normally, is highly suggestive of a membrane. Supravalvular aortic stenosis, although exceedingly rare, also occurs.

The emergence of transcatheter aortic valve intervention as a therapeutic option for patients with severe aortic stenosis who are not optimal candidates for surgical replacement has resulted in a very important clinical role for multimodality imaging. Imaging plays a critical role in preprocedural planning, intraprocedural implantation optimization, and follow-up of these patients. CT plays an important role in defining the eligibility of the proposed access site (CTA of the aorta and iliac arteries) and in defining the anatomic relationships between the aortic valve and aortic root, left ventricle, and coronary ostia. Cardiac CT and transesophageal echocardiography are also used to define the device size. Transesophageal echocardiography is used during the device implantation to ensure the best prosthesis–patient match, to assess prosthesis position and function after deployment, and to identify immediate complications (e.g., aortic insufficiency, paravalvular leak resulting from patient–prosthesis mismatch). Echocardiography is the imaging modality of choice for long-term surveillance.

Assessment of Aortic Regurgitation Assessment of aortic regurgitation requires qualitative assessment of the aortic valve structure. Aortic regurgitation is common with congenital abnormalities of the aortic valve, the most common of which is bicuspid aortic valve. Aortic regurgitation often coexists with aortic stenosis, and it is not uncommon for patients to have both severe aortic stenosis and regurgitation. Congenital abnormalities of the aortic leaflets, such as bicuspid aortic valve, are common causes of aortic insufficiency. Dilatation of the aortic root, as occurs in patients with hypertension and other disorders in which aortic dilatation can occur, can also lead to aortic regurgitation even when the valve leaflets are intrinsically normal due to malcoaptation of the leaflets. Aortic root dilatation is common in patients with aortic regurgitation, both as a cause or coexisting lesion, and the aortic root and ascending aorta should be measured and followed in these patients **(Fig. 248-19)**.

Because aortic regurgitation can result in dilatation of the left ventricle over time with ultimate reduction in ventricular function, caring for the patient with aortic regurgitation requires serial assessment of ventricular size and function. Patients whose ventricles dilate beyond an end-systolic diameter of 5.5 cm or whose LVEF declines below normal are at significantly higher risk of death or heart failure, and these measures are often used to decide the need for valve surgery. Quantitation of regurgitation itself can be performed using a number of methods. Semiquantitative visual assessment of aortic regurgitant jet width and depth by color flow Doppler remains the most used. The jet diameter as a ratio of the left ventricular outflow tract diameter proximal to the valve represents one of the most reliable indices of severity and correlates well with angiographic assessment. Similarly, the vena contracta, which represents the smallest diameter of the regurgitant flow at the level of the valve, can be used to assess the severity of aortic regurgitation. Other Doppler-based methods include assessing the pressure half-time, or rate of decline of the pressure gradient between the aorta and left ventricle, a measure of acuity of aortic regurgitation, and assessing aortic flow reversal in the descending aorta.

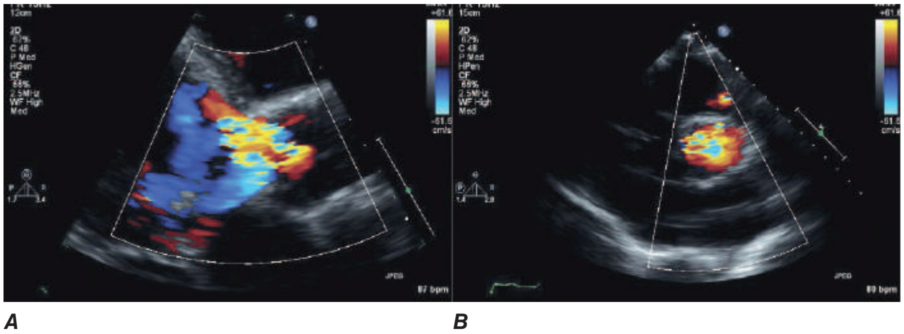

FIGURE 248-19 Aortic regurgitation visualized by color flow Doppler in the parasternal long-axis view (***A***) and the parasternal short-axis view (***B***).

The regurgitant volume can be calculated by comparing the flow across the aortic and pulmonic valves, assuming the pulmonic valve is competent. Central or perivalvular aortic regurgitation is common in patients following transcatheter aortic valve replacement and is generally assessed immediately after the procedure and on follow-up echocardiography.

CMR offers several advantages over echocardiography in the assessment of aortic regurgitation. CMR is more accurate than echocardiography for assessing small changes in cardiac size or function longitudinally in patients with aortic insufficiency. In addition, CMR can accurately quantify aortic regurgitant volume secondary to aortic insufficiency better than echocardiography. CMR can also capture aortic size in 3D that may be helpful in determining the etiology of the aortic regurgitation or in monitoring progression of the condition **(Fig. 248-20 and Video 248-6)**. For this reason, CMR is often used for serial follow-up of patients with aortic regurgitation.

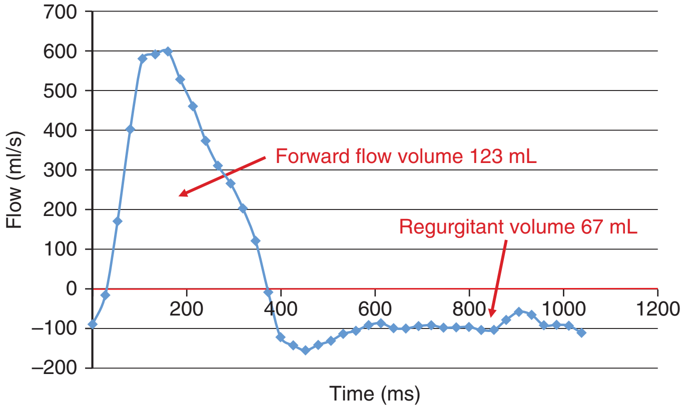

FIGURE 248-20 The resultant flow curve generated from phase contrast imaging demonstrates a forward flow of 123 mL and a regurgitant volume of 67 mL, yielding a regurgitant fraction of 54% indicating severe aortic regurgitation.

Assessment of Mitral Regurgitation The normal mitral valve consists of an anterior and posterior leaflet in a saddle shape configuration **(Fig. 248-21)**. The leaflets are attached to the papillary muscles via chordae tendineae that insert on the ventricular side of the leaflets. Mitral regurgitation can occur due to abnormalities of the leaflets, the chordal structures, or the ventricle, or any combination of these **(Fig. 248-22)**.

Mitral valve prolapse, in which one leaflet moves behind the plane of the other leaflet, can be due to myxomatous degeneration of the valves and leaflet redundancy, disruption of chordal structures secondary to degenerative disease, or papillary muscle rupture or dysfunction following myocardial infarction. Regurgitant jets can be visualized using color flow Doppler. The velocity of regurgitant jets is driven by the pressure gradient between the two chambers. This velocity will thus be quite high for left-sided regurgitant lesions, including mitral regurgitation and aortic regurgitation, resulting in turbulent jets on color flow Doppler (Fig. 248-22). Visual estimation of color flow Doppler is generally sufficient for qualitative assessment of regurgitant severity but can dramatically under- or overestimate regurgitation severity, particularly when regurgitant jets are quite eccentric. For this reason, quantitative assessment is generally recommended, especially when making clinical decisions about surgical intervention. The proximal isovelocity surface area (PISA) method is generally used for quantitative assessment of severity of mitral regurgitation. This method relies on estimation of the velocity of flow acceleration at a specific distance proximal to the valve with the assumption that the flow accelerates in concentric hemispheres.

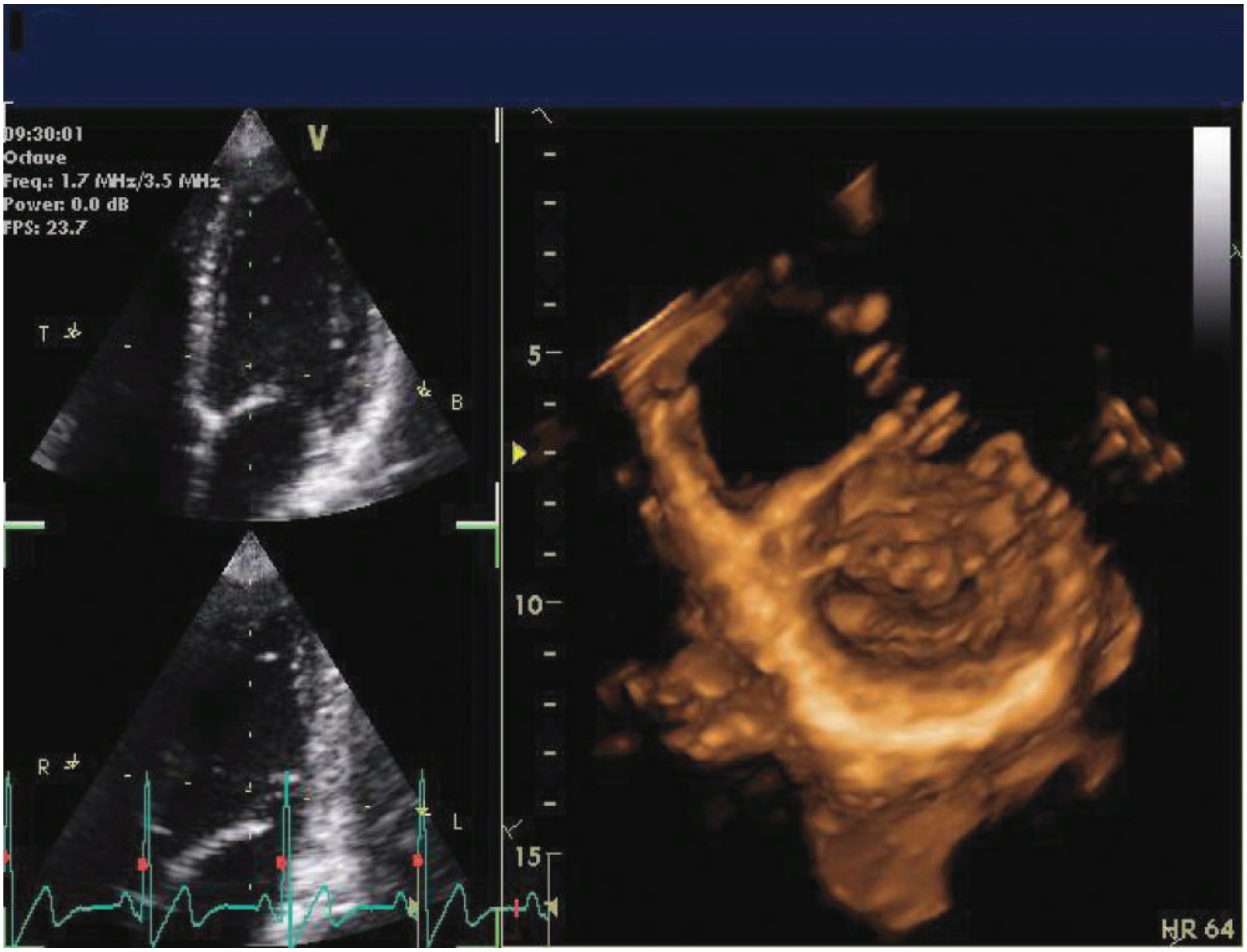

FIGURE 248-21 Normal mitral valve in two-dimensional views (*left*) and with three-dimensional imaging (*right*).

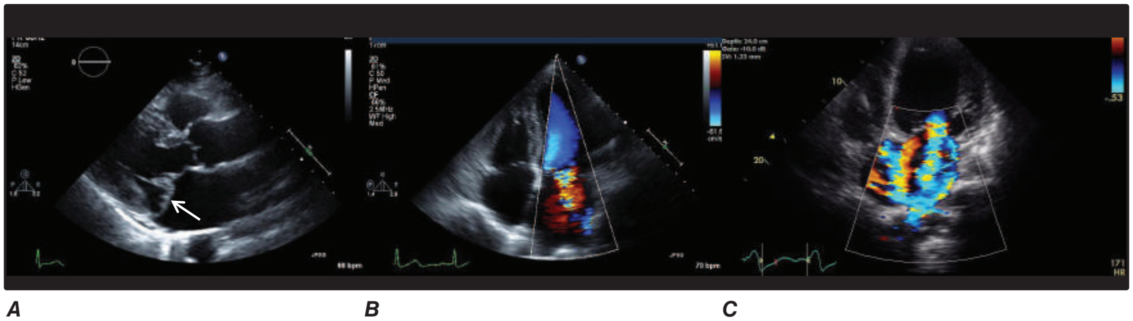

FIGURE 248-22 ***A.*** Mitral valve prolapse with posterior leaflet visualized prolapsing behind the plane of the anterior leaflet (*arrow*). ***B.*** Color flow Doppler showing mitral regurgitation in a patient with mitral valve prolapse. ***C.*** Severe functional mitral regurgitation in a patient with a dilated left ventricle.

As with aortic insufficiency, assessment of ventricular structure and function is also integral in the evaluation of mitral regurgitation. Although some patients have mitral regurgitation due to intrinsic abnormalities of the valve itself, in others, the valve can be relatively normal, but the mitral regurgitation can be secondary to dilatation and remodeling of the left ventricle. So-called functional mitral regurgitation is generally secondary to apical displacement of the papillary muscles in a dilated ventricle, resulting in the leaflets of the mitral valve being pulled toward the apex of the heart, resulting in poor coaptation during systole and resultant relatively central mitral regurgitation. This type of mitral regurgitation can generally be distinguished from intrinsic mitral valve disease, and the surgical or procedural treatment of these conditions can be different. Knowledge of the etiology of mitral regurgitation can be important for a surgeon planning mitral valve surgery. Moreover, new procedural approaches to mitral valve disease may be different depending on the etiology.

Ventricular dilatation is an important predictor of outcome in patients with mitral regurgitation of any cause. It is important to realize that in a patient with significant mitral regurgitation, a large portion of the blood being ejected from the left ventricle with every beat is regurgitant, thus artificially increasing the ejection fraction. Thus, an ejection fraction of 55% in a patient with severe mitral regurgitation may represent substantial reduction in myocardial systolic function.

CMR can be helpful in evaluating mitral regurgitation in a subset of patients when echocardiographic assessment is inadequate. CMR can directly quantify regurgitant volume of the mitral regurgitant jet or indirectly quantify regurgitant volume by measuring the difference of left ventricular stroke volume and aortic forward flow.

Assessment of Mitral Stenosis Rheumatic mitral disease remains the most common cause of mitral stenosis, although mitral stenosis can also result from severe calcification of the mitral leaflets. Rheumatic mitral stenosis has a distinct appearance characterized by tethering at the leaflet tips and relative pliability of the leaflets themselves, resulting in a hockey stick–type deformation particularly of the anterior leaflet **(Fig. 248-23)**. Narrowing of the mitral orifice impedes flow from the left atrium to the left ventricle, resulting in increased pressures in the left atrium, which are then transmitted backward into the pulmonary vasculature and the right side of the heart. When mitral stenosis is suspected, echocardiography can be useful for determining etiology (specifically whether it is rheumatic or not), estimating the valve areas and gradients across the valve, assessing the left atrium, and assessing right ventricular size and function. Assessment of left atrial size and right ventricular size and function is particularly useful in helping determine the severity of the mitral stenosis.

■ MYOCARDIAL INFARCTION AND HEART FAILURE

Role of Imaging after Myocardial Infarction Imaging can be useful in the immediate and long-term follow-up of patients with myocardial infarction. As discussed earlier in the chapter, LGE imaging by CMR is the best technique for imaging for presence or the extent of infarcted myocardium. In a recent multicenter study, LGE imaging identified infarct location accurately and detected acute and chronic infarcts at a sensitivity of 99 and 94%, respectively. In addition, regions of microvascular obstruction (no-reflow) can be seen as dense hypoenhanced areas within the core of a bright region of infarction **(Fig. 248-24)**. Both the presence of LGE and microvascular obstruction are markers of increased clinical risk.

While echocardiography is often used to assess myocardial function immediately after myocardial infarction, myocardial stunning is common in the early post–myocardial infarction period, especially in patients who undergo reperfusion therapy. In these patients, either partial or complete recovery of ventricular function is common within several days, so that early estimation of ejection fraction may be misleading. In patients with uncomplicated myocardial infarction, imaging can generally be deferred for several days so that a more accurate assessment of cardiac function, including regional wall motion, can be assessed **(Fig. 248-25)**.

Echocardiography is the best method for assessment of patients with suspected mechanical complications after myocardial infarction. These include mitral regurgitation secondary to either papillary muscle dysfunction or rupture of papillary muscle head, ventricular septal defect, or cardiac rupture. A new severe systolic murmur should raise

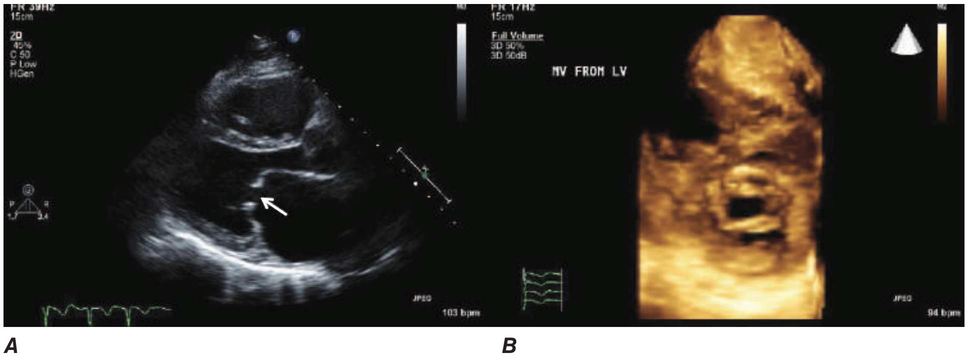

FIGURE 248-23 ***A.*** Rheumatic mitral stenosis showing pliable leaflets tethered at the tips (*arrow*). Note the characteristically enlarged left atrium. ***B.*** Mitral stenosis visualized from a three-dimensional echocardiogram.

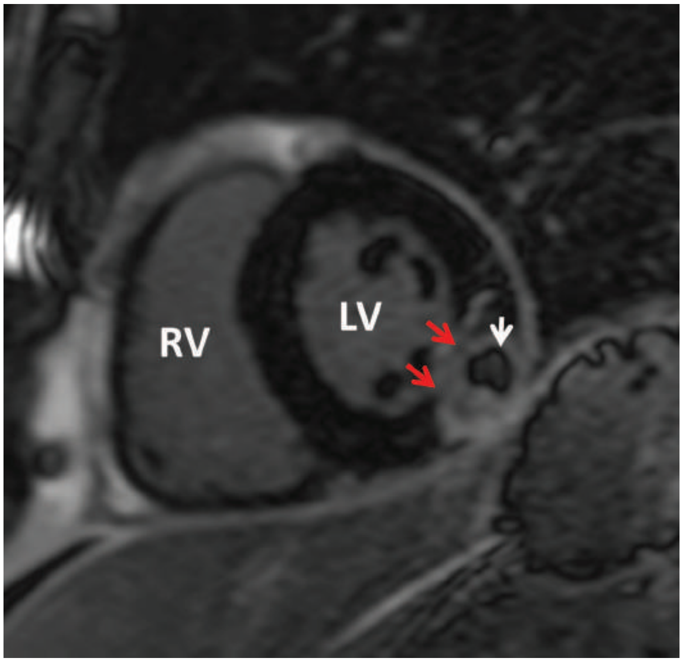

FIGURE 248-24 Example of a patient who presented with inferior ST-segment elevation myocardial infarction (MI) after several days of intermittent chest pain. Magnetic resonance imaging (MRI) confirmed an inferior MI by the location of late gadolinium enhancement (LGE; *red arrows*). In addition, there is a central area of microvascular obstruction (dark region surrounded by the bright LGE, *white arrow*). LV, left ventricle; RV, right ventricle.

suspicions for either severe mitral regurgitation or ventricular septal defect. While cardiac rupture is often catastrophic, contained ruptures, in the form of pseudoaneurysms, can occur, and early diagnosis and surgical treatment are the best way to maximize survival. The presence of thrombus within the pericardial space following myocardial infarction should immediately raise suspicion of myocardial rupture and represents a surgical emergency.

Some patients demonstrate progressive left ventricular dilatation and dysfunction, known as cardiac remodeling, after myocardial infarction. Assessment of cardiac function and regional wall motion is useful in the follow-up period, generally between 1 and 6 months following infarction. The persistence of left ventricular systolic dysfunction following infarction is used to determine the type of therapy (e.g., angiotensin-converting enzyme inhibitors or angiotensin receptor blockers are typically used in patients with systolic dysfunction following myocardial infarction).

In patients with acute or subacute myocardial infarction, investigation of residual ischemia and/or viability is occasionally an important clinical question, especially among those with recurrent symptoms after myocardial infarction **(Fig. 248-26)**. All cardiac imaging techniques can provide information regarding myocardial viability and ischemia. The available data suggest that radionuclide imaging, especially PET, is highly sensitive, with higher negative predictive value than dobutamine echocardiography. In contrast, dobutamine echocardiography tends to be associated with higher specificity and positive predictive accuracy than the radionuclide imaging methods. The experience with CMR suggests that it offers similar predictive accuracies as those seen with dobutamine echocardiography.

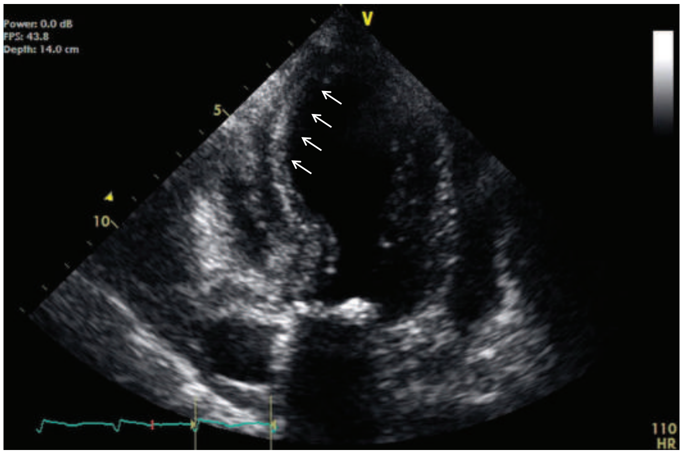

FIGURE 248-25 Acute left anterior descending artery distribution myocardial infarction at end systole showing akinetic region (*arrows*).

Role of Imaging in New-Onset Heart Failure Echocardiography is usually a first-line test in patients presenting with new-onset heart failure. As discussed above, this test provides a direct assessment of ventricular function and can help distinguish patients with reduced from those with mildly reduced or preserved ejection fraction, particularly important because of the different therapeutic options available for patients with each of these types of heart failure. In addition, it provides additional structural information including an assessment of valves, myocardium, and pericardium.

Although coronary angiography is commonly performed in patients with reduced ejection fraction, the determination of heart failure etiology in an individual patient may be difficult even if angiographically obstructive CAD is present. Indeed, patients with heart failure and no angiographic CAD may have typical angina or regional wall motion abnormalities on noninvasive imaging, whereas patients with angiographically obstructive CAD may have no symptoms of angina or history of myocardial infarction. Thus, the appropriate classification for any given patient is not always clear, and it often requires the complementary information of coronary angiography and noninvasive imaging. As discussed above, stress radionuclide imaging and echocardiography can be helpful in delineating the extent and severity of inducible myocardial ischemia and viability. Multiparametric CMR can be quite helpful in the differential diagnosis of heart failure etiologies. Apart from quantifying left and right ventricular volumes and function, CMR can provide information about myocardial ischemia and scar. The pattern of LGE helps differentiate infarction (typically starting in the subendocardium and involving a coronary territory) from other forms of infiltrative or inflammatory cardiomyopathies (typically involving the mid- or subepicardial layers without following a coronary distribution) **(Fig. 248-27)**. In addition, it can assess the presence of myocardial edema using T1 or T2 tissue mapping methods to inform toward chronicity of acute coronary syndromes or noncoronary inflammatory conditions (e.g., myocarditis). Other methods such as T2* mapping of myocardial iron deposition assess for the extent of iron infiltration that leads to cardiac toxicity. Infiltrative cardiomyopathy such as amyloidosis typically has a restrictive cardiomyopathy pattern characterized by biventricular increased wall thickness and bilateral atrial enlargement, as assessed by both echocardiography and CMR. CMR of patients with cardiac amyloidosis often also demonstrates a characteristic pattern of diffuse endocardial infiltration of the left ventricle and the atria (Fig. 248-27). On strain echocardiography, patients with cardiac amyloidosis show reduction in systolic function which typically spares the apical left ventricular segments. Bone scintigraphy complements the use of echocardiography and CMR in cardiac amyloidosis by allowing accurate noninvasive distinction of transthyretin (ATTR) from light chain (AL) cardiac amyloidosis. ATTR typically shows intense ^{99m}Tc tracer uptake compared to patients with AL amyloidosis **(Fig. 248-28)**. Hypertrophic cardiomyopathy has variable degree of increased ventricular thickness and often is seen to have outflow obstruction and intense LGE in regions with marked hypertrophy **(Fig. 248-29)**. CMR also can quantify myocardial iron content in patients at risk of iron-overload cardiomyopathy **(Video 248-7)**.

PET metabolic imaging has a complementary role in the evaluation of inflammatory cardiomyopathies, especially sarcoidosis where the presence of focal and/or diffuse glucose uptake can help identify areas of active inflammation. In addition, for patients undergoing immunosuppressive therapy, PET is frequently used to monitor therapeutic response **(Fig. 248-30)**. In patients with ischemic cardiomyopathy,

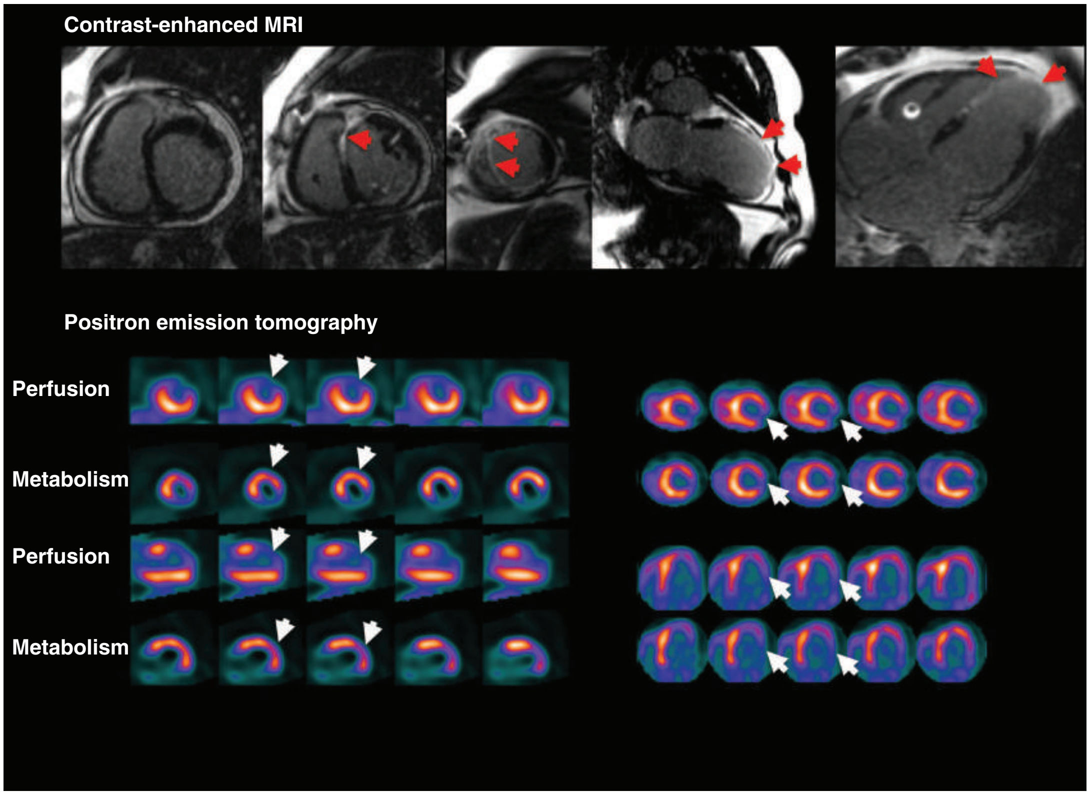

FIGURE 248-26 Examples of myocardial viability patterns obtained with cardiac magnetic resonance imaging (MRI) and positron emission tomography (PET) in three different patients with coronary artery disease. The *top panel* demonstrates extensive late gadolinium enhancement (*bright white areas*) involving the anterior, anteroseptal, and apical left ventricular walls (*arrows*), consistent with myocardial scar and nonviable myocardium. The *lower left panel* demonstrates rubidium-82 myocardial perfusion and ^{18}F-fluorodeoxyglucose (FDG) images showing a large and severe perfusion defect in the anterior, anterolateral, and apical walls, indicating preserved glucose metabolism (so-called *perfusion-metabolic mismatch*) consistent with viable myocardium. The *right lower panel* shows similar PET images demonstrating concordant reduction in perfusion and metabolism (so-called *perfusion-metabolic match*) in the lateral wall, consistent with nonviable myocardium.

radionuclide imaging in general and PET in particular are frequently used to quantify the presence and extent of myocardial ischemia and viability to assist with clinical decision-making related to myocardial revascularization (Fig. 248-25).

ASSESSING CARDIAC FUNCTION IN PATIENTS UNDERGOING CANCER TREATMENT

Therapies used to treat cancer can adversely affect the cardiovascular system. As the efficacy of cancer treatment and survival improve, many patients are presenting with late adverse consequences from chemotherapy and/or radiation therapy on cardiovascular function. Thus, the morbidity and mortality from late cardiovascular complications threaten to offset the early gains in cancer survival, especially among children and young adults. Early recognition and treatment of cardiomyocyte injury are critical for successful application of preventative therapies, but difficult because the adverse effects on cardiac function are a relatively late manifestation after exposure to anticancer therapy.

The accepted standard for clinical diagnosis of cardiotoxicity is defined as a >5% reduction in LVEF to <55% with symptoms of heart failure, or a >10% drop in LVEF to <55% in patients who are asymptomatic. Thus, noninvasive imaging plays a major role in diagnosing and monitoring for cardiac toxicity in patients undergoing cancer treatment. Radionuclide angiography has been the technique of choice for quite some time. However, echocardiography now plays a major role in this application.

Recently, more novel imaging approaches have been advocated, including deformation imaging with echocardiography and fibrosis imaging with CMR. These techniques have shown promising results in experimental animal models and in humans. In addition, there are also proof-of-concept studies in animal models using molecular imaging approaches targeting the mechanisms of cardiac toxicity (e.g., apoptosis and oxidant stress), which can presumably provide the earliest signs of the off-target effects of these therapies. However, these techniques are currently considered experimental.

PERICARDIAL DISEASE

The fibroelastic pericardial sac surrounding the heart consists of a visceral, or epicardial, layer and a parietal layer, with a generally small amount of pericardial fluid in between layers. The pericardium is generally quite pliable and moves easily with the heart during contraction and relaxation. Abnormalities of the pericardium can affect cardiac function primarily by impairing the heart's ability to fill. Inflammation of the pericardium can lead to an accumulation of fluid between the two layers, or *pericardial effusion*, which can be visualized by echocardiography, CMR, or CT. Other reasons for accumulation of pericardial fluid include infection, malignancy, and bleeding into the pericardium. The latter can be the result of catastrophic processes such as trauma, cardiac rupture, perforation in the setting of a cardiac procedure, cardiac surgery, or dissection of the aorta with extension in the pericardium.

Echocardiography remains the initial test of choice for assessing pericardial disease, especially effusions (Fig. 248-31). Moreover, echocardiography can be useful in evaluating for pericardial constrictive physiology, in which a thick noncompliant pericardium impairs cardiac filling. The location, size, and physiologic consequences of accumulated pericardial effusion can generally easily be determined by echocardiography. Pericardial tamponade occurs when enough pericardial fluid accumulates so that the intrapericardial pressure exceeds filling pressures of the heart, generally the right ventricle. The balance between intrapericardial pressure and ventricular pressure is more important than the extent of fluid accumulation. Conditions in which pericardial effusions accumulate over a long period of time, as can be the case in the setting of malignant effusions, can lead to large pericardial fluid accumulations without the classic hemodynamic findings associated with pericardial tamponade. In contrast, rapid accumulations of pericardial fluid, such as those that occur due to cardiac rupture or perforation, can lead to tamponade physiology without very large effusions. In patients with suspected pericardial effusion or tamponade, echocardiography can usually be performed rapidly, at the

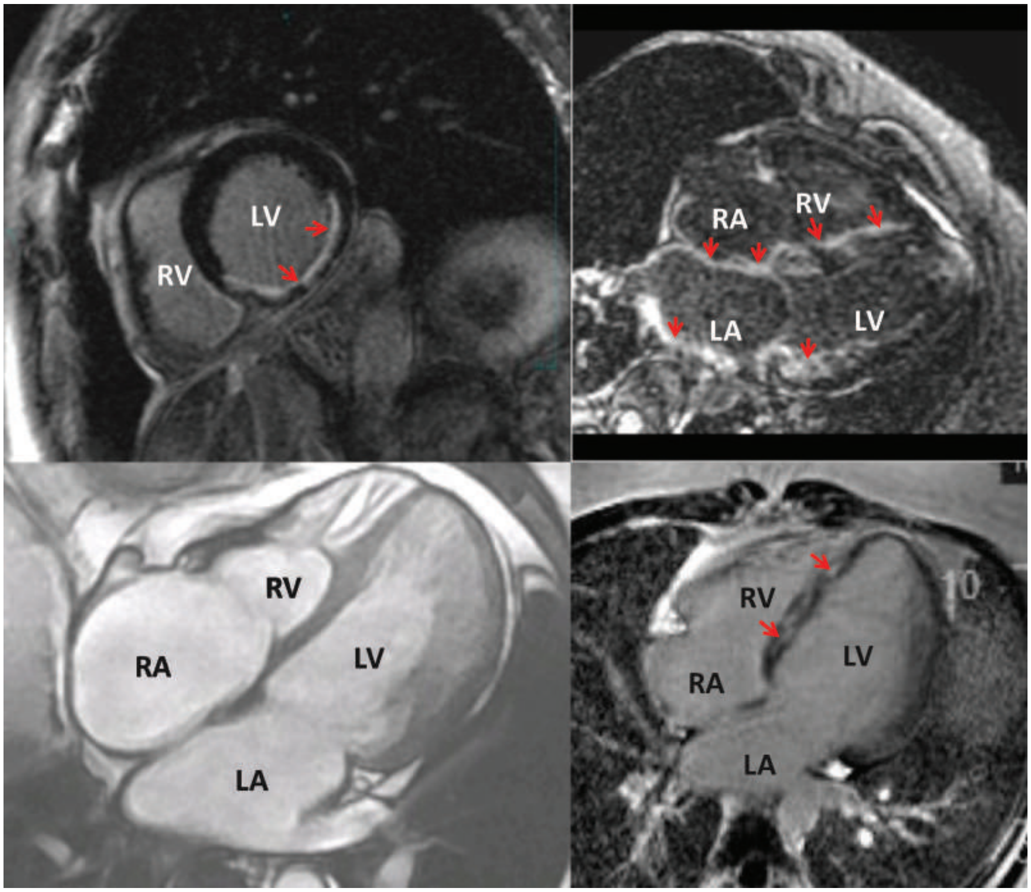

FIGURE 248-27 Differentiation of various cardiomyopathies by cardiac magnetic resonance (CMR). The *left upper panel* shows the short-axis late gadolinium enhancement (LGE) imaging of a patient who suffered an acute myocardial infarction. Note LGE of the endocardial myocardium in the inferior wall extending from the septum to the lateral wall associated with myocardial thinning (*arrows*). The *right upper panel* shows the long-axis LGE imaging of a patient who has cardiac amyloidosis. Note the diffuse LGE throughout left ventricular myocardium, the left atrium, and the interatrial septum (*arrows*). In addition, the blood pool is characteristically dark in signal, indicating sequestration of gadolinium contrast out of the blood pool after injection due to a high burden of amyloidosis in other organs. The *left lower panel* shows a cine diastolic long-axis image of a patient with a nonischemic dilated cardiomyopathy. Note that there is extensive sponge-like noncompacted myocardium of the LV as well as dilatation of all four cardiac chambers. This patient has a nonischemic dilated cardiomyopathy secondary to LV noncompaction. The *right lower panel* shows a 22-year-old female patient with a recent episode of acute chest pain and troponin elevation. Note the multiple mid-wall foci of LGE, which suggests acute myocarditis (*arrows*). LA, left atrium; LV, left ventricle; RA, right atrium; RV, right ventricle.

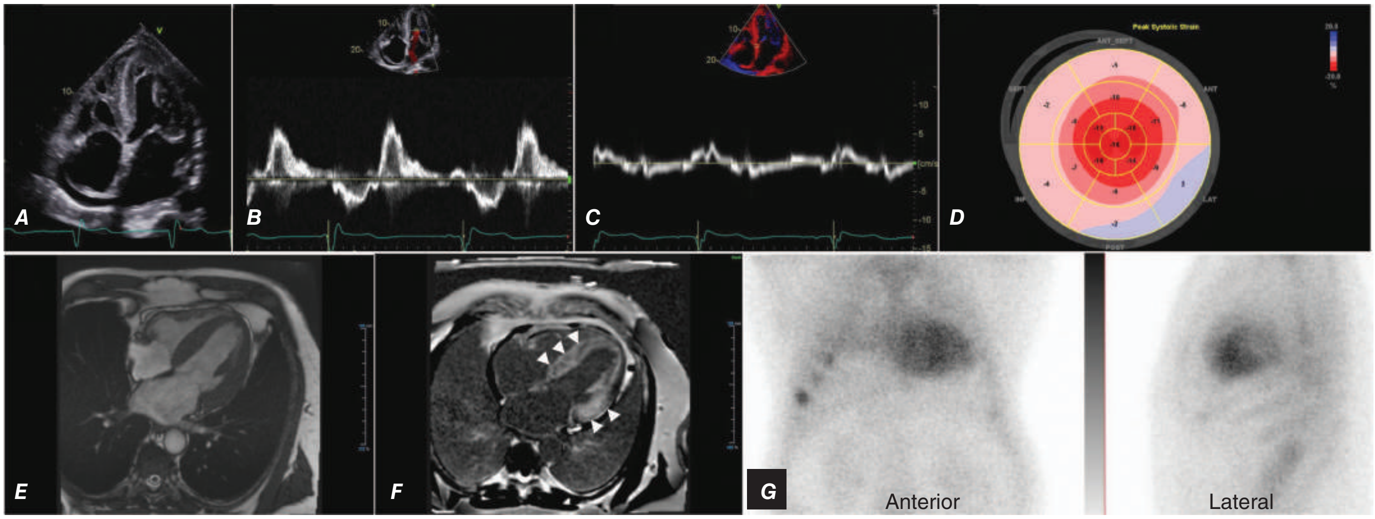

FIGURE 248-28 Multimodality cardiac imaging in a 71-year-old man with history of atrial fibrillation, heart failure with preserved ejection fraction, numbness in his fingers, a history of bilateral carpal tunnel surgery, and remote history of back surgery with L3 and S1 fusion. *A.* Two-dimensional echocardiogram demonstrating moderate increase in left ventricular thickness with a visually estimated left ventricular ejection fraction of 45–50% and small pericardial effusion. ***B, C.*** Mitral inflow and tissue Doppler demonstrating abnormal diastolic function (E/e′ was >20, consistent with elevated left ventricular filling pressure). ***D.*** Bull's-eye strain map demonstrating abnormal global longitudinal strain at the basal segments (*pink and blue colors*) with characteristic apical sparing (*red color*). ***E.*** Cardiac magnetic resonance imaging (MRI) study showing normal left ventricular (LV) cavity size and mildly increased LV mass and mildly enlarged right and left atria. ***F.*** The MRI shows a large amount of diffuse late gadolinium enhancement of all myocardial LV segments but more prominent in a circumferential subendocardial pattern (*arrowheads*). ***G.*** The ^{99m}Tc-pyrophosphate images demonstrate increased myocardial uptake of the radiotracer (uptake in the heart is greater than ribs, grade 3). *(Images courtesy of Dr. Sarah Cuddy, Brigham and Women's Hospital.)*

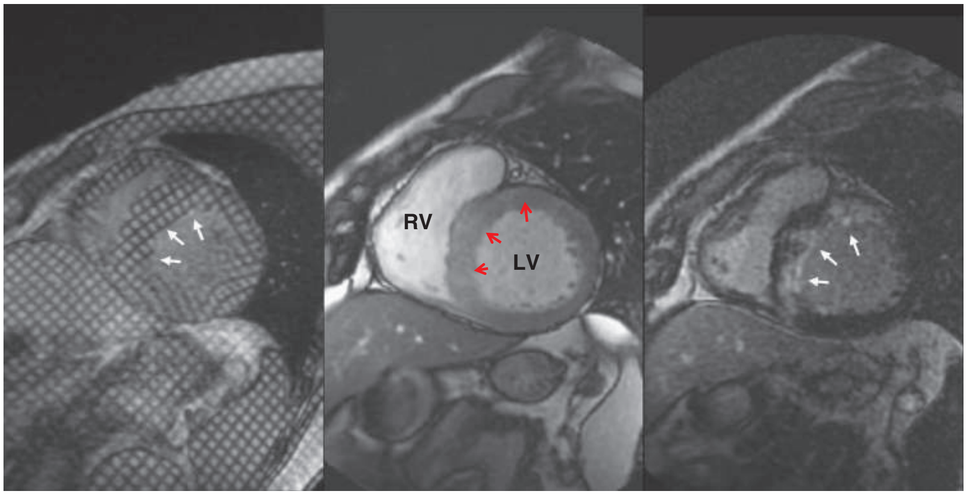

FIGURE 248-29 This figure demonstrates three pulse sequence techniques by cardiac magnetic resonance that are often used to assess patients with hypertrophic cardiomyopathy, all displayed in the mid short-axis scan plane. The *center panel* demonstrates that the left ventricle (LV) was markedly thickened in its wall thickness especially in the LV septum (*red arrows*). This finding was matched by marked regions of late gadolinium enhancement (LGE), which was consistent with fibrosis in these segments (*right panel, white arrows*). The *left panel* was cine myocardial tagging in the same slice plane. Myocardial tagging is used to assess the normal intramyocardial strain by assessing distortion of the myocardial grids during systole. In this case, despite normal-appearing systolic radial wall thickening, the myocardial strain as assessed by the distortion of grids was markedly reduced (*left panel, white arrows*). This finding is consistent with substantial myofibril disarray in the anterior and anteroseptal segments in this patient. RV, right ventricle.

bedside, and even by operators with limited skill. The distance from the parietal to the visceral pericardial layer can be measured, and when this exceeds ~1 cm, an effusion is considered significant. Echocardiographic features suggestive of tamponade include diastolic collapse of the right ventricular free wall, suggestive of pericardial pressures that exceed right ventricular filling pressures, and Doppler evidence of respiratory flow variation, which is the Doppler equivalent of pulsus paradoxus. Despite the benefits of echocardiography in suspected pericardial tamponade, the diagnosis of tamponade remains a clinical diagnosis, and other important features, such as patient's blood pressure in the presence of pulsus paradoxus, need to be considered when weighing therapeutic options.

Chronic inflammation of the pericardium leads to thickening and calcification of the parietal pericardium, resulting in pericardial constriction in which diastolic filling can be severely impaired. In these cases, filling of the ventricles comes to an abrupt halt when the volume of ventricular filling is impaired by the constricting pericardium. Assessment of pericardial thickness in these patients is important, but it is just as important to note that approximately one in five patients with severe pericardial constriction have no significant pericardial thickening by imaging or at surgery. Thus, a lack of thickened pericardium does not rule out pericardial constriction, and patients' signs and symptomatology and physiologic evidence of constriction should be assessed independently. Pericardial constriction typically demonstrates

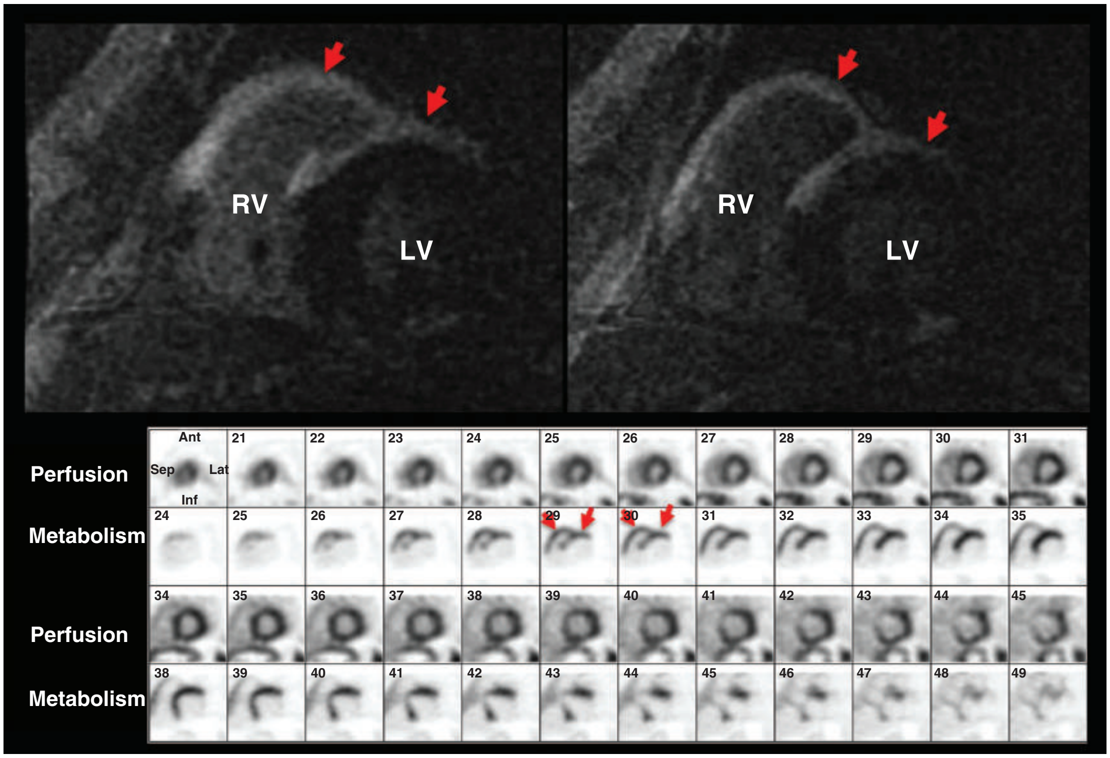

FIGURE 248-30 Representative cardiac magnetic resonance (CMR; *top panel*) and positron emission tomography (PET; *lower panel*) images from a 45-year-old male presenting with complete heart block. The CMR images demonstrate extensive late gadolinium enhancement in the subepicardial left ventricular (LV) anterior and anteroseptal walls and also in the right ventricular (RV) free wall (*arrows*). The PET images demonstrate extensive fluorodeoxyglucose uptake in the same areas, most consistent with active inflammation due to sarcoidosis.

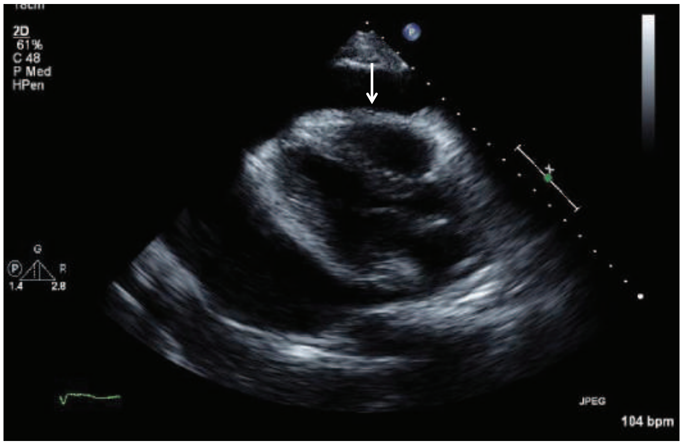

FIGURE 248-31 Pericardial effusion with tamponade physiology. The right ventricle (*arrow*) is small and collapsing in end diastole due to increased pericardial pressure.

marked respiratory changes in diastolic flow on Doppler echocardiography, in contrast to restrictive cardiomyopathy, but substantial overlap exists. CT and CMR offer tomographic, whole-heart assessment of pericardial thickening and other anatomy abnormalities in pericardial constriction enlarged atria, vena cava, and pleural and pericardial effusions **(Figs. 248-32 and 248-33, and Videos 248-8 and 248-9)**. CMR offers the additional information of pericardial fibrosis and inflammation by LGE imaging and evidence of constrictive physiology (e.g., regional relaxation concordance due to myocardial adhesions, abnormal septal bounce with Valsalva maneuver) (Fig. 248-33).

CARDIAC THROMBUS AND MASS

Echocardiography is usually the modality that first detects a cardiac mass with differential diagnoses including thrombus, tumor, or vegetation. Given their unrestricted tomographic views and multiplanar 3D imaging, CMR and CT can complement echocardiography by further characterizing the physical features of the cardiac mass. Compared to CT, CMR has the advantage of higher tissue contrast differentiation, more robust cine imaging, and the use of multifaceted techniques within the same imaging session to determine the physiologic characteristics of the mass. Gadolinium contrast enhancement patterns of increased capillary perfusion can detect vascularity within a mass that differentiates a tumor from a thrombus. Structures that are known to mimic a cardiac mass include (1) anatomic variants, such as the Eustachian valve, Chiari network, crista sagittalis or terminalis, and the right ventricular moderator band, and (2) "pseudotumors," such as interatrial septal aneurysm, coronary or aortic aneurysm, lipomatous hypertrophy of interatrial septum, hiatal hernia, or a catheter/pacemaker lead. Coexisting abnormalities that raise the likelihood of a cardiac thrombus **(Fig. 248-34)** include regional wall motion abnormality from an infarction or ventricular aneurysm, atrial fibrillation leading to slow flow in the left atrial appendage, or presence of venous catheters or recent endovascular injury. CMR has the advantage of being able to assess regional wall motion and infarction or ventricular aneurysm in matching scan planes, adjacent to the cardiac thrombus, using cine and LGE imaging, respectively. For ventricular thrombus, gadolinium-enhanced LGE imaging can detect thrombus at a higher sensitivity than echocardiography by depicting high-contrast difference between the dark thrombus and its adjacent structures and by imaging in three dimensions. In addition, mural thrombus does not enhance on first-pass perfusion and often has a characteristic "etched" appearance (black border surrounding a bright center) on LGE imaging, thus providing higher diagnostic specificity than anatomic information alone **(Fig. 248-35)**. Comparing the signal intensities of a mass before and after contrast injection may confirm the lack of tissue vascularity (i.e., thrombus) by the lack of signal enhancement after contrast administration. Like intracardiac thrombus, regions of microvascular obstruction also appear dark, but microvascular obstruction is confined within the myocardium and surrounded by infarction and thus can be differentiated from intracardiac thrombus. Cardiac CT imaging is ideally suited for small thrombus in the left atrial appendage especially in cases where transesophageal echocardiography is suboptimal or not feasible.

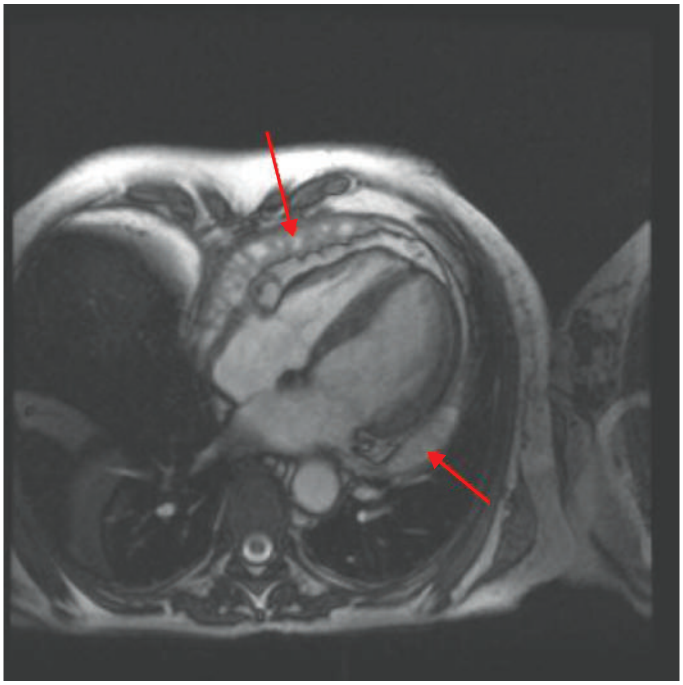

FIGURE 248-32 A female patient developed pericardial constriction and right heart failure, secondary to radiation therapy for breast cancer. Note the multiple pericardial adhesions (*red arrows*).

Metastatic cardiac malignancy is about 20-fold more common than primary cardiac tumors. However, with the exception of involvement of the pericardium, even metastatic tumors of the heart are rare. Metastasis to the heart can be the result of direct invasion (e.g., lung and breast), lymphatic spread (e.g., lymphomas and melanomas), or hematogenous spread (e.g., renal cell carcinoma). Primary benign cardiac tumors are seen mostly in children and young adults and include atrial myxoma, rhabdomyoma, fibroma, and endocardial fibroelastoma **(Fig. 248-36)**. Atrial myxomas are often seen as a round or multilobar mass in the left atrium (75%), right atrium (20%), or ventricles or mixed chambers (5%). They typically have inhomogeneous brightness in the center on cine steady-state free precession imaging due to their gelatinous contents and may have a pedunculated attachment to the fossa ovalis. Primary malignant cardiac tumors are rare and may include angiosarcoma, fibrosarcoma, rhabdomyosarcoma, and liposarcoma.

ROLE OF IMAGING IN INFECTIOUS AND INFLAMMATORY DISEASE

Patients with suspected endocarditis often undergo echocardiography for the purpose of identifying vegetations or intramyocardial abscesses. Vegetations are generally highly mobile structures that most typically are attached to valves or present in areas of the heart with turbulent flow. The absence of a vegetation on echocardiography does not rule out endocarditis, because small vegetations below the resolution of the imaging techniques can be present. Echocardiography remains the best technique for assessment of vegetations because its high temporal resolution allows visualization of the typical oscillating motion, although large vegetations can be visualized with other techniques **(Fig. 248-37)**. The size and location of a vegetation do not necessarily provide any specific information about the type of infection. Abscesses, particularly around the aortic and mitral annuli, are particularly concerning in patients with endocarditis and should be suspected in patients with prolongation of cardiac intervals in the setting of endocarditis. Visualization of both vegetations and possible abscesses is best done with transesophageal echocardiography, particularly in patients with prosthetic valves. Indeed, transesophageal echocardiography is the first test of choice in a patient with a mechanical mitral or aortic valve and

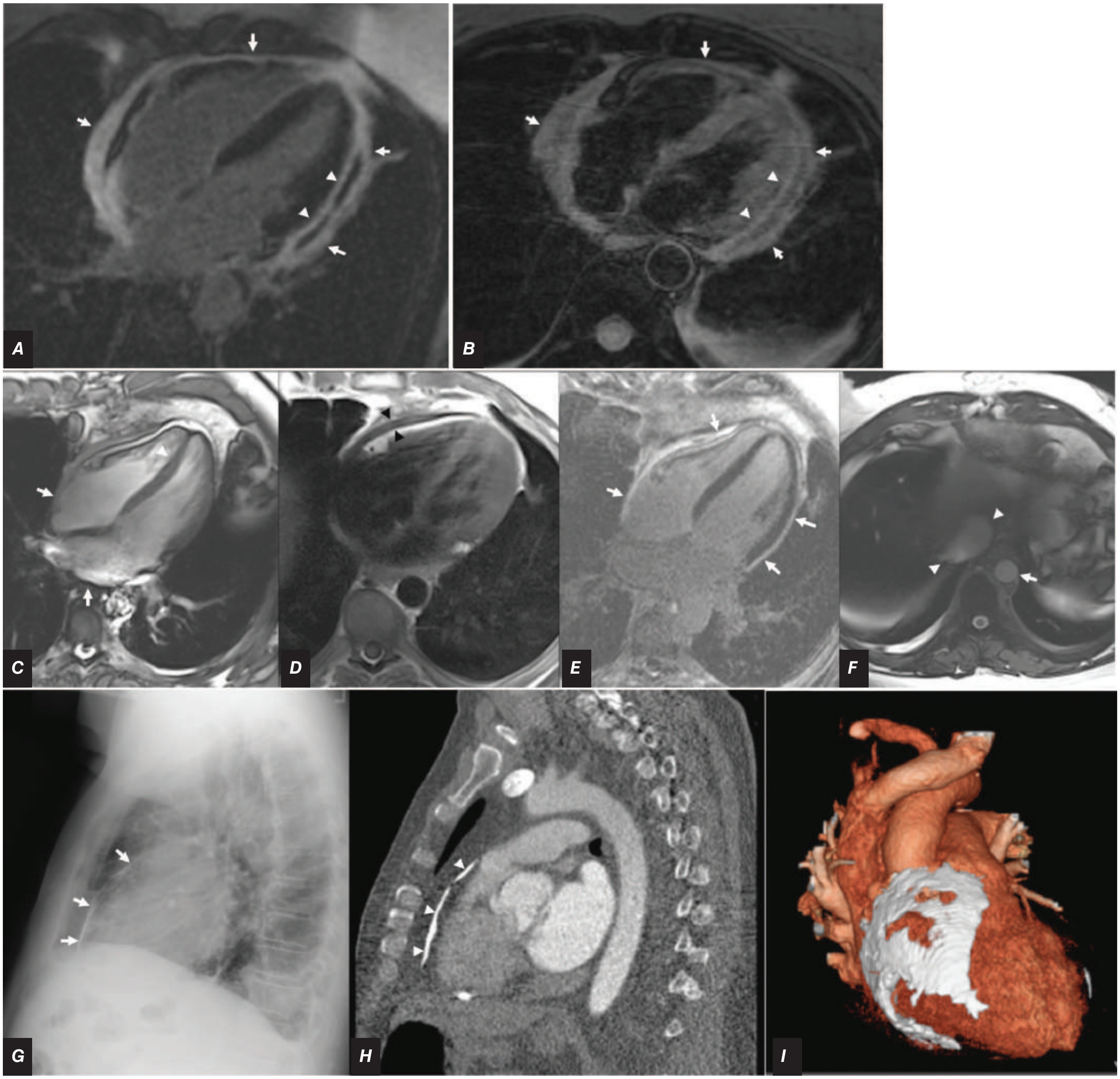

FIGURE 248-33 Representative cardiac magnetic resonance (CMR) and computed tomography (CT) features of pericardial disease. ***A.*** Four-chamber late gadolinium-enhanced image showing severe diffuse enhancement of both the visceral (*arrowheads*) and parietal (*arrows*) layers of the pericardium. The black signal between the two layers represents effusion. ***B.*** Axial T2-weighted image showing severe thickening and increased signal of both the visceral (*arrowheads*) and parietal (*arrows*) layers of the pericardium. ***C.*** Four-chamber steady-state free precession (bright-blood) image in early diastole showing the apical interventricular septum bowed to the left (*white arrowhead*) due to the increased right-sided ventricular volume at the expense of the left ventricular volume. Dilated atria (*white arrows*) are also a feature of constriction. ***D.*** Axial T1-weighted double inversion recovery (black-blood) image showing increased thickness of the pericardium >4 mm (*black arrowhead*). ***E.*** Four-chamber late gadolinium-enhanced image showing diffuse enhancement of the thickened pericardium (*white arrows*). ***F.*** Axial steady-state free precession (bright-blood) image showing dilated inferior vena cava (IVC; *white arrowheads*) that is greater than twice the size of the normal aorta (*white arrow*). Under normal conditions, the IVC should be similar in size to the aorta. ***G.*** Lateral chest radiograph showing calcification of the pericardium anteriorly (*arrows*). ***H.*** Corresponding sagittal view from CT angiography (CTA) showing the calcified pericardium (*arrowheads*). ***I.*** Three-dimensional rendered segmented image from the CTA showing the extent of the pericardial calcification. *(Images courtesy of Dr. Michael Steigner, Brigham and Women's Hospital.)*

suspected endocarditis (Fig. 248-37). Vegetations should be measured because their size has prognostic importance and can be used to decide whether a patient should be taken to surgery.

PET metabolic imaging is emerging as a potentially useful imaging technique to identify the source of infection in patients with prosthetic valves, vascular grafts, and implantable pacemakers/defibrillators, especially in patients in whom echocardiography and/or blood cultures are negative. There is an emerging literature documenting the potential value of macrophage-targeted metabolic imaging with ^{18}F-FDG and PET **(Fig. 248-38)**. Likewise, FDG-PET is also useful to identify vascular inflammation and monitor the response to immunosuppressive therapy **(Fig. 248-39)**.

EVALUATION OF COMMON CONGENITAL ABNORMALITIES IN THE ADULT

While a discussion of complex congenital heart disease is beyond the scope of this chapter, several common congenital abnormalities are present in adults, and cardiac imaging is essential to diagnosing and managing these conditions. Abnormalities of the interatrial septum probably represent the most common adult congenital cardiac

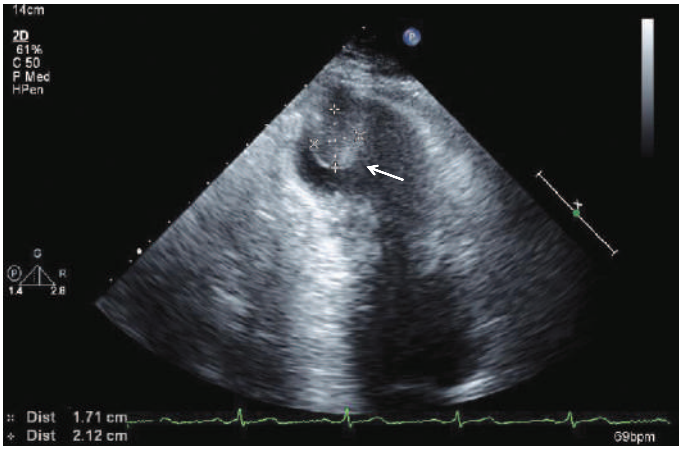

FIGURE 248-34 Cardiac thrombus (*arrow*) in an apical aneurysmal region following acute myocardial infarction.

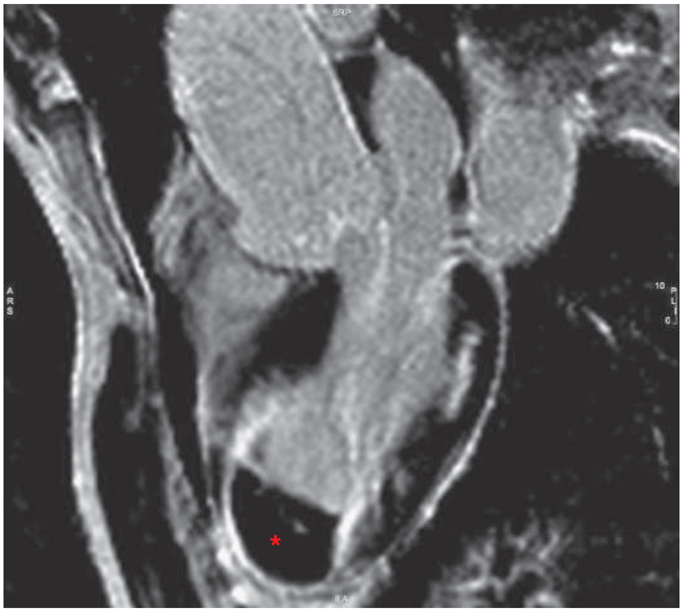

FIGURE 248-35 Late gadolinium enhancement image of a massive anterior infarction complicated by a dyskinetic left ventricular aneurysm and intracavitary thrombus (*red asterisk*).

abnormalities. Patent foramen ovale (PFO) can be identified in almost 25% of patients. In patients with PFO, a one-way flap in the region of the fossa ovalis is normally kept close by the left atrial pressure, which is generally higher than right atrial pressure for much of the cardiac cycle. However, right-to-left flow through a PFO can occur any time the right atrial pressure exceeds the left atrial pressure, including with maneuvers or conditions in which intrathoracic pressure is increased. The presence of a PFO can increase the likelihood of the paradoxical embolus, and thus the presence of a PFO should be determined in patients with stroke or systemic embolus of unknown etiology. Because the one-way flap of the PFO will be closed during much of the cardiac cycle, color flow Doppler will usually not reveal a PFO. Instead, agitated saline (bubble study) is the best way to assess for PFO or atrial septal defect. Saline is agitated and injected peripherally and then enters the right atrium. If no shunt is present, only the right side of the heart will be pacified because the air bubbles will be too small to traverse the lungs. Because PFO is a one-way flap, maneuvers should be used to temporarily increase right atrial pressure. Either a Valsalva maneuver or sniff maneuver can be effective.

Atrial septal defects occur most commonly in the region of the fossa ovalis, referred to as secundum-type defects **(Fig. 248-40)**. Additional atrial septal defects include defects of the sinus venosus and atrium primum. Color flow Doppler echocardiography is usually sufficient for diagnosis of a secundum-type atrial septal defect, but agitated saline is generally needed for the diagnosis of other types of atrial septal defects.

Ventricular septal defects can generally be visualized by color flow Doppler as turbulent high-velocity jets from the left to the right ventricle. In cases where the jet origin is unclear, continuous wave Doppler can estimate the velocities. These would be expected to be extremely high to reflect the pressure gradient between the left and right ventricles. Defects can occur in both the muscular and membranous portions of the ventricular septum.

In patients with either atrial or ventricular septal defects, estimation of the severity of the left-to-right shunt is essential and can be an important determinant in management decisions. Shunts are generally assessed by echocardiography by assessing the relationship between pulmonary flow and aortic flow, the Qp/Qs ratio. Shunts and cardiac anatomy of most congenital heart diseases can also be accurately evaluated by CMR **(Fig. 248-41)**.

ARTIFICIAL INTELLIGENCE IN CARDIAC IMAGING

Artificial intelligence (AI) is poised to revolutionize cardiac imaging over the next decade. AI methods have been applied to image acquisition, especially in echocardiography where a probe needs to

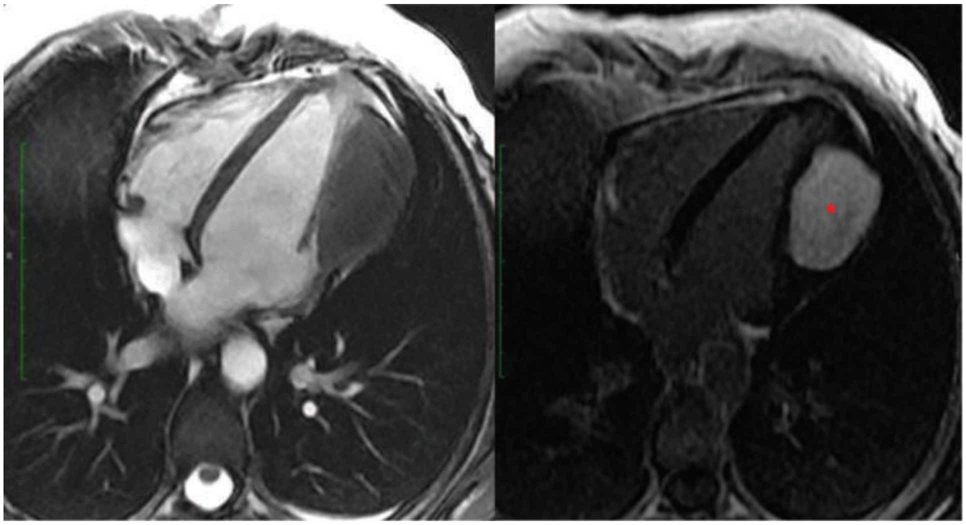

FIGURE 248-36 A case of a cardiac fibroma. A patient presented with shortness of breath and was found to have a large myocardial mass on echocardiography. Cine cardiac magnetic resonance imaging confirmed the large myocardial mass involving the anterolateral wall. Shortly after gadolinium contrast was injected, the myocardial mass demonstrated intense accumulation of contrast on late gadolinium enhancement imaging (*right panel, asterisk*). This is a case of cardiac fibroma. The patient also has gingival hyperplasia and bifid thoracic ribs, a part of the rare Gorlin's syndrome.

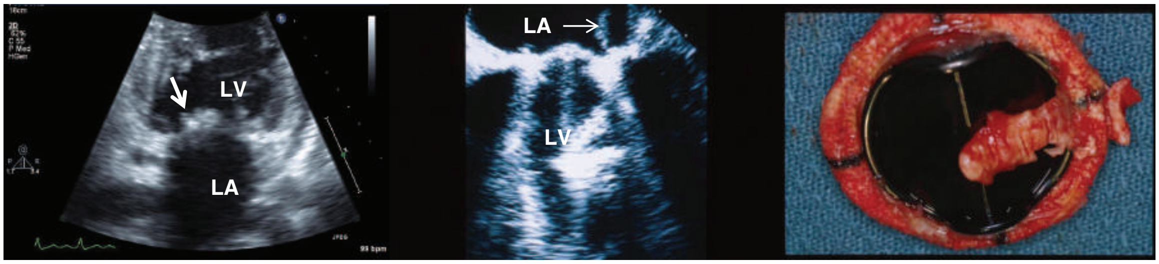

FIGURE 248-37 Vegetation on native mitral valve (*left panel, arrow*). Left atrium (LA) and left ventricle (LV) are indicated. *Middle panel* shows a vegetation on a mechanical prosthesis (St. Jude) indicated by an *arrow*; *right panel* shows vegetation on prosthesis after excision.

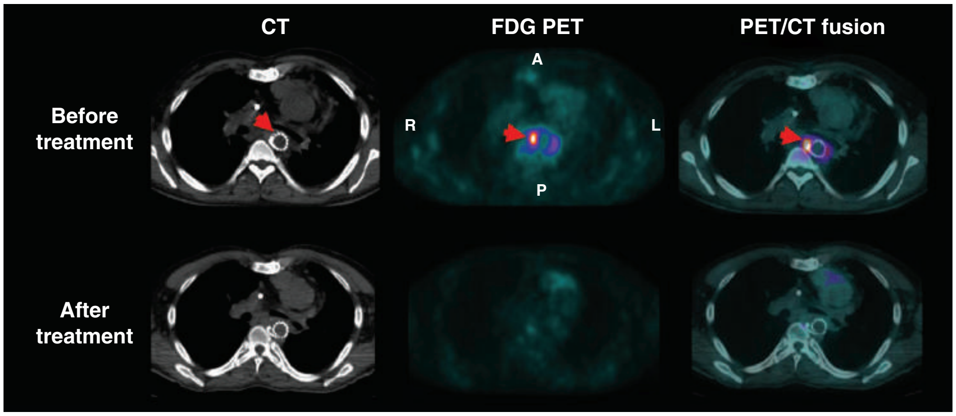

FIGURE 248-38 Representative cross-sectional computed tomography (CT; *left*), fluorodeoxyglucose (FDG) positron emission tomography (PET; *middle*), and fused CT and PET (*right*) images before and after antibiotic treatment in a patient with fever and suspected infection of the stent placed in the descending portion of the aortic arch (*arrow*) for treatment of aortic coarctation. The FDG images before treatment demonstrate intense glucose uptake within the stent, consistent with inflammation/infection. The *lower panel* demonstrates significant attenuation of the FDG signal after treatment. *(Images used with permission from Dr. Sharmila Dorbala, Brigham and Women's Hospital.)*

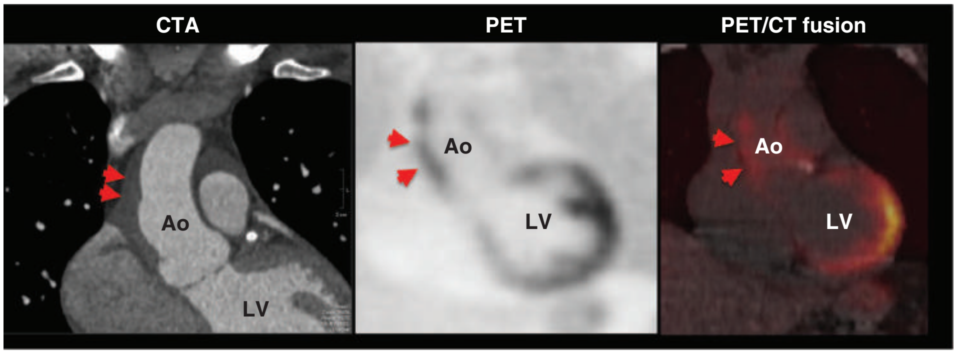

FIGURE 248-39 Representative coronal computed tomography (CT) angiographic (CTA; *left panel*), fluorodeoxyglucose (FDG) positron emission tomography (PET; *middle panel*), and fused CT and PET (*right panel*) images in a patient with suspected aortitis. The CTA images demonstrate thickening of the ascending aorta (Ao), which correlates with intense, focal FDG uptake consistent with active inflammation. LV, left ventricle.

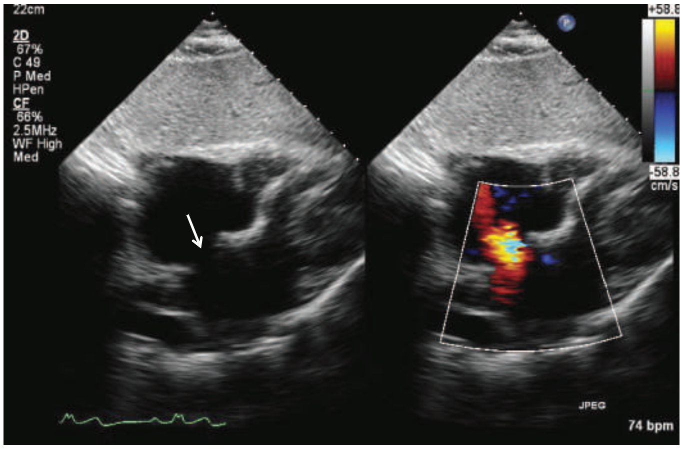

FIGURE 248-40 Large secundum-type atrial septal defect (*arrow*) noted in the subcostal view with color flow Doppler showing flow through the defect (*right*).

be manually applied to a patient's chest and the images acquired. AI software can direct a user to obtain optimal images. Likewise, cardiac MRI is a multicomponent examination that consists of numerous pulse sequences, and it is limited by the availability of an experienced scanning operator, the need for manual contouring and analysis, and a relatively long scan time of 45–60 min. FDA-approved commercial AI solutions are now available to guide the cardiac MRI scanning procedure, greatly reducing the complexities of image acquisition and reconstruction, and improving patient throughput and image quality.

AI is also being embedded in postprocessing and routine quantification. This includes image segmentation and routine measurements of cardiac dimension/volumes and function with echocardiography and CMR, quantification of CAC on nongated CT images (e.g., PET/CT and SPECT/CT), and plaque burden on coronary CTA. There is also growing evidence that AI may be able to refine risk prediction by integrating imaging markers of risk.

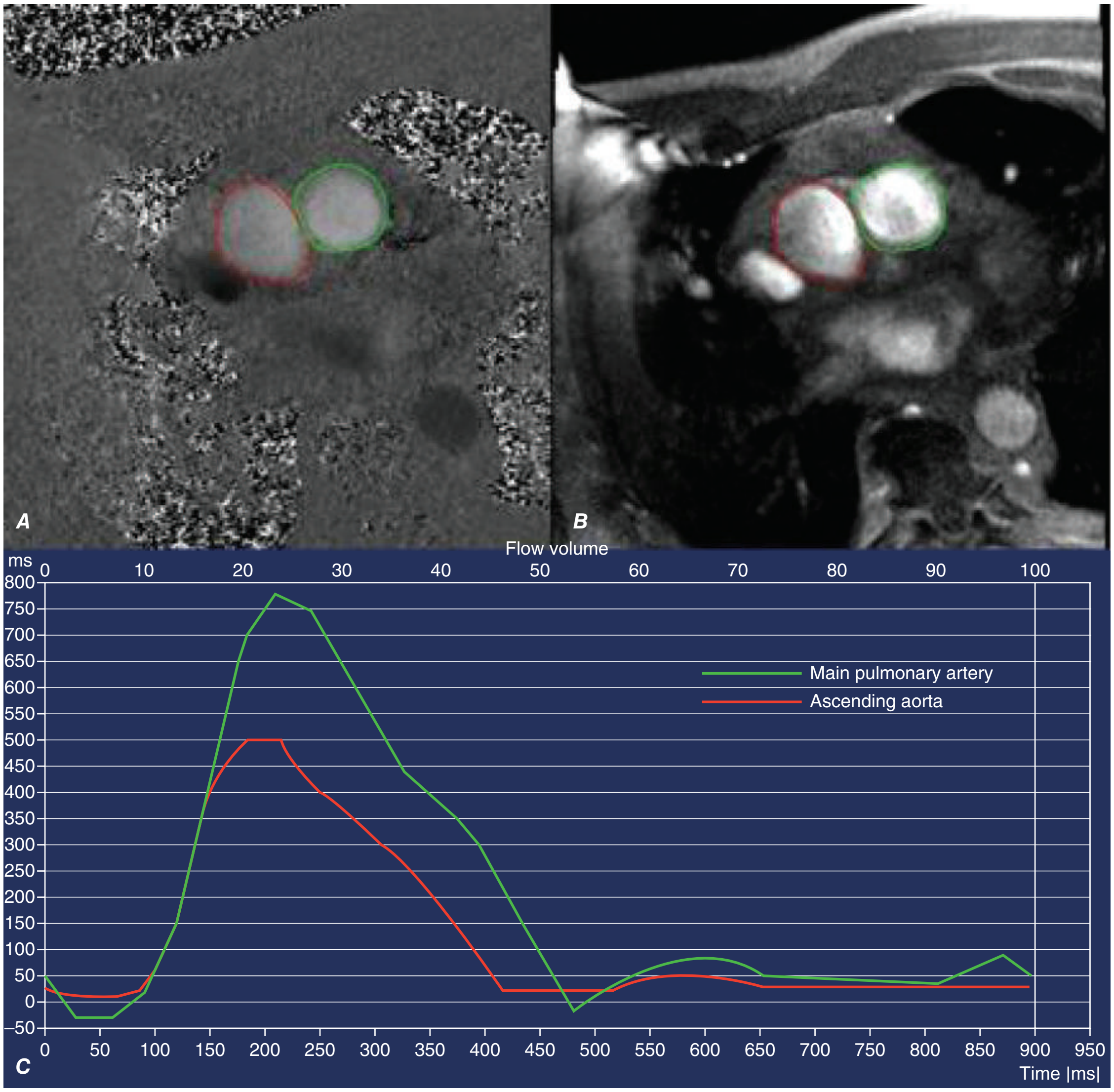

FIGURE 248-41 ***A*** and ***B*** are phase contrast images that display blood flow (phase images on ***A***) and anatomy (structural images on ***B***) of the aorta (*red*) and pulmonary artery (*green*). ***C*** demonstrates the flow curves of the aorta (*red*) and the pulmonary artery (*green*). Note that the total flow (area under the curve) was substantially higher in the pulmonary artery than the aorta, indicative of a marked elevated pulmonary-to-systemic shunt ratio, as a result of the partial anomalous pulmonary venous return that drained into the superior vena cava.

FURTHER READING

Abbara S et al (eds): *CT and MR in Cardiology*. Philadelphia, PA, Elsevier, 2019.

Di Carli MF (ed): *Nuclear Cardiology and Multimodality Imaging: A Companion to Braunwald's Heart Disease*. Philadelphia, PA, Elsevier, 2022.

Kwong RY et al (eds): *Cardiovascular Magnetic Resonance Imaging*, 2nd ed. New York, Humana Press, 2019.

Solomon SD et al (eds): *Essential Echocardiography: A Companion to Braunwald's Heart Disease*. Philadelphia, PA, Elsevier, 2018.

Vandoorne K, Nahrendorf M: Multiparametric imaging of organ system interfaces. Circ Cardiovasc Imaging 10:e005613, 2017.

VIDEO 248-1 Cine steady-state free precession (SSFP) imaging (*left*) in short axis in a patient who had a large anterior myocardial infarction. Only one cut of a stack of short axis is shown. This method allows quantification of left ventricular (LV) and right ventricular (RV) volumes in diastole and systole and calculation of the LV ejection fraction, stroke volumes, and cardiac output (a product of LV stroke volume and heart rate). Note that in this case there is anterior and anteroseptal akinesia (lack of systolic wall thickening, as shown by the left cine movie, *red arrows*) matching by a near-transmural myocardial infarction as seen by the matching late gadolinium enhancement (LGE) image (*right picture, white arrows*).

VIDEO 248-2 This is cine cardiac magnetic resonance (CMR) imaging of a patient in the long-axis four-chamber view. Note that the basal aspect of the right ventricular (RV) free wall is thickened, aneurysmal, and akinetic (*red arrows*). The global RV systolic function is mildly reduced, and the RV is dilated. CMR can image the RV using tomographic views and can quantify the RV volumes and ejection fraction volumetrically. This is a patient who presented with syncopal spells and inducible ventricular tachycardia on subsequent workup. He was diagnosed to have arrhythmogenic right ventricular dysplasia.

VIDEO 248-3 Exercise echocardiogram showing rest images on left and poststress images on right, with parasternal long-axis, upper panel, and apical four-chamber, lower panel, end-systolic frames. Following exercise, the distal septal/apical region becomes akinetic. A = upper left (UL); B = upper right (UR); C = lower left (LL); D = lower right (LR).

VIDEO 248-4 The video shows cardiac magnetic resonance (CMR) myocardial perfusion imaging during vasodilating stress, in three parallel short-axis views. A bolus of gadolinium contrast was injected intravenously while rapid imaging acquisition occurred. The contrast enhances the right ventricle first, then travels through the pulmonary circulation, enters the left ventricle (LV), and then perfuses the LV myocardium. Myocardial perfusion defects with this technique show as black subendocardial rims, reflecting lack of contrast accumulation due to ischemia and/or scar. In this case, the anterior wall has a severe perfusion defect (*red arrow*). Figure 248-14 shows the late gadolinium enhancement (LGE) image of a mid short-axis view. There is no evidence of infarction in the anterior wall, which would be seen as bright white areas, indicating that the stress perfusion defect primarily represents myocardial ischemia. This patient had a significant stenosis of the left anterior descending coronary artery.

VIDEO 248-5 A 60-year-old female presented with intermittent chest pain of 3 days in duration but was pain-free at the time of assessment in the emergency room. Admission electrocardiogram (ECG) demonstrated T-wave inversion in the anterior precordial lead, but cardiac enzymes were normal. A resting cardiac magnetic resonance (CMR) study reviewed a large area of anteroseptal hypokinesia (*left picture*, region of hypokinesia shown by the *red arrows*), matching with a large resting perfusion defect (*middle picture*, perfusion defect shown by the *blue arrows*). Late gadolinium enhancement (LGE) imaging (*right picture*), however, did not show any enhancement to indicate any infarction in the anteroseptal wall, suggesting that the hypocontractile and hypoperfused anteroseptal wall was viable. Urgent coronary angiography demonstrated an acute thrombus in the mid left anterior descending coronary artery, which required coronary stenting. This case represents an example of acute coronary syndrome with hibernating but viable myocardium in the anteroseptal wall. The anteroseptal wall recovered contractile function when reassessed 6 months later.

VIDEO 248-6 A patient with severe aortic regurgitation quantified by cardiac magnetic resonance (CMR). Notice the dark flow jet during diastolic across the aortic valve. For quantitation of the aortic regurgitation severity, a cross-sectional cut was made just below the aortic valve, perpendicular to the aortic regurgitation jet, using phase contrast flow imaging. Apart from aortic regurgitation fraction and volume, CMR also can volumetrically quantify ventricular sizes and dimensions of the aorta, which are useful in monitoring patients with aortic valve diseases.

VIDEO 248-7 These are T2* images of the heart (*left panel*) and the liver (*right panel*) of a patient who has hemochromatosis. Note that iron and the liver are markedly darkened in these movies, indicating high load of iron in the heart muscle and liver. The rate of signal reduction (decay) in the myocardium and liver can be calculated as T2* value expressed in milliseconds. In this case, the T2* was at 10 ms. T2* <20 ms in patients with cardiomyopathy has been shown to indicate iron toxicity as the etiology of the cardiomyopathy, and it carries prognostic valve for such patients at risk of cardiac iron toxicity.

VIDEO 248-8 This video shows the heart in long and short axis. Note the large atria, thickened pericardium, and extensive pericardial adhesions. Given the extensive pericardial adhesions, there is little shearing motion of the ventricles against the parietal pericardium.

VIDEO 248-9 This video shows the heart in long axis with a 4-chamber view. The steady-state free precession (bright-blood) cine show the early diastolic filling bounce of the apical interventricular septum due to constrictive physiology.

249

Diagnostic Cardiac Catheterization and Coronary Angiography

Jane A. Leopold, Ajar Kochar

Diagnostic cardiac catheterization and coronary angiography are considered the gold standard in the assessment of the anatomy and physiology of the heart and its associated vasculature. In 1929, Forssmann demonstrated the feasibility of cardiac catheterization in humans when he passed a urological catheter from a vein in his arm to his right atrium and documented the catheter's position in the heart by x-ray. In the 1940s, Cournand and Richards applied this technique to patients with cardiovascular disease to evaluate cardiac function. These three physicians were awarded the Nobel Prize in 1956. In 1958, Sones inadvertently performed the first selective coronary angiography when a catheter in the left ventricle slipped back across the aortic valve, engaged the right coronary artery, and power-injected 40 mL of contrast down the vessel. The resulting angiogram provided superb anatomic detail of the artery, and the patient suffered no adverse effects. Sones went on to develop selective coronary catheters, which were modified further by Judkins, who developed preformed catheters and allowed coronary artery angiography to gain widespread use as a diagnostic tool. In the United States, cardiac catheterization is the second most common operative procedure, with more than 1.0 million procedures performed annually.

CARDIAC CATHETERIZATION

■ INDICATIONS, RISKS, AND PREPROCEDURE MANAGEMENT

Cardiac catheterization and coronary angiography are indicated to evaluate the extent and severity of cardiac disease in symptomatic patients and to determine if medical, surgical, or catheter-based interventions are warranted (Table 249-1). They are also used to exclude

TABLE 249-1 Indications for Cardiac Catheterization and Coronary Angiography

Asymptomatic or Symptomatic Coronary Artery Disease
High risk for adverse outcome based on noninvasive testing
Sudden cardiac death
Sustained (>30 s) monomorphic ventricular tachycardia
Nonsustained (<30 s) polymorphic ventricular tachycardia
Symptomatic Coronary Artery Disease
Canadian Cardiology Society Class II, III, or IV stable angina on medical therapy
Acute coronary syndrome (unstable angina and non-ST-segment elevation myocardial infarction)
Chest-pain syndrome of unclear etiology and equivocal findings on noninvasive tests
Acute Myocardial Infarction
Reperfusion with primary percutaneous coronary intervention
Persistent or recurrent ischemia
Pulmonary edema and/or reduced ejection fraction
Cardiogenic shock or hemodynamic instability
Risk stratification or positive stress test after acute myocardial infarction
Mechanical complications—mitral regurgitation, ventricular septal defect
Valvular Heart Disease
Suspected severe valve disease in symptomatic patients—dyspnea, angina, heart failure, syncope
Infective endocarditis with need for cardiac surgery
Asymptomatic patients with aortic regurgitation and cardiac enlargement or ↓ ejection fraction
Prior to cardiac surgery or transcatheter aortic valve replacement or other percutaneous valvular interventions in patients with suspected coronary artery disease
Congestive Heart Failure
New-onset angina or suspected undiagnosed coronary artery disease
New-onset cardiomyopathy of uncertain cause or suspected to be due to coronary artery disease
Congenital Heart Disease
Prior to surgical correction or percutaneous interventions, when symptoms or noninvasive testing suggests coronary disease
Suspicion for congenital coronary anomalies
Pericardial Disease
Symptomatic patients with suspected cardiac tamponade or constrictive pericarditis
Cardiac Transplantation
Preoperative and postsurgical evaluation
Other Conditions
Hypertrophic cardiomyopathy with angina
Diseases of the aorta when knowledge of coronary artery involvement is necessary for management
Pulmonary hypertension
Unexplained dyspnea

severe disease in symptomatic patients with equivocal findings on noninvasive studies and in patients with chest-pain syndromes of unclear etiology for whom a definitive diagnosis is necessary for management. Cardiac catheterization is not mandatory prior to cardiac surgery in some younger patients who have uncomplicated congenital or valvular heart disease that is well defined by noninvasive imaging and who do not have symptoms or risk factors that suggest concomitant coronary artery disease.

The risks associated with elective cardiac catheterization are relatively low, with a reported risk of <0.1% for myocardial infarction, 0.01% for stroke, and 0.05% for death. For elective and emergent procedures, the in-hospital mortality is 1.4%. These risks increase substantially if the catheterization is performed emergently, during acute myocardial infarction, or in hemodynamically unstable patients. Additional risks of the procedure include tachy- or bradyarrhythmias that require countershock, temporary pacing or pharmacologic therapy, acute renal failure leading to transient or permanent dialysis, air embolus, hemodynamic compromise, acute pulmonary edema, aortic dissection or other peripheral vascular complications that necessitate surgical repair or percutaneous intervention, and significant access-site bleeding. Of these risks, vascular access-site bleeding is the most common complication, occurring in 1.5–2.0% of patients, with major bleeding events associated with a worse short- and long-term outcome. These vascular access site bleeding complications are less common with the radial approach as compared with the femoral access site.

In patients who understand and accept the risks associated with cardiac catheterization, there are no absolute contraindications when the procedure is performed in anticipation of a life-saving intervention. Relative contraindications for elective procedures do, however, exist; these include decompensated congestive heart failure; acute renal failure; severe chronic renal insufficiency, unless dialysis is planned; bacteremia; acute stroke; active gastrointestinal bleeding; excessive anticoagulation or recent lytic administration; severe, uncorrected electrolyte abnormalities; a history of an anaphylactic/anaphylactoid reaction to iodinated contrast agents without premedication; and a history of allergy/anaphylaxis/bronchospasm to aspirin in patients for whom progression to a percutaneous coronary intervention is likely and aspirin desensitization has not been performed. For right heart catheterization, contraindications include right heart tumor, endocarditis, thrombus, or clot in transit.

Contrast allergy and contrast-induced acute kidney injury merit further consideration because these adverse events may occur in otherwise healthy individuals and prophylactic measures exist to reduce risk. Allergic reactions to contrast agents occur in <5% of cases, with severe anaphylactoid (clinically indistinguishable from anaphylaxis, but not mediated by an IgE mechanism) reactions occurring in 0.1–0.2% of patients. Mild reactions manifest as nausea, vomiting, and urticaria, while severe anaphylactoid reactions lead to hypotensive shock, pulmonary edema, and cardiorespiratory arrest. Patients with a history of significant contrast allergy should be premedicated prior to planned coronary angiography with corticosteroids (prednisone 50 mg orally 13, 7, and 1 hour prior to the procedure) and antihistamines (H_1-blockers), and studies performed with nonionic, low-osmolar contrast agents that have a lower reported rate of allergic reactions.

Contrast-induced acute kidney injury, defined as an increase in creatinine >0.5 mg/dL or 25% above baseline that occurs 48–72 h after contrast administration, occurs in ~2–7% of patients, with rates of 20–30% reported in high-risk patients. Risk factors for contrast-induced kidney injury are diabetes mellitus, congestive heart failure, chronic kidney disease, anemia, older age, and ST-segment elevation myocardial infarction. Dialysis is required in 0.3–0.7% of patients and is associated with a fivefold increase in in-hospital mortality. For all patients, adequate intravascular volume expansion with intravenous 0.9% saline (1.0–1.5 mL/kg per hour) for 3–12 h before and continued 6–24 h after the procedure limits the risk of contrast-induced acute kidney injury by >50%. Other strategies to decrease risk include the administration of sodium bicarbonate (3 mL/kg per hour) 1 h before and 6 h after the procedure (similar outcome to saline infusion); use of low- or iso-osmolar contrast agents; and limiting the volume of contrast to <50 mL per procedure. Diabetic patients treated with metformin should stop the drug 24 h prior to the procedure and not restart until 48 h after contrast administration to limit the associated risk of lactic acidosis.

Cardiac catheterization is performed after the patient has fasted for 6 h and has received intravenous conscious sedation to remain awake but sedated during the procedure. Diabetic patients who have a tendency to develop hypoglycemia should hold oral hypoglycemic medications. All patients with suspected coronary artery disease are pretreated with 325 mg of aspirin. In patients in whom the procedure is likely to progress to a percutaneous coronary intervention, an additional $P2Y_{12}$ antiplatelet inhibitor agent should be started: clopidogrel (600-mg loading dose and 75 mg daily), prasugrel (60-mg loading dose and 10 mg daily), or ticagrelor (180-mg loading dose and 90 mg twice daily). Prasugrel should not be selected for individuals with prior stroke or transient ischemic attack and is not recommended for patients 75 years of age or older. Ticagrelor should not be chosen for patients with bradycardia; additionally, up to 10% of patients treated with ticagrelor may experience dyspnea. For those patients who will be treated with an oral anticoagulant after the procedure, clopidogrel is the antiplatelet agent of choice. Intravenous cangrelor may be used for patients who are unable to tolerate an oral $P2Y_{12}$ inhibitor medication. Warfarin is held starting 2–3 days prior to the catheterization to allow the international normalized ratio (INR) to fall to <1.8 and limit access-site bleeding complications. For patients at high risk of thrombotic complications (mechanical heart valve, ventricular assist device), a heparin bridge may be necessary. The direct oral anticoagulants (DOACs) should be stopped 24–48 h prior to the test. Cardiac catheterization is a sterile procedure, so antibiotic prophylaxis is not required.

TECHNIQUE

Cardiac catheterization and coronary angiography provide a detailed hemodynamic and anatomic assessment of the heart and coronary arteries. The selection of procedures is dependent on the patient's symptoms and clinical condition, with some direction provided by noninvasive studies.

Vascular Access Cardiac catheterization procedures are performed using a percutaneous technique to enter the femoral, radial, or ulnar artery and femoral, brachial, or internal jugular vein as the access sites for left and right heart catheterization, respectively. A flexible sheath is inserted into the vessel over a guidewire, allowing diagnostic catheters to be introduced into the vessel and advanced toward the heart using fluoroscopic guidance. The radial artery (or rarely the ulnar or brachial artery) access site is advantageous in patients with peripheral arterial disease that involves the abdominal aorta, iliac, or femoral vessels; severe iliac artery tortuosity, morbid obesity, or preference for early postprocedure ambulation. Use of radial artery access is the preferred access route due to a lower rate of access-site bleeding complications and improved patient comfort. The femoral artery should be accessed using a micropuncture needle with ultrasound guidance to ensure that the insertion site is above the bifurcation of the profunda femoris and superficial femoral artery and below the inferior epigastric artery. Femoral artery access outside this zone increases the risk of retroperitoneal hemorrhage, pseudoaneurysm formation, and/or arteriovenous fistula. Complications associated with radial artery access include severe radial vasospasm that may result in device entrapment, forearm hematoma, and rarely, compartment syndrome. The internal jugular or antecubital veins serve as the preferred access sites to the right heart when the patient has an inferior vena cava filter in place or requires prolonged hemodynamic monitoring.

Right Heart Catheterization This procedure measures pressures in the right heart and pulmonary artery. Right heart catheterization is no longer a routine part of diagnostic cardiac catheterization, but it is reasonable in patients with unexplained dyspnea, pulmonary hypertension, valvular heart disease, pericardial disease, right and/or left ventricular dysfunction, congenital heart disease, and suspected intracardiac shunts. Right heart catheterization most commonly uses a balloon-tipped flotation catheter that is advanced sequentially to the right atrium, right ventricle, pulmonary artery, and pulmonary wedge position (as a surrogate for left atrial pressure) using fluoroscopic guidance; in each cardiac chamber, pressure is measured and blood samples are obtained for oxygen saturation analysis to screen for intracardiac shunts and calculate a cardiac output.

Left Heart Catheterization This procedure measures pressures in the left heart as a determinant of left ventricular performance. With the aid of fluoroscopy, a catheter is guided to the ascending aorta and across the aortic valve into the left ventricle to provide a direct measure of left ventricular pressure. In patients with a tilting-disc prosthetic aortic valve, crossing the valve with a catheter is contraindicated, and the left heart may be accessed via a transseptal technique from the right atrium using a needle-tipped catheter to puncture the atrial septum at the fossa ovalis. Once the catheter crosses from the right to the left atrium, it can be advanced across the mitral valve to the left ventricle. This technique is also used for mitral valvuloplasty. Heparin is given for prolonged procedures to limit the risk of stroke from embolism of clots that may form on the catheter. For patients with heparin-induced thrombocytopenia, the direct thrombin inhibitors bivalirudin (0.75 mg/kg bolus, 1.75 mg/kg per hour for the duration of the procedure) or argatroban (350 μg/kg bolus, 15 μg/kg per min for the duration of the procedure) may be used.

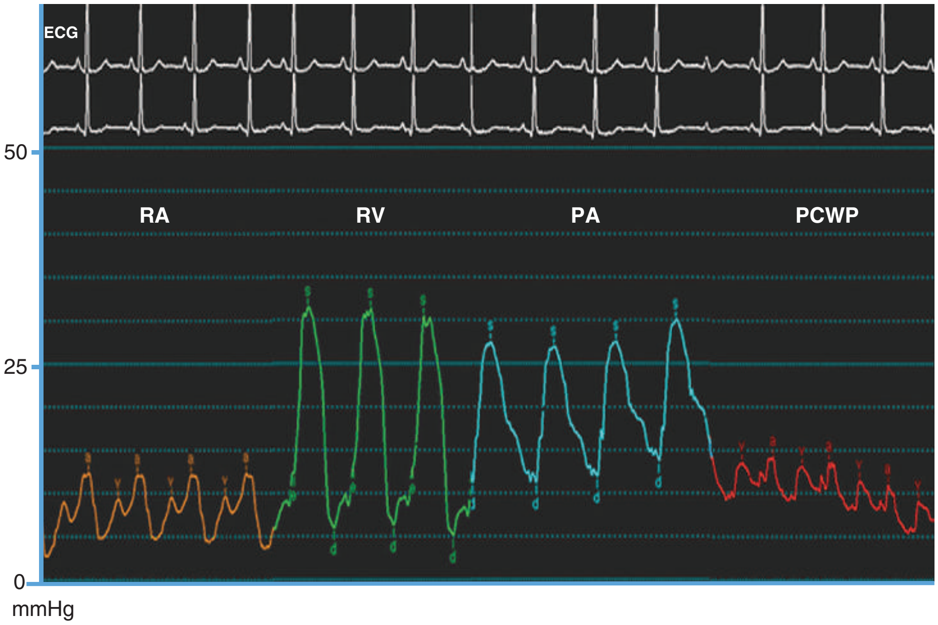

FIGURE 249-1 Normal hemodynamic waveforms recorded during right heart catheterization. Atrial pressure tracings have a characteristic "*a*" wave that reflects atrial contraction and a "*v*" wave that reflects pressure changes in the atrium during ventricular systole. Ventricular pressure tracings have a low-pressure diastolic filling period and a sharp rise in pressure that occurs during ventricular systole. d, diastole; PA, pulmonary artery; PCWP, pulmonary capillary wedge pressure; RA, right atrium; RV, right ventricle; s, systole.

HEMODYNAMICS

A comprehensive hemodynamic assessment involves obtaining pressure measurements in the right and left heart and peripheral arterial system and determining the cardiac output (Table 249-2). The shape and magnitude of the pressure waveforms provide important diagnostic information; an example of normal pressure tracings is shown in Fig. 249-1. In the absence of valvular heart disease, the atria and ventricles are "one chamber" during diastole when the tricuspid and mitral valves are open, whereas in systole, when the pulmonary and aortic valves are open, the ventricles and their respective outflow tracts are considered "one chamber." These concepts form the basis by which hemodynamic measurements are used to assess valvular stenosis. When aortic stenosis is present, there is a systolic pressure gradient between the left ventricle and the aorta; when mitral stenosis is present, there is a diastolic pressure gradient between the pulmonary capillary wedge (left atrial) pressure and the left ventricle (Fig. 249-2). Hemodynamic measurements also discriminate between aortic stenosis and hypertrophic obstructive cardiomyopathy where the asymmetrically hypertrophied septum creates a dynamic intraventricular pressure gradient during ventricular systole. The magnitude of this obstruction is measured using an end-hole catheter positioned at the left ventricular apex that is pulled back while recording pressure; once the catheter has passed the septal obstruction and is positioned in the apex of the left ventricle, a gradient can be measured between the left ventricular apex and the aorta. Hypertrophic obstructive cardiomyopathy is confirmed by the Brockenbrough-Braunwald sign: following a premature ventricular contraction, there is an increase in the left ventricular–aorta pressure gradient with a simultaneous decrease in the aortic pulse pressure. The finding of a decrease in pulse pressure is absent in aortic stenosis.

TABLE 249-2 Normal Values for Hemodynamic Measurements

Pressures (mmHg)	
Right atrium (mean)	1–8
Right ventricle	
Peak systolic/end diastolic	15–30/1–8
Pulmonary artery	
Peak systolic/end diastolic	15–30/4–12
Mean	9–19
Pulmonary capillary wedge (mean)	4–12
Left atrium (mean)	4–12
Left ventricle	
Peak systolic/end diastolic	90–129/5–12
Aorta	
Peak systolic/end diastolic	90–129/60–79
Mean	70–100
Resistances ([dyn-s]/cm^5)	
Systemic vascular resistance	900–1400
Pulmonary vascular resistance	40–120
Oxygen Consumption Index ([L-min]/m^2)	**115–140**
Arteriovenous oxygen difference (vol %)	3.5–4.8
Cardiac index ([L-min]/m^2)	2.8–4.2

Regurgitant valvular lesions increase volume (and pressure) in the "receiving" cardiac chamber. In severe mitral and tricuspid regurgitation, the increase in blood flow to the atria takes place during ventricular systole, leading to an increase in the *v* wave (often two times greater than the mean pressure). The size of the *v* wave is a measure of the compliance of the left atrium but is not a reliable measure of the severity of the mitral regurgitation. Severe aortic regurgitation leads to a decrease in aortic diastolic pressure with a concomitant rise in left ventricular end-diastolic pressure, resulting in equalization of pressures between the two chambers at end diastole.

Hemodynamic measurements are also used to differentiate between cardiac tamponade, constrictive pericarditis, and restrictive cardiomyopathy (Table 249-3). In cardiac tamponade, right atrial pressure is increased with a decreased or absent "y" descent, indicative of impaired right atrial emptying in diastole, and there is diastolic equalization of pressures in all cardiac chambers. In constrictive pericarditis, right atrial pressure is elevated with a prominent "y" descent, indicating rapid filling of the right ventricle during early diastole. A diastolic dip and plateau, or "square root sign," in the ventricular waveforms due to an abrupt halt in ventricular filling during diastole, elevated right

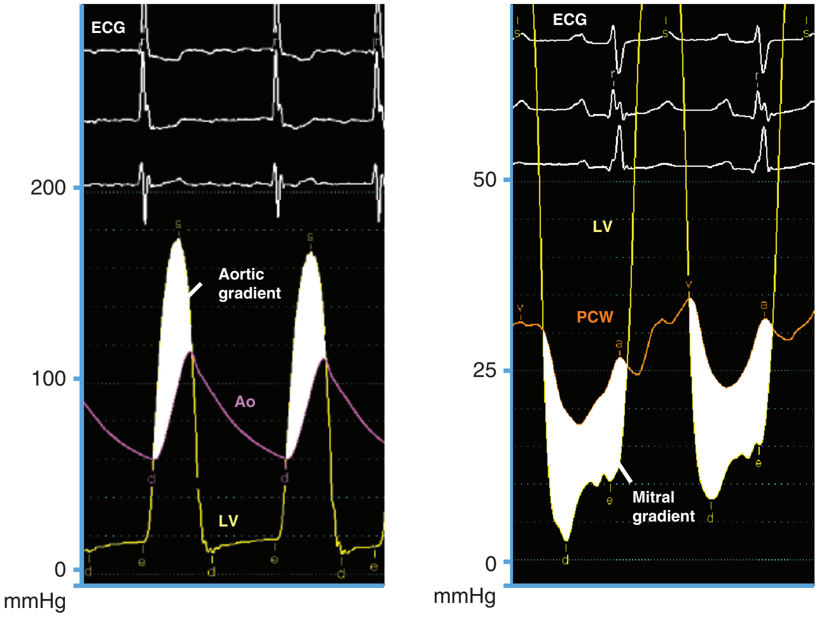

FIGURE 249-2 Severe aortic and mitral stenosis. Simultaneous recording of left ventricular (LV) and aortic (Ao) pressure tracings demonstrates a 62-mmHg mean systolic gradient (*shaded area*) that corresponds to an aortic valve area of 0.6 cm^2 (*left*). Simultaneous recording of LV and pulmonary capillary wedge (PCW) pressure tracings reveals a 14-mmHg mean diastolic gradient (*shaded area*) that is consistent with critical mitral stenosis (mitral valve area = 0.5 cm^2). d, diastole; e, end diastole; s, systole.

ventricular and pulmonary artery pressures, and discordant pressure changes in the right and left ventricles with inspiration (right ventricular systolic pressure increases while left ventricular systolic pressure decreases) are observed. In the absence of constriction, the two ventricular pressures are concordant. The latter hemodynamic phenomenon is the most specific for constriction. Restrictive cardiomyopathy may be distinguished from constrictive pericarditis by a marked increase in right ventricular and pulmonary artery systolic pressures (usually >60 mmHg), a separation of the left and right ventricular diastolic pressures by >5 mmHg (at baseline or with acute volume loading), and concordant changes in left and right ventricular diastolic filling pressures with inspiration (both increase).

Cardiac Output Cardiac output is measured by the Fick method or the thermodilution technique. Typically, the Fick method and thermodilution technique are both performed during cardiac catheterization, although the Fick method is considered more reliable in the presence of tricuspid regurgitation and in low-output states. The Fick method uses oxygen as the indicator substance and is based on the principle that the amount of a substance taken up or released by an organ (oxygen consumption) is equal to the product of its blood flow (cardiac output) and the difference in the concentration of the substance in the arterial and venous circulation (arterial-venous oxygen difference). Thus, the formula for calculating the Fick cardiac output is:

Cardiac output (L/min) = (oxygen consumption [mL/min])/
(arterial-venous oxygen difference [mL/L])

If oxygen consumption is not measured directly, then it is estimated as 125 mL oxygen/minute × body surface area, and the arterial-venous oxygen difference is determined by first calculating the oxygen-carrying capacity of blood (hemoglobin [g/100 mL] × 1.36 [mL oxygen/g hemoglobin] × 10) and multiplying this product by the fractional oxygen saturation. The indicator dilution method measures the concentration of a substance that is injected proximally, adequately mixes with blood, and is then sampled distally. In contemporary practice, thermodilution cardiac outputs are measured using temperature as the indicator. Measurements are made with a thermistor-tipped catheter that detects temperature deviations in the pulmonary artery after the injection of 10 mL of room-temperature normal saline into the right atrium.

Vascular Resistance Resistance across the systemic and pulmonary circulations is calculated by extrapolating from Ohm's law of electrical resistance and is equal to the mean pressure gradient divided by the mean flow (cardiac output). Therefore, systemic vascular resistance is ([mean aortic pressure – mean right atrial pressure]/cardiac output) multiplied by 80 to convert the resistance from Wood units to dyn-s-cm^{-5}. Similarly, the pulmonary vascular resistance is ([mean pulmonary artery – mean pulmonary capillary wedge pressure]/cardiac output) × 80. Pulmonary vascular resistance is lowered by oxygen, nitroprusside, calcium channel blockers, prostacyclin infusions, and inhaled nitric oxide; these therapies may be administered during catheterization to determine if increased pulmonary vascular resistance is fixed or reversible.

Valve Area Hemodynamic data may also be used to calculate the valve area using the Gorlin formula that equates the area to the flow

TABLE 249-3 Hemodynamic Findings in Tamponade, Constrictive Pericarditis, and Restrictive Cardiomyopathy

	CARDIAC TAMPONADE	CONSTRICTIVE PERICARDITIS	EFFUSIVE-CONSTRICTIVE PERICARDITIS	RESTRICTIVE CARDIOMYOPATHY
Pericardial pressure	↑	↑	↑	Normal
Right atrium pressure	↑	↑	↑ (Fails to decrease by 50% or to <10 mmHg after pericardiocentesis)	↑
Right atrium pressure waveform	Prominent “x” descent Diminished or absent “y” descent	Prominent “x” descent Prominent “y” descent	Prominent “x” descent “y” descent less prominent than expected	Prominent “y” descent
Right ventricle systolic pressure	<50 mmHg	<50 mmHg	<50 mmHg	>60 mmHg
Right ventricle end-diastolic pressure	Equals left ventricular end-diastolic pressure within 5 mmHg	>1/3 right ventricular systolic pressure Equals left ventricular end-diastolic pressure within 5 mmHg	>1/3 right ventricular systolic pressure Equals left ventricular end-diastolic pressure within 5 mmHg	<1/3 right ventricular systolic pressure Less than left ventricular end-diastolic pressure by ≥5 mmHg
Right ventricle pressure waveform		Dip and plateau or “square root” sign	Dip and plateau or “square root” sign	Dip and plateau or “square root” sign
Right ventricle–left ventricle systolic pressure relationship with inspiration	Discordant	Discordant	Discordant	Concordant

across the valve divided by the pressure gradient between the cardiac chambers surrounding the valve. The formula for the assessment of valve area is: Area = (cardiac output [cm^3/min]/[systolic ejection period or diastolic filling period][heart rate])/44.3 C × square root of the pressure gradient, where C = 1 for aortic valve and 0.85 for the mitral valve. A valve area of <1.0 cm^2 and a mean gradient of >40 mmHg indicate severe aortic stenosis, while a valve area of <1.5 cm^2 and a mean gradient >5–10 mmHg are consistent with moderate-to-severe mitral stenosis; in symptomatic patients with a mitral valve area >1.5 cm^2, a mean gradient >15 mmHg, pulmonary artery pressure >60 mmHg, or a pulmonary artery wedge pressure >25 mmHg after exercise is also considered significant and may warrant intervention. The modified Hakki formula has also been used to estimate aortic valve area. This formula calculates the valve area as the cardiac output (L/min) divided by the square root of the pressure gradient. Aortic valve area calculations based on the Gorlin formula are flow-dependent, and therefore, for patients with low cardiac outputs, it is imperative to determine if a decreased valve area actually reflects a fixed stenosis or is overestimated by a low cardiac output and stroke volume that is insufficient to open the valve leaflets fully. In these instances, cautious hemodynamic manipulation using dobutamine to increase the cardiac output and recalculation of the aortic valve area may be necessary.

Intracardiac Shunts In patients with congenital heart disease or unexplained hypoxemia, detection, localization, and quantification of the intracardiac shunt should be evaluated. A shunt should be suspected when there is unexplained arterial desaturation or increased oxygen saturation of venous blood. A "step up" or increase in oxygen content indicates the presence of a left-to-right shunt while a "step down" indicates a right-to-left shunt. The shunt is localized by detecting a difference in oxygen saturation levels of 5–7% between adjacent cardiac chambers. The severity of the shunt is determined by the ratio of pulmonary blood flow (Q_p) to the systemic blood flow (Q_s), or Q_p/Q_s = ([systemic arterial oxygen content − mixed venous oxygen content]/pulmonary vein oxygen content − pulmonary artery oxygen content). For an atrial septal defect, a shunt ratio of 1.5 is considered significant and factored with other clinical variables to determine the need for intervention. When a congenital ventricular septal defect is present, a shunt ratio of ≥1.5 with evidence of left ventricular volume overload is a strong indication for surgical correction.

VENTRICULOGRAPHY AND AORTOGRAPHY

Ventriculography to assess left ventricular function may be performed during cardiac catheterization. A pigtail catheter is advanced retrograde across the aortic valve into the left ventricle, and 30–45 mL of contrast is power-injected to visualize the left ventricular chamber during the cardiac cycle. The ventriculogram is usually performed in the right anterior oblique projection to examine wall motion and mitral valve function. Normal wall motion is observed as symmetric contraction of all segments; hypokinetic segments have decreased contraction, akinetic segments do not contract, and dyskinetic segments appear to bulge paradoxically during systole (**Fig. 249-3**). Ventriculography may also reveal a left ventricular aneurysm, pseudoaneurysm, or diverticulum, or an apical ballooning pattern consistent with takotsubo cardiomyopathy, and can also be used to assess mitral valve prolapse and the severity of mitral regurgitation. The degree of mitral regurgitation is estimated by comparing the density of contrast opacification of the left atrium with that of the left ventricle. Minimal contrast reflux into the left atrium is considered 1+ mitral regurgitation, while contrast density in the left atrium that is greater than that in the left ventricle with reflux of contrast into the pulmonary veins within three beats defines 4+ mitral regurgitation. When it takes more than three beats but fully fills the atrium, it is 3+. Both 3+ and 4+ are considered severe regurgitation. Ventriculography performed in the left anterior oblique projection can be used to identify a ventricular septal defect. Calculation of the ventricular volumes in systole and diastole allows calculation of stroke volume and cardiac output.

Aortography in the cardiac catheterization laboratory visualizes abnormalities of the ascending aorta, including aneurysmal dilation and involvement of the great vessels, as well as dissection with compression of the true lumen by an intimal flap that separates the true and false lumina. Aortography can also be used to identify patent saphenous vein grafts that elude selective cannulation, identify shunts that involve the aorta such as a patent ductus arteriosus, evaluate the takeoff or proximal anatomy of the great vessels, and provide a qualitative assessment of aortic regurgitation using a 1+–4+ scale similar to that used for mitral regurgitation.

CINEFLUOROSCOPY OF PROSTHETIC MECHANICAL VALVES

Prosthetic valve leaflet dysfunction may occur as a result of thrombus or obstruction of leaflet excursion by pannus (**Fig. 249-4**). The incidence of prosthetic valve thrombosis in left-sided valves is 0.1–6.0% per patient-year with differences in rates attributable to valve type, position, anticoagulation status, and left ventricular function. Prosthetic valve dysfunction should be suspected in patients with subtherapeutic anticoagulation with a low mean INR, a prothrombotic state, recent-onset heart failure, cardiogenic shock, cardiac arrest, thromboembolic

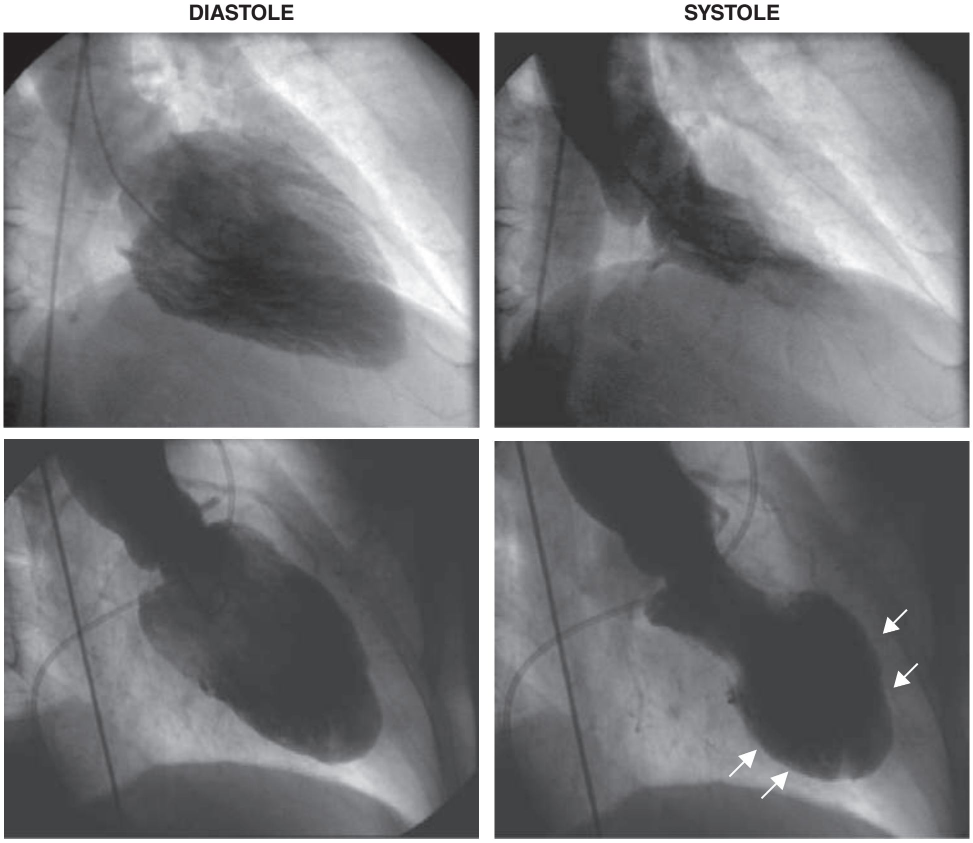

FIGURE 249-3 Left ventriculogram at end diastole (left) and end systole (right). In patients with normal left ventricular function, the ventriculogram reveals symmetric contraction of all walls (*top*). Patients with coronary artery disease may have wall motion abnormalities on ventriculography as seen in this 60-year-old male following a large anterior myocardial infarction. In systole, the anterior, apical, and inferior walls are akinetic (*white arrows*) (*bottom*).

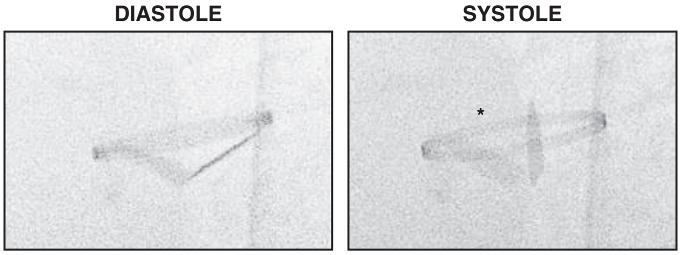

FIGURE 249-4 Cinefluoroscopic detection of mechanical valve leaflet dysfunction. Images of a bileaflet mechanical valve in the aortic position taken during diastole (*left*) and systole (*right*) show that one leaflet opens normally during systole while the other leaflet (below asterisk) remains immobile and fixed consistent with valve leaflet thrombosis.

event, or, in asymptomatic patients, an increasing gradient across the valve. Cinefluoroscopy visualizes the motion of mechanical valve leaflets, is noninvasive, is available in most centers, and can be performed rapidly with minimal radiation exposure. Prosthetic mechanical valves should be imaged *en face* and at a 90° angle over several cardiac cycles to document opening and closing of the valve leaflets as well as motion of the base ring. Each type of prosthetic valve has leaflet opening and closing angles that are reported by the manufacturer and can be used to determine if movement or closure of the valve leaflets is restricted, suggestive of mechanical obstruction.

CORONARY ANGIOGRAPHY

Selective coronary angiography is almost always performed during cardiac catheterization and is used to define the coronary anatomy and determine the extent of epicardial coronary artery and coronary artery bypass graft disease. Specially shaped coronary catheters are used to engage the left and right coronary ostia. Injection of radiopaque contrast agents by hand or automated contrast injector system creates a coronary "luminogram" that is recorded as radiographic images (cine angiography). Because the coronary arteries are three-dimensional objects that are in motion with the cardiac cycle, angiograms of the vessels using several different orthogonal projections are taken to best visualize the vessels without overlap or foreshortening.

The normal coronary anatomy is highly variable between individuals, but, in general, there are two coronary ostia and three major coronary vessels—the left anterior descending, the left circumflex, and the right coronary arteries with the left anterior descending and left circumflex arteries arising from the left main coronary artery **(Fig. 249-5)**. When the right coronary artery is the origin of the atrioventricular nodal branch, the posterior descending artery, and the posterior lateral vessels, the circulation is defined as right dominant; this is found in ~85% of individuals. When these branches arise from the left circumflex artery, as occurs in ~5% of individuals, the circulation is defined as left dominant. The remaining ~10% of patients have a codominant circulation with the posterior descending vessel arising from both the right coronary and the posterior lateral vessels from left coronary circulation. In some patients, a ramus intermedius branch arises directly from the left main coronary artery that trifurcates into the left anterior descending, ramus, and circumflex arteries; this finding is a normal variant. Coronary artery anomalies occur in 1–2% of patients, with separate ostia for the left anterior descending and left circumflex arteries (0.41%), ectopic origin of the right coronary artery from the left coronary sinus (0.92%), ectopic origin of the left circumflex artery from the right coronary sinus (0.67%), and fistulas between the coronary arteries and cardiac chambers or the pulmonary artery or vein (0.87%) being the most common. Certain anomalous coronary artery patterns increase the risk of sudden cardiac death, especially if the artery travels between the pulmonary artery and aorta.

Coronary angiography visualizes coronary artery stenoses as luminal narrowings on the cine angiogram. The degree of narrowing is referred to as the percent stenosis and is determined visually by comparing the most severely diseased segment with a proximal or distal "normal segment;" a stenosis >50% is considered significant **(Fig. 249-6)**. Online quantitative coronary angiography can provide a more accurate assessment of the percent stenosis and lessen the tendency to overestimate lesion severity visually. The presence of a myocardial bridge, which most commonly involves the left anterior descending artery, may be mistaken for a significant stenosis; this occurs when a portion of the vessel dips below the epicardial surface into the myocardium and is subject to compressive forces during ventricular systole. The key to differentiating a myocardial bridge from a fixed stenosis is that the "stenosed" part of the vessel returns to normal during diastole. Coronary angiography may also reveal spontaneous coronary artery dissections that appear similar to a stenosis but tends to occur in an angiographically normal-appearing coronary artery; a linear dissection flap may be present. Coronary spasm appears similar to a fixed stenosis but resolves with the administration of intracoronary nitroglycerin. Coronary calcification is also seen during angiography prior to the injection of contrast agents. Collateral blood vessels may be seen traversing from one vessel to the distal vasculature of a severely stenosed or totally occluded vessel. Thrombolysis in myocardial infarction (TIMI) flow grade, a measure of the relative duration of time that it takes for contrast to opacify the coronary artery fully, may provide an additional clue to the degree of lesion severity, and the presence of TIMI grade 1 (minimal filling) or 2 (delayed filling) suggests that a severe coronary artery stenosis is present.

INTRAVASCULAR ULTRASOUND, OPTICAL COHERENCE TOMOGRAPHY, FRACTIONAL FLOW RESERVE, AND CORONARY FLOW RESERVE

During coronary angiography, intermediate stenoses (40–70%), indeterminate findings, or anatomic findings that are incongruous with the patient's symptoms may require further interrogation. In these cases, intravascular ultrasound (IVUS) provides a more accurate anatomic

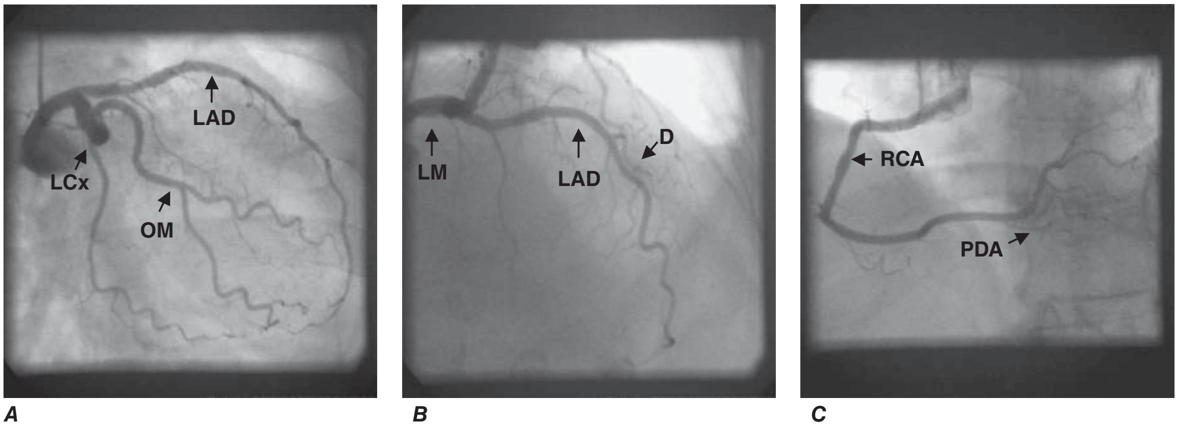

FIGURE 249-5 Normal coronary artery anatomy. ***A.*** Coronary angiogram showing the left circumflex (LCx) artery and its obtuse marginal (OM) branches. The left anterior descending (LAD) artery is also seen but may be foreshortened in this view. ***B.*** The LAD and its diagonal (D) branches are best seen in cranial views. In this angiogram, the left main (LM) coronary artery is also seen. ***C.*** The right coronary artery (RCA) gives off the posterior descending artery (PDA), so this is a right dominant circulation.

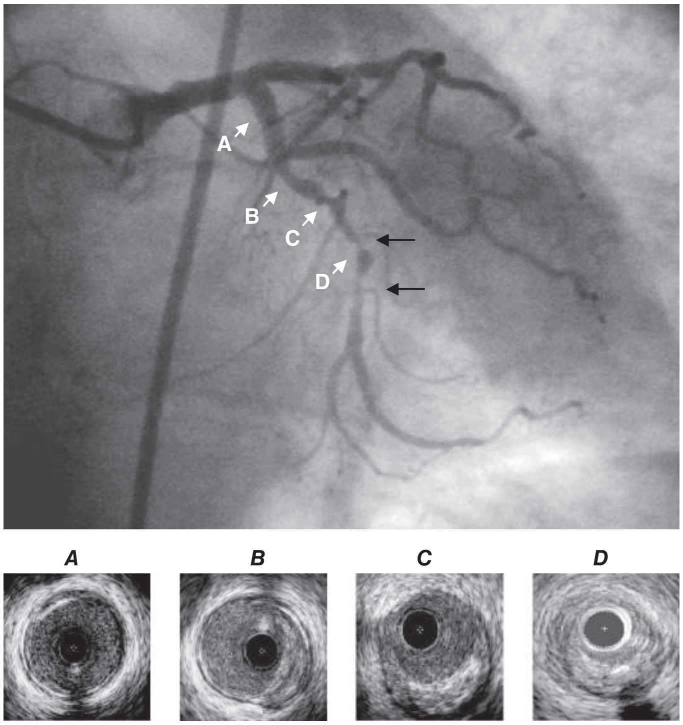

FIGURE 249-6 Coronary stenoses on cine angiogram and intravascular ultrasound. Significant stenoses in the coronary artery are seen as narrowings (*black arrows*) of the vessel. Intravascular ultrasound shows a normal segment of artery (***A***), areas with eccentric plaque (***B, C***), and near total obliteration of the lumen at the site of the significant stenosis (***D***). Note that the intravascular ultrasound catheter is present in the images as a black circle.

assessment of the coronary artery and the degree of coronary atherosclerosis (Fig. 249-6). IVUS is performed using a small flexible catheter with a transducer at its tip that is advanced into the coronary artery over a guidewire. Data from IVUS studies may be used to image atherosclerotic plaque precisely, determine luminal cross-sectional area, characterize plaque, including quantifying the arc and length of calcium, and measure vessel size. It is also used during or following percutaneous coronary intervention to assess the stenosis and determine the adequacy of the stent implantation procedure (stent expansion, apposition, geographic miss, or edge dissection). Optical coherence tomography (OCT) is a catheter-based imaging technique that uses near-infrared light to generate images with better spatial resolution than IVUS (10–15 microns vs 100–150 microns); however, the depth of field is smaller. The advantage of OCT imaging over IVUS lies in its ability to image characteristics of the atherosclerotic plaque (lipid, fibrous cap, plaque rupture or erosion, thrombus) with high definition and to assess coronary stent placement, apposition, and patency **(Fig. 249-7)**.

Measurement of the fractional flow reserve provides a functional assessment of the stenosis and is more accurate in predicting long-term clinical outcome than imaging techniques. The fractional flow reserve is the ratio of the pressure in the coronary artery distal to the stenosis divided by the pressure in the artery proximal to the stenosis at maximal vasodilation. Fractional flow reserve is measured using a coronary pressure–sensor guidewire at rest and at maximal hyperemia following the infusion of adenosine **(Fig. 249-8)**. A fractional flow reserve of <0.80 indicates a hemodynamically significant stenosis that would benefit from intervention. The instantaneous wave-free ratio, which measures the gradient across the stenosis during the latter part of diastole, does not require the use of adenosine and may be preferred for some patients with asthma or documented allergy to adenosine. An instantaneous wave-free ratio of <0.89 is considered positive for ischemia. Resting gradients have also been shown to predict a significant stenosis.

Microvascular dysfunction can be evaluated by assessing coronary flow reserve, the ratio between coronary blood flow at maximal hyperemia and rest. Coronary flow reserve is measured using a Doppler wire– or pressure wire–based thermodilution technique in patients with unexplained chest pain or ischemia and no obstructive coronary artery disease. A coronary flow reserve <2.0 is considered abnormal. An index of microcirculatory resistance can also be calculated by multiplying the pressure wire thermodilution-derived mean transit time by the distal pressure. An index of microcirculatory resistance of <25 is considered normal. Studies have shown this derivative parameter to be an important predictor of outcome as well.

POSTPROCEDURE CARE

Once the procedure is completed, vascular access sheaths are removed. If the femoral approach is used, direct manual compression or vascular closure devices that immediately close the arteriotomy site with a staple/clip, collagen plug, or sutures are used to achieve hemostasis. These devices decrease the length of supine bed rest (from 6 h to 2–4 h) and improve patient satisfaction but have not been shown definitively to be superior to manual compression with respect to access-site complications. The overall risk of complications for vascular closure devices is low, but potential complications include infection, embolization, and limb ischemia. Vascular closure devices should be avoided in heavily diseased vessels where the risk of complications is increased. With radial-artery access, the sheath is removed and a plastic wristband with an air pillow is used to keep pressure on the access site while maintaining flow through the radial artery. Bed rest is needed for only 2 hours while the air in the wristband is released. When cardiac catheterization is performed as an elective outpatient procedure, the patient completes postprocedure bed rest in a monitored setting and is discharged home with instructions to liberalize fluids because contrast agents promote an osmotic diuresis, to avoid strenuous activity, and to observe the vascular access site for signs of complications. Overnight hospitalization may be required for high-risk patients with significant comorbidities, patients with complications occurring during the catheterization, or patients who have undergone a complicated percutaneous coronary intervention. Hypotension early after the procedure may be due to inadequate fluid replacement or retroperitoneal bleeding from the femoral access site. Patients who received >2 Gy of radiation during the procedure should be examined for signs of erythema. For patients who received higher doses (>5 Gy), clinical follow-up within 1 month to assess for skin injury is recommended.

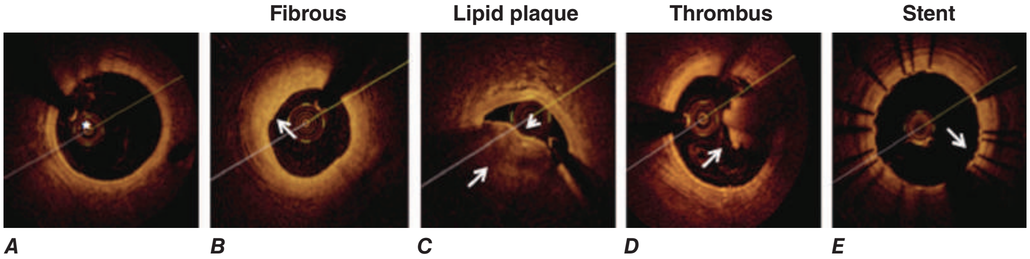

FIGURE 249-7 Optical coherence tomography imaging. ***A.*** The optical coherence tomography (OCT) catheter (*) in the lumen of a coronary artery with limited neointima formation. The intima is seen with high definition, but unlike intravascular ultrasound imaging, the vessel media and adventitia are not well visualized. ***B.*** A fibrous plaque (*arrow*) is characterized by a bright signal. ***C.*** A large, eccentric, lipid-rich plaque obscures part of the vessel lumen. Because lipid in the plaque absorbs light, the lipid-rich plaque appears as a dark area with irregular borders (*arrow*). The plaque is covered by a thin fibrous cap (*arrowhead*) typical of a vulnerable plaque. ***D.*** A thrombus (*arrow*) adherent to a ruptured plaque that is protruding into the vessel lumen. ***E.*** A coronary stent is that is well apposed to the vessel wall. The stent struts appear as short bright lines with dropout behind the struts (*arrow*).

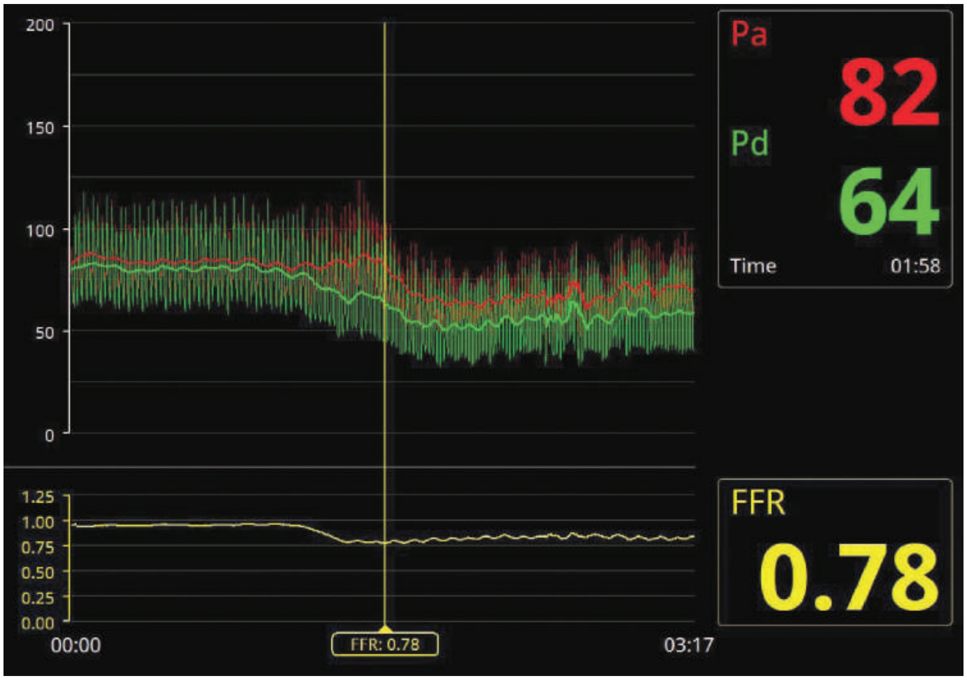

FIGURE 249-8 Fractional flow reserve. The fractional flow reserve is measured using a coronary pressure-sensor guidewire that measures the ratio of the pressure in the coronary artery distal to the stenosis (Pd, *green*) divided by the pressure in the artery proximal to the stenosis (Pa, *red*) at maximal hyperemia following the injection of adenosine. A fractional flow reserve of <0.80 indicates that revascularization would be beneficial.

FURTHER READING

Bangalore S et al: Evidence-based practices in the cardiac catheterization laboratory. Circulation 144:e107, 2021.

Moscucci M (ed): *Grossman & Baim's Cardiac Catheterization, Angiography, and Intervention*, 9th ed. Philadelphia, Lippincott Williams & Wilkins, 2020.

Nishimura R et al: Hemodynamics in the cardiac catheterization laboratory of the 21st century. Circulation 125:2138, 2012.

Räber L et al: Clinical use of intracoronary imaging. Part 1: guidance and optimization of coronary interventions. An expert consensus document of the European Association of Percutaneous Cardiovascular Interventions. Eur Heart J 39:3281, 2018.

Samuels BA et al: Comprehensive management of angina with nonobstructive coronary artery disease (ANOCA), Part 1 definition, patient population, and diagnosis. J Am Coll Cardiol 82:1245, 2023.

Section 3 Disorders of Rhythm

250 Principles of Clinical Cardiac Electrophysiology

William H. Sauer, Bruce A. Koplan, Paul C. Zei

HISTORICAL PERSPECTIVE

Clinical cardiac electrophysiology is the subspecialty of cardiology that focuses on the study and management of heart rhythm disorders. The development of the modern surface electrocardiogram (ECG) by Willem Einthoven more than 100 years ago enabled an understanding of the relationship between cardiac electrical potentials, mechanical cardiac function, and pathophysiology of cardiac arrhythmias. In the mid-twentieth century, the recording of cellular membrane currents enabled the understanding that the surface ECG represents the sum of cellular cardiac electrical activity. An understanding of cellular electrophysiology also ushered in the development of antiarrhythmic drugs utilized by cardiac electrophysiologists.

The modern era of clinical cardiac electrophysiology began with the first recordings of human intracardiac electrograms in the 1960s. Initially, invasive electrophysiology studies were limited to diagnostic tools. This included serial electrophysiologic testing to evaluate arrhythmia mechanisms and evaluate arrhythmia suppression by antiarrhythmic drugs, and programmed stimulation of the ventricle for risk stratification of sudden cardiac death. In the 1960s and 1970s, cardiac surgery was the only available invasive treatment for cardiac arrhythmias. The subsequent development of radiofrequency catheter ablation in the 1980s ushered in the era of interventional cardiac electrophysiology. In addition, with the development of implanted cardiac rhythm management devices including pacemakers and defibrillators, clinical cardiac electrophysiology became a distinct medical subspecialty. Developments in catheter ablation techniques, cardiac resynchronization and conduction system pacing, subcutaneous defibrillator implantation, leadless pacemaker implantation, left atrial appendage closure, and laser-assisted lead extraction have broadened the procedural aspects of the specialty over the past 30 years; however, the principles of arrhythmia patient management have remained the same.

CELLULAR ELECTROPHYSIOLOGY

The cardiac action potential (AP) drives the electrophysiologic behavior of all cardiac myocytes. The AP is characterized morphologically by five distinct phases, termed phases 0–4, as shown in Fig. 250-1. Moreover, as ventricular electrophysiologic activity accounts for the QRS and T complexes of the surface ECG, each AP phase in ventricular tissues corresponds to distinct phases in the surface ECG: Phase 0, the rapid upstroke, corresponds to the QRS deflection; phases 1–2 account for the ST segment; phase 3 accounts for the T wave; while phase 4 corresponds to the segment between the end of the T wave and the subsequent QRS deflection. In addition, the P wave corresponds to atrial depolarization, while the PR interval corresponds to the time between

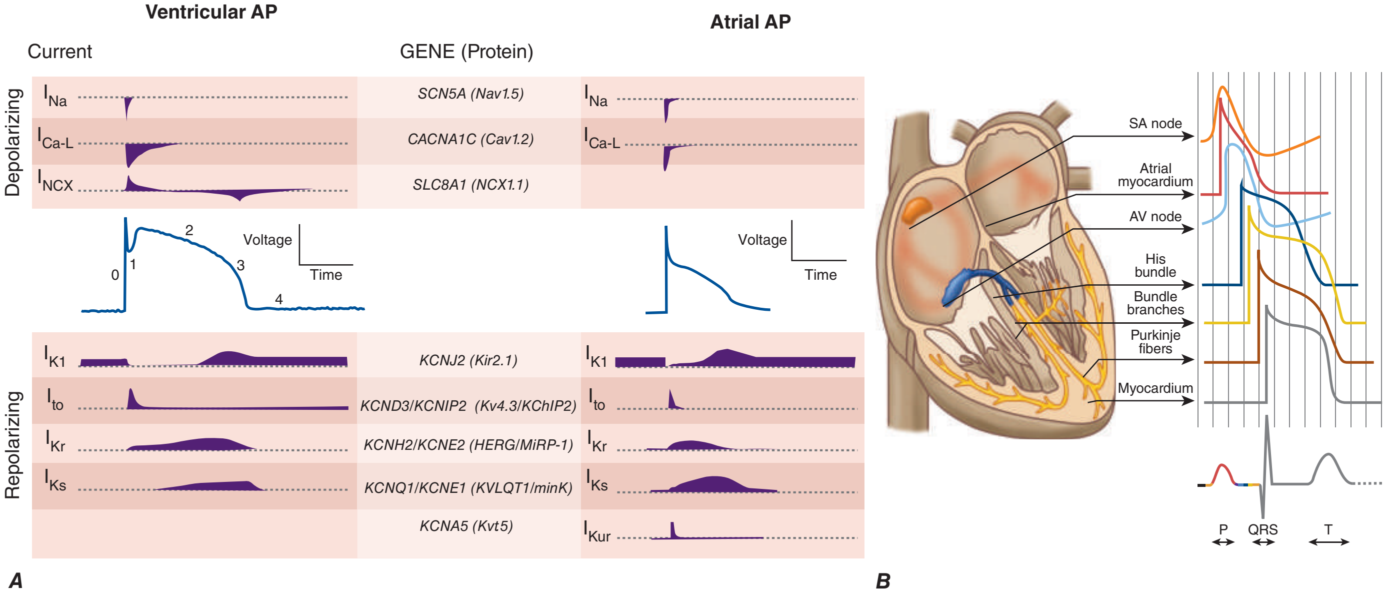

FIGURE 250-1 ***A.*** Cellular atrial and ventricular action potentials. Phases 0–4 are the rapid upstroke, early repolarization, plateau, late repolarization, and diastole, respectively. The ionic currents and their respective genes are shown above and below the action potentials. The currents that underlie the action potentials vary in atrial and ventricular myocytes. Potassium current (I_{K1}) is the principal current during phase 4 and determines the resting membrane potential of the myocyte. Sodium current generates the upstroke of the action potential (phase 0); activation of I_{to} with inactivation of the Na current inscribes early repolarization (phase 1). The plateau (phase 2) is generated by a balance of repolarizing potassium currents and depolarizing calcium current. Inactivation of the calcium current with persistent activation of potassium currents (predominantly I_{Kr} and I_{Ks}) causes phase 3 repolarization. Currents that result in membrane depolarization are grouped at the top of the figure above the action potentials, while repolarizing currents are shown below the action potentials. ***B.*** A surface electrocardiogram (ECG) representation of sinus rhythm is shown with respective intracardiac action potentials that are active during each phase of the ECG. Each cardiac conduction region's action potential is shown in the upper portion of the panel, with colors reflected in the ECG segment shown in the lower portion of the panel. Note that during the P wave, atrial depolarization is active. During the PR interval, the atrioventricular (AV) nodal, His, bundle branches, and Purkinje fibers are active (in sequence), although these action potentials are not discernible on the surface ECG. During the QRS interval, ventricular action potentials are active, with the QRS morphology most reflective of the sequence of ventricular tissue action potential activation. The ST segment is predominantly determined by the plateau phase 2 of the ventricular action potential. The T wave is determined largely by ventricular repolarization (phase 3), while the isoelectric segment is the result of the electrically neutral phase 4 of the ventricular action potential.

the initiation of atrial depolarization to the initiation of ventricular depolarization, comprised (typically) for the most part by the conduction time through the atrioventricular (AV) node.

AP morphologies are the result of the precise and carefully timed sequences of opening, closing, and inactivation of an array of membrane ion channels in response to cellular membrane potential changes, ligands that bind to the ion channel complex, or membrane stretch in a time-dependent fashion. The open ion channel allows flux of specific charged ions through a central pore, resulting in electrical (ionic) currents that drive the AP. The activity of different subsets of ion channels drives the different phases of the AP. Specific ionic currents that flux through an open channel are driven by the electrochemical gradient of that particular ion across the membrane, which in turn are driven by ion pumps or transporters/exchangers, which in turn are catalyzed by ATP **(Fig. 250-2)**.

Ion channels are complex, multi-subunit transmembrane glycoproteins that contain a central pore that is selective for particular ionic species (selectivity); a "gating" apparatus that regulates the opening, closing, and inactivation apparatus; and often one or more regulatory subunits. Most channels gate in response to changes in membrane potential, a specific ligand, or mechanical deformation. The molecular underpinnings of these specific functional properties of channels have become well understood through decades of basic electrophysiologic study using the tools of voltage clamp and patch clamp techniques, and more recently, molecular, genetic, and structural/crystallographic techniques.

The structural makeup of most ion channels contains several common motifs. All channels form a central conducting pore, with ionic selectivity determined by specific amino acids that line the central pore. The central pore of most channels is formed by the P domain, a series of hydrophilic amino acid residues, with one of several structural variants: four separate homologous alpha subunits, each with homologous P domains (voltage-gated K channels); a single alpha subunit with four internally homologous P domains (voltage-gated Na or Ca channels); or two internally homologous P domains from two separate subunits (most ligand-gated K channels). A series of one or more transmembrane segments surrounds the central pore. In voltage-gated channels, the fourth of six segments, the S4 segment, contains a series of charged amino acid residues that functions as a voltage sensor, responding to changes in membrane potential by facilitating protein conformational changes that result in channel opening or closing (gating). In ligand-activated channels, the binding of a ligand (transmitters, molecules, or other ions) results in channel opening or closing, while deformations in membrane shape determine gating in stretch-activated channels. In addition, in many ion channels, a complex of auxiliary proteins is associated with the primary alpha subunit; most auxiliary subunits appear to facilitate regulation of ion channel expression and activity. A distinct type of transmembrane protein complex is the gap junction complex. A large multimeric complex of connexin subunits forms a large, nonselective pore that spans and thereby connects adjacent myocytes. This allows free flux of ions between adjacent myocytes, facilitating impulse propagation across myocardial tissues.

Due to the physiologic gradient of their respective ions across the cell membrane, Na and Ca channels account for most inward, or depolarizing, currents in cardiac myocytes, and these channels respond to membrane depolarization with rapid opening, relatively rapid closing, and inactivation. Na and Ca currents therefore drive phase 0 depolarization of the AP. Potassium channels, on the other hand, account for most of the repolarizing currents seen in cardiac myocytes. Relatively slow K channel opening, as well as Na and Ca channel closing and inactivation, drives the plateau of phases 1–2 as well as the repolarizing phase 3 of the AP. Mutations in K channel subtypes are causative of many inherited channelopathies. Mutations that either inherently delay the closing or inactivation of K channels result in prolongation of the QT interval, leading to many forms of inherited long QT syndrome.

The morphologic and functional properties of APs vary across different regions of the heart. These variations are the result of variations in the active ionic currents during each phase of the AP, which in turn reflects regional variation in ion channel expression. In atrial and ventricular myocytes, Na currents dominate the rapid upstroke

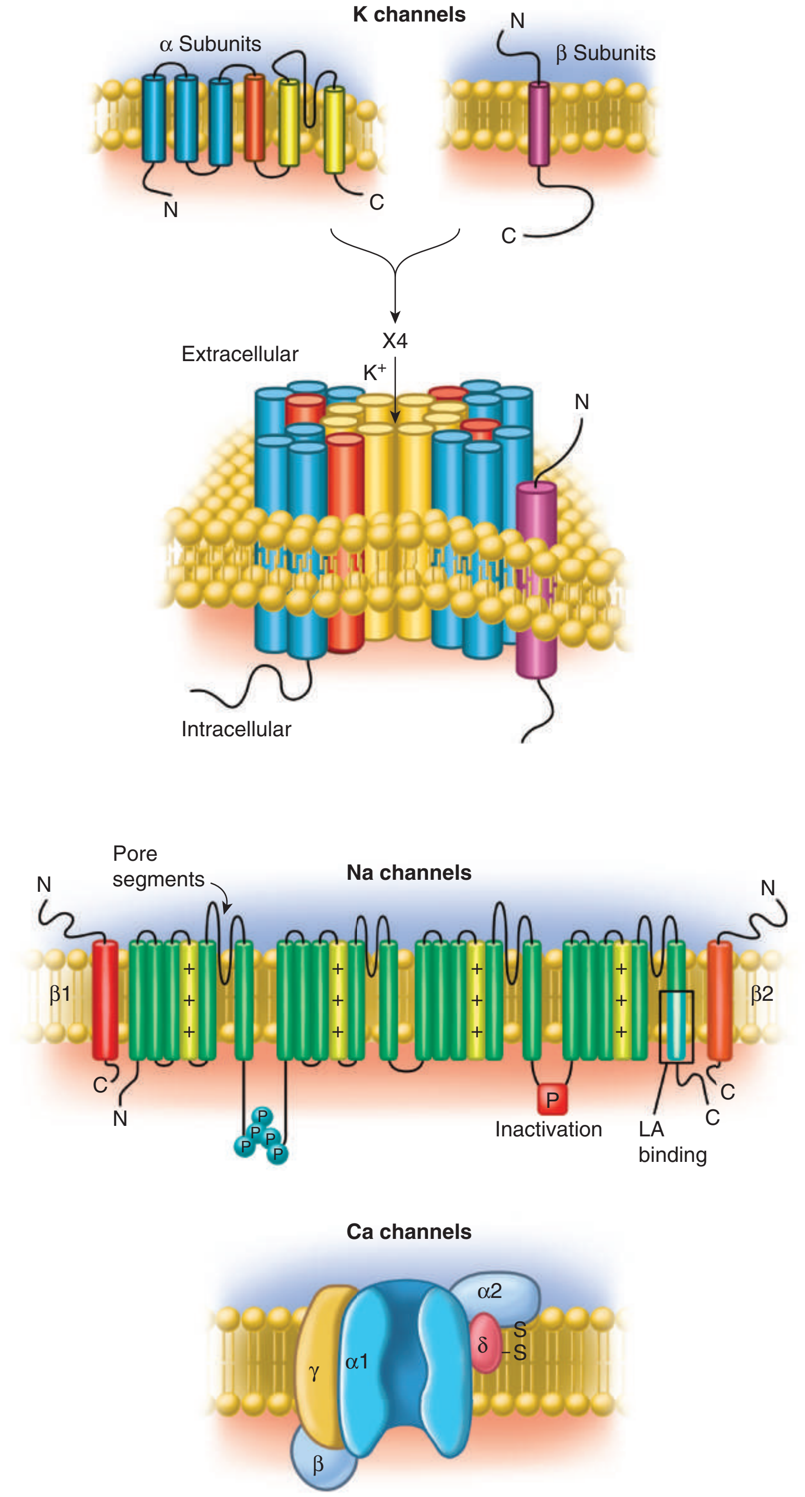

FIGURE 250-2 Topology and subunit composition of the voltage-dependent ion channels. Potassium channels are formed by the tetramerization of α or pore-forming subunits and one or more β subunits; only single β subunits are shown for clarity. Sodium and calcium channels are composed of α subunits with four homologous domains and one or more ancillary subunits. In all channel types, the loop of protein between the fifth and sixth membrane-spanning repeat in each subunit or domain forms the ion-selective pore. In the case of the sodium channel, the channel is a target for phosphorylation, the linker between the third and fourth homologous domain is critical to inactivation, and the sixth membrane-spanning repeat in the fourth domain is important in local anesthetic antiarrhythmic drug binding. The Ca channel is a multi-subunit protein complex with the α_1 subunit containing the pore and major drug-binding domain.

(phase 1) of the AP, while in nodal tissues, Ca currents, which activate more slowly, dominate phase 1. Hence, for instance, drugs that bind and block the cardiac Na channel demonstrate efficacy in treating tachyarrhythmias arising from the atria and ventricles, whereas Ca channel blocking agents demonstrate efficacy at nodal tissues. During the pre-depolarizing phase 4 of the AP, ionic currents remain relatively quiescent in atrial and ventricular myocytes as they await local depolarization that triggers the next AP. In contrast, in sinus nodal tissues, which possess the property of automaticity, or intrinsic rhythmic depolarization, there is gradual depolarization observed during phase 4, until a threshold is reached that initiates the next AP. In these nodal tissues, this depolarizing phase 4 current is generated by a semiselective Na/Ca channel, termed the "funny current" or I_f, which is the target for the medication ivabradine.

NORMAL CARDIAC IMPULSE PROPAGATION

The normal cardiac impulse initiates and travels through specialized conduction fibers, often referred to as the cardiac conduction system. Each impulse is initiated in the sinoatrial (SA) node, located at the lateral junction between the superior vena cava (SVC) and right atrium (RA). SA nodal tissues exhibit automaticity, such that a reliable, rhythmic impulse emanates from the SA node. The SA node (along with the AV node) is richly innervated by autonomic fibers, allowing precise and dynamic control of heart rate and overall function by the central nervous system. The normal impulse then travels across the RA then the LA across preferential conduction pathways, initiating atrial systole. Once the impulse reaches the AV node, conduction occurs in a relatively slow time frame through the AV nodal tissues. This conduction time not only serves to provide physiologic AV synchrony but also is reflected in the surface ECG as the PR interval, or time between the atrial inscription and the subsequent ventricular, or QRS, complex. In normal hearts, the AV node serves as the only electrical connection between atria and ventricles. Both the SA and AV nodes respond exquisitely to autonomic input; for instance, with exercise and increased adrenergic tone, the PR interval physiologically shortens. After the AV node, the impulse travels through a network of specialized conduction fibers: the bundle of His divides into a right and left bundle branch, which transmit conduction to the right and left ventricles, respectively. The left bundle then divides further into the left anterior and posterior fascicles. The fascicles then further divide into a network of Purkinje fibers. The conduction velocity of electrical impulses is much higher in Purkinje fibers (2–3 m/s) than in myocardial cells (0.3–0.4 m/s). Different connexins in gap junctions of Purkinje networks are partially responsible for more rapid conduction. This network of conductive Purkinje fibers is located endocardially and serves to rapidly transmit depolarization throughout the ventricles, such that myocardial depolarization, and hence mechanical contraction, occur rapidly and in a coordinated, synchronized fashion, optimizing mechanical contraction of the ventricles. Repolarization of the ventricular myocardium, on the other hand, occurs relatively slowly and progresses from the epicardial surface back toward the endocardium. Hence, the T wave inscription in most ECG leads is concordant with the QRS complex.

MECHANISMS OF CARDIAC ARRHYTHMIAS

Cardiac arrhythmias are the manifestation of abnormalities in the initiation and/or propagation of the cardiac electrical impulse. Bradyarrhythmias result most commonly from abnormalities in the specialized conduction tissues. Abnormal function of the SA node may result in pathologic sinus bradycardia; AV node disease may result in conduction block; pathology in the His-Purkinje system may result in conduction block as well. Tachyarrhythmias may arise from not only

TABLE 250-1 Overview of the Mechanisms of Cardiac Tachyarrhythmias

TACHYARRHYTHMIA CATEGORY	MECHANISM	PROTOTYPICAL ARRHYTHMIAS
Abnormal automaticity	Enhanced (acceleration of phase 4 repolarization)	Idiopathic VT; AT
	Suppressed (absent or decelerated phase 4 repolarization)	Sinus node dysfunction
Triggered activity	EADs	TdP in long QT syndrome, PVCs
	DADs	Reperfusion PVCs/VT, AT and VT with digitalis toxicity
Reentry	(1) Anatomic or functional confinement of a circuit (i.e., scar, accessory pathway); (2) unidirectional block after a premature impulse; (3) wave of excitation that travels in a single direction returning to its point of origin	AVNRT, AVRT, atrial flutter, scar-related VT

Abbreviations: AT, atrial tachycardia; AVRT, atrioventricular reentry tachycardia; AVNRT, atrioventricular nodal reentry tachycardia; DADs, delayed afterdepolarizations; EADs, early afterdepolarizations; PVC, premature ventricular contraction; TdP, torsades des pointes; VT, ventricular tachycardia.

nearly every location within the conduction tissues, but also within atrial or ventricular tissues. Tachyarrhythmias are typically classified by mechanism: enhanced automaticity refers to abnormal spontaneous depolarization, which can occur along the conduction system, the atria, or ventricles; triggered arrhythmias result from abnormal afterdepolarizations that occur in either phase 2/3 (early afterdepolarizations) or phase 4 (delayed afterdepolarizations) of the AP; reentry results from circus movement of an electrical impulse **(see Table 250-1 and Fig. 250-3)**.

ENHANCED AUTOMATICITY

Automaticity, defined as spontaneous depolarizations occurring during phase 4 of the AP, is a normal property of several myocardial tissues, including the SA node, AV node, and the His-Purkinje system. The automaticity of the SA node triggers the normal cardiac impulse. When the automaticity of a more proximal conduction system tissue is unreliable or slow, the automaticity of a more distal aspect of the conduction system may result in an "escape rhythm" that may maintain cardiac output. Automaticity in these tissues results from phase 4 depolarization of cellular membranes driven by several ionic currents. In the SA node, the nonselective Na/Ca I_f current drives this depolarization, while in other tissues, K currents, Ca currents, or even the Na/Ca and ATP-driven Na/K exchangers contribute.

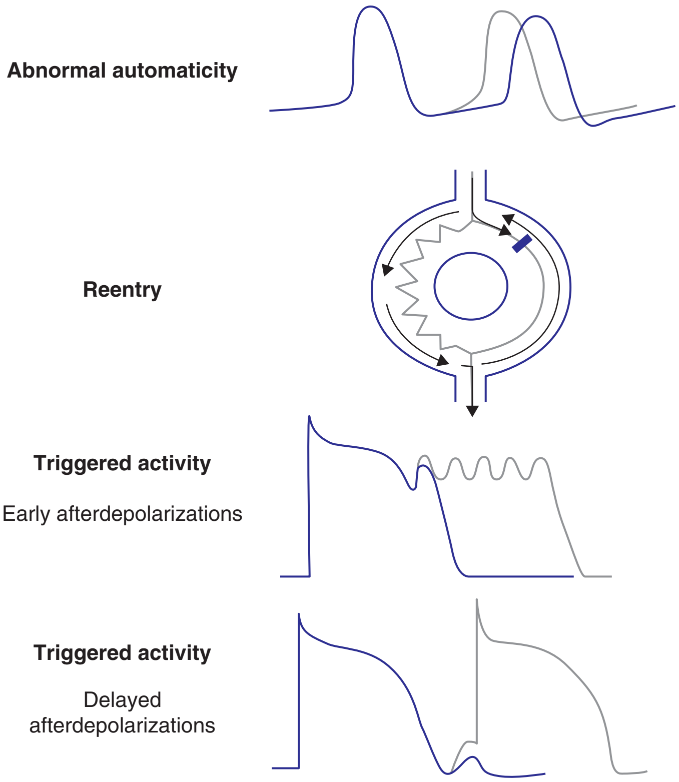

FIGURE 250-3 Schematic action potentials with early afterdepolarizations (EADs) and delayed afterdepolarizations (DADs).

The rate of depolarization during phase 4 drives the frequency of APs and hence automaticity rate of these tissues. In nodal tissues, this rate of depolarization is highly regulated by the autonomic system. Parasympathetic input results in local acetylcholine (ACh) release, which then binds the IKACh potassium channel complex (specifically via a G protein–mediated mechanism). The opening of IKACh channels, resulting in K efflux, hyperpolarizes these cells, resulting in slowing of phase 4 depolarization, thereby slowing the automaticity rate. Sympathetic input, via catecholamines, activates both alpha- and beta-adrenergic receptors. Beta-1 adrenergic stimulation results in activation of L-type Ca channels, Ca influx, and as a result, enhanced depolarization rates during phase 4, and increased automaticity rates. The normal range of SA automaticity rates is between 30 and 220 beats/min, corresponding to the normal range of rates during sinus rhythm. The sinus rate at any instant is therefore a dynamic balance between sympathetic and parasympathetic input, with the latter dominating in the restful state. The intrinsic heart rate (IHR) is defined as the "native" automaticity rate of the SA node, absent any autonomic input.

Abnormally enhanced automaticity may occur at any site that exhibits automaticity, including the SA node, AV node, or His-Purkinje system, resulting in pathologic tachycardia. In addition, in pathologic states, other stereotyped regions in the heart may exhibit enhanced automaticity, including the pulmonary veins, coronary sinus superior vena cava, and ventricular outflow tracts. Injury to myocardium, whether through ischemia or other mechanisms, may alter its cellular membrane properties, resulting in automaticity in these tissues. For instance, the border zones of infarcted ventricular myocardium, or rapidly reperfused ischemic myocardium, often exhibit automatic arrhythmias including premature ventricular contractions (PVCs) or automatic idioventricular rhythms (AIVR). Abnormal automaticity in the pulmonary veins is believed to underpin the triggers that drive paroxysmal atrial fibrillation, while automaticity elsewhere in the atria drives atrial tachycardias.

AFTERDEPOLARIZATIONS AND TRIGGERED ARRHYTHMIAS

Afterdepolarizations and triggered arrhythmias refer to abnormal depolarizations that occur in the late phases of the AP (afterdepolarizations) that can initiate sustained arrhythmias. Early afterdepolarizations (EADs) occur typically during phases 2–3 of the AP and may be facilitated by intracellular Ca loading. When the QT interval prolongs, typically in a heterogeneous fashion across the ventricles, EADs may trigger wavefronts of abnormal depolarizations, resulting in torsades des pointes (TdP), a nonsustained or sustained ventricular arrhythmia that may result in cardiac arrest. Medications that prolong QT interval, as well as other QT-prolonging factors including hypokalemia, hypomagnesemia, and bradycardia, predispose the ventricles to EADs, leading to TdP. Electrical remodeling in cardiomyopathies may also predispose to QT prolongation and risk of EADs and TdP.

Delayed afterdepolarizations (DADs) are abnormal depolarizations occurring in phase 4 of the AP. The mechanism underlying DADs is increased intracellular Ca, which then enhances repetitive depolarizations during the late phases of the AP. As a result, repetitive depolarizations ensue, including the well-described phenomenon of bidirectional ventricular tachycardia (VT). Digitalis glycoside toxicity, ischemia, and catecholamines are the most commonly described causes for DADs.

REENTRY

Reentry refers to the circus movement of a wavefront of electrical activation. Reentry can occur around a fixed anatomic barrier, referred to as anatomic reentry, or around a functionally blocked or refractory barrier or anchor, termed functional reentry. Initiation and maintenance of a reentrant arrhythmia require (1) unidirectional block, where the electrical wavefront can only propagate in one direction, and (2) slow conduction, a zone within the reentrant circuit where conduction is relatively slow, allowing the remainder of the circuit to repolarize and recover from refractoriness (the inability to reexcite).

The more common form of reentry is anatomic, which requires a defined electrical/anatomic circuit with a pathway around a fixed barrier. A wavefront of depolarization encounters a barrier to conduction that allows propagation in only one direction (unidirectional block), forcing activation preferentially along one limb or pathway. Due to slow conduction, the depolarization wavefront travels through the remaining circuit and continually encounters tissues that have recovered from refractoriness and are hence excitable. This results in perpetual circus movement. Moreover, if the total length of the circuit exceeds a distance determined by the product of the conduction velocity (theta) of the tissue and the refractory period (duration) of that tissue (tr), referred to as the wavelength of tachycardia (lambda = theta × tr), an excitable gap, where tissue is recovered from refractory and able to depolarize, is created, allowing reentry. Reentry is the mechanism for several clinically important and common cardiac arrhythmias, including atrial flutter, AV nodal reentry, AV reciprocating tachycardia utilizing an accessory pathway, and scar-based reentrant VT.

When reentry occurs in the absence of a fixed anatomic barrier, it is termed *functional reentry*. A nidus of partially refractory tissue anchors the depolarization wavefront, resulting in a circular or rotational reentrant wavefront. In this case, the reentrant circuit or activity tends to be less stable than that from anatomic reentry, resulting in variations in depolarization rate and propensity to easily terminate and/or reinitiate. There is evidence that functional reentry is the underlying mechanism for perpetuation and maintenance of both atrial fibrillation (AF) and ventricular fibrillation (VF). In both of these apparently chaotic and disorganized arrhythmias, multiple wavefronts resulting from multiple functional reentrant circuits appear to drive arrhythmia in many, if not most, instances. Underlying pathology of the myocardium resulting in heterogeneous electrophysiologic properties, altered activation, and repolarization properties predispose myocardial tissues to initiation and propagation of functional reentry-based arrhythmias.

In addition to intrinsic alterations in cellular membrane electrophysiologic properties that underpin most arrhythmias, extrinsic factors may precipitate other architectural and tissue changes that contribute to proarrhythmia. Ischemia and infarct may create regions of heterogeneous fibrosis, resulting in islands of scar surrounded by injured tissue. This creates the anatomic substrate that can sustain anatomic reentry, which underlies scar-based VT, as well as many macroreentrant atrial arrhythmias. Peri-infarct border zones often contain injured myocardium as well, and the resultant alterations in cellular membrane properties may promote enhanced automaticity or triggered arrhythmias. Chronic ischemia also results in downregulation of connexin proteins and gap junctions, resulting in slowed impulse propagation, which is one of the factors required for reentrant arrhythmias. Alterations in ion channel function, either through inherited mutations or through drug effect, can promote arrhythmias. QT prolongation can occur when the closing of potassium channels that should hyperpolarize cells is delayed or slowed, or when the closing or inactivation of Na channels is impaired.

■ UNDERPINNINGS OF THE TREATMENT OF ARRHYTHMIAS

Pharmacologic therapies for arrhythmias are directed toward the specific underlying mechanism. For enhanced automaticity-based arrhythmia, medications that target phase 4 depolarization, including Ca channel blockers, beta-adrenergic blockers (via indirect action on adrenergic input), or ivabradine may be used. For triggered activity-based arrhythmia, correcting the precipitating factor is most effective. This includes, among other therapies, removal of digitalis glycosides from the body, discontinuation of QT-prolonging medications, or even increasing heart rate, thereby shortening QT interval. For reentrant arrhythmia, medications that increase the refractory period, in particular K channel blocking agents, will increase the wavelength of conduction beyond the circuit length of tachycardia, resulting in the inability to sustain reentry. Medications that slow conduction velocity may have the paradoxical effect of promoting reentrant arrhythmias due to the creation of a larger excitable gap. This explains much of the proarrhythmic effect of many antiarrhythmic medications. Therefore, for these agents, which typically include Na channel blockers, sufficient dosing is required to slow conduction velocity to the point of extinguishing meaningful arrhythmia circuit conduction.

A major aspect of clinical cardiac electrophysiology that has evolved over several decades is the ability to disrupt arrhythmic substrates through catheter-based (or rarely surgical) myocardial ablation. For automaticity-based arrhythmias, precise localization and elimination or isolation through ablation of the site of focal automaticity is effective in eliminating arrhythmia. For anatomically bound reentrant arrhythmias, interruption of the reentrant circuit with a series of ablation lesions is effective. In contrast, given the lack of a fixed anatomic circuit, and perhaps also due to the presence of multiple, often migratory, circuits, it appears that mechanical disruption, typically through catheter-based ablation, of identified sites of functional reentry appears to be ineffective in eliminating arrhythmia.

CARE OF THE ARRHYTHMIA PATIENT

■ EVALUATION AND DIAGNOSIS

The evaluation of a patient with suspected arrhythmia begins with a directed history and physical examination, which must include an ECG. The history and examination should focus on determining the nature of symptoms attributable to the arrhythmia itself and clues to potential underlying cardiac, medical, or metabolic conditions that may predispose to specific arrhythmias, and hence direct further studies and evaluations, ultimately directing appropriate therapy, prognosis, and counseling. Family history may also provide clues toward possible inherited arrhythmia syndromes. Symptoms attributable to arrhythmia can vary from a vague sensation of fatigue, chest pain, dyspnea, or lightheadedness to more specific sensations of rapid, slow, or irregular heart rate. Premature contractions, whether atrial or ventricular, may be sensed as extra beats, or if these extrasystoles result in diminished stroke volume for that particular beat, a sensation of a missed beat. Second, the hemodynamic sequelae of impaired cardiac output may result in symptoms, from presyncope to frank syncope, dyspnea, chest discomfort, or generalized weakness. Importantly, as the cadence and duration of arrhythmia episodes are highly variable, the temporal manifestations of arrhythmia symptoms may vary significantly. Sporadic episodes of arrhythmia will result in intermittent symptoms, including syncope if hemodynamic compromise is significant; protracted episodes of arrhythmia may cause persistent symptoms. In patients with underlying compromise in cardiac function, most typically in patients with structural heart disease, arrhythmia leading to diminished cardiac output may trigger or exacerbate symptoms associated with the underlying condition such as angina, congestive heart failure, or hypoxia-associated symptoms.

Inciting factors or associations may also provide clues to the diagnosis. Arrhythmias associated with activities that increase adrenergic tone, such as exercise, stimulant intake, or emotional stress, may suggest not only tachyarrhythmias but also automaticity-triggered arrhythmias. However, keep in mind that exceptions will occur. Medication use may be highly suggestive of an etiology: use of Ca channel blockers or beta blockers may suggest bradycardia exacerbated by these medications. Medications known to potentially prolong QT interval may suggest a malignant ventricular arrhythmia, specifically TdP. Eliciting a thorough family history, not only for known arrhythmia diagnoses, but of unexplained sudden death, may point toward a heritable syndrome. Demographic factors may point toward or away from certain diagnoses. For instance, AF rarely occurs in children and young adults, save rare familial forms, or AF associated with structural heart disease; a strong male predominance, as well as a higher prevalence in certain ethnic populations such as Southeast Asians, is seen in Brugada syndrome; inappropriate sinus tachycardia is nearly exclusively a condition affecting young women; degenerative conduction system disease leading to symptomatic bradycardia is most commonly a condition seen in older patients.

Arrhythmias may run the gamut from benign to malignant, life-threatening etiologies. Therefore, an important aspect of the evaluation of suspected arrhythmia is to discern patient prognosis, which then informs treatment. Arrhythmias that result in more significant hemodynamic compromise, and therefore more profound symptoms, tend to correlate with more malignant disease. In turn, the higher the suspicion for a malignant arrhythmia, the more aggressive the evaluation will likely be. Loss of consciousness, which may be the result of cardiac arrhythmia but also other etiologies that may be more benign, presents a particularly challenging yet common diagnostic dilemma. Therefore, careful thought into the appropriate evaluation for a patient with syncope is critical. In general, the presence of underlying structural abnormality of ventricular myocardium favors more malignant arrhythmias, both due to the increased risk of lethal ventricular arrhythmias and to potential inability to hemodynamically tolerate any particular arrhythmia. A careful history of circumstances, symptoms, and associated findings during the syncopal episode(s) can be very helpful in formulating a differential diagnosis.

The ECG is the cornerstone and most important diagnostic test that should be performed on every patient with suspected arrhythmia. A 12-lead resting ECG may offer clues to the diagnosis. Most simply, if active arrhythmia is captured on the ECG, a definitive diagnosis can be made. In addition, evidence suggesting underlying cardiac disease, such as prior myocardial infarction, left ventricular hypertrophy and possible hypertrophic cardiomyopathy, atrial disease, or baseline conduction system disease, may suggest a diagnosis. A subset of conditions that predispose to arrhythmia, both inherited or acquired, may be discerned as well, including ventricular preexcitation, prolonged or shortened QT interval, or ECG findings suggestive of specific inherited conditions such as Brugada syndrome or right ventricular cardiomyopathy.

However, the ECG typically records only 6 s of cardiac electrical activity, and therefore more intermittent and transient arrhythmias, particularly those not typically associated with abnormalities on the resting ECG, may not be seen. Many arrhythmias, including forms of both supraventricular tachycardia (SVT) and VT, can only be diagnosed definitively if an ECG is performed during active arrhythmia and/or symptoms from arrhythmia or, alternatively, provoked in the electrophysiology laboratory. Therefore, various forms of ambulatory monitoring may be performed to attempt to capture ECG activity during active arrhythmia. A growing variety of monitoring options are available; the most appropriate option should be primarily guided by the cadence of suspected arrhythmia episodes. For instance, if daily symptoms occur, a 24- or 48-h continuous Holter monitor is appropriate. On the other hand, a patient-activated event recorder is inappropriate in a patient with syncope, as the arrhythmic event will likely have passed once the patient reawakens.

Attempts at provoking arrhythmia may be warranted in the appropriate circumstances. An ECG-monitored treadmill test may elicit exercise-induced arrhythmias, or if long QT is suspected, a QT interval that fails to shorten appropriately with increased heart rate may be helpful. Pharmacologic provocation may be indicated for certain suspected diagnoses, such as Brugada syndrome. Judicious and appropriate use of carotid sinus massage or other means to enhance vagal tone may be helpful to diagnose carotid hypersensitivity or overall vagally mediated syncope.

Tilt table testing (TTT) involves having a patient strapped to a tiltable table. While monitoring heart rate and blood pressure, the patient is moved from a supine to upright position. In patients with suspected autonomic dysfunction–mediated syncope or presyncope, this provocation may elicit a paradoxical vagal response, resulting in bradycardia and/or sinus pauses as well as hypotension, and perhaps frank syncope. However, given the significant lack of both sensitivity and specificity, the current role of TTT is unclear, and it is seldom indicated after a careful history elucidates a neurally mediated cause for syncope.

Invasive electrophysiologic (EP) testing is the most useful diagnostic modality for many arrhythmias. Catheter-based recordings of intracardiac electrograms, with or without provocative pacing or pharmacologic maneuvers, may elicit the clinical arrhythmia. This will, in turn, help to define the mechanism of arrhythmia. However, one must keep in mind that for certain arrhythmia mechanisms, such as automaticity-driven tachycardia, EP study may fail to elicit arrhythmia due to the often transient and multifactorial nature of initiation of these arrhythmias. The nature of arrhythmia elicited will aid in determination of the patient's prognosis. In a typical EP study, catheters are placed within the heart via femoral venous access. Baseline conduction properties are measured. Provocative maneuvers including electrical pacing maneuvers, programmed stimulation, and pharmacologic provocation are performed. In the modern era, the vast majority of invasive EP studies are performed in conjunction with planned catheter ablation, although programmed ventricular stimulation for risk stratification of sudden death may still be utilized. EP study during catheter ablation is performed to confirm the diagnosis, localize appropriate ablation targets, and evaluate the efficacy of ablation performed during the procedure.

Depending on the suspected arrhythmia diagnosis, further testing may be indicated. If structural heart disease is suspected, echocardiography is most often the best next test, as it can assess for underlying structural disease, evaluate left ventricular function, and assess atrial dimensions and mitral valve function if AF is suspected, both of which are fair prognostic indicators. In patients in whom underlying coronary artery disease is suspected, an evaluation for coronary ischemia is indicated. Further evaluation for underlying structural heart disease will be directed based on the differential diagnosis. Cardiac computed tomography provides broad diagnostic utility, depending on the scanning protocol, including evaluation for ischemia, ventricular scar, anatomic evidence of coronary artery disease, congenital anomalies, and left atrial anatomy. Cardiac magnetic resonance imaging provides significant resolution of soft-tissue characteristics and may be used to assess for ischemia, infarct, cardiomyopathy, or infiltrative disease. Cardiac positron emission tomography can also discern underlying ischemia, as well as metabolic/inflammatory/infiltrative conditions.

TREATMENT

Cardiac Arrhythmias

ANTIARRHYTHMIC DRUG THERAPY

The effects of pharmacologic agents on cardiac electrophysiologic properties are often complex and, in some instances, remain incompletely understood. The complexity is the result of complex pharmacodynamics and pharmacokinetics, in particular significant cross-reactivity of certain drugs across different targets as well as variable effects on drug targets across drugs within the same category. There are regional differences in drug effect within the myocardium, and interpatient variations in drug metabolism play important roles. This has, in part, led to many instances of harm that have come from the adverse effects of many agents used over the years. In fact, many antiarrhythmic agents currently in use carry significant risks of side effects, some of which may be significant and even lethal. Therefore, judicious use of antiarrhythmic medications by those with appropriate knowledge base and experience is warranted. The practical result of the narrow therapeutic index of this class of medications has rendered their use increasingly as ancillary options **(Table 250-2)**.

The traditional nomenclature of antiarrhythmic drugs (AADs) is known as the Vaughan-Williams classification schema. In this schema, there are four classes (I–IV; Table 250-2). Class I AADs primarily target the Na channel, class II agents target the

TABLE 250-2 Antiarrhythmic Drug Actions

	CLASS ACTIONS				
DRUG	I	II	III	IV	OTHER ACTIONS/COMMON SIDE EFFECTS
Quinidine	++		++		Anticholinergic
Flecainide	+++		+		Can promote reentrant arrhythmias (atrial flutter, ventricular tachycardia)
Propafenone	++	+			Mild beta-blocker effect
Amiodarone	++	++	+++	+	Multiorgan toxicity with long-term use
Sotalol		++	+++		Prominent beta-blocker effect
Dofetilide			+++		Prolongation of QT at slower heart rates
Dronedarone	+	+	+	+	Mild effect
Ibutilide			+++		Used only for acute cardioversion
Ranolazine	++		++		Late sodium channel blockade
Lidocaine	++				Used for reperfusion arrhythmias

beta-adrenergic receptor, class III agents target potassium channels, and class IV agents target Ca channels. Class I agents are further subdivided into three subclasses based on the kinetics of drug to Na channel interactions. Class IA agents, including procainamide and quinidine, possess intermediate binding kinetics and potency. Class IB agents, including lidocaine and mexiletine, possess rapid binding kinetics and relatively low potency. Class IC agents (flecainide, propafenone) possess slow kinetics and high potency. Class II agents consist entirely of beta-adrenergic blocking agents. Class III agents (sotalol, dofetilide, ibutilide) specifically target the Kv11.1 potassium channel (encoded by the *KCNH2* gene) and risk prolongation of the QT interval through these effects on phases 2/3 of the AP and hence ventricular repolarization. Class IV agents are cardioselective Ca channel blockers including verapamil and diltiazem. This classification has significant limitations, however. Many AADs interact with multiple ion channels, and as a result, many exhibit behavior consistent with multiple classes. Amiodarone, in particular, exhibits properties of all AAD classes. Adenosine, with primary antiarrhythmic effects as an acute and transient, intravenously administered AV nodal blocking agent, as well as digitalis glycoside, which blocks the Na^+/K^+ pump, which in turn inhibits the Na^+/Ca^{++} exchanger, resulting in antiarrhythmic effect, do not neatly fit into this classification schema.

CATHETER ABLATION

The rationale that underlies catheter ablation for cardiac arrhythmia is that an anatomic substrate can be identified and localized, and disruption or isolation of that substrate will eliminate the cardiac arrhythmia. For automaticity-driven arrhythmias, a focal source of automaticity is identified, localized, and ablated. For anatomic reentrant arrhythmias, a critical zone of slow conduction that sustains arrhythmia and can be reasonably targeted is ablated. Moreover, the ablation target must be in a location deemed at acceptable risk of not damaging critical structures such as the native conduction system, coronary arteries, or extracardiac structures including the esophagus and phrenic nerve. Advances in electroanatomic mapping, a technology that uses alterations in electrical impedance and a magnetic field as measured by an intracardiac mapping catheter, have allowed for real-time reconstruction of cardiac chambers and identification of arrhythmogenic tissue to be targeted for ablation while safely avoiding nontargeted critical structures. Intracardiac echocardiography has also been used to enhance the safety and efficacy of invasive electrophysiologic procedures with real-time visualization of cardiac structures **(Fig. 250-4)**.

In the 1950s–1960s, as the underlying anatomic substrates for arrhythmias became better understood, open surgical disruption of arrhythmia circuits was the only available interventional and curative therapy for many arrhythmias. Surgical ligation of accessory pathways or resection of ischemic VT substrates was performed at specialized surgical centers. The first attempts at clinical catheter ablation utilized direct current (DC) electrical energy. This resulted in a high-voltage pulse of electrical energy that would ablate cardiac tissue, but with a difficult-to-control scope, often leading to significant complications. Radiofrequency (RF) energy was adapted to catheter-based cardiac ablation in the 1980s. RF alternating electrical current (300–550 kHz) delivered through a catheter tip results in local tissue heating and permanent injury, rendering the targeted tissue electrically inert. This type of ablation is similar to the technology used in electrosurgical techniques using a Bovie electrocautery device. For >35 years, RF delivery via catheters has been iteratively optimized such that it has become the most common and mainstay energy source for catheter ablation. Catheter ablation is indicated for a wide variety of clinical arrhythmias, including SVT, accessory pathways, atrial flutter, AF, PVCs, and VT. Alternative

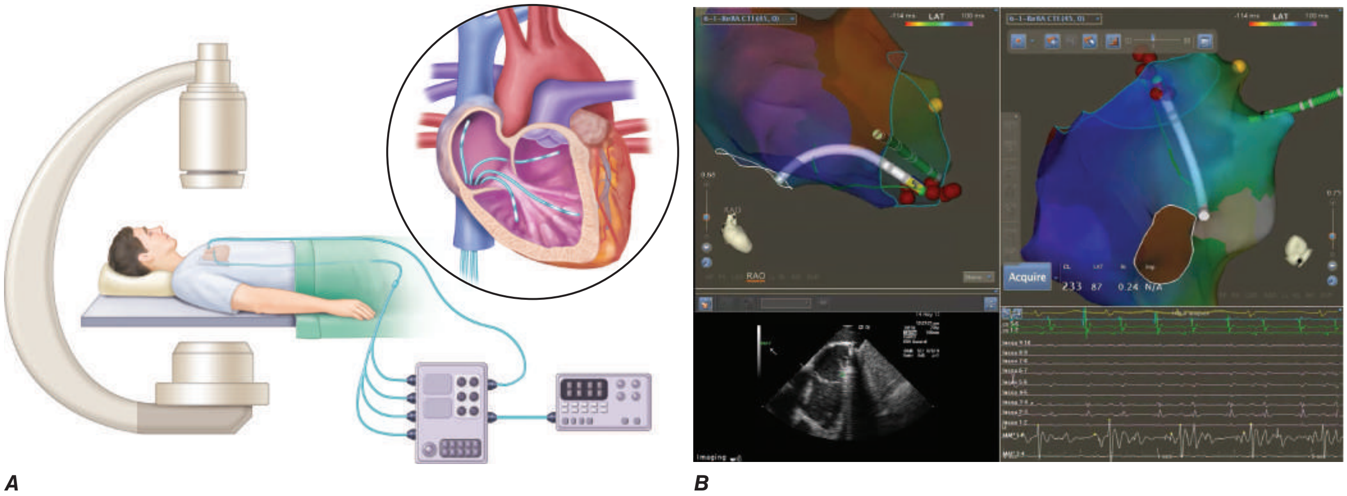

FIGURE 250-4 Catheter ablation of cardiac arrhythmias. ***A.*** A schematic of the catheter system and generator in a patient undergoing radiofrequency catheter ablation (RFCA); the circuit involves the catheter in the heart and a dispersive patch placed on the body surface (usually the back). The inset shows a diagram of the heart with a series of intracardiac catheters placed via the inferior vena cava (IVC), typically through femoral venous access. Catheters are located at the high right atrium, His bundle location, right ventricular (RV) apex, and through a transseptal puncture within the left atrium. ***B.*** Images from an electroanatomic mapping system are shown during mapping and ablation of typical cavo-tricuspid isthmus-dependent atrial flutter. This system allows three-dimensional real-time localization and annotation of catheter position and cardiac anatomy to guide mapping and ablation. In this instance, two projections of the map are shown at the top of the right atrium (RA), a right anterior oblique (RAO) and left anterior oblique (LAO) caudal view. Annotations of ablation lesion delivery are shown as *red dots*. In the left lower aspect of this panel, a simultaneous image from intracardiac echocardiography (ICE) is shown of the RA, with the ablation catheter in view in all three images. In the lower right aspect of this panel, surface electrocardiogram and intracardiac electrograms acquired in real time are shown.

ablative energy sources have been explored over the years, including cryothermy, light spectrum (laser), microwave, ultrasound, and more recently pulsed field electroporation, which injures targeted myocardium through high-energy, ultra-short pulses of electrical current that disrupts the lipid cell membrane, resulting in permanent cell death. Recently, the well-established ablative technique of stereotactic (focused and directed) external beam ionizing radiation has been applied to the heart to treat various arrhythmias, including VT and AF. This particular treatment modality holds promise given its ability to target regions of the heart that may be inaccessible to catheters, as well as the completely noninvasive nature of the procedure.

A widely applied non-RF ablative energy source today is cryothermy, where an ablative catheter tip is cooled to a temperature range (typically below −40°C) that results in permanent tissue death. Cryothermy is most widely applied to ablation of paroxysmal atrial ablation, via an expandable balloon introduced sequentially into each pulmonary vein and cooled to produce a circumferential ablative lesion at the ostium/antrum of each pulmonary vein. Similar catheter-based tools utilizing pulsed field ablation (PFA) have also been introduced for the purpose of electrically isolating pulmonary veins during AF ablation.

IMPLANTED ELECTRICAL DEVICE THERAPY

Implanted cardiac rhythm management devices are commonly utilized to manage arrhythmia. The first definitive pacemaker was implanted in 1958, and this technology has evolved to be the mainstay in the management of bradyarrhythmias. Sinus node dysfunction and AV conduction disease, particularly with symptoms, are the primary indications for most implanted pacemakers. Pacemakers are typically implanted percutaneously, with insulated wires, or leads, inserted through the upper extremity venous system into the right atrial and/or ventricular myocardium, with the lead tip secured to the myocardium mechanically. The leads are connected to a pulse generator placed in the prepectoral space, which contains electronic circuitry and a battery, allowing sensing and/or delivery of pacing stimuli to maintain adequate heart rate. More recently, a completely leadless pacemaker inserted through a large femoral venous sheath directly into the RA or right ventricle endocardium has become available. Although these devices possess more limited pacing options, they likely reduce the risks associated with transvenous lead systems, including infection or lead fracture requiring extraction.

Implanted cardioverter-defibrillators (ICDs) are placed in a similar fashion to pacemakers. However, ICDs have the ability to sense abnormal ventricular arrhythmias and deliver either antitachycardia pacing or defibrillation to prevent sudden death. In patients who experience a potentially lethal ventricular arrhythmia, ICD therapy may be lifesaving. Indications for ICD therapy are considered for either primary prevention of sudden cardiac death (SCD) due to arrhythmia in an at-risk patient or as secondary prevention in a patient who has survived an SCD event. More recently, a completely subcutaneous ICD system has become available, avoiding intravenous leads that increase risk for systemic infection, and potentially the procedure to extract a potentially fibrosed lead in cases of lead malfunction or endovascular infection.

FURTHER READING

Callans DJ: *Josephson's Clinical Cardiac Electrophysiology: Techniques and Interpretations*, 7th ed. Philadelphia, Wolters Kluwer, 2024.

Ellenbogen K et al (eds): *Clinical Cardiac Pacing, Defibrillation, and Resynchronization Therapy*, 5th ed. Philadelphia, Elsevier, 2016.

Jalife J, Stevenson W (eds): *Zipes and Jalife's Cardiac Electrophysiology: From Cell to Bedside*, 8th ed. Philadelphia, Elsevier, 2021.

251

The Bradyarrhythmias: Disorders of the Sinoatrial Node

William H. Sauer, Bruce A. Koplan

The sinoatrial (SA) node serves as the natural pacemaker of the heart and has variable rates in response to parasympathetic and sympathetic stimulation. If the sinus node is dysfunctional or suppressed, a subsidiary pacemaker in the atrioventricular node or specialized conduction system will take over, leading to a junctional or ventricular rhythm. Symptoms of sinus node dysfunction can vary but typically present as fatigue, exercise intolerance, or dyspnea. The diagnostic evaluation includes an investigation into reversible causes of sinus bradycardia, confirmation of sinus node dysfunction with outpatient telemetry monitoring or exercise testing, and possibly cardiac imaging if structural heart disease is suspected. Once irreversible sinus node dysfunction is confirmed, permanent pacemaker implantation is the only reliable long-term therapy for symptomatic bradycardia.

STRUCTURE AND PHYSIOLOGY OF THE SA NODE

The SA node region is complex in structure. Clusters of myocytes with pacemaker activity are surrounded by fibroblasts, endothelial cells, and transitional cells. These clusters of small fusiform cells in the sulcus terminalis on the epicardial surface of the heart at the right atrial–superior vena cava junction envelop the SA nodal artery. The SA node is structurally heterogeneous, but the central prototypic nodal cells have fewer distinct myofibrils than does the surrounding atrial myocardium, no intercalated disks visible on light microscopy, a poorly developed sarcoplasmic reticulum, and no T tubules. Cells in the peripheral regions of the SA node are transitional in both structure and function. The SA nodal artery arises from the right coronary artery in 55–60% and the left circumflex artery in 40–45% of persons. This feature, along with a protective extracellular matrix of connective tissue, insulates the SA node from the hyperpolarizing influence of the larger atrium. In addition, the alignment of this complex matrix is associated with nearly unidirectional electrical propagation to the atrium **(Fig. 251-1)**.

Pacemaker cells spontaneously depolarize in a continuous manner setting the natural rate of depolarization and myocardial contraction. Action potential depolarization in the SA node is normally at a resting rate of 60–100 beats/min. The autonomic nervous system exhibits control over the sinus node, with a preponderance of parasympathetic innervation at baseline. Removal of parasympathetic tone or an increase in sympathetic innervation leads to an increase in rate of depolarization. In denervated hearts, the rate of electrical depolarization (intrinsic heart rate) is approximately 100 beats/min, reflecting the rate of automaticity of the sinus node uninhibited by parasympathetic tone. The complement of ionic currents present in nodal cells results in a less negative resting membrane potential compared with atrial or ventricular myocytes. Electrical diastole in nodal cells is characterized by slow diastolic depolarization (phase 4), which generates an action potential as the membrane voltage reaches threshold. The action potential upstrokes (phase 0) are slow compared with atrial or ventricular myocytes, being mediated by calcium rather than sodium current.

Cells with properties of SA nodal tissue are electrically connected to the remainder of the myocardium by cells with an electrophysiologic phenotype between that of nodal cells and that of atrial or ventricular myocytes. Cells in the SA node exhibit the most rapid phase 4 depolarization and thus are the dominant pacemakers in a normal heart.

Myocytes within the SA node complex include specialized cells surrounded by fibrous tissue. Unlike atrial and ventricular cells, sinus node pacemaker cells have no true resting potential, but instead depolarize automatically and repetitively after the end of an action potential, and the depolarizing current in the SA node myocytes results primarily from slow calcium currents instead of fast sodium channels, which are

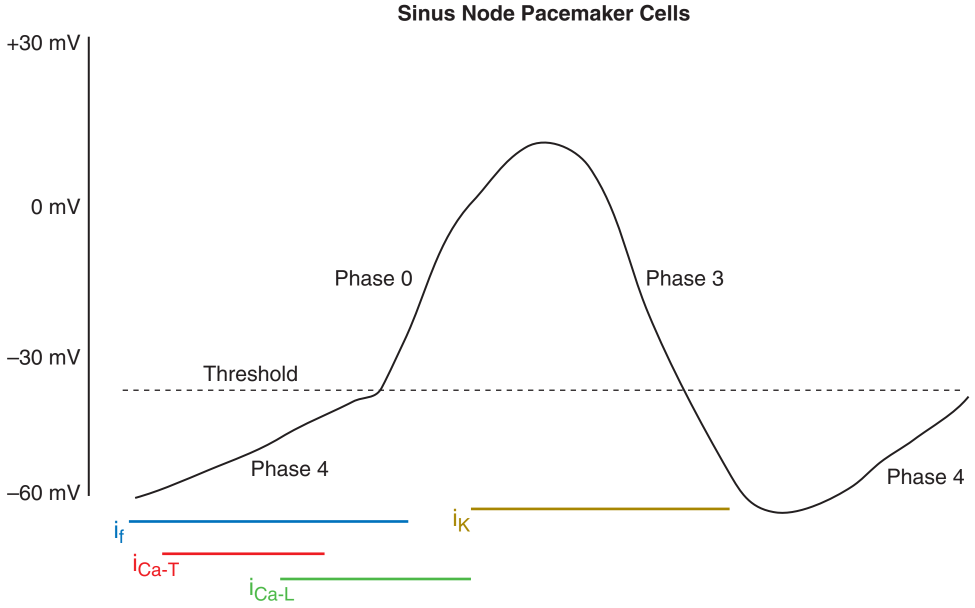

FIGURE 251-1 Cellular ion currents involved in depolarization and automaticity of sinoatrial (SA) nodal pacemaker cells. Phase 4 spontaneous depolarization results from i_f (funny) current, along with T- and L-type calcium channels. Phase 0 is the depolarization phase of the action potential. This is followed by phase 3 repolarization, which results from the outward directed hyperpolarizing K+ currents. i_f, funny current; i_{Ca-T}, T-type calcium current; i_{Ca-L}, L-type calcium current; i_K, potassium current.

absent in SA node cells. Spontaneous phase 4 depolarization results from a combination of slow inward depolarizing sodium current (I_f, "funny current"), along with inward calcium current controlled by T-type and L-type calcium channels. The upstroke of depolarization in SA node myocytes is slower and lower in amplitude than in ventricular myocytes.

In patients <85 years of age, the resting heart rate is strongly influenced by parasympathetic tone at baseline. The absence or elimination of autonomic influence on the SA node (e.g., after atropine administration) leads to an intrinsic heart rate that is normally 100–110 beats/min. The myocytes within the SA node that initiate pacing will change with different rates with a superior shift at higher heart rates and an inferior shift at lower rates. This shift may lead to a slightly different P wave inscribed on electrocardiograms (ECGs) recorded during different rates of sinus rhythm.

In addition, a progressive decline in maximum heart rate occurs with age, although the resting heart rate normally remains unchanged. Intrinsic heart rate declines 5–6 beats/min for each decade of age. However, the constancy of resting heart rate is associated with a gradual decrease in parasympathetic tone and a transition to predominant sympathetic tone by the ninth decade.

DIAGNOSIS OF SA NODAL DISEASE

Intrinsic sinus node disease is sometimes referred to as sick sinus syndrome or sinus node dysfunction (SND) and can manifest as fatigue, exercise intolerance, or syncope resulting from either reduced heart rate or pauses. Electrocardiographic recording plays a central role in the diagnosis and management of SA node dysfunction. The correlation between symptoms and slow heart rate or pauses is essential in determining whether bradycardia may be considered pathologic and necessitating intervention. Baseline ECG can detect baseline sinus bradycardia but may not indicate symptom correlation in certain settings. To address the limitations of the resting ECG, longer-term recording employing mobile telemetry devices such as Holter monitors or mobile cardiac telemetry can also be helpful in correlating symptoms with rate abnormalities **(Fig. 251-2)**.

In addition, commercially available wearable devices, such as watches with ECG recording capabilities, can have electrograms with excellent fidelity that may also be utilized. Contemporary event monitors may be automatically triggered to record the ECG when certain programmed heart rate criteria are met and implantable monitors permit very long-term recording (years) in particularly challenging patients. Treadmill testing can be utilized to assess for maximum heart rate. It is worth noting, however, that standard Bruce protocol treadmill testing may be helpful in detecting abnormalities in maximum heart rate, but more insidious chronotropic incompetence that manifests as abnormalities of rate increase during submaximal exercise may be more evident with treadmill protocols that have more gradual effort increases.

Once there is evidence of SND, it is important to rule out reversible causes of resting sinus bradycardia or chronotropic incompetence. **Table 251-1** lists the potentially reversible causes of sinus node disease and includes hypothyroidism and rate-slowing medications. Many patients with sleep apnea will have high vagal tone during sleep, especially during apneic events. Sinus bradycardia and sinus pauses frequently are seen if a patient is being monitored during this period. Sleep apnea, a common reversible cause, should be suspected if marked sinus bradycardia and prolonged sinus pauses are observed in a telemetry monitoring period during sleep. It is also worth noting that asymptomatic sinus bradycardia and pauses during sleep are typically not an indication for pacing, and it is, therefore, important when interpreting a wearable monitor to determine the timing of bradycardia events with regard to the sleep versus awake periods.

If structural heart disease is suspected, transthoracic echocardiography should be used to detect potential cardiac abnormalities associated with SND **(Fig. 251-3)**. Advanced cardiac imaging is indicated for evaluation of possible myocardial diseases such as amyloidosis, infiltrative cardiomyopathy, or myocarditis. Invasive electrophysiology testing solely to assess sinus node function is rarely utilized beyond the noninvasive techniques mentioned. In patients who are undergoing electrophysiology studies (EPS) for other indications, evaluation of sinus node function as part of the EPS may be considered. In symptomatic patients with suspected SND, EPS may rarely be considered when the diagnosis remains uncertain and after initial noninvasive evaluation is inconclusive. Investigation of the sinus node during EPS can consist of determination of sinus node recovery time (SNRT) and sinoatrial conduction time (SACT). In addition, the intrinsic heart rate [118.1 – (0.57 × age)] can be assessed via pharmacologic blockade of autonomic tone with intravenous propranolol and atropine. EPS is not widely used, however, as there is no evidence that abnormal SNRT or SACT alone can be used as an absolute indication for permanent pacing (PPM). There is no indication for EPS in asymptomatic patients with sinus bradycardia.

SA NODAL DYSFUNCTION SUBTYPES

SND can be categorized into problems with impulse formation and problems with impulse conduction. The term *sick sinus syndrome* may be used interchangeably with SND and refers to a group of related

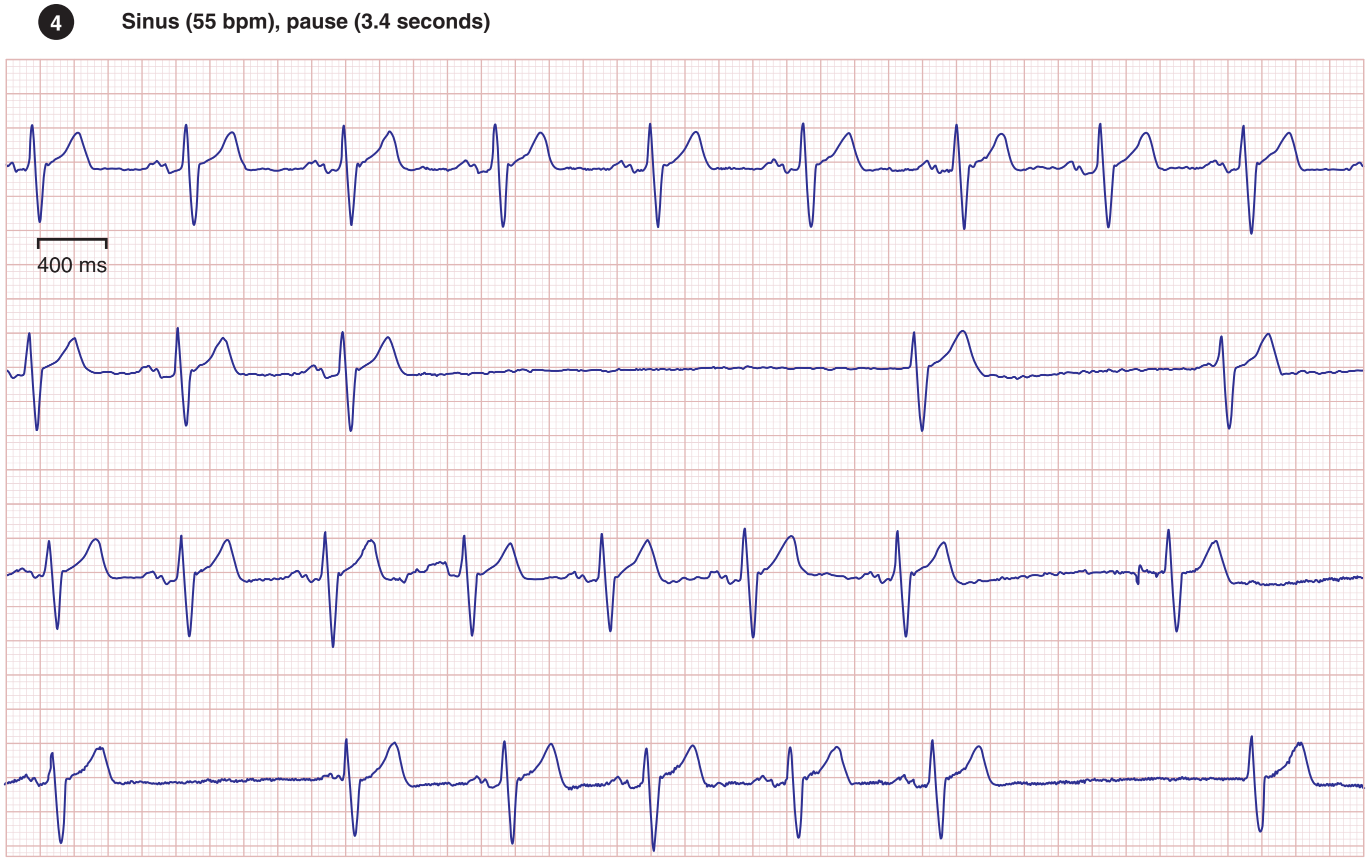

FIGURE 251-2 Sinoatrial exit block. A pause in the heart rhythm is seen that results from a sinus pause. On the second line of the tracing, there is a pause that results from the absence of a sinus beat (absent P wave) and no subsequent QRS. This is followed by a junctional escape beat and eventually recovery of the presence of sinus rhythm P waves.

conditions comprising problems of both impulse formation and impulse conduction.

Sinus Node Exit Block (See Fig. 251-4) "Sinus arrest" results from failure of impulse formation within the sinus node. Sinoatrial exit block results from failure of sinus node activity to propagate to the atrium. Sinoatrial exit block can have similar pattern characteristics of types of atrioventricular (AV) node block. It can manifest as complete SA block. Type I SA block involves fixed delay out of the sinus node. Type II SA block can occur with either progressive delay and then intermittent failure to propagate to the atrium (Mobitz I type) or fixed delay with intermittent failure to conduct (Mobitz II). The mass of the sinus node is not large enough to have an appearance on the ECG. Instead, the P waves that result from atrial depolarization can provide information that reflects the health of the sinus node. Type II second-degree SA block can be inferred on the ECG if the sinus rate abruptly transitions to a sinus rate that is half the previous rate (every other sinus depolarization is blocked from exiting to the atrium). Sinoatrial Wenckebach can be inferred from the ECG in the setting of progressive shortening of the P-P interval leading up to a sinus pause. This is due to progressive prolongation of SA conduction, but to a lesser extent with each successive prolongation. This is similar to the typical progressive shortening of the R-R interval that is observed with AV nodal Wenckebach. Other types of SA block require invasive EPS to decipher. The exercise of determining the type of SA block with invasive electrophysiology testing is typically not necessary because it does not alter management.

Tachy-Brady Syndrome Tachycardia-bradycardia (tachy-brady) syndrome is a subset of sick sinus syndrome/sinus node disease that consists of high heart rates (most commonly atrial fibrillation) with alternating symptomatic bradycardia or offset pauses (Fig. 251-5). Commonly, medications that are needed for rate control of tachycardia exacerbate bradycardia episodes, and thus the presence of tachy-brady syndrome is often a reason to consider pacemaker implantation.

Chronotropic Incompetence Chronotropic incompetence (CI) is broadly defined as the inability of the heart to increase its rate to meet activity or demand. Compared to an increased stroke volume, the increase in heart rate is a stronger contributor to the increase in oxygen uptake (VO_2) during aerobic exercise. Therefore, CI can be the primary cause of severe exercise intolerance and increased cardiovascular events and overall morality. Unfortunately, CI lacks a consistent definition, leading to a lack of clarity on its overall prevalence. CI can take many forms, including failure to achieve a percentage (e.g., 85%) of age-predicted maximal heart rate [(220 – age) × 0.85] during stress testing. Other definitions that have been used include an overall a maximum heart rate [208 – (0.7 × age)], heart rate instability with exercise, or failure to achieve submaximal heart rate. Due to this latter category, standard exercise testing can, at times, fail to recognize a patient with CI because some patients can achieve an appropriate maximum heart rate but may exhibit heart rate instability or inadequate heart rate during activities of daily living (ADLs). Ambulatory heart rate monitoring along with a diary can be helpful to correlate symptoms with abnormally slow heart rates. Because CI can be insidious and multiple definitions exist, it can be easily overlooked.

Sinus Node Fibrosis Clinical SND is most common in older adults. This is due to normally occurring age-associated increase in fibrotic tissue in the SA node, which can exacerbate any degree of SND. A loss of pacemaker cells in the sinus node is also seen with age. It is worth noting, however, that while increased fibrosis in the SA node and decreased numbers of pacemaker myocytes are part of a normal process of aging, SND is pathologic and there are many elderly patients with extensive fibrosis and normal heart rate.

SA Nodal Ischemia and Infarction Sinus bradycardia is common in patients with acute inferior or posterior myocardial infarction (MI) and can be exacerbated by increased vagal tone (Bezold-Jarisch reflex) or with the use of drugs such as morphine and beta blockers. Ischemia of the SA nodal artery occurs in acute coronary syndromes

TABLE 251-1 Reversible Causes of Sinus Node Dysfunction

Medical Conditions Associated with Sinus Bradycardia
• Hypothyroidism
• Sleep apnea
• Hypoxia
• Hypothermia
• Increased intracranial pressure
• Lyme disease
• Myocarditis
• COVID-19
• Vagal reflex (cough, pain, etc.)
Medications Associated with Sinus Node Dysfunction
Antihypertensive Medications
• Beta-adrenergic receptor blockers
• Clonidine
• Methyldopa
• Nondihydropyridine calcium channel blockers
Antiarrhythmic Medications
• Amiodarone
• Dronedarone
• Flecainide
• Procainamide
• Propafenone
• Quinidine
• Sotalol
• Ivabradine
Psychiatric Medications
• Donepezil
• Lithium
• Opioid analgesics
• Phenothiazine antiemetics and antipsychotics
• Phenytoin
• Selective serotonin reuptake inhibitors
• Tricyclic antidepressants
Other
• Anesthetic drugs (propofol)
• Cannabis
• Digoxin
• Muscle relaxants

more typically with involvement with the right coronary artery, and even with infarction, the effect on SA node function most often is transient. However, there are rare cases where SA infarction can affect sinus node function. One potential rare complication of atrial fibrillation catheter ablation is the inadvertent injury to the SA nodal artery that may be coursing over a targeted ablation region in the right and left atrium. SND and arrest have been described following ablation of atrial fibrillation and flutter.

Carotid Sinus Hypersensitivity and Neurally Mediated Bradycardia Sinus bradycardia is a prominent feature of carotid sinus hypersensitivity and neurally mediated bradycardia associated with the cardioinhibitory variant of vasovagal syncope. Carotid hypersensitivity with recurrent syncope or presyncope associated with a predominant cardioinhibitory component responds to pacemaker implantation. Although the vasodepressor effect of the enhanced vagal tone may be unaffected by the pacing support, the lack of bradycardia often prevents injury with this subtype of vasovagal syncope. Several randomized trials have investigated the efficacy of permanent pacing in patients with drug-refractory vasovagal syncope, with mixed results. Although initial trials suggested that patients undergoing pacemaker implantation have fewer recurrences and a longer time to recurrence of symptoms, at least one follow-up study did not confirm these results.

TREATMENT

SA Nodal Disease

TEMPORARY PACING FOR TRANSIENT SUPPORT

In symptomatic patients presenting with sinus node disease, removing any possible reversible cause remains the initial strategy. Acute myocardial infarction, electrolyte abnormalities, medications, and hypothyroidism should all be considered as potentially reversible causes. Unnecessary medications that may be causing bradycardia should be eliminated. Beta blockers, calcium channel blockers, and digoxin are some of the more common medications in use that may cause bradycardia. These drugs may have a wide range of indications in patients after MI and with chronic systolic dysfunction. If stopping the medication or decreasing the dose is an option, this should be tried first. If the medication is felt to be unavoidable, a pacemaker may be indicated.

In patients with tachy-brady syndrome, alleviation of the tachycardia, whether it is atrial fibrillation or other forms of supraventricular tachyarrhythmias, can prevent bradycardia events. Treatment of the tachycardia can sometimes be accomplished with antiarrhythmic drug therapy or catheter ablation. If arrhythmia control cannot be achieved, permanent pacing may be necessary.

Hypoxia from decrease in blood flow to the SA node, which can occur with cardiac ischemia or MI, can lead to slowing of phase 4 depolarization and resultant bradycardia. Further ischemia and necrosis of pacemaker cells can cause irreversible sinus node disease. On occasion, reversal of ischemia with revascularization can alleviate bradycardia. Sinus pauses in the setting of tachy-brady syndrome may be eliminated if atrial tachyarrhythmias can be successfully treated. It is also important to recognize when bradycardia may be transient. Acute illness associated with episodes of extreme vagal tone may lead to transient SA node abnormalities. Typically, this may be observed as sinus slowing, followed by transient sinus arrest and/or AV block. Although a pacemaker may be needed in extreme instances of prolonged arrest, recovery from the acute illness may make the pacemaker unnecessary in follow-up.

Sinus bradycardia may also be observed after heart transplantation and cardiac surgery. Due to cardiac denervation, a normal resting heart rate in heart transplant recipients is generally 90–110 beats/min. Therefore, a heart rate that may be normal in a nontransplant patient may represent CI in a transplanted patient. In the case of heart transplantation, sinus bradycardia may be due to accumulated drugs such as amiodarone that affect the donor heart or ischemic injury to the SA node upon transplantation. If the SA nodal artery is injured at the time of right atriotomy during cardiac surgery, sinus arrest with junctional rhythm may be observed. Temporary pacing or pharmacologic support with beta-1 adrenergic agonists may be needed in these circumstances while awaiting SA nodal recovery.

In addition, sinus bradycardia and sinus pauses are common after spinal cord injury. The mechanism of bradycardia is enhanced parasympathetic tone and autonomic dysreflexia. Common triggers can be tracheal suctioning and turning the patient. Atropine and inotropes have shown mixed success. Adenosine blockade with theophylline or aminophylline can sometimes be successful. Temporary and sometimes permanent pacing may be necessary in extreme circumstances.

PERMANENT PACEMAKER IMPLANTATION

Pacing in SA nodal disease is indicated to alleviate symptoms of bradycardia. Consensus guidelines published by the American Heart Association (AHA)/American College of Cardiology (ACC)/Heart Rhythm Society (HRS) outline the indications for the use of pacemakers and categorize them by class based on levels of evidence **(Fig. 251-6)**. Since the first implementation of permanent pacing in the 1950s, many advances in technology have resulted in miniaturization, increased longevity of pulse generators, improvement in leads, and increased functionality. To better understand

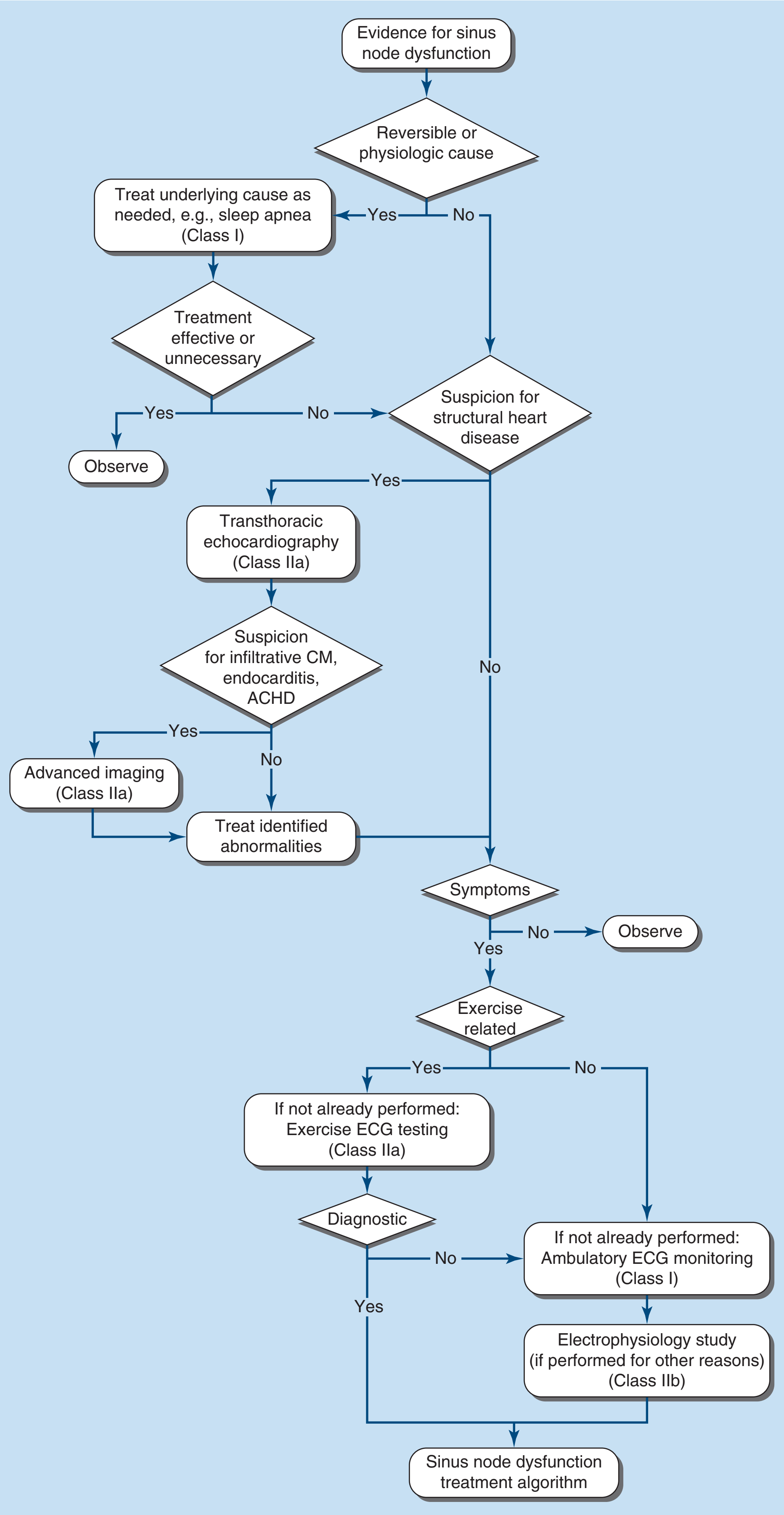
Evidence for sinus node dysfunction
Reversible or physiologic cause
Yes
No
Treat underlying cause as needed, e.g., sleep apnea (Class I)
Treatment effective or unnecessary
Yes
No
Observe
Suspicion for structural heart disease
Yes
No
Transthoracic echocardiography (Class IIa)
Suspicion for infiltrative CM, endocarditis, ACHD
Yes
No
Advanced imaging (Class IIa)
Treat identified abnormalities
Symptoms
No
Observe
Yes
Exercise related
Yes
No
If not already performed: Exercise ECG testing (Class IIa)
Diagnostic
No
Yes
If not already performed: Ambulatory ECG monitoring (Class I)
Electrophysiology study (if performed for other reasons) (Class IIb)
Sinus node dysfunction treatment algorithm

FIGURE 251-3 Evaluation of bradycardia and conduction disease. In patients with sinus node dysfunction, reversible causes should be identified and eliminated when possible. If no reversible cause can be identified, structural heart disease should be considered and evaluated for, if appropriate. If no symptoms are present, an observation strategy is appropriate. In patients who are symptomatic, further evaluation with ambulatory monitoring or exercise testing to identify symptom-rhythm correlation should be considered. ACHD, adult congenital heart disease; CM, cardiomyopathy. *(Reproduced with permission from FM Kusumoto et al: 2018 ACC/AHA/HRS guideline on the evaluation and management of patients with bradycardia and cardiac conduction delay. Heart Rhythm 16:e128, 2019.)*

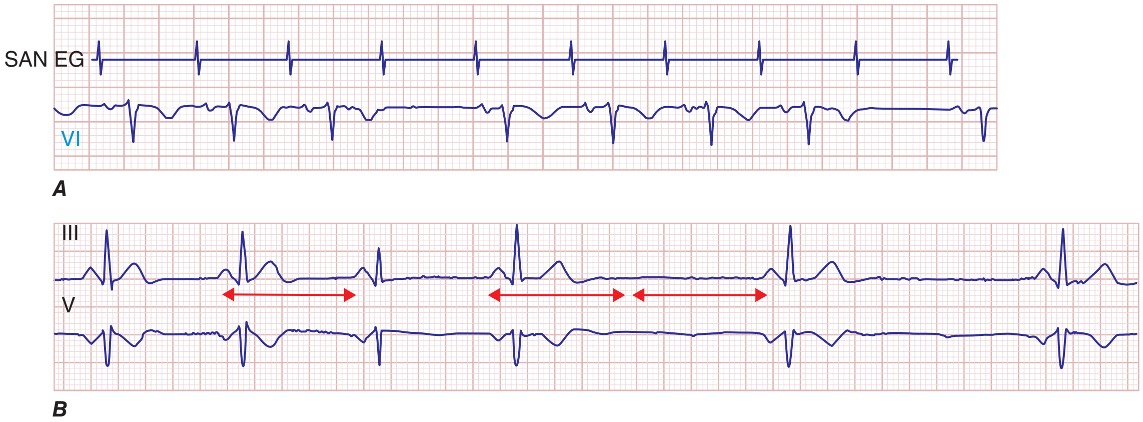

FIGURE 251-4 ***A.*** Mobitz type I sinoatrial (SA) nodal exit block. A theoretical SA node electrogram (SAN EG) is shown. Note that there is grouped beating producing a regularly irregular heart rhythm. The SAN EG rate is constant with progressive delay in exit from the node and activation of the atria, inscribing the P wave. This produces subtly decreasing P-P intervals before the pause, and the pause is less than twice the cycle length of the last sinus interval. ***B.*** Mobitz type II SA nodal exit block. This panel shows sinus rhythm in the first four beats followed by a sinus pause with the absence of a P wave. The interval comprising the absent P wave is exactly twice as long as the preceding P-P interval consistent with type II SA exit block.

pacemaker therapy for bradycardias, it is important to be familiar with the fundamentals of pacemaker function.

There is no established heart rate below which pacemaker treatment is indicated **(Table 251-2)**. Well-conditioned athletes can have resting sinus rates below 40 beats/min, and some individuals can have similar levels of bradycardia during sleep. Permanent pacing is typically not indicated for sleep-related pauses felt secondary to high vagal tone in the absence of other symptoms. Asymptomatic sinus bradycardia has not been associated with adverse outcomes and does not typically warrant permanent pacing. In situations such as asymptomatic sinus bradycardia, sinus pauses secondary to physiologically elevated parasympathetic tone, transient pauses during sleep, or asymptomatic SND where symptoms have been documented to occur in the absence of bradycardia, a pacemaker is generally not indicated.

Medications to improve heart rate in order to avoid PPM are very rarely utilized. Medications such as methylxanthines (e.g., theophylline) or beta agonists (e.g., terbutaline) are sometimes utilized on a temporary basis when a pacemaker may need to be delayed due to unique circumstances such as active infection. In addition, oral theophylline may be considered to determine if an increase in heart rate is associated with improvement in symptoms in a patient with sinus bradycardia to suggest that a PPM may be beneficial. This latter strategy is rarely utilized in more equivocal situations.

PPM is the principal treatment for SND, and the decision to pursue this treatment is largely driven by a correlation between symptoms and bradycardia. The stronger the correlation between symptoms and bradycardia, the greater is the likelihood of improvement. PPM is most commonly achieved through transvenous implantation of one or more leads through the left or right subclavian veins into the cardiac chambers. The leads are attached to a pacemaker generator that is placed subcutaneously in the pectoral chest region. Less commonly, pacing leads can be placed in the epicardium via surgical approaches including sternotomy or thoracotomy. This latter approach can be accomplished as a standalone procedure but is more commonly performed concomitantly

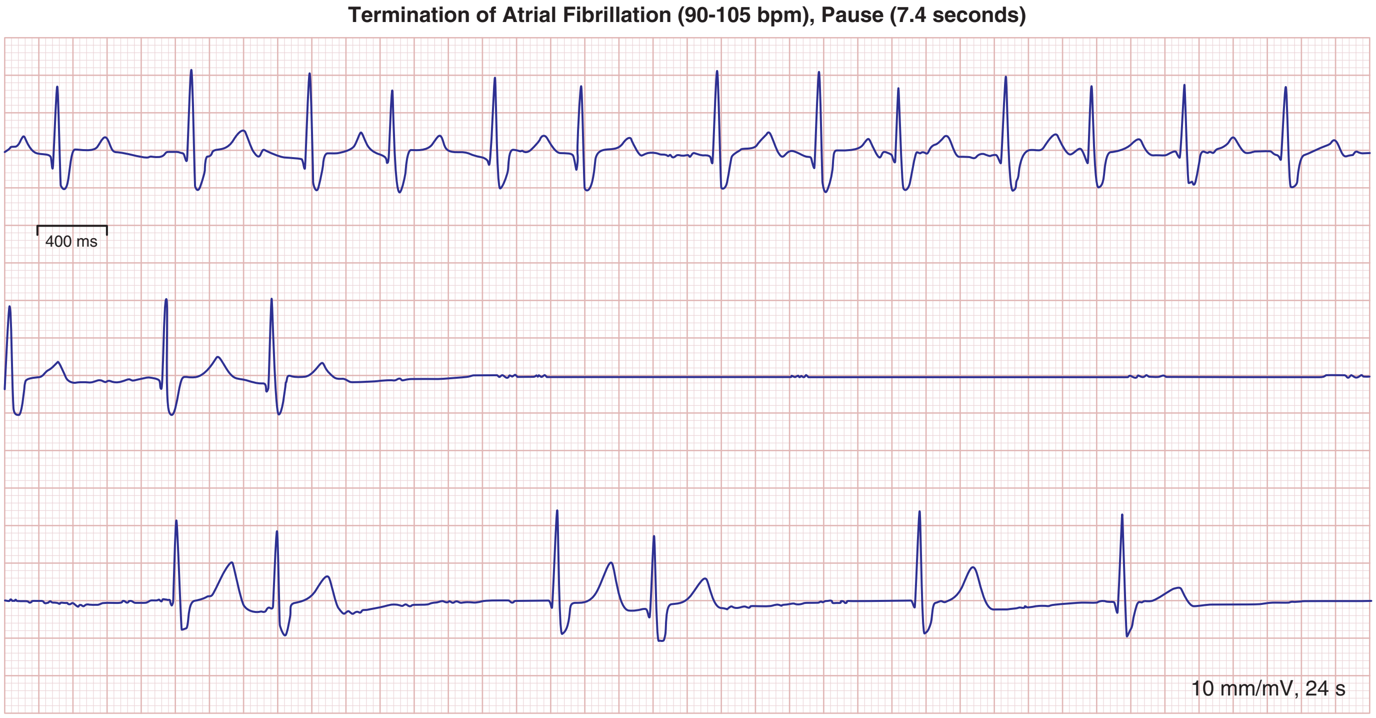
Termination of Atrial Fibrillation (90-105 bpm), Pause (7.4 seconds)

FIGURE 251-5 Offset pause and tachy-brady syndrome. An offset pause after termination of atrial fibrillation is seen and is consistent with tachy-brady syndrome.

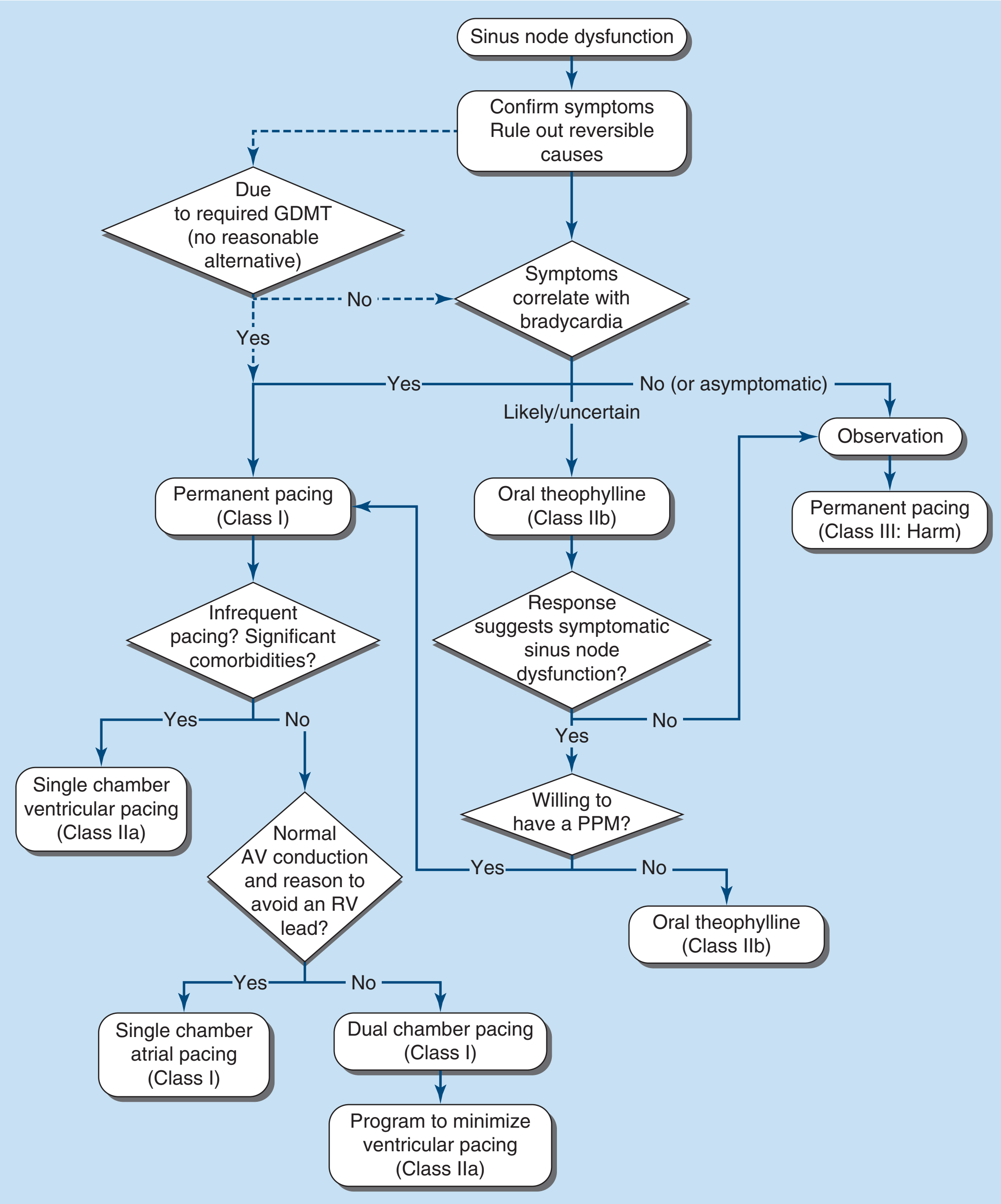
Sinus node dysfunction
Confirm symptoms
Rule out reversible causes
Due to required GDMT (no reasonable alternative)
No
Yes
Symptoms correlate with bradycardia
Yes
No (or asymptomatic)
Likely/uncertain
Observation
Permanent pacing (Class I)
Oral theophylline (Class IIb)
Permanent pacing (Class III: Harm)
Infrequent pacing? Significant comorbidities?
Response suggests symptomatic sinus node dysfunction?
Yes
No
No
Yes
Single chamber ventricular pacing (Class IIa)
Normal AV conduction and reason to avoid an RV lead?
Willing to have a PPM?
Yes
No
Oral theophylline (Class IIb)
Yes
No
Single chamber atrial pacing (Class I)
Dual chamber pacing (Class I)
Program to minimize ventricular pacing (Class IIa)

FIGURE 251-6 Management of sinus node dysfunction. Management of sinus node dysfunction begins with eliminating reversible causes and confirming whether symptoms correlate with bradycardia. If symptoms are clearly correlated, permanent pacing should be offered. If it is unclear, a trial of oral theophylline can be considered diagnostically. If there is no correlation between symptoms and bradycardia, then observation is appropriate. Class I recommendations should be performed or are indicated. Class IIa recommendations are considered reasonable to perform. Class IIb recommendations may be considered. Class III recommendations are associated with harm more than benefit. AV, atrioventricular; GDMT, guideline-directed management and therapy; PPM, permanent pacemaker; RV, right ventricular. *(Reproduced with permission from FM Kusumoto et al: 2018 ACC/AHA/HRS guideline on the evaluation and management of patients with bradycardia and cardiac conduction delay. Heart Rhythm 16:e128, 2019.)*

TABLE 251-2 Indications for Permanent Pacing in Sinus Node Dysfunction (SND)

- Symptoms that are directly attributable to SND
- Symptomatic sinus bradycardia because of essential medication therapy for which there is no alternative treatment
- Tachy-brady syndrome and symptoms attributable to bradycardia
- Symptomatic chronotropic incompetence
- In patients with symptoms that are possibly attributable to SND, a trial of oral theophylline may be considered to increase heart rate and determine if permanent pacing may be beneficial

Source: FM Kusumoto et al: Heart Rhythm 16:e128, 2019.

during another primary cardiac surgery. Leadless pacemakers that are totally self-contained pacing devices can also be placed in the right atrium and right ventricle to provide dual chamber pacing. Some leadless pacemakers can also incorporate technology to sense atrial activity to attempt to coordinate atrial sensing with ventricular pacing.

A standard nomenclature for pacing mode programming utilizes a four-letter code. The first letter indicates the chamber(s) paced (O, none; A, atrium; V, ventricular; D, dual; S, single). The second letter indicates the chamber(s) sensed. The third letter is the response to a sensed event (O, none; I, inhibited; T, triggered;

D, inhibition and triggered). The fourth letter refers to whether rate response (R) is turned on. Therefore, a dual-chamber pacemaker programmed in a DDDR mode provides atrial and ventricular pacing, A and V sensing, can be inhibited or triggered by a sensed beat, and is programmed to provide rate responsiveness to activity via either a built-in accelerometer, minute ventilation sensor, or both. Rate response is essential for the treatment of CI as it attempts to mimic the natural physiologic increase in heart rate in response to exertion.

A single-chamber atrial pacemaker can be a consideration in patients with pure SND who are felt to be at low risk for developing AV nodal block. However, fibrosis of the sinus node is associated with fibrosis of the AV node on pathology series in older patients and many patients with SND will develop AV node disease. Therefore, although a single-chamber atrial pacemaker can be a consideration in younger patients with pure SND, a majority of patients receiving a pacemaker for sinus node disease (particularly older individuals) often receive a dual-chamber pacemaker with backup ventricular pacing if needed.

Class I indications for pacing in SA node dysfunction include documented symptomatic bradycardia, SND-associated long-term drug therapy for which there is no alternative, and symptomatic CI. Class IIa indications include those outlined previously in which SND is suspected but not documented and for syncope of unexplained origin in the presence of major abnormalities of SA node dysfunction. Mildly symptomatic individuals with heart rates consistently <40 beats/min constitute a class IIb indication for pacing. Pacing is not indicated in patients with SA node dysfunction who do not have symptoms and in those in whom bradycardia is associated with the use of nonessential drugs.

COMPLICATIONS RELATED TO PACEMAKER IMPLANTATION

Although pacemakers are highly reliable, they are subject to a number of complications related to implantation and electronic function. In adults, permanent pacemakers are most commonly implanted with access to the heart by way of the subclavian superior vena cava venous system. Rare but possible acute complications of transvenous pacemaker implantation include infection, hematoma, pneumothorax, cardiac perforation, diaphragmatic/phrenic nerve stimulation, and lead dislodgment. Limitations of chronic pacemaker therapy include infection, erosion, lead failure, and abnormalities resulting from inappropriate programming or interaction with the patient's native electrical cardiac function. Rotation of the pacemaker pulse generator in its subcutaneous pocket, either intentionally or inadvertently, often referred to as "twiddler's syndrome," can wrap the leads around the generator and produce dislodgment with failure to sense or pace the heart. The small size and light weight of contemporary pacemakers make this a rare complication. Transvenous leads are considered the least reliable component of permanent pacing systems. Enhancements in battery technology and component design have produced a pacing system small enough to be implanted in the heart without the need for a transvenous lead. The first "leadless" pacemakers were only appropriate for patients with indications for single-chamber ventricular (right ventricle) pacing. More recently, these devices have been modified to detect mechanical atrial contraction and thus can be programmed to preserve AV synchrony in patients with heart block but without SND. Dual-chamber leadless pacemakers can also be implanted to preserve AV synchrony in patients with AV block.

FURTHER READING

Callans DJ: *Josephson's Clinical Cardiac Electrophysiology: Techniques and Interpretations*, 7th ed. Philadelphia, Wolters Kluwer, 2024.

Jalife J, Stevenson W (eds): *Zipes and Jalife's Cardiac Electrophysiology: From Cell to Bedside*, 8th ed. Philadelphia, Elsevier, 2021.

Sidhu S, Marine JE: Evaluating and managing bradycardia. Trends Cardiovasc 30:265, 2020.

252

The Bradyarrhythmias: Disorders of the Atrioventricular Node

William H. Sauer, Bruce A. Koplan

Impulses generated in the sinoatrial (SA) node are conducted to the ventricles through the electrically and anatomically complex atrioventricular (AV) node. The AV node is specialized for slow conduction of the action potential to create a delay between atrial and ventricular activation. There is a similar pattern of arrangement of gap junctions and expression of connexins in the AV node as in the sinus node. This allows for weak electrical coupling in the center of the node to slow conduction and protect it from the periphery. This unique arrangement of gap junctions, along with extracellular matrix and fibroblasts and a lack of conductance in adjacent valvular tissue, allow the AV node to slow conduction and serve as the electrical "gatekeeper" to the ventricle.

As described in previous chapters, cells located in the AV node sit at a relatively higher resting membrane potential than surrounding atrial and ventricular myocytes, exhibit spontaneous depolarization during phase 4 of the action potential, and have slower phase 0 depolarization (mediated by calcium influx in nodal tissue with a lack of the type of fast sodium channels found in atrial and ventricular myocytes) compared to ventricular tissue (mediated by sodium influx). Although the AV node has the potential for pacemaker activity, the normal automaticity rate is 20–60 beats/min, which is overridden by the higher intrinsic rate of the SA node (60–100 beats/min). Therefore, the AV node can provide a backup heart rate when the SA node fails to depolarize. There is a progressive decrease in the frequency of spontaneous depolarization down the His and Purkinje fibers. This progressive decrease in rate of depolarization allows for unidirectional flow of impulses through the conduction system.

Bradycardia may occur when conduction across the AV node is compromised, resulting in slow ventricular rates. The consequences can be fatigue, syncope, and, if a reliable escape rhythm does not occur, death from asystole. Transient AV conduction block is common in the young and is most likely the result of high vagal tone found in up to 10% of young adults. Acquired and persistent failure of AV conduction is rare in healthy adult populations, with an estimated incidence of 1 per 5000 in the U.S. population per year. In the setting of myocardial ischemia, aging and fibrosis, or cardiac infiltrative diseases, however, persistent AV block is much more common. As with symptomatic bradycardia arising from SA node dysfunction, permanent pacing is the only reliable therapy for symptoms arising from AV conduction block. While pacemakers for sinus node dysfunction are mainly indicated when symptoms correlate with bradycardia, pacemakers may be indicated in some forms of AV block in the absence of symptoms if the type of AV block is lower in the conduction system (infra-Hisian block) due to an increased risk of life-threatening asystole even if symptoms are not present leading up to such an event.

STRUCTURE AND PHYSIOLOGY OF THE AV NODE

The normal AV junctional area can be divided into a transition cell zone (which results from approaches from the atrium to the AV node), the compact AV node, and the penetrating part of the His bundle. Conduction from the SA node to the AV node occurs in a preferential manner via intra-atrial pathways with higher conduction velocity than the remainder of atrial tissue. The AV node itself is a small region (~1 × 3 × 5 mm) that lies beneath the right atrial endocardium at the apex of the triangle of Koch, a region defined by three landmarks: the coronary sinus ostium posteriorly, the septal tricuspid valve annulus anteriorly, and the tendon of Todaro superiorly. The AV node can be further subdivided into the lower nodal bundle and compact node. A rightward inferior node extension spreads along the tricuspid valve toward the coronary sinus, and the leftward nodal extension spreads from the compact node along the tendon of Todaro. In some people, there are two functional pathways in the AV node: a slow pathway located in the inferior node extension and a fast pathway that is less well defined but is superior to the slow pathway. The role of these pathways in supraventricular tachycardia is described in another chapter.

The compact AV node continues anteriorly and superiorly as the penetrating AV bundle where it traverses the central fibrous body in close proximity to the aortic, mitral, and tricuspid valve annuli. The penetrating AV bundle continues through the annulus fibrosis and emerges as the His bundle along the ventricular septum. The right bundle branch (RBB) emerges from the distal AV bundle and terminates to a band that traverses the right ventricle (moderator band). In contrast, the left bundle branch (LBB) is a broad subendocardial sheet of tissue on the septal left ventricle. The Purkinje fiber network emerges from the RBB and LBB and extensively ramifies on the endocardial surfaces of the right and left ventricles, respectively.

The cells that constitute the AV node complex are heterogeneous with a range of action potential profiles. The AV junction has distinct regions including a transitional cell zone (atrionodal cells), the compact AV node, and the cells of the penetrating part of the His bundle. In the transitional zones, the cells have an electrical phenotype between those of atrial myocytes and cells of the compact node. Atrionodal transitional connections exhibit *decremental conduction*, defined as slowing of conduction with increasingly rapid rates of stimulation. Myocytes that constitute the compact AV node are similar to sinus node myocytes, having a resting membrane potential of ~–60 mV; exhibit action potentials with low amplitudes, slow upstrokes of phase 0 (<10 V/s), and spontaneous phase 4 diastolic depolarization; and have high-input resistance and relative insensitivity to external [K$^+$]. The action potential phenotype is explained by the complement of ionic currents expressed. AV nodal cells lack a robust inward rectifier potassium current (I_{K1}) and fast sodium current (I_{Na}); L-type calcium current ($I_{Ca\text{-}L}$) is responsible for phase 0; and phase 4 depolarization reflects the composite activity of the depolarizing currents—funny current (I_f), $I_{Ca\text{-}L}$, T-type calcium current ($I_{Ca\text{-}T}$), and sodium-calcium exchanger current (I_{NCX})—and the repolarizing currents—delayed rectifier (I_{Kr}) and acetylcholine-gated (I_{KACh}) potassium currents. Electrical coupling between cells in the AV node is tenuous due to the relatively sparse expression of gap junction channels (predominantly connexin-40) and increased extracellular volume.

The His bundle and the bundle branches are insulated from ventricular myocardium. The most rapid conduction in the heart is observed in Purkinje cells (1–3 m/s), with action potentials exhibiting a very rapid upstroke (phase 0), prolonged plateau (phase 2), and modest automaticity (phase 4 depolarization). Gap junctions, composed largely of connexin-40, are abundant, but bundles are not connected transversely to ventricular myocardium. The AV node is innervated by both sympathetic and parasympathetic autonomic input that can either slow or enhance conduction.

The blood supply to the penetrating AV bundle is from the AV nodal artery and first septal perforator of the left anterior descending coronary artery. The AV node artery arises from the right coronary artery (80–90% of the time) or the left circumflex (10%) with the assigned artery associated with the dominance of the coronary artery circulation. The bundle branches also have a dual blood supply from the septal perforators of the left anterior descending coronary artery (LBB and proximal RBB) and branches of the posterior descending coronary artery. The AV node is highly innervated with postganglionic sympathetic and parasympathetic nerves; however, the bundle of His and distal conducting system are not influenced by autonomic tone.

ELECTROCARDIOGRAPHIC DEFINITIONS OF AV CONDUCTION BLOCK (TABLE 252-1)

Conduction block in the AV node is classified based on the appearance on electrocardiography (ECG), which may also be a reflection of the location of block along the AV conduction axis. First-degree AV block involves a fixed prolongation of the PR interval (>200 ms). In first-degree AV block, delay usually occurs within the AV node, although the atria, His bundle, and Purkinje system may also be involved. Although the term *block* is a misnomer of sorts because electrical conduction is delayed and not interrupted, it remains commonly in use. Second-degree AV block involves intermittent failure of conduction between the atrium and ventricle. There are two subtypes of second-degree AV block. Type I (Mobitz I, Wenckebach) AV block manifests as progressive prolongation of the PR interval prior to one or more "dropped" QRS complexes. Progressive shortening of the RR interval and grouped beating of QRS complexes are classically seen in Mobitz I AV block. In addition, comparison of the last PR interval before and the first PR interval after the absent QRS complex will often reveal the greatest discrepancy of PR interval, making the diagnosis of Mobitz I AV block most evident. Mobitz I AV block typically involves the AV node, is hemodynamically stable, and in the absence of symptoms, does not typically warrant pacing. Type II second-degree AV block manifests on ECG as failed AV conduction preceded by fixed PR interval (no prolongation of PR interval prior to a dropped beat). Type II block has more serious implications, including a risk of sudden death. It is infranodal in location and associated with a less reliable escape rhythm, and thus, permanent pacing is required even in the absence of symptoms. In the setting of 2:1 AV block, ECG differentiation of type I versus type II block is not possible. If the PR interval is <160 ms prior to the AV conduction and QRS is wider than normal, infranodal (type II) block is most likely. Complete heart block (third-degree block) involves complete AV dissociation with a ventricular rate that is slower than the atrial rate **(Fig. 252-1)**.

TABLE 252-1 Electrocardiographic Classification of Atrioventricular (AV) Block

First-Degree AV Block
All atrial impulses are conducted to the ventricle
PR interval is abnormally long (>200 ms)
AV delay usually occurs within the AV node
Second-Degree AV Block (intermittent failure of conduction between atrium and ventricle)
Two subtypes
Type I/Mobitz I/Wenckebach block: progressive prolongation of the PR interval until loss of conduction occurs
Type II/Mobitz II: fixed PR interval precedes loss of conduction
Usually associated with QRS widening
Third-Degree AV Block (complete heart block)
Complete interruption of conduction between atria and ventricles

In the absence of a preexisting bundle branch block, a wide QRS escape rhythm implies a block in the distal His or bundle branches; in contrast, a narrow QRS rhythm implies a block in the AV node or proximal His and an escape rhythm originating in the AV junction. Narrow QRS escape rhythms are typically faster and more stable than wide QRS escape rhythms and originate more proximally in the AV conduction system.

ETIOLOGY OF AV CONDUCTION DISEASE

There are numerous causes of intrinsic AV node dysfunction. Fibrosis and sclerosis of the conduction system are the most common causes of acquired conduction disease, accounting for ~50% of AV block. Numerous conditions that may not be distinguishable from one another can lead to conduction system fibrosis. Senile degeneration of the conduction system is most commonly seen in the elderly and results from idiopathic fibrosis and calcification. Lev's disease results from proximal bundle branch fibrosis. Lenègre's disease results from a sclerodegenerative process that occurs in a younger age group and involves the more distal portions of the bundle branches. Calcification of the aortic valve annulus can encroach on the conduction system. Compared to aortic valve calcification, mitral calcification is less commonly a cause of AV block **(Table 252-2)**.

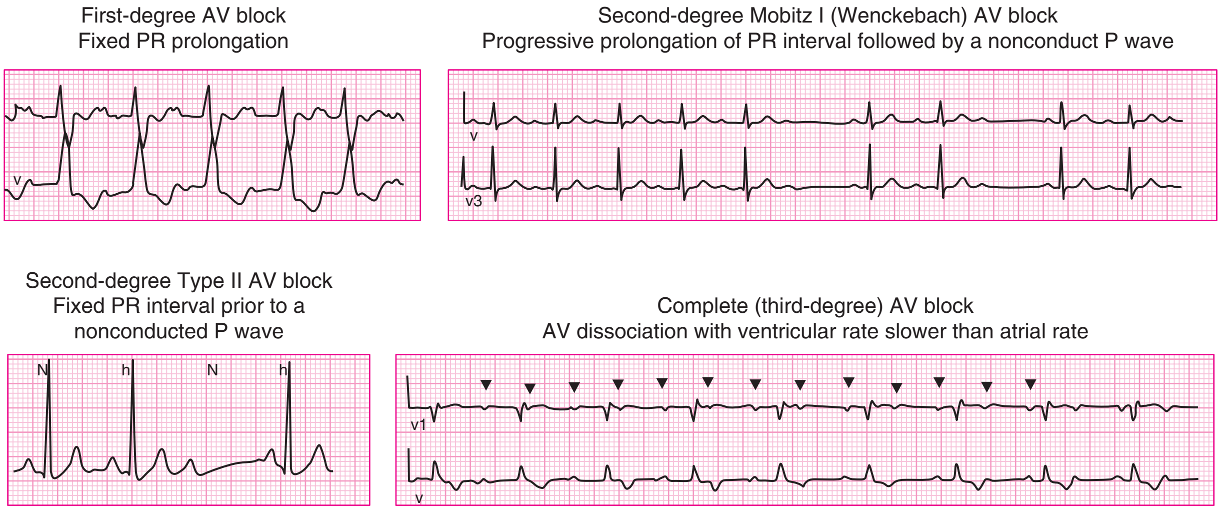
First-degree AV block
Fixed PR prolongation

Second-degree Mobitz I (Wenckebach) AV block
Progressive prolongation of PR interval followed by a nonconduct P wave

Second-degree Type II AV block
Fixed PR interval prior to a
nonconducted P wave

Complete (third-degree) AV block
AV dissociation with ventricular rate slower than atrial rate

FIGURE 252-1 Types of atrioventricular (AV) block. The *upper left* figure displays fixed prolongation of the PR interval. The *upper right* figure demonstrates Mobitz I block (Wenckebach AV block) manifested as progressive prolongation of the PR interval followed by a nonconducted P wave ("dropped beat"). The *lower left* figure displays AV block with P wave with no QRS complex and no associated PR prolongation prior to the dropped beat (Mobitz type II AV block). The *lower right* figure demonstrates complete heart block manifested as dissociation between P waves and QRS complexes (AV dissociation).

IATROGENIC CAUSES

AV block may also be from iatrogenic causes. Cardiac surgery, most commonly cardiac valve surgery, can result in damage to the AV conduction system, with the highest risk occurring in aortic valve and tricuspid valve surgery. Percutaneous transcatheter aortic valve replacement, septal myectomy, and alcohol septal ablation also carry a risk of AV block. Percutaneous catheter ablation for atrial arrhythmias, particularly AV nodal reentry tachycardia ablation, is associated with a small (<1%) risk of AV block. Medications, including beta blockers, verapamil, diltiazem, and digoxin, are also common iatrogenic causes of AV block. Many patients with drug-induced AV block have preexisting conduction system disease. Iatrogenic AV block may occur rarely in the setting of thoracic radiation or chemotherapy. AV block is a rare complication of the surgical repair of ventricular septal defects (VSDs) or atrial septal defects (ASDs) but may complicate repairs of transposition of the great arteries.

TABLE 252-2 Causes of AV Block

- Fibrosis/sclerosis/calcification of the conduction system
  - Senile degeneration of the conduction system (Lev's disease)
  - Lenègre's disease
  - Calcification of the aortic valve annulus (mitral—less common)
- Iatrogenic
  - After cardiac surgery (including valve surgery)
  - TAVR/alcohol septal ablation
  - Complication from catheter ablation
  - Medication (beta blockers, verapamil, diltiazem, digoxin)
  - Toxin/overdose/poisoning
- Acute MI/coronary ischemia
- Infectious causes
  - Lyme carditis
  - Bacterial endocarditis with perivalvular abscess
  - Viral myocarditis
  - Chagas' disease
  - Toxoplasmosis
- Infiltrative heart disease/inflammatory disease
  - Sarcoidosis
  - Amyloid
  - Rheumatologic disease: reactive arthritis (Reiter's syndrome), SLE, RA, systemic sclerosis
- Congenital AV block
  - Maternal lupus
  - Idiopathic congenital AV block
  - Congenital heart defects
  - Genetic
- Endocrine (e.g., thyroid disease, hypoadosteronism)
- Autoimmune disease
- Neuromuscular disease (e.g., myotonic dystrophy, Kearns-Sayre syndrome, Erb's dystrophy)
- Lymphoma
- Enhanced vagal tone/neurocardiogenic

Abbreviations: AV, atrioventricular; MI, myocardial infarction; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; TAVR, transcatheter aortic valve replacement.

AV BLOCK IN THE SETTING OF MYOCARDIAL ISCHEMIA

Coronary artery disease may produce transient or persistent AV block. In the setting of coronary spasm, ischemia, particularly in the right coronary artery distribution, may produce transient AV block. In acute myocardial infarction (MI), AV block transiently develops in 10–25% of patients; most commonly, this is first- or second-degree AV block, but complete heart block (CHB) may also occur. Second-degree and higher-grade AV block tends to occur more often in inferior than in anterior acute MI; however, the level of block in inferior MI tends to be in the AV node with more stable, narrow escape rhythms. In contrast, acute anterior MI is associated with block in the distal AV nodal complex, His bundle, or bundle branches and results in wide complex, unstable escape rhythms and a worse prognosis with high mortality rates.

AV conduction abnormalities may be caused by either direct ischemia to the conduction system or enhanced autonomic tone (Bezold-Jarisch reflex). Conduction abnormalities can be considered based on infarct location, and this may also predict which conduction abnormalities may be reversible. High-grade AV block associated with inferior MI is often located proximal to the His bundle in 90% of patients. Narrow junctional escape typically occurs with rates >40 beats/min, and a temporary pacemaker is typically not required as the AV block is often reversible and successfully managed with pharmacologic therapy. High-grade AV block in the setting of anterior MI is typically indicative of extensive infarction, is more often distal to the AV node, and is associated with an increased mortality. Temporary pacing in this circumstance is typically indicated. CHB in the setting of anterior MI may also be preceded by RBB block due to the arterial supply of the proximal RBB.

INFECTIOUS CAUSES OF AV BLOCK

Infection can also cause AV block. AV block is a common manifestation of Lyme carditis due to infection with *Borrelia burgdorferi*.

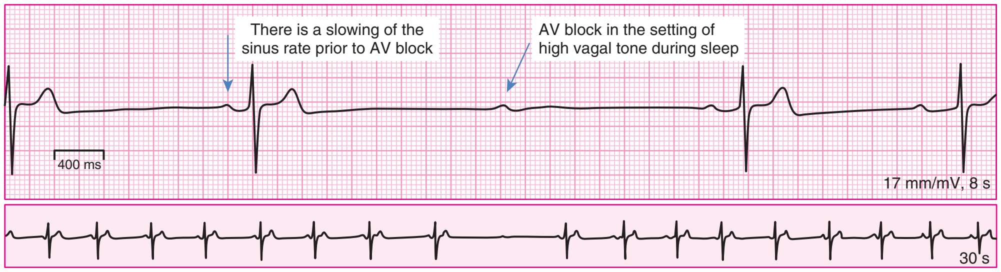
There is a slowing of the sinus rate prior to AV block
AV block in the setting of high vagal tone during sleep
400 ms
17 mm/mV, 8 s
30 s

FIGURE 252-2 Evidence of atrioventricular (AV) block during sleep. During sleep, increased vagal tone leads to sinus bradycardia with associated Mobitz I (Wenckebach) AV block. See Fig. 252-1 for an explanation of Mobitz I block.

AV block is typically at the level of the AV node with narrow junctional escape rhythm >40 beats/min. Less commonly, conduction abnormalities can occur below the level of the AV node or in the sinus node. In Lyme carditis, AV block typically improves within 1 week of antibiotic therapy, although a longer time frame can occur in some patients. AV block in the setting of infective endocarditis should raise concern for perivalvular abscess, which requires urgent surgical consultation. Viral myocarditis, Chagas' disease, and toxoplasmosis are less common infectious causes of AV block. Infiltrative heart disease such as cardiac sarcoid, amyloid, and hemochromatosis can also present as AV block. Autoimmune diseases, including systemic lupus erythematosus (SLE), rheumatoid arthritis, mixed connective tissue disease, and scleroderma, may cause AV block due to infiltration of the conduction system. Rare malignancies also may impair AV conduction.

AUTONOMIC AND FUNCTIONAL CAUSES OF AV BLOCK

Functional causes of AV block (autonomic, metabolic/endocrine, and drug-related) tend to be reversible. Most other etiologies produce structural changes, typically fibrosis, in segments of the AV conduction axis that are generally permanent. Heightened vagal tone during sleep or in well-conditioned individuals can be associated with all grades of AV block **(Fig. 252-2)**.

Carotid sinus hypersensitivity, vasovagal syncope, and cough and micturition syncope may be associated with SA node slowing and AV conduction block. Transient metabolic and endocrinologic disturbances and a number of pharmacologic agents also may produce reversible AV conduction block.

ACQUIRED AV BLOCK FROM FIBROSIS AND INFILTRATIVE CARDIOMYOPATHIES

Idiopathic progressive fibrosis of the conduction system is one of the more common and degenerative causes of AV conduction block. Aging is associated with degenerative changes in the summit of the ventricular septum, central fibrous body, and aortic and mitral annuli and has been described as "sclerosis of the left cardiac skeleton." The process typically begins in the fourth decade of life and may be accelerated by atherosclerosis, hypertension, and diabetes mellitus. Accelerated forms of progressive familial heart block have been identified in families with mutations in the cardiac sodium channel gene (*SCN5A*) and other loci that have been mapped to chromosomes 1 and 19.

AV conduction block has been associated with heritable neuromuscular diseases, including the nucleotide repeat disease myotonic dystrophy, the mitochondrial myopathy Kearns-Sayre syndrome, and several of the monogenic muscular dystrophies. Patients with these neuromuscular diseases and evidence of AV block should be considered for permanent pacing support.

CONGENITAL AV BLOCK

Congenital AV block may be observed in complex congenital cardiac anomalies, such as transposition of the great arteries, ostium primum ASDs, VSDs, endocardial cushion defects, and some single-ventricle defects. Congenital AV block in the setting of a structurally normal heart can occur sporadically and can also be seen in children born to mothers with SLE and other autoimmune diseases.

DIAGNOSTIC TESTING

Patients with conduction abnormalities should be evaluated for the presence or absence of structural heart disease. Physical exam may reveal valvular heart disease. ECG may suggest concomitant disease that predisposes to conduction abnormalities. Echocardiography is also indicated to evaluate for structural heart disease including valvular abnormalities, ejection fraction, and ventricular wall motion. Because age-dependent progressive fibrosis of the conduction system is the most common cause, AV node block that develops at a younger age (≤60 years) warrants advanced imaging such as chest computed tomography (CT), cardiac magnetic resonance imaging (MRI), or CT/positron emission tomography (PET) scan to further evaluate for infiltrative heart disease such as sarcoidosis. Advanced imaging may also be warranted based on other factors in the history and testing that suggest the need for evaluation of infiltrative heart disease. Evaluation for cardiac ischemia should be driven by the clinical suspicion at presentation (e.g., symptoms of ischemia, ECG abnormalities, etc.) **(Fig. 252-3)**.

Diagnostic testing in the evaluation of AV block is aimed at determining the level of conduction block, particularly in asymptomatic patients, since the prognosis and therapy depend on whether the block is in or below the AV node. Vagal maneuvers, carotid sinus massage, exercise, and administration of drugs such as atropine and isoproterenol may be diagnostically informative. Owing to the differences in the innervation of the AV node and infranodal conduction system, vagal stimulation and carotid sinus massage slow conduction in the AV node but have less of an effect on infranodal tissue and may even appear to improve conduction due to a reduced rate of activation of distal tissues. Conversely, atropine, isoproterenol, and exercise improve conduction through the AV node and may appear to impair infranodal conduction. In patients with congenital CHB and a narrow QRS complex, exercise typically increases heart rate; by contrast, those with acquired CHB, particularly with wide QRS, do not respond to exercise with an increase in heart rate.

Additional diagnostic evaluation, including electrophysiologic testing, may be indicated in patients with syncope and suspected high-grade AV block. This is particularly relevant if noninvasive testing does not reveal the cause of syncope or if the patient has structural heart disease with ventricular tachyarrhythmias as a cause of symptoms. Electrophysiologic testing provides more precise information regarding the location of AV conduction block and permits studies of AV conduction under conditions of pharmacologic stress and exercise. Recording of the His bundle electrogram by a catheter positioned at the superior margin of the tricuspid valve annulus provides information about conduction at all levels of the AV conduction axis. A properly recorded His bundle electrogram reveals local atrial activity, the His electrogram, and local ventricular activation; when it is monitored simultaneously with recorded body surface ECG traces, intra-atrial, AV nodal, and infranodal conduction times can be assessed. The time from the most rapid deflection of the atrial electrogram in the His bundle recording to the His electrogram (*AH interval*) represents conduction through the AV node and is normally <130 ms. The time from the His electrogram to the earliest onset of the QRS on the surface ECG (*HV interval*) represents the conduction time through the His-Purkinje system and is normally ≤55 ms **(Fig. 252-4)**.

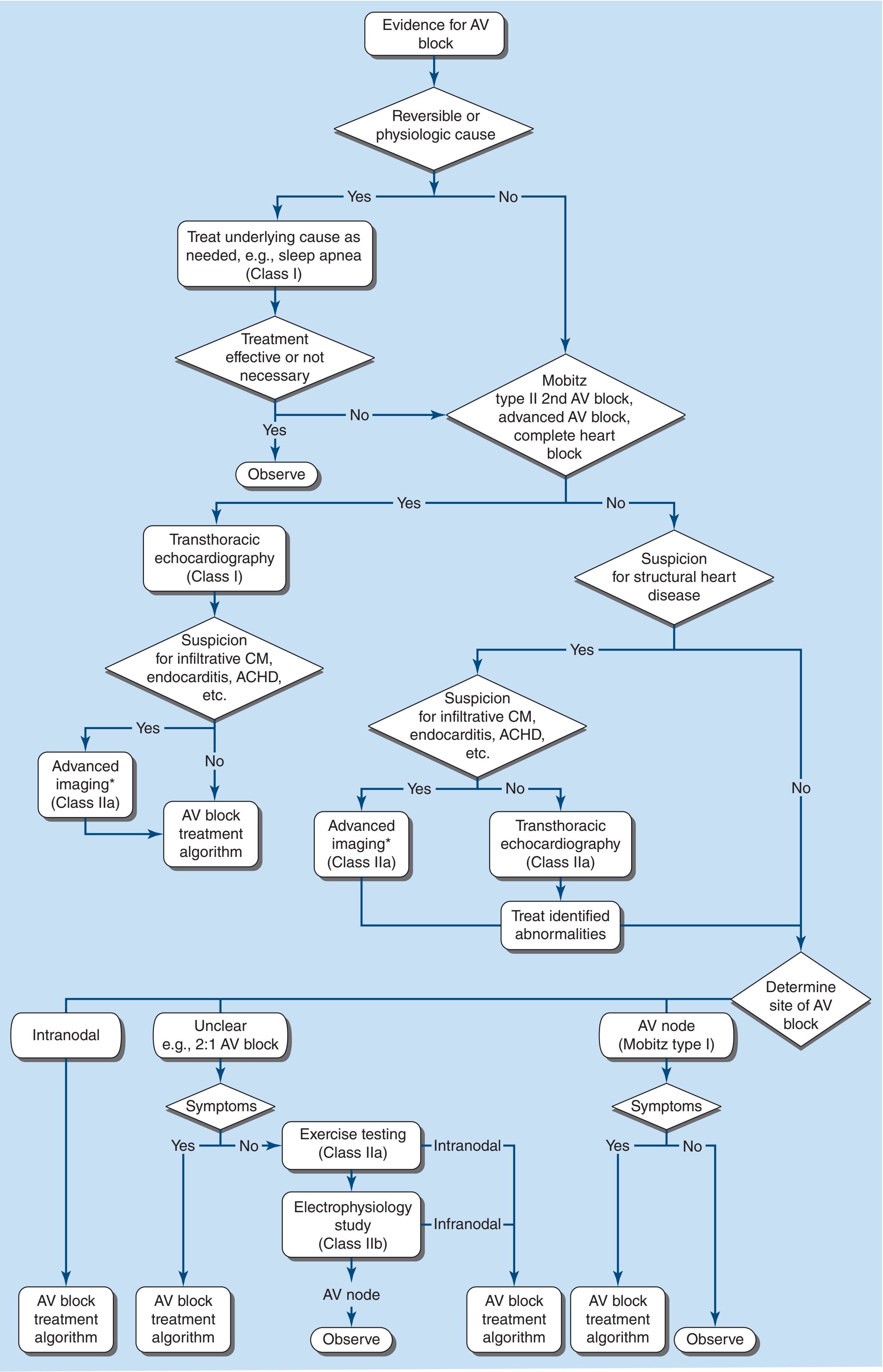
Evidence for AV block
Reversible or physiologic cause
Yes
No
Treat underlying cause as needed, e.g., sleep apnea (Class I)
Treatment effective or not necessary
No
Yes
Observe
Mobitz type II 2nd AV block, advanced AV block, complete heart block
Yes
No
Transthoracic echocardiography (Class I)
Suspicion for structural heart disease
Suspicion for infiltrative CM, endocarditis, ACHD, etc.
Yes
No
Advanced imaging* (Class IIa)
AV block treatment algorithm
Yes
Suspicion for infiltrative CM, endocarditis, ACHD, etc.
Yes
No
Advanced imaging* (Class IIa)
Transthoracic echocardiography (Class IIa)
No
Treat identified abnormalities
Determine site of AV block
Intranodal
Unclear e.g., 2:1 AV block
AV node (Mobitz type I)
Symptoms
Yes
No
Exercise testing (Class IIa)
Intranodal
Electrophysiology study (Class IIb)
Infranodal
AV node
Symptoms
Yes
No
AV block treatment algorithm
AV block treatment algorithm
Observe
AV block treatment algorithm
AV block treatment algorithm
Observe

FIGURE 252-3 Initial evaluation of suspected atrioventricular (AV) block algorithm. *Targeted advanced imaging—magnetic resonance imaging (MRI): amyloidosis, myocarditis, hemochromatosis, sarcoidosis, congenital heart disease (CHD), sinus of Valsalva aneurysm, aortic dissection, arrhythmogenic right ventricular cardiomyopathy; fluorodeoxyglucose-positron emission tomography (FDG-PET): sarcoidosis; technetium-99m pyrophosphate (Tc PYP) or 99m technetium 3,3-diphosphono-1,2-propanodicarboxylic acid (TC-DPD): transthyretin (TTR) amyloidosis; cardiac computed tomography (CT): CHD, sinus of Valsalva. ACHD, adult CHD; CM, cardiomyopathy. *(Reproduced with permission from FM Kusumoto et al: 2018 ACC/AHA/HRS guideline on the evaluation and management of patients with bradycardia and cardiac conduction delay. Heart Rhythm 16:e128, 2019.)*

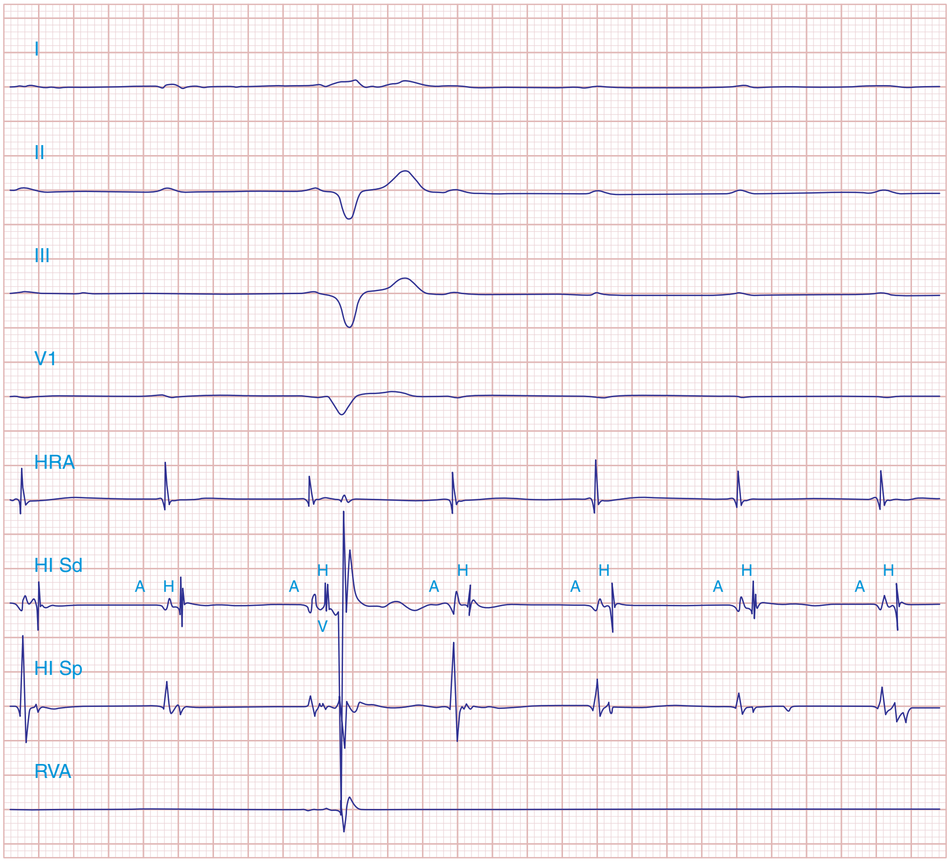

FIGURE 252-4 High-grade atrioventricular (AV) block below the His. The AH interval is normal and is not changing before the block. Atrial and His bundle electrograms are recorded consistent with block below the distal AV junction. I, II, III, and V_1 are surface electrocardiogram leads. HISp, HISd, and RVA are the proximal HIS, distal HIS, and right ventricular apical electrical recordings, respectively. A, H, and V represent the atrial, His, and ventricular electrograms on the His bundle recording, respectively. *(Courtesy of Dr. Joseph Marine.)*

Rate stress produced by pacing can unmask abnormal AV conduction. Mobitz I second-degree AV block at short atrial paced cycle lengths is a normal response. However, when it occurs at atrial cycle lengths >500 ms (<120 beats/min) in the absence of high vagal tone, it is abnormal. Typically, type I second-degree AV block is associated with prolongation of the AH interval, representing conduction slowing and block in the AV node. AH prolongation occasionally is due to the effect of drugs (beta blockers, calcium channel blockers, digitalis) or increased vagal tone. Atropine can be used to reverse high vagal tone; however, if AH prolongation and AV block at long pacing cycle lengths persist, intrinsic AV node disease is likely. Type II second-degree block is typically infranodal, often in the His-Purkinje system. Block below the node with prolongation of the HV interval or a His bundle electrogram with no ventricular activation is abnormal unless it is elicited at fast pacing rates or short coupling intervals with extra stimulation. It is often difficult to determine the type of second-degree AV block when 2:1 conduction is present; however, the finding of a His bundle electrogram after every atrial electrogram indicates that block is occurring in the distal conduction system.

Intracardiac recording at an electrophysiologic study that reveals prolongation of conduction through the His-Purkinje system (i.e., long HV interval) is associated with an increased risk of progression to higher grades of block and is generally an indication for pacing. In the setting of bundle branch block, the HV interval may reveal the condition of the unblocked bundle and the prognosis for developing more advanced AV conduction block. Prolongation of the HV interval in patients with asymptomatic bundle branch block is associated with an increased risk of developing higher-grade AV block. The risk increases with greater prolongation of the HV interval such that in patients with an HV interval >100 ms, the annual incidence of complete AV block approaches 10%, indicating a need for pacing. In patients with acquired CHB, even if intermittent, there is little role for electrophysiologic testing, and pacemaker implantation is almost always indicated.

TREATMENT

Acute Management of AV Conduction Block

The first-line strategies for management of AV block should be to eliminate reversible causes and to determine the immediate safety and reliability of the heart rhythm (e.g., escape rhythm) and whether or not temporary or permanent pacing is warranted. The need for temporary pacing is determined by the symptoms of the patient, hemodynamic status, and the estimate of the level at which AV block is present. In a general sense, the lower in the conduction system that an escape rhythm is occurring, the lower is the reliability of the escape rhythm. A narrow-complex junctional escape of 45 beats/min with no symptoms does not warrant urgent temporary pacing, whereas a wide-complex (implying block lower in the conduction system) escape rhythm at 30 beats/min does. Elimination of unnecessary medications known to slow AV conduction (e.g., beta blockers, diltiazem, verapamil, digoxin), correction of electrolyte abnormalities, ischemia, and inhibition of excessive vagal tone may increase the heart rate. Adjunctive pharmacologic treatment with atropine or isoproterenol may be useful if the block is in the AV node. When pacing is indicated, the most expeditious technique is the use of transcutaneous pacing, where pacing patches are placed anteriorly over the cardiac apex (cathode) and posteriorly between the spine and the scapula or above the right nipple (anode). Acutely, transcutaneous pacing is highly effective, but its duration is limited by patient discomfort and longer-term failure to capture the ventricle owing to changes in lead impedance. Transvenous temporary pacing is more reliable, and a pacing wire can be placed from the jugular, subclavian, or femoral venous system and advanced to the right ventricle, permitting stable temporary pacing.

AV conduction abnormalities may be reversible in certain circumstances including removal of unnecessary medication or toxins,

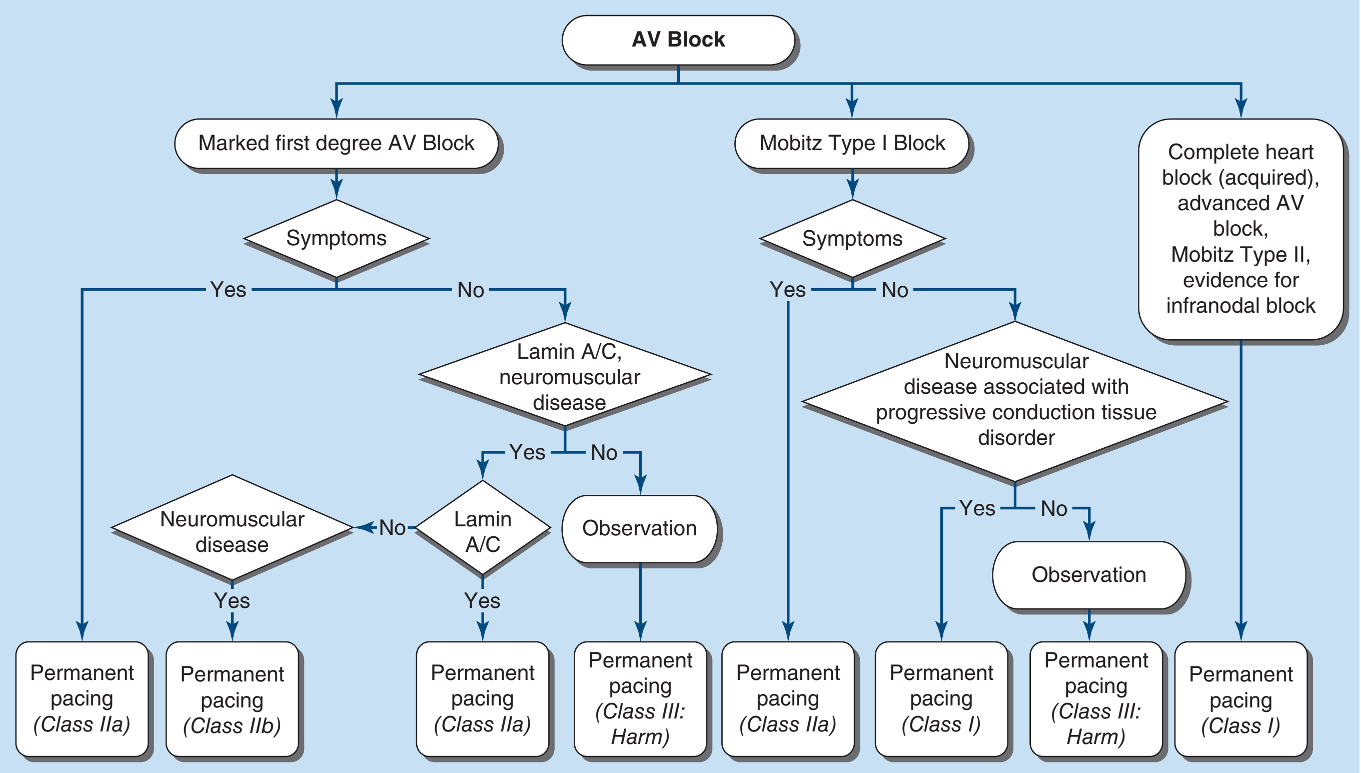
AV Block
Marked first degree AV Block
Mobitz Type I Block
Complete heart block (acquired), advanced AV block, Mobitz Type II, evidence for infranodal block
Symptoms
Yes
No
Symptoms
Yes
No
Lamin A/C, neuromuscular disease
Yes
No
Neuromuscular disease associated with progressive conduction tissue disorder
Neuromuscular disease
No
Lamin A/C
Observation
Yes
No
Observation
Yes
Yes
Permanent pacing (Class IIa)
Permanent pacing (Class IIb)
Permanent pacing (Class IIa)
Permanent pacing (Class III: Harm)
Permanent pacing (Class IIa)
Permanent pacing (Class I)
Permanent pacing (Class III: Harm)
Permanent pacing (Class I)

FIGURE 252-5 Indications for pacing in patients with atrioventricular (AV) block. In patients presenting with AV block, the category of AV block should be determined (first-degree, second-degree, or complete heart block). In first-degree AV block, permanent pacing may be indicated in the setting of symptoms or higher-risk systemic disease such as neuromuscular disease or Lamin A/C cardiomyopathy. In Mobitz I AV block, pacing may be considered in the setting of symptoms or the additional disease mentioned with first-degree AV block. In complete heart block or Mobitz II AV block, permanent pacing is generally indicated. Class I recommendations should be performed or are indicated. Class IIa recommendations are considered reasonable to perform. Class IIb recommendations may be considered. Class III recommendations are associated with harm more than benefit. *(Reproduced with permission from FM Kusumoto et al: 2018 ACC/AHA/HRS guideline on the evaluation and management of patients with bradycardia and cardiac conduction delay. Heart Rhythm 16:e128, 2019.)*

correction of electrolyte abnormalities, relief of ischemia, treatment of certain infiltrative heart disease (e.g., immunosuppression in cardiac sarcoidosis), and treatment of sleep apnea in patients with nocturnal vagally mediated AV block. When symptomatic or infranodal AV block is not reversible, which is often the case, permanent pacing is warranted.

PERMANENT PACEMAKER IMPLANTATION

The indications for pacing in AV conduction block are shown in **Fig. 252-5**. In patients with acquired Mobitz type II AV block, high-grade AV block, or third-degree AV block that is not reversible or physiologic, permanent pacing is recommended regardless of symptoms. For all other types of AV block, in the absence of conditions associated with progressive AV conduction abnormalities, permanent pacing should generally be considered only in the presence of symptoms that correlate with block. In patients with neuromuscular disease and other progressive cardiomyopathies affecting the conduction system, permanent pacemaker implantation is recommended for marked first-degree AV block and Mobitz I AV block. Pacemaker implantation should be performed in any patient with symptomatic bradycardia and irreversible second- or third-degree AV block, regardless of the cause or level of block in the conducting system. Symptoms may include those directly related to bradycardia and low cardiac output or to worsening heart failure, angina, or intolerance to an essential medication. Pacing in patients with asymptomatic AV block should be individualized; situations in which pacing should be considered are patients with acquired CHB, particularly in the setting of cardiac enlargement; left ventricular dysfunction; and waking heart rates ≤40 beats/min. Patients who have asymptomatic second-degree AV block of either type should be considered for pacing if the block is demonstrated to be intra- or infra-His or is associated with a wide QRS complex. Pacing may be indicated in asymptomatic patients in special circumstances, in patients with profound first-degree AV block and left ventricular dysfunction in whom a shorter AV interval produces hemodynamic improvement, and in the setting of milder forms of AV conduction delay (first-degree AV block, intraventricular conduction delay) in patients with neuromuscular diseases that have a predilection for the conduction system, such as myotonic dystrophy and other muscular dystrophies and Kearns-Sayre syndrome.

AV block in acute MI is often transient, particularly in inferior infarction. The circumstances in which pacing is indicated in acute MI are persistent second- or third-degree AV block, particularly if symptomatic, and transient second- or third-degree AV block associated with bundle branch block. Pacing is generally not indicated in the setting of transient AV block in the absence of intraventricular conduction delays or in the presence of fascicular block or first-degree AV block that develops in the setting of preexisting bundle branch block. Fascicular blocks that develop in acute MI in the absence of other forms of AV block also do not require pacing.

Distal forms of AV conduction block may require pacemaker implantation in certain clinical settings. Patients with bifascicular or trifascicular block and symptoms, particularly syncope that is not attributable to other causes, should undergo pacemaker implantation. Permanent pacemaker implantation is indicated in asymptomatic patients with bifascicular or trifascicular block who experience intermittent third-degree block, type II second-degree AV block, or alternating bundle branch block. In patients with fascicular block who are undergoing electrophysiologic study, a markedly prolonged HV interval or block below the His at long cycle lengths also may constitute an indication for permanent pacing. Patients with fascicular block and the neuromuscular diseases previously described should also undergo pacemaker implantation.

SELECTION OF PACING MODE AND SYSTEM

In general, a pacing mode that maintains AV synchrony reduces complications of pacing such as pacemaker syndrome and pacemaker-mediated tachycardia. This is particularly true in younger patients; the importance of dual-chamber pacing in the elderly, however, is

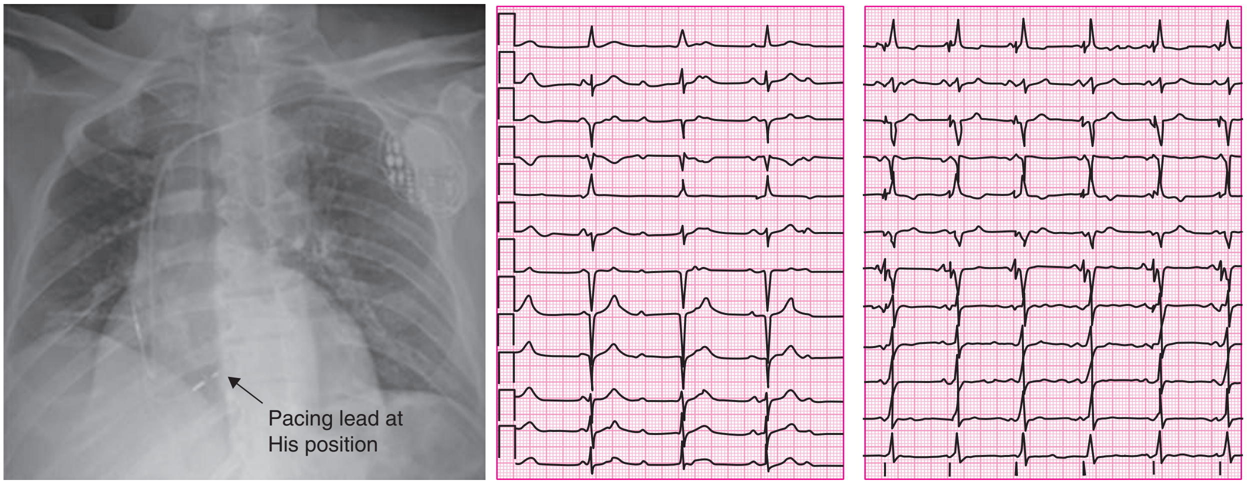

FIGURE 252-6 Left bundle branch area pacing. The chest radiograph on the left shows a dual-chamber pacemaker with a pacing lead in the right atrium (*upper left*) and a pacing lead in the region of proximal ventricular septum to achieve physiologic pacing by pacing deep enough in the septum to achieve left bundle branch area pacing. The electrocardiograms demonstrated intrinsic conduction with complete heart block on the left and atrial sensed, ventricular paced rhythm with a narrow QRS complex by capturing the specialized conduction system of the heart.

less well established, although AV synchrony in patients with sinus rhythm and AV block is typically desired.

Physiologic Ventricular Pacing In patients with left ventricular ejection fraction <50% and AV block who have an indication for permanent pacing and are expected to require ventricular pacing >40% of the time, techniques to provide more physiologic ventricular activation are preferred to right ventricular pacing to prevent heart failure. Cardiac resynchronization therapy (CRT) involves placement of an additional pacing lead in a lateral or anterolateral branch of the coronary sinus to allow for simultaneous right ventricle and lateral left ventricle pacing leading to a more physiologic left ventricular contraction. CRT pacing has been shown to improve outcomes and mortality in appropriately selected patients. Physiologic ventricular pacing has also been achieved with placement of a ventricular pacing lead in the region of the His bundle. His bundle pacing recruits the specialized conduction system, leading to a more physiologic cardiac contraction. In addition to His bundle pacing, left bundle branch area pacing in the proximal interventricular septal region has also been shown to achieve a more physiologic pacing response. The selection of pacing lead location should be individualized based on expected pacing requirements, underlying heart disease, and level of conduction block **(Fig. 252-6)**.

The availability of leadless miniaturized pacing systems may be appropriate in selected patients. Leadless pacemakers are completely self-contained devices that are implanted via the femoral vein into the right atrium and ventricle. The technology for these devices continues to evolve, and the most recent models are capable of detecting atrial mechanical contraction to allow for the preservation of AV synchrony. The device can provide single- or dual-chamber pacing in addition to containing technology that can sense atrial activity (utilizing the accelerometer in the pacemaker) to coordinate an atrial sensed, ventricular paced rhythm. Leadless pacemakers can be particularly useful in patients with vascular access limitations. Because there is no intravascular pacing wire or implanted subcutaneous pacemaker generator, the long-term infection rate is lower and there is no risk of lead fracture **(Fig. 252-7)**.

Several studies have failed to demonstrate a difference in mortality rate in older patients with AV block treated with a single- (VVI) compared with a dual- (DDD) chamber pacing mode. In some of the studies that randomized pacing mode, the risk of chronic atrial fibrillation and stroke risk decreased with physiologic pacing. In patients with sinus rhythm and AV block, the very modest increase in risk with dual-chamber pacemaker implantation appears to be justified to avoid the possible complications of single-chamber pacing.

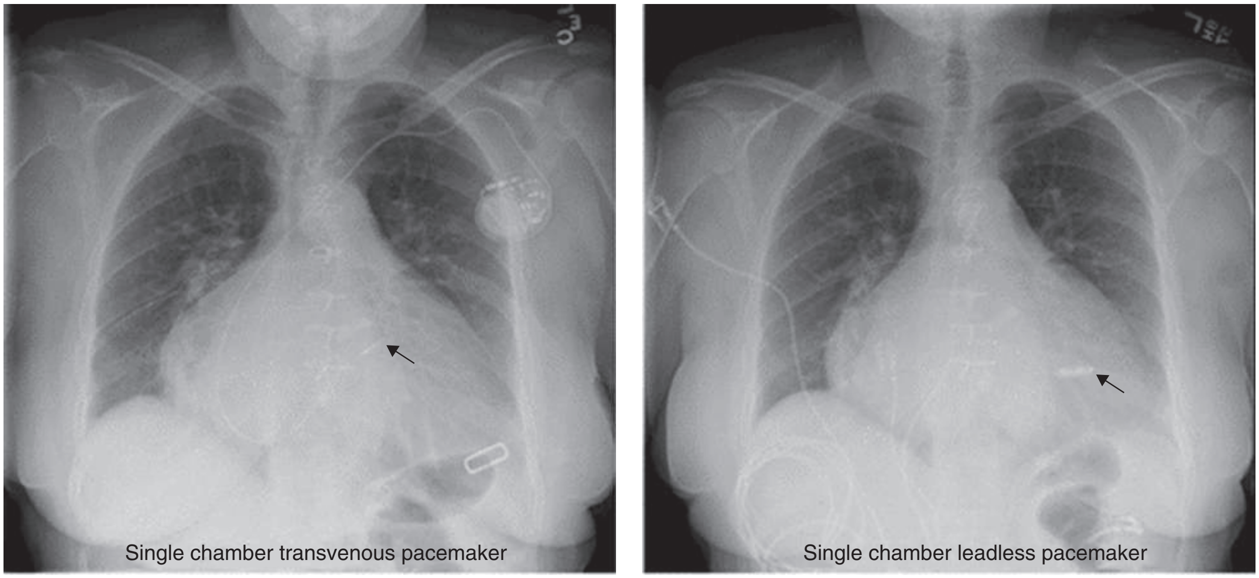
Single chamber transvenous pacemaker

Single chamber leadless pacemaker

FIGURE 252-7 Types of pacemakers. A single-chamber pacemaker with the pacing lead in the right ventricular outflow tract (*arrow*) is shown on the left. A single-chamber leadless pacemaker (*arrow*) is shown on the right.

FURTHER READING

Jalife J, Stevenson W (eds): *Zipes and Jalife's Cardiac Electrophysiology: From Cell to Bedside*, 8th ed. Philadelphia, Elsevier, 2022.

Kusumoto FM et al: 2018 ACC/AHA/HRS guideline on the evaluation and management of patients with bradycardia and cardiac conduction delay: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. Heart Rhythm 16:e128, 2019.

Sidhu S, Marine JE: Evaluating and managing bradycardia. Trends Cardiovasc Med 30:265, 2020.

253

Approach to Supraventricular Arrhythmias

William H. Sauer, Paul C. Zei

Supraventricular arrhythmias refer to all abnormal heart rhythms originating above the level of the ventricle including the atria and the atrioventricular (AV) junction. Supraventricular arrhythmias include premature atrial contractions (PACs), wandering atrial pacemaker, sinus arrhythmia, nonphysiologic sinus tachycardia, atrial tachycardia, junctional tachycardia, atrial flutter, and atrial fibrillation. Supraventricular arrhythmias that result in an elevated heart rate (>100 beats/min) are broadly defined as *supraventricular tachycardias* (SVTs). SVTs originate from or are dependent on conduction through the atrium or AV node to the ventricles. Most produce narrow QRS complex tachycardia (QRS duration <120 ms) characteristic of ventricular activation over the Purkinje system and, thus, are sometimes referred to as a narrow-complex tachycardias. The QRS morphology of the SVT is usually identical to the sinus rhythm QRS. Conduction block in the left or right bundle branch or activation of the ventricles from an accessory pathway produces a wide QRS complex during SVT that must be distinguished from ventricular tachycardia (VT). Mechanisms of supraventricular tachyarrhythmia can be divided into physiologic sinus tachycardia and pathologic tachycardia **(Table 253-1)**.

Pathologic tachycardia can be further subclassified by mechanism as reentrant arrhythmias dependent on AV nodal conduction (e.g., AV nodal reentry tachycardia), large reentry circuits within the atrial tissue alone (e.g., atrial flutter), or focal atrial tachycardias that can be due to automaticity or small reentry circuits. The prognosis and treatment vary considerably depending on the mechanism and underlying heart disease. SVT can be of brief duration, termed *nonsustained*, or can be sustained such that an intervention, such as cardioversion, catheter ablation, or drug administration, is required for termination and maintenance of sinus rhythm. Episodes that occur with sudden onset and termination are referred to as paroxysmal. *Paroxysmal supraventricular tachycardia* (PSVT) refers to a family of tachycardias including AV node reentry, AV reciprocating tachycardia using an accessory pathway, and atrial tachycardia described in subsequent chapters **(Fig. 253-1)**.

Other supraventricular arrhythmias that may or may not be symptomatic include PACs, sinus arrhythmia, ectopic atrial rhythm, and accelerated junctional rhythm. Atrial flutter and atrial fibrillation may present as a tachycardia or may be adequately rate controlled with or without AV nodal blocking agents; these entities are discussed separately **(see Chaps. 255, 256, 257, and 258)**.

TABLE 253-1 Mechanisms of Supraventricular Arrhythmias

Physiologic Sinus Tachycardia
Defining feature: normal sinus mechanism precipitated by exertion, stress, exogenous or endogenous stimulants, concurrent illness
Pathologic Supraventricular Tachycardia (SVT)
A. Tachycardias originating from the atrium
Defining feature: tachycardia may continue despite beats that fail to conduct to the ventricles, indicating that the atrioventricular (AV) node is not participating in the tachycardia circuit
1. Inappropriate sinus tachycardia
Defining feature: tachycardia from the normal sinus node area that occurs without an identifiable precipitating factor as a result of dysfunctional autonomic regulation
2. Focal atrial tachycardia (AT)
Defining feature: regular atrial tachycardia with defined P wave; may be sustained, nonsustained, paroxysmal, or incessant; frequent sites of origin occur along the valve annuli of left or right atrium, pulmonary veins, coronary sinus musculature, superior vena cava
3. Atrial flutter and macroreentrant atrial tachycardia
Defining feature: macroreentry reflected as organized atrial activity on an electrocardiogram (ECG), commonly seen as sawtooth flutter waves at rates typically faster than 200 beats/min
4. Atrial fibrillation
Defining feature: chaotic rapid atrial electrical activity with variable ventricular rate; the most common sustained cardiac arrhythmia in older adults
5. Multifocal atrial tachycardia
Defining feature: multiple discrete P waves often seen in patients with pulmonary disease during acute exacerbations of pulmonary insufficiency
B. AV nodal reentry tachycardia (AVNRT)
Defining feature: paroxysmal regular tachycardia with P waves visible at the end of the QRS complex or not visible at all; the most common paroxysmal sustained tachycardia in healthy young adults; more common in women
C. Tachycardias associated with accessory atrioventricular pathways
1. Orthodromic AV reciprocating tachycardia (AVRT)
Defining feature: paroxysmal sustained tachycardia similar to AV nodal reentry; during sinus rhythm, evidence of ventricular preexcitation may be present (Wolff-Parkinson-White syndrome) or absent (concealed accessory pathway)
2. Preexcited tachycardia
Defining feature: wide QRS tachycardia with QRS morphology similar to ventricular tachycardia
a. Antidromic AV reciprocating tachycardia—regular paroxysmal tachycardia
b. Atrial fibrillation with preexcitation—irregular wide-complex or intermittently wide-complex tachycardia, some with dangerously rapid rates faster than 250/min
c. Atrial tachycardia or flutter with preexcitation
Other Supraventricular Arrhythmias
A. Premature atrial contractions (PACs)
Defining feature: sinus rhythm with an early atrial complex distinct from the sinus P wave resulting in an irregular rhythm. A pattern of ectopy (i.e., trigeminy, bigeminy) is sometimes seen with PACs
B. Sinus arrhythmia
Defining feature: irregular rhythm with a sinus P wave and with P-P intervals varying with respiration
C. Accelerated junctional rhythm
Defining feature: paroxysmal regular rhythm with P waves visible at the end of the QRS complex or not visible at all
D. Ectopic atrial rhythm
Defining feature: regular rhythm with a rate <100 beats/min but usually faster than sinus rhythm and with a P wave distinct from sinus that may or may not sustain

CLINICAL PRESENTATION

Symptoms of supraventricular arrhythmia vary depending on the rate, duration, associated heart disease, and comorbidities and include palpitations, chest pain, dyspnea, diminished exertional capacity, and occasionally syncope. Rarely, a supraventricular arrhythmia precipitates cardiac arrest in patients with Wolff-Parkinson-White (WPW) syndrome or severe heart disease, such as hypertrophic cardiomyopathy. Asymptomatic supraventricular arrhythmias are often captured on routine electrocardiographic (ECG) recordings and sometimes prompt

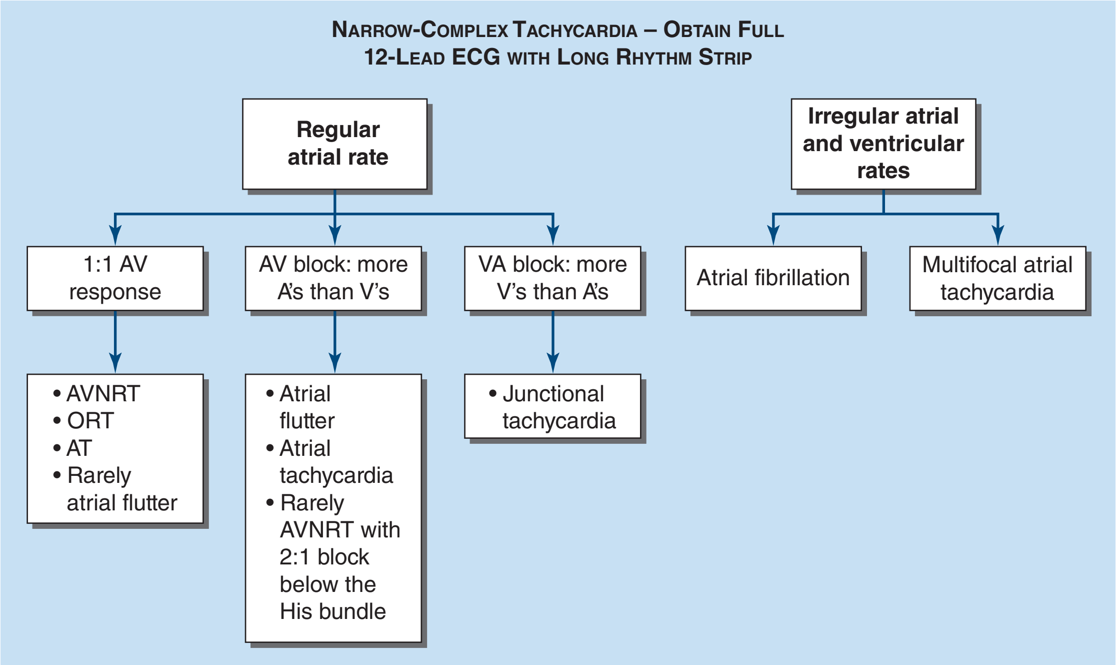
Narrow-Complex Tachycardia – Obtain Full 12-Lead ECG with Long Rhythm Strip
Regular atrial rate
Irregular atrial and ventricular rates
1:1 AV response
AV block: more A's than V's
VA block: more V's than A's
Atrial fibrillation
Multifocal atrial tachycardia
• AVNRT
• ORT
• AT
• Rarely atrial flutter
• Atrial flutter
• Atrial tachycardia
• Rarely AVNRT with 2:1 block below the His bundle
• Junctional tachycardia

FIGURE 253-1 Diagnostic possibilities based on the appearance of the 12-lead electrocardiogram (ECG) recorded during an episode of supraventricular tachycardia (SVT). AT, focal atrial tachycardia; AVNRT, atrioventricular (AV) nodal reentry tachycardia; ORT, orthodromic AV reentry tachycardia.

referral to a cardiologist. Most asymptomatic supraventricular arrhythmias do not require treatment or further evaluation.

INITIAL EVALUATION

The diagnosis of SVT is most often entertained when evaluating a patient for arrhythmia-related symptoms or when evidence of ventricular preexcitation is seen on an ECG as an outpatient. Diagnosis of SVT requires obtaining an ECG at the time of symptoms **(Fig. 253-2)**. Ventricular preexcitation on the resting ECG suggests AV reciprocating tachycardia using an accessory pathway. When the arrhythmia is ongoing at the time of recording, the ECG usually establishes or suggests the diagnosis. In the urgent care or inpatient setting, treatment of SVT will often involve vagal maneuvers or carotid sinus massage (CSM) to achieve AV block **(Table 253-2)**. In the appropriate patient, CSM should be used cautiously, if at all, if there is concern for carotid atherosclerosis that may be embolized during manipulation. If this is unsuccessful, the administration of 6 or 12 mg of adenosine to cause transient AV block is usually successful in terminating an AV nodal–dependent SVT or diagnosing a non-AV nodal–dependent SVT such as atrial tachycardia or atrial flutter. There are some atrial tachycardias that are adenosine sensitive, and thus, termination of an SVT with adenosine does not exclude this potential diagnosis.

For transient arrhythmias, ambulatory ECG recording is warranted. Patients will often have access to ECG recording devices, such as a watch or smartphone-enabled ECG recording electrode pair. Therefore, a patient may have an ECG diagnosis before seeing a physician **(Fig. 253-3)**.

Exercise testing is useful for assessing exercise-related symptoms and potentially evoking the arrhythmia. Additional evaluation for underlying cardiac disease and to exclude potentially dangerous arrhythmias should be performed based on the clinical scenario. Occasionally, an invasive electrophysiology study is warranted to provoke the arrhythmia with pacing, confirm the mechanism, and risk stratify the patient, but most commonly, this is performed at the time of intended catheter ablation to treat the arrhythmia.

Paroxysmal SVT is most commonly encountered in patients who do not have structural heart disease. Other supraventricular arrhythmias, particularly atrial fibrillation, are often associated with a variety of heart diseases. At initial evaluation, history and examination should assess possible underlying heart disease. Any abnormal findings may warrant further cardiac evaluation.

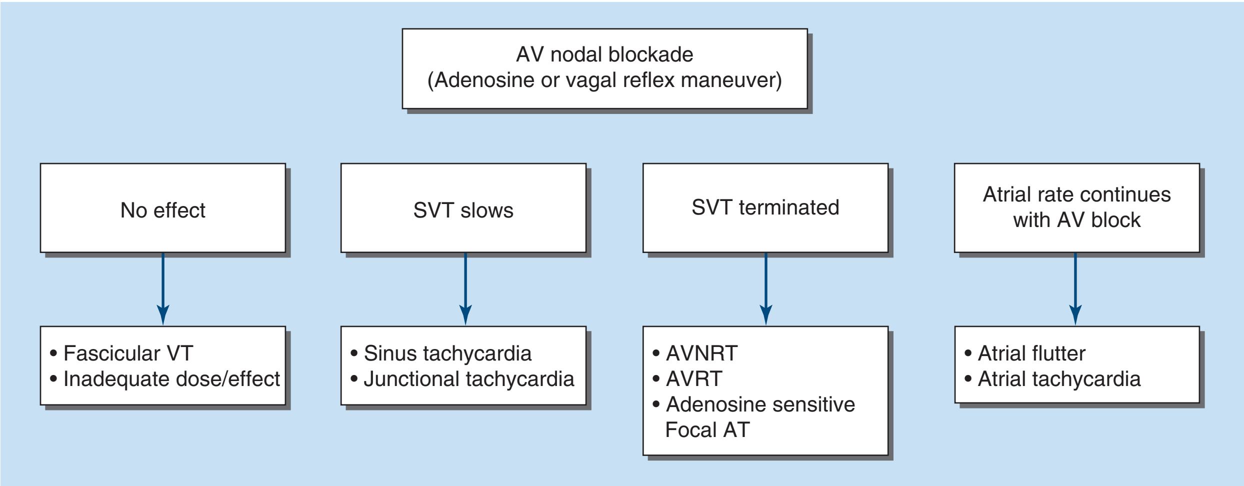
AV nodal blockade
(Adenosine or vagal reflex maneuver)
No effect
SVT slows
SVT terminated
Atrial rate continues with AV block
• Fascicular VT
• Inadequate dose/effect
• Sinus tachycardia
• Junctional tachycardia
• AVNRT
• AVRT
• Adenosine sensitive Focal AT
• Atrial flutter
• Atrial tachycardia

FIGURE 253-2 Diagnostic effect of increasing atrioventricular (AV) node blockade with vagal maneuvers, carotid sinus massage, adenosine, verapamil, or beta blockers. AT, focal atrial tachycardia; AVNRT, atrioventricular nodal reentry tachycardia; AVRT, atrioventricular reciprocating tachycardia; SVT, supraventricular tachycardia.

TABLE 253-2 Vagal Maneuvers

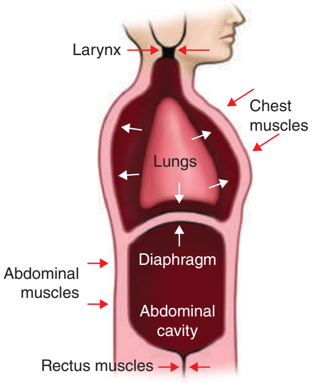 Larynx Chest muscles Lungs Diaphragm Abdominal muscles Abdominal cavity Rectus muscles **Holding breath while bearing down to increase intrathoracic pressure**	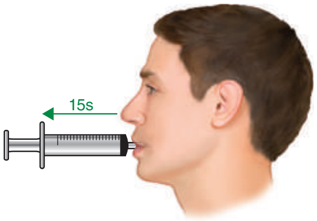 **Breathing hard into a syringe against pressure to increase intrathoracic pressure**	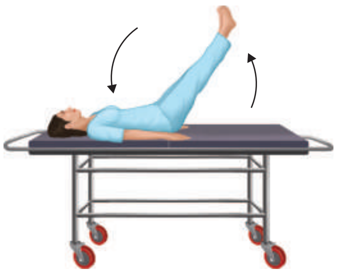 **Raise legs abruptly to increase venous return**
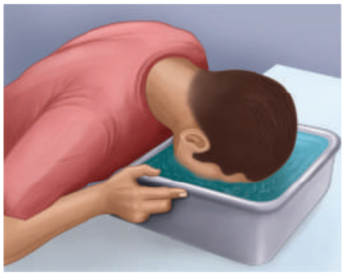 **Submerge face into cold water (diver's reflex)**	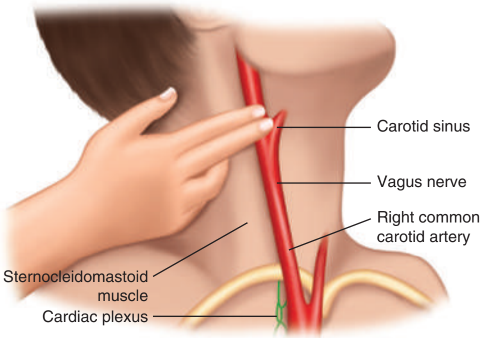 Carotid sinus Vagus nerve Right common carotid artery Sternocleidomastoid muscle Cardiac plexus **Carotid sinus massage**	**Adenosine**

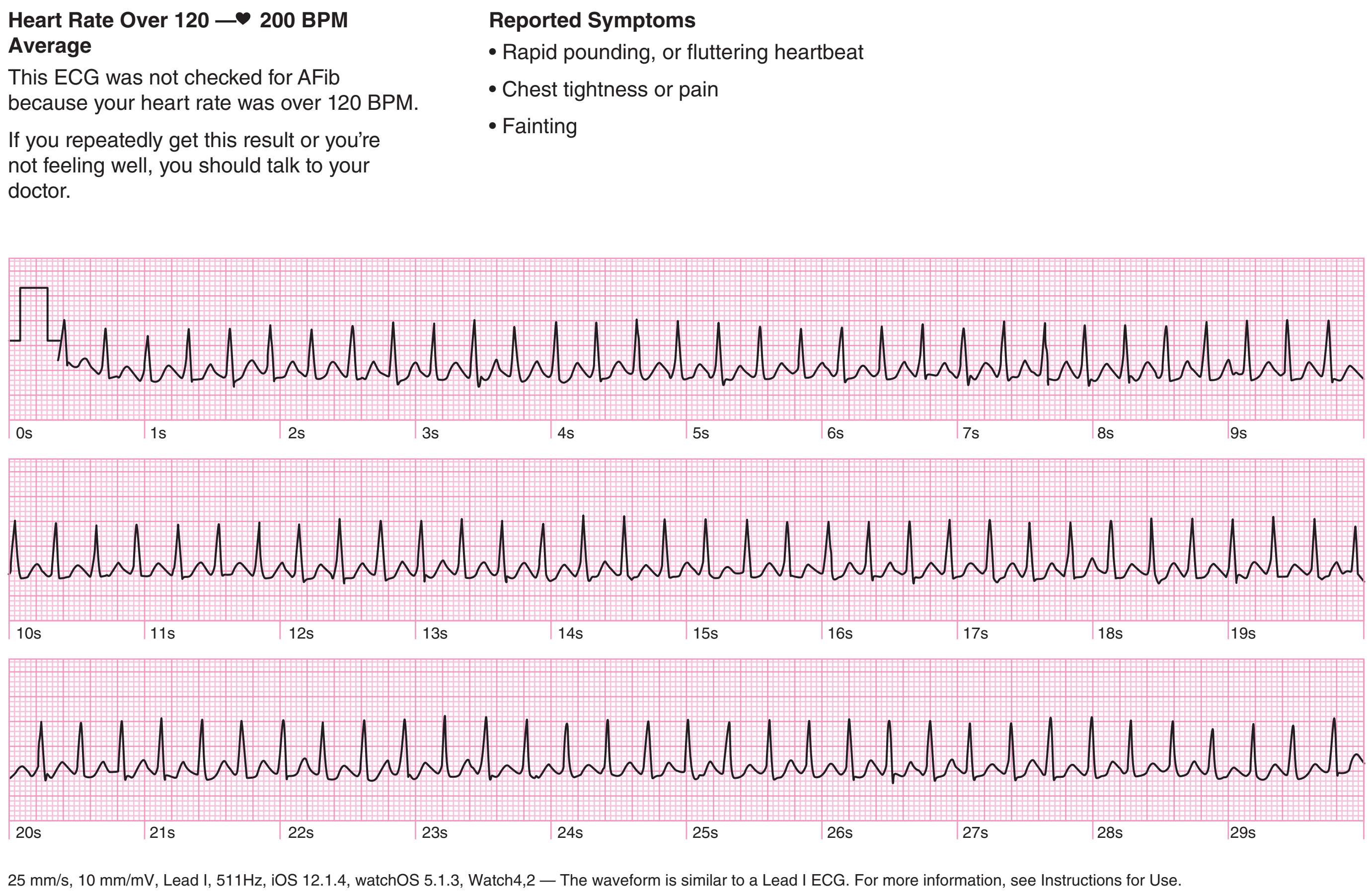

Heart Rate Over 120 —♥ 200 BPM Average

This ECG was not checked for AFib because your heart rate was over 120 BPM.

If you repeatedly get this result or you're not feeling well, you should talk to your doctor.

Reported Symptoms

- Rapid pounding, or fluttering heartbeat
- Chest tightness or pain
- Fainting

0s
1s
2s
3s
4s
5s
6s
7s
8s
9s
10s
11s
12s
13s
14s
15s
16s
17s
18s
19s
20s
21s
22s
23s
24s
25s
26s
27s
28s
29s

25 mm/s, 10 mm/mV, Lead I, 511Hz, iOS 12.1.4, watchOS 5.1.3, Watch4,2 — The waveform is similar to a Lead I ECG. For more information, see Instructions for Use.

FIGURE 253-3 Narrow-complex tachycardia recorded by a consumer wearable monitor (Apple watch). Afib, atrial fibrillation; ECG, electrocardiogram.

The most common SVT is sinus tachycardia in response to physiologic stress, such as exercise, but it can also be a manifestation of acute illness. The first step in diagnosis of SVT is to consider the possibility of sinus tachycardia. Therapy is then determined by the clinical findings and probable diagnosis. If sinus tachycardia is diagnosed, treatment of the underlying inciting cause is the primary approach. If the arrhythmia is ongoing and is not due to sinus tachycardia, initial assessment determines whether immediate therapy is needed to terminate the arrhythmia or slow the rate. Arrhythmias that cause hypotension, impaired consciousness, angina, or heart failure warrant immediate therapy, guided by the type of arrhythmia. Treatment options for specific types of SVT are discussed in more detail in subsequent chapters and include pharmacologic and procedural interventions.

■ FURTHER READING

Brugada J et al: 2019 ESC guidelines for the management of patients with supraventricular tachycardia. The task force for the management of patients with supraventricular tachycardia of the European Society of Cardiology (ESC) developed in collaboration with the Association for European Paediatric and Congenital Cardiology (AEPC). Eur Heart J 41:655, 2020.

Callans DJ: *Josephson's Clinical Cardiac Electrophysiology: Techniques and Interpretations*, 7th ed. Philadelphia, Wolters Kluwer, 2024.

254

Physiologic and Nonphysiologic Sinus Rhythm

William H. Sauer, Paul C. Zei

The sinus node is composed of a group of cells located in the lateral superior aspect of the junction between the right atrium and superior vena cava, within the superior aspect of the thick ridge of muscle known as the crista terminalis where the posterior smooth atrial wall derived from the sinus venosus meets the trabeculated anterior portion of the right atrium. Patients with sinus tachycardia will often seek medical attention with the uncomfortable awareness of their heartbeat as their chief complaint. Often, an arrhythmia is suspected because of the similar constellation of symptoms that accompanies supraventricular and ventricular tachycardia or atrial and ventricular ectopy. However, a careful review of the 12-lead electrocardiogram (ECG) reveals a characteristic P wave originating from the superior and lateral aspect of the right atrium with a positive deflection in leads I, II, and III and a biphasic morphology in lead V_1. Sinus P waves are characterized by a frontal plane axis directed inferiorly and leftward, with positive P waves in leads II, III, and aVF; a negative P wave in aVR; and an initially positive biphasic P wave in V_1. Normal sinus rhythm has a range of rates between 60 and 100 beats/min **(Fig. 254-1)**.

SINUS ARRHYTHMIA

Sinus arrhythmia is a common finding that is usually asymptomatic and related to normal physiology in healthy individuals. The rhythm is defined as arising from a sinus node origin but with irregularity between P-P intervals of >120 ms. When there is an irregularity in the heart rhythm and there are different P-wave morphologies observed, then this arrhythmia is most likely due to premature atrial contractions (PACs) and not sinus arrhythmia. Sinus arrhythmia usually occurs at rest and is often eliminated with higher rates observed with exertion due to removal of vagal tone. The three types of sinus arrhythmias observed are respirophasic, ventriculophasic, and nonphasic. Respirophasic sinus arrhythmia occurs when vagal tone is inhibited reflexively during inspiration and is restored with expiration. A similar phenomenon is seen with breath-holding leading to exaggerated pauses most often seen with obstructive sleep apnea. Ventriculophasic sinus arrhythmia is most often observed with heart block or after a premature ventricular contraction (PVC). The P-P interval is shortened when there is an interposed ventricular complex seen in these conditions possibly related to the triggered baroreceptor reflex from the subsequent beat after a longer ventricular filling time and increased stroke volume. Nonphasic sinus arrhythmia refers to variations in sinus P-P intervals unrelated to the cardiac or respiratory cycle. Regardless of the mechanism, asymptomatic sinus arrhythmia does not warrant further cardiac evaluation and is not considered pathogenic.

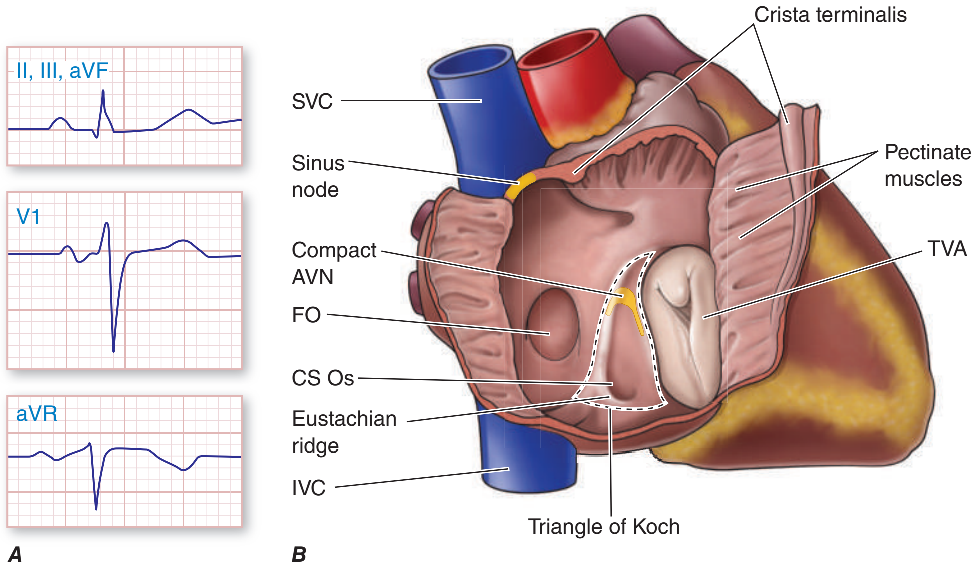

FIGURE 254-1 Right atrial anatomy pertinent to normal sinus rhythm and supraventricular tachycardia. ***A.*** Typical P-wave morphology during normal sinus rhythm based on standard 12-lead electrocardiogram. There is a positive P wave in leads II, III, and aVF and a biphasic, initially positive P wave in aVR. ***B.*** Right atrial anatomy seen from a right lateral perspective with lateral wall opened to view the septum. AVN, atrioventricular node; CS Os, coronary sinus ostium; FO, fossa ovalis; IVC, inferior vena cava; TVA, tricuspid valve annulus.

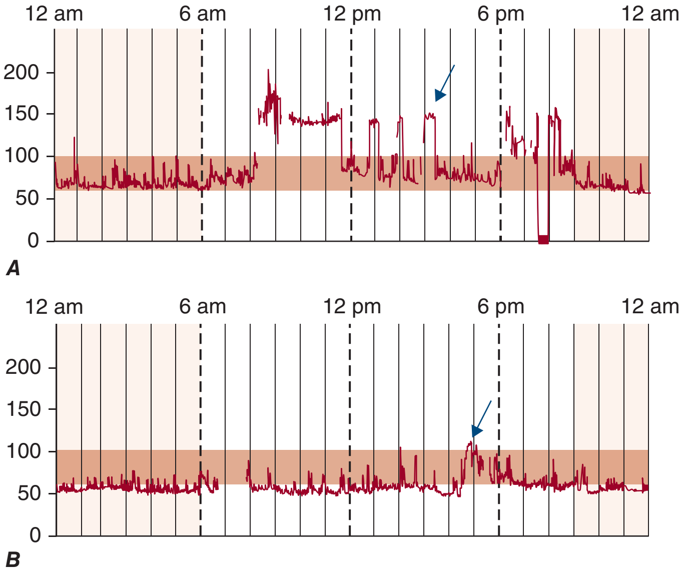

FIGURE 254-2 Outpatient telemetry monitor in a patient with intermittent atrial tachycardia (***A***) and normal physiologic sinus tachycardia (***B***).

PHYSIOLOGIC SINUS TACHYCARDIA

Sinus tachycardia (>100 beats/min) typically occurs in response to sympathetic stimulation and/or vagal withdrawal, whereby the rate of spontaneous depolarization of the sinus node increases and the focus of earliest activation within the node typically shifts more leftward and closer to the superior septal aspect of the crista terminalis, thus producing taller P waves in the inferior limb leads when compared to normal sinus rhythm. Sinus bradycardia is defined as rates <60 beats/min; however, bradycardia can be normal during sleep and in fit individuals.

Sinus tachycardia is considered physiologic when it is an appropriate response to exercise, stress, or illness. Sinus tachycardia can be difficult to distinguish from focal atrial tachycardia (see below) that originates near the sinus node. A causative factor (e.g., exertion) and a gradual rate increase favor a diagnosis of sinus tachycardia, whereas abrupt tachycardia onset and offset favor atrial tachycardia (Fig. 254-2).

The distinction can be difficult and occasionally requires extended ECG monitoring or invasive electrophysiology study. Treatment for physiologic sinus tachycardia is aimed at the underlying condition, but frequently, no therapy is necessary. Consideration to abnormal thyroid conditions and anemia should be given in patients with sinus tachycardia as these represent reversible causes. In addition, structural and functional cardiovascular abnormalities can present as sinus tachycardia, especially pulmonary embolism, and thus must be ruled out in the appropriate clinical scenario before considering sinus tachycardia as nonphysiologic. Finally, as sinus rate varies widely between individuals, a relatively elevated sinus rate (whether at rest or during exercise) without underlying cause, particularly without symptoms, typically does not warrant treatment (Table 254-1).

TABLE 254-1 Common Causes of Sinus Tachycardia

Physiologic Causes
Emotion, physical exercise, sexual intercourse, pain, pregnancy
Pathologic Causes
Anxiety, panic attack, anemia, fever, dehydration, infection, malignancies, hyperthyroidism, hypoglycemia, pheochromocytoma, Cushing's disease, diabetes mellitus with evidence of autonomic dysfunction, pulmonary embolus, myocardial infarction, pericarditis, valve disease, decompensated heart failure, shock, alcohol withdrawal
Drugs
Epinephrine, norepinephrine, dopamine, dobutamine, atropine, β_2-adrenergic receptor agonists (salbutamol), methylxanthines, doxorubicin, daunorubicin, beta blocker withdrawal, caffeine, alcohol
Illicit Drugs
Amphetamines, cocaine, lysergic acid diethylamide, psilocybin, ecstasy, cocaine

NONPHYSIOLOGIC SINUS TACHYCARDIA

Inappropriate sinus tachycardia is an uncommon condition in which the sinus rate increases spontaneously at rest or out of proportion to physiologic stress or exertion and is within a spectrum of ill-defined conditions associated with autonomic dysregulation. The underlying mechanism remains elusive, but it may be related to imbalance between sympathetic and parasympathetic inputs to the sinus node, altered membrane automaticity of sinus node cells, or a combination of both. Affected individuals are often women in the third or fourth decade of life. Fatigue, dizziness, and even syncope may accompany palpitations, which can be disabling. Additional symptoms of chest pain, headaches, and gastrointestinal upset are common. Inappropriate sinus tachycardia must be distinguished from appropriate sinus tachycardia and from focal atrial tachycardia arising from a region near the sinus node. The distinction between physiologic sinus tachycardia due to an anxiety disorder and inappropriate sinus tachycardia can be difficult. Therapy is often ineffective or poorly tolerated. Careful titration of beta blockers may reduce symptoms. Clonidine and serotonin reuptake inhibitors have also been used. Ivabradine, a drug that blocks the I_f current that causes spontaneous sinus node depolarization, is approved in the United States for use in heart failure but has also been effective in the treatment of inappropriate sinus tachycardia. Catheter ablation of the sinus node to modify and thereby decrease the sinus rate has been performed, but long-term control of symptoms is usually poor and can result in a permanent pacemaker requirement due to resultant symptomatic sinus bradycardia or arrest, or chronotropic incompetence (Fig. 254-3).

Postural orthostatic tachycardia syndrome (POTS) is characterized by symptomatic sinus tachycardia that occurs with postural change from a supine position to standing. The sinus rate increases by 30 beats/min or to >120 beats/min within 10 min of standing and in the absence of hypotension. Symptoms are often similar to those in patients with inappropriate sinus tachycardia. POTS is sometimes due to autonomic dysfunction following a viral illness and may resolve spontaneously over 3–12 months. Prolonged postviral symptoms after COVID-19 infection, sometimes referred to as "long COVID," have been ascribed to autonomic dysfunction and a POTS-like presentation. Volume expansion with salt supplementation, oral fludrocortisone, compression stockings, and the α-agonist midodrine, often in combination, can be helpful. Exercise training has also been shown to improve symptoms

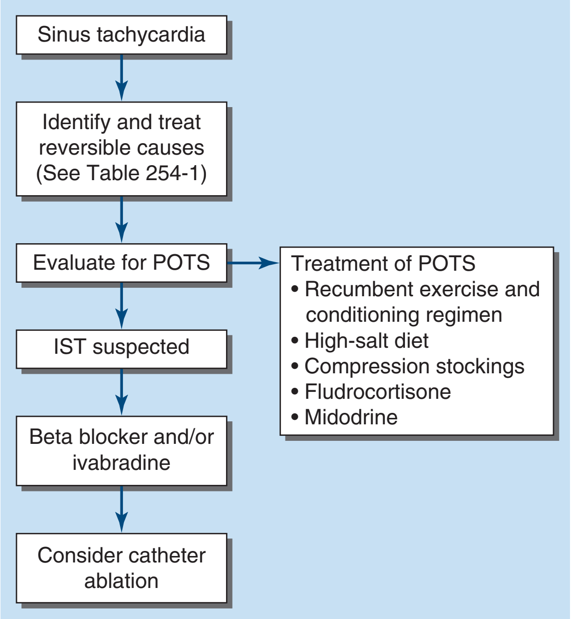
Sinus tachycardia
Identify and treat reversible causes (See Table 254-1)
Evaluate for POTS
IST suspected
Beta blocker and/or ivabradine
Consider catheter ablation
Treatment of POTS
• Recumbent exercise and conditioning regimen
• High-salt diet
• Compression stockings
• Fludrocortisone
• Midodrine

FIGURE 254-3 Evaluation and treatment of sinus tachycardia. For the patient who presents with sinus tachycardia, reversible causes of appropriate sinus tachycardia must be excluded and treated as indicated. Otherwise, evaluation for a spectrum of syndromes resulting in inappropriate sinus tachycardia should be undertaken. Potential directed therapies are shown. IST, inappropriate sinus tachycardia; POTS, postural orthostatic tachycardia syndrome.

and should be a part of a treatment strategy to reduce symptoms. While it is sometimes difficult to differentiate inappropriate sinus tachycardia from POTS, recognition of these distinct clinical syndromes is critical for treatment. Sinus node modification will be ineffective for the treatment of POTS and will possibly exacerbate the nontachycardia components of the condition. Likewise, treatment strategies aimed at increasing blood pressure will not be appropriate for inappropriate sinus tachycardia.

■ FURTHER READING

Mayuga KA et al: Sinus tachycardia: A multidisciplinary expert focused review. Circ Arrhythm Electrophysiol 15:e007960, 2022.

Vernino S et al: Postural orthostatic tachycardia syndrome (POTS): State of the science and clinical care from a 2019 National Institutes of Health Expert Consensus Meeting-Part 1. Auton Neurosci 235:102828, 2021.

255

Focal Atrial Tachycardia

William H. Sauer, Paul C. Zei

The underlying mechanisms of focal atrial tachycardia (AT) include abnormal automaticity, triggered automaticity, or a small reentry circuit in diseased atrial tissue. The term *focal* is used to differentiate this form of AT from typical and atypical atrial flutter but does not define a mechanism of the arrhythmia. ATs can originate from most regions of the atria, including atrial tissue extending into a pulmonary vein, the coronary sinus, or vena cava. It can be sustained, nonsustained, paroxysmal, or incessant. Focal AT accounts for ~10% of paroxysmal supraventricular tachycardia (PSVTs) in patients referred for catheter ablation. Nonsustained focal AT is commonly observed on ambulatory electrocardiogram (ECG) recordings, and the prevalence increases with age. Asymptomatic nonsustained ATs are often labeled as "SVT" on monitor reports, prompting patients to seek advice on catheter ablation. However, treatment is not recommended for asymptomatic nonsustained AT identified on ECG monitoring. Frequent atrial ectopy and nonsustained AT are often precursors to more significant arrhythmias such as atrial fibrillation and atrial flutter. Occasionally, nonsustained, frequent atrial ectopy or short bursts of AT may be symptomatic and require therapy similar to that required for focal AT **(Fig. 255-1)**.

AT can occur in the absence of structural heart disease or may be associated with any condition that causes atrial fibrosis, including atrial tissue targeted with prior catheter ablation. Areas of fibrosis can act as a nidus for abnormal automaticity from injured but partially living cells or microreentry in zones of slow conduction within and on the border of fibrotic areas. Sympathetic stimulation is a promoting factor, and the emergence of AT can be a sign of underlying illness or drug toxicity. In particular, AT with atrioventricular (AV) block is characteristic of digitalis toxicity. Symptoms from AT are highly variable but similar to other supraventricular tachycardias (SVTs), and incessant AT can cause tachycardia-induced cardiomyopathy.

AT typically presents with 1:1 AV conduction or with AV block in a Wenckebach or fixed (e.g., 2:1 or 3:1) pattern. Because it is not dependent on AV nodal conduction, AT will not terminate with AV block, and the atrial rate will not be affected, which distinguishes AT from most AV nodal–dependent SVTs, such as AV nodal reentry and AV reentry using an accessory pathway (see below). A so-called warm-up phase when the atrial activation rate increases after initiation or a cool-down phase when the rate slows prior to termination also favors AT rather than AV nodal–dependent SVT, as this is a common observation with enhanced automaticity. P waves are often discrete, with an intervening isoelectric segment, in contrast to atrial flutter and macroreentrant AT because atrial activation from a focal source occurs through a small portion of the tachycardia cycle **(Fig. 255-2)**.

When 1:1 conduction to the ventricles is present, the arrhythmia can resemble sinus tachycardia typically with a P-R interval shorter than the R-P interval, *particularly when sympathetic tone results in rapid AV nodal conduction*. It can be distinguished from sinus tachycardia by the P-wave morphology, which usually differs from sinus P waves depending on the location of the focus. Focal AT tends to originate in areas of complex atrial anatomy, such as the crista terminalis, valve annuli, atrial septum, and atrial muscle extending along cardiac thoracic veins (superior vena cava, coronary sinus, and pulmonary veins), and the location can often be approximated by the P-wave morphology. AT from the atrial septum will frequently have a narrower P-wave duration than sinus rhythm. AT from the left atrium will usually have a monophasic, positive P wave in lead V_1 and negative P waves in I and aVL, indicating an activation wavefront away from the left atrial free wall. AT that originates from superior atrial locations, such as the superior vena cava or superior pulmonary veins, will be positive in the inferior limb leads II, III, and aVF, whereas AT from a more inferior location, such as the ostium of the coronary sinus, will inscribe negative P waves in these same leads. When the focus is in the superior aspect of the crista terminalis, close to the sinus node, however, the P wave will resemble that of sinus tachycardia. Abrupt onset and offset then favor AT rather than sinus tachycardia. Depending on the atrial rate, the P wave may fall on top of the T wave, or during 2:1 conduction,

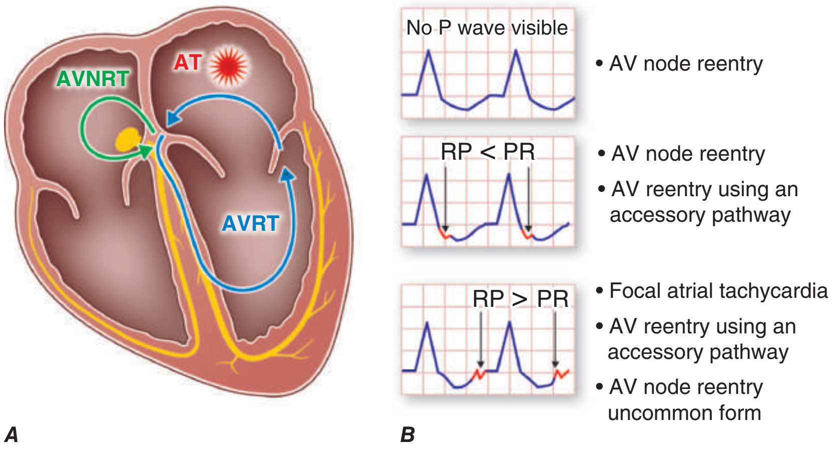
AT
AVNRT
AVRT
No P wave visible
• AV node reentry
RP < PR
• AV node reentry
• AV reentry using an accessory pathway
RP > PR
• Focal atrial tachycardia
• AV reentry using an accessory pathway
• AV node reentry uncommon form
A
B

FIGURE 255-1 Common mechanisms underlying paroxysmal supraventricular tachycardia along with typical R-P relationships. ***A.*** Schematic showing a four-chamber view of the heart with atrioventricular (AV) node and specialized conduction tissue (His-Purkinje) in *yellow.* Atrial tachycardia (AT; *red circuit*) is confined completely to atrial tissue. Atrioventricular nodal reentry tachycardia (AVNRT; *green circuit*) uses AV nodal and perinodal atrial tissue. Atrioventricular reentry tachycardia (AVRT; *blue circuit*) uses atrial and ventricular tissue, accessory pathway between the ventricle and atrium, AV node, and His-Purkinje tissue as part of the reentry circuit. ***B.*** Typical relation of the P wave to QRS, commonly described as the R-P to P-R relationship, for the different tachycardia mechanisms.

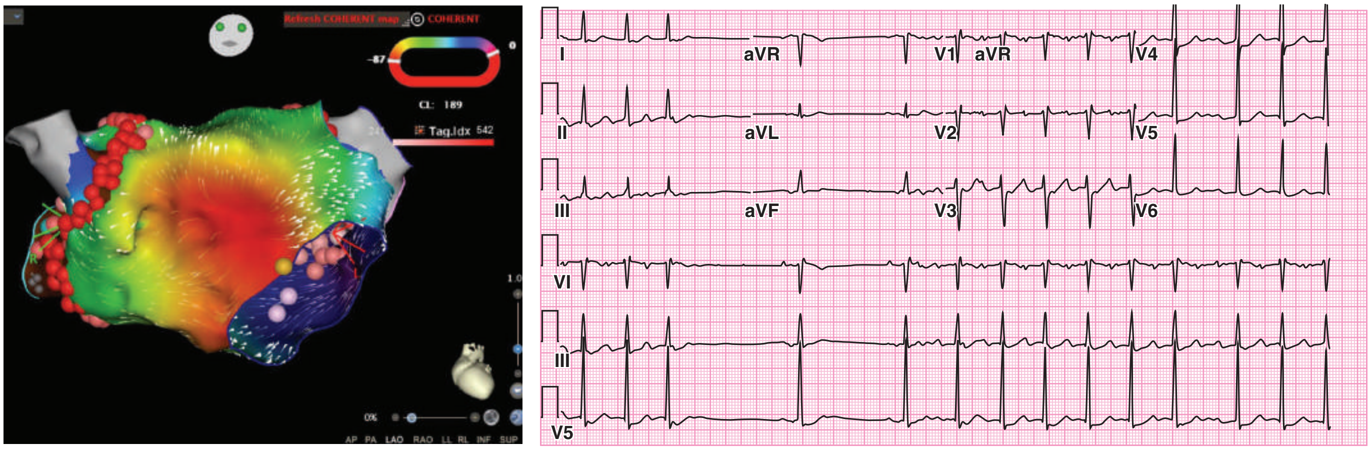

FIGURE 255-2 Focal atrial tachycardia. In the right panel, a surface 12-lead electrocardiogram shows focal intermittent atrial tachycardia. Note the discrete P waves, with isoelectric segments between, as well as the sinus rhythm. The left panel shows an electroanatomic map of the same focal atrial tachycardia originating from the anterior interatrial septum, as viewed in an anterior-posterior (AP) view of the left atrium obtained during electrophysiology study and ablation. The colors represent the timing of local electrical activation during each tachycardia atrial activation, showing a focal early (*red*) site. Additional markers of *white "flecks"* represent conduction direction, demonstrating activation of the atrium dispersing from this focal site. Of note, the *pink* and *red dots* represent ablation lesions, in this case, for pulmonary vein isolation. *(Adapted from J Brugada et al: 2019 ESC Guidelines for the management of patients with supraventricular tachycardia. The Task Force for the management of patients with supraventricular tachycardia of the European Society of Cardiology (ESC). Eur Heart J 41:655, 2020.)*

every other P wave may fall coincident with the QRS. Maneuvers that increase AV block, such as carotid sinus massage, Valsalva maneuver, or administration of AV nodal–blocking agents, such as adenosine, are useful to create AV block that will expose the P wave.

Acute management of sudden-onset, sustained AT is the same as for other forms of PSVT, but the response to pharmacologic therapy is variable, likely depending on the mechanism **(Fig. 255-3)**.

For AT due to reentry, administration of adenosine or vagal maneuvers may transiently increase AV block without terminating tachycardia. Some ATs terminate with a sufficient dose of adenosine, consistent with triggered activity as the mechanism. Cardioversion can be effective in some but fails in others because of immediate recurrence, suggesting automaticity as the mechanism in these cases. Beta blockers and calcium channel blockers may slow the ventricular rate by increasing AV block, which can improve tolerance of the arrhythmias, but large doses are sometimes required. Potential precipitating factors and intercurrent illness should be sought and corrected. Underlying heart disease should be considered and excluded.

For patients with recurrent episodes, beta blockers, calcium channel blockers such as diltiazem or verapamil, and antiarrhythmic drugs such as flecainide, propafenone, disopyramide, sotalol, and amiodarone can be effective, but potential toxicities and adverse effects often warrant avoidance of long-term use.

Catheter ablation targeting the AT focus is effective in >80% of patients and is recommended for recurrent symptomatic AT when drugs fail or are not desired or for incessant AT causing tachycardia-induced cardiomyopathy. Although AT is often a precursor to atrial fibrillation or atrial flutter, the associated risk for stroke and, hence, indications for long-term anticoagulation are unclear but not considered equivalent.

FURTHER READING

Brugada J et al: 2019 ESC Guidelines for the management of patients with supraventricular tachycardia. The Task Force for the Management of Patients with Supraventricular Tachycardia of the European Society of Cardiology (ESC) developed in collaboration with the Association for European Paediatric and Congenital Cardiology (AEPC). Eur Heart J 41:655, 2020.

Callans DJ: *Josephson's Clinical Cardiac Electrophysiology: Techniques and Interpretations*, 7th ed. Philadelphia, Wolters Kluwer, 2024.

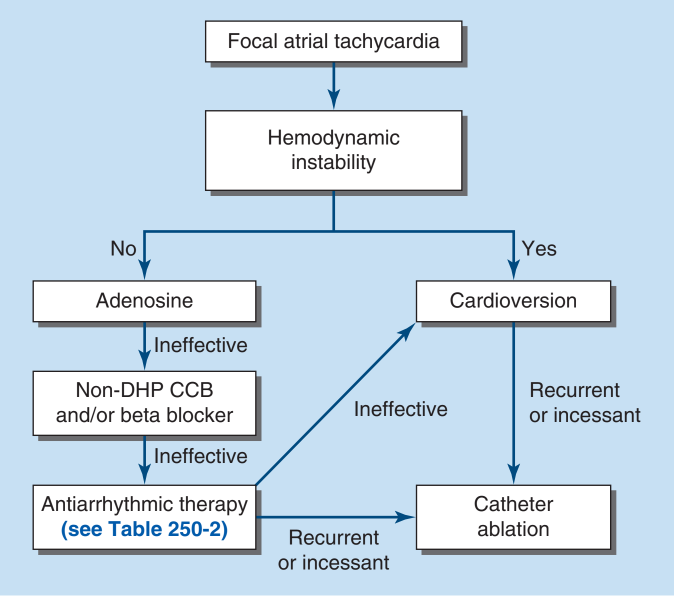

FIGURE 255-3 Clinical approach and treatment algorithm for management of focal atrial tachycardia. CCB, calcium channel blocker; DHP, dihydropyridine. *(Adapted from J Brugada et al: 2019 ESC Guidelines for the management of patients with supraventricular tachycardia The Task Force for the management of patients with supraventricular tachycardia of the European Society of Cardiology (ESC) [published correction appears in Eur Heart J 41:4258, 2020]. Eur Heart J 41:655, 2020.)*

256

Paroxysmal Supraventricular Tachycardias

William H. Sauer, Paul C. Zei

In this chapter, sustained supraventricular tachycardias (SVTs) dependent on the atrioventricular (AV) node are discussed. These include AV nodal reentry tachycardia (AVNRT), junctional tachycardia, AV reciprocating tachycardia (AVRT) utilizing an accessory pathway, and a group of additional various SVTs that involve an accessory

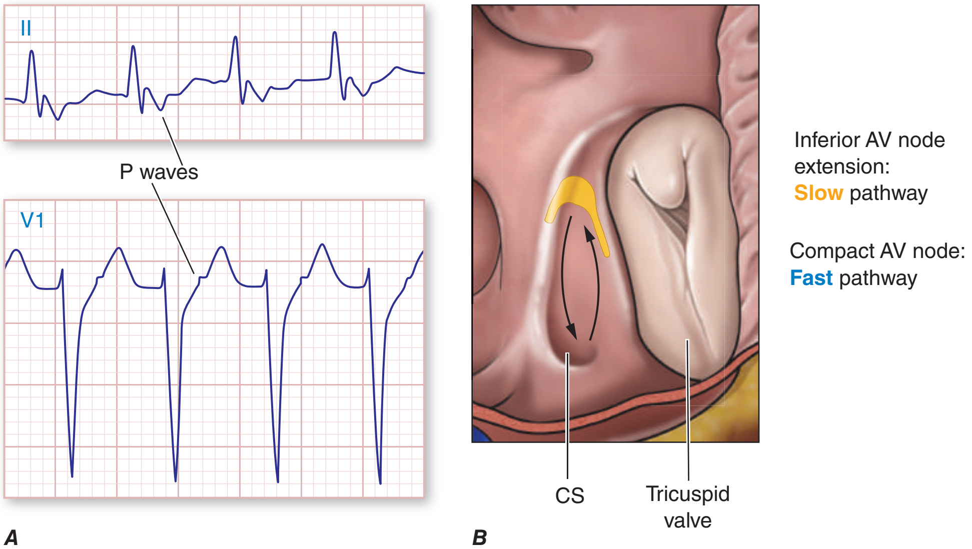
II
P waves
V1
A
Inferior AV node extension: **Slow** pathway
Compact AV node: **Fast** pathway
CS
Tricuspid valve
B

FIGURE 256-1 Atrioventricular (AV) node reentry. ***A.*** Leads II and V_1 are shown. P waves are visible at the end of the QRS complex and are negative in lead II and may give the impression of S waves in the inferior limb leads II, III, and aVF and an R′ in lead V_1. ***B.*** Stylized version of the AV nodal reentry circuit within the triangle of Koch (see **Fig. 254-1**) that involves AV node and its extensions along with perinodal atrial tissue. CS, coronary sinus.

pathway, termed *preexcited tachycardias*. The term *SVT* encompasses a broad group of tachyarrhythmias based on anatomic origin and technically includes sinus tachycardia, atrial tachycardia (AT), atrial flutter, and atrial fibrillation; however, for the purposes of describing an organized approach to diagnosis and treatment of SVT, a separate discussion for these non-AV nodal–dependent SVTs are discussed elsewhere.

ATRIOVENTRICULAR NODAL REENTRY TACHYCARDIA

AVNRT is the most common form of paroxysmal supraventricular tachycardia (PSVT), representing ~60% of cases referred for catheter ablation. It most commonly manifests in the second to fourth decades of life, more often in women. It is usually well tolerated, but rapid tachycardia, particularly in the elderly, may cause angina, pulmonary edema, hypotension, or syncope. It is not usually associated with structural heart disease. In patients without associated heart disease, AVNRT is not a life-threatening arrhythmia; however, it may cause significant symptoms.

The mechanism is reentry involving the AV node and the perinodal atrium, made possible by the existence of multiple pathways for conduction from the atrium into the AV node that are capable of conduction in two directions **(Fig. 256-1)**.

Most forms of AVNRT utilize a slowly conducting AV nodal pathway (right inferior extension) that extends from the compact AV node near the His bundle, inferiorly along the tricuspid valve annulus to the floor of the coronary sinus. The reentry wavefront propagates up this slowly conducting pathway to the compact AV node and then exits from the fast pathway at the top of the AV node. The path back to the slow pathway probably involves the left atrial septum, which has connections to the coronary sinus musculature. More unusual forms of AVNRT utilize a left inferior extension that connects to the compact AV node through the roof of the coronary sinus or, in extremely rare cases, directly from the mitral valve annulus avoiding the coronary sinus musculature altogether. In typical forms, the conduction time from the compact AV node region to the atrium is similar to that from the compact node to the His bundle and ventricles, such that atrial activation occurs at about the same time as ventricular activation. The P wave is therefore inscribed during, slightly before, or slightly after the QRS and can be difficult to discern. Often the P wave is seen at the end of the QRS complex as a pseudo-r′ in lead V_1 and pseudo-S waves in leads II, III, and aVF **(Fig. 256-2)**.

More unusual forms of AVNRT have P waves falling later, anywhere between QRS complexes, in which case, an inverted P wave is seen in the inferior limb leads with the inverted P wave seen in the subsequent T wave. The rate can vary with sympathetic tone through its effect on the conduction time of AV nodal tissues. Simultaneous atrial and ventricular contraction results in atrial contraction against a closed tricuspid valve, producing a cannon A wave visible in the jugular venous pulse often perceived as a fluttering sensation in the neck. Elevated venous pressures may also lead to release of natriuretic peptides that cause post-tachycardia diuresis. In contrast to ATs, maneuvers or medications that produce AV nodal block terminate the arrhythmia. Acute treatment is the same as for other forms of PSVT (discussed below). Whether ongoing therapy is warranted depends on the severity of symptoms and frequency of episodes. Reassurance and instruction as to how to perform the Valsalva maneuver or other vagal

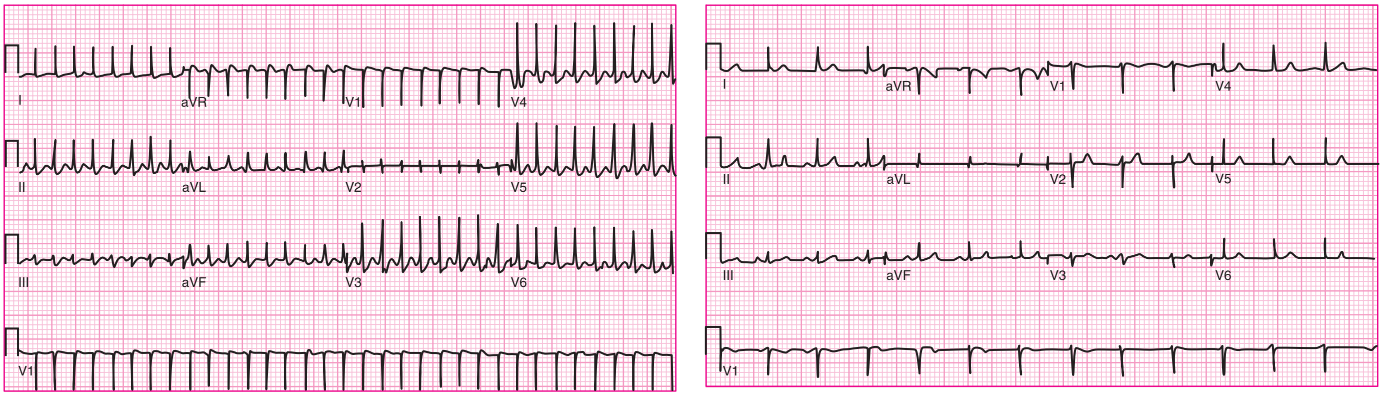

FIGURE 256-2 Atrioventricular nodal reentry tachycardia with retrograde P waves before and after adenosine termination.

nerve–stimulating maneuvers to terminate episodes are sufficient for many patients. Administration of an oral beta blocker, verapamil, or diltiazem at the onset of an episode can be used to facilitate termination. Chronic therapy with these medications or flecainide is an option if prophylactic therapy is needed. Catheter ablation of the slow AV nodal pathway is recommended for patients with recurrent or severe episodes or when drug therapy is ineffective, not tolerated, or not desired by the patient. Catheter ablation is curative in >95% of patients. The major risk is AV block requiring permanent pacemaker implantation, which occurs in <1% of patients.

JUNCTIONAL TACHYCARDIA

Junctional ectopic tachycardia (JET) is due to automaticity within the AV node. It is rare in adults and more frequently encountered as an incessant tachycardia in children, often in the perioperative period of surgery for congenital heart disease. It presents as a narrow QRS tachycardia, often with ventriculoatrial (VA) block, such that AV dissociation is present. JET can occur as a manifestation of increased adrenergic tone and may be seen after administration of isoproterenol, particularly after catheter ablation in the perinodal region. It may also occur for a short period of time after ablation for AVNRT. *Accelerated junctional rhythm* is a junctional automatic rhythm between 50 and 100 beats/min. Initiation may occur with gradual acceleration in rate, suggesting an automatic focus, or after a premature ventricular contraction, suggesting a focus of triggered automaticity. VA conduction is usually present, with P-wave morphology and timing such that it resembles AVNRT at a slow rate. It can be related to increased sympathetic tone and may produce palpitations. It usually does not require specific therapy.

ACCESSORY PATHWAYS AND THE WOLFF-PARKINSON-WHITE SYNDROME

Accessory pathways (APs) occur in 1 in 1500–2000 people and are associated with a variety of arrhythmias including narrow-complex PSVT, wide-complex tachycardias, and, rarely, sudden death. Most patients have structurally normal hearts, but APs are associated with Ebstein's anomaly of the tricuspid valve and forms of hypertrophic cardiomyopathy including *PRKAG2* mutations, Danon's disease, and Fabry's disease **(Fig. 256-3)**.

APs are abnormal connections that allow conduction between the atrium and ventricles across the AV ring. They are present from birth and are due to failure of complete partitioning of atrium and ventricle by the fibrous AV rings. They occur across either an AV valve annulus or the septum, most frequently between the left atrium and free wall of the left ventricle, followed by posteroseptal, right free wall, and anteroseptal locations. If the impulse from the sinus node conducts through the AP to the ventricle (antegrade) before the impulse conducts through the AV node and His bundle, then the ventricles are preexcited during sinus rhythm, and the electrocardiogram (ECG) shows a short P-R interval (<0.12 s), slurred initial portion of the QRS (delta wave), and prolonged QRS duration produced by slow conduction through direct activation of ventricular myocardium over the AP. The morphology of the QRS and delta wave is determined by the AP location and associated site of earliest ventricular activation, and the degree of fusion between the excitation wavefronts from conduction over the AV node and conduction over the AP **(Fig. 256-4)**.

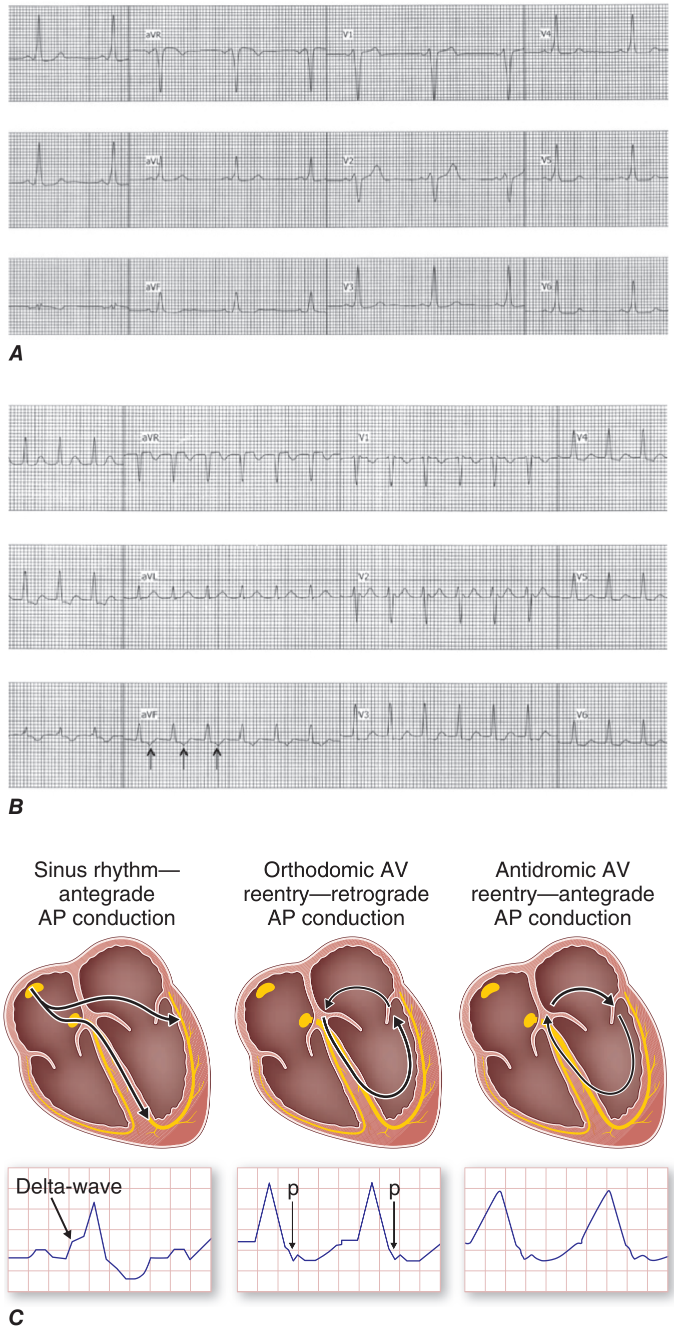
aVR
V1
V4
aVL
V2
V5
aVF
V3
V6
A
aVR
V1
V4
aVL
V2
V5
aVF
V3
V6
B
Sinus rhythm—antegrade AP conduction
Orthodomic AV reentry—retrograde AP conduction
Antidromic AV reentry—antegrade AP conduction
Delta-wave
p
p
C

FIGURE 256-3 Wolff-Parkinson-White (WPW) syndrome. ***A.*** A 12-lead electrocardiogram in sinus rhythm (SR) of a patient with WPW demonstrating short P-R interval, delta waves, and widened QRS complex. This patient had an anteroseptal location of the accessory pathway (AP). ***B.*** Orthodromic atrioventricular (AV) reentry in a patient with WPW syndrome using a posteroseptal AP. Note the P waves in the ST segment (*arrows*) seen in lead III and normal appearance of QRS complex. ***C.*** Three most common rhythms associated with WPW syndrome: sinus rhythm demonstrating antegrade conduction over the AP and AV node; orthodromic AV reentry tachycardia (AVRT) using retrograde conduction over the AP and antegrade conduction over the AV node; and antidromic AVRT using retrograde conduction over the AV node and antegrade conduction over the AP.

Right-sided pathways preexcite the right ventricle, producing a left bundle branch block–like configuration in lead V_1, and often create marked preexcitation because of their relatively close proximity of the AP to the sinus node (Fig. 256-4). Left-sided pathways preexcite the left ventricle and may produce a right bundle branch–like configuration in lead V_1 and a negative delta wave in aVL, indicating initial depolarization of the lateral portion of the left ventricle that can mimic Q waves of lateral wall infarction (Fig. 256-4). Because of the relatively large distance between the sinus node and left free wall APs, preexcitation may be minimal or absent on 12-lead ECG. Preexcitation due to an AP at the diaphragmatic surface of the heart, typically in the paraseptal region, produces delta waves that are negative in leads III and aVF, mimicking the Q waves of inferior wall infarction (Fig. 256-4). Preexcitation can be intermittent and disappear during exercise as conduction over the AV node accelerates and may take over ventricular activation completely.

Wolff-Parkinson-White (WPW) syndrome is defined as a preexcited QRS during sinus rhythm and episodes of PSVT. There are a number of variations of APs that may not cause preexcitation and/or arrhythmias. Concealed APs allow only retrograde conduction, from ventricle to atrium, so no preexcitation is present during sinus rhythm, but SVT

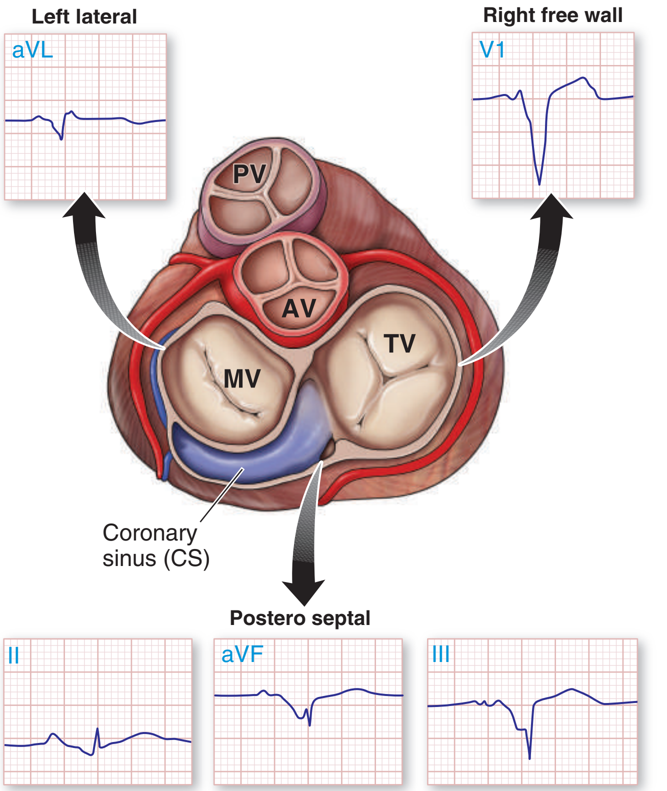

FIGURE 256-4 Potential locations for accessory pathways in patients with Wolff-Parkinson-White syndrome and typical QRS appearance of delta waves that can mimic underlying structural heart disease such as myocardial infraction of bundle branch block. AV, aortic valve; MV, mitral valve; PV, pulmonary valve; TV, tricuspid valve.

can occur. Other unusual forms of APs exist. Fasciculoventricular connections between the His bundle and ventricular septum produce preexcitation but do not cause arrhythmia, probably because the circuit is too short to promote reentry. Atriofascicular pathways, also known as Mahaim fibers, probably represent a duplicate AV node and His-Purkinje system that connect the right atrium to fascicles of the right bundle branch and produce a wide-complex tachycardia having a left bundle branch block configuration.

ATRIOVENTRICULAR RECIPROCATING TACHYCARDIA

The most common tachycardia caused by an AP is the PSVT designated *orthodromic AV reciprocating tachycardia*. The circulating reentry wavefront propagates from the atrium anterogradely over the AV node and His-Purkinje system to the ventricles and then reenters the atria via retrograde conduction over the AP. The QRS is narrow or may have typical right or left bundle branch block, but without preexcitation during tachycardia. Because excitation through the AV node and AP are necessary, AV or VA block results in tachycardia termination. During sinus rhythm, preexcitation is seen if the pathway also allows anterograde conduction. Most commonly, during tachycardia, the R-P interval is shorter than the P-R interval and can resemble AVNRT. Unlike typical AVNRT, P waves always follow the QRS and are never simultaneous with a narrow QRS complex because the ventricles must be activated before the reentry wavefront reaches the AP and conducts back to the atrium. The morphology of the P wave during tachycardia is determined by the pathway location, but it can be difficult to assess because it is usually inscribed during the ST segment. The P wave in posteroseptal APs is negative in leads II, III, and aVF, similar to that in AV nodal reentry, but P-wave morphology will differ from AV nodal reentry for pathways in other locations. Occasionally, an AP conducts extremely slowly in the retrograde direction, resulting in tachycardia with a long R-P interval, similar to most ATs. These pathways are usually located in the septal region and have negative P waves in leads II, III, and aVF. Slow AP conduction facilitates reentry, often leading to nearly incessant tachycardia, known as *permanent junctional reciprocating tachycardia* (PJRT). Tachycardia-induced cardiomyopathy can occur. Without an invasive electrophysiology study, it may be difficult to distinguish this form of orthodromic AV reentry from atypical AV nodal reentry or AT.

PREEXCITED TACHYCARDIAS

Preexcited tachycardia occurs when the ventricles are activated by antegrade conduction over the AP. The most common mechanism is *antidromic AV reciprocating tachycardia* in which activation propagates from atrium to ventricle via the AP and then conducts retrogradely to the atria via the His-Purkinje system and the AV node (or rarely a second AP). The wide QRS complex is produced entirely via ventricular excitation over the AP because there is no contribution of ventricular activation over more rapidly conducting specialized His-Purkinje fibers. This tachycardia is often indistinguishable from monomorphic ventricular tachycardia. The presence of preexcitation in sinus rhythm suggests the diagnosis.

Preexcited tachycardia also occurs if an AP allows antegrade conduction to the ventricles during AT, atrial flutter, atrial fibrillation (AF), or AV nodal reentry, otherwise known as bystander AP conduction. AF and atrial flutter are potentially life-threatening if the AP allows very rapid repetitive conduction **(Fig. 256-5)**.

Approximately 25% of APs causing preexcitation allow minimum R-to-R intervals of <250 ms during AF and are associated with a higher risk of inducing ventricular fibrillation and sudden death. Preexcited AF presents as a wide-complex, very irregular rhythm. During AF, the ventricular rate is determined by the conduction properties of the AP and AV node. The QRS complex can appear quite bizarre and change on a beat-to-beat basis due to the variability in the degree of fusion from activation over the AV node and AP, or all beats may be due to conduction over the AP. Ventricular activation from the Purkinje system may depolarize the ventricular aspect of the AP and prevent atrial wavefront conduction over the AP. Slowing AV nodal conduction without slowing AP conduction can thereby facilitate AP conduction and dangerously accelerate the ventricular rate. Administration of AV nodal–blocking agents, including oral or intravenous verapamil, diltiazem, beta blockers, intravenous adenosine, and intravenous amiodarone, is contraindicated during preexcited AF. Rapid preexcited tachycardia should be treated with electrical cardioversion or intravenous procainamide or ibutilide, which may terminate the arrhythmia or slow the ventricular rate.

MANAGEMENT OF PATIENTS WITH ACCESSORY PATHWAYS

Acute management of orthodromic AV reentry is discussed below for PSVT. Patients with WPW syndrome may have wide-complex tachycardia due to antidromic AV reentry, orthodromic AV with bundle branch block, or a preexcited tachycardia, and treatment depends on the underlying rhythm. Initial patient evaluation should include assessment for aggravating factors, including intercurrent illness and factors that increase sympathetic tone. Examination should focus on excluding underlying heart disease. An echocardiogram is reasonable to exclude Ebstein's anomaly, forms of hypertrophic cardiomyopathy that can be associated with APs, or tachycardia-mediated cardiomyopathy.

Patients with preexcitation who have symptoms of arrhythmia are at risk for developing AF and sudden death if they have an AP that allows rapid antegrade conduction. The risk of cardiac arrest is in the range of 2 per 1000 patients in adults but is likely greater in children. An invasive electrophysiology study is recommended to assess whether the pathway can support dangerously rapid heart rates if AF were to occur, and it is usually combined with potentially curative catheter ablation. Catheter ablation is warranted for recurrent arrhythmias when drugs are ineffective, not tolerated, or not desired by the patient. Efficacy is in the range of 95% depending on the location of the AP. Serious complications occur in <3% of patients but can include AV

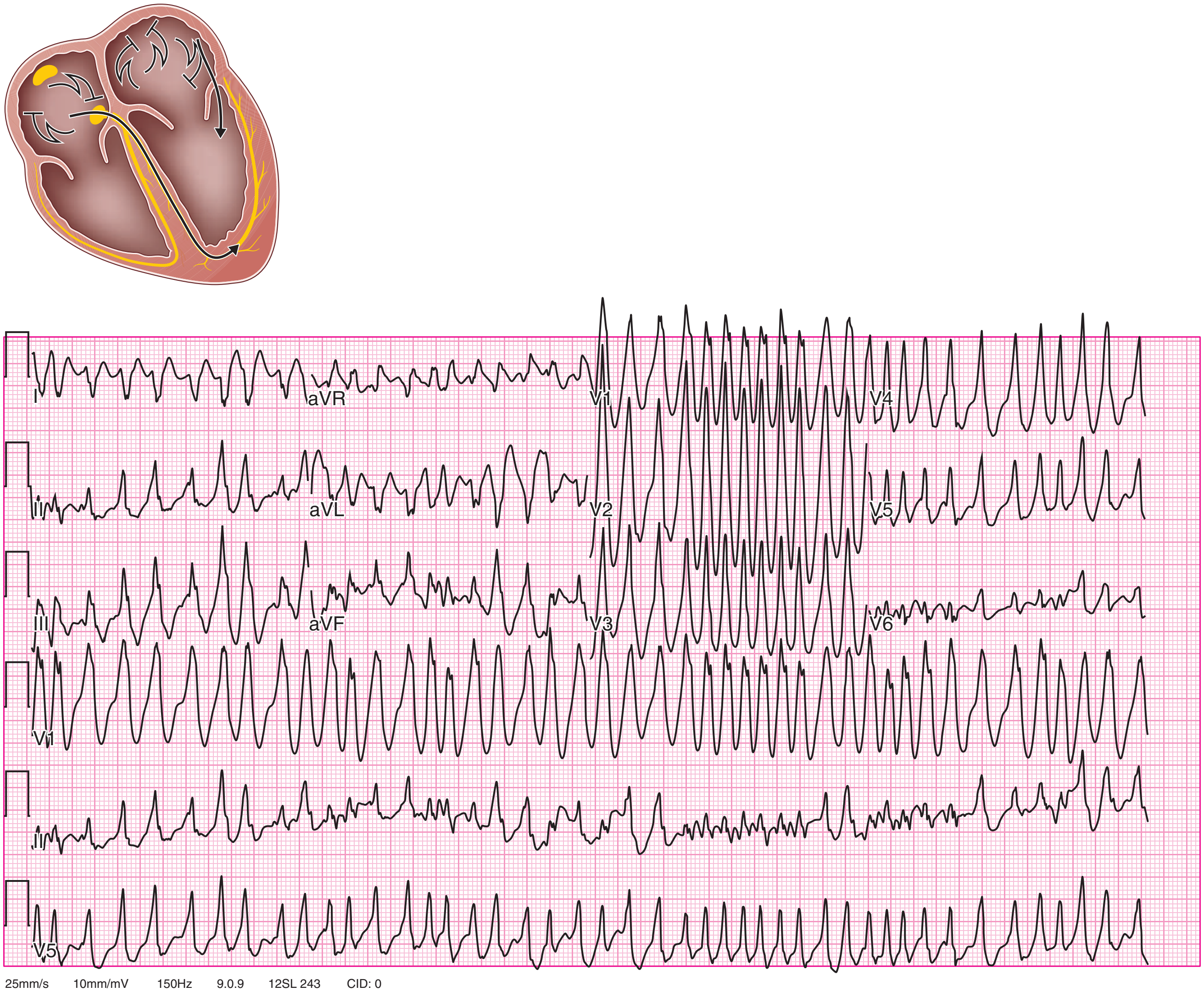

FIGURE 256-5 Preexcited atrial fibrillation (AF) due to conduction over a left free wall accessory pathway (AP). The electrocardiogram shows rapid irregular QRS complexes that represent fusion between conduction over the atrioventricular node and left free wall AP. Shortest R-R intervals between preexcited QRS complexes of <250 ms, as in this case, indicate a risk of sudden death with this arrhythmia.

block, cardiac tamponade, thromboembolism, coronary artery injury, and vascular access complications. The risk of AV block is higher when the AP is located near the AV node and/or His bundle, in the so-called anteroseptal or mid-septal locations. Procedure mortality is <1 in 1000 patients. Ambulatory monitoring or exercise testing is often used to gain reassurance that the AP is not high risk, evaluating for abrupt loss of conduction (preexcitation) at physiologic heart rates consistent with a low-risk pathway, but this is not completely reliable. Gradual loss of AP conduction with increased sympathetic tone does not reliably indicate low risk since this can occur as AV nodal conduction time shortens, and therefore, the possibility of rapid antegrade AP conduction is not excluded definitively.

For patients with concealed APs or known low-risk APs causing orthodromic AVRT, chronic therapy is guided by symptoms and frequency of events. Vagal maneuvers may terminate episodes, as may a dose of beta blocker, verapamil, or diltiazem taken at the onset of an episode. Chronic therapy with these agents or flecainide can reduce the frequency of episodes in some patients.

Adults who have preexcitation but no arrhythmia symptoms have a risk of sudden death estimated to be 1 per 1000 patient-years. Electrophysiology study is usually advised for people in occupations for which an arrhythmia occurrence would place them or others at risk, such as police, military, and pilots, or for individuals who desire evaluation for risk. Routine follow-up without therapy is reasonable in others. Children are at greater risk of sudden death, ~2 per 1000 patient-years.

TREATMENT

Paroxysmal Supraventricular Tachycardia

Acute management of narrow QRS PSVT is guided by the clinical presentation. Continuous ECG monitoring should be implemented, and a 12-lead ECG should always be obtained when possible, since this may be useful in determining the mechanism. In the presence of hypotension with unconsciousness or respiratory distress, QRS-synchronous direct current cardioversion is warranted, but this is rarely needed, because intravenous adenosine works promptly in most situations (see below). For stable individuals, initial therapy takes advantage of the fact that most PSVTs are dependent on AV nodal conduction (AV nodal reentry or orthodromic AV reentry) and, therefore, likely to respond to sympatholytic and vagotonic maneuvers and drugs. As these are administered, the ECG should be continuously recorded because the response can establish the diagnosis. AV block with only transient slowing of tachycardia may expose ongoing P waves, indicating AT or atrial flutter as the mechanism **(Fig. 256-6)**.

Carotid sinus massage is reasonable provided the risk of carotid vascular disease is low, as indicated by absence of carotid bruits and no prior history of stroke. A Valsalva maneuver should be attempted in cooperative individuals, and if effective, the patient can be taught to perform this maneuver as needed. If vagal maneuvers fail or

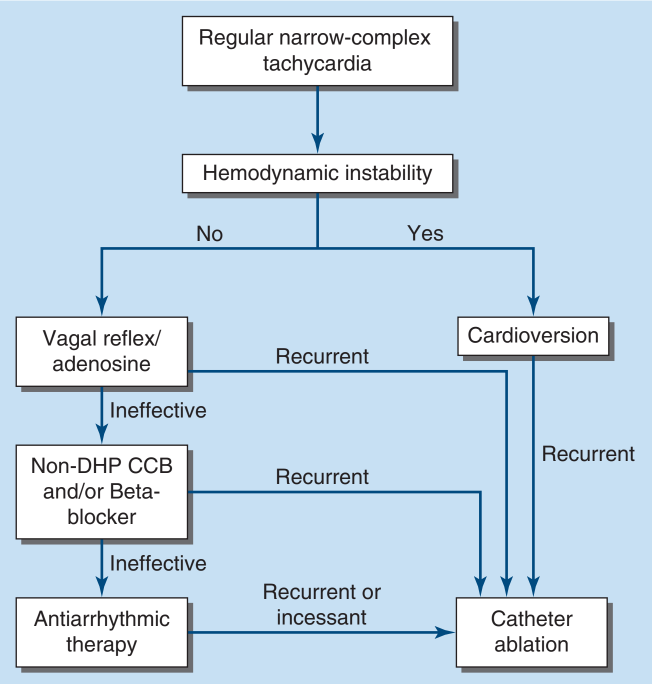

FIGURE 256-6 Treatment algorithm for patients presenting with hemodynamically stable paroxysmal supraventricular tachycardia. CCB, calcium channel blocker; DHP, dihydropyridine.

cannot be performed, intravenous adenosine will terminate the vast majority of PSVT episodes by transiently blocking conduction in the AV node. Adenosine may produce transient chest pain, dyspnea, and anxiety. It is contraindicated in patients with prior cardiac transplantation due to potential hypersensitivity due to surgical sympathetic denervation. It can theoretically aggravate bronchospasm. Adenosine precipitates AF, which is usually brief, in up to 15% of patients, so it should be used cautiously in patients with WPW syndrome in whom AF may produce hemodynamic instability. Intravenous beta blockers and calcium channel blockers (verapamil or diltiazem) are also effective but may cause hypotension before and after arrhythmia termination and have a longer duration of action. These agents can also be given orally and can be taken by the patient on an as-needed basis to slow ventricular rate and facilitate termination by Valsalva maneuver.

The differential diagnosis of wide-complex tachycardia includes ventricular tachycardia, PSVT with bundle branch block aberrancy, and preexcited tachycardia (see above). In general, these should be managed as ventricular tachycardia until proven otherwise. If the tachycardia is regular and the patient is stable, a trial of intravenous adenosine is reasonable. Very irregular wide-complex tachycardia is most likely preexcited AF or flutter (see above) and should be managed with cardioversion, intravenous procainamide, or ibutilide. If the diagnosis of PSVT with aberrancy is unequivocal, as may be the case in patients with prior episodes, treatment for PSVT with vagal maneuvers and adenosine is reasonable. In all cases, continuous ECG monitoring should be implemented, and emergency cardioversion and defibrillation should be available.

For management of PSVT outside of the acute setting, most patients with recurrent episodes are candidates for catheter ablation (see above).

FURTHER READING

Brugada J et al: 2019 ESC Guidelines for the management of patients with supraventricular tachycardia. The task force for the management of patients with supraventricular tachycardia of the European Society of Cardiology (ESC) Developed in collaboration with the Association for European Paediatric and Congenital Cardiology (AEPC). Eur Heart J 41:655, 2020.

Callans DJ: *Josephson's Clinical Cardiac Electrophysiology: Techniques and Interpretations*, 7th ed. Philadelphia, Wolters Kluwer, 2024.

Jalife J, Stevenson W (eds): *Zipes and Jalife's Cardiac Electrophysiology: From Cell to Bedside*, 8th ed. Philadelphia, Elsevier, 2022.

257 Common Atrial Flutter and Macroreentrant and Multifocal Atrial Tachycardias

William H. Sauer, Jorge E. Romero, Paul C. Zei

Macroreentrant atrial tachycardia is characterized by a large anatomic reentry circuit, often associated with areas of atrial scar tissue with slow conduction. *Common or typical right atrial flutter* is attributed to a wavefront of electrical propagation encircling the tricuspid valve annulus, bounded anteriorly by the annulus and posteriorly by the inferior vena cava and the Eustachian ridge. The wavefront passes between the inferior vena cava and the tricuspid valve annulus, known as the sub-Eustachian or cavotricuspid isthmus, where it is susceptible to interruption by catheter ablation. Thus, common atrial flutter is also known as *cavotricuspid isthmus-dependent atrial flutter*. This circuit most commonly revolves in a counterclockwise direction (as viewed looking toward the tricuspid annulus from the ventricular apex), which produces the characteristic negative sawtooth flutter waves in leads II, III, and aVF and positive P waves in lead V_1. When the direction is reversed, clockwise rotation produces the opposite P-wave vector in those leads and is called "reverse typical atrial flutter." The atrial rate is typically 240–300 beats/min but may be slower in the presence of atrial disease or antiarrhythmic drugs. It often conducts to the ventricles with 2:1 atrioventricular (AV) block, creating a regular tachycardia at 140–150 beats/min, with flutter waves that may be difficult to discern from the T wave. Maneuvers that increase AV nodal block will typically expose flutter waves, allowing diagnosis. AV nodal disease or AV nodal–blocking agents may render the conduction ratio between atrium and ventricle higher, resulting in more obvious flutter waves **(Fig. 257-1)**. However, controlling ventricular rates during atrial flutter with AV nodal–blocking agents is frequently challenging.

Common or typical right atrial flutter often occurs alongside atrial fibrillation and is frequently associated with atrial scarring from senescence or prior cardiac surgery. Nonetheless, the onset of typical atrial flutter has been proposed to be preceded by atrial fibrillation, indicating a potential transition from atrial fibrillation to atrial flutter due to a line of functional block or the coalescence of fibrillatory wavelets surrounding anatomic barriers such as the crista terminalis and eustachian valve. Interestingly, some patients with atrial fibrillation treated with antiarrhythmic drugs, particularly flecainide, propafenone, or amiodarone, may present with atrial flutter rather than atrial fibrillation, because these agents slow atrial conduction velocity and can promote reentry, in addition to suppressing ectopic atrial triggers. This mechanism highlights the complex interplay between atrial fibrillation and atrial flutter, as well as the influence of antiarrhythmic medications on arrhythmia presentation.

Macroreentrant atrial tachycardias (ATs) that are not dependent on conduction through the cavotricuspid isthmus are referred to as *atypical atrial flutters*, also known as "non-cavotricuspid isthmus–dependent" atrial flutters. They can occur in either atrium and are almost universally associated with areas of atrial scar. Right atrial atypical flutter often occurs after cardiac surgery if an atriotomy is performed in the right atrium as part of the surgery. Left atrial flutter and perimitral left

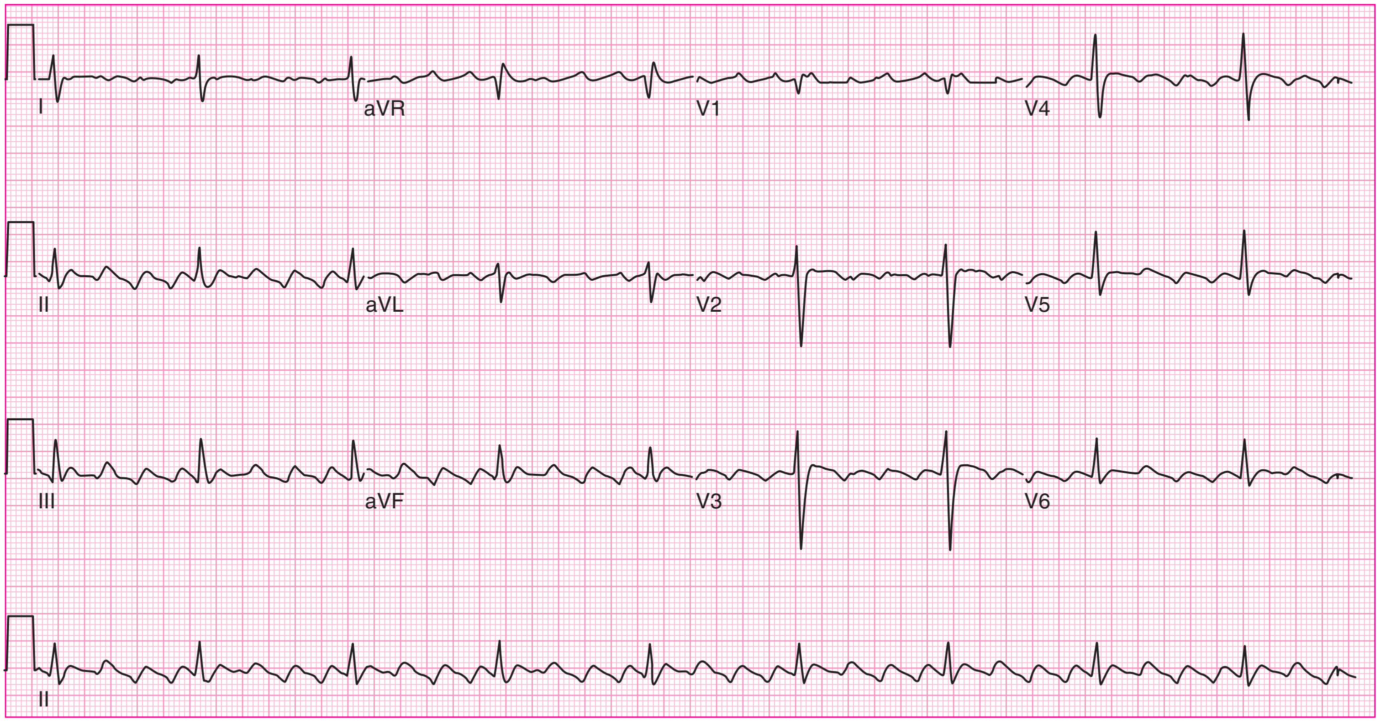

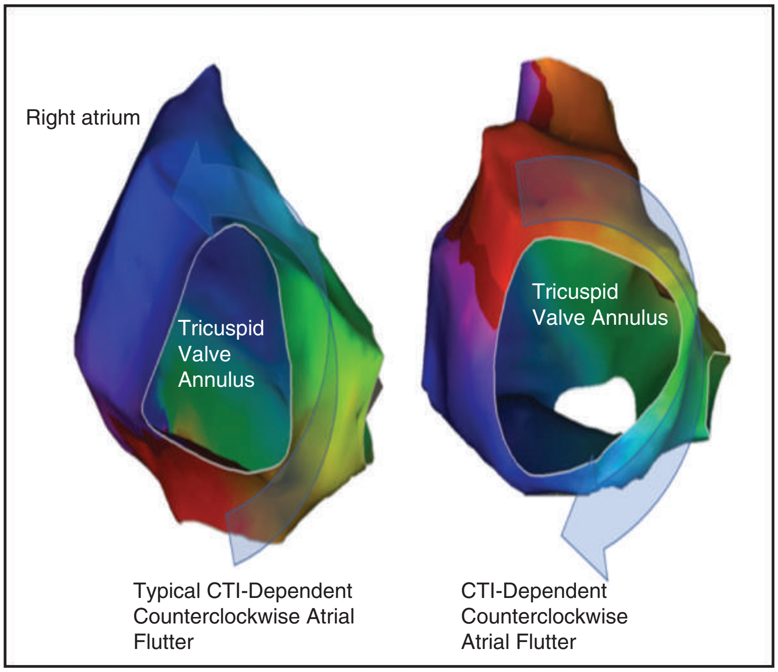
Right atrium
Tricuspid Valve Annulus
Tricuspid Valve Annulus
Typical CTI-Dependent Counterclockwise Atrial Flutter
CTI-Dependent Counterclockwise Atrial Flutter

FIGURE 257-1 A 12-lead electrocardiogram (ECG) of typical or common atrial flutter (counterclockwise activation) is shown. Note the sawtooth pattern of atrial activation, with negative flutter (F) waves in inferior limb lead and positive flutter waves in lead V_1, as well as 4:1 atrioventricular (AV) conduction during flutter. CTI, cavotricuspid isthmus.

atrial flutter are commonly seen after extensive left atrial ablation for atrial fibrillation or mitral valve surgery **(Fig. 257-2)**. While the clinical presentation resembles common atrial flutter, it exhibits different P-wave morphologies. They can be difficult to distinguish from focal AT, and in most cases, the underlying mechanism can only be confirmed by an electrophysiology study **(Fig. 257-3)**.

ATRIAL FLUTTER

Initial management of atrial flutter is similar to that for atrial fibrillation, discussed in more detail in **Chap. 258**. Anticoagulation therapy plays a pivotal role in preventing thromboembolic events in this population and should be determined based on individual thromboembolic risk profiles with the CHA_2DS_2-VASc scoring system and bleeding risk assessment. Nonetheless, anticoagulation is crucial before considering conversion to sinus rhythm, unless electrical cardioversion is warranted for hemodynamic instability or severe symptoms.

Rate control can be achieved with administration of AV nodal–blocking agents. However, achieving optimal rate control in atrial flutter may be more challenging compared to atrial fibrillation. Once rate control is achieved or the patient is hemodynamically stable, consideration is given to rhythm control strategies for restoring and maintaining sinus rhythm. Antiarrhythmic drug therapy such as sotalol, dofetilide, disopyramide, ibutilide, and amiodarone may be considered, but >70% of patients experience recurrences. Therefore, catheter ablation is recognized as first-line therapy due to its high success rate (i.e., >95%), low risk of complications, and low recurrence rates. However, up to half of the patients undergoing cavotricuspid isthmus ablation go on to develop atrial arrhythmias, with atrial fibrillation being the most

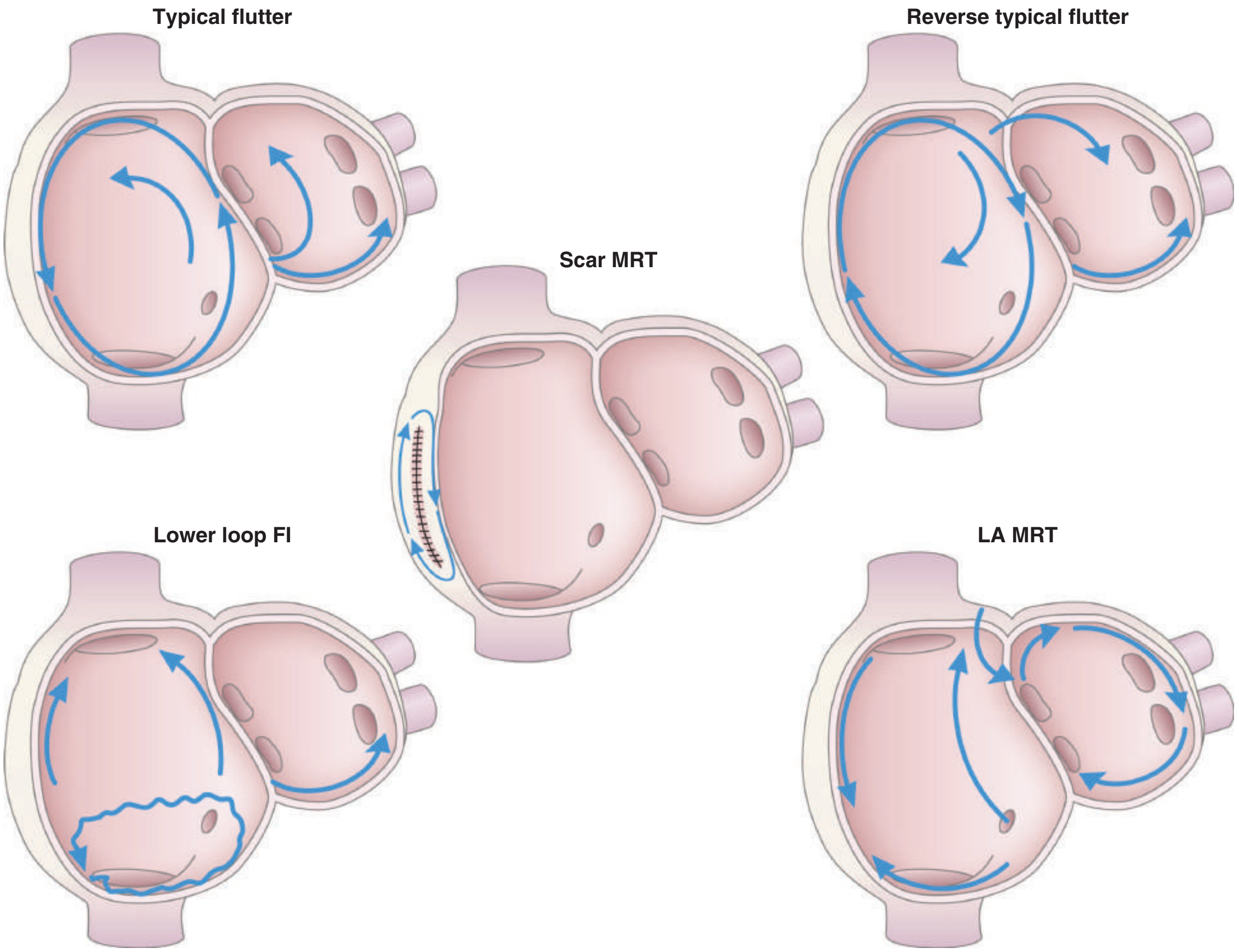

FIGURE 257-2 Types of atrial flutter and macroreentrant atrial tachycardia. The typical, reverse typical, and lower-loop flutter circuits all involve the low right atrial isthmus. Additionally, other forms of macroreentrant flutters encompass scar-mediated reentrant tachycardia and left atrial mitral isthmus flutter. Fl, flutter; LA, left atrium; MRT, macroreentrant. *(Reprinted with permission Circulation. 149:e1-e156, 2024. ©2023 American Heart Association, Inc.)*

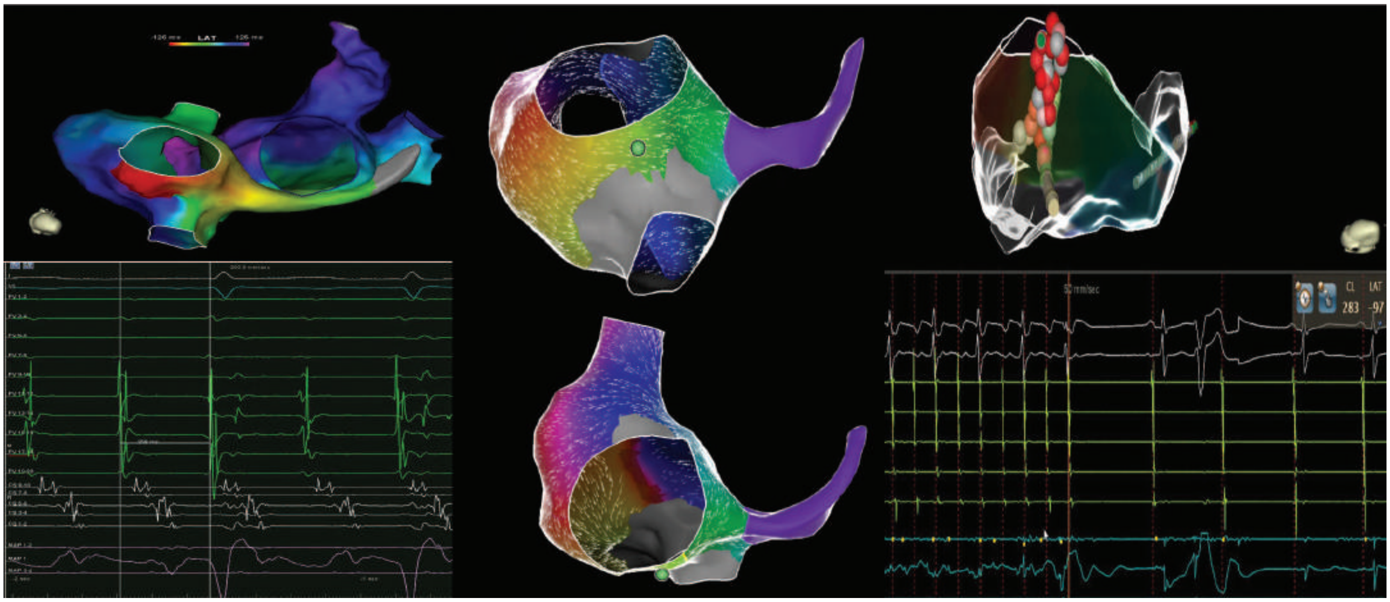

FIGURE 257-3 Electroanatomic mapping of typical atrial flutter. Electroanatomic maps reveal counterclockwise (more common) cavotricuspid isthmus (CTI)–dependent atrial flutter. These maps depict the electrical activation pattern during flutter, obtained during electrophysiologic study and catheter ablation. Colors refer to local activation time where "early meets late" (red-purple). In the *left panel*, electrograms display atrial flutter with a cycle length of 260 ms. The *middle panel* shows an activation map with vector integration, while the *right panel* illustrates termination of the atrial flutter into sinus rhythm by delivering radiofrequency ablation lesions in the CTI.

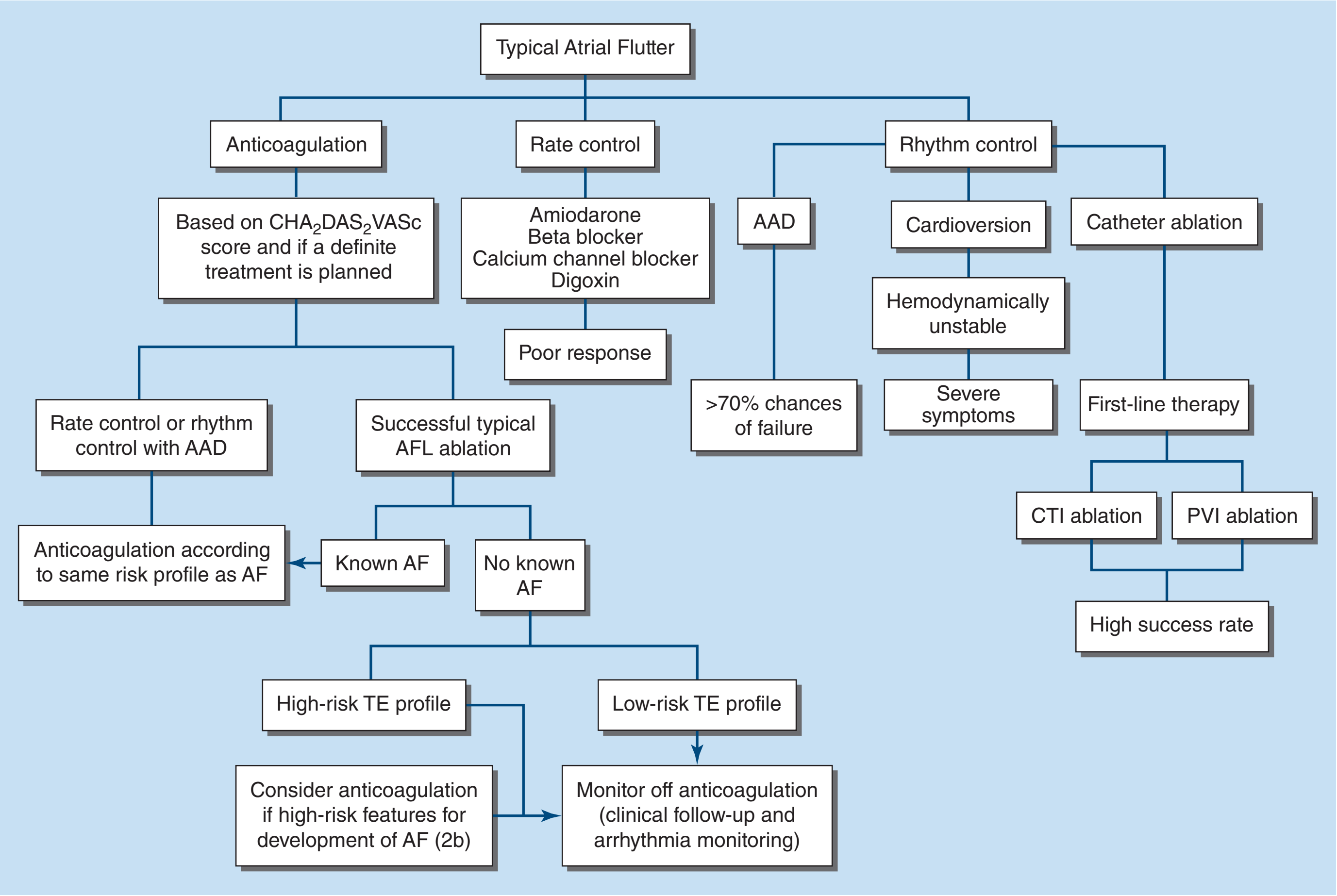
Typical Atrial Flutter
Anticoagulation
Rate control
Rhythm control
Based on CHA_2DAS_2VASc score and if a definite treatment is planned
Amiodarone
Beta blocker
Calcium channel blocker
Digoxin
Poor response
AAD
Cardioversion
Catheter ablation
Hemodynamically unstable
Severe symptoms
>70% chances of failure
First-line therapy
CTI ablation
PVI ablation
High success rate
Rate control or rhythm control with AAD
Successful typical AFL ablation
Anticoagulation according to same risk profile as AF
Known AF
No known AF
High-risk TE profile
Low-risk TE profile
Consider anticoagulation if high-risk features for development of AF (2b)
Monitor off anticoagulation (clinical follow-up and arrhythmia monitoring)

FIGURE 257-4 Approach to the patient with atrial flutter. AAD, antiarrhythmic drug; AF, atrial fibrillation; AFL, atrial flutter; CTI, cavotricuspid isthmus; PVI, pulmonary vein isolation; TE, thromboembolic event. *(Modified from JA Joglar et al: 2023 ACC/AHA/ACCP/HRS Guideline for the Diagnosis and Management of Atrial Fibrillation: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 149:e1, 2024.)*

common among them. While prophylactic cavotricuspid isthmus ablation may not be beneficial in patients with atrial fibrillation, a recent study suggested that treating atrial flutter with pulmonary vein isolation as first-line therapy can prevent recurrent atrial flutter, in addition to reducing the risk of new-onset atrial fibrillation.

In summary, the management of atrial flutter involves a stepwise approach beginning with anticoagulation for thromboembolic prevention, followed by rate control to manage symptoms and hemodynamic stability and, more importantly, rhythm control to maintain sinus rhythm. Catheter ablation is increasingly favored as the first-line therapy for achieving long-term rhythm control and improving quality of life in these patients. More importantly, additional studies are needed to substantiate a more effective ablation strategy aimed at preventing the recurrence of atrial flutter and reducing the incidence of new-onset atrial fibrillation episodes. Lastly, postablation anticoagulation remains a subject of controversy, contingent upon the likelihood of developing atrial fibrillation **(Fig. 257-4)**.

MULTIFOCAL ATRIAL TACHYCARDIA

Multifocal AT (MAT) is characterized by a rhythm with at least three distinct P-wave morphologies, with rates typically from 100 to 150 beats/min. Unlike atrial fibrillation, MAT exhibits clear isoelectric intervals between P waves, and the atrial rate tends to be slower. The underlying mechanism is often attributed to triggered automaticity originating from multiple atrial foci. This condition is commonly observed in patients with chronic pulmonary disease and acute illnesses **(Fig. 257-5)**.

Therapy for MAT is directed at treating the underlying disease and correcting any metabolic abnormalities. Electrical cardioversion is ineffective. Calcium channel blockers such as verapamil or diltiazem may be used to slow the atrial and ventricular rate. Patients with severe pulmonary disease often do not tolerate beta-blocker therapy well. MAT may respond to amiodarone, but long-term therapy with this agent is usually avoided due to its potential toxicities, particularly pulmonary fibrosis. The associated risk of thromboembolism in MAT remains unclear but is not considered to be the same as atrial fibrillation or atrial flutter.

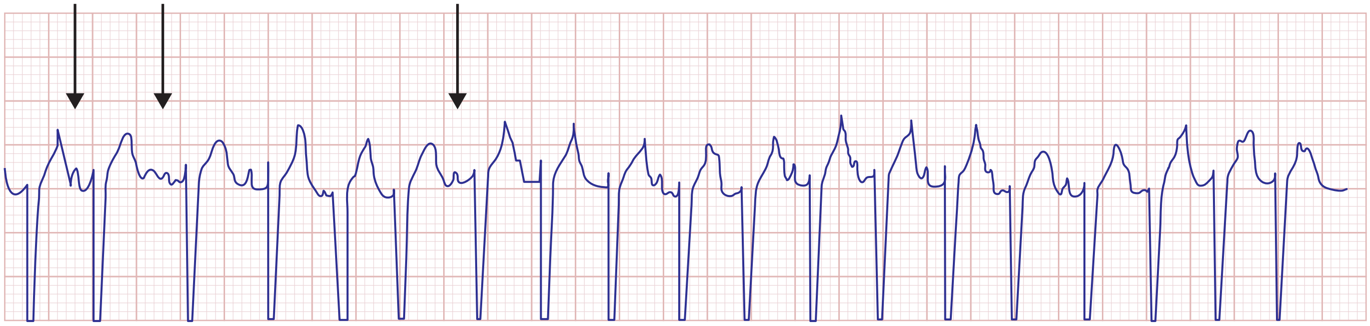

FIGURE 257-5 Multifocal atrial tachycardia. Rhythm strip obtained from a patient with severe pulmonary disease during an acute illness. Note the irregularity of the rhythm but with distinct P waves. *Arrows* note three distinct P-wave morphologies.

■ FURTHER READING

Brugada J et al: 2019 ESC Guidelines for the management of patients with supraventricular tachycardia. The task force for the management of patients with supraventricular tachycardia of the European Society of Cardiology (ESC) developed in collaboration with the Association for European Paediatric and Congenital Cardiology (AEPC). Eur Heart J 41:655, 2020.

Callans DJ: *Josephson's Clinical Cardiac Electrophysiology: Techniques and Interpretations*, 7th ed. Philadelphia, Wolters Kluwer, 2024.

Jalife J, Stevenson W (eds): *Zipes and Jalife's Cardiac Electrophysiology: From Cell to Bedside*, 8th ed. Philadelphia, Elsevier, 2022.

Joglar JA et al: 2023 ACC/AHA/ACCP/HRS Guideline for the Diagnosis and Management of Atrial Fibrillation: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. J Am Coll Cardiol 83:109, 2024.

258

Atrial Fibrillation

William H. Sauer, Jorge E. Romero, Paul C. Zei

PATHOPHYSIOLOGY AND EPIDEMIOLOGY

Atrial fibrillation (AF) is a cardiac arrhythmia characterized by seemingly disorganized, rapid, and irregular atrial electrical activation, resulting in loss of organized atrial mechanical contraction. These rapid and irregular electrical signals input into the atrioventricular (AV) node, which determines ventricular activation and rate. The conducted ventricular rate is variable, resulting in an irregular, usually rapid ventricular rate, ranging typically between 110 and 160 beats/min in most. In some patients, the sustained ventricular rate can exceed 200 beats/min, whereas in others with either high vagal tone or AV nodal conduction disease, the ventricular rate may be excessively slow **(Fig. 258-1)**.

The disorganized atrial activation is best appreciated in lead V_1 for this patient. AF is the most common sustained arrhythmia; as a result, it is a major public health issue. Prevalence increases with age, with >95% of AF patients >60 years of age. The prevalence in humans over age 80 is ~20%. The lifetime risk of developing AF for men aged 40 years old is ~25%. AF is slightly more common in men than women and more common in whites than blacks. Risk factors for developing AF in addition to age and underlying cardiac disease include hypertension, diabetes mellitus, cardiac disease, family history of AF, obesity, thyroid disease, and sleep-disordered breathing. AF is not a benign condition, with a 1.5- to 1.9-fold increased risk of mortality after controlling for underlying cardiac disease. Perhaps the most important consequence of AF is a significantly increased risk of stroke compared to the general population, causing ~25% of all strokes. AF has been detected up to 8.9% of patients within 6 months following cryptogenic stroke using insertable cardiac monitors.

The risk of dementia is increased in patients with AF, as is the risk of magnetic resonance imaging (MRI)-detected asymptomatic embolic infarct. AF, most often when ventricular rate remains uncontrolled for prolonged periods, increases the risk of developing congestive heart failure and cardiomyopathy. Moreover, as a corollary, patients with underlying heart disease, in particular cardiomyopathy and congestive heart failure, are at higher risk for developing AF. AF is a marker for worsened morbidity and mortality in patients with existing heart disease, although the precise extent of the independent risk increase associated with AF in heart disease is unclear. AF may, on occasion, be associated with an identifiable precipitating factor, such as hyperthyroidism, acute alcohol intoxication, myocardial infarction, pulmonary embolism, pericarditis, and cardiac surgery, where AF occurs in up to 50% of patients postoperatively.

AF is clinically most typically defined by the pattern of episodes. Paroxysmal AF is defined as a pattern of AF episodes that occur and terminate with a relatively short duration either spontaneously or by pharmacologic or electrical cardioversion, most commonly defined as 7 days or less. Persistent AF refers to AF that occurs continuously for >7 days but <1 year, whereas long-standing persistent AF refers to AF that has been persistent for >1 year. These descriptors for AF correlate somewhat with the underlying pathophysiology of AF. AF tends to be a progressive condition, with, at this point, no definitive "cure" that will completely eliminate AF durably in a predictable fashion. The pathophysiology of AF, however, remains incompletely understood. Most data support a multifactorial process that leads to the development of manifest AF. Clinical and epidemiologic studies have demonstrated that, in addition to cardiovascular disease, age, alcohol use, obesity, hypertension, diabetes mellitus, and sleep-disordered breathing are associated with higher risk of developing AF. The proposed pathophysiology suggests a "final common pathway" of these risk factors leading to electrophysiologic changes in atrial tissues. Alterations in regulation of membrane channels and other proteins result in abnormal electrical excitability. Atrial tissues, in particular pulmonary vein musculature, exhibit enhanced automaticity, resulting in ectopic beats (premature atrial contractions), as shown in **Fig. 258-2**. Bouts of rapid atrial ectopy may then initiate either atrial tachycardia or frank AF. Additional cellular and, eventually, tissue remodeling results in abnormal conduction properties throughout the atria, including, in particular, shortening of atrial tissue refractory periods. This enables sustained AF through a combination of rapid automaticity-based "drivers" and areas of functional reentry. Further remodeling leads to the development of fibrosis and left atrial enlargement **(Table 258-1)**.

These functional and anatomic changes in atrial tissues appear to correlate with the progression of clinical AF. AF tends to be a progressive disease in most, although exceptions occur. Typically, for a period of time, patients experience sporadic ectopic beats and short runs of atrial tachycardia, likely originating from the pulmonary veins, preceding the onset of frank AF.

Other regions of the atria have been demonstrated to produce ectopic depolarizations that may trigger AF; these include the posterior wall of the left atrium and muscular tissue sleeves within the superior vena cava, coronary sinus, or the remnant of the vein of Marshall. When enough frequent bursts of ectopic beats/tachycardia and/or changes in underlying substrate support the maintenance of AF for short periods, the patient develops episodes of paroxysmal AF. In the untreated patient, over time, as the electrical, contractile, and structural remodeling continues to progress, episodes of paroxysmal AF may be prolonged to the point of not terminating spontaneously, the hallmark of persistent AF. After further remodeling, not only do patients continue to long-standing persistent AF but also the efficacy of therapeutic interventions to restore sinus rhythm diminishes.

CLINICAL PRESENTATION AND MANIFESTATIONS

The clinical manifestations of AF result from (1) symptoms related to the irregular, often rapid but sometimes slow ventricular rates that result; (2) the hemodynamic consequences of altered cardiac function; (3) the consequences of cardioembolic phenomena; and/or (4) the impact of AF on cardiovascular function over time. AF is diagnosed by electrocardiogram (ECG), either by 12-lead standard ECG, limited lead ambulatory monitor ECG and implantable loop recorders, with findings of lack of organized atrial activity (no P wave), with an irregular ventricular response. The role of screening populations for AF is evolving with the use of wearable monitors and home ECG capabilities.

With irregular, rapid ventricular rates, there is variable cardiac displacement and contraction, resulting in the sensation of palpitations and awareness of the heartbeat, when of course, in a normal rhythm, most humans do not sense each heartbeat. Interestingly, many patients are, for the most part, unaware of the irregular ventricular beating for unknown reasons.

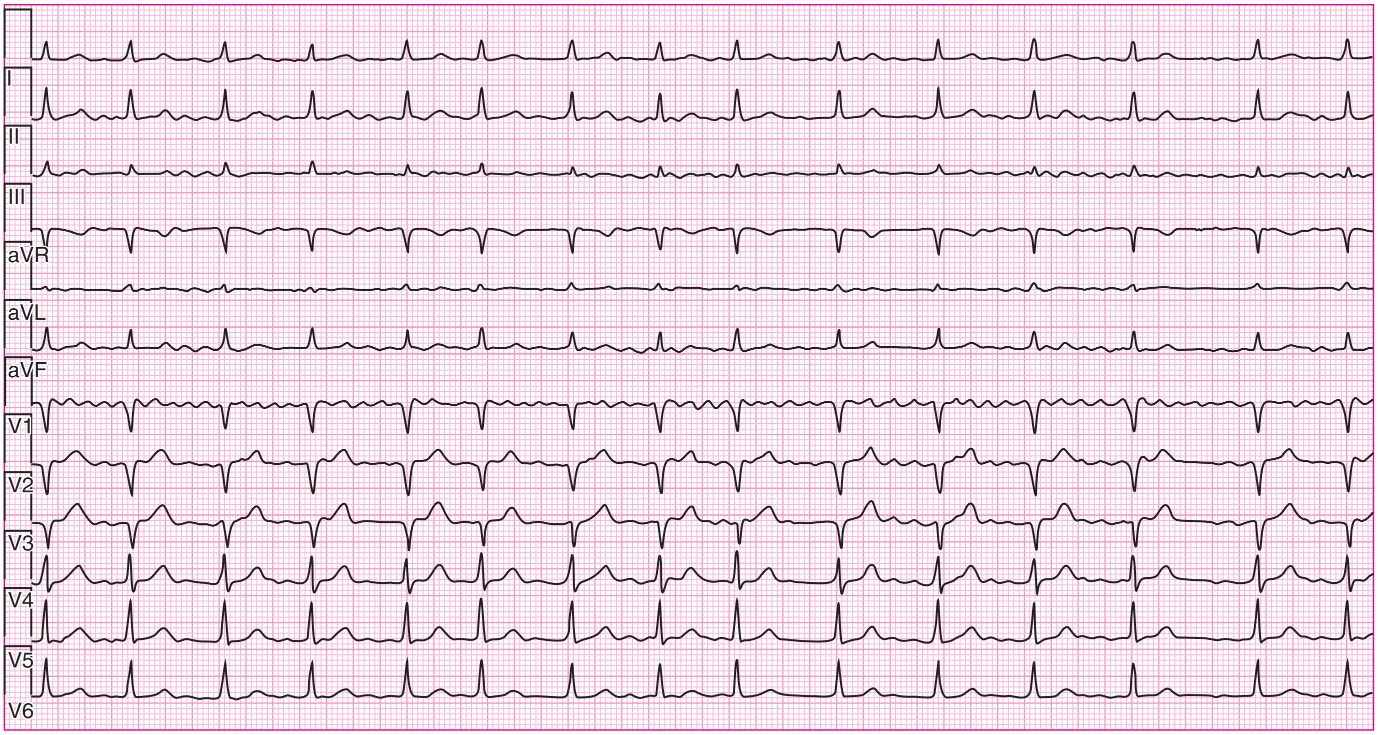

FIGURE 258-1 Electrocardiogram of an irregularly irregular heart rhythm without discernable P waves. The disorganized atrial activation is best appreciated in lead V_1 for this patient.

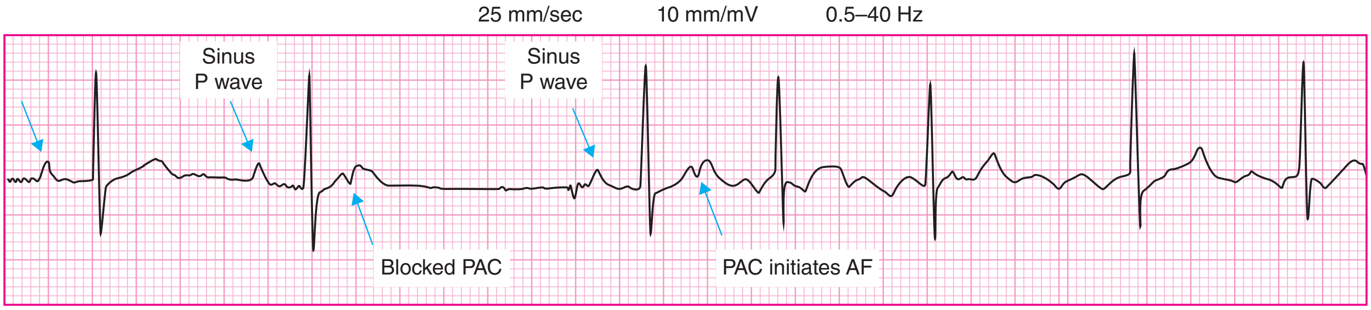

FIGURE 258-2 Surface electrocardiogram (ECG) of atrial ectopy initiating atrial fibrillation (AF). In this single-lead surface ECG recording, the tracing begins with two conducted sinus beats. A nonconducted premature atrial contraction (PAC) (labeled "blocked PAC") is shown after the second QRS complex. After the next sinus P wave and QRS, an ectopic beat (PAC) initiates atrial fibrillation, as demonstrated by (somewhat organized) erratic atrial activity and an irregular ventricular response.

TABLE 258-1 Categorization of Atrial Fibrillation (AF) by Clinical Temporal Characteristics and Associated Features

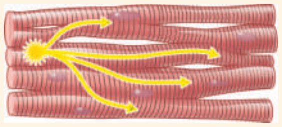

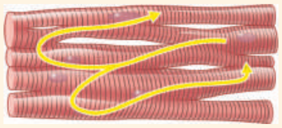

	PAROXYSMAL AF	PERSISTENT AF	LONG-STANDING PERSISTENT AF
Definition	Episodes self-terminate or via pharmacologic or electrical CV in <7 days	Episodes lasting >7 days and <1 year	Persistent AF >1 year
LA size	Normal to mildly enlarged	Mild to severely enlarged	Typically, severely enlarged
LA scar burden	Low	Moderate	High
Efficacy of AAD	Often effective	Not as effective	Usually refractory
When to offer ablation?	First-line therapy reasonable	First-line appropriate but usually offered after AAD failure	After AAD failure, not always a good option
Ablation technique	PV isolation alone usually effective	PV isolation and any identified non-PV AF source	PV isolation; additional ablation for substrate modification likely needed

Note: With paroxysmal, persistent, and long-standing persistent AF, definitions are based on duration of events and diagnosis overall. These categorizations correlate with LA size, LA scar burden, and resultant efficacy of medical and ablative therapies.

Abbreviations: AAD, antiarrhythmic drugs; CV, cardioversion; LA, left atrium; PV, pulmonary vein.

During AF, there is loss of the contribution of atrial systole to overall cardiac output and, with irregular ventricular rates, variable ventricular filling and, consequently, variable stroke volume. The resultant impact on overall cardiac output may result in exercise intolerance, fatigue, weakness, presyncope, or dyspnea. In patients with underlying cardiac disease, the additional hemodynamic compromise resulting from AF may result in exacerbation of the disease and/or heart failure symptoms. Patients with hypertrophic cardiomyopathy, coronary artery disease, valvular disease, heart failure with either depressed or preserved ejection fraction, or amyloidosis are particularly susceptible. In patients with concomitant AV nodal conduction disease, bradycardia during AF may result in presyncope or syncope. Pauses at the time of spontaneous conversion from AF to sinus rhythm, a manifestation of sinus node dysfunction that commonly occurs in patients with AF, may result in presyncope or syncope as well.

With the loss of atrial mechanical contraction, blood stasis may promote in situ thrombosis, which, when embolized, may result in a range of clinical consequences, most importantly, ischemic stroke. Thrombus formation occurs primarily in the left atrial appendage. Over time, recurrent thromboembolism to the brain, even if asymptomatic, may result in debilitating neurologic sequelae, including cognitive impairment. An increased risk of dementia in patients with AF may be the consequence of this phenomenon, although the contribution of chronic hypoperfusion in patients with long-standing persistent atrial fibrillation is unclear.

In patients with prolonged periods of rapid ventricular rates resulting from AF, there is risk of developing a tachycardia-induced cardiomyopathy, with associated depressed left ventricular function. Tachycardia-induced myopathy appears generally to be reversible once ventricular rates are controlled. In patients with long-standing persistent AF, the atria, especially the left atrium, tend to be more dilated and to contain a higher burden of fibrotic, noncontractile atrial tissue. The hemodynamic consequences of a noncompliant, fibrotic left atrium, including elevated left atrial filling pressures, volume overload, and congestive heart failure, have been described as "stiff left atrial syndrome."

TREATMENT

Atrial Fibrillation

The treatment and management of the patient with AF centers on three aims: (1) control of patient symptoms through a strategy of rate control and/or rhythm control; (2) appropriate mitigation of thromboembolism risk; and (3) addressing modifiable risk factors for progression of AF. In the acute onset of AF, if significant hemodynamic compromise, pulmonary edema, or evidence of coronary ischemia is present, emergent cardioversion is recommended. Electrical cardioversion can be achieved with a QRS synchronous shock, preferably in a sedated patient, or via pharmacologic cardioversion, most typically with the intravenous administration of the class III antiarrhythmic ibutilide. Ibutilide should be avoided in patients with baseline prolonged QT interval or severe left ventricular dysfunction, given the risk of torsades des pointes. In the hemodynamically stable patient with new-onset AF, therapy should focus on control of ventricular rate to prevent hemodynamic sequelae, consideration of anticoagulation to mitigate thromboembolic risk, and consideration of restoration and maintenance of sinus rhythm—a so-called rhythm control strategy. If restoration of sinus rhythm is being considered, a more immediate risk of thromboembolism must be factored into the management strategy. Although there is a lack of definitive data, it is presumed that if the presenting episode of AF is >48 h or if the episode duration is unknown, there is risk for precipitating a thromboembolic complication through cardioversion, whether electrically or pharmacologically achieved. Therefore, in this circumstance, the patient should be either initiated on anticoagulation, with cardioversion deferred for at least 3 weeks after uninterrupted anticoagulation, or evaluated to exclude the presence of left atrial appendage thrombus. Most commonly, transesophageal echocardiography (TEE) is used to evaluate for left atrial appendage thrombus, although cardiac computed tomography (CT) angiography using delayed acquisition imaging has been demonstrated to have excellent sensitivity and specificity as well.

CARDIOVERSION AND ANTICOAGULATION

The major source of thromboembolism and stroke in nonvalvular AF is formation of thrombus in the left atrial appendage where flow is relatively stagnant, although thrombus occasionally forms in other locations as well, particularly in patients with mitral valvular disease and severely dilated left atrium. Following conversion from prolonged AF to sinus rhythm, atrial mechanical function can be delayed for weeks (i.e., atrial stunning), such that thrombi can form even during sinus rhythm. When AF has been present for >48 h and in patients at high risk for thromboembolism, such as those with mitral stenosis or hypertrophic cardiomyopathy, conversion to sinus rhythm is associated with an increased risk of thromboembolism. Thromboembolism can occur soon or several days after restoration of sinus rhythm if appropriate anticoagulation measures are not taken. In patients with AF and left atrial appendage closure devices (e.g., Watchman device), electrical cardioversion is feasible without the need for oral anticoagulation if preprocedural transesophageal echocardiography shows good device position, absence of device-related thrombus, and peri-device leak of ≤5 mm.

Cardioversion *within 48 h of the onset of AF* without TEE or cardiac CT is common practice in patients who have not been anticoagulated, provided that they are not at high risk for stroke due to a prior history of embolic events, rheumatic mitral stenosis, or hypertrophic cardiomyopathy with marked left atrial enlargement. These low-risk patients with occasional episodes of AF can be instructed to notify their physician when AF occurs to arrange for cardioversion to be done within 48 h.

If the duration of AF exceeds 48 h or is unknown, there is greater concern for thromboembolism after cardioversion, even in patients considered low risk (CHA_2DS_2-VASc of 0 or 1 [see below]) for stroke. There are two approaches to mitigate the risk related to cardioversion. One option is to anticoagulate continuously for 3 weeks before and a minimum of 4 weeks after cardioversion. A second more frequently used approach is to start anticoagulation and perform a TEE or high-resolution cardiac CT scan to detect the presence of thrombus in the left atrial appendage. If thrombus is absent, electrical or pharmacologic cardioversion can be performed and anticoagulation continued for a minimum of 4 weeks to allow time for recovery of atrial mechanical function. In either case, cardioversion of AF is associated with a substantial risk of recurrence, which may not be symptomatic. It should be noted that these recommendations for short-term anticoagulation and thrombus exclusion at the time of cardioversion lack contemporary robust data to support these strategies. Longer-term maintenance of anticoagulation is considered based on the patient's individual risk for stroke, commonly assessed using the CHA_2DS_2-VASc score.

ACUTE RATE CONTROL

The goal of rate control in AF is to allow more diastolic filling time, improving cardiac output and reducing patient symptoms. In the longer term, adequate rate control will minimize the risk of congestive heart failure and tachycardia-induced cardiomyopathy. Acute rate control can be achieved with beta blockers and/or the calcium channel blockers verapamil and diltiazem administered either intravenously or orally, as warranted by the urgency of the clinical situation. Digoxin has been used for several years for rate control, particularly in patients with labile blood pressure and in patients with cardiomyopathy susceptible to congestive heart failure, because it lacks the negative inotropic effect seen in calcium channel blockers and beta blockers. It acts synergistically with beta blockers and calcium channel blockers and, therefore, may be useful as an added agent when rate control is inadequate. However, recent evidence suggests increased mortality with its chronic use, and so its utilization has declined.

CHRONIC RATE CONTROL

For patients who remain in AF chronically, the goal of rate control is to both alleviate symptoms and prevent deterioration of ventricular function from excessive rates. β-Adrenergic blockers and calcium channel blockers are often used either alone or in combination. Exertion-related symptoms are often an indication of inadequate rate control. Rate should be assessed with exertion and medications adjusted accordingly. Adequate rate control is defined as a resting heart rate of <80 beats/min that increases to <100 beats/min with light exertion, such as walking. If it is difficult to slow the ventricular rate to that degree, allowing a resting rate of up to 110 beats/min is acceptable provided it does not cause symptoms and ventricular function is normal; however, periodic assessment of ventricular function is warranted because some patients develop tachycardia-induced cardiomyopathy. In patients with permanent atrial fibrillation, a lenient rate-control strategy (resting heart rate <110 beats/min) is as effective as strict rate-control strategy (resting heart rate <80 beats/min and heart rate during moderate exercise <110 beats/min) in terms of death from cardiovascular causes, hospitalization for heart failure and stroke, systemic embolism, bleeding, and life-threatening arrhythmic events, and this strategy is easier to achieve.

If adequate rate control in AF is difficult to achieve, further consideration should be given to restoring sinus rhythm (see below). Catheter ablation of the AV junction to create permanent AV block and implantation of a permanent pacemaker reliably achieve rate control without the need for AV nodal–blocking agents, a so-called "ablate and pace" strategy. These patients not only remain in AF but also become dependent on the pacemaker to support ventricular rate. The typical pacing configuration with placement of a ventricular lead in the right ventricular apex may induce dyssynchronous ventricular activation that can depress ventricular function in some patients. Biventricular pacing or direct pacing of the left bundle branch area may be used to minimize the degree of ventricular dyssynchrony. AV nodal ablation and cardiac resynchronization therapy have been demonstrated to be superior to pharmacologic therapy in improving quality of life and in reducing the development of heart failure, heart failure hospitalizations, and all-cause mortality in patients with permanent AF and a narrow QRS, irrespective of their baseline left ventricular ejection fraction.

STROKE PREVENTION IN ATRIAL FIBRILLATION

CHA$_2$DS$_2$-VASc			**HAS-BLED**	
Risk Criteria				
Congestive heart failure	1		Hypertension	1
Age >75	2		Abnormal renal or liver function	1 each
Hypertension	1		Stroke history	1
			Bleeding diasthesis	1
Diabetes mellitus	1		Labile INR (on warfarin)	1
Prior stroke or TIA	2		Elderly (Age >65)	1
Vascular disease	1		Drugs that predispose to bleeding or alcohol	1 each
Age >65	1			
Sex category (F)	1			

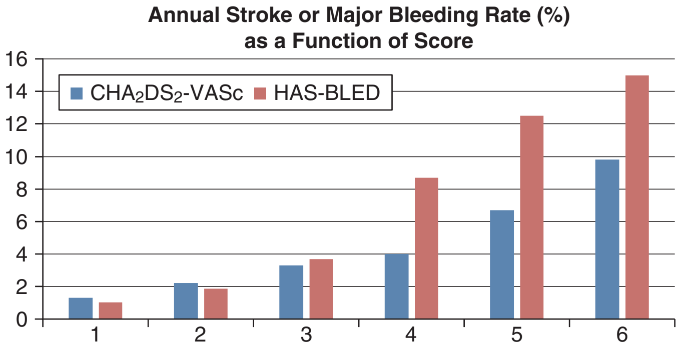

FIGURE 258-3 CHA$_2$DS$_2$-VASc and HAS-BLED systems. The CHA$_2$DS$_2$-VASc scoring system gives a point for each outlined stroke risk factor, whereas the HAS-BLED scoring system gives a point for each bleeding risk factor, as outlined in the table. In the chart below the table, the corresponding risk of stroke (CHA$_2$DS$_2$-VASC) or major bleed event (HAS-BLED) is plotted as a percent risk per annum as a function of score. F, female; INR, international normalized ratio; TIA, transient ischemic attack.

Thromboembolic complications, in particular, stroke, are the most significant and potentially life-threatening sequelae of AF. Therefore, appropriate stroke prevention strategies are a key aspect of AF management. The mainstay of stroke prevention is continuous anticoagulation therapy, most commonly using an oral medication. Specific patient populations have a high risk of stroke, including patients with hypertrophic cardiomyopathy, mitral stenosis, and prior stroke history, and therefore, anticoagulation is recommended, barring contraindications. AF in patients without mitral stenosis is commonly referred to as nonvalvular AF. In most patients with AF, the decision about whether a stroke prevention regimen is indicated is largely based on an assessment of stroke risk, balanced by the risk of the preventative therapy. The risk of stroke appears to be most accurately predicted by the presence of underlying risk factors known to increase stroke risk. The CHA$_2$DS$_2$-VASc scoring system **(Fig. 258-3)** is a widely used tool to estimate stroke risk. Anticoagulation is currently recommended in the United States and Europe for patients with a score of ≥1 unless the lone risk factor is female gender. Stroke risk increases with increasing CHA$_2$DS$_2$-VASc score, such that annual stroke risk may be as high as nearly 20% without anticoagulation. On the other hand, anticoagulation carries a risk of serious and potentially life-threatening bleeding complications, in particular, intracranial hemorrhage and gastrointestinal bleed. Bleeding risk is often assessed using the HAS-BLED scoring system (Fig. 258-3). If bleeding risk is deemed to be outweighed by stroke risk, anticoagulation is recommended. It is important to note the conventional wisdom that the *perceived* burden of AF has not been shown to predict stroke risk. The approach to patients with paroxysmal AF is therefore the same as for persistent AF. It is recognized that many patients who appear to have infrequent AF episodes based on office visits often have asymptomatic episodes that put them at risk. Absence of AF during periodic monitoring is not sufficient to indicate low risk. The role of continuous monitoring with implanted recorders or pacemakers as a guide for anticoagulation in patients with a borderline risk profile is not clear. Subclinical AF is short-lasting and asymptomatic and can usually be detected only by long-term continuous monitoring with implantable loop recorders, pacemakers, or defibrillators. Subclinical AF is associated with an increased risk of stroke by a factor of 2.5. In patients with subclinical AF lasting 6 min to 24 h, apixaban has been shown to decrease the risk of stroke or systemic embolism but increases the risk of major bleeding. Therefore, a more accurate accounting for the impact of AF burden on stroke risk remains to be clarified.

Antiplatelet agents alone are generally not sufficient. In nonvalvular AF, warfarin reduces the annual risk of stroke by 64% compared to placebo and by 37% compared to antiplatelet therapy. Patients with AF with an increased risk of stroke also have an increased risk of venous thromboembolism, which appears to be lower with oral anticoagulation. The direct-acting anticoagulants dabigatran, rivaroxaban, apixaban, and edoxaban were noninferior to warfarin in individual trials of nonvalvular AF patients, and intent-to-treat analysis of pooled data suggests superiority to warfarin by small absolute margins of 0.4–0.7% in reduction of mortality, stroke, major bleeding, and intracranial hemorrhage. Warfarin is required for patients with rheumatic mitral stenosis or mechanical heart valves. Among patients with rheumatic heart disease–associated AF, warfarin therapy has led to a lower rate of a composite of cardiovascular events or death than rivaroxaban therapy, without a higher rate of bleeding. Similarly, apixaban and dabigatran have failed to demonstrate noninferiority to warfarin and are less effective than warfarin for the prevention of valve thrombosis or thromboembolism in patients with mechanical heart valves.

Warfarin can be an inconvenient agent that requires several days to achieve a therapeutic effect (prothrombin time [PT]/international normalized ratio [INR] >2), requires monitoring of PT/INR to adjust dose, and has many drug and food interactions that can hinder patient compliance and render maintaining a therapeutic effect challenging. The direct-acting agents are easier to use and achieve reliable anticoagulation promptly without requiring dosage adjustment based on blood tests. Dabigatran, rivaroxaban, and apixaban have renal excretion and require dose adjustment for modest renal impairment, which is of particular concern in the elderly, who are at increased bleeding risk. Limited experience with apixaban and rivaroxaban demonstrates safety and efficacy in patients undergoing chronic hemodialysis for end-stage kidney disease. Excretion can also be influenced by P-glycoprotein inducers and inhibitors. Warfarin anticoagulation can be reversed by administration of fresh frozen plasma, prothrombin complex concentrate, and vitamin K. Reversal agents are available for dabigatran (idarucizumab), and Xa inhibitors are available (andexanet alfa), and both are administered intravenously. These agents may be prothrombotic, and administration must be judicious. The antiplatelet agents aspirin and clopidogrel are inferior to warfarin for stroke prevention in AF and do not have less risk of bleeding. Clopidogrel combined with aspirin is better than aspirin alone for stroke prevention, but this combination is inferior to warfarin and has a greater bleeding risk than aspirin alone.

Bleeding is the major risk of anticoagulation. Major bleeding requiring transfusion and intracranial bleeding occur in ~1% of patients per year with warfarin. Direct-acting anticoagulants appear to have a lower risk of intracranial bleeding compared with warfarin without sacrificing protective effects against thromboembolism. Risk factors for bleeding include age >65–75 years, heart failure, renal insufficiency, prior bleeding, and excessive alcohol or nonsteroidal anti-inflammatory drug use. In patients who require dual antiplatelet therapy (e.g., aspirin and clopidogrel) after coronary or peripheral arterial stenting, there is a substantially increased bleeding risk when standard oral anticoagulation with warfarin or a direct-acting anticoagulant is added. In AF patients undergoing percutaneous coronary intervention, the combination of oral platelet inhibition with a $P2Y_{12}$ inhibitor (preferably clopidogrel) is recommended. Triple antithrombotic therapy, preferably including a direct-acting anticoagulant, should be considered in patients with high ischemic risk (e.g., acute coronary syndrome) and for up to 30 days.

Chronic anticoagulation is contraindicated in some patients due to bleeding risks. Because most atrial thrombi likely originate in the left atrial appendage, surgical removal of the appendage, combined with atrial maze surgery, may be considered for patients undergoing surgery, although removal of the appendage has not been unequivocally shown to reduce the risk of thromboembolism. Percutaneously deployed devices that occlude or ligate the left atrial appendage are also available, appear to be noninferior to warfarin in reducing stroke risk, and are considered in patients who have a high risk of thromboembolism but serious bleeding risk from chronic oral anticoagulation (**Table 258-2**). Importantly, left atrial appendage closure devices (i.e., Watchman) provide stroke prevention comparable to warfarin, with additional significant reductions in major bleeding, particularly hemorrhagic stroke, and all-cause mortality. Furthermore, left atrial appendage closure devices appear to be noninferior to direct-acting anticoagulants in preventing major AF-related cardiovascular, neurologic, and bleeding events.

RHYTHM CONTROL

The decision to administer antiarrhythmic drugs or perform catheter ablation to attempt maintenance of sinus rhythm (commonly referred to as the *rhythm control strategy*) is mainly guided by patient symptoms and preferences regarding the benefits and risks of therapies. In general, patients who maintain sinus rhythm have better survival than those who continue to have AF. This may be because continued AF is a marker of disease severity or that AF promotes deterioration in cardiac function. In older randomized trials, administration of antiarrhythmic medications to maintain sinus rhythm did not improve survival or symptoms compared to a rate control strategy, and the drug therapy group had more hospitalizations. Disappointing efficacy and toxicities of available antiarrhythmic drugs, in retrospect inappropriate discontinuation of anticoagulation in the rhythm control arms, and patient selection bias may be factors that influenced the results of these trials. Recently, a randomized trial evaluating an early rhythm control strategy (within 1 year of initial presentation) compared to standard rate control demonstrated a reduction in cardiovascular events, including death from cardiovascular causes and stroke. Differences between this study and earlier randomized trials that failed to show a significant difference in outcomes in rate versus rhythm control included the use of catheter ablation and a high adherence rate to anticoagulation despite apparent rhythm control. In patients with heart failure due to depressed left ventricular function, a catheter ablation–based strategy to maintain sinus rhythm appears to provide all-cause mortality benefit compared with a medical rhythm control strategy. Furthermore, the combination of catheter ablation and guideline-directed medical therapy in patients with symptomatic AF and end-stage heart failure who are referred for heart transplantation evaluation is associated with a lower likelihood of a composite of death from any cause, implantation of a left ventricular assist device, or urgent heart transplantation than medical therapy alone.

A rhythm control strategy is usually selected for patients with symptomatic paroxysmal AF, recurrent episodes of symptomatic persistent AF, AF with difficult rate control, and AF that has resulted in depressed ventricular function or that aggravates heart failure. A rhythm control strategy is more likely to be favored in younger patients than in sedentary or elderly patients in whom rate control is more easily achieved. Even if sinus rhythm is apparently maintained, anticoagulation is recommended according to the CHA_2DS_2-VASc stroke risk profile because asymptomatic episodes of AF are common. Following a first episode of persistent AF, a strategy using AV nodal–blocking agents, cardioversion, and anticoagulation is reasonable, in addition to addressing possible aggravating factors. If recurrences are infrequent, periodic cardioversion is reasonable. However, if a patient has frequent symptomatic AF despite rate control, then a rhythm control strategy incorporating

TABLE 258-2 Novel Oral Anticoagulant Dosing

	DABIGATRAN	RIVAROXABAN	APIXABAN	EDOXABAN
Standard dose	150 mg bid	20 mg qd	5 mg bid	60 mg qd
Reduced dose	110 mg bid	15 mg qd	2.5 mg bid	30 mg qd
Dose reduction criteria	Dabigatran 110 mg bid in patients with: age ≥80 years, concomitant use of verapamil, or increased bleeding risk	Creatine clearance 15–49 mL/min	At least 2 of 3 criteria: age ≥80 years, body weight ≤60 kg, or serum creatinine ≥1.5 mg/dL (133 mol/L)	If any of the following: creatinine clearance 30–50 mL/min, body weight ≤60 kg, or concomitant use of dronedarone, cyclosporine, erythromycin, or ketoconazole

Note: As of publication, four novel or direct oral anticoagulants are available and indicated for stroke prevention for atrial fibrillation. The standard dosing, reduced dosing, and criteria for reduced dosing are shown for each agent.

catheter ablation and/or antiarrhythmic medications is indicated. Based on recent randomized trial data demonstrating superiority of ablation over medications for maintenance of sinus rhythm and benefits of an early rhythm control strategy, catheter ablation is considered first-line therapy, especially for individuals with paroxysmal AF.

Pharmacologic Therapy for Maintaining Sinus Rhythm The goal of pharmacologic therapy is to maintain sinus rhythm or reduce episodes of AF. Risks and side effects of antiarrhythmic drugs are a major consideration in selecting therapy. Drug therapy can be instituted once sinus rhythm has been established or in anticipation of cardioversion. However, antiarrhythmic medications may in some instances pharmacologically cardiovert the patient into sinus rhythm. Therefore, an appropriate anticoagulation strategy approach similar to electrical cardioversion is recommended, particularly at the time of initiation of therapy. β-Adrenergic blockers and calcium channel blockers help control ventricular rate, improve symptoms, and possess a low-risk profile, but have low efficacy for preventing or terminating AF episodes. Class I sodium channel–blocking agents (e.g., flecainide, propafenone, disopyramide) are options for patients without significant structural heart disease, but negative inotropic and proarrhythmic effects warrant avoidance in patients with coronary artery disease or heart failure. The class III agents sotalol and dofetilide can be administered to patients with coronary artery disease or structural heart disease but have ~3% risk of inducing excessive QT prolongation and torsades des pointes. Dofetilide should be initiated only in a hospital with ECG monitoring, and many physicians take this approach with sotalol as well. Dronedarone increases mortality in patients with heart failure or long-standing persistent AF. All these agents have modest efficacy in patients with paroxysmal AF, of whom ~30–50% will benefit. Amiodarone is more effective, maintaining sinus rhythm in approximately two-thirds of patients. It can be administered to patients with heart failure and coronary artery disease. However, >40% of patients experience amiodarone-related toxicities during long-term therapy, and thus, careful monitoring of potential toxicities, including skin, liver, lung, and thyroid abnormalities, must be accompanied with this therapy.

Catheter and Surgical Ablation for Maintaining Sinus Rhythm Successful catheter ablation avoids antiarrhythmic drug toxicities, but procedural risks and efficacy depend on operator experience. For patients with previously untreated but recurrent paroxysmal AF, catheter ablation has superior efficacy compared to antiarrhythmic drug therapy, and ablation is even more clearly superior to antiarrhythmic drugs for patients who have recurrent AF despite drug treatment. Long-term control of AF is more difficult to achieve in patients with persistent and long-standing persistent AF, likely because of more extensive atrial abnormalities and associated greater comorbidities in these patients that may promote ongoing progression of atrial abnormalities that in turn promote AF recurrence.

Catheter ablation involves percutaneous venous access (typically via the femoral veins), trans (atrial) septal puncture, and radiofrequency ablation or cryoablation to electrically isolate the left atrial regions around the pulmonary vein antra, abolishing the ability of triggering foci in these regions to initiate AF and also likely impacting the substrate for reentry in the left atrium **(Fig. 258-4)**. Gaps in healed ablation areas or emergence of new trigger sites outside the pulmonary veins necessitate a repeat procedure in 10–30% of patients. Several alternative energy sources to create ablative lesions are being evaluated for ablation of AF and other arrhythmias, including laser, external beam radiation, and pulsed field electroporation.

Pulsed-field ablation uses electric fields generated by short pulses of high energy and has shown promise by specifically targeting myocardium without generating heat or damaging adjacent tissue **(Fig. 258-5)**. Myocardial cells are uniquely sensitive to high-voltage, short-duration electric fields with electroporation thresholds of 268–375 V/cm compared to other tissue types including nerves, endothelium, vascular smooth muscle, and blood cells, all of which have electroporation thresholds >1600 V/cm. The pulse waveforms used to generate an electric field can have many different characteristics including voltage amplitude, pulse width, cycle period, voltage polarity (monophasic vs. biphasic), electrode polarity (unipolar vs. bipolar), and the number of pulses delivered in a train. There are limited data evaluating the impact of how each of these variables affect lesion safety and efficacy (Fig. 258-5). Therefore, unlike radiofrequency ablation, pulsed-field ablation in its current clinical iteration lacks the ability to titrate and tailor energy delivery during ablation.

In patients with paroxysmal AF, sinus rhythm is maintained for >1 year after a single ablation procedure in ~80% of patients and is achieved in >90% of patients after multiple procedures in some studies. Among patients with paroxysmal AF receiving a catheter-based therapy, pulsed-field ablation has demonstrated to be noninferior to conventional thermal ablation (i.e., radiofrequency ablation and cryoablation) with respect to freedom from a composite of initial procedural failure, documented atrial tachyarrhythmia after a 3-month blanking period, antiarrhythmic drug use, cardioversion, or repeat ablation and with respect to device- and procedure-related serious adverse events at 1 year.

Many patients become more responsive to antiarrhythmic drugs or become less symptomatic with a reduced AF burden after a pulmonary vein isolation procedure, and thus, repeat ablation may not be required for symptom control in some. Ablation is less effective in patients with persistent AF, particularly long-standing persistent AF, especially when associated with more extensive cardiac disease, comorbidities, and evidence of moderate and severe left atrial enlargement. More extensive ablation is often required, targeting areas that likely support reentry and/or AF maintenance and regions outside but adjacent to the pulmonary venous antra.

Most ablation targets and strategies beyond pulmonary vein isolation have failed to show systematic outcome improvement in randomized controlled clinical trials. However, individualized

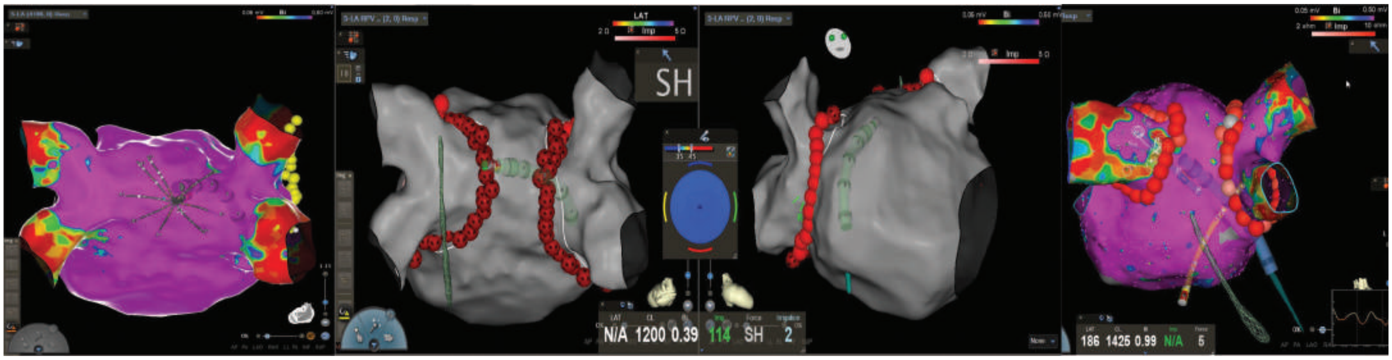

FIGURE 258-4 ***A.*** (*Left*) Electroanatomic map superimposed on a cardiac computed tomography reconstruction of a left atrium with mapping catheter in the posterior wall of this chamber. (*Middle*) Final radiofrequency lesion set around the pulmonary veins. (*Right*) Multipolar catheter in the right inferior pulmonary vein ***B.*** Spontaneous pulmonary vein (PV) ectopy initiating fibrillatory conduction contained within the isolated vein while 12-lead electrocardiogram shows normal sinus rhythm.

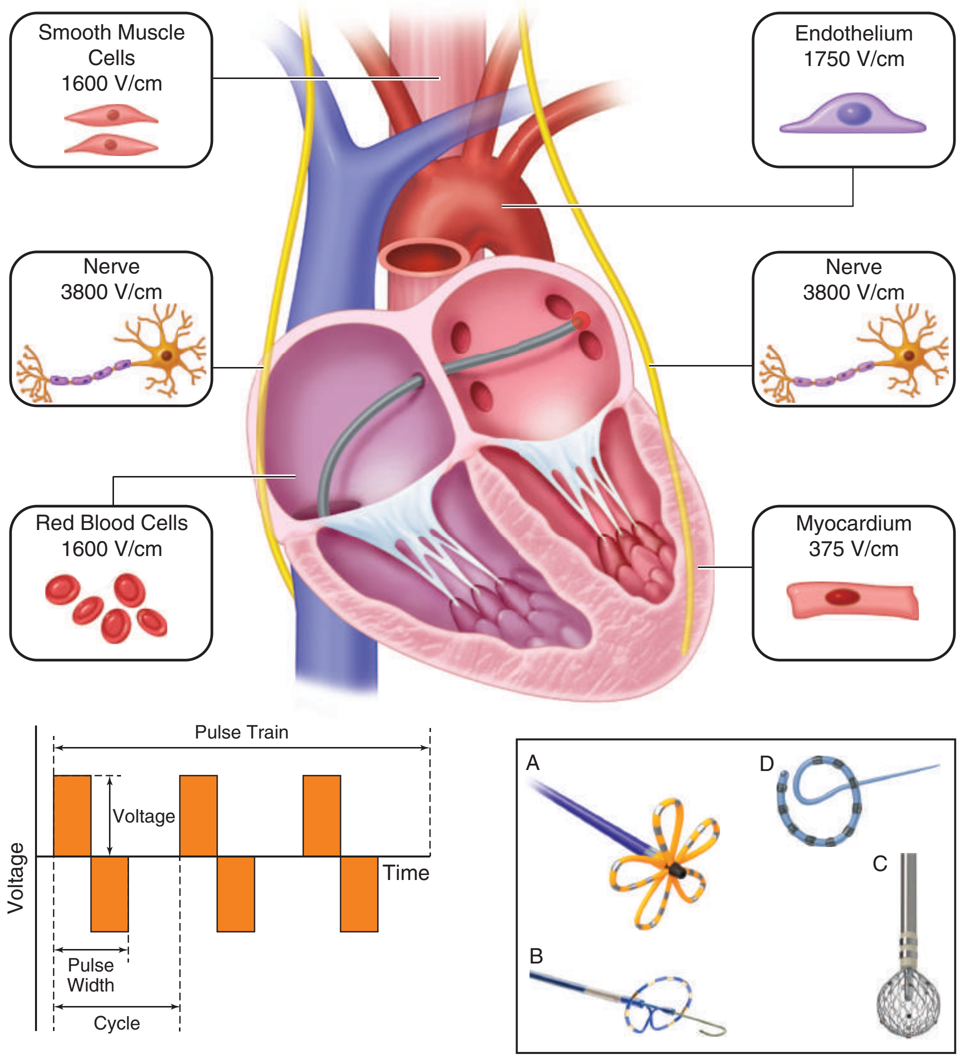

Waveform Variables

- Pulse amplitude (voltage)
- Pulse polarity (monophasic-biphasic)
- Number of pulses in a train
- Pulse width
- Cycle period

Catheter Variables

- Contact force
- Electrode surface area
- Electrode polarity (uni vs bipolar)
- Electrode shape (torus vs ring)
- Electrode and tissue orientation

FIGURE 258-5 Pulsed-field electroporation. (*Top*) Pulsed-field ablation has the potential to target specifically myocardial tissue without negatively affecting adjacent structures or cells such as red blood cells, nerves, the esophagus, or arteries. (*Bottom*) Numerous factors are involved in creating long-lasting transmural lesions with pulse-field ablation; a combination of most of these parameters will eventually help in delivering electroporation effectively and safely into the myocardial tissue. Catheters currently undergoing clinical evaluation for pulsed-field ablation (PFA). *(Reproduced from CD Matos et al: Pulsed Field Ablation of Atrial Fibrillation: A Comprehensive Review. Rev Cardiovasc Med 24:337, 2023 and Reproduced with permission from NA Steiger, JE Romero. Pulsed-field ablation: What are the unknowns and when will they cease to concern us. J Cardiovasc Electrophysiol 33:1489, 2022.)* (***A***) Farawave, reproduced with permission from Boston Scientific; (***B***) PVAC, reproduced with permission from Medtronic; (***C***) Sphere-9, reproduced with permission from Medtronic; (***D***) Varipulse.

ablation of atrial low-voltage myocardium in addition to pulmonary vein isolation significantly improved outcomes in patients with persistent AF in one study. Similarly, in patients with persistent AF, treatment with combined catheter ablation and vein of Marshall ethanol infusion had better outcomes compared with catheter ablation alone. Ablation of areas of rapid activity during AF and creation of empiric ablation lines to block conduction across regions of the atria have not been proven to improve outcomes in unselected patients. Other ablation targets include non–pulmonary vein foci that fire in response to high-dose isoproterenol and regions with repetitive rotational or focal activation during AF. More than one ablation procedure is often required to maintain sinus rhythm in patients with persistent and long-standing persistent AF because of lack of lesion durability and complex atrial substrate with non–pulmonary vein sources that may be incompletely treated at the initial ablation session (Table 258-3).

Catheter ablation has a 2–7% risk of major procedure-related complications, with the long-term trend suggesting steady improvement in complication rates. Complication rates are clearly lowest with high-volume operators and centers. Complications include stroke (0.5–1%), cardiac tamponade (1%), phrenic nerve paralysis, bleeding from femoral access sites, and fluid overload with heart failure, which can emerge 1–3 days after the procedure. It is important to recognize the potential for delayed presentation of some complications. Ablation within the pulmonary vein can lead to pulmonary vein stenosis, presenting weeks to months after the procedure with dyspnea or hemoptysis. The esophagus abuts the posterior wall of the left atrium where it is subject to injury, and esophageal ulcers can form immediately after the procedure and may rarely lead to a fistula between the left atrium and esophagus (estimated incidence of <0.1%) that presents as endocarditis and stroke 10 days to 3 weeks after the procedure. Early diagnosis

TABLE 258-3 Recommendations for Catheter Ablation in Patients with Atrial Fibrillation (AF)

CLASS	LEVEL OF EVIDENCE	RECOMMENDATIONS
1	A	In patients with symptomatic AF in whom antiarrhythmic drugs have been ineffective, contraindicated, not tolerated or not preferred, and continued rhythm control is desired, catheter ablation is useful to improve symptoms.
1	A	In selected patients (generally younger with few comorbidities) with symptomatic paroxysmal AF in whom rhythm control is desired, catheter ablation is useful as first-line therapy to improve symptoms and reduce progression to persistent AF.
1	A	In patients with symptomatic or clinically significant atrial flutter, catheter ablation is useful for improving symptoms.
2a	B	In patients who are undergoing ablation for AF, ablation of additional clinically significant supraventricular arrhythmias can be useful to reduce the likelihood of future arrhythmia.
2a	B	In patients (other than younger with few comorbidities) with symptomatic paroxysmal or persistent AF who are being managed with a rhythm-control strategy, catheter ablation as first-line therapy can be useful to improve symptoms.

Source: Reproduced with permission from JA Joglar et al: 2023 ACC/AHA/ACCP/HRS Guideline for the Diagnosis and Management of Atrial Fibrillation: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. J Am Coll Cardiol 83:109, 2024.

of atrioesophageal fistula is important because delayed diagnosis leads to likely death. Diagnosis is made by chest CT scan with water-soluble oral and IV contrast. Endoscopy should be avoided in patients with a suspected fistula because of the risk of air/esophageal fluid embolus. Definitive repair of the atrioesophageal fistula with emergent surgery is required. Pulsed-field ablation is associated with a significantly reduced risk of pulmonary vein stenosis, phrenic nerve injury, and left atrial esophageal fistula.

Surgical ablation of AF is most frequently performed concomitant with cardiac valve or coronary artery surgery and less commonly as a stand-alone procedure. There is no significant difference in the rate of freedom from AF between patients who undergo pulmonary vein isolation and those who undergo the biatrial maze procedure.

However, for patients with persistent AF, surgical or hybrid procedures (a combination of a surgical and catheter-based approach, most often in separate procedures) appear to have comparable efficacy to catheter ablation. Risks include sinus node injury requiring pacemaker implantation and higher morbidity with surgical ablation. Surgical removal of the left atrial appendage may reduce stroke risk, although thrombus can form in the remnant of the appendage or if the appendage is not completely ligated.

RISK FACTORS FOR AND LIFESTYLE IMPACT ON ATRIAL FIBRILLATION

There is strong evidence that AF is associated with a sedentary lifestyle, obesity, hypertension, smoking, alcohol use, and sleep apnea. Aggressive treatment of these risk factors can substantially reduce AF episodes in some patients and is warranted in all patients, as additional benefits to the patient are likely beyond AF improvement. The amount of exercise appears to have a complex relationship with the risk of AF development. In males, a U-shaped curve exists, where AF risk is high among those with sedentary lifestyles and those who participate extensively in endurance athletics such as long-distance running or cycling. Moderate exercise appears to confer a lower risk of AF. On the other hand, in females, a linear relationship exists between exercise and AF risk, with risk of AF decreasing continuously with increasing exercise activity. Although caffeine intake is often invoked as a risk for AF development or as a trigger for AF episodes in patients with a known AF diagnosis, large cohort studies have demonstrated, in contrast, a modest decrease in AF risk with modest caffeine intake. Other proposed risk factors are being evaluated, including psychological stress. Genetic predisposition to AF is seen in those with first-degree relatives with AF, and a small subset of AF patients can be determined have a familial form of AF.

There is emerging emphasis on an integrated approach to management of AF patients, with coordinated management of risk factor modification, stroke prevention, rate control, rhythm control, and management of associated comorbidities of critical importance. The previous classification of AF, which was based only on arrhythmia duration (i.e., paroxysmal, persistent, and long-standing persistent), although useful, tended to emphasize therapeutic interventions. A more recent classification using four stages (i.e., at risk of AF, pre-AF, AF, and permanent AF) has been proposed and recognizes AF as a disease continuum that requires a variety of strategies at the different stages, including prevention, lifestyle and risk factor modification, screening, and therapy.

FURTHER READING

Joglar JA et al: 2023 ACC/AHA/ACCP/HRS Guideline for the Diagnosis and Management of Atrial Fibrillation: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. J Am Coll Cardiol 83:109, 2024.

Packer DL et al: Effect of catheter ablation vs antiarrhythmic drug therapy on mortality, stroke, bleeding, and cardiac arrest among patients with atrial fibrillation: The CABANA randomized clinical trial. JAMA 321:1261, 2019.

259

Approach to Ventricular Arrhythmias

William H. Sauer, Usha B. Tedrow

There are myriad types of ventricular arrhythmias (VAs), ranging from benign to life-threatening, occurring both in patients with normal hearts and in those with structural heart disease. An understanding of an approach to these arrhythmias is critical to being appropriately parsimonious with benign forms, while understanding timely workup and management of the malignant forms.

TYPES OF VAs

VAs can arise from focal sites of origin or from reentrant circuits. Focal VAs can originate from myocardial or specialized Purkinje cells capable of automaticity or triggered activity. Reentrant VAs often involve areas of scar such as old myocardial infarction or a cardiomyopathic process. Less commonly, diseased Purkinje conduction pathways can also result in reentrant circuits. VAs are characterized by their electrocardiographic appearance and duration. Conduction away from the ventricular focus or reentrant circuit exit, propagating through the ventricular myocardium, is slower than activation of the ventricles over the normal Purkinje system. For this reason, the QRS complex duration during VAs will be wide, typically >0.12 s, though there are unusual situations that can arise with narrow QRS duration as well.

Premature ventricular beats (also referred to as a *premature ventricular contractions* or *premature ventricular complexes* [PVCs]) are single ventricular beats that occur earlier than the next anticipated supraventricular beat **(Fig. 259-1)**. PVCs that originate from the same focus will have the same QRS morphology and are referred to as unifocal **(Fig. 259-1*A*)**. PVCs that originate from different ventricular sites have different QRS morphologies and are referred to as multifocal **(Fig. 259-1*B*)**. Two consecutive ventricular beats are ventricular couplets.

Ventricular tachycardia (*VT*) is three or more consecutive beats at a rate faster than 100 beats/min. Three or more consecutive beats at

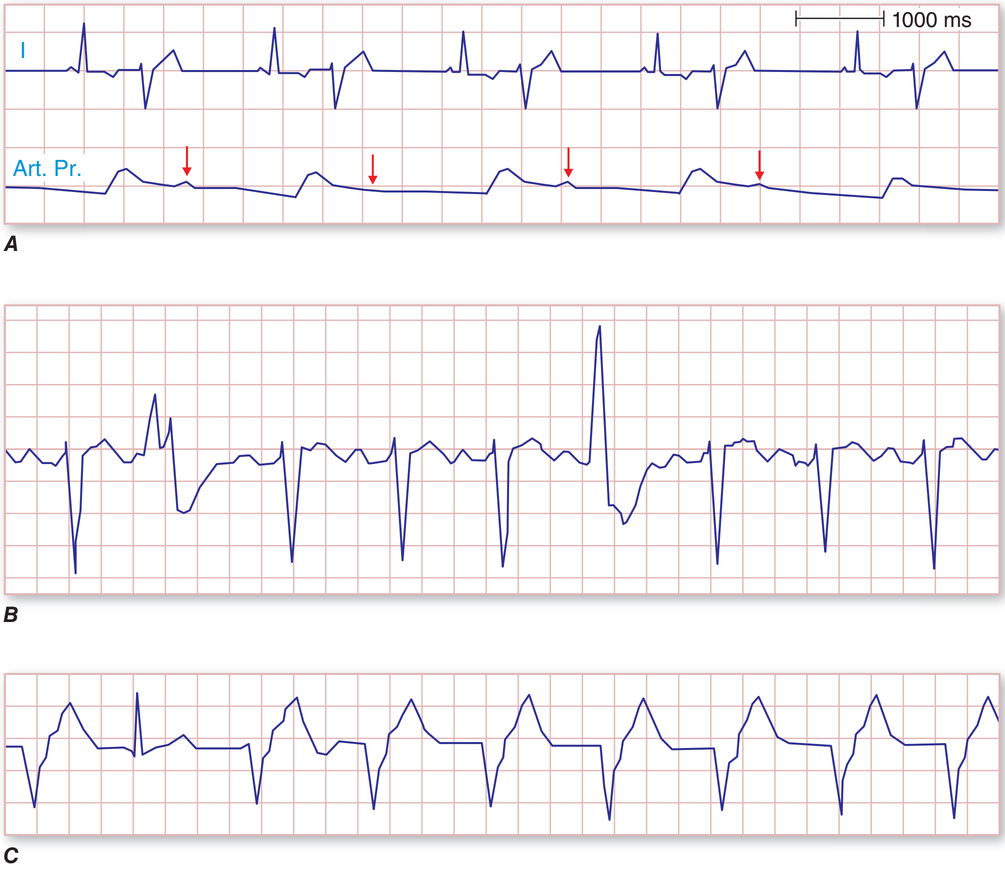

FIGURE 259-1 ***A.*** Unifocal premature ventricular contractions (PVCs) at bigeminal frequency. Trace shows electrocardiogram lead 1 and arterial pressure (Art. Pr.). Sinus rhythm beats are followed by normal arterial waveform. The arterial pressure following premature beats is attenuated (*arrows*) and imperceptible to palpation. The pulse in this patient is registered at half the heart rate. ***B.*** Multifocal PVCs. The two PVCs shown have different morphologies. ***C.*** Example of accelerated idioventricular rhythm. (See text for details.)

slower rates are designated an idioventricular rhythm. VT that terminates spontaneously within 30 s is designated *nonsustained*, whereas sustained VT persists for >30 s or is terminated by an active intervention, such as administration of an intravenous medication, external cardioversion, antitachycardia pacing, or a shock from an implanted cardioverter-defibrillator **(Fig. 259-2)**.

Monomorphic VT has the same QRS complex from beat to beat, indicating that the activation sequence is the same from beat to beat

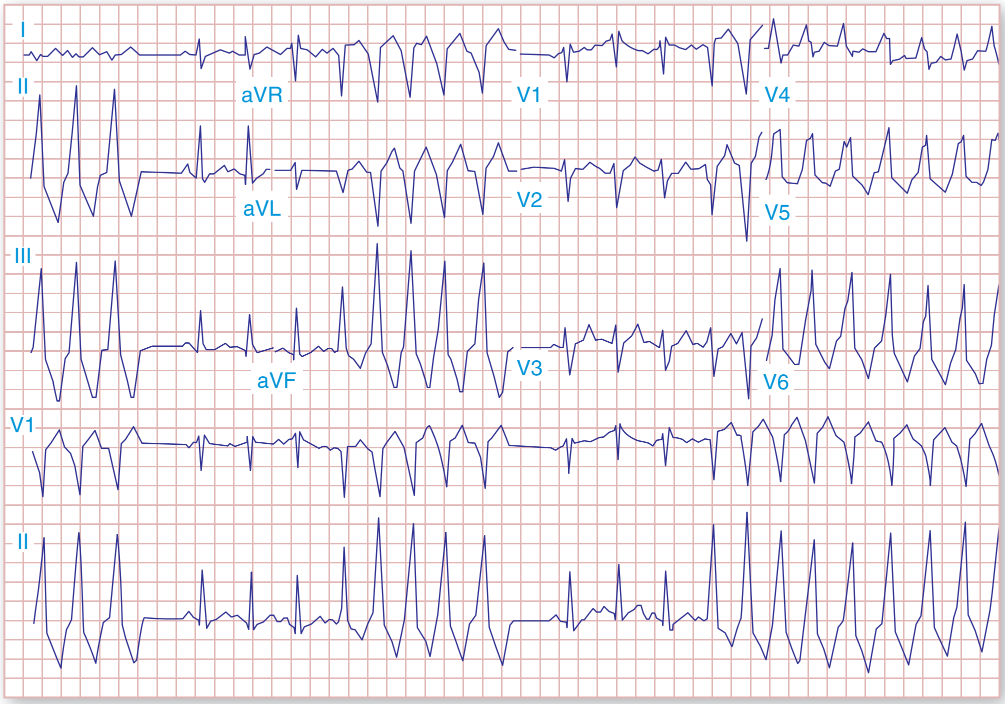

FIGURE 259-2 Repetitive monomorphic nonsustained ventricular tachycardia (VT) of right ventricular outflow tract origin. The VT has a left bundle branch block pattern with inferior axis with tall QRS complexes in the inferior leads.

and that each beat likely originates from the same source **(Fig. 259-3*A*)**. The initial site of ventricular activation largely determines the sequence of ventricular activation. Therefore, the QRS morphology of PVCs and monomorphic VT provides an indication of the site of origin within the ventricles **(Fig. 259-4)**. The likely origin often suggests whether an arrhythmia is idiopathic or associated with structural disease. Arrhythmias that originate from the right ventricle or septum result in late activation of much of the left ventricle, thereby producing a prominent S wave in V_1 referred to as a left bundle branch block–like configuration. Arrhythmias that originate from the free wall of the left ventricle have a prominent positive deflection in V_1, thereby producing a right bundle branch block–like morphology in V_1. The frontal plane axis of the QRS is also useful. An axis that is directed inferiorly, as indicated by dominant R waves in leads II, III, and AVF, suggests initial activation of the cranial portion of the ventricle, whereas a frontal plane axis that is directed superiorly (dominant S waves in II, III, and AVF) suggests initial activation at the inferior wall.

Very rapid monomorphic VT has a sinusoidal appearance, also called *ventricular flutter*, because it is not possible to distinguish the QRS complex from the T wave **(Fig. 259-3*B*)**. Relatively slow *sinusoidal* VTs have a wide QRS indicative of slowed ventricular conduction **(Fig. 259-3*C*)**. Hyperkalemia, toxicity from excessive effects of drugs that block sodium channels (e.g., flecainide, propafenone, or tricyclic antidepressants), and severe global myocardial ischemia are possible causes.

Polymorphic VT has a continually changing QRS morphology indicating a changing ventricular activation sequence. Polymorphic VT that occurs in the context of congenital or acquired prolongation of the QT interval often has a waxing and waning QRS amplitude, creating a characteristic shifting axis referred to as torsades des pointes after the classic ballet sequence **(Fig. 259-3*D*)**.

Ventricular fibrillation (VF) has continuous irregular activation with no discrete QRS complexes. Monomorphic or polymorphic VT may transition to VF in susceptible patients. Cardiac ischemia is the most common cause of VF **(Fig. 259-3*E*)**.

The term *idiopathic ventricular arrhythmia* generally refers to PVCs or VT that occurs in patients with a normal electrocardiogram (ECG), without structural heart disease, and not associated with an underlying genetic syndrome or risk of sudden death.

CLINICAL MANIFESTATIONS

Common symptoms of VAs include palpitations, dizziness, exercise intolerance, episodes of lightheadedness, syncope, or sudden cardiac arrest leading to sudden death if not resuscitated. VAs can also be asymptomatic and encountered unexpectedly as an irregular pulse or heart sounds on examination or may be seen on a routine ECG, exercise test, or cardiac ECG monitoring. Occasionally when every other beat is a PVC (*bigeminy*), pulse measurements for heart rate can be erroneously low (*pseudobradycardia*) because the PVCs may not generate a separate pulse wave.

Syncope is a concerning symptom, particularly when occurring without prodrome, during exercise, or in the setting of abnormal ECG or structural heart disease. Such episodes can be due to VT that produces severe hypotension, warranting concern for risk of cardiac arrest and sudden death with arrhythmia recurrence. Although benign processes such as reflex-mediated neurocardiogenic (vasovagal) episodes and orthostatic hypotension are the most common causes of syncope, it is important to consider the possibility of underlying heart disease or a genetic syndrome causing VT. When these are suspected, hospitalization for further evaluation and monitoring is often appropriate.

Sustained VT may present as a wide QRS complex tachycardia that must be distinguished from supraventricular tachycardia with aberrancy **(Chap. 253)**. Symptoms can be minor but more commonly include hypotension with syncope and even imminent cardiac arrest, particularly in patients with structural heart disease. Sustained VT may degenerate to VF, most commonly if it is rapid and polymorphic. Many patients who are at risk for VT have known heart disease, and many have an implantable cardioverter-defibrillator (ICD). In patients with an ICD, VT episodes may cause transient lightheadedness, palpitations, or syncope followed by a shock from the ICD (see below).

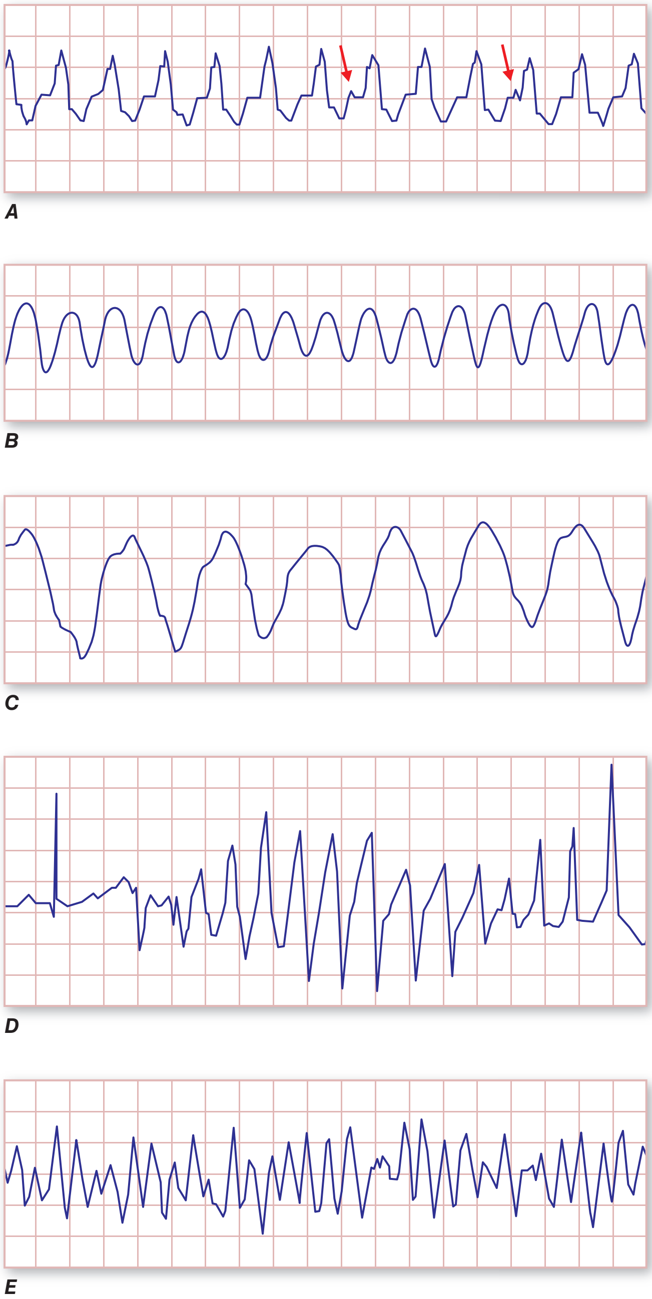

FIGURE 259-3 ***A.*** Monomorphic ventricular tachycardia (VT) with dissociated P waves (*short arrows*). ***B.*** Ventricular flutter. ***C.*** Sinusoidal VT due to electrolyte disturbance or drug effects. ***D.*** Polymorphic VT resulting from prolongation of QT interval (torsade de pointe VT). ***E.*** Ventricular fibrillation. (See text for details.)

EVALUATION OF PATIENTS WITH DOCUMENTED OR SUSPECTED VENTRICULAR ARRHYTHMIAS

There are several important considerations that guide evaluation of patients with documented or suspected cardiac arrhythmias. First, establish whether a VA is the cause of the symptoms or clinical presentation. Second, determine whether the arrhythmia is associated with a cardiac disease, and establish the prognostic significance of that disease and, in particular, whether it is associated with a risk of sudden cardiac death. Finally, define the likelihood of arrhythmia recurrence and the symptoms and risk imposed by the recurrence. The risks of cardiac arrest and sudden cardiac death are largely determined by the cause of the arrhythmia and the associated underlying heart disease.

The diagnosis of VAs can be established by recording the arrhythmia on an ECG, by an ambulatory or implanted cardiac monitor, by an implanted rhythm management device such as a pacemaker or ICD, or in some cases, initiation of the arrhythmia during an electrophysiologic

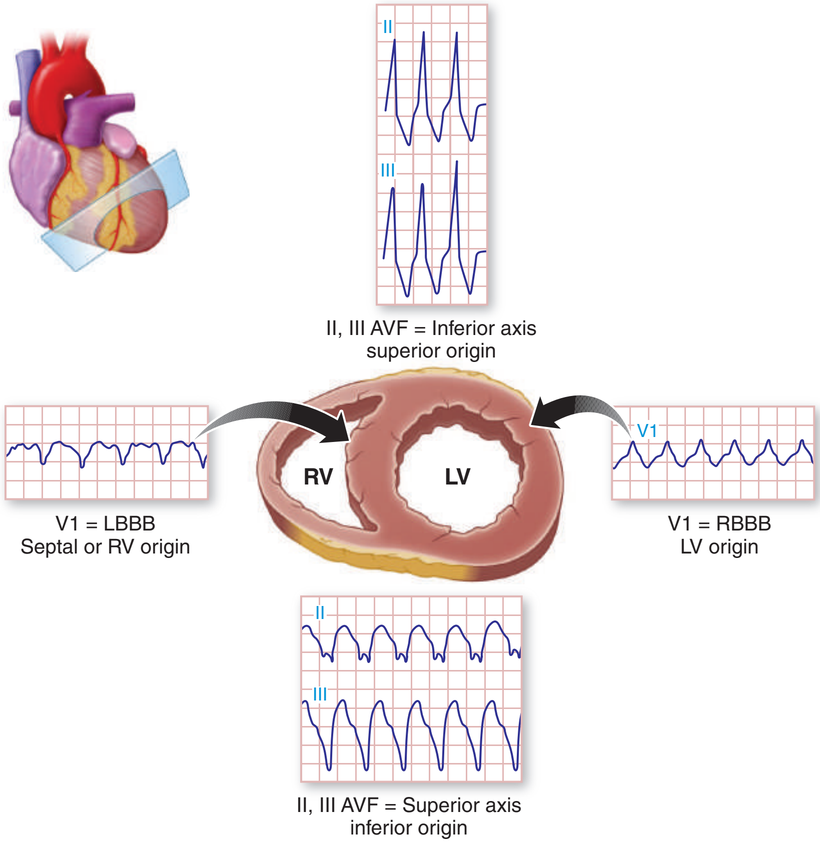

FIGURE 259-4 Site of ventricular tachycardia origin based on QRS morphology. (See text for details.) LBBB, left bundle branch block; LV, left ventricle; RBBB, right bundle branch block; RV, right ventricle.

study. A 12-lead ECG of the arrhythmia should be obtained when possible and often provides clues to the potential site of origin and possible presence of underlying heart disease (see above) (Fig. 259-4).

For patients with sustained wide-complex tachycardia, initial management is guided by the patient's hemodynamic stability. The approach to sustained wide-complex tachycardia is discussed in **Chap. 261**. The management of VT that causes cardiac arrest is discussed in **Chap. 317**. Once hemodynamic stability is restored, further management is guided by the possibility of a recurrence and the risk imposed by a recurrence.

EVALUATION OF THE PATIENT WITH ARRHYTHMIA SYMPTOMS

When symptoms are intermittent, initial evaluation aims to establish symptom severity, provocative factors, and presence of underlying heart disease. Syncope or near syncope raises concern that an arrhythmia is causing episodes of hypotension and that there may be a risk of cardiac arrest. Symptoms that occur with exertion suggest arrhythmias that are provoked by sympathetic stimulation but can also be related to exertional ischemia in patients with coronary artery disease, although nonarrhythmia causes must also be considered. A past history of any cardiac disease is important. A review of all medications is relevant. Medications that prolong the QT interval predispose to polymorphic VT **(Chap. 262)**. Adrenergic stimulants can provoke PVCs.

Family history should determine the presence of premature coronary artery disease, cardiomyopathy, or cardiac arrhythmias, particularly a history of sudden death. Family history may also suggest that the possibility of a genetic cause of an arrhythmia warrants careful consideration. Details of premature deaths are relevant. Sudden death victims are often said to have died of a "massive heart attack" despite absence of definite confirmation of thrombotic myocardial infarction and when other causes such as arrhythmia may have been possible.

The physical examination focuses on evidence of structural heart disease with assessment of pulse, jugular venous pressure lung fields, and cardiac auscultation. Stigmata of neuromuscular disease or dysmorphic features may suggest a genetic arrhythmia syndrome.

A 12-lead ECG should be obtained even if the patient is not having symptoms at the time of evaluation. Occasionally, premature ventricular contractions will be detected. Patients with benign idiopathic arrhythmias usually have a completely normal ECG during sinus rhythm. Any ECG abnormality warrants further evaluation. Particularly relevant findings include Q waves that indicate prior myocardial infarction, which may have been silent, and ventricular hypertrophy, which may indicate hypertrophic cardiomyopathy or other ventricular disease. An ECG finding is the major diagnostic manifestation of several genetic arrhythmia syndromes in patients without structural heart disease, including the long QT syndrome, Brugada syndrome, and short QT syndrome.

If there is suspicion of structural heart disease, cardiac imaging is warranted to assess ventricular function and structure. Transthoracic echocardiography is most frequently employed for initial evaluation. Depressed ventricular function increases concern for a risk of sudden death and warrants further evaluation to establish the cause, which may be cardiomyopathy, coronary artery disease, or valvular heart disease. Ventricular thickening may indicate hypertrophic cardiomyopathy or infiltrative diseases such as amyloidosis. Cardiac magnetic resonance imaging with gadolinium contrast imaging provides similar assessment but also can detect areas of ventricular scar, evident as regions of delayed hyperenhancement, which are usually present in patients who have sustained monomorphic VT **(Fig. 259-5)**. The nature and location of abnormalities are helpful in assessing the type of heart disease. Evaluation to exclude atherosclerotic coronary artery disease should be performed in patients at risk, guided by age and other risk factors.

TREATMENT OPTIONS FOR VENTRICULAR ARRHYTHMIAS

Treatment of VAs is guided by the severity and frequency of symptoms. For those with structurally normal hearts and normal ECGs, reassurance and removal of aggravating factors (e.g., caffeine or alcohol) may be all that is needed. For arrhythmias associated with a sudden death risk, ICD implantation is usually indicated and will provide a "safety net" to terminate life-threatening VT or VF, preventing sudden death but without preventing the arrhythmia. When suppression of the arrhythmia is required, antiarrhythmic drug therapy or catheter ablation is a major consideration.

ANTIARRHYTHMIC DRUGS

Use of antiarrhythmic drugs is based on consideration of the risks and potential benefit for the individual patient. Efficacy and side effects for the individual patient are not always predictable and are assessed by individual therapeutic trial. Causes of drug intolerance are mostly noncardiac and minor but can sometimes be severe enough to limit their use. Cardiac adverse events, however, include the potential for "proarrhythmia," whereby a drug can increase the frequency of arrhythmia or cause a new arrhythmia. Aggravation of bradyarrhythmias is also a common concern. Although antiarrhythmic drugs are classified based on their actions on receptors or ion channels, most have multiple effects, affecting more than one channel.

β-ADRENERGIC BLOCKERS

Many VAs are sensitive to sympathetic stimulation, and β-adrenergic stimulation may also interact with the electrophysiologic effects of many membrane-active antiarrhythmic drugs. The safety of β-blocking agents makes them the first choice of therapy for most VAs. β-Blockers are particularly useful for exercise-induced arrhythmias and idiopathic arrhythmias but have limited efficacy for most arrhythmias associated

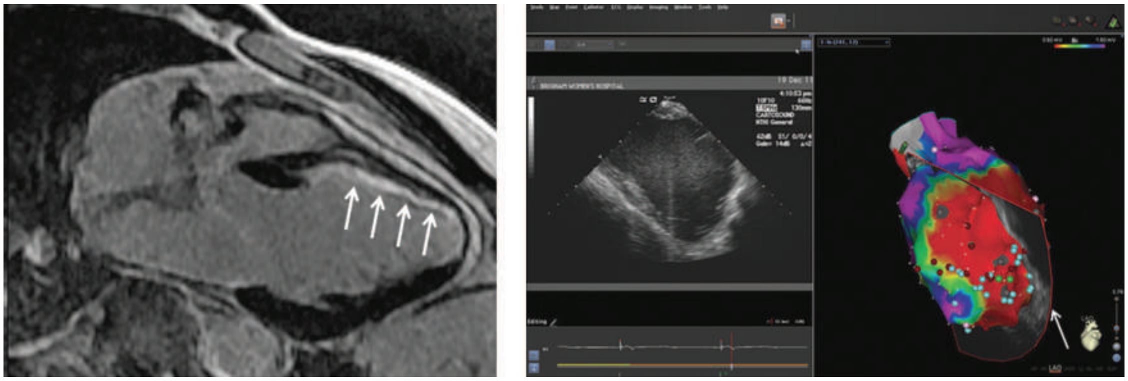

FIGURE 259-5 Imaging studies of the left ventricle (LV) used to assist ablation for ventricular tachycardia (VT). *Left panel* is a magnetic resonance image of a longitudinal section demonstrating thinning of the anterior wall and late gadolinium enhancement in a subendocardial scar (*white arrows*). The *middle panel* shows a two-dimensional image of the LV in long axis corresponding to the sector through the mid-LV (*arrow in figure on right panel*) obtained by an intracardiac echocardiography probe positioned in the right ventricle. An electroanatomic three-dimensional map of the LV in the left anterior oblique projection is displayed in the *right panel*. The purple areas depict areas of normal voltage (>1.5 mV). Blue, green, and yellow represent progressively lower voltages, with the red areas indicating scar (<0.5 mV). Channels of viable myocardium with slow conduction within the scar are identified with the light blue dots. Areas of ablation delivered to regions involved in reentrant VT are indicated by maroon dots.

with heart disease. Bradyarrhythmias and negative inotropic effects are the major cardiac adverse effects.

CALCIUM CHANNEL BLOCKERS

The nondihydropyridine calcium channel blockers diltiazem and verapamil can be effective for some idiopathic VTs. The risk of proarrhythmia is low, but they have negative inotropic and vasodilatory effects that can aggravate hypotension.

SODIUM CHANNEL–BLOCKING AGENTS

Drugs whose major effect is mediated through sodium channel blockade include mexiletine, quinidine, disopyramide, flecainide, and propafenone, which are available for chronic oral therapy. Blockade of the fast inward sodium current has been referred to as a class I antiarrhythmic drug effect. Antiarrhythmic actions are the result of depressing cardiac conduction and membrane excitability. Conduction slowing can be manifest as a prolongation of QRS duration. Lidocaine, quinidine, and procainamide are available as intravenous formulations. Quinidine, disopyramide, and procainamide also have potassium channel–blocking effects that may prolong the QT interval (class III antiarrhythmic drug action), contributing to their antiarrhythmic effect. Quinidine also blocks a particular potassium current, I_{to}, the blockade of which can be important in Brugada syndrome. These agents have potential proarrhythmic effects and, with the possible exception of quinidine, also have negative inotropic effects that may have contributed to the increased mortality observed when some were administered chronically to patients with prior myocardial infarction. Long-term therapy is generally avoided in patients with structural heart disease but may be used to reduce symptomatic arrhythmias in patients with ICDs.

POTASSIUM CHANNEL BLOCKING AGENTS

Sotalol and dofetilide block the delayed rectifier potassium channel I_{Kr}, thereby prolonging action potential duration (QT interval) and the cardiac refractory period, known as the class III antiarrhythmic drug effect. Sotalol also has nonselective β-adrenergic–blocking activity. It has been shown to have a modest effect on reducing ICD shocks due to ventricular and atrial arrhythmias. Proarrhythmia due to the polymorphic VT torsade de pointe that is associated with QT prolongation occurs in 3–5% of patients. Both sotalol and dofetilide are excreted via the kidneys, necessitating dose adjustment or avoidance in renal insufficiency. These drugs must be avoided in patients with other risk factors for torsades de pointes, including QT prolongation, other administered medications that prolong the QT interval, hypokalemia, and significant bradycardia.

AMIODARONE

Amiodarone blocks multiple cardiac ionic currents and has sympatholytic activity. It is the most effective antiarrhythmic drug for suppressing VAs. It is administered intravenously for life-threatening arrhythmias. During chronic oral therapy, electrophysiologic effects develop over several days. It is more effective than sotalol in reducing ICD shocks and is often used for VAs in patients with heart disease. Bradyarrhythmias are the major cardiac adverse effect. Ventricular proarrhythmia can occur, but torsades de pointes VT is rare. Noncardiac toxicities are a major problem and contribute to drug discontinuation in at least a third of patients during long-term therapy. Hyper- and hypothyroidism are related to the iodine content of the drug. Pneumonitis or pulmonary fibrosis occurs in ~1% of patients. Photosensitivity is common, and neuropathy and ocular toxicity can occur. Systematic monitoring is recommended during chronic therapy including assessment for thyroid, liver, and pulmonary toxicity. Intravenous administration of amiodarone via a peripheral vein for >24 h can cause severe peripheral thrombophlebitis. Dronedarone has structural similarities to amiodarone but without the iodine moiety. Efficacy for VAs is poor, and dronedarone increases mortality in patients with heart failure, so dronedarone is not typically used for treatment of VAs.

IMPLANTABLE CARDIOVERTER-DEFIBRILLATORS

ICDs detect sustained VT, largely based on heart rate, and then terminate the arrhythmia. In transvenous devices, VF is terminated by a shock applied between a lead in the right ventricle and the ICD pulse generator. The lead can provide pacing for bradycardia if needed. This transvenous form of ICD has the disadvantages of vascular occlusion, risk of lead fracture, endocarditis in the event of infection, and difficulty with removal. Monomorphic VT can also be terminated by a burst of rapid pacing faster than the VT, known as antitachycardia pacing (ATP) **(Fig. 259-6*A*)**. If ATP fails or is not a programmed treatment, as is often the case for rapid VT or VF, a shock is delivered **(Fig. 259-6*B*)**. ICDs can also be subcutaneous or extravascular, without a transvenous lead. The rhythm is also sensed by this lead, in a manner similar to a surface ECG. The lead is placed overlying the left chest with a coil parallel and next to or underneath the sternum. No matter the type of ICD, shocks are uncomfortable and distressing if the patient is conscious, sometimes leading to a posttraumatic stress disorder (PTSD). The most common ICD complication is the delivery of unnecessary therapy (either ATP or shocks) in response to an inappropriately detected rapid supraventricular tachycardia or electrical noise as a result of an ICD lead fracture or electromagnetic interference from an external source. ICDs record and store electrograms from arrhythmia episodes that can be retrieved by interrogation of the ICD, which can be performed remotely and communicated via the internet. This assessment is critical after an ICD shock to determine the arrhythmia diagnosis and exclude unnecessary therapy. Device infection is an important problem long term and occurs in ~1% of patients. This risk may be less for subcutaneous or extravascular implants.

ICDs decrease mortality in patients at risk for sudden death due to structural heart diseases. In all cases, ICDs are recommended only if there is also an expectation for survival of at least a year with acceptable

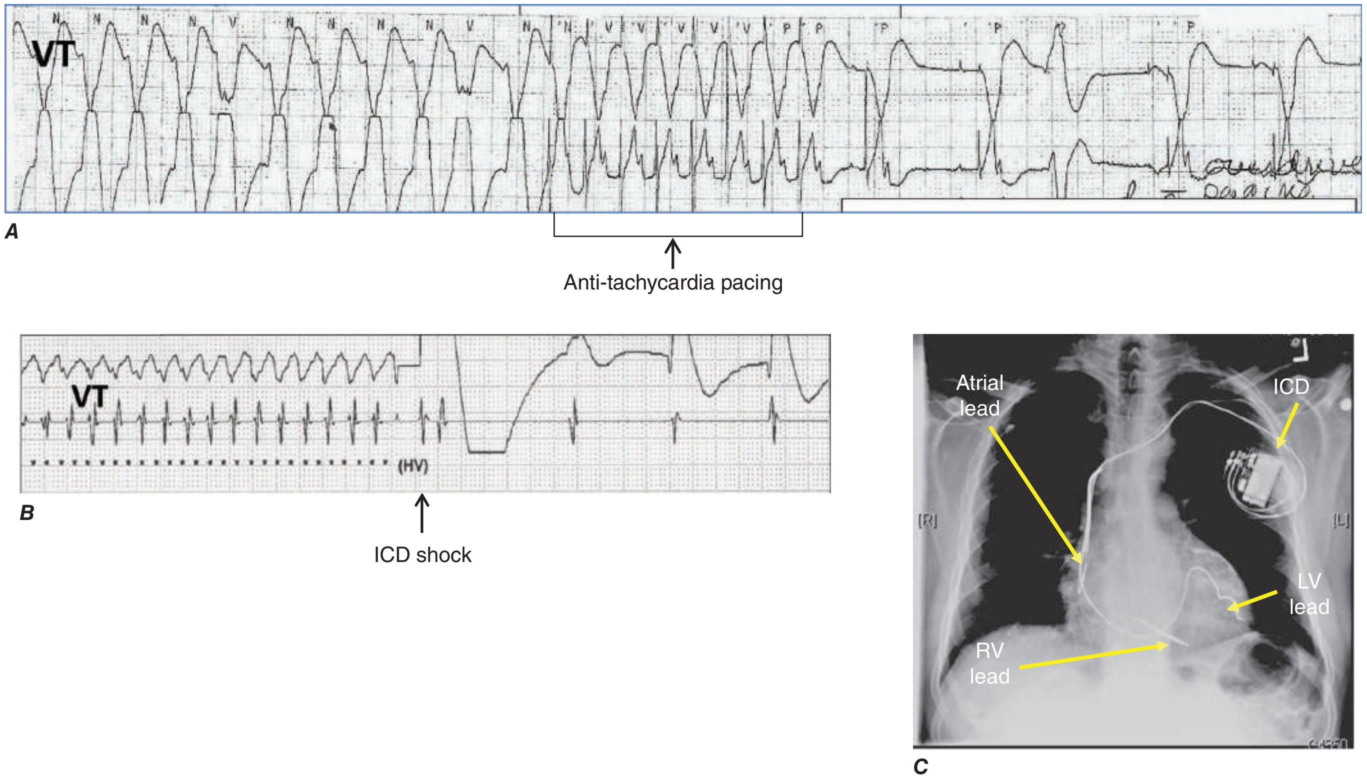

FIGURE 259-6 Implantable cardioverter-defibrillator (ICD) and therapies for ventricular arrhythmias. ***A.*** A monomorphic ventricular tachycardia (VT) is terminated by a burst of pacing impulses at a rate faster than VT (antitachycardia pacing). ***B.*** A rapid VT is converted with a high-voltage shock (*arrow*). The chest x-ray in the panel ***C*** shows the components of an ICD capable of biventricular pacing (*yellow arrows*). ICD generator in the subcutaneous tissue of the left upper chest, pacing leads in the right atrium and the left ventricular (LV) branch of the coronary sinus (LV lead), and a pacing/defibrillating lead in the right ventricle (RV lead) are shown.

functional capacity. The exception is in cases of patients with end-stage heart disease who are awaiting cardiac transplantation outside the hospital or who have left bundle branch block QRS prolongation such that they are likely to have improvement in ventricular function with cardiac resynchronization therapy from a biventricular ICD or conduction system pacing **(Fig. 259-6*C*)**. In these cases, an ICD may be warranted despite a guarded prognosis. A wearable ICD system with electrodes incorporated into a vest and an external battery pack is also available for short-term use in patients pending a decision regarding a permanent implanted system, or when a permanent system cannot be implanted for other reasons such as an ongoing infection.

Despite prompt termination of VT or VF by an ICD, the occurrence of these arrhythmias predicts subsequent increased mortality and risk of heart failure. Occurrence of VT or VF should therefore prompt assessment for potential causes including worsening heart failure, electrolyte abnormalities, and ischemia. Repeated shocks, even if appropriate, often induce posttraumatic stress disorder. Antiarrhythmic drug therapy, most commonly amiodarone, or catheter ablation is often required for suppression of recurrent arrhythmias. Antiarrhythmic drug therapy can alter the VT rate and the energy required for defibrillation, thereby necessitating programming changes in the ICD's algorithms for detection and therapy.

■ CATHETER ABLATION FOR VT

Catheter ablation is usually performed by applying radiofrequency (RF) current to cause thermal injury by resistive heating of cardiac tissue responsible for the arrhythmia. An electrode catheter with an electroanatomic mapping system is used to map local electrical activity to identify the ventricular myocardium that is causing the arrhythmia, referred to as the arrhythmia substrate. The size and location of the arrhythmia substrate determine the ease and likely effectiveness of the procedure, as well as the potential complications. When the arrhythmia originates from the endocardium, as is most commonly the case, it can be reached from an endovascular approach via a femoral vein or artery. Less commonly, arrhythmias originate from the subepicardium, and percutaneous pericardial puncture, similar to pericardiocentesis, is required to insert a catheter into the pericardial space for mapping and ablation. In patients with scar-related VT due to prior infarction or cardiomyopathy, ablation typically targets abnormal regions in the scar or region of fibrosis. Because these scars often contain multiple reentry circuits over relatively large regions, extensive areas of ablation can be required, and these areas are often identified as regions of low voltage displayed on electroanatomic reconstructions of the ventricle (Fig. 259-5).

Catheter ablation is often performed in patients with recurrent VAs associated with poor cardiac function, and the procedure-related mortality in this situation is 0.5–3%. Outcomes are better for patients with prior infarction and VT than for patients with nonischemic cardiomyopathies in which the scar locations are more variable and often intramural or subepicardial. Ablation can be lifesaving for patients with very frequent or incessant VT. Methods of delivering ablative energy to intramural areas or areas requiring very extensive ablation are under development. These include needle catheters capable of delivering ablative energy into intramural sources, and bipolar ablation where radiofrequency energy is delivered across two ablation catheters. Stereotactic body radiation therapy (SBRT), classically used for treating thoracic tumors, has been used to direct radiation therapy to a specific portion of the scar substrate to noninvasively ablate VT with encouraging early studies.

Idiopathic VTs and PVCs that occur in the absence of structural heart disease usually originate from a small focus, for which catheter ablation typically has a higher success rate for preventing recurrent arrhythmia. Long-term arrhythmia-free survival in these patients is excellent.

ARRHYTHMIA SURGERY

When antiarrhythmic drug therapy and catheter ablation fail or are not an option, surgical cryoablation, often combined with aneurysmectomy, can be effective therapy for recurrent VT due to prior myocardial infarction and has also been used successfully in a few patients with nonischemic heart disease. Few centers now maintain the expertise for this therapy, though some use this therapy as an adjunct to ventricular assist device implantation.

FURTHER READING

Al-Khatib SM et al: 2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. Heart Rhythm 15:e73, 2018.

Deyell MW et al: 2020 Canadian Cardiovascular Society/Canadian Heart Rhythm Society position statement on the management of ventricular tachycardia and fibrillation in patients with structural heart disease. Can J Cardiol 36:822, 2020.

Jalife J, Stevenson W (eds): *Zipes and Jalife's Cardiac Electrophysiology: From Cell to Bedside*, 8th ed. Philadelphia, Elsevier, 2021.

Zeppenfeld K et al: 2022 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: Developed by the task force for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death of the European Society of Cardiology (ESC) Endorsed by the Association for European Paediatric and Congenital Cardiology (AEPC). Eur Heart J 43:3997, 2022.

260

Premature Ventricular Contractions, Nonsustained Ventricular Tachycardia, and Accelerated Idioventricular Rhythm

William H. Sauer, Usha B. Tedrow

Ventricular ectopic beats are very common and may be identified during outpatient or inpatient telemetry monitoring either due to symptoms of palpitations or as an incidental finding. In most cases, ventricular ectopy, presenting as premature ventricular contractions (PVCs), nonsustained ventricular tachycardia (NSVT), and accelerated idioventricular rhythm (AIVR), is asymptomatic and does not require specific treatment. While most commonly benign when presenting in patients with structurally normal hearts and normal electrocardiograms (ECGs), these ventricular arrhythmias can rarely be associated with structural heart disease and a risk of sudden death.

PREMATURE VENTRICULAR CONTRACTIONS AND NONSUSTAINED VENTRICULAR TACHYCARDIA

PVCs are very common and can be due to enhanced automaticity, triggered automaticity, or localized reentry. PVCs are often sensitive to sympathetic stimulation and can be a sign of increased sympathetic tone, myocardial ischemia, hypoxia, electrolyte abnormalities, or underlying heart disease. During myocardial ischemia or in association with other structural heart disease, PVCs can be a harbinger of sustained ventricular tachycardia (VT) or ventricular fibrillation (VF).

The ECG characteristics of the arrhythmia are often suggestive of whether structural heart disease is present. PVCs with smooth uninterrupted contours and sharp QRS deflections may suggest an ectopic focus in relatively normal myocardium, whereas broad notching and slurred QRS deflections may suggest a diseased myocardial substrate. The QRS morphology also suggests the likely site of origin within the ventricle. PVCs that have a dominant S wave in V_1, referred to as a left bundle branch block–like configuration, originate from the right ventricle or interventricular septum. Those with a dominant R wave in V_1 originate from the left ventricle. A superior frontal plane axis (negative in II, III, aVF) indicates initial depolarization of the inferior wall (diaphragmatic aspect of the heart), whereas an inferior frontal plane axis (positive in II, III, aVF) indicates an origin in the cranial aspect of the heart. The location of arrhythmia origin often suggests the nature of underlying heart disease. Most common ventricular arrhythmias that are not associated with structural heart disease have a left bundle branch block–like configuration. PVCs with a right bundle branch block configuration are more likely to be associated with structural heart disease. Multiple morphologies of PVCs (multifocal PVCs) are also more likely to indicate structural heart disease or a myopathic disease process. In patients with heart disease, greater frequency and complexity (couplets and NSVT) of these arrhythmias are associated with more severe disease.

PVCs AND NSVT DURING ACUTE ILLNESS

These arrhythmias are often encountered in patients who are being evaluated in the emergency department or who have been hospitalized and are on an ambulatory cardiac monitor. When encountered during acute illness or as a new finding, evaluation should focus on detection and correction of potential aggravating factors and causes, specifically myocardial ischemia, ventricular dysfunction, and electrolyte abnormalities, most commonly hypokalemia. If there is a suspicion of underlying heart disease, then this should be evaluated. Otherwise, asymptomatic PVCs and NSVT in the hospitalized patient do not indicate any specific treatment outside the patient's presenting illness.

PVCs AND NSVT IN PATIENTS WITHOUT HEART DISEASE

Idiopathic ventricular arrhythmias frequently originate from the left or right ventricular outflow tracts near the valve annuli, giving rise to PVCs or VT that have a left bundle branch block–like configuration, with an inferiorly directed frontal plane axis, as discussed below. Other regions that give rise to PVCs in normal hearts include the papillary muscles and fascicular tissue. NSVT from a benign idiopathic source is usually monomorphic, with rates <200 beats/min. NSVT that is very rapid, polymorphic, or with a first beat that occurs prior to the peak of the T wave ("short-coupled") is uncommon and should prompt concern and careful evaluation for underlying disease or genetic syndromes associated with sudden death.

A family history of sudden death should prompt evaluation for genetic syndromes associated with sudden death, including cardiomyopathy, long QT syndrome, and arrhythmogenic right ventricular cardiomyopathy (ARVC) (see below). Any abnormality on the 12-lead ECG warrants further evaluation. Repolarization abnormalities are seen in a number of genetically determined syndromes associated with sudden death, including the long QT syndrome, Brugada syndrome, ARVC, and hypertrophic cardiomyopathy **(Fig. 260-1)**. Structural abnormalities such as mitral valve prolapse and mitral annular disjunction can be associated with papillary muscle PVCs and sudden death. An echocardiogram is often necessary to assess ventricular function, wall motion abnormalities, and valvular heart disease. Contrast-enhanced cardiac magnetic resonance imaging (MRI) is also useful for this purpose and for the detection of ventricular scarring that is the substrate for sustained VT. Exercise stress testing should be performed in patients with effort-related symptoms and for those at risk for coronary artery disease.

TREATMENT OF IDIOPATHIC ARRHYTHMIAS

For PVCs and NSVT in the absence of structural heart disease or a genetic sudden death syndrome, no specific therapy is needed unless the patient has significant symptoms or evidence that frequent PVCs are depressing ventricular function (see below). Reassurance that the arrhythmia is benign is often sufficient to allow the patient to cope with the symptoms, which will often wax and wane in frequency over years. Avoiding stimulants, such as caffeine and alcohol, is helpful in some patients. If symptoms require treatment, β-adrenergic blockers and nondihydropyridine calcium channel blockers (verapamil and diltiazem) are sometimes helpful. If these fail, more membrane active antiarrhythmic drugs and catheter ablation are options. The antiarrhythmic agents flecainide, propafenone, mexiletine, and amiodarone can be effective, but the potential for side effects warrants careful consideration prior to prescribing these agents for long-term use. Catheter ablation is effective at suppressing this arrhythmia in ~90% of patients. Failure of ablation is usually due to inability to provoke the arrhythmia for mapping in the electrophysiology laboratory or if the site of origin is near a vital structure, such as the coronary arteries or His-Purkinje system, or is not accessible due to a site of origin deep within the myocardium.

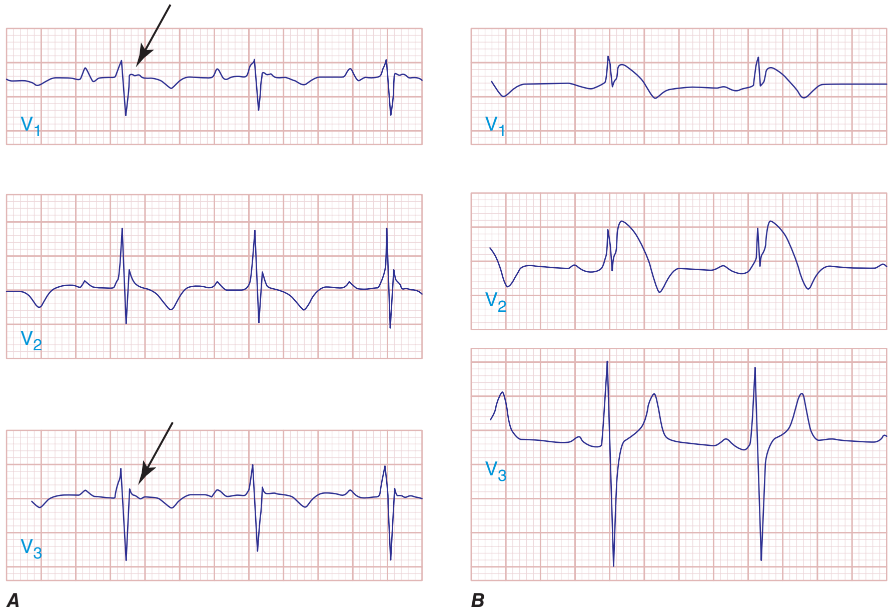

FIGURE 260-1 Electrocardiogram leads depicting Brugada syndrome. Precordial chest leads V_1–V_3 showing typical abnormalities of arrhythmogenic right ventricular cardiomyopathy (ARVC) (***A***) and Brugada syndrome (**B**). In ARVC, there is T inversion and delayed ventricular activation manifest as epsilon waves (*arrows*). Panel ***B*** shows ST elevation in V_1 and V_2 typical of the Brugada syndrome.

PVCs AND NSVT ASSOCIATED WITH ACUTE CORONARY SYNDROMES

In the peri-infarct period, PVCs and NSVT are common and can be an early manifestation of ischemia and a harbinger of subsequent VF. Treatment with β-adrenergic blockers and correction of hypokalemia and hypomagnesemia reduce the risk of VF. Routine administration of antiarrhythmic drugs such as lidocaine or amiodarone does not reduce mortality and is not indicated for suppression of PVCs or asymptomatic NSVT but may be implemented transiently if an episode of sustained VT or VF occurs, with the goal of reducing the likelihood of a subsequent episode.

Following recovery from acute myocardial infarction (MI), frequent PVCs (typically >10 PVCs/h), repetitive PVCs with couplets, and NSVT are markers for depressed ventricular function and increased mortality, but routine antiarrhythmic drug therapy to suppress these arrhythmias has not been shown to improve mortality. Therefore, amiodarone is an option for treatment of symptomatic arrhythmias in this population when the potential benefit outweighs its potential toxicities. β-Adrenergic blockers reduce sudden death but have limited effect on spontaneous arrhythmias.

For survivors of an acute MI, an implantable cardioverter-defibrillator (ICD) reduces mortality in certain high-risk groups: patients who have survived >40 days after the acute MI and have a left ventricular ejection fraction of <30%, or who have an ejection fraction <35% and have symptomatic heart failure (functional class II or III); and patients >5 days after MI who have a reduced left ventricular ejection fraction, NSVT, and inducible sustained VT or VF on electrophysiologic testing. ICDs do not reduce total mortality when routinely implanted early after MI and have not been demonstrated to improve mortality when implanted early after coronary artery revascularization.

PVCs AND NSVT ASSOCIATED WITH DEPRESSED VENTRICULAR FUNCTION AND HEART FAILURE

PVCs and NSVT are common in patients with depressed ventricular function and heart failure and are markers for disease severity and increased mortality, but antiarrhythmic drug therapy to suppress these arrhythmias has not been shown to improve survival. The use of antiarrhythmic drugs whose major action is blockade of the cardiac sodium channel (flecainide, propafenone, mexiletine, quinidine, and disopyramide) is avoided in patients with structural heart disease because of a risk of proarrhythmia and negative inotropic effects. Amiodarone suppresses ventricular ectopy and reduces sudden death but does not improve overall survival. ICDs are the major therapy to protect against sudden death in patients at high risk and are recommended for those with a left ventricular ejection fraction <35% and New York Heart Association class II or III heart failure, in whom they reduce mortality from 36 to 29% over 5 years.

PVC AND NSVT ASSOCIATED WITH OTHER CARDIAC DISEASES

Ventricular ectopy is associated with increased mortality in patients with hypertrophic cardiomyopathy or with congenital heart disease associated with right or left ventricular dysfunction. In these patients, management is similar to that for patients with ventricular dysfunction. Pharmacologic suppression of the arrhythmia has not been shown to improve mortality. ICDs are indicated for patients considered at high risk for sudden cardiac death.

PVC-INDUCED VENTRICULAR DYSFUNCTION

Very frequent ventricular ectopy and repetitive NSVT can depress ventricular function, possibly through an effect similar to chronic tachycardia or by inducing ventricular dyssynchrony. Depression of ventricular function rarely occurs unless PVCs account for at least 10–20% of total beats over a 24-h period, and only a minority of patients with PVCs will have a reversible cardiomyopathy. Often the PVCs are idiopathic and unifocal, most commonly originating from the outflow tract regions or left ventricular papillary muscles **(Fig. 260-2)**, where they can be targeted for ablation.

Other sites of origin such as the mitral and tricuspid valve annuli, right ventricular moderator band, and even the epicardial surface of the heart also occur **(Fig. 260-3)**. The factors that can potentially predict development of heart failure and increased risk of adverse outcomes include PVC frequency, characteristics of the PVC morphology, and timing of the PVC coupling interval. In addition, the presence of late gadolinium enhancement on cardiac MRI may suggest the presence of an additional underlying cardiomyopathic process. The degree of expected recovery of ventricular function with PVC suppression is difficult to predict. Even in the setting of known underlying cardiomyopathy, controlling frequent ventricular ectopy can be helpful to improve ejection fraction and improve other factors such as delivery of resynchronization pacing.

ACCELERATED IDIOVENTRICULAR RHYTHMS

Three or more ventricular beats at a rate slower than 100 beats/min are termed an AIVR **(Fig. 260-4)**. Automaticity is the likely mechanism, although in some rare cases, a reentrant circuit utilizing diseased

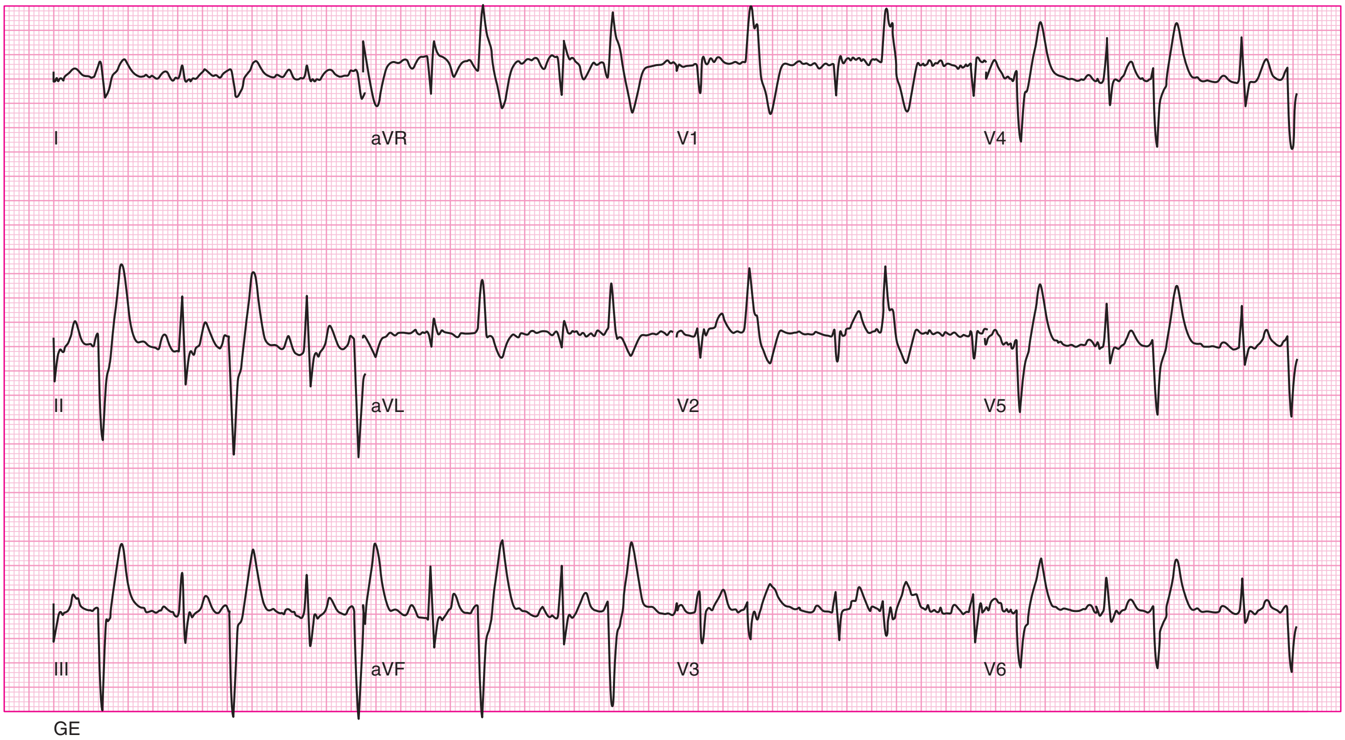
I
aVR
V1
V4
II
aVL
V2
V5
III
aVF
V3
V6
GE

FIGURE 260-2 Ventricular bigeminy with premature ventricular contractions (PVCs) originating from the posteromedial papillary muscle. Twelve-lead electrocardiogram showing normal sinus rhythm with ventricular bigeminy. The PVCs have a right bundle branch block configuration in V_1 with superiorly directed axis. The leads V_4–V_6 are negative. The configuration is consistent with a PVC origin in the posteromedial papillary muscle of the left ventricle.

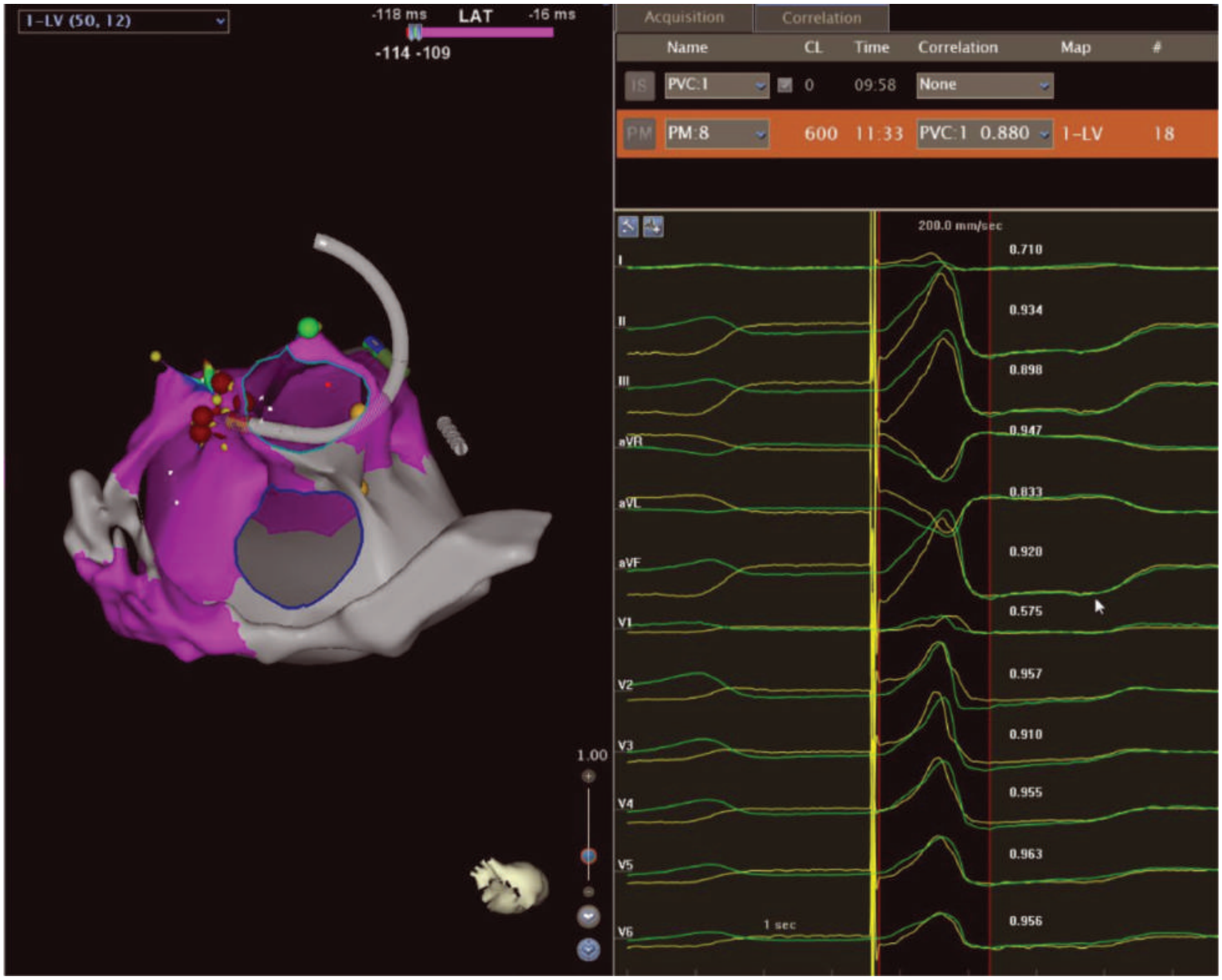

	Name	CL	Time	Correlation	Map	#
IS	PVC:1	0	09:58	None		
PM	PM:8	600	11:33	PVC:1 0.880	1-LV	18

1-LV (50, 12)
-118 ms
LAT
-16 ms
-114 -109
Acquisition
Correlation
200.0 mm/sec
I
0.710
II
0.934
III
0.898
aVR
0.947
aVL
0.833
aVF
0.920
V1
0.575
V2
0.957
V3
0.910
V4
0.955
V5
0.963
V6
0.956
1 sec
1.00

FIGURE 260-3 Catheter ablation of premature ventricular contractions (PVCs) from the left ventricular outflow tract. Shown is an electroanatomic map on the left and PVC morphology with superimposed pacing morphology on the right. The left ventricle is seen from the atrial side (posteriorly), and an ablation catheter is seen passing through the aortic valve at the top of the electroanatomic map and contacting the anterolateral portion of the left ventricular outflow tract. Maroon dots are ablation lesions that have been delivered at the site of interest. Pacing from the site of interest generates a QRS complex very similar to the clinical PVC as seen on the right.

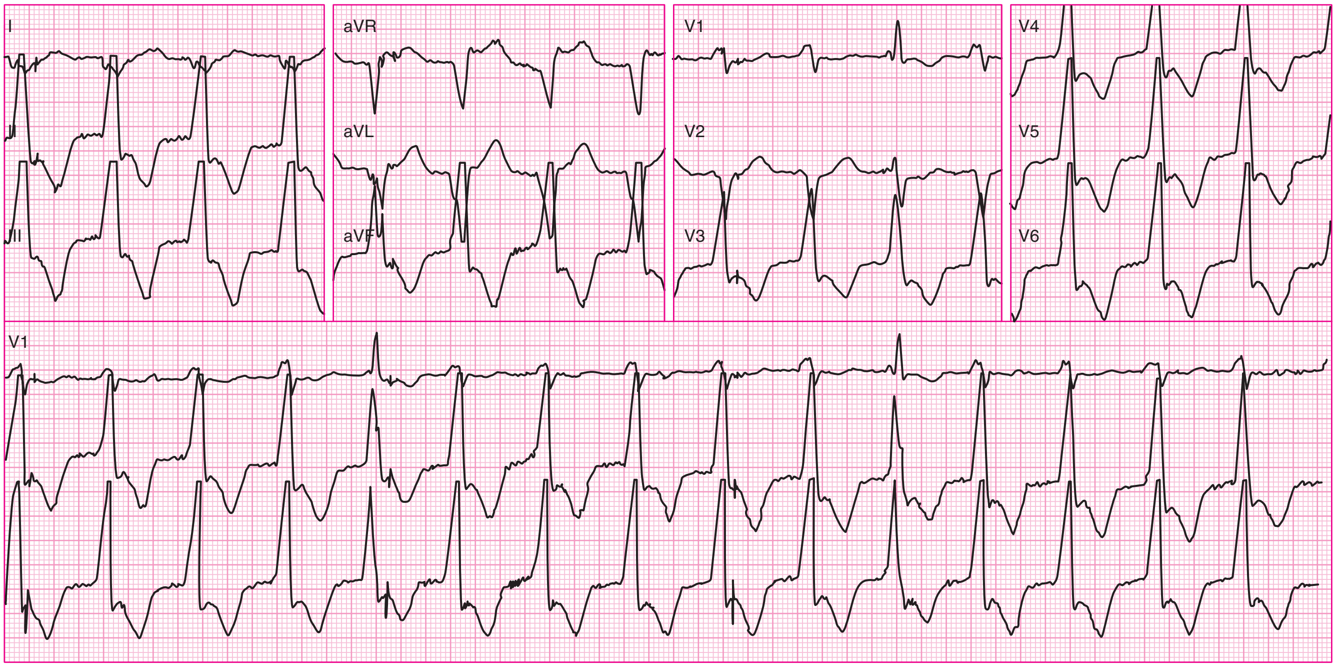

FIGURE 260-4 Accelerated idioventricular rhythm. Shown is an example of a slow regular wide-complex rhythm. Fusion beats are seen on complexes 4 and 10, which are more positive in lead V_1 and narrower than the rest of the beats. These features are consistent with an accelerated idioventricular rhythm.

myocardium can cause AIVR. Idioventricular rhythms are common during acute MI and may emerge during sinus bradycardia. Often, they are not symptomatic, but hemodynamic compromise may occur with the loss of atrioventricular synchrony in susceptible patients. Atropine may be administered to increase the sinus rates if this is a concern. This rhythm is also common in patients with cardiomyopathies or sleep apnea. It can also be idiopathic, often emerging when the sinus rate slows during sleep. Therapy should target any underlying cause and correction of bradycardia. Specific antiarrhythmic therapy for asymptomatic idioventricular rhythm is not necessary.

FUTURE DIRECTIONS

Recently, it has been appreciated that inflammation plays a role in the genesis of PVCs in specific patients with inflammatory cardiomyopathies and even in inherited cardiomyopathies. The roles of early identification of this process and targeted treatment are areas of active research.

■ FURTHER READING

Al-Khatib SM et al: 2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. Heart Rhythm 15:e73, 2018.

Callans DJ: *Josephson's Clinical Cardiac Electrophysiology: Techniques and Interpretations*, 7th ed. Philadelphia, Wolters Kluwer, 2024.

Cronin EM et al: 2019 HRS/EHRA/APHRS/LAHRS expert consensus statement on catheter ablation of ventricular arrhythmias. EP Europace 21:1143, 2019.

Jalife J, Stevenson W (eds): *Zipes and Jalife's Cardiac Electrophysiology: From Cell to Bedside*, 8th ed. Philadelphia, Elsevier, 2022.

Zeppenfeld K et al: 2022 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: Developed by the task force for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death of the European Society of Cardiology (ESC) Endorsed by the Association for European Paediatric and Congenital Cardiology (AEPC). Eur Heart J 43:3997, 2022.

261

Sustained Ventricular Tachycardia

William H. Sauer, Usha B. Tedrow

Sustained monomorphic ventricular tachycardia (VT) is a ventricular arrhythmia with a wide QRS lasting for at least 30 s or requiring an intervention such as antitachycardia pacing from a defibrillator or a cardioversion for termination. Each QRS complex resembles the others, indicating either a focal site of origin or a repetitive exit from a fixed arrhythmia reentry circuit. In structural heart disease, the arrhythmia substrate is most often an area of patchy replacement fibrosis due to infarction, fibrosis, inflammation, or prior cardiac surgery that creates anatomic or functional reentrant pathways. Less commonly, VT is related to reentry or automaticity in diseased conduction pathways in the Purkinje system. While scar-related reentrant VTs are associated with risk of sudden death, idiopathic VT is a more benign form of VT that occurs in structurally normal hearts and can be due to a focal region of automaticity in the myocardium or reentry involving a portion of the Purkinje system.

The clinical presentation varies depending on the rate of the arrhythmia, underlying cardiac function, and autonomic adaptation in response to the arrhythmia. Rapid VT can produce hypotension that may present as syncope, particularly in patients with significant ventricular dysfunction. In contrast, patients with normal cardiac function might tolerate sustained VT, even presenting with simple palpitations, despite rapid rates. Monomorphic VT that is rapid or associated with structural heart disease may eventually deteriorate to ventricular fibrillation (VF), which may be the initial cardiac rhythm recorded at the time of resuscitation of an out-of-hospital cardiac arrest.

DIAGNOSIS

Sustained monomorphic VT **(Table 261-1)** has to be distinguished from other causes of uniform wide QRS tachycardia. These include supraventricular tachycardia with left or right bundle branch block

TABLE 261-1 Sustained Ventricular Arrhythmias

1. **Idiopathic ventricular tachycardia (VT) without structural heart disease**
   - A. Outflow tract origin
     - Right ventricular (RV) outflow tract: left bundle branch block pattern in V_1 with inferior axis (tall QRS in inferior leads) and late transition in the precordial leads
     - Left ventricular (LV) outflow tract: similar inferiorly directed axis but with early precordial transition with prominent R wave in V_2–V_3
   - B. LV fascicular VT: Typical right bundle branch block pattern in V_1 with sharp intrinsicoid deflection and left axis deviation (arising from left posterior fascicle in its most common form)
   - C. Papillary muscle VT
     - Posteromedial: atypical right bundle branch block pattern in V_1 with monophasic R wave and left axis deviation
     - Anterolateral: atypical right bundle branch block pattern in V_1 with positive deflection in lead III and negative deflection in lead I
2. **Ischemic cardiomyopathy**
   - Monomorphic VT is common with prior large myocardial infarction
   - Polymorphic VT and ventricular fibrillation (VF) should prompt ischemia evaluation
3. **Nonischemic cardiomyopathy**
   - Fibrotic scars can cause monomorphic VT, especially with sarcoidosis or other inflammatory cardiomyopathies, Chagas' disease, and familial arrhythmogenic cardiomyopathies such as Lamin A/C genetic cardiomyopathy
   - Polymorphic VT and VF can also occur independently or related to degeneration of monomorphic VT
4. **Arrhythmogenic RV cardiomyopathy**
   - Monomorphic VT usually of RV origin (left bundle branch morphology in V_1)
   - Polymorphic VT and VF can occur independently or related to degeneration of monomorphic VT
5. **Repaired tetralogy of Fallot**
   - Monomorphic VT of RV origin (usually left bundle branch morphology in V_1)
6. **Hypertrophic cardiomyopathy**
   - Polymorphic VT or ventricular fibrillation
   - Less commonly, monomorphic VT associated with myocardial scars, particularly apical aneurysms
7. **Genetic arrhythmia syndromes**
   - A. Long QT syndrome
     - Torsades des pointes VT
   - B. Brugada syndrome
     - Ventricular fibrillation episodes, often nocturnal
   - C. Catecholaminergic polymorphic VT
     - Polymorphic VT or bidirectional VT
   - D. Short QT and early repolarization syndromes
     - Ventricular fibrillation
8. **Idiopathic polymorphic VT or ventricular fibrillation**
   - Usually triggered by recurrent premature ventricular contractions; the most common site of origin is the left posterior fascicle (right bundle branch block/left anterior fascicular block pattern)

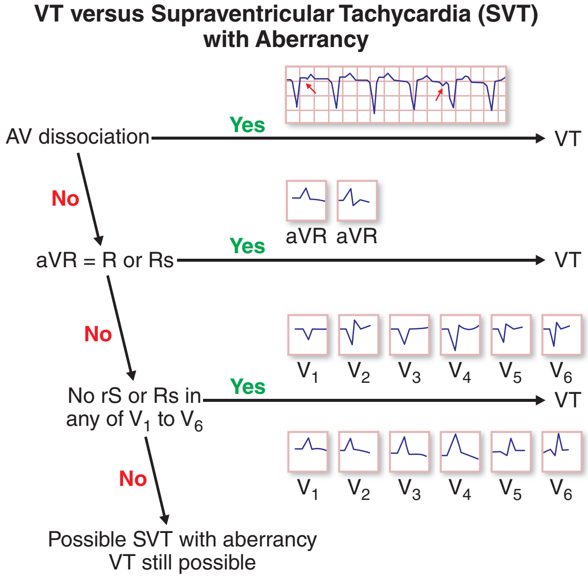
VT versus Supraventricular Tachycardia (SVT) with Aberrancy
AV dissociation
Yes
VT
No
aVR aVR
aVR = R or Rs
Yes
VT
No
V_1 V_2 V_3 V_4 V_5 V_6
No rS or Rs in any of V_1 to V_6
Yes
VT
V_1 V_2 V_3 V_4 V_5 V_6
No
Possible SVT with aberrancy
VT still possible

FIGURE 261-1 Algorithm for differentiation of ventricular tachycardia (VT) from supraventricular tachycardia with aberration. AV, atrioventricular.

aberrant conduction, supraventricular tachycardias conducted to the ventricles over an accessory pathway, and rapid cardiac pacing, appropriate or inappropriate, in a patient with a ventricular pacemaker or defibrillator. In the presence of known heart disease, VT is the most likely diagnosis of a wide QRS tachycardia, independent of QRS morphology. When left ventricular (LV) function is depressed or there is evidence of structural myocardial disease, scar-related reentry is the most likely cause of sustained monomorphic VT. Scars are suggested by pathologic Q waves on the electrocardiogram (ECG), segmental LV or right ventricular wall motion abnormalities on echocardiogram or nuclear imaging, and areas of delayed gadolinium enhancement during magnetic resonance imaging (MRI).

Hemodynamic stability during the arrhythmia does not help distinguish between VT and other mechanisms of wide-complex tachycardia. A number of ECG criteria have been evaluated to distinguish supraventricular tachycardia with aberrancy from VT. The presence of ventriculoatrial (VA) dissociation is a reliable marker for VT, provided the atrial rate is slower than the ventricular rate. Sometimes, P waves can be difficult to define, and the VA relationship cannot be assessed in a patient with an ongoing atrial arrhythmia such as atrial fibrillation. A P wave following each QRS does not exclude VT because 1:1 conduction from ventricle to atrium can occur. A monophasic R wave or Rs complex in aVR or concordance from V_1 to V_6 of monophasic R or S waves is also relatively specific for VT **(Fig. 261-1)**. A number of other QRS morphology criteria have also been described, but all have limitations and are not very reliable in patients with severe heart disease. In patients with known bundle branch block, the same QRS morphology during tachycardia as during sinus rhythm suggests supraventricular tachycardia rather than VT, but even this is not absolutely reliable. Patients with reentry involving the bundle branches of the Purkinje system can have a VT morphology that resembles their native QRS in sinus rhythm. An electrophysiologic study is sometimes required for definitive diagnosis. Occasionally, noise and movement artifacts on telemetry recordings can simulate VT, and prompt recognition of this can avoid unnecessary tests and interventions.

TREATMENT AND PROGNOSIS

Initial management follows Advanced Cardiac Life Support (ACLS) guidelines. If hypotension, impaired consciousness, or pulmonary edema is present, QRS synchronous electrical cardioversion should be performed, ideally after sedation if the patient is conscious. For stable tachycardia, a trial of adenosine is reasonable as this may clarify a supraventricular tachycardia with aberrancy. Adenosine should not be used if the patient has a heart transplant or if the wide-complex rhythm is irregular or unstable. Intravenous amiodarone is the drug of choice if heart disease is present. Following restoration of sinus rhythm, hospitalization and evaluation to define underlying heart disease are required. Assessment of cardiac biomarkers for evidence of myocardial infarction (MI) is appropriate, but acute MI is rarely a cause of sustained monomorphic VT. Elevations in troponin or creatine kinase (CK)-MB are more likely to indicate myocardial injury that is secondary to hypotension and ischemia from fixed coronary lesions during the VT rather than an acute coronary event. Subsequent management is determined by the underlying heart disease and frequency of VT. If VT recurs frequently or is incessant, administration of antiarrhythmic medications or catheter ablation may be required to restore stability. More commonly, sustained monomorphic VT occurs as a single episode but with a high risk of recurrence. Implantable cardioverter-defibrillators (ICDs) are warranted for secondary prevention of sudden death in patients who present with sustained VT associated with structural heart disease **(Fig. 261-2)**.

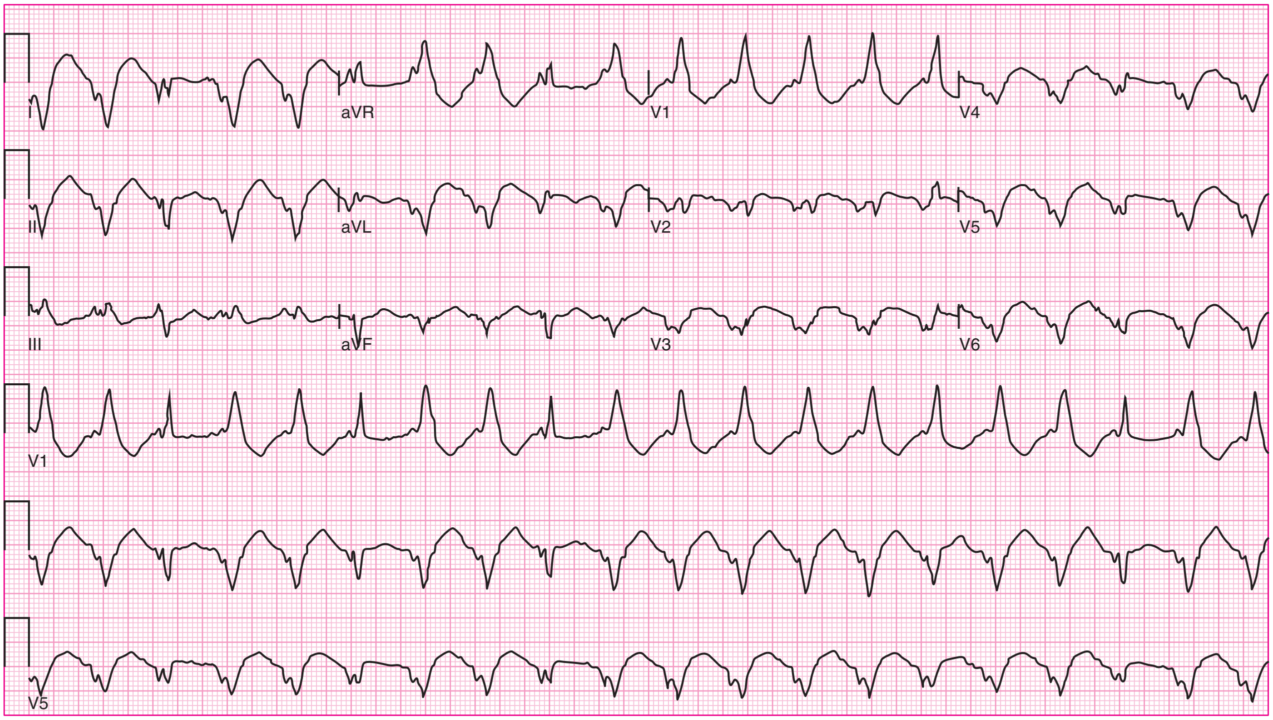

FIGURE 261-2 Monomorphic ventricular tachycardia in a patient with prior myocardial infarction. Shown is a wide-complex tachycardia. Complexes 3, 6, 9, and 18 are narrower and are examples of fusion beats, proving ventriculoatrial (VA) dissociation and proving that this rhythm is in fact ventricular tachycardia.

SUSTAINED MONOMORPHIC VT IN SPECIFIC DISEASES

■ CORONARY ARTERY DISEASE

Patients who present with sustained monomorphic VT associated with coronary artery disease typically have a history of a remote prior large MI. Patients typically present years after the acute infarct with a remodeled ventricle and markedly depressed LV function. Even when there is biomarker evidence of acute MI, a preexisting scar from previous MI should be suspected as the cause of the VT. Infarct scars provide a durable substrate for sustained VT, and up to 70% of patients have a recurrence of the arrhythmia within 2 years. Scar-related reentry is not usually dependent on recurrent acute myocardial ischemia, so coronary revascularization is unlikely to prevent recurrent VT, although it may be appropriate for treatment of angina or other indications. Depressed ventricular function, which is a risk factor for sudden death, is usually present. Implantation of an ICD is clearly indicated for secondary prevention provided that there is a reasonable expectation of survival for 1 year with acceptable functional status. Compared with antiarrhythmic drug therapy, ICDs reduce annual mortality from 12.3 to 8.8% and lower arrhythmic deaths by 50% in patients with hemodynamically significant sustained VT or a history of cardiac arrest. Antiarrhythmic drugs may have some utility for palliation of VT symptoms and prevention of ICD therapies, such as shocks and antitachycardia pacing; however, without an ICD, these drugs do not improve survival.

Following ICD implantation, patients with depressed ejection fraction remain at risk for clinical heart failure, recurrent ischemic events, and recurrent VT, with a 5-year mortality that exceeds 30%. Attention to guideline-directed medical therapy for patients with heart failure and coronary artery disease, including β-adrenergic blocking agents and angiotensin-converting enzyme inhibitors, is important.

ICD therapies, whether shocks or antitachycardia pacing, constitute an adverse event for the patient and are associated with increased rates of heart failure, mortality, and psychological stress. For this reason, recurrent VT episodes in patients with an ICD warrant treatment with medications or catheter ablation. In a randomized study of catheter ablation versus escalated medical therapy (Ventricular Tachycardia Ablation versus Escalation of Antiarrhythmic Drugs [VANISH]), patients receiving catheter ablation fared better than those receiving increasing doses of antiarrhythmic drugs, in particular, amiodarone. Another randomized trial (BERLIN VT) examined a preventative versus deferred ablation strategy in patients who had not yet failed an antiarrhythmic drug. This trial was stopped early for futility, with more procedural complications but fewer VT episodes in the catheter ablation group. For this reason, the most recent consensus statement most strongly recommends catheter ablation for patients with ischemic cardiomyopathy failing or intolerant of antiarrhythmic drugs but also allows for consideration of catheter ablation when long-term therapy with an antiarrhythmic drug (such as amiodarone, which has significant long-term toxicities) is not desired **(Table 261-2)**.

■ NONISCHEMIC DILATED CARDIOMYOPATHY

Sustained monomorphic VT associated with nonischemic cardiomyopathy is usually due to scar-related reentry. The etiology of scar is often unclear, but progressive replacement fibrosis is the likely cause. Patients with nonischemic cardiomyopathy (NICM) have historically been presumed to have a postviral etiology, although increasingly, genetic causes are found in many. Inflammatory etiologies (myocarditis, sarcoidosis) are also increasingly appreciated. On cardiac MRI, scars are detectable as areas of delayed gadolinium enhancement and are more often intramural **(Fig. 261-3)** or subepicardial in location as compared with patients with prior MI. Scars that cause VT are often located adjacent to a valve annulus and can occur in either ventricle. Any cardiomyopathic process can cause scars and VT, but cardiac sarcoidosis, Chagas' disease, and cardiomyopathy due to Lamin A/C mutations are particularly associated with monomorphic VT. An ICD is indicated for patients with a history of sustained VT, syncope, or New York Heart Association class II or III heart failure symptoms, with additional drugs or catheter ablation for control of recurrent VT. In addition, for patients with malignant familial arrhythmogenic cardiomyopathies, an ICD may be considered earlier in the clinical course.

Overall, there are fewer studies of catheter ablation for VT in NICM. Reported success rates are lower than VT ablation in ischemic cardiomyopathy in most observational series. Additionally, inability to reproduce the clinical VT at ablation attempts and epicardial and intramural reentry circuits are important causes of failure of endocardial VT ablation in NICM. Imaging with MRI or computed tomography (CT) scans with late contrast administration to define areas of fibrosis can be useful to guide ablation (Table 261-2).

TABLE 261-2 Summary of Randomized Controlled Studies Assessing Catheter Ablation of Ventricular Tachycardia in Patients with Structural Heart Disease

STUDY, YEAR	STUDY PERIOD	SAMPLE SIZE		INCLUSION CRITERIA	CONTROL ARM	AGE (years)	MALE (%)	ICM (%)	BASELINE LVEF (%)	FOLLOW-UP (months)
		CA	OC							
SMASH-VT, 2007	2000–2006	64	64	Prior MI, ICD for VF or unstable VT	Medical therapy	67 ± 10	87	100	32 ± 9	22.5 ± 5.5
VTACH, 2010	2002–2006	52	55	Prior MI, LVEF ≤50%, ICD indicated for stable VT	ICD + medical therapy	66 ± 8	93	100	34 ± 9	22.5 ± 9
CALYPSO, 2015	2012–2014	13	14	IHD, ≥1 ICD shock or ≥3 ATP therapies for monomorphic VT in last 6 months	AAD therapy	63 ± NR	93	100	30 ± NR	6
VANISH, 2016	2009–2015	132	127	Prior MI, ICD in situ, ≥1 episode of VT while on a class I/III AAD	Escalation of AAD therapy	69 ± 8	93	100	31 ± 11	27.9 ± 17.1
SMS, 2017	2002–2010	54	57	IHD, LVEF ≤40%, unstable VT	ICD + medical therapy	67 ± 8	84	100	31 ± 7	27.0 ± 13.2
BERLIN-VT, 2020	2015–2018	76	83	Prior MI, LVEF 30–50%, ICD in situ for life-threatening VT	Medical therapy until third ICD shock (then VT ablation)	66 ± 10	87	100	41 ± 6	13.2 ± 9.5
PARTITA, 2022	2012–2021	23	24	Cardiomyopathy, had first ICD shock	Medical therapy	68 ± 9	85	81	32 ± 9	24.2 (8.5–24.4)
SURVIVE-VT, 2022	2010–2017	71	73	Prior MI, sustained VT causing ICD shock or syncope	AAD therapy	70 ± 9	96	100	33 ± 11	23
PAUSE-SCD, 2022	2015–2020	60	61	LVEF <50%, ICD indicated for secondary prevention or inducible monomorphic VT on EPS	ICD + medical therapy	55 (46–64)	81	35	40 (30–49)	31.3 (20.1–40)

Abbreviations: AAD, antiarrhythmic drug; ATP, antitachycardia pacing; BERLIN VT, Preventive Ablation of Ventricular Tachycardia in Patients with Myocardial Infarction; CA, catheter ablation; CALYPSO, Catheter Ablation for Ventricular Tachycardia in Patients with an Implantable Cardioverter Defibrillator; EPS, electrophysiology study; ICD, implantable cardioverter-defibrillator; ICM, ischemic cardiomyopathy; IHD, ischemic heart disease; LVEF, left ventricular ejection fraction; MI, myocardial infarction; NR, not reported; OC, other care; PARTITA, Does Timing of VT Ablation Affect Prognosis in Patients with an Implantable Cardioverter-Defibrillator?; PAUSE-SCD, Pan-Asia United States Prevention of Sudden Cardiac Death Catheter Ablation Trial; SMASH VT, Substrate Mapping and Ablation in Sinus Rhythm to Halt Ventricular Tachycardia; SMS, Substrate Modification Study; SURVIVE-VT, Substrate Ablation Versus Antiarrhythmic Drug Therapy for Symptomatic Ventricular Tachycardia; VANISH, Ventricular Tachycardia Ablation Versus Escalated Antiarrhythmic Drug Therapy in Ischemic Heart Disease; VF, ventricular fibrillation; VT, ventricular tachycardia; VTACH, Ventricular Tachycardia Ablation in Coronary Heart Disease.

Source: Reproduced with permission from SA Virk, S Kumar: Catheter ablation of ventricular tachycardia in patients with structural heart disease: A meta-analysis. JACC Clin Electrophysiol 9: 255; 2023.

ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOMYOPATHY

Arrhythmogenic right ventricular cardiomyopathy (ARVC) is a rare genetic disorder most commonly due to mutations in genes encoding for cardiac desmosomal proteins; however, it is increasingly appreciated that other cardiomyopathic processes may produce a similar phenotype. Approximately 50% of patients have a familial transmission with autosomal dominant inheritance. A less common, autosomal recessive form is classically associated with cardiocutaneous syndromes that include Naxos disease and Carvajal syndrome. The former is most commonly related to mutations in plakophilin-2 and plakoglobin, while the latter is most commonly due to a mutation in desmoplakin. Patients are typically diagnosed between the second and fifth decades with palpitations, syncope, or cardiac arrest owing to sustained monomorphic VT, although polymorphic VT can also occur. Fibrosis and fibrofatty replacement most commonly involve the right ventricular myocardium and provide the substrate for reentrant VT that usually has a left bundle branch block–like configuration in ECG lead V_1, consistent with the right ventricular origin, and can resemble idiopathic VT. The sinus rhythm ECG suggests the disease in >85% of patients, most often showing T-wave inversions in V_1–V_3. Delayed activation of the right ventricle may cause a widened QRS (>110 ms) in the right precordial leads (V_1–V_3) and a prolonged S-wave upstroke in those leads and, occasionally, a notched deflection at the end of the QRS known as an epsilon wave. Cardiac imaging may show right ventricular enlargement or areas of abnormal motion or reveal areas of scar on contrast-enhanced MRI.

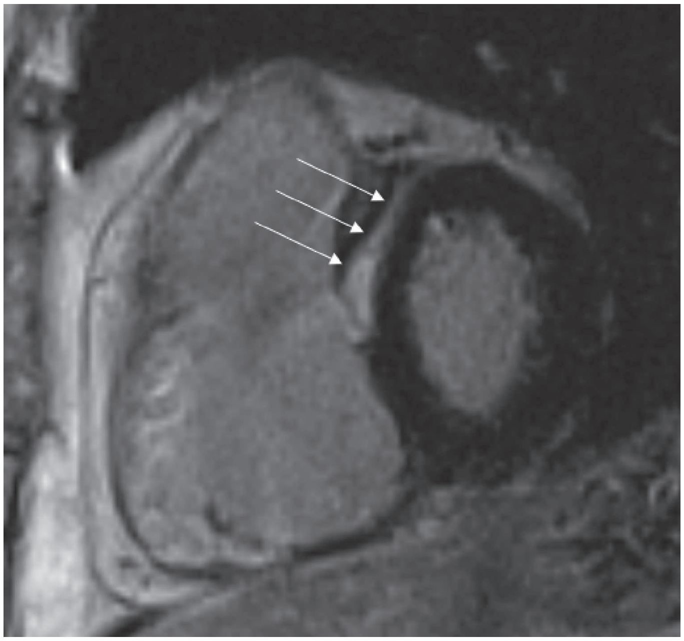

FIGURE 261-3 Cardiac magnetic resonance image (MRI). Shown is an MRI of the heart with the right ventricle on the left and the left ventricle on the right. Between the ventricles (*arrows*) is a stripe of late gadolinium enhancement, indicating midmyocardial fibrosis in the interventricular septum. This type of scar pattern is often seen in patients with nonischemic cardiomyopathies and ventricular tachycardia.

LV involvement can occur and can occasionally precede manifest right ventricular disease. Clinical heart failure is rare except in late stages, and survival to advanced age can be anticipated provided that VT can be controlled. An ICD is recommended. When VT is exercise induced, it may respond to β-adrenergic blockers and limiting exercise. Sotalol, flecainide, and amiodarone have been used to reduce ventricular arrhythmias. Catheter ablation prevents or reduces VT episodes in 70% of patients, but epicardial mapping and ablation are often required.

ADULT CONGENITAL HEART DISEASE

Among all patients with adult congenital heart disease (ACHD), sustained monomorphic VT is quite rare. However, the most common substrate for sustained VT is seen in those with repairs of a ventricular septal defect, in particular tetralogy of Fallot (TOF). The prevalence of VT after TOF repair is estimated to be 3–14%, and risk of sudden cardiac death may reach as high as 1% per year in adulthood by the fourth or fifth decade of life. The greatest risk for ventricular arrhythmias is posed via two potential mechanisms: (1) those who have undergone repair involving a ventriculotomy and (2) those with long-standing hemodynamic overload causing ventricular dysfunction and/or hypertrophy independent of surgical incisions.

Monomorphic VT in TOF most commonly occurs in stereotyped circuits due to reentry around areas of surgically created scar in the right ventricle. Factors associated with VT risk include age >5 years at the time of repair, high-grade ventricular ectopy, inducible VT on an electrophysiologic study, abnormal right ventricular hemodynamics, and sinus rhythm QRS duration >180 ms. An ICD is usually warranted for patients who have a spontaneous episode of VT, but ICDs are also considered for patients with multiple risk factors. Catheter ablation or antiarrhythmic drug therapy is used to control recurrent episodes.

BUNDLE BRANCH REENTRY VT

Reentry through the Purkinje system occurs in ~5% of patients with monomorphic VT in the presence of structural heart disease. The reentry circuit typically revolves retrograde via the left bundle and anterograde down the right bundle, thereby producing VT that has a left bundle branch block configuration. The VT QRS morphology may closely resemble the QRS morphology in sinus rhythm. Catheter ablation of the right bundle branch abolishes this VT. Bundle branch reentry is usually associated with severe underlying heart disease. Other scar-related VTs are often present and often require additional therapy.

IDIOPATHIC MONOMORPHIC VT

Idiopathic VT in patients without structural heart disease usually presents with palpitations, lightheadedness, and, rarely, syncope. Episodes can be provoked either by sympathetic stimulation or acute withdrawal of sympathetic tone, as in the immediate postexercise period. The QRS morphology of the arrhythmia suggests the diagnosis (see below). The sinus rhythm ECG is normal. Family history suggests no familial cardiomyopathy or sudden death. Cardiac imaging, including echocardiography and cardiac MRI, shows normal ventricular function and no evidence of ventricular scar. Occasionally, a patient with structural heart disease is found to have concomitant idiopathic VT unrelated to the structural disease, in which case, the underlying disease should be treated as per the guidelines, separate from the VT. Repeated bursts of nonsustained VT, which may occur incessantly, are known as repetitive monomorphic VT and can cause a tachycardia-induced cardiomyopathy with depressed ventricular function that recovers after suppression of the arrhythmia. Sudden death in isolated idiopathic VT is rare, and an ICD is not recommended.

Outflow tract VTs originate from a focus near the pulmonic or aortic valve annuli, usually with features consistent with triggered automaticity. The arrhythmia may present with sustained VT, nonsustained VT, or premature ventricular contractions (PVCs). Most originate in the right ventricular outflow tract, which gives rise to VT that has a left bundle branch block configuration in V_1 and an axis that is directed inferiorly, with tall R waves in II, III, and aVF. Idiopathic VT can also arise in the LV outflow tract or in sleeves of myocardium that extend along the aortic root. LV origin is suspected when lead V_1 or V_2 has prominent R waves. Although this typical outflow tract QRS morphology favors idiopathic VT, some cardiomyopathies, notably ARVC, can cause PVCs or VT from this region. Excluding these diseases is an initial focus of evaluation **(Fig. 261-4)**.

LV fascicular VT, sometimes referred to as Belhassen's VT or verapamil-sensitive VT, is the second most common form of idiopathic VT after outflow tract VTs. It often presents with sustained VT that has a right bundle branch block–like configuration and is negative in the inferior leads. It is often exercise induced and occurs more often in men than women. The mechanism was originally thought to be focal but has been demonstrated to be due to a small reentry circuit in or near the septal ramifications of the LV Purkinje system. There can be an LV false tendon associated with this rhythm. Despite its reentrant nature, the course of this type of VT is typically benign.

Other sites of origin for idiopathic VT exist, including papillary muscles, mitral and tricuspid valve annuli, and the moderator band in the right ventricle. Even focal sites from the epicardial surface have been described. The presence of VT from these more unusual sites should prompt even more careful assessment for structural heart disease.

MANAGEMENT OF IDIOPATHIC VT

Treatment is required for symptoms or when frequent or incessant arrhythmias depress ventricular function. Symptoms can be controlled with medications including beta blockers, calcium channel blockers,

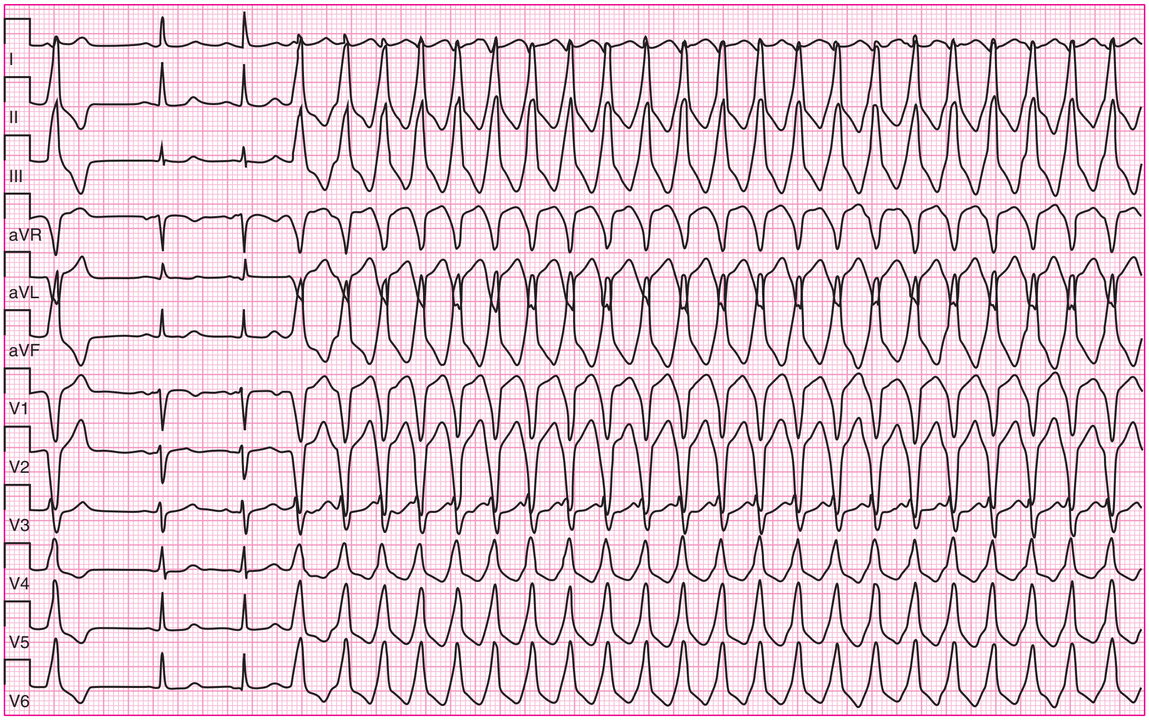

FIGURE 261-4 Idiopathic monomorphic ventricular tachycardia (VT). This is a 12-lead electrocardiogram showing the onset of idiopathic VT in a young patient without structural heart disease. The VT has a left bundle branch block configuration in V_1 and an inferiorly directed axis consistent with an outflow tract origin. Note that the narrow (normal sinus) beats have a normal QRS configuration, consistent with the patient's lack of structural heart disease.

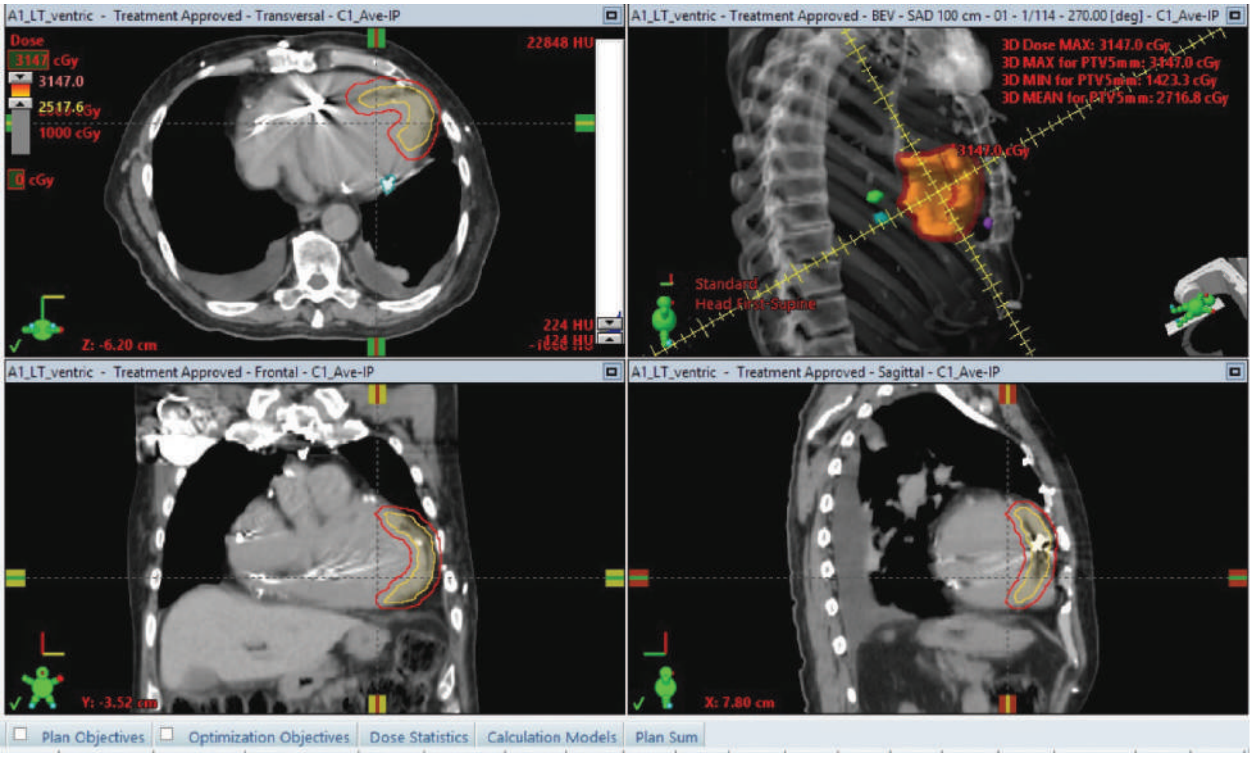
A1_LT_ventric - Treatment Approved - Transversal - C1_Ave-IP
Dose
3147 cGy
3147.0
2517.6
1000 cGy
0 cGy
22848 HU
224 HU
Z: -6.20 cm
A1_LT_ventric - Treatment Approved - BEV - SAD 100 cm - 01 - 1/114 - 270.00 [deg] - C1_Ave-IP
3D Dose MAX: 3147.0 cGy
3D MAX for PTV5mm: 3147.0 cGy
3D MIN for PTV5mm: 1423.3 cGy
3D MEAN for PTV5mm: 2716.8 cGy
3147.0 cGy
Standard
Head First-Supine
A1_LT_ventric - Treatment Approved - Frontal - C1_Ave-IP
Y: -3.52 cm
A1_LT_ventric - Treatment Approved - Sagittal - C1_Ave-IP
X: 7.80 cm
☐ Plan Objectives
☐ Optimization Objectives
Dose Statistics
Calculation Models
Plan Sum

FIGURE 261-5 Stereotactic body radiation therapy. Shown is a planning computed tomography scan for noninvasive cardiac ablation delivered with stereotactic radiation. The region of interest is determined by analysis of presenting arrhythmias, underlying structural heart disease, and proximity to adjacent structures such as coronary arteries, the phrenic nerve, and the gastrointestinal tract. Radiation is delivered to the chosen treatment volume in order to create conduction block in the culprit scar region.

and sodium channel blockers such as flecainide. Although flecainide is not typically recommended in patients with structural heart disease, it has been used successfully to resolve tachycardia-induced cardiomyopathy in the setting of idiopathic PVCs and VT in patients with a normal underlying ECG and no fibrosis evident on cardiac MRI. Catheter ablation is also indicated for control of symptoms, has an overall success rate of 80%, and is recommended for those with symptomatic VT in whom medications are ineffective or not preferred by the patient. Efficacy and risks of catheter ablation vary with the specific site of origin of the VT, being most favorable for arrhythmias originating in the right ventricular outflow tract. Failure of ablation is most often due to inability to initiate the arrhythmia for mapping in the electrophysiology laboratory.

LV interfascicular VT can be terminated by intravenous administration of verapamil, although chronic therapy with oral verapamil is not always effective. Catheter ablation is recommended if β-adrenergic blockers or calcium channel blockers are ineffective or not desired.

FUTURE DIRECTIONS

Treatment of monomorphic VT, especially in the setting of structural heart disease, is an important cause of morbidity and mortality. Limitations of current therapies include toxicities of antiarrhythmic drugs and inability to successfully perform catheter ablation of the substrate for arrhythmias. Advances in imaging and intracardiac mapping techniques are likely to improve success rates over time. In addition, the inability of ablative energy to reach deep intramural substrates is a current limitation. Innovations in delivery of ablative energy, including bipolar ablation, needle catheter ablation, and electroporation, are ongoing. In addition, noninvasive ablation is a promising area of investigation. Proton beam or stereotactic radiation can be used to target VTs identified by advanced cardiac imaging or by a multielectrode ECG vest. Early multicenter studies suggest durable control of VT with noninvasive ablation **(Fig. 261-5)**.

■ FURTHER READING

Al-Khatib SM et al: 2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. Heart Rhythm 15:e73, 2018.

Callans DJ: *Josephson's Clinical Cardiac Electrophysiology: Techniques and Interpretations*, 7th ed. Philadelphia, Wolters Kluwer, 2024.

Cronin EM et al: 2019 HRS/EHRA/APHRS/LAHRS expert consensus statement on catheter ablation of ventricular arrhythmias. EP Europace 21:1143, 2019.

Jalife J, Stevenson W (eds): *Zipes and Jalife's Cardiac Electrophysiology: From Cell to Bedside*, 8th ed. Philadelphia, Elsevier, 2022.

Zeppenfeld K et al: 2022 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: Developed by the task force for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death of the European Society of Cardiology (ESC) Endorsed by the Association for European Paediatric and Congenital Cardiology (AEPC). Eurn Heart J 43:3997, 2022.

262

Polymorphic Ventricular Tachycardia and Ventricular Fibrillation

William H. Sauer, Usha B. Tedrow

POLYMORPHIC VENTRICULAR TACHYCARDIA

Sustained polymorphic ventricular tachycardia (VT) has a continuously changing QRS configuration from beat to beat, indicating a continually changing ventricular activation sequence. However, unlike sustained monomorphic VT, polymorphic VT does not necessarily indicate a fixed structural abnormality or focus of automaticity. Reentry can occur with variable reentrant paths, spiral wave reentry, and multiple automatic foci as potential mechanisms. This type of reentry can occur near fibrotic areas of myocardium, potentiated by proximity to damaged Purkinje cells or ventricular hypertrophy. Abnormal transmural dispersion of repolarization can occur in the setting of channelopathies or antiarrhythmic drugs in the absence of structural heart disease. Sustained polymorphic VT usually degenerates into ventricular fibrillation (VF). Polymorphic VT is typically seen in association with acute myocardial infarction or ischemia (MI), ventricular hypertrophy, and a number of genetic mutations that affect cardiac ion channels.

POLYMORPHIC VT ASSOCIATED WITH ACUTE MYOCARDIAL INFARCTION/ ISCHEMIA

Acute MI or ischemia is a common cause of polymorphic VT and should be the initial consideration in management. Approximately 10% of patients with acute MI develop VT that degenerates to VF, likely related to reentry through the infarct border zone. The risk is greatest in the first hour of acute MI. More rarely, surviving Purkinje cells with automaticity can initiate polymorphic VT. Following defibrillation as per the Advanced Cardiac Life Support (ACLS) guidelines, management is as for acute MI. β-Adrenergic blockers, correction of electrolyte abnormalities, and prompt myocardial reperfusion are required. Repeated episodes of polymorphic VT may suggest ongoing MI and warrant assessment of adequacy of myocardial reperfusion. Polymorphic VT and VF that occur within the first 48 h of acute MI are associated with greater in-hospital mortality, but patients who survive past hospital discharge are not at increased risk for arrhythmic sudden death. Long-term therapy for postinfarct ventricular arrhythmia is determined by residual left ventricular (LV) function, with an implantable cardioverter defibrillator (ICD) indicated for persistent severe LV dysfunction (LV ejection fraction <35%).

REPOLARIZATION ABNORMALITIES AND GENETIC ARRHYTHMIA SYNDROMES

■ ACQUIRED LONG QT SYNDROME

Abnormal prolongation of the QT interval is associated with the polymorphic VT torsades des pointes. The VT often has a characteristic initiation sequence of a premature ventricular beat that induces a pause, followed by a sinus beat that has a longer QT interval and interruption of the T wave by the premature ventricular contraction (PVC) that is the first beat of the polymorphic VT **(Fig. 262-1)**. This characteristic initiation is termed *pause-dependent*. Causes of QT prolongation include electrolyte abnormalities, bradycardia, and a number of medications that block repolarizing potassium currents, notably the antiarrhythmic drugs sotalol, dofetilide, and ibutilide, but also a number of other medications used for noncardiac diseases, including erythromycin, pentamidine, haloperidol, phenothiazines, and methadone. Individual susceptibility may be related to genetic polymorphisms or mutations that influence repolarization. The website *crediblemeds.org* is an excellent resource for the clinician to determine whether a given medication has been reported to prolong the QT interval.

Patients typically present with near-syncope, syncope, or cardiac arrest. Sustained episodes degenerate to VF requiring defibrillation. PVCs and nonsustained VT often precede episodes of sustained VT. Intravenous administration of 1–2 g of magnesium sulphate usually suppresses recurrent episodes. If magnesium alone is ineffective, increasing heart rate with isoproterenol infusion or pacing, to a rate of 100–120 depolarizations/min as required to suppress PVCs, usually suppresses VT recurrences. These maneuvers allow time for correction of associated electrolyte disturbance (hypokalemia and hypocalcemia) and bradycardia and removal of any causative drugs **(Table 262-1)**. Drug interactions that elevate levels of the offending agent are often a precipitating factor. Patients who experience a polymorphic VT induced by QT prolongation should be considered to have a susceptibility to the arrhythmia and should avoid all future exposure to medications known to prolong the QT interval.

■ CONGENITAL LONG QT SYNDROME

The congenital long QT syndrome (LQTS) is caused by mutations in genes coding for cardiac ion channels responsible for ventricular repolarization. The corrected QT (QTc) is typically prolonged to >440 ms in men and 460 ms in women. Symptoms are due to torsades des pointes VT. Several forms of congenital LQTS have been identified, but three groups of mutations that lead to LQTS-1, LQTS-2, and LQTS-3

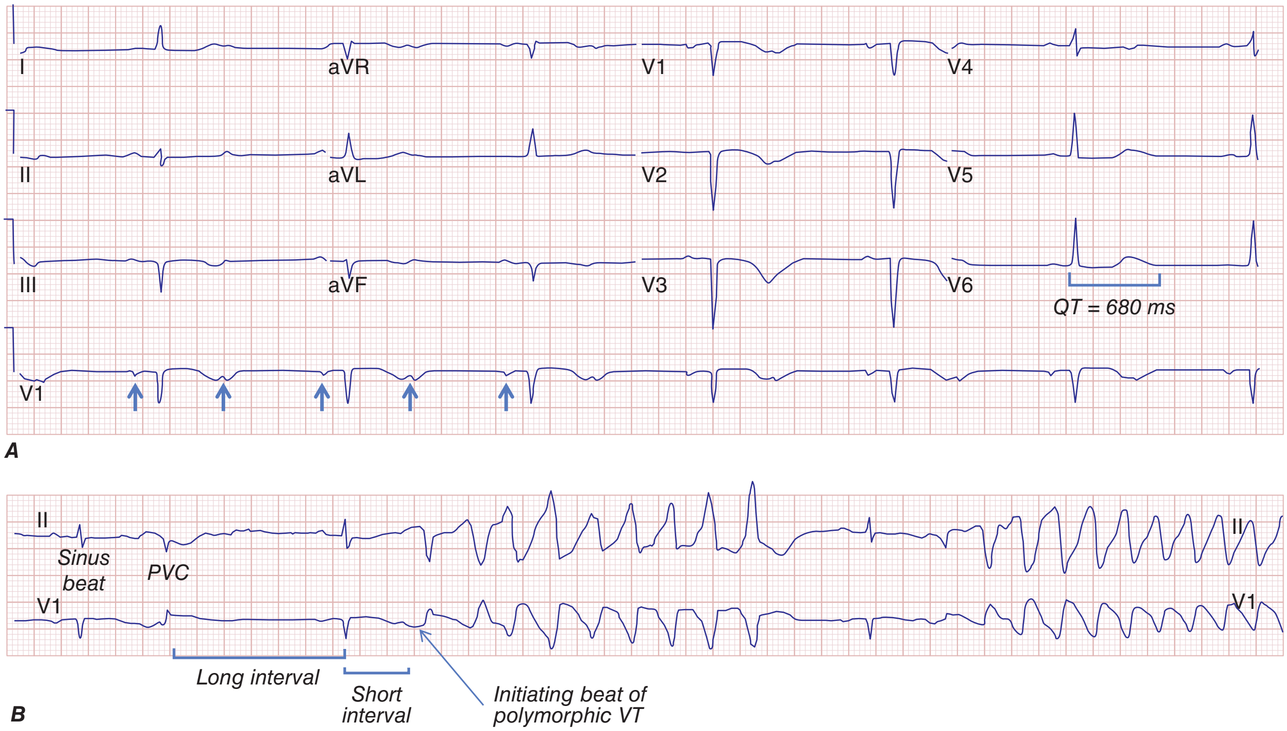

FIGURE 262-1 Torsades des pointes ventricular tachycardia (VT) in a patient with bradycardia and marked QT prolongation. *A.* Twelve-lead electrocardiogram showing 2:1 atrioventricular block (P waves marked by *blue arrows*) with a heart rate of 40 beats/min and QT interval of 680 ms and corrected QT of 550 ms. ***B.*** The bottom panel shows a telemetry rhythm strip with periods of self-limiting torsades des pointes polymorphic VT. Following a normally conducted sinus beat, a premature ventricular contraction (PVC) causes a compensatory pause, leading to a long RR interval. A PVC after the next sinus beat initiates VT. This is the classic pause-dependent mode of initiation of torsades des pointes VT with long–short intervals.

TABLE 262-1 Causes of QT Prolongation and Torsades des Pointes

1. Congenital long QT syndromes
    - Long QT syndrome type 1: Reduced repolarizing current I_{Ks} due to mutation in *KCNQ1* gene
    - Long QT syndrome type 2: Reduced repolarizing current I_{Kr} due to mutation in *KCNH2* gene
    - Long QT syndrome type 3: Delayed inactivation of the I_{Na} due to mutations in *SCN5A* gene
    - Others: Several other types of long QT syndromes have been described; long QT syndrome types 1, 2, and 3 account for 80–90% of cases
2. Electrolyte abnormalities: Hypokalemia, hypomagnesemia, hypocalcemia
3. Drug-induced acquired prolongation of QT interval
    - Antiarrhythmic drugs
        - Class IA: Quinidine, disopyramide, procainamide
        - Class III: Sotalol, dronedarone, ranolazine, amiodarone, ibutilide, dofetilide
    - Antibiotics
        - Macrolides: Erythromycin, clarithromycin, azithromycin
        - Fluroquinolones: Levofloxacin, moxifloxacin
        - Trimethoprim-sulfamethoxazole
        - Clindamycin
        - Pentamidine
        - Chloroquine
        - Antifungals: Ketoconazole, itraconazole
        - Antivirals: Amantadine
    - Antipsychotics
        - Haloperidol, phenothiazines, thioridazine, trifluoperazine, sertindole, zimelidine, ziprasidone
        - Tricyclic and tetracyclic antidepressants
    - Antihistamines (histamine 1-receptor antagonists)
        - Astemizole, diphenhydramine, hydroxyzine
    - Other drugs
        - Citrate (massive blood transfusions)
        - Cocaine
        - Methadone
        - Hydroxychloroquine
4. Cardiac conditions
    - Myocardial ischemia and infarction
    - Myocarditis
    - Marked bradycardia
    - Stress cardiomyopathy
5. Endocrine disorders
    - Hypothyroidism
    - Hyperparathyroidism
    - Pheochromocytoma
    - Hyperaldosteronism
6. Intracranial disorders
    - Subarachnoid hemorrhage
    - Thalamic hematoma
    - Cerebrovascular accident
    - Encephalitis
    - Head injury
7. Nutritional disorders
    - Anorexia nervosa
    - Starvation
    - Liquid protein diets
    - Gastroplasty and ileojejunal bypass
    - Celiac disease

syndromes account for 90% of cases. The most frequently encountered mutations (LQTS-1 and LQTS-2) are due to abnormalities of potassium channels, but mutations affecting the sodium channel (LQTS-3) and calcium channels have also been described.

Typical presentation is with syncope or cardiac arrest, usually during childhood. In LQTS-1, episodes tend to occur during exertion, particularly swimming. In LQTS-2, sudden auditory stimuli or emotional upset predisposes to events. In LQTS-3, sudden death tends to occur during sleep. Asymptomatic patients may be discovered in the course of family screening or on a routine electrocardiogram (ECG). Genotyping can be helpful for family screening and to provide reassurance regarding the diagnosis. Correlations of genotype with risk and response to therapy are beginning to emerge. In most patients with LQTS-1 or LQTS-2, adequate doses of beta-blocker therapy (the nonselective agents nadolol and propranolol are favored) are sufficient protection from arrhythmia episodes. Markers of increased risk include QTc interval exceeding 500 ms, female gender, and a history of syncope or cardiac arrest. Recurrent syncope despite beta-blocker therapy or a high-risk profile merits consideration of an ICD. Avoidance of QT-prolonging drugs is critical for all patients with LQTS, including those who are genotype positive but have normal QT intervals.

SHORT QT SYNDROME

Short QT syndrome is very rare compared to LQTS. The QTc is <360 ms and usually <300 ms. The genetic abnormality causes a gain of function of the potassium channel (I_{Kr}) or reduced inward depolarizing currents. The abnormality is associated with atrial fibrillation, polymorphic VT, and sudden death.

BRUGADA SYNDROME

Brugada syndrome is a rare syndrome characterized by >0.2 mV of ST-segment elevation with a coved ST segment and negative T wave in more than one anterior precordial lead (V_1–V_3) **(see Fig. 260-1)** and episodes of syncope or cardiac arrest due to polymorphic VT in the absence of structural heart disease. The term *Brugada syndrome* is distinguished from *Brugada pattern* observed in asymptomatic individuals with the characteristic ECG findings, similar to how Wolff-Parkinson-White syndrome is distinguished from the Wolff-Parkinson-White pattern. Cardiac arrest may occur during sleep or be provoked by febrile illness. The mechanism of the ventricular arrhythmias associated with Brugada syndrome are related to the dispersion of repolarization across the ventricular wall with epicardial and M cells that are more affected than endocardial cells. The characteristic coved ST segment in the anterior precordial leads is a consequence of this difference in transmembrane action potentials localized to the right ventricular outflow tract. This regionalized substrate is apparent with epicardial mapping and allows for ablation as a potential treatment option for patients with recurrent VT/VF despite pharmacologic therapy.

Males are more commonly affected than females and men are much more likely to have Brugada pattern on ECG compared to women. In cases where there is a family history, Brugada syndrome demonstrates an autosomal dominant inheritance pattern with variable penetrance. Mutations involving cardiac sodium channels (*SCN5A*, *SCN1B*, and *SCN10A*) are identified in ~25% of cases. There are non–sodium channel genes that have also been described including loss-of-function mutations in calcium and gain-of-function mutations in potassium channel component genes affecting the transient outward current (I_{to}). Structural and trafficking proteins have also been implicated in Brugada syndrome, revealing a higher level of complexity compared to other monogenetic arrhythmia substrates.

Distinction from patients with similar ST elevation owing to LV hypertrophy, pericarditis, myocardial ischemia or MI hyperkalemia, hypothermia, right bundle branch block, and arrhythmogenic right ventricular cardiomyopathy is often difficult. Furthermore, the characteristic ST-segment elevation can wax and wane over time and may become pronounced during acute illness and fever or with changes in autonomic tone. Administration of the sodium channel–blocking drugs flecainide, ajmaline, or procainamide can augment or unmask ST elevation in affected individuals. An ICD is indicated for individuals who have had unexplained syncope or have been resuscitated from cardiac arrest. Patients with Brugada syndrome must avoid drugs known to affect sodium channels including cocaine, cannabis, some sodium channel antiepileptic agents, and other psychotropic medications. Quinidine and catheter ablation of abnormal regions in the epicardial right ventricular outflow tract have been used successfully to suppress frequent episodes of VT.

■ EARLY REPOLARIZATION SYNDROME

Patients resuscitated from VF who have no structural heart disease or other identified abnormality have a higher prevalence of J-point elevation with notching in the terminal QRS. A family history of sudden death is present in some patients, suggesting a potential genetic basis. J-point elevation is also seen in some patients with the Brugada syndrome and is associated with a higher risk of arrhythmias. An ICD is recommended for those who have had prior cardiac arrest. It should be noted that J-point elevation is commonly seen as a normal variant in patients without arrhythmias, and in the absence of specific symptoms, the clinical relevance is not known.

■ CATECHOLAMINERGIC POLYMORPHIC VT

This rare familial syndrome is due to mutations in the cardiac ryanodine receptor and, less commonly, the sarcoplasmic calcium binding protein calsequestrin 2. These mutations result in abnormal sarcoplasmic calcium handling and polymorphic ventricular arrhythmias that resemble those seen with digitalis toxicity. The VT is polymorphic or has a characteristic alternating QRS morphology termed bidirectional VT. Patients usually present during childhood with exercise or emotion-induced palpitations, syncope, or cardiac arrest. β-Adrenergic blockers (e.g., nadolol and propranolol) and an ICD are usually recommended. Verapamil, flecainide, or surgical left cardiac sympathetic denervation reduces or prevents recurrent VT in some patients. The use of ICDs is controversial because of the fear that an ICD shock could initiate a vicious cycle of adrenergic output and escalated ventricular arrhythmias, leading to death.

■ HYPERTROPHIC CARDIOMYOPATHY

Hypertrophic cardiomyopathy (HCM) is the most common genetic cardiovascular disorder, occurring in 1 in 500 individuals, and is a prominent cause of sudden death before the age of 35 years. Sudden death can be due to polymorphic VT/VF. Rarely, sustained monomorphic VT occurs related to areas of ventricular scar, most commonly in patients who develop an apical aneurysm. Risk factors for sudden death in this disease include young age, nonsustained VT, failure of blood pressure to increase during exercise, recent (within 6 months) syncope, ventricular wall thickness >3 cm, and possibly the severity of LV outflow obstruction. An ICD is generally indicated for high-risk subjects, but the specific risk profile warranting an ICD continues to be debated. Surgical myectomy, performed to relieve outflow obstruction, has been associated with a sudden death rate of <1% per year. The reported annual rate of sustained VT or sudden death after transcoronary ethanol septal ablation done to relieve outflow obstruction has been reported to range between 1 and 5%.

■ GENETIC DILATED CARDIOMYOPATHIES

Genetic dilated cardiomyopathies account for 30–40% of cases of nonischemic dilated cardiomyopathies. Some are associated with muscular dystrophy. Autosomal dominant, recessive, X-linked, and mitochondrial inheritance patterns are recognized. Mutations in genes coding for structural proteins of the nuclear lamina (Lamin A/C) and the *SCN5A* gene are particularly associated with conduction system disease and ventricular arrhythmias. Patients can experience polymorphic VT and cardiac arrest or develop areas of scar causing sustained monomorphic VT. ICDs are recommended for those who have had a sustained VT, are at high risk due to significantly depressed ventricular function (LV ejection fraction <35% and associated with heart failure), or have a malignant family history of sudden death.

■ VENTRICULAR FIBRILLATION

VF is characterized by disordered electrical ventricular activation without identifiable QRS complexes **(Fig. 262-2)**. Spiral wave reentry and multiple circulating reentry wavefronts are possible mechanisms. Sustained polymorphic or monomorphic VT that degenerates to VF is a common cause of out-of-hospital cardiac arrest. Treatment follows ACLS guidelines with defibrillation to restore sinus rhythm. If resuscitation is successful, further evaluation is performed to identify and treat underlying heart disease and potential causes of the arrhythmia, including the possibility that monomorphic or polymorphic VT could have initiated VF. If a transient reversible cause such as acute MI is not identified, therapy to reduce the risk of sudden death with an ICD is often warranted.

FUTURE DIRECTIONS

The role for catheter ablation in polymorphic VT and VF is rapidly evolving, with some investigators making use of simultaneous electrical recordings from basket catheters to define critical sites for the arrhythmia and correlating these with detailed cardiac imaging to identify typical vulnerable sites for ablation.

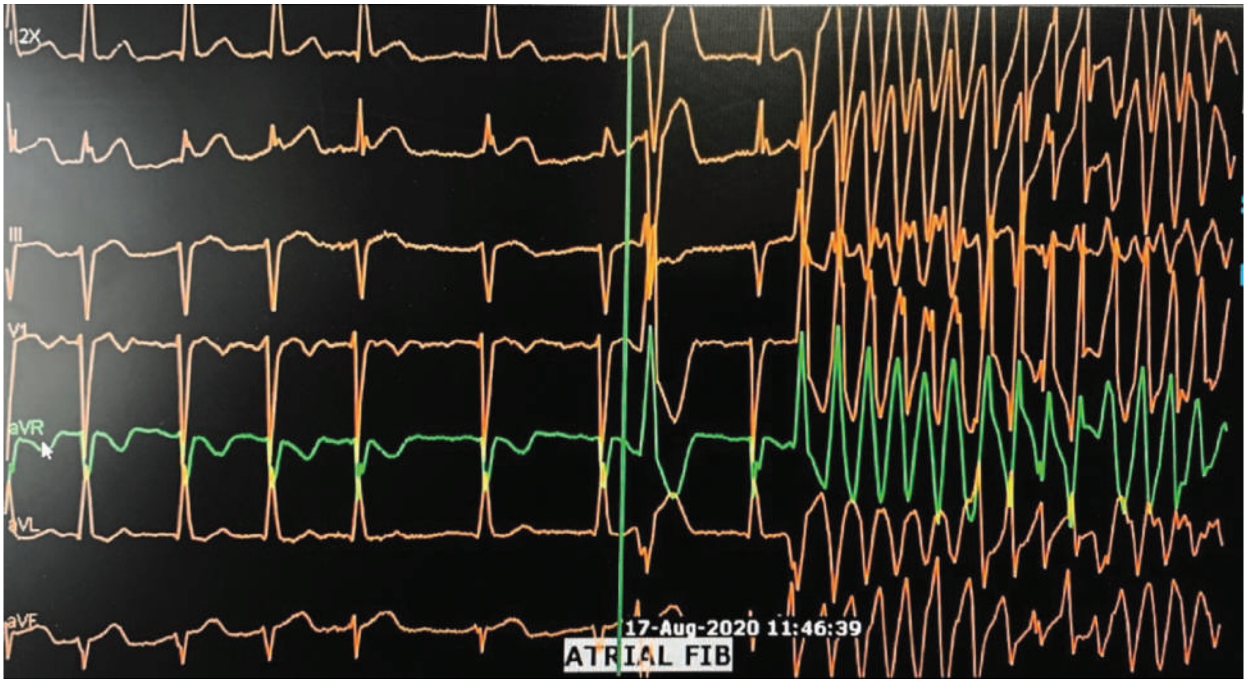

FIGURE 262-2 Fascicular ectopy triggering ventricular fibrillation. Shown is a multilead monitor from a patient with recent inferoposterior myocardial infarction and surgical revascularization. Purkinje fibers can often survive acute infarction due to greater cellular glycogen stores and oxygenation from the left ventricular cavity. Upon revascularization, these now surviving but poorly coupled Purkinje fibers can trigger premature ventricular contractions and ventricular fibrillation as shown in this strip.

FURTHER READING

Callans DJ: *Josephson's Clinical Cardiac Electrophysiology: Techniques and Interpretations*, 7th ed. Philadelphia, Wolters Kluwer, 2024.

Jalife J, Stevenson W (eds): *Zipes and Jalife's Cardiac Electrophysiology: From Cell to Bedside*, 8th ed. Philadelphia, Elsevier, 2022.

Zeppenfeld K et al: 2022 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: Developed by the task force for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death of the European Society of Cardiology (ESC) Endorsed by the Association for European Paediatric and Congenital Cardiology (AEPC). Eur Heart J 43:3997, 2022.

263

Electrical Storm and Incessant Ventricular Tachycardia

William H. Sauer, Usha B. Tedrow

ELECTRICAL STORM

Electrical storm or ventricular tachycardia (VT) storm refers to the occurrence of three or more episodes of VT or ventricular fibrillation (VF) within 24 h requiring intervention for termination. Although this disorder is uncommon in the general population, it is of great concern to internists because of its rapid course leading to patient death in the absence of treatment. Therefore, prompt recognition and therapeutic intervention are required. Electrical storms occur in 4% of patients with a primary prevention implantable cardioverter-defibrillator (ICD) but in as many as 20% of patients with a history of known VT or resuscitated sudden death. Catheter ablation, antiarrhythmic drug therapy, and other adjunctive therapies in electrical storms can be life-saving.

INCESSANT VT

VT is designated incessant when VT continues to recur shortly after electrical, pharmacologic, or spontaneous conversion to sinus rhythm **(Fig. 263-1)**. Typically, VT is monomorphic. Rarely, a slow incessant monomorphic VT will fail detection by an ICD because it falls outside of the programmed detection parameters. If the arrhythmia is hemodynamically stable acutely, patients can present with symptoms of gradual cardiac decompensation. VT may become incessant due to the proarrhythmic effect of an antiarrhythmic drug such as amiodarone or a sodium channel blocker such as flecainide. Hemodynamic support may be required acutely until the precipitating factors can be corrected. Urgent catheter ablation is often warranted.

MANAGEMENT OF PATIENTS PRESENTING WITH ICD SHOCKS

A substantial number of patients who receive an ICD can be expected to have an arrhythmia that is terminated by the ICD, either by a shock or antitachycardia pacing. Although this is an expected event, it can be a sign of impending instability, deterioration of cardiac function, or emergence of a new arrhythmia and therefore requires evaluation. Interrogation of the ICD is crucial after a patient reports a shock or symptoms of arrhythmia to confirm that the therapy was indeed delivered for a ventricular arrhythmia and not for lead malfunction or an atrial arrhythmia. After a shock and in the absence of other symptoms to suggest arrhythmia or ischemia, patients have the option of waiting until the next working day or using remote monitoring to transmit device interrogation data to their physician. However, occurrence of multiple ICD shocks constitutes a medical emergency and warrants immediate medical attention by activating the emergency medical system. Patients should never drive to the hospital themselves after receiving a shock from their ICD.

Spontaneous arrhythmias, particularly those that are converted with a shock, are associated with a subsequent increased risk of death and hospitalization in patients with depressed ventricular function. The occurrence of an arrhythmia, therefore, warrants a reevaluation for possible decline in cardiac function, emergence of ischemia, or intercurrent illness.

If the ICD therapy is appropriate for VT or VF, consideration is given to whether therapy is warranted to reduce further episodes with either antiarrhythmic drug therapy or catheter ablation. Patients who have a rare episode of VT that is appropriately terminated and who have no other evidence of instability may not need any additional therapy, particularly if the VT is terminated by antitachycardia pacing rather than a shock. Shocks reduce quality of life and can lead to posttraumatic stress disorder. In many patients, the possibility of a shock can be reduced with appropriate ICD programming. Studies have shown that antitachycardia pacing effectively terminates >70% of VT episodes, even when VT is very rapid. Most ICDs can be programmed to attempt overdrive pace termination during capacitor charge. If the arrhythmia then terminates, the shock is aborted. Appropriate programming of antitachycardia pacing is therefore critical for reducing shocks. For patients implanted with ICDs as primary prevention, programming of VF detection zones >220 beats/min significantly reduces unnecessary and inappropriate shocks. Long detection times will also help avoid unnecessary therapies for VT episodes liable to terminate spontaneously.

Recurrent symptomatic episodes of VT or VF **(Fig. 263-2)** warrant specific therapy with antiarrhythmic drugs or ablation as discussed for the specific arrhythmia. The beta blockers sotalol and amiodarone are the most common pharmacologic options. Amiodarone combined with beta blockers is more effective than sotalol or beta blockers alone. It is important to recognize that although VT/VF episodes may represent a deterioration of clinical status in these patients, interventions to control the arrhythmia itself may have adverse effects on outcome. Most antiarrhythmic drugs have the potential to induce bradycardia to the point of requiring pacing from the ICD that, in itself, may have deleterious effects on ventricular function. Catheter ablation is an important option for patients with monomorphic VT.

MANAGEMENT OF THE PATIENT WITH ELECTRICAL STORM

Patients should be adequately sedated to allay anxiety and provide pain relief. Recurrent VT/VF is treated using standard Advanced Cardiac Life Support guidelines; treatment includes the use of medications such as beta blockers, amiodarone, and lidocaine with correction of any metabolic abnormalities. Recordings from electrocardiogram (ECG) monitoring or an implanted ICD are important to assess whether VT

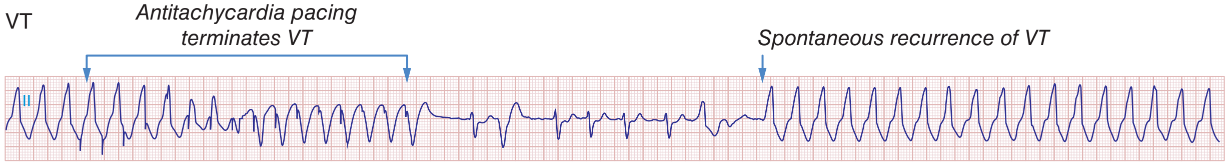
VT
Antitachycardia pacing terminates VT
Spontaneous recurrence of VT
II

FIGURE 263-1 Example of incessant monomorphic ventricular tachycardia (VT). In the initial portion of this electrocardiogram tracing, monomorphic VT is present. A train of antitachycardia pacing (area bracketed by *arrows*) that is initiated at the fourth VT complex results in ventricular capture with fusion by the eighth beat and termination of VT at cessation of pacing. The patient has underlying atrial fibrillation. Multifocal premature ventricular contractions are present. VT similar in morphology to the initial VT restarts spontaneously toward the latter part of the trace (*arrow*).

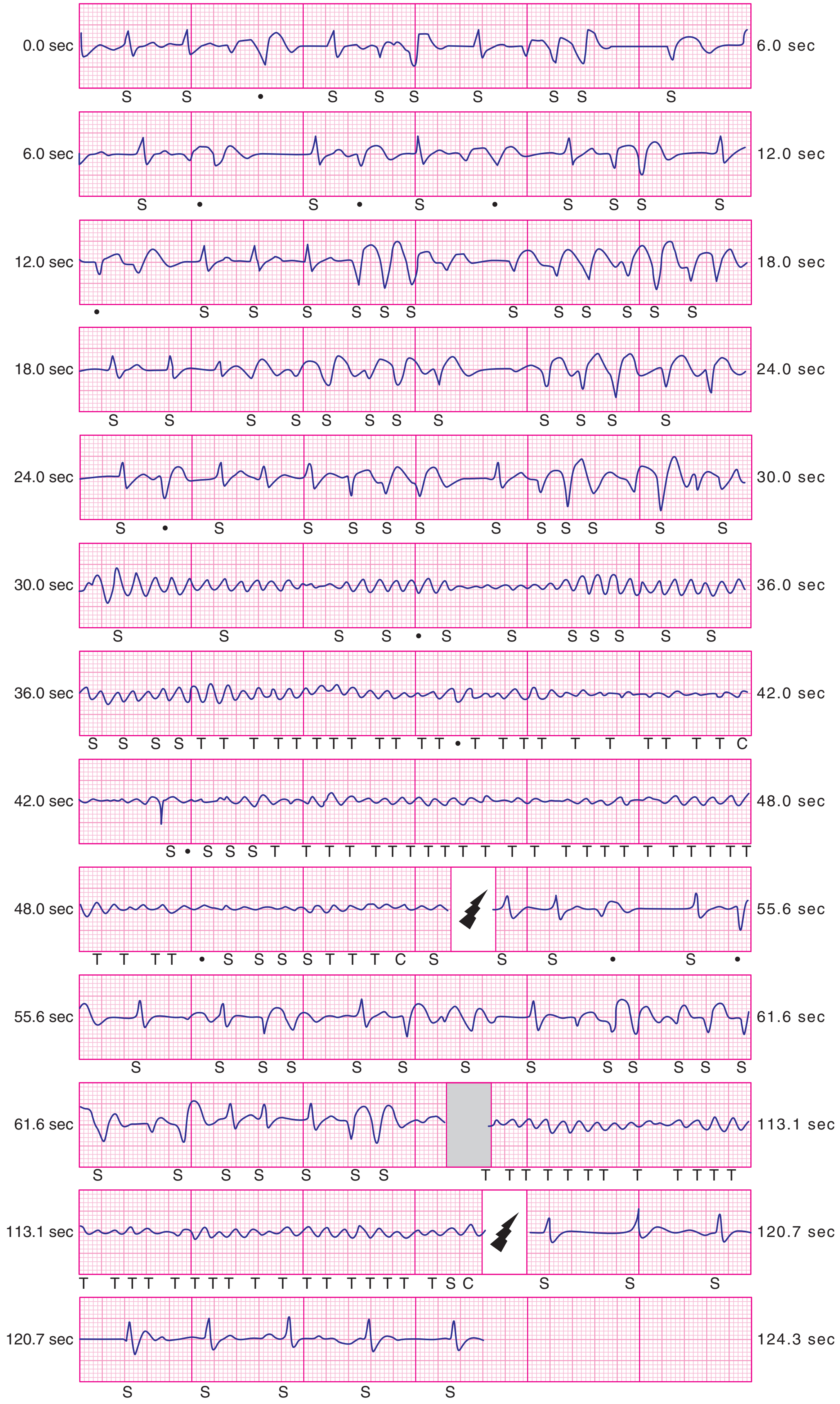
0.0 sec
6.0 sec
S S • S S S S S S S
6.0 sec
12.0 sec
S • S • S • S S S S
12.0 sec
18.0 sec
• S S S S S S S S S S S S
18.0 sec
24.0 sec
S S S S S S S S S S S S
24.0 sec
30.0 sec
S • S S S S S S S S S S S
30.0 sec
36.0 sec
S S S S • S S S S S S S
36.0 sec
42.0 sec
S S S S T T T T T T T T T T T T • T T T T T T T T T T C
42.0 sec
48.0 sec
S • S S S T T T T T T T T T T T T T T T T T T T T T T T T
48.0 sec
55.6 sec
T T T T • S S S S T T T C S S S • S •
55.6 sec
61.6 sec
S S S S S S S S S S S S S
61.6 sec
113.1 sec
S S S S S S S T T T T T T T T T T T
113.1 sec
120.7 sec
T T T T T T T T T T T T T T T T T S C S S S
120.7 sec
124.3 sec
S S S S S

FIGURE 263-2 Multiple implantable cardioverter-defibrillator (ICD) shocks from a subcutaneous ICD. Shown is a tracing from a patient with a subcutaneous ICD with recurrent episodes of ventricular fibrillation. The first five lines show gradually increasing amounts of ventricular ectopy and then ventricular fibrillation on the sixth line, which is terminated by a shock (*thunderbolt*) on the ninth line. The sequence repeats itself, and the patient receives a second shock that successfully terminates the arrhythmia.

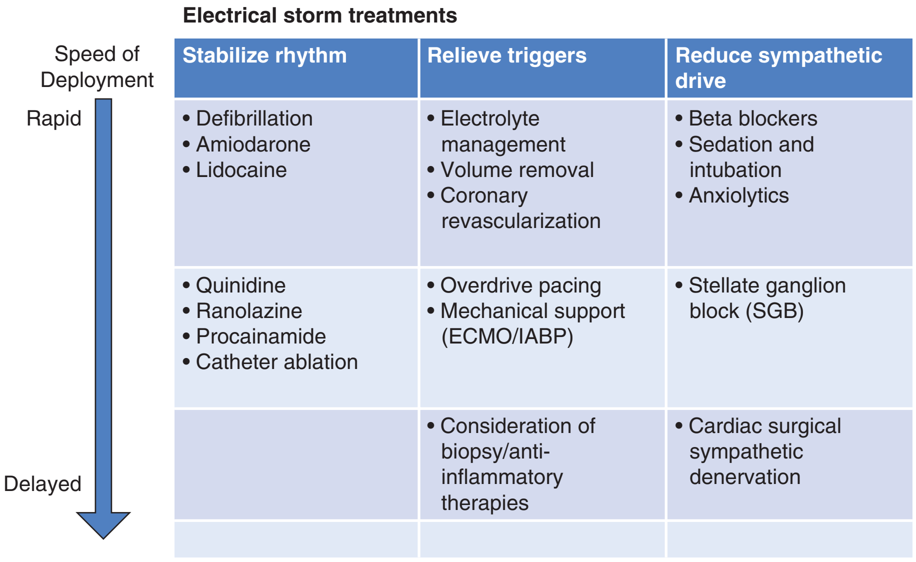
Electrical storm treatments
Speed of Deployment
Rapid
Delayed

Stabilize rhythm	Relieve triggers	Reduce sympathetic drive
• Defibrillation • Amiodarone • Lidocaine	• Electrolyte management • Volume removal • Coronary revascularization	• Beta blockers • Sedation and intubation • Anxiolytics
• Quinidine • Ranolazine • Procainamide • Catheter ablation	• Overdrive pacing • Mechanical support (ECMO/IABP)	• Stellate ganglion block (SGB)
	• Consideration of biopsy/anti-inflammatory therapies	• Cardiac surgical sympathetic denervation

FIGURE 263-3 Global strategy for managing electrical storm. Shown are considerations for stabilization of electrical storm with medication strategies and procedures. ECMO, extracorporeal membrane oxygenation; IABP, intra-aortic balloon pump.

is monomorphic or polymorphic. The initiation and termination of tachycardia in the stored ICD electrograms may also suggest possible precipitating or aggravating factors. Sedation or general anesthesia should be considered for suppression of recurrent hemodynamically unstable ventricular arrhythmia. Percutaneous stellate ganglion block and upper thoracic epidural anesthesia may reduce cardiac sympathetic outflow and have been used to restore stability in some patients. Rarely, mechanical ventricular support with extracorporeal membrane oxygenation, percutaneous left ventricular assist device, or intra-aortic balloon pump may be considered (Fig. 263-3).

In addition to this global strategy of stabilizing the heart rhythm, reducing sympathetic drive, and relieving any triggering mechanisms for the management of electrical storm, there are some specific therapies to be considered for patients with unique electrophysiologic substrate (Fig. 263-4).

VT/VF IN THE SETTING OF MYOCARDIAL ISCHEMIA

Ischemia should be considered especially if polymorphic VT or VF is identified as the primary arrhythmia. If electrical storm is occurring in the setting of an acute coronary syndrome, emergent revascularization and alleviation of anginal symptoms should be attempted. Within the infarcted myocardium, surviving Purkinje cells can exhibit triggered automaticity and lead to recurrent episodes of polymorphic VT/VF requiring frequent cardioversions before and after revascularization. Catheter ablation of premature ventricular contractions (PVCs) that are observed to repeatedly initiate the arrhythmia can be effective (Fig. 263-5).

PVC-INITIATED POLYMORPHIC VT/VF

Similar to the post–myocardial infarction electrical storm, patients without myocardial infarction or ischemia can have PVC-initiated polymorphic VT/VF storm. This idiopathic form of VF is usually caused by triggering PVCs originating from fascicular tissue or papillary muscles. Often, the ventricular ectopy is from scarred myocardial tissue detected on cardiac magnetic resonance imaging. Catheter ablation is indicated for this condition when antiarrhythmic medication is ineffective.

ACQUIRED OR CONGENITAL LONG QT SYNDROME

If QT prolongation causing torsades des pointes (TdP) is possible, intravenous magnesium should be administered for its immediate effect on repolarization. In addition, electrolyte repletion, especially potassium, should be aggressively pursued. Increasing the heart rate can sometimes normalize the QT interval, and thus, pharmacologic or pacing support should be considered. Isoproterenol can be used to increase a patient's sinus rate, but there is the possibility of increased ectopy with high doses of isoproterenol possibly exacerbating ventricular

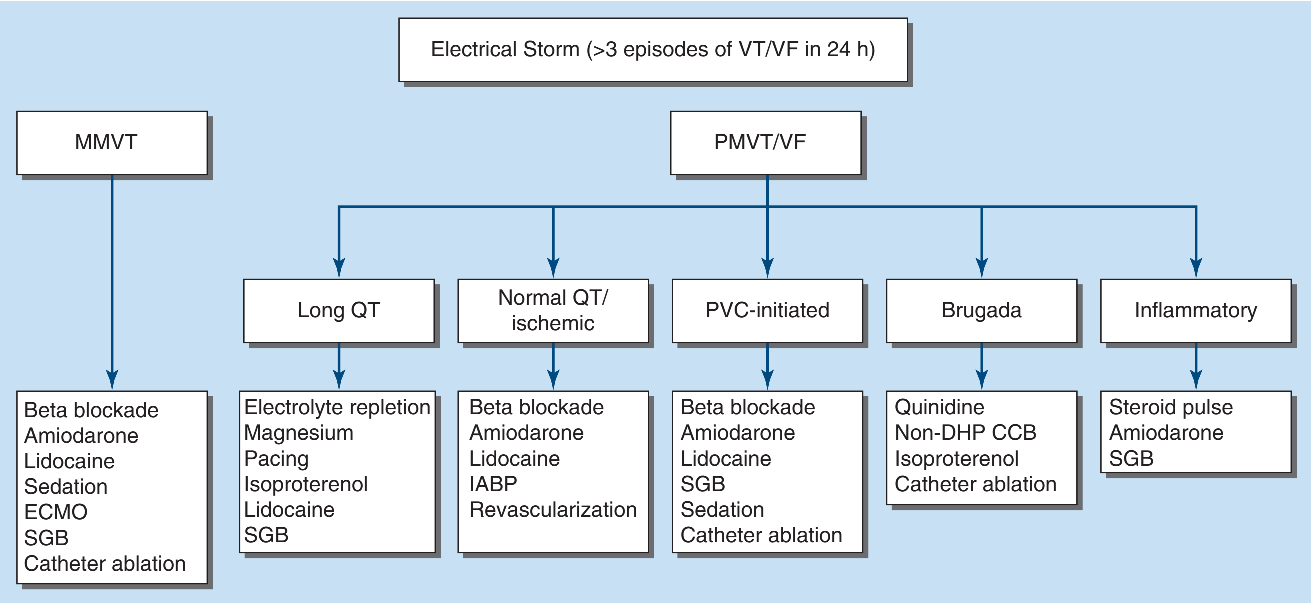
Electrical Storm (>3 episodes of VT/VF in 24 h)
MMVT
PMVT/VF
Long QT
Normal QT/ischemic
PVC-initiated
Brugada
Inflammatory
Beta blockade
Amiodarone
Lidocaine
Sedation
ECMO
SGB
Catheter ablation
Electrolyte repletion
Magnesium
Pacing
Isoproterenol
Lidocaine
SGB
Beta blockade
Amiodarone
Lidocaine
IABP
Revascularization
Beta blockade
Amiodarone
Lidocaine
SGB
Sedation
Catheter ablation
Quinidine
Non-DHP CCB
Isoproterenol
Catheter ablation
Steroid pulse
Amiodarone
SGB

FIGURE 263-4 Management algorithm for electrical storm. Shown is a suggested strategy for managing electrical storm based on the underlying rhythm and substrate. CCB, calcium channel blocker; DHP, dihydropyridine; ECMO, extracorporeal membrane oxygenation; IABP, intra-aortic balloon pump; MMVT, monomorphic ventricular tachycardia; PMVT, polymorphic ventricular tachycardia; PVC, premature ventricular contraction; SGB, stellate ganglion block; VF, ventricular fibrillation; VT, ventricular tachycardia.

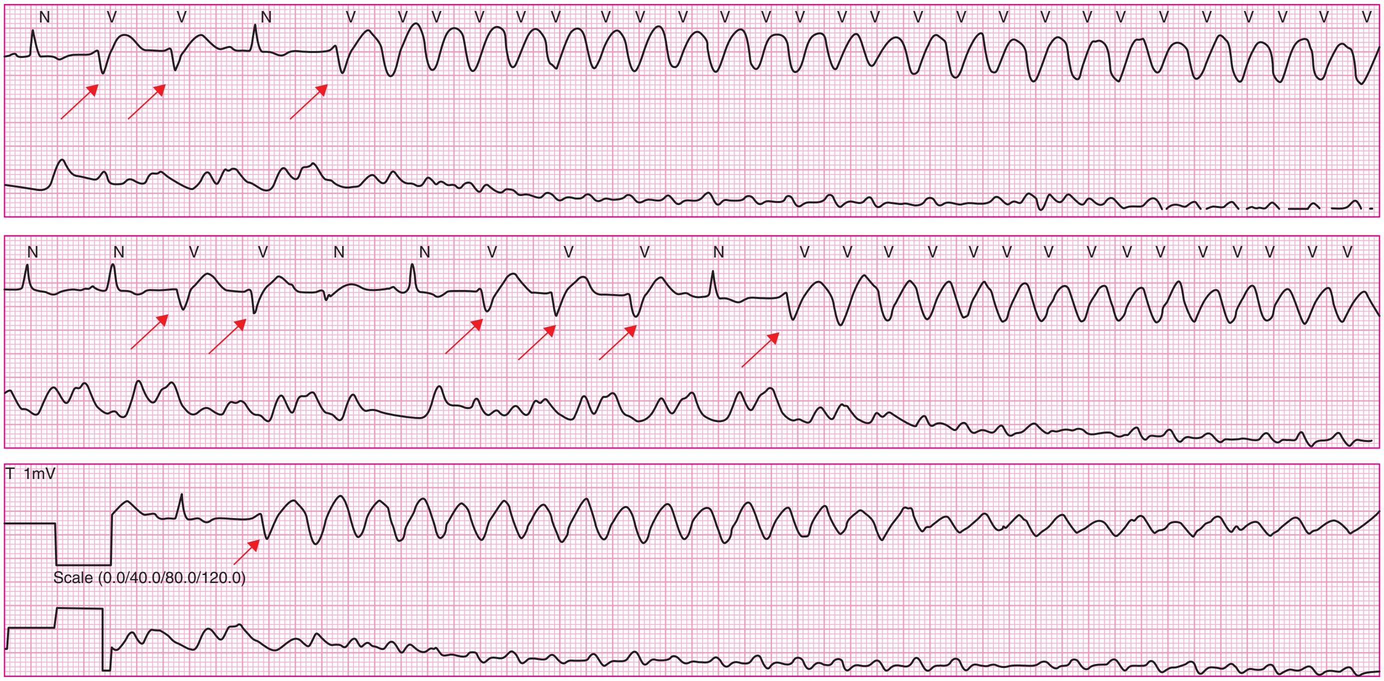

FIGURE 263-5 Premature ventricular contraction (PVC)–triggered ventricular fibrillation (VF) after myocardial infarction electrical storm. Shown is a series of monitoring strips from a patient with VF occurring in a PVC-triggered fashion after myocardial infarction. A single electrocardiogram lead and the blood pressure tracings are shown. The triggering PVCs are indicated with *red arrows*. The VF results in prompt hemodynamic collapse as evidenced by the blood pressure tracing.

arrhythmias. Although lidocaine can reduce the QT interval, other antiarrhythmic agents should be avoided because of their effect on repolarization.

BRUGADA SYNDROME

If the QT interval is not prolonged and a Brugada pattern of Rsr′ with ST elevation in leads V_1 or V_2 is seen on resting ECG, administration of quinidine and/or isoproterenol may abolish recurrent polymorphic VT/VF episodes. Nondihydropyridine calcium channel blockers and isoproterenol have also been used to reduce arrhythmic events. An epicardial substrate-based catheter ablation over the right ventricular outflow tract has been described as a strategy for drug-refractory ventricular tachyarrhythmias in Brugada syndrome.

INFLAMMATORY CARDIOMYOPATHY

If the patient has no known previous cardiac disease, consideration should be given to an inflammatory myocarditis causing the frequent ventricular arrhythmias. Giant cell myocarditis, cardiac sarcoidosis, and certain viral myocarditis can present with VT/VF storm. An endomyocardial biopsy should be considered to potentially identify new-onset inflammatory cardiomyopathies that may require urgent anti-inflammatory therapy. Once the acute episode is controlled, strategies to prevent recurrent VT or VF should be considered.

FURTHER READING

Callans DJ: *Josephson's Clinical Cardiac Electrophysiology: Techniques and Interpretations*, 7th ed. Philadelphia, Wolters Kluwer, 2024.

Jalife J, Stevenson W (eds): *Zipes and Jalife's Cardiac Electrophysiology: From Cell to Bedside*, 8th ed. Philadelphia, Elsevier, 2022.

Zeppenfeld K et al: 2022 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: Developed by the task force for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death of the European Society of Cardiology (ESC) Endorsed by the Association for European Paediatric and Congenital Cardiology (AEPC). Eur Heart J 43:3997, 2022.

Section 4 Disorders of the Heart, Muscles, Valves, and Pericardium

264

Heart Failure: Pathophysiology and Diagnosis

Michael M. Givertz, Mandeep R. Mehra

CLINICAL DEFINITIONS, EPIDEMIOLOGY, AND PHENOTYPES

DEFINITIONS

Heart failure (HF) is a common final pathway for most chronic cardiovascular diseases including hypertension, coronary artery disease, and valvular heart disease. The American Heart Association/American College of Cardiology/Heart Failure Society (AHA/ACC/HFSA) guideline defines HF as a complex clinical syndrome with symptoms and signs that result from any structural or functional impairment of ventricular filling or ejection of blood. The European Society of Cardiology's (ESC) definition emphasizes cardinal symptoms (e.g., breathlessness, ankle swelling, and fatigue) that may be accompanied by signs (e.g., elevated jugular venous pressure, pulmonary crackles, and peripheral edema) due to a structural and/or functional abnormality of the heart that results in elevated intracardiac pressures and/or inadequate cardiac output at rest and/or during exercise. An international consensus conference proposed a universal definition of HF that is comprehensive and practical enough to encompass formal disease stages, with universal applicability, prognostic and therapeutic validity, and acceptable sensitivity and specificity **(Fig. 264-1)**.

Because some patients present without signs or symptoms of volume overload, the term *heart failure* is preferred over the older term

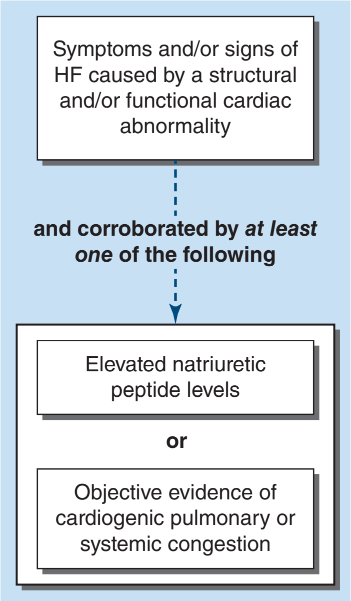
Symptoms and/or signs of HF caused by a structural and/or functional cardiac abnormality
and corroborated by *at least one* of the following
Elevated natriuretic peptide levels
or
Objective evidence of cardiogenic pulmonary or systemic congestion

FIGURE 264-1 Universal definition of heart failure (HF). This contemporary universal definition of HF is simple but conceptually comprehensive, with near universal applicability, prognostic and therapeutic validity, and acceptable sensitivity and specificity. *(Reproduced with permission from B Bozkurt et al: Universal Definition and Classification of Heart Failure. J Card Fail 27:387, 2021.)*

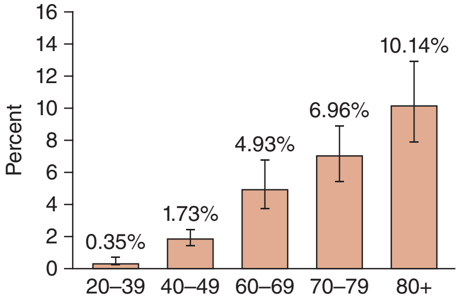

FIGURE 264-2 Heart failure prevalence by age categories. Prevalence of heart failure among U.S. adults ≥20 years of age by age, from the National Health and Nutrition Examination Survey (NHANES), 2017–2020. *(Reproduced with permission from B Bozkurt et al: J Card Fail 29:1412, 2023.)*

congestive heart failure. *Cardiomyopathy* and *left ventricular dysfunction* are more general terms that describe disorders of myocardial structure and/or function, which may lead to HF. In pathophysiologic terms, HF has been defined as a syndrome characterized by elevated cardiac filling pressure and/or inadequate peripheral oxygen delivery, at rest or during stress, caused by cardiac dysfunction.

Chronic heart failure describes patients with longstanding (e.g., months to years) symptoms and/or signs of HF typically treated with medical and device therapy as described in **Chap. 265**. Such patients are at risk of *worsening heart failure,* and when the episode resolves, the use of the term *remission* rather than *stable* HF is preferred, since these patients continue to retain the risk for further decompensation and sudden death. *Acute heart failure*, previously termed acute decompensated HF, refers to the rapid onset or worsening of symptoms of HF. Most episodes of acute HF result from worsening of chronic HF, but ~20% are due to new-onset HF that can occur in the setting of acute coronary syndrome, acute valvular dysfunction, hypertensive urgency, or postcardiotomy syndrome. Similarly, *acute pulmonary edema* in HF describes a clinical scenario in which a patient presents with rapidly worsening signs and symptoms of pulmonary congestion, typically due to severe elevation of left heart filling pressures.

EPIDEMIOLOGY

Global Incidence and Prevalence HF is a major cause of morbidity and mortality worldwide. An estimated 6.7 million American adults are being treated for HF, with >600,000 new cases diagnosed each year. Globally, it is estimated that 56.2 million people are living with HF with prevalence varying greatly by country. The prevalence of HF increases significantly with age, occurring in 1–2% of the population aged 40–49 years and 10% or more in adults >80 years old **(Fig. 264-2)**. The lifetime risk of HF has increased to 24%; approximately one in four persons will develop HF in their lifetime. Projections based on National Health and Nutrition Examination Survey and U.S. Census Bureau data show that the prevalence of HF is expected to rise to 8.5 million Americans by 2030. According to the Nationwide Readmission Database, rates of HF hospitalizations in the United States declined from 2010 to 2014, followed by an increase from 2014 to 2017. HF readmissions after index hospitalization followed a similar trend. While prevalence of HF continues to rise, incidence may be decreasing due to improved recognition and treatment of cardiovascular disease and its comorbidities as well as disease prevention. However, as rates of obesity rise globally, these favorable trends in HF incidence may reverse.

There are distinct racial and ethnic differences in HF epidemiology **(Fig. 264-3)**. In community-based studies, black individuals have the highest risk of developing HF, followed by Hispanic, white, and Chinese Americans. These differences are attributed to disparities in cardiometabolic risk factors (e.g., obesity, hypertension, diabetes) as well as social determinants of health including socioeconomic status and access to health care. Similarly, studies have shown that age-adjusted rates of HF hospitalization are highest for black men, followed by black women, white men, and white women. Accurate data on HF prevalence from emerging nations are lacking. As developing nations undergo socioeconomic development, the epidemiology of HF is becoming like that of Western Europe and North America, with coronary artery disease emerging as the most common cause of HF, although hypertension remains the highest population attributable risk for HF occurrence.

Morbidity and Mortality In primary care, the overall 5-year survival following the diagnosis of HF is ~50%. For patients with severe HF, the 1-year mortality may be as high as 40%. In the United States, one in eight deaths list HF on the death certificate. The majority of these patients die of cardiovascular causes, most commonly progressive HF or sudden cardiac death. A number of clinical and laboratory parameters are independent predictors of mortality **(Table 264-1)**. In a population-based study, hospitalizations were common after an HF diagnosis, with 83% hospitalized at least once, and 67%, 54%, and 43% hospitalized at least two, three, and four times, respectively. Following an HF admission, mortality rates range from 8–14% at 30 days to 26–37% at 1 year to up to 75% at 5 years. Readmission with HF is also common, ranging from 20–25% at 60 days to nearly

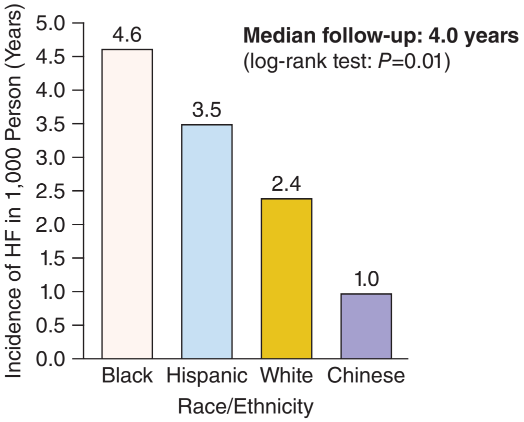

FIGURE 264-3 Incidence of heart failure (HF). HF incidence rates by race/ethnicity in the United States. *(Reproduced with permission from IL Pina et al: J Am Coll Cardiol 78:2589, 2021.)*

TABLE 264-1 Independent Predictors of Adverse Outcomes in Heart Failure

Clinical	Male sex Older age Diabetes mellitus Chronic kidney disease Coronary artery disease Advanced NYHA class[a] Presence of third heart sound or elevated JVP Decreased exercise capacity Cardiac cachexia Depression
Structural	Reduced left ventricular ejection fraction Reduced right ventricular ejection fraction Increased ventricular volumes and mass Secondary mitral or tricuspid regurgitation
Hemodynamic	Elevated pulmonary capillary wedge pressure Reduced cardiac index Reduced peak oxygen consumption Pulmonary hypertension Diastolic dysfunction
Biochemical	Worsening renal function Hyponatremia Hyperuricemia Elevated cardiac biomarkers (troponin and natriuretic peptides) Elevated plasma neurohormones (norepinephrine, renin, aldosterone, and endothelin-1)
Electrophysiologic	Tachycardia Widened QRS interval or LBBB Atrial fibrillation Ventricular ectopic activity Ventricular tachycardia and sudden death

[a] See Table 264-4.

Abbreviations: JVP, jugular venous pressure; LBBB, left bundle branch block; NYHA, New York Heart Association.

TABLE 264-2 Selected Causes of Heart Failure

Heart Failure with Reduced Ejection Fraction	
Coronary artery disease Myocardial infarction Myocardial ischemia	Nonischemic cardiomyopathy Infiltrative disorders Familial disorders Tachycardia induced
Valvular heart disease Aortic stenosis or regurgitation Mitral or tricuspid regurgitation	Toxic cardiomyopathy Chemotherapy, immunotherapy Drugs such as hydroxychloroquine Alcohol, cocaine
Congenital heart disease Intracardiac shunts Repaired defects Systemic right ventricular failure	Chronic lung/pulmonary vascular disease Cor pulmonale Pulmonary arterial hypertension
Infectious Chagas HIV	Autoimmune disease Giant cell myocarditis Lupus myocarditis
Heart Failure with Preserved Ejection Fraction	
Hypertension	Coronary artery disease
Valvular heart disease Aortic stenosis Mitral stenosis	Restrictive cardiomyopathy Amyloidosis Sarcoidosis Hemochromatosis Glycogen storage disease
Hypertrophic cardiomyopathy	Radiation therapy
Constrictive pericarditis	Aging
Myocarditis	Endomyocardial fibroelastosis
Obesity	End-stage renal disease
High-Output Heart Failure	
Thyrotoxicosis	Arteriovenous shunt
Obesity	Cirrhosis
Anemia	Vitamin B deficiency (beriberi)
Chronic lung disease	Myeloproliferative disorder

Abbreviation: HIV, human immunodeficiency virus.

50% at 6 months. With each subsequent admission, the risk of death rises. There are racial disparities in outcomes, with black patients having higher case–fatality rates compared to white patients. Men have higher age-adjusted mortality rates for death related to HF than women; and in the United States, mortality rate from HF varies by region (highest in Midwest) and population density (highest in rural areas). Despite these statistics, the overall prognosis for patients with HF is improving due to treatment of risk factors and increased use of guideline-directed therapies.

Costs The overall cost of HF care is high (estimated $22.3 billion in the United States in 2018) and rising. Projections for 2030 are that hospitalization costs for HF in the United States will increase to $70 billion. Indirect costs due to lost work and productivity may equal or exceed this amount. The global economic burden of HF in 2012 was estimated at $108 billion, with direct costs accounting for 60%. For pediatric patients with acute HF, inpatient costs are estimated at $1 billion annually and rising.

PHENOTYPES AND CAUSES

HF with Reduced Versus Preserved Ejection Fraction Epidemiologic studies have shown that approximately one-half of patients who develop HF have reduced left ventricular ejection fraction (EF; ≤40%), while the other half have near normal or preserved EF (≥50%) or are classified as having HF with mildly reduced EF (41–49%). Because most patients with HF (regardless of EF) have abnormalities in both systolic and diastolic function, the older terms of *systolic heart failure* and *diastolic heart failure* have been abandoned. Classifying patients based on their EF (HF with reduced EF [HFrEF], HF with mildly reduced EH [HFmrEF], or HF with preserved EF [HFpEF]) is important due to differences in demographics, comorbidities, and response to therapies **(Chap. 265)**. Underlying causes of HF may be associated with reduced or preserved EF and include disorders of the coronary arteries, myocardium, pericardium, heart valves, and great vessels **(Table 264-2)**. The diagnosis of HFpEF is often more challenging due to the need to rule out noncardiac causes of shortness of breath and/or fluid retention.

HF with Recovered EF A subgroup of patients who are diagnosed with HFrEF and treated with guideline-directed therapy have rapid or gradual improvement in EF to the normal range and are referred to as having HF with recovered EF (HFrecEF). Predictors of HFrecEF include younger age, shorter duration of HF, nonischemic etiology, smaller ventricular volumes, and absence of myocardial fibrosis. Specific clinical examples include fulminant myocarditis, stress cardiomyopathy, peripartum cardiomyopathy, and tachycardia-induced cardiomyopathy, as well as reversible toxin exposures such as chemotherapy, immunotherapy, or alcohol. Despite recovery of EF, patients may remain symptomatic due to persistent abnormalities in diastolic function, exercise-induced pulmonary hypertension, or related comorbidities (e.g., obesity). For patients who become asymptomatic, withdrawal of disease-modifying therapy can lead to recurrence of HF symptoms and decrease in EF in up to half of such patients within 6 months. In general, prognosis of patients with HFrecEF is superior to that of patients with either HFrEF or HFpEF.

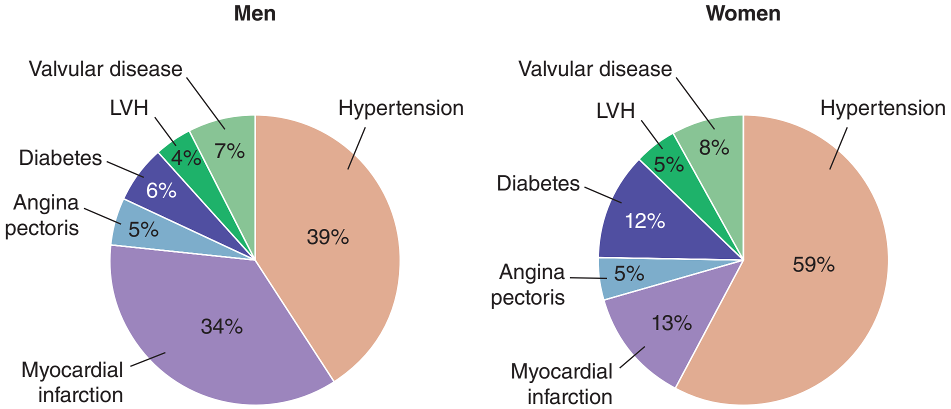

FIGURE 264-4 Population attributable risk of heart failure (HF) incidence. Based on longitudinal data from the Framingham Heart Study, the risk factors contributing most significantly to the population attributable risk (PAR) of HF in men were previous myocardial infarction and hypertension (in men, both represented equal contributions to HF PAR). In contrast, hypertension was the risk factor accounting for the majority of total PAR in women. In women, previous myocardial infarction accounted for only 13% of the PAR of HF compared with 34% in men. PAR values are developed based on individual calculations for each variable using hazard ratio and prevalence statistics. Thus, they may not, in aggregate, equal 100%. LVH, left ventricular hypertrophy. *(Reproduced with permission from Givertz MM and Colucci WS. Heart failure. In: Libby P, editor. Essential Atlas of Cardiovascular Disease. Philadelphia: Current Medicine; 2009.)*

HF with Mildly Reduced EF (HFmrEF) Patients with HF and an EF between 40 and 50% represent an intermediate group that are often treated for risk factors and comorbidities and with guideline-directed medical therapy similar to patients with HFrEF. They are felt to have primarily mild systolic dysfunction, but with features of diastolic dysfunction. They may also include either patients with reduced EF who experience partial improvement in their EF or those with initially preserved EF who suffer a decline in their systolic performance. The AHA/ACC/HFSA guideline requires evidence of spontaneous or provokable increased left ventricular filling pressures in their classification of HFmrEF.

Acquired Versus Familial, Congenital, and Other Disorders In developed countries, coronary artery disease is responsible for approximately two-thirds of the cases of HF, with hypertension as a principal contributor in up to 75% and diabetes mellitus in 10–40% (Fig. 264-4). Notably, population attributable risk for hypertension, obesity, diabetes mellitus, and coronary artery disease varies according to sex, race, and ethnicity (Fig. 264-5). While most cardiovascular disease underlying HF is acquired in mid and later life (Chaps. 272, 284, and 288), a wide range of congenital and inherited disorders leading to HF may be diagnosed in children and younger adults. It is currently estimated that 13.3 million people globally and approximately 467,000 U.S. adults are living with congenital heart disease (CHD). In general, adults with CHD who develop HF can be divided into one of three pathophysiologic groups: uncorrected defects with late presentation due to missed diagnosis, nonintervention, or lack of access to care; repaired or palliated defects with late valvular and/or ventricular failure; or failing single-ventricle physiology. In addition, each adult with CHD often presents with unique anatomic and physiologic challenges that affect HF and its treatment.

Inherited cardiomyopathies are also increasingly recognized in adults presenting with HF. These include more common disorders, such as hypertrophic and arrhythmogenic cardiomyopathies, and lesser known heart muscle disease related to pathogenic variants in genes encoding lamin and titin, muscular dystrophies, and mitochondrial disease. Most forms of familial cardiomyopathy are inherited in an autosomal dominant fashion. Society guidelines have been published documenting the importance of taking a detailed three-generational family history and indications for (and limitations of) clinical genetic testing.

A myriad of systemic diseases with cardiac and extracardiac manifestations (e.g., amyloidosis, sarcoidosis), autoimmune disorders (e.g., systemic lupus erythematosus, rheumatoid arthritis), infectious diseases (e.g., Chagas, HIV), and drug toxicities (chemotherapy, other prescribed or illicit agents) can result in HF with either reduced or preserved EF. In Africa and Asia, rheumatic heart disease remains a major cause of HF, especially in the young. Finally, disorders associated with a high cardiac output state (e.g., anemia, thyrotoxicosis) are seldom associated with HF in the absence of underlying structural heart disease. However, diagnosis and treatment of high-output HF will be missed if not considered in the differential diagnosis of patients with predisposing conditions (e.g., severe obesity, severe anemia, cirrhosis, end-stage renal disease with arteriovenous fistula, Paget's disease, or nutritional deficiency such as beriberi).

PATHOPHYSIOLOGY

PROGRESSIVE DISEASE

HFrEF is a progressive disease that typically involves an index event followed by months to years of structural and functional cardiovascular remodeling (Fig. 264-6). The primary event may be sudden in onset, such as an acute myocardial infarction; more gradual, as occurs in the setting of chronic pressure or volume overload (with valvular heart disease); inherited, as seen with genetic cardiomyopathies; or congenital disease in origin. Despite an initial reduction in cardiac performance, patients may be asymptomatic or mildly symptomatic for prolonged periods due to the activation of compensatory mechanisms (described below) that ultimately contribute to disease progression.

Ventricular Remodeling As demonstrated in both animal and human studies, different patterns of ventricular remodeling occur in response to excess cardiac workload. Concentric hypertrophy, in which increased mass is out of proportion to chamber volume, effectively reduces wall stress under conditions of pressure overload (e.g., hypertension, aortic stenosis). By contrast, an increase in cavity size or volume (eccentric hypertrophy) occurs in volume overload conditions (e.g., aortic regurgitation, mitral regurgitation). In both forms of remodeling, an increase in ventricular mass is accompanied by changes at the *cellular level* with myocyte hypertrophy and interstitial fibrosis, at the *protein level* with alterations in calcium-handling and cytoskeletal protein abundance and/or function, and at the *molecular level* by reexpression of fetal genes (Table 264-3). In addition to cell loss from necrosis, myocytes that are unable to adapt to remodeling stimuli may be triggered to undergo apoptosis or programmed cell death. Further impairment in pump function and increased wall stress in the face of systemic vasoconstriction and loss of neurohormonal adaptation (discussed below) can lead to afterload mismatch. These events feed back on remodeling stimuli, setting up a cycle of deleterious processes resulting in clinical HF.

While our understanding of ventricular remodeling in HFrEF is well supported by animal and human studies, the mechanisms underlying HFpEF are less clear. The original descriptions of HFpEF focused on diastolic dysfunction as the primary mediator of HF

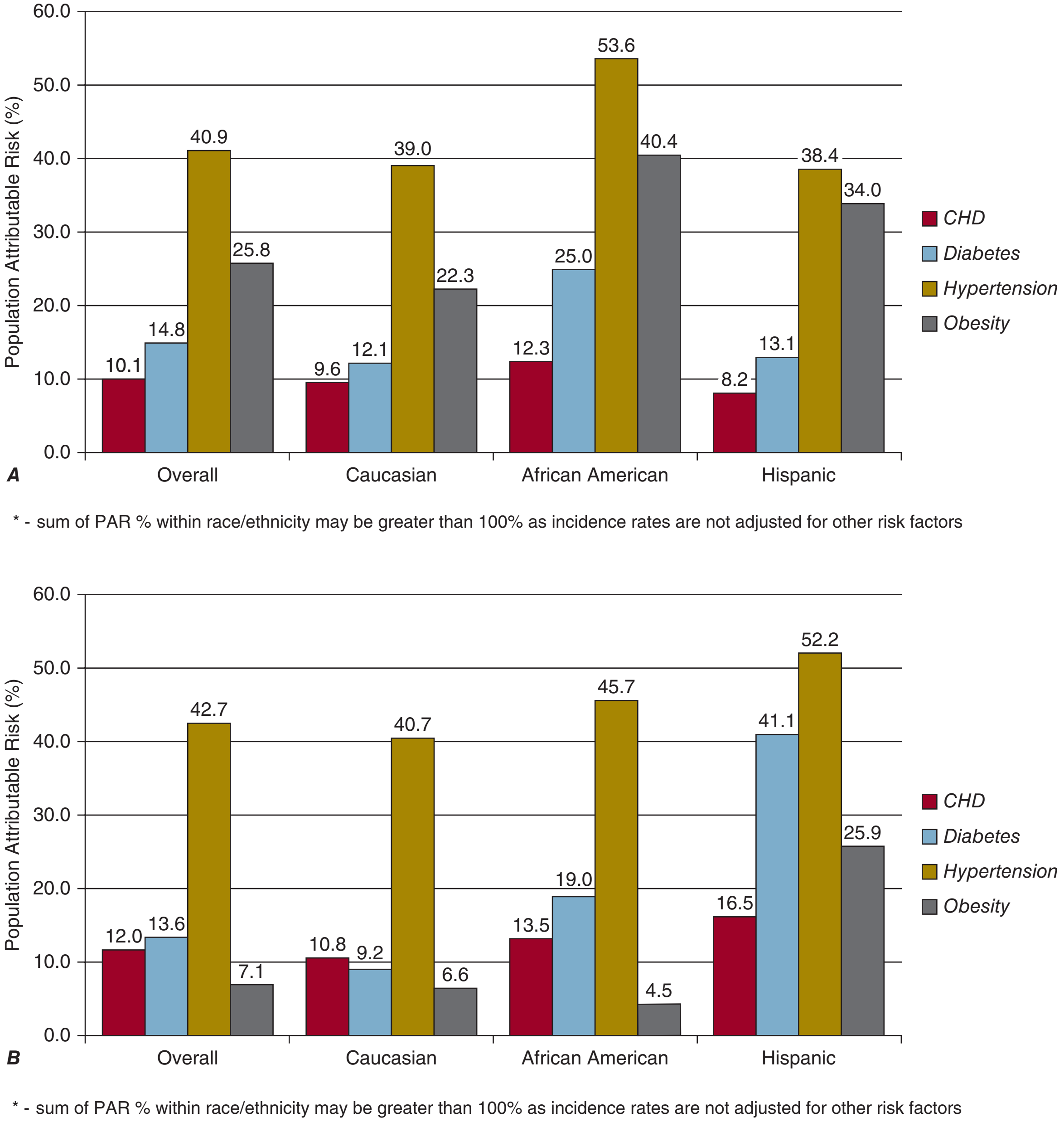

FIGURE 264-5 Population attributable risk (PAR) for heart failure (HF) by race and ethnicity. PAR* by race and ethnicity for (***A***) HF with preserved ejection fraction (HFpEF) and (***B***) HF with reduced ejection fraction (HFrEF). *Sum of PAR% within race/ethnicity may be >100% as incidence rates are not adjusted for other risk factors. CHD, coronary heart disease. *(Reproduced with permission from CB Eaton, M Pettinger, J Rossouw, LW Martin et al: Risk factors for incident hospitalized heart failure with preserved versus reduced ejection fraction in a multiracial cohort of postmenopausal women. Circ Heart Fail 9(10): e002883, 2016.)*

signs and symptoms as exemplified in older women with hypertension. At the myocyte level, impaired uptake of cytosolic calcium into the sarcoplasmic reticulum by reductions in adenosine triphosphate explained abnormalities in myocardial relaxation. As different phenotypes of HFpEF have emerged, many pathophysiologic processes that work in concert with each other, beyond diastolic dysfunction, have been implicated in disease progression, including vascular stiffness, renal dysfunction, sodium avidity, and inflammation related to regional adiposity. Furthermore, biologic alterations including oxidative stress, impaired nitric oxide signaling leading to nitrosative stress, and insulin resistance may play a role in disease activity and inform future therapies. Autophagy is a natural process that removes unwanted cellular components by forming autophagosomes that fuse with lysosomes. While autophagy is deemed cytoprotective and adaptive, unchecked induction of autophagy may be maladaptive and a target by which disease-modifying therapy may exert beneficial effects.

MECHANISMS OF DISEASE PROGRESSION

Several compensatory mechanisms become activated during the development of HF and contribute to disease progression. Our understanding of these mechanisms derives from preclinical studies, in vivo human studies, and randomized clinical trials demonstrating benefit of therapies targeted to attenuating or reversing these biologic processes.

Neurohormonal Activation Activation of the sympathetic nervous system (SNS) and renin-angiotensin-aldosterone system (RAAS) plays a critical role in the development and progression of HF. Initially, neurohormonal activation leads to increases in heart rate, blood pressure, and cardiac contractility and retention of sodium and water to augment preload and maintain cardiac output at rest and during exercise. Over time, these unchecked compensatory responses lead to excessive vasoconstriction and volume retention, electrolyte and renal abnormalities, baroreceptor dysfunction, direct myocardial toxicity, and cardiac arrhythmias. At the tissue level, neurohormonal activation

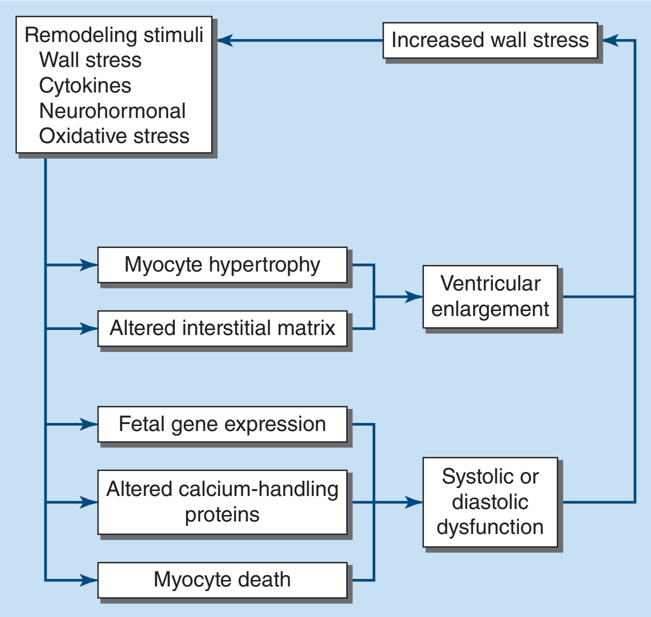

FIGURE 264-6 Remodeling stimuli in heart failure. Chronic hemodynamic stimuli such as pressure and volume overload lead to ventricular remodeling through increases in myocardial wall stress, inflammatory cytokines, signaling peptides, neuroendocrine signals, and oxidative stress. The myocardium responds with adaptive as well as maladaptive changes. Reexpression of fetal contractile proteins and calcium handling proteins may contribute to impaired contraction and relaxation. Myocytes unable to adapt might be triggered to undergo programmed cell death (apoptosis). The net result of these changes is further impairment in pump function and increased wall stress, thus completing a vicious cycle that leads to further progression of myocardial dysfunction. *(Reproduced with permission from Givertz MM and Colucci WS. Heart failure. In: Libby P, editor. Essential Atlas of Cardiovascular Disease. Philadelphia: Current Medicine; 2009.)*

contributes to remodeling of the heart, blood vessels (atherosclerosis), kidneys, and other organs **(Fig. 264-7)** and the development of symptomatic HF. Landmark clinical trials in HF have demonstrated that antagonism of the RAAS and SNS with renin-angiotensin system inhibitors, mineralocorticoid receptor antagonists, and beta blockers attenuates or reverses ventricular and vascular remodeling and reduces morbidity and mortality **(Chap. 265)**.

Vasodilatory Hormones While RAAS and SNS activation contributes to disease progression in HF, a number of counterregulatory hormones are upregulated and exert beneficial effects on the heart, kidney, and vasculature. These include the natriuretic peptides (atrial natriuretic peptide [ANP] and B-type natriuretic peptide [BNP]), prostaglandins (prostaglandin E$_1$ [PGE$_1$] and prostacyclin [PGI$_2$]), bradykinin, adrenomedullin, and nitric oxide. ANP and BNP are stored and released primarily from the atria and ventricles, respectively, in response to increased stretch or pressure. Beneficial actions are mediated through stimulation of guanylate cyclase and include systemic and pulmonary vasodilation, increased sodium and water excretion, inhibition of renin and aldosterone, and baroreceptor modulation. Bradykinin and natriuretic peptides are inactivated by neprilysin, a membrane-bound peptidase, which explains in part the beneficial clinical impact of angiotensin receptor–neprilysin inhibition in HF **(Chap. 265)**. As described below, natriuretic peptide levels can be used to assist in the diagnosis and risk stratification of patients with HF.

TABLE 264-3 Mechanisms of Ventricular Remodeling

Changes in Myocyte Biology
Abnormal excitation-contraction coupling and crossbridge interaction
Fetal gene expression (e.g., β-myosin heavy chain)
β-Adrenergic receptor desensitization
Myocyte hypertrophy
Impaired cytoskeletal proteins
Changes in Myocardial Composition
Myocyte necrosis, apoptosis, and autophagy
Interstitial and perivascular fibrosis
Matrix degradation
Changes in Ventricular Geometry
Ventricular dilation and wall thinning
Increased sphericity and displacement of papillary muscles
Atrioventricular valve regurgitation

Endothelin, Inflammatory Cytokines, and Oxidative Stress Endothelin is a potent vasoconstrictor peptide with growth-promoting effects that may play an important role in pulmonary hypertension and right ventricular failure. Endothelin is released from a variety of vascular and inflammatory cells within the pulmonary circulation and myocardium in response to increased pressure and has direct deleterious effects on the heart, leading to myocyte hypertrophy and interstitial fibrosis. Unlike RAAS and SNS inhibition, however, endothelin blockade has not been shown to slow the progression of clinical HF due to left ventricular failure but is beneficial for treatment of pulmonary arterial hypertension and consequent right HF **(Chap. 294)**. Other factors that have the potential to cause or contribute to ventricular remodeling in HF include inflammatory cytokines such as tumor necrosis factor (TNF) α and interleukin (IL) 1β and reactive oxygen species such as superoxide and peroxynitrite. Potential sources of these biologically active substances are the liver and gastrointestinal tract, as described below. The role of anti-inflammatory and antioxidant therapies remains unproven.

Novel Biologic Targets Sodium-glucose cotransporter 2 (SGLT-2) is a protein located on the proximal tubule of the kidney that is responsible for reabsorption of up to 90% of filtered glucose. In patients with HF, activity of SGLT-2 contributes to sodium and water retention, endothelial dysfunction, abnormal myocardial metabolism, and impaired calcium handling. Inhibitors of SGLT-2 were developed for the treatment of type 2 diabetes mellitus to take advantage of their glycosuric and metabolic effects **(Chap. 416)**. Subsequent large clinical trials in cardiovascular disease including HF (with or without diabetes mellitus) have demonstrated not only safety of these agents (as required by the U.S. Food and Drug Administration) but also, more importantly, beneficial effects on morbidity and mortality. Whether benefits of SGLT-2 inhibitors in HF are due primarily to diuretic effects or to effects on cardiac and vascular remodeling, proarrhythmia, renal function, and/or metabolic function, inflammation or dysregulated autophagy remains to be conclusively determined. Another pathway that is downregulated in HF and contributes to endothelial dysfunction involves cyclic guanosine monophosphate (cGMP). Oral soluble guanylate cyclase stimulators enhance the cGMP pathway and exert beneficial myocardial and vascular effects in experimental and clinical HF.

Dyssynchrony and Electrical Instability In up to one-third of patients with HF, disease progression is associated with prolongation of the QRS interval. Electrical dyssynchrony in the form of left bundle branch block (LBBB) or intraventricular conduction delay results in abnormal ventricular contraction. As discussed in **Chap. 265**, correction of electrical dyssynchrony with left or biventricular pacing can improve contractile function, decrease mitral regurgitation, and reverse ventricular remodeling. In patients with symptomatic HFrEF and LBBB on guideline-directed medical therapy, cardiac resynchronization therapy or LBBB (or His bundle) pacing is suggested to reduce morbidity and mortality. Other forms of electrical instability, including atrial fibrillation with inadequate rate control and frequent premature ventricular complexes, can also contribute to worsening HF. In addition to the direct impact of tachycardia and irregular rhythm on disease progression, the link between these arrhythmias and cardiac

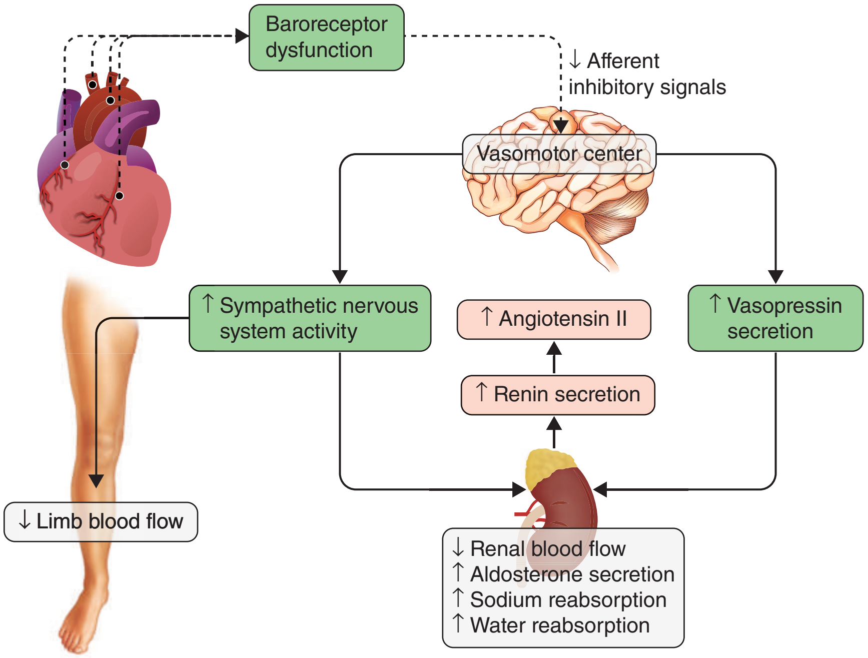

FIGURE 264-7 Activation of neurohormonal systems in heart failure. Decreased cardiac output in heart failure (HF) results in an "unloading" of high-pressure baroreceptors (*circles*) in the left ventricle, carotid sinus, and aortic arch, which in turn causes reduced parasympathetic tone. This decrease in afferent inhibition results in a generalized increase in efferent sympathetic tone and nonosmotic release of arginine vasopressin from the pituitary. Vasopressin is a powerful vasoconstrictor that also leads to reabsorption of free water by the kidney. Afferent signals to the central nervous system also activate sympathetic innervation of the heart, kidney, peripheral vasculature, and skeletal muscles. Sympathetic stimulation of the kidney leads to the release of renin, with a resultant increase in circulating levels of angiotensin II and aldosterone. The activation of the renin-angiotensin-aldosterone system promotes salt and water retention, peripheral vasoconstriction, myocyte hypertrophy, cell death, and myocardial fibrosis. Although these neurohormonal mechanisms facilitate short-term adaptation by maintaining blood pressure and organ perfusion, they also result in end-organ changes in the heart and circulation. *(Modified from A Nohria et al: Atlas of Heart Failure: Cardiac Function and Dysfunction, 4th ed, WS Colucci [ed]. Philadelphia, Current Medicine Group, 2002, p. 104, and J Hartupee, DL Mann: Nat Rev Cardiol 14:30, 2017.)*

remodeling (atrial and ventricular) involves increased wall stress, neurohormonal activation, and inflammation.

Secondary Mitral Regurgitation A large number of patients with HFrEF demonstrate evidence of mitral regurgitation. This occurs due to a distortion in the mitral valve apparatus and includes the effects of various pathophysiologic mechanisms including reduced contractile force, which leads to decreased coaptation of the leaflets, a spherical shape of the ventricle that influences length and function of the chordal-papillary muscle structure, increased dimension of the mitral annulus (and inability of the annulus to contract during systole) with reduced leaflet alignment, and dilation of the posterior wall of the left atrium, which distorts the posterior leaflet of the valve. This worsening in regurgitant volume contributes to progression in HF and adversely influences prognosis. Ensuring that this vicious cycle is interrupted is now a therapeutic target in HF. Some success has been noted by treating the mitral valve using transcatheter techniques when patients are carefully selected after exposure to optimal medical therapy when residual and significant secondary mitral regurgitation persists. Similarly, progressive tricuspid regurgitation can result from and promote adverse right ventricular remodeling. Current interventional studies are underway to assess the impact of transcatheter tricuspid valve repair or replacement in patients with advanced HF.

CARDIORENAL AND ABDOMINAL INTERACTIONS

An important concept underlying the pathophysiology of HF recognizes the systemic nature of disease. Thus, while the primary hemodynamic problem in HF is related to abnormalities in myocardial function (preload, afterload, and contractility), many of the presenting signs and symptoms are related to end-organ failure, including dysfunction of the kidneys, liver, and lungs. The heart and kidney interaction increases circulating volume, worsens symptoms of HF, and results in disease progression, referred to as the cardiorenal syndrome. Traditionally, this relationship was deemed to be a consequence of an impairment in forward flow (cardiac output) leading to a decrease in renal arterial perfusion, worsening renal function, and neurohormonal activation with release of arginine vasopressin, resulting in water and sodium retention. However, evidence has emerged that renal dysfunction may not be adequately explained simply by arterial underfilling and a decline in cardiac output. Systemic venous congestion in HF with increased backward pressure may be operative in determining the development of the cardiorenal syndrome, and relief of venous congestion is associated with significant improvement in renal function in HF. Increased intraabdominal pressure, as noted in right-sided HF, and a rise in abdominal congestion are correlated with renal dysfunction in worsening HF. The interaction is not only confined to the renal component of the abdominal compartment but also involves the liver and spleen. The splanchnic veins serve as a blood reservoir and actively function in regulation of cardiac preload during changes in volume status, regulated by transmural pressure changes or mechanisms of systemic sympathetic activation. The liver and spleen participate in determining volume regulation in HF in addition to several additional interactive pathways. Splanchnic congestion results in portal vein distension and activation of the hepatorenal reflex as well as the splenorenal reflex, which induces renal vasoconstriction. Thus, decongestion in HF by diuretic therapy or mechanical means such as ultrafiltration reduces volume, but also facilitates a decrease in pressure within the abdominal compartment, and this combination of therapeutic effect may serve to improve or stabilize renal function in HF.

GUT CONGESTION, THE MICROBIOME, AND INFLAMMATION

As noted above, circulating levels of proinflammatory cytokines are elevated in a number of cardiovascular disease states, including HF, and have been associated with disease progression. While the primary source of inflammation is unknown, emerging evidence suggests that an alteration in gut microbial composition and loss of microbial diversity may play an important role. The potential role of gut congestion and also altered gut microbial composition may propagate the chronic state of inflammation and immune system dysregulation, eventually leading to progression of HFrEF. Lipopolysaccharide (LPS) is a gram-negative bacterial cell wall product whose levels are increased in patients with HF in the setting of increased intestinal permeability during periods of congestion, which is reduced with diuretic treatment. LPS is a strong stimulator of the immune system and can lead to dysregulated systemic inflammation via macrophage activation. Resulting increases in cytokines such as TNF-α, IL-1, and IL-6 in these pathways can cause progressive loss of cardiac function and also contribute to cardiac cachexia. A mechanistic link has been shown between gut microbe–dependent generation of trimethylamine *N*-oxide derived from specific dietary nutrients such as choline and carnitine and poor outcomes in patients with both acute and chronic HF. Microbe-generated uremic toxins, such as indoxyl sulfate, may play an important role in the development of HF, particularly in interaction with renal insufficiency. Thus, bowel ischemia and/or congestion depending on HF severity may be associated with morphologic and functional alterations in the intestines and result in bacterial endotoxemia and a proinflammatory state.

HIGH-OUTPUT STATES

Although most patients with HF, with either reduced or preserved EF, have low or normal cardiac output (CO) accompanied by elevated systemic vascular resistance (SVR), a minority of patients with HF present with a high-output state with low SVR (Table 264-2). High-output states by themselves are seldom responsible for HF, but their development in the presence of underlying cardiovascular disease can precipitate HF. For example, chronic anemia is associated with high CO when hemoglobin reduces significantly, for example, to a level that is ≤8 g/dL. An increase in vasodilatory metabolites and arteriolar vasodilation in response to decreased oxygen-carrying capacity of the blood in addition to a decrease in blood viscosity contributes to low SVR. Even when severe, anemia rarely causes high-output HF in the absence of a specific cardiac abnormality such as ischemic or valvular heart disease. Patients with end-stage renal disease **(Chap. 323)** are at particular risk of developing high-output HF when chronic anemia is exacerbated by increased flow through an arteriovenous fistula. In a recent analysis of the National Readmission Database, the most common causes of high-output HF included pulmonary disease (19.8%), severe obesity (9.9%), sepsis (9.6%), cirrhosis (8.9%), myelodysplastic syndrome (7.9%), hyperthyroidism (5.5%), and sickle cell disease (3.3%).

EVALUATION

HISTORY

Symptoms of Congestion: Pulmonary Versus Systemic The most common symptoms of HF are related to volume overload with elevation in pulmonary and/or systemic venous pressures. Shortness of breath is a cardinal manifestation of left HF and may arise with increasing severity as exertional dyspnea, orthopnea, paroxysmal nocturnal dyspnea, and dyspnea at rest. Mechanisms of dyspnea include pulmonary venous congestion and transudation of fluid into the interstitium and/or alveoli, leading to decreased lung compliance, increased airway resistance, hypoxemia, and ventilation/perfusion mismatch. Stimulation of juxtacapillary J receptors leading to an increased ventilatory drive and reduced blood flow to respiratory muscles may cause lactic acidosis and a sensation of dyspnea. The New York Heart Association (NYHA) functional classification **(Table 264-4)** may be used to categorize patients based on the amount of effort required to provoke breathlessness. Notably, however, NYHA class does not correlate well with other objective measures of cardiac structure (e.g., left ventricular size, EF) or function (e.g., peak oxygen consumption).

Shortness of breath when bending forward (e.g., to put on socks or tie a shoe) has been associated with an increase in cardiac filling pressure, especially in the presence of a low CO, a symptom referred to as "bendopnea." Orthopnea refers to dyspnea that occurs in the recumbent position and is due to redistribution of fluid from the abdomen and lower body into the chest, increased work of breathing due to decreased lung compliance, and, in patients with ascites or hepatomegaly, an elevation of the diaphragm. Orthopnea typically occurs in the awake patient within 1–2 min of lying down and may be relieved by raising the head and chest with pillows or an adjustable bed. With more severe HF, patients may end up sleeping in a recliner chair or sitting up, although for some, orthopnea may diminish as symptoms of right HF appear. Orthopnea may be accompanied by nocturnal cough related to pulmonary congestion.

TABLE 264-4 New York Heart Association Functional Classification

FUNCTIONAL CLASS	LIMITATION	CLINICAL ASSESSMENT
Class 1	None	Ordinary physical activity does not cause undue fatigue, dyspnea, palpitations, or angina.
Class II	Mild	Comfortable at rest. Ordinary physical activity (e.g., carrying heavy packages) may result in fatigue, dyspnea, palpitations, or angina.
Class III	Moderate	Comfortable at rest. Less than ordinary physical activity (e.g., getting dressed) leads to symptoms.
Class IV	Severe	Symptoms of heart failure or angina are present at rest and worsen with any activity.

Paroxysmal nocturnal dyspnea (PND) refers to episodes of shortness of breath that awaken a patient suddenly from sleep with feelings of anxiety and suffocation and require sitting upright for relief. In contrast to orthopnea, PND usually occurs after prolonged recumbency, is less predictable in occurrence, and may require 30 min or longer in the upright position for relief. Episodes are often accompanied by coughing and wheezing (so-called cardiac asthma) thought to be due to increased bronchial arterial pressure leading to airway compression and interstitial pulmonary edema causing increased airway resistance. Acute pulmonary edema, due to marked elevation of the pulmonary capillary wedge pressure, is manifested by severe shortness of breath and pink, frothy sputum **(Chap. 316)**. Cheyne-Stokes respiration and central sleep apnea may precipitate episodes of PND in HF and are related to increased sensitivity of the respiratory center to arterial PCO_2 and a prolonged circulatory time. Unlike obstructive sleep apnea, which can be treated with positive airway pressure, oral appliance, or surgical therapy, central sleep apnea has no proven therapy beyond the directed treatment of HF **(Chap. 308)**.

In contrast to symptoms of left HF due to pulmonary venous congestion, symptoms of right HF are typically related to systemic venous congestion. Weight gain and lower extremity edema may be the initial manifestations followed by a range of gastrointestinal symptoms due to edema of the bowel wall and hepatic congestion. Abdominal bloating, anorexia, and early satiety are common. Some patients develop right upper quadrant pain related to stretching of the hepatic capsule with consequent nausea and vomiting. When these symptoms are associated with abnormal liver function tests (see below), misdiagnosis of biliary tract disease may occur. For patients with refractory right HF, the development of massive edema involving the entire body with recurrent pleural effusions and/or ascites is termed *anasarca*.

Symptoms of Reduced Perfusion Some patients with advanced HF present with symptoms related to decreased CO, sometimes referred to as *low-output syndrome*. Fatigue and weakness, particularly of the lower extremities, are nonspecific symptoms that can occur with exertion or at rest. Pathophysiology includes reduced blood flow to exercising muscles with endothelial dysfunction and increased SVR from neurohormonal activation. Chronic alterations in skeletal muscle structure and metabolism have also been demonstrated. In older patients with HF and cerebrovascular disease, reduced systemic perfusion may result in mental dullness, depressed affect, and confusion. In addition to low CO, fatigue may be caused by volume depletion, hyponatremia, iron deficiency, and medication effect (e.g., beta blockers).

Other Symptoms Patients with HF may present with mood disturbances and poor sleep, both of which may be exacerbated by nocturnal

TABLE 264-5 Precipitating Factors in Heart Failure

Patient-Related
Excess exertion or emotional stress
Excess fluid and/or sodium intake
Nonadherence with medications
Heavy alcohol use
Provider-Related
Use of medications that cause salt and water retention (e.g., NSAIDs)
Prescribed use of medications with negative inotropic properties (e.g., CCBs)
Unrecognized congestion and inadequate use of diuretics
Heart Failure–Related
Uncontrolled hypertension
Myocardial ischemia or infarction
Atrial or ventricular arrhythmias
Pulmonary embolism
Other Disease States
Systemic infection
Worsening renal or hepatic failure
Hyperthyroidism
Untreated sleep apnea
Anemia or iron deficiency syndrome

Abbreviations: CCB, calcium channel blocker; NSAID, nonsteroidal anti-inflammatory drug.

TABLE 264-6 Diagnostic Criteria for Cachexia in Adults

- Underlying disease and body weight loss $\geq$5% in $\leq$12 months (or BMI <20 kg/m^2)
- Plus at least three of the following five criteria
  - Decrease in muscle strength
  - Fatigue
  - Anorexia
  - Low fat-free mass index
  - Abnormal biochemistry: inflammation, anemia, low serum albumin levels

Abbreviations: BMI, body mass index.

Source: Reproduced with permission from S von Haehling et al: Nat Rev Cardiol 14:323, 2017.

dyspnea and obstructive and/or central sleep apnea. Nocturia due to improved CO and renal perfusion in the supine position, in addition to delayed diuretic effects, can also contribute to sleep disturbances. Oliguria due to severe reductions in renal blood flow may be a sign of advanced-stage HF.

Precipitating Factors Patients with HF may be asymptomatic or mildly symptomatic either because the cardiac impairment is mild or because compensatory mechanisms help to balance or normalize cardiac function. Symptoms of HF may develop when one or more precipitating factors increase cardiac workload and disrupt the balance in favor of decompensation. Specific factors may be identified in 50–90% of admissions and can be divided into patient-related factors, provider-related factors, HF-related disease states, and other causes (Table 264-5). Inability to recognize and correct these factors promptly may lead to persistent HF despite adequate treatment.

PHYSICAL EXAMINATION

General Appearance Most patients with mild-moderate HF will appear well nourished and comfortable at rest. Even patients with more advanced disease may be in no distress after resting for a few minutes but may demonstrate immediate dyspnea with limited exertion such as walking across the room. In contrast, patients with severe HF may need to sit upright and appear anxious, diaphoretic, and dyspneic at rest with pallor due to anemia or duskiness due to low output. Other signs of severe HF include cool extremities and peripheral cyanosis. Cardiac cachexia (Table 264-6), defined partially as unintentional edema-free weight loss of >5% over 12 months, may be observed in patients with longstanding, severe HF as bitemporal or upper body muscle wasting. Contributing factors include poor oral intake due to anorexia, decreased fat absorption due to bowel wall edema, and catabolic/metabolic imbalance from activation of inflammatory cytokines (see above) and dysregulation of the growth hormone–insulin-like growth factor 1 pathway. Rarely, scleral icterus and jaundice may result from severe right HF. Others may demonstrate frailty, which can be diagnosed in the presence of sarcopenia and is exemplified by poor hand-grip strength or severely reduced gait speed.

Vital Signs With new-onset HF, heart rate rises and blood pressure may initially be increased due to sympathetic activation. In patients with chronic HF on guideline-directed medical therapy, resting heart rate ideally should be <70–75 beats/min, and blood pressure should be in the normal to low-normal range. An irregular rhythm may be due to atrial fibrillation or flutter or frequent premature atrial or ventricular complexes. Severe HF may be associated with hypotension and narrow pulse pressure along with a rapid, thready pulse. An alternating strong and weak pulse, known as *pulsus alternans*, is attributed to reduced left ventricular contraction in every other cardiac cycle due to incomplete recovery causing reduction in the left ventricular stroke volume with each alternate beat. Respiratory rate may be normal at rest but may increase on lying down or on minimal exertion. Advanced HF may be associated with periodic breathing or Cheyne-Stokes respirations. The patient is usually unaware of the altered breathing pattern, but family members or friends may become alarmed or attribute this incorrectly to anxiety. Oxygen saturation is typically normal on room air unless there is acute pulmonary edema, underlying CHD with shunting, severe pulmonary arterial hypertension, or concomitant acute or chronic lung disease. A low-grade fever resulting from cytokine activation may occur in severe HF and subside when compensation is restored.

Jugular Venous Pulse Examination of the jugular veins provides an estimate of the right atrial pressure. Typically, the patient is examined at a 45° angle, and jugular venous pressure (JVP) is quantified in centimeters of water by estimating the height of the venous column of blood above the sternal angle in centimeters and then adding 5. In patients with mild right HF, JVP may be normal at rest (≤8 cm H_2O) but increase with compression of the right upper quadrant. *Hepatojugular reflux* is elicited by applying firm continuous pressure over the liver for 15–30 s while observing the neck veins. The patient must breathe normally and not strain during the maneuver. Higher levels of venous pressure approaching the angle of the jaw are common in chronic right HF. If significant tricuspid regurgitation is present, prominent V waves and Y descents may be noted. The *abdominojugular test*, defined as an increase in right atrial pressure during 10 s of firm midabdominal compression followed by an abrupt drop on pressure release, suggests elevated left-sided filling pressure. Elevation in JVP during inspiration or Kussmaul's sign may be due to severe biventricular HF and is a marker of poor outcome; it can also be seen with constrictive pericarditis or restrictive cardiomyopathy.

Lung Examination Pulmonary rales result from transudation of fluid from the intravascular space into the alveoli and airways. In general, rales are heard at the lung bases, but in severe HF or acute pulmonary edema, they may be heard throughout the lung fields. Wheezing and rhonchi can occur with congestion of the bronchial mucosa and sometimes lead to a misdiagnosis (and inappropriate treatment) of asthma or chronic obstructive pulmonary disease (COPD). Rales may be absent in patients with longstanding HF and chronically elevated pulmonary capillary wedge pressures due to increased lymphatic drainage, which prevents spillage from the interstitium into the alveoli. In biventricular or predominant right HF, bilateral pleural effusions are recognized as dullness to percussion and decreased breath sounds at the bases. When pleural effusions are unilateral, they typically involve the right side.

Cardiac Examination As discussed above, chronic HF with ventricular remodeling is typically accompanied by cardiac enlargement. The apical impulse is displaced downward and to the left and may be diffuse in dilated cardiomyopathy or sustained in pressure overloaded states such as aortic stenosis. In biventricular or severe right HF, a right ventricular heave or parasternal lift may be palpated along the left sternal border. Uncommonly, a palpable third heart sound may be present. In patients with HFpEF, precordial palpation is often normal. On auscultation, an S_3 gallop is most commonly present in patients with volume overload and tachycardia, suggests severe hemodynamic compromise, and carries negative prognostic significance. An S_4 gallop is not specific to HF but may be present in patients with HFpEF due to hypertension. Holosystolic murmurs of mitral and tricuspid regurgitation are present in the setting of advanced HF, often in the absence of structural valvular abnormalities. In patients with secondary pulmonary hypertension, a loud pulmonary component of the second heart sound may be heard.

Abdomen and Extremities Hepatomegaly is an early sign of systemic venous congestion. The liver edge may be tender due to stretching of the capsule, but with progression of right HF, tenderness may disappear. The liver edge may be pulsatile in patients with tricuspid regurgitation. Longstanding hepatic congestion may result in cardiac cirrhosis with congestive splenomegaly and mild-moderate ascites. The presence of massive ascites should lead to a search for other causes such as constrictive pericarditis or primary liver failure. Dependent lower extremity edema is common in chronic HF and is typically symmetric and pitting. Over time, chronic edema may cause reddening and induration of the skin, become weeping, or lead to cellulitis. Anasarca is used to describe massive, generalized edema involving the legs, sacrum, and abdominal wall. In patients with acute HF or younger adults with chronic HF, lower extremity edema may be absent despite marked systemic venous hypertension. Unilateral lower extremity edema may be due to deep venous thrombosis, prior trauma, or history of vein harvest for bypass surgery. Nonpitting edema that does not respond to increasing doses of diuretics may represent lymphedema that requires alternative diagnostic workup and treatment.

DIAGNOSIS

The diagnosis of HF is relatively straightforward when the patient presents with typical signs and symptoms; however, the signs and symptoms of HF are neither specific nor sensitive. It is therefore important for clinicians to have a high index of suspicion for HF, particularly in patients who are at increased risk, including older patients with underlying cardiovascular disease and those with comorbidities such hypertension, diabetes, and chronic kidney disease. In this setting, additional laboratory testing and imaging should be performed **(Fig. 264-8)**.

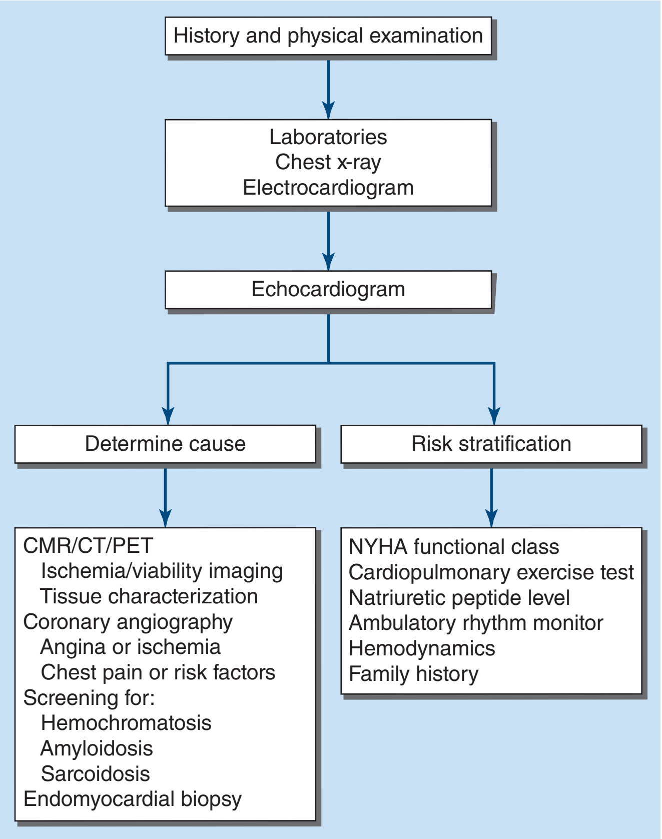

FIGURE 264-8 Initial assessment of patients presenting with heart failure. The initial evaluation starts with a thorough history and physical examination, focusing on detection of comorbidities including hypertension, diabetes, and hyperlipidemia. In addition, identification of valvular heart disease, vascular disease, history of mediastinal radiation, or exposure to cardiotoxins (e.g., chemotherapy, alcohol, or illicit drugs) may help determine underlying cause. A family history of sudden death, heart failure, arrhythmias, or cardiomyopathy is also useful. Routine laboratory evaluation (see text) should also be performed. Chest x-ray is useful to detect cardiomegaly and fluid overload and to rule out pulmonary disease. A 12-lead electrocardiogram should be performed to detect abnormalities of cardiac rhythm and conduction, left ventricular hypertrophy, and evidence of myocardial ischemia or infarction. Two-dimensional echocardiography with Doppler imaging is indicated to assess cardiovascular structure and function and detect abnormalities of the myocardium, heart valves, or pericardium. Further imaging and laboratory studies aimed at identifying a specific cause of cardiomyopathy depend on information obtained from the history and physical examination. In all patients, risk stratification should be performed to assess severity of illness, guide therapy, and provide prognosis to patient and family. CMR, cardiac magnetic resonance imaging; CT, computed tomography; NYHA, New York Heart Association; PET, positron emission tomography.

Routine Laboratories Standard laboratory testing in patients with HF includes a comprehensive metabolic panel, complete blood count, coagulation studies, and urinalysis. Selected patients should have assessment for diabetes, hyperlipidemia, and thyroid function. Blood urea nitrogen and creatinine levels are often elevated in moderate-severe HF due to reduced renal blood flow and/or increased renal venous pressure. Worsening renal function **(Chaps. 321 and 322)** due to diuretics, RAAS inhibitors, and noncardiac medications (e.g., nonsteroidal anti-inflammatory drugs) is also common. Proteinuria may be present in the setting of longstanding hypertension or diabetes or suggest an underlying systemic disease. Chronic right HF with congestive hepatomegaly can lead to modest elevations in transaminases, alkaline phosphatase, and bilirubin that should not be confused with biliary tract disease. Marked elevation in transaminases and lactic acid suggests cardiogenic shock with severe low output. In patients with cardiac cirrhosis, hypoalbuminemia may exacerbate fluid accumulation, whereas hyperammonemia contributes to altered mental status. In general, inflammatory markers such as erythrocyte sedimentation rate, C-reactive protein, and uric acid are nonspecific and do not aid in the diagnosis of HF. Other laboratories, including antinuclear antibodies, rheumatoid factor, serum free light chains, serum protein electrophoresis, ferritin, ceruloplasmin, hepatitis C, and HIV, are reserved for targeted testing.

Electrolyte abnormalities seen in HF include hyponatremia due to sodium restriction, diuretic therapy, and vasopressin-mediated free water retention. Hyponatremia is a negative prognostic indicator at the time of HF hospitalization and predicts decreased long-term survival (Table 264-1). Hypokalemia is most often due to thiazide or loop diuretics given without oral potassium supplementation but may also result from increased aldosterone levels. Hyperkalemia may result from marked reductions in glomerular filtration rate and is exacerbated by use of RAAS inhibitors, potassium-sparing diuretics, and potassium supplements **(Chap. 265)**. Hypo- or hyperkalemia may lead to atrial or ventricular arrhythmias. Hypophosphatemia and hypomagnesemia are commonly associated with chronic alcohol use.

Anemia is not diagnostic of HF, but when present, it may exacerbate underlying ischemic heart disease or decrease quality of life in patients with HF due to any cause and should be corrected. Rarely, severe anemia may cause high-output HF typically in the presence of underlying cardiovascular disease. The presence of iron deficiency (with or without anemia) is increasingly recognized in patients with chronic HF and has been attributed to decreased gut absorption, impaired hepatic storage, and chronic blood loss. Repletion with IV iron (but not oral iron) results in improved symptoms and exercise capacity and reduced HF hospitalizations, but its effect on survival remains uncertain.

Chest X-Ray Major abnormalities on chest imaging associated with left HF include enlarged cardiac silhouette (cardiothoracic ratio >0.5) and pulmonary venous congestion. Early radiologic signs of acute HF include upper zone venous redistribution and thickening of interlobular septa. When the pulmonary capillary wedge pressure is moderate to severely elevated, alveolar edema can present as diffuse haziness extending downward toward the lower lung fields. The absence of these findings in patients with chronic HF reflects the increased capacity of the lymphatics to remove interstitial and/or pulmonary fluid. Pleural effusions of varying size and distribution are common in biventricular HF. Chest x-ray can also be used to identify noncardiac causes of dyspnea (e.g., pneumonia, COPD).

Electrocardiogram No specific electrocardiographic (ECG) pattern is diagnostic of HF. Rather, the ECG may provide important information regarding presence of underlying cardiac disease. For example, left ventricular hypertrophy and left atrial enlargement suggest HFpEF due to hypertension, aortic stenosis, or hypertrophic cardiomyopathy. The presence of Q waves or infarction is suggestive of ischemic heart disease, whereas Q waves with reduced QRS voltage (pseudo-infarct pattern) may be seen with restrictive or infiltrative cardiomyopathies (e.g., amyloid). Conduction system disease should raise concern for cardiac sarcoid or Chagas cardiomyopathy in the right clinical setting.

Paroxysmal or persistent atrial fibrillation is present in up to 40% of patients with chronic HF and is an indication for anticoagulation. Premature ventricular complexes (PVCs) and nonsustained ventricular tachycardia can reflect worsening HF and are markers of increased risk. Conversely, frequent PVCs can cause cardiomyopathy that may be treated successfully with ablation **(Chap. 260)**. Finally, determination of the QRS width and presence of LBBB is used to ascertain whether the patient may benefit from cardiac resynchronization or left bundle branch (or His bundle) pacing therapy.

Noninvasive Imaging Noninvasive cardiac imaging **(Chap. 248)** is essential for the diagnosis, evaluation, and management of HF. Two-dimensional echocardiography provides an accurate and rapid determination of ventricular size and function and valvular morphology and function and can detect intracavitary thrombi and pericardial effusions. When left ventricular ejection fraction (LVEF) is ≥50%, systolic function is deemed to be normal. Myocardial strain rate imaging using speckle tracking can add incremental value to LVEF and carries prognostic value. Doppler techniques can be used to estimate CO, pulmonary artery pressures, and valve areas, and may detect abnormalities in left ventricular diastolic filling in patients with HFpEF. For patients with end-stage HF, echocardiography is critical for assessment of right ventricular function before and after mechanical circulatory support and heart transplant. Transesophageal echocardiogram (TEE) or cardiac computed tomography (CT) with contrast is indicated to rule out left atrial appendage thrombus prior to cardioversion and can assess aortic or mitral valve pathology in planning for transcatheter valvular replacement or repair. TEE can also be used to assess for endocarditis in patients with bacteremia or acute valvular regurgitation.

Cardiac magnetic resonance imaging (CMR) has emerged as a highly accurate and quantitative tool for evaluation of left ventricular mass, volumes, and function and for determining specific causes of HF (e.g., ischemic cardiomyopathy, myocarditis, amyloidosis, hemochromatosis). CMR is particularly helpful in defining multiple anatomic and functional abnormalities in adults with CHD. Serial CMR studies can assess ventricular remodeling in response to therapy and are useful in clinical research. For patients who cannot undergo CMR (e.g., due to implantable devices), cardiac CT is particularly helpful to rule out pericardial disease or left ventricular apical thrombus. Coronary CT angiography has also emerged as a useful noninvasive test to rule out obstructive coronary artery disease as a cause of HF. While limited by availability and cost, cardiac positron emission tomography (PET) plays a role in evaluating the extent of ischemia, infarction, or hibernating myocardium in patients with coronary artery disease and, in the case of sarcoidosis, can reliably determine the severity and distribution of cardiac inflammation.

Cardiopulmonary Exercise Testing While not routinely performed in HF, cardiopulmonary exercise testing using a symptom-limited, ramp protocol can provide an objective assessment of peak functional capacity in patients being evaluated for mechanical circulatory support or heart transplant **(Chap. 271)**. Several parameters including absolute and percent-predicted peak oxygen consumption (VO_2) and ventilatory efficiency (VE; assessed by the VE/VCO_2 slope) are independent predictors of survival. Additional data including heart rate and blood pressure response to exercise and exercise-induced arrhythmias can also be assessed. This test may also be useful in defining the cause of dyspnea when the diagnosis is uncertain.

Biomarkers Circulating levels of natriuretic peptides are useful, adjunctive tools in the diagnosis of HF. BNP and N-terminal pro-BNP (NT-proBNP) are released from the atria and ventricles in response to increased wall stress. Patients with HFrEF tend to have higher levels than patients with HFpEF, whereas levels may be falsely low in obesity. In ambulatory patients with dyspnea, the measurement of BNP or NT-proBNP is useful to support clinical decision-making regarding the diagnosis of HF, especially in the setting of clinical uncertainty or with concomitant lung disease. Moreover, natriuretic peptide levels can be used to establish disease severity and prognosis in chronic HF and may help to guide optimal dosing of medical therapy in stable outpatients. Importantly, many noncardiac factors, including age, female sex, and chronic kidney disease, increase natriuretic peptide levels. Other cardiovascular diseases including atrial fibrillation, pulmonary embolism, and pulmonary arterial hypertension can also increase BNP levels. Galectin-3 and soluble ST2 (suppression of tumorigenicity 2 protein) are newer biomarkers that have been approved for assessment of prognosis in HF but are not widely used. Biomarkers of renal injury require further study in HF, although use of cystatin C to measure renal function may be more accurate than a creatinine, which can be influenced by muscle mass.

Invasive Studies In the intensive care setting, assessment of cardiac filling pressures and CO may be necessary to differentiate cardiogenic from noncardiogenic pulmonary edema and manage hemodynamic instability. Placement of a pulmonary artery catheter can be performed safely at the bedside and used to determine response to intravenous vasoactive and diuretic therapy in severe HF. Simultaneous measurement of right and left heart filling pressures in the cardiac catheterization laboratory can be used to distinguish restrictive cardiomyopathy from constrictive pericarditis. Coronary angiography is indicated to exclude ischemic heart disease as an underlying, potentially reversible cause of left ventricular dysfunction. The management of coronary artery disease in the setting of chronic HF is discussed in **Chaps. 285–287**. If echocardiographic windows are suboptimal, left ventriculography can provide an assessment of left ventricular size and function and severity of mitral regurgitation. The role of right ventricular endomyocardial biopsy in the management of HF and cardiomyopathy remains controversial. Indications include detection of myocarditis (lymphocytic, eosinophilic, sarcoid, or giant cell), diagnosis of cardiac amyloidosis and chemotherapy- or immunotherapy-related left ventricular failure, and screening for cardiac allograft rejection following heart transplant.

COMORBIDITIES

■ DIABETES

Type 2 diabetes mellitus is a risk factor for the development of HF **(Table 264-7)** and increases the risk of morbidity and mortality in

TABLE 264-7 Mechanisms That Contribute to Development of Heart Failure in Patients with Type 2 Diabetes Mellitus

Altered myocardial substrate
Abnormal mitochondrial bioenergetics
Oxidative stress and inflammation
Lipotoxicity
Endoplasmic reticulum stress
Impaired insulin signaling
β_2-Adrenergic receptor signaling
G protein–coupled receptor kinase 2 signaling
RAAS activation
Advanced glycation end products
Autophagy

Abbreviation: RAAS, renin-angiotensin-aldosterone system.

Source: Modified with permission from TA Zelniker: Mechanisms of cardiorenal effects of sodium-glucose cotransporter 2 inhibitors: JACC state-of-the-art review. J Am Coll Cardiol 75:422, 2020.

TABLE 264-8 Differential Diagnosis of Heart Failure

SYMPTOM OR SIGN	DIFFERENTIAL DIAGNOSIS
Dyspnea	Chronic lung disease Pulmonary arterial hypertension Neuromuscular disease Anemia Iron-deficiency anemia
Edema	Venous insufficiency Nephrotic syndrome Deep vein thrombosis Lymphedema
Ascites	Hepatic cirrhosis Portal vein thrombosis Malignant carcinomatosis
Pleural effusion(s)	Chronic infection Lung cancer Collagen vascular or rheumatologic disease
Jugular venous distension	Constrictive pericarditis Pericardial effusion Superior vena cava syndrome

patients with established disease. In ambulatory HF cohorts, the prevalence of diabetes ranges from 10 to 40%, with prevalence even higher in patients hospitalized with HF. When the two diseases coexist, patients are at increased risk for adverse outcomes, worse quality of life, and higher costs of care. Recent data from cardiovascular outcomes trials demonstrate that HF is a critical outcome in patients with diabetes and that glucose-lowering therapies can impact morbidity and mortality. As discussed above, SGLT-2 inhibitors in particular have not only been shown to be safe in patients with HF but can also improve renal function, enhance quality of life, increase LVEF, and decrease the risk of hospitalization and death. Use of other guideline-directed medical therapy is indicated in patients with HF regardless of diabetes status.

■ SLEEP APNEA

Sleep-disordered breathing is common in HF, with increased incidence of both obstructive sleep apnea and central sleep apnea **(Chap. 308)**. The pathophysiologic link between these disorders has been studied in both animal models and humans and includes increased afterload, decreased preload, intermittent hypoxia, and sympathetic activation. Increase in sympathetic tone can provoke ischemia and arrhythmias and complicate blood pressure management. Approximately one-third of patients with HF and sleep-disordered breathing have central sleep apnea, which is associated with increased mortality independent of other known risk factors. In patients with HFrEF and obstructive sleep apnea, continuous positive airway pressure has been shown to improve quality of life, decrease blood pressure and arrhythmias, and increase LVEF. Unlike obstructive sleep apnea, there is no proven therapy for central sleep apnea, although phrenic or hypoglossal nerve pacing may provide benefit in some cases.

■ OBESITY

Similar to diabetes, obesity is both a risk factor for the development of HF and highly prevalent in patients with HF. In particular, obesity is common in patients with HFpEF and complicates the assessment of volume status in both ambulatory and inpatient settings. Unlike diabetes, the risk of morbidity and mortality in patients with obesity and HF is complex. The obesity paradox refers to the observation that obese patients diagnosed with HF have a more favorable prognosis than patients with low or even normal body mass index. While weight loss has been shown to improve quality of life and exercise capacity and may contribute to reverse ventricular remodeling in patients with HF, the impact on survival is unknown. Large randomized clinical trials are assessing safety and efficacy of glucagon-like peptide-1 (GLP-1) agonists such as semaglutide in patients with heart failure and obesity.

■ DEPRESSION

Depression is an independent risk factor for adverse outcomes in HF (Table 264-1), especially in older women. The mechanisms underlying this risk remain unknown but may involve neuroendocrine dysfunction and systemic inflammation, as well as contributions from poor sleep, decreased appetite, and adverse effects of medications and alcohol. The AHA recommends screening for depression among patients with cardiovascular disease including HF using validated patient health questionnaires. Selective serotonin reuptake inhibitors are safe for treating depression in HF but do not appear to affect the natural history of disease. The effects of cognitive behavioral therapy and the collaborative care model, as well as newer therapies such as transcranial magnetic stimulation, on HF morbidity and mortality require further study.

DIFFERENTIAL DIAGNOSIS

Many symptoms and signs suggesting HF may be caused by other conditions **(Table 264-8)**. In a patient with dyspnea, the clinician must distinguish cardiac from pulmonary causes, although the differentiation may be difficult. For example, orthopnea may be a well-established symptom in some patients with severe chronic lung disease. Patients with underlying pulmonary disease may also experience episodic shortness of breath during sleep that mimics PND. In chronic lung disease, this is usually due to accumulation of tracheobronchial secretions and is relieved by coughing and expectoration, whereas in cardiac disease, the patient has to sit upright. Wheezing caused by bronchoconstriction may be a prominent symptom when left ventricular failure supervenes in individuals with reactive airways disease. Patients with cardiac asthma may be more likely to exhibit diaphoresis and varying degrees of cyanosis compared to patients with bronchial asthma. Differentiating dyspnea related to HF versus pulmonary disease may be impossible when the diseases coexist, a situation that is common in chronically ill older patients with active or prior smoking. Following effective diuresis, pulmonary function tests may help to determine the predominant cause of dyspnea. In ambulatory patients with advanced HF, cardiopulmonary exercise testing can also help to make this distinction. Finally, a very low BNP or NT-proBNP level may be helpful in excluding HF as the cause of dyspnea in nonobese patients.

Apart from pulmonary disease, HF needs to be distinguished from conditions in which congestion results from abnormal salt and water retention but in which cardiac structure and function are normal (e.g., acute or chronic kidney disease) and from noncardiac causes of pulmonary edema (e.g., acute respiratory distress syndrome). Non-HF causes of lower extremity edema such as venous insufficiency, lymphedema, and obesity should also be considered.

FURTHER READING

Aimo A et al: Cardiac remodeling - Part 2: Clinical, imaging and laboratory findings. A review from the Study Group on Biomarkers of the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail 24:944, 2022.

Alcaide P et al: Myocardial inflammation in heart failure with reduced and preserved ejection fraction. Circ Res 134:1752, 2024.

Boorsma EM et al: Congestion in heart failure: A contemporary look at physiology, diagnosis and treatment. Nat Rev Cardiol 17:641, 2020.

Bozkurt B et al: Heart failure epidemiology and outcomes statistics: A report of the Heart Failure Society of America. J Card Fail 29:1412, 2023.

Bozkurt B et al: Universal definition and classification of heart failure: A report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure. J Card Fail 27:387, 2021.

Campbell P et al: Heart failure with preserved ejection fraction: Everything the clinician needs to know. Lancet 403:1083, 2024.

Dunlay SM et al: Type 2 diabetes mellitus and heart failure: A scientific statement from the American Heart Association and the Heart Failure Society of America. Circulation 140:e294, 2019.

Heidenreich PA et al: 2022 AHA/ACC/HFSA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 145:e895, 2022.

Houston BA et al: Right ventricular failure. N Engl J Med 388:1111, 2023.

Lam CSP et al: Classification of heart failure according to ejection fraction. J Am Coll Cardiol 77:3217, 2021.

Lembo M et al: Obesity: The perfect storm for heart failure. ESC Heart Fail 11:1841, 2024.

McDonagh TA et al: 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 42:3599, 2021.

Pandey A et al: Exercise intolerance in older adults with heart failure with preserved ejection fraction: JACC State-of-the-Art Review. J Am Coll Cardiol 78:1166, 2021.

Talha KM et al: Frailty and heart failure: State-of-the-art review. J Cachexia Sarcopenia Muscle 14:1959, 2023.

Xanthopoulos A et al: Heart failure and liver disease: Cardiohepatic interactions. JACC Heart Fail 7:87, 2019.

265

Heart Failure: Management

Akshay Desai, Mandeep R. Mehra

Clinical management of patients with heart failure (HF) varies widely based on the clinical phenotype at presentation. Those in the earliest stage of disease with asymptomatic ventricular dysfunction (American College of Cardiology [ACC]/American Heart Association [AHA] stage B) may be amenable to treatment with neurohormonal antagonists, including angiotensin-converting inhibitors and β-adrenergic receptor antagonists, with the goal of facilitating ventricular recovery and preventing the development of clinical HF (not further discussed). Those with symptomatic HF (ACC/AHA stage C) comprise a heterogeneous group in whom the approach to therapy is differentiated largely based on measurement of the left ventricular ejection fraction (LVEF). Data from prospective, randomized clinical outcomes trials enrolling patients with symptomatic chronic HF and reduced ejection fraction (HFrEF; LVEF ≤40%) have provided a rich evidence base that supports the efficacy of stepped pharmacologic therapy with renin-angiotensin-aldosterone system (RAAS) antagonists (including angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, and mineralocorticoid receptor antagonists), neprilysin inhibitors, β-adrenergic receptor antagonists, and sodium-glucose cotransporter 2 (SGLT-2) inhibitors as a complement to device-based treatment with cardiac resynchronization therapy and implantable cardioverter-defibrillators. By contrast, treatment of patients with symptomatic chronic HF and mildly reduced (HFmrEF; LVEF 41–49%) or preserved ejection fraction (HFpEF; LVEF ≥50%) has, until recently, been symptom-focused owing to the lack of evidence-based therapies but is evolving in the wake of favorable results from recently reported trials of SGLT-2 inhibitors, glucagon-like peptide 1 (GLP-1) agonists, and angiotensin-neprilysin inhibitors. Even with effective therapy, patients with HFrEF, HFmrEF, and HFpEF are at risk for clinical deterioration, typically because of progressive sodium and fluid retention that fuels the development of congestive symptoms and acute decompensated HF (ADHF). Management of these exacerbations (frequently hospital-based) is heavily focused on hemodynamic stabilization, decongestion, and institution of appropriate disease-modifying therapy in the transition back to chronic ambulatory management. Recurrent episodes of ADHF despite careful follow-up and effective treatment may signal the onset of an advanced or refractory HF phenotype (ACC/AHA stage D) in which the risk of mortality from sudden death or end-stage HF is high, and consideration of salvage therapies including cardiac transplant or mechanical circulatory support may be appropriate prior to escalation to palliative measures **(Chap. 271)**.

HEART FAILURE WITH MILDY REDUCED OR PRESERVED EJECTION FRACTION

GENERAL PRINCIPLES

Although clinical trials of RAAS antagonists, digoxin, β-adrenergic receptor blockers, and neprilysin inhibitors have been conducted in patients with HFpEF, none has conclusively demonstrated a mortality reduction. Variable benefits, measured in the form of reduced HF hospitalizations, have been seen with mineralocorticoid receptor antagonists, angiotensin receptor blockers, and angiotensin receptor-neprilysin inhibitors, but benefits are principally confined to those with LVEF below the normal range (<0.60). This has led many to recommend that patients with HFmrEF should be treated with the same therapies as those with HFrEF. Patients with higher ejection fraction (EF) remain a therapeutic challenge. In the absence of specific pharmacologic therapies proven to improve clinical outcomes, management of patients with HFpEF has historically been focused on improving symptoms and effort tolerance through lifestyle modification, control of congestion, stabilization of heart rhythm (particularly in those with atrial fibrillation), control of blood pressure to guideline-recommended targets, and management of comorbidities that may contribute to disease progression (including, for example, obesity, obstructive lung disease, obstructive sleep apnea, diabetes/insulin resistance, anemia, iron deficiency, and chronic kidney disease). More recently, however, targeted pharmacologic therapy is emerging, with clinical trials of SGLT-2 inhibitors supporting use of these agents to reduce cardiovascular mortality and HF hospitalizations in HFmrEF and HFpEF. Trials of angiotensin receptor-neprilysin inhibitors and GLP-1 agonists (in obese patients) also suggest possible benefits in selected HFpEF patients.

CLINICAL TRIALS IN HFPEF

Attempts to export the benefits of drugs that improve clinical outcomes in patients with HFrEF, including angiotensin-converting enzyme (ACE) inhibitors, angiotensin receptor blockers (ARBs), β-adrenergic receptor blockers, digoxin, and mineralocorticoid receptor antagonists, to those with HFpEF have generally been unsuccessful. The Candesartan in Heart Failure—Assessment of Mortality and Morbidity (CHARM) Preserved study showed a statistically significant reduction in HF hospitalizations but no difference in all-cause mortality in patients with HFpEF who were treated with the ARB candesartan.

Similarly, the Irbesartan in Heart Failure with Preserved Systolic Function (I-PRESERVE) trial demonstrated no differences in the composite of cardiovascular death or HF hospitalization during treatment with the ARB irbesartan compared with placebo. Apparent early benefits of the ACE inhibitor perindopril on HF hospitalizations and functional capacity in the Perindopril in Elderly People with Chronic Heart Failure (PEP-CHF) study were attenuated over longer-duration follow-up. The Digitalis Investigation Group (DIG) Ancillary Trial found no impact of digoxin on all-cause mortality or on all-cause or cardiovascular hospitalization among patients with chronic HF, EF >45%, and sinus rhythm, although a modest reduction in HF hospitalizations was noted. While no dedicated study of beta blockers has been conducted in HFpEF, the subgroup of elderly patients with prior hospitalization and HFpEF enrolled in the Study of the Effects of Nebivolol Intervention on Outcomes and Rehospitalization in Seniors with Heart Failure (SENIORS) trial of nebivolol, a vasodilating beta blocker, did not appear to experience significant reductions in all-cause or cardiovascular mortality.

Regarding *mineralocorticoid receptor antagonists*, which have potent antifibrotic effects in HFrEF, the Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (TOPCAT) trial explored the potential benefit of spironolactone compared to placebo in HFpEF. This trial demonstrated no improvement in the primary composite endpoint of cardiovascular death, HF hospitalizations, or aborted cardiac arrest but did show a reduction in HF hospitalizations among those allocated to spironolactone. Post hoc analyses of the study suggested significant regional differences in the baseline characteristics, event rates, adverse effects, and adherence to spironolactone among patients randomized in Russia and the Republic of Georgia compared with those randomized in the Americas that raised concerns about study conduct in Russian and Georgian sites. Apparent reductions in cardiovascular death and HF hospitalization associated with spironolactone among the subgroup of patients randomized in the Americas suggest that these study design issues may have obscured a signal of spironolactone benefit. These data have supported a weak recommendation for spironolactone in patients with HFpEF who meet the inclusion criteria for the TOPCAT trial and are at low risk for adverse effects, including hyperkalemia and worsening renal function, in the most recent U.S. and European guidelines. However, the results of the Aldosterone Receptor Blockade in Diastolic Heart Failure (ALDO-DHF) study in which spironolactone improved echocardiographic indices of diastolic dysfunction but failed to improve exercise capacity, symptoms, or quality-of-life (QOL) measures highlight the need for further study. Ongoing trials, including the registry-based Spironolactone Initiation Registry Randomized Interventional Trial in Heart Failure with Preserved Ejection Fraction (SPIRRIT-HFpEF) (SPIRRIT-HFpEF; clinicaltrials.gov identifier NCT02901184) and the randomized Study to Evaluate the Efficacy and Safety of Finerenone on Morbidity and Mortality in Participants with Heart Failure and Left Ventricular Ejection Fraction Greater than or Equal to 40% (FINEARTS-HF, *clinicaltrials.gov* identifier: NCT04435626) may provide additional insight in this regard.

Addition of the SGLT-2 inhibitors dapagliflozin and empagliflozin to guideline-directed medical therapy of HFrEF has been associated with reductions in cardiovascular mortality and HF hospitalization among patients with and without diabetes mellitus enrolled in the Dapagliflozin and Prevention of Adverse Outcomes in Heart Failure (DAPA-HF) and Empagliflozin Outcome Trial in Patients with Chronic Heart Failure with Reduced Ejection Fraction (EMPEROR-REDUCED) trials. These benefits have recently been extended to the population of patients with symptomatic HF and LVEF >40% based on observed reductions in the composite of cardiovascular death or HF hospitalization among those assigned to SGLT-2 inhibitors compared to placebo in the Dapagliflozin Evaluation to Improve the Lives of Patients with Preserved Ejection Fraction Heart Failure (DELIVER) and Empagliflozin Outcome Trial in Patients with Chronic Heart Failure with Preserved Ejection Fraction (EMPEROR-PRESERVED) trials. Pooled analysis of the data from these trials is associated with a reduction in the composite or cardiovascular death or HF hospitalization. These results have driven updates to treatment guidelines in the United States and Europe, which now embrace use of SGLT-2 inhibitors as foundational therapy in patients with symptomatic HF regardless of EF. However, cost-effectiveness analyses suggest that these agents may be of low to intermediate economic value, which may limit access until price reductions ensue.

■ OTHER THERAPEUTIC TARGETS

Beyond pharmacologic therapy, small studies of exercise training and cardiac rehabilitation in patients with HFpEF have suggested benefits on functional capacity and QOL, indicating a possible role for lifestyle interventions to improve cardiorespiratory fitness in this population. For obese patients with HFpEF, aggressive efforts at weight loss (including bariatric surgery) may also be associated with improvements in hemodynamics and exercise capacity. Recently, the randomized STEP-HFpEF trial of the once-weekly GLP-1 agonist semaglutide in obese (body mass index ≥30 kg/m^2) HFpEF patients without diabetes mellitus demonstrated improvements in functional capacity and QOL associated with incremental weight loss of nearly 12% relative to placebo. Several ongoing trials including STEP-HFpEF DM (Semaglutide Treatment Effect in People with Obesity and HFpEF and Type 2 Diabetes; clinicaltrials.gov identifier: NCT04916470) and the SUMMIT trial of the dual GLP-1/glucose-dependent insulinotropic polypeptide (GIP) agonist tirzepatide (*clinicaltrials.gov* identifier: NCT04847557) will provide further data for clinical utility in this population.

A novel paradigm for understanding the pathophysiology of HFpEF has focused on the role of microvascular endothelial inflammation driven by comorbidities that results in impaired nitric oxide (NO) signaling and associated increases in myocardial stiffening. This paradigm has emphasized the potential for improving outcomes in HFpEF by enhancing NO bioavailability and improving downstream protein kinase G–based signaling. In this regard, a small trial demonstrated that the phosphodiesterase-5 inhibitor *sildenafil* improved filling pressures and right ventricular function in a cohort of HFpEF patients with pulmonary venous hypertension. This finding led to the phase 2 trial, Phosphodiesterase-5 Inhibition to Improve Clinical Status and Exercise Capacity in Diastolic Heart Failure (RELAX), in HFpEF patients (LVEF >50%) with New York Heart Association (NYHA) functional class II or III symptoms, who received sildenafil at 20 mg three times daily for 3 months, followed by 60 mg three times daily for another 3 months, compared with a placebo. There was no improvement in functional capacity, QOL, or other clinical and surrogate parameters in those allocated to sildenafil compared to placebo. On the premise that *nitrates*, which are NO donors, might improve preload, coronary perfusion, endothelial function, and exercise tolerance, the Nitrate's Effect on Activity Tolerance in Heart Failure with Preserved Ejection Fraction (NEAT-HFpEF) study was conducted. Isosorbide mononitrate did not improve QOL or submaximal exercise capacity and decreased overall activity levels in treated patients. *Inorganic nitrate* compounds have also been shown to enhance NO signaling but did not improve functional capacity compared to placebo among patients with HFpEF randomized in the Inorganic Nitrite Delivery to Improve Exercise Capacity in Heart Failure with Preserved Ejection Fraction (INDIE-HFpEF) trial.

Neprilysin inhibition is known to increase circulating levels of various vasoactive peptides, including the natriuretic peptides, which may facilitate cyclic guanosine 3′,5′-monophosphate–based signaling, enhance myocardial relaxation, and reduce ventricular hypertrophy. Composite *angiotensin receptor-neprilysin inhibition* (ARNI) with sacubitril-valsartan reduced cardiovascular mortality, overall mortality, and HF hospitalization compared with enalapril among patients with HFrEF randomized in the PARADIGM-HF trial. The PARAGON-HF trial randomized 4822 patients with symptomatic HFpEF (LVEF ≥45%), elevated natriuretic peptides, and structural heart disease to treatment with either sacubitril-valsartan or valsartan with the novel composite primary endpoint of cardiovascular death and total hospitalizations for HF. Although there was a 13% reduction in the rate of the primary composite endpoint in those allocated to sacubitril-valsartan, this result narrowly missed the margin for statistical significance in the

Heart Failure with Preserved Ejection Fraction: Pathology and Management

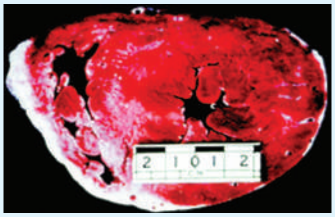

Pathology

Hypertrophy

Fibrosis/altered collagen

Infarction/ischemia

↔

Risk markers

Hypertension

Aging

Atherosclerosis

Diabetes

Obesity

General Therapeutic Principles

- **Reduce the congestive state**
  - Caution to not reduce preload excessively
  - Use of implantable hemodynamics monitors to guide management is useful
- **Control blood pressure**
  - Central aortic blood pressure control may be more relevant
- **Maintain atrial contraction and prevent tachycardia**
  - Efforts to maintain sinus rhythm in atrial fibrillation may be beneficial (atrial fibrillation ablation may reduce morbidity and mortality)
- **Treat and prevent myocardial ischemia**
  - May mimic HF as an “angina equivalent”
- **Detect and treat sleep apnea**
  - Common co-morbidity causing systemic hypertension, pulmonary hypertension, and right heart dysfunction (adaptive servo-ventilation ineffective)
- **Lifestyle modification**
  - Diet and exercise to promote weight reduction and improve functional capacity

Specific Therapy Targets
(beyond general management)

- **Renin-angiotensin-aldosterone–directed therapy**
  - ACEIs and ARBs ineffective (except in “prevention”)
  - Aldosterone antagonists (may be beneficial)
- **Digoxin**
  - Ineffective (may reduce hospitalizations)
- **Beta blockers and calcium channel blockers**
  - Ineffective (useful in preventing tachycardia in patients with AF)
- **Phosphodiesterase-5 inhibitors**
  - Sildenafil ineffective
- **Novel Therapy**
  - ARNIs (may be effective in selected patients)
  - SGLT-2 inhibitors (reduce HF hospitalization)
- **Chronotropic insufficiency**
  - ? Targeted pacing (likely ineffective)
- **Obesity treatment**
  - GLP-1 agonists (improve QOL irrespective of diabetes mellitus)

FIGURE 265-1 Pathophysiologic correlations, general therapeutic principles, and results of specific “directed” therapy in heart failure (HF) with preserved ejection fraction. ACEI, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; ARNI, angiotensin receptor-neprilysin inhibitor; GLP-1, glucagon-like peptide 1; SGLT-2, sodium-glucose cotransporter 2; QOL, quality of life.

primary statistical analysis ($p = .06$). Directional benefits in secondary endpoints including QOL, NYHA class, and renal function favoring sacubitril-valsartan support a possible modest benefit of neprilysin inhibition in this population, particularly among patients with lower (i.e., mildly reduced) EF and in women, subgroups who appeared to derive greater benefit. Based on these data, sacubitril-valsartan has recently been approved in the United States for treatment of symptomatic HF across the full spectrum of EF, with benefits acknowledged to be greatest in those with LVEF below normal. The PARAGLIDE-HF trial randomly assigned patients with HF and EF >40% within 30 days of a worsening HF event to treatment with sacubitril-valsartan or valsartan. Although those assigned to sacubitril-valsartan experienced greater reductions in N-terminal prohormone of brain natriuretic peptide (NTproBNP) at 8 weeks, clinical benefits were modest and appeared to be confined to those with EF ≤60%, suggesting this therapy may be most effective in that subgroup.

CLINICAL GUIDING PRINCIPLES

Beyond disease-modifying therapy, treatment of HFpEF should focus on decongestion, aggressive management of medical comorbidities, and relief of exacerbating factors. A careful diagnostic approach is critical, since patients with HF and a normal or near-normal LVEF compose a heterogenous group that includes patients with infiltrative heart disease (amyloidosis, hemochromatosis, sarcoidosis), storage disease (Fabry’s disease, Gaucher’s disease), hypertrophic cardiomyopathy, pericardial disease, pulmonary arterial hypertension, valvular heart disease, and primary right ventricular failure who may require a different management approach. For those with true HFpEF, aggressive control of blood pressure to guideline-recommended targets and relief of volume overload with diuretics are critical to symptom relief. Excessive decrease in preload with diuretics and vasodilators may lead to underfilling the ventricle and subsequent azotemia, hypotension, and syncope. For patients at risk for coronary heart disease, deliberate evaluation for ischemia and consideration of coronary revascularization may be important. Since clinical outcomes in HFpEF are worse in the setting of atrial fibrillation, aggressive rate control, anticoagulation, and early consideration of sinus rhythm restoration are important. Comorbidities such as obesity, obstructive lung disease, sleep apnea, chronic kidney disease, and anemia/iron deficiency are increasingly recognized as important contributors to diminished functional capacity and QOL in patients with HFpEF and may be additional targets for therapy. Some investigators have suggested that the exercise intolerance in HFpEF is a manifestation of chronotropic insufficiency. While this hypothesis appears to be supported by small trials randomized trials of beta blocker withdrawal, improving chronotropic response through rate-adaptive atrial pacing did not improve functional capacity in the randomized RAPID-HF trial (Fig. 265-1).

ACUTE DECOMPENSATED HEART FAILURE

GENERAL PRINCIPLES

ADHF is a heterogeneous clinical syndrome most often resulting in need for hospitalization due to confluence of interrelated abnormalities of decreased cardiac performance, renal dysfunction, and alterations in vascular compliance. Admission with a diagnosis of ADHF is associated with excessive morbidity and mortality, with nearly half of these patients readmitted for management within 6 months, and a high short-term (5% in-hospital) and long-term cardiovascular mortality

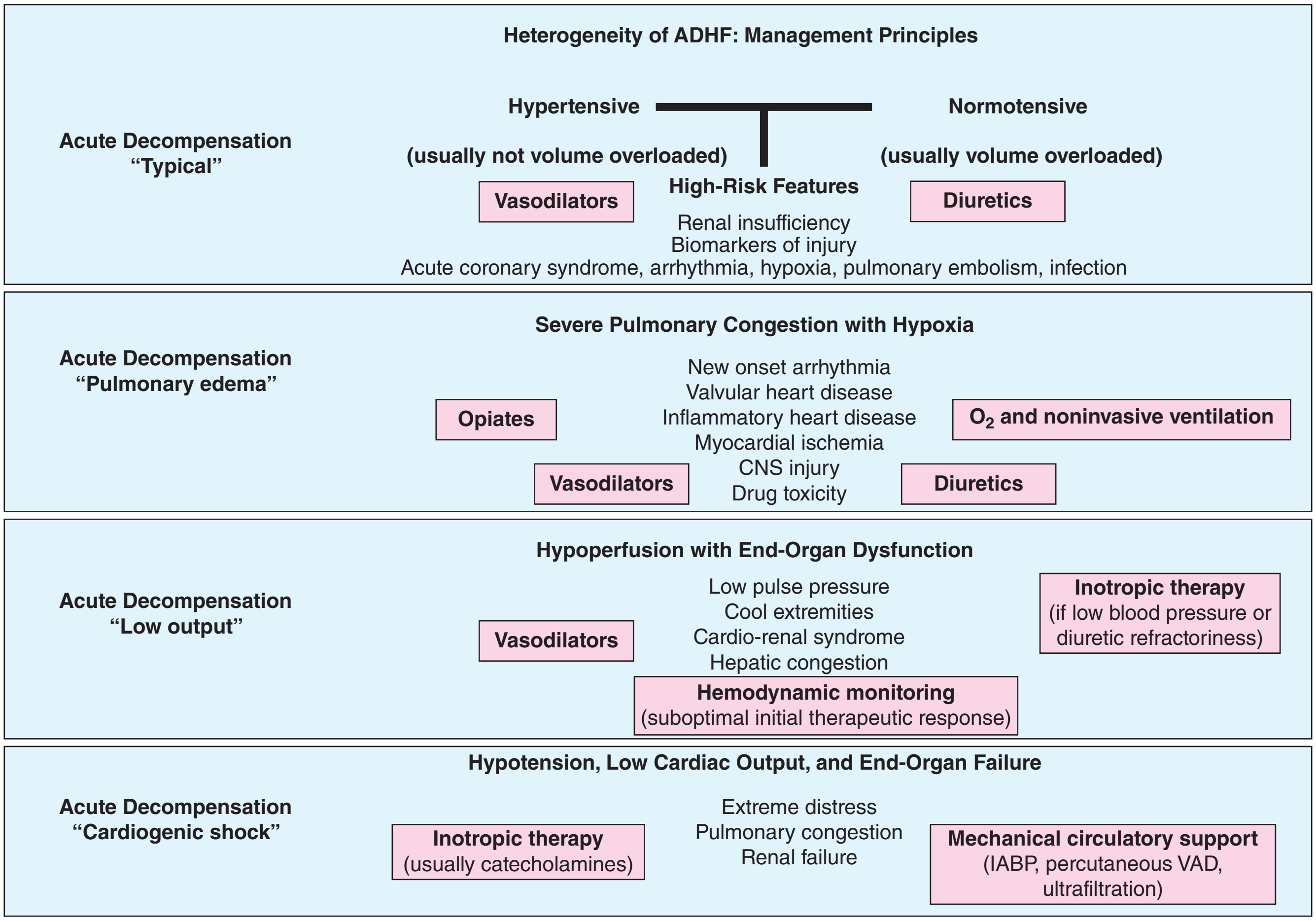
Heterogeneity of ADHF: Management Principles
Acute Decompensation "Typical"
Hypertensive
(usually not volume overloaded)
Normotensive
(usually volume overloaded)
Vasodilators
High-Risk Features
Diuretics
Renal insufficiency
Biomarkers of injury
Acute coronary syndrome, arrhythmia, hypoxia, pulmonary embolism, infection
Acute Decompensation "Pulmonary edema"
Severe Pulmonary Congestion with Hypoxia
Opiates
New onset arrhythmia
Valvular heart disease
Inflammatory heart disease
Myocardial ischemia
CNS injury
Drug toxicity
O2 and noninvasive ventilation
Vasodilators
Diuretics
Acute Decompensation "Low output"
Hypoperfusion with End-Organ Dysfunction
Low pulse pressure
Cool extremities
Cardio-renal syndrome
Hepatic congestion
Vasodilators
Inotropic therapy (if low blood pressure or diuretic refractoriness)
Hemodynamic monitoring (suboptimal initial therapeutic response)
Acute Decompensation "Cardiogenic shock"
Hypotension, Low Cardiac Output, and End-Organ Failure
Inotropic therapy (usually catecholamines)
Extreme distress
Pulmonary congestion
Renal failure
Mechanical circulatory support (IABP, percutaneous VAD, ultrafiltration)

FIGURE 265-2 The distinctive phenotypes of acute decompensated heart failure (ADHF), their presentations, and suggested therapeutic routes. (Unique causes of ADHF, such as isolated right heart failure and pericardial disease, and rare causes, such as aortic and coronary dissection or ruptured valve structures or sinuses of Valsalva, are not delineated and are covered elsewhere.) CNS, central nervous system; IABP, intraaortic balloon pump; VAD, ventricular assist device.

(20% at 1 year). Importantly, long-term outcomes remain poor, with a combined incidence of cardiovascular deaths, HF hospitalizations, myocardial infarction, strokes, or sudden death reaching 50% at 12 months after hospitalization. The management of these patients remains difficult and principally revolves around volume control and hemodynamic optimization to maximize end-organ perfusion.

The first principle of management in ADHF is to identify and address the factors that precipitated decompensation. Important historical factors to consider are nonadherence to medications, dietary salt indiscretion, and usage of medications (including over-the-counter preparations) that may exacerbate HF, including nonsteroidal anti-inflammatory drugs, thiazolidinediones, tumor necrosis factor inhibitors, selected antidepressants, selected cancer therapies, cold and flu preparations with cardiac stimulants, and some herbal preparations. Coronary ischemia frequently drives HF exacerbation in patients with atherosclerotic cardiovascular disease and should be systematically investigated (either invasively or noninvasively) in all patients at risk to identify candidates for revascularization. Atrial and ventricular arrhythmias are common contributors to HF exacerbation and may trigger the need for antiarrhythmic drug suppression, cardioversion, or catheter ablation. Valvular heart disease is increasingly recognized as a target for therapy in patients with recurrent HF exacerbations and can be readily identified through echocardiography. Systemic infection and pulmonary thromboembolism are additional triggers of HF decompensation and should be routinely considered.

Concurrent with the identification of HF precipitants, effective management of ADHF requires pharmacologic therapy directed at hemodynamic optimization, including relief of congestion, reduction in afterload, and maximization of vital organ perfusion. The routine use of a pulmonary artery catheter is not recommended and should be restricted to those who present with features typical of low-output HF or cardiogenic shock who may require vasopressor or mechanical circulatory support, those who are resistant or refractory to diuretic therapy, those with combined cardiorenal dysfunction in whom therapeutic goals are difficult to define at the bedside, and those with known or suspected pulmonary arterial hypertension in whom vasodilator therapy may be appropriate. Analysis of in-hospital registries has identified several parameters associated with worse outcomes: a blood urea nitrogen level >43 mg/dL (to convert to mmol/L, multiply by 0.357), systolic blood pressure <115 mmHg, a serum creatinine level >2.75 mg/dL (to convert to μmol/L, multiply by 88.4), and elevated cardiac biomarkers including natriuretic peptides and cardiac troponins. A useful clinical schema to identify treatment targets for the various phenotypic presentations and management goals in ADHF is depicted in Fig. 265-2.

VOLUME MANAGEMENT

Intravenous Diuretic Agents Intravenous loop diuretic agents rapidly and effectively relieve symptoms of congestion and are essential when oral drug absorption is impaired. When high doses of diuretic agents are required or when the effect of bolus dosing is suboptimal, a continuous infusion may be needed to reduce toxicity and maintain stable serum drug levels. Randomized clinical trials of high- versus low-dose and bolus versus continuous infusion diuresis have not provided clear justification for the best diuretic strategy in ADHF, and as such, the use of diuretic regimens remains an art rather than science. For those refractory to loop diuretic treatment alone, addition of a thiazide diuretic agent such as chlorothiazide or metolazone to provide sequential nephron blockade may enhance natriuresis and facilitate decongestion, but also increases the risk of significant hypokalemia. Addition of acetazolamide to loop diuretic therapy in the randomized ADVOR trial was demonstrated to facilitate greater decongestion but was not associated with reduction in HF readmissions or mortality.

Change in weight is often used as a surrogate for adequate diuresis, but this objective measure of volume status may be surprisingly difficult to interpret, and weight loss during hospitalization does not necessarily correlate closely with outcomes. Effective decongestion may also be confirmed by improvement in clinical symptoms as well as the bedside examination documenting normalization of the jugular venous pressure, clearance of pulmonary rales, suppression of cardiac gallops, and resolution of peripheral edema, hepatomegaly, and abdominal ascites. It is generally advisable to continue diuresis until euvolemia has been achieved, since residual congestion or volume overload is strongly associated with risk for recurrent decompensation. Predischarge measurement of natriuretic peptide levels, which are highly correlated with risk for postdischarge mortality and readmission, may also be useful in assessing the adequacy of therapy and stratifying risk. Chronic oral loop diuretic therapy is appropriate at discharge for most patients to maintain decongestion. The TRANSFORM-HF trial did not demonstrate any mortality or morbidity advantage of torsemide compared to furosemide for this purpose despite its greater oral bioavailability, longer half-life, and other potential beneficial effects on myocardial fibrosis, aldosterone production, sympathetic activation, ventricular remodeling, and natriuretic peptides noted in small studies.

The Cardiorenal Syndrome The cardiorenal syndrome is increasingly being recognized as a complication of ADHF. Multiple definitions have been proposed for the cardiorenal syndrome, but at its simplest, it can be thought to reflect the interplay between abnormalities of heart and kidney function, with deteriorating function of one organ while therapy is administered to preserve the other. Approximately 30% of patients hospitalized with ADHF exhibit abnormal renal function at baseline, and this is associated with longer hospitalizations and increased mortality. However, mechanistic studies have been largely unable to find correlation between deterioration in renal function, cardiac output, left-sided filling pressures, and reduced renal perfusion; most patients with cardiorenal syndrome demonstrate a preserved cardiac output. It is hypothesized that in patients with established HF, this syndrome represents a complex interplay of neurohormonal factors, potentially exacerbated by “backward failure” resulting from increased intraabdominal pressure and impairment in return of renal venous blood flow. Continued use of diuretic therapy may be associated with a reduction in glomerular filtration rate and a worsening of the cardiorenal syndrome when right-sided filling pressures remain elevated. In patients in the late stages of disease characterized by profound low cardiac output state, inotropic therapy or mechanical circulatory support has been shown to preserve or improve renal function in selected individuals in the short term until more definitive therapy such as assisted circulation or cardiac transplantation is implemented.

Ultrafiltration Ultrafiltration (UF) is an invasive fluid removal technique that may supplement the need for diuretic therapy. Proposed benefits of UF include controlled rates of fluid removal, neutral effects on serum electrolytes, and decreased neurohormonal activity. This technique has also been referred to as aquapheresis in recognition of its electrolyte depletion–sparing effects. In an initial study evaluating UF versus conventional therapy, fluid removal was improved, and subsequent HF hospitalizations and urgent clinic visits were reduced with UF; however, no improvement in renal function and no subjective differences in dyspnea scores or adverse outcomes were noted. In the Cardiorenal Rescue Study in Acute Decompensated Heart Failure (CARRESS-HF) trial, 188 patients with ADHF and worsening renal failure were randomized to stepped pharmacologic care or UF. The primary endpoint was a change in serum creatinine and change in weight (reflecting fluid removal) at 96 h. Although similar weight loss occurred in both groups (~5.5 kg), there was a rise in serum creatinine among patients allocated to the UF group. Deaths and hospitalizations for HF were no different between groups, but there were more adverse events in the UF group, mainly due to kidney failure, bleeding complications, and intravenous catheter-related complications. This investigation argues against using UF as a primary strategy in patients with ADHF who are diuretic-responsive. Whether UF is useful as a rescue strategy in diuretic refractory patients with advanced renal disease remains an open question, and this strategy continues to be employed judiciously in such situations.

■ VASOACTIVE THERAPY

Vasodilators including *intravenous nitroglycerin, sodium nitroprusside*, and *nesiritide* (a recombinant brain-type natriuretic peptide) are frequently used in ADHF to lower intracardiac filling pressures and reduce systemic vascular tone. Rapid reduction in ventricular preload and afterload with these therapies may be effective in providing symptom relief in patients with pulmonary edema and in restoring end-organ perfusion for those with low cardiac output and high systemic vascular resistance. Nitroglycerine principally impacts venous tone and ventricular preload, whereas sodium nitroprusside is a potent arterial and venous vasodilator with more comprehensive effects on both preload and afterload. While intravenous nitroglycerine is commonly utilized as an adjunct to diuretics for acute management of symptomatic HF and pulmonary edema, nitroprusside is typically reserved for use in those with adequate arterial pressure or invasive hemodynamic monitoring due to the risk for hypotension. The hemodynamic effects of nesiritide are intermediate between those of nitroglycerine and nitroprusside, with head-to-head comparisons with nitroglycerine suggesting more rapid reduction in pulmonary capillary wedge pressure and pulmonary vascular resistance. Clinical utilization of nesiritide has waned due to concerns raised regarding heightened risks of renal insufficiency and mortality identified in early trials. The Acute Study of Clinical Effectiveness of Nesiritide in Decompensated Heart Failure (ASCEND-HF) randomizing 7141 patients with ADHF to nesiritide or placebo did not confirm this risk but identified no clear clinical benefit regarding subsequent HF admissions, mortality, or symptom relief (reduction in dyspnea). Renal function did not worsen, but increased rates of hypotension were noted. A smaller study of low-dose nesiritide in acute HF (Renal Optimization Strategies Evaluation Acute Heart Failure Study [ROSE-AHF]) also showed no incremental benefit over intravenous diuretics for relief of congestion or preservation of renal function. Despite apparent safety in ADHF, the routine use of nesiritide is accordingly not recommended.

Other novel vasodilators have been explored for the management of ADHF. Recombinant human relaxin-2, or *serelaxin*, is a vasodilatory hormone known to contribute to cardiovascular and renal adaptations during pregnancy. In the Relaxin in Acute Heart Failure (RELAX-AHF) trial, 1161 patients hospitalized with ADHF, evidence of congestion, and systolic pressure >125 mmHg were randomized to treatment with serelaxin or placebo in addition to standard HF therapy. Serelaxin improved dyspnea, reduced signs and symptoms of congestion, and was associated with less early worsening of HF. A positive signal of reduced mortality identified in an exploratory analysis prompted a second study (RELAX-AHF2), which did not confirm an effect on cardiovascular death or worsening HF. Accordingly, this agent was not approved for use in clinical practice.

One hypothesis for the failure of vasodilator therapies to improve clinical outcomes in ADHF despite favorable hemodynamic effects is related to the acute injury hypothesis; in this model, acute HF is analogous to presentation with an acute coronary syndrome, with the initial hours of presentation representing a period of vulnerability to myocardial damage (reflected in a rise in markers of myocyte injury such as cardiac troponins) as a consequence of abrupt increases in ventricular wall stress related to acute plasma volume expansion. To test this hypothesis, the Trial of Ularitide Safety and Efficacy in Acute Heart Failure (TRUE-AHF) randomly allocated 2157 patients with acute HF to early treatment with the synthetic natriuretic peptide *ularitide* (at a dose sufficient to reduce ventricular wall stress) or placebo. Despite a very short duration between initial clinical presentation and pharmacologic intervention (<6 h) and early hemodynamic benefits, no improvement in clinical outcomes was observed in patients allocated to ularitide at 6 months. Ularitide was associated with a higher rate of hypotension and worsening serum creatinine. These data undermine the notion that acute myocardial damage related to ventricular distension associated with HF exacerbation drives subsequent clinical outcomes and argue against the clinical importance of early vasodilator therapy in ADHF.

INOTROPIC THERAPY

Impairment of myocardial contractility often accompanies ADHF, and pharmacologic agents that increase intracellular concentration of cyclic adenosine monophosphate via direct or indirect pathways, such as sympathomimetic amines (*dopamine, dobutamine*) and phosphodiesterase-3 inhibitors (*milrinone*), respectively, serve as positive inotropic agents. Their activity leads to an increase in cytoplasmic calcium. Inotropic therapy in those with a low-output state augments cardiac output, reduces systemic vascular resistance, improves perfusion, and relieves congestion acutely. Although systematic head-to-head comparisons are available to identify a "best" agent, slight variations in the hemodynamic effects of inotropic drugs may condition selection of the appropriate drug for a given clinical context. Dopamine exhibits dose-dependent effects on dopaminergic, α-, and β-adrenergic receptors, with vasodilatory effects predominating at lower doses (<2 μg/kg per min), β-adrenergic (inotropic) effects at moderate doses, and α-adrenergic effects (vasoconstriction) at higher doses (typically >10 μg/kg per min). Low-dose ("renal dose") dopamine has been explored as an adjunctive strategy for preservation of renal function and augmentation of diuresis in acute HF but does not appear to provide incremental advantage over routine therapy with intravenous diuretics (ROSE-AHF).

Milrinone is typically associated with a greater reduction in systemic and pulmonary vascular resistance than dobutamine and, accordingly, carries a higher risk of systemic hypotension. Moreover, because milrinone has a longer half-life and is renally excreted, it requires dose adjustments in the setting of kidney dysfunction. Because milrinone acts downstream from the β_1-adrenergic receptor, it may provide an advantage in patients receiving beta blockers when admitted to the hospital.

Long-term inotropic therapy is associated with a heightened risk of mortality in HF, perhaps due to the increased risk of arrhythmia and sudden death. Routine, short-term use of milrinone in patients hospitalized with ADHF in the Outcomes of a Prospective Trial of Intravenous Milrinone for Exacerbations of Chronic Heart Failure (OPTIME-CHF) trial was associated with increased risk of atrial arrhythmias and prolonged hypotension, but no benefit regarding subsequent mortality or HF hospitalization. Accordingly, routine use of inotropic support in ADHF is discouraged, and these agents are currently indicated principally for short-term use as bridge therapy (to either left ventricular assist device support or to transplant) in cardiogenic shock or as selectively applied palliation in end-stage HF.

Novel inotropic agents that leverage the concept of myofilament calcium sensitization rather than increasing intracellular calcium levels have been introduced. *Levosimendan* is a calcium sensitizer that provides inotropic activity but also possesses phosphodiesterase-3 inhibition properties that are vasodilatory. Two trials, the second Randomized Multicenter Evaluation of Intravenous Levosimendan Efficacy (REVIVE II) and Survival of Patients with Acute Heart Failure in Need of Intravenous Inotropic Support (SURVIVE), have tested this agent in ADHF. SURVIVE compared levosimendan with dobutamine, and despite an initial reduction in circulating B-type natriuretic peptide levels in the levosimendan group compared with patients in the dobutamine group, this drug did not reduce all-cause mortality at 180 days or affect any secondary clinical outcomes. The second trial compared levosimendan against traditional noninotropic therapy and found a modest improvement in symptoms with worsened short-term mortality and ventricular arrhythmias. Although levosimendan has been approved for use to support management of HF in several countries worldwide, it is not approved for use in the United States, largely owing to the lack of compelling data for incremental efficacy in comparison with conventional inotropic drugs or standard HF therapies.

(**Table 265-1** depicts typical inotropic, vasodilator, and diuretic drugs used in ADHF.)

OTHER THERAPIES FOR ADHF

Other trials testing unique agents have yielded disappointing results in the situation of ADHF. Adenosine has been implicated as a mediator of worsening renal function and diuretic resistance, and accordingly, treatment with *adenosine receptor antagonists* was postulated to be potentially beneficial in relieving symptoms and preserving renal function in patients with acute HF. Among patients with acute HF and renal dysfunction enrolled in the Placebo-Controlled Randomized Study of the Selective A1 Adenosine Receptor Antagonist Rolofylline for Patients Hospitalized with Acute Decompensated Heart Failure and Volume Overload to Assess Treatment Effect on Congestion and Renal Function (PROTECT) trial, no cardiovascular or renal benefit was observed. Similarly, despite compelling theoretical benefit of vasopressin receptor antagonism in acute HF (based on the central role of vasopressin in mediating the fluid retention that contributes to worsening HF), no benefit of the oral selective vasopressin-2 antagonist *tolvaptan* was seen regarding mortality or HF-associated morbidity in the Efficacy of Vasopressin Antagonism in Heart Failure Outcome Study with Tolvaptan (EVEREST) trial.

CLINICAL GUIDING PRINCIPLES

In the absence of data to support specific pharmacologic interventions in ADHF, management is largely goal-directed and focused on decongestion to relieve symptoms, investigation and suppression of triggers for recurrent decompensation, and careful transition to longitudinal HF management. Patients who fail to respond adequately to medical therapy or who develop hemodynamic instability may benefit from pulmonary artery catheter placement to guide titration of vasoactive therapy or inotropic support; in those with hemodynamics suggestive of cardiogenic shock, mechanical assist devices may be required **(Chap. 271)**. Following stabilization, all patients should receive education regarding HF self-management prior to discharge, including guidance regarding diet and lifestyle modification, identification of worsening HF symptoms, and whom to contact in the event of clinical deterioration. Early postdischarge follow-up of patients following hospitalization for management of worsening HF is associated with lower rates of hospital readmission. For patients with HFrEF hospitalized with ADHF, data suggest that institution of appropriate guideline-directed medical therapy prior to hospital discharge is associated with higher rates of adherence to appropriate pharmacologic treatment in longitudinal follow-up and may be associated with improved outcomes in the early postdischarge interval. In the Comparison of Sacubitril-Valsartan Versus Enalapril on Effect on NTproBNP in Patients Stabilized from an Acute Heart Failure Episode (PIONEER-HF) study of patients with HFrEF stabilized after hospital admission for ADHF, predischarge initiation of sacubitril-valsartan compared with enalapril was associated with greater reductions in natriuretic peptides as well as lower rates of composite death and HF readmission at 8 weeks. Similarly, in the EMPULSE trial, predischarge initiation of the SGLT-2 inhibitor empagliflozin in hospitalized HF patients across the spectrum of EF was associated with reductions in a hierarchical clinical composite outcome at 90 days. More recently, in the STRONG-HF trial, an intensive postdischarge treatment strategy of frequent follow-up and rapid uptitration of guideline-recommended medical therapy was associated with lower rates of death and HF readmission at 180 days compared with usual care. These results were consistent in those with EF ≤40% and those with EF >40%, underscoring the need for early postdischarge follow-up and medical optimization in high-risk patients following a worsening HF event, regardless of EF.

HEART FAILURE WITH REDUCED EJECTION FRACTION

The past 50 years have witnessed great strides in the management of HFrEF. Treatment of symptomatic HF has evolved from a renocentric (diuretics) and hemodynamic therapy model (digoxin, inotropic therapy) to an era of disease-modifying therapy with neurohormonal antagonism. In this regard, RAAS blockers (including ARNIs), β-adrenergic receptor blockers, and, most recently, SGLT-2 inhibitors form pillars of pharmacotherapy and facilitate stabilization and even improvement in cardiac structure and function with consequent reduction in symptoms, improvement in QOL, decreased burden of hospitalizations, and a decline in mortality from both pump failure and arrhythmic deaths **(Fig. 265-3)**.

TABLE 265-1 Vasoactive Therapy in Acute Decompensated Heart Failure

DRUG CLASS	GENERIC DRUG	USUAL DOSING	SPECIAL CAUTION	COMMENTS
Inotropic therapy				Use in hypotension, end-organ hypoperfusion, or shock states
	Dobutamine	2–20 μg/kg per min	Increased myocardial oxygen demand, arrhythmia	Short acting, an advantage; variable efficacy in presence of beta blockers (requires higher doses); clinical tolerance to prolonged infusions; concerns with hypersensitivity carditis (rare)
	Milrinone	0.375–0.75 μg/kg per min	Hypotension, arrhythmia	Decrease dose in renal insufficiency; avoid initial bolus; effectiveness retained in presence of beta blockers
	Levosimendan	0.1 μg/kg per min; range, 0.05–0.2 μg/kg per min	Hypotension, arrhythmia	Long acting; should not be used in presence of low blood pressure; similar effectiveness as dobutamine but effectiveness retained in presence of beta blockers
Vasodilators				Use in presence of pulmonary congestion for rapid relief of dyspnea, in presence of a preserved blood pressure
	Nitroglycerin	10–20 μg/min, increase up to 200 μg/min	Headache, flushing, tolerance	Most common vasodilator but often underdosed; effective in higher doses
	Nesiritide	Bolus 2 μg/kg and infusion at 0.01 μg/kg per min	Hypotension	Decrease in blood pressure may reduce renal perfusion pressure; bolus may be avoided since it increases hypotension predilection
	Nitroprusside	0.3 μg/kg per min titrated to 5 μg/kg per min	Thiocyanate toxicity in renal insufficiency (>72 h)	Requires arterial line placement for titration for precise blood pressure management and prevention of hypotension
	Serelaxin	N/A (tested at 30 μg/kg per d)	Baseline blood pressure should be >125 mmHg	Not widely commercially available; ineffective in confirmatory trials
	Ularitide	15 ng/kg per min (48 h)	Baseline blood pressure >116 mmHg	Excess hypotension and increased serum creatinine
Diuretics				First line of therapy in volume overload with congestion; may use bolus or continuous dosing; initial low dose (1 × home dose) or high dose (2.5 × home dose) equally effective with higher risk of renal worsening with higher dose
	Furosemide	20–240 mg daily	Monitor for electrolyte loss	In severe congestion, use intravenously and consider continuous infusion (not trial supported)
	Torsemide	10–100 mg daily	Monitor for electrolyte loss	High bioavailability, can be given orally; anecdotally more effective in advanced heart failure states if furosemide less bioavailable (due to gut congestion)
	Bumetanide	0.5–5 mg daily	Monitor for electrolyte loss	Can be used orally; intermediate bioavailability
	Adjuvant diuretics for augmentation	N/A	Metolazone, chlorthalidone, spironolactone, acetazolamide	Acetazolamide is useful in presence of alkalosis (bicarbonate level >27mEq/L); metolazone given in 2.5- to 10-mg doses; concomitant use of loop diuretics and thiazides associated with risk for severe hypokalemia, careful laboratory monitoring advised; spironolactone is useful in presence of severe hypokalemia and normal renal function

Abbreviation: N/A, not applicable.

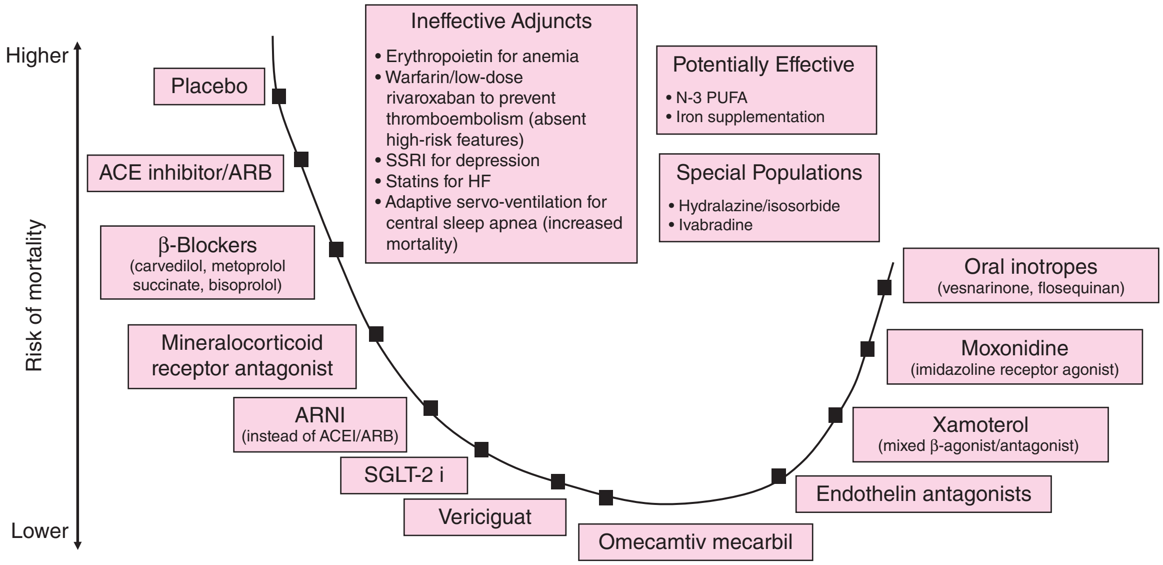
Higher
Lower
Risk of mortality
Placebo
ACE inhibitor/ARB
β-Blockers
(carvedilol, metoprolol succinate, bisoprolol)
Mineralocorticoid receptor antagonist
ARNI
(instead of ACEI/ARB)
SGLT-2 i
Vericiguat
Omecamtiv mecarbil
Endothelin antagonists
Xamoterol
(mixed β-agonist/antagonist)
Moxonidine
(imidazoline receptor agonist)
Oral inotropes
(vesnarinone, flosequinan)
Ineffective Adjuncts
• Erythropoietin for anemia
• Warfarin/low-dose rivaroxaban to prevent thromboembolism (absent high-risk features)
• SSRI for depression
• Statins for HF
• Adaptive servo-ventilation for central sleep apnea (increased mortality)
Potentially Effective
• N-3 PUFA
• Iron supplementation
Special Populations
• Hydralazine/isosorbide
• Ivabradine

FIGURE 265-3 Progressive decline in mortality with angiotensin-converting enzyme (ACE) inhibitors (ACEIs) or angiotensin receptor blockers (ARBs), angiotensin receptor-neprilysin inhibitors (ARNIs), beta blockers, mineralocorticoid receptor antagonists, sodium-glucose cotransporter-2 (SGLT-2) inhibitors, and balanced vasodilators (*selected populations such as African Americans). Addition of selected therapies (ivabradine, vericiguat) may further reduce heart failure (HF) hospitalization but does not substantially impact mortality. Further stack-on neurohormonal therapy is ineffective or results in worse outcome. Management of comorbidity (e.g., iron deficiency, sleep apnea) is of unproven efficacy. HFrEF, heart failure with reduced ejection fraction; PUFA, polyunsaturated fatty acid; SSRI, selective serotonin reuptake inhibitor.

NEUROHORMONAL ANTAGONISM

Meta-analyses suggest a 23% reduction in mortality and a 35% reduction in the combined endpoint of mortality and hospitalizations for HF in patients with symptomatic HFrEF treated with ACE inhibitors (ACEIs). Addition of β-adrenergic receptor blockers to background therapy with ACEIs provides a further 35% reduction in mortality. Although placebo-controlled studies are lacking, several noninferiority trials have demonstrated comparable efficacy of ARBs and ACEIs in patients with HFrEF, making ARBs a suitable alternative for patients who are intolerant to ACEIs due to cough or angioedema. Abundant data support the efficacy across the full spectrum of HF severity (including those with NYHA class III–IV functional capacity), as well as the safety data of these agents. These observations demonstrate the basis for the tolerability of these agents even in subgroups at higher risk for adverse effects such as those with mild-moderate chronic kidney disease. In diabetes mellitus and chronic obstructive lung disease, these agents have been established as foundational therapy for HFrEF as directed by consensus guidelines. Both agents are generally recommended for all patients with HFrEF, independent of symptom burden, and should be titrated to the doses proven to provide clinical benefit or to the maximally tolerated dose. The inability to tolerate initiation or dose titration of neurohumoral antagonists due to hypotension, worsening HF, or progressive renal insufficiency is a poor prognostic marker and may be a cardinal manifestation of transition to an advanced HF phenotype.

Class Effect and Sequence of Administration ACEIs and ARBs exert their beneficial effects in HFrEF as a class; however, the beneficial effects of beta blockers are thought to be limited to specific drugs. Beta blockers with intrinsic sympathomimetic activity (xamoterol) and other agents, including bucindolol, have not demonstrated a survival benefit. Based on the available data, beta blocker use in HFrEF should ideally be restricted to carvedilol, bisoprolol, and metoprolol succinate—agents tested and proven to improve survival in clinical trials. Whether beta blockers or ACEIs should be started first was answered by the Cardiac Insufficiency Bisoprolol Study (CIBIS) III, in which outcomes did not vary based on the sequence of drug initiation. Thus, it matters little which agent is initiated first; what does matter is that optimally titrated doses of both ACEIs and beta blockers be established in a timely manner.

Dose and Outcome In general, the benefits of neurohumoral antagonists in HFrEF are closely related to the dose achieved, girding the rationale for aggressive titration to target doses as defined by clinical trials. Prospective trials of high- versus low-dose ACEIs (ATLAS), ARBs (HEAAL), and beta blockers (MOCHA) consistently favor the higher dose, with lower rates of death and HF hospitalization seen in the higher-dose group. Clinical experience suggests that, in the absence of symptoms to suggest hypotension (fatigue and dizziness), pharmacotherapy may be uptitrated every 2 weeks in stable ambulatory patients as tolerated. Notably, data from large registries in the United States and Europe suggest that guideline-directed medical therapy for patients with HFrEF is frequently underutilized and underdosed, leaving considerable room for quality improvement.

MINERALOCORTICOID RECEPTOR ANTAGONISTS

Addition of mineralocorticoid receptor antagonists to treatment with ACEI/ARBs and beta blockers in patients with symptomatic HFrEF (NYHA class II–IV) is associated with further reductions in morbidity and mortality. Elevated aldosterone levels in HFrEF promote sodium retention, electrolyte imbalance, and endothelial dysfunction and may directly contribute to myocardial fibrosis. Hyperkalemia and worsening renal function are concerns, especially in patients with underlying chronic kidney disease, and renal function and serum potassium levels must be closely monitored. Spironolactone is the most utilized agent in this class based on efficacy demonstrated in the Randomized Aldactone Evaluation Study (RALES) in patients with HFrEF and NYHA class III–IV symptoms. Eplerenone (studied principally in patients with milder NYHA class II symptoms and those with HF or left ventricular dysfunction complication myocardial infarction) lacks the antiandrogen effects of spironolactone and may be a suitable alternative for patients who experience sexual side effects (gynecomastia, erectile dysfunction, diminished libido).

RAAS THERAPY AND NEUROHORMONAL “ESCAPE”

Since angiotensin II can be generated by non-ACE pathways, levels of angiotensin II may recover to pretreatment levels during long-term ACEI therapy. This phenomenon of neurohormonal “escape” has fueled interest in dual blockade of the RAAS using ACEI and ARBs in combination. In both the Valsartan Heart Failure Trial (Val-HeFT) and the Candesartan in Heart Failure Assessment of Reduction in Mortality and Morbidity (CHARM-Added) trial, addition of an ARB to an ACEI and other HF therapy was associated with a lower risk of HF hospitalizations. Since neither trial mandated an evidence-based dose of an ACEI, it remains unclear whether combination therapy was clearly superior to a strategy of maximizing a single agent through dose titration. Subsequent data from the Valsartan in Acute Myocardial Infarction (VALIANT) trial suggested that the addition of the ARB valsartan to an evidence-based dose of the ACEI captopril in patients with HF complicating myocardial infarction was associated with an increase in adverse events without any added benefit compared with monotherapy for either group. The findings of the VALIANT trial are also informed by the Aliskiren Trial to Minimize Outcomes in Patients with Heart Failure (ATMOSPHERE), which randomly allocated 7016 patients with HFrEF to treatment with enalapril (targeted dose 10 mg twice daily as recommended by guidelines), the plasma renin inhibitor aliskiren, or the combination on top of standard HF therapy. In that study, combination treatment with aliskiren and enalapril was associated with higher rates of hyperkalemia, hypotension, and worsening renal function, but no incremental benefit regarding HF hospitalization or cardiovascular mortality. Together, these data argue for a ceiling of benefit of angiotensin inhibition in HFrEF, beyond which further inhibition brings more adverse effects without additional efficacy. Guidelines discourage the combination of an ACEI, ARB, and spironolactone in HFrEF due to the risks of hyperkalemia and renal dysfunction, and for patients, treatment with either an ACEI or ARB and spironolactone is deemed most appropriate.

ALTERNATIVE VASODILATORS

The combination of hydralazine and nitrates has been demonstrated to improve survival in HFrEF. Hydralazine reduces systemic vascular resistance and induces arterial vasodilatation by affecting intracellular calcium kinetics; nitrates are transformed in smooth muscle cells into NO, which stimulates cyclic guanosine monophosphate production and consequent arterial-venous vasodilation. This combination improves survival but to a lesser extent than ACEIs. However, in individuals with HFrEF unable to tolerate RAAS-based therapy for reasons such as renal insufficiency or hyperkalemia, this combination is preferred as a disease-modifying approach. A trial conducted in self-identified African Americans, the African-American Heart Failure Trial (A-HeFT), studied a fixed dose of isosorbide dinitrate with hydralazine in patients with advanced symptoms of HFrEF who were receiving standard background therapy including an ACEI and beta blocker. The study demonstrated improvements in survival and hospital admission for HF in the treatment group. Adherence to this regimen is limited by the thrice-daily dosing schedule.

NOVEL NEUROHORMONAL ANTAGONISTS

Despite an abundance of animal and clinical data demonstrating deleterious effects of activated neurohormonal pathways beyond the RAAS and sympathetic nervous system, targeting such pathways with incremental blockade has been largely unsuccessful. As an example, the endothelin antagonist bosentan is associated with worsening HF in HFrEF despite demonstrating benefits in right-sided HF due to pulmonary arterial hypertension. Similarly, the centrally acting sympatholytic agent moxonidine worsens outcomes in left HF. The combined drug omapatrilat hybridizes an ACEI with a neutral endopeptidase (neprilysin) inhibitor, and this agent was tested in the Omapatrilat

TABLE 265-2 Guideline-Directed Pharmacologic Therapy and Target Doses in Heart Failure with Reduced Ejection Fraction

DRUG CLASS	GENERIC DRUG	MEAN DAILY DOSE IN CLINICAL TRIALS (mg)	INITIATION (mg)	TARGET DOSE (mg)
Angiotensin-Converting Enzyme Inhibitors				
	Lisinopril	4.5–33	2.5–5 qd	20–35 qd
	Enalapril	17	2.5 bid	10–20 bid
	Captopril	123	6.25 tid	50 tid
	Trandolapril	N/A	0.5–1 qd	4 qd
Angiotensin Receptor Blockers				
	Losartan	129	50 qd	150 qd
	Valsartan	254	40 bid	160 bid
	Candesartan	24	4–8 qd	32 qd
Aldosterone Antagonists				
	Eplerenone	42.6	25 qd	50 qd
	Spironolactone	26	12.5–25 qd	25–50 qd
Beta Blockers				
	Metoprolol succinate CR/XL	159	12.5–25 qd	200 qd
	Carvedilol	37	3.125 bid	25–50 bid
	Bisoprolol	8.6	1.25 qd	10 qd
Arteriovenous Vasodilators				
	Hydralazine isosorbide dinitrate	270/136	37.5/20 tid	75/40 tid
	Fixed-dose hydralazine/isosorbide dinitrate	143/76	37.5/20 qid	75/40 qid
Angiotensin Receptor-Neprilysin Inhibitor				
	Sacubitril-valsartan	375	100 bid	200 bid
SGLT-2 Inhibitor				
	Dapagliflozin	10	10 qd	10 qd
	Empagliflozin	10	10 qd	10 qd
	Sotagliflozin	200	200 qd	200 qd
Novel Therapies				
	Vericiguat (sGC stimulator)	9.2	2.5 qd	10 qd
	Omecamtiv mecarbil (myosin activator)	Not reported	25 bid	Up to 50 mg bid (based on plasma concentrations)

Abbreviations: sGC, soluble guanylyl cyclase; SGLT-2, sodium-glucose cotransporter 2.

Versus Enalapril Randomized Trial of Utility in Reducing Events (OVERTURE) trial. This drug did not favorably influence the primary outcome measure of the combined risk of death or hospitalization for HF requiring intravenous treatment compared with enalapril alone, and notably, the risk of angioedema was increased in patients assigned to omapatrilat.

The risk of angioedema with composite ACE/neprilysin inhibition appears to be related to excessive blockade of bradykinin breakdown by this combination. Blockade of angiotensin at the receptor level with an ARB leaves the ACE pathway for bradykinin breakdown intact and is associated with lower angioedema risk. Recently, a composite ARNI, sacubitril-valsartan (formerly LCZ696), was developed and applied to the treatment of patients with HFrEF. In the PARADIGM-HF trial, 8399 patients with HFrEF treated with guideline-directed medical therapy were randomly allocated to treatment with either enalapril or sacubitril-valsartan after a run-in period designed to ensure tolerability of both drugs at target doses. Compared to those assigned to enalapril, patients assigned to sacubitril-valsartan experienced a dramatic 20% reduction in the composite primary endpoint of cardiovascular death or HF hospitalization and a 16% reduction in all-cause mortality, as well as clinically important improvements in QOL measures. Sacubitril-valsartan was well tolerated and associated with lower rates of hyperkalemia and worsening renal function, but greater rates of symptomatic hypotension, than enalapril. Guidelines now advocate a switch to ARNI for patients with symptomatic HFrEF who tolerate ACEIs and ARBs, and emerging data suggest that up-front utilization of ARNI in patients with de novo HF naïve to ACEIs/ARBs may also be appropriate for those with adequate blood pressure to tolerate it. Given ongoing concern for angioedema, use of ARNI is contraindicated in patients with prior history of angioedema, and those being transitioned from ACEIs should receive ARNI only after a 36-h gap to limit the risk of overlap. Table 265-2 lists the common neurohormonal and vasodilator regimens for HFrEF.

HEART RATE MODIFICATION

Distinct from β-adrenergic receptor blockers, ivabradine, an inhibitor of the I_f current in the sinoatrial node, selectively reduces heart rate without affecting cardiac contractility or vascular tone. The Systolic Heart Failure Treatment with Ivabradine Compared with Placebo Trial (SHIFT) was conducted in patients with NYHA class II or III HFrEF, prior HF hospitalization, sinus rhythm, and heart rate >70 beats/min. Ivabradine reduced the combined endpoint of cardiovascular-related death and HF hospitalization in proportion to the degree of heart rate reduction, which supports the notion that heart rate may be a therapeutic target in patients with HFrEF in sinus rhythm. Importantly, despite a protocol requirement for patients to be treated with a maximally tolerated dose of a beta blocker prior to study entry, 10% of patients randomized were not treated with a beta blocker, and 75% were treated at subtarget doses. Accordingly, it remains unclear whether this agent would have been effective in patients receiving robust, guideline-recommended therapy for HF; however, these data do support the potential value of ivabradine as an adjunct or alternative in those who are intolerant to initiation or dose titration of beta blockers. Clinical guidelines have been adapted to encourage consideration of ivabradine in patients with HFrEF who remain symptomatic after treatment with guideline-based ACEI/ARB/ARNIs, beta blockers, and mineralocorticoid receptor antagonists; are in sinus rhythm; and have a residual heart rate >70 beats/min.

SGLT-2 INHIBITION

Inhibitors of SGLT-2 have been shown to reduce cardiovascular events and mortality among patients with type 2 diabetes mellitus at high cardiovascular risk or with established atherosclerotic cardiovascular disease. A particular signal of benefit has been seen regarding the incidence of HF hospitalization, which was reduced by 35% in comparison to placebo in the Empagliflozin Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients (EMPA-REG OUTCOMES) study. Because the cardiovascular benefits of SGLT-2 inhibition appear to be unrelated to the degree of reduction in hemoglobin A_{1c}, it has been postulated that the HF benefits of this therapy might be extended to patients without diabetes mellitus. Recently, the Dapagliflozin in Heart Failure (DAPA-HF) study randomized 4744 patients with symptomatic HFrEF treated with guideline-directed medical therapy (including a beta blocker, ACEI/ARB/ARNI, and spironolactone in >70% of patients) to treatment with either the SGLT-2 inhibitor dapagliflozin (dosage 10 mg daily) or placebo over a median duration of follow-up of 18.2 months. Patients allocated to dapagliflozin experienced a highly significant 26% reduction in the primary composite endpoint of worsening HF or death from cardiovascular causes, an effect that was consistent in both patients with (42%) and without diabetes mellitus at baseline. These results have been reinforced by the results of the EMPEROR-Reduced trial, in which 3730 patients with symptomatic HF and EF of ≤40% were randomized to treatment with empagliflozin (dosage 10 mg once daily) or placebo in addition to recommended therapy. Over a median 16-month follow-up, patients allocated to empagliflozin experienced a 25% reduction in the primary composite endpoint of cardiovascular death or hospitalization for HF, an effect that was again consistent regardless of the presence or absence of diabetes mellitus. Together, these studies have driven consensus guidelines to consider use of SGLT-2 inhibitors as foundational therapy for HF alongside ARNIs, beta blockers, and mineralocorticoid receptor antagonists.

SOLUBLE GUANYLYL CYCLASE STIMULATION

Soluble guanylyl cyclase (sGC) is a key enzyme of the NO signaling pathway that catalyzes synthesis of cyclic guanosine monophosphate (GMP), producing vasodilation. Vericiguat is a novel oral sGC stimulator that enhances cyclic GMP and NO signaling by directly stimulating sGC and sensitizing sGC to endogenous NO. The Vericiguat Global Study in Subjects with Heart Failure with Reduced Ejection Fraction (VICTORIA) randomly assigned 5050 patients with chronic HF, NYHA class II–IV symptoms, LVEF <45%, elevated natriuretic peptide levels, and evidence of worsening HF (requiring recent hospitalization or intravenous diuretic therapy) despite guideline-directed medical therapy to treatment with vericiguat (target dose 10 mg) or matching placebo over a median follow-up of 11 months. The primary study results were notable for a modest 10% relative risk reduction in the primary composite outcome of cardiovascular death or HF hospitalization among those assigned to vericiguat, an effect driven principally by effects on HF hospitalization, rather than cardiovascular death. As vericiguat was generally well tolerated with a low rate of serious adverse events, these data suggest a potential role for sGC stimulation as an adjunct to guideline-directed medical therapy in the high-risk group of HFrEF patients with recent congestive exacerbations requiring treatment, although these data await further review by regulatory agencies and guidelines committees.

MYOSIN ACTIVATION

A novel approach to augmentation of cardiac output is to prolong ventricular systole without increasing myocardial contractility. As a selective myosin activator, *omecamtiv mecarbil* prolongs the ejection period and increases fractional shortening without altering the force of contraction. This agent, distinct from inotropic agents, is not associated with an increase in myocardial oxygen demand. Importantly, the drug is available for oral, rather than intravenous, administration, enabling chronic use in the ambulatory setting. In the COSMIC-HF (Chronic Oral Study of Myosin Activation to Increase Contractility in Heart Failure) trial of 448 patients with chronic HF and left ventricular systolic dysfunction, treatment with omecamtiv mecarbil for 20 weeks was associated with significant improvements in cardiac function and indices of left ventricular remodeling, as well as reductions in natriuretic peptide levels. Notably, the safety profile was comparable to placebo, with no increase in cardiac adverse events despite a modest increase in cardiac troponins in patients allocated to omecamtiv mecarbil. These promising preliminary data fueled a larger clinical outcomes trial (GALACTIC-HF), in which 8256 patients with symptomatic chronic HF and EF of ≤35% were randomized to treatment with omecamtiv mecarbil (25–50 mg twice daily based on plasma levels) or placebo in addition to standard HF therapy. Over a median follow-up of 21.8 months, patients allocated to omecamtiv mecarbil experienced a 14% reduction in the primary composite endpoint of death from cardiovascular causes or first HF event (hospitalization or urgent visit for HF), an outcome driven principally by reduction in HF events (no measurable effect on cardiovascular death alone). A possible signal of greater benefit in patients with features of advanced HF (lower EF, higher natriuretic peptide levels, more severe symptoms) combined with a favorable safety and tolerability profile suggests a possible role for this agent in patients with late-stage disease, although additional study is needed.

DIGOXIN

Digitalis glycosides exert a mild inotropic effect, attenuate carotid sinus baroreceptor activity, and are sympatho-inhibitory. These effects decrease serum norepinephrine levels, plasma renin levels, and possibly aldosterone levels. The Digitalis Investigation Group (DIG) trial demonstrated a reduction in HF hospitalizations in the treatment group (patients with HF and sinus rhythm) but no reduction in mortality or improvement in QOL. Importantly, treatment with digoxin resulted in a higher mortality rate and hospitalizations in women than men. It should be noted that low doses of digoxin are sufficient to achieve any potentially beneficial outcomes, and higher doses breach the therapeutic safety index. Although digoxin levels should be checked to minimize toxicity and although dose reductions are indicated for higher levels, no adjustment is made for low levels. Generally, digoxin is now relegated as late-line therapy for patients who remain profoundly symptomatic despite optimal neurohormonal blockade and adequate volume control.

ORAL DIURETICS

Neurohormonal activation results in avid salt and water retention. Diuretic therapy is typically required in patients with symptomatic HF to remedy congestive symptoms as a prelude to initiation and titration of neurohormonal therapy. Because of their potent effect on renal sodium excretion, loop diuretic agents are the preferred agents, with thiazide diuretics reserved for use in combination with loop diuretics for those with refractory volume overload. Frequent dose adjustments of loop diuretics may be necessary during longitudinal follow-up of patients with HF because of variable oral absorption and fluctuations in renal function. Patients who fail to respond to furosemide at high doses may benefit from transition to torsemide or bumetanide, which have greater oral bioavailability, but recent studies have not confirmed a greater benefit for alternatives to furosemide as a loop diuretic. Importantly, clinical trial data confirming efficacy are limited, and no data suggest that these agents improve survival. Since loop diuretics do enhance neurohumoral activation, dosing should be minimized as possible to maximize the balance of risk and benefit.

CALCIUM CHANNEL ANTAGONISTS

Amlodipine and felodipine, second-generation calcium channel–blocking agents, safely and effectively reduce blood pressure in HFrEF but do not affect morbidity, mortality, or QOL. The first-generation agents, including verapamil and diltiazem, may exert negative inotropic effects and destabilize previously asymptomatic patients. Accordingly, their use should be discouraged in patients with HFrEF.

ANTI-INFLAMMATORY THERAPY

Targeting inflammatory cytokines such as tumor necrosis factor α (TNF-α) for the management of HF by using anticytokine agents such as infliximab and etanercept has been unsuccessful and associated with worsening HF. Use of intravenous immunoglobulin therapy in nonischemic etiology of HF has not been shown to result in beneficial outcomes. Nonspecific immunomodulation has been tested in the Advanced Chronic Heart Failure Clinical Assessment of Immune Modulation Therapy (ACCLAIM-HF) trial where ex vivo exposure of a blood sample from systolic HF patients to controlled oxidative stress was hypothesized to initiate apoptosis of leukocytes soon after intramuscular gluteal injection of the treated sample. The physiologic response to apoptotic cells results in a reduction in inflammatory cytokine production and upregulation of anti-inflammatory cytokines. This promising hypothesis was not proven, although certain subgroups (those with no history of previous myocardial infarction and those with mild HF) showed signals in favor of immunomodulation. Most recently, in the Canakinumab Anti-inflammatory Thrombosis Outcomes Study (CANTOS), treatment of post–myocardial infarction patients with elevated high-sensitivity C-reactive protein using a monoclonal antibody targeted at interleukin 1β was associated with a dose-dependent reduction in hospitalization for HF and HF-associated mortality in post hoc analyses. Whether this approach might have relevance for patients with established HF remains unclear.

HMG-CoA REDUCTASE INHIBITORS (STATINS)

Potent lipid-altering and pleiotropic effects of statins reduce major cardiovascular events and improve survival in non-HF populations. Once HF is well established, this therapy may not be as beneficial and theoretically could even be detrimental by depleting ubiquinone in the electron transport chain. Two trials, Controlled Rosuvastatin Multinational Trial in Heart Failure (CORONA) and Gruppo Italiano per lo Studio della Sopravvivenza nell'Insufficienza Cardiac (GISSI-HF), have tested low-dose rosuvastatin in patients with HFrEF and demonstrated no improvement in aggregate clinical outcomes. If statins are required to treat progressive atherosclerotic vascular disease or significant dyslipidemia in the background setting of HF, then they should be employed. However, no clear rationale appears to exist for routine use of statin therapy in nonischemic HF.

ANTICOAGULATION AND ANTIPLATELET THERAPY

HFrEF is accompanied by a hypercoagulable state and therefore a high risk of thromboembolic events, including stroke, pulmonary embolism, and peripheral arterial embolism. Although the value of long-term oral anticoagulation is established in certain groups, including patients with atrial fibrillation, the data are insufficient to support the use of warfarin in patients in normal sinus rhythm without a history of thromboembolic events or echocardiographic evidence of left ventricular thrombus. In the large Warfarin versus Aspirin in Reduced Cardiac Ejection Fraction (WARCEF) trial, full-dose aspirin or international normalized ratio–controlled warfarin was tested with follow-up for 6 years. Among patients with reduced LVEF in sinus rhythm, there was no significant overall difference in the primary outcome between treatment with warfarin and treatment with aspirin. A reduced risk of ischemic stroke with warfarin was offset by an increased risk of major hemorrhage. A recent trial of the direct oral anticoagulant rivaroxaban at low dose (2.5 mg daily) for patients with ischemic heart disease, HFrEF, and sinus rhythm also indicated no reduction in stroke or ischemic events compared with placebo. Aspirin blunts ACEI-mediated prostaglandin synthesis, but the clinical importance of this finding remains unclear. Current guidelines support the use of aspirin in patients with ischemic cardiomyopathy who do not have a contraindication, although the risk-benefit balance in older patients at higher risk for bleeding complications may be less favorable.

FISH OIL

Treatment with long-chain omega-3 polyunsaturated fatty acids (ω-3 PUFAs) has been shown to be associated with modestly improved clinical outcomes in patients with HFrEF. This observation from the GISSI-HF trial was extended to measurements of ω-3 PUFAs in plasma phospholipids at baseline and after 3 months. Three-month treatment with ω-3 PUFAs enriched circulating eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). Low EPA levels are inversely related to total mortality in patients with HFrEF.

MICRONUTRIENTS

A growing body of evidence suggests an association between HF and micronutrient status. Reversible HF has been described because of severe thiamine and selenium deficiency. Thiamine deficiency has received attention in HF because malnutrition and diuretics are prime risk factors for thiamine loss. Small exploratory randomized studies have suggested a benefit of supplementation with thiamine in HFrEF with evidence of improved cardiac function. This finding is restricted to chronic HF states and does not appear to be beneficial in the ADHF phenotype. Due to the exploratory nature of the evidence, no recommendations for routine supplementation or testing for thiamine deficiency can be made.

ENHANCED EXTERNAL COUNTERPULSATION

Peripheral lower extremity therapy using graded external pneumatic compression at high pressure is administered in 1-h sessions for 35 treatments (7 weeks) and has been proposed to reduce angina symptoms and extend time to exercise-induced ischemia in patients with coronary artery disease. The Prospective Evaluation of Enhanced External Counterpulsation in Congestive Heart Failure (PEECH) study assessed the benefits of enhanced external counterpulsation in the treatment of patients with mild-to-moderate HF. This randomized trial improved exercise tolerance, QOL, and NYHA functional classification but without an accompanying increase in peak oxygen consumption. A placebo effect due to the nature of the intervention simply cannot be excluded.

EXERCISE

The Heart Failure: A Controlled Trial Investigating Outcomes of Exercise Training (HF-ACTION) study investigated short-term (3-month) and long-term (12-month) effects of a supervised exercise training program in patients with moderate HFrEF. Exercise was safe, improved patients' sense of well-being, and correlated with a trend toward mortality reduction. Maximal changes in 6-min walk distance were evident at 3 months with significant improvements in cardiopulmonary exercise time and peak oxygen consumption persisting at 12 months. Therefore, exercise training is recommended as an adjunctive treatment in patients with HF.

MANAGEMENT OF SELECTED COMORBIDITY

Sleep-disordered breathing is common in HF and particularly in HFrEF. A range of presentations exemplified by obstructive sleep apnea, central sleep apnea, and its extreme form of Cheyne-Stokes breathing are noted. Frequent periods of hypoxia and repeated micro- and macro-arousals trigger adrenergic surges, which can worsen hypertension and impair systolic and diastolic function. A high index of suspicion is required, especially in patients with difficult-to-control hypertension or with predominant symptoms of fatigue despite reverse remodeling in response to optimal medical therapy. Worsening of right heart function with improvement of left ventricular function noted on medical therapy should immediately trigger a search for underlying sleep-disordered breathing or pulmonary complications such as occult embolism or pulmonary hypertension. Treatment with nocturnal positive airway pressure improves oxygenation, LVEF, and 6-min walk distance. However, no conclusive data exist to support this therapy as a disease-modifying approach with reduction in mortality. A recent trial using adaptive servo-ventilation triggered by a minute ventilation sensor in patients who had HFrEF and predominantly central sleep apnea increased all-cause and cardiovascular mortality, so this approach should be avoided; however, modification of this approach using lower background pressures and an alternative trigger based on peak airflow appeared to be safe in patients with sleep apnea and HFrEF enrolled in the ADVENT-HF trial, although there was no discernible benefit with regard to clinical outcomes. Whether this modified approach to adaptive servo-ventilation will be appropriate for management of selected patients with HFrEF and central sleep apnea requires further study.

Anemia is common in HF patients, reduces functional status and QOL, and is associated with increased proclivity for hospital admissions and mortality. Anemia in HF is more common in the elderly, in those with advanced stages of HFrEF, in the presence of renal insufficiency, and in women and African Americans. The mechanisms include iron deficiency, dysregulation of iron metabolism, and occult gastrointestinal bleeding. Intravenous iron using either iron sucrose or carboxymaltose (Ferric Carboxymaltose Assessment in Patients with Iron Deficiency and Chronic Heart Failure [FAIR-HF] trial) has been shown to correct anemia and improve functional capacity. Another trial, CONFIRM-HF, enrolled similar patients with iron deficiency (ferritin <100 ng/mL or 100–300 ng/mL if transferrin saturation <20%) and demonstrated that use of ferric carboxymaltose in a simplified high-dose schedule resulted in improvement in functional capacity, symptoms, and QOL. These symptomatic benefits of iron repletion, however, do not appear to translate clearly to benefits on longer term clinical outcomes. In the AFFIRM-AHF trial of patients with HF, LVEF <50%, and iron deficiency, randomization to ferric carboxymaltose compared with placebo was associated with a lower rate of cardiovascular death and total HF hospitalizations that narrowly missed the margin for statistical significance. Similar results were seen in the IRONMAN trial of another iron polysaccharide, ferric derisomaltose. The results of both of these trials may have been confounded by the coronavirus disease 2019 (COVID-19) pandemic, and pooled data suggested a possible favorable effect on HF hospitalizations. However, the HEART-FID trial, which randomized 3065 ambulatory HF patients with LVEF ≤40% and iron deficiency to treatment with ferric carboxymaltose or placebo, showed no benefit with regard to the primary hierarchical composite of death, hospitalizations for HF, or 6-minute walk distance. Accordingly, it seems that benefits of iron repletion in this population are likely confined to reduction in symptoms, and perhaps HF hospitalizations, particularly among those with iron deficiency. Oral iron supplementation does not appear to be effective in treating iron deficiency in HF. Erythropoiesis-regulating agents such as erythropoietin analogues have been studied with disappointing results. The Reduction of Events by Darbepoetin Alfa in Heart Failure (RED-HF) trial demonstrated that treatment with darbepoetin alfa did not improve outcomes in patients with systolic HF but increased the risk of thromboembolism-related adverse events.

Depression is common in HFrEF, with a reported prevalence of one in five patients, and is associated with a poor QOL, limited functional status, and increased risk of morbidity and mortality in this population. However, the largest randomized study of depression in HFrEF, the Sertraline Against Depression and Heart Disease in Chronic Heart Failure (SADHART-CHF) trial, showed that although sertraline was safe, it did not provide greater reduction in depression or improve cardiovascular status among patients with HF and depression compared with nurse-driven multidisciplinary management.

Atrial arrhythmias, especially atrial fibrillation, are common and serve as a harbinger of worse prognosis in patients with HF. When rate control is inadequate or symptoms persist, pursuing a rhythm control strategy is reasonable. Rhythm control may be achieved via pharmacotherapy or by percutaneous or surgical techniques, and referral to practitioners or centers experienced in these modalities is recommended. Antiarrhythmic drug therapy should be restricted to amiodarone and dofetilide, both of which have been shown to be safe and effective but do not alter the natural history of the underlying disease. The Antiarrhythmic Trial with Dronedarone in Moderate-to-Severe Congestive Heart Failure Evaluating Morbidity Decrease (ANDROMEDA) studied the effects of the novel antiarrhythmic agent dronedarone and found an increased mortality due to worsening HF. Given the potential adverse effects and limited efficacy of pharmacologic strategies for maintenance of sinus rhythm, catheter ablation has been explored as an alternative. In the CASTLE-AF trial, among 363 patients with HF, LVEF ≤35%, and paroxysmal or persistent atrial fibrillation randomized to treatment with catheter ablation or medical therapy, those assigned to catheter ablation experienced a significantly lower rate of death or hospitalization. These results are supported by similar outcomes in the CASTLE-HTX trial, which sought to include patients with advanced HF; however, this aspect is disputed. These data argue for consideration of restoration of sinus rhythm with catheter ablation as a preferred strategy to antiarrhythmic drugs in patients with HF and reduced EF.

Diabetes mellitus is a frequent comorbidity in HF. Prior studies using thiazolidinediones (activators of peroxisome proliferator-activated receptors) have been associated with worsening HF. GLP-1 agonists such as liraglutide have also been tested and do not lead to worsening in HF but require more study. The role of SGLT-2 inhibitors in HF has been previously discussed, and they represent an important disease-modifying therapy in diabetic patients with HF.

■ NEUROMODULATION USING DEVICE THERAPY

Autonomic dysfunction is common in HF, and attempts at using devices to modulate the sympathetic and parasympathetic systems have been undertaken. Broadly, devices that achieve vagal nerve stimulation, baroreflex activation, renal sympathetic denervation, spinal cord stimulation, or left cardiac sympathetic denervation have been employed. While small preclinical and clinical studies have demonstrated benefits, large, randomized trials, when conducted, have failed. The INOVATE-HF study tested vagal nerve stimulation versus optimal medical therapy among individuals with stable HF. Vagus nerve stimulation did not reduce the rate of death or hospitalization for HF. However, functional capacity and QOL were favorably affected by vagus nerve stimulation. The ANTHEM-HFrEF trial, which enrolled only half its intended population, did not show a signal for benefit of vagal nerve stimulation.

■ CARDIAC CONTRACTILITY MODULATION

Cardiac contractility modulation (CCM) is a device-based therapy for HF that involves nonexcitatory electrical stimulation to the right ventricular septal wall during the absolute myocardial refractory period to augment the strength of subsequent myocardial contraction. A series of small, randomized, prospective clinical trials, as well as several real-world observational registries, have suggested that application of CCM to selected patients with HF may improve symptoms, functional capacity, and QOL, although no effect on hard clinical outcomes such as HF hospitalization or mortality has been established. The predominant benefits of CCM appear to accrue to those with symptomatic HFrEF (EF 25–45%) and narrow QRS duration (for whom cardiac resynchronization therapy is not an option), and the approach can be combined with an implantable defibrillator. The device is not currently endorsed for routine use by clinical treatment guidelines in the United States or Europe and is deemed to require more evidence prior to widespread adoption.

CARDIAC RESYNCHRONIZATION THERAPY

Nonsynchronous contraction between the walls of the left ventricle (intraventricular) or between the ventricular chambers (interventricular) impairs systolic function, decreases mechanical efficiency of contraction, and adversely affects ventricular filling. Mechanical dyssynchrony results in an increase in wall stress and worsens functional mitral regurgitation. The single most important association of extent of dyssynchrony is a widened QRS interval on the surface electrocardiogram, particularly in the presence of a left bundle branch block pattern. With placement of a pacing lead via the coronary sinus to the lateral wall of the ventricle, cardiac resynchronization therapy (CRT) enables a more synchronous ventricular contraction by aligning the timing of activation of the opposing walls. Early studies showed improved exercise capacity, reduction in symptoms, and evidence of reverse remodeling. The Cardiac Resynchronization in Heart Failure Study (CARE-HF) trial was the first study to demonstrate a reduction in all-cause mortality with CRT placement in patients with HFrEF on optimal therapy with continued moderate-to-severe residual symptoms of

NYHA class III or IV HF. More recent clinical trials have demonstrated disease-modifying properties of CRT in even minimally symptomatic patients with HFrEF, including the Resynchronization–Defibrillation for Ambulatory Heart Failure Trial (RAFT) and Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy (MADIT-CRT), both of which sought to use CRT in combination with an implantable defibrillator. Most benefit in mildly symptomatic HFrEF patients accrues from applying this therapy in those with a QRS width of >149 ms and a left bundle branch block pattern. Attempts to further optimize risk stratification and expand indications for CRT using modalities other than electrocardiography have proven disappointing. Echocardiographically derived measures of dyssynchrony vary tremendously, and narrow QRS dyssynchrony has not proven to be a good target for treatment. Uncertainty surrounds the benefits of CRT in those with ADHF, a predominant right bundle branch block pattern, atrial fibrillation, and evidence of scar in the lateral wall, which is the precise location where the CRT lead is positioned. His-Purkinje conduction system or left bundle branch area pacing are evolving alternatives to biventricular pacing; however, the evidence supporting their benefits over conventional CRT remains less certain.

SUDDEN CARDIAC DEATH PREVENTION IN HEART FAILURE

Sudden cardiac death (SCD) due to ventricular arrhythmias is the mode of death in approximately half of patients with HF and is particularly proportionally prevalent in HFrEF patients with early stages of the disease. Patients who survive an episode of SCD are at very high risk and qualify for placement of an implantable cardioverter-defibrillator (ICD). Although primary prevention is challenging, the degree of residual left ventricular dysfunction despite optimal medical therapy (≤35%) to allow for adequate remodeling and the underlying etiology (post–myocardial infarction or ischemic cardiomyopathy) are the two single most important risk markers for stratification of need and benefit. Currently, patients with NYHA class II or III symptoms of HF and an LVEF <35%, irrespective of etiology of HF, are appropriate candidates for ICD prophylactic therapy. In patients with a myocardial infarction and optimal medical therapy with residual LVEF ≤30% (even when asymptomatic), placement of an ICD is appropriate. A recent Danish trial suggested that prophylactic ICD implantation in patients with symptomatic systolic HF not caused by coronary artery disease was not associated with a significantly lower long-term rate of death from any cause than was usual clinical care. In this trial, benefits were noted in those aged <60 years. In patients with a terminal illness and a predicted life span of <6 months or in those with NYHA class IV symptoms who are refractory to medications and who are not candidates for transplant, the risks of multiple ICD shocks must be carefully weighed against the survival benefits. If a patient meets the QRS criteria for CRT, combined CRT with ICD is often employed **(Table 265-3)**.

TABLE 265-3 Principles of ICD Implantation for Primary Prevention of Sudden Death

PRINCIPLE	COMMENT
Arrhythmia–sudden death mismatch	Sudden death in heart failure patients is generally due to progressive LVD, not a focal arrhythmia substrate (except in patients with post-MI HF with scar)
Diminishing returns with advanced disease	Intervention at early stages of HF most successful since sudden death incidence diminishes as cause of death with increasing severity of HF
Timing of benefits	LVEF should be evaluated on optimal medical therapy or after revascularization before ICD therapy is employed; no benefit to ICD implant within 40 days of an MI (unless for secondary prevention)
Estimation of benefits and prognosis	Patients and clinicians often overestimate benefits of ICDs; an ICD discharge is not equivalent to an episode of sudden death (some ventricular arrhythmias terminate spontaneously); appropriate ICD discharges are associated with a worse near-term prognosis; recent trials with uncertain benefit in nonischemic cardiomyopathy (especially in those >68 years old)

Abbreviations: HF, heart failure; ICD, implantable cardioverter-defibrillator; LVD, left ventricular disease; LVEF, left ventricular ejection fraction; MI, myocardial infarction.

SURGICAL THERAPY IN HEART FAILURE

Coronary artery bypass grafting (CABG) is considered in patients with ischemic cardiomyopathy with multivessel coronary artery disease. The recognition that hibernating myocardium, defined as myocardial tissue with abnormal function but maintained cellular function, could recover after revascularization led to the notion that revascularization with CABG would be useful in those with living myocardium. Revascularization is most robustly supported in individuals with ongoing angina and left ventricular failure. Revascularizing those with left ventricular failure in the absence of angina remains controversial. The Surgical Treatment for Ischemic Heart Failure (STICH) trial in patients with an EF of ≤35% and coronary artery disease amenable to CABG demonstrated no significant initial benefit compared to medical therapy. However, patients assigned to CABG had lower rates of death from cardiovascular causes and of death from any cause or hospitalization for cardiovascular causes over 10 years than among those who received medical therapy alone. An ancillary study of this trial also determined that the detection of hibernation (viability) before revascularization did not materially influence the efficacy of this approach, nor did it help to define a population unlikely to benefit if hibernation was not detected. Notably, a strategy of percutaneous coronary revascularization in patients with ischemic left ventricular dysfunction in the REVIVED-BCIS2 study did not show clinical benefit.

Surgical ventricular restoration (SVR), a technique characterized by infarct exclusion to remodel the left ventricle by reshaping it surgically in patients with ischemic cardiomyopathy and dominant anterior left ventricular dysfunction, has been proposed. However, in a 1000-patient trial in patients with HFrEF who underwent CABG alone or CABG plus SVR, the addition of SVR to CABG had no disease-modifying effect. However, left ventricular aneurysm surgery is still advocated in those with refractory HF, ventricular arrhythmias, or thromboembolism arising from an akinetic aneurysmal segment of the ventricle. Other remodeling procedures, such as use of an external mesh-like net attached around the heart to limit further enlargement, have not been shown to provide hard clinical benefits, although favorable cardiac remodeling was noted.

Functional (or secondary) mitral regurgitation (MR) occurs with varying degrees in patients with HFrEF and dilated ventricles, and its severity is correlated inversely with prognosis. Annular dilatation and leaflet noncoaptation related to distorted papillary muscle geometry in the context of ventricular remodeling is typically responsible, although in patients with ischemic heart disease and prior myocardial infarction, leaflet tethering and displacement may contribute. The primary approach to management of functional MR is optimization of guideline-directed medical therapy, followed by CRT in eligible patients, but relief may be incomplete for many patients with advanced HF. In these patients with HF and severe left ventricular dysfunction who are not candidates for surgical coronary revascularization, surgical mitral valve repair (MVR) to remedy functional MR carries significant risk and remains controversial. The development of percutaneous approaches to edge-to-edge MVR has provided a less invasive approach that enables reduction in functional MR at lower risk than conventional surgery. Recently, two large, randomized trials of transcatheter MVR using this approach have been conducted in patients with symptomatic HFrEF and moderate-severe functional MR. In the Cardiovascular Outcomes Assessment of the MitraClip Percutaneous Therapy for Heart Failure Patients with Functional Mitral Regurgitation (COAPT) study, patients allocated to MVR versus standard HF therapy experienced a marked reduction in both HF hospitalizations and mortality at 2 years, supporting the efficacy of this approach. In the second trial, Percutaneous Repair with the MitraClip Device for Severe Functional/Secondary Mitral Regurgitation (MITRA-FR), which employed a similar design, the rates of death or HF hospitalization did not differ between the percutaneous MVR and medical therapy groups. The precise reason for discrepant results between these studies remains unclear but may be related to differences in background utilization of guideline-directed medical therapy, procedural success rates, and patient selection (particularly whether the severity of MR is proportionate or disproportionate to the degree of left ventricular cavity dilation). Because mortality rates at 2 years remain high even with percutaneous MVR, patients with advanced symptoms of HF in whom MR severity is driven principally by end-stage left ventricular remodeling should also be considered for advanced therapies such as mechanical circulatory support if symptoms remain refractory to medical therapy.

CELLULAR AND GENE-BASED THERAPY

The cardiomyocyte possesses regenerative capacity, and such renewal is accelerated under conditions of stress and injury, such as an ischemic event or HF. Investigations that use either bone marrow–derived precursor cells or autologous cardiac-derived cells have gained traction but have not generally improved clinical outcomes in a convincing manner. More promising, however, are cardiac-derived stem cells. Two preliminary pilot trials delivering cells via an intracoronary approach have been reported. In one, autologous c-kit–positive cells isolated from the atria obtained from patients undergoing CABG were cultured and reinfused. In another, cardiosphere-derived cells grown from endomyocardial biopsy specimens were used. These small trials demonstrated improvements in left ventricular function but require far more work to usher in a clinical therapeutic success. Efforts to utilize mesenchymal stem cells to facilitate left ventricular recovery and weaning from mechanical circulatory support in patients with left ventricular assist devices have also been disappointing. The appropriate route of administration, the quantity of cells to achieve a minimal therapeutic threshold, the constitution of these cells (single source or mixed), the mechanism by which benefit accrues, and short- and long-term safety remain to be elucidated.

Targeting molecular aberrations using gene transfer therapy, mostly with an adenoviral vector, has been tested in HFrEF. A cellular target includes calcium cycling proteins such as inhibitors of phospholamban such as SERCA2a, which is deficient in patients with HFrEF. Primarily responsible for reincorporating calcium into the sarcoplasmic reticulum during diastole, this target was tested in the CUPID (Efficacy and Safety Study of Genetically Targeted Enzyme Replacement Therapy for Advanced Heart Failure) trial. This study used coronary arterial infusion of adeno-associated virus type 1 carrying the gene for SERCA2a and initially demonstrated that natriuretic peptides were decreased, reverse remodeling was noted, and symptomatic improvements were forthcoming. However, a confirmatory trial failed to meet its primary efficacy endpoint. The DREAM-HF trial was a randomized, double-blind, multicenter study of a single transendocardial administration of mesenchymal precursor cells in patients in HFrEF. The primary and secondary endpoints of the trial were negative.

More advanced therapies for late-stage HF such as left ventricular assist devices and cardiac transplantation are covered in detail in Chap. 271.

DISEASE MANAGEMENT AND SUPPORTIVE CARE

Despite stellar outcomes with medical therapy, admission rates following HF hospitalization remain high, with nearly half of all patients readmitted to hospital within 6 months of discharge. Recurrent HF and related cardiovascular conditions account for only half of readmissions in patients with HF, whereas other comorbidity-related conditions account for the rest. The key to achieving enhanced outcomes must begin with the attention to transitional care at the index hospitalization with facilitated discharge through comprehensive discharge planning, patient and caregiver education, appropriate use of visiting nurses, and planned follow-up. Early postdischarge follow-up, whether by telephone or clinic-based, may be critical to ensuring stability because most HF-related readmissions tend to occur within the first 2 weeks after discharge. The results of the recent STRONG-HF trial suggest that a strategy of intensive titration of medical therapy to guideline-recommended targets within 2 weeks of hospital discharge and frequent ambulatory follow-up through 2 months is associated with a substantial reduction in all-cause mortality and HF readmission at 180 days, underscoring the importance of timely application of targeted pharmacologic therapy and intensive clinical surveillance in the early postdischarge interval.

Although routinely advocated, intensive surveillance of weight and vital signs with use of telemonitoring has not decreased hospitalizations. Serial measurement of intrathoracic impedance has been utilized to identify early signals of worsening congestion to guide preemptive management to obviate the need for hospitalization. However, when systematically studied in randomized trials, this approach has not been proven to improve outcomes in comparison with routine HF care and may even enhance the rate of hospitalization due to the high frequency of impedance threshold crossings and device alerts. Implantable hemodynamic monitoring systems that directly measure pulmonary artery pressure tend to provide signals for early decompensation, and in patients with HF and moderately advanced symptoms across the full spectrum of EF, such systems have been shown to provide information that can allow implementation of therapy to avoid hospitalizations by as much as 39% (in the CardioMEMS Heart Sensor Allows Monitoring of Pressure to Improve Outcomes in NYHA Class III Heart Failure Patients [CHAMPION] trial). More recently, in the larger Hemodynamic Guided Management of Heart Failure (GUIDE-HF) Trial, hemodynamic-guided management with the same sensor did not result in a lower rate of composite mortality and total HF events in the primary analysis, the results of which were confounded by the impact of the COVID-19 pandemic on overall hospitalization rates. In a prespecified analysis of patients enrolled prior to the pandemic onset, there was a statistically significant reduction in the composite endpoint driven by lower rates of HF hospitalization in those assigned to the hemodynamic-guided therapy arm. Benefits were consistent across patients with prior HF hospitalization within 12 months and in those with elevated natriuretic peptide levels but no recent hospitalization. These data in aggregate suggest that hemodynamic-guided management is a useful adjunct to routine clinical care in selected high-risk patients with chronic HF across the spectrum of EF. Alternate approaches to longitudinal HF monitoring that leverage multiparameter signals derived from implantable cardiac rhythm devices such as pacemakers and defibrillators to provide a global index of congestion are also being explored as adjuncts to longitudinal HF management.

Once HF becomes advanced, regularly scheduled review of the disease course and options with the patient and family is recommended, including discussions surrounding end-of-life preferences when patients are comfortable in an outpatient setting. As the disease state advances further, integrating care with social workers, pharmacists, and community-based nursing may be critical in improving patient satisfaction with the therapy, enhancing QOL, and avoiding HF hospitalizations. Equally important is attention to seasonal influenza vaccinations and periodic pneumococcal vaccines that may obviate non-HF hospitalizations in these ill patients. When nearing end of life, facilitating a shift in priorities to outpatient and hospice palliation is key, as are discussions around advanced therapeutics and continued use of ICD prophylaxis, which may worsen QOL and prolong death. Small, randomized trials have suggested that systematic integration of palliative care considerations in high-risk HF patients by a specialized team has been demonstrated to improve QOL, anxiety, depression, and spiritual well-being and to facilitate goal-concordant care.

GLOBAL CONSIDERATIONS

Substantial differences exist in the practice of HF therapeutics and outcomes by geographic location. The penetrance of CRT and ICD is higher in the United States than in Europe. Conversely, therapy unavailable in the United States, such as levosimendan, is designated as useful in Europe, although the usefulness is disputable. Variation in the benefits of beta blockers based on world region remains an area of controversy. In oral pharmacologic therapy trials of HFrEF, patients from southwest Europe have a lower incidence of ischemic cardiomyopathy and those in North America tend to have more diabetes and prior coronary revascularization. There is also regional variation in medication use even after accounting for indication. In trials of HF, disparate effects are noted across populations. As a recent example, in TOPCAT, the drug spironolactone was effective when used in the U.S. population, whereas patients recruited from Russia and contiguous territories showed no difference. Whether this represents population differences or trial conduct disparity remains a serious question. ADHF patients in Eastern Europe tend to be younger, with higher EFs and lower natriuretic peptide levels. Patients from South America tend to have the lowest rates of comorbidities, revascularization, and device use. In contrast, patients from North America have the highest comorbidity burden with high revascularization and device use rates. Given geographic differences in baseline characteristics and clinical outcomes, the generalizability of therapeutic outcomes in patients in the United States and Western Europe may require verification.

■ FURTHER READING

Anker SD et al: Empagliflozin in heart failure with a preserved ejection fraction. N Engl J Med 385:1451, 2021.

Borlaug BA: The pathophysiology of heart failure with preserved ejection fraction. Nat Rev Cardiol 11:507, 2014.

Braunwald E: Heart failure. JACC Heart Fail 1:1, 2013.

Heidenreich PA et al: 2022 AHA/ACC/HFSA guideline for the management of heart failure: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 145:e895, 2022.

Hein AM et al: Medical management of heart failure with reduced ejection fraction in patients with advanced renal disease. JACC Heart Fail 7:371, 2019.

Hollenberg SM et al: 2019 ACC Expert Consensus Decision Pathway on Risk Assessment, Management, and Clinical Trajectory of Patients Hospitalized with Heart Failure: A report of the American College of Cardiology Solution Set Oversight Committee. J Am Coll Cardiol 74:1966, 2019.

Hussein AA, Wilkoff BL: Cardiac implantable electronic device therapy in heart failure. Circ Res 124:1584, 2019.

Kittleson MM et al: 2023 ACC Expert Consensus Decision Pathway on Management of Heart Failure With Preserved Ejection Fraction: A report of the American College of Cardiology Solution Set Oversight Committee. J Am Coll Cardiol 81:1835, 2023.

Kusumoto FM et al: HRS/ACC/AHA expert consensus statement on the use of implantable cardioverter-defibrillator therapy in patients who are not included or not well represented in clinical trials. Circulation 130:94, 2014.

McMurray JJ et al: Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med 371:993, 2014.

McMurray JJV et al: Dapagliflozin in patients with heart failure and reduced ejection fraction. N Engl J Med 381:1995, 2019.

Obadia JF et al: Percutaneous mitral valve repair or medical therapy for secondary mitral regurgitation. N Engl J Med 379:2297, 2018.

Packer M, Grayburn PA: Neurohormonal and transcatheter repair strategies for proportionate and disproportionate functional mitral regurgitation in heart failure. JACC Heart Fail 7:518, 2019.

Packer M et al: Cardiovascular and renal outcomes with empagliflozin in heart failure. N Engl J Med 383:1413, 2020.

Solomon SD et al: Angiotensin-neprilysin inhibition in heart failure with preserved ejection fraction. N Engl J Med 381:1609, 2019.

Solomon SD et al: Dapagliflozin in heart failure with mildly reduced or preserved ejection fraction. N Engl J Med 387:1089, 2022.

Stone GW et al: Transcatheter mitral valve repair in patients with heart failure. N Engl J Med 379:2307, 2018.

Teerlink JR et al: Cardiac myosin activation with omecamtiv mecarbil in systolic heart failure. N Engl J Med 384:105, 2021.

Velazquez EJ et al: Coronary-artery bypass surgery in patients with ischemic cardiomyopathy. N Engl J Med 374:1511, 2016.

266

Classification of Cardiomyopathy

Lynne Warner Stevenson, Neal K. Lakdawala, Joseph Loscalzo

The term *cardiomyopathy* describes primary disease of the heart muscle itself, originally excluding myocardial dysfunction resulting from other cardiovascular disease. Common usage, however, often includes diagnoses of ischemic cardiomyopathy, valvular cardiomyopathy, and hypertensive cardiomyopathy. The traditional morphologic classification defines the three major phenotypes of hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), and restrictive cardiomyopathy (RCM) **(Table 266-1)**.

Left ventricular wall thickness is increased (≥13–15 mm depending on context) and ejection fraction is normal or high with HCM. DCM is defined when the left ventricular ejection fraction (LVEF) is ≤0.50, but most clinical presentations are with LVEF ≤0.40. RCMs typically present with mildly decreased ejection fraction and variably increased wall thickness and are often defined less by morphology than by evidence of abnormal diastolic physiology on echocardiography or invasive hemodynamic measurement.

Although these phenotypes are helpful to guide initial evaluation of clinical disease, the phenotypes increasingly overlap over time as there is increasing recognition that DCMs can respond to recommended therapies with "reverse remodeling" to higher ejection fraction and smaller ventricles, while HCM evolves in ~5% of patients to a reduced ejection fraction (HCM-rEF) with more restrictive physiology. A fourth evolving phenotype with predominantly genetic causes is arrhythmogenic right ventricular cardiomyopathy (ARVC), originally termed *arrhythmogenic right ventricular dysplasia*, characterized by life-threatening arrhythmias and abnormal right ventricular structure and function, with variable expression in the left ventricle. As new phenotype-genotype connections are revealed, the terminology continues to evolve, as arrhythmogenic right ventricular dysplasia may also be termed ACM-RV (arrhythmogenic cardiomyopathy–right ventricle predominant), in line with ACM-LV, in which the arrhythmias and structural changes are predominantly in the left ventricle. If the ACM terminology is expanded to include acquired disease, the granulomatous disease of sarcoidosis and the protozoal myocarditis of Chagas' disease can both cause phenotypes with cardiomyopathy and ventricular arrhythmias that qualify as ACM-RV and ACM-LV.

Reliance upon phenotypic presentation is diminishing as more is learned about the underlying causes of cardiomyopathy. The catalog is rapidly growing of pathogenic genetic variants that can lead to heritable cardiomyopathies, which new imaging techniques can now sometimes identify prior to clinical disease. Expanding knowledge of immune response pathways reveals how some aspects of myocarditis may also be inherited and how they contribute to infectious and noninfectious inflammation that can cause clinical myocarditis and cardiomyopathy. Diagnosis and outcomes of clinical cardiomyopathies are further complicated by the frequent two-hit models where the clinical expression of a genetic predisposition to cardiomyopathy may

TABLE 266-1 Classification of Cardiomyopathies

Mid-range LVEF (0.40–0.50)

CARDIOMYOPATHY (CM) PHENOTYPE	DIAGNOSTIC CRITERIA	OTHER MORPHOLOGY	COMMON CHALLENGES IN DIAGNOSIS
Hypertrophic cardiomyopathy (HCM) *Mid-range LVEF includes rare transition to HCM with LV systolic dysfunction* (LVEF <0.50)	Septal thickness in men ≥15 mm, ≥13 mm in women. 13–14 mm may be diagnostic in relatives of proband with known HCM or with positive genetic test. LVEF ≥ normal, usually >0.60 LV chamber volume ≤normal.	Patterns of LV hypertrophy (LVH) in HCM: • Asymmetric septal hypertrophy • Inverse (sigmoid pattern) septal hypertrophy • Concentric LVH • Apical hypertrophy	Distinction from athlete's heart and severe chronic hypertension The storage diseases of *GLA* (Anderson-Fabry), *PRKAG2,* and *LAMP2* (Danon's), which can mimic HCM morphology Exclude aortic stenosis In older patients, exclude amyloidosis, which can also cause asymmetric septal thickening
Restrictive cardiomyopathy (RCM) Least common cardiomyopathy (CM) phenotype	Functional diagnosis based on moderate-severe diastolic dysfunction and/or elevated cardiac filling pressures. Wall thickness often increased but can appear normal. LVEF usually mildly reduced, occasionally normal.	Usually marked atrial enlargement Although both ventricles affected, clinical right heart failure often dominates. Marked wall thickness suggests: • Amyloidosis • Inherited storage diseases • Inborn metabolic diseases	Common etiologies of radiation, scleroderma-type connective tissue diseases, doxorubicin and other medications Some sarcomeric variants and storage diseases can appear with RCM as well as HCM phenotypes Although traditionally listed as restrictive, sarcoidosis more often has morphology of regional wall motion abnormalities, DCM phenotype, or right ventricular (RV) involvement with systolic dysfunction
Mid-range LVEF is often a transition in DCM, either deterioration from early DCM or improvement into DCM remission LV dilated cardiomyopathy (DCM)	• Early: LVEF ≤0.50 and/or LVEDV >112% normal for age/sex or LVEDD >95% predicted sex/height • LVEF ≤0.40: threshold for traditionally recommended therapies for heart failure with low LVEF • Persistent LVEF ≤0.30–0.35: threshold for primary prevention ICD	Can involve LV alone or with RV involvement either from primary cause or from secondary RV failure due to chronically elevated pulmonary artery pressures	Structural heart disease such as coronary artery disease or infarction from other cause, primary valve disease
Arrhythmogenic CM Dominant in LV (ACM-LV) Usually DCM, occasionally RCM	Morphologic criteria generally those for DCM. *Ventricular tachyarrhythmias dominate without or before severely reduced LVEF and heart failure.* Primary prevention ICD considered even when LVEF >0.35.	Suggestive but not necessarily conclusive differences in patterns of late gadolinium enhancement between different variants Occasional ACM-LV with RCM phenotype	For frequent PVCs or NSVT, may be difficult to distinguish PVC-related CM from genetic CM causing both the arrhythmias and the CM
Arrhythmogenic CM Dominant in RV (ACM-RV; also termed ARVC) *Biventricular arrhythmogenic CM*	Modified task force criteria from 2010 include combinations of major and minor criteria *Ventricular arrhythmias and evidence of both ACM-RV and ACM-LV*	Abnormal RV function or structure *Biventricular ACM often diagnosed from LGE in ventricle with less involvement*	Cardiac sarcoidosis can cause predominantly RV involvement with ventricular arrhythmias, RV wall motion abnormalities, aneurysms, and dilation
Trait of LV noncompaction (LVNC) Implications determined by CM phenotype and genotype	Often assessed by maximum ratio of noncompacted/compacted LV myocardium >2.3 and other criteria	Can occur with HCM, DCM, some dystrophies, and other syndromic presentations	Can occur in normal hearts, pregnancy Increased prevalence in athletic hearts

Abbreviations: ICD, implantable cardioverter-defibrillator; LGE, late gadolinium enhancement; LV, left ventricle; LVEDD, left ventricular end-diastolic dimension; LVEDV, left ventricular end-diastolic volume; LVEF, left ventricular ejection fraction; NSVT, nonsustained ventricular tachycardia; PVC, premature ventricular contraction.

be triggered by acquired conditions such as infections, toxic exposures, pregnancy, or tachycardia.

CLINICAL PRESENTATION AND EVALUATION OF CARDIOMYOPATHY

Early symptoms of cardiomyopathy often reflect exertional intolerance with breathlessness or fatigue. Arrhythmias are often presenting events of unrecognized cardiomyopathy, which can also present with embolic events related to atrial fibrillation or apical ventricular thrombi. Cardiomyopathy may often present first with the syndrome of congestion, with fluid retention and elevated left heart filling pressures causing shortness of breath with minimal activity or even at rest, particularly with orthopnea. It may be accompanied by elevated right-sided filling pressures causing edema and often abdominal symptoms. The nonspecific historical term *congestive heart failure* describes the syndrome of congestion, which is common to diverse cardiac diagnoses such as congenital heart disease, primary pulmonary hypertension, and structural valve disease. "Congestive heart failure" should not be considered a diagnosis or an etiology of cardiomyopathy but a nonspecific syndrome requiring thorough evaluation of possible etiology/ies. Initial evaluation of possible cardiomyopathy begins with a detailed clinical history and examination seeking clues to cardiac, genetic, and systemic causes of heart disease, which help to guide subsequent evaluation **(Table 266-2)**.

Echocardiography remains the initial imaging modality to define morphology and function, with increasing use of magnetic resonance imaging to provide further information on myocardial tissue characterization, patterns of fibrosis indicated by late gadolinium enhancement, and T1 and T2 mapping for evidence of focal and diffuse inflammation.

TABLE 266-2 Initial Evaluation of Cardiomyopathy

Clinical Evaluation
Thorough history and physical examination to identify cardiac and noncardiac disorders
Detailed family history of heart failure, cardiomyopathy, skeletal myopathy, conduction disorders, tachyarrhythmias, and sudden death
History of alcohol, illicit drugs, chemotherapy or radiation therapy
Assessment of changing ability to perform routine and desired activities
Assessment of jugular venous pressure, edema, orthostatic blood pressure, adequacy of perfusion
Laboratory Evaluation
Electrocardiogram
Chest radiograph
Two-dimensional and Doppler echocardiogram
Magnetic resonance imaging for evidence of myocardial inflammation and fibrosis
Chemistry:
Serum sodium, potassium, calcium, magnesium
Fasting glucose (glycohemoglobin in diabetes mellitus)
Creatinine, blood urea nitrogen
Albumin, total protein, liver function tests
Lipid profile
Thyroid-stimulating hormone
Serum iron, transferrin saturation
Urinalysis
Creatine kinase isoforms
Cardiac troponin level
Hematology:
Hemoglobin/hematocrit
White blood cell count with differential
Total eosinophil count if abnormal % on differential
Erythrocyte sedimentation rate
Evaluation When Specific Diagnoses Are Suspected
Respiratory pathogen panel during acute respiratory syndromes
Diagnosis of other specific infections such as:
Human immunodeficiency virus
Chagas' disease (*Trypanosoma cruzi*)
Lyme disease (*Borrelia burgdorferi*) and other tick-borne diseases
Toxoplasmosis
Trichinosis
Genetic counseling and testing with multigene cardiomyopathy panel
Serologies for active rheumatologic disease
Endomyocardial biopsy including sample for electron microscopy when suspecting specific diagnosis with therapeutic implications
Catheterization with coronary angiography in patients who have evidence of ischemia/infarction and are candidates for intervention

FURTHER READING

Arbelo E et al: 2023 ESC guidelines for the management of cardiomyopathies. Eur Heart J 44:3503, 2023.

Elliott P: Towards a new classification of cardiomyopathies. Curr Cardiol Rep 25:229, 2023.

Kontorovich AR: Approaches to genetic screening in cardiomyopathies: Practical guidance for clinicians. JACC Heart Fail 11:133, 2023.

Merlo M et al: Clinical application of CMR in cardiomyopathies: Evolving concepts and techniques: A position paper of myocardial and pericardial diseases and cardiac magnetic resonance working groups of Italian Society of Cardiology. Heart Fail Rev 28:77, 2023.

267 Genetic Cardiomyopathies

Neal K. Lakdawala, Lynne Warner Stevenson, Joseph Loscalzo

Each of the traditional morphologic forms of cardiomyopathy, hypertrophic, dilated, and restrictive, can be caused or modified by underlying genetic factors **(Table 267-1)**. Estimates for the prevalence of a genetic etiology for cardiomyopathy continue to rise, with increasing availability of genetic testing and attention to the family history. Well-recognized in hypertrophic cardiomyopathy, heritability is also present in at least 30% of dilated cardiomyopathy (DCM) without other clear etiology. Careful family history should elicit information about not only known cardiomyopathy and heart failure, but also family members who have had sudden death, often incorrectly attributed to a "heart attack," who have had atrial fibrillation or pacemaker implantation by middle age, or who have muscular dystrophy.

Most familial cardiomyopathies are inherited in an autosomal dominant pattern, with occasional autosomal recessive, matrilineal (mitochondrial), and X-linked inheritance (Table 267-1). Missense variants with amino acid substitutions and truncating variants are the most common genetic abnormalities in cardiomyopathy. Expressed mutant proteins may interfere with function of the normal allele through a dominant negative mechanism. Variants introducing a premature stop codon (nonsense) or shift in the reading frame (frameshift) may create a truncated or unstable protein, the lack of which causes cardiomyopathy (haploinsufficiency). In some forms of genetic DCM, both haploinsufficiency and the dominant negative effect of a truncated allele may contribute to pathophysiology. Deletions or duplications of an entire exon or gene are uncommon causes of cardiomyopathy, except for the dystrophinopathies.

Many different genes have been implicated in human cardiomyopathy (locus heterogeneity), and many pathogenic or likely pathogenic variants (i.e., variants) within those genes have been associated with disease (allelic heterogeneity). Although most identified variants are "private" to individual families, several specific variants are found repeatedly, either due to a founder effect or recurrent variants at a common residue. While most patients with genetic cardiomyopathy have a single rare, "high effect," disease allele, there is a growing appreciation for the effect of multiple less rare "intermediate effect" alleles on penetrance and expression. While the additive effects of multiple common alleles/variants captured as a polygenic risk score have shown greatest utility in common diseases such as atherosclerosis, they may also provide insight into the subset of patients with primary cardiomyopathy without an identifiable "high effect" allele.

Genetic cardiomyopathy is characterized by age-dependent and incomplete penetrance. The defining phenotype of cardiomyopathy is rarely present at birth and, in some individuals, may never manifest. Related individuals who carry the *same* variant may differ in the severity and rate of progression of cardiac dysfunction and associated rhythm disorders, indicating the important role of other genetic, epigenetic, and environmental modifiers in disease expression. Sex appears to play a role, as penetrance and clinical severity may be greater in men for most cardiomyopathies. The clinical course of a patient usually cannot be predicted based on which variant is present; thus, current therapy is based on the phenotype rather than the genetic defect. Currently, the greatest utility of genetic testing for cardiomyopathy is to inform family evaluations. However, genetic testing can provide risk stratification in some forms of cardiomyopathy and occasionally enables the detection of a disease for which specific therapy is indicated, such as the replacements for defective metabolic enzymes in Fabry's disease and Gaucher's disease. Moreover, clinical trials are interrogating the role of gene therapies for cardiomyopathy, which will provide greater impetus for genetic testing.

TABLE 267-1 Selected Genetic Defects Associated with Cardiomyopathy

	GENE PRODUCT	INHERITANCE	CARDIAC PHENOTYPE	ISOLATED CARDIAC PHENOTYPE[a]	EXTRACARDIAC MANIFESTATIONS
Sarcomere	*ACTC1* (cardiac actin)	AD	HCM, DCM	Yes	
	MYH7 (β myosin heavy chain)	AD	HCM, DCM, LVNC	Yes	Skeletal myopathy
	MYBPC3 (myosin binding protein C)	AD	HCM	Yes	
	TNNT2 (cardiac troponin T)	AD	HCM, DCM, LVNC	Yes	
	TNNI3 (cardiac troponin I)	AD, AR	HCM, DCM, RCM	Yes	
	TTN (Titin)	AD	DCM	Yes	
	TPM1 (α-tropomyosin)	AD	HCM, DCM	Yes	
	TNNC1 (cardiac troponin C)	AD	DCM	Yes	
	MYL2 (myosin regulatory light chain)	AD	HCM	Yes	Skeletal myopathy
	MYL3 (myosin essential light chain)	AD	HCM	Yes	
Z-disk and cytoskeleton	*DES* (desmin)	AD	RCM, DCM	Yes	Skeletal myopathy
	FLNC (filamin C)	AD	DCM	Yes	Skeletal myopathy
	NEXN (nexilin)	AD	DCM	Yes	
	VCL (vinculin)	AD	DCM	Yes	
Nuclear membrane	*LMNA* (lamin A/C)	AD, AR	CDDC	Yes	Skeletal myopathy
	EMD (emerin)	X-linked	CDDC	No	Skeletal myopathy, contractures
Excitation-contraction coupling	*PLN* (phospholamban)	AD	DCM, ARVC	Yes	
	SCN5A (NAV 1.5)	AD	CDDC	Yes	Note other variants associated with Brugada syndrome
	RYR2 (cardiac ryanodine receptor)	AD	ARVC	Yes	
	CASQ2 (calsequestrin 2)	AR	ARVC	Yes	
Cellular metabolism	*PRKAG2* (γ-subunit of AMP kinase)	AD	HCM+	Yes	
	LAMP2 (lysosomal associated membrane protein)	X-linked	HCM+	No[b]	Danon's disease: skeletal myopathy, cognitive impairment
	TAZ (tafazzin)	X-linked	DCM, LVNC	No	Barth's syndrome: skeletal myopathy, cognitive impairment, neutropenia
	FXN (frataxin)	AR	HCM	No	Friedreich's ataxia: ataxia, diabetes mellitus type 2
	TMEM43 (transmembrane protein 43)	AD	ARVC	Yes	
	GLA (α-galactosidase-A)	X-linked	HCM+	No	Fabry's disease: renal failure, angiokeratomas and painful neuropathy
Mitochondria	Mitochondrial DNA	Maternal transmission	DCM, HCM	No	MELAS, MERRF, Kearns-Sayre syndrome, ocular myopathy
Sarcolemmal membrane	*DMD* (dystrophin)	X-linked	DCM	No[b]	Duchenne's and Becker's muscular dystrophy
	DMPK (dystrophica myotonica protein kinase)	AD	DCM	No	Myotonic dystrophy type 1
Desmosome	*DSP* (desmoplakin), *JUP* (plakoglobin)	AD, AR	ARVC, DCM	Yes	Carvajal syndrome (AR), Naxos syndrome (AR), "woolly hair" and hyperkeratosis of palms and soles
	DSG2 (desmoglein 2), *DSC2* (desmocollin 2), *PKP2* (plakophilin 2)	AD	ARVC	Yes	
Other examples	*RBM20* (RNA binding motif 20)	AD	DCM	Yes	
	BAG3 (BCL2-associated athanogene 3)	AD	DCM	Yes	
	ALPK3 (α-kinase 3)	AR	HCM	Yes	

[a] Indicates that the usual clinical presentation is of isolated cardiomyopathy; however, occasionally present extracardiac manifestations are also provided. [b] Indicates that isolated cardiac phenotype can occur in women with the X-linked defects.

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; ARVC, arrhythmogenic right ventricular cardiomyopathy; CDDC, conduction disease with dilated cardiomyopathy; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; HCM+, HCM with preexcitation; LVNC, left ventricular noncompaction; MELAS, (mitochondrial) myopathy, encephalopathy, lactic acidosis, and strokelike episodes syndrome; MERRF, myoclonic epilepsy with ragged red fibers; RCM, restrictive cardiomyopathy.

The genetic architecture of hypertrophic, dilated, arrhythmogenic, and restrictive cardiomyopathy are overlapping, summarized in Table 267-1 and Fig. 267-1 and detailed in the subsequent sections organized by phenotype.

For any patient with suspected or proven genetic disease, family members should be considered and evaluated in a longitudinal fashion. Screening generally includes both an echocardiogram and electrocardiogram (ECG). The indications and implications for confirmatory specific genetic testing vary depending on the specific variant. The profound questions raised by families about diseases shared and passed down merit serious and sensitive discussion, ideally provided by a trained genetic counselor.

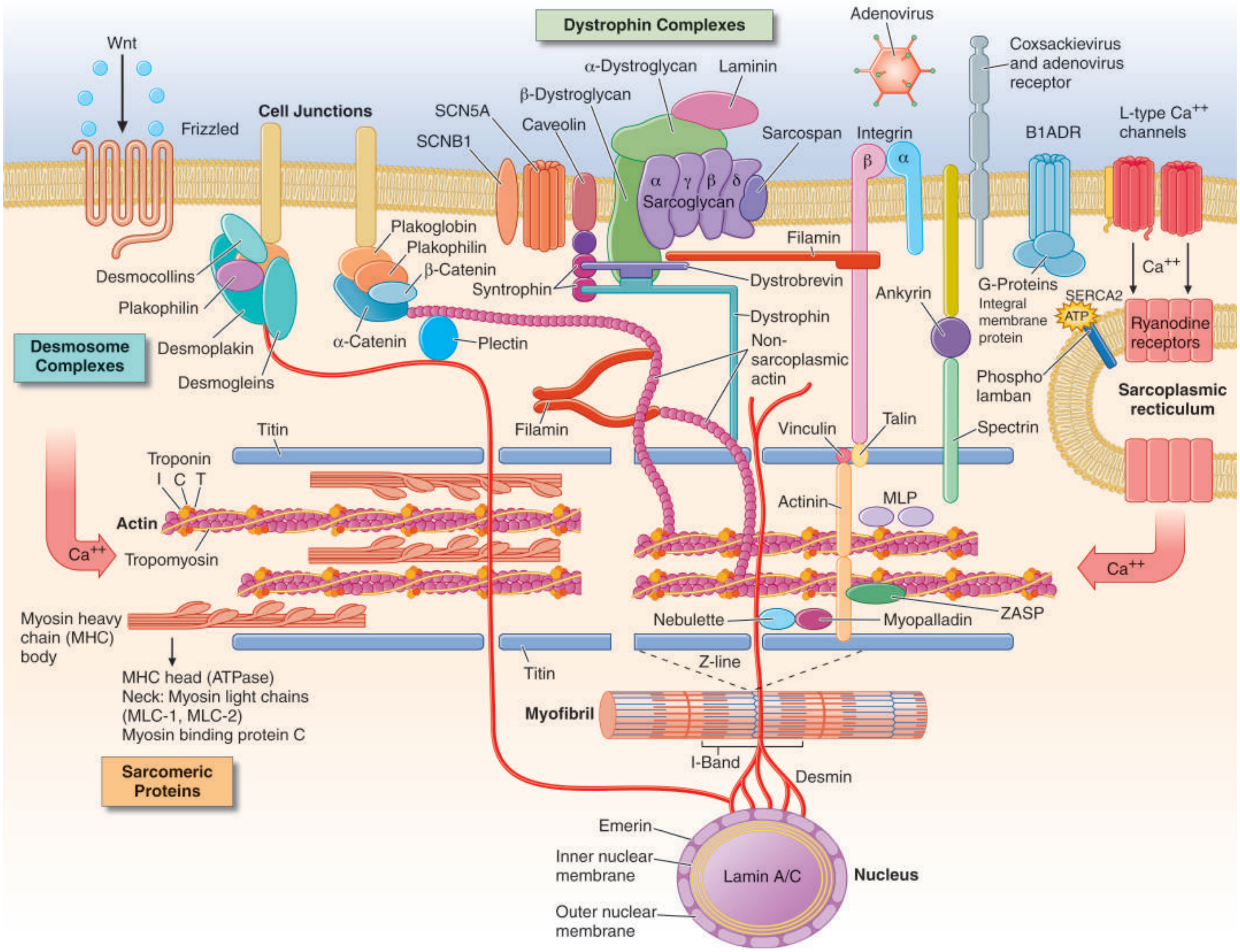
Dystrophin Complexes
Adenovirus
Wnt
α-Dystroglycan
Laminin
Coxsackievirus and adenovirus receptor
β-Dystroglycan
Frizzled
Cell Junctions
SCN5A
Caveolin
SCNB1
Sarcospan
Integrin
β
α
B1ADR
L-type Ca++ channels
α γ β δ
Sarcoglycan
Plakoglobin
Plakophilin
Filamin
β-Catenin
Desmocollins
Syntrophin
Dystrobrevin
Ankyrin
G-Proteins
Ca++
Integral membrane protein
SERCA2
ATP
Plakophilin
Dystrophin
Ryanodine receptors
Desmosome Complexes
Desmoplakin
α-Catenin
Plectin
Non-sarcoplasmic actin
Desmogleins
Phospholamban
Sarcoplasmic recticulum
Talin
Titin
Filamin
Vinculin
Spectrin
Troponin
I C T
Actin
Actinin
MLP
Ca++
Tropomyosin
Ca++
ZASP
Myosin heavy chain (MHC) body
Nebulette
Myopalladin
Z-line
MHC head (ATPase)
Neck: Myosin light chains (MLC-1, MLC-2)
Myosin binding protein C
Titin
Myofibril
I-Band
Desmin
Sarcomeric Proteins
Emerin
Inner nuclear membrane
Lamin A/C
Nucleus
Outer nuclear membrane

FIGURE 267-1 Drawing of myocyte indicating multiple sites of abnormal gene products associated with cardiomyopathy. Major functional groups include the sarcomeric proteins (actin, myosin, tropomyosin, and the associated regulatory proteins), the dystrophin complex stabilizing and connecting the cell membrane to intracellular structures, the desmosome complexes associated with cell-cell connections and stability, and multiple cytoskeletal proteins that integrate and stabilize the myocyte. ATP, adenosine triphosphate. *(Figure adapted from Jeffrey A. Towbin, MD, University of Tennessee Health Science Center.)*

HYPERTROPHIC CARDIOMYOPATHY

EPIDEMIOLOGY, ETIOLOGY, AND PATHOLOGY

Hypertrophic cardiomyopathy is defined as left ventricular hypertrophy that develops in the absence of causative hemodynamic factors, such as hypertension, aortic valve disease, or systemic infiltrative or storage diseases **(Figs. 267-2 and 267-3)**. It has previously been termed *hypertrophic obstructive cardiomyopathy* (HOCM); however, the accepted terminology is now hypertrophic cardiomyopathy with or without obstruction. Prevalence in North America, Africa, and Asia is about 1:500. It is a leading cause of sudden death in the young and is an important cause of heart failure. Pediatric presentation is associated with increased early morbidity and mortality, and patients diagnosed as adults have decreased survival compared to age-matched individuals without hypertrophic cardiomyopathy.

A sarcomere gene variant is present in ~40–50% of patients with hypertrophic cardiomyopathy and is more common in those with familial disease and characteristic asymmetric septal hypertrophy. More than nine different genes with >1500 variants have been implicated, although ~80% of patients have a variant in either *MYH7* or *MYBPC3* (Table 267-1), which encode the thick myofilament of the sarcomere.

Hypertrophic cardiomyopathy is characterized by age-dependent and incomplete penetrance. The defining phenotype of left ventricular hypertrophy is rarely present at birth and usually develops later in life. Women appear to have lower penetrance of sarcomere variants and an older age at hypertrophic cardiomyopathy diagnosis but subsequently increased rates of heart failure and mortality thereafter. Accordingly, screening of family members should begin in childhood, usually at adolescence, and extend through adulthood. In *MYBPC3* variant carriers, the average age of disease development is ~40 years, while 30% remain free from hypertrophy after 70 years. Related individuals who carry the *same* variant may have a different extent and pattern of hypertrophy (e.g., asymmetric vs concentric), occurrence of outflow tract obstruction, and associated clinical outcomes, although sudden death and progression to heart failure occur more commonly in families with that history.

At the level of the sarcomere, hypertrophic cardiomyopathy variants lead to enhanced calcium sensitivity, maximal force generation, and ATPase activity. Calcium handling is affected through modification of regulatory proteins. Sarcomere variants lead to abnormal energetics and impaired relaxation, both directly and as a result of hypertrophy. Hypertrophic cardiomyopathy is characterized by misalignment and disarray of the enlarged myofibrils and myocytes **(Fig. 267-4)**, which can also occur to a lesser extent in other cardiac diseases. Although hypertrophy is the defining feature of hypertrophic cardiomyopathy, fibrosis and microvascular disease are also present. Interstitial fibrosis is detectable before overt hypertrophy develops and likely results from early activation of profibrotic pathways. In the majority of patients with overt cardiomyopathy, focal areas of replacement fibrosis can be readily

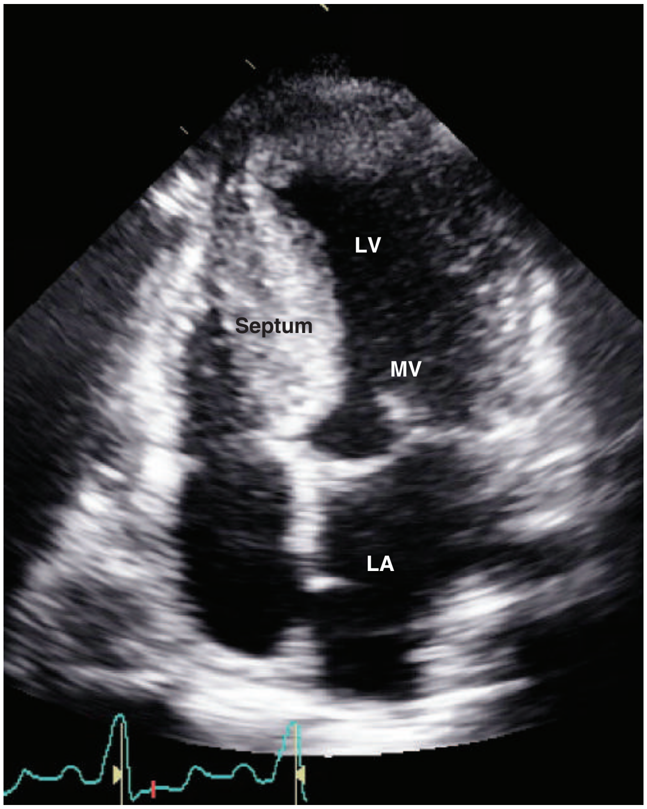

FIGURE 267-2 Hypertrophic cardiomyopathy. This echocardiogram of hypertrophic cardiomyopathy shows asymmetric hypertrophy of the septum compared to the lateral wall of the left ventricle (LV). The mitral valve (MV) is moving anteriorly toward the hypertrophied septum in systole. The left atrium (LA) is enlarged. Note that the echocardiographic and pathologic images are vertically opposite, such that the LV is by convention on the top right in the echocardiographic image and bottom right in the pathologic images. *(Image courtesy of Justina Wu, MD, Brigham and Women's Hospital, Boston.)*

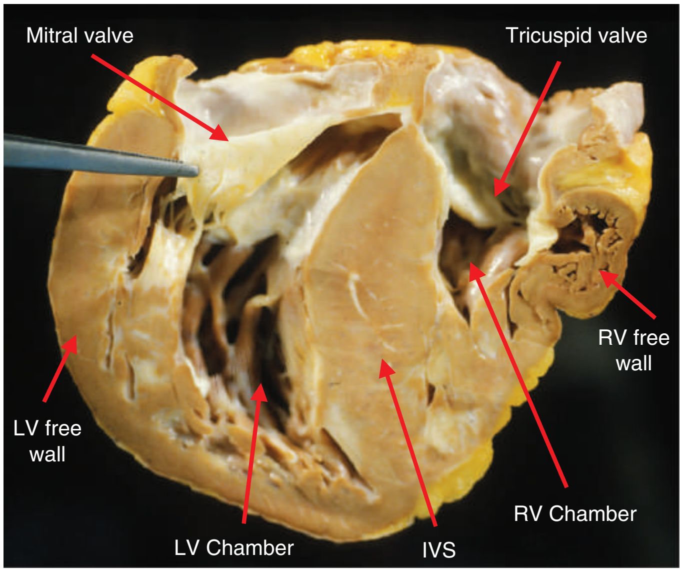

FIGURE 267-3 Hypertrophic cardiomyopathy. Gross specimen of a heart with hypertrophic cardiomyopathy removed at the time of transplantation, showing asymmetric septal hypertrophy (septum much thicker than left ventricular free wall) with the septum bulging into the left ventricular outflow tract causing obstruction. The forceps are retracting the anterior leaflet of the mitral valve, demonstrating the characteristic plaque of systolic anterior motion, manifest as endocardial fibrosis on the interventricular septum in a mirror-image pattern to the valve leaflet. There is patchy replacement fibrosis, and small thick-walled arterioles can be appreciated grossly, especially in the interventricular septum. IVS, interventricular septum; LV, left ventricle; RV, right ventricle. *(Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.)*

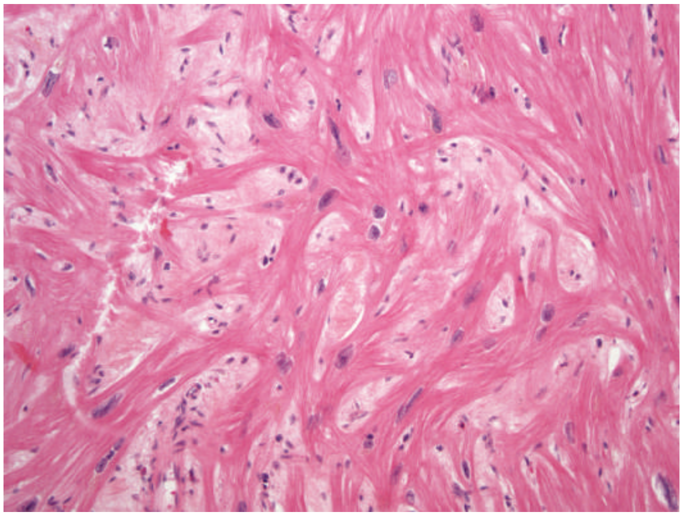

FIGURE 267-4 Hypertrophic cardiomyopathy. Microscopic image of hypertrophic cardiomyopathy showing the characteristic disarrayed myocyte architecture with swirling and branching rather than the usual parallel arrangement of myocyte fibers. Myocyte nuclei vary markedly in size, and interstitial fibrosis is present. *(Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.)*

detected with magnetic resonance imaging (MRI). These areas of "scar" may represent substrate for the development of ventricular arrhythmias. Increased thickness and decreased luminal area of the intramural vessels in hypertrophied myocardium contribute to microvascular ischemia and angina. Microinfarction of hypertrophied myocardium is a hypothesized mechanism for replacement scar formation.

Macroscopically, hypertrophy is typically manifest as nonuniform ventricular thickening (Fig. 267-3). The interventricular septum is the typical location of maximal hypertrophy, although other patterns of hypertrophic remodeling include concentric and midventricular. Hypertrophy confined to the ventricular apex (apical hypertrophic cardiomyopathy) is less often familial and has a different genetic substrate, with sarcomere variants present in only ~15%. Left ventricular outflow tract obstruction represents the most common focus of diagnosis and intervention, although diastolic dysfunction, myocardial fibrosis, and microvascular ischemia also contribute to contractile dysfunction and elevated intracardiac pressures. Obstruction is present in ~30% of patients at rest and can be provoked by exercise in another ~30%. Systolic obstruction is initiated by drag forces, which push an anteriorly displaced and enlarged anterior mitral leaflet into contact with the hypertrophied ventricular septum. Mitral leaflet coaptation may ensue, leading to posteriorly directed mitral regurgitation. To maintain stroke volume across outflow tract obstruction, the ventricle generates higher pressures, leading to higher wall stress and myocardial oxygen demand. Smaller chamber size and increased contractility exacerbate the severity of obstruction. Conditions of low preload, such as dehydration, and low afterload, such as arterial vasodilation, may lead to transient hypotension and near-syncope. The systolic ejection murmur of left ventricular outflow tract obstruction is harsh and late peaking and can be enhanced by bedside maneuvers that diminish ventricular volume and transiently worsen obstruction, such as the Valsalva maneuver or standing from a squatting position.

DIAGNOSIS

The substantial variability of hypertrophic cardiomyopathy pathology is reflected in the diversity of clinical presentations. Patients may be diagnosed after undergoing evaluations triggered by the abnormal physical findings (murmur) or by symptoms of exertional dyspnea, angina, or syncope. Alternatively, diagnosis may follow evaluations prompted by the detection of disease in family members. Cardiac imaging (Fig. 267-2) is central to diagnosis, for which the physical examination and ECG are insensitive. The identification of a disease-causing variant in a proband can focus family evaluations on variant carriers, but this strategy requires a high degree of certainty that the variant is truly pathogenic and not a benign DNA variant. Biopsy is not recommended to diagnose hypertrophic cardiomyopathy but can be used to exclude infiltrative and metabolic diseases. Rigorous athletic training (athlete's heart) may cause intermediate degrees of physiologic hypertrophy difficult to differentiate from mild hypertrophic cardiomyopathy. Unlike hypertrophic cardiomyopathy, hypertrophy in the athlete's heart regresses with cessation of training and is accompanied by supernormal exercise capacity (VO_{2max} >50 mL/kg per min), mild ventricular dilation, and normal diastolic function.

TREATMENT

Hypertrophic Cardiomyopathy

Management focuses on treatment of symptoms and prevention of sudden death and stroke **(Fig. 267-5)**. Left ventricular outflow tract obstruction can be controlled medically in the majority of patients. β-Adrenergic blocking agents and L-type calcium channel blockers (e.g., verapamil) are first-line agents that reduce the severity of obstruction by slowing heart rate, enhancing diastolic filling, and decreasing contractility. Persistent symptoms of exertional dyspnea or chest pain can be controlled occasionally with the addition of disopyramide, an antiarrhythmic agent with potent negative inotropic properties. The recently introduced small-molecule cardiac myosin inhibitors mavacamten (U.S. Food and Drug Administration approved) and aficamten (under investigation) have shown high efficacy in the treatment of symptomatic obstructive hypertrophic cardiomyopathy, including in patients with persistent symptoms despite treatment with a beta blocker and/or those considering septal reduction therapy. The principal risk of cardiac myosin inhibitors is a transient reduction in left ventricular ejection fraction (LVEF), which warrants frequent echocardiographic monitoring.

Patients with or without obstruction may develop heart failure symptoms due to fluid retention and require diuretic therapies for venous congestion. Severe medically refractory symptoms develop in ~5% of patients, for whom septal reduction therapy with surgical myectomy or alcohol septal ablation may be effective. Developed over 60 years ago, surgical myectomy effectively relieves outflow tract obstruction by excising part of the septal myocardium involved in the dynamic obstruction. In selected patients, perioperative mortality is extremely low with excellent long-term survival free from recurrent obstruction and symptoms. Mitral valve repair or replacement is usually unnecessary as associated eccentric mitral regurgitation resolves with myectomy alone. Alcohol septal ablation in patients with suitable coronary anatomy can relieve outflow tract obstruction via a controlled infarction of the proximal septum, which produces similar periprocedural outcomes and gradient reduction as surgical myectomy. Although head-to-head comparisons between the two septal reduction therapies do not exist, septal ablation is relegated primarily to patients who wish to avoid surgery or who have limiting comorbidities. Neither procedure has been shown to improve outcomes other than symptoms. With both procedures, the most common complication is the development of complete heart block necessitating permanent pacing. However, ventricular pacing as a primary therapy for outflow tract obstruction is ineffective and not generally advised.

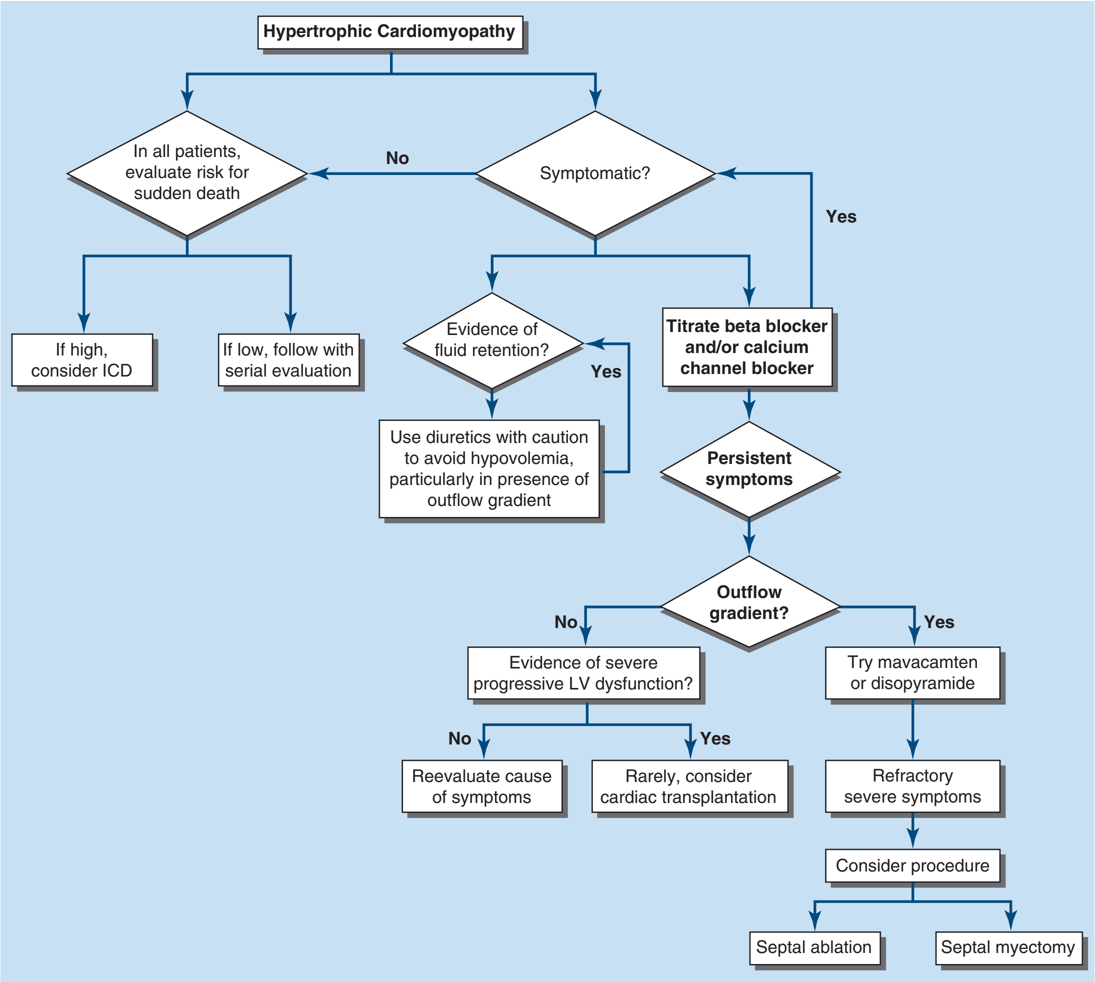
Hypertrophic Cardiomyopathy
In all patients, evaluate risk for sudden death
No
Symptomatic?
Yes
If high, consider ICD
If low, follow with serial evaluation
Evidence of fluid retention?
Yes
Titrate beta blocker and/or calcium channel blocker
Use diuretics with caution to avoid hypovolemia, particularly in presence of outflow gradient
Persistent symptoms
Outflow gradient?
No
Yes
Evidence of severe progressive LV dysfunction?
Try mavacamten or disopyramide
No
Yes
Reevaluate cause of symptoms
Rarely, consider cardiac transplantation
Refractory severe symptoms
Consider procedure
Septal ablation
Septal myectomy

FIGURE 267-5 Treatment algorithm for hypertrophic cardiomyopathy depending on the presence and severity of symptoms and the presence of an intraventricular gradient with obstruction to outflow. Note that all patients with hypertrophic cardiomyopathy should be evaluated for atrial fibrillation and risk of sudden death, whether or not they require treatment for symptoms. ICD, implantable cardioverter-defibrillator; LV, left ventricular.

Patients with hypertrophic cardiomyopathy have an increased risk of sudden cardiac death from ventricular tachyarrhythmias. Vigorous physical activity and competitive sports have been historically prohibited; however, recent studies have failed to identify a relationship between exertion and ventricular arrhythmias in hypertrophic cardiomyopathy, empowering patients and providers to make shared decisions about exercise. Factors that increase the risk of sudden death from a baseline of 0.5% per year are presented in **Table 267-2**. As sudden death has not been reduced by medical or procedural interventions, traditionally an implantable cardioverter-defibrillator has been advised for patients with one or more major risk factors and advised on a selected basis for patient with more than one modifying risk factor. Nevertheless, the positive predictive value of most risk factors is low, and many patients receiving a defibrillator never receive an appropriate device therapy. A complementary approach to sudden death risk stratification and discussion with patients is the application of an externally validated European Society of Cardiology risk score using major criteria from Table 267-2 and continuous variables such as outflow tract gradient and left atrial size. Shared decision-making around implantable cardioverter-defibrillator implantation for primary prevention has emphasized discussions of estimated risk levels rather than dichotomous yes–no criteria. Long-term use of a defibrillator may be associated with serious device-related complications, particularly in young active patients. Refinement of sudden death risk through the application of contemporary technologies such as cardiac MRI is ongoing.

Atrial fibrillation is common in patients with hypertrophic cardiomyopathy and may lead to hemodynamic deterioration and embolic stroke. Rapid ventricular response is poorly tolerated and may worsen outflow tract obstruction. β-Adrenergic blocking agents and L-type calcium channel blockers slow atrioventricular (AV) nodal conduction and improve symptoms; cardiac glycosides should be avoided, as they may increase contractility and worsen obstruction. Even with adequate rate control, symptoms exacerbated by atrial fibrillation may persist due to loss of AV synchrony and may require restoration of sinus rhythm. Disopyramide and amiodarone are the preferred antiarrhythmic agents, with radiofrequency ablation considered for medically refractory cases. Anticoagulation to prevent embolic stroke in atrial fibrillation is recommended.

TABLE 267-2 Risk Stratification for Sudden Death in Hypertrophic Cardiomyopathy

MAJOR RISK FACTOR		SCREENING TECHNIQUE
History of cardiac arrest or spontaneous sustained ventricular tachycardia[a]		History
Syncope	Nonvagal, often with or after exertion	History
Family history of sudden cardiac death		Family history
Left ventricular apical aneurysm	Generally applicable to patients with apical hypertrophy	Echocardiography with contrast, cardiac magnetic resonance imaging
LV thickness >30 mm	Present in <10% of patients	Echocardiography or cardiac magnetic resonance imaging
LV systolic dysfunction (ejection fraction <50%)	Present in <10% of patients	Echocardiography or cardiac magnetic resonance imaging
Variables Utilized in the European Society of Calculator for Estimated Risk of Sudden Death		
LV outflow tract gradient	Peak gradient measured at rest or with the Valsalva maneuver, mmHg	Echocardiography
Left atrial diameter	Diameter measured in the parasternal long axis, mm	Echocardiography
LV thickness	Maximal wall thickness, mm	Echocardiography
Age		
Syncope, family history, nonsustained ventricular tachycardia	As above	As above
Modifying Risk Factors		
Late gadolinium enhancement	As a percentage of myocardial mass	Cardiac magnetic resonance imaging
Spontaneous nonsustained ventricular tachycardia	>3 beats at rate >120	Exercise or 24-h to 48-h ambulatory recording

[a] Implantable cardioverter-defibrillator advised for patients with prior arrest or sustained ventricular tachycardia regardless of other risk factors if life expectancy is estimated to be >1 year. The European Society of Cardiology risk calculator can be found at *https://doc2do.com/hcm/webHCM.html* and provides an estimated 5-year risk of cardiac arrest. Patients with estimated risk of ≥6% are generally advised placement of an implantable cardioverter-defibrillator; those with risk between 4 and 6% can be considered for implant, and implant is not advised when risk is <4%. Emerging risk factors merit further clinical validation.

Abbreviation: LV, left ventricle.

PROGNOSIS

The general prognosis for hypertrophic cardiomyopathy is better than in early studies of referral populations, but mortality remains higher than in an age-matched population without cardiomyopathy. The sudden death risk is <1% per year; however, up to 1 in 20 patients will progress to overt systolic dysfunction with a reduced ejection fraction (<50%) with or without dilated remodeling (i.e., "burned out" or end-stage hypertrophic cardiomyopathy). These patients may suffer from low cardiac output and have an increased risk of death from progressive heart failure and sudden death unless they undergo timely cardiac transplantation.

GENETIC DILATED AND ARRHYTHMOGENIC CARDIOMYOPATHY

EPIDEMIOLOGY, ETIOLOGY, AND PATHOLOGY

As outlined in **Chap. 266**, an enlarged left ventricle with reduced systolic function as measured by LVEF characterizes DCM, which is considered to be present when the LVEF is ≤0.50 and/or the left ventricular diastolic dimension is >95% predicted for age and sex. In some patients with reduced LVEF, the left ventricular dilation is minimal, sometimes referred to as nondilated or minimally dilated cardiomyopathy **(Figs. 267-6, 267-7, and 267-8)**.

While diverse etiologies may cause DCM and arrhythmogenic cardiomyopathy (ACM), familial clustering is present in ~30–40% of cases and monogenic etiologies can be identified in ~25%.

Sarcomere variants are most associated with hypertrophic cardiomyopathy; however, they are also implicated in DCM. The most common genetic causes of DCM are truncating variants of the giant protein titin, encoded by *TTN*, which maintains sarcomere structure and acts as a key signaling molecule.

As cytoskeletal proteins play crucial roles in the structure, connection, and stability of the myocyte, multiple defects in these proteins can lead to dilated cardiomyopathy (Fig. 267-1). For example, desmin forms intermediate filaments that connect the nuclear and plasma membranes, Z-lines, and the intercalated disks between muscle cells. Desmin variants impair the transmission of force and signaling for both cardiac and skeletal muscle and may cause combined cardiac and skeletal myopathy, more commonly with a restrictive phenotype (RCM) than dilated phenotype (DCM).

Defects in the sarcolemmal membrane proteins are associated with DCM. The best known is dystrophin, encoded by the X chromosome gene *DMD*, abnormalities of which cause Duchenne's and Becker's muscle dystrophy. This protein provides a network that supports the sarcolemma and also connects to the sarcomere. The progressive functional defect in both cardiac and skeletal muscle reflects vulnerability

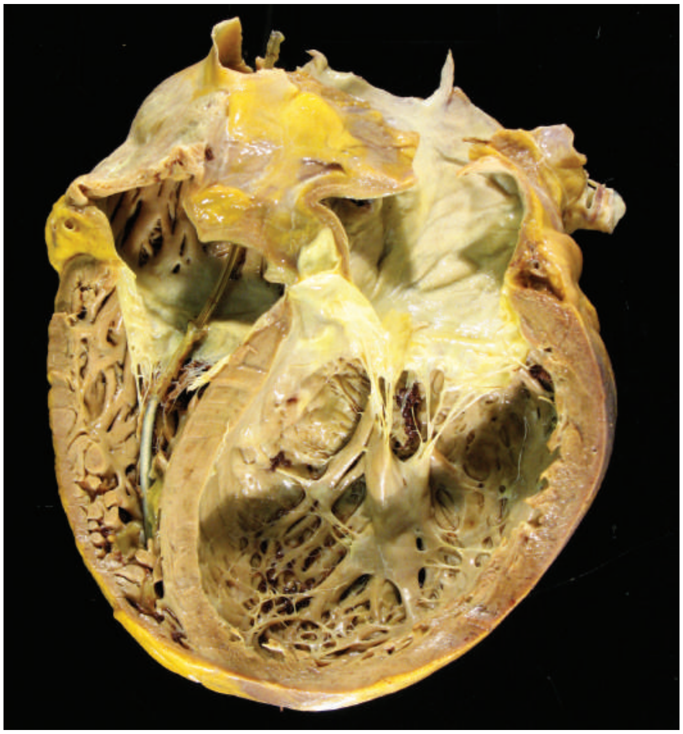

FIGURE 267-6 Dilated cardiomyopathy. This gross specimen of a heart removed at the time of transplantation shows massive left ventricular dilation and moderate right ventricular dilation. Although the left ventricular wall in particular appears thinned, there is significant hypertrophy of this heart, which weighs >800 g (upper limit of normal = 360 g). A defibrillator lead is seen traversing the tricuspid valve into the right ventricular apex. *(Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.)*

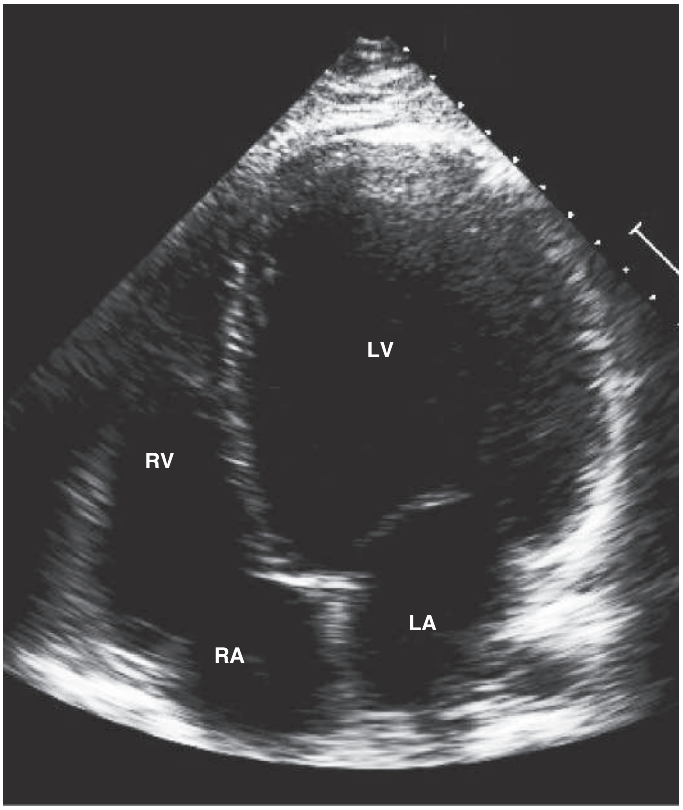

FIGURE 267-7 Dilated cardiomyopathy. This echocardiogram of a young man with dilated cardiomyopathy shows massive global dilation and thinning of the walls of the left ventricle (LV). The left atrium (LA) is also enlarged compared to normal. Note that the echocardiographic and pathologic images are vertically opposite, such that the LV is by convention on the top right in the echocardiographic image and bottom right in the pathologic images. RA, right atrium; RV, right ventricle. *(Image courtesy of Justina Wu, MD, Brigham and Women's Hospital, Boston.)*

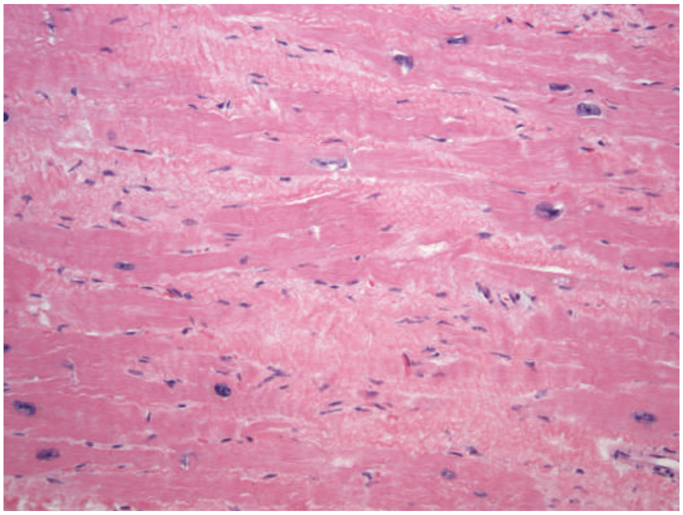

FIGURE 267-8 Dilated cardiomyopathy. Microscopic specimen of a dilated cardiomyopathy showing the nonspecific changes of interstitial fibrosis and myocyte hypertrophy characterized by increased myocyte size and enlarged, irregular nuclei. Hematoxylin and eosin–stained section, 100× original magnification. *(Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.)*

to mechanical stress. Defects in the sarcolemmal channel proteins (*channelopathies*) are generally associated with primary arrhythmias, but variants in *SCN5A*, the α subunit of the Nav 1.5 ion channel protein, distinct from those that cause the Brugada or long QT syndromes, have been implicated in DCM with conduction disease.

Nuclear membrane protein defects in cardiac and skeletal muscle occur in either autosomal (lamin A/C) or X-linked (emerin) patterns. These defects are associated with a high prevalence of atrial and ventricular arrhythmias and conduction system disease, which can occur in some family members without or before detectable cardiomyopathy and underlie gene-specific risk stratification for sudden death and different utilization criteria for primary prevention implantable cardioverter-defibrillators in these patients.

Intercalated disks contribute to intracellular connections, allowing mechanical and electrical coupling between cells and also connections to desmin filaments within the cell. Variants in proteins of the desmosomal complex compromise attachment of the myocytes, which can become disconnected and die via activation of Wnt/β-catenin and proinflammatory signaling pathways, to be replaced by fat and fibrous tissue. These areas are highly arrhythmogenic and may dilate to form aneurysms. Although more often noted in the right ventricle (arrhythmogenic right ventricular cardiomyopathy), this condition can be restricted to the left ventricle (especially when secondary to truncating variants in *DSP*, which encodes desmoplakin) or affect both ventricles and has also been termed "arrhythmogenic cardiomyopathy."

The recognized frequency of familial involvement in DCM has increased to >30%. Truncating variants in *TTN*, encoding the giant sarcomeric protein titin, are the most common cause of DCM, accounting for up to 25% of familial disease. On average, men with *TTN* truncating variants develop cardiomyopathy a decade before women, without distinctive clinical features. Variants in thick and thin filament genes account for ~8% of DCM and may manifest in early childhood.

■ PROGNOSIS

Prognosis and therapy of DCM and ACM are dictated primarily by the stage of clinical disease and the risk for sudden death, which also varies based on disease gene (higher for patients with variants in *DES*, *DSP*, *DSC2*, *DCG2*, *FLNC*, *LMNA*, *PKP2*, *PLN*, *RBM20*, *SCN5A*, and *TMEM43*). The rate of progression of disease is also heritable, with marked variation based on disease gene observed (e.g., patients with *TTN* truncating variants frequently experience recovery with medical therapy). Medical therapy is generally guided by the phenotype, such that patients with DCM and reduced LVEF are treated with therapies recommended for heart failure with reduced ejection fraction

(Chap. 264). As for HCM, shared decision-making is needed regarding implantable cardioverter-defibrillators. However, the specific genotype plays a greater role in decisions regarding DCM, for which some features may warrant implantable cardioverter-defibrillator placement before the LVEF has declined to 0.35, such as the presence of pathogenic *LMNA* or *FLNC* variants or a family history of sudden death at early ages.

Arrhythmogenic right ventricular cardiomyopathy (ARVC) has an established diagnostic framework (task force criteria) that rests upon identifying and quantifying right ventricular dilation, dyskinesis, and ECG abnormalities (repolarization, depolarization, and arrhythmias) in the context of a suggestive family history of genetic test results. Overlapping with ARVC and DCM is ACM, in which ventricular arrhythmias may precede or supersede the severity of predominantly left ventricular remodeling. Unlike ARVC, consensus diagnostic criteria for ACM are lacking. Early stages of ARVC and ACM may be restricted to ventricular arrhythmias, and over years, ventricular dilation, hypokinesis, and failure may ensue. Patients with an initial presentation of right ventricular cardiomyopathy who progress to include left ventricular dysfunction are at high risk for adverse events.

CARDIOMYOPATHY DUE TO INHERITED DISORDERS OF METABOLISM

Multiple genetic disorders of metabolic pathways can cause myocardial disease, due to infiltration of abnormal products or cells containing them between the myocytes, and storage disease, due to their accumulation within cells (Table 267-1). Hypertrophic cardiomyopathy may be mimicked by the myocardium thickened with these abnormal products causing "pseudohypertrophy," usually with an abnormally short PR interval. The pseudohypertrophic phenotype is most common, but restrictive cardiomyopathy and DCM may occur. Most of these diseases are diagnosed during childhood.

Fabry's disease results from a deficiency of the lysosomal enzyme alpha-galactosidase A caused by variants in *GLA*. This disorder of glycosphingolipid metabolism is an X-linked disorder that may also cause clinical disease in female carriers. Glycolipid accumulation may be limited to the cardiac tissues but usually also involves the skin, peripheral nerve, and kidney. Electron microscopy of endomyocardial biopsy tissue shows diagnostic vesicles containing concentric lamellar figures **(Fig. 267-9)**. Diagnosis can be made through assessment of enzyme activity and/or *GLA* sequencing and is crucial because enzyme replacement can reduce abnormal deposits and improve cardiac and clinical function. The magnitude of clinical impact has not been well established for this therapy, which requires frequent infusions of the enzyme at a cost of >$100,000 a year. The oral chaperone therapy, migalastat, stabilizes mutant forms of alpha-galactosidase, increases enzymatic activity, and was approved for use in a subset of patients with Fabry's disease bearing variants amenable to this therapy.

Carnitine is an essential cofactor in long-chain fatty acid metabolism. Multiple defects have been described that lead to carnitine deficiency, causing intracellular lipid inclusions and restrictive cardiomyopathy or DCM, often presenting in children. Fatty acid oxidation requires many metabolic steps with specific enzymes that can be deficient, with complex interactions with carnitine. Depending on the defect, cardiac and skeletal myopathy can be ameliorated with replacement of fatty acid intermediates and carnitine.

Two monogenic metabolic cardiomyopathies cause increased ventricular wall thickness without an increase of muscle subunits or an increase in contractility. Variants in the gamma-2 regulatory subunit of the adenosine monophosphate (AMP)-activated protein kinase important for glucose metabolism (*PRKAG2*) have been associated with a high prevalence of conduction abnormalities, such as AV block and ventricular preexcitation. Several defects have been reported in an X-linked lysosome-associated membrane protein (*LAMP2*). This defect can be maternally transmitted or sporadic and has occasionally been isolated to the heart, although it often leads to a syndrome of skeletal myopathy, intellectual disability, and hepatic dysfunction referred to as *Danon disease.* Extreme left ventricular hypertrophy appears early, often in childhood, and can progress rapidly to end-stage heart failure with low ejection fraction. Electron microscopy of these metabolic disorders shows that the myocytes are enlarged by multiple intracellular vacuoles of metabolic by-products. Gene therapy using a viral vector to deliver functional *LAMP2* to cardiomyocytes is currently under study in humans with Danon disease.

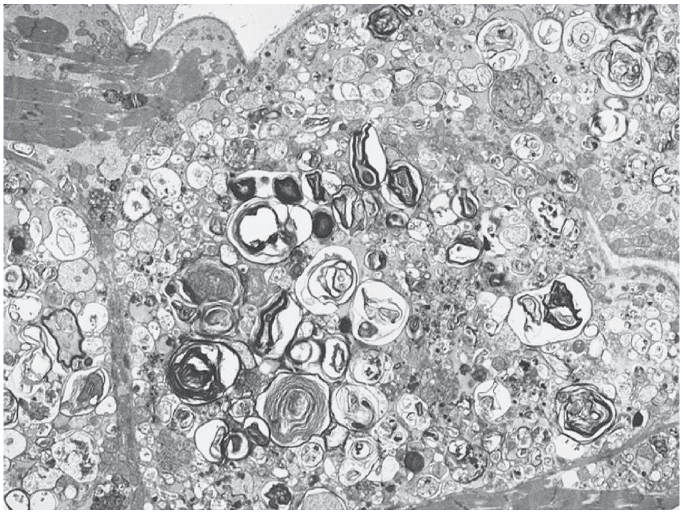

FIGURE 267-9 Fabry's disease. Transmission electron micrograph of a right ventricular endomyocardial biopsy specimen at high magnification showing the characteristic concentric lamellar inclusions of glycosphingolipids accumulating as a result of deficiency of the lysosomal enzyme alpha-galactosidase A. Image taken at 15,000× original magnification. *(Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.)*

RESTRICTIVE CARDIOMYOPATHY

Most restrictive cardiomyopathy (RCM) is due to acquired causes, and there is increasing emphasis to diagnose amyloidosis due to transthyretin variants **(see Chap. 266)**. Inherited metabolic and storage diseases can cause RCM, as can variants in *DES* causing combined cardiac and skeletal myopathy and sarcomere variants causing an overlap of RCM and hypertrophic cardiomyopathy.

■ FURTHER READING

Arbelo E et al: ESC guidelines for the management of cardiomyopathies. Eur Heart J 44:3503, 2023.

Gasperetti A et al: Arrhythmic risk stratification in arrhythmogenic right ventricular cardiomyopathy. Europace 25:euad312, 2023.

Groh WJ et al: 2022 HRS expert consensus statement on evaluation and management of arrhythmic risk in neuromuscular disorders. Heart Rhythm 19:e61, 2022.

Helms AS et al: Translation of new and emerging therapies for genetic cardiomyopathies. JACC Basic Transl Sci 7:70, 2021.

Heymans S et al: Dilated cardiomyopathy: Causes, mechanisms, and current and future treatment approaches. Lancet 402:998, 2023.

Ho CY et al: Genotype and lifetime burden of disease in hypertrophic cardiomyopathy insights from the Sarcomeric Human Cardiomyopathy Registry (SHaRe). Circulation 138:1387, 2018.

Lampert R et al: Vigorous exercise in patients with hypertrophic cardiomyopathy. JAMA Cardiol 8:595, 2023.

Mazzarotto F et al: Reevaluating the genetic contribution of monogenic dilated cardiomyopathy. Circulation 141:387, 2020.

Olivotto I et al: Mavacamten for treatment of symptomatic obstructive hypertrophic cardiomyopathy (EXPLORER-HCM): A randomized, double-blind, placebo-controlled, phase 3 trial. Lancet 396:759, 2020.

Orsborne C et al: Disease-specific therapy for the treatment of cardiovascular manifestations of Fabry disease: A systemic review. Heart 110:19, 2023.

268

Acute and Chronic Myocarditis

Neal K. Lakdawala,
Lynne Warner Stevenson, Joseph Loscalzo

Myocarditis is defined as inflammation of the heart muscle, which may present acutely, subacutely, or insidiously. Outcomes can include resolution, relapsing course, or progression to chronic cardiomyopathy. Myocarditis is generally considered acute when presenting with less than a month of symptoms, often only a few days. The profile of acute myocarditis is changing as the use of sensitive troponin assays and cardiac magnetic resonance imaging (MRI) are increasing recognition of mild cases with good outcomes. Acute myocarditis is most often attributed to active infection but can be caused by multiple types of noninfectious inflammation, such as sarcoidosis, giant cell and eosinophilic myocarditis, and systemic autoimmune disease or immunotherapies. It is uncertain as to how often chronic myocardial inflammation contributes to dilated cardiomyopathy. As with other cardiomyopathies, there is increasing implication of genetic predisposition to inflammatory responses and of independent pathogenic variants as primary cause of myocyte pathology, with almost one in five patients with inflammatory myocarditis harboring genetic variants considered pathogenic for cardiomyopathy.

MODELS OF INFECTIOUS MYOCARDITIS

Infectious agents can injure the myocardium through direct invasion with disruption of normal cellular processes and through activation of different phases of immune responses with or without persistent infection. Although myocarditis has been reported with most types of infectious agents, it is most often associated with viruses and the protozoal disease *Trypanosoma cruzi* (Chagas' disease). The pathogenesis of viral myocarditis has been traditionally portrayed in three phases, arising from extensive study in murine models of enteroviruses, with much less known about human infection. As defined by infection with coxsackie B virus, phase I includes direct invasion of the myocardium. Cellular entry may be enhanced by some genetic variants of the coxsackie/adenovirus receptors. The enteroviral protease A degrades the myocyte structural protein dystrophin and interacts with other proteins to induce apoptosis and interfere with autophagy. Direct myocardial invasion with entry of the viral genome has been shown to occur with the enteroviruses, HIV, and dengue virus. However, most common viruses do not appear to inflict cardiac damage directly. Early cardiac effects may result primarily from cytokine storm and other nonspecific immune responses leading to myocardial depression. This initial immune response appears to be crucial to recovery, as early immunosuppression increased viral replication and worsened cardiac injury in animal models.

The second phase is host response to infection in recognition of common antigenic patterns, triggering macrophage activation and expansion of specific T- and B-cell populations. Myocarditis due to this phase of immune response without direct viral-mediated injury is considered to be virally "triggered" myocarditis, implicated with respiratory viruses including adenovirus, influenza, and COVID-19. Molecular mimicry between viral and cardiac antigens may be the cause of some cases of myocardial injury caused by autoreactive T cells in the absence of ongoing viral infection. There is increasing study of the role of viral "hijacking" of cellular machinery to produce extracellular vesicles containing viral RNA that may then gain protected entry into other cells. A third phase of progression to chronic cardiomyopathy has been demonstrated in animal models of enteroviruses. However, myocarditis is rarely due to enteroviruses in adults. There are limited data providing direct links between chronic cardiomyopathy and previous viral infection (see "Chronic Myocarditis and Cardiomyopathy" below). The MRI pattern of intramyocardial late gadolinium enhancement in chronic cardiomyopathy was previously considered as evidence of prior viral myocarditis, but this pattern is now recognized to be common in genetic cardiomyopathy as well.

ACUTE MYOCARDITIS

PRESENTATION AND DIAGNOSIS

Acute myocarditis in adults typically occurs between 30 and 45 years and is more common in men than women, who tend to present at over 45 years of age. Prodromal symptoms typical of influenza, gastroenteritis, or upper respiratory infection may have occurred days to a few weeks earlier. Fever is present in over half of cases. Chest pain is the most common symptom of acute myocarditis, occurring in >80% of patients, who also often present with dyspnea or arrhythmias. Arrhythmias can manifest as palpitations or syncope but can also cause sudden cardiac death, as in an autopsy series in which inflammatory myocarditis was diagnosed 3–10% of previously healthy young adults. At presentation, about one-fourth of acute myocarditis cases include left ventricular ejection fraction (LVEF) <0.50, ventricular arrhythmias, or clinical shock, which is reported in ~3–9% of cases and termed *fulminant myocarditis*.

Viral titers are not often helpful to guide initial therapy of acute myocarditis. Respiratory viral panels can confirm recent infection with influenza and adenovirus, most implicated in acute myocarditis in adults with a viral prodrome. Finding of COVID-19 may alert to extracardiac involvement. Specific infectious serologies for HIV, Chagas' disease, cytomegalovirus, dengue fever, and Lyme disease and serologies for systemic autoimmune disease should be sent in selected patients. Eosinophil counts should be routinely checked because hypereosinophilia is present in most cases of eosinophilic myocarditis.

The common definition for "probable myocarditis" includes new appearance of least one symptom listed above and at least one supporting finding, which can be elevated troponin levels, consistent ECG findings, or imaging findings of decreased LVEF or obvious wall motion abnormality. Other laboratory findings may include elevated creatine phosphokinase levels, indicating cardiac or skeletal muscle involvement, and elevated C-reactive protein levels. The most common electrocardiogram (ECG) findings are ST elevation suggesting infarction, but also include conduction block, tachyarrhythmias, and nonspecific ST-T changes. Because the combination of chest pain, ECG changes, and elevated troponin is typical of both acute myocarditis and myocardial infarction, the first step in the differential diagnosis is often coronary artery imaging. Subsequent diagnosis of acute myocarditis has been reported in up to one-third of patients with myocardial infarction and nonobstructed coronary arteries (MINOCA).

"Definite myocarditis" previously required positive biopsy findings of typical inflammatory lymphocytic infiltrate on myocardial biopsy **(Fig. 268-1)**.

Newer techniques of immunohistochemistry may increase sensitivity of biopsies and advance our understanding of the inflammatory cell populations. However, the diagnosis can now be confirmed noninvasively when typical symptomatic presentation is accompanied by combination of elevated troponin levels and consistent findings on MRI, which include evidence of both myocardial edema and nonischemic myocardial injury in the absence of other known cause **(Fig. 268-2)**.

Although cardiac MRI may often be sufficient to diagnose acute myocarditis, endomyocardial biopsy should be performed when ventricular arrhythmias, conduction block, elevated eosinophil count, or evidence of systemic autoimmune disease suggests causes of acute myocarditis with specific implications for therapy and prognosis, as discussed below in "Noninfective Inflammatory Myocarditis."

THERAPY AND OUTCOMES

The prognosis is good for "uncomplicated" acute myocarditis, presenting with LVEF ≥0.50 without ventricular arrhythmias or circulatory shock. In the largest series, this low-risk group accounted for about three-fourths of patients with acute myocarditis, in which spontaneous recovery was common without specific therapies other than diuretics and nonsteroidal analgesics for chest pain. Almost all of these patients

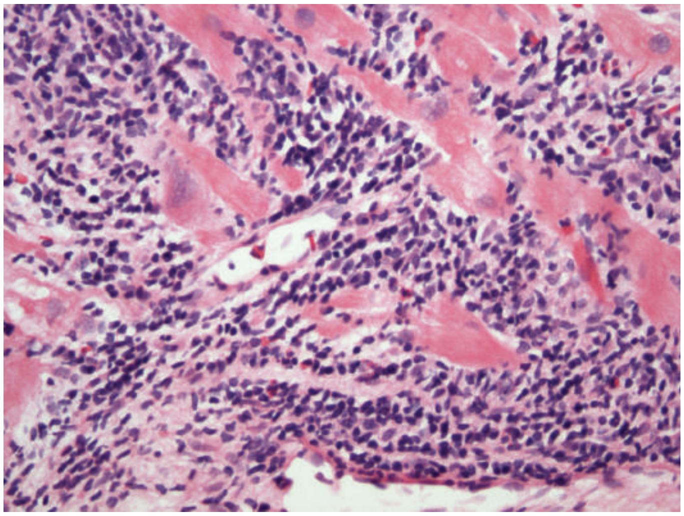

FIGURE 268-1 Acute myocarditis. Microscopic image of an endomyocardial biopsy showing massive infiltration with mononuclear cells and occasional eosinophils associated with clear myocyte damage. The myocyte nuclei are enlarged and reactive. Such extensive involvement of the myocardium would lead to extensive replacement fibrosis even if the inflammatory response could be suppressed. Hematoxylin and eosin–stained section, 200× original magnification. *(Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.)*

still had LVEF ≥0.50 on follow-up imaging and survived without transplantation at 5 years of follow-up. In the one-fourth of patients with complicated presentations, ~18% died or underwent heart transplantation in the next 5 years. Patients with reduced LVEF are generally discharged on the recommended therapies for heart failure with reduced ejection fraction (HFrEF). There are no data supporting routine immunosuppression in patients with acute myocarditis with presumed viral etiology. Fulminant myocarditis requiring high-dose inotropic therapy or mechanical circulatory support occurs in 3–9% of acute myocarditis, with early outcomes of refractory cardiogenic shock but recovery to near-normal LVEF in >50% of affected patients.

CHRONIC MYOCARDITIS AND CARDIOMYOPATHY

Patients with reduced LVEF at the time of presentation with acute myocarditis are at higher risk of developing chronic cardiomyopathy. When the LVEF does not resolve in the initial months after diagnosis, particularly with persistent elevation of troponin, cardiac magnetic resonance imaging (CMRi) or endomyocardial biopsy may reveal ongoing inflammation and developing fibrosis. Evaluation should again be considered for specific etiologies of myocarditis including noninfectious causes of inflammatory disease and for other causes of dilated cardiomyopathy (Chap. 266, Table 266-2).

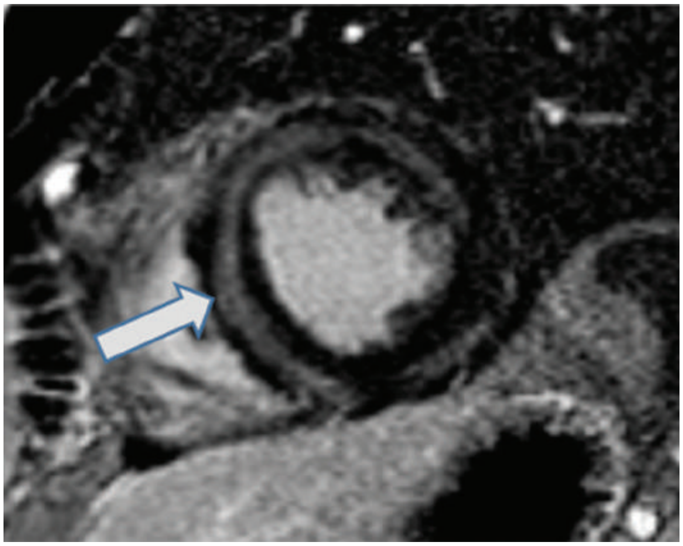

FIGURE 268-2 Magnetic resonance image of myocarditis showing the typical midwall location (*arrow*) for late gadolinium enhancement from cardiac inflammation and scarring. *(Image courtesy of Ron Blankstein, MD, and Marcelo Di Carli, MD, Division of Nuclear Medicine, Brigham and Women's Hospital, Boston.)*

Evidence of chronic inflammatory disease may also be found in some patients presenting with gradual onset of dilated cardiomyopathy without prior acute myocarditis. Immunosuppression for nonspecific lymphocytic myocarditis in the chronic setting has been tried in multiple series but has not demonstrated convincing benefit. Additional findings of anticardiac antibodies and viral genome fragments are of uncertain significance, as they could be either causes of ongoing cardiac injury or neutral bystanders. Assay for anticardiac antibodies and viral genome analysis in chronic cardiomyopathy continue to be investigated but are not in routine clinical use.

In the past, dilated cardiomyopathy without other cause was generally assumed to result from previously unrecognized viral myocarditis, in part because CMRi patterns of fibrosis suggested prior inflammation. However very similar patterns are typical of genetic cardiomyopathies. The expanding access to genetic testing reveals that 20–40% of patients with dilated cardiomyopathy have pathogenic variants sufficient to cause their cardiomyopathy, and the prevalence and nature of these variants are similar in acute myocarditis. Mapping of the interconnections between genetic predisposition, inflammation, and cardiomyopathy is evolving rapidly.

SPECIFIC CAUSES OF INFECTIOUS MYOCARDITIS

The first viruses implicated in infectious myocarditis were the picornavirus family of RNA viruses, principally the enteroviruses, coxsackie virus, echovirus, and poliovirus. Influenza, another RNA virus, is implicated with varying frequency every winter and spring as epitopes change. Of the DNA viruses, adenovirus, vaccinia, and the herpesviruses (varicella-zoster virus, cytomegalovirus, Epstein-Barr virus, and human herpesvirus 6 [HHV6]) can cause myocarditis but also occur commonly in the healthy population. Polymerase chain reaction (PCR) detects viral genomes in some patients with dilated cardiomyopathy and persistent lymphocytic infiltrates defined as "chronic myocarditis," such as the DNA viruses parvovirus B19 and HHV6, but these are also found in normal "control" hearts, so their contribution to chronic cardiomyopathy is uncertain.

COVID-19 is an RNA virus for which infection was associated with substantial rates of cardiac involvement in patients hospitalized with severe disease, including acute coronary syndrome, thromboemboli, pericarditis, and myocarditis, for which the diagnosis was limited to clinical findings and elevated enzymes rather than demonstration of lymphocytic infiltrates, as biopsies were rarely performed. From broad population data, the rate of myocarditis during the pandemic was estimated at 150/100,000 people, a 15-fold increase over pre-COVID incidence. As for most causes of viral myocarditis, the most common adult patients are young men under the age of 40. As for other viruses, there are multiple potential mechanisms of cardiac injury. Viral particles were identified by PCR in myocardial samples but in low levels at which pathogenicity was uncertain. Other mechanisms include cytokine storm, activated T cells and macrophages, and potential molecular mimicry by which autoimmune reactions may be triggered, as for some animal models of myocarditis with other viruses. As with multiple other viruses, extracellular vesicles containing viral RNA have been identified in myocardium during COVID infection, potentially promoting protected entry into other cells for further injury. In addition, the endothelium is also vulnerable, leading to prothrombotic endotheliopathy that may contribute to myocardial ischemia and stroke, along with the systemic coagulopathy.

HIV was associated with a dilated cardiomyopathy incidence of 1–2%, but this is declining with the availability of highly active antiretroviral therapy. Cardiomyopathy in HIV may also result from other associated viruses, such as cytomegalovirus and hepatitis C. Antiviral drugs to treat chronic HIV can cause cardiomyopathy, both directly and through drug hypersensitivity. The clinical picture may be complicated by pericardial effusions and pulmonary hypertension. There is an increased frequency of lymphocytic myocarditis found at autopsy; HIV viral particles have been demonstrated in the myocardium in some cases.

For any serious infection, the systemic inflammatory response can cause nonspecific depression of cardiac function, which is generally reversible if the patient survives. Other viruses implicated specifically in myocarditis include mumps, respiratory syncytial virus, the arboviruses (dengue fever and yellow fever), and arenaviruses (Lassa fever).

OTHER INFECTIOUS CAUSES

Parasitic Myocarditis Chagas' disease is the third most common parasitic infection in the world and the most common infective cause of cardiomyopathy. The protozoan *T. cruzi* is transmitted by the bite of the reduviid bug, endemic in the rural areas of South and Central America. Transmission can also occur through blood transfusion, organ donation, from mother to fetus, and occasionally orally. Programs to eradicate the insect vector have decreased the global prevalence from about 18 million in 1990 to 6 million, with the highest number of cases in Brazil and Argentina and increased prevalence also in Spain. It is estimated that >300,000 residents in the United States are infected with Chagas' disease, a minority of whom have acquired it locally.

Initial infection with *T. cruzi* is usually silent but, in 5–10% of cases, can present with systemic illness and acute myocarditis with systemic dense parasitic infiltration in the myocardium forming pseudocysts that rupture and lead to diffuse myocyte necrosis. After initial infection, patients enter an "indeterminate phase," which can last for decades. The silent progression during this phase was at one time attributed to secondary immune activation, but a degree of persistent parasitemia can now be detected and is recognized as the major determinant of ongoing inflammation and progression to chronic heart failure. Depopulation of parasympathetic neurons in the heart and gastrointestinal tract likely contributes to clinical disease. Understanding of the disease patterns has improved in part from the ability to test blood donors for infection. Better detection of the organism has revealed that both the prevalence and survival of patients in the indeterminate phase are higher than previously recognized, with over half of patients remaining asymptomatic.

T. cruzi infection should be considered during evaluation of acute myocarditis or cardiomyopathy in all patients from endemic regions and for other patients with suggestive features. These include sinus and atrioventricular node conduction system disease and right bundle branch block, with frequent atrial and ventricular arrhythmias, some of which may cause sudden death before symptomatic disease. The dilated ventricles are thrombogenic and sometimes have aneurysms. For acute myocarditis, PCR is the most sensitive test and trypomastigotes may also be detected in blood when parasite levels are high. In the chronic phase with low parasite burden, multiple serologic tests may be necessary for diagnosis, but their sensitivity and specificity are improving.

Antiparasitic treatment with nifurtimox or benznidazole is recommended for acute myocarditis with Chagas' cardiomyopathy. During the indeterminate phase prior to clinical symptoms, therapy is widely recommended for children. For the indeterminate phase, there is controversy over the routine use of antiparasitic treatment, which is associated with significant toxicity. Therapy is more often advocated for patients <50 years old, who are more likely to tolerate chronic suppressive therapy. Treatment is recommended for premenopausal women and for patients on immunosuppression, such as after cardiac transplantation. A large multinational trial showed no benefit for benznidazole given to patients with chronic heart failure from Chagas' cardiomyopathy. In these patients, general therapy is as indicated for other HFrEF, with increased indications for anticoagulation and for pacemaker defibrillators. Improvement of LVEF is uncommon and survival is lower than for other cardiomyopathies after the onset of overt clinical heart failure.

Trichinellosis (trichinosis) is caused by the *Trichinella* genus of nematodes (roundworms). The larvae are ingested with undercooked meat, with an initial intestinal phase after which larvae migrate into skeletal muscles, causing myalgias, weakness, and fever. Periorbital and facial edema and conjunctival and retinal hemorrhage may also be seen. Although the larva may occasionally invade the myocardium, clinical heart failure is rare and, when observed, attributed to the eosinophilic inflammatory response. The diagnosis is made from the serologies and is further supported by the presence of eosinophilia. Treatment includes anthelminthic drugs and glucocorticoids if inflammation is severe.

Bacterial Infections Most bacterial infections can involve the heart occasionally through direct invasion and abscess formation but do so rarely. More commonly, systemic inflammatory responses to severe infection and sepsis depress myocardial contractility.

Diphtheria is caused by bacillus *Corynebacterium diphtheriae*, usually presenting with an upper respiratory illness particularly affecting the pharynx, where a pseudomembrane is formed in response to the diphtheria toxin. This toxin interferes with protein synthesis in the heart and may particularly affect the conduction system. Cardiac involvement is the most common cause of death from this infection but rarely occurs with the occasional dominant cutaneous presentation. The prevalence of vaccines has shifted the incidence of diphtheria to countries without routine immunization and to older populations who have lost their immunity. The diagnosis is made from pharyngeal bacterial culture, which requires special culture media and a positive assay for the toxin but can be suspected from gram-positive rods on a Gram stain. Clinical suspicion is sufficient indication for the specific antitoxin, which should be administered as soon as possible, with higher priority than antibiotic therapy.

Streptococcal infection with β-hemolytic streptococci is most commonly associated with acute rheumatic fever and is characterized by inflammation and fibrosis of cardiac valves and systemic connective tissue, but it can also lead to a myocarditis with focal or diffuse infiltrates of mononuclear cells. Other systemic bacterial infections that can involve the heart include brucellosis, legionella, meningococcus, mycoplasma, psittacosis, and salmonellosis, for which specific treatment is directed at the systemic infection.

Tuberculosis can involve the myocardium directly as well as through tuberculous pericarditis but rarely does so when the disease is treated with antibiotics. Whipple's disease is caused by *Tropheryma whipplei*. The usual manifestations are in the gastrointestinal tract, but pericarditis, coronary arteritis, valvular lesions, and occasionally clinical heart failure may also occur. Multidrug antituberculous regimens are effective, but the disease tends to relapse even with appropriate treatment.

Tick-Borne Infections Spirochetal myocarditis has been diagnosed from myocardial biopsies containing *Borrelia burgdorferi*, which causes Lyme disease. Lyme carditis most often presents with arthritis and conduction system disease that resolves within 1–2 weeks of antibiotic treatment and is only rarely implicated in chronic heart failure. Other borrelia species carried by either ticks or lice can cause relapsing fever. Additional tick-borne illnesses associated with febrile illnesses and myocarditis include Rocky Mountain spotted fever, Q fever, and ehrlichiosis, all of which are treated with doxycycline alone or in combination with other agents.

NONINFECTIVE INFLAMMATORY MYOCARDITIS

Myocardial inflammation can occur in the absence of infectious causes. The paradigm of noninfective inflammatory myocarditis is cardiac transplant rejection, from which we have learned that myocardial depression can develop and reverse quickly, that noncellular mediators such as antibodies and cytokines play a major role in addition to lymphocytes, and that myocardial antigens are exposed by prior physical injury and viral infection. In the largest registry of acute myocarditis, chronic systemic inflammatory diseases were implicated in 7% of cases and in 15% of those with high-risk features.

The most commonly diagnosed noninfective inflammatory process affecting the myocardium is sarcoidosis. Sarcoidosis, as discussed in **Chap. 379**, is a multisystem granulomatous disease most commonly affecting the lungs, but involving many organs including skin, eyes, liver, nervous system, and bones, as well as the heart. The epidemiology appears to be changing, now recognized in men and women of all races and ethnic groups, typically between 30 and 50 years old but often after the age of 50 in women. Sarcoidosis is now understood as a combination of foreign antigen presentation and a dysregulated immune response that leads to ongoing inflammation, including activated macrophages. Occupational exposures include agriculture, firefighting, metal-working, and construction work, with silica dust, pesticides, mold, and other inhaled particles as examples of implicated antigens. The occurrence of sarcoidosis is increased in family members, reflecting potential sharing of environments and of genetic variants, which have been identified in loci affecting immune responses.

Cardiac sarcoidosis often accompanies pulmonary sarcoidosis but frequently occurs without detectable lung disease. The time course, burden and activity of cardiac granulomata, and the degree of extracardiac involvement are remarkably variable. Patients may present with rapid-onset heart failure and ventricular tachyarrhythmias, conduction block, chest pain syndromes, or minor cardiac findings in the setting of pulmonary sarcoidosis, ocular involvement, an infiltrative skin rash, or a nonspecific febrile illness. Chronic sarcoidosis may go unrecognized for months or years. When ventricular tachycardia or conduction block dominates the initial presentation of heart failure without coronary artery disease, suspicion should be high for sarcoidosis as a cause of cardiomyopathy. Right bundle branch block should raise suspicion for sarcoidosis but is uncommon with other cardiomyopathies, in which left bundle branch block is more common.

The pathology of cardiac sarcoidosis often shows a spectrum from metabolically active granulomas to bland fibrosis at the sites of previous inflammation. Regional wall-motion abnormalities are common, but global ventricular function may be preserved early during mild disease. When left ventricular function is only mildly reduced, ventricular size may be near normal, sometimes described as "nondilated" or "minimally dilated cardiomyopathy." Although occasionally listed with restrictive cardiomyopathies, sarcoidosis with a reduced ejection fraction generally has the phenotype of dilated or minimally dilated cardiomyopathy. There may be a right ventricular predominance of both dilation and ventricular arrhythmias, in which case potential diagnoses include genetic arrhythmogenic right ventricular cardiomyopathy, with which sarcoidosis shares multiple features.

Diagnosis of cardiac sarcoidosis is easiest in the presence of biopsy-proven sarcoidosis of other organs. Imaging of the heart can show regional wall motion abnormalities or small ventricular aneurysms. MRI of the heart can identify late gadolinium enhancement in a pattern of fibrosis not compatible with myocardial infarction. Computed tomography of the chest often reveals pulmonary lymphadenopathy even in the absence of clinical lung disease. Positron emission tomography (PET) of the whole chest can highlight active sarcoid lesions that are avid for glucose in heart or lung. When metabolically active adenopathy is detected, biopsy is often necessary to rule out malignancy or chronic granulomatous infections such as tuberculosis or histoplasmosis before treating with immunosuppression for sarcoidosis. The scattered granulomata of sarcoidosis are commonly missed on cardiac biopsy **(Fig. 268-3)**.

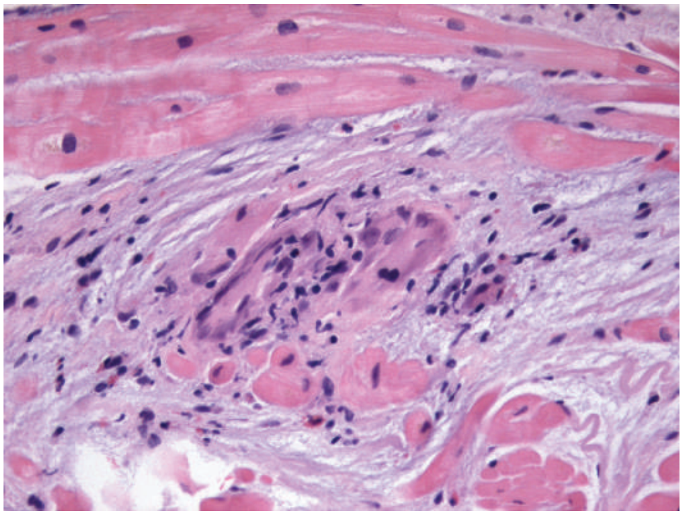

FIGURE 268-3 Sarcoidosis. Microscopic image of an endomyocardial biopsy showing a noncaseating granuloma and associated interstitial fibrosis typical of sarcoidosis. No microorganisms were present on special stains, and no foreign material was identified. Hematoxylin and eosin–stained section, 200× original magnification. *(Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.)*

Immunosuppression for sarcoidosis is generally initiated with glucocorticoids. Initial dosing was traditionally high but now is more often started at 30–40 mg daily or less, with early addition or substitution of steroid-sparing agents like methotrexate or mycophenolate. Third-line treatment may include tumor necrosis factor-alpha inhibitors or other immunomodulators under investigation. The impact of treatment is usually more apparent for suppression of arrhythmias than for improvement of markedly impaired left ventricular dysfunction. Devices are often indicated for tachyarrhythmias or for conduction disease and usually include both pacing and defibrillation capability. Patients with sarcoidosis are immunosuppressed, and their management should be supervised by an experienced multidisciplinary team. Cardiac sarcoidosis may respond to the initial immunosuppression course without recurrence, may fluctuate over a chronic course, or occasionally may lead to end-stage heart failure requiring heart transplantation.

Giant cell myocarditis is less common than sarcoidosis, but accounts for 10–20% of biopsy-proven myocarditis, typically presenting acutely with rapidly progressive heart failure and tachyarrhythmias. Giant cell myocarditis occurs equally in men and women, usually of older age than those with acute viral myocarditis, and is more often associated with systemic autoimmune disorders. Unlike sarcoidosis, the more diffuse involvement with giant cell myocarditis usually leads to diagnostic endomyocardial biopsy, revealing granulomatous lesions surrounded by extensive inflammation infiltrate, often with eosinophilic infiltration. However, the occasional finding of giant cell myocarditis in explanted hearts after a previous diagnosis of sarcoidosis suggests that they may share the same disease spectrum. Glucocorticoid therapy alone is rarely effective, but in combination with other immunosuppression therapies similar to those used for severe transplant rejection, the rate of death or heart transplantation has decreased for patients who are stable at the time of presentation. Most patients presenting with cardiogenic shock from giant cell myocarditis progress to need urgent mechanical support or transplantation, for which they may be rendered ineligible by severe infection resulting from intensive immunosuppression.

Eosinophilic myocarditis is one of the causes of fulminant myocarditis but may be missed on milder presentation that does not include assessment for hypereosinophilia, which is present in about three-fourths of diagnosed cases. Myocardial toxicity of eosinophils may lead to apical thrombi on cardiac imaging, and MRI reveals inflammation acutely and subendocardial fibrosis in chronic cases. Endomyocardial biopsies show infiltration with lymphocytes, neutrophils, and a high proportion of eosinophils. Hypersensitivity to chronic medications accounts for up to one-third of cases and is often cured by withdrawal of the agent and acute steroid therapy. Particularly in Mediterranean and African countries, hypereosinophilia can result from chronic parasitic infection, for which treatment may cure the myocarditis. About 12% of cases are due to eosinophilic granulomatosis with polyangiitis (EGPA, formerly Churg-Strauss syndrome), which responds to immunosuppression but may have a relapsing course. The hypereosinophilic syndrome (HES) accounts for 10–15% of cases, resulting from myeloproliferative variants or overproduction of eosinophil production, for which combined immunosuppression may soon include therapies under investigation to decrease eosinophil production. About 30% of HES cases are currently unexplained. Without curative therapy, the initial inflammatory necrosis can lead to thrombosis, valve disease, and progressive endomyocardial fibrosis with restrictive physiology, termed *Loeffler endocarditis*, which presents as a restrictive cardiomyopathy.

Myocarditis is often associated with systemic inflammatory diseases, such as *polymyositis* and *dermatomyositis*, which can affect cardiac as well as proximal skeletal muscle and are usually diagnosed from autoantibodies. Myocarditis with lymphocytic infiltrates on endomyocardial biopsy can be seen in some patients with systemic lupus erythematosus, but multiple cardiac manifestations also include accelerated coronary artery disease, valvular involvement from sterile endocarditis, pericarditis, vasculitis, and chloroquine cardiotoxicity.

Vaccines have occasionally been implicated in myocarditis. This has been best studied after the smallpox vaccine in military populations. As this is a live vaccine, it is unclear whether this is direct injury, but it is generally assumed to be a hypersensitivity response. More recent concern relates to COVID-19 vaccines, after which the overall rate of myocarditis is estimated to be about 1/100,000 vaccine doses (increased to 2–3/100,000 for recipients age 18–39 years). Most cases of vaccine-induced myocarditis resolve without hospitalization. Male adults under 40 are at highest risk of myocarditis from the COVID-19 vaccines, as they are for primary infectious myocarditis after COVID-19 and other viruses. The risk for men under 40 increases after repeat COVID-19 vaccines, for which the benefit/risk for individuals should be considered.

The most dramatic form of noninfectious inflammatory myocarditis is that seen with combined immune checkpoint inhibitors. Targeted monoclonal antibody therapy to unblock the host immune response has produced remarkable remission of multiple advanced tumors. Inhibitory receptors on T lymphocytes (such as CTLA-4 and PD-1) and the programmed death ligands, such as PD-L1, interact in normal self-regulation to inhibit overactivation of immune responses. Tumor cells can upregulate these ligands to hide from immune recognition. Therapeutic antibodies against these inhibitory receptors or ligands can heighten host defenses against the tumor but also unleash immune attack against host tissues expressing PD-L1, which include skeletal and cardiac muscle, endothelial cells, and multiple other organs, such as lung, liver, pancreas, thyroid, and skin. With cardiac involvement, troponin is often elevated, B-type natriuretic peptide may be elevated, and creatine phosphokinase may be high, particularly with skeletal involvement. The diagnosis should be suspected immediately with acute cardiac presentation in patients treated with checkpoint inhibitors. Admission to the coronary care unit is currently recommended for an elevated troponin levels and ECG changes, which may be nonspecific but can also include conduction blocks and bizarre arrhythmias. Echocardiography may suggest myocardial edema, but initial ejection fraction may not be markedly reduced. Initial care should generally not be delayed for endomyocardial biopsy, which typically shows extensive lymphocytic infiltration. Patients may also present initially with other acute organ system involvement, which warrants urgent multidisciplinary management in intensive care. Therapy with high-dose glucocorticoids should be initiated rapidly and may be followed soon by more targeted immune inhibition. Reports of fatality in checkpoint inhibition myocarditis were initially very high, but outcomes are improving with earlier recognition and therapy.

■ FURTHER READING

Ammirati E et al: Management of acute myocarditis and chronic inflammatory cardiomyopathy. Circ Heart Failure 13:e007405, 2020.

Drent M et al: Challenges of sarcoidosis and its management. N Engl J Med 385:1018, 2021.

Fairweather D et al: COVID-19, myocarditis and pericarditis. Circ Res 132: 1302, 2023.

Moslehi J, Salem JE: Immune checkpoint inhibitor myocarditis treatment strategies and future directions. JACC CardioOncol 4:704, 2022.

269 Dilated Cardiomyopathies

Neal K. Lakdawala, Lynne Warner Stevenson, Joseph Loscalzo

As described in **Chap. 266**, the phenotype of dilated cardiomyopathy (DCM) is characterized by decreased left ventricular systolic function, typically with increased left ventricular dimensions, although dilation may be minimal in some cases. Multiple causes and contributing factors have been implicated **(Table 269-1)**.

As discussed in **Chaps. 267 and 268**, many cases are attributed to prior myocarditis, and an increasing number are associated with pathogenic genetic variants, but other causes include toxic and metabolic disorders and peripartum cardiomyopathy, which are discussed in this chapter, along with takotsubo cardiomyopathy, which has a distinct presentation and phenotype that often resolves.

Despite the multiple etiologies and variable initial presentations, DCM often progresses into a convergent clinical phenotype similar to that of other injury such as acute myocardial infarction. Some myocytes may die early in the course, while others survive only to have later programmed cell death (apoptosis), and remaining myocytes develop hypertrophy in response to increased wall stress. Local and circulating neurohormonal factors stimulate deleterious secondary responses that contribute to progression of disease. Dynamic remodeling of the interstitial scaffolding affects diastolic function and the amount of ventricular dilation. Mitral regurgitation commonly develops as the ventricle dilates and the valvular apparatus is distorted and is commonly moderate to severe by the time heart failure is advanced. There is increasing evidence for the role of chronic inflammation in disease progression even when not initially implicated. Many cases that present "acutely" have progressed silently through these stages of injury over months to years. Dilation and decreased function of the right ventricle may result directly from the initial injury but more often develop later in response to elevated afterload presented by secondary pulmonary hypertension and in relation to mechanical interactions with the failing left ventricle.

The secondary responses are often modifiable or reversible. Almost a third of patients with recent-onset cardiomyopathy in the absence of coronary artery disease demonstrate substantial spontaneous recovery to normal ejection fraction. Partial recovery to left ventricular ejection fraction (LVEF) >0.40 ("heart failure with improved ejection fraction") is common in chronic DCM during recommended therapy with neurohormonal modulation, cardiac resynchronization therapy for left bundle branch block, and diuretics as needed to maintain fluid balance, as recommended for other heart failure with reduced LVEF **(Chap. 265)** regardless of the initial cause of DCM. Additional aspects of diagnosis, therapy, and outcomes for specific etiologies of DCM other than the infectious, inflammatory, and genetic cardiomyopathies are discussed below.

CARDIOTOXICITY AND CARDIOMYOPATHY

Cardiotoxicity has been reported with multiple environmental and pharmacologic agents. Often these associations are seen only with very high levels of exposure or acute overdoses, in which acute electrocardiographic and hemodynamic abnormalities may reflect both direct drug effect and systemic toxicity. *Alcohol* is the most common toxin implicated in chronic DCM. Although excess consumption may contribute to >10% of cases of heart failure, including exacerbation of heart failure with structural heart disease, alcoholic cardiomyopathy is relatively rare and remains a diagnosis of exclusion. Moreover, Mendelian randomization studies have not identified a link between genetically predicated alcohol consumption and heart failure, suggesting that the population attributable risk of alcohol to the overall heart failure epidemic is modest. Alcoholic cardiomyopathy causes

TABLE 269-1 Major Causes of Dilated Cardiomyopathy (with Common Examples)

Inflammatory Myocarditis (see Chap. 268)

- Infective
  - Viral (coxsackie,[a] adenovirus,[a] COVID-19, HIV)
  - Parasitic (*Trypanosoma cruzi*—Chagas' disease, trypanosomiasis, toxoplasmosis)
  - Bacterial (diphtheria)
  - Spirochetal (*Borrelia burgdorferi*—Lyme disease)
  - Rickettsial (Q fever)
  - Fungal (with systemic infection)
- Noninfective
  - Granulomatous inflammatory disease
    - Sarcoidosis
    - Giant cell myocarditis
  - Eosinophilic myocarditis
  - Polymyositis, dermatomyositis
  - Collagen vascular disease
  - Checkpoint inhibitor chemotherapy
  - Transplant rejection

Toxic

- Alcohol
- Catecholamines: amphetamines, cocaine
- Chemotherapeutic agents (anthracyclines, trastuzumab)
- Interferon
- Other therapeutic agents (hydroxychloroquine, chloroquine)
- Drugs of misuse (testosterone and other anabolic steroids emetine)
- Heavy metals: lead, mercury
- Occupational exposure: hydrocarbons, arsenicals

Metabolic[a]

- Nutritional deficiencies: thiamine, selenium, carnitine
- Electrolyte deficiencies: calcium, phosphate, magnesium
- Endocrinopathy
  - Thyroid disease
  - Pheochromocytoma
  - Diabetes mellitus
- Obesity
- Hemochromatosis
- Inherited metabolic pathway defects (see Chap. 267)

Familial[a] (see Table 267-1)

- Cardiomyopathies without extracardiac involvement
- Cardiomyopathy with skeletal myopathy, for example:
  - Dystrophin-related dystrophy (Duchenne's, Becker's)
  - Mitochondrial myopathies (e.g., Kearns-Sayre syndrome)
- Hemochromatosis
- Susceptibility to immune-mediated myocarditis
- Associated with other systemic diseases

Miscellaneous (Shared Elements of Above Etiologies)

- Arrhythmogenic ventricular cardiomyopathy
- Peripartum cardiomyopathy
- Left ventricular noncompaction[a]
- Tachycardia-related cardiomyopathy
  - Supraventricular arrhythmias with uncontrolled rate
  - Very frequent nonsustained ventricular tachycardia or high burden of premature ventricular complexes

[a] Some specific cases can be linked now to specific genetic mutation in a familial cardiomyopathy; others with similar phenotypes that appear to be acquired or idiopathic may represent genetic factors not yet identified.

many more hospital admissions in men than women, but prevalence is similar between men and women with alcoholism, with left ventricular dysfunction detected in about a third of asymptomatic patients. Estimates of the alcohol intake necessary to cause cardiomyopathy have been 3–4 ounces or ≥60–80 g of pure ethanol daily for ≥5 years, about 750 mL of wine, 6 beers, or a half pint of hard liquor, although women may develop cardiomyopathy with lower amounts of consumption. Frequent binge drinking may also be sufficient. Toxicity is attributed both to alcohol and to its primary metabolite, acetaldehyde. Chronic heavy exposure may cause neurohormonal activation and alter metabolism, protein synthesis, substrate utilization, and oxidative stress. Polymorphisms of the genes encoding alcohol dehydrogenase and the angiotensin-converting enzyme may influence the likelihood of alcoholic cardiomyopathy. Superimposed vitamin deficiencies and toxic alcohol additives are rarely implicated currently. Mutations in *TTN* and other DCM disease genes can be identified in ~10% of patients with presumed alcohol cardiomyopathy.

Many patients with alcoholic cardiomyopathy are fully functional in their daily lives without apparent stigmata of alcoholism. The cardiac impairment in severe alcoholic cardiomyopathy is the sum of both permanent damage and a substantial component that is reversible after cessation of alcohol consumption. Atrial fibrillation occurs commonly both early in the disease ("holiday heart") and in advanced stages. Medical therapy includes conventional guideline-directed medical therapies with neurohormonal, mineralocorticoid receptor, and β adrenoreceptor antagonists as well as sodium-glucose cotransporter 2 inhibitors with diuretics as needed for fluid management and careful attention to electrolyte repletion. Withdrawal should be supervised to avoid exacerbations of heart failure or arrhythmias and ongoing support arranged. Even with severe disease, marked improvement can occur within 3–6 months of abstinence, but the prognosis is grim if heavy alcohol consumption continues.

Cocaine, amphetamines, and related catecholaminergic stimulants can produce chronic cardiomyopathy as well as acute ischemia, tachyarrhythmias, malignant hypertension, aortic dissection, and stroke. Cardiac pathology reveals microinfarcts consistent with small vessel ischemia, similar to those seen with pheochromocytoma, and thrombosis secondary to endothelial dysfunction in the case of cocaine. Regular cannabis use has been linked to increased risk of atrial fibrillation, myocardial infarction, and stroke in patients with and without other known risk factors. Cannabis is not specifically implicated as a cause of cardiomyopathy in population studies but is of concern as the potency of both inhaled and edible products continues to increase.

Chemotherapy agents are the most common drugs implicated in toxic cardiomyopathy. Judicious use balances risks of the malignancy and the risks of cardiotoxicity presented not only by the drug regimens but also by the patient's cardiovascular profile and possibly genetic factors influencing myocyte response to injury. Receipt of cardiotoxic drugs or radiation may warrant designation as "stage B" heart failure, with asymptomatic changes in cardiac structure and biomarkers. Clinical recognition can be delayed as some symptoms can overlap with cancer and the prognosis with heart failure may be worse than for the underlying cancer.

Anthracyclines (e.g., doxorubicin) cause characteristic histologic changes of vacuolar degeneration and myofibrillar loss. Multiple mechanisms have been implicated, involving reactive oxygen species and iron compounds, mitochondrial damage, transcription factors such as hypoxia-induced factor, and, most recently, inhibition of topoisomerase II involved in DNA repair. Risk for cardiotoxicity increases with older age, obesity, hypertension, diabetes mellitus, preexisting cardiac disease, higher doses or combination therapies, and left chest irradiation. Systolic dysfunction can occur acutely with symptoms of heart failure noted soon after drug administration, but more often is detected by surveillance echocardiography during the first year after exposure. Doxorubicin cardiotoxicity generally does not result in marked left ventricular dilation, such that stroke volume and systemic perfusion can be low with only a modest reduction of ejection fraction. Therapy for reduced ejection fraction due to anthracycline therapy includes β-adrenergic receptor blockade and inhibition of the renin-angiotensin system, with conflicting data on whether these agents decrease toxicity when given in parallel with chemotherapy. The use of dexrazoxane, an intracellular iron chelating agent, can prevent anthracycline cardiomyopathy, but there is no consensus on when it should be used owing to concerns that it might attenuate the efficacy of cancer therapies. Once thought to have an inexorable downward course, many patients with symptomatic heart failure can improve to near-normal function with careful management, including prevention of "second-hit" insults such as atrial fibrillation or hypertension. The course differs for some children treated with these agents before puberty, in whom inadequate growth of the heart may lead to refractory heart failure as they reach their twenties.

Trastuzumab (Herceptin) is one of the humanized monoclonal antibodies that interfere with human epidermal growth receptor 2 (HER2), which is crucial for growth of some tumors, such as breast cancer, and for cardiac adaptation. Cardiotoxicity is highest when anthracyclines are administered in conjunction with trastuzumab; however, less toxicity is seen now when these agents are combined compared with the toxicity observed previously with paclitaxel for breast cancer. Although more often reversible than anthracycline cardiotoxicity, trastuzumab cardiomyopathy may persist in about a third of affected patients and can progress to clinical heart failure and death. For cardiotoxicity with anthracyclines or trastuzumab, therapy is recommended as for other causes of reduced ejection fraction.

Cardiotoxicity with *cyclophosphamide and ifosfamide* generally occurs acutely and with very high doses. 5-Fluorouracil, cisplatin, and some other alkylating agents can cause recurrent coronary spasm that occasionally leads to depressed contractility. Acute administration of interferon-α, interleukin 2, and other cytokine-based therapies can cause pericarditis, hypotension, and arrhythmias. Clinical heart failure occurring during their chronic administration usually resolves after discontinuation.

Vascular endothelial growth factor (VEGF), produced endogenously or by tumors, enhances angiogenesis by activating the VEGF signaling pathways. Monoclonal antibodies and many small-molecule *tyrosine kinase inhibitors* that affect VEGF are in use for different malignancies. Although these agents are "targeted" at specific tumor receptors or pathways, the biologic conservation of signaling pathways means that some of these drugs also find targets in the cardiovascular and other organ systems. Blood pressures increase in most patients during therapy, attributed to an imbalance between endogenous vasodilators and vasoconstrictors and alteration of glomerular function. Hypertension and proteinuria can develop with these agents, similar to preeclampsia, and presentation is associated with increased risk of future cardiac disease. Recognition of cardiotoxicity during therapy with these agents is complicated because they occasionally cause peripheral fluid accumulation (ankle edema, periorbital swelling, pleural effusions) due to local factors rather than elevated central venous pressures. Therapeutic approaches include management of associated hypertension, withdrawal of the tyrosine kinase inhibitor (when possible), and conventional treatment for heart failure. Newer tyrosine kinase inhibitors effective against multiple kinases may have more complex off-target effects. This includes the Bruton tyrosine kinase inhibitors (e.g., ibrutinib), which are used as primary therapy for lymphoid malignancies, with predominant cardiovascular risks of atrial and ventricular arrhythmias in addition to heart failure.

Proteasome inhibitors used to treat multiple myeloma are associated with an increased risk of hypertension, ischemic events, thromboembolism, and heart failure. The more potent agent, carfilzomib, appears more cardiotoxic than bortezomib. Other treatments for myeloma include immunomodulatory drugs including lenalidomide and thalidomide, which may cause heart failure in addition to risks of venous thromboembolism. Mitogen-activated extracellular signal regulated kinase (MEK) inhibitors used for metastatic melanoma may cause hypertension and cardiomyopathy, especially when co-administered with rapidly accelerated fibrosarcoma (RAF) inhibitors.

The most dramatic toxicity of contemporary cancer therapy results from combined immune checkpoint inhibitors, which block the natural counterregulatory T-cell suppression and unleash potentially fatal inflammation directed toward multiple organs that can include the heart and vessels. These are discussed in **Chap. 268** on noninfectious myocarditis.

Other therapeutic drugs that can cause cardiotoxicity during chronic use include tumor necrosis factor α antagonists for rheumatologic conditions, and carbamazepine, clozapine, and lithium for neurologic and psychiatric diagnoses. Antiretroviral therapies for HIV have been implicated in cardiomyopathy. Chloroquine and hydroxychloroquine, which are widely used for systemic lupus erythematosus and rheumatoid arthritis, can decrease ejection fraction with either restrictive or dilated phenotype, often in association with conduction block. The presumed mechanism of toxicity is impaired lysosomal function, with accumulation of inclusion bodies that can be seen on cardiac biopsy.

Toxic exposures can cause arrhythmias or respiratory injury acutely during accidents. Chronic exposures implicated in cumulative cardiotoxicity include cobalt, arsenicals, lead, and mercury. Treatment for these disorders includes removing exposure to the toxin and standard medical therapy for heart failure with reduced ejection fraction. Cardiomyopathy secondary to cobalt toxicity may be secondary to impaired myocardial energetics and is more common in the setting of hypothyroidism and dietary protein and thiamine deficiency. Cobalt cardiomyopathy usually presents with polycythemia, hypothyroidism, and goiter due to the effects of cobalt on red cell production and thyroxine. Most historical causes of pathologic cobalt exposure are no longer relevant (e.g., treatment of anemia associated with end-stage disease, beer foam stabilization), and recent cases have been attributed to industrial exposures and cobalt alloy prosthetic hips. Diagnosis occurs in the setting of cobalt exposure and can be confirmed by the presence of electron microscopy dense intramitochondrial particles.

PERIPARTUM CARDIOMYOPATHY

Peripartum cardiomyopathy (PPCM) develops during the last trimester or within the first 5–6 months after pregnancy, most commonly within the first 2 weeks after delivery. Between 1:1000 and 1:4000 deliveries in the United States are affected. Risk factors include older maternal age, increased parity, twin pregnancy, malnutrition, and tocolytic therapy for premature labor. Up to half of cases occur in the setting of hypertensive disorders of pregnancy, including preeclampsia. Risk of PPCM is fourfold higher in black women, in whom recovery of normal LVEF takes longer and is less likely than in white women.

Several of the risk factors contribute to antiangiogenic signaling through secreted inhibitors of VEGF, such as soluble FLT1 (sFLT1), high levels of which predict worse outcome. An abnormal cleavage fragment of the nursing hormone prolactin has also been implicated in decreased vascular response during oxidative stress, and investigation is ongoing with the prolactin inhibitor bromocriptine after mixed results of small trials. Multiple other hormonal changes of pregnancy and secretory products from the placenta may interact to cause cardiac dysfunction, suggesting a "vasculohormonal model" in the development of PPCM. Genetic contribution has also been recognized. Similar to other DCM populations, truncating mutations in *TTN* are found in ~15% of cases of PPCM and associated with lower rates of recovered systolic function. Pregnancy thus represents another environmental trigger for accelerated phenotypic expression of genetic cardiomyopathy.

PPCM usually presents with evidence of congestion, and criteria generally include an LVEF ≤0.45 presenting toward the end of pregnancy in the absence of another cardiac diagnosis. Pregnancy can unmask previously unrecognized heart disease but usually does so by the second trimester, when the major circulatory changes have already occurred. Toward the end of pregnancy, edema and dyspnea may be mistakenly attributed to the pregnancy itself, but elevated levels of brain natriuretic peptide or troponin or evidence of elevated central venous pressures are clues to cardiac dysfunction, particularly in the setting of hypertension. Echocardiography is usually sufficient for diagnosis, but in complex cases, cardiac magnetic resonance imaging (MRI) can be considered, but gadolinium should not be used.

Most cases of PPCM present within the first week after delivery, usually with increasing edema and dyspnea when urine output does not keep up with mobilization of fluid. Both atrial and ventricular arrhythmias can occur. It is important to exclude other complications of pregnancy such as pulmonary emboli and coronary artery dissection. Genetic testing should be considered as the results may impact the mother, in whom positive genetic testing predicts less recovery, and also other family members, too.

Initial treatment for PPCM includes loop diuretics as needed to restore normal volume status. Prior to delivery, close collaboration with the maternal-fetal medicine team is necessary to adjust therapies to stabilize the gravid mother while protecting the fetus. Digoxin and beta blockers can be used if needed for arrhythmias, and hydralazine/nitrate combinations can be used for hypertension, but renin-angiotensin system inhibitors should not be given due to adverse fetal effects. Hemodynamic instability may require ongoing hemodynamic monitoring, and plans should be in place for emergency delivery if necessary. When the mother is hemodynamically stable in the postpartum period, metoprolol tartrate, enalapril, and spironolactone have been shown to be compatible with breastfeeding. PPCM with LVEF <0.35 or marked dilation carries increased incidence of left ventricular thrombus and embolic risk, so anticoagulation is usually prescribed for the first 6 weeks once obstetric bleeding has resolved. Breastfeeding was once prohibited but now is generally encouraged in patients in whom fluid balance can be maintained through the high oral fluid intake required. For patients who are not breastfeeding, there is an ongoing large, randomized trial to determine the impact of bromocriptine on PPCM outcomes.

Improvement of LVEF to ≥0.50 occurs in 50–80% of PPCM, often within 6 months, when other cardiac diagnoses have been carefully excluded. Recovery is less likely with LVEF <0.30, left ventricular end-diastolic dimension ≥0.60, black race, and presentation >6 weeks after delivery. Elevated levels of natriuretic peptides, troponin, and sFLT1 have also been associated with less recovery. Patients with LVEF <0.35 have a higher risk of life-threatening arrhythmias during initial presentation and early after discharge, for which consideration of wearable defibrillators is reasonable. This is burdensome both physically and psychologically to a new mother, and more data are needed to help stratify risk. One-year mortality rates after PPCM have ranged in the United States from 4% in one study to 11% in a population of black women, and have been reported as up to twofold higher in Africa.

METABOLIC CAUSES OF CARDIOMYOPATHY

Endocrine disorders affect multiple organ systems, including the heart. *Hyperthyroidism and hypothyroidism* do not often cause clinical heart failure in an otherwise normal heart but commonly exacerbate heart failure. Clinical signs of thyroid disease may be masked, so tests of thyroid function are part of the routine evaluation of cardiomyopathy. Hyperthyroidism should always be considered with new-onset atrial fibrillation or ventricular tachycardia or atrial fibrillation in which the rapid ventricular response is difficult to control. The most common current reason for thyroid abnormalities in the cardiac population is the treatment of tachyarrhythmias with amiodarone, a drug with substantial iodine content. Hypothyroidism should be treated with very slow escalation of thyroid supplements to avoid exacerbating tachyarrhythmias and heart failure. Hyperthyroidism and heart failure create a dangerous combination that merits very close supervision, often hospitalization, during titration of antithyroid medications, during which decompensation of heart failure may occur precipitously and fatally.

Pheochromocytoma is rare but should be considered when a patient has heart failure and very labile blood pressure and heart rate, sometimes with episodic palpitations **(Chap. 399)**. Patients with pheochromocytoma often have postural hypotension. In addition to α-adrenergic receptor antagonists, definitive therapy requires surgical extirpation. Very high renin states, such as those caused by renal artery stenosis, can lead to modest depression in ejection fraction with little or no ventricular dilation and markedly labile symptoms with flash pulmonary edema, related to sudden shifts in vascular tone and intravascular volume.

Controversies remain regarding whether diabetes mellitus and obesity are sufficient to cause cardiomyopathy with reduced ejection fraction. Most heart failure in diabetes mellitus results from epicardial coronary disease, with further increase in coronary artery risk due to accompanying hypertension and renal dysfunction. Cardiomyopathy may result in part from insulin resistance and increased advanced-glycosylation end products, which impair both systolic and diastolic function. However, much of the dysfunction can be attributed to scattered focal ischemia resulting from distal coronary artery tapering and limited microvascular perfusion even without proximal focal stenoses. Diabetes mellitus is a typical factor in heart failure with “preserved” ejection fraction, along with hypertension, advanced age, and female gender.

The existence of a cardiomyopathy due to obesity is generally accepted, but the overlap is not well defined with the syndrome of heart failure with preserved ejection fraction (HFpEF). In addition to cardiac involvement from associated diabetes mellitus, hypertension, and vascular inflammation of the metabolic syndrome, obesity alone is associated with impaired excretion of excess volume load, which, over time, can lead to increased wall stress and secondary adaptive neurohumoral responses. Fluid retention may be aggravated by large fluid intake and the rapid clearance of natriuretic peptides by adipose tissue. In the absence of another obvious cause of cardiomyopathy in an obese patient with systolic dysfunction without marked ventricular dilation, effective weight reduction is often associated with major improvement in ejection fraction and clinical function. Improvement in cardiac function has been described after successful bariatric surgery, although all major surgical therapy poses increased risk for patients with heart failure. Postoperative malabsorption and nutritional deficiencies, such as calcium and phosphate deficiencies, may be particularly deleterious for patients with cardiomyopathy.

Nutritional deficiencies can occasionally cause DCM but are not commonly implicated in developed countries. Beriberi heart disease due to thiamine deficiency can result from poor nutrition in undernourished populations and in patients deriving most of their calories from alcohol and has been reported in teenagers subsisting only on highly processed foods. This disease is initially a vasodilated state with very-high-output heart failure that can later progress to a low-output state; thiamine repletion can lead to prompt recovery of cardiovascular function. Abnormalities in carnitine metabolism can cause dilated or restrictive cardiomyopathies, usually in children. Deficiency of trace elements such as selenium can cause cardiomyopathy (Keshan’s disease).

Calcium is essential for excitation-contraction coupling. Chronic deficiencies of calcium, such as can occur with hypoparathyroidism (particularly postsurgical) or intestinal dysfunction (from diarrheal syndromes and following extensive resection), can cause severe chronic heart failure that responds over days or weeks to vigorous calcium repletion. Phosphate is a component of high-energy compounds needed for efficient energy transfer and multiple signaling pathways. Hypophosphatemia can develop during starvation and early refeeding following a prolonged fast and occasionally during hyperalimentation.

Hemochromatosis is variably classified as a metabolic or storage disease **(Chap. 426)**. It is included among the causes of restrictive cardiomyopathy, but the clinical presentation is often that of a DCM. The autosomal recessive form is related to the *HFE* gene. With up to 10% of the population heterozygous for one mutation, the clinical prevalence might be as high as 1 in 500. The lower observed rates highlight the limited penetrance of the disease, suggesting the role of additional genetic and environmental factors such as alcoholism affecting clinical expression. Cardiac siderosis can also be acquired from iron overload due to hemoglobinopathies in patients treated with recurrent transfusions. Excess iron is deposited in the perinuclear compartment of cardiomyocytes, with resulting disruption of intracellular architecture and

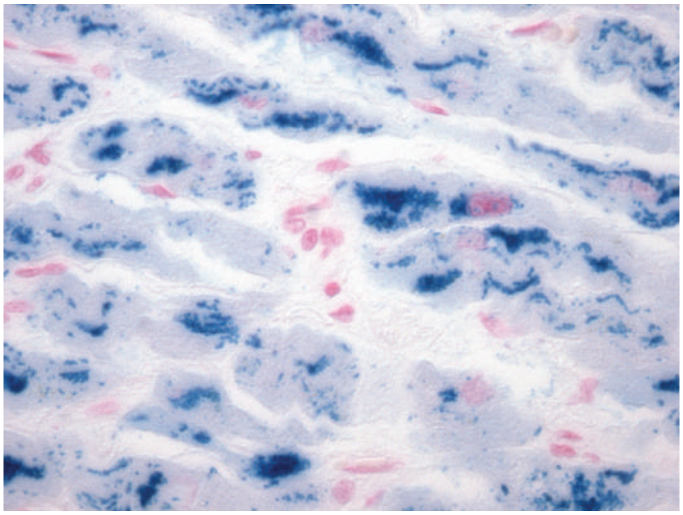

FIGURE 269-1 Hemochromatosis. Microscopic image of an endomyocardial biopsy showing extensive iron deposition within the cardiac myocytes with the Prussian blue stain (400× original magnification). *(Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.)*

mitochondrial function. A diagnosis of systemic iron overload is made from measurement of serum iron and transferrin saturation, with a threshold of >60% for men and >45–50% for women. MRI is used to quantitate iron stores in the liver and heart. While endomyocardial biopsy tissue can be stained for iron **(see Chap. 270 and Fig. 269-1)**, a diagnosis of cardiac iron overload is made principally by MRI and biopsy is not usually needed. If diagnosed early, hemochromatosis can often be managed by repeated phlebotomy to remove iron. For more severe iron overload, iron chelation therapy with desferrioxamine (deferoxamine) or deferasirox can help to improve cardiac function if myocyte loss and replacement fibrosis are not too severe.

Inborn disorders of metabolism occasionally present with DCM, although they are most often associated with restrictive cardiomyopathy **(Chap. 267, Table 267-1)**.

TAKOTSUBO SYNDROME OR CARDIOMYOPATHY

Apical ballooning or "takotsubo" syndrome, also referred to as stress-induced cardiomyopathy or "broken heart syndrome," is often classified as a cardiomyopathy, although its distinct acute presentation, typical ventricular shape, and frequent rapid recovery differ from most other cardiomyopathies. It occurs typically in older women after sudden intense emotional or physical stress. The ventricle shows global ventricular dilation with basal contraction, forming the shape of the narrow-necked jar (*takotsubo*) used in Japan to trap octopuses. Originally described in Japan, it is well recognized elsewhere during emergency cardiac catheterization and intensive care unit admissions for noncardiac conditions. Presentations include pulmonary edema, hypotension, and chest pain with electrocardiogram (ECG) changes mimicking an acute infarction. The left ventricular dysfunction extends beyond a specific coronary artery distribution and generally resolves within days to weeks. Animal models and ventricular biopsies suggest that this acute cardiomyopathy may result from intense sympathetic activation with heterogeneity of myocardial autonomic innervation, diffuse microvascular spasm, and/or direct catecholamine toxicity. Cardiac MRI **(Fig. 269-2)** demonstrates diffuse myocardial edema without necrosis and abnormal myocardial calcium handling. Coronary angiography may be required to rule out acute coronary occlusion.

A similar picture to takotsubo can also be caused a coronary embolus in the absence of atherosclerotic coronary artery disease. No therapies have been proven beneficial, but reasonable strategies include nitrates for pulmonary edema; combined alpha and beta blockers rather than selective beta blockade if hemodynamically stable; and magnesium for arrhythmias related to QT prolongation. An intra-aortic balloon pump is occasionally employed to improve critically low cardiac output, but only if there is no left ventricular outflow tract obstruction. The long-term prognosis is generally good, with the lowest mortality associated with episodes triggered by emotional rather than physical triggers. In-hospital complications and mortality are similar to those with acute myocardial infarction. Recurrence occurs in 10% of patients at an estimated rate of 2%/year.

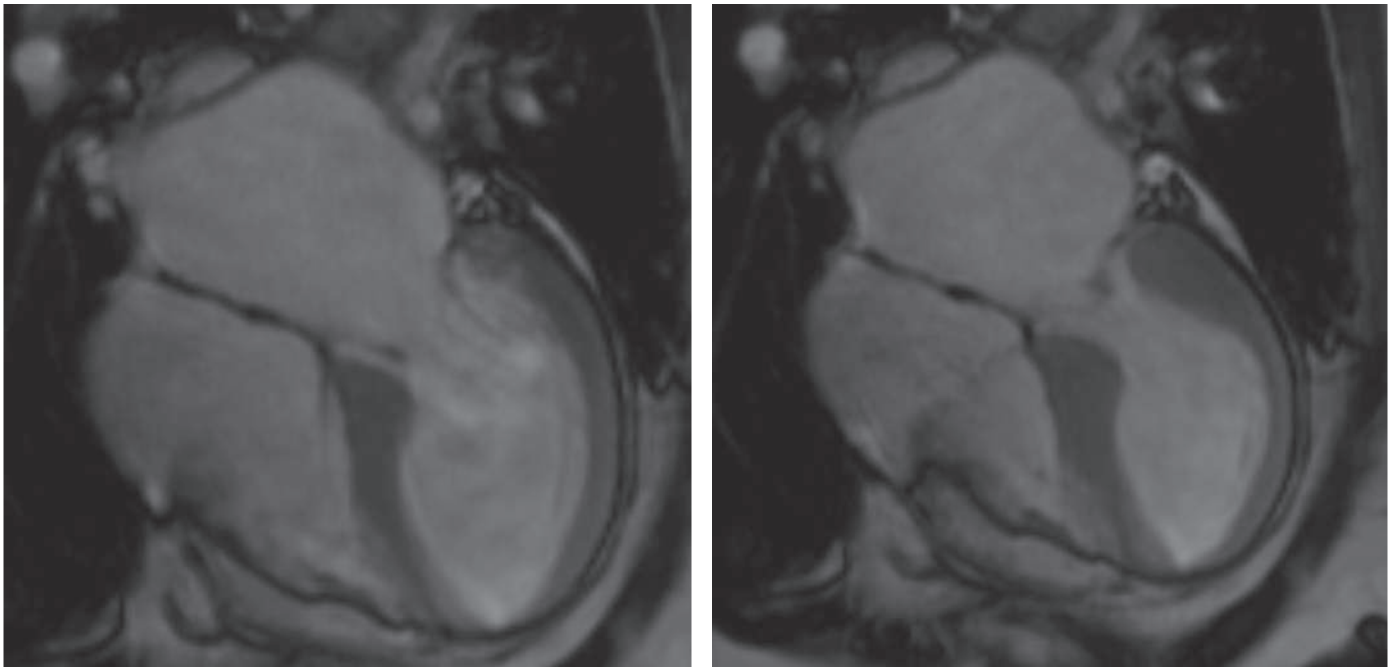

FIGURE 269-2 Takotsubo cardiomyopathy. Four-chamber view of cardiac magnetic resonance imaging demonstrating a mildly dilated left ventricle in diastole (*left panel*) with diffuse hypokinesis of the mid and apical segments and relative sparing of the basal segment wall motion at end systole (*right panel*) with left ventricular ejection fraction 46%. *(Image courtesy of Raymond Kwong, MD, and Zariyat M. Mannan, MD, Cardiovascular Imaging, Brigham and Women's Hospital, Boston.)*

FURTHER READING

Arany Z: Peripartum cardiomyopathy. N Engl J Med 390:154, 2024.

Cardinale D et al: Early detection of anthracycline cardiotoxicity and improvement with heart failure therapy. Circulation 131:1981, 2015.

Davis MB et al: Peripartum cardiomyopathy. J Am Coll Cardiol 75: 207, 2020.

Lyon AR et al: 2022 ESC Guidelines on cardio-oncology developed in collaboration with the European Hematology Association (EHA), the European Society for Therapeutic Radiology and Oncology (ESTRO) and the International Cardio-Oncology Society (IC-OS): Developed by the task force on cardio-oncology of the European Society of Cardiology (ESC). Eur Heart J 43:4229, 2022.

Packer M: Cobalt cardiomyopathy: A critical reappraisal in light of a recent resurgence. Circ Heart Fail 9:e003604. 2016.

Rasoul D et al: Alcohol and heart failure. Eur Cardiol 18:e65, 2023.

Quiroz-Aldave JE et al: Thyrotoxic cardiomyopathy: State of the art. touchREV Endocrinol 19:78, 2023.

Singh T et al: Manganese-enhanced magnetic resonance imaging in takotsubo syndrome. Circulation 146:1823, 2022.

Wood JC: Guidelines for quantifying iron overload. Hematology Am Soc Hematol Educ Program 1:210, 2014.

270 Amyloidosis and Other Restrictive Cardiomyopathies

Rodney H. Falk, Neal K. Lakdawala, Lynne Warner Stevenson, Joseph Loscalzo

RESTRICTIVE CARDIOMYOPATHY

Restrictive cardiomyopathy (RCM) is dominated by abnormal diastolic function within a noncompliant ventricle, often with mildly decreased contractility and ejection fraction (usually 30–50%). Both atria are enlarged, sometimes massively. Mild left ventricular dilation can be present. End-diastolic pressures are usually elevated in both ventricles, with preservation of resting cardiac output until late in the disease. Subtle exercise intolerance is usually the first symptom but is often not recognized until after clinical presentation with congestive symptoms. The restrictive diseases often present with relatively more right-sided symptoms, such as edema, abdominal discomfort, and ascites, although filling pressures are elevated in both ventricles. The cardiac impulse is less displaced than in dilated cardiomyopathy and less dynamic than in hypertrophic cardiomyopathy. A fourth heart sound is more common than a third heart sound in sinus rhythm, but atrial fibrillation is common. Jugular venous pressures often show rapid Y descents and may increase during inspiration (positive Kussmaul sign).

Most causes of RCM are "infiltrative," due to infiltration of abnormal substances between myocytes, with the amyloidoses being the most common causes of RCM. RCM can also result from storage of abnormal metabolic products within myocytes. RCM due to acquired fibrotic injury is most commonly caused by radiation or connective tissue disease and less commonly by hypereosinophilic injury **(see Chap. 268 and Table 270-1)**.

RCM can also be secondary to genetic causes of primary cardiomyopathy, including variants in *DES* or as a forme fruste of hypertrophic cardiomyopathy caused by sarcomere variants. The most common differential diagnosis is between RCM and constrictive pericardial disease, as both often present with dominant right-sided heart failure.

In the absence of specific therapy for the etiology of RCM, such as is now available for some patients with amyloidosis, general strategies are focused on maintenance of optimal volume status, which is generally a compromise between adequate ventricular filling and systemic decongestion. As ventricular filling and stroke volume are often restricted, therapy with beta blockers can reduce heart rate, contractility, cardiac output, and functional reserve. Vasodilation may not be well tolerated. Other than diuretics, therapies proven effective for dilated cardiomyopathy and other heart failure with reduced ejection fraction are not recommended for routine use in RCM.

TABLE 270-1 Causes of Restrictive Cardiomyopathies (RCM)

Infiltrative (Between Myocytes)
Amyloidosis
Light chain (AL) amyloid
Familial (variant transthyretin)[a]
Wild-type (normal) transthyretin
Inherited metabolic defects[a]
Storage (Within Myocytes)
Hemochromatosis (iron),[a] also with dilated cardiomyopathy phenotype when advanced
Inherited metabolic defects[a]
Fabry's disease
Glycogen storage disease (II, III)
Fibrotic
Radiation
Scleroderma
Endomyocardial
Possibly related fibrotic diseases
Tropical endomyocardial fibrosis
Hypereosinophilic syndrome (Löffler's endocarditis)
Carcinoid syndrome
Radiation
Drugs: e.g., serotonin, ergotamine
Genetic Variants Affecting Cardiomyocyte Function
Occasional RCM due to genetic variants more commonly associated with dilated or hypertrophic cardiomyopathy
RCM and skeletal muscle involvement with desminopathy due to pathogenic *DES* variants

[a] Contribution of genetic background.

AMYLOIDOSIS

Amyloidosis is a systemic disease characterized by the deposition of fibrillar proteinaceous material primarily in the extracellular space of one or more organs. Although there are many precursor proteins that have been shown to cause amyloidosis **(see Chap. 117)**, the unifying feature of all types of amyloid is its staining characteristics. Amyloid is a noncellular material deposited in the extracellular space that stains with Congo red and exhibits "apple-green birefringence" when viewed under polarized light microscopy. When viewed with electron microscopy, amyloid deposits are seen to consist of nonbranching fibrils ~10 nm in diameter and a few micrometers in length. From a cardiac standpoint, there are two main precursor proteins that cause most cases of amyloid cardiomyopathy: transthyretin (TTR), produced in the liver, and light chains produced by abnormal plasma cells. TTR amyloidosis may be caused by a genetic variant of TTR (ATTRv) or by wild-type TTR (ATTRwt), and variant TTR is generally considered to be less stable than wild-type TTR. Once a diagnosis of amyloidosis has been made, it is critically important to determine the precise precursor protein because treatment is specific to the type of amyloid and incorrect typing will lead to inappropriate therapy.

PATHOLOGIC MECHANISMS OF HEART FAILURE IN CARDIAC AMYLOIDOSIS

All forms of cardiac amyloidosis have a similar appearance on gross and microscopic examination **(Figs. 270-1 and 270-2)**.

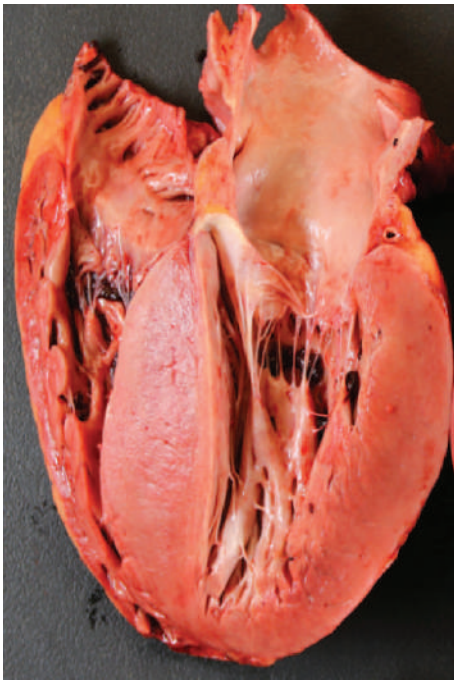

FIGURE 270-1 Gross pathology of the heart in amyloid cardiomyopathy. The left ventricle (LV) is severely thick due to amyloid infiltration. The LV cavity is small with markedly thick walls due to amyloid infiltration. The right ventricle is mildly thickened, and the left atrium (partly excised) is dilated. *(Courtesy of Dr. Richard Mitchell, Brigham and Women's Hospital Department of Pathology.)*

The extracellular deposits of amyloid result in an increase in the mass of the heart, due to expansion of the extracellular space. Both left and right ventricular walls demonstrate increased thickness, and decreased compliance and the left ventricular cavity may be relatively small, leading to a restrictive pathophysiology. The right ventricular cavity is often normal in size but may dilate in advanced disease. The atria are usually dilated due to the chronic elevation of biventricular diastolic filling pressure and may dilate further with the onset of atrial arrhythmias. Histologically, amyloid deposits are widespread throughout the heart so that endomyocardial biopsy is almost always diagnostic in cases of cardiac amyloidosis.

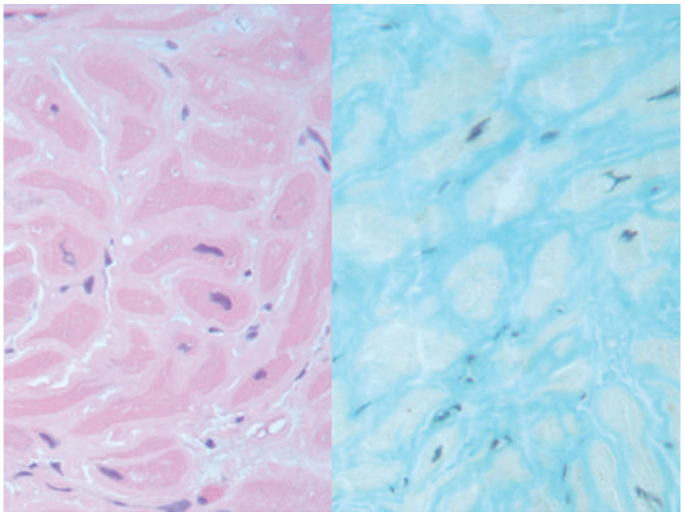

FIGURE 270-2 Histology of amyloid cardiomyopathy. *Left panel:* Hematoxylin and eosin stain showing extensive extracellular cardiac amyloid deposition resulting in separation and distortion of cardiomyocytes. The amyloid stains light pink and the myocytes are nucleated and stain a deeper pink. *Right panel:* Specimen from same patient stained with sulfated Alcian blue. The amyloid stains "sea green," and the myocytes are pale yellow-gray. Magnification ×60. *(Courtesy of Dr. Richard Mitchell, Brigham and Women's Hospital Department of Pathology.)*

CLINICAL FEATURES OF CARDIAC AMYLOIDOSIS

There are some clinical features of light chain (AL) amyloidosis and TTR amyloidosis that suggest one or the other diagnosis, but the cardiac features are similar enough that the final diagnosis of amyloid type often relies upon precise typing of a cardiac or extracardiac tissue biopsy. The most common clinical presentation is congestive heart failure. Although both ventricles are usually heavily infiltrated by the time congestive heart failure occurs, patients tend to present with symptoms of predominant right heart failure, namely peripheral edema and, not infrequently, ascites. Despite the clinical predominance of right ventricular failure, cardiac catheterization will reveal considerable elevation of both left and right ventricular filling pressure. Atrial arrhythmia, particularly atrial fibrillation, is a common presenting feature, particularly in TTR amyloidosis, and may occur at a time when the left ventricle is only mildly thick, precipitating heart failure in a previously asymptomatic patient. The onset of atrial fibrillation may be associated with symptoms of dyspnea, often more severe than is seen in nonamyloid patients with this arrhythmia. Refractory atrial arrhythmia, whether to antiarrhythmic drugs or characterized by failure to maintain sinus rhythm after repeated ablation procedures, should raise the possibility of TTR cardiac amyloidosis, particularly if occurring in a patient in the mid-60s or older without other apparent reason for the atrial fibrillation.

Cardiovascular physical examination in the symptomatic patient with cardiac amyloidosis reveals a normal or low pulse pressure. The apex beat may be difficult to palpate but, if felt, is usually only mildly displaced. The jugular venous pressure is almost always elevated and frequently demonstrates a paradoxical rise on inspiration (Kussmaul sign). The first and second heart sounds tend to be normal. Despite congestive heart failure, a third heart sound is very rarely heard, due to the restrictive pathophysiology and its associated impairment of ventricular relaxation. Similarly, a fourth heart sound is also unusual as the infiltrated atrium is rarely able to generate much contractile pressure. Pleural effusions, primarily caused by heart failure (but occasionally aggravated by associated hypoalbuminemia from nephrotic syndrome or by the presence of pleural amyloidosis) are quite common. Cardiac murmurs due to amyloid-associated valve disease are relatively uncommon, although a tricuspid regurgitation murmur may occasionally be present. There is an association between aortic valve stenosis and TTR amyloidosis, and a murmur of aortic stenosis may be present. Whether amyloid causes aortic stenosis or simply aggravates or precipitates heart failure in a patient with unrelated aortic valve disease is unclear. In addition to the cardiovascular examination, a careful systemic physical examination is mandatory. In advanced heart failure, hepatomegaly due to congestive heart failure is common, but an unusually hard liver may represent amyloid infiltration of the liver—something that occurs in AL amyloidosis but not TTR. The blood pressure should be checked in the supine, seated, and standing positions, seeking an abnormal postural drop. If present, this likely represents associated autonomic neuropathy, which is a feature of AL amyloidosis and some forms of variant TTR amyloidosis but not wild-type TTR amyloidosis. Macroglossia or periorbital bruising, if present, strongly suggest AL amyloidosis, whereas a history of bilateral carpal tunnel syndrome or the finding of a ruptured biceps tendon points toward TTR amyloidosis.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of a patient with biventricular failure manifesting predominantly as edema, in association with a thick left ventricle but without a history of severe hypertension, is small. Hypertensive heart disease is usually associated with left ventricular hypertrophy on the electrocardiogram (ECG), whereas this ECG finding is uncommon in amyloidosis, and it requires several years of severe uncontrolled hypertension to produce the degree of wall thickening equal to that seen in amyloidosis. While amyloidosis, particularly ATTR, can have asymmetric left ventricular thickening, the clinical distinction between hypertrophic cardiomyopathy and amyloidosis is usually fairly clear, since the former rarely presents with peripheral edema and usually has left ventricular hypertrophy (LVH) on the ECG. Other infiltrative cardiomyopathies such as Fabry disease may have typical skin manifestations and usually LVH on the ECG. The rare mitochondrial cardiomyopathies may have an echocardiographic appearance similar to amyloidosis but have different noncardiac clinical features, such as maternally inherited diabetes, deafness, and recurrent strokes. Patients with advanced diabetes may have heavy proteinuria and heart failure, but their history of diabetes is obvious, and the echocardiogram does not show features suggestive of amyloid heart disease. In the era before echocardiography, constrictive pericarditis, often presenting with edema, ascites, and elevated jugular venous pressure with a Kussmaul sign, was the major differential of amyloid heart disease. Constrictive pericarditis is less common nowadays, and echocardiography can easily distinguish between the two entities.

AL AMYLOIDOSIS OF THE HEART

AL amyloidosis is a plasma cell disorder closely related to multiple myeloma. It is described in detail in **Chap. 117**. An excessive production of abnormal lambda, or (less commonly) kappa, free light chains produced by abnormal clonal bone marrow plasma cells results in amyloid deposits in multiple organs, with heart and kidney being the most common organs involved. AL amyloidosis is closely related to multiple myeloma, with which it may occasionally overlap. It is an aggressive disease, and cardiac involvement, either alone or in combination with other major organ involvement, carries the worst survival, with a median survival from onset of heart failure to death in untreated patients of ~6 months. This short survival underscores the urgency of pursuing a diagnosis of AL amyloidosis once suspected, as therapy can markedly improve duration and quality of life and must be instituted as soon as the diagnosis is made. In addition to the general features of heart failure noted above, there are several clinical noncardiac clues to AL amyloidosis as the etiology of congestive heart failure. These indicate the presence of a multisystem disease and include significant proteinuria, neuropathy, periorbital purpura (virtually pathognomonic of AL amyloidosis and occurring in 20–30% of cases), and macroglossia (~10% of cases).

TTR AMYLOIDOSIS OF THE HEART

Patients with wild-type TTR amyloidosis cardiomyopathy and those with amyloid cardiomyopathy caused by variant TTR may have subtle differences in presentation. The less common variant TTR cardiomyopathy may be associated with amyloid neuropathy, which can vary between mild sensory neuropathy and a rapidly progressive severe sensorimotor neuropathy with or without autonomic nervous system involvement. Myocardial infiltration may be extensive in patients with severe variant transthyretin neuropathy, but cardiac symptoms can be relatively mild in these patients, either because the limitation of physical activity due to the neuropathy masks exercise intolerance or because autonomic neuropathy acts to reduce peripheral resistance, decreasing the work of the heart. If autonomic neuropathy is present, postural hypotension can be highly symptomatic and postural measurements of blood pressure are mandatory. A family history of neuropathy or cardiomyopathy may be elicited in these patients and, when neuropathy is present, may have been misdiagnosed as other neurologic conditions. In the United States, the most common TTR variant causing TTR amyloid cardiomyopathy is the substitution of valine for isoleucine at position 122, Val122Ile. This variant is present in ~3.5% of the U.S. population of African descent and presents as a late-onset restrictive cardiomyopathy usually in the seventh decade onward. The penetrance of clinical disease is incomplete, and the family history is often lacking or unknown. Neuropathy in this mutation is very mild or absent, although carpal tunnel syndrome, often preceding the onset of heart failure by several years, is commonly present. The clinical features of other forms of variant TTR amyloidosis can vary considerably, somewhat dependent on the transthyretin mutation, but in the case of the Val122Ile variant, it tends to be a relatively rapid progressive infiltrative cardiomyopathy with a median survival of 2–3 years following the onset of heart failure in untreated patients.

Once considered a rare form of cardiac amyloidosis, TTR deposition derived from native (wild-type) transthyretin is now considered the most common form of amyloidosis. It affects mainly men in the latter half of their seventh decade and eighth decades but may occasionally present at a younger age and often presents in the ninth decade onward. Men are almost 20 times more likely than women to be diagnosed with wild-type TTR amyloidosis, and when it does occur in women, it tends to present at a later age and to have a more indolent course. On average, the untreated median survival from the onset of heart failure to death is around 4–5 years, with death being due primarily to progressive congestive heart failure rather than to sudden death. Current therapies have slowed the disease progression and improved survival. Both wild-type and mutant TTR cardiomyopathy have a subclinical phase, probably of several years, during which there is a progressive infiltration of the heart with amyloid, and by the time heart failure occurs, 30–50% of the involved heart weight is due to amyloid. Atrial arrhythmias, most commonly atrial fibrillation or atrial flutter, may be the presenting feature of the disease, and when they occur, they often precipitate or worsen heart failure. In contrast to AL amyloidosis, the heart in ATTRwt is the sole major organ to be clinically involved, although histologic examination of the lungs and gut often shows fairly extensive amyloid deposits. TTR also had a predilection for ligaments and tendons, and a history of carpal tunnel syndrome, spinal stenosis, or ruptured bicep tendon, often occurring 5–8 years before heart failure, can be elicited in about half the patients with this disorder.

DIAGNOSIS

The diagnosis of cardiac amyloidosis rests upon a clinical suspicion combined with imaging findings and, in many cases, a confirmatory biopsy. ECG voltage may be abnormally low, particularly in AL amyloidosis, but this is not always the case. Echocardiography is an extremely useful imaging modality. The echocardiogram in all forms of cardiac amyloidosis with heart failure shows increased left ventricular wall thickness without ventricular dilation and, frequently, biatrial enlargement **(Fig. 270-3)**.

The right ventricle may also show increased thickness of the free wall. Diastolic function is abnormal, and a restrictive pattern, consistent with elevated left ventricular diastolic pressures, is often seen. Although the left ventricular ejection fraction may be normal or near-normal, left ventricular longitudinal strain imaging, a tool for measuring the contraction of the left ventricle in the longitudinal plane, often shows a pattern in which strain is relatively preserved at the left ventricular apex and severely impaired at the base of the heart. Color-coding of this pattern results in a "bull's-eye" appearance, which is highly suggestive of cardiac amyloidosis **(Fig. 270-4)**.

Atrial function is often impaired, particularly in AL amyloidosis, and this is characterized by a very small transmitral A wave in patients

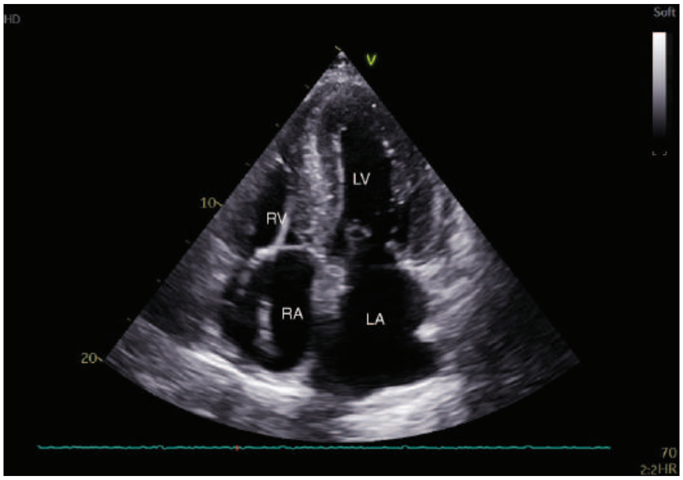

FIGURE 270-3 Echocardiogram (apical four-chamber view) showing typical appearance of amyloidosis. The left ventricular cavity is normal in size with thick walls, and there is biatrial enlargement. LA, left atrium; LV, left ventricle; RA, right atrium; RV, right ventricle.

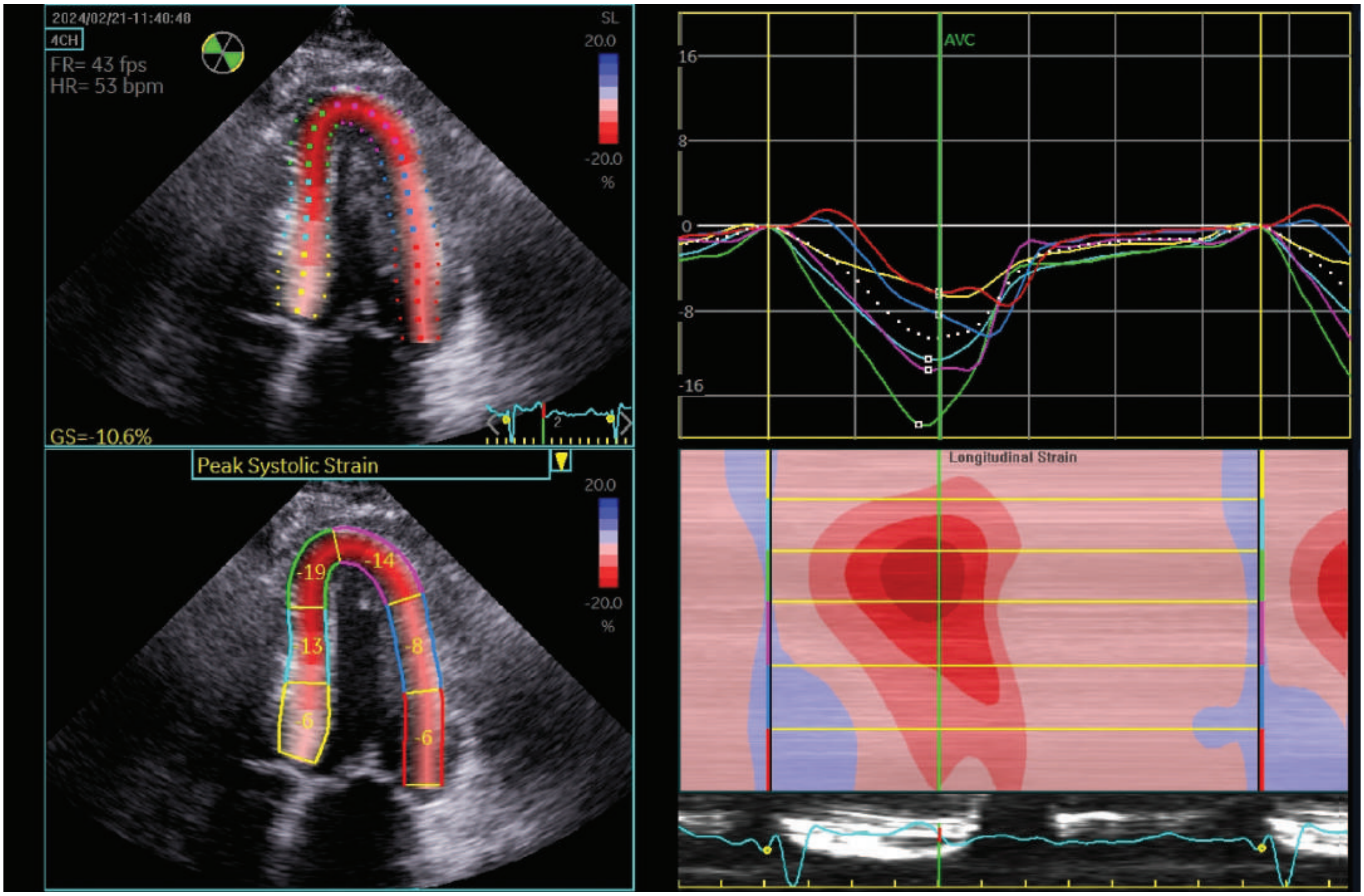

FIGURE 270-4 Left ventricular strain in amyloidosis. *Left panels* show outline of left ventricle divided into seven segments with numbers in bottom panel representing the percentage longitudinal systolic shortening of each segment. *Right top panel* shows corresponding curves for each segment, and *bottom right panel* shows color coding in which the apex codes bright red and base pink due to near-normal apical contraction, resulting in typical "bull's-eye" pattern of amyloidosis.

in sinus rhythm. This appearance may be associated with an increased risk of thromboembolism.

Echocardiography with the typical appearance may be enough to proceed to further laboratory evaluation for amyloidosis, but cardiac magnetic resonance is increasingly used in the diagnosis of suspected cardiomyopathy. The ventricle is seen to be abnormally thick, measurement of extracellular volume demonstrates marked expansion of the extracellular space due to amyloid deposition, and there is often extensive delayed gadolinium enhancement of the myocardium, associated with abnormal nulling of the images **(Figs. 270-5 and 270-6)**.

Taken together, these features are virtually pathognomonic of amyloid cardiomyopathy. Once amyloid cardiomyopathy is suspected, it is critical to determine the specific type of amyloid deposition. Tracers developed for bone imaging (technetium pyrophosphate or technetium 3,3-diphosphono-1,2-propanodicarboxylic acid [DPD]) are not usually taken up by the normal heart or in other forms of cardiac pathology, but there may be avid tracer uptake in patients with transthyretin amyloid cardiomyopathy **(Fig. 270-7)**.

Because patients with AL amyloid cardiomyopathy may occasionally have myocardial uptake of these tracers, it is critically important to measure serum free light chains, serum and urine protein electrophoresis, and serum and urine immunofixation to rule out a plasma cell dyscrasia. If these lab tests are all negative and the nuclear scan is positive, a diagnosis of transthyretin amyloidosis can be confidently made, provided that echocardiography or cardiac magnetic resonance is consistent with this diagnosis. Monoclonal gammopathy of unknown significance not uncommonly coexists with transthyretin amyloidosis and will be associated with abnormalities in testing for plasma cell dyscrasia. Thus, if light chain amyloidosis remains a possibility, despite the presence of a strongly positive pyrophosphate scan, consultation with a hematologist is important to help clarify the diagnosis. If uncertainty remains, an endomyocardial biopsy (or biopsy from other tissue) with definitive tissue typing (ideally by mass spectrometry) is required.

Almost all patients with light chain amyloidosis will have evidence of a plasma cell dyscrasia, and almost all of them will have elevation of either serum free lambda light chains or (less commonly) serum free kappa light chains. If light chain amyloidosis is clinically suspected and serum free light chains are elevated, it is mandatory to obtain a tissue biopsy to confirm the presence of amyloid deposits. Most patients with light chain amyloidosis have widespread systemic involvement, and a subcutaneous fat pad aspirate or deep skin biopsy will show amyloid deposits in the majority of cases, thus avoiding the necessity for a cardiac biopsy. If skin/fat biopsy is negative, it is reasonable to proceed to an endomyocardial biopsy. If amyloid heart disease is present, this will be positive in well over 90% of cases. In all cases, appropriate typing of the amyloid should be done to confirm the appropriate therapy. An algorithm with a suggested workup of suspected amyloidosis is shown in **Fig. 270-8**.

DIAGNOSIS AND TYPING OF CARDIAC AMYLOIDOSIS

Laboratory evaluation in patients with cardiac amyloidosis is generally nonspecific but may give some support to a suspected diagnosis. Electrolytes and complete blood count are usually normal or near normal. N-terminal prohormone of brain natriuretic peptide (NTproBNP) may be higher than anticipated for the degree of congestive heart failure, often in the range of 1000 to 2000 pg/mL even in the presence of relatively mild congestive heart failure. High-sensitivity troponin is frequently mildly elevated and may lead to a misdiagnosis of ischemic heart disease. In AL amyloidosis with renal involvement, heavy

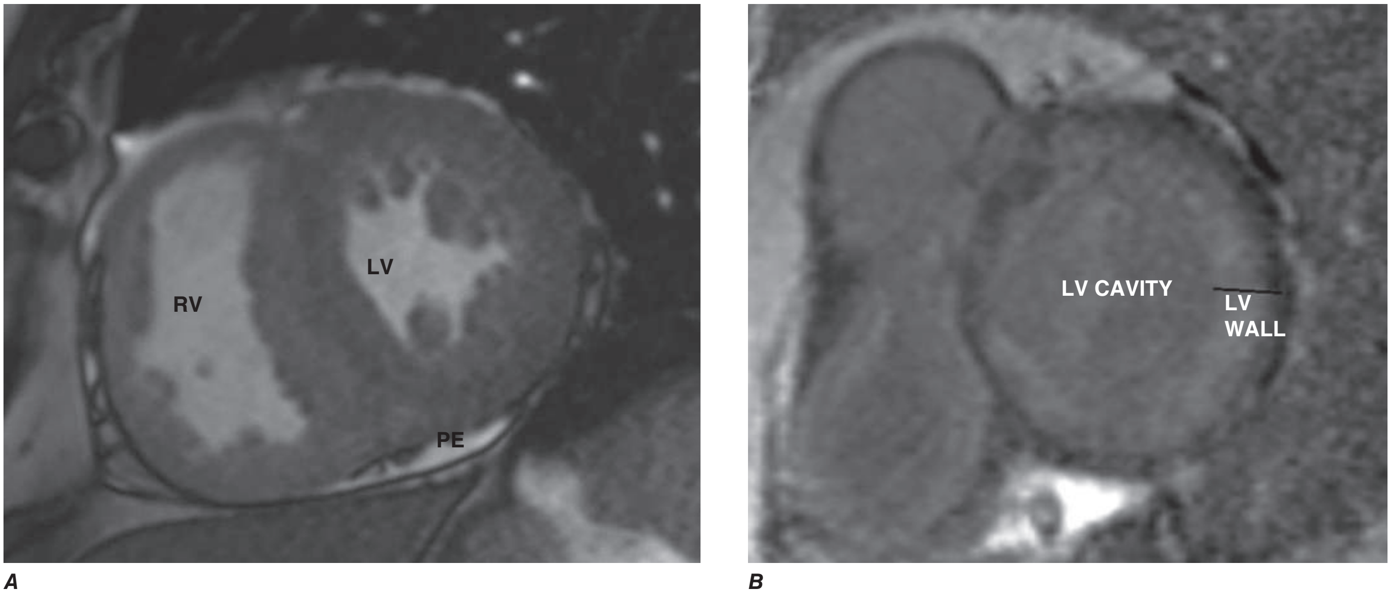

FIGURE 270-5 Typical cardiac magnetic resonance imaging (MRI) in amyloidosis. ***A.*** A typical cardiac MRI appearance in amyloid cardiomyopathy. The left ventricular (LV) cavity is small with a mildly dilated right ventricular (RV) cavity and thick LV walls. ***B.*** The appearance of late gadolinium enhancement after injection of gadolinium in the same patient. Two typical appearances are seen: the distinction between myocardium (*straight line*) and LV cavity is poor, due to delayed nulling of the myocardium, a near pathognomonic feature of cardiac amyloidosis. There is also extensive, transmural delayed gadolinium myocardial uptake due to severe amyloid infiltration. Another aspect of cardiac amyloidosis quantifiable by MRI (not shown) is measurement of extracellular volume, which is markedly increased in cardiac amyloidosis and helps to differentiate amyloid infiltration from true LV hypertrophy. PE, pericardial effusion. *(Image courtesy of Dr. Sarah A. M. Cuddy, Brigham and Women's Hospital, Section of Cardiology.)*

proteinuria is often found, and this may cause hypoalbuminemia. Renal involvement is not a feature of transthyretin amyloidosis. Rarely, in AL amyloidosis with hepatic involvement, the alkaline phosphatase is significantly elevated. Although AL amyloidosis is a plasma cell disorder closely related to multiple myeloma, the sedimentation rate is usually normal or only minimally elevated, unlike the marked elevation that may be seen in myeloma. Hypogammaglobulinemia may be present in AL amyloidosis, but serum and urine protein electrophoresis often do not reveal a monoclonal gammopathy, as the paraprotein level is low. Thus, in suspected AL amyloidosis, serum and urine immunofixation should be performed as these are more sensitive (although nonquantitative) than serum or urine protein electrophoresis. In almost all patients with AL amyloidosis, the level of circulating serum free light chains, either lambda or, less commonly, kappa free light chains, is elevated. It is important to recognize that findings suggestive of a plasma cell dyscrasia in a patient with a cardiomyopathy needs to be interpreted in the context of the clinical picture, as monoclonal gammopathy of unknown significance is common in the older population and may be unrelated to the cardiomyopathy. Unlike TTR cardiomyopathy, there is no definitive noninvasive test to diagnose AL amyloid cardiomyopathy, and thus a tissue biopsy is always needed. A subcutaneous fat pad aspirate or deep-skin needle aspiration biopsy has a high yield of positivity, but needs to be evaluated by a skilled pathologist. A biopsy positive for AL amyloid from a noncardiac site in the presence of imaging consistent with amyloid cardiomyopathy is adequate to conclude that cardiac involvement is present. Immunohistochemistry, performed by many pathology labs, is fraught with diagnostic uncertainty and should be used with great

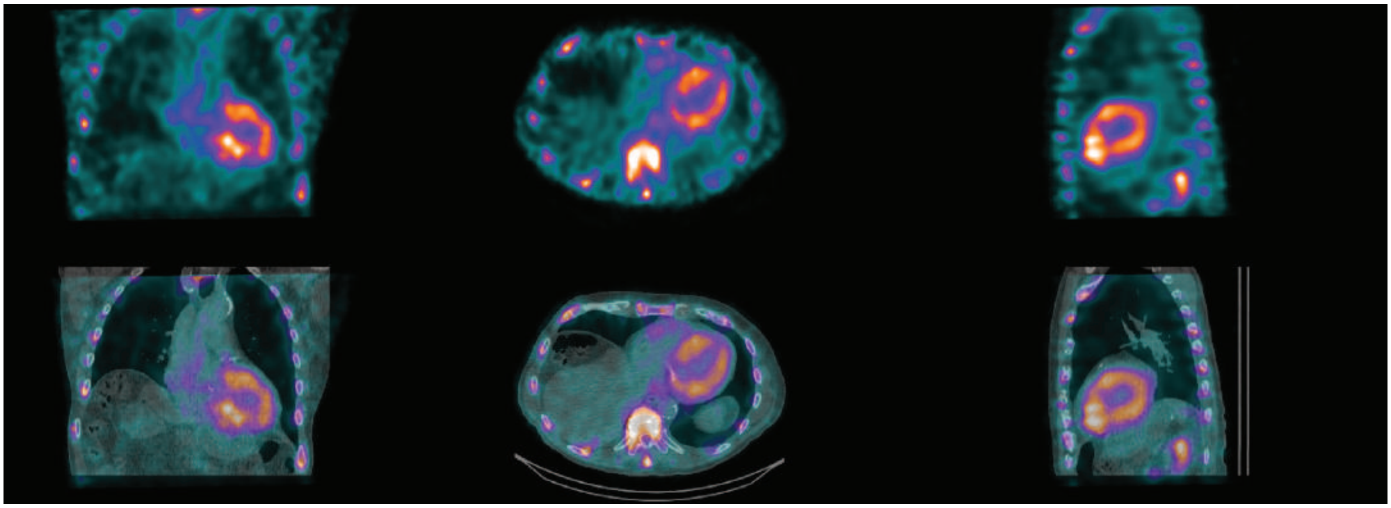

FIGURE 270-6 Technetium pyrophosphate (PYP) scan in a patient with wild-type transthyretin amyloidosis. The top three images are obtained in three separate views with single-photon emission computed tomography (CT) imaging alone, and the bottom three views with added cardiac CT imaging. Both modalities show marked myocardial uptake (*orange*) with absence of blood pool uptake (central clearing) and normal bone uptake. The lower images more precisely localize the uptake to the fused CT images of the heart, which in more equivocal cases helps to localize uptake to the myocardium and prevents misinterpretation as a positive scan due to blood pool uptake without myocardial uptake. PYP uptake is highly sensitive and specific for transthyretin amyloidosis but can occasional be present in AL amyloidosis; thus, exclusion of a plasma cell dyscrasia is mandatory when interpreting this imaging. In the normal heart and almost all other pathologies, there is no myocardial uptake of bone tracers such as PYP.

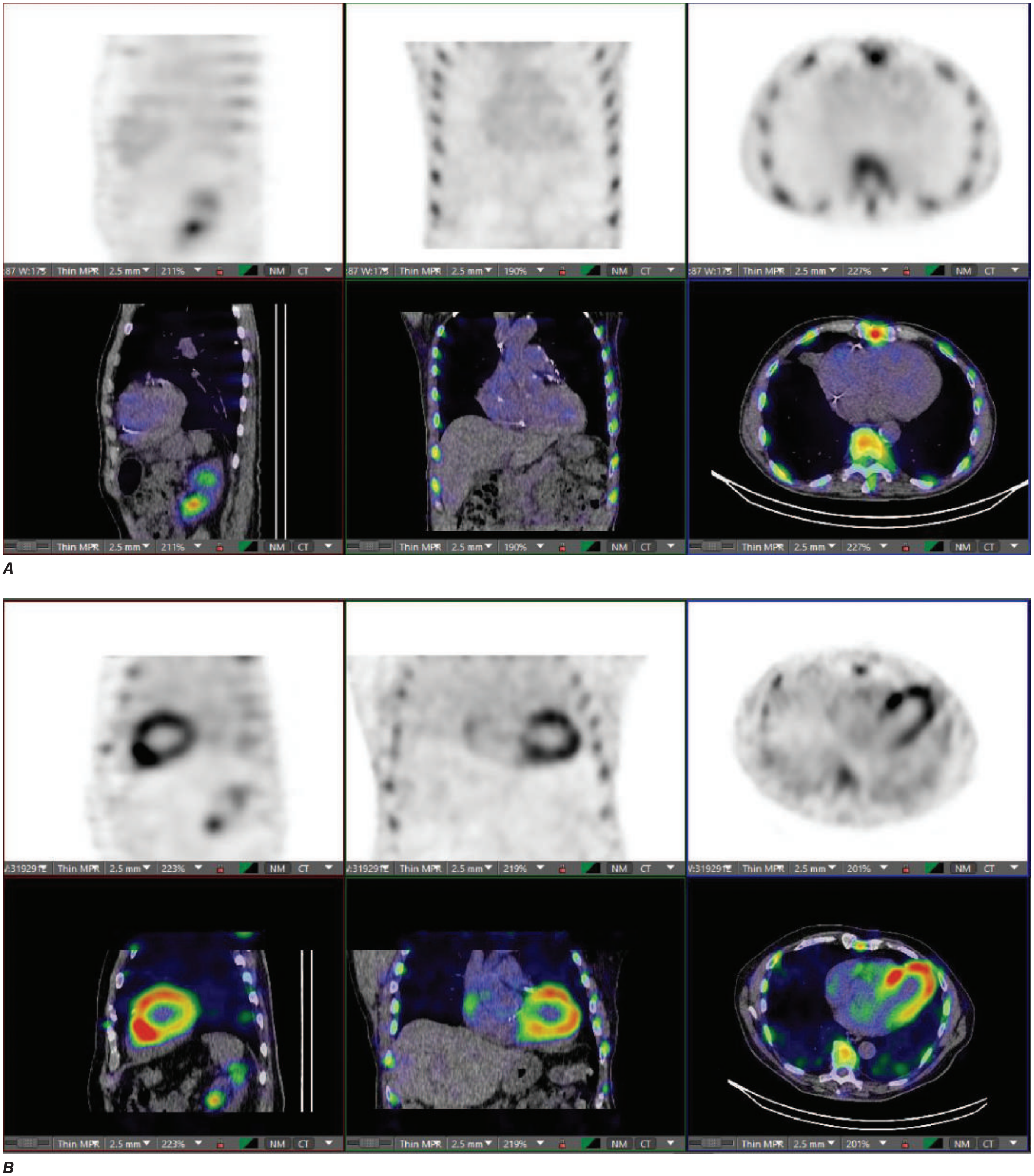

FIGURE 270-7 Technetium pyrophosphate (PYP) scan. *A.* A normal (negative) PYP scan. The planar images (*top*) show rib and sternal isotope uptake, but there is no uptake in the heart. *Bottom* panel shows single-photon emission computed tomography–computed tomography (SPECT-CT) fusion showing normal appearance of nonamyloid heart after PYP imaging (no isotope uptake). ***B.*** Positive PYP scan in a patient with cardiac transthyretin amyloidosis. There is intense cardiac uptake on the planar imaging (*top*) which is confirmed to be in the wall of the heart on feud SPECT-CT images (*lower panel*). *(Courtesy of Dr. Sharmila Dorbala, Brigham and Women's Hospital.)*

caution and always interpreted in the setting of the clinical picture. It is thus strongly recommended that typing of the amyloid ideally be done by mass spectrometry (available as a "send-out test" in the United States). Unless a diagnosis of AL amyloidosis is clear, sequencing of the TTR gene should be done to rule out familial transthyretin amyloidosis as this has obvious ramifications for the family.

TREATMENT

Cardiac Amyloidosis

The treatment of systemic amyloidosis is addressed in detail in Chap. 117. Here, we will concentrate on specific issues related to

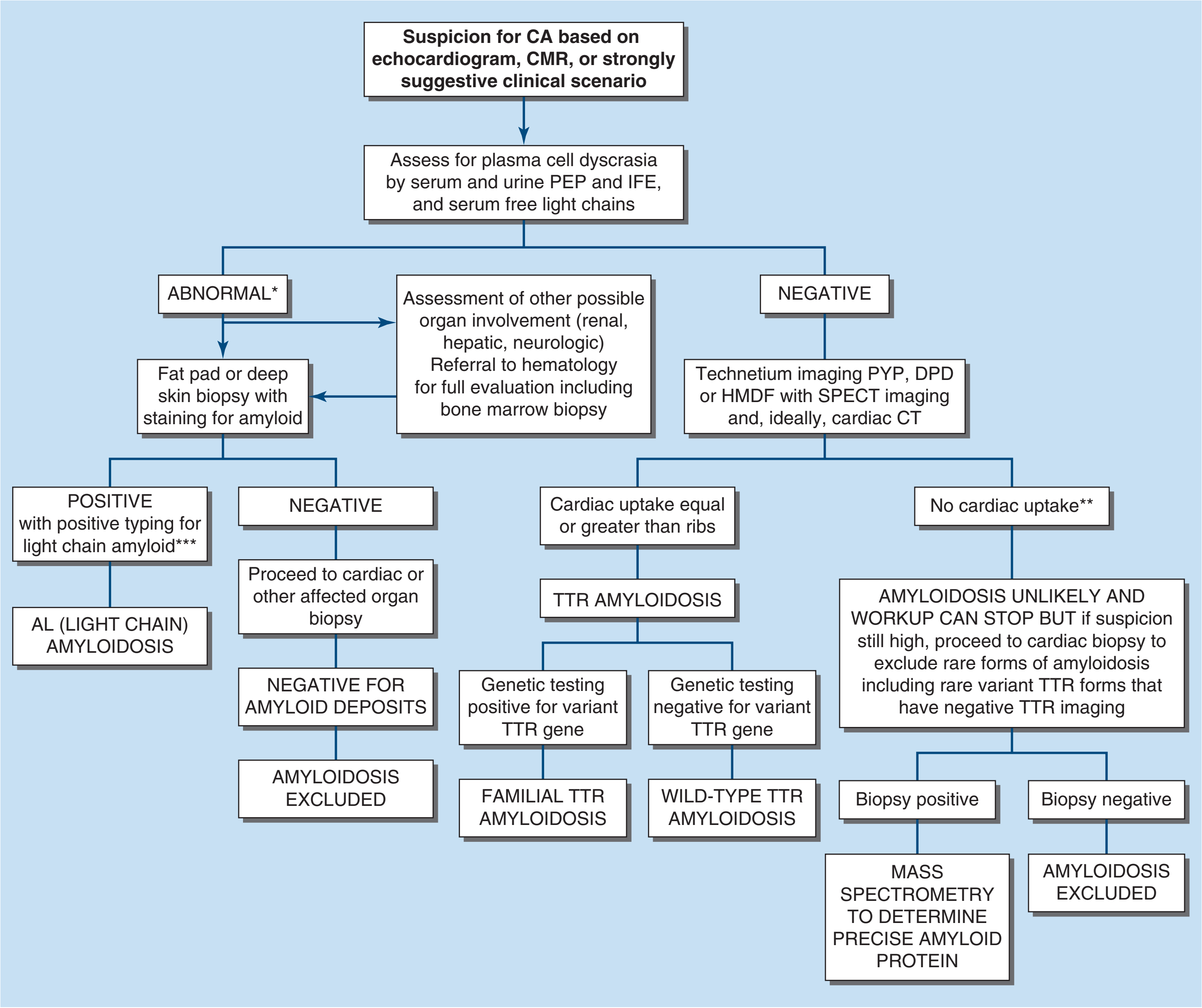
Suspicion for CA based on echocardiogram, CMR, or strongly suggestive clinical scenario
Assess for plasma cell dyscrasia by serum and urine PEP and IFE, and serum free light chains
ABNORMAL*
Assessment of other possible organ involvement (renal, hepatic, neurologic) Referral to hematology for full evaluation including bone marrow biopsy
NEGATIVE
Fat pad or deep skin biopsy with staining for amyloid
Technetium imaging PYP, DPD or HMDF with SPECT imaging and, ideally, cardiac CT
POSITIVE with positive typing for light chain amyloid***
NEGATIVE
Cardiac uptake equal or greater than ribs
No cardiac uptake**
AL (LIGHT CHAIN) AMYLOIDOSIS
Proceed to cardiac or other affected organ biopsy
TTR AMYLOIDOSIS
AMYLOIDOSIS UNLIKELY AND WORKUP CAN STOP BUT if suspicion still high, proceed to cardiac biopsy to exclude rare forms of amyloidosis including rare variant TTR forms that have negative TTR imaging
NEGATIVE FOR AMYLOID DEPOSITS
Genetic testing positive for variant TTR gene
Genetic testing negative for variant TTR gene
AMYLOIDOSIS EXCLUDED
FAMILIAL TTR AMYLOIDOSIS
WILD-TYPE TTR AMYLOIDOSIS
Biopsy positive
Biopsy negative
MASS SPECTROMETRY TO DETERMINE PRECISE AMYLOID PROTEIN
AMYLOIDOSIS EXCLUDED

FIGURE 270-8 Diagnostic algorithm. Algorithm for the workup of suspected cardiac amyloidosis (CA). This is a suggested algorithm but, depending on initial degree of suspicion for amyloidosis type, may vary slightly in order. Asterisk footnotes are as follows: *Normal values in workup for a plasma cell dyscrasia are negative serum and urine protein electrophoresis and immunofixation, with normal values for free serum kappa and lambda light chains adjusted for renal function. Abnormalities may not always reflect light chain amyloidosis, as monoclonal gammopathy of unknown significance not uncommonly coexists with TTR amyloidosis and thus expert evaluation of all abnormalities is mandatory. **Patients with bone-imaging uptake present but less than rib uptake should be carefully evaluated to ensure that this is not just blood pool isotope uptake and, if not, should be considered as possible amyloidosis with further evaluation based on expert evaluation of likely cardiac amyloid. ***Typing for amyloidosis should ideally be performed by mass spectrometry, generally only available in specialized centers. Immunohistochemistry, while often used, may be misleading and should be interpreted cautiously within the context of the clinical features of the disease. CMR, cardiac magnetic resonance; CT computed tomography; DPD, technetium-99m 3,3-diphosphono-1,2-propanodicarboxylic acid; HMDP, technetium-99m–hydroxymethylene diphosphonate; IFE, immunofixation electrophoresis; PYP, technetium-99m–pyrophosphate; PEP, protein electrophoresis; SPECT, single-photon emission computed tomography.

the cardiac disease. It is important, when considering treatment of cardiac amyloidosis, not simply to focus on therapies aimed at slowing or stopping the production of the precursor protein but also to address the manifestations of the disease. Sodium restriction is an integral part of management of heart failure in all forms of cardiac amyloidosis with congestive heart failure. The restrictive pathophysiology renders the patient particularly vulnerable to small changes in blood volume, and sodium restriction can significantly improve the response to diuretics. There are no specific clinical trials that evaluate the best way to treat congestive heart failure in cardiac amyloidosis, but it is generally recognized that diuretics are the mainstay of therapy. The combination of a loop diuretics such as furosemide or (preferably) torsemide with spironolactone or eplerenone is usually well tolerated. Renal function should be monitored while adjusting diuretic dose, particularly in patients with ATTRw amyloidosis who tend to be older and to have more baseline renal impairment. The use of other drugs commonly used in congestive heart failure is more controversial. In light chain amyloidosis, autonomic function is often impaired even in the absence of postural hypotension, and the use of angiotensin-converting enzyme inhibitors or angiotensin receptor blockers may be associated with profound hypotension. Thus, they are generally avoided. Beta-blockade is also often problematic and associated with decreased exercise tolerance. If a patient is receiving beta blockers when initially seen, we routinely taper and stop their use and reassess the patient's sense of well-being, which, not infrequently, improves. For patients with atrial fibrillation and a rapid ventricular response, beta-blockade may be needed but digoxin can also be used. However, restoration of sinus rhythm is preferable. The use of sodium-glucose cotransporter 2 (SGLT2) inhibitors appears to be clinically well-tolerated. This class of drug has efficacy across the spectrum of patients with heart failure and preserved ejection fraction and those with heart failure and reduced ejection fraction and may be of value in cardiac amyloidosis.

Cardiac arrhythmias in both AL and TTR amyloidosis are common. Most frequently seen is atrial fibrillation or flutter, the onset of which is often associated with clinical deterioration. There is a high risk of thromboembolism in patients with cardiac amyloidosis, and anticoagulation in patients with atrial fibrillation or flutter is mandatory. There does not appear to be increased risk of major bleeding among patients taking anticoagulants in either light chain or transthyretin amyloidosis. Restoration of sinus rhythm should be considered in all patients, particularly since rate-controlling agents are poorly tolerated. Amiodarone appears to be the most effective antiarrhythmic drug for atrial arrhythmias and should ideally be given, starting with a loading dose, prior to elective electrical cardioversion. If a patient is adequately anticoagulated before and during electrical cardioversion, there is no reason to perform transesophageal echocardiography prior to the procedure. If atrial arrhythmias are associated with clinical deterioration and sinus rhythm cannot be maintained despite antiarrhythmic therapy, consideration should be given to an invasive electrophysiologic procedure. Pulmonary venous isolation/atrial fibrillation ablation is often more complex than in nonamyloid patients, with a higher recurrence rate, and we therefore tend to limit ablation procedures to patients with atrial flutter and no evidence of atrial fibrillation. For those with symptomatic atrial fibrillation, atrioventricular nodal ablation with the implantation of a biventricular pacemaker is an effective way of both controlling and regularizing the ventricular rate. Sudden cardiac death may occur in cardiac amyloidosis, being more common in advanced AL amyloidosis, but a prophylactic implantable defibrillator is generally ineffective in preventing this. Thus, consideration of an implantable defibrillator should be highly selective and probably limited to patients with a prior resuscitated cardiac arrest or a history of sustained ventricular tachycardia.

Patients with TTR amyloidosis are at a high risk for the development of high-degree atrioventricular block. This is usually preceded on ECG by bifascicular block or left bundle branch block. Patients should be questioned about unexplained, nonpostural dizzy spells or syncope, and if these occur in the setting of one of these ECG abnormalities, strong consideration should be given to cardiac pacing. If a pacemaker is to be placed, we would generally recommend biventricular pacing, as right ventricular pacing produces abnormalities of septal contraction that, in the presence of a small cavity ventricle, may decrease cardiac output.

SPECIFIC THERAPY OF CARDIAC AMYLOIDOSIS: AL AMYLOIDOSIS

The most common regimen for treating patients with light chain amyloidosis includes daratumumab, bortezomib, cyclophosphamide, and dexamethasone. Dexamethasone may aggravate sodium retention, and it is critical that the cardiologist experienced in amyloidosis co-manages AL amyloidosis patients with cardiac involvement along with the treating hematologist. During therapy, an increased diuretic dose may be needed and/or a reduction in the dexamethasone dose. The other medications are generally well tolerated, but rarely, bortezomib may cause acute, reversible cardiac toxicity. Bortezomib may also cause a neuropathy and aggravate autonomic neuropathy of AL amyloidosis. High-dose melphalan chemotherapy with autologous stem cell transplantation is used in some patients with light chain amyloidosis and may be associated with considerable fluid retention or atrial arrhythmias in patients with cardiac involvement. For patients with AL cardiac amyloidosis, it is mandatory that the patient being considered for this therapy have a pretherapy cardiac evaluation to assess the risk-benefit ratio, along with regular posttransplant follow-up in the acute stage. Following successful chemotherapy (defined as a hematologic remission, with normalization of the serum free light chains), congestive heart failure often improves and is associated with an improvement in left ventricular longitudinal strain. However, some patients remain with very poor cardiac function. Such patients may benefit from cardiac transplantation, but this requires a very careful evaluation at an expert center to determine the extent of extracardiac amyloidosis and its likely effect on posttransplant prognosis.

SPECIFIC THERAPY OF CARDIAC AMYLOIDOSIS: TTR AMYLOIDOSIS

Currently, the aim of treatment of TTR cardiac amyloidosis is focused on reducing or stopping any further amyloid deposition. Tafamidis, a small molecule that stabilizes the tetrameric structure of transthyretin, is effective at slowing disease progression and may halt progression among patients treated early in the clinical presentation of the disease. It is extremely well tolerated. A similar drug, acoramidis, has shown similar efficacy in a pivotal clinical trial. An alternative class of drugs, the TTR silencers, significantly reduce the production of transthyretin by the liver and are approved in the United States for the treatment of familial amyloid polyneuropathy. At the time of writing, two drugs (both of which are only approved for amyloid neuropathy), eplontersen and vutrisiran, both administered subcutaneously, are completing clinical trials for the treatment of transthyretin amyloid cardiomyopathy.

Rarer forms of cardiac amyloidosis are listed in **Table 270-2**. Their rarity relates to the very uncommon nature of the type of amyloid with which they are associated or, as in the example of AA amyloidosis, the uncommon nature of cardiac involvement despite extensive other involvement. The diagnosis of most of these forms of cardiac amyloidosis will be made on mass spectrometry of a tissue biopsy after cardiac amyloidosis is suspected but when there is no evidence of a plasma cell dyscrasia or abnormality on technetium imaging.

TABLE 270-2 Rare Forms of Amyloidosis

AMYLOID TYPE	PRECURSOR PROTEIN	FREQUENCY OF CARDIAC DISEASE	OTHER ORGAN INVOLVEMENT	COMMENT
AA	Serum amyloid A (an inflammatory protein)	<5%	Kidney, liver	Usually associated with longstanding chronic inflammation. Uncommon in developed countries, and cardiac involvement rarely the predominant factor.
AApoA1	Apolipoprotein 1	25–30%	Kidney, liver, spleen, nervous system, larynx	Rare condition, with family history in many cases. Cardiac disease, when present, may be predominantly right-sided with tricuspid valve involvement.
AapoA4	Apolipoprotein 4	65–70%	Kidney	Family history often not present. Renal involvement without proteinuria is common, and cardiac involvement is usually present but often relatively mild.
Afib	Fibrinogen A-alpha (gene mutation)	Not described	Kidney, spleen	Rare. Proteinuria that progresses to renal failure. Typical renal biopsy with glomerular obliteration by amyloid. Typical amyloid cardiomyopathy absent.
ALECT2	Leukocyte cell-derived chemotaxin 2	Not described	Kidney	Liver involvement common. ALECT2 predominantly found in Hispanics.
Gelsolin	Gelsolin	Unknown prevalence. Usually mild, and manifesting as conduction disease	Corneal lattice dystrophy Skin and neurologic features	Predominantly found in Finnish patients.

SUMMARY

Over the past several years, cardiac amyloidosis has come to be recognized as a much more common disease than previously appreciated. In part, this is due to the aging population, in whom TTR amyloidosis is a significant cause of heart failure. The ability to diagnose TTR amyloidosis with noninvasive nuclear scanning rather than needing cardiac biopsy has made the diagnosis much easier, and the availability of effective disease-modifying treatment has added an incentive to diagnose previously untreated disease. AL amyloidosis, although more complex to diagnose, is seen in younger patients, and now effective lifesaving treatments and therapies for both types of cardiac amyloidosis are expanding. While a high degree of suspicion is needed to diagnose these diseases, the benefits of diagnosis to the patient are considerable, and the satisfaction of being able to treat a previously untreatable condition is one of the most gratifying aspects of the practice of medicine.

FURTHER READING

Dorbala S et al: How to image cardiac amyloidosis: A practical approach. JACC Cardiovasc Imaging 13:1368, 2020.

Ioannou A et al: Rare forms of cardiac amyloidosis: Diagnostic clues and phenotype in Apo AI and AIV amyloidosis. Circ Cardiovasc Imaging 16:523, 2023.

Kittleson MM et al: 2023 ACC Expert Consensus Decision Pathway on comprehensive multidisciplinary care for the patient with cardiac amyloidosis: A report of the American College of Cardiology Solution Set Oversight Committee. J Am Coll Cardiol 81;1076, 2023.

Ruberg FL, Maurer MS: Cardiac amyloidosis due to transthyretin protein: A review. JAMA 331:778, 2024.

Wechalekar AD et al: AL amyloidosis for cardiologists: Awareness, diagnosis, and future prospects: JACC: CardioOncology state-of-the-art review. JACC CardioOncol 4:427, 2022.

271

Cardiac Transplantation and Prolonged Assisted Circulation

Mandeep R. Mehra

Advanced heart failure, a distinct syndrome, is characterized by refractoriness to conventional therapy and represents a clinical dilemma associated with an increased symptom burden, frequent hospitalization, a poor quality of life, and high risk of death. Such individuals do not tolerate neurohormonal antagonists at recommended doses, exhibit worsening renal function, maintain poor cardiac reserve on cardiopulmonary stress testing, and typically display a low cardiac output state with elevated pulmonary pressures on invasive hemodynamic assessment. In general, therapeutic targets shift away from disease-modifying or neurohormonal therapy to surgical options that attend directly to supporting the dysfunctional heart and address myocardial stress and strain relationships. Most often, prolonged circulatory assistance using mechanical ventricular assist devices or cardiac transplantation is required to reliably improve quality of life and long-term survival.

CARDIAC TRANSPLANTATION

A decade after Norman Shumway had accomplished the technique of a successful heart transplant in canines, Christiaan Barnard successfully performed the first human-to-human transplant on December 3, 1967. Now, >5 decades later, this surgery has become entrenched in the standard armamentarium for treating patients with advanced heart failure who are otherwise healthy enough to receive such a life-altering treatment. Globally, >150,000 patients have undergone cardiac transplantation with a 1-year survival >80% and median survival of 12.5 years and conditional survival of 14.8 years if the recipient survives the first year after transplant. These gains have been ushered in by advances in immunosuppression, identification and management of allograft rejection, and a comprehensive appreciation for late complications including accelerated coronary artery disease, malignancy, and renal failure.

CANDIDATES FOR CARDIAC TRANSPLANTATION

The demand for cardiac transplantation outstrips the availability of organ donors. Hence, attention to the optimal utility, equitable allocation, and patient autonomy must dominate decisions to identify and list candidates for transplantation. Simultaneously, attempts at expanding the donor pool have surfaced. However, vigilance to evaluating candidates most likely to have a successful outcome from transplantation takes pre-eminence. In 2006, the International Society for Heart and Lung Transplantation identified a set of criteria to guide listing of patients. These criteria were updated in 2016 and include additional attention to the growing epidemiology of candidates suffering from congenital heart disease, restrictive and infiltrative cardiomyopathy (such as amyloidosis), and chronic infections in recipients (such as Chagas' disease, tuberculosis, and viral hepatitis). Selected general principles for listing candidates for cardiac transplantation are enumerated in Table 271-1.

PRINCIPLES OF DONOR RECOVERY AND ALLOCATION

Although listing criteria for candidates are typically adjudicated at an individual center level, organ allocation is handled by national regulatory processes in most countries. The allocation of donor hearts is based on (1) the urgency of the recipient clinical situation, (2) the length of time spent on the waiting list, (3) the distance from the recipient center, and (4) blood group. Candidates who are hospitalized in critical status and require extracorporeal membrane oxygenation

TABLE 271-1 Principles for Listing Candidates for Cardiac Transplantation

PRINCIPLE	COMMENT
Advanced disease severity	Refractory heart failure with a VO$_2$ of <14 mL/kg per min (<12, if on beta blockers) or percent predicted VO$_2$ <50%; combination of intolerance to disease-modifying therapy, cardiorenal syndrome, use of inotropic therapy to maintain stability, or need for a left ventricular assist system.
Comorbidity	Age is not an absolute contraindication, but frailty should be considered a relative contraindication; a BMI >35 kg/m^2 should require weight loss; cancer should be dealt with on an individual basis (e.g., low-grade prostate cancer may not be a contraindication); poorly controlled diabetes mellitus or end-organ damage may be a contraindication; eGFR <30 mL/min/1.73 m^2 is a relative contraindication (and requires consideration for a combined kidney transplant), if persistent; severe cerebrovascular disease or peripheral vascular disease (which will limit rehabilitation or function) is also a relative contraindication.
Donor-recipient matching	Sensitized individuals with circulating antibodies should have a prospective or virtual cross match; pulmonary vascular resistance with a transpulmonary gradient >15, PVR >3 Wood units, and absolute PA systolic pressure >50 mmHg (provided the systolic blood pressure is >85 mmHg) is a relative contraindication unless reactive to therapy.
Psychosocial issues	Tobacco use in any form limits posttransplant survival and should be stopped for at least 6 months; substance abuse, including marijuana, should be a contraindication if the individual cannot demonstrate control and cessation; patients with severe cognitive-behavioral disabilities or dementia (inability to ever understand and cooperate with medical care) have the potential for self-harm and should not receive a transplant.

Abbreviations: BMI, body mass index; eGFR, estimated glomerular filtration rate; PA, pulmonary artery; PVR, pulmonary vascular resistance; VO$_2$, peak oxygen consumption.

or temporary mechanical circulatory support devices to support both ventricles are given the highest urgency status, followed by those requiring daily invasive hemodynamic evaluation and intravenous inotropic therapy to maintain stability, or those with complications of a durable left ventricular assist device. Others stable at home while supported by a left ventricular assist device or those able to ambulate and live at home receive a lower urgency status. The geographical regional reach for allocation is based not only on territorial considerations but also on the time that a donor heart would be in transit and therefore in out of body "ischemia time," which is typically limited to 4 h, unless novel techniques for organ transport are employed. A key feature that is included in the allocation offer relates to the ABO blood group as well as the size match of a potential donor with the recipient. Donor organs are offered based on these initial characteristics and then a more detailed donor assessment ensues, resulting in acceptance or decline for any given donor heart. It is important to note that the time constraints imposed on the retrieval process make it difficult to invoke HLA matching of the donor and recipient. In cases where there is a high likelihood of allosensitization in the recipient (preformed circulating antibodies against donor antigens), a prospective or virtual cross match is entertained prior to acceptance. Other clinical criteria that are employed in the decision on accepting an offered donor include the donor-recipient size match (donor body weight no greater than 30% below that of the recipient if it is a same sex transplant, or no greater than 20% for a female to male transplant or assessed using a predicted heart mass calculator), the age of the donor (typically restricted to <55 years, but is often exceeded due to organ shortages), and presence or absence of concomitant pathology such as coronary artery disease, left ventricular hypertrophy, or severe injury to the allograft manifest by excess leak of injury markers (troponins) or poor contractile performance that does not improve in serial testing. In many cases, the prospective cardiac allograft can be reconditioned over time and used for transplantation even if the initial evaluation suggests poor contractile function. In efforts to enhance the donor pool, systems that allow ex situ normothermic perfusion to evaluate and reanimate organs with a prolonged out of body time have being developed that allow use of organs that may otherwise have been discarded. The classical heart donor is derived from a donor with established brain death; however, donors after circulatory death are being increasingly evaluated as candidates for cardiac reanimation using a variety of techniques including ex situ reanimation and subsequent transplantation. Such donor organs obtained after circulatory death are gaining widespread acceptance, and studies suggest that their outcomes after transplantation are no different from those of organs retrieved from brain death donors. However, this donor source is subject to ethical controversy in some geographical regions of the world. To increase the number of viable donors, organs from those infected with hepatitis C virus are increasingly being utilized since preventive or curative antiviral therapy can be used in the recipient shortly after transplantation.

SURGERY FOR CARDIAC TRANSPLANTATION

The most common contemporary operation is referred to as a "bicaval" orthotopic cardiac transplant that mimics the natural anatomic position. In this operation, the donor and recipient superior and inferior venae cavae are connected as are the aortic and pulmonary great vessels. The left atrium of the recipient retains its roof including the draining pulmonary veins, and the donor left atrium is then sutured to the retained atrial tissue. This technique maintains function of the donor right atrium, important for governing early postoperative right heart output, and may prevent atrial arrhythmias. The recipient is left with a surgical denervation, and the allograft is not responsive to any physiologic sympathetic or parasympathetic stimuli. Therefore, early in the adaptive postoperative phase, high-dose catecholamines are required to maintain adequate function. Due to denervation, bradycardia in a cardiac allograft cannot be treated with atropine and the drug of choice is isoproterenol, or temporary electrical pacing is used. Once the cardiac allograft adapts to its host circulation, the function is usually adequate at rest and with exercise, to provide normal physical activity and quality of life.

CARDIAC ALLOGRAFT REJECTION AND IMMUNOSUPPRESSION

The ability to perform endomyocardial biopsies and evaluate rejection pathologically and the introduction of the immunosuppression agent cyclosporine heralded cardiac transplantation as a viable clinical therapy. Triple-drug immunosuppression, which includes a calcineurin inhibitor (cyclosporine or tacrolimus), corticosteroids, and antiproliferative immunosuppression (azathioprine, mycophenolate mofetil, sirolimus, or everolimus), is now the standard combination used. The combination immunosuppression strategy that is most used and that achieves the best standard outcomes includes tacrolimus, mycophenolate mofetil, and prednisone. In those at high risk for rejection (multiparous women, allosensitized individuals, retransplants) or in situations where use of calcineurin inhibitors needs to be delayed (renal dysfunction), induction therapy using monoclonal (basiliximab) or polyclonal antibodies (antithymocyte globulin) to provide augmented immunosuppression is used, although this strategy remains poorly studied. Over several months, as surveillance endomyocardial biopsies are regularly performed and clinical as well as subclinical pathologic quiescence is established, gradual weaning of steroids is undertaken, in some cases to the point of elimination. Table 271-2 describes the immunosuppression drugs in common use.

TABLE 271-2 Immunoprophylaxis Drugs in Cardiac Transplantation

DRUG CLASS	GENERIC DRUG	CELLULAR TARGET	MAJOR SIDE EFFECTS
Calcineurin inhibitors	Cyclosporine	Binds to cyclophilin, which then inhibits calcineurin	Hypertension, dyslipidemia, gum hypertrophy, hypertrichosis
	Tacrolimus	Binds to immunophilin FK506 binding protein, which inhibits calcineurin	Hypertension, dyslipidemia, alopecia, diabetes mellitus
Antithymocyte globulin (ATG)	Rabbit ATG	T-cell depletion in blood and peripheral lymphoid tissues through complement-dependent lysis and T-cell activation and apoptosis	Cytokine release syndrome, leukopenia, thrombocytopenia, serum sickness
	Horse ATG	Same as above	Same as above
Interleukin-2 receptor antagonists	Basiliximab	Inhibition of CD-25 of interleukin 2 receptor	Well tolerated; rare hypersensitivity; increased infection risk if used with calcineurin inhibitors
Antimetabolites	Azathioprine	Imidazolyl derivative and prodrug of 6-mercaptopurine (cell cycle inhibitor)	Bone marrow suppression, pancreatitis, hepatitis
	Mycophenolate Mofetil	Inhibits inosine monophosphate dehydrogenase, which controls guanine monophosphate in the de novo pathway of purine synthesis (inhibits T- and B-cell proliferation)	Leukopenia, gastrointestinal toxicity
Proliferation signal inhibitors	Sirolimus	Binds with FKBP12 and complex inhibits the mechanistic target of rapamycin (mTOR)	Delayed wound healing, nonspecific pneumonia, pericardial effusion, hyperlipidemia (hypertriglyceridemia)
	Everolimus	Binds to FKBP12, which inhibits mTORC1 (and not mTORC2)	Dyslipidemia, stomatitis, pericardial effusions, and pancytopenia

Acute cellular rejection (ACR) and antibody-mediated rejection (AMR) are two separate forms of cardiac allograft rejection that are recognized and can sometimes coexist. ACR occurs early after transplantation and then tends to decline in incidence after 6 months. This occurs due to a T cell–mediated assault on the donor allograft tissue and histologically is characterized by lymphocytic infiltrates in the myocardium. In mild cases, these infiltrates are localized to the perivenular regions, and in severe cases, they progress diffusely into the cardiac interstitium. In late stages of severe ACR, most often associated with hemodynamic compromise, multiclonal cells such as macrophages, neutrophils, and eosinophils are observed with intramyocardial hemorrhage, myocyte injury, and myocyte necrosis. Subclinical ACR is typically treated with high doses of corticosteroid pulses, although some centers choose to simply observe mild forms of infiltration since it is known that many of these patients may recover spontaneously over time. If hemodynamic compromise occurs, rescue polyclonal antibodies are used in tandem with corticosteroids. Conversely, AMR is immunologically described as a noncellular antibody-driven phenomenon associated with a pattern of immunopathologic findings of immunoglobulin deposition and complement fixation on immunofluorescence, along with histopathologic findings of endothelial swelling and interstitial edema and cardiac allograft arteriolar vasculitis. AMR is characterized by the emergence of circulating donor-specific antibodies that are thought to fix complement and bind to the allograft. Commonly, AMR leads to allograft dysfunction, increases the risk for development of cardiac allograft vasculopathy, and is associated with worsened cardiac allograft survival compared with ACR. In this form of rejection, therapy is directed toward suppression and removal of circulating antibodies using plasmapheresis and drugs such as rituximab (chimeric monoclonal antibody directed against the CD20 antigen) or, in refractory cases, bortezomib (a proteasome inhibitor) or eculizumab (a terminal complement inhibitor). Intravenous immunoglobulin is also used as an immune-modulatory agent. The treatment with immunosuppression requires prophylaxis for opportunistic infections and ongoing surveillance and expertise in recognizing the more common clinical presentations of infections caused by cytomegalovirus (CMV), *Aspergillus*, and other opportunistic agents such as *Nocardia* and toxoplasmosis.

LATE COMPLICATIONS AFTER CARDIAC TRANSPLANTATION

The long-term consequences of exposure to chronic immunosuppression result in a variety of nonimmunologic cardiometabolic effects such as hypertension, hyperlipidemia, and hyperglycemia, as well as systemic disorders of bone loss and renal dysfunction. One aggressive complication that limits late survival of cardiac allografts includes the development of an accelerated form of coronary artery disease, referred to as cardiac allograft vasculopathy (CAV). This is characterized by a proliferative thickening of the vascular intima of the vasculature that is initiated as a diffuse endothelialitis in the setting of the confluence of the consequences of brain death, ischemia reperfusion injury during the transplant process, and early immunologic insults. Chronically, the metabolic consequences of hypertension, hyperlipidemia, and disordered glucose regulation result in worsening of vascular lesions that are diffuse and noted throughout the coronary tree. Early diagnosis and preventative therapy are critical since it is commonly silent in its development. Intravascular ultrasound or optical coherence tomography is more sensitive than routine coronary angiography for the diagnosis of CAV. A standardized grading scheme for CAV was proposed by the International Society for Heart and Lung Transplantation in 2010. CAV is graded as absent (CAV0), mild (CAV1), moderate (CAV2), or severe (CAV3) based on extent of angiographic disease and presence of allograft dysfunction by echocardiography or restrictive cardiac physiology. The grade of CAV correlates with patient prognosis. Angiographic CAV is present in ~30% of patients by 5 years after transplant and ~50% of patients by 10 years after transplant. Use of noninvasive testing for coronary flow reserve has been demonstrated to arbitrate prognosis in CAV. Use of statins may help prevent CAV development after heart transplantation, and targeting lipid lowering using nonstatin drugs (e.g., PCSK9 inhibitors) has shown promise. Antihypertensive agents and anti-CMV therapy have demonstrated some benefits in reducing CAV with varying degrees of supportive evidence. Antiproliferative immunosuppressive therapy such as mycophenolate mofetil and sirolimus or everolimus prevents vascular intimal thickening compared with azathioprine-based regimens. Use of the chimeric monoclonal antibody directed against the CD20 antigen, rituximab, may accelerate CAV. Options for medical treatment of established CAV remain limited. Revascularization using percutaneous coronary intervention or coronary artery bypass grafting may be employed in selected cases with focal lesions but does not improve survival in patients with CAV. However, retransplantation is the only definitive form of therapy for advanced CAV **(Fig. 271-1)**.

Another concern in cardiac transplantation is the development of malignancy with a greater frequency than in the normal population, suggesting that immunosuppression plays a sentinel role in its generation. Posttransplant lymphoproliferative disorders, typically driven by Epstein-Barr virus, occur most frequently and require a reduction in immunosuppression, administration of targeted antiviral agents, and traditional chemo- and radiotherapy. Specific antilymphocyte (targeted against CD20) therapy has also shown promise. Solid cancers most often manifest as skin malignancies (both basal cell and squamous cell carcinomas), and regular use of sunscreen is advised. Future research is required to define strategies for immune modulation, immune suppression, and malignancy prevention; however, the impact of decreasing immunosuppression in the treatment of these cancers remains unclear.

PROLONGED ASSISTED CIRCULATION

The quest for a prolonged and durable implantable mechanical circulatory support device has led to the development of continuous flow left ventricular assist systems (LVAS). Initially designed for short-term support as a bridge to recovery or to cardiac transplantation, the most frequent use today entails permanent support for lifetime therapy ("destination therapy"). The decision to implant LVAS dichotomously as either a bridge to transplantation or for destination therapy is not always clear, and in several instances, these devices are used as a "bridge to decision" (in those with potentially reversible underlying relative contraindications such as renal insufficiency or pulmonary hypertension, who may become future candidates for transplantation).

LEFT VENTRICULAR ASSIST SYSTEMS AND CLINICAL TRIALS

A pivotal trial, REMATCH, published in 2001, was the first study to reliably demonstrate that survival of patients with transplant-ineligible, refractory, predominantly inotropic therapy–supported heart failure is improved by implantation of an LVAS. This study used an early-generation pulsatile flow device and demonstrated a 48% reduction in risk of death. However, the LVAS used was of limited durability, and meaningful "out of hospital" survival was prolonged by a median of only 5 months. Furthermore, complications of strokes, multisystem organ failure, and infections reduced enthusiasm for widespread adoption. Over time, continuous flow systems that had small turbo-pumps with minimal moving parts and no valves were introduced, leading to greater durability and more generalized worldwide adoption.

A landmark trial compared the older bulky pulsatile LVAS studied in the REMATCH trial with a newer generation axial continuous flow LVAS, the HeartMate II, and demonstrated a marked improvement in short- and long-term survival, along with an improvement in functional capacity and meaningful quality-of-life enhancement. A centrifugal continuous flow LVAS, the HeartWare HVAD, was also introduced and demonstrated noninferiority to the HeartMate II pump. The HeartMate II is no longer used, and the HVAD has been removed from global distribution. A newer centrifugal device with a fully magnetically levitated system, the HeartMate 3 LVAS, is now the most used pump due to demonstration of reduced adverse events and improved long-term survival. The HeartMate 3 is a centrifugal, continuous flow pump that is placed in the thorax and is engineered to be a more hemocompatible LVAS. This device is constructed with

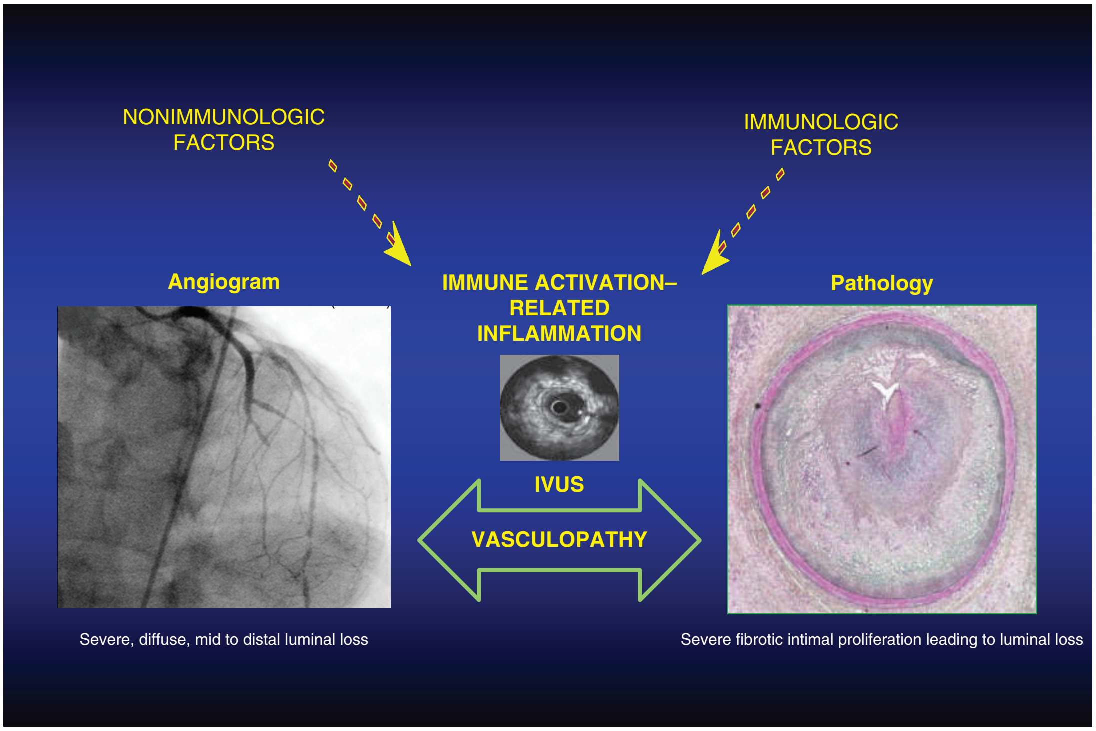
NONIMMUNOLOGIC FACTORS
IMMUNOLOGIC FACTORS
Angiogram
IMMUNE ACTIVATION–RELATED INFLAMMATION
Pathology
IVUS
VASCULOPATHY
Severe, diffuse, mid to distal luminal loss
Severe fibrotic intimal proliferation leading to luminal loss

FIGURE 271-1 Cardiac allograft vasculopathy is initiated and propagated by the combined influence of immunologic and nonimmunologic insults on the allograft vasculature. An inflammatory milieu determines the development of diffuse, aggressive luminal blockages that in early forms exhibit intimal thickening and fibrosis. IVUS, intravascular ultrasound (can be used to diagnose early forms of intimal thickening).

a fully magnetically levitated motor, offers wider blood flow paths, and exhibits a fixed intrinsic pulse (by the motor ramping its speed up and down at 2-s intervals). These features have been shown to reduce rates of hemolysis and decrease the shearing of high-molecular-weight multimers of von Willebrand factor. This pump has been tested in the MOMENTUM 3 trial, which reported its final results in 1028 patients randomly allocated to either the HeartMate 3 pump or the HeartMate II LVAS. The fully magnetically levitated HeartMate 3 pump was associated with less frequent need for pump replacement than the HeartMate II device and was superior with respect to survival free of disabling stroke or reoperation to replace or remove a malfunctioning device. The need for pump replacement and occurrence of stroke of any severity, major bleeding, or gastrointestinal hemorrhage were lower in the centrifugal-flow pump group than in the axial-flow pump group. Experience at 5 years with this LVAS has demonstrated an improved long-term survival (Fig. 271-2). This LVAS has been shown to nearly eliminate the complication of pump thrombosis and markedly reduce stroke rates, as well as decrease bleeding complications. Long-term follow-up with this pump has demonstrated a survival of nearly 60% at 5 years, which has been confirmed by real-world experience from registry analyses. In younger patients (<50 years) or those without comorbidity and an uncomplicated surgical course, the 5-year survival rivals that achieved with cardiac transplantation (75%). The patients for whom LVAS should ideally be employed include those with severe persistent systolic heart failure symptoms who have failed to respond to optimal medical management. Commonly, these patients have marked

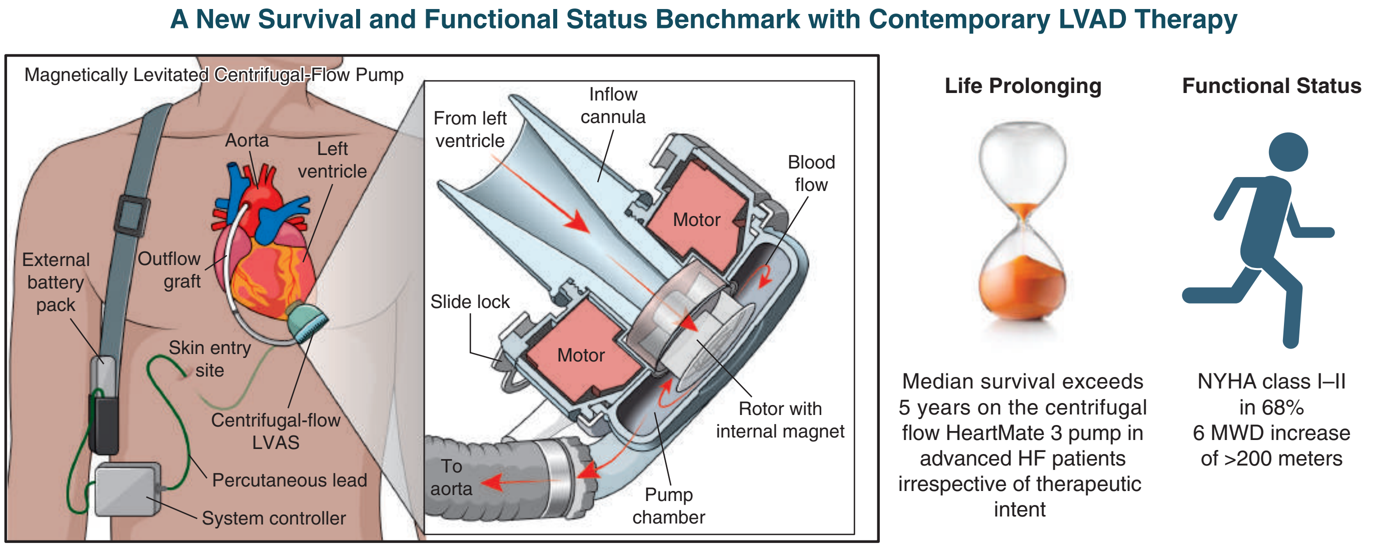
A New Survival and Functional Status Benchmark with Contemporary LVAD Therapy
Magnetically Levitated Centrifugal-Flow Pump
Aorta
Left ventricle
External battery pack
Outflow graft
Skin entry site
Centrifugal-flow LVAS
Percutaneous lead
System controller
From left ventricle
Inflow cannula
Blood flow
Motor
Slide lock
Motor
Rotor with internal magnet
To aorta
Pump chamber
Life Prolonging
Functional Status
Median survival exceeds 5 years on the centrifugal flow HeartMate 3 pump in advanced HF patients irrespective of therapeutic intent
NYHA class I–II in 68%
6 MWD increase of >200 meters

FIGURE 271-2 Due to the technological advances conferred by the HeartMate 3 pump, a new survival and functional status benchmark with contemporary left ventricular assist device (LVAD) therapy has been established. The centrifugal flow HeartMate 3 pump is frictionless and fully magnetically levitated, which does not induce hemolysis or pump thrombosis. This pump is associated with a median survival exceeding 5 years. 6 MWD, 6-min walk distance; HF, heart failure; NYHA, New York Heart Association. *(Left illustration: reproduced with permission from MR Mehra et al: N Engl J Med 376:440, 2017.)*

functional limitation indicated by a peak oxygen consumption of <12 mL/kg per min, or the patient is bound to continuous intravenous inotropic therapy owing to symptomatic hypotension or demonstrates worsening renal function or persistent refractory congestion. In such patients, despite improvements in medical therapy, the median survival is often less than 1 year, and substantial life prolongation can be anticipated with use of the HeartMate 3 LVAS. Currently, the role of LVAS in "less sick" patients (those with moderate symptoms) is less well supported since sufficient equipoise does not exist due to the need to be tethered to an external driveline that connects to a power source.

MANAGEMENT OF LVAS AND THEIR COMPLICATIONS

Continuous flow LVAS rely on pressure gradients between the left ventricular cavity and the aorta. As such, forward flow is critically dependent on management of systemic blood pressure. Due to the low pulsatile nature of the blood flow, blood pressure is measured by using a Doppler ultrasound (which measures mean or opening blood pressure, which is lower than the systolic blood pressure) since a peripheral pulse is usually not detectable. The ideal mean arterial blood pressure should be kept to ≤90 mmHg and antihypertensive drug therapy prescribed using renin-angiotensin-aldosterone system drugs or other vasodilators. The blood flow path through current devices results in increased shear stress, which may manifest in the form of low-grade hemolysis and the development of an acquired von Willebrand disease due to loss of high-molecular-weight multimers of von Willebrand factor. This hematologic aberration has been associated with a risk of gastrointestinal bleeding, particularly resulting from arteriovenous malformations in the intestines. Therefore, a common complication encountered in patients is that of an anemia, often due to iron deficiency.

The unsupported right ventricle often demonstrates worsening function and results in congestion requiring diuretic therapy. While unloading of the left ventricle decreases right-sided afterload, increased device flow results in a greater right heart preload, and effects of the LVAS on the septum reduce right ventricular contractile efficiency, leading to development of right ventricular dilatation and maladaptation between the right ventricle and pulmonary circuit. Cardiac arrhythmias are common in patients supported with LVAS and often require antiarrhythmic therapy or ablation since such events can trigger low flow through the device.

Hemocompatibility-related adverse outcomes include neurologic events (ischemic and hemorrhagic strokes), device-related thrombosis leading to pump malfunction, and nonsurgical bleeding complications **(Fig. 271-3)**. Elimination of aspirin as part of the antithrombotic regimen used (which includes a vitamin K antagonist targeting international normalized ratio to 2–3) has been shown to substantially reduce the risk of nonsurgical bleeding with consequent decrease in bleeding-related hospitalizations and cost of care. This outcome was demonstrated in the ARIES HM3 study with evidence of generalizability among most patients implanted with the HeartMate 3 LVAS. Strokes can occur with the HeartMate 3 LVAS and are generally noted during the surgical implant course with later occurrence in a manner similar to that observed for patients with advanced heart failure who are not supported with LVAS. This complication is an important reason for lack of adoption of device therapy in the less sick population. Another cause of morbidity with legacy devices was pump thrombosis requiring reoperation for device malfunction, which has been largely eliminated with the introduction of the HeartMate 3 LVAS. Lactate dehydrogenase is an excellent (although nonspecific) biomarker of hemolysis and can herald impending or established pump thrombosis, but this is uncommonly encountered except in circumstances where a circulating thrombus is entrained within the pump chamber. Patients who have suspected left ventricular assist device (LVAD) thrombosis and do not undergo LVAD exchange or cardiac transplantation have a high mortality rate, and reoperation (pump exchange) carries a modest perioperative mortality risk and is preferred as the therapy of choice, although this is extremely uncommon.

Infection is common, most often involving the driveline (the conduit connecting the device to the external controller and batteries), and occurs in 20–50% of patients following LVAS implant at 5 years. Such an infection is treated with local internal exploration and requires long-term suppressive antibiotics unless the patient undergoes cardiac transplantation or the device is exchanged. Infection and its inflammatory sequelae predispose to thrombosis and heighten the risk of neurologic complications, leading to a worsening milieu in hemocompatibility.

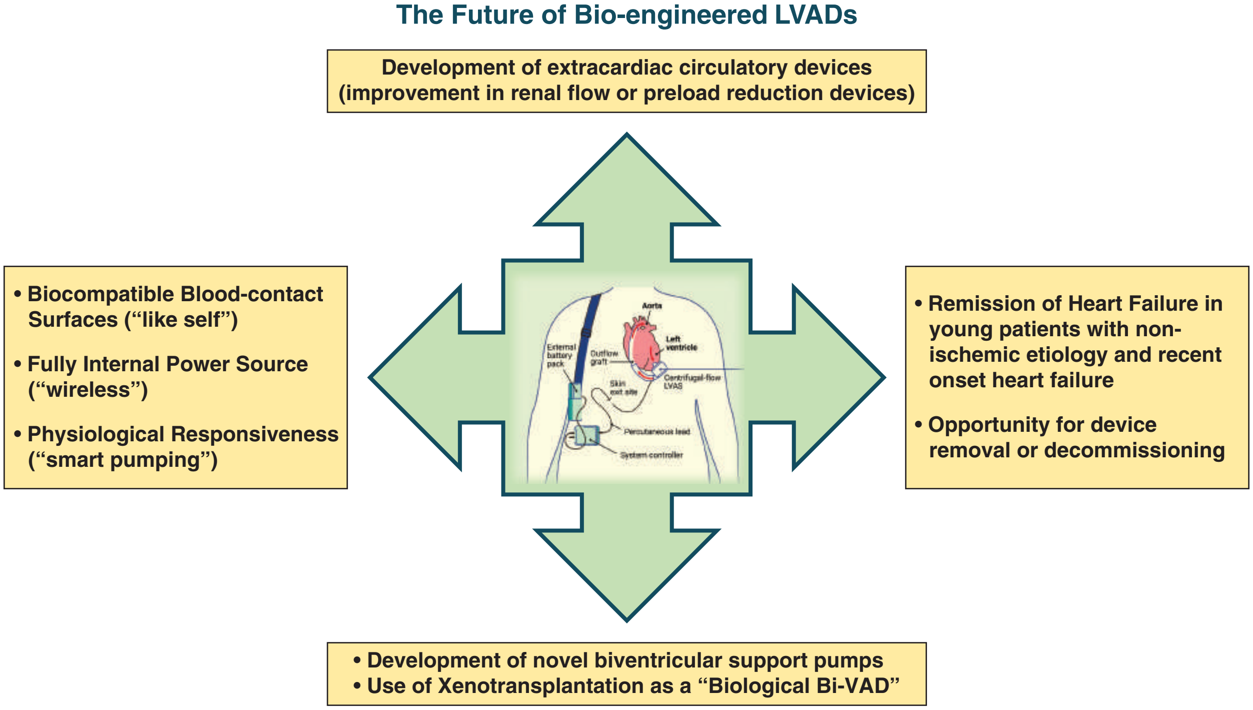
The Future of Bio-engineered LVADs
Development of extracardiac circulatory devices
(improvement in renal flow or preload reduction devices)
• Biocompatible Blood-contact Surfaces ("like self")
• Fully Internal Power Source ("wireless")
• Physiological Responsiveness ("smart pumping")
Aorta
Left ventricle
External battery pack
Outflow graft
Skin exit site
Centrifugal-flow LVAS
Percutaneous lead
System controller
• Remission of Heart Failure in young patients with non-ischemic etiology and recent onset heart failure
• Opportunity for device removal or decommissioning
• Development of novel biventricular support pumps
• Use of Xenotransplantation as a "Biological Bi-VAD"

FIGURE 271-3 The future of bioengineered left ventricular assist devices (LVADs). Engineering advances will require development of more biocompatible and fully internally implanted systems with physiologic responsive function. *(Adapted from Boulet J, Wanderley MRB Jr, Mehra MR. Contemporary Left Ventricular Assist Device Therapy as a Bridge or Alternative to Transplantation. Transplantation 108(6):1333-1341: 2024.)*

NOVEL DEVICES

Newer pumps are in development that are designed to overcome challenges inherent in current-generation LVAS. Engineering continues to advance in this field, and we await devices that provide physiologic and synchronized pulsatile flow (rather than the unnatural transapical to aortic flow with current LVADs). The next paradigm shift will likely require a return to natural and pulsatile flow LVAS that are more biocompatible (as opposed to hemocompatible), responsive to physiologic requirements (smart pumps), and forgettable (without an external driveline to power its components).

TOTAL ARTIFICIAL HEART

Not all patients are candidates for an LVAS, particularly those with severe right-sided heart failure or conditions that do not allow placement of an LVAS (restrictive cardiomyopathy, cardiac amyloidosis, massive anterior myocardial infarction, complex congenital heart disease). In such patients, either a biventricular assist device approach or a total artificial heart pump can be considered. The SynCardia total artificial heart is a pulsatile, implantable pump that consists of two polyurethane ventricles with pneumatically driven diaphragms and four tilting disc valves. This requires excision of the native ventricles and thus cannot be employed as a myocardial recovery strategy. There are specific clinical issues that are unique to the total artificial heart management. This device operates on a steep physiologic curve and has little adaptability to tolerate either systemic blood pressure changes or large shifts in blood volume. As the ventricles are excised, most patients exhibit a sharp decline in renal function due to the loss of natriuretic peptide expression by the myocardium. Severe hemolysis is common due to the presence of four mechanical valves, and aberrant erythropoiesis is noted, leading to a severe anemia. Newer artificial hearts using biocompatible surfaces are under study (CARMAT), as well as those that use continuous flow technology (BIVACOR).

XENOTRANSPLANTATION

On January 7, 2022, the first genetically edited pig-to-human heart xenotransplantation was performed. The porcine xenograft was derived from a 10-gene edited animal with four genes that were knocked out (targeting three carbohydrate antigens associated with hyperacute rejection and one anticardiac growth gene) and six genes that were knocked in (targeting human complement regulation, coagulation, and anti-inflammatory pathways). Two transplants have been performed in living human recipients with limited survival of 2 months or less, with death occurring due to delayed graft dysfunction and subsequent loss. There are substantial ongoing concerns with continued immunologic barriers (despite gene modification), costs of donor organ development and recovery, ethical considerations, and considerations of transmission of zoonoses.

GLOBAL CONSIDERATIONS

While LVAS are available worldwide, their use and indications vary from country to country. In the United States, payers used to require discrete discrimination of indication into either a bridge to transplant or destination therapy, whereas in most European countries, this artificial segregation was not used. Cost-effectiveness studies suggest improvement with the newer devices, yet some countries only allow use of this technology as a bridge to transplantation (United Kingdom) while awaiting more definitive long-term studies for lifetime use. The use of LVAS in moderately symptomatic ambulatory patients with chronic systolic heart failure is still discouraged throughout the world, awaiting the availability of devices that can be fully internalized without the need for an external driveline. Globally, the rates of myocardial recovery allowing for decommissioning or removal of devices remain low, although in young patients with nonischemic heart failure of relatively recent onset, this could be an important consideration.

FURTHER READING

Crespo-Leiro MG et al: Heart transplantation: Focus on donor recovery strategies, left ventricular assist devices, and novel therapies. Eur Heart J 43:2237, 2022.

Mehra MR et al: International Society for Heart and Lung Transplantation working formulation of a standardized nomenclature for cardiac allograft vasculopathy-2010. J Heart Lung Transplant 29:717, 2010.

Mehra MR et al: The 2016 International Society for Heart Lung Transplantation listing criteria for heart transplantation: A 10-year update. J Heart Lung Transplant 35:1, 2016.

Mehra MR et al: A fully magnetically levitated left ventricular assist device: Final report. N Engl J Med 380:1618, 2019.

Mehra MR et al: Five-year outcomes in patients with fully magnetically levitated vs axial-flow left ventricular assist devices in the MOMENTUM 3 randomized trial. JAMA 328:1233, 2022.

Mehra MR et al: The panvascular interplay in pathophysiology and prognosis of cardiac allograft vasculopathy. J Am Coll Cardiol 80:1629, 2022.

Mehra MR et al: Aspirin and hemocompatibility events with a left ventricular assist device in advanced heart failure: The ARIES-HM3 randomized clinical trial. JAMA 330:2171, 2023.

Mehra MR et al: Life-prolonging benefits of LVAD therapy in advanced heart failure: A clinician's action and communication aid. JACC Heart Fail 11:1011, 2023.

Schroder JN et al: Transplantation outcomes with donor hearts after circulatory death. N Engl J Med 388:2121, 2023.

272 Aortic Stenosis

Patrick T. O'Gara, Joseph Loscalzo

GLOBAL BURDEN OF VALVULAR HEART DISEASE

Valvular heart disease ranks well below ischemic heart disease, stroke, hypertension, obesity, and diabetes as a major threat to the public health. Nevertheless, it can cause significant morbidity and lead to premature death. Rheumatic fever **(Chap. 371)** is the dominant cause of valvular heart disease in low- and middle-income countries. Its prevalence has been estimated to range from as low as 1 per 100,000 school-age children in Costa Rica to as high as 150 per 100,000 in China **(Fig. 272-1)**. Prevalence is higher among females than males, especially for individuals age 20–40 years. Rheumatic heart disease accounts for 12–65% of hospital admissions related to cardiovascular disease and 2–10% of hospital discharges in some endemic countries. Prevalence and mortality rates vary among communities even within the same country as a function of overcrowding, the availability of medical resources, education level, and population-wide programs for detection and treatment of group A streptococcal pharyngitis. In economically deprived areas, tropical and subtropical climates (particularly on the Indian subcontinent and in Southeast Asia), Central America, and the Middle East, rheumatic valvular disease progresses more rapidly than in more developed nations and frequently causes serious symptoms in patients aged <20 years. This accelerated natural history may be due to repeated infections with more virulent strains of rheumatogenic streptococci. Approximately 45–50 million people (575.5 per 100,000) live with rheumatic heart disease worldwide, an estimated prevalence characterized by 300,000 new cases and 233,000 case fatalities (5 per 100,000) per year, with the highest prevalence and age-adjusted mortality rates in sub-Saharan Africa, South Asia, Central Asia, and Oceania. In the United States, rheumatic heart disease accounted for 3876 deaths in 2020. Although globally the age-standardized mortality rate from rheumatic heart disease declined by nearly 50% between 1990 and 2022, the prevalence of heart failure attributable to rheumatic heart disease increased by nearly 90% over the same time interval.

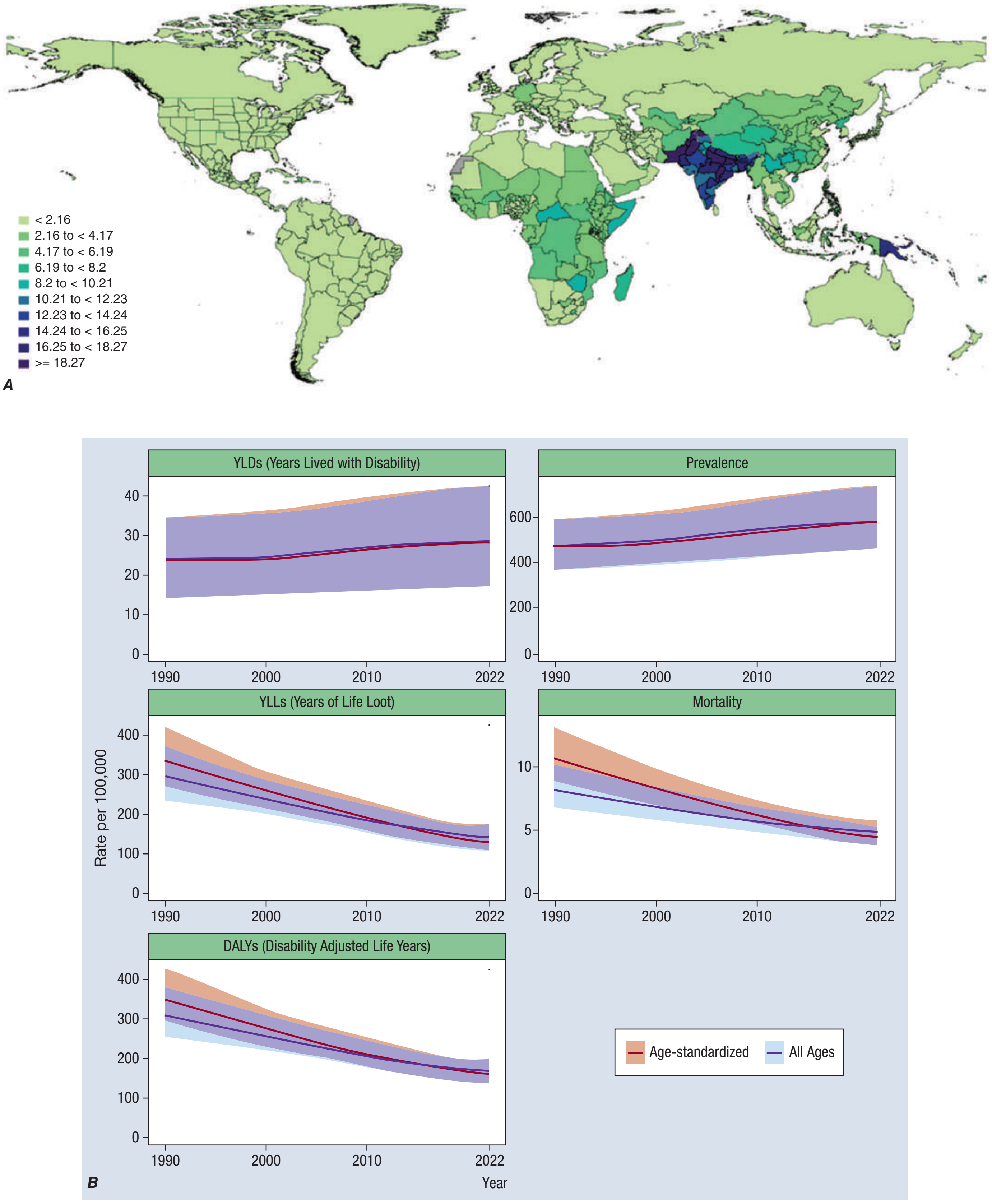

FIGURE 272-1 The global burden of rheumatic heart disease. (***A***) Global map of age-standardized rheumatic heart disease morality rate per 100,000 in 2022. Mortality rates are highest in South Asia and Oceania. (***B***) Global rheumatic heart disease estimates per 100,000 by measure with shaded 95% uncertainty interval, 1990–2022. DALYs, disability-adjusted life-years; YLDs, years lived with disability; YLLs, years of life lost. *(Reproduced with permission from GA Mensah et al: Global burden of cardiovascular diseases and risks, 1990-2022. J Am Coll Cardiol 82:2350, 2023.)*

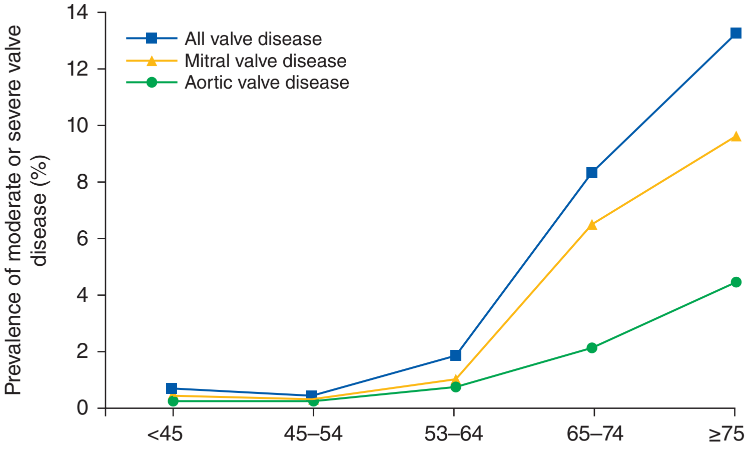

FIGURE 272-2 The burden of moderate or severe mitral and aortic valve disease in the United States. Prevalence estimates are derived from three population-based studies comprising a total of 11,911 individuals: The Coronary Artery Risk Development in Young Adults (CARDIA), the Atherosclerosis Risk in Communities (ARIC), and the Cardiovascular Health Study (CHS). *(Reproduced with permission from VT Nkomo, JM Gardin, TN Skelton, et al: Burden of valvular heart diseases: a population-based study, Lancet 368(9540):1005-1011, 2006.)*

Valve disease in high-income countries is dominated by degenerative or nonrheumatic inflammatory processes that lead to valve thickening, fibrosis, calcification, and dysfunction. The prevalence of valvular heart disease increases significantly with age. Community echocardiographic screening identifies previously undiagnosed, predominantly mild valvular heart disease in ~50% of the population aged >65 years. In this age group, the prevalence of previously undiagnosed moderate or severe valvular heart disease is ~6%. Significant left-sided valve disease may affect as many as 12–13% of adults aged >75 years **(Fig. 272-2)**. Severe aortic stenosis (AS) is estimated to affect 3.5% of the population aged >75 years. A Swedish epidemiologic study estimated the incidence of newly diagnosed valvular heart disease at 64 per 100,000 person-years, with approximately 70% of incident disease observed in individuals 65 years of age or older. AS and mitral regurgitation contributed approximately one-half and one-quarter, respectively, of the valvular heart disease diagnoses in this study.

The incidence of infective endocarditis **(Chap. 133)** has increased with the aging of the population, the more widespread prevalence of vascular grafts and intracardiac devices, the emergence of more virulent multidrug-resistant microorganisms, and the growing epidemic of injection drug use. North American age-standardized incidence rates for endocarditis increased from in 10.1 per 100,000 population in 1990 to 12.54 per 100,000 population in 2019. The more restricted use of antibiotic prophylaxis since 2007 has not been convincingly associated with an increase in incidence rates for infective endocarditis cases attributable to oropharyngeal pathogens. Infective endocarditis has become a relatively more frequent cause of acute valvular regurgitation. Valve surgery during the acute phase of infective endocarditis is performed in ~50–60% of hospitalized patients. Duration of intravenous antibiotic use may be shortened in selected cases.

Bicuspid aortic valve (BAV) disease affects as many as 0.5–1.4% of the general population and is accompanied by an associated aortopathy in ~30–40% of individuals, a disease process expressed as root or ascending aortic aneurysm formation or descending thoracic aortic coarctation. An increasing number of childhood survivors of congenital heart disease present later in life with valvular dysfunction. The global burden of valvular heart disease will continue to progress.

As is true for many other chronic health conditions, disparities in access to and quality of care for patients with valvular heart disease have been well documented, especially for those patients with rheumatic heart disease in low- and middle-income countries. In the Society for Thoracic Surgeons (STS)/American College of Cardiology (ACC) Transcatheter Valve Therapy (TVT) registry, black patients compose <5% of patients in the United States who have received a transcatheter valve for AS. Management decisions and outcome differences based on age, sex, race, geography, and other social determinants of health require intensification of educational efforts and prioritization of resources.

The role of the physical examination in the evaluation of patients with valvular heart disease is also considered in **Chaps. 44 and 246**; of electrocardiography (ECG) in **Chap. 247**; of echocardiography and other noninvasive imaging techniques in **Chap. 248**; and of cardiac catheterization and angiography in **Chap. 249**.

AORTIC STENOSIS

AS is the most common valve lesion among adult patients with chronic valvular heart disease; the majority of adult patients with symptomatic, valvular AS are male.

ETIOLOGY AND PATHOGENESIS

(Table 272-1) AS in adults is due to degenerative calcification of the aortic cusps and occurs most commonly on a substrate of congenital disease (BAV), chronic (trileaflet) deterioration, or previous rheumatic inflammation. A pathologic study of specimens removed at the time of aortic valve replacement (AVR) for AS in adults showed that 53% were bicuspid and 4% were unicuspid. The process of aortic valve deterioration and calcification is not a passive one, but, rather, one that shares many features with vascular atherosclerosis, including endothelial dysfunction, lipid accumulation, inflammatory cell activation, cytokine release, and upregulation of several signaling pathways **(Fig. 272-3)**. Eventually, a fibrocalcific response is established wherein collagen is deposited and valvular myofibroblasts differentiate phenotypically into osteoblasts and actively produce bone matrix proteins that allow for the deposition of calcium hydroxyapatite crystals. Genetic polymorphisms involving the vitamin D receptor, the estrogen receptor in postmenopausal women, interleukin 10, and apolipoprotein E4 have been linked to the development of calcific AS, and a strong familial clustering of cases with trileaflet valves has been reported from western France. Several traditional atherosclerotic risk factors have also been associated with the development and progression of calcific AS, including hypertension, low-density lipoprotein (LDL) cholesterol, lipoprotein(a) (Lp[a]), diabetes mellitus, smoking, chronic kidney disease, and the metabolic syndrome. In a Canadian observational cohort study, the incidence of severe AS was 144 per 100,000 person-years. Hypertension, diabetes mellitus, and dyslipidemia accounted for approximately one-third of the population-attributable risk for severe AS. The presence of aortic valve sclerosis (focal thickening and calcification of the leaflets not severe enough to cause obstruction) is associated with an excess risk of cardiovascular death and myocardial infarction (MI) among persons aged >65. Approximately 30% of persons aged >65 years exhibit some degree of aortic valve sclerosis. Rate and extent of progression to valve obstruction (stenosis) vary among individual patients.

Rheumatic disease of the aortic leaflets produces commissural fusion, sometimes resulting in a bicuspid-appearing valve. This condition, in turn, makes the leaflets more susceptible to trauma and ultimately leads to fibrosis, calcification, and further narrowing. By the time obstruction to left ventricular (LV) outflow causes serious clinical disability, the valve is usually a rigid calcified mass, and careful examination may make it difficult or even impossible to determine

TABLE 272-1 Major Causes of Aortic Stenosis

VALVE LESION	ETIOLOGIES
Aortic stenosis	Congenital (bicuspid, unicuspid) Degenerative calcific disease Rheumatic fever Radiation

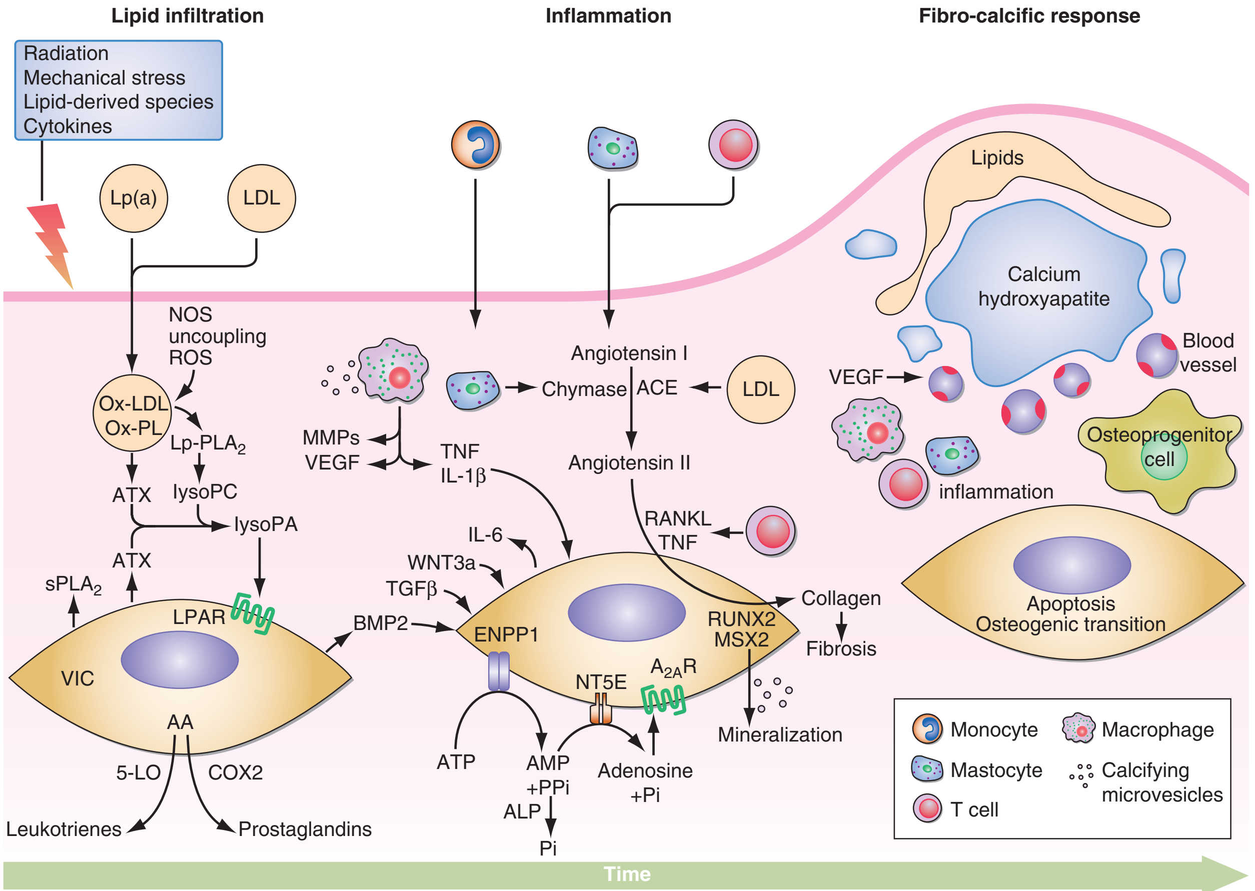
Lipid infiltration
Inflammation
Fibro-calcific response
Radiation
Mechanical stress
Lipid-derived species
Cytokines
Lp(a)
LDL
NOS
uncoupling
ROS
Ox-LDL
Ox-PL
Lp-PLA$_2$
ATX
lysoPC
lysoPA
ATX
sPLA$_2$
LPAR
VIC
AA
5-LO
COX2
Leukotrienes
Prostaglandins
MMPs
VEGF
TNF
IL-1β
Chymase
ACE
LDL
Angiotensin I
Angiotensin II
IL-6
WNT3a
TGFβ
BMP2
ENPP1
RANKL
TNF
RUNX2
MSX2
NT5E
A$_{2A}$R
ATP
AMP
+PPi
ALP
Pi
Adenosine
+Pi
Collagen
Fibrosis
Mineralization
Lipids
Calcium
hydroxyapatite
VEGF
Blood
vessel
Osteoprogenitor
cell
inflammation
Apoptosis
Osteogenic transition
Monocyte
Macrophage
Mastocyte
Calcifying
microvesicles
T cell
Time

FIGURE 272-3 Pathogenesis of calcific aortic stenosis. Lipid and inflammatory cell infiltration occurs across damaged endothelium. A cascade of events follows that leads eventually to formation of disorganized collagen (fibrosis) and calcium hydroxyapatite (bone) deposition. Valvular interstitial cells (VIC) are critical participants in this active process. AA, arachidonic acid; ACE, angiotensin-converting enzyme; ALP, alkaline phosphatase; ApoB, apolipoprotein B; AMP, adenosine monophosphate; ATP, adenosine triphosphate; ATX, autotaxin; A2AR, adenosine A2A receptor; BMP, bone morphogenetic protein; COX2, cyclooxygenase 2; ENPP, ectonucleotide pyrophosphatase/phosphodiesterase; IL, interleukin; 5-LO, 5-lipoxygenase; LDL, low-density lipoprotein; Lp(a), lipoprotein(a); LPAR, lysophosphatidic acid receptor; Lp-PLA2, lipoprotein-associated phospholipase A2; lysoPA, lysophosphatidic acid; lysoPC, lysophosphatidylcholine; MMP, matrix metalloproteinase; NOS, nitric oxide synthase; Ox-PL, oxidized phospholipid; Ox-LDL, oxidized LDL; RANKL, receptor activator of nuclear factor-κB ligand; ROS, reactive oxygen species; RUNX2, runt-related transcription factor 2; sPLA2, secreted PLA2; TGFβ, transforming growth factor β; TNF, tumor necrosis factor; VEGF, vascular endothelial growth factor; VIC, valvular interstitial cell. *(Reproduced with permission from B Lindman et al: Calcific aortic stenosis. Nat Rev Dis Primers 2:16006, 2016.)*

the etiology of the underlying process. Rheumatic AS is almost always associated with involvement of the mitral valve and with aortic regurgitation (AR). Mediastinal radiation can also result in late scarring, fibrosis, and calcification of the aortic leaflets. In this context, the calcification process also affects the mitral annulus.

BICUSPID AORTIC VALVE DISEASE

A bicuspid aortic valve (BAV) is the most common congenital heart valve defect and occurs in 0.5–1.4% of the population with a 2–4:1 male-to-female predominance. The inheritance pattern appears to be autosomal dominant with incomplete penetrance, although some have questioned an X-linked component as suggested by the prevalence of BAV disease among patients with Turner's syndrome. The prevalence of BAV disease among first-degree relatives of an affected individual is ~10%. A single gene defect to explain the majority of cases has not been identified, although mutations in the *NOTCH1*, *GATA5*, and *GATA4* genes have been described in some families. Abnormalities in endothelial nitric oxide synthase and NKX2.5 have been implicated as well. Medial degeneration with ascending aortic aneurysm formation occurs commonly among patients with BAV disease; aortic coarctation is less frequently encountered. Patients with BAV disease have larger aortas than patients with comparable tricuspid aortic valve disease. The aortopathy develops independently of the hemodynamic severity of the valve lesion, but directional shear forces dictated by the anatomic configuration of the valve appear to influence its expression. For example, enlargement of the ascending aorta along its greater curvature is most often associated with right-left cusp fusion (Sievers classification type 1), the most common bicuspid variant. Patients with BAV disease are at risk for aneurysm formation and/or dissection. A BAV can be a component of more complex congenital heart disease with or without other left heart obstructing lesions, as seen in Shone's complex (supravalvar mitral membrane, parachute mitral valve, subvalvar AS, and aortic coarctation).

OTHER FORMS OF OBSTRUCTION TO LEFT VENTRICULAR OUTFLOW

In addition to valvular AS, three other lesions may be responsible for obstruction to LV outflow: *hypertrophic obstructive cardiomyopathy* **(Chaps. 266–270)**, *discrete fibromuscular/membranous subaortic stenosis*, and *supravalvular AS* **(Chap. 280)**. The causes of LV outflow obstruction can usually be differentiated on the basis of the cardiac examination and Doppler echocardiographic findings.

PATHOPHYSIOLOGY

The obstruction to LV outflow produces a systolic pressure gradient between the LV and aorta. When severe obstruction is suddenly produced experimentally, the LV responds by dilation and reduction of stroke volume. However, in some patients, the obstruction may be present at birth and/or increase gradually over the course of many years, and LV contractile performance is maintained by the presence of concentric LV hypertrophy. Initially, this serves as an adaptive mechanism because it reduces toward normal the systolic stress developed by the myocardium, as predicted by the Laplace relation wall tension normalized to wall thickness ($S = Pr/h$, where S = systolic wall stress, P = pressure, r = radius, and h = wall thickness). A large transaortic valve pressure gradient may exist for many years without a reduction in cardiac output (CO) or the development of LV dilation. Ultimately, however, excessive hypertrophy becomes maladaptive, LV systolic function declines because of afterload mismatch, abnormalities of diastolic function progress, and irreversible myocardial fibrosis develops.

A mean systolic pressure gradient >40 mmHg with a normal CO or an effective aortic orifice area of ~<1 cm^2 (or ~<0.6 cm^2/m^2 body surface area in a normal-sized adult)—i.e., less than approximately one-third of the normal orifice area—is generally considered to represent severe obstruction to LV outflow. The elevated LV end-diastolic pressure observed in many patients with severe AS and preserved ejection fraction (EF) signifies the presence of diminished compliance of the hypertrophied LV. Although the CO at rest is within normal limits in most patients with severe AS, it usually fails to rise normally during exercise. Loss of an appropriately timed, vigorous atrial contraction, as occurs in atrial fibrillation (AF) or atrioventricular dissociation, may cause rapid progression of symptoms. Late in the course, contractile function deteriorates because of afterload excess, the CO and LV–aortic pressure gradient declines, and the mean left atrial (LA), pulmonary artery (PA), and right ventricular (RV) pressures rise. LV performance can be further compromised by superimposed epicardial coronary artery disease (CAD). Stroke volume (and thus CO) can also be reduced in patients with significant hypertrophy and a small LV cavity despite a normal EF. Low-flow (defined as a stroke volume index <35 mL/m^2), low-gradient (defined as a mean pressure gradient <40 mmHg) AS (with either reduced or normal LV systolic function) is both a diagnostic and therapeutic challenge.

The hypertrophied LV causes an increase in myocardial oxygen requirements. In addition, even in the absence of obstructive CAD, coronary blood flow is impaired to the extent that ischemia can be precipitated under conditions of excess demand. Capillary density is reduced relative to wall thickness, compressive forces are increased, and the elevated LV end-diastolic pressure reduces the coronary driving pressure. The subendocardium is especially vulnerable to ischemia by this mechanism.

■ SYMPTOMS

AS is rarely of clinical importance until the valve orifice has narrowed to ~1 cm^2. Even severe AS may exist for many years without producing any symptoms because of the ability of the hypertrophied LV to generate the elevated intraventricular pressures required to maintain a normal stroke volume. Once symptoms occur, or the LV ejection fraction falls below normal, valve replacement is indicated.

Most patients with pure or predominant AS have gradually increasing obstruction over years but do not become symptomatic until the sixth to eighth decades. Adult patients with BAV disease, however, develop significant valve dysfunction and symptoms one to two decades sooner. Exertional dyspnea, angina pectoris, and syncope are the three cardinal symptoms. Often, there is a history of insidious progression of fatigue and dyspnea associated with gradual curtailment of activities and reduced effort tolerance. *Dyspnea* results primarily from elevation of the pulmonary capillary pressure caused by elevations of LV diastolic pressures secondary to impaired relaxation and reduced LV compliance. *Angina pectoris* usually develops somewhat later and reflects an imbalance between the increased myocardial oxygen requirements and reduced oxygen availability. CAD may or may not be present, although its coexistence is common among AS patients age >65. *Exertional syncope* may result from a decline in arterial pressure caused by vasodilation in exercising muscles and inadequate vasoconstriction in nonexercising muscles in the face of a fixed CO, or from a sudden fall in CO produced by an arrhythmia.

Because the CO at rest is usually well maintained until late in the course, marked fatigability, weakness, peripheral cyanosis, cachexia, and other clinical manifestations of a low CO are usually not prominent until this stage is reached. Orthopnea, paroxysmal nocturnal dyspnea, and pulmonary edema, i.e., symptoms of LV failure, also occur only in the advanced stages of the disease. Severe pulmonary hypertension leading to RV failure and systemic venous hypertension, hepatomegaly, AF, and tricuspid regurgitation (TR) are usually late findings in patients with isolated severe AS.

When AS and mitral stenosis (MS) coexist, the reduction in flow (CO) caused by MS lowers the pressure gradient across the aortic valve and, thereby, masks many of the clinical findings produced by AS. The transaortic pressure gradient can be increased in patients with concomitant AR due to higher aortic valve flow rates.

■ PHYSICAL FINDINGS

The heart rhythm is generally regular until late in the course; at other times, AF should suggest the possibility of associated mitral valve disease. Hypertension occurs commonly among older adults with AS. In the late stages, however, when stroke volume declines, the systolic pressure may fall and the pulse pressure narrow. The carotid arterial pulse rises slowly to a delayed peak (*pulsus parvus et tardus*). A thrill or anacrotic "shudder" may be palpable over the carotid arteries, more commonly the left. In the elderly, the stiffening of the arterial wall may mask this important physical sign. In many patients, the *a* wave in the jugular venous pulse is accentuated. This results from the diminished distensibility of the RV cavity caused by the bulging, hypertrophied interventricular septum.

The LV impulse is sometimes displaced laterally in the later stages of the disease. A double apical impulse (with a palpable S_4) may be appreciated, particularly with the patient in the left lateral recumbent position. A systolic thrill may be present at the base of the heart to the right of the sternum when leaning forward or in the suprasternal notch.

Auscultation An early systolic ejection sound is frequently audible in children, adolescents, and young adults with congenital BAV disease. This sound usually disappears when the valve becomes calcified and rigid. As AS increases in severity, LV systole may become prolonged so that the aortic valve closure sound no longer precedes the pulmonic valve closure sound, and the two components may become synchronous, or aortic valve closure may even follow pulmonic valve closure, causing paradoxical splitting of S_2 **(Chap. 246)**. The sound of aortic valve closure can be heard most frequently in patients with AS who have pliable valves; calcification diminishes the intensity of this sound. Frequently, an S_4 is audible at the apex and reflects the presence of LV hypertrophy and an elevated LV end-diastolic pressure; an S_3 generally occurs late in the course when the LV dilates and its systolic function becomes severely compromised.

The murmur of AS is described as an ejection (mid) systolic murmur that commences shortly after the S_1, increases in intensity to reach a peak toward the middle of ejection, and ends just before aortic valve closure. It is characteristically low-pitched, rough, and rasping in character, and loudest at the base of the heart, most commonly in the second right intercostal space. It is transmitted upward along the carotid arteries. Occasionally, it is transmitted downward and to the apex, where it may be confused with the systolic murmur of mitral regurgitation (MR) (Gallavardin effect). In almost all patients with severe obstruction and preserved CO, the murmur is at least grade III/VI. In patients with mild degrees of obstruction or in those with severe stenosis with heart failure and low CO in whom the stroke volume and, therefore, the transvalvular flow rate are reduced, the murmur may be relatively soft and brief.

■ LABORATORY EXAMINATION

ECG In most patients with severe AS, there is LV hypertrophy. In advanced cases, ST-segment depression and T-wave inversion (LV "strain") in standard leads I and aVL and in the left precordial leads are evident. However, there is no close correlation between the ECG and the hemodynamic severity of obstruction, and the absence of ECG signs of LV hypertrophy does not exclude severe obstruction. Systemic hypertension can coexist and also contribute to the development of hypertrophy.

Echocardiogram The key findings on transthoracic echocardiogram are thickening, calcification, and reduced systolic opening of the aortic valve leaflets and LV hypertrophy. Eccentric closure of the aortic valve cusps is characteristic of congenitally bicuspid valves. Transesophageal echocardiography imaging can display the obstructed orifice extremely well, but it is not routinely required for accurate characterization of AS. The valve gradient and aortic valve area can be estimated by Doppler measurement of the transaortic velocity. Severe AS is defined by a valve area <1 cm^2, whereas moderate AS is defined by a valve area of 1–1.5 cm^2 and mild AS by a valve area of 1.6–2 cm^2. Aortic valve sclerosis, conversely, is accompanied by a jet velocity of <2.5 m/s (peak gradient <25 mmHg). LV dilation and reduced systolic shortening reflect impairment of LV function. There is a robust experience with the use of longitudinal strain to characterize earlier changes in LV systolic function before a decline in EF can be appreciated. Doppler indices of impaired diastolic function are frequently seen. The frequency with which echocardiography should be repeated during follow-up is dictated by the severity of the stenosis (Table 272-2).

Echocardiography is useful for identifying coexisting valvular abnormalities, differentiating valvular AS from other forms of LV outflow obstruction, and measuring the aortic root and proximal ascending aortic dimensions. These aortic measurements are particularly important for patients with BAV disease. Dobutamine stress echocardiography can be useful for the evaluation of patients with AS and severe LV systolic dysfunction (low-flow, low-gradient, severe AS with reduced EF), in whom the severity of the AS can often be difficult to judge. Patients with severe AS (i.e., valve area <1 cm^2) with a relatively low mean gradient (<40 mmHg) despite a normal EF (low-flow, low-gradient, severe AS with normal EF) are often hypertensive, and efforts to control their systemic blood pressure should be optimized before Doppler echocardiography is repeated. The use of dobutamine stress echocardiography in this setting is not advised. When there is continued uncertainty regarding the severity of AS in patients with reduced CO and reduced or normal LVEF, quantitative analysis of the amount of aortic valve calcium with chest computed tomography (CT) can be helpful. Aortic valve calcium scores that define severe AS differ for men and women, as men tend to have relatively more calcification and women more fibrosis of the valve leaflets. There is increasing use of chest CT angiography to assess aortic valve morphology and function. It has become the imaging method of choice to plan for transcatheter aortic valve implantation (TAVI). Finally, the use of cardiac magnetic resonance (CMR) imaging to screen for the presence of increased extracellular volume (interstitial fibrosis) and late gadolinium enhancement (replacement fibrosis) in patients with severe AS is an area of active investigation. Future management pathways related to the asymptomatic AS patient are likely to include an integrated assessment of the findings from multimodality imaging studies.

Chest X-Ray The chest x-ray may show no or little overall cardiac enlargement for many years. Hypertrophy without dilation may produce some rounding of the cardiac apex in the frontal projection and slight backward displacement in the lateral view. A dilated proximal ascending aorta may be seen along the upper right heart border in the frontal view. Aortic valve calcification may be discernible in the lateral view, but it is usually readily apparent on fluoroscopic examination or by echocardiography; the absence of valvular calcification on fluoroscopy in an adult suggests that severe valvular AS is *not* present. In later stages of the disease, as the LV dilates, there is increasing roentgenographic evidence of LV enlargement, pulmonary congestion, and enlargement of the LA, PA, and right-sided heart chambers.

TABLE 272-2 Frequency of Follow-Up Echocardiography in Aortic Stenosis

STAGE OF DISEASE	FREQUENCY OF ECHOCARDIOGRAPHY
Progressive (stage B)	Every 3–5 years (mild severity, V_{max} 2.0–2.9 m/s) Every 1–2 years (moderate severity, V_{max} 3.0–3.9 m/s)
Severe asymptomatic (stage C1)	Every 6–12 months (V_{max} >4 m/s)

Catheterization Right- and left-sided heart catheterization for invasive assessment of AS is performed infrequently but can be useful when there is a discrepancy between the clinical and noninvasive findings. Concern has been raised that attempts to cross the aortic valve for measurement of LV pressures are associated with a risk of cerebral embolization. Catheterization can also be useful in three distinct categories of patients: (1) *patients with multivalvular disease*, in whom the role played by each valvular deformity should be defined to aid in the planning of operative treatment; (2) *young, asymptomatic patients with noncalcific congenital AS*, to define the severity of obstruction to LV outflow, because operation or percutaneous aortic balloon valvuloplasty (PABV) may be indicated in these patients if severe AS is present, even in the absence of symptoms; and (3) *patients in whom it is suspected that the obstruction to LV outflow may not be at the level of the aortic valve* but rather at the sub- or supravalvular level.

Coronary angiography is indicated to screen for CAD in appropriate patients with severe AS who are being considered for surgical or transcatheter valve intervention. Angiography can be performed invasively at the time of catheterization for hemodynamic assessment or with noninvasive CT techniques. Decision-making regarding the need for coronary artery revascularization at the time of aortic valve intervention is individualized.

NATURAL HISTORY

Death in patients with severe AS occurs most commonly in the seventh and eighth decades. Based on data obtained at postmortem examination in patients before surgical treatment became widely available, the average time to death after the onset of various symptoms was as follows: angina pectoris, 3 years; syncope, 3 years; dyspnea, 2 years; and heart failure, 1.5–2 years. Moreover, in >80% of patients who died with AS, symptoms had existed for <4 years. Among adults dying with valvular AS, sudden death, which presumably resulted from an arrhythmia, occurred in 10–20%; however, most sudden deaths occurred in patients who had previously been symptomatic. Sudden death as the first manifestation of severe AS is very uncommon (~1% per year) in asymptomatic adult patients. Calcific AS is a progressive disease, with an annual reduction in valve area averaging 0.1 cm^2 and annual increases in peak jet velocity and mean valve gradient averaging 0.3 m/s and 7 mmHg, respectively.

TREATMENT

Aortic Stenosis (Fig. 272-4)

MEDICAL TREATMENT

In patients with severe AS (valve area <1 cm^2), strenuous physical activity and competitive sports should be avoided, even in the asymptomatic stage. Care must be taken to avoid dehydration and hypovolemia to protect against a significant reduction in CO. Medications used for the treatment of hypertension or CAD, including beta blockers and angiotensin-converting enzyme (ACE) inhibitors, are generally safe for asymptomatic patients with preserved LV systolic function. Control of blood pressure is important to attenuate the deleterious pathophysiologic effects of two resistance circuits (valve, arterial circulation) in series. Nitroglycerin is helpful in relieving angina pectoris in patients with CAD. Neither HMG-CoA reductase inhibitors (“statins”) nor

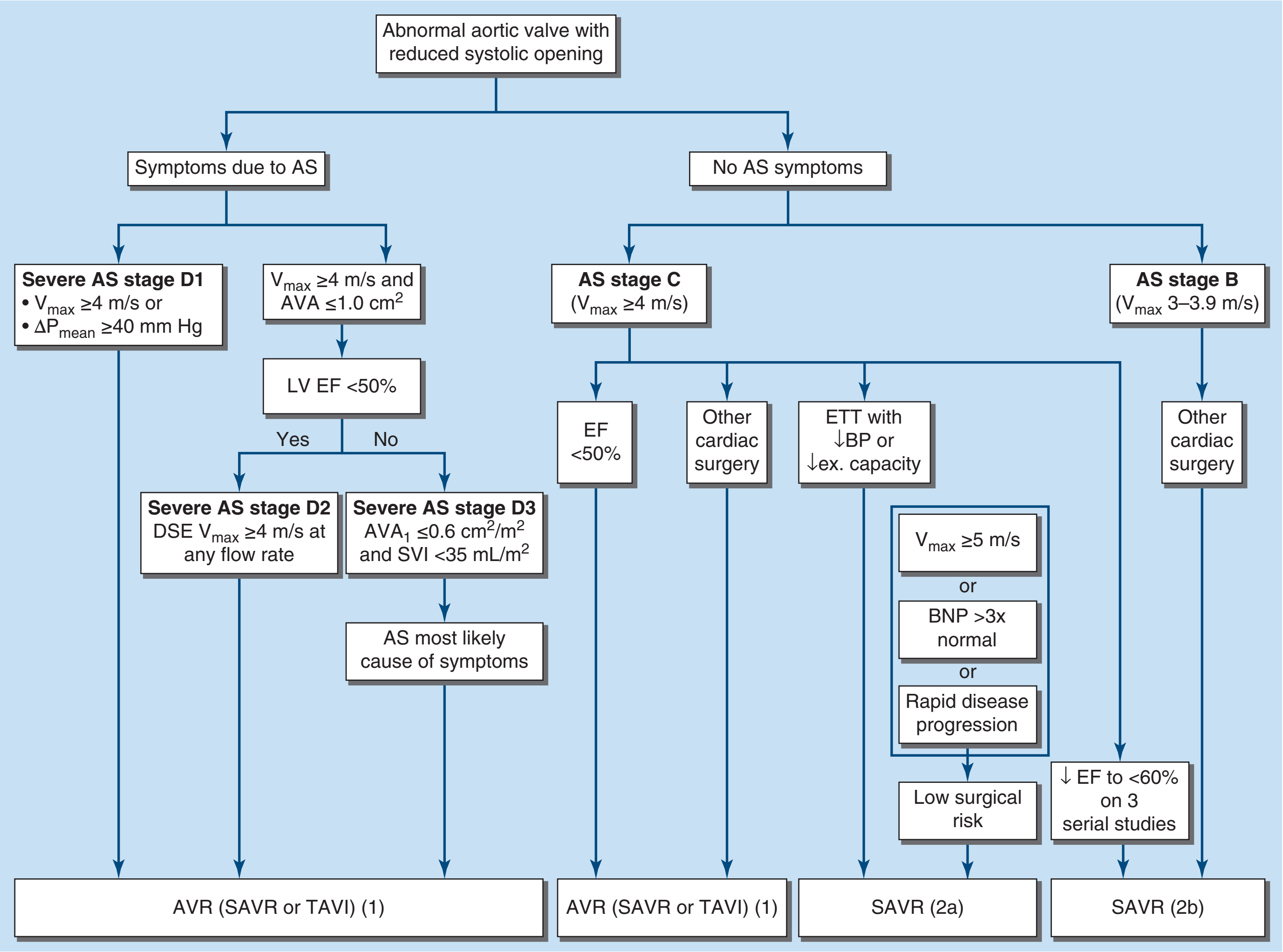
Abnormal aortic valve with reduced systolic opening
Symptoms due to AS
No AS symptoms
Severe AS stage D1
• V_{max} ≥4 m/s or
• ΔP_{mean} ≥40 mm Hg
V_{max} ≥4 m/s and AVA ≤1.0 cm^2
LV EF <50%
Yes
No
Severe AS stage D2
DSE V_{max} ≥4 m/s at any flow rate
Severe AS stage D3
AVA_1 ≤0.6 cm^2/m^2 and SVI <35 mL/m^2
AS most likely cause of symptoms
AS stage C
(V_{max} ≥4 m/s)
AS stage B
(V_{max} 3–3.9 m/s)
EF <50%
Other cardiac surgery
ETT with ↓BP or ↓ex. capacity
V_{max} ≥5 m/s
or
BNP >3x normal
or
Rapid disease progression
Low surgical risk
↓ EF to <60% on 3 serial studies
Other cardiac surgery
AVR (SAVR or TAVI) (1)
AVR (SAVR or TAVI) (1)
SAVR (2a)
SAVR (2b)

FIGURE 272-4 Management strategy for patients with aortic stenosis. Preoperative coronary angiography should be performed routinely as determined by age, symptoms, and coronary risk factors. Cardiac catheterization and angiography may also be helpful when there is a discrepancy between clinical and noninvasive findings. Patients who do not meet criteria for intervention should be monitored with clinical and echocardiographic follow-up. The class designations refer to the American Heart Association/American College of Cardiology methodology for treatment recommendations. Class I recommendations should be performed or are indicated; Class IIa recommendations are considered reasonable to perform; Class IIb recommendations may be considered. The stages refer to the stages of progression of the disease. At disease stage A, risk factors are present for the development of valve dysfunction; stage B refers to progressive, mild-moderate, asymptomatic valve disease; stage C disease is severe in nature but clinically asymptomatic; stage C1 characterizes asymptomatic patients with severe valve disease but compensated ventricular function; stage C2 refers to asymptomatic, severe disease with ventricular decompensation; stage D refers to severe, symptomatic valve disease. With aortic stenosis, stage D1 refers to symptomatic patients with severe aortic stenosis and a high valve gradient (>40 mmHg mean gradient); stage D2 comprises patients with symptomatic, severe, low-flow, low-gradient aortic stenosis and low left ventricular ejection fraction (LVEF); and stage D3 characterizes patients with symptomatic, severe, low-flow, low-gradient aortic stenosis and preserved LVEF (paradoxical, low-flow, low-gradient severe aortic stenosis). Patients with symptomatic severe AS (left side of the diagram, jet velocity ≥4m/s) should be referred for AVR (SAVR or TAVI). Asymptomatic patients with severe AS (jet velocity ≥4m/s) should be referred for AVR (SAVR or TAVI) for LVEF <50% or when other cardiac surgery is needed (e.g., aneurysm repair). There are several findings for which referral for AVR would be reasonable related to results of exercise testing, the presence of a jet velocity >5 m/s, or elevated B-type natriuretic peptide (BNP), provided the patient is considered low risk for complications related to AVR. AS, aortic stenosis; AVA, aortic valve area; AVR, aortic valve replacement; BP, blood pressure; DSE, dobutamine stress echocardiography; EF, ejection fraction; ETT, exercise treadmill test; ΔP_{mean}, mean pressure gradient; SAVR, surgical AVR; TAVI, transcatheter aortic valve implantation; V_{max}, maximum velocity. *(Reproduced with permission from CM Otto et al: 2020 AHA/ACC Guideline for management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 143(5):e72, 2021.)*

inhibitors of the renin-angiotensin-aldosterone system slow the rate of progression of AS. The use of statin medications should be driven by considerations regarding primary and secondary prevention of atherosclerotic cardiovascular disease (ASCVD) events. Studies with agents targeted to Lp(a) are ongoing. The need for endocarditis prophylaxis is restricted to AS patients with a prior history of endocarditis.

SURGICAL TREATMENT

Asymptomatic patients with calcific AS and severe obstruction should be followed carefully for the development of symptoms and for evidence of deteriorating LV function on serial echocardiography. Operation is indicated in patients with severe AS (valve area <1 cm^2 or 0.6 cm^2/m^2 body surface area) who are symptomatic, those who exhibit LV systolic dysfunction (EF <50%), and those with AS due to BAV disease and an aneurysmal root or ascending aorta (maximal dimension >5.5 cm). Operation for aneurysm disease is recommended at smaller aortic diameters (4.5–5.0 cm) for patients with a family history of an aortic catastrophe and for patients who exhibit rapid aneurysm growth (>0.5 cm/year). Patients with asymptomatic moderate or severe AS who are referred for coronary artery bypass grafting surgery should also have AVR. The majority (~80%) of patients with symptomatic severe AS referred for surgery are considered low risk for perioperative death or major complication. Operative risk increases as a function of age, comorbidities, and the need for concomitant aortic or other heart valve surgery or coronary artery bypass grafting. A 2023 analysis from the STS Adult Cardiac Surgery Database reported a 5-year survival rate of 95% following isolated surgical AVR (SAVR) in low-risk AS patients of mean age 74 years. The indications for SAVR in the asymptomatic patient have been the subject of intense debate, as surgical outcomes in selected patients have continued to improve. Relative indications for which surgery is reasonable include an abnormal response to treadmill exercise; rapid progression of AS, especially when urgent access to medical care might be compromised; very severe AS, defined by an aortic valve jet velocity >5 m/s or mean gradient >60 mmHg; excessive LV hypertrophy in the absence of systemic hypertension; and a brain natriuretic peptide level >3 times the upper reference limit, with low surgical risk. Exercise testing can be safely performed in asymptomatic patients, as many as one-third of whom will show signs of functional impairment. In a small randomized controlled trial (RCT) of early surgery versus conservative care for asymptomatic patients with very severe AS (defined by a transaortic valve jet velocity ≥4.5 m/s, mean gradient ≥50 mmHg, or aortic valve area ≤0.75 cm^2), the rate of operative death or death from cardiovascular causes during follow-up was reduced with early surgery. In the conservative care group, the cumulative incidence of sudden death was 4% at 4 years and 14% at 8 years. In another randomized trial of early surgery versus conservative care for asymptomatic patients with lesser degrees of AS (jet velocity ≥4 m/s, mean gradient ≥40 mmHg, aortic valve areas ≤1.0 cm^2) and normal LV systolic function, early surgery resulted in a significant reduction in a composite endpoint of death, MI, stroke, and heart failure hospitalization.

Operation should be carried out promptly (1–3 months) after symptom onset. Clinical decision-making is straightforward for patients with normal-flow (>35 mL/m^2), high-gradient (≥40 mmHg) severe AS. In patients with low-flow, low-gradient severe AS with reduced LVEF, perioperative mortality rates are high (15–20%), and evidence of LV dysfunction usually persists even after a technically successful operation. Long-term postoperative survival correlates with preoperative LV function. Nonetheless, in view of the even worse prognosis of such patients when they are treated medically, there is usually little choice but to advise valve replacement, especially in patients in whom flow reserve can be demonstrated by dobutamine stress echocardiography (defined by a ≥20% increase in stroke volume after dobutamine challenge). Patients in this high surgical risk group are treated with TAVI whenever feasible (see below), but robust data from RCTs in this subpopulation of severe AS patients are lacking. The management of patients with low-flow, low-gradient severe AS with normal LVEF is also challenging. Outcomes are improved with surgery or TAVI compared with conservative care for symptomatic patients with this type of "paradoxical" low-flow AS, but more research is needed to guide therapeutic decision-making for individual patients. In patients in whom severe AS and CAD coexist, relief of the AS and revascularization may sometimes result in striking clinical and hemodynamic improvement.

Because many patients with calcific AS are elderly, particular attention must be directed to the adequacy of hepatic, renal, and pulmonary function before AVR is recommended. Age alone is not a contraindication to SAVR for AS. The perioperative mortality rate depends to a substantial extent on the patient's preoperative clinical and hemodynamic state. Assessment of frailty is a critical component of preprocedural evaluation. Treatment decisions for AS patients who are not at low operative risk are made by a multidisciplinary heart team with representation from general cardiology, interventional cardiology, multimodality imaging, cardiac surgery, and other subspecialties as needed, including geriatrics. The 8-year survival rate of older adult (mean age 74), low surgical risk patients following isolated SAVR is 85–90%. Recommendations regarding the type of valve prosthesis (biological or mechanical) must weigh the trade-offs between limited bioprosthetic valve durability and the risks of thromboembolism and bleeding with a mechanical valve and are heavily influenced by patient age, expected longevity, and individual preferences. Bioprostheses are generally favored for patients age >65 years. Shared decision-making with younger patients must be individualized, although increasing numbers of patients age <65 now opt for a biological valve replacement. Approximately 10–20% of bioprosthetic valves evidence primary valve failure by 15 years, requiring re-replacement (or valve-in-valve TAVI, see below), and an approximately equal percentage of patients with mechanical prostheses develop hemorrhagic complications as a consequence of treatment with vitamin K antagonists. In a large observational study of patients who underwent SAVR in California between 1996 and 2013, receipt of a biological versus a mechanical prosthesis in patients <55 years old was associated with an excess hazard of death over 15 years of follow-up. Homograft AVR is usually reserved for patients with aortic valve endocarditis.

The Ross procedure involves replacement of the diseased aortic valve with the autologous pulmonic valve and implantation of a homograft in the native pulmonic position. It is a technically complex procedure that may be considered in selected young or middle-aged adult patients when surgical and institutional expertise are available. Late postoperative complications include aortic root dilation, AR, and pulmonary homograft stenosis.

PERCUTANEOUS AORTIC BALLOON VALVULOPLASTY

This procedure is preferable to operation in many children and young adults with congenital, noncalcific AS **(Chap. 280)**. It is not recommended as definitive therapy in adults with severe calcific AS because of a very high restenosis rate (80% within 1 year) and the risk of procedural complications, although on occasion, it has been used successfully as a bridge to operation or TAVI in patients with severe LV dysfunction and shock. It is performed routinely as part of the TAVI procedure (see below).

TRANSCATHETER AORTIC VALVE IMPLANTATION

TAVI surpassed SAVR for treatment of isolated AS in the United States in 2016 and is now available to symptomatic patients across the entire surgical risk spectrum (prohibitive, high, intermediate, and low) on the basis of the favorable results observed in a series of landmark RCTs reported over the past decade. The results of a randomized trial of TAVI versus conventional care in asymptomatic AS patients will be released around 2024–2025. TAVI is most commonly performed using one of two systems, a balloon-expandable valve (BEV) or a self-expanding valve (SEV), both of which incorporate a pericardial bioprosthesis **(Fig. 272-5*A, B*)**. TAVI is most frequently undertaken via the transfemoral route, although trans-LV apical, subclavian, carotid, and ascending aortic routes have been used. Nonfemoral access is associated with higher complication rates. Aortic balloon valvuloplasty under rapid RV (or LV) pacing is performed as a first step to create an orifice of sufficient size for the prosthesis. Procedural success rates exceed 95% in appropriately selected patients. Among low surgical risk patients with symptomatic severe AS, randomized trials with follow-up through 4–5 years have demonstrated similar valve performance and clinical outcomes for SAVR versus TAVI using either a BEV or SEV platform **(Fig. 272-6)**. Outcomes achieved with TAVI technology have been very favorable and have allowed the extension of AVR to groups of patients previously considered poor candidates for conventional surgery. Nevertheless, some prohibitive or high surgical risk patients are not candidates for this procedure because their comorbidity profile, frailty, and expected longevity would make its undertaking inappropriate. The heart team is specifically charged with making challenging decisions of this nature. The use of these devices for treatment of patients with structural deterioration of bioprosthetic aortic valves (valve-in-valve TAVI), as an alternative to reoperative valve replacement, has increased sharply over the past 5 years. The technology has also been increasingly applied to selected BAV patients despite the fact that patients with this anatomy were excluded from the landmark RCTs.

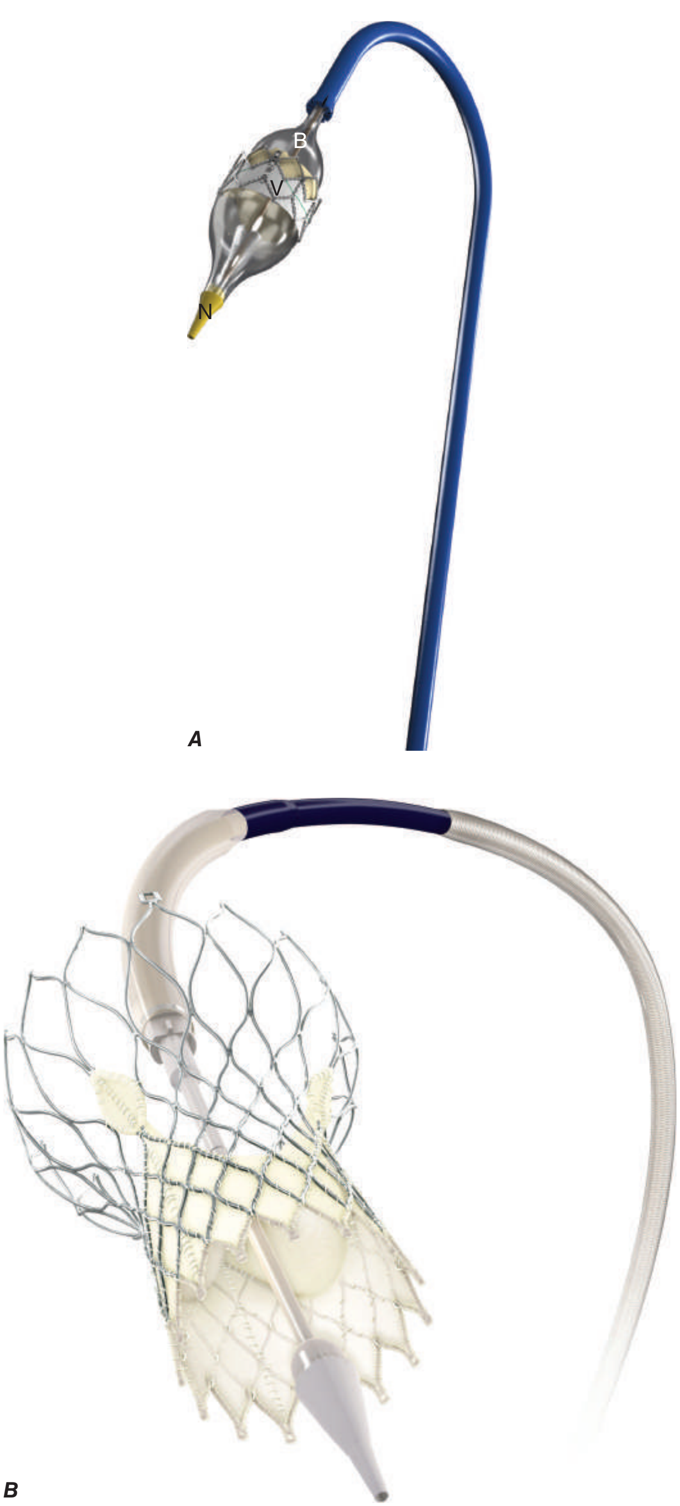

FIGURE 272-5 Balloon-expandable (*A*) and self-expanding (*B*) valves for transcatheter aortic valve replacement (TAVR). B, inflated balloon; N, nose cone; V, valve. *(Part A, courtesy of Edwards Lifesciences, Irvine, CA; with permission. NovaFlex+ is a trademark of Edwards Lifesciences Corporation. Part B, © Medtronic, Inc. 2015. Medtronic CoreValve Transcatheter Aortic Valve. CoreValve is a registered trademark of Medtronic, Inc.)*

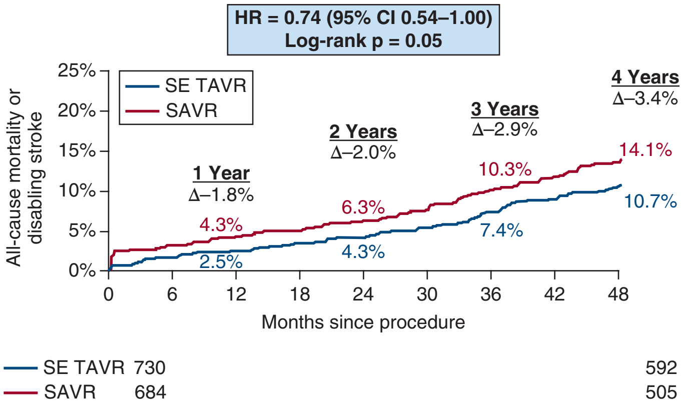

FIGURE 272-6 Four-year cumulative incidence of all-cause mortality or disabling stroke for low surgical risk aortic stenosis patients assigned to self-expanding transcatheter aortic valve implantation (SE-TAVR; n = 730) or surgical aortic valve replacement (SAVR; n = 684). In this study, TAVR was noninferior to SAVR and marginally superior to SAVR for the combined endpoint. *(Reproduced with permission from JK Forrest et al: 4-year outcomes of patients with aortic stenosis in the EVOLUT low risk trial. J Am Coll Cardiol 82:2163, 2023.)*

Compared with SAVR, transfemoral TAVI results in fewer periprocedural deaths and confers lower risks of strokes, major bleeding, and AF. Hospital lengths of stay are significantly shorter and return to normal activity more rapid with TAVI. Rates of permanent pacemaker use, perivalvular AR, bioprosthetic leaflet thrombosis, and vascular complications are lower with SAVR. The choice between TAVI versus SAVR for patients with trileaflet AS who prefer a biological prosthesis rests on several clinical, imaging, and technical considerations **(Fig. 272-7 and Table 272-3)**. Because there are scant RCT data on TAVI outcomes in patients <65 years, SAVR is recommended in this age group. Aortic valve/root anatomy, as well as the extent, severity, and distribution of calcium, and the distance of the coronary arteries from the plane of the annulus, may dictate a surgical approach, as could the need to perform a concomitant procedure such as ascending aortic replacement. Lastly, inability to achieve transfemoral access is a relative impediment to TAVI given the higher complication rates observed when this procedure is undertaken from other vascular access sites.

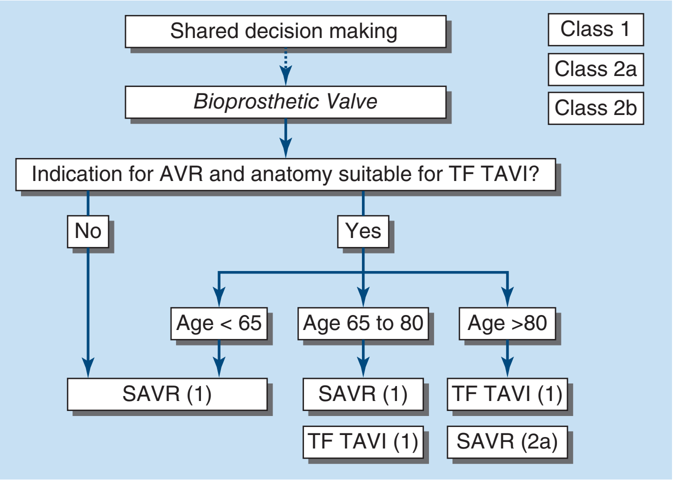

FIGURE 272-7 Suggested decision-making algorithm for the elective choice of transcatheter aortic valve implantation (TAVI) versus surgical aortic valve replacement (SAVR) for aortic stenosis patients with an indication for valve intervention. The pathway emphasizes the premium placed on transfemoral (TF) TAVI access and the age-related differences in recommendations. Patients younger than age 65 are recommended to undergo SAVR given the paucity of prospective randomized data on intermediate- and long-term TAVI outcomes for individuals younger than age 70. AVR, aortic valve replacement. *(Reproduced and abridged with permission from CM Otto et al: 2020 AHA/ACC Guideline for management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 143:e72, 2021.)*

TABLE 272-3 Factors Favoring SAVR, TAVI, or Palliative Care in Patients with Aortic Stenosis

	FAVORS SAVR	FAVORS TAVI	FAVORS PALLIATION
Age/life expectancy[a]	Younger age/longer life expectancy	Older age/fewer expected remaining years of life	Limited life expectancy
Valve anatomy	Bicuspid aortic valve Subaortic (LVOT) calcification Rheumatic valve disease Small or large aortic annulus[b]	Calcific trileaflet AS	
Prosthetic valve preference	Mechanical or surgical bioprosthetic valve preferred Concern for patient-prosthesis mismatch (annular enlargement might be considered)	Bioprosthetic valve preferred Favorable ratio of life expectancy to valve durability TAVI provides larger valve area than same-sized SAVR	
Concurrent cardiac conditions	Aortic dilation[c] Severe primary MR Severe CAD requiring bypass grafting Septal hypertrophy requiring myectomy Atrial fibrillation	Severe calcification of the ascending aorta ("porcelain" aorta)	Irreversible severe LV systolic dysfunction Severe MR due to annular calcification
Noncardiac conditions		Severe lung, liver, or renal disease Mobility issues (high risk for sternotomy)	Symptoms likely due to noncardiac conditions Severe dementia Moderate to severe involvement of 2 or more other organ systems
Frailty	Not frail or few frailty measures	Frailty likely to improve after TAVI	Severe frailty unlikely to improve after TAVI
Estimated risk of SAVR or TAVI	SAVR risk low TAVI risk high	TAVI risk low to medium SAVR risk high to prohibitive	Prohibitive SAVR risk (>15%) or post-TAVI life expectancy <1 year
Procedure-specific impediments	Valve anatomy, annular size, or low coronary ostial height precludes TAVI Vascular access does not allow transfemoral TAVI	Previous cardiac surgery with at-risk coronary grafts Previous chest irradiation	Valve anatomy, annular size, or coronary ostial height precludes TAVI Vascular access does not allow transfemoral TAVI
Goals of care and patient preferences and values	Less uncertainty about valve durability Avoid repeat intervention Lower risk of permanent pacer Life prolongation Symptom relief Improved long-term exercise capacity and QOL Avoid vascular complications Accepts longer hospital stay, pain in recovery period	Accepts uncertainty about valve durability and possible repeat intervention Higher risk of permanent pacer Life prolongation Symptom relief Improved exercise capacity and QOL Prefers shorter hospital stay, less postprocedure pain	Life prolongation not an important goal Avoid futile or unnecessary diagnostic or therapeutic procedures Avoid procedural stroke risk Avoid possibility of cardiac pacer

[a]Data on bioprosthetic valve durability are more robust for SAVR valves than for TAVI valves. Mechanical valves are very durable but require lifelong anticoagulation. Choice of prosthesis is a shared decision-making process accounting for individual patient values and preferences. [b]Surgical root enlargement can be performed at time of SAVR to allow a use of a larger prosthesis and reduce the occurrence of prosthesis-patient mismatch. [c]Aortic root or ascending aortic enlargement may require surgical correction at time of SAVR.

Abbreviations: AS, aortic stenosis; CAD, coronary artery disease; LV, left ventricular; LVOT, left ventricular outflow tract; MR, mitral regurgitation; QOL, quality of life; SAVR, surgical aortic valve replacement; TAVI, transcatheter aortic valve implantation.

Source: Reproduced with permission from CR Burke et al: Goals of care in patients with severe aortic stenosis. Eur Heart J 41:929, 2020.

FURTHER READING

Banovic M et al: Aortic valve replacement versus conservative treatment in asymptomatic severe aortic stenosis: The AVATAR trial. Circulation 145:648, 2022.

Carapetis JR et al: Acute rheumatic fever and rheumatic heart disease. Nat Rev Dis Primers 2:15084, 2016.

Forrest JK et al: 4-year outcomes of patients with aortic stenosis in the EVOLUT low risk trial. J Am Coll Cardiol 82:2163, 2023.

Kang D-H et al: Early surgery or conservative care for asymptomatic aortic stenosis. N Engl J Med 382:111, 2020.

Lindman B et al: Calcific aortic stenosis. Nat Rev Dis Primers 2:16006, 2016.

Mack MJ et al: Transcatheter aortic valve replacement in low-risk patients at 5 years. N Engl J Med 389:1949, 2023.

Mensah GA et al: Global burden of cardiovascular diseases and risks, 1990-2022. J Am Coll Cardiol 82:2350, 2023.

Otto CM et al: 2020 AHA/ACC Guideline for management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 143:e72, 2021.

Siontis GCM et al: Transcatheter aortic valve implantation versus surgical aortic valve replacement for treatment of symptomatic severe aortic stenosis: An updated meta-analysis. Eur Heart J 40:3143, 2019.

Thourani VH: Survival after surgical aortic valve replacement in low-risk patients: A contemporary trial benchmark. Ann Thor Surg 117:106, 2024.

Tsao CW et al: Heart disease and stroke statistics–2023 update. A report from the American Heart Association. Circulation 147:e93, 2023.

Watkins DA et al: Global, regional, and national burden of rheumatic heart disease, 1990-2015. N Engl J Med 377:713, 2017.

Zühlke L et al: Clinical outcomes in 3343 children and adults with rheumatic heart disease from 14 low- and middle-income countries: Two-year follow-up of the global Rheumatic Heart Disease Registry (the REMEDY Study). Circulation 134:1456, 2016.

273 Aortic Regurgitation

Patrick T. O'Gara, Joseph Loscalzo

ETIOLOGY

(Table 273-1) Aortic regurgitation (AR) may be caused by primary valve disease, aortic root disease, or their combination.

Primary Valve Disease Rheumatic disease results in thickening, deformity, and shortening of the individual aortic valve cusps, changes that prevent their proper opening during systole and closure during diastole. A rheumatic origin is much less common in patients with isolated AR who do not have associated rheumatic mitral valve disease. Patients with congenital bicuspid aortic valve (BAV) disease may develop predominant AR, and ~20% of these patients will require aortic valve surgery between 10 and 40 years of age. Congenital fenestrations of the aortic valve occasionally produce mild AR. Membranous subaortic stenosis results in a high-velocity systolic jet that often leads to thickening and scarring of the aortic valve leaflets and secondary AR. Prolapse of an aortic cusp, resulting in progressive chronic AR, occurs in ~15% of patients with ventricular septal defect (Chap. 280), but may also occur as an isolated phenomenon or as a consequence of myxomatous degeneration sometimes associated with mitral and/or tricuspid valve involvement.

AR may result from infective endocarditis (IE), which can develop on a valve previously affected by rheumatic disease, a congenitally deformed valve, or on a normal aortic valve, and may lead to perforation or destruction of one or more leaflets. The aortic valve leaflets may become scarred and retracted during the course of syphilis or ankylosing spondylitis and contribute further to the AR that derives primarily from associated root dilation. Although traumatic rupture or avulsion of an aortic cusp is an uncommon cause of acute AR, it represents the most frequent serious lesion in patients surviving nonpenetrating cardiac injuries. The coexistence of hemodynamically significant aortic stenosis (AS) with AR usually excludes all the rarer forms of AR because it occurs almost exclusively in patients with rheumatic or congenital AR. In patients with AR due to primary valvular disease, dilation of the aortic annulus may occur secondarily and lead to worsening regurgitation.

Primary Aortic Root Disease AR also may be due entirely to marked aortic annular dilation, i.e., aortic root disease, without primary involvement of the valve leaflets; widening of the aortic annulus and lack of diastolic coaptation of the aortic leaflets are responsible for the AR (Chap. 291). Medial degeneration of the ascending aorta, which may or may not be associated with other manifestations of Marfan's syndrome; idiopathic dilation of the aorta; annuloaortic ectasia; osteogenesis imperfecta; and severe, chronic hypertension may all widen the aortic annulus and lead to progressive AR. Occasionally AR is caused by retrograde dissection of the aorta involving the aortic annulus. Syphilis and ankylosing spondylitis, both of which may also affect the aortic leaflets, may be associated with cellular infiltration and scarring of the media of the thoracic aorta, leading to aortic dilation, aneurysm formation, and severe regurgitation. In syphilis of the aorta (Chap. 187), now a very rare condition, the involvement of the intima may narrow the coronary ostia, which in turn may be responsible for myocardial ischemia. Takayasu's aortitis and giant cell aortitis can also result in aneurysm formation and secondary AR.

TABLE 273-1 Major Causes of Aortic Regurgitation

VALVE LESION	ETIOLOGIES
Aortic regurgitation	**Valvular**
	Congenital (bicuspid)
	Endocarditis
	Rheumatic fever
	Myxomatous (prolapse)
	Radiation
	Trauma
	Syphilis
	Ankylosing spondylitis
	Aortic root disease
	Aortic dissection
	Medial degeneration
	Marfan syndrome
	Bicuspid aortic valve
	Nonsyndromic familial aneurysm
	Aortitis
	Hypertension

PATHOPHYSIOLOGY

The total stroke volume ejected by the left ventricle (LV) (i.e., the sum of the effective forward stroke volume and the volume of blood that regurgitates back into the LV) is increased in patients with AR. In patients with severe AR, the volume of regurgitant flow may equal the effective forward stroke volume. In contrast to MR, in which a portion of the LV stroke volume is delivered into the low-pressure left atrium (LA), in AR, the entire LV stroke volume is ejected into a high-pressure zone, the aorta. An increase in the LV end-diastolic volume (increased preload) constitutes the major hemodynamic compensation for AR. The dilation and eccentric hypertrophy of the LV allow this chamber to eject a larger stroke volume without requiring any increase in the relative shortening of each myofibril. Therefore, severe AR may occur with a normal effective forward stroke volume and a normal LV ejection fraction (LVEF, total [forward plus regurgitant] stroke volume/end-diastolic volume), together with an elevated LV end-diastolic pressure and volume. However, through the operation of Laplace's law, LV dilation increases the LV systolic tension required to develop any given level of systolic pressure. Chronic AR is, thus, a state in which LV preload and afterload are both increased. Ultimately, these adaptive measures fail. As LV function deteriorates, the end-diastolic volume rises further and the forward stroke volume and ejection fraction (EF) decline. Deterioration of LV function often precedes the development of symptoms. Considerable thickening of the LV wall also occurs with chronic AR, and at autopsy, the hearts of these patients may be among the largest encountered, sometimes weighing >1000 g.

The reverse diastolic pressure gradient from aorta to LV, which drives the AR flow, decreases progressively during diastole, accounting for the typical decrescendo nature of the diastolic murmur. Equilibration between aortic and LV pressures may occur toward the end of diastole in patients with chronic severe AR, particularly when the heart rate is slow. In patients with acute severe AR, the LV is unprepared for the regurgitant volume load. LV compliance is normal or reduced, and LV diastolic pressures rise rapidly, occasionally to levels >40 mmHg. The LV pressure may exceed the LA pressure toward the end of diastole, and this reversed pressure gradient closes the mitral valve prematurely.

In patients with chronic severe AR, the effective forward cardiac output (CO) usually is normal or only slightly reduced at rest, but often it fails to rise normally during exercise. An early sign of LV dysfunction is a reduction in the EF. In advanced stages, there may be considerable elevation of the LA, pulmonary artery (PA) wedge, PA, and right ventricular (RV) pressures and reduced forward CO at rest.

Myocardial ischemia may occur in patients with AR because myocardial oxygen requirements are elevated by LV dilation, hypertrophy, and elevated LV systolic tension, and coronary blood flow may be compromised. A large fraction (the majority) of coronary blood flow occurs during diastole, when aortic pressure is low, thereby reducing coronary perfusion or driving pressure. This combination of increased oxygen demand and reduced supply may cause myocardial ischemia, particularly of the subendocardium, even in the absence of epicardial coronary artery disease (CAD).

HISTORY

Approximately three-fourths of patients with pure or predominant valvular AR are men; women predominate among patients with primary valvular AR who have associated rheumatic mitral valve disease. A history compatible with IE may sometimes be elicited from patients with rheumatic or congenital involvement of the aortic valve, and the infection often precipitates or seriously aggravates preexisting symptoms.

In patients with *acute severe AR*, as may occur in IE, aortic dissection, or trauma, the LV cannot dilate sufficiently to maintain stroke volume, and LV diastolic pressure rises rapidly with associated marked elevations of LA and PA wedge pressures. Pulmonary edema and/or cardiogenic shock may develop rapidly.

Chronic severe AR may have a long latent period, and patients may remain relatively asymptomatic for as long as 10–15 years. Uncomfortable awareness of the heartbeat, especially on lying down, may be an early complaint. Sinus tachycardia, during exertion or with emotion, or premature ventricular contractions may produce particularly uncomfortable palpitations as well as head pounding. These complaints may persist for many years before the development of exertional dyspnea, usually the first symptom of diminished cardiac reserve. The dyspnea is followed by orthopnea, paroxysmal nocturnal dyspnea, and excessive diaphoresis. Anginal chest pain even in the absence of CAD may occur in patients with severe AR, even in younger patients. Anginal pain may develop at rest as well as during exertion. Nocturnal angina may be a particularly troublesome symptom, and it may be accompanied by marked diaphoresis. The anginal episodes can be prolonged and often do not respond satisfactorily to sublingual nitroglycerin. Systemic fluid accumulation, including congestive hepatomegaly and ankle edema, may develop late in the course of the disease.

PHYSICAL FINDINGS

In chronic severe AR, the jarring of the entire body and the bobbing motion of the head with each systole can be appreciated, and the abrupt distention and collapse of the larger arteries are easily visible. The examination should be directed toward the detection of conditions predisposing to AR, such as bicuspid valve, IE, Marfan's syndrome, or ankylosing spondylitis.

Arterial Pulse A rapidly rising "water-hammer" pulse, which collapses suddenly as arterial pressure falls rapidly during late systole and diastole (Corrigan's pulse), and capillary pulsations, an alternate flushing and paling of the skin at the root of the nail while pressure is applied to the tip of the nail (Quincke's pulse), are characteristic of chronic severe AR. A booming "pistol-shot" sound can be heard over the femoral arteries (Traube's sign), and a to-and-fro murmur (Duroziez's sign) is audible if the femoral artery is lightly compressed with a stethoscope.

The arterial pulse pressure is widened as a result of both systolic hypertension and a lowering of the diastolic pressure. The measurement of arterial diastolic pressure with a sphygmomanometer may be complicated by the fact that systolic sounds are frequently heard with the cuff completely deflated. However, the level of cuff pressure at the time of muffling of the Korotkoff sounds (phase IV) generally corresponds fairly closely to the true intra-arterial diastolic pressure. As the disease progresses and the LV end-diastolic pressure rises, the arterial diastolic pressure may actually rise as well, because the aortic diastolic pressure cannot fall below the LV end-diastolic pressure. For the same reason, acute severe AR may also be accompanied by only a slight widening of the pulse pressure. Such patients are invariably tachycardic as the heart rate increases in an attempt to preserve the CO.

Palpation In patients with chronic severe AR, the LV impulse is heaving and displaced laterally and inferiorly. The systolic expansion and diastolic retraction of the apex are prominent. A diastolic thrill may be palpable along the left sternal border in thin-chested individuals, and a prominent systolic thrill may be palpable in the suprasternal notch and transmitted upward along the carotid arteries. This systolic thrill and the accompanying murmur do not necessarily signify the coexistence of AS. In some patients with AR or with combined AS and AR, the carotid arterial pulse may be bisferiens, i.e., with two systolic waves separated by a trough **(see Fig. 246-2*C*)**.

Auscultation In patients with severe AR, the aortic valve closure sound (A_2) is usually absent. A systolic ejection sound is audible in patients with BAV disease, and occasionally an S_4 also may be heard. The murmur of chronic AR is typically a high-pitched, blowing, decrescendo diastolic murmur, heard best in the third intercostal space along the left sternal border **(see Fig. 246-5*B*)**. In patients with mild AR, this murmur is brief, but as the severity increases, it generally becomes louder and longer, indeed holodiastolic. When the murmur is soft, it can be heard best with the diaphragm of the stethoscope and with the patient sitting up, leaning forward, and with the breath held in forced expiration. In patients in whom the AR is caused by primary valvular disease, the diastolic murmur is usually louder along the left than the right sternal border. However, when the murmur is louder along the right sternal border, it suggests that the AR is caused by aneurysmal dilation of the aortic root. "Cooing" or musical diastolic murmurs suggest eversion of an aortic cusp vibrating in the regurgitant stream.

A mid-systolic ejection murmur is frequently audible in isolated AR. It is generally heard best at the base of the heart and is transmitted along the carotid arteries. This murmur may be quite loud without signifying aortic valve obstruction. A third murmur sometimes heard in patients with severe AR is the *Austin Flint murmur*, a soft, low-pitched, rumbling mid-to-late diastolic murmur. It is probably produced by the diastolic displacement of the anterior leaflet of the mitral valve by the AR stream and is not associated with hemodynamically significant mitral valve obstruction. The auscultatory features of AR are intensified by strenuous and sustained handgrip, which augments systemic vascular resistance and increases LV afterload.

In acute severe AR, the elevation of LV end-diastolic pressure may lead to early closure of the mitral valve, a soft S_1, a pulse pressure that is not particularly wide, and a soft, short, early diastolic murmur of AR.

LABORATORY EXAMINATION

ECG In patients with chronic severe AR, ECG signs of LV hypertrophy are common **(Chap. 247)**. In addition, these patients frequently exhibit ST-segment depression and T-wave inversion in leads I, aVL, V_5, and V_6 ("LV strain"). Left axis deviation and/or QRS prolongation may also be present.

Echocardiogram (Fig. 273-1) LV size is increased in chronic AR, and systolic function is normal or even supernormal until myocardial contractility declines, as signaled by a decrease in EF or increase in the end-systolic dimension. A rapid, high-frequency diastolic fluttering of the anterior mitral leaflet produced by the impact of the regurgitant jet is a characteristic finding. The echocardiogram is also useful in determining the cause of AR, by detecting dilation of the aortic annulus and root, aortic dissection or primary leaflet pathology. With severe AR, the central jet width assessed by color flow Doppler imaging exceeds 65% of the width of the LV outflow tract, the regurgitant volume is ≥60 mL/beat, the regurgitant fraction is ≥50%, and there is diastolic flow reversal in the proximal portion of the descending thoracic aorta. The continuous-wave Doppler profile of the AR jet shows a rapid deceleration time in patients with acute severe AR, due to the rapid increase in LV diastolic pressure. Surveillance transthoracic echocardiography (TTE) forms the cornerstone of longitudinal follow-up and allows for the early detection of changes in LV size and/or function. Assessment of LV global longitudinal strain (GLS; a measure of myocardial thickening in systole) with speckle track imaging may demonstrate changes in LV systolic function that precede a fall in EF. There is increasing experience with the use of three-dimensional echocardiography to measure LV volumes. Transesophageal echocardiography (TEE) can provide detailed anatomic assessment of the valve, root, and portions of the aorta. For patients in whom TTE is limited by poor acoustical windows or inadequate characterization of LV function or the cause or severity of the regurgitation, cardiac magnetic resonance (CMR) imaging can be performed. This modality also allows for accurate assessment of LV volumes, as well as aortic size and contour. It can also be utilized to screen for increased LV interstitial (extracellular volume fraction) and replacement fibrosis (late gadolinium enhancement). Both CMR imaging and cardiac computed tomography (CT) can also provide detailed assessment of aortic valve, root, and thoracic aortic anatomy.

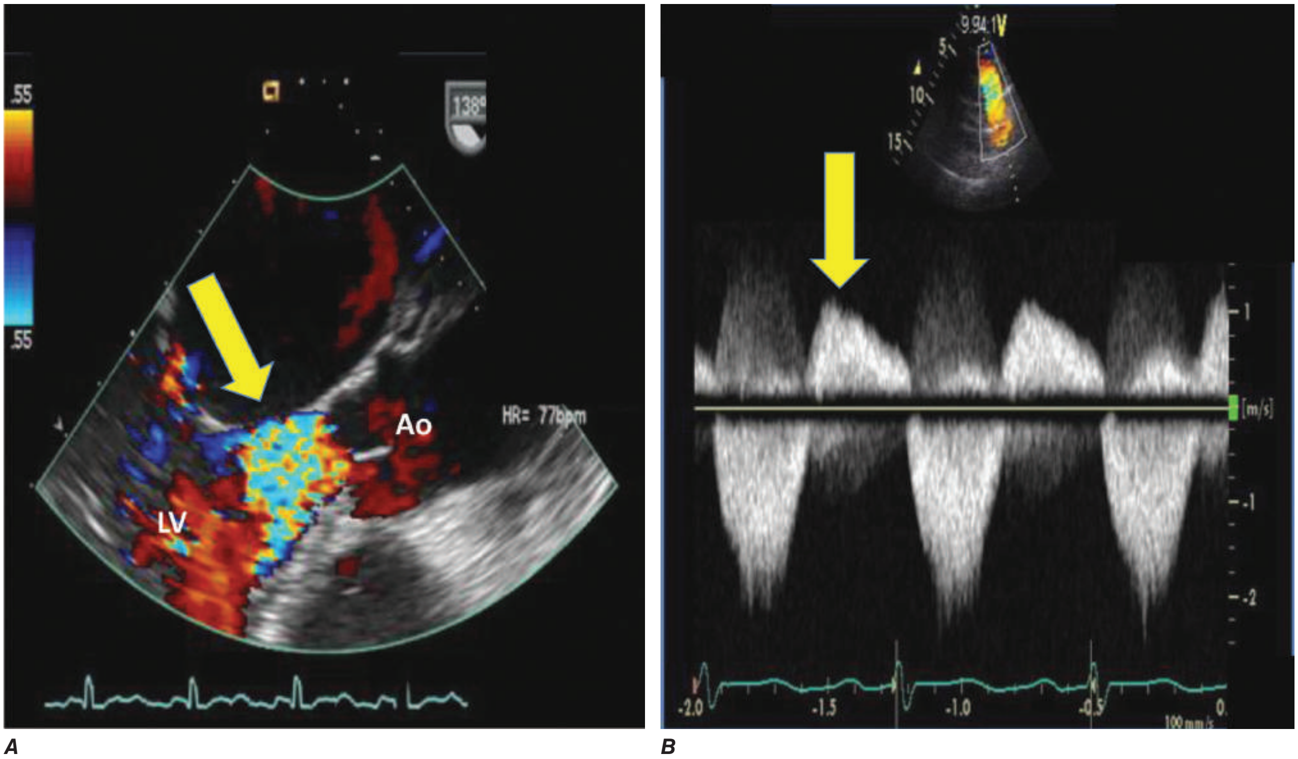

FIGURE 273-1 Echocardiographic and Doppler depiction of severe aortic regurgitation. (***A***) Color flow transesophageal echocardiographic long axis image in diastole shows a broad jet of severe aortic regurgitation (AR, yellow arrow) directed into the left ventricle. ECG rhythm strip below. Ao, aorta; BPM, beats per minute; HR, heart rate; LV, left ventricle. (***B***) Continuous wave Doppler tracing (middle image) obtained from the suprasternal window (top image) during transthoracic echocardiography shows dense, holodiastolic flow reversal in the descending thoracic aorta (yellow arrow) indicative of severe AR. ECG rhythm strip below.

Chest X-Ray In chronic severe AR, the apex is displaced downward and to the left in the frontal projection. In the left anterior oblique and lateral projections, the LV is displaced posteriorly and encroaches on the spine. When AR is caused by primary disease of the aortic root, aneurysmal dilation of the aorta may be noted, and the aorta may fill the retrosternal space in the lateral view. Echocardiography, CMR imaging, and chest CT angiography are more sensitive than a chest x-ray for the detection of root and ascending aortic enlargement.

Cardiac Catheterization and Angiography When needed, right and left heart catheterization with contrast aortography can provide confirmation of the magnitude of regurgitation and the status of LV function. Coronary angiography is performed routinely in patients at risk of coronary artery disease prior to surgery, although this anatomic information can also be obtained in many patients with coronary CT angiography.

TREATMENT

Aortic Regurgitation

ACUTE AORTIC REGURGITATION (FIG. 273-2)

Patients with acute severe AR may respond to intravenous diuretics and vasodilators (such as sodium nitroprusside), but stabilization is usually short-lived and operation is indicated urgently. Intra-aortic balloon counterpulsation is contraindicated. Beta blockers are best avoided so as not to reduce the CO further or slow the heart rate, thus allowing more time for diastolic filling of the LV. Surgery is the treatment of choice and is usually necessary within 24 h of diagnosis.

CHRONIC AORTIC REGURGITATION

The onset of symptoms, or LV systolic dysfunction, is an indication for surgery. Medical treatment with diuretics and vasodilators (angiotensin-converting enzyme inhibitors, angiotensin receptor blockers [ARBs], dihydropyridine calcium channel blockers, or hydralazine) may be useful as a temporizing measure. Surgery can then be performed in a more controlled setting. The use of vasodilators to extend the compensated phase of chronic severe AR in asymptomatic patients before the onset of symptoms or the development of LV dysfunction is not useful, although these agents should be employed to treat hypertension (systolic blood pressure >140 mmHg). It is often difficult to achieve adequate blood pressure control because of the increased stroke volume and enhanced LV ejection that accompany severe AR. Cardiac arrhythmias and systemic infections are poorly tolerated in patients with severe AR and must be treated promptly and vigorously. Although nitroglycerin and long-acting nitrates are not as helpful in relieving anginal pain as they are in patients with coronary heart disease, they are worth a trial. Patients with syphilitic aortitis should receive a full course of penicillin therapy **(Chap. 187)**. Beta blockers and the ARB losartan may be useful to retard the rate of aortic root enlargement in young patients with Marfan's syndrome and aortic root dilation. A randomized controlled trial showed no difference in efficacy between atenolol and losartan for this indication. Whether beta blockers or ARBs are useful in retarding the rate of growth of aortic aneurysms in other patient subsets (e.g., BAV disease with aortopathy, Takayasu's disease) has not been demonstrated. Beta blockers in patients with valvular AR were previously considered relatively contraindicated due to concern that slowing of the heart rate would allow more time for diastolic regurgitation and LV filling. Observational reports, however, have suggested that beta blockers may provide functional benefit in some patients with chronic AR. Beta blockers can sometimes provide incremental blood pressure lowering in patients with chronic AR and hypertension. They can also lessen the sense of forceful heart action that many patients find uncomfortable. Patients with severe AR, particularly those with an associated aortopathy, should avoid isometric exercises.

SURGICAL TREATMENT

In deciding on the advisability and proper timing of surgical treatment, two points should be kept in mind: (1) patients with chronic severe AR usually do not become symptomatic until *after* the development of myocardial dysfunction; and (2) when delayed too long

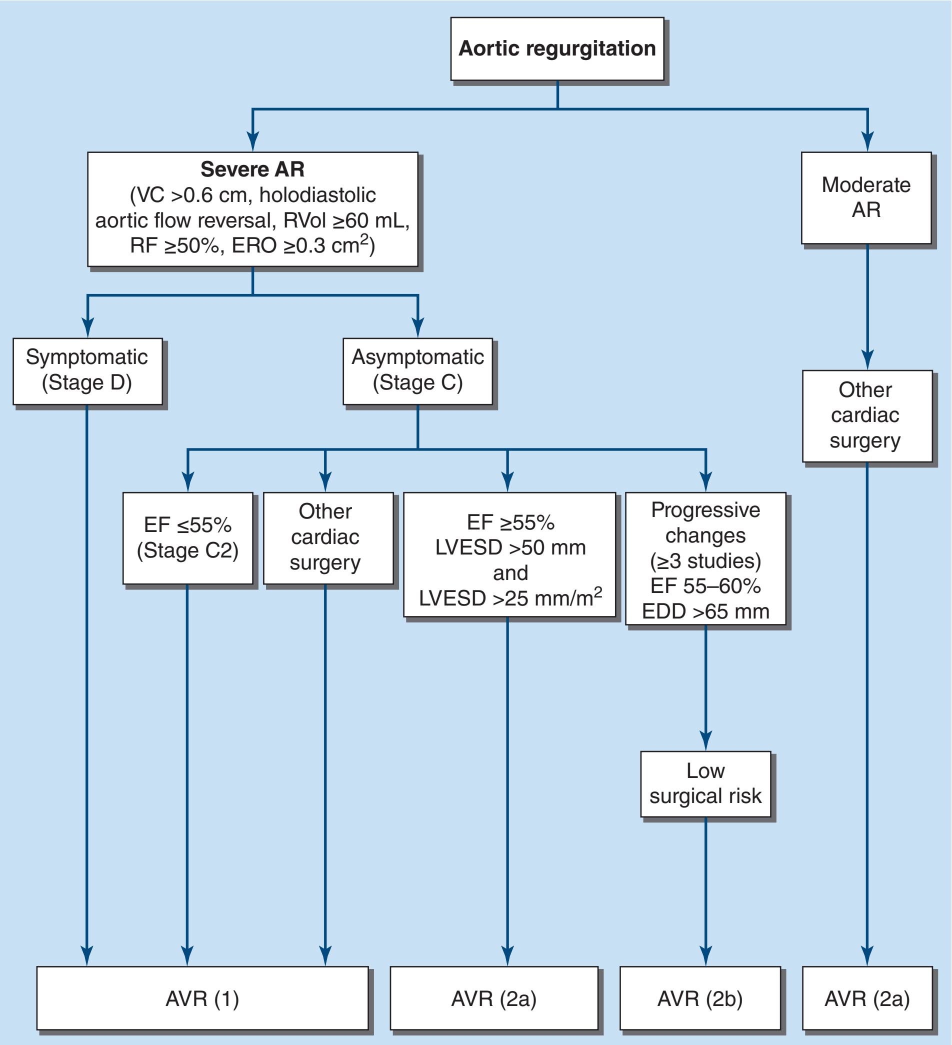

FIGURE 273-2 Management of patients with aortic regurgitation. See legend for **Fig. 272-4** for explanation of treatment recommendations (Class I, IIa, and IIb) and disease stages (B, C1, C2, and D). Preoperative coronary angiography should be performed routinely as determined by age, symptoms, and coronary risk factors. Cardiac catheterization and angiography may also be helpful when there is a discrepancy between clinical and noninvasive findings. Surgery is indicated for patients with severe AR and symptoms, LV dysfunction, or other indications for operation (e.g., aneurysm disease). Surgery is also reasonable once the LV indexed end-systolic dimension reaches or exceeds 25 mm/m². Patients who do not meet criteria for intervention should be monitored periodically with clinical and echocardiographic follow-up. AR, aortic regurgitation; AVR, aortic valve replacement (valve repair may be appropriate in selected patients); EDD, end-diastolic dimension; EF, ejection fraction; ERO, effective regurgitant orifice; LV, left ventricular; LVEDD, left ventricular end-diastolic dimension; LVEF, left ventricular ejection fraction; LVESD, left ventricular end-systolic dimension; RF, regurgitant fraction; RVol, regurgitant volume; VC, vena contracta. *(Reproduced with permission from CM Otto et al: 2020 AHA/ACC Guideline for management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2021;143(5):e72.)*

(defined as >1 year from onset of symptoms or LV dysfunction), surgical treatment often does not restore normal LV size and function. Therefore, in patients with chronic severe AR, careful clinical follow-up and noninvasive testing with echocardiography at ~6- to 12-month intervals are necessary if operation is to be undertaken at the optimal time, i.e., *coincident with or after* the onset of LV dysfunction but *prior to* the development of symptoms. Exercise testing may be helpful to assess effort tolerance more objectively and should be employed whenever questions about symptoms arise. Operation can be deferred as long as the patient both remains asymptomatic and retains normal LV function without severe or progressive chamber dilation.

Aortic valve replacement (AVR) is indicated for the treatment of severe AR in symptomatic patients irrespective of LV function. In general, the operation should be carried out in asymptomatic patients with severe AR and progressive LV dysfunction defined by an LVEF <55% on serial studies, an LV end-systolic dimension >50 mm (>25 mm/m²), or an LV diastolic dimension >65 mm. Smaller dimensions may be appropriate thresholds in individuals of smaller stature or when there is evidence of progressively decreasing LV function or increasing LV size on serial studies and the anticipated risks for surgical morbidity and mortality are low. Two case series from surgical referral centers have suggested that surgery should be performed at an even lower threshold for LV end-systolic dimension index (≥20 mm/m²), but data from randomized controlled trials are lacking. Abnormal LV GLS (≥ −18%) has been associated with an excess hazard for death in single-center studies. Observational studies using either three-dimensional echocardiography or CMR imaging have also suggested that event-free survival is reduced in asymptomatic patients with an LV end-systolic volume index ≥45 mL/m², compared with patients with an LV end-systolic volume index <45 mL/m². Patients with severe AR without indications for operation should be followed by clinical and echocardiographic examination every 6–12 months. Transcatheter aortic valve implantation (TAVI) is not recommended for patients with severe AR who are surgical candidates. Technical success with TAVI in patients with chronic AR is limited by the degree of aortic annular dilation and the relative paucity of valvular and annular calcium.

Surgical options for management of aortic valve and root disease have expanded considerably over the past decade. AVR with a suitable mechanical or tissue (biological) prosthesis is generally necessary in patients with rheumatic AR and in many patients with other causes of valvular AR. Rarely, when a leaflet has been perforated during IE

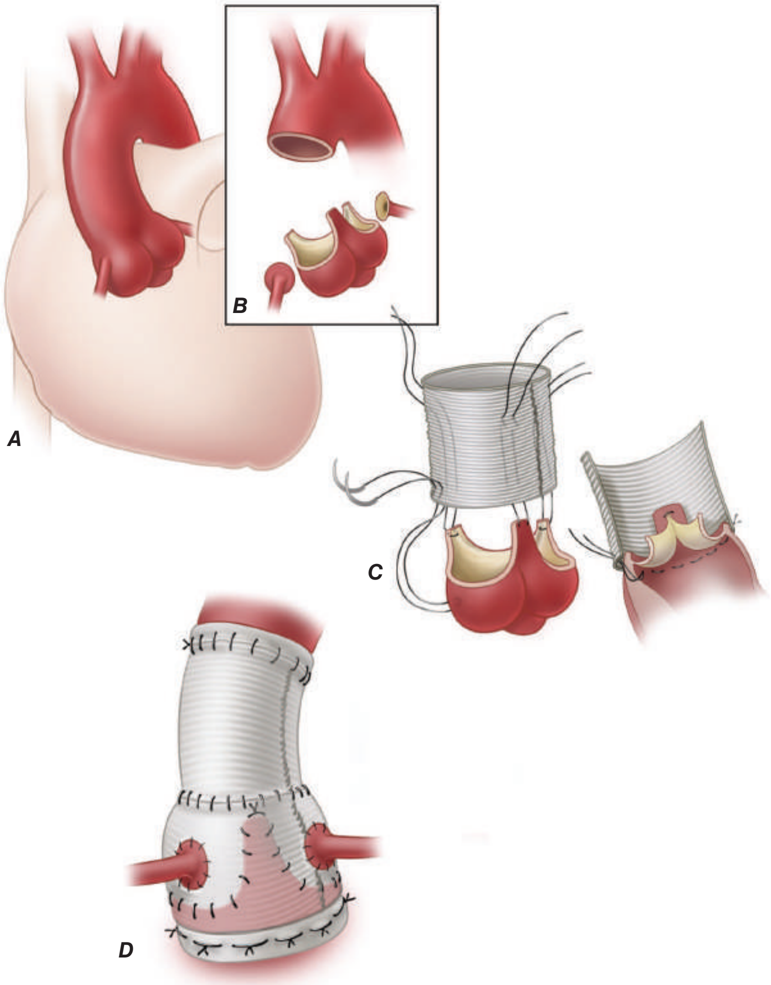

FIGURE 273-3 Valve-sparing aortic root reconstruction (David procedure). Aortic root and proximal ascending aorta (***A***) are resected (***B***) with sinuses of Valsalva and mobilized coronary artery buttons remaining. Subannular sutures (***C***) are placed, commissural posts are drawn up inside the valve, and the annular sutures are passed through the proximal end of the graft. The annular sutures are tied (***D***), the valve is reimplanted inside the graft, aortic continuity is reestablished with another graft of appropriate size, and the coronary buttons are attached to the side of the graft. *(From P Steltzer et al [eds]: Valvular Heart Disease: A Companion to Braunwald's Heart Disease, 3rd ed, Fig. 12-27, p. 200.)*

or torn from its attachments to the aortic annulus by thoracic trauma, primary surgical repair may be possible. When AR is due to aneurysmal dilation of the root or proximal ascending aorta rather than to primary valve involvement, it may be possible to reduce or eliminate the regurgitation by narrowing the annulus or by excising a portion of the aortic root without replacing the valve. Elective, valve-sparing aortic root reconstruction generally involves reimplantation of the valve in a contoured graft with reattachment of the coronary artery buttons into the side of the graft and is best undertaken in specialized surgical centers by experienced operators **(Fig. 273-3)**. Resuspension of the native aortic valve leaflets is possible in ~50% of patients with acute AR in the setting of type A aortic dissection. In other conditions, however, AR can be effectively eliminated only by replacing the aortic valve, as well as the dilated or aneurysmal ascending aorta responsible for the regurgitation, often using a composite prosthetic valve-graft conduit. Pure AR is not by itself a contraindication to the Ross procedure, although the presence of aortic annular dilation and/or an associated aortopathy would eliminate this option.

As is true in patients with other valvular heart disease, both operative and late mortality risks are largely dependent on the stage of the disease and myocardial function at the time of operation. The overall operative mortality rate for isolated AVR performed for pure AR is ~1–2%. The mortality risk doubles when aortic surgery is added to the operation. Patients with AR, marked cardiac enlargement, and established LV dysfunction experience an operative mortality rate of ~5–10% and a late mortality rate of ~3–5% per year due to LV failure despite a technically satisfactory operation. Nonetheless, because of the very poor prognosis with medical management, even patients with advanced LV systolic dysfunction should be considered for operation.

Patients with acute severe AR require prompt (24–48 h) surgical treatment, which may be lifesaving.

■ FURTHER READING

Hashimoto G et al: Association of left ventricular remodeling assessment by cardiac magnetic resonance with outcomes in patients with chronic aortic regurgitation. JAMA Cardiol 7:924, 2022.

Lacro RV et al: Atenolol versus losartan in children and young adults with Marfan's syndrome. N Engl J Med 371:2061, 2014.

Malaisrie SC, McCarthy PM: Surgical approach to disease of the aortic valve and the aortic root, in *Valvular Heart Disease: A Companion to Braunwald's Heart Disease*, 5th ed. CM Otto, RO Bonow (eds). Philadelphia, Elsevier Saunders, 2020, pp 267–288.

Otto CM et al: 2020 ACC/AHA guideline for the management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 143:e72, 2021.

Yang L-T et al: Association of echocardiographic left ventricular end-systolic volume and volume-derived ejection fraction with outcome in asymptomatic chronic aortic regurgitation. JAMA Cardiol 6:189, 2021.

274 Mitral Stenosis

Patrick T. O'Gara, Joseph Loscalzo

The role of the physical examination in the evaluation of patients with valvular heart disease is also considered in **Chaps. 44 and 246**; of electrocardiography (ECG) in **Chap. 247**; of echocardiography and other noninvasive imaging techniques in **Chap. 248**; and of cardiac catheterization and angiography in **Chap. 249**.

MITRAL STENOSIS

ETIOLOGY AND PATHOLOGY

Rheumatic fever is the leading cause of mitral stenosis (MS) **(Table 274-1; see also Chap. 371)**. Other less common etiologies of obstruction to left ventricular inflow include congenital mitral valve stenosis, cor triatriatum, mitral annular calcification with extension onto the leaflets, systemic lupus erythematosus, rheumatoid arthritis, left atrial myxoma, and infective endocarditis with large vegetations. Pure or predominant MS occurs in ~40% of all patients with rheumatic heart disease and a history of rheumatic fever **(Chap. 371)**. In other patients with rheumatic heart disease, lesser degrees of MS may accompany mitral regurgitation (MR) and aortic valve disease. With reductions in the incidence of acute rheumatic fever, particularly in temperate climates and middle- to high-income countries, the incidence of MS has declined considerably over the past several decades. However, it remains a major problem in low-income countries, especially in sub-Saharan Africa, India, Southeast Asia, and Oceania **(Chap. 272)**.

In rheumatic MS, chronic inflammation leads to diffuse thickening of the valve leaflets with formation of fibrous tissue often with calcific deposits. The mitral commissures fuse, the chordae tendineae fuse and shorten, the valvular cusps become rigid, and the pathologic process eventually leads to narrowing at the apex of the funnel-shaped ("fish-mouth") valve. Although the initial insult to the mitral valve is rheumatic, later changes may be exacerbated by inflammation, fibrosis, and trauma due to altered flow patterns. Calcification of the stenotic mitral valve immobilizes the leaflets and narrows the orifice further. Thrombus formation and arterial embolization may arise from the calcific valve itself, but in patients with atrial fibrillation (AF), thrombi arise more frequently from the dilated left atrium (LA), particularly from within the LA appendage.

PATHOPHYSIOLOGY

In normal adults, the area of the mitral valve orifice is 4–6 cm^2. In the presence of significant obstruction, i.e., when the orifice area is reduced to <~2 cm^2, blood can flow from the LA to the left ventricle (LV) only if propelled by an abnormally elevated left atrioventricular pressure gradient, the hemodynamic hallmark of MS. When the mitral valve opening is reduced to <1.5 cm^2, referred to as "severe" MS, an LA pressure of ~25 mmHg is required to maintain a normal cardiac output (CO). The elevated pulmonary venous and pulmonary arterial (PA) wedge pressures reduce pulmonary compliance, contributing to exertional dyspnea. The first bouts of dyspnea are usually precipitated by clinical events that increase the rate of blood flow across the mitral orifice, resulting in further elevation of the LA pressure (see below).

To assess the severity of obstruction hemodynamically, both the transvalvular pressure gradient and the flow rate must be measured **(Chap. 249)**. The latter depends not only on the CO but on the heart rate, as well. An increase in heart rate shortens diastole proportionately more than systole and diminishes the time available for flow across the mitral valve. Therefore, at any given level of CO, tachycardia, including that associated with rapid AF, augments the transvalvular pressure gradient and elevates further the LA pressure. Similar considerations apply to the pathophysiology of tricuspid stenosis (TS).

The LV diastolic pressure and ejection fraction (EF) are normal in isolated MS. In MS and sinus rhythm, the elevated LA and PA wedge pressures exhibit a prominent atrial contraction pattern (*a* wave) and a gradual pressure decline after the *v* wave and mitral valve opening (*y* descent). In severe MS and whenever pulmonary vascular resistance is significantly increased, the PA pressure (PAP) is elevated at rest and rises further during exercise, often causing secondary elevations of right ventricular (RV) end-diastolic pressure and volume.

Cardiac Output In patients with severe MS (mitral valve orifice 1–1.5 cm^2), the CO is normal or almost so at rest, but rises subnormally during exertion. In patients with very severe MS (valve area <1 cm^2), particularly those in whom pulmonary vascular resistance is markedly elevated, the CO is subnormal at rest and may fail to rise or may even decline during activity.

Pulmonary Hypertension The clinical and hemodynamic features of MS are influenced importantly by the level of the PAP. Pulmonary hypertension results from (1) passive backward transmission of the elevated LA pressure; (2) pulmonary arteriolar constriction (the so-called "second stenosis"), which presumably is triggered by LA and pulmonary venous hypertension (reactive pulmonary hypertension); (3) interstitial edema in the walls of the small pulmonary vessels; and (4) at end stage, organic obliterative changes in the pulmonary vascular bed. Severe pulmonary hypertension results in RV enlargement, secondary tricuspid regurgitation (TR), and pulmonic regurgitation (PR), as well as right-sided heart failure.

TABLE 274-1 Major Causes of Mitral Stenosis

Etiologies
Rheumatic fever
Congenital (parachute valve, cor triatriatum)
Severe mitral annular calcification with leaflet involvement
SLE, RA
Myxoma
IE with large vegetations

Abbreviations: IE, infective endocarditis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus.

SYMPTOMS

In temperate climates, the latent period between the initial attack of rheumatic carditis (in the increasingly rare circumstances in which a history of one can be elicited) and the development of symptoms due to MS is generally about two decades; most patients begin to experience disability in the fourth decade of life. Studies carried out before the development of surgical mitral valvotomy revealed that once a patient with MS became seriously symptomatic, the disease progressed inexorably to death within 2–5 years.

In patients whose mitral orifices are large enough to accommodate a normal blood flow with only mild elevations of LA pressure, marked elevations of this pressure leading to dyspnea and cough may be precipitated by sudden changes in the heart rate, volume status, or CO, as, for example, with severe exertion, excitement, fever, severe anemia, paroxysmal AF and other tachycardias, sexual intercourse, pregnancy, and thyrotoxicosis. As MS progresses, lesser degrees of stress precipitate dyspnea, the patient becomes limited in daily activities, and orthopnea and paroxysmal nocturnal dyspnea develop. The development of persistent AF often marks a turning point in the patient's course and is generally associated with acceleration of the rate at which symptoms progress. *Hemoptysis* **(Chap. 41)** results from rupture of pulmonary-bronchial venous connections secondary to pulmonary venous hypertension. It occurs most frequently in patients who have elevated LA pressures without markedly elevated pulmonary vascular resistances and is rarely fatal. *Recurrent pulmonary emboli* **(Chap. 290)**, sometimes with infarction, are an important cause of morbidity and mortality late in the course of MS and often arise from right atrial mural thrombus. *Pulmonary infections*, i.e., bronchitis, bronchopneumonia, and lobar pneumonia, commonly complicate untreated MS, especially during the winter months. In patients with significant left atrial enlargement, cardiovocal syndrome (Ortner's syndrome) may develop with hoarseness secondary to left recurrent laryngeal nerve compression.

Pulmonary Changes In addition to the aforementioned changes in the pulmonary vascular bed, fibrous thickening of the walls of the alveoli and pulmonary capillaries occurs commonly in MS. The vital capacity, total lung capacity, maximal breathing capacity, and oxygen uptake per unit of ventilation are reduced **(Chap. 296)**. Pulmonary compliance falls further as pulmonary capillary pressure rises during exercise.

Thrombi and Emboli *Thrombi* may form in the left atria, particularly within the enlarged atrial appendages of patients with MS. Systemic embolization, the incidence of which is 10–20%, occurs more frequently in patients with AF, in patients >65 years of age, and in those with a reduced CO. However, systemic embolization may be the presenting feature in otherwise asymptomatic patients with only mild MS.

PHYSICAL FINDINGS

(See also Chaps. 44 and 246).

Inspection and Palpation In patients with severe MS, there may be a malar flush with pinched and blue facies. In patients with sinus rhythm and severe pulmonary hypertension or associated TS, the jugular venous pulse reveals prominent *a* waves due to vigorous right atrial systole. The systemic arterial pressure is usually normal or slightly low. A parasternal lift signifies an enlarged RV. A diastolic thrill may very rarely be present at the cardiac apex, with the patient in the left lateral recumbent position.

Auscultation The first heart sound (S_1) is usually accentuated in the early stages of the disease and slightly delayed. The pulmonic component of the second heart sound (P_2) also is often accentuated with elevated PAPs, and the two components of the second heart sound (S_2) are closely split. The opening snap (OS) of the mitral valve is most readily audible in expiration at, or just medial to, the cardiac apex. This sound generally follows the sound of aortic valve closure (A_2) by 0.05–0.12 s. The time interval between A_2 and OS varies inversely with the severity of the MS. The OS is followed by a low-pitched, rumbling, diastolic murmur, heard best at the apex with the patient in the left lateral recumbent position **(see Fig. 246-5)**; it is accentuated by mild exercise (e.g., a few rapid sit-ups) carried out just before auscultation. In general, the duration of this murmur correlates with the severity of the stenosis in patients with preserved CO. In patients with sinus rhythm, the murmur often reappears or becomes louder during atrial systole (presystolic accentuation). Soft, grade I or II/VI systolic murmurs may be heard at or medial to the apex and may signify mixed mitral valve disease with regurgitation. Hepatomegaly, ankle edema, ascites, and pleural effusion, particularly in the right pleural cavity, may occur in patients with MS and RV failure.

Associated Lesions With severe pulmonary hypertension, a pansystolic murmur produced by functional TR may be audible along the left sternal border. This murmur is usually louder during inspiration and diminishes during forced expiration (Carvallo's sign). When the CO is markedly reduced in MS, the typical auscultatory findings, including the diastolic rumbling murmur, may not be detectable (silent MS), but they may reappear as compensation is restored. The *Graham Steell murmur* of PR, a high-pitched, diastolic, decrescendo blowing murmur along the left sternal border, results from dilation of the pulmonary valve ring and occurs in patients with mitral valve disease and severe pulmonary hypertension. This murmur may be indistinguishable from the more common murmur produced by aortic regurgitation (AR), although it may increase in intensity with inspiration and is accompanied by a loud and often palpable P_2.

LABORATORY EXAMINATION

ECG In MS and sinus rhythm, the P wave usually suggests LA enlargement **(see Fig. 247-8)**. It may become tall and peaked in lead II and upright in lead V_1 when severe pulmonary hypertension or TS complicates MS and right atrial (RA) enlargement develops. The QRS complex is usually normal. However, with severe pulmonary hypertension, right axis deviation and RV hypertrophy are often present.

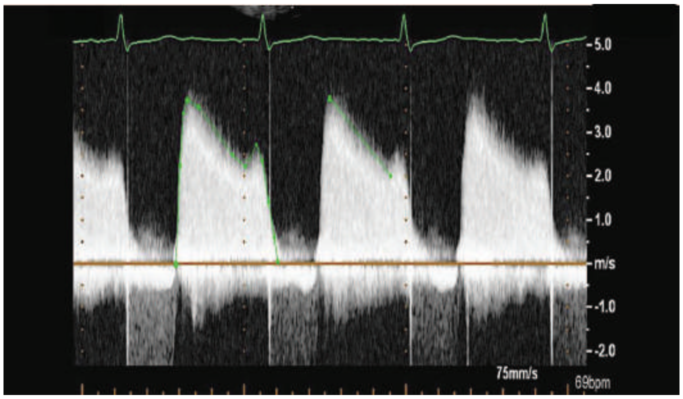

FIGURE 274-1 Continuous wave Doppler interrogation of transmitral valve velocities in a patient with severe rheumatic mitral stenosis. Electrocardiogram on top. Vertical scale in meters/second. Horizontal scale in seconds at sweep speed of 75 mm/se. Velocity is converted to pressure using the Bernoulli equation. In this example, the mean mitral valve gradient (area under the green tracing) is calculated to 38 mmHg and mitral valve area to 0.7 cm^2.

Echocardiogram (See also Chap. 248) Transthoracic echocardiography (TTE) with color flow and spectral Doppler imaging provides critical information, including measurements of mitral inflow velocity during early (E wave) and late (A wave in patients in sinus rhythm) diastolic filling, estimates of the transvalvular peak and mean gradients and mitral orifice area **(Fig. 274-1)**, the presence and severity of any associated MR, the extent of leaflet calcification and restriction, the degree of distortion of the subvalvular apparatus, and the anatomic suitability for percutaneous mitral balloon commissurotomy (PMBC; see below). In addition, TTE provides an assessment of LV and RV function, chamber sizes, an estimation of the PA systolic pressure based on the tricuspid regurgitant jet velocity, and an indication of the presence and severity of any associated valvular lesions, such as aortic stenosis (AS) and/or regurgitation. Transesophageal echocardiography (TEE) provides superior images and should be used when TTE is inadequate for guiding management decisions. TEE is especially indicated to exclude the presence of LA thrombus prior to PMBC. The performance of TTE with exercise to evaluate the mean mitral diastolic gradient, PAPs, and RV function can be very helpful in the evaluation of patients with MS when there is a discrepancy between the clinical findings and the resting hemodynamics.

Chest X-Ray The earliest changes are straightening of the upper left border of the cardiac silhouette, prominence of the main PAs, dilation of the upper lobe pulmonary veins, and posterior displacement of the esophagus by an enlarged LA. Kerley B lines are fine, dense, opaque, horizontal lines that are most prominent in the lower and mid-lung fields that result from distention of interlobular septae and lymphatics with edema when the resting mean LA pressure exceeds ~20 mmHg.

DIFFERENTIAL DIAGNOSIS

Like MS, significant MR may also be associated with a prominent diastolic murmur at the apex due to increased antegrade transmitral flow, but in patients with isolated MR, this diastolic murmur commences slightly later than in patients with MS, and there is often clear-cut evidence of LV enlargement. An OS and increased P_2 are absent, and S_1 is soft or absent. An apical pansystolic murmur of at least grade III/VI intensity as well as an S_3 suggests significant MR. Similarly, the apical mid-diastolic murmur associated with severe AR (*Austin Flint murmur*) may be mistaken for MS but can be differentiated from it because it is not intensified in pre-systole and becomes softer with administration of amyl nitrite or other arterial vasodilators. TS, which occurs rarely in the absence of MS, may mask many of the clinical features of MS or be clinically silent; when present, the diastolic murmur of

TS increases with inspiration and the *y* descent in the jugular venous pulse is delayed.

Atrial septal defect **(Chap. 280)** may be mistaken for MS; in both conditions, there is often clinical, ECG, and chest x-ray evidence of RV enlargement and accentuation of pulmonary vascularity. However, the absence of LA enlargement and of Kerley B lines and the demonstration of fixed splitting of S_2 with a grade II or III mid-systolic murmur at the mid to upper left sternal border all favor atrial septal defect over MS. Atrial septal defects with large left-to-right shunts may result in functional TS because of the enhanced diastolic flow. An incomplete right bundle branch block pattern on ECG is often present.

Left atrial myxoma **(Chap. 282)** may obstruct LA emptying, causing dyspnea, a diastolic murmur, and hemodynamic changes resembling those of MS. However, patients with an LA myxoma often have features suggestive of a systemic disease, such as weight loss, fever, anemia, systemic emboli, and elevated serum IgG and interleukin 6 (IL-6) concentrations. The auscultatory findings may change markedly with body position. The diagnosis can be established by the demonstration of a characteristic echo-producing mass in the LA with TTE.

■ CARDIAC CATHETERIZATION

Left and right heart catheterization can be useful when there is a discrepancy between the clinical and noninvasive findings, including those from TEE and exercise echocardiographic testing when appropriate. Catheterization can also be helpful in assessing associated lesions, such as AS and AR, and in patients with recurring or worsening symptoms late after mitral valve intervention. Computed tomographic coronary angiography is increasingly used to screen preoperatively for the presence of coronary artery disease in appropriate patients prior to heart valve surgery or transcatheter treatment.

TREATMENT

Mitral Stenosis (Fig. 274-2)

Penicillin prophylaxis of group A β-hemolytic streptococcal infections **(Chap. 371)** for secondary prevention of rheumatic fever is important for at-risk patients with rheumatic MS. Recommendations for infective endocarditis prophylaxis are similar to those for other valve lesions and are restricted to patients at high risk for complications from infection, including patients with a history of endocarditis. In symptomatic patients, some improvement usually occurs with restriction of sodium intake and small doses of oral diuretics. Beta blockers, nondihydropyridine calcium channel blockers (e.g., verapamil or diltiazem), and digitalis glycosides are useful in slowing the ventricular rate of patients with AF. Vitamin K antagonist therapy (such as warfarin) targeted to an international normalized ratio (INR) of 2–3 should be administered indefinitely to patients with MS who have AF, a history of thromboembolism, or demonstrated LA thrombus. The routine use of a vitamin K antagonist in patients in sinus rhythm with LA enlargement (maximal dimension >5.5 cm) with or without spontaneous echo contrast is more controversial. In a randomized trial of patients with rheumatic MS and AF, there was a significantly higher incidence of death among patients treated with rivaroxaban than with vitamin K antagonist therapy. A vitamin K antagonist is

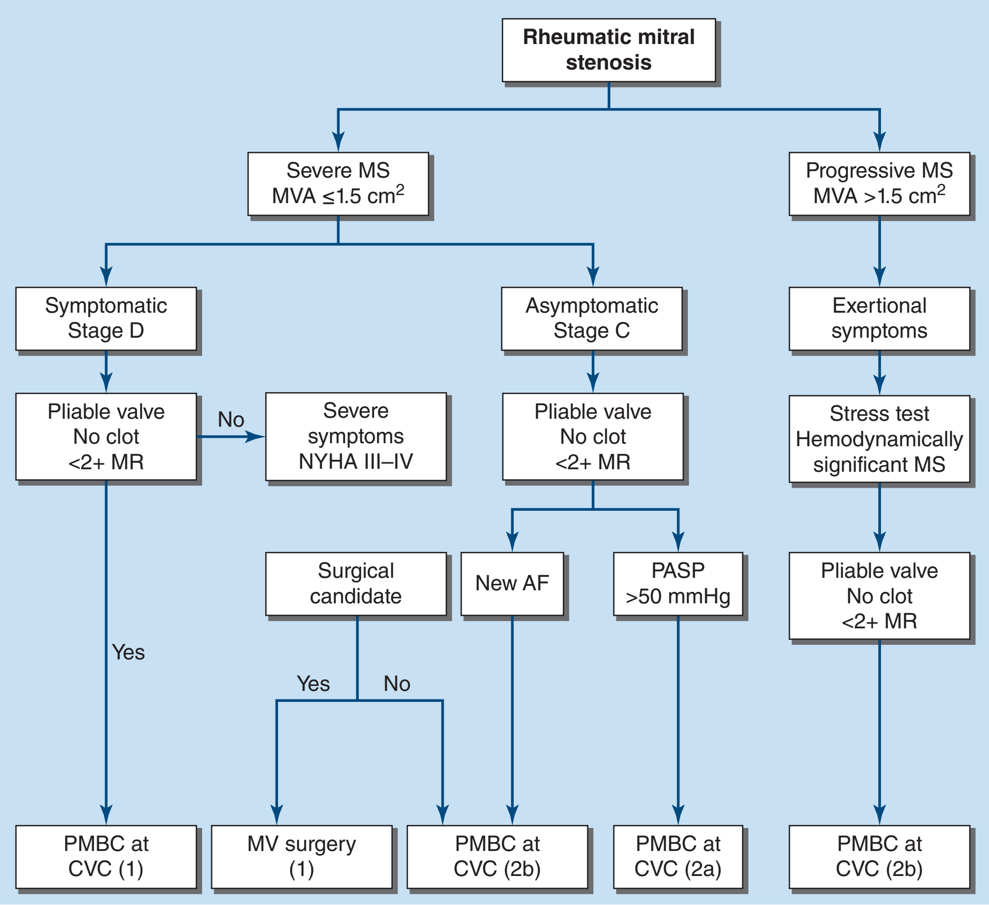
Rheumatic mitral stenosis
Severe MS MVA ≤1.5 cm²
Progressive MS MVA >1.5 cm²
Symptomatic Stage D
Asymptomatic Stage C
Exertional symptoms
Pliable valve No clot <2+ MR
No
Severe symptoms NYHA III–IV
Pliable valve No clot <2+ MR
Stress test Hemodynamically significant MS
Surgical candidate
New AF
PASP >50 mmHg
Pliable valve No clot <2+ MR
Yes
Yes
No
PMBC at CVC (1)
MV surgery (1)
PMBC at CVC (2b)
PMBC at CVC (2a)
PMBC at CVC (2b)

FIGURE 274-2 Management of rheumatic mitral stenosis. See legend for **Fig. 272-4** for explanation of treatment recommendations (Class I, IIa, IIb) and disease stages (C, D). Preoperative coronary angiography should be performed routinely as determined by age, symptoms, and coronary risk factors. Cardiac catheterization and angiography may also be helpful when there is a discrepancy between clinical and noninvasive findings. AF, atrial fibrillation; CVC, comprehensive valve center; MR, mitral regurgitation; MS, mitral stenosis; MV, mitral valve; MVA, mitral valve area; MVR, mitral valve surgery (repair or replacement); NYHA, New York Heart Association; PASP, pulmonary arterial systolic pressure; PMBC, percutaneous mitral balloon commissurotomy. *(Reproduced with permission from CM Otto et al: ACC/AHA guideline for the management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 143:e72, 2021.)*

recommended to reduce the risk for stroke or systemic embolism in at-risk patients with rheumatic MS.

If AF is of relatively recent onset in a patient whose MS is not severe enough to warrant PMBC or surgical intervention, reversion to sinus rhythm pharmacologically or by means of electrical countershock is indicated. Usually, cardioversion should be undertaken after the patient has had at least 3 consecutive weeks of anticoagulant treatment to a therapeutic INR. If cardioversion is indicated more urgently, then intravenous heparin should be provided and TEE performed to exclude the presence of LA thrombus before the procedure. Conversion to sinus rhythm is rarely successful or sustained in patients with severe MS, particularly those in whom the LA is significantly enlarged or in whom AF has been present for >1 year, conditions that favor the development of an LA myopathy.

MITRAL COMMISSUROTOMY

Unless there is a contraindication, mitral commissurotomy is indicated in symptomatic (New York Heart Association [NYHA] functional class II–IV) patients with isolated severe MS, whose effective orifice (valve area) is <~1 cm^2/m^2 body surface area or <1.5 cm^2 in normal-sized adults. Mitral commissurotomy can be carried out either percutaneously or surgically. In PMBC **(Figs. 274-3 and 274-4)**, a catheter is directed into the LA after transseptal puncture, and a single balloon is directed across the valve and inflated in the valvular orifice. Ideal patients have relatively pliable leaflets with little or no commissural calcium. In addition, the subvalvular structures should not be significantly scarred or thickened, and there should be no LA thrombus. Any associated MR should be of ≤2+/4+ severity. The short- and long-term results of this procedure in appropriate patients are similar to those of surgical commissurotomy, but with less morbidity and a lower periprocedural mortality rate. Event-free survival in younger (<45 years) patients with pliable valves is excellent, with rates as high as 80–90% over 3–7 years. Therefore, PMBC is the procedure of choice for such patients when it can be performed by a skilled operator in a high-volume center.

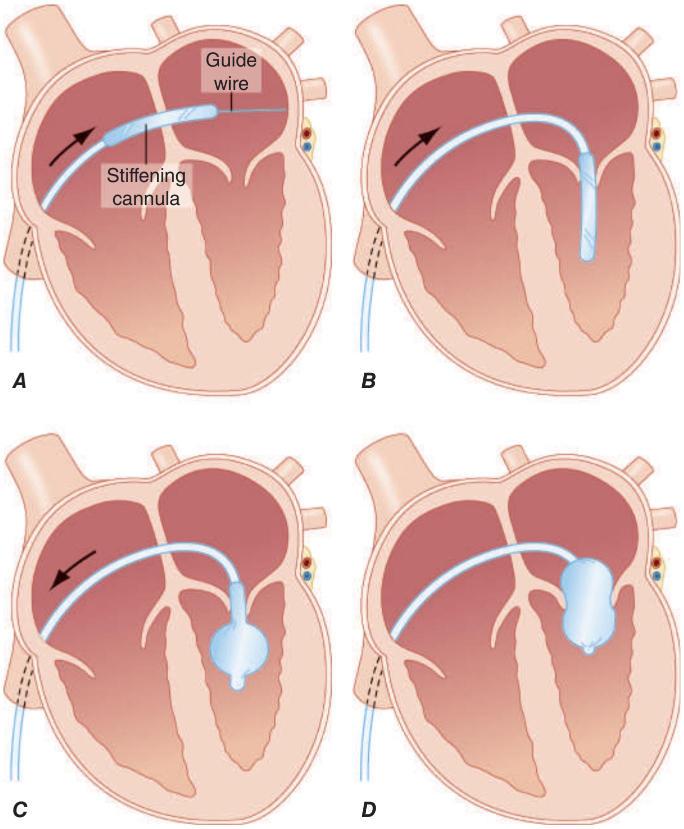

FIGURE 274-3 Inoue balloon technique for percutaneous mitral balloon commissurotomy. ***A.*** After transseptal puncture, the deflated balloon catheter is advanced across the interatrial septum, then across the mitral valve and into the left ventricle. ***B–D.*** The balloon is inflated stepwise within the mitral orifice.

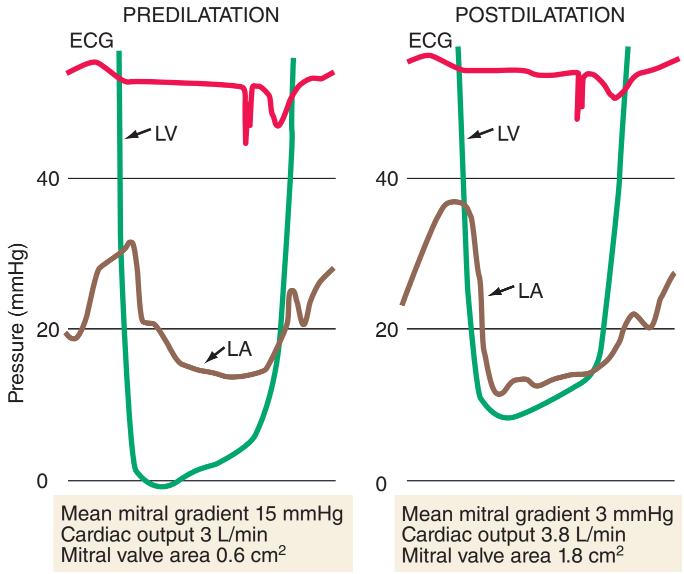

FIGURE 274-4 Simultaneous left atrial (LA) and left ventricular (LV) pressure before and after percutaneous mitral balloon commissurotomy (PMBC) in a patient with severe mitral stenosis. ECG, electrocardiogram. *(Courtesy of Raymond G. McKay, MD.)*

TTE is helpful in identifying patients for the percutaneous procedure; TEE is performed routinely to exclude LA thrombus and to assess the degree of MR at the time of the scheduled procedure. An "echo score" has been developed to help guide decision-making. The score accounts for the degree of leaflet thickening, calcification, and mobility, and for the extent of subvalvular thickening. A lower score predicts a higher likelihood of successful PMBC.

In patients in whom PMBC is not possible or unsuccessful, or in many patients with restenosis after previous surgery, an "open" surgical commissurotomy using cardiopulmonary bypass is necessary. In addition to opening the valve commissures, it is important to loosen any subvalvular fusion of papillary muscles and chordae tendineae; to remove large deposits of calcium, thereby improving valvular function; and to remove atrial thrombi. The perioperative mortality rate for this type of mitral valve repair procedure is ~2%.

Successful commissurotomy is defined by a 50% reduction in the mean mitral valve gradient and a doubling of the mitral valve area. Successful commissurotomy, whether balloon or surgical, usually results in striking symptomatic and hemodynamic improvement and prolongs survival. However, there is no evidence that the procedure improves the prognosis of patients with slight or no functional impairment. Therefore, unless recurrent systemic embolization or severe pulmonary hypertension has occurred (PA systolic pressures >50 mmHg at rest or >60 mmHg with exercise), commissurotomy is *not* recommended for patients who are asymptomatic and/or who have mild or moderate stenosis (mitral valve area >1.5 cm^2). When there is little symptomatic improvement after commissurotomy, it is likely that the procedure was ineffective, that it induced MR, or that associated valvular or myocardial disease was present. About half of all patients undergoing surgical mitral commissurotomy require reoperation by 10 years. In the pregnant patient with MS, commissurotomy should be carried out if pulmonary congestion occurs despite intensive medical treatment. PMBC is the preferred strategy in this setting and is performed with TEE and no or minimal x-ray exposure.

Mitral valve replacement (MVR) is necessary in patients with MS and significant associated MR, those in whom the valve has been severely distorted by previous transcatheter or operative manipulation, or those in whom the surgeon does not find it possible to improve valve function significantly with commissurotomy. MVR is now routinely performed with preservation of the chordal attachments to optimize LV functional recovery. Perioperative mortality rates with MVR vary with age, LV function, the presence of CAD, and associated comorbidities. They average 2–5% overall but are lower in young patients and may be twice as high in patients >65 years of age with significant comorbidities. Because there are also long-term complications of valve replacement, patients in whom preoperative evaluation suggests the possibility that MVR may be required should be operated on only if they have severe MS—i.e., an orifice area ≤1.5 cm^2—and are in NYHA class III, i.e., symptomatic with ordinary activity despite optimal medical therapy. The overall 10-year survival of surgical survivors is ~70%. Long-term prognosis is worse in patients >65 years of age and those with marked disability and marked depression of the CO preoperatively. Pulmonary hypertension and RV dysfunction are additional risk factors for poor outcome.

FURTHER READING

Connolly SJ et al: Rivaroxaban in rheumatic heart disease-associated atrial fibrillation. N Engl J Med 387:978, 2022.

Nishimura RA et al: Mitral valve disease: Current management and future challenges. Lancet 387:1324, 2016.

Otto CM et al: 2020 ACC/AHA guideline for the management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 143:e72, 2021.

275

Mitral Regurgitation

Patrick T. O'Gara, Joseph Loscalzo

The role of the physical examination in the evaluation of patients with valvular heart disease is also considered in **Chaps. 44 and 246**; of electrocardiography (ECG) in **Chap. 247**; of echocardiography and other noninvasive imaging techniques in **Chap. 248**; and of cardiac catheterization and angiography in **Chap. 249**.

ETIOLOGY

Mitral regurgitation (MR) may result from an abnormality or disease process that affects any one or more of the five functional components of the mitral valve apparatus (leaflets, annulus, chordae tendineae, papillary muscles, and subjacent myocardium) **(Table 275-1)**. Acute MR can occur in the setting of acute myocardial infarction (MI) with papillary muscle rupture **(Chap. 286)**, following blunt chest wall trauma, or during the course of infective endocarditis (IE) owing to leaflet perforation or destruction. With acute MI, the posteromedial papillary muscle is involved much more frequently than the anterolateral papillary muscle because of its singular blood supply. Transient, acute MR can occur during periods of active ischemia and bouts of angina pectoris. Rupture of chordae tendineae can result in "acute-on-chronic MR" in patients with myxomatous degeneration of the valve apparatus.

TABLE 275-1 Major Causes of Mitral Regurgitation (MR)

Etiologies
Acute
IE
Papillary muscle rupture (post-MI)
Chordal rupture/leaflet flail (MVP, IE)
Blunt trauma
Chronic
Primary (affecting leaflets, chordae)
Myxomatous (MVP, Barlow's, *forme fruste*)
Rheumatic fever
IE (healed)
Congenital (cleft, AV canal)
Radiation
Secondary (leaflets, chordae are "innocent bystanders")
Ischemic cardiomyopathy
Dilated cardiomyopathy
HOCM (with SAM)
AF with LA enlargement and annular dilation (atrial functional MR)
Mitral annular calcification[a]

[a] Mitral annular calcification may include elements of both primary and secondary MR (mixed) as the disease process may encroach on the leaflets, impair the normal sphincteric function of the annulus, or both. There are additional examples of "mixed" secondary MR such as the coexistence of MVP with an ischemic cardiomyopathy.

Abbreviations: AF, atrial fibrillation; AV, atrioventricular; HOCM, hypertrophic obstructive cardiomyopathy; IE, infective endocarditis; LA, left atrial; LV, left ventricular; MI, myocardial infarction; MVP, mitral valve prolapse; SAM, systolic anterior motion.

Chronic MR can result from several disease processes (Table 275-1). Distinction should be drawn between primary (degenerative) MR, in which the leaflets and/or chordae tendineae are primarily responsible for abnormal valve function, and secondary (functional) MR, in which the leaflets and chordae tendineae are usually normal but the regurgitation is caused by left ventricular (LV) and/or left atrial remodeling, annular dilation, papillary muscle displacement, dyssynchrony, leaflet tethering, or their combination. Patient assessment, treatment approach, and long-term prognosis differ significantly between primary and secondary MR. Mitral valve prolapse (MVP) is discussed more extensively in **Chap. 276**. The rheumatic process produces rigidity, deformity, and retraction of the valve cusps and commissural fusion, as well as shortening, contraction, and fusion of the chordae tendineae. MR can persist after resolution of the acute phase of infection and inflammation. MR may occur as a congenital anomaly **(Chap. 280)**, most commonly as a defect of the endocardial cushions (atrioventricular cushion defects). A cleft anterior mitral valve leaflet accompanies ostium primum atrial septal defect. Radiation can result in leaflet thickening, retraction, and calcification, often in association with annular and chordal involvement and some degree of mitral stenosis. Chronic MR occurs frequently after prior MI(s) associated with changes in LV size, shape, and function. Similar mechanisms of annular dilation and ventricular remodeling contribute to the MR that occurs among patients with nonischemic forms of dilated cardiomyopathy once the LV end-diastolic dimension reaches 6 cm. The MR associated with hypertrophic obstructive cardiomyopathy (HOCM) is usually dynamic in nature and dependent on systolic anterior motion of the anterior mitral valve leaflet into a narrowed LV outflow tract. Patients with chronic persistent atrial fibrillation (AF) may develop atrial remodeling and annular dilation with inadequate leaflet lengthening and MR (atrial functional MR). Secondary MR due to LV remodeling is more frequently encountered in the community than secondary MR that occurs in association with AF and annular dilation. Annular calcification can result in MR when it encroaches on the leaflets or results in decreased sphincteric function, is especially prevalent among patients with advanced renal disease, and is commonly observed in women >65 years of age with hypertension and diabetes mellitus. Irrespective of cause, chronic severe MR is often progressive because enlargement of the left atrium (LA) places tension on the posterior mitral leaflet, pulling it further away from the mitral orifice and thereby aggravating the valvular dysfunction. Similarly, LV dilation increases the regurgitation, which, in turn, enlarges the LA and LV further, resulting in a vicious cycle; hence the aphorism, "MR begets MR."

PATHOPHYSIOLOGY

The resistance to LV emptying (LV afterload) is reduced in patients with MR. As a consequence, the LV is decompressed into the LA during ejection, and with the reduction in LV size during systole, there is a rapid decline in LV tension. The initial compensation to MR is more complete LV emptying. However, LV volume increases progressively with time as the severity of the regurgitation increases and as LV contractile function deteriorates. This increase in LV volume is often accompanied by a reduced forward cardiac output (CO). LV compliance is often increased, and thus, LV diastolic pressure does not increase until late in the course. The regurgitant volume varies directly with the LV systolic pressure and the size of the regurgitant orifice; the latter, in turn, is influenced by the extent of LV and mitral annular dilation, as well as by leaflet morphology. Because ejection fraction (EF) rises in severe MR in the presence of normal LV function, even a modest reduction in this parameter (<60%) reflects significant contractile dysfunction.

During early diastole, as the distended LA empties, there is a particularly rapid *y* descent in the absence of accompanying MS. A brief, early diastolic LA-LV pressure gradient (often generating a rapid filling sound [S_3] and mid-diastolic murmur masquerading as MS) may occur in patients with pure, severe MR as a result of the very rapid flow of blood across a normal-sized mitral orifice.

Measurements of LV ejection fraction (LVEF), CO, pulmonary arterial (PA) systolic pressure, regurgitant volume, regurgitant fraction (RF), and the effective regurgitant orifice area can be obtained during a careful Doppler echocardiographic examination. These measurements can also be obtained accurately with cardiac magnetic resonance (CMR) imaging, although this technology is not widely available. Left and right heart catheterization with contrast ventriculography is used less frequently. Chronic, severe MR is defined by a regurgitant volume ≥60 mL/beat, RF ≥50%, and effective regurgitant orifice area ≥0.40 cm^2. In patients with secondary MR, in whom the severity of MR can be underappreciated using echocardiographic/Doppler techniques, lesser degrees of regurgitation may carry relatively greater prognostic weight. The adverse prognosis in secondary MR related to adverse LV remodeling is intimately related to the degree of myocardial dysfunction.

LA Compliance In acute severe MR, the regurgitant volume is delivered into a normal-sized LA having normal or reduced compliance. As a result, LA pressures rise markedly for any increase in LA volume. The *v* wave in the LA pressure pulse is usually prominent, LA and pulmonary venous pressures are markedly elevated, and pulmonary edema is common. Because of the rapid rise in LA pressures during ventricular systole, the murmur of acute MR is early in timing and decrescendo in configuration ending well before S_2, as a reflection of the progressive diminution in the LV-LA pressure gradient. LV systolic function in acute MR may be normal, hyperdynamic, or reduced, depending on the clinical context.

Patients with chronic severe MR, on the other hand, develop marked LA enlargement and *increased* LA compliance with little if any increase in LA and pulmonary venous pressures for any increase in LA volume. The LA *v* wave is relatively less prominent. The murmur of chronic MR is classically holosystolic in timing and plateau in configuration, as a reflection of the near-constant LV-LA pressure gradient. These patients usually complain of severe fatigue and exhaustion secondary to a low forward CO, whereas symptoms resulting from pulmonary congestion are less prominent initially; AF is almost invariably present once the LA dilates significantly.

■ SYMPTOMS

Patients with chronic mild-to-moderate, isolated MR are usually asymptomatic. This form of LV volume overload is well tolerated. Fatigue, exertional dyspnea, and orthopnea are the most prominent complaints in patients with chronic severe MR. Palpitations are common and may signify the onset of AF. Late-onset right-sided heart failure, with painful hepatic congestion, ankle edema, distended neck veins, ascites, and secondary tricuspid regurgitation (TR), occurs in patients with MR who have associated pulmonary vascular disease and pulmonary hypertension. Acute pulmonary edema is common in patients with acute severe MR.

■ PHYSICAL FINDINGS

In patients with chronic severe MR, the arterial pressure is usually normal, although the carotid arterial pulse may show a sharp, low-volume upstroke owing to the reduced forward CO. A systolic thrill is often palpable at the cardiac apex, the LV is hyperdynamic with a brisk systolic impulse and a palpable rapid-filling wave (S_3), and the apex beat is often displaced laterally.

In patients with acute severe MR, the arterial pressure may be reduced with a narrow pulse pressure, the jugular venous pressure and waveforms may be normal or increased and exaggerated, the apical impulse is not displaced, and signs of pulmonary congestion are prominent.

Auscultation S_1 is generally absent, soft, or buried in the holosystolic murmur of chronic, severe MR. In patients with severe MR, the aortic valve may close prematurely (due to the reduced forward cardiac output), resulting in wide but physiologic splitting of S_2. A low-pitched S_3 occurring 0.12–0.17 s after the aortic valve closure sound, i.e., at the completion of the rapid-filling phase of the LV, is believed to be caused by the sudden tensing of the papillary muscles, chordae tendineae, and valve leaflets. It may be followed by a short, rumbling, mid-diastolic murmur, even in the absence of structural MS. In patients with ischemic or dilated cardiomyopathy, however, a third sound (S_3) may also signify ventricular dysfunction. A fourth heart sound is often audible in patients with *acute* severe MR who are in sinus rhythm. A presystolic murmur is not ordinarily heard with isolated MR.

A systolic murmur of at least grade III/VI intensity is the most characteristic auscultatory finding in chronic severe MR. It is usually holosystolic **(see Fig. 246-5*A*)**, but as previously noted, it is decrescendo and ceases in mid-to-late systole in patients with acute severe MR. The systolic murmur of chronic MR is usually most prominent at the apex and radiates to the axilla. However, in patients with ruptured chordae tendineae or primary involvement of the posterior mitral leaflet with prolapse or flail, the regurgitant jet is eccentric, directed anteriorly, and strikes the LA wall adjacent to the aortic root. In this situation, the systolic murmur is transmitted to the base of the heart and, therefore, may be confused with the murmur of AS. The murmur associated with anterior leaflet prolapse or flail is directed to the axilla. In patients with ruptured chordae tendineae, the systolic murmur may have a cooing or "seagull" quality, whereas a flail leaflet may produce a murmur with a musical quality. The systolic murmur of chronic MR not due to MVP is intensified by isometric exercise (handgrip) but is reduced during the strain phase of the Valsalva maneuver because of the associated decrease in LV preload.

■ LABORATORY EXAMINATION

ECG In patients with sinus rhythm, there is evidence of LA enlargement, but right atrial (RA) enlargement also may be present when pulmonary hypertension is significant and affects right ventricular function and size. Chronic severe MR is frequently associated with AF. In many patients, there is no clear-cut ECG evidence of enlargement of either ventricle. In others, the signs of eccentric LV hypertrophy are present.

Echocardiogram (Fig. 275-1) Transthoracic echocardiography (TTE) is indicated to assess the mechanism of the MR and its hemodynamic severity. LV function can be assessed from LV end-diastolic and end-systolic volumes and EF. Observations can be made regarding leaflet structure and function, chordal integrity, LA and LV size, annular calcification, and regional and global LV systolic function. Doppler imaging should demonstrate the width or area of the color flow MR jet within the LA, the duration and intensity of the continuous wave Doppler signal, the pulmonary venous flow contour, the early peak mitral inflow velocity, and quantitative measures of regurgitant volume, RF, and effective regurgitant orifice area. In addition, the PA pressures (PAPs) can be estimated from the TR jet velocity. TTE is also indicated to follow the course of patients with chronic MR and to provide rapid assessment for any clinical change. Transesophageal echocardiography (TEE) provides greater anatomic detail than TTE **(see Fig. 248-5)**. Exercise testing with TTE can be useful to assess exercise capacity as well as any dynamic change in MR severity, PA systolic pressures, and biventricular function, for patients in whom there is a discrepancy between clinical findings and the results of other noninvasive testing.

Chest X-Ray The LA and LV are the dominant chambers in chronic MR. Late in the course of the disease, the LA may be massively enlarged and forms the right border of the cardiac silhouette. Pulmonary venous congestion, interstitial edema, and Kerley B lines

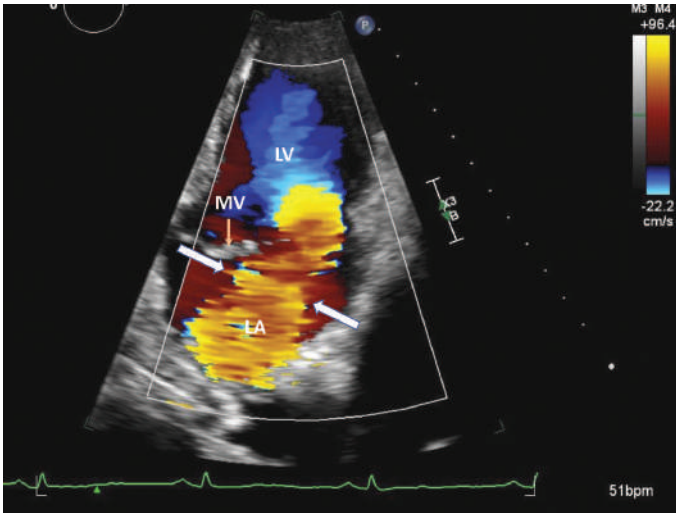
M3 M4
+96.4
LV
MV
LA
-22.2
cm/s
51bpm

FIGURE 275-1 Mitral regurgitation. Apical two-chamber transthoracic color flow Doppler echocardiographic display of a broad systolic jet of severe mitral regurgitation (yellow stream between the two white arrows) from the left ventricle directed posteriorly into the left atrium. LA, left atrium; LV, left ventricle; MV, anterior leaflet of the mitral valve.

are sometimes noted. Marked calcification of the mitral leaflets occurs commonly in patients with long-standing, combined rheumatic MR and MS, as well as in patients with radiation-induced mitral valve disease. Calcification of the mitral annulus may be visualized, particularly on the lateral view of the chest. Patients with acute severe MR may have asymmetric pulmonary edema if the regurgitant jet is directed predominantly to the orifice of an upper lobe pulmonary vein.

TREATMENT (FIGS. 275-2 AND 275-3)

Mitral Regurgitation

MEDICAL TREATMENT

The management of chronic severe MR depends to some degree on its cause. Anticoagulation with either warfarin or a direct oral agent (e.g., apixaban, rivaroxaban) should be provided if AF intervenes, as guided by the CHA_2DS_2-VASc risk score. The direct oral anticoagulants should not be used if moderate or severe rheumatic mitral stenosis is also present **(Ch. 274)**; they are also not approved for use in patients with mechanical prosthetic heart valves. Cardioversion should be considered depending on the clinical context, AF chronicity, and LA size. In contrast to the acute setting, there are no large, long-term prospective studies to substantiate the use of vasodilators for the treatment of chronic, isolated severe primary MR with preserved LV systolic function *in the absence of systemic hypertension.* However, the severity of secondary MR in the setting of an ischemic or dilated cardiomyopathy may diminish with aggressive guideline-directed medical therapy (GDMT) of heart failure including the use of diuretics for decongestion, beta blockers, angiotensin-converting enzyme (ACE) inhibitors/angiotensin receptor blockers, angiotensin-neprilysin inhibitors, mineralocorticoid receptor antagonists, sodium-glucose cotransporter

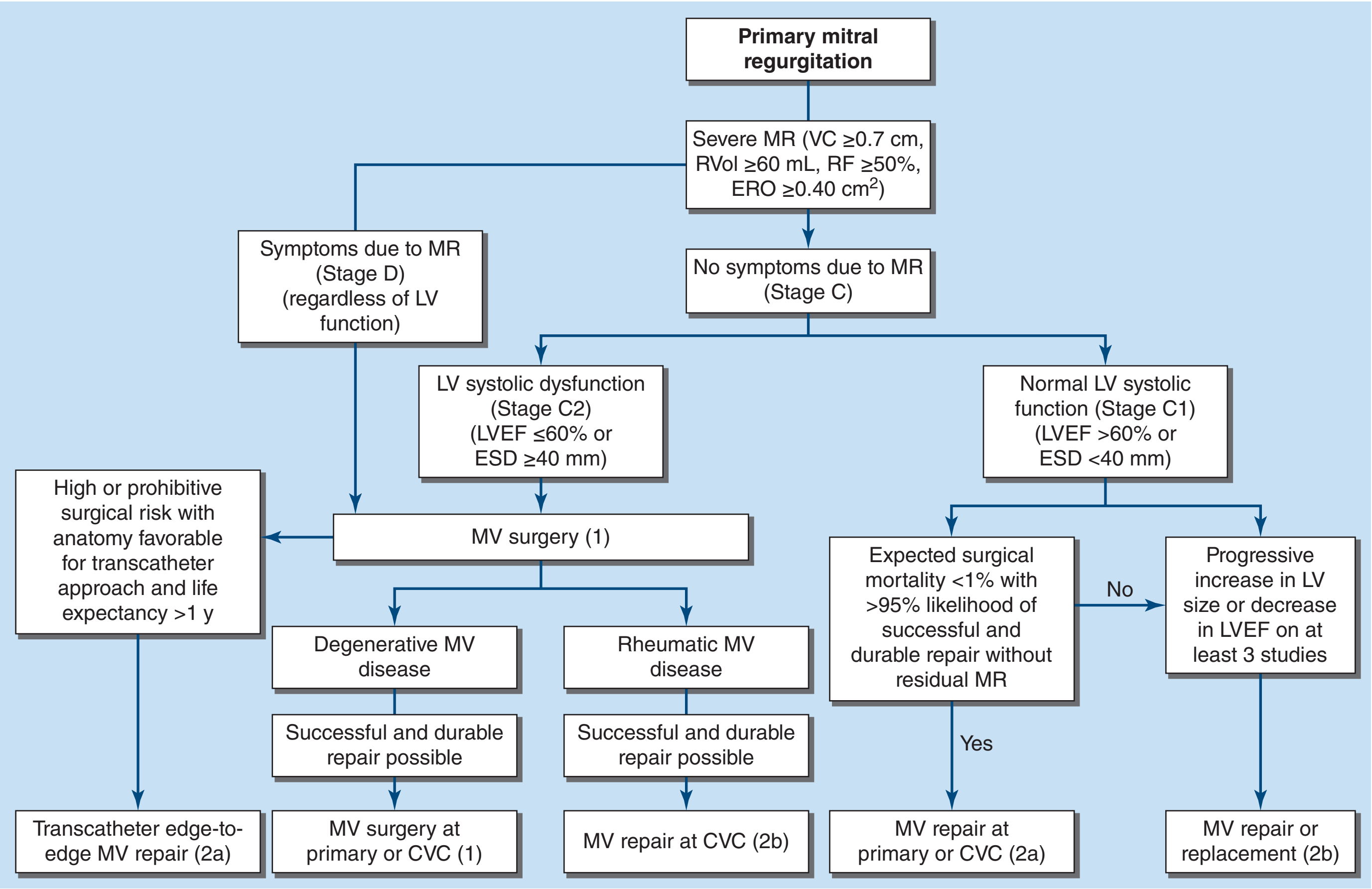
Primary mitral regurgitation
Severe MR (VC ≥0.7 cm, RVol ≥60 mL, RF ≥50%, ERO ≥0.40 cm²)
Symptoms due to MR (Stage D) (regardless of LV function)
No symptoms due to MR (Stage C)
LV systolic dysfunction (Stage C2) (LVEF ≤60% or ESD ≥40 mm)
Normal LV systolic function (Stage C1) (LVEF >60% or ESD <40 mm)
High or prohibitive surgical risk with anatomy favorable for transcatheter approach and life expectancy >1 y
MV surgery (1)
Expected surgical mortality <1% with >95% likelihood of successful and durable repair without residual MR
No
Progressive increase in LV size or decrease in LVEF on at least 3 studies
Degenerative MV disease
Rheumatic MV disease
Successful and durable repair possible
Successful and durable repair possible
Yes
Transcatheter edge-to-edge MV repair (2a)
MV surgery at primary or CVC (1)
MV repair at CVC (2b)
MV repair at primary or CVC (2a)
MV repair or replacement (2b)

FIGURE 275-2 Management of primary mitral regurgitation (MR). See legend for **Fig. 272-4** for explanation of treatment recommendations (Class I, IIa, IIb) and disease stages (B, C1, C2, D). Preoperative coronary angiography should be performed routinely as determined by age, symptoms, and coronary risk factors. Cardiac catheterization and angiography may also be helpful when there is a discrepancy between clinical and noninvasive findings. Mitral valve repair is strongly preferred over valve replacement whenever feasible for surgical treatment of primary MR. Transcatheter edge-to-edge repair (TEER) is reserved for high or prohibitive surgical risk patients with appropriate anatomy on transesophageal imaging. CVC, comprehensive valve center; EF, ejection fraction; ERO, effective regurgitant orifice; ESD, end-systolic dimension; LV, left ventricular; MV, mitral valve; RF, regurgitant fraction; RVol, regurgitant volume; VC, vena contracta. *(Reproduced with permission from CM Otto et al: ACC/AHA guideline for the management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2021; 143(5):e72.)*

2 inhibitors, and biventricular pacing (cardiac resynchronization therapy [CRT]) when indicated. Antibiotic prophylaxis for prevention of IE is indicated for MR patients with a prior history of IE. Asymptomatic patients with severe MR in sinus rhythm with normal LV size and systolic function should avoid isometric forms of exercise.

Patients with acute severe MR require urgent stabilization and preparation for surgery. Diuretics, intravenous vasodilators (particularly sodium nitroprusside), and even mechanical circulatory support may be needed for patients with post-MI papillary muscle rupture or other forms of acute severe MR.

SURGICAL TREATMENT

In the selection of patients with chronic, severe, primary MR for surgical treatment, the often slowly progressive nature of the condition must be balanced against the immediate and long-term risks associated with operation. These risks are significantly lower for primary valve repair than for valve replacement **(Table 275-2)**. In a 2023 analysis of outcomes reported to the Society of Thoracic Surgeons Adult Cardiac Surgical Database, investigators reported a mean perioperative mortality risk of 1.16% (<0.5% for patients <65 years of age) for repair of primary MR. Repair usually consists of valve reconstruction using a variety of valvuloplasty techniques and insertion of an annuloplasty ring. Repair spares the patient the long-term adverse consequences of valve replacement, including thromboembolic and hemorrhagic complications in the case of mechanical prostheses and late valve failure necessitating repeat valve replacement in the case of bioprostheses. In addition, by preserving the integrity of the papillary muscles, subvalvular apparatus, and chordae tendineae, mitral repair and valvuloplasty maintain LV function to a relatively greater degree than does valve replacement.

TABLE 275-2 Mortality Rates After Mitral Valve Surgery[a]

OPERATION	NUMBER	UNADJUSTED OPERATIVE MORTALITY (%)
MVR (isolated)	10,699	4.5
MVR + CAB	3509	9.6
MVRp	12,424	1.2
MVRp + CAB	4093	5.4

[a] Data are for calendar year 2018 during which 1088 participant groups reported a total of 287,872 procedures. Surgical mitral valve commissurotomy cases are included in the mitral valve repair procedures.

Abbreviations: CAB, coronary artery bypass; MVR, mitral valve replacement; MVRp, mitral valve repair.

Source: Adapted from ME Bowdish et al: Ann Thorac Surg 109:1646, 2020.

Surgery for chronic severe primary MR is indicated once symptoms occur, especially if valve repair is feasible (Fig. 275-2). Surgery should also be recommended for asymptomatic patients with LV dysfunction characterized by an EF ≤60% or an LV end-systolic dimension (LV ESD) ≥40 mm. Other indications for early consideration of mitral valve repair in asymptomatic patients include a progressive decrease in LVEF or increase in LV ESD on serial imaging as well as mitral valve anatomy that would predict a >95% of a successful and durable repair in a low surgical risk patient.

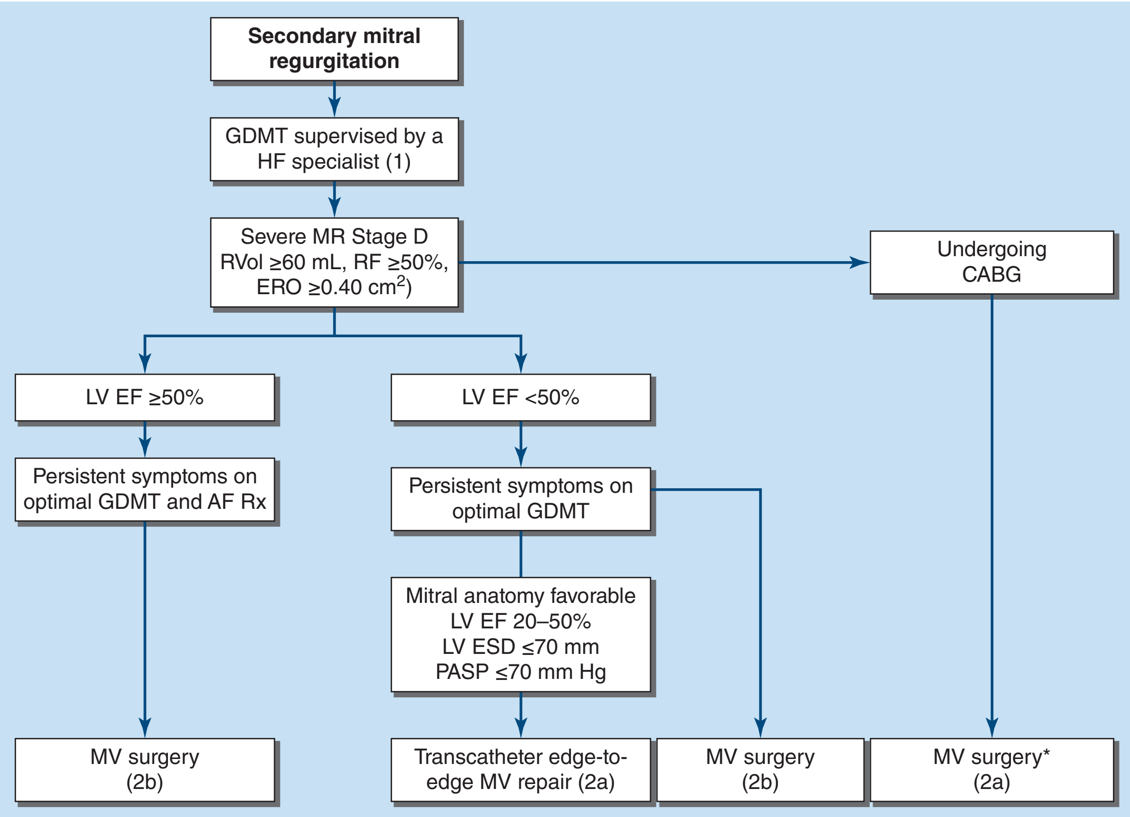
Secondary mitral regurgitation
GDMT supervised by a HF specialist (1)
Severe MR Stage D RVol ≥60 mL, RF ≥50%, ERO ≥0.40 cm2)
Undergoing CABG
LV EF ≥50%
LV EF <50%
Persistent symptoms on optimal GDMT and AF Rx
Persistent symptoms on optimal GDMT
Mitral anatomy favorable LV EF 20–50% LV ESD ≤70 mm PASP ≤70 mm Hg
MV surgery (2b)
Transcatheter edge-to-edge MV repair (2a)
MV surgery (2b)
MV surgery* (2a)

FIGURE 275-3 Management of secondary mitral regurgitation. See legend for **Fig. 272-4** for explanation of treatment recommendations (Class I, IIa, IIb) and disease stages (B, C1, C2, D). Preoperative coronary angiography should be performed routinely as determined by age, symptoms, and coronary risk factors. Cardiac catheterization and angiography may also be helpful when there is a discrepancy between clinical and noninvasive findings. Surgery is recommended for patients with left ventricular ejection fraction (LVEF) >50%. Transcatheter edge-to-edge repair (TEER) is reasonable in selected patients after guideline-directed management and therapy (GDMT) has been optimized. *MV replacement may be preferred over MV repair for ischemic MR; AF, atrial fibrillation; CABG, coronary artery bypass grafting; EF, ejection fraction; ERO, effective regurgitant orifice; ESD, end-systolic dimension; HF, heart failure; LV, left ventricular; MR, mitral regurgitation, MV, mitral valve; PASP, pulmonary artery systolic pressure; RF, regurgitant fraction; RVol, regurgitant volume; Rx, treatment. *(Reproduced with permission from CM Otto et al: ACC/AHA guideline for the management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2021; 143(5):e72.)*

These aggressive recommendations for surgery are predicated on the adverse long-term consequences of waiting for LV function to decline further as well as the outstanding results achievable with mitral valve repair by reference surgeons at high-volume centers. Indeed, the majority of patients <75 years of age with degenerative MR, normal LV systolic function, and no coronary artery disease (CAD) can now undergo successful and durable repair of degenerative MR (e.g., prolapse, flail) by an experienced surgeon at very low risk for perioperative death or major complication. Repair is feasible in up to 95% of patients with degenerative disease operated on by a high-volume surgeon in a referral center of excellence. Repair techniques include chordal transfer, creation of neochords, limited leaflet resection, and insertion of an annuloplasty band. Long-term durability is excellent; the incidence of reoperative surgery for failed primary repair is ~1% per year for the first 10 years after surgery. For patients with paroxysmal or persistent AF, left or biatrial maze surgery, or radiofrequency or cryoablative isolation of the pulmonary veins, along with left atrial appendage amputation, is performed to reduce the risk of recurrent postoperative AF and associated thrombus formation.

The surgical management of patients with secondary MR is more complicated. Surgery for patients with ischemic MR most often involves simultaneous coronary artery revascularization. Current surgical practice includes either annuloplasty repair with an undersized, rigid ring or chord-sparing valve replacement for patients with moderate or greater degrees of MR. Valve repair for ischemic MR is associated with lower perioperative mortality rates than valve replacement but significantly higher rates of recurrent MR over time. Thus, replacement may be preferred over repair in this context. In patients with ischemic MR and significantly impaired LV systolic function (EF <30%), the risk of surgery is higher, recovery of LV performance is incomplete, and long-term survival is reduced. Referral for surgery must be individualized and made only after aggressive attempts to improve symptoms with GDMT and CRT, when indicated. The routine performance of surgical valve repair in patients with significant secondary MR due to a dilated cardiomyopathy has not been shown to improve long-term survival compared with optimal GDMT, especially in the era of quadruple medical therapy for heart failure with reduced EF. Patients with acute severe MR can often be stabilized temporarily with appropriate medical therapy, but surgical correction will be necessary emergently in the case of papillary muscle rupture and within days in most other settings.

When surgical treatment is contemplated, left and right heart catheterization and left ventriculography *may* be helpful in confirming the presence of severe MR in patients in whom there is a discrepancy between the clinical and TTE findings that cannot be resolved with TEE or CMR. Coronary angiography identifies patients who require concomitant coronary revascularization.

TRANSCATHETER MITRAL VALVE REPAIR AND REPLACEMENT

A transcatheter approach to the treatment of either primary or secondary MR may be feasible in selected patients with appropriate mitral valve anatomy. One approach involves the deployment of a clip delivered via transseptal puncture that grasps the leading edges of the mitral leaflets in their mid-portion (anterior scallop to posterior scallop). The length and width of the gap between the leading edges of the leaflets, as well as other considerations such as leaflet thickening and calcification, dictate patient eligibility. There are two commercially available transcatheter systems employing clip technology for the treatment of degenerative MR in symptomatic patients considered to be at high or prohibitive surgical risk **(Fig. 275-4)**. Randomized trials of transcatheter

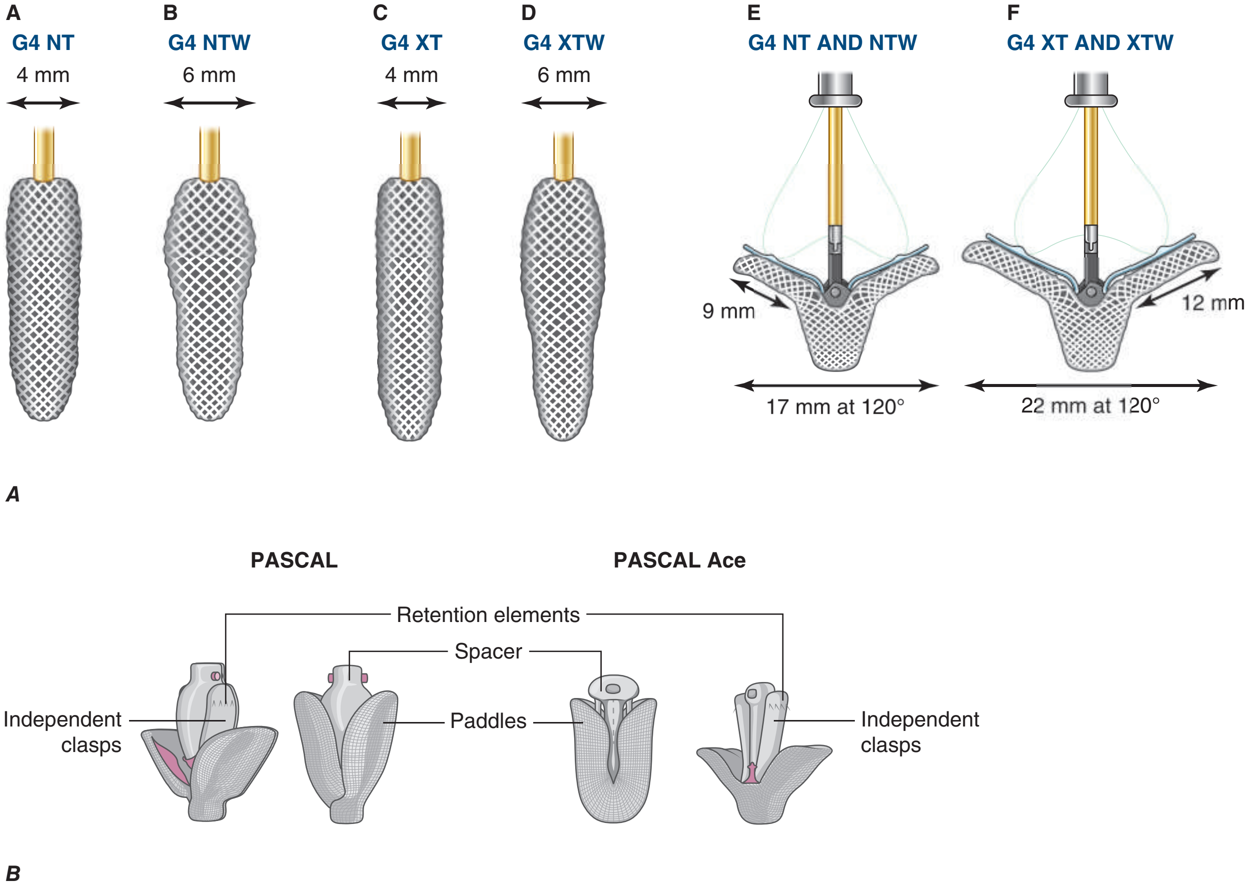

FIGURE 275-4 Clip devices used to grasp the free edges of the anterior and posterior leaflets in their midsections during transcatheter mitral valve repair of selected patients with mitral regurgitation. ***A.*** Fourth-generation of the first approved edge-to-edge repair device. ***B.*** System approved in 2022 includes a central spacer device and two paddles. The spacer is designed to reduce stress on the leaflets and preserve mitral valve area. *(MitraClip is a trademark of Abbott or its related companies. PASCAL is an Edwards Lifesciences tradename.)*

edge-to-edge repair (TEER) versus surgical repair are ongoing to determine if the extension of clip technology to low- and intermediate-risk surgical patients with significant degenerative MR is appropriate. The TEER system is also approved for the treatment of heart failure patients with secondary MR. The results of transthoracic and transesophageal echocardiographic imaging are critical to patient selection, along with an assessment of the adequacy of GDMT for heart failure. The use of TEER with a clip device in addition to medical therapy was shown to be superior to medical therapy alone in a trial involving symptomatic heart failure patients with reduced EF and at least moderately severe secondary MR followed through 5 years. Patients treated with the clip device had significantly lower rates of heart failure hospitalizations and all-cause mortality than those treated medically. This was the first randomized trial to show a survival benefit in patients with heart failure and secondary MR. There has been a rapid expansion over the past 5 years in both the number of sites offering mitral valve TEER and the number of cases reported to the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry.

Other transcatheter approaches to mitral valve repair have included the deployment of a device within the coronary sinus that can be adjusted to reduce mitral annular circumference and the effective orifice area of the valve much like a surgically implanted ring. Variations in the anatomic relationship of the coronary sinus to the mitral annulus and circumflex coronary artery have limited the applicability of this technique. Attempts to reduce the septal-lateral dimension of a dilated annulus using adjustable cords placed across the LV in a subvalvular location have been investigated. Construction of neochords to the mitral leaflets under TEE guidance using a system delivered via the cardiac apex has also been studied. Investigational experience with transcatheter mitral valve replacement systems remains in early clinical stages, although the field is evolving rapidly. Many high surgical risk patients are not candidates for transcatheter mitral valve repair, and thus, there is keen interest in refining this technology. Challenges with transseptal delivery and LV outflow tract obstruction from the devices used have prompted iterative changes in the systems utilized in early feasibility studies.

■ FURTHER READING

Badhwar V et al: Risk of mitral valve repair for primary mitral regurgitation. J Am Coll Cardiol 81:636, 2023.

Bonow RO et al: 2020 focused update of the 2017 expert consensus decision pathway on the management of mitral regurgitation. J Am Coll Cardiol 75:2236, 2020.

Lim DS et al: Randomized comparison of transcatheter edge-to-edge repair for degenerative mitral regurgitation in prohibitive surgical risk patients. J Am Coll Cardiol Interv 15:2523, 2022.

Otto CM et al: 2020 ACC/AHA guideline for the management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 143:e72, 2021.

Stone GW et al: Five-year follow-up after transcatheter repair of secondary mitral regurgitation. N Engl J Med 388:2037, 2023.

von Bardeleben RS et al: 1-year outcomes with fourth generation mitral valve transcatheter edge-to-edge repair from the EXPAND G4 study. J Am Coll Cardiol Interv 16:2600, 2023.

Zahr F et al: 1-year outcomes following transfemoral transseptal transcatheter mitral valve replacement: Intrepid TMVR early feasibility study results. J Am Coll Cardiol Interv 16:2868, 2023.

276

Mitral Valve Prolapse

Patrick T. O'Gara, Joseph Loscalzo

The role of the physical examination in the evaluation of patients with valvular heart disease is also considered in **Chaps. 44 and 246**; of electrocardiography (ECG) in **Chap. 247**; of echocardiography and other noninvasive imaging techniques in **Chap. 241**; and of cardiac catheterization and angiography in **Chap. 249**.

MITRAL VALVE PROLAPSE

Mitral valve prolapse (MVP), also variously termed the *systolic click-murmur syndrome, Barlow's syndrome* **(Fig. 276-1)**, *floppy-valve syndrome*, and *billowing mitral leaflet syndrome*, is a relatively common but highly variable clinical syndrome resulting from diverse pathologic mechanisms affecting the mitral valve apparatus. Among these are excessive or redundant mitral leaflet tissue, which is commonly associated with myxomatous degeneration and greatly increased concentrations of certain glycosaminoglycans. MVP is the most common abnormality leading to primary mitral regurgitation (MR) and mitral valve repair surgery **(see Chap. 275)**.

In most patients with MVP, the cause is unknown, but in some, it appears to be genetically determined. A reduction in the production of type III collagen has been implicated, and electron microscopy has revealed fragmentation of collagen fibrils. A meta-analysis of six

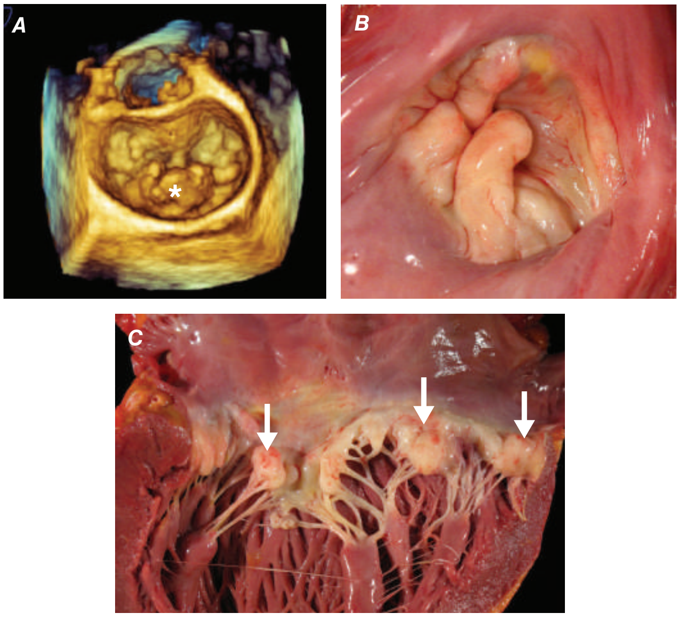

FIGURE 276-1 Mitral valve prolapse. Myxomatous thickening and prolapse of the mitral valve can occur in isolation in 2–3% of the general population or may be associated with heritable connective tissue disorders, such as Marfan syndrome. Myxomatous degeneration of the valve predisposes to severe regurgitation and chordal rupture and is a frequent indication for mitral valve repair or replacement. Prolapse can affect one or both leaflets, to varying degrees. ***A.*** Three-dimensional transesophageal echocardiogram showing a myxomatous mitral valve from the left atrial *en face* aspect. There is billowing and prolapse of the entire middle scallop of the posterior leaflet (*asterisk*). *(Figure courtesy of Douglas C. Shook, MD, Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital.)* ***B.*** The posterior leaflet of the mitral valve demonstrates marked prolapse and hooding in all segments and severe redundancy in this postmortem photograph taken from the vantage point of the left atrium. ***C.*** Opening the left heart reveals prominent mitral leaflet hooding (*arrows*). The chordae are focally thickened but are not fused as would be the case in rheumatic valve disease. *(Used with permission from JC Wu, RF Padera: Clinicopathologic correlates, in Atlas of Echocardiography, 2nd ed, SD Solomon [ed], E Braunwald [series ed]. Philadelphia, Current Medicine Group LLC, 2008. p 363.)*

genome-wide association studies and 4884 MVP cases identified 14 genetic loci associated with MVP. Candidate genes included *LMCD1, SPTBN1, LTBP2, TGFB2, NMB,* and *ALPK3*. A polygenic risk score improved the performance of a clinical risk prediction model for the development of MVP, although overall performance was modest. Other work has identified genetic associations with *LMNA, FLNC,* and *FLNA*, which are often expressed as cardiomyopathies.

MVP is a frequent finding in patients with heritable disorders of connective tissue, including Marfan syndrome **(Chap. 425)**, osteogenesis imperfecta, and Ehlers-Danlos syndrome. MVP may be associated with thoracic skeletal deformities similar to but not as severe as those in Marfan syndrome, such as a high-arched palate and alterations of the chest and thoracic spine, including kyphosis and the so-called straight back syndrome. Other associated features can include a history of inguinal hernias, joint dislocations, meniscal tears, and easy bruisability.

In most patients with MVP, myxomatous degeneration is confined to the mitral valve, although the tricuspid and aortic valves may also be affected. Prolapse can affect one or both leaflets. The posterior mitral leaflet is usually more affected than the anterior, and the mitral valve annulus is often dilated. In many patients, elongated, redundant, or ruptured chordae tendineae cause or contribute to the regurgitation.

MVP also may occur rarely as a sequel to acute rheumatic fever, in ischemic heart disease, and in various cardiomyopathies (see above), as well as in 20% of patients with ostium secundum atrial septal defect.

MVP may lead to excessive stress on the papillary muscles, leading to localized ischemia, infarction, and replacement fibrosis. The latter, which may be a nidus for ventricular arrhythmias, may be visible on cardiac magnetic resonance imaging as late gadolinium enhancement and occurs in the absence of coronary artery disease. Rupture of chordae tendineae and progressive annular dilation and calcification contribute to valvular regurgitation, which then places more stress on the diseased mitral valve apparatus, thereby creating a vicious cycle.

CLINICAL FEATURES

MVP is more common in women than men and occurs most frequently between the ages of 15 and 30 years; the clinical course is most often benign. MVP may also be observed in older (>50 years) patients, often men, in whom MR is often more severe because of chordal rupture and requires surgical treatment. There is an increased familial incidence for some patients, suggesting an autosomal dominant form of inheritance with incomplete penetrance. MVP varies in its clinical expression, ranging from only a systolic click and murmur with mild prolapse of the posterior leaflet to severe MR due to chordal rupture and leaflet flail. The degree of myxomatous change of the leaflets can also vary widely. In many patients, the condition progresses over years or decades; in others, it worsens rapidly as a result of chordal rupture or endocarditis.

Most patients are asymptomatic and remain so for their entire lives. However, in North America, MVP is now the most common cause of isolated severe MR requiring surgical treatment. Arrhythmias, most commonly ventricular premature contractions and paroxysmal supraventricular and ventricular tachycardia, as well as atrial fibrillation (AF), have been reported and may cause palpitations, light-headedness, and syncope. Sudden death is a very rare complication and occurs most often in patients with severe MR and depressed left ventricle (LV) systolic function, although it can occur in individuals with normal LV size and function. A small subset of MVP patients with high-grade ventricular ectopy has been identified with phenotypic features including electrocardiographic inferior-apical T-wave abnormalities, high-density premature ventricular complexes at rest, mitral annular disjunction (defined as abnormal atrial displacement of the mitral valve leaflet hinge point), and papillary muscle fibrosis on cardiac magnetic resonance imaging (see above). In addition, there may be an excess risk of sudden death among patients with a flail leaflet. Most of these patients have severe MR. Many patients with MVP have chest pain that can be difficult to evaluate; it is often substernal, prolonged, and not related to exertion, but may rarely resemble angina pectoris. Transient cerebral ischemic attacks secondary to emboli from the mitral valve due to endothelial disruption have been reported. Infective endocarditis may occur in patients with MR and/or leaflet thickening.

Auscultation A frequent finding is the mid- or late (nonejection) systolic click, which occurs 0.14 s or more after S_1 and is thought to be generated by the sudden tensing of slack, elongated chordae tendineae or by the prolapsing mitral leaflet when it reaches its maximal excursion. Systolic clicks may be multiple and may be followed by a high-pitched, mid-late systolic crescendo–decrescendo murmur, which occasionally is "whooping" or "honking" and is heard best at the apex. Radiation of the murmur will depend on the involved leaflet. With posterior leaflet prolapse, the jet of MR is directed anteriorly and the murmur will radiate to the base of the heart. With anterior leaflet involvement, the jet of MR is directed posteriorly and the murmur will radiate to the axilla and back. The click and murmur occur earlier with standing, during the strain phase of the Valsalva maneuver and with any intervention that decreases LV volume (preload), exaggerating the propensity of the leaflet to prolapse. Conversely, squatting and isometric exercises, which increase LV volume, diminish MVP; the click-murmur complex is delayed, moves away from S_1, and may even disappear. Some patients have a mid-systolic click without a murmur; others have a murmur without a click. Still others have both sounds at different times.

LABORATORY EXAMINATION

The ECG most commonly is normal but may show biphasic or inverted T waves in leads II, III, and aVF and, occasionally, supraventricular or ventricular premature beats. Transthoracic echocardiography (TTE) is particularly effective in identifying the abnormal position and prolapse of the mitral valve leaflets. A useful echocardiographic definition of MVP is systolic displacement (in the parasternal long axis view) of the belly of the mitral valve leaflets by at least 2 mm into the left atrium (LA) superior to the plane of the mitral annulus. There can be prolapse of one or both leaflets **(Fig. 276-2)**. Color flow and continuous wave Doppler imaging is helpful to evaluate the associated MR and provide estimates of severity. The jet lesion of MR due to MVP is most often eccentric, and assessment of the effective regurgitant orifice area and regurgitant volume can be difficult with standard techniques. Both three-dimensional echocardiography and cardiac magnetic resonance imaging can provide more precise determinations of LV volumes. Transesophageal echocardiography (TEE) is indicated when more accurate anatomic information is required and is performed routinely for intraoperative guidance during surgical or transcatheter valve repair. Exercise testing can be performed when there is uncertainty regarding functional capacity. It is often combined with rest and immediate poststress TTE to assess LV and right ventricular (RV) function and the dynamic nature of MR and pulmonary artery pressures. Left ventriculography done at the time

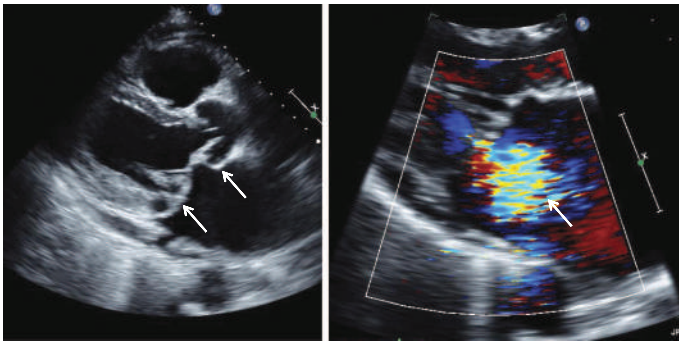

FIGURE 276-2 Barlow's valve with classic mitral valve prolapse, as seen on transthoracic echocardiogram in parasternal long-axis windows. Left: parasternal long-axis window, showing both myxomatous leaflets (*arrows*) billowing into the left atrium in late systole. Right: same window with color Doppler showing significant mitral regurgitation (*arrow*) in systole. *(Courtesy of Justina Wu, MD, PhD.)*

of right and left heart catheterization is rarely necessary but can also show prolapse of the posterior and sometimes of both mitral valve leaflets.

TREATMENT

Mitral Valve Prolapse

Infective endocarditis prophylaxis is indicated for patients with a prior history of endocarditis. Beta blockers sometimes relieve chest pain and control palpitations. Decisions regarding anticoagulation for stroke prevention in AF should be based on the CHA_2DS_2-VASc score and an assessment of bleeding risk. If the patient is symptomatic from severe MR, mitral valve repair is indicated **(see Fig. 275-2)**. Other indications for surgery for MVP with severe primary MR include findings of established or progressive LV systolic dysfunction. Surgery can also be considered for low-risk asymptomatic patients in whom a successful and durable repair can be achieved with at least 95% likelihood by an expert surgeon. Mitral valve repair is preferred over replacement in patients with MVP or flail mitral leaflet **(see Table 275-2)**; technical success is dependent not only on the anatomic findings but also on the skill and experience of the surgeon. Repair of isolated posterior leaflet prolapse is usually straightforward, but increasingly more complex pathologies (e.g., anterior leaflet prolapse, bileaflet prolapse, Barlow's deformity) require advanced skills. Careful pre- and intraoperative TEE imaging is an important component of patient evaluation and surgical planning. Transcatheter edge-to-edge repair (TEER) using one of two commercially available systems to grasp the anterior and posterior leaflets together can be considered for treatment of symptomatic patients at prohibitive or high surgical risk with severe primary MR due to MVP **(see Chap. 275 and Fig. 275-4)**. Most often, the MR will be reduced in severity but not eliminated. Nevertheless, symptom status and indices of LV size and function can be improved with this approach, which is now offered at >540 specialized sites in the United States. Reported hospital mortality rates following the procedure are <2%. Transcatheter replacement devices remain under active investigation but are not yet approved for use in the United States **(see Chap. 275)**.

FURTHER READING

Essayagh B et al: Arrhythmic mitral valve prolapse and mitral annular disjunction: Pathophysiology, risk stratification, and management. Eur Heart J 44:3121, 2023.

Otto CM et al: 2020 ACC/AHA guideline for the management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 143:e72, 2021.

Roselli C et al: Genome-wide association study identifies novel genetic loci: A new polygenic risk score for mitral valve prolapse. Eur Heart J 43:1668, 2022.

277

Tricuspid Valve Disease

Patrick T. O'Gara, Joseph Loscalzo

TRICUSPID STENOSIS

Tricuspid stenosis (TS), which is much less prevalent than mitral stenosis (MS) in North America and Western Europe, is generally rheumatic in origin and is more common in women than men **(Table 277-1)**. It does not occur as an isolated lesion and is usually associated with MS.

TABLE 277-1 Causes of Tricuspid Valve Diseases

VALVE LESION	ETIOLOGIES
Tricuspid stenosis	Rheumatic Congenital
Tricuspid regurgitation	**Primary (organic, leaflet, or chordal related)** Rheumatic Endocarditis Myxomatous (TVP) Carcinoid Radiation Congenital (Ebstein's) Trauma (including that due to RV endomyocardial biopsy) **Secondary (functional, atrial, and/or ventricular)** RV and/or tricuspid annular dilation due to multiple causes (e.g., long-standing pulmonary HTN, remodeling post-RV MI, left-sided heart disease, cardiomyopathy, AF (atrial functional tricuspid regurgitation) **CIED related** Leaflet impingement, adherence, laceration, perforation, avulsion; chordal entrapment

Abbreviations: AF, atrial fibrillation; CIED, cardiac implanted electronic device; HTN, hypertension; MI, myocardial infarction; RV, right ventricular; TVP, tricuspid valve prolapse.

Hemodynamically significant TS occurs in 5–10% of patients with severe MS; rheumatic TS is commonly associated with some degree of tricuspid regurgitation (TR). Nonrheumatic causes of TS are rare.

PATHOPHYSIOLOGY

A diastolic pressure gradient between the right atrium (RA) and right ventricle (RV) defines TS. It is augmented when the transvalvular blood flow increases during inspiration and declines during expiration. A mean diastolic pressure gradient of 4 mmHg is usually sufficient to elevate the mean RA pressure to levels that result in systemic venous congestion. Unless sodium intake has been restricted and diuretics administered, this venous congestion is associated with hepatomegaly, ascites, and edema, sometimes severe. In patients with sinus rhythm, the RA *a* wave may be extremely tall and may even approach the level of the RV systolic pressure. The *y* descent is prolonged. The cardiac output (CO) at rest is usually depressed, and it fails to rise during exercise. The low CO is responsible for the normal or only slightly elevated left atrial (LA), pulmonary artery (PA), and RV systolic pressures despite the presence of MS. Thus, the presence of TS can mask the hemodynamic and clinical features of any associated MS.

SYMPTOMS

Because the development of MS generally precedes that of TS, many patients initially have symptoms of pulmonary congestion and fatigue. Characteristically, patients with severe TS complain of relatively little dyspnea for the degree of hepatomegaly, ascites, and edema that they have. However, fatigue secondary to a low CO and discomfort due to signs of right-sided congestion such as hepatomegaly and ascites are common in patients with advanced TS and/or TR. In some patients, TS may be suspected for the first time when symptoms of right-sided failure persist after an adequate mitral commissurotomy.

PHYSICAL FINDINGS

Because TS usually occurs in the presence of other obvious valvular disease, the diagnosis may be missed unless it is considered. Severe TS is associated with marked hepatic congestion, often resulting in cirrhosis, jaundice, serious malnutrition, anasarca, and ascites. Congestive hepatomegaly and, in cases of severe tricuspid valve disease, splenomegaly are present. The jugular veins are distended, and in patients with sinus rhythm, there may be giant *a* waves. The *v* waves are less conspicuous, and because tricuspid obstruction impedes RA emptying during diastole, there is a slow *y* descent. In patients with sinus rhythm, there may be prominent presystolic pulsations of the enlarged liver as well.

On auscultation, an opening snap (OS) of the tricuspid valve may rarely be heard ~0.06 s after pulmonic valve closure. The diastolic murmur of TS has many of the qualities of the diastolic murmur of MS, and because TS almost always occurs in the presence of MS, it may be missed. However, the tricuspid murmur is generally heard best along the left lower sternal border and over the xiphoid process and is most prominent during presystole in patients with sinus rhythm. The murmur of TS is augmented during inspiration, and it is reduced during expiration and particularly during the strain phase of the Valsalva maneuver, when tricuspid transvalvular flow is reduced.

■ LABORATORY EXAMINATION

The electrocardiogram (ECG) features of RA enlargement **(see Fig. 247-8)** include tall, peaked P waves in lead II, as well as prominent, upright P waves in lead V_1. The *absence* of ECG evidence of RV hypertrophy (RVH) in a patient with right-sided heart failure who is believed to have MS should suggest associated tricuspid valve disease. The chest x-ray in patients with combined TS and MS shows particular prominence of the RA and superior vena cava without much enlargement of the PA and with less evidence of pulmonary vascular congestion than occurs in patients with isolated MS; engorgement of the azygos vein can often be appreciated. On transthoracic echocardiographic (TTE) examination, the tricuspid valve is usually thickened and domes in diastole; the transvalvular gradient can be estimated by continuous wave Doppler echocardiography. Severe TS is characterized by a valve area ≤1 cm^2 or pressure half-time of ≥190 ms. The RA and inferior vena cava (IVC) are enlarged. TTE provides additional information regarding the severity of any associated TR, mitral valve structure and function, left ventricular (LV) and RV size and function, and PA pressure. Cardiac catheterization is not routinely necessary for assessment of TS.

TREATMENT

Tricuspid Stenosis

Patients with TS generally exhibit marked systemic venous congestion; salt restriction, bed rest, and diuretic therapy are required during the preoperative period. Such a preparatory period may diminish hepatic congestion and thereby improve hepatic function sufficiently so that the risks of operation, particularly bleeding, are diminished. Surgical relief of the TS should be carried out, preferably at the time of surgical mitral commissurotomy or mitral valve replacement (MVR) for mitral valve disease, in patients with moderate or severe TS who have mean diastolic pressure gradients exceeding ~4 mmHg and tricuspid orifice areas <1.5–2 cm^2. TS is almost always accompanied by significant TR. Operative repair may permit substantial improvement of tricuspid valve function. If repair cannot be accomplished, the tricuspid valve may have to be replaced. Meta-analysis has shown no difference in overall survival between mechanical and tissue valve replacement. Mechanical valves in the tricuspid position are more prone to thromboembolic complications than in other positions. Percutaneous tricuspid balloon commissurotomy for isolated severe TS without significant TR is very rarely performed.

TRICUSPID REGURGITATION

More than 85% of TR cases encountered in clinical practice are secondary (functional) in nature and related to tricuspid annular dilation and leaflet tethering in the setting of RV remodeling caused by pressure or volume overload (or both), myocardial infarction (MI), or trauma (Table 277-1). Secondary TR is commonly seen in the late stages of heart failure due to rheumatic or congenital heart disease with severe PA hypertension (PA systolic pressure >55 mmHg), as well as in other types of left-sided valvular (e.g., mitral regurgitation) or myocardial diseases (e.g., ischemic and idiopathic dilated cardiomyopathies). TR can often emerge in the setting of new-onset atrial fibrillation (AF), particularly in older patients (atrial functional TR). Rheumatic fever may produce primary TR, often associated with TS. Tricuspid valve prolapse, carcinoid heart disease, endomyocardial fibrosis, radiation, infective endocarditis, and leaflet trauma can also produce primary TR. Less commonly, primary TR results from congenitally deformed tricuspid valves and can occur with defects of the atrioventricular canal, as well as with Ebstein's malformation of the tricuspid valve **(Chap. 280)**. A third category of TR is that associated with cardiac implantable electronic device (CIED) leads from pacemakers or defibrillators. When placed across the tricuspid valve, the leads can result

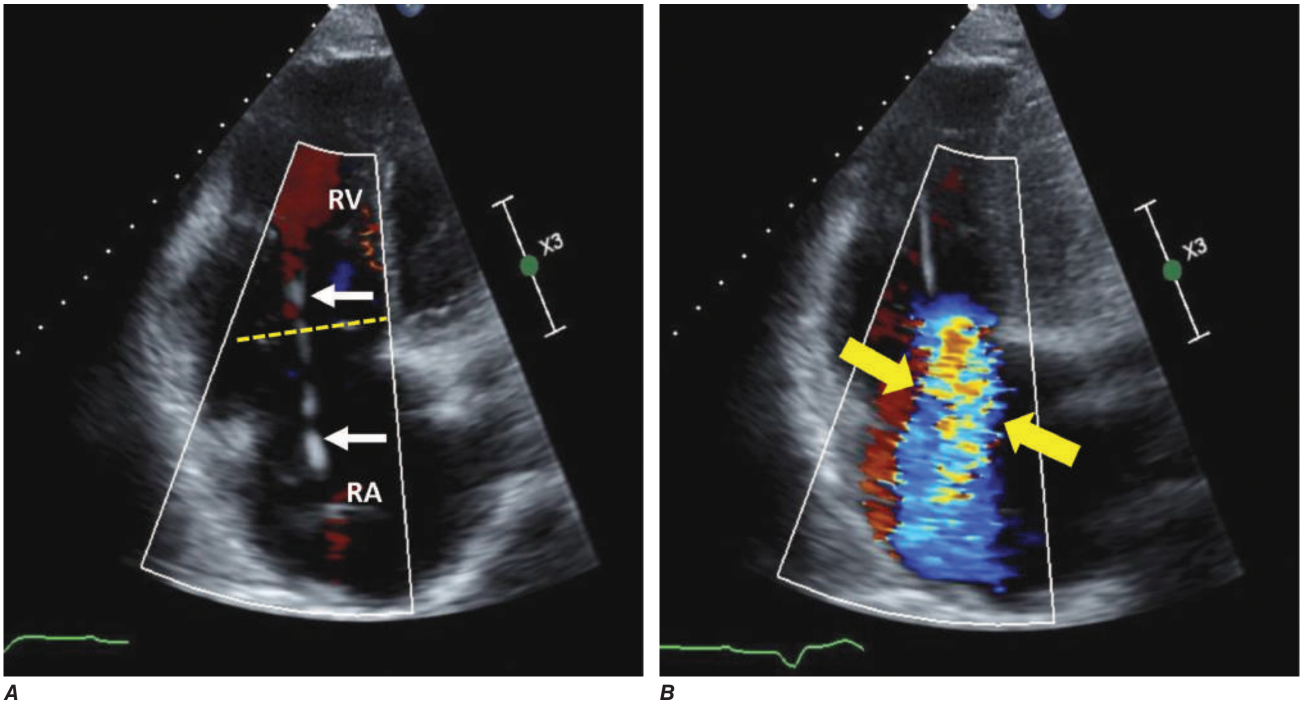

FIGURE 277-1 Tricuspid regurgitation. ***A.*** Transthoracic apical two-chamber view showing right atrium on bottom and right ventricle on top. There is a pacemaker lead (*white arrows*) traversing the tricuspid valve. The plane of tricuspid valve closure is shown by the *dotted yellow line*. ***B.*** Corresponding color flow Doppler image in systole shows a broad jet of severe tricuspid regurgitation (*between the yellow arrows*) refluxing into the right atrium. RA, right atrium; RV right ventricle.

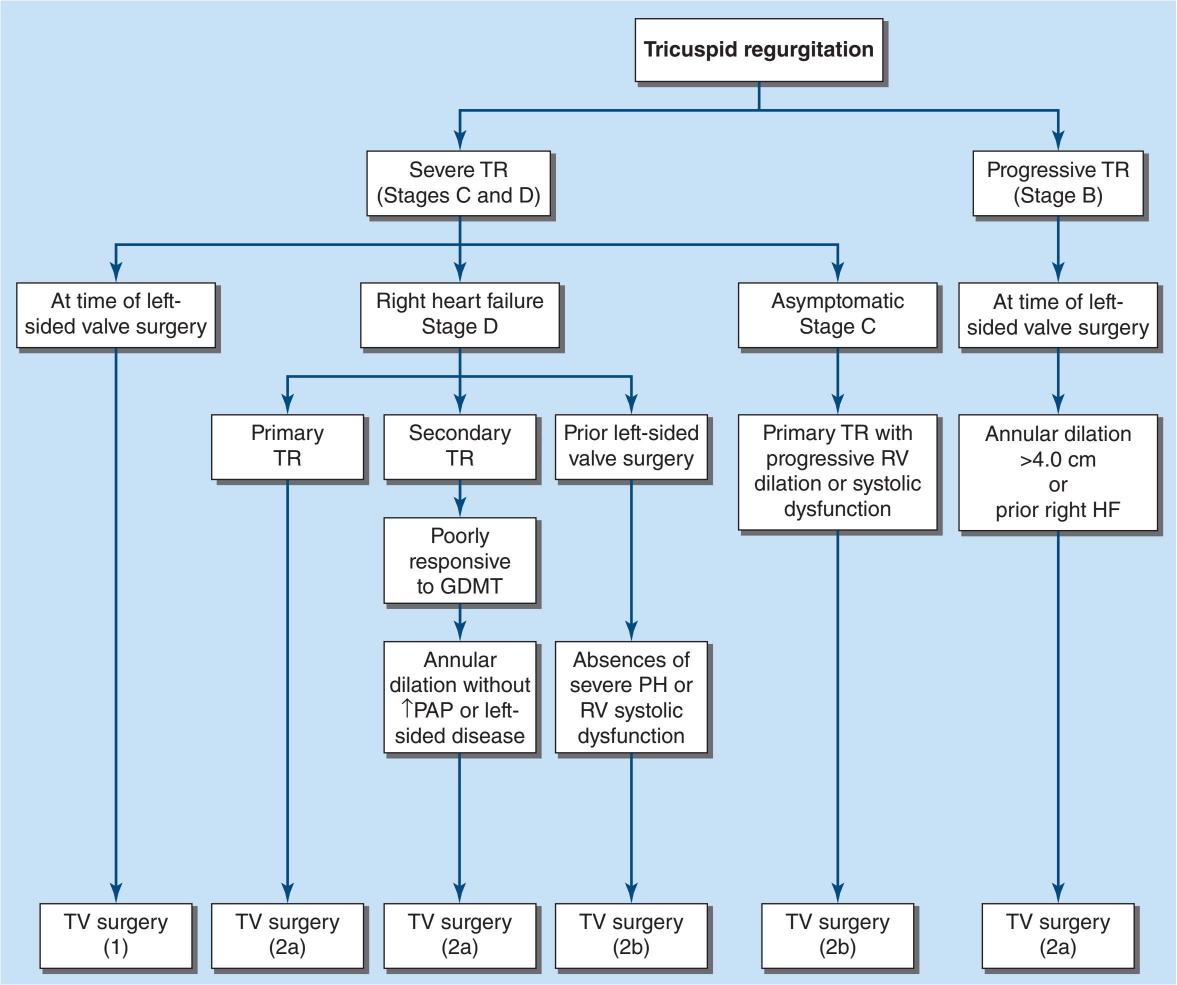
Tricuspid regurgitation
Severe TR (Stages C and D)
Progressive TR (Stage B)
At time of left-sided valve surgery
Right heart failure Stage D
Asymptomatic Stage C
At time of left-sided valve surgery
Primary TR
Secondary TR
Prior left-sided valve surgery
Primary TR with progressive RV dilation or systolic dysfunction
Annular dilation >4.0 cm or prior right HF
Poorly responsive to GDMT
Annular dilation without ↑PAP or left-sided disease
Absences of severe PH or RV systolic dysfunction
TV surgery (1)
TV surgery (2a)
TV surgery (2a)
TV surgery (2b)
TV surgery (2b)
TV surgery (2a)

FIGURE 277-2 Management of tricuspid regurgitation. See legend for Fig. 272-4 for explanation of treatment recommendations (Class I, IIa, IIb) and disease stages (B, C, D). Preoperative coronary angiography should be performed routinely as determined by age, symptoms, and coronary risk factors. Cardiac catheterization and angiography may also be helpful when there is a discrepancy between clinical and noninvasive findings. GDMT, guideline-directed management and therapy; HF, heart failure; PAP, pulmonary artery pressure; PH, pulmonary hypertension; RV, right ventricular; TR, tricuspid regurgitation; TV, tricuspid valve. Annular dilation is defined by >40 mm on transthoracic echocardiography (>21 mm/m^2) or >70 mm on direct intraoperative measurement. *(Reproduced with permission from CM Otto et al: 2020 AHA/ACC Guideline for management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 143:e72, 2021.)*

in leaflet entrapment or perforation. In addition, pacing from the right ventricular apex can result in dyssynchronous ventricular contraction and functional TR.

■ PATHOPHYSIOLOGY

The incompetent tricuspid valve allows blood to flow backward from the RV into the RA, the volume of which is dependent on the driving pressure (i.e., RV systolic pressure) and the size of the regurgitant orifice. The severity and physical signs of TR can vary as a function of PA systolic pressure (in the absence of RV outflow tract stenosis), the dimensions of the tricuspid valve annulus, the respiratory cycle-dependent changes in RV preload, and RA compliance. RV filling is increased during inspiration. With TR, forward CO is reduced and does not augment with exercise. Significant degrees of TR will lead to RA enlargement and elevation of the RA and jugular venous pressures with prominent *c-v* waves in the pulse tracings. Progressively severe TR can lead to "ventricularization" of the RA wave form **(see Fig. 246-1*B*)**. Severe TR is also characterized by RV dilation (RV volume overload) and eventual systolic dysfunction, the progression of which can be accelerated by a concomitant pressure load from PA hypertension or by myocardial fibrosis from previous injury.

■ SYMPTOMS

Mild or moderate degrees of TR are usually well tolerated in the absence of other hemodynamic disturbances. Because TR often coexists with left-sided valve lesions, LV dysfunction, and/or PA hypertension, symptoms related to these lesions may dominate the clinical picture. Atrial functional TR due to atrial fibrillation is usually accompanied by palpitations. Fatigue and exertional dyspnea owing to reduced forward CO are early symptoms of isolated, severe TR. As the disease progresses and RV function declines, patients may report cervical pulsations, abdominal fullness/bloating, diminished appetite, and muscle wasting, although with progressive weight gain and painful swelling of the lower extremities.

■ PHYSICAL FINDINGS

The neck veins in patients with severe TR are distended with prominent *c-v* waves and rapid *y* descents (in the absence of TS). TR is more often diagnosed by examination of the neck veins than by auscultation of the heart sounds. Other findings may include marked hepatomegaly with systolic pulsations, ascites, pleural effusions, edema, and a positive hepatojugular reflux sign. A prominent RV pulsation in the left parasternal region and a blowing holosystolic murmur along the lower left sternal margin, which may be intensified during inspiration (Carvallo's sign) and reduced during expiration or the strain phase of the Valsalva maneuver, are characteristic findings. The murmur of TR may sometimes be confused with that of mitral regurgitation (MR) unless attention is paid to its variation during the respiratory cycle and the extent of RV enlargement is appreciated. AF may be the driver of TR or may emerge later in the disease with ventricular and atrial remodeling.

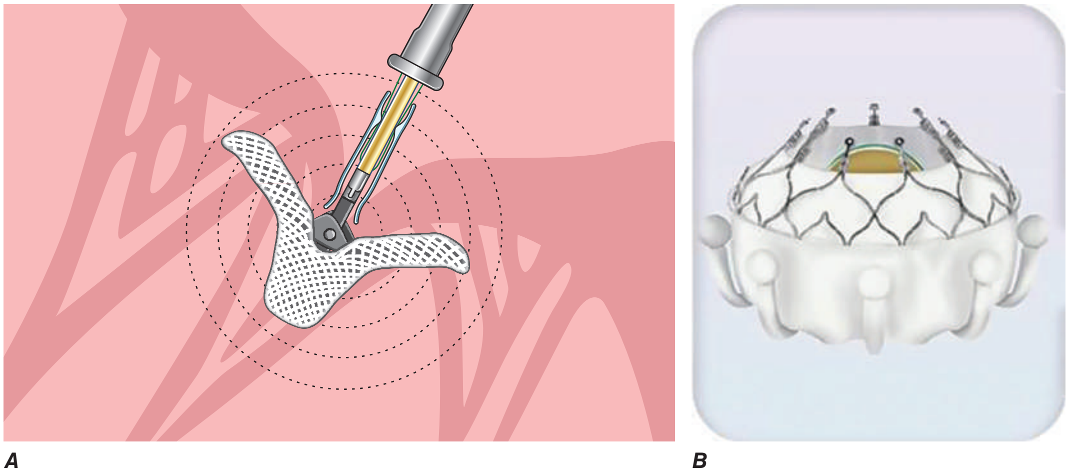

FIGURE 277-3 Transcatheter tricuspid valve (TV) repair and replacement. ***A.*** Shown is a clip device that can be used to grasp the leading edges of the TV leaflets to reduce the severity of tricuspid regurgitation (TR) by decreasing the effective orifice area of the valve. *(TriClip is a trademark of Abbott or its related companies.)* ***B.*** Bioprosthetic heart valve replacement approved for percutaneous treatment of severe TR is selected for high or prohibitive surgical risk patients. *(EVOQUE is a trademark of Edwards Lifesciences.)*

■ LABORATORY EXAMINATION

The ECG may show changes characteristic of the lesion responsible for the TR, e.g., an inferior Q-wave MI suggestive of a prior RV MI, RVH, or a bizarre right bundle branch block–type pattern with preexcitation in patients with Ebstein's anomaly. ECG signs of RA enlargement may be present in patients with sinus rhythm; AF is frequently noted. The chest x-ray may show RA and RV enlargement, depending on the chronicity and severity of TR. TTE is usually definitive with demonstration of RA dilation and RV volume overload and prolapsing, flail, scarred, or displaced/tethered tricuspid leaflets with annular dilatation; the diagnosis and assessment of TR can be made by color flow Doppler imaging **(see Figs. 248-8 and 277-1)**. Severe TR is accompanied by hepatic vein systolic flow reversal. Continuous wave Doppler of the TR velocity profile is useful in estimating PA systolic pressure, except when the TR is very severe and the jet velocity is blunted by rapidly increasing RA pressure. Accurate assessment of TR severity, PA pressures, and RV size and systolic function with TTE can be quite challenging in many patients. Real-time three-dimensional echocardiography and cardiac magnetic resonance (CMR) imaging provide alternative imaging modalities to aid in the assessment of TR severity, although they are not widely available. In patients with severe TR, the CO is usually markedly reduced, and the RA pressure pulse may not exhibit an *x* descent during early systole but rather show a prominent *c-v* wave with a rapid *y* descent. The mean RA and RV end-diastolic pressures are often elevated. Exercise testing can be used to assess functional capacity in patients with asymptomatic severe TR. The prognostic significance of exercise-induced changes in TR severity and RV function has not been well studied. (See Fig. 277-1.)

TREATMENT

Tricuspid Regurgitation (Fig. 277-2)

Diuretics can be useful for patients with severe TR and signs of right heart failure. An aldosterone antagonist may be particularly helpful because many patients have secondary hyperaldosteronism from marked hepatic congestion. Therapies to reduce elevated PA pressures and/or pulmonary vascular resistance, including those targeted at left-sided heart disease, can also be considered for patients with PA hypertension and severe secondary TR. Tricuspid valve surgery is recommended for patients with *severe TR* who are undergoing left-sided valve surgery and is also undertaken frequently for treatment of even moderate TR in patients undergoing left-sided valve surgery especially those with tricuspid annular dilation (>40 mm), a history of right heart failure, or PA hypertension. Operation most often comprises repair rather than replacement in these settings and has become more routine in many surgical centers. In a randomized trial of tricuspid valve annuloplasty surgery at the time of mitral valve surgery for degenerative MR in patients with *moderate TR* or *mild TR with annular dilation*, there was a significant reduction in the 2-year incidence of a composite endpoint of reoperation, progression of TR, or all-cause death among those who underwent tricuspid surgery versus those who had mitral valve surgery alone. The composite endpoint was driven by a reduction in the progression of TR. Surgery may also infrequently be required for treatment of severe, primary TR with right heart failure not responsive to standard medical therapy or because of progressively declining RV systolic function. Reported perioperative mortality rates for isolated tricuspid valve surgery (repair and replacement) are high (~8–9%) and likely are influenced by the hazards encountered during reoperation on patients who have undergone previous left-sided valve surgery and have reduced RV function.

There have been several recent advances in the application of transcatheter tricuspid valve repair and replacement techniques in patients with severe TR and high or prohibitive surgical risk. In a randomized trial of percutaneous transcatheter edge-to-edge repair versus medical therapy, clip repair resulted in a significant reduction in TR severity and improvement in quality-of-life scores, but no differences in all-cause mortality or hospitalization for heart failure over 1-year follow-up. The U.S. Food and Drug Administration approved the use of a transcatheter tricuspid valve replacement system in early 2024 based on favorable hemodynamic, echocardiographic, clinical, and quality-of-life results from prospective single-arm studies. **(See Figure 277-3.)**

■ FURTHER READING

Hahn RT et al: Tricuspid regurgitation management for heart failure. J Am Coll Cardiol HF 11:1084, 2023.

Kadri AN et al: Outcomes of patients with severe tricuspid regurgitation and congestive heart failure. Heart 105:1813, 2019.

Kodali S et al: Transfemoral tricuspid valve replacement and one-year outcomes: The TRISCEND study. Eur Heart J 44:4862, 2023.

Otto CM et al: 2020 ACC/AHA guideline for the management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 143:e72, 2021.

Sorajja P et al: Transcatheter repair for patients with tricuspid regurgitation. N Engl J Med 388:1833, 2023.

278 Pulmonic Valve Disease

Patrick T. O'Gara, Joseph Loscalzo

PULMONIC STENOSIS

Pulmonic valve stenosis (PS) is essentially a congenital disorder **(Table 278-1)**. With isolated PS, the valve is typically domed. Dysplastic pulmonic valves are seen as part of the Noonan syndrome **(Chap. 292)**, which maps to chromosome 12. Mutations in the *PTPN1* gene are associated with about half of all cases of Noonan syndrome. Much less common etiologies include carcinoid and obstructing tumors or bulky vegetations. The pulmonic valve is only very rarely affected by the rheumatic process.

■ PATHOPHYSIOLOGY

PS is defined hemodynamically by a systolic pressure gradient between the right ventricle (RV) and the main pulmonary artery (PA). RV hypertrophy (RVH) develops as a consequence of sustained obstruction to RV outflow, and systolic ejection is prolonged. Compared with the ability of the LV to compensate for the pressure overload imposed by aortic stenosis (AS), RV dysfunction from afterload mismatch occurs earlier in the course of PS and at lower peak systolic pressures, because the RV adapts less well to this type of hemodynamic burden. With normal systolic function and cardiac output (CO), severe PS is defined by a peak systolic gradient across the pulmonic valve of >64 mmHg (mean gradient >35 mmHg, Doppler jet velocity >4 m/s); moderate PS correlates with a peak gradient of 36–64 mmHg (Doppler jet velocity 3–4 m/s). Mild PS is characterized by a jet velocity <3 m/s (peak gradient <36 mmHg). PS rarely progresses in patients with mild PS but may worsen with age in those with moderate disease due to valve thickening and calcification. The right atrial (RA) *a* wave elevates in relation to the higher pressures needed to fill a noncompliant, hypertrophied RV. A prominent RA *v* wave signifies functional tricuspid regurgitation (TR) from RV and annular dilation. The CO is maintained until late in the course of the disease.

■ SYMPTOMS

Patients with mild or even moderate PS are usually asymptomatic and first come to medical attention because of a heart murmur (or early systolic click) that leads to echocardiography. With severe PS, patients may report exertional dyspnea or early-onset fatigue. Anginal chest pain from RV oxygen supply-demand mismatch and syncope may occur with very severe forms of obstruction, particularly in the presence of a destabilizing trigger such as atrial fibrillation, fever, infection, anemia, or pregnancy.

■ PHYSICAL FINDINGS

The murmur of mild or moderate PS is mid-systolic in timing, crescendo–decrescendo in configuration, heard best in the left second interspace, and usually introduced by an ejection sound (click) in younger adults whose valves are still pliable. The ejection sound is the only right-sided acoustic event that decreases in intensity with inspiration. This phenomenon reflects premature opening of the pulmonic valve by the elevated RV end-diastolic (post-atrial *a* wave) pressure. The systolic murmur increases in intensity during inspiration. With progressively severe PS, the ejection sound moves closer to the first heart sound and eventually becomes inaudible. A right-sided fourth heart sound may emerge. The systolic murmur peaks later and may persist through the aortic component of the second heart sound (A_2). Pulmonic valve closure is delayed, and the pulmonic component of the second heart sound (P_2) is reduced or absent. A prominent *a* wave, indicative of the higher atrial pressure necessary to fill the noncompliant RV, may be seen in the jugular venous pulse. A parasternal or RV lift can be felt with significant pressure overload. Signs of right heart failure due to RV systolic dysfunction, such as hepatomegaly, ascites, and edema, are uncommon but may appear very late in the disease.

TABLE 278-1 Causes of Pulmonic Valve Disease

VALVE LESION	ETIOLOGIES
Pulmonic stenosis	Congenital Carcinoid Tumor Endocarditis

Pulmonic regurgitation	Primary valve disease
	  Congenital
	  Post-valvotomy
	  Endocarditis
	  Carcinoid
	Annular enlargement
	  Pulmonary hypertension
	  Idiopathic dilation
	  Marfan syndrome

■ LABORATORY EXAMINATION

The electrocardiogram (ECG) will show right axis deviation, RVH, and RA enlargement in adult patients with severe PS. Chest x-ray findings include poststenotic dilation of the main PA in the frontal plane projection and filling of the retrosternal airspace due to RV enlargement on the lateral film. In some patients with RVH, the cardiac apex appears to be lifted off the left hemidiaphragm. The RA may also be enlarged. Transthoracic echocardiography (TTE) allows definitive diagnosis and characterization in most cases, with depiction of the valve and assessment of the jet velocity, gradient, RV function, PA pressures (which should be low), and any associated cardiac lesions. Transesophageal echocardiography (TEE) may be useful in some patients for improved delineation of the RV outflow tract (RVOT) and assessment of infundibular hypertrophy. Cardiac catheterization is not usually necessary for diagnostic purposes, but if performed, pressures should be obtained from just below and above the pulmonic valve with attention to the possibility that a dynamic component to the gradient may exist. The correlation between Doppler assessment of peak instantaneous gradient and catheterization-measured peak-to-peak gradient is weak. The latter may correlate better with the Doppler mean gradient.

TREATMENT

Pulmonic Stenosis

Diuretics can be used to treat symptoms and signs of right heart failure. Provided there is less than moderate pulmonic regurgitation (PR), percutaneous pulmonic balloon valvuloplasty is recommended for symptomatic patients with moderate or severe PS and for asymptomatic patients with a peak gradient >64 mmHg (or mean gradient >35 mmHg). Surgery may be required when the valve is dysplastic (as seen in patients with Noonan's syndrome and other disorders). A multidisciplinary heart team is best positioned to make treatment decisions of this nature.

PULMONIC REGURGITATION

PR may develop as a consequence of primary valve pathology, annular enlargement, or their combination; after surgical treatment of RVOT obstruction in children with such disorders as tetralogy of Fallot; or after percutaneous pulmonic balloon valvotomy (Table 278-1). Carcinoid usually causes mixed pulmonic valve disease with PR and PS. Long-standing severe PA hypertension from any cause can result in dilation of the pulmonic valve ring and PR. Trace or mild PR of no hemodynamic or clinical consequence is frequently observed on TTE in the absence of structural pulmonic valve disease **(Fig. 278-1)**.

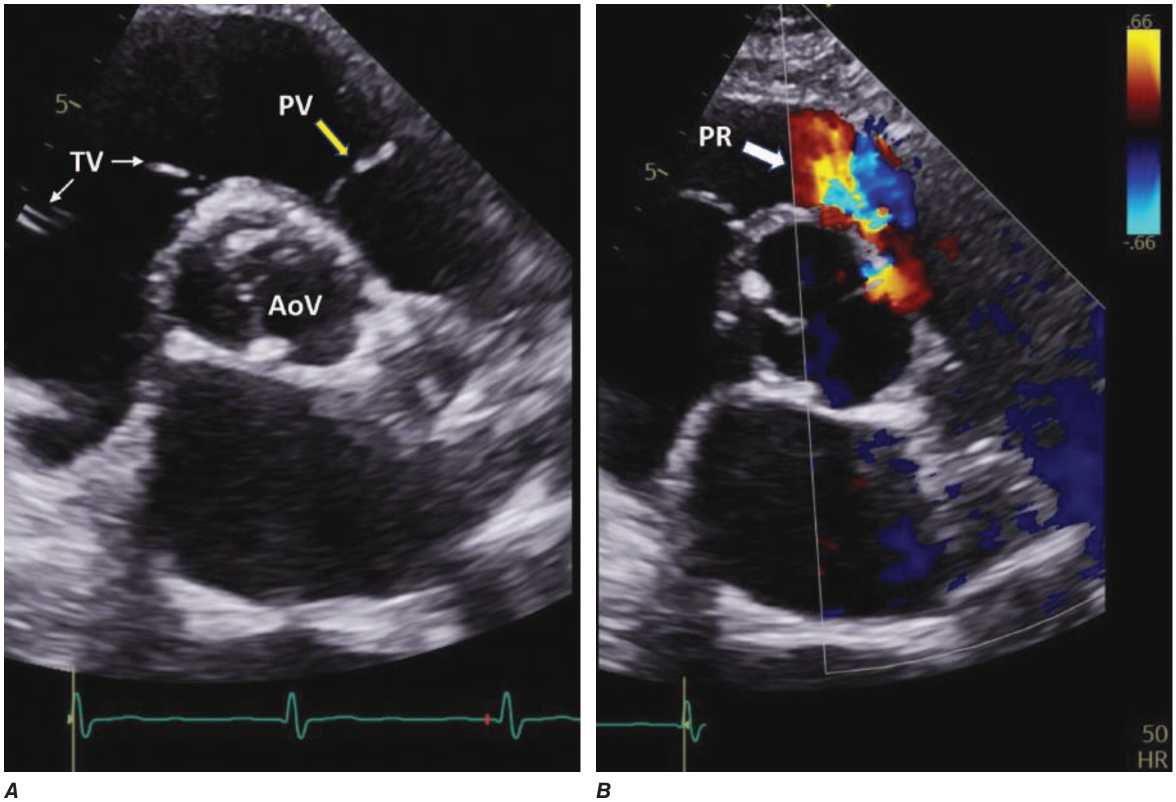

FIGURE 278-1 Pulmonic regurgitation. ***A.*** Transthoracic short axis window at the level of the aortic valve (AoV). The pulmonic valve (PV) is shown with the *yellow arrow*. Tricuspid valve (TV) leaflets are partially visualized (*white arrows*). ***B.*** Color flow Doppler image shows moderate pulmonic regurgitation (PR, *white arrow*).

■ PATHOPHYSIOLOGY

Severe PR results in RV chamber enlargement and eccentric hypertrophy. As is the case for aortic regurgitation (AR), PR is a state of increased preload and afterload. The diastolic pressure gradient between the PA to the RV, which drives the PR, progressively decreases throughout diastole and accounts for the decrescendo nature of the murmur. As RV diastolic pressure increases, the murmur becomes shorter in duration. The forward CO is preserved during the early stages of the disease but may not increase normally with exercise and declines over time. A reduction in RV ejection fraction may be an early indicator of hemodynamic compromise. In advanced stages, there is significant enlargement of the RV and RA with marked elevation of the jugular venous pressure.

■ SYMPTOMS

Mild or moderate degrees of PR do not, by themselves, result in symptoms. Other issues, such as PA hypertension, may dominate the clinical picture. With progressively severe PR and RV dysfunction, fatigue, exertional dyspnea, abdominal fullness/bloating, and lower extremity swelling may be reported.

■ PHYSICAL FINDINGS

The physical examination hallmark of PR is a high-pitched, decrescendo diastolic murmur (Graham Steell murmur) heard along the left sternal border that can be difficult to distinguish from the more frequently appreciated murmur of AR. The Graham Steell murmur may become louder with inspiration and is usually associated with a loud and sometimes palpable P_2 and an RV lift, as would be expected in patients with significant PA hypertension of any cause. Survivors of childhood surgery for tetralogy of Fallot or PS/pulmonary atresia may have an RV-PA conduit that is freely regurgitant because it does not contain a valve. PA pressures in these individuals are not elevated, and the diastolic murmur can be misleadingly low pitched and of short duration despite significant degrees of PR and RV volume overload.

■ LABORATORY EXAMINATION

Depending on both the etiology and severity of PR, the ECG may show findings of RVH and RA enlargement. On chest x-ray, the RV and RA may be enlarged. Pulmonic valve morphology and function can be assessed with transthoracic Doppler echocardiography. RV systolic pressure can be estimated from the tricuspid valve systolic jet velocity. Cardiac magnetic resonance (CMR) imaging can provide greater anatomic detail, particularly in patients with repaired congenital heart disease, and more precise assessment of RV volumes and function. Cardiac catheterization is not routinely necessary but would be performed as part of a planned transcatheter PV procedure.

TREATMENT

Pulmonic Regurgitation

In patients with significant functional PR due to PA hypertension and annular dilation, efforts to reduce pulmonary vascular resistance and pressure should be pursued. Such efforts may include pharmacologic/vasodilator and/or surgical/interventional strategies, depending on the cause of the PA hypertension (e.g., idiopathic PA hypertension, left-sided heart valve disease). Diuretics can be used to treat the manifestations of right heart failure. Surgical valve replacement for primary, severe, pulmonic valve disease, such as carcinoid or endocarditis, is rarely undertaken. Transcatheter pulmonic valve replacement has been successfully performed in many patients with severe PR after childhood repair of tetralogy of Fallot or pulmonic valve stenosis or atresia. This procedure was introduced clinically prior to transcatheter aortic valve replacement. Earlier concerns regarding an excess hazard of infective endocarditis with first-generation valves have led to procedural modifications.

■ FURTHER READING

Otto CM et al: 2020 AHA/ACC guideline for the management of patients with valvular heart disease. A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 143:e72, 2021.

Stout KK et al: 2018 ACC/AHA guidelines for the management of adults with congenital heart disease. J Am Coll Cardiol 73:e81, 2019.

279 Multiple and Mixed Valvular Heart Disease

Patrick T. O'Gara, Joseph Loscalzo

Many acquired and congenital cardiac lesions may result in stenosis and/or regurgitation of one or more heart valves. For example, rheumatic heart disease can involve the mitral (mitral stenosis [MS], mitral regurgitation [MR], or MS and MR), aortic (aortic stenosis [AS], aortic regurgitation [AR], or AS and AR), and tricuspid (tricuspid stenosis [TS], tricuspid regurgitation [TR], or TS and TR) valve, alone or in combination. The common association of functional TR with significant mitral valve disease is discussed in Chap. 277. Severe mitral annular calcification can result in regurgitation (due to decreased annular shortening during systole) and mild or moderate stenosis (caused by extension of the calcification onto the leaflets resulting in restricted valve opening). Patients with severe AS and left ventricular (LV) remodeling may develop functional MR that may not improve after isolated aortic valve replacement (AVR). Primary MR due to mitral valve prolapse or chordal rupture has been noted in patients with severe AS. Aortic valve infective endocarditis (IE) may secondarily involve the mitral apparatus either by abscess formation and contiguous spread via the intervalvular fibrosa or by "drop metastases" from the aortic leaflets onto the anterior leaflet of the mitral valve. Mediastinal radiation may result in aortic, mitral, and even tricuspid valve disease, most often with mixed stenosis and regurgitation. Carcinoid heart disease may cause mixed lesions of either or both the tricuspid and pulmonic valves. Ergotamines, and the previously used combination of fenfluramine and phentermine, can rarely result in mixed lesions of the aortic and/or mitral valve. Patients with Marfan syndrome may have both AR from aortic root dilation and MR due to mitral valve prolapse (MVP). Myxomatous degeneration causing prolapse of multiple valves (mitral, aortic, tricuspid) can also occur in the absence of an identifiable connective tissue disorder. Bicuspid aortic or pulmonic valve disease can result in mixed stenosis and regurgitation. The former is also associated with aortic aneurysm disease and a predisposition to aortic dissection.

PATHOPHYSIOLOGY

In patients with multivalvular heart disease, the pathophysiologic derangements associated with the more proximal valve disease can mask the full expression of the attributes of the more distal valve lesion. For example, in patients with rheumatic mitral and aortic valve disease, the reduction in cardiac output (CO) imposed by the mitral valve disease will decrease the magnitude of the hemodynamic derangements related to the severity of the aortic valve lesion (stenotic, regurgitant, or both). Alternatively, the development of atrial fibrillation (AF) during the course of MS can lead to sudden worsening in a patient whose aortic valve disease was not previously felt to be significant. The development of reactive pulmonary vascular disease, sometimes referred to as a "secondary obstructive lesion in series," can impose an additional challenge in these settings. As CO falls with progressive tricuspid valve disease, the severity of any associated mitral or aortic disease can be underestimated.

One of the most common examples of multivalve disease is that of functional TR in the setting of significant mitral valve disease. Functional TR occurs as a consequence of right ventricular and annular dilation; pulmonary artery (PA) hypertension may be present. The tricuspid leaflets are morphologically normal. Progressive degrees of TR lead to right ventricular volume overload and continued chamber and annular dilation. The TR is usually central in origin; reflux into the right atrium (RA) is expressed as large, systolic *c-v* waves in the RA pressure pulse. The height of the *c-v* wave is dependent on RA compliance and the volume of regurgitant flow. The RA waveform may become "ventricularized" in advanced stages of chronic, severe TR. CO falls and the severity of the associated mitral valve disease may become more difficult to appreciate. Findings related to advanced right heart failure (e.g., ascites, edema) predominate. Primary rheumatic tricuspid valve disease may occur with rheumatic mitral disease and cause hemodynamic changes reflective of TR, TS, or their combination. With TS, the *y* descent in the RA pressure pulse is prolonged. Typically, however, findings related to the mitral valve disease predominate over those related to the tricuspid valve disease.

Another example of rheumatic, multivalve disease involves the combination of mitral and aortic valve pathology, frequently characterized by MS and AR. In isolated MS, LV preload and diastolic pressure are reduced as a function of the severity of inflow obstruction. With concomitant AR, however, LV filling is enhanced and diastolic pressure may rise depending on the compliance characteristics of the chamber. Because the CO falls with progressive degrees of MS, transaortic valve flows will decline, masking the potential severity of the aortic valve lesion (AR, AS, or its combination). As noted above, onset of AF in such patients can be especially deleterious. The loss of atrial systole with AF may result in a critical reduction in CO, a rise in left atrial (LA) and LV diastolic pressures, and a further deleterious increase in heart rate.

Secondary (functional) MR may complicate the course of some patients with severe AS. The mitral valve leaflets and chordae tendineae are usually normal. Incompetence is related to changes in LV geometry (remodeling) and abnormal systolic tethering of the leaflets in the context of markedly elevated LV systolic pressures. Relief of the excess afterload with surgical or transcatheter AVR may result in reduction of the secondary MR. Persistence of significant, secondary MR following AVR is associated with impaired functional outcomes and reduced survival. Identification of patients who would benefit from concomitant treatment of their secondary MR at time of AVR is quite challenging. Most surgeons perform mitral valve repair of moderate-to-severe or severe secondary MR at time of surgical AVR. Significant primary MR may also coexist with AS and is routinely managed with repair (or replacement) at the time of AVR. There is increasing experience with the combination of transcatheter aortic valve implantation (TAVI) and transcatheter edge-to-edge mitral valve repair (TEER) in high surgical risk patients with severe AS and moderate-severe primary or secondary MR. Decision-making regarding the need for mitral valve intervention can be challenging.

In patients with mixed AS and AR, assessment of valve stenosis can be influenced by the magnitude of the regurgitant valve flow. Because transvalvular systolic flow velocities are augmented in patients with AR and preserved LV systolic function, the LV-aortic Doppler-derived pressure gradient and the intensity of the systolic murmur will be elevated to values higher than expected for the true systolic valve orifice size as measured by planimetry. Uncorrected, the Gorlin formula, which relies on forward CO (systolic transvalvular flow) and the mean pressure gradient for calculation of valve area, is not accurate in the setting of mixed aortic valve disease. Similar considerations apply to patients with mixed mitral valve disease. The peak mitral valve Doppler E wave velocity (v_0) is increased in the setting of severe MR because of enhanced early diastolic flow and may not accurately reflect the contribution to LA hypertension from any associated MS. When either AR or MR is the dominant lesion in patients with mixed aortic or mitral valve disease, respectively, the LV is dilated. When AS or MS predominates, LV chamber size will be normal or small. It can sometimes be difficult to ascertain whether stenosis or regurgitation is the dominant lesion in patients with mixed valve disease, although an integrated clinical and noninvasive assessment can usually provide clarification for purposes of patient management. For patients with moderate, mixed AS and AR in whom stenosis is the dominant lesion, the natural history tends to parallel what might be expected for isolated severe AS, and the treatment approach should be accordingly aligned.

Patients with significant AS, a nondilated LV chamber, and concentric hypertrophy will poorly tolerate the abrupt development of aortic regurgitation, as may occur, for example, with IE or after surgical AVR or TAVI complicated by paravalvular leakage. The noncompliant LV is not prepared to accommodate the sudden volume load, and as a result, LV diastolic pressure rises rapidly and severe heart failure develops.

Indeed, significant paravalvular regurgitation is a significant risk factor for short- to intermediate-term death following transcatheter AVR. Conditions in which the LV may not be able to dilate in response to chronic AR (or MR) include radiation heart disease, cardiac amyloid, and, in some patients, the cardiomyopathy associated with obesity and diabetes. Noncompliant ventricles of small chamber size predispose to earlier onset diastolic dysfunction and heart failure in response to any further perturbation in valve function.

■ SYMPTOMS

Compared with patients with isolated, single-lesion valve disease, patients with multiple or mixed valve disease may develop symptoms at a relatively earlier stage in the natural history of their disease. Symptoms such as exertional dyspnea and fatigue are usually related to elevated filling pressures, reduced CO, or their combination. Palpitations may signify AF and identify mitral valve disease as an important component of the clinical presentation, even when not previously suspected. Chest pain compatible with angina could reflect left or right ventricular oxygen supply/demand mismatch on a substrate of hypertrophy and pressure/volume overload with or without superimposed coronary artery disease. Symptoms related to right heart failure (abdominal fullness/bloating, edema) are late manifestations of advanced disease.

■ PHYSICAL FINDINGS

Mixed disease of a single valve is most often manifested by systolic and diastolic murmurs, each with the attributes expected for the valve in question. Thus, patients with AS and AR will have characteristic mid-systolic, crescendo-decrescendo and blowing, decrescendo diastolic murmurs at the base of the heart in the second right interspace and along the left sternal edge, respectively. Many patients with significant AR have mid-systolic outflow murmurs even in the absence of valve sclerosis/stenosis, and other findings of AS must be sought. The separate murmurs of AS and AR can occasionally be difficult to distinguish from the continuous murmurs associated with either a patent ductus arteriosus (PDA) or ruptured sinus of Valsalva aneurysm. With mixed aortic valve disease, the systolic murmur should end before, and not envelope or extend through, the second heart sound (S_2). The murmur associated with a PDA is heard best to the left of the upper sternum. The continuous murmur heard with a ruptured sinus of Valsalva aneurysm is often first appreciated after an episode of acute chest pain. An early ejection click, which usually defines bicuspid aortic valve disease in young adults, is often not present in patients with congenital mixed AS and AR. As noted above, both the intensity and duration of these separate murmurs can be influenced by a reduction in CO and transvalvular flow due to coexistent mitral valve disease or AF. In patients with isolated MS and MR, expected findings would include a blowing, holosystolic murmur and a mid-diastolic rumble (with or without an opening snap) best heard at the cardiac apex. An irregularly irregular heart rhythm in such patients would likely signify AF. Findings with TS and TR would mimic those of left-sided MS and MR, save for the expected changes in the murmurs with respiration. The murmurs of pulmonic stenosis and regurgitation behave in a fashion directionally similar to AS and AR; dynamic changes during respiration should be noted. Specific attributes of these cardiac murmurs are reviewed in **Chaps. 44 and 277.**

■ LABORATORY EXAMINATION

The electrocardiogram (ECG) may show evidence of ventricular hypertrophy and/or atrial enlargement. ECG signs indicative of right-sided cardiac abnormalities in patients with left-sided valve lesions should prompt additional assessment for PA hypertension and/or right-sided valve disease. The presence of AF in patients with aortic valve disease may be a clue to the presence of previously unsuspected mitral valve disease in the appropriate context. The chest x-ray can be reviewed for evidence of cardiac chamber enlargement, valve and/or annular calcification, and any abnormalities in the appearance of the pulmonary vasculature. The latter could include enlargement of the main and proximal pulmonary arteries with PA hypertension and pulmonary venous redistribution/engorgement or Kerley B lines with increasing degrees of LA and pulmonary venous hypertension. An enlarged azygos vein in the frontal projection indicates RA hypertension. Roentgenographic findings not expected based on a single or mixed valve lesion may reflect other valve disease.

Transthoracic echocardiography (TTE) is the most commonly used imaging modality for the diagnosis and characterization of multiple and/or mixed valvular heart disease and may often demonstrate findings not clinically suspected. Transesophageal echocardiography (TEE) may sometimes be required for more accurate assessment of valve anatomy (specifically, the mitral valve) and when IE is considered responsible for the clinical presentation. TTE findings of particular interest include those related to valve morphology and function, calcification, chamber size, ventricular wall thickness, biventricular function, estimated PA systolic pressure, and the dimensions of the great vessels, including the root and ascending aorta, PA, and inferior vena cava. Exercise testing (with or without echocardiography) can be useful when the degree of functional limitation reported by the patient is not adequately explained by the findings on TTE performed at rest. An integrated assessment of the clinical and TTE findings is needed to help determine the dominant valve lesion(s) and establish an appropriate plan for treatment and follow-up. Natural history is usually influenced to a relatively greater degree by the dominant lesion.

Cardiac magnetic resonance (CMR) imaging can be used to provide additional anatomic and physiologic information when echocardiography proves suboptimal but is less well suited to the evaluation of valve morphology. Cardiac computed tomography (CT) has been used to assess intracardiac structures in patients with complicated IE. It is invaluable in planning for transcatheter valve implantation. Coronary CT angiography provides a noninvasive alternative for the assessment of coronary artery anatomy prior to surgery or transcatheter intervention.

Invasive hemodynamic evaluation with right and left heart catheterization may be required to characterize more completely the individual contributions of each lesion in patients with either multiple or mixed valvular heart disease. It is strongly recommended when there is a discrepancy between the clinical and noninvasive findings in a symptomatic patient. Measurement of PA pressures and calculation of pulmonary vascular resistance (PVR) can help inform clinical decision-making in certain patient subsets, such as those with advanced mitral and tricuspid valve disease. It is important to identify any potential contribution to the clinical picture from pulmonary vascular disease. Attention to the accurate assessment of CO is essential. Coronary angiography (if indicated) can be performed as part of the procedure. Contrast ventriculography and great vessel angiography are performed infrequently.

TREATMENT

Multiple and Mixed Valve Disease

Management of patients with multiple or mixed valve disease can be challenging. As noted above, it is helpful to determine the dominant valve lesion and proceed according to the treatment and follow-up recommendations for it **(Chaps. 272–278)**, being mindful of deviations from the expected course due to the contributions of more than one valve lesion. For example, AF that emerges in the course of moderate mitral valve disease may precipitate heart failure in patients with concomitant, severe aortic valve disease that was previously asymptomatic.

Medical therapies are limited and include diuretics when indicated for relief of congestion and anticoagulation to prevent stroke and thromboembolism in patients with AF. Blood pressure–lowering medications may be needed to treat systemic hypertension, which may aggravate left-sided regurgitant valve lesions, but should be initiated and titrated carefully. Pulmonary vasodilators to lower PVR are not generally effective in this context.

There is a paucity of evidence to inform practice guidelines for surgical and/or transcatheter valve intervention in patients with multiple or mixed valve disease. When there is a clear, dominant lesion, as for example in a patient with severe AS and mild AR, indications for intervention are straightforward and follow those recommended for patients with AS **(Chap. 272)**. In other patients, however, there is less clarity, and decisions regarding intervention should be based on several considerations, including those related to lesion severity, ventricular remodeling, functional capacity, and PA pressures. In this regard, it is important to realize that patients with multiple and/or mixed valve disease may develop limiting symptoms or signs of physiologic impairment even with moderate valve lesions.

Concomitant aortic and mitral valve replacement surgery is associated with a significantly higher perioperative mortality risk than replacement of either valve alone; therefore, operation should be carefully considered. Double valve replacement surgery is usually performed for treatment of severe (unrepairable) valve disease at both locations and for the combination of severe disease at one location with moderate disease at the other to avoid the hazards of reoperation in the intermediate to late term for progressive disease of the unoperated valve. In addition, the presence of a prosthesis in the aortic position significantly restricts surgical exposure of the native mitral valve. The need for double valve replacement may also impact the decision regarding the type of prosthesis (i.e., mechanical vs tissue). In selected patients, TAVI for severe AS followed by edge-to-edge clip repair for severe MR can be accomplished during the same procedure.

Tricuspid valve repair for moderate or severe secondary (functional) TR at the time of left-sided valve surgery is now commonplace, particularly if there is dilation of the tricuspid annulus (>40 mm). The addition of tricuspid valve repair, consisting usually of insertion of an annuloplasty ring, adds little time or complexity to the procedure and is well tolerated, but may be associated with an excess risk of heart block and the need for a permanent pacemaker. Reoperation for repair (or replacement) of progressive TR years after initial surgery for left-sided valve disease, on the other hand, is associated with a relatively high perioperative mortality risk. Mitral valve repair or replacement for moderate or severe secondary MR at time of AVR for AS can usually be undertaken with acceptable risk for perioperative death or major complication.

The presence of moderate or severe MR in patients with rheumatic MS is a contraindication to percutaneous mitral balloon commissurotomy (PMBC). TAVI can be performed for mixed AS and AR when the anatomic findings related to annulus size, coronary height, and the distribution of calcium are favorable. Transcatheter management of both severe AS and severe primary or secondary MR (with deployment of an edge-to-edge clip) has been undertaken with increasing frequency in appropriately selected patients with prohibitive or high surgical risk. Further advances in transcatheter treatments for multiple and mixed valve disease are anticipated.

FURTHER READING

Alaour B et al: Combined significant aortic stenosis and mitral regurgitation: Challenges in timing and type of intervention. Can J Cardiol 40:235, 2024.

Egbe AC et al: Outcomes in moderate mixed aortic valve disease: Is it time for a paradigm shift? J Am Coll Cardiol 67:2321, 2016.

Gammie JS et al: Concomitant tricuspid repair in patients with degenerative mitral regurgitation. N Engl J Med 286:327, 2022.

Otto CM et al: 2020 AHA/ACC guidelines for management of patients with valvular heart disease. A report of the American Heart Association Joint Commission on Clinical Practice Guidelines. Circulation 143:e72, 2021.

280

Congenital Heart Disease in the Adult

Anne Marie Valente, Michael J. Landzberg

PREVALENCE

The number of adults with congenital heart disease (CHD) living in the United States is estimated to be at least 1.4 million. It is now projected that >10% of adults living with CHD in Europe will be over 60 years old by 2030. The majority of adults with CHD were diagnosed in childhood, although a substantial percentage may have CHD first recognized as adults. Lifelong follow-up in coordination with, or directly by, clinicians with expertise in adult congenital heart disease (ACHD) is recommended. In this chapter, we will review the current field of ACHD, with an introduction to CHD nomenclature and cardiac development. This is followed by a summary of the more common CHD lesions that may be diagnosed in adulthood. Lastly, some of the common repaired CHD lesions that are encountered in adults are discussed. Throughout the chapter, to aid in the understanding of congenital cardiac anatomy and physiology, we include figures displaying the passage of blood flow between blood vessels and cardiac chambers in various disorders **(Fig. 280-1)**.

THE CHANGING LANDSCAPE OF ADULT CHD

A Relatively New Subspecialty in Cardiovascular Disease

Over the past two decades, the field of caring for adults with CHD has blossomed, and several nationwide initiatives have been initiated to standardize care. The American College of Cardiology and American Heart Association developed guidelines for the care of adults with CHD, first published in 2008 and revised in 2018, which emphasize the need for collaboration among primary care practitioners, cardiologists, and ACHD subspecialty cardiologists. The body of medical knowledge and competencies attendant with ACHD combined with skill acquisition in coordination of complex care over a patient's medical lifetime led in 2015 to ACHD board certification examinations by the American Board of Medical Subspecialties, as well as the establishment of requirements for advanced fellowship training in ACHD care by the Accreditation Council for Graduate Medical Education. In temporal association, the Adult Congenital Heart Association (ACHA) developed a process for ACHD care program accreditation based on standardization of infrastructural components felt requisite to achieve quality outcomes for ACHD.

SPECIAL CONSIDERATIONS FOR THE ACHD PATIENT

Due to the need for lifelong care, it is essential that pediatric cardiology programs partner with ACHD programs to provide successful transition of patients. However, gaps in care are common during transition, much of which may be due to disparities in social determinants of health, such as race, ethnicity, socioeconomic status, access to insurance, and residence in geographically remote locations. Additionally, adults with CHD may not recognize subtle changes in their exercise capacity, some of which are associated with worse survival; by the time symptoms are recognized, irreversible physiologic changes may have occurred. ACHD patients are, therefore, advised to undergo regular evaluations for surveillance of anatomic, hemodynamic, and electrophysiologic sequelae that may be present. In addition, specific situations may arise in which it is prudent to review care in consultation with an ACHD specialist, several of which are outlined below.

Noncardiac Surgery Nearly all adults with CHD can be classified with stage A (harboring risk) or greater degrees of heart failure. As such, adults with CHD may demonstrate limited hemodynamic reserve to altered myocardial perfusion or loading conditions and may have subclinical organ dysfunction that is not recognized by standard laboratory assessment. Comprehensive, multispecialty assessment and care strategy review are recommended in advance of invasive or

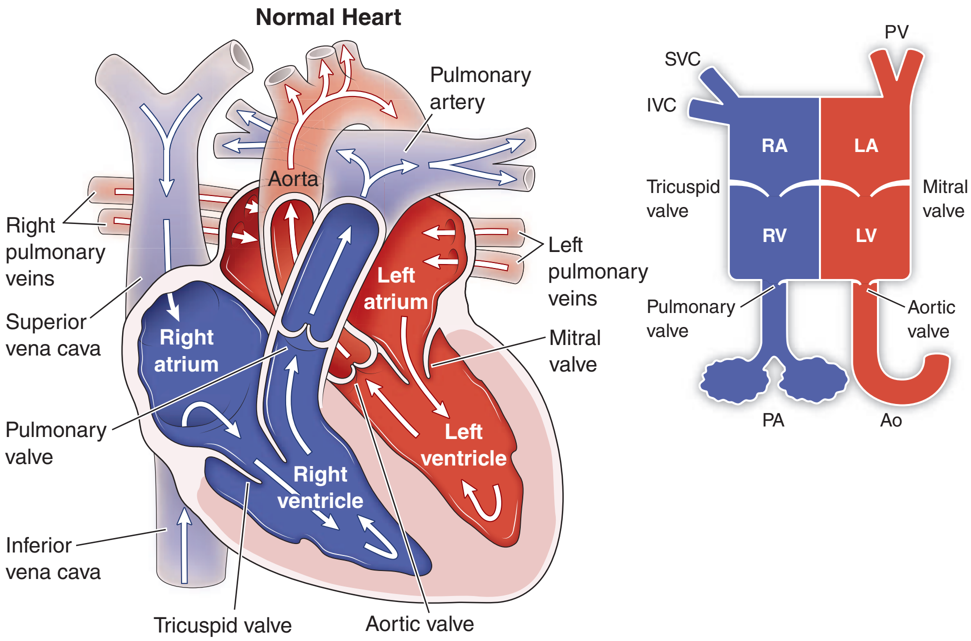
Normal Heart
Pulmonary artery
Aorta
Right pulmonary veins
Superior vena cava
Right atrium
Left atrium
Left pulmonary veins
Mitral valve
Pulmonary valve
Left ventricle
Right ventricle
Inferior vena cava
Tricuspid valve
Aortic valve
SVC
PV
IVC
RA
LA
Tricuspid valve
Mitral valve
RV
LV
Pulmonary valve
Aortic valve
PA
Ao

FIGURE 280-1 Normal heart. Understanding of congenital cardiac anatomy and physiology is facilitated by use of box diagrams, displaying passage of blood flow between blood vessels and cardiac chambers. Labeling (e.g., structure names, arrows to denote direction of flow, coloring to represent oxygen saturation, connections or obstructions, chamber or vascular pressures, oxygen saturations) can aid in representation. Ao, aorta; IVC, inferior vena cava; LA, left atrium; LV, left ventricle; PA, pulmonary artery; PV, pulmonary veins; RA, right atrium; RV, right ventricle; SVC, superior vena cava.

operative procedures for adults with CHD. Figure 280-2 illustrates the multiorgan considerations that should be considered in adults with CHD during perioperative resuscitation and convalescence. Anesthetic management requires knowledge of anatomy, physiologic consequence of underlying defects, myocardial and vascular performance, presence and nature of previous palliative procedures and residual shunts, alteration of venous or arterial pathways within the circulation, and status of noncardiovascular organ physiology.

Pregnancy Women with CHD should receive counseling regarding both maternal and fetal risks prior to conceiving a pregnancy and should be cared for in institutions with experience in treating CHD during pregnancy. Preconception evaluation includes detailed medical history, with particular attention to the women's functional capacity, which is closely linked to maternal and fetal outcomes. Table 280-1 lists the World Health Organization classification of risk during pregnancy in women with heart disease; women at risk should be strongly counseled about the significant risks of morbidity and mortality during pregnancy and the postpartum period. Normal physiologic hemodynamic changes of pregnancy are significant, occur over a relatively condensed period of time, and may be compounded in adults with CHD. Silversides and colleagues have developed a weighted-risk score for pregnant women with heart disease, based on a large registry known as CARPREG 2. The highest-weighted risk factors (weight of 3 points) include a prior history of cardiac events or arrhythmias, decreased functional status (New York Heart Association class ≥III), and presence of a mechanical heart valve. Risk factors that account for 2 points include ventricular dysfunction, high-risk left-sided valve disease/left ventricular outflow tract obstruction, pulmonary hypertension, coronary artery disease, and high-risk aortopathy. One point is assigned for late pregnancy assessment or no prior cardiac intervention. In this cohort, 16% of women experienced an adverse cardiac outcome, primarily heart failure and arrhythmia related. The predicted risks for cardiac events stratified according to point score were as follows: ≤1 point, 5%; 2 points, 10%; 3 points, 15%; 4 points, 22%; and >4 points, 41%.

Prepregnancy medications should be reviewed to ensure their safety in pregnancy. Alternatives to angiotensin-converting enzyme (ACE) inhibitors, angiotensin receptor blockers, direct oral anticoagulants, and endothelin receptor blockers should be considered, as these agents are teratogenic and contraindicated during pregnancy and should be discontinued. Women requiring anticoagulation must be advised of the challenges of managing anticoagulation during pregnancy, and individualized strategies should be developed. A fetal echocardiogram between 18 and 22 weeks of gestation is advised for patients with CHD. Additionally, both men and women with CHD should be counseled regarding the risk of CHD in their offspring.

CONGENITAL TERMINOLOGY, DEVELOPMENT, AND GENETICS

Congenital Nomenclature One of the challenges in caring for adults with CHD is the inconsistent terminology used to describe the congenital heart lesions. Several classification systems have been proposed, from the initial descriptions by Maude Abbott, Maurice Lev, and Jesse Edwards, to the extensive characterizations by Stella and Richard Van Praagh and Robert Anderson. In this chapter, we follow a segmental approach. The heart is composed of several segments that are

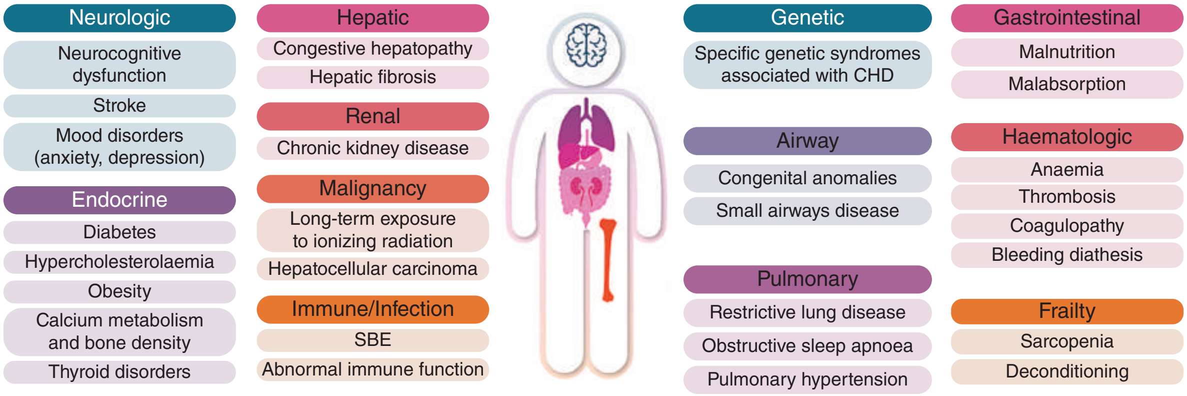
Neurologic
Neurocognitive dysfunction
Stroke
Mood disorders (anxiety, depression)
Endocrine
Diabetes
Hypercholesterolaemia
Obesity
Calcium metabolism and bone density
Thyroid disorders
Hepatic
Congestive hepatopathy
Hepatic fibrosis
Renal
Chronic kidney disease
Malignancy
Long-term exposure to ionizing radiation
Hepatocellular carcinoma
Immune/Infection
SBE
Abnormal immune function
Genetic
Specific genetic syndromes associated with CHD
Airway
Congenital anomalies
Small airways disease
Pulmonary
Restrictive lung disease
Obstructive sleep apnoea
Pulmonary hypertension
Gastrointestinal
Malnutrition
Malabsorption
Haematologic
Anaemia
Thrombosis
Coagulopathy
Bleeding diathesis
Frailty
Sarcopenia
Deconditioning

FIGURE 280-2 Noncardiac considerations in adults with congenital heart disease (CHD). SBE, subacute bacterial endocarditis. *(Modified from J Buber et al: Common congenital heart problems in acute and intensive care. Eur Heart J Acute Cardiovasc Care 12:267, 2023.)*

TABLE 280-1 Modified World Health Organization (mWHO) Classification of Heart Disease in Pregnancy

	mWHO I	mWHO II	mWHO II–III	mWHO III	mWHO IV
Diagnosis (if otherwise well and uncomplicated)	Small or mild Pulmonary stenosis Patent ductus arteriosus Mitral valve prolapse Successfully repaired simple lesions (atrial or ventricular septal defect, patent ductus arteriosus, anomalous pulmonary venous drainage) Atrial or ventricular ectopic beats, isolated	Unoperated atrial or ventricular septal defect Repaired tetralogy of Fallot Most arrythmias (supraventricular arrhythmias) Turner syndrome without aortic dilatation	Mild left ventricular impairment (EF >45%) Hypertrophic cardiomyopathy Native or tissue valve disease not considered WHO I or IV (mild mitral stenosis, moderate aortic stenosis) Marfan or other HTAD syndrome without aortic dilatation Aorta <45 mm in bicuspid aortic valve pathology Repaired coarctation Atrioventricular septal defect	Moderate left ventricular impairment (EF 30–45%) Previous peripartum cardiomyopathy without any residual left ventricular impairment Mechanical valve Systemic right ventricle with good or mildly decreased ventricular function Fontan circulation Fontan circulation with good clinical course and without associated comorbidities Unrepaired cyanotic heart disease Other complex heart disease Moderate mitral stenosis Severe asymptomatic aortic stenosis Moderate aortic dilatation (40–45 mm in Marfan syndrome or other HTAD; 45–50 mm in bicuspid aortic valve, Turner syndrome ASI 20–25 mm/m^2, tetralogy of Fallot <50 mm) Ventricular tachycardia	Pulmonary arterial hypertension Severe systemic ventricular dysfunction (EF <30% or NYHA class III–IV) Previous peripartum cardiomyopathy with any residual left ventricular impairment Severe mitral stenosis Severe symptomatic aortic stenosis Systemic right ventricle with moderate or severely decreased ventricular function Severe aortic dilatation (>45 mm in Marfan syndrome or other HTAD, >50 mm in bicuspid aortic valve, Turner syndrome ASI >25 mm/m^2, tetralogy of Fallot >50 mm) Vascular Ehlers-Danlos Severe (re)coarctation Fontan with any complication
Risk	No detectable increased risk of maternal mortality and no/mild increased risk in morbidity	Small increased risk of maternal mortality or moderate increase in morbidity	Intermediate increased risk of maternal mortality or moderate to severe increase in morbidity	Significantly increased risk of maternal mortality or severe morbidity	Extremely high risk of maternal mortality or severe morbidity

Abbreviations: ASI, aortic size index; EF, ejection fraction; HTAD, heritable thoracic aortic disease.

analyzed separately before formulating a comprehensive diagnosis. The principal segments are the atria, the ventricles, and the great arteries, which are joined together by the atrioventricular canal and the conus (infundibulum). In the normal heart, the right ventricle (RV) is right-sided and organized inflow-to-outflow from right to left, while the left ventricle (LV) is left-sided and organized inflow-to-outflow from left to right. It is important to determine the segmental alignments, that is, what drains into what. For example, in the normal heart, the right atrium (RA) is aligned with the RV and the LV with the aorta. Finally, the segmental connections, the way in which adjacent segments are physically linked to each other, are described. For example, in the normal heart, the pulmonary artery (PA) is connected to the RV by a complete muscular conus (infundibulum), while the aorta is connected to the LV by aortic-mitral fibrous continuity (without a complete conus). Alignment and connection are different concepts, and both are important, especially in complex defects.

Cardiac Development The heart starts to form in the third week of gestation and is nearly fully formed by 8 weeks' gestation. Mesodermal precardiac cells migrate to form the cardiac crescents (primary heart fields) in anterior lateral plate mesoderm, which are then brought together to form a primary linear heart tube by ventral closure of the embryo. Cells of the second heart field continue to proliferate outside the heart and are added to the heart tube over the course of embryogenesis, contributing to the atria, the RV, and outflow tract. Additionally, cardiac neural crest cells migrate into the developing heart in the 5th–6th weeks and are essential for septation of the outflow, formation of the semilunar valves, and patterning of the aortic arches. Once formed, the heart tube grows and elongates by addition of cells from the second heart field. The ends of the heart tube are relatively fixed by the pericardial sac so that as it elongates it must loop (bend), and in the vast majority of hearts, the loop falls to the right (D-loop). Further elongation pushes the mid-portion of the tube (future ventricles) inferior or caudal to the inflow, resulting in the normal relationship between the atria and ventricles. Further growth pushes the outflow medially and is associated with outflow rotation, both processes essential for normal alignment of the outflow. Finally, the proximal part of the outflow is incorporated in the RV, shortening the outflow in association with further rotation. While this remodeling is occurring, the outflow is undergoing septation under the influence of cardiac neural crest cells. Septation proceeds from distal to proximal, culminating in formation and muscularization of the infundibular, or muscular, outflow septum, which inserts onto the superior endocardial cushion at the rightward rim of the outflow foramen, walling the aorta into the LV via the outflow foramen and the PA directly into the RV.

Genetic Considerations CHD is the most commonly occurring birth defect; etiologic contributors are increasingly recognized, although often speculated to be multifactorial. Children born with trisomy 21 have a 50% chance of having CHD, most commonly defects in the atrioventricular canal. Conotruncal defects are associated with several chromosomal abnormalities, most notably a deletion at chromosome 22q11 (DiGeorge syndrome). Echocardiographic clues to this association in patients with a conotruncal defect include an associated right aortic arch or aberrant subclavian artery. Many adults currently living with conotruncal defects may not have undergone testing for DiGeorge syndrome. This condition is important to recognize because a variety of psychiatric disorders and disabilities in cognitive function may be present and go untreated. Patients with Noonan syndrome commonly have a dysplastic pulmonary valve and have facial and lymphatic abnormalities. Several defects in specific genes have been associated with Noonan syndrome, most notably *PTPN11*. Adults with Williams syndrome (7q11.23 deletion) commonly have supravalvar aortic stenosis and diffuse arteriopathy, with a "cocktail-like" personality and hypercalcemia. There is a growing importance of genome-wide analyses in subjects with CHD.

TABLE 280-2 Congenital Etiologies of Right Heart Dilation

- Congenital tricuspid valve disease
  - Tricuspid valve dysplasia with regurgitation
  - Ebstein anomaly
- Congenital pulmonary valve regurgitation
- Pulmonary arterial hypertension
- Myocardial abnormalities
  - Arrhythmogenic RV cardiomyopathy
  - Uhl's anomaly
- Shunt lesions
  - Partial anomalous pulmonary venous return
  - Primum ASD
  - Secundum ASD
  - Sinus venosus defect
  - Coronary sinus septal defect
  - Gerbode defect (LV-RA shunt)
  - Coronary artery fistula to the RA, CS
  - Postoperative residual shunts

Abbreviations: ASD, atrial septal defect; CS, coronary sinum; LV, left ventricle; RA, right atrium; RV, right ventricle.

SPECIFIC CHD LESIONS

Dilated Right Heart There are many congenital etiologies for right heart dilation (Table 280-2). These include congenital valvular anomalies (such as Ebstein anomaly or pulmonary regurgitation), intrinsic RV myocardial anomalies (arrhythmogenic RV dysplasia, Uhl's anomaly), or shunt lesions occurring proximal to the tricuspid valve (atrial septal defects or partial anomalous pulmonary veins). Cardiac imaging is critical in determining the etiology of right heart dilation, and knowledge of the anatomy and physiology of various shunt lesions is essential.

Atrial Septal Defect One of the most common etiologies of right heart dilation is presence of an atrial septal defect (ASD; Fig. 280-3*A*). Intracardiac communications allow blood transmission between chambers or spaces based on relative resistance, propulsion, and flow patterns. Patients with large ASDs often present in childhood; however, many ASDs are not discovered until adult life. The physiology of an ASD is predominantly that of a "left-to-right" shunt (flow of pulmonary venous, or oxygenated, blood toward systemic venous, or deoxygenated, chambers or vessels). The degree of left-to-right shunting determines the amount of right heart volume loading and is dictated by the size of the defect as well as the diastolic properties of the heart. As patients age, several factors, such as diabetes mellitus, systemic hypertension, and atherosclerosis, may contribute to decreased compliance of the left-sided cardiac chambers and contribute to increased left-to-right shunting and symptomatology. The classic physical examination finding is a wide, fixed splitting of the second heart sound, which is due to prolonged RV ejection and increased PA capacitance, which, in turn, delay pulmonary valve closure. The surface electrocardiogram (ECG) commonly displays an incomplete right bundle branch block. Symptoms, when they occur, most commonly include exercise intolerance, arrhythmia, and dyspnea with exertion. It is not uncommon for adults to have incidentally noted asymptomatic ASD during evaluation of other comorbid issues. Right heart dilation, without additional etiology for such, in the setting of unrepaired ASD is considered a risk for progression toward symptomatic right heart failure, atrial arrhythmias, and potential development of pulmonary arterial hypertension (if such is not already present). Therefore, a patient with an ASD and right heart dilation, particularly with symptoms attributable to such, should be offered ASD closure. Pulmonary vascular disease leading to pulmonary hypertension develops in up to 10% of patients with unrepaired ASD, and Eisenmenger syndrome (ES) is a rare complication (see below). Management of patients with concomitant ASD and pulmonary hypertension should be coordinated with both ACHD and pulmonary hypertension experts.

Figure 280-3*B* illustrates the locations of various ASDs. The most common type of an ASD is a secundum ASD, which is a defect, or true deficiency in the atrial septum, in the region of the fossa ovalis. This should be differentiated from a patent foramen ovale (PFO), which is persistence of patency of the flap valve of the fossa ovalis (not associated with right-sided cardiac dilation) and persists in up to 25% of adults. Secundum ASDs can often be closed with occluder devices placed percutaneously. However, certain anatomic determinants make percutaneous closure less favorable, including large defects, inadequate tissue rims surrounding the defect, and concomitance of anomalous draining pulmonary veins. A primum ASD is a deficiency of the atrioventricular (AV) canal portion of the atrial septum; primum ASD is always associated with abnormal development of the AV valves,

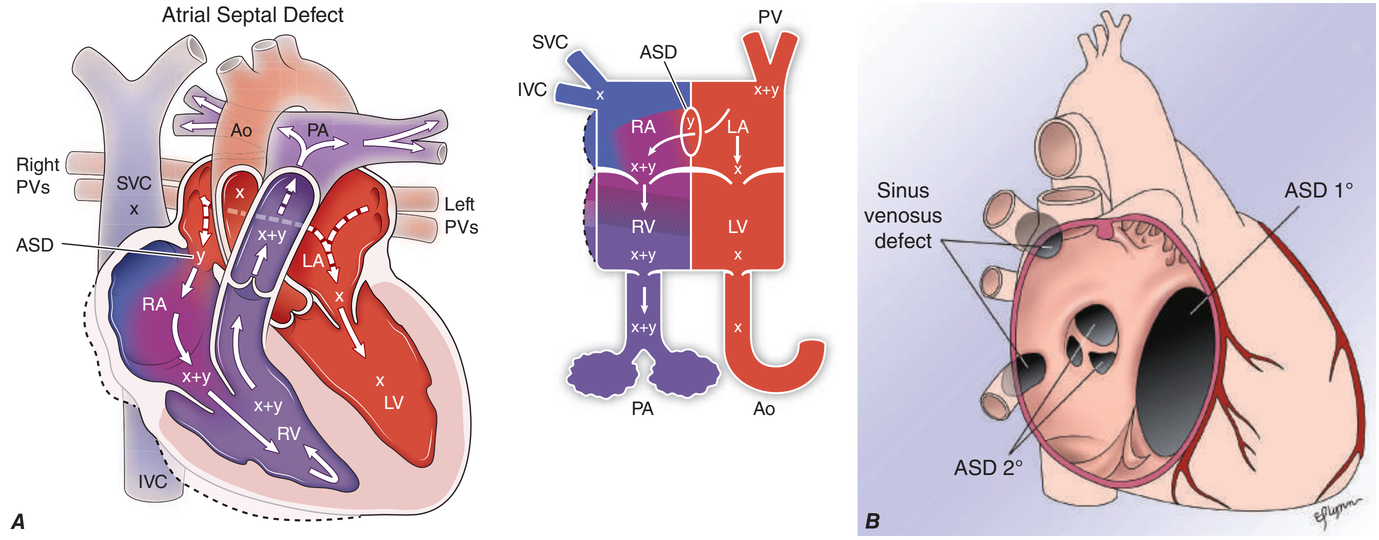

FIGURE 280-3 ***A.*** Atrial septal defect. In the presence of an atrial septal defect, the difference in compliance between the (RA + RV) as compared to the (LA + LV), combined with the size of the defect itself, allows for a "shunt" of flow ("y") of "red" (oxygenated) blood from the left side of the heart to the right side (deoxygenated). Systemic venous return of pure deoxygenated blood ("x") is increased by the oxygenated shunted blood ("y") to increase volume of blood ("x + y") in the RA, RV, and total blood flow to the lungs. If the volume or the sequelae of the shunted blood are sufficient, RA and RV can dilate (*hashed lines*), and arrhythmias or shortness of breath (and occasionally pulmonary hypertension) can ensue. Ao, aorta; ASD, atrial septal defect; IVC, inferior vena cava; LA, left atrium; LV, left ventricle; PA, pulmonary artery; PV, pulmonary veins; RA, right atrium; RV, right ventricle; SVC, superior vena cava. ***B.*** Diagrammatic representation of the location of various atrial septal defects. ASD 1, primum atrial septal defect; ASD 2, secundum atrial septal defect. *(Part B used with permission from Emily Flynn McIntosh, illustrator.)*

most commonly resulting in a cleft in the mitral valve. A coronary sinus defect is rare and involves an opening between the coronary sinus and the left atrium. A sinus venosus defect is not a defect in the atrial septum but, rather, a defect between either the right superior vena caval–atrial junction and the right upper pulmonary vein(s) or, less commonly, the inferior vena caval–atrial junction and the right lower pulmonary veins. Surgical closure is required for primum ASDs, sinus venosus defects, and coronary sinus septal defects.

Partial Anomalous Pulmonary Venous Return Partial anomalous pulmonary venous return (PAPVR) is occasionally discovered in adults with right heart dilation or incidentally on cross-sectional imaging (Fig. 280-4). There are several possible anomalous connections, with the most common being a left upper pulmonary vein to an ascending vertical vein into the innominate vein or the right upper pulmonary vein draining to the superior vena cava. In the latter case, careful attention should be paid to ensure that there is not an associated sinus venosus defect. Concomitant pulmonary hypertension can occur but is uncommon. Symptomatology may be absent, and a decision to repair isolated PAPVR should include variance in anatomy, lung ventilation and perfusion, hemodynamic response to shunt, symptoms, and surgical experience.

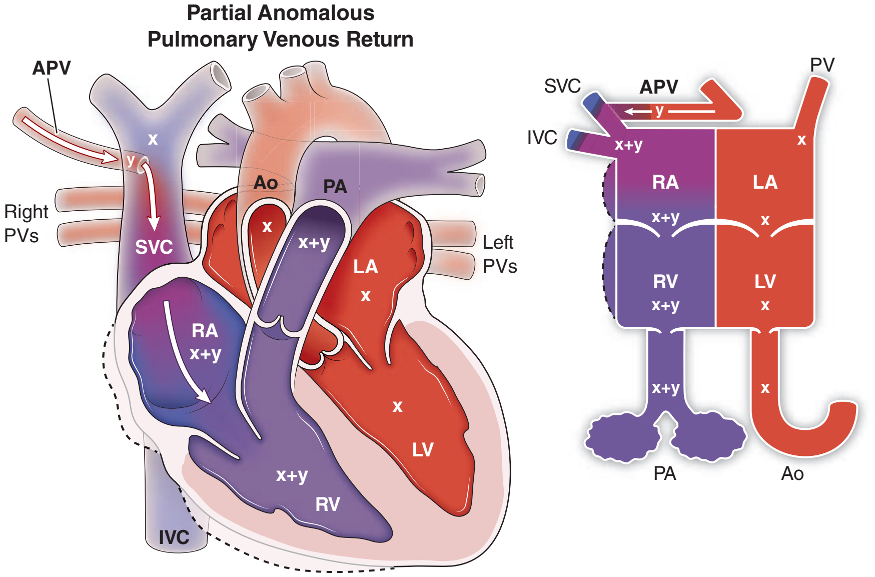

FIGURE 280-4 Partial anomalous pulmonary venous return. In the presence of an anomalously draining pulmonary vein (typically to a systemic vein such as the left innominate vein, SVC, or rarely IVC), an obligate "shunt" of flow ("y") of "red" (oxygenated) blood from the affected pulmonary vein to the right heart (deoxygenated) ensues. Systemic venous return of pure deoxygenated blood ("x") is increased by the oxygenated shunted blood ("y") to increase volume of blood ("x + y") in the SVC, RA, RV, and total blood flow to the lungs. If the volume or the sequelae of the shunted blood are sufficient, RA and RV can dilate (*hashed lines*), or shortness of breath can ensue. Ao, aorta; APV, anomalous pulmonary vein; IVC, inferior vena cava; LA, left atrium; LV, left ventricle; PA, pulmonary arteries; PV, pulmonary veins; RA, right atrium; RV, right ventricle; SVC, superior vena cava.

Ebstein Anomaly Ebstein anomaly (Fig. 280-5) is the result of embryologic failure of delamination, or "peeling away," of the tricuspid valve leaflets from the ventricular myocardium, resulting in adherence of the valve leaflets to the underlying myocardium. This results in a wide variety of abnormalities, including apical and posterior displacement of the dilated tricuspid valve annulus, dilation of the "atrialized" portion of the RV, and fenestrations, redundancy, and tethering typically of the anterior leaflet of the tricuspid valve. The malformed tricuspid valve is usually regurgitant but may occasionally be stenotic. The clinical presentation of Ebstein anomaly in the adult depends on several factors, including the extent of tricuspid valve leaflet distortion, degree of tricuspid regurgitation (TR), right atrial pressure, and presence of an atrial level shunt. The physical examination of a patient with Ebstein anomaly may vary depending on the severity of disease. In more severe cases, the first heart sound may be split and the second component of the first heart sound may have a distinctive snapping quality (known as the sail sign, due to the redundancy of the anterior tricuspid valve leaflet). Patients with significant TR may have prominent "v" waves of the jugular venous pulsations; however, this finding is often absent due to abnormal right atrial compliance. The ECG is often abnormal, with right atrial and ventricular enlargement. Up to 20% of patients have evidence of ventricular preexcitation (Wolff-Parkinson-White pattern). Surgical treatment includes a tricuspid valve repair or replacement, closure of any atrial level defects, and arrhythmia ablative procedures.

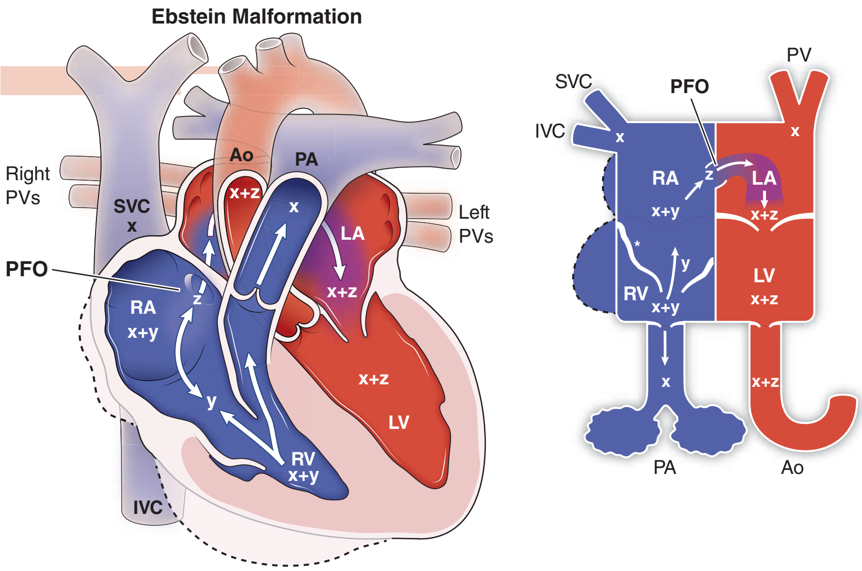

FIGURE 280-5 Ebstein malformation. In the presence of Ebstein anomaly, the tricuspid valve leaflets can be redundant, fenestrated, and sail-like (typically seen in the anterior leaflet*) or adherent to the underlying myocardium with apical displacement of the nonadherent components (typically the septal and posterior leaflets). Location and degree of leaflet coaptation are variable and account for varying degrees of tricuspid regurgitation, shift of the functional tricuspid valve anterior from the anatomic annulus into the right ventricle, "atrialization" of the right ventricle, and most commonly angulation of the tricuspid valve into the RV outflow tract. RA and RV dilation (*hashed lines*) can occur due to the effects of combined volume from systemic venous return ("x") and tricuspid regurgitant flow ("y"). PFO is frequent; worsening compliance and elevation of pressure in the RA as compared to the LA can lead to increasing "right-to-left" (deoxygenated to oxygenated) shunt and cyanosis. RV myocardial function may be abnormal. Ao, aorta; IVC, inferior vena cava; LA, left atrium; LV, left ventricle; PA, pulmonary arteries; PFO, patent foramen ovale; PV, pulmonary veins; RA, right atrium; RV, right ventricle; SVC, superior vena cava; *, anterior tricuspid valve leaflet.

Shunt Lesions Causing Left Heart Dilation Intracardiac shunts or intravascular passages that occur below the level of the tricuspid valve result in left heart dilation. The two major types of congenital shunts that result in left heart dilation are a ventricular septal defect (VSD; Fig. 280-6A) and patent ductus arteriosus (PDA; Fig. 280-7).

Ventricular Septal Defects VSDs are the most common congenital anomaly recognized at

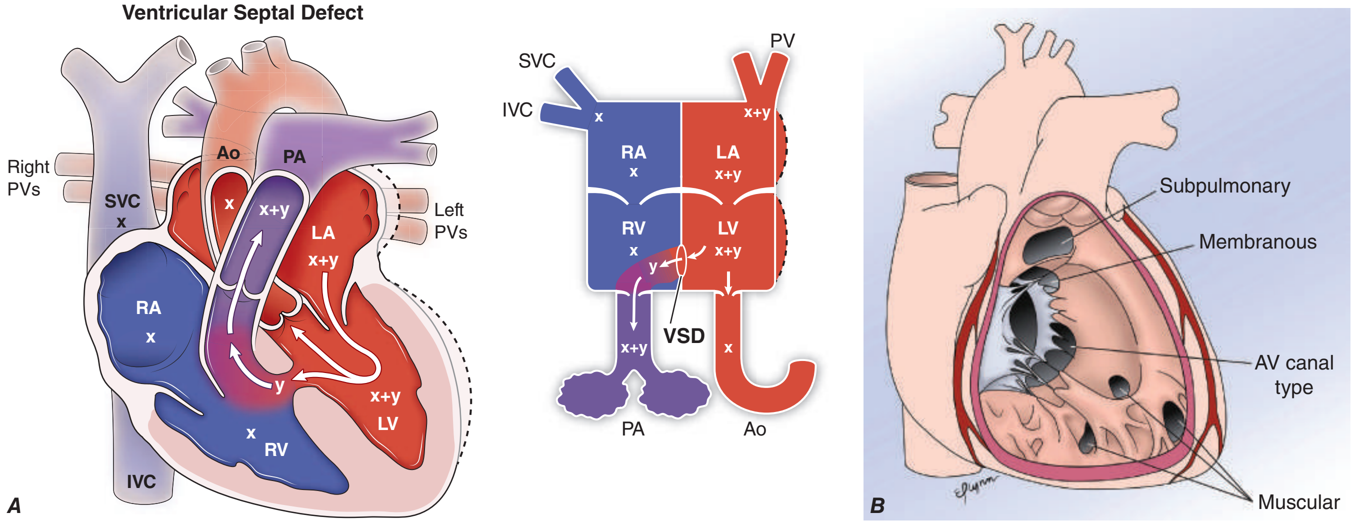

FIGURE 280-6 ***A.*** Ventricular septal defect. In the presence of a ventricular septal defect, the difference in pressure and outflow resistance in systole (and the difference in compliance in diastole) between the RV and LV, combined with the size of the defect itself, allow for a "shunt" of flow ("y") of "red" (oxygenated) blood from the left side of the heart to the right side (deoxygenated). Systemic venous return of pure deoxygenated blood ("x") is increased by the oxygenated shunted blood ("y") to increase volume of blood ("x + y") through the outflow of the RV into the lungs, and in the left atrium and left ventricle. If the volume or the sequelae of the shunted blood are sufficient, LA and LV can dilate (*hashed lines*), and arrhythmias or shortness of breath (and occasionally pulmonary hypertension) can ensue. Ao, aorta; IVC, inferior vena cava; LA, left atrium; LV, left ventricle; PA, pulmonary arteries; PV, pulmonary veins; RA, right atrium; RV, right ventricle; SVC, superior vena cava; VSD, ventricular septal defect. ***B.*** Diagrammatic representation of the location of various ventricular septal defects. AV, atrioventricular. *(Part B used with permission from Emily Flynn McIntosh, illustrator.)*

birth; however, they account for only ~10% of CHD in the adult, due to the high rate of spontaneous closure of small VSDs during the early years of life. Large VSDs usually cause symptoms of heart failure and poor somatic growth and are most often surgically closed before adulthood. Several classification systems for VSDs exist. Figure 280-6*B* illustrates various locations of VSDs; the most common location is in the membranous septum (also referred to as perimembranous or outlet defects). Muscular defects that persist into adult life are often pressure and flow restricted, resulting in no significant hemodynamic consequence. AV canal defects, also referred to as inlet defects, are located in the crux of the heart and are associated with abnormalities of the AV valve leaflets. Subpulmonary defects, also known as *conal septal defects*, are commonly associated with prolapse of the right coronary cusp and aortic insufficiency. The outcome for adults with small VSDs without evidence of ventricular dilation or pulmonary hypertension is generally excellent.

Patent Ductus Arteriosus A PDA courses between the aortic isthmus and the origin of one of the branch PAs. Small PDAs are often silent to auscultation and do not cause hemodynamic changes. The classic murmur is heard best just below the left clavicle and typically extends from systole past the second heart sound into diastole, reflecting flow turbulence and gradient between the aorta and the PAs (resulting in left-to-right shunting). Large PDAs will lead to left heart dilation and may lead to chronically elevated pulmonary vascular resistance, including the potential for ES.

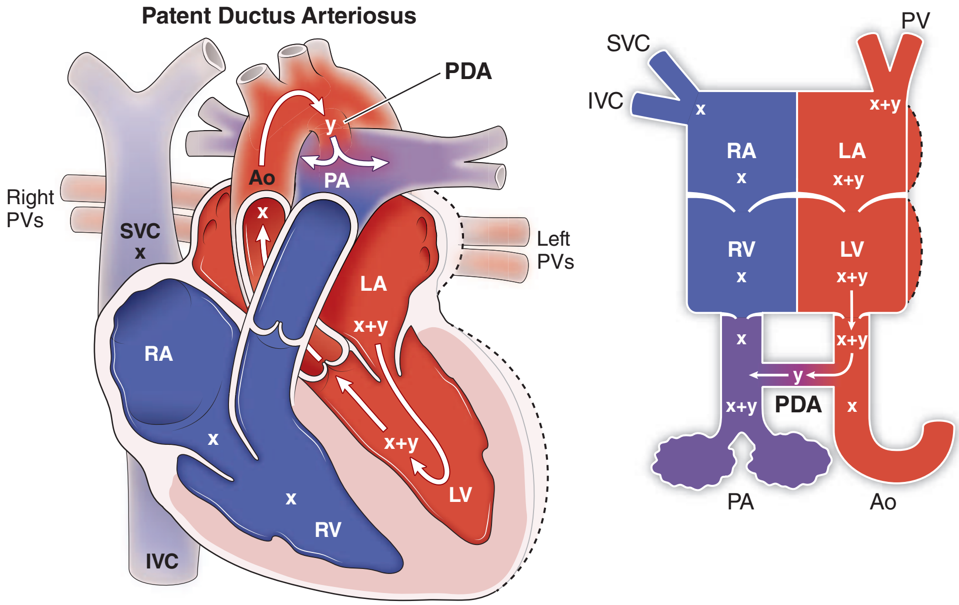

FIGURE 280-7 Patent ductus arteriosus. In the presence of a patent ductus arteriosus, the difference in pressure and resistance in both systole and diastole between the pulmonary arteries and the aorta, combined with the size of the ductus itself, allow for a "shunt" of flow ("y") of "red" (oxygenated) blood from the aorta to the pulmonary arteries (deoxygenated). Systemic venous return of pure deoxygenated blood ("x") is increased by the oxygenated shunted blood ("y") to increase volume of blood ("x + y") in the lungs, the left atrium, the left ventricle, and out the aortic valve. If the volume or the sequelae of the shunted blood are sufficient, LA and LV can dilate (*hashed lines*), and arrhythmias or shortness of breath (and occasionally pulmonary hypertension) can ensue. Ao, aorta; IVC, inferior vena cava; LA, left atrium; LV, left ventricle; PA, pulmonary arteries; PDA, patent ductus arteriosus; PV, pulmonary veins; RA, right atrium; RV, right ventricle; SVC, superior vena cava.

MODERATE AND COMPLEX CHD

Tetralogy of Fallot Tetralogy of Fallot (TOF) is the most common form of cyanotic CHD, occurring in 0.5 per 1000 live births. It involves anterior deviation of the conal septum, resulting in RV outflow tract (RVOT) obstruction, a VSD, RV hypertrophy, and an overriding aorta (**Fig. 280-8*A*, *B***). There is a large spectrum of severity of disease in TOF, from patients who have only mild pulmonary stenosis to those with complete pulmonary atresia (TOF/PA). Current surgical strategies involve primary repair in infancy (**Fig. 280-8*C***); however, many adults may have first undergone palliative procedures (Blalock-Taussig, Potts, Waterston shunts) prior to a complete repair. The goal of surgical repair is to alleviate the pulmonary stenosis and close the VSD. Up to 10% of patients with TOF have an anomalous coronary artery, most commonly, an anomalous left anterior descending coronary artery from the right coronary cusp. Patients with an anomalous coronary as well as those with TOF/PA may require an RV-to-PA conduit.

Adults with repaired TOF often have hemodynamic sequelae that may require reintervention in adulthood (**Table 280-3**). Pulmonary

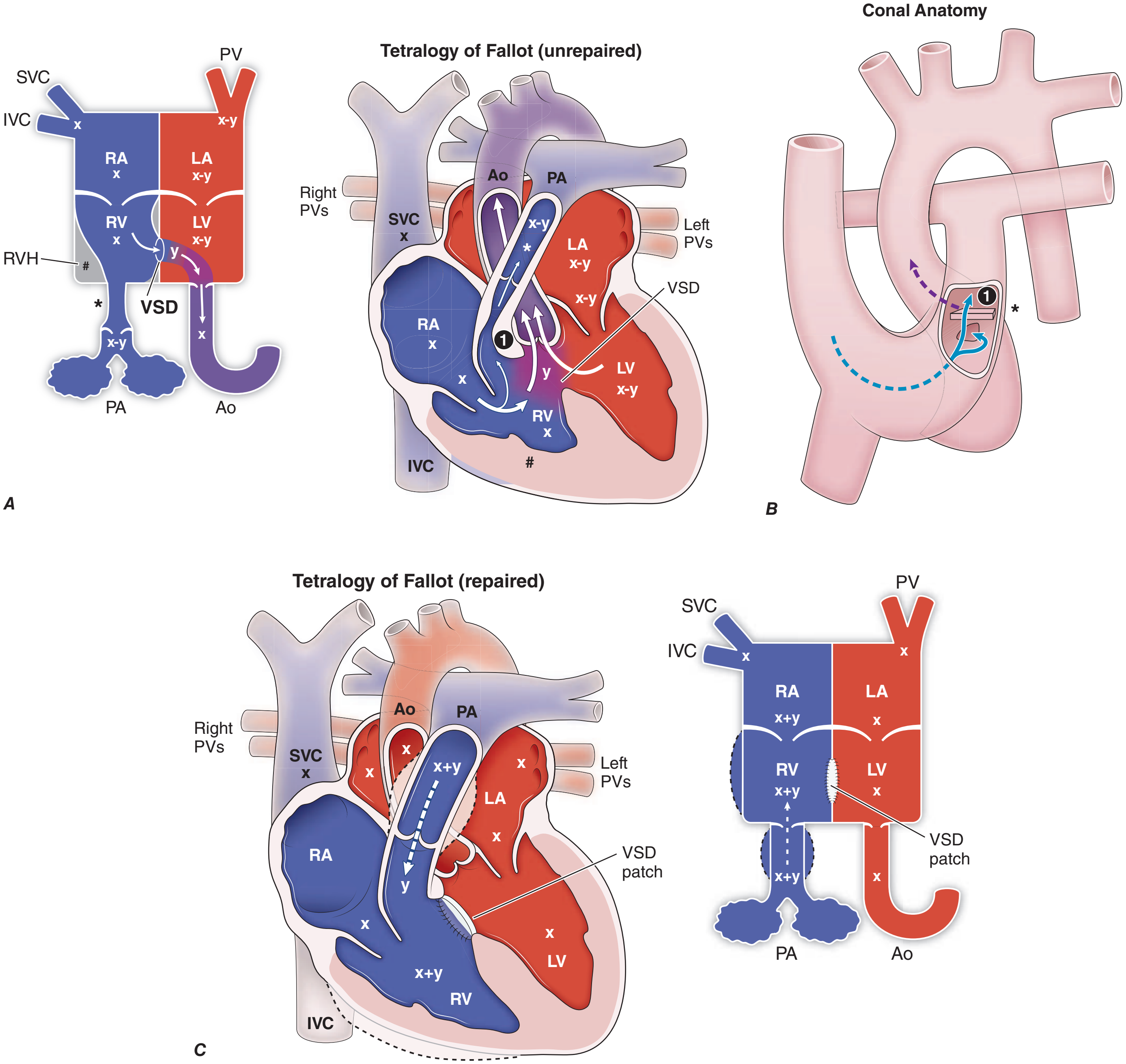

FIGURE 280-8 ***A.*** Tetralogy of Fallot involves anterior and superior malalignment of a bar of tissue (conal septum) (see *in part ***B***, which presents a cut-away view through the anterior surface of the RV, into the RV outflow), partially obstructing the right ventricular outflow (under the pulmonary valve, i.e., "subpulmonary stenosis"; labeled as 1), and leaving a gap in the interventricular septum (VSD). The pulmonary valve annulus is typically hypoplastic. Outflow obstruction prevents regression of right ventricular hypertrophy ($^{\#}$), which was present in utero. The difference in pressure and outflow resistance in systole (and the difference in compliance in diastole) between the obstructed RV and the LV allows for a "shunt" of flow ("y") of "blue" (deoxygenated) blood from the right side of the heart to the left side (oxygenated). Systemic venous return of pure deoxygenated blood ("x") is decreased by the shunted blood ("y"), leading to a total decrease in the volume of blood ("x – y") passing beyond into the lungs. The deoxygenated shunted blood ("y") mixes with fully oxygenated blood in the LV, contributing to systemic arterial cyanosis. ***C.*** Tetralogy of Fallot—repaired. After modern repair of tetralogy of Fallot, VSD has been patched closed, and outflow tract obstruction has been surgically removed, frequently at the expense of a patch enlarging the pulmonary valve annulus at the expense of sacrificing the integrity of the pulmonary valve (causing pulmonary regurgitation). The pulmonary regurgitant volume ("y") is added to systemic venous return ("x"), contributing to RV chamber enlargement (*hashed lines*) and may be associated with tricuspid annular dilation and valve regurgitation, resulting in RA enlargement. Ao, aorta; IVC, inferior vena cava; LA, left atrium; LV, left ventricle; PA, pulmonary arteries; PV, pulmonary veins; RA, right atrium; RV, right ventricle; RVH, right ventricular hypertrophy; SVC, superior vena cava; VSD, ventricular septal defect.

TABLE 280-3 Potential Sequelae of Repaired Tetralogy of Fallot

Right atrial dilation
Right ventricular dilation
Right ventricular dysfunction
Right ventricular outflow tract obstruction
Pulmonary regurgitation
Branch pulmonary artery stenosis
Tricuspid regurgitation
Residual ventricular septal defect
Left ventricular dysfunction
Aortic root dilation
Atrial arrhythmias
Ventricular arrhythmias
Sudden cardiac death

regurgitation is common following TOF repair and is usually associated with RV dilation. Accurate quantification of RV size, function, and mass is particularly important in adults after repair of TOF, as RV dilation, dysfunction, and hypertrophy are associated with adverse outcomes in these patients. Patients may also have residual RVOT obstruction, which may occur beneath the pulmonary valve, at the valve level, above the valve, or in the branch PAs. Cardiac magnetic resonance imaging is routinely used in the surveillance of these patients. Left ventricular dysfunction is present in at least 20% of adults with repaired TOF, particularly those who were repaired later in life, had prior palliative shunts, or have concomitant RV dysfunction.

As patients age with repaired TOF, both atrial and ventricular arrhythmias occur with increasing frequency. A QRS duration on a resting ECG of 180 ms or more has been associated with increased risk of ventricular tachycardia and sudden death in this patient population. In one prospective follow-up study of 144 adults with repaired TOF, there was a 72% survival at 40 years, but only a 25% cumulative event-free survival. These events include need for reintervention (most commonly pulmonary valve replacement [PVR]), symptomatic arrhythmias, and heart failure.

The most common reintervention in a repaired TOF patient is a PVR. However, optimal timing of PVR in asymptomatic patients with repaired TOF remains unclear. Traditionally, PVR has been accomplished with a surgical procedure; however, percutaneous implantation of pulmonary valves is becoming increasingly utilized in clinical practice.

Patients with repaired TOF may also undergo interventions including closure of residual VSDs, dilation and/or stenting of the RVOT or branch PAs, and tricuspid valve repair. Patients with clinically significant arrhythmias may benefit from catheter ablation.

Transposition of the Great Arteries Transposition of the great arteries (TGA) is defined by the great arteries arising from the opposite side of the ventricular septum than normal; as such, the aorta arises from the RV and the PA from the LV. The more common form of TGA, known as *D-loop TGA*, involves AV concordance and ventricular-arterial discordance, resulting in a physiology that allows two circuits to be in parallel rather than in series **(Fig. 280-9*A*)** and intense cyanosis shortly after birth. This physiology is not compatible with long-term survival without surgical intervention. Patients with TGA may be born with additional congenital defects (most commonly a VSD).

The surgical repairs for D-loop TGA have evolved over time. In the late 1950s through the 1970s, the atrial switch procedure (Mustard, Senning procedures) was performed **(Fig. 280-9*B*)**. These atrial switch

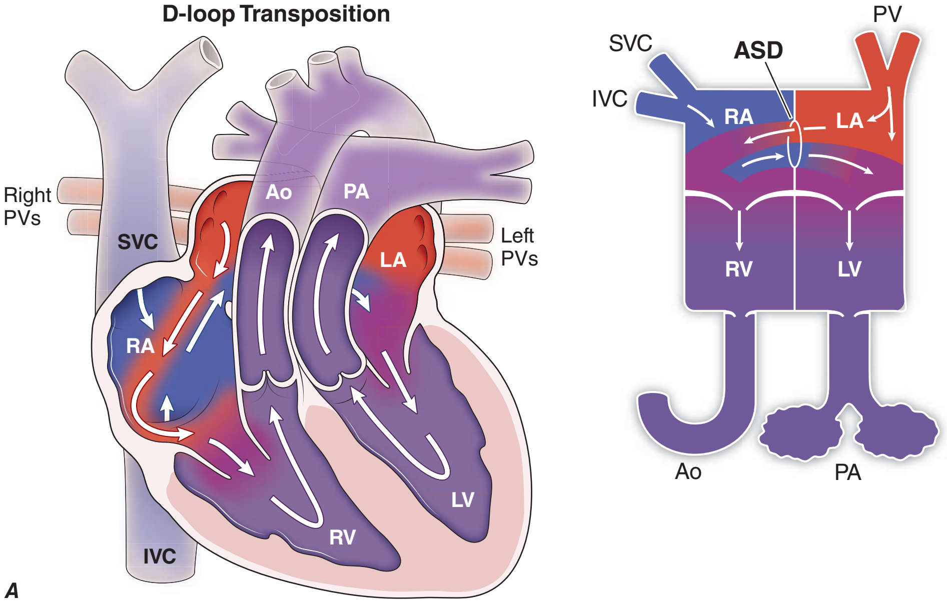

FIGURE 280-9 ***A.*** Transposition of the great arteries. When the great arteries are transposed, the aorta arises from the RV, and the pulmonary artery arises from the LV, leaving deoxygenated blood circulating from systemic veins to systemic arteries in separated fashion from oxygenated blood, which circulates from pulmonary veins to pulmonary arteries. Without interchamber or intravascular communications, this circulation is incompatible with life. Presence of an atrial septal defect (ASD), depicted here, ventricular septal defect (VSD), or patent ductus arteriosus (PDA) allows for some interchamber or intravascular mixing and, at best, partial relief of cyanosis and sustenance of life, at the expense of increased pulmonary blood flow. ***B.*** Atrial switch. Atrial level switch procedures (Mustard and Senning) were the first standardized surgeries to alter the natural course of complex congenital heart disease, utilizing intracardiac rerouting via a "baffle" to redirect blood flow. The atrial switch simulates inverted trousers, with each "pants leg" (*) attaching to either the SVC or the IVC, transporting deoxygenated blood through the interior of the trousers to the "waist of the trousers" and directing blood through the mitral valve to the LV and out the PA. Surgical removal of the atrial septum allows pulmonary venous return to traverse from posterior left atrium through the space between the pants legs of the baffle, through the tricuspid valve to the RV (serving as the "systemic ventricle," i.e., that pumps to the systemic arterial circulation), through the aorta. Non-infrequent sequelae include sinus node dysfunction, atrial arrhythmias, systolic dysfunction of the RV, tricuspid regurgitation (from RV to LA), leaks in the baffle material allowing shunting of blood, and obstruction of the systemic or pulmonary venous baffles. ***C.*** Arterial switch. The arterial switch operation allowed both anatomic and physiologic correction for D-loop transposition of the great arteries. Successful surgical switching of the PA and the Ao above the level of the native roots (*hashed lines*) necessitated ability to transfer coronary artery origins contained within a button of tissue (*) back to the neo-aorta (now supported by the LV). Deoxygenated blood flow from SVC and IVC passes from RA to RV to PA, and oxygenated blood passes from PV to LA to LV to Ao. Uncommon sequelae include obstruction at any of the surgical sites (supravalvar PA or Ao stenosis, coronary orifice obstruction) or more distal obstructions due to tension placed on the PA, Ao, or coronary arteries. Ao, aorta; IVC, inferior vena cava; LA, left atrium; LV, left ventricle; PA, pulmonary arteries; PV, pulmonary veins; RA, right atrium; RV, right ventricle; SVC, superior vena cava.

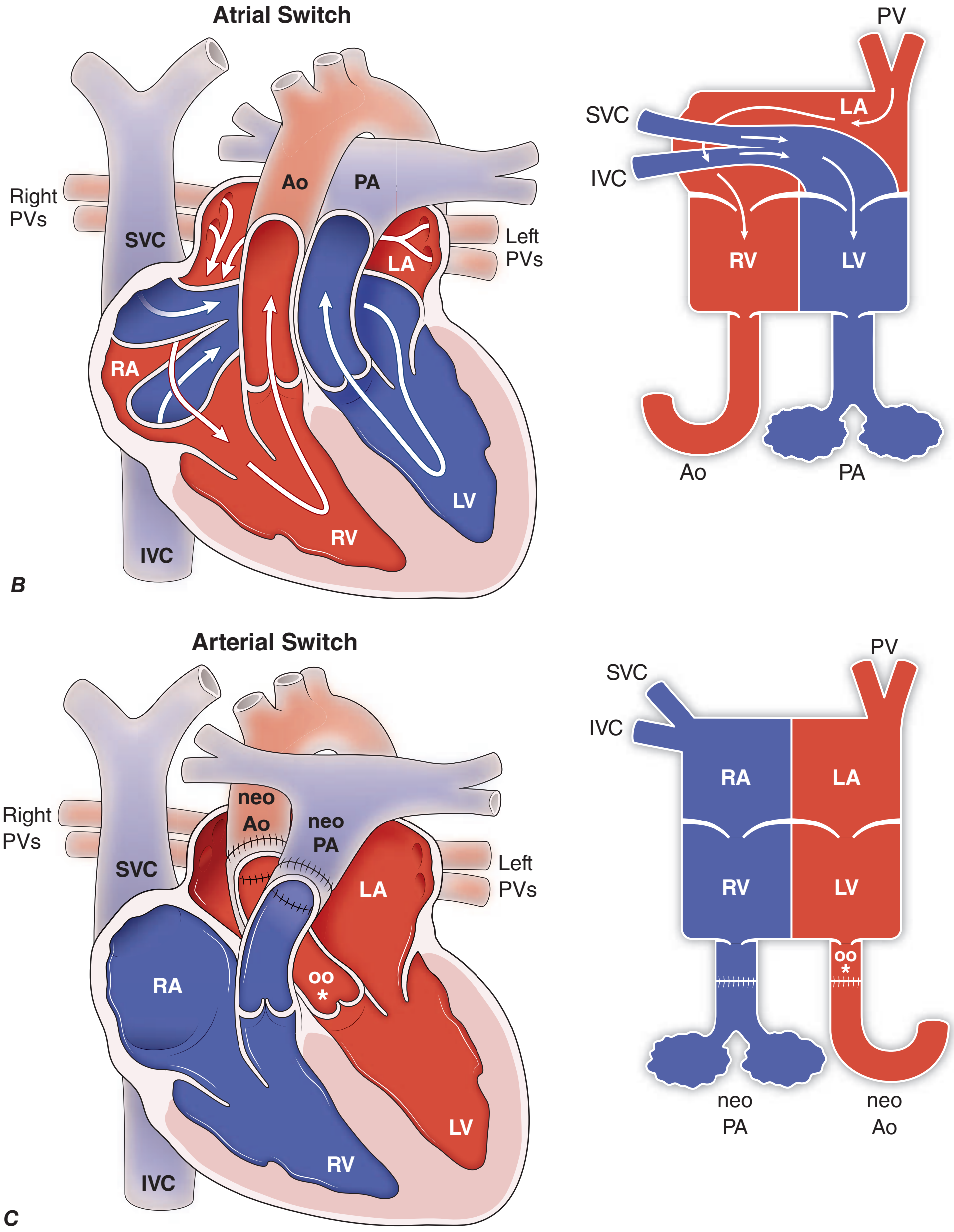

FIGURE 280-9 (Continued)

procedures relieved the cyanosis but left the patient with a systemic RV. Despite moderate-term survival over decades, there are multiple long-term sequelae that may present following the atrial switch procedure. The most worrisome complication is that of systemic RV dysfunction. The prevalence of RV dysfunction in this population is not well defined. Limited study has failed to reveal medical therapies effective for systemic RV dysfunction.

A subset of patients with D-loop TGA, VSD, and PS may have undergone a Rastelli procedure. This intervention involves placing an RV-to-PA conduit and routing the LV to the aorta through the VSD, which results in relief of cyanosis and the benefit of a systemic LV.

In the 1980s, the arterial switch operation (ASO; Fig. 280-9C) became the surgical procedure of choice for D-loop TGA. This procedure involves transecting the great arteries above the sinuses and placing the PAs anteriorly to come into alignment with the RV, resulting in draping of the branch PAs over the ascending aorta. A coronary artery translocation is performed. The ASO has resulted in substantial long-term survival.

The potential long-term sequelae of the various surgical procedures for D-loop TGA are listed in Table 280-4.

The less common form of TGA, known as *L-loop TGA* (physiologically corrected or "congenitally corrected" TGA; Fig. 280-10), may not require surgical intervention but is presented here in relation to other forms of TGA. L-loop TGA involves both AV discordance (RA allowing passage of deoxygenated systemic venous return to the LV, and

TABLE 280-4 Long-Term Sequelae of D-Loop TGA Surgery

ATRIAL SWITCH	ARTERIAL SWITCH	RASTELLI PROCEDURE
Systemic venous baffle	Arterial anastomosis stenosis	Subaortic stenosis
Pulmonary venous baffle	Branch PA stenosis	RV-PA conduit obstruction
RV (systemic) dysfunction	Neo-aortic root dilation	Pulmonary regurgitation
Tricuspid regurgitation	Neo-aortic regurgitation	Ventricular dysfunction
Baffle leaks	Coronary artery stenosis	
LVOT obstruction (PS)	LV dysfunction	

Abbreviations: LV, left ventricle; LVOT, left ventricular outflow tract; PA, pulmonary artery; PS, pulmonary stenosis; RV, right ventricle; TGA, transposition of the great arteries.

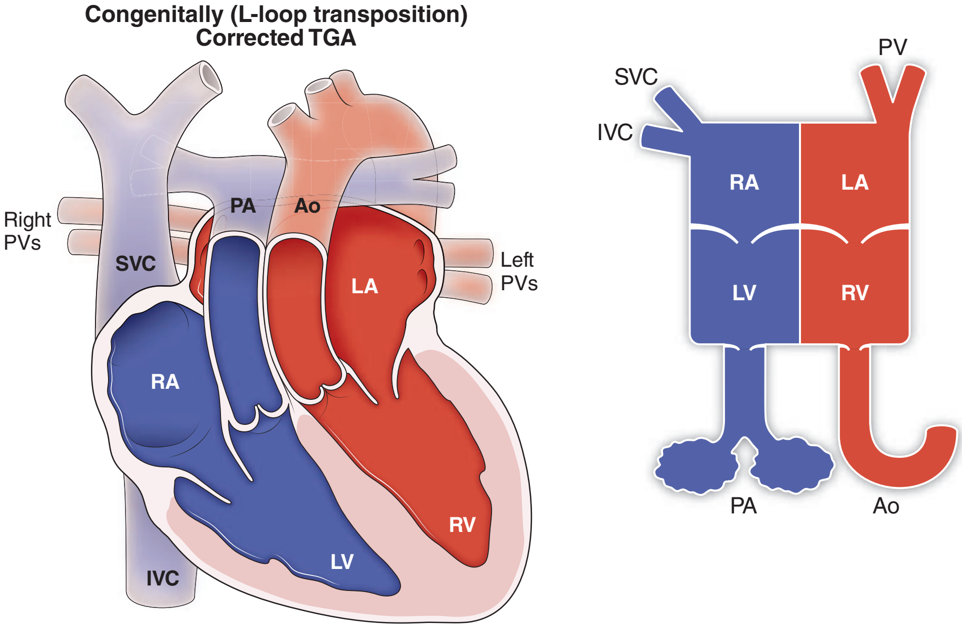
Congenitally (L-loop transposition) Corrected TGA

FIGURE 280-10 Congenitally corrected transposition of the great arteries. Physiologically corrected transposition of the great arteries (also known as *congenitally corrected transposition of the great arteries*) is characterized by atrioventricular discordance and ventriculoarterial discordance. Systemic venous blood passes from the right atrium (RA) through the mitral valve into the morphologic left ventricle (LV) to the pulmonary artery (PA). Oxygenated blood then returns to the lungs to the left atrium (LA) through the tricuspid valve into the morphologic right ventricle (RV) and then out the aorta (Ao). IVC, inferior vena cava; PV, pulmonary veins; SVC, superior vena cava.

conversely, the left atrium conducting oxygenated pulmonary venous blood to the RV) as well as ventriculoarterial discordance (connections of LV to PA, RV to aorta). This results in normal arterial oxygen saturation, yet an RV associated with the aorta. Patients with L-loop TGA commonly have associated congenital anomalies, including dextrocardia, ASDs, a dysplastic tricuspid valve, and pulmonary stenosis. Conduction disturbances are common, and complete heart block occurs in up to 30% of patients. Those patients without associated defects may not present until later in life, most commonly with heart failure, TR, or newly recognized conduction disease.

Coarctation of the Aorta Adults with coarctation of the aorta (Fig. 280-11) typically have a shelf-like obstruction at the level of the descending aorta that passes just posterior to the junction of the main and left PA; obstruction less commonly involves the transverse aortic arch. On physical examination, the lower extremity blood pressure and pulses are lower than (and delayed in timing, in contrast to) the upper extremity values, unless significant aortic collaterals have developed. A continuous murmur over the scapula may be present due to the collateral blood flow. Significant coarctation increases afterload to all proximal structures in the path of oxygenated blood, from LV and coronary arteries to ascending and transverse aorta, to cerebral and arm vessels and proximal descending aorta. Bicuspid aortic valve (typically with right-left commissural fusion) is a common association. In women with short stature, webbed neck, lymphedema, and primary amenorrhea, a concomitant diagnosis of Turner syndrome should be considered, the presence of which indicates greater degree of, and risks from, sequelae from seemingly similar anatomy and physiology. Patients who have undergone surgical repair in general have a good prognosis; however, they remain at risk for systemic hypertension, premature atherosclerosis, LV failure, and aortic aneurysm, dissection, and recurrent coarctation.

Single Ventricle The term *single ventricle heart disease* is imprecise but useful in some settings, as it refers to congenital heart conditions in which one ventricle or its valves preclude surgical creation of a biventricular circulation. Common congenital diagnoses in this category include tricuspid atresia, double inlet LV, and hypoplastic left heart syndrome. Most patients with single ventricle physiology undergo a series of surgeries culminating in a Fontan procedure (Fig. 280-12*A*, *B*). Since its initial use for tricuspid valve atresia in 1971, multiple modifications of this procedure have occurred, with common features of near complete separation of the pulmonary and systemic circulations. The Fontan procedure utilizes the single ventricle to pump pulmonary venous (oxygenated) blood through the aorta to the body and allows for "passive" flow of systemic venous return of deoxygenated blood through surgically created connections to the lungs. Patients who have undergone a Fontan procedure are at risk for multiple comorbidities in adulthood, including atrial arrhythmias, heart failure, renal and hepatic dysfunction, and both venous and arterial thrombosis and embolism.

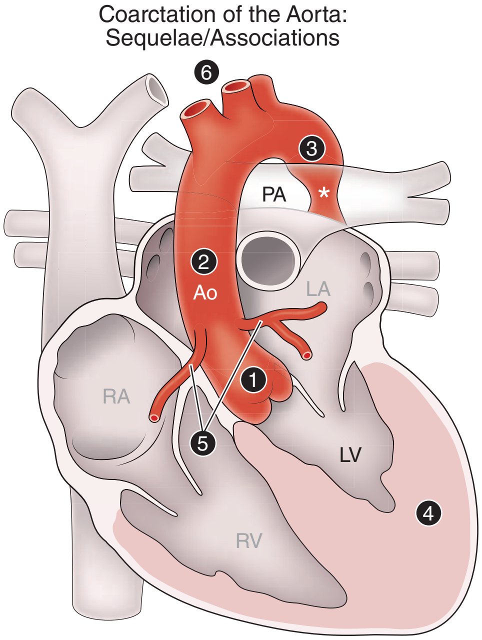
Coarctation of the Aorta: Sequelae/Associations

FIGURE 280-11 Aortic coarctation (*). Bicuspid aortic valve (1) is most common concomitant lesion. Sequelae from aortic coarctation (unrepaired or repaired) include systemic arterial hypertension, ascending (2) or descending (3) aortic enlargement or aneurysm formation, left ventricular (LV) hypertrophy (4), LV diastolic and systolic heart failure, accelerated coronary (5) or cerebral (6) atherosclerosis, cerebral aneurysm formation, and recurrence of coarctation after repair. Ao, aorta; PA, pulmonary arteries.

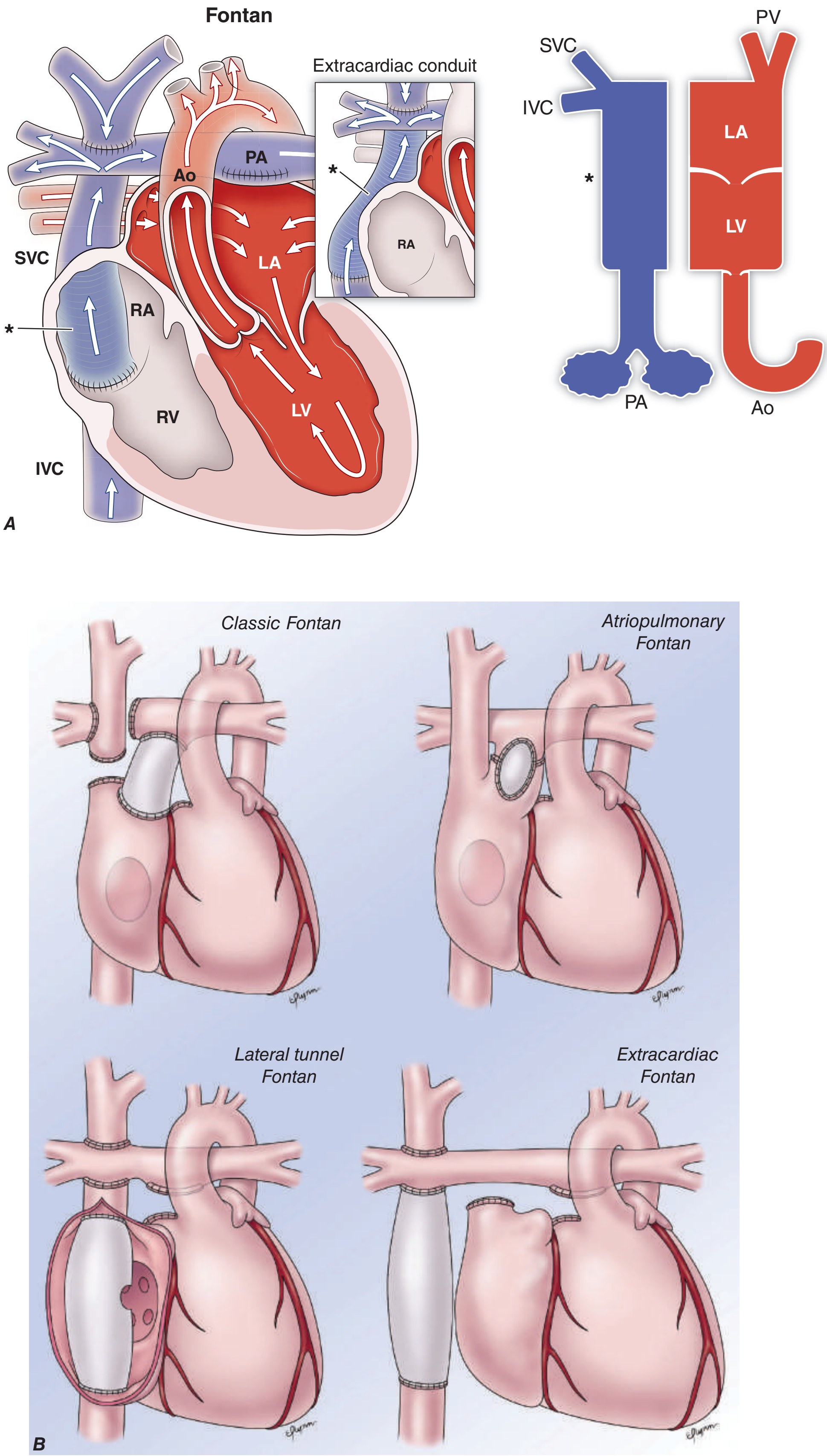
Fontan
Extracardiac conduit
SVC
PA
Ao
LA
RA
RV
LV
IVC
*
RA
A
SVC
IVC
PV
LA
LV
PA
Ao
Classic Fontan
Atriopulmonary Fontan
Lateral tunnel Fontan
Extracardiac Fontan
B

FIGURE 280-12 ***A.*** Fontan surgery creates a unique circulation in which deoxygenated blood is directed to the PAs from the SVC and IVC in a fashion that bypasses any pumping chamber. The SVC and IVC are connected (*) via either an internal "tunnel" or an extracardiac conduit that guides flow to the PA. Pulmonary venous (oxygenated) return courses from PV to LA to LV to aorta. In contrast to physiology in normal adults (where pressure is generated by an RV to propel blood flow from a lower pressure RA to a higher pressure LA), in Fontan circulation, by definition, due to the absence of a pumping chamber to the PA, RA pressure is greater than LA pressure, permitting flow through the lungs. Ao, aorta; IVC, inferior vena cava; LA, left atrium; LV, left ventricle; PA, pulmonary arteries; PV, pulmonary veins; SVC, superior vena cava; *, Fontan baffle. ***B.*** Diagrammatic representation of the location of various types of Fontan operations. *(Part B used with permission from Emily Flynn McIntosh, illustrator.)*

UNREPAIRED CYANOTIC CHD

Eisenmenger Syndrome ES is felt to be the consequence of a long-standing high-volume or pressurized left-to-right shunt in which excessive blood flow to the pulmonary vasculature leads to severely increased pulmonary vascular resistance that eventually results in reversal of the shunt, creating bidirectional or right-to-left flow. ES is a multiple-organ condition and may occur with any CHD with an initial left-to-right shunt. The natural history of ES is variable, and although there is significant morbidity, in general, adults with ES appear to survive longer than those with other forms of pulmonary arterial hypertension. Medical care recommendations have included sustaining adequate hydration, avoiding, and treating anemia including iron supplementation when appropriate, and anticoagulation (although this remains controversial due to predisposition to bleeding and occurrence of clinical hemoptysis, which has frequently been associated with pulmonary vascular thrombosis). Elevation of hematocrit above that considered appropriate for the degree of cyanosis can be managed in symptomatic patients by hydration alone or, on occasion, by performing phlebotomy with isovolumic replenishment. Routine phlebotomy in the asymptomatic adult with ES is contraindicated. Appropriate optimization of iron stores has been demonstrated to improve quality of life and functional performance in iron-deficient adults with ES. Contraception for women with ES who are of childbearing age is strongly recommended, avoiding use of estrogen, which may be thrombogenic. Pregnancy is contraindicated in these women due to the high risk of maternal mortality.

Selective pulmonary vasodilators, such as bosentan or sildenafil, are efficacious in ES. Select patients may be candidates for combined heart–lung transplantation or preferably lung transplantation with concomitant repair of the intracardiac defect, if feasible.

The Role of Palliative Care in ACHD In aggregate, adults with CHD demonstrate both quality-of-life–limiting comorbidities and premature mortality far in excess of age-matched controls. The reported prevalence of pain, anxiety, depression, dyspnea, and fatigue appears similar to that reported for adults who are decades older and engaged in palliative care for acquired cardiovascular disease at end of life (EOL). Similarly, at EOL with ACHD, frequencies of hospitalization, intensive care admission, 30-day readmission, and increased length of hospital stay appear greater (despite younger age) than for adults with cancer. In a retrospective study of ACHD patients who died during a hospital admission, only a minority had engaged in EOL discussions with their providers. Surveys of both adults with CHD and their providers suggest that the overwhelming majority of patients wish to participate in advanced care planning and discussion of palliative care; this contrasts with statements from ACHD care providers noting their uncertainties regarding EOL prognostication and concerns over discussion about EOL. Palliative care specialists who are embedded within or aligned with ACHD care teams can play an important and iterative role in defining and addressing alignment of patient and clinician goals within the boundaries of frequently complex care decisions over the adult life span.

Global Considerations As survival patterns improve for all medically complex patients, the internist and general practitioner are faced with particular challenges and dilemmas; foremost is accrual of sufficient knowledge and competency so as to be able both to engage in patient care provision as well as to seek greater expertise, guidance, and support, when such is appropriate. Across the globe, lifelong care for adults with CHD typifies this growing demand. Care for adults with CHD within medical care centers that contain an ACHD specialty care program has been associated with improved overall survival. However, current analyses suggest that the majority of adults with CHD seek and receive their medical care outside of such ACHD specialty care centers and within the hands of the general practitioner, internist, and cardiologist. Under a surface of adaptability and determination, adults with CHD present a wide spectrum of cognitive and functional performance, multiple organ system comorbidities, abnormalities of systemic and pulmonary vasculature, and a near universal presence of heart failure of one stage or another, all over a lifetime. It appears incumbent on the ACHD specialist and ACHD specialty care centers to serve as a hub for partnering practitioners, encouraging engagement to the level of highest competencies, and providing education, oversight, and support, so as to achieve optimal outcomes.

FURTHER READING

Gilboa SM et al: Congenital heart defects in the United States: Estimating the magnitude of the affected population in 2010. Circulation 134:101, 2016.

Gurvitz M et al: Emerging research directions in adult congenital heart disease: A report from an NHLBI/ACHA Working Group. J Am Coll Cardiol 67:1956, 2016.

Holzer R et al: PICS/AEPC/APPCS/CSANZ/SCAI/SOLACI: Expert consensus statement on cardiac catheterization for pediatric patients and adults with congenital heart disease. JACC Cardiovasc Interv 17:115, 2024.

Regitz-Zagrosek V et al: 2018 ESC guidelines for the management of cardiovascular diseases during pregnancy. Eur Heart J 39:3165, 2018.

Sachdeva R et al: ACC/AHA/ASE/HRS/ISACHD/SCAI/SCCT/SCMR/SOPE 2020 Appropriate Use Criteria for non-invasive multimodality imaging during the follow-up care of patients with congenital heart disease: A report of the American College of Cardiology Solution Set Oversight Committee and Appropriate Use Criteria Task Force, American Heart Association, American Society of Echocardiography, Heart Rhythm Society, International Society for Adult Congenital Heart Disease, Society for Cardiovascular Angiography and Interventions, Society of Cardiovascular Computed Tomography, Society for Cardiovascular Magnetic Resonance, and Society of Pediatric Echocardiography. J Am Coll Cardiol 75:657, 2020.

Silversides CK et al: Pregnancy outcomes in women with heart disease: The CARPREG II Study. J Am Coll Cardiol 71:2419, 2018.

Stout KK et al: 2018 AHA/ACC guidelines for the management of adults with congenital heart disease: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 139:e698, 2019.

281

Pericardial Disease

Joseph Loscalzo

NORMAL FUNCTIONS OF THE PERICARDIUM

The normal pericardium is a double-layered sac of the visceral pericardium and parietal pericardium. The visceral pericardium is a serous membrane that is separated from the fibrous parietal pericardium by a small quantity (15–50 mL) of fluid, an ultrafiltrate of plasma. The normal pericardium, by exerting a restraining force, prevents sudden dilation of the cardiac chambers, especially the right atrium and ventricle, e.g., during exercise. It also restricts the anatomic position of the heart and likely retards the spread of infections from the lungs and pleural cavities to the heart. Nevertheless, *total* absence of the pericardium, either congenital or after surgery, does not produce obvious clinical disease. In *partial* left pericardial defects, the main pulmonary artery and left atrium may bulge through the defect; very rarely, herniation and subsequent strangulation of the left atrium may cause sudden death.

ACUTE PERICARDITIS

Acute pericarditis, by far the most common pathologic process involving the pericardium **(Table 281-1)**, has four principal diagnostic features:

1. *Chest pain* is usually present in acute infectious pericarditis and in many of the forms presumed to be related to hypersensitivity,

TABLE 281-1 Classification of Pericarditis

Clinical Classification
I. Acute pericarditis (<6 weeks)
A. Fibrinous
B. Effusive (serous or sanguineous)
II. Subacute pericarditis (6 weeks to 6 months)
A. Effusive-constrictive
B. Constrictive
III. Chronic pericarditis (>6 months)
A. Constrictive
B. Adhesive (nonconstrictive)
Etiologic Classification
I. Infectious pericarditis
A. Viral (coxsackievirus A and B, echovirus, herpesviruses, mumps, adenovirus, hepatitis, HIV, post-acute COVID, mpox)
B. Pyogenic (pneumococcus, *Streptococcus, Staphylococcus, Neisseria, Legionella, Chlamydia*)
C. Tuberculous
D. Fungal (histoplasmosis, coccidioidomycosis, *Candida*, blastomycosis)
E. Other infections (syphilitic, protozoal, parasitic)
II. Noninfectious pericarditis
A. Acute idiopathic
B. Renal failure
C. Neoplasia
1. Primary tumors (benign or malignant, mesothelioma)
2. Tumors metastatic to pericardium (lung and breast cancer, lymphoma, leukemia)
D. Trauma (penetrating chest wall, nonpenetrating)
E. Aortic dissection (with leakage into pericardial sac)
F. Acute myocardial infarction
G. Postirradiation
H. Familial Mediterranean fever and other autoinflammatory syndromes
I. Familial pericarditis
1. Mulibrey nanism[a]
J. Metabolic (myxedema, cholesterol)
III. Pericarditis presumably related to autoimmunity
A. Rheumatic fever
B. Collagen vascular disease (systemic lupus erythematosus, rheumatoid arthritis, ankylosing spondylitis, scleroderma, acute rheumatic fever, granulomatosis with polyangiitis, IgG4 disease)
C. Drug-induced (e.g., procainamide, hydralazine, phenytoin, isoniazid, minoxidil, anticoagulants, methysergide)
D. Postcardiac injury
1. Postpericardiotomy
2. Posttraumatic
3. Postmyocardial infarction (Dressler's syndrome)

[a] An autosomal recessive syndrome characterized by growth failure, muscle hypotonia, hepatomegaly, ocular changes, enlarged cerebral ventricles, intellectual disability, ventricular hypertrophy, and chronic constrictive pericarditis.

Abbreviation: COVID, coronavirus disease.

autoimmunity, or of unknown cause (idiopathic). The pain of acute pericarditis is often severe, retrosternal and/or left precordial, and referred to the neck, arms, or left shoulder. Frequently the pain is pleuritic, consequent to accompanying pleural inflammation (i.e., sharp and aggravated by inspiration and coughing); however, at times, it is steady, radiates to the trapezius ridge or into either arm, and resembles that of myocardial ischemia. For this reason, confusion with acute myocardial infarction (AMI) is common. Characteristically, pericardial pain may be intensified by lying supine and relieved by sitting up and leaning forward **(Chap. 15)**. Pain is often absent in slowly developing tuberculous, postirradiation, neoplastic, and uremic pericarditis.

The differentiation of AMI from acute pericarditis may be challenging when, with the latter, serum biomarkers of myocardial damage rise, presumably because of concomitant involvement of the epicardium in the inflammatory process (an epi-myocarditis) with resulting myocyte necrosis. If they occur, however, these elevations are quite modest compared to those in AMI, given the superficial (sub-)epicardial injury despite often extensive electrocardiographic ST-segment elevation in pericarditis. This dissociation is useful in differentiating between these conditions.

2. A *pericardial friction rub* is audible at some point in the illness in approximately 85% of patients with acute pericarditis. The rub may have up to three components per cardiac cycle and is described as rasping, scratching, or grating **(Chap. 246)**; it is heard most frequently at end expiration with the patient upright and leaning forward.
3. The *electrocardiogram* (ECG) in acute pericarditis without massive effusion usually displays changes secondary to acute subepicardial inflammation **(Fig. 281-1*A*)**, and typically evolves through four stages. In stage 1, there is widespread elevation of the ST segments, often with upward concavity, involving two or three standard limb leads and V_2–V_6, with reciprocal depressions only in aVR and occasionally V_1. In addition, there is depression of the PR segment below the TP segment, reflecting atrial involvement, an early change that may occur prior to ST segment elevation. Usually there are no significant changes in QRS complexes unless a large pericardial effusion develops (see below). After several days, the ST segments return to normal (stage 2), and only then, or even later, do the T waves become inverted (stage 3). Weeks or months after the onset of acute pericarditis, the ECG returns to normal (stage 4). In contrast, in AMI, ST elevations are upwardly convex, and reciprocal depression is usually more prominent; these changes may return to normal within a day or two **(Chaps. 285 and 286)**.
4. *Pericardial effusion* is usually associated with pain and/or the ECG changes mentioned above and, if the effusion is large, with electrical alternans **(Fig. 281-1*B*)**. Pericardial effusion is especially important clinically when it develops within a relatively short time because it may lead to cardiac tamponade (see below). Differentiation from cardiac enlargement on physical examination may be difficult, but heart sounds may be fainter with large pericardial effusion. The friction rub and the apex impulse may disappear. The base of the left lung may be compressed by pericardial fluid, producing *Ewart's sign*, a patch of dullness, increased fremitus, and egophony beneath the angle of the left scapula. The chest x-ray may show enlargement of the cardiac silhouette, with a "water bottle" configuration, but may be normal in patients with small effusions.

Diagnosis *Echocardiography* **(Chap. 248)** is the most widely used imaging technique. It is sensitive, specific, simple, and noninvasive; may be performed at the bedside; and allows localization and estimation of the quantity of pericardial fluid. The presence of pericardial fluid is recorded by two-dimensional transthoracic echocardiography as a relatively echo-free space between the posterior pericardium and left ventricular epicardium and/or as a space between the anterior right ventricle and the parietal pericardium just beneath the anterior chest wall **(Fig. 281-2)**.

The diagnosis of pericardial fluid or thickening may be confirmed by computed tomography (CT) or magnetic resonance imaging (MRI). These techniques may be superior to echocardiography in detecting loculated pericardial effusions and pericardial thickening, and in the identification of pericardial masses. MRI is also helpful in detecting pericardial inflammation **(Fig. 281-3)**.

TREATMENT

Acute Pericarditis

There is no specific therapy for acute idiopathic pericarditis, but bed rest may be recommended, and anti-inflammatory treatment with aspirin (2–4 g/d) or nonsteroidal anti-inflammatory drugs (NSAIDs), such as ibuprofen (600–800 mg tid) or indomethacin (25–50 mg tid), should be administered along with gastric protection (e.g., omeprazole 20 mg/d). In responsive

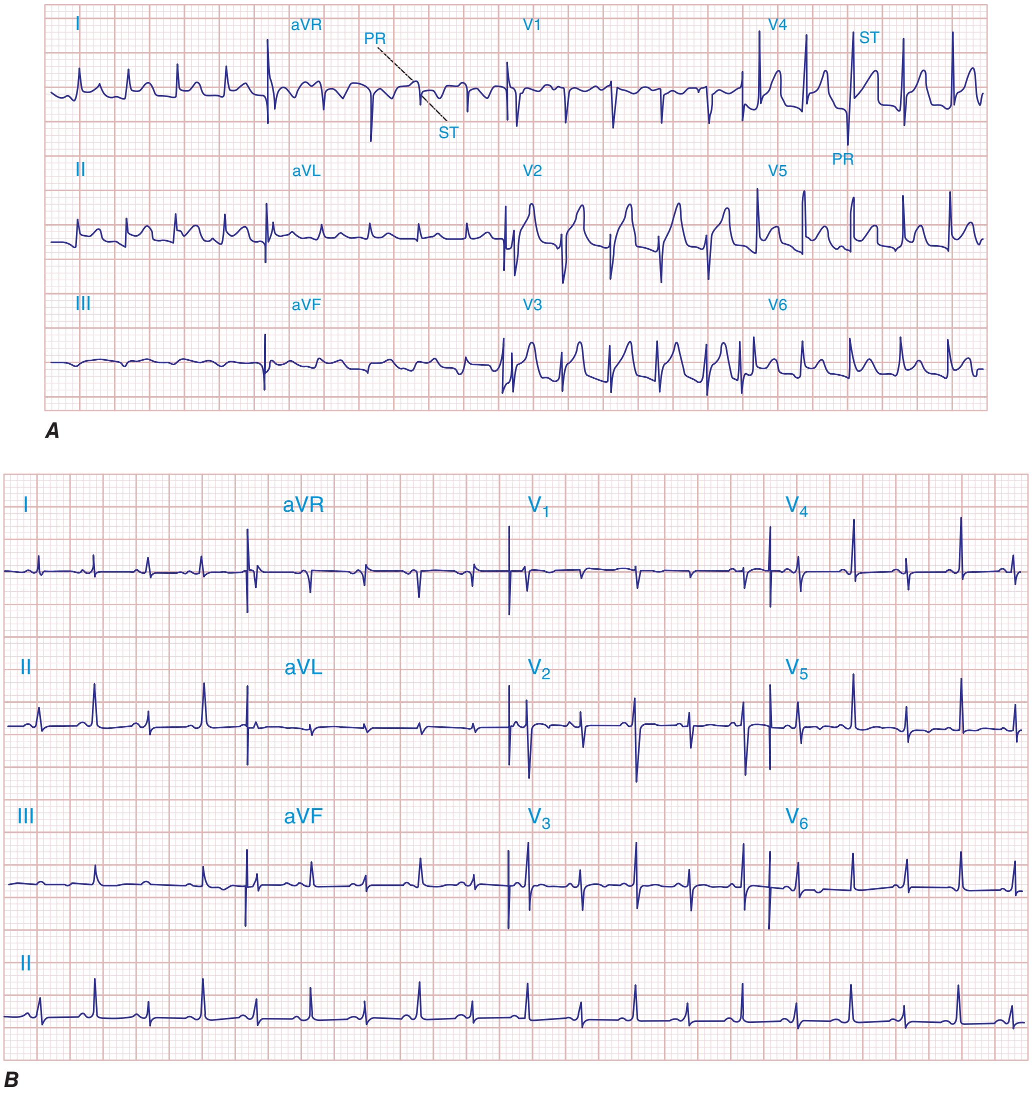

FIGURE 281-1 ***A.*** Acute pericarditis. There are diffuse ST-segment elevations in leads I, II, aVF, and V_2–V_6). There is PR-segment depression due to a concomitant atrial injury current. ***B.*** Electrical alternans. This tracing was obtained from a patient with a large pericardial effusion with cardiac tamponade.

patients, these doses should be continued for 1–2 weeks and then tapered over several weeks. In addition, colchicine (0.5 to 0.6 mg qd [<70 kg] or 0.5 to 0.6 mg bid [>70 kg]) should be administered for 3 months. Colchicine enhances the response to NSAIDs and also aids in reducing the risk of recurrent pericarditis. This drug is concentrated in and interferes with the migration of neutrophils, may cause diarrhea and other gastrointestinal side effects, and is contraindicated in patients with hepatic or renal dysfunction. Glucocorticoids (e.g., prednisone 1 mg/kg per day) usually suppress the clinical manifestations of acute pericarditis in patients who have failed therapy with or do not tolerate NSAIDs and colchicine. However, since they increase the risk of subsequent recurrence, full-dose corticosteroids should be given for only 2–4 days and then tapered. Anticoagulants should be avoided because their use could cause bleeding into the pericardial cavity and tamponade.

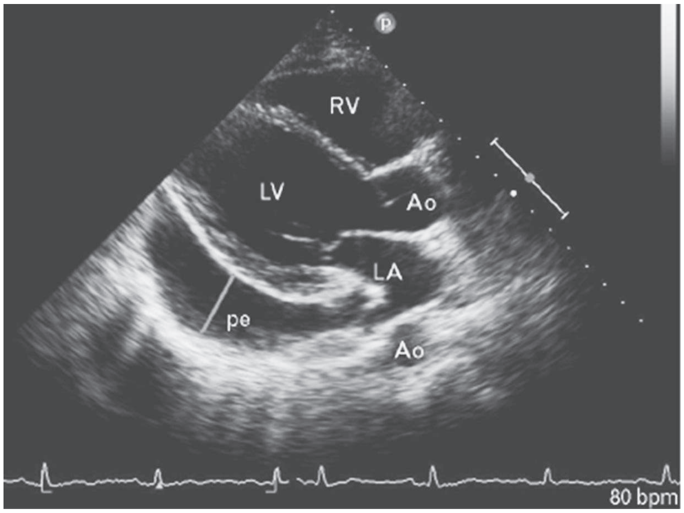

FIGURE 281-2 Two-dimensional echocardiogram in lateral view in a patient with a large pericardial effusion. Ao, aorta; LA, left atrium; LV, left ventricle; pe, pericardial effusion; RV, right ventricle. *(Reproduced with permission from Imazio M: Contemporary management of pericardial diseases. Curr Opin Cardiol 27:308, 2012.)*

In patients with multiple, frequent, and disabling recurrences that continue for >2 years, are not prevented by continuing colchicine and other NSAIDs, and are not controlled by glucocorticoids, treatment with azathioprine or anakinra (an interleukin 1β receptor antagonist) has been reported to be of benefit. Rarely, pericardial stripping may be necessary; however, this procedure may not always terminate the recurrences.

The majority of patients with acute pericarditis can be managed as outpatients with careful follow-up. However, when specific causes (tuberculosis, neoplastic disease, bacterial infection) are suspected, or if any of the predictors of poor prognosis (fever >38°C, subacute onset, large pericardial effusion, immunosuppression, or evidence of pericardial tamponade) are present, hospitalization is advisable.

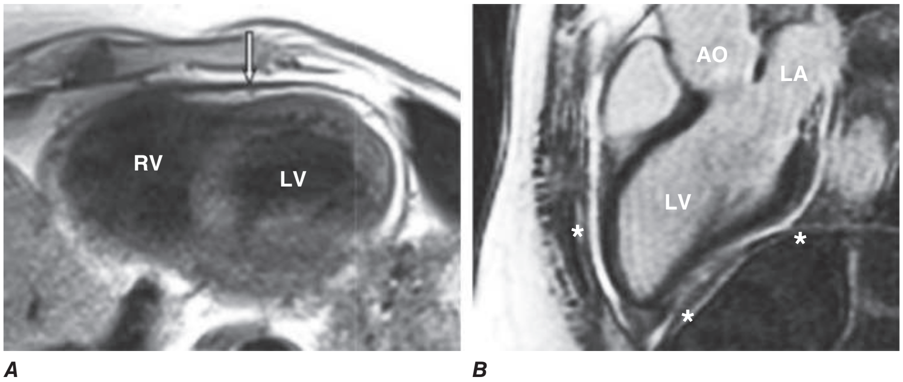

FIGURE 281-3 Pericardial inflammation by cardiac magnetic resonance imaging. ***A.*** Short axis view. The pericardium is thickened and enhanced on T2 magnetic images. Note thickened white line denoted by arrow. ***B.*** Long axis view. Late gadolinium enhancement of thickened, inflamed pericardium. AO, aorta; LA, left atrium; LV, left ventricle; RV, right ventricle. *(From RY Kwong: Cardiovascular magnetic resonance imaging, in Braunwald's Heart Disease, 10th ed, Mann DL et al [eds]. Philadelphia: Elsevier, 2015, pp 320–40.)*

CARDIAC TAMPONADE

The accumulation of fluid in the pericardial space in a quantity sufficient to cause serious obstruction of the inflow of blood into the ventricles results in cardiac tamponade. This complication may be fatal if it is not recognized and treated promptly. The most common causes of tamponade are idiopathic pericarditis and pericarditis secondary to neoplastic disease, tuberculosis, or bleeding into the pericardial space after leakage from an aortic dissection, cardiac operation, trauma, or treatment with anticoagulants.

The three principal features of tamponade (*Beck's triad*) are hypotension, soft or absent heart sounds, and jugular venous distention with a prominent *x* (early systolic) descent but an absent *y* (early diastolic) descent. The limitations to ventricular filling are responsible for reductions of cardiac output and arterial pressure. The quantity of fluid necessary to produce cardiac tamponade may be as small as 200 mL when the fluid develops rapidly or be as much as >2000 mL in slowly developing effusions when the pericardium has had the opportunity to stretch and adapt to an increasing volume.

A high index of suspicion for cardiac tamponade is required because in many instances no obvious cause for pericardial disease is apparent. This diagnosis should be considered in any patient with otherwise unexplained sudden enlargement of the cardiac silhouette, hypotension, and elevation of jugular venous pressure. Reductions in amplitude of the QRS complexes and *electrical alternans* of the P, QRS, or T waves should also raise the suspicion of cardiac tamponade (Fig. 281-1).

Table 281-2 lists the features that distinguish acute cardiac tamponade from constrictive pericarditis.

Paradoxical Pulse This important clue to the presence of cardiac tamponade consists of a greater than normal (10 mmHg) inspiratory decline in systolic arterial pressure. When severe, it may be detected by palpating weakness or even disappearance of the arterial pulse during inspiration, but usually sphygmomanometric measurement of systolic pressure during slow respiration is required.

Because both ventricles share a tight incompressible covering, i.e., the pericardial sac, the inspiratory enlargement of the right ventricle

TABLE 281-2 Features That Distinguish Cardiac Tamponade from Constrictive Pericarditis and Similar Clinical Disorders

CHARACTERISTIC	TAMPONADE	CONSTRICTIVE PERICARDITIS	RESTRICTIVE CARDIOMYOPATHY	RIGHT VENTRICULAR MYOCARDIAL INFARCTION	EFFUSIVE CONSTRICTIVE PERICARDITIS
Clinical					
Pulsus paradoxus	+++	+	+	+	+++
Jugular veins					
Prominent *y* descent	–	++	+	+	–
Prominent *x* descent	+++	++	+++	+	+++
Kussmaul's sign	–	+++	+	+++	++
Third heart sound	–	–	+	+	+
Pericardial knock	–	++	–	–	–
Electrocardiogram					
Low ECG voltage	++	++	+	–	+
Electrical alternans	++	–	–	–	+
Echocardiogram					
Thickened pericardium	–	+++	–	–	++
Pericardial calcification	–	++	–	–	_
Pericardial effusion	+++	–	–	–	++
RV size	Usually small	Usually normal	Usually normal	Enlarged	Usually normal
Exaggerated respiratory variation in flow velocity	+++	+++	–	+++	+
CT/MRI					
Thickened pericardium	–	+++	–		++
Equalization of diastolic pressures	+++	+++	–	++	++

Abbreviations: +++, always present; ++, usually present; +, rare; –, absent; CT, computed tomography; ECG, electrocardiogram; MRI, magnetic resonance imaging; RV, right ventricle.

Source: Reproduced with permission from GM Brockington et al: Constrictive pericarditis. Cardiol Clin 8:645, 1990.

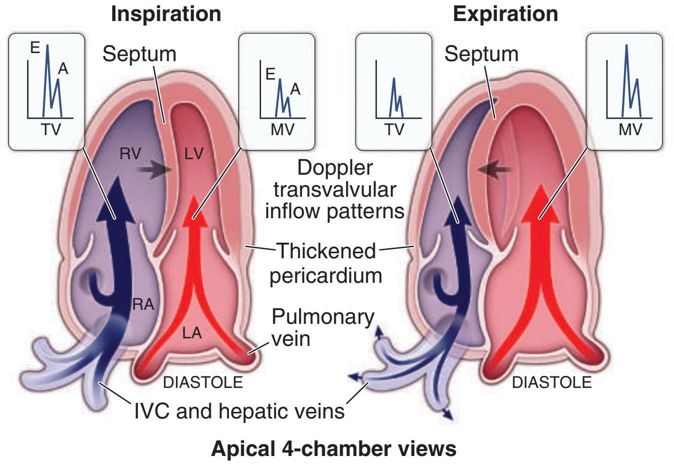

FIGURE 281-4 Constrictive pericarditis. Doppler schema of respirophasic changes in mitral and tricuspid inflow. Reciprocal patterns of ventricular filling are assessed on pulsed Doppler examination of mitral valve (MV) and tricuspid valve (TV) inflow. IVC, inferior vena cava; LA, left atrium; LV, left ventricle; RA, right atrium; RV, right ventricle. *(Courtesy of Bernard E. Bulwer, MD.)*

causes leftward bulging of the interventricular septum, reducing left ventricular volume, stroke volume, and arterial systolic pressure. Paradoxical pulse also occurs in approximately one-third of patients with constrictive pericarditis (see below), and in some cases of hypovolemic shock, acute and chronic obstructive airway disease, and pulmonary embolism. Right ventricular infarction **(Chap. 286)** may resemble cardiac tamponade with hypotension, elevated jugular venous pressure, a slow *y* descent in the jugular venous pulse, and, occasionally, a paradoxical pulse (Table 281-2).

Diagnosis Because immediate treatment of cardiac tamponade may be lifesaving, prompt establishment of the diagnosis, usually by echocardiography, should be undertaken. When pericardial effusion causes tamponade, Doppler ultrasound shows that tricuspid and pulmonic valve flow velocities increase markedly during inspiration, whereas pulmonic vein, mitral, and aortic flow velocities decrease (as in constrictive pericarditis, see below) **(Fig. 281-4)**. In tamponade, there is late diastolic inward motion (collapse) of the right ventricular free wall and the right atrium. Transesophageal echocardiography, CT, or cardiac MRI may be necessary to diagnose a loculated effusion responsible for cardiac tamponade.

TREATMENT

Cardiac Tamponade

Patients with acute pericarditis should be observed frequently for the development of an effusion. If a large effusion is present, pericardiocentesis should be performed or the patient watched closely for signs of tamponade with serial echocardiography and monitoring of arterial and venous pressures.

PERICARDIOCENTESIS

If manifestations of tamponade appear, pericardiocentesis using an apical, parasternal, or, most commonly, subxiphoid approach must be carried out at once because if left untreated, tamponade may be rapidly fatal. Whenever possible, this procedure should be carried out under echocardiographic guidance. Intravenous saline may be administered as the patient is being readied for the procedure, but the pericardiocentesis must not be delayed. If possible, intrapericardial pressure should be measured before fluid is withdrawn, and the pericardial cavity should be drained as completely as possible. A small, multiholed catheter may be advanced over the needle inserted into the pericardial cavity and left in place to allow draining of the pericardial space if fluid reaccumulates. Surgical drainage through a limited (subxiphoid) thoracotomy may be required in recurrent tamponade to remove loculated effusions and/or when it is necessary to obtain tissue for diagnosis.

Pericardial fluid obtained from an effusion may have the physical characteristics of an exudate. In developed nations, bloody fluid is most commonly due to neoplasm, renal failure, or cardiac trauma. In developing nations, tuberculosis may also cause exudative and/or bloody effusion.

The pericardial fluid should be analyzed for red and white blood cells and cytology for neoplastic cells. Cultures should be obtained. The presence of DNA of *Mycobacterium tuberculosis* determined by the polymerase chain reaction and/or of an elevated adenosine deaminase strongly supports the diagnosis of tuberculous pericarditis; however, it is often necessary to obtain pericardial tissue to make this diagnosis **(Chap. 183)**.

■ VIRAL OR IDIOPATHIC ACUTE PERICARDITIS

In many instances, acute pericarditis occurs in association with or following illnesses of known or presumed viral origin and probably is caused by the same agent. There may be an antecedent infection of the respiratory tract, but viral isolation and serologic studies are usually negative. In some cases, coxsackievirus A or B or the virus of influenza, echovirus, mumps, herpes simplex, varicella-zoster, adenovirus, or cytomegalovirus has been isolated from pericardial fluid, and/or appropriate elevations in viral antibody titers have been observed. Frequently, a viral cause cannot be established, and the term *idiopathic acute pericarditis* is appropriate.

Viral or idiopathic acute pericarditis occurs at all ages but is most common in young adult males and is often associated with pleural effusion and pneumonitis. The almost simultaneous development of fever and precordial pain, often 10–12 days after a presumed viral illness, constitutes an important feature in the differentiation of acute pericarditis from AMI, in which chest pain precedes fever. The constitutional symptoms are usually mild to moderate, and a pericardial friction rub is often audible. The disease ordinarily runs its course in a few days to 4 weeks. Elevations of C-reactive protein and of the white blood cell count are common. The ST-segment alterations in the ECG usually disappear after 1 or more weeks, but the abnormal T waves may persist for as long as several years and be a source of confusion in persons without a clear history of pericarditis. Accumulation of some pericardial fluid is common, and both tamponade and constrictive pericarditis are possible, but infrequent, complications.

The most frequent complication is recurrent (relapsing) pericarditis, which occurs in approximately one-fourth of patients with acute idiopathic pericarditis. A smaller number of individuals have multiple recurrences.

Postcardiac Injury Syndrome Acute pericarditis may appear in a variety of circumstances that have one common feature—previous injury to the myocardium with blood in the pericardial cavity. The syndrome may develop after a cardiac operation (postpericardiotomy syndrome), after blunt or penetrating cardiac trauma **(Chap. 283)**, or after perforation of the heart with a catheter; rarely, it follows AMI.

The clinical picture mimics acute viral or idiopathic pericarditis. The principal symptom is the pain of acute pericarditis, which usually develops 1–4 weeks after the cardiac injury. Recurrences are common and may occur up to 2 years or more following the injury. Fever, pleuritis, and pneumonitis are accompanying features, and the illness usually subsides in 1 or 2 weeks. The pericarditis may be of the fibrinous variety, or it may be a pericardial effusion, which is often serosanguinous and rarely causes tamponade. ECG changes typical of acute pericarditis may also occur. This syndrome is probably the result of a hypersensitivity (or autoimmune) reaction to antigen(s) that originate from injured myocardial tissue and/or pericardium.

Often no treatment is necessary aside from aspirin and analgesics. When the illness is severe or followed by a series of disabling recurrences, therapy with another NSAID, colchicine, or a glucocorticoid, such as described for treatment of acute pericarditis, is usually effective.

■ DIFFERENTIAL DIAGNOSIS

Because there is no specific test for *acute idiopathic pericarditis*, the diagnosis is one of exclusion. Consequently, all other disorders that may be associated with acute fibrinous pericarditis must be considered. A common diagnostic error is mistaking acute viral or idiopathic pericarditis for AMI and vice versa.

Pericarditis secondary to postcardiac injury is differentiated from acute idiopathic pericarditis chiefly by timing. If it occurs within a few days or weeks of a chest blow, a cardiac perforation, a cardiac operation, or an AMI, the two are probably related.

It is important to distinguish *pericarditis due to collagen vascular disease* from acute idiopathic pericarditis. Most important in the differential diagnosis is the pericarditis due to systemic lupus erythematosus (SLE; **Chap. 368**) or drug-induced (hydralazine or procainamide) lupus. When pericarditis occurs in the absence of any obvious underlying disorder, the diagnosis of SLE may be suggested by a rise in the titer of antinuclear antibodies. Acute pericarditis is an occasional complication of *rheumatoid arthritis*, *scleroderma*, and *polyarteritis nodosa*, and other evidence of these diseases is usually obvious at the time of presentation with acute pericarditis.

Pyogenic (purulent) pericarditis is usually secondary to cardiothoracic operations, by extension of infection from the lungs or pleural cavities, from rupture of the esophagus into the pericardial sac, or from rupture of a valvular ring abscess in a patient with infective endocarditis. It may also complicate the viral, bacterial, mycobacterial, and fungal infections that occur with HIV infection. It is generally accompanied by fever, chills, septicemia, and evidence of infection elsewhere, and generally has a poor prognosis. The diagnosis is made by examination of the pericardial fluid. It requires immediate drainage as well as vigorous antibiotic treatment.

Pericarditis of renal failure (*uremic pericarditis*) occurs in up to one-third of patients with severe renal dysfunction and is also seen in patients undergoing chronic dialysis who have normal levels of blood urea nitrogen (*dialysis-associated pericarditis*). These two forms of pericarditis may be fibrinous and are generally associated with serosanguinous effusions; frank hemorrhagic effusions may be seen in some cases of uremic pericarditis prior to the onset of dialysis. A pericardial friction rub is common, but pain is usually absent or mild. Treatment with an NSAID and intensification of dialysis are usually adequate. Occasionally, tamponade occurs and pericardiocentesis is required. When the pericarditis of renal failure is recurrent or persistent, a pericardial window should be created or pericardiectomy may be necessary.

Pericarditis due to *neoplastic diseases* results from extension or invasion of metastatic tumors (most commonly carcinoma of the lung and breast, malignant melanoma, lymphoma, and leukemia) to the pericardium. The pain of pericarditis, tamponade, and atrial arrhythmias are complications that occur occasionally. Diagnosis is made by pericardial fluid cytology or pericardial biopsy. *Mediastinal irradiation* for neoplasm may cause acute pericarditis and/or chronic constrictive pericarditis. Unusual causes of acute pericarditis include syphilis, fungal infection (histoplasmosis, blastomycosis, aspergillosis, and candidiasis), and parasitic infestation (amebiasis, toxoplasmosis, echinococcosis, and trichinosis) (Table 281-1).

CHRONIC PERICARDIAL EFFUSIONS

Chronic pericardial effusions are sometimes encountered in patients without an antecedent history of acute pericarditis. They may cause few symptoms per se, and their presence may be detected by finding an enlarged cardiac silhouette on a chest roentgenogram. *Tuberculosis* and myxedema may be causal. Neoplasms, SLE, rheumatoid arthritis, mycotic infections, radiation therapy to the chest, and chylopericardium may also cause chronic pericardial effusion and should be considered and specifically sought in such patients. Aspiration and analysis of the pericardial fluid are often helpful in diagnosis. Pericardial fluid should be analyzed as described under pericardiocentesis. Grossly sanguineous pericardial fluid results most commonly from a neoplasm, tuberculosis, renal failure, or slow leakage from an aortic dissection. Pericardiocentesis may resolve large effusions, but pericardiectomy may be required in patients with recurrence. Intrapericardial instillation of sclerosing agents may be used to prevent reaccumulation of fluid, most commonly in recurrent neoplastic effusions.

CHRONIC CONSTRICTIVE PERICARDITIS

This disorder results when the healing of an acute fibrinous or serofibrinous pericarditis or the resorption of a chronic pericardial effusion is followed by obliteration of the pericardial cavity with the formation of granulation tissue. The latter gradually contracts and forms a firm scar encasing the heart, which may become calcified. In developing nations, a high percentage of cases are of tuberculous origin, but this is now an uncommon cause in North America or Western Europe. Chronic constrictive pericarditis may follow acute or relapsing viral or idiopathic pericarditis, trauma with organized blood clot, or cardiac surgery of any type, or results from mediastinal irradiation, purulent infection, histoplasmosis, neoplastic disease (especially breast cancer, lung cancer, and lymphoma), rheumatoid arthritis, SLE, or chronic renal failure treated by chronic dialysis. In many patients, the cause of the pericardial disease is undetermined, and in these patients, an asymptomatic or forgotten bout of viral pericarditis, idiopathic or acute, may have been the inciting event.

The basic physiologic abnormality in patients with chronic constrictive pericarditis is the inability of the ventricles to fill owing to the limitations imposed by the rigid, thickened pericardium. Ventricular filling is unimpeded during early diastole but is reduced abruptly when the elastic limit of the pericardium is reached, whereas in cardiac tamponade, ventricular filling is impeded throughout diastole. In both conditions, ventricular end-diastolic and stroke volumes are reduced and the end-diastolic pressures in both ventricles and the mean pressures in the atria, pulmonary veins, and systemic veins are all elevated to similar levels (i.e., within 5 mmHg of one another). Despite these hemodynamic changes, systolic function may be normal or only slightly impaired at rest. However, in advanced cases, the fibrotic process may extend into the myocardium and cause myocardial scarring and atrophy, and venous congestion may then be due to the combined effects of the pericardial and myocardial lesions.

In constrictive pericarditis, the right and left atrial pressure pulses display an M-shaped contour, with prominent x and y descents. The y descent, which is absent or diminished in cardiac tamponade, is the most prominent deflection in constrictive pericarditis; it reflects rapid early filling of the ventricles. The y descent is interrupted by a rapid rise in atrial pressure during early diastole, when ventricular filling is impeded by the constricting pericardium. These characteristic changes are transmitted to the jugular veins where they may be recognized by inspection. In constrictive pericarditis, the ventricular pressure pulses in both ventricles exhibit the characteristic "square root" sign during diastole. These hemodynamic changes, although characteristic, are not pathognomonic of constrictive pericarditis and may also be observed in restrictive cardiomyopathies **(Chaps. 266–270, Table 266-1)**.

CLINICAL AND LABORATORY FINDINGS

Weakness, fatigue, weight gain, increased abdominal girth, abdominal discomfort, and edema are common. The patient often appears chronically ill, and in advanced cases, anasarca, skeletal muscle wasting, and cachexia may be present. Exertional dyspnea is common, and orthopnea may occur, although it is usually not severe. The neck veins are distended and may remain so even after intensive diuretic treatment, and venous pressure may fail to decline during inspiration (*Kussmaul's sign*). The latter is common in chronic pericarditis but may also occur in tricuspid stenosis, right ventricular infarction, and restrictive cardiomyopathy.

The pulse pressure is normal or reduced. A paradoxical pulse can be detected in about one-third of cases. Congestive hepatomegaly is pronounced, may impair hepatic function, and may cause jaundice; ascites is common and is usually more prominent than dependent edema. Pleural effusions and splenomegaly may also be present. The apical pulse is reduced and may retract in systole (*Broadbent's sign*). The heart sounds may be distant; an early third heart sound (i.e., a pericardial knock) occurring at the cardiac apex with the abrupt cessation of ventricular filling is often conspicuous.

The ECG frequently displays low voltage of the QRS complexes and diffuse flattening or inversion of the T waves. Atrial fibrillation is present in about one-third of patients. The *chest roentgenogram* shows a normal or slightly enlarged heart. Pericardial calcification is most common in tuberculous pericarditis. Pericardial calcification may, however, occur in the absence of constriction, and constriction may occur without calcification.

Inasmuch as the common physical signs of cardiac disease (murmurs, cardiac enlargement) may be inconspicuous or absent in chronic constrictive pericarditis, hepatic enlargement and dysfunction associated with jaundice and intractable ascites may lead to a mistaken diagnosis of hepatic cirrhosis. This error can be avoided if the neck veins are inspected and found to be distended with the characteristic waveform features.

The transthoracic *echocardiogram* often shows pericardial thickening, dilation of the inferior vena cava and hepatic veins, and a sharp halt to rapid left ventricular filling in early diastole, with normal ventricular systolic function and flattening of the left ventricular posterior wall. There is a distinctive pattern of transvalvular flow velocity on Doppler echocardiography (Fig. 281-4). During inspiration, there is an exaggerated reduction in blood flow velocity in the pulmonary veins and across the mitral valve, and a leftward shift of the ventricular septum; the opposite occurs during expiration. Diastolic flow velocity in the inferior vena cava into the right atrium and across the tricuspid valve increases in an exaggerated manner during inspiration and declines during expiration. However, echocardiography cannot definitively establish or exclude the diagnosis of constrictive pericarditis; CT and MRI are more accurate, with the latter useful in evaluating myocardial involvement.

DIFFERENTIAL DIAGNOSIS

As with chronic constrictive pericarditis, cor pulmonale **(Chap. 264)** may be associated with marked systemic venous hypertension, little pulmonary congestion, a (left) heart that is not enlarged, and a paradoxical pulse. However, in cor pulmonale, advanced parenchymal pulmonary disease is usually apparent and venous pressure *falls* during inspiration (i.e., Kussmaul's sign is negative). *Tricuspid stenosis* **(Chap. 277)** may also simulate chronic constrictive pericarditis with congestive hepatomegaly, splenomegaly, ascites, and venous distention. However, the characteristic murmur and that of accompanying mitral stenosis are usually present.

Because it can be corrected surgically, it is important to distinguish chronic constrictive pericarditis from restrictive cardiomyopathy **(Chap. 266)**, which has a similar pathophysiologic underpinning (i.e., restriction of ventricular filling). The differentiating features are summarized in Table 281-2. When a patient has progressive, disabling, and unresponsive congestive heart failure and displays any of the features of constrictive heart disease, Doppler echocardiography to record respiratory effects on transvalvular flow (Fig. 281-4) should be performed and an MRI or CT scan should be obtained to detect or exclude constrictive pericarditis because the latter is usually correctable.

TREATMENT

Constrictive Pericarditis

Pericardial resection is the only definitive treatment of constrictive pericarditis and should be as complete as possible. Coronary arteriography should be carried out preoperatively in patients aged >50 years to exclude accompanying coronary artery disease. The benefits derived from cardiac decortication are usually progressive over a period of months. The risk of this operation depends on the extent of penetration of the myocardium by the fibrotic and calcific process, the severity of myocardial atrophy, the extent of secondary impairment of hepatic and/or renal function, and the patient's general condition. Operative mortality is in the range of 5–10% even in experienced centers; the patients with the most severe disease, especially secondary to radiation therapy, are at highest risk. Therefore, surgical treatment should, if possible, be carried out as early as possible.

Subacute Effusive-Constrictive Pericarditis This form of pericardial disease is characterized by the combination of a tense effusion in the pericardial space and constriction of the heart by thickened pericardium. As such, it shares a number of features with both chronic pericardial effusion producing cardiac compression and with pericardial constriction. It may be caused by tuberculosis (see below), multiple attacks of acute idiopathic pericarditis, radiation, traumatic pericarditis, renal failure, scleroderma, and neoplasms. The heart is generally enlarged, and a paradoxical pulse is usually present. After pericardiocentesis, the physiologic findings may change from those of cardiac tamponade to those of pericardial constriction. Furthermore, the intrapericardial pressure and the central venous pressure may decline, but not to normal. The diagnosis can be established by pericardiocentesis followed by pericardial biopsy. Wide excision of both the visceral and parietal pericardium is usually effective therapy.

Tuberculous Pericardial Disease This chronic infection is a common cause of chronic pericardial effusion, especially in the developing world where active tuberculosis and HIV are endemic. Tuberculous pericarditis may present as pericardial effusion, chronic constrictive pericarditis, or subacute effusive-constrictive pericarditis (see above). The clinical picture is that of a chronic, systemic illness in a patient with pericardial effusion. It is important to consider this diagnosis in a patient with known tuberculosis, with HIV, and with fever, chest pain, weight loss, and enlargement of the cardiac silhouette of undetermined origin. If the etiology of chronic pericardial effusion remains obscure despite detailed analysis including culture of the pericardial fluid, a pericardial biopsy, preferably by a limited thoracotomy, should be performed. If definitive evidence is still lacking but the specimen shows granulomas with caseation, antituberculous chemotherapy **(Chap. 183)** is indicated.

If the biopsy specimen shows a thickened pericardium after 2–4 weeks of antituberculous therapy, pericardiectomy should be performed to prevent the development of constriction. Tubercular cardiac constriction should be treated surgically while the patient is receiving antituberculous chemotherapy.

FURTHER READING

Arula S, Madsen N: Management of acute pericarditis. Curr Opin Caradiol 38:364, 2023.

Bayes-Coenis A et al: Colchicine in pericarditis. Eur Heart J 38:1706, 2017.

Garcia MJ: Constrictive pericarditis versus restrictive cardiomyopathy? J Am Coll Cardiol 67:2061, 2016.

Imazio M et al: New developments in the management of recurrent pericarditis. Can J Cardiol 2023.

Janus SE, Hoit BD: Effusive-constrictive pericarditis in the spectrum of pericardial comprehensive syndromes. Heart 15:heartjnl-2020-316664, 2021.

LeWinter MM: Acute pericarditis. N Engl J Med 371:2410, 2014.

Mircanda WR, Oh JK: Effusive-constrictive pericarditis. Cardiol Clin 3:551, 2017.

Vistarini N et al: Pericardiectomy for constrictive pericarditis. Ann Thorac Surg 100:107, 2015.

Welch TD: Constrictive pericarditis: Diagnosis, management, and clinical outcomes. Heart 104:725, 2018.

282

Atrial Myxoma and Other Cardiac Tumors

Stephen Tsaur, Eric H. Awtry

Cardiac tumors can be broadly classified into those that arise primarily in the heart and those that reflect metastatic disease from a distant primary source. Primary cardiac tumors can be further divided into those that are pathologically benign and those that are malignant.

TABLE 282-1 Imaging Modalities and Their Utility in the Evaluation of Cardiac Tumors

MODALITY	UTILITY IN CARDIAC TUMOR EVALUATION
Transthoracic echocardiography (TTE) (including two-dimensional, three-dimensional, and contrast)	Assessment of tumor location and size and its impact on adjacent structures (e.g., valves, pericardium).
Transesophageal echocardiography (TEE)	Improved tumor characterization and spatial resolution compared with TTE. May aid in determining surgical approach.
Cardiac magnetic resonance imaging (MRI) with gadolinium contrast	Improved tissue characterization, definition of tumor size, and identification of local invasion when compared with TTE or TEE. May differentiate tumor from thrombus.
Gated cardiac computed tomography (CT)	Provides anatomic assessment and tissue characterization of the tumor. Useful when patients cannot tolerate MRI or when MRI is not feasible (e.g., patients with implantable cardiac devices). Allows for better assessment of calcified lesions and evaluation of extracardiac tumor involvement.
Nuclear imaging (including ^{18}F-fluorodeoxyglucose positron emission tomography [FDG-PET])	Definition of extracardiac disease. May be useful in diagnosis of certain cardiac tumors (e.g., neuroendocrine tumors), but assessment of smaller tumors may be limited by surrounding myocardial FDG uptake.

Overall, primary cardiac tumors are relatively uncommon, whereas secondary involvement of the heart or pericardium occurs in as many as 20% of patients with end-stage metastatic cancer. While patients with cardiac tumors may present with a variety of symptoms, many patients are asymptomatic at the time of diagnosis, with the tumor being identified incidentally on imaging studies performed for other reasons. Cardiac tumors need to be differentiated from other cardiac masses such as vegetation, thrombus, inflammatory myofibroblastic tumors, or myocardial hypertrophy. Echocardiography is usually the initial imaging modality used to evaluate cardiac tumors; however, a variety of imaging modalities are now available, and a multimodality approach is often necessary for accurate diagnosis and clarification of treatment options **(Table 282-1)**. When diagnosis is uncertain after echocardiography, cardiac magnetic resonance imaging (MRI) is often useful to help further differentiate lesions. Imaging characteristics, including size, location, presence of myocardial infiltration, and first pass perfusion of gadolinium contrast agents, can help differentiate benign from malignant tumors. However, tissue histology remains the gold standard for diagnosis of cardiac tumors.

■ PRIMARY TUMORS

Primary tumors of the heart are rare, occurring in ~1 in 2000 patients in autopsy series. Approximately three-quarters are histologically benign, the majority of which are myxomas. Malignant tumors, almost all of which are sarcomas, account for 25% of primary cardiac tumors. All cardiac tumors, regardless of pathologic type, have the potential to cause life-threatening complications. Many tumors are now surgically curable; thus, early diagnosis is imperative.

Clinical Presentation Cardiac tumors may present with a wide array of cardiac and noncardiac manifestations. These manifestations, which depend in large part on the location and size of the tumor as well as its impact on surrounding cardiac structures, are often nonspecific features of more common forms of heart disease, and include chest pain, syncope, congestive heart failure (CHF), murmurs, arrhythmias, conduction disturbances, pericardial effusion, and pericardial tamponade. Additionally, embolic phenomena and constitutional symptoms may occur.

Myxoma Myxomas are the most common type of primary cardiac tumor in adults, accounting for one-third to one-half of all cases at postmortem examination, and approximately three-quarters of the tumors treated surgically. They occur at all ages most commonly in the third through sixth decades, with a female predilection. Approximately 90% of myxomas are sporadic; the remainder are familial with autosomal dominant transmission. The familial variety often occurs as part of a syndrome complex (Carney complex) that includes (1) myxomas (cardiac, skin, and/or breast), (2) lentigines and/or pigmented nevi, and (3) endocrine overactivity (primary nodular adrenal cortical disease with or without Cushing's syndrome, testicular tumors, and/or pituitary adenomas with gigantism or acromegaly). The genetic basis of this complex has not been elucidated completely; however, inactivating mutations in the tumor-suppressor gene *PRKAR1A*, which encodes the protein kinase A type I-α regulatory subunit, have been identified in ~70% of patients with Carney complex.

Pathologically, myxomas are gelatinous structures that consist of myxoma cells embedded in a stroma rich in glycosaminoglycans. Most sporadic tumors are solitary, arise from the interatrial septum in the vicinity of the fossa ovalis (particularly in the left atrium), and are often pedunculated on a fibrovascular stalk. In contrast, familial or syndromic tumors tend to occur in younger individuals, are often multiple, may be ventricular in location, and are more likely to recur after initial resection.

Myxomas commonly present with obstructive signs and symptoms. The most common clinical presentation mimics that of mitral valve disease: either stenosis owing to tumor prolapse into the mitral orifice or regurgitation resulting from tumor-induced valvular trauma or distortion. Ventricular myxomas may cause outflow tract obstruction similar to that caused by subaortic or subpulmonic stenosis. The symptoms and signs of myxoma may be sudden in onset or positional in nature, owing to the effects of gravity on tumor position. A characteristic low-pitched sound, a "tumor plop," may be appreciated on auscultation during early or mid-diastole and is thought to result from the impact of the tumor against the mitral valve or ventricular wall. Myxomas also may present with peripheral or pulmonary embolic phenomenon (resulting from embolization of tumor fragments or tumor-associated thrombus) or with constitutional signs and symptoms, including fever, weight loss, cachexia, malaise, arthralgias, rash, digital clubbing, and Raynaud's phenomenon. These constitutional symptoms are likely the result of cytokines (e.g., interleukin 6) secreted by the myxoma. Laboratory abnormalities, such as hypergammaglobulinemia, anemia, polycythemia, leukocytosis, thrombocytopenia or thrombocytosis, elevated erythrocyte sedimentation rate, and elevated C-reactive protein level are often present. These features account for the frequent misdiagnosis of patients with myxomas as having endocarditis, collagen vascular disease, or a paraneoplastic syndrome.

Two-dimensional and three-dimensional transthoracic and/or transesophageal echocardiography are useful in the diagnosis of cardiac myxoma and allow for assessment of tumor size and determination of the site of tumor attachment, both of which are important considerations in the planning of surgical excision **(Fig. 282-1)**.

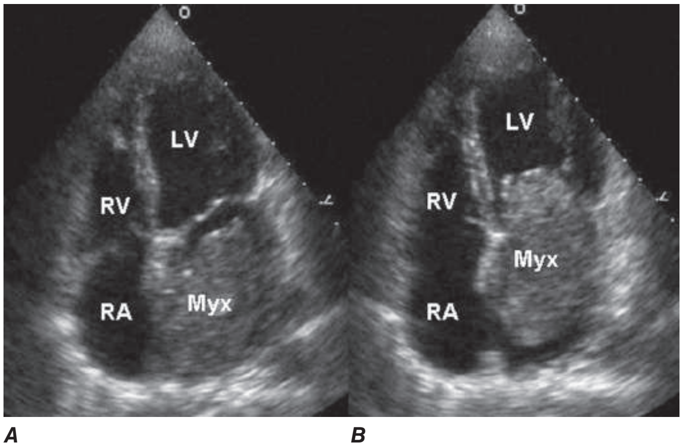

FIGURE 282-1 Transthoracic echocardiogram demonstrating a large atrial myxoma. The myxoma (Myx) fills the entire left atrium in systole (***A***) and prolapses across the mitral valve and into the left ventricle (LV) during diastole (***B***). RA, right atrium; RV, right ventricle. *(Courtesy of Dr. Michael Tsang; with permission.)*

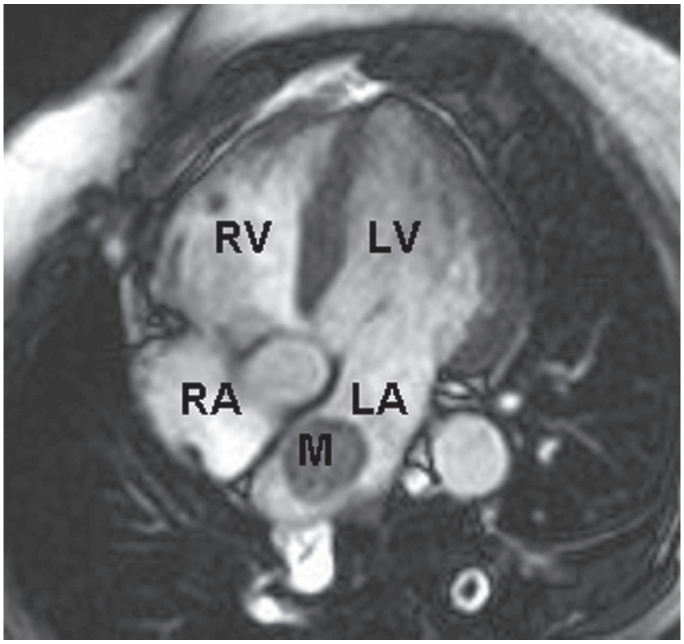

FIGURE 282-2 Cardiac magnetic resonance imaging demonstrating a rounded mass (M) within the left atrium (LA). Pathologic evaluation at the time of surgery revealed it to be an atrial myxoma. LV, left ventricle; RA, right atrium; RV, right ventricle.

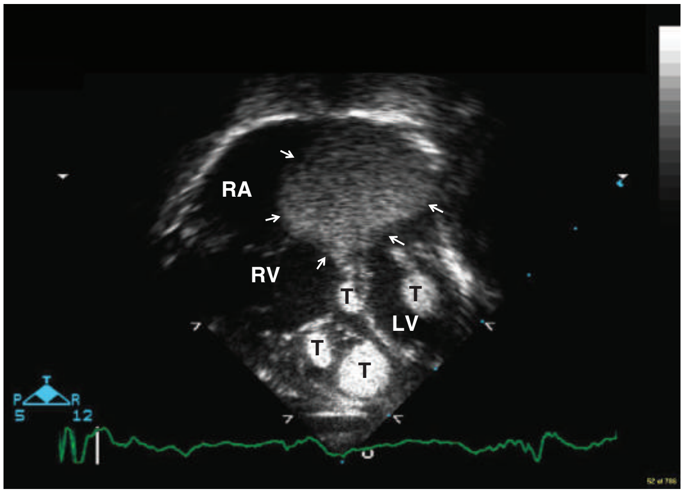

FIGURE 282-3 Transthoracic echocardiogram revealing multiple tumors (T) consistent with rhabdomyomas in a 1-day-old infant. The largest tumor (*arrows*) was located in the left antrioventricular groove and measured 2 cm × 2 cm. LV, left ventricle; RA, right atrium; RV, right ventricle.

Computed tomography (CT) and MRI may provide important additional information regarding size, shape, composition, and surface characteristics of the tumor **(Fig. 282-2)**.

Although cardiac catheterization and angiography were previously performed routinely before tumor resection, they no longer are considered mandatory when adequate noninvasive information is available and other cardiac disorders (e.g., coronary artery disease) are not considered likely. Additionally, catheterization of the chamber from which the tumor arises carries the risk of tumor embolization. Because myxomas may be familial, echocardiographic screening of first-degree relatives is appropriate, particularly if the patient is young and has multiple tumors or features of a myxoma syndrome.

TREATMENT

Myxoma

Surgical excision using cardiopulmonary bypass is indicated regardless of tumor size and is generally curative. Myxomas recur in 12–22% of familial cases but in only 1–2% of sporadic cases. Tumor recurrence most likely results from multifocal lesions in the former setting and incomplete tumor resection in the latter.

Other Benign Tumors Cardiac *lipomas*, although relatively common, are usually incidental findings at postmortem examination; however, they may grow as large as 15 cm, may present as an abnormality of the cardiac silhouette on chest x-ray, and should be resected if they produce symptoms owing to mechanical interference with cardiac function, arrhythmias, or conduction disturbances. *Papillary fibroelastomas* are friable tumors with frond-like projections that are usually solitary and are the most common tumors of the cardiac valves. Although usually clinically silent, they can cause valve dysfunction and may embolize distally, resulting in transient ischemic attacks, stroke, or myocardial infarction. In general, these tumors should be resected even when asymptomatic, although a more conservative approach may be considered for small, right-sided lesions. *Rhabdomyomas* and *fibromas* are the most common cardiac tumors in infants and children and usually occur in the ventricles, where they may produce mechanical obstruction to blood flow, thereby mimicking valvular stenosis, CHF, restrictive or hypertrophic cardiomyopathy, or pericardial constriction. Rhabdomyomas are probably hamartomatous growths, are multiple in 90% of cases, occur in ~50% of children with tuberous sclerosis, and are associated with mutations in the tumor-suppressor genes *TSC1* and *TSC2* **(Fig. 282-3)**. These tumors have a tendency to regress completely or partially; only tumors that cause obstruction require surgical resection. Fibromas are usually single, universally ventricular in location, often calcified, and may be associated with mutations in the tumor-suppressor gene *PTCH1*. Fibromas tend to grow and cause arrhythmias and obstructive symptoms and should be completely resected when possible. *Paragangliomas* are rare chromaffin cell tumors that represent extra-adrenal pheochromocytomas. Most are located in the roof of the left atrium and can be identified on cardiac CT or MRI or with nuclear scanning using ^{131}I-metaiodobenzylguanidine. They are highly vascular and may be hormonally active, resulting in uncontrolled hypertension. Extensive surgical resection is usually required. *Hemangiomas* and *mesotheliomas* are generally small tumors, most often intramyocardial in location, and may cause atrioventricular (AV) conduction disturbances and even sudden death as a result of their propensity to develop in the region of the AV node. Other benign tumors arising from the heart include *teratoma*, *chemodectoma*, *neurilemoma*, and *granular cell myoblastoma*.

Malignant Tumors Almost all malignant primary cardiac tumors are sarcomas, which may be of several histologic types; angiosarcomas are the most common type in adults, whereas rhabdomyosarcomas are the most common type in children. In general, sarcomas are characterized by rapid progression that culminates in the patient's death within weeks to months from the time of presentation as a result of hemodynamic compromise, local invasion, or distant metastases. Almost one-third are metastatic at the time of initial diagnosis, usually involving the lungs. Sarcomas commonly involve the right side of the heart, are rapidly growing, frequently invade the pericardial space, and may obstruct the cardiac chambers or venae cavae. Sarcomas also may occur on the left side of the heart and may be mistaken for myxomas. Isolated cardiac lymphomas have been rarely described but more commonly occur in the context of systemic disease. They are more common in men and in the elderly; usually involve the right heart; may present with arrhythmias, syncope, CHF, or constitutional symptoms; and are usually of the large B-cell type.

TREATMENT

Malignant Tumors

The optimal therapy for cardiac sarcoma is complete resection, often with neoadjuvant and postoperative chemotherapy; however, at the time of presentation, many of these tumors have spread too extensively to allow for surgical excision. Although there are scattered reports of palliation with radiotherapy and/or chemotherapy, the response of cardiac sarcomas to these therapies is generally poor. The one exception appears to be cardiac lymphosarcomas, which may respond to a combination of chemo- and radiotherapy. Primary cardiac lymphoma is the most chemotherapy-sensitive primary cardiac malignancy, with long-term survival achieved in ~40% of treated individuals.

TUMORS METASTATIC TO THE HEART

Metastatic cardiac tumors are much more common than primary cardiac tumors, and their incidence is likely to increase as the life expectancy of patients with various forms of malignant neoplasms is extended by more effective therapy and improved imaging modalities allow earlier identification of metastatic disease. Although cardiac metastases may occur with any tumor type, the relative incidence is especially high in malignant melanoma and, to a somewhat lesser extent, leukemia and lymphoma **(Fig. 282-4)**. In absolute terms, the most common primary sites from which cardiac metastases originate are carcinoma of the breast and lung, reflecting the high incidence of these malignancies. Cardiac metastases almost always occur in the setting of widespread primary disease; most often, there is either primary or metastatic disease elsewhere in the thoracic cavity.

Cardiac metastases may occur via hematogenous or lymphangitic spread or by direct tumor invasion. While they generally manifest as small, firm nodules, diffuse infiltration also may occur, especially with sarcomas or hematologic neoplasms. The pericardium is most often involved, followed by myocardial involvement of any chamber and, rarely, by involvement of the endocardium or cardiac valves.

Cardiac metastases are clinically apparent only ~10% of the time, are usually not the cause of the patient's presentation, and rarely are the cause of death. The vast majority occur in the setting of a previously recognized malignant neoplasm. As with primary cardiac tumors, the clinical presentation reflects more the location and size of the tumor than its histologic type. When symptomatic, cardiac metastases may result in a variety of clinical features, including dyspnea, acute pericarditis, cardiac tamponade, ectopic tachyarrhythmias, heart block, and CHF. Importantly, many of these signs and symptoms may also result from myocarditis, pericarditis, or cardiomyopathy induced by radiotherapy or chemotherapy, and a high index of suspicion for cardiac involvement should be maintained for patients with malignant disease who develop these symptoms.

Electrocardiographic (ECG) findings are nonspecific but may reveal features consistent with pericarditis or may demonstrate low QRS voltage and electrical alternans in the setting of a large pericardial effusion. On chest x-ray, the cardiac silhouette is most often normal but may be enlarged or exhibit a bizarre contour. Echocardiography is useful for identifying and assessing the significance of pericardial effusions and visualizing larger metastases, although CT and radionuclide imaging may define the tumor burden more clearly. Cardiac MRI offers superb image quality and plays a central role in the diagnostic evaluation of cardiac metastases and cardiac tumors in general. Pericardiocentesis may allow for a specific cytologic diagnosis in patients with malignant pericardial effusions with a reported sensitivity of 67–92%. Angiography is rarely necessary but may help to delineate discrete myocardial lesions.

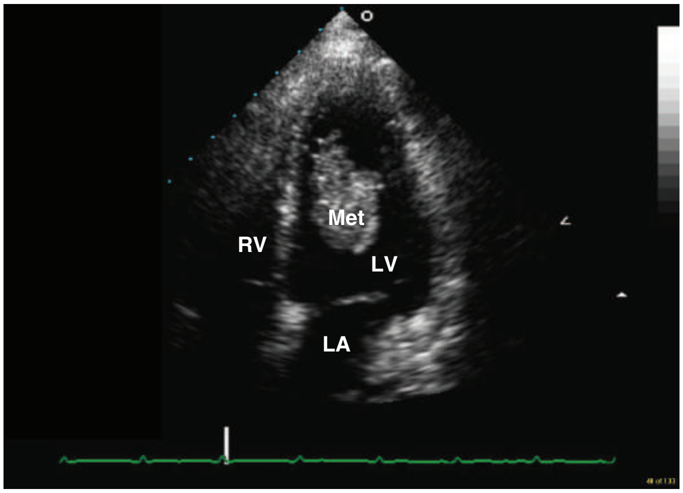

FIGURE 282-4 Large metastatic lesion (Met) in the left ventricle (LV) of a patient with diffusely metastatic bladder cancer. The mass arose from the interventricular septum and prolapsed into the aortic outflow tract during systole.

TREATMENT

Tumors Metastatic to the Heart

Most patients with cardiac metastases have advanced malignant disease; thus, therapy is generally palliative and consists of controlling symptoms and treatment of the primary tumor. Symptomatic malignant pericardial effusions should be drained by pericardiocentesis. For patients with refractory or recurrent malignant pericardial effusion, the surgical creation of a pericardial window allowing for drainage of the effusion to the adjacent pleural or peritoneal space may prevent recurrent pericardial tamponade. Prolonged drainage (3–5 days) and concomitant instillation of a sclerosing agent (e.g., bleomycin) can be considered as a palliative measure in terminally ill patients. Given the overall poor prognosis of these patients, discussions regarding goals of care and involvement of palliative care services are often appropriate.

FURTHER READING

Bussani R et al: Cardiac tumors: Diagnosis, prognosis and treatment. Curr Cardiol Rep 22:169, 2020.

Mousavi N et al: Assessment of cardiac masses by cardiac magnetic resonance imaging: Histological correlation and clinical outcomes. J Am Heart Assoc 8:e007829, 2019.

Shapira O et al: Tumors of the heart, in *Sabiston and Spenser Surgery of the Chest*, 9th ed, FW Sellke et al (eds). Philadelphia, Elsevier, 2016, pp 1849–1857.

Tamin SS et al: Prognostic and bioepidemiologic implications of papillary fibroelastomas. J Am Coll Cardiol 65:2420, 2015.

Young PM et al: Computed tomography imaging of cardiac masses. Radiol Clin N Am 57:75, 2019.

283

Cardiac Trauma

Alexandra R. Pipilas, Eric H. Awtry

CARDIAC TRAUMA

Traumatic cardiac injury may be caused by either penetrating or nonpenetrating trauma; the latter is often referred to as blunt cardiac injury (BCI). *Penetrating injuries* most often result from gunshot or knife wounds, and the site of entry is usually obvious. *Blunt cardiac injuries* most often occur during motor vehicle accidents, either from rapid deceleration or from impact of the chest, but can also result from falls from heights, crush injuries, blast injuries, violent assault, or significant physical contact during sporting events. Importantly, rapid deceleration following motor vehicle accidents may be associated with significant cardiac injury even in the absence of external signs of thoracic trauma.

BLUNT CARDIAC INJURY

Myocardial contusion is a nonspecific term that has been used to describe a broad spectrum of nonpenetrating cardiac injuries that

TABLE 283-1 Spectrum of Cardiac Abnormalities Following Blunt Cardiac Injury

ABNORMALITY	COMMENTS
ECG abnormalities	Sinus tachycardia, RBBB, heart block, ST-T wave abnormalities, atrial and ventricular arrhythmias
Elevated cardiac biomarkers	Troponin I or T are most specific
Focal wall motion abnormality or hematoma	Most commonly involving RV free wall, LV apex, and interventricular septum
Valvular insufficiency	Most commonly involving mitral and tricuspid valves and occasionally the aortic valve
Myocardial rupture	Ventricular septal defect or free wall rupture
Coronary artery injury	Most commonly involving the LAD; usually presents as STEMI
Pericardial effusion and tamponade	Resulting from free wall rupture or coronary artery laceration

Abbreviations: ECG, electrocardiogram; LAD, left anterior descending coronary artery; LV, left ventricle; RBBB, right bundle branch block; RV, right ventricle; STEMI, ST-segment elevation myocardial infarction.

result in abnormalities on electrocardiogram (ECG), elevation in cardiac biomarkers, and acute structural cardiac abnormalities **(Table 283-1)**. Importantly, the cardiac injury may initially be overlooked in trauma patients as the clinical focus is directed toward other, more obvious injuries. Unfortunately, there is no one sign or symptom that confirms the diagnosis of BCI, and clinical, laboratory, and radiographic findings may be nonspecific in the setting of significant trauma. The physical examination may be challenging in the setting of chest wall injury; however, patients should be carefully examined to detect pericardial rubs, cardiac murmurs, and evidence of pericardial tamponade **(Chap. 281)**. The mechanism of injury and the presence of other chest trauma should be considered when determining the index of suspicion for BCI; however, there is no proven association between sternal or rib fractures and the presence of BCI, and significant cardiac injury may be present in the absence of chest wall abnormalities.

Chest pain is common following thoracic trauma, and while it can indicate cardiac ischemia or pericardial injury, it often reflects musculoskeletal trauma. Nonetheless, myocardial necrosis may occur as a direct result of the blunt injury or as a result of traumatic coronary laceration, dissection, or thrombosis. The injured myocardium is pathologically similar to infarcted myocardium and may be associated with atrial or ventricular arrhythmias, conduction disturbances including bundle branch and atrioventricular (AV) nodal block, or abnormalities on ECG resembling those of infarction or pericarditis. Thus, it is important to obtain an ECG in all patients presenting with chest trauma.

Serum creatine kinase, myocardial band (CK-MB) isoenzyme levels are increased in ~20% of patients who experience blunt chest trauma but may be falsely elevated in the presence of massive skeletal muscle injury and should not be relied upon to confirm the diagnosis of BCI in the setting of trauma. Cardiac troponin levels are more specific for identifying cardiac damage; patients with normal serial troponin levels after chest trauma are very unlikely to have sustained cardiac injury. When combined with a normal ECG, a normal troponin level at 6–8 h after chest trauma essentially excludes BCI. Echocardiography is the most useful modality for the detection of structural and functional sequelae of BCI, including regional wall motion abnormalities (most commonly involving the right ventricle, interventricular septum, or left ventricular apex), pericardial effusion, valvular dysfunction, and ventricular or ventricular septal rupture. A transthoracic echocardiogram (TTE) should be performed in all patients with suspected BCI, especially in those with an abnormal ECG, elevated troponin, or hemodynamic instability; transesophageal echocardiography should be considered for patients in whom adequate TTE images cannot be obtained.

Traumatic rupture of the cardiac valves or their supporting structures, most commonly of the tricuspid or aortic valves, leads to acute valvular incompetence. This complication is usually heralded by the development of a loud murmur, may be associated with rapidly progressive heart failure, and can be diagnosed by either TTE or transesophageal echocardiography. BCI may also result in dissection, occlusion, or laceration of the coronary arteries. ST elevation may be seen on the ECG, and prompt intervention is often required.

The most serious consequence of nonpenetrating cardiac injury is myocardial rupture. Free wall rupture may result in hemopericardium and tamponade while a ventricular septal rupture can result in significant intracardiac shunting. Although generally fatal, up to 40% of patients with cardiac rupture have been reported to survive long enough to reach a specialized trauma center. Hemopericardium also may result from traumatic rupture of a pericardial vessel or a coronary artery. Additionally, pericarditis and/or pericardial effusion may develop weeks or even months after blunt chest trauma as a manifestation of the post–cardiac injury syndrome, an inflammatory condition that resembles the postpericardiotomy syndrome **(Chap. 281)**.

Blunt, nonpenetrating, often innocent-appearing injuries to the chest may trigger ventricular fibrillation even in the absence of structural myocardial damage. This syndrome, referred to as *commotio cordis*, occurs most often in adolescents during sporting events (e.g., baseball, hockey, football, and lacrosse) and is an electrical phenomenon that probably results from an impact to the chest wall overlying the heart during the susceptible phase of repolarization (just before the peak of the T wave). Survival depends on prompt defibrillation. Sudden emotional or physical trauma, even in the absence of direct cardiac trauma, may precipitate a transient catecholamine-mediated cardiomyopathy referred to as *takotsubo syndrome* or *apical ballooning syndrome* **(Chaps. 266–270)**.

Rupture or transection of the aorta, usually just above the aortic valve or at the site of the tethering ligamentum arteriosum, is a common consequence of nonpenetrating chest trauma and is the most common vascular deceleration injury. The clinical presentation may be similar to that of aortic dissection **(Chap. 291)**; the arterial pressure and pulse amplitude may be increased in the upper extremities and decreased in the lower extremities, and chest x-ray may reveal mediastinal widening. Aortic rupture into the left thoracic space is almost universally fatal; however, the rupture may occasionally be contained by the aortic adventitia, resulting in a false, or *pseudo*-, aneurysm that may be discovered months or years after the initial injury.

TREATMENT

Blunt Cardiac Injury

The treatment of BCI depends on the specific injury sustained. Hemodynamically stable patients with a normal ECG and normal serial troponin levels are at low risk for BCI and usually do not require hospital admission for cardiac issues. Patients with an abnormal ECG, including those with conduction disturbances, and/or elevated troponin but normal echocardiogram usually warrant 24–48 h of telemetry monitoring; however, other specific cardiac treatment is not usually required in the absence of the development of arrhythmias, as conduction disturbances are often transient. Patients with mechanical complications (acute valvular insufficiency, myocardial rupture) require urgent operative correction.

PENETRATING CARDIAC INJURY

Penetrating injuries of the heart produced by knife or bullet wounds usually result in rapid clinical deterioration and frequently in death as a result of hemopericardium/pericardial tamponade or massive hemorrhage. Nonetheless, up to half of such patients may survive long enough to reach a specialized trauma center if immediate resuscitation is performed. Prognosis in these patients relates to the mechanism of injury, the specific cardiac chamber(s) involved, and their clinical condition at presentation. In general, gunshot wounds are associated with a higher mortality than are knife wounds. Twenty percent of shooting victims survive versus up to 65% of stabbing victims. This is likely in part because ballistic wounds are more frequently associated with multichamber cardiac injury. As a result of its anterior position in the chest, the right ventricle (RV) is the most frequently injured cardiac chamber,

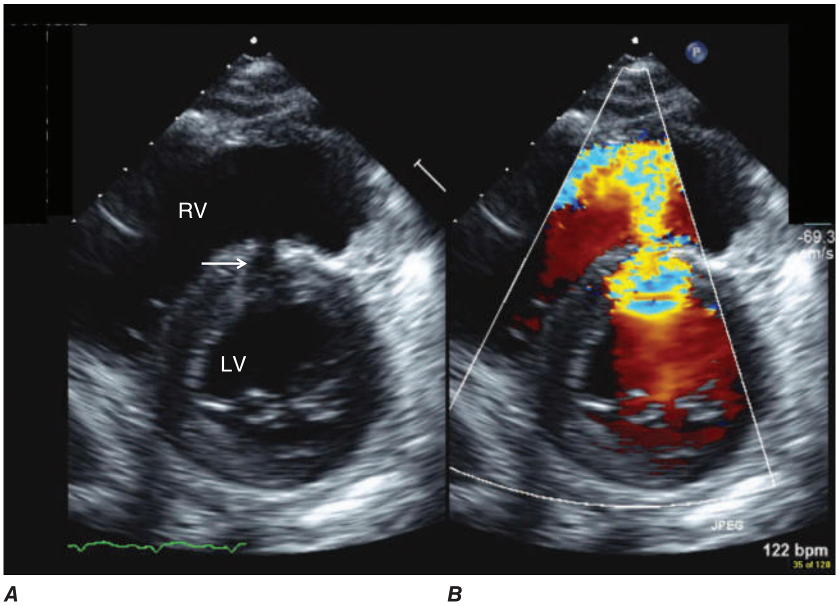

FIGURE 283-1 Transthoracic echocardiogram demonstrating a traumatic ventricular septal defect. The patient underwent emergent repair of the right ventricle following a self-inflicted stab wound to the chest. Subsequent two-dimensional imaging (***A***) revealed a laceration of the interventricular septum (*arrow*) with color flow Doppler (***B***) demonstrating prominent left-to-right shunting across the defect. LV, left ventricle; RV, right ventricle.

followed by the left ventricle (LV); isolated atrial injury is uncommon. Some studies suggest that RV injuries may be associated with a better prognosis than LV injuries, and most reports indicate that multichamber involvement carries a worse prognosis than single-chamber injury. Patients who are in hemodynamic collapse at presentation to the emergency department have a particularly poor prognosis with a mortality rate approaching 90%, whereas ~75% of patients who are stable enough to be brought to the operating room will survive.

Cardiac perforation of the right atrium, the RV free wall, or the interventricular septum may occur as a complication of cardiac procedures including placement of central venous/intracardiac catheters, insertion of pacemaker/defibrillator leads, or performance of RV endomyocardial biopsies; and coronary arterial perforation can occur during deployment of intracoronary stents. These iatrogenic injuries are associated with a better prognosis than are other forms of penetrating cardiac trauma, likely related to a more limited degree of cardiac injury and the rapid availability of corrective therapies.

Traumatic rupture of a great vessel from penetrating injury is usually associated with hemothorax and, less often, hemopericardium, both of which are associated with significant mortality. Local hematoma formation may compress adjacent vessels and produce ischemic symptoms, and arteriovenous fistulas may develop, occasionally resulting in high-output heart failure.

Some patients with penetrating chest injuries are hemodynamically stable at presentation and without symptoms to suggest cardiac injury; however, as many as 20% of these patients will have occult penetrating cardiac trauma. As a result, there should always be a high index of suspicion for cardiac injury in any patient with penetrating chest trauma, irrespective of clinical stability. TTE should be performed in all of these patients to assess for the presence of pericardial effusion or hematoma.

Occasionally, patients who survive penetrating cardiac injuries may subsequently present days or weeks later with a new cardiac murmur or heart failure as a result of mitral or tricuspid regurgitation or an intracardiac shunt (i.e., ventricular or atrial septal defect, aortopulmonary fistula, or coronary arteriovenous fistula) that was undetected at the time of the initial injury or developed subsequently **(Fig. 283-1)**. Therefore, trauma patients should be examined carefully several weeks after the injury. If a mechanical complication is suspected, it can be confirmed by echocardiography or cardiac catheterization.

TREATMENT

Penetrating Cardiac Injury

Penetrating cardiac injury associated with hemodynamic instability is a surgical emergency and requires immediate resuscitative measures including endotracheal intubation, establishment of large-bore intravenous access to facilitate massive volume resuscitation, and immediate thoracotomy to allow for pericardial drainage and repair of cardiac injuries. Occasionally, cross-clamping of the descending aorta is required to perfuse the heart and brain preferentially until hemodynamic stability can be achieved. Hemodynamically stable patients in whom echocardiography reveals even a small pericardial effusion require urgent surgical exploration to evaluate for occult cardiac perforation. Pericardiocentesis may be lifesaving in patients with tamponade but is usually only a temporizing measure while awaiting definitive surgical therapy. In some survivors of penetrating cardiac injury, the pericardial hemorrhage predisposes to the development of constriction **(Chap. 281)**, which may require surgical decortication.

FURTHER READING

Ali H et al: Clinical and electrocardiographic features of complete heart block after blunt cardiac injury: A systematic review of the literature. Heart Rhythm 14:1561, 2017.

Biffl WL et at: Diagnosis and management of blunt cardiac injury: What you need to know. J Trauma Acute Care Surg 96:685, 2024.

Crawford T et al: Thoracic trauma, in *Sabiston and Spencer Surgery of the Chest*, 9th ed, FW Sellke et al (eds). Philadelphia, Elsevier, 2016, pp 100–130.

Ismailov RM et al: Trauma associated with cardiac dysrhythmias: Results from a large, matched case-control study. J Trauma 62:1186, 2007.

Morse BC et al: Penetrating cardiac injuries: A 36-year perspective at an urban level 1 trauma center. J Trauma Acute Care Surg 81:623, 2016.

Wu Y et al: Imaging of cardiac trauma. Radiol Clin N Am 57:795, 2019.

Yousef R, Carr JA: Blunt cardiac trauma: A review of the current knowledge and management. Ann Thorac Surg 98:1134, 2014.

Section 5 Coronary and Peripheral Vascular Disease

284

Ischemic Heart Disease

Robert P. Giugliano, Elliott M. Antman, Joseph Loscalzo

Ischemic heart disease (IHD) is a condition in which there is an inadequate supply of blood and oxygen to a portion of the myocardium; it typically occurs when there is an imbalance between myocardial oxygen supply and demand. The most common cause of myocardial ischemia is atherosclerotic disease of an epicardial coronary artery (or arteries) sufficient to cause a regional reduction in myocardial blood flow and inadequate perfusion of the myocardium supplied by the involved coronary artery. This chapter focuses on the chronic manifestations and treatment of IHD (sometimes referred to as chronic coronary disease or chronic coronary syndrome), while the subsequent chapters address the acute phases of IHD.

EPIDEMIOLOGY AND GLOBAL TRENDS

IHD causes more deaths and disability and incurs greater economic costs than any other illness in the developed world. IHD is the most common, serious, chronic, life-threatening illness in the United States, where 20.5 million persons have IHD. Although there is regional variation, ~3–4% of the population has sustained a myocardial infarction. Genetic factors, a high-fat and energy-rich diet, smoking, and a sedentary lifestyle are associated with the emergence of IHD. In the United States and Western Europe, IHD is growing among low-income groups, but primary prevention has delayed the disease to later in life across socioeconomic groups. Despite these sobering statistics, it is worth noting that epidemiologic data show a decline in the rate of deaths due to IHD, about half of which is attributable to treatments and half to prevention by risk factor modification.

Obesity, insulin resistance, and type 2 diabetes mellitus are increasing and are powerful risk factors for IHD. These trends are occurring in the general context of population growth and as a result of the increase in the average age of the world's population. With urbanization in countries with emerging economies and a growing middle class, elements of the energy-rich Western diet are being adopted. As a result, the prevalence of risk factors for IHD and the prevalence of IHD itself are both increasing rapidly, so that in analyses of the global burden of disease, there is a shift from communicable to noncommunicable diseases, and it is estimated that globally over 200 million people live with IHD. Population subgroups that appear to be particularly affected are men in South Asian countries, especially India and the Middle East. IHD is a major contributor to the number of disability-adjusted life-years (DALYs) experienced globally.

PATHOPHYSIOLOGY

Central to an understanding of the pathophysiology of myocardial ischemia is the concept of myocardial supply and demand. In normal conditions, for any given level of a demand for oxygen, the myocardium will control the supply of oxygen-rich blood to prevent underperfusion of myocytes and the subsequent development of ischemia and infarction. The major determinants of myocardial oxygen demand (MVO_2) are heart rate, myocardial contractility, and myocardial wall tension (stress). An adequate supply of oxygen to the myocardium requires a satisfactory level of oxygen-carrying capacity of the blood (determined by the inspired level of oxygen, pulmonary function, and hemoglobin concentration and function) and an adequate level of coronary blood flow. Blood flows through the coronary arteries in a phasic fashion, with the majority occurring during diastole. About 75% of the total coronary resistance to flow occurs across three sets of arteries: (1) large epicardial arteries (Resistance 1 = R_1), (2) prearteriolar vessels (R_2), and (3) arteriolar and intramyocardial capillary vessels (R_3). In the absence of significant flow-limiting atherosclerotic obstructions, R_1 is trivial; the major determinant of coronary resistance is found in R_2 and R_3 **(Fig. 284-1)**. The normal coronary circulation is dominated and controlled by the heart's requirements for oxygen. This need is met by the ability of the coronary vascular bed to vary its resistance (and, therefore, blood flow) considerably while the myocardium extracts a high and relatively fixed percentage of oxygen. Normally, intramyocardial resistance vessels demonstrate a great capacity for dilation (R_2 and R_3 decrease). The changing oxygen needs of the heart with exercise and emotional stress affect coronary

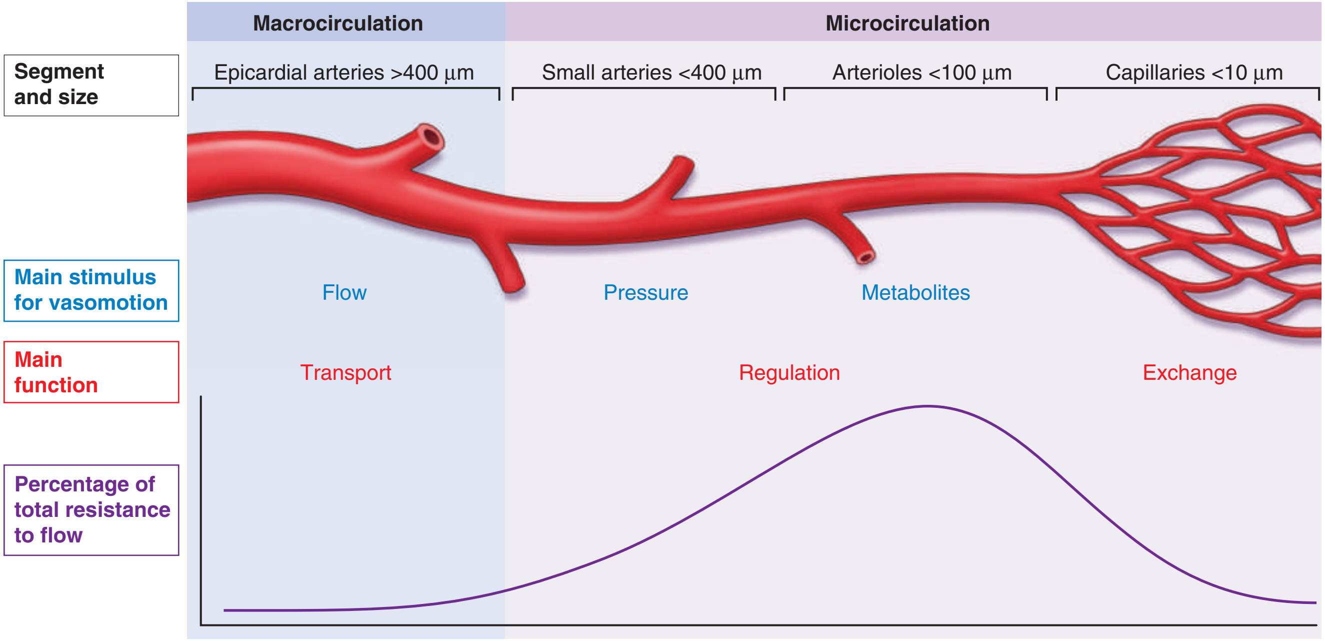

FIGURE 284-1 Macrocirculation and microcirculation across segments and sizes of the arteries. The location and size of the arteries supplying blood to the heart is shown at the top. Vasomotion of the arterial segments occurs in response to the stimuli shown. The main function of each of the arterial segments is shown next, followed by a depiction of the relative resistance to antegrade flow. (*Adapted from J Knuuti et al: 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes. Eur Heart J 41:407, 2020/.*)

vascular resistance and, in this manner, regulate the supply of oxygen and substrate to the myocardium (*metabolic regulation*). The coronary resistance vessels also adapt to physiologic alterations in blood pressure to maintain coronary blood flow at levels appropriate to myocardial needs (*autoregulation*).

By reducing the lumen of the coronary arteries, atherosclerosis limits appropriate increases in perfusion when the demand for more coronary flow occurs. When the luminal reduction is severe, myocardial perfusion in the basal state is reduced. Coronary blood flow also can be limited by spasm (see vasospastic angina in **Chap. 285**), arterial thrombi, and, rarely, coronary emboli as well as by ostial narrowing due to aortitis. Congenital abnormalities such as the origin of the left anterior descending coronary artery from the pulmonary artery may cause myocardial ischemia and infarction in infancy, but this cause is very rare in adults.

Myocardial ischemia also can occur if myocardial oxygen demands are markedly increased and particularly when coronary blood flow may be limited, as occurs in severe left ventricular hypertrophy (LVH) due to aortic stenosis. The latter can present with angina that is indistinguishable from that caused by coronary atherosclerosis largely owing to subendocardial ischemia **(Chap. 272)**. A reduction in the oxygen-carrying capacity of the blood, as in extremely severe anemia or in the presence of carboxyhemoglobin, rarely causes myocardial ischemia by itself but may lower the threshold for ischemia in patients with moderate coronary obstruction.

Not infrequently, two or more causes of ischemia coexist in a patient, such as an increase in oxygen demand due to LVH secondary to hypertension and a reduction in oxygen supply secondary to coronary atherosclerosis and anemia. Abnormal constriction or failure of normal dilation of the coronary resistance vessels also can cause ischemia. When it causes angina, this condition is referred to as *microvascular angina*.

CORONARY ATHEROSCLEROSIS

Epicardial coronary arteries are the major site of atherosclerotic disease. The major risk factors for atherosclerosis (high levels of plasma low-density lipoprotein [LDL], cigarette smoking, hypertension, and diabetes mellitus) vary in their relative impact on disturbing the normal functions of the vascular endothelium. These functions include local control of vascular tone, maintenance of an antithrombotic surface, and control of inflammatory cell adhesion and diapedesis. The loss of these defenses leads to inappropriate constriction, luminal thrombus formation, and abnormal interactions between blood cells, especially monocytes and platelets, and the activated vascular endothelium. Functional changes in the vascular milieu ultimately result in the subintimal collections of fat, smooth muscle cells, fibroblasts, and intercellular matrix that define the atherosclerotic plaque. Rather than viewing atherosclerosis strictly as a vascular problem, it is useful to consider it in the context of alterations in the nature of the circulating blood (hyperglycemia; increased concentrations of LDL cholesterol, tissue factor, fibrinogen, von Willebrand factor, coagulation factor VII, and platelet microparticles). The combination of a "vulnerable vessel" in a patient with "vulnerable blood" promotes a state of hypercoagulability and hypofibrinolysis. This is especially true in patients with diabetes mellitus.

Atherosclerosis develops at irregular rates in different segments of the epicardial coronary tree and leads eventually to segmental reductions in cross-sectional area, i.e., plaque formation. There is also a predilection for atherosclerotic plaques to develop at sites of increased turbulence in coronary flow, such as at branch points in the epicardial arteries. When a stenosis reduces the diameter of an epicardial artery by 50%, there is a limitation of the ability to increase flow to meet increased myocardial demand. When the diameter is reduced by ~80%, blood flow at rest may be reduced, and further minor decreases in the stenotic orifice area can reduce coronary flow dramatically to cause myocardial ischemia at rest or with minimal stress.

Segmental atherosclerotic narrowing of epicardial coronary arteries is caused most commonly by the formation of a plaque, which is subject to rupture or erosion of the cap separating the plaque from the bloodstream. Upon exposure of the plaque contents to blood, two important and interrelated processes are set in motion: (1) platelets are activated and aggregate, and (2) the coagulation cascade is activated, leading to deposition of fibrin strands. A thrombus composed of platelet aggregates and fibrin strands traps red blood cells and can reduce coronary blood flow, leading to the clinical manifestations of myocardial ischemia.

The location of the obstruction influences the quantity of myocardium rendered ischemic and determines the severity of the clinical manifestations. Thus, critical obstructions in vessels, such as the left main coronary artery and the proximal left anterior descending coronary artery, are particularly hazardous. Chronic severe coronary narrowing and myocardial ischemia frequently are accompanied by the development of collateral vessels, especially when the narrowing develops gradually. When well developed, such vessels can by themselves provide sufficient blood flow to sustain the viability of the myocardium at rest but not during conditions of increased demand.

With progressive worsening of a stenosis in a proximal epicardial artery, the distal resistance vessels (when they function normally) dilate to reduce vascular resistance and maintain coronary blood flow. A pressure gradient develops across the proximal stenosis, and poststenotic pressure falls. When the resistance vessels are maximally dilated, myocardial blood flow becomes dependent on the pressure in the coronary artery distal to the obstruction. In these circumstances, ischemia, manifest clinically by angina or electrocardiographically by ST-segment deviation, can be precipitated by increases in myocardial oxygen demand caused by physical activity, emotional stress, and/or tachycardia. Changes in the caliber of the stenosed coronary artery resulting from physiologic vasomotion, loss of endothelial control of dilation (as occurs in atherosclerosis), pathologic spasm (vasospastic angina), or small platelet-rich plugs also can upset the critical balance between oxygen supply and demand and thereby precipitate myocardial ischemia.

■ EFFECTS OF ISCHEMIA

During episodes of inadequate perfusion caused by coronary atherosclerosis, myocardial tissue oxygen tension falls and may cause transient disturbances of the mechanical, biochemical, and electrical functions of the myocardium **(Fig. 284-2)**. Coronary atherosclerosis is a focal process that usually causes nonuniform ischemia. During ischemia, regional disturbances of ventricular contractility cause segmental hypokinesia, akinesia, or, in severe cases, bulging (dyskinesia), which can reduce myocardial pump function.

The abrupt development of severe ischemia, as occurs with total or subtotal coronary occlusion, is associated with near instantaneous failure of normal muscle relaxation and then diminished contraction. The relatively poor perfusion of the subendocardium causes more intense ischemia of this portion of the wall (compared with the subepicardial region). Ischemia of large portions of the ventricle causes transient left ventricular (LV) failure, and if the papillary muscle apparatus is involved, mitral regurgitation can occur. When ischemia is transient, it may be associated with angina pectoris; when it is prolonged, it can lead to myocardial necrosis and scarring with or without the clinical picture of acute myocardial infarction **(Chap. 286)**.

A wide range of abnormalities in cell metabolism, function, and structure underlie these mechanical disturbances during ischemia. The normal myocardium metabolizes fatty acids and glucose to carbon dioxide and water. With severe oxygen deprivation, fatty acids cannot be oxidized, and glucose is converted to lactate; intracellular pH is reduced, as are the myocardial stores of high-energy phosphates, i.e., ATP and creatine phosphate. Impaired cell membrane function leads to the leakage of potassium and the uptake of sodium by myocytes as well as an increase in cytosolic calcium. The severity and duration of the imbalance between myocardial oxygen supply and demand determine whether the damage is reversible (≤20 min for total occlusion in the absence of collaterals) or permanent, with subsequent myocardial necrosis (>20 min).

Ischemia also causes characteristic changes in the electrocardiogram (ECG) such as repolarization abnormalities, as evidenced by inversion

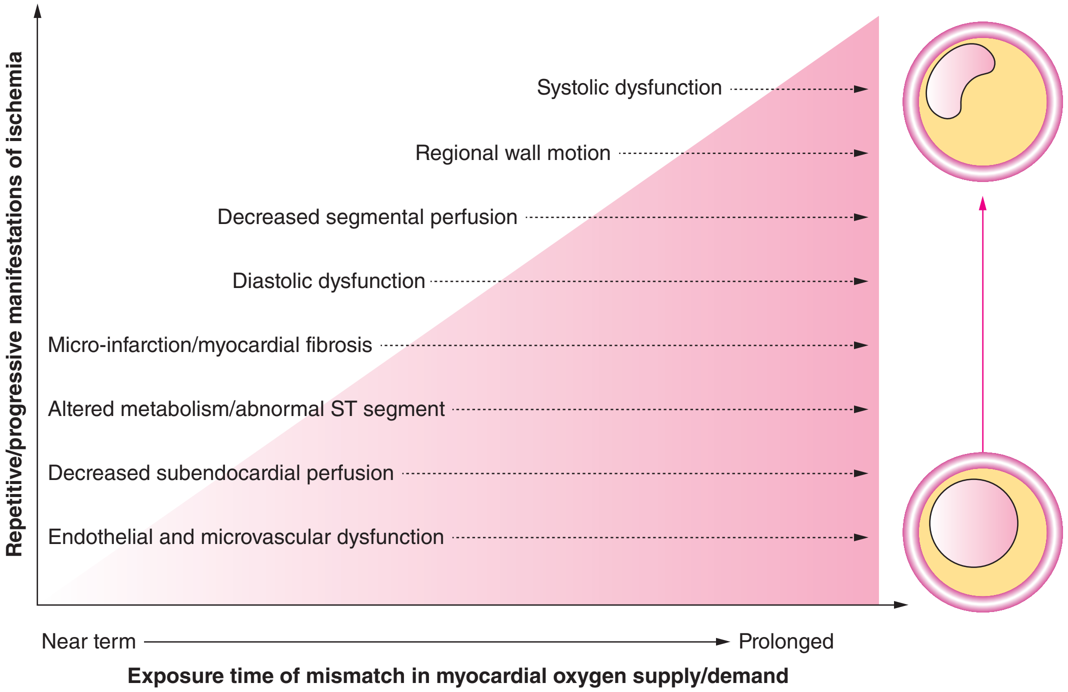
Repetitive/progressive manifestations of ischemia
Systolic dysfunction
Regional wall motion
Decreased segmental perfusion
Diastolic dysfunction
Micro-infarction/myocardial fibrosis
Altered metabolism/abnormal ST segment
Decreased subendocardial perfusion
Endothelial and microvascular dysfunction
Near term
Prolonged
Exposure time of mismatch in myocardial oxygen supply/demand

FIGURE 284-2 Cascade of mechanisms and manifestations of ischemia. (*Reproduced with permission from LJ Shaw et al: Women and ischemic heart disease: Evolving knowledge. J Am Coll Cardiol 54:1561, 2009.*)

of T waves and, when more severe, displacement of ST segments **(Chap. 247)**. Transient T-wave inversion probably reflects nontransmural, intramyocardial ischemia; transient ST-segment depression often reflects patchy subendocardial ischemia; and ST-segment elevation is thought to be caused by more severe transmural ischemia. Another important consequence of myocardial ischemia is electrical instability, which may lead to isolated ventricular premature beats or even ventricular tachycardia or ventricular fibrillation **(Chaps. 261 and 262)**. Most patients who die suddenly from IHD do so as a result of ischemia-induced ventricular tachyarrhythmias **(Chap. 317)**.

■ ASYMPTOMATIC VERSUS SYMPTOMATIC IHD

Although the prevalence is decreasing, postmortem studies of accident victims and military casualties in Western countries show that coronary atherosclerosis can begin before age 20 and is present among adults who were asymptomatic during life. Exercise stress tests in asymptomatic persons may show evidence of silent myocardial ischemia, i.e., exercise-induced ECG changes not accompanied by angina pectoris; coronary angiographic studies of such persons may reveal coronary artery plaques and previously unrecognized obstructions **(Chap. 249)**. Coronary artery calcifications (CACs) may be seen on computed tomography (CT) images of the heart, can be quantified in a CAC score, and may be used as adjunctive information to support a diagnosis of IHD. However, they should not be used as the primary screening modality or as the isolated basis on which to formulate therapeutic decisions. (See further discussion below.) Postmortem examination of patients with such obstructions without a history of clinical manifestations of myocardial ischemia often shows macroscopic scars secondary to myocardial infarction in regions supplied by diseased coronary arteries, with or without collateral circulation. According to population studies, ~25% of patients who survive acute myocardial infarction may not come to medical attention, and these patients have the same adverse prognosis as do those who present with the classic clinical picture of acute myocardial infarction **(Chap. 286)**. Sudden death may be unheralded and is a common presenting manifestation of IHD **(Chap. 317)**.

Patients with IHD also can present with cardiomegaly and heart failure secondary to ischemic damage of the LV myocardium that may have caused no symptoms before the development of heart failure; this condition is referred to as *ischemic cardiomyopathy*. In contrast to the asymptomatic phase of IHD, the symptomatic phase is characterized by chest discomfort due to either angina pectoris or acute myocardial infarction **(Chap. 286)**. Having entered the symptomatic phase, the patient may exhibit a stable or progressive course, revert to the asymptomatic stage, or die suddenly.

STABLE ANGINA PECTORIS

This episodic clinical syndrome is a result of transient myocardial ischemia. Various diseases that cause myocardial ischemia and the numerous forms of discomfort with which it may be confused are discussed in **Chap. 15**. Males constitute ~70% of all patients with angina pectoris and an even greater proportion of those aged <50 years. It is, however, important to note that descriptions of angina pectoris in women may be different from that in men.

■ HISTORY

The typical patient with angina is a man >50 years or a woman >60 years of age who complains of episodes of chest discomfort, usually described as heaviness, pressure, squeezing, smothering, or choking and only rarely as frank pain. When the patient is asked to localize the sensation, they typically place a hand over the sternum, sometimes with a clenched fist, to indicate a squeezing, central, substernal discomfort (Levine's sign). Angina is usually crescendo-decrescendo in nature (typically with the severity of the discomfort not at its most intense level at the outset of symptoms), typically lasts 2–5 min, and can radiate to either shoulder and to both arms (especially the ulnar aspects of the forearm and hand). It also can arise in or radiate to the back, interscapular region, root of the neck, jaw, teeth, and epigastrium. Angina is rarely localized below the umbilicus or above the mandible. A useful finding in assessing a patient with chest discomfort is the fact that myocardial ischemic discomfort does not radiate to the trapezius muscles; that radiation pattern is more typical of pericarditis.

Although episodes of angina typically are caused by exertion (e.g., exercise, hurrying, or sexual activity) or emotion (e.g., stress, anger, fright, or frustration) and are relieved by rest, they also may occur at rest and while the patient is recumbent (angina decubitus). The patient may be awakened at night by typical chest discomfort and dyspnea. Nocturnal angina may be due to episodic tachycardia, diminished oxygenation as the respiratory pattern changes during sleep, or expansion of the intrathoracic blood volume that occurs with recumbency; the latter causes an increase in cardiac size (end-diastolic volume), wall tension, and myocardial oxygen demand that can lead to ischemia and transient LV failure.

The threshold for the development of angina pectoris may vary by time of day and emotional state. Many patients report a fixed threshold for angina, occurring predictably at a certain level of activity, such as climbing two flights of stairs at a normal pace. In these patients, coronary stenosis and myocardial oxygen supply are fixed, and ischemia is precipitated by an increase in myocardial oxygen demand; they are said to have stable exertional angina. In other patients, the threshold for angina may vary considerably within any particular day and from day to day. In such patients, variations in myocardial oxygen supply, most likely due to changes in coronary vasomotor tone, may play an important role in defining the pattern of angina. A patient may report symptoms upon minor exertion in the morning yet by midday be capable of much greater effort without symptoms. Angina may also be precipitated by unfamiliar circumstances, a heavy meal, exposure to cold, or a combination of these factors.

Exertional angina typically is relieved in 1–5 min by slowing or ceasing activities and even more rapidly by rest and sublingual nitroglycerin (see below). Indeed, the diagnosis of angina should be suspect if it does not respond to the combination of these measures. The severity of angina can be conveniently summarized by the Canadian Cardiac Society functional classification **(Table 284-1)**. Its impact on the patient's functional capacity can be described by using the New York Heart Association functional classification (Table 284-1).

Sharp, fleeting chest pain or a prolonged, dull ache localized to the left submammary area is rarely due to myocardial ischemia. However, especially in women and patients with diabetes mellitus, angina pectoris may be atypical in location and not strictly related to provoking factors. In addition, this symptom may exacerbate and remit over days, weeks, or months. Its occurrence can be seasonal, occurring more frequently in the winter in temperate climates. Anginal "equivalents" are symptoms of myocardial ischemia other than angina. They include dyspnea, nausea, fatigue, and faintness and are more common in the elderly and in patients with diabetes mellitus.

TABLE 284-1 Cardiovascular Disease Classification Chart

CLASS	NEW YORK HEART ASSOCIATION FUNCTIONAL CLASSIFICATION	CANADIAN CARDIOVASCULAR SOCIETY FUNCTIONAL CLASSIFICATION
I	Patients have cardiac disease but *without* the resulting *limitations* of physical activity. Ordinary physical activity does not cause undue fatigue, palpitation, dyspnea, or anginal pain.	Ordinary physical activity, such as walking and climbing stairs, *does not cause angina.* Angina present with strenuous or rapid or prolonged exertion at work or recreation.
II	Patients have cardiac disease resulting in *slight limitation* of physical activity. They are comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnea, or anginal pain.	*Slight limitation* of ordinary activity. Walking or climbing stairs rapidly, walking uphill, walking or stair climbing after meals, in cold, or when under emotional stress or only during the few hours after awakening. Walking more than two blocks on the level and climbing more than one flight of stairs at a normal pace and in normal conditions.
III	Patients have cardiac disease resulting in *marked limitation* of physical activity. They are comfortable at rest. Less than ordinary physical activity causes fatigue, palpitation, dyspnea, or anginal pain.	*Marked limitation* of ordinary physical activity. Walking one to two blocks on the level and climbing one flight of stairs at normal pace.
IV	Patients have cardiac disease resulting in *inability* to carry on any physical activity without discomfort. Symptoms of cardiac insufficiency or of the anginal syndrome may be present even at rest. If any physical activity is undertaken, discomfort is increased.	*Inability* to carry on any physical activity without discomfort—anginal syndrome *may* be present at rest.

Source: Reproduced with permission from L Goldman et al: Comparative reproducibility and validity of systems for assessing cardiovascular functional class: Advantages of a new specific activity scale. Circulation 64:1227, 1981.

Systematic questioning of a patient with suspected IHD is important to uncover the features of an unstable syndrome associated with increased risk, such as angina occurring with less exertion than in the past, occurring at rest, or awakening the patient from sleep. Since coronary atherosclerosis often is accompanied by similar lesions in other arteries, a patient with angina should be questioned and examined for peripheral arterial disease (intermittent claudication **[Chap. 292]**), stroke, or transient ischemic attacks **(Chap. 437)**. It is also important to uncover a family history of premature IHD (<55 years in first-degree male relatives and <65 years in female relatives) and the presence of diabetes mellitus, hyperlipidemia, hypertension, cigarette smoking, and other risk factors for coronary atherosclerosis.

The history of typical angina pectoris establishes the diagnosis of IHD until proven otherwise. Given the importance of the history, clinicians should move beyond unstructured interviews with the patient and consider using a validated questionnaire (e.g., Seattle Angina Questionnaire) to establish the presence and severity of IHD. The coexistence of advanced age, male sex, the postmenopausal state, and risk factors for atherosclerosis increases the likelihood of hemodynamically significant coronary disease. A particularly challenging problem is the evaluation and management of patients with persistent ischemic-type chest discomfort but no flow-limiting obstructions in their epicardial coronary arteries. This situation arises more often in women than in men. Potential etiologies include microvascular coronary disease (detectable on coronary reactivity testing in response to vasoactive agents such as intracoronary adenosine, acetylcholine, and nitroglycerin) and abnormal cardiac nociception. Treatment of microvascular coronary disease should focus on efforts to improve endothelial function, including nitrates, beta blockers, calcium antagonists, statins, and angiotensin-converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARBs). Abnormal cardiac nociception is more difficult to manage and may be ameliorated in some cases by imipramine.

■ PHYSICAL EXAMINATION

The physical examination is often normal in patients with stable angina when they are asymptomatic. However, because of the increased likelihood of IHD in patients with diabetes and/or peripheral or cerebrovascular arterial disease, clinicians should search for evidence of atherosclerotic disease at other sites, such as an abdominal aortic aneurysm, carotid arterial bruits, and diminished arterial pulses in the lower extremities. The physical examination also should include a search for evidence of risk factors for atherosclerosis such as xanthelasmas and xanthomas. Evidence for peripheral and cerebrovascular arterial disease should be sought by evaluating the pulse contour at multiple locations and comparing the blood pressure between the arms and between the arms and the legs (ankle-brachial index). Examination of the fundi may reveal an increased light reflex and arteriovenous nicking as evidence of hypertension. There also may be signs of anemia, thyroid disease, and nicotine stains on the fingertips from cigarette smoking.

Palpation may reveal cardiac enlargement and abnormal contraction of the cardiac impulse (LV dyskinesia). Auscultation can uncover arterial bruits, a third and/or fourth heart sound, and, if acute ischemia or previous infarction has impaired papillary muscle function, an apical systolic murmur due to mitral regurgitation. These auscultatory signs are best appreciated with the patient in the left lateral decubitus position. Aortic stenosis, aortic regurgitation **(Chap. 272)**, pulmonary hypertension **(Chap. 294)**, and hypertrophic cardiomyopathy **(Chaps. 266–270)** must be excluded, since these disorders may cause angina in the absence of coronary atherosclerosis. Examination during an anginal attack is useful, since ischemia can cause transient LV failure with the appearance of a third and/or fourth heart sound, a dyskinetic cardiac apex, mitral regurgitation, and even pulmonary edema. Tenderness of the chest wall, localization of the discomfort with a single fingertip on the chest, or reproduction of the pain with palpation of the chest makes it unlikely that the pain is caused by myocardial ischemia. A protuberant abdomen may indicate that the patient has the metabolic syndrome and is at increased risk for atherosclerosis.

LABORATORY EXAMINATION

Although the diagnosis of IHD can be made with a high degree of confidence from the history and physical examination, a number of simple laboratory tests can be helpful. The urine should be examined for evidence of diabetes mellitus and renal disease (including microalbuminuria) since these conditions accelerate atherosclerosis. Similarly, examination of the blood should include measurements of lipids (cholesterol—total, LDL, high-density lipoprotein [HDL]—triglycerides, and lipoprotein (a)), glucose (hemoglobin A_{1C}), creatinine, hematocrit, and, if indicated based on the physical examination, thyroid function. A chest x-ray may be helpful in demonstrating the consequences of IHD, i.e., cardiac enlargement, ventricular aneurysm, or signs of heart failure. These signs can support the diagnosis of IHD and are important in assessing the degree of cardiac damage. Evidence exists that an elevated level of high-sensitivity C-reactive protein (CRP) (specifically, between 1 and 3 mg/L) is an independent risk factor for IHD and may be useful in therapeutic decision-making about the initiation of hypolipidemic treatment. The major benefit of high-sensitivity CRP is in reclassifying the risk of IHD in patients in the "intermediate" risk category on the basis of traditional risk factors.

ELECTROCARDIOGRAM

A 12-lead ECG recorded at rest may be normal in patients with typical angina pectoris, but there may also be signs of an old myocardial infarction **(Chap. 247)**. Although repolarization abnormalities, i.e., ST-segment and T-wave changes, as well as LVH and disturbances of cardiac rhythm or intraventricular conduction, are suggestive of IHD, they are nonspecific, since they also can occur in pericardial, myocardial, and valvular heart disease or, in the case of the former, transiently with anxiety, changes in posture, drugs, or esophageal disease. The presence of LVH is a significant indication of increased risk of adverse outcomes from IHD. Of note, even though LVH and cardiac rhythm disturbances are nonspecific indicators of the development of IHD, they may be contributing factors to episodes of angina in patients in whom IHD has developed as a consequence of conventional risk factors. Dynamic ST-segment and T-wave changes that accompany episodes of angina pectoris and disappear thereafter are more specific.

STRESS TESTING

Electrocardiographic The most widely used test for both the diagnosis of IHD and the estimation of risk and prognosis involves recording of the 12-lead ECG before, during, and after exercise, usually on a treadmill **(Fig. 284-3)**. The test consists of a standardized incremental increase in external workload **(Table 284-2)** while symptoms, the ECG, and arm blood pressure are monitored. Exercise duration is usually symptom-limited, and the test is discontinued upon evidence of chest discomfort, severe shortness of breath, dizziness, severe fatigue, ST-segment depression >0.2 mV (2 mm), a fall in systolic blood pressure >10 mmHg, or the development of a ventricular tachyarrhythmia. This test is used to discover any limitation in exercise performance, detect typical ECG signs of myocardial ischemia, and establish their relationship to chest discomfort. The ischemic ST-segment response generally is defined as flat or downsloping depression of the ST segment >0.1 mV below baseline (i.e., the PR segment) and lasting longer than 0.08 s (Fig. 284-2). Upsloping or junctional ST-segment changes are not considered characteristic of ischemia and do not constitute a positive test. Although T-wave abnormalities, conduction disturbances, and ventricular arrhythmias that develop during exercise should be noted, they are also not diagnostic. Negative exercise tests in which the target heart rate (85% of maximal predicted heart rate for age and sex) is not achieved are considered nondiagnostic.

In interpreting ECG stress tests, the probability that coronary artery disease (CAD) exists in the patient or population under study (i.e., pretest probability) should be considered. A positive result on exercise indicates that the likelihood of CAD is 98% in males who are >50 years with a history of typical angina pectoris and who develop chest discomfort during the test. The likelihood decreases if the patient has atypical or no chest pain by history and/or during the test.

The incidence of false-positive tests is significantly increased in patients with low probabilities of IHD, such as asymptomatic men age <40 or premenopausal women with no risk factors for premature atherosclerosis. It is also increased in patients taking cardioactive drugs, such as digitalis and antiarrhythmic agents, and in those with intraventricular conduction disturbances, resting ST-segment and T-wave abnormalities, ventricular hypertrophy, or abnormal serum potassium levels. Obstructive disease limited to the circumflex coronary artery may result in a false-negative stress test since the posterolateral portion of the heart that this vessel supplies is not well represented on the surface 12-lead ECG. Since the overall sensitivity of an exercise stress ECG is only ~75%, a negative result does not exclude CAD, although it makes the likelihood of three-vessel or left main CAD extremely unlikely.

A medical professional should be present throughout the exercise test. It is important to measure total duration of exercise, the times to the onset of ischemic ST-segment change and chest discomfort, the external work performed (generally expressed as the stage of exercise), and the internal cardiac work performed, i.e., by the heart rate–blood pressure product. The depth of the ST-segment depression and the time needed for recovery of these ECG changes are also important. Because the risks of exercise testing are small but real—estimated at one fatality and two nonfatal complications per 10,000 tests—equipment for resuscitation should be available. Modified (heart rate–limited rather than symptom-limited) exercise tests can be performed safely in patients as early as 6 days after uncomplicated myocardial infarction (Table 284-2). Contraindications to exercise stress testing include rest angina within 48 h, unstable rhythm, severe aortic stenosis, acute myocarditis, uncontrolled heart failure, severe pulmonary hypertension, and active infective endocarditis.

The normal response to graded exercise includes progressive increases in heart rate and blood pressure. Failure of the blood pressure to increase or an actual decrease with signs of ischemia during the test is an important adverse prognostic sign, since it may reflect ischemia-induced global LV dysfunction. The development of angina and/or severe (>0.2 mV) ST-segment depression at a low workload, i.e., before completion of stage II of the Bruce protocol, and/or ST-segment depression that persists >5 min after the termination of exercise increases the specificity of the test and suggests severe IHD and a high risk of future adverse events.

Cardiac Imaging **(See also Chap. 248)** When the resting ECG is abnormal (e.g., preexcitation syndrome, >1 mm of resting ST-segment depression, left bundle branch block, paced ventricular rhythm), information gained from an exercise test can be enhanced by stress myocardial radionuclide perfusion imaging after the intravenous administration of thallium-201 or 99m-technetium sestamibi during exercise (or with pharmacologic) stress. Contemporary data also suggest positron emission tomography (PET) imaging (with exercise or pharmacologic stress) using N-13 ammonia or rubidium-82 as another technique for assessing perfusion. Images obtained immediately after cessation of exercise to detect regional ischemia are compared with those obtained at rest to confirm reversible ischemia and regions of persistently absent uptake that signify infarction.

A sizable fraction of patients who need noninvasive stress testing to identify myocardial ischemia and increased risk of coronary events cannot exercise because of peripheral vascular or musculoskeletal disease, exertional dyspnea, or deconditioning. In these circumstances, an intravenous pharmacologic challenge is used in place of exercise. For example, adenosine can be given to create a coronary "steal" by temporarily increasing flow in nondiseased segments of the coronary vasculature at the expense of diseased segments. Alternatively, a graded incremental infusion of dobutamine may be administered to increase MVO_2. A variety of imaging options are available to accompany these pharmacologic stressors (Fig. 284-3). The development of a transient perfusion defect with a tracer such as thallium-201 or 99m-technetium sestamibi is used to detect myocardial ischemia.

Echocardiography is used to assess LV function in patients with chronic stable angina and patients with a history of a prior myocardial infarction, pathologic Q waves, or clinical evidence of heart failure. Two-dimensional echocardiography can assess both global and regional wall motion abnormalities of the left ventricle that are transient when due to ischemia. Stress (exercise or dobutamine) echocardiography may cause the emergence of regions of akinesis or dyskinesis that are not present at rest. Stress echocardiography, like stress myocardial perfusion imaging, is more sensitive than exercise electrocardiography in the diagnosis of IHD. Cardiac magnetic resonance (CMR) stress testing is also evolving as an alternative to radionuclide, PET, or echocardiographic stress imaging. CMR stress testing performed with dobutamine infusion can be used to assess wall motion abnormalities accompanying ischemia, as well as myocardial perfusion. CMR can be used to provide more complete ventricular evaluation using multislice magnetic resonance imaging (MRI) studies.

Atherosclerotic plaques become progressively calcified over time, and coronary calcification in general increases with age. For this

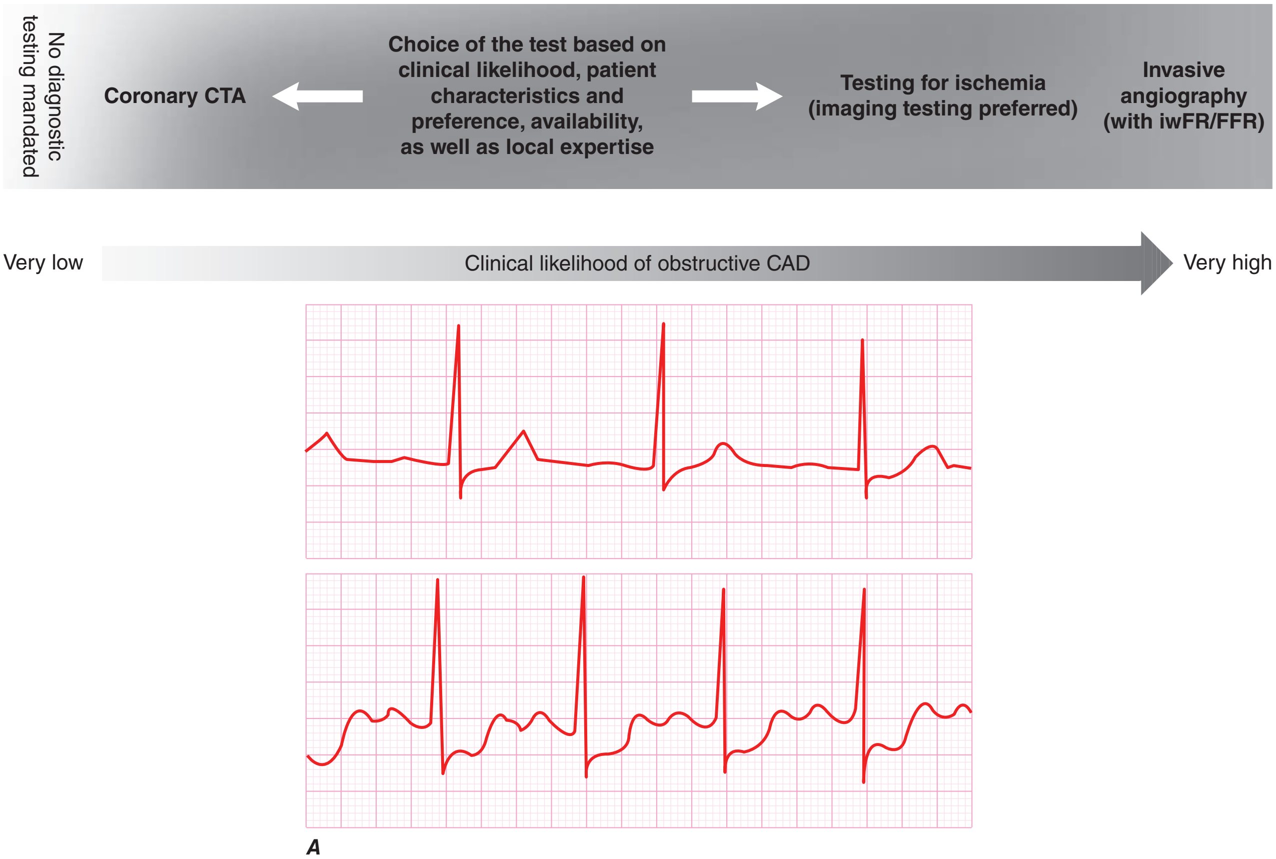
No diagnostic testing mandated
Coronary CTA
Choice of the test based on clinical likelihood, patient characteristics and preference, availability, as well as local expertise
Testing for ischemia (imaging testing preferred)
Invasive angiography (with iwFR/FFR)
Very low
Clinical likelihood of obstructive CAD
Very high
A

FIGURE 284-3 Selecting appropriate testing patients with angina and suspected coronary artery disease (CAD). On the left of the figure is an algorithm for selecting from among testing options. In patients who are at low risk, in whom prior testing was equivocal, or in whom the diagnosis of is CAD uncertain, noninvasive functional stress testing with imaging for myocardial ischemia or computed tomography angiography (CTA) is reasonable to establish the diagnosis of CAD prior to initiation of treatment. Patients with a high clinical likelihood of CAD, patients with symptoms despite antianginal therapy or with low-level activities, and patients with high-risk features based on the initial clinical evaluation may proceed directly to invasive coronary angiography without further diagnostic testing. *(Adapted from J Knuuti et al: 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes. Eur Heart J 41:407, 2020.*) Panels ***A–F*** are examples of the data obtained with electrocardiogram (ECG) monitoring and specialized imaging procedures. CMR, cardiac magnetic resonance; EBCT, electron beam computed tomography; ECHO, echocardiography; FFR, fractional flow reserve; IHD, ischemic heart disease; iwFR, instantaneous wave-free ration; MIBI, methoxyisobutyl isonitrite; MR, magnetic resonance; PET, positron emission tomography. ***A.*** Lead V_4 at rest (*top panel*) and after 4.5 min of exercise (*bottom panel*). There is 3 mm (0.3 mV) of horizontal ST-segment depression, indicating a positive test for ischemia. *(Adapted from BR Chaitman, in E Braunwald et al [eds]: Heart Disease, 8th ed, Philadelphia, Saunders, 2008.)* ***B.*** A 45-year-old avid jogger who began experiencing classic substernal chest pressure underwent an exercise echo study. With exercise, the patient's heart rate increased from 52 to 153 beats/min. The left ventricular chamber dilated with exercise, and the septal and apical portions became akinetic to dyskinetic (*red arrow*). These findings are strongly suggestive of a significant flow-limiting stenosis in the proximal left anterior descending artery, which was confirmed at coronary angiography. *(Modified from SD Solomon, in E Braunwald et al [eds]: Primary Cardiology, 2nd ed, Philadelphia, Saunders, 2003.)* ***C.*** Stress and rest myocardial perfusion single-photon emission computed tomography images obtained with 99m-technetium sestamibi in a patient with chest pain and dyspnea on exertion. The images demonstrate a medium-size and severe stress perfusion defect involving the inferolateral and basal inferior walls, showing nearly complete reversibility, consistent with moderate ischemia in the right coronary artery territory (*red arrows*). *(Images provided by Dr. Marcello Di Carli, Nuclear Medicine Division, Brigham and Women's Hospital, Boston, MA.)* ***D.*** A patient with a prior myocardial infarction presented with recurrent chest discomfort. On cardiac magnetic resonance (CMR) cine imaging, a large area of anterior akinesia was noted (marked by the *arrows* in the top left and right images, systolic frame only). This area of akinesia was matched by a larger extent of late gadolinium-DTPA enhancements consistent with a large transmural myocardial infarction (marked by *arrows* in the middle left and right images). Resting (*bottom left*) and adenosine vasodilating stress (*bottom right*) first-pass perfusion images revealed reversible perfusion abnormality that extended to the inferior septum. This patient was found to have an occluded proximal left anterior descending coronary artery with extensive collateral formation. This case illustrates the utility of different modalities in a CMR examination in characterizing ischemic and infarcted myocardium. DTPA, diethylenetriamine penta-acetic acid. *(Images provided by Dr. Raymond Kwong, Cardiovascular Division, Brigham and Women's Hospital, Boston, MA.)* ***E.*** Stress and rest myocardial perfusion PET images obtained with rubidium-82 in a patient with chest pain on exertion. The images demonstrate a large and severe stress perfusion defect involving the mid and apical anterior, anterolateral, and anteroseptal walls and the left ventricular apex, showing complete reversibility, consistent with extensive and severe ischemia in the mid-left anterior descending coronary artery territory (*red arrows*). *(Images provided by Dr. Marcello Di Carli, Nuclear Medicine Division, Brigham and Women's Hospital, Boston, MA.)* ***F.*** A 58-year-old woman with psoriasis presented with chest discomfort and underwent coronary CT angiography (CCTA) for further evaluation. There was a large amount of mostly noncalcified plaque resulting in severe (>70%) stenosis of the mid left anterior descending artery (LAD) and the mid-right coronary artery (RCA). There was minimal plaque in the left circumflex (LCx). CCTA data can be used to reconstruct three-dimensional images to aid visualization. In addition, each coronary artery can be visualized en face by creating a short axis along the length of the coronary segment being visualized. In this case, the stenosis of the LAD is seen in the red box (corresponding to the level of the *dashed red line*) and compared with the proximal normal segment in the green box (corresponding to the level of the *dashed green line*). (*Images provided by Dr. Ron Blankstein, Cardiovascular Division, Brigham and Women's Hospital, Boston, MA.*)

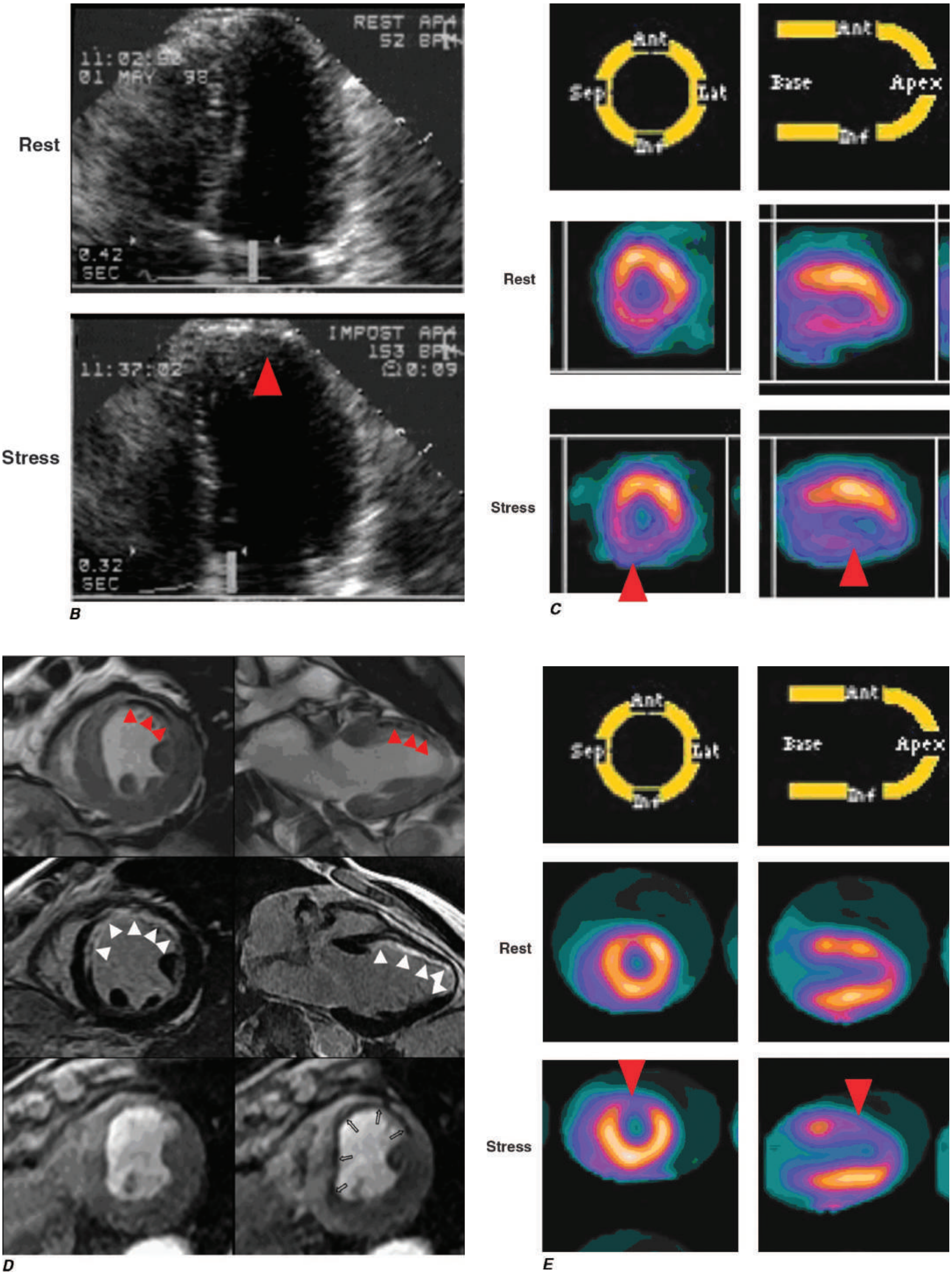
Rest
REST AP4
52 BPM
11:02:52
01 MAY 98
0.42
SEC
Stress
IMPOST AP4
153 BPM
0:09
11:37:02
0.32
SEC
B
Ant
Sep
Lat
Inf
Base
Apex
Rest
Stress
C
D
Ant
Sep
Lat
Inf
Base
Apex
Rest
Stress
E

FIGURE 284-3 (Continued)

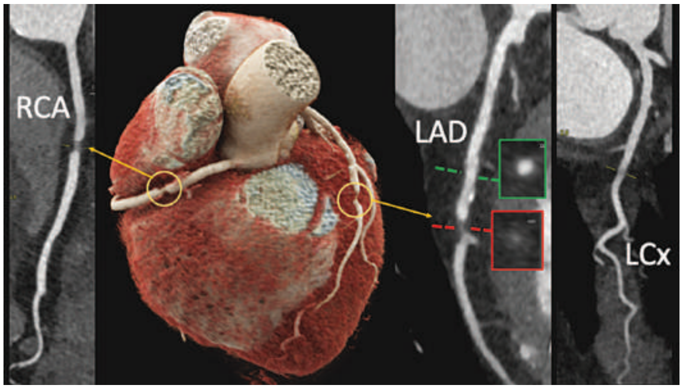

F

FIGURE 284-3 (Continued)

reason, methods for detecting coronary calcium have been developed as a measure of the presence of coronary atherosclerosis. These methods involve CT applications that achieve rapid acquisition of images (electron beam [EBCT] and multidetector [MDCT] detection). Coronary calcium detected by these imaging techniques most commonly is quantified by using the Agatston score, which is based on the area and density of calcification.

Coronary CT angiography (CCTA) (Fig. 284-3) is helpful in diagnosing the extent and severity of nonobstructive and obstructive IHD, as well as assessing plaque composition. Thus, CCTA is an attractive option when the pretest likelihood of IHD is intermediate in a young patient (< 65 years) with chest pain or less obstructive IHD is suspected. Stress imaging is the preferred diagnostic test when the goal is assessing the adequacy of ischemia-guided management.

CORONARY ARTERIOGRAPHY

(See also Chap. 249) This diagnostic method outlines the lumina of the coronary arteries and can be used to detect or exclude serious coronary obstruction. However, coronary arteriography provides no information about the arterial wall, and severe atherosclerosis that does not encroach on the lumen may go undetected. Of note, atherosclerotic plaques characteristically are scattered throughout the coronary tree, tend to occur more frequently at branch points, and grow progressively in the intima and media of an epicardial coronary artery at first without encroaching on the lumen, causing an outward bulging of the artery—a process referred to as remodeling. Later in the course of the disease, further growth causes luminal narrowing.

Indications The ISCHEMIA trial informs decision-making about referral for coronary arteriography (with intent to perform revascularization) in patients with stable IHD and an ejection fraction >35% even in the presence of moderate-severe ischemia on noninvasive functional testing. Over the course of 4 years of follow-up, early referral for an invasive strategy was not associated with a reduction in the risk of myocardial infarction or death but was more effective than an initial conservative, medical strategy in relieving angina. Thus, coronary arteriography is indicated in (1) patients with chronic stable angina pectoris who are severely symptomatic despite medical therapy and are

TABLE 284-2 Relation of Metabolic Equivalent Tasks (METs) to Stages in Various Testing Protocols

FUNCTIONAL CLASS	CLINICAL STATUS				O_2 COST mL/Kg/min	METs	TREADMILL PROTOCOLS			
							BRUCE Modified 3 min Stages		BRUCE 3 min Stages	
							MPH	%GR	MPH	%GR
NORMAL AND I							6.0	22	6.0	22
	HEALTHY, DEPENDENT ON AGE, ACTIVITY						5.5	20	5.2	20
							5.0	18	5.0	18
					56.0	16				
					52.5	15				
					49.0	14				
					45.5	13	4.2	16	4.2	16
					42.0	12				
					38.5	11	3.4	14	3.4	14
		SEDENTARY HEALTHY			35.0	10				
					31.5	9				
					28.0	8				
			LIMITED		24.5	7	2.5	12	2.5	12
II				SYMPTOMATIC	21.0	6				
					17.5	5	1.7	10	1.7	10
III					14.0	4				
					10.5	3	1.7	5		
					7.0	2	1.7	0		
IV					3.5	1				

Note: The standard Bruce treadmill protocol (right-hand column) begins at 1.7 MPH and 10% gradient (GR) and progresses every 3 min to a higher speed and elevation. The corresponding oxygen consumption and clinical status of the patient are shown in the center and left-hand columns.

Abbreviations: GR, grade; MPH, miles per hour.

Source: Reproduced with permission from GF Fletcher et al: Exercise standards for testing and training. Circulation 104:1694, 2001.

being considered for revascularization, i.e., a percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG); (2) patients with troublesome symptoms that present diagnostic difficulties in whom there is a need to confirm or rule out the diagnosis of IHD; (3) patients with known or possible angina pectoris who have survived cardiac arrest; and (4) patients with angina or evidence of ischemia on noninvasive testing with clinical or laboratory evidence of ventricular dysfunction.

Examples of other indications for coronary arteriography include the following:

1. Patients with chest discomfort suggestive of angina pectoris but a negative or nondiagnostic stress test who require a definitive diagnosis for guiding medical management, alleviating psychological stress, career or family planning, or insurance purposes.
2. Patients who have been admitted repeatedly to the hospital for a suspected acute coronary syndrome **(Chaps. 285 and 286)**, but in whom this diagnosis has not been established and in whom it is considered clinically important to determine the presence or absence of CAD.
3. Patients with careers that involve the safety of others (e.g., pilots, firefighters, police) who have questionable symptoms or suspicious or positive noninvasive tests and in whom there are reasonable doubts about the state of the coronary arteries.
4. Patients with aortic stenosis or hypertrophic cardiomyopathy and angina in whom the chest pain could be due to IHD.
5. Male patients >45 years and females >55 years who are to undergo a cardiac operation such as valve replacement or repair and who may or may not have clinical evidence of myocardial ischemia.
6. Patients after myocardial infarction, especially those who are at high risk after myocardial infarction because of the recurrence of angina or the presence of heart failure, frequent ventricular premature contractions, or signs of ischemia on the stress test.
7. Patients in whom coronary spasm or another nonatherosclerotic cause of myocardial ischemia (e.g., coronary artery anomaly, Kawasaki disease) is suspected.

Noninvasive alternatives to diagnostic coronary arteriography include CT angiography and CMR angiography **(Chap. 248)**. Important aspects of their use that should be noted include the substantially higher radiation exposure with CT angiography compared to conventional diagnostic arteriography and the limitations on CMR imposed by cardiac movement during the cardiac cycle, especially at high heart rates.

PROGNOSIS

The principal prognostic indicators in patients known to have IHD are age, the functional state of the left ventricle, the location(s) and severity of coronary artery narrowing, and the severity or activity of myocardial ischemia. Angina pectoris of recent onset, unstable angina **(Chap. 285)**, early postmyocardial infarction angina, angina that is unresponsive or poorly responsive to medical therapy, and angina accompanied by symptoms of congestive heart failure all indicate an increased risk for adverse coronary events. The same is true for the physical signs of heart failure, episodes of pulmonary edema, transient third heart sounds, and mitral regurgitation and for echocardiographic or radioisotopic (or roentgenographic) evidence of cardiac enlargement and reduced (<0.40) ejection fraction.

Most important, any of the following signs during noninvasive testing indicates a high risk for coronary events: inability to exercise for 6 min, i.e., stage II (Bruce protocol) of the exercise test; a strongly positive exercise test showing onset of myocardial ischemia at low workloads (≥0.1 mV ST-segment depression before completion of stage II, ≥0.2 mV ST-segment depression at any stage, ST-segment depression for >5 min after the cessation of exercise, a decline in systolic pressure >10 mmHg during exercise, or the development of ventricular tachyarrhythmias during exercise); the development of large or multiple perfusion defects or increased lung uptake during stress radioisotope perfusion imaging; and a decrease in LV ejection fraction during exercise on radionuclide ventriculography or during stress echocardiography. Conversely, patients who can complete stage III of the Bruce exercise protocol and have a normal stress perfusion scan or negative stress echocardiographic evaluation are at very low risk for future coronary events. The finding of frequent episodes of ST-segment deviation on ambulatory ECG monitoring (even in the absence of symptoms) is also an adverse prognostic finding.

On cardiac catheterization, elevations of LV end-diastolic pressure and ventricular volume and reduced ejection fraction are the most important signs of LV dysfunction and are associated with a poor prognosis. Patients with chest discomfort but normal LV function and normal coronary arteries have an excellent prognosis. Obstructive lesions of the left main (>50% luminal diameter) or left anterior descending coronary artery proximal to the origin of the first septal artery are associated with a greater risk than are lesions of the right or left circumflex coronary artery because of the greater quantity of myocardium at risk. Atherosclerotic plaques in epicardial arteries with fissuring or filling defects indicate increased risk. These lesions go through phases of inflammatory cellular activity, degeneration, endothelial dysfunction, abnormal vasomotion, platelet aggregation, and fissuring or hemorrhage. These factors can temporarily worsen the stenosis and cause thrombosis and/or abnormal reactivity of the vessel wall, thus exacerbating the manifestations of ischemia. The recent onset of symptoms, the development of severe ischemia during stress testing (see above), and unstable angina pectoris **(Chap. 285)** all reflect episodes of rapid progression in coronary lesions.

With any degree of obstructive CAD, mortality is greatly increased when LV function is impaired; conversely, at any level of LV function, the prognosis is influenced importantly by the quantity of myocardium perfused by critically obstructed vessels. Therefore, it is essential to collect all the evidence substantiating past myocardial damage (evidence of myocardial infarction on ECG, echocardiography, radioisotope imaging, or left ventriculography), residual LV function (ejection fraction and wall motion), and risk of future damage from coronary events (extent of coronary disease and severity of ischemia defined by noninvasive stress testing). The larger the quantity of established myocardial necrosis is, the less the heart is able to withstand additional damage and the poorer the prognosis. Risk estimation must include age, presenting symptoms, all risk factors, signs of arterial disease, existing cardiac damage, and signs of impending damage (i.e., ischemia).

The greater the number and severity of risk factors for coronary atherosclerosis (advanced age [>75 years], hypertension, dyslipidemia, diabetes, morbid obesity, accompanying peripheral and/or cerebrovascular disease, previous myocardial infarction), the worse the prognosis of an angina patient. Evidence exists that elevated levels of CRP in the plasma, extensive coronary calcification on EBCT (see above), and increased carotid intimal thickening on ultrasound examination also indicate an increased risk of coronary events.

TREATMENT

Stable Angina Pectoris

Once the diagnosis of IHD has been made, each patient must be evaluated individually with respect to their level of understanding, expectations and goals, control of symptoms, and prevention of adverse clinical outcomes such as myocardial infarction and premature death. The degree of disability and the physical and emotional stress that precipitates angina must be recorded carefully to set treatment goals. The management plan should include the following components: (1) explanation of the problem and reassurance about the ability to formulate a treatment plan, (2) identification and treatment of aggravating conditions, (3) recommendations for adaptation of activity as needed, (4) treatment of risk factors that will decrease the occurrence of adverse coronary outcomes, (5) drug therapy for angina, and (6) consideration of revascularization. Emphasis should be placed on a patient-centric, team-based approach to care, with recognition of the importance of social determinants of health.

EXPLANATION AND REASSURANCE

Patients with IHD need to understand their condition and realize that a long and productive life is possible even though they have angina pectoris or have experienced and recovered from an acute myocardial infarction. Offering results of clinical trials showing improved outcomes can be of great value in encouraging patients to resume or maintain activity and return to work. A planned program of rehabilitation can encourage patients to lose weight, improve exercise tolerance, and control risk factors with more confidence.

IDENTIFICATION AND TREATMENT OF AGGRAVATING CONDITIONS

A number of conditions may increase oxygen demand or decrease oxygen supply to the myocardium and may precipitate or exacerbate angina in patients with IHD. LVH, aortic valve disease, and hypertrophic cardiomyopathy may cause or contribute to angina and should be excluded or treated. Obesity, hypertension, and hyperthyroidism should be treated aggressively to reduce the frequency and severity of anginal episodes. Decreased myocardial oxygen supply may be due to reduced oxygenation of the arterial blood (e.g., in pulmonary disease or, when carboxyhemoglobin is present, due to cigarette or cigar smoking) or decreased oxygen-carrying capacity (e.g., in anemia). Correction of these abnormalities, if present, may reduce or even eliminate angina pectoris.

ADAPTATION OF ACTIVITY

Myocardial ischemia is caused by a discrepancy between the demand of the heart muscle for oxygen and the ability of the coronary circulation to meet that demand. Most patients can be helped to understand this concept and utilize it in the rational programming of activity. Many tasks that ordinarily evoke angina may be accomplished without symptoms simply by reducing the speed at which they are performed. Patients must appreciate the diurnal variation in their tolerance of certain activities and reducing their energy requirements in the morning, immediately after meals, and in cold or inclement weather. On occasion, it may be necessary to recommend a change in employment or residence to avoid physical stress.

Physical conditioning usually improves the exercise tolerance of patients with angina and has substantial psychological benefits. A regular program of isotonic exercise that is within the limits of the individual patient's threshold for the development of angina pectoris and that does not exceed 80% of the heart rate associated with ischemia on exercise testing should be strongly encouraged. Based on the results of an exercise test, the number of metabolic equivalent tasks (METs) performed at the onset of ischemia can be estimated (Table 284-2) and a practical exercise prescription can be formulated to permit daily activities that will fall below the ischemic threshold **(Table 284-3)**.

TREATMENT OF RISK FACTORS

A *family history* of premature IHD is an important indicator of increased risk and should trigger a search for treatable risk factors such as hyperlipidemia, hypertension, and diabetes mellitus. *Obesity* impairs the treatment of other risk factors and increases the risk of adverse coronary events. In addition, obesity often is accompanied by three other risk factors: diabetes mellitus, hypertension, and hyperlipidemia. The treatment of obesity and these accompanying risk factors is an important component of any management plan. A diet low in saturated and *trans*-unsaturated fatty acids and a reduced caloric intake to achieve optimal body weight are a cornerstone in the management of chronic IHD. It is especially important to emphasize weight loss and regular exercise in patients with the metabolic syndrome or overt diabetes mellitus.

Cigarette smoking accelerates coronary atherosclerosis in both sexes and at all ages and increases the risk of thrombosis, plaque instability, myocardial infarction, and death. In addition, by increasing myocardial oxygen needs and reducing oxygen supply, it aggravates angina. Smoking cessation studies have demonstrated important benefits with a significant decline in the occurrence of these adverse outcomes. Noncombustible tobacco in the form of electronic cigarettes (nicotine delivery systems) may also increase the frequency of anginal episodes. The physician's message must be clear and strong and supported by programs that achieve and monitor abstinence from all tobacco product use **(Chap. 465)**.

TABLE 284-3 Energy Requirements for Some Common Activities

LESS THAN 3 METs	3–5 METs	5–7 METs	7–9 METs	MORE THAN 9 METs
Self-Care				
Washing/shaving Dressing Light housekeeping Desk work Driving auto	Cleaning windows Raking Power lawn mowing Bed making/stripping Carrying objects (15–30 lb)	Easy digging in garden Level hand lawn mowing Carrying objects (30–60 lb)	Heavy shoveling Carrying objects (60–90 lb)	Carrying loads upstairs (objects >90 lb) Climbing stairs (quickly) Shoveling heavy snow
Occupational				
Sitting (clerical/assembly) Desk work Standing (store clerk)	Stocking shelves (light objects) Light welding/carpentry	Carpentry (exterior) Shoveling dirt Sawing wood	Digging ditches (pick and shovel)	Heavy labor
Recreational				
Golf (cart) Knitting	Dancing (social) Golf (walking) Sailing Tennis (doubles)	Tennis (singles) Snow skiing (downhill) Light backpacking Basketball Stream fishing	Canoeing Mountain climbing	Squash Ski touring Vigorous basketball
Physical Conditioning				
Walking (2 mph) Stationary bike Very light calisthenics	Level walking (3–4 mph) Level biking (6–8 mph) Light calisthenics	Level walking (4.5–5.0 mph) Bicycling (9–10 mph) Swimming, breast stroke	Level jogging (5 mph) Swimming (crawl stroke) Rowing machine Heavy calisthenics Bicycling (12 mph)	Running more than 6 mph Bicycling (more than 13 mph) Rope jumping Walking uphill (5 mph)

Abbreviation: METs, metabolic equivalent tasks.

Source: Modified from WL Haskell: Rehabilitation of the coronary patient, in NK Wenger, HK Hellerstein (eds): *Design and Implementation of Cardiac Conditioning Program.* New York, Churchill Livingstone, 1978.

Hypertension **(Chap. 288)** may coexist with other risk factors for IHD and is associated with an increased risk of adverse clinical events from coronary atherosclerosis as well as stroke. In addition, the LVH that results from sustained hypertension aggravates ischemia. There is evidence that long-term effective treatment of hypertension can decrease the occurrence of adverse coronary events **(Chap. 288)**.

Diabetes mellitus **(Chap. 415)** accelerates coronary, cerebrovascular, and peripheral atherosclerosis and is frequently associated with dyslipidemia and increases in the risk of angina, myocardial infarction, and sudden coronary death. Aggressive control of the dyslipidemia (target LDL cholesterol <70 mg/dL) and hypertension (blood pressure <130/80 mmHg), that are frequently found in patients with diabetes mellitus, is highly effective and therefore essential, as described below. Use of either a sodium-glucose cotransporter 2 (SGLT-2) inhibitor or glucagon-like peptide 1 (GLP-1) receptor agonist in patients with IHD and type 2 diabetes is recommended in current guidelines to reduce the risk of major adverse cardiovascular events.

DYSLIPIDEMIA

The treatment of dyslipidemia is central in aiming for long-term relief from angina, reduced need for revascularization, and reduction in myocardial infarction and death. The control of lipids can be achieved by the combination of a diet low in saturated and *trans*-unsaturated fatty acids, exercise, and weight loss. Nearly always, HMG-CoA reductase inhibitors (statins) are required and can lower LDL cholesterol (25–50%), raise HDL cholesterol (5–9%), and lower triglycerides (5–30%). A powerful treatment effect of statins on atherosclerosis, IHD, and outcomes is seen regardless of the pretreatment LDL cholesterol level. In patients with IHD on a maximally tolerated statin with an LDL cholesterol ≥70 mg/dL, addition of ezetimibe is recommended as the next step, followed by a PCSK9 inhibitor (in patients at very high risk) should the LDL cholesterol goal not be achieved. When these therapies are not sufficient or tolerated, bempedoic acid may be considered for further LDL cholesterol reduction. Icosapent ethyl (purified eicosapentaenoic acid), fibrates, and resin-binding agents can be used to lower triglycerides **(Chap. 419)**, but only icosapent ethyl has been shown to reduce cardiovascular risk in the statin era. Controlled trials with lipid-regulating regimens have shown equal proportional benefit for men, women, the elderly, patients with diabetes mellitus, and smokers. Evidence exists that a high rate of control of LDL over time is associated with a lower cumulative LDL exposure and lower rate of major adverse cardiovascular events.

Compliance with the health-promoting behaviors listed above is generally very poor, and a conscientious physician must not underestimate the major effort required to meet this challenge. Many patients who are discharged from the hospital with proven coronary disease do not receive adequate treatment for dyslipidemia. In light of the proof that treating dyslipidemia brings major benefits, physicians need to establish treatment pathways, monitor compliance, and follow up regularly.

RISK REDUCTION IN WOMEN WITH IHD

The incidence of clinical IHD in premenopausal women is very low; however, after menopause, the atherogenic risk factors increase (e.g., increased LDL) and the rate of clinical coronary events accelerates to the levels observed in men. Diabetes mellitus, which is more common in women, greatly increases the occurrence of clinical IHD and amplifies the deleterious effects of hypertension, hyperlipidemia, and smoking. Cardiac catheterization and coronary revascularization are underused in women and are performed at a later and more severe stage of the disease than in men. When cholesterol lowering, beta blockers after myocardial infarction, and CABG are applied in the appropriate patient groups, women benefit to the same degree as men.

DRUG THERAPY

The commonly used drugs for the treatment of angina pectoris are summarized in **Tables 284-4 through 284-6**. Pharmacotherapy for IHD is designed to reduce the frequency of anginal episodes, myocardial infarction, and coronary death. Trial data emphasize how important medical management is when added to the health-promoting behaviors discussed above. To achieve maximum benefit from medical therapy for IHD, it is frequently necessary to combine agents from different classes and titrate the doses as guided by the individual profile of risk factors, symptoms, hemodynamic responses, and side effects.

TABLE 284-4 Nitrate Therapy in Patients with Ischemic Heart Disease

PREPARATION OF AGENT	DOSE	SCHEDULE
Nitroglycerin[a]		
Ointment	0.5–2 in.	Two or three times daily
Transdermal patch	0.2–0.8 mg/h	Every 24 h; remove at bedtime for 12–14 h
Sublingual tablet	0.3–0.6 mg	As needed, up to three doses 5 min apart
Spray	One or two sprays	As needed, up to three doses 5 min apart
Isosorbide dinitrate[a]		
Oral	10–40 mg	Two or three times daily
Oral sustained release	80–120 mg	Once or twice daily (eccentric schedules)
Isosorbide 5-mononitrate		
Oral	20 mg	Twice daily (given 7–8 h apart)
Oral sustained release	30–240 mg	Once daily

[a] A 10- to 12-h nitrate-free interval is recommended.

Source: Reproduced with permission from DA Morrow, WE Boden: Stable ischemic heart disease. In RO Bonow et al (eds): *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*, 9th ed. Philadelphia, Saunders, 2012.

NITRATES

The organic nitrates are a valuable class of drugs in the management of angina pectoris (Table 284-4). Their major mechanisms of action include systemic venodilation with concomitant reduction in LV end-diastolic volume and pressure, thereby reducing myocardial wall tension and oxygen requirements; dilation of epicardial coronary vessels; and increased blood flow in collateral vessels. When metabolized, organic nitrates release nitric oxide (NO) that binds to guanylyl cyclase in vascular smooth muscle

TABLE 284-5 Properties of Beta Blockers in Clinical Use for Ischemic Heart Disease

DRUGS	SELECTIVITY	PARTIAL AGONIST ACTIVITY	USUAL DOSE FOR ANGINA
Acebutolol	β_1	Yes	200–600 mg twice daily
Atenolol	β_1	No	50–200 mg/d
Betaxolol	β_1	No	10–20 mg/d
Bisoprolol	β_1	No	10 mg/d
Esmolol (intravenous)[a]	β_1	No	50–300 μg/kg/min
Labetalol[b]	None	Yes	200–600 mg twice daily
Metoprolol	β_1	No	50–200 mg twice daily
Nadolol	None	No	40–80 mg/d
Nebivolol	β_1 (at low doses)	No	5–40 mg/d
Pindolol	None	Yes	2.5–7.5 mg 3 times daily
Propranolol	None	No	80–120 mg twice daily
Timolol	None	No	10 mg twice daily

[a] Esmolol is an ultra-short-acting beta blocker that is administered as a continuous intravenous infusion. Its rapid offset of action makes esmolol an attractive agent to use in patients with relative contraindications to beta blockade. [b] Labetalol is a combined alpha and beta blocker.

Note: This list of beta blockers that may be used to treat patients with angina pectoris is arranged alphabetically. It is preferable to use a sustained-release formulation that may be taken once daily to improve the patient's compliance with the regimen.

Source: Data from RJ Gibbons et al: J Am Coll Cardiol 41:159, 2003.

TABLE 284-6 Calcium Channel Blockers in Clinical Use for Ischemic Heart Disease

DRUGS	USUAL DOSE	DURATION OF ACTION	SIDE EFFECTS
Dihydropyridines			
Amlodipine	5–10 mg qd	Long	Headache, edema
Felodipine	5–10 mg qd	Long	Headache, edema
Isradipine	2.5–10 mg bid	Medium	Headache, fatigue
Nicardipine	20–40 mg tid	Short	Headache, dizziness, flushing, edema
Nifedipine	Immediate release:[a] 30–90 mg daily orally Slow release: 30–180 mg orally	Short	Hypotension, dizziness, flushing, nausea, constipation, edema
Nisoldipine	20–40 mg qd	Short	Similar to nifedipine
Nondihydropyridines			
Diltiazem	Immediate release: 30–80 mg 4 times daily	Short	Hypotension, dizziness, flushing, bradycardia, edema
	Slow release: 120–320 mg qd	Long	
Verapamil	Immediate release: 80–160 mg tid	Short	Hypotension, myocardial depression,
	Slow release: 120–480 mg qd	Long	heart failure, edema, bradycardia

[a] May be associated with increased risk of mortality if administered during acute myocardial infarction.

Note: This list of calcium channel blockers that may be used to treat patients with angina pectoris is divided into two broad classes, dihydropyridines and nondihydropyridines, and arranged alphabetically within each class. Among the dihydropyridines, the greatest clinical experience has been obtained with amlodipine and nifedipine. After the initial period of dose titration with a short-acting formulation, it is preferable to switch to a sustained-release formulation that may be taken once daily to improve patient compliance with the regimen.

Source: Data from RJ Gibbons et al: J Am Coll Cardiol 41:159, 2003.

cells, leading to an increase in cyclic guanosine monophosphate, which causes relaxation of vascular smooth muscle. Nitrates also exert antithrombotic activity by NO-dependent activation of platelet guanylyl cyclase, impairment of intraplatelet calcium flux, and platelet activation.

The absorption of these agents is rapid and complete through mucous membranes. For this reason, nitroglycerin is most commonly administered sublingually in tablets of 0.4 or 0.6 mg. Patients with angina should be instructed to take the medication both to relieve angina and also ~5 min before activities that are likely to induce an episode.

Nitrates improve exercise tolerance in patients with chronic angina and relieve ischemia in patients with unstable angina as well as patients with vasospastic variant angina **(Chap. 285)**. A diary of angina and nitroglycerin use may be valuable for detecting changes in the frequency, severity, or threshold for discomfort that may signify the development of unstable angina pectoris and/or herald an impending myocardial infarction.

Long-Acting Nitrates None of the long-acting nitrates is as effective as sublingual nitroglycerin for the acute relief of angina. These organic nitrate preparations can be swallowed, chewed, or administered as a patch or paste by the transdermal route (Table 284-4). They provide effective plasma levels for up to 24 h, but the therapeutic response is highly variable. Different preparations and/or administration during the daytime should be tried only to prevent discomfort while avoiding side effects such as headache and dizziness. Individual dose titration is important to prevent side effects. To minimize the effects of nitrate tolerance, the minimum effective dose should be used and a minimum of 8 h each day kept free of the drug to restore any useful response(s).

Beta-Adrenergic Blockers These drugs represent an important component of the pharmacologic treatment of angina pectoris (Table 284-5). They reduce myocardial oxygen demand by inhibiting the increases in heart rate, arterial pressure, and myocardial contractility caused by adrenergic activation. Beta blockade reduces these variables most strikingly during exercise but causes only small reductions at rest. Long-acting beta-blocking drugs or sustained-release formulations offer the advantage of once-daily dosing (Table 284-5). The therapeutic aims include relief of angina and ischemia. These drugs also can reduce mortality and reinfarction rates in patients after myocardial infarction and are moderately effective antihypertensive agents. Use of beta blockers beyond 1 year after myocardial infarction may be reassessed in the absence of LV systolic dysfunction, angina, arrhythmias, or uncontrolled hypertension, as their long-term benefit in such patients is unclear.

Relative contraindications include asthma and reversible airway obstruction in patients with chronic lung disease, atrioventricular conduction disturbances, severe bradycardia, Raynaud's phenomenon, and a history of mental depression. Side effects include fatigue, reduced exercise tolerance, nightmares, impotence, cold extremities, intermittent claudication, bradycardia (sometimes severe), impaired atrioventricular conduction, LV failure, bronchial asthma, worsening claudication, and intensification of the hypoglycemia produced by oral hypoglycemic agents and insulin. Reducing the dose or even discontinuation may be necessary if these side effects develop and persist. Since sudden discontinuation can intensify ischemia, the doses should be tapered over 2 weeks. Beta blockers with relative β_1-receptor specificity such as metoprolol and atenolol may be preferable in patients with mild bronchial obstruction and insulin-requiring diabetes mellitus.

Calcium Channel Blockers Calcium channel blockers (Table 284-6) are coronary vasodilators that produce variable and dose-dependent reductions in myocardial oxygen demand, contractility, and arterial pressure. These combined pharmacologic effects are advantageous and make these agents as effective as beta blockers in the treatment of angina pectoris. They are indicated when beta blockers are contraindicated, poorly tolerated, or ineffective. Because of differences in the dose-response relationship on cardiac electrical activity between the dihydropyridine and nondihydropyridine calcium channel blockers, verapamil and diltiazem may produce symptomatic disturbances in cardiac conduction and bradyarrhythmias. They also exert negative inotropic actions and are more likely to aggravate LV failure, particularly when used in patients with LV dysfunction, especially if the patients are also receiving beta blockers. Although useful effects usually are achieved when calcium channel blockers are combined with beta blockers and nitrates, individual titration of the doses is essential with these combinations. Vasospastic angina responds particularly well to calcium channel blockers (especially members of the dihydropyridine class), supplemented when necessary by nitrates **(Chap. 285)**.

Verapamil ordinarily should not be combined with beta blockers because of the combined adverse effects on heart rate and contractility. Diltiazem can be combined with beta blockers in patients with normal ventricular function and no conduction disturbances. Amlodipine and beta blockers have complementary actions on coronary blood supply and myocardial oxygen demands. Whereas the former decreases blood pressure and dilates coronary arteries, the latter slows heart rate and decreases contractility. Amlodipine and the other second-generation dihydropyridine calcium antagonists (nicardipine, isradipine, long-acting nifedipine, and felodipine) are potent vasodilators and are useful in the simultaneous treatment of angina and hypertension. Short-acting dihydropyridines should be avoided because of the risk of precipitating infarction, particularly in the absence of concomitant beta blocker therapy.

Choice Between Beta Blockers and Calcium Channel Blockers for Initial Therapy Since beta blockers have been shown to improve life expectancy after acute myocardial infarction **(Chaps. 285 and 286)** and calcium channel blockers have not, the former may also be preferable in patients with angina and a damaged left ventricle. However, calcium channel blockers are indicated in patients with the following: (1) inadequate responsiveness to the combination of beta blockers and nitrates; many of these patients do well with a combination of a beta blocker and a dihydropyridine calcium channel blocker; (2) adverse reactions to beta blockers such as depression, sexual disturbances, and fatigue; (3) angina and a history of asthma or chronic obstructive pulmonary disease; (4) sick-sinus syndrome or significant atrioventricular conduction disturbances; (5) vasospastic angina; or (6) symptomatic peripheral arterial disease.

Ranolazine, a piperazine derivative, may be useful for patients with chronic angina despite standard medical therapy **(Table 284-7)**. Its antianginal action is believed to occur via inhibition of the late inward sodium current (I_{Na}). The benefits of I_{Na} inhibition include limitation of the Na overload of ischemic myocytes and prevention of Ca^{2+} overload via the Na^{+}–Ca^{2+} exchanger. A dose of 500–1000 mg orally twice daily is usually well tolerated. Ranolazine is contraindicated in patients with hepatic impairment or with conditions or drugs associated with QT_c prolongation, and when drugs that inhibit the CYP3A metabolic system (e.g., ketoconazole, diltiazem, verapamil, macrolide antibiotics, HIV protease inhibitors, and large quantities of grapefruit juice) are being used.

A comparison of the common side effects, contraindications, and potential drug interactions of many of the frequently presented antianginal agents is shown in Table 284-7.

Antiplatelet Drugs Aspirin is an irreversible inhibitor of platelet cyclooxygenase and thereby interferes with platelet activation. Chronic administration of 75–325 mg orally per day has been shown to reduce coronary events in asymptomatic adult men over age 50, patients with chronic stable angina, and patients who have or have survived unstable angina and myocardial infarction. There is a dose-dependent increase in bleeding when aspirin is used chronically. It is preferable to use an enteric-coated formulation in the range of 75–162 mg/d. Administration of this drug should be considered in all patients with IHD in the absence of gastrointestinal bleeding, allergy, or dyspepsia. Clopidogrel (300–600 mg loading and 75 mg/d) is an oral agent that blocks $P2Y_{12}$ ADP receptor–mediated platelet aggregation. It provides benefits similar to those of aspirin in patients with stable chronic IHD and may be substituted for aspirin if aspirin causes the side effects listed above. Clopidogrel combined with aspirin reduces death and coronary ischemic events in patients with an acute coronary syndrome **(Chap. 285)** and also reduces the risk of thrombus formation in patients undergoing implantation of a stent in a coronary artery **(Chap. 287)**. Alternative antiplatelet agents that block the $P2Y_{12}$ platelet receptor such as prasugrel and ticagrelor have been shown to be more effective than clopidogrel for prevention of ischemic events after placement of a stent for an acute coronary syndrome but are associated with an increased risk of bleeding. Although combined treatment with clopidogrel and aspirin for at least a year is recommended in patients with an acute coronary syndrome treated with implantation of a drug-eluting stent, studies have not shown any benefit from the routine addition of clopidogrel to aspirin in patients with chronic stable IHD.

OTHER THERAPIES

ACE inhibitors and ARBs are widely used in the treatment of survivors of myocardial infarction, patients with hypertension or chronic IHD including angina pectoris, and those at high risk of vascular diseases such as diabetes. The benefits of ACE inhibitors and ARBs are most evident in IHD patients at increased risk, especially if diabetes mellitus or LV dysfunction is present, and those who have not achieved adequate control of blood pressure and LDL cholesterol on beta blockers and statins. However, the routine administration of ACE inhibitors or ARBs to IHD patients who have normal LV function and have achieved blood pressure and LDL goals on other therapies does not reduce the incidence of events and therefore is not cost-effective.

Despite treatment with nitrates, beta blockers, calcium channel blockers, and ranolazine, some patients with IHD continue to experience angina, and additional medical therapy is now available to alleviate their symptoms.

Originally introduced for the management of diabetes mellitus, the SGLT-2 inhibitor drugs have emerged as important agents with cardiovascular and renal protective effects. They promote weight loss, lower blood pressure, and reduce plasma volume—all of which are desirable in patients with IHD. In addition, they decrease

TABLE 284-7 Antianginal Agents

AGENT	COMMON SIDE EFFECTS	CONTRAINDICATIONS	POTENTIAL DRUG INTERACTIONS
Agents That Have a Physiologic Effect			
Short-acting and long-acting nitrates	Headache, flushing, hypotension, syncope and postural hypotension, reflex tachycardia, methemoglobinemia	Hypertrophic obstructive cardiomyopathy	Phosphodiesterase type 5 inhibitors (sildenafil and similar agents), beta-adrenergic blockers, calcium channel blockers
Beta blockers	Fatigue, depression, bradycardia, heart block, bronchospasm, peripheral vasoconstriction, postural hypotension, impotence, masked signs of hypoglycemia	Low heart rate or heart conduction disorder, cardiogenic shock, asthma, severe peripheral vascular disease, decompensated heart failure, vasospastic angina; use with caution in patients with COPD (cardioselective beta blockers may be used if patient receives adequate treatment with long-acting beta agonists)	Heart rate–lowering calcium channel blockers, sinus node or AV conduction depressors
Calcium-channel blockers			
Heart rate–lowering agents	Bradycardia, heart conduction defect, low ejection fraction, constipation, gingival hyperplasia	Cardiogenic shock, severe aortic stenosis, obstructive cardiomyopathy	CYP3A4 substrates (digoxin, simvastatin, cyclosporine)
Dihydropyridine	Headache, ankle swelling fatigue, flushing, reflex tachycardia	Low heart rate or heart rhythm disorder, sick-sinus syndrome, congestive heart failure, low blood pressure	Agents with cardiodepressant effects (beta blockers, flecainide), CYP3A4 substrates
Agents That Affect Myocardial Metabolism			
Ranolazine	Dizziness, constipation, nausea, QT interval prolongation	Liver cirrhosis	CYP3A4 substrates (digoxin, simvastatin, cyclosporine), drugs that prolong the corrected QT interval

Abbreviations: AV, atrioventricular; COPD, chronic obstructive pulmonary disease; CYP3A4, cytochrome P450 3A4.

Source: Data from SE Husted: Lancet 386:691, 2015, and EM Ohman: N Engl J Med 374:1167, 2016.

intraglomerular hypertension and hyperfiltration. Evidence exists that they are helpful in patients with and without diabetes who have a reduced LV ejection fraction.

Colchicine exhibits a number of broad cellular effects (interferes with chemotaxis and phagocytosis of inflammatory cells, reduces the expression of adhesion molecules, modifies cytokine production) that result in an anti-inflammatory effect and may favorably affect the progression of atherosclerosis. In placebo-controlled randomized trials after myocardial infarction, low-dose colchicine (0.5 mg daily) prevented future cardiovascular events; however because it has a narrow therapeutic window, has a long half-life dependent upon renal clearance, is metabolized by CYP3A4, and is a substrate for P-glycoprotein (both of which that may result in drug-drug interactions), additional monitoring is required.

In contrast, nonsteroidal anti-inflammatory drug (NSAID) use in patients with IHD may be associated with a small but finite increased risk of myocardial infarction and mortality. For this reason, they generally should be avoided in IHD patients. If they are required for symptom relief, it is advisable to coadminister aspirin and strive to use an NSAID associated with the lowest risk of cardiovascular events, in the lowest dose required, and for the shortest period of time.

Nicorandil opens ATP-sensitive potassium channels in myocytes, leading to a reduction of free intracellular calcium ions. It is typically administered orally in a dose of 20 mg twice daily for prevention of angina. Nicorandil is not available for use in the United States but is used in several other countries and is recommended as second-line treatment in the European chronic coronary disease (CCD) guidelines. Similarly, trimetazidine, which improves mitochondrial metabolism through inhibition of myocardial fatty acid uptake and oxidation and consequent stimulation of glucose oxidation, is recommended as a second-line antianginal in CCD and is available in many countries outside the United States.

Ivabradine (2.5–7.5 mg orally twice daily) is a specific sinus node inhibiting agent that may be helpful for preventing cardiovascular events in patients with IHD who have a resting heart rate ≥70 beats/min (alone or in combination with a beta blocker) *and* LV systolic dysfunction. However, the data are mixed regarding its clinical benefit, with the most recent U.S. CCD guideline recommending against its use in patients with normal LV function due to an increased risk of death or myocardial infarction).

Angina and Heart Failure Transient LV failure with angina can be controlled by the use of nitrates. For patients with established congestive heart failure, the increased LV wall tension raises myocardial oxygen demand. See **Chap. 265** for the "pillars" of treatment for heart failure. Treatment of congestive heart failure **(Chap. 264)** reduces heart size, wall tension, and myocardial oxygen demand, which helps control angina and ischemia. If the symptoms and signs of heart failure are controlled, an effort should be made to use beta blockers not only for angina but because trials in heart failure have shown significant improvement in survival. A trial of the intravenous ultra-short-acting beta blocker esmolol may be useful to establish the safety of beta blockade in selected patients. Nocturnal angina often can be relieved by the treatment of heart failure.

The combination of congestive heart failure and angina in patients with IHD usually indicates a poor prognosis and warrants serious consideration of cardiac catheterization and coronary revascularization.

CORONARY REVASCULARIZATION

Clinical trials have confirmed that with the initial diagnosis of stable IHD, it is first appropriate to initiate a medical regimen as described above. Revascularization should be considered in the presence of unstable phases of the disease, intractable symptoms, high-risk coronary anatomy, diabetes, and impaired LV function. *Revascularization should be employed in conjunction with but not replace the continuing need to modify risk factors and assess medical therapy.* An algorithm for integrating medical therapy and revascularization options in patients with IHD is shown in **Fig. 284-4**.

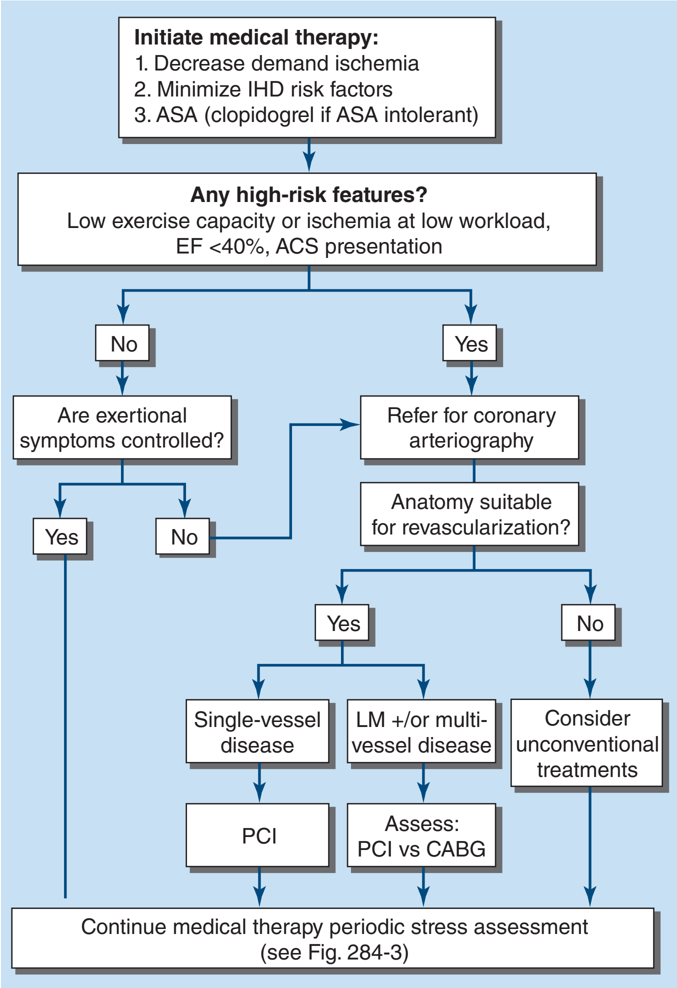
Initiate medical therapy:
1. Decrease demand ischemia
2. Minimize IHD risk factors
3. ASA (clopidogrel if ASA intolerant)
Any high-risk features?
Low exercise capacity or ischemia at low workload, EF <40%, ACS presentation
No
Yes
Are exertional symptoms controlled?
Refer for coronary arteriography
Yes
No
Anatomy suitable for revascularization?
Yes
No
Single-vessel disease
LM +/or multi-vessel disease
Consider unconventional treatments
PCI
Assess: PCI vs CABG
Continue medical therapy periodic stress assessment (see Fig. 284-3)

FIGURE 284-4 Algorithm for management of a patient with ischemic heart disease. All patients should receive the core elements of medical therapy as shown at the top of the algorithm. If high-risk features are present, as established by the clinical history, exercise test data, and imaging studies, the patient should be referred for coronary arteriography. Based on the number and location of the diseased vessels and their suitability for revascularization, the patient is treated with a percutaneous coronary intervention (PCI) or coronary artery bypass graft (CABG) surgery or should be considered for unconventional treatments. See text for further discussion. ACS, acute coronary syndrome; ASA, aspirin; EF, ejection fraction; IHD, ischemic heart disease; LM, left main.

■ PERCUTANEOUS CORONARY INTERVENTION

(See also Chap. 287) PCI involving balloon dilatation usually accompanied by coronary stenting is widely used to achieve revascularization of the myocardium in patients with symptomatic IHD and suitable stenoses of epicardial coronary arteries. Whereas patients with stenosis of the left main coronary artery and those with three-vessel IHD (especially with diabetes and/or impaired LV function) who require revascularization are best treated with CABG, PCI is widely employed in patients with symptoms and evidence of ischemia due to stenoses of one or two vessels and even in selected patients with three-vessel disease (and, perhaps, in some patients with left main disease) and may offer many advantages over surgery.

Indications and Patient Selection The most common clinical indication for PCI is symptom-limiting angina pectoris, despite medical therapy, accompanied by evidence of ischemia during a stress test. PCI is more effective than medical therapy for the relief of angina. PCI improves outcomes in patients with unstable angina or when used early in the course of myocardial infarction with and without cardiogenic shock. However, in patients with stable exertional angina, clinical trials have confirmed that PCI does not reduce the occurrence of death or myocardial infarction compared to optimum medical therapy. PCI can be used to treat stenoses in native coronary arteries as well as in bypass grafts in patients who have recurrent angina after CABG.

Risks When coronary stenoses are discrete and symmetric, two and even three vessels can be treated in sequence. However, case selection is essential to avoid a prohibitive risk of complications, which are usually due to dissection or thrombosis with vessel occlusion, uncontrolled ischemia, and ventricular failure **(Chap. 287)**. Oral aspirin, a $P2Y_{12}$ antagonist, and an antithrombin agent are given to reduce coronary thrombus formation. Left main coronary artery stenosis generally is regarded as a lesion that should be treated with CABG. In selected cases such as patients with prohibitive surgical risks, PCI of an unprotected left main can be considered, but such a procedure should be performed only by a highly skilled operator; importantly, there are regional differences in the use of this approach internationally.

Efficacy Primary success, with relief of angina, is achieved in >95% of cases. In diseased vein grafts, procedural success has been improved by the use of an embolic protection device, when feasible, to prevent ischemia and infarction. The use of current-generation drug-eluting stents that locally deliver antiproliferative drugs have a lower rate of restenosis (<5%) than the initial bare metal stents and are now the standard of care in PCI. Of note, however, the delayed endothelial healing in the region of a drug-eluting stent also extends the period during which the patient is at risk for subacute stent thrombosis. Aspirin should be administered indefinitely and a $P2Y_{12}$ antagonist daily (dual antiplatelet therapy [DAPT]) ideally for 1 year after implantation of a drug-eluting stent. Evidence exists of a benefit of continuing DAPT for up to 30 months, albeit at the cost of a higher risk of bleeding. Shorter courses of DAPT may be used in patients who are at high bleeding risk **(see Chap. 69)** or who have a long-term indication for oral anticoagulation **(see Chap. 121)**.

Efforts are underway to develop new antithrombotic regimens to reduce the risk of bleeding. These include (1) shortening the duration of DAPT by eliminating aspirin after 1 or 3 months and continuing a $P2Y_{12}$ antagonist alone, (2) de-escalating from a potent $P2Y_{12}$ inhibitor (i.e., prasugrel, ticagrelor) after 1 month to clopidogrel or reduced-dose prasugrel 5 mg daily, and (3) switching from DAPT to dual pathway inhibition with an antiplatelet agent and a low-dose direct oral anticoagulant. While each of these has been compared to standard DAPT regimens, such studies tend to be unpowered for ischemic events, and these alternative strategies have not been compared to one another.

When a situation arises in which temporary discontinuation of antiplatelet therapy is necessary, the clinical circumstances should be reviewed with the operator who performed the PCI and a coordinated plan should be established for minimizing the risk of late stent thrombus; central to this plan is the discontinuation of antiplatelet therapy for the shortest acceptable period. The risk of stent thrombosis is dependent on stent size and length, complexity of the lesions, age, diabetes, and technique. However, compliance with DAPT and individual responsiveness to platelet inhibition are very important factors as well.

Successful PCI produces effective relief of angina in >95% of cases. The majority of patients with symptomatic IHD who require revascularization can be treated initially by PCI. Successful PCI is less invasive and expensive than CABG and permits savings in the *initial* cost of care. Successful PCI avoids the risk of stroke associated with CABG surgery and allows earlier return to work and resumption of an active life. However, the early health-related and economic benefits of PCI are reduced over time because of the greater need for follow-up and the increased need for repeat procedures. When directly compared in patients with diabetes or three-vessel or left main CAD, CABG was superior to PCI in preventing major adverse cardiac or cerebrovascular events over a 12-month follow-up.

■ CORONARY ARTERY BYPASS GRAFTING

Anastomosis of one or both of the internal mammary arteries or a radial artery to the coronary artery distal to the obstructive lesion is the preferred procedure. For additional obstructions that cannot be bypassed by an artery, a section of a vein (usually the saphenous) is used to form a venous bypass conduit between the aorta and the coronary artery distal to the obstructive lesion.

Although some indications for CABG are controversial, certain areas of agreement exist:

1. The operation is relatively safe, with mortality rates <1% in patients without serious comorbid disease and normal LV function and when the procedure is performed by an experienced surgical team.
2. Intraoperative and postoperative mortality rates increase with the severity of ventricular dysfunction, comorbidities, age ≥80 years, and lack of surgical experience. The effectiveness and risk of CABG vary widely depending on case selection and the skill and experience of the surgical team.
3. Occlusion of *venous* grafts is observed in 10–20% of patients during the first postoperative year and in ~2% per year during 5- to 7-year follow-up and 4% per year thereafter. Long-term patency rates are considerably higher for internal mammary and radial artery implantations than for saphenous vein grafts. In patients with left anterior descending coronary artery obstruction, survival is better when coronary bypass involves the internal mammary artery rather than a saphenous vein. Graft patency and outcomes are improved by meticulous treatment of risk factors, particularly dyslipidemia.
4. Angina is abolished or greatly reduced in ~90% of patients after complete revascularization. Although this usually is associated with graft patency and restoration of blood flow, the pain may also have been alleviated as a result of infarction of the ischemic segment, denervation due to median sternotomy, or a placebo effect.
5. Survival may be improved by operation in patients with stenosis of the left main coronary artery as well as in patients with three- or two-vessel disease with significant obstruction of the proximal left anterior descending coronary artery. The survival benefit is greater in patients with abnormal LV function (ejection fraction <35). Survival *may* also be improved in the following patients: (a) patients with obstructive CAD who have survived sudden cardiac death or sustained ventricular tachycardia; (b) patients who have undergone previous CABG and have multiple saphenous vein graft stenoses, especially of a graft supplying the left anterior descending coronary artery; and (c) patients with recurrent stenosis after PCI and high-risk criteria on noninvasive testing.
6. Minimally invasive CABG through a small thoracotomy and/or off-pump surgery can reduce morbidity and shorten convalescence in suitable patients but does not appear to reduce significantly the risk of neurocognitive dysfunction postoperatively.
7. Among patients with type 2 diabetes mellitus and multivessel coronary disease, CABG surgery plus optimal medical therapy is superior to optimal medical therapy alone in preventing major cardiovascular events, a benefit mediated largely by a significant reduction in nonfatal myocardial infarction. The benefits of CABG are especially evident in patients with diabetes mellitus treated with an insulin-sensitizing strategy as opposed to an insulin-providing strategy. CABG has also been shown to be superior to PCI (including the use of drug-eluting stents) in preventing death, myocardial infarction, and repeat revascularization in patients with diabetes mellitus and multivessel IHD.

Indications for CABG usually are based on the severity of symptoms, coronary anatomy, and ventricular function. The ideal candidate has no other complicating disease and has troublesome or disabling angina that is not adequately controlled by medical therapy or does not tolerate medical therapy. Great symptomatic benefit can be anticipated if a patient wishes to lead a more active life and has severe stenoses of two or three epicardial coronary arteries with objective evidence of myocardial ischemia as a cause of the chest discomfort. Congestive heart failure and/or LV dysfunction, advanced age (≥80 years), reoperation, urgent need for surgery, and the presence of diabetes mellitus are all associated with a higher perioperative mortality rate.

LV dysfunction can be due to noncontractile or hypocontractile segments that are viable but are chronically ischemic (hibernating myocardium). As a consequence of chronic reduction in myocardial blood flow, these segments downregulate their contractile function.

They can be detected by using radionuclide scans of myocardial perfusion and metabolism, PET, cardiac MRI, or delayed scanning with thallium-201 or by improvement of regional functional impairment provoked by low-dose dobutamine. In such patients, revascularization improves myocardial blood flow, can return function, and can improve survival.

The Choice Between PCI and CABG All the clinical characteristics of each individual patient must be used to decide on the method of revascularization (e.g., LV function, diabetes, lesion complexity). A number of randomized clinical trials have compared PCI and CABG in patients with multivessel CAD who were suitable technically for both procedures. The redevelopment of angina requiring repeat coronary angiography and repeat revascularization is higher with PCI. This is a result of restenosis in the stented segment (a problem largely solved with drug-eluting stents) and the development of new stenoses in unstented portions of the coronary vasculature. It has been argued that PCI with stenting focuses on culprit lesions, whereas a bypass graft to the target vessel also provides a conduit around future culprit lesions proximal to the anastomosis of the graft to the native vessel **(Fig. 284-5)**. By contrast, stroke rates are lower with PCI.

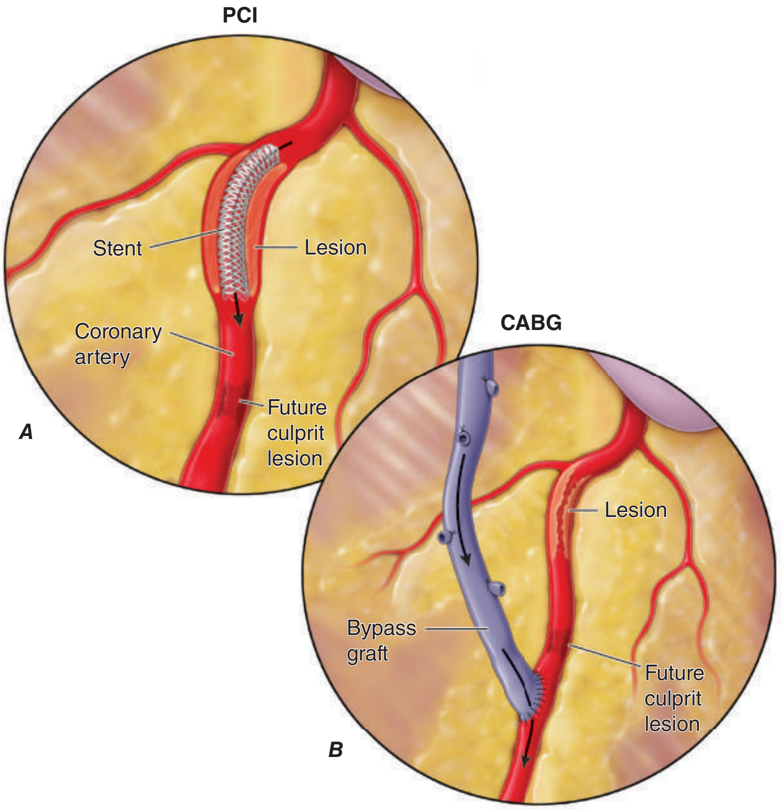

FIGURE 284-5 Difference in the approach to the lesion with percutaneous coronary intervention (PCI) and coronary artery bypass grafting (CABG). PCI is targeted at the "culprit" lesion or lesions, whereas CABG is directed at the epicardial vessel, including the culprit lesion or lesions and future culprits, proximal to the insertion of the vein graft, a difference that may account for the superiority of CABG, at least in the intermediate term, in patients with multivessel disease. (*From The New England Journal of Medicine, Quantitative Determinants of the Outcome of Asymptomatic Mitral Regurgitation, M Enriquez-Sarano et al. 352, 2235. Copyright © 2005. Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.*)

Based on available evidence, it is now recommended that patients with lifestyle-limiting angina despite guideline-directed medical management and therapy be considered for coronary revascularization. Patients with single- or two-vessel disease with normal LV function and anatomically suitable lesions ordinarily are advised to undergo PCI **(Chap. 287)**. For patients who are poor candidates for surgery, it is reasonable to choose PCI over CABG to improve symptoms and reduce major adverse cardiac events.

In patients with left main disease associated with high-complexity CAD, CABG remains preferred to PCI to improve survival. Similarly, in patients with diabetes and multivessel disease involving the left anterior descending artery, CABG (with a left internal mammary conduit) is preferred over PCI, although in patients with low- or intermediate-complexity CAD, PCI may be considered.

Patients with three-vessel disease (or two-vessel disease that includes the proximal left descending coronary artery) and impaired global LV function (LV ejection fraction <50%) or diabetes mellitus and those with left main CAD or other lesions unsuitable for catheter-based procedures should be considered for CABG as the initial method of revascularization. In light of the complexity of the decision-making, it is desirable to have a multidisciplinary team, including a cardiologist and a cardiac surgeon in conjunction with the patient's primary care physician, provide input along with ascertaining the patient's preferences before committing to a particular revascularization option.

■ UNCONVENTIONAL TREATMENTS FOR IHD

On occasion, clinicians will encounter a patient who has persistent, disabling angina despite maximally tolerated medical therapy and for whom revascularization is not an option (e.g., small diffusely diseased vessels not amenable to stent implantation or acceptable targets for bypass grafting). In such situations, unconventional treatments should be considered.

Enhanced external counterpulsation utilizes pneumatic cuffs on the lower extremities to provide diastolic augmentation and systolic unloading of blood pressure to decrease cardiac work and oxygen consumption while enhancing coronary blood flow. Clinical trials have shown that regular application improves angina, exercise capacity, and regional myocardial perfusion.

Neuromodulation (e.g., spinal cord simulation, transcutaneous or subcutaneous electrical neural stimulation, sympathectomy) and coronary sinus constriction (using a reducer device) have shown promise in small studies. Experimental approaches, such as stem cell therapies and cardiac repair with small noncoding RNA molecules (miRNA), are also under active study.

ASYMPTOMATIC (SILENT) ISCHEMIA

Obstructive CAD, acute myocardial infarction, and transient myocardial ischemia can occur in the absence of symptoms. During continuous ambulatory ECG monitoring, the majority of ambulatory patients with typical chronic stable angina are found to have objective evidence of myocardial ischemia (ST-segment depression) during episodes of chest discomfort while they are active outside the hospital. In addition, many of these patients also have more frequent episodes of asymptomatic ischemia. Frequent episodes of ischemia (symptomatic and asymptomatic) during daily life appear to be associated with an increased likelihood of adverse coronary events (death and myocardial infarction). In addition, patients with asymptomatic ischemia after a myocardial infarction are at greater risk for a second coronary event. The widespread use of exercise ECG during routine examinations has also identified some of these previously unrecognized patients with asymptomatic CAD. Longitudinal studies have demonstrated an increased incidence of coronary events in asymptomatic patients with positive exercise tests.

TREATMENT

Asymptomatic Ischemia

The management of patients with asymptomatic ischemia must be individualized. When coronary disease has been confirmed, the aggressive treatment of hypertension and dyslipidemia is essential and will decrease the risk of infarction and death. In addition, the physician should consider the following: (1) the degree of positivity of the stress test, particularly the stage of exercise at which ECG signs of ischemia appear; the magnitude and number of the ischemic zones of myocardium on imaging; and the change in LV ejection fraction that occurs on radionuclide ventriculography or echocardiography during ischemia and/or during exercise; (2) the ECG leads showing a positive response, with changes in the anterior precordial leads indicating a less favorable prognosis than changes in the inferior leads; and (3) the patient's age, occupation, and general medical condition.

Most would agree that an asymptomatic 45-year-old commercial airline pilot with significant (0.4-mV) ST-segment depression in leads V_1 to V_4 during mild exercise should undergo coronary arteriography, whereas an asymptomatic, sedentary 85-year-old retiree with 0.1-mV ST-segment depression in leads II and III during maximal activity need not. However, there is no consensus about the most appropriate approach in the large majority of patients for whom the situation is less extreme. Asymptomatic patients with silent ischemia, three-vessel CAD, and impaired LV function may be considered appropriate candidates for CABG.

The treatment of risk factors, particularly lipid lowering and blood pressure control as described above, and the use of aspirin, statins, and beta blockers after infarction have been shown to reduce events and improve outcomes in asymptomatic as well as symptomatic patients with ischemia and proven CAD. Although the incidence of asymptomatic ischemia can be reduced by treatment with beta blockers, calcium channel blockers, and long-acting nitrates, it is not clear whether this is necessary or desirable in patients who have not had a myocardial infarction.

■ FURTHER READING

American Diabetes Association Professional Practice Committee: Cardiovascular disease and risk management: Standards of care in diabetes—2024. Diabetes Care 47:S179, 2024.

Antman EM, Braunwald E: Managing stable ischemic heart disease. N Engl J Med 382:1468, 2020.

Gaba PK et al: Association between achieved low-density lipoprotein cholesterol levels and long-term cardiovascular and safety outcomes: An analysis of FOURIER-OLE. Circulation 147:1192, 2023.

Gulati M et al: 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR Guideline for the Evaluation and Diagnosis of Chest Pain: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 144:e368, 2021.

Hokimoto S et al: JCS/CVIT/JCC 2023 guideline focused update on diagnosis and treatment of vasospastic angina (coronary spastic angina) and coronary microvascular dysfunction. Circ J 87:879, 2023.

Knuuti J et al: 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes. Eur Heart J 41:407, 2020.

Lloyd-Jones DM et al: 2022 ACC expert consensus decision pathway on the role of nonstatin therapies for LDL-cholesterol lowering in the management of atherosclerotic cardiovascular disease risk. J Am Coll Cardiol 80:1366, 2022.

Maron DJ et al: Initial invasive or conservative strategy for stable coronary disease. N Engl J Med 382:1395, 2020.

Martin SS et al: 2024 Heart disease and stroke statistics: A report of US and global data from the American Heart Association. Circulation 149:e347, 2024.

Marx N et al: 2023 ESC Guidelines for the management of cardiovascular disease in patients with diabetes. Eur Heart J 44:4043, 2023.

Mensah GA et al: Global burden of cardiovascular diseases and risks, 1990-2022. J Am Coll Cardiol 82:2350, 2023.

Rao SV et al: Chronic coronary disease guidelines. Circulation 148:729, 2023.

Virani SS et al: 2023 AHA/ACC/ACCP/ASPC/NLA/PCNA Guideline for the Management of Patients With Chronic Coronary Disease: A report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines [published correction appears in Circulation 148:e148, 2023] [published correction appears in Circulation 148:e186, 2023]. Circulation 148:e9, 2023.

285

Non-ST-Segment Elevation Acute Coronary Syndrome (Non-ST-Segment Elevation Myocardial Infarction and Unstable Angina)

Robert P. Giugliano, Christopher P. Cannon, Eugene Braunwald

Patients with acute coronary syndrome (ACS) are commonly classified into two groups to facilitate evaluation and management, namely patients with acute myocardial infarction (MI) with ST-segment elevation (STEMI) on their presenting electrocardiogram (ECG) **(see Chap. 286)** and those with non-ST-segment elevation acute coronary syndrome (NSTE-ACS). The latter include patients with non-ST-segment elevation MI (NSTEMI), who, by definition, have evidence of myocyte necrosis, and those with unstable angina (UA), who do not **(Fig. 285-1)**.

The incidence of NSTEMI is rising due to the increasing burden of obesity, diabetes, and chronic kidney disease in an aging population and the increasing detection of myocardial necrosis by troponin (see below), whereas the incidence of STEMI is declining due to greater use of aspirin, statins, and less smoking. Among patients with NSTE-ACS, the proportion with NSTEMI is increasing while that with UA is falling because of the wider use of highly sensitive troponin (hsTn) assays (see below) with enhanced detection of myocyte necrosis, thereby reclassifying UA to NSTEMI.

■ PATHOPHYSIOLOGY

NSTE-ACS is caused by an imbalance between myocardial oxygen supply and demand, most commonly resulting from one or more of three processes that lead to coronary arterial thrombosis: (1) plaque fissure with inflammation—the inflammatory response is reflected by an increased activity of effector T cells as part of an adaptive immunity dysregulation; (2) plaque fissure without inflammation; and (3) plaque erosion, which is present in at least one-third of ACS and is recognized with increasing frequency **(Fig. 285-2)**. The so-called "vulnerable plaques" responsible for ACS may show an eccentric stenosis with scalloped or overhanging edges and a narrow neck on coronary angiography. Such plaques usually are composed of a lipid-rich core with a thin fibrous cap. Patients with NSTE-ACS frequently have multiple such plaques that are at risk of disruption. A fourth process, without thrombosis, may be caused by epicardial or microvascular spasm or increased myocardial oxygen demand, typically in the presence of fixed epicardial coronary obstruction.

Among patients with NSTE-ACS studied at angiography, <10% have stenosis of the left main coronary artery, and approximately 35% have three-vessel coronary artery disease, 25% have two-vessel disease, 20% have single-vessel disease, and 10% have no apparent critical epicardial coronary artery stenosis; some of the latter may have obstruction of the coronary microcirculation and/or spasm of the epicardial vessels.

■ CLINICAL PRESENTATION

Diagnosis The diagnosis of NSTE-ACS is based largely on a combination of the history and clinical findings (age, ECG, and circulating troponin) **(Table 285-1, Fig. 285-3)**.

History and Physical Examination Typically, chest discomfort is severe and has at least one of three features: (1) occurrence at rest (or with minimal exertion), lasting >10 min; (2) of relatively recent onset (i.e., within the prior 2 weeks); and/or (3) a crescendo pattern, i.e., distinctly more severe, prolonged, or frequent than previous episodes.

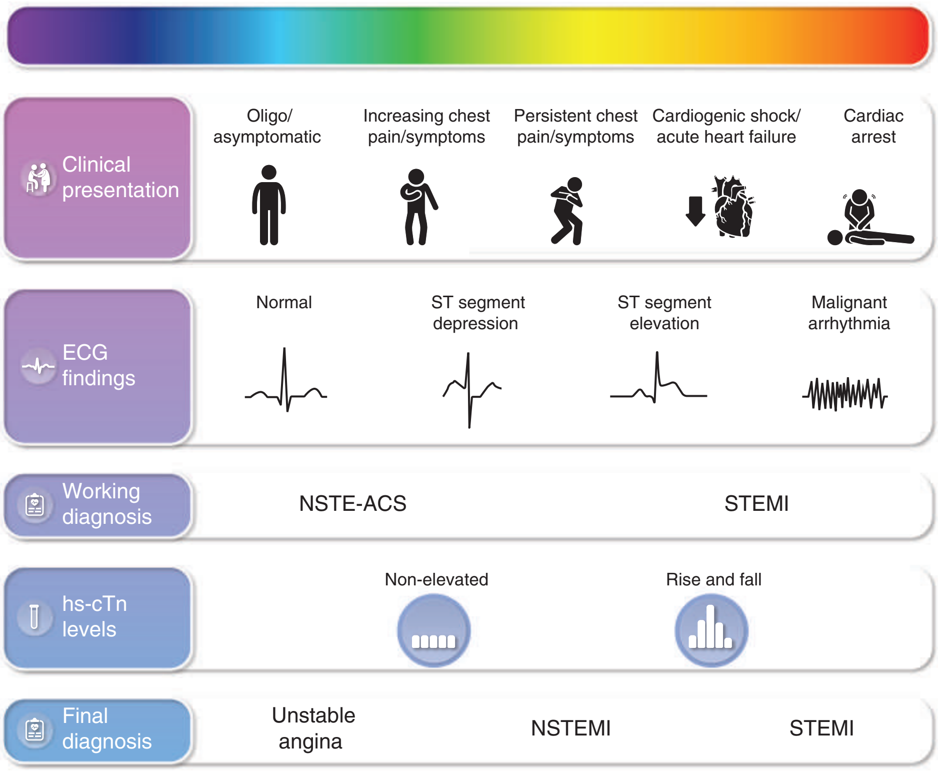
Clinical presentation
Oligo/ asymptomatic
Increasing chest pain/symptoms
Persistent chest pain/symptoms
Cardiogenic shock/ acute heart failure
Cardiac arrest
ECG findings
Normal
ST segment depression
ST segment elevation
Malignant arrhythmia
Working diagnosis
NSTE-ACS
STEMI
hs-cTn levels
Non-elevated
Rise and fall
Final diagnosis
Unstable angina
NSTEMI
STEMI

FIGURE 285-1 Spectrum of acute coronary syndromes. The spectrum of clinical presentations, electrocardiographic findings, and high-sensitivity cardiac troponin levels in patients with acute coronary syndrome. ACS, acute coronary syndrome; ECG, electrocardiogram; hs-cTn, high-sensitivity cardiac troponin; NSTE-ACS, non-ST-segment elevation acute coronary syndrome; NSTEMI, non-ST-segment elevation myocardial infarction; STEMI, ST-segment elevation myocardial infarction. *(RA Byrne et al: 2023 ESC guideline for the management of acute coronary syndromes. Eur Heart J 44:3720, 2023, by permission of Oxford university Press.)*

The diagnosis of NSTEMI is established if a patient with any of these features (without ECG ST-segment elevations) develops evidence of myocardial necrosis, as reflected in abnormally elevated levels of circulating troponin, in the absence of another explanation (see below). The chest discomfort is typically located in the substernal region and radiates to the left arm, left shoulder, and/or superiorly to the neck and jaw. Anginal equivalents such as dyspnea, epigastric discomfort, nausea, or weakness may occur instead of chest discomfort. These equivalents are more frequently reported by women, the elderly, and patients with diabetes mellitus. The physical examination resembles that in patients with stable angina (see Chap. 284) and may be unremarkable. However, if the patient has a large area of myocardial ischemia or a large NSTEMI, the physical findings can include diaphoresis; pale, cool skin; sinus tachycardia; a third and/or fourth heart sound; basilar rales; hypotension; and in severe cases, cardiogenic shock. An important goal of the physical examination is to identify potential alternative life-threatening diagnoses, e.g., pulmonary embolism (see Chap. 290), aortic dissection (see Chap. 291), and cardiac tamponade (see Chap. 281),

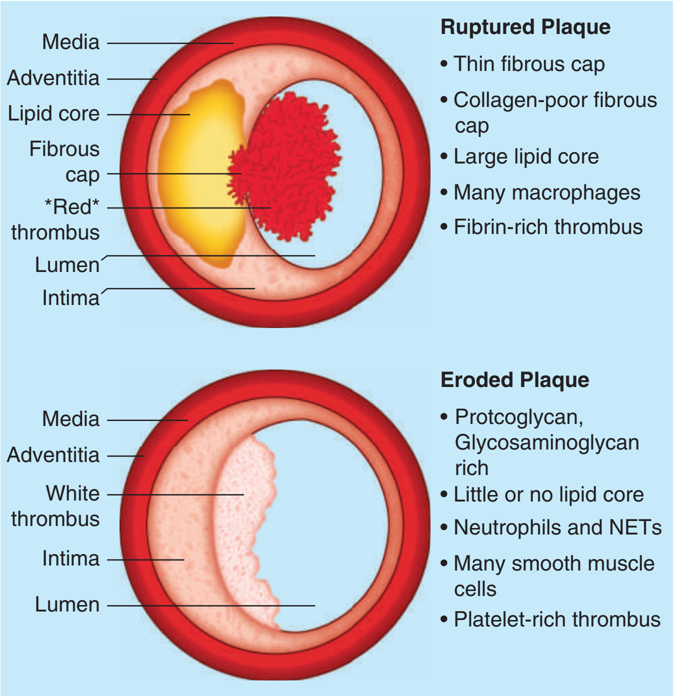
Media
Adventitia
Lipid core
Fibrous cap
Red thrombus
Lumen
Intima
Ruptured Plaque
• Thin fibrous cap
• Collagen-poor fibrous cap
• Large lipid core
• Many macrophages
• Fibrin-rich thrombus
Media
Adventitia
White thrombus
Intima
Lumen
Eroded Plaque
• Protcoglycan, Glycosaminoglycan rich
• Little or no lipid core
• Neutrophils and NETs
• Many smooth muscle cells
• Platelet-rich thrombus

FIGURE 285-2 Comparison of the characteristics of human atheromata complicated by thrombosis and causing acute coronary syndrome. The column on the left highlights some of the characteristics demonstrated by analyses of human coronary arterial lesions that have undergone thrombosis by these two diverse mechanisms. NETs, neutrophil extracellular traps. *(Reproduced with permission from P Libby et al: Reassessing the mechanisms of acute coronary syndromes the "vulnerable plaque" and superficial erosion. Circ Res 124:150, 2019.)*

TABLE 285-1 TIMI Risk Score for NSTE-ACS

RISK MARKERS	
• Age ≥65 years • Known CAD (≥50% stenosis) • ST deviation >0.5 mm on presenting ECG	• ↑ cardiac markers • ≥2 original episodes in prior 24 h • Prior angina • ≥3 CAD risk factors
NO. OF RISK MARKERS	**INCIDENCE OF ADVERSE CARDIAC EVENTS[a] (%)**
0/1	5
2	8
3	13
4	20
5	26
6/7	41

[a] Risk at 14 days of death, new or recurrent myocardial infarction, or severe recurrent ischemia requiring urgent revascularization.

Abbreviations: CAD, coronary artery disease; ECG, electrocardiogram; NSTE-ACS, non-ST-segment elevation acute coronary syndrome; TIMI, Thrombolysis in Myocardial Infarction.

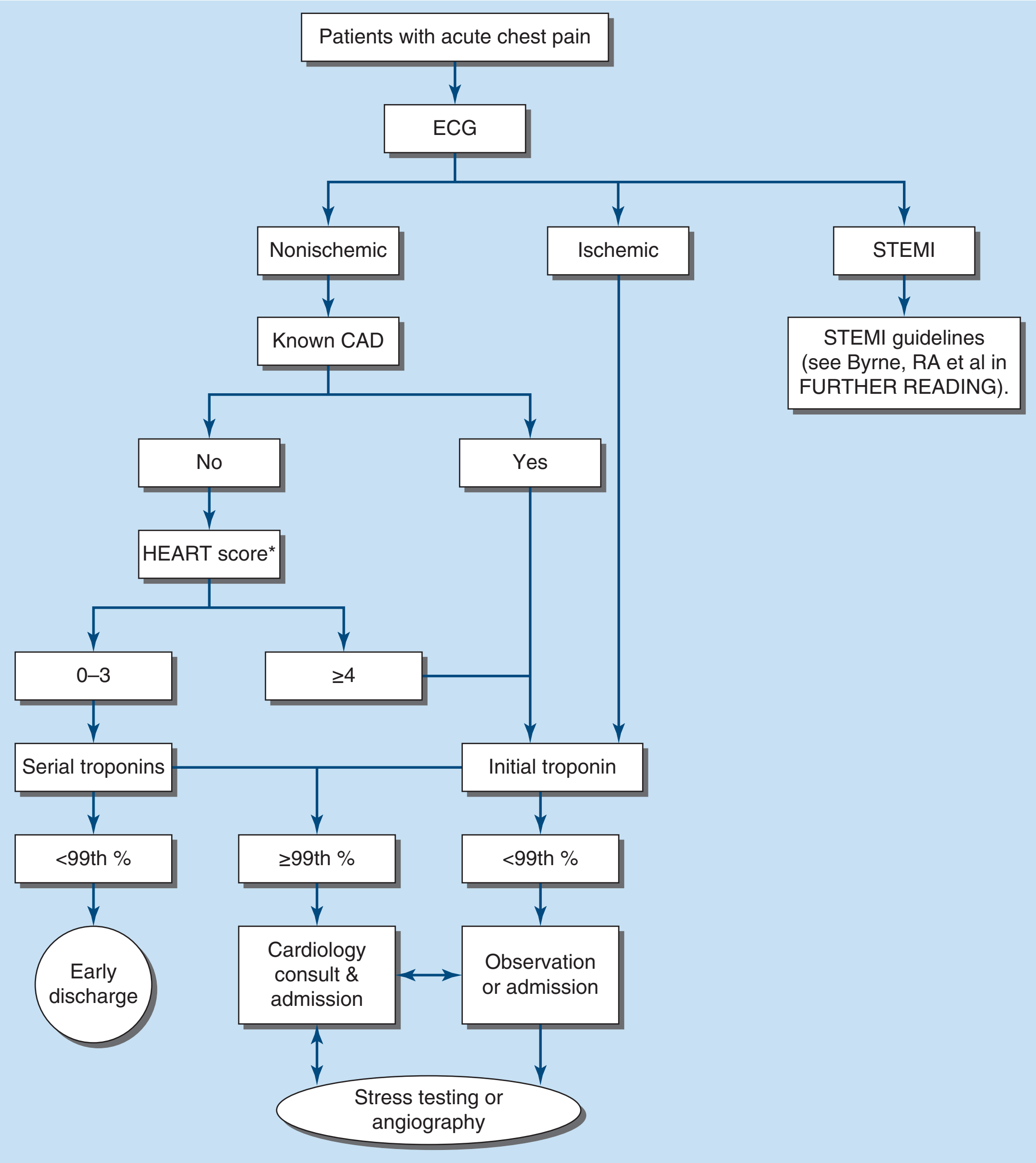

FIGURE 285-3 The HEART pathway for evaluation of acute chest pain. CAD, coronary artery disease; ECG, electrocardiogram; STEMI, ST-segment elevation myocardial infarction. *The Heart Score assigns 0, 1, or 2 points depending on the extent of abnormality for each of the History, ECG, Age, Risk factors, and Troponin (*Six AJ, Backus BE, and Kelder JC. Chest pain in the emergency room: value of the HEART score. Neth Heart J 16:191, 2008*). *(Reproduced with permission from JL Januzzi et al: Recommendations for institutions transitioning to high-sensitivity troponin testing: JACC scientific expert panel. J Am Coll Cardiol 73:1059-1077, 2019.)*

given the vastly different treatment plans required to avoid major morbidity and death.

Electrocardiogram New ST-segment depression occurs in about one-third of patients with NSTE-ACS. It may be transient but can persist for as long as several days following NSTEMI. T-wave changes are more common but are a less specific sign of ischemia unless they are new and deep T-wave inversions (≥0.3 mV).

Cardiac Biomarkers Patients with NSTEMI have elevated biomarkers of necrosis, such as cardiac troponin (cTn) I or T (cTnI or cTnT). High-sensitive (hs) cTns are sensitive, relatively specific, and the preferred markers of myocardial necrosis. Elevated levels of cTn with a dynamic early change distinguish patients with NSTEMI from those with UA. In patients with NSTEMI, there is a characteristic temporal rise of the plasma concentration, peaking at 12–24 h after onset of symptoms and gradually decreasing thereafter. There is a direct relationship between the degree of elevation and mortality. The 1-h rapid rule-out MI algorithm (no abnormal elevation of hsTn at 0 or 1 h after presentation) has been recommended by recent practice guidelines. It is important to distinguish myocardial injury from myocardial necrosis; the former is defined by elevations of cTn >99th percentile of the upper reference limit in patients *without* a clear clinical history or ECG features of acute myocardial ischemia. Myocardial injury may be caused by a variety of noncardiac and cardiac conditions other than MI **(Table 285-2)**.

■ RISK STRATIFICATION

Patients with documented NSTE-ACS exhibit a wide spectrum of early (30 days) risk of death, ranging from 1 to 10%, and a recurrent ACS rate of 5–15% during the first year. Assessment of risk can be accomplished by one of several clinical risk scoring systems, including those developed from the Thrombolysis in Myocardial Infarction (TIMI) Trials, the Global Registry of Acute Coronary Event (GRACE), and the HEART (*h*istory, *e*lectrocardiogram, *a*ge, *r*isk factors, *t*roponin) score (Fig. 285-3). Multimarker strategies including hsTn and brain natriuretic peptides are now gaining favor, both to define more fully the pathophysiologic mechanisms underlying a patient's presentation and to stratify the patient's risk further. Early risk assessment is useful in identifying patients who would derive the greatest benefit from an early invasive strategy (see below).

TABLE 285-2 Reasons for the Elevation of Cardiac Troponin Values as a Result of Myocardial Injury

Myocardial injury related to acute myocardial infarction
Atherosclerotic plaque disruption or erosion with thrombosis
Myocardial injury related to acute myocardial ischemia because of oxygen supply/demand imbalance
Reduced myocardial perfusion
• Coronary artery spasm, microvascular dysfunction
• Coronary embolism
• Coronary artery dissection
• Sustained bradyarrhythmia
• Hypotension or shock
• Respiratory failure
• Severe anemia
Increased myocardial oxygen demand
• Sustained tachyarrhythmia
• Severe hypertension
Other causes of myocardial injury
Cardiac conditions
• Heart failure
• Myocarditis
• Cardiomyopathy (any type)
• Takotsubo syndrome
• Recent coronary revascularization
• Cardiac procedure other than revascularization
• Catheter ablation
• Defibrillator shocks
• Cardiac contusion
Systemic conditions
• Sepsis
• Chronic kidney disease
• Stroke, subarachnoid hemorrhage
• Pulmonary embolism
• Infiltrative diseases, e.g., amyloidosis, sarcoidosis
• Chemotherapeutic agents
• Critical illness
• Strenuous exercise

Note: For a more comprehensive listing, see the Fourth Universal Definition of Myocardial Infarction (source).

Source: Reproduced with permission from K Thygesen et al: Fourth universal definition of myocardial infarction (2018). Circulation 138:2231, 2018.

TREATMENT

Non-ST-Segment Elevation Acute Coronary Syndrome

Patients with a low likelihood of ischemia can usually be managed in an emergency department or a dedicated "chest pain unit." Evaluation of such patients includes clinical monitoring for recurrent ischemic discomfort and continuous monitoring of ECGs, serial monitoring of hsTn at presentation, 1 h, and, if necessary, 3 h. In patients for whom there is diagnostic uncertainty or intermediate-risk patients, stress testing to detect and grade ischemia or cardiac computed tomography angiography to assess epicardial coronary artery obstruction may be considered **(see Chaps. 248 and 284)**.

MEDICAL TREATMENT

Patients who "rule-in" for NSTE-ACS by clinical features, hsTn, or ST-T–wave changes on the ECG should be admitted to the hospital. Patients should be placed on bed rest with continuous ECG monitoring for ST-segment deviation and cardiac arrhythmias, preferably on a specialized cardiac unit. Ambulation is permitted if the patient shows no recurrence of ischemia (symptoms or ECG changes) and does not develop an elevation of cTn for 24 h.

Medical therapy consists of an acute phase focused on the clinical symptoms and stabilization of the culprit lesion(s) and a longer-term phase that involves therapies directed at the prevention of disease progression and future recurrent NSTE-ACS.

ANTI-ISCHEMIC TREATMENT (TABLE 285-3)

To provide relief of pain and discomfort, initial treatment, in addition to bed rest, includes nitrates, β-adrenergic blockers, and inhaled oxygen in patients with hypoxemia (arterial O_2 saturation <90%) and/or in those with heart failure and rales.

Nitrates Nitroglycerin should first be given sublingually or by buccal spray (0.3–0.6 mg) if the patient is experiencing ischemic discomfort. If symptoms persist after three doses given 5 min apart, intravenous nitroglycerin (5–10 µg/min, using nonabsorbing tubing) is recommended. The infusion rate may be increased by 10 µg/min every 3–5 min until symptoms are relieved, systolic arterial pressure falls to <90 mmHg, or the dose reaches 200 µg/min. Topical or oral nitrates **(see Chap. 284)** can be used when the pain has resolved, or they may replace intravenous nitroglycerin when the patient has been symptom-free for 12–24 h. The only absolute contraindications to the use of nitrates are hypotension or the recent use of a phosphodiesterase type 5 (PDE-5) inhibitor, sildenafil or vardenafil (within 24 h), or tadalafil (within 48 h).

β-Adrenergic Blockers and Other Agents Beta blockers are the other mainstay of anti-ischemic treatment because they reduce myocardial oxygen needs. They may be started by the intravenous route in patients with severe ischemia but should be avoided in the presence of acute or severe heart failure, low cardiac output,

TABLE 285-3 Recommendations for Anti-Ischemic Drugs in the Acute Phase of NSTE-ACS

THERAPY	RECOMMENDATION	WHEN TO AVOID
Nitrates	• Use sublingual or intravenous nitrates to relieve angina • Use intravenous nitrates if recurrent angina, uncontrolled hypertension, or signs of heart failure • Consider in patients with vasospastic angina	• Recent use of a PDE-5 inhibitor[a] • Hypotension • Right ventricular infarct • Severe aortic stenosis
Beta blockers	• Initiate early for ischemic symptoms • Continue chronic therapy	• PR interval >0.24 s • 2nd or 3rd atrioventricular block • Heart rate <50 beats/min • Systolic pressure <90 mmHg • Shock or Killip class III or IV heart failure • Severe reactive airways disease
Calcium channel blockers	• Consider in patients with vasospastic angina • Consider in patients with contraindications to beta blockers	• Systolic pressure <90 mmHg • Pulmonary edema • Left ventricular dysfunction[b] • Avoid short-acting nifedipine
Morphine or other narcotic analgesics[c]	• Continued severe angina despite 3 sublingual nitroglycerin tablets • Recurrent ischemia despite adequate anti-ischemic therapy	• Hypotension • Respiratory depression • Confusion or obtundation

[a]Sildenafil or vardenafil <24 h or tadalafil <48 h. [b]Diltiazem or verapamil. [c]Concomitant administration may delay the absorption and blunt the antiplatelet effect of oral $P2Y_{12}$ inhibitors.

Abbreviations: NSTE-ACS, non-ST-segment elevation acute coronary syndrome; PDE-5, phosphodiesterase type 5.

Source: Modified from M Roffi et al: Eur Heart J 37:267, 2017.

hypotension, or contraindications (e.g., high-degree atrioventricular block, active bronchospasm). Ordinarily, oral beta blockade targeted to a heart rate of 50–60 beats/min is recommended.

Heart rate–slowing calcium channel blockers, e.g., verapamil or diltiazem, are recommended for patients who have persistent symptoms or ECG signs of ischemia after treatment with full-dose nitrates and beta blockers and in patients with contraindications to either class of these agents. Patients who have continuing severe chest pain despite maximal anti-ischemic therapy and are without contraindications to morphine may receive this drug intravenously (1–5 mg every 5–30 min). Additional medical therapy includes angiotensin-converting enzyme (ACE) inhibitors or angiotensin receptor blockers. Early administration of intensive HMG-CoA reductase inhibitors (statins), such as atorvastatin 80 mg/d or rosuvastatin 40 mg/d, prior to percutaneous coronary intervention (PCI), and continued thereafter has been shown to reduce both early and late recurrent ischemic events. In patients who do not have an adequate response to maximally tolerated statin (i.e., <50% decrease in low-density lipoprotein cholesterol [LDL-C]), addition of ezetimibe 10 mg daily and/or a PCSK9 inhibitor (alirocumab, evolocumab) early after ACS have been shown to further reduce the LDL-C and prevent future cardiovascular events.

ANTITHROMBOTIC THERAPY

Antithrombotic therapy consisting of antiplatelet and anticoagulant drugs represents the second major cornerstone of treatment **(Table 285-4)**.

Antiplatelet Drugs (See Chap. 123) Initial treatment should begin with the cyclooxygenase inhibitor aspirin with a dose of 150–325 mg of an oral non-enteric-coated preparation or 75–250 mg intravenously. Thereafter, lower oral doses (75–100 mg/d) are recommended since they maintain efficacy while causing less bleeding. Contraindications are severe active bleeding and aspirin allergy.

In the absence of a high risk for bleeding, patients with NSTE-ACS, irrespective of whether an invasive or conservative strategy (see below) is selected, should also receive a platelet $P2Y_{12}$ receptor blocker to inhibit platelet activation. There are now three oral and one intravenous $P2Y_{12}$ inhibitors available. The thienopyridine clopidogrel is an inactive prodrug that is converted into an active metabolite that causes irreversible blockade of the platelet $P2Y_{12}$ receptor. The loading dose of clopidogrel is 300–600 mg, whereas the maintenance dose is 75 mg daily. When clopidogrel is added to aspirin, so-called dual antiplatelet therapy (DAPT), in patients with NSTE-ACS, it confers a 20% relative reduction in cardiovascular death, MI, or stroke, compared to aspirin alone but is associated with a moderate (absolute 1%) increase in major bleeding.

Two other $P2Y_{12}$ inhibitors have been shown to be superior to clopidogrel in preventing recurrent cardiac ischemic events but both increase bleeding. Prasugrel, also a thienopyridine, achieves a more rapid onset and higher level of irreversible platelet inhibition than clopidogrel. Compared to clopidogrel, prasugrel significantly reduced the combined risk of cardiovascular death, MI, stroke, and stent thrombosis, but increased bleeding, in patients with ACS with scheduled PCI. It has been approved for ACS patients following angiography when PCI is planned; it should be administered at a loading dose of 60 mg followed by 10 mg/d. Prasugrel is contraindicated in patients with prior stroke or transient ischemic attack or at high risk for bleeding, and should be administered once the coronary anatomy is known (i.e., post coronary angiography).

Ticagrelor, a potent, *reversible* platelet $P2Y_{12}$ inhibitor, reduces the risk of cardiovascular death, total mortality, or MI compared to clopidogrel across a broad spectrum of patients with ACS. After a loading dose of 180 mg, 90 mg bid is administered as maintenance. Like prasugrel, ticagrelor increases the risk of bleeding. Unlike prasugrel, ticagrelor demonstrated benefit whether patients were managed conservatively or with an early invasive strategy (see below). Some patients may develop dyspnea soon after administration, although the symptoms are often transient and infrequently serious.

Up to one-third of patients have an inadequate response to clopidogrel, and a substantial proportion of these cases are related to a genetic variant of the cytochrome P450 system involving the *2C19* gene that leads to reduced conversion of clopidogrel into its active metabolite. Thus, alternate $P2Y_{12}$ blockers (prasugrel or ticagrelor) should be considered in patients with NSTE-ACS who develop a new coronary event while receiving clopidogrel and aspirin. An intravenous, direct, and rapidly acting $P2Y_{12}$ inhibitor, cangrelor, has also shown benefit relative to clopidogrel in patients who underwent PCI following a NSTE-ACS and reduced the risk of cardiovascular ischemic events, but at the expense of increased major bleeding events.

TABLE 285-4 Clinical Use of Antithrombotic Therapy

Oral Antiplatelet Therapy	
Aspirin	Loading dose of 150–325 mg orally of nonenteric formulation followed by 75–100 mg/d of an enteric or a nonenteric formulation
Clopidogrel	Loading dose of 600 mg (if PCI planned) or 300 mg (if no PCI planned), followed by 75 mg/d
Prasugrel	Pre-PCI: Loading dose of 60 mg followed by 10 mg/d In patient with body weight <60 kg or >75 years old, maintenance dose of 5 mg/d
Ticagrelor	Loading dose of 180 mg followed by 90 mg twice daily
Intravenous Antiplatelet Therapy at the Time of PCI	
Cangrelor	30 μg/kg bolus followed by 4 μg/kg/min infusion for 2 h or duration of procedure
Eptifibatide	180 μg/kg bolus followed in 10 min by second bolus of 180 μg/kg with infusion of 2.0 μg/kg/min for up to 18 h
Tirofiban	25 μg/kg per min over 3 minutes, followed by infusion of 0.15 μg/kg/min for up to 18 h
Parenteral Anticoagulants[a]	
Unfractionated heparin	Bolus 70–100 U/kg (maximum 5000 U) IV followed by infusion of 12–15 U/kg per h titrated to ACT 250–300 s
Enoxaparin	0.5 mg/kg IV bolus at the time of PCI *or* 1 mg/kg subcutaneous every 12 h; the first dose may be preceded by a 30-mg IV bolus; renal adjustment to 1 mg/kg once daily if creatine clearance <30 mL/min
Bivalirudin	Initial IV bolus of 0.75 mg/kg followed by an infusion of 1.75 mg/kg per h
Fondaparinux	2.5 mg subcutaneously daily (only prior to PCI)
Oral Anticoagulant Drugs (concomitant treatment after PCI)	
VKA	Dosing based on INR value and the respective clinical indication
Apixaban	Maintenance dose 5 mg (dose reduced to 2.5 mg) twice daily
Dabigatran	Maintenance dose 150 mg (dose reduced to 75 mg) twice daily in the United States; outside the United States, either 110 or 150 mg twice daily may be used
Edoxaban	Maintenance dose 60 mg (dose reduced to 30 mg) once daily
Rivaroxaban	Maintenance dose 20 mg (dose reduced to 15 mg) once daily In patients without atrial fibrillation or venous thromboembolism 2.5 mg bid may be used

Note: All dose reductions for patients meeting criteria. **(See also Chap. 123.)**

Abbreviations: ACT, activated clotting time; INR, international normalized ratio; PCI, percutaneous coronary intervention; VKA, vitamin K antagonists (e.g., warfarin).

Source: Modified from FJ Neumann: Eur Heart J 40:137, 2019.

In the 1990s and early 2000s, several trials of glycoprotein IIb/IIIa inhibitors in patients with NSTE-ACS had shown modest benefit counterbalanced by an increase in major bleeding. However, the majority of earlier studies were performed without concomitant $P2Y_{12}$ inhibitor treatment, and more recent studies in patients receiving the latter failed to show a benefit of routine early initiation of a glycoprotein IIb/IIIa inhibitor. Because of the increased risk of bleeding, the addition of a glycoprotein IIb/IIIa inhibitor to aspirin and a $P2Y_{12}$ inhibitor (i.e., triple antiplatelet therapy) should be reserved for patients undergoing PCI with a high coronary thrombus burden on angiography or thrombotic complications of PCI.

Long-term DAPT should continue preferably for 12 months in patients with NSTE-ACS without an indication for long-term full-dose anticoagulation; the duration of DAPT is dependent upon the risk of bleeding versus thrombosis. Clinicians should select the antiplatelet regimen that provides the best balance between efficacy and safety based on the individual patient characteristics and clinical scenario. A shorter duration of DAPT may be considered for patients at high risk of bleeding who are not high at high ischemic risk (see below). Longer duration DAPT, i.e., >12 months, is favored in patients with high atherothrombotic risk (e.g., due to stenting of the left main coronary artery or proximal left anterior descending or proximal bifurcating coronary arteries, recurrent MI, stent thrombosis).

Anticoagulants (See Chap. 123) Four parenteral options are available for anticoagulant therapy to be added to antiplatelet agents: (1) unfractionated heparin (UFH), long the mainstay of therapy; (2) the low-molecular-weight heparin (LMWH) enoxaparin, which has been shown to be superior to UFH in reducing recurrent cardiac events, especially in patients managed by a conservative strategy (however, it is accompanied by a modest increase in bleeding); (3) bivalirudin, a direct thrombin inhibitor that is similar in efficacy to either UFH or LMWH and is used just prior to and/or during PCI; and (4) fondaparinux, a synthetic factor Xa inhibitor that is equivalent in efficacy to enoxaparin but has a lower risk of major bleeding. While UFH and enoxaparin have been widely studied in patients managed either with an early conservative or invasive strategy, bivalirudin is rarely used in conservatively managed patients, while fondaparinux requires supplemental UFH to prevent procedure-related thrombosis. In patients who develop NSTE-ACS while receiving treatment with a direct oral anticoagulant (DOAC), the DOAC **(see Chap. 123)** should be held when one of the four parenteral anticoagulants is begun.

If an early invasive strategy is indicated (see below), radial arterial access is recommended to reduce the risk of bleeding. Excessive bleeding is the most important adverse effect of all antithrombotic agents, including both antiplatelet agents and anticoagulants. Therefore, attention must be directed to the doses of antithrombotic agents, accounting for age, body weight, creatinine clearance, and a previous history of excessive bleeding. Patients who have experienced a stroke are at higher risk of intracranial bleeding with potent antiplatelet agents and combinations of antithrombotic drugs. In patients with atrial fibrillation (including patients with NSTE-ACS) treated with an oral anticoagulant who undergo PCI, the duration of DAPT should be shortened (e.g., stop aspirin after hospital discharge or at most at 1 month after PCI, except in patients at very high risk for ischemic events), and continue $P2Y_{12}$ inhibitor plus DOAC for 1 year. After 1 year, the majority of patients should be transitioned to oral anticoagulation monotherapy without concomitant antiplatelet treatment.

INVASIVE VERSUS CONSERVATIVE STRATEGY (FIG. 285-4)

In an invasive strategy, following initiation of anti-ischemic and antithrombotic agents as described above, coronary arteriography is recommended within 2 h of presentation for patients are unstable or within 24 h in high-risk stable patients, followed by coronary revascularization (PCI or coronary artery bypass grafting), depending on the coronary anatomy (Fig. 285-4). Multiple clinical trials have demonstrated the benefit of this strategy in high-risk patients (i.e., patients with multiple clinical risk factors, ST-segment deviation, and/or positive biomarkers). Two studies comparing an immediate invasive strategy (median time to intervention of 1.4 and 4.7 h after presentation) reduced the rate of death or new MI compared to a delayed invasive strategy (median time of intervention of 61 and 62 h), with a greater benefit among patients with a high-risk score. In patients at low risk, the outcomes from an

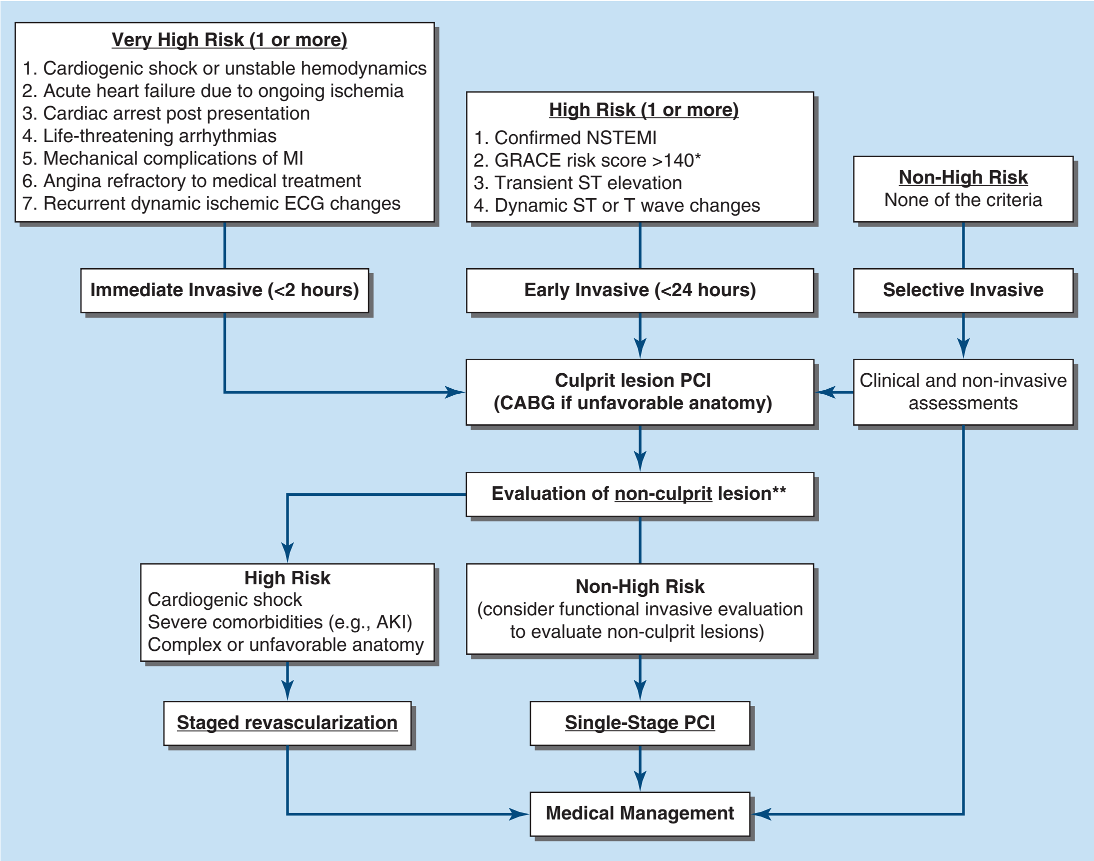
Very High Risk (1 or more)
1. Cardiogenic shock or unstable hemodynamics
2. Acute heart failure due to ongoing ischemia
3. Cardiac arrest post presentation
4. Life-threatening arrhythmias
5. Mechanical complications of MI
6. Angina refractory to medical treatment
7. Recurrent dynamic ischemic ECG changes
High Risk (1 or more)
1. Confirmed NSTEMI
2. GRACE risk score >140*
3. Transient ST elevation
4. Dynamic ST or T wave changes
Non-High Risk
None of the criteria
Immediate Invasive (<2 hours)
Early Invasive (<24 hours)
Selective Invasive
Culprit lesion PCI
(CABG if unfavorable anatomy)
Clinical and non-invasive assessments
Evaluation of non-culprit lesion**
High Risk
Cardiogenic shock
Severe comorbidities (e.g., AKI)
Complex or unfavorable anatomy
Non-High Risk
(consider functional invasive evaluation to evaluate non-culprit lesions)
Staged revascularization
Single-Stage PCI
Medical Management

FIGURE 285-4 Selection, timing, and staging of invasive strategy in patients with non-ST-elevation acute coronary syndrome. AKI, acute kidney injury; CABG, coronary artery bypass grafting; GRACE, Global Registry of Acute Coronary Syndrome Events; PCI, percutaneous coronary intervention. Patients with non-ST-elevations acute coronary syndrome who have one or more very-high risk features should undergo immediate invasive strategy with coronary angiography as soon as possible (within 2 hours), unless absolute contraindications exist. An early invasive strategy should be considered within 24 hours in stable patients with one or more high risk features. Stable patients with no very-high or high risk features should be managed with a selective invasive approach in which clinical, non-invasive assessments, and patient preferences are used to guide management. *Fox KA, BMJ Open 2014;4:e004425. https://doi.org/10.1136/bmjopen-2013-004425. **In patients with multivessel disease, single-stage PCI during the index procedure to achieve complete coronary revascularization should be considered in the absence of cardiogenic shock, severe comorbidities such as AKI, or coronary anatomy that is unfavorable or high risk requiring evaluation for The Heart Team for CABG vs complex coronary intervention.

invasive strategy are similar to those obtained from a conservative strategy. The latter consists of anti-ischemic and antithrombotic therapy followed by a "selective invasive approach," in which the patient is observed closely and coronary arteriography is conducted if coronary computed angiography shows the presence of epicardial coronary stenosis, rest pain or ST-segment changes recur, a biomarker of necrosis becomes positive, or there is evidence of severe ischemia on a stress test.

LONG-TERM MANAGEMENT

The time of hospital discharge is a "teachable moment" for the patient with NSTE-ACS, when the caregiver can review and optimize the medical regimen. Risk factor modification is crucial, and the importance of smoking cessation, following an appropriate diet, achieving and maintaining optimal weight, daily exercise, an LDL-C <55 mg/dL, blood pressure control, and control of hyperglycemia should be emphasized. There is evidence of benefit with long-term therapy with several classes of drugs. Beta blockers, intensive lipid-lowering therapies (high-intensity statin, ezetimibe, and if needed, adding a PCSK9 inhibitor or bempedoic acid) to achieve an LDL-C <55 mg/dL, ACE inhibitors or angiotensin receptor blockers, and sodium-glucose cotransporter 2 or glucagon-like peptide 1 receptor agonists in selected patients with type 2 diabetes mellitus **(see Chap. 416)** are recommended. The recommended antiplatelet regimen consists of the combination of low-dose (75–100 mg/d) aspirin and a $P2Y_{12}$ inhibitor (prasugrel or ticagrelor is favored over clopidogrel) for 12 months unless there is a high risk of bleeding. After 12 months, antiplatelet monotherapy with a single antiplatelet should be continued thereafter **(see Chap. 284)**, unless long-term full-dose anticoagulation is indicated, in which case anticoagulant without antiplatelet therapy is recommended long-term (see above).

Several trials have explored shorter courses and less intensive DAPT in patients not on anticoagulation who are at high risk of bleeding and low risk of ischemic events. Recent guidelines support a shorter-duration DAPT regimens of 3–6 months in these patients, followed by stopping aspirin and continuing with $P2Y_{12}$ inhibitor alone. Another option is de-escalation of $P2Y_{12}$ intensity (i.e., switching from prasugrel or ticagrelor to clopidogrel at 3 months) while continuing aspirin. De-escalation of DAPT is not recommended in the first 30 days following ACS as this has been associated with increased risk of ischemic events.

In patients who require long-term anticoagulation (e.g., patients with atrial fibrillation), the duration of so-called "triple therapy" consisting of aspirin, clopidogrel, and an oral anticoagulant should be kept to a minimum. In these patients, aspirin should be discontinued at discharge or 1 week, and oral anticoagulation and clopidogrel continued until month 12, followed thereafter by oral anticoagulant alone. Conversely, in patients at high ischemic risk (e.g., those with prior MI, diabetes mellitus, coronary vein graft, heart failure) who are also at low risk of bleeding and not on an anticoagulant, continuation of DAPT for up to 3 years has been shown to be beneficial. If ticagrelor is used, the dose should be reduced to 60 mg twice daily 1 year after ACS as this reduces bleeding without an increase in ischemic events.

Registries have shown that women and racial minorities, as well as patients with NSTE-ACS at high risk, including the elderly and patients with diabetes or chronic kidney disease, are less likely to receive evidence-based pharmacologic and interventional therapies with resultant poorer clinical outcomes and quality of life. Special attention should be directed to these groups.

VASOSPASTIC ANGINA

In 1959, Prinzmetal and colleagues described a syndrome of severe ischemic pain that usually occurs at rest and is associated with ST-segment elevation originally known as Prinzmetal's variant angina, but since renamed vasospastic angina (VA). VA is caused by focal spasm of an epicardial coronary artery with resultant transmural ischemia and abnormalities in left ventricular function that may lead to acute MI, ventricular tachycardia or fibrillation, and sudden cardiac death. The cause of the spasm is not well defined, but it may be related to hypercontractility of coronary arterial smooth muscle due to adrenergic vasoconstrictors, leukotrienes, or serotonin. For reasons that are not clear, the prevalence of VA has decreased substantially during the past few decades, although it remains much more frequent in Japan than in North America or Western Europe.

Clinical and Angiographic Manifestations Patients with VA are generally younger and, except for cigarette smoking, have fewer coronary risk factors than do patients with NSTE-ACS. Cardiac examination is usually unremarkable in the absence of ischemia. However, a minority of patients have a generalized vasospastic disorder associated with migraine and/or Raynaud's phenomenon. The clinical diagnosis of VA is made by the detection of transient ST-segment *elevation* with rest pain, although many patients may also exhibit episodes of silent ischemia.

Coronary angiography demonstrates transient coronary spasm as the diagnostic hallmark of VA. Atherosclerotic plaques in at least one proximal coronary artery occur in about half of patients. Hyperventilation and intracoronary acetylcholine have been used to provoke focal coronary stenosis on angiography or to provoke rest angina with ST-segment elevation to establish the diagnosis. In patients with no obstructive coronary atherosclerosis and suspected coronary vasomotor abnormalities, a positive provocative test for spasm has been shown to be safe, identifies a high-risk subgroup, and is endorsed by guidelines since it permits selection of therapy most appropriate for the underlying pathophysiology.

TREATMENT

Vasospastic Angina

Calcium channel blockers are the main therapeutic agents. Nitroglycerin, either sublingual or intravenous, is effective to relieve coronary arterial spasm and angina, but the efficacy of long-term nitrates is limited by nitrate tolerance. Aspirin may actually increase the severity of ischemic episodes, possibly as a result of the sensitivity of coronary tone to modest changes in the synthesis of prostacyclin, and should be used only in patients with significant atherosclerosis. Statin therapy has been shown to reduce the risk of major adverse events, although the precise mechanism is not established. The response to beta blockers is variable. Cilostazol and nicorandil (not available in the United States) may also prevent recurrent episodes. Coronary revascularization may be helpful in patients who also have discrete, flow-limiting, proximal fixed obstructive lesions. Patients who have had ischemia-associated ventricular fibrillation despite maximal medical therapy should receive an implantable cardioverter-defibrillator.

Prognosis Many patients with VA pass through an acute, active phase, with frequent episodes of angina and cardiac events during the first 6 months after presentation, after which there may be a tendency for symptoms and cardiac events to diminish over time. Survival at 5 years is excellent (~90–95%), but as many as 20% of patients experience an MI. Patients with no or mild fixed coronary obstruction experience a lower incidence of cardiac death or MI compared to patients with associated severe obstructive lesions. Patients with VA who develop serious arrhythmias during spontaneous episodes of pain are at a higher risk for sudden cardiac death. In most patients who survive an infarction or the initial 3- to 6-month period of frequent episodes, there is a tendency for symptoms and cardiac events to diminish over time.

GLOBAL CONSIDERATIONS

Ischemic heart disease (IHD) and its most dangerous manifestation, ACS, remain the most frequent cause of death and disability worldwide. In the mid-twentieth century, these conditions were most common in high-income countries. Although improvements in primary prevention and treatment of ACS have reduced cardiovascular morbidity and mortality due to IHD over the past 70 years, these advances have not affected all population groups equally. In Europe, there remains a northeast to southwest gradient, with higher prevalence in northern Russia and the Baltic nations and considerably lower prevalence in Spain, France, and Italy. In the United States, there remain racial and economic disparities, with poorer outcomes in minorities and low-income populations.

Simultaneous with these important advances in the high-income countries, the low- and middle-income countries have moved in the opposite direction. Improvements in agriculture, nutrition, sanitation, prevention and treatment of infections, and management of maternal/early childhood disorders, urbanization, and a reduction of physical labor have, in combination, led to marked increases in coronary risk factors—hypertension, cigarette smoking, obesity, diabetes mellitus, and elevations of circulating LDL-C. These changes have been most prominent in Central and South Asia, as well as in the more developed regions of sub-Saharan Africa and the Middle East.

The current challenge is to apply what was learned in high-income countries to the large populations in the low- and middle-income countries that are now at high risk. This will require large educational efforts directed at both the populations and their caregivers. An additional challenge will be to provide the trained specialized personnel, facilities, drugs, and devices to deal with these threats. The successful implementation of these measures is now principally a socio-politico-economic issue. One mitigating factor is that many of the important drugs to prevent and treat these disorders, such as statins, ezetimibe, ACE inhibitors, diuretics, beta blockers, and calcium antagonists, are off patent and are now inexpensive.

■ FURTHER READING

Byrne RA et al: 2023 ESC guidelines for the management of acute coronary syndromes. Eur Heart J 23:44, 2023.

Capodanno D et al: Defining strategies of modulation of antiplatelet therapy in patients with coronary artery disease. Circulation 147: 1933, 2023.

Hokimoto S et al: JCS/CVIT/JCC 2023 guideline focused update on diagnosis and treatment of vasospastic angina (coronary spastic angina) and coronary microvascular dysfunction. Circ J 87:879, 2023.

Kontos MC et al: 2022 ACC Expert Consensus Decision Pathway on the evaluation and disposition of acute chest pain in the emergency department. J Am Coll Cardiol 80:1925, 2022.

Lawton JS et al: 2021 ACC/AHA/SCAI guideline for coronary artery revascularization. J Am Coll Cardiol 79:e21, 2021.

Mach F et al: ESC/EAS guidelines for the management of dyslipidaemias: Lipid modification to reduce cardiovascular risk. Eur Heart J 41:111, 2020.

Martin SS et al: 2024 Heart disease and stroke statistics: A report of US and global data from the American Heart Association. Circulation 149:e347, 2024.

Thygesen K et al: Fourth universal definition of myocardial infarction. J Am Coll Cardiol 72:2231, 2018.

286

ST-Segment Elevation Myocardial Infarction

David A. Morrow, Elliott M. Antman, Joseph Loscalzo

Acute myocardial infarction (AMI) is a common diagnosis in hospitalized patients in industrialized countries. In the United States, ~605,000 patients experience a new AMI and 200,000 experience a recurrent AMI each year. About half of AMI-related deaths occur before the stricken individual reaches the hospital. The in-hospital mortality rate after admission for AMI has declined from 10 to ~5%. The 1-year mortality rate after AMI is ~15%. Mortality is approximately fourfold higher in patients aged >75 years as compared with younger patients.

When patients with prolonged ischemic discomfort at rest are first seen, the working clinical diagnosis is that they are suffering from an acute coronary syndrome **(Fig. 286-1)**. The 12-lead electrocardiogram (ECG) is a pivotal diagnostic and triage tool because it is at the center of the decision pathway for management, permitting distinction of those patients presenting with ST-segment elevation from those presenting without ST-segment elevation. Circulating cardiac biomarkers of myocardial injury are measured to distinguish unstable angina (UA) from non-ST-segment elevation myocardial infarction (NSTEMI) and to estimate preliminarily the magnitude of myocardial necrosis in an ST-segment elevation myocardial infarction (STEMI). Epidemiologic studies indicate there has been a shift in the pattern of AMI over the past several decades with more patients with NSTEMI than STEMI. This chapter focuses on the evaluation and management of patients with STEMI, while **Chap. 286** discusses UA/NSTEMI.

PATHOPHYSIOLOGY: ROLE OF ACUTE PLAQUE RUPTURE/EROSION

STEMI usually occurs when coronary blood flow decreases abruptly after a thrombotic occlusion of a coronary artery previously affected by atherosclerosis. Slowly developing, high-grade coronary artery stenoses do not typically precipitate STEMI because of the development of a rich collateral network over time. Instead, STEMI occurs when a coronary artery thrombus develops rapidly at a site of vascular injury. This injury is produced or facilitated by factors such as cigarette smoking, hypertension, lipid accumulation, and inflammation. In most cases, STEMI occurs when the surface of an atherosclerotic plaque becomes disrupted either through erosion or rupture (exposing its contents to the blood) and conditions (local or systemic) favor thrombogenesis. Histologic studies indicate that the coronary plaques prone to disruption are those with a rich lipid core and a thin fibrous cap. However, current clinical data have demonstrated that less than 5% of such thin-capped fibroatheromas are a nidus for AMI during long-term follow-up. Other morphologic characteristics associated with rupture-prone plaque include expansive remodeling, neovascularization (angiogenesis), plaque hemorrhage, adventitial inflammation, and a “spotty” pattern of calcification.

A mural thrombus forms at the site of plaque disruption, and the involved coronary artery becomes occluded. After an initial platelet monolayer forms at the site of the disrupted plaque, various agonists (collagen, ADP, epinephrine, serotonin) promote platelet activation. After agonist stimulation of platelets, thromboxane A_2 (a potent local vasoconstrictor) is released, further platelet activation occurs, and potential resistance to fibrinolysis develops.

In addition to the generation of thromboxane A_2, activation of platelets by agonists promotes a conformational change in the glycoprotein IIb/IIIa receptor **(Chap. 120)**. Once converted to its functional state, this receptor develops a high affinity for soluble adhesive proteins (i.e., integrins) such as fibrinogen. Since fibrinogen is a multivalent molecule, it can bind to two different platelets simultaneously, resulting in platelet cross-linking and aggregation.

The coagulation cascade is activated on exposure of tissue factor in damaged endothelial cells at the site of the disrupted plaque. Factors VII and X are activated, ultimately leading to the conversion of prothrombin to thrombin, which then converts fibrinogen to fibrin **(Chap. 121)**. Fluid-phase and clot-bound thrombin participate in an autoamplification reaction, leading to further activation of the coagulation cascade. The culprit coronary artery eventually becomes occluded by a thrombus containing platelet aggregates and fibrin strands **(Fig. 286-2)**.

In rare cases, STEMI may be due to coronary artery occlusion caused by coronary emboli, congenital coronary abnormalities, coronary spasm, or spontaneous coronary artery dissection. The amount of myocardial damage caused by coronary occlusion depends on (1) the territory supplied by the affected vessel, (2) whether or not the vessel becomes totally occluded, (3) the duration of coronary occlusion, (4) the quantity of blood supplied by collateral vessels to the affected tissue, (5) the demand for oxygen of the myocardium whose blood supply has been suddenly limited, (6) endogenous factors that can produce early spontaneous lysis of the occlusive thrombus, and (7) the adequacy of myocardial perfusion in the infarct zone when flow is restored in the occluded epicardial coronary artery.

Patients at increased risk for developing STEMI include those with multiple coronary risk factors and those with UA **(Chap. 285)**. Less

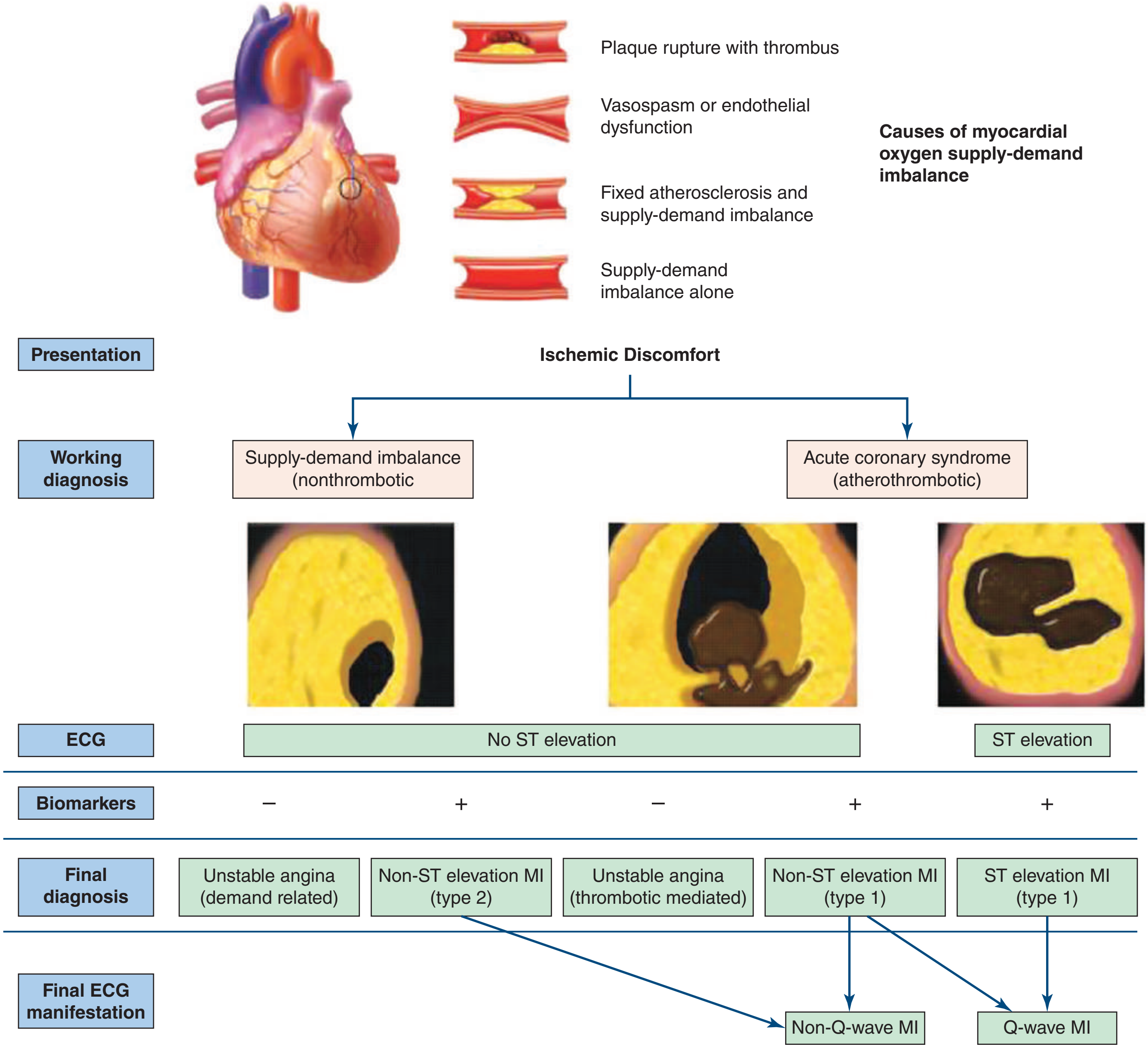
Plaque rupture with thrombus
Vasospasm or endothelial dysfunction
Fixed atherosclerosis and supply-demand imbalance
Supply-demand imbalance alone
Causes of myocardial oxygen supply-demand imbalance
Presentation
Ischemic Discomfort
Working diagnosis
Supply-demand imbalance (nonthrombotic)
Acute coronary syndrome (atherothrombotic)
ECG
No ST elevation
ST elevation
Biomarkers
−
+
−
+
+
Final diagnosis
Unstable angina (demand related)
Non-ST elevation MI (type 2)
Unstable angina (thrombotic mediated)
Non-ST elevation MI (type 1)
ST elevation MI (type 1)
Final ECG manifestation
Non-Q-wave MI
Q-wave MI

FIGURE 286-1 Acute coronary syndromes. Following disruption of a vulnerable plaque, patients experience ischemic discomfort resulting from a reduction of flow through the affected epicardial coronary artery. The flow reduction may be caused by a completely occlusive thrombus (*right*) or subtotally occlusive thrombus (*middle*). Patients with ischemic discomfort may present with or without ST-segment elevation. Of patients with ST-segment elevation, the majority ultimately develop a Q wave on the electrocardiogram (ECG; Q-wave myocardial infarction [MI]), while a minority do not develop Q waves and, in the past, were said to have sustained a non-Q-wave MI (NQMI). Patients who present without ST-segment elevation are suffering from either unstable angina or a non-ST-segment elevation MI (NSTEMI), a distinction that is ultimately made based on the presence or absence of a biomarker of myocardial injury such as cardiac troponin detected in the blood. *(Reproduced with permission from Hamm Scirica BM, Libby P, Morrow DA: ST-elevation myocardial infarction: Pathophysiology and clinical evolution, in Braunwald's Heart Disease, 12th ed, Libby P et al (eds). New York, Elsevier, 2022, Figure 37-1, pp 636-661.)*

common underlying medical conditions predisposing patients to STEMI include hypercoagulability, collagen vascular disease, systemic inflammatory diseases, cocaine abuse, and intracardiac thrombi or masses that produce coronary emboli.

There have been major advances in the management of STEMI with recognition that the "chain of survival" involves a highly integrated system starting with prehospital care and extending to early hospital management so as to provide expeditious implementation of a reperfusion strategy.

CLINICAL PRESENTATION

In up to one-half of cases, a precipitating factor appears to be present before STEMI, such as vigorous physical exercise, emotional stress, or a medical or surgical illness. Although STEMI may commence at any time of the day or night, circadian variations have been reported such that clusters are seen in the morning within a few hours of awakening.

Pain is the most common presenting complaint in patients with STEMI. The pain is deep and visceral; adjectives commonly used to describe it are *heavy*, *squeezing*, and *crushing*; although, occasionally, it is described as stabbing or burning **(Chap. 15)**. It is similar in character to the discomfort of angina pectoris **(Chap. 284)** but commonly occurs at rest, is usually more severe, and lasts longer. Typically, the pain involves the central portion of the chest and/or the epigastrium, and, on occasion, it radiates to the arms. Less common sites of radiation include the abdomen, back, lower jaw, and neck. The frequent location of the pain beneath the xiphoid and epigastrium and the patients' denial that they may be suffering a heart attack are chiefly responsible for the common mistaken impression of indigestion. The pain of STEMI may radiate as high as the occipital area but not below the umbilicus. It is often accompanied by sweating, nausea, vomiting, anxiety, and a sense of impending doom. The pain may commence when the patient is at rest, but when it begins during a period of exertion, it does not usually subside with cessation of activity, in contrast to angina pectoris.

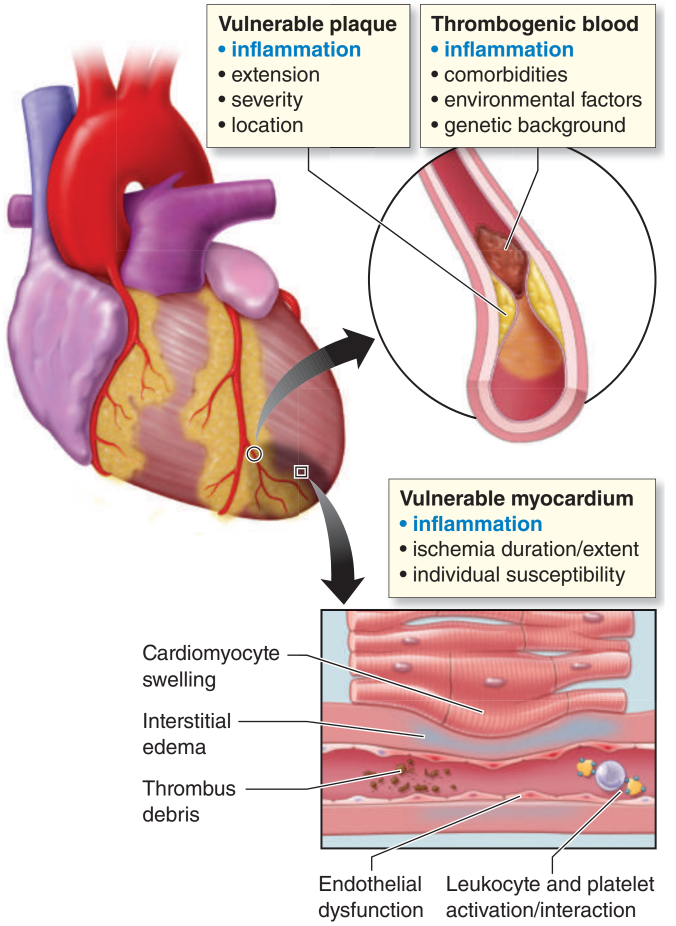

FIGURE 286-2 Critical determinants of myocardial infarction injury. The overlapping of vulnerable plaque and thrombogenic blood are critical determinants for myocardial infarction occurrence and extension. In addition, myocardial vulnerability, which is largely due to coronary microvascular dysfunction, contributes to extension and severity of ischemic injury. In the most severe form (known as no-reflow), structural and functional impairments sustain vascular obstruction. Endothelial dysfunction triggers leukocyte and platelet activation/interaction, whereas thrombotic debris may worsen the obstruction. Furthermore, cardiomyocyte swelling, interstitial edema, and tissue inflammation promote extravascular compression. *(Modified from F Montecucco, F Carbone, TH Schindler. Pathophysiology of ST-segment elevation myocardial infarction: Novel mechanisms and treatments. Eur Heart J 37:1268, 2016.)*

The pain of STEMI can simulate pain from acute pericarditis **(Chap. 281)**, pulmonary embolism **(Chap. 290)**, acute aortic dissection **(Chap. 291)**, costochondritis, and gastrointestinal disorders. These conditions should therefore be considered in the differential diagnosis. Radiation of discomfort to the trapezius is not seen in patients with STEMI and may be a useful distinguishing feature that suggests pericarditis is the correct diagnosis. However, *pain is not uniformly present in patients with STEMI*. The proportion of painless STEMIs is greater in patients with diabetes mellitus, and it increases with age. STEMI may present as sudden-onset breathlessness. Other less common presentations, with or without pain, include sudden loss of consciousness, a confusional state, a sensation of profound weakness, the appearance of an arrhythmia, evidence of peripheral embolism, or merely an unexplained drop in arterial pressure.

■ PHYSICAL FINDINGS

Most patients are anxious and restless, attempting unsuccessfully to relieve the pain by moving about in bed, altering their position, and stretching. Pallor associated with perspiration and coolness of the extremities occurs commonly. The combination of substernal chest pain persisting for >30 min and diaphoresis strongly suggests STEMI. Although many patients have a normal pulse rate and blood pressure within the first hour of STEMI, patients with anterior infarction may have manifestations of sympathetic nervous system hyperactivity (tachycardia and/or hypertension), and those with inferior infarction may show evidence of parasympathetic hyperactivity (bradycardia and/or hypotension).

The precordium is usually quiet, and the apical impulse may be difficult to palpate. Other physical signs of ventricular dysfunction include fourth and third heart sounds, decreased intensity of the first heart sound, and paradoxical splitting of the second heart sound **(Chap. 246)**. A transient midsystolic or late systolic apical systolic murmur due to dysfunction of the mitral valve apparatus may be present. A pericardial friction rub may be heard in patients with transmural STEMI at some time in the course of the illness. The carotid pulse is often decreased in volume, reflecting reduced stroke volume. Temperature elevations up to 38°C may be observed during the first week after STEMI.

LABORATORY FINDINGS

STEMI progresses through the following temporal stages: (1) acute (first few hours–7 days), (2) healing (7–28 days), and (3) healed (≥29 days). The myocardium undergoes a series of cellular responses in the infarct zone, beginning with recruitment of polymorphonuclear leukocytes (for removal of dead cells and clearance of extracellular macromolecules) followed by monocytes. Experimental work has now detailed this sequence of accumulation of subpopulations of mononuclear phagocytes with an initial wave of monocytes characterized by high proteolytic and phagocytic capacity and elaboration of proinflammatory cytokines followed by the repair monocytes that release a mediators that stimulate angiogenesis and extracellular matrix production **(Fig. 286-3)**. New microvessels and fibrosis are key constituents of forming granulation tissue, and these processes establish a foundation for myocardial scar formation, ventricular remodeling, and infarct healing. In addition, emerging evidence reveals a role for an "emergency hematopoiesis" that mobilizes leukocyte progenitor cells that ultimately participate in myocardial healing.

When evaluating the results of diagnostic tests for STEMI, the temporal phase of the infarction must be considered. The tests of value in confirming the diagnosis may be divided into four groups: (1) ECG, (2) blood-based cardiac biomarkers, (3) cardiac imaging, and (4) nonspecific indices of tissue necrosis and inflammation.

■ ELECTROCARDIOGRAM

The electrocardiographic manifestations of STEMI are described in **Chap. 247**. During the initial stage, total occlusion of an epicardial coronary artery produces ST-segment elevation. Most patients initially presenting with ST-segment elevation ultimately evolve Q waves on the ECG. However, Q waves in the leads overlying the infarct zone may vary in magnitude and appear only transiently, depending on the reperfusion status of the ischemic myocardium and restoration of transmembrane potentials over time. A small proportion of patients initially presenting with ST-segment elevation will not develop Q waves when the obstructing thrombus is not totally occlusive, obstruction is transient, or a rich collateral network is present. Among patients presenting with ischemic discomfort but *without* ST-segment elevation, if a cardiac biomarker of necrosis (see below) is detected, the diagnosis of NSTEMI is ultimately made (Fig. 286-1). A minority of patients who present initially without ST-segment elevation may develop a Q-wave myocardial infarction (MI). Previously, it was believed that transmural MI is present if the ECG demonstrates Q waves or loss of R waves, and nontransmural MI may be present if the ECG shows only transient ST-segment and T-wave changes. However, electrocardiographic-pathologic correlations are far from perfect, and terms such as *Q-wave MI*, *non-Q-wave MI*, *transmural MI*, and *nontransmural MI* have been replaced by STEMI and NSTEMI (Fig. 286-1). Contemporary studies using magnetic resonance imaging (MRI) suggest that the development of a Q wave on the ECG is more dependent on the volume of infarcted tissue rather than the transmurality of infarction.

■ CARDIAC BIOMARKERS OF MYOCARDIAL INJURY

Certain proteins, referred to as cardiac biomarkers, are released from necrotic heart muscle after STEMI. The rate of liberation of specific

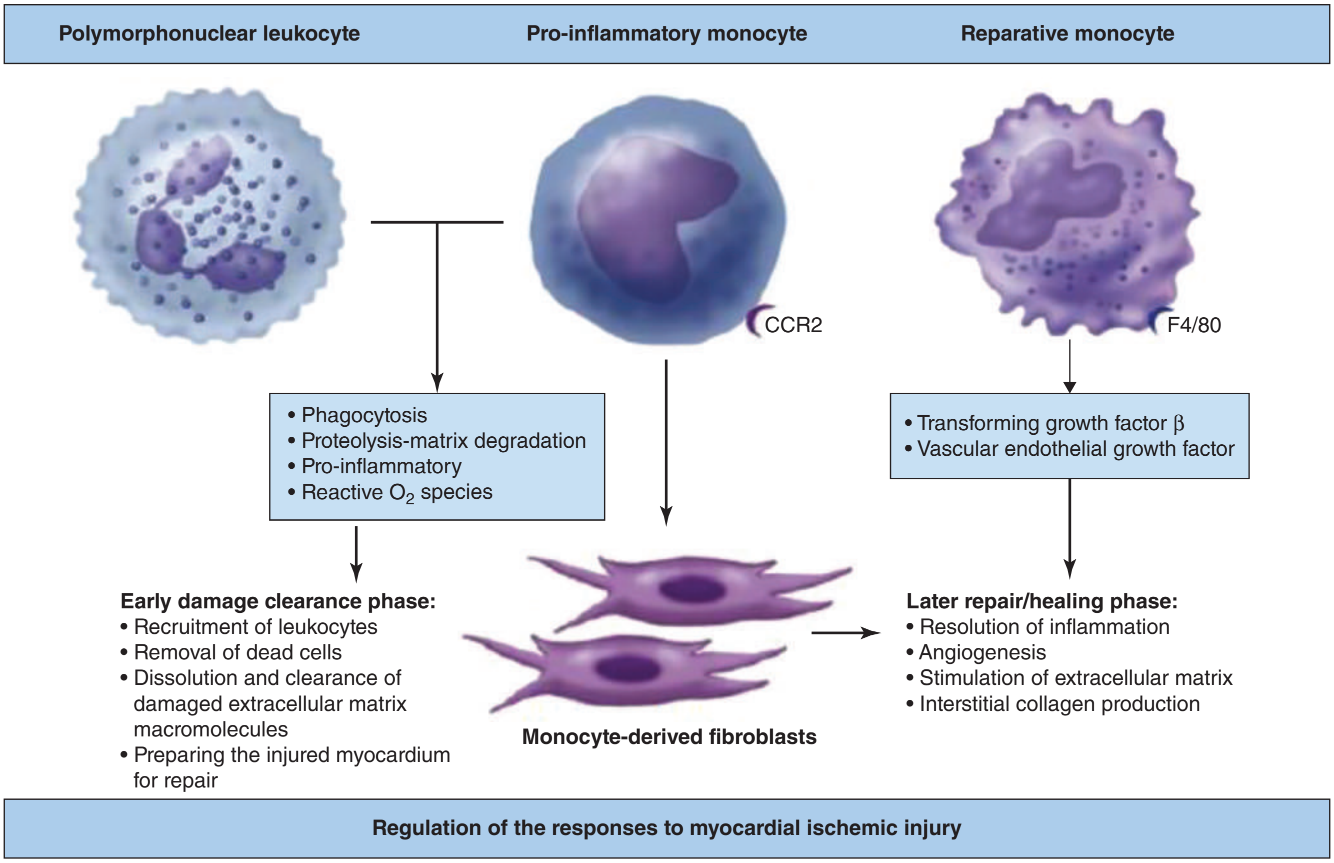
Polymorphonuclear leukocyte
Pro-inflammatory monocyte
Reparative monocyte
CCR2
F4/80
• Phagocytosis
• Proteolysis-matrix degradation
• Pro-inflammatory
• Reactive O2 species
• Transforming growth factor β
• Vascular endothelial growth factor
Early damage clearance phase:
• Recruitment of leukocytes
• Removal of dead cells
• Dissolution and clearance of damaged extracellular matrix macromolecules
• Preparing the injured myocardium for repair
Monocyte-derived fibroblasts
Later repair/healing phase:
• Resolution of inflammation
• Angiogenesis
• Stimulation of extracellular matrix
• Interstitial collagen production
Regulation of the responses to myocardial ischemic injury

FIGURE 286-3 Phases of myocardial injury and healing. Immediate recruitment of polymorphonuclear leukocytes precedes the accumulation of proinflammatory monocytes (typically bearing the chemokine receptor 2 [CCR2]). These phagocytic cells release the mediators of the early phase response to ischemic injury by clearing dead cells and debris and prompting additional inflammatory cells to enter the injured area. Reparative monocytes stimulate the production of extracellular matrix macromolecules that reinforce the cardiac skeleton during healing as well as microvessel formation characteristic of granulation tissue. Monocyte-derived fibroblasts are capable of interstitial collagen synthesis that reinforces repair of the myocardial skeleton, potentially mitigating expansive remodeling of the infarct and promoting healing with less risk for chronic heart failure. *(Reproduced with permission from P Libby et al: The myocardium: More than just myocytes. J Am Coll Cardiol 74:3137, 2019.)*

proteins differs depending on their intracellular location, their molecular weight, and the local blood and lymphatic flow. Cardiac biomarkers become detectable in the peripheral blood once the capacity of the cardiac lymphatics to clear the interstitium of the infarct zone is exceeded and spillover into the venous circulation occurs. The criteria for AMI require a rise and/or fall in cardiac biomarker values with at least one value above the 99th percentile of the upper reference limit for normal individuals.

Cardiac-specific troponin T (cTnT) and *cardiac-specific troponin I* (cTnI) have amino-acid sequences that differ from those of the skeletal muscle forms of these proteins. These differences permitted the development of quantitative assays for cTnT and cTnI using highly specific monoclonal antibodies. cTnT and cTnI may increase after STEMI to levels many times higher than the upper reference limit (set at the 99th percentile of values in a reference population). The measurement of cTnT or cTnI is of considerable diagnostic usefulness, and they are now the preferred biochemical markers for MI when measured with high-sensitivity assays **(Fig. 286-4 and Chap. 15)**. In practical terms, the high-sensitivity troponin assays are of less immediate value in patients with STEMI. Contemporary urgent reperfusion strategies necessitate making a decision (based largely on a combination of clinical and ECG findings) before the results of blood tests have returned from the laboratory. Levels of cTnI and cTnT typically remain elevated for at least 7–10 days after STEMI.

Historically, creatine phosphokinase (CK) and its MB isoenzyme (CK-MB) were used for the diagnosis of AMI. A ratio (relative index) of CK-MB mass to CK activity ≥2.5 suggests but is not diagnostic of a myocardial rather than a skeletal muscle source for the CK-MB elevation. It is *not* cost-effective to measure both a cardiac-specific troponin and CK-MB. Because of its more rapid decline after the onset of AMI, CK-MB may be useful for the discrimination of early reinfarction during the period that cardiac troponin remains elevated following the index event (Fig. 286-3).

While it has long been recognized that the total quantity of protein released correlates with the size of the infarct, the peak protein concentration correlates only weakly with infarct size. Recanalization of a coronary artery occlusion (either spontaneously or by mechanical or pharmacologic means) in the early hours of STEMI causes earlier peaking of biomarker measurements (Fig. 286-4) because of a rapid washout from the interstitium of the infarct zone, quickly overwhelming lymphatic clearance of the proteins.

The *nonspecific reaction* to myocardial injury is associated with polymorphonuclear leukocytosis, which appears within a few hours after the onset of pain and persists for 3–7 days; the white blood cell count often reaches levels of 12,000–15,000/μL. The erythrocyte sedimentation rate rises more slowly than the white blood cell count, peaking during the first week and sometimes remaining elevated for 1 or 2 weeks.

■ CARDIAC IMAGING

Abnormalities of wall motion on *two-dimensional echocardiography* **(Chap. 248)** are almost universally present in patients with STEMI. Although acute STEMI cannot be distinguished from an old myocardial scar or from acute severe ischemia by echocardiography, the ease and safety of the procedure make its use appealing as a screening tool in the emergency department. When the ECG is not diagnostic of STEMI, early detection of the presence or absence of wall motion abnormalities by echocardiography can aid in management decisions, such as whether the patient should receive reperfusion therapy (e.g., percutaneous coronary intervention [PCI] or fibrinolysis). Echocardiographic estimation of left ventricular (LV) function is useful prognostically; detection of reduced function serves as an indication for

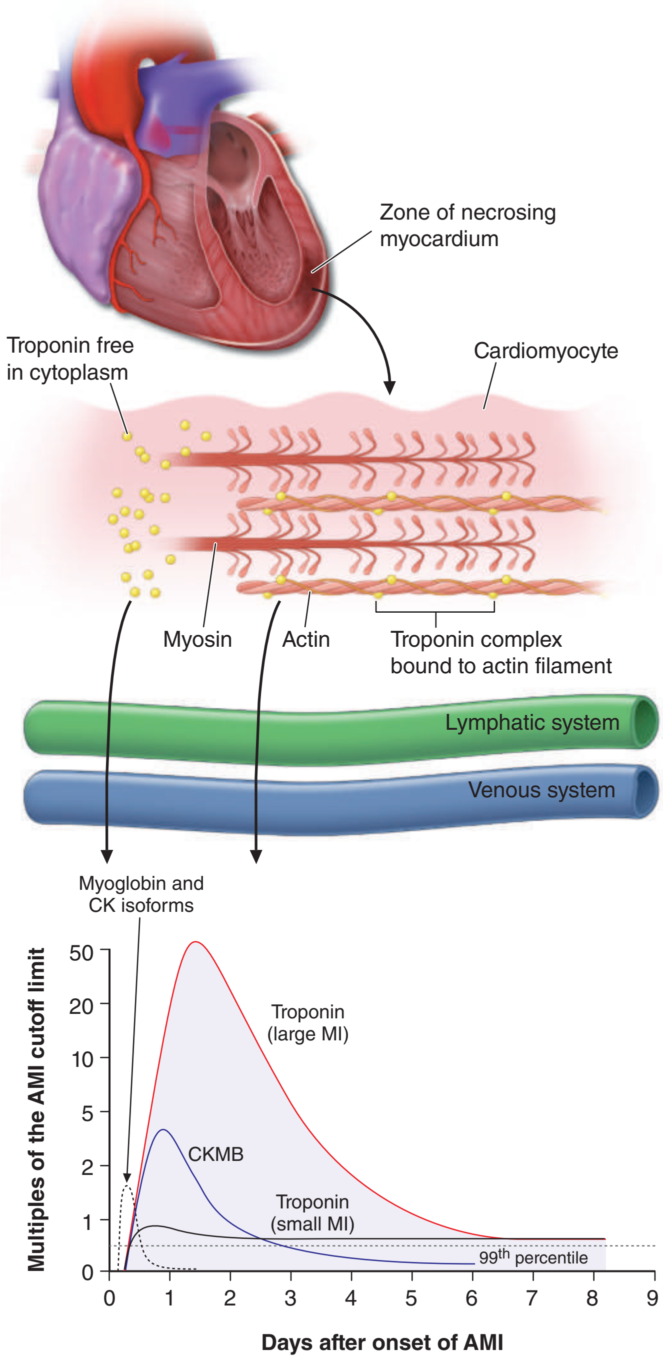

FIGURE 286-4 The zone of necrosing myocardium is shown at the top of the figure, followed in the middle portion of the figure by a diagram of a cardiomyocyte that is in the process of releasing biomarkers. The biomarkers that are released into the interstitium are first cleared by lymphatics followed subsequently by spillover into the venous system. After disruption of the sarcolemmal membrane of the cardiomyocyte, the cytoplasmic pool of biomarkers is released first (*left-most arrow in bottom portion of figure*). Markers such as myoglobin and CK isoforms are rapidly released, and blood levels rise quickly above the cutoff limit; this is then followed by a more protracted release of biomarkers from the disintegrating myofilaments that may continue for several days. Cardiac troponin levels rise to about 20–50 times the upper reference limit (the 99th percentile of values in a reference control group) in patients who have a "classic" acute myocardial infarction (MI) and sustain sufficient myocardial necrosis to result in abnormally elevated levels of the MB fraction of creatine kinase (CK-MB). Clinicians can now diagnose episodes of microinfarction by sensitive assays that detect cardiac troponin elevations above the upper reference limit, even though CK-MB levels may still be in the normal reference range (not shown). CV, coefficient of variation. *(Modified from EM Antman: Decision making with cardiac troponin tests. N Engl J Med 346:2079, 2002, and; bottom image: Reproduced with permission from AS Jaffe: Biomarkers in acute cardiac disease: The present and the future. J Am Coll Cardiol 48:1, 2006.)*

therapy with an inhibitor of the renin-angiotensin-aldosterone system. Echocardiography may also identify the presence of right ventricular (RV) infarction, ventricular aneurysm, pericardial effusion, and LV thrombus. In addition, Doppler echocardiography is useful in the detection and quantitation of a ventricular septal defect and mitral regurgitation, two serious complications of STEMI.

Radionuclide imaging techniques **(Chap. 248)** are available but rarely used for evaluating patients with suspected STEMI. Myocardial perfusion imaging with $[^{99m}\text{Tc}]$-sestamibi, which is distributed in proportion to myocardial blood flow and concentrated by viable myocardium **(Chap. 285)**, reveals a defect ("cold spot") in most patients during the first few hours after development of a transmural infarct. However, the technique cannot distinguish acute infarcts from chronic scars and, thus, is not specific for the diagnosis of *acute* MI.

An Expert Consensus Task Force for the Universal Definition of Myocardial Infarction has provided a comprehensive set of criteria for the definition of MI that integrates the clinical and laboratory findings discussed earlier as well as a classification of MI into five types that reflect the clinical circumstances in which it may occur **(Table 286-1)**.

INITIAL MANAGEMENT

■ PREHOSPITAL CARE

The prognosis in STEMI in the era of primary PCI is largely related to the occurrence of two general classes of complications: (1) electrical complications (arrhythmias) and (2) mechanical complications ("pump failure"). Most out-of-hospital deaths from STEMI result from the sudden development of ventricular fibrillation. The vast majority of deaths due to ventricular fibrillation occur within the first 24 h of the onset of symptoms, and of these, over half occur in the first hour. Therefore, the major elements of prehospital care of patients with suspected STEMI include (1) recognition of symptoms by the patient and prompt seeking of medical attention; (2) rapid deployment of an emergency medical team capable of performing resuscitative maneuvers, including defibrillation; (3) expeditious transportation of the patient to a hospital facility that is continuously staffed by physicians and nurses skilled in managing arrhythmias and providing advanced cardiac life support; and (4) expeditious implementation of reperfusion therapy. The greatest delay usually occurs not during transportation to the hospital but, rather, between the onset of pain and the patient's decision to call for help. This delay can best be reduced by health care professionals educating the public concerning the significance of chest discomfort and the importance of seeking early medical attention. Regular office visits with patients having a history of, or who are at risk for, ischemic heart disease are important "teachable moments" for clinicians to review the symptoms of AMI and the appropriate action plan.

Increasingly, monitoring and treatment are carried out by trained personnel in the ambulance, further shortening the time between the onset of the infarction and appropriate treatment. In areas remote from PCI centers, fibrinolytic therapy may be administered prehospital. General guidelines for initiation of fibrinolysis in the prehospital setting include the ability to transmit 12-lead ECGs to confirm the diagnosis, the presence in the ambulance of personnel trained in the interpretation of ECGs and management of STEMI, and online medical command and control that can authorize the initiation of treatment in the field.

MANAGEMENT IN THE EMERGENCY DEPARTMENT

In the emergency department, the goals for the management of patients with suspected STEMI include control of cardiac discomfort, rapid identification of patients who are candidates for urgent reperfusion therapy, triage of lower-risk patients to the appropriate location in the hospital, and avoidance of inappropriate discharge of patients with STEMI. Many aspects of the treatment of STEMI are initiated in the emergency department and then continued during the in-hospital phase of management. The overarching goal is to minimize the time from first medical contact to initiation of reperfusion therapy **(Fig. 286-5)**. This may involve transfer from a non-PCI hospital to one

TABLE 286-1 Definitions of Myocardial Injury and Infarction

Criteria for Myocardial Injury

The term *myocardial injury* should be used when there is evidence of elevated cardiac troponin (cTn) levels with at least one value above the 99th percentile upper reference limit (URL). The myocardial injury is considered acute if there is a rise and/or fall of cTn values.

Criteria for Acute Myocardial Infarction (types 1, 2, and 3 MI)

The term *acute myocardial infarction* (MI) should be used when there is acute myocardial injury with clinical evidence of acute myocardial ischemia and with detection of a rise and/or fall of cTn values with at least one value above the 99th percentile URL and at least one of the following:

- Symptoms of myocardial ischemia
- New ischemic electrocardiographic (ECG) changes
- Development of pathologic Q waves
- Imaging evidence of new loss of viable myocardium or new regional wall motion abnormality in a pattern consistent with an ischemic etiology
- Identification of a coronary thrombus by angiography or autopsy (not for types 2 or 3 MIs)

Postmortem demonstration of acute atherothrombosis in the artery supplying the infarcted myocardium meets criteria for *type 1 MI*. Evidence of an imbalance between myocardial oxygen supply and demand unrelated to acute atherothrombosis meets criteria for *type 2 MI*. Cardiac death in patients with symptoms suggestive of myocardial ischemia and presumed new ischemic ECG changes before cTn values became available or abnormal meets criteria for *type 3 MI*.

Criteria for Coronary Procedure–Related MI (types 4 and 5 MI)

Percutaneous coronary intervention (PCI)–related MI is termed *type 4a MI*.

Coronary artery bypass grafting (CABG)–related MI is termed *type 5 MI*. Coronary procedure–related MI <48 h after the index procedure is arbitrarily defined by an elevation of cTn values >5 times for *type 4a MI* and >10 times for *type 5 MI* of the 99th percentile URL in patients with normal baseline values. Patients with elevated preprocedural cTn values, in whom the preprocedural cTn levels are stable (<20% variation) or falling, must meet the criteria for a >5- or >10-fold increase and manifest a change from the baseline value of >20%. In addition, they must have at least one of the following:

- New ischemic ECG changes (this criterion is related to *type 4a MI* only)
- Development of new pathologic Q waves
- Imaging evidence of loss of viable myocardium that is presumed to be new and in a pattern consistent with an ischemic etiology
- Angiographic findings consistent with a procedural flow-limiting complication such as coronary dissection, occlusion of a major epicardial artery or graft, side-branch occlusion-thrombus, disruption of collateral flow, or distal embolization

Isolated development of new pathologic Q waves meets the *type 4a MI* or *type 5 MI* criteria with either revascularization procedure if cTn levels are elevated and rising, but less than the prespecified thresholds for PCI and CABG.

Other types of type 4 MI include *type 4B MI* stent thrombosis and *type 4C MI* restenosis that both meet *type 1 MI* criteria.

Postmortem demonstration of a procedure-related thrombus meets the *type 4a MI* and *type 5 MI* criteria if associated with a stent.

Criteria for Prior or Silent/Unrecognized MI

Any one of the following criteria meets the diagnosis for prior or silent/unrecognized MI:

- Abnormal Q waves with or without symptoms in the absence of nonischemic causes
- Imaging evidence of loss of viable myocardium in a pattern consistent with ischemic etiology
- Pathoanatomical findings of a prior MI

Source: Reproduced with permission from K Thygesen et al: Fourth universal definition of myocardial infarction (2018). Circulation 138:e618, 2018.

that is PCI capable, with a goal of initiating PCI within 120 min of first medical contact (Fig. 286-5).

Aspirin is essential in the management of patients with suspected STEMI and is effective across the entire spectrum of acute coronary syndromes. Rapid inhibition of cyclooxygenase-1 in platelets followed by a reduction of thromboxane A_2 levels is achieved by buccal absorption of a chewed 160–325-mg tablet in the emergency department. This measure should be followed by daily oral administration of aspirin in a dose of 75–162 mg.

In patients whose arterial oxygen (O_2) saturation is normal, supplemental O_2 is not recommended. However, when hypoxemia is present (O_2 saturation <90%), O_2 should be administered to correct the hypoxemia; the patient should then be reassessed to determine if there is a continued need for such treatment.

CONTROL OF DISCOMFORT

Sublingual *nitroglycerin* can be given safely to most patients with STEMI. Up to three doses of 0.4 mg should be administered at about 5-min intervals in patients with persistent ischemic symptoms. In addition to diminishing or abolishing chest discomfort, nitroglycerin may be capable of both decreasing myocardial O_2 demand (by lowering preload) and increasing myocardial O_2 supply (by dilating infarct-related coronary vessels or collateral vessels and by improving subendocardial perfusion). In patients whose initially favorable response to sublingual nitroglycerin is followed by the return of chest discomfort, particularly if accompanied by other evidence of ongoing ischemia such as further ST-segment or T-wave shifts, the use of intravenous nitroglycerin may be considered. Therapy with nitrates should be avoided in patients who present with low systolic arterial pressure (<90 mmHg) or in whom there is clinical suspicion of RV infarction (inferior infarction on ECG, elevated jugular venous pressure, clear lungs, and hypotension). Nitrates should not be administered to patients who have taken a phosphodiesterase-5 inhibitor for erectile dysfunction within the preceding 24 h because it may potentiate the hypotensive effects of nitrates. An idiosyncratic reaction to nitrates, consisting of sudden marked hypotension, sometimes occurs but can usually be reversed promptly by the rapid administration of intravenous atropine.

Morphine is a very effective analgesic for the pain associated with STEMI. However, it may reduce sympathetically mediated arteriolar and venous constriction, and the resulting venous pooling may reduce cardiac output and arterial pressure. These hemodynamic disturbances usually respond promptly to elevation of the legs, but in some patients, volume expansion with intravenous saline is required. The patient may experience diaphoresis and nausea, but these events usually pass and are replaced by a feeling of well-being associated with the relief of pain. Morphine also has a vagotonic effect and may cause bradycardia or advanced degrees of heart block, particularly in patients with inferior infarction. These side effects usually respond to atropine (0.5 mg intravenously). Morphine may also slow the gastrointestinal absorption of oral medicines, which may delay the onset of action of orally administered antiplatelet therapy; however, currently available clinical data have not demonstrated an increase in the risk of adverse clinical outcomes as a result of any interaction between morphine and antiplatelet agents. Therefore, it is reasonable to use morphine in patients with STEMI to relieve pain. Morphine is administered by repetitive (every 5 min) intravenous injection of small doses (2–4 mg).

Intravenous *beta blockers* are also useful in mitigating the ischemic pain of STEMI. These drugs control pain effectively in some patients, presumably by diminishing myocardial O_2 demand and hence ischemia. More important, there is evidence that intravenous beta blockers reduce the risks of reinfarction and ventricular fibrillation (see "Beta-Adrenoceptor Blockers" below). A commonly employed regimen is metoprolol, 5 mg every 2–5 min for a total of three doses, provided the patient has a heart rate >60 beats/min, systolic pressure >100 mmHg, a PR interval <0.24 s, and no signs of acute heart failure.

Patient selection is important when considering beta blockers for STEMI. Oral beta blocker therapy should be initiated in the first 24 h for patients who do *not* have any of the following: (1) signs of acute heart failure, (2) evidence of a low-output state, (3) increased risk for cardiogenic shock, or (4) other relative contraindications to beta blockade (bradycardia, PR interval >0.24 s, second- or third-degree heart block, active asthma, or reactive airway disease). In eligible patients, 15 min after the last intravenous dose, an oral regimen is initiated with 50 mg every 6 h for 48 h, followed by 100 mg every 12 h.

Unlike beta blockers, calcium antagonists are of little value in the acute setting, and there is evidence that short-acting dihydropyridines may be associated with an increased mortality risk.

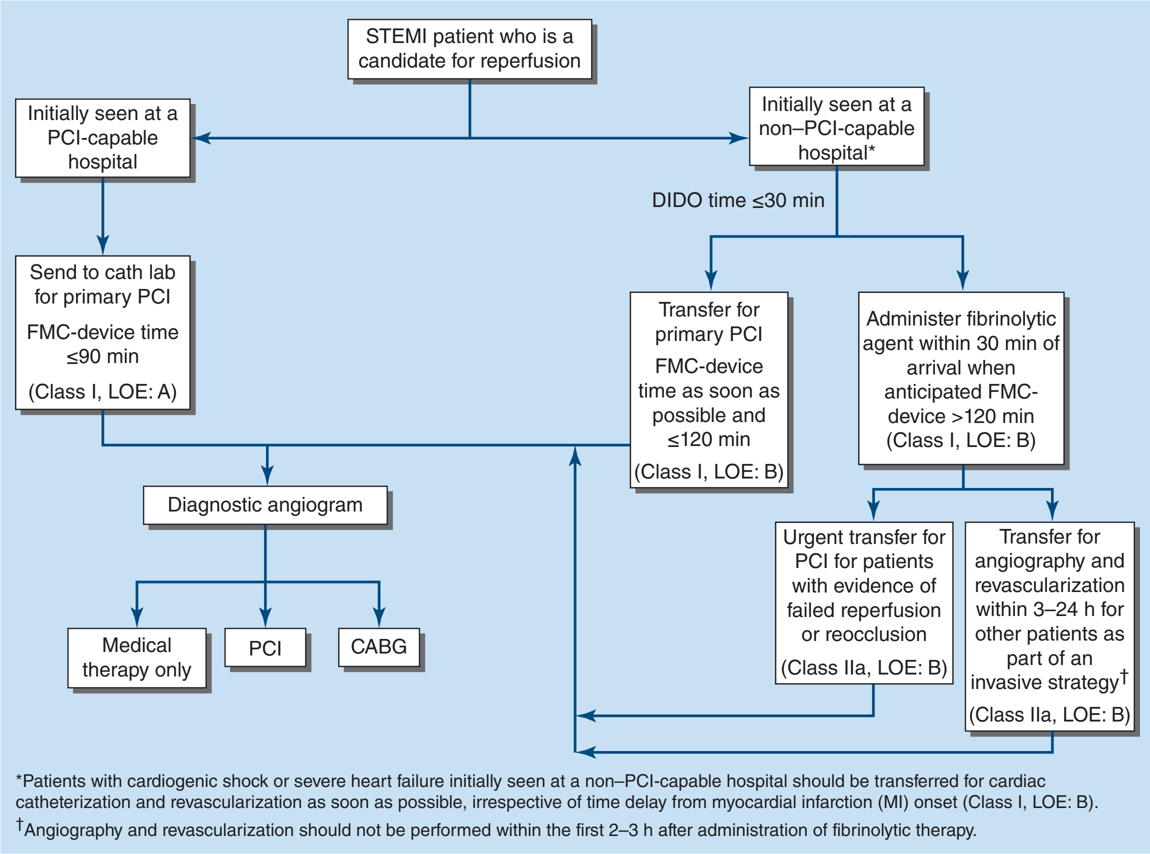
STEMI patient who is a candidate for reperfusion
Initially seen at a PCI-capable hospital
Initially seen at a non–PCI-capable hospital*
DIDO time ≤30 min
Send to cath lab for primary PCI
FMC-device time ≤90 min
(Class I, LOE: A)
Transfer for primary PCI
FMC-device time as soon as possible and ≤120 min
(Class I, LOE: B)
Administer fibrinolytic agent within 30 min of arrival when anticipated FMC-device >120 min
(Class I, LOE: B)
Diagnostic angiogram
Urgent transfer for PCI for patients with evidence of failed reperfusion or reocclusion
(Class IIa, LOE: B)
Transfer for angiography and revascularization within 3–24 h for other patients as part of an invasive strategy†
(Class IIa, LOE: B)
Medical therapy only
PCI
CABG

*Patients with cardiogenic shock or severe heart failure initially seen at a non–PCI-capable hospital should be transferred for cardiac catheterization and revascularization as soon as possible, irrespective of time delay from myocardial infarction (MI) onset (Class I, LOE: B).

†Angiography and revascularization should not be performed within the first 2–3 h after administration of fibrinolytic therapy.

FIGURE 286-5 Reperfusion therapy for patients with ST-segment elevation myocardial infarction (STEMI). Performance of percutaneous coronary intervention (PCI) is dictated by an anatomically appropriate culprit stenosis. CABG, coronary artery bypass graft; DIDO, door-in–door-out; FMC, first medical contact; LOE, level of evidence. *(Reproduced with permission from PT O'Gara: 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction. Circulation 127:e362, 2013.)*

MANAGEMENT STRATEGIES

The primary tool for screening patients and making triage decisions is the initial 12-lead ECG. When ST-segment elevation meeting the Universal Definition of MI criteria **(Table 286-2)** is present, a patient should be considered a candidate for *reperfusion therapy* (Figs. 286-1 and 286-5). The process of selecting patients for fibrinolysis versus primary PCI **(Chap. 287)** is discussed below. In the absence of ST-segment elevation, fibrinolysis is not helpful, and evidence exists suggesting that it may be harmful.

LIMITATION OF INFARCT SIZE

The quantity of myocardium that becomes necrotic as a consequence of a coronary artery occlusion is determined by factors other than just the location of occlusion. While the central zone of the infarct contains necrotic tissue that is irretrievably lost, the fate of the surrounding ischemic myocardium (ischemic penumbra) may be improved by timely restoration of coronary perfusion, reduction of myocardial O_2 demands, prevention of the accumulation of noxious metabolites, and blunting of the impact of mediators of reperfusion injury (e.g., calcium overload and oxygen-derived free radicals). Up to one-third of patients with STEMI may achieve *spontaneous* reperfusion of the infarct-related coronary artery within 24 h and experience improved healing of infarcted tissue.

TABLE 286-2 Electrocardiographic Criteria for ST-Elevation Myocardial Infarction

New ST-elevation at the J-point in 2 contiguous leads with the cut-point:

- In all leads other than leads V_2–V_3: ≥0.1 mV (1 mm on standard scale)
- In leads V_2–V_3:
  - ≥0.2 mV (2 mm) in men ≥40 years old
  - ≥0.25 mV (2.5 mm) in men <40 years
  - ≥0.15 mV (1.5 mm) in women regardless of age

Source: Reproduced from K Thygesen et al: Fourth universal definition of myocardial infarction (2018). Circulation 138:e618, 2018.

Reperfusion, either pharmacologically (by fibrinolysis) or by primary PCI, accelerates the opening of infarct-related arteries in those patients in whom spontaneous fibrinolysis ultimately would have occurred and also greatly increases the number of patients in whom restoration of flow in the infarct-related artery is accomplished. Timely restoration of flow in the epicardial infarct–related artery combined with improved perfusion of the downstream zone of infarcted myocardium results in a limitation of infarct size. Protection of the ischemic myocardium by the maintenance of an optimal balance between myocardial O_2 supply and demand through pain control, treatment of heart failure (HF), and minimization of tachycardia and hypertension extends the "window" of time for the salvage of myocardium by reperfusion strategies.

Glucocorticoids and nonsteroidal anti-inflammatory agents, with the exception of aspirin, should be avoided in patients with STEMI. They can impair infarct healing and increase the risk of myocardial rupture, and their use may result in a larger infarct scar. In addition, they can increase coronary vascular resistance, thereby potentially reducing flow to ischemic myocardium.

PRIMARY PERCUTANEOUS CORONARY INTERVENTION

(See also Chap. 287) PCI, usually stenting without preceding fibrinolysis, referred to as *primary PCI*, is effective in restoring perfusion in STEMI when carried out on an emergency basis in the first few hours of MI. It has the advantage of being applicable to patients who have contraindications to fibrinolytic therapy (see below) but otherwise are considered appropriate candidates for reperfusion. Primary PCI is more effective than fibrinolysis in opening occluded coronary arteries and, *when performed by experienced teams in dedicated medical centers*, is associated with better short-term and long-term clinical outcomes. Compared with fibrinolysis, primary PCI is preferred unless the time to delivery of the first coronary device is anticipated to be longer than 120 min. Primary PCI is also preferred when the diagnosis is in doubt, cardiogenic shock is present, bleeding risk is high, or symptoms have been present for longer than 2–3 h, at which time the clot is less easily lysed by fibrinolytic drugs. Contemporary studies (PRAMI, CvLPRIT, COMPLETE) provide evidence that, in patients with STEMI without shock, performing PCI on nonculprit coronary vessels, either during the index procedure or within 45 days, results in a lower rate of cardiovascular events and a lower need for subsequent ischemia-driven revascularization. In contrast, among patients with cardiogenic shock, routine PCI of nonculprit arteries during the initial primary PCI is contraindicated due to an increase in the rate of death or severe renal failure compared with culprit-only PCI.

■ FIBRINOLYSIS

If timely primary PCI is not available and no contraindications are present (see below), fibrinolytic therapy should ideally be initiated within 30 min of presentation (i.e., door-to-needle time ≤30 min). The principal goal of fibrinolysis is prompt restoration of full coronary arterial patency. The fibrinolytic agents tissue plasminogen activator (tPA), streptokinase, tenecteplase (TNK), and reteplase (rPA) have been approved by the U.S. Food and Drug Administration for intravenous use in patients with STEMI. These drugs all act by promoting the conversion of plasminogen to plasmin, which subsequently lyses fibrin thrombi. Although considerable emphasis was first placed on a distinction between more fibrin-specific agents, such as tPA, and non-fibrin-specific agents, such as streptokinase, it is now recognized that these differences are only relative, as some degree of systemic fibrinolysis occurs with the former agents. TNK and rPA are referred to as *bolus fibrinolytics* since their administration does not require a prolonged intravenous infusion.

When assessed angiographically, flow in the culprit coronary artery is described by a simple qualitative scale called the *Thrombolysis in Myocardial Infarction (TIMI) grading system*: grade 0 indicates complete occlusion of the infarct-related artery; grade 1 indicates some penetration of the contrast material beyond the point of obstruction, but without perfusion of the distal coronary bed; grade 2 indicates perfusion of the entire infarct vessel into the distal bed, but with flow that is delayed compared with that of a normal artery; and grade 3 indicates full perfusion of the infarct vessel with normal flow. The latter is the goal of reperfusion therapy, because full perfusion of the infarct-related coronary artery yields far better results in terms of limiting infarct size, maintenance of LV function, and reduction of both short- and long-term mortality rates.

tPA and the other relatively fibrin-specific plasminogen activators, rPA and TNK, are more effective than streptokinase at restoring full perfusion—i.e., TIMI grade 3 coronary flow—and have a small edge in improving survival as well. The recommended regimen of tPA consists of a 15-mg bolus followed by 50 mg intravenously over the first 30 min, followed by 35 mg over the next 60 min. Streptokinase is administered as 1.5 million units (MU) intravenously over 1 h. rPA is administered in a double-bolus regimen consisting of a 10-MU bolus given over 2–3 min, followed by a second 10-MU bolus 30 min later. TNK is given as a single weight-based intravenous bolus of 0.53 mg/kg over 10 s. In addition to the fibrinolytic agents discussed earlier, pharmacologic reperfusion typically involves adjunctive antiplatelet and antithrombotic drugs, as discussed subsequently.

Clear contraindications to the use of fibrinolytic agents include a history of cerebrovascular hemorrhage at any time, a nonhemorrhagic stroke or other cerebrovascular event within the past year, marked hypertension (a reliably determined systolic arterial pressure >180 mmHg and/or a diastolic pressure >110 mmHg) at any time during the acute presentation, suspicion of aortic dissection, and active internal bleeding (excluding menses). While advanced age is associated with an increase in hemorrhagic complications, the benefit of fibrinolytic therapy in the elderly appears to justify its use if no other contraindications are present, the amount of myocardium in jeopardy appears to be substantial, and timely access to primary PCI is not available.

Relative contraindications to fibrinolytic therapy, which require assessment of the risk-to-benefit ratio, include current use of anticoagulants (international normalized ratio ≥2), a recent (<2 weeks) invasive or surgical procedure or prolonged (>10 min) cardiopulmonary resuscitation, known bleeding diathesis, pregnancy, a hemorrhagic ophthalmic condition (e.g., hemorrhagic diabetic retinopathy), active peptic ulcer disease, and a history of severe hypertension that is currently adequately controlled. Because of the risk of an allergic reaction, patients should not receive streptokinase if that agent had been received within the preceding 5 days to 2 years.

Allergic reactions to streptokinase occur in ~2% of patients who receive it. While a minor degree of hypotension occurs in 4–10% of patients given this agent, marked hypotension occurs, although rarely, in association with severe allergic reactions.

Hemorrhage is the most frequent and potentially the most serious complication. Because bleeding episodes that require transfusion are more common when patients require invasive procedures, unnecessary venous or arterial interventions should be avoided in patients receiving fibrinolytic agents. Hemorrhagic stroke is the most serious complication and occurs in ~0.5–0.9% of patients being treated with these agents. This rate increases with advancing age, with patients >70 years experiencing roughly twice the rate of intracranial hemorrhage as those <65 years. Large-scale trials have suggested that the rate of intracranial hemorrhage with tPA or rPA is slightly higher than with streptokinase.

■ INTEGRATED REPERFUSION STRATEGY

Timely performance of PCI is the preferred reperfusion strategy in the management of STEMI. Prior approaches that segregated the pharmacologic and catheter-based approaches to reperfusion have now been replaced with an integrated approach to triage and transfer STEMI patients to receive PCI (Fig. 286-5). To achieve the degree of integration required to care for a patient with STEMI, all communities should create and maintain a regional system of STEMI care that includes assessment and continuous quality improvement of emergency medical services and hospital-based activities.

In patients who have been treated with a fibrinolytic, urgent coronary angiography should be performed if there is evidence of either (1) failure of reperfusion (persistent chest pain and ST-segment elevation >90 min), in which case a *rescue PCI* should be considered; or (2) coronary artery reocclusion (re-elevation of ST segments and/or recurrent chest pain) or the development of recurrent ischemia. Moreover, transfer to a PCI-capable center is recommended in all patients immediately after fibrinolysis with intent for angiography and PCI of the infarct-related artery, if indicated, between 3 and 24 h after successful fibrinolysis. Coronary artery bypass surgery should be reserved for patients whose coronary anatomy is unsuited to PCI but in whom revascularization appears to be advisable because of extensive jeopardized myocardium or recurrent ischemia.

HOSPITAL PHASE MANAGEMENT

■ CARDIAC INTENSIVE CARE UNITS

These units, previously described as coronary care units, are routinely equipped with continuous monitoring of the cardiac rhythm of each patient and hemodynamic monitoring in selected patients. Equally important is the organization of a highly trained team of nurses who can recognize arrhythmias and perform cardiac resuscitation, including electroshock, when necessary. The availability of electrocardiographic monitoring and trained personnel outside the coronary care unit has made it possible to admit lower-risk patients (e.g., those not hemodynamically compromised and without active arrhythmias) to "intermediate care units." However, it remains reasonable to admit patients with STEMI to a cardiac intensive care unit when high-risk features are present (e.g., hemodynamic or electrical instability) or when suitably equipped and staffed intermediate care units are not available.

The duration of stay in the coronary care unit is dictated by the ongoing need for intensive care. In general, patients with STEMI who remain at low risk (no persistent chest discomfort, HF, hypotension, or cardiac arrhythmias) may be safely transferred out of the coronary care unit within 24–48 h.

Activity Factors that increase the work of the heart during the initial hours of infarction may increase the size of the infarct. Therefore, patients with STEMI should generally be kept at bed rest for the first 6–12 h. However, in the absence of complications, patients should be encouraged, under supervision, to resume an upright posture by dangling their feet over the side of the bed and sitting in a chair within the first 24 h. This practice is psychologically beneficial and usually results in a reduction in the pulmonary capillary wedge pressure. In the absence of hypotension and other complications, patients typically are ambulating with increasing duration, anticipating that they may be discharged after 3–5 days.

Diet Because of the risk of emesis and aspiration soon after STEMI, patients should receive either nothing or only clear liquids by mouth for the first 4–12 h. The typical diet after AMI should provide ≤30% of total calories as fat and have a cholesterol content of ≤300 mg/d. Complex carbohydrates should make up 50–55% of total calories. Portions should not be unusually large, and the menu should be enriched with foods that are high in potassium, magnesium, and fiber, but low in sodium. Diabetes mellitus and hypertriglyceridemia are managed by restriction of concentrated sweets in the diet.

PHARMACOTHERAPY

■ ANTITHROMBOTIC AGENTS

The use of antiplatelet and anticoagulant therapy during the initial phase of STEMI is based on extensive laboratory and clinical evidence that thrombosis plays an important role in the pathogenesis of this condition. The primary goal of treatment with antiplatelet and anticoagulant agents is to maintain patency of the infarct-related artery, in conjunction with reperfusion strategies. A secondary goal is to reduce the patient's tendency to thrombosis and, thus, the likelihood of mural thrombus formation or deep-venous thrombosis. The degree to which antiplatelet and anticoagulant therapy achieves these goals partly determines how effectively it reduces the risk of mortality from STEMI.

As noted previously (see "Management in the Emergency Department" earlier), aspirin is the standard antiplatelet agent for patients with STEMI. The most compelling evidence for the benefits of antiplatelet therapy (mainly with aspirin) in STEMI is found in the comprehensive overview by the Antiplatelet Trialists' Collaboration. Data from nearly 20,000 patients with MI enrolled in 15 randomized trials were pooled and revealed a relative reduction of 27% in the mortality rate, from 14.2% in control patients to 10.4% in patients receiving antiplatelet agents.

Inhibitors of the $P2Y_{12}$ ADP receptor prevent activation and aggregation of platelets. The addition of the $P2Y_{12}$ inhibitor clopidogrel to background treatment with aspirin to STEMI patients reduces the risk of clinical events (death, reinfarction, stroke) and, in patients receiving fibrinolytic therapy, has been shown to prevent reocclusion of a successfully reperfused infarct artery. Third-generation oral $P2Y_{12}$ ADP receptor antagonists, such as prasugrel and ticagrelor, are more effective than clopidogrel in preventing ischemic complications in STEMI patients undergoing PCI but are associated with an increased risk of bleeding. Glycoprotein IIb/IIIa receptor inhibitors may be used when thrombotic complications occur during PCI.

The anticoagulant agent most commonly used in clinical practice is unfractionated heparin (UFH). Use of UFH is recommended in patients with STEMI undergoing primary PCI. In addition, the available data suggest that when UFH is added to a regimen of aspirin and a non-fibrin-specific thrombolytic agent such as streptokinase, additional mortality benefit occurs (about 5 lives saved per 1000 patients treated). The immediate administration of intravenous UFH, in addition to a regimen of aspirin and relatively fibrin-specific fibrinolytic agents (tPA, rPA, or TNK), helps to maintain patency of the infarct-related artery. This effect is achieved at the cost of a small increased risk of bleeding. The recommended dose of UFH is an initial bolus of 60 U/kg (maximum 4000 U) followed by an initial infusion of 12 U/kg per h (maximum 1000 U/h). The activated partial thromboplastin time during maintenance therapy should be 1.5–2 times the control value.

Alternatives to UFH for anticoagulation of patients with STEMI are the low-molecular-weight heparin (LMWH) preparations, a synthetic version of the critical pentasaccharide sequence (fondaparinux), and the direct antithrombin bivalirudin. Advantages of LMWHs include high bioavailability permitting administration subcutaneously, reliable anticoagulation without monitoring, and greater anti-Xa:IIa activity. Enoxaparin has been shown to reduce significantly the composite endpoints of death/nonfatal reinfarction and death/nonfatal reinfarction/urgent revascularization compared with UFH in STEMI patients who receive fibrinolysis. Treatment with enoxaparin is associated with higher rates of serious bleeding, but net clinical benefit—a composite endpoint that combines efficacy and safety—still favors enoxaparin over UFH. Interpretation of the data on fondaparinux is difficult because of the complex nature of the pivotal clinical trial evaluating it in STEMI (OASIS-6). Fondaparinux appears superior to placebo in STEMI patients not receiving reperfusion therapy, but its relative efficacy and safety compared with UFH are less certain. Owing to the risk of catheter thrombosis, fondaparinux should not be used alone at the time of coronary angiography and PCI but should be combined with another anticoagulant with antithrombin activity such as UFH or bivalirudin. Trials of bivalirudin used an open-label design to evaluate its efficacy and safety compared with UFH plus a glycoprotein IIb/IIIa inhibitor. Bivalirudin was associated with a lower rate of bleeding, largely driven by reductions in vascular access site hematomas ≥5 cm or the administration of blood transfusions. Bivalirudin is an alternative to UFH in patients undergoing primary PCI.

Patients with an anterior location of the infarction, severe LV dysfunction, HF, a history of embolism, or two-dimensional echocardiographic evidence of mural thrombus are at increased risk of systemic or pulmonary thromboembolism. Patients with suspected mural thrombus should receive full therapeutic levels of anticoagulant therapy (LMWH or UFH) while hospitalized, followed by at least 3 months of oral anticoagulant therapy. It is reasonable to consider extended oral anticoagulant therapy (i.e., up to 3 months after presentation with STEMI) in patients with a high risk of developing systemic thromboembolism because of a large area of akinetic myocardium.

■ BETA-ADRENOCEPTOR BLOCKERS

The benefits of beta blockers in patients with STEMI can be divided into those that occur immediately when the drug is given acutely and those that accrue over the long term when the drug is given for secondary prevention after an infarction. Acute intravenous beta blockade, although no longer routinely recommended because of the risk of precipitating heart failure, improves the myocardial O_2 supply-demand relationship, decreases pain, reduces infarct size, and decreases the incidence of serious ventricular arrhythmias. In patients who undergo fibrinolysis soon after the onset of chest pain, no incremental reduction in mortality rate is seen with beta blockers, but recurrent ischemia and reinfarction are reduced.

Oral beta-blocker therapy after STEMI is useful for most patients (including those treated with an angiotensin-converting enzyme [ACE] inhibitor) except those in whom it is specifically contraindicated (patients with HF or severely compromised LV function, heart block, orthostatic hypotension, or a history of asthma) and perhaps those whose excellent long-term prognosis (defined as an expected mortality rate of <1% per year, patients <55 years, no previous MI, with normal ventricular function, no complex ventricular ectopy, and no angina) markedly diminishes any potential benefit.

■ INHIBITION OF THE RENIN-ANGIOTENSIN-ALDOSTERONE SYSTEM

ACE inhibitors reduce the mortality rate after STEMI, and the mortality benefits are additive to those achieved with aspirin and beta blockers. The maximum benefit is seen in high-risk patients (those who have an anterior infarction, a prior infarction, and/or globally depressed LV function), but evidence suggests that a short-term benefit occurs when ACE inhibitors are prescribed unselectively to all hemodynamically stable patients with STEMI (i.e., those with a systolic pressure >100 mmHg). The mechanism involves a reduction in ventricular remodeling after infarction (see

“Ventricular Dysfunction” later) with a subsequent reduction in the risk of HF. The rate of recurrent infarction may also be lower in patients treated chronically with ACE inhibitors after infarction.

ACE inhibitors should be continued indefinitely in patients who have clinically evident HF, in patients in whom an imaging study shows a reduction in global LV function or a large regional wall motion abnormality, or in those who are hypertensive.

Angiotensin receptor blockers (ARBs) should be administered to STEMI patients who are intolerant of ACE inhibitors and who have either clinical or radiologic signs of HF. Long-term mineralocorticoid receptor inhibition (spironolactone, eplerenone) should be prescribed for STEMI patients who are already receiving therapeutic doses of an ACE inhibitor, have an LV ejection fraction ≤40%, and have either symptomatic HF or diabetes mellitus and do not have significant renal dysfunction (creatinine ≥2.5 mg/dL in men and ≥2.0 mg/dL in women) or hyperkalemia (potassium ≥5.0 mEq/L).

Although angiotensin receptor–neprilysin inhibition with sacubitril/valsartan is more effective than ACE inhibitor therapy in patients with symptomatic HF with a reduced ejection fraction, sacubitril/valsartan was not more effective than an ACE inhibitor in preventing the development of incident HF in patients early post-MI.

■ OTHER AGENTS

Favorable effects on the ischemic process and ventricular remodeling (see below) previously led many physicians to routinely use *intravenous nitroglycerin* (5–10 μg/min initial dose and up to 200 μg/min as long as hemodynamic stability is maintained) for the first 24–48 h after the onset of infarction. However, a subsequent large, randomized trial demonstrated no benefit of intravenous nitroglycerin with respect to major outcomes in patients with STEMI. Use of intravenous nitroglycerin may be considered in patients with STEMI who have persistent hypertension, HF, or ongoing ischemia.

Results of multiple trials of different calcium antagonists have failed to establish a role for these agents in the treatment of most patients with STEMI. Therefore, the routine use of calcium antagonists cannot be recommended. Targeted control of blood glucose in diabetic patients with STEMI has been shown to reduce the mortality rate. Serum magnesium should be measured in all patients on admission, and any demonstrated deficits should be corrected to minimize the risk of arrhythmias.

COMPLICATIONS AND THEIR MANAGEMENT

Recognition and management of the complications of STEMI are essential elements of the care of this patient population **(Table 286-3)**.

■ VENTRICULAR DYSFUNCTION

After STEMI, the left ventricle undergoes a series of changes in shape, size, and thickness in both the infarcted and noninfarcted segments. This process is referred to as *ventricular remodeling* and generally precedes the development of clinically evident HF in the months to years after infarction. Soon after STEMI, the left ventricle begins to dilate. Acutely, this results from expansion of the infarct, i.e., slippage of muscle bundles, disruption of normal myocardial cells, and tissue loss within the necrotic zone, resulting in disproportionate thinning and elongation of the infarct zone. Later, lengthening of the noninfarcted segments occurs as well. The overall chamber enlargement that occurs is related to the size and location of the infarct, with greater dilation following infarction of the anterior wall and apex of the left ventricle and causing more marked hemodynamic impairment, more frequent HF, and a poorer prognosis. Progressive dilation and its clinical consequences may be ameliorated by therapy with ACE inhibitors. In patients with an ejection fraction <40%, regardless of whether or not HF is present, ACE inhibitors or ARBs, and eventually a beta blocker, should be prescribed (see “Inhibition of the Renin-Angiotensin-Aldosterone System” earlier).

■ HEMODYNAMIC ASSESSMENT

Pump failure is now the primary cause of in-hospital death from STEMI. The extent of infarction correlates well with the degree of pump failure and with mortality, both early (within 10 days of infarction) and later. The most common clinical signs are pulmonary rales and an S_3 gallop sound. Pulmonary congestion is also frequently seen on the chest roentgenogram. Elevated LV filling pressure and elevated pulmonary artery pressure are the characteristic hemodynamic findings, but these findings may result from a reduction of ventricular compliance (diastolic failure) and/or a reduction of stroke volume with secondary cardiac dilation (systolic failure) **(Chap. 264)**.

A classification originally proposed by Killip divides patients into four groups: class I, no signs of pulmonary or venous congestion; class II, moderate HF as evidenced by rales at the lung bases, S_3 gallop, tachypnea, or signs of failure of the right side of the heart, including venous and hepatic congestion; class III, severe HF and pulmonary edema; and class IV, shock with systolic pressure <90 mmHg and evidence of peripheral vasoconstriction, peripheral cyanosis, mental confusion, and oliguria. When this classification was established in 1967, the expected hospital mortality rate of patients in these classes was as follows: class I, 0–5%; class II, 10–20%; class III, 35–45%; and class IV, 85–95%. With advances in management, the mortality rate in each class has fallen, perhaps by as much as

TABLE 286-3 Mechanical Complications of ST-Elevation Myocardial Infarction

CHARACTERISTIC	VENTRICULAR SEPTAL RUPTURE	RUPTURE OF THE VENTRICULAR FREE WALL	PAPILLARY MUSCLE RUPTURE
Incidence	0.2–3% without reperfusion therapy, 0.2–0.34% with fibrinolytic therapy, 3.9% in patients with cardiogenic shock	Approximately 0.3–1%; fibrinolytic therapy does not reduce risk; primary PCI seems to reduce risk	Approximately 0.1–1% (posteromedial more frequent than anterolateral papillary muscle rupture)
Time course	Bimodal peak; within 24 h and 3–5 days; range, 1–14 days	Bimodal peak; within 24 h and 3–5 days; range, 1–14 days	Bimodal peak; within 24 h and 3–5 days; range, 1–14 days
Clinical manifestations	Chest pain, shortness of breath, hypotension	Anginal, pleuritic, or pericardial chest pain; syncope; hypotension; restlessness; sudden death	Abrupt onset of shortness of breath and pulmonary edema; hypotension
Physical findings	Harsh holosystolic murmur, thrill, S_3, accentuated S_2, pulmonary edema, RV and LV failure, cardiogenic shock	Jugular venous distention (29% of patients), pulsus paradoxus (47%), electromechanical dissociation, cardiogenic shock	A soft murmur in some cases, no thrill, variable signs of RV overload, severe pulmonary edema, cardiogenic shock
Echocardiographic findings	Ventricular septal rupture, left-to-right shunt on color flow Doppler echocardiography through the ventricular septum, pattern of RV overload	>5 mm pericardial effusion not visualized in all cases; layered, high-acoustic echoes within the pericardium (blood clot); direct visualization of tear; signs of tamponade	Hypercontractile LV, torn papillary muscle or chordae tendineae, flail leaflet, severe mitral regurgitation on color flow Doppler echocardiography
Right-heart catheterization	Increase in oxygen saturation from the RA to RV, large *v* waves	Ventriculography insensitive, classic signs of tamponade not always present (equalization of diastolic pressures in the cardiac chambers)	No increase in oxygen saturation from the RA to RV, large *v* waves, very high PCWP

Abbreviations: LV, left ventricle; PCI, percutaneous coronary intervention; PCWP, pulmonary capillary wedge pressure; RA, right atrium; RV, right ventricle.

Source: Reproduced with permission from Bohula EA, Morrow DA: ST-elevation myocardial infarction: Management, in *Braunwald’s Heart Disease,* 12th ed, Libby P et al (eds). New York, Elsevier, 2022, pp 662-713.

one-third to one-half. However, the mortality rate for patients with cardiogenic shock (class IV) has plateaued without additional improvement for the past 25 years.

Hemodynamic evidence of abnormal global LV function appears when contraction is seriously impaired in 20–25% of the left ventricle. Infarction of ≥40% of the left ventricle usually results in cardiogenic shock **(Chap. 316)**. Positioning of a balloon flotation (Swan-Ganz) catheter in the pulmonary artery permits monitoring of LV filling pressure as well as estimation of the cardiac output; this technique may be useful in patients who exhibit hypotension and/or clinical evidence of HF. With the addition of intraarterial pressure monitoring, systemic vascular resistance can be calculated as a guide to adjusting vasopressor and vasodilator therapy. Some patients with STEMI have markedly elevated LV filling pressures (>22 mmHg) and normal cardiac indices (2.6–3.6 L/[min/m^2]), while others have relatively low LV filling pressures (<15 mmHg) and reduced cardiac indices. The former patients usually benefit from diuresis, while the latter may respond to volume expansion.

HYPOVOLEMIA

This is an easily corrected condition that may contribute to the hypotension and vascular collapse associated with STEMI in some patients. It may be secondary to previous diuretic use, to reduced fluid intake during the early stages of the illness, and/or to vomiting associated with pain or medications. Consequently, hypovolemia should be identified and corrected in patients with STEMI and hypotension by cautious fluid administration during continuous monitoring of oxygenation before more vigorous forms of therapy are begun. Eventually, the cardiac output plateaus, and further increases in LV filling pressure only increase congestive symptoms and decrease systemic oxygenation without raising arterial pressure.

TREATMENT

Heart Failure

The management of HF in association with STEMI is similar to that of acute HF secondary to other forms of heart disease (avoidance of hypoxemia, diuresis, afterload reduction, inotropic support) **(Chap. 264)**, except that the benefits of digitalis administration to patients with STEMI are unimpressive. By contrast, diuretic agents are extremely effective, as they diminish pulmonary congestion in the presence of systolic and/or diastolic HF. LV filling pressure falls, and orthopnea and dyspnea improve after the intravenous administration of furosemide or other loop diuretics. Nitrates in various forms may be used to decrease preload and congestive symptoms. Oral isosorbide dinitrate, topical nitroglycerin ointment, and intravenous nitroglycerin all have the advantage over a diuretic of lowering preload through venodilation without decreasing the total plasma volume. In addition, nitrates may improve ventricular compliance if ischemia is present, as ischemia causes an elevation of LV filling pressure. Vasodilators must be used with caution to prevent serious hypotension. As noted earlier, ACE inhibitors are an ideal class of drugs for management of ventricular dysfunction after STEMI, especially for the long term. (See "Inhibition of the Renin-Angiotensin-Aldosterone System" earlier.)

CARDIOGENIC SHOCK

Prompt reperfusion, efforts to reduce infarct size, and treatment of ongoing ischemia and other complications of MI appear to have reduced the incidence of cardiogenic shock from 20 to ~7%. Among those who exhibit cardiogenic shock, only 10% of patients with this condition present with it on admission, while 90% develop it during hospitalization. Typically, patients who develop cardiogenic shock have severe multivessel coronary artery disease with evidence of "piecemeal" necrosis extending outward from the original infarct zone. **The evaluation and management of cardiogenic shock after STEMI are discussed in detail in Chap. 316.**

RIGHT VENTRICULAR INFARCTION

Approximately one-third of patients with inferior infarction demonstrate at least a minor degree of RV necrosis. An occasional patient with inferoposterior LV infarction also has extensive RV infarction, and rare patients present with infarction limited primarily to the RV. Clinically significant RV infarction causes signs of severe RV failure (jugular venous distention, Kussmaul's sign, hepatomegaly **[Chap. 246]**) with or without hypotension. ST-segment elevations of right-sided precordial ECG leads, particularly lead V_4R, are frequently present in the first 24 h in patients with RV infarction. Two-dimensional echocardiography is helpful in determining the degree of RV dysfunction. Catheterization of the right side of the heart often reveals a distinctive hemodynamic pattern resembling constrictive pericarditis (steep right atrial "y" descent and an early diastolic dip and plateau in RV waveforms) **(Chap. 281)**. Therapy consists of volume expansion to maintain adequate RV preload and efforts to improve LV performance with attendant reduction in pulmonary capillary wedge and pulmonary arterial pressures.

ARRHYTHMIAS

(See also Chaps. 251 and 253) The incidence of arrhythmias after STEMI is higher in patients seen early after the onset of symptoms. The mechanisms responsible for infarction-related arrhythmias include autonomic nervous system imbalance, electrolyte disturbances, ischemia, and slowed conduction in zones of ischemic myocardium. An arrhythmia can usually be managed successfully if trained personnel and appropriate equipment are available when it develops. Since most deaths from arrhythmia occur during the first few hours after infarction, the effectiveness of treatment relates directly to the speed with which patients come under medical observation. The prompt management of arrhythmias constitutes a significant advance in the treatment of STEMI.

Ventricular Premature Beats Infrequent, sporadic ventricular premature depolarizations occur in almost all patients with STEMI and do not require therapy. Whereas in the distant past, frequent, multifocal, or early diastolic ventricular extrasystoles (so-called warning arrhythmias) were routinely treated with antiarrhythmic drugs to reduce the risk of development of ventricular tachycardia and ventricular fibrillation, pharmacologic therapy is now reserved for patients with sustained ventricular arrhythmias. Prophylactic antiarrhythmic therapy (either intravenous lidocaine early or oral agents later) is contraindicated for ventricular premature beats in the absence of clinically important ventricular tachyarrhythmias because such therapy may increase the mortality rate. Beta-adrenoceptor blocking agents are effective in abolishing ventricular ectopic activity in patients with STEMI and in the prevention of ventricular fibrillation. As described earlier (see "Beta-Adrenoceptor Blockers"), they should be used routinely in patients without contraindications. In addition, hypokalemia and hypomagnesemia are risk factors for ventricular fibrillation in patients with STEMI; to reduce the risk, the serum potassium concentration should be adjusted to ~4.5 mmol/L and magnesium to ~2.0 mmol/L.

Ventricular Tachycardia and Fibrillation Sustained ventricular tachycardia that is well tolerated hemodynamically should be treated with an intravenous regimen of amiodarone (bolus of 150 mg over 10 min, followed by infusion of 1.0 mg/min for 6 h and then 0.5 mg/min). A less desirable but alternative regimen is procainamide (bolus of 15 mg/kg over 20–30 min; infusion of 1–4 mg/min). If ventricular tachycardia does not stop promptly, electrical cardioversion should be used **(Chap. 253)**. An unsynchronized discharge of 200 J (biphasic waveform) is used immediately in patients with ventricular fibrillation or when ventricular tachycardia causes hemodynamic deterioration. Ventricular tachycardia or fibrillation that is refractory to electroshock may be more responsive after the patient is treated with amiodarone (150-mg bolus).

Ventricular arrhythmias, including the unusual form of ventricular tachycardia known as *torsades des pointes* **(Chaps. 259 and 261)**, may occur in patients with STEMI as a consequence of ongoing ischemia or other concurrent problems (e.g., hypoxia, hypokalemia, or other electrolyte disturbances) or of the toxic effects of a QT-prolonging agent being administered to the patient. A search for such secondary causes should always be undertaken.

Although the in-hospital mortality rate is increased, the long-term survival is excellent in patients who survive to hospital discharge after *primary* ventricular fibrillation; i.e., ventricular fibrillation that is a primary response to acute ischemia that occurs during the first 48 h and is not associated with predisposing factors such as HF, shock, bundle branch block, or ventricular aneurysm. This natural history is in sharp contrast to the poor prognosis for patients who develop ventricular fibrillation *secondary* to severe pump failure. For patients who develop ventricular tachycardia or ventricular fibrillation late in their hospital course (i.e., after the first 48 h), the mortality rate is increased both in-hospital and during long-term follow-up. Such patients should be considered for implantation of a cardioverter-defibrillator (ICD) **(Chap. 259)**. A more challenging issue is the prevention of sudden cardiac death from ventricular fibrillation late after STEMI in patients who have not exhibited sustained ventricular tachyarrhythmias during their index hospitalization. An algorithm for selection of patients who warrant prophylactic implantation of an ICD is shown in **Fig. 286-6**.

Accelerated Idioventricular Rhythm Accelerated idioventricular rhythm (AIVR, "slow ventricular tachycardia"), a ventricular rhythm with a rate of 60–100 beats/min, often occurs transiently during fibrinolytic therapy at the time of reperfusion. For the most part, AIVR, whether it occurs in association with fibrinolytic therapy or spontaneously, is benign, and does not presage the development of classic ventricular tachycardia. Most episodes of AIVR do not require treatment if the patient is monitored carefully, as degeneration into a more serious arrhythmia is rare; overly aggressive treatment can lead to complete heart block.

Supraventricular Arrhythmias Sinus tachycardia is the most common supraventricular arrhythmia. If it occurs secondary to another cause (such as anemia, fever, HF, or a metabolic derangement), the primary problem should be treated first. However, if it appears to be due to sympathetic overstimulation (e.g., as part of a hyperdynamic state), then treatment with a beta blocker is indicated. Other common arrhythmias in this group are atrial flutter and atrial fibrillation, which are often secondary to LV failure. Digoxin may be considered for supraventricular arrhythmias if HF with reduced ejection fraction is present. If left ventricular systolic dysfunction is absent, beta blockers, verapamil, and diltiazem are suitable alternatives for controlling the ventricular rate, as they may also help to control ischemia. If the abnormal rhythm persists for a prolonged period with a high ventricular rate (e.g., >120 beats/min) or if tachycardia induces HF, shock, or ischemia (as manifested by recurrent pain or ECG changes), a synchronized electroshock should be used.

Accelerated junctional rhythms have diverse causes but may occur in patients with inferoposterior infarction. Digitalis excess must be ruled out. In some patients with severely compromised LV function, the loss of appropriately timed atrial systole results in a marked reduction of cardiac output. Right atrial or coronary sinus pacing is indicated in such instances.

Sinus Bradycardia Treatment of sinus bradycardia is indicated if hemodynamic compromise results from the slow heart rate. Atropine is the most useful drug for increasing heart rate and should be given intravenously in doses of 1.0 mg initially. If the rate remains

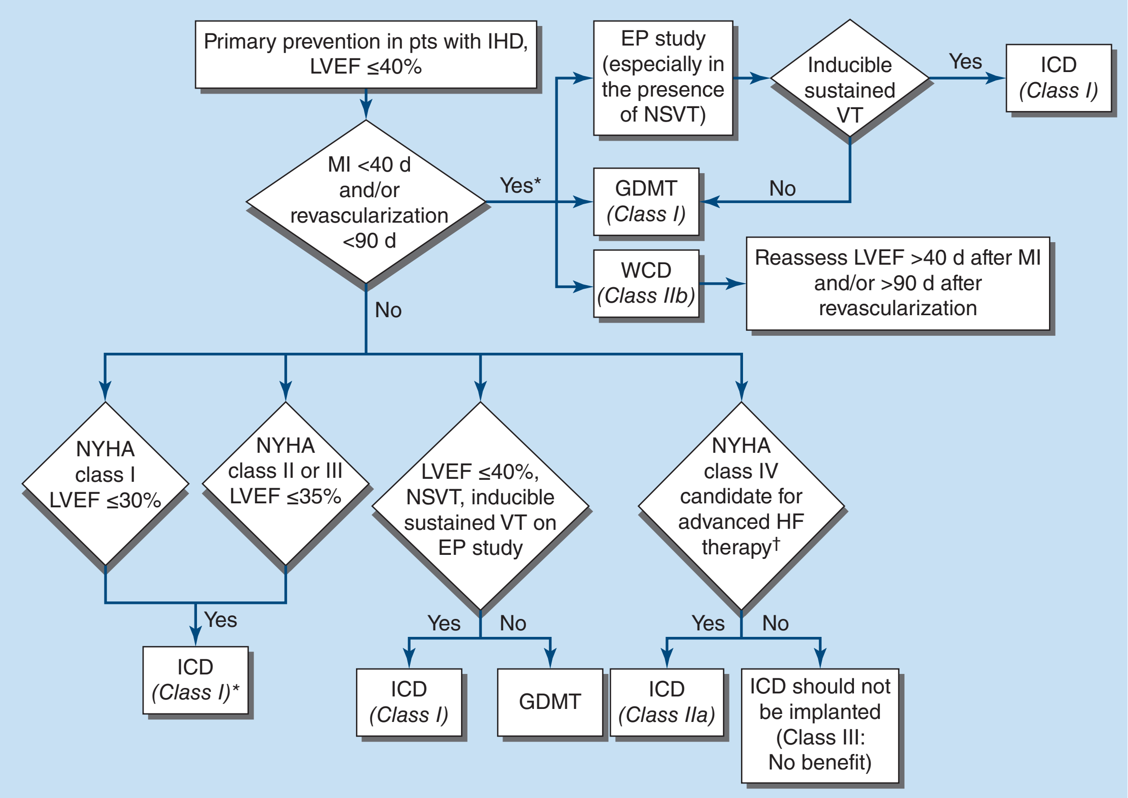
Primary prevention in pts with IHD, LVEF ≤40%
MI <40 d and/or revascularization <90 d
Yes*
EP study (especially in the presence of NSVT)
Inducible sustained VT
Yes
ICD (Class I)
No
GDMT (Class I)
WCD (Class IIb)
Reassess LVEF >40 d after MI and/or >90 d after revascularization
No
NYHA class I LVEF ≤30%
NYHA class II or III LVEF ≤35%
LVEF ≤40%, NSVT, inducible sustained VT on EP study
NYHA class IV candidate for advanced HF therapy†
Yes
ICD (Class I)*
Yes
No
ICD (Class I)
GDMT
Yes
No
ICD (Class IIa)
ICD should not be implanted (Class III: No benefit)

FIGURE 286-6 Primary prevention of SCD in patients with ischemic heart disease, including recent MI. *Scenarios exist for early ICD placement in select circumstances such as patients with a pacing indication or syncope. †Advanced HF therapy includes CRT, cardiac transplant, and left ventricular assist device. CRT, cardiac resynchronization therapy; EP, electrophysiologic; GDMT, guideline-directed management and therapy; HF, heart failure; ICD, implantable cardioverter-defibrillator; IHD, ischemic heart disease, LVEF, left ventricular ejection fraction; MI, myocardial infarction; NSVT, nonsustained ventricular tachycardia; NYHA, New York Heart Association; pts, patients; SCD, sudden cardiac death; VT, ventricular tachycardia; WCD, wearable cardioverter-defibrillator. The available evidence does not suggest there is a survival advantage to the use of an ICD early after MI, and the WCD is a potential option while waiting until the ejection fraction is reassessed (see figure). While the WCD appears to be effective in patients who wear the device, it is associated with frequent alarms, skin irritation, and emotional distress, which results in reduced wear time in a large number of patients. *(Reproduced with permission from SM Al-Khatib et al: 2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death. Circulation 138:e272, 2018.)*

<50–60 beats/min, additional doses, up to a total of 3.0 mg, may be given. Persistent bradycardia (<40 beats/min) despite atropine may be treated with electrical pacing.

Atrioventricular and Intraventricular Conduction Disturbances **(See also Chap. 251)** Both the in-hospital mortality rate and the postdischarge mortality rate of patients who have complete atrioventricular (AV) block in association with anterior infarction are markedly higher than those of patients who develop AV block with inferior infarction. This difference is related to the fact that heart block in inferior infarction is commonly a result of increased vagal tone and/or the release of adenosine and, therefore, is transient. In anterior wall infarction, however, heart block is usually related to ischemic malfunction of the conduction system, which is commonly associated with extensive myocardial necrosis.

Temporary electrical pacing provides an effective means of increasing the heart rate of patients with bradycardia due to AV block. However, acceleration of the heart rate may have only a limited impact on prognosis in patients with anterior wall infarction and complete heart block in whom the large size of the infarct is the major factor determining outcome. It should be carried out if it improves hemodynamics. Pacing does appear to be beneficial in patients with inferoposterior infarction who have complete heart block associated with HF, hypotension, marked bradycardia, or significant ventricular ectopic activity. A subgroup of these patients, those with RV infarction, often respond poorly to ventricular pacing because of the loss of the atrial contribution to ventricular filling. In such patients, dual-chamber AV sequential pacing may be required.

External noninvasive pacing electrodes should be positioned in a "demand" mode for patients with sinus bradycardia (rate <50 beats/min) that is unresponsive to drug therapy, Mobitz II second-degree AV block, third-degree heart block, or bilateral bundle branch block (e.g., right bundle branch block plus left anterior fascicular block). Retrospective studies suggest that permanent pacing may reduce the long-term risk of sudden death due to bradyarrhythmias in the rare patient who develops combined persistent bifascicular and transient third-degree heart block during the acute phase of MI.

■ OTHER COMPLICATIONS

Recurrent Chest Discomfort Because recurrent or persistent ischemia often heralds extension of the original infarct or reinfarction in a new myocardial zone and is associated with a near tripling of mortality after STEMI, patients with these symptoms should be referred for prompt coronary arteriography and mechanical revascularization. Administration of a fibrinolytic agent is an alternative to early mechanical revascularization in patients with recurrent ischemic ST-segment elevation.

Pericarditis **(See also Chap. 281)** Pericardial friction rubs and/or pericardial pain are frequently encountered in patients with STEMI involving the epicardium. This complication can usually be managed with aspirin (650 mg four times daily). It is important to diagnose the chest pain of pericarditis accurately because failure to recognize it may lead to the erroneous diagnosis of recurrent ischemic pain and/or infarct extension, with resulting inappropriate use of anticoagulants, nitrates, beta blockers, or coronary arteriography. When it occurs, complaints of pain radiating to either trapezius muscle is helpful because such a pattern of discomfort is typical of pericarditis but rarely occurs with ischemic discomfort. Anticoagulants potentially could cause tamponade in the presence of acute pericarditis (as manifested by either pain or persistent rub) and therefore should not be used unless there is a compelling indication.

Thromboembolism Clinically apparent thromboembolism complicates STEMI in ~10% of cases, but embolic lesions are found in 20% of patients in necropsy series, suggesting that thromboembolism is often clinically silent. Thromboembolism is considered to be an important contributing cause of death in 25% of patients with STEMI who die after admission to the hospital. Arterial emboli originate from LV mural thrombi, while most pulmonary emboli arise in the leg veins.

Thromboembolism typically occurs in association with large infarcts (especially anterior), HF, and an LV thrombus detected by echocardiography. The incidence of arterial embolism from a clot originating in the ventricle at the site of an infarction is small but real. Two-dimensional echocardiography reveals LV thrombi in about one-third of patients with anterior wall infarction but in few patients with inferior or posterior infarction. Arterial embolism often presents as a major complication, such as hemiparesis when the cerebral circulation is involved. When a thrombus has been clearly demonstrated by echocardiographic or other techniques, systemic anticoagulation should be undertaken (in the absence of contraindications), as the incidence of embolic complications appears to be markedly lowered by such therapy. The appropriate duration of therapy is unknown, but 3–6 months is reasonable.

Left Ventricular Aneurysm The term *ventricular aneurysm* is usually used to describe *dyskinesis* or local expansile paradoxical wall motion. Normally functioning myocardial fibers must shorten more if stroke volume and cardiac output are to be maintained in patients with ventricular aneurysm; if they cannot, overall ventricular function is impaired. True aneurysms are composed of scar tissue and neither predispose to nor are associated with cardiac rupture.

The complications of LV aneurysm do not usually occur for weeks to months after STEMI; they include HF, arterial embolism, and ventricular arrhythmias. Apical aneurysms are the most common and the most easily detected by clinical examination. The physical finding of greatest value is a double, diffuse, or displaced apical impulse. Ventricular aneurysms are readily detected by two-dimensional echocardiography, which may also reveal a mural thrombus in an aneurysm.

Rarely, myocardial rupture may be contained by a local area of pericardium, along with organizing thrombus and hematoma. Over time, this *pseudoaneurysm* enlarges, maintaining communication with the LV cavity through a narrow neck. Because a pseudoaneurysm often ruptures spontaneously, it should be surgically repaired if recognized.

POSTINFARCTION RISK STRATIFICATION AND MANAGEMENT

Many clinical and laboratory factors have been identified that are associated with an increase in cardiovascular risk after initial recovery from STEMI. Some of the most important factors include persistent ischemia (spontaneous or provoked), depressed LV ejection fraction (<40%), rales above the lung bases on physical examination or congestion on chest imaging, and symptomatic ventricular arrhythmias. Other features associated with increased risk include a history of previous MI, age >75, diabetes mellitus, prolonged sinus tachycardia, hypotension, ST-segment changes at rest without angina ("silent ischemia"), an abnormal signal-averaged ECG, nonpatency of the infarct-related coronary artery (if angiography is undertaken), and persistent advanced heart block or a new intraventricular conduction abnormality on the ECG. Therapy must be individualized on the basis of the relative importance of the risk(s) present.

The goal of preventing reinfarction, pump failure, and death after recovery from STEMI has led to strategies to evaluate risk after infarction. The risk of early recurrent ischemic events has diminished substantially in the era of primary PCI, with further reduction in the setting of current practice recommendations for routine revascularization of suitable severely stenosed nonculprit coronary arteries. In contrast, in stable patients after fibrinolysis who have not undergone revascularization, submaximal exercise stress testing may be carried out before hospital discharge to detect residual ischemia and ventricular ectopy and to provide the patient with a guideline for exercise in the early recovery period. Alternatively, or in addition, a maximal (symptom-limited) exercise stress test may be carried out 4–6 weeks after infarction. Patients in whom angina is induced at relatively low workloads, those who have a large reversible defect on perfusion imaging, those with demonstrable ischemia, and those in whom exercise provokes symptomatic ventricular arrhythmias should be considered at high risk for recurrent MI or death from arrhythmia. Cardiac catheterization with coronary angiography and/or invasive electrophysiologic evaluation is advised.

Evaluation of LV function should be performed in all patients with STEMI. Recognition of a depressed LV ejection fraction by echocardiography, cardiac magnetic resonance imaging (MRI), or nuclear myocardial perfusion imaging **(Chap. 248)** identifies patients who should receive medications to inhibit the renin-angiotensin-aldosterone system. Cardiac MRI **(Chap. 248)** can provide high-resolution imaging of the extent of myocardial injury. A standard imaging agent (gadolinium) is administered, and images are obtained after a 10-min delay. Since little gadolinium enters the normal myocardium, where there are tightly packed myocytes, but does percolate into the intercellular region of the infarct zone, there is a bright signal in areas of infarction that appears in stark contrast to the dark areas of normal myocardium.

In many hospitals, a cardiac rehabilitation program with progressive exercise is initiated in the hospital and continued after discharge. Such programs should include an educational component that informs patients about their disease and its risk factors. Exercise testing performed at entry into cardiac rehabilitation also aids in formulating an individualized exercise prescription.

The usual duration of hospitalization for an uncomplicated STEMI is about 3–5 days. The remainder of the convalescent phase may be accomplished at home. During the first 1–2 weeks, the patient should be encouraged to increase activity by walking about the house and outdoors in good weather. Normal sexual activity may be resumed during this period. After 2 weeks, the physician should regulate the patient's activity on the basis of exercise tolerance. Most patients will be able to return to work within 2–4 weeks.

SECONDARY PREVENTION

Various secondary preventive measures are at least partly responsible for the improvement in the long-term mortality and morbidity rates after STEMI. Long-term treatment with an antiplatelet agent (usually aspirin) after STEMI is associated with a 25% reduction in the risk of recurrent infarction, stroke, or cardiovascular mortality (36 fewer events for every 1000 patients treated). Dual antiplatelet therapy with aspirin and a $P2Y_{12}$ receptor antagonist (clopidogrel, prasugrel, or ticagrelor) is recommended for 12 months after a STEMI, in the absence of high bleeding risk or the occurrence of bleeding complications. ACE inhibitors or ARBs and, in appropriate patients, aldosterone antagonists should be used indefinitely by patients with clinically evident HF, a reduced LV ejection fraction, or a large regional wall motion abnormality to prevent late ventricular remodeling and recurrent ischemic events.

After several decades of evaluation, there is substantial evidence that suggests that administration of an anticoagulant for secondary prevention has a favorable effect on late mortality, stroke, and reinfarction in patients hospitalized with STEMI. Nevertheless, given the multiple alternatives for antithrombotic therapy, including potent antiplatelet agents, for this indication, clinicians must weigh the potential benefits of treatment with an oral anticoagulant based on established indications for anticoagulation, the use of other antithrombotic therapies, including long-term dual antiplatelet therapy, and the risk for bleeding.

The chronic routine use of oral beta-adrenoceptor blockers for at least 1 year after STEMI is supported by well-conducted, placebo-controlled trials.

Anti-inflammatory therapy with low-dose colchicine (0.5 mg once daily) may be considered as part of a secondary preventive regimen for patients with STEMI, particularly if the patient has had a recurrent atherosclerotic vascular event despite guideline-directed medical therapy.

Finally, risk factors for *atherosclerosis* **(Chap. 244)** should be discussed with the patient and, when possible, favorably modified.

■ FURTHER READING

Gershlick AH, Price MJ: Full revascularization in the patient with ST-segment elevation myocardial infarction: The story so far. J Am Coll Cardiol 74:2724, 2019.

Krumholz HM et al: Twenty-year trends in outcomes for older adults with acute myocardial infarction in the United States. JAMA Netw Open 2:e191938, 2019.

Lawton JS et al: 2021 ACC/AHA/SCAI Guideline for Coronary Artery Revascularization: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 145:e18, 2022.

Matter MA et al: Inflammation in acute myocardial infarction: The good, the bad and the ugly. Eur Heart J 45:89, 2024.

O'Gara PT et al: 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: Executive summary: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation 127:529, 2013.

Park J et al: Prognostic implications of door-to-balloon time and onset-to-door time on mortality in patients with ST-segment-elevation myocardial infarction treated with primary percutaneous coronary intervention. J Am Heart Assoc 8:e012188, 2019.

Stahli BE et al: Timing of complete revascularization with multivessel PCI for myocardial infarction. N Engl J Med 389:1368, 2023.

Szummer K et al: From early pharmacology to recent pharmacology interventions in acute coronary syndromes: JACC state-of-the-art review. J Am Coll Cardiol 74:1618, 2019.

Thiele H et al: PCI strategies in patients with acute myocardial infarction and cardiogenic shock. N Engl J Med 377:2419, 2017.

Tsao CW et al: Heart disease statistics–2023 update: A report from the American Heart Association. Circulation 147:e93, 2023.

287

Percutaneous Coronary Interventions and Other Interventional Procedures

Deepak L. Bhatt, Pinak B. Shah

Percutaneous transluminal coronary angioplasty (PTCA) was first introduced by Andreas Gruentzig in 1977 as an alternative to coronary bypass surgery. The concept was initially demonstrated by Charles Dotter in 1964 in peripheral vessels. The development of a small inelastic balloon catheter by Gruentzig allowed expansion of the technique into smaller peripheral and coronary vessels. Initial coronary experience was limited to single-vessel coronary disease and discrete proximal lesions due to the technical limitations of the equipment. Advances in technology with smaller profile balloon catheters and movable steerable guidewires and greater operator experience allowed the procedure to grow rapidly with expanded use in patients with more complex lesions and multivessel disease. The introduction of coronary stents in 1994 was one of the major advances in the field. These devices reduced acute complications and reduced by half the significant problems of acute thrombosis and late restenosis (or recurrence of stenosis). Further reductions in restenosis were achieved by the introduction of drug-eluting stents in 2003. These stents slowly release antiproliferative drugs directly into the plaque over a few months. Percutaneous coronary intervention (PCI) is the most common revascularization procedure in the United States and is performed more than twice as often as coronary artery bypass surgery: >900,000 patients a year.

Interventional cardiology is a separate discipline that began with coronary interventions. The discipline has now expanded to include interventions for structural heart disease including treatment of congenital heart disease and valvular heart disease; it also includes interventions to treat peripheral vascular disease, including atherosclerotic and nonatherosclerotic lesions in the carotid, renal, aortic, and peripheral arterial and venous circulations.

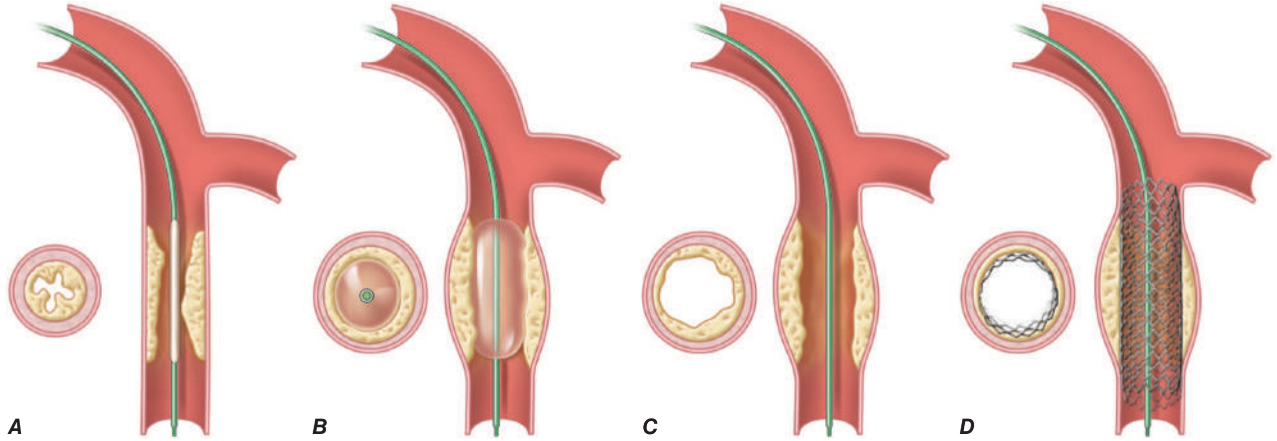

FIGURE 287-1 Schematic diagram of the primary mechanisms of balloon angioplasty and stenting. ***A.*** A balloon angioplasty catheter is positioned into the stenosis over a guidewire under fluoroscopic guidance. ***B.*** The balloon is inflated, temporarily occluding the vessel. ***C.*** The lumen is enlarged primarily by stretching the vessel, often resulting in small dissections in the neointima. ***D.*** A stent mounted on a deflated balloon is placed into the lesion and pressed against the vessel wall with balloon inflation (not shown). The balloon is deflated and removed, leaving the stent permanently against the wall acting as a scaffold to hold the dissections against the wall and prevent vessel recoil. *(Reproduced with permission from EJ Topol: Textbook of Cardiovascular Medicine, 2nd ed. Philadelphia, Lippincott Williams & Wilkins, 2002.)*

TECHNIQUE

The initial procedure is performed in a similar manner as a diagnostic cardiac catheterization **(Chap. 249)**. Arterial access is obtained via the radial or femoral artery. To prevent thrombotic complications during the procedure, patients who are anticipated to need PCI are given aspirin (325 mg) and may be given a platelet $P2Y_{12}$ inhibitor such as clopidogrel (loading dose of 600 mg), prasugrel (loading dose of 60 mg), or ticagrelor (loading dose of 180 mg) before the procedure. Cangrelor, a potent IV $P2Y_{12}$ inhibitor, is approved for use in patients who have not received an oral agent prior to the procedure. During the procedure, anticoagulation is achieved by administration of unfractionated heparin, enoxaparin (a low-molecular-weight heparin), or bivalirudin (a direct thrombin inhibitor). In patients with ST-segment elevation myocardial infarction (STEMI), high-risk acute coronary syndrome, or a large thrombus in the coronary artery, an intravenous glycoprotein IIb/IIIa inhibitor (tirofiban or eptifibatide) may also be given, although cangrelor appears to be as effective with less bleeding risk.

Following placement of an introducing sheath into the artery, preformed guiding catheters are used to cannulate selectively the origins of the coronary arteries. Through the guiding catheter, a flexible, steerable guidewire is negotiated into the coronary artery lumen using fluoroscopic guidance; it is then advanced through the stenosis and into the vessel beyond. This guidewire then serves as a "rail" over which angioplasty balloons, stents, or other therapeutic devices can be advanced to enlarge the narrowed segment of coronary artery. The artery is usually dilated with a balloon catheter followed by placement of a stent. The catheters and introducing sheath are removed and the artery manually held, or in the case of radial access, an inflatable cuff is used. One of several femoral arterial closure devices can also be used to achieve hemostasis. Because PCI is usually performed under local anesthesia and mild sedation, procedural recovery is generally less than 24 h. Most patients undergoing elective PCI can be discharged the same day of the procedure.

Balloon angioplasty works by stretching the artery and displacing the plaque away from the lumen, enlarging the entire vessel **(Figs. 287-1 and 287-2)**. The procedure rarely results in embolization of atherosclerotic material. Owing to inelastic elements in the plaque, the stretching of the vessel by the balloon results in small, localized dissections that can protrude into the lumen and be a nidus for acute thrombus formation. If the dissections are severe, then they can obstruct the lumen or induce a thrombotic occlusion of the artery (acute closure). Stents have largely prevented this complication by holding the dissection flaps up against the vessel wall (Fig. 287-1).

Stents are currently used in >90% of coronary angioplasty procedures. Stents are wire meshes (usually made of stainless steel or other

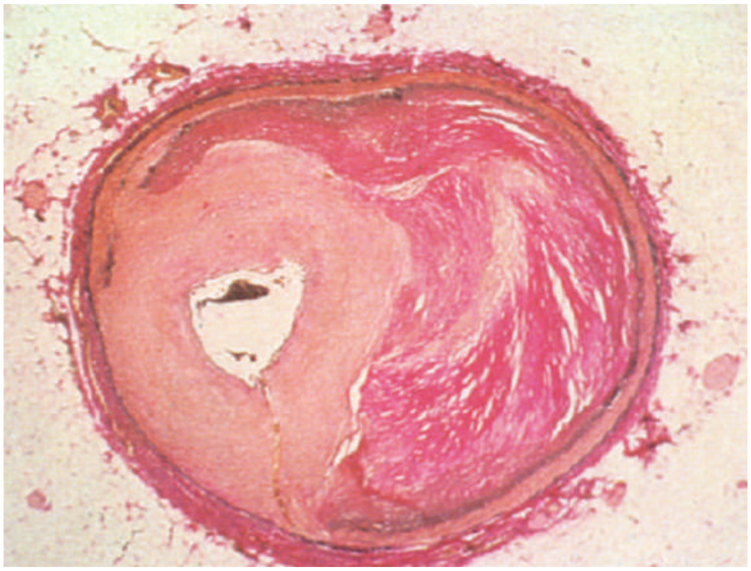

FIGURE 287-2 An example of neointimal hyperplasia and restenosis showing narrowing of the vessel. *(Reprinted from CE Essed, M Van den Brand, AE Becker: Transluminal coronary angioplasty and early restenosis. Fibrocellular occlusion after wall laceration. Br Heart J 49:393, 1983; with permission.)*

metals, such as cobalt chromium or nitinol) that are compressed over a deflated angioplasty balloon. When the balloon is inflated, the stent is enlarged to approximate the "normal" vessel lumen. The balloon is then deflated and removed, leaving the stent behind to provide a permanent scaffold in the artery. Owing to the design of the struts, these devices are flexible, allowing their passage through diseased and tortuous coronary vessels. Stents are rigid enough to prevent elastic recoil of the vessel and have dramatically improved the success and safety of the procedure as a result.

Drug-eluting stents (DESs) further enhanced the efficacy of PCI. An antiproliferative agent is attached to the metal stent by use of a thin polymer coating. The antiproliferative drug elutes from the stent over a 1- to 3-month period or longer after implantation. DESs have been shown to reduce clinical restenosis by 50%, so that in uncomplicated lesions, symptomatic restenosis occurs in 5–10% of patients. Not surprisingly, this benefit led to the rapid acceptance of these devices, and given their safety profile, nearly 100% of stents placed today are DESs. The first-generation DESs were coated with either sirolimus or paclitaxel. Second-generation DESs use newer agents such as everolimus, biolimus, and zotarolimus. These second-generation DESs appear to be more effective with fewer complications, such as early or late stent thrombosis, than the first-generation devices and, therefore, have replaced the first-generation stents. Biodegradable polymers that are used to attach the drugs to the stents may be theoretically superior to permanent polymers in preventing late stent thrombosis. In addition, the first-generation everolimus-eluting biodegradable vascular scaffold (BVS) stent had been shown to be reasonably safe with gradual degradation over several years, although concerns about late and very late stent thrombosis ultimately prevented introduction into clinical practice in the United States. Additional bioresorbable stents are under investigation. Drug-coated balloons (DCBs) are covered with an antiproliferative drug that can also reduce restenosis and are used primarily to treat in-stent restenosis.

Other interventional devices include atherectomy devices, thrombectomy catheters, and intravascular lithotripsy (IVL). These devices are designed to modify complex plaque or thrombus to facilitate and optimize stent placement. Rotational atherectomy is the most commonly used adjunctive device and is modeled after a dentist's drill, with small round burrs of 1.25–2.5 mm at the tip of a flexible wire shaft. The burr is passed over the guidewire up to the stenosis and drills atherosclerotic material. Because the atherosclerotic particles are ≤25 μm, they pass through the coronary microcirculation and rarely result in complications. The device is particularly useful in heavily calcified plaques that are resistant to balloon dilatation. Orbital atherectomy is a newer approach to calcified lesions that also relies on a spinning burr. IVL catheters emit pressure waves that fracture calcium within vessel walls, thereby increasing the compliance of the vessel and enhancing stent deployment. In acute STEMI, specialized catheters without a balloon can be used to aspirate thrombus in order to prevent embolization down the coronary vessel and to improve blood flow before angioplasty and stent placement. Current studies show that manual catheter thrombus aspiration should not be used routinely but, in certain cases of a large thrombus burden, can improve blood flow in primary PCI.

PCI of degenerated saphenous vein graft lesions has been associated with a significant incidence of distal embolization of atherosclerotic material, unlike PCI of native vessel disease. A number of distal protection devices have been shown to significantly reduce embolization and myocardial infarction in this setting. Most devices work by using a collapsible wire filter at the end of a guidewire that is expanded in the distal vessel before PCI. If atherosclerotic debris is dislodged, the basket captures the material, and at the end of the PCI, the basket is pulled into a delivery catheter, and the debris safely removed from the patient.

Intravascular imaging is an adjunct to angiographic imaging for PCI that provides information regarding optimal treatment strategy and improves the long-term durability and safety of PCI. There are two modalities available for intravascular imaging: intravascular ultrasound (IVUS) and optical coherence tomography (OCT). IVUS catheters have piezoelectric ultrasound transducers at the distal end that can provide an image from within the vessel, as well as provide information regarding plaque morphology. OCT utilizes near-infrared light to provide intraluminal imaging of the vessel. The resolution of OCT is better than IVUS; however, both modalities have been shown in meta-analyses of clinical trials to optimize stent diameters and reduce the likelihood of delayed complications, such as in-stent restenosis and stent thrombosis.

SUCCESS AND COMPLICATIONS

A successful procedure (angiographic success), defined as a reduction of the stenosis to less than a 20% diameter narrowing, occurs in 95–99% of patients. Lower success rates are seen in patients with tortuous, small, or calcified vessels or chronic total occlusions. Chronic total occlusions have the lowest success rates, and their recanalization is significantly better if the occlusion is recent (within 3 months) or there are favorable anatomic features. Improvements in equipment and complex antegrade and retrograde wire crossing techniques have increased the success rates of recanalization of chronic total occlusions to 70–80%.

Serious complications are rare but include a mortality rate of 0.1–0.3% for elective cases, a large myocardial infarction in <3%, and stroke in 0.1–0.4%. Patients who are older (>65 years), undergoing an emergent or urgent procedure, have chronic kidney disease, present with STEMI, or are in shock have significantly higher risk. Scoring systems can help to estimate the risk of the procedure. Myocardial infarction during PCI can occur for multiple reasons including an acute occluding thrombus, severe coronary dissection, embolization of thrombus or atherosclerotic material with resulting (micro) vascular occlusion, or closure of a side branch vessel at the site of angioplasty or stent placement. Most myocardial infarctions are small and only detected by a rise in the creatine phosphokinase (CPK) or troponin level after the procedure. Only those with significant enzyme elevations (>10 times the upper limit of normal) are associated with a less favorable long-term outcome. Coronary stents have largely prevented occlusive coronary dissections from angioplasty due to the scaffolding effect of the stent.

All types of stents are prone to stent thrombosis (1–3%), either acute (<24 h) or subacute (1–30 days), which can be ameliorated by greater attention to full initial stent deployment and the use of dual antiplatelet therapy (DAPT) (aspirin plus a platelet $P2Y_{12}$ receptor blocker [clopidogrel, prasugrel, or ticagrelor]). Late (30 days–1 year) and very late (>1 year) stent thromboses occur very infrequently with stents but are slightly more common with first-generation DESs. Use of the second-generation stents is associated with lower rates of late and very late stent thromboses, and shorter durations of DAPT (6 months) are recommended for the stent, although longer durations may be useful depending on the underlying atherothrombotic and bleeding risks. Longer durations of DAPT up to 1 year should also be considered for patients undergoing PCI for acute coronary syndromes. Premature discontinuation of DAPT, particularly in the first month after implantation, is associated with a significantly increased risk for stent thrombosis (three- to nine-fold greater). Stent thrombosis results in death in 10–20% and myocardial infarction in 30–70% of patients. Elective surgery that requires discontinuation of antiplatelet therapy after DES implantation should be postponed until after 3 months and preferably after 6 months, if at all possible.

Restenosis, or renarrowing of the dilated coronary stenosis, is the most common complication of PCI and occurs in 20–50% of patients with balloon angioplasty alone, 10–30% of patients with bare metal stents, and 5–15% of patients with DESs within the first year. The fact that stent placement provides a larger acute luminal area than balloon angioplasty alone reduces the incidence of subsequent restenosis. DESs further reduce restenosis through a reduction in excessive neointimal growth over the stent. If restenosis does not occur, the long-term outcome is excellent **(Fig. 287-3)**. Clinical restenosis is recognized by recurrence of angina or symptoms within 12 months of the procedure. Less frequently, patients with restenosis can present with non-ST-segment elevation myocardial infarction (NSTEMI) (10%) or STEMI (2%) as well. Very late stent thrombosis and restenosis after 1 year are more likely to be due to neoatherosclerosis than intimal hyperplasia seen within the first year. Clinical restenosis requires confirmation of a significant stenosis at the site of the prior PCI. The

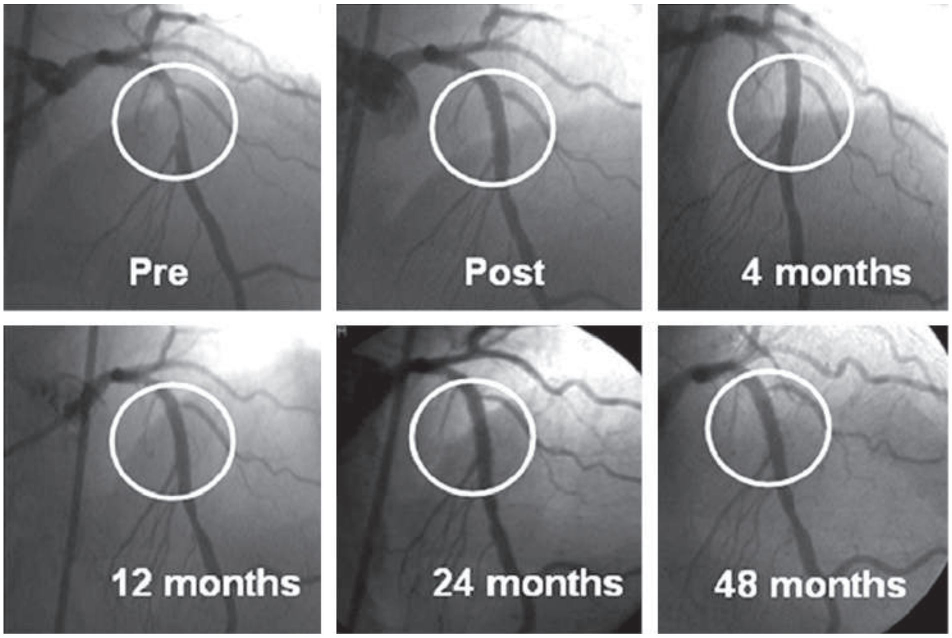

FIGURE 287-3 Long-term results from one of the first patients to receive a sirolimus-eluting stent from early Sao Paulo experience. *(From GW Stone, in D Baim [ed]: Cardiac Catheterization, Angiography and Intervention, 7th ed. Philadelphia, Lippincott Williams & Wilkins, 2006; with permission.)*

management of clinical restenosis is usually to repeat the PCI with balloon dilatation and placement of another DES. Once a patient has had restenosis, the risk of a second restenosis is further increased. The risk factors for restenosis are diabetes, myocardial infarction, long lesions, small-diameter vessels, and suboptimal initial PCI result. Treatments for symptomatic recurrent restenosis include re-stenting, brachytherapy, DCBs, or coronary bypass surgery.

INDICATIONS

The American College of Cardiology (ACC)/American Heart Association (AHA) guidelines extensively review the indications for PCI in patients with stable angina, unstable angina, NSTEMI, and STEMI and should be referred to for a comprehensive discussion of the indications. Briefly, the two principal indications for coronary revascularization in patients with *chronic stable angina* **(Chap. 284)** are (1) to improve angina symptoms in patients who remain symptomatic despite adequate medical therapy and (2) to reduce mortality rates in patients with severe and extensive coronary disease. In patients with stable angina who are well controlled on medical therapy, studies such as the Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation (COURAGE) and Bypass Angioplasty Revascularization Investigation 2 Diabetes (BARI 2D) trials have shown that initial revascularization does not lead to better outcomes (death or myocardial infarction) and can be safely delayed until symptoms worsen or evidence of severe ischemia on noninvasive testing occurs. The International Study of Comparative Health Effectiveness with Medical and Invasive Approaches (ISCHEMIA) trial was the largest trial comparing optimal medical therapy to revascularization with PCI or coronary artery bypass grafting (CABG) in stable patients with moderate ischemia on stress testing but without left main disease or reduced left ventricular function (<35% ejection fraction). It showed that optimal medical therapy was similar to revascularization in a composite of cardiovascular death or hospitalization or in cardiovascular death and myocardial infarction at a median of 3.3 years. This trial confirms prior studies and supports conservative management in most stable patients. When revascularization is indicated due to unacceptable symptoms, the choice of PCI or CABG depends on a number of clinical and anatomic factors.

The Synergy between Percutaneous Coronary Intervention with Taxus and Cardiac Surgery (SYNTAX) trial compared PCI with the paclitaxel DES to CABG in 1800 patients with three-vessel coronary disease or left main disease. The study found no difference in death or myocardial infarction at 1 year, but repeat revascularization was significantly higher in the stent-treated group (13.5 vs 5.9%), while stroke was significantly higher in the surgical group (2.2 vs 0.6%). The primary endpoint of death, myocardial infarction, stroke, or revascularization was significantly better with CABG, particularly in those with the most extensive coronary artery disease such as three-vessel disease. The 10-year results confirm these findings. The Future Revascularization Evaluation in Patients with Diabetes Mellitus: Optimal Management of Multivessel Disease (FREEDOM) trial randomized 1900 patients with diabetes and multivessel disease and found a significantly lower primary endpoint of death, myocardial infarction, or stroke with CABG than PCI. Recent trials comparing PCI with CABG have shown similar outcomes for those with less extensive disease, but a better outcome with CABG when the coronary disease is severe and extensive. These studies support CABG for those with the most severe left main and three-vessel disease or those with diabetes. Lesser degrees of multivessel disease in patients with or without diabetes have an equal outcome with PCI, including left main disease with favorable angiographic characteristics.

The choice of PCI versus CABG is also related to the anticipated procedural success and complications of PCI and the risks of CABG. For PCI, the characteristics of the coronary anatomy are critically important. The location of the lesion in the vessel (proximal or distal), the degree of tortuosity, and the size of the vessel are considered. In addition, the lesion characteristics, including the degree of the stenosis, the presence of calcium, lesion length, and presence of thrombus, are assessed. The most common reason to decide not to do PCI is that the lesion(s) felt to be responsible for the patient's symptoms is not treatable. This is most commonly due to the presence of a chronic total occlusion (>3 months in duration) with unfavorable characteristics. A lesion classification to characterize the likelihood of success or failure of PCI has been developed by the ACC/AHA. Lesions with the highest success are called type A lesions (such as proximal noncalcified subtotal lesions), and those with the lowest success or highest complication rate are type C lesions (such as chronic total occlusions). Intermediate lesions are classified as type B1 or B2 depending on the number of unfavorable characteristics. Approximately 25–30% of patients will not be candidates for PCI due to unfavorable anatomy, whereas only 5% of CABG patients will not be candidates for surgery due to coronary anatomy. The primary reason for being considered inoperable with CABG is the presence of severe comorbidities such as advanced age, frailty, severe chronic obstructive pulmonary disease (COPD), poor left ventricular function or lack of suitable surgical conduits, or poor distal targets for bypass.

Another consideration in choosing a revascularization strategy is the degree of revascularization needed. In patients with multivessel disease, bypass grafts can usually be placed to all vessels >2 mm with significant stenosis, whereas PCI may be able to treat only some of the lesions due to the presence of unfavorable anatomy. Assessment of the significance of intermediate lesions using fractional flow reserve (FFR) or the instantaneous wave-free ratio (iFR) **(Chap. 249)** can assist in determining which lesions should be revascularized. The Fractional Flow Reserve versus Angiography for Multivessel Evaluation (FAME) trial showed a 30% reduction in adverse events when revascularization by PCI was restricted to those lesions that were hemodynamically significant (FFR ≤0.80) rather than when guided by angiography alone. Trials have shown that iFR is as predictive as FFR but is quicker and easier to perform, especially if there are sequential lesions or multivessel disease. Thus, complete revascularization of all functionally significant lesions should be favored and considered when choosing the optimal revascularization strategy. Given the multiple factors that need to be considered in choosing the best revascularization for an individual patient with multivessel disease, it is optimal to have a discussion among the cardiac surgeon, interventional cardiologist, and the physicians caring for the patient (so-called heart team) to weigh the choices properly.

Patients with acute coronary syndrome are at excess risk of short- and long-term mortality. Randomized clinical trials have shown that PCI is superior to intensive medical therapy in reducing mortality and myocardial infarction, with the benefit largely confined to those patients who are high risk. High-risk non-ST-segment elevation acute coronary syndrome patients are defined as those with any one of the following: refractory ischemia, recurrent angina, positive cardiac-specific enzymes, new ST-segment depression, transient ST-segment elevation, low ejection fraction, severe arrhythmias, or a recent PCI or CABG. PCI is preferred over surgical therapy in most high-risk patients with acute coronary syndromes unless they have severe multivessel disease or the culprit lesion responsible for the unstable presentation cannot be adequately determined or treated. In STEMI, thrombolysis and PCI (primary PCI) are effective methods to restore coronary blood flow and salvage myocardium within the first 12 h after onset of chest pain. Because PCI is more effective in restoring flow than thrombolysis, it is preferred if readily available within 90 min of presentation to the hospital. PCI is also performed following thrombolysis to facilitate adequate reperfusion or as a rescue procedure in those who do not achieve reperfusion from thrombolysis, who cannot be rapidly transferred to a hospital that can perform primary PCI, or who develop cardiogenic shock. The Complete Versus Culprit-Only Revascularization Strategies to Treat Multivessel Disease After Early PCI for STEMI (COMPLETE) trial supports complete revascularization of nonculprit lesions in STEMI either in the hospital or in the few weeks after discharge.

OTHER INTERVENTIONAL TECHNIQUES

■ STRUCTURAL HEART DISEASE

Interventional treatment for structural heart disease (adult congenital heart disease and valvular heart disease) is a significant and growing component of the field of interventional cardiology.

The most common adult congenital lesion to be treated with percutaneous techniques is closure of atrial septal defects **(Chap. 280)**. The procedure is done as in a diagnostic right heart catheterization with the passage of a catheter up the femoral vein into the right atrium. With echo and fluoroscopic guidance, the size and location of the defect can be accurately defined, and closure is accomplished using one of several approved devices. All devices use a left atrial and right atrial wire mesh or covered disk that are pulled together to capture the atrial septum around the defect and seal it off. The Amplatzer Septal Occluder (AGA Medical, Minneapolis, MN) and the Gore ASD Occluder (W.L. Gore, Flagstaff, AZ) are the most commonly used devices in the United States **(Fig. 287-4)**. The success rate in selected patients is 85–95%, and the device complications are rare and include device embolization, infection, or erosion. Closure of patent foramen ovale (PFO) is done in a similar fashion. PFO closure may be considered in patients who have had recurrent paradoxical stroke or transient ischemic attack (TIA) despite adequate medical therapy including anticoagulation or antiplatelet therapy or who are at high risk for recurrent stroke. Ten-year follow-up from the Randomized Evaluation of Recurrent Stroke Comparing PFO Closure to Established Current Standard of Care Treatment (RESPECT) trial showed a benefit of closure in reducing the risk for recurrent cryptogenic stroke, as have the REDUCE and CLOSE trials. The use in the treatment of migraine is not supported by the current data.

Similar devices can also be used to close patent ductus arteriosus and ventricular septal defects. Other congenital diseases that can be treated percutaneously include coarctation of the aorta, pulmonic stenosis, peripheral pulmonary stenosis, and other abnormal communications between the cardiac chambers or vessels.

The treatment of valvular heart disease is the most rapidly growing area in interventional cardiology. In the past, the only available techniques were balloon valvuloplasty for the treatment of aortic, mitral, or pulmonic stenosis **(Chap. 272)**. Mitral valvuloplasty is the preferred treatment for symptomatic patients with rheumatic mitral stenosis

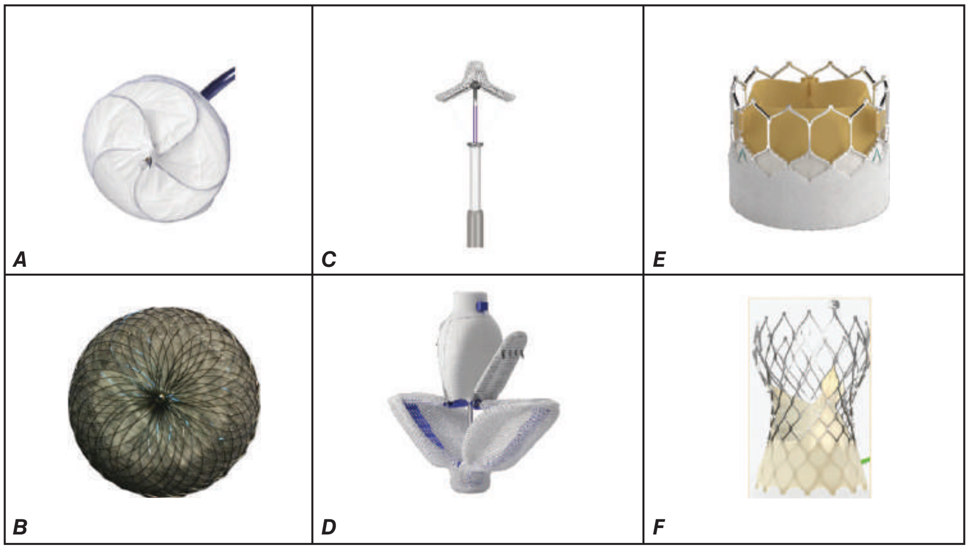

FIGURE 287-4 Devices commonly used for structural heart procedures. ***A.*** Gore Atrial Septal Occluder. *(Reproduced with permission from W.L. Gore.)* ***B.*** Amplatzer ASD Occluder. *(Reproduced with permission from Abbott.)* ***C.*** MitraClip. *(Reproduced with permission from Abbott.)* ***D.*** Pascal. *(Reproduced with permission from Edwards Lifesciences.)* ***E.*** Sapien 3 Ultra. *(Reproduced with permission from Edwards Lifesciences.)* ***F.*** Evolut FX. *(Reproduced with permission from Medtronic.)*

who have favorable anatomy. The outcome in these patients is equal to that of surgical commissurotomy. The success is highly related to the echocardiographic appearance of the valve. The most favorable setting is commissural fusion without calcification or subchordal fusion and the absence of significant mitral regurgitation. Access is obtained from the femoral vein using a transseptal technique in which a long metal catheter with a needle tip is advanced from the femoral vein through the right atrium and atrial septum at the level of the foramen ovale into the left atrium. A balloon-dilatation catheter is negotiated across the mitral valve and inflated to a predetermined size to enlarge the valve. The most commonly used dilatation catheter is the Inoue balloon. The technique splits the commissural fusion and commonly results in a doubling of the mitral valve area. The success of the procedure in favorable anatomy is 95%, and severe complications are rare (1–2%). The most common complications are tamponade due to puncture into the pericardium during the transseptal puncture or the creation of severe mitral regurgitation due to damage to the valve leaflets.

Severe mitral regurgitation can be treated percutaneously using transcatheter edge-to-edge repair (TEER). There are two devices approved for TEER including the MitraClip (Abbott, Abbott Park, IL) and the PASCAL device (Edwards Lifescience, Irvine, CA) (Fig. 287-4). The procedure involves the passage of a catheter into the left atrium using the transseptal technique. A special catheter with a metallic clasping device on the end is passed through the mitral valve and retracted to catch and bring together the mid portion of the anterior and posterior mitral valve leaflets. The device creates a double opening in the mitral valve and thereby reduces mitral regurgitation similar to the surgical Alfieri repair. In the Endovascular Valve Edge-to-Edge Repair Study (EVEREST II) trial, the MitraClip device was less effective than surgical repair or replacement but was shown to be safe. Subsequent trials have shown it to be reasonably effective for patients who are not good candidates for surgical repair, particularly when the regurgitation is due to functional causes. The Cardiovascular Outcomes Assessment of the MitraClip Percutaneous Therapy for Heart Failure Patients with Functional Mitral Regurgitation (COAPT) trial showed that, in patients with heart failure and functional mitral regurgitation who were carefully selected based on clinical and echocardiographic features, the procedure can reduce mortality.

At the present time, there are no approved devices for transcatheter mitral valve replacement (TMVR of native mitral valves). There are many devices presently in clinical trials. TMVR device development has been challenging due to the interior location of the mitral valve, the large size of the mitral valve, and potential for left ventricular outflow tract obstruction.

Severe aortic stenosis can be treated with balloon valvuloplasty as well. In this setting, the valvuloplasty balloon catheter is placed retrograde across the aortic valve from the femoral artery and briefly inflated to stretch open the valve. The success is much less favorable, with only 50% achieving an aortic valve area of >1 cm^2 and a restenosis rate of 25–50% after 6–12 months. This poor success rate has limited its use to patients who are not surgical candidates or as a bridge to surgery or transcatheter aortic valve replacement (TAVR). In this setting, the intermediate-term mortality rate of the procedure is high (10%). Repeat aortic valvuloplasty as a treatment for aortic valve restenosis has been reported.

TAVR has been shown to be an effective treatment for low-, intermediate-, and high-risk patients and inoperable patients with aortic stenosis. Currently, three valve devices are approved in the United States: the Edwards SAPIEN valve (Edwards Lifescience, Irvine, CA), the Evolut platform (Medtronic, Minneapolis, MN), and the Navitor system (Abbott Cardiovascular, Plymouth, MN) (Fig. 287-4). Data to date show excellent durability of the valves, although long-term outcomes of 10 years are not yet available. The Evolut and Navitor valves are self-expanding, whereas the Sapien valve is balloon expanded. The cannulas are large (14–18 French), and retrograde access via the femoral artery is most commonly utilized. In patients with peripheral artery disease, access via the axillary or carotid artery may be utilized. The valve is positioned across the valve and deployed, occasionally requiring postdeployment balloon inflation to ensure full contact with the aortic annulus. The success rate is >90%, and the 30-day mortality rate is 2–15% based on preoperative risk. The Placement of Aortic Transcatheter Valve (PARTNER) randomized trial of the Edwards valve showed a 55% reduction in 1-year mortality and major adverse events in the extreme-risk group randomized to TAVR compared with medical therapy. In separate randomized trials comparing the Sapien platform to surgery and the Evolut platform to surgery, low-, moderate-, and high-risk patients had similar outcomes to surgical valve replacement at 1 year. As a result, these valves have been approved for low-, intermediate-, high-, and extreme-risk patients with severe symptomatic aortic stenosis. At the present time, the Navitor platform is only approved for patients at high risk for surgery.

Aortic and mitral bioprosthetic valve degeneration can be treated with repeat surgery or, in high-risk patients, with a transcatheter valve-in-valve procedure where a percutaneous valve is placed inside of the prior surgical valve. It has been shown to be effective for aortic and mitral valves.

Pulmonic stenosis can also be effectively treated with balloon valvuloplasty and percutaneously replaced with the Melody valve (Medtronic). Tricuspid valve interventions are increasingly being performed utilizing TEER and tricuspid valve replacement devices, which are now approved for use.

PERIPHERAL ARTERY INTERVENTIONS

The use of percutaneous interventions to treat symptomatic patients with arterial obstruction in the carotid, renal, aortic, and peripheral vessels is an effective alternative to vascular surgery. Randomized clinical trial data support the use of carotid stenting in patients at high risk of complications from carotid endarterectomy **(Fig. 287-5)**. Recent trials suggest similar outcomes with carotid stenting and carotid endarterectomy in patients at average risk, although depending on the patient's risk for periprocedural stroke or myocardial infarction, one procedure may be preferred over the other. The success rate of peripheral artery interventional procedures has been improving, including treatment for long segments of occlusive disease historically treated by peripheral bypass surgery **(Fig. 287-6)**. The use of DCBs and DESs has been shown to reduce restenosis when compared with balloon angioplasty alone. Peripheral intervention is increasingly part of the training of an interventional cardiologist, and most programs now require an additional year of training after the interventional cardiology training year. The techniques and outcomes are described in detail in the chapter on peripheral vascular disease **(Chap. 292)**.

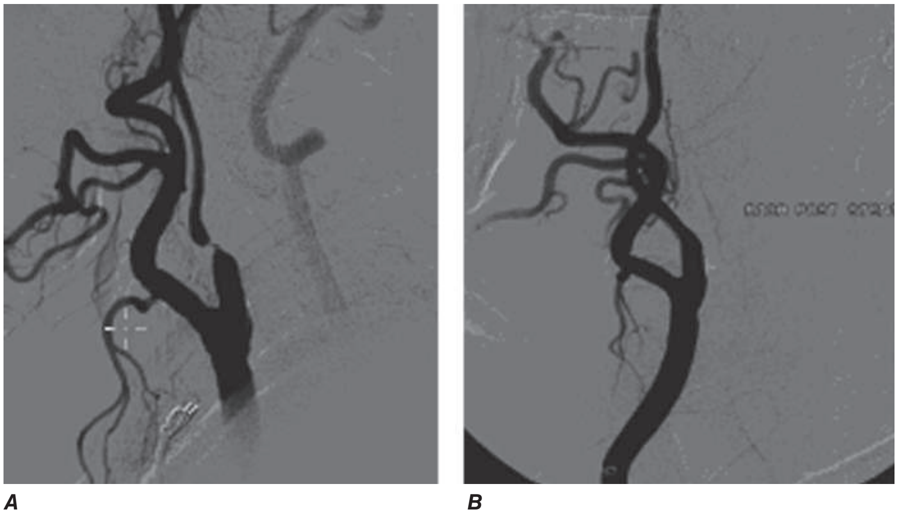

FIGURE 287-5 ***A.*** An example of a high-risk patient who requires carotid revascularization but who is not a candidate for carotid endarterectomy. ***B.*** Carotid artery stenting resulted in an excellent angiographic result. *(From M Belkin, DL Bhatt: Carotid stenting in the elderly: Is 80 the new 60? Circulation 119:2302; 2009; with permission.)*

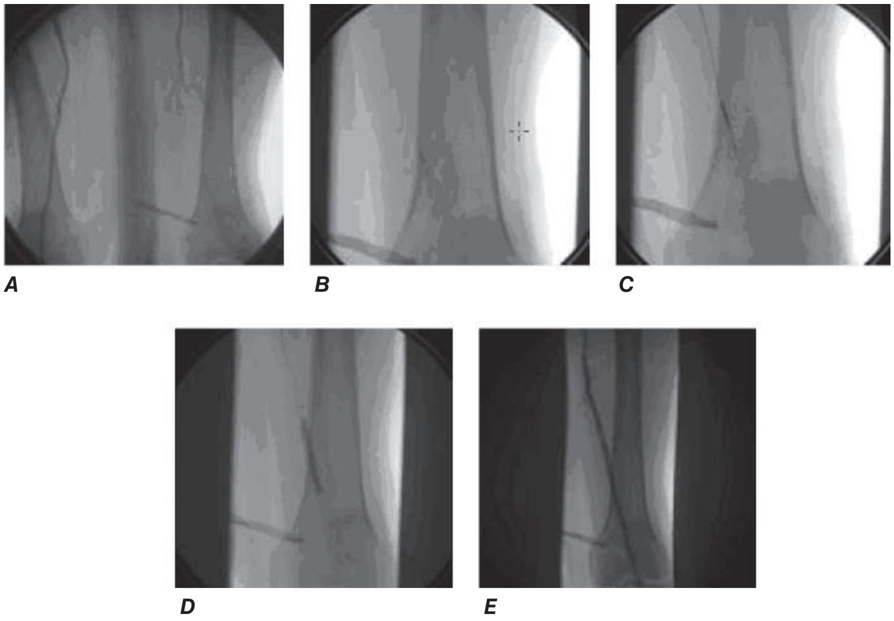

FIGURE 287-6 Peripheral interventional procedures have become highly effective at treating anatomic lesions previously amenable only to bypass surgery. ***A.*** Complete occlusion of the left superficial femoral artery. ***B.*** Wire and catheter advanced into subintimal space. ***C.*** Intravascular ultrasound positioned in the subintimal space to guide retrograde wire placement through the occluded vessel. ***D.*** Balloon dilation of the occlusion. ***E.*** Stent placement with excellent angiographic result. *(Reprinted from A Almahameed, DL Bhatt: Contemporary management of peripheral arterial disease: III. Endovascular and surgical management. Cleve Clin J Med 2006;73(suppl 4):S45-S51. With permission from The Cleveland Clinic Foundation. © 2006. The Cleveland Clinic Foundation. All rights reserved.)*

■ CIRCULATORY SUPPORT TECHNIQUES

The use of circulatory support techniques is indicated for the management of patients with shock or hemodynamic instability and occasionally is needed in order to safely perform PCI on hemodynamically unstable patients. It also can be useful in helping to stabilize patients before surgical interventions. The most commonly used device is the percutaneous intraaortic balloon pump developed in the early 1960s. A 7- to 8-French, 25- to 50-mL balloon catheter is placed retrograde from the femoral artery into the descending aorta between the aortic arch and the abdominal aortic bifurcation. It is connected to a helium gas inflation system that synchronizes the inflation to coincide with early diastole with deflation by mid-diastole. As a result, it increases early diastolic pressure, lowers systolic pressure, and lowers late diastolic pressure through displacement of blood from the descending aorta (counterpulsation). This results in an increase in coronary blood flow and a decrease in afterload. It is contraindicated in patients with aortic regurgitation, aortic dissection, or severe peripheral artery disease. The major complications are vascular and thrombotic. Intravenous heparin is given in order to reduce thrombotic complications.

Another support device is the Impella (Abiomed, Danvers, MA). The Impella CP catheter is placed percutaneously from the femoral artery into the left ventricle. The catheter has a small microaxial pump at its tip that can pump up to 2.5–5 L/min from the left ventricle to the aorta. The larger Impella 5.5 device provides greater support for longer periods of time but requires surgical implantation typically through the axillary artery. Patients can also be placed on peripheral extracorporeal membrane oxygenation (ECMO) using large cannulas placed in the femoral artery and vein. This technique can be performed emergently or electively in the catheterization laboratory and is useful for support of patients with acute respiratory failure or cardiac failure.

■ INTERVENTIONS FOR PULMONARY EMBOLISM

The treatment of deep vein thrombosis is intravenous anticoagulation, with placement of an inferior vena cava filter if recurrent pulmonary emboli (PE) occur or anticoagulation is not possible. Postphlebitic syndrome is a serious condition due to chronic venous obstruction that can lead to chronic leg edema and venous ulcers. Randomized data show that mechanical treatments may have a selective role in treatment of large iliofemoral deep-vein thrombosis.

PE should be treated with fibrinolytic agents if massive and in some cases if submassive. Surgical pulmonary embolectomy is an option for the treatment of massive PE with hemodynamic instability in patients who have contraindications for systemic fibrinolysis or those in whom it has failed. Catheter-based therapies for submassive and massive PEs are still evolving, but studies have shown promise. The techniques employed include catheter directed thrombolysis with intraclot infusion of a thrombolytic agent, potentially enhanced with ultrasound-assisted catheter-directed thrombus fragmentation. The success of these techniques has been reported to be 80–90%, with major complications occurring in 2–4% of patients. Thrombus removal with large-bore aspiration devices include the FlowTriever (Inari Medical, Irvine, CA) and the CAT12 (Penumbra, Alameda, CA).

■ INTERVENTIONS FOR REFRACTORY HYPERTENSION

The recent recognition of the importance of the renal sympathetic nerves in modulating blood pressure has led to a technique to selectively denervate renal sympathetic nerves in patients with refractory hypertension. The procedure involves applying low-power radiofrequency treatment via a catheter along the length of both renal arteries. Despite promising nonrandomized data, in the randomized Symplicity HTN-3 trial, renal denervation did not significantly reduce blood pressure compared with medical therapy. Further optimization of the technique is ongoing with some evidence from small, randomized trials of modest efficacy, which was sufficient to lead to regulatory approval.

CONCLUSION

Interventional cardiology continues to expand its borders. Treatment for coronary artery disease, including complex anatomic subsets, continues to advance. Technological advances such as DESs, now already in their second generation, are improving the results of PCI. PCI is the treatment of choice for patients with acute coronary syndromes. For patients with stable coronary disease, PCI is effective in symptom alleviation. Use of a Heart Team is the best way to make decisions concerning which revascularization—PCI or CABG—is best for an individual patient. Treatment of peripheral and cerebrovascular disease can be effective with percutaneous techniques. Structural heart disease is increasingly being treated with percutaneous options, with interventional approaches such as TAVR becoming preferred over surgical aortic valve replacement. Further growth of invasive procedures is anticipated in years to come.

FURTHER READING

Bhatt DL: *Cardiovascular Intervention: A Companion to Braunwald's Heart Disease*. Philadelphia, Elsevier, 2024.

Kumar V et al: Transcatheter aortic valve replacement programs: Clinical outcomes and developments. J Am Heart Assoc 9:120.015921, 2020.

Lawton JS et al: 2021 ACC/AHA/SCAI guideline for coronary artery revascularization: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 145:e18, 2022.

288

Hypertension

Paul K. Whelton

High blood pressure (BP) is a leading risk factor for cardiovascular disease (CVD), including ischemic and hemorrhagic stroke, coronary heart disease (CHD), heart failure (HF), peripheral arterial disease (PAD), chronic kidney disease (CKD)/end-stage kidney disease (ESKD), dementia, and all-cause mortality. Hypertension, a subset of high BP, is very common no matter how it is defined and results in a huge burden due to death, morbidity, disability, social and workplace disruption, and cost to the individual and society. According to data from the National Ambulatory Medical Care Survey (NAMCS), hypertension is a component for more than one-third of all visits to office-based physicians by U.S. adults. Likewise, global estimates have ranked hypertension as the most common reason for primary care visits worldwide. Hypertension is often associated with other CVD risk factors, resulting in a higher risk of complications and related burden of illness. The optimal approach to high BP is to prevent its development, but diagnosis, treatment, and control of hypertension also provide an effective means to reduce the risk of BP-related CVD. Unfortunately, both strategies are poorly implemented worldwide, resulting in a high burden of preventable disease.

BLOOD PRESSURE PHYSIOLOGY AND PATHOPHYSIOLOGY

Complex and incompletely understood mechanisms control blood flow in individual organs and the arteriolar system. At the most basic level, arterial BP is controlled by cardiac output and peripheral resistance **(Fig. 288-1)**. However, arterial pressure is largely thought to be controlled by a renal-volume-endocrine pressure control system, in which the blood volume and total peripheral resistance are manipulated slowly to adjust arterial BP.

Cardiac output is influenced by stroke volume and heart rate, with stroke volume being related to myocardial contraction and the size of the intravascular compartment. Changes in cardiac output play an important role in acute BP responses to stressors. Peripheral vascular resistance is determined by functional and anatomic changes in small arteries and arterioles. The vascular endothelium is composed of a single layer of endothelial cells that constitute the inner cellular lining of arterioles. It serves as a direct contact with circulating blood and regulates exchanges between the bloodstream and surrounding tissues. Endothelial cells control vascular tone and, thereby, blood flow by synthesizing and releasing relaxing and contracting factors such as nitric oxide; metabolites of arachidonic, lipoxygenase, and cytochrome pathways; peptides, including endothelin; adenosine; purines; and reactive oxygen species; and by generating endothelial enzymes that produce vasoactive hormones such as angiotensin II. Endothelial dysfunction may play an important role in the initiation or progression of hypertension and atherosclerosis. A wide variety of systems interplay to keep cardiac output and peripheral resistance in balance, including sodium handling, primarily by the kidney, and many neural and hormonal systems that influence peripheral resistance by stimulating vascular constriction (e.g., the sympathetic system and the renin-angiotensin-aldosterone system [RAAS]) or inducing vasodilation (e.g., bradykinin system), and by modulating hormones that result in excretion or retention of sodium (natriuretic peptides and aldosterone, respectively).

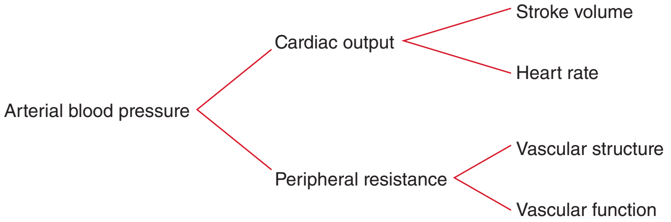

FIGURE 288-1 Schematic depiction of factors that influence the control of blood pressure.

Most vascular beds have the capacity to autoregulate blood flow. When peripheral arterial resistance is increased, the autoregulatory systems tend to increase vascular resistance within the vascular bed to maintain constant blood flow. During the 1960s, Arthur Guyton hypothesized that the capacity of the kidney to excrete sodium ultimately dictates long-term changes in levels of BP. Although the Guyton model has been challenged and alternatives proposed, it has had a profound influence on the understanding of BP regulation. A commonly accepted theory is that initial elevation of BP is due to increased cardiac output and expanded intravascular volume; peripheral resistance increases and cardiac output reverts toward normal over time. Whether or not this is the typical sequence of events in the pathogenesis of hypertension, salt can activate a number of neural, endocrine, paracrine, and vascular mechanisms that have the potential to increase arterial pressure. As arterial pressure increases in response to a high intake of sodium chloride, urinary sodium excretion increases, and sodium balance is maintained at the expense of an increase in arterial pressure. The mechanism for this "pressure-natriuresis" phenomenon may involve a subtle increase in glomerular filtration rate (GFR), decreased sodium absorption capacity in the renal tubules, and hormonal factors such as atrial natriuretic factor. In individuals with an impaired capacity to excrete sodium, greater increases in arterial pressure are required to achieve natriuresis and sodium balance.

Many of the drugs used to manage high BP work though influences on the various BP control mechanisms. For example, direct vasodilators and calcium channel blockers (CCBs) work by inducing arteriolar vasodilation. Agents that block the RAAS, such as angiotensin-converting enzyme inhibitors (ACEIs), angiotensin receptor blockers (ARBs), and renin inhibitors, act by blocking the vasoconstricting effects of angiotensin II and sodium-retaining effects of aldosterone. ACEIs also inhibit inactivation of the vasodilator bradykinin, resulting in angioedema in some patients treated with this agent. Beta blockers tend to affect BP by impairing the action of sympathetic system neurotransmitters that stimulate vasoconstriction and heart rate. For example, beta blockers can antagonize the neurotransmitter acetylcholine as well as epinephrine and norepinephrine. In addition to being the most important site for the effect of agents that block the RAAS activity, many drugs affect electrolyte and fluid exchange in the kidneys. Carbonic anhydrase inhibitors have a diuretic effect in the proximal tubules, but due to side effects and diminishing efficacy over time, they are rarely used for BP control. Thiazide and thiazide-like diuretics work to enhance sodium excretion in the distal convoluted tubule and are a mainstay of drug therapy for BP control. They work by reducing intravascular volume, but over the long term, their effect on BP is through reducing peripheral arteriolar resistance. Loop diuretics are more potent diuretics that produce sodium excretion in the loop of Henle. They tend to be relatively short-acting agents that are better suited for sodium excretion than BP reduction. Mineralocorticoid receptor antagonists (MRAs) or aldosterone antagonists are diuretic drugs that work in the distal tubule of the kidney, including the distal convoluted tubule, connecting the tubule and cortical collecting duct, to antagonize the action of aldosterone at mineralocorticoid receptor sites. These are examples of presumed principal mechanisms of action, but many drugs have other effects on the control of vascular tone and cardiac output that play a role in their effects on BP.

BLOOD PRESSURE MEASUREMENT

Commonly, BP measurements are used to estimate an individual's average level of BP; estimate their risk of CVD, in combination with other indicators; and determine the need for hypertension prevention or treatment. Office and clinic BP measurements are among the most common procedures in clinical practice and arguably provide the best assessment of BP because they have been used in almost all of the landmark BP-CVD risk prediction cohort studies and hypertension prevention and treatment trials. The need for standardization of the BP measurement procedure was recognized by early advocates of BP assessment. Subsequently, the effects of factors that can systematically (predictably) result in falsely high or low BP estimates were quantified, which formed the basis for BP measurement recommendations by professional societies and clinical practice guideline committees. BP also varies randomly (nonpredictably) within and between visits, resulting in corresponding recommendations to rely on an average of repeated readings obtained at more than one visit for classification of an individual's average level of BP. For many years, U.S. clinical practice guidelines have recommended using an average of two or more BP measurements obtained at two or more visits to estimate the usual level of BP. The most important elements for obtaining accurate office BP measurements are visually depicted in **Fig. 288-2** and outlined below.

- Instruct the patient to avoid a full bladder and abstain from caffeine, smoking, alcohol, or exercise for 30 min prior to the measurements.
- Use a quiet room with a comfortable ambient temperature.
- Explain the procedure to the patient, including the number of readings.
- Instruct the patient to rest for 3–5 min prior to the first measurement and avoid talking or distractions such as cell phone use during the rest period and subsequent BP measurements.
- Seat the patient in a chair with upright back support. The patient's feet should be flat on the ground and the measurement arm comfortably supported such that the cuff is at heart level.
- Use a clinically validated, preferably automated BP measurement device. Choose a cuff that is the correct size for the patient's arm, and measure their BP at the mid-brachial level.
- The BP measurements should be obtained by a trained, preferably certified, member of the health care team using the arm with the highest pressure at the first visit.
- Use an average of two or more readings at two or more visits to estimate the patient's usual level of BP.
- Provide the patient with the results and an interpretation of their clinical implications.

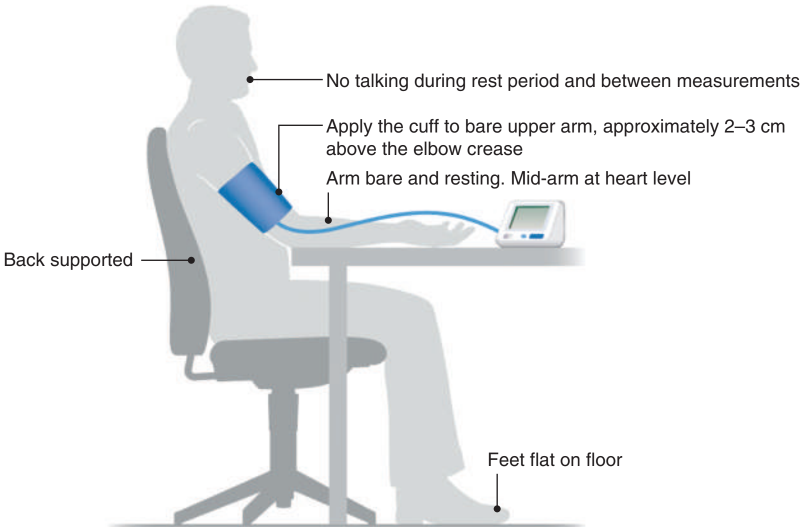
No talking during rest period and between measurements
Apply the cuff to bare upper arm, approximately 2–3 cm above the elbow crease
Arm bare and resting. Mid-arm at heart level
Back supported
Feet flat on floor

FIGURE 288-2 Schematic depiction of the important elements for accurate measurement of office blood pressure. *(Reproduced with permission from AK Cheung et al: International consensus on standardized clinic blood pressure measurement: A call to action. Am J Med 136:438, 2023.)*

TABLE 288-1 American College of Cardiology/American Heart Association Blood Pressure Classification System in Adults

BP CATEGORY	SYSTOLIC BP, mmHg		DIASTOLIC BP, mmHg
Normal BP	<120	and	<80
Elevated BP	120–129	and	<80
Stage 1 hypertension	130–139	or	80–89
Stage 2 hypertension	≥140	or	≥90

Abbreviation: BP, blood pressure.

Most clinical practice guidelines recommend use of a clinically validated oscillometric BP measurement device, a method that eliminates several sources of measurement error and does not require periodic recalibration. Several websites provide a listing of BP measurement devices that have been clinically validated to international standards. These include the Validate BP website in the United States (*https://www.validatebp.org/*) and the STRIDE BP website in Europe (*https://stridebp.org/bp-monitors*). Similarly, many free office and home BP measurement training and certification courses are available, including an easy-to-access course sponsored by the Pan American Health Organization, World Hypertension League, and others (*https://campus.paho.org/en/node/29166*).

DEFINITION OF HYPERTENSION

BP classifications systems differ by country and have changed substantially over time. The current classification proposed by the American College of Cardiology (ACC) and American Heart Association (AHA) for office/clinic BP measurements in U.S. adults is displayed in **Table 288-1**.

Correct classification presumes accurate BP measurement and averaging of two or more readings obtained at two or more occasions. When the systolic BP (SBP) and diastolic BP (DBP) readings are in different categories of BP, the higher classification should be chosen. In children and adolescents <13 years of age, hypertension is generally based on a comparison with age-, sex-, and height-specific normative data. Hypertension is defined as an average of three SBP or DBP readings at or above the 95th percentile or an SBP or DBP ≥130 or 80 mmHg, respectively. This chapter is focused on BP in adults ≥18 years.

Despite the importance and relative simplicity of accurate BP readings, errors in BP measurement are common in routine clinical practice. Generally, SBP is overestimated by an average of ~7 mmHg, which results in a 15–20% overestimation of hypertension prevalence. However, underestimation of BP and failure to diagnose hypertension are also common. Clinical practice errors vary by level of BP and practice settings and are inconsistent over time, making it impossible to use formulae to correct the errors. For those who aspire to practice evidence-based medicine, the only solution is to follow the recommendations for accurate and precise BP measurements that are advocated by the AHA, ACC, and many other organizations. Having a trained nonphysician staff member obtain routine BP measurements provides an efficient and cost-effective means of obtaining high-quality observations. This is the common practice in research studies and in many large U.S. systems of care.

OUT-OF-OFFICE BLOOD PRESSURE MEASUREMENT

Office BP measurements provide a limited and potentially biased depiction of an individual's usual level of BP. Guidelines commonly recommend complementing office BP readings with out-of-office BP measurements to confirm high office readings and to probe for higher or lower BPs outside the office. White coat hypertension, in which office BPs meet the criteria for hypertension but out-of-office BPs are nonhypertensive, and masked hypertension, in which office BPs are nonhypertensive but out-of-office BPs meet the criteria for hypertension, are common, with prevalence estimates of 15–25% for both conditions. Adults with white coat hypertension have a CVD risk profile that is more like those without than those with sustained high BPs in and out of the office and they may be best treated with nonpharmacologic interventions, with careful monitoring to recognize a transition to sustained hypertension. In contrast, adults with masked hypertension have a CVD risk profile that is like those with sustained hypertension and may be better treated with a combination of antihypertensive drugs in addition to nonpharmacologic therapy. Masked hypertension should be suspected especially in adults with an elevated but nonhypertensive office BP and evidence of end-organ damage such as left ventricular hypertrophy or proteinuria. Home BP measurements provide the most practical approach to obtaining out-of-office BP readings and provide the additional benefit of engaging patients in their own care. When home BP measurements are desired, patients should be instructed to use a clinically validated BP measurement device and trained to measure their BPs accurately and precisely using the same approaches outlined for measurement of office BPs. Most guidelines recommend using an average of two to three measurements obtained in the morning (before taking antihypertensive medications) and evening. In the United States, obtaining home BP measurements for about 3 days prior to an office visit represents a reasonable and practical option for estimating a patient's usual level of home BP. The BP for recognition of hypertension is the same for office and home BP measurements (SBP ≥130 and DBP ≥80 mmHg). Likewise, the BP control is defined as an SBP/DBP <130/80 mmHg in both settings. Many home BP measurement devices allow for storage of the readings on a memory chip and for transfer to the physician's office by use of telemetry. This approach thwarts the potential for the bias of more favorable readings that can occur with written patient self-reports. Ambulatory BP measurements are best obtained by practitioners with special expertise in this technique and interpretation of the results. The equivalent SBP/DBP readings for an office average of 130/80 mmHg are 130/80, 110/65, and 125/75 mmHg for daytime, nighttime, and 24-h ambulatory readings, respectively. Ambulatory BP measurements provide the potential for scrutiny of nighttime readings, which may provide the best BP prediction of CVD risk. Specifically, those whose BPs fail to follow the usual pattern of a pronounced decrement during the nighttime (nondippers) tend to have a higher risk of CVD. BP nondipping is more common in non-Hispanic blacks than the other major racial/ethnic groups in the United States. Ambulatory BP measurement is relatively expensive and intrusive. Patients in the United States are usually reluctant to undergo repeat ambulatory BP measurements, making this measurement method most useful in confirming the diagnosis of hypertension, especially where uncertainty is a special concern. Wearable devices, including watches, hold great promise for the provision of convenient ways to obtain a more comprehensive assessment of daytime and nighttime BPs, but none of the currently available options are clinically valid.

PREVALENCE OF HYPERTENSION

Prevalence estimates vary depending on the criteria for definition of hypertension, the methods for BP measurement, and the population being studied; hypertension is very common, especially at older age. Age-related increases in BP are noted in almost all countries. Generally, SBP increases progressively until about the eighth decade of life. DBP also increases with age, but less steeply than SBP, until about the fifth decade of life and remains stable or declines thereafter. Isolated systolic hypertension is common late in life due to a widening of the pulse pressure (difference between SBP and DBP). Generally, those who have the highest or lowest BPs early in life tend to "track" in the same extremes of BP over life, with high pressures early in life providing a crude opportunity to identify those at higher risk for hypertension in adulthood. Little or no age-related change in BP has been observed in many isolated populations and in subsets of populations in countries where age-related increases in BP are common. This observation indicates that there is no biological necessity for the commonly observed age-related increase in BP and underscores the value of interventions aimed at prevention of hypertension. Migrant studies that have tracked isolated populations from their native environment to nonnative settings have uniformly reported progression to the more common pattern of age-related increases in BP. These BP changes have been associated with increases in dietary sodium and decreases in potassium intake, consumption of a less healthy diet, decreased physical activity, increases in body weight, and increased intake of alcohol.

Based on the ACC/AHA criteria for diagnosis, >103 million adults in the United States have hypertension, making it one of the most common health conditions reported. Hypertension is present in ~20–30% of U.S. adults aged 20–44 years, but the prevalence increases to 80–85% in those aged 75 years or older. The overall prevalence of hypertension in U.S. adults has remained fairly stable in recent decades (~46% overall). Men have slightly higher BPs compared to women during the first half of life, but the opposite is true in later life. In the United States, non-Hispanic black adults have a prevalence of hypertension (59%) that is substantially higher compared to whites (45%), Hispanics (47%), or Asians (46%). Adjusted estimates of hypertension in adult non-Hispanic black men and women are reported to be 59 and 56%, respectively, whereas the corresponding estimates for non-Hispanic white men and women are 47 and 41%, respectively. In addition, hypertension in non-Hispanic black adults tends to begin at a younger age and result in more CVD and kidney complications compared to the other major racial/ethnic groups in the United States. The prevalence of hypertension and CVD complications (especially stroke) is higher in the southeastern United States compared to other regions of the country, especially the northwest. No matter how defined, hypertension is very common in all parts of the world, with a progressively increasing prevalence in low- and middle-income countries and a slight decline in high-income countries. Using the hypertension definition of SBP ≥140 mmHg, DBP ≥90 mmHg, or taking antihypertensive medication, the prevalence in adults was estimated to be 31.5% (1.04 billion) in low- and middle-income countries and 28.5% (349 million) in high-income countries in 2010. As in the United States, hypertension prevalence is very age-dependent and varies by region, race/ethnicity, and other factors, including importantly socioeconomic factors, in most countries.

LABORATORY AND OTHER INVESTIGATIONS

In patients with newly diagnosed hypertension, basic laboratory testing is indicated to (1) recognize the presence and extent of target organ damage; (2) facilitate the identification of secondary causes of hypertension, including kidney disease and primary aldosteronism; (3) recognize comorbid conditions, including diabetes mellitus and hyperlipidemia; (4) estimate atherosclerotic CVD (ASCVD) risk in patients without a history of a CVD; and (5) assist in the optimal choice of antihypertensive drug therapy. At a minimum, a complete blood count, serum electrolytes (sodium, potassium, calcium), serum creatinine and estimated GFR (eGFR), lipid profile, glycemic status (hemoglobin A_{1c} or fasting blood glucose), thyroid-stimulating hormone level, urinalysis and urine albumin-to-creatinine ratio, and 12-lead electrocardiogram (ECG) should be obtained. These basic tests can be complemented by other evaluations as clinically indicated. In some countries, it is not possible to obtain the minimally recommended laboratory results. Guidelines written for clinicians in such countries, including the 2021 World Health Organization (WHO) guideline for the treatment of hypertension in adults, recommend that this should not be an impediment to BP reduction.

ESTIMATION OF CARDIOVASCULAR DISEASE RISK

Abundant studies have shown that BP is associated with CVD risk in a continuous, progressive, log-linear fashion from low to high levels of SBP and DBP. For SBP, the risk relationship has been documented from as low as 90 mmHg to more than 180 mmHg, suggesting relatively low levels of BP, if physiological, may be optimal to prevent the genesis of BP-related atherosclerotic complications. It also suggests that one might reasonably expect a "lower BP is better" finding in antihypertensive randomized controlled treatment (RCT) trials. In addition to CVD, the level of BP is strongly associated with other diseases, including CKD, ESKD, and dementia. The BP-CVD risk association in observational studies is equally true for men and women. The BP-related slope for relative risk of CVD is steeper at younger age when BP elevation is often an isolated risk predictor. In older adults, the corresponding slope for CVD risk is less steep, but the absolute risk of CVD is far higher because other CVD risk factors are common in those with an elevated BP. At any level of BP, the risk of a CVD complication varies dramatically with more than a 30-fold difference in 10-year predicted risk of CVD for those with an isolated elevation of BP compared to their counterparts with multiple CVD risk factors in addition to an elevated BP. This observation has special relevance for clinical decision-making in patients with a usual SBP between 130 and 139 mmHg. In individuals with stage 1 hypertension as an isolated CVD risk factor, the 5- or 10-year risk of a CVD event may be quite small and limit enthusiasm for introducing antihypertensive drug therapy. However, when stage 1 hypertension is accompanied by other CVD risk factors, the corresponding risk of ASCVD may be quite high and make prescription of antihypertensive drug therapy much more appealing. In antihypertensive clinical trials, the relative risk reductions for CVD events are similar in groups with different levels of underlying CVD risk, but the absolute benefit is much greater in the groups at higher risk for ASCVD. For both these reasons, CVD risk assessment should be part of the initial evaluation.

Patients with a history of a prior CVD complication are known to be at high risk for recurrent CVD events. In those without a history of prior CVD, ASCVD risk should be estimated using a risk calculator. For U.S. adults 40–75 years of age, use of the ACC/AHA pooled cohort equations estimator is recommended because it has been validated in non-Hispanic white and black adults. Those with a 10-year risk of ASCVD ≥10% should be considered at high risk. PREVENT is an updated risk predicting model that is based on a larger and more current dataset than that used for the pooled cohort equations model. It incorporates an assessment of kidney (glomerular) function and allows for inclusion of urinary albumin/creatinine ratio, zip code for assessment of the social determinants of health, and hemoglobin A_{1c} where indicated and has been introduced by the AHA as a replacement for the pooled cohort equations. It can be used in U.S. adults aged 30–70 years to predict 10- and 30-year CVD risk. It includes the prediction of heart failure risk, an important feature in the care of patients with hypertension. Over time, the PREVENT instrument is expected to replace the current pooled cohort equations risk prediction model.

CAUSES OF HYPERTENSION

PRIMARY HYPERTENSION

Most adults have "primary" hypertension, with no obvious underlying anatomic cause of their high BP. Adoption, twin, clinical, and family studies provide evidence for a heritable component of high BP and hypertension. However, much of this may be due to a shared environment because, with rare exceptions, genetic investigations have only identified modest polygenic associations between multiple genes and BP. While genetic research remains an important area for investigation, the clinical implications of genetic studies for diagnosis and management of high BP are currently very limited.

Strong associations with level of BP have been reported for several general environmental exposures, including heavy metals such as lead, mercury, cadmium, and arsenic. Several indicators of air pollution, commonly including levels of particulate matter with a diameter of 2.5 microns ($PM_{2.5}$) or less, have been associated with higher levels of SBP and DBP. In most studies, the increase in SBP has been ~3–5 mmHg, but the magnitude can vary depending on quantity and duration of the exposure. Air pollution is a challenge worldwide, but the level of exposure to pollutants varies widely by geography, climate, season, and extent of economic development. In addition to air pollution, clinically important changes in BP are recognized in geographic regions that experience large seasonal variations in ambient temperature, with lower and higher BPs during the warmest and coldest seasons, respectively. BP also tends to be higher in those who are acutely exposed to high altitudes, possibly due to a combination of hypoxia and cold temperatures.

Personal environmental exposures related to components of diet, physical activity, and alcohol consumption seem to be responsible for much of the age-related increases in BP that are common in most countries. A large body of ecological, migrant, and longitudinal cohort studies have documented higher BPs in those who consume unhealthy diets, have an excessive intake of sodium or an inadequate intake of potassium, are physically inactive, are overweight or obese, or consume alcohol. Many of these exposures have also been linked to CVD complications. For example, insufficient dietary intake of potassium has been identified as a risk factor for stroke in many studies. Stress and other psychosocial indicators have also been associated with higher BPs and CVD events. The six personal exposures identified in Table 288-2 (diet quality, body weight, excessive dietary sodium intake, insufficient dietary potassium intake, physical inactivity, and alcohol consumption) are all highly prevalent and strongly associated with higher BP and age-related increases in BP. About half of U.S. adults report attempts to eat a heart healthy diet, but this is likely to represent an overestimate. Almost all adults worldwide exceed the intake of dietary sodium and fail to meet the level of dietary potassium recommended by the WHO and national organizations. Likewise, excess body weight is very common in many countries. For example, approximately three-quarters of U.S. adults are either overweight (body mass index [BMI] 25–29 kg/m^2) or obese (BMI ≥30 kg/m^2). More than a quarter of U.S. adults report no physical activity outside their workplace, and about half report less than the recommended level of physical activity (≥150 min of aerobic exercise/week). About 85% of U.S. adults report consumption of alcohol at some point during their life, with >60% reporting alcohol consumption during the previous year and 25% reporting binge drinking during the previous month. Thus, these six exposures are not only closely related to level of BP and targets for interventions aimed at BP lowering but also extremely common exposures.

SECONDARY HYPERTENSION

A minority of patients have "secondary" hypertension, with an overt underlying anatomic or biochemical cause for their high BP. Secondary hypertension should be considered during the evaluation of all patients with new-onset hypertension and in selected patients with prevalent hypertension who have indicators that suggest the possibility of a complicating secondary cause of hypertension, including (1) treatment-resistant hypertension; (2) abrupt worsening of hypertension; (3) disproportionate target organ damage for level of BP; and (4) laboratory or diagnostic findings such as unprovoked hypokalemia, proteinuria, or left ventricular hypertrophy. The approximate prevalence, pathophysiology, possible clinical and physical examination findings, common screening and confirmatory tests, and potential treatments for the six most common causes of secondary hypertension are summarized in Table 288-3.

Obstructive Sleep Apnea Obstructive sleep apnea (OSA) is probably the most common cause of secondary hypertension. It is closely associated with overweight, obesity, and other comorbidities. The reported prevalence of OSA varies substantially depending on the criteria used for definition and methods of ascertainment. However, it is very common and may be present in as many as 15–30% of adult men and 10–15% of adult women in the United States. More than half of U.S. adults with OSA have hypertension and >30% of U.S. adults with hypertension have OSA. OSA and hypertension share many common pathophysiologic features, including being overweight, obesity, unhealthy lifestyles, and abnormalities of the renin-angiotensin system and fluid distribution. Sympathetic system activation due to intermittent hypoxia is also an important component of the underlying pathophysiology in OSA. Almost three-quarters of patients with OSA are overweight or obese. In general, there is a direct association between the severity of OSA and an individual's level of BP as well as their resistance to antihypertensive therapy. Resistant hypertension,

TABLE 288-2 Summary Information for the Six Best Proven Nonpharmacologic Interventions That Lower Blood Pressure

CHARACTERISTIC	BP ASSOCIATION	INTERVENTION	REGIMEN	EXPECTED CHANGE IN BP
Diet	BPs are lower in those who consume heart-healthy diets.	Consumption of a heart-healthy diet, especially the Dietary Approaches to Stop Hypertension (DASH) diet.	DASH diet meal plans are readily available.	With good adherence to the DASH diet plan, an average SBP reduction of about 5 mmHg can be expected in patients with and 2–3 mmHg in those without hypertension.
Body weight	Increased body weight is directly associated with SBP/DBP. Almost 75% of U.S. adults are overweight (BMI 25–<30 kg/m^2; 31%) or obese (BMI ≥30 kg/m^2; 42%).	In most patients, weight loss should be achieved by behavioral counseling. In a minority, drug therapy should be considered. Bariatric surgery should be confined to adults with a BMI ≥40 kg/m^2 or ≥35 kg/m^2 and hypertension or another obesity-related comorbidity.	Behavioral interventions are aimed at a combination of calorie reduction and increased physical activity.	Generally, SBP is reduced by ~1 mmHg for every kilogram reduction in body weight. In high-quality RCTs, an average reduction of about 10 lb (4.5 kg) has been common at 6–12 months.
Sodium intake	Dietary sodium is directly associated with SBP, and excessive intake may be important in age-related increases in BP.	Behavioral counseling, use of salt substitutes, public health messaging, and policy to reduce the amount of sodium added during food processing and commercial preparation.	The sodium-BP dose response is almost linear, so any reduction in sodium intake is beneficial. Optimal target recommendations vary from <1500 to 2300 mg sodium intake per day.	In high-quality behavior change trials, average sodium intake is reduced by ~25% and SBP by ~5 mmHg and 2–3 mmHg for adults with and without hypertension, respectively. Use of salt substitutes has resulted in prevention of stroke (14%), CVD (13%), and all-cause mortality (12%) in addition to BP lowering.
Potassium intake	Potassium intake is inversely associated with BP and a lower risk of CVD, especially stroke.	Dietary or pill supplementation, but the former is preferred. An additional intake of 2000 mg/d (~50 mmol/d).	The recommended intake of potassium in adults is ~3500 mg/d. A dose-response meta-analysis suggests supplementing usual intake by ~1200 mg/d may be optimal. Potassium supplementation is contraindicated in patients with hyperkalemia or advanced kidney disease or who are taking medicines that increase the risk of hyperkalemia.	Clinical trial meta-analyses have repeatedly shown that potassium supplementation by diet or pill use lowers BP. The average reduction in SBP has been ~5 and 3 mmHg in adults with and without hypertension, respectively. Greater reductions are observed in adult black patients and those consuming large amounts of dietary sodium.
Physical activity	Observational studies identify a strong inverse association between physical activity and BP, hypertension, and CVD.	Aerobic, dynamic resistance, and isometric resistance exercises.	Most trials have evaluated aerobic exercise interventions, and this type of exercise is commonly prescribed. Most guidelines recommend ≥150 min of aerobic exercise per week. However, any increase in physical activity and any type of exercise are likely to be beneficial.	In adults with hypertension, aerobic exercise is likely to reduce SBP by ~5 mmHg and, when combined with dynamic resistance exercise, by ~7 mmHg. The extent of BP lowering depends on the starting level of BP and success of the intervention.
Alcohol consumption	Alcohol intake is associated with BP and hypertension in a direct and roughly linear fashion, with no threshold for the association.	Most commonly, substitution of lower alcoholic or nonalcoholic beverages, but behavioral counseling and abstinence, without access to alcohol, have been used in some trials.	Most interventions have targeted a reduction in alcohol consumption, often to ≤2 and ≤1 standard drinks per day in men and women, respectively. However, any reduction in alcohol consumption is helpful.	The extent of BP lowering depends on the starting level of alcohol consumption and the magnitude of the reduction in alcohol intake. Commonly, alcohol reduction trials have resulted in an SBP reduction of >5 mmHg in adults with hypertension.

Abbreviations: BMI, body mass index; BP, blood pressure; CVD, cardiovascular disease; DASH, Dietary Approaches to Stop Hypertension; DBP, diastolic BP; RCT, randomized controlled trials; SBP, systolic BP.

snoring, poor quality sleep, breathing pauses during sleep, and daytime sleepiness are common clinical indicators of OSA. Several simple questionnaire-based tests are available to screen for those who should be evaluated more carefully with a sleep study (polysomnography). Lifestyle improvements, especially those resulting in weight loss, are an important component of treatment for both OSA and any associated hypertension. OSA-specific treatment includes continuous positive airway pressure (CPAP), but meta-analyses demonstrate limited BP lowering due to CPAP treatment. As a result, hypertension in patients with OSA should be treated with antihypertensive drug therapy in addition to nonpharmacologic therapy.

Medication and Other Substances Prescription and over-the-counter medications, herbal and food additives, and illegal substances are often identified as secondary causes of new-onset hypertension or diminished BP control in patients being treated for hypertension. Typically, the increase in BP results from a direct pressor effect of the agents, but these exposures can also have an indirect effect on BP as a result of drug-drug or drug-nondrug exposure interactions. Some of the most common prescription drugs that can increase BP include amphetamines, angiogenesis inhibitors, antidepressants and antipsychotic agents, decongestants, oral contraceptives, nonsteroidal anti-inflammatory drugs (NSAIDs), and systemic corticosteroids. Examples of illicit drugs that can result in a higher BP include cocaine, marijuana, amphetamines, and methylenedioxymethamphetamine. Herbal treatments that have been reported to raise BP include arnica, ephedra (Ma-Huang), ginseng, guarana, consumption of large quantities of licorice, and St. John's wort. Finally, caffeine and several anesthetic agents can result in short-term elevation of BP. A careful history will identify most of these agents and will allow for use of alternatives or a reduced dosage in most patients. The exact prevalence of exposure to any of these agents is uncertain but high. For example, in the nationally representative 1999–2004 National Health and Nutrition Examination Survey (NHANES), >26% of U.S. adults reported regular NSAID use, and the percentage was much higher in older adults and in non-Hispanic whites.

TABLE 288-3 Summary Description of the Most Common Secondary Causes of Hypertension

CAUSE	PREVALENCE	PATHOPHYSIOLOGY	CLINICAL INDICATORS	PHYSICAL EXAMINATION	SCREENING TESTS	CONFIRMATORY TESTS	TREATMENT
Obstructive sleep apnea (OSA)	OSA is very common (10–30% of U.S. adults). More than 50% of U.S. adults with OSA have hypertension, and ~30% with hypertension have OSA.	OSA and hypertension share many common pathophysiologic features, including overweight and obesity. Activation of the sympathetic system due to intermittent hypoxia is an important component of the pathophysiology for OSA.	Resistant hypertension, snoring, poor quality sleep, breathing pauses during sleep, daytime sleepiness, overweight and obesity	Obesity and signs of upper airway narrowing	OSA screening questionnaires (several have been validated)	Polysomnography (sleep study)	Lifestyle improvement, especially aimed at weight loss, is indicated for both OSA and associated hypertension. Antihypertensive drug therapy. Continuous positive airway pressure (CPAP) treatment for OSA has little effect on BP.
Medication and other substances	High prevalence. Numerous potential exposures. For NSAID alone, >26% of U.S. adults reporting regular use.	Varies depending on exposure.	Usually exposure best detected by medical history report	Occasionally, specific physical exam findings, e.g., signs suggesting cocaine use	History taking	History taking	Discontinuation, substitution of alternative therapy, or maintenance of exposure but management of high BP.
Renal parenchymal disease.	One of the most common causes of secondary hypertension (>5%).	Various kidney diseases, including polycystic renal disease, chronic glomerulonephritis, diabetic and hypertensive kidney disease, recurrent stones, and kidney infections.	Resistant hypertension, fatigue, edema, report of hematuria	Proteinuria, renal mass	Serum creatinine, albuminuria, glucose abnormality	Imaging by sonography, CT, or MRI Renal biopsy	Treatment of ASCVD risk factors, including hypertension. Management of the specific kidney disease, if feasible.
Primary aldosteronism	One of the more common causes of secondary hypertension, responsible for the high BP in ~5% of adults with hypertension.	Idiopathic bilateral adrenal hyperplasia (60–70%) Benign adrenal adenoma (30–40%) Occasionally, familial Rarely due to other causes	Severe or resistant hypertension Hypokalemia Nonspecific fatigue, muscle cramps, or weakness Incidental finding of adrenal adenoma Obstructive sleep apnea Family history of early-onset hypertension or stroke	No specific findings on clinical examination	High plasma aldosterone/renin ratio, when measured under standardized conditions CT of the abdomen	Saline suppression test or captopril challenge test are helpful, but adrenal vein sampling for aldosterone levels is generally accepted as the gold standard test.	Adrenalectomy (laparoscopic or surgical) for many patients with an adrenal adenoma. Hypertension due to adrenal hyperplasia best treated medically, with inclusion of a mineralocorticoid receptor antagonist, and lifestyle changes to reduce sodium intake and body weight.
Renovascular disease	Generally reported to approximate 5%.	Usually unilateral disease due to ASCVD (90%) with >70% arterial lumen narrowing. About 10% due to fibromuscular disease, typically occurring in middle-aged white women.	Early, abrupt-onset hypertension (consider fibromuscular disease) Rapidly worsening or resistant hypertension in older adults (consider ASCVD)	Bruits, especially abdominal renal artery bruit	Various imaging options, including Doppler ultrasonography, magnetic resonance, or CT angiography	Digital subtraction angiography	PTA may cure or substantially improve hypertension in patients with discrete fibromuscular disease lesions. Medical treatment is preferred in patients with ASCVD and those with very diffuse fibromuscular disease.
Hypertensive disorders of pregnancy	During pregnancy, approximate prevalence of hypertension 10%, preeclampsia 3%, and eclampsia 0.1–0.3%.	Varies depending on whether the hypertension is new onset or chronic. New-onset hypertension seems to result from reduced placental perfusion and subsequent elevation in peripheral resistance, which usually disappears after delivery.	High BP Preeclampsia if high BP is accompanied by proteinuria or other evidence of target organ damage	High BP, with or without proteinuria	No specific screening tests	Diagnosis by recognition of high BP, based on an average of ≥2 accurately measured BP readings taken on ≥2 occasions	Lifestyle counseling and antihypertensive medication. ACEIs and ARBs should not be used. Initial use of the β-blocker labetalol and/or the CCB nifedipine is commonly recommended because of their safety record and known efficacy in pregnancy.

Abbreviations: ACEI, angiotensin converting enzyme inhibitor; ARB, angiotensin receptor blocker; ASCVD, atherosclerotic cardiovascular disease; BP, blood pressure; CCB, calcium channel blocker; CT, computed tomography; MRI, magnetic resonance imaging; NSAID, nonsteroidal anti-inflammatory drug; PTA, percutaneous transluminal angioplasty.

Renal Parenchymal Disease This is also relatively common and probably accounts for >5% of all non-OSA and medication-related secondary causes of hypertension. It can result from almost any type of underlying kidney disease but is somewhat more common with glomerular than interstitial disorders. Most patients with CKD disease have hypertension. Importantly, it is often difficult to determine the extent to which a patient's high BP is the cause or a consequence of their kidney disease. Activation of the RAAS is often a part of the underlying pathophysiology that leads to hypertension. Clinical indicators of hypertension due to renal parenchymal disease include treatment resistance, fatigue, palpation of a renal mass, hematuria, proteinuria, elevated creatinine, or hyperglycemia. Imaging with sonography, computed tomography (CT), or magnetic resonance imaging (MRI) can be helpful for the diagnosis of some forms of underlying kidney disease. Renal biopsy is most helpful when glomerulonephritis is suspected. Most patients with CKD have a variety of both conventional and novel risk factors for ASCVD placing them at high risk for CVD complications. Improvement of CVD risk factors, including hypertension, is an essential element of treating patients with CKD. Often, control of hypertension may require use of three or more antihypertensive agents in addition to lifestyle management. Diuretics and RAAS inhibitors such as ACEIs and ARBs are obvious components of the antihypertensive drug treatment regimen. Traditionally, thiazide and thiazide-like diuretics were thought to be relatively ineffective for the management of patients with a serum creatinine >3 mg/dL, but a recent well-conducted trial documented substantial BP and albuminuria reduction in patients with advanced CKD who were treated with the long-acting diuretic chlorthalidone. Finerenone, a newer MRA, has also been shown to enhance BP reduction in patients with advanced CKD. Gliflozins or sodium-glucose cotransporter 2 (SGLT2 inhibitors), which are often very useful in the management of patients with CKD, are diuretics and may provide a small reduction in BP. If possible, the underlying kidney disease should be treated with specific therapy.

Primary Aldosteronism A relatively common cause of secondary hypertension, primary aldosteronism may be the underlying cause of high BP in ~5% of adults with non-OSA and medication-related secondary hypertension. However, the reported prevalence varies widely depending on the criteria for diagnosis and the population studied. Most patients with primary aldosteronism (60–70%) have idiopathic bilateral adrenal hyperplasia but ~30–40% have a benign adenoma, usually unilateral. Often, the clinical presentation is subtle with no more than resistant hypertension or nonspecific symptoms. However, in some patients with more pronounced aldosteronism, unprovoked hypokalemia or hypokalemia-related cardiac arrhythmias, such as atrial fibrillation, may be present. Common screening tests for primary aldosteronism include abdominal imaging by CT to detect an adenoma and measurement, under standardized conditions, of a high aldosterone/renin ratio in conjunction with high plasma aldosterone levels. Several noninvasive investigations, including the saline suppression and captopril challenge tests, can provide more definitive evidence of primary aldosteronism, but adrenal vein aldosterone sampling is generally accepted as the gold standard test for diagnosis. The most important goal is to recognize the minority of patients with a benign unilateral adenoma whose hypertension can be cured or substantially improved by adrenalectomy. The latter can be accomplished by minimally invasive laparoscopy or by use of traditional surgical techniques. Medical management of hypertension in patients with aldosteronism should include the use of MRA agents. In most patients, spironolactone is well tolerated but occasionally adverse effects, including hyperkalemia, breast tenderness, or gynecomastia in men, require the substitution of newer, more expensive, alternative MRAs. The screening tests and especially the more definitive diagnostic tests for primary aldosteronism require careful attention to detail and are generally best conducted by those who have specialized experience in the evaluation of patients who may have primary aldosteronism. Likewise, most adrenal adenomas are very small, and removal by laparoscopy or surgery is most successful when performed by an experienced treatment team. Rarely, primary aldosteronism can be caused by an adrenal carcinoma, ectopic malignancy, or rare genetic disorders.

Renovascular Hypertension Renovascular obstructions that result in substantial and hemodynamically important occlusion of the renal artery can result in hypertension and impaired renal function. Renovascular disease is found in as many as 25% of patients at autopsy, but renovascular hypertension only accounts for ~1% of the high BP in all patients with hypertension. In ~90% of patients with renovascular hypertension, renal artery obstruction is caused by atherosclerosis. Commonly, atherosclerotic renovascular disease is unilateral and the principal obstruction tends to present as a discrete lesion found close to the origin of the renal artery. Patients with atherosclerotic renovascular disease tend to be older and have other indications of ASCVD, including an abnormal risk factor profile and atherothrombotic lesions in other vascular beds. At times, a bruit can be heard over the renal artery or over the carotid or femoral arteries. A minority of patients with renovascular disease have fibromuscular disease. They tend to be younger, mostly white, women. Their renal artery lesions are more diffuse, with an irregular sawtooth appearance, and they are frequently located in more distal parts of the renal artery compared with atherosclerotic lesions. In addition, bilateral real artery disease is more common in patients with fibromuscular disease than is the case for renovascular disease due to ASCVD. Fibromuscular disease should be suspected in any patient with abrupt-onset hypertension, especially in young white women, and in those with resistant hypertension, or recent worsening of hypertension. Underlying atherosclerosis should be suspected in older patients with other indications of ASCVD and in those with otherwise unexplained worsening of kidney function. The threshold for additional diagnostic investigations should be lower when fibromuscular disease is suspected.

A variety of screening tests, including renal ultrasound studies, CT and magnetic resonance angiography imaging studies, renograms, and standardized measurement of plasma renin activity, can be used to exclude other causes of hypertension and to screen for evidence of renovascular disease. Renal angiography provides more definitive evidence of the extent and type of renovascular disease and is currently considered to be the gold standard diagnostic test. Skilled interventional radiologists can employ digital subtraction angiography to generate high-quality studies with less contrast exposure compared to standard angiography. Hypertension due to renovascular disease is more likely in patients with lesions that occlude >70% of the vessel lumen. For many years, renal venous renin sampling was a gold standard diagnostic test but is now rarely performed due to a high frequency of false-positive and false-negative results.

In many patients with hypertension due to fibromuscular disease, percutaneous transluminal angioplasty results in a cure for their hypertension or substantially improves the control of their high BP. For this reason, the most important goal in evaluating patients with suspected renovascular hypertension is to recognize individuals with underlying fibromuscular disease. As is the case for some other forms of secondary hypertension, screening, diagnosis, and treatment results tend to be better when conducted by teams with experience in caring for patients with renovascular disease. Several well-conducted randomized controlled trials, with relatively large sample sizes, have failed to demonstrate any special benefit for BP control or renal preservation following percutaneous transluminal angioplasty or renovascular surgery compared to medical care in patients with atherosclerotic renovascular hypertension. Most patients with suspected atherosclerotic renovascular hypertension should be managed using a combination of nonpharmacologic and antihypertensive drug therapy, including the use of agents that block the RAAS. In rare instances, arteritis and other inflammatory disease can be the underlying cause of renovascular hypertension. The underlying cause should be treated in addition to medical management of hypertension.

Hypertensive Disorders of Pregnancy Hypertension occurs in >10% of pregnancies in the United States and about 3–5% are complicated by preeclampsia, in which hypertension is accompanied by proteinuria or other evidence of target organ damage. The HELLP syndrome, in which hypertension is accompanied by hemolysis, elevated liver enzymes, and low platelets, represents a more severe form of preeclampsia. Progression to eclampsia, with seizures as the cardinal feature, occurs in <1% of women with mild and about 3% of those with severe preeclampsia. Hypertension during pregnancy is associated with a higher rate of fetal complications and of hypertension and CVD complications for the mother later in life. Hypertensive disorders of pregnancy include the following: (1) chronic hypertension, in which the high BP proceeds pregnancy; (2) gestational hypertension, in which new-onset hypertension is recognized after 20 weeks of gestation; (3) preeclampsia potentially culminating in eclampsia; and (4) chronic hypertension with superimposed preeclampsia. Management of hypertension in pregnancy often includes bed rest, nonpharmacological interventions, and pharmacotherapy. Clinical trials and meta-analyses provide strong evidence that antihypertensive therapy during pregnancy reduces the risk of progression to more severe hypertension, compared to placebo, but less convincing evidence for prevention of preeclampsia and other fetal complications. ACEI and ARB therapy should be avoided due to the potential for teratogenic side effects. Few high-quality studies have compared the benefits and risks of treatment with other commonly used antihypertensive drugs during pregnancy. Usually, treatment with the beta blocker labetalol, the CCB nifedipine, or the centrally acting α_2-adrenergic agonist methyldopa is recommended as providing safe first-line antihypertensive drug therapy, with one trial suggesting that labetalol and nifedipine may be superior to methyldopa for prevention of preeclampsia. A review of prescribing habits in the United States identified labetalol, nifedipine, and the direct vasodilator hydralazine as the three most commonly used antihypertensive agents during pregnancy. Clinical trials have failed to demonstrate a benefit for more intensive therapy to a SBP/DBP <130/80 mmHg compared to <140/90 mmHg.

Less Common Causes of Secondary Hypertension There are many other secondary causes of hypertension, but they tend to be infrequent and are generally best considered when patients are referred to consultants with expertise in recognizing rare causes of secondary hypertension. These include Cushing's syndrome in which large amounts of cortisol are produced in response to excessive levels of adrenocorticotropic hormone (ACTH) produced by a pituitary gland adenoma or ectopic ACTH-producing tumor, an adrenal adenoma, or treatment with glucocorticoids. Clinical signs of note include weight gain in the face (moon face) and trunk, a fatty lump between the shoulders (buffalo hump), thin arms and legs, pink or purple stretch marks on the stomach, hips, thighs, breasts, and underarms, easy bruising, slow healing, and acne. Screening and diagnosis are usually based on an assessment of 24-h urinary corticosteroid levels, imaging tests, and an overnight dexamethasone-suppression test. More than 75% of patients with Cushing's syndrome have hypertension. Treatment depends on the underlying cause of the syndrome. *Pheochromocytoma* is another rare form of secondary hypertension and is caused by excessive production of catecholamines, usually due to a benign neuroendocrine tumor composed of chromaffin cells located in the adrenal medulla (80–85%) or in extra-adrenal paraganglion tissue (paraganglioma). Most patients with pheochromocytoma present with hypertension at a relatively young age (often 30–50 years). Many have nonspecific signs and symptoms, but some have paroxysmal attacks characterized by high BP, sweating, headaches, and cardiac arrhythmias. Screening and diagnosis are usually based on blood and 24-h urinary catecholamine values and abdominal CT, MRI, or positron emission tomography imaging studies. Typically, pheochromocytomas are treated by complete or partial adrenalectomy using minimally invasive surgical techniques. A small minority of patients with pheochromocytoma have an inherited genetic predisposition such as multiple endocrine neoplasia type 2 (MEN 2), von Hippel-Landau disease, neurofibromatosis, and hereditary paraganglioma syndromes. Coarctation of the aorta is the most common congenital cause of hypertension and results from a birth defect in which a portion of the aorta is narrower than usual. When the defect is severe, the coarctation is likely to be detected early in life and may be associated with other congenital abnormalities. When the defect is mild, the disorder may not be recognized until early adulthood. The classical physical examination findings are delayed and diminished femoral and distal pulses. A systolic mummer may be heard in the posterior intrascapular area, and signs of left ventricular hypertrophy may be detected. Screening and diagnostic tests include imaging studies such as echocardiography, CT, and angiography. Treatment options include balloon angioplasty, surgery, and medical management. Other rare causes of hypertension include hyper- and hypothyroidism, acromegaly, and hypercalcemia.

PREVENTION AND TREATMENT OF PRIMARY HYPERTENSION

Both nonpharmacologic interventions, mostly lifestyle improvements, and antihypertensive medication play a role in the management of high BP.

The overall approach to prevention and treatment of hypertension, which is depicted in **Fig. 288-3**, is to encourage a healthy lifestyle in adults with a normal BP (SBP/DBP <120/80 mmHg) and actively advise nonpharmacologic therapy in adults with an elevated BP (SBP 120–129 mmHg and DBP <80 mmHg). Adults with stage 2 hypertension (SBP ≥140 mmHg or DBP ≥90 mmHg) should be treated with a combination of nonpharmacologic and antihypertensive drug therapy. Most adults with stage 1 hypertension (SBP 130–139 mmHg or DBP 80–89 mmHg) should be managed by active application of nonpharmacologic interventions, but in the minority (~30% in the United States) who have a history of CVD or are at high risk for ASCVD, antihypertensive medication should be added to the nondrug approaches to lower BP. In RCTs, low-dose pharmacotherapy has been effective for BP lowering, prevention of hypertension, regression of left ventricular mass, and prevention of left ventricular hypertrophy in nonhypertensive patients. However, drug therapy is not recommended for this purpose in any major clinical practice guideline. The next two sections provide a more detailed guide to the selection and application of nonpharmacologic and antihypertensive drug treatments.

■ NONPHARMACOLOGICAL INTERVENTIONS TO PREVENT AND TREAT HYPERTENSION

Many nondrug interventions have been reported to lower BP. Most of them are based on changing personal exposures associated with higher BP. The best proven approaches to nonpharmacologic lowering of BP are displayed in Table 288-2.

Generally, the nonpharmacologic interventions used in clinical practice settings require a change in behavior. However, pill supplementation to enhance potassium intake, renal denervation therapy in selected patients with resistant hypertension, and baroreceptor activation therapy are examples of nonpharmacologic treatments that are not primarily based on achieving behavior change. For the nonpharmacologic interventions based on behavior modification, the difficulty of achieving the desired changes varies by intervention and patient, but generally, behavior change can be hard to achieve and maintain over long periods of time. Behavior change interventions are most successful when patients accept the need for change, pledge to embrace the behavioral changes that are necessary, and can be counseled by a member of the health care team who is knowledgeable in the techniques for behavior change. For weight loss, abstinence from alcohol, physical activity, and high-quality counseling programs that are available in convenient locations close to a patient's home or workplace may provide an alternative option to office/clinic-based counseling. Behavior change can be successful in any patient but is especially likely in those with or at high risk for CVD and in patients who have the time, resources, and desire to concentrate on improving their health behaviors. Not surprisingly, older adults and those with higher socioeconomic status (SES) tend to be more successful than their younger and lower SES counterparts. The following are brief summaries of specific aspects of the six interventions identified in Table 288-2.

Diet Quality Heart healthy diets tend to lower BP and improve other diet-related CVD risk factors, such as low-density lipoprotein cholesterol and blood glucose. For example, a reduction in BP has

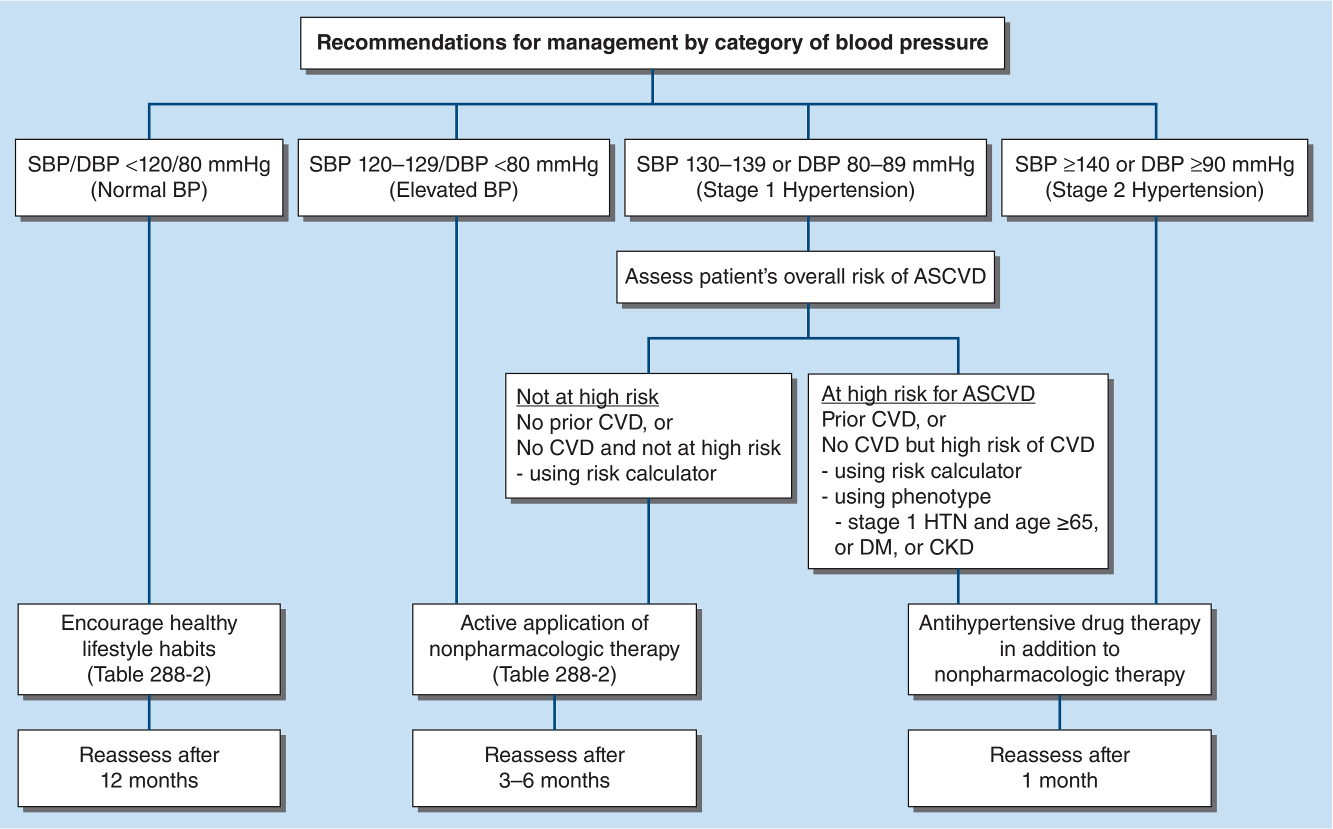
Recommendations for management by category of blood pressure
SBP/DBP <120/80 mmHg (Normal BP)
SBP 120–129/DBP <80 mmHg (Elevated BP)
SBP 130–139 or DBP 80–89 mmHg (Stage 1 Hypertension)
SBP ≥140 or DBP ≥90 mmHg (Stage 2 Hypertension)
Assess patient's overall risk of ASCVD
Not at high risk
No prior CVD, or
No CVD and not at high risk
- using risk calculator
At high risk for ASCVD
Prior CVD, or
No CVD but high risk of CVD
- using risk calculator
- using phenotype
- stage 1 HTN and age ≥65, or DM, or CKD
Encourage healthy lifestyle habits (Table 288-2)
Active application of nonpharmacologic therapy (Table 288-2)
Antihypertensive drug therapy in addition to nonpharmacologic therapy
Reassess after 12 months
Reassess after 3–6 months
Reassess after 1 month

FIGURE 288-3 Recommendations for nonpharmacologic and antihypertensive drug therapy by category of blood pressure. ASCVD, atherosclerotic CVD; BP, blood pressure; CKD, chronic kidney disease; CVD, cardiovascular disease; DBP, diastolic BP; DM, diabetes mellitus; HTN, hypertension; SBP, systolic BP.

been a consistent finding in trials that have evaluated the Mediterranean diet, low-carbohydrate diets, and vegetarian and vegan diets. The DASH (Dietary Approaches to Stop Hypertension) diet is a heart-healthy diet that was specifically formulated to lower BP. The DASH diet is high in vegetables, fruits, whole grains, and low-fat dairy products and low in fats, saturated fats, and cholesterol. It has been extensively evaluated, and many DASH resources and support systems are readily available to health care providers and the public. For these reasons, it is commonly recommended as the best heart-healthy diet choice for BP reduction. Many of the DASH evaluation trials have been conducted as feeding studies in which patients are given specially prepared foods as a replacement for their usual diet. Compared to trials based on behavior change, feeding studies require less effort by the patient, achieve greater intervention change, and generally report a greater reduction in BP. However, the results of feeding trials are less generalizable to clinical practice where provision of patient meals is infeasible. The DASH trials that have employed a behavior change strategy provide a better indication of what can be expected in clinical practice settings. In the original short-term (8-week) feeding trial, the DASH diet reduced SBP/DBP by 5.5/3.0 mmHg compared with the usual U.S. diet. In longer-term (6-month) behavior change trials, the DASH diet combined with other established nonpharmacologic recommendations for BP lowering reduced SBP and DBP by ~4 mmHg and 3 mmHg, respectively, compared to an advice-only intervention, and by 0.6 mmHg and 0.9 mmHg, respectively, compared to other established recommendations on their own. With good adherence, one could reasonably expect use of the DASH diet on its own to result in an average SBP reduction of ~5 mmHg in patients with and 2–3 mmHg in patients without hypertension. Community-based trials have demonstrated that the DASH diet components can be purchased at supermarket locations without the need for patients to pay for higher priced specialty food products. The DASH diet can be combined with other nonpharmacologic interventions such as weight loss, sodium reduction, or increased physical activity. Likewise, the DASH diet eating plan can be customized to meet an individual's other dietary requirements, meal preferences, and lifestyle.

Weight Loss Weight loss not only lowers BP but can also improve lipid levels and is effective for prevention and management of diabetes mellitus. It can be achieved through a combination of calorie reduction and increased physical activity. The most sustainable approach to weight loss is targeting gradual and progressive loss in weight. As with all behavior change strategies, weight loss can be difficult to achieve and even more difficult to maintain over long periods of follow-up. However, it can be successful, especially in motivated patients who receive high-quality counseling and clinician reinforcement. In addition to clinic-based interventions, many commercial and noncommercial providers offer weight loss programs. The commercial programs usually provide some or all of the recommended meals. The optimal goal is achievement of an ideal body weight, but any weight reduction is beneficial. Application of high-quality weight loss counseling usually results in 6- to12-month weight loss of ~10 lb (4.5 kg). Typically, a 1-mmHg reduction in SBP can be expected for about every kilogram of weight loss. This translates to a 6- to 12-month average reduction in SBP of ~5 mmHg. The average reduction varies depending on the success of the intervention and the starting level of BP, with SBP reductions in those with and without hypertension tending to be about 5 and 2–3 mmHg, respectively. Several medications can be used to assist in weight loss, but some of the most popular agents, like sodium-glucose cotransporter 2 (SGLT2) inhibitors, are off-label options and are best reserved for U.S. Food and Drug Administration (FDA)-approved indications. Injectable glucagon-like peptide 1 (GLP-1) receptor agonists are equivalent agents that have been approved for weight loss by the FDA and have been shown to reduce weight in clinical trials. Certain patients with hypertension and a body mass index ≥35 kg/m^2 may be candidates for bariatric surgery. Selection of candidates for bariatric surgery is best conducted by multidisciplinary teams with experience in managing morbid obesity.

Dietary Sodium Reduction Many RCTs have demonstrated reductions in BP following decreases in dietary sodium intake. In a large dose-response meta-analysis of RCTs, there was a striking almost linear relationship between reduction in dietary sodium intake and the corresponding decreases in BP. This finding was quite robust, being noted for every subgroup studied. An SBP/DBP difference of ~15/10 mmHg was noted between those with the highest (~7000 mg/d) and lowest (<1000 mg/d) intakes of sodium. An average reduction of ~5.5 mmHg was noted for every 2300-mg (100-mmol) decrement in dietary sodium intake. As expected, the effect was greater in adults with compared to those without hypertension and in SBP compared with DBP. The results were consistent with a "lower is better than higher" conclusion for intake of dietary sodium. In healthy U.S. adults, consumption of a diet with 1500 mg sodium provides adequate opportunity to meet the other recommended dietary requirements. Recommendations for daily sodium intake in adults vary considerably. The U.S. federal *Dietary Guidelines for Americans* recommends a daily sodium intake <2300 mg, whereas the World Health Organization recommends <2000 mg and others recommend much lower intakes. More than 95% of U.S. adults exceed the 2000 mg/d threshold, and almost 100% exceed the 1500 mg/d threshold. However, any reduction in dietary sodium intake is likely to be beneficial.

The traditional approach to reducing dietary sodium intake in clinical settings is by means of behavior change interventions. In addition to helping patients estimate and track their sodium intake, trying to avoid high-sodium foods and maximizing home preparations using fresh, natural foods are desirable. In countries like the United States, where most of the population consume commercial food products, >80% of an individual's sodium intake is generally due to sodium added during food processing or commercial preparation of foods and only ~10% is due to naturally occurring sources of sodium. Careful scrutiny of food labels for packaged foods while shopping, reducing meal portion size, choosing condiments and seasonings with low sodium content, salt substitution using herbs and spices, and use of salt substitutes that replace ~25% of the sodium content with potassium are other well-proven strategies to reduce dietary sodium intake. Salt substitutes provide a relatively easy and well-accepted way to achieve a modest reduction in sodium intake and an increase in potassium intake, especially in those who primarily eat foods prepared in the home and/or add a lot of salt at the table. A large cluster-designed clinical trial conducted in Chinese adults with stroke or at high risk for stroke demonstrated an impressive and statistically significant 14% reduction in stroke (the primary outcome), other major CVD events, and all-cause mortality in adults randomized to use a salt with 25% substitution of potassium for sodium compared to those who used regular salt. There was no evidence of a difference in hyperkalemic events in the two treatment groups. Public health messaging based on simple educational and action recommendations is also useful. Policy aimed at reducing the amount of sodium added to foods has great potential but has been difficult to implement in the United States. Long-term follow-up for 10–15 years in participants who had been randomized to a behavioral change sodium reduction intervention or usual care has also suggested a significant reduction of 30% in new-onset CVD events.

Potassium Supplementation A higher intake of potassium is associated with a lower BP and a reduced risk of CVD, especially stroke. U.S. Dietary Reference Intake committees have not made a dietary allowance intake recommendation due to insufficient evidence but have identified 2600 and 3400 mg/d in women and men, respectively, as adequate intake levels. Individual RCTs and meta-analyses provide strong evidence that potassium supplementation reduces BP. This has been achieved with pill therapy and dietary supplementation, with the latter being the preferred approach in clinical practice because this usually results in consumption of a more heart-healthy diet and a lower intake of dietary sodium. In most RCTs, routine potassium intake has been supplemented by about 2300 mg (60 mmol) per day, and the average reduction in SBP has been ~5 mmHg in trials with a successful intervention (net change in urinary potassium ≥20 mmol [870 mg] per day). As expected, the SBP reduction is greater in adults with (~5 mmHg) than without hypertension (~2 mmHg) trials. The SBP lowering is almost three times greater in black compared with white adults. Finally, there is a strong linear association between BP reduction with potassium supplementation and consumption of dietary sodium, possibly reflecting the well-known natriuretic effects of potassium. A dose-response meta-analysis has suggested a U-shaped BP response to potassium intake, with an optimal supplemental intake of ~1200 mg/d. However, the unexpected BP response findings at the upper and lower end of the U-shaped curve were based on indirect estimates and a relatively small number of trials.

Many foods contain potassium, but fruits, vegetables, legumes, and leafy greens are especially high in potassium content. By design, the DASH diet targets provision of 4700 mg of potassium per day, exceeding the recommended intake of 3500 mg/d in U.S. adults. Potassium supplementation is contraindicated in patients with hyperkalemia or advanced kidney disease and in those taking pills that can block the urinary excretion of potassium, including potassium-sparing diuretics, inhibitors of the RAAS, and MRA agents, such as aldosterone, eplerenone, and finerenone.

Physical Activity Observational studies identify a strong inverse association between physical activity (leisure time or total) and BP as well as CVD. In like manner, high-quality RCTs have demonstrated that increased physical activity lowers BP in adults with and without hypertension. Most of the trials have been based on aerobic exercise interventions such as brisk walking, swimming, or dancing. Meta-analyses of aerobic exercise RCTs have been consistent in documenting substantial BP-lowering efficacy, but the quantitative estimates for average reduction in BP have varied considerably, with reports of 5–10 mmHg for SBP in studies restricted to adults with hypertension and 3–5 mmHg for studies in those without hypertension. Dynamic resistance exercises such as climbing stairs, push-ups, weightlifting, and squats have also been shown to lower BP and are often employed concurrently with aerobic exercise. Isometric resistance exercises, during which muscles or muscle groups are tightened without any visible movement of the surrounding joints, include yoga poses, wall sit, plank, and bench press exercises. Fewer high-quality RCTs have assessed the efficacy of isometric resistance exercise, but individual trials and meta-analyses indicate this approach is effective for BP lowering. In the largest meta-analysis (~400 trials) of physical activity, aerobic exercise reduced SBP by ~5 mmHg in adults with hypertension and when combined with dynamic resistance exercise by ~6.5 mmHg. In trials that allowed for a direct comparison with antihypertensive drug therapy, physical activity yielded an almost identical level of BP lowering. Most guidelines recommend moderate-intensity aerobic exercise as the principal approach to lowering BP but acknowledge that dynamic and isometric resistance exercises are a useful complement. Based on clinical trial evidence, an aerobic exercise duration of 40–60 min three or more times per week for a total ≥150 min per week may be optimal. Although more intensive physical activity may be ideal for overall cardiac fitness, less intensive physical activity is sufficient for BP reduction. Less frequent but more intensive physical activity (e.g., "weekend warrior" exercise) has been shown to lower BP.

Alcohol Consumption Alcohol has long been known to be a pressor. Observational dose-response meta-analyses demonstrate a roughly linear association between alcohol consumption and both BP and hypertension, with no threshold for the association. For each 14-mg increased intake of alcohol (amount in a U.S. standard alcoholic drink), SBP is ~1.5 mmHg higher. RCTs have studied the efficacy of reducing alcohol intake on BP by substitution of lower-alcohol or no-alcohol drinks, use of behavior change interventions, or abstinence without access to alcohol. The overall effect of alcohol reduction on SBP in the largest meta-analysis was a reduction of ~5.5 mmHg. However, the magnitude of the effect is greatly influenced by baseline level of BP and alcohol intake and by the success of the intervention. Most current U.S. guidelines recommend an alcohol intake goal of ≤2 and ≤1 standard drinks per day for men and women, respectively. In part, this reflects the possibility that a modest intake of alcohol may favorably influence other CVD risk factors, such as high-density lipoprotein cholesterol. The WHO recommends abstinence from alcohol.

Summary for Nonpharmacologic Interventions to Lower BP In summary, a strong body of evidence supports the value of interventions based on the six exposures highlighted in Table 288-2. In general, the interventions exhibit a linear-type dose-response for BP reduction. Consequently, more successful interventions result in greater BP reduction, and even smaller than desired changes in the exposure are likely to be beneficial. In addition, greater BP lowering can be expected in adults with compared to adults without hypertension, in those who start with more abnormality in the exposure, and when two or more interventions are combined. For example, a greater reduction in BP can be expected when the intervention combines the DASH diet with weight loss or sodium reduction. However, the behavioral changes needed for combination interventions are typically more challenging and require a greater commitment by the therapist and patient. Finally, nonpharmacologic interventions often enhance the effect of antihypertensive medications. This has been especially well demonstrated for reductions in sodium intake. A practical approach is to focus initially on the behavior that is most abnormal and most amenable to change, with a longer-term goal of more comprehensive improvements. With the exception of salt substitution, change in BP has been the primary outcome in most of the nonpharmacologic treatment trials, but some have demonstrated improvements in intermediate CVD outcomes, especially left ventricular mass and left ventricular hypertrophy. In addition, randomized comparisons following long periods of posttrial follow-up have suggested CVD benefits for both sodium reduction and weight loss.

PHARMACOLOGIC THERAPY TO PREVENT CARDIOVASCULAR DISEASE

Efficacy of Pharmacologic Therapy in Uncomplicated Primary Hypertension In most adults, primary hypertension is uncomplicated and relatively easy to manage. The introduction of diuretics in the late 1950s and early 1960s revolutionized pharmacotherapy of high BP, and these drugs were the principal agents in the early antihypertensive drug treatment trials. The first antihypertensive drug RCT reported a substantial benefit of active therapy compared to placebo in 1966, but it was quickly superseded by two larger placebo-controlled multicenter Veterans Administration Cooperative Group RCT analyses. The first analysis was for adults with an average DBP between 115 and 129 mmHg, and this component was stopped early due to benefit and published in 1967. The second analysis, for adults with an average DBP between 90 and 114 mmHg, was published in 1970. Both analyses demonstrated that antihypertensive drug therapy resulted in a dramatic CVD benefit compared to placebo. Subsequent trials confirmed these findings and documented benefits when SBP reductions were targeted, either in combination with an elevated DBP or as an isolated elevation of SBP in older adults. In a group meta-analysis of 123 trials, active treatment resulted in a reduction of major CVD (20%), CHD (17%), stroke (27%), HF (28%), and all-cause mortality (13%). In this and other meta-analyses, the relative reduction in risk has been almost identical for those with or without prior CVD, albeit only adults at high risk for CVD could participate in the included trials, and benefit has accrued to those with a baseline SBP from 130 to 139 mmHg and above.

Choice of Antihypertensive Drug Classes Characteristics of the 11 most commonly used classes of antihypertensive medication, including examples, usual daily dosage, frequency of administration, proposed mechanism of action, approximate BP lowering compared to placebo, most common side effects, frequency of adverse events, and special aspects of the drug class are shown in Table 288-4.

Initial therapy with agents from five classes of drug therapy (diuretics, beta blockers, CCB, ACEI, and ARB) has been shown to prevent CVD compared to placebo. However, in head-to-head RCTs, beta blockers have been inferior to agents from the other four antihypertensive drug classes, especially for prevention of stroke. CCBs are good for prevention of stroke but inferior for prevention of HF, especially compared to diuretics. Generally, meta-analyses have identified diuretics as being the "best in class" for first-step prevention of CVD. Recognizing these differences in antihypertensive drug classes, RCTs and meta-analyses also provide evidence that BP lowering, however achieved, is more important than the drugs used to achieve it. Most clinical practice guidelines, including the U.S. ACC/AHA BP guideline, recommend an approach similar to that outlined in Fig. 288-4 in which diuretics, CCBs, and ACEIs or ARBs, alone or in combination, are used for initial drug therapy in patients without a compelling indication for use of a beta blocker.

One or more nonpharmacologic therapies should be used to manage adults with stage 1 hypertension who are not at high risk for CVD (no CVD and a calculated 10-year risk of ASCVD <10% based on use of the ACC/AHA pooled cohort equations calculator). If BP cannot be controlled to an SBP/DBP <130/80 mmHg after 6 months of nonpharmacologic therapy, it may be reasonable to add an antihypertensive medication, especially in younger adults with a high lifetime risk of ASCVD. The latter can be estimated using a calculator that is readily available on the ACC website and elsewhere. Antihypertensive drug therapy should be used in addition to nonpharmacologic therapy in adults with stage 1 hypertension who are at high risk for ASCVD and for all individuals with stage 2 hypertension. A minority of such persons with an average SBP/DBP close to 130/80 mmHg can be managed successfully with single antihypertensive agent combined with nonpharmacologic therapy. However, most patients with hypertension require treatment with more than one antihypertensive agent. This is especially the case for adults with stage 2 hypertension with an SBP ≥140 mmHg and all non-Hispanic black adults with hypertension.

In clinical practice settings, it is common for patients with hypertension to have other comorbid conditions, including diabetes mellitus, CHD, and CKD, requiring use of agents from other drug classes that concurrently lower BP. Longer-acting versions of the recommended four first-step drug classes should be used to ensure the adequacy of once-daily treatment. For example, longer-acting diuretics such as chlorthalidone or indapamide are preferred over shorter acting agents such as hydrochlorothiazide because the long half-life of these agents is better suited for the provision of nighttime as well as daytime control of BP and because chlorthalidone has been used in almost all of the landmark U.S. trials of antihypertensive drug treatment. The most common side effects with thiazide and thiazide-like diuretics are biochemical changes, such as hypokalemia and slight increases in blood glucose. Potassium-sparing agents, such as amiloride or triamterene, can be combined with diuretics to counter these effects. The biochemical changes with diuretics are generally mild and do not counteract their well-proven CVD benefits. Although there are differences between individual ACEI agents, their clinical effects are more similar than different. Use of longer-acting agents such as lisinopril is preferred. The most common problematic side effect with ACEIs is a dry cough. This side effect is usually resolved quickly by switching to a corresponding dosage of an ARB. Side effects with ARBs are rare. ACEIs and ARBs should not be used in combination because this unnecessarily increases the risk of hyperkalemia, especially in vulnerable patients and provides little if any additional benefit over using either drug class on its own. Neither an ACEI nor an ARB should be used in pregnancy or planned pregnancy due to the risk of teratogenic side effects in the fetus. In contrast to the other antihypertensive drug classes, there are substantial clinical differences between dihydropyridine and nondihydropyridine CCB agents, with the former being primarily vasoactive and the latter having greater effect on the heart. Typically, dihydropyridine CCBs should be used in the treatment of hypertension. Peripheral edema occurs in ~10% of patients taking dihydropyridine CCBs but is often mild and can be minimized by dose reduction or eliminated by switching to an agent from another class. Nondihydropyridine CCBs should rarely be employed to treat hypertension unless there is another compelling indication for their use. Particular caution is necessary in patients at risk for bradycardia, heart block, or HF.

The earliest beta blocker drugs were relatively nonselective blockers of β-adrenergic receptors, including the heart (β_1 receptors) and lungs (β_2 receptors). They are rarely prescribed for hypertension due to the potential for bronchospasm. The second class of beta blocker agents, including metoprolol, contain more cardioselective and longer-acting drugs. Although cardioselective, atenolol is not recommended because several meta-analyses have reported that it has been less

TABLE 288-4 Summary Information for the Major Antihypertensive Drug Classes Used to Manage High Blood Pressure

CLASS	EXAMPLES	DOSE RANGE, mg	DOSE FREQUENCY/D	METHOD OF ACTION	SBP REDUCTION VERSUS PLACEBO, mmHg (APPROXIMATE)	MOST COMMON ADVERSE EFFECTS	FREQUENCY OF ADVERSE EVENTS	COMMENTS
First-Step Drugs								
Diuretics (thiazide and thiazide-like).	Chlorthalidone Hydrochlorothiazide Indapamide	12.5–25 12.5–50 1.25–2.5	1 1 1	Block sodium reabsorption in distal convoluted tubule	12	Hyponatremia Hypokalemia Hypercalcemia Hyperuricemia Hyperglycemia Dyslipidemia	Similar to placebo	Chlorthalidone preferred due to long half-life and use in most U.S. trials
ACE inhibitors	Enalapril Lisinopril Benazepril Ramipril Trandolapril	5–40 10–40 10–40 2.5–20 1–4	1 or 2 1 1 or 2 1 1	Inhibit ACE activity and Ang II production	12.0	Cough Hyperkalemia in CKD AKF in severe bilateral RAS Hypotension Angioedema	Similar to placebo	Do not use in combination with ARB; contraindicated in pregnancy
ARB	Losartan Valsartan Azilsartan Candesartan Olmesartan	50–100 80–320 40–80 8–32 20–40	1 or 2 1 1 1 1	Block Ang II binding to Ang receptors	12	Hyperkalemia in CKD Reduced GFR AKF in severe bilateral RAS Hypotension	Similar to placebo	Do not use in combination with ACE inhibitor; contraindicated in pregnancy
CCB (DHP)	Amlodipine Felodipine Nifedipine LA	2.5–10 2.5–10 30–90	1 1 1	Blocks calcium from entering cells, primarily inhibiting vasoconstriction	10	Peripheral edema (dose-dependent) Gingival hyperplasia	Similar to placebo	Amlodipine may be preferred CCB if tolerated Not recommended in HFrEF
CCB (non-DHP)	Diltiazem ER Verapamil ER	120–360 100–300	1 1 (evening administration recommended)	Blocks calcium from entering cells, reducing heart rate and vasoconstriction	9.0	Bradycardia Nausea Constipation.	Similar to placebo	Do not use in combination with β-blockers Do not use in HFrEF or in high-grade AV or SA block
Other Drugs								
β-Blocker	Metoprolol Carvedilol Nebivolol	50–200 20–80 5–40	1 1 1	Block β-adrenergic receptors	10	Asthma Bradycardia Fatigue Exercise intolerance Impaired concentration	Similar to placebo	First-step drugs when clinically indicated, e.g., in IHD and HF Contraindicated in high-grade heart block
MRA	Spironolactone Eplerenone Finerenone	25–100 50–100 10–20	1 2 1	Block distal tubule mineralocorticoid receptor activity	9	Hyperkalemia, especially in CKD and with potassium-sparing agents Spironolactone may cause tender breasts, gynecomastia, and erectile dysfunction in men	9–20%	Especially useful in resistant hypertension and low-renin states
K^+-sparing diuretics	Amiloride Triamterene	5–10 50–100	1 or 2 1 or 2	Block distal tubule sodium reabsorption	NA	Hyperkalemia, especially in CKD or with other agents that favor potassium retention	Similar to placebo	Minimal effect on BP Used to counteract hypokalemic effect of diuretics

(*Continued*)

TABLE 288-4 Summary Information for the Major Antihypertensive Drug Classes Used to Manage High Blood Pressure (Continued)

CLASS	EXAMPLES	DOSE RANGE, mg	DOSE FREQUENCY/D	METHOD OF ACTION	SBP REDUCTION VERSUS PLACEBO, mmHg (APPROXIMATE)	MOST COMMON ADVERSE EFFECTS	FREQUENCY OF ADVERSE EVENTS	COMMENTS
Loop diuretics	Furosemide Torsemide	20–80 5–10	2 1	Inhibits reabsorption of sodium in loop of Henle	NA	Hypokalemia Volume depletion Hyperuricemia	Infrequent	Preferred in CKD with GFR <30, in symptomatic HF, and when using potent direct vasodilator minoxidil
α_1-Receptor blockers	Doxazosin	1–16	1	Inhibit α_1-adrenergic receptors	5	Orthostatic hypotension, especially in older adults	9%	Less cardioprotective than diuretics Mostly used in men with prostrate hypertrophy
Direct vasodilator	Hydralazine Minoxidil	100–200 5–100	2 or 3 1–3	Dilate peripheral arterioles	NA	Reflex tachycardia Fluid retention Hydralazine-induced lupus syndrome (rare) Minoxidil-induced hirsutism in women	Up to 80%	Potent antihypertensive agents Usually reserved for fourth- or fifth-step treatment in resistant hypertension Usually combined with diuretic (loop agents for minoxidil) and β-blocker to minimize side effects
Central α_2-agonist and other centrally acting drugs	Clonidine (oral) Clonidine patch Guanfacine Methyldopa	0.1–0.8 0.1–0.3 0.5–2.0 250–1000	2 1/wk 1 2–3	Stimulate central nervous system α_2-adrenergic receptors	NA	Sedation Dry mouth Bradycardia Fatigue Constipation Orthostatic hypotension.	6–20%	Infrequently used due to side effects and potential for hypertensive crisis following abrupt withdrawal Methyldopa has a good safety record in pregnancy

Note: Adverse event estimates based on U.S. Food and Drug Administration labeling at *http://dailymed.nlm.nih.gov/dailymed/index.cfm.* Estimates of BP lowering versus placebo are "artificially" higher with older agents than with newer agents because the former were compared to placebo alone, whereas the latter were evaluated in patients who were already being treated with other antihypertensive medications.

Abbreviations: ACE, angiotensin-converting enzyme; AKF, acute kidney failure; Ang II, angiotensin II; ARB, angiotensin receptor blocker; AV, atrioventricular; BP, blood pressure; CCB, calcium channel blocker; CKD, chronic kidney disease; DHP, dihydropyridine; ER, extended release; GFR, glomerular filtration; HF, heart failure; HFrEF, heart failure with reduced ejection fraction; IHD, ischemic heart disease; LA, long acting; MRA, mineralocorticoid receptor antagonist; NA, not applicable; RAS, renal artery stenosis; SA, sinoatrial.

cardioprotective than other drug classes, including diuretics, when used for treatment of hypertension. The third-generation beta blocker drugs are cardioselective and have additional vasodilatory properties (nebivolol), used alone or in combination with α-receptor blockade (carvedilol). No RCT has documented improved CVD event protection with use of the newer agents. The remaining agents in Table 288-4 are covered in other sections of this chapter.

It has become increasingly clear that starting with lower doses of a two-drug combination provides superior BP control and results in fewer side effects compared to use of a single drug at a high dose. Initial therapy is especially important because surveys of clinical practice have repeatedly demonstrated that the initial drug choice is often materially unaltered due to therapeutic inertia despite unsatisfactory BP control. RCTs using stepped-care drug therapy, in which antihypertensive medications are added sequentially if BP control has not been achieved at full doses of the proceeding drugs, demonstrate that this approach can be successful if well applied. However, initial combination drug therapy has proven to be more effective for rapid achievement of target BP and medication adherence compared to the stepped-care approach.

Antihypertensive drug combinations should be based on use of drugs with complementary physiologic actions. For example, a diuretic or CCB combined with an ACEI or ARB is suitable for dual therapy and a diuretic combined with a CCB and an ACEI or ARB for triple-therapy combinations. Use of single-pill combinations results in better adherence compared to treatment with multiple drugs. In RCTs, the average number of drugs to achieve an SBP/DBP <140/90 mmHg has been two and to achieve an SBP/DBP <130/80 mmHg has been three.

Antihypertensive Treatment Blood Pressure Target Many clinical trials have reported better prevention of CVD in groups randomized to lower compared with higher achieved treatment BPs. Meta-analyses of RCTs to compare different achieved BPs have typically pooled the experience in trials that compared active versus placebo treatments, more versus less intensive treatments, and groups randomized to different BP targets. Most of these meta-analyses have reported greater benefit for a lower versus higher achieved BP, with the benefit being greatest for those who start at a higher BP and achieve a greater reduction in BP that their counterparts who achieve the

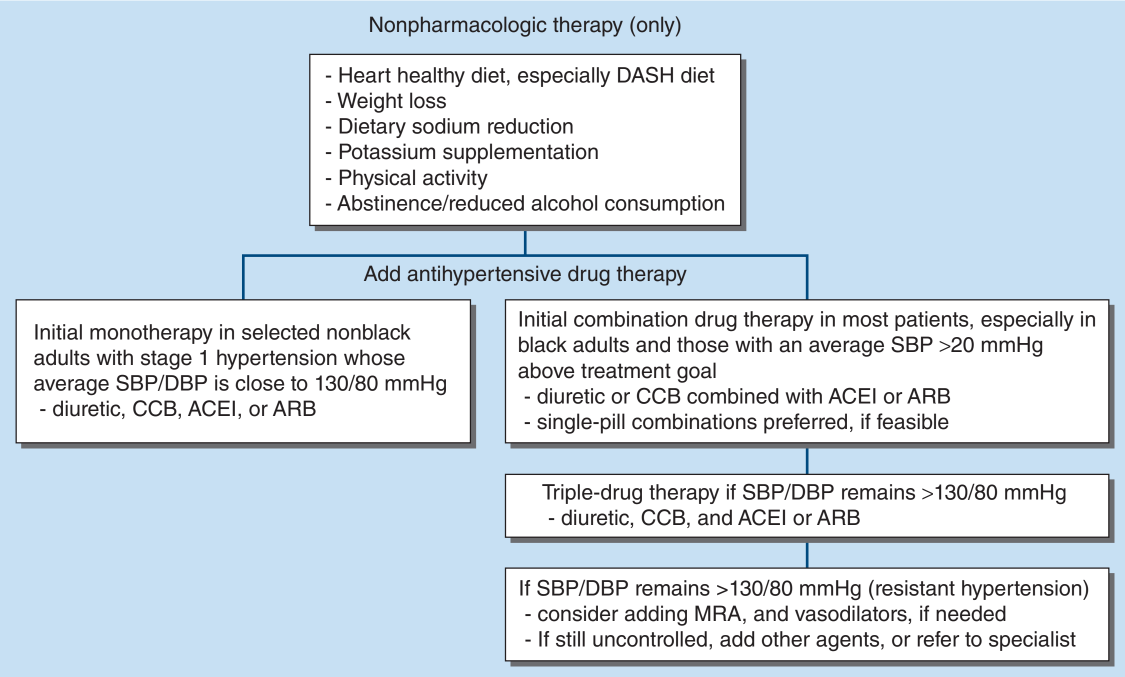
Nonpharmacologic therapy (only)
- Heart healthy diet, especially DASH diet
- Weight loss
- Dietary sodium reduction
- Potassium supplementation
- Physical activity
- Abstinence/reduced alcohol consumption
Add antihypertensive drug therapy
Initial monotherapy in selected nonblack adults with stage 1 hypertension whose average SBP/DBP is close to 130/80 mmHg
- diuretic, CCB, ACEI, or ARB
Initial combination drug therapy in most patients, especially in black adults and those with an average SBP >20 mmHg above treatment goal
- diuretic or CCB combined with ACEI or ARB
- single-pill combinations preferred, if feasible
Triple-drug therapy if SBP/DBP remains >130/80 mmHg
- diuretic, CCB, and ACEI or ARB
If SBP/DBP remains >130/80 mmHg (resistant hypertension)
- consider adding MRA, and vasodilators, if needed
- If still uncontrolled, add other agents, or refer to specialist

FIGURE 288-4 Overview of specific antihypertensive treatments in adults with uncomplicated hypertension. ACEI, angiotensin converting enzyme inhibitor; ARB, angiotensin receptor blocker; BP, blood pressure; CCB, calcium channel blocker; DASH, Dietary Approaches to Stop Hypertension; DBP, diastolic BP; MRA, mineralocorticoid receptor antagonists; SBP, systolic BP.

same BP but start from a lower level of BP. As one would expect from the observational association between BP and CVD, the incremental benefit for CVD protection diminishes as the category of achieved BP gets progressively lower. Almost all of the meta-analyses, including one conducted for the ACC/AHA BP clinical practice guideline committee, have reported that an achieved SBP of <130 mmHg is superior to higher categories of achieved SBP. The major English-language clinical practice guidelines recommend SBP treatment targets that vary from 120 to 135 mmHg in adults with hypertension and a high risk of ASCVD. Surveys indicate that almost all patients with hypertension who are seen in clinical practice are at high risk for ASCVD. At the current time, an SBP/DBP treatment goal of 130/80 mmHg in adults with hypertension, as recommended in the ACC/AHA BP guideline, seems reasonable. The only exception is in noninstitutionalized community-dwelling older adults, arbitrarily defined as ≥65 years, where the goal should be to achieve an SBP <130 mmHg. For older adults with hypertension and a high burden of comorbidity who have a limited life expectancy, clinical judgment and patient preference should form the basis for decisions regarding the potential value and intensity of BP treatment. An SBP/DBP treatment goal of <130/80 mmHg should not preclude the possibility of achieving a greater reduction in BP in those who tolerate their BP-lowering treatment without evidence of adverse effects. In all four of the RCTs that have compared randomization to an SBP goal of 120 versus 140 mmHg, fewer CVD complications have been noted in the group assigned to the lower BP, albeit this has only been statistically significant in two of the four trials. Three additional large 120 versus 140 mmHg RCTs are expected to report in the near future, which should result in more definitive guidance for a 120 versus 130 mmHg SBP target.

Resistant Hypertension Failure to control SBP/DBP to <130/80 mmHg with three antihypertensive medications, preferably including a diuretic, or taking four or more antihypertensive medications to achieve an SBP/DBP <130/80 mmHg is designated as apparent resistant hypertension. A general approach to the investigation and management of apparent resistant hypertension is outlined in Fig. 288-5.

First, pseudo-resistance due to inaccurate BP measurement should be excluded. Next, accurately measured out-of-office BP readings should be obtained to exclude white coat hypertension. After this important issue is addressed, lifestyle, OSA, use of medications that raise BP, and other secondary causes of hypertension should be excluded. Management of resistant hypertension starts with confirmation of adherence to the prescribed treatment regimen because lack of adherence to lifestyle change advice and medication nonadherence are common explanations for apparent resistant hypertension. It is also important to ensure that the medication regimen is based on use of long-acting, once-daily agents. This is especially true for diuretics, where shorter acting agents such as hydrochlorothiazide should be replaced by chlorthalidone or indapamide. If the antihypertensive medications have been prescribed as separate pills, conversion to a single pill combination is highly desirable. After this issue is addressed, an MRA such as spironolactone should be added to the regimen (Table 288-4). Most patients tolerate spironolactone, but in a minority, laboratory abnormalities such as hyperkalemia or side effects such as dizziness, leg cramps, and gynecomastia can be troublesome, especially in men. Many of the spironolactone side effects can be avoided by use of newer but more expensive MRA agents such as eplerenone and, especially, the nonsteroidal finerenone. In 2023, the U.S. FDA approved the use of ultrasound and radiofrequency renal denervation therapy as adjunctive treatments in patients with resistant hypertension. Consideration of renal nerve denervation (RDN) should be restricted to patients with true resistant hypertension and an office SBP ≥160 mmHg.

Treatment of Hypertension in Special Groups Certain patient groups have characteristics that warrant special attention. The following are some of the more important groups.

PATIENTS WITH CVD The combination of hypertension and CVD is very common. Choice of antihypertensive agents is commonly influenced by the specific type of CVD, but use of ACEI, ARB, or beta blockers is common, often accompanied by diuretics, CCB, and MRA. Additional treatment with SGLT2 inhibitors may be indicated, especially in those with HF and diabetes mellitus. Nondihydropyridines should be used with caution in patients with bradycardia or the potential for heart block. In general, the standard treatment goal of SBP/DBP <130/80 mmHg is indicated if well tolerated.

KIDNEY DISEASE Hypertension is very common in patients with CKD, becoming more prevalent with increasing severity of CKD, such that 80–85% of those with advanced kidney disease have concurrent hypertension. High BP is a risk factor for CKD and ESKD, but the interaction between BP and kidney disease is complex, making it difficult

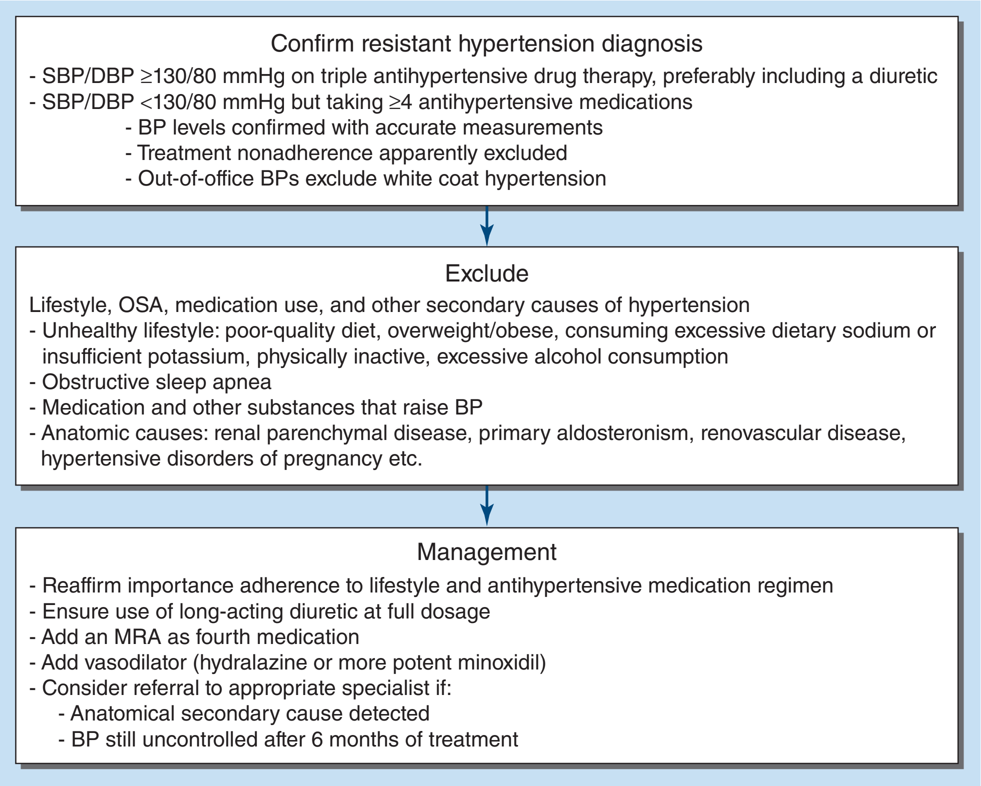

Confirm resistant hypertension diagnosis

- SBP/DBP ≥130/80 mmHg on triple antihypertensive drug therapy, preferably including a diuretic
- SBP/DBP <130/80 mmHg but taking ≥4 antihypertensive medications
  - BP levels confirmed with accurate measurements
  - Treatment nonadherence apparently excluded
  - Out-of-office BPs exclude white coat hypertension

Exclude

Lifestyle, OSA, medication use, and other secondary causes of hypertension

- Unhealthy lifestyle: poor-quality diet, overweight/obese, consuming excessive dietary sodium or insufficient potassium, physically inactive, excessive alcohol consumption
- Obstructive sleep apnea
- Medication and other substances that raise BP
- Anatomic causes: renal parenchymal disease, primary aldosteronism, renovascular disease, hypertensive disorders of pregnancy etc.

Management

- Reaffirm importance adherence to lifestyle and antihypertensive medication regimen
- Ensure use of long-acting diuretic at full dosage
- Add an MRA as fourth medication
- Add vasodilator (hydralazine or more potent minoxidil)
- Consider referral to appropriate specialist if:
  - Anatomical secondary cause detected
  - BP still uncontrolled after 6 months of treatment

FIGURE 288-5 Confirmation and management of resistant hypertension. BP, blood pressure; DBP, diastolic blood pressure; MRA, mineralocorticoid receptor antagonist; OSA, obstructive sleep apnea; SBP, systolic blood pressure.

to determine cause and effect. Patients with kidney disease have a high prevalence of both traditional and novel risk factors for CVD, including markers of inflammation and endothelial dysfunction. As a result, the burden of CVD complications is high in adults with kidney disease, and patients with ESKD are more likely to die from CVD than from other complications of kidney failure. Managing hypertension in adults with advanced kidney disease can be difficult because they frequently have a combination of increased vascular resistance and intravascular volume overload and can exhibit vascular instability with wide swings in BP that complicates management. Normally, preglomerular vasoconstriction decreases the pressure to which the glomerulus is exposed and optimizes filtration without damaging the glomerulus. The efferent glomerular arteriole can play a supplementary role, as needed, to regulate intraglomerular pressure. When kidney disease is present, the ability to regulate intraglomerular pressure is diminished, with the result that it rises and results in physical damage, a pathophysiologic process that often manifests in proteinuria and diminished glomerular filtration as the number of functioning nephrons declines.

Generally, patients with kidney disease and hypertension require several antihypertensive agents in addition to nonpharmacologic counseling, especially focused on sodium reduction. Antihypertensive drugs that block the RAAS such as ACEIs and ARBs play a central role because they can preserve glomerular function by inducing efferent arteriolar dilatation and reducing intraglomerular pressure. They are particularly important in patients with proteinuria and advanced kidney disease. Diuretics also play an important role because they can concurrently reduce peripheral resistance and intravascular volume. Traditionally, loop diuretics have been recommended in place of thiazide or thiazide-like diuretics in patients with an eGFR of ~30 mL/min per 1.73 m^2 or lower, but a recent high-quality RCT in patients with stage 4 kidney disease (average eGFR <23 mL/min per 1.73 m^2) demonstrated substantially lower (10.5 mmHg) reduction in ambulatory BP and a halving of urinary albumin/creatinine ratio in those randomized to chlorthalidone (12.5–50 mg) compared to placebo. Although CCBs tend to dilate the preglomerular afferent artery, they have a good safety record in kidney disease patients when used in combination with an ACEI or ARB. Single-pill antihypertensive combinations are especially useful in patients with kidney disease because they are usually taking a large number of medications. Electrolyte levels should be monitored carefully, especially serum potassium levels, in kidney disease patients who are treated with potassium-sparing agents, including ACEI, ARB, and MRA drugs. Dual therapy with an ACEI, ARB, or ACEI-ARB combination is dangerous because of the risk of hyperkalemia and decline in kidney function.

PATIENTS WITH DIABETES MELLITUS The combination of hypertension and diabetes is common, resulting in a substantial risk for CVD and kidney disease. In patients with longstanding severe diabetes mellitus, autonomic neuropathy or volume depletion can cause orthostatic hypotension (10 mmHg difference in SBP between sitting and upright pressures), making management of hypertension difficult and requiring careful monitoring. Symptomatic hypotension (syncope) is a more important finding and may require a stepdown in antihypertensive therapy. Numerous RCTs in adults with diabetes have demonstrated that antihypertensive drug therapy reduces the risk of ASCVD, HF, and microvascular complications. The recommended antihypertensive treatment target is an SBP/DBP <130/90 mmHg. Almost all patients with hypertension and diabetes mellitus require antihypertensive combination therapy, with many requiring three or four agents to achieve satisfactory BP control. Inclusion of an ACEI or ARB in the combination is recommended, especially in those with heavy proteinuria or kidney insufficiency. In addition, most patients require a diuretic. CCBs are commonly part of the antihypertensive drug regimen as well. In patients with diabetes mellitus and resistant hypertension, the addition of an MRA has been effective; however, serum potassium should be monitored carefully. Many drugs used to treat diabetes mellitus affect BP. Perhaps the best documented are SGLT2 inhibitors, which in meta-analyses result in an SBP/DBP reduction of ~5/2 mmHg, a modest reduction that does not meet the FDA requirement for classification as an antihypertensive agent. Given that patients with diabetes mellitus are often taking many medications, use of single-pill dual- or triple-combination therapy is especially useful.

OLDER ADULTS Vascular stiffening increases with aging, and as a result, isolated elevation of SBP and a wide pulse pressure are common in older adults. Comorbidity is common, resulting in the potential for a compelling nonhypertensive indications for use of BP-lowering agents. Drug-drug interactions and delayed drug excretion due to kidney or liver disease are also possible. Despite these challenges, most older adults with hypertension can be treated in a manner similar to younger patients. Indeed, the RCT evidence for antihypertensive drug therapy generally comes from studies in which the average age was close to 70 years at baseline. Older adults tend to be at high risk for CVD and all-cause mortality and have a disproportionately large treatment-related reduction in the absolute CVD and all-cause mortality benefit. A natural concern in treating hypertension in older adults is the potential for excessive BP reduction resulting in syncope and resultant falls, or organ hypoperfusion with resultant infarction, especially in the distribution of the coronary arteries. Thus far, RCTs of antihypertensive drug treatment in older adults have been reassuring, resulting in substantial prevention of CVD and all-cause mortality with limited evidence of adverse events. An increased incidence of hypotension has been noted in some but not all trials of antihypertensive treatment in older adults. However, there has been no evidence of a resultant increase in syncope, falls, or infarctions in the RCTs. Nonrandomized analyses of data sets have identified a J-shaped association between BP and CVD risk. However, J-curves are common in observational epidemiology, and randomized comparisons, which are more helpful in guiding treatment decisions, have revealed prevention of CVD and all-cause mortality in those randomized to more versus less intensive antihypertensive therapy at any point in the J-curve. A reasonable approach is to initiate or enhance existing treatment using lower doses of medications than in younger patients. Because congestive heart failure becomes increasingly common at older age, use of a diuretic rather than a CCB for dual-therapy combinations with an ACEI or ARB is sensible. Many older patients require triple therapy or have comorbidities that necessitate treatment with a CCB. As is the case generally, BP lowering is more important than the treatment combination used to achieve the desired SBP goal of <130 mmHg. Recent trials have provided compelling evidence that more intensive antihypertensive treatment in older adults reduces the risk of cognitive impairment compared to less intensive treatment, and increasing evidence suggests that it reduces the risk of dementia.

RACE/ETHNICITY Non-Hispanic blacks have disproportionately high BP levels and higher prevalence of hypertension compared to whites and other major race/ethnicity groups in the United States. In addition, hypertension tends to occur earlier in life and is associated with a higher risk of CVD and kidney disease in blacks compared to whites and the other major race/ethnicity groups. Non-Hispanic blacks have the highest awareness of hypertension of any race/ethnicity group in the United States and have an antihypertensive drug treatment rate that is at least as high as that of whites. However, their rate of control to <140/90 mmHg in the 2017–2020 NHANES survey was 15% lower than in whites (37 vs 52%) and 6% lower for the 130/80 mmHg cut point (20 vs 26%). Genetic differences seem to explain a small fraction of the high prevalence of hypertension in black Americans, but most of the disparity is likely due to differences in the social determinants of health, including access to and quality of education and health care, neighborhood of residence and the related built environment factors, economic instability, and the social and community context. The BP, CVD, and kidney disease health discrepancies between black and nonblack U.S. adults are not only sizable but have also been persistent, with little improvement in recent decades. RCT experience demonstrates that the gap can be eliminated with high-quality care, but following termination of RCT care, there has been a relatively rapid reversion to previous patterns of inadequate BP control. Resolution of the situation is unlikely without structural change in the approach to provision of care and a multilevel effort that involves patient engagement, a team-based approach to care that frees up clinicians to address social determinants of health, health system policies that enhance the provision of high-quality care, and payors with a focus on quality and use of contemporary BP goals for BP control. Diuretics and CCBs are especially effective for lowering BP in U.S. blacks, but many patients require triple antihypertensive therapy, and use of BP-lowering agents for management of comorbid conditions is common.

SEX Although there are well-described differences in the underlying pathophysiology of hypertension in women and men, RCT experience indicates a similar benefit in prevention of CVD in both sexes in the landmark antihypertensive drug treatment trials. In the United States, women are more likely to be aware of hypertension, to be treated with antihypertensive medication, and to achieve the desired level of BP control compared with men, especially at younger ages. ACEIs and ARBs should not be used in pregnancy or women of child-bearing age who are contemplating pregnancy. Antihypertensive agents with a good safety record in pregnancy are highlighted in the section on hypertension in pregnancy. Observational studies and RCTs, including meta-analyses, have reported a reduced risk of fractures in women being treated with thiazide diuretics. This is postulated to result from decreased urinary calcium excretion and increased osteoblast cell formation. In contrast to thiazide diuretics, loop diuretics increase urinary calcium excretion. However, analyses of large observational studies have failed to associate loop diuretic use in postmenopausal women with fractures or a decrease in bone marrow density. Women report more cough with ACEIs and more pedal edema with CCB treatment compared with men.

Hypertensive Urgencies and Emergencies Adults presenting with a very high level of BP, usually designated as an SBP/DBP ≥180/100 mmHg, should be classified as having a hypertensive urgency if they are asymptomatic or a hypertensive emergency if there is evidence of active ongoing hypertensive end-organ damage such as hypertensive encephalopathy, often manifesting with headaches, vision defects, nausea, vomiting, seizures, acute left ventricular failure, or acute kidney failure. Hypertensive urgency is far more common than hypertensive emergency and should be treated with institution, reinstitution, or intensification of oral antihypertensive agents in an outpatient setting. In contrast, a hypertensive emergency requires immediate, carefully supervised management with intravenous antihypertensive agents in an emergency room or inpatient setting. A hypertensive emergency associated with acute aortic dissection, eclampsia or severe preeclampsia, or a pheochromocytoma in crisis signals the need for particularly rapid care. Several drug classes can be used intravenously to achieve a rapid reduction in BP. Intravenous sodium nitroprusside is a potent vasodilator with a long history of success in managing hypertensive emergencies but must be monitored carefully to avoid overshooting the goal BP and causing hypotension. The dihydropyridine CCB nicardipine, given intravenously, is perhaps the most commonly utilized agent for acute BP reduction in patients with a hypertensive emergency in the contemporary practice of medicine. No dose adjustment is required for its use in older adults, but it is contraindicated in patients with severe aortic stenosis. Intravenous administration of the beta blocker labetalol is also effective but is contraindicated in patients with obstructive airways diseases, chronic obstructive pulmonary disease, bradycardia, or second- or third-degree heart block. Whichever drug is employed, a reduction rather than immediate normalization of BP is sufficient, with subsequent management of the patient's hypertension with oral agents. Secondary causes of hypertension, especially renovascular disease, are relatively common in patients with a hypertensive emergency and should be considered after the immediate reduction in BP.

PREVALENCE, AWARENESS, TREATMENT, AND CONTROL OF HYPERTENSION

NHANES provides a way to monitor prevalence, awareness, treatment, and control of hypertension in the U.S. general population. The NHANES definition of hypertension is based on BPs measured at a single visit, which results in an overestimation of hypertension, and the HHANES definition of treatment and control do not take into account nondrug therapy of hypertension. However, the methods used in NHANES have been consistent and careful over time, providing very helpful estimates for the temporal trends in hypertension prevalence and BP control. NHANES hypertension prevalence estimates have been fairly stable at ~46 and 32% for definitions based on use of the

SBP/DBP cut points of 130/80 and 140/90 mmHg, respectively, with higher prevalence rates in older versus younger adults, non-Hispanic blacks versus all other major race/ethnicity groups, and men compared with women. Treatment (with antihypertensive agents) and hypertension control rates increased progressively until about 2009–2012, with close to 53% of U.S. adults being controlled to an SBP/DBP <140/90 mmHg (the recommended goal at the time) and ~26% for an SBP/DBP <130/90 mmHg. After that, there was a progressive decline in control rates, with ~48 and 24% having an SBP/DBP <140/90 and <130/80 mmHg, respectively, by 2017–2020. The percentage controlled was higher when the analysis was confined to adults being treated for hypertension, but the temporal trends pattern was similar. The decrement in hypertension control was especially prominent in adults ≥75 years old. There has been a persistent gap in hypertension control rates between non-Hispanic black adults and all other race/ethnicity groups in the United States, with the percentage of non-Hispanic blacks and whites controlled to an SBP/DBP <140/90 mmHg in 2017–2020 being ~37 and 52%, respectively (20 and 26% for an SBP/DBP <130/80 mmHg). Worldwide, the situation is worse, with a recent global estimate reporting control to an SBP/DBP <140/90 mmHg in <14% of adults and in <8% of adults in low- and middle-income countries, where most of the world's population resides.

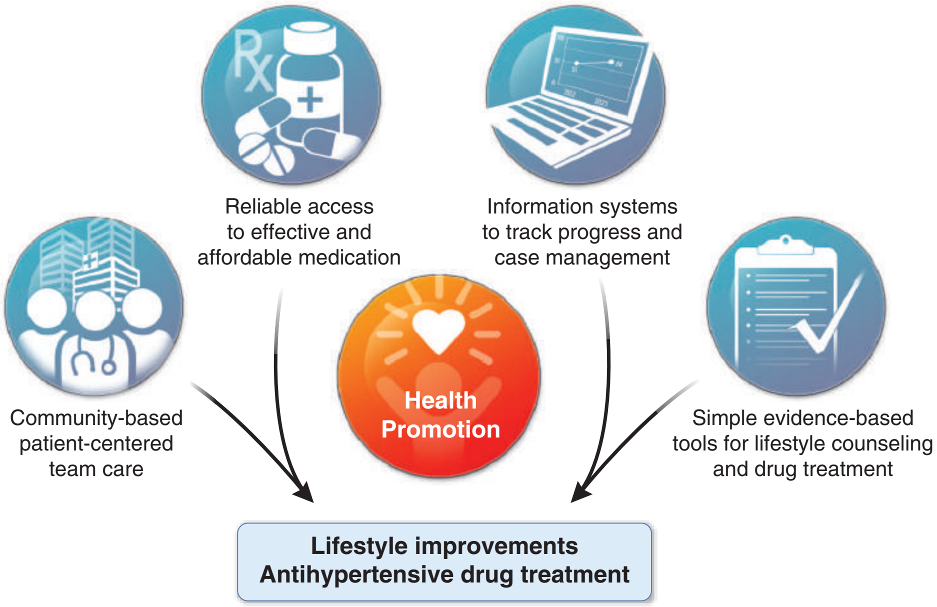
Reliable access to effective and affordable medication
Information systems to track progress and case management
Community-based patient-centered team care
Health Promotion
Simple evidence-based tools for lifestyle counseling and drug treatment
Lifestyle improvements
Antihypertensive drug treatment

- Commitment by provider, system of care, or country to specific goals
- Health promotion for BP-lowering and antihypertensive medication augementation
- Manage uncomplicated hypertension with patient-centered, team-based care
- Ensure access to effective and affordable meds, preferably at the point of care
- Use simple evidence-based protocols and algorithms for team management
- Engage patients in their own care with HBPM and other options
- Track progress and employ case management for immediate corrective actions

FIGURE 288-6 Best practices for control of hypertension. BP, blood pressure; HBPM, home blood pressure monitoring.

IMPROVING THE CONTROL OF HYPERTENSION

Clearly, the current hypertension control rates are unacceptable, and the traditional model of one-on-one physician-patient care for hypertension is not yielding a satisfactory outcome. Primary care physicians, who care for most patients with hypertension, are overburdened with responsibilities and have very limited time for direct provision of care in those with "routine" uncomplicated hypertension. A number of approaches have been shown to improve hypertension control rates in RCTs and clinical practice settings. Among these, team-based care has resulted in the biggest improvement in RCTs and meta-analyses. In the typical model of team-based care for hypertension management, a physician coordinates care provided by a team that may include nurses, pharmacists, community health care workers, social workers, lifestyle counselors, medical technicians, or others who have been specially trained in the provision of hypertension care. The specifics of the team composition depend on what is feasible in a practice setting and can be extended to engage front-office staff, spouses/partners, and close friends. In RCTs, team-based care has resulted in an average SBP reduction of ~7 mmHg, but in a recent high-quality, well-executed RCT, the SBP reduction exceeded 23 mmHg. It has been especially effective when trained nonphysician professionals have had the authority, usually with physician oversight, to prescribe antihypertensive medications in patients with uncomplicated hypertension. This frees up physician time to address patients with more complex care requirements. Most team-based care interventions have been multifaceted, incorporating health counseling, home BP monitoring (HBPM) and electronic support systems, and use of simple algorithms for health counseling and management of antihypertensive drug therapy. HBPM provides an excellent way to engage patients in their own care and to complement office BP management (OBPM) for monitoring long-term control of hypertension. Use of HBPM mandates advising patients on the purchase of clinically valid BP measurement devices and training in accurate measurement of BP (see BP measurement section). In U.S. adults, getting three morning (prior to taking BP medications) and three evening HBPM measurements for 3 days prior to a clinic visit is a reasonable strategy for assessing average home BP. Team-based care approaches have not only been successful in RCTs but also yielded excellent results in clinical practice settings, with Kaiser-Permanente, Northern California reporting an improvement in SBP/DBP control to <140/90 mmHg from 44% in 2001 to 90% in 2015 based on institution of a multifaceted systemwide program of health care quality improvement that incorporated many of the previously mentioned approaches. Likewise, many of these approaches have been used to good effect in the U.S. Department of Veterans Affairs Health System. An overall best practices approach is outlined in Fig. 288-6.

While it may not be possible for an individual practitioner to implement all of the suggestions in the figure, they should advocate for those that are beyond their direct control. There is great need to improve the management and control of hypertension in the United States and worldwide. Hypertension is among the most important, prevalent, modifiable, and cost-effective risk factors for CVD, kidney disease, and all-cause mortality. Its detection, treatment, and control should be a high priority for individual clinicians, systems of care, and countries worldwide.

FURTHER READING

Blood Pressure Lowering Treatment Trialists, Collaboration: Pharmacological blood pressure lowering for primary and secondary prevention of cardiovascular disease across different levels of blood pressure: An individual participant-level data meta-analysis. Lancet 397:1625, 2021.

Bundy JD et al: Systolic blood pressure reduction and risk of cardiovascular disease and mortality: A systematic review and network meta-analysis. JAMA Cardiol 2:775, 2017.

Carey RM et al: Treatment of hypertension. A review. JAMA 328:1848, 2022.

Fuchs FD, Whelton PK: High blood pressure and cardiovascular disease. Hypertension 75:285, 2020.

Panagiotis G, Agarwal R: Hypertension in chronic kidney disease: Treatment standard 2023. Nephrol Dial Transplant 3:2694, 2023.

Whelton P et al: 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 71:e13, 2018.

Wright JT Jr. et al: SPRINT revisited: Updated results and implications. Hypertension 78:1701, 2021.

289 Renovascular Disease

Stephen C. Textor[*]

The renal vasculature is unusually complex with high arteriolar flow to the cortex in excess of metabolic requirements, consistent with its primary function as a filtering organ. After delivering blood to cortical glomeruli, the postglomerular circulation supplies deeper medullary segments that support energy-dependent solute transport at multiple levels of the renal tubule. These postglomerular vessels deliver less blood and, coupled with high oxygen consumption, leave the deeper medullary regions at the margin of hypoxemia. Vascular disorders that commonly threaten the blood supply of the kidney include large-vessel atherosclerosis, fibromuscular diseases, and embolic disorders. **Microvascular injury, including inflammatory and primary hematologic disorders, is described in Chap. 329**.

MECHANISMS OF VASCULAR INJURY AND HYPERTENSION

The glomerular capillary endothelium shares susceptibility to oxidative stress, pressure injury, and inflammation with other vascular territories. Endothelial injury can be manifest by urinary albumin excretion (UAE), which is predictive of systemic atherosclerotic disease events. Increased UAE may develop years before cardiovascular events. UAE and the risk of cardiovascular events are both reduced with pharmacologic therapy such as antihypertensive drugs and statins. Experimental studies demonstrate functional changes and rarefaction of renal microvessels under conditions of accelerated atherosclerosis and/or compromise of proximal perfusion pressures with large-vessel disease **(Fig. 289-1)**.

Large-vessel renal artery occlusive disease can result from multiple conditions, including extrinsic compression of the vessel, intimal dissection, aortic stent graft placement, fibromuscular dysplasia (FMD), or, most commonly, atherosclerotic disease. Any disorder that reduces perfusion pressure to the kidney can activate mechanisms that tend to restore renal pressures at the expense of developing systemic hypertension. Because restoration of perfusion pressures can reverse these pathways, renovascular disease is considered a specifically treatable "secondary" cause of hypertension.

Renal artery stenosis is common, usually gradually progressive, and often has only minor hemodynamic effects. FMD is reported in 3–5% of normal subjects presenting as potential kidney donors without hypertension. It may present clinically with hypertension in younger individuals (between age 15 and 50), most often women. FMD does not often threaten kidney function, but sometimes produces total occlusion and can be associated with renal artery aneurysms. Atherosclerotic renal artery stenosis (ARAS) is common in the general population (6.8% of a community-based sample above age 65). The prevalence increases with age and for patients with other vascular conditions such as coronary artery disease (18–23%) and/or peripheral aortic or lower extremity disease (>30%). If untreated, ARAS progresses in nearly 50% of cases over a 5-year period, sometimes to total occlusion. Intensive treatment of arterial blood pressure and statin therapy can slow these rates and improve clinical outcomes.

Critical levels of stenosis (usually >70–80% luminal obstruction) lead to a reduction in perfusion pressure that activates the renin-angiotensin system, reduces sodium excretion, and activates sympathetic adrenergic pathways. These events lead to systemic hypertension characterized by angiotensin dependence in the early stages, widely varying pressures, loss of circadian blood pressure (BP) rhythms, and accelerated target organ injury, including left ventricular hypertrophy and renal fibrosis. Renovascular hypertension can be treated with agents that block the renin-angiotensin system and other drugs that modify these pressor pathways. It can also be treated with restoration of renal blood flow by either endovascular or surgical revascularization. Most patients require continued antihypertensive drug therapy due to preexisting hypertension and because revascularization alone rarely lowers BP to normal.

ARAS and systemic hypertension tend to affect both the poststenotic and contralateral kidneys, reducing overall glomerular filtration rate (GFR) in ARAS. When kidney function is threatened by large-vessel disease primarily, it has been labeled ischemic nephropathy. Moderately reduced blood flow that develops gradually is associated with reduced GFR and limited oxygen consumption with preserved tissue oxygenation. Hence, kidney function often remains reduced but stable during medical therapy, sometimes for years. With more advanced disease, reductions in cortical perfusion and overt tissue hypoxia develop. Unlike FMD, ARAS develops in patients with other risk factors for atherosclerosis and is commonly superimposed upon preexisting small-vessel disease in the kidney resulting from hypertension, aging, and diabetes. Nearly 85% of patients considered for renal revascularization have stage 3–5 chronic kidney disease (CKD) with estimated GFR <60 mL/min per 1.73 m^2. The presence of ARAS is a strong predictor of morbidity- and mortality-related cardiovascular events, independent of whether renal revascularization is undertaken.

DIAGNOSIS OF RENOVASCULAR DISEASE

Diagnostic approaches to renovascular disease necessarily include evaluation of the kidney vasculature and depend on the specific clinical questions to be addressed. Noninvasive characterization of the renal vasculature may be achieved by several techniques, summarized in **Table 289-1**. Although activation of the renin-angiotensin system is a key step in developing renovascular hypertension, it is transient. Levels of renin activity are therefore subject to timing, the effects of drugs, and sodium intake, and do not reliably predict the response to vascular therapy. Peak systolic renal artery velocities by Doppler ultrasound >200 cm/s generally predict lesions with more than >60% vessel lumen occlusion, although some treatment trials have required velocity >300 cm/s to avoid false positives. The renal resistive index has predictive value regarding the viability of the kidney. It remains operator- and institution-dependent, however. Contrast-enhanced computed tomography (CT) with vascular reconstruction provides excellent vascular images and functional assessment but carries a small risk of contrast toxicity. It provides a more reliable evaluation of accessory vessels and the distal vasculature than duplex or magnetic resonance imaging (MRI). Magnetic resonance angiography (MRA) is less often used than previously, as gadolinium contrast has been associated with nephrogenic systemic fibrosis particularly in patients with reduced GFR. Captopril-enhanced renography has a strong negative predictive value when entirely normal.

TREATMENT

Renal Artery Stenosis

While restoring renal blood flow and perfusion seems intuitively beneficial for high-grade occlusive lesions, revascularization procedures also pose hazards and expense. Patients with FMD are commonly younger females with otherwise normal vessels and a long

---

[*] Deceased

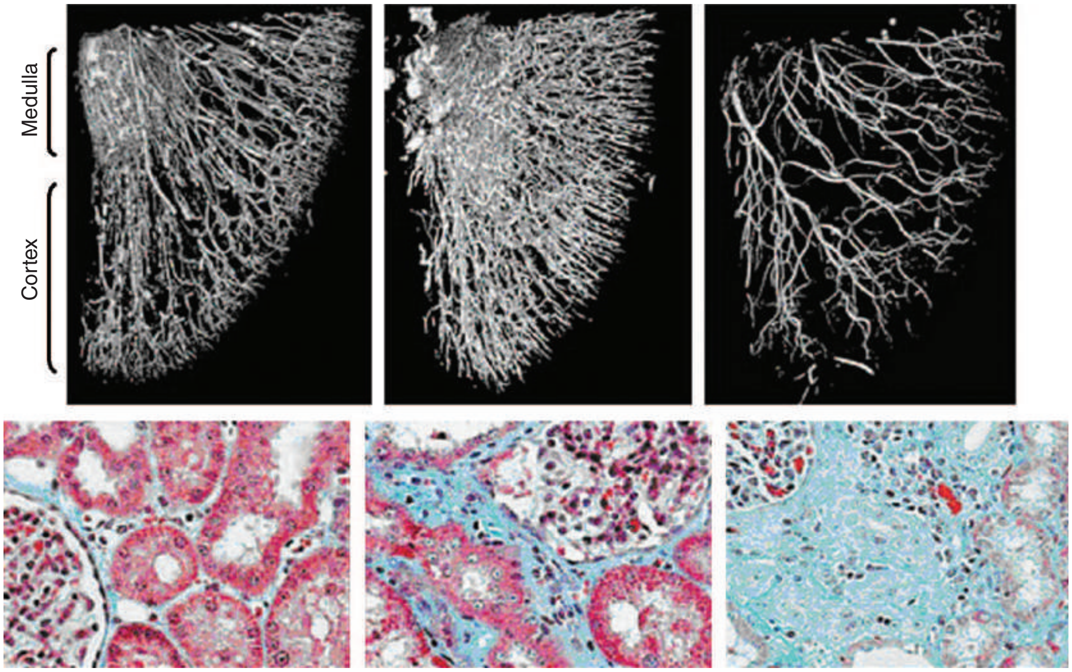

Normal

MV proliferation
(early atherosclerosis)

MV rarefaction
(chronic renal ischemia)

FIGURE 289-1 Examples of micro-CT images from vessels defined by radiopaque casts injected into the renal vasculature. These illustrate the complex, dense cortical capillary network supplying the kidney cortex that can either proliferate or succumb to rarefaction under the influence of atherosclerosis and/or occlusive disease. Changes in blood supply are followed by tubulointerstitial fibrosis and loss of kidney function. MV, microvascular. *(Reproduced with permission from LO Lerman, AR Chade. Angiogenesis in the kidney: A new therapeutic target? Curr Opin Nephrol Hypertens 18:160, 2009.)*

life expectancy. These patients often respond well to percutaneous renal artery angioplasty. If BP can be controlled to goal levels and kidney function remains stable in patients with ARAS, it may be argued that medical therapy with follow-up for disease progression is equally effective over periods of 3–5 years. Multiple prospective randomized controlled trials for individuals with moderate stenosis have failed to identify compelling additional benefits for interventional revascularization procedures regarding short-term results of BP and renal function. Studies of cardiovascular outcomes, including stroke, congestive heart failure, myocardial infarction, and end-stage renal failure, suggest a small mortality benefit for stented patients without proteinuria. Medical therapy should include blockade of the renin-angiotensin system, attainment of goal BPs, cessation of tobacco, statins, and aspirin. Follow-up requires surveillance for progressive occlusion manifest by worsening renal function and/or loss of BP control. Renal revascularization should be considered for patients with rapidly progressive clinical syndromes, failing medical therapy, and/or developing additional complications.

Techniques of renal revascularization are improving. With experienced operators, major complications occur in <5% of cases, including renal artery dissection, capsular perforation, hemorrhage, and occasional atheroembolic disease. Although not common, atheroembolic disease can be catastrophic and accelerate both hypertension and kidney failure, precisely the events that revascularization is intended to prevent. Although renal blood flow usually can be restored by endovascular stenting, recovery of renal function is limited to ~25% of cases, with no change in 50% and some deterioration evident in others. Patients with rapid loss of kidney function, sometimes associated with antihypertensive drug therapy, or with vascular disease affecting the entire functioning kidney mass

TABLE 289-1 Summary of Imaging Modalities for Evaluating the Kidney Vasculature

Perfusion Studies to Assess Differential Renal Blood Flow			
Captopril renography with technetium-99m mertiatide (^{99m}Tc MAG3)	Captopril-mediated fall in filtration pressure amplifies differences in renal perfusion	Normal study excludes renovascular hypertension	Multiple limitations in patients with advanced atherosclerosis or creatinine >2.0 mg/dL (177 μmol/L)
Vascular Studies to Evaluate the Renal Arteries			
Duplex ultrasonography	Shows the renal arteries and measures flow velocity as a means of assessing the severity of stenosis	Inexpensive; widely available, suitable for follow-up studies	Heavily dependent on operator's experience; less useful than invasive angiography for the diagnosis of fibromuscular dysplasia and abnormalities in accessory renal arteries
Computed tomographic angiography	Shows the renal arteries and perirenal aorta	Provides excellent images; stents do not cause artifacts	Expensive, moderate volume of contrast required
Magnetic resonance angiography	Shows the renal arteries and perirenal aorta	Not nephrotoxic, but concerns for gadolinium toxicity exclude use in GFR <30 mL/min/1.73 m^2; provides excellent images	Expensive; gadolinium excluded in renal failure, unable to visualize stented vessels
Intraarterial angiography	Shows location and severity of vascular lesion	Considered "gold standard" for diagnosis of large-vessel disease, usually performed simultaneously with planned intervention	Expensive, associated hazard of atheroemboli, contrast toxicity, procedure-related complications, e.g., dissection

Abbreviation: GFR, glomerular filtration rate.

TABLE 289-2 Clinical Factors That Determine the Role of Revascularization in Addition to Medical Therapy for Renal Artery Stenosis

Factors Favoring Medical Therapy with Revascularization for Renal Artery Stenosis
• Progressive decline in GFR during treatment of systemic hypertension
• Failure to achieve adequate blood pressure control with optimal medical therapy (medical failure)
• Rapid or recurrent decline in the GFR in association with a reduction in systemic pressure
• Decline in the GFR during therapy with ACE inhibitors or ARBs
• Recurrent congestive heart failure in a patient in whom left ventricular dysfunction does not fully explain the cause
Factors Favoring Medical Therapy and Surveillance of Renal Artery Disease
• Controlled blood pressure with stable renal function (e.g., stable renal insufficiency)
• Stable renal artery stenosis without progression on surveillance studies (e.g., serial duplex ultrasound)
• Advanced age and/or limited life expectancy
• Extensive comorbidity that make revascularization too risky
• High risk for or previous experience with atheroembolic disease
• Other concomitant renal parenchymal diseases that cause progressive renal dysfunction (e.g., interstitial nephritis, diabetic nephropathy), particularly with proteinuria

Abbreviations: ACE, angiotensin-converting enzyme; ARBs, angiotensin receptor blockers; GFR, glomerular filtration rate.

are more likely to recover function after restoring blood flow. When hypertension is refractory to effective therapy, revascularization offers real benefits. Table 289-2 summarizes currently accepted guidelines for considering renal revascularization in addition to optimal medical therapy.

ATHEROEMBOLIC RENAL DISEASE

Emboli to the kidneys arise most frequently as a result of cholesterol crystals breaking free of atherosclerotic vascular plaque and lodging in downstream microvessels. Most clinical atheroembolic events follow angiographic procedures, often of the coronary vessels. It has been argued that nearly all arterial interventional procedures lead to plaque fracture and release of microemboli, but clinical manifestations develop only in a fraction of these. The incidence of clinical atheroemboli has been increasing with more vascular procedures and longer life spans. Atheroembolic renal disease is suspected in >3% of elderly subjects with end-stage renal disease (ESRD) and is likely underdiagnosed. It is more frequent in males with a history of diabetes, hypertension, and ischemic cardiac disease. Atheroemboli in the kidney are strongly associated with aortic aneurysmal disease and renal artery stenosis. Most clinically evident cases can be linked to precipitating events, such as angiography, vascular surgery, anticoagulation with heparin, thrombolytic therapy, or trauma. Clinical manifestations of this syndrome commonly develop between 1 and 14 days after an inciting event and may continue to develop for weeks thereafter. Systemic embolic disease manifestations, such as fever, abdominal pain, and weight loss, are present in less than half of patients, although cutaneous manifestations including livedo reticularis and localized toe gangrene may be more common. Worsening hypertension and deteriorating kidney function are common, sometimes reaching a malignant phase. Progressive renal failure can occur and require dialytic support. These cases often develop after a stuttering onset over many weeks and have an ominous prognosis. Mortality rate after 1 year exceeds 38%, and although some may eventually recover sufficiently to no longer require dialysis, many do not.

Beyond the clinical manifestations above, laboratory findings include rising creatinine, transient eosinophilia (60–80%), elevated sedimentation rate, and hypocomplementemia (15%). Establishing this diagnosis can be difficult and is often by exclusion. Definitive diagnosis depends on kidney biopsy demonstrating microvessel occlusion with cholesterol crystals that leave a "cleft" in the vessel. Biopsies obtained from patients undergoing surgical revascularization of the kidney indicate that silent cholesterol emboli are frequently present before any further manipulation is performed.

No effective therapy is available for atheroembolic disease once it has developed. Withdrawal of anticoagulation is recommended. Late recovery of kidney function after supportive measures sometimes occurs, and statin therapy may improve outcome. The role of embolic protection devices in the renal circulation during angiography is unclear, but a few prospective trials have failed to demonstrate major benefits. The effect of such devices is limited to distal protection during the endovascular procedure, and they offer no protection from embolic debris developing after removal.

THROMBOEMBOLIC RENAL DISEASE

Thrombotic occlusion of renal vessels or branch arteries can lead to declining renal function and hypertension. It is difficult to diagnose and is often overlooked, especially in elderly patients. Thrombosis can develop as a result of local vessel abnormalities, such as local dissection, trauma, inflammatory vasculitis, or systemic infections, such as COVID-19. Local microdissections sometimes lead to patchy, transient areas of infarctions labeled "segmental arteriolar mediolysis." Although hypercoagulability conditions sometimes present as renal artery thrombosis, this is rare. It can also derive from distant embolic events, e.g., the left atrium in patients with atrial fibrillation or from fat emboli originating from traumatized tissue, most commonly large bone fractures. Cardiac sources include vegetations from subacute bacterial endocarditis. Systemic emboli to the kidneys may also arise from the venous circulation if right-to-left shunting occurs, e.g., through a patent foramen ovale.

Clinical manifestations vary depending on the rapidity of onset and extent of occlusion. Acute arterial thrombosis may produce flank pain, fever, leukocytosis, nausea, and vomiting. If kidney infarction results, enzymes such as lactate dehydrogenase (LDH) rise transiently to extreme levels. If both kidneys are affected, renal function will decline precipitously with a drop in urine output. If a single kidney is involved, renal functional changes may be minor. Hypertension related to sudden release of renin from ischemic tissue can develop rapidly, as long as some viable tissue in the "peri-infarct" border zone remains. If the infarct zone demarcates precisely, the rise in BP and renin activity may resolve. Diagnosis of renal infarction may be established by vascular imaging with CT angiography, MRI, or arteriography **(Fig. 289-2)**.

■ MANAGEMENT OF ARTERIAL THROMBOSIS OF THE KIDNEY

Options for interventions of newly detected arterial occlusion include surgical reconstruction, anticoagulation, thrombolytic therapy, endovascular procedures, and supportive care, particularly antihypertensive drug therapy. Application of these methods depends on the patient's overall condition, the precipitating factors (e.g., local trauma or systemic illness), the magnitude of renal tissue and function at risk, and the likelihood of recurrent events in the future. For unilateral disease, for example, arterial dissection with thrombosis and supportive care with anticoagulation may suffice. Acute, bilateral occlusion is potentially catastrophic, producing anuric renal failure. Depending on the precipitating event, surgical or thrombolytic therapies can sometimes restore kidney viability if undertaken early in the course of the acute event.

MICROVASCULAR INJURY ASSOCIATED WITH HYPERTENSION

■ ARTERIOLONEPHROSCLEROSIS

"Malignant" Hypertension Although BP rises with age, it has long been recognized that some individuals develop rapidly progressive BP elevations with target organ injury including retinal hemorrhages, encephalopathy, and declining kidney function. Placebo arms during the early controlled trials of hypertension therapy identified progression to severe levels in 20% of subjects over 5 years. If untreated, patients with target organ injury including papilledema

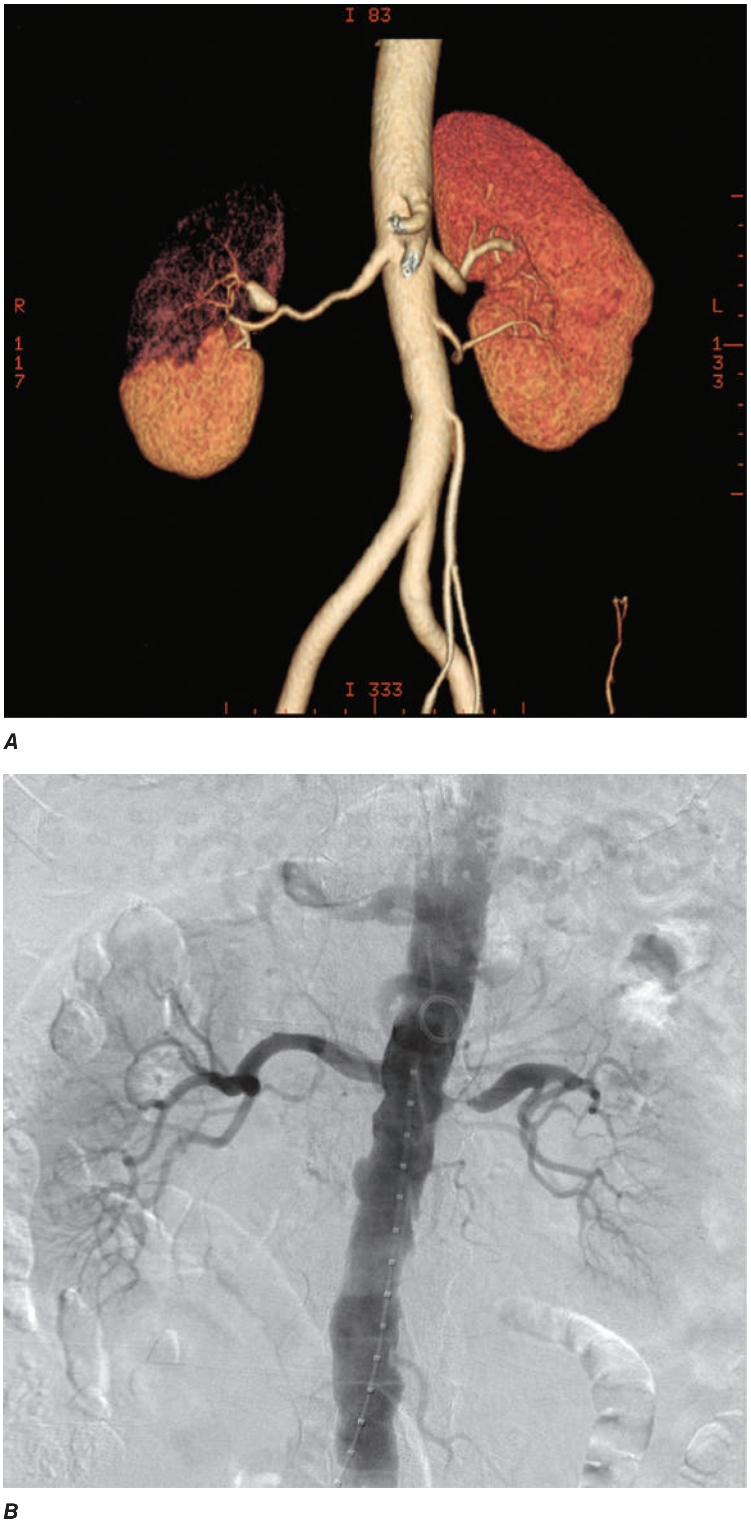

FIGURE 289-2 ***A.*** CT angiogram illustrating loss of circulation to the upper pole of the right kidney in a patient with fibromuscular disease and a renal artery aneurysm. Activation of the renin-angiotensin system produced rapidly developing hypertension. ***B.*** Angiogram illustrating high-grade renal artery stenosis affecting the left kidney. This lesion is often part of widespread atherosclerosis and sometimes is an extension of aortic plaque. This lesion develops in older individuals with preexisting atherosclerotic risk factors.

and declining kidney function suffered mortality rates in excess of 50% over 6–12 months, hence the designation "malignant." Postmortem studies of such patients identified vascular lesions, designated "fibrinoid necrosis," with breakdown of the vessel wall, deposition of eosinophilic material including fibrin, and a perivascular cellular infiltrate. A separate lesion was identified in the larger interlobular arteries in many patients with hyperplastic proliferation of the vascular wall cellular elements, deposition of collagen, and separation of layers, designated the "onionskin" lesion. For many of these patients, fibrinoid necrosis led to obliteration of glomeruli and loss of tubular structures. Progressive kidney failure ensued and, without dialysis support, led to early mortality in untreated malignant-phase hypertension. These vascular changes could develop with pressure-related injury from a variety of hypertensive pathways, including but not limited to activation of the renin-angiotensin system and severe vasospasm associated with catecholamine release. Occasionally, endothelial injury is sufficient to induce microangiopathic hemolysis, as discussed below.

Antihypertensive therapy is the mainstay of therapy for malignant hypertension. With effective BP reduction, manifestations of vascular injury, including microangiopathic hemolysis and renal dysfunction, can improve over time. Whereas prior reports before the era of drug therapy suggested that 1-year mortality rates exceeded 90%, current survival over 5 years exceeds 50%.

Malignant hypertension is less common in Western countries, although it persists in parts of the world where medical care and antihypertensive drug therapy are less available. It most commonly develops in patients with treated hypertension who neglect to take medications or who may use vasospastic drugs, such as cocaine. Renal abnormalities typically include rising serum creatinine and occasionally hematuria and proteinuria. Biochemical findings may include evidence of hemolysis (anemia, schistocytes, and reticulocytosis) and changes associated with kidney failure. African-American males are more likely to develop rapidly progressive hypertension and kidney failure than are whites in the United States. Genetic polymorphisms for *APOL1* that are common in the African-American population predispose to focal sclerosing glomerular disease, with severe hypertension developing at younger ages secondary to renal disease in this instance.

"Hypertensive Nephrosclerosis" Based on experience with malignant hypertension and epidemiologic evidence linking BP with long-term risks of kidney failure, it has long been believed that lesser degrees of hypertension induce less severe, but prevalent, changes in kidney vessels and loss of kidney function. As a result, a large portion of patients reaching ESRD without a specific etiologic diagnosis are assigned the designation "hypertensive nephrosclerosis." Pathologic examination commonly identifies afferent arteriolar thickening with deposition of homogeneous eosinophilic material (hyaline arteriolosclerosis) associated with narrowing of vascular lumina. Clinical manifestations include retinal vessel changes associated with hypertension (arteriolar narrowing, arteriovenous crossing changes), left ventricular hypertrophy, and elevated BP. The role of these vascular changes in kidney function is unclear. Multiple studies of nephrectomy and biopsy samples from normotensive kidney donors demonstrate changes of nephrosclerosis associated with aging, dyslipidemia, and glucose intolerance, with only modest association with BP. Antihypertensive drug therapy does not alter the course of kidney dysfunction identified specifically as hypertensive nephrosclerosis, although BP reduction does slow progression of proteinuric kidney diseases and is warranted to reduce the excessive cardiovascular risks associated with CKD.

FURTHER READING

Bhalla V et al: Revascularization for renovascular disease: A scientific statement from the American Heart Association. Hypertension 79:e128, 2022.

De Mast Q, Beutler JJ: The prevalence of atherosclerotic renal artery stenosis in risk groups: A systemic literature review. J Hypertens 27:1333, 2009.

Freedman BI, Cohen AH: Hypertension-attributed nephropathy: What's in a name? Nat Rev Nephrol 12:27, 2016.

Gornik HL et al: First International Consensus on the diagnosis and management of fibromuscular dysplasia. Vasc Med 24:164. 2019.

Herrmann SM et al: Management of atherosclerotic renovascular disease after Cardiovascular Outcomes in Renal Atherosclerotic Lesions (CORAL). Nephrol Dial Transplant 30:366, 2015.

Modi KS, Rao VK: Atheroembolic renal disease. J Am Soc Nephrol 12:1781 2001.

Parikh SA et al: SCAI expert consensus statement for renal artery stenting appropriate use. Catheter Cardiovasc Interv 84:1163, 2014.

Persu A et al: European consensus on the diagnosis and management of fibromuscular dysplasia. J Hypertens 32:1367, 2014.

Textor SC, Lerman LO: The role of hypoxia in ischemic chronic kidney disease. Semin Nephrol 39:589, 2019.

290

Deep-Venous Thrombosis and Pulmonary Thromboembolism

Samuel Z. Goldhaber

EPIDEMIOLOGY

Venous thromboembolism (VTE) encompasses deep-venous thrombosis (DVT) and pulmonary embolism (PE) and causes cardiovascular death, chronic disability, and emotional distress. Hospitalizations in the United States have decreased for acute myocardial infarction (MI), heart failure, and stroke among Medicare beneficiaries, but PE hospitalizations increased.

Pulmonary Embolism Mortality PE-related mortality in the United States decreased from 6 per 100,000 in 2000 and plateaued at approximately 4.5 per 100,000 in 2017. In contrast, Europe's age-standardized annual PE-related mortality rate has continued to decrease linearly since 2000. In the United States, PE-related mortality is increasing among young and middle-aged adults. Consequently, the median age at death from PE decreased from 73 years in 2000 to 68 years in 2018. During this period, the annual PE mortality was up to 50% higher in African Americans compared with Caucasians.

Socioeconomic Status To evaluate the relationship between socioeconomic status and PE, as well as other cardiovascular diseases, 10,942,483 Medicare beneficiaries were studied to determine whether they were hospitalized for PE, MI, heart failure, or stroke between 2003 and 2019. Hospitalizations for MI, heart failure, and stroke declined in socioeconomically disadvantaged and nondisadvantaged communities. In contrast, PE hospitalizations increased in both disadvantaged and nondisadvantaged communities. The heat map **(Fig. 290-1)** shows marked regional differences in the rates of PE hospitalization and degrees of social and community disadvantage throughout the United States. By 2019, 30-day mortality was similar between hospitalized beneficiaries from socioeconomically disadvantaged and nondisadvantaged communities for MI, heart failure, and ischemic stroke. In contrast, mortality among patients hospitalized with PE was consistently higher in disadvantaged areas **(Fig. 290-2)**. This finding may be due, in part, to poor outpatient follow-up and impeded access to anticoagulant medications.

Pulmonary Embolism Readmissions From 2010 to 2018, 721,710 fee-for-service Medicare beneficiaries were hospitalized with PE. The 30-day all-cause readmission rate was 11%, and the 90-day all-cause readmission rate was 15% **(Fig. 290-3)**. The most common reason for readmission was recurrent PE, accounting for about 11% of readmissions within 30 days.

COVID-19 In 2020, COVID-19 erupted and caused a global pandemic. The most notable clinical feature was a life-threatening acute respiratory syndrome requiring prolonged mechanical ventilation and resulting in a high case–fatality rate. This viral illness was associated with a high incidence of DVT and PE. At autopsy, about one-fourth of PE patients with COVID had both macrovascular and microvascular PE. Contributing etiologies included excessive inflammation with cytokine storm, platelet activation, endothelial dysfunction, and stasis **(Fig. 290-4)**.

Long-Term Sequelae of Pulmonary Embolism With respect to quality of life, about half of PE patients report persistent dyspnea, fatigue, and reduced exercise capacity, and about one-quarter have persistent right ventricular dysfunction on echocardiogram. This constellation of findings is being recognized more frequently and is called the "post-PE syndrome" **(Fig. 290-5)**. A less frequent complication of PE is the development of chronic thromboembolic pulmonary hypertension (CTEPH), which usually causes breathlessness, especially with exertion.

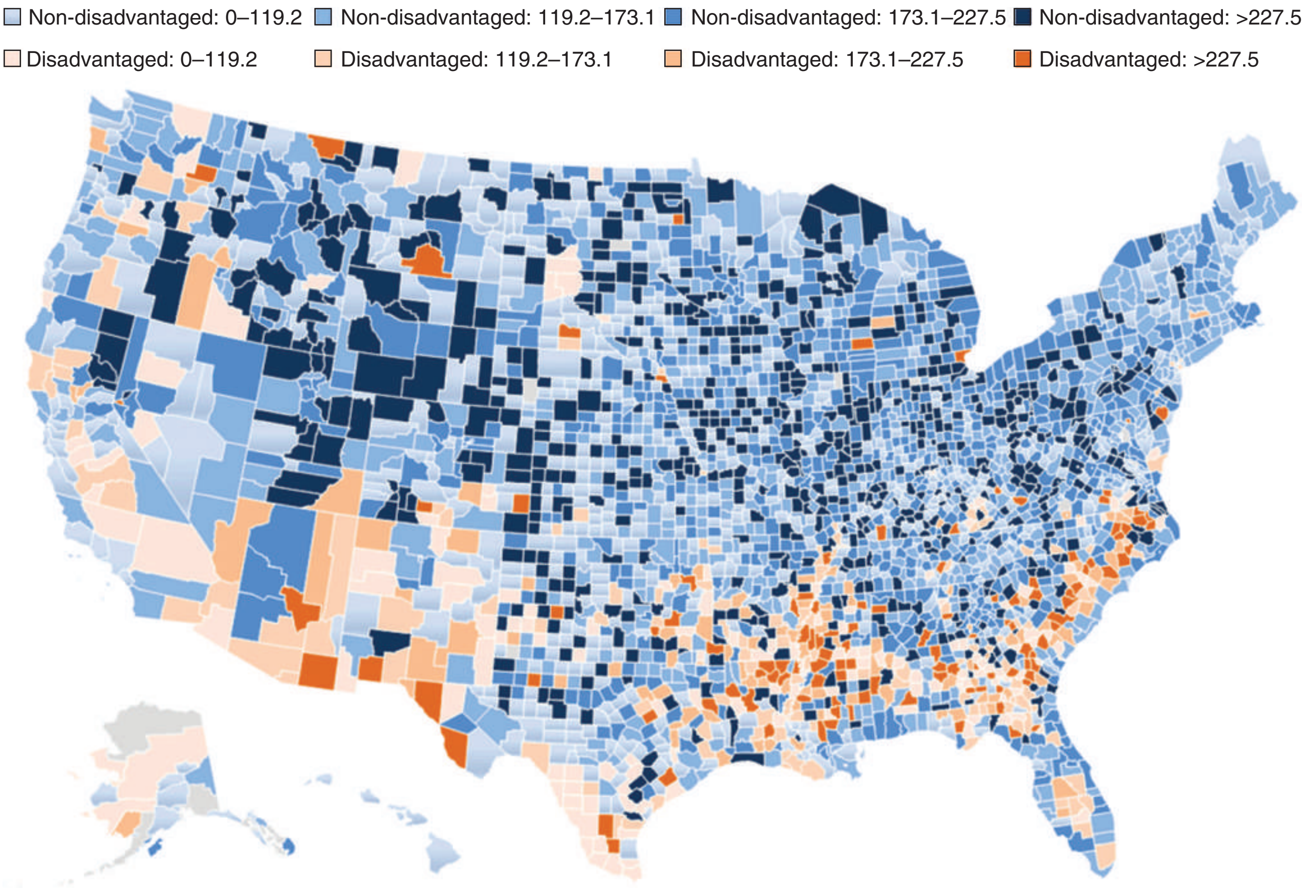
Non-disadvantaged: 0–119.2
Non-disadvantaged: 119.2–173.1
Non-disadvantaged: 173.1–227.5
Non-disadvantaged: >227.5
Disadvantaged: 0–119.2
Disadvantaged: 119.2–173.1
Disadvantaged: 173.1–227.5
Disadvantaged: >227.5

FIGURE 290-1 Observed hospitalizations for PE are shown per 100,000 Medicare beneficiaries aged 65 year or older residing in socioeconomically disadvantaged (*orange*) and nondisadvantaged (*blue*) counties of the United States in 2019. Lower hospitalization rates are represented with lighter shades of each color, and higher hospitalization rates are represented with darker shades. The range of hospitalizations per 100,000 represented by each shade of orange and each shade of blue are shown in the legend. *(Reproduced with permission from RK Wadhera et al: Circ Cardiovasc Qual Outcomes 17:e010090, 2024.)*

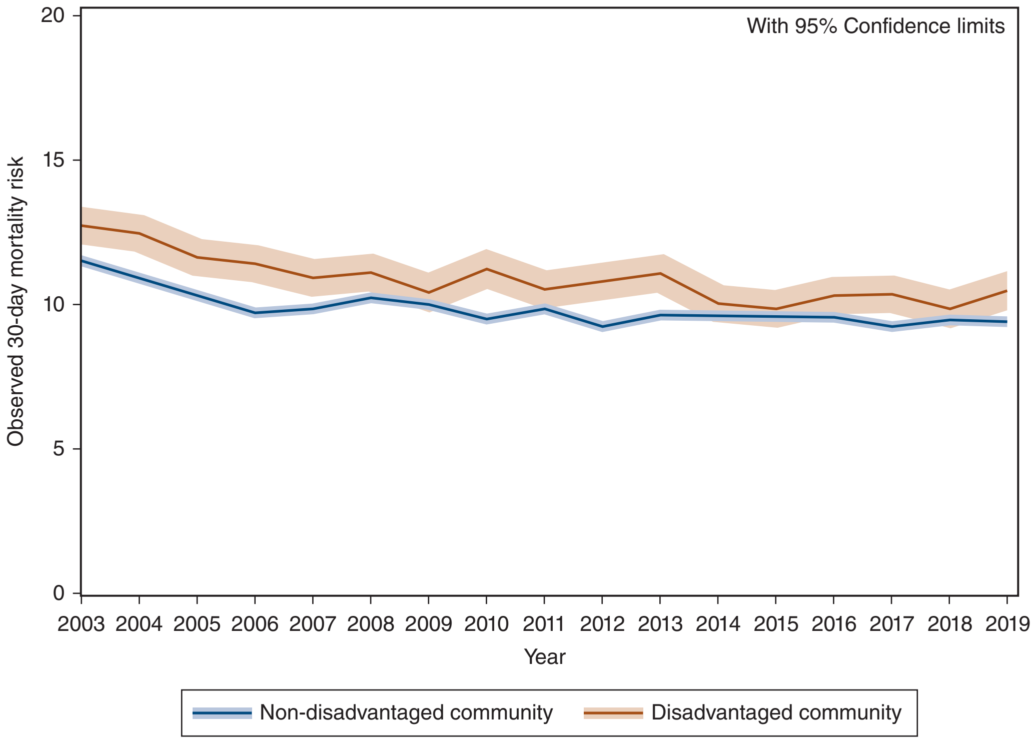

FIGURE 290-2 Trends in 30-day mortality among Medicare beneficiaries hospitalized with pulmonary embolism by community socioeconomic disadvantage from 2003 to 2019. The red lines represent beneficiaries from disadvantaged communities, and the blue line, beneficiaries from nondisadvantaged communities. The bands represent 95% confidence intervals for estimates. Trends in observed 30-day mortality among Medicare beneficiaries hospitalized with cardiovascular conditions by community socioeconomic disadvantage from 2003 to 2019. *(Reproduced with permission from RK Wadhera et al: Circ Cardiovasc Qual Outcomes 17:e010090, 2024.)*

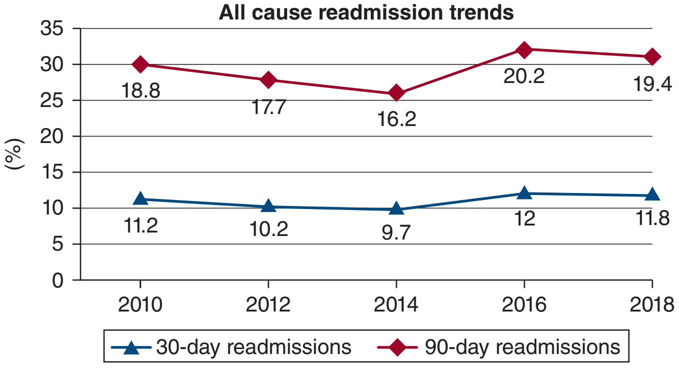

FIGURE 290-3 Graph showing all-cause readmission rates for pulmonary embolism (PE) patients within 30 days and within 90 days from 2010 to 2018. *(Reproduced from M Murthi: Am J Cardiol 184:133, 2022.)*

Postthrombotic Syndrome Postthrombotic syndrome (also known as *chronic venous insufficiency*) damages the venous valves of the leg and worsens quality of life by causing ankle or calf swelling and leg aching, especially after prolonged standing. In its most severe form, postthrombotic syndrome causes deep skin ulceration **(Fig. 290-6)**.

PATHOPHYSIOLOGY

Inflammation Inflammation takes center stage as a trigger of acute PE and DVT. Inflammation-related risk factors and medical illnesses are now linked as precipitants of VTE **(Table 290-1)**.

Prothrombotic States The two most common autosomal dominant genetic mutations are (1) factor V Leiden, which causes resistance to the endogenous anticoagulant, protein C, and (2) the prothrombin gene mutation, which increases the plasma prothrombin concentration **(Chaps. 69 and 122)**. Antithrombin, protein C, and protein S are naturally occurring coagulation inhibitors. Deficiencies of these inhibitors are associated with VTE but are rare. Antiphospholipid syndrome (APS) is an acquired (not genetic) thrombophilic disorder that predisposes to both venous and arterial thrombosis. Patients with APS often warrant lifetime anticoagulation, even if the initial VTE was provoked by trauma or surgery.

Clinical Risk Factors Common comorbid VTE risk factors include cancer, obesity, cigarette smoking, systemic arterial hypertension, chronic obstructive pulmonary disease, chronic kidney disease, long-haul air travel, air pollution, estrogen-containing contraceptives, pregnancy, the first 6–12 weeks postpartum, postmenopausal hormone replacement, surgery, and trauma. A sedentary lifestyle is an increasingly prevalent risk factor. A Japanese study found that each 2 hours per day increment of television watching is associated with a 40% increased likelihood of fatal PE.

Polygenic Risk Scores Ghouse and colleagues reported a genome-wide association study of VTE that incorporated 81,190 cases and 1,419,671 controls from six cohorts. They identified 93 risk loci, 62 of which were previously unreported. Many risk loci affected the coagulation cascade or platelet function. A VTE *polygenic risk score (PRS)* enabled identification of both high- and low-risk individuals. Individuals within the top 0.1% of PRS distribution had a VTE risk similar to homozygous or compound heterozygous carriers of factor V Leiden or the prothrombin gene mutation. In contrast, in the bottom 10% of the PRS distribution, factor V Leiden and prothrombin gene mutation carriers had a VTE risk similar to that of the general population. PRS improved individual risk prediction beyond that of genetic and clinical risk factors. Some risk factors for arterial thrombosis, such as body mass index and smoking, were concordant with VTE risk, whereas others, such as blood pressure and triglyceride levels, were discordant.

Activated Platelets Virchow's triad of venous stasis, hypercoagulability, and endothelial injury, usually coupled with an inflammatory trigger, leads to recruitment of activated platelets, which release microparticles. These microparticles contain proinflammatory mediators

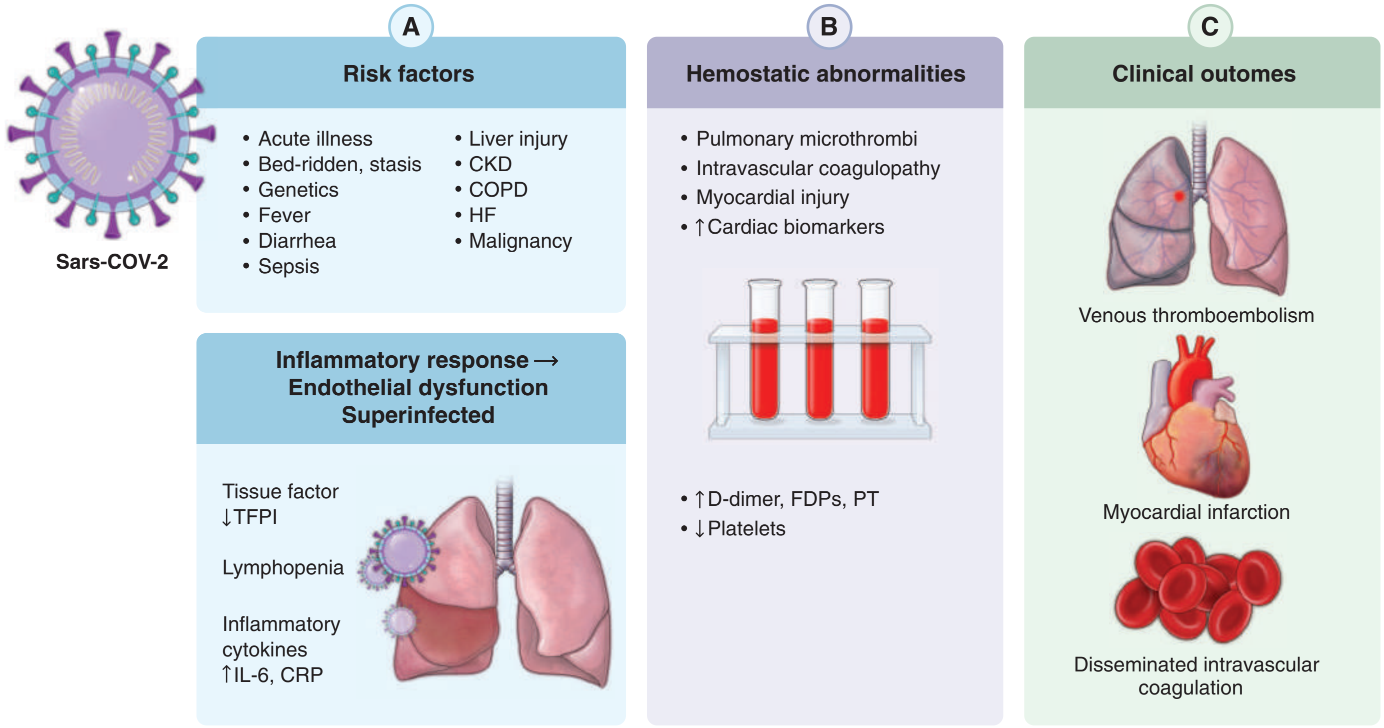
A
Risk factors
• Acute illness
• Bed-ridden, stasis
• Genetics
• Fever
• Diarrhea
• Sepsis
• Liver injury
• CKD
• COPD
• HF
• Malignancy
Sars-COV-2
Inflammatory response ⟶
Endothelial dysfunction
Superinfected
Tissue factor
↓TFPI
Lymphopenia
Inflammatory
cytokines
↑IL-6, CRP
B
Hemostatic abnormalities
• Pulmonary microthrombi
• Intravascular coagulopathy
• Myocardial injury
• ↑Cardiac biomarkers
• ↑D-dimer, FDPs, PT
• ↓Platelets
C
Clinical outomes
Venous thromboembolism
Myocardial infarction
Disseminated intravascular
coagulation

FIGURE 290-4 Postulated mechanisms of coagulopathy and pathogenesis of thrombosis in COVID-19. ***A.*** Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection activates an inflammatory response, leading to release of inflammatory mediators. Endothelial and hemostatic activation ensues, with decreased levels of TFPI and increased tissue factor. The inflammatory response to severe infection is marked by lymphopenia and thrombocytopenia. Liver injury may lead to decreased coagulation and antithrombin formation. ***B.*** COVID-19 may be associated with hemostatic derangement and elevated troponin. ***C.*** Increased thromboembolic state results in venous thromboembolism, myocardial infarction, or, in case of further hemostatic derangement, disseminated intravascular coagulation. CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; FDP, fibrin degradation product; HF, heart failure; IL, interleukin; LDH, lactate dehydrogenase; PT, prothrombin time; TFPI, tissue factor pathway inhibitor. *(Reproduced with permission from B Bikdeli et al: COVID-19 and thrombotic or thromboembolic disease: Implications for prevention, antithrombotic therapy, and follow-up. J Am Coll Cardiol 75:2950, 2020.)*

PVFS scale grade		Description
0	No functional limitations	All usual duties/activities at home or at work can be carried out at the same level of intensity. Symptoms, pain and anxiety are absent.
1	Negligible functional limitations	All usual duties/activities at home or at work can be carried out at the same level of intensity, despite some symptoms, pain, or anxiety.
2	Slight functional limitations	Some usual duties/activities at home or at work are carried out at a lower level of intensity or are occasionally avoided due to symptoms, pain, or anxiety.
3	Moderate functional limitations	Usual duties/activities at home or at work have been structurally modified (reduced) due to symptoms, pain, or anxiety.
4	Severe functional limitations	Assistance needed in activities of daily living due to symptoms, pain, or anxiety: nursing care and attention are required.
D	Death	Death occurred before the scheduled assessment.

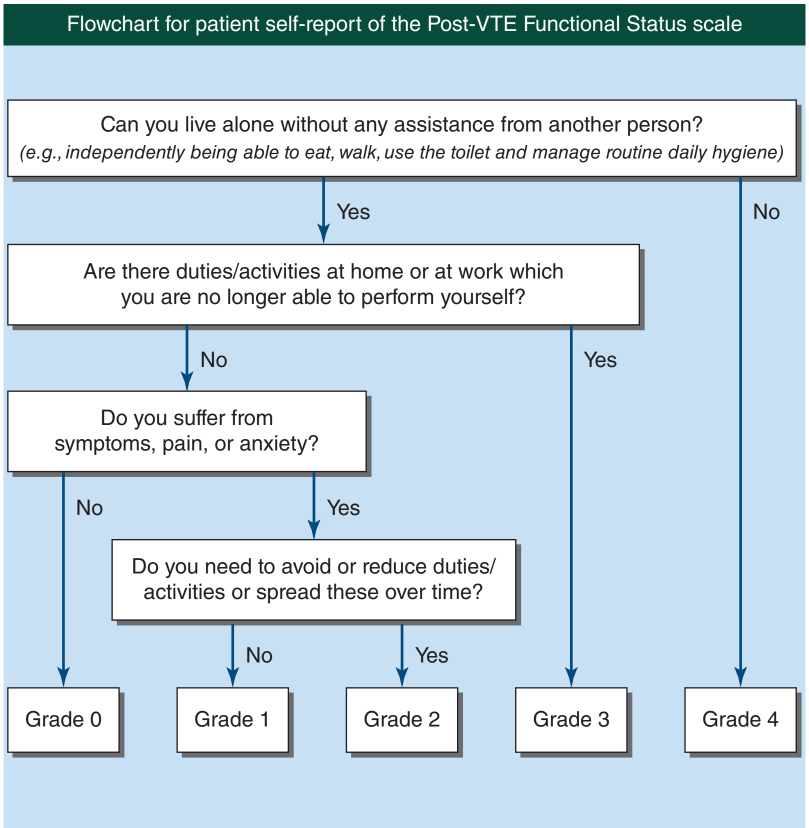
Flowchart for patient self-report of the Post-VTE Functional Status scale
Can you live alone without any assistance from another person?
(e.g., independently being able to eat, walk, use the toilet and manage routine daily hygiene)
Yes
No
Are there duties/activities at home or at work which you are no longer able to perform yourself?
No
Yes
Do you suffer from symptoms, pain, or anxiety?
No
Yes
Do you need to avoid or reduce duties/activities or spread these over time?
No
Yes
Grade 0
Grade 1
Grade 2
Grade 3
Grade 4

FIGURE 290-5 Flow chart for patient self-report of the Post-Venous Thromboembolism (VTE) Functional Status scale. (*Reproduced with permission from GJAM Boon et al: Thromb Res 190:45, 2020, Figure 2.)*

that bind neutrophils, stimulating them to release their nuclear material and form web-like extracellular networks called *neutrophil extracellular traps*. These prothrombotic networks contain histones that stimulate platelet aggregation and promote platelet-dependent thrombin generation. Venous thrombi form and flourish in this environment of stasis, low oxygen tension, and upregulation of proinflammatory genes.

Interaction Between Venous Thromboembolism and Atherothrombosis VTE (red clot) and arterial thrombotic (white clot) events were previously considered separate entities. However, the presence of carotid artery plaque is associated with double the risk of incident VTE. This observation led to the discovery of a broad interaction of VTE with acute coronary syndrome and with acute ischemic stroke **(Fig. 290-7)**. These three vascular conditions share similar risk factors and similar pathophysiology: inflammation, hypercoagulability, and endothelial injury. Patients who suffer VTE are more than twice

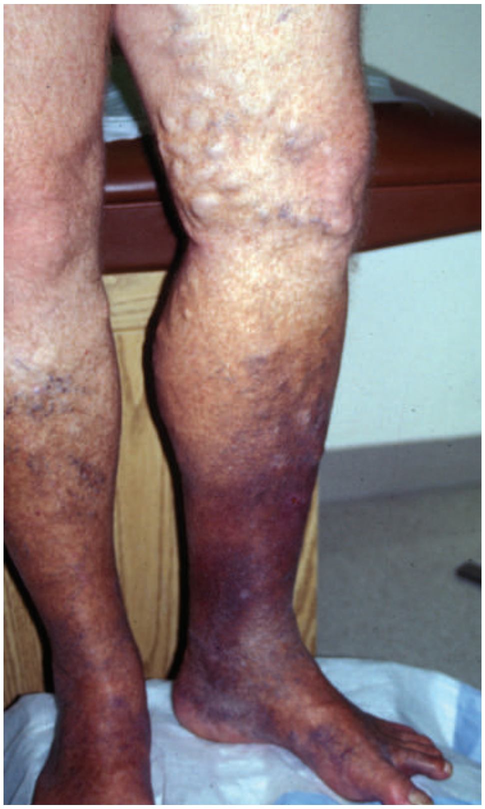

FIGURE 290-6 Skin ulceration in the medial malleolus from postthrombotic syndrome of the leg.

TABLE 290-1 Inflammation-Linked Conditions That Can Trigger PE or DVT

Ulcerative colitis
Crohn's disease
Rheumatoid arthritis
Psoriasis
Diabetes mellitus, type 2
Obesity/metabolic syndrome
Hypercholesterolemia, especially elevated LDL cholesterol
Lipoprotein(a)
Pneumonia
Acute coronary syndrome
Acute stroke
Cigarette smoking
Sepsis/septic shock
Erythropoiesis-stimulating agents
Blood transfusion
Cancer

Abbreviations: DVT, deep-venous thrombosis; LDL, low-density lipoprotein; PE, pulmonary embolism.

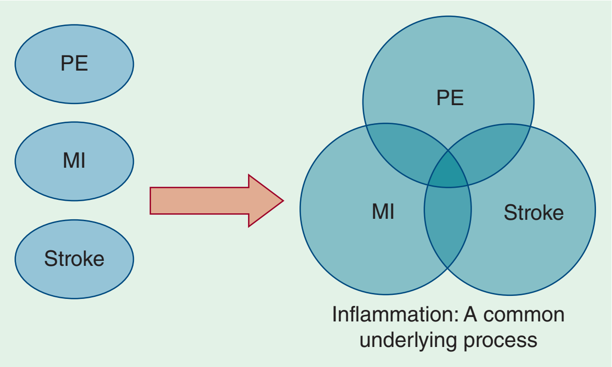

FIGURE 290-7 Broad interaction between venous thromboembolism and atherothrombosis. MI, myocardial infarction; PE, pulmonary embolism.

as likely to have a future MI or stroke. Conversely, patients with MI or stroke are more than twice as likely to suffer a future VTE.

Embolization When DVTs **(Fig. 290-8)** detach from their site of formation, they embolize to the vena cava, right atrium, and right ventricle and lodge in the pulmonary arterial circulation, thereby causing acute PE. Many patients with PE have no evidence of DVT because the leg thrombus has already embolized to the lungs. Paradoxically, these thrombi occasionally embolize to the arterial circulation through a patent foramen ovale or atrial septal defect.

Physiology The most common gas exchange abnormalities are arterial hypoxemia and an increased alveolar-arterial O_2 tension gradient, which represent the inefficiency of oxygen transfer across the lungs. Anatomic dead space increases because breathed gas does not enter gas exchange units of the lung. Physiologic dead space increases because ventilation to gas exchange units exceeds venous blood flow through the pulmonary capillaries **(Fig. 290-9)**.

Other pathophysiologic abnormalities include:

1. *Increased pulmonary vascular resistance* due to vascular obstruction or platelet secretion of vasoconstricting neurohumoral agents such as serotonin. Release of vasoactive mediators can produce ventilation-perfusion mismatching at sites remote from the embolus, thereby accounting for discordance between a small PE and a large alveolar-arterial O_2 gradient.
2. *Impaired gas exchange* due to increased alveolar dead space from vascular obstruction, hypoxemia from alveolar hypoventilation relative to perfusion in the nonobstructed lung, right-to-left shunting, or impaired carbon monoxide transfer due to loss of gas exchange surface.
3. *Alveolar hyperventilation* due to reflex stimulation of irritant receptors.

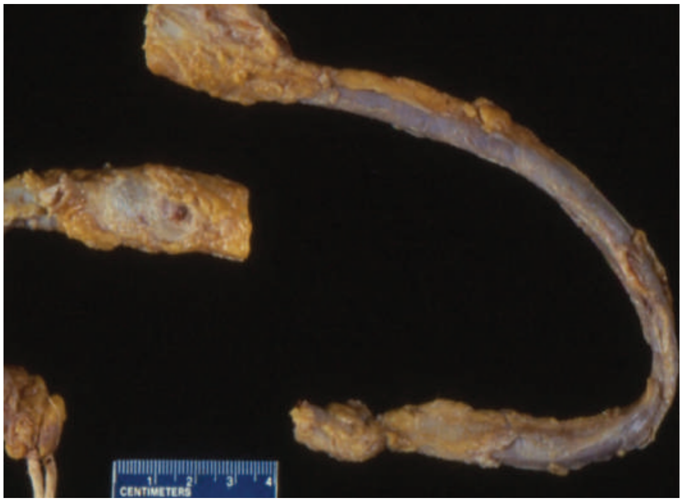

FIGURE 290-8 Deep-venous thrombosis at autopsy.

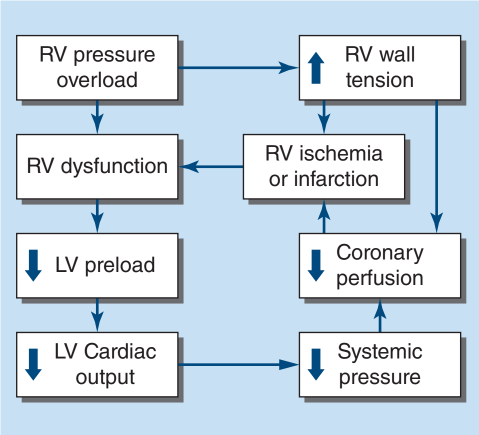

FIGURE 290-9 Pathophysiology of pulmonary embolism (PE). LV, left ventricular; RV, right ventricular.

4. *Increased airway resistance* due to constriction of airways distal to the bronchi.
5. *Decreased pulmonary compliance* due to lung edema, lung hemorrhage, or loss of surfactant.

Pulmonary Hypertension, Right Ventricular (RV) Dysfunction, and RV Microinfarction Pulmonary artery obstruction and neurohumoral mediators cause a rise in both pulmonary artery pressure and pulmonary vascular resistance. When RV wall tension rises, RV dilation, stretch, and dysfunction ensue, with release of the cardiac biomarker brain natriuretic peptide. The interventricular septum bulges into and compresses an intrinsically normal left ventricle (LV). Diastolic LV dysfunction reduces LV distensibility and impairs LV filling. Increased RV wall tension also compresses the right coronary artery, limits myocardial oxygen supply, and precipitates right coronary artery ischemia and RV microinfarction, with release of cardiac biomarkers such as troponin. Underfilling of the LV may lead to a fall in LV cardiac output and systemic arterial pressure, with consequent circulatory collapse and death (Fig. 290-9).

CLASSIFICATION OF PULMONARY EMBOLISM AND DEEP-VENOUS THROMBOSIS

Pulmonary Embolism *Massive (high-risk) PE* accounts for 5–10% of cases and is usually characterized by systemic arterial hypotension and extensive thrombosis affecting at least half of the pulmonary vasculature. Dyspnea, syncope, hypotension, and cyanosis are hallmarks of massive PE. Patients with massive PE may present in cardiogenic shock and can die from multisystem organ failure. *Submassive (intermediate-risk) PE* accounts for 20–25% of patients and is characterized by RV dysfunction despite normal systemic arterial pressure. The combination of right heart failure and release of cardiac biomarkers such as troponin indicates a high risk of clinical deterioration. *Low-risk PE* constitutes about 65–75% of cases. These patients have an excellent prognosis.

Deep-Venous Thrombosis *Lower extremity DVT* usually begins in the calf and can propagate proximally to the popliteal, femoral, and iliac veins. Leg DVT is ~10 times more common than *upper extremity DVT,* which is often precipitated by placement of pacemakers, internal cardiac defibrillators, or indwelling central venous catheters. The likelihood of upper extremity DVT increases as the catheter diameter and number of lumens increase.

Superficial venous thrombosis usually presents with erythema, tenderness, and a "palpable cord"; it is most common in the lower extremities because of varicose veins and in the upper extremities because of indwelling intravenous catheters. Superficial veins may be visible in the chest wall (*Urschel's sign*). Patients are at risk for extension of the superficial vein thrombosis to the deep-venous system.

DIAGNOSIS

Clinical Evaluation PE is known as "the great masquerader." Diagnosis is difficult because symptoms and signs are nonspecific.

The most common symptom of PE is unexplained breathlessness. When occult PE occurs concomitantly with overt congestive heart failure or pneumonia, clinical improvement often fails to ensue despite standard medical treatment of the concomitant illness. This scenario provides a clinical clue to the possible coexistence of PE.

With DVT, the most common symptom is a cramp or "charley horse" in the lower calf that persists and intensifies over several days. Wells Point Score criteria, validated in ambulatory patients but not in hospitalized patients, help estimate the clinical likelihood of DVT and PE **(Table 290-2)**. Patients with a low likelihood of DVT or a low-to-moderate likelihood of PE should undergo initial diagnostic evaluation with D-dimer testing alone (see "Blood Tests"), without obligatory imaging tests if the D-dimer test result is normal **(Fig. 290-10)**. However, patients with a high clinical likelihood of VTE should skip D-dimer testing and undergo imaging as the next step in the diagnostic algorithm.

Clinical Pearls Not all leg pain is due to DVT, and not all dyspnea is due to PE **(Table 290-3)**. Sudden, severe calf discomfort suggests a ruptured Baker's cyst. Fever and chills often herald cellulitis rather than DVT. Physical findings, if present, may consist only of mild palpation discomfort in the lower calf. However, massive DVT often presents with marked thigh swelling, tenderness, and erythema. If a leg is diffusely edematous, DVT is unlikely, except for the most serious form of DVT, *phlegmasia cerulea dolens*. Recurrent left thigh edema, especially in young women, raises the possibility of May-Thurner syndrome, with right proximal iliac artery compression of the left proximal iliac vein. This syndrome may be associated with a DVT. Upper extremity venous thrombosis may present with asymmetry in the supraclavicular fossa or in the circumference of the upper arms.

Pulmonary infarction usually indicates a small PE. This condition is exquisitely painful because the thrombus lodges peripherally, near the innervation of pleural nerves. *Nonthrombotic PE* etiologies include fat embolism after pelvic or long bone fracture, tumor embolism, bone marrow, and air embolism. Cement embolism and bony fragment embolism can occur after total hip or knee replacement. Intravenous drug users may inject themselves with a wide array of substances that can embolize, such as hair, talc, and cotton. *Amniotic fluid embolism* occurs when fetal membranes leak or tear at the placental margin.

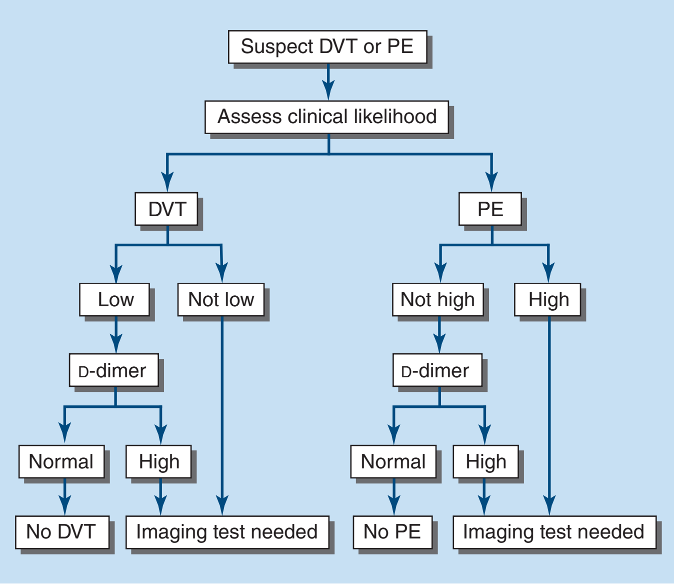

FIGURE 290-10 How to decide whether diagnostic imaging is needed. For assessment of clinical likelihood, see Table 290-2. DVT, deep-venous thrombosis; PE, pulmonary embolism.

Nonimaging Diagnostic Modalities • **BLOOD TESTS** The quantitative *plasma* D-*dimer enzyme-linked immunosorbent assay (ELISA)* rises in the presence of DVT or PE because of the breakdown of fibrin by plasmin. Elevation of D-dimer indicates endogenous, although often clinically ineffective, thrombolysis. The sensitivity of an abnormally elevated D-dimer is >95% for PE. A normal D-dimer is a useful "rule out" test for PE in patients with low pretest probability. However, the D-dimer assay is not specific. Levels increase in patients with MI, pneumonia, sepsis, cancer, and the postoperative state, and those in the second or third trimester of pregnancy. Therefore, D-dimer testing rarely has a useful role among hospitalized patients because levels are frequently elevated due to systemic illness.

The upper limit of normal for D-dimer was considered to be 500 ng/mL; however, guidelines now recommend use of an age-adjusted D-dimer when ruling out acute PE. The age-adjusted D-dimer applies to patients older than 50 years of age with low or intermediate clinical probability of PE. To calculate the upper limit of normal D-dimer in these patients, multiply the age by 10. For example, a 70-year-old patient suspected of PE would have 700 ng/mL as the upper limit of normal for D-dimer. Implementing routine use of the

TABLE 290-2 Clinical Decision Rules

Low Clinical Likelihood of DVT if Point Score Is Zero or Less; Moderate Likelihood if Score Is 1 to 2; High Likelihood if Score Is 3 or Greater	
CLINICAL VARIABLE	**DVT SCORE**
Active cancer	1
Paralysis, paresis, or recent cast	1
Bedridden for >3 days; major surgery <12 weeks	1
Tenderness along distribution of deep veins	1
Entire leg swelling	1
Unilateral calf swelling >3 cm	1
Pitting edema	1
Collateral superficial nonvaricose veins	1
Alternative diagnosis at least as likely as DVT	–2
High Clinical Likelihood of PE if Point Score Exceeds 4	
CLINICAL VARIABLE	**PE SCORE**
Signs and symptoms of DVT	3.0
Alternative diagnosis less likely than PE	3.0
Heart rate >100/min	1.5
Immobilization >3 days; surgery within 4 weeks	1.5
Prior PE or DVT	1.5
Hemoptysis	1.0
Cancer	1.0

Abbreviations: DVT, deep-venous thrombosis; PE, pulmonary embolism.

TABLE 290-3 Differential Diagnosis of DVT and PE

DVT
Ruptured Baker's cyst
Muscle strain/injury
Cellulitis
Acute postthrombotic syndrome/venous insufficiency
PE
Pneumonia, asthma, chronic obstructive pulmonary disease
Congestive heart failure
Pericarditis
Pleurisy: "viral syndrome," costochondritis, musculoskeletal discomfort
Rib fracture, pneumothorax
Acute coronary syndrome
Anxiety
Vasovagal syncope

Abbreviations: DVT, deep-venous thrombosis; PE, pulmonary embolism.

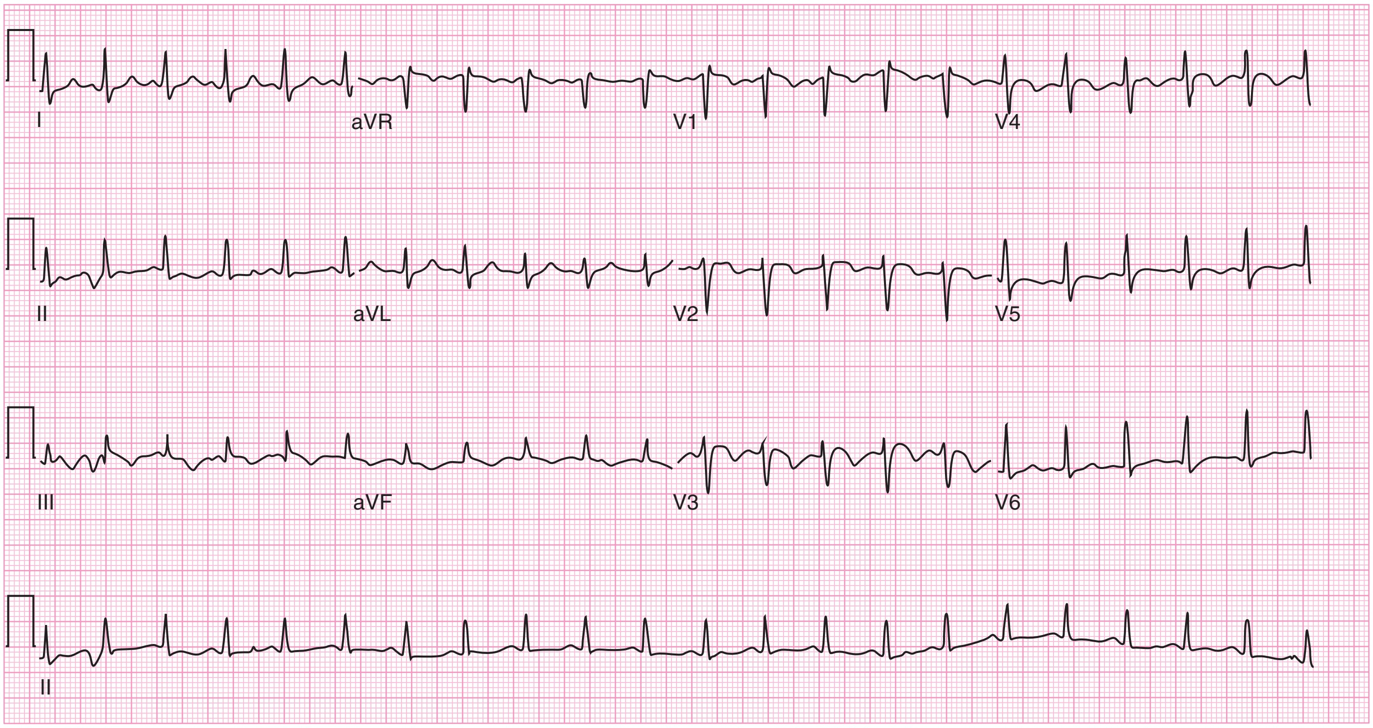

FIGURE 290-11 Electrocardiogram with both the S1Q3T3 sign and T-wave inversions in leads V_1-V_4—typical of an anatomically large pulmonary embolism. This patient's CT pulmonary angiogram is shown as **Fig. 290-13*A* and *B***.

age-adjusted D-dimer reduces the number of computed tomography (CT) pulmonary angiograms that are ordered. However, the age-adjusted D-dimer does not apply to patients suspected of acute DVT.

ELEVATED CARDIAC BIOMARKERS Serum troponin increases because of RV microinfarction. Myocardial stretch causes release of brain natriuretic peptide or NT-pro-brain natriuretic peptide.

ELECTROCARDIOGRAM The most frequently cited abnormality, in addition to sinus tachycardia, is the S1Q3T3 sign: an S wave in lead I, a Q wave in lead III, and an inverted T wave in lead III **(Chap. 247)**. This finding is relatively specific but insensitive. RV strain and ischemia cause the most common abnormality, T-wave inversion in leads V_1 to V_4 **(Fig. 290-11)**.

Noninvasive Imaging Modalities • **VENOUS ULTRASONOGRAPHY** Ultrasonography of the deep-venous system relies on loss of vein compressibility as the primary diagnostic criterion for DVT. When a normal vein is imaged in cross-section, it readily collapses with gentle manual pressure on the ultrasound transducer. This creates the illusion of a "wink." With acute DVT, the vein loses its compressibility because of passive distention by acute thrombus. The diagnosis of acute DVT is even more secure when thrombus is directly visualized; it appears homogeneous, is located in the center of the vein, and has low echogenicity **(Fig. 290-12)**. The vein itself often appears mildly dilated, and collateral channels may be absent.

Venous flow dynamics can be examined with Doppler imaging. Normally, manual calf compression causes augmentation of the Doppler flow pattern. Loss of normal respiratory variation is caused by an obstructing DVT or by any obstructive process within the pelvis. For patients with a technically poor or nondiagnostic venous ultrasound, consider alternative imaging modalities such as CT and magnetic resonance imaging.

CHEST ROENTGENOGRAPHY A normal or nearly normal chest x-ray often occurs in PE. Well-established abnormalities include focal oligemia (*Westermark's sign*), a peripheral triangular-shaped density usually located at the pleural base (*Hampton's hump*—which is usually due to pulmonary infarction), and an enlarged right descending pulmonary artery (*Palla's sign*).

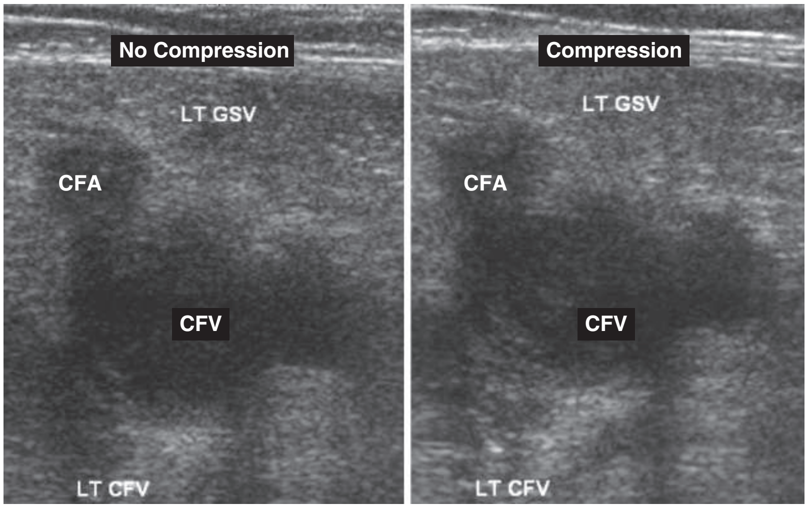

FIGURE 290-12 Venous ultrasound, with and without compression of the leg veins. CFA, common femoral artery; CFV, common femoral vein; GSV, great saphenous vein; LT, left.

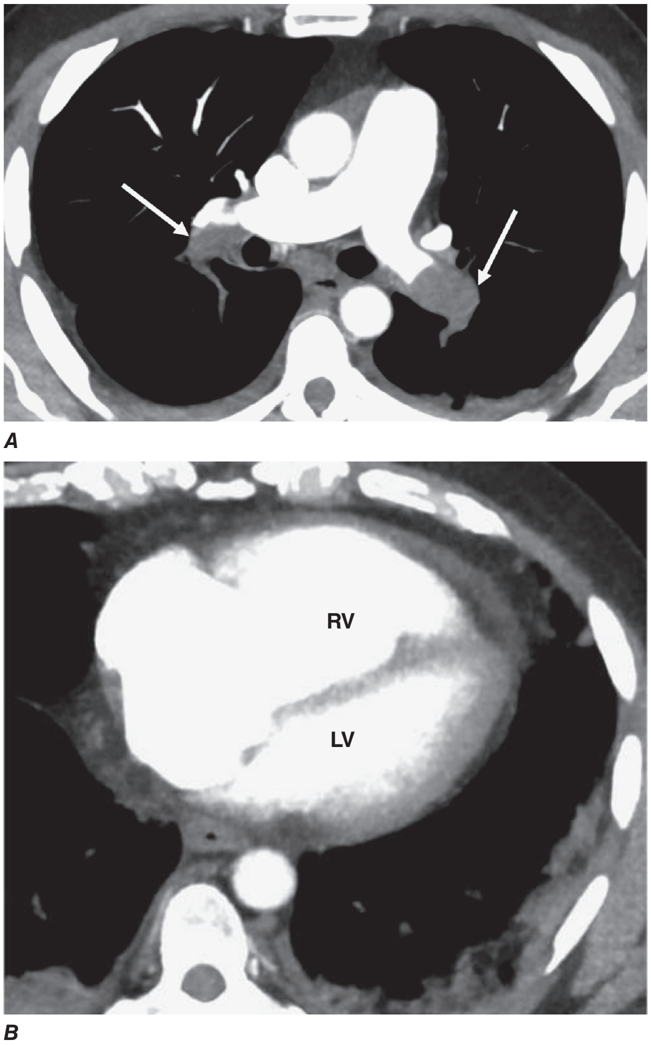

FIGURE 290-13 ***A.*** Massive bilateral proximal pulmonary embolism on an axial chest CT image in a 53-year-old man (whose electrocardiogram is shown in Fig. 290-11) with filling defects in the right and left main pulmonary arteries (*white arrows*). ***B.*** Four-chamber view in the same patient showing the right ventricle (RV) larger than the left ventricle (LV).

CHEST CT CT pulmonary angiography with intravenous contrast is the principal imaging test for the diagnosis of PE **(Fig. 290-13*A*)**. Thin-cut chest CT images can provide exquisite detail, with ≤1 mm of resolution during a short breath hold. Sixth-order branches can be visualized with resolution superior to that of conventional invasive contrast pulmonary angiography. The CT scan also provides an excellent four-chamber view of the heart **(Fig. 290-13*B*)**. RV enlargement on chest CT indicates an increased likelihood of death within the next 30 days compared with PE patients who have normal RV size (commonly defined as a right to left ventricular diameter ratio >1). In patients without PE, the lung parenchymal images may establish alternative diagnoses not apparent on chest x-ray that explain the presenting symptoms and signs, such as pneumonia, emphysema, pulmonary fibrosis, pulmonary mass, and aortic pathology.

LUNG SCANNING Ventilation/perfusion lung scanning has become a second-line diagnostic test for PE, used mostly for patients who cannot tolerate intravenous contrast. Small particulate aggregates of albumin labeled with a gamma-emitting radionuclide are injected intravenously and trapped in the pulmonary capillary bed. The perfusion scan defect indicates absent or decreased blood flow, possibly due to PE. Ventilation scans, obtained with a radiolabeled inhaled gas such as xenon or krypton, improve the specificity of the perfusion scan. Abnormal ventilation scans indicate abnormal nonventilated lung, thereby providing possible explanations for perfusion defects other than acute PE, such as asthma or chronic obstructive pulmonary disease. A high-probability scan for PE is defined as two or more segmental perfusion defects in the presence of normal ventilation. The diagnosis of PE is very unlikely in patients with normal and nearly normal scans and, in contrast, is ~90% certain in patients with high-probability scans.

MAGNETIC RESONANCE (MR) (CONTRAST-ENHANCED) IMAGING When pelvic or leg ultrasound is equivocal, MR venography with gadolinium contrast is an excellent imaging modality to diagnose DVT. MR pulmonary angiography may detect large proximal PE but is not reliable for smaller segmental and subsegmental PE.

ECHOCARDIOGRAPHY Echocardiography is *not* a reliable diagnostic imaging tool for acute PE because most patients with PE have normal echocardiograms. However, echocardiography is very useful for detecting conditions that may mimic PE, such as acute MI, pericardial tamponade, and aortic dissection. Transthoracic echocardiography rarely images thrombus directly in the main pulmonary artery or the right or left main branches. The best-known indirect sign of PE on transthoracic echocardiography is McConnell's sign: hypokinesis of the RV free wall with normal or hyperkinetic motion of the RV apex.

Invasive Diagnostic Modalities • **PULMONARY ANGIOGRAPHY** Chest CT pulmonary angiography with contrast (see above) has virtually replaced invasive pulmonary angiography as a diagnostic test. Invasive catheter-based diagnostic testing is reserved for patients with technically unsatisfactory chest CTs and for those in whom an interventional procedure such as catheter-directed thrombolysis or embolectomy is planned. A definitive diagnosis of PE requires visualization of an intraluminal filling defect in more than one projection. Secondary signs of PE include abrupt occlusion ("cut-off") of vessels, segmental oligemia, avascularity, and a prolonged arterial phase with slow filling and tortuous, tapering peripheral vessels.

CONTRAST PHLEBOGRAPHY Venous ultrasonography has virtually replaced contrast phlebography as the principal diagnostic test for suspected DVT. However, contrast phlebography is used when an interventional procedure is planned.

Integrated Diagnostic Approach An integrated diagnostic approach streamlines the workup of suspected DVT and PE **(Fig. 290-14)**.

TREATMENT

Deep-Venous Thrombosis

UPPER EXTREMITY DVT

As peripherally inserted central catheter (PICC) use has increased, so has the rate of upper extremity DVT. This rate can be decreased by more judicious selection of patients who require a PICC, use of single-lumen rather than double- or triple-lumen PICCs, and use of the smallest possible lumen size, ideally 4 French rather than 5 or 6 French.

ISOLATED CALF DVT

The GARFIELD-VTE Registry recruited 2145 patients with isolated calf DVT and 3846 patients with proximal DVT with or without calf DVT. Isolated calf DVT patients were more likely to have either undergone surgery or have experienced leg trauma, and they were less likely to have active cancer or a prior history of VTE. Almost all isolated calf DVT patients received anticoagulation, and nearly half were anticoagulated for at least 1 year.

FIBRINOLYSIS

If advanced therapy beyond anticoagulation is warranted, one can utilize clot dissolution or clot extraction therapy with catheter-directed therapy. The open vein hypothesis postulates that patients who receive primary therapy will sustain less long-term damage to venous valves, with consequent lower rates of postthrombotic syndrome. However, the ATTRACT trial randomized 692 patients with femoral or iliofemoral DVT to catheter-directed therapy plus

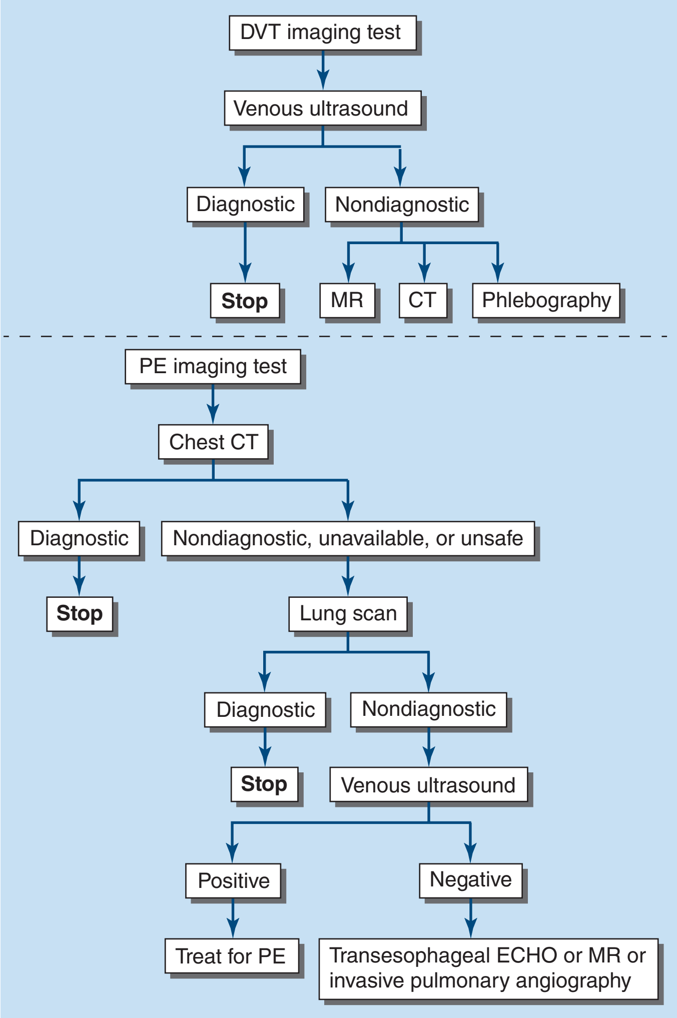
DVT imaging test
Venous ultrasound
Diagnostic
Nondiagnostic
Stop
MR
CT
Phlebography
PE imaging test
Chest CT
Diagnostic
Nondiagnostic, unavailable, or unsafe
Stop
Lung scan
Diagnostic
Nondiagnostic
Stop
Venous ultrasound
Positive
Negative
Treat for PE
Transesophageal ECHO or MR or invasive pulmonary angiography

FIGURE 290-14 Integrated diagnostic approach. CT, computed tomography; DVT, deep-venous thrombosis; ECHO, echocardiography; MR, magnetic resonance; PE, pulmonary embolism.

anticoagulation versus usual care with anticoagulation alone. After 2 years of follow-up, there was no overall reduction in postthrombotic syndrome in the thrombolysis group. However, there was a trend toward less postthrombotic syndrome among patients with iliofemoral DVT (compared with only femoral DVT) who received catheter-directed thrombolysis compared with anticoagulation alone.

COMPRESSION STOCKINGS

For patients with swelling of the legs when acute DVT is diagnosed, below-knee graduated compression stockings may be prescribed, usually 30–40 mmHg or 20–30 mmHg, to lessen patient discomfort. They should be replaced every 3–6 months because they lose their elasticity. However, prescription of vascular compression stockings in asymptomatic newly diagnosed acute DVT patients does not prevent the development of postthrombotic syndrome.

TREATMENT

Pulmonary Embolism

RISK STRATIFICATION

Hemodynamic instability, RV dysfunction on echocardiography, RV enlargement on chest CT, and elevation of the troponin level due to RV microinfarction portend a high risk of adverse clinical outcomes despite anticoagulation. However, in most cases, RV

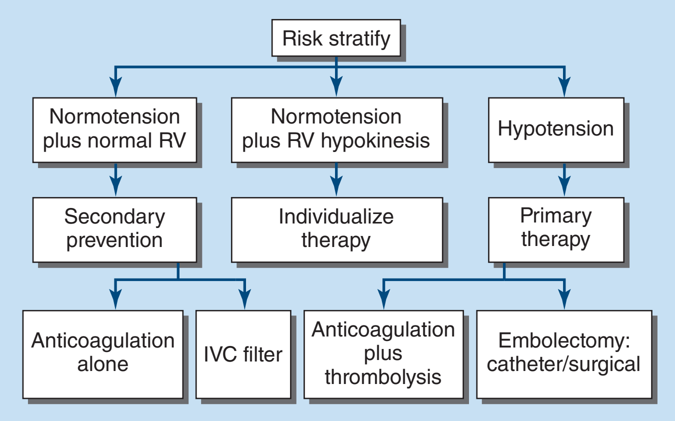
Risk stratify
Normotension plus normal RV
Normotension plus RV hypokinesis
Hypotension
Secondary prevention
Individualize therapy
Primary therapy
Anticoagulation alone
IVC filter
Anticoagulation plus thrombolysis
Embolectomy: catheter/surgical

FIGURE 290-15 Acute management of pulmonary thromboembolism. IVC, inferior vena cava; PE, pulmonary embolism; RV, right ventricular.

function and hemodynamic function remain normal, which portend an excellent prognosis and clinical outcome **(Fig. 290-15)**.

ANTICOAGULATION

There are three major strategies for anticoagulation: (1) the well-established—but waning—strategy of parenteral anticoagulation with unfractionated heparin (UFH), low-molecular-weight heparin (LMWH), or fondaparinux either as monotherapy or more often "bridged" to warfarin; (2) parenteral anticoagulation, switched after 5 days to dabigatran (a direct thrombin inhibitor) or edoxaban (an anti-Xa agent); or (3) oral anticoagulation monotherapy with rivaroxaban or apixaban (both anti-Xa agents) with a 3-week or 1-week loading dose, respectively. For patients with VTE in the setting of suspected or proven heparin-induced thrombocytopenia, one can choose between two intravenous direct thrombin inhibitors: argatroban (metabolized by the liver) or bivalirudin (metabolized by the kidney), or subcutaneously administered fondaparinux **(Table 290-4)**.

Unfractionated Heparin UFH binds to and accelerates the activity of antithrombin, thus preventing additional thrombus formation. Intravenous UFH is typically dosed to achieve a target activated partial thromboplastin time (aPTT) of 60–80 s, which usually corresponds to a chromogenic anti-Xa level of 0.3–0.7 U/mL. When switching from a direct oral anticoagulant (DOAC) to intravenous UFH, monitor the aPTT, not the chromogenic anti-Xa assay. However, the anti-Xa assay is more accurate than the aPTT in critically ill patients, APS, pregnancy, and suspected heparin resistance.

In patients with normal liver function, use an initial bolus of 80 U/kg as a loading dose of UFH, followed by an initial infusion rate of 18 U/kg per h in patients with normal liver function. The short half-life of UFH is especially useful in patients in whom hour-to-hour control of anticoagulation intensity is desired. Heparin also has pleiotropic effects that may decrease systemic and local inflammation.

Low-Molecular-Weight Heparins These fragments of UFH exhibit less binding to plasma proteins and endothelial cells and consequently have greater bioavailability, a more predictable dose response, and a longer half-life than UFH. No monitoring or dose adjustment is needed unless the patient is markedly obese or has chronic kidney disease.

Fondaparinux Fondaparinux, an anti-Xa pentasaccharide, is essentially an ultra-low-molecular-weight heparin. It is administered as a weight-based, once-daily subcutaneous injection in a prefilled syringe. No laboratory monitoring is required. Fondaparinux is synthesized in a laboratory and, unlike LMWH or UFH, is not derived from animal products. It does not cause heparin-induced thrombocytopenia. For patients with renal dysfunction, the fondaparinux dose must be adjusted downward.

TABLE 290-4 Anticoagulation of VTE

Non-Warfarin Anticoagulation
Unfractionated heparin, bolus and continuous infusion, to achieve aPTT 2–3 times the upper limit of the laboratory normal, *or*
Enoxaparin 1 mg/kg twice daily with normal renal function, *or*
Dalteparin 200 U/kg once daily or 100 U/kg twice daily, with normal renal function, *or*
Tinzaparin 175 U/kg once daily with normal renal function, *or*
Fondaparinux weight-based once daily; adjust for impaired renal function
Direct thrombin inhibitors: argatroban or bivalirudin (with suspected or proven heparin-induced thrombocytopenia)
Rivaroxaban 15 mg twice daily for 3 weeks, followed by 20 mg once daily with the dinner meal thereafter. For CrCl 15–50 mL/min: 10 mg twice daily.
Apixaban 10 mg twice daily for 1 week, followed by 5 mg twice daily thereafter. In patients with at least 2 of the following characteristics: age >80 years, body weight <60 kg, or serum creatinine >1.5 mg/dL, the recommended dose is 2.5 mg orally twice daily.
Dabigatran: 5 days of unfractionated heparin, LMWH, or fondaparinux followed by dabigatran 150 mg twice daily with CrCl >30 mL/min. For CrCl 15–30 mL/min: 75 mg twice daily.
Edoxaban: 5 days of unfractionated heparin, LMWH, or fondaparinux followed by edoxaban 60 mg once daily with normal renal function, weight >60 kg, in the absence of potent P-glycoprotein inhibitors. For CrCl 15–50 mL/min: 30 mg once daily.
Warfarin Anticoagulation
Requires 5–10 days of administration to achieve effectiveness as monotherapy
Use full-dose unfractionated heparin, LMWH, or fondaparinux as "bridging agents" when initiating warfarin. Continue parenteral anticoagulation for a minimum of 5 days and until two sequential INR values, at least 1 day apart, achieve the target INR range.
Usual start dose is 5 mg
Titrate to INR, target 2.0–3.0

Abbreviations: aPTT, activated partial thromboplastin time; CrCl, creatinine clearance; INR, international normalized ratio; LMWH, low-molecular-weight heparin; VTE, venous thromboembolism.

Warfarin This vitamin K antagonist prevents carboxylation-dependent activation of coagulation factors II, VII, IX, and X **(Chap. 69)**. The full effect of warfarin is attained after daily therapy for 5–7 days. Overlapping UFH, LMWH, fondaparinux, or parenteral direct thrombin inhibitors during initiation of warfarin provides therapeutic levels of anticoagulation. Parenteral anticoagulation is discontinued when warfarin has achieved a target international normalized ratio (INR) of 2.5. Warfarin can cause major hemorrhage, including intracranial hemorrhage, even when the INR remains within the desired therapeutic range. Warfarin can also cause "off-target" side effects such as alopecia or arterial vascular calcification. Centralized anticoagulation clinics have improved the efficacy and safety of warfarin dosing. Some patients can self-monitor their INR with a home point-of-care fingerstick machine, and a few can be taught to self-dose their warfarin.

Direct Oral Anticoagulants DOACs are administered in a fixed dose, establish effective anticoagulation within hours of ingestion, do not require laboratory coagulation monitoring or limitation of green leafy vegetables, and have few drug-drug interactions. DOACs should not be prescribed during pregnancy or breastfeeding. Warfarin, not DOACs, is the standard treatment for APS. The efficacy and safety of DOACs remain uncertain in patients with end-stage kidney disease or cirrhosis.

Management of Bleeding from Anticoagulants For life-threatening or intracranial hemorrhage due to heparin or LMWH, administer protamine sulfate. The dabigatran antibody, idarucizumab, is an effective, rapidly acting antidote for dabigatran. Andexanet alfa or protein complex concentrate reverses major bleeding caused by anti-Xa DOACs, but beware of rebound thrombosis. With minor bleeding, fresh-frozen plasma, vitamin K, or observation without pharmacologic intervention can be considered.

TABLE 290-5 Take-Home Points from the European Society of Cardiology 2019 Pulmonary Embolism Guidelines

1. Terminology such as "provoked" versus "unprovoked" PE/DVT is no longer supported by the guidelines, as it is potentially misleading and not helpful for decision-making regarding the duration of anticoagulation.
2. Extended oral anticoagulation without an end date should be considered for patients with a first episode of PE and:
   a. No identifiable risk factor
   b. A persistent risk factor
   c. A minor transient or reversible risk factor

Abbreviations: DVT, deep-venous thrombosis; PE, pulmonary embolism.

Cancer and Venous Thromboembolism For patients with cancer and VTE, prescribe LMWH as monotherapy or a DOAC in the absence of a gastrointestinal cancer, and continue extended-duration anticoagulation until the patient is declared cancer-free. The National Comprehensive Cancer Network guidelines endorse treatment of VTE with apixaban or edoxaban.

Duration of Anticoagulation Data-driven guidelines from the European Society of Cardiology are changing our conceptual approach to determining the optimal duration of anticoagulation **(Tables 290-5 and 290-6)**. We should no longer classify VTE as "provoked" or "unprovoked," given that many types of provoked VTE have a similar recurrence rate to unprovoked VTE once anticoagulation is discontinued. Enduring risk factors such as heart failure, chronic lung disease, and inflammatory bowel diseases can heighten the risk of recurrence.

INFERIOR VENA CAVA FILTERS

The two principal indications for insertion of an inferior vena cava filter are (1) active bleeding that precludes anticoagulation and (2) recurrent venous thrombosis despite intensive anticoagulation. Prevention of PE in patients who are not candidates for anticoagulation is a much "softer" indication for filter placement. The filter itself may fail by permitting the passage of small- to medium-size clots via collateral veins that develop. Paradoxically, by providing a nidus for clot formation, filters may increase the DVT rate even though they usually prevent PE. In most situations, place a retrievable rather than a permanent filter.

MANAGEMENT OF MASSIVE PE

For patients with massive PE and hypotension, replete volume with 500 mL of normal saline. Excessive fluid administration exacerbates RV wall stress, causes more profound RV ischemia, and worsens LV compliance and filling by causing further interventricular septal shift toward the LV. Norepinephrine and dobutamine are first-line vasopressor and inotropic agents, respectively, for treatment of PE-related shock. Norepinephrine increases systemic arterial pressure. Dobutamine increases RV inotropy and lowers filling pressures. If

TABLE 290-6 Risk of Recurrent Venous Thromboembolism After Discontinuing Anticoagulation (European Society of Cardiology 2019 Pulmonary Embolism Guidelines)

RISK OF RECURRENCE	EXAMPLES
Low risk (<3% per year)	Major surgery or trauma
Intermediate risk (3–8% per year)	Minor surgery Hospitalized with acute medical illness Pregnancy/estrogens Long-haul flight Inflammatory bowel disease Autoimmune disease No identifiable risk factor (formerly called "unprovoked")
High risk (>8% per year)	Active cancer Antiphospholipid syndrome

additional hemodynamic circulatory support is required, available, and within the scope of the patient's goals of care, consider venoarterial extracorporeal membrane oxygenation (ECMO).

FIBRINOLYSIS

Successful fibrinolytic therapy rapidly reverses right heart failure and may result in a lower rate of death and recurrent PE by (1) dissolving much of the anatomically obstructing pulmonary arterial thrombus, (2) preventing the continued release of serotonin and other neurohumoral factors that exacerbate pulmonary hypertension, and (3) lysing much of the source of the thrombus in the pelvic or deep leg veins, thereby potentially decreasing the likelihood of recurrent PE.

For massive PE, the U.S. Food and Drug Administration has approved systemically administered 100 mg of recombinant tissue plasminogen activator (tPA) (alteplase) prescribed as a continuous intravenous infusion over 2 h. The sooner thrombolysis is administered, the more effective it is. However, systemic fibrinolysis can be utilized effectively for at least 14 days after the PE has occurred. A popular but controversial off-label dosing regimen is 50 mg of tPA administered over 2 h. This lower dose appears less effective than full-dose tPA but may be associated with fewer bleeding complications.

Contraindications to fibrinolysis include intracranial disease, recent surgery, and trauma. The overall major bleeding rate is ~10%, including a 2–3% risk of intracranial hemorrhage. Careful screening of patients for contraindications to fibrinolytic therapy **(Chap. 286)** is the best way to minimize bleeding risk.

For patients with submassive PE who have preserved systolic blood pressure but moderate or severe RV dysfunction, use of fibrinolysis remains controversial. A 2019 American Heart Association Scientific Statement suggests consideration of thrombolysis or embolectomy in patients with a lack of clinical improvement, clinical deterioration, clot in transit, severe or persistent RV strain, or signs of low cardiac output.

CATHETER-DIRECTED THERAPY

There are three types of catheter-directed therapy to treat VTE: (1) pharmacologic only, (2) mechanical only, or (3) pharmacomechanical. *Pharmacologic therapy* usually utilizes a low-dose continuous infusion of tPA. *Mechanical therapy* serves as a percutaneous thrombectomy. The procedure usually employs suction, snaring, and extraction of thrombus without using thrombolytic therapy. One mechanical device is a catheter with expandable nitinol disks that traps the thrombus, which is then removed. *Pharmacomechanical therapy* is epitomized by low-energy ultrasound-facilitated low-dose thrombolysis. The low-dose thrombolysis strategy most likely explains the lower major bleeding complication rate compared with systemically administered fibrinolysis.

SURGICAL PULMONARY EMBOLECTOMY

The substantial risk of major hemorrhage with systemically administered fibrinolysis reawakened interest in alternative treatment with surgical embolectomy, an operation that had almost become extinct. More rapid referral for surgery before the onset of irreversible multisystem organ failure, improved surgical technique, and utilization of ECMO have increased the survival rate of surgical pulmonary embolectomy and have markedly improved right ventricular function.

PULMONARY THROMBOENDARTERECTOMY

CTEPH develops in about 2% of acute PE patients. Therefore, PE patients who have initial pulmonary hypertension (often detected initially during echocardiography) should be followed up at about 6 weeks and, if necessary, at 6 months, with repeat echocardiograms to determine whether pulmonary arterial pressure has worsened, plateaued, or abated. Patients impaired by dyspnea due to CTEPH should be considered for pulmonary thromboendarterectomy, which can markedly reduce, and sometimes even cure, pulmonary hypertension **(Chap. 294)**. The operation requires median sternotomy, cardiopulmonary bypass, deep hypothermia, and periods of hypothermic circulatory arrest. The mortality rate at experienced centers is 1–5%. Inoperable patients should be managed with pulmonary vasodilator therapy and balloon angioplasty of pulmonary arterial webs.

TABLE 290-7 Prevention of Venous Thromboembolism Among Hospitalized Patients

CONDITION	PROPHYLAXIS STRATEGY
High-risk nonorthopedic surgery	Unfractionated heparin 5000 units SC bid or tid Enoxaparin 40 mg daily Dalteparin 2500 or 5000 units daily
Medical oncology (ambulatory)	Apixaban or rivaroxaban for up to 6 months in high-risk patients with cancer starting a new chemotherapy regimen
Cancer surgery, including gynecologic cancer surgery	Enoxaparin 40 mg daily, consider 1 month of prophylaxis
Major orthopedic surgery	Aspirin 81 mg twice daily Rivaroxaban 10 mg daily, beginning 6–10 h postoperatively Dabigatran 110 mg first day, then 220 mg daily Apixaban 2.5 mg bid, beginning 12–24 h postoperatively Warfarin (target INR 2.0) Enoxaparin 40 mg daily Dalteparin 2500 or 5000 units daily Fondaparinux 2.5 mg daily
Medically ill patients, especially if immobilized, with a history of prior VTE, with an indwelling central venous catheter, or with cancer (but without active gastroduodenal ulcer, major bleeding within 3 months, or platelet count <50,000)	Unfractionated heparin 5000 units bid or tid Enoxaparin 40 mg daily Dalteparin 2500 or 5000 units daily Fondaparinux 2.5 mg daily
Medically ill patients about to be discharged from hospital	Rivaroxaban 10 mg daily for about 5 weeks (rarely prescribed for this indication)
Anticoagulation contraindicated	Intermittent pneumatic compression devices and/or graduated vascular below-knee compression stockings

Abbreviations: INR, international normalized ratio; VTE, venous thromboembolism.

PREVENTION OF VTE

Prevention of DVT and PE **(Table 290-7)** is of paramount importance because VTE is difficult to detect and poses a profound medical and economic burden. Low-dose UFH or LMWH is the most common form of in-hospital prophylaxis. Computerized reminder systems in ambulatory and in-hospital practices can increase the use of preventive measures and, at Brigham and Women's Hospital, have reduced the symptomatic VTE rate.

EMOTIONAL SUPPORT

Patients with VTE may feel overwhelmed when they learn that they are suffering from PE or DVT. Some have never previously encountered serious cardiovascular illness. They often assume they will have a reduced life expectancy. They worry about the health of their families and the genetic implications of their illness. At Brigham and Women's Hospital, a physician-nurse–facilitated PE support group was initiated to address these concerns and has met monthly for >30 years. The nonprofit North American Thrombosis Forum (*www.thrombosis.org*) runs monthly online support groups that garner worldwide participation.

FURTHER READING

Barco S et al: Age-sex specific pulmonary embolism-related mortality in the USA and Canada, 2000–18: An analysis of the WHO Mortality Database and of the CDC Multiple Cause of Death database. Lancet Respir Med 9:33, 2021.

Bejjani A et al: When direct oral anticoagulants should not be standard treatment. JACC State-of-the-Art Review. J Am Coll Cardiol 83:444, 2024.

Bikdeli B et al: COVID-19 and thrombotic or thromboembolic disease: Implications for prevention, antithrombotic therapy, and follow-up. J Am Coll Cardiol 75:2590, 2020.

Ghouse J et al: Genome-wide meta-analysis identifies 93 risk loci and enables risk prediction equivalent to monogenic forms of venous thromboembolism. Nat Genet 55:399, 2023.

Goldberg JB et al: Survival and right ventricular function after surgical management of acute pulmonary embolism. J Am Coll Cardiol 76:903, 2020.

Kobayashi T et al: Contemporary management of outcomes of patients with high-risk pulmonary embolism. J Am Coll Cardiol 83:35, 2024.

Konstantinides SV et al: 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS). Eur Heart J 41:543, 2020.

Luijten D et al: Post-pulmonary embolism syndrome and functional outcomes after acute pulmonary embolism. Semin Thromb Hemost 49:848, 2023.

Murthi M et al: Pulmonary embolism readmission trend over the years (from a National Readmission Database). Am J Cardiol 184:133, 2022.

Wadhera RK et al: Community socioeconomic status, acute cardiovascular hospitalizations and mortality among Medicare beneficiaries, 2003 to 2019. Circ Cardovasc Qual Outcomes 17:e010090. 2024.

291

Diseases of the Aorta

Mark A. Creager, Joseph Loscalzo

The aorta is the conduit through which blood ejected from the left ventricle is delivered to the systemic arterial bed. In adults, its diameter is ~3.5 cm at the origin and in the ascending portion, ~2.5 cm in the descending portion in the thorax, and 1.8–2.0 cm in the abdomen. The aortic wall consists of a thin intima composed of endothelium, subendothelial connective tissue, and an internal elastic lamina; a thick tunica media composed of smooth muscle cells, elastic fibers, collagen, and extracellular matrix; and an adventitia composed primarily of connective tissue enclosing the vasa vasorum and nervi vascularis. In addition to the conduit function of the aorta, its viscoelastic and compliant properties serve a buffering function. The aorta is distended during systole to allow a portion of the stroke volume and elastic energy to be stored, and it recoils during diastole so that blood continues to flow to the periphery. Owing to its continuous exposure to high pulsatile pressure and shear stress, the aorta is particularly prone to injury and disease resulting from mechanical trauma. The aorta is also more prone to rupture than is any other vessel, especially with the development of aneurysmal dilation, since its wall tension, as governed by Laplace's law (i.e., proportional to the product of pressure and radius), will be increased.

CONGENITAL ANOMALIES OF THE AORTA

Congenital anomalies of the aorta usually involve the aortic arch and its branches. Symptoms such as dysphagia, stridor, and cough may occur if an anomaly causes a ring around or otherwise compresses the esophagus or trachea. Anomalies associated with symptoms include double aortic arch, origin of the right subclavian artery distal to the left subclavian artery, and right-sided aortic arch with an aberrant left subclavian artery. A Kommerell diverticulum is an anatomic remnant of a right aortic arch. Most congenital anomalies of the aorta do not cause symptoms and are detected during catheter-based procedures. The diagnosis of suspected congenital anomalies of the aorta typically is confirmed by computed tomographic (CT) or magnetic resonance (MR) angiography. Surgery is used to treat symptomatic anomalies.

Coarctation of the aorta **(Chap. 280)** typically occurs near the insertion of the ligamentum arteriosum, adjacent to the left subclavian artery. It may be associated with a bicuspid aortic valve, aortic arch hypoplasia, other congenital heart defects, and intracranial aneurysms. A pulse delay or pressure differential between the upper and lower extremities should raise suspicion of aortic coarctation. Imaging modalities, including echocardiography, CT, and MR angiography are used to confirm the diagnosis. If untreated, hypertension develops in the arteries proximal to the coarctation. Treatment of hemodynamically significant aortic coarctation includes endovascular stent implantation if feasible or surgical repair.

AORTIC ANEURYSM

An *aneurysm* is defined as a pathologic dilation of a segment of a blood vessel. A *true aneurysm* involves all three layers of the vessel wall and is distinguished from a *pseudoaneurysm*, in which the intimal and medial layers are disrupted and the dilated segment of the aorta is lined by adventitia only and, at times, by perivascular clot. Aneurysms also may be classified according to their gross appearance. A *fusiform aneurysm* affects the entire circumference of a segment of the vessel, resulting in a diffusely dilated artery. In contrast, a *saccular aneurysm* involves only a portion of the circumference, resulting in an outpouching of the vessel wall. Aortic aneurysms also are classified according to location, i.e., abdominal versus thoracic. Aneurysms of the descending thoracic aorta are usually contiguous with infradiaphragmatic aneurysms and are referred to as *thoracoabdominal aortic aneurysms*.

■ ETIOLOGY

Aortic aneurysms result from conditions that cause degradation or abnormal production of the structural components of the aortic wall: elastin and collagen. The causes of aortic aneurysms may be broadly categorized as degenerative or sporadic, heritable, congenital, aortitis, infective, and trauma **(Table 291-1)**. Inflammation, oxidative stress, proteolysis, and biomechanical wall stress contribute to the degenerative processes that characterize most aneurysms of the abdominal and descending thoracic aorta. These are mediated by B-cell and T-cell lymphocytes, macrophages, inflammatory cytokines, and matrix metalloproteinases that degrade elastin and collagen and alter the tensile strength and ability of the aorta to accommodate pulsatile stretch. The associated histopathology demonstrates destruction of elastin and collagen, decreased vascular smooth muscle, in-growth of new blood vessels, and inflammation. Factors associated with degenerative aortic aneurysms include aging, cigarette smoking, hypercholesterolemia, hypertension, and male sex.

The most common pathologic condition associated with degenerative aortic aneurysms is *atherosclerosis*. Many patients with aortic aneurysms have coexisting risk factors for atherosclerosis, as well as atherosclerosis in other blood vessels.

The pathologic condition of aortic aneurysms associated with genetic or developmental diseases is medial degeneration, a histopathologic term used to describe the degeneration of collagen and elastic fibers in the tunica media of the aorta as well as the loss of medial cells that are replaced by multiple clefts of mucoid material, such as proteoglycans. Medial degeneration characteristically affects the proximal aorta, results in circumferential weakness and dilation, and leads to the development of fusiform aneurysms involving the ascending aorta and the sinuses of Valsalva. It is found in patients with Marfan syndrome, Loeys-Dietz syndrome, vascular Ehlers-Danlos syndrome **(Chap. 425)**, hypertension, bicuspid aortic valves, Turner syndrome, and familial thoracic aortic aneurysm syndromes. It sometimes appears as an isolated condition in patients without any other apparent disease. Thoracic and abdominal aortic aneurysms also occur in patients with fibromuscular dysplasia, although the nature of the aortic pathology is not established.

TABLE 291-1 Diseases of the Aorta: Etiology and Associated Factors

- Aortic aneurysm
  - Degenerative/sporadic
    - Aging
    - Cigarette smoking
    - Hypercholesterolemia
    - Hypertension
    - Atherosclerosis
  - Heritable
    - Marfan syndrome
    - Loeys-Dietz syndrome
    - Ehlers-Danlos syndrome type IV
    - Aneurysm-osteoarthritis syndrome
    - Smooth muscle dysfunction syndrome
    - Familial (nonsyndromic)
  - Congenital
    - Bicuspid aortic valve
    - Turner syndrome
    - Aortic coarctation
  - Fibromuscular dysplasia
  - Chronic aortic dissection
  - Aortitis (see below)
  - Infective (see below)
  - Trauma
- Acute aortic syndromes (aortic dissection, acute intramural hematoma, penetrating atherosclerotic ulcer)
  - Degenerative disorders (see above)
  - Heritable/disorders (see above)
  - Congenital disorders (see above)
  - Hypertension
  - Aortitis (see below)
  - Pregnancy
  - Trauma
- Aortic occlusion
  - Atherosclerosis
  - Thromboembolism
- Aortitis
  - Vasculitis
    - Takayasu's arteritis
    - Giant cell arteritis
    - IgG4-related aortitis
    - Isolated aortitis
  - Rheumatic
    - Rheumatoid aortitis
    - HLA-B27–associated spondyloarthropathies
    - Behçet syndrome
    - Cogan syndrome
  - Infective
    - Syphilis
    - Tuberculosis
    - Mycotic (*Salmonella*, staphylococcal, streptococcal, fungal)

Familial clusterings of aortic aneurysms occur in 20% of patients, suggesting a hereditary basis for the disease. Mutations of the gene that encodes fibrillin-1 are present in patients with Marfan syndrome. Fibrillin-1 is an important component of extracellular microfibrils, which support the architecture of elastic fibers and other connective tissue. Deficiency of fibrillin-1 in the extracellular matrix leads to excessive signaling by transforming growth factor β (TGF-β). Loeys-Dietz syndrome is caused by mutations in the genes that encode TGF-β isoforms (*TGFB2* and *TGFB3*), TGF-β receptors 1 (*TGFBR1*) and 2 (*TGFBR2*), and SMAD3, which encodes a downstream signaling protein involved with TGF binding to its receptors. Increased signaling by TGF-β and mutations of *TGFBR1*, *TGFBR2*, as well as *TGFB2* and *TGFB3*, may cause thoracic aortic aneurysms. Thoracic aortic aneurysm is associated with autosomal dominant polycystic kidney disease, which is caused by mutations in *PKD1*. Mutations of the genes encoding the smooth muscle–specific alpha-actin (*ACTA2*), smooth muscle cell–specific myosin heavy chain 11 (*MYH11*), myosin light chain kinase (*MYLK*), and type I cGMP-dependent protein kinase (*PRKG1*), as well as mutations of *TGFBR2* and *SMAD3* and several other pathogenic variants, have been reported in some patients with nonsyndromic familial thoracic aortic aneurysms. Mutations in type III procollagen (*COL3A1*) have been implicated in vascular Ehlers-Danlos syndrome.

The infectious causes of aortic aneurysms include syphilis, tuberculosis, and other bacterial infections. *Syphilis* **(Chap. 187)** is a relatively uncommon cause of aortic aneurysm. Syphilitic periaortitis and mesoaortitis damage elastic fibers, resulting in thickening and weakening of the aortic wall. Approximately 90% of syphilitic aneurysms are located in the ascending aorta or aortic arch. *Tuberculous aneurysms* **(Chap. 183)** typically affect the thoracic aorta and result from direct extension of infection from hilar lymph nodes or contiguous abscesses as well as from bacterial seeding. Loss of aortic wall elasticity results from granulomatous destruction of the medial layer. A *mycotic aneurysm* is a rare condition that develops as a result of staphylococcal, streptococcal, *Salmonella*, or other bacterial or fungal infections of the aorta, usually at an atherosclerotic plaque. These aneurysms are usually saccular. Blood cultures are often positive and reveal the nature of the infective agent.

Vasculitides associated with aortic aneurysm include Takayasu arteritis and giant cell arteritis, which may cause aneurysms of the aortic arch and descending thoracic aorta. Thoracic and abdominal aortic aneurysms also occur in patients with IgG4-related systemic disease or those with isolated aortitis. Spondyloarthropathies such as ankylosing spondylitis, rheumatoid arthritis, psoriatic arthritis, relapsing polychondritis, and reactive arthritis are associated with dilation of the ascending aorta. Aortic aneurysms also occur in patients with Behçet syndrome **(Chap. 376)** and Cogan syndrome. *Traumatic aneurysms* may occur after penetrating or nonpenetrating chest trauma and most commonly affect the descending thoracic aorta just beyond the site of insertion of the ligamentum arteriosum. Chronic aortic dissections are associated with weakening of the aortic wall that may lead to the development of aneurysmal dilatation.

THORACIC AORTIC ANEURYSMS

A contemporary definition of an aneurysm affecting the aortic root and ascending aorta is that of the diameter ≥4.5 cm, whereas a diameter of 4.0 cm to 4.4 cm is defined as aortic dilation. An aneurysm of the descending thoracic aorta and abdominal aorta is designated when their diameters are dilated 50% more than the adjacent normal diameter. The clinical manifestations and natural history of thoracic aortic aneurysms depend on their cause and location. Medial degeneration is the most common pathology associated with ascending aortic aneurysms, whereas atherosclerosis is the condition most frequently associated with aneurysms of the descending thoracic aorta. The average growth rate of thoracic aortic aneurysms is 0.1–0.2 cm per year. The risk of rupture is related to the size of the aneurysm and the presence of symptoms, ranging approximately from 2–3% per year for thoracic aortic aneurysms <4.0 cm in diameter to 7% per year for those >6 cm in diameter. Thoracic aortic aneurysms associated with Marfan syndrome may expand at a greater rate. Some patients with Loeys-Dietz syndrome, such as those with the *TGFBR1* and *TGFBR2* pathogenic variants and some with nonsyndromic familial thoracic aortic disease, are at greater risk of rupture at smaller aortic root sizes than those with Marfan syndrome. Most thoracic aortic aneurysms are asymptomatic; however, compression or erosion of adjacent tissue by aneurysms may cause symptoms such as chest pain, shortness of breath, cough, hoarseness, and dysphagia. Aneurysmal dilation of the ascending aorta may cause congestive heart failure as a consequence of aortic regurgitation,

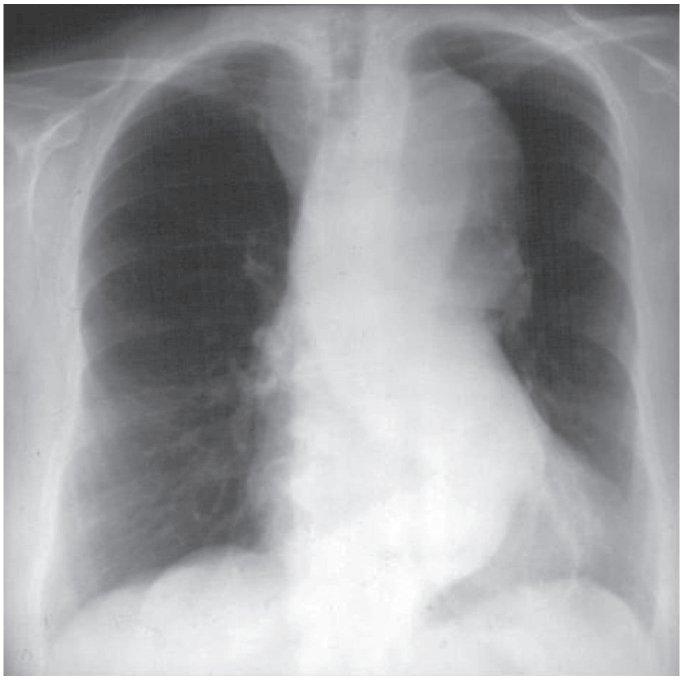

FIGURE 291-1 A chest x-ray of a patient with a thoracic aortic aneurysm.

and compression of the superior vena cava may produce congestion of the head, neck, and upper extremities.

A chest x-ray may be the first test that suggests the diagnosis of a thoracic aortic aneurysm **(Fig. 291-1)**. Findings include widening of the mediastinal shadow and displacement or compression of the trachea or left main stem bronchus. Echocardiography, particularly transesophageal echocardiography, can be used to assess the proximal ascending aorta and descending thoracic aorta. Contrast-enhanced CT, magnetic resonance imaging (MRI), and conventional invasive aortography are sensitive and specific tests for assessment of aneurysms of the thoracic aorta and involvement of branch vessels **(Fig. 291-2)**. In asymptomatic patients whose aneurysms are too small to justify surgery, noninvasive testing with either contrast-enhanced CT or MRI should be performed initially at least every 6–12 months to monitor expansion. Subsequent surveillance imaging should be performed every 6–24 months depending on the size, stability of prior measurements, and the underlying cause.

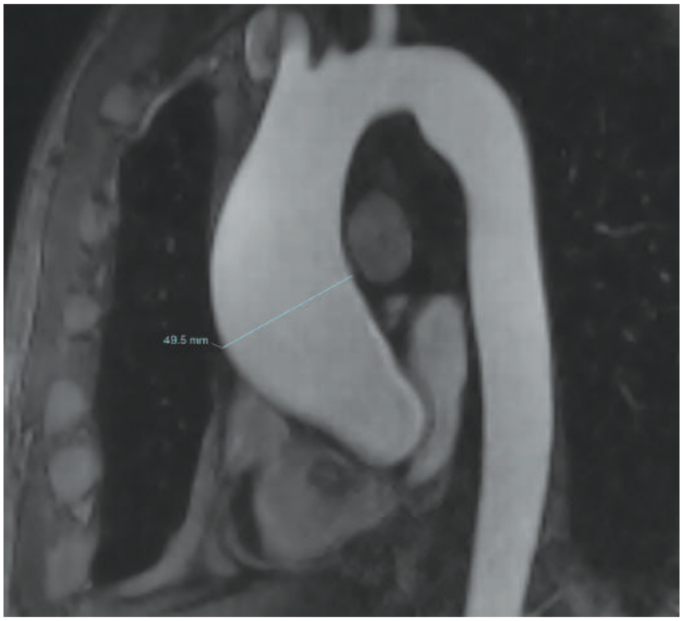

FIGURE 291-2 A magnetic resonance angiogram demonstrating a fusiform aneurysm of the ascending thoracic aorta. *(Courtesy of Dr. Michael Steigner, Brigham and Women's Hospital, Boston, MA, with permission.)*

Genetic testing is recommended in patients with ascending thoracic aortic disease who have features of Marfan syndrome, Loeys-Dietz syndrome, or vascular Ehlers-Danlos syndrome, or who present when they are <60 years of age.

TREATMENT

Thoracic Aortic Aneurysms

Many of the treatment recommendations herein are derived from the recent American Heart Association (AHA)/American College of Cardiology (ACC) Aortic Disease Clinical Practice Guidelines. Treatment with either a β-adrenergic blocker or angiotensin receptor antagonist currently is recommended for patients with thoracic aortic aneurysms associated with Marfan syndrome and Loeys-Dietz syndrome who have evidence of aortic root dilatation to reduce the rate of further expansion. Clinical outcome trials have found that the rate of aortic root enlargement in patients with Marfan syndrome was similar with a β-adrenergic blocker, such as atenolol, and an angiotensin receptor antagonist, such as losartan. These classes of drugs likely confer beneficial effects by different mechanisms. β-adrenergic blockers may decrease the rate of aortic dilation in these patients by reducing stiffness, and angiotensin receptor antagonists may reduce the rate of aortic dilation by blocking TGF-β signaling. Combination therapy with both a β-adrenergic blocker and angiotensin receptor antagonist may be considered. In patients with Ehlers-Danlos syndrome, celiprolol, a β-adrenergic blocker with vasodilator properties, has shown efficacy, but is not available in the United States. For these and other patients with thoracic aortic aneurysms, additional medical therapy should be given as necessary to control hypertension. Risk factor modification with lipid-lowering therapy, smoking cessation interventions, as well as low-dose aspirin, is advised for patients with degenerative thoracic aortic aneurysms associated with atherosclerosis. Operative repair with placement of a prosthetic graft is indicated in patients with symptomatic ascending thoracic aortic aneurysms, and for most asymptomatic aneurysms, including those associated with bicuspid aortic valves, when the aortic root or ascending aortic diameter is ≥5.0 to ≥5.5 cm. The lower threshold applies to patients with high-risk features, such as a family history of aortic dissection or an aortic growth rate ≥0.3 cm per year for 2 years or when the growth rate is >0.5 cm per year, and when performed under the direction of an experienced multidisciplinary aortic team. Replacement of the ascending aorta ≥4.5 cm is reasonable in patients with bicuspid aortic valves undergoing aortic valve replacement because of severe aortic stenosis or aortic regurgitation. In patients with Marfan syndrome, aortic root and ascending thoracic aortic aneurysms of ≥4.5 to ≥5.0 cm should be considered for surgery, with the threshold also depending on the presence of high-risk features. Due to the higher risk of dissection at smaller aortic sizes in patients with Loeys-Dietz syndrome and those with nonsyndromic familial thoracic aortic aneurysms, the aortic diameter threshold for operative repair ranges from ≥4.0 to ≥5.0 cm, depending on the genetic variant and presence of high-risk features. Repair is indicated for patients with isolated aortic arch aneurysms and descending thoracic aortic aneurysms when the diameter is ≥5.5 cm. The threshold for repair of a descending thoracic aortic aneurysm of ≥5.0 cm may be considered in patients with syndromic and nonsyndromic heritable aortopathies or when the diameter of a descending thoracic aortic aneurysm has increased >0.5 cm per year.

■ ABDOMINAL AORTIC ANEURYSMS

Abdominal aortic aneurysms occur more frequently in males than in females, and the incidence increases with age. Cigarette smoking is a potent modifiable risk factor. Abdominal aortic aneurysms ≥4.0 cm may affect 1–2% of men aged >50 years. At least 90% of all abdominal aortic aneurysms >4.0 cm are related to atherosclerotic disease, and most of these aneurysms are below the level of the renal arteries.

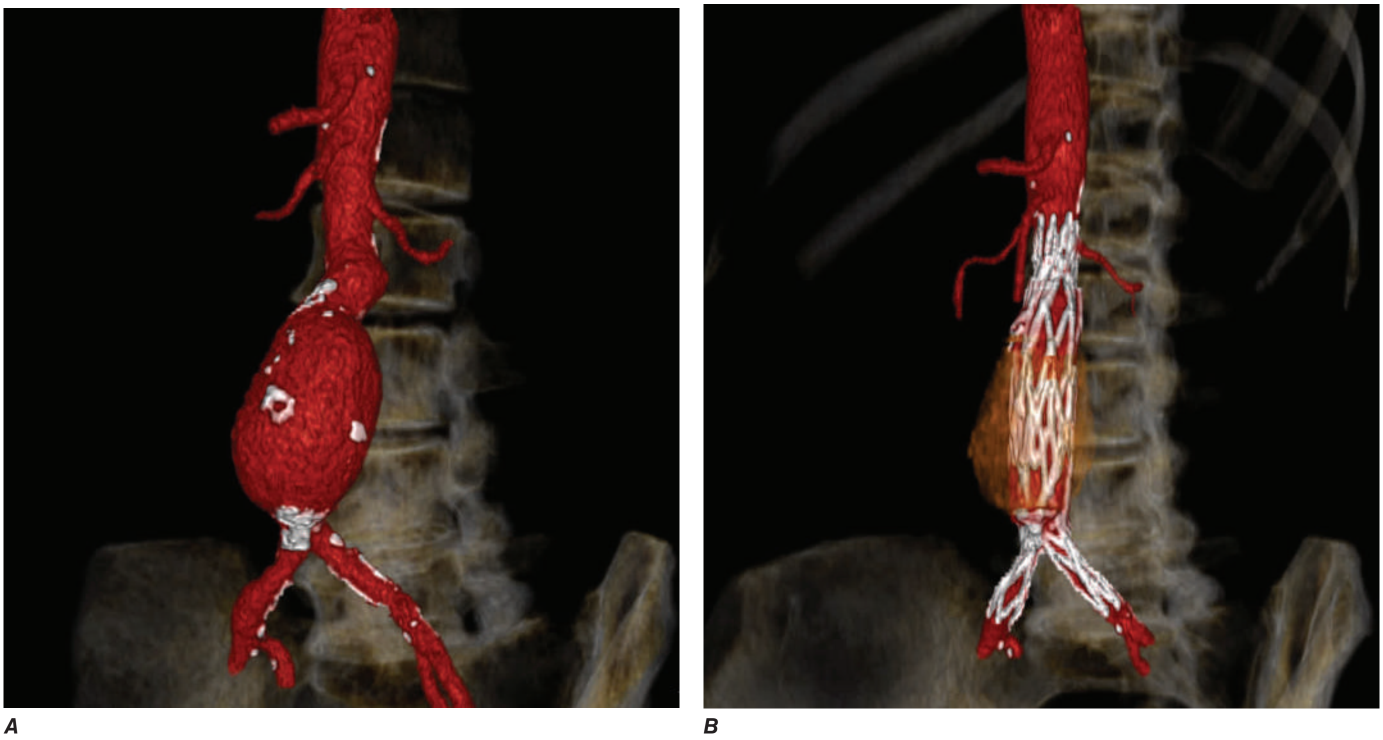

FIGURE 291-3 A computed tomographic angiogram depicting a fusiform abdominal aortic aneurysm before (***A***) and after (***B***) treatment with a bifurcated stent graft. *(Courtesy of Drs. Elizabeth George and Frank Rybicki, Brigham and Women's Hospital, Boston, MA, with permission.)*

Prognosis is related to both the size of the aneurysm and the severity of coexisting coronary artery and cerebrovascular disease. The risk of rupture increases with the size of the aneurysm: the 5-year risk for aneurysms <5 cm is 1–2%, whereas it is 20–40% for aneurysms >5 cm in diameter. The formation of mural thrombi within aneurysms may predispose to peripheral embolization.

An abdominal aortic aneurysm commonly produces no symptoms. It usually is detected on routine examination as a palpable, pulsatile, expansile, and nontender mass, or it is an incidental finding observed on an abdominal imaging study performed for other reasons. As abdominal aortic aneurysms expand, however, they may become painful. Some patients complain of strong pulsations in the abdomen; others experience pain in the chest, lower back, or scrotum. Aneurysmal pain is usually a harbinger of rupture and represents a medical emergency. More often, acute rupture occurs without any prior warning, and this complication is always life-threatening. Rarely, there is leakage of the aneurysm with severe pain and tenderness. Acute pain and hypotension occur with rupture of the aneurysm, which requires an emergency operation or endovascular repair.

Abdominal radiography may demonstrate the calcified outline of the aneurysm; however, ~25% of aneurysms are not calcified and cannot be visualized by x-ray imaging. An abdominal ultrasound can delineate the transverse and longitudinal dimensions of an abdominal aortic aneurysm and may detect mural thrombus. Abdominal ultrasound is useful for serial documentation of aneurysm size and can be used to screen patients at risk for developing an aortic aneurysm. In a large study, ultrasound screening of men aged 65–74 years was associated with a risk reduction in aneurysm-related death of 42%. In a meta-analysis of population-based randomized clinical trials, ultrasound screening of men aged 65 years or older was associated with a 35% risk reduction in aneurysm-related death over 12–15 years. According to the AHA/ACC Guidelines on Aortic Disease, screening by ultrasonography is recommended for men aged ≥65 years who have ever smoked. Although the prevalence of abdominal aortic aneurysms is less in women than in men, the risk of abdominal aortic aneurysms is still substantially increased in women who have smoked. Accordingly, it is reasonable to consider abdominal aortic aneurysm screening in women aged ≥65 years who have ever smoked. In addition, screening for abdominal aortic aneurysms is recommended for males and females ≥65 years who have siblings or offspring with abdominal aortic aneurysms. Individuals with thoracic aortic or peripheral arterial aneurysms should also undergo screening for abdominal aortic aneurysms. CT with contrast and MRI are accurate noninvasive tests to determine the location and size of abdominal aortic aneurysms and to plan endovascular or open surgical repair **(Fig. 291-3*A*)**. Contrast aortography may be used for the evaluation of patients with aneurysms, but the procedure carries a small risk of complications such as bleeding, allergic reactions, and atheroembolism. Since the presence of mural thrombi may reduce the luminal size, aortography may underestimate the diameter of an aneurysm.

TREATMENT

Abdominal Aortic Aneurysms

Many of the treatment recommendations for abdominal aortic aneurysms are derived from the AHA/ACC Aortic Disease Practice Guidelines. Statins are indicated to reduce the risk of cardiovascular events related to atherosclerosis. Medical therapies, such as β-adrenergic blockers and renin-angiotensin inhibitors, have not proven effective in reducing the rate of aneurysm growth. Nonetheless, antihypertensive therapy to lower the blood pressure to <130/<80 mmHg is recommended to reduce the risk of adverse cardiovascular events. Similarly, antiplatelet agents, such as aspirin, are recommended in patients with abdominal aortic aneurysms to reduce the risk of adverse cardiovascular outcomes, but their effects on abdominal aortic aneurysm–specific outcomes are not established. Operative repair of the aneurysm with insertion of a prosthetic graft or endovascular placement of an aortic stent graft **(Fig. 291-3*B*)** is indicated for abdominal aortic aneurysms of any size that are expanding rapidly or are associated with symptoms. For asymptomatic aneurysms, abdominal aortic aneurysm repair is indicated if the diameter is ≥5.5 cm in men and ≥5.0 cm in women. In randomized trials of patients with abdominal aortic aneurysms <5.5 cm, there was no difference in the long-term (>8-year) mortality rate between those followed with ultrasound surveillance and those undergoing elective endovascular or surgical repair. The risk of rupture, however, was higher in woman than in men at smaller diameters. Thus, serial noninvasive follow-up of smaller aneurysms

(<5.5 cm in men and <5.0 cm in women) is an alternative to immediate repair. The decision to perform an open surgical operation or endovascular repair is based in part on the vascular anatomy and comorbid conditions. Endovascular repair of abdominal aortic aneurysms has a lower short-term morbidity rate but a comparable long-term mortality rate with open surgical reconstruction. Long-term surveillance with CT or MR aortography is indicated after endovascular repair to detect leaks and possible aneurysm expansion.

In surgical candidates, careful preoperative cardiac and general medical evaluations (followed by appropriate therapy for complicating conditions) are essential. Preexisting coronary artery disease, congestive heart failure, pulmonary disease, diabetes mellitus, and advanced age add to the risk of surgery. With careful preoperative cardiac evaluation and postoperative care, the operative mortality rate approximates 1–2%. After acute rupture, the mortality rate of emergent operation is 45–50%. Endovascular repair with stent placement is an alternative approach to treat ruptured aneurysms and may be associated with a lower mortality rate.

ACUTE AORTIC SYNDROMES

The four major acute aortic syndromes are aortic rupture (discussed earlier), aortic dissection, intramural hematoma, and penetrating atherosclerotic ulcer. Aortic dissection is caused by a tear of the intima. It often occurs along the right lateral wall of the ascending aorta where the hydraulic shear stress is high. Another common site is the descending thoracic aorta just below the ligamentum arteriosum. The initiating event is either a primary intimal tear with secondary dissection into the media or a medial hemorrhage that dissects into and disrupts the intima. The pulsatile aortic flow then dissects along the elastic lamellar plates of the aorta and creates a false lumen. The dissection usually propagates distally down the descending aorta and into its major branches, but it may propagate proximally. Distal propagation may be limited by atherosclerotic plaque. In some cases, a secondary distal intimal disruption occurs, resulting in the reentry of blood from the false to the true lumen.

There are at least two important pathologic and radiologic variants of aortic dissection: intramural hematoma without an intimal flap and penetrating atherosclerotic ulcer. Acute intramural hematoma is thought to result from rupture of the vasa vasorum with hemorrhage into the wall of the aorta. Most of these hematomas occur in the descending thoracic aorta. Acute intramural hematomas may progress to dissection and rupture. Penetrating atherosclerotic ulcers are caused by erosion of a plaque into the aortic media, are usually localized, but may evolve into an intramural hematoma and also progress to dissection and rupture. They are found primarily in the middle and distal portions of the descending thoracic aorta and are associated with extensive atherosclerotic disease. The ulcer can erode beyond the internal elastic lamina, leading to medial hematoma, and may progress to false aneurysm formation or rupture.

Several classification schemes have been developed for thoracic aortic dissections. DeBakey and colleagues initially classified aortic dissections as type I, in which an intimal tear occurs in the ascending aorta but the dissection may propagate to the aortic arch, the descending thoracic aorta, and even the abdominal aorta; type II, in which the dissection is limited to the ascending aorta; and type III, in which the intimal tear is located in the descending aorta with distal propagation of the dissection **(Fig. 291-4)**. Another classification (Stanford) is that of type A, in which the dissection involves the ascending aorta (proximal dissection), and type B, in which it is limited to the arch and/or descending aorta (distal dissection). From a management standpoint, classification of aortic dissections and intramural hematomas into type A or B is more practical and useful, since DeBakey types I and II are managed in a similar manner.

The factors that predispose to aortic dissection include those associated with medial degeneration and others that increase aortic wall stress (Table 291-1). Systemic hypertension is a coexisting condition in 70% of patients. Aortic dissection is the major cause of morbidity and mortality in patients with Marfan syndrome **(Chap. 425)** or Loeys-Dietz syndrome, and similarly may affect patients with Ehlers-Danlos syndrome. The incidence also is increased in patients with inflammatory aortitis (i.e., Takayasu's arteritis, giant cell arteritis), congenital aortic valve anomalies (e.g., bicuspid valve), coarctation of the aorta, and a history of aortic trauma. In addition, the risk of dissection is increased in otherwise normal women during the third trimester of pregnancy. Aortic dissection also may occur as a consequence of weightlifting, cocaine use, or deceleration injury.

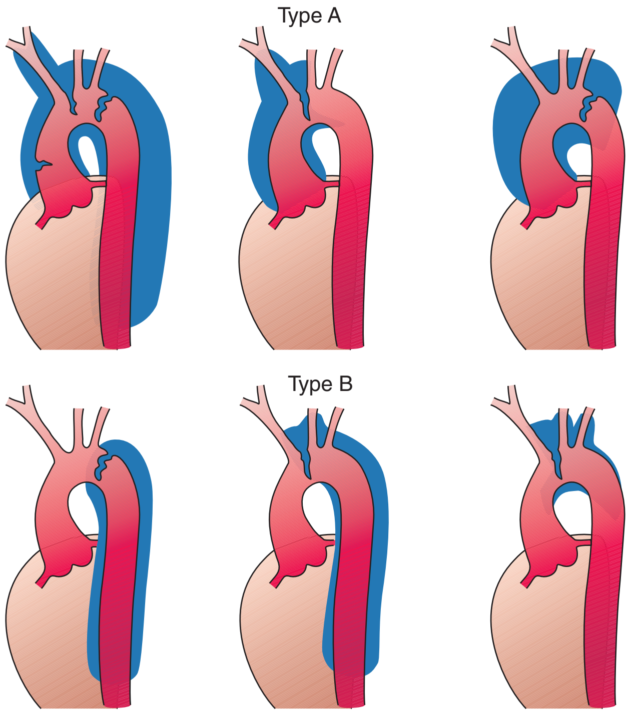

FIGURE 291-4 Classification of aortic dissections. Stanford classification: Type A dissections (*top*) involve the ascending aorta independent of site of tear and distal extension; type B dissections (*bottom*) involve transverse and/or descending aorta without involvement of the ascending aorta. DeBakey classification: Type I dissection involves ascending to descending aorta (*top left*); type II dissection is limited to ascending or transverse aorta, without descending aorta (*top center + top right*); type III dissection involves descending aorta only (*bottom left*). *(Reproduced with permission from DC Miller, in RM Doroghazi, EE Slater [eds]: Aortic Dissection. New York, McGraw-Hill, 1983.)*

CLINICAL MANIFESTATIONS

The peak incidence of aortic dissection is in the sixth and seventh decades. Men are more affected than women by a ratio of 2:1. The presentations of aortic dissection and its variants are the consequences of intimal tear, dissecting hematoma, occlusion of involved arteries, and compression of adjacent tissues. Acute aortic dissection presents with the sudden onset of pain **(Chap. 15)**, which often is described as very severe and tearing and is associated with diaphoresis. The pain may be localized to the front or back of the chest, often the interscapular region, and typically migrates with propagation of the dissection. Other symptoms include syncope, dyspnea, and weakness. Physical findings may include hypertension or hypotension, loss of pulses, aortic regurgitation, pulmonary edema, and neurologic findings due to carotid artery obstruction (hemiplegia, hemianesthesia) or spinal cord ischemia (paraplegia). Bowel ischemia, hematuria, and myocardial ischemia all may occur. These clinical manifestations reflect complications resulting from the dissection occluding the major arteries. Furthermore, clinical manifestations may result from the compression of adjacent structures (e.g., superior cervical ganglia, superior vena cava, bronchus, esophagus) by the expanding dissection causing aneurysmal dilation and include Horner's syndrome, superior vena cava syndrome, hoarseness, dysphagia, and airway compromise. Hemopericardium and cardiac tamponade may complicate a type A lesion with retrograde dissection. Acute aortic regurgitation is an important and common (>50%) complication of proximal dissection. It is the outcome of either a circumferential tear that widens the aortic root or a disruption of the annulus by a dissecting hematoma that tears a leaflet(s) or displaces it inferior to the line of closure. Signs of aortic regurgitation include bounding pulses, a wide pulse pressure, a diastolic murmur often radiating along the right sternal border, and evidence of congestive heart failure. The clinical manifestations depend on the severity of the regurgitation.

In dissections involving the ascending aorta, the chest x-ray often reveals a widened superior mediastinum. A pleural effusion (usually left-sided) also may be present. This effusion is typically serosanguineous and not indicative of rupture unless accompanied by hypotension and falling hematocrit. In dissections of the descending thoracic aorta, a widened mediastinum may be observed on chest x-ray. In addition, the descending aorta may appear to be wider than the ascending portion. An electrocardiogram that shows no evidence of myocardial ischemia is helpful in distinguishing aortic dissection from myocardial infarction among patients who present with chest pain. Rarely, the dissection involves the right or, less commonly, left coronary ostium and causes acute myocardial infarction.

The diagnosis of aortic dissection can be established by noninvasive techniques such as echocardiography, CT, and MRI. Aortography is used less commonly because of the accuracy of these noninvasive techniques. Transthoracic echocardiography can be performed simply and rapidly and has an overall sensitivity of 60–85% for aortic dissection. For diagnosing proximal ascending aortic dissections, its sensitivity exceeds 80%; it is less useful for detecting dissection of the arch and descending thoracic aorta. Transesophageal echocardiography requires greater skill and patient cooperation but is very accurate in identifying dissections of the ascending and descending thoracic aorta but not the arch, achieving 98% sensitivity and ~90% specificity. Echocardiography also provides important information regarding the presence and severity of aortic regurgitation and pericardial effusion. CT and MRI are both highly accurate in identifying the intimal flap and the extent of the dissection and involvement of major arteries; each has a sensitivity and specificity >90%. They are useful in recognizing intramural hemorrhage and penetrating ulcers. The relative utility of CT, MRI, and transesophageal echocardiography depends on the availability and expertise in individual institutions, although among these, CT is the imaging method most often used due to the presence of CT scanners near emergency rooms and the rapidity of performance.

TREATMENT

Aortic Dissection

Medical therapy should be initiated as soon as the diagnosis is considered. The patient should be admitted to an intensive care unit for hemodynamic monitoring. Unless hypotension is present, therapy should be aimed at reducing cardiac contractility and systemic arterial pressure, and thus shear stress. For acute dissection, unless contraindicated, β-adrenergic blockers should be administered parenterally, using intravenous propranolol, metoprolol, or the short-acting esmolol to achieve a heart rate of 60–80 beats/min. This should be accompanied by intravenous dilators, such as sodium nitroprusside, if needed to lower systolic blood pressure to ≤120 mmHg. Labetalol **(Chap. 288)**, a drug with both β- and α-adrenergic blocking properties, also may be used as a parenteral agent in acute therapy for dissection.

The calcium channel antagonists verapamil and diltiazem may be used intravenously if nitroprusside or β-adrenergic blockers cannot be employed. The addition of a parenteral angiotensin-converting enzyme (ACE) inhibitor such as enalaprilat to a β-adrenergic blocker also may be considered. Isolated use of a direct vasodilator such as hydralazine is contraindicated because these agents can increase hydraulic shear and heart rate and may propagate the dissection.

Emergent or urgent surgical correction is the preferred treatment for acute ascending aortic dissections and intramural hematomas (type A). Surgery involves excision of the intimal flap, obliteration of the false lumen, and placement of an interposition graft. Aortic valve repair or aortic root replacement with a composite valve-graft conduit is used if the aortic valve is disrupted. The overall in-hospital mortality rate after surgical treatment of patients with aortic dissection is reported to be 15–25%. The major causes of perioperative mortality and morbidity include myocardial infarction, paraplegia, renal failure, tamponade, hemorrhage, and sepsis. Thoracic endovascular aortic repair with an endoluminal stent graft is indicated for complicated type B dissections, including those characterized by propagation, compromise of major aortic branches, impending rupture, or continued pain. Other transcatheter techniques, such as fenestration of the intimal flaps and stenting of narrowed branch vessels to increase flow to compromised organs, are used in selected patients. Surgical correction is indicated for complicated type B dissections, particularly if endovascular repair is not feasible. Hybrid procedures consisting of both surgery and endovascular repair may be used when the dissection involves both the aortic arch and the descending thoracic aorta. For uncomplicated and stable distal dissections and intramural hematomas (type B), medical therapy is the preferred treatment. The in-hospital mortality rate of medically treated patients with type B dissection is ~12%. Long-term therapy for patients with aortic dissection and intramural hematomas (with or without surgery) consists of control of hypertension and reduction of cardiac contractility with the use of β-adrenergic blockers plus other antihypertensive agents, such as ACE inhibitors or calcium antagonists. Patients with chronic type B dissection and intramural hematomas should be followed on an outpatient basis initially at 1 month, 6 months, and 12 months, and then, if stable, every 12 months with contrast-enhanced CT or MRI to detect propagation or expansion. Patients with Marfan syndrome are at high risk for postdissection complications. The long-term prognosis following hospital discharge for patients with treated dissections is generally good with careful follow-up; the 10-year survival rate is ~60%.

■ CHRONIC ATHEROSCLEROTIC OCCLUSIVE DISEASE

Atherosclerosis may affect the thoracic and abdominal aorta. Occlusive aortic disease caused by atherosclerosis usually is confined to the distal abdominal aorta below the renal arteries. Frequently the disease extends to the iliac arteries **(Chap. 292)**. Claudication characteristically involves the buttocks, thighs, and calves and may be associated with impotence in males (Leriche syndrome). The severity of the symptoms depends on the adequacy of collaterals. With sufficient collateral blood flow, a complete occlusion of the abdominal aorta may occur without the development of ischemic symptoms. The physical findings include the absence of femoral and other distal pulses bilaterally and the detection of an audible bruit over the abdomen (usually at or below the umbilicus) and the common femoral arteries. Atrophic skin, loss of hair, and coolness of the lower extremities usually are observed. In advanced ischemia, rubor on dependency and pallor on elevation can be seen.

The diagnosis usually is established by physical examination and noninvasive testing, including leg pressure measurements, Doppler velocity analysis, pulse volume recordings, and duplex ultrasonography. The anatomy may be defined by MRI, CT, or conventional contrast angiography, typically performed when one is considering revascularization. Catheter-based endovascular or operative treatment is indicated in patients with lifestyle-limiting or debilitating symptoms of claudication and patients with chronic limb-threatening ischemia.

■ ACUTE AORTIC OCCLUSION

Acute occlusion in the distal abdominal aorta constitutes a medical emergency because it threatens the viability of the lower extremities; it usually results from an occlusive (saddle) embolus that almost always originates from the heart. Rarely, acute occlusion may occur as the result of in situ thrombosis in a preexisting severely narrowed segment of the aorta.

The clinical picture is one of acute ischemia of the lower extremities. Severe rest pain, coolness, and pallor of the lower extremities and the absence of distal pulses bilaterally are the usual manifestations.

Diagnosis should be established rapidly by MRI, CT, or aortography. Emergency thrombectomy or revascularization is indicated.

AORTITIS

Aortitis, a term referring to inflammatory disease of the aorta, may be caused by large vessel vasculitides such as Takayasu arteritis, giant cell arteritis, IgG4-related systemic disease, isolated aortitis, rheumatic and HLA-B27–associated spondyloarthropathies, Behçet syndrome, antineutrophil cytoplasmic antibody (ANCA)–associated vasculitides, Cogan syndrome, Erdheim-Chester disease, and infections such as syphilis, tuberculosis, and *Salmonella*, or it may be associated with retroperitoneal fibrosis. Aortitis may result in aneurysmal dilation and aortic regurgitation, occlusion of the aorta and its branch vessels, or acute aortic syndromes.

TAKAYASU ARTERITIS

(See also Chap. 375) This inflammatory disease often affects the ascending aorta and aortic arch, causing obstruction of the aorta and its major arteries. Takayasu arteritis is also termed *pulseless disease* because of the frequent occlusion of the large arteries originating from the aorta. It also may involve the descending thoracic and abdominal aorta and occlude large branches such as the renal arteries. Aortic aneurysms also may occur. The pathology is a panarteritis characterized by mononuclear cells and occasionally giant cells, with marked intimal hyperplasia, medial and adventitial thickening, and, in the chronic form, fibrotic occlusion. The disease is most prevalent in young females of Asian descent but does occur in women of other geographic and ethnic origins and also in young men. During the acute stage, fever, malaise, weight loss, and other systemic symptoms may be evident. Elevations of the erythrocyte sedimentation rate and C-reactive protein are common. The chronic stages of the disease, which is intermittently active, present with symptoms related to large artery occlusion, such as upper extremity claudication, cerebral ischemia, and syncope. The process is progressive, and there is no definitive therapy. Glucocorticoids are effective in most patients during the acute phase. Other immunosuppressive agents, such as methotrexate, azathioprine, leflunomide, or mycophenolate, are prescribed to some patients to lower glucocorticoid requirements and treat relapses. Biologically targeted agents, such as the tumor necrosis factor (TNF) inhibitors etanercept and infliximab, are also used, but efficacy has not been established in randomized clinical trials. Surgical bypass or endovascular intervention of a critically stenotic artery may be necessary.

GIANT CELL ARTERITIS

(See also Chap. 375) This vasculitis occurs in older individuals and affects women more often than men. Primarily large and medium-size arteries are affected. The pathology is that of focal granulomatous lesions involving the entire arterial wall; it frequently is associated with polymyalgia rheumatica. Obstruction of medium-size arteries (e.g., temporal and ophthalmic arteries) and major branches of the aorta and the development of aortitis and aortic regurgitation are important complications of the disease. High-dose glucocorticoid therapy should be administered early and then gradually tapered. Immunosuppressive therapy with methotrexate may allow reduction in steroid dosage and reduce the risk of relapse. Tocilizumab, an interleukin-6 antagonist, demonstrated efficacy in several randomized trials. Other biologically targeted therapies are under investigation.

IGG4-RELATED AORTITIS

Aortitis may occur in patients with IgG4-related disease **(Chap. 380)** and is associated with retroperitoneal fibrosis and hydronephrosis. It can affect the thoracic and abdominal aorta, as well as the iliac arteries. Serum IgG4 levels may be elevated but are not diagnostic. Histopathologic characteristics include a lymphoplasmacytic infiltrate that comprises IgG4-positive plasma cells, fibrosis, and obliterative adventitial phlebitis; it affects men more than women and typically occurs in middle age. Glucocorticoids are used for initial treatment. Case series have reported efficacy with rituximab, an anti-CD20 monoclonal antibody. Immunosuppressive agents such as azathioprine are steroid sparing and may be effective.

ISOLATED AORTITIS

Isolated abdominal aortitis is characterized by adventitial and periaortic inflammation with thickening of the aortic wall; it is associated with abdominal aortic aneurysms and idiopathic retroperitoneal fibrosis. Affected individuals may present with vague constitutional symptoms, fever, and abdominal pain. Retroperitoneal fibrosis can cause ureteral obstruction and hydronephrosis. Glucocorticoids and immunosuppressive agents may reduce the inflammation.

RHEUMATIC AORTITIS

Rheumatoid arthritis **(Chap. 370)**, ankylosing spondylitis **(Chap. 374)**, psoriatic arthritis **(Chap. 374)**, reactive arthritis (formerly known as Reiter's syndrome) **(Chap. 374)**, relapsing polychondritis, and inflammatory bowel disorders may all be associated with aortitis involving the ascending aorta. The inflammatory lesions usually involve the ascending aorta and may extend to the sinuses of Valsalva, the mitral valve leaflets, and adjacent myocardium. The clinical manifestations are aneurysm, aortic regurgitation, and involvement of the cardiac conduction system.

INFECTIVE AORTITIS

Infective aortitis may result from direct invasion of the aortic wall by bacterial pathogens such as *Staphylococcus*, *Streptococcus*, and *Salmonella* or by fungi. These bacteria cause aortitis by infecting the aorta at sites of atherosclerotic plaque. Bacterial proteases lead to degradation of collagen, and the ensuing destruction of the aortic wall leads to the formation of a saccular aneurysm referred to as a mycotic aneurysm. Mycotic aneurysms have a predilection for the suprarenal abdominal aorta. The pathologic characteristics of the aortic wall include acute and chronic inflammation, abscesses, hemorrhage, and necrosis. Mycotic aneurysms typically affect the elderly and occur in men three times more frequently than in women. Patients may present with fever, sepsis, and chest, back, or abdominal pain; there may have been a preceding diarrheal illness. Blood cultures are positive in the majority of patients. Both CT and MRI are useful to diagnose mycotic aneurysms. Treatment includes antibiotic therapy and surgical removal of the affected part of the aorta and revascularization of the lower extremities with grafts placed in uninfected tissue.

Syphilitic aortitis is a late manifestation of luetic infection **(Chap. 187)** that usually affects the proximal ascending aorta, particularly the aortic root, resulting in aortic dilation and aneurysm formation. Syphilitic aortitis occasionally may involve the aortic arch or the descending aorta. The aneurysms may be saccular or fusiform and are usually asymptomatic, but compression of and erosion into adjacent structures may result in symptoms; rupture also may occur.

The initial lesion is an obliterative endarteritis of the vasa vasorum, especially in the adventitia. This is an inflammatory response to the invasion of the adventitia by the spirochetes. Destruction of the aortic media occurs as the spirochetes spread into this layer, usually via the lymphatics accompanying the vasa vasorum. Destruction of collagen and elastic tissues leads to dilation of the aorta, scar formation, and calcification. These changes account for the characteristic radiographic appearance of linear calcification of the ascending aorta.

The disease typically presents as an incidental chest radiographic finding 15–30 years after initial infection. Symptoms may result from aortic regurgitation, narrowing of coronary ostia due to syphilitic aortitis, compression of adjacent structures (e.g., esophagus), or rupture. Diagnosis is established by a positive serologic test, such as rapid plasmin regain (RPR), Venereal Disease Research Laboratory (VDRL), or fluorescent treponemal antibody absorption (FTA-ABS). Treatment includes penicillin and surgical excision and repair.

FURTHER READING

Chou E et al: Genetics and mechanisms of thoracic aortic disease. Nat Rev Cardiol 20:168, 2023.

Fletcher AJ et al: Inherited thoracic aortic disease: New insights and translational targets. Circulation 141:1570, 2020.

Guirguis-Blake JM et al: Primary care screening for abdominal aortic aneurysm: Updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 322:2211, 2019.

Hensley SE, Upchurch GR Jr: Repair of abdominal aortic aneurysms: *JACC* focus seminar, part 1. J Am Coll Cardiol 80:831, 2022.

Isselbacher EM et al: 2022 ACC/AHA guideline for the diagnosis and management of aortic disease: A report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines. Circulation 146:e334, 2022.

Kadian-Dodov D et al: Inflammatory diseases of the aorta: *JACC* focus seminar, part 2. J Am Coll Cardiol 80:832, 2022.

Pitcher A et al: Angiotensin receptor blockers and β blockers in Marfan syndrome: An individual patient data meta-analysis of randomised trials. Lancet 400:822, 2022.

Vilacosta I et al: Acute aortic syndrome revisited: *JACC* state-of-the-art review. J Am Coll Cardiol 78:2106, 2021.

292

Arterial Diseases of the Extremities

Mark A. Creager, Joseph Loscalzo

PERIPHERAL ARTERY DISEASE

Peripheral artery disease (PAD) is defined as a clinical disorder in which there is a stenosis or occlusion in the aorta or the arteries of the limbs. Atherosclerosis is the leading cause of PAD in patients >40 years old. Other causes include thrombosis, embolism, vasculitis, fibromuscular dysplasia, entrapment, cystic adventitial disease, and trauma. The highest prevalence of atherosclerotic PAD occurs in the sixth and seventh decades of life. Prevalence of PAD is similar in men and women, is greater in persons identified as black than non-Hispanic white, and is associated with lower socioeconomic status. As in patients with atherosclerosis of the coronary and cerebral vasculature, there is an increased risk of developing PAD in cigarette smokers and in persons with diabetes mellitus, hypercholesterolemia, elevated lipoprotein(a), hypertension, or renal insufficiency.

Pathology Segmental lesions that cause stenosis or occlusion are usually localized to large and medium-size vessels. The pathology of the lesions includes atherosclerotic plaques with calcium deposition, thinning of the media, patchy destruction of muscle and elastic fibers, fragmentation of the internal elastic lamina, and thrombi composed of platelets and fibrin. The primary sites of involvement are the abdominal aorta and iliac arteries (30% of symptomatic patients), the femoral and popliteal arteries (80–90% of patients), and the more distal vessels, including the tibial and peroneal arteries (40–50% of patients). Atherosclerotic lesions occur preferentially at arterial branch points, which are sites of increased turbulence, altered shear stress, and intimal injury. Involvement of the distal vasculature is most common in elderly individuals and patients with diabetes mellitus.

Clinical Evaluation Fewer than 50% of patients with PAD are symptomatic, although many have a slow or impaired gait. The most typical *symptom* is intermittent claudication, which is defined as a pain, ache, cramp, numbness, or a sense of fatigue in the muscles; it occurs during exercise and is relieved by rest. The site of claudication is distal to the location of the occlusive lesion. For example, buttock, hip, thigh, and calf discomfort occurs in patients with aortoiliac disease, whereas calf claudication develops in patients with femoral-popliteal disease. Symptoms are far more common in the lower than in the upper extremities because of the higher incidence of obstructive lesions in the former region. In patients with severe arterial occlusive disease in whom resting blood flow cannot accommodate basal nutritional needs of the tissues, chronic limb-threatening ischemia may develop. Patients complain of rest pain or a feeling of cold or numbness in the foot and toes. Frequently, these symptoms occur at night when the legs are horizontal and improve when the legs are in a dependent position. With severe ischemia, rest pain may be persistent.

A comprehensive vascular examination includes palpation of the peripheral pulses, including the femoral, popliteal, dorsalis pedis, and posterior tibial arteries; auscultation of the abdomen and groin for bruits; and inspection of the legs and feet. Important *physical findings* of PAD include decreased or absent pulses distal to the obstruction, the presence of bruits over the narrowed artery, and muscle atrophy. With more severe disease, hair loss, thickened nails, smooth and shiny skin, reduced skin temperature, and pallor or cyanosis are common physical signs. In patients with chronic limb-threatening ischemia, ulcers or gangrene may occur. Elevation of the legs and repeated flexing of the calf muscles produce pallor of the soles of the feet, whereas rubor, secondary to reactive hyperemia, may develop when the legs are dependent. The time required for rubor to develop or for the veins in the foot to fill when the patient's legs are transferred from an elevated to a dependent position is related to the severity of the ischemia and the presence of collateral vessels. Patients with severe ischemia may develop peripheral edema because they keep their legs in a dependent position much of the time. Ischemic neuropathy can result in numbness and hyporeflexia.

Noninvasive Testing The history and physical examination are often sufficient to establish the diagnosis of PAD. An objective assessment of the presence and severity of disease is obtained by noninvasive techniques. Arterial pressure can be recorded noninvasively in the legs by placement of sphygmomanometric cuffs at the ankles and the use of a Doppler device to auscultate or record blood flow from the dorsalis pedis and posterior tibial arteries. Normally, systolic blood pressure in the legs and arms is similar. Indeed, ankle pressure may be slightly higher than arm pressure due to pulse-wave amplification. In the presence of hemodynamically significant stenoses, the systolic blood pressure in the leg is decreased. Thus, the ratio of the ankle and brachial artery systolic pressures (termed the *ankle-brachial index*, or ABI) is 1.00–1.40 in normal individuals. ABI values of 0.91–0.99 are considered "borderline," and those ≤0.90 are abnormal and diagnostic of PAD. ABIs >1.40 indicate noncompressible arteries secondary to vascular calcification, as may occur in patients with diabetes mellitus or chronic kidney disease. When an ABI cannot be accurately assessed due to vascular calcification, the presence of PAD can be detected by measuring the systolic pressure of the great toe and calculating *a toe-brachial index* (TBI). A TBI of ≤0.70 is considered abnormal.

Other noninvasive tests include segmental pressure measurements, segmental pulse volume recordings, duplex ultrasonography (which combines B-mode imaging and Doppler flow velocity waveform analysis), transcutaneous oximetry, and stress testing (usually using a treadmill). Placement of pneumatic cuffs enables assessment of systolic pressure along the legs. The presence of pressure gradients between sequential cuffs provides evidence of the presence and location of hemodynamically significant stenoses. In addition, the amplitude of the pulse volume contour becomes blunted in the presence of significant PAD. Duplex ultrasonography is used to image and detect stenotic lesions in native arteries and bypass grafts.

Treadmill testing allows the physician to assess functional limitations objectively. Decline of the ABI immediately after exercise provides further support for the diagnosis of PAD in patients with equivocal symptoms and findings on examination.

Magnetic resonance angiography (MRA), computed tomographic angiography (CTA), and conventional catheter-based angiography should not be used for routine diagnostic testing, but are performed before potential revascularization **(Fig. 292-1)**. Each test is useful in defining the anatomy to assist planning for endovascular and surgical revascularization procedures.

Prognosis The natural history of patients with PAD is influenced primarily by the extent of coexisting coronary artery and cerebrovascular disease. Approximately one-third to one-half of patients with symptomatic PAD have evidence of coronary artery disease (CAD) based on clinical presentation and electrocardiogram, and over one-half have

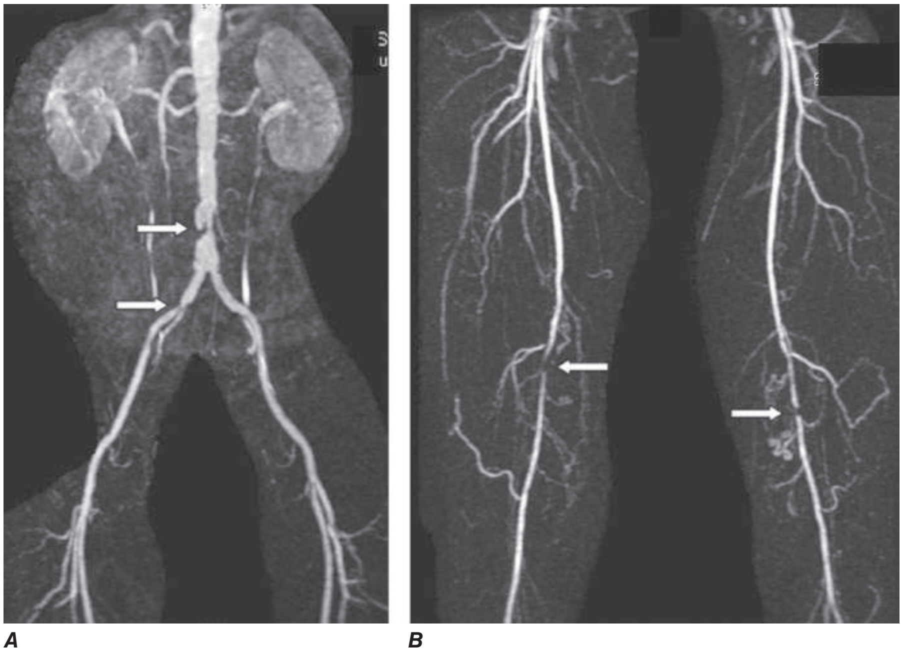

FIGURE 292-1 Magnetic resonance angiography of a patient with intermittent claudication, showing stenoses of the distal abdominal aorta and right common iliac artery (***A***) and stenoses of the right and left superficial femoral arteries (***B***). *(Courtesy of Dr. Edwin Gravereaux, with permission.)*

significant CAD by coronary angiography. Patients with PAD have a 15–25% 5-year mortality rate and a two- to fourfold increased risk of death from cardiovascular disease. Measurement of ABI is useful for detecting PAD and identifying persons at risk for adverse cardiovascular and limb events. Mortality rates are highest in those with the most severe PAD. The ABI worsens in almost 40% of patients, and symptoms progress in ~20–25% when assessed over a period of 5 years. Approximately 11% of patients with symptomatic PAD ultimately develop chronic limb-threatening ischemia, and without revascularization, approximately 25% of patients with chronic limb-threatening ischemia undergo amputation within 1 year. The prognosis is worse in patients who continue to smoke cigarettes or have diabetes mellitus.

TREATMENT

Peripheral Artery Disease

Patients with PAD should receive therapies to reduce the risk of associated cardiovascular events, such as myocardial infarction and death, and to improve limb symptoms, prevent progression to chronic limb-threatening ischemia, and preserve limb viability. Risk factor modification and antithrombotic therapy should be initiated to improve cardiovascular outcomes. The importance of discontinuing cigarette smoking cannot be overemphasized. The physician must assume a major role in this lifestyle modification. Counseling and adjunctive drug therapy with the nicotine patch, bupropion, or varenicline increase smoking cessation rates and reduce recidivism. It is important to control blood pressure in hypertensive patients. Angiotensin-converting enzyme inhibitors and angiotensin receptor blockers may reduce the risk of cardiovascular events in patients with symptomatic PAD. β-Adrenergic blockers do not worsen claudication and may be used to treat hypertension, especially in patients with coexistent CAD. Treatment of hypercholesterolemia with statins and, if needed, adjunctive lipid-lowering agents such as ezetimibe or a PCSK9 inhibitor, are advocated to reduce the risk of myocardial infarction, stroke, and death. Statins and PCSK9 inhibitors also are associated with a decreased risk of adverse limb events, including amputation, in patients with PAD. The 2018 American Heart Association (AHA)/American College of Cardiology (ACC) Guideline on the Management of Blood Cholesterol recommends high-intensity statin treatment in patients with atherosclerotic disorders, including PAD, with the aim of achieving a 50% or greater reduction in low-density lipoprotein cholesterol. Intensive glucose lowering reduces the risk of amputations in patients with diabetes mellitus. Among patients with diabetes and established cardiovascular disease, including PAD, the glucagon-like protein (GLP)-1 agonists and the sodium-glucose cotransporter 2 (SGLT2) inhibitors have beneficial cardiovascular effects. GLP-1 agonists also may reduce the risk of amputation. Platelet inhibitors, including aspirin and the adenosine diphosphate (ADP) antagonist clopidogrel, reduce the risk of adverse cardiovascular events in patients with atherosclerosis and are recommended for patients with symptomatic PAD, including those with intermittent claudication or chronic limb-threatening ischemia or prior lower extremity revascularization. Outcomes with ticagrelor are similar to those with clopidogrel. The benefit of dual antiplatelet therapy with both aspirin and clopidogrel compared with aspirin alone in reducing cardiovascular morbidity and mortality rates in patients with PAD is uncertain. When added to other antiplatelet therapy, vorapaxar, a protease-activated receptor-1 antagonist that inhibits thrombin-mediated platelet activation, decreases the risk of adverse cardiovascular events in patients with atherosclerosis, including PAD. It also reduces the risk of acute limb ischemia and peripheral revascularization; however, it is associated with an increased rate of moderate bleeding. The anticoagulant warfarin is as effective as antiplatelet therapy in preventing adverse cardiovascular events but causes more major bleeding; therefore, it is not indicated to improve outcomes in patients with chronic PAD. The combination of a low dose of the oral factor Xa inhibitor rivaroxaban and aspirin improves cardiovascular and limb outcomes in patients with atherosclerosis, including those with established PAD, as well as those who have undergone peripheral revascularization, but is associated with increased risk of bleeding.

Therapies for intermittent claudication and chronic limb-threatening ischemia include supportive measures, medications, exercise training, endovascular interventions, and surgery. Supportive measures include meticulous care of the feet, which should be kept clean and protected against excessive drying with moisturizing creams. Well-fitting and protective shoes are advised to reduce trauma. Elastic support hose should be avoided, as it reduces blood flow to the skin. In patients with chronic limb-threatening ischemia, shock blocks under the head of the bed together with a canopy over the feet may improve perfusion pressure and ameliorate some of the rest pain.

Patients with claudication should be encouraged to exercise regularly and at progressively more strenuous levels. Supervised exercise training programs for 30- to 45-min sessions, at least three times per week for 12 weeks, prolong walking distance. The beneficial effect of supervised exercise training on walking performance in patients with claudication often is similar to or greater than that realized after a revascularization procedure. Structured home and community-based exercise programs are also effective. Pharmacologic treatment of PAD has not been as successful as the medical treatment of CAD **(Chap. 284)**. In particular, vasodilators as a class have not proved to be beneficial. During exercise, peripheral vasodilation occurs distal to sites of significant arterial stenoses. As a result, perfusion pressure falls, often to levels lower than those generated in the interstitial tissue by the exercising muscle. Drugs such as α-adrenergic blocking agents, calcium channel antagonists, and other vasodilators have not been shown to be effective in patients with PAD.

Cilostazol, a phosphodiesterase-3 inhibitor with vasodilator and antiplatelet properties, increases claudication distance by 40–60% and improves measures of quality of life. The mechanism of action accounting for its beneficial effects is not known. Pentoxifylline, a substituted xanthine derivative, increases blood flow to the microcirculation and enhances tissue oxygenation. Although several placebo-controlled studies have found that pentoxifylline modestly increases the duration of exercise, its efficacy has not been confirmed in other clinical trials. Statins appeared effective for treatment of intermittent claudication in initial clinical trials, but more studies are needed to confirm the efficacy of this class of drugs.

There is no definitive medical therapy for chronic limb-threatening ischemia. Vasodilators, including prostaglandins, are not effective in relieving symptoms or preventing limb loss. Enthusiasm for therapy with angiogenic growth factors abated when clinical trials of intramuscular gene transfer of DNA encoding vascular endothelial growth factor, fibroblast growth factor, hepatocyte growth factor, or hypoxia-inducible factor 1α failed to demonstrate improvement in symptoms or outcomes in patients with intermittent claudication or critical limb ischemia. Most clinical trials of bone marrow–derived vascular progenitor cells to promote angiogenesis and preserve limb viability in patients with critical limb ischemia have failed to demonstrate benefit, although a meta-analysis of these trials suggested a modest reduction in the risk of amputation.

REVASCULARIZATION

Revascularization procedures, including catheter-based, open surgical, and hybrid interventions, are indicated for patients with chronic limb-threatening ischemia to relieve pain and prevent limb loss. These procedures are also indicated for patients with disabling, progressive, or severe symptoms of intermittent claudication despite medical therapy in order to improve walking distance and functional capacity. When revascularization is performed in conjunction with a supervised exercise program, walking distance improves more than with exercise training alone. MRA, CTA, or conventional angiography should be performed to assess vascular anatomy in patients who are being considered for revascularization.

Endovascular interventions include percutaneous transluminal balloon angioplasty (PTA) (including drug-coated balloons), stent placement (including drug-eluting stents), stent grafts, and atherectomy **(Chap. 287)**. Several operative procedures are available for treating patients with PAD. The preferred operative procedure depends on the location and extent of the obstruction(s) and the general medical condition of the patient. Operative procedures for aortoiliac disease include aortobifemoral bypass, axillofemoral bypass, femoro-femoral bypass, and aortoiliac endarterectomy. Operative therapy for femoral-popliteal and tibioperoneal artery disease includes in situ and reverse autogenous saphenous vein bypass grafts, placement of polytetrafluoroethylene (PTFE) or other synthetic grafts, and thromboendarterectomy. The decision to use an endovascular, open surgical, or a hybrid revascularization strategy depends on the vascular anatomy, including the location and extent of the arterial occlusions, the availability of suitable saphenous vein segments, comorbidities, and the skill and experience of the operator.

Preoperative cardiac risk assessment may identify individuals who are especially likely to experience an adverse cardiac event during the perioperative period. Patients with angina, prior myocardial infarction, heart failure, diabetes, or renal insufficiency are among those at increased risk. Stress testing with treadmill exercise (if feasible), radionuclide myocardial perfusion imaging, or echocardiography permits further stratification of risk in these patients, particularly those with poor or unknown functional capacity **(Chap. 287)**. Patients with abnormal test results require close supervision and adjunctive management with anti-ischemic medications. Coronary angiography and coronary artery revascularization compared with optimal medical therapy do not improve outcomes in most patients undergoing peripheral vascular surgery, but cardiac catheterization should be considered in patients with unstable angina and angina refractory to medical therapy as well as those suspected of having left main or three-vessel CAD.

FIBROMUSCULAR DYSPLASIA

Fibromuscular dysplasia (FMD) is a hyperplastic disorder that typically affects medium-size and small arteries, but it can also affect larger arteries. The most common histologic characteristics include medial deposition of collagen causing fibromuscular ridges alternating with areas of thinned media. FMD occurs predominantly in females and usually involves the renal and carotid/vertebral arteries but can involve coronary and mesenteric arteries, as well as extremity vessels such as the iliac and subclavian arteries. FMD may cause stenosis, dissection, aneurysm, or thrombosis in affected arteries. When limb vessels are involved, clinical manifestations are similar to those for atherosclerosis, including claudication and rest pain.

A contemporary classification based on the angiographic appearance divides FMD into two types: multifocal, identified by a “string of beads” appearance caused by the thickened fibromuscular ridges contiguous with thin, less-involved portions of the arterial wall; and focal, identified as either unifocal (<1 cm) or tubular (≥1 cm) stenoses.

Aspirin is recommended to reduce the risk of thrombosis in affected vessels. PTA and surgical reconstruction are beneficial in patients with debilitating symptoms or threatened limbs.

THROMBOANGIITIS OBLITERANS

Thromboangiitis obliterans (Buerger’s disease) is an inflammatory occlusive vascular disorder involving small and medium-size arteries and veins in the distal upper and lower extremities. Cerebral, visceral, and coronary vessels may be affected rarely. This disorder develops most frequently in men <40 years of age. The prevalence is higher in Asians and individuals of Eastern European descent. Although the cause of thromboangiitis obliterans is not known, there is a definite relationship to cigarette smoking in patients with this disorder.

In the initial stages of thromboangiitis obliterans, polymorphonuclear leukocytes infiltrate the walls of the small and medium-size arteries and veins. The internal elastic lamina is preserved, and a cellular, inflammatory thrombus develops in the vascular lumen. As the disease progresses, mononuclear cells, fibroblasts, and giant cells replace the neutrophils. Later stages are characterized by perivascular fibrosis, organized thrombus, and recanalization.

The clinical features of thromboangiitis obliterans often include a triad of claudication of the affected extremity, Raynaud phenomenon, and migratory superficial vein thrombophlebitis. Claudication usually is confined to the calves and feet or the forearms and hands because this disorder primarily affects distal vessels. In the presence of severe digital ischemia, trophic nail changes, painful ulcerations, and gangrene may develop at the tips of the fingers or toes. The physical examination shows normal brachial and popliteal pulses but reduced or absent radial, ulnar, and/or tibial pulses. MRA, CTA, and conventional arteriography are helpful in making the diagnosis. Smooth, tapering segmental lesions in the distal vessels are characteristic, as are collateral vessels at sites of vascular occlusion. Proximal atherosclerotic disease is usually absent. The diagnosis can be confirmed by excisional biopsy and pathologic examination of an involved vessel.

There is no specific treatment except abstention from tobacco. The prognosis is worse in individuals who continue to smoke, but results are discouraging even in those who stop smoking. Arterial bypass of the larger vessels may be used in selected instances, as well as local debridement, depending on the symptoms and severity of ischemia. Antibiotics may be useful; anticoagulants and glucocorticoids are not helpful. If these measures fail, amputation may be required.

■ OTHER VASCULITIDES

Other vasculitides may affect the arteries that supply the upper and lower extremities. **Takayasu arteritis and giant cell (temporal) arteritis are discussed in Chap. 375.**

■ ACUTE LIMB ISCHEMIA

Acute limb ischemia occurs when arterial occlusion results in the sudden cessation of blood flow to an extremity. The severity of ischemia and the viability of the extremity depend on the location and extent of the occlusion and the presence and subsequent development of collateral blood vessels. Principal causes of acute arterial occlusion include embolism, thrombus in situ, arterial dissection, and trauma.

The most common sources of arterial emboli are the heart, aorta, and large arteries. Cardiac disorders that cause thromboembolism include atrial fibrillation; acute myocardial infarction; ventricular aneurysm; cardiomyopathy; infectious and marantic endocarditis; thrombi associated with prosthetic heart valves; and atrial myxoma. Emboli to the distal vessels may also originate from proximal sites of atherosclerosis and aneurysms of the aorta and large vessels. Less frequently, an arterial occlusion results paradoxically from a venous thrombus that has entered the systemic circulation via a patent foramen ovale or another septal defect. Arterial emboli tend to lodge at vessel bifurcations because the vessel caliber decreases at those sites; in the lower extremities, emboli lodge most frequently in the femoral artery, followed by the iliac artery, aorta, and popliteal and tibioperoneal arteries.

Acute arterial thrombosis in situ occurs most frequently in atherosclerotic vessels at the site of an atherosclerotic plaque or aneurysm and in arterial bypass grafts. Trauma to an artery may disrupt continuity of blood flow and cause acute limb ischemia via formation of an acute arterial thrombus or by disruption of an artery's integrity and extravasation of blood. Arterial occlusion may complicate arterial punctures and placement of catheters; it also may result from arterial dissection if the intimal flap obstructs the artery. Less common causes include thoracic outlet compression syndrome, which causes subclavian artery occlusion, and entrapment of the popliteal artery by abnormal placement of the medial head of the gastrocnemius muscle. Polycythemia and hypercoagulable disorders **(Chaps. 108 and 121)** are also associated with acute arterial thrombosis.

■ CLINICAL FEATURES

The symptoms of an acute arterial occlusion depend on the location, duration, and severity of the obstruction. Often severe pain, paresthesia, numbness, and coldness develop in the involved extremity within 1 h. Paralysis may occur with severe and persistent ischemia. Physical findings include loss of pulses distal to the occlusion, cyanosis or pallor, mottling, decreased skin temperature, muscle stiffening, loss of sensation, weakness, and/or absent deep tendon reflexes. If acute arterial occlusion occurs in the presence of an adequate collateral circulation, as is often the case in acute graft occlusion, the symptoms and findings may be less severe. In this situation, the patient complains about an abrupt decrease in the distance walked before claudication occurs or of modest pain and paresthesia. Pallor and coolness are evident, but sensory and motor functions generally are preserved. The clinical evaluation includes Doppler assessment of peripheral blood flow. The diagnosis of acute limb ischemia is usually apparent from the clinical presentation. In most circumstances, duplex ultrasound, MRA, CTA, or catheter-based arteriography is used to confirm the diagnosis and demonstrate the location and extent of arterial occlusion.

TREATMENT

Acute Limb Ischemia

Once the diagnosis is made, the patient should be anticoagulated with intravenous heparin to prevent propagation of the clot and recurrent embolism. In cases of severe ischemia of recent onset, particularly when limb viability is jeopardized, immediate intervention to ensure reperfusion is indicated. Catheter-directed thrombolysis/thrombectomy, surgical thromboembolectomy, and arterial bypass procedures are used to restore blood flow to the ischemic extremity promptly, particularly when a large proximal vessel is occluded.

Intraarterial thrombolytic therapy with recombinant tissue plasminogen activator, reteplase, or tenecteplase is most effective when acute arterial occlusion is recent and caused by a thrombus in an atherosclerotic vessel, arterial bypass graft, or occluded stent. Thrombolytic therapy is also indicated when the patient's overall condition contraindicates surgical intervention or when smaller distal vessels are occluded, thus preventing surgical access. Meticulous observation for hemorrhagic complications is required during intraarterial thrombolytic therapy. Ultrasound-emitting catheters may accelerate reperfusion by improving thrombus permeability to thrombolytic agents. Another endovascular approach to thrombus removal is percutaneous mechanical thrombectomy using devices that employ hydrodynamic forces or rotating baskets to fragment and remove the clot. These treatments may be used alone but usually are used in conjunction with pharmacologic thrombolysis. Surgical revascularization is preferred when restoration of blood flow must occur within 24 h to prevent limb loss. Amputation is performed when the limb is not viable, as characterized by loss of sensation, paralysis, and the absence of Doppler-detected blood flow in both arteries and veins.

Long-term anticoagulation is indicated when acute limb ischemia is caused by cardiac thromboembolism. Emboli resulting from infective endocarditis, the presence of prosthetic heart valves, or atrial myxoma often require surgical intervention to remove the cause.

■ ATHEROEMBOLISM

Atheroembolism is another cause of limb ischemia. In this condition, multiple small deposits of fibrin, platelets, and cholesterol debris embolize from proximal atherosclerotic lesions or aneurysmal sites. Large protruding aortic atheromas are a source of emboli that may lead to limb ischemia, as well as stroke and renal insufficiency. Atheroembolism may occur after intraarterial procedures. Since atheroemboli to limbs tend to lodge in the small vessels of the muscle and skin and may not occlude the large vessels, distal pulses usually remain palpable. Patients complain of acute pain and tenderness at the site of embolization. Digital vascular occlusion may result in ischemia and the "blue toe" syndrome; digital necrosis and gangrene may develop **(Fig. 292-2)**. Localized areas of tenderness, pallor, and livedo reticularis (see below) occur at sites of emboli. Skin or muscle biopsy may demonstrate cholesterol crystals.

Ischemia resulting from atheroemboli is notoriously difficult to treat. Local foot care and occasionally amputation may be needed to treat necrotic areas. Analgesics are indicated for pain relief. Usually neither surgical revascularization procedures nor thrombolytic therapy is helpful because of the multiplicity, composition, and distal location of the emboli. Therapy with antiplatelet drugs and statins improves cardiovascular outcome in patients with atherosclerosis, but it is not established whether either class of drugs prevents recurrent atheroembolism. Similarly, it is not known whether anticoagulant therapy is effective. Endovascular or surgical intervention to exclude or bypass the atherosclerotic vessel or aneurysm that causes the recurrent atheroemboli may be necessary.

■ THORACIC OUTLET COMPRESSION SYNDROME

This is a symptom complex resulting from compression of the neurovascular bundle (artery, vein, or nerves) at the thoracic outlet as it

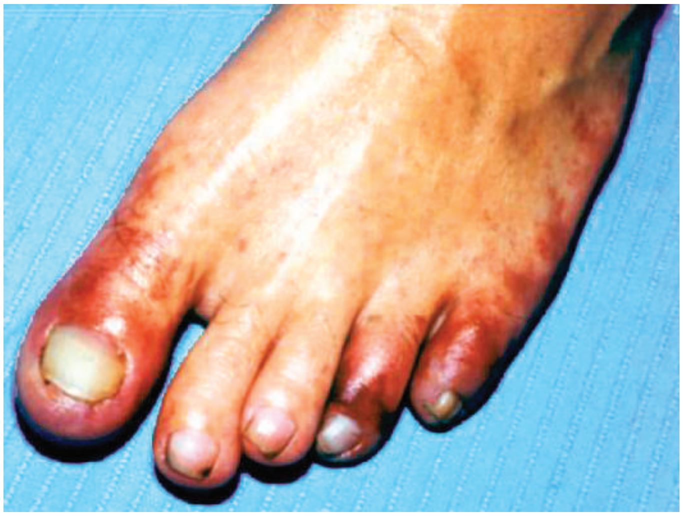

FIGURE 292-2 Atheroembolism causing cyanotic discoloration and impending necrosis of the toes ("blue toe" syndrome).

courses through the neck and shoulder. Cervical ribs, abnormalities of the scalenus anticus muscle, proximity of the clavicle to the first rib, or abnormal insertion of the pectoralis minor muscle may compress the subclavian artery, subclavian vein, and brachial plexus as these structures pass from the thorax to the arm. Depending on the structures affected, thoracic outlet compression syndrome is divided into arterial, venous, and neurogenic forms. Patients with neurogenic thoracic outlet compression may develop shoulder and arm pain, weakness, and paresthesias. Patients with arterial compression may experience claudication, Raynaud phenomenon, and even ischemic tissue loss and gangrene. Venous compression may cause thrombosis of the subclavian and axillary veins; this is often associated with effort and is referred to as *Paget-Schroetter syndrome.*

APPROACH TO THE PATIENT

Arterial Thoracic Outlet Compression Syndrome

Examination of a patient with arterial thoracic outlet compression syndrome is often normal unless provocative maneuvers are performed. Occasionally, distal pulses are decreased or absent and digital cyanosis and ischemia may be evident.

Several maneuvers that support the diagnosis of arterial thoracic outlet compression syndrome may be used to precipitate symptoms, cause a subclavian artery bruit, and diminish arm pulses. These maneuvers include the abduction and external rotation test, in which the affected arm is abducted by 90° and the shoulder is externally rotated; the scalene maneuver (extension of the neck and rotation of the head to the side of the symptoms); the costoclavicular maneuver (posterior rotation of shoulders); and the hyperabduction maneuver (raising the arm 180°). A chest x-ray will indicate the presence of cervical ribs. Duplex ultrasonography, CTA, MRA, and contrast angiography can be performed during provocative maneuvers to demonstrate thoracic outlet compression of the subclavian artery. Neurophysiologic tests such as the electromyogram, nerve conduction studies, and somatosensory evoked potentials may be abnormal if the brachial plexus is involved, but the diagnosis of neurogenic thoracic outlet syndrome is not necessarily excluded if these tests are normal owing to their low sensitivity.

Most patients can be managed conservatively. They should be advised to avoid the positions that cause symptoms. Many patients benefit from shoulder girdle exercises. Surgical procedures such as removal of the first rib and resection of the scalenus anticus muscle are necessary occasionally for relief of symptoms or treatment of ischemia.

POPLITEAL ARTERY ENTRAPMENT

Popliteal artery entrapment typically affects young athletic men and women when the gastrocnemius or popliteus muscle compresses the popliteal artery and causes intermittent claudication. Thrombosis, embolism, or popliteal artery aneurysm may occur. The pulse examination may be normal unless provocative maneuvers such as ankle dorsiflexion and plantar flexion are performed. The diagnosis is confirmed by duplex ultrasound, CTA, MRA, or conventional angiography. Treatment involves surgical release of the popliteal artery or vascular reconstruction.

POPLITEAL ARTERY ANEURYSM

Popliteal artery aneurysms are the most common peripheral artery aneurysms. Approximately 50% are bilateral. Patients with popliteal artery aneurysms often have aneurysms of other arteries, especially the aorta. The most common clinical presentation is limb ischemia secondary to thrombosis or embolism. Rupture occurs less frequently. Other complications include compression of the adjacent popliteal vein or peroneal nerve. Popliteal artery aneurysm can be detected by palpation and confirmed by duplex ultrasonography. Repair is indicated for symptomatic aneurysms or when the diameter exceeds 2–3 cm, owing to the risk of thrombosis, embolism, or rupture.

ARTERIOVENOUS FISTULA

Abnormal communications between an artery and a vein, bypassing the capillary bed, may be congenital or acquired. Congenital arteriovenous fistulas are a result of persistent embryonic vessels that fail to differentiate into arteries and veins; they may be associated with birthmarks, can be located in almost any organ of the body, and frequently occur in the extremities. Acquired arteriovenous fistulas either are created to provide vascular access for hemodialysis or occur as a result of a penetrating injury such as a gunshot or knife wound or as complications of arterial catheterization or surgical dissection. An uncommon cause of arteriovenous fistula is rupture of an arterial aneurysm into a vein.

The clinical features depend on the location and size of the fistula. Frequently, a pulsatile mass is palpable, and a thrill and a bruit lasting throughout systole and diastole are present over the fistula. With long-standing fistulas, clinical manifestations of chronic venous insufficiency, including peripheral edema; large, tortuous varicose veins; and stasis pigmentation become apparent because of the high venous pressure. Evidence of ischemia may occur in the distal portion of the extremity. Skin temperature is higher over the arteriovenous fistula. Large arteriovenous fistulas may result in an increased cardiac output with consequent cardiomegaly and high-output heart failure **(Chap. 264)**.

The diagnosis is often evident from the physical examination. Compression of a large arteriovenous fistula may cause reflex slowing of the heart rate (Nicoladoni-Branham sign). Duplex ultrasonography may detect an arteriovenous fistula, especially one that affects the femoral artery and vein at the site of catheter access. CTA and conventional angiography can confirm the diagnosis and are useful in demonstrating the site and size of the arteriovenous fistula.

Management of arteriovenous fistulas may involve surgery, radiotherapy, or embolization. Congenital arteriovenous fistulas are often difficult to treat because the communications may be numerous and extensive, and new communications frequently develop after ligation of the most obvious ones. Many of these lesions are best treated conservatively using elastic support hose to reduce the consequences of venous hypertension. Occasionally, embolization with autologous material, such as fat or muscle, or with hemostatic agents, such as gelatin sponges or silicon spheres, is used to obliterate the fistula. Acquired arteriovenous fistulas are usually amenable to surgical treatment that involves division or excision of the fistula. Occasionally, autogenous or synthetic grafting is necessary to reestablish continuity of the artery and vein.

RAYNAUD PHENOMENON

Raynaud phenomenon is characterized by episodic digital ischemia, manifested clinically by the sequential development of digital

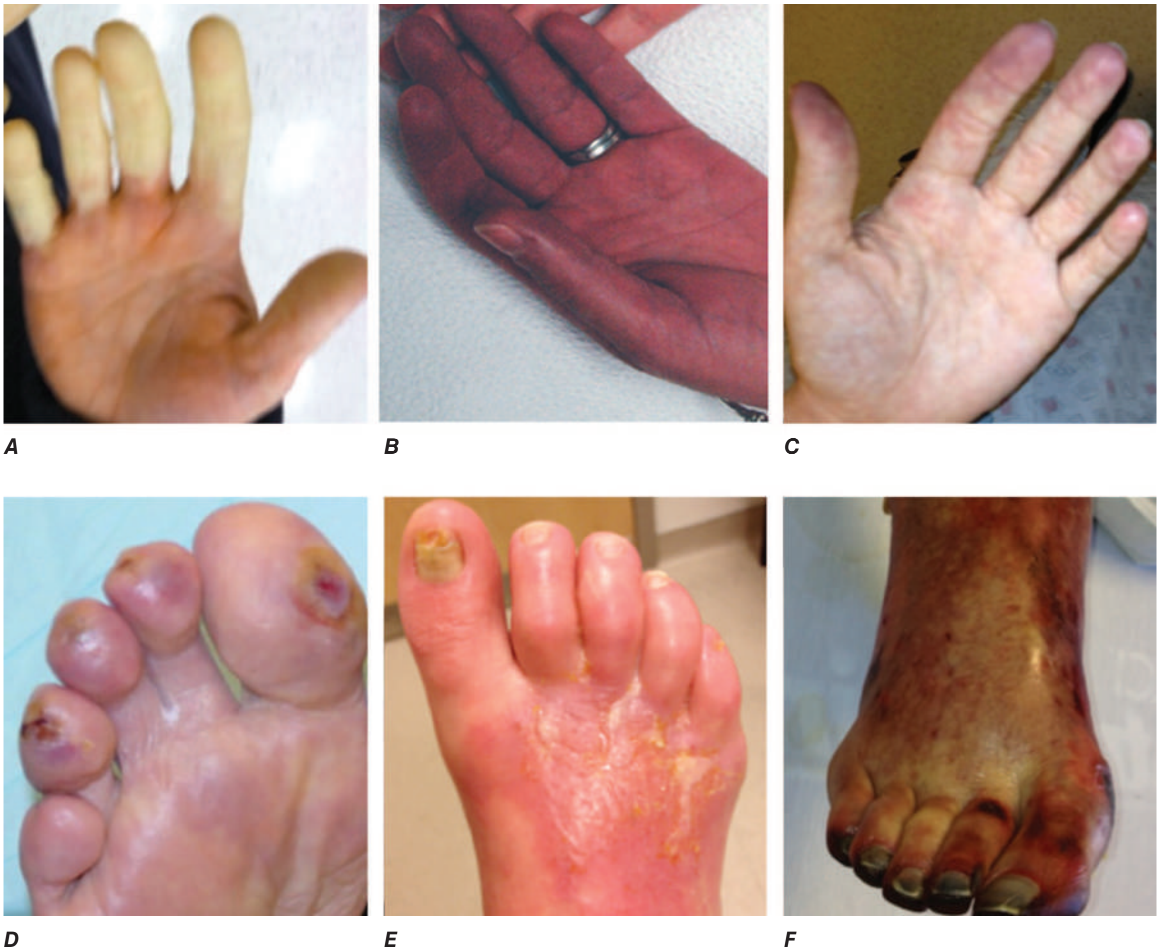

FIGURE 292-3 Vascular diseases associated with temperature: ***A.*** Raynaud phenomenon; ***B.*** acrocyanosis; ***C.*** livedo reticularis; ***D.*** pernio; ***E.*** erythromelalgia; and ***F.*** frostbite.

blanching, cyanosis, and rubor of the fingers or toes after cold exposure and subsequent rewarming. Emotional stress may also precipitate Raynaud phenomenon. The color changes are usually well demarcated and are confined to the fingers or toes. Typically, one or more digits will appear white when the patient is exposed to a cold environment or touches a cold object **(Fig. 292-3*A*)**. The blanching, or pallor, represents the ischemic phase of the phenomenon and results from vasospasm of digital arteries. During the ischemic phase, capillaries and venules dilate, and cyanosis results from the deoxygenated blood that is present in these vessels. A sensation of cold or numbness or paresthesia of the digits often accompanies the phases of pallor and cyanosis.

With rewarming, the digital vasospasm resolves, and blood flow into the dilated arterioles and capillaries increases dramatically. This "reactive hyperemia" imparts a bright red color to the digits. In addition to rubor and warmth, patients often experience a throbbing, painful sensation during the hyperemic phase. Although the triphasic color response is typical of Raynaud phenomenon, some patients may develop only pallor and cyanosis; others may experience only cyanosis.

Raynaud phenomenon is broadly separated into two categories: idiopathic, termed primary Raynaud phenomenon, and secondary Raynaud phenomenon, which is associated with other disease states or known causes of vasospasm **(Table 292-1)**.

Primary Raynaud Phenomenon This appellation is applied when the secondary causes of Raynaud phenomenon have been excluded. Over 50% of patients with Raynaud phenomenon have the primary form. Women are affected about five times more often than men, and the age of presentation is usually between 20 and 40 years. The fingers are involved more frequently than the toes. Initial episodes may involve only one or two fingertips, but subsequent attacks may involve the entire finger and may include all the fingers. The toes are affected in 40% of patients. Although vasospasm of the toes usually occurs in patients with symptoms in the fingers, it may happen alone. Rarely, the earlobes, the tip of the nose, tongue, nipple, or penis are involved. Raynaud phenomenon occurs frequently in patients who also have migraine headaches or variant angina from coronary vasospasm. These associations suggest that there may be a common predisposing cause for the vasospasm.

Results of physical examination are often entirely normal; the radial, ulnar, and pedal pulses are normal. The fingers and toes may be cool between attacks and may perspire excessively. Nailfold capillaroscopy reveals normal superficial capillaries, which appear as regularly spaced hairpin loops. Thickening and tightening of the digital subcutaneous tissue (*sclerodactyly*) develop in 10% of patients. Angiography of the digits for diagnostic purposes is not indicated.

TABLE 292-1 Classification of Raynaud Phenomenon

Primary or idiopathic Raynaud phenomenon

Secondary Raynaud phenomenon

- Collagen vascular diseases: scleroderma, systemic lupus erythematosus, rheumatoid arthritis, dermatomyositis, polymyositis, mixed connective tissue disease, Sjögren syndrome
- Arterial occlusive diseases: atherosclerosis of the extremities, thromboangiitis obliterans, acute arterial occlusion, thoracic outlet syndrome
- Pulmonary hypertension
- Neurologic disorders: intervertebral disk disease, syringomyelia, spinal cord tumors, stroke, poliomyelitis, carpal tunnel syndrome, complex regional pain syndrome
- Blood dyscrasias: cold agglutinins, cryoglobulinemia, cryofibrinogenemia, myeloproliferative disorders, lymphoplasmacytic lymphoma
- Trauma: vibration injury, hammer hand syndrome, electric shock, cold injury, typing, piano playing
- Drugs and toxins: ergot derivatives, methysergide, β-adrenergic receptor blockers, bleomycin, vinblastine, cisplatin, gemcitabine, vinyl chloride

In general, patients with primary Raynaud disease have milder clinical manifestations. Fewer than 1% of these patients lose a part of a digit. After the diagnosis is made, the disease improves spontaneously in ~15% of patients and progresses in ~30%.

Secondary Causes of Raynaud Phenomenon Raynaud phenomenon occurs in 80–90% of patients with systemic sclerosis (scleroderma) and is the presenting symptom in 30% **(Chap. 372)**. It may be the only symptom of scleroderma for many years. Abnormalities of the digital vessels may contribute to the development of Raynaud phenomenon in this disorder. Ischemic fingertip ulcers may develop and progress to gangrene and autoamputation. About 20% of patients with systemic lupus erythematosus (SLE) have Raynaud phenomenon **(Chap. 368)**. Occasionally, persistent digital ischemia develops and may result in ulcers or gangrene. In most severe cases, the small vessels are occluded by a proliferative endarteritis. Raynaud phenomenon occurs in ~30% of patients with dermatomyositis or polymyositis **(Chap. 377)**. It frequently develops in patients with rheumatoid arthritis and may be related to the intimal proliferation that occurs in the digital arteries.

Atherosclerosis of the extremities is a common cause of Raynaud phenomenon in men aged >50 years. Thromboangiitis obliterans is an uncommon cause of Raynaud phenomenon but should be considered in young men, particularly those who are cigarette smokers. The development of cold-induced pallor in these disorders may be confined to one or two digits of the involved extremity. Occasionally, Raynaud phenomenon may follow acute occlusion of large and medium-sized arteries by a thrombus or embolus. Embolization of atheroembolic debris may cause digital ischemia. The latter situation often involves one or two digits and should not be confused with Raynaud phenomenon. In patients with thoracic outlet compression syndrome, Raynaud phenomenon may result from diminished intravascular pressure, stimulation of sympathetic fibers in the brachial plexus, or a combination of both. Raynaud phenomenon occurs in patients with primary pulmonary hypertension **(Chap. 294)**; this is more than coincidental and may reflect a neurohumoral abnormality that affects both the pulmonary and digital circulations.

A variety of blood dyscrasias may be associated with Raynaud phenomenon. Cold-induced precipitation of plasma proteins, hyperviscosity, and aggregation of red cells and platelets may occur in patients with cold agglutinins, cryoglobulinemia, or cryofibrinogenemia. Hyperviscosity syndromes that accompany myeloproliferative disorders and lymphoplasmacytic lymphoma (Waldenström macroglobulinemia) should also be considered in the initial evaluation of patients with Raynaud phenomenon.

Raynaud phenomenon occurs often in patients whose vocations require the use of vibrating hand tools, such as chain saws or jackhammers. The frequency of Raynaud phenomenon also seems to be increased in pianists and keyboard operators. Electric shock injury to the hands or frostbite may lead to the later development of Raynaud phenomenon.

Several drugs have been causally implicated in Raynaud phenomenon. They include ergot preparations, methysergide, β-adrenergic receptor antagonists, and the chemotherapeutic agents bleomycin, vinblastine, cisplatin, and gemcitabine.

TREATMENT

Raynaud Phenomenon

Most patients with Raynaud phenomenon experience only mild and infrequent episodes. These patients need reassurance and should be instructed to dress warmly and avoid unnecessary cold exposure. In addition to gloves and mittens, patients should protect the trunk, head, and feet with warm clothing to prevent cold-induced reflex vasoconstriction. Tobacco use is contraindicated.

Drug treatment should be reserved for severe cases. Dihydropyridine calcium channel antagonists such as nifedipine and amlodipine decrease the frequency and severity of Raynaud phenomenon. Diltiazem may be considered but is less effective. The postsynaptic α_1-adrenergic antagonist prazosin has been used with favorable responses; doxazosin and terazosin may also be effective. Phosphodiesterase type 5 inhibitors such as sildenafil, tadalafil, and vardenafil may improve symptoms in patients with secondary Raynaud phenomenon, as occurs with systemic sclerosis. There is also evidence that topical nitroglycerin preparations are effective. Digital sympathectomy is helpful in some patients who are unresponsive to medical therapy. Injection of botulinum toxin into the perivascular tissue of the wrist or palm improved ischemic manifestations of severe Raynaud phenomenon in case series, although controlled clinical trials have yielded inconsistent results.

ACROCYANOSIS

In this condition, there is arterial vasoconstriction and secondary dilation of the capillaries and venules with resulting persistent cyanosis of the hands and, less frequently, the feet. Cyanosis may be intensified by exposure to a cold environment. Acrocyanosis may be categorized as primary or secondary to an underlying condition. In primary acrocyanosis, women are affected much more frequently than men, and the age of onset is usually <30 years. Generally, patients are asymptomatic but seek medical attention because of the discoloration. The prognosis is favorable, and pain, ulcers, and gangrene do not occur. Examination reveals normal pulses, peripheral cyanosis, and moist palms **(Fig. 292-3*B*)**. Trophic skin changes and ulcerations do *not* occur. The disorder can be distinguished from Raynaud phenomenon because it is persistent and not episodic, the discoloration extends proximally from the digits, and blanching does not occur. Ischemia secondary to arterial occlusive disease can usually be excluded by the presence of normal pulses. Central cyanosis and decreased arterial oxygen saturation are not present. Patients should be reassured and advised to dress warmly and avoid cold exposure. Pharmacologic intervention is not indicated.

Secondary acrocyanosis may result from hypoxemia, vasopressor medications, connective tissue diseases, atheroembolism, antiphospholipid antibodies, cold agglutinins, or cryoglobulins and is associated with anorexia nervosa and postural orthostatic tachycardia syndrome. Treatment should be directed at the underlying disorder.

LIVEDO RETICULARIS

In this condition, localized areas of the extremities develop a mottled or rete (netlike) appearance of reddish to blue discoloration **(Fig. 293-3*C*)**. There are primary and secondary forms of livedo reticularis. The primary, or idiopathic, form of this disorder may be benign or associated with ulcerations. The benign form occurs more frequently in women than in men, and the most common age of onset is the third decade. The mottling typically is symmetric and uniform and may be more prominent after cold exposure and improve with warming. Patients with the benign form are usually asymptomatic and seek attention for cosmetic reasons. These patients should be reassured and advised to avoid cold environments. No drug treatment is indicated. Primary livedo reticularis with ulceration is also called *atrophie blanche en plaque*. The ulcers are painful and may take months to heal. Secondary livedo reticularis can occur with atheroembolism (see above), SLE and other vasculitides, antiphospholipid antibodies, hyperviscosity, cryoglobulinemia, and Sneddon's syndrome (ischemic stroke and livedo reticularis). Livedo racemosa is the term used to characterize secondary livedo reticularis, when the mottling is irregular and disrupted, and does not improve with warming. Rarely, skin ulcerations develop.

PERNIO (CHILBLAINS)

Pernio is a vasculitic disorder associated with exposure to cold; acute forms have been described. Raised erythematous lesions develop most commonly on the toes or fingers in cold weather **(Fig. 292-3*D*)**. They are associated with pruritus and a burning sensation, and they may blister and ulcerate. Pathologic examination demonstrates angiitis characterized by intimal proliferation and perivascular infiltration of mononuclear and polymorphonuclear leukocytes. Giant cells may be present in the subcutaneous tissue. Patients should avoid exposure to cold, and ulcers should be kept clean and protected with sterile dressings. Sympatholytic drugs and dihydropyridine calcium channel antagonists may be effective in some patients.

ERYTHROMELALGIA

This disorder is characterized by burning pain and erythema of the extremities **(Fig. 292-3*E*)**. The feet are involved more frequently than the hands, and males are affected more frequently than females. Erythromelalgia may occur at any age but is most common in middle age. It may be primary (also termed erythermalgia) or secondary. Mutations in the *SCN9A* gene, which encodes the Nav1.7 voltage-gated sodium channel expressed in sensory and sympathetic nerves, have been described in inherited forms of erythromelalgia. The most common causes of secondary erythromelalgia are myeloproliferative disorders such as polycythemia vera and essential thrombocytosis. Less common causes include drugs, such as calcium channel blockers, bromocriptine, and pergolide; neuropathies; connective tissue diseases such as SLE; and paraneoplastic syndromes. Patients complain of burning in the extremities that is precipitated by exposure to a warm environment and aggravated by a dependent position. The symptoms are relieved by exposing the affected area to cool air or water or by elevation. Erythromelalgia can be distinguished from ischemia secondary to peripheral arterial disorders because the peripheral pulses are present. There is no specific treatment; aspirin may produce relief in patients with erythromelalgia secondary to myeloproliferative disease. Topical anesthetics, such as lidocaine, or a combination of amitriptyline and ketamine may be considered to relieve pain. Topical midodrine may also be considered. Treatment of associated disorders in secondary erythromelalgia may be helpful.

FROSTBITE

In this condition, tissue damage results from severe environmental cold exposure or from direct contact with a very cold object. Tissue injury results from both freezing and vasoconstriction. Frostbite usually affects the distal aspects of the extremities or exposed parts of the face, such as the ears, nose, chin, and cheeks. Superficial frostbite involves the skin and subcutaneous tissue. Patients experience pain or paresthesia, and the skin appears white and waxy. After rewarming, there is cyanosis and erythema, wheal-and-flare formation, edema, and superficial blisters. Deep frostbite involves muscle, nerves, and deeper blood vessels. It may result in edema of the hand or foot, vesicles and bullae, tissue necrosis, and gangrene **(Fig. 292-3*F*)**.

Initial treatment is rewarming, performed in an environment where reexposure to freezing conditions will not occur. Rewarming is accomplished by immersion of the affected part in a water bath at temperatures of 40°–44°C (104°–111°F). Massage, application of ice water, and extreme heat are contraindicated. The injured area should be cleansed with soap or antiseptic, and sterile dressings should be applied. Analgesics are often required during rewarming. Antibiotics are used if there is evidence of infection. The efficacy of sympathetic blocking drugs is not established. After recovery, the affected extremity may exhibit increased sensitivity to cold.

FURTHER READING

Aboyans V et al: Antithrombotic therapies in aortic and peripheral arterial diseases in 2021: A consensus document from the ESC working group on aorta and peripheral vascular diseases, the ESC working group on thrombosis, and the ESC working group on cardiovascular pharmacotherapy. Eur Heart J 42:4013, 2021.

Aday AW, Matsushita K: Epidemiology of peripheral artery disease and polyvascular disease. Circ Res 128:1818, 2021.

Bonaca M et al: Contemporary medical management of peripheral artery disease. Circ Res 128:1868, 2021.

Choi E, Henkin S: Raynaud's phenomenon and related vasospastic disorders. Vasc Med 26:56, 2021.

Creager MA et al: Reducing nontraumatic lower-extremity amputations by 20% by 2030: Time to get to our feet: A policy statement from the American Heart Association. Circulation 143:e875, 2021.

GBD 2019 Peripheral Artery Disease Collaborators: Global burden of peripheral artery disease and its risk factors, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. Lancet Glob Health 11:e1553, 2023.

Gornik HL et al: First International Consensus on the diagnosis and management of fibromuscular dysplasia. Vasc Med 24:164, 2019.

Gornik HL et al: 2024 AHA/ACC guideline on the management of patients with lower extremity peripheral artery disease: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation. JACC 83:2497, 2024.

Polonsky TS, McDermott MM: Lower extremity peripheral artery disease without chronic limb-threatening ischemia. A review. JAMA 325:2188, 2021.

293

Chronic Venous Disease and Lymphedema

Mark A. Creager, Robert T. Eberhardt,
Joseph Loscalzo

CHRONIC VENOUS DISEASE

Chronic venous diseases range from telangiectasias and reticular veins, to varicose veins, to chronic venous insufficiency with edema, skin changes, and ulceration. This section of the chapter will focus on identification and treatment of varicose veins and chronic venous insufficiency, since these problems are encountered frequently by the internist. The estimated prevalence of varicose veins in the United States is ~15% in men and 30% in women. Chronic venous insufficiency with edema affects ~7.5% of men and 5% of women, and the prevalence increases with age ranging from 2% among those <50 years of age to 10% of those 70 years of age. Approximately 20% of patients with chronic venous insufficiency develop venous ulcers.

VENOUS ANATOMY

Veins in the extremities can be broadly classified as either superficial or deep. The superficial veins are located between the skin and deep fascia. In the legs, these include the great and small saphenous veins and their tributaries. The great saphenous vein is the longest vein in the body. It originates on the medial side of the foot and ascends anterior to the medial malleolus and then along the medial side of the calf and thigh and drains into the common femoral vein. The small saphenous vein originates on the dorsolateral aspect of the foot, ascends posterior to the lateral malleolus and along the posterolateral aspect of the calf, and drains into the popliteal vein. The deep veins of the leg accompany the major arteries. There are usually paired peroneal, anterior tibial, and posterior tibial veins in the calf, which converge to form the popliteal vein. Soleal tributary veins drain into the posterior tibial or peroneal veins, and gastrocnemius tributary veins drain into the popliteal vein. The popliteal vein ascends in the thigh as the femoral vein. The confluence of the femoral vein and deep femoral vein form the common femoral vein, which ascends in the pelvis as the external iliac and then common iliac vein, which converges with the contralateral common iliac vein at the inferior vena cava. Perforating veins connect the superficial and deep systems in the legs at multiple locations, normally allowing blood to flow from the superficial to deep veins. In the arms, the superficial veins include the basilic, cephalic, and median cubital veins and their tributaries. The basilic and cephalic veins course along the medial and lateral aspects of the arm, respectively, and these are connected via the median cubital vein in the antecubital fossa. The deep veins of the arms accompany the major arteries and include the radial, ulnar, brachial, axillary, and subclavian veins. The subclavian vein converges with the internal jugular vein to form the brachiocephalic vein, which joins the contralateral brachiocephalic vein to form the superior vena cava. Bicuspid valves are present throughout the venous system to direct the flow of venous blood centrally.

Pathophysiology of Chronic Venous Disease *Varicose veins* are dilated, bulging, tortuous superficial veins, measuring at least 3 mm in diameter. The smaller and less tortuous reticular veins are dilated intradermal veins, which appear blue-green, measure 1–3 mm in diameter, and do not protrude from the skin surface. Telangiectasias, or spider veins, are small, dilated veins, <1 mm in diameter, located near the skin surface, and form blue, purple, or red linear, branching, or spider-web patterns.

Varicose veins can be categorized as primary or secondary. Primary varicose veins originate in the superficial system and result from defective structure and function of the valves of the saphenous veins, intrinsic weakness of the vein wall, and high intraluminal pressure. Approximately one-half of these patients have a family history of varicose veins. Other factors associated with primary varicose veins include aging, pregnancy, hormonal therapy, obesity, and prolonged standing. Secondary varicose veins result from venous hypertension, associated with deep-venous insufficiency or deep-venous obstruction, and incompetent perforating veins that cause enlargement of superficial veins. Arteriovenous fistulas also cause varicose veins in the affected limb.

Chronic venous insufficiency is a consequence of incompetent veins in which there is venous hypertension and extravasation of fluid and blood elements into the tissue of the limb. It may occur in patients with varicose veins and superficial venous disease, but more advanced manifestations are usually caused by disease in the deep veins. It also is categorized as primary or secondary. Primary venous insufficiency is a consequence of an intrinsic structural or functional abnormality in the vein wall or venous valves leading to valvular reflux. Secondary deep-venous insufficiency is caused by obstruction and/or valvular incompetence from previous deep-vein thrombosis **(Chap. 290)**. Deep-venous insufficiency occurs following deep-vein thrombosis, as the delicate valve leaflets become thickened and contracted and can no longer prevent retrograde flow of blood and the vein itself becomes rigid and thick walled. Although most veins recanalize after an episode of thrombosis, the large proximal veins may remain occluded. Secondary incompetence develops in valves distal to the obstruction because high pressures distend the vein and separate the leaflets. Other causes of secondary deep-venous insufficiency include May-Thurner syndrome, where the left iliac vein is occluded or stenosed by extrinsic compression from the overlapping right common iliac artery; extrinsic compression from tumor or retroperitoneal fibrosis; arteriovenous fistulas resulting in increased venous pressure; congenital deep-vein agenesis or hypoplasia; and venous malformations as may occur in Klippel-Trénaunay and Parkes-Weber syndromes.

Clinical Presentation Patients with venous varicosities are often asymptomatic but still concerned about the cosmetic appearance of their legs. Superficial venous thrombosis may be a recurring problem, and rarely, a varicosity ruptures and bleeds. Symptoms in patients with varicose veins or venous insufficiency, when they occur, include a dull ache, throbbing or heaviness, or pressure sensation in the legs typically after prolonged standing; these symptoms usually are relieved with leg elevation. Additional symptoms may include cramping, burning, pruritus, leg swelling, and skin ulceration.

The legs are examined in both the supine and standing positions. Visual inspection and palpation of the legs in the standing position confirm the presence of varicose veins. The location and extent of the varicose veins should be noted. Chronic venous insufficiency is characterized by the development of edema or other skin manifestations. Edema, stasis dermatitis, and skin ulceration near the ankle may be present if there is superficial venous insufficiency and venous hypertension. Findings of deep-venous insufficiency include increased leg circumference, venous varicosities, edema, and skin changes. The edema, which is usually pitting, may be confined to the ankles, extend above the ankles to the knees, or involve the thighs in severe cases. Over time, the edema may become less pitting and more indurated, particularly with the secondary development of lymphatic dysfunction. Dermatologic findings associated with venous stasis include hyperpigmentation, erythema, eczema, lipodermatosclerosis, *atrophie blanche*, and a phlebectasia corona. Lipodermatosclerosis is the combination of induration, hemosiderin deposition, and inflammation, and typically occurs in the lower part of the leg just above the ankle. Atrophie blanche is a white patch of scar tissue, often with focal telangiectasias and a hyperpigmented border; it usually develops near the medial malleolus. A phlebectasia corona is a fan-shaped pattern of intradermal veins near the ankle or on the foot. Skin ulceration may occur near the medial and lateral malleoli. A venous ulcer is often shallow and characterized by an irregular border, a base of granulation tissue, and the presence of exudate **(Fig. 293-1)**.

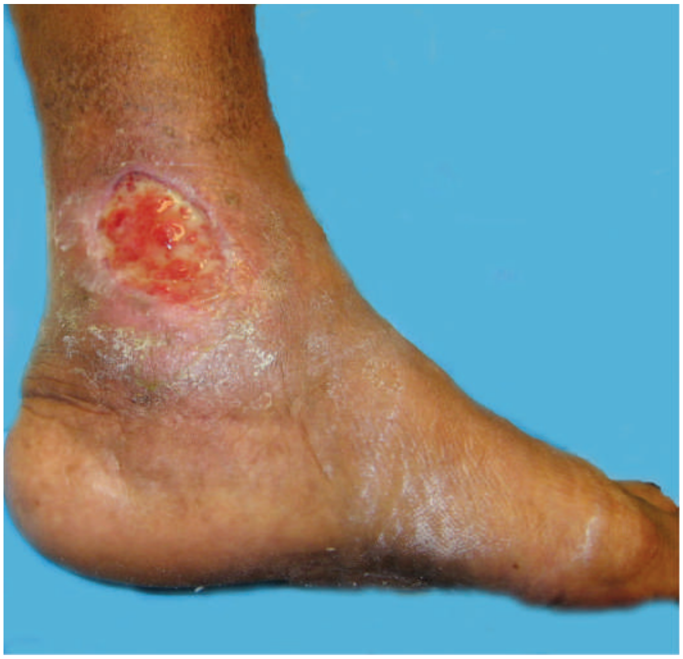

FIGURE 293-1 Venous insufficiency with active venous ulcer near the medial malleolus. *(Courtesy of Dr. Steven Dean, with permission.)*

Bedside maneuvers can be used to distinguish primary varicose veins from secondary varicose veins caused by deep-venous insufficiency. With the contemporary use of venous ultrasound (see below), however, these maneuvers are employed infrequently. The Brodie-Trendelenburg test is used to determine whether varicose veins are secondary to deep-venous insufficiency. As the patient is lying supine, the leg is elevated and the veins allowed to empty. Then, a tourniquet is placed on the proximal part of the thigh and the patient is asked to stand. Filling of the varicose veins within 30 s indicates that the varicose veins are caused by deep-venous insufficiency and incompetent perforating veins. Primary varicose veins with superficial venous insufficiency are the likely diagnosis if venous refilling occurs promptly after tourniquet removal. The Perthes test assesses the possibility of deep-venous obstruction. A tourniquet is placed on the midthigh after the patient has stood, and the varicose veins are filled. The patient is then instructed to walk for 5 min. A patent deep-venous system and competent perforating veins enable the superficial veins below the tourniquet to collapse. Deep-venous obstruction is likely to be present if the superficial veins distend further with walking.

Differential Diagnosis The duration of leg edema helps to distinguish chronic venous insufficiency from acute deep-vein thrombosis. Lymphedema, as discussed later in this chapter, is often confused with chronic venous insufficiency, and both may occur together when venous insufficiency impairs lymphatic function leading to phlebolymphedema. Other disorders that cause leg swelling should be considered and excluded when evaluating a patient with presumed venous insufficiency. Bilateral leg swelling occurs in patients with congestive heart failure, hypoalbuminemia secondary to nephrotic syndrome or severe hepatic disease, or myxedema caused by hypothyroidism or pretibial myxedema associated with Graves' disease, and with drugs such as dihydropyridine calcium channel blockers and thiazolidinediones. Unilateral causes of leg swelling also include ruptured leg muscles, hematomas secondary to trauma, and popliteal cysts. Cellulitis may cause erythema and swelling of the affected limb. Leg ulcers may be caused by severe peripheral artery disease and chronic limb threatening ischemia; neuropathies, particularly those associated with diabetes; and less commonly, skin cancer, vasculitis, or rarely as a complication of hydroxyurea. The location and characteristics of venous ulcers help to differentiate these from other causes.

Classification of Chronic Venous Disease The CEAP (clinical, etiologic, anatomic, pathophysiologic) classification schema incorporates the range of symptoms and signs of chronic venous disease to characterize its severity. It also broadly categorizes the etiology as primary, secondary, or congenital; identifies the affected veins as superficial, deep, or perforating; and characterizes the pathophysiology as reflux, obstruction, both, or neither **(Table 293-1)**.

Diagnostic Testing The principal diagnostic test to evaluate patients with chronic venous disease is venous duplex ultrasonography. A venous duplex ultrasound examination uses a combination of B-mode imaging and spectral Doppler to detect the presence of venous obstruction and venous reflux in superficial and deep veins. Color-assisted Doppler ultrasound is useful to visualize venous flow patterns. Obstruction may be diagnosed by the absence of flow, the presence of an echogenic thrombus within the vein, or failure of the vein to collapse when a compression maneuver is applied by the sonographer, the last implicating the presence of an intraluminal thrombus. Venous reflux is detected by prolonged reversal of venous flow direction during a Valsalva maneuver, particularly for the common femoral vein or saphenofemoral junction, or after compression and release of a cuff placed on the limb distal to the area being interrogated.

TABLE 293-1 CEAP (Clinical, Etiologic, Anatomic, Pathophysiologic) Classification

Clinical Classification	
C_0	No visible or palpable signs of venous disease
C_1	Telangiectasias or reticular veins
C_2	Varicose veins
C_{2r}	Recurrent varicose veins
C_3	Edema
C_4	Changes in skin and subcutaneous secondary to CVD
C_{4a}	Pigmentation or eczema
C_{4b}	Lipodermatosclerosis or atrophie blanche
C_{4c}	Corona phlebectatica
C_5	Healed venous ulcer
C_6	Active venous ulcer
C_{6r}	Recurrent active venous ulcer
Etiologic Classification	
E_p	Primary
E_s	Secondary
E_{si}	Secondary – intravenous
E_{se}	Secondary – extravenous
E_c	Congenital
E_n	No cause identified
Anatomic Classification	
A_s	Superficial
A_p	Perforator
A_d	Deep
A_n	No venous anatomic location identified
Pathophysiologic Classification	
P_r	Reflux
P_o	Obstruction
$P_{r,o}$	Reflux and obstruction
P_n	No pathophysiology identified

Abbreviation: CVD, chronic venous disease.

Source: Data from F Lurie et al: J Vasc Surg 8:342, 2020.

Some vascular laboratories use air or strange gauge plethysmography to assess the severity of venous reflux and complement findings from the venous ultrasound examination. Venous volume and venous refilling time are measured when the legs are placed in a dependent position and after calf exercise to quantify the severity of venous reflux and the efficiency of the calf muscle pump to affect venous return.

Magnetic resonance, computed tomographic, and conventional venography are rarely required to determine the cause and plan treatment for chronic venous insufficiency unless there is suspicion for pathology that might warrant intervention. These modalities are used to identify obstruction or stenosis of the inferior vena cava and iliofemoral veins, as may occur in patients with previous proximal deep-vein thrombosis; occlusion of inferior vena cava filters; extrinsic compression from tumors; and May-Thurner syndrome.

TREATMENT

Chronic Venous Disease

SUPPORTIVE MEASURES

Varicose veins usually are treated with conservative measures. Symptoms often decrease when the legs are elevated periodically, prolonged standing is avoided, and elastic support hose are worn. External compression with elastic stockings, multilayer elastic wraps, stretch bandages, or inelastic garments provides a counterbalance to the hydrostatic pressure in the veins. Although compression garments may improve symptoms and are recommended in patients with healed or active venous ulceration, they are not curative and do not prevent progression of varicose veins. Graduated compression stockings with pressures of 20–30 mmHg are suitable for most patients with simple varicose veins, although higher pressures may be required for patients with varicose veins and manifestations of venous insufficiency such as edema and ulcers.

Patients with chronic venous insufficiency also should be advised to avoid prolonged standing or sitting; frequent leg elevation is helpful. Graded compression therapy consisting of stockings or multilayered compression bandages is the standard of care for advanced chronic venous insufficiency characterized by edema, skin changes, or venous ulcers defined as CEAP clinical class C3–C6. Graduated compression stockings of 30–40 mmHg are more effective than lesser grades for healing venous ulcers. The length of stocking depends on the distribution of edema. Calf-length stockings are tolerated better by most patients, particularly elderly patients; for patients with varicose veins or edema extending to the thigh, thigh-length stockings or panty hose should be considered. Exercise training, including leg muscle strengthening, may improve calf muscle pump function and antegrade venous flow, and reduce the severity of chronic venous insufficiency. Overweight and obese patients should be advised to lose weight via caloric restriction and exercise or consult an obesity specialist for more advanced therapies.

In addition to a compression bandage or stocking, patients with venous ulcers also may be treated with low-adherent absorbent dressings that take up exudates while maintaining a moist environment. Other types of dressings include hydrocolloid (an adhesive dressing composed of polymers such as carboxymethylcellulose that absorbs exudates by forming a gel), hydrogel (a nonabsorbent dressing comprising >80% water or glycerin that moisturizes wounds), foam (an absorbent dressing made with polymers such as polyurethane), and alginate (an absorbent, biodegradable dressing that is derived from seaweed), but there is little evidence that these are more effective than low-adherent absorbent dressings. The choice of specific dressing depends on the amount of drainage, presence of infection, and integrity of the skin surrounding the ulcer. Ulcers should be debrided of necrotic tissue. Antibiotics are not indicated unless the ulcer is infected. The multilayered compression bandage or graduated compression garment is then put over the dressing.

MEDICAL THERAPIES

There are no drugs approved by the U.S. Food and Drug Administration for the treatment of chronic venous insufficiency. Diuretics may reduce edema, but at the risk of volume depletion and compromise in renal function. Topical steroids may be used for a short period of time to treat inflammation associated with stasis dermatitis. Several herbal supplements, such as horse chestnut seed extract (aescin); ruscus extract; flavonoids, including diosmin, hesperidin, or the two combined as micronized purified flavonoid fraction; and French maritime pine bark extract, are touted to have venoconstrictive and anti-inflammatory properties. Several meta-analyses have suggested that ruscus extract and micronized purified flavonoid fraction reduce pain, leg heaviness, and/or sensation of swelling, and that micronized purified flavonoid fraction, and perhaps pentoxifylline, in conjunction with compression therapy facilitates venous ulcer healing; however, there remains insufficient evidence to recommend the general use of these substances in patients with chronic venous insufficiency.

INTERVENTIONAL AND SURGICAL THERAPIES

Ablative procedures, including endovenous thermal and nonthermal ablation, sclerotherapy, and surgery, are used to treat varicose veins in selected patients who have persistent symptoms, great saphenous vein incompetency, and complications of venous insufficiency including dermatitis, edema, and ulcers in order to treat symptoms, accelerate healing, and prevent recurrence. Ablative therapy may also be indicated for cosmetic reasons.

Endovenous thermal ablation procedures of the saphenous veins include endovenous laser and radiofrequency ablation. To ablate the great saphenous vein, a catheter is placed percutaneously and advanced from the level of the knee to just below the saphenofemoral junction via ultrasound guidance. Thermal energy is then delivered as the catheter is pulled back. The heat injures the endothelium and media and promotes thrombosis and fibrosis, resulting in venous occlusion. Average 1- and 5-year occlusion rates exceed 90% following endovenous laser therapy and are slightly less after radiofrequency ablation. Deep-vein thrombosis of the common femoral vein adjacent to the saphenofemoral junction is an uncommon but potential complication of endovenous thermal ablation. Other adverse effects of thermal ablation procedures include pain, paresthesias, bruising, hematoma, and hyperpigmentation.

Nonthermal ablation procedures of the saphenous veins include endovenous delivery of a cyanoacrylate tissue adhesive, which causes fibrosis, and mechanochemical ablation, which involves insertion of a rotating wire to injure the endothelium and infusion of a liquid sclerosant. One-year occlusion rates approximate or exceed 90%, respectively. Adverse effects of nonthermal ablation procedures include superficial thrombophlebitis, deep vein thrombosis, ecchymoses, hematomas, and hyperpigmentation.

Sclerotherapy involves the injection of a chemical into a vein to cause fibrosis and obstruction. Sclerosing agents approved by the U.S. Food and Drug Administration include sodium tetradecyl sulfate, polidocanol, sodium morrhuate, and glycerin. The sclerosing agent is administered as a liquid or mixed with air or CO_2/O_2 to create a foam. It may be used to treat the great saphenous vein, but most commonly sclerotherapy is used to treat the affected tributaries of the great saphenous vein or incompetent perforating veins either as a standalone procedure or concomitant or staged with ablative procedures of the truncal superficial veins. Following completion of the procedure, elastic bandages are applied, or 30–40 mmHg compression stockings are worn for 1–2 weeks. Average 1- and 5-year occlusion rates are 81 and 74%, respectively, following sclerotherapy. Complications are uncommon and include deep-vein thrombosis, hematomas, damage to adjacent saphenous or sural nerves, and infection. Anaphylaxis is a very rare but severe complication.

Surgical therapy usually involves ligation and stripping of the great and small saphenous veins. The procedure is performed under general anesthesia. Incisions are made at the groin and the upper calf. The great saphenous vein is ligated below the saphenofemoral junction, and a wire is inserted into the great saphenous vein and advanced distally. The proximal part of the great saphenous vein is secured to the wire and retrieved, i.e., stripped, via the calf incision. Stripping of the great saphenous vein below the knee and stripping of the small saphenous vein usually are not performed because of the respective risks of saphenous and sural nerve injury. Complications of great saphenous vein ligation and stripping include deep-vein thrombosis, bleeding, hematoma, infection, and nerve injury. Recurrent varicose veins occur in up to 50% patients by 5 years, due to technical failures, deep-venous insufficiency, and incompetent perforating veins. Endovenous ablation techniques are generally favored over ligation and stripping.

Stab phlebectomy is another surgical treatment for varicose veins. A small incision is made alongside the varicose vein, and it is avulsed by means of a forceps or hook. This procedure may be performed in conjunction with saphenous vein ligation and stripping or thermal or nonthermal ablation. Subfascial endoscopic perforator surgery (SEPS) uses endoscopy to identify and occlude incompetent perforating veins. It has largely been replaced, however, by ablative techniques applied directly to the perforating veins and performed in a staged manner for persistent or recurrent symptoms after treatment for the truncal veins.

Endovascular interventions, surgical bypass, and reconstruction of the valves of the deep veins are performed when feasible to treat patients with advanced chronic venous insufficiency who have not responded to other therapies. Catheter-based interventions, usually involving placement of endovenous stents, may be considered to treat some patients with chronic occlusion or stenosis of the iliac veins. Technical success rates exceed 85% in most series, and long-term patency is achieved in ~75% of these patients. Iliocaval bypass, femoroiliac venous bypass, and femorofemoral crossover venous bypass are procedures used occasionally to treat iliofemoral vein occlusion; saphenopopliteal vein bypass can be used to treat chronic femoropopliteal vein obstruction. Long-term patency rates for venous bypass procedures generally exceed 60% and are associated with improvement in symptoms. Surgical reconstruction of the valves of the deep veins and valve transfer procedures rarely are used to treat valvular incompetence. Valvuloplasty involves tightening the valve by commissural apposition. With valve transfer procedures, a segment of vein with a competent valve, such as a brachial or axillary vein, or adjacent saphenous or deep femoral vein, is inserted as an interposition graft in the incompetent vein. Both valvuloplasty and vein transfer operations may result in ulcer healing in the majority of patients, although they are uncommonly performed and the success rates are somewhat better with valvuloplasty.

Lymphedema Lymphedema is a chronic condition caused by impaired transport of lymph and characterized by swelling of one or more limbs and occasionally the trunk and genitalia. Fluid accumulates in interstitial tissues when there is an imbalance between lymph production and lymph absorption, a process governed in large part by Starling forces. Deficiency, reflux, or obstruction of lymph vessels perturbs the ability of the lymphatic system to reabsorb proteins that had been filtered by blood vessels, and the tissue osmotic load promotes interstitial accumulation of water. Persistent lymphedema leads to inflammatory and immune responses characterized by infiltration of mononuclear cells, fibroblasts, and adipocytes, leading to adipose and collagen deposition in the skin and subcutaneous tissues.

Lymphatic Anatomy Lymphatic capillaries are blind-ended tubes formed by a single layer of endothelial cells. The absent or widely fenestrated basement membrane of lymphatic capillaries allows access to interstitial proteins and particles. Lymphatic capillaries merge to form microlymphatic precollector vessels, which contain few smooth muscle cells. The precollector vessels drain into collecting lymphatic vessels, which comprise endothelial cells, a basement membrane, smooth muscle, and bileaflet valves. The collecting lymphatic vessels in turn merge to form larger lymphatic conduits. Analogous to venous anatomy, there are superficial and deep lymphatic vessels in the legs, which communicate at the popliteal and inguinal lymph nodes. Pelvic lymphatic vessels drain into the thoracic duct, which ascends from the abdomen to the thorax and connects with the left brachiocephalic vein. Superficial and deep lymphatic vessels of the arm communicate with axillary lymph nodes, and the efferent lymphatic vessels join lymphatic vessels from the head to form the right and left subclavian lymphatic trunks, which ultimately enter the right brachiocephalic vein and thoracic duct, respectively. Lymph is propelled centrally by the phasic contractile activity of lymphatic smooth muscle and facilitated by the contractions of contiguous skeletal muscle. The presence of lymphatic valves ensures unidirectional flow.

Etiology Lymphedema may be categorized as primary or secondary **(Table 293-2)**. The prevalence of primary lymphedema is ~1.15 per 100,000 persons <20 years of age. Females are affected more frequently than males. Primary lymphedema may be caused by agenesis, hypoplasia, hyperplasia, or obstruction of the lymphatic vessels. There are three clinical subtypes: congenital lymphedema, which appears shortly after birth; lymphedema praecox, which has its onset at the time of puberty; and lymphedema tarda, which usually begins after age 35. Familial forms of congenital lymphedema (Milroy's disease) and lymphedema praecox (Meige's disease) may be inherited in an autosomal dominant manner with variable penetrance; autosomal or sex-linked recessive forms are less common. At least 19 genes are associated with inherited forms of lymphedema. Mutations in the FLT4 gene expressing vascular endothelial growth factor receptor 3 (*VEGFR3*), which is a determinant of lymphangiogenesis, cause Milroy's disease; and a mutation of the gene encoding *VEGF-C*, a ligand for *VEGFR3*, may cause a Milroy's disease-like phenotype. A mutation of the *LSC1* gene is associated with the cholestasis-lymphedema syndrome. Mutations in the *FOXC2* gene, which encodes a transcription factor that interacts with a signaling pathway involved in the development of lymphatic vessels, cause the lymphedema-distichiasis syndrome, in which lymphedema praecox occurs in patients who also have a double row of eyelashes. A mutation of *SOX18*, a transcription factor upstream of lymphatic endothelial cell differentiation, has been described in patients with lymphedema, alopecia, and telangiectasias (hypotrichosis, lymphedema, telangiectasia syndrome). Mutations of the *CCBE1* gene, which enhances the lymphangiogenic effects of VEGF-C, cause Hennekam's lymphangiectasia-lymphedema syndrome, and *KIF11* gene mutations are associated with microcephaly-lymphedema syndrome. Mutations of the *GATA2* gene, which is involved in the development of lymphatic valves, cause lymphedema and a predisposition to acute myeloid leukemia. Patients with a chromosomal aneuploidy, such as Turner's syndrome, Klinefelter's syndrome, or trisomy 18, 13, or 21, may develop lymphedema. Syndromic vascular anomalies associated with lymphedema also include Klippel-Trénaunay syndrome and Parkes-Weber syndrome. Other disorders associated with lymphedema include Noonan's syndrome, yellow nail syndrome, intestinal lymphangiectasia syndrome, lymphangiomyomatosis, and neurofibromatosis type 1.

TABLE 293-2 Causes of Lymphedema

Primary

- Sporadic (no identified cause)
- Genetic disorders
  - Milroy's disease (*VEGFR3, VEGF-C*)
  - Meige's disease (gene mutation not established)
  - Lymphedema-distichiasis syndrome (*FOXC2*)
  - Cholestasis-lymphedema (*LSC1*)
  - Hennekam's lymphangiectasia-lymphedema syndrome (*LCCBE1*)
  - Emberger's syndrome-lymphedema and predisposition to AML (*GATA2*)
  - Microcephaly-lymphedema syndrome (*KIF11*)
  - Hypotrichosis-lymphedema-telangiectasia (*SOX18*)
- Chromosomal aneuploidies
  - Turner's syndrome
  - Klinefelter's syndrome
  - Trisomy 13, 18, or 21
- Other disorders associated with primary lymphedema
  - Noonan's syndrome
  - Klippel-Trénaunay syndrome
  - Parkes-Weber syndrome
  - Yellow nail syndrome
  - Intestinal lymphangiectasia syndrome
  - Lymphangiomyomatosis
  - Neurofibromatosis type 1

Secondary

- Infection
  - Bacterial lymphangitis (*Streptococcus pyogenes, Staphylococcus aureus*)
  - Lymphogranuloma venereum (*Chlamydia trachomatis*)
  - Filariasis (*Wucheria bancrofti, Brugia malayi, B. timori*)
  - Tuberculosis
- Neoplastic infiltration of lymph nodes
  - Lymphoma
  - Prostate
  - Others
- Surgery or irradiation of axillary or inguinal lymph nodes for treatment of cancer Iatrogenic
  - Lymphatic division (during peripheral bypass surgery, varicose vein surgery, or harvesting of saphenous veins)
- Miscellaneous
- Chronic venous insufficiency
  - Contact dermatitis
  - Podoconiosis
  - Rheumatoid arthritis
  - Pregnancy
  - Factitious

Secondary lymphedema is an acquired condition that results from damage to or obstruction of previously normal lymphatic channels. Recurrent episodes of bacterial lymphangitis, usually caused by streptococci, are a very common cause of lymphedema. The most common etiology of secondary lymphedema worldwide is lymphatic filariasis, affecting >120 million children and adults and causing lymphedema and elephantiasis in 14 million of these affected individuals **(Chap. 240)**. Recurrent bacterial lymphangitis by *Streptococcus* may result in chronic lymphedema. Other infectious causes include lymphogranuloma venereum and tuberculosis. A common acquired cause of lymphedema in tropical countries is podoconiosis, which results from barefoot exposure and absorption of silicate particles in soil derived from volcanic rock. In developed countries, the most common secondary cause of lymphedema is surgical excision or irradiation of axillary and inguinal lymph nodes for treatment of cancers, such as breast, cervical, endometrial, and prostate cancer, sarcomas, and malignant melanoma. Lymphedema of the arm occurs in 13% of breast cancer patients after axillary node dissection and in 22% after both surgery and radiotherapy. Lymphedema of the leg affects ~15% of patients with cancer after inguinal lymph node dissection. Tumors, such as prostate cancer and lymphoma, also can infiltrate and obstruct lymphatic vessels. Chronic venous insufficiency is gaining recognition as a cause of lymphedema termed phlebolymphedema. Less common causes include contact dermatitis, rheumatoid arthritis, pregnancy, and self-induced or factitious lymphedema after application of tourniquets.

Clinical Presentation Lymphedema is generally a painless condition, but patients may experience a chronic dull, heavy sensation in the leg, and most often, they are concerned about the appearance of the leg. Lymphedema of the lower extremity initially involves the foot and gradually progresses up the leg so that the entire limb becomes

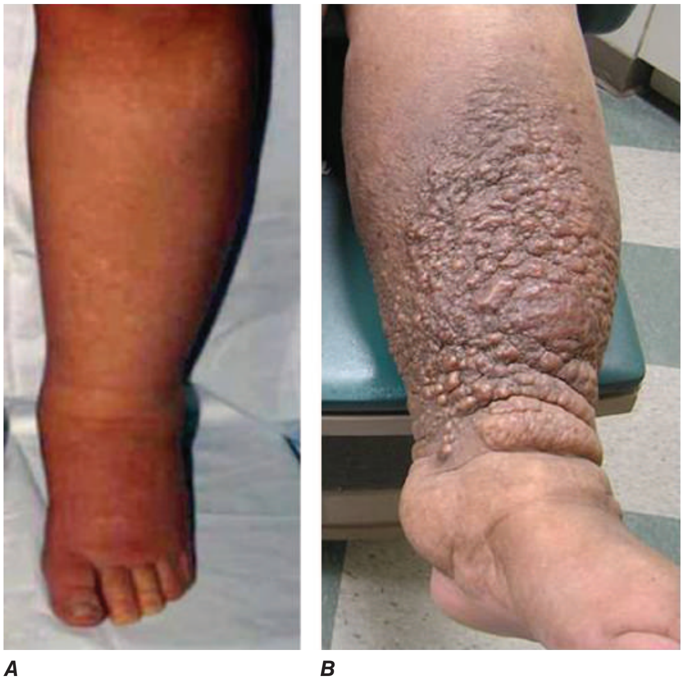

FIGURE 293-2 ***A.*** Lymphedema characterized by swelling of the leg, nonpitting edema, and squaring of the toes. *(Courtesy of Dr. Marie Gerhard-Herman, with permission.)* ***B.*** Advanced chronic stage of lymphedema illustrating the woody appearance of the leg with acanthosis and verrucous overgrowths. *(Courtesy of Dr. Jeffrey Olin, with permission.)*

edematous **(Fig. 293-2)**. In the early stages, the edema is soft and pits easily with pressure. Over time, subcutaneous adipose tissue accumulates, the limb enlarges further and loses its normal contour, and the toes appear square. Thickening of the skin is detected by Stemmer's sign, which is the inability to tent the skin at the base of the toes. *Peau d'orange* is a term used to describe dimpling of the skin, resembling that of an orange peel, caused by lymphedema. In the chronic stages, the edema no longer pits and the limb acquires a woody texture as the tissues become indurated and fibrotic. The International Society of Lymphology describes four clinical stages of lymphedema **(Table 293-3)**.

Differential Diagnosis Lymphedema should be distinguished from other disorders that cause unilateral leg swelling, such as deep-vein thrombosis and chronic venous insufficiency. In the latter condition, the edema is softer, and there is often evidence of a stasis dermatitis, hyperpigmentation, and superficial venous varicosities, as described earlier. Other causes of leg swelling that resemble lymphedema are myxedema and lipedema. Lipedema usually occurs in women and is caused by accumulation of adipose tissue in the leg from the thigh to the ankle with sparing of the feet.

TABLE 293-3 Stages of Lymphedema

Stage	
Stage 0 (or Ia)	A latent or subclinical condition where swelling is not evident despite impaired lymph transport. It may exist for months or years before overt edema occurs.
Stage I	Early accumulation of fluid relatively high in protein content that subsides with limb elevation. Pitting may occur. An increase in proliferating cells may also be seen.
Stage II	Limb elevation alone rarely reduces tissue swelling, and pitting is manifest. Late in stage II, the limb may or may not pit as excess fat and fibrosis supervene.
Stage III	Lymphostatic elephantiasis where pitting can be absent and trophic skin changes such as acanthosis, further deposition of fat and fibrosis, and warty overgrowths have developed.

Source: Adapted from The 2013 Consensus Document of the International Society of Lymphology: Lymphology 46:1, 2013.

Diagnostic Testing Physical examination is usually sufficient to establish a diagnosis of lymphedema. The evaluation of patients with lymphedema should include diagnostic studies to clarify the cause. Abdominal and pelvic ultrasound and computed tomography (CT) can be used to detect obstructing lesions such as neoplasms. Magnetic resonance imaging (MRI) of the affected limb may reveal a honeycomb pattern characteristic of lymphedema in the epifascial compartment and identify enlarged lymphatic channels and lymph nodes. MRI also is useful to distinguish lymphedema from lipedema. Lymphoscintigraphy and lymphangiography are rarely indicated, but either can be used to confirm the diagnosis or differentiate primary from secondary lymphedema. Lymphoscintigraphy involves the injection of radioactively labeled technetium-containing colloid into the distal subcutaneous tissue of the affected extremity, which is imaged with a scintigraphic camera to visualize lymphatic vessels and lymph nodes. Findings indicative of primary lymphedema include absent or delayed filling of the lymphatic vessels or dermal back flow caused by lymphatic reflux. Findings of secondary lymphedema include dilated lymphatic vessels distal to an area of obstruction. In lymphangiography, iodinated radiocontrast material is injected into a distal lymphatic vessel that has been isolated and cannulated. In primary lymphedema, lymphatic channels are absent, hypoplastic, or ectatic. In secondary lymphedema, lymphatic channels often appear dilated beneath the level of obstruction. The complexities of lymphatic cannulation and the risk of lymphangitis associated with the contrast agent limit the utility of lymphangiography. Optical imaging with a near-infrared fluorescence dye enables quantitative imaging of peripheral superficial lymph flow and is useful in the operative setting.

TREATMENT

Lymphedema

Patients with lymphedema of the lower extremities must be instructed to take meticulous care of their feet and legs to prevent cellulitis and lymphangitis. Skin hygiene is important, and emollients can be used to prevent drying. Prophylactic antibiotics are often helpful, and fungal infection should be treated aggressively. Patients should be encouraged to participate in physical activity, as exercise may minimize symptoms and maintain functionality; frequent leg elevation can reduce the amount of edema. Psychosocial support is indicated to assist patients cope with anxiety or depression related to body image, self-esteem, functional disability, and fear of limb loss.

Physical therapy, including massage to facilitate lymphatic drainage, may be helpful. The type of massage, termed manual lymphatic drainage, is part of comprehensive decongestive physiotherapy for lymphedema and involves mild compression of the skin of the affected extremity to dilate the lymphatic channels and enhance lymphatic motility. Multilayered, compressive bandages are applied after each massage session to reduce recurrent edema. After optimal reduction in limb volume by decongestive physiotherapy, patients can be fitted with graduated compression hose or other adjustable compression garments. Sequential intermittent pneumatic compression devices can also be applied at home to facilitate reduction of the edema, improve quality of life, and reduce recurrent cellulitis. Diuretics are contraindicated and may cause depletion of intravascular volume and metabolic abnormalities. Failure to control limb edema may lead to recurrent infections, progressive trophic skin changes, and rarely a highly lethal complication of lymphangiosarcoma.

Liposuction in conjunction with decongestive physiotherapy may be considered to treat lymphedema, particularly postmastectomy lymphedema. Other surgical interventions, including lymph node transfer, are rarely used and often not successful in ameliorating lymphedema. Microsurgical lymphaticovenous anastomotic procedures have been performed to rechannel lymph flow from obstructed lymphatic vessels into the venous system and may improve lymphatic transport and quality of life in secondary lymphedema. Limb reduction procedures to resect subcutaneous tissue and excessive skin are performed occasionally in severe cases of lymphedema to improve mobility. Modification of surgical techniques to minimize impact upon the lymphatic system or enhance drainage with prophylactic lymphatic-venous shunts are being employed during operations that require lymph node dissection to reduce the risk of lymphedema.

Therapeutic lymphangiogenesis has intriguing potential as it has been studied in animal models of lymphedema. Overexpression of VEGF-C generates new lymphatic vessels and improves lymphedema in a murine model of primary lymphedema. The administration of recombinant VEGF-C or VEGF-D stimulated lymphatic growth in preclinical models of postsurgical lymphedema. There may be additional benefits when administered in conjunction with lymph node transfer. However, clinical trials in patients with lymphedema are still required to determine efficacy of gene transfer and cell-based therapies for lymphedema.

FURTHER READING

Aspelund A et al: Lymphatic system in cardiovascular medicine. Circ Res 118:515, 2016.

Brouillard P et al: Genetics of lymphatic anomalies. J Clin Invest 124:898, 2014.

Executive Committee: The diagnosis and treatment of peripheral lymphedema: 2020 consensus document of the International Society of Lymphology. Lymphology 53:3, 2020.

Gloviczki P et al: The 2022 Society for Vascular Surgery, American Venous Forum, and American Vein and Lymphatic Society clinical practice guidelines for the management of varicose veins of the lower extremities. Part I. Duplex Scanning and Treatment of Superficial Truncal Reflux. J Vasc Surg Venous Lymphat Disord 11:231.e6, 2023.

Jayaraj A, Gloviczki P: Chronic venous insufficiency, in *Vascular Medicine*, MA Creager, JA Beckman, J Loscalzo (eds). Philadelphia, Elsevier, 2020, pp 709-727.

Kahn SR et al: The postthrombotic syndrome: Evidence-based prevention, diagnosis, and treatment strategies: A scientific statement from the American Heart Association. Circulation 130:1636, 2014.

Lurie F et al: The American Venous Forum, American Vein and Lymphatic Society and the Society for Vascular Medicine expert opinion consensus on lymphedema diagnosis and treatment. Phlebology 37:252, 2022.

Masuda E et al: The 2020 appropriate use criteria for chronic lower extremity venous disease of the American Venous Forum, the Society for Vascular Surgery, the American Vein and Lymphatic Society, and the Society of Interventional Radiology. J Vasc Surg Venous Lymphat Disord 8:505.e4, 2020.

Rabinovich A, Kahn SR: The postthrombotic syndrome: Current evidence and future challenges. J Thromb Haemost 15:230, 2017.

Rockson SG: Diseases of the lymphatic circulation, in *Vascular Medicine*, MA Creager, JA Beckman, J Loscalzo (eds). Philadelphia, Elsevier, 2020, pp 771-784.

Schleimer K et al: Update on diagnosis and treatment strategies in patients with post-thrombotic syndrome due to chronic venous obstruction and role of endovenous recanalization. J Vasc Surg Venous Lymphat Disord 7:592, 2019.

Sharma A, Wasan S: Varicose veins, in *Vascular Medicine*, MA Creager, JA Beckman, J Loscalzo (eds). Philadelphia, Elsevier, 2020, pp 693-708.

294 Pulmonary Hypertension

Bradley A. Maron, Joseph Loscalzo

Pulmonary hypertension (PH) is a heterogeneous disease involving pathogenic remodeling of the pulmonary vasculature, which increases pulmonary artery pressure and pulmonary vascular resistance (PVR). The most common causes of PH are left heart or primary lung disease. Additionally, PH is observed in some patients as a later complication of luminal pulmonary embolism and may be observed variably in patients with various hematologic, myeloproliferative, and other systemic diseases. Pulmonary arterial hypertension (PAH) is an uncommon, but distinct, PH subtype characterized by the interplay between molecular and genetic events that cause an obliterative arteriopathy and symptoms of dyspnea, chest pain, and syncope. If left untreated, PH carries a high mortality rate, largely owing to decompensated right heart failure.

There have been significant advances in diagnosis, classification, and treatment of PH patients. For example, the mean pulmonary artery pressure (mPAP) used to diagnose PH has been lowered from ≥25 mmHg to >20 mmHg, and the PVR level used to identify patients with PAH and other forms of precapillary PH has been lowered from ≥3 to >2 Wood units (WU). These adjustments emphasize earlier disease detection, as a substantial delay in diagnosis of up to 2 years is common in PAH and has important implications for both quality of life and life span. Clinicians should be able to recognize the signs and symptoms of PH and complete a systematic evaluation in at-risk patients. In this way, prompt diagnosis, appropriate treatment, and optimized patient outcome are achievable.

PATHOBIOLOGY

Apoptosis resistance, cell proliferation, dysregulated metabolism, and increased oxidant stress involving pulmonary vascular cells, pericytes, and adventitial fibroblasts underlie the pathogenesis of PAH. These events lead to hypertrophic, fibrotic, and plexogenic remodeling of distal (small) pulmonary arterioles, which decreases vascular compliance and promotes in situ thrombosis **(Fig. 294-1)**. The remodeling pattern also includes extracellular matrix expansion, which is rich in integrins, nonfibrillar collagens, fibronectin, and other tensile proteins that stiffen affected blood vessels further. A minority of patients appear to have a vasoconstriction-dominant phenotype, which, if present, requires a unique treatment strategy discussed in greater detail below. There is now greater understanding that in PAH, right ventricular dysfunction occurs due not only to increased right ventricular afterload but also as a consequence of depressed right ventricular sarcomere function from chronic inflammation or autoimmune mechanisms.

Genetic Drivers Several germline variants in genes that regulate fundamental cellular processes such as growth, proliferation, or phenotype switching are associated with incident PAH and follow a familial pattern. A variant in the gene encoding bone morphogenetic protein receptor-2 (*BMPR2*) is the most common cause of hereditary PAH, although penetrance is variable and may be as low as 20% in some populations. Sporadic mutations involving *BMPR2* are also reported in PAH patients. BMPR2 is within the (super)family of transforming growth factor- β (TGF-β) receptors that transduce a wide range of cellular events via stimulation by activins, inhibins, and bone morphogenetic proteins **(Fig. 294-2)**. Indeed, imbalance between proliferative and antiproliferative TGF-β receptor signaling in PAH has emerged as a bona fide therapeutic target based on clinical trials testing the effect of activin signal inhibitor therapy on outcome in PAH (see section on sotatercept).

The advance of genome-wide association studies has expanded the range of genetic drivers linked to pathogenic mechanisms underlying PAH, to include *TBX4* and *SOX17*, which regulate activation of fibroblast growth factor-10 and hepatocyte growth factor, respectively. Genetic predisposition to PH in the setting of left heart disease is unresolved, although single-gene polymorphisms that target actin

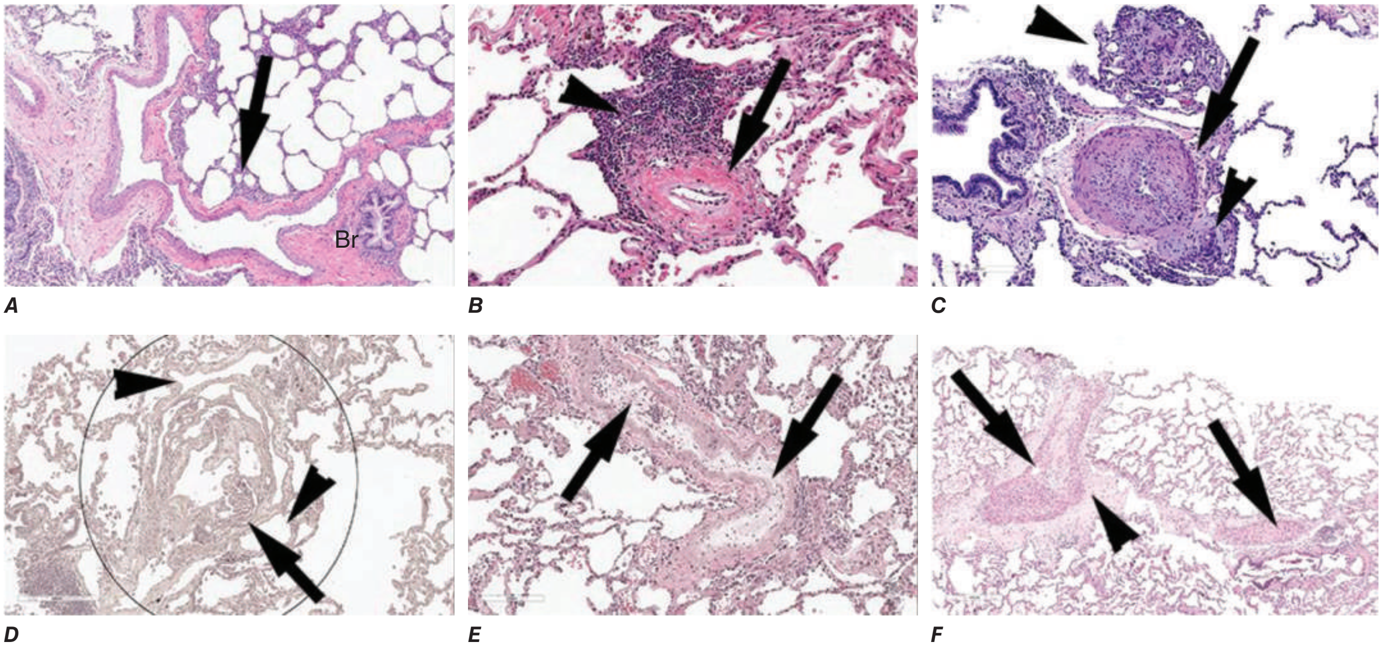

FIGURE 294-1 Panels on the left show examples of plexogenic pulmonary arteriopathy. Representative images of a normal lung (***A***) and examples of pulmonary vascular remodeling in pulmonary arterial hypertension (***B–F***), including idiopathic pulmonary arterial hypertension (***B–E***) and pulmonary venoocclusive disease (***F***), are shown. ***A.*** Normal pulmonary artery (*arrow*) adjacent to a terminal bronchiole (Br). ***B.*** Marked media and intima thickening of small vessels (*arrow*), partly surrounded by lymphoid cells, form a cluster reminiscent of a primary follicle (*arrowhead*). ***C.*** Idiopathic pulmonary hypertension lung with a markedly muscularized medium-sized pulmonary artery (*arrow*), which distally branches into a plexiform lesion (*lower arrowhead*) and an adjacent plexiform lesion (*upper arrowhead*). ***D.*** Complex vascular lesion (*circle*) with a combination of telangiectatic-like dilations of the pulmonary artery (*arrowheads*) and a plexiform lesion (*arrow*). ***E.*** Medium-sized pulmonary artery with complete lumen obliteration with a loose collagen, poorly cellular matrix (*arrows*). ***F.*** Interlobular septal, medium-sized vein (*arrowhead*) obliterated by loose connective tissue (*arrows*), likely the result of an organized thrombus, characteristic of venoocclusive disease. *(These representative images were provided courtesy of Dr. Rubin Tuder. The samples were obtained through the evaluation of lungs collected by the Pulmonary Hypertension Breakthrough Initiative, with similar pulmonary vascular pathology spectrum as reported in reference E Stacher et al: Modern age pathology of pulmonary arterial hypertension. Am J Respir Crit Care Med 186:261, 2012. Adapted with permission of the American Thoracic Society. Copyright © 2021 American Thoracic Society. All rights reserved. Reproduced with permission from BA Maron et al: Pulmonary Arterial Hypertension: Diagnosis, Treatment, and Novel Advances. Am J Respir Crit Care Med 203:1472, 2021.)*

binding, basement membrane, and major histocompatibility complex II proteins have been reported in lung tissue of affected patients. A variant in *EIF2AK4* is an established genetic risk factor for pulmonary venoocclusive disease with a recessive mode of transmission, although clinical expression among carriers is evident initially across a wide age range (10–60 years).

Epigenetics in PH The baseline DNA damage profile is increased in pulmonary artery endothelial cells and circulating blood mononuclear cells from PAH patients. This finding implicates DNA itself as a target of injury and is accompanied by epigenetic disease mechanisms involving methylation and demethylation that include translocation methylcytosine dioxygenase 2 (*TET2*) and methyltransferase 3B (*DNMT3B*). The extent to which these events are transmissible across generations remains limited to experimental disease models. Nevertheless, bromodomain-containing 4 and histone deacetylation inhibition, particularly for HDAC-1 and HDAC-5, are promising future PAH therapeutics. Long noncoding RNAs, which regulate epigenetic (among other)

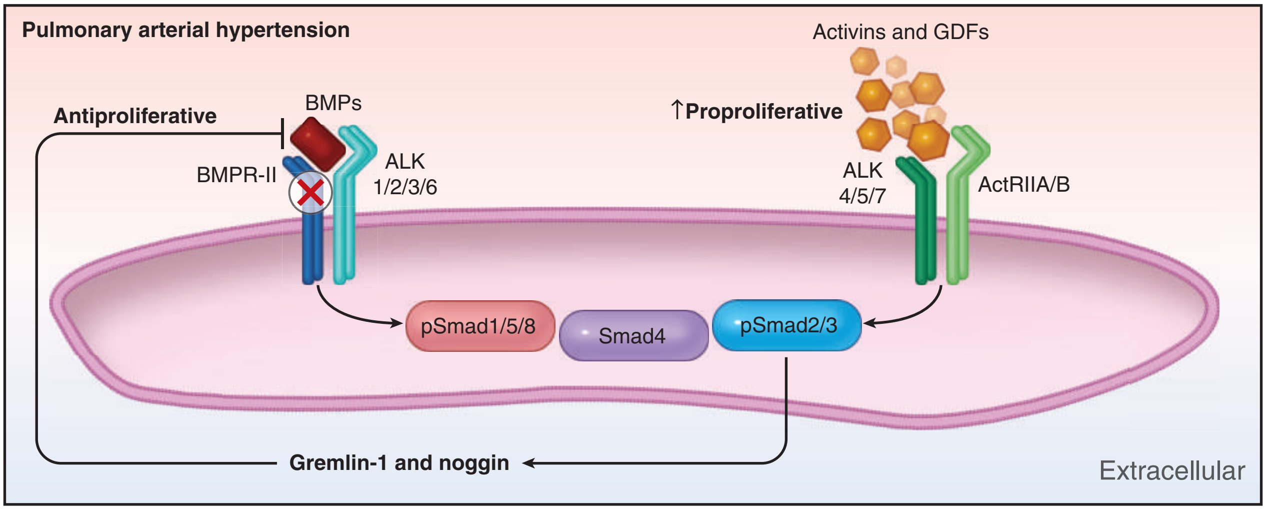

FIGURE 294-2 Pulmonary arterial hypertension is associated with dysregulation of the bone morphogenetic protein (BMP) receptor type II (BMPR-II)–Smad1/5/8 pathway in pulmonary vascular smooth muscle and endothelial cells, causing an imbalance between proproliferative and antiproliferative signaling pathways. BMPR-II–Smad1/5/8 pathway downregulation leads to increased production of activin ligands, such as activin A, growth differentiation factor 8 (GDF8), and GDF11, which contributes to activin receptor type IIA (ActRIIA)–Smad2/3 pathway upregulation. Increased phosphorylated Smad (pSmad) 2/3 activity promotes expression of the endogenous BMP antagonists gremlin-1 and noggin. Gremlin-1 and noggin further reduce BMP–Smad1/5/8 signaling. The overall result is that antiproliferative signaling is reduced, shifting the balance toward proproliferative activin–Smad2/3 signaling, which leads to pulmonary vascular remodeling. *(Adapted from M Humbert et al: N Engl J Med 384:1204, 2021.)*

processes that underpin hypertrophic vascular remodeling in PAH, include the tyrosine kinase receptor–inducing long noncoding RNA through its effects on p53/platelet-derived growth factor receptor β in lung pericytes. At present, hypoxia is the most well-studied epigenetic driver linked to PH, but a practical framework to predict disease onset through this mechanism in individual patients is unresolved.

Endothelial-Mesenchymal Transition Plexigenic and fibrotic pulmonary arterial remodeling in PAH is linked to cellular phenotype switching, particularly endothelial-mesenchymal transition (End-MT) for which specific markers are identified in ~6% of cells in explanted lungs from patients or at autopsy. Inflammation as well as hypoxia plus vascular endothelial growth factor receptor inhibition have been shown to increase End-MT, which in affected cells is characterized by migration, gain of smooth muscle actin fibers, and loss of cell-cell contacts. Interestingly, End-MT is related to a loss of BMPR2 expression, tying together a monogenic risk to a key endophenotype that drives PAH pathology.

Hypoxia, Metabolism, and Matrix Remodeling Abnormalities in multiple molecular pathways, genes, and cell types are associated with potentially modifying therapeutics, particularly those that focus on extracellular matrix remodeling (Fig. 294-3). The tyrosine kinase Janus kinase 2 (JAK2) is overactivated in PAH, which is inhibited by ruxolitinib to attenuate cellular proliferation and experimental PH. Hypoxia-inducible factor (HIF) signaling is a potentially modifiable disease-causing pathway, as treatment with the novel HIF-2α inhibitor PT2567 improves hemodynamics and vascular remodeling in PAH. Pulmonary artery smooth muscle cells isolated from PAH patients preferentially synthesize lactic acid in the presence of abundant molecular oxygen (Warburg effect). This metabolic change results in a proliferative cellular phenotype and has stimulated intense interest on metabolic modulating therapies. Dichloroacetate, which inhibits pyruvate dehydrogenase kinase to normalize voltage-gated K^+ channels and promote pulmonary artery smooth muscle cell apoptosis, has been tested clinically with overall mixed results but a signal toward predicting treatment response by *SIRT3* and *UCP2* genotype. Nonetheless, numerous alternative metabolic targets show therapeutic promise, including SOD2, PPARγ, and NFAT. Finally, there is greater interest in reverse remodeling pathways in PAH. For example, matrix metalloproteinase (MMP)-8 is protective against PAH by stabilizing mechanosensitive focal adhesion kinase–yes-associated protein/transcriptional coactivator with PDZ-binding motif (FAK-YAP/TAZ). This stabilization, in turn, suppresses extracellular matrix expansion and vascular fibrosis.

PATHOPHYSIOLOGY

In PAH, plexogenic and fibrotic remodeling of pulmonary arterioles impairs pulmonary arterial compliance and results in a progressive increase in total PVR. The resting PVR increases through the temporal progression of PAH, corresponding to a rise in mPAP. To preserve cardiac output (CO) in the face of elevated right ventricular afterload, right ventricular work must increase. A sustained (or progressive) increase in right ventricular work causes a shift in the efficiency of right ventricular systolic function by which maintaining pulmonary

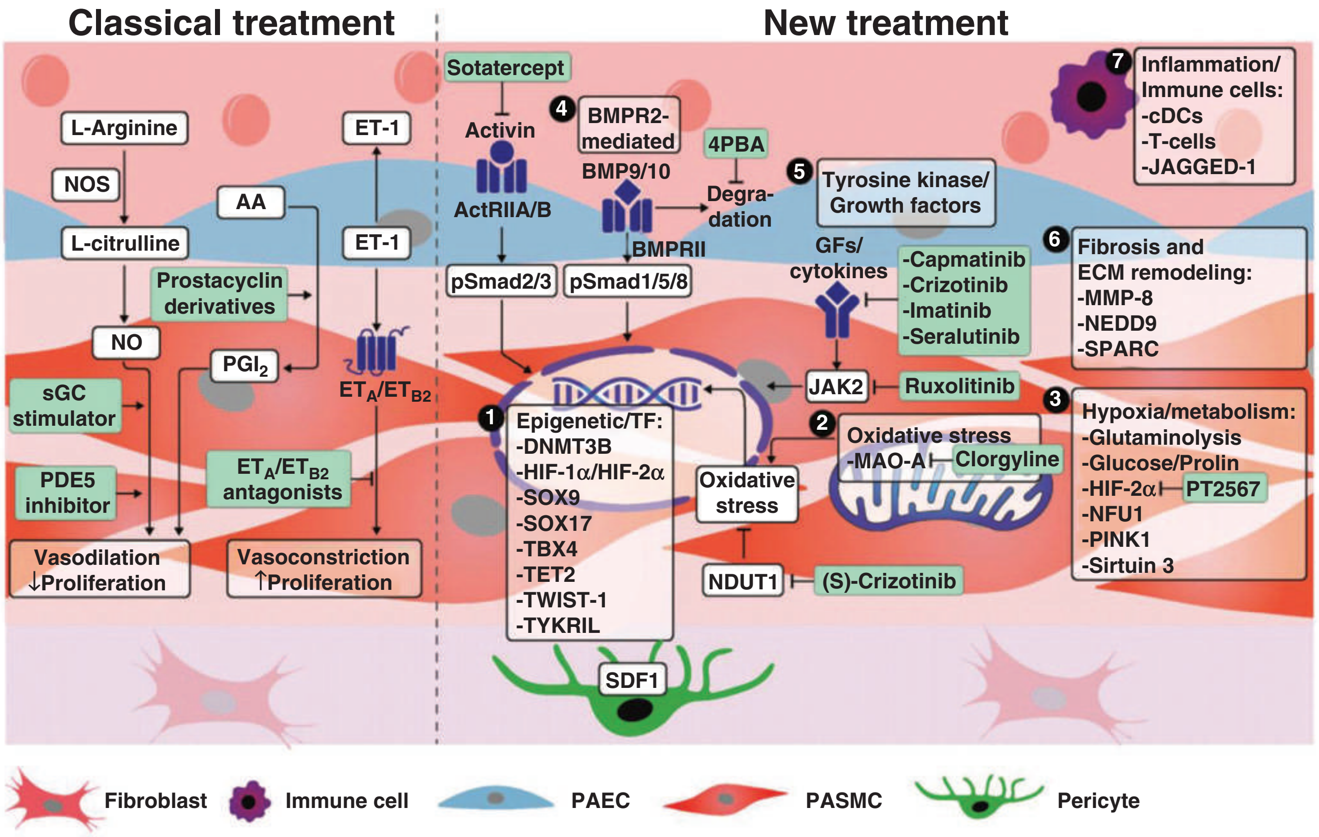
Classical treatment
New treatment
L-Arginine
NOS
L-citrulline
NO
AA
Prostacyclin derivatives
PGI$_2$
sGC stimulator
PDE5 inhibitor
ET-1
ET-1
ET$_A$/ET$_{B2}$
ET$_A$/ET$_{B2}$ antagonists
Vasodilation ↓Proliferation
Vasoconstriction ↑Proliferation
Sotatercept
Activin
ActRIIA/B
pSmad2/3
4
BMPR2-mediated
BMP9/10
BMPRII
pSmad1/5/8
4PBA
Degra-dation
5
Tyrosine kinase/ Growth factors
GFs/ cytokines
-Capmatinib
-Crizotinib
-Imatinib
-Seralutinib
JAK2
Ruxolitinib
1
Epigenetic/TF:
-DNMT3B
-HIF-1α/HIF-2α
-SOX9
-SOX17
-TBX4
-TET2
-TWIST-1
-TYKRIL
2
Oxidative stress
-MAO-A
Clorgyline
Oxidative stress
NDUT1
(S)-Crizotinib
SDF1
7
Inflammation/ Immune cells:
-cDCs
-T-cells
-JAGGED-1
6
Fibrosis and ECM remodeling:
-MMP-8
-NEDD9
-SPARC
3
Hypoxia/metabolism:
-Glutaminolysis
-Glucose/Prolin
-HIF-2α
PT2567
-NFU1
-PINK1
-Sirtuin 3
Fibroblast
Immune cell
PAEC
PASMC
Pericyte

FIGURE 294-3 Modifiable pathobiological targets in pulmonary arterial hypertension. Current treatment approaches focus on endothelin, nitric oxide, and prostacyclin signaling. Novel treatment targets address (1) epigenetic mechanisms/transcriptional factors; (2) oxidative stress; (3) hypoxia and metabolic signaling; (4) BMPR2-mediated signaling; (5) tyrosine kinase and growth factor signaling; (6) fibrosis and extracellular matrix remodeling; and (7) inflammation and immune cell infiltration. 4PBA, 4-phenylbutyric acid; AA, arachidonic acid; ActRII, activin receptor type II; BMP, bone morphogenetic protein; BMPR2, bone morphogenetic protein receptor type 2; cDC, conventional dendritic cell; DNMT3B, DNA methyltransferase 3 b; ET, endothelin; GDFs, growth and differentiation factors; HIF = hypoxia-inducible factor; JAGGED-1, jagged canonical Notch ligand 1; JAK, Janus kinase; MAO-A, monoaminoxidase A; MMP, matrix metalloproteinase; NUDT1, nudix hydrolase 1; NEDD9, neural precursor cell–expressed developmentally downregulated protein 9; NFU1, NFU1 iron-sulfur cluster scaffold; NO, nitric oxide; NOS, nitric oxide synthase; PAEC, pulmonary artery endothelial cell; PASMC, pulmonary arterial smooth muscle cell; PDE5, phosphodiesterase 5; PGI2, prostacyclin; PINK1, phosphatase and tensin homolog–induced kinase 1; SDF1, stromal cell–derived factor 1; sGC, soluble guanylate cyclase; SOX9, SRY-box transcription factor 9; SOX17, SRY-box transcription factor 17; SPARC, secreted protein acidic and rich in cysteine; TBX4, T-box transcription factor 4; TET2, Tet methylcytosine dioxygenase 2; TWIST1, Twist family BHLH transcription factor 1; TYKRIL, tyrosine kinase receptor–inducing long noncoding RNA. *(Reproduced with permission from SJ Johnson et al: Am J Respir Crit Care Med 208:528, 2023.)*

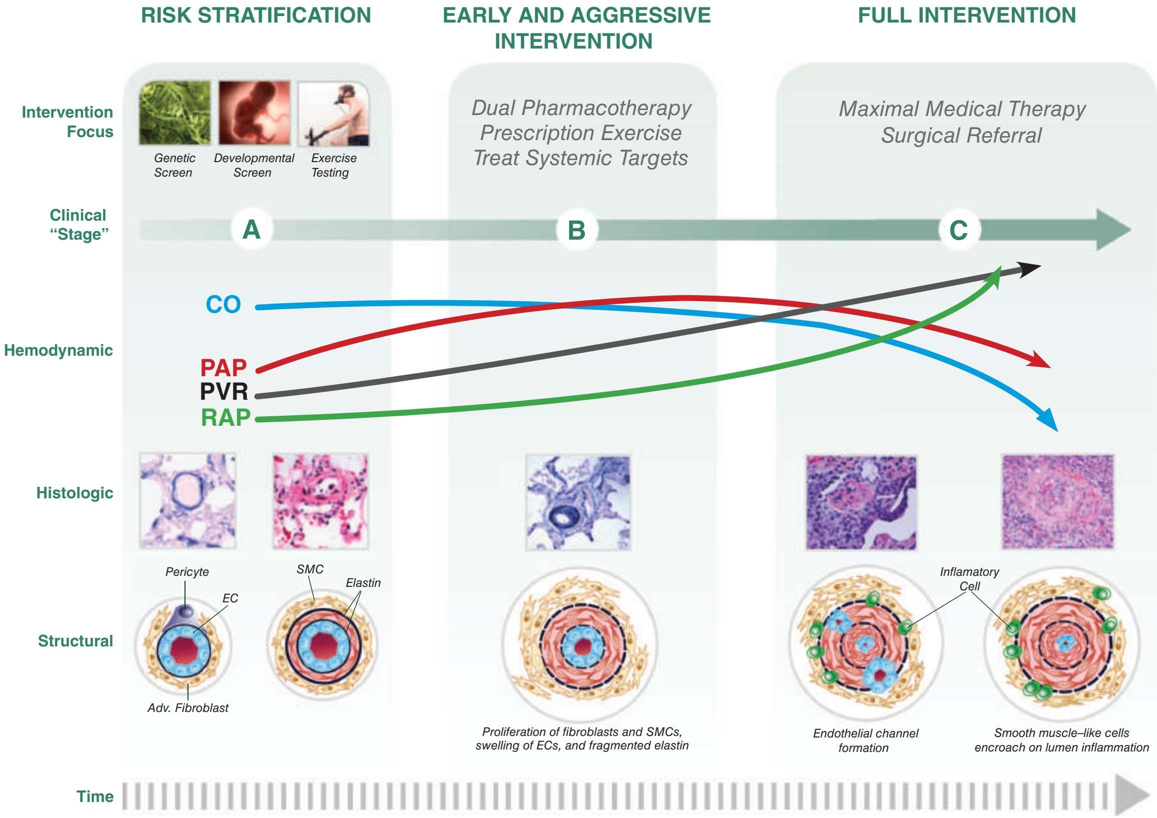
RISK STRATIFICATION
EARLY AND AGGRESSIVE INTERVENTION
FULL INTERVENTION
Intervention Focus
Genetic Screen
Developmental Screen
Exercise Testing
Dual Pharmacotherapy
Prescription Exercise
Treat Systemic Targets
Maximal Medical Therapy
Surgical Referral
Clinical "Stage"
A
B
C
Hemodynamic
CO
PAP
PVR
RAP
Histologic
Structural
Pericyte
EC
SMC
Elastin
Adv. Fibroblast
Proliferation of fibroblasts and SMCs, swelling of ECs, and fragmented elastin
Inflamatory Cell
Endothelial channel formation
Smooth muscle–like cells encroach on lumen inflammation
Time

FIGURE 294-4 An integrated overview of pulmonary arterial hypertension (PAH). In PAH, initial changes in the histopathophenotype of distal pulmonary arterioles precede significant changes in hemodynamics or the development of symptoms in most patients (clinical stage A). As vascular remodeling progresses, there is an increase in pulmonary vascular resistance (PVR), pulmonary artery pressure (PAP), and right atrial pressure (RAP). In clinical stage B, symptoms are evident and, when diagnosed, prompt early, aggressive treatment. Effacement of pulmonary arterioles results in severely increased PVR that promotes right heart failure, defined by a decrease in cardiac output (CO) and PAP. Patients in clinical stage C have severe symptoms and require full therapeutic intervention. Identifying clinical stage A patients remains challenging, although genetic risk factors or symptoms with exercise may be informative. EC, endothelial cell; SMC, smooth muscle cell. *(Reproduced with permission from Maron BA, Abman SH: Focusing on Developmental Origins and Disease Inception for the Prevention of Pulmonary Hypertension. Am J Respir Crit Care Med 195:292, 2017.)*

circulatory pressure depletes myocardial energy. These changes occur at the expense of energy normally reserved to maintain optimal blood perfusion through the alveolar-capillary interface for blood oxygenation, a process termed right ventricular–pulmonary arterial uncoupling. In end-stage PAH, the CO declines, leading to a decrease in mPAP **(Fig. 294-4)**, with extrapulmonary vascular manifestations. These include overactivation of neurohumoral signaling, renal failure, abnormal episcleral vessels (seen in patients with hereditary PAH due to *BMPR2* mutation), and hyper- or hypothyroidism. Volitional muscle atrophy in PAH is associated with reduced muscle strength, type I fiber switching to fatigable type II fibers, and decreased capillary density. This myopathy affects respiratory function, as well, including impairment of maximal inspiratory and expiratory pressure. Overall, sarcopenia in PAH with right ventricular heart failure is likely multifactorial, driven by deconditioning and alternative pathophysiological mechanisms linked to impaired cardiac output (i.e., neurohumoral signaling) and possibly through the ectopic effects of molecular intermediaries such as miR-136 downregulation **(Fig. 294-5)**.

■ DIAGNOSIS

The diagnosis of PH can be missed without a reasonable index of suspicion. Indeed, findings from clinical registries suggest that PH is often overlooked, even among patients with numerous risk factors. This shortcoming may be because PH symptoms are nonspecific, insidious, and overlap considerably with many common conditions, such as asthma, heart failure with preserved ejection fraction, and deconditioning. Additionally, there is a misconception that in patients with comorbid cardiopulmonary conditions (e.g., interstitial lung disease, mitral valve disease), PH is merely an extension of the underlying disease rather than a specific clinical entity.

Most patients will present with dyspnea and/or fatigue, whereas edema, chest pain, presyncope, and syncope are less common and associated with more advanced disease. In early phases of PAH, the physical examination is often unrevealing. As the disease progresses, there may be evidence of right ventricular failure with elevated jugular venous pressure, lower extremity edema, and ascites. Additionally, the cardiovascular examination may reveal an accentuated P_2 component of the second heart sound, a right-sided S_3 or S_4, and a holosystolic tricuspid regurgitant murmur. It is also important to seek signs of the diseases that are commonly concurrent with PH: clubbing may be seen in some chronic lung diseases, sclerodactyly and telangiectasia may signify scleroderma (or the limited cutaneous form, CREST [calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia]), and crackles on examination of the lungs and systemic hypertension may be clues to left-sided systolic or diastolic heart failure.

Overview of the Diagnostic Clinical Evaluation Once clinical suspicion is raised, a systematic approach to diagnosis and assessment is essential. In advanced disease, electrocardiography may show right ventricular hypertrophy or strain, and enlargement of pulmonary arteries and obliteration of the retrosternal space are often observed on chest roentgenography **(Fig. 294-6)**. In turn, echocardiography with agitated saline (bubble) study is the most important initial screening test. Elevated estimated pulmonary artery systolic pressure (>35 mmHg), notched waveform on continuous wave Doppler interrogation of the right ventricular outflow tract, or a hypertrophied or dilated right ventricle supports the diagnosis of PH. Important additional information can be gleaned about specific etiologies of PH, such as valvular disease, left ventricular systolic

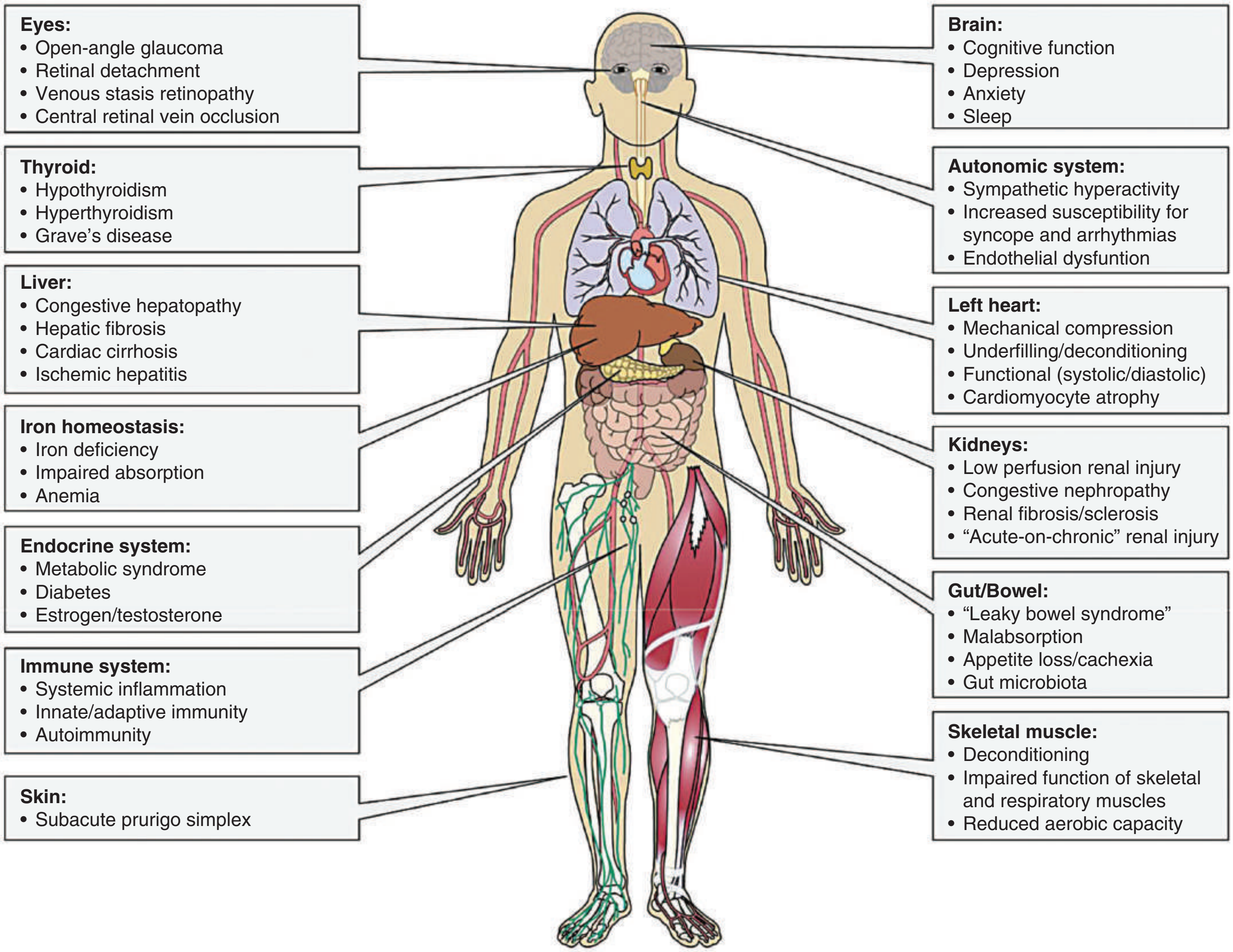
Eyes:
• Open-angle glaucoma
• Retinal detachment
• Venous stasis retinopathy
• Central retinal vein occlusion
Thyroid:
• Hypothyroidism
• Hyperthyroidism
• Grave's disease
Liver:
• Congestive hepatopathy
• Hepatic fibrosis
• Cardiac cirrhosis
• Ischemic hepatitis
Iron homeostasis:
• Iron deficiency
• Impaired absorption
• Anemia
Endocrine system:
• Metabolic syndrome
• Diabetes
• Estrogen/testosterone
Immune system:
• Systemic inflammation
• Innate/adaptive immunity
• Autoimmunity
Skin:
• Subacute prurigo simplex
Brain:
• Cognitive function
• Depression
• Anxiety
• Sleep
Autonomic system:
• Sympathetic hyperactivity
• Increased susceptibility for syncope and arrhythmias
• Endothelial dysfuntion
Left heart:
• Mechanical compression
• Underfilling/deconditioning
• Functional (systolic/diastolic)
• Cardiomyocyte atrophy
Kidneys:
• Low perfusion renal injury
• Congestive nephropathy
• Renal fibrosis/sclerosis
• "Acute-on-chronic" renal injury
Gut/Bowel:
• "Leaky bowel syndrome"
• Malabsorption
• Appetite loss/cachexia
• Gut microbiota
Skeletal muscle:
• Deconditioning
• Impaired function of skeletal and respiratory muscles
• Reduced aerobic capacity

FIGURE 294-5 Systemic manifestations of right heart failure. *(Reproduced with permission from S Rosenkranz et al: Systemic consequences of pulmonary hypertension and right-sided heart failure. Circulation 141:67, 2020.)*

dysfunction, left atrial enlargement, and intracardiac shunt. In addition, hypertrophic cardiomyopathy (HCM), amyloid cardiomyopathy, and sarcoid cardiomyopathy, which are specific forms of left heart disease that predispose to PH, are linked to disease-specific approaches to treatment.

A high-quality echocardiogram that is absolutely normal may obviate the need for further PH evaluation. However, this is distinct from an echocardiogram in which tricuspid regurgitation is not detected. In this scenario, the information required to estimate pulmonary artery pressure is lacking, and PH is observed in one-third of such patients. Patients with evidence of PH on echocardiography or in whom unexplained dyspnea or hypoxemia is evident despite an unremarkable echocardiogram often require further assessment.

Additional tests focusing on functional capacity are useful for quantifying disease burden, such as a 6-minute walk distance (6-MWD) assessment, which also aids in assessing prognosis. Cardiopulmonary exercise testing (CPET) may differentiate between cardiac and pulmonary limitations to exercise thereby providing pathophysiologic insights on the cause of dyspnea. Peak volume of oxygen consumption (pVO_2), which is an integrated parameter of cardiopulmonary fitness, is prognostic in PH patients when <15 mL/kg/min. However, low pVO_2 is an important clinical finding across the spectrum of heart-lung disease and, thus, is not diagnostic of PH per se. In patients with a normal CPET, further invasive testing is often unnecessary. One exception to this approach is in patients with reassuring CPET results but in whom a significant decrease in exercise tolerance from baseline is nonetheless reported, often observed in elite athletes or highly conditioned individuals with early-stage PH or subacute pulmonary embolism.

Invasive hemodynamic monitoring with right heart catheterization (RHC) is the gold standard for PH diagnosis and severity assessment. Interpretation of RHC data, however, is often optimized by information from diagnostic tests that support and frame the clinical context of pulmonary vascular disease.

Stepwise Approach to Diagnosing PH One common PH diagnostic strategy is outlined in Figure 294-7; however, the approach should be individualized in practice according to a particular patient's clinical and risk factor profile. For example, patients with a strong history of inhaled tobacco use may benefit from prioritizing diagnostic tests assessing pulmonary function and the lung parenchyma, whereas a myocardial ischemia evaluation should be considered early in the evaluation of patients with left-sided cardiomyopathy.

PULMONARY FUNCTION AND LUNG IMAGING Pulmonary function testing results may suggest restrictive or obstructive lung diseases as the cause of dyspnea or PH. In PAH, an isolated reduction in diffusing capacity of the lungs for carbon monoxide (DL_{CO}) is a classic finding. High-resolution computed tomography (CT) provides useful information, particularly enlargement of the main pulmonary artery, right ventricle, and atria, as well as peripheral pruning of small vessels; however, high-resolution CT may also reveal signs of venous congestion, including centrilobular ground glass infiltrate and thickened septa. In the absence of left heart disease, these findings suggest pulmonary venous disease, a rare cause of PAH that can be quite challenging to diagnose. CT is also critical for distinguishing comorbid interstitial lung disease, emphysema, or overlap syndromes that include fibrosis and obstructive pulmonary disease.

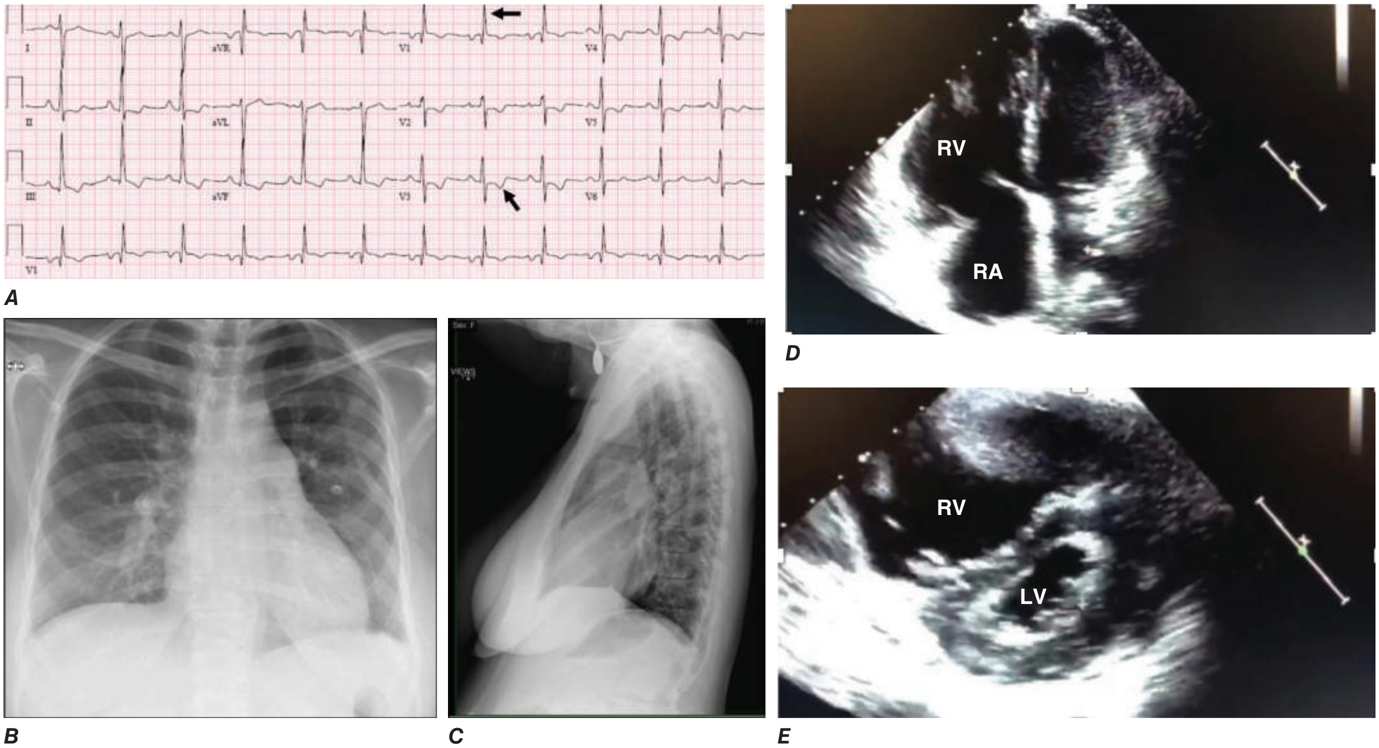
I
aVR
V1
V4
II
aVL
V2
V5
III
aVF
V3
V6
V1
A
RV
RA
D
B
C
RV
LV
E

FIGURE 294-6 Electrocardiography, chest roentgenography, and two-dimensional echocardiography in advanced pulmonary arterial hypertension. ***A.*** Standard 12-lead electrocardiogram shows peaked R waves in lead V_1 and ST-segment depression in leads V_2–V_3, suggestive of right ventricular hypertrophy with strain (*arrows*). ***B, C.*** Anterior-posterior and lateral chest roentgenogram demonstrating enlargement of central pulmonary arteries and obliteration of the retrosternal space, indicative of right ventricular hypertrophy. ***D, E.*** Apical four-chamber and two-chamber short axis views acquired by transthoracic echocardiography demonstrate right ventricular (RV) and right atrial (RA) enlargement, as well as interventricular septal flattening in diastole consistent with pressure overload. LV, left ventricle.

SLEEP STUDIES Nocturnal desaturation is a common finding in PH, even in the absence of sleep-disordered breathing. Thus, all patients should undergo nocturnal oximetry screening, regardless of whether classic symptoms of obstructive sleep apnea or obesity-hypoventilation syndrome are present.

ASSESSMENT OF PULMONARY ARTERIAL THROMBOSIS Patients with prior luminal pulmonary embolism are at increased risk for chronic thromboembolic pulmonary hypertension (CTEPH), which is a specific PH subtype characterized by vascular fibrosis and arterial microthrombus. Although CTEPH is curable in many patients by surgical endarterectomy, it is also widely underdiagnosed. Ventilation-perfusion ($\dot{V}/\dot{Q}$) scanning is still the primary test used to screen and diagnose CTEPH, which should be considered in any patient with PH of unclear etiology. Nonetheless, CT angiography is increasingly used in clinical

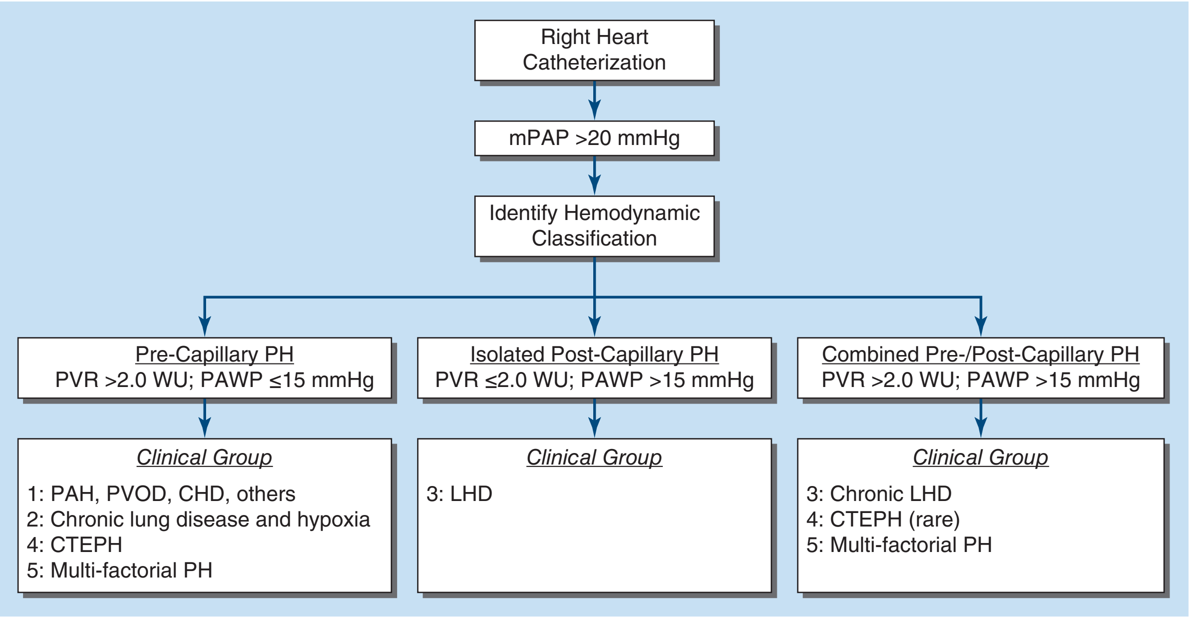
Right Heart Catheterization
mPAP >20 mmHg
Identify Hemodynamic Classification
Pre-Capillary PH
PVR >2.0 WU; PAWP ≤15 mmHg
Isolated Post-Capillary PH
PVR ≤2.0 WU; PAWP >15 mmHg
Combined Pre-/Post-Capillary PH
PVR >2.0 WU; PAWP >15 mmHg
Clinical Group
1: PAH, PVOD, CHD, others
2: Chronic lung disease and hypoxia
4: CTEPH
5: Multi-factorial PH
Clinical Group
3: LHD
Clinical Group
3: Chronic LHD
4: CTEPH (rare)
5: Multi-factorial PH

FIGURE 294-7 Integrating the cardiopulmonary hemodynamic and clinical profile of patients with pulmonary hypertension (PH). Diagnosing PH is achieved by right heart catheterization and requires a mean pulmonary artery pressure (mPAP) >20 mmHg. Patients are then classified by hemodynamic category, which together with the clinical profile and other supporting data (e.g., chest imaging, serology, genetic testing) is used to determine the PH clinical group. Certain clinical groups are incompatible with hemodynamic classifications; for example, pulmonary arterial hypertension (PAH) cannot be diagnosed in patients with isolated postcapillary or combined pre-/postcapillary PH hemodynamics. CHD, congenital heart disease; CTEPH, chronic thromboembolic PH; LHD, left heart disease; PAWP, pulmonary artery wedge pressure; PVOD, pulmonary venoocclusive disease; PVR, pulmonary vascular resistance. *(Reproduced with permission from BA Maron et al: Cardiopulmonary hemodynamics in pulmonary hypertension and heart failure: JACC Review Topic of the Week. J Am Coll Cardiol 76:2671, 2020.)*

practice to manage CTEPH especially for staging anatomic thromboembolic burden, which may be ultimately necessary to determine operative candidacy. The definitive diagnostic procedure is digital subtraction pulmonary angiography since contrast enhancement in this study provides detailed information on webbing, stricture, and vascular tapering patterns pathognomonic for CTEPH.

SEROLOGY Laboratory data that are important for screening include a human immunodeficiency virus (HIV) test when clinically indicated. In addition, all patients should have antinuclear antibodies, rheumatoid factor, and anti-Scl-70 antibodies assessed to screen for the most common rheumatologic diseases associated with PH. Liver function and hepatitis serology tests are important to screen for underlying liver disease. Methamphetamine use is recognized increasingly as a cause of PAH, and screening should be considered in patients from endemic regions or in whom the cause of PAH is not otherwise established. Finally, brain natriuretic peptide (BNP) and the N-terminus of its propeptide (NT-proBNP) correlate with right ventricular dysfunction, hemodynamic severity, and functional status in PAH. Medical therapy also lowers NT-proBNP levels in PAH, and therefore, this test may be used as a biomarker for assessing treatment response in clinical practice once a baseline level is established.

INVASIVE CARDIOPULMONARY HEMODYNAMICS The RHC remains the gold standard test to both establish the diagnosis of PH and guide selection of appropriate medical therapy. The hemodynamic criteria for diagnosing PH requires, first, an mPAP >20 mmHg. Precapillary and postcapillary PH are then distinguished by virtue of a pulmonary artery wedge pressure (PAWP) (or left ventricular end-diastolic pressure [LVEDP]) ≤15 mmHg or >15 mmHg, respectively. Isolated precapillary PH also requires a PVR >2.0 WU, whereas isolated postcapillary PH is defined by PVR ≤2.0 WU. Increasingly, combined pre- and postcapillary PH is recognized, defined by elevated mPAP >20 mmHg, PVR >2.0 WU, and PAWP >15 mmHg (Fig. 294-7).

These hemodynamic profiles inform PH clinical categorization. For example, isolated precapillary PH is most often due to primary lung disease, PAH, or CTEPH. Isolated postcapillary PH occurs in patients with mitral valvular disease, left ventricular systolic dysfunction, heart failure with preserved ejection faction, HCM, and amyloid cardiomyopathy. The same etiologies for isolated postcapillary PH also underlie combined pre- and postcapillary PH. When present, this indicates that chronic vascular congestion due to left atrial hypertension has resulted in substantial pulmonary vascular remodeling, although the precise mechanisms that transition patients from isolated postcapillary PH to combined pre-/postcapillary PH are not known. Patients with left heart disease risk factors who are diuresed aggressively prior to RHC may appear to have an isolated precapillary PH profile, tempting a diagnosis of PAH. However, these patients require specific consideration given the potential effects of diuresis on masking the true underlying PH hemodynamic classification.

Vasoreactivity testing should be reserved for patients with idiopathic or hereditary PAH. Vasodilators with a short duration of action, such as inhaled nitric oxide ($NO^{\bullet}$) or inhaled epoprostenol, are preferred for testing. A decrease in mPAP by ≥10 mmHg to an absolute level ≤40 mmHg without a decrease in CO is defined as a positive pulmonary vasodilator response, and such responders are considered for long-term treatment with calcium channel blockers. Less than 5% of patients are deemed vasoreactive, although prognosis among these patients is particularly favorable.

PULMONARY HYPERTENSION CLASSIFICATION

The current classification system, last revised in 2018 during the Sixth World Symposium on Pulmonary Hypertension (WSPH), recognizes five PH categories listed here sequentially as groups 1–5: PAH; PH due to left heart disease; PH due to chronic lung disease or sleep-disordered breathing; CTEPH; and a group of miscellaneous diseases that rarely (or inconsistently) cause PH **(Table 294-1)**.

TABLE 294-1 World Symposium on Pulmonary Hypertension Clinical Classification

GROUP 1 Pulmonary arterial hypertension (PAH)
- 1.1 Idiopathic
    - 1.1.1 Non-responders at vasoreactivity testing
    - 1.1.2 Acute responders at vasoreactivity testing
- 1.2 Heritable[a]
- 1.3 Associated with drugs and toxins[a]
- 1.4 Associated with:
    - 1.4.1 Connective tissue disease
    - 1.4.2 HIV infection
    - 1.4.3 Portal hypertension
    - 1.4.4 Congenital heart disease
    - 1.4.5 Schistosomiasis
- 1.5 PAH with features of venous/capillary (PVOD/PCH) involvement
- 1.6 Persistent PH of the newborn

GROUP 2 PH associated with left heart disease
- 2.1 Heart failure:
    - 2.1.1 with preserved ejection fraction
    - 2.1.2 with reduced or mildly reduced ejection fraction[b]
- 2.2 Valvular heart disease
- 2.3 Congenital/acquired cardiovascular conditions leading to post-capillary PH

GROUP 3 PH associated with lung diseases and/or hypoxia
- 3.1 Obstructive lung disease or emphysema
- 3.2 Restrictive lung disease
- 3.3 Lung disease with mixed restrictive/obstructive pattern
- 3.4 Hypoventilation syndromes
- 3.5 Hypoxia without lung disease (*e.g.* high altitude)
- 3.6 Developmental lung disorders

GROUP 4 PH associated with pulmonary artery obstructions
- 4.1 Chronic thrombo-embolic PH
- 4.2 Other pulmonary artery obstructions[c]

GROUP 5 PH with unclear and/or multifactorial mechanisms
- 5.1 Hematological disorders[d]
- 5.2 Systemic disorders[e]
- 5.3 Metabolic disorders[f]
- 5.4 Chronic renal failure with or without hemodialysis
- 5.5 Pulmonary tumor thrombotic microangiopathy
- 5.6 Fibrosing mediastinitis

Note: Patients with heritable PAH or PAH associated with drugs and toxins might be acute responders. Left ventricular ejection fraction for HF with reduced ejection fraction: ≤40%; for HF with mildly reduced ejection fraction: 41–49%. Other causes of pulmonary artery obstructions include: sarcomas (high or intermediate grade or angiosarcoma), other malignant tumors (e.g., renal carcinoma, uterine carcinoma, germ-cell tumors of the testis), non-malignant tumors (e.g., inherited and acquired chronic hemolytic anemia and chronic myeloproliferative disorders, sarcoidosis, pulmonary Langerhans's cell histiocytosis, and neurofibromatosis type 1, glycogen storage diseases and Gaucher disease.)

Abbreviations: HF, heart failure; HIV, human immunodeficiency virus; PAH, pulmonary arterial hypertension; PCH, pulmonary capillary hemangiomatosis; PH, pulmonary hypertension; PVOD, pulmonary veno-occlusive disease.

Source: Reproduced with permission of the © ERS 2024: Eur Resp J 61:2200879, 2023; DOI:10.1183/13993003.00879-2022.

Pulmonary Arterial Hypertension WSPH group 1 PH, or PAH, involves marked pulmonary arterial precapillary remodeling, including intimal fibrosis, increased medial thickness, pulmonary arteriolar occlusion, and classic plexiform lesions (described in detail above in the Pathobiology section). The hemodynamic criteria for PAH are sustained elevation in resting mPAP >20 mmHg, PVR >2.0 WU, and PAWP or LVEDP of ≤15 mmHg based on RHC. Idiopathic PAH (IPAH) is a progressive disease that leads to right heart failure and early mortality. From the original National Institutes of Health registry on IPAH in 1987, the average age at diagnosis was 36 years, with only 9% of patients with IPAH over the age of 60. However, contemporary data now inclusive of numerous international registries suggest a different clinical profile. The mean age of PAH patients is reported to be 54–68 years old across studies. This reflects, in part, rising awareness of this disease in the elderly. The prevalence of IPAH favors women to men by ~3.1-fold; however, the hemodynamics at diagnosis are more severe, and the prognosis is less favorable in men compared to women. Patients with hereditary PAH from a *BMPR2* mutation tend to be younger at diagnosis with more severe cardiopulmonary hemodynamics and are associated with comparatively greater clinical risk compared to IPAH. The rate of conversion to PAH among carriers without clinical evidence of disease is ~2.3% per year.

Diseases Associated with PAH Other forms of PAH that deserve specific consideration are those associated with congenital heart disease with intracardiac shunt, connective tissue disease, portal hypertension, and HIV.

CONGENITAL HEART DISEASE PAH in the setting of congenital heart disease is important to recognize since surgical correction may be indicated and when successful is associated with favorable prognosis. This is particularly salient today, as more congenital heart disease patients live to adulthood and populate general medical practices. Still, referral to adult congenital heart disease centers should be considered for patients with suspected PAH, which in this population is subclassified into four groups: Eisenmenger's syndrome, systemic-to-pulmonary shunts, coincidental or small defects causing shunts, and postoperative/closed defects causing shunts. Surgical repair of congenital anatomic lesions may be indicated prior to elevation in PVR >3.0 WU to avoid the development of Eisenmenger's syndrome, a pathophysiologic consequence of progressive pulmonary vascular remodeling due to a large-volume left-to-right shunt that is associated with cyanosis, hyperviscosity, weakness, and shortened life span. Indeed, the effect of changing the PVR threshold to define precapillary PH from 3.0 to 2.0 WU on the timing of shunt repair is an evolving concept.

CONNECTIVE TISSUE DISEASE Patients with connective tissue disease–associated PAH are encountered relatively commonly in clinical practice. Although case series link rheumatoid arthritis and systemic lupus erythematosus with pulmonary vascular disease, the predominant clinical phenotype is systemic sclerosis-associated PAH. It is important to distinguish patients with limited cutaneous scleroderma from those with diffuse scleroderma because PH in the former is likely PAH and PH in the latter often occurs in the setting of interstitial lung disease. Although the average age of scleroderma onset is between 30 and 50 years old, patients who eventually develop scleroderma-associated PAH tend to be older at the time of scleroderma diagnosis. The development of PAH in scleroderma is particularly worrisome prognostically, although implementation of modern therapies improves outcome. Overlap between pulmonary circulatory and interstitial lung disease is encountered commonly in clinical practice. It is tempting to focus on the hemodynamic derangement of these patients as a focal point of clinical care, although the efficacy of implementing PAH-specific treatment to these overlap syndrome patients is unproven.

PORTOPULMONARY HYPERTENSION Among patients with established portal hypertension, 2–10% develop portopulmonary hypertension independent of the cause of liver disease. Furthermore, portopulmonary hypertension is observed in patients with nonhepatic etiologies of portal hypertension. A hyperdynamic circulatory state is common, as in most patients with advanced liver disease; however, the same pulmonary vascular remodeling observed in other forms of PAH is seen in the pulmonary vascular bed in portopulmonary hypertension. It is important to distinguish this process from hepatopulmonary syndrome, which can also manifest with dyspnea and hypoxemia but is pathophysiologically distinct from portopulmonary hypertension in that abnormal vasodilation of the pulmonary vasculature leads to intrapulmonary shunting. Portopulmonary hypertension is an established marker of adverse outcome in the post–liver transplant period with 100% mortality reported in one study among patients with mPAP ≥50 mmHg. The optimal preoperative PVR threshold used to predict elevated postoperative risk must also consider morbidity linked to prolonged transplant wait time. At present, PVR >3.0 WU is considered a strong predicter of unfavorable outcome after transplant, although PVR >2.0 WU also appears to capture potential risk.

HIV-PAH The true prevalence of HIV-PAH is not known; however, this PAH subtype is an important cause of mortality in the HIV-infected population, and prognosis in these patients is among the least favorable for all PH subgroups. There is no correlation between the stage of HIV infection and the development of PAH. By contrast, population data from the U.S. Veterans Administration database suggest a positive and inverse correlation, respectively, between CD4+ count and viral load with estimated pulmonary artery systolic pressure (ePASP) on echocardiography. An ePASP >40 mmHg in HIV-infected patients is associated with a 40% increase in all-cause mortality compared to uninfected counterparts. Although delineating patients with PAH from within this cohort and other unselected populations is not possible, there is a pressing need to understand further the role of HIV-PH in prognosis patients, particularly in endemic areas of the world.

Pulmonary Hypertension Associated with Left Heart Disease Patients with PH due to left ventricular systolic dysfunction, aortic and mitral valve disease, and heart failure with preserved ejection fraction (HFpEF) are classified in WSPH group 2. The hallmark of this PH phenotype is elevated left atrial pressure with resulting pulmonary venous hypertension. In left-sided systolic heart failure or HFpEF, even mildly elevated mPAP is associated with adverse clinical outcome. It should be noted that PH in the setting of mitral stenosis or regurgitation is an indication for surgical (or percutaneous) valve intervention. Recent data suggest that PH is common in obstructive HCM and associated with fibroproliferative remodeling of distal pulmonary arterials even when mPAP is only mildly elevated **(Fig. 294-8)**. In amyloid cardiomyopathy, ~75% of patients have elevated mPAP, and combined pre-/postcapillary PH is the most common hemodynamic subgroup. Regardless of the cause of elevated left atrial pressure, left atrial hypertension causes pulmonary venule sclerosis and thickening that leads to PH and, ultimately, pathogenic changes to pulmonary arterioles.

Pulmonary Hypertension Associated with Lung Disease Intrinsic lung disease is the second most common cause of PH and has been observed in both chronic obstructive pulmonary disease (COPD) and interstitial lung disease. Additionally, PH is also diagnosed in diseases of mixed obstructive/restrictive pathophysiology: bronchiectasis, cystic fibrosis, mixed obstructive-restrictive disease marked by fibrosis in the lower lung zones, and emphysema predominantly in the upper lung zones. When associated with chronic lung disease, PH is usually modest. For example, 90% of COPD patients have mPAP >20 mmHg, but an mPAP >35 mmHg is observed in only 5% of patients. Nonetheless, the subgroup of patients with primary lung disease and severe PH is challenging clinically, as extensive pulmonary arterial involvement, very low DL_{CO} on pulmonary function testing, and inhibition of normal vasoreactivity are observed and are associated with poor outcome. Sleep-disordered syndromes generally result in mild PH.

Pulmonary Hypertension Associated with Chronic Thromboembolic Disease The development of PH after chronic thromboembolic obstruction of the pulmonary arteries, termed CTEPH, is well described. The incidence of CTEPH following a single pulmonary embolic event is difficult to determine accurately, but probably is between 3 and 7% of patients. Importantly, 25% of patients with CTEPH have no history of clinical venous thromboembolism, suggesting that CTEPH may develop following a subclinical pulmonary embolism or through a diverse range of mechanisms. Obstruction of the proximal pulmonary vasculature due to webbing, stricture, or focal fibrotic occlusion signifies proximal vessel involvement. Distal pulmonary arterioles remodel by luminal narrowing or obliteration. Approximately 10–15% of patients will develop a disease very similar clinically and pathologically to PAH after resection of the proximal thrombus **(Fig. 294-9)**. Chronic thromboembolic disease refers to patients with thrombotic pulmonary arterial remodeling and diminished exercise capacity with correlative symptoms in the absence of PH. This subtype is less common than CTEPH but requires consideration in patients with prior PE and dyspnea even if echocardiography is reassuring.

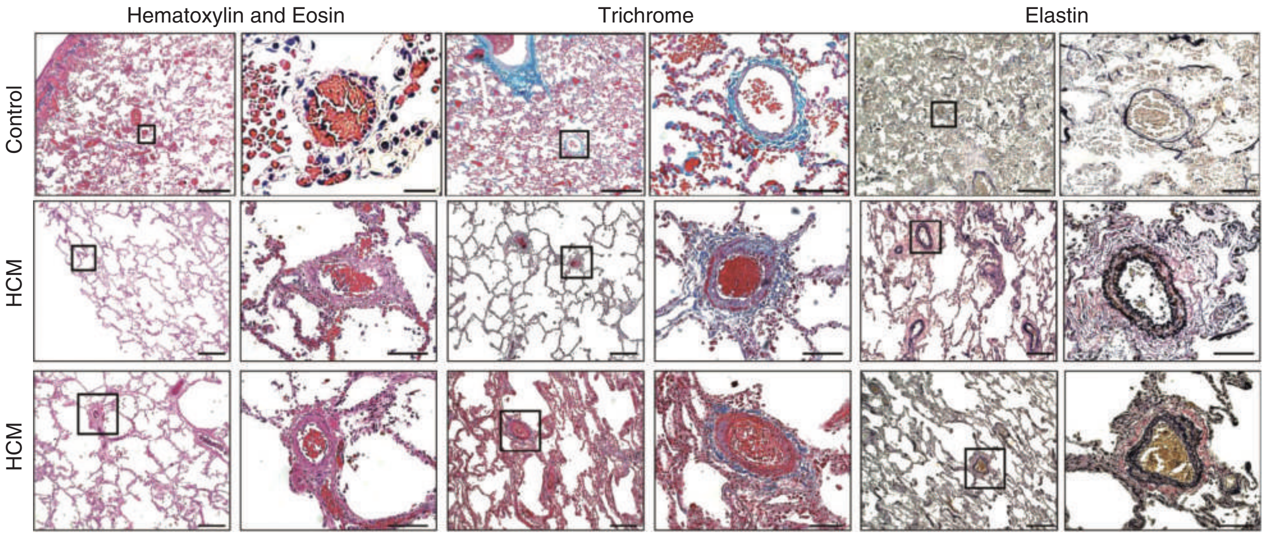

FIGURE 294-8 Pulmonary vascular remodeling in patients with obstructive hypertrophic cardiomyopathy (HCM) and mild pulmonary hypertension. Photomicrographs showing pulmonary arterial histopathophenotype of HCM. Paraffin-embedded tissue sections from control donors without lung disease and example autopsy specimens from patients with symptomatic obstructive HCM were analyzed after hematoxylin and eosin, Masson trichrome, and Verhoeff-Van Gieson staining. *(Adapted with permission from BA Maron et al: Chest 163:678, 2023.)*

OTHER DISORDERS AFFECTING THE PULMONARY VASCULATURE

Sarcoidosis Patients with sarcoidosis can develop PH as a result of lung involvement, and those who present with progressive dyspnea and PH require a thorough evaluation. In sarcoidosis, PH develops mainly due to granulomatous inflammation of the pulmonary vessels, although mechanical compression of pulmonary arteries by enlarged lymph nodes is also reported.

Sickle Cell Disease Cardiovascular system abnormalities are prominent in the clinical spectrum of sickle cell disease (and other hemoglobinopathies), including PH, which occurs in 6–10% of patients. The etiology is multifactorial, including hemolysis, hypoxemia, thromboembolism, chronically high CO, and chronic liver disease.

Schistosomiasis Schistosomiasis affects >230 million people worldwide, of whom 5% develop PAH. Thus, this infection is among the most common causes of PAH worldwide. The development of PAH occurs in the setting of hepatosplenic disease and portal hypertension. Indeed, even in the absence of schistosomiasis, splenectomy is a risk factor for PAH, presumably through impaired platelet sequestration and associated microthrombosis of pulmonary arterioles. Studies suggest that inflammation from a schistosomiasis infection induces pulmonary vascular injury through a combination of the following mechanisms: luminal obstruction by worm eggs, pulmonary artery endothelial cell inflammation, or portal hypertension that promotes a portopulmonary hypertension-PH phenotype The diagnosis is confirmed by finding the parasite ova in the urine or stool of patients with symptoms, which can be difficult. The efficacy of therapies directed toward PAH in these patients is unknown.

PHARMACOLOGIC TREATMENT OF PAH

There are 14 U.S. Food and Drug Administration (FDA)–approved medical therapies for PAH, and standardized treatment strategies have been developed that emphasize early, aggressive pharmacotherapy initiated at a specialty clinical center. Among optimally treated patients, the 1-, 3-, and 5-year survival estimates are 82, 67, and 58%, respectively, but this may overestimate risk since outcome data in the era of routine dual (and in some cases triple) therapy are lacking. All approved medical therapies target the prostacyclin, NO$^{\bullet}$, or endothelin receptor signaling pathways. Drug delivery methods now include oral, inhaled, subcutaneous (including via surgically implanted devices), and intravenous routes.

Prostanoids In PAH, endothelial dysfunction and platelet activation cause an imbalance of arachidonic acid metabolites with reduced prostacyclin levels and increased thromboxane A_2 production. Prostacyclin (PGI_2) activates cyclic adenosine monophosphate

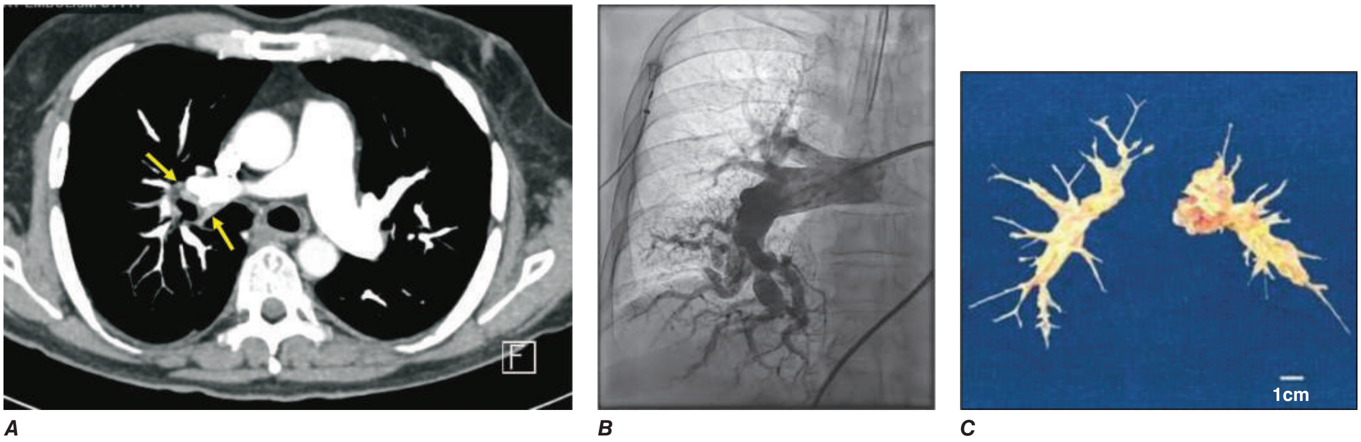

FIGURE 294-9 Chronic thromboembolic pulmonary hypertension (CTEPH) imaging findings and surgical endarterectomy specimen. *A.* Contrast-enhanced computed tomography of the chest shows an obstructive vascular pattern involving segmental pulmonary arteries (*yellow arrows*) in a 63-year-old man with exertional dyspnea and remote history of pulmonary embolism. ***B.*** Still image of a pulmonary angiography of the right lung (submaximal injection shown) shows pulmonary artery stricture, webbing, and severe dearborization that is classic for CTEPH. ***C.*** Fibrotic, chronic clot specimens resected during surgical pulmonary endarterectomy, which is curative in most CTEPH patients. *(Panel C is reproduced with permission from IM Lang, M Madani: Update on chronic thromboembolic pulmonary hypertension. Circulation 130:508, 2014.)*

(cAMP)–dependent pathways that mediate vasodilation. PGI_2 also has antiproliferative effects on vascular smooth muscle and inhibits platelet aggregation. Protein levels of prostacyclin synthase are decreased in pulmonary arteries of patients with PAH. This imbalance of mediators is offset therapeutically by the administration of either exogenous prostacyclin (and analogues, termed prostanoids) or a prostacyclin receptor agonist.

Epoprostenol was the first prostanoid available for the management of PAH. Epoprostenol delivered as a continuous intravenous infusion improves functional capacity and survival in PAH. The efficacy of epoprostenol in World Health Organization (WHO) Functional Class (FC) III and IV PAH patients was demonstrated in a clinical trial that showed improved quality of life, mPAP, PVR, 6-MWD, and mortality. Treprostinil has a longer half-life than epoprostenol (~4 h vs ~6 min), which allows for subcutaneous administration. Treprostinil has been shown to improve pulmonary hemodynamics, symptoms, exercise capacity, and survival in PAH. Inhaled prostacyclin provides the beneficial effects of infused prostacyclin therapy without the inconvenience and side effects of infusion catheters (e.g., risk of infection and infusion site reactions). Both inhaled iloprost and treprostinil have been approved for patients with PAH and severe heart failure symptoms. Oral prostacyclin is also efficacious in clinical trials, but the maximal dose is modest and, therefore, generally reserved as a second-line therapy.

Selexipag is an oral nonprostanoid diphenylpyrazine derivative that binds the prostaglandin I_2 (IP) receptor with high affinity. The active metabolite of selexipag has a prolonged half-life in comparison with prostanoid analogues and permits twice-daily dosing. The efficacy of selexipag was evaluated in patients with PAH in New York Heart Association (NYHA) FC II to III on background therapy with either an endothelin-1 (ET-1) receptor antagonist or sildenafil, or both. This trial represents the largest randomized placebo-controlled trial among patients with PAH ever completed, enrolling 1156 patients treated for a median of 1.4 years. Selexipag reduced the risk of hospitalization and the risk of disease progression by 43% ($p < .0001$) compared to those who received placebo. There were no significant differences in mortality between the two study groups, and the side effect profile was similar to that of prostacyclins.

Endothelin Receptor Antagonists Endothelin receptor antagonists (ERAs) inhibit the detrimental effects of ET-1, a potent endogenous vasoconstrictor and vascular smooth muscle mitogen. In PAH, ET-1 associates positively with PVR and mPAP and inversely with CO and 6-MWD. The ET-1 signaling axis is complex and cell type–specific: ET type A (ET_A) and type B (ET_B) receptors expressed in pulmonary artery smooth muscle cells mediate vasoconstriction, whereas human pulmonary artery endothelial cells express ET_B receptors that promote ET-1 clearance and vasodilation through endothelial nitric oxide synthase activation and prostacyclin release.

The three ERAs approved for use in the United States are the nonselective $ET_{A/B}$ receptor antagonists bosentan and macitentan and the selective ET_A antagonist ambrisentan. Studies have shown that bosentan improves hemodynamics and exercise capacity and delays clinical worsening. The randomized, placebo-controlled, phase 3 Bosentan Randomized Trial of Endothelin Antagonist Therapy (BREATHE)-1 trial comparing bosentan to placebo demonstrated improved symptoms, 6-MWD, and WHO FC in patients treated with bosentan. The Endothelin Antagonist Trial in Mildly Symptomatic Pulmonary Arterial Hypertension Patients (EARLY) study comparing bosentan to placebo demonstrated improved PVR and 6-MWD in patients with WHO FC II.

Several studies, including the phase 3, placebo-controlled Ambrisentan in Pulmonary Arterial Hypertension, (ARIES)-1 trial, have demonstrated that ambrisentan improves exercise tolerance, WHO FC, hemodynamics, and quality of life in patients with PAH. More recently, the Study with an Endothelin Receptor Antagonist in Pulmonary Arterial Hypertension to Improve Clinical Outcome (SERAPHIN) trial randomized 742 PAH patients to receive placebo or macitentan, which is an $ET_{A/B}$ antagonist with optimized receptor binding affinity. The majority of patients were on some form of background PAH therapy. Over an average treatment duration of 85 weeks, the hazard ratio for achieving the composite primary endpoint of PAH-related clinical worsening, which included death or disease progression, was decreased by 45% in the 10-mg dose arm.

Nitric Oxide Pathway Effectors The gaseous, lipophilic molecule $NO^{\bullet}$ is generated by endothelial nitric oxide synthase in endothelial cells and activates soluble guanylyl cyclase (sGC) to generate cyclic guanosine monophosphate (cGMP) in vascular smooth muscle cells and platelets. The cyclic nucleotide cGMP is a second messenger that induces vasodilation through relaxation of arterial smooth muscle cells and inhibits platelet activation. Phosphodiesterase type 5 (PDE5) enzymes are highly expressed in lung vascular tissue (and the corpus carvernosum of the penis). The PDE5 inhibitors prevent hydrolysis (inactivation) of cGMP to maximize $NO^{\bullet}$-dependent vasodilation, serving as the basis for use of this drug class in the treatment of PH (and erectile dysfunction). The two PDE5 inhibitors used for the treatment of PAH are sildenafil and tadalafil. Both agents have been shown to improve hemodynamics and 6-MWD.

Riociguat increases bioactive cGMP by (1) stabilizing the molecular interaction between $NO^{\bullet}$ and sGC, and (2) directly stimulating sGC independent of $NO^{\bullet}$ bioavailability. Riociguat significantly improved exercise capacity, pulmonary hemodynamics, WHO FC, and time to clinical worsening in patients with PAH and is the sole approved pharmacotherapy for CTEPH patients for whom surgical pulmonary endarterectomy is ineffective or contraindicated.

Activin Signal Inhibitor Therapy Motivated by basic and translational data suggesting that treatment with a TGF-β ligand trap inhibits the proproliferative activity of activin ligands in pulmonary vascular cells, sotatercept was developed as a novel fusion protein of the Fc domain of human IgG and the extracellular domain of activin receptor type II. Sotatercept blocks activin receptor II–ALK 4/5/7 signal transduction to upregulate BMPR2 bioactivity. The putative effect of this drug's reaction in PAH is rebalancing of the TGF-β–BMPR2 pathway to restore normal cellular survival and growth patterns.

The PULSAR trial was a large, phase 2, placebo-controlled trial demonstrating a dose-dependent decrease in PVR by sotatercept (0.3 and 0.7 mg/kg administered as an injection) at study week 24 compared to baseline of –1.8 WU and 3.0 WU, respectively. The STELLAR trial was a phase 3 placebo-controlled trial that studied the effect of sotatercept 0.3 mg/kg with target dose of 0.7 mg/kg on 6-MWD. The patients in the study had moderate or severe symptom burden despite ongoing pulmonary vasodilator treatment. In fact, ~60% of patients were on triple therapy, including 40% who were on treatment with intravenous prostacyclin treatment, typically reserved for high-risk or very-high-risk patients (see below). Despite this prior treatment, sotatercept therapy resulted in an average of +40 m improvement in 6-MWD compared to –1.4 m for placebo and was associated with significant improvement in nearly all secondary endpoints including WHO FC improvement, reduction in NT-proBNP level, and shorter time to death or PAH worsening. Telangiectasia development was observed in 10.4% of patients in the sotatercept group, which tracks with the angioproliferative bioactivity reported for TGF-β, although the clinical relevance of these vascular anomalies is not known. Finally, the long-term open-label study that tracked clinical progress up to 1 year in patients from PULSAR demonstrated a sustained (but not progressive) benefit in PVR reduction by sotatercept without an unexpected change in the proportion of side effects given the duration of the follow-up period.

APPROACH TO PAH TREATMENT

Treatment aims to achieve a low clinical risk profile, defined as a 1-year mortality risk of <5%. Generally, this describes a patient with minimal symptoms, WHO FC I or II, 6-WMD >440 m, and cardiac index ≥2.5 L/min per m^2. To accomplish this goal, early referral to an expert center is advised since most patients will ultimately require two or more PAH pharmacotherapies in addition to risk factor modification (such as a low-sodium diet), diuretic use to include mineralocorticoid receptor antagonists if tolerated, supplemental oxygen, and prescription (or supervised) exercise **(Fig. 294-10)**. It is clear that targeting the diverse pathobiologic and pathophysiologic events involved in vascular

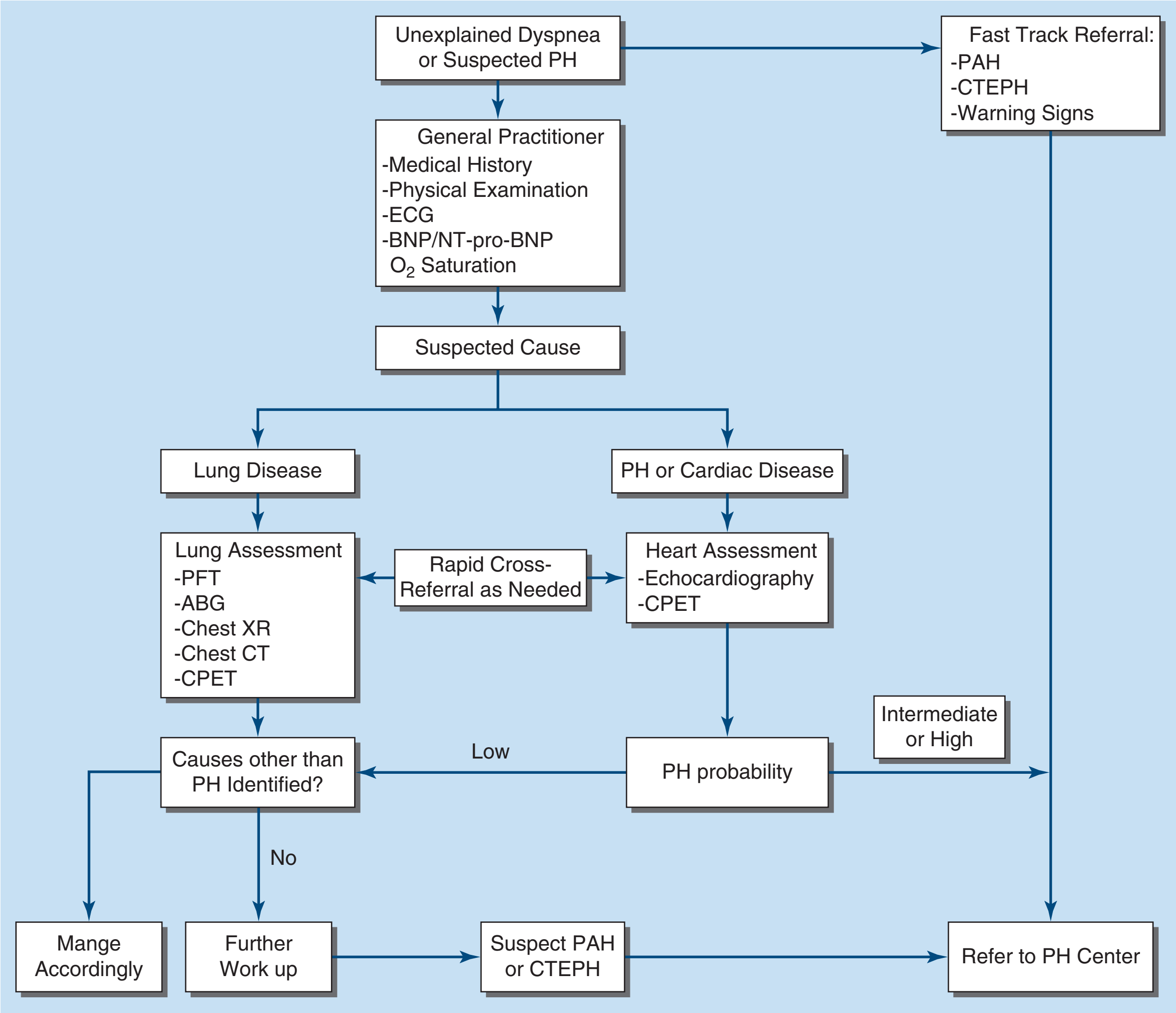
Unexplained Dyspnea or Suspected PH
Fast Track Referral:
-PAH
-CTEPH
-Warning Signs
General Practitioner
-Medical History
-Physical Examination
-ECG
-BNP/NT-pro-BNP
O2 Saturation
Suspected Cause
Lung Disease
PH or Cardiac Disease
Lung Assessment
-PFT
-ABG
-Chest XR
-Chest CT
-CPET
Rapid Cross-Referral as Needed
Heart Assessment
-Echocardiography
-CPET
Intermediate or High
Causes other than PH Identified?
Low
PH probability
No
Mange Accordingly
Further Work up
Suspect PAH or CTEPH
Refer to PH Center

FIGURE 294-10 Strategy for diagnosing pulmonary hypertension (PH) in clinical practice. Diagnostic algorithm of patients with unexplained dyspnea and/or suspected pulmonary hypertension. ABG, arterial blood gas analysis; BNP, brain natriuretic peptide; CPET, cardiopulmonary exercise testing; CT, computed tomography; CTEPH, chronic thromboembolic pulmonary hypertension; ECG, electrocardiogram; NT-proBNP, N-terminal pro-brain natriuretic peptide; PAH, pulmonary arterial hypertension; PFT, pulmonary function tests; PH, pulmonary hypertension. Warning signs include rapid progression of symptoms, severely reduced exercise capacity, presyncope or syncope on mild exertion, and signs of right heart failure. Lung and heart assessment by specialist as per local practice. CT pulmonary angiography recommended if PH suspected. Includes connective tissue disease (especially systemic sclerosis), portal hypertension, HIV infection, and family history of PAH. *(Reproduced with permission of the © ERS 2024: Eur Resp J 61:2200879, 2023; DOI:10.1183/13993003.00879-2022.)*

remodeling is needed to optimize treatment. The concept of combination therapy in PAH is modeled after other complex diseases in which a similar approach has been effective, including HIV, cancer, and left heart failure.

The approach to therapy in a patient with incident PAH is outlined in Fig. 294-11. The role of early, aggressive therapy with combination oral treatments was addressed in the landmark Initial Use of Ambrisentan plus Tadalafil in Pulmonary Arterial Hypertension (AMBITION) trial. Treatment-naïve, incident PAH patients (n = 500) were randomized to a combination of ambrisentan and tadalafil, ambrisentan monotherapy, or tadalafil monotherapy. Up-front combination therapy with ambrisentan and tadalafil was associated with a 50% lower risk of clinical worsening (composite of death, lung transplantation, hospitalization for PAH worsening, and worsening PAH) when compared with the monotherapy groups. This difference was driven primarily by the delay in time to first hospitalization. Importantly, initial combination therapy was not associated with an increase in adverse events. Registry data suggest that patients on dual therapy with a PDE5 inhibitor plus ERA combinations alternative to the drugs studied in AMBITION also have better outcomes compared to patients treated with monotherapy, suggesting that the attendant benefit from combination therapy may not be drug specific.

The paradigm shift toward early, aggressive pharmacotherapy in PAH is expanding to up-front triple combination therapy. Clinical studies on up-front triple therapy remain mixed, which is likely driven by challenges surrounding patient selection given the precarious balance between cumulative off-target effects occurring when multiple vasoactive drugs are initiated at the same time, frailty and limited hemodynamic reserve associated with advanced PAH, and need for aggressive treatment to mitigate severely abnormal hemodynamics or right ventricular heart failure. Generally, treatment deescalation, escalation, or class switching for patients with established PAH should be decided in partnership with a PAH clinical center of excellence.

Treatment of HFpEF Disease-specific treatments for HFpEF-PH are lacking. The focus of care should be to optimize contemporary treatment of HFpEF, including minimizing excessive salt and fluid intake while advancing SGLT-2 therapy, mineralocorticoid therapy, and diuretics, and encouraging regular exercise to promote weight loss and conditioning. Invasive PA pressure monitoring may be helpful for tracking elevation in LVEDP, detected as PH, to avert hospitalization.

Treatment of PH due to Lung Disease Inhaled iloprost improved 6-MWD on average +33 m in a placebo-controlled randomized clinical trial studying patients with interstitial lung disease and PH and moderate-to-severe exercise limitation. Although a promising future strategy for the routine management of patients, additional data are needed to confirm efficacy and clarify the criteria defining optimal patients for treatment since dosing is frequent and a substantial rate

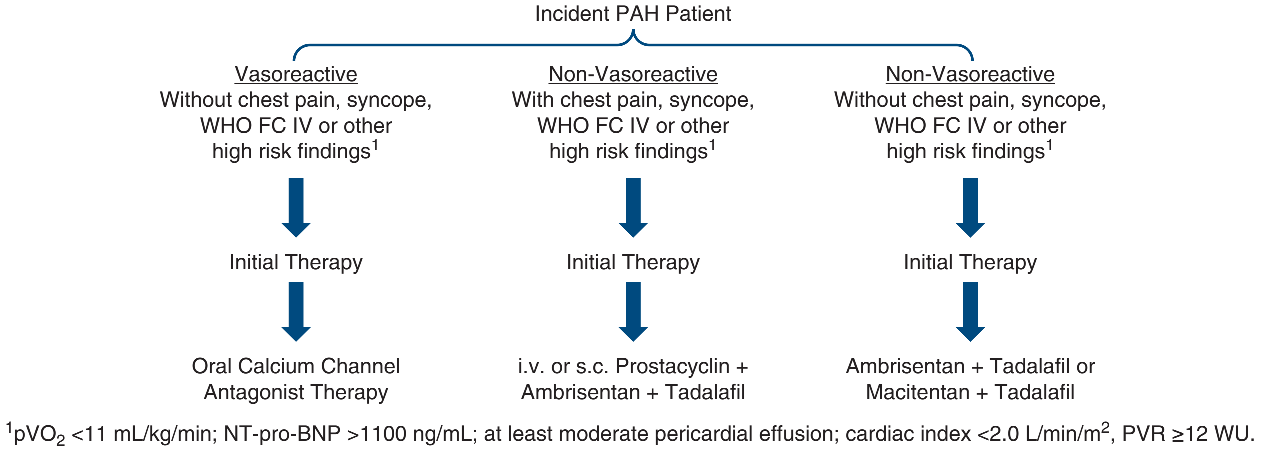
Incident PAH Patient
Vasoreactive
Without chest pain, syncope, WHO FC IV or other high risk findings[1]
Non-Vasoreactive
With chest pain, syncope, WHO FC IV or other high risk findings[1]
Non-Vasoreactive
Without chest pain, syncope, WHO FC IV or other high risk findings[1]
Initial Therapy
Initial Therapy
Initial Therapy
Oral Calcium Channel Antagonist Therapy
i.v. or s.c. Prostacyclin + Ambrisentan + Tadalafil
Ambrisentan + Tadalafil or Macitentan + Tadalafil
[1]pVO_2 <11 mL/kg/min; NT-pro-BNP >1100 ng/mL; at least moderate pericardial effusion; cardiac index <2.0 L/min/m^2, PVR ≥12 WU.

FIGURE 294-11 Treatment strategy overview for patients with newly diagnosed pulmonary arterial hypertension (PAH). Newly diagnosed patients with idiopathic, hereditary, or drug-induced PAH should be considered for vasoreactivity testing in the cardiac catheterization laboratory at an expert pulmonary hypertension center. In the absence of a high-risk clinical profile, patients who demonstrate a positive vasoreactivity response, defined by decrease in mean pulmonary artery pressure ≥10 mmHg from baseline to ≤40 mmHg without a decrease in cardiac output, should be initiated on calcium channel antagonist therapy dose titrated to optimal clinical benefit/adverse effect balance. For patients with PAH without evidence of vasoreactivity but with high-risk findings, consideration of up-front therapy with the prostacyclin analogue treprostinil administered by intravenous (IV) or subcutaneous (SC) route plus the phosphodiesterase type 5 inhibitor tadalafil and endothelin receptor antagonist ambrisentan is indicated. For patients with PAH without vasoreactivity or high-risk findings, initial combination therapy with tadalafil and ambrisentan or the alternate endothelin receptor antagonist macitentan should be considered. NT-proBNP, N-terminal pro-B-type natriuretic peptide; pVO_2, peak volume of oxygen consumption; PVR, pulmonary vascular resistance; WHO-FC, World Health Organization Functional Class; WU, Wood units. *(Reproduced with permission from BA Maron: Revised definition of pulmonary hypertension and approach to management: A clinical primer. J Am Heart Assoc 12:e029024, 2023.)*

of treatment attrition is notable. It is also unclear if inhaled iloprost is generalizable to other lung-PH phenotypes, as at least one study focusing on COPD-PH was terminated early, presumably owing to a lack of benefit or adverse effects.

Treatment of CTEPH Pulmonary endarterectomy performed at a high-volume surgical referral center is the preferred therapy in patients with favorable anatomy and profile. Distal fibrothrombotic remodeling, PVR >12 WU, right heart failure, and NYHA FC IV are viewed as high-risk findings and may be prohibitive in terms of postoperative risk. Patients who are ineligible for pulmonary endarterectomy or do not experience a complete clinical result in the postoperative phase should be considered for treatment with riociguat (sGC stimulator) therapy, balloon pulmonary angioplasty (BPA), or a combination of both. Compared to riociguat, BPA modulates a decrease in mPAP of −9.3 mmHg but requires on average five procedural attempts and is associated with some important potential complications, including hemoptysis and reperfusion pulmonary edema **(Fig. 294-12)**.

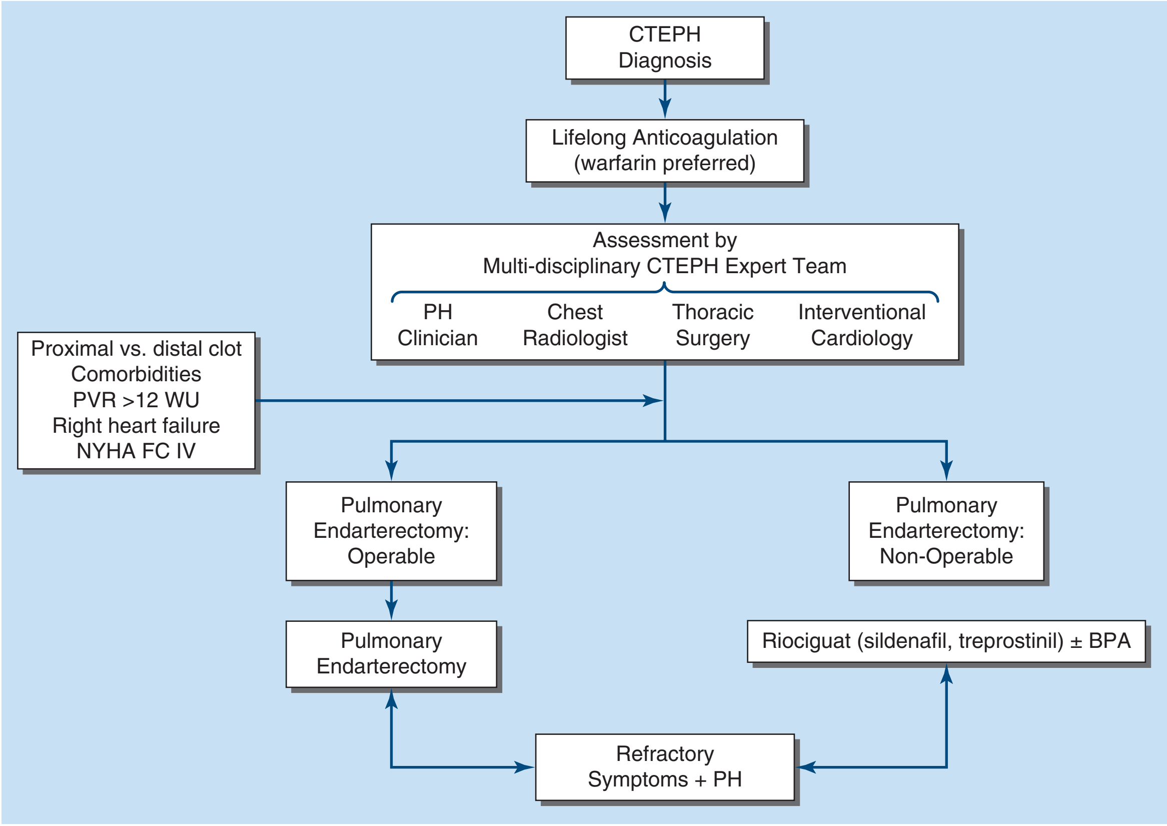
CTEPH Diagnosis
Lifelong Anticoagulation (warfarin preferred)
Assessment by Multi-disciplinary CTEPH Expert Team
PH Clinician
Chest Radiologist
Thoracic Surgery
Interventional Cardiology
Proximal vs. distal clot
Comorbidities
PVR >12 WU
Right heart failure
NYHA FC IV
Pulmonary Endarterectomy: Operable
Pulmonary Endarterectomy: Non-Operable
Pulmonary Endarterectomy
Riociguat (sildenafil, treprostinil) ± BPA
Refractory Symptoms + PH

FIGURE 294-12 Algorithm for the management of chronic thromboembolic pulmonary hypertension (CTEPH). BPA, balloon pulmonary angioplasty; NYHA FC, New York Heart Association functional class; PH, pulmonary hypertension; PVR, pulmonary vascular resistance. *(Adapted from NH Kim et al: Eur Respir J 53:1801915, 2019.)*

TABLE 294-2 FDA-Approved Therapies for the Treatment of Pulmonary Arterial Hypertension (PAH)

GENERIC NAME	ROUTE OF ADMINISTRATION	DRUG CLASS	INDICATION
Epoprostenol	IV	Prostacyclin derivative	Treatment of PAH to improve exercise capacity
Iloprost	Inhaled	Prostacyclin derivative	Treatment of PAH to improve a composite endpoint consisting of exercise tolerance, symptoms (NYHA class), and lack of deterioration
Treprostinil	IV or SC	Prostacyclin derivative	Treatment of PAH to diminish symptoms associated with exercise
Treprostinil	Inhaled	Prostacyclin derivative	Treatment of PAH to improve exercise ability
Treprostinil	Oral	Prostacyclin derivative	Treatment of PAH to improve exercise ability
Selexipag	Oral	Selective IP receptor agonist	Treatment of PAH to improve a composite endpoint lack of clinical deterioration
Bosentan	Oral	Endothelin receptor antagonist	Treatment of PAH to improve exercise capacity and to decrease clinical worsening
Ambrisentan	Oral	Endothelin receptor antagonist	Treatment of PAH to improve exercise capacity and delay clinical worsening
Macitentan	Oral	Endothelin receptor antagonist	Treatment of PAH to improve a composite endpoint of delay of clinical worsening
Sildenafil	Oral or IV	PDE5 inhibitor	Treatment of PAH to improve exercise capacity and delay clinical worsening
Tadalafil	Oral	PDE5 inhibitor	Treatment of PAH to improve exercise ability
Riociguat	Oral	Soluble guanylyl cyclase stimulator	Treatment of PAH to improve exercise ability

Abbreviations: FDA, U.S. Food and Drug Administration; IV, intravenous; NYHA, New York Heart Association; PAH, pulmonary arterial hypertension; PDE5, phosphodiesterase-5; SC, subcutaneous.

UNMET AND FUTURE RESEARCH NEEDS IN PULMONARY HYPERTENSION

Delayed diagnosis is a major barrier to improving outcome in PAH. Improved awareness among clinicians and patients could lead to more timely diagnosis that will affect the response to therapy and survival. At-risk patients should be referred early to a specialty center that focuses on treatment of patients with pulmonary vascular disease, which will ensure their access to state-of-the-art (multidisciplinary) care. In addition, the role of currently available drugs in early-stage disease is not known and requires further investigation **(Table 294-2)**. The anticipated approval of sotatercept is exciting and paves a path toward the availability of the first novel drug class in PAH in nearly two decades. However, clinical trial data continue to lack information from diverse patient subgroups, suggesting that health inequity in PH remains a critical opportunity for progress.

FURTHER READING

Alba GA et al: NEDD9 is a novel and modifiable mediator of platelet-endothelial adhesion in the pulmonary circulation. Am J Respir Crit Care Med 203:1533, 2021.

Bernardo RJ et al: Health care disparities in pulmonary arterial hypertension. Clin Chest Med 44:543, 2023.

Hemnes AR et al: Clinical characteristics and transplant-free survival across the spectrum of pulmonary vascular disease. J Am Coll Cardiol 80:697, 2022.

Ho JE et al: Exercise pulmonary hypertension predicts clinical outcomes in patients with dyspnea on effort. J Am Coll Cardiol 75:17, 2020.

Hoeper MM et al: Phase 3 trial of sotatercept for treatment of pulmonary arterial hypertension. N Engl J Med 388:1478, 2023.

Johnson S et al: Pulmonary hypertension: A contemporary review. Am J Respir Crit Care Med 208:528, 2023.

Sommer N et al: Current and future treatments of pulmonary arterial hypertension. Br J Pharmacol 178:6, 2021.

Walters R et al: SOX17 enhancer variants disrupt transcription factor binding and enhancer inactivity drives pulmonary hypertension. Circulation 147:1606, 2023.

Yung LM et al: ACTRIIA-Fc rebalances activin/GDF versus BMP signaling in pulmonary hypertension. Sci Transl Med 12:eaaz5660, 2020.

Zeder K et al: Elevated pulmonary vascular resistance predicts mortality in COPD patients. Eur Respir J 58:2100944, 2021.

This page intentionally left blank

Disorders of the Respiratory System PART 7

Section 1 Diagnosis of Respiratory Disorders

295 Approach to the Patient with Disease of the Respiratory System

Bruce D. Levy

The majority of diseases of the respiratory system present with cough and/or dyspnea and fall into one of three major categories: (1) obstructive; (2) restrictive; and (3) vascular diseases. Obstructive pathophysiology is most common and primarily results from airway diseases, such as asthma, chronic obstructive pulmonary disease (COPD), bronchiectasis, and bronchiolitis. Diseases resulting in restrictive pathophysiology include parenchymal lung diseases, abnormalities of the chest wall and pleura, and neuromuscular disease. Pulmonary embolism, pulmonary hypertension, and pulmonary venoocclusive disease are examples of disorders of the pulmonary vasculature. Although many specific diseases fall into these major categories, both infective and neoplastic processes can affect the respiratory system and result in myriad pathologic findings, including those listed in the three categories above (Table 295-1).

Disorders can also be grouped according to gas exchange abnormalities, including hypoxemia, hypercarbia, or combined impairment; however, many respiratory disorders do not manifest as gas exchange abnormalities.

As with the evaluation of most patients, the approach to a patient with a respiratory system disorder begins with a thorough history and a focused physical examination. Many patients will subsequently undergo pulmonary function testing, chest imaging, blood and sputum analysis, a variety of serologic or microbiologic studies, and diagnostic procedures, such as bronchoscopy. This stepwise approach is discussed in detail below.

TABLE 295-1 Categories of Respiratory Disease

CATEGORY	EXAMPLES
Obstructive pathophysiology—airway disease	Asthma Chronic obstructive pulmonary disease (COPD) Bronchiectasis Bronchiolitis
Restrictive pathophysiology—parenchymal disease	Idiopathic pulmonary fibrosis (IPF) Asbestosis Desquamative interstitial pneumonitis (DIP) Sarcoidosis
Restrictive pathophysiology—neuromuscular weakness	Amyotrophic lateral sclerosis (ALS) Guillain-Barré syndrome Myasthenia gravis
Restrictive pathophysiology—chest wall/pleural disease	Kyphoscoliosis Ankylosing spondylitis Chronic pleural effusions
Pulmonary vascular disease	Pulmonary embolism Pulmonary arterial hypertension (PAH) Pulmonary venoocclusive disease Vasculitis
Malignancy	Bronchogenic carcinoma (non-small-cell and small-cell lung cancer) Metastatic disease
Infectious diseases	Pneumonia Bronchitis Tracheitis

HISTORY

Dyspnea and Cough The cardinal symptoms of respiratory disease are dyspnea and cough (Chaps. 39 and 40). Dyspnea has many causes, some of which are not predominantly due to lung pathology. The words a patient uses to describe shortness of breath can suggest certain etiologies for dyspnea. Patients with obstructive lung disease often complain of "chest tightness" or "inability to get a deep breath," whereas patients with congestive heart failure more commonly report "air hunger" or a sense of suffocation.

The tempo of onset and the duration of a patient's dyspnea are likewise helpful in determining the etiology. Acute shortness of breath is usually associated with sudden physiologic changes, such as acute airway narrowing (e.g., laryngeal edema, bronchospasm, or mucus plugging), acute hypoxemia (e.g., pulmonary edema, pneumonia, or pulmonary embolism), or sudden changes in the work of breathing (e.g., pneumothorax). Patients with COPD and idiopathic pulmonary fibrosis (IPF) experience a gradual progression of dyspnea on exertion, punctuated by acute exacerbations of shortness of breath. In contrast, most asthmatics do not have daily symptoms, but experience intermittent episodes of dyspnea, cough, and chest tightness that are usually associated with specific triggers, such as an upper respiratory tract infection or exposure to allergens.

Specific questioning should focus on factors that incite dyspnea as well as on any intervention that helps resolve the patient's shortness of breath. Asthma is commonly exacerbated by specific triggers, although this can also be true of COPD. Many patients with lung disease report dyspnea on exertion. Determining the degree of activity that results in shortness of breath gives the clinician a gauge of the patient's degree of disability. Many patients adapt their level of activity to accommodate progressive limitation. For this reason, it is important, particularly in older patients, to delineate the activities in which they engage and how these activities have changed over time. Dyspnea on exertion is often an early symptom of underlying lung or heart disease and warrants a thorough evaluation.

For cough, the clinician should inquire about the duration of the cough, whether or not it is associated with sputum production, and any specific triggers that induce it. Acute cough productive of phlegm is often a symptom of infection of the respiratory system, including processes affecting the upper airway (e.g., sinusitis, tracheitis), the lower airways (e.g., bronchitis, bronchiectasis), and the lung parenchyma (e.g., pneumonia). Both the quantity and quality of the sputum, including whether it is blood-streaked or frankly bloody, should be determined. Hemoptysis warrants urgent evaluation as delineated in Chap. 41.

Chronic cough (defined as that persisting for >8 weeks) is commonly associated with obstructive lung diseases, particularly asthma, COPD, and chronic bronchiectasis, as well as "nonrespiratory" diseases, such as gastroesophageal reflux and postnasal drip. Diffuse parenchymal lung diseases, including IPF, frequently present as a persistent, nonproductive cough. All causes of cough are not respiratory in origin, and assessment should encompass a broad differential, including cardiac and gastrointestinal diseases as well as psychogenic causes.

Additional Symptoms Patients with respiratory disease may report wheezing, which is suggestive of airways disease, particularly asthma. Hemoptysis can be a symptom of a variety of lung diseases, including infections of the respiratory tract, bronchogenic carcinoma, and pulmonary embolism. In addition, chest pain or discomfort can be respiratory in origin. As the lung parenchyma is not innervated with pain fibers, pain in the chest from respiratory disorders usually results from either diseases of the parietal pleura (e.g., pneumothorax) or pulmonary vascular diseases (e.g., pulmonary hypertension). As many diseases of the lung can result in strain on the right side of the heart, patients may also present with symptoms of cor pulmonale, including abdominal bloating or distention and pedal edema (Chap. 264).

Additional History A thorough social history is an essential component of the evaluation of patients with respiratory disease. All patients should be asked about current or previous cigarette smoking, as this exposure is associated with many diseases of the respiratory system, including COPD, bronchogenic lung cancer, and select parenchymal lung diseases (e.g., desquamative interstitial pneumonitis and pulmonary Langerhans cell histiocytosis). For most of these disorders, increased cigarette smoke exposure (i.e., cigarette pack-years) increases the risk of disease. E-cigarette or vaping use can lead to acute or subacute lung injury (i.e., E-cigarette or vaping use-associated lung injury [EVALI]). Secondhand smoke also increases risk for some respiratory disorders, so patients should also be asked about parents, spouses, or housemates who smoke. Possible inhalational exposures at work (e.g., asbestos, silica) or home (e.g., wood smoke, excrement from pet birds) should be explored (Chap. 300). Travel predisposes one to certain infections of the respiratory tract, most notably tuberculosis. Potential exposure to fungi is increased in specific geographic regions or climates (e.g., *Histoplasma capsulatum*), so exposures to these regions should be determined.

Associated symptoms of fever and chills should raise the suspicion of infective etiologies, both pulmonary and systemic. A comprehensive review of systems may suggest rheumatologic or autoimmune disease presenting with respiratory tract manifestations. Questions should focus on joint pain or swelling, rashes, dry eyes, dry mouth, or constitutional symptoms. In addition, carcinomas from a variety of primary sources commonly metastasize to the lung and cause respiratory symptoms. Finally, therapy for other conditions, including both irradiation and medications, can result in diseases of the chest.

Physical Examination The clinician's suspicion of respiratory disease often begins with the patient's vital signs. The respiratory rate is informative, whether elevated (tachypnea) or depressed (hypopnea). In addition, pulse oximetry should be measured, as many patients with respiratory disease have hypoxemia, either at rest or with exertion.

The first step of the physical examination is inspection. Patients with respiratory disease may be in distress and using accessory muscles of respiration to breathe. Severe kyphoscoliosis can result in restrictive pathophysiology. Inability to complete a sentence in conversation is generally a sign of severe impairment and should result in an expedited evaluation of the patient.

Percussion of the chest is used to establish diaphragm excursion and lung size. In the setting of decreased breath sounds, percussion is used to distinguish between pleural effusions (dull to percussion) and pneumothorax (hyper-resonant note).

The role of palpation is limited in the respiratory examination. Palpation can demonstrate subcutaneous air in the setting of barotrauma. It can also be used as an adjunctive assessment to determine whether an area of decreased breath sounds is due to consolidation (increased tactile fremitus) or a pleural effusion (decreased tactile fremitus). To detect unilateral disorders of ventilation, the symmetry and degree of chest wall expansion can be assessed during a deep inspiration by placing one's thumbs together at the midline over the lower posterior chest while grasping the lateral rib cage.

The majority of the manifestations of respiratory disease present as abnormalities of auscultation. Wheezes are a manifestation of airway obstruction. While most commonly a sign of asthma, peribronchial edema in the setting of congestive heart failure can also result in diffuse wheezes, as can any other process that causes narrowing of small airways. Wheezes can be polyphonic, involving multiple different size airways (e.g., asthma), or monophonic, involving one size airway (e.g., bronchogenic carcinoma). For these reasons, clinicians must take care not to attribute all wheezing to asthma.

Rhonchi are a manifestation of obstruction of medium-sized airways, most often with secretions. In the acute setting, this manifestation may be a sign of viral or bacterial bronchitis. Chronic rhonchi suggest bronchiectasis or COPD. In contrast to expiratory wheezes and rhonchi, stridor is a high-pitched, focal inspiratory wheeze, usually heard over the neck as a manifestation of upper airway obstruction.

Crackles, or rales, are commonly a sign of alveolar disease. Processes that fill the alveoli with fluid may result in crackles, including pulmonary edema and pneumonia. Crackles in pulmonary edema are generally more prominent at the bases. Interestingly, diseases that result in fibrosis of the interstitium (e.g., IPF) also result in crackles that sound like Velcro being ripped apart. Although some clinicians make a distinction between "wet" and "dry" crackles, this distinction has not been shown to be a reliable way to differentiate among etiologies of respiratory disease.

One way to help distinguish between crackles associated with alveolar fluid and those associated with interstitial fibrosis is to assess for egophony. *Egophony* is the auscultation of the sound "AH" instead of "EEE" when a patient phonates "EEE." This change in note is due to abnormal sound transmission through consolidated parenchyma and is present in pneumonia but not in IPF. Similarly, areas of alveolar filling have increased whispered *pectoriloquy* as well as transmission of larger-airway sounds (i.e., bronchial breath sounds in a lung zone where vesicular breath sounds are expected).

The lack or diminution of breath sounds can also help determine the etiology of respiratory disease. Patients with emphysema often have a quiet chest with diffusely decreased breath sounds. A pneumothorax or pleural effusion may present with an area of absent breath sounds.

Other Systems Pedal edema, if symmetric, may suggest cor pulmonale; if asymmetric, it may be due to deep venous thrombosis and associated pulmonary embolism. Jugular venous distention may also be a sign of volume overload associated with right heart failure. *Pulsus paradoxus* is an ominous sign in a patient with obstructive lung disease, as it is associated with significant negative intrathoracic (pleural) pressures required for ventilation and impending respiratory failure.

As stated earlier, rheumatologic disease may manifest primarily as lung disease. Owing to this association, particular attention should be paid to joint and skin examination. Clubbing can be found in many lung diseases, including cystic fibrosis, IPF, and lung cancer. Cyanosis is seen in hypoxemic respiratory disorders that result in >5 g of deoxygenated hemoglobin/dL.

DIAGNOSTIC EVALUATION

The sequence of studies is dictated by the clinician's differential diagnosis, as determined by the history and physical examination. Acute respiratory symptoms are often evaluated with multiple tests performed at the same time in order to diagnose any life-threatening diseases rapidly (e.g., pulmonary embolism or multilobar pneumonia). In contrast, chronic dyspnea and cough can be evaluated in a more protracted, stepwise fashion.

Pulmonary Function Testing (See also Chap. 297) The initial pulmonary function test obtained is spirometry. This study is an effort-dependent test used to assess for obstructive pathophysiology as seen in asthma, COPD, and bronchiectasis (Table 295-1). A diminished-forced expiratory volume in 1 second (FEV_1)/forced vital capacity (FVC) (often defined as <70%) is diagnostic of airflow obstruction. In addition to measuring FEV_1 and FVC, the clinician should examine the flow-volume loop (which is less effort-dependent). A plateau of the inspiratory and expiratory curves suggests large-airway obstruction in extrathoracic and intrathoracic locations, respectively.

Spirometry with symmetric decreases in FEV_1 and FVC warrants further testing, including measurement of lung volumes and the diffusion capacity of the lung for carbon monoxide (DL_{CO}). A total lung capacity <80% of the patient's predicted value defines restrictive pathophysiology. Restriction can result from parenchymal disease, neuromuscular weakness, or chest wall or pleural diseases (Table 295-1). Restriction with impaired gas exchange, as indicated by a decreased DL_{CO}, suggests parenchymal lung disease. Additional testing, such as measurements of maximal inspiratory and expiratory pressures, can help diagnose neuromuscular weakness. Normal spirometry, normal lung volumes, and a low DL_{CO} should prompt further evaluation for pulmonary vascular disease.

Arterial blood gas testing is often helpful in assessing respiratory disease. Hypoxemia, while usually apparent with pulse oximetry, can be further evaluated with the measurement of arterial PO_2 and the calculation of an alveolar gas and arterial blood oxygen tension difference ($[A–a]DO_2$). Patients with diseases that cause ventilation-perfusion mismatch or shunt physiology have an increased $(A–a)DO_2$ at rest. Arterial blood gas testing also allows the measurement of arterial PCO_2. Hypercarbia can accompany disorders of ventilation, as seen in severe airway obstruction (e.g., COPD) or progressive restrictive physiology.

Chest Imaging (See Chap. A12) Most patients with disease of the respiratory system undergo imaging of the chest as part of the initial evaluation. Clinicians should generally begin with ultrasound of the chest or a plain chest radiograph, preferably posterior-anterior and lateral films. Ultrasound is often readily available and can help rapidly diagnose pneumothorax, pleural effusion, and consolidation of lung parenchyma. Chest radiographs give additional detail and can reveal findings including opacities of the parenchyma, blunting of the costophrenic angles, mass lesions, and volume loss. Of note, many diseases of the respiratory system, particularly those of the airways and pulmonary vasculature, are associated with a normal chest radiograph.

Computed tomography (CT) scan of the chest can also be useful to delineate parenchymal processes, pleural disease, masses or nodules, and large airways. If the test includes administration of intravenous contrast, the pulmonary vasculature can be assessed with particular utility for determination of pulmonary emboli. Intravenous contrast also allows lymph nodes to be examined in greater detail. When coupled with positron emission tomography (PET), lesions of the chest can be assessed for metabolic activity, helping differentiate between malignancy and scar.

FURTHER STUDIES

Depending on the clinician's suspicion, a variety of other studies may be done. Concern about large-airway lesions may warrant bronchoscopy. This procedure may also be used to sample the alveolar space with bronchoalveolar lavage or to obtain nonsurgical lung biopsies. Blood testing may include assessment for hypercoagulable states in the setting of pulmonary vascular disease, serologic testing for infectious or rheumatologic disease, or assessment of inflammatory markers or leukocyte counts (e.g., eosinophils). Genetic testing is increasingly used for heritable lung diseases such as cystic fibrosis. Sputum evaluation for malignant cells or microorganisms may be appropriate. An echocardiogram to assess right- and left-sided heart function is often obtained. Finally, at times, a surgical lung biopsy is needed to diagnose certain diseases of the respiratory system. All of these studies will be guided by the preceding history, physical examination, pulmonary function testing, and chest imaging.

FURTHER READING

Bohadana A et al: Fundamentals of lung auscultation. N Engl J Med 370:744, 2014.

Chung KF et al: Cough hypersensitivity and chronic cough. Nat Rev Dis Primers 8:45, 2022.

García-de-Acilu M et al: Use of thoracic ultrasound in acute respiratory distress syndrome. Ann Transl Med 11:320, 2023.

Mojoli F et al: Lung ultrasound for critically ill patients. Am J Resp Crit Care Med 199:701, 2019.

Parshall MB et al: An official American Thoracic Society statement: Update on the mechanisms, assessment, and management of dyspnea. Am J Respir Crit Care Med 185:435, 2012.

Pellegrino R et al: Interpretive strategies for lung function tests. Eur Respir J 26:948, 2005.

Stanojevic S et al: ERS/ATS technical standard on interpretative strategies for routine lung function tests. Eur Respir J 60:2101499, 2022.

296

Disturbances of Respiratory Function

George R. Washko, William M. Oldham

The primary functions of the respiratory system—to oxygenate blood and eliminate carbon dioxide—require virtual contact between blood and fresh air, which facilitates diffusion of respiratory gases between blood and gas. This process occurs in the lung alveoli, where blood flowing through alveolar wall capillaries is separated from alveolar gas by an extremely thin membrane of flattened endothelial and epithelial cells, across which respiratory gases diffuse and equilibrate. Blood flow through the lung is unidirectional via a continuous vascular path along which venous blood absorbs oxygen from and loses CO_2 to inspired gas. The path for airflow, in contrast, reaches a dead end at the alveolar walls; thus, the alveolar space must be ventilated tidally, with inflow of fresh gas and outflow of alveolar gas alternating periodically at the respiratory rate (RR). To provide an enormous alveolar surface area (typically 70 m^2) for blood-gas diffusion within the modest volume of a thoracic cavity (typically 7 L), nature has distributed both blood flow and ventilation among millions of tiny alveoli through multigenerational branching of both pulmonary arteries and bronchial airways. Ideally, for the lung to be most efficient in exchanging gas, the fresh gas ventilation of a given alveolus must be matched to its perfusion. However, as a consequence of variations in tube lengths and calibers along these pathways as well as the effects of gravity, tidal pressure fluctuations, and anatomic constraints from the chest wall, the alveoli vary in their relative ventilations and perfusions even in health.

For the respiratory system to succeed in oxygenating blood and eliminating CO_2, it must be able to ventilate the lung tidally and thus to freshen alveolar gas; it must provide for perfusion of the individual alveolus in a manner proportional to its ventilation; and it must allow adequate diffusion of respiratory gases between alveolar gas and capillary blood. Furthermore, it must accommodate several-fold increases in the demand for oxygen uptake or CO_2 elimination imposed by metabolic needs or acid-base derangement. Given these multiple requirements for normal operation, it is not surprising that many diseases disturb respiratory function. This chapter considers in some detail the physiologic determinants of lung ventilation and perfusion, elucidates how the matching distributions of these processes and rapid gas diffusion allow normal gas exchange, and discusses how common diseases derange these normal functions, thereby impairing gas exchange—or at least increase the work required by the respiratory muscles or heart to maintain adequate respiratory function.

VENTILATION

It is useful to conceptualize the respiratory system as three independently functioning components: the lung, including its airways; the neuromuscular system; and the chest wall, which includes everything that is not lung or active neuromuscular system. Accordingly, the mass of the respiratory muscles is part of the chest wall, while the force these muscles generate is part of the neuromuscular system; the abdomen (especially an obese abdomen) and the heart (especially an enlarged heart) are, for these purposes, part of the chest wall. Each of these three components has mechanical properties that relate to its enclosed volume (or—in the case of the neuromuscular system—the respiratory system volume at which it is operating) and to the rate of change of its volume (i.e., flow). The work of breathing required of the neuromuscular system is the sum of the work due to volume-related mechanical properties and the work from flow-related mechanical properties required to move air throughout the airways to create this volume change.

Volume-Related Mechanical Properties—Statics Figure 296-1 shows the volume-related properties of each component of the respiratory system. The natural tendency of the lung is to collapse because

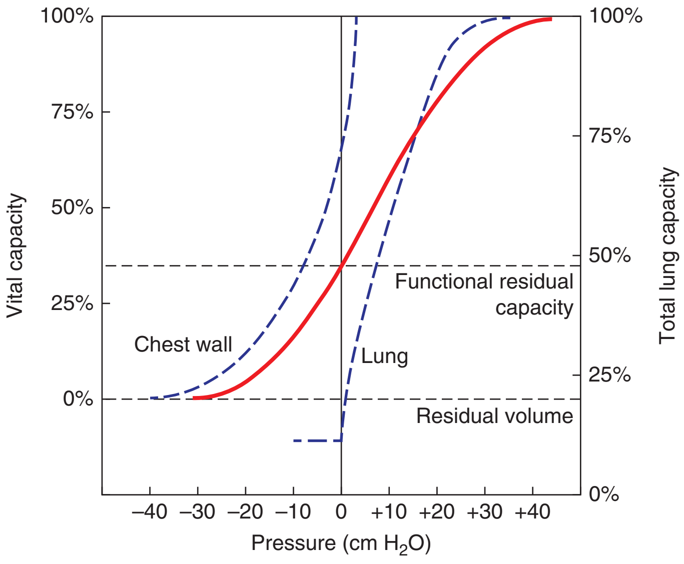

FIGURE 296-1 Pressure-volume curves of the isolated lung, isolated chest wall, combined respiratory system, inspiratory muscles, and expiratory muscles. FRC, functional residual capacity; RV, residual volume; TLC, total lung capacity.

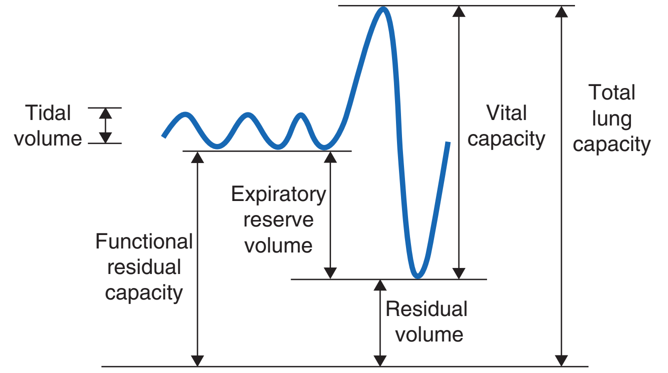

FIGURE 296-2 Spirogram demonstrating a slow vital capacity maneuver and various lung volumes.

of both surface tension at the air-liquid interface between alveolar wall lining fluid and alveolar gas and elastic recoil of the lung tissue itself. To stay inflated, the pressure within the alveolus must equal or exceed the pressure at the pleural surface. This difference in pressures ($P_{alveolus} - P_{pleura}$) is expressed as the transpulmonary pressure.

The elastic recoil of the lung is not constant, increasing with inflation. At high lung volume, the lung becomes rather stiff, so that large changes in transpulmonary pressure are required for relatively small changes in lung volume. In contrast, the lung is compliant at lower volumes, including those at which tidal breathing normally occurs. At zero inflation pressure, even normal lungs retain some air in the alveoli. Because the small peripheral airways are tethered open by outward radial pull from inflated lung parenchyma attached to adventitia, as the lung deflates during exhalation, those small airways are pulled open progressively less, and eventually close, trapping some gas in the alveoli. This effect can be exaggerated with age and especially with obstructive airway diseases, resulting in gas trapping at quite large lung volumes.

Functional residual capacity (FRC) is the passive resting point of the respiratory system attained when the outward recoil of the chest wall is balanced exactly by the inward recoil of the lung. The elastic behavior of the passive chest wall (i.e., in the absence of neuromuscular activation), differs markedly from that of the lung. While the lungs become stiff at high volumes, the chest wall stiffens at low volumes due to squeezing together of ribs and intercostal muscles, diaphragm stretch, displacement of abdominal contents, and straining of ligaments and bony articulations. The normal lung and chest wall function in mechanical series, and the pressure required to displace the passive respiratory system (lungs plus chest wall) at any volume is simply the sum of the elastic recoil pressure of the lungs and the transmural pressure across the chest wall. When plotted against respiratory system volume, this relationship assumes a sigmoid shape, exhibiting stiffness at high lung volumes (imparted by the lung), stiffness at low lung volumes (imparted by the chest wall or sometimes by airway closure), and compliance in the middle range of lung volumes where normal tidal breathing occurs. As these recoils are transmitted through the pleural fluid, the lung is pulled both outward and inward simultaneously at FRC, where the negative intrapleural pressure is exactly offset by the positive intrapulmonary pressure yielding an airway pressure of 0 mmHg.

The normal passive respiratory system would equilibrate at the FRC and remain there were it not for the actions of the respiratory muscles. Gas flows from high to low pressure and the lung inflates when pressure at the airway opening exceeds pressure in the alveoli. The lung deflates when pressure in the alveoli exceeds pressure at the airway opening. The inspiratory muscles act on the chest wall to expand the volume of the thorax, decreasing pleural pressure, lowering pressure in the alveoli below pressure at the airway opening. In contrast, the expiratory muscles raise the alveolar gas pressure above pressure at the airway opening, leading to an outflow of gas from the lung.

The maximal pressures these sets of muscles can generate vary with the lung volume at which they operate. This variation is due to length-tension relationships in striated muscle sarcomeres and to changes in mechanical advantage as the angles of insertion change with lung volume (Fig. 296-1). Nonetheless, under normal conditions, the respiratory muscles are substantially "overpowered" for their roles and generate more than adequate force to drive the respiratory system to its stiffness extremes, as determined by the lung (total lung capacity [TLC]) or by chest wall or airway closure (residual volume [RV]); the airway closure always prevents the adult lung from emptying completely under normal circumstances. The excursion between full and minimal lung inflation is called *vital capacity* (VC; **Fig. 296-2**). The VC is easy to measure (see below), but it provides little information about the intrinsic properties of the respiratory system. As will become clear, it is much more useful for the clinician to consider TLC and RV individually.

Flow-Related Mechanical Properties—Dynamics The passive chest wall and active neuromuscular system both exhibit mechanical behaviors related to the rate of change of volume, but these behaviors become quantitatively important only at markedly supraphysiologic breathing frequencies (e.g., during high-frequency mechanical ventilation), and thus will not be addressed here. In contrast, the dynamic airflow properties of the lung substantially affect its ability to ventilate and contribute importantly to the work of breathing, and these properties are often deranged by disease. Understanding dynamic airflow properties is, therefore, worthwhile.

As with the flow of any fluid (gas or liquid) in any tube, maintenance of airflow within the pulmonary airways requires a pressure gradient that falls along the direction of flow, the magnitude of which is determined by the flow rate and the frictional resistance to flow. During quiet tidal breathing, the pressure gradients driving inspiratory or expiratory flow are small owing to the very low frictional resistance of normal pulmonary airways (R_{aw}, normally <2 $cmH_2O/L/s$). However, during rapid exhalation, another phenomenon reduces flow below that which would have been expected if frictional resistance were the only impediment to flow. This phenomenon is called *dynamic airflow limitation*, and it occurs because the bronchial airways through which air is exhaled are collapsible rather than rigid **(Fig. 296-3)**. An important anatomic feature of the structure of the pulmonary airways is their tree-like branching pattern. While the individual airways in each successive generation, from most proximal (trachea) to most distal (respiratory bronchioles), are smaller than those of the parent generation, their number increases exponentially such that the summed cross-sectional area of the airways becomes very large toward the lung periphery. Because flow (volume/time) is constant along the airway tree, the velocity of airflow (flow/summed cross-sectional area)

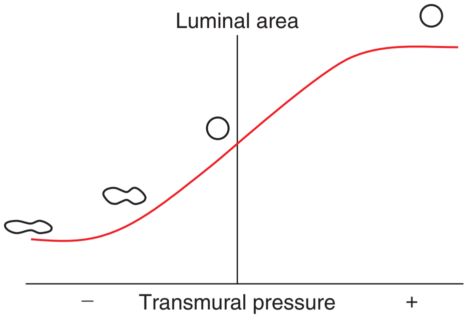

FIGURE 296-3 Luminal area versus transmural pressure relationship. Transmural pressure represents the pressure difference across the airway wall from inside to outside.

is much greater in the central airways than in the peripheral airways. During exhalation, gas leaving the alveoli must, therefore, gain velocity as it proceeds toward the mouth. This acceleration reduces intraluminal gas pressure and airway transmural pressure leading to a reduction in airway size. Referred to as the Bernoulli effect, this process may be best appreciated in the example of the airplane. As the flow of air accelerates over the curved surface of its wings, it provides lift to the plane (Fig. 296-3). If an individual attempts to exhale more forcefully, the local velocity increases further (increasing "lift") and airway size grows smaller, resulting in no net increase in flow. Under these circumstances, flow has reached its maximum possible value, or its *flow limit*.

Lungs normally exhibit such dynamic airflow limitation. This limitation can be assessed by spirometry, in which an individual inhales fully to TLC and then forcibly exhales to RV. Maximal expiratory flow at any lung volume is determined by gas density, airway cross-section and distensibility, elastic recoil pressure of the lung, and frictional pressure loss to the flow-limiting airway site. Under normal conditions, maximal expiratory flow falls with lung volume **(Fig. 296-4)**, primarily because of the dependence of lung recoil pressure on lung volume (Fig. 296-1). In pulmonary fibrosis, lung recoil pressure is increased at any lung volume, and thus the maximal expiratory flow is elevated when considered in relation to lung volume. Conversely, in emphysema, lung recoil pressure is reduced; this reduction is a principal mechanism by which maximal expiratory flows fall. Diseases that narrow the airway lumen at any transmural pressure (e.g., asthma or chronic bronchitis) or that cause excessive airway collapsibility (e.g., tracheomalacia) also reduce maximal expiratory flow.

The Bernoulli effect also applies during inspiration, but the more negative pleural pressures during inspiration lower the pressure outside of the airways, thereby increasing transmural pressure and promoting airway expansion. Thus, inspiratory airflow limitation seldom occurs due to diffuse pulmonary airway disease. Conversely, extrathoracic airway narrowing (e.g., due to a tracheal adenoma or posttracheostomy stricture) can lead to inspiratory airflow limitation (Fig. 296-4).

The Work of Breathing In health, the elastic (volume change-related) and dynamic (flow-related) loads that must be overcome to ventilate the lungs at rest are small, and the work required of the respiratory muscles is minimal. However, the work of breathing can increase considerably due to a metabolic requirement for substantially increased ventilation, an abnormally increased mechanical load, or both. As discussed below, the rate of ventilation is primarily set by the need to eliminate carbon dioxide, and thus, ventilation increases during exercise (sometimes by >20-fold) and during metabolic acidosis as a compensatory response. Naturally, the work rate required to overcome the elasticity of the respiratory system increases with both the depth and the frequency of tidal breaths, while the work required to overcome the dynamic load increases with total ventilation. A modest increase of ventilation is most efficiently achieved by increasing tidal volume but not RR, which is the normal ventilatory response to lower-level exercise. At higher levels of exercise, deep breathing persists, but RR also increases.

The work of breathing also increases when disease reduces the compliance of the respiratory system or increases the resistance to airflow. The former occurs commonly in diseases of the lung parenchyma (interstitial processes or fibrosis, alveolar filling diseases such as pulmonary edema or pneumonia, or substantial lung resection), and the latter occurs in obstructive airway diseases such as asthma, chronic bronchitis, emphysema, and cystic fibrosis. Furthermore, severe airflow obstruction can functionally reduce the compliance of the respiratory system by leading to dynamic hyperinflation. In this scenario, expiratory flows slowed by the obstructive airways disease may be

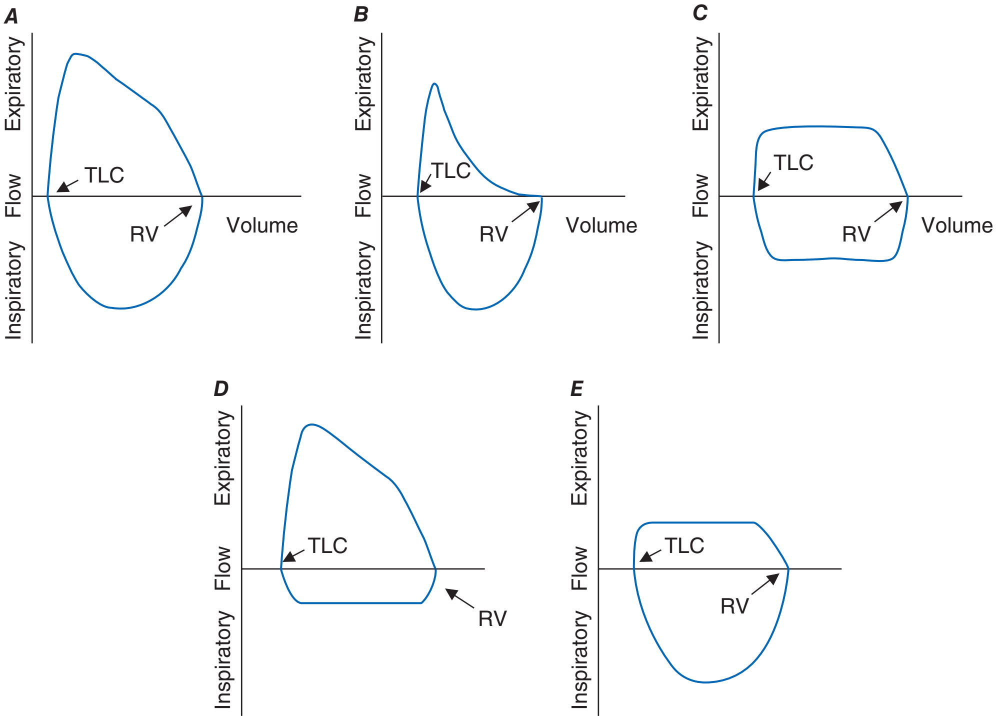

FIGURE 296-4 Flow-volume loops. ***A.*** Normal. ***B.*** Airflow obstruction. ***C.*** Fixed central airway obstruction (either above or below the thoracic inlet). ***D.*** Variable upper airway obstruction (above the thoracic inlet) ***E.*** Variable lower airway obstruction (below the thoracic inlet). RV, residual volume; TLC, total lung capacity.

insufficient to allow complete exhalation during the expiratory phase of tidal breathing; as a result, the "functional residual capacity (FRC)" from which the next breath is inhaled is greater than the static FRC. With repetition of incomplete exhalations of each tidal breath, the operating FRC becomes dynamically elevated, sometimes to a level that approaches TLC. At these high lung volumes, the respiratory system is much less compliant than at normal breathing volumes, and thus, the elastic work of each tidal breath is also increased. The dynamic pulmonary hyperinflation that accompanies severe airflow obstruction causes patients to sense difficulty in inhaling—even though the root cause of this pathophysiologic abnormality is expiratory airflow obstruction.

Adequacy of Ventilation As noted above, the respiratory control system that sets the rate of ventilation responds to chemical signals, including arterial CO_2 and oxygen tensions and blood pH, and to volitional needs, such as the need to inhale deeply before playing a long phrase on the trumpet. Disturbances in ventilation are discussed in **Chap. 307**. The focus of this chapter is on the relationship between ventilation of the lung and CO_2 elimination.

At the end of each tidal exhalation, the conducting airways are filled with alveolar gas that did not reach the mouth when expiratory flow stopped. During the ensuing inhalation, fresh gas immediately enters the airway tree at the mouth, but the gas first entering the alveoli at the start of inhalation is that same alveolar gas in the conducting airways that had just left the alveoli. Accordingly, fresh gas does not enter the alveoli until the volume of the conducting airways has been inspired. This volume is called the *anatomic dead space* (V_D). Quiet breathing with tidal volumes smaller than the anatomic dead space introduces no fresh gas into the alveoli at all; only that part of the inspired tidal volume (V_T) that is greater than the V_D introduces fresh gas into the alveoli. The dead space can be further increased functionally if some of the inspired tidal volume is delivered to a part of the lung that receives no pulmonary blood flow and thus cannot contribute to gas exchange (e.g., the portion of the lung distal to a large pulmonary embolus). In this situation, exhaled minute ventilation ($\dot{V}_E = V_T \times RR$) includes a component of dead space ventilation ($\dot{V}_D = V_D \times RR$) and a component of fresh gas alveolar ventilation ($\dot{V}_A = [V_T - V_D] \times RR$). Carbon dioxide elimination from the alveoli is equal to V_A times the difference in CO_2 fraction between inspired air (essentially zero) and alveolar gas (typically ~5.6% after correction for humidification of inspired air, corresponding to 40 mmHg).

In the steady state, the alveolar fraction of CO_2 is equal to metabolic CO_2 production divided by alveolar ventilation. Because, as discussed below, alveolar and arterial CO_2 tensions are equal, and because the respiratory controller normally strives to maintain arterial P_{CO_2} (Pa_{CO_2}) at ~40 mmHg, the adequacy of alveolar ventilation is reflected in Pa_{CO_2}. If the Pa_{CO_2} falls much below 40 mmHg, alveolar hyperventilation is present; if the Pa_{CO_2} exceeds 40 mmHg, alveolar hypoventilation is present. Ventilatory failure is characterized by extreme alveolar hypoventilation. In vivo, the production and clearance of CO_2 can be assessed through sampling of the arterial and central venous blood. CO_2 clearance can also be estimated noninvasively using capnography. Capnography enables visualization of respirophasic changes in CO_2 concentration at the airway opening, which at end of a tidal breath ($EtCO_2$) provides an estimate of ventilation.

As a consequence of oxygen uptake of alveolar gas into capillary blood, alveolar oxygen tension falls below that of inspired gas. The rate of oxygen uptake (determined by the body's metabolic oxygen consumption) is related to the average rate of metabolic CO_2 production, and their ratio—the "respiratory quotient" ($R = \dot{V}_{CO_2}/\dot{V}_{O_2}$)—depends largely on the fuel being metabolized. For a typical American diet, R is usually around 0.85. Together, these phenomena allow the estimation of alveolar oxygen tension, according to the following relationship, known as the *alveolar gas equation*:

$$Pa_{O_2} = F_{IO_2} \times (P_{bar} - P_{H_2O}) - Pa_{CO_2}/R$$

The alveolar gas equation also highlights the influences of inspired oxygen fraction F_{IO_2} barometric pressure (P_{bar}), and vapor pressure of water (P_{H_2O} = 47 mmHg at 37°C) in addition to alveolar ventilation (which sets Pa_{CO_2}) in determining Pa_{O_2}. An implication of the alveolar gas equation is that severe arterial hypoxemia rarely occurs as a pure consequence of alveolar hypoventilation at sea level while an individual is breathing air. The potential for alveolar hypoventilation to induce severe hypoxemia with otherwise normal lungs increases as P_{bar} falls with increasing altitude.

GAS EXCHANGE

Diffusion For oxygen to be delivered to the peripheral tissues, it must pass from alveolar gas into alveolar capillary blood by diffusing through alveolar membrane. The aggregate alveolar membrane is highly optimized for this process, with a very large surface area and minimal thickness. Diffusion through the alveolar membrane is so efficient in the human lung that in most circumstances hemoglobin of a red blood cell becomes fully oxygen saturated by the time the cell has traveled just one-third the length of the alveolar capillary. Thus, the uptake of alveolar oxygen is ordinarily limited by the amount of blood transiting the alveolar capillaries rather than by the rapidity with which oxygen can diffuse across the membrane; consequently, oxygen uptake from the lung is said to be "perfusion limited" rather than diffusion limited. CO_2 also equilibrates rapidly across the alveolar membrane. Therefore, the oxygen and CO_2 tensions in capillary blood leaving a normal alveolus are essentially equal to those in alveolar gas. Only in rare circumstances (e.g., at high altitude or in high-performance athletes exerting maximal effort) is oxygen uptake from normal lungs diffusion limited. Diffusion limitation can also occur in interstitial lung disease if substantially thickened alveolar walls remain perfused.

Ventilation/Perfusion Heterogeneity As noted above, for gas exchange to be most efficient, ventilation ($\dot{V}$) to each individual alveolus (among the millions of alveoli) should match perfusion ($\dot{Q}$) to its accompanying capillaries. Because of the differential effects of gravity on lung mechanics and blood flow throughout the lung and because of differences in airway and vascular architecture among various respiratory paths, there is minor ventilation/perfusion heterogeneity even in the normal lung; however, $\dot{V}/\dot{Q}$ heterogeneity can be particularly marked in disease. Two extreme examples are (1) ventilation of unperfused lung distal to a pulmonary embolus, in which ventilation of the physiologic dead space is "wasted" in the sense that it does not contribute to gas exchange; and (2) perfusion of nonventilated lung (a "shunt"), which allows venous blood to pass through the lung unaltered. When mixed with fully oxygenated blood leaving other well-ventilated lung units, shunted venous blood disproportionately lowers the mixed arterial Pa_{O_2} as a result of the nonlinear oxygen content versus PO_2 relationship of hemoglobin **(Fig. 296-5)**. Furthermore, the resulting arterial hypoxemia is refractory to supplemental inspired oxygen. The reason is that (1) raising the inspired F_{IO_2} has no effect on alveolar gas tensions in nonventilated alveoli and (2) while raising inspired F_{IO_2} increases Pa_{CO_2} in ventilated alveoli, the oxygen content of blood exiting ventilated units increases only slightly, as hemoglobin will already have been nearly fully saturated. Furthermore, the solubility of oxygen in plasma is quite small, and in normobaric conditions, the dissolved amount of oxygen in blood offers little additional physiologic benefit.

A more common occurrence than the two extreme examples given above is a widening of the distribution of ventilation/perfusion ratios; such $\dot{V}/\dot{Q}$ heterogeneity is a common consequence of lung disease. In this circumstance, perfusion of relatively underventilated alveoli results in the incomplete oxygenation of exiting blood. When mixed with well-oxygenated blood leaving higher $\dot{V}/\dot{Q}$ regions, this partially reoxygenated blood disproportionately lowers arterial Pa_{O_2}, although to a lesser extent than does a similar perfusion fraction of blood leaving regions of pure shunt. In addition, in contrast to shunt regions, inhalation of supplemental oxygen raises the Pa_{O_2} even in relatively underventilated low $\dot{V}/\dot{Q}$ regions, and so the arterial hypoxemia induced by $\dot{V}/\dot{Q}$ heterogeneity is typically responsive to oxygen therapy (Fig. 296-5).

In sum, arterial hypoxemia can be caused by substantial reduction of inspired oxygen tension, severe alveolar hypoventilation, perfusion of relatively underventilated (low $\dot{V}/\dot{Q}$) or completely unventilated

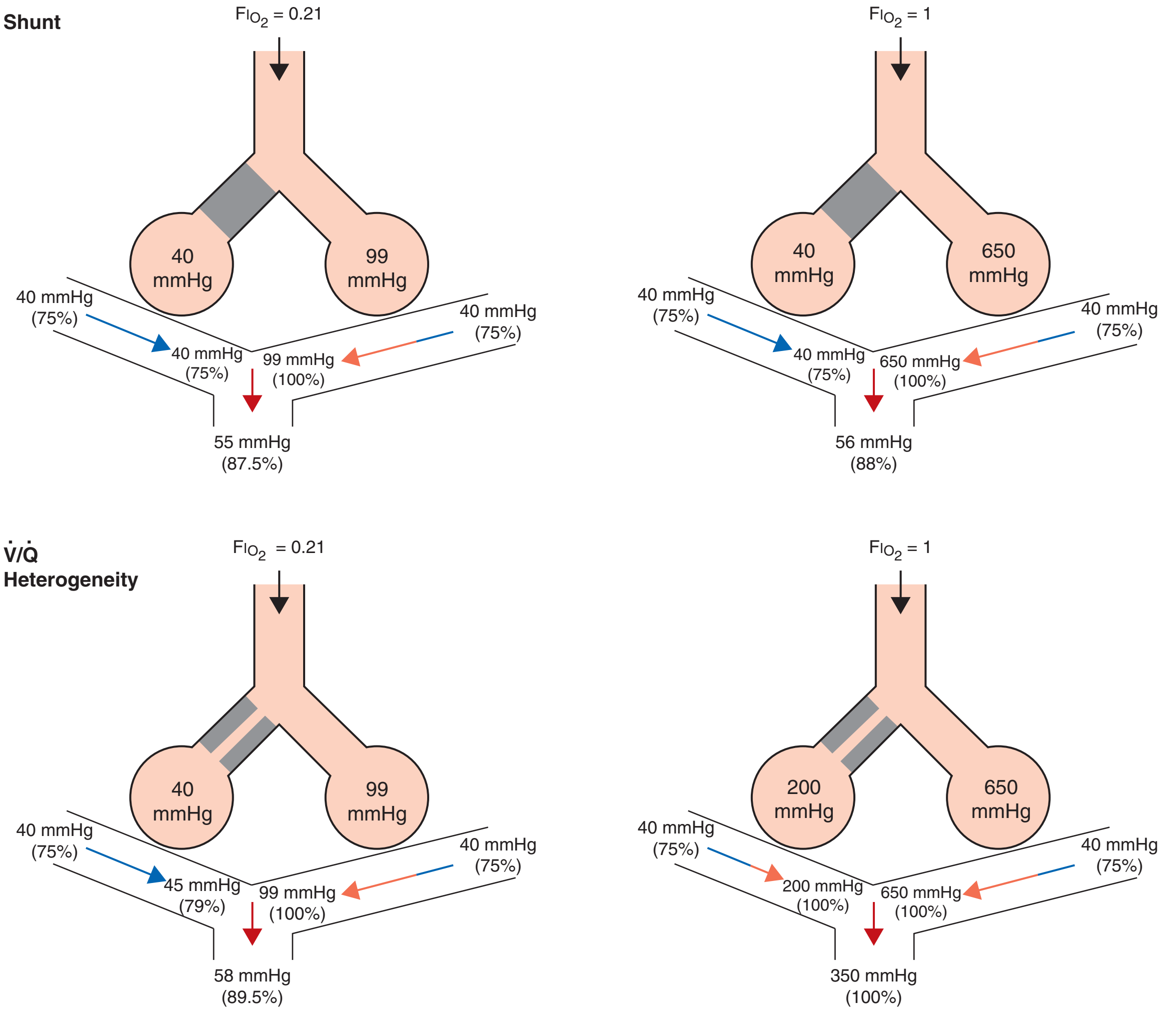

FIGURE 296-5 Influence of air versus oxygen breathing on mixed arterial oxygenation in shunt and ventilation/perfusion heterogeneity. Partial pressure of oxygen (mmHg) and oxygen saturations are shown for mixed venous blood, for end capillary blood (normal vs affected alveoli), and for mixed arterial blood. FI_{O_2} fraction of inspired oxygen; $\dot{V}/\dot{Q}$, ventilation/perfusion.

(shunt) lung regions, and, in very unusual circumstances, limitation of gas diffusion.

PATHOPHYSIOLOGY

Although many diseases injure the respiratory system, this system responds to injury in relatively few ways. For this reason, the pattern of physiologic abnormalities may or may not provide sufficient information by which to discriminate among conditions.

Figure 296-6 lists abnormalities in pulmonary function testing that are typically found in a number of common respiratory disorders and highlights the simultaneous occurrence of multiple physiologic abnormalities. The coexistence of some of these respiratory disorders results in more complex superposition of these abnormalities. Methods to measure respiratory system function clinically are described later in this chapter.

Ventilatory Restriction due to Increased Elastic Recoil—Example: Idiopathic Pulmonary Fibrosis Idiopathic pulmonary fibrosis raises lung recoil at all lung volumes, thereby lowering TLC, FRC, and RV as well as forced vital capacity (FVC). Maximal expiratory flows are also reduced from normal values but are elevated when considered in relation to lung volumes. Increased flow occurs both because the increased lung recoil drives greater maximal flow at any lung volume and because airway diameters are relatively increased due to greater radially outward traction exerted on bronchi by the stiff lung parenchyma. For the same reason, airway resistance is also normal. Destruction of the pulmonary capillaries by the fibrotic process results in a marked reduction in diffusing capacity (see below). Oxygenation is often severely reduced by persistent perfusion of alveolar units that are relatively underventilated due to fibrosis of nearby (and mechanically linked) lung due to those alveolar units already being stretched to their maximum volume with little further increase in volume with inspiration. The flow-volume loop (see below) looks like a miniature version of a normal loop but is shifted toward lower absolute lung volumes and displays maximal expiratory flows that are increased for any given volume over the normal tracing.

Ventilatory Restriction due to Chest Wall Abnormality—Example: Moderate Obesity As the size of the average American continues to increase, this pattern may become the most common of pulmonary function abnormalities. In moderate obesity, the outward recoil of the chest wall is blunted by the weight of chest wall adipose tissue and the space occupied by intraabdominal fat. In this situation, preserved inward recoil of the lung overbalances the reduced outward recoil of the chest wall, and FRC falls. Because respiratory muscle strength and lung recoil remain normal, TLC is typically unchanged (although it may fall in massive obesity) and RV is normal (but may be reduced in massive obesity). Mild hypoxemia may be present due to perfusion of alveolar units that are poorly ventilated because of airway closure in dependent portions of the lung during breathing near the reduced FRC. Flows remain normal, as does the diffusion capacity of the lung for carbon monoxide (DL_{CO}) unless obstructive sleep apnea (which often accompanies obesity) and associated chronic intermittent hypoxemia have induced pulmonary arterial hypertension, in which case DL_{CO} may be low.

Ventilatory Restriction due to Reduced Muscle Strength—Example: Myasthenia Gravis In this circumstance, FRC remains

	Restriction due to increased lung elastic recoil (pulmonary fibrosis)	Restriction due to chest wall abnormality (moderate obesity)	Restriction due to respiratory muscle weakness (myasthenia gravis)	Obstruction due to airway narrowing (acute asthma)	Obstruction due to decreased elastic recoil (severe emphysema)
TLC	60%	95%	75%	100%	130%
FRC	60%	65%	100%	104%	220%
RV	60%	100%	120%	120%	310%
FVC	60%	92%	60%	90%	60%
FEV$_1$	75%	92%	60%	35% pre-b.d. 75% post-b.d.	35% pre-b.d. 38% post-b.d.
R$_{aw}$	1.0	1.0	1.0	2.5	1.5
DL$_{CO}$	60%	95%	80%	120%	40%

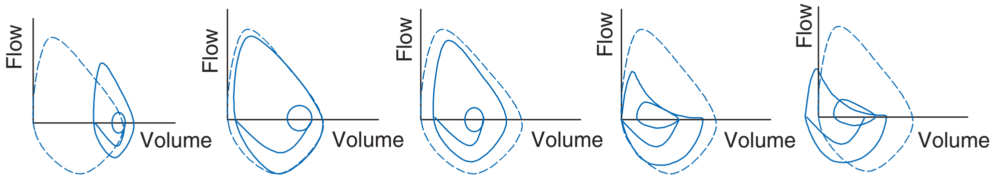

FIGURE 296-6 Common abnormalities of pulmonary function (see text). Pulmonary function values are expressed as a percentage of normal predicted values, except for R$_{aw}$, which is expressed as cmH$_2$O/L/s (normal, <2 cmH$_2$O/L/s). The figures at the bottom of each column show the typical configuration of flow-volume loops in each condition, including the flow-volume relationship during tidal breathing. b.d., bronchodilator; DL$_{CO}$ diffusion capacity of lung for carbon monoxide; FEV$_1$, forced expiratory volume in 1 s; FRC, functional residual capacity; FVC, forced vital capacity; R$_{aw}$, airways resistance; RV, residual volume; TLC, total lung capacity.

normal, as both lung recoil and passive chest wall recoil are normal. However, TLC is low and RV is elevated because respiratory muscle strength is insufficient to push the passive respiratory system fully toward either volume extreme. Caught between the low TLC and the elevated RV, FVC and forced expiratory volume in 1 s (FEV$_1$) are reduced as "innocent bystanders." As airway size and lung vasculature are unaffected, both R$_{aw}$ and DL$_{CO}$ are normal. Oxygenation is normal unless weakness becomes so severe that the patient has insufficient strength to reopen collapsed alveoli during sighs, with resulting atelectasis.

Airflow Obstruction due to Decreased Airway Diameter—Example: Acute Asthma During an episode of acute asthma, luminal narrowing due to smooth muscle constriction as well as inflammation and thickening within the small- and medium-sized bronchi raise frictional resistance and reduce airflow. "Scooping" of the flow-volume loop is caused by reduction of airflow, especially at lower lung volumes. Often, airflow obstruction can be improved through the administration of short-acting β_2-adrenergic or muscarinic agonists acutely, or by treatment with longer-acting β_2-adrenergic or muscarinic agonists, inhaled corticosteroids, and new systemically administered biologic immunotherapies chronically. TLC usually remains normal (although elevated TLC is sometimes seen in long-standing asthma), but FRC may be dynamically elevated. RV is often increased due to exaggerated airway closure at low lung volumes, and this elevation of RV reduces FVC. Because central airways are narrowed, airway resistance (R$_{aw}$) is usually elevated. Mild arterial hypoxemia is often present due to perfusion of relatively underventilated alveoli distal to obstructed airways (and is responsive to oxygen supplementation), but DL$_{CO}$ is normal or mildly elevated.

Airflow Obstruction due to Decreased Elastic Recoil—Example: Severe Emphysema Loss of lung elastic recoil in severe emphysema results in pulmonary hyperinflation, of which elevated TLC is the hallmark. FRC is more severely elevated due to both loss of lung elastic recoil and dynamic hyperinflation—the same phenomenon as auto-PEEP (auto–positive end-expiratory pressure), which is the positive end-expiratory alveolar pressure that occurs when a new breath is initiated before the lung volume is allowed to return to FRC. RV is very severely elevated because of airway closure and because exhalation toward RV may take so long that RV cannot be reached before the patient must inhale again. Both FVC and FEV$_1$ are markedly decreased, the former because of the severe elevation of RV and the latter because loss of lung elastic recoil reduces the pressure driving maximal expiratory flow and also reduces tethering open of small intrapulmonary airways. The flow-volume loop demonstrates marked scooping, with an initial transient spike of flow attributable largely to expulsion of air from collapsing central airways at the onset of forced exhalation. Otherwise, the central airways remain relatively unaffected, so R$_{aw}$ is normal in "pure" emphysema. Loss of alveolar surface and capillaries in the alveolar walls reduces DL$_{CO}$; however, because poorly ventilated emphysematous acini are also poorly perfused (due to loss of their capillaries), arterial hypoxemia usually is not seen at rest until emphysema becomes very severe. However, during exercise, PaO_2 may fall precipitously if extensive destruction of the pulmonary vasculature prevents a sufficient increase in cardiac output and mixed venous oxygen content falls substantially. Under these circumstances, any venous admixture through low $\dot{V}/\dot{Q}$ units has a particularly marked effect in lowering mixed arterial oxygen tension.

FUNCTIONAL MEASUREMENTS

Measurement of Ventilatory Function • LUNG VOLUMES Figure 296-2 demonstrates a spirometry tracing in which the volume of air entering or exiting the lung is plotted over time. In a slow vital capacity maneuver, the patient inhales from FRC, fully inflating the lungs to TLC, and then exhales slowly to RV; VC, the difference between TLC and RV, represents the maximal excursion of the respiratory system. Spirometry discloses relative volume changes during these maneuvers but cannot reveal the absolute volumes at which they occur. To determine absolute lung volumes, two approaches are commonly used: inert gas dilution and body plethysmography. In the former, a known amount of a nonabsorbable inert gas (usually helium or neon) is inhaled in a single large breath or is rebreathed from a closed circuit; the inert gas is diluted by the gas resident in the lung at the time of inhalation, and its final concentration reveals the volume of resident gas contributing to the dilution. A drawback of this method is that regions of the lung that ventilate poorly (e.g., due to airflow obstruction) may not receive much inspired inert gas and so do not contribute to its dilution. Therefore, inert gas dilution (especially in the single-breath method) often underestimates true lung volumes.

In the second approach, FRC is determined by measuring the compressibility of gas within the chest, which is proportional to the volume of gas being compressed. The patient sits in a body plethysmograph (a chamber usually made of transparent plastic to minimize claustrophobia) and, at the end of a normal tidal breath (i.e., when lung volume is at FRC), is instructed to pant against a closed shutter, thus periodically compressing air within the lung slightly. Pressure fluctuations at the mouth and volume fluctuations within the body box (equal but opposite to those in the chest) are determined, and from these measurements, the thoracic gas volume is calculated by means of Boyle's law ($P_1V_1 = P_2V_2$). Once FRC is obtained, TLC and RV are calculated by adding the value for inspiratory capacity and subtracting the value for expiratory reserve volume, respectively (both values having been obtained during spirometry) (Fig. 296-2). The most important determinants of healthy individuals' lung volumes are height, age, and sex, but there is considerable additional normal variation beyond that accounted for by these parameters. In practice, a mean "normal" value is predicted by multivariate regression equations using height, age, and sex, and the patient's value is divided by the predicted value to determine "percent predicted." For most measures of lung function, 85–115% of the predicted value can be normal; however, in health, the various lung volumes tend to scale together. For example, if one is "normal big" with a TLC 110% of the predicted value, all other lung volumes and spirometry values will also approximate 110% of their respective predicted values. This pattern is particularly helpful in evaluating airflow, as discussed below.

AIR FLOW As noted above, spirometry plays a key role in lung volume determination. Even more often, spirometry is used to measure airflow, which reflects the dynamic properties of the lung. During an FVC maneuver, the patient inhales to TLC and then exhales rapidly and forcefully to RV; this method ensures that flow limitation has been achieved, so that the precise effort made has little influence on actual flow. The total amount of air exhaled is the FVC, and the amount of air exhaled in the first second is the FEV_1; the FEV_1 is a flow rate, revealing volume change per time. Like lung volumes, an individual's maximal expiratory flows should be compared with predicted values based on height, age, and sex. While the FEV_1/FVC ratio is typically reduced in airflow obstruction, this condition can also reduce FVC by raising RV, sometimes rendering the FEV_1/FVC ratio "artifactually normal" with the erroneous implication that airflow obstruction is absent.

The relationships among volume, flow, and time during spirometry are best displayed in two plots—the spirogram (volume vs time) and the flow-volume loop (flow vs volume) (Fig. 296-4). In conditions that cause airflow obstruction, the site of obstruction is sometimes correlated with the shape of the flow-volume loop. In diseases that cause lower airway obstruction, such as asthma and emphysema, flows decrease more rapidly with declining lung volumes, leading to a characteristic scooping of the flow-volume loop. In contrast, fixed upper-airway obstruction typically leads to inspiratory and/or expiratory flow plateaus (Fig. 296-4).

RESPIRATORY MUSCLE STRENGTH To measure respiratory muscle strength, the patient is instructed to exhale or inhale with maximal effort against a closed shutter while pressure is monitored at the mouth. Pressures >±60 cmH_2O at FRC are considered adequate and make it unlikely that respiratory muscle weakness accounts for any other resting ventilatory dysfunction that is identified. A more sensitive and better tolerated approach to identify inspiratory muscle weakness is performing spirometry in the supine position. This position increases diaphragmatic work by neutralizing the assistance of gravity. FVC in normal subjects decreases approximately 3–8% from upright to supine position, and patients with diaphragmatic weakness, hemidiaphragmatic paralysis, or neuromuscular disease suffer decrements from 10 to >25%.

Measurement of Gas Exchange • **DIFFUSING CAPACITY (DL_{CO})** This test uses a small (and safe) amount of carbon monoxide (CO) to measure gas exchange across the alveolar membrane during a 10-s breath hold. CO in exhaled breath is analyzed to determine the quantity of CO crossing the alveolar membrane and combining with hemoglobin in red blood cells. This "single-breath diffusing capacity" (DL_{CO}) value increases with the surface area available for diffusion and the amount of hemoglobin within the capillaries, and it varies inversely with alveolar membrane thickness. Thus, DL_{CO} decreases in diseases that thicken or destroy alveolar membranes (e.g., pulmonary fibrosis, emphysema), curtail the pulmonary vasculature (e.g., pulmonary hypertension), or reduce alveolar capillary hemoglobin (e.g., anemia). Single-breath diffusing capacity may be elevated in asthma, polycythemia, and pulmonary hemorrhage.

Arterial Blood Gases The effectiveness of gas exchange can be assessed by measuring the partial pressures of oxygen and CO_2 in a sample of blood obtained by arterial puncture. The oxygen content of blood (CaO_2) depends on arterial saturation ($\%O_2Sat$), which is set by Pao_2, pH, and $Paco_2$ according to the oxyhemoglobin dissociation curve. CaO_2 can also be measured by oximetry (see below):

$$CaO_2\ (mL/dL) = 1.39\ (mL/dL) \times [hemoglobin]\ (g) \times \%O_2Sat + 0.003\ (mL/dL/mmHg) \times Pao_2\ (mmHg)$$

If hemoglobin saturation alone needs to be determined, this task can be accomplished noninvasively with pulse oximetry.

ACKNOWLEDGMENT

Edward T. Naureckas and Julian Solway contributed to this chapter in the 21st edition and some material from that chapter has been retained here.

FURTHER READING

Alotaibi NM et al: Mild airways obstruction: Spirometer diagnostic pitfalls. Curr Opin Pulm Med 30:121, 2024.

Bates JH: Systems physiology of the airways in health and obstructive pulmonary disease. Wiley Interdiscip Rev Sys Biol Med 8:423, 2016.

Haverkamp HC, Balmain BN: Ventilatory response to exercise by age, sex, and health status. Curr Sports Med Rep 23:79, 2024.

Hughes JM et al: Effect of lung volume on the distribution of pulmonary blood flow in man. Respir Physiol 4:58, 1968.

Macklem PT, Murphy BR: The forces applied to the lung in health and disease. Am J Med 57:371, 1974.

Pederson OF, Ingram RH: Configuration of maximal expiratory flow-volume curve: Model experiments with physiologic implications. J Appl Physiol 58:1305, 1985.

Prange HD: *Respiratory Physiology: Understanding Gas Exchange*. New York, Chapman and Hill, 1996.

Weibel ER: Morphometric estimation of pulmonary diffusion capacity, I. Model and method. Respir Physiol 11:54, 1970.

West JB: *Respiratory Physiology, The Essentials*, 9th ed. Philadelphia, Lippincott Williams & Wilkins, 2012.

Wiley Online Library: Comprehensive physiology: The respiratory system. Available from *http://www.comprehensivephysiology.com/WileyCDA/Section/id-420557.html*. Accessed August 12, 2016.

297

Diagnostic Procedures in Respiratory Disease

George R. Washko, Hilary J. Goldberg, Mājid Shafiq

Diagnostic procedures in respiratory disease encompass a wide array of invasive and noninvasive modalities. Methods for acquiring diagnostic specimens are described in this chapter, as are the various imaging modalities at hand. **Pulmonary function tests and measurements of gas exchange are described in Chap. 295.**

BEDSIDE PLEURAL PROCEDURES

THORACENTESIS

Thoracentesis, also known as pleurocentesis, refers to percutaneous aspiration of fluid from the pleural space. The right and left pleural spaces do not normally communicate with each other, and either can be directly accessed between the thoracic ribs. The current standard of care entails using point-of-care ultrasonography to mark the site of needle puncture; this reduces the risks of "dry tap" as well as complications such as pneumothorax. Beside palliation of symptoms associated with pleural effusion (most commonly dyspnea), thoracentesis may be performed for diagnostic purposes. The range of hematologic, biochemical, microbiologic, and cytologic pleural fluid studies has largely remained unchanged over the past few decades, as has the widespread adoption of Light's criteria for distinguishing exudates from transudates that were described in 1972. However, newer assays such as mesothelin-1 testing for neoplastic diseases (chiefly mesothelioma) have also become available more recently. **More details on pleural fluid testing are described in Chap. 305.**

CLOSED PLEURAL BIOPSY

Closed pleural biopsy involves percutaneous sampling of the parietal pleural lining. This procedure can be performed either "blindly" (typically with an Abrams needle) or by using imaging guidance such as computed tomography (CT) or ultrasound. Closed pleural biopsy without ultrasound guidance is highly sensitive for pleural tuberculosis, owing to the diffuse pleural involvement that is typically seen in those cases.

Image-guided closed pleural biopsy is most helpful in case of focal pleural abnormalities such as pleural nodules, which are virtually pathognomonic of malignant involvement. Limited studies have shown high diagnostic yields of around 80–90% with this modality, but patient selection is key as the diagnostic performance may be considerably lower in the absence of a specific pleural abnormality that could be visualized. Between CT and ultrasound imaging, only ultrasound is typically performed in real time during the act of obtaining the biopsy.

THORACIC SURGICAL PROCEDURES

THORACOSCOPY AND THORACOTOMY

Thoracoscopy and thoracotomy encompass a spectrum of surgical procedures that involve accessing and operating within the pleural space, either via one or more small entry ports using thoracoscopic tools or via larger incisions as in thoracotomy **(Fig. 297-1)**. Thoracoscopy varies in its scope considerably. An interventional pulmonologist typically performs a pleuroscopy (also known as *medical thoracoscopy*) and accesses the pleural space through a single port for parietal pleural biopsy or for limited therapeutic purposes such as minor lysis of adhesions, thoracoscopic pleurodesis, or indwelling pleural catheter placement. This procedure can usually be performed safely under conscious sedation. On the other hand, video-assisted thoracoscopic surgery (VATS) and robotic-assisted thoracoscopic surgery (RATS) represent more invasive procedures but with more controlled environments entailing general anesthesia with single-lung ventilation, creation of multiple entry ports, and several additional diagnostic and therapeutic possibilities including, but not limited to, lung biopsy, lymph node sampling, lobectomy, decortication, and creation of a pericardial window. Open thoracotomy uses wider incisions and more conventional surgical techniques for performing all of the above as well as additional tasks such as creation of a Clagett window for chronic bronchopleural fistula with empyema.

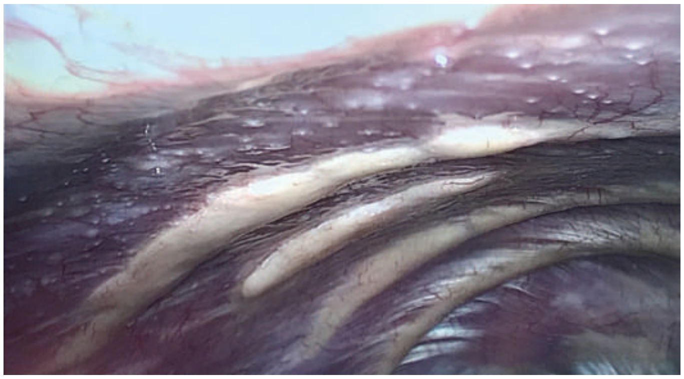

FIGURE 297-1 Thoracoscopy demonstrating numerous parietal pleural nodules in a patient with sarcoidosis-related pleural disease. Pleural biopsy revealed nonnecrotizing granulomas. *(Source: Mājid Shafiq, MD, MPH.)*

MEDIASTINOSCOPY AND MEDIASTINOTOMY

Surgical access to the mediastinum, either through a small port (mediastinoscopy) or a wider incision (mediastinotomy), enables diagnostic sampling of mediastinal structures such as mediastinal lymph nodes as part of lung cancer staging. With the advent of endoscopic needle-based techniques (see below), surgery is no longer considered the first-line option for diagnostic lymph node sampling but is recommended in cases of negative needle-based sampling where suspicion for malignant nodal involvement remains sufficiently high.

BRONCHOSCOPY

Bronchoscopy, which entails passing a tube with a lighted camera inside the lower respiratory tract, includes flexible and rigid bronchoscopy (termed after the physical properties of each bronchoscope). Flexible bronchoscopy is by far the more commonly used form and enables access to more distal parts of the respiratory tract. The rigid bronchoscope, although limited to the central airways, has the added advantage of providing a secure airway for ventilation; artificial breaths can then be administered through the scope itself as part of a closed circuit or through open jet ventilation. The rigid bronchoscope also provides a conduit for diagnostic or therapeutic instruments to be passed freely, rather than through the relatively constrained working channel of a flexible bronchoscope. When bronchoscopy is limited to diagnostic indications, the rigid bronchoscope is seldom used except on occasion as a precautionary measure for anticipated severe bleeding where having a more secure airway might be particularly advantageous (e.g., in transbronchial cryobiopsy). Different types of diagnostic bronchoscopic procedures are described below.

Bronchoalveolar Lavage Bronchoalveolar lavage (BAL) is the gold standard method for obtaining respiratory secretions for hematologic, biochemical, microbiological, and/or cytologic analyses. It avoids the risk of salivary contamination, which may be seen in a sputum specimen and is particularly useful when sputum cannot be obtained or when sampling of a specific pulmonary lobe or segment is desired. After wedging the bronchoscope in a distal airway in order to prevent fluid escape around the scope, sterile saline or distilled water is instilled through the scope's working channel (typically in one to three aliquots of ~50 mL each). Immediately thereafter, suction is applied to aspirate as much of the fluid as possible. This allows sampling of distal airways and lung parenchyma—areas not directly viewable or accessible. If there is concern for alveolar hemorrhage, serial BALs from the same site may show rising red blood cell counts and even visibly bloodier returns with subsequent lavages.

Brushing and Endobronchial Biopsy Bronchoscopic brushing is a minimally invasive sampling technique that can be used to sample the mucosal biofilm for microbiologic analyses as well as the bronchial epithelial layer for cytologic analyses. Endobronchial biopsy allows sampling of abnormal bronchial mucosa and submucosa for histopathologic analysis (as may be indicated in cases of endobronchial amyloidosis or sarcoidosis, for example). Among cigarette smokers with one or more lung nodules and a nondiagnostic bronchoscopy, bronchial brushings can be used with a commercially available classifier that estimates lung cancer probability based on a gene expression signature. Patients with intermediate pretest probability who end up with low posttest probability can more confidently opt for imaging surveillance, thus avoiding further invasive testing and related complications.

Transbronchial Biopsy Including Cryobiopsy Transbronchial biopsy involves removing a piece of alveolated lung tissue by passing a sampling tool into the alveolar space. The most commonly employed biopsy tool is flexible forceps, typically 2.0 mm or 2.8 mm in caliber. When a specific pulmonary lesion such as a lung nodule is being biopsied, various imaging and navigation tools (described below) may be used to help guide the site of forceps biopsy. When random sampling of the lung parenchyma is desired, e.g., to assess for posttransplant lung rejection, either fluoroscopic guidance or tactile feedback is commonly used to position the forceps in the subpleural lung parenchyma. Limited data point to three biopsy samples being adequate for optimizing sensitivity in case of malignant lung nodules. On the other hand, at least five distinct pieces of alveolated lung tissue are needed for formal diagnosis of acute cellular rejection among lung transplant recipients per current recommendations. An increasingly popular biopsy tool is the cryoprobe, a flexible catheter with a blunt tip that delivers liquid nitrogen or carbon dioxide over a few seconds to freeze a portion of lung parenchyma and make it adhere to the probe itself. Before the tissue can thaw and detach, the probe is pulled back (typically along with the bronchoscope itself), and a frozen piece of lung tissue is removed alongside. Cryobiopsy has a higher diagnostic yield than forceps biopsy for diffuse parenchymal illnesses such as idiopathic pulmonary fibrosis but comes with a higher risk of major bleeding and pneumothorax.

Transbronchial Needle Aspiration Transbronchial needle aspiration (TBNA) involves using a hollow-bore needle for obtaining aspirated specimens. This may be accompanied by suction or simply rely on capillary action, with data not pointing to suction impacting diagnostic sensitivity. TBNA has diagnostic sensitivity superior to that of transbronchial biopsy for malignant peripheral nodules. This makes intuitive sense given that the lesion may lie extraluminally and require traversing the airway wall, which only the needle may be able to accomplish. Furthermore, combining TBNA with conventional transbronchial biopsy appears to increase pooled diagnostic sensitivity.

Endobronchial Ultrasound-Guided Transbronchial Needle Aspiration Endobronchial ultrasound (EBUS) and EBUS-guided transbronchial needle aspiration (EBUS-TBNA) represent a major advance in diagnostic bronchoscopy over the turn of the twentieth century, largely replacing surgical methods for lymph node sampling. EBUS-TBNA involves using a specialized flexible bronchoscope that simultaneously operates a video camera and a convex ultrasound probe (which is installed at its distal end). Under real-time ultrasonographic visualization, the aspiration needle is inserted through the airway wall into the mediastinal target and the aspirate is sent for microbiologic or cytologic analyses as indicated (Fig. 297-2). Newer variants of this technique involve the use of core needles or mini-forceps, providing tissue specimens rather than aspirates that can be sent for histopathologic analysis. EBUS-TBNA has a sensitivity of ~90% for epithelial malignancies and ~70% for lymphoma (higher for detecting cases of lymphoma recurrence than for de novo lymphoma). For sarcoidosis, estimates point to a sensitivity of at least 80% (higher if combined with endobronchial and transbronchial biopsies). EBUS-TBNA has been shown to provide adequate amounts of material to provide ancillary testing in cases of malignancy, such as immunostaining or genetic testing. A related needle-based technique, also using ultrasound guidance, involves sampling mediastinal structures through the esophagus, which can be a useful adjunct to EBUS-TBNA as it may provide better access to certain mediastinal lymph node stations. The combined sensitivity of these two techniques is slightly higher compared to either one alone. Esophageal sampling can be accomplished either with the standard endoscope used by gastroenterologists for endoscopic ultrasound (EUS) or by inserting the same EBUS bronchoscope through the esophagus (EUS-B).

At many centers, EBUS-TBNA is accompanied by rapid on-site cytologic evaluation (ROSE), wherein a portion of the aspirated specimen is immediately examined by a cytotechnologist or pathologist using rapid staining. This rapid assessment, while often inadequate for

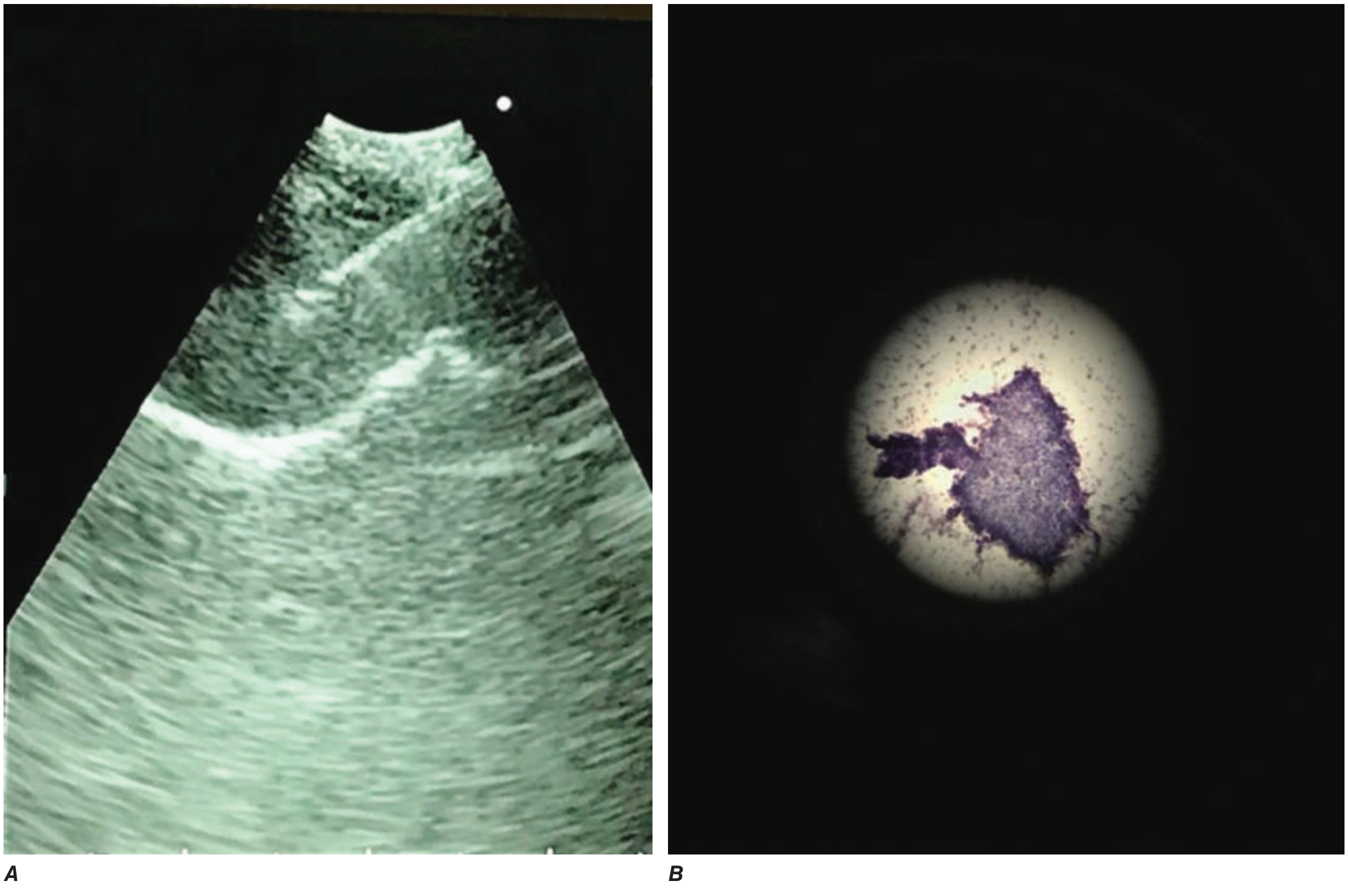

FIGURE 297-2 ***A.*** Endobronchial ultrasound-guided transbronchial needle aspiration of a mediastinal lymph node. ***B.*** Rapid on-site evaluation (ROSE) using Diff-Quik stain indicative of noncaseating granuloma. *(Source: Mājid Shafiq, MD, MPH.)*

a definitive final diagnosis, can be helpful in establishing adequacy of sampled material by providing the bronchoscopist with real-time feedback on whether additional sampling is advisable.

The optimal way to process samples obtained via EBUS-TBNA is unknown. Some centers practice the tissue coagulum clot method, in which multiple aspirates are emptied onto a single piece of filter paper to form a clot that can help with preparation of a cell block. Other centers simply use the residue from spun specimens for this purpose. There is no conclusive evidence that one technique is superior to the other, but this question has not been well studied to date.

Guided Peripheral Bronchoscopy Including Robotic Bronchoscopy Guided peripheral bronchoscopy involves the use of advanced tools to aid with successful bronchoscopic sampling of peripherally located lesions in the lung parenchyma, such as lung nodules. Prior to 2018, this entailed the use of navigation software and/or real-time imaging while utilizing conventional flexible bronchoscopes that were already commercially available. Robotic bronchoscopic platforms, first approved for commercial use by the U.S. Food and Drug Administration (FDA) in 2018, were the first to offer additionally improved bronchoscope stability and reach within the peripheral airways compared to conventional flexible bronchoscopes (Fig. 297-3). Early data on the diagnostic performance of robotic bronchoscopy, including from small-sized multicenter prospective studies, are encouraging.

Guided peripheral bronchoscopy comprises three crucial steps:

1. Navigation: Electromagnetic navigational bronchoscopy (which involves GPS-like feedback as the bronchoscope is advanced toward the target) and virtual bronchoscopy (which overlays live endoscopic images onto a CT-derived virtual bronchoscopic map) can help with successful navigation through the airways to the appropriate lobar/segmental/subsegmental airway. Shape-sensing technology, used as part of one robotic bronchoscopy platform, also aims to achieve the same purpose.

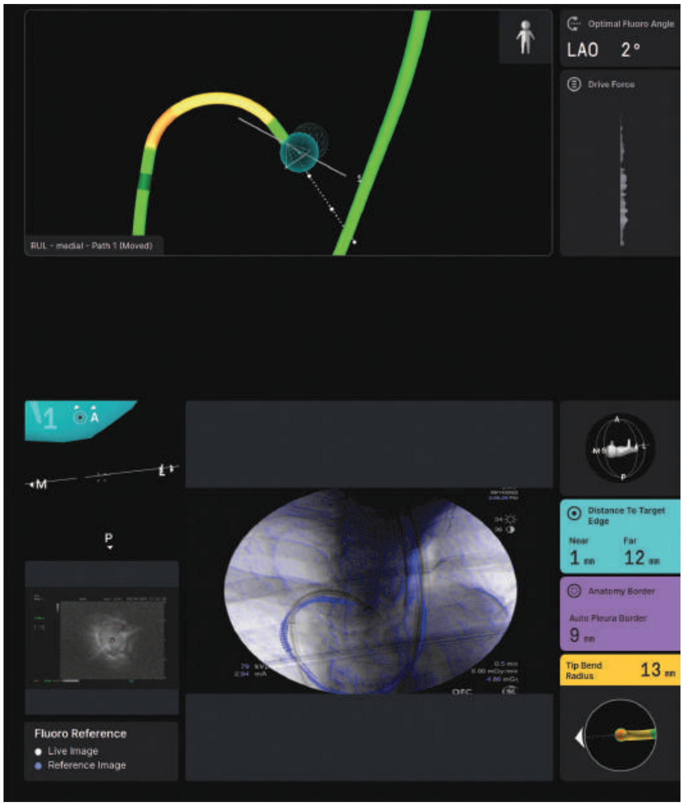

FIGURE 297-3 Use of a robotic navigational bronchoscopic platform for sampling of a 9-mm apicomedial right upper lobe nodule. The navigational software indicates that the lesion is accurately localized (*bottom right*), as does the concentric image on radial endobronchial ultrasound (*bottom left*). Biopsy showed non-small cell lung carcinoma. *(Source: Mājid Shafiq, MD, MPH.)*

2. Localization: The aforementioned technologies can also help localize a lesion, although they are limited by relying on previously acquired CT images that may or may not accurately represent precisely where the lesion is currently located in a three-dimensional space. Radial EBUS uses a thin ultrasound-tipped catheter that can be passed through a bronchoscope's working channel all the way to the lung periphery. This provides real-time images of structures beyond airway walls. A concentric image of the target, indicating a lesion with the airway going through its center, is associated with a high diagnostic yield. Alternatively, radiographic including fluoroscopic imaging can be used to recalibrate the precise target location on navigational bronchoscopic platforms, potentially improving localization as well. Cone-beam CT, which is a distilled version of CT imaging that has been used intraprocedurally in multiple other fields such as interventional radiology, can be used for confirmation of optimal tool-in-lesion (with the patient undergoing a breath hold) prior to sampling.
3. Sampling: The tools available for peripheral sampling include biopsy forceps, brushes, and aspiration needles, as described above, with TBNA having the highest diagnostic sensitivity for malignant lung nodules. Evidence for use of cryobiopsy for sampling discrete lesions in the lung periphery is currently limited but promising. Recent innovations also include real-time ultrasound-guided peripheral sampling (similar to EBUS-TBNA of mediastinal structures) and steerable sampling tools, which hold promise for more optimal sampling of the target lesion.

MEDICAL IMAGING

Imaging has revolutionized the practice of medicine. Technologies such as x-ray, CT, magnetic resonance imaging (MRI), and positron emission tomography (PET) can provide noninvasive assessments of alveolar perfusion, the metabolic activity of a lung nodule, the bronchovascular source of hemoptysis, or the earliest disease-related changes in parenchymal structure. Given the breadth of advances in respiratory system imaging and increasingly specialized applications across diseases, the following section is organized by technology. The final part of this section is dedicated to deep learning and the role it is increasingly playing in medical image interpretation.

CHEST X-RAY

The field of medical imaging can be traced back to work done by Wilhelm Roentgen in the 1890s. Roentgen noted that after connecting a cathode ray tube to a power supply, material in his lab would fluoresce even if the emission of visible light from the tube was blocked. He quickly deduced the presence of additional invisible "x-rays" and subsequently observed that their passage through solid material was attenuated in proportion to the material's density. Within weeks of its discovery, x-ray technology was being widely leveraged to guide surgical exploration and the extraction of foreign objects such as shrapnel from the battlefield. Chest x-ray (CXR) has since become the foundation of clinical practice for respiratory medicine and is a widely available technology even in resource-limited settings.

The most commonly used CXR images for respiratory medicine are the posteroanterior (PA) and lateral films in the outpatient setting and anteroposterior (AP) films for those studies obtained at the bedside. These are two-dimensional representations of three-dimensional structures, and the differing views can be used to examine superimposed structures (e.g., a parenchymal opacity in the retrocardiac space). The contours of the chest wall, the silhouette of the heart, great vessels, and mediastinum, as well as the appearance of the parenchyma and bronchovascular bundle are all used to detect and classify disease as well as monitor its progression or response to therapeutic intervention. An example of a normal PA and lateral CXR is provided in Fig. 297-4.

In this image of the normal lung, many of the smaller structures such as the lymphatics and distal airways are beyond the ability of conventional x-ray technology to resolve. Larger structures such as the pulmonary vasculature may also be indistinct because of body position

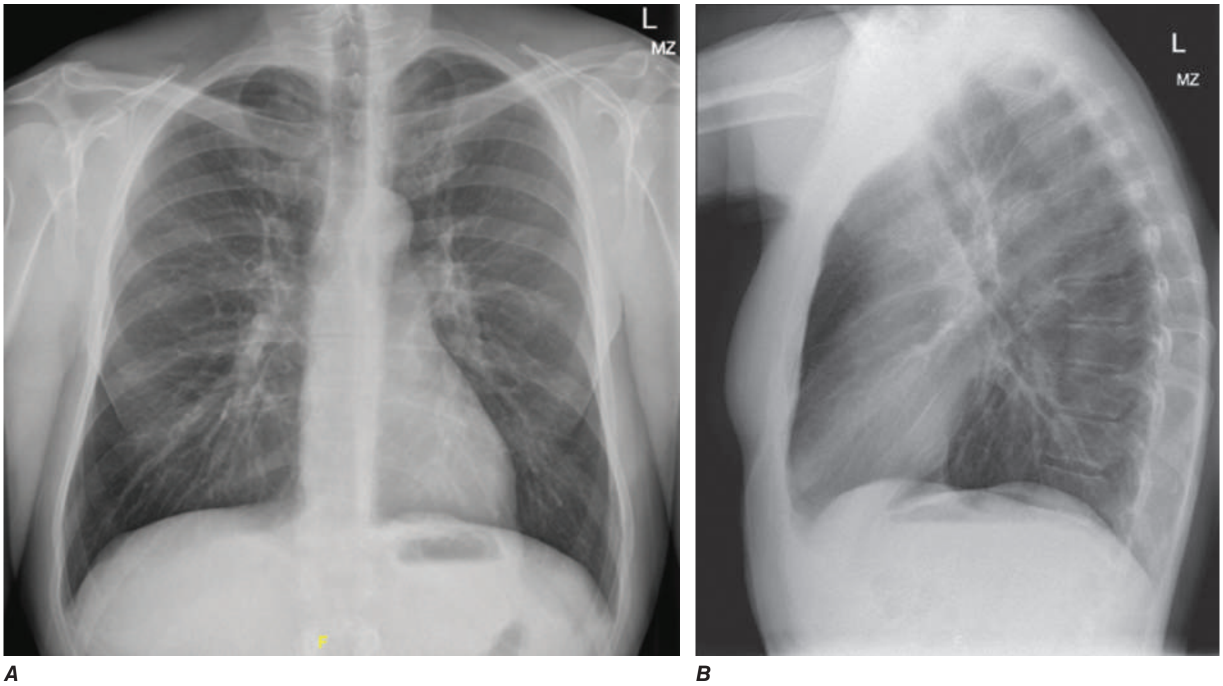

FIGURE 297-4 Posteroanterior (***A***) and lateral (***B***) chest x-ray of a normal healthy subject. *(Source: George Washko.)*

and the redistribution of blood flow to more gravitationally dependent regions. Diseases involving these structures may enhance or obscure their appearance. An example of these diseases is congestive heart failure where the lymphatics become engorged (Kerley B lines), the nondependent vasculature more prominent (cephalization), and the outer boundaries of the bronchial walls blurred (bronchial cuffing). Each of these findings must be clinically contextualized, and while a thickened interstitium may be due to hydrostatic pulmonary edema, it may also be indicative of interstitial lung disease or carcinomatosis. CXR can also be used to discriminate pulmonary and extrapulmonary disease, and because of that, it is an excellent initial diagnostic for nonspecific symptoms. An elevated hemidiaphragm, fibrosis of the mediastinum, or hyperlucency of the lung parenchyma all reflect processes that cause dyspnea, but their treatment and prognosis differ markedly.

■ COMPUTED TOMOGRAPHY

CT was introduced to clinical care in the 1980s and quickly became one of the most heavily leveraged modalities for medical imaging. While CXR provides one or two views of the thorax from which an experienced clinician must disambiguate overlying structures, CT provides spatially resolved reconstructions of all structures in the thorax. The acquisition of a CT scan involves the same basic process as an x-ray with a patient placed between a source of photons and a detector, but the image reconstruction and advanced analytics that can be applied to those images differ markedly. The passage of photons through the body is impeded in proportion to tissue density. This absorption or attenuation of photon passage is measured in Hounsfield units (HU), and clinical CT scanners are regularly calibrated to a standard scale with water having an HU of 0 and air −1000 HU. The broad range of tissue densities (reflected as attenuation values) in the thorax and the limited human ability to visually discriminate between two structures of similar densities are addressed by modifying the image display. A window width and level (the range and center of the range of HU values to display) is selected to optimize viewing structures of interest. For example, lung windows are optimized for visual inspection of the low-density lung parenchyma and all of the surrounding higher-density structures appear white, whereas the mediastinal windows are optimized to view the higher-density structures and anything of lower tissue density such as the lung parenchyma appears black. This does not change the HU values of the voxels (three-dimensional pixels) in the image, just their presentation for visual inspection.

The visual interpretation of thoracic CT is based on the appearance of the secondary pulmonary lobule. This structure is a fundamental subunit of the lung consisting of a central airway and pulmonary artery, parenchyma, and then surrounding interstitium with the lymphatics and pulmonary veins **(Fig. 297-5)**.

Processes affecting the small airways such as respiratory bronchiolitis may appear as centrilobular nodules. Parenchymal diseases such as emphysema may begin by effacing the centroid of the lobule (centrilobular emphysema [CLE]), the periphery of the lobule (paraseptal emphysema [PSE]), or diffusely across the lobule (panlobular emphysema [PLE]). Pathology of the lymphatics or interstitium (interstitial lung disease [ILD]) results in beading and/or thickening of the interlobular septa. **Examples of many of these processes are provided in Chaps. 303 and 304.**

The diagnostic information provided by the appearance of the secondary pulmonary lobule is further augmented by the distribution of these patterns of injury across the lung. Whereas CLE tends to first appear in the upper lung zones, PLE has a predilection to be basilar predominant. Interstitial thickening in the apices is more likely to be nonspecific interstitial pneumonitis (NSIP), while a basal and dependent predominant distribution of that same process is more consistent with idiopathic pulmonary fibrosis (IPF).

Finally, morphology of the central airways and vessels can be used to diagnose disease and estimate its severity. Bronchiectatic dilation of the airways may be cylindrical and predominantly in the lower lobes, as is seen in chronic obstructive pulmonary disease (COPD), or be cystic dilation in the upper lobes (cystic fibrosis), or there may be a focal nonspecific dilation of an airway due to prior infection. Pathologic dilation of the airways may also be due to disease of the surrounding parenchyma. Because of the mechanical interdependence of the bronchial tree and parenchyma, conditions that reduce lung compliance may result in traction bronchiectasis. This may be a local process or more diffuse depending on the distribution of the underlying parenchymal disease and likely provides further insight into disease severity.

The caliber of the central pulmonary arterial (PA) trunk proximal to its first bifurcation is directly related to pulmonary arterial pressure. A measure of >3 cm is suggestive of the presence of elevated pulmonary vascular pressures, and more recent studies have demonstrated that an increased ratio of the PA diameter to the diameter of the adjacent aorta (PA/A) provides a metric of disease severity and, in the case of chronic respiratory diseases such as COPD, is prognostic for both

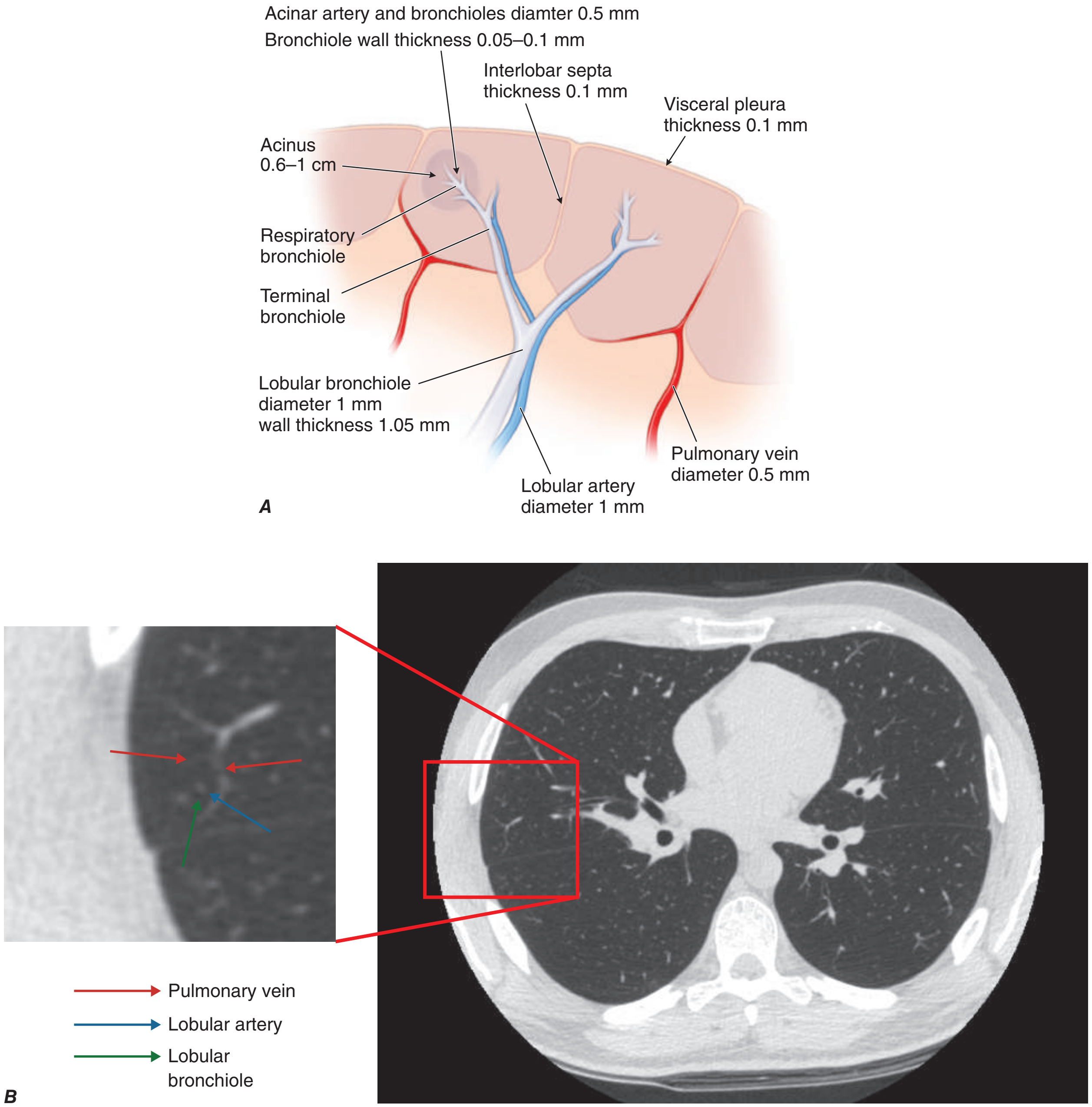
Acinar artery and bronchioles diamter 0.5 mm
Bronchiole wall thickness 0.05–0.1 mm
Interlobar septa thickness 0.1 mm
Visceral pleura thickness 0.1 mm
Acinus 0.6–1 cm
Respiratory bronchiole
Terminal bronchiole
Lobular bronchiole diameter 1 mm wall thickness 1.05 mm
Pulmonary vein diameter 0.5 mm
Lobular artery diameter 1 mm
A
Pulmonary vein
Lobular artery
Lobular bronchiole
B

FIGURE 297-5 ***A.*** Illustration of the anatomy of the secondary pulmonary lobule. ***B.*** Computed tomography image showing the visible anatomy of the secondary pulmonary lobule. *(Panel A adapted from WR Webb: Thin-section CT of the secondary pulmonary lobule: Anatomy and the Image–The 2004 Fleischner Lecture. Radiology 2006;239:322; Panel B source from Samuel Yoffe Ash, MD.)*

acute respiratory exacerbations and death. Assessment of the intraparenchymal pulmonary vasculature is typically augmented through the intravenous infusion or bolus of iodinated contrast. This bolus and subsequent image acquisition may be timed to visualize passage of contrast through the pulmonary arteries to enable detection of thromboembolic disease, which appears as dark filling voids in otherwise bright white vessels.

It must be noted that the acquisition of CXRs and thoracic CT scans involves exposing the patient to ionizing radiation. Several studies have estimated the excess numbers of cancer due to CT scanning, and extensive efforts have been made by both CT manufacturers and clinicians to reduce the radiation dose to the lowest possible amount that does not jeopardize image quality and interpretability.

■ MAGNETIC RESONANCE IMAGING

MRI is based on the behavior of protons in a magnetic field. A strong magnetic field is applied to align the protons, and then a pulse of radiofrequency current is applied to the subject. This perturbs the protons, and the speed at which they subsequently realign differs based on the properties of the tissues within the region of interest. While this technique provides exquisite imaging data for the chest wall or solid organs such as the brain or heart, the abundance of air in the lung creates an artifact that impairs direct assessment of the parenchyma. For this reason, MRI of the lung leverages intravenous contrast agents such as gadolinium and is increasingly exploring the use of inhaled agents such as hyperpolarized noble gas. These respective agents enable in vivo assessments of organ perfusion and detailed measures of the morphology of the distal airspaces. An example of noble gas–enhanced MRI is shown in **Fig. 297-6**. The inhaled agent is ^{3}He, and because it is proton rich, it can be used to examine lung ventilation visually and objectively. Regions of the lung that are poorly ventilated due to disease of the airways or distal airspaces have low concentrations of ^{3}He and appear as dark regions in an otherwise bright blue organ.

While an MRI may have a longer acquisition time than CT, and the geometry of the equipment often leads to a sense of claustrophobia, it does not involve the administration of ionizing radiation. This makes it a modality of choice in the pediatric population or clinical situations where repeated assessments are required.

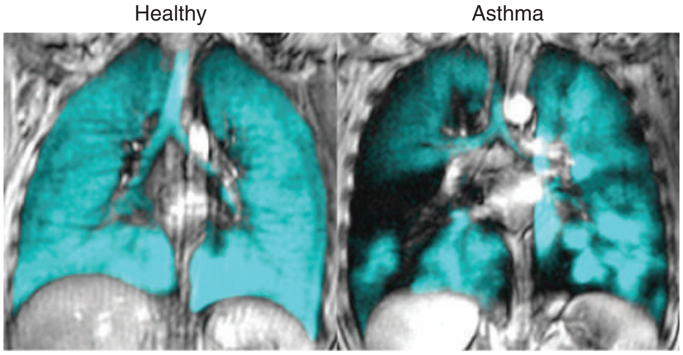

FIGURE 297-6 Noble gas magnetic resonance. Healthy control on left and asthma on right. *(Images courtesy of Grace Parraga, PhD, Department of Medical Biophysics, Department of Medicine, School of Biomedical Engineering, Robarts Research Institute, Western University, London, Ontario, Canada.)*

POSITRON EMISSION TOMOGRAPHY

PET generates an image based on the aggregation of radiolabeled tracers. The most common agent used for these purposes is [^{18}F]-fluoro-2-deoxyglucose (FDG). This radiolabeled glucose analogue is administered intravenously and is taken up by cells in direct proportion to their metabolic activity. In the clinical setting, it is most commonly used for the discrimination of benign and malignant lung nodules, as well as lung cancer staging. Given the relatively low resolution of PET, co-registration with CT is common and the aligned imaging modalities allow the reader to determine the structural source of heightened metabolic activity.

There is increasing interest in the use of PET imaging in the biomedical community. These applications are largely still confined to research, but advances in areas such as in vivo assessments of vascular biology in acute and chronic disease have been impressive.

ARTIFICIAL INTELLIGENCE/DEEP LEARNING

The final aspect to thoracic imaging that must be discussed is the growing field of artificial intelligence and deep learning applied to image analysis. Classic machine learning approaches to medical image interpretation involve the development of advanced algorithms to detect structures of interest, segment their boundaries, and then extract metrics related to size, shape, texture, and so on. The massive increase in processing capacity afforded by graphical processing units (GPUs), the increasing availability of large amounts of data, and the wide dissemination of open-source software libraries allowing developers to create powerful work environments have led to explosive growth in the utilization of deep learning for image analytics. Some of the first medical applications of deep learning were in the field of dermatology, and more recently, this advanced form of pattern recognition has been reported to excel at the discrimination of benign and malignant lung nodules in thoracic CT scan. The breadth of application of these tools continues to expand to include image navigation and feature detection, biomarker development, and direct prediction of clinical outcomes. An example of deep learning–enabled segmentation of the heart and pulmonary vasculature from non–contrast-enhanced non–cardiac-gated CT scan is shown in Fig. 297-7.

TRANSTHORACIC NEEDLE ASPIRATION

Radiologically guided needle biopsy has served as a long-standing mechanism for evaluation of parenchymal lung lesions, both malignant and infectious. In the setting of published guidelines recommending low-dose screening CT scan for lung malignancy in high-risk patients, and with evolving guidelines for monitoring and assessment of incidental lung lesions arising in this setting, radiologically guided sampling of lung lesions has become an increasingly important mechanism to address parenchymal lung abnormalities concerning for cancer. Moreover, as novel immune modulators and biologic agents are increasingly utilized for the management of systemic disease and transplantation, effective interventions are becoming progressively more important in assessing for potential pulmonary infections arising as complications of immune suppression. Transthoracic needle aspiration (TTNA) remains one important arm in the assessment of these pulmonary complications.

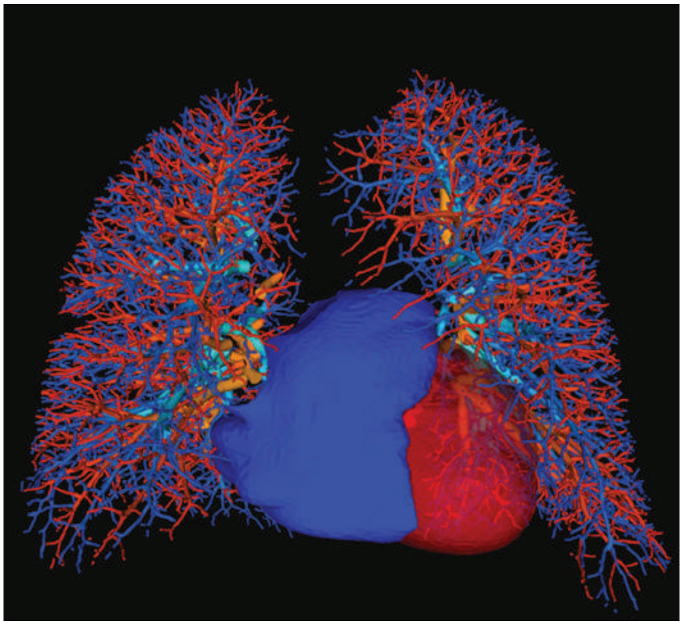

FIGURE 297-7 Arterial/venous segmentation of the pulmonary vasculature (*blue:* arteries; *red:* veins) and epicardial surface of the right (*blue*) and left ventricles (*red*). *(Image courtesy of Raul San Jose Estepar, PhD, Applied Chest Imaging Laboratory, Department of Radiology, Brigham and Women's Hospital, Boston, MA.)*

TTNA can be accomplished with a variety of complementary imaging mechanisms, including under fluoroscopic, CT, ultrasound, or MRI guidance. Overall adequacy of sampling as well as adequacy for epidermal growth factor receptor (EGFR) analysis of >90% and for cytologic analysis of >80%. CT is currently the most common imaging modality used to assess parenchymal lung lesions, with sensitivity and specificity reported to be >90%. Sensitivity of CT-guided TTNA is increased in more peripheral lesions. Transthoracic ultrasound has the advantages of a low complication rate in the setting of fine-needle aspiration (FNA) and portability, allowing for more logistical simplicity in the setting of lung lesion assessment. In a prospective study of ultrasound-guided percutaneous FNA compared with CT-guided FNA, diagnostic rates were comparable between the two groups, with shorter procedure time associated with ultrasound guidance, numerical suggestion of decreased complication rate using ultrasound guidance, and lower costs associated with ultrasound guidance. Use of elastography to better characterize lung lesions has also been proposed in the context of ultrasound, although additional diagnostic yield has not yet been proven. Color Doppler ultrasonographic imaging has been demonstrated to have a high sensitivity and specificity and a low complication rate in another study. Electromagnetic guidance, unlike CT imaging, can be used in combination with endobronchial ultrasound and/or navigational bronchoscopy in the operative setting, theoretically allowing for a multimodal approach that could increase diagnostic yield and allow for a combined staging procedure. Electromagnetic TTNA alone has demonstrated an 83% diagnostic yield in a pilot study, with an increase to 87% when combined with navigational bronchoscopy. Conflicting data are available regarding the diagnostic superiority of TTNA compared with alternative biopsy modalities such as endobronchial ultrasound for diagnosis of lung lesions, and results may depend on center experience.

Transthoracic sampling can be obtained using FNA or core needle biopsy. In one retrospective study, FNA was found to have an inferior diagnostic rate, compared with core needle sampling, as well as lower specificity. In this study, a method involving two FNA passes was compared to core needle sampling with six cores obtained from a single pass. No significant differences in complication rates were noted. In another retrospective study, in which procedure was determined by operator preferences, core needle aspirate samples were more likely to provide sufficient material for molecular testing than FNA. A systematic review of these techniques concluded that insufficient evidence was available to support a difference between FNA and core needle biopsy in diagnostic efficiency, though core needle biopsy may be more specific in diagnosing benign lung lesions. Given the negative predictive value estimate of 70%, negative results from TTNA are less reliable than positive results and should not be considered definitive to eliminate the concern for malignancy. Further assessment is needed to directly compare imaging modalities for TTNA guidance and to compare TTNA with other diagnostic modalities to determine the optimal choice of procedure in particular settings. Choice of procedure should be considered in the context of the size and location of the lesion, the experience of the center and operator, and patient-specific factors.

In regard to the safety of TTNA, in a retrospective study from 2015, the presence of mild to moderate pulmonary hypertension in patients did not increase complication rates in the setting of TTNA. The complication rates noted in this report were substantial, however, with hemorrhage occurring in one-third to one-quarter of patients, and pneumothorax in 17–28%. The majority of pneumothoraces did not require chest tube placement. Other complications included hemoptysis and hemothorax, though these were uncommon. These complication rates are consistent with those reported in other studies. In a meta-analysis of complication rates of CT-guided TTNA, complication rates were higher with core needle aspirates than FNA (38.8% [95% confidence interval (CI) 34.3–43.5%] vs 24% [95% CI 18.2–30.8%]). The majority of these complications were minor. Risk factors for complications with FNA included smaller nodule diameter, larger needle diameter, and increased traversed lung parenchyma. No clear risk factors were noted for complications after core needle biopsies in this publication. More generally, the risks of TTNA increase for more centrally located lesions and those residing in close proximity to intrathoracic vasculature. In a study of patient claims in the Medicare and a subset of the commercial population between 2016 and 2020, the use of TTNA decreased in both groups, with the use of endobronchial ultrasound guidance for sampling increasing in the Medicare population. In this study, TTNA presented a higher risk for pneumothorax than the use of alternate modalities.

Despite the outstanding questions regarding the context and optimal approach for TTNA, this modality has been shown to be effective in cancer diagnosis in the thorax. Adenocarcinoma has become the most prevalent parenchymal lung malignancy in reported studies and also the most common malignant diagnosis found on TTNA of the lung. TTNA can also be effective in diagnosing less common tumors of the lung, both malignant and benign, including squamous and small cell carcinomas, lymphomas, and others, as well as in assessing tumors of the mediastinum. The diagnostic utility of TTNA is consistent across solid, subsolid, and partially calcified lung nodules. Immunocytochemistry markers can be utilized in TTNA samples to assist with diagnosis, prognosis, and prediction of response to therapy, and samples should be preserved whenever possible to allow for these studies, if needed. RNA extraction has also proven feasible in the setting of a single FNA sample, which could be instrumental in gene expression profiling, though this has thus far only been successfully accomplished in a research context.

The utility of TTNA in diagnosing pulmonary infections is variable in published literature. Some publications have reported that TTNA establishes a diagnosis of infection in 60–70% of cases, with a particularly high yield in the setting of *Aspergillus* infections. TTNA has also been shown to be particularly effective in the diagnosis of pulmonary tuberculosis, though a wide variety of infections have been diagnosed using this method. The presence of necrosis in lung lesions makes establishing an infectious diagnosis more likely using TTNA. Numerous staining techniques are available to assist with infectious diagnoses, and immunohistochemistry can also aid in the diagnosis of infection. Cytology should be correlated with histopathology and culture results, when available. Metagenomics using next-generation sequencing for detection of infection is evolving but requires further study. TTNA has also been useful in identifying granulomatous inflammation, which can provide supportive evidence of a granulomatous parenchymal lung disease in the appropriate clinical setting.

In summary, TTNA is an important element of diagnostic algorithms in the setting of lung nodules and masses, particularly when concern for malignancy is not high enough to warrant immediate excision, when the patient is not a surgical candidate, or the lesion or disease is not amenable to surgical resection. Further study is needed, however, to better understand the role of TTNA and other diagnostic modalities in the evaluation of parenchymal lung lesions.

MISCELLANEOUS TESTING

SPUTUM TESTING

Sputum microscopy and culture are commonly utilized to diagnose respiratory tract infections and identify the causative organisms. In patients with productive cough, the sampling is simple and noninvasive; however, it is subject to patient technique and the potential for oropharyngeal and/or upper respiratory tract contamination. In those who are not expectorating, sputum induction can be considered using provocative nebulization with saline. This technique has been demonstrated to be generally safe and well tolerated even in patients with airflow limitation. Numerous studies have attempted to define criteria for reliability and reproducibility of sputum samples. The majority include quantification of number of epithelial cells and white blood cells per low-power field, and many assess the ratio of the two for adequacy of sampling. None has been confirmed as superior in establishing the reliability of sampling to reflect lower respiratory tract growth. The quality of the sputum sample directly impacts the diagnostic reliability in the setting of bacterial pneumonia. Growth of *Mycobacterium tuberculosis, Legionella,* or pneumocystis should raise concern for infection, even in the setting of a poor sample. In a prospective study of the use of a multiplex polymerase chain reaction (PCR) assay in conjunction with sputum sampling, good-quality samples had a higher proportion of bacterial detections than low-quality samples but equivalent frequency of samples with bacterial growth in patients who received treatment. In this study, 40% of bacterial detections would have been missed if only high-quality samples were analyzed. The authors conclude that all samples submitted for PCR-based testing should be analyzed, regardless of sputum sample quality. In systematic analyses and meta-analyses, multiplex PCR assays have been shown to decrease the time to diagnosis and length of stay in the setting of viral infections and in patients with COVID-19 and bacterial co-infection. Endotracheal aspirates have not been demonstrated to be clearly superior to expectorated sputum in terms of diagnostic reliability, but such sampling may be required if spontaneous coughing is nonproductive and induced sputum is not feasible or successful. The use of multiplex fluorescence PCR may allow for assessment of multiple pathogens with a reduced time to pathogen identification.

As in the context of infection, sputum cytologic analysis has been utilized to assist in the diagnosis of malignancy, mainly because it can be obtained noninvasively. While sputum cytology demonstrating malignant cells is highly specific for a diagnosis of lung malignancy, its sensitivity has been reported at <40%. A systematic review of screening methods demonstrated no added benefit from sputum cytology when combined with CXR to screen for lung cancer. Advanced molecular techniques such as PCR, DNA methylation markers, micro-RNA assessment, and tumor-related protein analysis have been proposed in sputum assessment for diagnostic purposes and risk stratification. At present, however, sputum cytology is recommended only when more invasive techniques cannot be pursued, such as in patients with prohibitive comorbidities or in resource-limited settings.

EXHALED BREATH CONDENSATE

Exhaled breath condensate includes gaseous, liquid, and water-soluble components, with numerous biomarker types and collection system varieties developed over time. Validation standards for many components are still being determined. Exhaled nitric oxide is the most highly validated of the biomarkers identified in exhaled breath condensate. The fraction of exhaled nitric oxide (FeNO) has been demonstrated in higher concentrations in exhaled breath condensate of patients with asthma than in healthy individuals, has been shown in a systematic review and meta-analysis to help differentiate asthma/chronic obstructive pulmonary disease overlap syndrome, and has been shown in some studies to correlate with the presence of eosinophils in the sputum and blood and with response to inhaled corticosteroids, although data are conflicting. For example, in a systematic review and meta-analysis, FeNO elevation increased the odds of having asthma in both children above the age of 5 years and adults. In another systematic review of FeNO utilization in the management of adults with asthma, the assessment was helpful in the management of severe exacerbations but had no significant impact on overall exacerbations or inhaled corticosteroid use. Moreover, evidence suggests that tailoring of asthma therapy based on sputum eosinophil levels was effective in decreasing asthma exacerbations, but tailoring of therapy based on FeNO was not beneficial in improving outcomes, and insufficient evidence was observed to advocate the use of either sputum analysis or FeNO in clinical practice. FeNO has also been shown to be influenced by ethnicity, and appropriate reference standards for different ethnic groups have yet to be established. While FeNO has been proposed as a potential clinical guide to management, its use has not been incorporated into all guideline recommendations, and it has not been formally approved for clinical use.

SWEAT TESTING

Assessment of chloride concentration in sweat using pilocarpine iontophoresis, or sweat testing, remains a key element in the diagnostic framework of cystic fibrosis (CF). This method utilizes pilocarpine to stimulate sweat production. As patients with CF suffer from alterations to the sodium chloride ion channel, measurement of electrolytes in their secretions such as sweat reveals elevated chloride concentrations, among other abnormalities. This testing has been considered the gold standard in the diagnosis of CF due to its functional nature, its relative noninvasiveness, the establishment of validated standards for its performance, and its ability to discriminate between healthy individuals and those with CF at a chloride concentration of ≥60 mmol/L. The likelihood of a diagnosis of CF at a concentration of <40 mmol/L has been observed to be low, and an indeterminate range was defined as 40–59 mmol/L, which could be consistent with the disease if genetic and clinical manifestations were supportive.

While functional testing such as sweat chloride testing remains an essential component of diagnostic algorithms in CF, the evolution of genetic analysis has led to identification of an extensive array of genetic mutations associated with varied phenotypic impacts in this disease. In this context, the indeterminate range of chloride concentrations of 40–59 mmol/L on sweat test analysis was found to inadequately identify milder or more heterogenous forms of the disease associated with newly identified genetic mutations. As a result, the Cystic Fibrosis Foundation provided updated guidance for the interpretation of sweat test results, with a decreased lower threshold to define an intermediate range of chloride concentration (changed from 40–50 mmol/L to 30–59 mmol/L), which could be consistent with the diagnosis of CF in the appropriate genetic and clinical context. In a subsequent analysis, utilization of the new guidance was found to enhance the probability of identifying patients with CF without increasing the false-positive diagnosis rate in the population. Sweat testing is a critical component of the CF diagnostic algorithm but should be interpreted in the context of clinical manifestations of disease and correlated with genetic testing in those suspected of the diagnosis.

ALLERGY TESTING

Allergy testing is often considered in the assessment of environmental exposures, including seasonal allergens, food allergens, and drug allergens. In the case of drug allergens in particular, drug reactions are often reported based on remote history and are often unconfirmed. The hesitancy to re-expose patients with an unconfirmed drug allergy can lead to limited options for treatment, to delay in treatment, and to utilization of treatments with more extended spectrum, potentially influencing the resistance patterns of these agents. Drug reactions can be mediated by IgE (immediate-type reactions, type I), IgG or IgM (type II), immune complex reactions (type III), and delayed-type hypersensitivity reactions mediated by cellular immune mechanisms (type IV).

Skin testing, including patch testing and/or delayed intradermal testing, is available to test exposure to particular allergens and determine reactivity. These tests have been shown to aid in clinical phenotyping of type I reactions and potentially in type IV reactions, though their role in type IV assessment remains more controversial. In the context of suspected type I reactions, patch testing is more cost effective and may be as effective as intradermal testing in identifying potential causative agents. The negative predictive value of intradermal skin testing in assessing for IgE-mediated drug allergies is high; however, the high sensitivity of this testing limits its specificity, and results must be interpreted in the context of the pretest probability and the clinical experience of the patient. Skin tests have also been demonstrated to assist in identifying the causative agent in type IV reactions and to assess cross-reactivity between structurally related drugs. Intradermal testing may be more sensitive than patch testing to assess for type IV drug reactions. Though some debate continues regarding a mandatory role for skin testing in the assessment of potential drug allergies, drug provocation testing or rechallenge is generally regarded as safe in low-risk individuals with history of urticaria or immediate rash, whereas skin testing has been proposed as a preliminary assessment in higher-risk individuals with a history of two or more reactions, angioedema, or anaphylaxis, prior to consideration of drug provocation testing.

FURTHER READING

Callister ME et al: British Thoracic Society guidelines for the investigation and management of pulmonary nodules. Thorax 70:ii1, 2015.

Deng CJ et al: Clinical updates of approaches for biopsy of pulmonary lesions based on systematic review. BMC Pulm Med 18:146, 2018.

Dragonieri S: Methodological aspects of induced sputum. Adv Resp Med 5:397, 2023.

Shepherd W: Image-guided bronchoscopy for biopsy of peripheral pulmonary lesions. In: UpToDate. Post TW (ed). UpToDate, Waltham, MA, 2023.

Silvestri G et al: Methods for staging non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence based clinical practice guidelines. Chest 143:e211s, 2013.

Thunnissen FBJM: Sputum examination for early detection of lung cancer. J Clin Pathol 56:805, 2003.

Webb WR: Thin-section CT of the secondary pulmonary lobule: Anatomy and the image. The 2004 Fleischner Lecture. Radiology 239:322, 2006.

Section 2 Diseases of the Respiratory System

298

Asthma

Elliot Israel

Asthma is a disease characterized by episodic airway obstruction and airway hyperresponsiveness usually accompanied by airway inflammation. In most cases, the airway obstruction is reversible, but in a subset of asthmatics, a component of the obstruction may become irreversible. In a large proportion of patients, the airway inflammation is eosinophilic, but some patients may present with differing types of airway inflammation, and in some cases, there is no obvious evidence of airway inflammation.

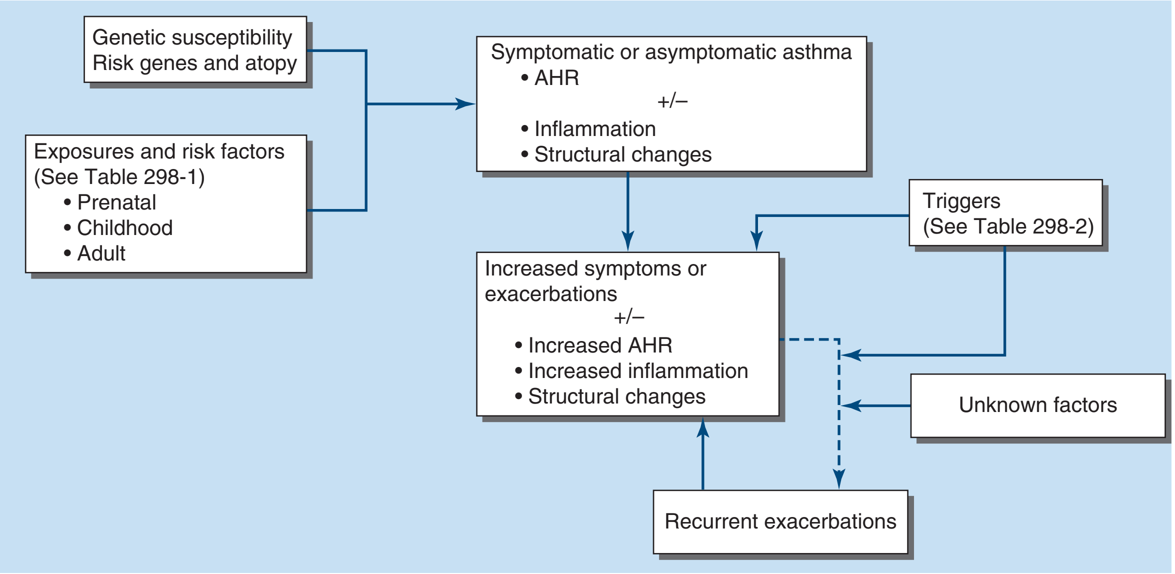

FIGURE 298-1 Asthma development pathway. Illustration of how genetic susceptibility and development and exposure during the life span interact to produce a disease that can vary in intensity and chronicity. Disease expression is characterized by airway hyperresponsiveness with varying degrees of airway inflammation and airway structural changes accompanied by varying degrees of symptoms that can be influenced by exposure to triggers that can cause acute deterioration as well as chronic symptoms. AHR, airway hyperresponsiveness.

MANIFESTATIONS

Asthma most frequently presents as episodic shortness of breath, wheezing, and cough, which can occur in relation to triggers but may also occur spontaneously. These symptoms can occur in combination or separately. Other symptoms can include chest tightness and/or mucus production. These symptoms can resolve spontaneously or with therapy. In some patients, wheezing and/or dyspnea can be persistent. Episodes of acute bronchospasm, known as exacerbations, may be severe enough to require emergency medical care or hospitalization and may result in death.

EPIDEMIOLOGY

Asthma is the most common chronic disease associated with significant morbidity and mortality, with ~262 million people affected globally. Cross-sectional studies suggest that 7.9% of the population in the United States is asthmatic as compared to ~4.3% prevalence worldwide. Prevalence continues to increase (starting at 7.3% in 2001 in the United States) and is associated with transition from rural to urban living. Asthma is more prevalent among children (8.4%) than adults (7.7%). In children, the prevalence is greatest among boys (2:1 male-to-female ratio), with a trend toward greater prevalence in women in adulthood. In some patients, asthma resolves as they enter adulthood only to "recur" later in life.

In the United States in 2016, prior to the effects of the COVID-19 pandemic, 1.8 million people visited an emergency department for asthma, and 189,000 were hospitalized. The total economic cost in the United States in 2013 was estimated at $82 billion. In the United States, asthma is more prevalent in blacks than Caucasians, and black race is associated with greater case morbidity. The ethnicity with the greatest prevalence in the United States is the Puerto Rican population.

Asthma mortality increased worldwide in the 1960s, apparently related to overuse of inhaled β_2-agonists. Reduction in mortality since then has been attributed to increased use of inhaled corticosteroids. Asthma mortality declined globally from 0.44 per 100,000 people in 1993 to 0.19 in 2006, but further reduction in mortality has not occurred since that time.

THE PATHWAY TO THE DEVELOPMENT OF ASTHMA

The pathway to development of asthma can be varied. As illustrated in Fig. 298-1, there is an interplay between genetic susceptibility (see below) and environmental exposure and endogenous developmental factors (e.g., aging and menopause [see "Etiologic Mechanisms, Risk Factors, Triggers, and Complicating Comorbidities" and Table 298-1]) that can lead to the development of asthma. Continued or additional exposures and triggers (Table 298-2) can affect the progression of disease and the degree of impairment.

PATHOPHYSIOLOGY

■ MECHANISMS LEADING TO ACUTE AND CHRONIC AIRWAY OBSTRUCTION

The pathobiologic processes in the airways that lead to episodic and chronic airway obstruction of asthma are discussed below. Their pathologic correlates are highlighted in Fig. 298-2, illustrating the pathologic changes that can occur in asthmatic airways. These processes can occur individually or simultaneously. There can be temporal variation of these processes in an individual based on exogenous factors, discussed later in this chapter, as well as the aging process itself. These processes can involve the entire airway (but not the parenchyma), but there can be significant spatial heterogeneity, as has now been demonstrated using hyperpolarized gas ventilation studies and high-resolution computed tomography (CT) of the thorax.

Airway Hyperresponsiveness Airway hyperresponsiveness is a hallmark of asthma. It is defined as an acute narrowing response of the airways in reaction to agents that do not elicit airway responses in nonaffected individuals or an excess narrowing response to inhaled agents as compared to that which would occur in nonaffected individuals. A component of the hyperresponsiveness occurs at the level of the airway smooth muscle itself as demonstrated by hyperresponsiveness to direct smooth-muscle–acting agents such as histamine or methacholine. In many patients, the apparent hyperresponsiveness is due

TABLE 298-1 Exposures and Risk Factors Related to the Development of Asthma

1. Allergen exposure in those with a predisposition to atopy
2. Occupational exposure
3. Air pollution
4. Infections (viral and *Mycoplasma*)
5. Tobacco
6. Obesity
7. Diet
8. Fungi in allergic airway mycoses
9. Acute irritants and reactive airway dysfunction syndrome (RADS)
10. High-intensity exercise in elite athletes

TABLE 298-2 Triggers of Airway Narrowing

1. Allergens
2. Irritants
3. Viral infections
4. Exercise and cold, dry air
5. Air pollution
6. Drugs
7. Occupational exposures
8. Hormonal changes
9. Pregnancy

to indirect activation of airway narrowing mechanisms as a result of stimulation of inflammatory cells (which release direct bronchoconstrictors and mediators that cause airway edema and/or mucus secretion) and/or stimulation of sensory nerves that can act on the smooth muscle or inflammatory cells. Agents and physical stimuli that elicit such responses are discussed later.

The apparent increased responsiveness of the airways in asthma may also have a structural etiology. In asthma, airway wall thickness is associated with disease severity and duration. This thickening, which may result from a combination of smooth-muscle hypertrophy and hyperplasia, subepithelial collagen deposition, airway edema, and mucosal inflammation, can result in a tendency for the airway to narrow disproportionately in response to stimuli that elicit increased airway muscle tension. A major therapeutic objective in asthma is to decrease the degree of airway hyperresponsiveness.

Inflammatory Cells While airway inflammation can be precipitated by acute exposure to inhalants, most asthmatics have evidence of chronic inflammation in the airways. Most commonly, this inflammation is eosinophilic in nature. In some patients, neutrophilic inflammation may be predominant, especially in those with more severe

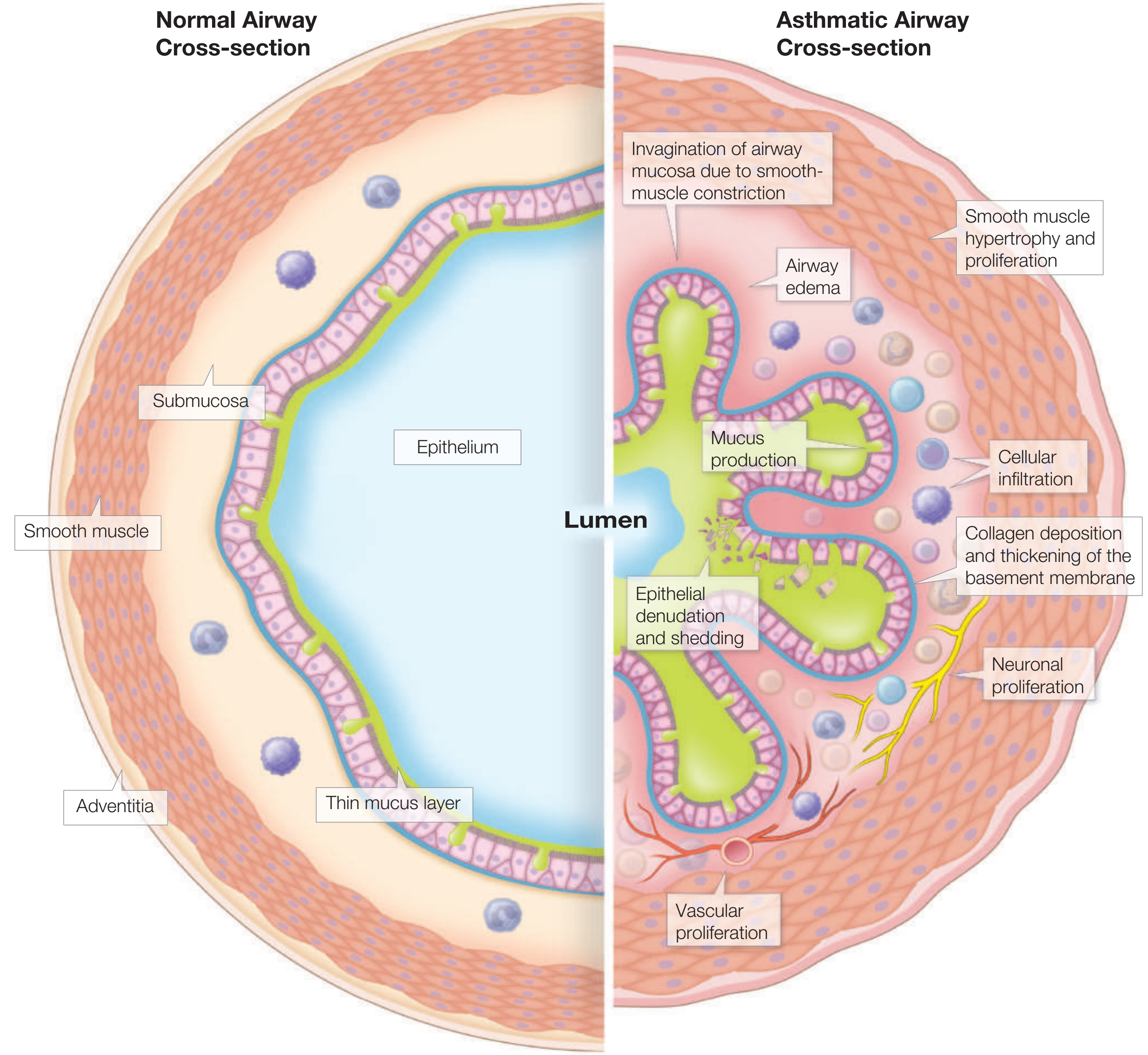
Normal Airway Cross-section
Asthmatic Airway Cross-section
Invagination of airway mucosa due to smooth-muscle constriction
Smooth muscle hypertrophy and proliferation
Airway edema
Submucosa
Epithelium
Mucus production
Cellular infiltration
Smooth muscle
Lumen
Collagen deposition and thickening of the basement membrane
Epithelial denudation and shedding
Neuronal proliferation
Adventitia
Thin mucus layer
Vascular proliferation

FIGURE 298-2 Pathologic changes that can be seen in asthmatic airways. Illustrated is a cross-sectional lumen of a bronchus. The left-hand side represents the normal airway, and the right represents an asthmatic airway highlighting the pathologic changes that can be seen in asthma. The asthmatic airway lumen is reduced by smooth-muscle contraction and hypertrophy, mucus in the airway lumen, and thickening of the submucosa due to edema and cellular infiltration. In addition, the ability of the lumen to increase in size with smooth-muscle relaxation may be impaired by deposition of collagen. The epithelium is disrupted, and there is evidence of vascular and neuronal proliferation. All these changes may not be present in one individual, and certain patients may have normal-appearing airways.

asthma. Patients with both eosinophilic and neutrophilic inflammation may present with the most severe phenotype. Mast cells are also more frequent. Many inflammatory cells are present in an activated state, as will be discussed in the section on inflammation.

Airway Smooth Muscle Airway smooth muscle can contribute to asthma in three ways. First, it can be hyperresponsive to stimuli, as noted above. Second, hypertrophy and hyperplasia can lead to airway wall thickening with consequences for hyperresponsiveness, as noted above. Lastly, airway smooth-muscle cells can produce chemokines and cytokines that promote airway inflammation and promote the survival of inflammatory cells, particularly mast cells.

Subepithelial Collagen Deposition and Matrix Deposition Thickening of the subepithelial basement membrane occurs as a result of deposition of repair-type collagens and tenascin, periostin, fibronectin, and osteopontin primarily from myofibroblasts under the epithelium. The deposition of collagen and matrix stiffens the airway and can result in exaggerated responses to increased circumferential tension exerted by the smooth muscle. Such deposition can also narrow the airway lumen and decrease its ability to relax and thus can contribute to chronic airway obstruction.

Airway Epithelium Airway epithelium disruption takes the form of separation of columnar cells from the basal cells. The damaged epithelium is hypothesized to form a trophic unit with the underlying mesenchyme. This unit elaborates multiple growth factors thought to contribute to airway remodeling. The airway epithelium is also a source of multiple cytokines, such as alarmins, and mediators that have been shown to promote a cascade of airway inflammation.

Vascular Proliferation In a subset of asthmatics, there is a significant degree of angiogenesis thought to be secondary to elaboration of angiogenic factors in the context of airway inflammation. Vascular leakage from postcapillary venules in response to inflammatory mediators can also contribute to the acute and chronic edema of the airways.

Airway Edema Submucosal edema can be present as an acute response in asthma and as a chronic contributor to airway wall thickening.

Epithelial Goblet Cell Metaplasia and Mucus Hypersecretion Chronic inflammation can result in the proliferation of mucus cells. Increased mucus production can reduce the effective airway luminal area. Mucus plugs can obstruct medium-size airways and can extend into the small airways. These mucus plugs can result in persistent airway obstruction.

Neuronal Proliferation Neurotrophins, which can lead to neuronal proliferation, are elaborated by smooth-muscle cells, epithelial cells, and inflammatory cells. Neuronal inputs can regulate smooth-muscle tone and mucus production, which may mediate acute bronchospasm and potentially chronically increased airway tone.

AIRWAY INFLAMMATION (TYPE 2 AND NON–TYPE 2 INFLAMMATION)

Most asthma is accompanied by airway inflammation. In the past, asthma had been divided into *atopic* and *nonatopic* (or *intrinsic*) asthma. The former was identified as relating to allergen sensitivity and exposure, with production of IgE, and occurring more commonly in children. The latter was identified as occurring in individuals with later onset asthma, with or without allergies, but frequently with eosinophilia. This paradigm is being superseded by a nosology that favors consideration of whether asthma is associated with type 2 or non–type 2 inflammation. This approach to immunologic classification is driven by a developing understanding of the underlying immune processes and by the development of therapeutic approaches that target type 2 inflammation (see later sections on asthma therapy).

Type 2 Inflammation Type 2 inflammation is an immune response involving the innate and adaptive arms of the immune system to promote barrier immunity on mucosal surfaces. It is called type 2 because it is associated with the type 2 subset of CD4+ T-helper cells, which produce the cytokines interleukin (IL) 4, IL-5, and IL-13. As shown in Fig. 298-3, these cytokines can have pleiotropic effects. IL-4 induces B-cell isotype switching to production of IgE. IgE, through its binding to basophils and mast cells, results in environmental sensitivity to allergens as a result of cross-linking of IgE on the surface of these mast cells and basophils. The products released from these cells include type 2 cytokines as well as direct activators of smooth-muscle constriction and edema. IL-5 has a critical role in regulating eosinophils. It controls formation, recruitment, and survival of these cells. IL-13 induces airway hyperresponsiveness, mucus hypersecretion, and goblet cell metaplasia. While allergen exposure in allergic individuals can elicit a cascade of activation of type 2 inflammation, it is now understood (see Fig. 298-3) that nonallergic stimuli can elicit production of type 2 cytokines, particularly due to stimulation of type 2 innate lymphoid cells (ILC2). These cells can produce IL-5 and IL-13. ILC2s can be activated by epithelial cytokines known as alarmins, which are produced in response to "nonallergic" epithelial exposures such as irritants, pollutants, oxidative agents, fungi, or viruses. Thus, these "nonallergic" stimuli can be associated with eosinophilia.

The development of anti–IL-5 drugs that dramatically reduce eosinophils has allowed us to determine that, in many asthmatics, eosinophils play a major role in asthma pathobiology. They may induce hyperresponsiveness through release of oxidative radicals and major basic protein, which can disrupt the epithelium. They produce cysteinyl-leukotrienes that directly stimulate smooth muscle contraction and resulting in airway constriction. In addition, recent CT imaging has suggested that mucus plugs, which may contain significant amounts of eosinophil aggregates, may accumulate in the airways and contribute to asthma severity.

Non–Type 2 Processes As shown in Fig. 298-2, multiple processes can contribute to airway narrowing and apparent airway hyperresponsiveness. While type 2 inflammatory processes are most common, non–type 2 processes can exist either in combination with type 2 inflammation or without type 2 inflammation. Neutrophilic inflammation, as shown in Fig. 298-3, can also occur. This type of inflammation is more commonly seen in severe asthma that has not responded to the common anti-inflammatory therapies, such as corticosteroids, that usually suppress type 2 inflammation. In some cases, it may also be associated with chronic infection, occasionally with atypical pathogens such as *Mycoplasma*, perhaps accounting for the response of some of these patients to macrolide antibiotics. It is also commonly seen in reactive airway dysfunction syndrome (see "Etiologic Mechanisms, Risk Factors, Triggers, and Complicating Comorbidities").

In a small subset of asthmatics, the pathologic changes seen in Fig. 298-2 may occur without any evidence of tissue infiltration by inflammatory cells. The etiology of such pauci-granulocytic asthma is unclear.

MEDIATORS

Many chemical substances or signaling factors can contribute to the pathobiologic picture of asthma. Some of them have been successfully targeted in developing asthma therapies.

Cytokines As illustrated in Fig. 298-3, and as discussed above, IL-4, IL-5, and IL-13 are the major cytokines associated with type 2 inflammation. They have all been targeted successfully in asthma therapies. Thymic stromal lymphopoietin (TSLP), IL-25, and IL-33 also play a role in the signaling cascade and are being actively studied as targets for treatment of asthma. IL-9 has been implicated as well. IL-6, IL-17, tumor necrosis factor α (TNF-α), IL-1β, and IL-8 have been implicated in non–type 2 inflammation.

Fatty Acid Mediators Proinflammatory mediators derived from arachidonic acid include leukotrienes and prostaglandins. The cysteinyl leukotrienes (leukotrienes C_4, D_4, and E_4) are produced by eosinophils and mast cells. They are potent smooth-muscle constrictors. They also stimulate mucus secretion, recruit allergic inflammatory cells, cause microvascular leakage, modulate cytokine production, and influence neural transmission. Cysteinyl leukotriene

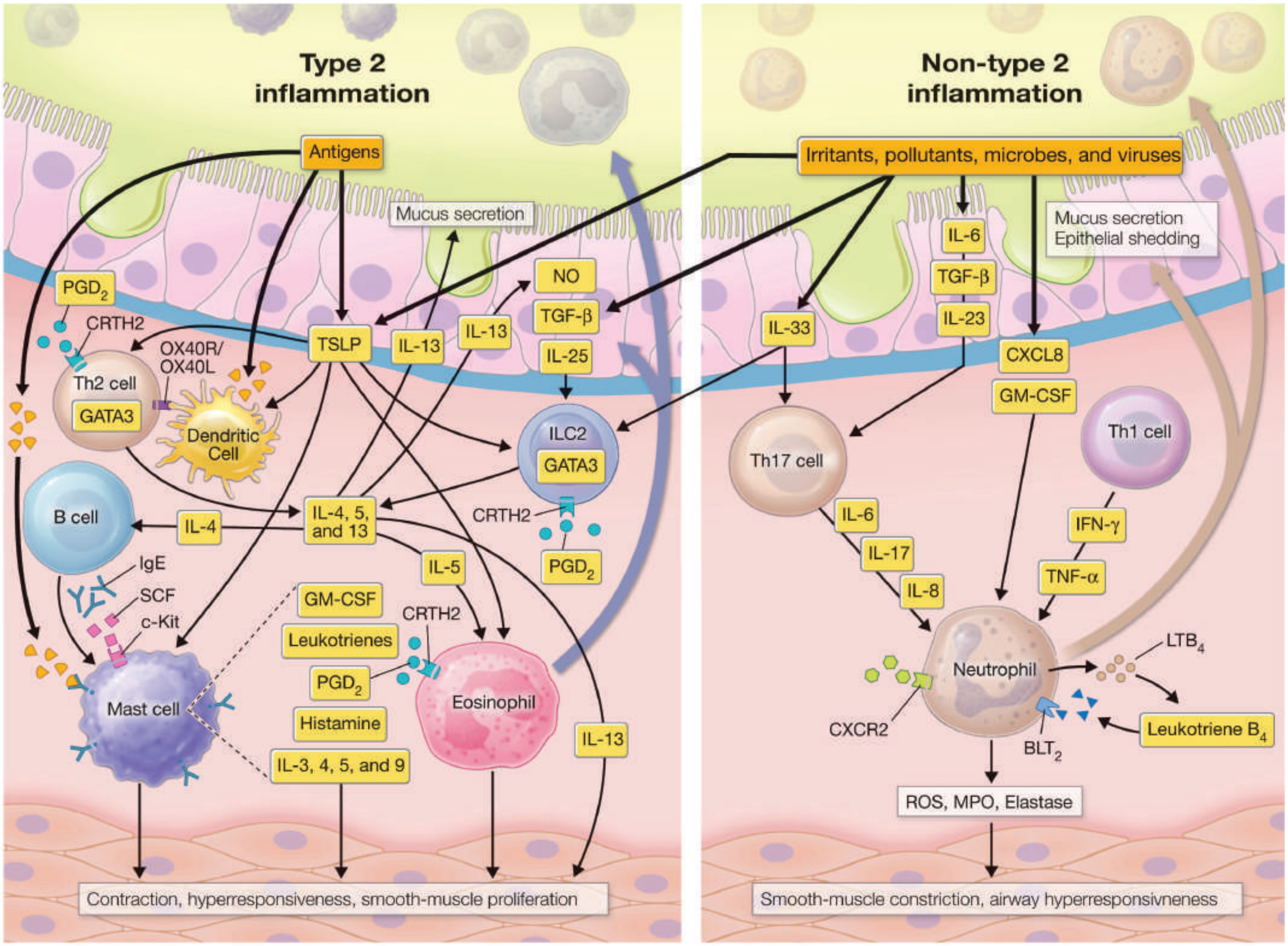
Type 2 inflammation
Antigens
Mucus secretion
PGD$_2$
CRTH2
OX40R/OX40L
Th2 cell
GATA3
Dendritic Cell
TSLP
IL-13
IL-13
NO
TGF-β
IL-25
ILC2
GATA3
B cell
IL-4
IL-4, 5, and 13
CRTH2
PGD$_2$
IgE
SCF
c-Kit
IL-5
GM-CSF
Leukotrienes
PGD$_2$
Histamine
IL-3, 4, 5, and 9
CRTH2
Mast cell
Eosinophil
IL-13
Contraction, hyperresponsiveness, smooth-muscle proliferation
Non-type 2 inflammation
Irritants, pollutants, microbes, and viruses
IL-6
TGF-β
IL-23
Mucus secretion
Epithelial shedding
IL-33
CXCL8
GM-CSF
Th17 cell
Th1 cell
IL-6
IL-17
IL-8
IFN-γ
TNF-α
LTB$_4$
Neutrophil
CXCR2
BLT$_2$
Leukotriene B$_4$
ROS, MPO, Elastase
Smooth-muscle constriction, airway hyperresponsivneness

FIGURE 298-3 Inflammatory cells and mediators involved in type 2 and non–type 2 inflammation. Allergens and nonallergic stimuli can trigger activation of multiple inflammatory cells and release of mediators that are responsible for recruiting and activating these cells. The mediators can affect airway smooth-muscle proliferation and hyperresponsiveness and fibroblast proliferation and matrix deposition. BLT2, leukotriene B$_4$ receptor 2; CRTH2, chemoattractant receptor-homologous molecule (PGD2 receptor); CXCL8, CXC motif chemokine ligand 8; CXCR2, CXC chemokine receptor 2; GATA3, GATA binding protein 3; GM-CSF, granulocyte-macrophage colony-stimulating factor; IFN-γ, interferon gamma; IL, interleukin; ILC2, innate lymphoid type 2 cells; c-Kit, mast/stem cell growth factor receptor; LTB4, leukotriene B4; MPO, myeloperoxidase; NO, nitric oxide; OX40L/OX40R, OX40 ligand/OX40 receptor; PGD2, prostaglandin D2; ROS, reactive oxygen species; SCF, stem cell factor; TGF-β, transforming growth factor β; Th, T helper; TNF-α, tumor necrosis factor α; TSLP, thymic stromal lymphopoietin.

modifiers have shown clinical benefit in asthma. The non-cysteinyl leukotriene, LTB$_4$, is produced primarily from neutrophils but can also be synthesized by macrophages and epithelial cells. It is a potent neutrophil chemoattractant. Prostaglandins are for the most part proinflammatory. Prostaglandin D$_2$ (PGD$_2$) is produced by mast cells. Receptors for PGD$_2$ (CRTH$_2$ receptors) are present on T$_H$2 cells, ILC2 cells, mast cells, eosinophils, macrophages, and epithelial cells, and the activation of these receptors upregulates type 2 inflammation. However, studies with drugs blocking CRTH$_2$ have been relatively disappointing thus far.

There are several classes of fatty acid–derived mediators that are responsible for the resolution of inflammation. These include the resolvins and lipoxins. Several studies suggest that deficiencies in these moieties may be responsible for the ongoing inflammation in asthma, especially in severe asthma.

Nitric Oxide Nitric oxide is a potent vasodilator, and in vitro studies suggest that it can increase mucus production and smooth-muscle proliferation. It is produced by epithelial cells, especially in response to IL-13, and by stimulated inflammatory cells including eosinophils, mast cells, and neutrophils. Its precise role in the asthmatic diathesis is unclear. However, its production is increased in the airways in the presence of IL-13-induced inflammation, and it can be detected in exhaled breath and is reduced by interventions that interfere with IL-13 production or activity.

Reactive Oxygen Species When allergens, pollutants, bacteria, and viruses activate inflammatory cells in the airway, they induce respiratory bursts that release reactive oxygen species that result in oxidative stress in the surrounding tissue. Increases in oxidative stress have been shown to affect smooth-muscle contraction, increase mucus secretion, produce airway hyperresponsiveness, and result in epithelial shedding.

Chemokines A variety of chemokines are secreted by the epithelium (as well as other inflammatory cells) and attract inflammatory cells into the airways. Those of particular interest include eotaxin (an eosinophil chemoattractant), TARC and MDC (which attract T$_H$2 cells), and RANTES (which has pluripotent pro-phlogistic effects).

ETIOLOGIC MECHANISMS, RISK FACTORS, TRIGGERS, AND COMPLICATING COMORBIDITIES

As illustrated in Fig. 298-1, the development of asthma involves an interplay between risk factors and exposures (see Table 298-1) and genetic predisposition.

■ HERITABLE PREDISPOSITION

Asthma has a strong genetic predisposition. Family and twin studies suggest a 25–80% degree of heritability. Genetic studies suggest complex polygenic inheritance complicated by interaction with environmental exposures. Further, epigenetic modifications related to environmental exposures may also produce heritable patterns of asthma. Many of the genes related to asthma have been associated with risk for atopy. However, there appear to be genetic modifications that predispose to asthma and its severity. Association studies have identified multiple candidate genes. In many cases, these genes vary from population to population. The most consistently identified include *ORMDL3/GSDMB* (in the 17q21 chromosomal region), *ADAM33*, *DPP-10*, *TSLP*, *IL-12*, *IL-33*, *ST2 (IL-33 receptor)*, *HLA-DQB1*, *HLA-DQB2*, *TLR1*, *IL-13,* and *IL6R*. Several of these genes have been identified in relation to the pathways involved in airway inflammation (Fig. 298-3). Regulation of ORMDL3 has been associated with altered T_H2 cytokine levels. It is important to note that there is considerable evidence of ancestry heterogeneity (difference in the strength of genetic association by ancestry) so that, for example, an association with PYHIN1 (interferon inducible protein X) is only found among people of African descent. In many cases, association studies have identified polymorphisms in noncoding regions of the genome, suggesting that the majority of the currently identified traits act as "enhancers" of biologic processes.

Genetic polymorphisms have also been associated with differential responses to asthma therapies. Variants in the β-receptor (Arg16Gly in $ADRB_2$), the glucocorticoid-induced transcript 1 gene, and genes in the leukotriene synthesis and receptor pathways have been associated with altered response to the pharmacologic agents acting at those receptors or through those pathways.

While genetic variation plays a key role in asthma susceptibility, it is important to understand that unraveling the complexities of the genetic contribution to asthma remains elusive. To wit, only 7.2% of asthma risk can be explained by the single nucleotide polymorphisms that have been associated with asthma. Recently, polygenic risk score analysis has been used as a tool to attempt to improve the prediction of asthma risk.

A significant proportion of the heritability of asthma relates to the heritability of atopy. Atopy is the genetic tendency toward specific IgE production in response to allergen exposure. Serum levels of IgE correlate closely with the development of asthma. The National Health and Nutrition Examination Survey (NHANES) III found that half of asthma in patients aged 6–59 could be attributed to atopy with evidence of allergic sensitization. The allergens most associated with risk include house dust mites, indoor fungi, cockroaches, and indoor animals.

EXPOSURES AND RISK FACTORS

Allergic Sensitization and Allergen Exposure Like asthma, the development of allergic sensitization involves an interplay between heritable susceptibility and allergen exposure. Allergen exposure during vulnerable developmental periods is believed to increase the risk of development of allergic sensitization in those with a tendency toward atopy. Allergic sensitization is increased in industrialized nations. Research comparing rural environments has suggested that early, varied microbiome exposure (exposure to bacteria and bacterial products) may reduce the development of atopy. Studies of the role of early allergen avoidance in reducing the risk of developing asthma have produced contradictory results.

Tobacco Maternal smoking and secondhand smoke exposure are associated with increased childhood asthma. Childhood secondhand smoke exposure increased asthma risk twofold. Active smoking is estimated to increase the incidence of asthma by up to fourfold in adolescents and young adults.

Air Pollution Early life exposure to pollution increases the risk of development of asthma. Proximity to major roadways increases the risk of early childhood asthma, thought to be attributed to levels of nitrogen dioxide exposure. Decreasing nitrogen dioxide exposure has been found to decrease asthma incidence in children. Studies of exposure to mixed pollutants suggest that most risk lies with carbon monoxide and nitric dioxide, with marginal effects of sulfur dioxide. Indoor air pollution from open fires and use of gas stoves has been associated with increased risk of children developing asthma symptoms. Mechanistically, pollutants are thought to cause oxidative injury to the airways, producing airway inflammation and leading to remodeling and increased risk of airway sensitization.

Infections Respiratory infections clearly can precipitate asthma deteriorations. However, the degree to which respiratory infections indicate susceptibility to asthma, represent a causal factor, or in some cases provide protection from asthma is unclear. Incidence and frequency of human rhinovirus and respiratory syncytial virus infections in children are associated with development of asthma, but whether they play a causal role is unclear. Evidence of prior *Mycoplasma pneumoniae* infection has been associated with the development of asthma in Taiwanese adults.

Occupational Exposures Occupational asthma is estimated to account for 10–25% of adult-onset asthma. The occupations associated with the most cases in European Community Health Surveys were nursing and cleaning. Two types of exposures are recognized: (1) an immunologic stimulus (further subdivided into high-molecular-weight [e.g., proteins, flour] and low-molecular-weight [e.g., formaldehyde, diisocyanate] stimuli based on whether they act as haptens or can directly stimulate a response), and (2) an irritative stimulus. The immunologic form is associated with a latency period between time of exposure and development of symptoms. The irritative form, known as reactive airway dysfunction syndrome (RADS), will be discussed below. A combination of genetic predisposition (including atopy), timing, intensity of exposure, and co-exposure (e.g., smoking) influences whether an individual will develop occupational asthma.

Diet There are suggestions that prenatal diet or vitamin deficiency may alter the risk of developing asthma. The evidence is not yet definitive, but vitamin D insufficiency may increase asthma risk in the progeny and supplementation may decrease such risk. Similarly, preliminary studies suggest that maternal supplementation with vitamins C and E and zinc may decrease asthma in children. One study suggested that maternal polyunsaturated fatty acid supplementation may decrease childhood asthma risk. Observational studies have suggested that increased maternal sugar intake may increase childhood asthma risk.

Obesity Multiple studies suggest that obesity may be a risk factor for development of childhood and adult asthma. Adipokines and IL-6 have been thought to play a pathobiologic role. Some have argued that the risk is overestimated, and a study from NHANES II found an association with dyspnea but not with airway obstruction.

Medications There are conflicting data regarding prenatal and early childhood risk for asthma posed by certain classes of medications. Use of H_2 blockers and proton pump inhibitors in pregnancy has been associated with an increased risk of asthma in children (relative risk, 1.36–1.45); however, another study found a small risk for H_2 blockers only. Conflicting data have been presented on the risk of perinatal acetaminophen and early childhood acetaminophen use. In a prospective study, the use of acetaminophen was not associated with an increased risk of exacerbations in young children with asthma, when compared to ibuprofen.

Prenatal and Perinatal Risk Factors Preeclampsia and prematurity have been associated with increased risk of asthma in the progeny. Babies born by cesarean section are at higher risk for asthma. Those with neonatal jaundice are also at increased risk. Breast-feeding reduced early wheezing but has a less clear effect on later incidence of asthma.

Endogenous Developmental Risk Factors Asthma is more prevalent among boys than girls, with the difference receding by age 20 and reversing (with increased prevalence among women) by age 40. Atopy is more prevalent among boys in childhood, and they have reduced airway size compared with girls. Both factors are thought to contribute to the sex discrepancy. A subset of women develop asthma around menopause. Such asthma tends to involve non–type 2 mechanisms. Pregnancy may precipitate or aggravate asthma as well.

High-Concentration Irritant Exposure and RADS A solitary exposure to a high concentration of irritant agents that rapidly (usually within hours) produces bronchospasm and bronchial hyperactivity is known as RADS. Causative agents include oxidizing and reducing agents in an aerosol or high levels of particulates. The acute pathology usually involves epithelial injury with neutrophilia. There is little evidence of type 2 inflammation. This syndrome differs from occupational asthma in that these patients have not become sensitized to the provocative agent and can return to work in that environment once they have recovered. However, the course of the disease may be variable, with some series showing documented abnormalities and persistent symptoms 10 years after exposure.

Fungi and Allergic Airway Mycoses One to 2% of patients with asthma may have an IgE-mediated sensitization to colonization of the airway by fungi, with the most common fungus causing such a reaction being *Aspergillus fumigatus*. So-called allergic bronchopulmonary aspergillosis (ABPA) is characterized by a type 2 airway inflammatory response to aspergillus with IgE >1000 IU/mL, eosinophils >500/μL, positive skin test to *Aspergillus,* and specific IgE and IgG antibodies to *Aspergillus*. Patients may have intermittent mucus plugging and central bronchiectasis. Up to two-thirds of patients will grow *Aspergillus* from the sputum. Treatment involves systemic antifungal treatment with itraconazole or voriconazole and oral corticosteroids. A role for biologics has also been suggested.

Exercise-Induced Symptoms in Elite Athletes Exercise-induced airway narrowing in elite athletes undertaking extreme exercise in strenuous condition. These athletes may have little, or no, airway hyperreactivity or asthma risk factors. The condition may involve additional mechanisms including direct epithelial injury. Such a syndrome has also been reported in swimmers possibly related to pool chlorination.

TRIGGERS OF AIRWAY NARROWING

The risk factors and exposures reviewed above lead to increased airway reactivity and a propensity to react to factors that trigger airway narrowing (see Fig. 298-1). Almost all asthmatics can identify triggers that will make their asthma worse. The major triggers are listed in Table 298-2. Many of them overlap with the risk factors and etiologic factors reviewed above. In some cases, elimination of these triggers may dramatically reduce the impairment caused by asthma. In a minority, abatement can lead to "remission" so that these patients no longer require asthma medications and do not experience bronchospasm with their daily activities and routines. While acute exposures to these triggers generally cause short-lived bronchospasm, the bronchospasm may be severe enough that treatment for an exacerbation is required. Chronic exposure may lead to permanent deterioration in asthma control, although this does not appear to be true for exercise or stress. It should be noted that evidence suggests that severe asthma exacerbations (those requiring systemic corticosteroids) may, in and of themselves, accelerate lung function decline.

Allergens In patients with sensitization to particular allergens through production of allergen-specific IgE, exposure to those allergens by inhalation can result in activation of mast cells and basophils with acute production of bronchoactive mediators (see Fig. 298-3). Such exposure can produce immediate bronchospasm (early response) and a late response (2–24 h after exposure) with bronchial narrowing and inflammation. These mechanisms can account for reactions to inhalation of pollens, mold, or dust; insects (especially cockroaches); animals; occupational materials; seasonal worsening of asthma; and so-called "thunderstorm asthma." Chronic exposure may lead to persistent symptoms. While food allergies can produce bronchospasm through anaphylaxis, food allergies are generally not etiologically linked to asthma.

Irritants Many asthmatics report increased symptoms on exposure to strong odors, smoke, combustion products, cleaning fluids, or perfumes. In general, the effects are short-lived, although chronic exposure (see "Occupational Exposures" above) and large-quantity exposures (see discussion of RADS above) can lead to long-lasting or permanent symptoms.

Viral Infections Most asthmatics report that asthma exacerbations can be triggered by upper respiratory infections. The inflammation that occurs may be neutrophilic as well as eosinophilic. There is some evidence that IgE generation may reduce production of interferon, possibly predisposing to the effects of upper respiratory viruses. Increased airway reactivity after viral infections generally persists for 4–6 weeks but, in some cases, may be associated with permanent changes and impairment. The almost 50% reduction in exacerbations during the COVID-19 pandemic has been attributed to decreased viral infections.

Exercise and Cold/Dry Air Exercise may be a trigger to asthmatic bronchoconstriction in patients with asthma. Hyperventilation that occurs with exercise dries the airway lining, changing the tonicity of lining cells and causing release of bronchoconstrictive mediators. This effect is more prominent the lower the moisture content of the air, and since cold air has a lower absolute moisture content, the lower the temperature of the inspired air, the less exercise is required to induce bronchoconstriction. In addition, cold air may produce airway edema during airway wall rewarming. At routine levels of exercise, these effects are short-lived.

Air Pollution Increased rates of exacerbations have been associated with increased ambient ozone, sulfur dioxide, and nitrogen dioxide, among other air pollutants.

Drugs Beta blockers may trigger bronchospasm even when used solely in ophthalmic preparations. While the more selective beta blockers are safe for most asthmatics, beta blocker use may be a cause of difficult-to-control asthma. Aspirin may precipitate bronchospasm in those with aspirin-exacerbated respiratory disease (see "Special Considerations"). Angiotensin-converting enzyme (ACE) inhibitors (and to a lesser extent angiotensin receptor blockers) may cause cough that may be attributed to poorly controlled asthma.

Occupational Exposures In addition to RADS (see above), episodic and/or recurrent exposures to workplace irritants and/or substances to which one has become sensitized can produce symptoms. These symptoms are usually reduced when patients are away from such exposures on weekends or vacation.

Stress Asthmatics may report increased symptoms with stress. The mechanisms are poorly understood.

Hormonal Factors A small proportion of women report a regular increase in perimenstrual symptoms, and symptoms may worsen during perimenopause. This may be related to rapid fluctuations in estrogen levels. Pregnancy can precipitate worsening of asthma in approximately one-third of pregnant patients.

COMORBIDITIES

Comorbidities may make asthma difficult to manage, and the common comorbidities are listed in Table 298-3.

Obesity Obese adults with asthma have more severe asthma symptoms than lean adults and are two to four times more likely to be hospitalized with an asthma exacerbation. Nonrandomized studies have shown an improvement and significant reduction in exacerbations after bariatric surgery.

Gastroesophageal Reflux Disease The presence of gastroesophageal reflux disease (GERD) predicts poor quality of life and is an independent predictor of asthma exacerbations. Treatment of symptomatic reflux disease has been shown to produce modest improvements in airway function, symptoms, and exacerbation frequency. Treatment of asymptomatic patients has not shown a benefit.

TABLE 298-3 Differential Diagnosis and Comorbidities That May Make Asthma Difficult to Control

Differential Diagnosis of Diseases with Overlapping Symptoms That Can Present with Obstructive Pulmonary Function Tests
1. Heart failure
2. Chronic obstructive pulmonary disease (COPD)
3. α_1 Antitrypsin deficiency
4. Airway obstruction from mass or foreign body
5. Inducible laryngeal dysfunction (vocal cord dysfunction)
6. Bronchiolitis obliterans
7. Bronchiectasis
8. Tracheobronchomalacia
Comorbidities That Can Make Asthma Difficult to Control
1. Chronic rhinosinusitis +/– nasal polyposis
2. Obesity
3. Gastroesophageal reflux disease
4. Inducible laryngeal obstruction (vocal cord dysfunction)
5. COPD
6. Anxiety/depression
7. Obstructive sleep apnea

Rhinosinusitis and/or Nasal Polyposis Rhinosinusitis may be a manifestation of the eosinophilic inflammation in the lower airway in asthma. In addition, poorly controlled rhinosinusitis is believed to aggravate asthma by several potential mechanisms including inflammatory and irritant effects of the secretions on the lower airway, neural reflexes, and production of inflammatory mediators and cells that produce systemic inflammation. Treatment with intranasal corticosteroids has been shown to decrease airway reactivity and emergency department visits and hospitalizations. Evidence for the benefit of surgical therapy is inconclusive. There is increasing evidence that biologics targeted at type 2 inflammation may also be particularly useful for asthma associated with rhinosinusitis and polyposis in particular.

Nasal polyposis is rare in children, and its presence in adults with asthma should raise suspicions of aspirin-exacerbated respiratory disease (see "Special Considerations").

Inducible Laryngeal Obstruction Previously known as vocal cord dysfunction, this process involves inappropriate narrowing of the larynx, producing resistance to airflow; it can complicate asthma as well as mimic it, and it is more commonly seen in women and patients with anxiety and depression. Definitive diagnosis involves laryngoscopy during symptomatic episodes or during induced obstruction.

Chronic Obstructive Pulmonary Disease (COPD) See "Asthma-COPD Overlap" under "Special Considerations."

Anxiety/Depression Increased rates of asthma exacerbations occur in asthmatics with anxiety, depression, or chronic stress. Some patients may be unable to distinguish anxiety attacks from asthma.

DIAGNOSIS AND EVALUATION

APPROACH

A presumptive diagnosis of asthma can usually be made based on a compatible history of recurrent wheezing, shortness of breath, chest tightness, or cough related to common bronchoconstrictor precipitants when appropriate components of the differential diagnosis have been considered and/or eliminated. In some cases, a therapeutic trial of low-dose inhaled corticosteroid (ICS) may be considered. In all but the mildest cases, the diagnosis should be confirmed with pulmonary function testing or demonstration of airway hyperresponsiveness. Unfortunately, the diagnosis may be difficult to confirm after initiation of therapy since airway obstruction and hyperresponsiveness may be mitigated with therapy. A trial of tapering medications may be necessary. Studies have shown that more than one-third of patients with a physician diagnosis of asthma do not meet the criteria for the diagnosis.

Adjunctive evaluation, as outlined below, should be undertaken to identify precipitating factors and underlying mechanisms that may be amenable to specific therapies (e.g., allergen avoidance). Cases that require more than a daily moderate-dose ICS combined with a long-acting β_2-agonist (LABA) (together known as ICS/LABA) should undergo more formal evaluation to assess comorbidities that may make asthma difficult to control and a reassessment of any possible confounding diagnoses that may mimic asthma symptoms (see Table 298-3).

PRIMARY ASSESSMENT TOOLS FOR ESTABLISHING A DIAGNOSIS

History Patients with asthma most commonly complain of episodes of wheezing, shortness of breath, chest tightness, mucus production, or cough upon exposure to triggers listed in Table 298-2. Symptoms may be worse on arising in the morning. Some may have nocturnal symptoms alone. The latter such patients should be evaluated for postnasal drip or GERD if that is their sole presenting symptom. Patients frequently complain of symptoms with rapid changes of temperature or humidity. Exercise-induced symptoms are common and frequently reported with increased sensitivity to cold air. As compared to cardiac sources of dyspnea, exercise symptoms tend to develop more slowly after initiation of exercise and tend to resolve more slowly unless a β_2-agonist is administered after the onset of symptoms. A careful exposure history should be obtained for home (e.g., pets, molds, dust, direct or secondhand smoke), work (work environment and exposure to occupational sensitizers), and recreational (e.g., hobbies, recreational inhalants) exposures. Allergen-sensitized patients may complain of symptoms on exposure to known allergens such as animals and may complain of increased symptoms during specific pollen seasons. Up to two-thirds of patients with asthma will be atopic (as opposed to half of the U.S. population), and almost half will have a history of rhinitis, with many complaining of intermittent sinusitis. In patients with adult-onset asthma, a careful occupational history should be obtained and a history of reactions to nonsteroidal anti-inflammatory drugs (NSAIDs) or use of new medications, such as beta blockers (including ophthalmic preparations) and ACE inhibitors (due to potential cough), should be elicited.

Physical Examination In between acute attacks, physical findings may be normal. Many patients will have evidence of allergic rhinitis with pale nasal mucus membranes. Five percent or more of patients may have nasal polyps, with increased frequency in those with more severe asthma and aspirin-exacerbated respiratory disease. Some patients will have wheezing on expiration (less so on inspiration). During an acute asthma attack, patients present with tachypnea and tachycardia, and use of accessory muscles can be observed. Wheezing, with a prolonged expiratory phase, is common during attacks, but as the severity of airway obstruction progresses, the chest may become "silent" with loss of breath sounds.

Pulmonary Function Tests Effective reduction of the airway lumen in asthma produces increased resistance to airflow, which can be detected as a reduction in expiratory airflow during forced expiratory maneuvers. The peak expiratory flow rate (PEFR), forced expiratory volume in 1 s (FEV_1), and the FEV_1/forced vital capacity (FVC) ratio are reduced below the lower limit of normal. The flow-volume loop may show a characteristic scalloping **(see Chap. 297)**. These findings may not be present during acute attacks or on therapy (especially after recent use of bronchodilators). Reversibility has been newly defined as ≥10% increase in the FEV_1 percent of predicted (e.g., 60% predicted to 70% predicted) at least 15 min after administration of a β_2-agonist or after several weeks of corticosteroid therapy. Many continue to use the prior definition of a 12% increase in the baseline FEV_1 with a minimum 200-mL change. Diurnal peak flow variability of >20% has also been proposed as an indicator of reversible airways disease, but it is less reliable due to difficulties with quality control and variability of home assessments. Lung volumes and diffusing capacity should be normal in uncomplicated asthma. In more severe asthma with severe airway obstruction, lung volumes may indicate air trapping, and total lung capacity may be under- or overestimated depending on whether body box or gas dilution methods, respectively, are utilized. Oscillometry, which does not require forced expiratory maneuvers, is gaining increased use and can be particularly useful for diagnosing large airway and variable tracheal obstruction.

Assessment of Airway Responsiveness In cases where pulmonary function tests are nonconfirmatory and the diagnosis remains in doubt, testing to demonstrate increased reactivity to provocative stimuli in the laboratory can be undertaken. Methacholine, a cholinergic agonist, inhaled in increasing concentrations is most commonly used. A provocative dose producing a 20% drop in FEV_1 (PD_{20}) is calculated, with a cumulative value ≤400 μg indicative of airway reactivity. Mannitol is used as well, and occasionally, hypertonic saline may be used. Challenge with exercise and/or cold, dry air can be performed, with a positive response recorded if there is a ≥10% drop in FEV_1 from baseline. In the case of suspected environmental/occupational exposures, specific allergen challenges may be undertaken in specialized labs.

■ ADJUNCTIVE ASSESSMENT TOOLS

Eosinophil Counts A large proportion of asthma patients not treated with oral or high-dose ICSs will have eosinophil counts ≥300 cells/μL. Eosinophil counts correlate with severity of disease in population studies. Their presence in patients with severe asthma indicates a likelihood that the patient would respond to medications targeted at type 2 inflammation. Extremely elevated levels should prompt consideration of eosinophilic granulomatosis with polyangiitis or primary eosinophilic disorders.

IgE, Skin Tests, and Radioallergosorbent Tests Total serum IgE levels are useful in considering whether patients with severe asthma would be eligible for anti-IgE therapy. Levels >1000 IU/mL should prompt consideration of ABPA. Skin tests, or their in vitro counterparts that detect IgE directed at specific antigens (radioallergosorbent test [RAST]), can be useful in confirming atopy and suggesting allergic rhinitis, which can complicate asthma management. Allergy investigations may be useful, when correlated with a history of reactions, in identifying environmental exposures that may be aggravating asthma.

Exhaled Nitric Oxide Fraction of exhaled nitric oxide (FeNO) in exhaled breath is an indicator of type 2 inflammation (particularly IL-13 induced) in the airways. It is easily suppressed by ICSs and, thus, can be used to assess adherence in patients in whom it was initially elevated. Elevated levels (>35–40 ppb) in untreated patients are indicative of type 2 inflammation. Levels >20–25 ppb in patients with severe asthma on moderate- to high-dose ICS indicate either poor adherence or persistent type 2 inflammation despite therapy.

■ ADDITIONAL EVALUATION IN SEVERE/POORLY RESPONSIVE ASTHMA

In patients with poorly responsive asthma, additional evaluations for comorbidities (see Table 298-3) may be necessary, including sinus radiographic studies (even in those who have no symptoms of sinus disease) and esophageal studies in those who have symptoms of reflux. In patients with nonreversible disease, many obtain a serum α_1 antitrypsin level. Additionally, the following evaluations may be of utility in poorly responsive asthma.

Chest CT Chest CT can be useful to assess for the presence of bronchiectasis and other structural abnormalities that could produce airway obstruction. New image analysis tools are being used in the research setting to assess airway properties such as airway wall thickness, airway diameter, mucus plugging, and evidence of air trapping.

Sputum Induced sputum may be used in more specialized centers to help characterize type 2 and non–type 2 inflammation by detection of eosinophils and neutrophils, respectively. In severe asthma, there is some evidence that some patients may have localized persistent eosinophilic airway inflammation despite lack of peripheral eosinophils on blood analysis.

TABLE 298-4 Goals of Asthma Therapy

1. Reduction in symptom frequency to ≤2 times/week
2. Reduction of nighttime awakenings to ≤2 times/month
3. Reduction of reliever use to ≤2 times a week (except before exercise)
4. No more than 1 exacerbation/year
5. Optimization of lung function
6. Maintenance of normal daily activities
7. Satisfaction with asthma care with minimal or no side effects of treatment

TREATMENT

Asthma

GOALS OF ASTHMA THERAPY AND ASSESSMENT OF CONTROL

Goals of asthma therapy in terms of achieving control of symptoms and reducing risk (as reflected in frequency of asthma exacerbations) are listed in **Table 298-4**. The therapeutic agents used in treatment are discussed below, and an integrated approach to care is discussed subsequently.

A comprehensive treatment approach involves avoiding and reducing asthma triggers and, if necessary, the adjunctive use of medications. Asthma medications are primarily divided into those that relax smooth muscle and produce a fairly rapid relief of acute symptoms and those that target inflammation or mediator production. The former medications are commonly referred to as reliever medications, and the latter are known as controller medications.

REDUCING TRIGGERS

Mitigation As shown in Tables 298-1 and 298-2, triggers and exposures can cause asthma and make it difficult to control. In the case of those with occupational exposures, removal from the offending environment may sometimes result in complete resolution of symptoms or significant improvement. Secondhand smoke exposure and frequent exposure to combustion products of cannabis are remediable environmental exposures as well. The removal of pets that are clearly associated with symptoms can reduce symptoms. Pest control at home and in the school in those with evidence of IgE-mediated sensitivity (skin test or IgE RAST) may also be beneficial. The effect of dust or mold control in reducing asthma symptoms has been more variable. There is moderate evidence that dust control (impermeable mattress and pillowcase covers) in those patients with symptoms and sensitization may be effective in reducing symptoms *only* when conducted as part of a comprehensive allergen mitigation strategy.

Allergen Immunotherapy Allergen immunotherapy reduces IgE-mediated reactions to the allergens administered. It clearly reduces the symptoms of allergic rhinitis and thus may be helpful in reducing this comorbidity. The evidence for its effectiveness in isolated asthma in those who are sensitized and have clinical symptoms is variable. Due to the risk of anaphylaxis, guidelines generally recommend immunotherapy only in patients whose asthma is under control and who have mild to moderate asthma. The evidence base for the effectiveness of sublingual allergen immunotherapy in the treatment of asthma is not substantial.

Vaccination Respiratory infections are a major cause of asthma exacerbations. Adult patients with asthma are strongly advised to receive the currently available pneumococcal vaccines (regardless of age) and yearly influenza vaccines. COVID-19 vaccination and the respiratory syncytial virus vaccines are advised, as well.

MEDICATIONS

Bronchodilators Bronchodilators relax airway smooth muscle. There are three major classes of bronchodilators, β_2-agonists, anticholinergics, and theophylline.

β_2-Agonists Available in inhaled or oral form, these agents activate β_2-receptors present on airway smooth muscle. Such receptors are also present on mast cells, but they contribute little to the efficacy of these agents in asthma. β_2-receptors are G protein–coupled receptors that activate adenyl cyclase to produce cyclic AMP, which results in relaxation of smooth muscle.

Use β_2-Agonists are primarily used in inhaled forms to provide relief of bronchospasm or to reduce the degree of bronchospasm anticipated in response to exercise or other provocative stimuli. Regular use has been associated with tachyphylaxis of the bronchoprotective effect and possible increased airway reactivity. This may be more common in patients with a polymorphism at the 16th amino acid position of the β_2-receptor. Frequent short-acting β_2-agonist use has been associated with increased asthma mortality, resulting in decreased enthusiasm for their use in isolation without inhaled corticosteroids.

Short-Acting β_2-Agonists (SABAs) Albuterol (also known as salbutamol) is the most commonly used agent. Bronchodilation begins within 3–5 min of inhalation, and effects generally last 4–6 h. It is most commonly administered by metered-dose inhaler. Solutions for nebulization are also used, especially for relief of bronchospasm in children. Oral forms are available but are not commonly used.

Long-Acting β_2-Agonists Salmeterol and formoterol are the two available LABAs. They have an ~12-h duration of action. Formoterol has a quick onset comparable to the SABAs. Salmeterol has a slower onset of action. These agents can be used for prophylaxis of exercise-induced bronchospasm. In contrast to their use in chronic obstructive pulmonary disease (COPD), these agents are *not* recommended for use as monotherapy in the treatment of asthma. Their use in asthma is generally restricted to use in combination with an ICS.

Ultra-Long-Acting β_2-Agonists These agents (indacaterol, olodaterol, and vilanterol) have a 24-h effect. They are only used in combination with ICSs in the treatment of asthma.

Safety β_2-Agonists are fairly specific for the β_2-receptors, but in some patients and especially at higher doses, they can produce tremor, tachycardia, palpitations, and hypertension. They promote potassium reentry into cells, and at high doses, they can produce hypokalemia. Type B (nonhypoxic) lactic acidosis can also occur and is thought to be secondary to increased glycogenolysis and glycolysis and increased lipolysis, leading to a rise in fatty acid levels, which can inhibit conversion of pyruvate to acetyl-coenzyme A.

Increased asthma mortality was associated with high-potency β_2-agonists in Australia and New Zealand. Increased use of β_2-agonists for relief of bronchospasm is a clear marker of poor asthma control and has been associated with increased mortality. Questions had been raised as to whether adding LABAs to ICS might be associated with severe adverse asthma outcomes, but several studies have not detected such outcomes in comparison to maintaining the ICS dose.

Anticholinergics Cholinergic nerve–induced smooth-muscle constriction plays a role in asthmatic bronchospasm. Anticholinergic medications can produce smooth-muscle relaxation by antagonizing this mechanism of airway narrowing. Agents that have been developed for asthma have been pharmacologically designed to be less systemically absorbed so as to minimize their systemic anticholinergic effects. The long-acting agents in this class are known as long-acting muscarinic antagonists (LAMAs).

Use The short-acting agents in this class can be used alone for acute bronchodilation. They appear to be somewhat less effective than β_2-agonists and have a slower onset of action as well.

Safety Dry mouth may occur. At higher doses and in the elderly, acute glaucoma and urinary retention have been reported. There was a numerical (but not significant) difference in mortality in African Americans treated with ICS/LAMA versus ICS/LABA for asthma.

Theophylline Theophylline, an oral compound that increases cyclic AMP levels by inhibiting phosphodiesterase, is now rarely used for asthma due to its narrow therapeutic window, drug-drug interactions, and reduced bronchodilation as compared to other agents.

Controller (Anti-Inflammatory/Antimediator) Therapies So-called "controller" therapies reduce asthma exacerbations and improve long-term control, decreasing the need for intermittent use of bronchodilator therapies. None of these therapies have yet been shown to prevent progression of airway remodeling or the more rapid decline in lung function that can occur in a subset of asthma patients.

Corticosteroids Corticosteroids are particularly effective in reducing type 2 inflammation and airway hyperresponsiveness. Corticosteroids bind to a cytoplasmic glucocorticoid receptor to form a complex that translocates to the nucleus. The complex binds to positive and negative response elements that result in inhibition of T-cell activation; eosinophil function, migration, and proliferation; and proinflammatory cytokine elaboration and activation of nuclear factor-κB. It also attaches to other transcription factors, resulting in deactivation of other proinflammatory pathways.

Use Corticosteroids reduce airway hyperresponsiveness, improve airway function, prevent asthma exacerbations, and improve asthma symptoms. Corticosteroid use by inhalation (ICSs) minimizes systemic toxicity and represents a cornerstone of asthma treatment.

ICS ICSs are the cornerstone of asthma therapy. They take advantage of the pleiotropic effects of corticosteroids to produce salutary impact at levels of systemic effect considerably lower than oral corticosteroids. Their use is associated with decreased asthma mortality. They are generally used regularly twice a day as first-line therapy for all forms of persistent asthma. Doses are increased, and they are combined with LABAs to control asthma of increasing severity (see next section "ICS/LABA and ICS/SABA"). European guidelines now recommend their intermittent use even in the mildest forms of asthma to reduce the likelihood of exacerbations. Longer-acting preparations permitting once-a-day use are available. Their effects can be noticeable in several days, but continued improvement may occur over months of therapy, with the majority of improvement evident within the first month of regular use. Adherence to regular therapy is generally poor, with as few as 25% of total annual prescriptions being refilled. Very high doses are sometimes used to reduce oral corticosteroid requirements. Not all patients respond to ICS. Increasing evidence suggests that the most responsive patients are those with significant type 2–mediated asthma.

ICS/LABA and ICS/SABA ICSs are available in combination with a LABA. The combination produces asthma control using a lower dose of ICS. Guidelines suggest that they be considered for daily maintenance therapy once more than low-dose ICS are required for control. The World Health Organization Global Initiative for Asthma (GINA) Guidelines now suggest that Anti-Inflammatory Reliever (AIR) (a combination of a quick onset LABA (formoterol only) or a short acting beta-agonist, either of these agents combined with an inhaled corticosteroid (ICS/formoterol or ICS/SABA), be used instead of albuterol alone at ALL levels of severity (see Table 298-5 and "Approach to the Patient"). In the case of long-acting beta-agonist, ONLY formoterol can be used in this manner since it acts quickly, whereas the other LABAs have a slower onset of action and therefore have not been evaluated. Currently, ICS/SABA is only available in the United States.

Oral Corticosteroids Chronic oral corticosteroids (OCSs) at the lowest doses possible (due to side effects) are used in patients who cannot achieve acceptable asthma control without them. Alternate-day dosing may be preferred. OCSs are also used to treat asthma exacerbations, frequently starting at a dose of 40–60 mg/d of prednisone or equivalent with a rapid taper over 1–2 weeks. Since they

TABLE 298-5 Step Therapy for the Treatment of Asthma Ages 12+ (Modified from GINA and NAEPP)

	Address exposures and comorbidities (see Tables 298-2 and 298-3) Confirm inhaler technique and optimize adherence Move up or down steps based on control (see Table 298-3)				
	STEP 1	**STEP 2**	**STEP 3**	**STEP 4**	**STEP 5**
Preferred regular therapy	None	None[a] or low-dose ICS[b]	Low-dose ICS/formoterol	Medium-dose ICS/formoterol	High-dose ICS/LABA + add-on LAMA
Alternative regular therapy	None	LTRA	Medium-dose ICS	High-dose ICS	Anti-IgE or anti–IL-5 or anti–IL4-Rα or anti-TSLP
Adjunctive therapy					LTM, azithromycin, OCS[c]
As-needed (PRN) reliever therapy	ICS/formoterol (low dose) or SABA[b] or ICS/SABA[f]	ICS/formoterol (low dose)[a], or PRN concomitant ICS and SABA[a,b] or SABA[e,b] or ICS/SABA[f]	ICS/formoterol (low or medium dose)[d] or ICS/SABA[f]		

[a] If using as-needed ICS/formoterol or PRN concomitant ICS and SABA, this is an option in which no regular daily therapy is prescribed. [b] National Asthma Education and Prevention Program (NAEPP) recommendation. [c] To be avoided as much as possible. [d] PRN ICS/formoterol only suggested for steps 3 and 4 by NAEPP. [e] If using low-dose ICS as regular therapy. [f] Alternative GINA recommendation.

Abbreviations: GINA, Global Initiative for Asthma; ICS, inhaled corticosteroid; ICS/LABA, combined ICS and LABA in one device; ICS/SABA, combined ICS and SABA in one device; IL, interleukin; LABA, long-acting β-agonist; LAMA, long-acting muscarinic antagonist; LTM, leukotriene modifier; LTRA, leukotriene receptor antagonist; OCS, oral corticosteroid; PRN, as needed; SABA, short-acting β-agonist.

are well absorbed, they may also be used for managing hospitalized patients.

Intravenous Corticosteroids Intravenous preparations are frequently used in hospitalized patients. Patients are rapidly transitioned to OCS once their condition has stabilized.

Intramuscular Corticosteroids In high-risk, poorly adherent patients, intramuscular triamcinolone acetonide has been used to achieve asthma control and reduce exacerbations.

Safety Chronic administration of systemic corticosteroids is associated with a plethora of side effects including diabetes, osteoporosis, cataracts and glaucoma, bruising, weight gain, truncal obesity, hypertension, ulcers, depression, and accelerated cardiac risk, among others. Appropriate monitoring and infectious (pneumocystis pneumonia prophylaxis for those treated chronically with ≥20 mg prednisone/d) and bone health prophylaxis are necessary. Intermittent "bursts" of systemic corticosteroids to treat asthma exacerbations are associated with reduced side effects, but observational studies have suggested that the cumulative dose over time is associated with deleterious side effects.

ICSs have dramatically reduced side effects as compared to OCSs. At higher doses, bruising occurs and osteoporosis can accelerate. There is a small increase in glaucoma and cataracts. Local effects include thrush, which can be reduced by use of a spacer and gargling. Hoarseness may be the result of a direct myopathic effect on the vocal cords. Rare patients exhibit side effects even at moderate doses of ICS. Children may experience growth suppression.

Leukotriene Modifiers Agents that inhibit production of leukotrienes (through inhibition of 5-lipoxygenase) or the action of leukotrienes at the CysLT$_1$ receptor are moderately effective in asthma. They are also effective in reducing symptoms of allergic rhinitis. Montelukast, a CysLT$_1$ antagonist, is frequently used in children with mild asthma due to concerns of ICS-related growth suppression.

Leukotriene modifiers are effective in preventing exercise-induced bronchoconstriction without the tachyphylactic effects that occur with regular use of LABAs. They are particularly effective in aspirin-exacerbated respiratory disease, which is characterized by significant leukotriene overproduction. They have also shown modest effect as add-on therapy in patients poorly controlled on high-dose ICS/LABA.

CysLT$_1$ Antagonists Montelukast and zafirlukast are administered orally once or twice daily, respectively. The onset of effect is rapid (hours), with the majority of chronic effectiveness seen within 1 month.

5-Lipoxygenase Inhibition Zileuton in its extended form is administered orally twice a day.

Safety Montelukast is well tolerated, but has been associated with suicidal ideation. Zileuton increases liver function tests (transaminases) in 3% of patients. It inhibits CYP1A2.

Cromolyn Sodium Cromolyn sodium is an inhaled agent believed to stabilize mast cells. It is only available by nebulization and must be administered two to four times a day. It is mildly to modestly effective and appears to be helpful for exercise-induced bronchospasm. It is used primarily in pediatrics in those concerned about ICS side effects.

Anti-IgE Omalizumab, a monoclonal antibody to the Fc portion of the IgE molecule, prevents the binding of IgE to mast cells and basophils and thus blocks antigen-related signaling, which is responsible for production or release of many of the mediators and cytokines critical to asthma pathobiology. Over time, reduction in IgE production occurs as well. Anti-IgE has been shown to increase interferon production in rhinovirus infections, decrease viral-induced asthma exacerbations, and reduce the duration and peak viral shedding, probably due to interference with IgE's ability to reduce interferon γ production in response to viral infections.

Use In asthma, anti-IgE has been tested in patients with a circulating IgE ≥30 IU/mL and a positive skin test or RAST to a perennial allergen. It is generally used in patients not responsive to moderate- to high-dose ICS/LABA. It reduces exacerbations by 25–50% and can reduce asthma symptoms, but has minimal effect on lung function. Anti-IgE is dosed based on body weight and circulating IgE and is administered subcutaneously every 2–4 weeks depending on the calculated dose. In the United States, the maximum dose is 300 mg every 2 weeks, which generally restricts the drug to those with a body weight ≤150 kg. Most effects are generally seen in 3–6 months. Retrospective studies have suggested that patients with an exhaled nitric oxide approximately ≥20 ppb or circulating eosinophils ≥260/μL have the greatest response as ascertained by reduction in exacerbations. FeNO is slightly reduced by treatment, but circulating IgE, as measured by available clinical tests, is not affected since these tests measure total circulating IgE, not free IgE. Anti-IgE has also been found to be effective in patients with chronic idiopathic urticaria and nasal polyposis.

Safety The incidence of side effects is low. Anaphylaxis has been reported in 0.2% of patients receiving the drug.

IL-5–Active Drugs Mepolizumab and reslizumab are monoclonal antibodies that bind to IL-5, and benralizumab binds to the IL-5 receptor. They rapidly (within a day) reduce circulating eosinophils.

Use In patients symptomatic on moderate- to high-dose ICS/LABA, generally with two or more exacerbations that require OCS per year and with an eosinophil count of ≥300/μL (unless they are on chronic OCS), IL-5–active drugs reduce exacerbations by about half or more. FEV_1 and symptoms improve moderately as well. In patients who are not on chronic OCSs, these drugs are less effective in those with eosinophil counts <300/μL. In patients on chronic OCS, they reduce the need for OCSs regardless of circulating eosinophil count (presumably due to the fact that many of those patients have type 2 inflammation but their circulating eosinophils have been suppressed by the systemic OCS). FeNO and IgE are relatively unaffected by these drugs. Most clinical effects are usually seen within 3–6 months. Mepolizumab has been approved for treatment of nasal polyposis.

Safety These drugs are associated with minimal side effects.

Anti–IL-4/13 The IL-4 and IL-13 receptors are heterodimers that share a common subunit, IL-4 receptor α. Dupilumab is an antibody that binds to this subunit and, thus, blocks signaling through both receptors.

Use In addition to effectiveness in the phenotype of patients who respond to anti–IL-5 therapies, poorly controlled patients on moderate- to high-dose ICS/LABA with an FeNO of 20–25 ppb also appear to respond to dupilumab even if their peripheral eosinophils are not elevated. Dupilumab reduces exacerbations by ≥50%, decreases symptoms, and may produce more of an effect on FEV_1 than anti–IL-5 drugs. It has been shown to be effective for patients requiring oral corticosteroids regardless of biomarker levels. It gradually reduces FeNO and IgE levels. Paradoxically, circulating eosinophil counts may initially temporarily increase. Most effects are seen by 3–6 months of therapy. It is also approved for nasal polyposis atopic dermatitis, eosinophilic esophagitis, and lupus pernio also approved for patients with COPD with exacerbations and elevated eosinophils.

Safety Side effects are minimal, but cases of serious systemic eosinophilia, most frequently associated with the reduction of oral corticosteroids, have been noted.

Anti-TSLP Tezepelumab binds to TSLP, a proximal alarmin in the epithelial response cascade (Fig. 298-3). As a result, it affects multiple pathways and decreases blood eosinophils, FeNO, and IgE levels.

Use In addition to effectiveness in patients who qualify for and respond to anti–IL-5 therapies and anti–IL-4/13, tezepelumab has some degree of effectiveness in patients with asthma who have recurrent exacerbations without evidence of elevated levels of FeNO or blood eosinophils. This agent reduces exacerbations by 50–70% and improves lung function and symptoms. Available studies have not been able to demonstrate effectiveness in atopic eczema or in reducing OCS dose. Most effects are seen within 3–6 months of therapy, and biomarker levels may continue to fall over time.

Safety Minimal side effects have been reported.

Bronchial Thermoplasty, Alternative Therapies, and Therapies Under Development • **Bronchial Thermoplasty** This procedure involves radiofrequency ablation of the airway smooth muscle in the major airways administered through a series of three bronchoscopies for patients with severe asthma. There is some evidence that it may reduce exacerbations in very select patients. The procedure may be accompanied by significant morbidity, and most guidelines do not recommend it other than in the context of clinical trials or registries. The manufacturer has discontinued production of the devices necessary for this procedure and it may no longer be available.

Alternative Therapies Macrolides appear to be effective in a subset of asthmatics with poor response to other therapies and are suggested as possible adjunctive therapies in step 5 asthma (Table 298-5). Alternative therapies such as acupuncture and yoga have not been shown to improve asthma in controlled trials. Studies with placebo have demonstrated that there may be a significant response to placebo.

Therapies in Development Trials are underway targeting pathways and receptors shown in Fig. 298-3. Those in more advanced stages of development include therapies targeting IL-33, OX-40, and mediators involved in mucin production. Studies targeting IL-17 and TNF-α have not shown efficacy, but it is unclear if patients with appropriate endotypes were targeted. Proof-of-concept studies targeting mast cells via inhibition of tyrosine kinase have suggested efficacy in severe asthma.

APPROACH TO THE PATIENT

Asthma

U.S. (National Asthma Education and Prevention Program [NAEPP]) and World Health Organization (Global Initiative for Asthma [GINA]) guidelines advise a symptomatic approach to asthma treatment assuming that appropriate measures have been taken to address asthma triggers, exposures, and comorbidities enumerated in Tables 298-2 and 298-3. Additionally, adherence and inhaler techniques need to be addressed. Poor adherence or poor inhaler technique has been identified as the cause of poor asthma control in up to 50% of patients referred for poorly controlled asthma.

The stepwise approach to intensifying and reducing asthma therapy is outlined in Table 298-5. It involves "stepping" therapy up or down based on assessment of whether asthma is controlled by the criteria listed in Table 298-4. Assuming comorbidities have been addressed, adherence has been evaluated, education regarding avoiding triggers has been performed, and inhaler technique is verified, the cornerstone of preferred therapy is the intensification of ICS therapy in conjunction with the use of a LABA to achieve greater control at lower ICS doses.

A major change in the stepwise approach, advocated for more than two decades, has occurred. Evidence has accumulated that as-needed ICS can be used instead of regular ICS in milder asthma and that the trigger for such use could be patient perception of the need to use a reliever inhaler. This approach is now labeled AIR for Antiinflammatory Reliever. Since formoterol is a LABA with a rapid onset, combination ICS/formoterol has been used as a single agent in multiple studies: as needed without background therapy in milder asthma, and as needed in addition to twice-daily ICS/formoterol in more severe asthma. Since asthma mortality can occur even in mild asthma (albeit at lower rates than more severe asthma), GINA, as part of a comprehensive strategy of asthma management, recommends ICS/formoterol be used as the reliever in all steps of asthma severity, including intermittent asthma (step 1). With recent introduction of a combination ICS/SABA in the United States, GINA recognizes use of this preparation as alternative antiinflammatory reliever that can be used in place of albuterol. NAEPP guidelines utilizing evidence-based studies recommend that ICS/formoterol be used as the reliever medication in patients requiring step 3 and 4 therapy (see Table 298-5) and that as-needed concomitant ICS and SABA can be used as a therapy in step 2. For the sake of simplicity, an adapted GINA approach is outlined in Table 298-5 with footnotes identifying the major differences from the NAEPP. Leukotriene receptor antagonists (LTRAs) are alternative medications in step 2, which may be used in those concerned about the minimal ICS side effects. However, recent warnings about suicidal ideation associated with montelukast may make this approach less appealing. Leukotriene modifiers and long-acting anticholinergics are possible add-on (adjunctive) therapies in those requiring step 4 and/or 5 therapies. Biologics are incredibly effective in their specific endotypes (type 2 with exacerbations and specific biomarkers, as previously described), but their high cost currently relegates them to step 5 therapy or beyond.

SPECIAL CONSIDERATIONS

ASTHMA ATTACKS

Asthma deteriorations of mild to moderate severity can be initially treated with a β_2-agonist administered up to every 1 h. Increasing the dose of ICSs by four- to fivefold may be helpful as well. If patients fail to achieve adequate control and continue to require β_2-agonists hourly for several hours, they should be referred for urgent care. In the urgent care setting, PEFR or FEV_1 should be assessed, and patients are usually treated with nebulized β_2-agonists up to every 20 min. Those with PEFR >60% of predicted will frequently respond to β_2-agonists alone. If they fail to respond in 1–2 h, intravenous corticosteroids should be administered. Supplemental oxygen is usually administered to correct hypoxemia. An LTRA and magnesium are sometimes given as well. Nebulized anticholinergics can be administered to produce additional bronchodilation. Failure to achieve PEFR >60% or persistent severe tachypnea over 4–6 h should prompt consideration of admission to the hospital. In-hospital treatment may include continuous bronchodilator nebulization. Noninvasive positive-pressure ventilation to assist with respiratory exhaustion is sometimes used to prevent a need for intubation, and helium-oxygen mixtures may be used to decrease the work of breathing. Antibiotics should be administered only if there are signs of infection.

Most patients with asthma attacks present with hypocapnia due to a high respiratory rate. Normal or near-normal Pco_2 in a patient with asthma in respiratory distress should raise concerns of impending respiratory failure and need for mechanical ventilation. Mechanical ventilation may be difficult in patients with status asthmaticus due to high positive pressures in the setting of high resistance to airflow due to airway obstruction. Mechanical ventilation should aim for low respiratory rates and/or low ventilation volumes to decrease peak airway pressures. This can frequently be achieved by "permissive hypercapnia"—allowing the Pco_2 to rise and, if necessary, temporarily correcting critical acidosis with administration of bicarbonate to increase the pH. Neuromuscular paralysis may sometimes be beneficial. Bronchoscopy to clear mucus plugs has been described but may be dangerous in the setting of difficulties with mechanical ventilation.

HIGH-RISK ASTHMA PATIENTS

Three to four thousand people die from asthma in the United States each year. Table 298-6 lists characteristics of patients at high risk for asthma death. These characteristics should be considered in evaluating and treating patients who present with asthma.

EXERCISE-INDUCED SYMPTOMS

In many cases, the degree of exercise intolerance may reflect poor asthma control. Treatment involves step therapy of asthma as outlined in Table 298-5. In other cases, however, asthma may be well controlled in all other respects, but patients may report that they cannot undertake the level of exercise they desire. Some increase in exercise capacity can be achieved by starting at lower levels of exercise (warming up) and by using a mask in colder weather to condition the air. Pretreatment with a SABA can increase the threshold of ventilation required to induce bronchoconstriction. LABAs may extend the period of protection, but their use alone in asthma is to be discouraged. For occasional exercise, ICS/LABA can be used, but regular use may expose the patient to unnecessary doses of ICS. If regular exercise is undertaken, then LTRAs may provide protection and can be used regularly. A SABA (or ICS/formoterol) should always be available for quick relief.

Exercise-induced airway narrowing in elite athletes may be related to direct epithelial injury. In addition to the above, conditioning of incoming air may be of major assistance. Ipratropium has been reported to be of utility as well.

TABLE 298-6 Patients at Greater Risk for Asthma Mortality

1. History of intensive care unit admission for asthma
2. History of intubation for asthma
3. Illicit drug use
4. Depression
5. New diagnosis within past year
6. ≥2 emergency unit visits in past 6 months
7. Severe psychosocial problems
8. Lower socioeconomic status
9. On daily prednisone prior to admission

PREGNANCY

Asthma may improve, deteriorate, or remain unchanged during pregnancy. Poor asthma control, especially exacerbations, is associated with poor fetal outcomes. The general principles of asthma management and its goals are unchanged. Avoidance of triggers, especially smoking environments, is critical in view of the risk of loss of control and, in the case of smoking, its clear effects on risk of development of asthma in the child. There is extensive experience suggesting the safety of inhaled albuterol, beclomethasone, budesonide, and fluticasone, with reassuring information on formoterol and salmeterol in pregnancy. Animal studies have not suggested toxicity for montelukast, zafirlukast, omalizumab, and ipratropium. Antibodies cross the placenta, and there are few human data on the safety of the asthma biologics. Chronic use of OCS has been associated with neonatal adrenal insufficiency, preeclampsia, low birth weight, and a slight increase in the frequency of cleft palate. However, it is clear that poorly controlled asthma during pregnancy carries greater risk to the fetus and mother than these effects. There should be no hesitancy in administering routine pharmacotherapy for acute exacerbations. Initiation of allergen immunotherapy during pregnancy is not recommended. In cases where prostaglandins are needed to manage pregnancy, PGF2-α should be avoided since it is associated with bronchoconstriction.

ASPIRIN-EXACERBATED RESPIRATORY DISEASE

A subset of patients (5–10%) present in adulthood with difficult-to-control asthma and type 2 inflammation with eosinophilia, sinusitis, nasal polyposis, and severe asthma exacerbations that are precipitated by ingesting inhibitors of cyclooxygenase, with aspirin being the most prominent of such inhibitors. Such patients, classified as having aspirin-exacerbated respiratory disease, overproduce leukotrienes in response to inhibition of cyclooxygenase-1, probably secondary to inhibition of PGE_2. These patients should avoid inhibitors of cyclooxygenase-1 (aspirin and NSAIDs) but can generally tolerate inhibitors of cyclooxygenase-2 and acetaminophen. They should be treated with leukotriene modifiers. Aspirin desensitization can be undertaken to decrease upper respiratory symptoms and to allow chronic administration of aspirin or NSAIDs for those that require it. Dupilumab and the IL-5–active biologics appear to be particularly helpful (some have been approved to treat nasal polyposis in this setting) and appear to be superseding aspirin desensitization in management except when chronic administration of aspirin or NSAIDs is required for another therapeutic indication.

SEVERE ASTHMA

Severe and difficult-to-treat asthma, which composes ~5–10% of asthma, is defined as asthma that, having undergone appropriate evaluation for comorbidities and mimics, education, and trigger mitigation, remains uncontrolled on step 5 therapy or requires step 5 therapy for its control. Severe asthma can account for almost 50% of the cost of asthma care in the United States. A significant proportion of these patients have trouble with adherence and/or inhaler technique, and these factors need to be investigated vigorously. More than half of these patients have evidence of persistent eosinophilic inflammation as evidenced by peripheral blood eosinophils and/or induced sputum. Those with recurrent exacerbations have a substantially increased likelihood of responding to the type 2 targeted biologics. Tezepelumab has been shown to have some degree of effectiveness in those without elevated type 2 biomarkers. Treatment for those with mixed inflammation, isolated neutrophilic inflammation, or pauci-granulocytic inflammation remains to be determined. Some data suggest that many of these patients may have aberrations in the pathways responsible for resolution of inflammation. A rare patient may have biochemical abnormalities that interfere with steroid response pathways. Macrolides are of use in a subset. Studies targeting mast cells, IL-6, IL-33, and other pathways illustrated in Fig. 298-3 are underway. Therapies aimed at improving pro-resolving pathways may also be promising.

ELDERLY PATIENTS WITH ASTHMA

Asthma may present at or persist into older age. The mortality of asthma in those >65 years old is five times greater than that of younger cohorts even when adjusting for comorbidities. Many of these patients had asthma as children, some with quiescent periods as they entered adulthood. Of those with new-onset asthma, almost half were smokers or are currently smoking. One-quarter of adult-onset asthma is believed to be due to occupational exposure. Patients presenting with eosinophilic inflammation appear to have more severe asthma. Besides investigations of comorbidities, these patients may require adjustment to step therapy based on intolerance of β_2-agonist therapy due to arrhythmia or tremulousness. The coexistence of COPD needs to be carefully considered (see below).

ASTHMA-COPD OVERLAP

Most clinicians agree that asthma-COPD overlap is not a syndrome, but rather recognize that it may be useful to identify patients who present with symptoms related to airway dysfunction that may be due to simultaneous coexistence of both asthma and COPD. From an asthma perspective, recognition that COPD and smoking can alter the response to asthma therapies may be important. Smoking can blunt the response to ICS. Dupilumab has been shown to be effective in patients with COPD with recurrent exacerbations and elevated eosinophils and has been approved for such use. Additionally, in patients with both diseases, earlier initiation of anticholinergics may be considered.

FURTHER READING

Brusselle GG, Koppelman GH: Biologic therapies for severe asthma. N Engl J Med 386:157, 2022.

Cloutier MM et al: 2020 focused updates to the asthma management guidelines: A report from the National Asthma Education and Prevention Program Coordinating Committee Expert Panel Working Group. J Allergy Clin Immunol 146:1217, 2020.

Global Initiative for Asthma: 2023 GINA Report, global strategy for asthma management and prevention. *https://ginasthma.org/2023-gina-main-report/*.

Kaur R, Chupp G: Phenotypes and endotypes of adult asthma: Moving toward precision medicine. J Allergy Clin Immunol 144:1, 2019.

299

Hypersensitivity Pneumonitis and Pulmonary Infiltrates with Eosinophilia

Praveen Akuthota, Michael E. Wechsler

HYPERSENSITIVITY PNEUMONITIS

INTRODUCTION AND DEFINITION

Hypersensitivity pneumonitis (HP), also referred to as extrinsic allergic alveolitis, is a pulmonary disease that occurs due to inhalational exposure to a variety of antigens leading to an inflammatory response of the alveoli and small airways. Systemic manifestations such as fever and fatigue can accompany respiratory symptoms. Although sensitization to an inhaled antigen as manifested by specific circulating IgG antibodies is necessary for the development of HP, sensitization alone is not sufficient as a defining characteristic, because many sensitized individuals do not develop HP. The incidence and prevalence of HP are variable, depending on geography, occupation, avocation, and environment of the cohort being studied. As yet unexplained is the decreased risk of developing HP in smokers.

OFFENDING ANTIGENS

HP can be caused by any of a large list of potential offending inhaled antigens **(Table 299-1)**. The various antigens and environmental

TABLE 299-1 Examples of Hypersensitivity Pneumonitis

DISEASE	ANTIGEN	SOURCE
Farming/Food Processing		
Farmer's lung	Thermophilic actinomycetes (e.g., *Saccharopolyspora rectivirgula*); fungus	Grain, moldy hay, silage
Bagassosis	Thermophilic actinomycetes	Sugarcane
Cheese washer's lung	*Penicillium casei; Aspergillus clavatus*	Cheese
Coffee worker's lung	Coffee bean dust	Coffee beans
Malt worker's lung	*Aspergillus* species	Barley
Miller's lung	*Sitophilus granarius* (wheat weevil)	Wheat flour
Mushroom worker's lung	Thermophilic actinomycetes; mushroom spores	Mushrooms
Potato riddler's lung	Thermophilic actinomycetes; *Aspergillus* species	Moldy hay around potatoes
Tobacco grower's lung	*Aspergillus* species	Tobacco
Wine maker's lung	*Botrytis cinerea*	Grapes
Birds and Other Animals		
Bird fancier's lung (also named by specific bird exposures)	Proteins derived by parakeets, pigeons, budgerigars	Bird feathers, droppings, serum proteins
Duck fever	Duck feathers, serum proteins	Ducks
Fish meal worker's lung	Fish meal dust	Fish meal
Furrier's lung	Dust from animal furs	Animal furs
Laboratory worker's lung	Rat urine, serum, fur	Laboratory rats
Pituitary snuff taker's lung	Animal proteins	Pituitary snuff from bovine and porcine sources
Poultry worker's lung	Chicken serum proteins	Chickens
Turkey handling disease	Turkey serum proteins	Turkeys
Other Occupational and Environmental Exposures		
Chemical worker's lung	Isocyanates	Polyurethane foam, varnish, lacquer
Detergent worker's lung	*Bacillus subtilis* enzymes	Detergent
Hot tub lung	*Cladosporium* species; *Mycobacterium avium* complex	Contaminated water, mold on ceiling
Humidifier fever (and air conditioner lung)	Several microorganisms including: *Aureobasidium pullulans*; *Candida albicans*; thermophilic actinomycetes; *Mycobacterium* species; *Klebsiella oxytoca*; *Naegleria gruberi*	Humidifiers and air conditioners (contaminated water)
Machine operator's lung	*Pseudomonas* species; *Mycobacteria* species	Metal working fluid
Sauna taker's lung	*Aureobasidium* species; other antigens	Sauna water
Suberosis	*Penicillium glabrum*; *Chrysonilia sitophila*	Cork dust
Summer-type pneumonitis	*Trichosporon cutaneum*	House dust mites, bird droppings
Woodworker's lung	*Alternaria* species; *Bacillus subtilis*	Oak, cedar, pine, mahogany dusts

conditions described to be associated with HP give rise to an expansive list of monikers given to specific forms of HP. Antigens derived from fungal, bacterial, mycobacterial, bird-derived, and chemical sources have all been implicated in causing HP.

Categories of individuals at particular risk in the United States include farmers, bird owners, industrial workers, and hot tub users. Farmer's lung occurs as a result of exposure to one of several possible sources of bacterial or fungal antigens such as grain, moldy hay, or silage. Potential offending antigens include thermophilic actinomycetes or *Aspergillus* species. Bird fancier's lung (also referred to by names corresponding to specific birds) must be considered in patients who give a history of keeping birds in their home and is precipitated by exposure to antigens derived from feathers, droppings, and serum proteins. Occupational exposure to birds may also cause HP, as is seen in poultry worker's lung. Chemical worker's lung is provoked by exposure to occupational chemical antigens such as diphenylmethane diisocyanate and toluene diisocyanate. Mycobacteria may cause HP rather than frank infection, a phenomenon observed in hot tub lung and in HP due to metalworking fluid.

PATHOPHYSIOLOGY

While much remains to be learned regarding the pathophysiology of HP, it has been established that HP is an immune-mediated condition that occurs in response to inhaled antigens that are small enough to deposit in distal airways and alveoli. From an immunologic perspective, HP is characterized by dysregulated T_H1 and T_H17 immune responses. Although the presence of precipitating IgG antibodies against specific antigens in HP suggests a prominent role for adaptive immunity in the pathophysiology of HP, innate immune mechanisms likely also make an important contribution. This is highlighted by the observation that Toll-like receptors and downstream signaling proteins such as MyD88 are activated in HP, leading to neutrophil recruitment. Although no clear genetic basis for HP has been established, in specific cohorts, polymorphisms in genes involved in antigen processing and presentation, including TAP1 and major histocompatibility complex type II, have been observed. In chronic HP, bone marrow–derived fibrocytes may contribute to lung inflammation and fibrosis. Chronic HP shares some genetic overlap with idiopathic pulmonary fibrosis (IPF), with a polymorphism in the mucin gene *MUC5B* that is associated with IPF also associated with chronic HP.

CLINICAL PRESENTATION

Given the heterogeneity among patients, variability in offending antigens, and differences in the intensity and duration of exposure to antigen, the presentation of HP is accordingly variable. Although these categories are not fully satisfactory in capturing this variability, HP has been traditionally categorized as having *acute, subacute*, and *chronic* forms. However, recent professional society guidelines advocate abandoning this terminology in favor of categorizing HP as either *nonfibrotic* or *fibrotic*. The HP Study Group found by cluster analysis that a cohort of HP patients is best described in bipartite fashion, with one group featuring recurrent systemic signs and symptoms and the other featuring more severe respiratory findings.

Symptoms of HP may be acute and can be intense in nature, although HP can also present in a more insidious and slowly progressive manner. Systemic symptoms, including fevers, chills, and malaise, can occur and are accompanied by dyspnea and cough. Acute symptoms may resolve within hours to days if no further exposure to the offending antigen occurs but may be recurrent and episodic depending on exposure patterns. A similar presentation may occur as a culmination of intermittent episodes of acute HP. Although respiratory impairment may be quite severe, antigen avoidance generally results in resolution of the symptoms, but with a slower time course, on the order of weeks to months, than that seen with acute HP. Chronic HP can present with an even more gradual onset of symptoms than subacute HP, with progressive dyspnea, cough, fatigue, weight loss, and clubbing of the digits. The insidious onset of symptoms and frequent lack of an anteceding episode of acute HP make diagnosing chronic HP a challenge. Unlike with the other forms of HP, there can be an irreversible component to the respiratory impairment that is not responsive to removal of the responsible antigen from the patient's environment. The disease progression of chronic HP to lung fibrosis with honeycombing on chest imaging and hypoxemic respiratory failure can mirror that seen in IPF, with a similar prognosis. Diagnostic uncertainty between these two entities is not uncommon. Fibrotic lung disease is a potential feature of chronic HP due to exposure to bird antigens, whereas an emphysematous phenotype may be seen in farmer's lung.

Concordant with the variability in the presentation of HP is the observed variability in outcome. HP that has not progressed to chronic lung disease has a more favorable outcome with likely resolution if antigen avoidance can be achieved. However, chronic HP resulting in lung fibrosis has a poorer prognosis, with patients with chronic pigeon breeder's lung having demonstrated a similar mortality as seen in IPF.

DIAGNOSIS

Although there is no set of universally accepted criteria for arriving at a diagnosis of HP, diagnosis depends foremost on establishing a history of exposure to an offending antigen that correlates with respiratory and systemic symptoms. A careful occupational and home exposure history should be taken and may be supplemented if necessary by a clinician visit to the work or home environment. Specific inquiries will be influenced by geography and the occupation of the patient. Questionnaires designed to identify inciting agents are under development but require further validation. When HP is suspected by history, the additional workup is aimed at establishing an immunologic and physiologic response to inhalational antigen exposure with chest imaging, pulmonary function testing (PFT), serologic studies, bronchoscopy, and, on occasion, lung biopsy. Reexposure to the offending environment may be performed to aid in confirming the diagnosis of HP.

Chest Imaging Chest x-ray findings in HP are nonspecific and can even lack any discernible abnormalities. In cases of nonfibrotic HP, findings may be transient and can include ill-defined micronodular opacities or hazy ground-glass airspace opacities. Findings on chest x-ray will often resolve with removal from the offending antigen, although the time course of resolution may vary. With fibrotic HP, the abnormalities seen on the chest radiograph are fibrotic in nature and may be difficult to distinguish from IPF.

High-resolution computed tomography (HRCT) is a common component in the diagnostic workup for HP. Although the HRCT may be normal in nonfibrotic forms of HP, this may be due to lack of temporal correlation between exposure to the offending antigen and obtaining the imaging. Additionally, because of the transient nature of acute symptoms, HRCT is not always performed. When HRCT changes are seen in nonfibrotic HP, ground-glass airspace opacities are characteristic, as is the presence of centrilobular nodules. Expiratory images may show areas of air trapping that are likely caused by involvement of the small airways **(Fig. 299-1)**. Reticular changes and traction bronchiectasis can be observed in fibrotic HP. Subpleural honeycombing similar to that seen in IPF may be present in advanced cases, although unlike in IPF, the lung bases are frequently spared.

Pulmonary Function Testing Either restrictive or obstructive PFTs can be present in HP, so the pattern of PFT change is not useful in establishing the diagnosis of HP. However, obtaining PFTs is of use in characterizing the physiologic impairment of an individual patient and in gauging the response to antigen avoidance and/or corticosteroid therapy. Diffusion capacity for carbon monoxide may be significantly impaired, particularly in cases of fibrotic HP with pulmonary parenchymal changes.

Serum Precipitins Assaying for IgG antibodies against specific antigens, either through precipitin testing or direct serum measurement, can be a useful adjunct in the diagnosis of HP. However, the presence of an immunologic response alone is not sufficient for establishing the diagnosis, because many asymptomatic individuals with high levels of exposure to antigen may display specific IgG, as has been observed in farmers and in pigeon breeders. It should also be noted that panels that test for several specific antigens often provide false-negative results, because they represent an extremely limited proportion of the universe of potential offending environmental antigens.

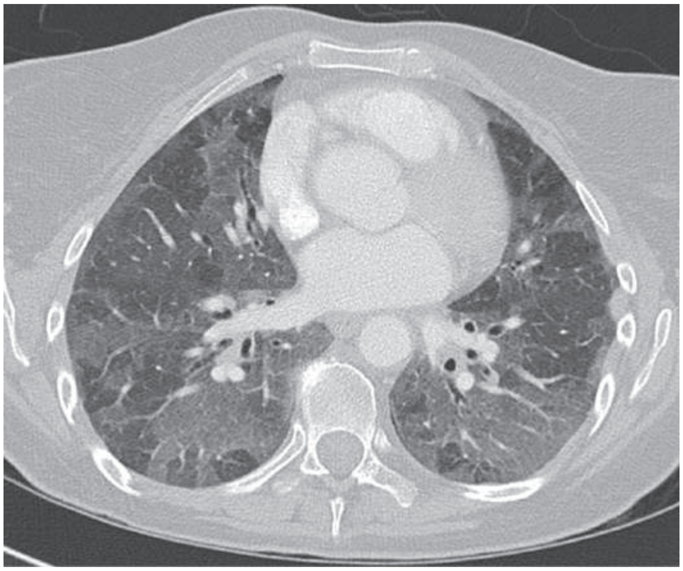

FIGURE 299-1 Chest computed tomography scan of a patient with subacute hypersensitivity pneumonitis in which scattered regions of ground-glass infiltrates in a mosaic pattern consistent with air trapping are seen bilaterally. This patient had bird fancier's lung. *(Courtesy of TJ Gross; with permission.)*

Bronchoscopy Bronchoscopy with bronchoalveolar lavage (BAL) may be used in the evaluation of HP. Although not a specific finding, BAL lymphocytosis is characteristic of HP. However, in active smokers, a lower threshold should be used to establish BAL lymphocytosis, because smoking will result in lower lymphocyte percentages. Most cases of HP have a CD4+/CD8+ lymphocyte ratio of <1, but again, this is not a specific finding and has limited utility in the diagnosis of HP.

Lung Biopsy Tissue samples may be obtained by a bronchoscopic approach using transbronchial biopsy, or more architecturally preserved specimens may be obtained by transbronchial cryobiopsy or by a surgical approach (video-assisted thoracoscopy or open approach). As is the case with BAL, histologic specimens are not absolutely necessary to establish the diagnosis of HP, but they can be useful in the correct clinical context. A common histologic feature in HP is the presence of noncaseating granulomas in the vicinity of small airways (**Fig. 299-2**). As opposed to pulmonary sarcoidosis, in which noncaseating granulomas are well defined, the granulomas seen in HP are loose and poorly defined in nature. Within the alveolar spaces and in the interstitium, a mixed cellular infiltrate with a lymphocytic predominance is observed that is frequently patchy in distribution. Bronchiolitis with the presence of organizing exudate is also often observed. Fibrosis may be present as well, particularly in fibrotic HP. Fibrotic changes may be focal but can be diffuse and severe with honeycombing in advanced cases, similar to findings in IPF.

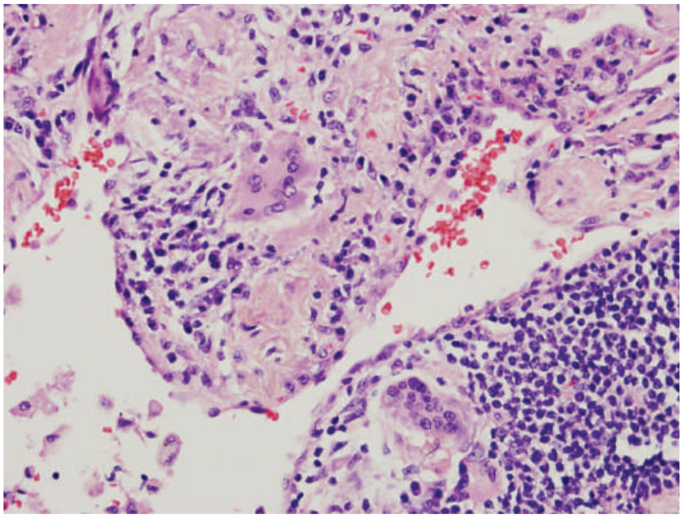

FIGURE 299-2 Open-lung biopsy from a patient with subacute hypersensitivity pneumonitis demonstrating a loose, nonnecrotizing granuloma made up of histiocytes and multinucleated giant cells. Peribronchial inflammatory infiltrate made up of lymphocytes and plasma cells is also seen. *(Courtesy of TJ Gross; with permission.)*

Clinical Prediction Rule Although not meant as a set of validated diagnostic criteria, a clinical prediction rule for predicting the presence of HP has been published by the HP Study Group. They identified six statistically significant predictors for HP, the strongest of which was exposure to an antigen known to cause HP. Other predictive criteria were the presence of serum precipitins, recurrent symptoms, symptoms occurring 4–8 h after antigen exposure, crackles on inspiration, and weight loss.

DIFFERENTIAL DIAGNOSIS

Differentiating HP from other conditions that cause a similar constellation of respiratory and systemic symptoms requires an increased index of suspicion based on obtaining a history of possible exposure to an offending antigen. Presentations of nonfibrotic HP can be mistaken for respiratory infection. In cases of fibrotic disease, HP must be differentiated from interstitial lung disease, such as IPF or nonspecific interstitial pneumonitis (NSIP); this can be a difficult task even with lung biopsy. Given the presence of pulmonary infiltrates and noncaseating granulomas on biopsy, sarcoidosis is also a consideration in the differential diagnosis of HP. Unlike in HP, however, hilar adenopathy may be prominent on chest x-ray, organs other than the lung may be involved, and noncaseating granulomas in pathologic specimens tend to be well formed. Other inhalational syndromes, such as organic toxic dust syndrome (OTDS), can be misdiagnosed as HP. OTDS occurs with exposure to organic dusts, including those produced by grains or mold silage, but neither requires prior antigen sensitization nor is characterized by positive specific IgG antibodies.

TREATMENT

Hypersensitivity Pneumonitis

The mainstay of treatment for HP is antigen avoidance, if possible. A careful exposure history must be obtained to attempt to identify the potential offending antigen and to identify the location where a patient is exposed. Once a potential antigen and location are identified, efforts should be made to modify the environment to minimize patient exposure. This may be accomplished with measures such as removal of birds, removal of molds, and improved ventilation. Personal protective equipment including respirators and ventilated helmets can be used but may not provide adequate protection for sensitized individuals. In some cases, fully avoiding specific environments may be necessary, although such a recommendation must be balanced against the effects to an individual's lifestyle or occupation. It is not uncommon for patients with HP due to exposure to household birds to be unwilling to remove them from the home.

Because nonfibrotic HP can be a self-limited disease after a discrete exposure to an offending antigen, pharmacologic therapy is often not necessary. However, in more severe nonfibrotic and fibrotic forms of the disease, there is a role for glucocorticoid therapy. In patients with particularly severe symptoms as a result of nonfibrotic HP, antigen avoidance may be insufficient after establishing the diagnosis. Although glucocorticoids do not change the long-term outcome in these patients, they can accelerate the resolution of symptoms. While there is significant variability in the approach to glucocorticoid therapy by individual clinicians, prednisone therapy can be initiated at 0.5–1 mg/kg of ideal body weight per day (not to exceed 60 mg/d or alternative glucocorticoid equivalent) over a duration of 1–2 weeks, followed by a taper over the next 2–6 weeks. In fibrotic HP, a similar trial of corticosteroids may be used, although a variable component of fibrotic disease may be irreversible. In cases with progressive lung fibrosis, antifibrotic agents such as nintedanib or pirfenidone may be considered, and in advanced cases, lung transplantation may be necessary.

GLOBAL CONSIDERATIONS

As the ever-expanding list of antigens and exposures associated with the development of HP suggests, populations at risk for HP will vary globally based on specifics of local occupational, avocational, and environmental factors. Specific examples of geographically limited HP include summer-type pneumonitis seen in Japan and suberosis seen in cork workers in Portugal and Spain.

PULMONARY INFILTRATES WITH EOSINOPHILIA

Although eosinophils are normal constituents of the lungs, there are several pulmonary eosinophilic syndromes that are characterized by pulmonary infiltrates on imaging along with an increased number of eosinophils in lung tissue, in sputum, and/or in BAL fluid, with resultant increased respiratory symptoms and the potential for systemic manifestations. Because the eosinophil plays such an important role in each of these syndromes, it is often difficult to distinguish between them, but there are important clinical and pathologic differences as well as differences in prognosis and treatment paradigms.

CLASSIFYING PULMONARY INFILTRATES WITH EOSINOPHILIA AND GENERAL APPROACH

Because there are so many different diagnoses associated with pulmonary infiltrates with eosinophilia, the first step in classifying pulmonary eosinophilic syndromes is distinguishing between primary pulmonary eosinophilic lung disorders and those with eosinophilia that are secondary to a specific cause such as a drug reaction, an infection, a malignancy, or another pulmonary condition such as asthma. Table 299-2 lists primary and secondary pulmonary eosinophilic disorders.

For each patient, a detailed history is of utmost importance and can help elucidate what the underlying disease is. Details regarding onset, timing, and precipitants of specific symptoms can help discern one diagnosis from another. History regarding pharmacologic, occupational, and environmental exposures is instructive, and family and travel history are crucial. In addition to details about the sinuses and lungs, it is important to inquire about systemic manifestations and assess for physical findings of cardiac, gastrointestinal (GI), neurologic, dermatologic, and genitourinary involvement, all of which may give clues to specific diagnoses. Once the details from history and physical are teased out, laboratory testing (including measurements of blood eosinophils, cultures, and markers of inflammation), spirometry, and radiographic imaging can help distinguish between different diseases. Often, however, BAL, transbronchial, or open lung biopsies are required. In many cases, biopsies or noninvasive diagnostic studies of other organs (e.g., echocardiogram, electromyogram, or bone marrow biopsy) can be helpful.

TABLE 299-2 Pulmonary Infiltrates with Eosinophilia

Primary Pulmonary Eosinophilic Disorders
Acute eosinophilic pneumonia
Chronic eosinophilic pneumonia
Eosinophilic granulomatosis with polyangiitis (Churg-Strauss syndrome)
Hypereosinophilic syndrome
Pulmonary Disorders of Known Cause Associated with Eosinophilia
Asthma and eosinophilic bronchitis
Allergic bronchopulmonary aspergillosis
Bronchocentric granulomatosis
Drug/toxin reaction
Infection (Table 299-4)
Parasitic/helminthic disease
Nonparasitic infection
Lung Diseases Associated with Eosinophilia
Cryptogenic organizing pneumonia
Hypersensitivity pneumonitis
Idiopathic pulmonary fibrosis
Pulmonary Langerhans cell granulomatosis
Malignant Neoplasms Associated with Eosinophilia
Leukemia
Lymphoma
Lung cancer
Adenocarcinoma of various organs
Squamous cell carcinoma of various organs
Systemic Disease Associated with Eosinophilia
Postradiation pneumonitis
Rheumatoid arthritis
Sarcoidosis
Sjögren's syndrome

PATHOPHYSIOLOGY

Pathologically, the pulmonary eosinophilic syndromes are characterized by tissue infiltration by eosinophils (Fig. 299-2). In eosinophilic granulomatosis with polyangiitis (EGPA), extravascular granulomas and necrotizing vasculitis may occur in the lungs, as well as in the heart, skin, muscle, liver, spleen, and kidneys, and may be associated with fibrinoid necrosis and thrombosis.

The exact etiology of the various pulmonary eosinophilic syndromes is unknown; however, it is felt that these syndromes result from dysregulated eosinophilopoiesis or an autoimmune process because of the prominence of allergic features and the presence of immune complexes, heightened T-cell immunity, and altered humoral immunity as evidenced by elevated IgE, IgG4, and rheumatoid factor. Because of its integral involvement in eosinophilopoiesis, interleukin 5 (IL-5) has been hypothesized to play an etiologic role. Monoclonal antibodies against IL-5 are now in clinical use for the treatment of eosinophilic asthma, eosinophilic granulomatosis with polyangiitis, and hypereosinophilic syndromes. Antineutrophil cytoplasmic antibodies (ANCAs) are present in about one-third to two-thirds of patients with EGPA; binding of ANCAs to vascular walls likely contributes to vascular inflammation and injury as well as chemotaxis of inflammatory cells.

ACUTE EOSINOPHILIC PNEUMONIA

Acute eosinophilic pneumonia is a syndrome characterized by fevers, acute respiratory failure that often requires mechanical ventilation, diffuse pulmonary infiltrates, and pulmonary eosinophilia in a previously healthy individual (Table 299-3).

Clinical Features and Etiology At presentation, acute eosinophilic pneumonia is often mistaken for acute lung injury or acute respiratory distress syndrome (ARDS), until a BAL is performed and reveals >25% eosinophils. Although the predominant symptoms of acute eosinophilic pneumonia are cough, dyspnea, malaise, myalgias, night sweats, and pleuritic chest pain, physical examination findings include high fevers, basilar rales, and rhonchi on forced expiration. Acute eosinophilic pneumonia most often affects males between age 20 and 40 with no history of asthma. Although no clear etiology has been identified, several case reports have linked acute eosinophilic pneumonia to recent initiation of tobacco smoking or even electronic cigarette inhalation (vaping), or exposure to other environmental stimuli including dust from indoor renovations.

TABLE 299-3 Diagnostic Criteria of Acute Eosinophilic Pneumonia

Acute febrile illness with respiratory manifestations of <1 month in duration
Hypoxemic respiratory failure
Diffuse pulmonary infiltrates on chest x-ray
Bronchoalveolar lavage eosinophilia >25%
Absence of parasitic, fungal, or other infection
Absence of drugs known to cause pulmonary eosinophilia
Quick clinical response to corticosteroids
Failure to relapse after discontinuation of corticosteroids

In addition to a suggestive history, the key to establishing a diagnosis of acute eosinophilic pneumonia is the presence of >25% eosinophilia on BAL fluid. While lung biopsies show eosinophilic infiltration with acute and organizing diffuse alveolar damage, it is generally not necessary to proceed to biopsy to establish a diagnosis. Although patients present with an elevated white blood cell count, in contrast to other pulmonary eosinophilic syndromes, acute eosinophilic pneumonia is often not associated with peripheral eosinophilia upon presentation. However, between 7 and 30 days of disease onset, peripheral eosinophilia often occurs with mean eosinophil counts of 1700. Erythrocyte sedimentation rate (ESR), C-reactive protein, and IgE levels are high but nonspecific, whereas HRCT is always abnormal with bilateral random patchy ground-glass or reticular opacities and small pleural effusions in as many as two-thirds of patients. Pleural fluid is characterized by a high pH with marked eosinophilia.

Clinical Course and Response to Therapy Although some patients improve spontaneously, most patients require admission to an intensive care unit and respiratory support with either invasive (intubation) or noninvasive mechanical ventilation. However, what distinguishes acute eosinophilic pneumonia from both other cases of acute lung injury as well as some of the other pulmonary eosinophilic syndromes is the absence of organ dysfunction or multisystem organ failure other than respiratory failure. One of the characteristic features of acute eosinophilic pneumonia is the high degree of corticosteroid responsiveness and the excellent prognosis. Another distinguishing feature of acute eosinophilic pneumonia is that complete clinical and radiographic recovery without recurrence or residual sequelae occurs in almost all patients within several weeks of initiation of therapy.

CHRONIC EOSINOPHILIC PNEUMONIA

In contrast to acute eosinophilic pneumonia, chronic eosinophilic pneumonia is a more indolent syndrome that is characterized by pulmonary infiltrates and eosinophilia in both the tissue and blood. Most patients are female nonsmokers with a mean age of 45, and patients do not usually develop the acute respiratory failure and significant hypoxemia appreciated in acute eosinophilic pneumonia. Similar to EGPA, a majority have asthma, with many having a history of allergies.

Patients present with a subacute illness over weeks to months, with cough, low-grade fevers, progressive dyspnea, weight loss, wheezing, malaise, and night sweats, and a chest x-ray with migratory bilateral peripheral or pleural-based opacities. Although this "photographic negative pulmonary edema" appearance on chest x-ray and chest CT is pathognomonic of chronic eosinophilic pneumonia, <25% of patients present with this finding. Other radiographic findings include atelectasis, pleural effusions, lymphadenopathy, and septal line thickening.

Almost 90% of patients have peripheral eosinophilia, with mean eosinophil counts of over 30% of total white blood cell count. BAL eosinophilia is also an important distinguishing feature with mean BAL eosinophil counts of ~60%. Both peripheral and BAL eosinophilia are very responsive to treatment with corticosteroids. Other laboratory features of chronic eosinophilic pneumonia include increased ESR, C-reactive protein, platelets, and IgE. Lung biopsy is also often not required to establish a diagnosis but may show accumulation of eosinophils and histiocytes in the lung parenchyma and interstitium, as well as cryptogenic organizing pneumonia, but with minimal fibrosis. Nonrespiratory manifestations are uncommon, but arthralgias, neuropathy, and skin and GI symptoms have all been reported; their presence may suggest EGPA or a hypereosinophilic syndrome. Another similarity is the rapid response to corticosteroids with quick resolution of peripheral and BAL eosinophilia and improvement in symptoms. In contrast to acute eosinophilic pneumonia, though, >50% of patients relapse, and many require prolonged courses of corticosteroids for months to years.

EOSINOPHILIC GRANULOMATOSIS WITH POLYANGIITIS (EGPA)

Previously known as allergic angiitis granulomatosis or Churg-Strauss syndrome, this complex syndrome is characterized by eosinophilic vasculitis that may involve multiple organ systems including the lungs, heart, skin, GI tract, and nervous system. Although EGPA is characterized by peripheral and pulmonary eosinophilia with infiltrates on chest x-ray, the primary features that distinguish EGPA from other pulmonary eosinophilic syndromes are the presence of eosinophilic vasculitis in the setting of asthma and involvement of multiple end organs (a feature it shares with hypereosinophilic syndrome). Although perceived to be quite rare, in the last few years, there has appeared to be an increased incidence of this disease, particularly in association with various asthma therapies, including leukotriene modifiers and anti-IGE therapy with omalizumab, possibly due to concurrent systemic corticosteroid withdrawal (*forme fruste* EGPA), as well as with the anti-IL-4R therapy dupilumab.

The primary features of EGPA include asthma, peripheral eosinophilia, neuropathy, pulmonary infiltrates, paranasal sinus abnormality, and presence of eosinophilic vasculitis. The mean age at diagnosis is 48 years, with a range of 14–74 years; the average length of time between diagnosis of asthma and vasculitis is 9 years. EGPA typically occurs in several phases. The prodromal phase is characterized by asthma and allergic rhinitis, and usually begins when the individual is in their twenties or thirties, typically persisting for many years. The eosinophilic infiltrative phase is characterized by peripheral eosinophilia and eosinophilic tissue infiltration of various organs including the lungs and GI tract. The third phase is the vasculitic phase and may be associated with constitutional signs and symptoms including fever, weight loss, malaise, and fatigue. This phasic progression supports the hypothesis that there is a pathophysiologic continuum between eosinophilic asthma, chronic eosinophilic asthma, and EGPA.

Similar to other pulmonary eosinophilic syndromes, constitutional symptoms are very common in EGPA and include weight loss of 10–20 lb, fevers, and diffuse myalgias and migratory polyarthralgias. Myositis may be present with evidence of vasculitis on muscle biopsies. In contrast to the eosinophilic pneumonias, EGPA involves many organ systems including the lungs, skin, nerves, heart, GI tract, and kidneys.

Symptoms and Clinical Manifestations • **RESPIRATORY** Most EGPA patients have asthma that arises later in life and in individuals who have no family history of atopy. The asthma can often be severe, and oral corticosteroids are often required to control symptoms but may lead to suppression of vasculitic symptoms. In addition to the more common symptoms of cough, dyspnea, sinusitis, and allergic rhinitis, alveolar hemorrhage and hemoptysis may also occur.

NEUROLOGIC Over three-fourths of EGPA patients have neurologic manifestations. Mononeuritis multiplex most commonly involves the peroneal nerve, but also involves the ulnar, radial, internal popliteal, and occasionally, cranial nerves. Cerebral hemorrhage and infarction may also occur and are important causes of death. Despite treatment, neurologic sequelae often do not completely resolve.

DERMATOLOGIC Approximately half of EGPA patients develop dermatologic manifestations. These include palpable purpura, skin nodules, urticarial rashes, and livedo.

CARDIOVASCULAR Granulomas, vasculitis, and widespread myocardial damage may be found on biopsy or at autopsy, and cardiomyopathy and heart failure may be seen in up to half of all patients but are often at least partially reversible. Acute pericarditis, constrictive pericarditis, myocardial infarction, and other electrocardiographic changes all may occur. The heart is a primary target organ in EGPA, and cardiac involvement often portends a worse prognosis.

GI GI symptoms are common in EGPA and likely represent an eosinophilic gastroenteritis characterized by abdominal pain, diarrhea, GI bleeding, and colitis. Ischemic bowel, pancreatitis, and cholecystitis have also been reported in association with EGPA and usually portend a worse prognosis.

RENAL Renal involvement is more common than once thought, and ~25% of patients have some degree of renal involvement. This may include proteinuria, glomerulonephritis, renal insufficiency, and rarely, renal infarct.

Lab Abnormalities Systemic eosinophilia is the hallmark laboratory finding in patients with EGPA and reflects the likely pathogenic role that the eosinophil plays in this disease. Eosinophilia >1000 cells/μL (or >10% of peripheral white blood cell [WBC] count) is one of the defining features of this illness and may be as high as 75% of the total WBC count. It is present at the time of diagnosis in >80% of patients but may respond quickly (often within 24 h) to initiation of systemic corticosteroid therapy. Even in the absence of systemic eosinophilia, tissue eosinophilia may be present.

Although not specific to EGPA, ANCAs are present in approximately one-third of patients, mostly in a perinuclear staining pattern, with specific antibodies against myeloperoxidase detected. Nonspecific lab abnormalities that may be present in patients with EGPA include a marked elevation in ESR, a normochromic normocytic anemia, an elevated IgE, hypergammaglobulinemia, and positive rheumatoid factor and antinuclear antibodies (ANA). Although BAL often reveals significant eosinophilia, this may be seen in other eosinophilic lung diseases. Similarly, PFT often reveals an obstructive defect similar to asthma.

Radiographic Features Chest x-ray abnormalities are extremely common in EGPA and consist of bilateral, nonsegmental, patchy infiltrates that often migrate and may be interstitial or alveolar in appearance. Reticulonodular and nodular disease without cavitation can be seen, as can pleural effusions and hilar adenopathy. The most common CT findings include bilateral ground-glass opacity and airspace consolidation that is predominantly subpleural. Other CT findings include bronchial wall thickening, hyperinflation, interlobular septal thickening, lymph node enlargement, and pericardial and pleural effusions. Angiography may be used diagnostically and may show signs of vasculitis in the coronary, central nervous system, and peripheral vasculature.

Treatment and Prognosis of EGPA Most patients diagnosed with EGPA have previously been diagnosed with asthma, rhinitis, and sinusitis, and have received treatment with inhaled or systemic corticosteroids. Because these agents are also the initial treatment of choice for EGPA patients, institution of these therapies in patients with EGPA who are perceived to have severe asthma may delay the diagnosis of EGPA because signs of systemic eosinophilia and vasculitis may be masked. Corticosteroids dramatically alter the course of EGPA: up to 50% of those who are untreated die within 3 months of diagnosis, whereas treated patients have a 6-year survival of >70%. Common causes of death include heart failure, cerebral hemorrhage, renal failure, and GI bleeding. Recent data suggest that clinical remission may be obtained in >90% of patients treated; ~25% of those patients may relapse, often due to corticosteroid tapering, with a rising eosinophil count heralding the relapse. Myocardial, GI, and renal involvement most often portend a poor prognosis. In such cases, induction treatment with higher doses of corticosteroids or the addition of cytotoxic agents such as cyclophosphamide or anti-CD20 therapy with rituximab is often warranted. Although survival does not differ between those treated or untreated with cyclophosphamide, cyclophosphamide is associated with a reduced incidence of relapse and an improved clinical response to treatment. Recent studies examining the efficacy of anti-IL-5 (mepolizumab) or anti-IL-5 receptor therapy (benralizumab) have shown promise, indicating that these therapies are safe and effective corticosteroid-sparing agents that can reduce relapses and induce remission. Other therapies that have been used successfully in the management of EGPA include azathioprine, methotrexate, rituximab, omalizumab, intravenous gamma globulin, and interferon α. Plasma exchange has not been shown to provide any additional benefit.

HYPEREOSINOPHILIC SYNDROMES

Hypereosinophilic syndromes (HES) constitute a heterogeneous group of disease entities manifest by persistent eosinophilia >1500 eosinophils/μL in association with end organ damage or dysfunction, in the absence of secondary causes of eosinophilia. In addition to familial, undefined, and overlap syndromes with incomplete criteria, the predominant HES subtypes are the myeloproliferative and lymphocytic variants. The myeloproliferative variants may have acquired genetic abnormalities, including of platelet-derived growth factor receptor α (PDGFRα), attributed to a constitutively activated tyrosine kinase fusion protein (Fip1L1-PDGFRα) due to a chromosomal deletion on 4q12; this variant is often responsive to imatinib. Myeloproliferative HES may also be associated with mutations involving platelet-derived growth factor β (PDGFRβ), Janus kinase 2 (JAK2), and fibroblast growth factor receptor 1 (FGFR1). Chronic eosinophilic leukemia with demonstrable cytogenetic abnormalities and/or blasts on peripheral smear is often categorized with the myeloproliferative HES. Clinical and laboratory findings in myeloproliferative HES may include dysplastic peripheral eosinophils, increased serum vitamin B_{12}, increased tryptase, anemia, thrombocytopenia, splenomegaly, bone marrow cellularity >80%, spindle-shaped mast cells, and myelofibrosis. The evaluation for lymphocytic HES includes searching for abnormal T-cell clonal populations.

Extrapulmonary Manifestations of HES More common in men than in women, HES occurs between the ages of 20 and 50 and is characterized by significant extrapulmonary involvement, including infiltration of the heart, GI tract, kidney, liver, joints, and skin. Cardiac involvement includes myocarditis and/or endomyocardial fibrosis, as well as a restrictive cardiomyopathy.

Pulmonary Manifestations of HES Similar to the other pulmonary eosinophilic syndromes, these HES are manifested by high levels of blood, BAL, and tissue eosinophilia. Lung involvement occurs in 40% of these patients and is characterized by cough and dyspnea, as well as pulmonary infiltrates. Although it is often difficult to discern the pulmonary infiltrates and effusions seen on chest x-ray from pulmonary edema resulting from cardiac involvement, computed tomography (CT) scan findings include interstitial infiltrates, ground-glass opacities, and small nodules. HES are typically not associated with ANCA. IgE may be elevated in lymphocytic HES variants.

Course and Response to Therapy Unlike the other pulmonary eosinophilic syndromes, less than half of patients with these HES respond to corticosteroids as first-line therapy. Although other treatment options include hydroxyurea, cyclosporine, and interferon, the tyrosine kinase inhibitor imatinib has emerged as an important therapeutic option for patients with the myeloproliferative variant, particularly in individuals with the *Fip1L1-PDGFRA* gene fusion. Anti-IL-5 therapy with mepolizumab has demonstrated efficacy for HES, while benralizumab also holds promise for these patients and is currently being investigated.

ALLERGIC BRONCHOPULMONARY ASPERGILLOSIS

Allergic bronchopulmonary aspergillosis (ABPA) is an eosinophilic pulmonary disorder that occurs in response to allergic sensitization to antigens from *Aspergillus* species fungi. The predominant clinical presentation of ABPA is an asthmatic phenotype, often accompanied by cough with production of brownish plugs of mucus. ABPA has also been well described as a complication of cystic fibrosis. A workup for ABPA may be beneficial in patients who carry a diagnosis of asthma but have proven refractory to usual therapy. ABPA is a distinct diagnosis from simple asthma, characterized by prominent peripheral eosinophilia and elevated circulating levels of IgE (>500 IU/mL, but often >1000 IU/mL). Establishing a diagnosis of ABPA also requires establishing sensitivity to *Aspergillus* antigens by skin test reactivity and/or direct measurement of circulating specific IgE to *Aspergillus*. Positive serum precipitins for *Aspergillus* or direct measurement of circulating specific IgG to *Aspergillus* can be used as an adjunct diagnostic criterion. Central bronchiectasis is described as a classic finding on chest imaging in ABPA but is not necessary for making a diagnosis. Other possible findings on chest imaging include patchy infiltrates and evidence of mucus impaction.

Systemic glucocorticoids may be used in the treatment of ABPA that is persistently symptomatic despite the use of inhaled therapies for asthma. Courses of glucocorticoids should be tapered over 3–6 months, and their use must be balanced against the risks of prolonged steroid therapy. Antifungal agents such as itraconazole and voriconazole given over a 4-month course reduce the antigenic stimulus in ABPA and may therefore modulate disease activity in selected patients. Newer azole agents may be used as well. The use of monoclonal antibody against IgE (omalizumab) has been described in treating severe ABPA, particularly in individuals with ABPA as a complication of cystic fibrosis. Other monoclonal antibodies used in severe eosinophilic asthma, such as those targeting IL-5 (or its receptor) or targeting IL-4-receptor-alpha have also demonstrated efficacy in small case series.

ABPA-like syndromes have been reported as a result of sensitization to several non-*Aspergillus* species fungi. However, these conditions are substantially rarer than ABPA, which may be present in a significant proportion of patients with refractory asthma.

■ INFECTIOUS PROCESSES

Infectious etiologies of pulmonary eosinophilia are largely due to helminths and are of particular importance in the evaluation of pulmonary eosinophilia in tropical environments and in the developing world **(Table 299-4)**. These infectious conditions may also be considered in recent travelers to endemic regions. Loffler syndrome refers to transient pulmonary infiltrates with eosinophilia that occurs in response to passage of helminthic larvae through the lungs, most commonly larvae of *Ascaris* species (roundworm). Symptoms are generally self-limited and may include dyspnea, cough, wheeze, and hemoptysis. Loffler syndrome may also occur in response to hookworm infection with *Ancylostoma duodenale* or *Necator americanus*. Chronic *Strongyloides stercoralis* infection can lead to recurrent respiratory symptoms with peripheral eosinophilia between flares. In immunocompromised hosts, including patients on glucocorticoids, a severe, potentially fatal, hyperinfection syndrome can result from *Strongyloides* infection. Paragonimiasis, filariasis, and visceral larval migrans can all cause pulmonary eosinophilia as well.

■ DRUGS AND TOXINS

A host of medications are associated with the development of pulmonary infiltrates with peripheral eosinophilia. Therefore, drug reaction must always be included in the differential diagnosis of pulmonary eosinophilia. Although the list of medications associated with pulmonary eosinophilia is ever expanding, common culprits include nonsteroidal anti-inflammatory medications and systemic antibiotics. Additionally, various and diverse environmental exposures such as particulate metals, scorpion stings, and inhalational drugs of abuse may also cause pulmonary eosinophilia. Radiation therapy for breast cancer has been linked with eosinophilic pulmonary infiltration as well. The mainstay of treatment is removal of the offending exposure, although glucocorticoids may be necessary if respiratory symptoms are severe.

TABLE 299-4 Infectious Causes of Pulmonary Eosinophilia

Löffler Syndrome
Ascaris
Hookworm
Schistosomiasis
Heavy Parasite Burden
Strongyloidiasis
Direct Pulmonary Penetration
Paragonimiasis
Visceral larval migrans
Immunologic Response to Organisms in Lungs
Filariasis
Dirofilariasis
Cystic Disease
Echinococcus
Cysticercosis
Other Nonparasitic
Coccidioidomycosis
Basidiobolomycosis
Paracoccidioidomycosis
Tuberculosis

Source: Adapted from P Akuthota, PF Weller: Clin Microbiol Rev 25:649, 2012.

■ GLOBAL CONSIDERATIONS

In the United States, drug-induced eosinophilic pneumonias are the most common cause of eosinophilic pulmonary infiltrates. A travel history or evidence of recent immigration should prompt the consideration of parasite-associated disorders. Tropical eosinophilia is usually caused by filarial infection; however, eosinophilic pneumonias also occur with other parasites such as *Ascaris* spp., *Ancylostoma* spp., *Toxocara* spp., and *Strongyloides stercoralis*. Tropical eosinophilia due to *Wuchereria bancrofti* or *Wuchereria malayi* occurs most commonly in southern Asia, Africa, and South America and is treated successfully with diethylcarbamazine. In the United States, *Strongyloides* is endemic to the southeastern and Appalachian regions.

■ FURTHER READING

Akuthota P, Weller PF: Eosinophilic pneumonias. Clin Microbiol Rev 25:649, 2012.

Fernández Pérez ER et al: Diagnosis and evaluation of hypersensitivity pneumonitis: CHEST guideline and expert panel report. Chest 160:e97, 2021.

Khoury P et al: HES and EGPA: Two sides of the same coin. Mayo Clin Proc 98:1054, 2023.

Wechsler ME et al: Mepolizumab or placebo for eosinophilic granulomatosis with polyangiitis. N Engl J Med 376:1921, 2017.

Wechsler ME et al: Benralizumab versus mepolizumab for eosinophilic granulomatosis with polyangiitis. N Engl J Med 390:911, 2024.

300

Occupational and Environmental Lung Disease

John R. Balmes, Mehrdad Arjomandi

Occupational and environmental lung diseases are difficult to distinguish from those of nonenvironmental origin. Virtually all major categories of pulmonary disease can be caused by environmental agents, and environmentally related disease usually presents clinically in a manner indistinguishable from that of disease not caused by such agents. In addition, the etiology of many diseases may be multifactorial; occupational and environmental factors may interact with other factors (such as smoking and genetic risk). It is often only after a careful exposure history is taken that the underlying workplace or general environmental exposure is uncovered.

Why is knowledge of occupational or environmental etiology so important? Patient management and prognosis are affected significantly by such knowledge. For example, patients with occupational asthma or hypersensitivity pneumonitis often cannot be managed adequately without cessation of exposure to the offending agent. Establishment of cause may have significant legal and financial implications for a patient who no longer can work in their usual job. Other exposed people may be identified as having the disease or prevented from getting it. In addition, new associations between exposure and disease may be identified (e.g., nylon flock worker's lung disease, diacetyl-induced bronchiolitis obliterans, military burn pit-related constrictive bronchiolitis).

Although the exact proportion of lung disease due to occupational and environmental factors is unknown, a large number of individuals are at risk. For example, 15–20% of the burden of adult asthma and chronic obstructive pulmonary disease (COPD) has been estimated to be due to occupational factors.

HISTORY AND EXPOSURE ASSESSMENT

The patient's history is of paramount importance in assessing any potential occupational or environmental exposure. Inquiry into specific work practices should include questions about the specific contaminants involved, the presence of visible dust, chemical odors, the size and ventilation of workspaces, the use of respiratory protective equipment, and whether coworkers have similar complaints. The temporal association of exposure at work and symptoms may provide clues to occupation-related disease. In addition, the patient must be questioned about alternative sources of exposure to potentially toxic agents, including hobbies, home characteristics, exposure to secondhand tobacco smoke, and proximity to traffic or industrial facilities. Short-term and long-term exposures to potential toxic agents in the distant past also must be considered.

In the United States, workers have the right to know about potential hazards in their workplaces under federal Occupational Safety and Health Administration (OSHA) regulations. Employers must provide specific information about potential hazardous agents in products being used through Safety Data Sheets as well as training in personal protective equipment and environmental control procedures. However, the introduction of new processes and/or new chemical compounds may change exposure significantly, and often only the employee on the production line is aware of the change. For the physician caring for a patient with a suspected work-related illness, a visit to the work site can be very instructive. Alternatively, an affected worker can request an inspection by OSHA. If reliable environmental sampling data are available, that information should be used in assessing a patient's exposure. Because chronic diseases may result from exposure over many years, current environmental measurements should be combined with work histories to arrive at estimates of past exposure.

LABORATORY TESTS

Exposures to inorganic and organic dust can cause interstitial lung disease that presents with a restrictive pattern and a decreased diffusing capacity **(Chap. 296)**. Similarly, exposure to various dusts or chemical agents may result in occupational asthma or COPD that is characterized by airway obstruction. Measurement of change in forced expiratory volume in 1 second (FEV_1) before and after a working shift can be used to detect an acute bronchoconstrictive response.

The chest radiograph is useful in detecting and monitoring the pulmonary response to mineral dusts, certain metals, and organic dusts capable of inducing hypersensitivity pneumonitis. The International Labour Organisation (ILO) International Classification of Radiographs of Pneumoconioses classifies chest radiographs by the nature and size of opacities seen and the extent of involvement of the parenchyma. In general, small, rounded opacities are seen in silicosis or coal worker's pneumoconiosis, and small linear opacities are seen in asbestosis. Although useful for epidemiologic studies and screening large numbers of workers, the ILO system can be problematic when applied to an individual worker's chest radiograph. With dust causing rounded opacities, the degree of involvement on the chest radiograph may be extensive, whereas pulmonary function may be only minimally impaired. In contrast, in pneumoconiosis causing linear, irregular opacities like those seen in asbestosis, the radiograph may lead to underestimation of the severity of the impairment until relatively late in the disease. For patients with a history of dust exposure, conventional computed tomography (CT) is more sensitive for the detection of lung opacities and pleural thickening, and high-resolution CT (HRCT) improves the detection of interstitial changes.

Other procedures that may be of use in identifying the role of environmental exposures in causing lung disease include skin prick testing or specific immunoglobulin type E (IgE) antibody titers for evidence of immediate hypersensitivity to agents capable of inducing occupational asthma (e.g., flour antigens in bakers), specific immunoglobulin type G (IgG) precipitating antibody titers for agents capable of causing hypersensitivity pneumonitis (e.g., pigeon antigen in bird handlers), and assays for specific cell-mediated immune responses (e.g., beryllium lymphocyte proliferation testing in nuclear workers). Sometimes a bronchoscopy to obtain transbronchial biopsies of lung tissue may be required for histologic diagnosis (chronic beryllium disease [CBD]). Rarely, video-assisted thoracoscopic surgery to obtain a larger sample of lung tissue may be required to determine the specific diagnosis of environmentally induced lung disease (hypersensitivity pneumonitis, constrictive bronchiolitis, or giant cell interstitial pneumonitis due to cobalt exposure).

DETERMINANTS OF INHALATIONAL EXPOSURE

The chemical and physical characteristics of inhaled agents affect both the dose and the site of deposition in the respiratory tract. Water-soluble gases such as ammonia and sulfur dioxide are absorbed in the lining fluid of the upper and proximal airways and thus tend to produce irritative and bronchoconstrictive responses. In contrast, nitrogen dioxide and phosgene, which are less soluble, may penetrate to the bronchioles and alveoli in sufficient quantities to produce acute chemical pneumonitis.

Particle size of air contaminants must also be considered. Because of their settling velocities in air, particles >10–15 μm in diameter do not penetrate beyond the nose and throat. Particles <10 μm in size are deposited below the larynx. These particles are divided into three size fractions on the basis of their size characteristics and sources. Particles ~2.5–10 μm (coarse-mode fraction) contain crustal elements such as silica, aluminum, and iron. These particles mostly deposit relatively high in the tracheobronchial tree. Although the total mass of an ambient sample is dominated by these larger respirable particles, the number of particles, and therefore the surface area on which potential toxic agents can deposit and be carried to the lower airways, is dominated by particles <2.5 μm (fine-mode fraction). These fine particles are created primarily by the burning of fossil fuels or high-temperature industrial processes resulting in condensation products from gases, fumes, or vapors. The smallest particles, those <0.1 μm in size, represent the ultrafine fraction and make up the largest number of particles; they tend to remain in the airstream and deposit in the lung only on a random basis as they come into contact with the alveolar walls. If they do deposit, however, particles of this size range may penetrate into the circulation and be carried to extrapulmonary sites. New technologies create particles of this size ("nanoparticles") for use in many commercial applications. Besides the size characteristics of particles and the solubility of gases, the actual chemical composition, mechanical properties, and immunogenicity or infectivity of inhaled material determine in large part the nature of the diseases found among exposed persons.

OCCUPATIONAL EXPOSURES AND PULMONARY DISEASE

Table 300-1 provides broad categories of exposure in the workplace and diseases associated with chronic exposure in those industries.

ASBESTOS-RELATED DISEASES

Asbestos is a generic term for several different mineral silicates, including chrysolite, amosite, anthophyllite, and crocidolite. In addition to workers involved in the production of asbestos products (mining, milling, and manufacturing), many workers in the shipbuilding and construction trades, including pipe fitters and boilermakers, were occupationally exposed because asbestos was widely used during the twentieth century for its thermal and electrical insulation properties. Asbestos also was used in the manufacture of fire-resistant textiles, in cement and floor tiles, and in friction materials such as brake and clutch linings.

Exposure to asbestos is not limited to persons who directly handle the material. Cases of asbestos-related diseases have been encountered in individuals with only bystander exposure, such as painters and electricians who worked alongside insulation workers in a shipyard. Community exposure resulted from the use of asbestos-containing mine and mill tailings as landfill, road surface, and playground material (e.g., Libby, Montana, the site of a vermiculite mine in which the

TABLE 300-1 Categories of Occupational Exposure and Associated Respiratory Conditions

OCCUPATIONAL EXPOSURES	NATURE OF RESPIRATORY RESPONSES	COMMENT
Inorganic Dusts		
Asbestos: mining, processing, construction, ship repair	Fibrosis (asbestosis), pleural disease, cancer, mesothelioma	Virtually all new mining and construction with asbestos done in developing countries
Silica: mining, stone cutting, sandblasting, quarrying, artificial stone manufacture and installation	Fibrosis (silicosis), progressive massive fibrosis (PMF), cancer, tuberculosis, chronic obstructive pulmonary disease (COPD)	Improved protection in United States; persistent risk in developing countries
Coal dust: mining	Fibrosis (coal worker's pneumoconiosis), PMF, COPD	Risk persists in certain areas of United States, increasing in countries where new mines open
Beryllium: processing alloys for nuclear power and weapons, aerospace, and electronics	Acute pneumonitis (rare), chronic granulomatous disease, lung cancer (highly suspect)	Risk in high-tech industries persists
Other metals: aluminum, chromium, cobalt, nickel, titanium, tungsten carbide, or "hard metal" (contains cobalt)	Wide variety of conditions from acute pneumonitis to lung cancer and asthma	New diseases appear with new process development
Organic Dusts		
Cotton dust: milling, processing	Byssinosis (an asthma-like syndrome), chronic bronchitis, COPD	Increasing risk in developing countries with drop in United States as jobs shift overseas
Grain dust: elevator agents, dock workers, milling, bakers	Asthma, chronic bronchitis, COPD	Risk shifting more to migrant labor pool
Other agricultural dusts: fungal spores, vegetable products, insect fragments, animal dander, bird and rodent feces, endotoxins, microorganisms, pollens	Hypersensitivity pneumonitis (farmer's lung), asthma, chronic bronchitis	Important in migrant labor pool but also resulting from in-home exposures
Toxic chemicals: wide variety of industries; see Table 300-2	Asthma, chronic bronchitis, COPD, hypersensitivity pneumonitis, pneumoconiosis, and cancer	Reduced risk with recognized hazards; increasing risk for developing countries where controlled labor practices are less stringent
Other Environmental Agents		
Uranium and radon daughters, secondhand tobacco smoke, polycyclic aromatic hydrocarbons (PAHs), biomass smoke, diesel exhaust, welding fumes, wood finishing	Occupational exposures estimated to contribute to up to 10% of all lung cancers; chronic bronchitis, COPD, and fibrosis	In-home exposures important; in developing countries, biomass smoke is a major risk factor for COPD among women in these countries

ore was contaminated with asbestos). Finally, exposure can occur from the disturbance of naturally occurring asbestos (e.g., from increasing residential development in the foothills of the Sierra Mountains in California).

Asbestos has largely been replaced in the developed world with synthetic mineral fibers such as fiberglass and refractory ceramic fibers, but it continues to be used in the developing world. The major health effects from exposure to asbestos are pleural and pulmonary fibrosis, cancers of the respiratory tract, and pleural and peritoneal mesothelioma.

Asbestosis is a diffuse interstitial fibrosing disease of the lung that is directly related to the intensity and duration of exposure. The disease resembles other forms of diffuse interstitial fibrosis **(Chap. 304)**. Usually, exposure has taken place for at least 10 years before the disease becomes manifest. The mechanisms by which asbestos fibers induce lung fibrosis are not completely understood but are known to involve oxidative injury due to the generation of reactive oxygen species by the transition metals on the surface of the fibers as well as from cells engaged in phagocytosis.

Past exposure to asbestos is specifically indicated by pleural plaques on chest radiographs, which are characterized by either thickening or calcification along the parietal pleura, particularly along the lower lung fields, the diaphragm, and the cardiac border. Without additional manifestations, pleural plaques imply only exposure, not pulmonary impairment. Benign pleural effusions also may occur.

Irregular or linear opacities that usually are first noted in the lower lung fields are the chest radiographic hallmark of asbestosis. An indistinct heart border or a "ground-glass" appearance in the lung fields may be seen. HRCT may show distinct changes of subpleural curvilinear lines 5–10 mm in length that appear to be parallel to the pleural surface **(Fig. 300-1)**.

Pulmonary function testing in asbestosis reveals a restrictive pattern with a decrease in both lung volumes and diffusing capacity. There may also be evidence of mild airflow obstruction (due to peribronchiolar fibrosis).

Because no specific therapy is available for asbestosis, supportive care is the same as that given to any patient with diffuse interstitial fibrosis of any cause. In general, newly diagnosed cases will have resulted from exposures that occurred many years before.

Lung cancer **(Chap. 83)** is the most common cancer associated with asbestos exposure. The excess frequency of lung cancer (all histologic types) in asbestos workers is associated with a minimum latency of 15–19 years between first exposure and development of the disease. Persons with more exposure are at greater risk of disease. In addition, there is a significant interactive effect of smoking and asbestos exposure that results in greater risk than what would be expected from the additive effect of each factor.

Mesotheliomas **(Chap. 305)**, both pleural and peritoneal, are also associated with asbestos exposure. In contrast to lung cancers, these tumors do not appear to be associated with smoking. Relatively short-term asbestos exposures of ≤1–2 years, occurring up to 40 years in the past, have been associated with the development of mesotheliomas (an observation that emphasizes the importance of obtaining a complete environmental exposure history). Although the risk of mesothelioma is much less than that of lung cancer among asbestos-exposed workers, ~3000 cases per year are diagnosed in the United States.

Because epidemiologic studies have shown that >80% of mesotheliomas may be associated with asbestos exposure, documented mesothelioma in a patient with occupational or environmental exposure to asbestos may be compensable.

■ SILICOSIS

Despite being one of the oldest known occupational pulmonary hazards, *free silica* (SiO_2), or crystalline quartz, is still a major cause of disease. The major occupational exposures include mining; stonecutting; sand blasting; glass and cement manufacturing; foundry work; packing of silica flour; and quarrying, particularly of granite. Most often, pulmonary fibrosis due to silica exposure (silicosis) occurs in a dose-response fashion after many years of exposure. Two recent outbreaks of silicosis have involved sandblasting of denim jeans to make them look "used" and manufacture and installation of artificial stone ("faux granite") kitchen countertops.

Workers heavily exposed through sandblasting in confined spaces, tunneling through rock with a high quartz content (15–25%), or the

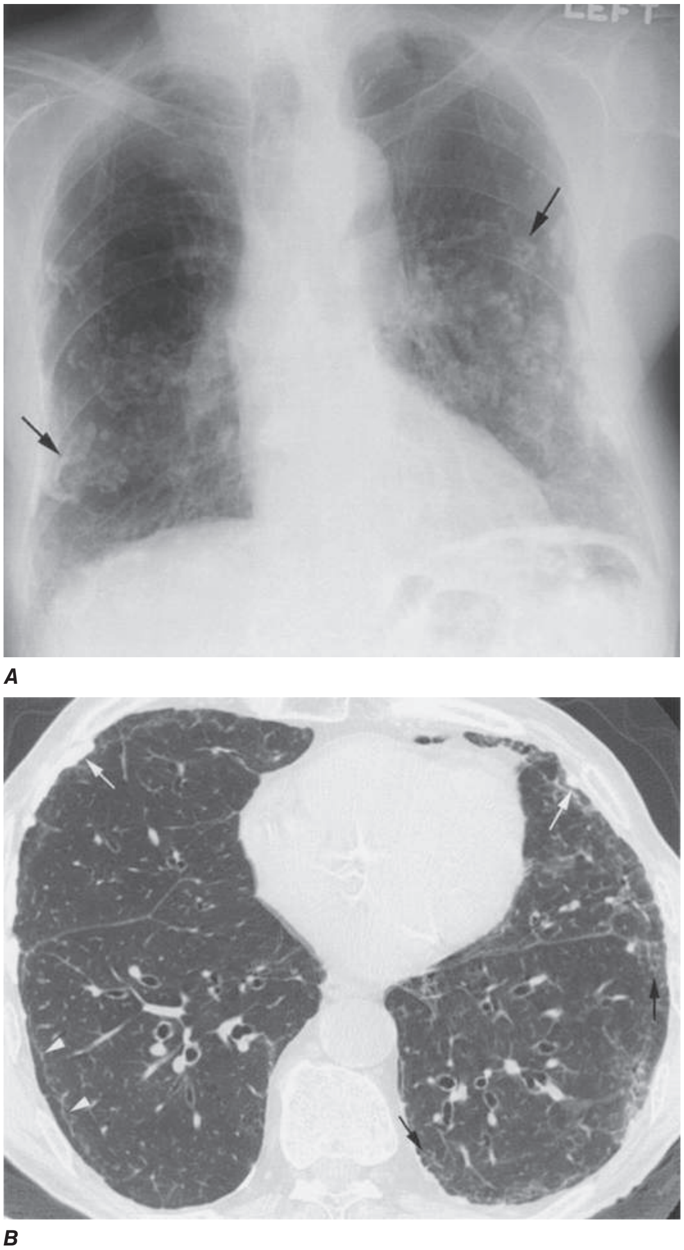

FIGURE 300-1 Asbestosis. ***A.*** Frontal chest radiograph shows bilateral calcified pleural plaques consistent with asbestos-related pleural disease. Poorly defined linear and reticular abnormalities are seen in the lower lobes bilaterally. ***B.*** Axial high-resolution computed tomography of the thorax obtained through the lung bases shows bilateral, subpleural reticulation (*black arrows*), representing fibrotic lung disease due to asbestosis. Subpleural lines are also present (*arrowheads*), characteristic of, though not specific for, asbestosis. Calcified pleural plaques representing asbestos-related pleural disease (*white arrows*) are also evident.

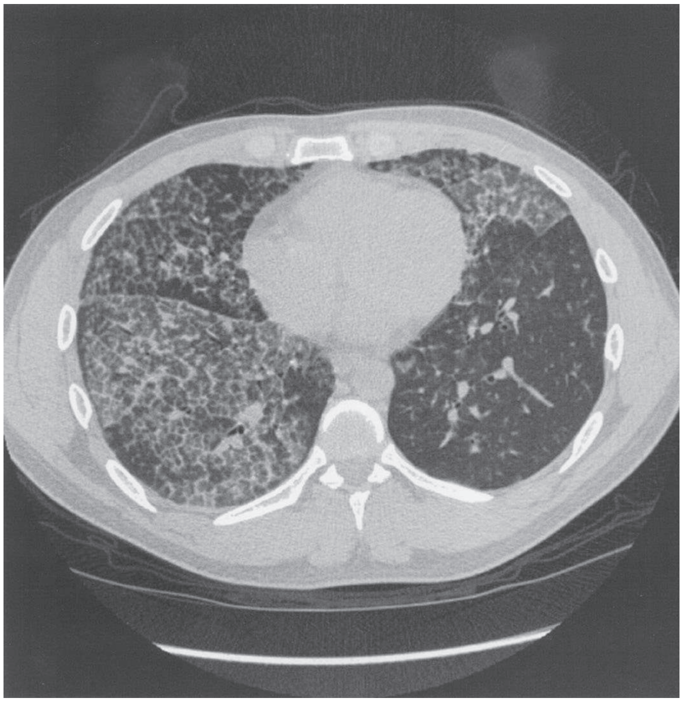

FIGURE 300-2 Acute silicosis. This high-resolution computed tomography scan shows multiple small nodules consistent with silicosis but also diffuse ground-glass densities with thickened intralobular and interlobular septa producing polygonal shapes. This has been referred to as "crazy paving."

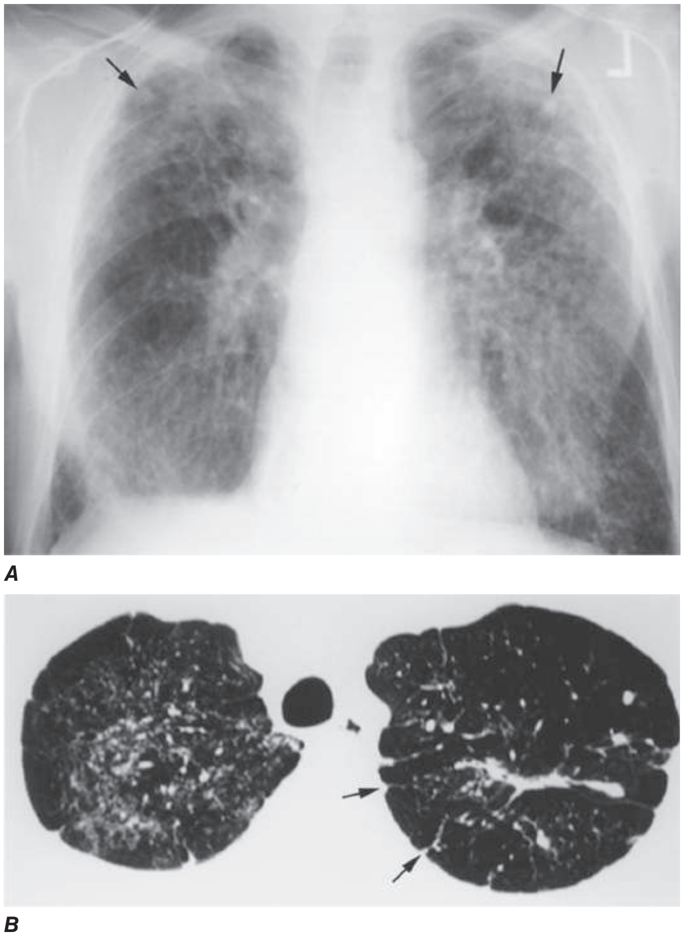

FIGURE 300-3 Chronic silicosis. ***A.*** Frontal chest radiograph in a patient with silicosis shows variably sized, poorly defined nodules (*arrows*) predominating in the upper lobes. ***B.*** Axial thoracic computed tomography image through the lung apices shows numerous small nodules, more pronounced in the right upper lobe. A number of the nodules are subpleural in location (*arrows*).

manufacture of artificial stone countertops may develop acute silicosis with only months of exposure. The clinical and pathologic features of acute silicosis are similar to those of pulmonary alveolar proteinosis **(Chap. 304)**. The chest radiograph may show profuse miliary infiltration or consolidation, and a characteristic HRCT pattern known as "crazy paving" could be present **(Fig. 300-2)**. The disease may be quite severe and progressive despite the discontinuation of exposure. Whole-lung lavage may provide symptomatic relief and slow the progression.

With long-term, less intense exposure, small rounded opacities in the upper lobes may appear on the chest radiograph after 15–20 years of exposure, usually without associated impairment of lung function (*simple silicosis*). Calcification of hilar nodes may occur in as many as 20% of cases and produces a characteristic "eggshell" pattern. Silicotic nodules may be identified more readily by HRCT **(Fig. 300-3)**. The nodular fibrosis may be progressive in the absence of further exposure, with coalescence and formation of nonsegmental conglomerates of irregular masses >1 cm in diameter (*complicated silicosis*). These masses can become quite large, and when this occurs, the term *progressive massive fibrosis* (PMF) is applied. Significant functional impairment with both restrictive and obstructive components may be associated with PMF.

Because silica causes alveolar macrophage dysfunction, patients with silicosis are at greater risk of acquiring lung infections that involve these cells as a primary defense (*Mycobacterium tuberculosis*, atypical mycobacteria, and fungi). Because of the increased risk of active tuberculosis, the recommended treatment of latent tuberculosis in these patients is longer. Silica has immunoadjuvant properties, and another potential clinical complication of silicosis is autoimmune connective tissue disorders such as rheumatoid arthritis and scleroderma. In addition, there are sufficient epidemiologic data that the International Agency for Research on Cancer lists silica as a probable lung carcinogen.

Other, less hazardous silicates include fuller's earth, kaolin, mica, diatomaceous earths, silica gel, soapstone, carbonate dusts, and cement dusts. The production of fibrosis in workers exposed to these agents is believed to be related either to the free silica content of these dusts or, for substances that contain no free silica, to the potentially large dust loads to which these workers may be exposed. Some silicates, including *talc* and *vermiculite*, may be contaminated with asbestos. Fibrosis of lung or pleura, lung cancer, and mesothelioma have been associated with chronic exposure to talc and vermiculite dusts.

COAL WORKER'S PNEUMOCONIOSIS (CWP)

Occupational exposure to *coal dust* can lead to CWP, which has enormous social, economic, and medical significance in every nation in which coal mining is an important industry. Simple radiographically identified CWP is seen in ~10% of all coal miners and in as many as 50% of anthracite miners with >20 years of work on the coal face. The prevalence of disease is lower in workers in bituminous coal mines.

With prolonged exposure to coal dust (i.e., 15–20 years), small, rounded opacities similar to those of silicosis may develop. As in silicosis, the presence of these nodules (*simple CWP*) usually is not associated with pulmonary impairment. In addition to CWP, coal dust can cause chronic bronchitis and COPD **(Chap. 303)**. The effects of coal dust are additive to those of cigarette smoking.

Complicated CWP is manifested by the appearance on the chest radiograph of nodules ≥1 cm in diameter generally confined to the upper half of the lungs. As in silicosis, this condition can progress to PMF that is accompanied by severe lung function deficits and associated with premature mortality. Silica is often present in anthracitic coal dust, and its presence may contribute to risk of PMF. Due to increased mechanization and narrower veins of coal with more silica contamination of mine dust, cases of PMF are occurring in the Appalachian coal belt at an alarming rate.

Caplan syndrome **(Chap. 370)**, first described in coal miners but subsequently in patients with silicosis, is the combination of pneumoconiotic nodules and seropositive rheumatoid arthritis.

CHRONIC BERYLLIUM DISEASE

Beryllium is a lightweight metal with tensile strength, good electrical conductivity, and value in the control of nuclear reactions through its ability to quench neutrons. Although beryllium may produce an acute pneumonitis, it is far more commonly associated with a chronic granulomatous inflammatory disease that is similar to sarcoidosis **(Chap. 379)**. Unless one inquires specifically about occupational exposures to beryllium in the manufacture of alloys, ceramics, or high-technology electronics in a patient with sarcoidosis, one may miss entirely the etiologic relationship to the occupational exposure. Combat-related embedded metal fragments (shrapnel) in military veterans may also contain beryllium and thus be a source of exposure to the metal. What distinguishes chronic beryllium disease (CBD) from sarcoidosis is evidence of a specific cell-mediated immune response (i.e., delayed hypersensitivity) to beryllium.

The test that usually provides this evidence is the beryllium lymphocyte proliferation test (BeLPT). The BeLPT compares the in vitro proliferation of lymphocytes from blood or bronchoalveolar lavage in the presence of beryllium salts with that of unstimulated cells. Proliferation is usually measured by lymphocyte uptake of radiolabeled thymidine.

Chest imaging findings are similar to those of sarcoidosis (nodules along septal lines) except that hilar adenopathy is somewhat less common. As with sarcoidosis, pulmonary function test results may show restrictive and/or obstructive ventilatory deficits and decreased diffusing capacity. With early disease, both chest imaging studies and pulmonary function tests may be normal. Fiberoptic bronchoscopy with transbronchial lung biopsy usually is required to make the diagnosis of CBD. In a beryllium-sensitized individual, the presence of noncaseating granulomas or monocytic infiltration in lung tissue establishes the diagnosis. Accumulation of beryllium-specific CD4+ T cells occurs in the granulomatous inflammation seen on lung biopsy. Susceptibility to CBD is highly associated with human leukocyte antigen DP (HLA-DP) alleles that have a glutamic acid in position 69 of the β chain.

OTHER METALS

Aluminum and titanium dioxide have been rarely associated with a sarcoid-like reaction in lung tissue. Exposure to dust containing tungsten carbide, also known as "hard metal," may produce giant cell interstitial pneumonitis. Cobalt is a constituent of tungsten carbide and is the likely etiologic agent of both the interstitial pneumonitis and the occupational asthma that may occur. The most common exposures to tungsten carbide occur in tool and dye, saw blade, and drill bit manufacture. Diamond polishing may also involve exposure to cobalt dust. In patients with interstitial lung disease, one should always inquire about exposure to metal fumes and/or dusts. Especially when sarcoidosis appears to be the diagnosis, one should always consider possible CBD.

OTHER INORGANIC DUSTS

Most of the inorganic dusts discussed thus far are associated with the production of either dust macules or interstitial fibrotic changes in the lung. Other inorganic and organic dusts (see categories in Table 300-1), along with some of the dusts previously discussed, are associated with chronic mucus hypersecretion (chronic bronchitis), with or without reduction of expiratory flow rates. Cigarette smoking is the major cause of these conditions, and any effort to attribute some component of the disease to occupational and environmental exposures must take cigarette smoking into account. Most studies suggest an additive effect of dust exposure and smoking. The pattern of the irritant dust effect is similar to that of cigarette smoking, suggesting that small airway inflammation may be the initial site of pathologic response in those cases and continued exposure may lead to chronic bronchitis and COPD.

ORGANIC DUSTS

Some of the specific diseases associated with organic dusts are discussed in detail in the chapters on asthma **(Chap. 298)** and hypersensitivity pneumonitis **(Chap. 299)**. Many of these diseases are named for the specific setting in which they are found, e.g., farmer's lung, malt worker's disease, and mushroom worker's disease. Often the temporal relation of symptoms to exposure furnishes the best evidence for the diagnosis. Three occupational exposures are singled out for discussion here because they affect the largest proportions of workers.

Cotton Dust (Byssinosis) Workers occupationally exposed to cotton dust (but also to flax, hemp, or jute dust) in the production of yarns for textiles and rope making are at risk for an asthma-like syndrome known as byssinosis. The risk of byssinosis is associated with both cotton dust and endotoxin levels in the workplace environment.

Byssinosis is characterized clinically as occasional (early-stage) and then regular (late-stage) chest tightness toward the end of the first day of the workweek ("Monday chest tightness"). Exposed workers may show a significant drop in FEV_1 over the course of a Monday work shift. Initially the symptoms do not recur on subsequent days of the week, but in a subset of workers, chest tightness may recur or persist throughout the workweek. After >10 years of exposure, workers with recurrent symptoms are more likely to have an obstructive pattern on pulmonary function testing.

Dust exposure can be reduced by the use of exhaust hoods, general increases in ventilation, and wetting procedures, but respiratory protective equipment may be required during certain operations. Regular surveillance of pulmonary function in cotton dust–exposed workers using spirometry before and after the work shift is required by OSHA. All workers with persistent symptoms or significantly reduced levels of pulmonary function should be moved to areas of lower risk of exposure.

Grain Dust Worldwide, many farmers and workers in grain storage facilities are exposed to grain dust. The presentation of obstructive airway disease in grain dust–exposed workers is virtually identical to the characteristic findings in cigarette smokers, i.e., persistent cough, mucus hypersecretion, wheeze and dyspnea on exertion, and reduced FEV_1 and FEV_1/FVC (forced vital capacity) ratio (Chap. 296).

Dust concentrations in grain elevators vary greatly but can be >10,000 μg/m^3 with many particles in the respirable size range. The effect of grain dust exposure is additive to that of cigarette smoking, with ~50% of workers who smoke having symptoms. Smoking grain dust–exposed workers are more likely to have obstructive ventilatory deficits on pulmonary function testing. As in byssinosis, endotoxin may play a role in grain dust–induced chronic bronchitis and COPD.

Farmer's Lung This condition results from exposure to moldy hay containing spores of thermophilic actinomycetes that produce a hypersensitivity pneumonitis (Chap. 299). A patient with acute farmer's lung presents 4–8 h after exposure with fever, chills, malaise, cough, and dyspnea without wheezing. The history of exposure is obviously essential to distinguish this disease from influenza or pneumonia with similar symptoms. In the chronic form of the disease, the history of repeated attacks after similar exposure is important in differentiating this syndrome from other causes of patchy fibrosis (e.g., sarcoidosis).

A wide variety of other organic dusts are associated with the occurrence of hypersensitivity pneumonitis (Chap. 299). For patients who present with hypersensitivity pneumonitis, specific and careful inquiry about occupations, hobbies, and other home environmental exposures is necessary to uncover the source of the etiologic agent.

TOXIC CHEMICALS

Exposure to toxic chemicals affecting the lung generally involves gases and vapors. A common accident is one in which the victim is trapped in a confined space where the chemicals have accumulated to harmful levels. In addition to the specific toxic effects of the chemical, the victim often sustains considerable anoxia, which can play a dominant role in determining whether the individual survives.

Table 300-2 lists a variety of toxic agents that can produce acute and sometimes life-threatening reactions in the lung. All these agents in sufficient concentrations have been demonstrated, at least in animal studies, to affect the lower airways and disrupt alveolar architecture, either acutely or as a result of chronic exposure.

Firefighters and fire victims are at risk of *smoke inhalation*, an important cause of acute cardiorespiratory failure. Smoke inhalation kills more fire victims than does thermal injury. Carbon monoxide poisoning with resulting significant hypoxemia can be life-threatening (Chap. 470).

TABLE 300-2 Selected Common Toxic Chemical Agents That Affect the Lung

AGENT(S)	SELECTED EXPOSURES	ACUTE EFFECTS FROM HIGH OR ACCIDENTAL EXPOSURE	CHRONIC EFFECTS FROM RELATIVELY LOW EXPOSURE
Acid anhydrides	Manufacture of resin esters, polyester resins, thermoactivated adhesives	Nasal irritation, cough	Asthma, chronic bronchitis, hypersensitivity pneumonitis
Acid fumes: H_2SO_4, HNO_3	Manufacture of fertilizers, chlorinated organic compounds, dyes, explosives, rubber products, metal etching, plastics	Mucous membrane irritation, followed by chemical pneumonitis 2–3 days later	Bronchitis and suggestion of mildly reduced pulmonary function in children with lifelong residential exposure to high levels
Acrolein and other aldehydes	By-product of burning plastics, woods, tobacco smoke	Mucous membrane irritant, decrease in lung function	Upper respiratory tract irritation
Ammonia	Refrigeration; petroleum refining; manufacture of fertilizers, explosives, plastics, and other chemicals	Same as for acid fumes, but bronchiectasis also has been reported	Upper respiratory tract irritation, chronic bronchitis
Cadmium fumes	Smelting, soldering, battery production	Mucous membrane irritant, acute respiratory distress syndrome (ARDS)	Chronic obstructive pulmonary disease (COPD)
Formaldehyde	Manufacture of resins, leathers, rubber, metals, and woods; laboratory workers, embalmers; emission from urethane foam insulation	Same as for acid fumes	Nasopharyngeal cancer
Halides and acid salts (Cl, Br, F)	Bleaching in pulp, paper, textile industry; manufacture of chemical compounds; synthetic rubber, plastics, disinfectant, rocket fuel, gasoline	Mucous membrane irritation, pulmonary edema; possible reduced forced vital capacity (FVC) 1–2 years after exposure	Upper respiratory tract irritation, epistaxis, tracheobronchitis
Hydrogen sulfide	By-product of many industrial processes, oil, other petroleum processes and storage	Increase in respiratory rate followed by respiratory arrest, lactic acidosis, pulmonary edema, death	Conjunctival irritation, chronic bronchitis, recurrent pneumonitis
Isocyanates (TDI, HDI, MDI)	Production of polyurethane foams, plastics, adhesives, surface coatings	Mucous membrane irritation, dyspnea, cough, wheeze, pulmonary edema	Upper respiratory tract irritation, cough, asthma, hypersensitivity pneumonitis, reduced lung function
Nitrogen dioxide	Silage, metal etching, explosives, rocket fuels, welding, by-product of burning fossil fuels	Cough, dyspnea, pulmonary edema may be delayed 4–12 h; possible result from acute exposure: bronchiolitis obliterans in 2–6 weeks	Emphysema in animals, chronic bronchitis, associated with reduced lung function growth in children with lifelong residential exposure
Ozone	Arc welding, flour bleaching, deodorizing, emissions from copying equipment, photochemical air pollutant	Mucous membrane irritant, reduced pulmonary function transiently in children and adults, asthma exacerbation	Excess cardiopulmonary mortality rates, increased risk for new-onset asthma in children
Phosgene	Organic compound, metallurgy, volatilization of chlorine-containing compounds	Delayed onset of bronchiolitis and pulmonary edema	Chronic bronchitis
Sulfur dioxide	Manufacture of sulfuric acid, bleaches, coating of nonferrous metals, food processing, refrigerant, burning of fossil fuels, wood pulp industry	Mucous membrane irritant, epistaxis, bronchospasm (especially in people with asthma)	Chronic bronchitis

Abbreviations: HDI, hexamethylene diisocyanate; MDI, methylene diphenyl diisocyanate; TDI, toluene diisocyanate.

Synthetic materials (plastic, polyurethanes), when burned, may release a variety of other toxic agents (such as cyanide and hydrochloric acid), and this must be considered in evaluating smoke inhalation victims. Exposed victims may have some degree of lower respiratory tract inflammation and/or pulmonary edema.

Exposure to certain highly reactive, low-molecular-weight agents used in the manufacture of synthetic polymers, paints, and coatings (*diisocyanates* in polyurethanes, *aromatic amines* and *acid anhydrides* in epoxies) is associated with a high risk of occupational asthma. Although this occupational asthma manifests clinically as if sensitization has occurred, an IgE antibody–mediated mechanism is not necessarily involved. Hypersensitivity pneumonitis–like reactions also have been described in diisocyanate and acid anhydride–exposed workers.

Fluoropolymers such as Teflon, which at normal temperatures produce no reaction, become volatilized upon high-temperature heating. The inhaled agents cause a characteristic syndrome of fever, chills, malaise, and occasionally mild wheezing, leading to the diagnosis of *polymer fume fever*. A similar self-limited, influenza-like syndrome—*metal fume fever*—results from acute exposure to fumes containing zinc oxide, typically from welding of galvanized steel. These inhalational fever syndromes may begin several hours after work and resolve within 24 h, only to return on repeated exposure.

Two other agents have been associated with potentially severe lung disease. Occupational exposure to nylon flock has been shown to induce a lymphocytic bronchiolitis, and workers exposed to diacetyl, which is used to provide "butter" flavor in the manufacture of microwave popcorn and other foods, have developed bronchiolitis obliterans **(Chap. 304)**.

World Trade Center Disaster A consequence of the attack on the World Trade Center (WTC) in New York City on September 11, 2001, was relatively heavy exposure of a large number of firefighters and other rescue workers to the dust generated by the collapse of the buildings. Environmental monitoring and chemical characterization of WTC dust have revealed a wide variety of potentially toxic constituents, although much of the dust was pulverized cement. Possibly because of the high alkalinity of WTC dust, significant cough, wheeze, and phlegm production occurred among firefighters and cleanup crews. New cough and wheeze syndromes also occurred among local residents. Heavier exposure to WTC dust among New York City firefighters was associated with accelerated decline of lung function over the first year after the disaster. More recently, concerns have been raised about risk of interstitial lung disease, especially of a granulomatous nature.

Burn Pit Emissions The U.S. military used open pits to burn waste of all types—so-called burn pits—during conflicts in the Middle Eastern and Southwest Asian theaters of operations. After deployment to these theaters, a considerable number of veterans complained of symptoms, primarily but not exclusively respiratory, that seem to have been chronologically attributable to exposure to burn pit emissions. A myriad of materials were burned using jet fuel, including plastics, metals, and human waste, generating multiple toxic agents in both particulate and gaseous form. While understanding the health effects of such exposures is an active area of ongoing research, the U.S. Congress recently passed legislation that provides disability compensation to military veterans for multiple burn pit and other toxic exposure presumptive conditions, including allergic rhinitis, asthma, COPD, vocal cord dysfunction, constrictive or obliterative bronchiolitis, and several interstitial lung diseases.

■ OCCUPATIONAL RESPIRATORY CARCINOGENS

Exposures at work have been estimated to contribute to 10% of all lung cancer cases. In addition to asbestos, other agents either proven or suspected to be respiratory carcinogens include acrylonitrile, arsenic compounds, beryllium, bis(chloromethyl) ether, chromium (hexavalent), formaldehyde (nasal), isopropanol (nasal sinuses), mustard gas, nickel carbonyl (nickel smelting), polycyclic aromatic hydrocarbons (coke oven emissions and diesel exhaust), secondhand tobacco smoke, silica (both mining and processing), talc (possible asbestos contamination in both mining and milling), vinyl chloride (sarcomas), wood (nasal), and uranium. Workers at risk of radiation-related lung cancer include not only those involved in mining or processing uranium but also those exposed in underground mining operations of other ores where radon daughters may be emitted from rock formations.

■ ASSESSMENT OF DISABILITY

Disability is the term used to describe the decreased ability to work due to the effects of a medical condition. Physicians are generally able to assess physiologic dysfunction, or *impairment*, but the rating of disability for compensation of loss of income also involves nonmedical factors such as the education and employability of the individual. The disability rating scheme differs with the compensation-granting agency. For example, the U.S. Social Security Administration requires that an individual be unable to do any work (i.e., *total* disability) before they will receive income replacement payments. Many state workers' compensation systems allow for payments for *partial* disability. In the Social Security scheme, no determination of cause is done, whereas work-relatedness must be established in workers' compensation systems.

For respiratory impairment rating, resting pulmonary function tests (spirometry and diffusing capacity) are used as the initial assessment tool, with cardiopulmonary exercise testing (to assess maximal oxygen consumption) used if the results of the resting tests do not correlate with the patient's symptoms. Methacholine challenge (to assess airway reactivity) can also be useful in patients with asthma who have normal spirometry when evaluated. Some compensation agencies (e.g., Social Security) have proscribed disability classification schemes based on pulmonary function test results. When no specific scheme is proscribed, the *Guidelines of the American Medical Association* should be used.

GENERAL ENVIRONMENTAL EXPOSURES

■ OUTDOOR AIR POLLUTION

Primary standards regulated by the U.S. Environmental Protection Agency (EPA) designed to protect the public health with an adequate margin of safety exist for sulfur dioxide, particulate matter (PM), nitrogen dioxide, ozone, lead, and carbon monoxide. Standards for each of these pollutants are updated regularly through an extensive review process conducted by the EPA. (For details on current standards, go to *https://www.epa.gov/criteria-air-pollutants/naaqs-table*.)

Pollutants are generated from both stationary sources (power plants and industrial facilities) and mobile sources (motor vehicles), and none of the regulated pollutants occurs in isolation. Furthermore, pollutants may be changed by chemical reactions after being emitted. For example, sulfur dioxide and PM emissions from a coal-fired power plant may react in air to produce acid sulfate aerosols, which can be transported long distances in the atmosphere. Oxides of nitrogen and volatile organic compounds from automobile exhaust react with sunlight to produce ozone. Although originally recognized in Los Angeles, photochemically derived pollution ("smog") is now known to be a problem throughout the United States and in many other countries. Both acute and chronic effects of pollutant exposures have been documented in large population studies.

The symptoms and diseases associated with air pollution are the same as conditions commonly associated with cigarette smoking. In addition, decreased growth of lung function and asthma have been associated with chronic exposure to only modestly elevated levels of traffic-related air pollution. Multiple population-based time-series studies within cities have demonstrated excess health care utilization for asthma and other cardiopulmonary conditions as well as increased mortality rates. Cohort studies comparing cities that have relatively high levels of particulate exposures with less polluted communities suggest excess morbidity and mortality rates from cardiopulmonary conditions in long-term residents of the former. The strong epidemiologic evidence that fine PM is a risk factor for cardiovascular morbidity and mortality has prompted toxicologic investigations into the underlying mechanisms. The inhalation of

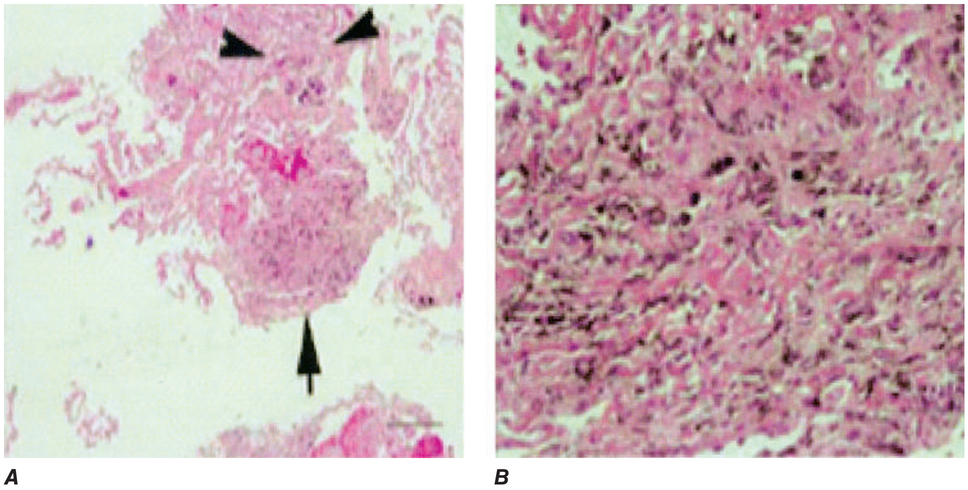

FIGURE 300-4 Histopathologic features of biomass smoke–induced interstitial lung disease. ***A.*** Anthracitic pigment is seen accumulating along alveolar septae (*arrowheads*) and within a pigmented dust macule (*single arrow*). ***B.*** A high-power photomicrograph contains a mixture of fibroblasts and carbon-laden macrophages.

fine particles from combustion sources generates oxidative stress followed by local injury and inflammation in the lungs that in turn lead to autonomic and systemic inflammatory responses. Recent research findings on the health effects of air pollutants have led to stricter U.S. ambient air quality standards for ozone, oxides of nitrogen, and PM as well as greater emphasis on publicizing pollution alerts to encourage individuals with cardiovascular and respiratory disorders to stay indoors during high-pollution episodes (e.g., from wildfires). In addition to staying indoors during episodes of poor air quality due to wildfires, creating clean air spaces in homes and buildings with central ventilation filtration and/or the use of portable HEPA air cleaners can reduce exposure to wildfire PM.

■ INDOOR EXPOSURES

Secondhand tobacco smoke **(Chap. 465)**, radon gas, wood smoke, and other biologic agents generated indoors must be considered. Several studies have shown that the respirable particulate load in any household is directly proportional to the number of cigarette smokers living in that home. Increases in prevalence of respiratory illnesses, especially asthma, and reduced levels of pulmonary function have been found in the children of smoking parents in a number of studies. Recent meta-analyses for lung cancer and cardiopulmonary diseases, combining data from multiple secondhand tobacco smoke epidemiologic studies, suggest an ~25% increase in relative risk for each condition, even after adjustment for major potential confounders.

Exposure to *radon gas* in homes is a risk factor for lung cancer. The main radon product (radon-222) is a gas that results from the decay series of uranium-238, with the immediate precursor being radium-226. The amount of radium in earth materials determines how much radon gas will be emitted. Levels associated with excess lung cancer risk may be present in as many as 10% of the houses in the United States. When smokers reside in the home, the problem is potentially greater, because the molecular size of radon particles allows them to attach readily to smoke particles that are inhaled. Fortunately, technology is available for assessing and reducing the level of exposure.

Other indoor exposures of concern are bioaerosols that contain antigenic material (fungi, cockroaches, dust mites, and pet danders) associated with an increased risk of atopy and asthma. Indoor chemical agents that have been associated with respiratory symptoms include strong cleaning agents (bleach, ammonia), formaldehyde, perfumes, and pesticides, Exposure to oxides of nitrogen from gas appliances, especially stoves has been associated with increased risk of asthma. Nonspecific responses associated with “tight-building syndrome,” perhaps better termed “building-associated illness,” in which no particular agent has been implicated, have included a wide variety of complaints, among them respiratory symptoms that are relieved only by avoiding exposure in the building in question. Indoor exposure to *household air pollution* from cooking or heating with solid fuels (wood, dung, crop residues, charcoal, coal) is estimated to be responsible for 4% of deaths worldwide, due to pneumonia in children, COPD and lung cancer in women, and cardiovascular disease among men. This burden of disease places exposure to household air pollution as one of the leading environmental hazards for poor health on a global scale.

Forty percent of the world’s population uses solid fuel for cooking, heating, or baking. Kerosene (similar to diesel fuel) is often used for lighting and sometimes cooking. This occurs predominantly in the rural areas of developing countries. Because many families burn coal or biomass fuels in open stoves, which are highly inefficient, and inside homes with poor ventilation, women and young children are exposed on a daily basis to high levels of smoke. In these homes, 24-h mean levels of fine PM have been reported to be 2–30 times higher than the National Ambient Air Quality Standard set by the U.S. EPA.

Epidemiologic studies have consistently shown associations between exposure to biomass smoke and both chronic bronchitis and COPD. Because of increased migration to the United States from developing countries, clinicians need to be aware of the chronic respiratory effects of exposure to biomass smoke, which can include interstitial lung disease **(Fig. 300-4)**.

Household air pollution (HAP) from domestic use of solid fuels also contributes substantially to outdoor air pollution. Contributions from HAP, coal-fired power plants without emission scrubbers, and increased traffic congestion involving motor vehicles without pollution controls can lead to high concentrations of outdoor air pollution, especially fine PM, in mega-cities in developing countries (e.g., Delhi).

ACKNOWLEDGMENT

The author acknowledges the contribution of Dr. Frank Speizer to the prior version of this chapter.

■ FURTHER READING

Blanc PD et al: The occupational burden of nonmalignant respiratory diseases. An official American Thoracic Society and European Respiratory Society statement. Am J Respir Crit Care Med 199:1312, 2019.

Caceres JD, Venkata AN: Asbestos-associated pulmonary disease. Curr Opin Pulm Med 29:76, 2023.

Fazio JC et al: Silicosis among immigrant engineered stone (quartz) countertop fabrication workers in California. JAMA Intern Med 183:991, 2023.

Lee KK et al: Adverse health effects associated with household air pollution: A systematic review, meta-analysis, and burden estimation study. Lancet Glob Health 8:e1427, 2020.

Mein SA et al: Lifetime exposure to traffic-related pollution and lung function in early adolescence. Ann Am Thorac Soc 19:1776, 2023.

Weissman DN: Progressive massive fibrosis: An overview of the recent literature. Pharmacol Ther 240:108232, 2022.

301 Bronchiectasis

Rebecca M. Baron, Beverly W. Baron, Miriam Baron Barshak

Bronchiectasis refers to an irreversible airway dilation that involves the lung in either a focal or a diffuse manner and that classically has been categorized as cylindrical or tubular (the most common form), varicose, or cystic. This chapter will focus largely on non–cystic fibrosis (CF) bronchiectasis. **The reader is referred to Chapter 302 for a more focused discussion on CF bronchiectasis.**

ETIOLOGY

Bronchiectasis can arise from infectious or noninfectious causes (Table 301-1). Clues to the underlying etiology often are provided by the pattern of lung involvement. *Focal bronchiectasis* refers to bronchiectatic changes in a localized area of the lung and can be a consequence of obstruction of the airway—either extrinsic (e.g., due to compression by adjacent lymphadenopathy or parenchymal tumor mass) or intrinsic (e.g., due to an airway tumor or aspirated foreign body, a scarred/stenotic airway, or bronchial atresia from congenital underdevelopment of the airway). *Diffuse bronchiectasis* is characterized by widespread bronchiectatic changes throughout the lung and often arises from an underlying systemic or infectious disease process.

TABLE 301-1 Major Etiologies of Bronchiectasis and Proposed Workup

PATTERN OF LUNG INVOLVEMENT	ETIOLOGY BY CATEGORY (EXAMPLES)	WORKUP
Focal	Obstruction (e.g., aspirated foreign body, tumor mass)	Chest imaging (chest x-ray and/or chest CT)[a]; bronchoscopy
Diffuse	Infection (e.g., bacterial, nontuberculous mycobacterial)	Sputum Gram's stain/culture; stains/cultures for acid-fast bacilli and fungi. If no pathogen is identified, consider bronchoscopy with bronchoalveolar lavage.
	Immunodeficiency (e.g., hypogammaglobulinemia, HIV infection, bronchiolitis obliterans after lung transplantation)	Complete blood count with differential; immunoglobulin measurement; HIV testing
	Genetic causes (e.g., cystic fibrosis, Kartagener's syndrome, α_1 antitrypsin deficiency)	Measurement of chloride levels in sweat (for cystic fibrosis), α_1 antitrypsin levels; nasal or respiratory tract brush/biopsy (for dyskinetic/immotile cilia syndrome); genetic testing
	Autoimmune or rheumatologic causes (e.g., rheumatoid arthritis, Sjögren's syndrome, inflammatory bowel disease); immune-mediated disease (e.g., allergic bronchopulmonary aspergillosis)	Clinical examination with careful joint exam, serologic testing (e.g., for rheumatoid factor). Consider workup for allergic bronchopulmonary aspergillosis, especially in patients with refractory asthma.[b]
	Recurrent aspiration	Test of swallowing function and general neuromuscular strength
	Miscellaneous (e.g., yellow nail syndrome, traction bronchiectasis from postradiation fibrosis or idiopathic pulmonary fibrosis)	Guided by clinical condition
	Idiopathic	Exclusion of other causes

[a]Chest imaging is included in the general workup for all etiologies of bronchiectasis as described in the text. [b]Skin testing for *Aspergillus* reactivity; measurement of serum precipitins for *Aspergillus*, serum IgE levels, serum eosinophils, etc.

More pronounced involvement of the upper lung fields is most common in CF and also is observed in postradiation fibrosis, corresponding to the lung region encompassed by the radiation port. Bronchiectasis with predominant involvement of the lower lung fields usually has its source in chronic recurrent aspiration (e.g., due to esophageal motility disorders like those in scleroderma), end-stage fibrotic lung disease (e.g., traction bronchiectasis from idiopathic pulmonary fibrosis), or recurrent immunodeficiency-associated infections (e.g., hypogammaglobulinemia). Bronchiectasis resulting from infection by nontuberculous mycobacteria (NTM), most commonly the *Mycobacterium avium-intracellulare* complex (MAC), often preferentially affects the midlung fields. Congenital causes of bronchiectasis with predominant midlung field involvement include the dyskinetic/immotile cilia syndrome. Finally, predominant involvement of the central airways is reported in association with allergic bronchopulmonary aspergillosis (ABPA), in which an immune-mediated reaction to *Aspergillus* damages the bronchial wall. Congenital causes of central airway–predominant bronchiectasis resulting from cartilage deficiency include tracheobronchomegaly (Mounier-Kuhn syndrome) and Williams-Campbell syndrome.

In many cases, the etiology of bronchiectasis is not determined. In case series, as many as 25–50% of patients referred for bronchiectasis have idiopathic disease. There is increasing appreciation for the need to define disease subphenotypes in this heterogeneous group of underlying causes of bronchiectasis, which might permit better targeting of clinical trials and treatment strategies.

EPIDEMIOLOGY

The overall reported prevalence of bronchiectasis in the United States has recently increased, but the epidemiology of bronchiectasis varies greatly with the underlying etiology. For example, patients with CF often develop significant clinical bronchiectasis in late adolescence or early adulthood, although atypical presentations of CF in adults in their thirties and forties also are possible. In contrast, bronchiectasis resulting from MAC infection classically affects nonsmoking women >50 years of age. In general, the incidence of bronchiectasis increases with age. Bronchiectasis is more common among women than among men. Bronchiectasis may also frequently be co-diagnosed with chronic obstructive pulmonary disease (COPD) or asthma.

In areas where tuberculosis is prevalent, bronchiectasis more frequently occurs as a sequela of granulomatous infection. Focal bronchiectasis can arise from extrinsic compression of the airway by enlarged granulomatous lymph nodes and/or from development of intrinsic obstruction as a result of erosion of a calcified lymph node through the airway wall (e.g., broncholithiasis). Especially in reactivated tuberculosis, parenchymal destruction from infection can result in areas of more diffuse bronchiectasis. Apart from cases associated with tuberculosis, an increased incidence of non-CF bronchiectasis with an unclear underlying mechanism has been reported as a significant problem in developing nations. It has been suggested that the high incidence of malnutrition in certain areas may predispose to immune dysfunction and development of bronchiectasis.

PATHOGENESIS AND PATHOLOGY

The most widely cited mechanism of infectious bronchiectasis is the "vicious cycle hypothesis," in which susceptibility to infection and poor mucociliary clearance result in microbial colonization of the bronchial tree. Some organisms, such as *Pseudomonas aeruginosa*, exhibit a particular propensity for colonizing damaged airways and evading host defense mechanisms. Impaired mucociliary clearance can result from inherited conditions such as CF or dyskinetic cilia syndrome, and it has been proposed that a single severe infection (e.g., pneumonia caused by *Bordetella pertussis* or *Mycoplasma pneumoniae*) can result in significant airway damage and poor secretion clearance. The presence of the microbes incites continued chronic inflammation, with consequent damage to the airway wall, continued impairment of secretions and microbial clearance, and ongoing propagation of the infectious/inflammatory cycle. Moreover, it has been proposed that mediators released directly from bacteria can interfere with mucociliary clearance. A recent study suggested that there exist molecular endotypes of bronchiectasis with differential inflammatory markers and microbiome signatures that correlate with risk of exacerbations. Classic studies of the pathology of bronchiectasis from the 1950s demonstrated significant small-airway wall inflammation and larger-airway wall destruction as well as dilation, with loss of elastin, smooth muscle, and cartilage. It has been proposed that inflammatory cells in the small airways release proteases and other mediators, such as reactive oxygen species and proinflammatory cytokines, that damage the larger airway walls. Furthermore, the ongoing inflammatory process in the smaller airways results in airflow obstruction. It is thought that antiproteases, such as α_1 antitrypsin, play an important role in neutralizing the damaging effects of neutrophil elastase and in enhancing bacterial killing. Bronchiectasis and emphysema have been observed in patients with α_1 antitrypsin deficiency. Interestingly, a recent phase 2 study demonstrated improved bronchiectasis outcomes with an oral inhibitor of neutrophil serine protease activity.

Proposed mechanisms for noninfectious bronchiectasis include immune-mediated reactions that damage the bronchial wall (e.g., those associated with systemic autoimmune conditions such as Sjögren's syndrome and rheumatoid arthritis). Recent studies suggest that there might exist a new bronchiectasis endophenotype of patients with sensitization to multiple environmental allergens. *Traction bronchiectasis* refers to dilated airways arising from parenchymal distortion as a result of lung fibrosis (e.g., postradiation fibrosis or idiopathic pulmonary fibrosis).

■ CLINICAL MANIFESTATIONS

The most common clinical presentation is a persistent productive cough with ongoing production of thick, tenacious sputum. Physical findings frequently include crackles and wheezing on lung auscultation, and some patients with bronchiectasis exhibit clubbing of the digits. Mild to moderate airflow obstruction often is detected on pulmonary function tests, overlapping with that seen at presentation with other conditions, such as COPD. Acute exacerbations of bronchiectasis usually are characterized by changes in the nature of sputum production, with increased volume and purulence. However, typical signs and symptoms of lung infection, such as fever and new infiltrates, may not be present.

■ DIAGNOSIS

The diagnosis usually is based on presentation with a persistent chronic cough and sputum production accompanied by consistent radiographic features. Although chest radiographs lack sensitivity, the presence of "tram tracks" indicating dilated airways is consistent with bronchiectasis. Chest CT is more specific for bronchiectasis and is the imaging modality of choice for confirming the diagnosis. CT findings include airway dilation (detected as parallel "tram tracks" or as the "signet-ring sign"—a cross-sectional area of the airway with a diameter at least 1.5 times that of the adjacent vessel), lack of bronchial tapering (including the presence of tubular structures within 1 cm from the pleural surface), bronchial wall thickening in dilated airways, inspissated secretions (e.g., the "tree-in-bud" pattern), or cysts emanating from the bronchial wall (especially pronounced in cystic bronchiectasis) **(Fig. 301-1)**. Recently, a group of international experts put forth consensus guidelines for clinical and radiologic diagnosis of bronchiectasis, proposing that a diagnosis of bronchiectasis should require radiologic criteria (at least one of the following on chest CT: [1] inner- or [2] outer-airway-artery diameter ratio ≥1.5; [3] lack of airway tapering; and, [4] visibility of airways in the periphery) along with the clinical syndrome (at least two of the following: [1] cough most days of the week; [2] sputum production most days of the week; and [3] history of exacerbations).

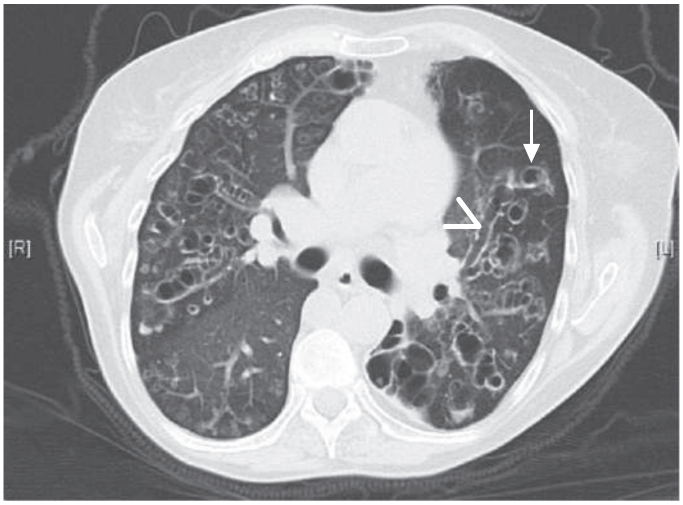

FIGURE 301-1 Representative chest CT image of severe bronchiectasis. This patient's CT demonstrates many severely dilated airways, seen both longitudinally (*arrowhead*) and in cross-section (*arrow*).

APPROACH TO THE PATIENT

Bronchiectasis

The evaluation of a patient with bronchiectasis entails elicitation of a clinical history, chest imaging, and a workup to determine the underlying etiology. Evaluation of focal bronchiectasis almost always requires bronchoscopy to exclude airway obstruction by an underlying mass or foreign body. A workup for diffuse bronchiectasis includes analysis for the major etiologies (Table 301-1), with an initial focus on excluding CF. Pulmonary function testing is an important component of a functional assessment of the patient.

TREATMENT

Bronchiectasis

Treatment of infectious bronchiectasis is directed at the control of active infection and improvements in secretion clearance and bronchial hygiene so as to decrease the microbial load within the airways and minimize the risk of repeated infections.

ANTIBIOTIC TREATMENT

Antibiotics targeting the causative or presumptive pathogen (with *Haemophilus influenzae* and *P. aeruginosa* isolated commonly) should be administered in acute exacerbations, usually for a minimum of 7–10 days and perhaps for as long as 14 days. Decisions about treatment of NTM infection can be difficult, given that these organisms can be colonizers as well as pathogens, and the prolonged treatment course often is not well tolerated. Consensus guidelines have advised that diagnostic criteria for true clinical infection with NTM should be considered in patients with symptoms and radiographic findings of lung disease who have at least two sputum samples positive on culture; at least one bronchoalveolar lavage (BAL) fluid sample positive on culture; a biopsy sample displaying histopathologic features of NTM infection (e.g., granuloma or a positive stain for acid-fast bacilli) along with one positive sputum culture; or a pleural fluid sample (or a sample from another sterile extrapulmonary site) positive on culture. MAC strains are the most common NTM pathogens, and the recommended regimen for HIV-negative patients infected with macrolide-sensitive MAC includes a macrolide combined with rifampin and ethambutol. Consensus guidelines recommend macrolide susceptibility testing for clinically significant MAC isolates.

BRONCHIAL HYGIENE

The numerous approaches used to enhance secretion clearance in bronchiectasis include hydration and mucolytic administration, aerosolization of bronchodilators and hyperosmolar agents (e.g., hypertonic saline), and chest physiotherapy (e.g., postural drainage, traditional mechanical chest percussion via hand clapping to the chest, or use of devices such as an oscillatory positive expiratory pressure flutter valve or a high-frequency chest wall oscillation vest). Pulmonary rehabilitation and a regular exercise program may assist with secretion clearance as well as with other aspects of bronchiectasis, including improved exercise capacity and quality of life. The mucolytic dornase (DNase) is recommended routinely in CF-related bronchiectasis but not in non-CF bronchiectasis, given concerns about lack of efficacy and potential harm in the non-CF population.

ANTI-INFLAMMATORY THERAPY

It has been proposed that control of the inflammatory response may be of benefit in bronchiectasis, and relatively small-scale trials have yielded evidence of alleviated dyspnea, decreased need for inhaled β-agonists, and reduced sputum production with inhaled glucocorticoids. However, no significant differences in lung function or bronchiectasis exacerbation rates have been observed. Risks of immunosuppression and adrenal suppression must be carefully considered with use of anti-inflammatory therapy in infectious bronchiectasis. Nevertheless, administration of oral/systemic glucocorticoids may be important in treatment of bronchiectasis due to certain etiologies, such as ABPA, or of noninfectious bronchiectasis due to underlying conditions, especially that in which an autoimmune condition is believed to be active (e.g., rheumatoid arthritis or Sjögren's syndrome). Patients with ABPA also may benefit from a prolonged course of treatment with an oral antifungal agent such as itraconazole.

REFRACTORY CASES

In select cases, surgery can be considered, with resection of a focal area of suppuration. In advanced cases, lung transplantation can be considered.

COMPLICATIONS

In more severe cases of infectious bronchiectasis, recurrent infections and repeated courses of antibiotics can lead to microbial resistance to antibiotics. In certain cases, combinations of antibiotics that have independent toxicity profiles may be necessary to treat resistant organisms.

Recurrent infections can result in injury to superficial mucosal vessels, with bleeding and, in severe cases, life-threatening hemoptysis. Management of massive hemoptysis usually requires intubation to stabilize the patient, identification of the source of bleeding, and protection of the nonbleeding lung. Control of bleeding often necessitates bronchial artery embolization and, in severe cases, surgery.

PROGNOSIS

Outcomes of bronchiectasis can vary widely with the underlying etiology and comorbid conditions and may also be influenced by the frequency of exacerbations and (in infectious cases) the specific pathogens involved (with worse outcomes associated with *P. aeruginosa* colonization). Increasing attention is being given to defining clinical subphenotypes of bronchiectasis in light of heterogeneous clinical, radiographic, and microbial features and to developing screening tools for the assessment of quality of life and disease severity. In one study, the decline of lung function in patients with non-CF bronchiectasis was similar to that in patients with COPD, with the forced expiratory volume in 1 s (FEV_1) declining by 50–55 mL per year as opposed to 20–30 mL per year for healthy controls.

PREVENTION

Reversal of an underlying immunodeficient state (e.g., by administration of gamma globulin for immunoglobulin-deficient patients) and vaccination of patients with chronic respiratory conditions (e.g., influenza, pneumococcal, COVID, and RSV vaccines) can decrease the risk of recurrent infections. Patients who smoke should be counseled about smoking cessation.

After resolution of an acute infection in patients with recurrences (e.g., ≥3 episodes per year), the use of suppressive antibiotics to minimize the microbial load and reduce the frequency of exacerbations has been proposed. Although there is less consensus about this approach in non-CF-associated bronchiectasis than in CF-related bronchiectasis, small studies have supported benefits of selected therapies, though with concerns for development of antibiotic resistance over time. Possible suppressive treatments include (1) administration of an oral antibiotic (e.g., ciprofloxacin) daily for 1–2 weeks per month; (2) use of a rotating schedule of oral antibiotics (to minimize the risk of development of drug resistance); (3) administration of a macrolide antibiotic (see below) daily or three times per week (with mechanisms of possible benefit related to non-antimicrobial properties, such as anti-inflammatory effects and reduction of gram-negative bacillary biofilms); (4) inhalation of aerosolized antibiotics (e.g., tobramycin inhalation solution) for select patients on a rotating schedule (e.g., 30 days on, 30 days off), with the goal of decreasing the microbial load without eliciting the side effects of systemic drug administration; other studies examining inhaled aztreonam and inhaled ciprofloxacin formulations have shown conflicting results, suggesting there might be subpopulations of patients with bronchiectasis who might benefit from specific therapies; and (5) intermittent administration of IV antibiotics (e.g., "clean-outs") for patients with more severe bronchiectasis and/or resistant pathogens. In relation to macrolide therapy (point 3 above), a number of double-blind, placebo-controlled, randomized trials have been published in non-CF bronchiectasis and support a benefit of long-term macrolides (6–12 months of azithromycin or erythromycin) in decreasing rates of bronchiectasis exacerbation, mucus production, and decline in lung function. However, two of these studies and a meta-analysis also reported increased macrolide resistance in commensal pathogens, dampening enthusiasm for universal use of macrolides in this setting and raising the question of whether there might be select non-CF bronchiectasis patients with higher morbidity for whom benefits of long-term macrolides might outweigh the risks of emergence of antibiotic resistance. In particular, development of macrolide-resistant NTM is a potential concern, making treatment of those pathogens much more difficult. Furthermore, patients with different patterns of microbial colonization may not all experience similar benefits with macrolide therapy. Therefore, before chronic macrolide therapy is considered, it is advisable to rule out NTM infection and carefully consider each patient's scenario closely, obtaining an electrocardiogram to rule out a prolonged QT interval that might place the patient at increased risk of arrhythmias.

In addition, ongoing consistent attention to bronchial hygiene can promote secretion clearance and decrease the microbial load in the airways.

FURTHER READING

Aliberti S et al: Criteria and definitions for the radiological and clinical diagnosis of bronchiectasis in adults for use in clinical trials: International consensus recommendations. Lancet Respir Med 10:298, 2022.

Chalmers JD, Chotirmall SH: Bronchiectasis: New therapies and new perspectives. Lancet Respir Med 6:715, 2018.

Choi H et al: Inflammatory molecular endotypes in bronchiectasis: A European multicenter cohort study. Am J Respir Crit Care Med 208:1166, 2023.

Guan W-J et al: A double-blind randomized placebo-controlled phase 3 trial of tobramycin inhalation solution in adults with bronchiectasis with *Pseudomonas aeruginosa* infection. Chest 163:64, 2023.

Herrero-Cortina B et al: European Respiratory Society statement on airway clearance techniques in adults with bronchiectasis. Eur Respir J 62:2202053, 2023.

302 Cystic Fibrosis

Eric J. Sorscher

CLINICAL FEATURES

Cystic fibrosis (CF) is an autosomal recessive exocrinopathy affecting multiple epithelial tissues. The gene product responsible for CF (the cystic fibrosis transmembrane conductance regulator [CFTR]) serves as an anion channel in the apical (luminal) plasma membranes of epithelial cells and regulates volume and composition of exocrine secretion. A highly sophisticated understanding of CFTR molecular genetics and membrane protein biochemistry has enabled recent and transformative drug discovery for patients with this disease.

Respiratory Manifestations The major morbidity and mortality associated with CF is attributable to pulmonary compromise, characterized by copious hyperviscous and adherent secretions that obstruct small and medium-sized airways. CF respiratory secretions are exceedingly difficult to clear, and a complex bacterial flora that includes *Staphylococcus aureus*, *Haemophilus influenzae*, and *Pseudomonas aeruginosa* (among other pathogens, see below) is routinely cultured from CF sputum. Robust pulmonary inflammation in the setting of inspissated mucus and chronic infection leads to collateral tissue injury and further aggravates respiratory decline. Organisms such as *P. aeruginosa* exhibit a stereotypic mode of pathogenesis; a sentinel and early colonization event often engenders lifelong pulmonary infection by the same genetic strain. Over a period of many years, *P. aeruginosa* evolves in CF lungs to adopt a mucoid phenotype (attributable to release of alginate exoproduct) that confers selective advantage for the pathogen and poor prognosis for the host. Radiographic evidence of sinusitis occurs in most patients with CF and can be associated with microorganisms similar to those recovered from lower airways, suggesting that CF sinus involvement may serve as a reservoir for bacterial seeding of the lung.

Pancreatic Findings The complete name of the disease, *cystic fibrosis of the pancreas*, refers to profound tissue destruction of the exocrine pancreas, with fibrotic scarring and/or fatty replacement, cyst formation, loss of acinar tissue, and ablation of normal pancreatic architecture. As in the lung, tenacious exocrine secretions (sometimes termed *concretions*) obstruct pancreatic ducts and impair production and flow of digestive enzymes to the duodenum. The sequelae of exocrine pancreatic insufficiency include chronic malabsorption, poor growth, fat-soluble vitamin deficiency, high levels of blood immunoreactive trypsinogen (a test used in newborn screening), low levels of fecal elastase-1, and loss of pancreatic islet cell mass. CF-related diabetes mellitus is a manifestation in >30% of adults with the disease and likely multifactorial in nature (attributable to progressive destruction/dysfunction of the endocrine pancreas and, in some cases, insulin resistance or other features).

Additional Organ System Damage As in CF lung and pancreas, thick and inspissated secretions compromise numerous exocrine tissues. Obstruction of intrahepatic bile ducts and parenchymal fibrosis are observed, with multilobular cirrhosis in 4–15% of patients with CF and significant hepatic insufficiency as a resulting manifestation among many adults. Hepatic steatosis, focal biliary fibrosis, and portal hypertension are other well-described findings of CF liver disease. Contents of the intestinal lumen are hyperviscous and often difficult to excrete, leading to meconium ileus (a clinical presentation in many newborns with CF) or distal intestinal obstructive syndrome in older individuals. Men typically exhibit complete involution of the vas deferens (despite functioning spermatogenesis), and ~99% of males with CF are infertile (i.e., unable to conceive children without in vitro fertilization). Abnormalities of female reproductive tract secretions are likely contributors to a higher incidence of infertility among women with the disease.

PATHOGENESIS

Cystic Fibrosis Transmembrane Conductance Regulator CFTR is an integral membrane protein that functions as an epithelial anion channel. The ~1480-amino-acid molecule encodes a passive conduit for chloride and bicarbonate transport across plasma membranes of epithelial tissues, with direction of ion flow dependent on the electrochemical driving force. Gating of CFTR involves conformational cycling between an open and closed configuration and is augmented by hydrolysis of adenosine triphosphate (ATP). Anion flux mediated by CFTR does not involve active transport against a concentration gradient but utilizes the energy provided from ATP hydrolysis as a central feature of ion channel mechanochemistry and gating.

CFTR is situated in the apical plasma membranes of acinar and other epithelial cells where it regulates the amount and composition of secretion by exocrine glands. In numerous epithelia, chloride and bicarbonate release via CFTR is followed passively by flow of water through other pathways, aiding mobilization and clearance of exocrine products. Along respiratory mucosa, CFTR is necessary to provide sufficient depth of the periciliary fluid layer (PCL), allowing normal ciliary extension and mucociliary transport. CFTR-deficient airway cells exhibit depleted PCL, causing ciliary collapse and failure to clear overlying mucus **(Video 302-1)**. In airway submucosal glands, CFTR is expressed in acini and may participate both in the formation of mucus and extrusion of glandular secretion onto the airway surface **(Fig. 302-1)**. In other exocrine glands characterized by abrogated mucus transport (e.g., pancreatic acini and ducts, as well as bile canaliculi, and intestinal tissues), similar pathogenic mechanisms have been implicated. In these cases, a driving force for apical chloride and/or bicarbonate secretion is believed to promote CFTR-mediated fluid and electrolyte release into the lumen, which confers proper rheologic properties and composition of mucins and other related exocrine products. Failure of this mechanism disrupts normal hydration and transport of glandular secretion—and is viewed as a proximate cause of obstruction, with concomitant tissue injury.

Pulmonary Inflammation and Remodeling The CF airway is characterized by an aggressive, unrelenting, neutrophilic inflammatory response with release of proteases and oxidants leading to airway remodeling and bronchiectasis. Intense pulmonary inflammation is largely driven by chronic respiratory pathogens, although some measure of inflammatory responsiveness may occur very early in the disease and prior to bacterial infection. Macrophages and other cells resident in CF lungs augment elaboration of proinflammatory cytokines, which contribute to innate and adaptive immune reactivity. CFTR-dependent abnormalities of airway surface fluid composition (e.g., low pH) have been reported as contributors to impaired bacterial killing in CF lungs. The role of CFTR activity as a direct mediator of cellular immune hyperresponsiveness and/or pulmonary remodeling (i.e., intensifying damage caused by mucoobstruction and bacterial infection) represent important areas of investigation.

MOLECULAR GENETICS

DNA sequencing from CF patients (and others) worldwide has revealed >2000 allelic mutations in *CFTR*; several hundred of these are well characterized as disease-causing variants. Distinguishing the single nucleotide transversions or other polymorphisms with causal relevance can sometimes present a significant challenge. The CFTR2 resource (*www.cftr2.org/*) helps delineate gene variants with a clear etiologic role.

CFTR defects known to elicit disease have been categorized based on molecular mechanism. For example, the common F508del mutation (nomenclature denotes omission of a single phenylalanine residue [F] at CFTR position 508) leads to a protein folding abnormality recognized by cellular quality control pathways. CFTR encoding F508del retains partial ion channel function, but protein maturation is arrested in the endoplasmic reticulum, and CFTR fails to arrive at the plasma membrane. Instead, F508del CFTR is misrouted and undergoes endoplasmic reticulum–associated degradation via the proteasome. *CFTR* mutations that disrupt protein maturation in this manner are termed

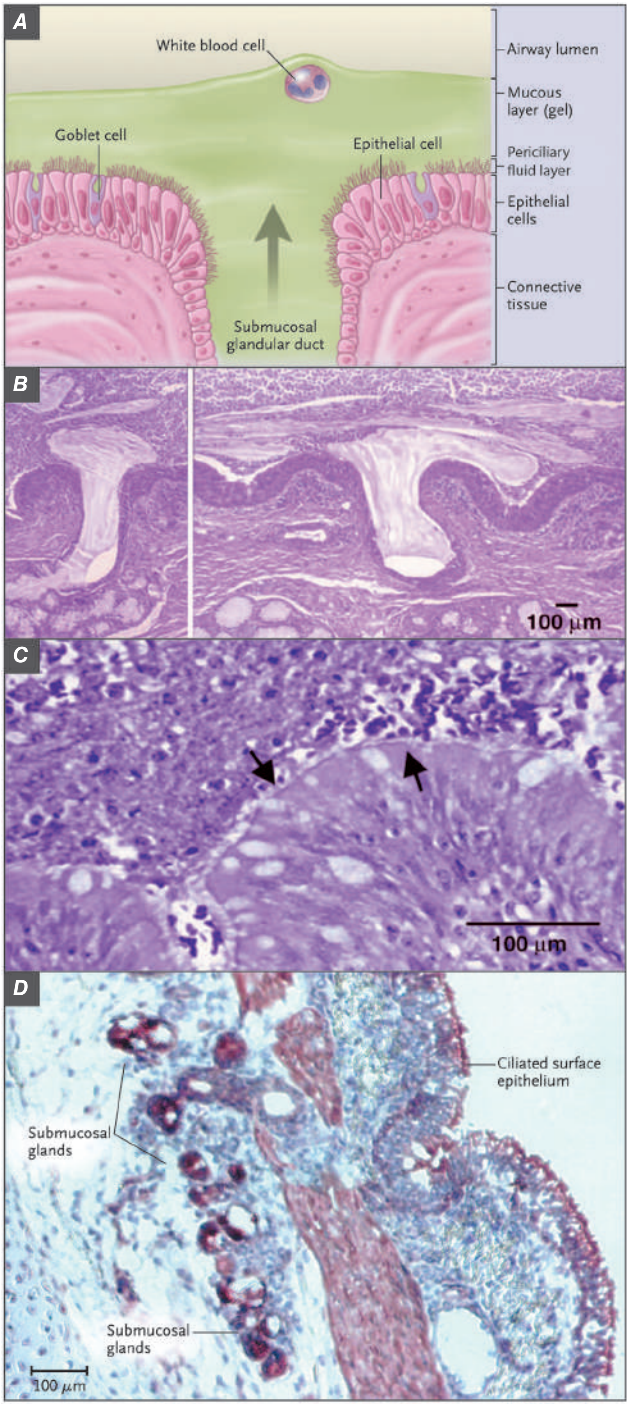
A
White blood cell
Airway lumen
Mucous layer (gel)
Goblet cell
Epithelial cell
Periciliary fluid layer
Epithelial cells
Connective tissue
Submucosal glandular duct
B
100 µm
C
100 µm
D
Ciliated surface epithelium
Submucosal glands
Submucosal glands
100 µm

FIGURE 302-1 Extrusion of mucus secretion onto the epithelial surface of airways in cystic fibrosis (CF). ***A.*** Schematic of the surface epithelium and supporting glandular structure of the human airway. ***B.*** The submucosal glands of a patient with CF are filled with mucus, and mucopurulent debris overlies the airway surfaces, essentially burying the epithelium. ***C.*** A higher magnification view of a mucus plug tightly adhering to the airway surface, with *arrows* indicating the interface between infected and inflamed secretions and the underlying epithelium to which the secretions adhere. (Both ***B*** and ***C*** were stained with hematoxylin and eosin, with the colors modified to highlight structures.) Infected secretions obstruct airways and, over time, dramatically disrupt the normal architecture of the lung. ***D.*** CFTR is expressed in surface epithelium and serous cells at the base of submucosal glands in a porcine lung sample, as shown by the dark staining, signifying binding by CFTR antibodies to epithelial structures (aminoethylcarbazole detection of horseradish peroxidase with hematoxylin counterstain). *(From SM Rowe, S Miller, EJ Sorscher: Cystic Fibrosis. N Engl J Med 352:1992, 2005. Copyright © 2005 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

class II defects and are by far the most common genetic abnormalities. F508del alone accounts for ~85% of defective *CFTR* alleles in the United States.

Other gene defects include CFTR ion channels properly trafficked to the apical cell surface but unable to open and/or gate. Such channel proteins include G551D (a glycine to aspartic acid replacement at CFTR position 551), which abrogates transport of Cl^- or HCO_3^- (a class III abnormality). Individuals with at least one G551D allele represent ~4% of patients with CF. *CFTR* nonsense mutations such as G542X, R553X, or W1282X (premature termination codon replaces glycine, arginine, or tryptophan at positions 542, 553, or 1282, respectively) are among the common class I defects, in addition to large deletions or other major disruptions of the gene. The W1282X mutation, for example, is prevalent among individuals of Ashkenazi descent and a predominant CF genotype in Israel. Additional categories of *CFTR* mutation include impairment of the ion channel pore (class IV), RNA splicing (class V), and increased plasma membrane turnover (class VI) **(Fig. 302-2)**.

DIAGNOSIS

CF classically presented in childhood with chronic productive cough, malabsorption including steatorrhea, and failure to thrive. During the past decade, newborn screening has led to most CF diagnoses, with confirmation through *CFTR* mutation analysis and/or sweat electrolyte measurements as cardinal tests. The findings of disease-causing

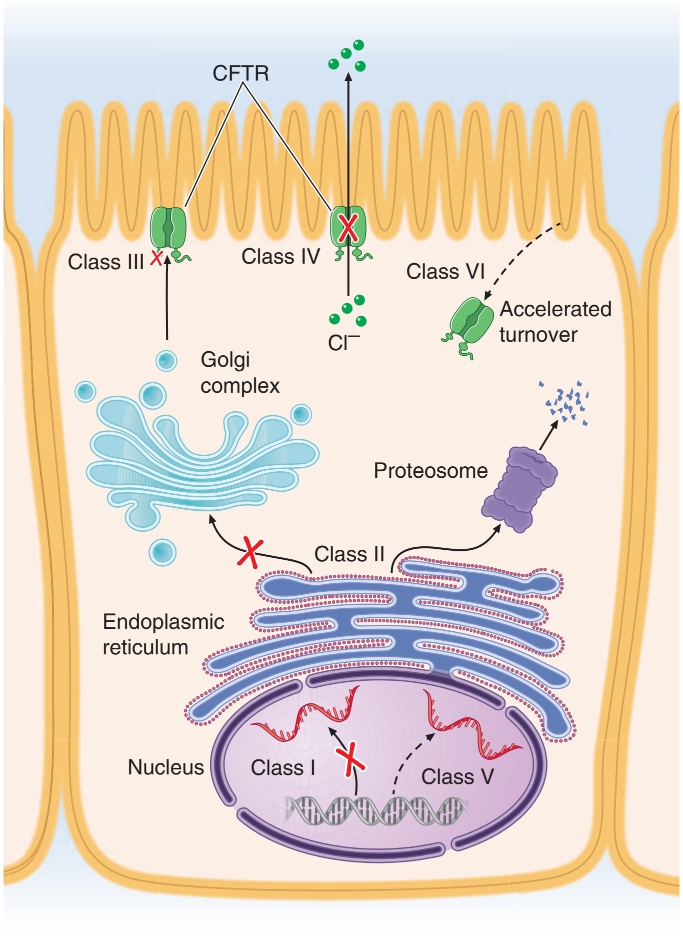
CFTR
Class III
Class IV
Class VI
Cl⁻
Accelerated turnover
Golgi complex
Proteosome
Class II
Endoplasmic reticulum
Nucleus
Class I
Class V

FIGURE 302-2 Categories of *CFTR* mutations. Classes of defects in the *CFTR* gene include the absence of synthesis (class I); defective protein maturation and premature degradation (class II); disordered gating/regulation, such as diminished adenosine triphosphate (ATP) binding and hydrolysis (class III); defective conductance through the ion channel pore (class IV); a reduced number of *CFTR* transcripts due to a promoter or splicing abnormality (class V); and accelerated turnover from the cell surface (class VI). *(From SM Rowe, S Miller, EJ Sorscher: N Engl J Med 352:1992, 2005.)*

variants on both *CFTR* alleles and/or sweat chloride ≥60 mEq/L, together with characteristic respiratory or other exocrine manifestations, are sufficient for confirming a diagnosis of CF. DNA-based evaluation typically surveys numerous disease-associated mutations; panels that identify several hundred *CFTR* variants are available through a variety of public health laboratories or commercial sources. Alternatively, complete *CFTR* DNA testing, or exonic sequencing together with analysis of splice junctions and key regulatory elements, can be obtained.

Sweat electrolytes following pilocarpine iontophoresis comprise an essential diagnostic element, with levels of chloride markedly elevated in CF compared to non-CF individuals. The sweat test result is highly specific and served as a mainstay of diagnosis for decades prior to availability of *CFTR* genotyping. Notably, hyperviscosity of eccrine sweat is not a clinical feature of the disease. Sweat ducts function to reabsorb chloride from a primary sweat secretion produced by the glandular coil. Malfunction of CFTR leads to diminished chloride uptake from the ductular lumen, and sweat emerges on the skin with elevated levels of chloride.

COMPLEXITY OF A CF-RELATED PHENOTYPE

Several "severe" defects that impair CFTR activity (including F508del, G551D, and nonsense alleles) are predictive of pancreatic insufficiency, which is evident in ~80% of those with the disease. In general, and with regard to projecting respiratory outcome for an individual patient, genotype has been of limited value for predicting the rate of clinical decline, respiratory prognosis, or longevity.

A spectrum of CFTR-related conditions with features resembling classic CF has been well described. In addition to multiorgan involvement, forme frustes, such as isolated congenital bilateral absence of the vas deferens or pancreatitis (without other organ system findings), are strongly associated with *CFTR* mutations in at least one allele. CF carrier status also predisposes to both non-CF bronchiectasis and chronic rhinosinusitis, indicating a contribution of relative CFTR deficiency in these prevalent illnesses. Although CF is a classic monogenic disease, the importance of non-*CFTR* gene modifiers and proteins that regulate ion flux, inflammatory pathways, and airway remodeling has been appreciated as influencing clinical course. For example, the magnitude of transepithelial sodium reabsorption in CF airways, which helps control periciliary fluid depth and composition, is strongly influenced by CFTR and has represented a molecular target for experimental intervention in the past.

CFTR MODULATORS

Potentiation of Mutant CFTR Gating A major effort directed toward high-throughput analysis of large compound libraries resulted in identification of effective new CFTR "modulator" therapies for CF. The first approved compound in this class, ivacaftor, robustly potentiates CFTR channel opening and stimulates ion transport. Ivacaftor overcomes the G551D CFTR gating defect, and individuals carrying this mutation exhibit pronounced improvement in lung function, respiratory outcomes (fewer hospital admissions for pulmonary exacerbation), weight gain, and other clinical benefits. Ivacaftor has been deemed "highly effective modulator therapy" (HEMT) for G551D-related CF and leads to substantial reduction of sweat chloride. Prior to ivacaftor, no clinical intervention of any sort had been shown to normalize the CF sweat phenotype. In addition to G551D, ivacaftor is approved in the United States for 96 other *CFTR* variants. Multiyear treatment analysis indicates durable respiratory palliation. The drug has been viewed as a harbinger of a new era for CF therapeutics directed toward addressing fundamental causes of the disease.

Correction of the F508del Processing Abnormality Lumacaftor and tezacaftor, two U.S. Food and Drug Administration (FDA)-approved "corrector" molecules that repair CFTR misfolding (as distinct from CFTR gating "potentiators" such as ivacaftor), partially overcome the F508del biogenesis defect. The drugs also promote cell surface localization of many other class II *CFTR* mutations. A different corrector molecule, elexacaftor, operates by a distinct mechanism of action and is FDA approved in combination with tezacaftor and ivacaftor for patients with CF encoding at least one F508del variant (irrespective of the second *CFTR* allele), as well as numerous less common CF mutations. This triple combination therapy (TCT) has been projected to benefit >90% of individuals with the disease. Marked enhancement of forced expiratory volume in 1 s (FEV_1), fewer respiratory exacerbations, improved quality of life, and diminished sweat chloride have all been demonstrated in patients following TCT, leading to designation as HEMT. For example, among individuals carrying one F508del together with a *CFTR* minimal function variant, TCT led to improved FEV_1 (% predicted) by ~14% over a 4- to 24-week treatment period. Monitoring liver function of patients on TCT and attention to pharmacologic interactions, including effects mediated by CYP3A, are required. **(See Video 302-2 A, B.)**

Personalized Molecular Therapies Based on the large number of disease-causing *CFTR* mutations, together with the ability to group these into molecular categories (Fig. 302-2), CF has been deemed a condition ideally suited for personalized (i.e., mechanistically tailored) drug treatment. That being said, many CFTR variants clearly exhibit multiple molecular abnormalities (across more than one mechanistic subclass), and modulator compounds can therefore provide benefit across numerous disease subcategories. CFTR drug discovery—while highly successful—might, therefore, be viewed as less "personalized" or "precise" than originally envisioned. Moreover, clinical data indicate that a subset of individuals with F508del respond poorly to TCTs. Understanding the multifactorial determinants that govern drug effectiveness and/or risk of toxicity (e.g., due to genomic loci other than *CFTR*, epigenetic/environmental features, or complex *CFTR* alleles with numerous polymorphisms) constitutes a major objective in the field.

Other Challenges Involving CFTR Modulators and Progress Toward Nucleotide-Based Therapeutics The high cost of modulator compounds has often restricted third-party reimbursement to include only the specific genotypes for which FDA or other regulatory approval has been obtained. As a consequence, access to potentially efficacious modulators among patients with very rare CFTR defects and off-label prescribing are largely precluded. Moreover, clinical trials intended to expand the drug label can be difficult to arrange based on small numbers of patients carrying a particular ultra-rare variant. In vitro models rigorously shown to predict clinical modulator benefit have proven useful in this setting (e.g., studies of primary airway and other well-validated epithelial monolayers, or organoid-type cultures) and represent a potential means to gain FDA approval or insurance reimbursement for those with uncommon *CFTR* abnormalities.

CF drug discovery is emblematic of what might be accomplished in other refractory inherited conditions using an approach grounded in molecular etiology and unbiased compound library screening. Beyond CFTR modulators, genetic manipulation (e.g., *CFTR* gene transfer, DNA editing) and airway progenitor cell engraftment comprise experimental approaches that may be less dependent on a specific (i.e., personalized) pathogenic mechanism. For example, efficient, safe, and durable delivery of wild-type *CFTR* using viral (e.g., adeno-associated), lipid nanoparticle, or other vehicles represents a potential means to address diverse molecular abnormalities, independent of the responsible *CFTR* mutation(s). In that context, rescue of secretory epithelial progenitor cells has been emphasized, with newer technologies such as single-cell RNA-seq or spatial transcriptomics available to track successful nucleotide-based transduction. Approaches to genetic correction are of particular urgency for patients with forms of CF unresponsive to CFTR modulation, such as disease caused by premature truncation codons or disruptive splice site abnormalities.

TREATMENTS DIRECTED TOWARD CF SEQUELAE

Chronic Outpatient Management, Including Relationship to Modulators Standard care for patients with CF is intensive, with outpatient regimens that include exogenous pancreatic enzymes taken with meals, nutritional supplementation, anti-inflammatory medication, bronchodilators, and chronic or periodic dosing of oral or aerosolized antibiotics (e.g., as maintenance therapy for patients with *P. aeruginosa*). Recombinant DNAse aerosols (degrade DNA strands that contribute to mucus viscosity) and nebulized hypertonic saline or mannitol (augment PCL depth, activate mucociliary clearance, and mobilize inspissated airway secretions) have traditionally been administered. Chest physiotherapy several times each day is routinely used as a means to promote clearance of airway mucus, although rigorous evidence for clinical benefit is limited. Among adults with CF, intestinal malabsorption, chronic inflammation, and endocrine abnormalities can lead to poor bone mineralization requiring treatment with vitamin D, calcium, and other measures. The time, complexity, and expense of home care are considerable and take a significant toll on patients and their families. Important CF longitudinal trials are reevaluating the need for interventions, such as recombinant DNAse, and examining optimal guidelines for frequency of outpatient clinic visits, home monitoring, sputum culture surveillance, and other aspects of clinical care in the era of highly effective modulators.

Chronic sequelae of CF have received continued attention following the advent of HEMT, particularly since patients with established lung disease given TCT or other formulations continue to exhibit respiratory infection and inflammation despite substantial clinical improvement. Certain disease hallmarks, such as CF sinus involvement, are palliated by HEMT, and both exocrine pancreatic function and glucose tolerance have been shown to benefit from modulators in specific clinical settings. (Annual testing for diabetes mellitus continues to be recommended for adults with CF.) The impact of CFTR modulation has not been fully characterized for other extrapulmonary manifestations of the disease. Improved treatments that address ongoing nutritional deficits, hepatic and endocrine abnormalities, mucostasis, or additional features that persist despite modulators remain a priority. Sexual and reproductive health have become areas of considerable interest, since pregnancies are markedly increased among women on HEMT. Understanding potential relationship(s) between clinical depression, overall mental health, and chronic modulator therapy remains an important and emerging topic.

Pulmonary Exacerbation Severe CF respiratory exacerbation is commonly managed by hospital admission for parenteral antibiotics and frequent chest physiotherapy directed against (often multidrug-resistant) bacterial pathogens. Aggressive intervention in this setting can restore a large component of lung function, but ongoing and cumulative loss of pulmonary reserve has traditionally reflected the natural history of the disease. Poor prognostic indicators such as sputum culture containing *Burkholderia cepacia* complex, *Stenotrophomonas maltophilia*, *Achromobacter*, mucoid *P. aeruginosa*, or atypical mycobacteria are rigorously monitored in the CF patient population. Methicillin-resistant *S. aureus* in CF lungs may also be associated with less favorable outcomes. As noted above, while HEMT may diminish bacterial density in sputum samples, infection commonly persists despite modulator treatment. Typical inpatient antibiotic coverage includes combination drug therapy with an aminoglycoside and β-lactam (if *Pseudomonas* is present) for ~14 days. Maximal improvement in lung function is often achieved by 8–10 days, although optimal duration of therapy is a subject of continuing investigation. Many families elect parenteral antibiotic treatment at home, but additional studies are needed to evaluate specific drug combinations, duration of therapy, and home versus inpatient management. Other CF respiratory sequelae that may require hospitalization include hemoptysis and pneumothorax. Hypersensitivity to *Aspergillus* (allergic bronchopulmonary aspergillosis) occurs in ~5% of individuals with CF and should be suspected in the absence of a beneficial response to aggressive inpatient antibiotics.

Lung Transplantation For end-stage CF pulmonary failure, transplantation is a viable therapeutic option with median survival >9 years among adults with the disease. Determining optimal timing for surgery presents a substantial challenge in patients with severe respiratory compromise, particularly since the rate of continued functional decline, as well as individualized mortality risk from transplantation, can be difficult to predict. FEV_1 measurements <30% predicted, together with an assortment of other clinical parameters (e.g., hospitalization frequency, need for supplemental oxygen, modulator treatment), are employed as thresholds for transplant referral, although patients with conditions such as significant pulmonary hypertension may merit consultation at higher FEV_1. In general, evaluation for pulmonary transplant has often been underutilized among patients with CF. The decision is best approached based on early input from providers specializing in both CF clinical management and transplant medicine.

CYSTIC FIBROSIS QUALITY IMPROVEMENT AND HEALTHSPAN

As a direct result of advances in basic research, modulator and other therapies have transformed CF from a disease that historically led to death in early childhood to a condition with frequent survival well into the fifth decade of life and beyond. Initiating modulator treatment in young children with CF is expected to further promote longevity by forestalling pulmonary damage, although this assertion will require formal testing in patients on long-term HEMT. Care of aging individuals with CF (in which older patients with the disease experience obesity, cancer, vascular, and other age-related comorbidities commonly observed in the general population) has become an emerging priority. As modulatory therapies advance, standardized approaches to treatment will be essential. Well-defined protocols for general CF management are already widely established, including guidance regarding hospital admission, antibiotic regimens, nutritional intervention, periodicity of diagnostic tests, and other clinical parameters. These recommendations are implemented by specialized CF care centers and similarly accredited programs. Such measures have led to markedly improved pulmonary function, weight gain, body mass index, and other clinical endpoints among patients with the disease. The same approach is being applied to help optimize care of individuals given new CFTR modulators, as well as older patients with CF. Standardized protocols for CF therapy can be accessed at *https://www.cff.org/managing-cf/new-era-cf-care-possible-future-changes* or through a number of excellent reviews.

GLOBAL CONSIDERATIONS

Newborn screening for CF is universal throughout the United States and Canadian provinces, Australia, New Zealand, and most of Europe, and facilitates early intervention. Although the disease has traditionally been viewed as most common among whites (~1 in 3300 live U.S. births), there is growing concern that CF has been underdiagnosed in certain parts of the world, including regions such as Eastern Europe, Latin America, Asia, and India. CF in these areas—as well as among minority populations in the United States—is sometimes associated with rare variants that are less well defined and/or poorly characterized. In many clinical settings, therefore, enhanced attention to diagnosis is needed, particularly when specialized DNA testing and sweat chloride measurements are unavailable and patients with CF may be missed. Failure to identify CF properly in early childhood has important implications, since nutritional and other therapies at a young age are believed to promote quality of life and increase longevity. As one example, median survival among individuals with CF is <30 years in much of Latin America (compared to >60 years in the United States). The less favorable prognosis is attributable in part to lack of widespread diagnostic capabilities (i.e., newborn screening, sweat testing, and genetic analysis tailored to ethnic background), together with insufficient access to leading-edge, interdisciplinary treatment. Efforts to apply state-of-the-art management to underdiagnosed and underserved CF patient populations will help improve outcomes and mitigate CF health disparities in the future.

FURTHER READING

Bell SC et al: The future of cystic fibrosis care: A global perspective. Lancet Respir Med 8:65, 2020.

Farrell PM et al: The impact of the CFTR gene discovery on cystic fibrosis diagnosis, counseling, and preventive therapy. Genes 11:401, 2020.

Grasemann H, Ratjen F: Cystic fibrosis. N Engl J Med 389:1693, 2023.

Keating D et al: VX-445-tezacaftor-ivacator in patients with cystic fibrosis and one or two Phe508del alleles. N Engl J Med 379:1612, 2018.

Manfredi C et al: Making precision medicine personal for cystic fibrosis. Science 365:220, 2019.

Middleton PG et al: Elexacaftor-tezacaftor-ivacaftor for cystic fibrosis with a single Phe508del allele. N Engl J Med 381:1809, 2019.

Ramos KF et al: Survival and lung transplant outcomes for individuals with advanced cystic fibrosis lung disease living in the United States and Canada: An analysis of national registries. Chest 160:843, 2021.

Sosnay PR et al: Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. Nat Genet 45:1160, 2013.

Stoltz DA et al: Origins of cystic fibrosis lung disease. N Engl J Med 372:351, 2015.

VIDEO 302-1 Role of CFTR during airway mucociliary clearance. Initial video sequences depict establishment of the normal periciliary fluid layer bathing the surface airway epithelium, with spheres representing chloride and bicarbonate ions secreted through CFTR and across the apical (mucosal) respiratory surface. Later video describes failure of CFTR anion transport and resulting depletion of the periciliary layer, "plastering" of cilia against the mucosal surface, and accumulation of mucus in the airway with resulting bacterial infection. *(Reproduced with permission from Cystic Fibrosis Foundation.)*

VIDEO 302-2 Pharmacologic modulation of mutant CFTR. Initial video (***A***) illustrates CFTR encoding an ion transport gating (class III) defect. The cystic fibrosis (CF) gene product is localized to the plasma membrane but incapable of conducting anions (*yellow spheres*) until a potentiator molecule (shown in *green*) binds and facilitates channel opening. Later video (***B***) describes CFTR encoding a maturational processing (protein biogenesis, class II) defect. The mutant protein is misfolded, fails to traffic to the cell surface, and is degraded by the proteasome. Binding of corrector molecules (*red spheres*) improves folding and facilitates CFTR stabilization and cell surface localization/function. *(Reproduced with permission from Cystic Fibrosis Foundation.)*

303

Chronic Obstructive Pulmonary Disease

Craig P. Hersh, Edwin K. Silverman, Dawn DeMeo

Chronic obstructive pulmonary disease (COPD) is defined as a disease state characterized by persistent respiratory symptoms and airflow obstruction (*https://goldcopd.org/2024-gold-report/*). COPD includes *emphysema*, an anatomically defined condition characterized by destruction of the lung alveoli with air space enlargement; *chronic bronchitis*, a clinically defined condition with chronic cough and phlegm; and/or *small airway disease*, a condition in which small bronchioles are narrowed and reduced in number. The *classic definition* of COPD requires the presence of chronic airflow obstruction, determined by spirometry, that usually occurs in the setting of noxious environmental exposures, most commonly products of combustion, including cigarette smoking in the United States and biomass fuels in some other countries. The increasing prevalence of vaping and use of inhaled cannabis are of increasing concern, especially in adolescent populations. Other factors including abnormal lung development, respiratory infections, asthma, and genetics, can lead to COPD. Emphysema, chronic bronchitis, and small airway disease are present in varying degrees in different COPD patients. Patients with a history of cigarette smoking without chronic airflow obstruction may have chronic bronchitis, emphysema, respiratory symptoms including dyspnea, and acute exacerbations. Although these patients are not included within the classic definition of COPD, they may have similar disease processes. Investigators in the COPDGene study proposed a multidimensional approach to COPD diagnosis, which is based on domains of environmental exposures, respiratory symptoms, imaging abnormalities, and physiologic abnormalities.

COPD is the fourth leading cause of death and affects >15 million persons in the United States. COPD is also a disease of increasing public health importance around the world. Globally, there are an estimated 480 million individuals with COPD, with a projection of 592 million by 2050.

PATHOGENESIS AND PATHOLOGY

Airflow obstruction, the primary physiologic marker of COPD, can result from airway disease and/or emphysema. Cigarette smoke exposure may affect the large airways, small airways (≤2 mm diameter), and alveoli. Changes in large airways cause cough and sputum production, while changes in small airways and alveoli are responsible for physiologic alterations. Airway inflammation, small airway destruction, and the development of emphysema are present in most persons with COPD; however, their relative contributions to airflow obstruction vary from one person to another. The early development of chronic airflow obstruction is driven by small airway disease. Small airways may become narrowed by cells (hyperplasia and accumulation), mucus, and fibrosis. Advanced stages of COPD are typically characterized by extensive emphysema, although there are a small number of subjects with very severe airflow obstruction with airway disease and virtually no emphysema. The subjects at greatest risk of progression in COPD are those with both aggressive airway disease and emphysema.

■ LARGE AIRWAY DISEASE

Cigarette smoking often results in mucus gland enlargement and goblet cell hyperplasia, leading to cough and mucus production that define chronic bronchitis. Mucus plugs have been frequently observed on chest computed tomography (CT) scans of COPD patients, and they have recently been associated with increased mortality risk. In response to cigarette smoking or other inhaled exposures, goblet cells increase not only in number but also in extent through the bronchial tree. Bronchi also undergo squamous metaplasia, predisposing to carcinogenesis and disrupting mucociliary clearance. Although not as prominent as in asthma, patients with COPD may have bronchial hyperreactivity leading to airflow obstruction. Neutrophil influx has been associated with purulent sputum during respiratory tract infections. Independent of its proteolytic activity, neutrophil elastase is among the most potent secretagogues identified.

■ SMALL AIRWAY DISEASE

The major site of increased airflow resistance in most individuals with COPD is in airways ≤2 mm diameter. Characteristic cellular changes include goblet cell metaplasia, with these mucus-secreting cells replacing surfactant-secreting club cells. Smooth-muscle hypertrophy may also be present. Luminal narrowing can occur by fibrosis, excess mucus, edema, and cellular infiltration. Reduced surfactant may increase surface tension at the air-tissue interface, predisposing to airway narrowing or collapse. Respiratory bronchiolitis with mononuclear inflammatory cells collecting in distal airway tissues may cause proteolytic destruction of elastic fibers in the respiratory bronchioles and alveolar ducts where the fibers are concentrated as rings around alveolar entrances. Narrowing and drop-out of small airways precede the onset of emphysematous destruction. Advanced COPD has been shown to be associated with a loss of many of the smaller airways and a similar significant loss of the lung microvasculature.

■ LUNG PARENCHYMAL DESTRUCTION

Emphysema is characterized by destruction of gas-exchanging air spaces, i.e., the respiratory bronchioles, alveolar ducts, and alveoli. Large numbers of macrophages accumulate in respiratory bronchioles of essentially all smokers. Neutrophils, B lymphocytes, and

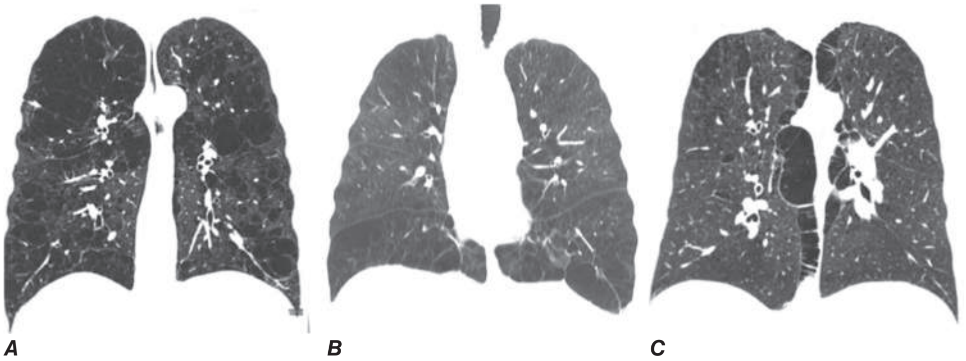

FIGURE 303-1 Computed tomography (CT) patterns of emphysema. ***A.*** Centrilobular emphysema with severe upper lobe involvement in a 68-year-old man with a 70-pack-year smoking history but forced expiratory volume in 1 s (FEV$_1$) 81% predicted (Global Initiative for Chronic Obstructive Lung Disease [GOLD] spirometry grade 1). ***B.*** Panlobular emphysema with diffuse loss of lung parenchymal detail predominantly in the lower lobes in a 64-year-old man with severe α_1-antitrypsin (α_1AT) deficiency. ***C.*** Paraseptal emphysema with marked airway inflammation in a 52-year-old woman with a 37-pack-year smoking history and FEV$_1$ 40% predicted.

T lymphocytes, particularly CD8+ cells, are also increased in the alveolar space of smokers. Alveolar walls become perforated and later obliterated with coalescence of the delicate alveolar structure into large emphysematous air spaces.

Emphysema is classified into distinct pathologic types, which include centrilobular, panlobular, and paraseptal **(Fig. 303-1)**. *Centrilobular emphysema*, the type most frequently associated with cigarette smoking, is characterized by enlarged air spaces found (initially) in association with respiratory bronchioles. Centrilobular emphysema is usually most prominent in the upper lobes and superior segments of lower lobes. *Panlobular emphysema* refers to abnormally large air spaces evenly distributed within and across acinar units. Panlobular emphysema is commonly observed in patients with α_1-antitrypsin (α_1AT) deficiency, which has a predilection for the lower lobes. Paraseptal emphysema is distributed along the pleural margins with relative sparing of the lung core or central regions. Except for α_1AT deficiency, pathobiological mechanisms related to different emphysema pathophenotypes are not well understood. Radiographic imaging can be used to quantify the amount and distribution of emphysema.

DISEASE MECHANISMS

Although the precise biological mechanisms leading to COPD have not been determined, a number of key cell types, molecules, and pathways have been identified from cell-based and animal model studies. The pathogenesis of emphysema (shown in **Fig. 303-2**) is more clearly defined than the pathogenesis of small airway disease. Pulmonary vascular destruction occurs in concert with small airway disease and emphysema.

The current dominant paradigm for the pathogenesis of emphysema comprises a series of four interrelated events: (1) Chronic inhaled exposures (typically cigarette smoke) in genetically susceptible individuals triggers inflammatory and immune cell recruitment within large and small airways and in the terminal air spaces of the lung. (2) Inflammatory cells release proteinases that damage the extracellular matrix supporting airways, vasculature, and gas exchange surfaces of the lung. (3) Structural cell death occurs through oxidant-induced

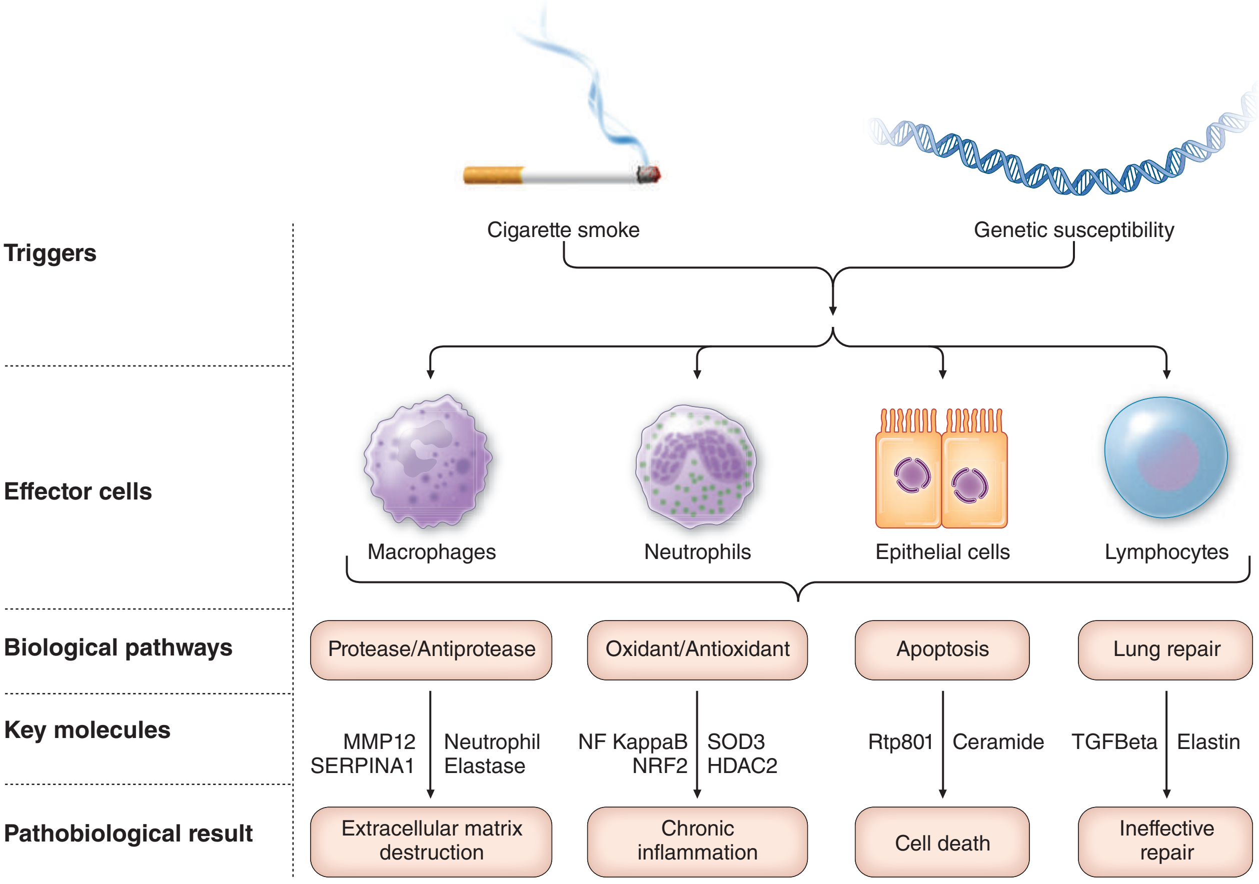
Triggers
Cigarette smoke
Genetic susceptibility
Effector cells
Macrophages
Neutrophils
Epithelial cells
Lymphocytes
Biological pathways
Protease/Antiprotease
Oxidant/Antioxidant
Apoptosis
Lung repair
Key molecules
MMP12
SERPINA1
Neutrophil
Elastase
NF KappaB
NRF2
SOD3
HDAC2
Rtp801
Ceramide
TGFBeta
Elastin
Pathobiological result
Extracellular matrix destruction
Chronic inflammation
Cell death
Ineffective repair

FIGURE 303-2 Pathogenesis of emphysema. Upon long-term exposure to cigarette smoke in genetically susceptible individuals, lung epithelial cells and T and B lymphocytes recruit inflammatory cells to the lung. Biological pathways of protease-antiprotease imbalance, oxidant/antioxidant imbalance, apoptosis, and lung repair lead to extracellular matrix destruction, cell death, chronic inflammation, and ineffective repair. Although most of these biological pathways influence multiple pathobiological results, only a single relationship between pathways and results is shown. A subset of key molecules related to these biological pathways is listed.

damage, cellular senescence, and proteolytic loss of cellular-matrix attachments leading to extensive loss of smaller airways, vascular pruning, and alveolar destruction. (4) Disordered repair of elastin and other extracellular matrix components contributes to air space enlargement and emphysema.

Inflammation and Extracellular Matrix Proteolysis Elastin, the principal component of elastic fibers, is a highly stable component of the extracellular matrix that is critical to the integrity of the lung. The elastase:antielastase hypothesis, proposed in the mid-1960s, postulated that the balance of elastin-degrading enzymes and their inhibitors determines the susceptibility of the lung to destruction, resulting in air space enlargement. This hypothesis was based on the clinical observation that patients with genetic deficiency in α_1AT, the inhibitor of the serine proteinase neutrophil elastase, were at increased risk of emphysema, and that instillation of elastases, including neutrophil elastase, into experimental animals resulted in emphysema. The elastase:antielastase hypothesis remains a prevailing mechanism for the development of emphysema. However, a complex network of immune and inflammatory cells and additional biological mechanisms that contribute to emphysema have subsequently been identified. Both innate and adaptive immune systems are likely involved in COPD pathogenesis. Upon exposure to oxidants from cigarette smoke, lung macrophages and epithelial cells become activated, producing proteinases and chemokines that attract other inflammatory and immune cells. Single-cell transcriptomics revealed that a subset of alveolar type 2 cells is the predominant site of lung gene expression of *HHIP*, one of the top COPD susceptibility genes identified by genome-wide association studies. Oxidative stress is a key component of COPD pathobiology; the transcription factor NRF2, a major regulator of oxidant-antioxidant balance, and SOD3, a potent antioxidant, have been implicated in emphysema pathogenesis by animal models. Mitochondrial dysfunction in COPD may worsen oxidative stress. One mechanism of macrophage activation occurs via oxidant-induced inactivation of histone deacetylase-2 (HDAC2), shifting the balance toward acetylated or open chromatin, exposing nuclear factor-κB sites, and resulting in transcription of matrix metalloproteinases and proinflammatory cytokines such as interleukin 8 (IL-8) and tumor necrosis factor α (TNF-α); this leads to neutrophil recruitment. CD8+ T cells are also recruited in response to cigarette smoke and release interferon-inducible protein-10 (IP-10, CXCL-7), which in turn leads to macrophage production of macrophage elastase (matrix metalloproteinase-12 [MMP12]).

Matrix metalloproteinases and serine proteinases, most notably neutrophil elastase, work together by degrading the inhibitor of the other, leading to lung destruction. Extracellular vesicles induced by cigarette smoke and expressing neutrophil elastase or MMP12 on their surface have been recently implicated as important sources of these destructive proteinases. Proteolytic cleavage products of elastin serve as a macrophage chemokine, and proline-glycine-proline (generated by proteolytic cleavage of collagen) is a neutrophil chemokine, fueling this damaging positive feedback loop. Elastin degradation and disordered repair are thought to be primary mechanisms in the development of emphysema.

There is some evidence that autoimmune mechanisms may promote the progression of disease. Lymphoid follicles composed of B cells and T cells are present around the airways of COPD patients, particularly in patients with advanced disease. T$_H$1 and T$_H$17 lymphocytes may activate innate immune cells (including neutrophils, macrophages, and innate lymphoid cells), contributing to a cycle of chronic inflammation. Antibodies have been found against elastin fragments as well; IgG autoantibodies with avidity for pulmonary epithelium and the potential to mediate cytotoxicity have been detected.

Concomitant cigarette smoke–induced loss of cilia in the airway epithelium and impaired macrophage phagocytosis predispose to bacterial infection with neutrophilia. In end-stage lung disease, long after smoking cessation, there remains an exuberant inflammatory response, suggesting that cigarette smoke–induced inflammation both initiates the disease and, in susceptible individuals, establishes a chronic process that can continue disease progression even after smoking cessation.

Cell Death Cigarette smoke oxidant-mediated structural cell death occurs via a variety of mechanisms including excessive ceramide production and Rtp801 inhibition of mammalian target of rapamycin (mTOR), leading to cell death as well as inflammation and proteolysis. Involvement of mTOR and other cellular senescence markers has led to the concept that emphysema resembles premature aging of the lung; emphysema has been identified in the lung tissue of lifetime never smokers of advanced age. Heterozygous gene targeting of hedgehog interacting protein (*HHIP*) in a murine model leads to aging-related emphysema.

Ineffective Repair The ability of the adult lung to replace lost smaller airways and microvasculature and to repair damaged alveoli appears limited. Uptake of apoptotic cells by macrophages normally results in production of growth factors and dampens inflammation, promoting lung repair. Cigarette smoke and other inhaled exposures impair macrophage uptake of apoptotic cells, limiting repair. It is unlikely that the intricate and dynamic process of septation that is responsible for alveologenesis during lung development can be reinitiated in the adult human lung.

PATHOPHYSIOLOGY

Persistent irreversible reduction in forced expiratory flow rates is the classic definition of COPD. Hyperinflation with increases in the residual volume and the residual volume/total lung capacity ratio, nonuniform distribution of ventilation, and ventilation-perfusion mismatching also occur.

■ AIRFLOW OBSTRUCTION

Airflow obstruction, also known as airflow limitation, is typically determined for clinical purposes by spirometry, which involves maximal forced expiratory maneuvers after the subject has inhaled to total lung capacity. Key parameters obtained from spirometry include the volume of air exhaled within the first second of the forced expiratory maneuver (FEV$_1$) and the total volume of air exhaled during the entire spirometric maneuver (forced vital capacity [FVC]). Patients with fixed airflow obstruction have a chronically reduced ratio of FEV$_1$/FVC that does not normalize following bronchodilator treatment.

■ HYPERINFLATION

Lung volumes are also commonly assessed in pulmonary function testing. In COPD, there is often "air trapping" (increased residual volume and increased ratio of residual volume to total lung capacity) and progressive hyperinflation (increased total lung capacity) in more advanced disease. Hyperinflation of the thorax during tidal breathing preserves maximum expiratory airflow because as lung volume increases, elastic recoil pressure increases and airways enlarge so that airway resistance decreases.

Despite compensating for airway obstruction, hyperinflation can push the diaphragm into a flattened position with a number of adverse effects. First, by decreasing the zone of apposition between the diaphragm and the abdominal wall, positive abdominal pressure during inspiration is not applied as effectively to the chest wall, hindering rib cage movement and impairing inspiration. Second, because the muscle fibers of the flattened diaphragm are shorter than those of a more normally curved diaphragm, they are less capable of generating normal inspiratory pressures. Third, the flattened diaphragm must generate greater tension to develop the transpulmonary pressure required to produce tidal breathing. Fourth, the thoracic cage is distended beyond its normal resting volume, and during tidal breathing, the inspiratory muscles must do work to overcome the resistance of the thoracic cage to further inflation instead of gaining the normal assistance from the chest wall recoiling outward toward its resting volume.

■ GAS EXCHANGE

Although there is considerable variability in the relationships between the FEV$_1$ and other physiologic abnormalities in COPD, certain generalizations may be made. The partial pressure of oxygen in arterial blood Pa$_{\text{O}_2}$ usually remains near normal until the FEV$_1$ is decreased to below 50% of predicted, and even much lower FEV$_1$ values can be associated with a normal PaO_2, at least at rest. An elevation of arterial level of carbon dioxide (PaCO_2) is not expected until the FEV$_1$ is <25% of predicted and even then may not occur. Pulmonary arterial hypertension severe enough to cause cor pulmonale and right ventricular failure due to COPD typically occurs in individuals who have marked decreases in FEV$_1$ (<25% of predicted) and chronic hypoxemia (PaO_2 <55 mmHg); however, some patients develop significant pulmonary arterial hypertension likely due to vascular destruction, independent of COPD severity **(Chap. 294)**.

Nonuniform ventilation and ventilation-perfusion mismatching are characteristic of COPD, reflecting the heterogeneous nature of the disease process within the airways and lung parenchyma. Physiologic studies are consistent with multiple parenchymal compartments having different rates of ventilation due to regional differences in compliance and airway resistance. Ventilation-perfusion mismatching accounts for essentially all of the reduction in PaO_2 that occurs in COPD; shunting is minimal. This finding explains the effectiveness of modest elevations of inspired oxygen in treating hypoxemia due to COPD and therefore the need to consider problems other than COPD when hypoxemia is difficult to correct with modest levels of supplemental oxygen.

RISK FACTORS

CIGARETTE SMOKING

By 1964, the Advisory Committee to the Surgeon General of the United States had concluded that cigarette smoking was a major risk factor for mortality from chronic bronchitis and emphysema. Subsequent longitudinal studies have shown accelerated decline in FEV$_1$ in a dose-response relationship to the intensity of cigarette smoking, which is typically expressed as pack-years (average number of packs of cigarettes smoked per day multiplied by the total number of years of smoking). This dose-response relationship between reduced pulmonary function and cigarette smoking intensity accounts, at least in part, for the higher prevalence rates of COPD with increasing age. The historically higher rate of smoking among males is the likely explanation for the historically higher prevalence of COPD among males; however, the prevalence of COPD among females has increased to the point that it is the same or exceeds the prevalence of COPD in men.

Although the causal relationship between cigarette smoking and the development of COPD has been absolutely proved, there is considerable individual variability in the response to smoking. Pack-years of cigarette smoking is the most highly significant predictor of FEV$_1$ **(Fig. 303-3)**, but only 15% of the variability in FEV$_1$ is explained by pack-years. This finding suggests that additional developmental, environmental, and/or genetic factors contribute to the impact of smoking on the development of chronic airflow obstruction. Nonetheless, many patients with a history of cigarette smoking with normal spirometry have evidence for worse health-related quality of life, reduced exercise capacity, and emphysema and/or airway disease on chest CT scans; thus, they have not escaped the harmful effects of cigarette smoking and may present with respiratory exacerbations despite normal spirometry. While they do not meet the classic definition of COPD based on population normals for FEV$_1$ and FEV$_1$/FVC, studies have shown that these subjects overall have a shift toward lower FEV$_1$ values, which is consistent with obstruction on an individual level.

Although cigar and pipe smoking may also be associated with the development of COPD, the evidence supporting such associations is less compelling, likely related to the lower dose of inhaled tobacco by-products during cigar and pipe smoking. The impact of electronic cigarettes and vaping on the development and progression of COPD is a growing area of concern; emerging literature identifies increased risk for pulmonary symptoms. Inhaled cannabis use is increasing in prevalence and may increase risk for cough, sputum production, and wheeze. However, it remains unclear whether there is an elevated risk for COPD from cannabis inhalation. A large study of lung function observed heavy cannabis use associated with steeper decline in FEV$_1$, although other studies have not observed accelerated lung function decline.

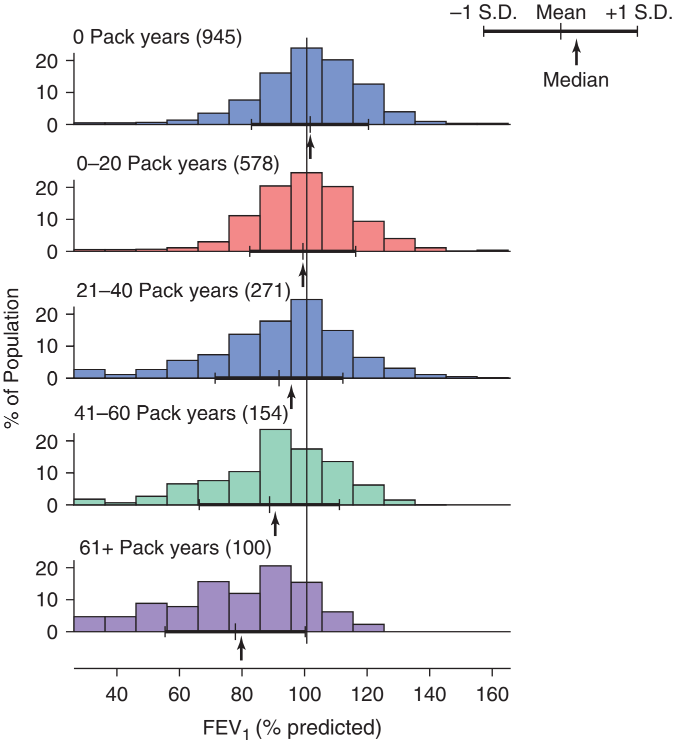

FIGURE 303-3 Distributions of forced expiratory volume in 1 s (FEV$_1$) values in a general population sample, stratified by pack-years of smoking. Means, medians, and ±1 standard deviation of percent predicted FEV$_1$ are shown for each smoking group. Although a dose-response relationship between smoking intensity and FEV$_1$ was found, marked variability in pulmonary function was observed among subjects with similar smoking histories. S.D., standard deviation. *(Reproduced with permission from B Burrows: Quantitative relationships between cigarette smoking and ventilatory function. Am Rev Respir Dis 115:195, 1977.)*

AIRWAY RESPONSIVENESS AND COPD

A tendency for increased reversible bronchoconstriction in response to a variety of exogenous stimuli, including methacholine and histamine, is one of the defining features of asthma **(Chap. 298)**. However, many patients with COPD also share this feature of airway hyperresponsiveness. In older subjects, there is considerable overlap between persons with a history of chronic asthma and COPD in terms of airway responsiveness, airflow obstruction, and pulmonary symptoms. The origin of asthma is viewed in many patients as an allergic disease, while COPD is thought to primarily result from smoking-related inflammation and damage; however, they likely share common environmental and genetic factors, and the chronic form in older subjects can present similarly. This is particularly relevant for childhood asthmatic subjects who chronically smoke throughout adulthood.

Longitudinal studies that compared airway responsiveness to subsequent decline in pulmonary function have demonstrated that increased airway responsiveness is clearly a significant predictor of subsequent decline in pulmonary function. A study from the Childhood Asthma Management Program identified four lung function trajectories in children with persistent asthma. Asthmatics with reduced lung function early in life were more likely to develop persistent airflow obstruction in early adulthood. Both asthma and airway hyperresponsiveness are risk factors for COPD.

RESPIRATORY INFECTIONS

The impact of adult respiratory infections on decline in pulmonary function is controversial, but significant long-term reductions in pulmonary function are not typically seen following an individual episode of acute bronchitis or pneumonia. However, respiratory infections are important causes of COPD exacerbations, and results from the COPDGene and ECLIPSE studies suggest that COPD exacerbations are associated with increased loss of lung function longitudinally, particularly among those individuals with better baseline lung function levels. The impact of the effects of childhood respiratory illnesses on the subsequent development of COPD has been difficult to assess due to a lack of adequate longitudinal data, but recent studies have suggested that childhood pneumonia may lead to increased risk for COPD later in life. Globally, tuberculosis infection has been associated with chronic airflow limitation and may be an important risk factor in the absence of a personal tobacco use history; tuberculosis-associated COPD may present with both emphysematous and fibrotic changes in the lung and with extensive airway remodeling. Infection with human immunodeficiency virus (HIV) leads to an increased risk for COPD and emphysema, especially with lower CD4 cell counts; COPD is frequently underdiagnosed in this setting.

■ OCCUPATIONAL EXPOSURES

Increased respiratory symptoms and airflow obstruction have been suggested to result from exposure to vapors, gas, dust, or fumes (VGDF). Several specific occupational exposures, including coal mining, gold mining, and cotton textile dust, have been implicated as risk factors for chronic airflow obstruction. Although nonsmokers in these occupations can develop some reductions in FEV$_1$, the importance of dust exposure as a risk factor for COPD, independent of cigarette smoking, is not certain for most of these exposures. However, among coal miners, coal mine dust exposure was a significant risk factor for emphysema in both smokers and nonsmokers. Given general underrecognition of VGDF and other occupational drivers and exacerbants of COPD, clinical history taking should include past and current occupational exposures, and referral to occupational health specialists as relevant.

■ AIR POLLUTION AND CLIMATE CHANGE

Some investigators have reported increased respiratory symptoms in those living in urban compared to rural areas, which may relate to increased ambient air pollution in urban settings, especially fine and ultrafine particulate pollution. However, the relationship of air pollution to chronic airflow obstruction remains unproved. Prolonged exposure to smoke produced by biomass combustion—a common mode of cooking in some countries—is a significant risk factor for COPD, particularly among women. The future impact of global warming on COPD is not clear, although increased extreme hot days are associated with respiratory exacerbations.

■ EARLY LIFE PASSIVE SMOKE EXPOSURE

Maternal tobacco smoking during in utero lung development contributes to significant reductions in postnatal lung development and pulmonary function. Exposure of children to environmental smoke related to tobacco and biomass fuel may also result in reduced lung growth and respiratory exacerbations. Early life exposure to environmental tobacco smoke has been linked to emphysema in adulthood.

■ GENETICS

Although cigarette smoking is the major environmental risk factor for the development of COPD, the development of airflow obstruction in smokers is highly variable. Severe α_1AT deficiency is a proven genetic risk factor for COPD; there is increasing evidence that other genetic determinants also exist.

α_1-Antitrypsin Deficiency Many variants of the protease inhibitor (PI or *SERPINA1*) gene that encodes α_1AT have been described. The common M allele is associated with normal α_1AT levels. The S allele, associated with slightly reduced α_1AT levels, and the Z allele, associated with markedly reduced α_1AT levels, also occur with frequencies of >1% in most white populations. Rare individuals inherit null alleles, which lead to the absence of any α_1AT production through a heterogeneous collection of mutations. Individuals with two Z alleles or one Z and one null allele are referred to as Pi^Z, which is the most common form of severe α_1AT deficiency.

Although only ~1% of COPD patients are found to have severe α_1AT deficiency as a contributing cause of COPD, these patients demonstrate that genetic factors can have a profound influence on the susceptibility for developing COPD. Pi^Z individuals often develop early-onset COPD, but the ascertainment bias in the published series of Pi^Z individuals—which have usually included many Pi^Z subjects who were tested for α_1AT deficiency because they had COPD—means that the fraction of Pi^Z individuals who will develop COPD and the age-of-onset distribution for the development of COPD in Pi^Z subjects remain unknown. Approximately 1 in 3000 individuals in the United States inherits severe α_1AT deficiency, but only a small minority of these individuals has been identified. The clinical laboratory test used most frequently to test for α_1AT deficiency is measurement of the immunologic level of α_1AT in serum (see "Laboratory Findings").

A significant percentage of the variability in pulmonary function among Pi^Z individuals is explained by cigarette smoking; cigarette smokers with severe α_1AT deficiency are more likely to develop COPD at early ages. However, the development of COPD in Pi^Z subjects, even among current or ex-smokers, is not absolute. Among Pi^Z nonsmokers, impressive variability has been noted in the development of airflow obstruction. Asthma and male gender also appear to increase the risk of COPD in Pi^Z subjects. Other genetic and/or environmental factors likely contribute to this variability.

Specific treatment in the form of α_1AT augmentation therapy is available for severe α_1AT deficiency as a weekly IV infusion (see "Treatment," below).

Several recent studies have demonstrated that heterozygous Pi^{MZ} subjects (who have intermediate serum levels of α_1AT) who smoke are at increased risk for the development of COPD. However, α_1AT augmentation therapy is not recommended for use in Pi^{MZ} subjects.

Other Genetic Risk Factors Studies of pulmonary function measurements performed in general population samples have indicated that genetic factors other than PI type influence variation in pulmonary function. Familial aggregation of airflow obstruction within families of COPD patients has also been demonstrated.

Genome-wide association studies (GWAS) have identified >80 regions of the genome that contain COPD susceptibility loci, including a region near the *HHIP* gene on chromosome 4, a cluster of genes on chromosome 15 (including components of the nicotinic acetylcholine receptor and another gene, *IREB2*, related to mitochondrial iron regulation), and *FAM13A* on chromosome 4, which is involved in Wnt/beta-catenin signaling. As with most other complex diseases, the risk associated with individual GWAS loci is modest, but these genetic determinants may identify important biological pathways related to COPD. Gene-targeted murine models for *HHIP*, *FAM13A*, and *IREB2* exposed to chronic cigarette smoke had altered emphysema susceptibility, suggesting that those genes are likely to be involved in COPD pathogenesis. An elevated polygenic risk score, which sums the effects of multiple genetic variants, is associated with increased COPD risk.

NATURAL HISTORY

The effects of cigarette smoking on pulmonary function appear to depend on the intensity of smoking exposure, the timing of smoking exposure during lung growth and development, and the baseline lung function of the individual; other environmental factors may have similar effects. Most individuals follow a steady trajectory of increasing pulmonary function with growth during childhood and adolescence, followed by a plateau in early adulthood, and then gradual decline with aging starting in the fourth decade of life. Individuals appear to track in their quantile of pulmonary function based on environmental and genetic factors that put them on different tracks. The risk of eventual mortality from COPD is closely associated with reduced levels of FEV$_1$ and the presence of emphysema. A graphic depiction of the natural history of COPD is shown as a function of the influences on tracking curves of FEV$_1$ in **Fig. 303-4**. Death or disability from COPD can result from a normal rate of decline after a reduced growth phase (curve *C*), an early initiation of pulmonary function decline after normal growth (curve *B*), or an accelerated decline after normal growth (curve *D*). Although accelerated rates of lung function decline have classically been associated with COPD, analyses of several population-based cohorts demonstrated that many subjects with fixed airflow obstruction had reduced growth but normal rates of lung function decline.

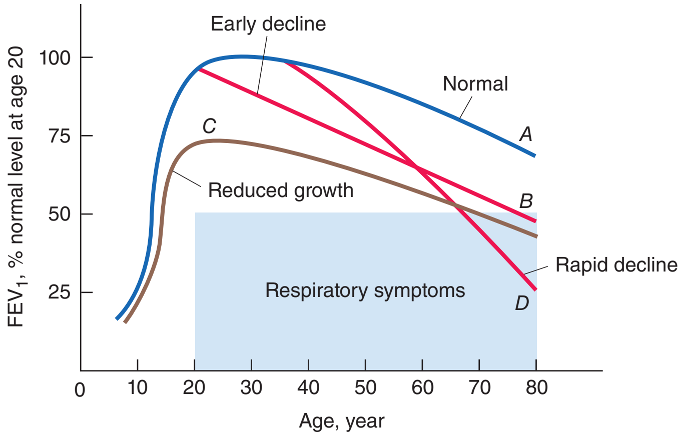

FIGURE 303-4 Hypothetical tracking curves of forced expiratory volume in 1 s (FEV_1) for individuals throughout their life spans. The normal pattern of growth and decline with age is shown by curve *A*. Significantly reduced FEV_1 (<65% of predicted value at age 20) can develop from a normal rate of decline after a reduced pulmonary function growth phase (curve *C*), early initiation of pulmonary function decline after normal growth (curve *B*), or accelerated decline after normal growth (curve *D*). *(From B Rijcken: Doctoral dissertation, p 133, University of Groningen, 1991.)*

The rate of decline in pulmonary function can be modified by changing environmental exposures (i.e., quitting smoking), with smoking cessation at an earlier age providing a more beneficial effect than smoking cessation after marked reductions in pulmonary function have already developed. The absolute annual loss in FEV_1 tends to be highest in mild COPD and lowest in very severe COPD. Multiple genetic factors influence the level of pulmonary function achieved during growth.

In chronic smokers, a group at risk for COPD (sometimes called pre-COPD) based on substantial chest CT changes (emphysema and airway wall thickening) has been identified in subjects with normal physiology (normal FEV_1 and FEV_1/FVC). COPD in these subjects commonly progresses in two primary patterns. Subjects with an emphysema-predominant pattern show emphysema early and classically progress through COPD severity grades. Subjects with an airway disease–predominant pattern typically show initial evidence of airway inflammation and progress with little emphysema early as FEV_1 falls while retaining a normal FEV_1/FVC ratio. This is termed *preserved ratio–impaired spirometry* (PRISm) physiology. The natural history of patients classified as having PRISm revealed an increased risk of mortality and respiratory and cardiovascular events. Although advanced age is an important risk factor for COPD, early COPD is an important area of research given advances in understanding cellular and molecular features and the importance of slowing lung function decline.

CLINICAL PRESENTATION

HISTORY

The three most common symptoms in COPD are cough, sputum production, and exertional dyspnea. Many patients have such symptoms for months or years before seeking medical attention. Although the development of airflow obstruction is a gradual process, many patients date the onset of their disease to an acute illness or exacerbation. A careful history, however, usually reveals the presence of respiratory symptoms prior to the acute exacerbation. The development of exertional dyspnea, often described as increased effort to breathe, heaviness, air hunger, or gasping, can be insidious. It is best elicited by a careful history focused on typical physical activities and how the patient's ability to perform them has changed. Activities involving significant arm work, particularly at or above shoulder level, are particularly difficult for many patients with COPD. Conversely, activities that allow the patient to brace the arms and use accessory muscles of respiration are better tolerated. Examples of such activities include pushing a shopping cart or walking on a treadmill. As COPD advances, the principal feature is worsening dyspnea on exertion with increasing intrusion on the ability to perform vocational or avocational activities. In the most advanced stages, patients are breathless doing basic activities of daily living. Validated questionnaires such as the COPD Assessment Test (CAT) and Modified Medical Research Council dyspnea scale may be used to reliably capture symptoms and activity limitations.

Accompanying worsening airflow obstruction is an increased frequency of exacerbations (described below). Patients may also develop resting hypoxemia and require institution of supplemental oxygen. A thorough history must consider symptoms of common comorbidities, such cardiovascular disease, gastroesophageal reflux, osteoporosis, frailty, depression, and anxiety.

PHYSICAL EXAMINATION

In the early stages of COPD, patients usually have an entirely normal physical examination. In patients with more severe disease, the physical examination of the lungs is notable for a prolonged expiratory phase and may include expiratory wheezing. In addition, signs of hyperinflation include a barrel chest and enlarged lung volumes with poor diaphragmatic excursion as assessed by percussion. Patients with severe airflow obstruction may also exhibit use of accessory muscles of respiration, sitting in the characteristic "tripod" position to facilitate the actions of the sternocleidomastoid, scalene, and intercostal muscles. Patients may develop cyanosis, visible in the lips and nail beds.

Advanced disease may be accompanied by cachexia, with significant weight loss and diffuse loss of subcutaneous adipose tissue. This syndrome has been associated with both inadequate oral intake and elevated levels of inflammatory cytokines (e.g., TNF-α). Such wasting is an independent poor prognostic factor in COPD.

Signs of overt right heart failure, termed *cor pulmonale*, are relatively infrequent since the advent of supplemental oxygen therapy, but pulmonary hypertension must be considered in patients with persistent activity limitations or lower extremity edema.

Clubbing of the digits is not a sign of COPD, and its presence should alert the clinician to initiate an investigation for causes of clubbing. In COPD patients, the development of lung cancer is the most likely explanation for newly developed clubbing.

LABORATORY FINDINGS

The hallmark of COPD is airflow obstruction (discussed above). Pulmonary function testing shows airflow obstruction with a reduction in FEV_1 and FEV_1/FVC **(Chap. 296)**. With worsening disease severity, lung volumes may increase, resulting in an increase in total lung capacity, functional residual capacity, and residual volume. In patients with emphysema, the diffusing capacity may be reduced, reflecting the lung parenchymal destruction characteristic of the disease. The degree of airflow obstruction is an important prognostic factor in COPD. For spirometric severity grading, the Global Initiative for Chronic Obstructive Lung Disease (GOLD) has consistently used thresholds of FEV_1 as a percentage of predicted value based on population reference equations. The American Thoracic Society and European Respiratory Society have recently advocated grading severity of lung function impairment using population-defined Z-scores **(Table 303-1)**. Pulmonary function interpretation has traditionally used race-specific reference values; however, race-neutral interpretation of pulmonary function tests may be preferable. Although the degree of airflow obstruction *generally* correlates with the presence and severity of respiratory symptoms, exacerbations, emphysema, and hypoxemia, the correlations are far from perfect. Thus, clinical features should be carefully assessed in each individual patient with COPD to determine the most appropriate therapies. It has been shown that a multifactorial index (BODE), incorporating body mass index, airflow obstruction, dyspnea, and exercise performance, is a better predictor of mortality. Exercise capacity can be quantified by the distance a patient is able to walk in 6 min. The GOLD COPD classification system incorporates respiratory symptoms and exacerbation history; these metrics are used to guide COPD treatment (see below).

Arterial blood gases and oximetry may demonstrate resting or exertional hypoxemia. Arterial blood gases provide additional information about alveolar ventilation and acid-base status by measuring arterial

TABLE 303-1 Comparison of GOLD and ATS/ERS Criteria for Severity of Airflow Obstruction

GOLD		ERS/ATS	
SEVERITY	**SPIROMETRY**	**SEVERITY**	**SPIROMETRY**
Mild	FEV$_1$/FVC <0.7 and FEV$_1$ $\geq 80\%$ predicted	[a]	FEV$_1$/FVC <LLN and zFEV$_1$ > -1.645
Moderate	FEV$_1$/FVC <0.7 and FEV$_1$ $\geq 50\%$ but $<80\%$ predicted	Mild	FEV$_1$/FVC <LLN and zFEV$_1$ between –1.65 and –2.5
Severe	FEV$_1$/FVC <0.7 and FEV$_1$ $\geq 30\%$ but $<50\%$ predicted	Moderate	FEV$_1$/FVC <LLN and zFEV$_1$ between –2.51 and –4
Very severe	FEV$_1$/FVC <0.7 and FEV$_1$ $<30\%$ predicted	Severe	FEV$_1$/FVC <LLN and zFEV$_1$ < -4.1

[a] The ERS/ATS document does not provide a term for this category. Some authors have called it "borderline" or "minimal."

Abbreviations: ATS, American Thoracic Society; COPD, chronic obstructive pulmonary disease; ERS, European Respiratory Society; FEV$_1$, forced expiratory volume in 1 s; FVC, forced vital capacity; GOLD, Global Initiative for Chronic Obstructive Lung Disease; LLN, lower limit of normal; zFEV$_1$, z-score of FEV$_1$.

$P\text{CO}_2$ and pH. The change in pH with $P\text{CO}_2$ is 0.08 units/10 mmHg acutely and 0.03 units/10 mmHg in the chronic state. Knowledge of the arterial pH therefore allows the classification of ventilatory failure, defined as $P\text{CO}_2$ >45 mmHg, into acute or chronic conditions with acute respiratory failure being associated with acidemia. The arterial blood gas is an important component of the evaluation of patients presenting with symptoms of an exacerbation. An elevated hematocrit suggests the presence of chronic hypoxemia, as does the presence of signs of right ventricular hypertrophy. The blood eosinophil count, measured on a complete blood count with differential, is used to guide therapy, with the goal of reducing exacerbation risk

Radiographic studies may assist in the classification of the type of COPD. Chest x-ray findings may be consistent with COPD but cannot be used reliably to make the diagnosis. Increased lung volumes and flattening of the diaphragm suggest hyperinflation but do not provide information about chronicity of the changes. Obvious bullae, paucity of parenchymal markings, or hyperlucency on chest x-ray suggests the presence of emphysema. Chest CT scan is the current definitive test for establishing the presence or absence of emphysema, the pattern of emphysema, and the presence of significant disease involving medium and large airways (Fig. 303-1). It also enables the discovery of coexisting interstitial lung disease and bronchiectasis. Smokers with COPD are at high risk for development of lung cancer; annual low-dose chest CT scans for lung cancer screening have been demonstrated to reduce mortality in selected current and former smokers. In advanced COPD, CT scans can help determine the possible value of surgical or bronchoscopic therapy (described below).

Guidelines suggest testing for α_1AT deficiency in all subjects with COPD or asthma with chronic airflow obstruction. Measurement of the serum α_1AT level is the usual initial test. For subjects with low α_1AT levels, the definitive diagnosis of α_1AT deficiency requires PI type determination. This is typically performed by isoelectric focusing of serum or plasma, which reflects the genotype at the PI locus for the common alleles and many of the rare PI alleles as well. Mass spectrometry or molecular genotyping can be performed for the common PI variants (M, S, and Z), and DNA sequencing can detect other rare deficiency variants.

TREATMENT

Chronic Obstructive Pulmonary Disease

STABLE COPD

The two main goals of COPD therapy are to provide symptomatic relief (reduce respiratory symptoms, improve exercise tolerance, and improve health status) and reduce future risk (prevent disease progression, prevent and treat exacerbations, and reduce mortality). The institution of therapies should be based on symptom assessment, benefits of therapy, potential risks, and costs. **Figure 303-5** provides the currently suggested assessment of COPD patients based on spirometry, respiratory symptoms and risk for exacerbations. Response to therapy should be assessed, and decisions should be made whether or not to continue or alter treatment.

Three interventions—smoking cessation, oxygen therapy in chronically hypoxemic patients, and lung volume reduction surgery (LVRS) in selected patients with emphysema—have been demonstrated to improve survival of patients with COPD. Evidence is less strong that other nonpharmacologic interventions also reduce mortality, such as pulmonary rehabilitation after a COPD hospitalization, noninvasive positive-pressure ventilation in severe hypercapnia, and possibly lung transplantation. Triple inhaled therapy (long-acting beta-agonist bronchodilator, long-acting muscarinic antagonist bronchodilator, and inhaled corticosteroid) reduces mortality in selected patients with COPD.

PHARMACOTHERAPY

***Smoking Cessation* (See also Chap. 465)** It has been shown that middle-aged smokers who were able to successfully stop smoking experienced a significant improvement in the rate of decline in pulmonary function, often returning to annual changes similar to that of nonsmoking patients. In addition, smoking cessation improves survival. Thus, all smokers with COPD should be strongly urged to quit smoking and educated about the benefits of quitting. An emerging body of evidence demonstrates that combining pharmacotherapy with traditional supportive approaches considerably enhances the chances of successful smoking cessation. There are three principal pharmacologic approaches to the problem: nicotine replacement therapy available as gum, transdermal patch, lozenge, inhaler, and nasal spray; bupropion; and varenicline, a nicotinic acid receptor agonist/antagonist. The use of electronic cigarettes as a harm reduction strategy remains controversial. Current recommendations from the U.S. Surgeon General are that all adult, nonpregnant smokers considering quitting be offered pharmacotherapy, in the absence of any contraindication to treatment. Smoking cessation counseling is also recommended, and free counseling is available through state Smoking Quitlines.

Bronchodilators In general, inhaled bronchodilators are the primary treatment for almost all patients with COPD and are used for symptomatic benefit and to reduce exacerbation risk. In symptomatic patients, both regularly scheduled use of long-acting agents and as-needed short-acting medications are indicated. **Figure 303-6** provides suggestions for prescribing inhaled medication therapy based on grouping patients by severity of symptoms and risk of exacerbations.

Muscarinic Antagonists Short-acting ipratropium bromide improves symptoms with acute improvement in FEV$_1$. Long-acting muscarinic antagonists (LAMA; including aclidinium, glycopyrrolate, glycopyrronium, revefenacin, tiotropium, and umeclidinium) improve symptoms and reduce exacerbations. Side effects are minor; dry mouth is the most frequent.

Beta Agonists Short-acting beta agonists ease symptoms with acute improvements in lung function. Long-acting beta agonists (LABAs) provide symptomatic benefit and reduce exacerbations, though to a lesser extent than a LAMA. Currently available long-acting inhaled beta agonists are arformoterol, formoterol, indacaterol, olodaterol, salmeterol, and vilanterol. The main side effects are tremor and tachycardia.

Combinations of Beta Agonist–Muscarinic Antagonist Combination inhaled LABA and LAMA therapy has been demonstrated to

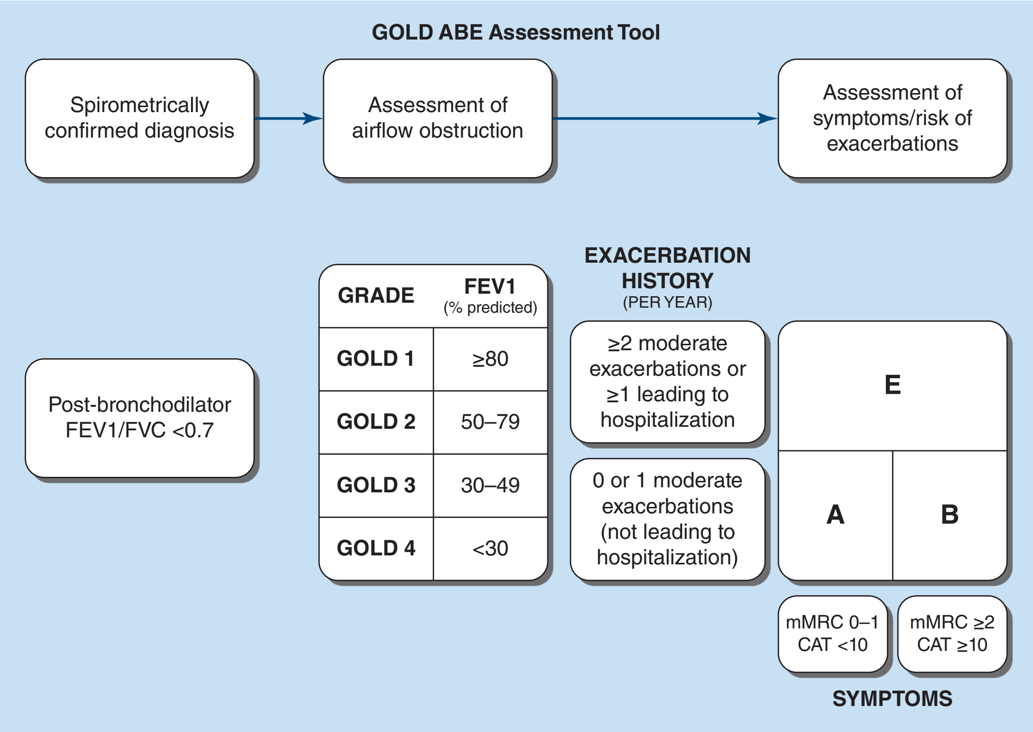
GOLD ABE Assessment Tool
Spirometrically confirmed diagnosis
Assessment of airflow obstruction
Assessment of symptoms/risk of exacerbations
Post-bronchodilator FEV1/FVC <0.7
GRADE
FEV1 (% predicted)
GOLD 1 ≥80
GOLD 2 50–79
GOLD 3 30–49
GOLD 4 <30
EXACERBATION HISTORY (PER YEAR)
≥2 moderate exacerbations or ≥1 leading to hospitalization
0 or 1 moderate exacerbations (not leading to hospitalization)
E
A
B
mMRC 0–1 CAT <10
mMRC ≥2 CAT ≥10
SYMPTOMS

FIGURE 303-5 Chronic obstructive pulmonary disease (COPD) severity assessment. COPD severity categories are defined using respiratory symptoms (based on the Modified Medical Research Council Dyspnea Scale [mMRC] or COPD Assessment Test [CAT]) and annual frequency of COPD exacerbations. The mMRC provides a single number for degree of breathlessness: 0—only with strenuous activity; 1—hurrying on level ground or walking up a slight hill; 2—walk slower than peers or stop walking at their own pace; 3—walking about 100 yards or after a few minutes on level ground; 4—too breathless to leave the house or when dressing. The CAT is an eight-item COPD health status measure with Likert scale responses for questions about cough, phlegm, chest tightness, dyspnea after climbing one flight of stairs, limitation in home activities, confidence in leaving the home, sleep, and energy. Range of total score is 0–40. Both mMRC and CAT are available from Global Strategy for the Diagnosis, Management and Prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2024. *(Reproduced with permission from 2025 GOLD Report-Global initiative for chronic obstructive lung disease; 2025.)*

provide improvement in lung function that is greater than either agent alone and reduces exacerbations. Dual bronchodilators are recommended as first-line treatments in patients with symptoms and/or increased exacerbation risk. There are multiple approved drug-device combinations, and some patients may find improved symptoms or increased ease of use for a particular inhaler or nebulizer device.

Inhaled Corticosteroids The main role of inhaled corticosteroids (ICS) is to reduce exacerbations. In population studies, patients with an eosinophil count of <100 cells per microliter do not benefit, while benefit increases as eosinophil counts rise above 100; ICS are recommended as part of the initial treatment regimen for eosinophil counts above 300. ICS are never used alone in COPD due to little symptomatic benefit but, rather, are combined with a LABA or used with a LABA and LAMA. Triple-therapy inhalers (LAMA-LABA-ICS) may be preferred over multiple inhalers. ICS use has been associated with increased rates of oropharyngeal candidiasis and pneumonia and, in some studies, an increased rate of loss of bone density and development of cataracts. ICS can be added when patients with eosinophils counts above 100 continue to have exacerbations despite dual bronchodilator therapy. In stable patients without exacerbations, ICS withdrawal may be considered. Although ICS withdrawal does not lead to an increase in exacerbations, there may be a small decline in lung function.

Oral Glucocorticoids The chronic use of oral glucocorticoids for treatment of COPD is not recommended because of an unfavorable benefit/risk ratio. The chronic use of oral glucocorticoids is associated with significant side effects, including osteoporosis, weight gain, cataracts, glucose intolerance, and increased risk of infection. A study demonstrated that patients tapered off chronic low-dose prednisone (~10 mg/d) did not experience any adverse effect on the frequency of exacerbations, health-related quality of life, or lung function.

PDE4 Inhibitors The selective phosphodiesterase 4 (PDE4) inhibitor roflumilast has been demonstrated to reduce exacerbation frequency in patients with severe COPD, chronic bronchitis, and a prior history of exacerbations; its effects on airflow obstruction and respiratory symptoms are modest, and side effects (including nausea, diarrhea, and weight loss) are common.

Antibiotics There are strong data implicating bacterial infection as a precipitant of COPD exacerbations. A randomized clinical trial of azithromycin, chosen for both its anti-inflammatory and antimicrobial properties, administered daily to subjects with a history of exacerbation in the past 6 months, demonstrated a reduced exacerbation frequency and longer time to first exacerbation. Azithromycin was most effective in older patients, former smokers, and milder COPD. Side effects include hearing loss and development of macrolide-resistant organisms; chronic azithromycin should be avoided in the setting of a prolonged QTc interval.

Oxygen Supplemental O_2 has been demonstrated to unequivocally decrease mortality in hypoxemic patients with COPD. For patients with resting hypoxemia (resting O_2 saturation ≤88% in any patient or ≤89% with signs of pulmonary arterial hypertension, right heart failure, or erythrocytosis), the use of O_2 has been demonstrated to have a significant impact on mortality. Patients meeting these criteria should be on continuous oxygen supplementation because the mortality benefit is proportional to the number of hours per day oxygen is used. Various delivery systems are available, including portable systems that patients may carry to allow mobility outside the home. A recent study failed to demonstrate mortality or symptomatic benefits to COPD patients with moderate hypoxemia at rest or with hypoxemia only with activity. Long-term nocturnal noninvasive mechanical ventilation may be beneficial in stable COPD patients with chronic hypercapnia; a sleep study is recommended prior to initiation to exclude concurrent obstructive sleep apnea.

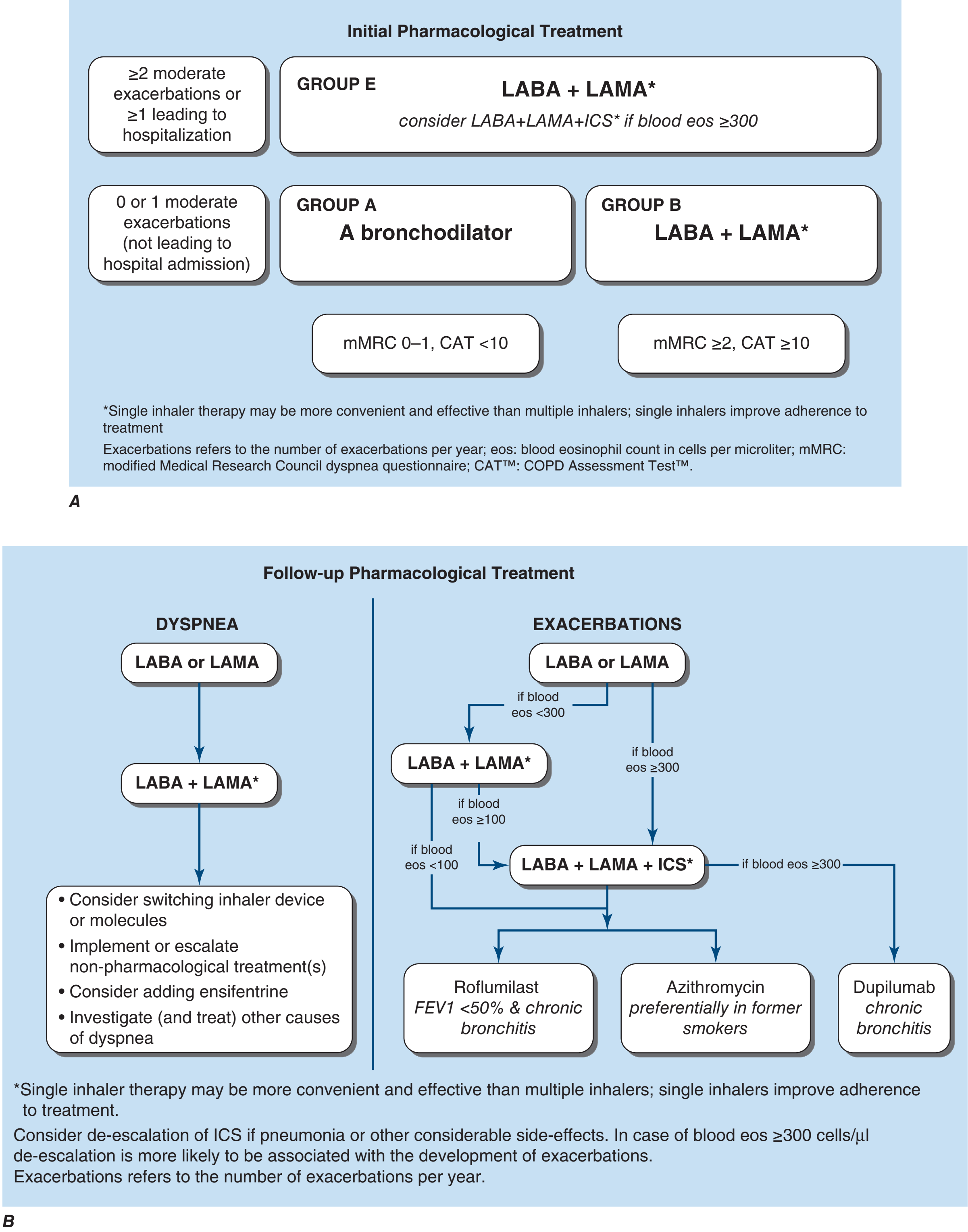
Initial Pharmacological Treatment
≥2 moderate exacerbations or ≥1 leading to hospitalization
GROUP E
LABA + LAMA*
consider LABA+LAMA+ICS* if blood eos ≥300
0 or 1 moderate exacerbations (not leading to hospital admission)
GROUP A
A bronchodilator
GROUP B
LABA + LAMA*
mMRC 0–1, CAT <10
mMRC ≥2, CAT ≥10
A
Follow-up Pharmacological Treatment
DYSPNEA
LABA or LAMA
LABA + LAMA*
• Consider switching inhaler device or molecules
• Implement or escalate non-pharmacological treatment(s)
• Consider adding ensifentrine
• Investigate (and treat) other causes of dyspnea
EXACERBATIONS
LABA or LAMA
if blood eos <300
if blood eos ≥300
LABA + LAMA*
if blood eos ≥100
if blood eos <100
LABA + LAMA + ICS*
if blood eos ≥300
Roflumilast FEV1 <50% & chronic bronchitis
Azithromycin preferentially in former smokers
Dupilumab chronic bronchitis
B

*Single inhaler therapy may be more convenient and effective than multiple inhalers; single inhalers improve adherence to treatment

Exacerbations refers to the number of exacerbations per year; eos: blood eosinophil count in cells per microliter; mMRC: modified Medical Research Council dyspnea questionnaire; CAT™: COPD Assessment Test™.

*Single inhaler therapy may be more convenient and effective than multiple inhalers; single inhalers improve adherence to treatment.

Consider de-escalation of ICS if pneumonia or other considerable side-effects. In case of blood eos ≥300 cells/μl de-escalation is more likely to be associated with the development of exacerbations.

Exacerbations refers to the number of exacerbations per year.

FIGURE 303-6 Medication therapy for stable chronic obstructive pulmonary disease (COPD). Initial pharmacologic therapy (***A***) is based on both COPD exacerbations and respiratory symptoms (assessed through the Modified Medical Research Council Dyspnea Scale [mMRC] or the COPD Assessment Test [CAT]). Follow-up pharmacologic therapy (***B***) is based on response to treatment initiation and reassessment of symptoms and exacerbations. Global Strategy for the Diagnosis, Management and Prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease (GOLD), 2024. *For ***B***: Single-inhaler therapy may be more convenient and effective than multiple inhalers. **Consider deescalation of ICS if pneumonia or other side effects occur. Eos, blood eosinophil count in cells per microliter; FEV$_1$, forced expiratory volume in 1 s; ICS, inhaled corticosteroid; LABA, long-acting beta agonist; LAMA, long-acting muscarinic antagonist. *(Reproduced with permission from 2025 GOLD Report-Global initiative for chronic obstructive lung disease; 2025.)*

α_1AT Augmentation Therapy Specific treatment in the form of IV α_1AT augmentation therapy is available for individuals with severe α_1AT deficiency. Although biochemical efficacy of α_1AT augmentation therapy has been shown, the benefits of α_1AT augmentation therapy are controversial. A randomized study suggested a reduction in emphysema progression in patients receiving α_1AT augmentation therapy. Eligibility for α_1AT augmentation therapy requires a serum α_1AT level <11 μ*M* (~55 mg/dL). Typically, Pi^Z individuals will qualify, although other rare types associated with severe deficiency (e.g., null-null) are also eligible. Because only a fraction of individuals with severe α_1AT deficiency will develop COPD, α_1AT augmentation therapy is not recommended for severely α_1AT-deficient persons with normal pulmonary function and a normal chest CT scan.

Other Biologic Therapies Dupilumab is a monoclonal antibody targeting IL-4 and IL-13, delivered by subcutaneous injection. It is approved for use in allergic diseases including asthma, atopic dermatitis, chronic rhinosinusitis with nasal polyps, and eosinophilic esophagitis. A recent study showed that dupilumab injections reduced the exacerbation rate in symptomatic COPD patients with high exacerbation risk and elevated blood eosinophils of at least 300 cells per microliter, when added to triple inhaled therapy.

NONPHARMACOLOGIC THERAPIES

Patients with COPD should receive the influenza vaccine annually. Pneumococcal, COVID-19, and respiratory syncytial virus (RSV) vaccines are recommended; vaccination for *Bordetella pertussis* is recommended for those not vaccinated in adolescence.

Pulmonary Rehabilitation Pulmonary rehabilitation is a comprehensive treatment program that incorporates exercise, education, and psychosocial and nutritional counseling. In COPD, pulmonary rehabilitation has been demonstrated to improve health-related quality of life, dyspnea, and exercise capacity. It has also been shown to reduce rates of hospitalization over a 6- to 12-month period.

Lung Volume Reduction In carefully selected patients with emphysema, surgery to remove the most emphysematous portions of lung improves exercise capacity, lung function, and survival. The anatomic distribution of emphysema and postrehabilitation exercise capacity are important prognostic characteristics. Patients with upper lobe–predominant emphysema and a low postrehabilitation exercise capacity are most likely to benefit from LVRS. Patients with an FEV_1 <20% of predicted and either diffusely distributed emphysema on CT scan or diffusing capacity of lung for carbon monoxide ($D_{L_{CO}}$) <20% of predicted have increased mortality after the procedure and thus are not candidates for LVRS.

Two different endobronchial one-way valves for bronchoscopic lung volume reduction (BLVR) have been approved by the U.S. Food and Drug Administration. The indications are largely similar to those for LVRS, although valves can be used to treat emphysema in any lung lobe. Patients should not have interlobar collateral ventilation, either measured by intact fissures on chest CT scan or a bronchoscopic device. The main postprocedure complication is pneumothorax, which usually occurs within the first 3 days. A small study suggested similar outcomes at 12 months for BLVR compared to LVRS.

***Lung Transplantation* (See Chap. 309)** COPD is currently the second leading indication for lung transplantation. Current recommendations are that candidates for lung transplantation should have very severe airflow obstruction; have severe disability despite maximal medical therapy; be free of significant comorbid conditions such as liver, renal, or cardiac disease; and not be a candidate for LVRS or BLVR. Currently, there is no cure for COPD except lung transplantation. Current efforts focus on early diagnosis and comprehensive treatment as well as mitigation of environmental risk factors.

EXACERBATIONS OF COPD

Exacerbations are often a prominent feature of the natural history of COPD. Exacerbations are episodic acute worsening of respiratory symptoms, including increased dyspnea, cough, and/or change in the amount and character of sputum, usually over a period of <14 days. They may or may not be accompanied by other signs of illness, including fever, myalgias, and sore throat. The strongest single predictor of exacerbations is a history of a previous exacerbation. The frequency of exacerbations increases as airflow obstruction worsens; patients with severe airflow obstruction (FEV_1 <50% predicted) on average have 1–3 episodes per year. However, some individuals with severe airflow obstruction do not have frequent exacerbations. Other factors, such as current smoking, an elevated ratio of the diameter of the pulmonary artery to aorta on chest CT, and gastroesophageal reflux, are also associated with increased risk of COPD exacerbations. Economic analyses have shown that a majority of the $50 billion annual COPD-related health care expenditures in the United States are due to COPD exacerbations.

Precipitating Causes and Strategies to Reduce Frequency of Exacerbations A variety of stimuli may result in the final common pathway of airway inflammation and increased respiratory symptoms that are characteristic of COPD exacerbations. Viral respiratory infections had been thought to be a less common cause of COPD exacerbations than bacterial infection, although polymerase chain reaction–based studies have shown that viral infections may be a cause of >50% of exacerbations. Studies suggest that acquiring a new strain of bacteria is associated with increased near-term risk of exacerbation. Other inciting factors include air pollution, allergens, pulmonary embolism, and medication nonadherence. In a significant minority of instances, no specific precipitant can be identified.

Patient Assessment An attempt should be made to establish the severity of the exacerbation as well as the severity of preexisting COPD. The more severe either of these two components, the more likely it is that the patient will require hospital admission. The history should include quantification of the degree and change in dyspnea by asking about breathlessness during activities of daily living and typical activities for the patient. The patient should be asked about fever; change in character of sputum; and associated symptoms such as wheezing, nausea, vomiting, diarrhea, myalgias, and chills. Inquiring about the frequency and severity of prior exacerbations can provide important information; the single greatest risk factor for hospitalization with an exacerbation is a history of previous hospitalization.

The physical examination should incorporate an assessment of the degree of distress of the patient. Specific attention should be focused on tachycardia, tachypnea, use of accessory muscles, signs of perioral or peripheral cyanosis, the ability to speak in complete sentences, and the patient's mental status. The chest examination should establish the presence or absence of focal findings, degree of air movement, presence or absence of wheezing, asymmetry in the chest examination (suggesting large airway obstruction or pneumothorax mimicking an exacerbation), and the presence or absence of paradoxical motion of the abdominal wall.

Patients with severe underlying COPD, who are in moderate or severe distress, or those with focal findings should have a chest x-ray or chest CT scan. Approximately 25% of x-rays in this clinical situation will be abnormal, with the most frequent findings being pneumonia and congestive heart failure, and occasionally pneumothorax. Patients with advanced COPD, a history of hypercarbia, or mental status changes (confusion, sleepiness) or those in significant distress should have an arterial blood gas measurement. The presence of hypercarbia, defined as a P_{CO_2} >45 mmHg, has important implications for treatment (discussed below). In contrast to its utility in the management of exacerbations of asthma, measurement of pulmonary function has not been demonstrated to be helpful in the diagnosis or management of exacerbations of COPD. Pulmonary embolus (PE) should also be considered because the incidence of PE is increased in COPD exacerbations.

The need for inpatient treatment of exacerbations is suggested by the presence of respiratory acidosis and hypercarbia, new or worsening hypoxemia, severe underlying COPD, significant comorbidities such as heart failure, and those whose living situation is not conducive to careful observation and the delivery of prescribed treatment. COPD exacerbation is a clinical diagnosis. A recent expert consensus has developed more objective criteria to diagnose and stage the severity of COPD exacerbations, based on dyspnea, vital signs, and testing of C-reactive protein and arterial blood gases. However, this schema may have more utility in the emergency department than in the outpatient setting.

TREATMENT OF ACUTE EXACERBATIONS

Bronchodilators Typically, patients are treated with inhaled beta agonists and muscarinic antagonists. These may be administered separately or together, and the frequency of administration depends on the severity of the exacerbation. Patients are often treated initially with nebulized therapy, as such treatment is often easier to administer in those in respiratory distress. It has been shown, however, that conversion to metered-dose inhalers is effective when accompanied by education and training of patients and staff. This approach has significant economic benefits and also allows an easier transition to outpatient care.

Antibiotics Patients with COPD are frequently colonized with potential respiratory pathogens, and it is often difficult to identify conclusively a specific species of bacteria responsible for a particular clinical event. Bacteria frequently implicated in COPD exacerbations include *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis,* and *Chlamydia pneumoniae;* viral pathogens are also common etiologies of exacerbations. The choice of antibiotic should be based on local patterns of antibiotic susceptibility of the above bacterial pathogens as well as the patient's clinical condition. Patients with moderate or severe exacerbations are usually treated with 5–7 days of antibiotics, even in the absence of data implicating a specific pathogen.

Glucocorticoids In patients admitted to the hospital, the use of systemic glucocorticoids reduces the length of stay, hastens recovery, and reduces the chance of subsequent exacerbation or relapse. Current recommendations suggest 40 mg of oral prednisone or its equivalent typically for a period of 5 days. Hyperglycemia, particularly in patients with preexisting diagnosis of diabetes, is the most frequently reported acute complication of glucocorticoid treatment.

Oxygen Supplemental O_2 should be supplied with a target oxygen saturation of 88–92%. Studies have demonstrated that in patients with both acute and chronic hypercarbia, the administration of supplemental O_2 does not reduce minute ventilation. It does, in some patients, result in modest increases in arterial P_{CO_2}, chiefly by altering ventilation-perfusion relationships within the lung. This should not deter practitioners from providing the oxygen needed to correct hypoxemia.

Mechanical Ventilatory Support The initiation of noninvasive positive-pressure ventilation (NIPPV) in patients with acute respiratory acidosis, defined as Pa_{CO_2} >45 mmHg and pH ≤7.35, results in a significant reduction in mortality rate, need for intubation, complications of therapy, and hospital length of stay. Contraindications to NIPPV include cardiovascular instability, impaired mental status, inability to cooperate, copious secretions or the inability to clear secretions, or craniofacial abnormalities or trauma precluding effective fitting of the mask.

Invasive (conventional) mechanical ventilation via an endotracheal tube is indicated for patients with severe respiratory distress, hypoxemia, severe hypercarbia and/or acidosis despite noninvasive ventilation, markedly impaired mental status, respiratory arrest, hemodynamic instability, or other complications. The goal of mechanical ventilation is to correct the aforementioned conditions. Factors to consider during mechanical ventilatory support include the need to provide sufficient expiratory time in patients with severe airflow obstruction and the presence of auto-PEEP (positive end-expiratory pressure), which can result in patients having to generate significant respiratory effort to trigger a breath during a demand mode of ventilation.

The mortality rate of patients requiring mechanical ventilatory support for a COPD exacerbation is 17–49% for that particular hospitalization. Owing to the high mortality of invasive mechanical ventilation in COPD exacerbations, patient preferences for advanced directives (e.g., do not resuscitate) should be discussed in the outpatient setting. Following a hospitalization for COPD, ~20% of patients are rehospitalized in the next 30 days and 45% in the next year. Mortality is ~20% in the year following hospital discharge.

ACKNOWLEDGMENT

James Crapo and Barry Make contributed to this chapter in the 21st edition and some material from that chapter has been retained here.

■ FURTHER READING

Agusti A, Hogg JC: Update on the pathogenesis of chronic obstructive pulmonary disease. N Engl J Med 381:1248, 2019.

Celli CR, Wedzicha JA: Update on clinical aspects of chronic obstructive pulmonary disease. N Engl J Med 381:1257, 2019.

Global Strategy for the Diagnosis, Management and Prevention of COPD: Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2024. Available at *http://goldcopd.org*.

Kheradmand F et al: Contribution of adaptive immunity to human COPD and experimental models of emphysema. Physiol Rev 103:1059, 2023.

Lynch D et al: CT definable subtypes of COPD: A statement of the Fleischner Society. Radiology 277:192, 2015.

Regan E et al: Clinical and radiologic disease in smokers with normal spirometry. JAMA Intern Med 175:1539, 2015.

Sakornsakoplat P et al: Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. Nat Genet 51:494, 2019.

Sandhaus RA et al: The diagnosis and management of alpha-1 antitrypsin deficiency in the adult. Chronic Obstr Pulm Dis 3:668, 2016.

Stanojevic S et al: ERS/ATS technical standard on interpretive strategies for routine lung function tests. Eur Respir J 60:2101499, 2022.

Stolz D et al: Towards the elimination of chronic obstructive pulmonary disease: A Lancet Commission. Lancet 400:921, 2022.

304 Interstitial Lung Disease

Gary M. Hunninghake, Ivan O. Rosas

Diffuse parenchymal lung diseases include a large number (>200) of heterogeneous conditions that affect the lung parenchyma with varying degrees of inflammation and fibrosis. While remodeling of the interstitial space, the region between the epithelium and endothelium, tends to be the dominant site of involvement for most of the interstitial lung diseases (ILDs), it is important to recognize the prominent role of the alveolar epithelium and endothelial cells (including both airways and vessels) in the pathogenesis of these ILDs.

Despite the diverse array of conditions, most patients ultimately diagnosed with an ILD will come to medical attention with reports of progressive exertional dyspnea or a persistent dry cough. However, because some ILDs are part of multisystem disorders, some patients will be identified based on nonrespiratory symptomatology (e.g., skin thickening in the setting of systemic sclerosis, **Chap. 372**) or physical examination findings (e.g., ulnar deviation of the fingers in the setting of rheumatoid arthritis [RA], **Chap. 370**). Additionally, ILDs can also be identified incidentally based on the results of abnormal pulmonary function tests, chest x-rays (CXRs), and computed tomography (CT) studies of both the chest and abdomen (which can both visualize at least a portion, of the lung parenchyma), and positron emission tomography (PET) scans. It is important to remember that ILDs can be associated with high rates of morbidity and mortality, and although prognosis depends on both disease extent and specificity, this fact makes these important disorders to recognize in a timely manner.

TABLE 304-1 Common Interstitial Lung Disease (ILD) Findings

	IPF	NONSPECIFIC INTERSTITIAL PNEUMONIA	RESPIRATORY BRONCHIOLITIS–ASSOCIATED ILD	SYSTEMIC SCLEROSIS–ASSOCIATED ILD	SARCOIDOSIS
Clinical symptoms	Gradual onset of SOB, dry cough. More common in older adults.	Subacute onset of SOB, dry cough. Frequently associated with other conditions.	Can be asymptomatic, or have SOB and cough.	Gradual onset of SOB, dry cough. Fatigue, tightening of skin, exaggerated cold response, reflux, and difficulty swallowing.	Can be asymptomatic, or have SOB and cough. Can also have fatigue, palpitations, and eye, skin, and joint findings.
Physical exam findings	Frequent rales at lung bases; digital clubbing is common.	Frequent rales. Clubbing is less common.	Rales common. Clubbing is rare.	Can have rales in isolation. Also skin thickening, joint swelling, and telangiectasias.	Can be normal; rales may be present. Can have skin findings, joint pain, and enlarged lymph nodes.
Exposures	Idiopathic but many exposed to smoke. Genetic findings may explain more than one-third of the risk of the disease.	Can be idiopathic but should prompt consideration for associated conditions.	Strong association with smoking.	Mostly unknown; some debate about solvent and silicate exposures.	Mostly unknown, although silicate dusts thought to play a role in some cases.
HRCT findings	Bilateral subpleural reticular changes most prominent in lower, posterior lung zones. Traction bronchiectasis and honeycombing common. Classic usual interstitial pneumonia (UIP) pattern is considered diagnostic.	Peripheral subpleural ground-glass and reticular patterns. Traction bronchiectasis is common, but honeycombing is rare. HRCT not diagnostic.	Diffuse patchy centrilobular ground glass nodules.	Can have UIP or nonspecific interstitial pneumonia (NSIP) patterns, also dilated esophagus, occasional mediastinal calcifications, and pulmonary vascular enlargement.	Can have mediastinal and hilar lymphadenopathy. Peribronchovascular reticular-nodular findings.
Histopathology	UIP pattern including fibroblastic foci, temporal and spatial heterogeneity, honeycombing.	Cellular or fibrotic pattern of NSIP. More uniform than a UIP pattern.	Respiratory bronchiolitis with adjacent inflammatory and fibrosing changes. Pigment-laden macrophages.	Both UIP or NSIP patterns can occur.	Noncaseating granulomas.
Clinical course	50% 3- to 5-year mortality.	18% 5-year mortality.	25% 7-year mortality.	20–30% 10-year mortality.	Generally low but varies by state.

Abbreviations: HRCT, high-resolution computed tomography; IPF, idiopathic pulmonary fibrosis; SOB, shortness of breath.

Owing to a variety of clinical presentations, as well as overlapping imaging and histopathologic findings (Table 304-1), ILDs can be difficult to diagnose. A generally accepted central tenet of ILD diagnosis is that the combined weight of clinical data, laboratory studies, pulmonary function testing, imaging findings, and histopathology (if obtained) are jointly required to make a confident diagnosis. No single piece of data confers a diagnosis alone. For example, a lung biopsy demonstrating the usual interstitial pneumonia (UIP) pattern is helpful in diagnosing a patient with idiopathic pulmonary fibrosis (IPF) but can also be present in some connective tissue diseases (CTDs) (e.g., RA-associated ILD, Chap. 370). In light of this challenge, most ILD centers recommend a multidisciplinary approach to the diagnosis (and, in some cases, the management) of ILDs. An example of a multidisciplinary approach is a conference attended by pulmonologists, rheumatologists, radiologists, and pathologists where all of the data generated on a patient can be discussed and reviewed jointly by those with unique sets of expertise in the care of patients with ILD.

While there are numerous ways to categorize the ILDs, one classic approach is to divide them into those of known and unknown causes (Fig. 304-1). Although even this approach has limitations (e.g., a growing number of genetic studies demonstrate that a significant portion of familial and sporadic pulmonary fibrosis or IPF may be explained, in part, by genetic factors), it is a useful place to start. Known causes of ILD include occupational exposures (e.g., asbestosis), medications (e.g., nitrofurantoin), and those related to an underlying systemic disease (e.g., cryptogenic organizing pneumonia [COP] in the setting of polymyositis). Unknown causes of ILD include groups of rare disorders often with classic presentations (e.g., a spontaneous pneumothorax in a young female with diffuse cystic changes on a chest CT might suggest lymphangioleiomyomatosis [LAM]) and the most common group of ILDs, the idiopathic interstitial pneumonias (IIPs). Granulomatous lung diseases straddle both known (e.g., hypersensitivity pneumonitis [HP] due to chronic bird exposure, Chap. 299) and unknown (e.g., sarcoidosis, Chap. 379) causes and are often separated due to their unique presentations, imaging findings, and diagnostic evaluation. Equally important to knowledge of disease classification is knowledge of disease prevalence. Although there is variability within different demographic groups, most studies demonstrate that IPF, sarcoidosis (Chap. 379), and ILDs related to CTDs (Chap. 425) as a group are among the most common forms of ILD.

DIAGNOSTIC APPROACH

The initial diagnostic approach to diffuse parenchymal lung disease is often broader than a focus on ILD and should include an evaluation for alternate causes, including cardiovascular disease (e.g., heart failure, Chap. 265), diffuse infections (e.g., *pneumocystis* pneumonia, Chap. 227), and malignancy (e.g., bronchoalveolar cell carcinoma). This chapter will focus on the diagnostic evaluation that helps to distinguish among the various forms of ILD.

■ HISTORY

Age The age of onset of clinical symptoms has a strong influence on the pretest probability that IPF, in particular, is present. For example, IPF occurs most commonly in patients aged >60 and is quite rare among patients aged <50. In fact, in patients aged >65 without strong evidence for an alternate diagnosis, atypical chest CT findings are still more likely to result in a histopathologic diagnosis of UIP (a pathologic hallmark of IPF) than they are to result in an alternate IIP diagnosis. Other common ILDs, such as sarcoidosis and CTD-associated ILD, and less common ILDs, such as LAM and pulmonary Langerhans cell histiocytosis (PLCH), tend to present between the ages of 20 and 40.

Sex Although less influential than age, sex has some influence on the likelihood of various ILDs. LAM (and the related disorder tuberous sclerosis) is a disorder that is frequently diagnosed in young women. Many CTD-associated ILDs are more common among women, except for RA-associated ILD, which is more common among men. IPF and occupational/exposure-related ILDs (likely due to work-related exposures that tend to differ between men and women) are more common among men.

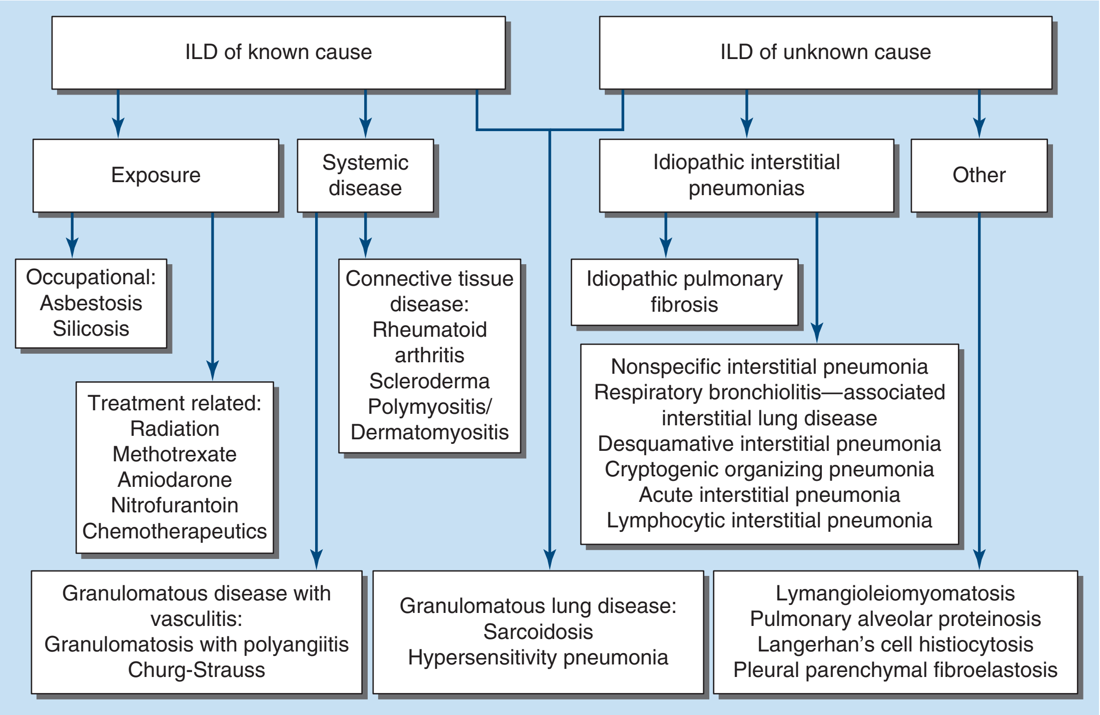
ILD of known cause
ILD of unknown cause
Exposure
Systemic disease
Idiopathic interstitial pneumonias
Other
Occupational: Asbestosis Silicosis
Connective tissue disease: Rheumatoid arthritis Scleroderma Polymyositis/ Dermatomyositis
Idiopathic pulmonary fibrosis
Treatment related: Radiation Methotrexate Amiodarone Nitrofurantoin Chemotherapeutics
Nonspecific interstitial pneumonia Respiratory bronchiolitis—associated interstitial lung disease Desquamative interstitial pneumonia Cryptogenic organizing pneumonia Acute interstitial pneumonia Lymphocytic interstitial pneumonia
Granulomatous disease with vasculitis: Granulomatosis with polyangiitis Churg-Strauss
Granulomatous lung disease: Sarcoidosis Hypersensitivity pneumonia
Lymangioleiomyomatosis Pulmonary alveolar proteinosis Langerhan's cell histiocytosis Pleural parenchymal fibroelastosis

FIGURE 304-1 Classification of interstitial lung disease. This algorithm represents a common approach to subclassifying the interstitial lung diseases. It is typical to divide the interstitial lung diseases into those of known and unknown causes (although it is important to note that genetic studies demonstrate that a significant portion of familial and idiopathic pulmonary fibrosis [classically described as diseases of unknown cause] may be explained, in part, by genetic factors). The idiopathic interstitial pneumonias were more precisely defined by a 2002 study as described in Am J Respir Crit Care Med 165:277, 2002, referenced in the Further Reading list.

Duration of Symptoms Acute presentations (*days to weeks*) of ILD are unusual and are commonly misdiagnosed as more common diseases such as pneumonia, a chronic obstructive pulmonary disease (COPD) exacerbation, or heart failure. ILDs that can present acutely include eosinophilic pneumonia, acute interstitial pneumonia (AIP), HP, and granulomatosis with polyangiitis (GPA). An acute exacerbation of IPF as the initial presentation of this disease should also be a consideration given its prevalence. ILDs most commonly have a chronic indolent presentation (*months to years*) typified by IPF. However, subacute presentations (*weeks to months*) can occur in most of the ILDs but, in the right context, could suggest sarcoidosis, CTD-associated ILD, drug-induced ILD, or COP.

Respiratory Symptoms Progressive dyspnea, most frequently noted with exertion, is the most common complaint in patients presenting with an ILD. Despite this fact, both research studies of general population samples and clinical experiences of asymptomatic patient referrals with abnormal chest CT imaging patterns have also demonstrated that some patients, even those with more extensive disease, may not report dyspnea. Cough, particularly a dry cough, is also common and can be the most prominent symptom in patients with IPF. Cough is often reported in other ILDs, particularly those with prominent airway involvement including sarcoidosis and HP. Cough with hemoptysis is rare and could suggest an ILD associated with diffuse alveolar hemorrhage (DAH) (e.g., Goodpasture's syndrome), GPA, or LAM. Cough with hemoptysis could also suggest a secondary pulmonary infection that can be seen in patients with traction bronchiectasis and in those receiving immunosuppressive therapy. Chest pain is rare in most of the ILDs, with the exception of sarcoidosis, where chest discomfort is not uncommon. Fatigue is common to all of the ILDs.

Past Medical History The most pertinent history includes a personal history of a CTD or a history of symptoms commonly associated with a CTD (e.g., Raynaud's phenomena). It is also important to remember that ILD associated with a CTD can be the initial presenting symptom of the disease and can precede the development of additional symptomatology by many years. A history of malignancy is important, because some malignancies can be associated with dermatomyositis-associated COP and sarcoid-like reactions. A history of asthma and allergic rhinitis might suggest a diagnosis of eosinophilic GPA.

Medications Many medications have been associated with ILD, and to complicate matters further, many medications commonly used to treat inflammatory and granulomatous lung disease are also associated with ILD development (e.g., methotrexate, azathioprine, rituximab, and the tumor necrosis factor α–blocking agents). Specific medications in many classes are also known to cause ILD, including antibiotics (e.g., nitrofurantoin), antiarrhythmics (e.g., amiodarone), and many of the antineoplastic agents (e.g., bleomycin).

Family History A family history of ILD (of almost any type) is important to ascertain. The percentage of pulmonary fibrosis that is familial, as opposed to idiopathic, varies by study, and could be as high as 20%. Despite the variability, most studies suggest the presence of a close relative with an IIP is among the strongest risk factors for IPF. Family studies have consistently noted familial aggregation of diverse forms of IIP (such as IPF, nonspecific interstitial pneumonia [NSIP], and desquamative interstitial pneumonia [DIP] running in the same family) and, in some cases, other forms of ILD. To date, the most well-replicated genetic factors for pulmonary fibrosis (a promoter variant of a mucin gene [*MUC5B*]) and various genetic determinants known to influence telomere length (e.g., variants in the telomerase reverse transcriptase gene [*TERT*]) **(Chap. 495)** appear to be associated with both familial and idiopathic forms of pulmonary fibrosis similarly.

Social History A history of smoking is nearly always present in some forms of ILD (e.g., respiratory bronchiolitis and DIP—sometimes referred to by pathologists jointly as smoking-related ILD) where it is felt to be causative. A history of smoking is also noted in approximately three-quarters of IPF patients. Occupational and environmental exposure histories are also important to obtain as they might identify exposures known to cause pulmonary fibrosis (e.g., significant asbestos exposure) or HP (pigeon breeder's lung).

■ PHYSICAL EXAMINATION

End-inspiratory fine crackles, or rales, noted at the lung bases are found in most patients with IPF and may be one of the earliest signs of the disease. However, rales are nonspecific and can be found in many forms of ILD and other disorders. Wheezing is uncommon in most forms of ILD but can be present in some disorders, such as sarcoidosis, HP, and eosinophilic GPA. Signs of advanced disease include cyanosis, digital clubbing, and cor pulmonale.

LABORATORY STUDIES

Laboratory studies can be particularly helpful in the workup for an underlying CTD-associated ILD. As noted previously, these tests can reveal the presence of an underlying CTD as the cause of an ILD (e.g., a positive anti-cyclic citrullinated peptide [anti-CCP] antibody for RA) even when no other symptomatology or physical examination findings suggestive of the disorder are present. However, the cost-effectiveness and the extent of laboratory testing that should be ordered in various clinical contexts have yet to be determined (as there is a relatively long list of autoantibody tests that could be ordered).

PULMONARY FUNCTION TESTS

Most forms of ILD will eventually result in a restrictive deficit in pulmonary function testing. A restrictive deficit is typified by a reduced total lung capacity (TLC) and symmetrically reduced measures of forced expiratory volume in 1 s (FEV_1) and forced vital capacity (FVC). A reduction in the diffusing capacity of the lung for carbon monoxide (DL_{CO}) is also common and may precede a reduction in lung volumes; however, there is more measurement variability in DL_{CO} measurement and the test is less specific for ILD. A reduced FEV_1 to FVC ratio, which is diagnostic of airway obstruction, is unusual in many forms of ILD but can be present as an isolated finding or in conjunction with an additional restrictive deficit in ILDs involving the airways such as sarcoidosis, HP, and LAM. Although pulmonary function testing is rarely diagnostic, reductions in lung function help to characterize the extent of disease, and evidence for a decline in repeated measures of pulmonary function (e.g., FVC) has been correlated with an elevated mortality rate.

CHEST IMAGING STUDIES

Chest X-Ray Findings on CXR can be the first clinical indication that an ILD might be present. For example, enlarged hilar lymph nodes and a pattern of central nodular opacities in the mid to upper lung zones can suggest sarcoidosis. A basilar reticular pattern, with small cystic spaces, in the absence of clinical evidence for heart failure, might suggest IPF. With a few exceptions, CXRs alone rarely lead to a specific diagnosis.

Chest CT High-resolution CT (HRCT) chest imaging is now considered to be standard of care in the initial evaluation of a patient with a suspected ILD. HRCT can be diagnostic for some ILDs (e.g., IPF) in the right clinical context and may preclude the need for, and spare the patient the risk of, a lung biopsy. HRCT also helps to define the extent of the ILD, determine the presence of more concerning features suggestive of advanced disease (e.g., honeycombing), provide information on coexisting diseases (e.g., emphysema and lung cancer), and when not diagnostic, provide the most useful locations for obtaining lung biopsy specimens.

LUNG BIOPSY

Fiberoptic Bronchoscopy A bronchoscopy can be helpful in establishing a specific ILD diagnosis, and can help to establish an alternate diagnosis, in select cases. Examination of serial lavage fluid can be helpful in establishing DAH, which can be present in ILDs with vasculitis (e.g., GPA), and in some cases, cellular examination can suggest a specific diagnosis (eosinophilia >25% in chronic eosinophilic pneumonia or fat globules in macrophages in lipoid pneumonia). Transbronchial lung biopsies and lymph node biopsies (particularly in sarcoidosis) can lead to a confident diagnosis in patients with likely granulomatous lung disease (e.g., sarcoidosis and HP). However, in general, bronchoscopically obtained tissue samples are often felt to be insufficient to diagnose most of the IIPs. To date, studies have been mixed on whether bronchoscopically obtained cryobiopsies, which can result in yields larger than those obtained by transbronchial forceps biopsies, could improve the diagnostic yield of bronchoscopy; however, the precise role of cryobiopsies in the diagnostic workup of ILD has yet to be determined.

Surgical Lung Biopsy A surgically obtained lung biopsy specimen can help solidify the diagnosis of ILD. In many cases, these are now obtained through a video-assisted thoracoscopic (VATS) approach (as compared to an open thoracotomy), which tends to reduce the length of operative times and hospital stays. The diagnostic yield of biopsies tends to be higher if obtained prior to treatment. The desire to obtain a surgical lung biopsy should be weighed against the risks, which can include a short-term mortality rate of as high as 5%. These risks are reported to be higher in biopsies of patients ultimately diagnosed with IPF and in those presenting acutely.

INDIVIDUAL FORMS OF ILD

The ILDs include a diverse group of lung pathologies that can be subclassified into those disorders of unknown cause (e.g., IIPs) and those of known cause (e.g., sometimes referred to as secondary interstitial pneumonias [CTD-associated ILDs]) (see Fig. 304-1). Although this remains a useful approach to classifying these disorders, it is important to recognize that genetic studies are challenging this classic categorization. For example, numerous ILDs commonly listed as having an "unknown cause" have been determined to have significant genetic underpinnings (e.g., IPF and LAM), while the pathophysiologic processes that result in ILDs of "known cause" (e.g., CTD) remain incompletely understood. Diagnosis is based on combined information obtained from a patient's clinical presentation, measures of pulmonary function, imaging, immune serologies, and histopathology. It is important to remember that prognosis and treatment vary widely by disorder (and disease extent). In some cases, medical therapy that is felt to be effective for some ILDs has been proven to be harmful for others. Medical treatments range from immune modulators to antifibrotic medications, whereas lung transplantation remains the standard of care for patients with advanced and rapidly progressive ILDs.

IDIOPATHIC INTERSTITIAL PNEUMONIAS

IDIOPATHIC PULMONARY FIBROSIS

Clinical Manifestations IPF is the most common ILD of unknown cause. Prevalence increases with age and is estimated at 50–200:100,000. IPF is commonly diagnosed in the fifth or sixth decade of life, affects men more than women, and is frequently associated with a history of smoking or other environmental exposures. IPF is a variably progressive disease that carries a poor prognosis with an estimated 50% 3- to 5-year survival.

HRCT Image Findings Chest CT findings include subpleural reticulation with a posterior basal predominance usually including more advanced fibrotic features, such as honeycombing and traction bronchiectasis. Collectively, these imaging findings are referred to as a UIP pattern. The presence of extensive ground-glass opacities, bronchovascular changes, micronodules, mosaic attenuation, or an upper lung predominance should raise suspicion for an alternative diagnosis (Fig. 304-2).

Histopathology Diagnostic VATS biopsy findings include subpleural reticulation associated with honeycomb changes and fibroblast foci (subepithelial collections of myofibroblasts and collagen). These fibrotic changes alternate with areas of preserved normal alveolar architecture consistent with temporal and spatial heterogeneity (Fig. 304-3). Collectively, these pathologic findings are referred to as UIP.

Treatment Historically, IPF was felt to be refractory to medical therapy, with lung transplantation the only viable therapeutic option. This dogma changed in 2014 with large clinical trials that demonstrated that antifibrotic therapy (pirfenidone and nintedanib) can slow the decline of lung function in IPF patients. Further meta-analyses

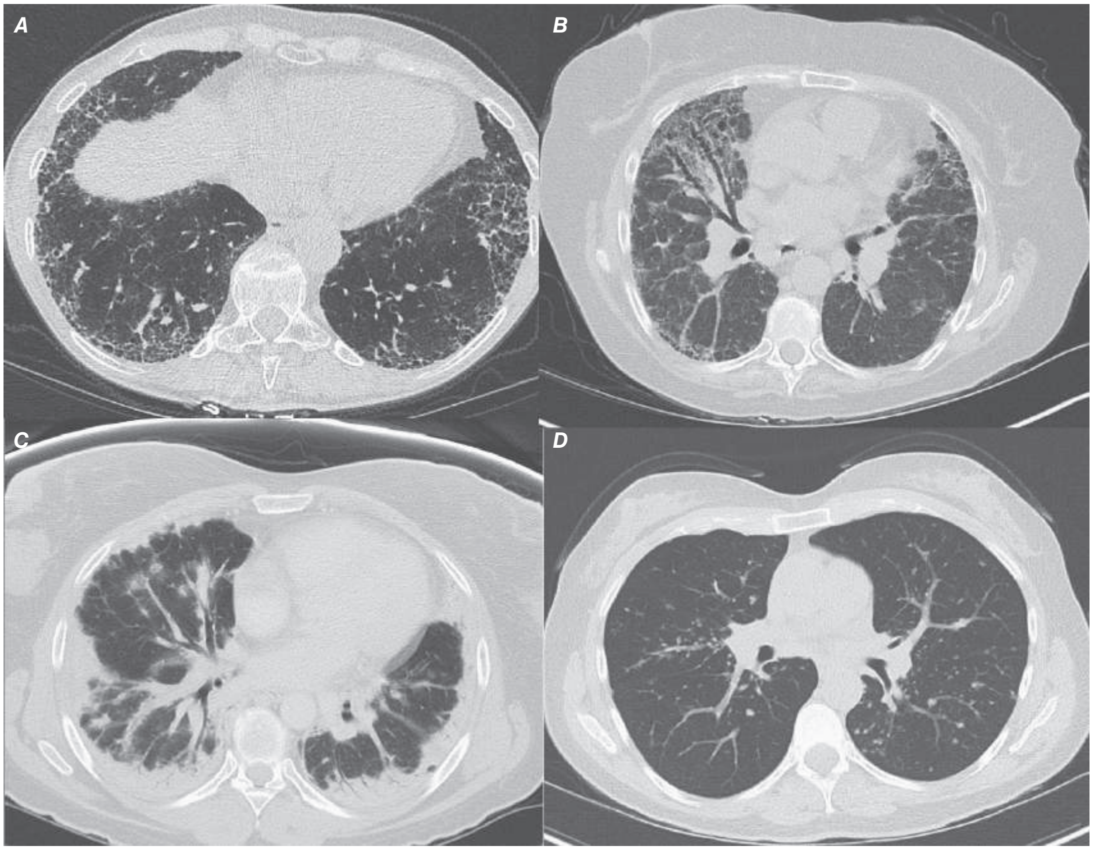

FIGURE 304-2 Chest CT imaging and interstitial lung disease. ***A.*** Idiopathic pulmonary fibrosis (IPF): Classic findings of IPF (apparent on this image) include a posterior, basilar predominance of subpleural reticular markings and more advanced features of pulmonary fibrosis including traction bronchiectasis and honeycombing. This constellation of findings is often referred to as a usual interstitial pneumonia (UIP) pattern. ***B.*** Nonspecific interstitial pneumonia (NSIP): Chest CT findings of NSIP can overlap with those of a UIP pattern but tend to include a bilateral, symmetric pattern that presents with a greater percentage of ground-glass opacities than is apparent in a UIP pattern. Additional unique findings include more diffuse imaging abnormalities with a predominance not limited to the lung bases, imaging abnormalities that spare the subpleural regions, and thickening of the bronchovascular bundles (as is apparent in the right mid lung zone on this image). ***C.*** Cryptogenic organizing pneumonia: Chest CT findings include patchy, sometimes migratory, subpleural consolidative opacities (as is apparent on this image) often with associated ground-glass opacities. Peribronchiolar or perilobar opacities can be present, and sometimes a rim of subpleural sparing (often referred to as a reversed halo or atoll sign) can be seen, which can help to aid in the diagnosis. ***D.*** Sarcoidosis: Sarcoidosis can present with varied imaging abnormalities, but a pattern of mediastinal and hilar lymphadenopathy with a pattern of reticular-nodular opacities involving the bronchovascular bundles (apparent in this image) is a common feature. Additional findings can include diffuse small nodules in a miliary pattern, larger nodular opacities, extensive ground-glass infiltrates, and mosaic attenuation suggestive of small airways involvement, and, in more advanced cases, signs of pulmonary fibrosis.

have suggested that antifibrotic therapy may also improve survival. More recent trials suggest that antifibrotic therapy may also be effective in other forms of progressive pulmonary fibrosis. In contrast, treatment with immunosuppression, which had been commonly prescribed to many IPF patients, has been shown to be associated with increased morbidity and mortality. Physical therapy and supplemental oxygen, when indicated, can improve exercise tolerance and reduce the likelihood of developing pulmonary hypertension. Lung transplantation can extend survival and improve the quality of life in a subset of IPF patients who meet the criteria to undergo transplant.

■ NONSPECIFIC INTERSTITIAL PNEUMONIA

Clinical Manifestations Idiopathic NSIP is a distinct clinical entity with characteristic clinical, radiologic, and pathologic features; however, NSIP is also commonly observed in patients with CTD and less frequently with familial interstitial pneumonia, drug toxicity, and infection. Although the prevalence of NSIP is not well established, it is commonly diagnosed in nonsmoking females in their fifth decade of life. Positive serologic tests for CTD are frequently observed. Idiopathic NSIP has a relatively good prognosis, with a 5-year survival of >80%; patients with a predominant cellular NSIP pattern have a more favorable prognosis than those with a fibrosing NSIP pattern.

HRCT Image Findings Diffuse subpleural, symmetric, ground-glass, and reticular opacities are common. Volume loss and traction bronchiectasis involving the lower lung zones can also be found. Occasionally subpleural sparing is noted, while peribronchiolar thickening and honeycombing are uncommon.

Histopathology Diagnostic lung biopsy findings include varying amounts of interstitial inflammation and fibrosis with a uniform appearance. Honeycomb changes are usually absent, and fibroblast foci are rare. NSIP is often referred to histopathologically as being either predominantly cellular (and potentially more responsive to medical therapy) or fibrotic (and potentially less likely to resolve with medical therapy).

Treatment Pulmonary fibrosis associated with CTD is commonly treated with immunosuppression despite the paucity of randomized clinical trials to demonstrate efficacy. Idiopathic NSIP is often treated with oral steroids (prednisone), cytotoxic agents (mycophenolate, azathioprine, and cyclophosphamide), or biologics (rituximab and tocilizumab). Recent trials suggest that NSIP patients with progressive pulmonary fibrosis may benefit from antifibrotic therapy. Oxygen therapy, pulmonary rehabilitation, and lung transplantation may be required in patients with progressive disease.

■ SMOKING-RELATED ILD

Although smoking-related ILDs, including respiratory bronchiolitis with interstitial lung disease (RB-ILD), and DIP are frequently subclassified with the IIPs, these disorders (along with PLCH, an ILD with unique clinical, imaging, and histopathologic manifestations) are commonly felt to be the result of active or prior tobacco smoke exposure.

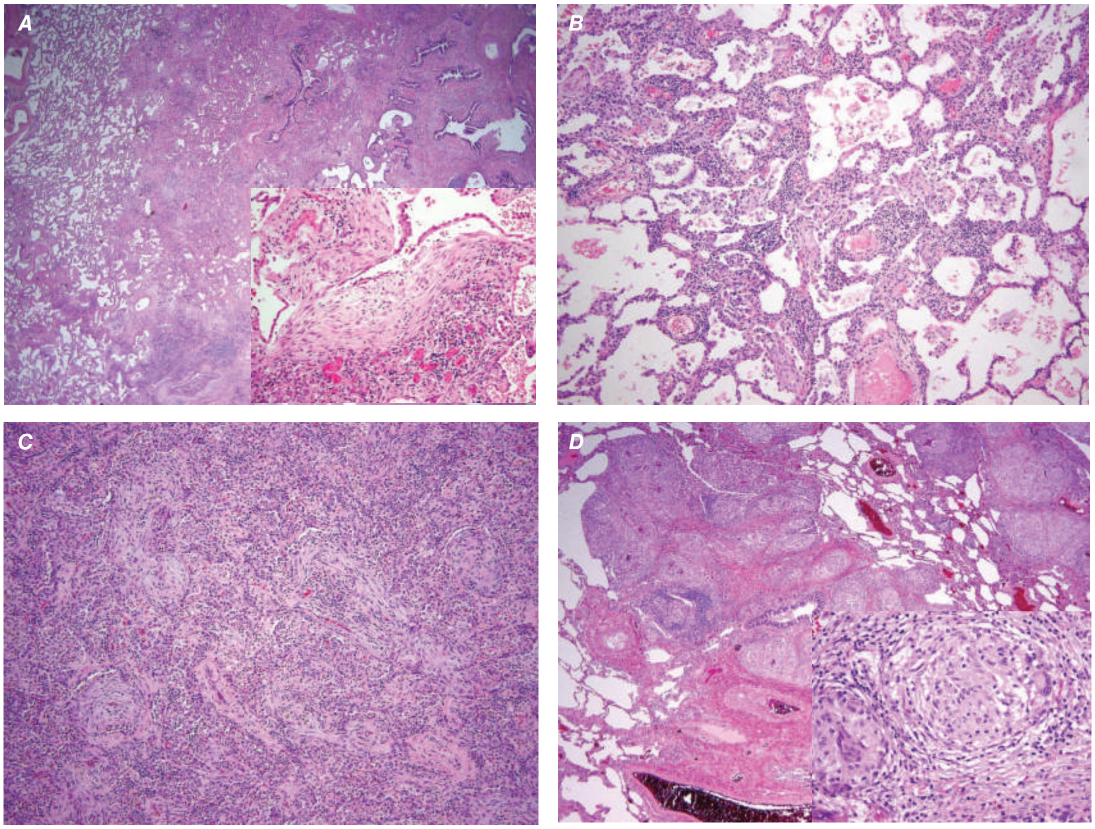

FIGURE 304-3 Histopathology of interstitial lung disease. ***A.*** Idiopathic pulmonary fibrosis (IPF): Histopathologic findings include subpleural reticulation associated with honeycomb changes alternating with areas of preserved normal lung architecture referred to as temporal and spatial heterogeneity (as is apparent in the low-power image above). Additional important diagnostic findings include fibroblast foci, which are subepithelial collections of myofibroblasts and collagen (as is apparent in the higher-powered inset of this image). Collectively, these pathologic findings are referred to as usual interstitial pneumonia (UIP). ***B.*** Nonspecific interstitial pneumonia (NSIP): Histopathologic findings of NSIP include varying amounts of interstitial inflammation and fibrosis with a uniform appearance (as is apparent in this image). Honeycomb changes are usually absent and fibroblast foci are rare. NSIP is often referred to histopathologically as being either predominantly cellular or fibrotic. ***C.*** Cryptogenic organizing pneumonia (COP): Histopathologic findings of COP include patchy regions of organizing pneumonia with granulation tissue that commonly involves the small airways, alveolar ducts, and alveoli with surrounding inflammation that can involve the alveolar walls (as is apparent in this image). ***D.*** Sarcoidosis: The hallmark histopathologic feature of sarcoidosis is presence of granulomas (as are apparent numerously in the low-powered image and more closely visualized in the higher-powered inset image). Typically, these are referred to as noncaseating, which suggests the absence of necrosis. Caseating granulomas are rare in sarcoid and should prompt additional evaluation for an underlying infection. Because malignancy can result in a granulomatous reaction, it is important to closely survey biopsy specimens with granulomatous involvement for additional signs of malignancy.

DIP has also been known to occur in children with familial pulmonary fibrosis (FPF). Smokers, particularly elderly smokers, frequently have radiologic (centrilobular) interstitial abnormalities. These interstitial abnormalities are often incidentally found on routine CXR or chest CT studies in asymptomatic or minimally symptomatic individuals. Respiratory bronchiolitis is felt to correlate histopathologically with these imaging findings. However, in some cases, these imaging findings can progress to more advanced radiologic changes where more diffuse signs of interstitial pneumonia tend to be present.

Clinical Manifestations These disorders predominantly occur in active, and in many cases heavy, smokers who are typically between 40 and 50 years of age. In those ultimately diagnosed with RB-ILD or DIP, dyspnea and cough are relatively common and symptomatic wheezing is not rare. The prevalence of smoking-related ILDs is not well understood, but they are generally felt to account for <10% of the IIPs. While there are minimal data on the natural histories and prognoses of these conditions, prolonged survival can be expected in most patients with RB-ILD and death secondary to progressive ILD is felt to be rare.

HRCT Image Findings Prominent and common findings in RB-ILD include central bronchial wall thickening, peripheral bronchial wall thickening, centrilobular nodules, and ground-glass opacities. Septal lines and a reticular pattern are also not uncommon. Honeycombing is generally felt to be rare (and indicates a worse prognosis). Similar findings are noted in patients with DIP where diffuse (or patchy) bilateral symmetric ground-glass opacities tend to be even more prominent.

Histopathology Common features of RB-ILD include the accumulation of pigmented macrophages within the lumens of respiratory bronchioles and alveolar ducts, accompanied by chronic inflammation of the respiratory bronchiolar walls and both bronchiolar and peribronchiolar alveolar fibrosis causing architectural distortion. These features are patchy and confined to the peribronchiolar region. DIP tends to include similar changes but has a more diffuse pattern characterized by pigmented macrophage accumulation, pneumocyte hyperplasia, and prominent interstitial thickening.

Treatment Patients with smoking-related ILD should be counseled to discontinue smoking and/or encouraged to enroll in a formal smoking cessation program. Small studies have evaluated treatment with immunosuppressive (e.g., prednisone) and cytotoxic (e.g., azathioprine and cyclophosphamide) agents and, in some cases, bronchodilators. To date, there is no strong evidence that these therapies result in significant improvements in symptoms or measures of pulmonary function or prevent clinical deterioration.

CRYPTOGENIC ORGANIZING PNEUMONIA

Clinical Manifestations COP typically involves patients in their 50–60s and often presents as a subacute flulike illness with cough, dyspnea, fever, and fatigue. Inspiratory rales are often present on examination, and most patients are noted to have restrictive lung deficits on pulmonary function testing with hypoxemia. COP is commonly mistaken for pneumonia. It is important to note that this syndrome can occur in isolation, can be secondary to an underlying CTD (e.g., polymyositis) or medications, or can result from an underlying malignancy. Laboratory testing for various CTDs is helpful as testing can both be diagnostic and suggest the need for prolonged medical therapy.

HRCT Image Findings The most common imaging findings include patchy, sometimes migratory, subpleural consolidative opacities often with associated ground-glass opacities. Peribronchiolar or perilobar opacities can be present, and sometimes a rim of subpleural sparing (often referred to as a reversed halo or atoll sign) can be seen, which can aid in the diagnosis.

Histopathology Surgical lung biopsy specimens tend to reveal patchy regions of organizing pneumonia with granulation tissue that commonly involves the small airways, alveolar ducts, and alveoli with surrounding inflammation that can involve the alveolar walls (see Fig. 304-3).

Treatment Corticosteroids can result in substantial clinical improvement in many patients but usually need to be continued for at least 6 months as relapse rates are high. Evidence is growing that alternate cytotoxic (e.g., mycophenolate, cyclophosphamide) or biologic (e.g., rituximab) therapies can be helpful in both treating the disease and reducing the need for steroids. In some patients with secondary forms of the disease, long-term therapy may be needed.

ACUTE OR SUBACUTE IIPS

ACUTE INTERSTITIAL PNEUMONIA (HAMMAN-RICH SYNDROME)

Clinical Manifestations AIP is a rare and often fatal lung disorder that is characterized by an acute onset of respiratory distress and hypoxemia. A prodromal period of symptoms consistent with an acute upper respiratory infection is common. The mortality rate within 6 months of presentation can be quite high (>50%), and recurrences are common. In those who recover, lung function improvement can be substantial. AIP can be difficult to distinguish from acute respiratory distress syndrome (ARDS) and an acute exacerbation of an unsuspected underlying pulmonary fibrotic process.

HRCT Image Findings The most common imaging findings are patchy bilateral ground-glass opacities. Dependent regions of air-space consolidation are also common.

Histopathology Similar to ARDS and acute exacerbations of underlying pulmonary fibrosis, AIP presents histopathologically as diffuse alveolar damage (DAD) demonstrated on a surgical lung biopsy.

Treatment Treatment is mostly supportive and often includes mechanical ventilation. There is no proven drug therapy for AIP. Glucocorticoids are often given, but they are not clearly effective, and data on their use in other forms of DAD (e.g., ARDS) is controversial.

ACUTE EXACERBATIONS OF IIPS

Clinical Manifestations Acute exacerbations are not separate disorders, but rather an accelerated phase of lung injury that can occur in any form of ILD with pulmonary fibrosis. Acute exacerbations are most common and most severe in patients with known IPF. Acute exacerbations are characterized by an acute onset (<30 days) of respiratory distress and hypoxemia occurring in a patient with underlying pulmonary fibrosis not explained by an alternate cause (e.g., pneumonia, left heart failure). Reported mortality rates are very high (>85%), and mean survival periods range from as little as days to months.

HRCT Image Findings The most common imaging findings include patchy bilateral ground-glass opacities and dependent regions of air-space consolidation that can be appreciated on the background of the imaging findings characteristic of the underlying IIP. However, at times, they obscure the preceding imaging findings.

Histopathology Acute exacerbations of underlying pulmonary fibrosis present histopathologically as DAD, although sometimes organizing pneumonia can also be demonstrated on a surgical lung biopsy.

Treatment Overall, treatment is supportive. Mechanical ventilation, when not being used as a bridge to lung transplantation, is controversial as the survival rate in these patients tends to be poor. There is some evidence that drug therapy (e.g., nintedanib) may reduce the rate of acute exacerbations in patients with IPF. Drug therapy, in the context of an acute exacerbation, is also controversial. Immunosuppressive (e.g., prednisone) and cytotoxic (e.g., cyclophosphamide) therapies are commonly used without proven benefit.

ILD ASSOCIATED WITH CONNECTIVE TISSUE DISEASE

ILD is a common disease manifestation of many CTDs. Disease progression, response to therapy, and survival are variable and associated with specific radiologic and histopathologic patterns. ILD occurs most commonly in patients with scleroderma (systemic sclerosis form, or SSc), RA, polymyositis/dermatomyositis, and less frequently Sjögren's syndrome and systemic lupus erythematosus (SLE). ILD may precede the development of extrapulmonary manifestations of a specific CTD or may present as part of a poorly defined CTD. In rare cases, lung manifestations may be the sole feature of the patient's clinical presentation.

SYSTEMIC SCLEROSIS

Clinical Manifestations (Chap. 372) ILD is the most common pulmonary manifestation of SSc. ILD occurs in ~50% of SSc patients with diffuse disease and in ~30% of patients with limited disease. Pulmonary hypertension can occur separately or concomitantly with ILD and is more frequent in patients with limited SSc.

HRCT Image Findings Imaging features observed in patients with NSIP and IPF can be present, although less common findings present in COP and DAD may also be found. Additional HRCT findings may include a dilated esophagus and pulmonary artery enlargement.

Histopathology Comparable to the imaging overlap, histopathologic changes commonly noted in patients with NSIP and IPF are frequently identified. COP and DAD patterns may be observed and could be secondary to aspiration due to esophageal dysmotility, which is common in SSc.

Treatment Cyclophosphamide has a modest benefit in the preservation of lung function and is associated with significant toxicity. Mycophenolate has recently been shown to have similar efficacy and improved tolerability. Recent trials suggest that SSc patients with ILD may benefit from anti–interleukin 6 therapy (e.g., tocilizumab) and antifibrotic therapy (e.g., nintedanib). Minimizing the risk of reflux by using high-dose proton pump inhibitors or antireflux surgery should be considered in SSc with progressive ILD, as gastroesophageal reflux disease may contribute to lung injury and fibrosis. Lung transplantation can potentially be offered to select patients without significant aspiration or chest wall restriction.

RHEUMATOID ARTHRITIS

Clinical Manifestations (Chap. 370) A common extraarticular complication of RA is ILD. Although RA is more common in females, RA-ILD is more frequent in males and in patients with a history of tobacco exposure. In a small subset of patients, ILD is the first disease manifestation of RA. Clinically evident RA-ILD occurs in nearly 10% of the RA population; however, up to 40–50% of RA patients have radiologic abnormalities on chest CT, suggesting that ILD in the context of RA may be underdiagnosed.

HRCT Image Findings The most common imaging pattern of ILD in patients with RA is a UIP pattern, although NSIP patterns are not uncommon. There is evidence that survival in patients with RA is decreased in patients with a UIP pattern and among those with more extensive fibrosis in general.

Histopathology Histopathologic findings of UIP and NSIP are the most common. Some studies suggest that UIP in the context of RA (as compared to IPF) may present with a reduced number of fibroblastic foci and an increased amount of germinal centers. Comparable to the imaging findings, UIP (and DAD) patterns in patients with RA are associated with reduced survival.

Treatment In contrast with SSc, there are no published clinical trials testing the role of immune suppression in RA-ILD. Extrapolating from the scleroderma experience, immunosuppressive (e.g., prednisone) and cytotoxic (e.g., mycophenolate, azathioprine, cyclophosphamide, and calcineurin inhibitors) agents have been used with variable success. However, RA patients with progressive pulmonary fibrosis may have less of a decline in lung function in response to antifibrotic therapy. Lung transplantation is a viable therapeutic approach for eligible patients with progressive disease that is not responsive to medical therapy.

DERMATOMYOSITIS/POLYMYOSITIS

Clinical Manifestations (Chap. 377) The idiopathic inflammatory myopathies are disorders characterized by immune-mediated destruction and dysfunction of muscle; however, these disorders can affect the skin, joints, cardiovascular system, and lungs. The prevalence of ILD associated with inflammatory myopathy varies by report; however, ILD is present in up to 45% of patients with positive anti-synthetase antibodies. The anti-synthetase syndrome is characterized by positive anti-synthetase antibodies, myositis, fever, Raynaud's phenomenon, mechanic's hands, arthritis, and progressive ILD. There is a subset of anti–Jo-1 antibody–positive individuals who develop a rapidly progressive form of ILD consistent with an acute exacerbation. Some studies have suggested that ILD may be more common in those with other antibodies (e.g., anti-PL-12). Dermatomyositis/polymyositis can occur as an isolated CTD or as a process associated with an underlying malignancy.

HRCT Image Findings Common imaging patterns of ILD in patients with dermatomyositis/polymyositis include those consistent with NSIP with or without evidence for COP. A UIP pattern can also occur. Some studies have suggested that a UIP pattern may be more common among those with anti-PL-12 antibodies.

Histopathology The anti-synthetase syndrome is associated with multiple histopathologic subtypes including NSIP, COP, and UIP. DAD, a histopathologic pattern observed in AIP and acute exacerbations, is associated with rapidly progressive ILD in myositis patients.

Treatment Immunosuppressive (e.g., prednisone) and cytotoxic (e.g., mycophenolate, azathioprine, cyclophosphamide, and calcineurin inhibitors) agents are often used in patients with progressive ILD. Some patients (particularly those with less fibrosis) have been noted to have improved or resolved ILD in response to medical therapy. In small studies, relapses have been more common in patients treated with prednisone alone. Patients who fail immune-suppressive therapy can benefit from lung transplantation.

GRANULOMATOUS ILDS

The most common granulomatous ILD is sarcoidosis, a multisystem disorder of unknown cause where lung involvement is often the most dominant feature; sarcoidosis is discussed in Chap. 379. HP, a granulomatous reaction due to inhalation of organic (e.g., bird fancier's lung secondary to exposure to bird feathers) and inorganic (e.g., coal worker's pneumoconiosis secondary to exposure to coal dusts) dusts, is also an important and common cause of ILD and is discussed in Chap. 299.

Granulomatous Vasculitides (See Chap. 67) These disorders are characterized by blood vessels with inflammatory infiltrates and associated granulomatous lesions with or without the presence of tissue necrosis. The lungs are commonly involved, and a unique feature of these disorders is that hemoptysis can be a presenting symptom. Although laboratory testing is helpful and can provide specific information, biopsies of involved tissue can be essential for making a diagnosis. Many of these disorders include additional systemic manifestations. GPA, previously referred to as Wegener's disease, is an example of a granulomatous vasculitis that commonly affects the lung (including inflammatory infiltrates in small to medium-sized vessels), ears, nose, throat, and kidney (resulting in glomerulonephritis). Common imaging abnormalities of GPA include nodules, patchy ground-glass and consolidative opacities that can be migratory, and hilar lymphadenopathy. Eosinophilic GPA (EG; also referred to as Churg-Strauss syndrome) is another example of a granulomatous vasculitis that affects the lung (including eosinophilic infiltrates in small to medium-sized vessels) and can result in numerous clinical manifestations but frequently includes chronic sinusitis, asthma, and peripheral blood eosinophilia. Common imaging abnormalities of EG include peripheral consolidative opacities that can be migratory and small pleural effusions.

GENETICS AND ILD

Studies of genetic epidemiology have led to important insights in our understanding of ILD. First, studies of families with FPF have demonstrated that unique IIPs can cosegregate with specific genetic variants known to be associated with IPF. This suggests that many genetic variants appear to predispose to interstitial lung injury patterns more broadly than to unique diagnoses specifically. Second, most of the genetic variants associated with FPF are also associated with more sporadic forms of the disease. Third, at least one of the genetic factors most strongly associated with FPF and IPF is common and confers a large increase in the risk of these diseases. At least one copy of a mucin 5B (*MUC5B*) promoter variant is present in ~20% of Caucasian populations and 35–45% of patients with IPF and confers an approximate sixfold increase in the risk of this disease. Fourth, studies of general population samples demonstrate that imaging abnormalities suggestive of an early stage of pulmonary fibrosis in research participants without known ILD are not uncommon (occurring in ~7–9% of adults) and are also associated with the same genetic variants known to be associated with IPF (e.g., the *MUC5B* promoter variant). This latter finding suggests a path forward toward an early detection of IPF. Additional genetic findings demonstrating replicable associations with pulmonary fibrosis include numerous genetic variants in, and adjacent to, genes known to be involved in the regulation of telomere length (e.g., the *TERT* gene, the telomerase RNA component [*TERC*] gene, and the regulator of telomere elongation helicase 1 [*RTEL1*] gene) and surfactant protein genes (e.g., surfactant protein A2 [*SFTPA2*] gene) (Chap. 495), and aggregates of effect estimates from the combined effect of thousands of genetic variants (polygenic risk) score.

Genetic studies have also provided some insights into other forms of ILD. Genome-wide association studies of sarcoidosis have demonstrated numerous variants in genes and in genomic regions that are associated with the disease. Some of these disease-associated variants in sarcoidosis fall in human leukocyte antigen (*HLA*) regions, in regions of genes involved in immune regulation (e.g., interleukin 12B [*IL12B*]), and in regions of genes that are less well understood (butyrophilin-like 2 [*BTNL2*]) but also appear to be involved in T-cell activation. LAM is often associated with genetic variants in the tuberous sclerosis complex genes (e.g., *TSC1* and *TSC2*), consistent with the known evidence that this disease can occur in isolation but also in patients with known tuberous sclerosis. Many genetic factors for rare diseases such as Hermansky-Pudlak syndrome (a rare autosomal recessive disorder that results in pulmonary fibrosis but also includes oculocutaneous albinism, bleeding diatheses, and horizontal nystagmus) have also been discovered (e.g., *HSP1* and *HSP3–7*).

GLOBAL CONSIDERATIONS

The prevalence, clinical presentation, and natural history of most ILDs in European countries resemble those described in the United States. However, as expected, there is growing evidence for racial differences in clinical (rate of acute exacerbations) and genetic (*MUC5B*) attributes between Caucasian and Asian populations. To date, there are limited data on the prevalence of ILD in Hispanics, subjects of African descent, and many other ethnic groups.

FURTHER READING

American Thoracic Society/European Respiratory Society: Consensus classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med 165:277, 2002.

Raghu G et al; ATS/ERS/JRS/ALAT Committee on Idiopathic Pulmonary Fibrosis: An official ATS/ERS/JRS/ALAT statement: Idiopathic pulmonary fibrosis: Evidence-based guidelines for the diagnosis and management. Am J Respir Crit Care Med 183:788, 2011.

Travis WD et al: Idiopathic nonspecific interstitial pneumonia: Report of an American Thoracic Society project. Am J Respir Crit Care Med 177:1338, 2008.

Travis WD et al: An official American Thoracic Society/European Respiratory Society Statement: Ten decade update on IIP's, potential areas for future investigation are proposed (ATS/ERS update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med 188:733, 2013.

305

Disorders of the Pleura

Khalid Ismail, Hilary J. Goldberg, Rebecca M. Baron

The pleural space lies between the lung and the chest wall and normally contains a very thin layer of fluid, which serves as a coupling system. This space is maintained around –3 to –5 cmH_2O because of a constant balance between the elastic recoil of the lung, with its attached visceral pleura, and the counter expansion of the chest wall with attached parietal pleura. Invasion of this space with organisms, inflammatory or cancer cells, or introduction of air, blood, or chyle can lead to significant disease.

PLEURAL EFFUSION

A pleural effusion is present when there is an excess quantity of fluid in the pleural space.

Etiology Under normal conditions, fluid enters the pleural space from the capillaries in the parietal pleura and is removed via the lymphatics in the parietal pleura. Fluid also can enter the pleural space from the interstitial spaces of the lung via the visceral pleura or from the peritoneal cavity via small holes in the diaphragm. The lymphatics have the capacity to absorb 20 times more fluid than is formed normally. Hydrostatic and oncotic pressures play a fundamental role. A pleural effusion develops when there is disturbance in hydrostatic-oncotic balance, leading to pleural fluid formation that overwhelms fluid removal by the lymphatics. It is estimated that 1.5 million Americans per year develop pleural effusion. The most common presentation is that of shortness of breath and occasionally chest pain.

Diagnostic Approach Patients suspected of having a pleural effusion should undergo chest imaging to evaluate its extent. Although chest x-ray with lateral decubitus films have been the standard diagnostic modality (and remains so in some parts of the world), advances in chest ultrasound and computed tomography (CT) imaging have made them invaluable, for both identifying characteristics of the effusion and guiding pleural sampling and/or drainage. When a patient is found to have a pleural effusion, an effort should be made to determine the cause **(Fig. 305-1)**. The first step is to determine whether the effusion is a transudate or an exudate, which requires thoracentesis. A transudative pleural effusion occurs when systemic factors (hydrostatic-oncotic pressures) that influence the formation and absorption of pleural fluid are altered. The leading causes of transudative pleural effusions in the United States are left ventricular failure and cirrhosis. An exudative pleural effusion occurs when local factors that influence the formation and absorption of pleural fluid are altered. The leading causes of exudative pleural effusions are bacterial pneumonia, malignancy, viral infection, and pulmonary embolism **(Table 305-1)**. The primary reason for making this differentiation is that additional diagnostic procedures are indicated with exudative effusions to define the cause.

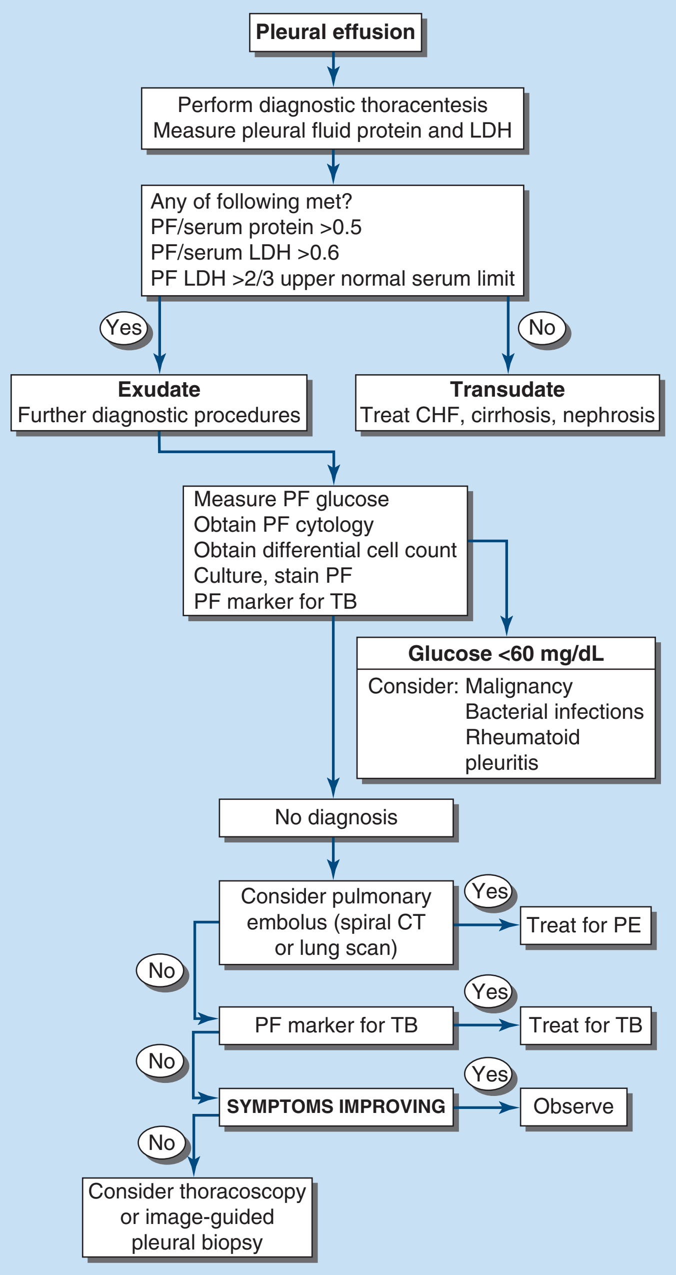
Pleural effusion
Perform diagnostic thoracentesis
Measure pleural fluid protein and LDH
Any of following met?
PF/serum protein >0.5
PF/serum LDH >0.6
PF LDH >2/3 upper normal serum limit
Yes
No
Exudate
Further diagnostic procedures
Transudate
Treat CHF, cirrhosis, nephrosis
Measure PF glucose
Obtain PF cytology
Obtain differential cell count
Culture, stain PF
PF marker for TB
Glucose <60 mg/dL
Consider: Malignancy
Bacterial infections
Rheumatoid
pleuritis
No diagnosis
Consider pulmonary
embolus (spiral CT
or lung scan)
Yes
Treat for PE
No
PF marker for TB
Yes
Treat for TB
No
SYMPTOMS IMPROVING
Yes
Observe
No
Consider thoracoscopy
or image-guided
pleural biopsy

FIGURE 305-1 Approach to the diagnosis of pleural effusions. CHF, congestive heart failure; CT, computed tomography; LDH, lactate dehydrogenase; PE, pulmonary embolism; PF, pleural fluid; TB, tuberculosis.

Transudative and exudative pleural effusions are distinguished by measuring the lactate dehydrogenase (LDH) and protein levels in the pleural fluid. Exudative pleural effusions meet at least one of the following criteria, whereas transudative pleural effusions meet none:

1. Pleural fluid protein/serum protein >0.5
2. Pleural fluid LDH/serum LDH >0.6
3. Pleural fluid LDH more than two-thirds the normal upper limit for serum

These criteria misidentify ~25% of transudates as exudates ("pseudoexudates"), often felt to be related to a "diuresed transudate." If one or more of the exudative criteria are met and the patient is clinically thought to have a condition producing a transudative effusion, alternative criteria can be used. Serum to pleural fluid protein gradient (SPPG) and serum to pleural fluid albumin gradient (SPAG) >3.1 g/dL and

TABLE 305-1 Differential Diagnoses of Pleural Effusions

Transudative Pleural Effusions

1. Congestive heart failure
2. Cirrhosis
3. Nephrotic syndrome
4. Peritoneal dialysis
5. Superior vena cava obstruction
6. Myxedema
7. Urinothorax

Exudative Pleural Effusions

1. Neoplastic diseases
    a. Metastatic disease
    b. Mesothelioma
2. Infectious diseases
    a. Bacterial infections
    b. Tuberculosis
    c. Fungal infections
    d. Viral infections
    e. Parasitic infections
3. Pulmonary embolization
4. Gastrointestinal disease
    a. Esophageal perforation
    b. Pancreatic disease
    c. Intraabdominal abscesses
    d. Diaphragmatic hernia
    e. After abdominal surgery
    f. Endoscopic variceal sclerotherapy
    g. After liver transplant
5. Collagen vascular diseases
    a. Rheumatoid pleuritis
    b. Systemic lupus erythematosus
    c. Drug-induced lupus
    d. Sjögren syndrome
    e. Granulomatosis with polyangiitis (Wegener)
    f. Churg-Strauss syndrome
6. Post–coronary artery bypass surgery
7. Asbestos exposure
8. Sarcoidosis
9. Uremia
10. Meigs' syndrome
11. Yellow nail syndrome
12. Drug-induced pleural disease
    a. Nitrofurantoin
    b. Dantrolene
    c. Methysergide
    d. Bromocriptine
    e. Procarbazine
    f. Amiodarone
    g. Dasatinib
13. Trapped lung
14. Radiation therapy
15. Post–cardiac injury syndrome
16. Hemothorax
17. Iatrogenic injury
18. Ovarian hyperstimulation syndrome
19. Pericardial disease
20. Chylothorax

TABLE 305-2 Disease-Specific Pleural Fluid Tests

SUSPECTED DISEASE	TESTS
Pancreatic disease or esophageal rupture	Pleural fluid amylase
Drug-induced pleural effusion	Pleural fluid eosinophils
Congestive heart failure	Pleural fluid N-terminal pro-brain natriuretic peptide (NT-proBNP)
Chylothorax	Pleural fluid cholesterol and triglycerides
Hemothorax	Pleural fluid hematocrit
Rheumatoid disease	Pleural fluid glucose and pH
Amyloidosis	Congo red staining
Lymphoma	Flow cytometry

>1.2 g/dL, respectively, together can identify pseudoexudates with 100% sensitivity in heart failure and 99% sensitivity in hepatothorax. Elevated pleural fluid cholesterol has also been used, especially when combined with elevated LDH, to favor a true exudate. In addition to describing the appearance of pleural fluid, the following pleural fluid tests should be obtained: glucose level, differential cell count, microbiologic studies, and cytology. More disease-specific diagnostic tests can be obtained depending on the clinical scenario (see examples in Table 305-2).

Effusion Due to Heart Failure The most common cause of pleural effusion is left ventricular dysfunction. Elevated left atrial pressure with elevated pulmonary venous pressure leads to increased amounts of fluid in the lung interstitial spaces exiting in part across the visceral pleura (a classic example of increased hydrostatic pressure), and this overwhelms the capacity of the lymphatics in the parietal pleura to remove the fluid. In patients with heart failure, a diagnostic thoracentesis should be performed if the effusions are not bilateral and comparable in size, if the patient is febrile, or if the patient has pleuritic chest pain, to verify that the patient has a transudative effusion. Otherwise, the patient's heart failure is treated. If the effusion persists despite therapy, a diagnostic thoracentesis should be performed. A pleural fluid N-terminal pro-brain natriuretic peptide (NT-proBNP) level >1500 pg/mL is suggestive of an effusion that is secondary to congestive heart failure and correlates well with serum values.

Parapneumonic Effusion Parapneumonic effusions (PPEs) can be seen in up to 50% of patients with community-acquired pneumonia and are generally the most common cause of exudative pleural effusion in the United States. PPEs are usually reactive (i.e., no organisms identified on culture). In 10% of patients with PPE, infection of the pleural space sets in, leading to a complicated PPE or an empyema, which refers to a fibrinopurulent or grossly purulent effusion, respectively.

Patients with aerobic bacterial pneumonia and pleural effusion may present with an acute febrile illness consisting of pleuritic chest pain, sputum production, and leukocytosis. Patients with anaerobic infections can present with a subacute illness with weight loss, leukocytosis, mild anemia, and a history of some factor that predisposes them to aspiration. These patients can have minimal parenchymal infiltrates with a large effusion.

The possibility of a PPE should be considered whenever a patient with bacterial pneumonia is initially evaluated. Thoracentesis is usually indicated to exclude infection of the pleural space if significant pleural fluid is present. The presence of free-flowing pleural fluid can be demonstrated with a lateral decubitus radiograph, CT of the chest, or ultrasound (Fig. 305-2*A*, *C*). Traditionally, a minimum of 10 mm of free fluid on a lateral decubitus film suggests the safety of thoracentesis. Presently, CT/ultrasound guidance permits a better determination, both for safety of thoracentesis as well as likelihood of a complicated PPE (usually indicated by loculation or septations in the fluid; Fig. 305-2*B*, *D*) that might make bedside thoracentesis more challenging. Factors indicating the likely need for evacuation of the pleural space by more advanced intervention (in increasing order of importance) include the following:

1. Loculated pleural fluid
2. Pleural fluid pH <7.20
3. Pleural fluid glucose <3.3 mmol/L (<60 mg/dL)
4. Positive Gram's stain or culture of the pleural fluid
5. Presence of gross pus in the pleural space

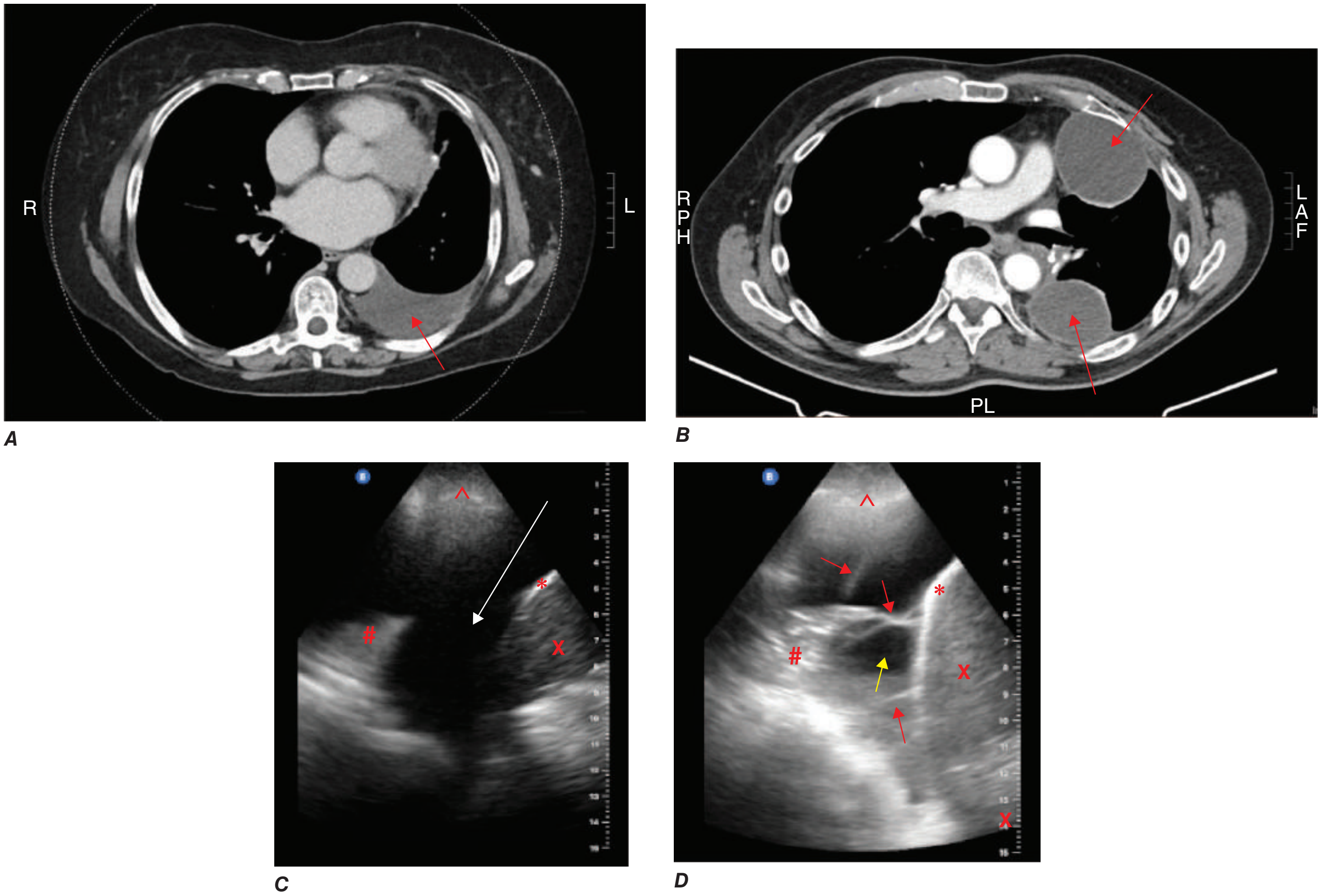

FIGURE 305-2 ***A.*** CT scan from a patient with simple pleural effusion, demonstrating free-flowing fluid in the left pleural space (red arrow). ***B.*** CT scan from a patient with loculated pleural effusion, demonstrating two pockets of fluids (*red arrows*). ***C.*** Ultrasound image of a simple pleural effusion, demonstrating free-flowing fluid (*arrow*), atelectatic lung (#), chest wall (^), diaphragm (*), and liver (x). ***D.*** Ultrasound image of a complex pleural effusion, with loculations (*yellow arrow*) and septations (*red arrows*), diaphragm (*), consolidated lung (#), chest wall (^), and liver (x).

An elevated pleural fluid LDH (>900 IU/L) is another feature that suggests the need for pleural fluid evacuation.

Effusions can progress from exudative to fibrinopurulent to an organizing phase (pleural peel) **(Fig. 305-3)**. If the fluid is fibrinopurulent and cannot be completely drained by thoracentesis, consideration should be given to insertion of a chest tube for drainage and possible instillation of the combination of a fibrinolytic agent (e.g., tissue plasminogen activator, 10 mg) and deoxyribonuclease (5 mg), or performing a thoracoscopy with the breakdown of adhesions if chest tube drainage does not completely evacuate the pleural space. Surgical decortication should be considered when these measures are ineffective.

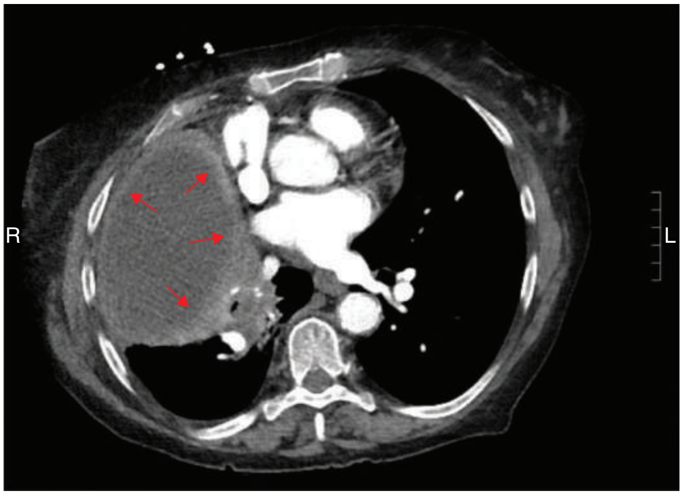

FIGURE 305-3 CT scan from a patient with empyema. Note thick pleural rind with that enhances with contrast (*red arrows*).

Trapped lung (i.e., a lung that cannot re-expand) might develop if a fibrous restrictive peel forms around the visceral pleura, usually in the setting of long-standing pleuro-pulmonary pathology. Although measuring pleural pressure at the time of thoracentesis (pleural manometry) can help confirm trapped lung, a thick pleural rind, presence of air on chest imaging after thoracentesis (termed pneumothorax ex vacuo), and recurrence of effusion shortly after drainage are suggestive of a lung that will not re-expand without decortication.

Effusion Secondary to Malignancy Malignant pleural effusions are the second most common type of exudative pleural effusion. The three tumors that cause ~75% of all malignant pleural effusions are lung carcinoma, breast carcinoma, and lymphoma. Its presence usually portends poor prognosis (<6-month survival). Most patients complain of dyspnea, which is frequently out of proportion to the size of the effusion. The pleural fluid is usually an exudate (although it can rarely be a transudate), and its glucose level may be reduced if the tumor burden in the pleural space is high.

The diagnosis usually is made via cytology of the pleural fluid. If the initial cytologic examination is negative, CT- or ultrasound-guided needle biopsy of pleural thickening or nodules can confirm the diagnosis. Ultimately, thoracoscopic biopsy is most definitive if malignancy is strongly suspected.

Patients with a malignant pleural effusion are treated symptomatically for the most part, since the presence of the effusion indicates disseminated disease, and most malignancies associated with pleural effusion are not curable with chemotherapy. If the patient's lifestyle is compromised by dyspnea and if the dyspnea is not relieved with a therapeutic thoracentesis, one of the following procedures should be considered: (1) insertion of a small indwelling catheter for regular home drainage (improves dyspnea) or (2) tube thoracostomy with the instillation of a sclerosing agent (chemical pleurodesis) versus thoracoscopy with drainage of the effusion and surgical pleurodesis.

Mesothelioma Malignant mesotheliomas are primary tumors that arise from the mesothelial cells that line the pleural cavity; most are related to asbestos exposure. Incidence has dropped because of asbestos remediation measures. Patients with mesothelioma often present with chest pain and shortness of breath. The chest radiograph reveals a pleural effusion, generalized pleural thickening, and a shrunken hemithorax. The diagnosis is usually established with image-guided needle biopsy or thoracoscopy.

Hepatic Hydrothorax Pleural effusions occur in ~5% of patients with cirrhosis and ascites. The predominant mechanism is the direct movement of peritoneal fluid through small openings in the diaphragm into the pleural space. Ascites is common, and the effusion is usually right-sided and often large enough to produce severe dyspnea. Like ascitic fluid, it is usually low in protein (transudative) but can develop into a spontaneous bacterial empyema. Treatment is that of ascites, but if recurrent or persistent, pleurodesis, transjugular intrahepatic portosystemic shunt, or liver transplant might be required.

Effusion Secondary to Pulmonary Embolism The diagnosis most commonly overlooked in the differential diagnosis of a patient with an undiagnosed pleural effusion is pulmonary embolism. Dyspnea is the most common symptom. The pleural fluid is almost always an exudate. The diagnosis is established most often by CT angiography **(Chap. 290)**. Treatment of a patient with a pleural effusion secondary to pulmonary embolism is the same as it is for any patient with pulmonary emboli. If the pleural effusion increases in size after anticoagulation, the patient may have recurrent emboli or another complication, such as a hemothorax or a pleural infection.

Tuberculous Pleuritis/Effusion (See also Chap. 183) In many parts of the world, the most common cause of an exudative pleural effusion is tuberculosis (TB), but tuberculous effusions are relatively uncommon in the United States. Tuberculous pleural effusions usually are associated with primary TB and are thought to be due primarily to a hypersensitivity reaction to tuberculous protein in the pleural space. Patients with tuberculous pleuritis present with fever, weight loss, dyspnea, and/or pleuritic chest pain. The pleural fluid is exudative and predominantly lymphocytic. The diagnosis is suggested by demonstrating high levels of TB markers in the pleural fluid (adenosine deaminase >40 IU/L or interferon γ >140 pg/mL).

Diagnosis is established by culture of the pleural fluid for acid-fast bacilli, needle biopsy of the pleura, or thoracoscopy. Guided percutaneous needle biopsy has largely replaced closed needle biopsy, except in parts of the world where TB is prevalent. The culture yield of pleural biopsy is higher than that of pleural fluid. The recommended treatment of pleural and pulmonary TB is identical **(Chap. 183)**.

Effusion Secondary to Viral Infection Viral infections are probably responsible for a sizable percentage of undiagnosed exudative pleural effusions. In many series, no diagnosis is established for ~20% of exudative effusions, and these effusions usually resolve spontaneously with no long-term residua. The importance of these effusions is that one should not be too aggressive in trying to establish a diagnosis for the undiagnosed effusion if the patient is improving clinically.

Chylothorax A chylothorax occurs when the thoracic duct is disrupted and chyle accumulates in the pleural space. The most common cause of chylothorax is trauma (most frequently thoracic surgery), but it also may result from tumors in the mediastinum. Patients with chylothorax may present with dyspnea, and a large pleural effusion is often present on the chest radiograph. Thoracentesis usually reveals milky fluid, and biochemical analysis reveals a triglyceride level that exceeds 1.2 mmol/L (110 mg/dL). Patients with chylothorax and no obvious trauma should have a lymphangiogram and a chest CT scan to assess the mediastinum for lymphadenopathy. The treatment of choice for most chylothoraces is insertion of a chest tube plus the administration of octreotide and modification of diet to eliminate enteral fat intake. If these modalities fail, percutaneous transabdominal thoracic duct blockage effectively controls most chylothoraces. An alternative treatment is ligation of the thoracic duct. Patients with chylothoraces should not undergo prolonged tube thoracostomy with chest tube drainage because this can lead to malnutrition and immunologic incompetence.

Hemothorax When a diagnostic thoracentesis reveals bloody pleural fluid, a hematocrit should be obtained on the pleural fluid. If the hematocrit is more than one-half of that in the peripheral blood, the patient is considered to have a hemothorax. Most hemothoraces are the result of trauma; other causes include rupture of a blood vessel or tumor. Most patients with hemothorax should be treated with tube thoracostomy, which allows continuous quantification of bleeding. If the bleeding emanates from a laceration of the pleura, apposition of the two pleural surfaces is likely to stop the bleeding. If the pleural hemorrhage exceeds 200 mL/h, consideration should be given to angiographic coil embolization, thoracoscopy, or thoracotomy.

Miscellaneous Causes of Pleural Effusion There are many other causes of pleural effusion (Table 305-2). Key features of some of these conditions are as follows: If the pleural fluid amylase level is elevated, the diagnosis of esophageal rupture or pancreatic disease is likely. If the patient is febrile, has predominantly polymorphonuclear cells in the pleural fluid, and has no pulmonary parenchymal abnormalities, an intraabdominal abscess should be considered. The diagnosis of an asbestos pleural effusion is one of exclusion. Benign ovarian tumors can produce ascites and a pleural effusion (Meigs' syndrome), as can the ovarian hyperstimulation syndrome. Several drugs can cause pleural effusion; the associated fluid may be eosinophilic. Pleural effusions commonly occur after coronary artery bypass graft (CABG) surgery. Effusions occurring within the first weeks after CABG are typically left-sided and bloody, with large numbers of eosinophils, and respond to one or two therapeutic thoracenteses. Effusions occurring after the first few weeks of CABG are typically left-sided and clear yellow, with lymphocytic predominance, and tend to recur. Other medical manipulations that induce pleural effusions include abdominal surgery; radiation therapy; liver, lung, or heart transplantation; and complications of the intravascular insertion of central lines.

PNEUMOTHORAX

Pneumothorax is the presence of gas in the pleural space. A spontaneous pneumothorax is one that occurs without antecedent trauma to the thorax. A primary spontaneous pneumothorax occurs in the absence of underlying lung disease, whereas a secondary pneumothorax occurs in its presence. A traumatic pneumothorax results from penetrating or nonpenetrating chest injuries. A tension pneumothorax is a pneumothorax in which the pressure in the pleural space is positive throughout the respiratory cycle.

Primary Spontaneous Pneumothorax Primary spontaneous pneumothoraces are usually due to rupture of apical pleural blebs, small cystic spaces that lie within or immediately under the visceral pleura. Approximately one-half of patients with an initial primary spontaneous pneumothorax will have a recurrence. Conservative management with careful observation can be considered in adults who are asymptomatic or minimally symptomatic. Outpatient observational management is an option for low-risk patients with a good support system. Otherwise, the initial recommended treatment is needle aspiration or tube drainage. If the lung does not expand or if the patient has a recurrent pneumothorax, thoracoscopy with stapling of blebs and pleurodesis is usually indicated. Thoracoscopy or thoracotomy with surgical pleurodesis is almost 100% successful in preventing recurrences.

Secondary Pneumothorax Most secondary pneumothoraces are due to chronic obstructive pulmonary disease, but pneumothoraces have been reported with virtually every lung disease. Pneumothorax in patients with lung disease can be more life-threatening than it is in normal individuals because of the lack of pulmonary reserve in these patients. Nearly all patients with secondary pneumothorax should be treated with tube drainage. Many will need thoracoscopy or thoracotomy with stapling of blebs and surgical pleurodesis. If the patient is not a good operative candidate or refuses surgery, chemical pleurodesis should be considered.

Traumatic Pneumothorax Traumatic pneumothoraces can result from both penetrating and nonpenetrating chest trauma. Traumatic pneumothoraces are usually treated with tube drainage unless they are very small. If a hemopneumothorax is present, one chest tube might be placed in the superior part of the hemithorax to evacuate the air and another in the inferior part of the hemithorax to remove the blood. Iatrogenic pneumothorax is a type of traumatic pneumothorax that is becoming more common. The leading causes are transthoracic needle aspiration, thoracentesis, and complications of the insertion of central intravenous catheters. Most can be managed with supplemental oxygen or aspiration, but if these measures are unsuccessful, tube drainage should be performed.

Tension Pneumothorax This condition usually occurs during mechanical ventilation or resuscitative efforts. The positive pleural pressure is life-threatening both because ventilation is severely compromised and because the positive pressure is transmitted to the mediastinum, resulting in decreased venous return to the heart and reduced cardiac output. Difficulty in ventilation during resuscitation or high peak inspiratory pressures during mechanical ventilation strongly suggest the diagnosis. The diagnosis is made by physical examination showing an enlarged hemithorax with no breath sounds, hyperresonance to percussion, shift of the mediastinum to the contralateral side, and hypotension. Tension pneumothorax must be treated as a medical emergency. If the tension in the pleural space is not relieved, the patient is likely to die from inadequate cardiac output or marked hypoxemia. A large-bore needle should be inserted into the pleural space through the second anterior intercostal space. If large amounts of gas escape from the needle after insertion, the diagnosis is confirmed. The needle should be left in place until a chest tube can be inserted.

ACKNOWLEDGMENT

Richard W. Light contributed to this chapter in the 21st edition and some material from that chapter has been retained here.

FURTHER READING

Feller-Koppman D, Light R: Pleural disease. N Engl J Med 378:740, 2018.

Light RW: *Pleural Diseases*, 6th ed. Lippincott, Williams and Wilkins, Baltimore, 2013.

Rahman NM et al: Intrapleural use of tissue plasminogen activator and DNase in pleural infection. N Engl J Med 365:518, 2011.

Roberts ME et al: British Thoracic Society Guideline for pleural disease. Thorax 78:1143, 2023.

306 Disorders of the Mediastinum

Stefi F. Lee, Rebecca M. Baron, Hilary J. Goldberg

The mediastinum is the region between the pleural sacs bound by the thoracic inlet superiorly and the diaphragm inferiorly. It is separated into three compartments, and diseases can arise from the anatomical structures residing in each respective compartment **(Fig. 306-1)**. The *anterior mediastinum* comprises the retrosternal space anterior to the pericardium and brachiocephalic vasculature; it contains the thymus gland, the anterior mediastinal lymph nodes, and the internal mammary arteries and veins, and can give rise to several neoplasms. The *middle mediastinum* lies between the anterior and posterior mediastina and contains the heart; the ascending and transverse arches of the aorta; the venae cavae; the brachiocephalic arteries and veins; the phrenic nerves; the trachea, the main bronchi, and their contiguous lymph nodes; and the pulmonary arteries and veins. The *posterior mediastinum* extends from the posterior border of the heart and trachea to the vertebral column posteriorly. It contains the descending thoracic aorta, the esophagus, the thoracic duct, the azygos and hemiazygos veins, and the posterior group of mediastinal lymph nodes. Disorders of the mediastinum encompass a broad range of diseases including but not exclusive to neoplastic and nonneoplastic masses, congenital or acquired malformations of anatomical structures, infections, and chronic fibrosing mediastinitis, which will be addressed in this chapter.

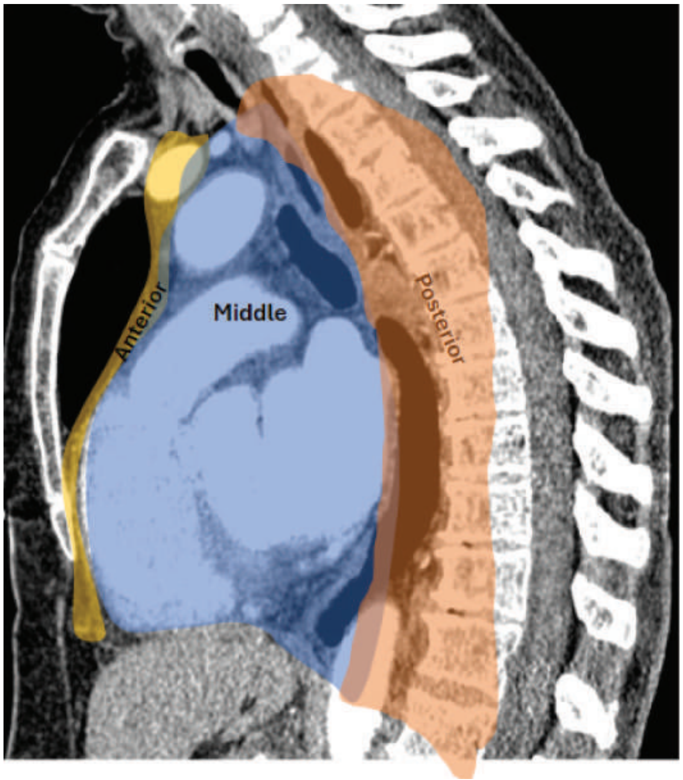

FIGURE 306-1 Sagittal chest CT of the mediastinum and its three separate compartments.

MEDIASTINAL MASSES

The first step in evaluating a mediastinal mass is to place it in one of the three mediastinal compartments, since each has different characteristic lesions **(Table 306-1)**. The anterior mediastinum can give rise to neoplasms including thymomas, germ cell tumors, teratomas, lymphomas, and thyroid goiter. Thymomas are the most common neoplasm in the anterior mediastinum and are closely tied to paraneoplastic syndromes, most notably, myasthenia gravis, the diagnosis of which is supported by the presence of serum anti-acetylcholine receptor antibodies. Teratomas, the most common type of germ cell tumor, are multilobulated cystic lesions that contain fat, soft tissue, and calcium. Elevated tumor markers, α fetoprotein and β-human chorionic gonadotropin, distinguish between seminomatous and nonseminomatous germ cell tumors, which can guide treatment. Seminomas are generally responsive to radiation therapy, whereas nonseminomatous germ cell tumors are treated with standard chemotherapy. Computed tomography (CT) scanning is an important imaging technique for evaluating mediastinal masses as it can attenuate fat, water, calcifications, and air in most instances. Diffusion-weighted magnetic resonance imaging aids in improved tumor characterization and differentiation between cystic and solid structures. Lymphomas appear as homogeneous soft

TABLE 306-1 The Three Compartments of the Mediastinum

	ANTERIOR COMPARTMENT	MIDDLE COMPARTMENT	POSTERIOR COMPARTMENT
Anatomical boundaries	Manubrium and sternum anteriorly; pericardium, aorta, and brachiocephalic vessels posteriorly	Anterior mediastinum anteriorly; posterior mediastinum posteriorly	Pericardium and trachea anteriorly; vertebral column posteriorly
Contents	Thymus gland, anterior mediastinal lymph nodes, internal mammary arteries, and veins	Pericardium, heart, ascending and transverse arch of aorta, superior and inferior vena cavae, brachiocephalic arteries and veins, phrenic nerves, trachea, and mainstem bronchi and their contiguous lymph nodes, pulmonary arteries, and veins	Descending thoracic aorta, esophagus, thoracic duct, azygos and hemiazygos veins, sympathetic chains, and the posterior group of mediastinal lymph nodes
Common abnormalities	Thymoma, lymphomas, teratomatous neoplasms, thyroid masses, parathyroid masses, mesenchymal tumors, giant lymph node hyperplasia, hernia through foramen of Morgagni	Metastatic lymph node enlargement, granulomatous lymph node enlargement, pleuropericardial cysts, bronchogenic cysts, masses of vascular origin	Neurogenic tumors, meningocele, meningomyelocele, gastroenteric cysts, esophageal diverticula, hernia through foramen of Bochdalek, extramedullary hematopoiesis

tissue structures and can present at various stages. At more advanced stages, lymphomas can present with anterior and middle mediastinal lymphadenopathy. Mediastinoscopy is often performed for sampling of paratracheal and subcarinal lymph nodes. Biopsy can also be achieved via percutaneous fine-needle aspiration or endoscopic transesophageal or endobronchial ultrasound. Scintigraphy with radioactive iodine scan can efficiently establish the diagnosis of an intrathoracic goiter that can also lie in the anterior mediastinum.

Surgical resection is the mainstay of treatment for most anterior mediastinal masses like in localized early stages of thymic carcinomas, symptomatic teratomas, and mediastinal goiters. Lymphoma is treated with multimodal therapy with a combination of radiation and chemotherapy; surgery is rarely indicated. Ultimately, the involvement of a multidisciplinary interprofessional team of oncologists, pulmonologists, radiologists, and cardiothoracic surgeons is necessary to guide best treatment practices and provide personalized care for each patient.

The middle mediastinum consists of central vascular and bronchogenic structures. Most middle mediastinal masses are benign and result from abnormalities in embryonic development, such as bronchogenic cysts (that arise from abnormal ventral foregut budding), duplication cysts (that arise from persistent diverticulum of the dorsal bud of the foregut), and pericardial cysts (that arise from abnormal fusion of pericardial recesses). In contrast, advanced stages of malignancy often do not remain localized within one compartment and can traverse borders through hematogenous and lymphatic spread (Fig. 306-2). It is also important to note that, due to the presence of central vasculature in the middle mediastinum, disruption of these structures can lead to acute life-threatening disease. For example, superior vena cava (SVC) syndrome is a medical emergency that presents with severe dyspnea and facial and upper extremity edema from venous distension because of compression of the SVC. SVC obstruction often occurs due to a malignant process; lung cancer, particularly small-cell lung carcinoma, accounts for up to 50% of cases. Endovenous recanalization, usually performed via an interventional radiology approach, restores venous return and reduces the risk of respiratory failure and death.

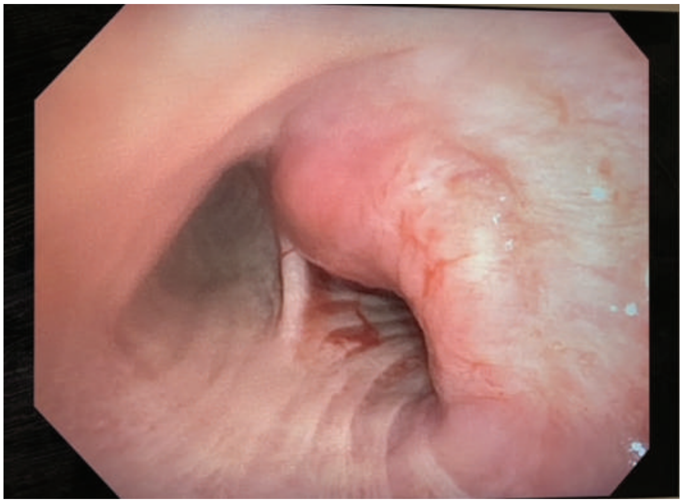

FIGURE 306-2 Bronchoscopic view of main carina with a malignant mass encroaching on the right mainstem airway. *(Courtesy of Dr. Majid Shafiq.)*

Neurogenic tumors are the most common cause of posterior mediastinal tumors. They may arise from peripheral, autonomic or paraganglionic nervous systems and are often benign. Other causes include diaphragmatic hernia, meningoceles, esophageal diverticula, or extramedullary hematopoiesis. Barium studies of the gastrointestinal tract are indicated in many patients with posterior mediastinal lesions because hernias, diverticula, and achalasia are readily diagnosed in this manner. Finally, mediastinal masses have a wide spectrum of disease severity, for which reason it may be difficult to determine the best course of action. When presented with a mediastinal mass, considering factors such as clinical symptoms, age of the patient, and location of the mediastinal tumor helps to determine the likelihood of malignancy and need for intervention.

ACUTE MEDIASTINITIS

Mediastinitis refers to an acute inflammatory and/or infectious condition involving the mediastinum. Cases of acute mediastinitis are usually due to esophageal perforation, occur after median sternotomy for cardiac surgery, or are seen related to infections descending from the neck, oral cavity, or facial area. Patients with esophageal rupture are acutely ill with chest pain and dyspnea due to the mediastinal infection. The esophageal rupture can occur spontaneously or as a complication of esophagoscopy or the insertion of a Blakemore tube. Appropriate treatment consists of exploration of the mediastinum with primary repair of the esophageal tear and drainage of the pleural space and the mediastinum.

The incidence of mediastinitis after median sternotomy is 0.4–5.0%. Patients most commonly present with wound drainage. Other presentations include sepsis and a widened mediastinum. The diagnosis usually is established with mediastinal needle aspiration. Treatment includes immediate drainage, debridement, and parenteral antibiotic therapy, but the mortality rate still exceeds 20%.

CHRONIC MEDIASTINITIS AND FIBROSING MEDIASTINITIS

The spectrum of chronic mediastinitis ranges from granulomatous inflammation of the lymph nodes in the mediastinum to fibrosing mediastinitis. Most cases of fibrosing mediastinitis follow a history of granulomatous infections, commonly due to histoplasmosis or tuberculosis, but sarcoidosis, silicosis, and fungal diseases are at times causative. Patients with granulomatous mediastinitis are usually asymptomatic. Those with fibrosing mediastinitis usually have variable symptoms depending on the extent and location of invasion of mediastinal structures such as the SVC or large airways, phrenic or recurrent laryngeal nerve paralysis, or obstruction of the pulmonary artery or proximal pulmonary veins. Fibrosing mediastinitis can be devastating in severe cases. Common clinical manifestations include dyspnea, chest pain, SVC syndrome, hemoptysis, dysphagia, and Horner's syndrome. Diagnosis involves a combination of clinical, radiologic, and histopathologic findings. Imaging modalities such as CT and magnetic resonance imaging can delineate the extent of fibrosis. Tissue biopsy may be necessary to rule out malignancy. Treatment in chronic fibrosing mediastinitis is focused on relieving symptoms and preventing further complications. Therapy often includes a combination of surgical debulking or resection and endovascular interventions, including stenting, if veins or arteries are involved. Glucocorticoids have been used with variable success in cases of autoimmune-related and fibro-inflammatory causes of fibrosing mediastinitis. In cases of active fungal infection, antifungal therapies can be used, although they are generally ineffective. Emerging reports have shown success with the use of rituximab in mitigating symptoms and reducing metabolic activity in refractory disease.

■ PNEUMOMEDIASTINUM

In this condition, gas is noted in the interstices of the mediastinum. The three main causes are (1) alveolar rupture with dissection of air into the mediastinum; (2) perforation or rupture of the esophagus, trachea, or main bronchi; and (3) dissection of air from the neck or the abdomen into the mediastinum. Typically, there is severe substernal chest pain with or without radiation into the neck and arms. The physical examination usually reveals subcutaneous emphysema in the suprasternal notch and *Hamman's sign*, which is a crunching or clicking noise synchronous with the beating heart and is best heard with a stethoscope applied to the anterior chest wall in the left lateral decubitus position. The diagnosis is confirmed with a chest radiograph. Usually no treatment is required, but the mediastinal air will be absorbed faster if the patient inspires high concentrations of oxygen. If mediastinal structures are compressed, the compression can be relieved with needle aspiration.

Acknowledgment

Richard W. Light contributed to this chapter in the 21st edition and some material from that chapter has been retained here.

■ FURTHER READING

Ahua J et al: Approach to imaging of mediastinal masses. Diagnostics (Basel) 13:3171, 2023.

Carter BW et al: ITMIG classification of mediastinal compartments and multidisciplinary approach to mediastinal masses. Radiographics 37:413, 2017.

Duwe B et al: Tumors of the mediastinum. Chest 128:2893, 2005.

Juanpere S et al: A diagnostic approach to the mediastinal masses. Insights Imaging 4:29, 2013.

Manyeruke FD et al: Idiopathic fibrosing mediastinitis. Afr J Thorac Crit Care Med 27:10.7196/AJTCCM.2021.v27i2.064, 2021.

Markman M: Diagnosis and management of superior vena cava syndrome. Cleve Clin J Med 66:59, 1999.

Ponamgi SP et al: Catheter-based intervention for pulmonary vein stenosis due to fibrosing mediastinitis: The Mayo Clinic experience. Int J Cardiol Heart Vasc 8:103, 2015.

Westerly BD et al: Targeting B lymphocytes in progressive fibrosing mediastinitis. Am J Respir Crit Care Med 190:1069, 2014.

307 Disorders of Ventilation

Khalid Ismail, George R. Washko

DEFINITION AND PHYSIOLOGY

In health, the arterial level of carbon dioxide (Paco$_2$) is maintained between 37 and 43 mmHg at sea level. All acute and chronic disorders of ventilation result in abnormal measurements of the partial pressure of carbon dioxide in the arterial blood, Paco$_2$. Paco$_2$ will be above the normal physiologic range if production exceeds clearance by the lung. This chapter reviews chronic ventilatory disorders.

The continuous production of carbon dioxide (CO$_2$) by cellular metabolism necessitates its efficient elimination by the respiratory system. The relationship between CO$_2$ production and Paco$_2$ is described by the equation: Paco$_2$ = (k) ($\dot{V}CO_2$)/$\dot{V}A$, where $\dot{V}CO_2$ represents the carbon dioxide production, k is a constant, and $\dot{V}A$ is fresh gas alveolar ventilation **(Chap. 296)**. Alveolar ventilation is not equivalent to minute ventilation (calculated as the product of respiratory rate and tidal breath volume) because not all parts of the ventilated respiratory system participate in gas exchange. The portion of each tidal breath that remains within the conducting airways is called the dead space fraction, represented as V_d/V_t. $\dot{V}A$ is the minute ventilation minus that portion of minute ventilation apportioned to dead space and is calculated as minute ventilation × (1 – V_d/V_t). As such, all disturbances of Paco$_2$ must reflect altered CO$_2$ production, minute ventilation, or dead space fraction.

Diseases that alter $\dot{V}CO_2$ are often acute (e.g., sepsis, burns, or pyrexia), and their contribution to ventilatory abnormalities and/or respiratory failure is reviewed elsewhere. Chronic ventilatory disorders typically involve inappropriate levels of minute ventilation or increased dead space fraction, both of which result in insufficient alveolar ventilation to maintain the Paco$_2$ in the normal physiologic range. This is called alveolar hypoventilation. Note that a person may exhibit abnormally rapid breathing and have an elevated minute ventilation but still suffer from alveolar hypoventilation (because of an increased fraction of dead space ventilation). An understanding of this apparent paradox and more general disorders of ventilation requires a review of the normal respiratory cycle.

The spontaneous cycle of inspiration and expiration is automatically generated in the brainstem. Two groups of neurons located within the medulla are particularly important: the dorsal respiratory group (DRG) and the ventral respiratory column (VRC). These neurons have widespread projections including the descending projections into the contralateral spinal cord where they perform many functions. They initiate activity in the phrenic nerve/diaphragm, project to the upper airway muscle groups and spinal respiratory neurons, and innervate the intercostal and abdominal muscles that participate in normal respiration. The DRG acts as the initial integration site for many of the afferent nerves relaying information about Pao$_2$, Paco$_2$, pH, and blood pressure from the carotid and aortic chemoreceptors and baroreceptors to the central nervous system (CNS). In addition, the vagus nerve relays information from stretch receptors and juxtapulmonary-capillary receptors in the lung parenchyma and chest wall to the DRG. The respiratory rhythm is generated within the VRC as well as the more rostrally located parafacial respiratory group (pFRG), which is particularly important for the generation of active expiration. One particularly important area within the VRC is the so-called pre-Bötzinger complex. This area is responsible for the generation of various forms of inspiratory activity, and lesioning of the pre-Bötzinger complex leads to the complete cessation of breathing. The neural output of these medullary respiratory networks can be voluntarily suppressed or augmented by input from higher brain centers and the autonomic nervous system. During normal sleep, there is an attenuated response to hypercapnia and hypoxemia, resulting in mild nocturnal hypoventilation that corrects upon awakening.

Once neural input has been delivered to the respiratory pump muscles, normal gas exchange requires an adequate amount of respiratory muscle strength to overcome the elastic and resistive loads of the respiratory system **(Fig. 307-1*A*) (also see Chap. 296)**. In health, the strength of the respiratory muscles readily accomplishes this important functional response, and normal respiration continues indefinitely. Reduction in respiratory drive or neuromuscular competence or substantial increase in respiratory load can diminish minute ventilation, resulting in hypercapnia **(Fig. 307-1*B*)**. Alternatively, if normal respiratory muscle strength is coupled with excessive respiratory drive, then alveolar hyperventilation ensues and leads to hypocapnia **(Fig. 307-1*C*)**.

HYPOVENTILATION

■ CLINICAL FEATURES

Diseases that reduce minute ventilation or increase dead space fall into four major categories: parenchymal lung and chest wall disease, obesity

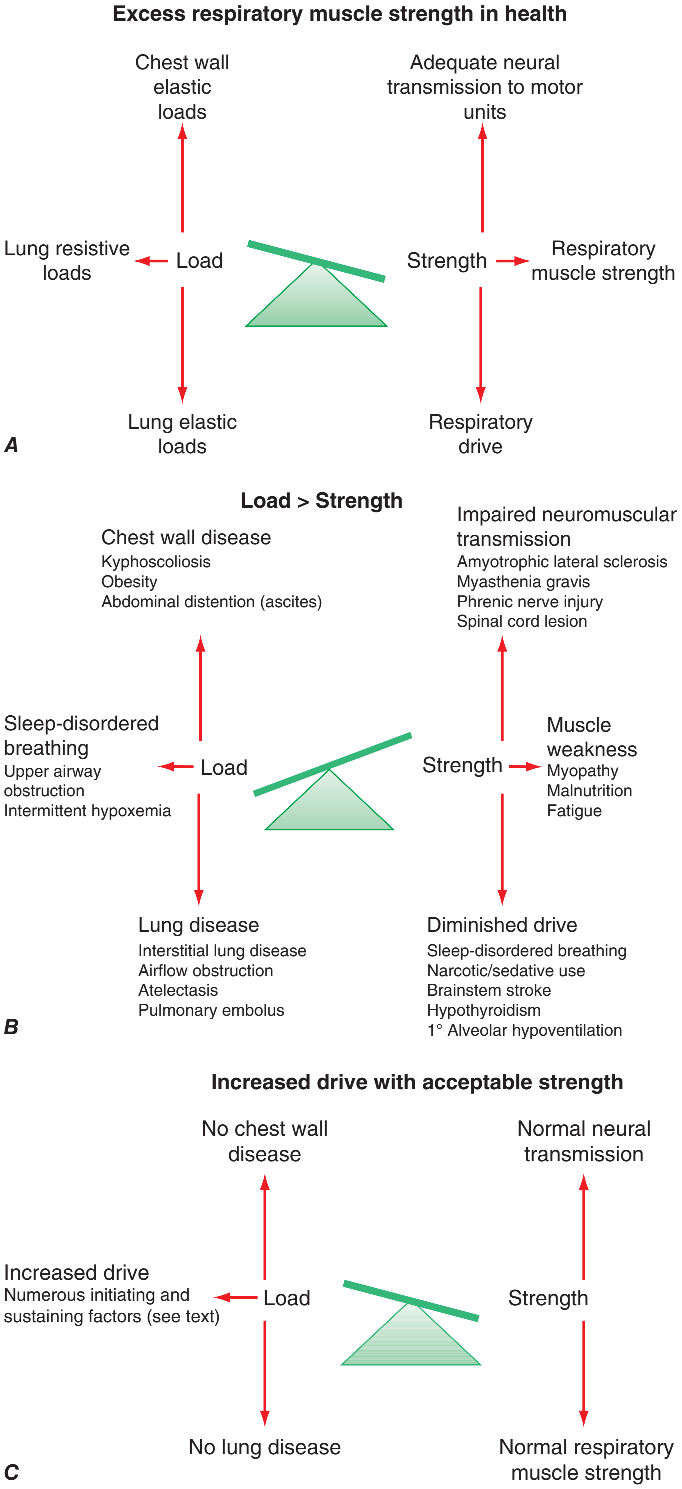
Excess respiratory muscle strength in health
Chest wall elastic loads
Adequate neural transmission to motor units
Lung resistive loads
Load
Strength
Respiratory muscle strength
Lung elastic loads
Respiratory drive
A
Load > Strength
Chest wall disease
Kyphoscoliosis
Obesity
Abdominal distention (ascites)
Impaired neuromuscular transmission
Amyotrophic lateral sclerosis
Myasthenia gravis
Phrenic nerve injury
Spinal cord lesion
Sleep-disordered breathing
Upper airway obstruction
Intermittent hypoxemia
Load
Strength
Muscle weakness
Myopathy
Malnutrition
Fatigue
Lung disease
Interstitial lung disease
Airflow obstruction
Atelectasis
Pulmonary embolus
Diminished drive
Sleep-disordered breathing
Narcotic/sedative use
Brainstem stroke
Hypothyroidism
1° Alveolar hypoventilation
B
Increased drive with acceptable strength
No chest wall disease
Normal neural transmission
Increased drive
Numerous initiating and sustaining factors (see text)
Load
Strength
No lung disease
Normal respiratory muscle strength
C

FIGURE 307-1 Examples of balance between respiratory system strength and load. ***A.*** Excess respiratory muscle strength in health. ***B.*** Load greater than strength. ***C.*** Increased drive with acceptable strength.

TABLE 307-1 Signs and Symptoms of Hypoventilation

Dyspnea during activities of daily living
Orthopnea in diseases affecting diaphragm function
Poor-quality sleep
Daytime hypersomnolence
Early morning headaches
Anxiety
Impaired cough in neuromuscular diseases

hypoventilation syndrome, neuromuscular disease, and respiratory drive disorders (Fig. 307-1*B*). "Pump failure" refers to disorders affecting the respiratory drive, neuronal pathways, neuromuscular junction, and muscles. The clinical manifestations of hypoventilation syndromes are nonspecific **(Table 307-1)** and vary depending on the severity of hypoventilation, the rate at which hypercapnia develops, the degree of compensation for respiratory acidosis, and the underlying disorder. Patients with parenchymal lung or chest wall disease typically present with shortness of breath and diminished exercise tolerance. Episodes of increased dyspnea and sputum production are hallmarks of obstructive lung diseases such as chronic obstructive pulmonary disease (COPD), whereas progressive dyspnea and cough are common in interstitial lung diseases. Excessive daytime somnolence, poor-quality sleep, and snoring are common among patients with sleep-disordered breathing and obesity hypoventilation syndrome. Sleep disturbance and orthopnea are also described in neuromuscular disorders. As neuromuscular weakness progresses, the respiratory muscles, including the diaphragm, are placed at a mechanical disadvantage in the recumbent position owing to the upward movement of the abdominal contents. New-onset orthopnea is one of the earliest symptoms and often heralds reduced respiratory muscle force generation. More commonly, however, extremity weakness or bulbar symptoms develop prior to respiratory muscle involvement in neuromuscular diseases such as amyotrophic lateral sclerosis (ALS) or muscular dystrophy. The characteristic finding of "paradoxical breathing" on physical examination (the abdomen moving inward instead of outward with deep inspiration while supine) indicates significant diaphragmatic weakness or paralysis.

Patients with respiratory drive disorders do not have symptoms distinguishable from other causes of chronic hypoventilation and are usually secondary in nature. A cause can usually be identified with a detailed medical history, including medications, or illicit drug use. The diagnosis of a primary respiratory drive disorder is otherwise that of exclusion.

The clinical course of patients with chronic hypoventilation follows a characteristic sequence: an asymptomatic stage where daytime $\mathrm{Pa{\small O}_2}$ and $\mathrm{Pa{\small CO}_2}$ are normal, then nocturnal hypoventilation develops, initially during rapid eye movement (REM) sleep and later in non-REM sleep. As disease progresses, the vital capacity and corresponding volume of the tidal breaths decrease, leading to a further drop in alveolar ventilation. Eventually daytime hypercapnia develops. Symptoms can develop at any point along this time course and often depend on the pace of underlying disease progression. Regardless of cause, the hallmark of all alveolar hypoventilation syndromes is an increase in alveolar $\mathrm{P{\small CO}_2}$ ($\mathrm{P{\small A}{\small CO}_2}$) and therefore in $\mathrm{Pa{\small CO}_2}$. The resulting continuous respiratory acidosis eventually leads to a compensatory increase in plasma bicarbonate concentration. The increase in $\mathrm{Pa{\small CO}_2}$ displaces oxygen in the alveolus, leading to a decrease in $\mathrm{P{\small A}{\small O}_2}$, often resulting in hypoxemia. If chronic, hypoxemia can stimulate erythropoiesis and thereby induce secondary erythrocytosis. The combination of chronic hypoxemia and hypercapnia may also induce pulmonary vasoconstriction, leading eventually to pulmonary hypertension, right ventricular hypertrophy, and right heart failure.

DIAGNOSIS

The hallmark of the diagnosis of hypoventilation is a $\mathrm{Pa{\small CO}_2}$ ≥45 mmHg, measured in arterial blood gas (ABG) analysis. A venous blood gas can be an alternative if ABG is not available, keeping in mind that venous $\mathrm{P{\small CO}_2}$ is normally several points higher than $\mathrm{Pa{\small CO}_2}$. Elevated serum bicarbonate (i.e., total serum CO_2, which equals calculated bicarbonate plus dissolved CO_2) in the absence of volume depletion is suggestive of hypoventilation. However, it is important to point out that a serum bicarbonate level <27 mmol/L in the setting of normal renal function makes the diagnosis of hypoventilation very unlikely. By contrast, a serum bicarbonate level ≥27 mmol/L should trigger clinicians to measure $\mathrm{Pa{\small CO}_2}$ as a confirmatory diagnostic test. Therefore, serum bicarbonate can be used as a sensitive but not specific test to detect hypercapnia. An ABG demonstrating elevated $\mathrm{Pa{\small CO}_2}$ combined with normal pH confirms chronic alveolar hypoventilation.

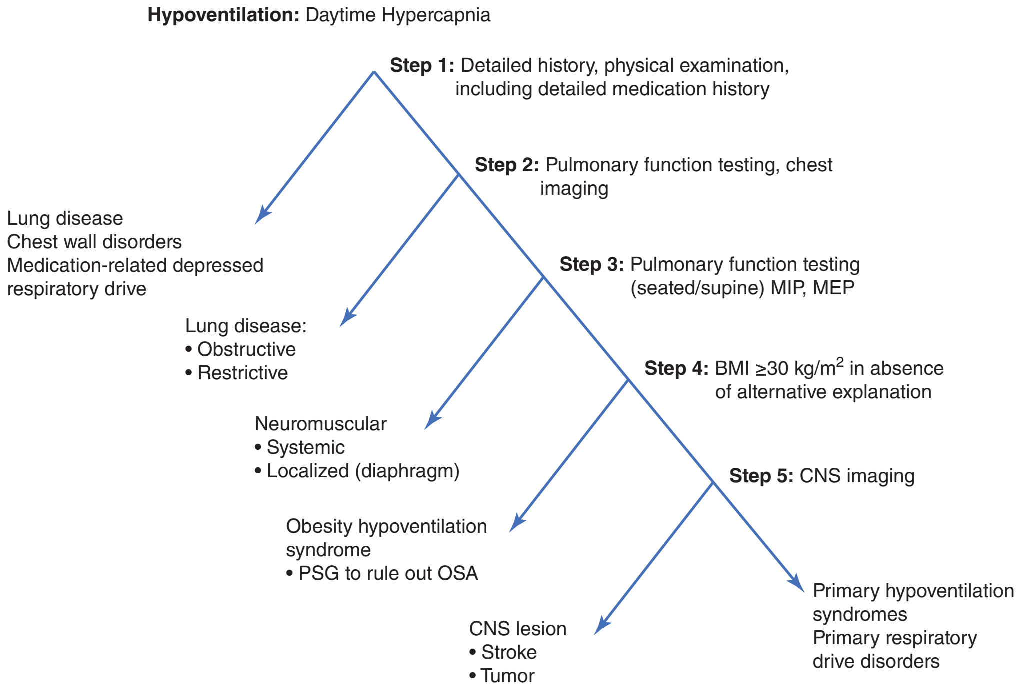
Hypoventilation: Daytime Hypercapnia
Step 1: Detailed history, physical examination, including detailed medication history
Lung disease
Chest wall disorders
Medication-related depressed respiratory drive
Step 2: Pulmonary function testing, chest imaging
Lung disease:
• Obstructive
• Restrictive
Step 3: Pulmonary function testing (seated/supine) MIP, MEP
Neuromuscular
• Systemic
• Localized (diaphragm)
Step 4: BMI ≥30 kg/m2 in absence of alternative explanation
Obesity hypoventilation syndrome
• PSG to rule out OSA
Step 5: CNS imaging
CNS lesion
• Stroke
• Tumor
Primary hypoventilation syndromes
Primary respiratory drive disorders

FIGURE 307-2 A diagnostic algorithm to determine the etiology of hypoventilation. Note that more than one condition may be present. BMI, body mass index; CNS, central nervous system; MEP, maximal expiratory pressure; MIP, maximal inspiratory pressure; OSA, obstructive sleep apnea; PSG, polysomnography.

The subsequent evaluation to identify an etiology should initially focus on whether the patient has lung disease or chest wall abnormalities. Physical examination, imaging studies (chest x-ray and/or computed tomography [CT] scan), and pulmonary function tests are sufficient to identify most lung/chest wall disorders leading to hypercapnia **(Fig. 307-2)**. If the ventilatory apparatus (lung, airways, chest wall) is not responsible for chronic hypercapnia, then the focus should shift to the respiratory pump (i.e., respiratory drive and neuromuscular disorders). In respiratory drive disorders, there is a blunted response in minute ventilation to elevated CO_2 and/or low O_2. As mentioned earlier, secondary forms of the disease are far more common and can be excluded by careful review of medications, including sedating medications and chronic narcotic use, or metabolic derangements including significant hypothyroidism. Brain imaging (CT scan or magnetic resonance imaging) can sometimes identify structural abnormalities in the pons or medulla that can result in hypoventilation. Primary forms of the disease are difficult to diagnose and should be suspected when patients with hypercapnia are found to have normal respiratory muscle strength, normal pulmonary function, and normal alveolar-arterial Po_2 difference. Hypoventilation is more marked during sleep in patients with respiratory drive defects, and polysomnography often reveals elevation in nocturnal CO_2, in addition to central apnea and hypopneas.

In neuromuscular disorders, typically physical examination reveals decreased strength in major muscle groups prior to the development of hypercapnia. Measurement of maximum inspiratory and expiratory pressures or forced vital capacity (FVC) can be used to assess for respiratory muscle involvement in diseases with progressive muscle weakness. Decrease in vital capacity is first seen in the recumbent position, which makes testing for a 15–25% drop in FVC, from seated to supine position in the pulmonary function test laboratory, an invaluable screening tool for impending chronic respiratory failure in these patients. Polysomnography, although unnecessary for diagnosis in most cases, can confirm nocturnal hypercapnia and demonstrate pseudo-central apnea/hypopnea (due to decreased muscle tone, especially in REM sleep) and obstructive apneas, especially with bulbar involvement.

If these evaluations are unrevealing, the clinician should screen for obesity hypoventilation syndrome (OHS). The diagnosis requires the following: body mass index (BMI) ≥30 kg/m^2 and chronic daytime alveolar hypoventilation, defined as $Paco_2$ ≥45 mmHg at sea level in the absence of other known causes of hypercapnia. In almost 90% of cases, obstructive sleep apnea (OSA) is present, with close to 70% exhibiting severe OSA. The population at risk for the development of OHS continues to rise as the worldwide obesity epidemic persists. Although no population-based prevalence studies of OHS have been performed, some estimates suggest it may be as high as 0.4% of the U.S. adult population (or 1 in 263 adults). Some, but not all, studies suggest that severe obesity (BMI >40 kg/m^2) and severe OSA (apnea-hypopnea index [AHI] >30 events per h) are risk factors for the development of OHS. Several screening tools have been developed to identify patients at risk for OSA, including the Epworth Sleepiness Scale (ESS), which measures daytime sleepiness, but it is neither sensitive enough nor specific enough to screen for OSA. The Berlin Questionnaire has been validated in a primary care setting and identifies patients likely to have OSA. Another often used screening questionnaire is the STOP-Bang survey, which has been used in preoperative anesthesia clinics to identify patients at risk of having OSA. Ultimately, polysomnography is needed for the diagnosis of OSA in that patient population and to screen for nocturnal hypoventilation if suspected. These diagnostic recommendations are summarized in Fig. 307-2.

TREATMENT

Hypoventilation

Nocturnal noninvasive positive-pressure ventilation (NIPPV) has been used successfully in the treatment of hypoventilation. It has been shown to improve daytime hypercapnia, prolong survival, and improve health-related quality of life when daytime hypercapnia is documented. There is accumulating evidence to guide timing of initiation of NIPPV and the choice of particular modes in specific disease processes, but more robust research is needed for more concrete recommendations.

NEUROMUSCULAR DISEASE

In patients with neuromuscular disease, the use of NIPPV has been best studied in the ALS patient population. ALS guidelines recommend consideration of nocturnal NIPPV if symptoms of hypoventilation exist and one of the following criteria are present: $Paco_2$ ≥45 mmHg; nocturnal oximetry demonstrates oxygen saturation ≤88% for

5 consecutive min; maximal inspiratory pressure <60 cmH$_2$O; or sniff nasal pressure <40 cmH$_2$O and FVC <50% predicted. However, at present, there is inconclusive evidence to support preemptive nocturnal NIPPV use in all patients with neuromuscular and chest wall disorders who demonstrate nocturnal but not daytime hypercapnia. Once NIPPV is initiated, many of these patients will go from using it only during sleep to needing support while awake, as their disease progresses. The advent of new modern home ventilation devices, with advanced modes of ventilation, and a multitude of comfortable mask interfaces, including mouthpiece ventilation, have all revolutionized the way we manage these patients at home. In addition to improved survival and quality of life, we are able to delay the need for tracheostomy or long-term institutionalization significantly. Smaller portable ventilators, with longer battery life, allow patients to remain functional until late in their disease. Cough assist devices have helped patients cope with ineffective cough and reduced the risk of recurrent pneumonia, a common cause for acute deterioration in these patients.

If NIPPV is initiated, the goal is to gradually correct hypoventilation and bring Paco$_2$ back to a level below 52 mmHg, or at least a 20% reduction from pretreatment levels. Excessive metabolic alkalosis should also be corrected, as serum bicarbonate levels elevated out of proportion for the degree of chronic respiratory acidosis can result in additional hypoventilation. When indicated, administration of supplemental oxygen is effective in attenuating hypoxemia, polycythemia, and pulmonary hypertension. However, in some patients, supplemental oxygen, even at low concentrations, can worsen hypercapnia.

In addition to NIPPV, if available, treatment, should be directed at the underlying disorder. Pharmacologic agents that stimulate respiration, such as medroxyprogesterone and acetazolamide, have been poorly studied in chronic hypoventilation and should not replace treatment of the underlying disease process. Phrenic nerve or diaphragm pacing is a potential therapy for patients with hypoventilation from high cervical spinal cord lesions or respiratory drive disorders. Prior to surgical implantation, patients should have nerve conduction studies to ensure normal bilateral phrenic nerve function. Small case series suggest that effective diaphragmatic pacing can improve quality of life in these patients.

OBESITY HYPOVENTILATION SYNDROME

The pathogenesis of hypoventilation in patients with OHS is multifactorial and includes OSA, increased work of breathing, respiratory muscle impairment relative to the increased load because of excess adiposity, ventilation-perfusion mismatching, and depressed central ventilatory responsiveness to hypoxemia and hypercapnia. Treatment requires addressing all these variables. Defects in central respiratory drive often improve with treatment of sleep-disordered breathing and correction of hypoventilation with continuous positive airway pressure (CPAP) or NIPPV without any significant change in body weight, which suggests that decreased ventilatory responsiveness is a consequence rather than a primary cause of OHS.

The treatment of OSA follows standard guidelines: weight reduction and positive airway pressure therapy during sleep with either CPAP or NIPPV. There is evidence that substantial weight loss (i.e., 20–25% of actual body weight) alone normalizes Paco$_2$ in patients with OHS. Unfortunately, achieving and sustaining this degree of weight loss without bariatric surgery are very challenging for most patients. CPAP improves daytime hypercapnia and hypoxemia in more than half of patients with OHS and concomitant severe OSA. Bilevel positive airway pressure (BiPAP) without a backup rate (BiPAP spontaneous mode) should be reserved for patients who are not able to tolerate high levels of CPAP support or when obstructive respiratory events persist despite reaching the maximum CPAP pressure of 20 cmH$_2$O. NIPPV in the form of BiPAP with a backup rate (BiPAP ST or spontaneous timed) or volume-assured pressure support modes should be strongly considered if hypercapnia persists after several weeks of CPAP therapy with objectively proven adherence. There is now evidence from randomized controlled trials that both CPAP and NIPPV equally improve nocturnal and daytime symptoms, long-term Paco$_2$ and bicarbonate levels, and cardiovascular as well as overall mortality in patients with OHS and severe OSA (AHI >30 events per h). Otherwise, patients with OHS who have no evidence of significant OSA are typically started on BiPAP ST or volume-assured pressure support modes, as are patients presenting with acute decompensated OHS.

CHRONIC HYPERCAPNIC COPD

Chronic hypercapnia in COPD is known to indicate a more advanced disease but has been shown to be associated with worse survival compared to patients with equal disease severity who were normocapnic. Use of NIPPV in patients with COPD has been well established in the acute setting, when patients present with acute-on-chronic hypercapnic respiratory failure. Home NIPPV, on the other hand, has been controversial. It has gained significant momentum over the past decade with the advent and success of high-intensity BiPAP. In patients with severe but stable chronic hypercapnic COPD (Paco$_2$ >52 mmHg and a normal pH), the use of high-intensity BiPAP (inspiratory positive airway pressure 24–28 cmH$_2$O, with a backup rate), aimed to reduce Paco$_2$ to <48 mmHg or a >20% drop from baseline, was associated with improved 1-year mortality in addition to improved physiologic and quality of life parameters, compared with standard of care (home oxygen).

If patients are hospitalized with acute-on-chronic exacerbation, evidence suggests retesting for Paco$_2$ elevation 2–4 weeks after discharge and only considering NIPPV in patients with persistent hypercapnia (Paco$_2$ >52 mmHg) after their exacerbation has resolved. This approach has been shown to reduce hospital readmissions and 1-year mortality.

CENTRAL HYPOVENTILATION SYNDROME

This syndrome can present later in life or in the neonatal period when it is often called Ondine's curse or congenital central hypoventilation syndrome (CCHS). Abnormalities in the gene encoding PHOX2b, a transcription factor with a role in neuronal development, have been implicated in the pathogenesis of CCHS. Regardless of the age of onset, these patients have absent respiratory response to hypoxia or hypercapnia, mildly elevated Paco$_2$ while awake, and markedly elevated Paco$_2$ during non-REM sleep. Interestingly, these patients are able to augment their ventilation and "normalize" Paco$_2$ during exercise and during REM sleep. These patients typically require NIPPV or mechanical ventilation as therapy and should be considered for phrenic nerve or diaphragmatic pacing at centers with experience performing these procedures.

HYPERVENTILATION

CLINICAL FEATURES

Hyperventilation is defined as ventilation in excess of metabolic requirements (CO$_2$ production) leading to a reduction in Paco$_2$. The physiology of patients with chronic hyperventilation is poorly understood, and there is no typical clinical presentation. Symptoms can include dyspnea, paresthesias, tetany, headache, dizziness, visual disturbances, and atypical chest pain. Because symptoms can be so diverse, patients with chronic hyperventilation present to a variety of health care providers, including internists, neurologists, psychologists, psychiatrists, and pulmonologists.

It is helpful to think of hyperventilation as having initiating and sustaining factors. Some investigators believe that an initial event leads to increased alveolar ventilation and a drop in Paco$_2$ to ~20 mmHg. The ensuing onset of chest pain, breathlessness, paresthesia, or altered consciousness can be alarming. The resultant increase in minute volume to relieve these acute symptoms only serves to exacerbate symptoms that are often misattributed by the patient and health care workers to cardiopulmonary disorders. An unrevealing evaluation for causes of these symptoms often results in patients being anxious and fearful of additional attacks. It is important to note that anxiety disorders and panic attacks are *not* synonymous with hyperventilation. Anxiety disorders can be both an initiating and sustaining factor in the pathogenesis of chronic hyperventilation, but these are not necessary for the development of chronic hypocapnia.

DIAGNOSIS

Respiratory symptoms associated with acute hyperventilation can be the initial manifestation of systemic illnesses such as diabetic ketoacidosis. Causes of acute hyperventilation need to be excluded before a diagnosis of chronic hyperventilation is considered. Arterial blood gas sampling that demonstrates a compensated respiratory alkalosis with a near normal pH, low PaCO_2, and low calculated bicarbonate is necessary to confirm chronic hyperventilation. Other causes of respiratory alkalosis, such as mild asthma, need to be diagnosed and treated before chronic hyperventilation can be considered. A high index of suspicion is required as increased minute ventilation can be difficult to detect on physical examination. Once chronic hyperventilation is established, a sustained 10% increase in alveolar ventilation is sufficient to perpetuate hypocapnia. This increase can be accomplished with subtle changes in the respiratory pattern, such as occasional sigh breaths or yawning 2–3 times per min.

TREATMENT

There are few well-controlled treatment studies of chronic hyperventilation owing to its diverse features and the lack of a universally accepted diagnostic process. Clinicians often spend considerable time identifying initiating factors, excluding alternative diagnoses, and discussing the patient's concerns and fears. In some patients, reassurance and frank discussion about hyperventilation can be liberating. Identifying and eliminating habits that perpetuate hypocapnia, such as frequent yawning or sigh breathing, can be helpful. Some evidence suggests that breathing exercises and diaphragmatic retraining may be beneficial for some patients. The evidence for using medications to treat hyperventilation is scant. Beta blockers may be helpful in patients with sympathetically mediated symptoms such as palpitations and tremors.

ACKNOWLEDGMENT

John F. McConville, Julian Solway, and Babak Mokhlesi contributed to this chapter in the 21st edition and some material from that chapter has been retained here.

FURTHER READING

Anderson PM et al: EFNS guidelines on the clinical management of amyotrophic lateral sclerosis (MALS)–revised report of the EFNS task force. Eur J Neurol 19:360, 2012.

Benditt JO: Pathophysiology of neuromuscular respiratory diseases. Clin Chest Med 39:297, 2018.

Chung F et al: STOP-Bang questionnaire: A practical approach to screen for obstructive sleep apnea. Chest 149:631, 2016.

Douglas IS: Acute-on-chronic respiratory failure, in *Principles of Critical Care*, 4th ed. Hall JB et al (eds). New York, McGraw-Hill, 2015, pp 482–495.

Gardner WN: The pathophysiology of hyperventilation disorders. Chest 109:516, 1996.

Macrea M et al: Long-term noninvasive ventilation in chronic stable hypercapnic chronic obstructive pulmonary disease. An official American Thoracic Society clinical practice guideline. Am J Respir Crit Care Med 202:e74, 2020.

Masa JF et al: Long-term clinical effectiveness of continuous positive airway pressure therapy versus non-invasive ventilation therapy in patients with obesity hypoventilation syndrome: A multicentre, open-label, randomised controlled trial. Lancet 393:1721, 2019.

Masa JF et al: Obesity hypoventilation syndrome. Eur Respir Rev 28:180097, 2019.

Mokhlesi B et al: Evaluation and management of obesity hypoventilation syndrome. An official American Thoracic Society clinical practice guideline. Am J Respir Crit Care Med 200:e6, 2019.

Rimmer KP et al: Home mechanical ventilation for patients with amyotrophic lateral sclerosis: A Canadian Thoracic Society clinical practice guideline. Canadian J Respir Crit Care Sleep Med 3:9, 2019.

308

Sleep Apnea

Andrew Wellman, Daniel J. Gottlieb,
Susan Redline

Obstructive sleep apnea (OSA) and central sleep apnea (CSA) are both classified as sleep-related breathing disorders. OSA and CSA share some risk factors and physiologic bases but also have unique features. Each disorder is associated with impaired ventilation during sleep and disruption of sleep, and each diagnosis requires careful elicitation of the patient's history, physical examination, and physiologic testing. OSA, the more common disorder, causes daytime sleepiness and impaired daily function. It is a cause of hypertension and is strongly associated with cardiovascular disease in adults and behavioral problems in children. CSA is less common and may occur alone or in combination with OSA. It can occur as a primary condition, as a response to high altitude, or secondary to a medical condition (such as heart failure) or medication (such as opioids). Patients with CSA often report frequent awakenings and daytime fatigue and are at increased risk for heart failure and atrial fibrillation.

OBSTRUCTIVE SLEEP APNEA/HYPOPNEA SYNDROME

Definition OSA is defined on the basis of nocturnal and daytime symptoms as well as sleep study findings. Diagnosis requires the patient to have (1) either symptoms of nocturnal breathing disturbances (snoring, snorting, gasping, or breathing pauses during sleep) that impair quality of life or daytime sleepiness or fatigue that occurs despite sufficient opportunity to sleep and is unexplained by other medical problems; and (2) five or more episodes of obstructive apnea or hypopnea per hour of sleep (the apnea-hypopnea index [AHI], calculated as the number of episodes divided by the number of hours of sleep) documented during a sleep study. OSA also may be diagnosed in the absence of symptoms if the AHI is ≥15 episodes/h. Each episode of apnea or hypopnea represents a reduction in breathing for at least 10 s and commonly results in a ≥3% drop in oxygen saturation or a brain cortical arousal. OSA severity can be characterized by the frequency of breathing disturbances (AHI), the amount of oxyhemoglobin desaturation with respiratory events, the duration of apneas and hypopneas, the degree of sleep fragmentation, and the level of reported daytime sleepiness or functional impairment.

Pathophysiology During inspiration, intraluminal pharyngeal pressure becomes increasingly negative, creating a "suctioning" force. Because the pharyngeal airway has no fixed bone or cartilage, airway patency is dependent on the stabilizing influence of the pharyngeal dilator muscles. Although these muscles are continuously activated during wakefulness, neuromuscular output declines with sleep onset. In patients with a collapsible airway, the reduction in neuromuscular output results in transient episodes of pharyngeal collapse (manifesting as an "apnea") or near collapse (manifesting as a "hypopnea"). The episodes of collapse are typically terminated when ventilatory reflexes are activated and cause arousal, thus stimulating an increase in neuromuscular activity and opening of the airway. The airway may collapse at different sites, such as the soft palate (most common), tongue base, lateral pharyngeal walls, and/or epiglottis **(Fig. 308-1)**. OSA may be most severe during rapid eye movement (REM) sleep, when neuromuscular output to the skeletal muscles is particularly low, and in the supine position due to gravitational forces.

Individuals with a small pharyngeal lumen require relatively high levels of neuromuscular activation to maintain patency during wakefulness and thus are predisposed to airway collapse following the normal sleep-related reduction in pharyngeal muscle activity during sleep. The airway lumen may be narrowed by enlargement of soft tissue structures (tongue, palate, and uvula) due to fat deposition, increased lymphoid tissue, or genetic variation. Craniofacial factors such as

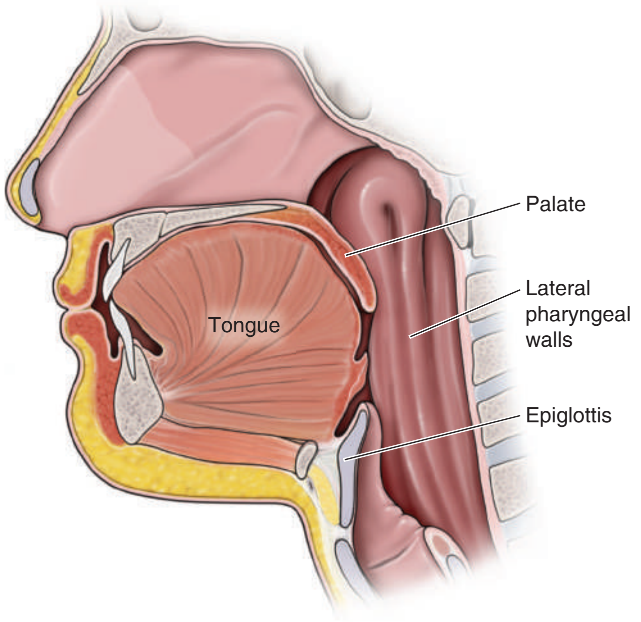

FIGURE 308-1 The structures causing airway collapse in obstructive sleep apnea include the palate, the tongue, and/or the epiglottis. In addition, collapse can also occur due to the lateral pharyngeal walls.

mandibular retroposition or micrognathia, reflecting genetic variation or developmental influences, also can reduce lumen dimensions. In addition, lung volumes influence the caudal traction on the pharynx and consequently the stiffness of the pharyngeal wall. Accordingly, low lung volume in the recumbent position, which is particularly pronounced in the obese, contributes to collapse (less caudal traction). A high degree of nasal resistance (e.g., due to nasal septal deviation or polyps) can contribute to airway collapse by reducing intraluminal pressure downstream in the pharynx. High-level nasal resistance also may trigger mouth opening during sleep, which breaks the seal between the tongue and the palate and allows the tongue to fall posteriorly and occlude the airway.

Pharyngeal muscle activation is integrally linked to ventilatory drive. Thus, factors related to ventilatory control, particularly ventilatory sensitivity, arousal threshold, and neuromuscular responses to carbon dioxide (CO_2), contribute to the pathogenesis of OSA. A buildup in CO_2 during sleep activates both the diaphragm and the pharyngeal muscles. Pharyngeal activation stiffens the upper airway and can counteract inspiratory suction pressure and maintain airway patency to an extent that depends on the anatomic predisposition to collapse. However, pharyngeal collapse can occur when the ventilatory control system is overly sensitive to CO_2, with resultant wide fluctuations in ventilation, ventilatory drive, and upper airway stiffness. Moreover, increasing levels of CO_2 during sleep result in central nervous system arousal, causing the individual to move from a deeper to a lighter level of sleep or to awaken. A low arousal threshold (i.e., awakening to a low level of CO_2 or ventilatory drive) can preempt the CO_2-mediated process of pharyngeal muscle compensation and prevent airway stabilization. A high arousal threshold, conversely, may prevent appropriate termination of apneas, prolonging apnea duration and exacerbating oxyhemoglobin desaturation. Finally, any impairment in the ability of the muscles to compensate during sleep can contribute to collapse of the pharynx. The relative contributions of risk factors vary by age, sex, body mass index, and other factors. Approaches to the measurement of these factors in clinical settings, with consequent enhancement of "personalized" therapeutic interventions, are being actively investigated.

Risk Factors and Prevalence The major risk factors for OSA are obesity, male sex, and older age. Additional risk factors include mandibular retrognathia and micrognathia, a positive family history of OSA, sedentary lifestyle, genetic syndromes that reduce upper airway patency (e.g., Down syndrome, Treacher-Collins syndrome), adenotonsillar hypertrophy (especially in children), menopause (in women), and various endocrine syndromes (e.g., acromegaly, hypothyroidism).

Approximately 40–60% of cases of OSA are attributable to excess weight. Obesity predisposes to OSA through the narrowing effects of upper airway fat on the pharyngeal lumen. Obesity also reduces chest wall compliance and decreases lung volumes, resulting in a loss of caudal traction on upper airway structures. Obese individuals are at a fourfold or greater risk for OSA than their normal-weight counterparts, although the association of OSA with obesity is weaker in the elderly than in younger adults. A 10% weight gain is associated with a >30% increase in AHI. Even modest weight loss or weight gain can influence the risk and severity of OSA. However, the absence of obesity does not exclude this diagnosis.

The prevalence of OSA is twofold higher among men than among women. Factors that predispose men to OSA include android pattern of obesity (resulting in upper-airway and abdominal fat deposition) and relatively greater pharyngeal length, which increases collapsibility. Premenopausal women are relatively protected from OSA by the influence of sex hormones on ventilatory drive. The decline in sex difference in older age reflects an increased OSA prevalence in women after menopause. The pathogenesis and presentation of OSA also differ in men and women: compared to men, women have a lower arousal threshold and less neuromuscular collapsibility. Women tend to have shorter duration of apneas and apneas that occur predominantly in REM sleep. Failure to recognize these differences can contribute to underrecognition of OSA in women.

Variations in craniofacial morphology that reduce the size of the posterior airway space increase OSA risk. The contribution of skeletal structural features to OSA is most evident in nonobese patients. Identification of features such as retrognathia can influence therapeutic decision-making.

OSA has a strong genetic basis, as evidenced by its significant familial aggregation and heritability. For a first-degree relative of a patient with OSA, the odds of having OSA are approximately twofold higher than that of someone without an affected relative. Several genetic variants have been associated with prevalence of OSA or with related traits, such as the frequency of apneas and hypopneas, the duration of respiratory events, and degree of overnight levels of hypoxemia.

OSA prevalence varies with age, from 5 to 15% among middle-aged adults to >20% among elderly individuals, although in a majority of affected adults, the disorder is undiagnosed. There is a peak due to lymphoid hypertrophy among children between the ages of 3 and 8 years; with airway growth and lymphoid tissue regression during later childhood, prevalence declines. Then, as obesity prevalence increases in adolescence and adulthood, OSA prevalence again increases.

The prevalence of OSA is especially high among patients with certain medical conditions, including diabetes mellitus, hypertension, and atrial fibrillation. Individuals of East Asian ancestry appear to be at increased risk of OSA at relatively low levels of body mass index, reflecting the greater influence of craniofacial risk factors. In the United States, African Americans, especially children and young adults, are at higher risk for OSA than their white counterparts.

Course of the Disorder The precise onset of OSA is usually hard to identify. A person may snore for many years, often beginning in childhood, before OSA is identified. Weight gain may precipitate an increase in symptoms, which in turn may lead the patient to pursue an evaluation. OSA may become less severe with weight loss, particularly after bariatric surgery. In adults, there is a gradual increase in AHI with age, although marked increases and decreases in the AHI are uncommon unless accompanied by weight change.

APPROACH TO THE PATIENT

Obstructive Sleep Apnea/Hypopnea Syndrome

An evaluation for OSA should be considered in patients with symptoms of OSA and one or more risk factors. Screening also should be considered in patients who report symptoms consistent with OSA and who are at high risk for OSA-related morbidities, such as hypertension, diabetes mellitus, and cardiac and cerebrovascular diseases.

SYMPTOMS AND HISTORY

When possible, a sleep history should be obtained with assistance from a bed partner or household member. Snoring is the most common symptom; however, its absence does not exclude the diagnosis, as pharyngeal collapse may occur without tissue vibration. Gasping or snorting during sleep may also be reported, reflecting termination of individual apneas with abrupt airway opening. Dyspnea is unusual, and its absence generally distinguishes OSA from paroxysmal nocturnal dyspnea, nocturnal asthma, and acid reflux with laryngospasm. Patients also may describe frequent awakening or sleep disruption, which is more common among women and older adults. The most common daytime symptom is excessive daytime sleepiness, identified by a history of difficulty maintaining alertness or involuntary periods of dozing. However, many women preferentially report fatigue rather than sleepiness. Other symptoms include a dry mouth, nocturnal heartburn, diaphoresis of the chest and neck, nocturia, morning headaches, trouble concentrating, irritability, and mood disturbances. Insomnia, which is common in the general population, may coexist with OSA. Although difficulty falling sleep is rarely caused by OSA, awakening at apnea termination may cause difficulty maintaining sleep, a symptom more likely to be reported by women than by men, and often responds to treatment of OSA. Several questionnaires that evaluate snoring frequency, self-reported apneas, and daytime sleepiness can facilitate OSA screening. The predictive ability of a questionnaire can be enhanced by a consideration of whether the patient is male or has risk factors such as obesity or hypertension.

PHYSICAL FINDINGS

Physical findings often reflect the etiologic factors for the disorder as well as comorbid conditions, particularly vascular disease. On examination, patients may exhibit hypertension and regional (central) obesity, as indicated by a large waist and neck circumference. The oropharynx may reveal a small orifice with crowding due to an enlarged tongue, a low-lying soft palate with a bulky uvula, large tonsils, a high-arched palate, or micro-/retrognathia. Since nasal resistance can increase the propensity to pharyngeal collapse, the nasal cavity should be inspected for polyps, septal deviation, allergic rhinitis, and other signs of obstruction. Because patients with heart failure are at increased risk for both OSA and CSA, a careful cardiac examination should be conducted to detect possible left- or right-sided cardiac dysfunction. Evidence of cor pulmonale suggests a comorbid cardiopulmonary condition; OSA alone is not thought to cause right-heart failure. A neurologic evaluation is needed to evaluate for conditions such as neuromuscular and cerebrovascular diseases, which increase OSA risk.

LABORATORY FINDINGS

Diagnostic Findings Since symptoms and signs do not accurately predict the severity of sleep-related breathing disturbances, specific diagnosis and categorization of OSA severity require objective measurement of breathing during sleep. The gold standard for diagnosis of OSA is an overnight polysomnogram (PSG). A negative in-laboratory PSG usually rules out OSA. However, false-negative studies can result from night-to-night variation in OSA severity, particularly if there was insufficient REM sleep or less supine sleep during testing than is typical for the patient. Home sleep tests that record only respiratory and cardiac channels are commonly used as a cost-effective means for diagnosing OSA. However, a home study may yield a false-negative result if sleep time is not accurately estimated or in individuals experiencing hypopneas with arousals rather than oxyhemoglobin desaturation. Therefore, if there is a high prior probability of OSA, a negative home study should be followed by PSG.

The key physiologic information collected during a sleep study for OSA assessment includes measurement of breathing (changes in airflow, respiratory excursion), oxygenation (hemoglobin oxygen saturation), body position, and cardiac rhythm. In addition, PSGs and some home sleep studies measure sleep continuity and sleep stages (by electroencephalography, chin electromyography, electro-oculography, and actigraphy), leg movements, and snoring intensity. This information is used to quantify the frequency and subtypes of abnormal respiratory events during sleep as well as associated changes in oxygen hemoglobin saturation, arousals, and sleep stage distributions. **Tables 308-1** and **308-2** define the respiratory events scored and the severity guidelines employed during a sleep study. **Fig. 308-2** shows examples of sleep-related respiratory events. A typical sleep study report provides quantitative data such as the AHI (number of apneas plus hypopneas per hour of sleep) and the profile of oxygen saturation over the night (mean, nadir, time at low levels). Reports may also include the respiratory disturbance index, which includes the number of respiratory effort–related arousals in addition to the AHI. In-laboratory PSG also quantifies sleep latency (time from "lights off" to first sleep onset), the frequency of periodic limb movements during sleep, sleep efficiency (percentage of time asleep relative to time in bed), arousal index (number of cortical arousals per hour of sleep), and time in each sleep stage. These metrics can further characterize the severity of OSA, which is associated with an increased arousal index, low sleep efficiency, and a reduction of time in deep (stage N3) and REM sleep and increase in light (stage N1) sleep. The detection of autonomic responses to apneas and hypopneas, such as surges in blood pressure, changes in heart rate, and abnormalities in cardiac rhythm, also provides relevant information on OSA severity.

While the AHI is the chief disease-defining measurement derived from sleep studies, metrics that quantify respiratory event–related hypoxemia, heart rate response, and ventilatory reduction have been shown to predict adverse cardiovascular outcomes and mortality and may soon be incorporated into clinical decision-making.

Other Laboratory Findings Various imaging studies, including cephalometric radiography, upper airway magnetic resonance imaging (MRI) and computed tomography (CT), and fiberoptic endoscopy, can be used to identify anatomic risk factors for OSA. While these may be useful for planning surgical interventions, they are not indicated in the routine evaluation of OSA. Cardiac testing may yield evidence of impaired systolic or diastolic ventricular function or abnormal cardiac structure. Overnight blood pressure monitoring often displays a "nondipping" pattern (absence of the typical 10% fall of blood pressure during sleep compared to wakefulness). Arterial blood gas measurements made during wakefulness are usually normal. Waking hypoxemia or hypercarbia suggests coexisting cardiopulmonary disease or hypoventilation syndromes. Patients with severe nocturnal hypoxemia may have elevated hemoglobin values. A multiple sleep latency test or a maintenance of wakefulness test can be useful in quantifying sleepiness and helping to distinguish OSA from narcolepsy.

TABLE 308-1 Respiratory Event Definitions

- *Apnea:* Cessation of airflow for ≥10 s during sleep, accompanied by:
- Persistent respiratory effort (obstructive apneas, **Fig. 308-2*A***), or
- Absence of respiratory effort (central apneas, **Fig. 308-2*B***)
- *Hypopnea:* A ≥30% reduction in airflow for at least 10 s during sleep that is accompanied by either a ≥3% desaturation or an arousal **(Fig. 308-2*C*)**
- *Respiratory effort–related arousal (RERA):* Partial obstruction that does not meet the criteria for hypopnea but provides evidence of increasing inspiratory effort (usually through pleural pressure monitoring) punctuated by an arousal (**Fig. 308-2*D***)
- *Flow-limited breath:* A partially obstructed breath, typically within a hypopnea or RERA, identified by a flattened or "scooped-out" inspiratory flow shape **(Fig. 308-3)**

TABLE 308-2 Obstructive Sleep Apnea/Hypopnea Syndrome (OSAHS): Quantification and Severity Scale

- *Apnea-hypopnea index (AHI)*[a]: Number of apneas plus hypopneas per hour of sleep
- *Respiratory disturbance index (RDI):* Number of apneas plus hypopneas plus RERAs per hour of sleep
- *Mild OSAHS:* AHI of 5–14 events/h
- *Moderate OSAHS:* AHI of 15–29 events/h
- *Severe OSAHS:* AHI of ≥30 events/h

[a]Each level of AHI can be further quantified by level of sleepiness and associated hypoxemia.

Abbreviation: RERAs, respiratory effort–related arousals.

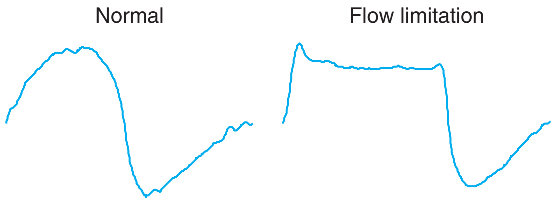

FIGURE 308-3 Example of flow limitation. The inspiratory flow pattern in a patent airway is rounded and peaks in the middle. In contrast, a partially obstructed airway exhibits an early peak followed by mid-inspiratory flattening, yielding a scooped-out appearance.

Health Consequences and Comorbidities OSA is the most common medical cause of daytime sleepiness and negatively influences quality of life. It is also strongly associated with cardiac, cerebrovascular, and metabolic disorders and with premature death and increased risk for certain cancers. This broad range of health effects is attributable to the impact of sleep fragmentation, cortical arousal, and intermittent hypoxemia and hypercapnia on vascular, cardiac, metabolic, and neurohumoral functions. OSA-related respiratory events stimulate sympathetic overactivity, leading to acute blood pressure surges during sleep and nocturnal as well as daytime hypertension. OSA-related hypoxemia also stimulates release of acute-phase proteins and reactive oxygen species that exacerbate insulin resistance and lipolysis and cause an augmented prothrombotic and proinflammatory state. Inspiratory effort against an occluded airway causes large intrathoracic negative pressure swings, altering cardiac preload and afterload and resulting in cardiac remodeling and reduced cardiac function. Hypoxemia and sympathetic-parasympathetic imbalance also may cause electrical remodeling of the heart and myocyte injury.

HYPERTENSION OSA can raise blood pressure to prehypertensive and hypertensive ranges, increase the prevalence of a nondipping overnight blood pressure pattern, and increase the risk of uncontrolled and resistant hypertension. Elevations in blood pressure are due to augmented sympathetic nervous system activation as well as alterations in the renin-angiotensin-aldosterone system and fluid balance. Treatment of OSA with nocturnal continuous positive airway pressure (CPAP) has been shown to reduce 24-h ambulatory blood pressure. Although the overall impact of CPAP on blood pressure levels is relatively modest (averaging 2–4 mmHg), larger improvements are observed among

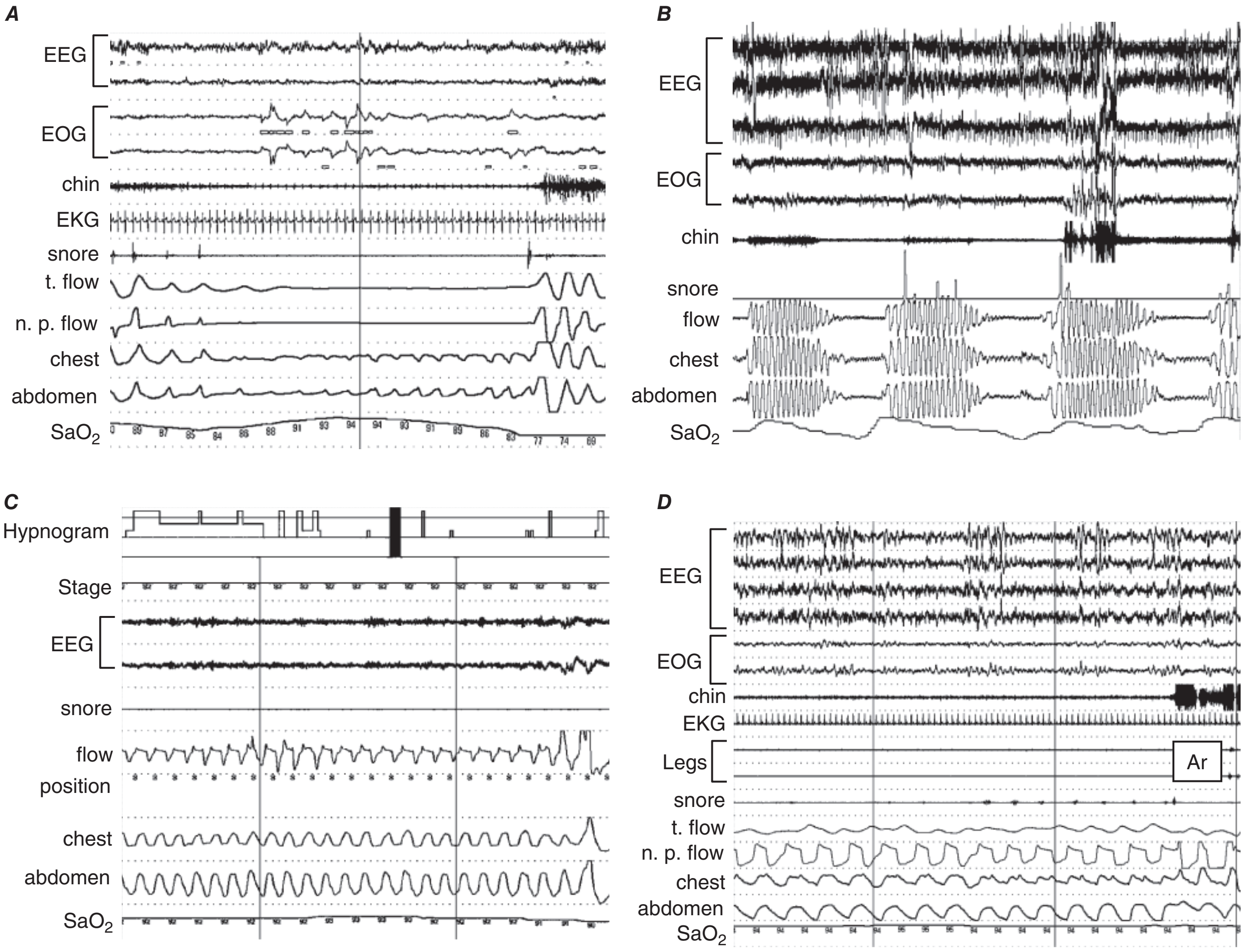

FIGURE 308-2 Obstructive apnea. ***A.*** There are 30 s of no airflow, as shown in the nasal pressure (n. p. flow) and thermistor-measured flow (t. flow). Note the presence of chest-abdomen paradox, indicating respiratory effort against an occluded airway. ***B.*** Central apnea in a patient with Cheyne-Stokes respiration due to congestive heart failure. The flat chest-abdomen tracings indicate the absence of inspiratory effort during the central apneas. ***C.*** Hypopnea. Partial obstruction of the pharyngeal airway can limit ventilation, leading to desaturation (a mild decrease in this patient, from 93 to 90%) and arousal. ***D.*** Respiratory effort–related arousal (RERA). Minimal flow reduction terminated by an arousal (Ar) without desaturation constitutes an RERA. EEG, electroencephalogram; EKG, electrocardiogram; EOG, electro-oculogram.

patients who have a high AHI, report daytime sleepiness, or have resistant hypertension.

CARDIOVASCULAR, CEREBROVASCULAR, AND METABOLIC DISEASES Among the most serious health consequences of OSA may be its impact on cardiac and metabolic functions. Strong epidemiologic evidence indicates that OSA significantly increases the risk of coronary artery disease, heart failure with and without reduced ejection fraction, atrial and ventricular arrhythmias, atherosclerosis and coronary artery disease, stroke, and diabetes. Treatment of OSA has been shown to reduce several markers of cardiovascular risk and improve insulin resistance and, in uncontrolled studies, is associated with a decreased recurrence rate of atrial fibrillation. Large randomized clinical trials, however, have failed to demonstrate that OSA treatment with CPAP reduces cardiovascular and stroke event rates or prolongs survival. These outcomes may reflect exclusion from these trials of patients with excessive sleepiness, as there is evidence that sleepy patients may have the greatest OSA-related cardiovascular risk. Limited adherence to treatment among trial participants or the widespread use of other effective secondary prevention measures, such as beta blockade, antiplatelet agents, and lipid-lowering therapy, may also limit the impact of CPAP on cardiovascular risk.

SLEEPINESS AND ASSOCIATED RISKS More than 50% of patients with moderate to severe OSA report daytime sleepiness. However, there is not a strong association between AHI level and degree of sleepiness. Patients with OSA symptoms have a twofold increased risk of occupational accidents. Individuals with elevated AHIs are involved in motor vehicle crashes approximately two to three times as often as persons with normal AHIs. Randomized controlled trials have shown that treatment of OSA with CPAP alleviates sleepiness as measured by either questionnaire or objective testing in patients with both mild and more severe disease. However, the degree of improvement varies widely. Residual sleepiness may be due to several factors, including suboptimal treatment adherence, insufficient sleep duration, other sleep disorders, or prior hypoxia-mediated damage in brain areas involved in alertness. Moreover, visceral adipose tissue, which is present in higher amounts in patients with OSA, releases somnogenic cytokines that may contribute to sleepiness. Thus, even after treatment, it is important to assess and monitor patients for residual sleepiness and to optimize treatment adherence, improve sleep patterns, and identify other disorders that may contribute to sleepiness. Careful and supervised use of alerting agents may be appropriate as adjunctive treatment in patients in whom sleepiness does not respond to CPAP alone.

QUALITY OF LIFE AND MOOD Reductions in health-related quality of life are common in patients with OSA, with the largest decrements observed in scales that measure physical functioning and energy levels. Work-related productivity also has been shown to improve in patients with moderate to severe OSA treated with CPAP. Numerous studies, including a large-scale trial of minimally symptomatic patients, have shown that treatment with CPAP can improve these patient-reported outcomes. Depressive symptoms, in particular somatic symptoms (irritability, fatigue, lack of energy), are commonly reported in OSA and improve with CPAP.

TREATMENT

Obstructive Sleep Apnea

A comprehensive approach to the management of OSA is needed to reduce risk factors and comorbidities. The clinician should seek to identify and address lifestyle and behavioral factors as well as comorbidities that may be exacerbating OSA. As appropriate, treatment should aim to reduce weight; optimize sleep duration (7–9 h per night); regulate sleep schedules (with similar bedtimes and wake times across the week); encourage the patient to avoid sleeping in the supine position; treat nasal allergies; increase physical activity; eliminate alcohol ingestion (which impairs pharyngeal muscle activity) within 3 h of bedtime; and minimize use of opiate medications. Sedative-hypnotic medications have inconsistent effects on OSA but should be avoided in most patients with moderate to severe OSA. Patients should be counseled to avoid drowsy driving.

CPAP is the standard medical therapy with the highest level of evidence for efficacy. Delivered through a nasal or nasal-oral mask, CPAP works as a mechanical splint to hold the airway open, thus maintaining airway patency during sleep. An overnight CPAP titration study can determine the optimal pressure setting that reduces the number of apneas/hypopneas during sleep, improves gas exchange, and reduces arousals; however, the use of "auto-titrating" CPAP (APAP) devices used in home settings has eliminated the need for titration sleep studies in many patients. Rates of adherence to CPAP treatment are highly variable (average, 50–80%) and may be improved with support by a skilled health care team who can address side effects, help the patient "problem solve," and provide motivational education (Table 308-3). Online CPAP support tools can provide the patient personalized support and feedback. Despite the limitations of CPAP, controlled studies have demonstrated its beneficial effect on alertness, mood, quality of life, work-related productivity, blood pressure, and insulin sensitivity. Uncontrolled studies also indicate a favorable effect on cardiovascular outcomes, cardiac ejection fraction, atrial fibrillation recurrence, and mortality risk.

Oral appliances for OSA work by advancing the mandible, thus opening the airway by repositioning the lower jaw and pulling the tongue forward. These devices generally work better when customized for patient use; maximal adaptation can take several weeks. Efficacy studies show that these devices can reduce the AHI by ≥50% in two-thirds of individuals, although these data are based largely on patients with mild OSA. Some patients with moderate or severe OSA respond to oral appliances as well, although no consistent predictors of success have been identified in these groups, and thus, follow-up sleep testing is recommended. Side effects of oral appliances include temporomandibular joint pain and tooth movement; thus, they require that the patient have adequate dental and periodontal structures. Oral appliances are most often used for treating patients with mild/moderate OSA or patients who do not tolerate CPAP. However, as some patients are more adherent to oral appliances than to CPAP, these devices are under investigation for treatment of more severe disease. In some patients, more consistent treatment of OSA may be obtained by alternating between CPAP and oral appliance therapy.

Upper airway surgery for OSA is less efficacious than CPAP and is mostly reserved for the treatment of patients who snore, have mild OSA, or cannot tolerate CPAP. Uvulopalatopharyngoplasty (UPPP, removal of the uvula and the margin of the soft palate) is the most commonly performed surgery for OSA. However, results vary greatly, and as a standalone procedure, UPPP often has limited efficacy, particularly in severe OSA and in obese patients. Thus, palatal surgery is often combined with other procedures ("multilevel surgery" involving more than one pharyngeal site/structure) performed by an experienced surgeon, but the selection of patients is an important factor and relies on careful targeting of culprit areas for surgical resection. Bariatric surgery is an option

TABLE 308-3 Side Effects of Continuous Positive Airway Pressure (CPAP) and Their Treatments

SIDE EFFECT	TREATMENT
Nasal congestion	Provide heated humidification, administer saline/steroid nasal sprays
Claustrophobia	Change mask interface (e.g., to nasal prongs), promote habituation (i.e., practice breathing on CPAP while awake)
Difficulty exhaling	Temporarily reduce pressure, provide bilevel positive airway pressure
Bruised nasal ridge	Change mask interface, provide protective padding
Aerophagia	Administer antacids

for obese patients with OSA and can improve not only OSA but also other obesity-associated health conditions. Other procedures that can decrease snoring but have minimal effects on OSA include injection of a hardening agent to the soft palate (resulting in stiffening), radiofrequency ablation, laser-assisted uvulopalatoplasty, and palatal implants.

Upper airway neurostimulation is a recently tested alternative treatment for OSA. Unilateral stimulation of the hypoglossal nerve through a surgically implanted device was shown to significantly decrease the AHI and improve a number of patient-reported outcomes, such as sleepiness and quality of life, for a duration of at least 5 years after treatment in carefully selected patients. This therapy is reserved for patients who cannot tolerate or fail CPAP therapy. Current inclusion criteria are moderate to severe OSA (AHI 15–65), body mass index <35 kg/m^2, and absence of complete concentric collapse at the level of the velum documented by awake and drug-induced endoscopy (a predictor of response to surgery). This therapy is also approved for use in children ages 13 or older and adults with Down's syndrome who have an AHI of 10–50. Additional research is underway to further evaluate longer-term effectiveness and potential utility of this treatment in other patient groups.

Vibrotactile positional therapy (devices that emit incremental vibratory stimuli in response to body position in the supine position) can reduce AHI by as much as 50% in individuals with OSA occurring predominantly in the supine position.

Supplemental oxygen can improve oxygen saturation, but there is little evidence that it improves OSA symptoms or the AHI in unselected patients. There is conflicting evidence regarding the effect of supplemental oxygen on blood pressure in patients with OSA.

Orofacial myofunctional therapy (OMT), which involves exercises designed to enhance the strength and coordination of orofacial muscles, has been tested in several small studies of OSA. These studies predominantly focused on individuals suffering from mild-to-moderate OSA and showed modest improvements in snoring and AHI, as well as in overall sleep quality. However, larger and more comprehensive studies are necessary to evaluate fully the efficacy of myofunctional therapy in treating OSA.

Currently, there are no U.S. Food and Drug Administration (FDA)-approved medications for the treatment of OSA. However, ongoing research is exploring the efficacy of weight loss medications and drugs that enhance pharyngeal muscle activity during sleep. Among weight loss drugs, anorectics are generally not favored for OSA due to their side effects and potential for abuse. However, glucagon-like peptide 1 (GLP-1) receptor agonists have garnered recent interest because of their substantial impact on weight loss. Given the strong correlation between weight and OSA, drugs that facilitate weight reduction, such as GLP-1 agonists, which slow gastric emptying and suppress appetite, are being investigated for their potential in improving OSA symptoms. Liraglutide, a GLP-1 agonist, has shown small effects on OSA severity, but tirzepatide, which targets both GLP-1 and glucose-dependent insulinotropic polypeptide (GIP) receptors, may offer more significant weight loss and is currently under trial for OSA. Sodium-glucose cotransporter 2 (SGLT-2) inhibitors also promote weight loss and are under investigation for the treatment of OSA.

Concerning drugs that stimulate pharyngeal muscle activity, studies have shown that sleep-related withdrawal of norepinephrine from cranial motor neurons leads to pharyngeal hypotonia and consequent upper airway obstruction in patients with OSA. Pharyngeal hypotonia is further exacerbated by muscarinic inhibition during REM sleep. Based on these mechanisms, drugs with noradrenergic and antimuscarinic properties have been tested for their potential in OSA management. Early studies have reported notable reductions in AHI, and more advanced phase 3 trials of these drug combinations are in progress.

CENTRAL SLEEP APNEA

CSA, which is less common than OSA, may occur in isolation or, more often, in combination with obstructive events in the form of "mixed" apneas. CSA is often caused by an increased sensitivity to P_{CO_2}, which leads to an unstable breathing pattern that manifests as hyperventilation alternating with apnea. A prolonged circulation delay between the pulmonary capillaries and carotid chemoreceptors is also a contributing cause; thus, individuals with congestive heart failure are at risk for CSA. With prolonged circulation delay, there is a crescendo-decrescendo breathing pattern known as *Cheyne-Stokes breathing* (Fig. 308-2*B*). Other risk factors for CSA include opioid medications (which appear to have a dose-dependent effect on CSA) and hypoxia (e.g., breathing at high altitude). In some individuals, CPAP—particularly at high pressures—seems to induce central apnea; this condition is referred to as *complex sleep apnea* or *treatment-emergent central sleep apnea*. Rarely, CSA may be caused by blunted chemosensitivity due to congenital disorders (congenital central hypoventilation syndrome) or acquired factors. CSA is associated with increased risk for the development of both heart failure and atrial fibrillation. This is possibly related to elevations in sympathetic nervous system activity that accompany this disorder; alternatively, CSA may be an early marker of subclinical myocardial dysfunction. Patients with CSA may report symptoms of frequent awakenings as well as daytime fatigue. Treatment of CSA is difficult and depends on the underlying cause. As treatment of CSA has not been shown to improve long-term health outcomes, specific treatment is generally indicated only for treatment of symptomatic CSA.

Limited data suggest that supplemental oxygen can reduce the frequency of central apneas, particularly in patients with hypoxemia; however, the effectiveness of supplemental oxygen on clinical outcomes is unknown.

In patients with CSA or Cheyne-Stokes breathing associated with heart failure, treatment is directed at optimizing therapy for heart failure. Device-based therapies that adjust pressure support on a breath-by-breath basis (adaptive servoventilation [ASV]) can be effective for regularizing breathing; however, two large international trials showed that ASV resulted in no cardiovascular or mortality benefit. The first trial unexpectedly reported increased mortality, leading to an FDA warning to avoid ASV in patients with CSA who have a left ventricular ejection fraction <45%. The second trial tested an ASV device that delivered lower pressure and reported no evidence of harm along with small improvements in symptoms, raising the possibility that ASV may have a role as adjunctive therapy for patients with heart failure and CSA.

Moderate to severe CSA in adults can be treated with an FDA-approved transvenous phrenic nerve stimulator. A clinical trial testing this therapy in patients with CSA from various etiologies showed an ~50% reduction in AHI and a near elimination of central apneas, along with improvements in sleep quality and quality of life. Ongoing research is addressing long-term safety and effectiveness, as well as subgroups who may benefit the most from this intervention.

FURTHER READING

Berry R, Wagner M: *Sleep Medicine Pearls*, 3rd ed. Philadelphia, Elsevier, 2015.

Gottesman RF et al: Impact of sleep disorders and disturbed sleep on brain health: A scientific statement from the American Heart Association. Stroke 55:e61, 2024.

Gottlieb DJ, Punjabi NM: Diagnosis and management of obstructive sleep apnea. A review. JAMA 323:1389, 2020.

Javaheri S et al: Sleep apnea: Types, mechanisms, and clinical cardiovascular consequences. J Am Coll Cardiol 769:841, 2017.

Kapur VK et al: Clinical Practice Guideline for Diagnostic Testing for Adult Obstructive Sleep Apnea: An American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med 13:479, 2017.

Marklund M et al: Update on oral appliance therapy. Eur Respir Rev 28:190083, 2019.

Redline S et al: Obstructive sleep apnoea heterogeneity and cardiovascular disease. Nat Rev Cardiol 20:560, 2023.

309 Lung Transplantation

Hilary J. Goldberg, Hari R. Mallidi

END-STAGE LUNG DISEASE AND INDICATIONS FOR LUNG TRANSPLANTATION

The Lung Allocation Score (LAS) was implemented in 2005 for the purpose of prioritizing organ allocation. In this model, the United Network for Organ Sharing (UNOS) divides advanced lung disease diagnoses into four categories: (A) obstructive lung disease (including non–cystic fibrosis-related bronchiectasis and obliterative bronchiolitis), (B) pulmonary vascular disease, (C) cystic fibrosis, and (D) restrictive lung disease. Historically, obstructive lung disease was the most common indication for transplantation, but after the implementation of the LAS system, idiopathic pulmonary fibrosis (IPF), the most common restrictive lung disease, has become an increasingly frequent indication. In 2023, UNOS transitioned to the lung Composite Allocation Score (CAS) system in the United States. This system is designed to incorporate factors beyond survival, including ethical considerations such as access and practical considerations such as efficiency, as well as efforts to optimize survival.

Prior to the era of antifibrotic therapy, the average life expectancy from the time of diagnosis of IPF was 3–5 years, making patients with this disease the cohort to experience most clearly a survival benefit from lung transplantation. As a result, the LAS prioritizes patients with IPF. Similarly, patients who experience secondary effects of their lung disease, including pulmonary hypertension, right heart dysfunction, and hypercarbia, are prioritized for allocation and should be considered for referral for transplant evaluation irrespective of other markers of disease severity.

Generally, the trajectory of decline and evolution of disease are key indicators of the appropriate timing of referral and listing for lung transplantation, rather than absolute thresholds of disease severity. However, suggested guidelines for referral have been elucidated for specific disease states. For example, in chronic obstructive pulmonary disease, the most common obstructive lung disease for which transplantation is considered, the Body-Mass Index, Airflow Obstruction, Dyspnea, and Exercise Capacity (BODE) Index is often used as a marker of disease severity, with an index of 5 being an appropriate indication for referral for evaluation, and 7 for listing for transplantation. Other suggested markers for transplant consideration in obstructive lung disease include pulmonary function test (PFT) data, such as a forced expiratory volume in 1 s (FEV_1) of <25% predicted. The frequency and severity of exacerbations of disease should also be considered in determining the appropriate timing of referral.

With the marked advances in medical therapy for pulmonary vascular disease over the past decade, transplantation for pulmonary vascular disease has become less frequent but is still an important consideration for patients who progress despite, or are refractory to, treatment. Functional assessments, such as New York Heart Association class III or IV limitations, and hemodynamic measurements, such as cardiac index <2 L/min per m^2, would suggest consideration for evaluation and listing. Patients with diagnoses generally poorly responsive to therapy such as pulmonary veno-occlusive disease should be referred early for evaluation.

Patients with cystic fibrosis (CF) have historically been considered for evaluation when the FEV_1 reaches ~30% predicted. However, with the exciting development of therapies targeting the CF transmembrane receptor, providers should keep in mind the potential for improvement in pulmonary function after treatment initiation. Despite this potential for pharmacotherapeutic response, referral and completion of testing should be considered so that patients are prepared for listing should they fail to see improvement or experience worsening of disease on therapy, and for patients who do not qualify for treatment. Moreover, patients who experience acceleration of acute exacerbation rate, recurrent hemoptysis, worsening functional and/or nutritional status, or colonization with resistant bacteria should also be assessed for transplantation irrespective of pulmonary function results.

Despite treatment progress with the development of antifibrotic therapy for IPF and other progressive fibrosing interstitial lung diseases, these therapies do not reverse the disease but only slow the rate of lung function decline. Therefore, transplant referral for patients with IPF and other fibrosing lung diseases should still be considered at the time of diagnosis. Forced vital capacity <80% predicted or diffusing capacity for carbon monoxide <40% predicted, failure to respond to medical therapy, decline in pulmonary function tests on therapy, and functional decline are additional indications for transplant consideration in patients with other restrictive lung diseases.

CONTRAINDICATIONS TO LUNG TRANSPLANTATION

Absolute Contraindications As experience with lung transplantation increases, and as lung allocation policy has prioritized patients with higher acuity of disease and with diseases affecting older age groups, recipient selection criteria have become more liberal compared to prior eras. While published guidelines suggest absolute and relative contraindications to transplantation, these criteria are in constant evolution, and each program ultimately establishes its own selection algorithms based upon clinical expertise, experience, program size and resources, and referral patterns.

Examples of absolute contraindications to lung transplantation (Table 309-1) include anatomic and technical considerations that would affect the ability to complete the transplant procedure, such as chest wall or spinal deformities or malacia of the large airways. Surgical input is critical in making such determinations. In addition, untreatable and/or irreversible organ dysfunction may preclude isolated lung transplantation. Cirrhosis of the liver, uncorrectable disease of the coronary arteries not amenable to combined surgical intervention during the transplant procedure, or other forms of uncorrectable atherosclerotic or vascular disease may make transplantation too high risk

TABLE 309-1 Contraindications to Lung Transplantation

	ABSOLUTE CONTRAINDICATIONS	RELATIVE CONTRAINDICATIONS
Surgical considerations	Anatomic abnormalities not amenable to transplant procedure	
Age		>65 years
Functional status	Immobility, inability to participate in physical therapy/rehabilitation	Limited functional status as defined by 6-minute walk distance
Medical comorbidities	Untreatable, irreversible organ dysfunction	Chronic kidney disease
	Active malignancy or malignancy with insufficient remission period	
	Active bacterial bloodstream infection	Infection resistant to treatment or of high risk for posttransplant morbidity/mortality (*Burkholderia cenocepacia, Mycobacterium abscessus*)
	Uncontrolled viral infection (HIV, hepatitis)	
Nutritional		BMI <18 or >30–35
Psychosocial	Untreatable, irreversible psychiatric disorder with potential to impact transplant outcome	
	Active substance abuse	Limited social supports
	Other circumstances that would impede ability to participate in and comply with posttransplant care	History of noncompliance with medical treatment

Abbreviations: BMI, body mass index; HIV, human immunodeficiency virus.

for consideration. Renal dysfunction is of particular concern given the known nephrotoxicity of calcineurin inhibitors, which are the mainstay of posttransplant immune suppression.

Relative Contraindications Age in and of itself is typically not a contraindication to transplantation at most centers. However, older patients with significant medical comorbidities may be at prohibitive risk for transplantation, and functional status and frailty may worsen in this setting. Published analyses of the Scientific Registry of Transplant Recipients, a comprehensive database of transplant outcomes, have consistently shown that functional capacity, as assessed by 6-minute walk distance, is inversely correlated with both wait list and posttransplant mortality. As a result, most programs utilize some assessment of functional status as a criterion for transplant candidacy.

Frailty, independent of walk distance, has also been recognized increasingly as a marker of poor outcome after lung transplantation and can be assessed using a number of instruments, including the Short Physical Performance Battery (SPPB), Fried Frailty Phenotype (FFP), and others. Most studies utilizing these instruments have been conducted at single centers. While both SPPB and FFP have been shown to correlate with the LAS, FFP has a stronger correlation, and SPPB and FFP may not correlate with each other. The Lung Transplant Frailty Scale (LT-FS) incorporates body composition and serum biomarker measurements and has been demonstrated to be a better predictor of outcomes in the lung transplant population compared to other assessments.

Patients with a history of malignancy are generally required to have experienced a period of remission prior to consideration for transplantation. The necessary length of disease-free survival should be determined in the context of the type of malignancy, stage at diagnosis, and likelihood of recurrence, and often varies by program.

A history of respiratory infection and colonization with resistant organisms is of particular concern in the CF and bronchiectasis populations, although it could affect patients with any advanced lung disease and a history of respiratory infection. Data on outcomes in the presence of resistant *Pseudomonas aeruginosa* infection are conflicting, but, in general, patients who have demonstrated a response to an antimicrobial regimen, even if colonized with resistant organisms, can be considered for transplantation. *Burkholderia cepacia* complex, another group of gram-negative organisms that can infect patients with CF, also presents a unique concern for transplantation. Data show that *B. cenocepacia* (formerly known as Genomovar III) portends the highest risk after transplant, often leading to bacteremia, abscess formation, and early mortality. *B. dolosa* and *B. gladioli* may cause similar posttransplant complications. Patients colonized with other *Burkholderia* species appear to have posttransplant outcomes comparable with the noncolonized population. Published guidelines suggest that those programs offering transplantation to patients colonized with *B. cenocepacia* do so under a research structure and after a specific discussion of the risks of transplantation in this setting with the patients. Other infectious considerations in lung transplant candidates include mycobacterial infections, particularly with rapidly growing organisms such as *M. abscessus*, which can lead to chronic, refractory infections and infections of the chest wall. In the case of fungal infection, assessment of the pathogenicity of the organism, resistance patterns, and, in some cases, responsiveness to pretransplant treatment is beneficial in determining the safety of transplantation.

A history of viral infections such as hepatitis and human immunodeficiency virus (HIV) is generally not considered a contraindication to transplantation. Demonstration of adequate control of infection and responsiveness to therapy are important in preparation for transplantation, and development of a treatment plan that minimizes toxicity and drug interactions, in consultation with transplant pharmacy, should be completed prior to placement on the wait list. Collaborative assessment with transplant infectious disease experts is beneficial whenever the infectious history is a concern for transplant safety.

Nutritional status is another important element to assess in determining candidacy for lung transplantation. Nutritional status has been shown to have a U-shaped relationship with transplant outcomes, with increased mortality risk associated with both underweight (body mass index [BMI] <18) and obesity (specifically BMI >35). Consultation with nutritional experts may allow for modification of this risk prior to transplant. In some underweight patients, placement of an enteral feeding tube and initiation of enteral feedings may be considered.

Psychosocial assessment is also a key component of the evaluation of patients under consideration for lung transplantation, and a multidisciplinary approach with transplant social work, psychiatry, and financial care coordination is often helpful. Assessment for and optimization of psychiatric disorders such as anxiety and depression, which can be exacerbated in the setting of transplantation, substance abuse disorders, and compliance with medical therapy recommendations are all important parts of the pretransplant evaluation. Perioperative pain management planning in the setting of medical therapy for opioid use disorder may require additional multidisciplinary input, but the need for this management plan alone should not preclude transplantation. Transplant candidates require a strong support system given their potential posttransplant care needs. Additionally, confirmation of insurance coverage for all phases of transplant care, expected medication copayments, and financial resources to support other expenses in the setting of transplantation should be completed during the transplant evaluation. Fundraising opportunities and subsidies for medications may need to be pursued in order to proceed safely with listing.

LUNG TRANSPLANT CANDIDATE MANAGEMENT

Lung transplant candidates benefit from meticulous medical care to ensure that they are in optimal condition at the time of transplant. Oxygen is prescribed to maintain adequate systemic oxygenation to allow for moderate physical activity and exertion. Patients should be enrolled in pulmonary rehabilitation programs, if available, and should continue to participate in daily physical exercises.

Patients with pulmonary vascular disease and severe pulmonary hypertension awaiting lung transplantation need special attention to maintain adequate right ventricular function. The use of pulmonary vasodilator therapy is recommended and should not be stopped prior to transplant. Patients who develop secondary pulmonary hypertension should also be assessed for the utility of direct pulmonary vasodilator therapy. Periodic assessment of right ventricular function with echocardiography is recommended, and in patients with clearly worsening ventricular function, right heart catheterization and assessment for responsiveness to short-acting vasodilator therapy should be considered.

In restrictive lung disease patients awaiting transplant, consideration should be given to continuation of immune modulators and/or antifibrotic therapy. Available literature does not indicate that continuation of antifibrotic therapy in lung transplant candidates before transplant portends an increased risk of wound dehiscence or worsened outcomes after transplant. Additionally, increased pulmonary vascular resistance can occur in these patients as the disease progresses, and acute exacerbations have been shown to be associated with severe acute decrease in right ventricular function. Steroids have been utilized in the management of acute exacerbations; however, the negative sequelae of chronic steroid use on wound healing is well established. Therefore, steroid use should be limited as much as possible and, if unavoidable, should be tapered rapidly.

Patients with CF can have pancreatic dysfunction leading to difficulty in maintaining normal blood glucose levels; uncontrolled diabetes mellitus can make the management of posttransplant blood glucose very challenging. Therefore, optimization of diabetes management should be pursued prior to transplantation.

Despite optimal medical therapy, the underlying disease in wait-listed patients will almost always continue to worsen. Prioritization of patients awaiting lung transplant is determined by the CAS system. The LAS system was introduced in 2005 with the goal of minimizing wait-list mortality and maximizing posttransplant survival. This system focused on prioritizing candidates most at risk for death while awaiting transplantation, and its key modeling variables focused on survival metrics. The CAS was introduced in 2023. While incorporating survival as a key component, the CAS also includes variables related to special biological considerations such as blood type and sensitization with human leukocyte antigen (HLA) antibodies, which can impact time to transplantation, as well as variables related to ethical and practical considerations such as patient access to transplant and efficiency of the planned procedure. In addition, the CAS aims to create an allocation system that addresses prioritization more continuously, rather than placing candidates within firm boundaries or groups that create hard cutoffs for transplant access. While under the LAS patients continued to die at a rate of 10–12 patient deaths per 100 patient-years on the wait list, models of CAS allocation suggest expected improvements in both wait-list mortality and post-transplant survival, as well as improvements in equity in access to transplantation. Further studies of the actual results after implementation will be needed to confirm these expectations.

A major consequence of improved efficiency in matching the sickest patients to the available pool of donors has been an increased use of extracorporeal membrane oxygenation (ECMO) devices to bridge the most critically ill patients to transplant. Mechanical circulatory support with ECMO allows for patients to be potentially weaned from the ventilator, to maintain physical activity and ambulation, and to be in a state of greater robustness as they await transplant. The posttransplant survival rate of patients bridged to transplant with ECMO is equivalent to those transplanted without the need for ECMO in experienced, high-volume centers and better than patients who had previously been transplanted directly from mechanical ventilatory support. Furthermore, with improvements in membrane oxygenator technology, platform miniaturization, and improvements in cannula design, outcomes continue to improve.

DONOR CONSIDERATIONS

The ideal lung donor has remained constant since the inception of lung transplantation in the 1980s **(Table 309-2)**. A donor between 25 and 40 years of age, with a $\text{Pao}_2/\text{Fio}_2$ ratio >350, no smoking history, a clear chest x-ray, clean bronchoscopy, and minimal ischemic time is considered the ideal donor; however, it is quite rare that a donor meets all of these criteria for transplantation. In fact, the vast majority of donor lungs used for transplant fall outside these ideal lung donor criteria as established more than three decades ago. Donors must have irreversible brain injury, and the majority of donors are brain dead. Only 20% of all donors with brain death are suitable lung donors due to the development of severe neurogenic pulmonary edema and increased susceptibility of potential lung allografts to infection and injury.

Absolute contraindications to lung donation include radiographic evidence of chronic lung disease such as emphysema and pulmonary fibrosis. Other absolute contraindications include active malignancy, a donor history of severe asthma requiring multiple hospitalizations, and positive HIV status. Relative contraindications include older donor age, severe thoracic trauma with extensive pulmonary contusions, the presence of pulmonary hypertension, and prolonged donor hypotension or acute hypoxemia.

The standard lung donor evaluation includes a donor medical and social history, physical examination, and laboratory examination. Chest imaging is mandatory, as are arterial blood gases, bronchoscopy, and serologic tests for cytomegalovirus (CMV), Epstein-Barr virus (EBV), hepatitis B and C, HIV, *Toxoplasma*, rapid plasma reagin, and herpes simplex virus. The presence of consolidation and atelectasis, while not absolute contraindications to transplant, are often difficult to assess with noncontrast radiographic imaging alone. Ventilation parameters must be evaluated to ensure adequate compliance of the donor lungs, with peak airway pressures <30 cmH_2O being ideal. Direct on-site inspection of the lungs and assessment for nodules, compliance, and full expansion are the final necessary steps before acceptance of donor lungs for transplant.

TABLE 309-2 Characteristics of the Ideal Lung Donor

Donor age	<55 years
ABO compatibility	Identical
Chest radiograph	Clear
Pao_2:Fio_2	>300 on PEEP 5 cmH_2O
Tobacco history	<20 pack-years
Chest trauma	Absent
Evidence of aspiration	Absent
Prior thoracic surgery	None
Sputum Gram stain	Negative
Bronchoscopy findings	No purulent secretions

Abbreviation: PEEP, positive end-expiratory pressure.

More recently, there has been an expanded use of allografts from donors after cardiac death (DCD) due to the ability to rehabilitate donor lungs using ex vivo lung perfusion (EVLP). DCD donors are patients who present with irreversible brain injury but without overt brain death. The potential donor allografts are often exposed to a period of prolonged warm ischemia during the donation process; there has been a concern about early graft dysfunction after DCD donation. Steen and colleagues in Lund demonstrated that EVLP could be used to assess these marginal donors prior to transplant. The landmark publication of the Normothermic Ex Vivo Lung Perfusion in Clinical Lung Transplantation trial in 2011 generated renewed interest in DCD lung donors. The group from the University of Toronto was also able to demonstrate that brain-dead donors with unacceptable donor lung parameters could be rehabilitated with the use of EVLP. They were able to salvage up to 50% of selected unsuitable donor lung allografts with the use of acellular normothermic hyperosmotic perfusion with excellent short-term outcomes.

Donor Management Brain death causes severe perturbations in the potential donor lung allograft function. The development of severe pulmonary edema often accompanies brain death. The hemodynamic instability and neurogenic shock that can accompany brain death are also major stressors on the preservation of donor allograft function. The primary goal of donor management is, therefore, the maintenance of hemodynamic stability and preservation of donor lung function. Judicious fluid resuscitation and avoidance of excessive resuscitation should be employed. Volume replenishment should be limited to maintain the central venous pressure between 5 and 8 mmHg. In general, crystalloid fluid boluses are to be avoided. Diabetes insipidus is common in donors and requires the use of intravenous vasopressin to prevent excessive urine loss. In general, blood transfusions should be avoided; however, if necessary, CMV-negative and leukocyte-filtered blood should be used whenever possible. Hypothermia should be avoided as it predisposes to ventricular arrhythmias and metabolic acidosis.

Excessive oxygen delivery should be minimized to prevent free-radical injury to the potential lung allograft. Positive end-expiratory pressures on the ventilator should be maintained to avoid the development of atelectasis. More recently, airway pressure release modes of ventilation have been utilized to preserve lung function and minimize barotrauma from prolonged ventilation.

PROCUREMENT OPERATION

Prior to incision, a thorough bronchoscopic evaluation is completed. The anatomy of the donor airways is defined. Any secretions that may be present are evacuated, and the airways are examined to rule out the presence of any lesions or masses. The epithelial lining is inspected for evidence of excessive friability and hemorrhage, which may indicate significant infection. A median sternotomy incision is employed to access the chest for lung procurement. The pleural spaces are opened and both lungs are inspected, palpated, and gently recruited to evaluate for suspicious nodules, consolidation, and/or pulmonary infarction.

The donor is systemically heparinized, and the main pulmonary artery is cannulated. Fifteen minutes prior to initiation of the explant, prostacyclin is introduced into the main pulmonary artery and allowed to circulate through the lungs. This vasoreactive prostanoid helps to ensure adequate pulmonary flush by dilating the pulmonary vasculature. The heart is arrested first, then the pulmonoplegia solution is instilled into the lungs at a low controlled pressure. Topical iced-saline solution is instilled into both pleural spaces. After the heart has been explanted, the individual pulmonic veins are flushed retrogradely. The lungs are then re-expanded, the trachea is clamped, and the explanted allograft is stored in ice-cold saline solution for transport. If the right and left lungs are being procured for different recipients, the posterior left atrium, the main pulmonary artery, and the left main-stem bronchus are divided to separate the right and left lungs, and the organs are stored and shipped separately. New preservation technologies and studies of optimal temperature for storage and transport may allow for longer transport times and time to implantation. Investigations of the impacts of these advances are ongoing.

RECIPIENT OPERATION AND EARLY POSTTRANSPLANT CONSIDERATIONS

Recipient Operation The recipient operation can be divided into two parts. The first part involves the explant of the native lung, and the second part involves the implant of the new lung. There are generally three main surgical approaches to the completion of the operation: a right or left thoracotomy, a transverse thoracosternotomy (clamshell), or a median sternotomy. These approaches are all favored by various centers for different benefits. The thoracotomy approach allows for explant and implant of donor lungs without the use of cardiopulmonary bypass (CPB) and is often the preferred approach for single-lung transplant. The clamshell incision offers the advantages of increased exposure compared to either thoracotomy or median sternotomy but comes at the cost of greater morbidity and postoperative wound complications. This incision can be used to perform bilateral lung transplants and allows for the possibility of avoiding CPB. A median sternotomy approach can be used to perform bilateral lung transplant. This approach offers the advantage of fewer wound complications, less postoperative pain, and flexibility with more complex or concomitant cardiac procedures at the time of lung transplant. This approach mandates the use of CPB. The routine use of CPB allows for early pneumonectomies without hemodynamic compromise and can significantly reduce the ischemic time to the second allograft. Additionally, overcirculation to the first allograft can be minimized with the routine use of CPB. Others prefer to avoid CPB as avoidance may be associated with decreased need for blood product administration and lower incidence of primary graft dysfunction.

Over the past 5 years, there has been a movement away from CPB for mechanical circulatory support toward ECMO. ECMO is an alternative strategy for providing hemodynamic and pulmonary support of the patient undergoing lung transplantation. The key difference between CPB and ECMO is the ability to salvage shed blood with CPB, the use of a blood reservoir with CPB, and the ability to filter air from the venous system before it is delivered to the arterial circuit with CPB. This flexibility is provided by exposing the blood in the system to a much larger foreign surface and has been associated with significantly increased inflammation and tendency for thrombosis. Therefore, ECMO requires much lower levels of anticoagulation than CPB. The advantages of ECMO-supported lung transplantation are a tendency toward less blood product administration and, in situations in which CPB support periods are longer (>180 min), a tendency to decreased rates of primary graft dysfunction. Data suggest that the use of ECMO for mechanical support during lung transplant is safer than lung transplant with no mechanical support, and the recommendation is to use ECMO for all cases. Its use may especially be helpful for cases that are expected to require longer support times. Careful attention to avoid the entrainment of air into the circuit is mandatory throughout the ECMO support period.

Anesthetic monitoring for lung transplant should include arterial pressure monitoring, pulse oximetry, continuous electrocardiographic monitoring, temperature monitoring, and urine output monitoring. Large-bore IV access and central venous access are vital to manage the patient safely. On a selective basis, pulmonary artery pressure monitoring and transesophageal echocardiographic monitoring may be useful. For patients without the planned use of CPB, double-lumen endotracheal tubes are mandatory, whereas they can be avoided for patients transplanted on CPB.

Once access to the thorax has been completed, the hilar structures are isolated and divided. The bronchial anastomosis is completed first, and the anastomosis is checked to ensure that it is secure by insufflating the lung gently while keeping the anastomosis under saline solution to observe for bubbles. The donor left atrial cuff incorporating the pulmonary vein is connected to the native left atrium, and the donor right or left pulmonary artery is connected to the native pulmonary artery. After completion of the vascular anastomoses, the lungs are gently reperfused. During this early reperfusion period, lung-protective ventilation strategies are employed and oxygen tension is reduced. The patient is transitioned to normal ventilation, drains are placed in the thoracic cavity, and the wounds are closed.

Induction of Immunosuppression Initiation of immunosuppression starts with induction of the patient under general anesthesia. Many programs utilize an induction agent (most commonly an interleukin 2 [IL-2] receptor/CD25 antagonist, but antithymocyte globulin, anti-CD52 monoclonal antibodies, or other induction agents may also be used), and systemic corticosteroids and purine modulators are administered after induction is complete. If an IL-2 receptor antagonist is utilized for induction, a second dose is administered 4 days after the original dose. An additional dose of methylprednisolone is administered after allograft reperfusion in the operating room. Three-drug immune suppression is initiated with a calcineurin inhibitor, purine modulator, and continued systemic corticosteroids. In patients with severe acute renal dysfunction, calcineurin inhibitor initiation may be delayed.

Perioperative Considerations and Complications Early morbidity and mortality after lung transplant most commonly are sequelae of primary graft dysfunction or infection. Very rarely, hyperacute rejection has been observed; however, with the implementation of robust systems to ensure ABO and HLA compatibility at the time of transplant, the occurrence of hyperacute rejection is extremely uncommon. Primary graft dysfunction (PGD) encompasses a constellation of findings that result in poor early graft function after transplant. This phenomenon is often the consequence of ischemia-reperfusion injury in the allograft and is not related to infection or rejection. It is characterized by a diffuse pattern of infiltrates on the chest x-ray and poor pulmonary gas exchange with $\mathrm{Pa{O_2}}$:$\mathrm{Fi{O_2}}$ ratios <300, with severe PGD characterized by diffuse severe infiltrates and a $\mathrm{Pa{O_2}}$:$\mathrm{Fi{O_2}}$ ratio of <100 at 72 h posttransplant. Most cases of PGD are mild and self-limiting, resolving with supportive care. However, if the PGD is severe and worsening despite maximal medical therapy, diuresis, inotropic therapy, maximal ventilation support, and paralysis of the patient, mechanical circulatory support with ECMO can become necessary. The incidence of severe PGD has been steady over the past two decades at approximately 10–15% in most programs. Severe PGD at 72 h posttransplant portends an increased mortality risk and is a risk factor for chronic lung allograft dysfunction (CLAD).

Bacterial, viral, and fungal infections are leading causes of morbidity and mortality in lung transplantation. The lung is one of the few solid organs that is in continuous contact with the environment. Each breath has the potential to introduce new organisms, and the reduced lymphatic function and mucociliary clearance in the transplanted lung increase the risk of serious infection. The highest incidence of infection is early after lung transplant and coincides with the intensity of immune suppression. Early infections, occurring within the first month after transplantation, are commonly bacterial (especially gram-negative bacilli) and manifest as pneumonia, mediastinitis, urinary tract infections, catheter sepsis, and skin infections. Patients can develop pathogenic infections with organisms associated with pretransplant colonization, and perioperative antibiotic regimens are often deployed to address this. Viral infections, and CMV infections in particular, can lead to severe recipient disease and early loss of graft and life. The majority of transplant programs employ antiviral prophylaxis in the early transplant period to avoid such complications. Invasive fungal infections peak in frequency between 10 days and 2 months after transplantation. Fungal prophylaxis regimens in the early posttransplant period vary widely. Treatment consists of inhaled amphotericin B in the setting of airway infection and/or azole therapy with more advanced or invasive disease. The institution of prophylaxis with oral trimethoprim-sulfamethoxazole (or atovaquone or inhalational pentamidine for sulfa-allergic patients) has effectively prevented *Pneumocystis* pneumonia. The risk of *Pneumocystis* infection is highest during the first year after transplant. However, as infections can also occur late after transplant, most centers recommend prophylactic therapy be continued for life.

LONG-TERM MANAGEMENT OF LUNG TRANSPLANT RECIPIENTS

While survival after lung transplantation continues to improve by era, the survival rates in this group are lower than in other solid-organ cohorts. Approximately 50% of lung transplant recipients will experience at least one episode of acute rejection in the first posttransplant year, and by 5 years posttransplant, approximately half will have developed chronic rejection. As a result, posttransplant immune suppression regimens may be more aggressive than in other solid-organ recipients, as described above. The immunosuppressive regimen must be balanced against the potential toxicities that accrue with these medications over time.

Acute cellular rejection in lung transplant recipients is most common in the first posttransplant year, with a decreased but not absent frequency thereafter. Infections can stimulate cellular rejection, most clearly demonstrated in the setting of CMV infection, but also noted after other infections. Most programs incorporate a schedule of routine surveillance bronchoscopy to assess for acute cellular rejection posttransplant. Donor-derived cell-free DNA, which is produced in the setting of cell turnover and parenchymal injury, is increasingly being explored as a noninvasive method of detection of allograft dysfunction, most importantly in the setting of acute and chronic rejection. Acute cellular rejection manifests as a lymphocytic infiltrate involving the distal small vessels and capillaries and/or a lymphocytic bronchiolitis involving the distal airways of the lung. Acute cellular rejection, a risk factor for the development of CLAD, is treated with augmented immune suppression. Antibody-mediated rejection in its classic form is a neutrophilic vasculitis associated with the small vessels and capillaries of the lung, with associated deposition of by-products of the complement cascade, in the setting of allograft dysfunction and circulating donor-specific HLA antibodies in the blood. The manifestations of antibody-mediated rejection in the lung allograft are less specific than in other organs. Further research is ongoing into the diagnostic and treatment considerations of this entity in lung transplantation.

CLAD is an overarching description of the syndrome of long-term allograft rejection. The classic manifestation of CLAD is obliterative bronchiolitis, the development of fibrinous material within the distal airways that leads to small-airways obstruction. As transbronchial biopsies are insensitive for diagnosing obliterative bronchiolitis, a clinical diagnostic designation of bronchiolitis obliterans syndrome can be made when specific PFT criteria are met and other causes of PFT decline are excluded. CLAD can also present as a restrictive phenotype, with imaging demonstrating upper lobe–predominant pleural thickening, small lung volumes, and interstitial changes on high-resolution computed tomography (CT). Numerous therapies for CLAD have been utilized, including azithromycin, montelukast, extracorporeal photopheresis, alemtuzumab, and others, with varying degrees of success.

Infection is a significant complication of lung transplantation, with persistent risk over the lifetime of the transplant recipient. As time progresses, the chance of opportunistic infection increases. The risk of bacterial infection and fungal infection remains, and can affect the lung parenchyma, airways and anastomotic sites, and other organs. Viral infections, such as CMV reactivation and infection, EBV-associated posttransplant lymphoproliferative disease, and other rarer infections, can also develop in the later posttransplant setting as well.

Numerous longer-term medical complications can be seen in lung transplant recipients. Essential hypertension, diabetes mellitus, chronic renal insufficiency, and bone loss are some examples of chronic medical conditions observed following transplantation. A multidisciplinary approach to care that involves the patient's primary care physician, local pulmonologist, and appropriate subspecialists, along with transplant pharmacy, as well as social work and care coordination, is beneficial in addressing the complex needs of lung transplant recipients over time. Predictors of short- and long-term outcomes after lung transplantation are outlined in **Table 309-3**.

TABLE 309-3 Predictors of Survival After Lung Transplantation

	1-YEAR SURVIVAL	≥10-YEAR SURVIVAL
Donor factors	HCV donor	
Recipient factors	Age <70 years Diagnosis other than pulmonary fibrosis, pulmonary hypertension, sarcoidosis, A1AT O_2 requirement <5 L CI >2 Outpatient at time of transplant Preserved recipient eGFR Total bilirubin <2	Age 18–35 years
Donor/recipient factors	Non–female-to-male transplant Donor/recipient weight ratio >0.7	Higher levels of HLA matching
Operative factors	Avoidance of unplanned conversion to cardiopulmonary bypass Decreased ischemic time	Bilateral lung transplant
Posttransplant factors	PaO_2/FIO_2 >260 at 72 h Absent need for postoperative ECMO support	Fewer hospitalizations for rejection
Other factors	Higher center volume	Higher center volume

Abbreviations: A1AT, α_1 antitrypsin deficiency; ECMO, extracorporeal membrane oxygenation; eGFR, estimated glomerular filtration rate; Fio_2, fraction of inspired oxygen; HCV, hepatitis C virus; HLA, human leukocyte antigen; Pao_2, partial pressure of oxygen.

FURTHER READING

Cypel M et al: Normothermic ex vivo lung perfusion in clinical lung transplantation. N Engl J Med 364:1431, 2011.

Leard LE et al: Consensus document for the selection of lung transplant candidates: An update from the International Society for Heart and Lung Transplantation. J Heart Lung Transplant 40:1349, 2021.

Lehr CJ et al: The impact of change in definition of increased-risk donors on survival after lung transplant. J Thorac Cardiovas Surg 160:572, 2020.

OPTN Lung Transplantation Committee: Establish continuous distribution of lungs. *https://optn.transplant.hrsa.gov/media/esjb4ztn/20211206-bp-lung-establish-cont-dist-lungs.pdf*. Accessed January 11, 2023.

Schwartz S et al: Procedural mechanical support for lung transplantation. Curr Opin Organ Transplant 26:309, 2021.

310

Interventional Pulmonary Medicine

Lonny Yarmus, David Feller-Kopman

Interventional pulmonary medicine is a subspecialty of pulmonary and critical care medicine focusing on the evaluation and management of patients with thoracic malignancy, central airway obstruction, pleural disease, and advanced obstructive lung disease such as chronic obstructive pulmonary disease (COPD)/emphysema and asthma. Novel minimally invasive interventions have drastically changed the way we care for patients. In this chapter, we will summarize recent developments and evolving technologies in interventional pulmonology (IP).

DIAGNOSTIC BRONCHOSCOPY

With the introduction of the rigid bronchoscope by Gustav Killian in 1897, the mortality associated with foreign-body aspiration dropped from over 90% to less than 5%, as patients no longer had to suffer from airway obstruction and postobstructive pneumonia. Shigeto Ikeda developed the flexible bronchoscope in 1967, allowing access to the peripheral airways and lung parenchyma. In 2018, the first robotic-assisted bronchoscopy platforms were introduced, providing a novel approach with the ability to reach further into the peripheral lung reliably with precise control. Bronchoscopy has remained an important diagnostic and therapeutic procedure, and recent technology has significantly increased its utility.

■ ENDOBRONCHIAL ULTRASOUND

The diagnosis and staging of lung cancer remain one of the most important roles of advanced diagnostic bronchoscopy and IP. Convex endobronchial ultrasound (cEBUS) is a flexible bronchoscope combined with ultrasound technology that allows for real-time visualization during transbronchial needle aspiration (TBNA) of mediastinal and hilar lymph nodes and masses adjacent to the airways **(Fig. 310-1)**.

With a sensitivity of 90% and a specificity of 100%, cEBUS is the gold standard for lung cancer staging and can also provide sufficient tissue to perform molecular profiling to guide targeted therapies in lung cancer with adequacy rates for testing that exceed 95%. cEBUS is also extremely helpful in diagnosing mediastinal and hilar adenopathy due to sarcoidosis. The use of endobronchial ultrasound to diagnose lymphomas has historically been of limited utility owing to lack of tissue architecture in needle aspirates. However, advances in cEBUS using a cryobiopsy technique in lieu of needle aspirate have shown promise in providing adequate tissue and histopathologic architecture from intranodal cryobiopsy.

■ PERIPHERAL BRONCHOSCOPY

Evaluations of pulmonary nodules and lung masses are frequent indications for bronchoscopy as a way to achieve a minimally invasive diagnosis. Historically, the diagnostic yield of bronchoscopy to target peripheral pulmonary lesions was <60%. Multiple guidance platforms now allow for improved access in the periphery of the lung.

Smaller or ultrathin bronchoscopes <4 mm in diameter can be combined with available imaging tools to improve target localization. Radial-probe endobronchial ultrasound utilizes a radial scanning ultrasound probe that is inserted through the bronchoscope and into the lung, producing a real-time image of the target lesion. Electromagnetic navigation bronchoscopy (ENB) involves image-guidance systems that manipulate thin-slice computed tomography (CT) images to create virtual airway reconstructions used as guided maps during bronchoscopy. Robotic-assisted bronchoscopic platforms offer the enhanced articulation and stability of a robotic arm, replacing the traditional flexible bronchoscope. Recent innovation has allowed for advances in intraprocedural imaging. Mobile cone beam CT scanners in combination with advanced peripheral bronchoscopy allow for image confirmation of a biopsy within the lesion of interest. Studies are currently underway to explore further the utility of these systems for peripheral lesion biopsy and the impact of advanced imaging techniques.

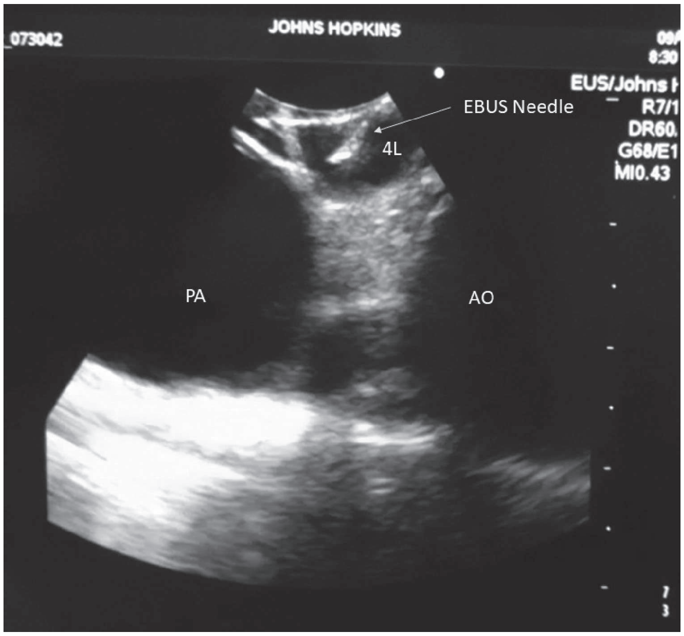

FIGURE 310-1 Endobronchial ultrasound transbronchial needle aspiration image of needle under ultrasound guidance sampling station 4L lymph node. AO, aorta; PA, pulmonary artery.

THERAPEUTIC BRONCHOSCOPY

Therapeutic bronchoscopy is indicated for the relief of malignant and nonmalignant central airway obstruction, asthma, and emphysema. Active research is also focusing on the utility of bronchoscopy for the ablation of early-stage lung cancer, as well as the treatment of chronic bronchitis.

■ CENTRAL AIRWAY OBSTRUCTION

Central airway obstruction (CAO) describes obstruction of the trachea, main stem bronchi, bronchus intermedius, and/or lobar bronchi, and can present as intrinsic (endoluminal), extrinsic (extraluminal), or mixed (extraluminal tumor resulting in mass effect and endoluminal involvement) **(Fig. 310-2)**. The differential diagnosis of CAO is shown in **Table 310-1**.

Patients often initially present with cough and exertional dyspnea, but then progress with increasing severity of obstruction to dyspnea at rest, stridor, and respiratory failure. Patients may also have wheezing, hemoptysis, or symptoms of postobstructive infection. Rigid bronchoscopy is the preferred tool to manage CAO in conjunction with ablative therapies, balloon bronchoplasty, and airway stenting to offer rapid symptomatic relief with immediate reductions in the level of required care. Therapeutic bronchoscopy for CAO has been shown to significantly improve both quality of life and survival.

■ ABLATIVE THERAPIES FOR CAO

Ablative therapy in the airway consists of both heat (laser, electrocautery, and argon plasma coagulation) and cold (cryotherapy) modalities. These techniques are most commonly used to destroy tumor and provide hemostasis. The cryoprobe can also be used for foreign-body removal. Other modalities, such as brachytherapy (BRT) and photodynamic therapy (PDT), have a delayed therapeutic effect and are often not suitable for situations where immediate relief of airway obstruction is desired.

■ BRONCHOPLASTY

Bronchoplasty (or bronchial dilation) can be achieved with the barrel of the rigid bronchoscope or with balloons that can be passed via the rigid or flexible bronchoscope. Bronchoplasty is most commonly used for dilation of stenotic airways or disruption of webs related to nonmalignant causes of airway diseases. Although dilation generally leads to immediate relief of the stenosis, results can be short-lived, and hence, this technique is often combined with airway stenting. Complications are rare but can include airway tears if proper techniques are not followed.

■ AIRWAY STENTING

After airway patency is achieved, airway stents can be utilized to prevent recurrence of CAO. Reports of endoscopically implantable stents for the airways date back to 1914. Airway stents are commonly used to treat patients with CAO due to extrinsic compression from a variety of malignant and nonmalignant disorders. Stents are effective and lead to symptomatic relief in >90% of patients. A variety of airway stents are available, each with its own benefits and detriments; it is important to choose the right stent for the specific indication. Stent complications are not uncommon and include migration, mucostasis, infection, and the development of granulation tissue. First-generation biodegradable stents, custom three-dimensional printed stents, and drug-coated

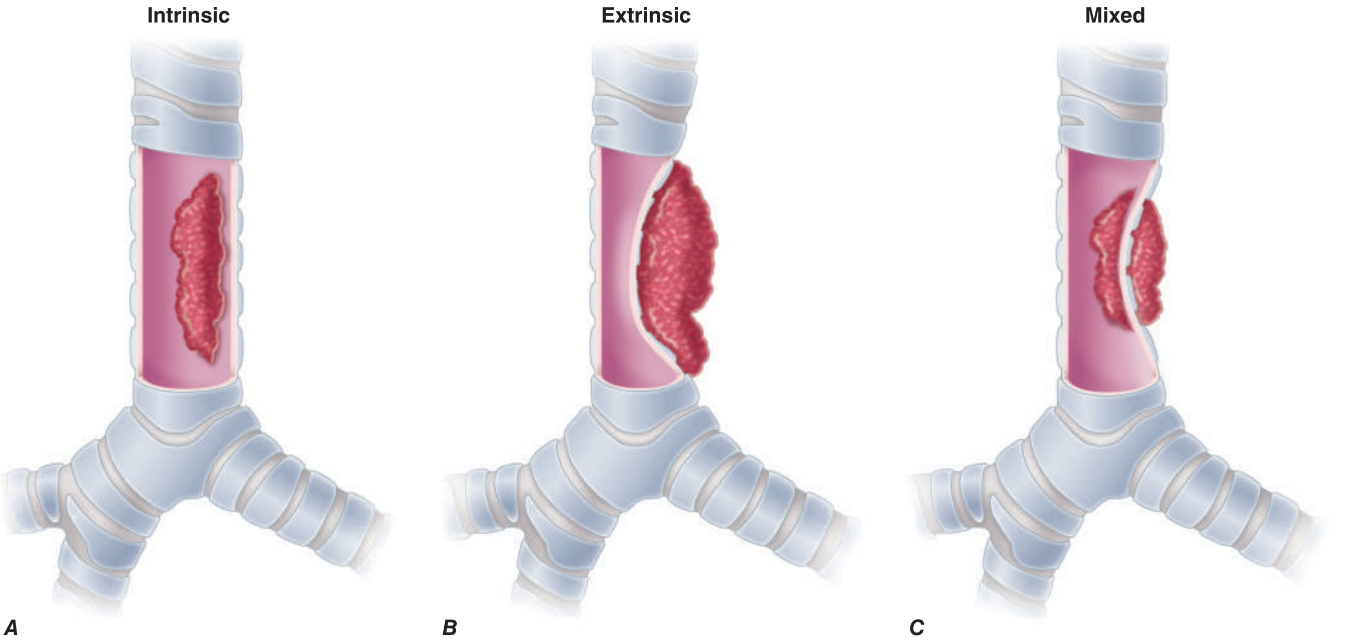

FIGURE 310-2 Types of central airway obstruction.

stents are currently being evaluated, working toward a personalized medicine approach wherein stents are tailored to an individual's airway anatomy and underlying disease.

ENDOBRONCHIAL INTRATUMORAL CHEMOTHERAPY

Endobronchial intratumoral chemotherapy (EITC) is an intervention aimed at improving and/or maintaining airway patency in patients with malignant CAO, with the potential to eliminate the need for airway stenting and its associated complications. Under bronchoscopic guidance, high-dose therapeutics can be safely injected directly into tumor to enhance response and limit systemic side effects. Multiple studies are ongoing to assess the efficacy of EITC.

TABLE 310-1 Differential Diagnosis of Central Airway Obstruction

MALIGNANT	NONMALIGNANT
Primary airway carcinoma	Lymphadenopathy
Bronchogenic	Sarcoidosis
Carcinoid adenoid cystic	Infectious (i.e., tuberculosis, histoplasmosis)
Mucoepidermoid	Cartilage
Metastatic carcinoma to the airway	Relapsing polychondritis
Bronchogenic	Granulation tissue from endotracheal tubes
Renal cell	Tracheostomy tubes
Breast	Airway stents
Thyroid	Foreign bodies
Colon	Surgical anastomosis
Sarcoma	Granulomatosis with polyangiitis
Melanoma	Pseudotumor
Laryngeal carcinoma	Hamartomas
Esophageal carcinoma	Amyloid
Mediastinal tumors	Papillomatosis
Thymus	Hyperdynamic
Thyroid	Tracheomalacia
Germ cell	Bronchomalacia
Lymphadenopathy	Idiopathic
Lymphoma	Tuberculosis
	Sarcoidosis
	Other
	Foreign-body goiter
	Mucus plug
	Blood clot

ABLATIVE THERAPIES FOR EARLY-STAGE LUNG CANCER

Bronchoscopic ablation of early-stage lung cancer has long been described as the "holy grail" of bronchoscopy due to the appeal of staging, diagnosing, and treating biopsy-proven early-stage lung cancer in one procedural setting. There is limited experience with bronchoscopic radiofrequency ablation (B-RFA) and microwave ablation (MWA) as a potential means to treat early-stage lung cancer. Ultimately, the efficacy of bronchoscopic ablation of early-stage nonoperable lung cancer must be proven in longitudinal studies demonstrating noninferiority in survival as compared to the current gold standard of stereotactic body radiation therapy (SBRT). To date, there are extremely limited safety and efficacy data for bronchoscopic ablation with multiple studies currently ongoing. Until there are adequate data to support the safe and effective use of this approach, bronchoscopic ablation is not recommended for clinical use.

BRONCHOSCOPIC THERAPIES FOR ASTHMA

Bronchial thermoplasty (BT) is a treatment for patients with severe persistent asthma who remain symptomatic despite maximal medical treatment that delivers radiofrequency energy to the airways to reduce their smooth muscle mass. A pivotal randomized clinical trial did not show a change in forced expiratory volume in 1 s (FEV_1) or airway hyperresponsiveness but was able to demonstrate an improvement in quality of life and reduction in exacerbation rates, visits to the emergency department, and days lost from school or work. At this time, the ideal asthma phenotypes and ideal candidates for this treatment modality remain to be determined, and thus, the utility of this approach remains limited.

BRONCHOSCOPIC THERAPIES FOR CHRONIC OBSTRUCTIVE PULMONARY DISEASE

The National Emphysema Treatment Trial (NETT), published in 2003, demonstrated that lung volume reduction (LVR) surgery for severe emphysema confers improved survival and exercise capacity in patients with upper lobe–predominant disease and poor exercise capacity. At the same time, it showed high perioperative morbidity and mortality. During the last decade, several bronchoscopic therapeutic modalities have been tested, including valves, coils, steam, stents, and foam, in patients with severe emphysema to mimic the physiologic effects of surgical lung volume reduction (SLVR) in a less invasive fashion.

BRONCHOSCOPIC LUNG VOLUME REDUCTION

Bronchoscopic lung volume reduction (BLVR) via valve placement involves placement of one-way valves in airways leading to areas of the lung with significant emphysema, allowing air and mucus to exit but blocking air entry to achieve lobar collapse. Several clinical trials on BLVR with valves have demonstrated improvements in lung function and overall improvement in quality of life and exercise tolerance. The overall safety profile of these valve systems compares favorably with SLVR with a lower rate of perioperative morbidity and mortality.

PLEURAL INTERVENTIONS

Thoracic ultrasound has become invaluable in the evaluation of patients with pleural effusion and pneumothorax. Medical thoracoscopy (also called *pleuroscopy*) is a minimally invasive technique most commonly used to evaluate recurrent exudative pleural effusions and is associated with a diagnostic yield of >95%.

Indwelling pleural catheters (IPCs) have gained tremendous popularity and have been declared by evidence-based guidelines to be as acceptable as chemical pleurodesis for the management of symptomatic malignant pleural effusions. When comparing IPC and pleurodesis via talc slurry, two multicentered, open-label, randomized controlled trials demonstrated IPC effectively relieved dyspnea, decreased the duration of hospital stay, and lessened the need for future procedures. A recent study in patients without significant lung entrapment has shown that the outpatient administration of talc through an indwelling pleural catheter for the treatment of malignant pleural effusion resulted in a significantly higher chance of pleurodesis at 35 days than an indwelling catheter alone, with no deleterious effects.

Pleural infection (empyema or complex parapneumonic effusion) is commonly encountered in clinical practice. The mainstay of therapy typically consisted of antibiotics, drainage of the infected pleural space with tube thoracostomy, and possible need for surgical decortication. The landmark Multicenter Intrapleural Sepsis Trial (MIST2) demonstrated that intrapleural sequential administration of recombinant tissue plasminogen activator (rtPA) and DNase resulted in significant radiographic and clinical improvements and allowed >90% of patients to avoid surgery. Randomized controlled trials are currently underway comparing intrapleural administration of rtPA and DNase to surgical decortication.

PNEUMOTHORAX AND PERSISTENT AIR LEAK

Persistent air leak is defined as a nonresolving pneumothorax with an air leak lasting more than 5–7 days. For over a decade, the U.S. Food and Drug Administration has maintained a humanitarian device exemption for compassionate use of the Spiration Valve System for management of persistent air leak following lobectomy, segmentectomy, or LVR surgery, although the device has also been used "off label" for the treatment of persistent air leak due to primary and secondary spontaneous pneumothoraces.

SUMMARY

IP provides diagnostic and therapeutic options that span the spectrum of benign and malignant airway and pleural disorders. The constant innovations in diagnostic and treatment modalities have continued to help push the boundaries of pulmonary medicine.

FURTHER READING

Roberts ME et al: British Thoracic Society Guideline for pleural disease. Thorax 78:114, 2023.

Wahidi MM et al: State of the art: Interventional pulmonology. Chest 157:734, 2020.

Section 1 Respiratory Critical Care

311 Approach to the Patient with Critical Illness

Rebecca M. Baron, Anthony F. Massaro

The care of critically ill patients requires a thorough understanding of pathophysiology and centers initially on the resuscitation of patients at the extremes of physiologic deterioration. This resuscitation is often fast-paced and occurs early when a detailed awareness of the patient's chronic medical problems may not yet be possible. While physiologic stabilization is taking place, intensivists attempt to gather important background medical information to supplement the real-time assessment of the patient's current physiologic condition. Numerous tools are available to assist intensivists in the assessment of pathophysiology and management of incipient organ failure, offering a window of opportunity for diagnosing and treating underlying disease(s) in a stabilized patient. However, despite these tools, ongoing clinical bedside assessment is imperative for care of the critically ill patient. Indeed, the use of interventions to support the patient, such as mechanical ventilation and renal replacement therapy, is commonplace in the intensive care unit (ICU). An appreciation of the risks and benefits of such aggressive and often invasive interventions is vital to ensure an optimal outcome. Nonetheless, intensivists must recognize when a patient's chances for recovery are remote or nonexistent and must counsel and comfort dying patients and their significant others if an initial trial of invasive supportive care is either not effective or is not appropriate for the patient's current condition. Critical care physicians often must redirect the goals of care from resuscitation and cure to comfort when the resolution of an underlying illness is not possible. The COVID-19 pandemic has heightened the need and priority for effective critical care practices **(Chap. 205)**, as well as the need for additional support for post-ICU care and recovery for ICU survivors

ASSESSMENT OF ILLNESS SEVERITY

In the ICU, illnesses are frequently categorized by degree of severity. Numerous severity-of-illness (SOI) scoring systems have been developed and validated over the past three decades. Although these scoring systems have been validated as tools to assess populations of critically ill patients, their utility in predicting individual patient outcomes at the bedside is not clear. Their utility may be more applicable toward defining patient populations for clinical trial outcomes and broader epidemiologic studies. SOI scores are also useful in guiding hospital administrative policies, directing the allocation of resources such as nursing and ancillary care, and assisting in assessments of quality of ICU care over time. Scoring system validations are based on the premise that age, chronic medical illnesses, and derangements from normal physiology are associated with increased mortality rates. All existing SOI scoring systems are derived from patients who have already been admitted to the ICU. Nevertheless, there has been increased recent clinical use of scoring systems due to revised consensus guidelines for definitions of sepsis, as will be detailed below.

The most commonly utilized scoring systems are the SOFA (Sequential Organ Failure Assessment) and the APACHE (Acute Physiology and Chronic Health Evaluation). There has been more recent interest in the use of a "quick" SOFA, or qSOFA, scoring system for prognostication of sepsis outcomes.

THE SOFA SCORING SYSTEM

The SOFA scoring system is composed of scores from six organ systems, graded from 0 to 4 according to the degree of dysfunction **(Table 311-1)**. The score accounts for clinical interventions; it can be measured repeatedly (i.e., each day), and rising scores correlate with increasing mortality. The most recent sepsis consensus conference guidelines incorporated an increase of at least two points in the SOFA score from baseline as diagnostic of sepsis in the setting of suspected or documented infection. Patients with suspected infection can be predicted to have poor outcomes typical of sepsis if they have at least two of the following clinical criteria: respiratory rate ≥22 breaths/min, altered mental status, or systolic blood pressure ≤100 mmHg. Recently, a new bedside clinical score using two or more of the above clinical criteria has emerged and is termed quick SOFA (qSOFA). qSOFA is intended to screen patients for risk of poor outcomes from sepsis from

TABLE 311-1 Calculation of SOFA Score[a]

	SCORE				
SYSTEM	**0**	**1**	**2**	**3**	**4**
Respiration Pao$_2$/Fio$_2$, mmHg (kPa)	≥400 (53.3)	<400 (53.3)	<300 (40)	<200 (26.7) with respiratory support	<100 (13.3) with respiratory support
Coagulation Platelets, × 10^3/μL	≥150	<150	<100	<50	<20
Liver Bilirubin, mg/dL (μmol/L)	<1.2 (20)	1.2–1.9 (20–32)	2.0–5.9 (33–101)	6.0–11.9 (102–204)	>12.0 (204)
Cardiovascular	MAP ≥70 mmHg	MAP <70 mmHg	Dopamine <5 or dobutamine (any dose)[b]	Dopamine 5.1–15 or epinephrine ≤0.1 or norepinephrine ≤0.1[b]	Dopamine >15 or epinephrine >0.1 or norepinephrine >0.1[b]
Central nervous system Glasgow Coma Scale[c]	15	13–14	10–12	6–9	<6
Renal Creatinine, mg/dL (μmol/L) or urine output, mL/d	<1.2 (110)	1.2–1.9 (110–170)	2.0–3.4 (171–299)	3.5–4.9 (300–440) or <500	>5.0 (440) or <200

[a]Adapted from JL Vincent et al: Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Med 22(7):707, 1996. [b]Catecholamine doses are given as μg/kg per min for at least 1 h. [c]Glasgow Coma Scale scores range from 3 to 15; higher score indicates better neurological function.

Abbreviations: Fio$_2$, fraction of inspired oxygen; MAP, mean arterial pressure; Pao$_2$, partial pressure of oxygen.

out-of-hospital, emergency department, and hospital ward settings. qSOFA was not developed, nor is it currently recommended, as a sepsis diagnostic screening tool, but studies are investigating its utility as such, especially in resource-poor settings that may not have the ability to measure all the components of the SOFA score.

THE APACHE II SCORING SYSTEM

The APACHE II system is the most commonly used SOI scoring system in North America. Age, type of ICU admission (after elective surgery vs nonsurgical or after emergency surgery), chronic health problems, and 12 physiologic variables (the worst values for each in the first 24 h after ICU admission) are used to derive a score. The predicted hospital mortality rate is derived from a formula that takes into account the APACHE II score, the need for emergency surgery, and a weighted, disease-specific diagnostic category **(Table 311-2)**. The relationship between APACHE II score and mortality risk is illustrated in **Fig. 311-1**. Updated versions of the APACHE scoring system (APACHE III and APACHE IV) have been published.

OTHER SCORING SYSTEMS

There are numerous other scoring systems that have been developed, and there are ongoing studies evaluating their utility. In particular, there is increasing interest in utilizing electronic health medical record scoring systems that might better incorporate larger and real-time data sets from patients, and that can alert providers to patients at risk for sepsis and/or poor outcomes from clinical illness.

SHOCK (SEE ALSO CHAP. 314)

INITIAL EVALUATION

Shock, a common condition necessitating ICU admission or occurring in the course of critical care, is defined by the presence of multisystem end-organ hypoperfusion. Clinical indicators include reduced mean arterial pressure (MAP), tachycardia, tachypnea, cool skin and extremities, acute altered mental status, and oliguria. The end result of multiorgan hypoperfusion is tissue hypoxia, often with accompanying lactic acidosis. Because the MAP is the product of cardiac output and systemic vascular resistance (SVR), reductions in blood pressure can be caused by decreases in cardiac output and/or SVR. Accordingly, once shock is contemplated, the initial evaluation of a hypotensive patient should include an early bedside assessment of the adequacy of cardiac output **(Fig. 311-2)**. Clinical evidence of *diminished* cardiac output includes a narrow pulse pressure (systolic blood pressure minus diastolic blood pressure)—a marker that correlates with stroke

TABLE 311-2 Calculation of Acute Physiology and Chronic Health Evaluation II (APACHE II) Score[a]

Acute Physiology Score

SCORE	+4	+3	+2	+1	+0	+1	+2	+3	+4
Rectal temperature (°C)	≥41	39.0–40.9		38.5–38.9	36.0–38.4	34.0–35.9	32.0–33.9	30.0–31.9	≤29.9
Mean blood pressure (mmHg)	≥160	130–159	110–129		70–109		50–69		≤49
Heart rate (beats/min)	≥180	140–179	110–139		70–109		55–69	40–54	≤39
Respiratory rate (breaths/min)	≥50	35–49		25–34	12–24	10–11	6–9		≤5
Arterial pH	≥7.70	7.60–7.69		7.50–7.59	7.33–7.49		7.25–7.32	7.15–7.24	<7.15
Oxygenation									
If FIO_2 >0.5, use $(A - a) DO_2$	≥500	350–499	200–349		<200				
If FIO_2 ≤0.5, use PaO_2					>70	61–70		55–60	<55
Serum sodium (meq/L)	≥180	160–179	155–159	150–154	130–149		120–129	111–119	≤110
Serum potassium (meq/L)	≥7.0	6.0–6.9		5.5–5.9	3.5–5.4	3.0–3.4	2.5–2.9		<2.5
Serum creatinine (mg/dL)	≥3.5	2.0–3.4	1.5–1.9		0.6–1.4		<0.6		
Hematocrit (%)	≥60		50–59.9	46–49.9	30–45.9		20–29.9		<20
WBC count (10^3/mL)	≥40		20–39.9	15–19.9	3–14.9		1–2.9		<1

Glasgow Coma Score[b,c]

EYE OPENING	VERBAL (NONINTUBATED)	VERBAL (INTUBATED)	MOTOR ACTIVITY
4—Spontaneous	5—Oriented and talks	5—Seems able to talk	6—Verbal command
3—Verbal stimuli	4—Disoriented and talks	3—Questionable ability to talk	5—Localizes to pain
2—Painful stimuli	3—Inappropriate words	1—Generally unresponsive	4—Withdraws from pain
1—No response	2—Incomprehensible sounds		3—Decorticate
	1—No response		2—Decerebrate
			1—No response

Points Assigned to Age and Chronic Disease

AGE, YEARS	SCORE
<45	0
45–54	2
55–64	3
65–74	5
≥75	6

CHRONIC HEALTH (HISTORY OF CHRONIC CONDITIONS)[d]	SCORE
None	0
If patient is admitted after elective surgery	2
If patient is admitted after emergency surgery or for reason other than after elective surgery	5

[a] The APACHE II score is the sum of the acute physiology score (vital signs, oxygenation, laboratory values), the Glasgow coma score, age, and chronic health points. The worst values during the first 24 h in the ICU should be used. For serum creatinine, double the point score for acute renal failure. [b] Glasgow coma score (GCS) = eye-opening score + verbal (intubated or nonintubated) score + motor score. [c] For GCS component of acute physiology score, subtract GCS from 15 to obtain points assigned. [d] Hepatic: cirrhosis with portal hypertension or encephalopathy; cardiovascular: class IV angina (at rest or with minimal self-care activities); pulmonary: chronic hypoxemia or hypercapnia, polycythemia, ventilator dependence; renal: chronic peritoneal or hemodialysis; immune: immunocompromised host.

Abbreviations: $(A - a) DO_2$, alveolar-arterial oxygen difference; FIO_2, fraction of inspired oxygen; PaO_2, partial pressure of oxygen; WBC, white blood cell count.

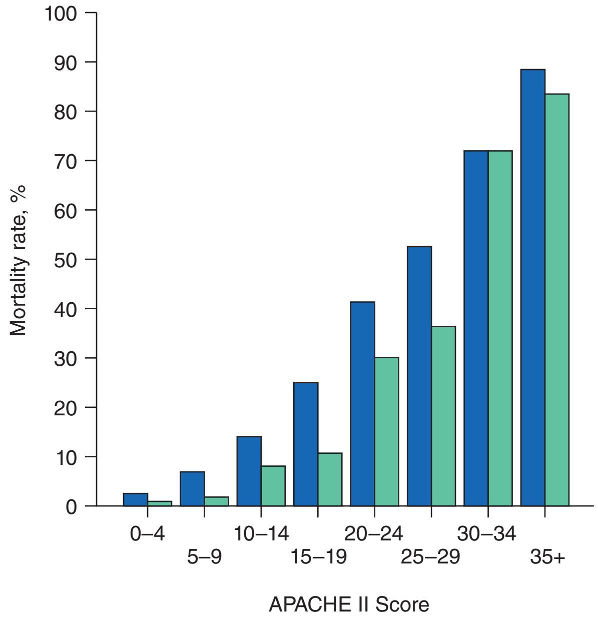

FIGURE 311-1 APACHE II survival curve. *Blue*, nonoperative; *green*, postoperative.

volume—and cool extremities with delayed capillary refill, colloquially termed "cold shock." It is important to palpate proximal extremities (e.g., thigh region) rather than distal extremities to determine relative "coolness," because patients with peripheral vascular disease may always have cool distal extremities. Signs of *increased* cardiac output include a widened pulse pressure (particularly with a reduced diastolic pressure), warm extremities with bounding pulses, and rapid capillary refill, colloquially termed "warm shock." If a hypotensive patient has clinical signs of increased cardiac output, it can be inferred that the reduced blood pressure is from decreased SVR.

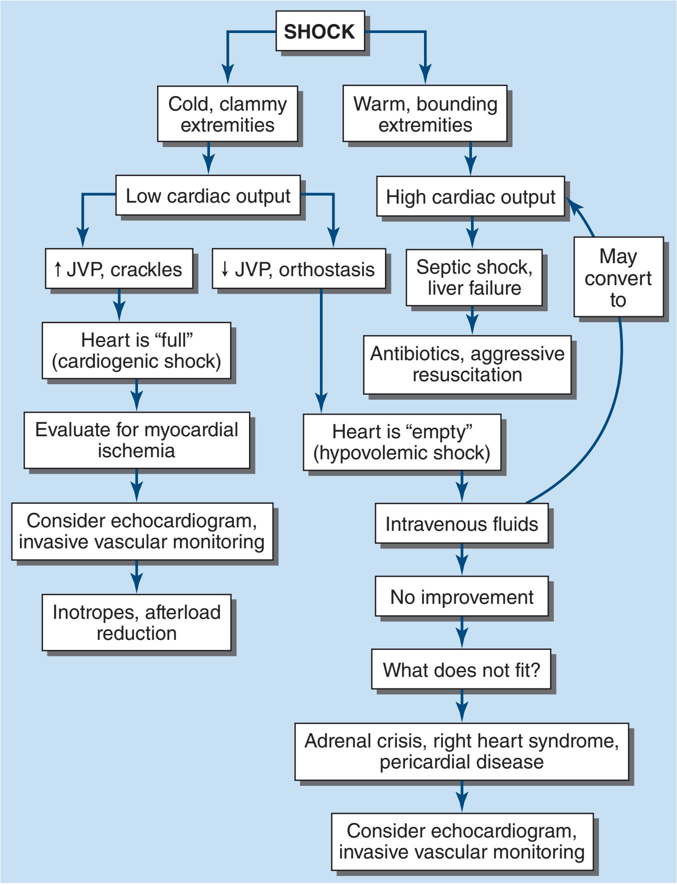
SHOCK
Cold, clammy extremities
Warm, bounding extremities
Low cardiac output
High cardiac output
↑ JVP, crackles
↓ JVP, orthostasis
Septic shock, liver failure
May convert to
Heart is "full" (cardiogenic shock)
Antibiotics, aggressive resuscitation
Evaluate for myocardial ischemia
Heart is "empty" (hypovolemic shock)
Consider echocardiogram, invasive vascular monitoring
Intravenous fluids
Inotropes, afterload reduction
No improvement
What does not fit?
Adrenal crisis, right heart syndrome, pericardial disease
Consider echocardiogram, invasive vascular monitoring

FIGURE 311-2 Approach to the patient in shock. JVP, jugular venous pressure.

In hypotensive patients with signs of reduced cardiac output, an assessment of intravascular volume status is appropriate. A hypotensive patient with decreased intravascular volume status may have a history suggesting hemorrhage or other volume losses (e.g., vomiting, diarrhea, polyuria). Although evidence of a reduced jugular venous pressure (JVP) is often sought, static measures of right atrial pressure do not predict fluid responsiveness reliably; the *change* in right atrial pressure as a function of spontaneous respiration is a better predictor of fluid responsiveness **(Fig. 311-3)**. Patients with fluid-responsive (i.e., hypovolemic) shock also may manifest large changes in pulse pressure as a function of respiration *during* mechanical ventilation **(Fig. 311-4)**. Other bedside metrics can help with judging whether a patient remains fluid-responsive, including responses to volume challenge or a straight leg raise (that increases venous return) that correlate with improved perfusion. Such tools include judging changes in JVP or central venous oxygen saturation, assessing changes in pulse pressure variation, determining changes in inferior vena cava collapse by ultrasound, and examining changes in left ventricular stroke volume using echocardiography. None of these measurements has been shown to be independently correlative, but a combination of these assessments with clinical judgment can help determine whether a patient remains volume-responsive. A hypotensive patient with increased intravascular volume and cardiac dysfunction may have S_3 and/or S_4 gallops on examination, increased JVP, extremity edema, and crackles on lung auscultation. The chest x-ray may show cardiomegaly, widening of the vascular pedicle, Kerley B lines, and pulmonary edema. There is increasing use of ultrasonography in place of chest radiography in resource-limited settings. Chest pain and electrocardiographic changes consistent with ischemia may be noted **(Chap. 316)**.

In hypotensive patients with clinical evidence of increased cardiac output, a search for causes of decreased SVR is appropriate. These patients usually require targeted initial volume resuscitation (as described above), after an initial fluid bolus, to achieve euvolemia, and often require vasopressors to maintain vascular tone. The most common cause of high-cardiac-output hypotension is sepsis **(Chap. 315)**. Patients with suspected sepsis should receive early broad-spectrum, appropriately dosed antibiotics and source control when feasible. Other causes of high-cardiac output hypotension include liver failure, severe pancreatitis, adrenal insufficiency, burns, trauma, anaphylaxis, thyrotoxicosis, and peripheral arteriovenous shunts.

Insertion of lines for monitoring and caring for critically ill patients may be necessary. Over the last two decades, management of shock has improved to the point where not all patients will require central venous and arterial lines, and recent sepsis guidelines support use of peripheral lines for pressors administration, if needed, while central access is being obtained. However, if a patient demonstrates that shock is not quickly resolving, as indicated by a persistent need for vasopressors and/or repeated measurement of the JVP and/or central venous O_2 saturation, then insertion of an arterial line for monitoring blood pressures and arterial blood gases, as well as a central venous line for administration of vasoactive agents and monitoring of the JVP and/or central venous O_2 saturation, may be required. Ideally, lines should be inserted under sterile conditions using a protocolized checklist approach, and lines should be removed as soon as they are no longer necessary to avoid risk of line-associated infection.

In summary, the most common categories of shock are hypovolemic, cardiogenic, and high-cardiac-output with decreased SVR (high-output hypotension). Certainly, more than one category can occur simultaneously (e.g., hypovolemic and septic shock). It may often be the case that an initial presentation with septic shock can present a cardiac strain, especially in patients with underlying heart dysfunction, such that later cardiac insufficiency may arise.

The initial assessment of a patient in shock should take only a few minutes. It is important that aggressive targeted resuscitation is instituted on the basis of the initial assessment, particularly since early resuscitation from septic and cardiogenic shock may improve survival (see below). If the initial bedside assessment yields equivocal or confounding data, more objective assessments such as ultrasound/echocardiography may be useful as described above. In spontaneously

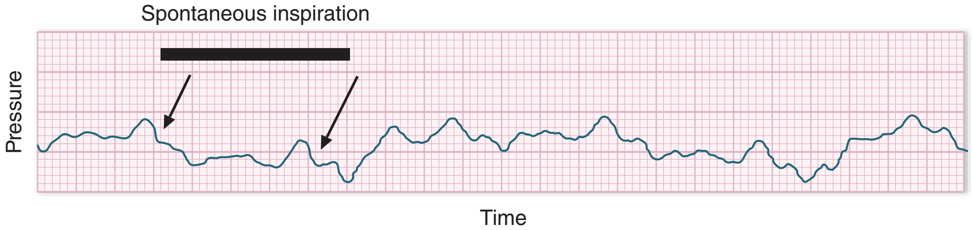

FIGURE 311-3 Right atrial pressure change during spontaneous respiration in a patient with shock whose cardiac output will increase in response to intravenous fluid administration. The right atrial pressure decreases from 7 mmHg to 4 mmHg. The horizontal bar marks the time of spontaneous inspiration.

breathing patients, inferior vena cava collapse seen on ultrasound may predict a fluid-responsive state. Increasingly, ultrasound of the thorax and abdomen is used by intensivists as an extension of the physical examination to assess rapidly imputed filling volumes, adequacy of cardiac performance, and for indices of other specific conditions (e.g., pericardial tamponade, pulmonary embolus, pulmonary edema, pneumothorax). The goal of aggressive resuscitation is to reestablish adequate tissue perfusion and thus to prevent or minimize end-organ injury. It is equally important not to over-resuscitate patients, as it is increasingly appreciated that excess fluid resuscitation is likely not beneficial. Thus, targeted fluid resuscitation is the goal.

■ MECHANICAL VENTILATORY SUPPORT (SEE ALSO CHAP. 313)

During the initial resuscitation of patients in shock, principles of advanced cardiac life support should be followed. An early assessment of the ability of a patient to protect their airway and to maintain adequate gas exchange is mandatory. Early intubation and mechanical ventilation often are required. Reasons for the institution of endotracheal intubation and mechanical ventilation include acute hypoxemic respiratory failure and ventilatory failure, which frequently accompany shock. Acute hypoxemic respiratory failure may occur in patients with cardiogenic shock and pulmonary edema **(Chap. 316)** as well as in those who are in septic shock with pneumonia or acute respiratory distress syndrome (ARDS) **(Chaps. 204, 312, and 315)**. Ventilatory failure often occurs as a consequence of an increased load on the respiratory system in the form of acute metabolic (often lactic) or respiratory acidosis, or decreased lung compliance (e.g., from pulmonary edema or pneumonia). Inadequate perfusion to respiratory muscles in the setting of shock may be another reason for early intubation and mechanical ventilation. Normally, the respiratory muscles receive a very small percentage of the cardiac output. However, in patients who are in shock with respiratory distress, the percentage of cardiac output dedicated to respiratory muscles may increase by 10-fold or more. Lactic acid production from inefficient respiratory muscle activity can present an additional ventilatory load.

Mechanical ventilation may relieve the work of breathing and allow redistribution of a limited cardiac output to other vital organs. Patients demonstrate respiratory distress by an inability to speak full sentences, accessory use of respiratory muscles, paradoxical abdominal muscle activity, extreme tachypnea (>40 breaths/min), and decreasing respiratory rate despite an increasing drive to breathe. When patients with shock are supported with mechanical ventilation, a major goal is for the ventilator to initially assume all or the majority of the work of breathing, facilitating a state of minimal respiratory muscle work. With the institution of mechanical ventilation for shock, further declines in MAP are frequently seen. The reasons include impeded venous return from positive-pressure ventilation, reduced endogenous catecholamine secretion once the stress associated with respiratory failure abates, and the actions of drugs used to facilitate endotracheal intubation (e.g., propofol, opiates). Patients with right heart dysfunction or preexisting pulmonary hypertension may also have diminished cardiac output related to the increases in right ventricular afterload resulting from positive-pressure ventilation. Accordingly, hypotension should be anticipated during and following endotracheal intubation. Because many of these patients may be fluid-responsive, IV volume administration should be considered, and vasopressor support peri-intubation may also be necessary. Figure 311-2 summarizes the diagnosis and treatment of different types of shock. **For further discussion of individual forms of shock, see Chaps. 314, 315, and 316.**

RESPIRATORY FAILURE

Respiratory failure is one of the most common reasons for ICU admission. In some ICUs, ≥75% of patients require mechanical ventilation during their stay. Respiratory failure can be categorized mechanistically on the basis of pathophysiologic derangements in respiratory function.

■ TYPE I: ACUTE HYPOXEMIC RESPIRATORY FAILURE

This type of respiratory failure occurs with alveolar flooding and subsequent ventilation-perfusion mismatch and intrapulmonary shunt physiology. Alveolar flooding may be a consequence of pulmonary edema, lung injury, pneumonia, or alveolar hemorrhage. Pulmonary edema can be further categorized as occurring due to elevated pulmonary microvascular pressures, as seen in heart failure and intravascular volume overload or ARDS ("low-pressure pulmonary edema," **Chap. 312**). This syndrome is defined by acute onset (≤1 week) of bilateral opacities on chest imaging that are not fully explained by cardiac failure or fluid overload and often includes ventilation-perfusion mismatch and shunt physiology requiring positive end-expiratory pressure (PEEP). A new global definition for ARDS has been proposed that does not rely upon chest radiography, arterial blood gases, or use of ventilatory support with PEEP. Type I respiratory failure occurs in clinical settings such as sepsis, gastric aspiration, pneumonia, COVID-19 **(Chap. 205)**, near-drowning, multiple blood transfusions, and pancreatitis. The mortality rate among patients with ARDS was traditionally very high (50–70%), although changes in patient care have led to mortality rates closer to 30% (see below). The COVID-19 pandemic resulted in a substantially increased incidence of viral-mediated ARDS.

It is well established that mechanical ventilation of patients with ARDS may propagate lung injury. As seen in **Fig. 311-5**, the pressure-volume relationship of the lung in ARDS is not linear. Alveoli may collapse at very low lung volumes. Animal studies have suggested that repeated stretching and overdistention of injured alveoli during mechanical ventilation can further injure the lung. Concern over this alveolar overdistention, termed *ventilator-induced "volutrauma,"* led to a multicenter, randomized, prospective trial comparing traditional ventilator strategies for ARDS (large tidal volume: 12 mL/kg of ideal body weight) with a low tidal volume (6 mL/kg of ideal body weight). This study showed a dramatic reduction in mortality rate in the low-tidal-volume group from that in the high-tidal-volume group (31 vs 39.8%). Other studies have suggested that large tidal volumes may lead to ARDS in patients who initially do

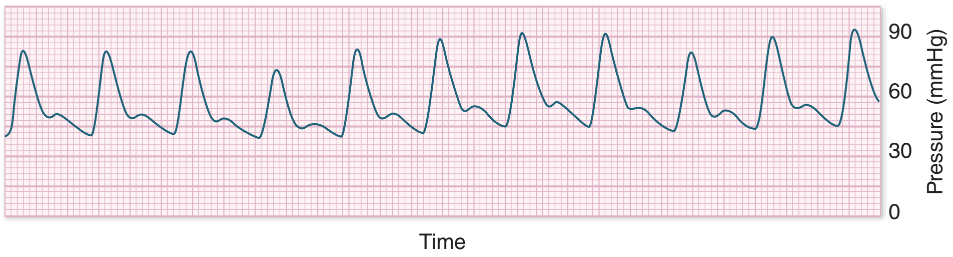

FIGURE 311-4 Pulse pressure change during mechanical ventilation in a patient with shock whose cardiac output will increase in response to intravenous fluid administration. The pulse pressure (systolic minus diastolic blood pressure) changes during mechanical ventilation in a patient with septic shock.

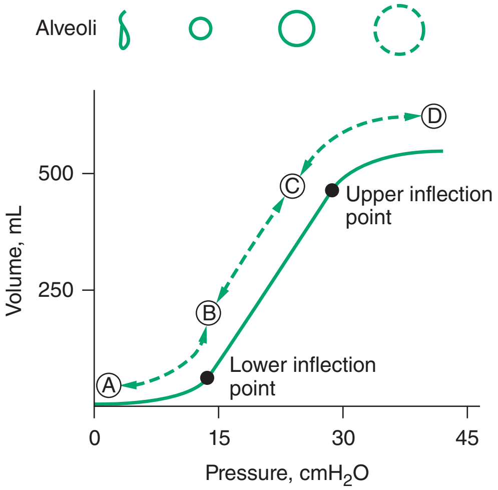

FIGURE 311-5 Pressure-volume relationship in the lungs of a patient with acute respiratory distress syndrome (ARDS). At the lower inflection point, collapsed alveoli begin to open and lung compliance changes. At the upper deflection point, alveoli become overdistended. The shape and size of alveoli are illustrated at the top of the figure.

not have this problem. Prone positioning has been shown to improve survival in those with severe ARDS and has been more broadly applied in many centers in COVID-19 ARDS. Select patients may benefit from neuromuscular blockade in ARDS. In addition, a "fluid-conservative" management strategy (maintaining a low central venous pressure [CVP] or pulmonary capillary wedge pressure [PCWP]) is associated with fewer days of mechanical ventilation than a "fluid-liberal" strategy (maintaining a relatively high CVP or PCWP) in ARDS in those patients who have been resuscitated from shock. There is growing interest in avoiding intubation in patients with ARDS by the use of a variety of devices, such as masks, high-flow oxygen delivery systems, and helmets for respiratory support that were more broadly applied during the COVID pandemic and likely more so in resource-limited settings when feasible; however, this must be balanced by concern that higher tidal volumes during spontaneous breathing with these devices could result in progression of preexisting lung injury.

■ TYPE II RESPIRATORY FAILURE: HYPERCAPNEIC RESPIRATORY FAILURE

This type of respiratory failure is a consequence of alveolar hypoventilation and results from the inability to eliminate carbon dioxide effectively. Mechanisms are categorized by impaired central nervous system (CNS) drive to breathe (colloquially termed, "won't breathe"), impaired strength with failure of neuromuscular function in the respiratory system, and increased load(s) on the respiratory system (with the latter two colloquially termed, "can't breathe"). Reasons for diminished CNS drive to breathe include drug overdose, brainstem injury, sleep-disordered breathing, and severe hypothyroidism. Reduced strength can be due to impaired neuromuscular transmission (e.g., myasthenia gravis, Guillain-Barré syndrome, amyotrophic lateral sclerosis) or respiratory muscle weakness (e.g., myopathy, electrolyte derangements, fatigue).

The overall load on the respiratory system can be subclassified into resistive loads (e.g., bronchospasm), loads due to reduced lung compliance (e.g., alveolar edema, atelectasis, intrinsic positive end-expiratory pressure [auto-PEEP]—see below), loads due to reduced chest wall compliance (e.g., pneumothorax, pleural effusion, abdominal distention), and loads due to increased minute ventilation requirements (e.g., pulmonary embolus with increased dead-space fraction, sepsis).

The mainstays of therapy for hypercapnic respiratory failure are directed at reversing the underlying cause(s) of ventilatory failure. Noninvasive positive-pressure ventilation with a tight-fitting facial or nasal mask, with avoidance of endotracheal intubation, may stabilize these patients in certain circumstances. This approach has been shown to be beneficial in treating patients with exacerbations of chronic obstructive pulmonary disease; it has been tested less extensively in other kinds of respiratory failure but may be attempted nonetheless with close monitoring in the absence of contraindications (e.g., hemodynamic instability, inability to protect the airway, respiratory arrest, significant airway secretions, significant aspiration risk).

■ TYPE III RESPIRATORY FAILURE: LUNG ATELECTASIS

This form of respiratory failure results from lung atelectasis. Because atelectasis occurs so commonly in the perioperative period, this form is also called *perioperative respiratory failure*. After general anesthesia, decreases in functional residual capacity lead to collapse of dependent lung units. Such atelectasis can be treated by frequent changes in position, chest physiotherapy, upright positioning, and control of incisional and/or abdominal pain. Noninvasive positive-pressure ventilation may also be used to reverse regional atelectasis.

■ TYPE IV RESPIRATORY FAILURE: METABOLIC DEMANDS

This form most often results from hypoperfusion of respiratory muscles in patients in shock. Normally, respiratory muscles consume <5% of total cardiac output and oxygen delivery. Patients in shock often experience respiratory distress due to pulmonary edema (e.g., in cardiogenic shock), lactic acidosis, and anemia. In this setting, up to 40% of cardiac output may be distributed to the respiratory muscles. Intubation and mechanical ventilation can allow redistribution of the cardiac output away from the respiratory muscles and back to vital organs while the shock is treated. In addition, other causes of significant metabolic acidosis might require ventilatory support while reversal of the underlying cause of the acidosis is addressed.

CARE OF THE MECHANICALLY VENTILATED PATIENT

Mechanically ventilated patients frequently require sedatives and analgesics. Opiates are the mainstay of therapy for analgesia in mechanically ventilated patients. After adequate pain control has been ensured, additional indications for sedation include anxiolysis; treatment of subjective dyspnea; reduction of autonomic hyperactivity, which may precipitate myocardial ischemia; and reduction of total O_2 consumption (Vo_2). Nonbenzodiazepine sedatives are preferred because benzodiazepines are associated with increased delirium and worse patient outcomes.

The neuromuscular blocking agent cisatracurium is occasionally used to facilitate mechanical ventilation in patients with profound ventilator dyssynchrony despite optimal sedation, particularly in the setting of severe ARDS. Use of these agents may result in prolonged weakness—a myopathy known as the *postparalytic syndrome*. For this reason, neuromuscular blocking agents typically are used as a last resort when aggressive sedation fails to achieve patient–ventilator synchrony. Because neuromuscular blocking agents result in pharmacologic paralysis without altering mental status, sedative-induced amnesia is mandatory when these agents are administered.

Amnesia can be achieved reliably with propofol and benzodiazepines such as lorazepam and midazolam. Outside the setting of pharmacologic paralysis, few data support the idea that amnesia is mandatory in all patients who require intubation and mechanical ventilation. Because many of these critically ill patients have impaired hepatic and renal function, sedatives and opiates may accumulate when given for prolonged periods. A nursing protocol–driven approach to sedation of mechanically ventilated patients or daily interruption of sedative infusions paired with daily spontaneous breathing trials has been shown to prevent excessive drug accumulation and shorten the duration of both mechanical ventilation and ICU stay (see below).

(See also Chap. 313.) Whereas a thorough understanding of the pathophysiology of respiratory failure is essential for optimal patient care, recognition of a patient's readiness to be liberated from mechanical ventilation is likewise important. Several studies have shown that daily spontaneous breathing trials can identify patients who are ready for extubation. Accordingly, all intubated, mechanically ventilated patients should undergo daily screening of respiratory function.

If oxygenation is stable (i.e., $Pa{O_2}/FI{O_2}$ [partial pressure of oxygen/fraction of inspired oxygen] >200 and PEEP ≤5 cmH_2O), cough and airway reflexes are intact, and no vasopressor agents or sedatives are being administered, the patient has passed the screening test and should undergo a spontaneous breathing trial (SBT). If sedatives are being administered, the patient can undergo a spontaneous awakening trial (SAT), as well, to determine if they are able to maintain adequate alertness and respiratory status without sedatives. The SAT/SBT trial consists of a period of breathing through the endotracheal tube without significant ventilator support (continuous positive airway pressure [CPAP] of 5 cmH_2O with or without low-level pressure support [e.g., 5 cmH_2O] and an open T-piece breathing system have all been validated) for 30–120 min. The spontaneous breathing trial is declared a failure and stopped if *any* of the following occur: (1) respiratory rate >35/min for >5 min, (2) O_2 saturation <90%, (3) heart rate >140/min or a 20% increase or decrease from baseline, (4) systolic blood pressure <90 mmHg or >180 mmHg, or (5) increased anxiety or diaphoresis. If, at the end of the spontaneous breathing trial, none of the above events has occurred, the patient can be considered for an extubation trial. Such protocol-driven approaches to patient care can have an important impact on the duration of mechanical ventilation and ICU stay. Despite such a careful approach to liberation from mechanical ventilation, up to 10% of patients develop respiratory distress after extubation and may require resumption of mechanical ventilation. Many of these patients will require reintubation. The use of noninvasive ventilation as a rescue strategy in patients in whom extubation fails may be associated in some patients with worse outcomes than are obtained with immediate reintubation. Some studies suggest that there are subgroups of patients who might benefit from administration of noninvasive ventilation or high-flow nasal oxygen therapy upon extubation, as it is believed that low levels of PEEP and/or inspiratory flow delivered by these devices after extubation may be helpful.

MULTIORGAN SYSTEM FAILURE

Multiorgan system failure, which is commonly associated with critical illness, is defined by the simultaneous presence of physiologic dysfunction and/or failure of two or more organs. Typically, this syndrome occurs in the setting of severe sepsis, shock of any kind, severe inflammatory conditions such as pancreatitis, and trauma. The fact that multiorgan system failure occurs commonly in the ICU is a testament to our current ability to stabilize and support single-organ failure. The ability to support single-organ failure aggressively (e.g., by mechanical ventilation or by renal replacement therapy) has reduced rates of early mortality in critical illness. As a result, it is less common for critically ill patients to die in the initial stages of resuscitation. Instead, many patients succumb to critical illness later in the ICU stay, after the initial presenting problem may have been stabilized.

Although there is debate regarding specific definitions of organ failure, several general principles governing the syndrome of multiorgan system failure apply. First, organ failure, no matter how it is defined, must persist beyond 24 h. Second, mortality risk increases with the accrual of failing organs. Third, the prognosis worsens with increased duration of organ failure. These observations remain true across various critical care settings (e.g., medical vs surgical).

MONITORING IN THE ICU

Because respiratory failure and circulatory failure are common in critically ill patients, monitoring of the respiratory and cardiovascular systems is undertaken frequently. Evaluation of respiratory gas exchange is routine in critical illness. The "gold standard" remains arterial blood-gas analysis, in which pH, $Pa{O_2}$, partial pressure of carbon dioxide ($P{CO_2}$), and O_2 saturation are measured directly. With arterial blood-gas analysis, the two main functions of the lung—oxygenation of arterial blood and elimination of CO_2—can be assessed directly. In fact, the arterial blood pH, which has a profound effect on the drive to breathe, can be assessed only by such sampling. Venous pH obtained through an indwelling central venous line can approximate arterial pH. Although sampling of arterial blood is generally safe and may be undertaken more frequently through insertion of a temporary indwelling arterial line, it may be painful and cannot provide continuous information. In light of these limitations, noninvasive monitoring of respiratory function is often employed.

PULSE OXIMETRY

The most commonly utilized noninvasive technique for monitoring respiratory function, pulse oximetry takes advantage of differences in the absorptive properties of oxygenated and deoxygenated hemoglobin. At wavelengths of 660 nm, oxyhemoglobin reflects light more effectively than does deoxyhemoglobin, whereas the reverse is true in the infrared spectrum (940 nm). A pulse oximeter passes both wavelengths of light through a perfused digit such as a finger, and the relative intensity of light transmission at these two wavelengths is recorded. From this information, the relative percentage of oxyhemoglobin is derived. Since arterial pulsations produce phasic changes in the intensity of transmitted light, the pulse oximeter is designed to detect only light of alternating intensity. This feature allows distinction of arterial and venous blood O_2 saturations. Studies during the COVID pandemic noted that pulse oximetry overestimates oxygen saturation in patients with darker skin, thus making correlation with arterial $Pa{O_2}$ more imperative.

RESPIRATORY SYSTEM MECHANICS

Respiratory system mechanics can be measured in patients during mechanical ventilation **(Chap. 313)**. When volume-controlled modes of mechanical ventilation are used, accompanying airway pressures can easily be measured as long as the patient is breathing passively. The peak airway pressure is determined by two variables: airway resistance and respiratory system compliance. At the end of inspiration, inspiratory flow can be stopped transiently. This end-inspiratory pause (*plateau pressure*) is a static measurement, affected only by respiratory system compliance and not by airway resistance. Therefore, during volume-controlled ventilation, the difference between the peak (airway resistance + respiratory system compliance) and plateau (respiratory system compliance only) airway pressures provides a quantitative assessment of airway resistance. Accordingly, during volume-controlled ventilation, patients with increases in airway resistance typically have increased peak airway pressures as well as abnormally high gradients between peak and plateau airway pressures (typically >10–15 cmH_2O) at a constant inspiratory flow rate of 1 L/s. The compliance of the respiratory system is defined by the change in volume of the respiratory system per unit change in pressure; thus, a quantitative assessment of compliance is provided by the tidal volume divided by the plateau pressure minus PEEP.

The respiratory system can be divided into two components: the lungs and the chest wall. Normally, respiratory system compliance is ~100 mL/cmH_2O. Pathophysiologic processes such as pleural effusions, pneumothorax, and increased abdominal girth all reduce chest wall compliance. Lung compliance may be reduced by pneumonia, pulmonary edema, alveolar hemorrhage, interstitial lung disease, or auto-PEEP. Accordingly, patients with abnormalities in compliance of the respiratory system (lungs and/or chest wall) typically have elevated peak *and* plateau airway pressures but a normal gradient between these two pressures. Auto-PEEP occurs when there is insufficient time for emptying of alveoli before the next inspiratory cycle. Because the alveoli have not decompressed completely, alveolar pressure remains positive at the end of exhalation (*functional residual capacity*). This phenomenon results most commonly from obstruction of distal airways in disease processes such as asthma and COPD. Auto-PEEP with resulting alveolar overdistention may result in diminished lung compliance, reflected by abnormally increased plateau airway pressures. Modern mechanical ventilators allow breath-to-breath display of pressure and flow, permitting detection of potential problems such as patient–ventilator dyssynchrony, airflow obstruction, and auto-PEEP **(Fig. 311-6)**.

CIRCULATORY STATUS

Oxygen delivery ($Q{O_2}$) is a function of cardiac output and the content of O_2 in the arterial blood ($Ca{O_2}$). The $Ca{O_2}$ is determined by the

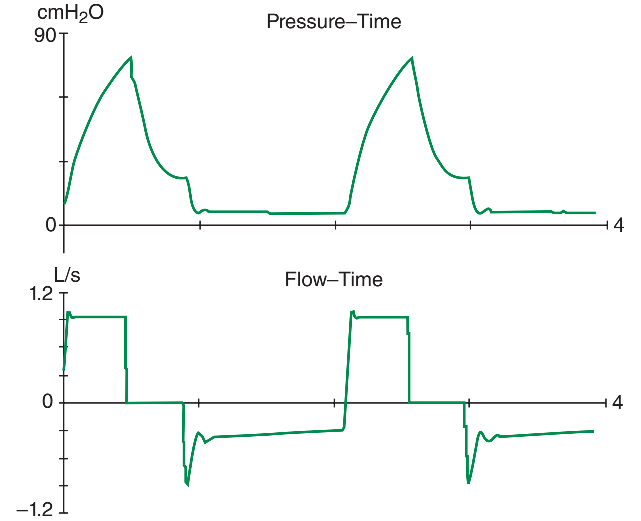

FIGURE 311-6 Increased airway resistance with auto–positive end-expiratory pressure (PEEP). The top waveform (airway pressure vs time) shows a large difference between the peak airway pressure (80 cmH_2O) and the plateau airway pressure (20 cmH_2O). The bottom waveform (flow vs time) demonstrates airflow throughout expiration (reflected by the flow tracing on the negative portion of the abscissa) that persists up to the next inspiratory effort.

hemoglobin concentration, the arterial hemoglobin saturation, and dissolved O_2 not bound to hemoglobin. For normal adults:

$$\begin{aligned} Qo_2 &= 50\text{ dL/min} \times (1.39 \times 15\text{ g/dL [hemoglobin concentration]} \\ &\quad \times 1.0\text{ [hemoglobin \% saturation]} + 0.0031 \times 10\text{ [Pao}_2\text{])} \\ &= 50\text{ dL/min (cardiac output)} \times 21.6\text{ mL } O_2 \text{ per dL blood (Cao}_2\text{)} \\ &= 1058\text{ mL } O_2 \text{ per min} \end{aligned}$$

It is apparent that nearly all the O_2 delivered to tissues is bound to hemoglobin and that the dissolved O_2 (Pao_2) contributes very little to O_2 content in arterial blood or to O_2 delivery. Normally, the content of O_2 in mixed venous blood ($C\bar{v}o_2$) is 15.76 mL/dL since the mixed venous blood is 75% saturated. Therefore, the normal tissue extraction ratio for O_2 is $Cao_2 - C\bar{v}o_2/Cao_2$ ([21.16 – 15.76]/21.16) or ~25%. A pulmonary artery catheter (see discussion below) allows measurements of O_2 delivery and the O_2 extraction ratio.

Information on the venous O_2 saturation allows assessment of global tissue perfusion. A reduced venous O_2 saturation may be caused by inadequate cardiac output, reduced hemoglobin concentration, and/or reduced arterial O_2 saturation. An abnormally high oxygen consumption (Vo_2) may also lead to a reduced venous O_2 saturation if O_2 delivery is not concomitantly increased. Abnormally increased Vo_2 in peripheral tissues may be caused by problems such as fever, agitation, shivering, and thyrotoxicosis.

The pulmonary artery catheter originally was designed as a tool to guide therapy for acute myocardial infarction but has been used in the ICU for evaluation and treatment of a variety of other conditions, such as ARDS, septic shock, congestive heart failure, and acute renal failure. This device has never been validated as a tool associated with reduction in morbidity and mortality rates. Indeed, despite numerous prospective studies, mortality or morbidity rate benefits associated with use of the pulmonary artery catheter have never been reported in any setting. Accordingly, it appears that routine pulmonary artery catheterization is not indicated as a means of monitoring and characterizing circulatory status in most critically ill patients, especially as monitoring of the venous O_2 saturation via an indwelling central venous line has proven helpful in many critical illness settings. However, there are still select circumstances where pulmonary artery catheterization may prove helpful when used by those with appropriate experience in its insertion and data interpretation.

PREVENTION OF COMPLICATIONS OF CRITICAL ILLNESS

SEPSIS IN THE CRITICAL CARE UNIT

(See also Chap. 315) Sepsis is defined as life-threatening organ dysfunction (i.e., an increase in SOFA of 2 points or more) caused by a dysregulated response to infection. Poor outcomes can be anticipated in patients with two or more of the following: respiratory rate ≥22 breaths/min, altered mentation, and systolic blood pressure ≤100 mmHg. Sepsis is a leading cause of death in noncoronary ICUs in the United States, with case rates expected to increase as the population ages and a higher percentage of people are vulnerable to infection.

NOSOCOMIAL INFECTIONS IN THE ICU

Many therapeutic interventions in the ICU are invasive and predispose patients to infectious complications. These interventions include endotracheal intubation, indwelling vascular catheters, transurethral bladder catheters, and other catheters placed into sterile body cavities (e.g., tube thoracostomy, percutaneous intraabdominal drainage catheterization). The longer such devices remain in place, the more prone to infections patients become from these devices. For example, ventilator-associated events such as ventilator-associated pneumonia correlate strongly with the duration of intubation and mechanical ventilation. Therefore, an important aspect of preventive care is the timely removal of invasive devices as soon as they are no longer needed. Moreover, multidrug-resistant organisms are commonplace in the ICU.

Infection control is critical in the ICU. Care bundles, which include measures such as frequent hand washing, are effective but underutilized strategies. Other components of care bundles, such as protective isolation of patients colonized or infected by drug-resistant organisms, are also commonly used. Studies evaluating multifaceted, evidence-based strategies to decrease catheter-related bloodstream infections have shown improved outcomes with strict adherence to measures such as hand washing, full-barrier precautions during catheter insertion, chlorhexidine skin preparation, avoidance of the femoral site, and timely catheter removal.

DEEP-VENOUS THROMBOSIS (DVT) (SEE ALSO CHAP. 290)

All ICU patients are at high risk for this complication because of their predilection for immobility. Therefore, all should receive some form of prophylaxis against DVT if feasible. The most commonly employed forms of prophylaxis are subcutaneous chemoprophylaxis (e.g., low-dose heparin) injections and sequential compression devices for the lower extremities. Observational studies report an alarming incidence of DVTs despite the use of these standard prophylactic regimens. Furthermore, heparin prophylaxis may result in heparin-induced thrombocytopenia, another nosocomial complication in critically ill patients.

Low-molecular-weight heparins such as enoxaparin are more effective than unfractionated heparin for DVT prophylaxis in high-risk patients (e.g., those undergoing orthopedic surgery) and are associated with a lower incidence of heparin-induced thrombocytopenia, although their use may be limited in patients with renal dysfunction given their renal clearance.

STRESS ULCERS

Prophylaxis against stress ulcers is not necessary for all ICU patients. It should only be administered to high-risk patients, such as those with coagulopathy or respiratory failure requiring mechanical ventilation. While there has been debate about the optimal agent for stress ulcer prophylaxis, a number of recent studies have supported improved efficacy of proton pump inhibitors (PPIs) in reducing bleeding risk compared with other agents (e.g., histamine-2 receptor antagonist [H2 blocker] or sucralfate). There exist concerns for increased risk of pneumonia and *Clostridium difficile* colitis with PPIs compared with other agents, although the data are not definitive, and the improved efficacy of PPIs in patients at high risk for stress ulcers may outweigh these potential infectious risks.

NUTRITION AND GLYCEMIC CONTROL

Nutrition and glycemic control are important issues that may be associated with respiratory failure, impaired wound healing, and dysfunctional immune response in critically ill patients. Early enteral feeding is reasonable, with some data suggesting that permissive underfeeding of nonprotein calories is not inferior to full-goal feeding. Certainly, enteral feeding, if possible, is preferred over parenteral nutrition, which is associated with numerous complications, including hyperglycemia, fatty liver, cholestasis, and sepsis. When parenteral feeding is necessary to supplement enteral nutrition, delaying this intervention until day 8 in the ICU results in better recovery and fewer ICU-related complications. Tight glucose control has been an area of controversy in critical care. Although one study showed a significant mortality benefit when glucose levels were aggressively normalized in a large group of surgical ICU patients, other studies of both medical and surgical ICU patients suggested that tight glucose control resulted in increased rates of mortality likely attributable, in part, to hypoglycemic episodes. Thus, current guidelines suggest targeting glucose levels of ≤180 mg/dL in critically ill patients, rather than targeting tighter control.

■ ICU-ACQUIRED WEAKNESS

ICU-acquired weakness occurs frequently in patients who survive critical illness. Both neuropathies and myopathies have been described, most commonly after ~1 week in the ICU. The mechanisms behind ICU-acquired weakness syndromes are poorly understood, and they are known to present with heterogeneous muscle pathophysiology. Very early physical and occupational therapy in mechanically ventilated patients reportedly results in significant improvements in functional independence at hospital discharge as well as in reduced durations of mechanical ventilation and delirium.

■ ANEMIA

Studies have shown that most ICU patients are anemic as a result of chronic inflammation. Phlebotomy also contributes to ICU anemia. A large multicenter study involving patients in many different ICU settings challenged the conventional notion that a hemoglobin level of 100 g/L (10 g/dL) is needed in critically ill patients, with similar outcomes noted in those whose transfusion trigger was 7 g/dL. Red blood cell transfusion is associated with impairment of immune function and increased risk of infections as well as of ARDS and volume overload, all of which may explain the findings in this study. A conservative transfusion strategy has shown similar outcomes in septic shock, postcardiac surgery, and post–hip surgery patients.

■ ACUTE KIDNEY FAILURE

(See also Chap. 321) Acute kidney failure occurs in a significant percentage of critically ill patients. The most common underlying etiology is acute tubular necrosis, usually precipitated by hypoperfusion and/or nephrotoxic agents. Currently, no pharmacologic agents are available for prevention of kidney injury in critical illness. Studies have shown convincingly that low-dose dopamine, fenoldopam, and vasopressin are *not* effective in protecting the kidneys from acute injury.

NEUROLOGIC DYSFUNCTION IN CRITICALLY ILL PATIENTS

■ DELIRIUM

(See also Chap. 29) Delirium is defined by (1) an acute onset of changes or fluctuations in mental status, (2) inattention, (3) disorganized thinking, and (4) an altered level of consciousness (i.e., a state other than alertness). Delirium is reported to occur in a wide range of mechanically ventilated ICU patients and can be detected by the Confusion Assessment Method for the ICU (CAM-ICU) or the Intensive Care Delirium Screening Checklist (ICDSC). These tools are used to ask patients to answer simple questions and perform simple tasks and can be used readily at the bedside. The differential diagnosis of delirium in ICU patients is broad and includes infectious etiologies (including sepsis), medications (particularly sedatives and analgesics), drug withdrawal, metabolic/electrolyte derangements, intracranial pathology (e.g., stroke, intracranial hemorrhage), seizures, hypoxia, hypertensive crisis, shock, and vitamin deficiencies (particularly thiamine). The etiology of a patient's ICU delirium impacts the prognosis. Those with persistent ICU delirium not related to sedatives have increases in length of hospital stay, time on mechanical ventilation, cognitive impairment at hospital discharge, and 6-month mortality rate. Interventions to reduce ICU delirium are limited. The sedative dexmedetomidine has been less strongly associated with ICU delirium than midazolam in some studies. In addition, very early physical and occupational therapy in mechanically ventilated patients has been demonstrated to reduce delirium.

■ ANOXIC CEREBRAL INJURY

(See also Chap. 318) This condition is common after cardiac arrest and often results in severe and permanent brain injury in survivors. Active cooling of patients to 33°C after cardiac arrest is controversial, with some studies showing improved neurologic outcomes and others showing no such improvement when compared to maintaining normothermia. Certainly, patients post cardiac arrest should have a temperature targeted to no higher than normothermia.

■ STROKE

(See also Chap. 437) Stroke is a common cause of neurologic critical illness. Hypertension must be managed carefully, because abrupt reductions in blood pressure may be associated with further brain ischemia and injury. Acute ischemic stroke treated with tissue plasminogen activator (tPA) has an improved neurologic outcome when treatment is given within 4.5 h of onset of symptoms, with likely increased benefit associated with earlier administration. The mortality rate is not reduced when tPA is compared with placebo, despite the improved neurologic outcome. The risk of cerebral hemorrhage is significantly higher in patients given tPA. No consistent overall benefit is seen when tPA therapy is given beyond 4.5 h after symptom onset. Heparin has not been convincingly shown to improve outcomes in patients with acute ischemic stroke. Decompressive craniectomy is a surgical procedure that relieves increased intracranial pressure in the setting of space-occupying brain lesions or brain swelling from stroke; available evidence suggests that this procedure may improve survival among select patients (e.g., ≤55 years of age), albeit at a cost of increased disability for some.

■ SUBARACHNOID HEMORRHAGE

(See also Chap. 437) Subarachnoid hemorrhage may occur secondary to aneurysm rupture and is often complicated by cerebral vasospasm, re-bleeding, and hydrocephalus. Vasospasm can be detected by either transcranial Doppler assessment or cerebral angiography; it is typically treated with the calcium channel blocker nimodipine, aggressive IV fluid administration to avoid hypovolemia, and therapy aimed at maintaining adequate central perfusion pressure, typically with vasoactive drugs such as phenylephrine. IV fluids and vasoactive drugs (hypertensive hypervolemic therapy) are used to overcome the cerebral vasospasm. Early surgical clipping or endovascular coiling of aneurysms is advocated to prevent complications related to re-bleeding. Hydrocephalus, typically heralded by a decreased level of consciousness, may require ventriculostomy drainage.

■ STATUS EPILEPTICUS (SEE ALSO CHAP. 436)

Recurrent or relentless seizure activity is a medical emergency. Cessation of seizure activity is required to prevent irreversible neurologic injury. Lorazepam is the most effective benzodiazepine for treating status epilepticus and is the treatment of choice for controlling seizures acutely. Maintenance of seizure control should be effected with a loading dose of fosphenytoin, valproate, or levetiracetam, as these agents have been shown to have similar efficacy and side effects.

■ BRAIN DEATH

(See also Chap. 318) Although deaths of critically ill patients usually are attributable to irreversible cessation of circulatory and respiratory function, a diagnosis of death also may be established by irreversible cessation of all functions of the entire brain, including the brainstem, even if circulatory and respiratory functions remain intact on artificial life support. Such a diagnosis requires demonstration of the absence of cerebral function (no response to any external stimulus) and brainstem functions (e.g., unreactive pupils, lack of ocular movement in response to head turning or ice-water irrigation of ear canals, positive apnea test [no drive to breathe]). Many U.S. institutions have a protocol based upon their state's requirements for declaration of brain death. Absence of brain function must have an established cause and be permanent without possibility of recovery; a sedative effect, hypothermia, hypoxemia, neuromuscular paralysis, and severe hypotension must be ruled out. If there is uncertainty about the cause of coma, studies of cerebral blood flow and electroencephalography should be performed.

WITHHOLDING OR WITHDRAWING CARE

(See also Chap. 13) Withholding or withdrawal of care occurs commonly in the ICU setting. The Task Force on Ethics of the Society of Critical Care Medicine reported that it is ethically sound to withhold or withdraw care if a patient or the patient's surrogate makes such a request or if the physician judges that the goals of therapy are not achievable. Because all medical treatments are justified by their expected benefits, the loss of such an expectation justifies the act of withdrawing or withholding such treatment; these two actions are judged to be fundamentally similar. An underlying stipulation derived from this report is that an informed patient should have their wishes respected with regard to life-sustaining therapy. Implicit in this stipulation is the need to ensure that patients are thoroughly and accurately informed regarding the plausibility and expected results of various therapies.

The act of informing patients and/or surrogate decision-makers is the responsibility of the physician and other health care providers. If a patient or surrogate desires therapy deemed futile by the treating physician, the physician is not obligated ethically to provide such treatment. Rather, arrangements may be made to transfer the patient's care to another care provider. Whether the decision to withdraw life support should be initiated by the physician or left to surrogate decision-makers alone is not clear. One study reported that slightly more than half of surrogate decision-makers preferred to receive such a recommendation, whereas the rest did not. Critical care providers should meet regularly with patients and/or surrogates to discuss prognosis when the withholding or withdrawal of care is being considered. After a consensus among caregivers has been reached, this information should be relayed to the patient and/or surrogate decision-maker. If a decision to withhold or withdraw life-sustaining care for a patient has been made, aggressive attention to analgesia and anxiolysis is needed. Often, an independent hospital ethics service can be of benefit in navigating complex decision-making.

Acknowledgment

John P. Kress and Jesse B. Hall contributed to this chapter in the 20th edition and some material from that chapter has been retained here.

FURTHER READING

Andersen-Ranberg NC et al: Haloperidol for the treatment of delirium in ICU patients. N Engl J Med 387:2425, 2022.

Evans L et al: Surviving Sepsis Campaign: International guidelines for management of sepsis and septic shock. Crit Care Med 49:e1063, 2021.

Girard TD et al: An official American Thoracic Society/American College of Chest Physicians Clinical Practice Guideline: Liberation from mechanical ventilation in critically ill adults. Rehabilitation protocols, ventilator liberation protocols, and cuff leak tests. Am J Respir Crit Care Med 195:120, 2017.

Guerin C et al: Prone positioning in severe acute respiratory distress syndrome. N Engl J Med 368:2159, 2013.

Kapur J et al: Randomized trial of three anticonvulsant medications for status epilepticus. N Engl J Med 381:2103, 2019.

Man S et al: Association between thrombolytic door-to-needle time and 1-year mortality and readmission in patients with acute ischemic stroke. JAMA 323:2170, 2020.

Matthay MA et al: A new global definition of acute respiratory distress syndrome. Am J Respir Crit Care Med 209:37, 2024.

National Heart, Lung, and Blood Institute Petal Clinical Trials Network et al: Early neuromuscular blockade in the acute respiratory distress syndrome. N Engl J Med 380:1997, 2019.

Singer M et al: The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 315:801, 2016.

Toews I et al: Interventions for preventing upper gastrointestinal bleeding in people admitted to intensive care units. Cochrane Database Syst Rev 6:CD008687, 2018.

312

Acute Respiratory Distress Syndrome

Rebecca M. Baron, Bruce D. Levy

Acute respiratory distress syndrome (ARDS) is a clinical syndrome of severe dyspnea of rapid onset, hypoxemia, and diffuse pulmonary infiltrates leading to respiratory failure. ARDS can be caused by diffuse lung injury from many underlying medical and surgical disorders. The lung injury may be direct (e.g., toxic inhalation) or indirect (e.g., sepsis) **(Table 312-1)**. The clinical features of ARDS are listed in **Table 312-2**. By expert consensus, ARDS was defined in 2012 by three categories based on the degrees of hypoxemia (Table 312-2). These stages of mild, moderate, and severe ARDS are associated with mortality risk and with the duration of mechanical ventilation in survivors. A more recent global definition of ARDS has been proposed that does not rely upon arterial blood gases, chest radiography, or use of positive end-expiratory pressure on invasive or noninvasive mechanical ventilation, recognizing the challenge of resource-poor settings and increasing use of high-flow nasal oxygen and noninvasive means of respiratory support.

The annual incidence of ARDS prior to the COVID-19 pandemic was estimated to be as high as 60 cases per 100,000 population. Approximately 10% of all intensive care unit (ICU) admissions involve patients with ARDS. This chapter will focus on non–COVID-19 ARDS. Please see **Chap. 205** for more information on COVID.

ETIOLOGY

While many medical and surgical illnesses have been associated with the development of ARDS, most cases (>80%) are caused by a relatively small number of clinical disorders: pneumonia and sepsis (~40–60%), followed in incidence by aspiration of gastric contents, trauma, multiple transfusions, and drug overdose. Among patients with trauma, the most frequently reported surgical conditions in ARDS are pulmonary contusion, multiple bone fractures, and chest wall trauma/flail chest, whereas head trauma, near-drowning, toxic inhalation, and burns are more rare causes. The risks of developing ARDS are increased in patients with more than one predisposing medical or surgical condition.

Several other clinical variables have been associated with the development of ARDS. These include older age, chronic alcohol abuse, pancreatitis, pneumonia and sepsis (40–60%, including pandemic COVID pneumonia and other respiratory viruses), and severity of critical illness. Trauma patients with an Acute Physiology and Chronic Health Evaluation (APACHE) II score ≥16 **(Chap. 311)** have a 2.5-fold increased risk of developing ARDS.

CLINICAL COURSE AND PATHOPHYSIOLOGY

The natural history of ARDS is marked by three phases—exudative, proliferative, and fibrotic—that each have characteristic clinical and pathologic features **(Fig. 312-1)**.

TABLE 312-1 Clinical Disorders Commonly Associated with ARDS

DIRECT LUNG INJURY	INDIRECT LUNG INJURY
Pneumonia	Sepsis
Aspiration of gastric contents	Severe trauma
Pulmonary contusion	Multiple bone fractures
Near-drowning	Flail chest
Toxic inhalation injury	Head trauma
	Burns
	Multiple transfusions
	Drug overdose
	Pancreatitis
	Postcardiopulmonary bypass

TABLE 312-2 Diagnostic Criteria for ARDS Based on 2012 Berlin Criteria

SEVERITY: OXYGENATION[a]	ONSET	CHEST RADIOGRAPH[b]	ABSENCE OF LEFT ATRIAL HYPERTENSION
Mild: 200 mmHg < PaO_2/FIO_2 ≤300 mmHg *Moderate:* 100 mmHg < PaO_2/FIO_2 ≤200 mmHg *Severe:* PaO_2/FIO_2 ≤100 mmHg	Acute: Within 1 week of a clinical insult or new or worsening respiratory symptoms	Bilateral opacities consistent with pulmonary edema not fully explained by effusions, lobar/lung collapse, or nodules	Hydrostatic edema is not the primary cause of respiratory failure. If no ARDS risk factor is present, then some objective evaluation is required (e.g., echocardiography) to rule out hydrostatic edema

[a] As assessed on at least 5 cmH_2O of positive end-expiratory pressure (PEEP). 2023 proposed updates to the Berlin criteria include: (1) consideration of ARDS in nonintubated patients using PaO_2/FIO_2 ≤300 mmHg or SpO_2/FIO_2 (S/F) ≤315 mmHg (if SpO_2 ≤97% as measured by pulse oximetry) on high-flow nasal oxygen with flow of ≥30 L/min or noninvasive ventilation/continuous positive airway pressure with at least 5 cmH_2O of PEEP; (2) addition of S/F ratios to consideration of ARDS in intubated patients (mild: 235 < SpO_2/FIO_2 ≤ 315 mmHg; moderate: 148 < SpO_2/FIO_2 ≤ 235 mmHg; severe: SpO_2/FIO_2 ≤148 mmHg [if SpO_2 ≤97%]); and (3) consideration of ARDS in resource-limited settings if SpO_2/FIO_2 ≤ 315 mmHg (if SpO_2 ≤97%) without requirement for PEEP, minimum oxygen flow rate, or specific respiratory support devices. It should be noted that pulse oximetry may overestimate the oxygen saturation in patients with darker skin tones, such that correlation with PaO_2 should be considered when feasible. [b] The 2023 proposed update to the Berlin criteria permits ultrasound as an alternative imaging modality, especially in resource-limited settings.

Abbreviations: ARDS, acute respiratory distress syndrome; FIO_2, inspired O_2 percentage; PaO_2, arterial partial pressure of O_2, SpO_2, peripheral saturation of O_2.

Exudative Phase In this phase, alveolar capillary endothelial cells and type I pneumocytes (alveolar epithelial cells) are injured, with consequent loss of the normally tight alveolar barrier to fluid and macromolecules. Edema fluid that is rich in protein accumulates in the interstitial and alveolar spaces (Fig. 312-2). Proinflammatory cytokines (e.g., interleukin 1, interleukin 6, interleukin 8, and tumor necrosis factor α [TNF-α]) and lipid mediators (e.g., leukotriene B_4) are increased in this acute phase, leading to the recruitment of leukocytes (especially neutrophils) into the pulmonary interstitium and alveoli. In addition, condensed plasma proteins aggregate in the air spaces with cellular debris and dysfunctional pulmonary surfactant to form hyaline membrane whorls. Pulmonary vascular injury also occurs early in ARDS, with vascular obliteration by microthrombi and fibrocellular proliferation (Fig. 312-3).

Alveolar edema often predominantly involves *dependent* portions of the lung with diminished aeration. Collapse of large sections of dependent lung can contribute to decreased lung compliance. Consequently, intrapulmonary shunting and hypoxemia develop and the work of breathing increases, leading to dyspnea. The pathophysiologic alterations in alveolar spaces are exacerbated by microvascular occlusion that results in reductions in pulmonary arterial blood flow to ventilated portions of the lung (and thus in increased dead space and pulmonary vascular resistance) and in pulmonary hypertension. Thus, in addition to severe hypoxemia, hypercapnia secondary to an increase in pulmonary dead space can be prominent in ARDS.

The exudative phase usually encompasses the first 7 days of illness after exposure to a precipitating ARDS risk factor, with the patient experiencing the onset of respiratory symptoms. Although usually presenting within 12–36 h after the initial insult, symptoms can be delayed by 5–7 days. Dyspnea develops, with a sensation of rapid shallow breathing and an inability to get enough air. Tachypnea and increased work of breathing result frequently in respiratory fatigue and ultimately in respiratory failure. Laboratory values are generally nonspecific and are primarily indicative of underlying clinical disorders. The chest radiograph usually reveals opacities consistent with pulmonary edema and often involves at least three-quarters of the lung fields (Fig. 312-2). While characteristic for ARDS, these radiographic findings are not specific and can be indistinguishable from cardiogenic pulmonary edema (Chap. 316). Unlike the latter, however, the chest x-ray in ARDS may not demonstrate cardiomegaly, pleural effusions, or pulmonary vascular redistribution as is often present in pure cardiogenic pulmonary edema. If no ARDS risk factor is present, then some objective evaluation is required (e.g., echocardiography) to exclude a cardiac etiology for hydrostatic edema. Chest computed tomography (CT) in ARDS also reveals the presence of bilateral pulmonary infiltrates and demonstrates extensive heterogeneity of lung involvement (Fig. 312-4).

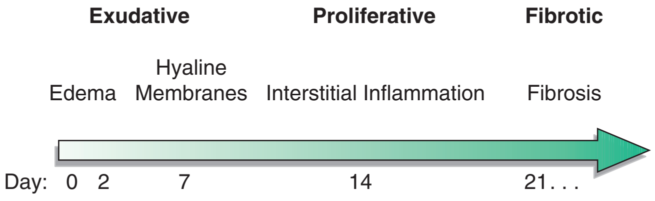

FIGURE 312-1 Diagram illustrating the time course for the development and resolution of acute respiratory distress syndrome (ARDS). The exudative phase is notable for early alveolar edema and neutrophil-rich leukocytic infiltration of the lungs, with subsequent formation of hyaline membranes from diffuse alveolar damage. Within 7 days, a proliferative phase ensues with prominent interstitial inflammation and early fibrotic changes. Approximately 3 weeks after the initial pulmonary injury, most patients recover. However, some patients enter the fibrotic phase, with substantial fibrosis and bullae formation.

Because the early features of ARDS are nonspecific, alternative diagnoses must be considered, although it is possible that there can be coexisting conditions with ARDS. In the differential diagnosis of ARDS, the most common disorders are cardiogenic pulmonary edema, bilateral pneumonia, and alveolar hemorrhage. Less common diagnoses to consider include acute interstitial lung diseases (e.g., acute interstitial pneumonitis; Chap. 304), acute immunologic injury (e.g., hypersensitivity pneumonitis; Chap. 299), toxin injury (e.g., radiation pneumonitis; Chap. 80), and neurogenic pulmonary edema (Chap. 39).

Proliferative Phase This phase of ARDS usually lasts from approximately day 7 to day 21. Many patients recover rapidly during this phase. Despite this improvement, many patients still experience dyspnea, tachypnea, and hypoxemia. Some patients develop progressive lung injury and early changes of pulmonary fibrosis during the proliferative phase. Histologically, the first signs of resolution are often

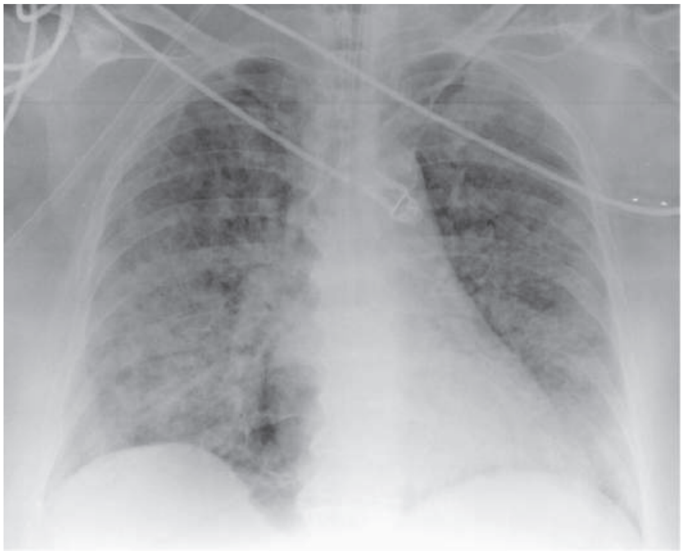

FIGURE 312-2 A representative anteroposterior chest x-ray in the exudative phase of acute respiratory distress syndrome (ARDS) shows bilateral opacities consistent with pulmonary edema that can be difficult to distinguish from left ventricular failure.

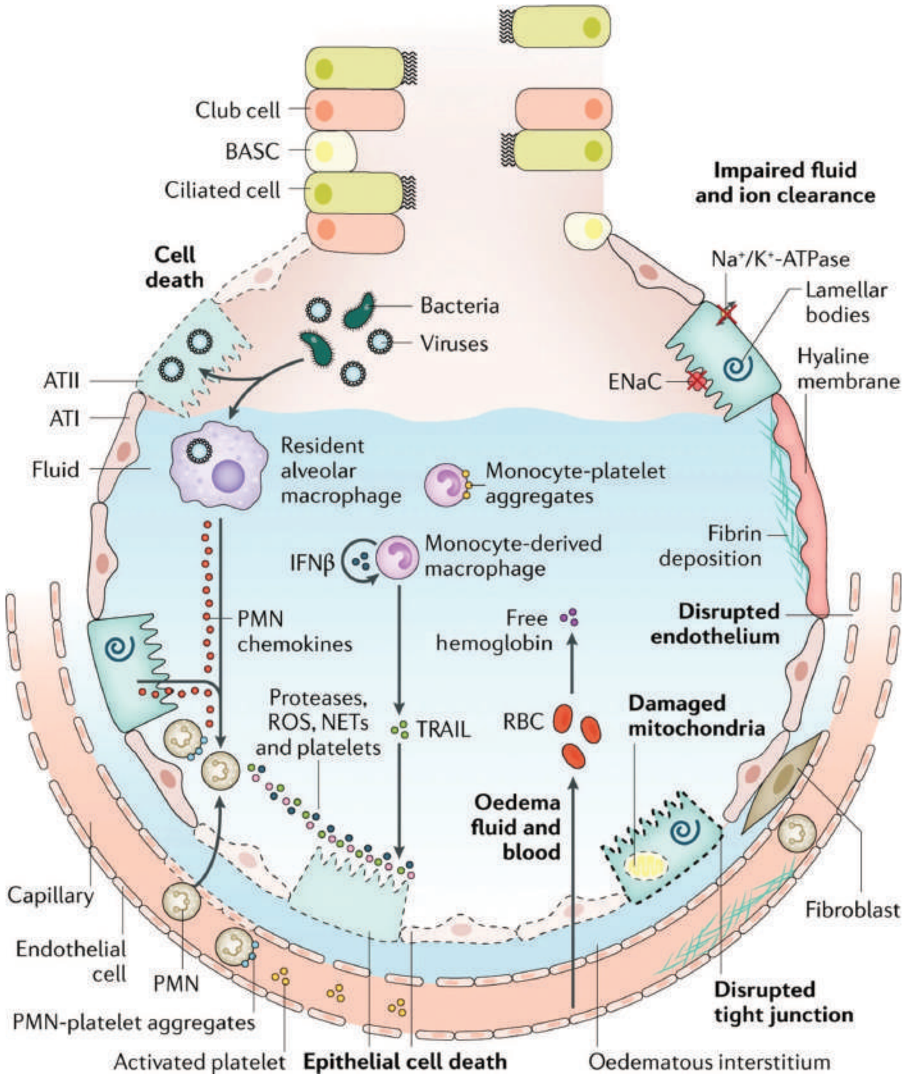
Club cell
BASC
Ciliated cell
Cell death
Bacteria
Viruses
ATII
ATI
Fluid
Resident alveolar macrophage
Monocyte-platelet aggregates
IFNβ
Monocyte-derived macrophage
PMN chemokines
Proteases, ROS, NETs and platelets
TRAIL
RBC
Free hemoglobin
Oedema fluid and blood
Impaired fluid and ion clearance
Na⁺/K⁺-ATPase
Lamellar bodies
ENaC
Hyaline membrane
Fibrin deposition
Disrupted endothelium
Damaged mitochondria
Fibroblast
Disrupted tight junction
Capillary
Endothelial cell
PMN
PMN-platelet aggregates
Activated platelet
Epithelial cell death
Oedematous interstitium

FIGURE 312-3 The injured alveolus in the acute phase of acute lung injury and the acute respiratory distress syndrome. A variety of insults (e.g., bacteria, viruses) can injure the epithelium, and this direct injury is propagated by subsequent activation of downstream pathways. Activation of Toll-like receptors (not shown) on alveolar type II (ATII) epithelial cells and resident macrophages induces the secretion of chemokines, which recruit circulating immune cells into the airspaces. As neutrophils migrate across the epithelium, they release toxic mediators, including proteases, reactive oxygen species (ROS) and neutrophil extracellular traps (NETs), which have an important role in host defense but also can exacerbate endothelial and epithelial injury. Monocytes also migrate into the lung and can cause injury, including epithelial cell apoptosis via interferon (IFN)-β-dependent release of tumor necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL), which activates death receptors. Activated platelets form aggregates with polymorphonuclear (PMN) leukocytes (which are involved in NET formation) and monocytes. Red blood cells (RBCs) release cell-free hemoglobin, which exacerbates injury via oxidant-dependent mechanisms. Epithelial injury also includes injury to the plasma membrane, which can be induced by bacterial pore-forming toxins or mechanical stretch (often related to mechanical ventilation), and mitochondrial dysfunction. Together, these and other effects result in endothelial and epithelial permeability, which further facilitate the transmigration of leukocytes and lead to the influx of edematous fluid and RBCs. Airspace filling with edematous fluid causes hypoxemia, often resulting in the need for mechanical ventilation. Vascular injury and alveolar edema can contribute to decreased ability to excrete CO_2 (hypercapnia), accounting for increased pulmonary dead space in acute respiratory distress syndrome. In turn, hypoxemia and hypercapnia impair sodium transport, reducing alveolar edema clearance. ATI, alveolar type I cell; BASC, bronchioalveolar stem cell; ENaC, epithelial sodium channel. *(Reproduced with permission from MA Matthay et al: Acute respiratory distress syndrome. Nat Rev Dis Primers 5:18, 2019.)*

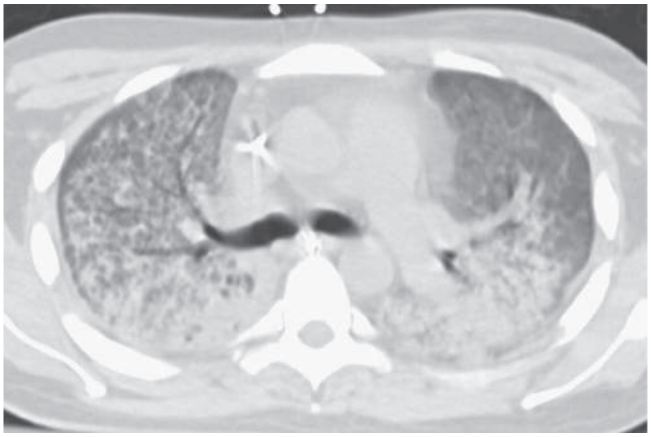

FIGURE 312-4 A representative CT scan of the chest during the exudative phase of acute respiratory distress syndrome (ARDS), in which *dependent* alveolar edema and atelectasis predominate.

evident in this phase, with the initiation of lung repair, the organization of alveolar exudates, and a shift from neutrophil- to lymphocyte-predominant pulmonary infiltrates. As part of the reparative process, type II pneumocytes proliferate along alveolar basement membranes. These specialized epithelial cells synthesize new pulmonary surfactant and differentiate into type I pneumocytes.

Fibrotic Phase While many patients with ARDS recover lung function 3–4 weeks after the initial pulmonary injury, some enter a fibrotic phase that may require long-term support on mechanical ventilators and/or supplemental oxygen. Histologically, the alveolar edema and inflammatory exudates of earlier phases convert to extensive alveolar-duct and interstitial fibrosis. Marked disruption of acinar architecture leads to emphysema-like changes, with large bullae. Intimal fibroproliferation in the pulmonary microcirculation causes progressive vascular occlusion and pulmonary hypertension. The physiologic consequences include an increased risk of pneumothorax, reductions in lung compliance, and increased pulmonary dead space. Patients in this late phase experience a substantial burden of excess morbidity. Lung biopsy evidence for pulmonary fibrosis in any phase of ARDS is associated with increased mortality risk.

TREATMENT

Acute Respiratory Distress Syndrome

GENERAL PRINCIPLES

Recent reductions in ARDS mortality rates are largely the result of general advances in the care of critically ill patients **(Chap. 311)**. Thus, caring for these patients requires close attention to (1) the recognition and treatment of underlying medical and surgical disorders (e.g., pneumonia, sepsis, aspiration, trauma); (2) the minimization of unnecessary procedures and their complications; (3) standardized "bundled care" approaches for ICU patients, including prophylaxis against venous thromboembolism, gastrointestinal bleeding, aspiration, excessive sedation, prolonged mechanical ventilation, and central venous catheter infections; (4) prompt recognition of nosocomial infections; and (5) provision of adequate nutrition via the enteral route when feasible.

MANAGEMENT OF MECHANICAL VENTILATION

(See also Chap. 313) Patients meeting clinical criteria for ARDS frequently become fatigued from increased work of breathing and progressive hypoxemia, requiring mechanical ventilation for support, although increasing use of high-flow nasal oxygen and noninvasive ventilation has enabled some patients to avoid mechanical ventilation.

Minimizing Ventilator-Induced Lung Injury Despite its life-saving potential, mechanical ventilation can aggravate lung injury from high tidal volumes. Experimental models have demonstrated that ventilator-induced lung injury can arise from at least two principal mechanisms: "volutrauma" from repeated alveolar overdistention from excess tidal volume (that might also coincide with increased alveolar pressures, or "barotrauma") and "atelectrauma" from recurrent alveolar collapse. As is evident from chest CT (Fig. 312-4), ARDS is a heterogeneous disorder, often principally involving dependent portions of the lung with relative sparing of other regions. Because compliance differs in affected versus more "normal" areas of the lung, attempts to fully inflate the consolidated lung may lead to overdistention of and injury to the more normal areas. Ventilator-induced injury can be demonstrated in experimental models of acute lung injury, in particular with high-tidal-volume (V_T) ventilation.

A large-scale, randomized controlled trial sponsored by the National Institutes of Health and conducted by the ARDS Network compared low V_T ventilation (6 mL/kg of predicted body weight) to conventional V_T ventilation (12 mL/kg predicted body weight). Lower airway pressures were also targeted in the low-tidal-volume group (i.e., plateau pressure measured on the ventilator after a 0.5-s pause after inspiration), with pressures targeted at ≤30 cmH$_2$O in the low-tidal-volume group versus ≤50 cmH$_2$O in the high-tidal-volume group. The mortality rate was significantly lower in the low V_T patients (31%) than in the conventional V_T patients (40%). This improvement in survival represents a substantial ARDS mortality benefit.

Minimizing Atelectrauma by Prevention of Alveolar Collapse In ARDS, the presence of alveolar and interstitial fluid and the loss of surfactant can lead to a marked reduction of lung compliance. Without an increase in end-expiratory pressure, significant alveolar collapse can occur at end-expiration, with consequent impairment of oxygenation. In most clinical settings, positive end-expiratory pressure (PEEP) is adjusted to minimize FIO_2 (inspired O_2 percentage) and provide adequate PaO_2 (arterial partial pressure of O_2) without causing alveolar overdistention. It should be noted that high-flow nasal cannula may provide low levels of PEEP. Currently, there is no consensus on the optimal method to set PEEP on the ventilator because numerous trials have proved inconclusive. Possible approaches include using the table of PEEP-FIO_2 combinations from the ARDS Network trial group, generating a static pressure-volume curve for the respiratory system and setting PEEP just above the lower inflection point on this curve to maximize respiratory system compliance, and measuring esophageal pressures to estimate transpulmonary pressure (which may be particularly helpful in patients with a stiff chest wall). Of note, a recent phase 2 trial in patients with moderate-to-severe ARDS demonstrated no benefit of routine use of esophageal pressure-guided PEEP titration over empirical high PEEP-FIO_2 titration. Until more data become available on how best to optimize PEEP settings in ARDS, clinicians can use these options or a practical approach to empirically measure "best PEEP" at the bedside to determine the optimal settings that best promote alveolar recruitment, minimize alveolar overdistention and hemodynamic instability, and provide adequate PaO_2 while minimizing FIO_2 **(Chap. 313)**.

Prone Positioning While several prior trials demonstrated that mechanical ventilation in the prone position improved arterial oxygenation without a mortality benefit, a 2013 trial demonstrated a significant reduction in 28-day mortality with prone positioning (32.8 to 16.0%) for patients with severe ARDS ($\text{PaO}_2/\text{FIO}_2$ <150 mmHg) early in their course of illness. Thus, many centers increased the use of prone positioning in severe ARDS, especially during the COVID pandemic, with the understanding that this maneuver requires a critical care team that is experienced in "proning," as repositioning critically ill patients can be hazardous, leading to accidental endotracheal extubation, loss of central venous catheters, and orthopedic injury.

OTHER STRATEGIES IN MECHANICAL VENTILATION

Recruitment maneuvers that transiently increase PEEP to high levels to "recruit" atelectatic lung can increase oxygenation, but a mortality benefit has not been established, and in fact, recruitment maneuvers were shown to increase mortality when combined with higher baseline PEEP settings. *Alternate modes of mechanical ventilation*, such as airway pressure release ventilation and high-frequency oscillatory ventilation, have not been proven beneficial over standard modes of ventilation in ARDS management. *In one study, lung-replacement therapy with extracorporeal membrane oxygenation* (ECMO) was shown to improve mortality for patients with ARDS in the United Kingdom who were referred to an ECMO center (though only 75% of referred patients received ECMO) and thus may have utility in select adult patients with severe ARDS as a rescue therapy. A subsequent study demonstrated that initial use of ECMO in patients with severe ARDS was not superior to use of ECMO as a rescue strategy for patients who failed standard ARDS management.

FLUID MANAGEMENT

(See also Chap. 311) Increased pulmonary vascular permeability leading to interstitial and alveolar edema fluid rich in protein is a central feature of ARDS. In addition, impaired vascular integrity augments the normal increase in extravascular lung water that occurs with increasing left atrial pressure. Maintaining a low left atrial filling pressure minimizes pulmonary edema and prevents further decrements in arterial oxygenation and lung compliance; improves pulmonary mechanics; and shortens ICU stay and the duration of mechanical ventilation. Thus, aggressive attempts to reduce left atrial filling pressures with fluid restriction and diuretics should be an important aspect of ARDS management, limited only by hypotension and hypoperfusion of critical organs such as the kidneys.

NEUROMUSCULAR BLOCKADE

In severe ARDS, sedation alone can be inadequate for the patient-ventilator synchrony required for lung-protective ventilation. In a multicenter, randomized, placebo-controlled trial of early neuromuscular blockade (with cisatracurium besylate) for 48 h, patients with severe ARDS had increased survival and ventilator-free days without increasing ICU-acquired paresis. A subsequent trial demonstrated no mortality benefit for early neuromuscular blockade for

48 h in patients with moderate-to-severe ARDS. This more recent study supports the notion that selective use of neuromuscular blockade might be beneficial in those ARDS patients with ventilatory dyssynchrony despite sedation.

GLUCOCORTICOIDS

Many attempts have been made to treat both early and late ARDS with glucocorticoids, with the goal of reducing potentially deleterious pulmonary inflammation. Few studies have shown any significant mortality benefit. Current evidence does *not* support the routine use of glucocorticoids in the care of ARDS patients. More recent guidelines have supported the use of low-dose hydrocortisone (200 mg over 24 h) in sepsis patients with refractory hypotension and in patients with severe community-acquired pneumonia, which are conditions that often coexist with ARDS.

OTHER THERAPIES

Clinical trials of surfactant replacement and multiple other medical therapies have proved disappointing. Pulmonary vasodilators such as inhaled nitric oxide and inhaled epoprostenol sodium can transiently improve oxygenation in some patients but have not been shown to improve survival or decrease time on mechanical ventilation.

RECOMMENDATIONS

Many clinical trials have been undertaken to improve the outcome of patients with ARDS; most have been unsuccessful in modifying the natural history. While results of large clinical trials must be judiciously applied to *individual* patients, evidence-based recommendations are summarized in **Table 312-3**, and an algorithm for the initial therapeutic goals and limits in ARDS management is provided in **Fig. 312-5**. Please note that these recommendations apply to non–COVID-19 ARDS. Please see recommendations for COVID-19 ARDS in **Chap. 205**.

TABLE 312-3 Evidence-Based Recommendations for ARDS Therapies

TREATMENT	RECOMMENDATION[a]
Mechanical ventilation	
Low tidal volume	A
Minimized left atrial filling pressures	B
High-PEEP or "open lung"	B[b]
Prone position	B[b]
Recruitment maneuvers	C[b]
High-frequency ventilation	D
ECMO	B[b]
Early neuromuscular blockade (routine use)	C[b]
Glucocorticoid treatment	D[c]
Inhaled vasodilators (e.g., inhaled NO, inhaled epoprostenol)	C
Surfactant replacement, and other anti-inflammatory therapy (e.g., ketoconazole, PGE_1, NSAIDs)	D

[a] Key: A, recommended therapy based on strong clinical evidence from randomized clinical trials; B, recommended therapy based on supportive but limited clinical data; C, recommended only as alternative therapy on the basis of indeterminate evidence; D, not recommended on the basis of clinical evidence against efficacy of therapy. [b] As described in the text, there is no consensus on optimal PEEP setting in ARDS, but general consensus supports an open lung strategy that minimizes alveolar distention with some studies favoring the high PEEP-FiO_2 ARDS table; prone positioning was shown to improve mortality in severe ARDS in one randomized controlled clinical trial; recruitment maneuvers combined with high PEEP were shown to increase mortality in one study; ECMO may be beneficial in select patients with severe ARDS; early neuromuscular blockade demonstrated a mortality benefit in one randomized controlled trial in patients with severe ARDS but was not replicated in a subsequent study, suggesting routine use of early neuromuscular blockade in all subjects with moderate-severe ARDS may not be beneficial. [c] While there are no direct supportive data for use of glucocorticoids in ARDS, there are recent data supporting the consideration of low-dose hydrocortisone in two conditions frequently encountered in patients with ARDS: (1) the 2021 Surviving Sepsis guidelines recommend consideration of low-dose hydrocortisone in patients with refractory hypotension due to septic shock, and (2) a recent study demonstrated a mortality benefit of low-dose hydrocortisone in patients with severe community-acquired pneumonia.

Abbreviations: ARDS, acute respiratory distress syndrome; ECMO, extracorporeal membrane oxygenation; NO, nitric oxide; NSAIDs, nonsteroidal anti-inflammatory drugs; PEEP, positive end-expiratory pressure; PGE_1, prostaglandin E_1.

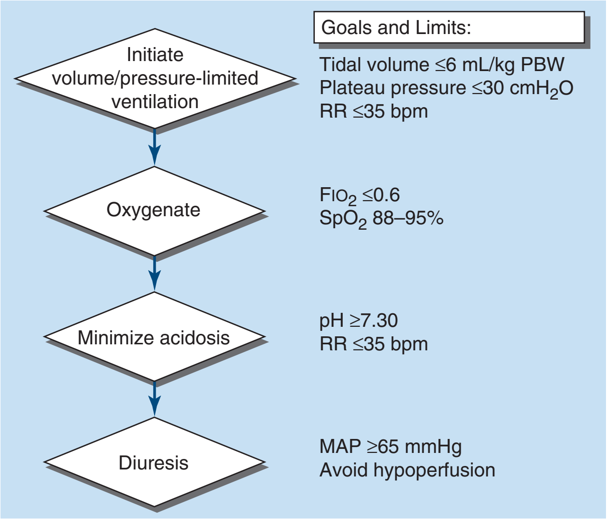

FIGURE 312-5 Algorithm for the initial management of acute respiratory distress syndrome (ARDS). Clinical trials have provided evidence-based therapeutic goals for a stepwise approach to the early mechanical ventilation, oxygenation, and correction of acidosis and diuresis of critically ill patients with ARDS. FiO_2, inspired O_2 percentage; MAP, mean arterial pressure; PBW, predicted body weight; RR, respiratory rate; SpO_2, arterial oxyhemoglobin saturation measured by pulse oximetry.

PROGNOSIS

Mortality In the Large Observational Study to Understand the Global Impact of Severe Acute Respiratory Failure (LUNG SAFE) trial, hospital mortality estimates for ARDS were 34.9% for mild ARDS, 40.3% for moderate ARDS, and 46.1% for severe ARDS. There is substantial variability, but a trend toward improved ARDS outcomes over time appears evident. Of interest, mortality in ARDS is largely attributable to nonpulmonary causes, with sepsis and nonpulmonary organ failure accounting for >80% of deaths. Thus, improvement in survival is likely secondary to advances in the care of septic/infected patients and those with multiple organ failure **(Chap. 311)**.

The major risk factors for ARDS mortality are nonpulmonary. Advanced age is an important risk factor. Patients aged >75 years have a substantially higher mortality risk (~60%) than those <45 (~20%). Moreover, patients >60 years of age with ARDS and sepsis have a threefold higher mortality risk than those <60 years of age. Other risk factors include preexisting organ dysfunction from chronic medical illness—in particular, chronic liver disease, chronic alcohol abuse, and chronic immunosuppression **(Chap. 311)**. Patients with ARDS arising from direct lung injury (including pneumonia, pulmonary contusion, and aspiration; Table 312-1) are nearly twice as likely to die as those with indirect causes of lung injury, while surgical and trauma patients with ARDS—especially those without direct lung injury—generally have a higher survival rate than other ARDS patients.

Increasing severity of ARDS, as defined by the consensus Berlin definition, predicts increased mortality. Surprisingly, there is little additional value in predicting ARDS mortality from other parameters of lung injury, including the level of PEEP (≥10 cmH_2O), respiratory system compliance (≤40 $\text{mL/cmH}_2\text{O}$), the extent of alveolar infiltrates on chest radiography, and the corrected expired volume per minute (≥10 L/min) (as a surrogate measure of dead space).

Functional Recovery in ARDS Survivors While it is common for patients with ARDS to experience prolonged respiratory failure and remain dependent on mechanical ventilation for survival, it is a testament to the resolving powers of the lung that the majority of patients who survive regain nearly normal lung function. Patients usually recover maximal lung function within 6 months. One year after endotracheal extubation, more than one-third of ARDS survivors have normal spirometry values and diffusion capacity. Most of the remaining patients have only mild abnormalities in pulmonary function. Unlike mortality risk, recovery of lung function is strongly associated with the extent of lung injury in early ARDS. Low static respiratory compliance, high levels of required PEEP, longer durations of mechanical ventilation, and high lung injury scores are all associated with less recovery of pulmonary function. Of note, when physical function is assessed 5 years after ARDS, exercise limitation and decreased physical quality of life are often documented despite normal or nearly normal pulmonary function. When caring for ARDS survivors, it is important to be aware of the potential for a substantial burden of psychological problems in patients and family caregivers, including significant rates of depression and posttraumatic stress disorder. Investigations into sequelae of COVID ARDS have provided additional insight into long-term ICU outcomes. Please see **Chap. 205** for information regarding COVID prognosis and recovery.

■ FURTHER READING

ARDS Definition Task Force: Acute respiratory distress syndrome: The Berlin definition. JAMA 307:2526, 2012.

ARDS Network: Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. N Engl J Med 342:1301, 2000.

Bellani G et al: Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. JAMA 315:788, 2016.

Dequin P-F et al: Hydrocortisone in severe community-acquired pneumonia. N Engl J Med 388:1931, 2023.

Gorman EA et al: Acute respiratory distress syndrome in adults: Diagnosis, outcomes, long-term sequelae, and management. Lancet 400:1157, 2022.

Matthay MA et al: A new global definition of acute respiratory distress syndrome. Am J Respir Crit Care Med 209:37, 2024.

Munshi L et al: Noninvasive respiratory support for adults with acute respiratory failure. N Engl J Med 387:1688, 2022.

The National Heart, Lung, and Blood Institute Petal Clinical Trials Network: Early neuromuscular blockade in the acute respiratory distress syndrome. N Engl J Med 380:1996, 2019.

■ WEBSITES

ARDS Foundation: *www.ardsusa.org*

ARDS Support Center for patient-oriented education: *www.ards.org*

National Health, Lung, and Blood Institute ARDS Clinical Trials information: *www.ardsnet.org* and *www.petalnet.org*

313

Mechanical Ventilatory Support

Scott Schissel

Mechanical ventilation refers to devices that deliver positive pressure gas, of varying oxygen content, to patients with acute or chronic respiratory failure. Hypoxemic respiratory failure refractory to supplemental oxygen and requiring mechanical ventilation is most often due to ventilation-perfusion mismatch or shunt caused by processes such as pneumonia, pulmonary edema, alveolar hemorrhage, acute respiratory distress syndrome (ARDS), and sequelae of trauma or surgery. Hypercapnic respiratory failure is most frequently caused by severe exacerbations of obstructive lung disease, including asthma and chronic obstructive pulmonary disease (COPD); loss of central respiratory drive from acute neurologic events, such as stroke, intracranial hemorrhage, or drug overdose; and respiratory muscle weakness from diseases such as Guillain-Barré syndrome. Mechanical ventilation may also be necessary if patients have an artificial airway placed (an endotracheal tube) due to poor airway protection, such as in coma or in the context of a large upper gastrointestinal hemorrhage and vomiting, or due to processes leading to large airway obstruction, such as laryngeal edema. Finally, since mechanical ventilation can lower the work of breathing compared to spontaneous ventilation, it is a useful adjunct therapy for shock and multiorgan system failure.

PRINCIPLES OF MECHANICAL VENTILATION

Although contemporary mechanical ventilators use positive pressure to inflate the lungs, a patient's response to pressure applied across the lung (transpulmonary pressure) depends on the elastic properties of their lungs and chest wall; the amount of pressure needed to inflate a lung is the same, therefore, whether applied positively via mechanical ventilation or negatively using the diaphragm and chest wall muscles. In ARDS, for example, lungs are "stiff" or poorly compliant and often require much more pressure to achieve a physiologic tidal volume **(Fig. 313-1)**, which, over time, may lead to respiratory muscle fatigue. If a patient with ARDS is on mechanical ventilation and makes no spontaneous respiratory effort, using sedation and neuromuscular blockade, the amount of positive pressure needed to inflate the lung is equal to the negative inflation pressure required if the patient were spontaneously breathing; however, the work of breathing is removed on a ventilator, allowing for sustainable ventilation.

Mechanical ventilation can be lifesaving by restoring adequate oxygenation and correcting hypercapnia. Optimal application of positive-pressure ventilation, however, requires avoiding underinflation, which can cause cycles of alveolar recruitment then collapse and, at the other extreme, alveolar overinflation **(Fig. 313-2)**; collectively, these processes can cause ventilator-induced lung injury by barotrauma and volume trauma. Optimal tidal volume ventilation occurs along the lung pressure–volume curve where respiratory system compliance is greatest, or where the smallest change in applied pressure leads to the greatest increase in lung volume (Fig. 313-2, *shaded box*). To prevent too low lung volumes at end-exhalation, where alveolar collapse occurs, the ventilator can be set to maintain a specified positive pressure at end-exhalation, or positive end-expiratory pressure (PEEP) (Fig. 313-2*B*). Lower tidal volume ventilation (goal 6 mL/kg of ideal body weight)

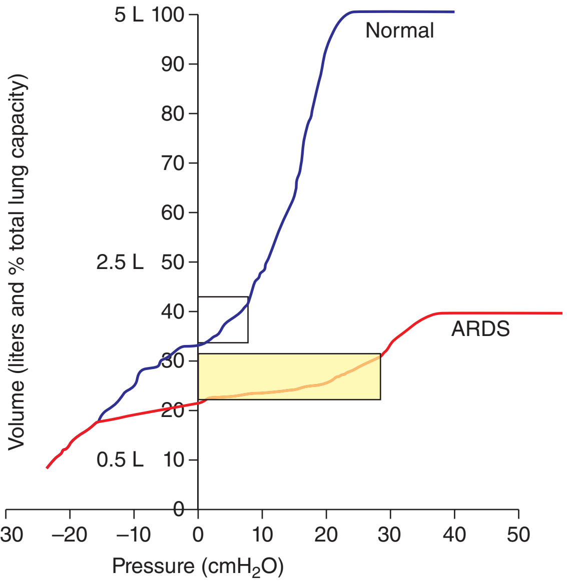

FIGURE 313-1 Hypothetical pressure-volume curves of patients with normal lung function (normal) and acute respiratory distress syndrome (ARDS). A tidal volume breath of 0.5 L in the normal lung requires 8 cmH$_2$O of pressure (*open box*), but in ARDS requires 28 cmH$_2$O (*shaded box*).

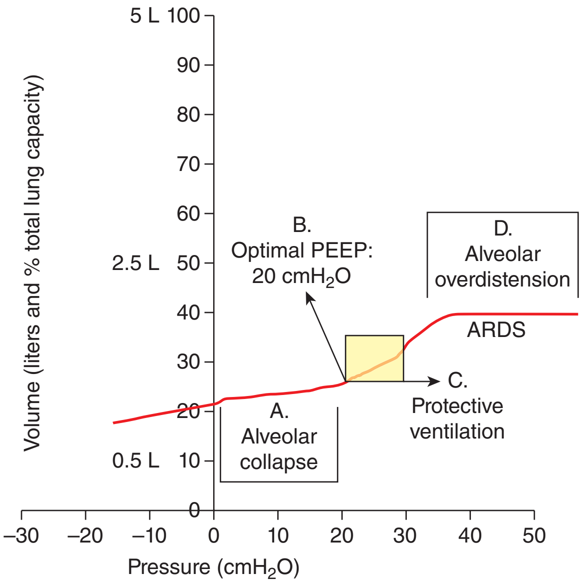

FIGURE 313-2 Hypothetical pressure-volume curve of a patient with acute respiratory distress syndrome (ARDS), demonstrating optimal positive end-expiratory pressure (PEEP) and protective ventilation. A tidal volume breath of 0.5 L initiated at a PEEP of 20 cmH$_2$O (*B*), after the area of greatest alveolar collapse (*A*). End inhalation occurs within the most compliant portion of the pressure-volume curve (*C*) and at a pressure <30 cmH$_2$O, before the area where lung overdistention occurs (*D*), minimizing lung injury.

can help prevent the end-inhalation, or "plateau," pressure (measured just after flow stops at end-inhalation) from exceeding 30 cmH$_2$O; this approach minimizes barotrauma and volume trauma–induced lung injury, especially in ARDS patients (Fig. 313-2C).

MECHANICAL VENTILATION MODES

Mechanical ventilation entails controlling or monitoring the same basic variables involved in spontaneous, negative-pressure breathing, including respiratory rate, tidal volume (V_T), inspiratory flow rate and time, and the fraction of inspired oxygen (FIO_2). In addition, the PEEP is a variable specific to positive-pressure ventilation and set by the clinician. The mechanical ventilation mode determines how much control the clinician and ventilator have over these variables versus the patient; for example, assist control (AC) mode allows for essentially complete operator control of all variables, whereas pressure support (PS) permits the patient to control important variables, such as respiratory rate, V_T, and flow rates (Table 313-1).

ASSIST CONTROL VENTILATION

AC allows the clinician to control nearly all ventilator variables and is widely used when patients cannot safely participate in their own ventilatory efforts, such as when deeply sedated or unstable from acute respiratory failure or other critical illness. Most AC ventilation is in *volume control* mode where the operator sets a specific V_T and respiratory rate, thereby assuring a minimum minute ventilation (V_E). In addition to the set rate, patients can get additional, fully supported breaths at the set V_T by making an inspiratory effort, which is sensed by the ventilator and triggers the breath. The inspiratory flow rate is set by the operator; thus, a dyspneic patient may meet resistance on inhalation if their desired flow rate is higher than the set rate, possibly leading to patient distress and increased work of breathing. In AC volume mode, the operator also sets the PEEP and FIO_2. Importantly, since V_T is an independent variable in volume control (i.e., set by the clinician), the end-inhalation (or plateau) pressure is a dependent variable not controlled by the clinician but rather determined by the compliance of the lung. Inspiratory pressures must be monitored, therefore, to minimize barotrauma.

Although AC is often volume controlled, it can be used in a *pressure control* mode, also referred to as pressure control ventilation (PCV). The key difference between volume control and PCV is that an inspiratory (or "driving") pressure is set instead of a tidal volume in PCV; thus, every time the ventilator delivers a breath, it raises the airway pressure to the set amount above PEEP until inspiratory flow decreases below a set threshold, therefore ending inhalation. The resulting tidal volume will, therefore, vary depending on the compliance of the lung. In a sedated and paralyzed patient (making no respiratory effort), the pressure required to generate a specific tidal volume (*x*) using PCV should, in the same patient, be equal to the plateau pressure in volume control mode where tidal volume is set at *x*. Importantly, since lung compliance can change dynamically, tidal volume may also change with PCV; tidal volume and minute ventilation, therefore, must be monitored since there is no assurance of delivered ventilation volumes as with volume control. PCV is often used to limit peak airway and lung distending (plateau) pressures in situations where high pressure can cause harm, such as in ARDS or after thoracic surgery with fresh suture lines in the airways or lung parenchyma. Importantly, however, inspiratory flow rate and volume are dependent variables in PCV, unlike in volume control ventilation, and not set by the clinician. Spontaneously breathing patients on PCV can generate a relative negative pressure in the

TABLE 313-1 Key Features of Commonly Used Mechanical Ventilation Modes

MODE	VARIABLES SET BY CLINICIAN (INDEPENDENT)	MONITORED VARIABLES (DEPENDENT)	ADVANTAGES	DISADVANTAGES
Assist control–volume control	V_T Respiratory rate PEEP FIO_2 Inspiratory flow rate	Peak inspiratory airway pressure End-inhalation (plateau) pressure V_E	Guarantee minimum V_T and V_E Control V_T, limiting volume trauma	Barotrauma from high plateau pressure Patient-ventilator dyssynchrony, increased work of breathing
Assist control–pressure control	Inspiratory driving pressure Respiratory rate PEEP FIO_2	Tidal volume V_E	Limit barotrauma (if patient respiratory efforts minimal) Inspiratory flow can vary with patient effort (improved comfort/synchrony)	V_t and V_E not mandated; must monitor closely Patient's respiratory effort can lead to large V_T and volume trauma
Pressure-regulated volume control	V_T Respiratory rate PEEP FIO_2	Peak inspiratory airway pressure End-inhalation (plateau) pressure V_E	Patient effort can vary inspiratory flow, increasing comfort, and ventilator synchrony Guarantee minimum V_T and V_E	Variable patient effort can lead to V_T larger than set V_T; monitor to prevent volume trauma
Pressure support	Inspiratory pressure PEEP FIO_2	V_T Respiratory rate V_E	Patient effort preserved and controls V_T, inspiratory flow, and respiratory rate, allowing for ventilator synchrony	Apnea and hypoventilation possible; must monitor respiratory rate, V_T, and V_E closely

Abbreviations: FIO_2, fraction of inspired oxygen; PEEP, positive end-expiratory pressure; V_E, minute ventilation; V_T, tidal volume.

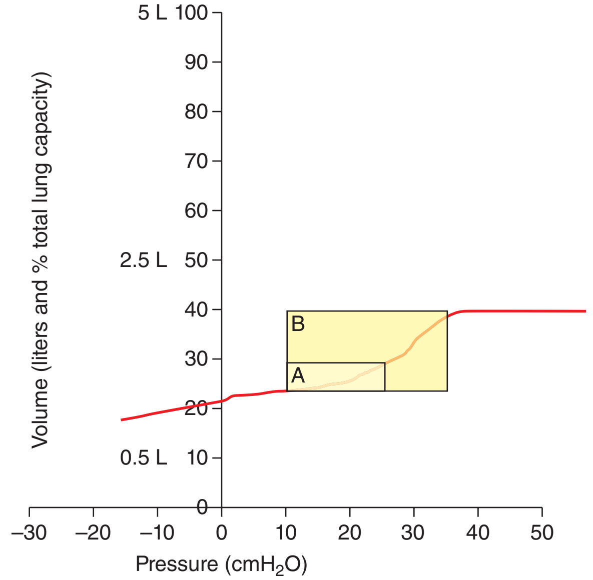

FIGURE 313-3 Hypothetical pressure-volume curve of a patient on pressure control ventilation, paralyzed (*A*) and breathing spontaneously (*B*). (*A*) Paralyzed patient (*light shaded box*): positive end-expiratory pressure (PEEP), 10 cmH$_2$O; inspiratory (driving) pressure, 15 cmH$_2$O; end-inhalation (plateau) pressure, 25 cmH$_2$O; tidal volume (V_T), 300 mL. (*B*) Breathing patient (*dark shaded box*): PEEP, 10 cmH$_2$O; inspiratory (driving) pressure, 15 cmH$_2$O; patient effort (negative "pulling" pressure), 10 cmH$_2$O; end-inhalation (plateau) pressure displayed on ventilator, 25 cmH$_2$O; net end-inhalation (transalveolar) pressure, 35 cmH$_2$O; V_T, 700 mL.

ventilator circuit, transiently decreasing the positive pressure below the set point; the ventilator responds by increasing gas flow until it restores the set pressure, resulting in higher inspiratory flow rates, a higher tidal volume for that breath, and importantly, increased pressure across the alveoli, equal to the absolute (negative) pressure generated by the patient plus the positive pressure set by the clinician (Fig. 313-3). Since mechanical ventilators do not routinely measure or graphically display the negative pressure generated by the patient, clinicians can be unaware of this additional transalveolar pressure and potential harm by volume and barotrauma; importantly, therefore, clinicians should monitor for increases in tidal volume on PCV.

■ PRESSURE-REGULATED VOLUME CONTROL VENTILATION

Advances in ventilator technology, such as flow and pressure sensors and microprocessors, allow for additional modes of mechanical ventilation that meld the benefits of volume and pressure control ventilation. Pressure-regulated volume control (PRVC) ventilation is a fully supported mode of ventilation where the clinician sets a target tidal volume, as in volume control ventilation, but it allows a patient to make spontaneous respiratory efforts and vary inspiratory flow rates, as in PCV, enhancing patient comfort and ventilator synchrony. PRVC senses patient inspiratory efforts and delivers the least amount of positive pressure to achieve the targeted tidal volume; since patient efforts can vary and ventilator adaptation is not instantaneous, tidal volumes can vary from breath to breath on PRVC. In disease states where tidal volume needs tight control to prevent volume trauma, such as in ARDS, PRVC must be used cautiously if the patient can make significant respiratory effort.

■ PRESSURE SUPPORT VENTILATION

Pressure support ventilation (PSV) and PCV are very similar except there is no mandated ventilation or set mechanical respiratory rate on PSV, and ventilator support is entirely patient triggered and controlled. The clinician sets the Fio$_2$, PEEP, and maximum inspiratory pressure. When patients make a negative-pressure inspiratory effort on PSV, the ventilator senses this pressure change and increases positive pressure to the set inspiratory pressure level, maintaining it until flow decreases below a set threshold (often ~20% of peak inspiratory flow); at this point, inhalation ends and pressure drops back to the set PEEP. The tidal volume on PSV is monitored but not assured, is determined by lung compliance, and depends on the patient's sustaining an inspiratory effort. Tidal volume, minute ventilation, and respiratory rate, therefore, must be closely monitored on PSV to detect hypopnea/apnea and hypoventilation. PSV is often used when patients are less sedated and able to participate in respiratory work, such as when transitioning off mechanical ventilation or on a ventilator only for airway support.

■ NONINVASIVE POSITIVE PRESSURE VENTILATION

Noninvasive ventilation (NIV) is historically referred to positive-pressure ventilation and is delivered via a nasal or full-face mask at a continuous pressure (continuous positive airway pressure [CPAP]) or at different inspiratory and expiratory pressures (bi-level positive airway pressure [BiPAP]). Most current noninvasive ventilators, however, can function in full support modes, including volume control ventilation. NIV is particularly beneficial for acute respiratory failure where the underlying cause responds quickly to treatment, minimizing the need for prolonged mechanical ventilatory support. For moderate acute hypercarbia, blood pH between 7.25 and 7.35, due to exacerbations of chronic obstructive pulmonary disease (COPD), NIV, for example, reduces the need for endotracheal intubation and shortens hospital length of stay; more severe acute respiratory acidosis from COPD exacerbations (blood pH <7.2) generally requires mechanical ventilation with an endotracheal tube. NIV can also be an important adjunct treatment for respiratory failure from acute cardiogenic pulmonary edema, where interventions, such as diuresis and vasodilator therapy, can rapidly improve gas exchange and respiratory mechanics. NIV, particularly with volume support modes, is effective in managing chronic respiratory failure from restrictive lung diseases, such as severe scoliosis and respiratory muscle weakness, and in COPD complicated by chronic hypercapnia, where nocturnal NIV reduces COPD-related hospital admissions. Despite the technical innovations in NIV and expanding clinical applications, several important contraindications to using mechanical ventilation without a secure airway, such as an endotracheal tube or tracheostomy tube, include delirium, difficulty managing respiratory secretions, and hemodynamic instability (Table 313-2).

STRATEGIES TO OPTIMIZE GAS EXCHANGE ON MECHANICAL VENTILATION

■ ARTERIAL OXYGENATION

The optimal partial pressure of arterial oxygen (Pao$_2$) and oxygen saturation measured by pulse oximetry (Spo$_2$) during mechanical ventilation remain uncertain. Although tissue hyperoxia can cause oxidative injury with some clinical studies of mechanically ventilated patients suggesting worse clinical outcomes with higher Fio$_2$ and when Pao$_2$ frequently reaches supraphysiologic levels, randomized studies comparing conservative oxygen delivery to a more liberal oxygen strategy have not demonstrated a clear advantage to conservative oxygen delivery. In ARDS, targeting a lower Pao$_2$ of 55–70 mmHg (or Spo$_2$ of 88–92%) versus a higher, but more physiologic, Pao$_2$ of 90–105 mmHg (or Spo$_2$ >96%) did not lower mortality, with adverse events being more frequent in the lower Pao$_2$ group, including mesenteric ischemia. Pao$_2$ and Spo$_2$ targets, therefore, should be individualized to patients considering circumstances where even mild hyperoxia may be harmful, such as in recovery from ischemic brain injury, and, conversely, where lower Pao$_2$ levels (<55–70 mmHg) may be less optimal, such as in patients with ARDS and evidence of bowel dysfunction. Regardless

TABLE 313-2 Common Contraindications to Noninvasive Ventilation

Inability to protect the airway, such as severe encephalopathy
High risk for aspiration, such as vomiting or severe upper gastrointestinal bleeding
Difficulty clearing respiratory secretions
Facial trauma or surgery
Upper airway obstruction or compromise
Significant hemodynamic instability

of the approach, there is no evidence that a supraphysiologic PaO_2 (>100 mmHg) has clinical benefit; thus, sustained hyperoxia should be avoided.

Arterial hypoxemia refractory to standard mechanical ventilation techniques is common in severe acute lung disease, especially ARDS. In general, if the FiO_2 requirement is >0.6 or the PaO_2:FiO_2 ratio is <150 mmHg, additional interventions should be considered to improve arterial oxygenation. The application of adequate PEEP to prevent alveolar collapse during exhalation improves oxygenation by decreasing ventilation/perfusion ($\dot{V}/\dot{Q}$) mismatch and shunt in areas of atelectatic lung. PEEP should ideally be set at the lower inflection point of the most compliant region of the lung pressure-volume curve (Fig. 313-2*B*). Although optimal PEEP may improve arterial oxygenation, achieving best PEEP has not been shown to improve clinical outcomes definitively and may have deleterious effects, including barotrauma with pneumothorax and hypotension from decreasing venous return to the right ventricle. Patients with refractory hypoxemia are often dyspneic on mechanical ventilation and make significant respiratory efforts dyssynchronous with the ventilator despite deep sedation, leading to poor ventilation and preventing optimal $\dot{V}/\dot{Q}$ matching. In this context, neuromuscular blockade can be very effective at restoring effective mechanical ventilation and optimizing gas exchange. Although a necessary intervention at times, neuromuscular blockade does not improve overall outcomes in ARDS, can contribute to critical illness myopathy, and requires adequately deep sedation to prevent conscious paralysis; thus, it should be used only when necessary to treat refractory hypoxemia. In ARDS, diseased lung is predominantly dependent, and placing the patient in a prone position for extended periods can significantly improve arterial oxygenation. The role of prone positioning in other disease states is unknown and can be associated with adverse events unless performed by a trained team, such as dislodging endotracheal tubes and central venous catheters. Delivery of pulmonary vasodilator medications through the airway can improve perfusion to ventilated alveolar units, therefore improving $\dot{V}/\dot{Q}$ matching and arterial oxygenation. Inhaled prostacyclins, such as epoprostenol, and nitric oxide are commonly used to treat refractory hypoxemia and can increase, on average, the PaO_2:FiO_2 ratio by 20–30 mmHg. Hypoxemia refractory to these multiple interventions may require consideration of transitioning to extracorporeal membrane oxygenation (ECMO; see below).

HYPERCAPNIA

Except for rare circumstances of excess CO_2 production (VCO_2), which can occur in the setting of fever, sepsis, overfeeding, and thyrotoxicosis, most hypercapnia is due to inadequate alveolar ventilation (V_A) from an increase in the fraction of dead space (V_D) [the volume of each breath not participating in CO_2 exchange] relative to the total minute ventilation (V_E), expressed as $V_A = V_E\,(1 - V_D/V_T)$. Normal physiologic dead space is approximately 150 mL (~2 mL/kg), making the V_D/V_T for a 500-mL tidal volume breath 0.3. In acute respiratory failure due to ARDS, for example, V_D may increase due to poorly perfused but ventilated portions of lung while ventilation strategies lead to low V_T; thus, a modest increase in V_D to 200 mL and a low V_T of 300 mL will result in a V_D/V_T of 0.66, a situation where hypercapnia may easily develop. Hypercapnia in the context of low tidal volume (6 mL/kg) ventilation for ARDS often causes acute respiratory acidosis that can be managed with higher respiratory rates, up to 30 breaths/min. Respiratory acidosis is often tolerated down to a pH of 7.2, so-called "permissive hypercapnia," but progressive acidosis may require intravenous alkalinizing therapy (e.g., sodium bicarbonate or tromethamine) or accepting an increase in V_T. In severe exacerbations of obstructive lung disease, COPD, and status asthmaticus, hypercapnia and acute respiratory acidosis are common despite mechanical ventilation, with average PaCO_2 values of 65 mmHg and blood pH of 7.20 after initial endotracheal intubation. Poor alveolar ventilation is primarily due to dead space created by alveolar capillary compression in areas of alveolar overdistension and lung hyperinflation. Increasing minute ventilation by increasing the respiratory rate or tidal volume will, therefore, often paradoxically worsen hypercapnia by increasing gas trapping and V_D/V_T. The optimal ventilator strategy for severe obstructive lung disease physiology entails using lower respiratory rates, usually 9–12 breaths/min, and moderate tidal volumes (7–9 mL/kg) to maintain a minute ventilation of ~10 L/min; higher minute ventilation usually worsens hyperinflation and can cause barotrauma. To prevent dyspneic patients from driving hyperventilation, deeper sedation and occasionally neuromuscular blockade are necessary until severe bronchial obstruction responds to medical therapy. Although permissive hypercapnia can minimize barotrauma and volume trauma during mechanical ventilation, hypercapnia has adverse effects including increased intracranial pressure, pulmonary artery vasoconstriction, and even depressed cardiac contractility (**Table 313-3**). The benefits and risks of a hypercapnia ventilatory strategy must, therefore, account for the individual patient's comorbid medical conditions, for example, acute neurologic injury and risk of critical increases in intracranial pressure.

TABLE 313-3 Adverse Effects of Hypercapnia[a]

Pulmonary arterial vasoconstriction (possible worsening of right heart failure)
Rightward shift of the oxyhemoglobin curve
Cerebral vasodilation
Increased intracranial pressure
Sympathetic-adrenal stimulation
Reduced cardiac contractility (especially in the presence of β-adrenergic blocking therapy)

[a]Some effects decrease if cellular pH is corrected.

COMPLICATIONS OF MECHANICAL VENTILATION

AIRWAY

Endotracheal intubation and mechanical ventilation can lead to several pulmonary and extrathoracic complications, especially when patients remain on mechanical ventilation for >7 days. Upper airway complications from endotracheal tube placement include vocal cord trauma (edema, avulsion, paralysis), tracheal stricture due to granulation tissue, and tracheomalacia. Vocal cord injury can lead to postextubation stridor (PES) and need for replacement of an endotracheal tube. PES risk factors include prolonged (>7 days) or traumatic intubation, large endotracheal tube size, previous episode of PES, and head/neck surgery or trauma. Patients with PES risk factors should have the balloon cuff deflated on their endotracheal tube and assessed for air passing across the balloon (so-called "cuff leak test"). Patients with no cuff leak have an approximate 30% risk of PES and may need further assessment for causes of PES, with endotracheal tube removal delayed until the underlying process is treated.

ADVERSE CARDIOPULMONARY EFFECTS OF POSITIVE-PRESSURE VENTILATION

High positive intrathoracic pressure, such as sustained inspiratory plateau pressures >30 cmH_2O or high PEEP, can cause several manifestations of lung barotrauma, including worsening of acute lung injury, pneumomediastinum, pneumothorax, and even pneumoperitoneum. Although positive-pressure ventilation can improve left-sided heart failure by decreasing left ventricular preload and afterload, right ventricular failure and pulmonary arterial hypertension can worsen due to inadequate right ventricular preload and an increase in right ventricular afterload and pulmonary vascular resistance; these effects on the right ventricular and pulmonary circulation should be considered when choosing a ventilatory strategy in patients with severe right-sided heart disease. In addition, blunted central venous return can cause upper and lower extremity edema, especially in the setting of aggressive IV fluid resuscitation and vascular leak related to the underlying critical illness.

VENTILATOR-ASSOCIATED PNEUMONIA

Several factors during mechanical ventilation, such as violation of natural airway defenses, sedation with depressed cough, and micro-aspiration, all increase the risk of bacterial entry into the lower respiratory tract and development of pneumonia. Ventilator-associated pneumonia (VAP) occurs in up to 15% of mechanically ventilated patients and causes death in nearly 50% of patients. VAP is a lower respiratory tract infection that occurs ≥48 h after initiating mechanical ventilation and requires the following: (1) new pulmonary opacities on chest x-ray, (2) a clinical change consistent with pneumonia (fever, increased sputum, leukocytosis, or increase in ventilator support, such as increased FIO_2 or PEEP), and (3) positive microbial culture obtained from the lower respiratory tract via deep endotracheal suctioning or bronchoscopy specimen (bronchoalveolar lavage or protected endobronchial brushing). Most VAP pathogens are typical hospital-acquired bacteria including *Staphylococcus aureus*, *Pseudomonas aeruginosa*, and several other enteric gram-negative rods. In cases of suspected VAP, early empiric antibiotic therapy generally requires an intravenous β-lactam with broad gram-negative rod activity, such as piperacillin-tazobactam, cefepime, or ceftazidime. Empiric therapy for methicillin-resistant *S. aureus* (MRSA) with vancomycin or linezolid or for multidrug-resistant enteric gram-negative rods with a carbapenem should depend on local intensive care unit (ICU) infection control data or individual patient risk for these resistant bacteria. If possible, based on respiratory cultures, empiric antibiotic regimens should be narrowed and total treatment duration should be 7 days. Given the significant morbidity and mortality for VAP, prevention strategies are paramount and should be part of standardized care or "bundles." VAP prevention interventions supported by clinical trial evidence include head-of-bed elevation to at least 30–45° (70% VAP reduction compared to supine position), specialized endotracheal tube use with a suction port above the cuff to minimize aspirated secretions (50% VAP reduction), minimization of ventilator circuit tubing changes (prevents bacterial entry), and hand hygiene before handling the ventilatory circuit. Practices with uncertain value in reducing VAP but still reasonable include limiting deep tracheal suctioning, daily sedation interruption, and routine mouth and dental care.

■ OTHER

The systemic physiologic stress associated with mechanical ventilation and necessary adjunctive therapies, such as sedation and neuromuscular blockade, can cause significant extrathoracic complications. The more common disorders include gastrointestinal stress ulcers and bleeding, deep venous thrombosis and pulmonary embolism, sleep disruption and delirium, and critical illness–associated myopathy that sometimes leads to prolonged mechanical ventilation. To minimize the risk of these adverse events, ICUs should institute care bundles including daily interruption of sedatives and assessment for extubation and prophylaxis for deep venous thrombosis.

LIBERATION FROM MECHANICAL VENTILATION

Discontinuing mechanical ventilation and transitioning a patient back to spontaneous breathing is often referred to as ventilator "weaning," implying dependency on positive-pressure ventilation once started. Although patients on prolonged mechanical ventilation can develop respiratory muscle weakness, this occurs in a minority of patients. Approaching removal of ventilator support as a "wean" extends unnecessary mechanical ventilation time up to 40%. Liberating a patient from mechanical ventilation, therefore, should be more active by frequently assessing a patient's readiness for spontaneous breathing, determined largely by resolution of the underlying process causing respiratory failure **(Fig. 313-4)**. Important criteria indicating a patient may be ready for extubation include the following: underlying disease process has improved, patient is awake and largely off sedative medications, FIO_2 ≤0.5, PEEP <8 $\mathrm{cmH_2O}$, and SaO_2 >88%, stable hemodynamics, and manageable respiratory secretions with adequate cough. These criteria should be assessed daily, and if achieved, patients should have a spontaneous breathing trial (SBT), which is a maneuver wherein positive pressure is set to a minimum to compensate for endotracheal tube resistance (usually 5–7 $\mathrm{cmH_2O}$) and the patient breathes spontaneously from 30 to 120 min. A patient "passes" the SBT if they appear comfortable overall (no marked anxiety or diaphoresis) and have a respiratory rate <35, SaO_2 >90%, systolic blood pressure between 90 and 180 mmHg, and heart rate change of <20%. Patients passing an SBT have a >70% chance of successful extubation. Incorporating extubation "readiness" screening followed by SBT into a care protocol leads to 25% fewer ventilator days and a 10% decrease in ICU length of stay compared to traditional ventilator weaning. Although many physiologic variables correlate with successful liberation from mechanical ventilation, such as minute ventilation, negative inspiratory force generation, and the respiratory rate–to–tidal volume ratio (Tobin index), overrelying on these measures versus the outcome of an SBT leads to unnecessary delays in extubation. Risk factors for failing extubation even after a successful SBT include age >65, congestive heart failure, COPD, Acute Physiology and Chronic Health Enquiry (APACHE-II) score >12, body mass index (BMI) >30, significant secretions, more than two medical comorbidities, and >7 days on mechanical ventilation. Patients with these risk factors transitioned immediately after extubation to noninvasive respiratory support using either high-flow oxygen or positive-pressure NIV have significantly lower rates of reintubation and need to resume mechanical ventilation. Although NIV

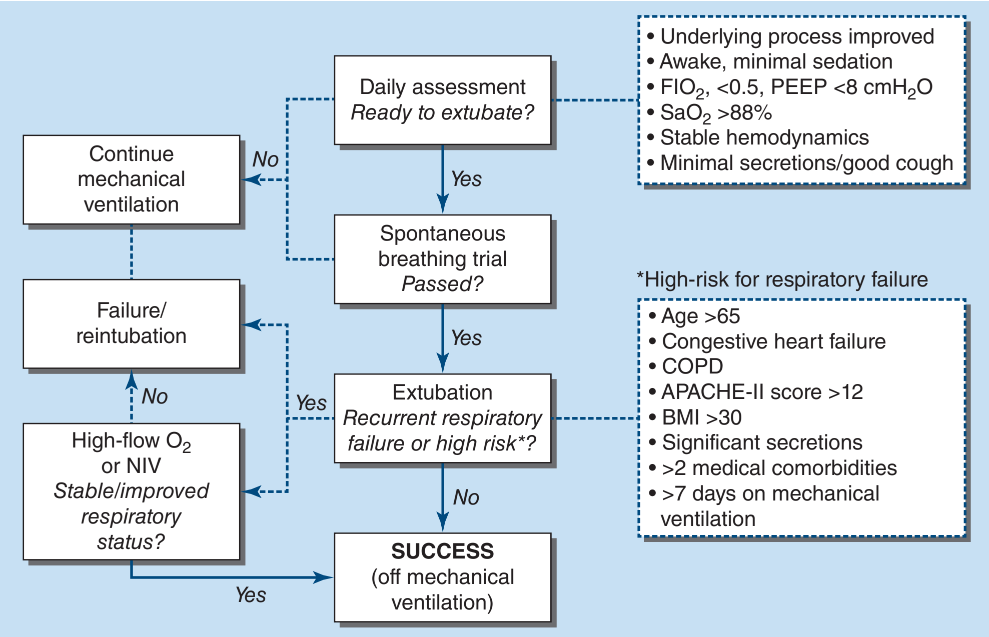
Daily assessment
Ready to extubate?
• Underlying process improved
• Awake, minimal sedation
• FIO_2, <0.5, PEEP <8 $\mathrm{cmH_2O}$
• SaO_2 >88%
• Stable hemodynamics
• Minimal secretions/good cough
Continue mechanical ventilation
No
Yes
Spontaneous breathing trial
Passed?
*High-risk for respiratory failure
Failure/reintubation
Yes
• Age >65
• Congestive heart failure
• COPD
• APACHE-II score >12
• BMI >30
• Significant secretions
• >2 medical comorbidities
• >7 days on mechanical ventilation
No
Yes
Extubation
Recurrent respiratory failure or high risk*?
High-flow O_2 or NIV
Stable/improved respiratory status?
No
SUCCESS
(off mechanical ventilation)
Yes

FIGURE 313-4 Algorithm for discontinuing mechanical ventilation. APACHE-II, Acute Physiology and Chronic Health Enquiry II; BMI, body mass index; COPD, chronic obstructive pulmonary disease; PEEP, positive end-expiratory pressure; NIV, noninvasive ventilation.

TABLE 313-4 Main Types and Key Features of Extracorporeal Gas Exchange

TERM	DESCRIPTION	KEY FEATURES	IMPORTANT TECHNICAL NOTES
VA-ECMO (venoarterial extracorporeal membrane oxygenation)	Deoxygenated blood drains via venous catheter to a pump and membrane oxygenator; blood is then returned to the arterial system	Circulatory and respiratory support	Requires large vascular catheters (16–30 Fr) Higher blood flow rates (2–6 L/min)
VV-ECMO (venovenous-ECMO)	Deoxygenated blood drains via venous catheter to a pump and membrane oxygenator; blood is then returned to the venous system	Respiratory support	Requires large vascular catheters (20–30 Fr) Higher blood flow rates (2–5 L/min)
ECCO$_2$R (extracorporeal CO$_2$ removal)	Venous catheter drains blood to a CO$_2$ removal device; blood then returns via a venous catheter	Partial respiratory support, CO$_2$ removal only	Requires smaller vascular catheters (14–18 Fr) Lower blood flow rates (0.25–2 L/min)

is indicated for patients with hypercapnia after extubation, high-flow oxygen support for all other patients may be preferable given similar efficacy to NIV in preventing reintubation and generally better patient comfort. Although many factors can cause a patient to fail an SBT or require reintubation and continued mechanical ventilation, common processes perpetuating mechanical ventilation include critical illness myopathy and polyneuropathy, myocardial ischemia, congestive heart failure, vascular and extravascular volume overload, delirium, malnutrition, and electrolyte abnormalities (hypophosphatemia, hypokalemia, and hypomagnesemia). These processes should be evaluated and treated, as necessary, in patients failing attempts to discontinue mechanical ventilation.

EXTRACORPOREAL GAS EXCHANGE

Despite interventions to optimize oxygenation and alveolar ventilation on mechanical ventilation, some patients suffer life-threatening hypoxemia, refractory respiratory acidosis, and barotrauma, and may be candidates for salvage therapy with extracorporeal gas exchange, a procedure whereby blood continuously circulates outside the body through a device that oxygenates it, removes CO$_2$, and then returns blood to the patient's circulation. Although often referred to as ECMO, modern gas exchange membranes both deliver oxygen and remove CO$_2$, replacing the gas exchange function of the lung. The main components of an ECMO "circuit" include vascular cannulas to remove and return blood to the patient, a pump to circulate blood, and a gas exchange membrane. ECMO can provide varying levels of both respiratory and circulatory support depending on the clinical situation **(Table 313-4)**. In a patient both in shock and requiring full respiratory support, the ECMO circuit would include a central venous cannula (V) to remove blood and a central arterial cannula (A) to return oxygenated blood at relatively high flow rates (up to 6 L/min) providing mechanical circulatory support, so-called VA-ECMO. In the absence of shock, both the draining and return vascular cannulas can be central venous, or VV-ECMO, but blood flow is still relatively high (2–5 L/min) to provide adequate oxygen delivery to tissues. In situations where a patient's lungs can provide adequate oxygenation but insufficient CO$_2$ removal, such as severe obstructive lung disease exacerbations, a venovenous circuit with low blood flows (0.25–2 L/min) is often adequate to remove CO$_2$ and treat refractory respiratory acidosis, a process called extracorporeal CO$_2$ removal (ECCO$_2$R). ECMO continues to evolve, including the use of hybrid circuits. For example, if patients on traditional VA-ECMO have asymmetric hypoxia in the upper body, an additional venous return catheter can be placed in an internal jugular vein to deliver additional oxygenated blood; this hybrid circuit would be V (removal)-VA (dual arterial and venous return)-ECMO. Moreover, several ECMO centers now have mobile ECMO equipment and teams, allowing patients who are too unstable for transfer to an ECMO center to start on ECMO and facilitate transfer to an ECMO center for further care.

Although technologic advances in the ECMO pumps, gas exchange membranes, and even vascular catheters have reduced ECMO-related complications, the procedure is resource-intensive and still associated with several adverse events, including cannula site hemorrhage and vascular injury, catheter-related infection, pneumothorax, pulmonary and gastrointestinal hemorrhage, limb ischemia, intracranial hemorrhage, and disseminated intravascular coagulation (DIC). Clinical outcomes for ECMO patients remain promising, including for patients with severe respiratory failure from SARS-CoV-2 infection treated with ECMO. However, the overall mortality benefit from ECMO, especially in ARDS, remains unclear. Selecting patients most likely to benefit from ECMO, therefore, is very important, and in addition to exhausting traditional mechanical ventilatory support, patients being considered for ECMO should have a reversible underlying illness or be eligible for organ transplant (heart and/or lung), no chronic severe end-organ disease (e.g., severe kidney disease), no contraindication to systemic anticoagulation, a good functional status before the acute illness requiring ECMO, and a good neurologic prognosis.

FURTHER READING

Acute Respiratory Distress Syndrome Network et al: Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. N Engl J Med 342:1301, 2000.

Barrot L et al: Liberal or conservative oxygen therapy for acute respiratory distress syndrome. N Engl J Med 328:999, 2020.

Bertini P et al: ECMO in COVID-19 patients: A systematic review and meta-analysis. J Cardiothorac Vasc Anesth 36:2700, 2022.

Girard T et al: An official American Thoracic Society clinical practice guideline: Liberation from mechanical ventilation in critically ill adults. Rehabilitation protocols, ventilator liberation protocols, and cuff leak tests. Am J Respir Crit Care Med 195:120, 2017.

Hernandez G et al: Effect of post extubation high-flow nasal cannula vs non-invasive ventilation on reintubation and post extubation respiratory failure in high-risk patients: A randomized clinical trial. JAMA 316:1565, 2016.

Moss M et al: Early neuromuscular blockade in the acute respiratory distress syndrome. N Engl J Med 380:1997, 2019.

Murphy PB et al: Effect of home noninvasive ventilation with oxygen therapy vs oxygen therapy alone on hospital readmission or death after an acute COPD exacerbation. A randomized clinical trial. JAMA 317:2177, 2017.

Tramm R et al: Extracorporeal membrane oxygenation for critically ill adults. Cochrane Database Syst Rev 1:CD010381, 2015.

Section 2 Shock and Cardiac Arrest

314

Approach to the Patient with Shock

Rebecca M. Baron, Anthony F. Massaro

Shock is the clinical condition of organ dysfunction resulting from an imbalance between cellular oxygen supply and demand resulting in cellular and tissue hypoxia. This life-threatening condition is common reason for requiring care in the intensive care unit (ICU).

A multitude of heterogeneous disease processes can lead to shock. The organ dysfunction seen in early shock is often reversible with restoration of adequate oxygen supply. Left untreated, shock transitions from this potentially reversible phase to an irreversible phase and death from irreversible multisystem organ dysfunction. The clinician is required to identify the patient with shock promptly, make a preliminary assessment of the type of shock present, and initiate therapy to prevent irreversible organ dysfunction and death. In this chapter, we review a commonly used classification system that organizes shock into four major types based on the underlying physiologic derangement. We discuss the initial assessment utilizing the history, physical examination, and initial diagnostic testing to confirm the presence of shock and determine the type of shock causing the organ dysfunction. Finally, we will discuss key principles of initial therapy with the aim of reducing the high morbidity and mortality associated with shock.

PATHOPHYSIOLOGY OF SHOCK

The cellular oxygen imbalance of shock is most commonly related to impaired oxygen delivery in the setting of circulatory failure. Shock can also develop during states of increased oxygen consumption or impaired oxygen utilization. An example of impaired oxygen utilization is cyanide poisoning, which causes uncoupling of oxidative phosphorylation. This chapter will focus on the approach to the patient with shock related to inadequate oxygen delivery.

In the setting of insufficient oxygen supply, the cell is no longer able to support aerobic metabolism. With adequate oxygen, the cell metabolizes glucose to pyruvate, which then enters the mitochondria where ATP is generated via oxidative phosphorylation. Without sufficient oxygen supply, the cell is forced into anaerobic metabolism, in which pyruvate is metabolized to lactate with much less ATP generation (per mole of glucose). Maintenance of the homeostatic environment of the cell is dependent on an adequate supply of ATP. ATP-dependent ion pumping systems, such as the Na^+/K^+ ATPase, consume 20–80% of the cell's energy. Inadequate oxygen delivery and subsequent decreased ATP disrupt the cell's ability to maintain osmotic, ionic, and intracellular pH homeostasis. Influx of calcium can lead to activation of calcium-dependent phospholipases and proteases, causing cellular swelling and death. In addition to direct cell death, cellular hypoxia can cause damage at the organ system level via leakage of the intracellular contents into the extracellular space activating inflammatory cascades and altering the microvascular circulation.

DETERMINANTS OF OXYGEN DELIVERY

Because shock is the clinical manifestation of inadequate oxygen delivery relative to cellular needs, we will review determinants of oxygen delivery (DO_2). Disease processes affecting any of the components of oxygen delivery have the potential to lead to the development of shock. Disturbances to key determinants of oxygen delivery form the basis of the four major shock types described below.

The two major components of DO_2 are cardiac output (CO) and arterial oxygen content (CaO_2):

$$DO_2 = CO \times CaO_2$$

The two components of CO are heart rate (HR) and stroke volume (SV), which can be substituted in the above equation as

$$DO_2 = (HR \times SV) \times CaO_2$$

The major determinants of SV are preload, afterload (systemic vascular resistance [SVR]), and cardiac contractility. The relationship can be represented as

$$SV\ \alpha\ (\text{Preload} \times \text{Contractility})/SVR$$

In this equation, preload refers to the myocardial fiber length before contraction (the ventricular end-diastolic volume). Contractility refers to the ability of the ventricle to contract independent of preload and afterload. The SVR represents the afterload, or the force against which the ventricle must contract.

The CaO_2 is composed of oxygen carried by convection with hemoglobin (arterial oxygen saturation, or SaO_2) and oxygen dissolved in arterial blood (partial pressure of oxygen, or PaO_2), given as

$$CaO_2\ (\text{in mL/dL}) = (Hb \times 1.34 \times SaO_2) + (PaO_2 \times 0.003)$$

A disease process that affects these variables (HR, preload, contractility, SVR, SaO_2, or hemoglobin (Hb)) has the potential to reduce oxygen delivery and cause cellular hypoxia. Each of the shock types described below has a distinctive physiologic hemodynamic profile corresponding with alterations in one of the variables affecting oxygen delivery described above.

CLASSIFICATION OF SHOCK

While there is a heterogeneous list of specific conditions that can cause shock, it is helpful to categorize these processes into four major shock types based on the primary physiologic derangement leading to reduced oxygen delivery and cellular hypoxia. The four major shock types are distributive, cardiogenic, hypovolemic, and obstructive. Table 314-1 outlines these major shock types, as well as specific disease processes that can result in that physiologic derangement. Each shock type has a distinct hemodynamic profile **(Table 314-2)**. Familiarity with the major shock types and their unique hemodynamic profile is essential so that when evaluating a patient presenting with shock, the clinician can use the history, physical examination, and diagnostic testing to determine the type of shock present and promptly begin appropriate initial therapy to restore oxygen delivery.

Distributive Shock Distributive shock is the condition of reduced oxygen delivery where the primary physiologic disturbance is a reduction in SVR. It is unique among the types of shock in that there is a compensatory increase in CO (Table 314-2). The central venous

TABLE 314-1 Physiologic Classification of Shock

- Distributive
  - Septic shock
  - Pancreatitis
  - Severe burns
  - Anaphylactic shock
  - Neurogenic shock
  - Endocrine shock
  - Adrenal crisis
- Cardiogenic
  - Myocardial infarction
  - Myocarditis
  - Arrhythmia
  - Valvular
  - i. Severe aortic valve insufficiency
  - ii. Severe mitral valve insufficiency
- Obstructive
  - Tension pneumothorax
  - Cardiac tamponade
  - Constrictive pericarditis
  - Pulmonary embolism
  - Aortic dissection
- Hypovolemic
  - Hemorrhagic
  - i. Trauma
  - ii. GI bleeding
  - iii. Ruptured ectopic pregnancy
  - GI losses
  - Burns
  - Polyuria
  - i. Diabetic ketoacidosis
  - ii. Diabetes insipidus

Abbreviation: GI, gastrointestinal.

TABLE 314-2 Hemodynamic Characteristics of the Major Types of Shock

TYPE OF SHOCK	CVP	PCWP	CARDIAC OUTPUT	SYSTEMIC VASCULAR RESISTANCE
Distributive	↓	↓	↑	↓
Cardiogenic	↑	↑	↓	↑
Obstructive	↑	↓↑	↓	↑
Hypovolemic	↓	↓	↓	↑

Abbreviations: CVP, central venous pressure; PCWP, pulmonary capillary wedge pressure.

pressure (CVP) and pulmonary capillary wedge pressure (PCWP) are usually reduced. The most common cause of distributive shock is sepsis. Sepsis has been redefined as the dysregulated host response to infection resulting in life-threatening organ dysfunction. When this process is accompanied by persistent hypotension requiring vasopressor support (despite adequate volume resuscitation) with end-organ hypoperfusion as measured by an elevated lactate level, it is classified as septic shock. Other processes that are manifest as cellular hypoxia related to a primary reduction of SVR include pancreatitis, severe burns, and liver failure. Anaphylaxis is predominantly an IgE-mediated allergic reaction that can rapidly develop after exposure to an allergen (e.g., food, medication, or insect bite), in which there is a profound distributive type of shock possibly mediated through histamine release. In this setting, there is evidence of both venous and arterial vasodilation. Studies have demonstrated extravasation of up to 35% of the circulating blood volume within 10 min. Patients with severe brain or spinal cord injury may have a reduction of SVR related to disruption of the autonomic pathways that regulate vascular tone. In these patients, there is pooling of blood in the venous system with a resulting decrease in venous return and decreased CO. A final category of patients who present with distributive shock consists of those with adrenal insufficiency. Adrenal insufficiency may be related to chronic steroid use, medications (immune checkpoint inhibitor–associated primary adrenal insufficiency), or other processes that might affect the adrenal glands, including metastatic malignancy, adrenal hemorrhage, infection (e.g., tuberculosis, HIV), autoimmune adrenalitis, or amyloidosis. In conditions of stress (such as infection or surgery), the deficit in adrenal function may become apparent with an inability to increase cortisol leading to vasodilation as well as aldosterone deficiency–mediated hypovolemia.

Cardiogenic Shock Cardiogenic shock is characterized by reduced oxygen delivery related to a reduction in CO owing to a primary cardiac problem. There is usually a compensatory increase in SVR in cardiogenic shock. When the cardiac process (e.g., myocardial infarction) affects the left ventricle (LV), there will be elevation of the PCWP and when it affects the right ventricle (RV), the CVP will be elevated. As detailed above, the CO (and accordingly the DO_2) can be reduced by alterations in the SV or HR. In cardiogenic shock, the SV may be reduced by processes that affect myocardial contractility (e.g., myocardial infarction, ischemic cardiomyopathies, and primary myocarditis) or mechanical valvular disease (acute mitral insufficiency or aortic insufficiency). Both bradyarrhythmias and tachyarrhythmias (from either an atrial or ventricular source) may have associated hemodynamic consequences with a reduction in CO.

Hypovolemic Shock Hypovolemic shock encompasses disease processes that reduce CO (and oxygen delivery) via a reduction in preload. In addition to the reduced CO, this shock type is characterized by an elevated SVR and low CVP and PCWP related to decreased intravascular volume. Any process causing a reduction in intravascular volume can cause shock of this type. Hypovolemic shock is most commonly related to hemorrhage, which may be external (secondary to trauma) or internal (most commonly upper or lower gastrointestinal [GI]) bleeding). Hypovolemic shock can also be seen with nonhemorrhagic processes. Examples include GI illnesses causing profound emesis or diarrhea, renal losses (e.g., osmotic diuresis associated with diabetic ketoacidosis or diabetes insipidus), or skin loss (e.g., severe burns, inflammatory conditions such as Stevens-Johnson syndrome).

Obstructive Shock Obstructive shock is also characterized by a reduction in oxygen delivery related to reduced CO, but in this case, the etiology of the reduced CO is an extracardiac pulmonary vascular or mechanical process impairing blood flow. Specific processes that can impede venous return to the heart and reduce CO include tension pneumothorax (PTX), cardiac tamponade, and restrictive pericarditis. Similarly processes that obstruct cardiac outflow, such as pulmonary embolism, venous air embolism, fat embolism (right heart), or aortic dissection (left heart), are included in this shock type category.

Mixed Shock The types of shock outlined in this classification scheme are not mutually exclusive; not uncommonly, a patient will present with more than one type of shock. For example, the initial physiologic disturbance leading to reduced perfusion and cellular hypoxia in sepsis might be distributive shock. In this setting, a sepsis-induced cardiomyopathy can develop, which reduces myocardial contractility, thus producing a cardiogenic component to what now would be described as a mixed type of shock.

Undifferentiated Shock Upon initial presentation, many patients have undifferentiated shock in which the shock type and specific disease process are not apparent. Using the history, physical examination, and initial diagnostic testing (including hemodynamic monitoring), the clinician attempts to classify a patient with one of the types of shock outlined above so that proper therapy can be initiated to restore tissue perfusion and oxygen delivery.

The epidemiology of shock is dependent on the clinical setting. A 2019 study of the etiology of shock in the emergency department (ED) of a university hospital in Denmark revealed that among 1553 patients with shock, 30.8% had hypovolemic shock, 27.2% had septic shock, 23.4% had distributive nonseptic shock, 14% had cardiogenic shock, and only 0.9% had obstructive shock. In the ICU setting, septic shock predominates. A 2010 study (from eight hospitals) demonstrated that 62% of ICU shock patients had septic shock, 16% hypovolemic shock, 15% cardiogenic shock, and only 2% obstructive shock. Among specialty ICUs, the distribution of shock type differentiates further. In the medical ICU, the largest number of patients have distributive shock related to sepsis. In contrast, a 2019 study of shock in 16 cardiac ICUs found that 66% of shock patients were assessed as having cardiogenic shock. Mortality associated with shock is high, but differences are seen between the types of shock. Septic shock and cardiogenic shock have the highest mortality rates. In the ED study from Denmark, the 90-day mortality of the septic and cardiogenic patients was 56.2% and 52.3%, respectively. These numbers coincide with other studies. Hypovolemic shock is associated with a lower mortality rate.

■ STAGES OF SHOCK

Regardless of type, shock progresses through a continuum of three stages. These stages are compensated shock (preshock), shock (decompensated shock), and irreversible shock. During compensated shock, the body utilizes a variety of physiologic responses to counteract the initial insult and attempts to reestablish the adequate perfusion and oxygen delivery. At this point, there are no overt signs of significant organ dysfunction. Laboratory evaluation may demonstrate mild organ dysfunction (i.e., elevated creatinine or troponin) or a mild elevation of lactate. The specific compensatory response is determined by the initial pathophysiologic defect. In early sepsis with reduction in SVR, there is a compensatory rise in HR (and CO). With early hemorrhagic volume loss, there will be a compensatory increase in SVR and HR. As the host compensatory responses are overwhelmed, the patient transitions into true shock with evidence of organ dysfunction. Appropriate interventions to restore perfusion and oxygen delivery during these initial two phases of shock can reverse the organ dysfunction. If untreated the patient will progress to the third phase of irreversible shock. At this point, the organ dysfunction is permanent and often the patient progresses to multisystem organ dysfunction.

TABLE 314-3 Key Principles in the Treatment of Shock

1. Recognize shock early
2. Assess for type of shock present
3. Initiate therapy simultaneous with the evaluation into the etiology of shock
4. Involve all members of the multidisciplinary team
5. Aim of therapy is to restore oxygen delivery
6. Identify etiologies of shock that require additional lifesaving interventions

EVALUATION OF THE PATIENT WITH SHOCK

The initial evaluation of the patient with shock utilizes the history, physical examination, and diagnostic testing toward two specific aims. The first aim is confirmation of the presence of shock. Given the reversible nature of the organ dysfunction in early shock, it is important that the clinician has a high clinical suspicion for this condition. The possibility of shock should be considered in all patients presenting with new organ dysfunction or hypotension. This early recognition of the presence of shock is an essential tenet of shock care **(Table 314-3)**. The second aim of the initial assessment (history, physical examination, and diagnostic testing) is to identify either a specific shock etiology or to determine the type of shock present. In some patients, the type of shock and etiology will be readily apparent (e.g., the patient with hypovolemic shock from a gunshot wound), but in many cases, the cause is determined only after further evaluation. We will discuss the role of the history, physical examination, and diagnostic testing toward these specific aims. While the assessment of shock etiology is ongoing, the initiation of therapy should not be delayed while the final diagnosis is being determined. Evaluation of shock etiology and initiation of therapy should be simultaneous and as expedient as possible.

History Obtaining a concise, focused history is essential. If the patient is unable to provide a history, ancillary information from family or friends accompanying the patient, emergency medical services (EMS) personnel, or nursing facility (if applicable) should be obtained, and a brief chart review should be performed. Often, a patient with shock will present with nonspecific symptoms such as weakness, malaise, or lethargy. When focal symptoms are reported, it is important to determine whether the symptom is related to the primary process causing shock or a result of inadequate oxygen delivery for cellular metabolic needs. For example, a patient with distributive shock from urosepsis could report chest discomfort in the setting of tissue hypoxia. As the history is being obtained, the clinician must be attentive to any details indicating new organ dysfunction. The most easily identified new organ dysfunction from the history is the presence of a newly altered mental status or decrease in urine output (oliguria). In some cases, the type of shock (and the specific disease process) is apparent from the initial history. Patients with distributive shock from sepsis may present with fever and a history suggesting a focal site of infection (e.g., cough, sputum production, abdominal discomfort, diarrhea, flank discomfort, or dysuria). Anaphylactic distributive shock may be suggested by the onset of pruritis, hives, dyspnea, and new facial edema after exposure to common allergens. Cardiogenic shock may be identified by the onset of exertional chest discomfort. The patient with significant arrhythmia may have an initial complaint of palpitations with syncope or presyncope. Hypovolemic shock may be identified in patients who present with a history of trauma (blunt or penetrating) or GI bleed (hematemesis, melena, or bright red blood per rectum). A patient with hypertension and tearing chest or back pain may be presenting with acute aortic dissection and obstructive-type shock. Asymmetric leg swelling, acute-onset chest pain with dyspnea in the setting of immobility, and/or underlying malignancy raises concern for obstructive shock due to pulmonary embolism.

For most patients, the specific etiology will be less clear, but the history can be helpful in raising the likelihood of a particular type of shock. As an example, a patient with a preexisting immune dysfunction or medication-induced neutropenia may present with hypoperfusion and new organ dysfunction, in which the clinician must have a high suspicion for septic shock. Similarly, a patient with extensive cardiac disease requires a higher suspicion for cardiogenic shock.

Physical Examination The physical examination can assist in the identification of shock (in both the compensated stage prior to overt evidence of organ dysfunction and decompensated stage). The examination also can add insight into what type of shock is present (distributive, cardiogenic, hypovolemic, or obstructive).

Shock is most commonly seen in the setting of circulatory failure. Vital signs are frequently abnormal. In most cases, this is manifest as hypotension (a systolic blood pressure [SBP] <90 mmHg or mean arterial pressure [MAP] <65 mmHg), but isolated blood pressure measurements below these values do not define shock on their own. Many patients may have underlying conditions such as peripheral vascular disease or autonomic dysfunction or are on medications that cause longstanding low blood pressure without any evidence of organ dysfunction. Alternatively, patients with underlying hypertension may develop shock and organ dysfunction at higher blood pressures. Evaluating the patient's current blood pressure in relation to the patient's baseline blood pressure and observing hemodynamic trends and correlation with end-organ perfusion over short time intervals are more useful than an absolute SBP or MAP value. Tachycardia is a common compensatory mechanism in shock. The absence of an elevated heart rate does not exclude shock as patients with underlying cardiac conduction system disease or on home nodal blocking medications may have a diminished or absent tachycardic response. Alternatively, one cannot be reassured by an elevated heart rate without hypotension, as many younger patients can compensate for an extended period of time before developing hypotension. Tachypnea is another vital sign abnormality seen early in shock as the body compensates for a developing metabolic acidosis. While these early compensatory responses are nonspecific, the clinician should recognize these findings early as they may herald the development of end-organ dysfunction if perfusion and oxygen delivery are not restored.

The physical examination can confirm the presence of shock prior to the return of laboratory testing. The central nervous system (CNS), kidney, and skin are the organ systems most easily assessed for evidence of organ dysfunction. These organ systems are considered the "windows" through which we can identify organ dysfunction. Decreased oxygen delivery to the brain is manifest as confusion and encephalopathy. In the early stage of shock, the body will redirect blood flow to the CNS to maintain adequate perfusion. In the patient with shock and altered mental status, all the usual compensatory mechanisms have been outstripped by the magnitude of shock pathophysiology. New encephalopathy represents a manifestation of decompensated shock. To assess renal function during the physical examination, one should evaluate the patient's urine output since the time of presentation. If not already present, a urinary catheter should be placed for accurate hourly assessment of urine output. In patients with normal baseline renal function, oliguria (<0.5 mL/kg per h) may indicate shock. Finally, cold and clammy skin is a sign of hypoperfusion with compensatory vasoconstriction to redirect blood flow centrally (brain, heart). Progressive vasoconstriction can lead to mottling of the skin. Capillary refill time (CRT) is the time it takes for color to return to an external capillary bed after pressure is applied. In the setting of shock, the CRT is delayed.

Many components of the examination provide insight into hemodynamics and assist in elucidating the type of shock present. The physical examination may be used to differentiate shock with high CO (distributive) from that with low CO (cardiogenic shock, hypovolemic shock, and obstructive shock). Examination findings suggestive of high-output shock (distributive) include warm peripheral extremities and large pulse pressure (with low diastolic pressure). Alternatively, cool extremities with narrow pulse pressure would indicate low CO forms of shock. Among types of shock with low CO, the examination can be used to distinguish between conditions with increased intravascular filling pressure (cardiogenic shock, obstructive shock) and intravascular volume depletion (hypovolemic shock). Elevation of jugular venous pressure (JVP) and presence of peripheral edema are seen with high right-sided cardiac pressures. The JVP may be elevated in cardiogenic shock (with right-sided failure) and obstructive shock (pulmonary embolism) but reduced (JVP <8 cm) in hypovolemic shock. Similarly, patients with cardiogenic shock and right-sided cardiac dysfunction may have peripheral edema, but this is not an examination finding typically present in acute hypovolemic shock. Distinguishing cardiogenic from obstructive shock can also be aided by physical examination. Rales on pulmonary auscultation may be related to left-sided cardiac dysfunction. The presence of cardiogenic shock would be further supported by an S3 gallop. One must remember, however, that it is well established that patients with chronic heart failure do not present with the classical findings of acute heart failure.

At times, the physical examination may identify the specific etiology of shock. This is particularly helpful in the patient who cannot provide a detailed history. The examination may demonstrate the site of an untreated infection (e.g., cellulitis, abscess). The examination may reveal a brady- or tachyarrhythmia leading to development of shock. Similarly, large ecchymosis may indicate a significant bleed related to trauma or spontaneous retroperitoneal bleeding. The rectal examination may reveal GI hemorrhage. Pulsus paradox and elevated JVP may suggest the presence of cardiac tamponade. Patients with a tension PTX may have a paucity of breath sounds over the affected side, deviation of the trachea away from the affected side, or subcutaneous emphysema.

Combinations of easily assessed examination components have been organized into a scoring system to identify high-risk patient populations. The shock index (SI) is defined as the HR/SBP, with a normal SI being 0.5–0.7. An elevated SI (>0.9) has been proposed to be a more sensitive indicator of transfusion requirement and of patients with critical bleeding among those with hypovolemic (hemorrhagic) shock than either HR or BP alone. The SI may also identify patients at risk for postintubation hypotension. This concept of use of a clinical score to identify at-risk patients has been extended to patients with distributive shock from sepsis. The quick Sequential Organ Failure Assessment (qSOFA) score is a rapid assessment scale that assigns a point for SBP <100, respiratory rate >22, or altered mental status (Glasgow Coma Scale <15). A qSOFA ≥2 (with a concern for infection) is associated with a significantly greater risk of death or prolonged ICU stay. The Third International Consensus Definition of Sepsis has recommended the use of the qSOFA to identify the most acutely ill subset of patients with sepsis (longer length of stay, increased need for ICU admission, and higher in-hospital mortality).

Diagnostic Testing Laboratory evaluation should be initiated promptly in all patients with suspected shock. The laboratory evaluation is directed toward the dual aim of assessing the extent of end-organ dysfunction and of gaining insight into the possible etiology of shock. Table 314-4 outlines the recommended initial laboratory evaluation of the patient with undifferentiated shock.

BLOOD TESTS Evaluation of lactate, blood urea nitrogen (BUN), creatinine, and transaminases provides an assessment of the extent of end-organ dysfunction related to shock. (See discussion of lactate below.) Urine electrolytes with subsequent calculation of the fractional excretion of sodium (FE_{Na}) or fractional excretion of urea (FE_{Urea}) may indicate states of hypovolemia or decreased effective circulating volume. Elevation of alkaline phosphatase may suggest biliary obstruction and may thereby identify a source of infection in patients with distributive shock. Elevation of cardiac enzymes can indicate a primary cardiac problem with myocyte damage related to ischemia, myocarditis, or a pulmonary embolism. An elevation of the white blood cell count may raise suspicion for an infective process, but this is certainly not diagnostic; an accompanying left shift may improve the sensitivity of this measure. Reductions in hemoglobin and hematocrit are seen in patients with hemorrhagic hypovolemic shock (although an actively bleeding patient may have normal values on initial presentation). While the extent of acidosis may be determined with a venous blood gas (VBG), if there is accompanying hypoxemia, an arterial blood gas should be obtained. For patients with undifferentiated shock, there should always be a high index of suspicion for possible infection. Urinalysis and urine sediment should be sent to evaluate for pyuria. Blood cultures, urine cultures, and sputum cultures should be obtained. Radiographic evaluation should be directed to seek sources of infection suggested by the history and physical examination.

TABLE 314-4 Initial Laboratory Evaluation of Undifferentiated Shock

1. Lactate
2. Renal function tests
3. Liver function tests
4. Cardiac enzymes
5. Complete blood count (with differential)
6. PT, PTT, and INR
7. Pregnancy test
8. Urinalysis and urine sediment
9. Arterial blood gas
10. ECG
11. CXR

Abbreviations: CXR, chest x-ray; ECG, electrocardiogram; INR, international normalized ratio; PT, prothrombin time; PTT, partial thromboplastin time.

Lactate measurement has a role in the diagnosis, risk stratification, and, potentially, the treatment of shock. Increased lactate (hyperlactemia) and lactic acidosis (hyperlactemia and pH <7.35) are common in shock. Lactate is a product of anaerobic glucose metabolism. In glycolysis, the enzyme phosphofructokinase metabolizes glucose to pyruvate. Under aerobic conditions, the pyruvate is then converted (in the mitochondria) to acetyl CoA and enters the Krebs cycle with resulting ATP generation through oxidative phosphorylation. In the setting of cellular hypoxia, the Krebs (tricarboxylic acid) cycle cannot oxidize the pyruvate, and thus the pyruvate is converted to lactate by the enzyme lactate dehydrogenase. Under normal conditions, lactate is produced from skeletal muscle, brain, skin, and intestine. In the setting of reduced oxygen delivery and cellular hypoxia, the amount of lactate produced from these tissues increases (and other tissue can begin to produce lactate). While most of the studies have been performed in patients with septic shock, there is evidence that elevated lactate correlates with a worse outcome. A recent systematic literature review evaluating the role of lactate measurement in a variety of critically ill populations supported the value of serial lactate measurements in the evaluation of critically ill patients and their response to therapy.

Electrocardiogram The electrocardiogram (ECG) is an essential part of the evaluation of the patient with shock. There may be a bradycardic or tachycardic arrhythmia causing a reduction in CO. ST-segment elevation myocardial infarction may be identified. The presence of the S1 Q3 T3 pattern would raise concerns for pulmonary embolism. Reduced voltage in the presence of electrical alternans raises the possibility of pericardial tamponade.

Chest X-Ray The chest x-ray (CXR) can demonstrate a new focal alveolar or interstitial infiltrate suggesting an infectious process (and possible distributive septic shock). Bilateral cephalization of the pulmonary vasculature, peribronchial cuffing, septal thickening, and intralobular thickening are typical of pulmonary edema and a cardiogenic process. The CXR can be used to confirm or exclude the presence of a pneumothorax. CXR findings are neither sensitive nor specific for pulmonary embolism. In select cases, there may be the finding of a peripheral wedge-shaped opacity indicating pulmonary infarction, an enlarged pulmonary artery, or regional vascular oliguria. A chest computed tomography (CT) angiogram may be needed to exclude the diagnosis of pulmonary embolism and further evaluate the thorax.

Point-of-Care Ultrasound Point-of-care ultrasound (POCUS) has an increasing role in the evaluation and treatment of shock. Benefits of POCUS include its low cost, rapidity with which it can be obtained, and noninvasive nature. It has diagnostic value in patients who present with undifferentiated shock. In patients with mixed shock, it can give insight into the relative contribution of the individual shock types. Several structured protocols exist for evaluation of undifferentiated shock including the Rapid Ultrasound for Shock and Hypotension (RUSH), the Abdominal and Cardiothoracic Evaluation with Sonography in Shock (ACES), and Sequential Echographic Scanning Assessing

Mechanism or Origin of Shock of Indistinct Cause (SESAME). These protocols share common components to assess cardiac function, evaluate intravascular volume status, and identify fluid collections. In a rapid and protocolized manner, views are obtained of the heart, lungs, pleural space, inferior vena cava, abdominal aorta, abdomen, and pelvis. Some of the protocols extend to view the deep veins of the lower extremity.

POCUS transthoracic echocardiography (TTE) is central to the POCUS evaluation of shock. TTE utilizes both the two-dimensional (2D) and M mode. It focuses the examination on LV function, RV function, and the pericardium. The 2D mode can evaluate LV size, wall thickness, and ventricular function. Ventricular size and thickness can suggest longer standing cardiac processes. Evaluation of LV function through estimation of left ventricular ejection fraction (LVEF) can identify shock with globally reduced LV function or regional wall motion abnormalities. Similarly, the assessment of RV function also examines RV size and wall thickness (to identify conditions such as elevated pulmonary pressures or suggest pulmonary embolism). An additional important assessment includes evaluation for pericardial tamponade. Two-dimensional echocardiography can also be used to assess valve function, including acute processes, such as mitral valve rupture. Assessment of valvular function is often an evaluation that requires a higher skilled practitioner. The performance of the bedside echocardiogram by the critical care practitioner does not replace formal examination by the echocardiography service or assessment by a cardiologist.

Another component of POCUS includes inferior vena cava (IVC) evaluation to assess intravascular filling. A collapsible IVC in spontaneously breathing patients at the end of expiration suggests reduced intravascular volume. Evaluation of the pleural space for effusion has been a longstanding role of ultrasound, and POCUS pleural space evaluation can be more sensitive than CXR for identifying a PTX. Defined views of the abdomen can identify significant intrabdominal fluid collections indicating hemorrhage or possible infection. Examinations that extend to the proximal deep veins of the lower extremity can identify deep-venous thrombosis, raising the possibility of pulmonary embolism as an etiology of shock. While POCUS can aid in determining the etiology of shock, a 2018 international randomized controlled study utilizing POCUS to evaluate undifferentiated shock in 273 ED patients did not demonstrate a benefit in survival at 30 days or hospital discharge. In addition, there was no difference in amount of IV fluids administered, inotrope use, CTs ordered, or need for ICU care or length of stay.

One significant limitation of POCUS is that performance and interpretation of testing are operator dependent. Familiarity with basic ultrasound techniques and interpretation is now expected in the ED and critical care setting. Accordingly, competency standards have been proposed for emergency medicine and critical care providers in both basic and advanced POCUS techniques.

INITIAL TREATMENT OF SHOCK

Because shock can progress rapidly to an irreversible stage, a key principle in shock management is to initiate treatment for circulatory shock simultaneously with efforts to elucidate shock etiology (Table 314-3). If the initial history, physical examination, and laboratory evaluation have identified the shock type or the specific etiology, then therapy is directed to reverse the underlying physiologic abnormality causing the hypoperfusion and reduced oxygen delivery. To expedite care, all members of the multidisciplinary team (physicians, nurses, pharmacists, and respiratory therapists) must be involved in the development and delivery of care. Details of the optimal care for the specific disease processes leading to shock may be found in other chapters of this text. As many patients will present with undifferentiated shock, in this section, we will discuss treatment directed at the patient with undifferentiated shock. At the conclusion of this section, we will highlight etiologies of shock that require initiation of lifesaving specific therapy.

A key early consideration is to ensure adequate intravenous access. Placement of two peripheral venous catheters (16 or 18 gauge) will provide initial access for the aggressive volume resuscitation that is often required for patients with distributive or hypovolemic shock. If there is concern for distributive shock with sepsis, this IV access will also permit prompt antibiotic administration. For patients with ongoing hypotension despite adequate volume resuscitation, placement of a central venous catheter (CVC) is indicated to provide therapy with vasopressors and inotropes. The CVC will provide a mechanism for hemodynamic monitoring (CVP), as well as a means to obtain central venous oxygen saturations ($ScvO_2$). The $ScvO_2$ is a surrogate of mixed venous oxygen saturation and thus can provide insight into the adequacy of oxygen delivery. Central venous access using a sheath will provide an access point for placement of a Swan-Ganz catheter if more detailed assessment of hemodynamic measurements is required (PCWP, CO, and SVR) and/or if larger bore central access is required for more aggressive volume or blood administration as in hemorrhagic shock. If the patient presents critically ill or in the midst of cardiopulmonary arrest, the quickest method of obtaining central access will be through the use of an intraosseous device. Placement of an arterial line allows for intravascular measurement of blood pressure and continuous determination of MAP. In addition, it can provide insight into the adequacy of volume resuscitation through the measurement of systolic or pulse pressure variation. The arterial line will provide access for determination of arterial oxygen tension, which is helpful since peripheral oximetry measurements (SpO_2) can be unreliable in states of tissue hypoperfusion. The arterial line facilitates repeated measures of acid-base status and lactate to assess the impact of treatment. All patients with shock should have a urinary catheter placed to permit hourly assessment of renal function as another potential indication of the adequacy of resuscitation.

Volume Resuscitation Initial volume resuscitation has the aim of restoring tissue perfusion and is crucial to optimal shock therapy. Assessment of current intravascular volume status and determination of the optimal amount of necessary volume resuscitation are challenging. The physiologic goal of volume resuscitation is to move the patient to the non–preload-dependent portion of the Starling curve. Most patients with any of the four shock types will benefit from an increase in intravascular volume. For patients with distributive shock, the need for early aggressive volume replacement is well established. In the past, the use of early goal-directed therapy (EGDT) in septic shock targeted specific measures of CVP, MAP, and SvO_2 to guide volume resuscitation (and initiation of vasopressors and inotropes). More recent studies have demonstrated that targeted resuscitation using invasive monitoring is not universally required, but in all of these studies, patients in the "usual care" arms of the study received early initial volume resuscitation. For patients with suspected septic shock, a minimum of 30 mL/kg as an initial volume of resuscitation is recommended by the Surviving Sepsis Campaign. While the need for volume resuscitation is most apparent for patients with distributive or hypovolemic shock, even some patients with cardiogenic shock may benefit from cautious volume replacement. In these patients, there should be a careful assessment of volume status prior to volume administration.

In general, volume replacement therapy should be given as a bolus with a predefined endpoint to assess the effect of the volume resuscitation. Most commonly, the volume resuscitation will begin with crystalloid. In patients with hypovolemic shock due to ongoing hemorrhage, volume replacement with packed red blood cells is warranted. In cases of massive transfusion, platelets and fresh frozen plasma should be provided to offset the dilution of these components during volume replacement. Because hemoglobin is a key determinant of CaO_2, red cell administration may be a part of volume replacement even without hemorrhage in order to optimize oxygen delivery if hemoglobin content is <7 g/dL.

Assessment of intravascular volume status (and the adequacy of volume resuscitation) begins with the physical examination (described above). The passive leg raise (PLR) test can predict responsiveness to additional IV fluid by providing the patient with an endogenous volume bolus. While the patient is resting in a semirecumbent position at a 45° angle, the bed is placed in Trendelenburg position such that the patient's head becomes horizontal and the legs are extended at a

45° angle. There is then an immediate (within 1 min) assessment of changes in CO (or pulse pressure variation as a surrogate). It is important to emphasize that one does not merely look for changes in blood pressure; if the shock patient is mechanically ventilated, there is the option of looking at changes in SV variation (or pulse pressure variation) during the respiratory cycle to assess volume responsiveness. A >12% SV variation suggests a volume-responsive state. This measurement requires that the patient be in a volume cycle mode of ventilation, without breath-to-breath variations in intrathoracic pressure and without arrhythmias. A final caveat to the use of these parameters to assess volume status is that these studies were performed on patients being ventilated with tidal volumes larger than currently used to minimize ventilator-induced lung injury.

There is also increased use of echocardiography to assist in determination of intravascular fluid status, with a variety of static and dynamic variables that the trained operator can assess. The most commonly used parameters to assess adequacy of volume resuscitation are IVC diameter and IVC collapse. Alternatively, serial assessments of LV function can be performed while volume is being administered. Placement of a pulmonary artery catheter (PAC) is another tool for assessment of volume status. This more invasive measure involves placement of the PAC into the central venous circulation and through the right heart. Ports in the PAC (Swan-Ganz catheter) allow for direct measurement of CVP, pulmonary artery (PA), and PCWPs. The PCWP is used as a surrogate for LA pressure. While studies have not identified a mortality or length-of-stay benefit with routine use of PA catheterization, there are cases where it may be beneficial. Patients with mixed shock (distributive and cardiogenic) or those with ongoing shock of unclear etiology are examples of situations in which it should be considered.

The need for continued volume replacement must be frequently reassessed. As the patient continues to receive treatment for shock, the initial proper strategy regarding volume management may change in light of development of processes that independently require a different volume-management strategy. For patients who initially present with shock but then develop respiratory failure related to acute respiratory distress syndrome (ARDS) or renal failure, it may be reasonable to begin volume removal.

Vasopressor and Inotropic Support If intravascular volume status has been optimized with volume resuscitation but hypotension and inadequate tissue perfusion persist, then vasopressor and/or inotropic support should be initiated. The use of vasopressors and inotropes must be tailored to the primary physiologic disturbance. The clinician must understand the receptor selectivity of various agents and that for some agents the selectivity may be dose dependent. In patients with distributive shock, the primary aim is to increase the SVR. Norepinephrine is the first-choice vasopressor in septic shock, with potent α_1 and β_1 adrenergic effects. The α_1 causes vasoconstriction, while β_1 has positive inotropic and chronotropic effects. At higher doses, epinephrine has a similar profile (at lower doses, the β effects predominate) but is associated with tachyarrhythmia, myocardial ischemia, decreased splanchnic blood flow, pulmonary hypertension, and acidosis. In distributive shock, vasopressin deficiency may be present. Vasopressin acts on the vasopressin receptors to reverse vasodilation and redistribute flow to the splanchnic circulation. In a randomized trial in patients with septic shock, the addition of low-dose vasopressin to norepinephrine did not reduce all-cause 28-day mortality compared to norepinephrine alone but suggested a potential benefit in the less "sick" population. Vasopressin is safe and has a role as a second agent for hypotension in septic shock. Dopamine does not have a role as a first-line agent in distributive shock. A randomized controlled study in patients with all-cause circulatory shock did not show a survival benefit from dopamine but did reveal an increase in adverse events (arrhythmia). In this study, the subgroup of patients with cardiogenic shock had increased mortality. For patients with cardiogenic shock, dobutamine is a first-line agent; it is a synthetic catecholamine with primarily β-mediated effects and minimal α adrenergic effects. The β_1 effect is manifest in increased inotropy, and the β_2 effect leads to vasodilation with decreased afterload; it can be used with norepinephrine in patients with mixed distributive and cardiogenic shock.

OXYGENATION AND VENTILATION SUPPORT

In addition to the cellular hypoxia caused by circulatory failure, patients with shock may present with hypoxemia. For patients with distributive shock, this may be related to a primary pulmonary process (e.g., pneumonia in a patient with septic shock). For patients with cardiogenic or obstructive shock, hypoxemia may be related to pulmonary edema as a result of LV dysfunction and elevations of PCWP. For patients with all types of shock, there can be development of ARDS and subsequent $\dot{V}/\dot{Q}$ (ventilation/perfusion) mismatch and shunt. Supplemental oxygen should be initiated and titrated to maintain SpO_2 of 92–95%. This may require intubation and initiation of mechanical ventilation. If the patient requires intubation and initiation of mechanical ventilation, this should be provided promptly so as to minimize the duration of tissue hypoxia. Patients with shock may have high minute ventilatory needs to compensate for metabolic acidosis. As shock progresses, they may not be able to maintain adequate respiratory compensation, which may be a second indication to initiate mechanical ventilator support. If mechanical ventilation support is initiated, it is important to provide ventilation with lung-protective strategies focused on low tidal volume ventilation and optimization of positive end-expiratory pressure to minimize ventilator-induced lung injury. In addition, there should be daily sedation cessation to assess underlying neurologic function and minimize time on mechanical ventilation. There are currently few data to support the use of noninvasive ventilation in the setting of shock.

Antibiotic Administration Sepsis is the most common cause of shock. For patients presenting with undifferentiated shock, if the diagnosis of septic shock is being entertained, then broad-spectrum antibiotics should be administered after obtaining appropriate cultures. For patients with sepsis, every hour of delay in appropriate antibiotic administration is associated with an increase in mortality. While it is ideal to initiate antibiotics after appropriate cultures, the inability to obtain cultures should not delay the start of treatment. When sepsis is excluded as a cause of shock, an important aspect of antibiotic stewardship is to stop all antibiotics.

Specific Causes of Shock Requiring Tailored Intervention The initial evaluation (history, physical examination, and diagnostic testing) may have identified an etiology of shock that requires urgent lifesaving intervention in addition to the initial treatment steps outlined above. Patients with distributive shock secondary to anaphylaxis require removal of the inciting allergen, administration of epinephrine, and vascular support with IV fluid resuscitation and vasopressors. Adrenal insufficiency requires replacement with IV stress-dose steroids. Cardiogenic shock patients with arrhythmia may require treatment as outlined in advanced cardiac life support algorithms or placement of an artificial pacemaker. In cases of acute ischemic events, consideration must be given to revascularization and temporary mechanical supportive measures. In the case of valve dysfunction, emergency surgery may be considered. Patients with hypovolemic shock due to hemorrhage may require surgical intervention in the case of trauma or endoscopic or interventional radiology procedures in the case of a GI source of blood loss. Sources of occult bleeding can include soft tissue injury sites including bleeding after long-bone fractures, retroperitoneal bleeding, and the GI tract. Among patients with obstructive shock, a tension PTX would necessitate immediate decompression. Proximal pulmonary embolism requires evaluation for thrombolytic therapy or surgical removal of the clot. Dissection of the ascending aorta may require surgical intervention.

FURTHER READING

Benham et al: A standardized and comprehensive approach to the management of cardiogenic shock. JACC Heart Fail 8:879, 2020.

Evans L et al: Surviving Sepsis Campaign: International guidelines for the management of sepsis and septic shock. Crit Care Med 49:e1063, 2021.

Gitz Holler et al: Etiology of shock in the emergency department: A 12-year population-based cohort study. Shock 51:60, 2019.

McGuire WC et al: Management strategies for acute pulmonary embolism in the intensive care unit. Chest 166:1532, 2024.

Munroe ES et al: Evolving management practices for early sepsis-induced hypoperfusion: A narrative review. Am J Respir Crit Care Med 207:1283, 2023.

Tehrani BN et al: A standardized and comprehensive approach to the management of cardiogenic shock. JACC Heart Fail 8:879, 2020.

Vincent JL, De Backer D: Circulatory shock. N Engl J Med 369:1726, 2013.

315

Sepsis and Septic Shock

Jeffrey R. Strich, Daniel S. Chertow

Sepsis is an infectious syndrome that results in considerable morbidity, mortality, and long-term sequelae among survivors. Sepsis, derived from the Greek word *sipsi*, "to make rotten," was first described by Hippocrates in the 400s B.C. In the 1800s, Sir William Osler opined on sepsis by saying, "except on few occasions, the patient appears to die from the body's response to infection rather than from it." The first consensus definition of sepsis, published in 1992, recognized sepsis as the body's systemic response to infection. To operationalize this definition, systemic inflammatory response syndrome (SIRS) clinical criteria were established, which included temperature <36°C or >38°C, heart rate >90 beats/min, respiratory rate >20 breaths/min or partial pressure of CO_2 <32 mmHg, and leukocyte count <4000/μL or >12,000/μL or >10% bands. Suspected infection with two or more SIRS criteria was classified as sepsis, while the term *severe sepsis* required evidence of hypoperfusion or end-organ dysfunction including oliguria, acute alteration in mental status, or lactic acidosis. Septic shock was defined as sepsis-induced hypotension, determined by systolic blood pressure <90 mmHg or reduced ≥40 mmHg from baseline, absent other causes, despite adequate volume resuscitation. Sepsis-2, the second consensus definition established in 2003, acknowledged clinical complexity beyond existing SIRS criteria and expanded the list of clinical and laboratory criteria to diagnose sepsis. Sepsis-3, established in 2016 as the current consensus definition, abandoned SIRS criteria, given poor specificity for distinguishing sepsis from other noninfectious inflammatory processes, and eliminated the term *severe sepsis*. Sepsis was redefined as life-threatening organ dysfunction caused by a dysregulated host response to infection, and septic shock was defined as a subset of sepsis in which profound circulatory, cellular, and metabolic abnormalities increase mortality risk beyond sepsis alone.

EPIDEMIOLOGY

Each year, an estimated 48.9 million sepsis cases occur globally, and 1.7 million cases occur in the United States. Data from >400 U.S. academic, community, and federal hospitals indicate that sepsis occurs in ~6% of hospitalized adults, with stable incidence over time **(Fig. 315-1)**. Approximately 11 million sepsis deaths occur globally each year, accounting for one in five total global deaths, with 85% occurring in low- and middle-income countries. An estimated 350,000 sepsis deaths or discharges to hospice occur annually in the United States, with an estimated 15% overall mortality among U.S. hospitalized adults with sepsis (Fig. 315-1) and up to 40% mortality in patients with septic shock. Data from the U.S. Centers for Medicare and Medicaid Services indicate the aggregate cost for inpatient hospital and skilled nursing facility sepsis admissions was an estimated $13.4 billion dollars in 2018. A retrospective evaluation of sepsis costs in U.S. hospitals during 2010–2016 suggests a cost of $16,324 and $38,298 per sepsis and septic shock admission, respectively.

■ PATHOGENS AND SITES OF INFECTION

Approximately 88% of sepsis cases are community onset, defined as being detected within 48 h of hospitalization, whereas an estimated 12% are hospital onset, detected after 48 h of hospitalization. While viruses, fungi, and other pathogens may induce sepsis, the role of bacterial infection is best described. An estimated 53% of sepsis cases in the United States are bacterial culture positive, with a relatively even split between gram-positive and gram-negative organisms.

A retrospective cohort study of 17,430 adult community-onset culture-positive sepsis cases from 104 U.S. hospitals during 2009 to 2015 identified *Staphylococcus aureus*, *Streptococcus* spp., and *Enterococcus* spp. as the most prevalent gram-positive organisms isolated, of which 13.6% were antibiotic-resistant, including methicillin-resistant *S. aureus* (MRSA) and vancomycin-resistant enterococci (VRE). *Escherichia coli*, *Klebsiella* spp., and *Pseudomonas aeruginosa* were the most frequently isolated gram-negative organisms, of which 13.2% were resistant to ceftriaxone, extended-spectrum β-lactams, or carbapenems. The most frequently reported anatomic site of primary infection was the urinary tract (48.9%), followed by the respiratory tract (32.9%), an intraabdominal site (13.6%), and skin or soft tissue (10.3%). Bacteria were most frequently isolated from urine (52.1%), blood (40.0%), and the respiratory tract (16.7%).

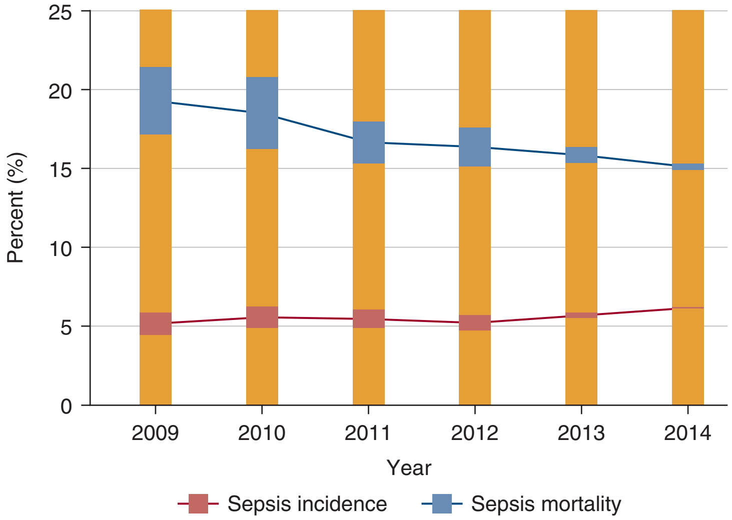

FIGURE 315-1 Sepsis incidence and mortality. Sepsis incidence and mortality. Sepsis incidence is expressed as a proportion of sepsis cases among 7,801,624 adult hospitalizations across 409 U.S. hospitals from 2009–2014. Mortality is the proportion of sepsis deaths among sepsis cases.

Patients with hospital-onset sepsis differ from community-onset sepsis patients in that they more often have comorbidities and experience intraabdominal infections and bacteremia with *Enterococcus*, *Candida*, and *Pseudomonas*. Patients with hospital-onset sepsis also experience higher intensive care unit (ICU) admission rates, longer hospital length of stay, and increased mortality.

RISK AND PROGNOSTIC FACTORS

U.S. Centers for Disease Control and Prevention (CDC) data from 2021 show that sepsis mortality increases with age and is higher among men than women in all age groups: 232.7 versus 173.0 (65–74 years), 477.3 versus 349.8 (75–84 years), and 1037.8 versus 755.5 (≥85 years) deaths per 100,000 population, respectively. Preexisting medical conditions including diabetes and obesity; neurologic, respiratory, or cardiac conditions; renal or hepatic insufficiency; and cancer or other immunosuppressing conditions increase sepsis mortality risk. Recent hospital admission for any reason has also been associated with threefold increased risk of developing sepsis in the following 90 days.

Multiple composite illness severity scoring systems have been applied to hospitalized septic patients to predict outcome or guide stratification or post-hoc analyses in clinical trials or observational studies of sepsis. For example, the Sequential Organ Failure Assessment (SOFA) score quantifies dysfunction in six organ systems including neurologic, using the Glasgow Coma Scale score; cardiovascular, using mean arterial blood pressure or use of vasoactive agents; respiratory, using the ratio of arterial blood partial pressure of oxygen ($Pa{\small O}_2$) to fraction of inspired oxygen ($F{\small IO}_2$) or use of mechanical ventilation; hepatic, using serum bilirubin levels; renal, using serum creatinine levels; and coagulation, using blood platelet levels. Elevated SOFA scores are associated with increased sepsis mortality at the population level but cannot accurately predict outcomes of individual patients.

Improved stratification of sepsis patients based on clinical "phenotype," determined by clustering of clinical variables using machine learning tools to better predict outcome, has been proposed. In a retrospective analysis of >45,000 patients who met Sepsis-3 criteria, statistical, machine learning, and simulation tools were applied to clinical data obtained within 6 h of emergency room presentation to identify α, β, γ, and δ phenotypes, with associated 2%, 5%, 15%, and 32% mortality rates, respectively. Beyond using immediately available clinical data to identify sepsis phenotypes, integration of transcriptional and proteomic data to define sepsis "endotypes" has also been attempted. While these approaches hold promise to guide improved clinical trial design and patient care, additional research is needed to contextualize and operationalize them across patients and pathogens.

PATHOGENESIS

The host-pathogen interaction during sepsis is heterogeneous, based on patient demographic and clinical factors and pathogen type and virulence. Consequently, the pathophysiology and molecular pathogenesis of sepsis vary across hosts and pathogens. However, a working framework for sepsis pathogenesis, with commonalities across hosts and pathogens, provides a basis for understanding which factors contribute to organ injury and dysfunction during severe infections. During local infection, a prototypical physiologic response is characterized by pathogen recognition followed by balanced inflammatory, anti-inflammatory, and repair responses resulting in pathogen clearance with minimal disruption to systemic homeostasis. During sepsis, however, pathogen components and exuberant cellular and soluble immune responses contribute to systemic illness, resulting in end-organ injury and dysfunction **(Fig. 315-2)**. Pathogen components and host responses to infection may also impair or delay adaptive immunity and tissue repair.

Pathologic responses in sepsis are mediated by myeloid lineage cells (i.e., neutrophils, monocytes, macrophages, and dendritic cells), lymphoid lineage cells (i.e., natural killer [NK] cells and lymphocytes), parenchymal cells (e.g., endothelial and epithelial cells), and soluble mediators (e.g., cytokines, chemokines, nitric oxide, histamine, prostaglandin, and bradykinin). Activated leukocytes adhere to endothelium, migrate into tissue, and perform end-effector functions, including reactive oxygen species (ROS) generation, that contribute to tissue injury. Activation of endothelial cells, platelets, the clotting cascade, and the complement system contribute to a prothrombotic state. When platelets and clotting factors are consumed, hemorrhage risk increases. Pathophysiologic manifestations of sepsis include systemic vasodilation, capillary leakage with interstitial fluid accumulation, and microvascular thrombosis resulting in impaired oxygen delivery, uptake, and utilization. Taken together, these factors drive cellular injury and end-organ dysfunction.

TRIGGERS OF HOST IMMUNE RESPONSES

Host immune responses in sepsis are initiated by pathogen components termed pathogen-associated molecular patterns (PAMPs) and propagated by host factors termed damage-associated molecular patterns (DAMPs). PAMPs and DAMPs interact with innate and adaptive immune cells and parenchymal cells early during sepsis, amplifying immune responses. PAMPs include bacterial lipopolysaccharide (LPS) or other bacterial cell wall or membrane components, pathogen nucleic acids, including single- or double-stranded RNA, and other pathogen-related molecules. DAMPs include proteins, lipids, nucleic acids, and other components released from injured cells. Examples include histones, high mobility group box 1 proteins, S100 proteins, oxidized phospholipids, double-stranded DNA, and adenosine triphosphate. PAMPs and DAMPs are recognized by extra- or intracellular receptors, termed pattern recognition receptors (PRRs). Multiple PRRs, including toll-like receptors (TLR), of which 10 have been described in humans, C-type lectin receptors (CLR), receptors for advanced glycation end products (RAGE), retinoic acid–inducible gene-I-like receptors (RIG-I), and nucleotide-binding oligomerization domain-like (NOD) receptors, have been implicated in sepsis signaling. Recognition of PAMPs and DAMPs by PRRs on innate immune cells such as neutrophils, monocytes, and macrophages activates inflammatory pathways, triggers release of proinflammatory cytokines and chemokines, and increases surface expression of vascular adhesion molecules.

MYELOID RESPONSES

Neutrophils, monocytes, and macrophages provide early innate immune defenses against invading pathogens but also contribute to sepsis pathogenesis by promoting inflammation and cellular injury and, in some cases, limiting adaptive immunity. Differentiated mature neutrophils are the most abundant circulating leukocytes in healthy individuals. Upon activation during sepsis, neutrophils upregulate surface adhesion molecule expression (e.g., CD11b) to bind endothelium. Within the microvasculature, neutrophils release neutrophil extracellular traps (NETs), web-like structures of DNA decorated with antimicrobial proteins such as cathepsin-G, myeloperoxidase, and neutrophil elastase to limit dissemination of invading pathogens. NETs interact with activated platelets, endothelial cells, and fibrin to form microvascular thromboses. Activated neutrophils also migrate into tissues to combat microbes through phagocytosis, degranulation, ROS generation, NET formation, and proinflammatory cytokine and chemokine release. Neutrophil granule components degrade extracellular matrix, while ROS oxidize proteins and lipids contributing to cellular injury and dysfunction. Soluble inflammatory mediators further recruit immune cells to sites of infection, exacerbating inflammation and cellular injury. Mediators such as interleukin (IL) 6, granulocyte-macrophage colony-stimulating factor, and granulocyte colony-stimulating factor trigger release of immature neutrophils from the bone marrow, a process termed emergency granulopoiesis that is associated with poor outcomes in septic patients.

Monocytes are innate myeloid cells that circulate in healthy individuals for up to 7 days following release from the bone marrow. Three main monocyte populations are defined by CD14 and CD16 surface expression, including classical (CD14+, CD16–), intermediate (CD14+, CD16+), and nonclassical ($CD14^{low}$CD16+) monocytes. Classical monocytes account for 85% of circulating monocytes in healthy individuals but rapidly migrate to infected tissues during sepsis, resulting in transient monocytopenia and an increased proportion of circulating nonclassical and intermediate monocytes. Nonclassical monocytes are more terminally differentiated blood-resident monocytes thought to patrol the endothelium during sepsis. Monocytes

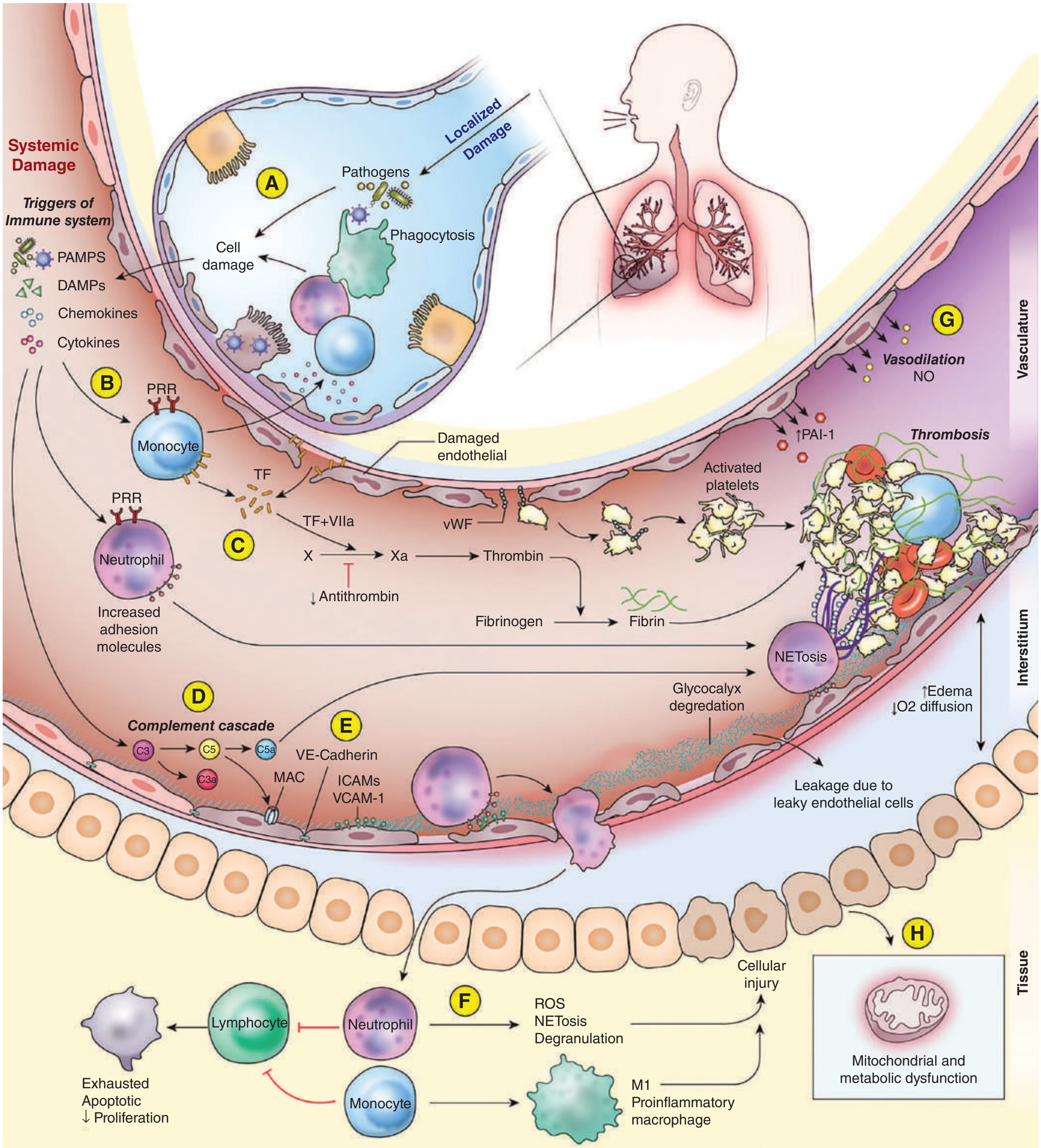
Systemic Damage
Triggers of Immune system
PAMPS
DAMPs
Chemokines
Cytokines
Localized Damage
A
Pathogens
Cell damage
Phagocytosis
B
PRR
Monocyte
PRR
Neutrophil
Increased adhesion molecules
TF
Damaged endothelial
C
TF+VIIa
X
Xa
Thrombin
↓Antithrombin
Fibrinogen
Fibrin
vWF
Activated platelets
↑PAI-1
Thrombosis
G
Vasodilation
NO
Vasculature
NETosis
Interstitium
↑Edema
↓O2 diffusion
Glycocalyx degredation
D
Complement cascade
C3
C5
C5a
C3a
MAC
E
VE-Cadherin
ICAMs
VCAM-1
Leakage due to leaky endothelial cells
H
Tissue
Cellular injury
Mitochondrial and metabolic dysfunction
F
Exhausted
Apoptotic
↓ Proliferation
Lymphocyte
Neutrophil
Monocyte
ROS
NETosis
Degranulation
M1
Proinflammatory
macrophage

FIGURE 315-2 Sepsis pathogenesis. **(A)** Invading pathogens and tissue-resident and recruited leukocyte responses induce localized cell damage. **(B)** Systemic damage is mediated by pathogen-associated molecular patterns (PAMPs), damage-associated molecular patterns (DAMPs), cytokines, and chemokines that activate pattern recognition receptors (PRRs) on circulating monocytes and neutrophils. **(C)** Increasing tissue factor (TF) on endothelium and leukocytes triggers the clotting cascade and fibrin generation. Adherent neutrophils release extracellular traps (NETs) and microthrombi form, composed of leukocytes, fibrin, and platelet aggregates, bound together by von Willebrand factor (vWF). Antithrombin levels are decreased, promoting clot formation, and plasminogen activator inhibitor-1 (PAI-1) levels are increased, impairing clot breakdown. **(D)** Complement activation generates C3a and C5a, which activate platelets and myeloid cells. The C5b-9 membrane attack complex (MAC) promotes endothelial injury. **(E)** Activated and injured endothelium increase intercellular adhesion molecules (ICAMs) and vascular cell adhesion molecule (VCAM)-1 expression, glycocalyx breakdown, apoptosis, and loss of intercellular vascular endothelial (VE)-cadherin and tight junctions contributing to interstitial edema, decreased oxygen diffusion, and leukocyte migration. **(F)** Activated neutrophils degranulate and release reactive oxygen species (ROS) and NETs, and monocytes differentiate into proinflammatory (M1) macrophages, contributing to cellular injury and dysfunction. Activated myeloid cells also suppress lymphocyte function, through exhaustion, apoptosis, and decreased proliferation. **(G)** Arterial vasodilation, partially mediated by excess nitric oxide, contributes to tissue hypoperfusion and ischemia. **(H)** With or without adequate oxygen delivery, cellular mitochondrial and metabolic functions are disrupted during sepsis.

that migrate into tissues are exposed to PAMPs, DAMPS, cytokines, and other mediators, stimulating them to differentiate into dendritic cells or macrophages. Dendritic cells are phenotypically and transcriptionally diverse professional antigen-presenting cells with a broad PRR and cytokine secretion repertoire. These cells contribute to pathogen clearance and inflammation and bridge innate and adaptive immune responses through antigen presentation and lymphocyte activation. Macrophages that originate from bone marrow–derived monocytes are distinct from tissue-resident macrophages, which derive from embryonic progenitors. Depending on environmental conditions within tissues, bone marrow–derived monocytes polarize to become proinflammatory (M1) or anti-inflammatory, reparative (M2) macrophages. M1 macrophages contribute to pathogen clearance through phagocytosis, extracellular trap formation, ROS generation, and inflammatory cytokine release (e.g., IL-1β, IL-6, and tumor necrosis factor [TNF]-α). However, these same effector functions contribute to tissue injury. Macrophages also process and present antigens on surface major histocompatibility complex (MHC) II molecules, which are required for development of antigen-specific effector and memory B- and T-cell responses. This essential link between myeloid and lymphoid cells can be disrupted by diverse pathogen-induced mechanisms including downregulation of MHC surface expression or interruption of antigen processing or presentation by pathogen proteins and other components.

While myeloid cells typically contribute to inflammation in sepsis, a subset of pathologically activated myeloid cells in tissues, termed myeloid-derived suppressor cells, functionally suppress NK-cell, T-cell, and B-cell function, contributing to immunosuppression in sepsis.

LYMPHOID RESPONSES

NK cells are a heterogenous group of innate lymphocytes that migrate to sites of infection, contribute to pathogen clearance, and mediate inflammatory responses during sepsis. NK cells have activating and inhibitory cell surface receptors that modulate their activity. NK-cell PRRs recognize PAMPs including LPS, peptidoglycan, and double-stranded RNA. NK-cell cytokine receptors recognize inflammatory cytokines including type 1 interferons (IFN) and IL-12. Activated NK cells release proinflammatory cytokines including IFN-γ and IL-32, which activate myeloid cells in a positive feedback loop. Activated NK cells also directly kill infected cells that have downregulated MHC class I surface molecule expression, through release of cytotoxic granules including perforin and granzyme. In studies of human sepsis, circulating NK cells with increased CD69 expression, indicative of activation, and increased plasma concentrations of IFN-γ and granzyme A and B support a proinflammatory role of NK cells. However, NK cells also release IL-10, known to suppress myeloid cell–mediated inflammatory responses, highlighting an anti-inflammatory role of NK cells in sepsis as well.

B- and T-cell responses in sepsis are essential for recognizing and clearing pathogens. B cells produce antibodies against foreign antigens and form antigen-specific memory cells. CD8+ T cells lyse infected cells that present foreign antigens in association with MHC surface molecules and form memory cells. CD4+ (T_H1) cells activate CD8+ T cells and support memory T-cell formation, CD4+ (T_H2) cells contribute to B-cell class switching, and CD4+ (T_H17) cells protect against extracellular fungal and bacterial infections. B- and T-cell dysfunction in sepsis impairs adaptive immunity and predisposes to subsequent infections among survivors. Contributing factors to adaptive immune dysfunction in sepsis include death of lymphocytes in the circulation and tissue, cellular exhaustion, decreased proliferation, and apoptosis-refractory regulatory T cells, all of which can inhibit inflammatory responses. Functional manifestations of T-cell exhaustion include impaired cytokine production and diminished cytotoxic activity, while diminished antibody production is observed in exhausted B cells.

ENDOTHELIAL ACTIVATION

During homeostasis, endothelial cells line blood vessels and regulate vascular tone and the exchange of cells, fluid, and molecules between the bloodstream and surrounding tissues. The endothelial glycocalyx, a network of proteoglycans, glycoproteins, glycolipids, glycosaminoglycans, and bound plasma proteins, lines luminal endothelium and maintains homeostatic functions. During sepsis, PAMPs and DAMPs activate endothelial cells, altering their structure and function. Glycocalyx breakdown disrupts blood flow and mediates proadhesive, procoagulant, and proinflammatory endothelial properties. Activated endothelial cells increase surface adhesion molecule expression, including intercellular adhesion molecules (ICAMs) and vascular cell adhesion molecule (VCAM)-1, that coordinates leukocyte adhesion, rolling, and diapedesis. Activated endothelium promotes clotting through increased tissue factor (TF) expression and platelet adhesion and impairs fibrinolysis through increased plasminogen activator inhibitor (PAI)-1 release, among other factors. Activated endothelial cells also produce proinflammatory cytokines, including IL-1β, IL-6, and IFN-α, contributing to local and systemic inflammation. Beyond glycocalyx breakdown, barrier function is impaired through endothelial cell apoptosis and loss of intercellular adherence (e.g., vascular endothelial [VE]-cadherin) and tight junctions. Loss of endothelial barrier function contributes to interstitial fluid accumulation, increased interstitial pressure, and tissue hypoperfusion. Alterations in endothelial cell nitric oxide (NO) metabolism in sepsis contribute to systemic vasodilation and cellular injury.

COAGULOPATHY

The coagulation system, composed of plasma proteins, platelets, and the endothelium, maintains vascular integrity. During sepsis, PAMPs and DAMPs interact with these components, innate immune cells, and the complement system to promote thrombus formation and impair clot breakdown. Increased TF expression on activated leukocytes and endothelium triggers the extrinsic clotting cascade, thereby driving thrombin-mediated conversion of fibrinogen to fibrin. Endothelial injury exposes von Willebrand factor (vWF) on subendothelial surfaces, which binds platelets via glycoprotein Ib receptors, and platelets bind fibrinogen via glycoprotein IIb/IIIa receptors. Activated platelets and endothelium release additional vWF from intracellular stores (i.e., Weibel-Palade bodies in endothelial cells), forming multimers that aggregate platelets at the endothelial surface. Activated monocytes and neutrophils bind to these platelet aggregates, and neutrophils release NETs, contributing to a meshwork of fibrin-rich microthrombi. The complement system, which is composed of >50 soluble or membrane bound proteins, also contributes to microthrombi formation. During sepsis all three complement pathways (classical, alternative, and lectin) become activated. Release of C3a and C5a anaphylatoxins further activates platelets and myeloid cells, and assembly of the membrane attack complex (MAC), composed of C5b-9, promotes endothelial injury. Counterregulatory anticoagulant molecules, including antithrombin, tissue factor pathway inhibitor, and activated protein C, are impaired in sepsis, and increased activity of antifibrinolytic molecules, including PAI and thrombin-activatable fibrinolysis inhibitor, limit clot breakdown. When clotting factors and platelets are consumed, spontaneous and provoked hemorrhage can occur.

CELLULAR INJURY AND DYSFUNCTION

A hallmark of sepsis is multiorgan cellular injury and dysfunction. Arterial hypotension due to systemic vasodilation, myocardial depression, and hypovolemia; increased interstitial edema due to vascular leakage resulting in increased oxygen diffusion distance; and microcirculatory dysfunction in part due to endothelial activation, injury, and microvascular thrombosis all contribute to inadequate oxygen delivery and cellular injury and dysfunction during sepsis. Cell death or dysfunction can also result from direct interactions with PAMPs, DAMPs, activated leukocytes, and their associated mediators including cytokines, chemokines, toxic granules, ROS, and reactive nitrogen species (RNS), including peroxynitrite, nitrogen dioxide, and dinitrogen trioxide. ROS and RNS contribute to protein and lipid oxidation and DNA damage of proximate cells. Even in the setting of adequate oxygen delivery, cellular mitochondrial and metabolic functions are disrupted during sepsis, including ATP generation through glycolysis and oxidative phosphorylation, ion (i.e., Na^+, K^+, and Ca^{2+}) homeostasis, and regulation of cell death pathways.

CLINICAL MANIFESTATIONS AND MANAGEMENT

Sepsis may present with nonspecific signs and symptoms including fever, tachycardia, lethargy, and myalgias with or without localizing end-organ findings such as cough, pyuria, or abdominal pain or evidence of end-organ dysfunction such as oliguria or altered mental status. There is no gold standard diagnostic test for sepsis, so epidemiologic, demographic, clinical, laboratory, radiologic, and microbiologic parameters must be considered in diagnosing sepsis. In addition to a focused history and physical examination to elicit signs, symptoms, and physical manifestations of sepsis, an initial laboratory evaluation should include complete blood count with differential, basic metabolic panel, liver function test, serum lactate, coagulation panel, urinalysis, and point-of-care pathogen testing when available. Microbiologic testing with culture of blood and other potentially infected sites (e.g., urine, sputum, wound), ideally prior to antimicrobial administration, should be performed to identify a specific infection and guide antimicrobial therapy. Focused imaging examination by x-ray, computed tomography, or ultrasound of potentially infected sites (e.g., lung, abdomen) to support a diagnosis of sepsis and guide efforts to control the source of infection should also be pursued. Sepsis mimics include but are not limited to noninfectious inflammatory febrile syndromes such as connective tissue diseases and vasculitides; heart failure and other noninfectious causes of pneumonitis or lung injury; noninfectious abdominal syndromes including mesenteric ischemia and inflammatory bowel disease; and hypotension due to noninfectious causes including hypovolemia, autonomic dysfunction, and adrenal insufficiency. In a recent study of septic patients admitted to the ICU, 25% were retrospectively deemed to have sepsis mimics. The most common mimics were cardiovascular (e.g., heart failure, cardiac arrest) and respiratory (e.g., noninfectious chronic obstructive pulmonary disease). In patients with suspected sepsis or septic shock in whom infection is not confirmed, continuous reevaluation for alternative diagnoses is imperative.

■ RECOGNIZING SEPSIS

Early recognition and treatment of bacterial septic shock with appropriate antibiotics has been associated with reduced mortality. Therefore, the 2023 CDC Hospital Sepsis Program Core Elements guidance document and the 2021 Surviving Sepsis Campaign guidelines recommend hospitals have dedicated sepsis improvement programs that include standard operating procedures for sepsis screening and early treatment. To operationalize updated Sepsis-3 definitions, consensus criteria for sepsis include an increase of ≥2 in SOFA score from baseline in patients with suspected or confirmed infection, and criteria for septic shock include septic patients requiring vasopressor therapy to maintain a mean arterial pressure >65 mmHg and lactate >2 mmol/L despite fluid resuscitation. To more rapidly screen patients for sepsis, the quick (q)SOFA score has been proposed, which includes respiratory rate ≥22 breaths/min, Glasgow Come Scale score <15, and systolic blood pressure ≤100 mmHg. A qSOFA score ≥2 is associated with poor outcome and is more specific, although less sensitive, than SIRS criteria for identifying patients with end-organ dysfunction due to infection. Multiple other sepsis screening tools have been proposed including the National Early Warning Score, the Modified Early Warning Score, and the artificial intelligence–based Targeted Real-Time Early Warning System. Given that each screening tool has advantages and limitations, none is preferentially endorsed for sepsis screening by the 2021 Surviving Sepsis Campaign guidelines.

■ INITIAL SEPSIS MANAGEMENT

Initial treatment of patients with sepsis or septic shock includes timely blood pressure and end-organ support, antimicrobial therapy, and identification and elimination of the source of infection. The 2021 Surviving Sepsis Campaign treatment guidelines provide the most up-to-date and evidenced-based approach for initial and subsequent treatment of patients with sepsis and septic shock. Initial management recommendations include obtaining intravenous access with peripheral or central venous catheters, administering appropriate and timely antibiotics, treating life-threating hypotension with intravenous crystalloid and vasopressor therapy, and managing respiratory insufficiency or failure with supplemental oxygen, airway support, and mechanical ventilation, when indicated. Following initial stabilization in patients who are critically ill or in shock, admission to the ICU within 6 h should be targeted.

Of these interventions, early appropriate antibiotic administration in patients with bacterial septic shock has been most clearly associated with improved survival. In patients with bacterial septic shock, there is an estimated 7–8% increase in mortality for every 1-h delay in appropriate antibiotic administration following shock recognition. Consequently, in patients with suspected or confirmed septic shock, immediate empiric antimicrobial therapy within 1 h of shock recognition is recommended. The association between time to antibiotics and mortality in patients with suspected or confirmed sepsis without shock has not been established. In patients in whom a diagnosis of sepsis is less certain and shock is absent, further time-limited clinical evaluation is recommended prior to empiric antibiotic administration. If an alternative diagnosis is not identified within 3 h of clinical presentation, empiric antibiotic administration is recommended.

Selection of initial empiric antibiotics should consider site of infection and potential etiologic organisms; community versus health care exposure; known prior infections, antibiotic usage, and local antimicrobial resistance profiles; and the patient's immune status, comorbidities, and illness severity. Recommendations for initial empiric antibiotic use based on site of infection and other pertinent factors are summarized in **Table 315-1**. In patients with undifferentiated sepsis, in whom the primary site of infection is unclear, use of broad-spectrum antibiotics with a high likelihood of in vitro susceptibility to all organisms likely causing infection is recommended. In patients in whom *Pseudomonas* is not considered a likely pathogen, then a third-generation cephalosporin such as ceftriaxone or cefotaxime is recommended for gram-negative bacteria coverage. If *Pseudomonas* is likely, then cefepime, piperacillin-tazobactam, or a carbapenem such as imipenem or meropenem is recommended for gram-negative bacteria coverage. In patients at risk for highly resistant gram-negative infections (e.g., patients with prior known highly resistant infections or colonization), use of two empiric gram-negative antibiotics is recommended. Finally, in patients with undifferentiated sepsis with risk factors for MRSA (e.g., frequent health care exposure or hospital-onset sepsis), vancomycin or linezolid administration is recommended.

Optimization of antibiotic delivery, such as administering β-lactam antibiotics prior to vancomycin, prolonged infusion of β-lactam antibiotics after the initial infusion, and optimization of pharmacokinetics/pharmacodynamics, should be considered in consultation with trained pharmacy and infectious diseases experts. Routine empiric antifungal therapy use in patients with undifferentiated sepsis is not recommended. However, in patients at increased risk of fungal infection (e.g., recent abdominal surgery, parenteral nutrition, liver failure, diabetes, colonization of multiple anatomic sites with *Candida* spp.), empiric echinocandin administration is recommended. While the Surviving Sepsis Campaign does not provide recommendations on antiviral use, remdesivir or neuraminidase inhibitor (e.g., oseltamivir) administration in septic patients with severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) and influenza infection, respectively, should be considered.

In addition to providing appropriate and timely antibiotics, identifying sources of infection amendable to control is imperative. Examples include, but are not limited to, intraabdominal abscess, bowel perforation, pyelonephritis, cholangitis, and necrotizing skin and soft tissue infections that are amendable to surgical or procedural source control. Source control should occur as rapidly as possible. Furthermore, removal of indwelling catheters should occur if the catheter appears infected (e.g., erythema or purulence at the catheter insertion site), if endovascular infection (e.g., endocarditis) is suspected or documented, and in critically ill patients without a clear alternate source of infection. Finally, antibiotic stewardship to limit development of resistant organisms and other antibiotic-associated complications (e.g., *Clostridioides difficile* infection, hypersensitivity/allergic reaction, and acute kidney

TABLE 315-1 Site-Specific Empiric Antibiotic Recommendations

SITE OF INFECTION	INITIAL EMPIRIC THERAPY	OTHER CONSIDERATIONS
Pulmonary		
CAP	Multidrug therapy with a β-lactam (ampicillin + sulbactam, ceftriaxone, or ceftaroline) and a macrolide (azithromycin or clarithromycin) Monotherapy with a respiratory fluoroquinolone (levofloxacin or moxifloxacin)	Risk factors for MRSA and/or *Pseudomonas aeruginosa*: add vancomycin or linezolid for MRSA coverage; replace standard CAP therapy with antipseudomonal coverage such as piperacillin-tazobactam, cefepime, meropenem, or imipenem. Recommendation based on "local validation" of risk factors for community-onset MRSA or *P. aeruginosa* or prior isolation of these organisms in the previous year, particularly from respiratory specimens.
HAP/VAP	Multidrug therapy with vancomycin or linezolid and piperacillin-tazobactam, cefepime, ceftazidime, imipenem, meropenem, or aztreonam	Two antipseudomonal antibiotics from different classes (addition of fluoroquinolones, aminoglycosides, or polymyxins) if prior intravenous antibiotic use within 90 days for HAP/VAP and septic shock at time of VAP, ARDS preceding VAP, 5 or more days of hospitalization prior to VAP, or acute renal replacement therapy prior to VAP. If prior colonization with carbapenem-resistant Enterobacterales or KPC-producing organism, ceftazidime-avibactam and meropenem-vaborbactam should be considered, but further efficacy data are needed. Empiric regimens should be informed by local distribution of pathogens and their antimicrobial susceptibilities.
Central Nervous System		
Health care–associated ventriculitis and meningitis	Vancomycin and cefepime, ceftazidime, or meropenem	β-Lactam choice based on local in vitro susceptibility patterns. If carbapenem-resistant *Acinetobacter* is suspected, addition of meropenem and colistin or polymyxin B.
Meningitis	Vancomycin and ceftriaxone	Age >50, alcohol abuse, or immunocompromised: add ampicillin. Penetrating head trauma, CSF shunt, or postneurosurgery, vancomycin and cefepime, ceftazidime, or meropenem. Clinical presentation suggestive of *Rickettsia* or *Ehrlichia*, add doxycycline.
Skin and Soft Tissue		
Necrotizing fasciitis including Fournier gangrene	Multidrug therapy with vancomycin or linezolid and piperacillin-tazobactam, a carbapenem, or ceftriaxone and metronidazole	Prompt surgical consultation is recommended for patients with aggressive infections associated with signs of systemic toxicity or suspicion of necrotizing fasciitis or gas gangrene.
Nonpurulent cellulitis/erysipelas (severe)	Vancomycin and piperacillin-tazobactam	Emergent surgical inspection to rule out necrotizing process.
Purulent furuncle/carbuncle/abscess (severe)	Vancomycin, daptomycin, linezolid, telavancin, or ceftaroline	Incision and drainage as indicated.
Intraabdominal		
Community-onset extrabiliary (mild)	Cefoxitin, ertapenem, moxifloxacin, or tigecycline	Health care setting with high prevalence of ESBL-producing Enterobacterales or >20% of *Pseudomonas* resistant to ceftazidime, consider carbapenem or piperacillin-tazobactam.
Community-onset extrabiliary (severe)	Imipenem-cilastatin, meropenem, doripenem, or piperacillin-tazobactam	Health care associated: imipenem-cilastatin, meropenem, or piperacillin-tazobactam, levofloxacin, or cefepime each along with metronidazole, vancomycin added to each regimen.
Community-onset biliary (mild to moderate)	Cefazolin, cefuroxime, or ceftriaxone	Empiric therapy should be driven by local microbiologic data and source control performed as indicated.
Community-onset biliary (severe) or cholangitis	Imipenem-cilastatin, meropenem, or piperacillin-tazobactam, levofloxacin, or cefepime each in combination with metronidazole	
Genitourinary		
Acute pyelonephritis (IDSA archived)	Ceftriaxone, trimethoprim-sulfamethoxazole, or ciprofloxacin	Requiring hospitalization: intravenous fluoroquinolone, aminoglycoside, extended-spectrum cephalosporin, extended-spectrum penicillin, or carbapenem with choice of agents based on local resistance data. Do not use fluoroquinolone if >10% resistance prevalence or trimethoprim-sulfamethoxazole in areas of high resistance.

Abbreviations: ARDS acute respiratory distress syndrome; CAP, community acquired pneumonia; CSF, cerebrospinal fluid; HAP, hospital-acquired pneumonia; IDSA, Infectious Diseases Society of America; KPC, *Klebsiella pneumoniae* carbapenemase; ESBL, extended spectrum beta-lactamase; MRSA, methicillin-resistant *Staphylococcus aureus*; VAP, ventilator-associated pneumonia.

injury) requires regular reassessment of the patient's clinical course and available microbiologic data to guide narrowing the spectrum of antibiotic therapy and using shorter rather than longer courses of therapy.

NEUROLOGIC COMPLICATIONS

Encephalopathy, manifested by altered consciousness, cognition, or attention, ranging from mild delirium to coma, that is not attributable to an alternative etiology, is the most common neurologic complication of sepsis **(Table 315–2)**. Sepsis-associated encephalopathy occurs in over half of septic patients and is associated with increased mortality and long-term functional and neuropsychiatric sequalae among survivors. Septic patients may also experience hyper- or hypoactive delirium, seizure, stoke, central nervous system infection, or coma, not directly attributable to sepsis. Thus, further diagnostic evaluation is needed in septic patients with clinical findings indicative of seizures (i.e., electroencephalogram), stroke (i.e., brain imaging), or meningoencephalitis (i.e., lumbar puncture) to identify and treat these complications. Underlying factors contributing to sepsis-associated encephalopathy include decreased cerebral perfusion, microcirculatory and blood-brain barrier disruption, and exposure of brain parenchyma to circulating inflammatory mediators and oxidative stress resulting in neuronal injury, dysfunction, and death. Neuronal and glial apoptosis

TABLE 315-2 Organ-Specific Clinical Findings and Management

	EPIDEMIOLOGY	CLINICAL FINDINGS	DIAGNOSIS	TREATMENT
Neurologic	54% of septic patients develop encephalopathy	Altered consciousness, cognition, or attention; seizure, stroke, or meningism	EEG, brain imaging (MRI or CT), lumbar puncture	Early recognition and supportive care; treat underlying cause; correct metabolic and electrolyte abnormalities; limit neurotoxic agents and sedatives
Cardiovascular	25% of septic patients develop shock and half have myocardial dysfunction	Tachycardia, hypotension, skin mottling, prolonged capillary refill, oliguria, altered mental status	Invasive blood pressure monitoring, dynamic assessment of volume status, echocardiogram	Intravenous fluid resuscitation with balanced crystalloid, ~30 mL/kg; vasopressors for persistent hypotension, norepinephrine (first agent), vasopressin (second agent), epinephrine (third agent); hydrocortisone 200 mg/d if ongoing vasopressor requirement; consider adding dobutamine to norepinephrine or switch to epinephrine in septic shock patients with decreased LVEF; consider use of pulmonary artery catheter in patients with mixed septic and cardiogenic shock
Respiratory	7% of septic patients develop ARDS	Tachypnea, hypoxia, increased work of breathing	Chest x-ray or ultrasound with noncardiogenic bilateral infiltrates and Pa_{O_2}/F_{IO_2} <300 mmHg or Sp_{O_2}/F_{IO_2} ≤315	Maintain Sp_{O_2} 90–96%; use high-flow nasal canula in patients with adequate neurologic status; target plateau pressure <30 cmH$_2$O and tidal volume of 6–8 mL; consider rescue therapies in patients with refractory hypoxia
Genitourinary	67% of septic patients have acute kidney injury	Oligura or anuria	Elevated serum creatinine and blood urea nitrogen, acidemia, hyperkalemia	Treat underlying infection, maintain renal perfusion, avoid nephrotoxic agents; start renal replacement therapy for progressive acidemia, hyperkalemia, uremia, or volume overload; infuse sodium bicarbonate if renal failure and pH ≤7.2.
Gastrointestinal	50% of septic shock patients have liver dysfunction	Right upper quadrant pain, asterixis, jaundice	Elevated bilirubin, alkaline phosphatase, and transaminases; right upper quadrant ultrasound	Treat underlying infection, avoid hypotension and hepatoxic agents; stress ulcer prophylaxis for high-risk patients; enteral feeding if shock controlled within 48 h; parenteral feeding if nutrition goal not met within 7 days; insulin if blood glucose ≥180 g/dL
Hematologic	35% of septic shock patients have disseminated intravascular coagulation	Clinical or subclinical thrombosis or hemorrhage	Thrombocytopenia, increased fibrin split products, decreased fibrinogen, prolonged PT/INR	Administer cryoprecipitate for fibrinogen <150 mg/dL, platelets for platelet count ≤50 × 10^9/L and evidence of bleeding, and fresh frozen plasma for prolongated PT/INR and evidence of bleeding; transfuse packed red blood cells for hemoglobin <7.0 g/dL

Abbreviations: ARDS, acute respiratory distress syndrome; CT, computed tomography; EEG, electroencephalogram; INR, international normalized ratio; LVEF, left ventricular ejection fraction; MRI, magnetic resonance imaging; PT, prothrombin time.

are correlative findings at autopsy among fatal cases. The management of sepsis-associated encephalopathy includes early recognition, treatment of the underlying infection, supportive care including correction of metabolic or electrolyte abnormalities, and limiting exposure to pharmacologic agents that are neurotoxic (e.g., cefepime) or have central nervous system effects including opiates and benzodiazepines that might exacerbate the condition or predispose to long-term disability or neurocognitive dysfunction.

CARDIOVASCULAR DYSFUNCTION

Cardiovascular comprise, manifested as hypotension or shock, occurs in ~25% of septic patients. Hypotension results from peripheral arteriolar vasodilation and decreased cardiac venous return due to venous vasodilation and intra- to extravascular fluid shifts. Clinical signs of diminished tissue perfusion include skin mottling, prolonged capillary refill time, oliguria, and altered mental status. During early compensated shock, heart rate and cardiac output increase. As septic shock progresses, loss of vascular smooth muscle contractility, despite endogenous neurohormonal stimuli and exogenous catecholamine administration, results in progressive or refractory shock.

Up to half of patients with septic shock also have myocardial dysfunction, which is associated with increased mortality. Sepsis-induced cardiomyopathy manifests as decreased left ventricular ejection fraction, increased end-diastolic volume index, and right ventricular impairment. Contributing factors to sepsis-induced cardiomyopathy include global ischemia, hypoxia, and impaired myocardial metabolism; endothelial damage, increased adhesion molecule expression, and coronary microcirculatory dysfunction; and direct cardiomyocyte suppression, mitochondrial dysfunction, or cardiomyocyte death from exposure to inflammatory mediators including cytokines (e.g., IL-1β, IL-6, and TNF-α), complement proteins, and NO.

Intravenously administered fluids and vasopressors are used to restore and maintain blood pressure and tissue perfusion in the setting of septic shock. An indwelling arterial catheter should be placed as soon as feasible to continuously monitor blood pressure invasively. Crystalloid solutions are preferred over colloid for initial volume resuscitation, although colloid such as albumin can be considered in patients who have already received large volumes of crystalloid. Balanced crystalloid solution, such as lactated Ringer's, may be preferrable to 0.9% normal saline, which is more likely to induce hyperchloremic metabolic acidosis associated with renal vasoconstriction and kidney injury. Use of pentastarch or hydroxyethyl starch is associated with severe kidney injury and death and should be avoided.

Crystalloid should be administered as a bolus over 5–10 min at a volume of ~30 mL/kg. However, 30 mL/kg may not be appropriate for all patients, including those with end-stage renal disease and systolic heart failure. Protocolized fluid resuscitation targets including central venous pressure (CVP) of 8–12 mmHg, central venous oxygen saturation >70%, and urine output ≥0.5 mL/kg per h have not been associated with improved mortality and so are not recommended to guide total fluid replacement. Instead, dynamic assessment of volume responsiveness using capillary refill time, passive leg raise maneuver, and point-of-care ultrasound, with iterative reevaluation, should guide total volume administered and is preferred over static measurements of volume responsiveness such as CVP. Notably, a 2017–2018 clinical trial conducted in 28 ICUs in five countries found that resuscitation guided by capillary refill time compared with lactate level–targeted resuscitation did not reduce all-cause 28-day mortality, an important finding for sepsis care in resource-limited settings.

In patients who have received adequate fluid resuscitation yet remain hypotensive, a continuous norepinephrine drip should be initiated as the first-line vasopressor to maintain a mean arterial pressure target of ≥65 mmHg. Vasopressin should be added as a second agent, and epinephrine as a third agent, if needed, to achieve the blood pressure target. A contemporary randomized clinical trial tested the efficacy of early restrictive (i.e., prioritizing vasopressors and lower intravenous fluid volumes) versus liberal (i.e., prioritizing higher volumes of intravenous fluids before vasopressors use) fluid management in the first 4 h of presentation in patients with septic shock. Across 1563 patients at 60 U.S. medical centers, there was no difference in 90-day mortality. In septic patients with evidence of myocardial dysfunction (e.g., low cardiac output, elevated filling pressures) and evidence of persistent hypoperfusion following adequate volume resuscitation, consideration should be given to adding dobutamine to norepinephrine or using epinephrine alone to increase inotropy. While routine use of pulmonary artery catheters to guide fluid management in sepsis has not been associated with improved outcomes, use may be considered in patients with mixed septic and cardiogenic shock.

In patients with septic shock and ongoing requirement for vasopressor therapy, it is recommended to start intravenous corticosteroids with hydrocortisone at a dose of 200 mg/d often provided as 50 mg every 6 h.

■ ACUTE LUNG INJURY

Lung injury, manifesting as acute hypoxic respiratory insufficiency or failure, and termed acute respiratory distress syndrome (ARDS), complicates ~7% of sepsis cases. ARDS is a syndrome that is classified by noncardiogenic diffuse pulmonary infiltrates and hypoxemia, in which infiltrates can be determined by x-ray or ultrasound imaging, and hypoxemia can be determined by $\text{Pao}_2/\text{Fio}_2$ <300 mmHg or percent oxygen saturation (Spo_2)/Fio_2 ≤315. The pathogenesis of ARDS overlaps with that of sepsis, in which activated myeloid cells and soluble inflammatory mediators disrupt pneumocytes' and endothelial cells' structure and function, resulting in leakage of plasma components and further recruitment of immune cells into lung alveolar and interstitial spaces. The histopathologic correlate of ARDS is diffuse alveolar damage, which progresses through early exudative, then proliferative, and finally fibrotic stages.

Initial management of acute lung injury in sepsis requires administration of oxygen to meet cellular metabolic demand while limiting cellular injury from oxidative stress, with a reasonable Spo_2 target being 90–96%. In patients with adequate neurologic status, and absent other specific contraindications, high-flow nasal canula is preferred over noninvasive ventilation to improve hypoxia in patients without hypercapnia. For patients with ARDS who require invasive mechanical ventilation, targeting a plateau pressure of <30 cmH_2O and a tidal volume of 6–8 mL/kg, based on ideal body weight, has been associated with reduced mortality and is recommended. In patients with refractory hypoxia despite low tidal volume ventilation, rescue measures including prone positioning, neuromuscular blockade, and venovenous extracorporeal membrane oxygenation should be considered.

■ ACUTE KIDNEY INJURY

Acute kidney injury (AKI) occurs in up to two-thirds of patients with sepsis or septic shock and is associated with increased mortality and risk of chronic kidney disease and disability among survivors. Sepsis-associated AKI presents with oliguria or anuria and elevated serum creatinine and blood urea nitrogen and accounts for all AKI cases occurring within 7 days of sepsis onset. Sepsis-associated AKI may be attributed to sepsis, associated complications (e.g., abdominal compartment syndrome), and clinical management including administration of nephrotoxic substances such as antibiotics and intravenous contrast dye. Factors that contribute to pathogenesis of AKI due to sepsis alone, termed sepsis-induced AKI, include renal hypoperfusion, microvascular injury and dysfunction, inflammatory cellular and soluble mediators, and altered renal tubular metabolic and mitochondrial function. Many serum and urine biomarkers have been evaluated to improve early recognition and treatment of sepsis-associated AKI. However, to date, none have been shown to improve outcomes.

Management of sepsis-associated AKI includes treating underlying infection, maintaining renal perfusion with fluid resuscitation and vasopressors to achieve blood pressure goals, avoiding nephrotoxic agents when possible, and identifying and treating reversible causes. Initiating early renal replacement therapy (RRT), including intermittent hemodialysis or continuous renal replacement therapy (CRRT), during sepsis-associated AKI is not associated with improved outcomes, and so RRT should be started for standard definitive indications including progressive acidemia, hyperkalemia, uremia, or volume overload. CRRT is better tolerated than intermittent hemodialysis in patients with septic shock, allowing for less dynamic intravascular volume shifts and so is preferred in these patients when available. Sodium bicarbonate infusion should be considered only in patients with AKI and severe metabolic acidemia (pH ≤7.2) to maintain vasopressor effectiveness and mitigate fatal ventricular arrythmia risk.

■ GASTROINTESTINAL COMPLICATIONS

Approximately 50% of patients with septic shock develop liver dysfunction, which is associated with increased mortality. Clinical findings may include right upper quadrant pain, jaundice, and asterixis depending on severity. Common laboratory findings include elevated serum bilirubin and alkaline phosphatase levels and elevated transaminases if marked hypotension occurs. Sepsis-induced cholestasis is attributed to impaired bile formation and decreased flow in a nonobstructive pattern. The liver plays an essential role in microbial clearance, in which the Kupffer cells phagocytose bacteria, release proinflammatory cytokines and chemokines, and bind platelets. Neutrophils are recruited to liver sinusoids, release NETs to trap pathogens, and contribute to a proinflammatory and prothrombotic environment. In autopsy studies of fatal sepsis, hepatitis and steatosis are detected in most patients, while portal inflammation, centrilobular necrosis, hepatocellular apoptosis, and cholangitis may also be detected. Management of liver injury and dysfunction during sepsis includes treating the underlying infection and avoiding hypotension and hepatoxic medications. In septic patients with elevated serum bilirubin, abdominal ultrasound should be performed to evaluate for biliary obstruction, cholecystitis, and cholangitis.

Sepsis has also been associated with alteration of the intestinal mucosa including increased epithelial permeability, changes in gut microbiota, and translocation of enteric microbes into circulation. In patients with sepsis and septic shock and risk factors for gastrointestinal bleeding (e.g., mechanical ventilation, coagulopathy, preexisting liver disease, high organ failure score, and need for RRT), stress ulcer prophylaxis is recommended. Decreased splanchnic perfusion and increased interstitial edema contribute to impaired intestinal absorption, which in combination with the catabolism of sepsis contributes to nutritional deficiencies. The route and timing of supplemental nutrition administration in patients with sepsis and septic shock are debated. Current recommendations are to initiate enteral nutrition early, within 48 h of diagnosis, in patients in whom shock is controlled with fluids and vasopressors and absent contraindications like bowel ischemia. The rationale for early enteral feeding is to maintain enteric epithelial integrity while limiting negative nitrogen and caloric balance. Supplemental use of parenteral nutrition is suggested if nutritional goals are unmet within 7 days by enteric feeding and supplemental glucose infusion. Insulin therapy is recommended for patients with blood glucose levels >180 mg/dL.

■ HEMATOLOGIC COMPLICATIONS

Coagulation abnormalities in sepsis are common, ranging from isolated thrombocytopenia to disseminated intravascular coagulation (DIC), and can manifest with clinical or subclinical thrombosis and hemorrhage. Up to 35% of septic shock patients meet DIC criteria, including thrombocytopenia, increased fibrin split products, decreased fibrinogen, and prolonged prothrombin time (PT)/international normalized ratio (INR), which is associated with increased mortality. Given that DIC is a late manifestation of coagulation abnormalities in sepsis, a scoring system for earlier detection of sepsis-induced coagulopathy (SIC) has been proposed. SIC scoring, which considers platelet counts, PT, and SOFA score, is more sensitive than DIC criteria for recognizing coagulation abnormalities in sepsis. However, the clinical utility of SIC for guiding interventions or improving outcomes in sepsis is unproven. The pathogenesis of coagulation abnormalities in sepsis, as discussed above, is attributed to activation of innate immune cellular and soluble responses including cytokines, chemokines, and complement; endothelial activation and injury; platelet activation and aggregation; and clotting cascade induction, suppression of antithrombotic molecules, and activation of antifibrinolytic molecules. This procoagulant milieu results in platelet and clotting factor consumption, which paradoxically increases bleeding risk. Current recommendations for DIC management focus on mitigating bleeding risk. Recommendations include administration of cryoprecipitate to patients with fibrinogen <150 mg/dL; administration of fresh frozen plasma to patients with prolonged PT/INR and evidence of bleeding; and administration of platelets to patients with platelet counts $\leq 50 \times 10^9$/L and evidence of bleeding. Platelet transfusion thresholds are lower in patients at increased bleeding risk including patients undergoing chemotherapy, post–hemopoietic stem cell transplantation patients, and postsurgical patients. Venous thromboembolism prophylaxis, with low-molecular-weight heparin preferred over unfractionated heparin, is recommended in patients with sepsis or septic shock. Full-dose anticoagulation is not recommended for prophylactic use in septic patients, but instead is reserved for standard treatment indications including deep-venous thrombosis and pulmonary embolism.

Anemia is common in septic patients, and transfusion of packed red blood cells to maintain hemoglobin >7.0 g/dL is recommended. However, a more conservative, higher hemoglobin target may be appropriate in some patients based on individual factors including degree of hypoxemia, myocardial ischemia, and active hemorrhage.

HOST-TARGETED THERAPIES

Despite years of effort and many promising preclinical studies, there are currently no U.S. Food and Drug Administration–approved therapies targeting pathologic immune responses during bacterial sepsis. Failed therapeutic targets include proinflammatory mediators (e.g., anti-TNF, anti-IL-1, anti-TLR-4, anti-C5a), components of the coagulation cascade (e.g., antithrombin III, activated protein C, thrombomodulin), and many others. Other adjunctive therapies that have been evaluated include polymyxin-B hemoperfusion, intravenous immunoglobulin, and vitamin C, all determined to be not beneficial. Future work aimed at developing host-targeted sepsis therapies will require a deeper understanding of cellular and soluble mediators contributing to its pathogenesis in blood and tissues and how these mediators vary across host, pathogen, and timing of infection. Emerging technologies including single-cell transcriptomics, proteomics, metabolomics, and spatial-transcriptomics should aid in identifying novel therapeutic targets, and innovative, adaptive clinical trial designs will help stratify heterogeneous septic patients into likely responders to specific therapies.

DEESCALATING CARE AND LIMITING LONG-TERM SEQUALAE

In sepsis survivors who remain hospitalized, care should focus on limiting complications and optimizing long-term outcomes. Indwelling central venous and urinary catheters should be removed when no longer needed. Early mobilization, deep-venous thrombosis prevention, discontinuing unnecessary intravenous fluids, and judicious use of diuretics in patients with significant fluid overload are all important interventions. Many sepsis survivors experience long-term complications including physical, cognitive, and psychological sequelae. Physical sequelae include prolonged fatigue, muscle loss, weakness, and diminished functional capacity. Cognitive and psychological sequelae include cognitive decline, dementia, depression, and decreased quality of life. Sepsis survivors have increased risk of cardiovascular events, including myocardial infarction and stroke, recurrent infection, readmission, and death. Fifty percent of initial sepsis survivors are rehospitalized within 1 year, and one in six die within the first year. Most deaths following sepsis occur in the first 6 months, but the risk of death remains elevated for up to 2 years. Common causes of readmission include heart failure, myocardial infarction, pneumonia, chronic obstructive pulmonary disease, and urinary tract infections. Given the high prevalence of long-term sequelae and complications among sepsis survivors, the 2021 Surviving Sepsis Campaign Guidelines recommend that hospital discharge plans include screening for economic and social support and establishing follow-up with providers who can assess and support physical, cognitive, and psychological issues.

FURTHER READING

Baghela A et al: Predicting sepsis severity at first clinical presentation: The role of endotypes and mechanistic signatures. EBioMedicine 75:103776, 2022.

Evans L et al: Surviving sepsis campaign: International guidelines for management of sepsis and septic shock 2021. Crit Care Med 49:e1063, 2021.

Habimana R et al: Sepsis-induced cardiac dysfunction: A review of pathophysiology. Acute Crit Care 35:57, 2020.

Raia L, Zafrani L: Endothelial activation and microcirculatory disorders in sepsis. Front Med (Lausanne) 9:907992, 2022.

Rhee C et al: Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. JAMA 318:1241, 2017.

Rhee C et al: Prevalence of antibiotic-resistant pathogens in culture-proven sepsis and outcomes associated with inadequate and broad-spectrum empiric antibiotic use. JAMA Netw Open 3:e202899, 2020.

Singer M et al: The Third International Consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 315:801, 2016.

Strich JR et al: Considerations for empiric antimicrobial therapy in sepsis and septic shock in an era of antimicrobial resistance. J Infect Dis 222:S119, 2020.

Wiersinga WJ, van der Poll T: Immunopathophysiology of human sepsis. EBioMedicine 86:104363, 2022.

Zarbock A et al: Sepsis-associated acute kidney injury: Consensus report of the 28th Acute Disease Quality Initiative workgroup. Nat Rev Nephrol 19:401, 2023.

316

Cardiogenic Shock and Pulmonary Edema

David H. Ingbar, Holger Thiele

Cardiogenic shock and pulmonary edema are each life-threatening high-acuity conditions that require treatment as medical emergencies, usually in an intensive care unit (ICU) or cardiac intensive care unit (CICU). The most common joint etiology is severe left ventricular (LV) dysfunction from myocardial infarction (MI) that leads to pulmonary congestion and/or systemic hypoperfusion **(Fig. 316-1)**. The pathophysiologies of pulmonary edema and shock are discussed in **Chaps. 39 and 314**, respectively.

CARDIOGENIC SHOCK

Cardiogenic shock (CS) is a low cardiac output state resulting in life-threatening end-organ hypoperfusion and hypoxia. The clinical presentation is typically characterized by persistent hypotension (<90 mmHg systolic blood pressure [BP]) or <60–65 mmHg mean arterial pressure that is unresponsive to volume replacement and/or requires use of vasopressors to maintain adequate BP) and is accompanied by clinical features of peripheral hypoperfusion, such as elevated arterial lactate (>2 mmol/L). Objective hemodynamic parameters such as cardiac index or pulmonary capillary wedge pressure can help confirm a cardiogenic cause of shock but are not mandatory. The in-hospital mortality rates range from 40 to 60%, depending on shock severity and the associated underlying cause. The Society for Cardiovascular Angiography and Interventions (SCAI) classification of CS

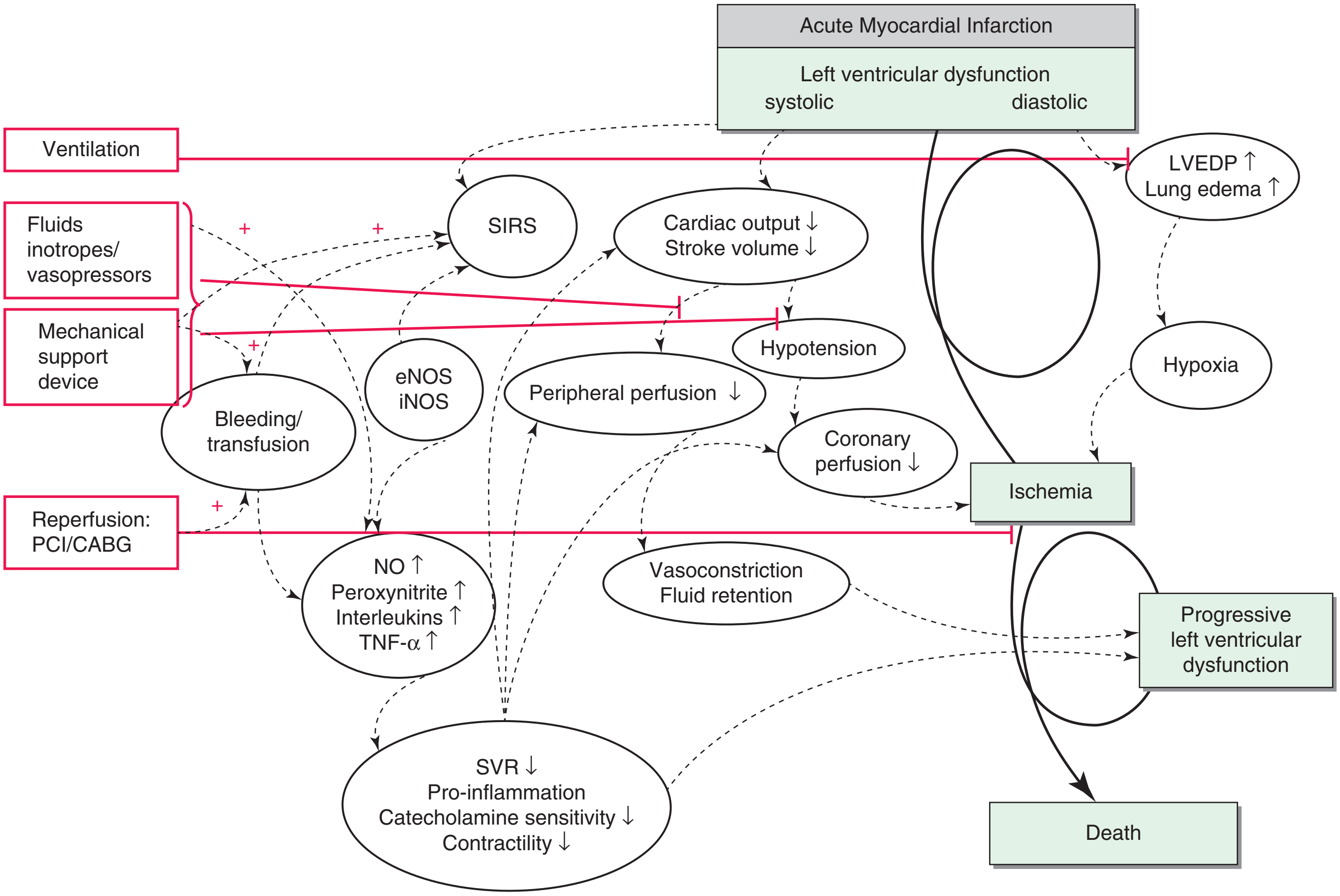
Acute Myocardial Infarction
Left ventricular dysfunction
systolic
diastolic
Ventilation
LVEDP ↑
Lung edema ↑
Fluids
inotropes/
vasopressors
Mechanical
support
device
SIRS
Cardiac output ↓
Stroke volume ↓
+
+
+
+
eNOS
iNOS
Hypotension
Hypoxia
Peripheral perfusion ↓
Bleeding/
transfusion
Coronary
perfusion ↓
Ischemia
Reperfusion:
PCI/CABG
NO ↑
Peroxynitrite ↑
Interleukins ↑
TNF-α ↑
Vasoconstriction
Fluid retention
Progressive
left ventricular
dysfunction
SVR ↓
Pro-inflammation
Catecholamine sensitivity ↓
Contractility ↓
Death

FIGURE 316-1 Pathophysiology of cardiogenic shock and potential treatment targets. The pathophysiologic concept of the expanded cardiogenic shock spiral and treatment targets. CABG, coronary artery bypass grafting; eNOS, endothelial nitric oxide synthase; iNOS, inducible nitric oxide synthase; LVEDP, left ventricular end-diastolic pressure; NO, nitric oxide; PCI, percutaneous coronary intervention; SIRS, systemic inflammatory response syndrome; SVR, systemic vascular resistance; TNF, tumor necrosis factor. *(Reproduced with permission from H Thiele et al: Shock in acute myocardial infarction: The Cape Horn for trials? Eur Heat J 31:1828, 2010.)*

that was introduced in 2019 includes five categories: (A) at risk, (B) beginning or preshock, (C) classical, (D) deteriorating, and (E) extremis CS **(Fig. 316-2)**. Preshock is defined as clinical evidence of relative hypotension or tachycardia without hypoperfusion. These patients should be monitored closely and treated early to avoid development of classical CS. Extremis CS includes cases in which considerations about futility of treatment should be done and possibly palliative care initiated. The SCAI definition recently was updated based on several validation studies; it still includes the stages A–E but also includes a three-axis model based on (1) shock severity, (2) phenotype and etiology, and (3) risk modifiers such as cardiac arrest.

Although declining in incidence, acute MI with LV dysfunction remains the most frequent cause of CS, with other causes listed in **Table 316-1**. Circulatory failure based on cardiac dysfunction may be caused by primary myocardial failure, most commonly secondary to acute MI **(Chap. 286)**, and less frequently by acute or chronic heart failure as a cause of cardiomyopathy or myocarditis **(Chaps. 266–270)**, cardiac tamponade **(Chap. 281)**, arrhythmias **(Chap. 261)**, or critical valvular heart disease **(Chap. 262)**.

Incidence The incidence of CS complicating acute MI has decreased to 5–10%, largely due to increasing use of early mechanical reperfusion therapy for acute MI. Shock is more common with ST-segment elevation MI (STEMI) than with non-STEMI **(Chap. 286)**.

LV failure accounts for ~80% of cases of CS complicating acute MI. Acute severe mitral regurgitation (MR), ventricular septal rupture (VSR), predominant right ventricular (RV) failure, and free wall rupture or tamponade account for the remainder. A recently recognized uncommon cause of transient CS is the Takotsubo syndrome.

Pathophysiology The understanding of the complex pathophysiology of CS has evolved over the past decades. In general, a profound depression of myocardial contractility results in a deleterious spiral of reduced cardiac output, low BP, and ongoing myocardial ischemia, followed by further contractility reduction (Fig. 316-1). This vicious cycle usually leads to death if not interrupted. CS can result in both acute and subacute derangements to the entire circulatory system. Hypoperfusion of vital organs and extremities remains a clinical hallmark. Although ineffective stroke volume is the inciting event, inadequate circulatory compensation also may contribute to shock. Initial peripheral vasoconstriction may improve coronary and peripheral perfusion at the cost of increased afterload, potentially worsening ischemia. However, over the course of CS, the systemic inflammation response triggered by acute cardiac injury often induces pathologic vasodilatation. Inflammatory cytokines and inducible (as well as endothelial) nitric oxide (NO) synthase may augment production of NO and its by-product, peroxynitrite, which has a negative inotropic effect and is cardiotoxic. Lactic acidosis and hypoxemia contribute to the vicious circle, as severe acidosis reduces the efficacy of endogenous and exogenous catecholamines. During ICU or CICU support, bleeding and/or transfusions may trigger inflammation and are usually associated with higher mortality (Fig. 316-1).

Patient Profile In patients with MI, older age, prior MI, diabetes mellitus, anterior MI location, and multivessel coronary artery disease with extensive coronary artery stenoses are associated with an increased risk of CS. Shock associated with a first inferior MI should prompt a search for a mechanical cause or RV involvement. CS may rarely occur in the absence of significant coronary stenosis, as seen in Takotsubo syndrome or fulminant myocarditis.

Timing Shock is present on admission in approximately one-quarter of MI patients who develop CS; of the remaining patients, one-quarter develop it rapidly thereafter, within 6 h of MI onset, and

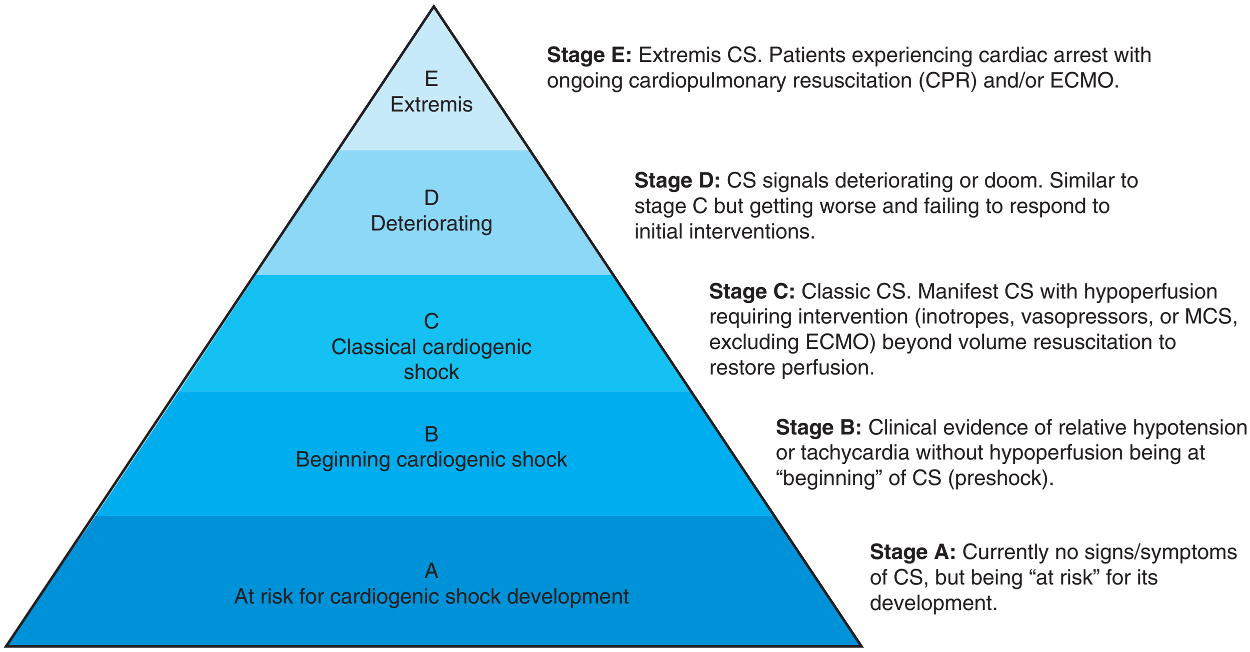
E
Extremis
D
Deteriorating
C
Classical cardiogenic shock
B
Beginning cardiogenic shock
A
At risk for cardiogenic shock development
Stage E: Extremis CS. Patients experiencing cardiac arrest with ongoing cardiopulmonary resuscitation (CPR) and/or ECMO.
Stage D: CS signals deteriorating or doom. Similar to stage C but getting worse and failing to respond to initial interventions.
Stage C: Classic CS. Manifest CS with hypoperfusion requiring intervention (inotropes, vasopressors, or MCS, excluding ECMO) beyond volume resuscitation to restore perfusion.
Stage B: Clinical evidence of relative hypotension or tachycardia without hypoperfusion being at "beginning" of CS (preshock).
Stage A: Currently no signs/symptoms of CS, but being "at risk" for its development.

FIGURE 316-2 Shock severity definition. Five categories of cardiogenic shock (CS). **Stage A:** At risk: Patients "at risk" for cardiogenic shock development but not currently experiencing signs/symptoms of cardiogenic shock. **Stage B:** Patients with clinical evidence of relative hypotension or tachycardia without hypoperfusion being at "beginning" of cardiogenic shock. **Stage C:** Patients in the state of "classic" cardiogenic shock. **Stage D:** Cardiogenic shock signals deteriorating or "doom." **Stage E:** Patients in "extremis," such as those experiencing cardiac arrest with ongoing cardiopulmonary resuscitation and/or extracorporeal membrane oxygenation (ECMO) cardiopulmonary resuscitation. MCS, mechanical circulatory support. *(Reproduced with permission from H Thiele et al: Management of cardiogenic shock complicating myocardial infarction: An update 2019. Eur Heart J 40:2671, 2019.)*

another quarter develop shock later on the first day. Later onset of CS may be due to reinfarction, marked infarct expansion, or mechanical complications.

Diagnosis For these unstable patients, supportive therapy must be initiated simultaneously with diagnostic evaluation (Fig. 316-3). A focused history and physical examination should be performed along with an electrocardiogram (ECG), chest x-ray, arterial blood gas (ABG) analysis, lactate measurement, and blood specimens for laboratory analysis. Initial echocardiography is an invaluable tool to elucidate the underlying cause of CS and also assess if it is predominantly left, right, or biventricular in origin.

CLINICAL FINDINGS Most patients initially are dyspneic, pale, apprehensive, and diaphoretic, and mental status may be altered. The pulse is typically weak and rapid, or occasionally, severe bradycardia due to high-grade heart block may be present. Systolic BP is typically reduced (<90 mmHg, or catecholamines are required to maintain BP >90 mmHg), but occasionally, BP may be maintained by very high systemic vascular resistance. Tachypnea and jugular venous distention may be present. Typically, there is a weak apical pulse and a soft S_1, and an S_3 gallop may be audible. Acute, severe MR and VSR usually are associated with characteristic systolic murmurs (Chap. 286). Crackles are audible in most patients with LV failure. Oliguria/anuria is common. CS patients often require early mechanical ventilation (~80%) for management of acute hypoxemia, increased work of breathing, and hemodynamic instability; vasopressors often are required to maintain adequate BP.

LABORATORY FINDINGS The white blood cell count and C-reactive protein typically are elevated. Renal function often is progressively impaired. Newer renal function markers such as cystatin C or neutrophil gelatinase–associated lipocalin (NGAL) do not add prognostic information over creatinine. Hepatic transaminases are elevated due to liver hypoperfusion in ~20% of patients, which is a marker of high mortality. By definition, in SCAI shock criteria, the arterial lactate level is elevated to >2 mmol/L; if higher, prognosis is worse, and those with lactate >8 mmol/L are in SCAI stage E. ABGs usually demonstrate hypoxemia and an anion gap metabolic acidosis. Glucose levels at admission are often elevated, a strong independent predictor for mortality. Cardiac markers, creatine kinase and its MB fraction, and troponins I and T are typically markedly elevated in acute MI.

ELECTROCARDIOGRAM In acute MI with CS, Q waves and/or ST elevation in multiple leads or left bundle branch block are usually present. Approximately one-half of MIs with CS are anterior infarctions. Global ischemia due to severe left main stenosis usually is accompanied by ST-segment elevation in lead aVR and ST depressions in multiple leads.

CHEST ROENTGENOGRAM The chest x-ray typically shows pulmonary vascular congestion and often pulmonary edema but may be normal in up to a third of patients. The heart size is usually normal when CS results from a first MI but may be enlarged when it occurs in a patient with a previous MI.

ECHOCARDIOGRAM An echocardiogram (Chap. 248) should be obtained promptly in patients with suspected/confirmed CS to help define its etiology. Echocardiography is able to delineate the extent of infarction/myocardium in jeopardy and the presence of mechanical complications such as VSR, MR, or cardiac tamponade. Furthermore, RV impairment, valvular obstruction or insufficiency, dynamic LV outflow tract obstruction, and proximal aortic dissection with aortic regurgitation or tamponade may be seen, or indirect evidence for pulmonary embolism may be obtained (Chap. 290) (Table 316-2).

PULMONARY ARTERY CATHETERIZATION The use of pulmonary artery catheter (PAC) hemodynamic monitoring had declined until recently because clinical trials have shown no mortality benefit. The recent increase in PAC use arose because hemodynamic data and waveforms can be helpful in both diagnosis and management. Recent observational data suggest better outcome with PAC use applied in this way. PAC hemodynamic data can confirm the presence and severity of CS, involvement of the right ventricle, left-to-right shunting, pulmonary artery pressures and transpulmonary gradient, and pulmonary and systemic vascular resistance. It can help in recognition of acute MR, decreased left atrial filling pressure, right or left dominance, and secondary septic causes and also can exclude left-to-right shunts. Equalization of diastolic pressures suggests cardiac tamponade, but echocardiogram is more definitive. The detailed hemodynamic profile can be used to individualize and monitor therapy and to provide prognostic information, such as cardiac index and cardiac power. The use of a PAC is currently recommended by the American Heart Association for potential utilization in cases of diagnostic or CS management uncertainty or in patients with severe CS who are unresponsive

TABLE 316-1 Etiologies of Cardiogenic Shock[a] and Cardiogenic Pulmonary Edema

Etiologies of Cardiogenic Shock or Pulmonary Edema
Acute myocardial infarction/ischemia
Left ventricular failure
Ventricular septal rupture
Papillary muscle/chordal rupture–severe mitral regurgitation
Ventricular free wall rupture
Other conditions complicating large myocardial infarctions
Excess negative inotropic or vasodilator medications
Post–cardiac arrest
Postcardiotomy
Refractory sustained supraventricular or ventricular tachyarrhythmias
Refractory sustained bradyarrhythmias
Acute fulminant myocarditis
End-stage cardiomyopathy
Takotsubo syndrome/apical ballooning syndrome
Hypertrophic cardiomyopathy with severe outflow obstruction
Aortic dissection with aortic insufficiency or tamponade
Severe valvular heart disease
Critical aortic or mitral stenosis
Acute severe aortic regurgitation or mitral regurgitation
Toxic/metabolic
β Blocker or calcium channel antagonist overdose
Pheochromocytoma
Scorpion venom
Hypertensive crisis
Post–cardiac arrest stunning
Myocardial depression in setting of septic shock or systemic inflammatory response syndrome
Myocardial contusion
Other Etiologies of Cardiogenic Shock[b]
Right ventricular failure due to:
Acute myocardial infarction
Acute or decompensated chronic cor pulmonale
Pericardial tamponade
Toxic/metabolic
Severe acidosis, severe hypoxemia

[a]The etiologies of cardiogenic shock are listed. Most of these can cause pulmonary edema instead of shock or pulmonary edema with cardiogenic shock. [b]These cause cardiogenic shock but not pulmonary edema.

to initial therapy. PAC use also can help differentiate noncardiogenic pulmonary edema.

ADVANCED HEMODYNAMIC MONITORING Recently, new central venous catheter systems linked to computer-based algorithms provide continuous monitoring of a variety of derived hemodynamic parameters, including cardiac output, stroke volume, stroke volume variation, and systemic vascular resistance **(Table 316-3)**. When combined with a femoral arterial catheter, calculated extravascular lung water and pulmonary permeability index can be monitored. The information allows for more rational therapy and assessment but has not yet shown improved clinical outcomes in patients with shock or pulmonary edema.

CARDIAC CATHETERIZATION AND CORONARY ANGIOGRAPHY The definition of the coronary anatomy provides useful information and is immediately indicated in all patients with CS complicating MI for further reperfusion treatment. Furthermore, cardiac catheterization should also be considered for resuscitated cardiac arrest survivors without ST-segment elevation in CS because ~70% of these patients have relevant coronary artery disease. However, routine early invasive coronary angiography did not show a survival benefit in hemodynamically stable patients after resuscitation from cardiac arrest without ST-segment elevation in two recent large, randomized trials. Consequently, guidelines were revised to avoid routine immediate cardiac catheterization in these patients.

TREATMENT

Acute Myocardial Infarction

GENERAL MEASURES

In addition to the usual treatment of acute MI **(Chap. 286)**, initial therapy is aimed at maintaining adequate systemic and coronary perfusion by raising the BP with vasopressors and adjusting volume status to a level that ensures optimum LV filling pressure (Fig. 316-3). There is some interpatient variability, but generally, adequate perfusion occurs with a mean arterial BP of 60–65 mmHg or a systolic BP of ~90 mmHg. Hypoxemia and acidosis need to be corrected, particularly since acidemia attenuates vasoconstriction by catecholamines. Up to 90% of patients require ventilatory support, decreasing the stress from increased work of breathing (see "Pulmonary Edema," below) (Fig. 316-3). Moderate glucose control (≤180 mg/dL or 10.0 mmol/L) should be a goal, and hypoglycemia must be avoided. Negative ionotropic agents should be discontinued. Bradyarrhythmias may require transvenous pacing. Recurrent ventricular tachycardia or rapid atrial fibrillation may require immediate treatment **(Chap. 253)**.

REPERFUSION-REVASCULARIZATION

Rapid revascularization of the infarct-related artery is the only evidence-based treatment strategy for mortality reduction in CS and forms the mainstay therapeutic intervention for CS due to MI (Fig. 316-2). In the SHOCK trial, 132 lives were saved per 1000 patients treated with early revascularization with percutaneous coronary intervention (PCI) or coronary artery bypass graft (CABG) compared with initial medical therapy. Outcome benefit correlates strongly with the time between symptom onset, first medical contact, and reperfusion. In general, PCI with drug-eluting stents of the infarct-related artery is the preferred reperfusion strategy. Approximately 80% of CS patients present with multivessel coronary artery disease. In these patients, culprit-only PCI with possible staged revascularization is the method of choice because it reduces mortality and requirement for renal replacement therapy at 30 days and 1 year in comparison to immediate multivessel PCI, as shown in the CULPRIT-SHOCK trial. The major driver for the reduction in the composite endpoint was a reduction in 30-day mortality. Updated recent clinical practice guidelines recommend avoiding immediate nonculprit PCI. Currently, vascular access for diagnostic angiography and PCI via the radial artery are preferred when feasible over femoral arterial access due to the greater safety of radial artery access. CABG is currently performed in only 5% of cases, mainly if coronary anatomy is not amenable to PCI.

VASOPRESSORS AND INOTROPES

Inotropic agents are theoretically appealing in CS treatment. However, current evidence is scarce. Vasoactive medications often are used in the management of patients with CS, and all have important disadvantages, including increases in myocardial oxygen consumption, afterload, lethal arrhythmias, and possible myocardial cell death. As a consequence, catecholamines should be used in the lowest possible doses for the shortest possible time. Despite their frequent use, little clinical outcome data prove their benefit or are available to guide the initial selection of vasoactive therapies in patients with CS. No vasopressor has been demonstrated to change outcome in large clinical trials. Norepinephrine is reasonable as the first-line vasopressor based on randomized trials compared to dopamine and also epinephrine. Norepinephrine was associated with fewer adverse events, including arrhythmias, compared to dopamine in a randomized trial of patients with several etiologies of circulatory shock and with improved survival in a prespecified subgroup of CS patients. Norepinephrine dosing is usually begun at

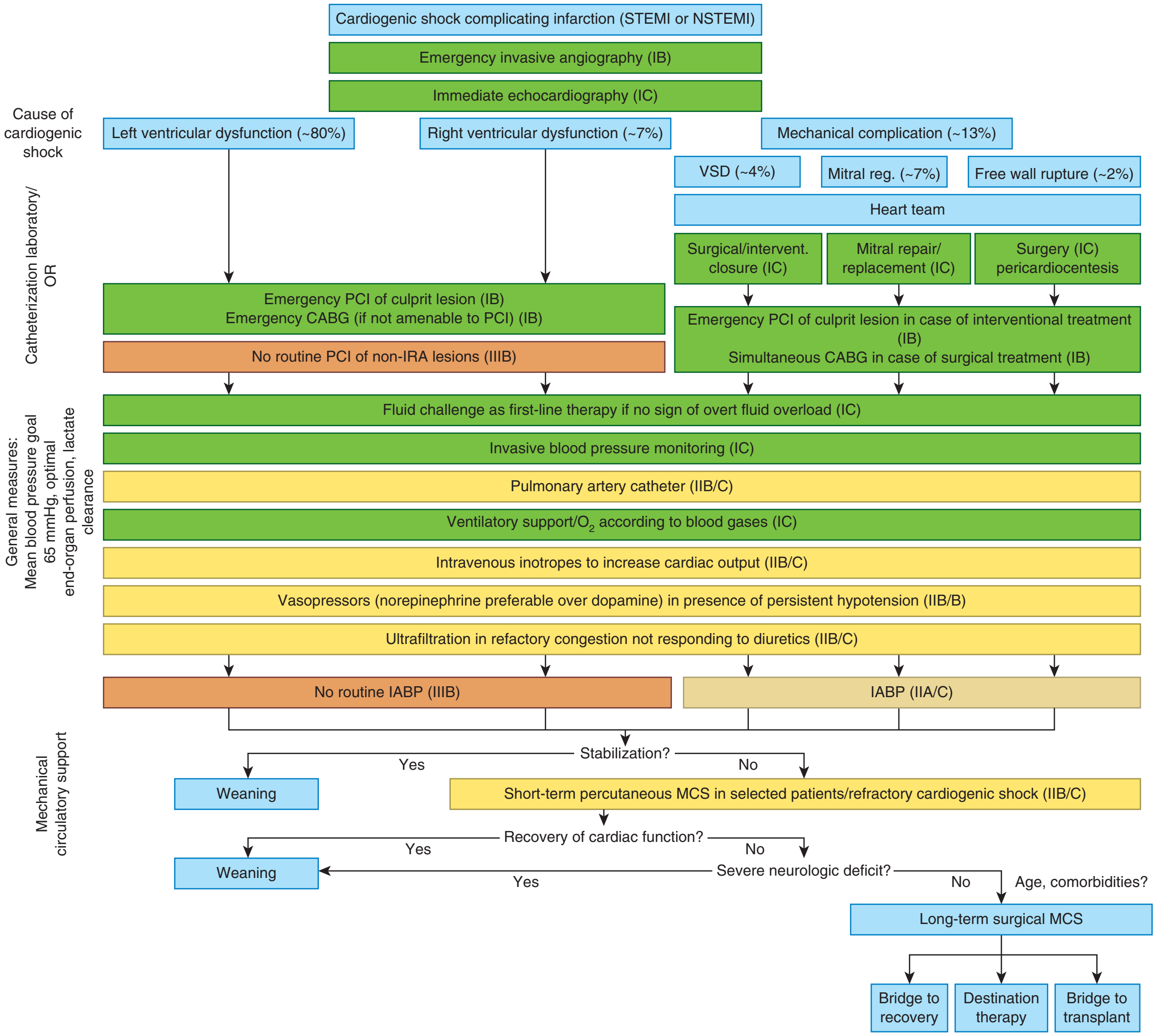
Cardiogenic shock complicating infarction (STEMI or NSTEMI)
Emergency invasive angiography (IB)
Immediate echocardiography (IC)
Cause of cardiogenic shock
Left ventricular dysfunction (~80%)
Right ventricular dysfunction (~7%)
Mechanical complication (~13%)
VSD (~4%)
Mitral reg. (~7%)
Free wall rupture (~2%)
Heart team
Surgical/intervent. closure (IC)
Mitral repair/ replacement (IC)
Surgery (IC) pericardiocentesis
Catheterization laboratory/ OR
Emergency PCI of culprit lesion (IB)
Emergency CABG (if not amenable to PCI) (IB)
No routine PCI of non-IRA lesions (IIIB)
Emergency PCI of culprit lesion in case of interventional treatment (IB)
Simultaneous CABG in case of surgical treatment (IB)
General measures: Mean blood pressure goal 65 mmHg, optimal end-organ perfusion, lactate clearance
Fluid challenge as first-line therapy if no sign of overt fluid overload (IC)
Invasive blood pressure monitoring (IC)
Pulmonary artery catheter (IIB/C)
Ventilatory support/O$_2$ according to blood gases (IC)
Intravenous inotropes to increase cardiac output (IIB/C)
Vasopressors (norepinephrine preferable over dopamine) in presence of persistent hypotension (IIB/B)
Ultrafiltration in refractory congestion not responding to diuretics (IIB/C)
No routine IABP (IIIB)
IABP (IIA/C)
Mechanical circulatory support
Stabilization?
Yes
No
Weaning
Short-term percutaneous MCS in selected patients/refractory cardiogenic shock (IIB/C)
Recovery of cardiac function?
Yes
No
Weaning
Yes
Severe neurologic deficit?
No
Age, comorbidities?
Long-term surgical MCS
Bridge to recovery
Destination therapy
Bridge to transplant

FIGURE 316-3 Emergency management of patients with cardiogenic shock (CS) complicating acute myocardial infarction (AMI). Treatment algorithm for patients with CS. The class of recommendation and level of evidence according to European Society of Cardiology guidelines are provided (see "Further Reading"). CABG, coronary artery bypass grafting; ECG, electrocardiogram; IABP, intraaortic balloon pump; IRA, infarct-related artery; MCS, mechanical circulatory support; NSTEMI, non–ST-segment elevation myocardial infarction; PCI, percutaneous coronary intervention; STEMI, ST-segment elevation myocardial infarction; VSD, ventricular septal defect. *(Reproduced with permission from H Thiele et al: Management of cardiogenic shock complicating myocardial infarction: An update 2019. Eur Heart J 40:2671, 2019.)*

2–4 μg/min and titrated upward based on BP. Norepinephrine was associated with lower lactate levels and less refractory CS compared to epinephrine. Dopamine's hemodynamic effects vary depending on dose, and there is interpatient variability in responses. Low doses stimulate renal dopaminergic receptors, and with increasing doses, there is stimulation of first β-adrenergic receptors and then α-adrenergic receptors. Dopamine should be avoided as first-line therapy for MI with CS based on hemodynamic and proarrhythmogenic effects.

Dobutamine is a synthetic sympathomimetic amine with positive inotropic action and minimal positive chronotropic activity at low doses (2.5 μg/kg per min) but moderate chronotropic activity at higher doses. Its vasodilating activity often precludes its use when a vasoconstrictor effect is required. Levosimendan may also be appealing despite a lack of randomized data but was not beneficial for organ dysfunction in sepsis and also in high-risk patients undergoing cardiovascular surgery. Milrinone—a phosphodiesterase-3 inhibitor and inodilator—was recently shown to have no benefit in comparison with dobutamine.

MECHANICAL CIRCULATORY SUPPORT

The most commonly used mechanical circulatory support (MCS) device has been the intraaortic balloon pump (IABP), which is inserted into the aorta via the femoral artery and provides passive hemodynamic support. However, routine IABP use in conjunction with early revascularization (predominantly with PCI) did not reduce 30-day, 12-month, or 6-year mortality in the IABP-SHOCK II trial. IABP also had no benefit on secondary endpoints (arterial lactate, catecholamine doses, renal function, or intensive care severity of illness unit scores). IABP is no longer recommended for CS with LV failure.

Active MCS devices to support the left, right, or both ventricles can be placed percutaneously or surgically. Temporary percutaneous MCS can be used as bridge to recovery, to surgically implanted durable devices, to heart transplantation, or as a temporizing measure when the neurologic status is uncertain. Percutaneous MCS, including the TandemHeart and Impella devices, and also venoarterial extracorporeal membrane oxygenation (VA-ECMO) have been used in patients not responding to standard treatment (catecholamines, fluids, and IABP) and also as a first-line treatment. Active percutaneous MCS results in better hemodynamic support compared to IABP. However, the appropriate role of MCS, in particular Impella, is uncertain because a positive impact on clinical outcomes or mortality has not yet been demonstrated in trials or meta-analyses. More recent observational data with matched comparisons comprising several ten thousands of patients even showed higher mortality and more complications with active devices such as Impella. Recently the results of the Danish-Germany (DanGer) shock trial

TABLE 316-2 Utility of the Echocardiogram in Cardiogenic Shock or Pulmonary Edema

CLINICAL QUESTION	INFORMATION
Ventricular function	Predominantly left, right, or biventricular involvement
Etiology	**Acute Myocardial Infarction** • Extent of infarction/myocardium in jeopardy • Status of the nonculprit infarct zone • Presence of mechanical complications **Acute/Chronic Valvular Insufficiency/Obstruction/Stenosis (Native/Prosthetic)** • Etiology: endocarditis; degenerative valve disease • Location and hemodynamic consequences **Dynamic Left Ventricular Tract Obstruction** **Takotsubo Syndrome** **Cardiac Tamponade** Circumferential versus localized effusion Route of pericardiocentesis if indicated **Acute Pulmonary Embolism** Right ventricular function Pulmonary artery pressure Presence of clot in transition/patent foramen ovale **Acute Aortic Syndrome** Nature and extent of dissection Degree of aortic insufficiency Presence of pericardial effusion
Hemodynamics	Volume assessment by inferior vena cava diameter and inspiratory collapse Estimated pulmonary artery systolic pressure Estimated left atrial pressure
Therapeutic guidance	Guide vasoactive support Monitor response to therapy Mechanical circulatory support decisions Catheter position and guidance
Pulmonary	Pleural effusion Lung edema Pneumothorax Pulmonary infiltration

were published: in 360 selected patients with anterior ST-elevation myocardial infarction without high risk of hypoxic brain injury comparing a microaxial flow pump with 3.5 L/min versus standard of care, the active MCS was associated with better 180-day outcome. Despite the long recruitment period, the narrow inclusion criteria, and several open questions such as a high increase in mortality from 30 days to 6-months in the control arm, a very short ICU time in the control group, and the highest ever reported renal replacement therapy frequency in the active MCS arm, this randomized trial is an important study supporting the use of MCS in selected patients.

Recent randomized data of VA-ECMO versus control in CS did not show a survival benefit in the ECLS-SHOCK trial. VA-ECMO was accompanied by significantly higher complications, such as moderate/severe bleeding or peripheral ischemic complications. The lack of mortality benefit and higher complication rates with VA-ECMO use were confirmed in an individual patient data meta-analysis.

Surgically implanted devices can support the circulation as bridging therapy for cardiac transplant candidates or as destination therapy **(Chap. 271)**. Assist devices should be used selectively in suitable patients based on decisions by a multidisciplinary team with expertise in the selection, implantation, and management of MCS devices (Fig. 316-3).

Prognosis The expected death rates for patients with MI complicated by CS range widely based on age, severity of hemodynamic abnormalities, severity of clinical hypoperfusion (arterial lactate, renal function), and performance of early revascularization. The recently introduced IABP-SHOCK II score predicts prognosis based on six readily available variables: age >73 years; prior stroke; glucose at admission >10.6 mmol/L (191 mg/dL); creatinine at admission >132.6 μmol/L (1.5 mg/dL); Thrombolysis in Myocardial Infarction (TIMI) flow grade after PCI <3; and arterial blood lactate at admission >5 mmol/L. It also may help guide treatment strategies. The SCAI CS severity definition with stages A to E is also helpful in prognosis estimation.

SHOCK SECONDARY TO RIGHT VENTRICULAR INFARCTION

Persistent CS due to predominant RV failure accounts for only 5% of CS complicating MI. It often results from proximal right coronary artery occlusion. The salient features are relatively high right atrial pressures, RV dilation and dysfunction, and only mildly or moderately depressed LV function. High right-sided pressures may be absent

TABLE 316-3 Hemodynamic Patterns[a]

	RA, mmHg	RVS, mmHg	RVD, mmHg	PAS, mmHg	PAD, mmHg	PCW, mmHg	CI, (L/min)/m^2	SVR, (dyn · s)/cm^5
Normal values	<6	<25	0–12	<25	0–12	<6–12	≥2.5	(800–1600)
MI without pulmonary edema[b]	—	—	—	—	—	~13 (5–18)	~2.7 (2.2–4.3)	—
Pulmonary edema	↔↑	↔↑	↔↑	↑	↑	↑	↔↓	↑
Cardiogenic shock								
LV failure	↔↑	↔↑	↔↑	↔↑	↑	↑	↓	↔↑
RV failure[c]	↑	↓↔↑[d]	↑	↓↔↑[d]	↔↓↑[d]	↓↔↑[d]	↓	↑
Cardiac tamponade	↑	↔↑	↑	↔↑	↔↑	↔↑	↓	↑
Acute mitral regurgitation	↔↑	↑	↔↑	↑	↑	↑	↔↓	↔↑
Ventricular septal rupture	↑	↔↑	↑	↔↑	↔↑	↔↑	↑PBF ↓SBF	↔↑
Hypovolemic shock	↓	↔↓	↔↓	↓	↓	↓	↓	↑
Septic shock	↓	↔↓	↔↓	↓	↓	↓	↑	↓

[a]There is significant patient-to-patient variation. Pressure may be normalized if cardiac output is low. [b]Forrester et al classified non-reperfused MI patients into four hemodynamic subsets. (From JS Forrester et al: N Engl J Med 295:1356, 1976.) PCW pressure and CI in clinically stable subset 1 patients are shown. Values in parentheses represent range. [c]"Isolated" or predominant RV failure. [d]PCW and pulmonary artery pressures may rise in RV failure after volume loading due to RV dilation and right-to-left shift of the interventricular septum, resulting in impaired LV filling. When biventricular failure is present, the patterns are similar to those shown for LV failure.

Abbreviations: CI, cardiac index; LV, left ventricular; MI, myocardial infarction; P/SBF, pulmonary/systemic blood flow; PAS/D, pulmonary artery systolic/diastolic; PCW, pulmonary capillary wedge; RA, right atrium; RV, right ventricular; RVS/D, right ventricular systolic/diastolic; SVR, systemic vascular resistance.

Source: Table prepared with the assistance of Krishnan Ramanathan, MD.

without volume loading. However, CS often has overlap combinations of both RV and LV ischemia, given a shared septum and the effect of ventricular interdependence on RV function. Management of isolated RV CS includes fluid administration to optimize right atrial pressure (10–15 mmHg); avoidance of excess fluids, which shifts the interventricular septum into the LV; catecholamines; early reestablishment of infarct-artery flow; and possibly right-sided MCS.

■ MITRAL REGURGITATION

(See also Chap. 286) Acute severe MR due to papillary muscle dysfunction and/or rupture may complicate MI and result in CS and/or pulmonary edema. This complication most often occurs on the first day, with a second peak several days later. The diagnosis is confirmed by echocardiography (Table 316-2). Afterload reduction with IABP and, if tolerated, vasodilators to reduce pulmonary edema is recommended as a bridge to surgery or interventional treatment. Mitral valve repair or reconstruction is the definitive therapy and should be performed early in the course in suitable candidates. Other options include percutaneous edge-to-edge repair, which has been successful in case series and registries (Fig. 316-3).

■ VENTRICULAR SEPTAL RUPTURE

(See also Chap. 286) VSR complicating MI is a relatively rare event associated with very high mortality if CS is present (>80%). The incidence of infarct-related VSR without reperfusion was 1–2% but has decreased to 0.2% in the era of reperfusion. VSR occurs a median of 24 h after infarction but may occur up to 2 weeks later. Echocardiography demonstrates shunting of blood from the left to the right ventricle and may visualize the opening in the interventricular septum. Current American guidelines recommend immediate surgical VSR closure, irrespective of the patient's hemodynamic status, to avoid further hemodynamic deterioration. European guidelines differ with a more selective approach based on heart team evaluation. IABP support as a bridge to surgery is recommended based on expert opinion. Active MCS may, however, be more appropriate for stabilization of the patient. Given high mortality, suboptimal surgical results, and the ineligibility for surgery of many patients, interventional percutaneous VSR umbrella device closure has been developed. Results of interventional VSR closure suggest a similar outcome as surgery. The heart team should decide how to close the VSR (Fig. 316-3).

■ FREE WALL RUPTURE

Myocardial rupture is a dramatic complication of MI that is most likely to occur during the first week after the onset of symptoms. The clinical presentation typically is a sudden loss of pulse, BP, and consciousness with ongoing sinus rhythm on ECG (pulseless electrical activity) due to cardiac tamponade **(Chap. 281)**. Free wall rupture may also result in CS due to subacute tamponade when the pericardium temporarily seals the rupture sites. Definitive surgical repair is required (Fig. 316-3).

■ ACUTE FULMINANT MYOCARDITIS

(See also Chaps. 266–270) Myocarditis can mimic acute MI with ST abnormalities or bundle branch block on the ECG and marked elevation of cardiac markers. Acute myocarditis causes CS in a small proportion of cases. These patients are typically younger than those with CS due to acute MI and often do not have typical ischemic chest pain. Echocardiography usually shows global LV dysfunction. Initial management is the same as for CS complicating acute MI but does not involve revascularization. Endomyocardial biopsy is recommended to determine the diagnosis and need for immunosuppressives for entities such as giant cell myocarditis. Refractory CS can be managed with MCS.

■ PULMONARY EDEMA

The etiologies and pathophysiology of pulmonary edema are discussed in Chap. 39.

Diagnosis Acute pulmonary edema usually presents with the rapid onset of dyspnea at rest, tachypnea, tachycardia, and severe hypoxemia. Crackles and wheezing due to alveolar flooding, increased airway fluid, and airway compression from peribronchial cuffing may be audible. Release of endogenous catecholamines often causes hypertension.

It is often difficult to distinguish between cardiogenic and noncardiogenic causes of acute pulmonary edema. *Echocardiography* may identify systolic and diastolic ventricular dysfunction and valvular lesions. ECG ST elevation and evolving Q waves are usually diagnostic of acute MI and should prompt immediate institution of MI protocols and coronary artery revascularization therapy **(Chap. 286)**. Brain natriuretic peptide levels, when substantially elevated, support heart failure as the etiology of acute dyspnea with pulmonary edema **(Chap. 264)**.

The use of a PAC permits measurement of pulmonary capillary wedge pressures (PCWP) and helps differentiate high-pressure (cardiogenic) from normal-pressure (noncardiogenic) causes of pulmonary edema. Pulmonary artery catheterization is indicated when the etiology of the pulmonary edema is uncertain, when edema is refractory to therapy, or when it is accompanied by refractory hypotension. Data derived from use of a PAC often alter the treatment plan, but no impact on mortality rates has been demonstrated.

TREATMENT

Pulmonary Edema

The treatment of pulmonary edema depends on the specific etiology. As an acute, life-threatening condition, a number of measures must be applied immediately to support the circulation, gas exchange, and lung mechanics. Simultaneously, conditions that frequently complicate pulmonary edema, such as infection, acidemia, anemia, and acute kidney dysfunction, must be corrected.

SUPPORT OF OXYGENATION AND VENTILATION

Patients with acute cardiogenic pulmonary edema generally have an identifiable cause of acute LV failure—such as arrhythmia, ischemia/infarction, or myocardial decompensation **(Chap. 264)**—that may be rapidly treated, with improvement in gas exchange. In contrast, noncardiogenic edema usually resolves much less quickly, and most patients require mechanical ventilation.

Oxygen Therapy Support of oxygenation is essential to ensure adequate O_2 delivery to peripheral tissues, including the heart. Generally, the goal is O_2 saturation of 92% or more, but very high saturation (>98%) may be detrimental. For non-CS acute hypoxemic respiratory failure patients with normal Pa_{CO_2}, O_2 administration by high-flow nasal cannula for acute hypoxemic respiratory failure has better outcomes than use of bilevel positive airway pressure (BiPAP).

Positive-Pressure Ventilation Pulmonary edema increases the work of breathing and the O_2 requirements of this work, imposing a significant physiologic stress on the heart. When oxygenation or ventilation is not adequate despite supplemental O_2, positive-pressure ventilation by face or nasal mask or by endotracheal intubation should be initiated. Noninvasive ventilation (NIV) **(Chap. 313)** can rest the respiratory muscles, improve oxygenation and cardiac function, and reduce the need for intubation. While NIV is believed effective for cardiogenic pulmonary edema, Cochrane analyses have not yet substantiated this benefit. In refractory cases, mechanical ventilation can relieve the work of breathing more completely than can NIV. Helmet ventilation is a new technique for ventilation with positive pressure without intubation. Mechanical ventilation with positive end-expiratory pressure can have multiple beneficial effects on pulmonary edema, as it: (1) decreases both preload and afterload, thereby improving cardiac function; (2) redistributes lung water from the intraalveolar to the extraalveolar space, where the fluid interferes less with gas exchange; and (3) increases lung volume to avoid atelectasis.

Renal Replacement Therapy For pulmonary edema patients with refractory volume overload, metabolic acidosis (pH <7.15–7.25), hypoxemia, and/or persistent hyperkalemia, renal replacement therapy should be considered. For patients who are hypotensive or require ionotropic support, continuous renal replacement therapy usually is better tolerated than intermittent hemodialysis.

REDUCTION OF PRELOAD

In most forms of pulmonary edema, the quantity of extravascular lung water is determined by a combination of the PCWP, the pulmonary vascular permeability, and the intravascular volume status.

Diuretics The loop diuretics furosemide, bumetanide, and torsemide are effective in most forms of pulmonary edema, even in the presence of hypoalbuminemia, hyponatremia, or hypochloremia. Furosemide is also a venodilator that rapidly reduces preload before any diuresis occurs and is the diuretic of choice. The initial dose of furosemide should be ≤0.5 mg/kg, but a higher dose (1 mg/kg) is required in patients with renal insufficiency, chronic diuretic use, or hypervolemia or after failure of a lower dose. Combinations of diuretics and/or continuous infusion are helpful to achieve the desired degree of diuresis in selected patients.

Nitrates Nitroglycerin and isosorbide dinitrate act predominantly as venodilators but have coronary vasodilating properties as well. Their onset is rapid, and they are effectively administered by a variety of routes. Sublingual nitroglycerin (0.4 mg × 3 every 5 min) is first-line therapy for acute cardiogenic pulmonary edema. If pulmonary edema persists in the absence of hypotension, sublingual may be followed by IV nitroglycerin, commencing at 5–10 μg/min. IV nitroprusside (0.1–5 μg/kg per min) is a potent venous and arterial vasodilator. It is useful for patients with pulmonary edema and hypertension but is not recommended in states of reduced coronary artery perfusion. It requires close monitoring and titration using an arterial catheter for continuous BP measurement.

Morphine Given in 2- to 4-mg IV boluses, morphine is a transient venodilator that reduces preload while relieving dyspnea and anxiety. These effects can diminish stress, catecholamine levels, tachycardia, and ventricular afterload in patients with pulmonary edema and systemic hypertension. However, some registry trials showed increased mortality with use of morphine.

Angiotensin-Converting Enzyme (ACE) Inhibitors ACE inhibitors reduce both afterload and preload and are recommended for hypertensive patients. A low dose of a short-acting agent may be initiated and followed by increasing oral doses. In acute MI with heart failure, ACE inhibitors reduce short- and long-term mortality rates. The optimal starting point of ACE inhibitors has not been tested so far.

Other Preload-Reducing Agents IV recombinant brain natriuretic peptide (nesiritide) is a potent arterial and venous vasodilator with diuretic properties and is effective in the treatment of cardiogenic pulmonary edema. It should be reserved for refractory patients and is not recommended in the setting of ischemia or MI. Endothelin antagonists are being studied as they inhibit vasoconstriction and can improve cardiac output and decrease PCWP.

Physical Methods In nonhypotensive patients, venous return can be reduced by use of the sitting position with the legs dangling along the side of the bed.

Inotropic and Inodilator Drugs The sympathomimetic amines dopamine and dobutamine (see above) are potent inotropic agents. The bipyridine phosphodiesterase-3 inhibitors (inodilators), such as milrinone (50 μg/kg followed by 0.25–0.75 μg/kg per min), stimulate myocardial contractility while promoting peripheral and pulmonary vasodilation. Inodilators may be helpful in selected patients with cardiogenic pulmonary edema and severe LV dysfunction, but there is little published clinical data. Angiotensin II is a vasoconstrictor and possible positive inotrope that can raise BP in many types of shock. It is expensive and has not been shown to have additive or superior benefit to other vasopressors in CS.

Digitalis Glycosides Once a mainstay of treatment because of their positive inotropic action **(Chap. 264)**, digitalis glycosides are rarely used at present. However, they may be useful for control of ventricular rate in patients with rapid ventricular response to atrial fibrillation or flutter and LV dysfunction with pulmonary edema, because they do not have the negative inotropic effects of other drugs that inhibit atrioventricular nodal conduction.

Intraaortic Balloon Counterpulsation IABP **(Chap. 271)** may be helpful in rare instances of acute MR but is not typically used for pulmonary edema with CS.

***Treatment of Tachyarrhythmias and Atrioventricular Resynchronization* (See also Chap. 259)** Sinus tachycardia or atrial fibrillation can result from elevated left atrial pressure and sympathetic stimulation. Tachycardia itself can limit LV filling time and raise left atrial pressure further. Although relief of pulmonary congestion will slow the sinus rate or ventricular response in atrial fibrillation, a primary tachyarrhythmia may require cardioversion. In patients with reduced LV function and without atrial contraction or with lack of synchronized atrioventricular contraction, placement of an atrioventricular sequential pacemaker should be considered **(Chap. 251)**.

Reduction in Pulmonary Vascular Permeability At present, no clinical therapies have been demonstrated as clinically effective to reduce the "leakiness" of the pulmonary capillaries.

Stimulation of Alveolar Fluid Clearance A variety of drugs and cellular therapies can stimulate alveolar epithelial ion transport and upregulate the clearance of alveolar solute and water, but this strategy has not been proven beneficial in clinical trials thus far.

SPECIAL CONSIDERATIONS

Risk of Iatrogenic Cardiogenic Shock In the treatment of pulmonary edema, vasodilators lower BP, and their use, particularly in combination, may lead to hypotension, coronary artery hypoperfusion, and shock (Fig. 316-1). In general, patients with a *hypertensive* response to pulmonary edema tolerate and benefit from these medications. In normotensive patients, low doses of single agents should be instituted sequentially, as needed, and with close monitoring.

***Acute Coronary Syndromes* (See also Chap. 286)** Acute STEMI complicated by pulmonary edema is associated with in-hospital mortality rates of 20–40%. After immediate stabilization, coronary artery blood flow must be reestablished rapidly. Early primary PCI is the method of choice; alternatively, a fibrinolytic agent should be administered. Early coronary angiography and revascularization by PCI or CABG also are indicated for patients with non–ST-segment elevation acute coronary syndrome.

Takotsubo Syndrome Takotsubo syndrome is an acute reversible heart failure syndrome characterized by acute onset of left-sided heart failure with reversible ST-segment elevation and some increase in troponin levels, usually triggered by a major physical or emotional, stressful event. At end systole, there often is the appearance of LV apical "ballooning." Most patients recover and return to normal ventricular function. However, prognosis is similar or even worse in comparison to patients with acute MI.

Extracorporeal Membrane Oxygenation (ECMO) For patients with acute, severe, noncardiogenic edema with a potential rapidly reversible cause, ECMO may be considered in highly selected patients as a temporizing supportive measure to achieve adequate gas exchange, with current survival to discharge rates of 50–60%. Usually, venovenous ECMO is used in this setting. ECMO can function as a bridge to transplantation or other interventions.

Unusual Types of Edema Specific etiologies of pulmonary edema may require particular therapy. Reexpansion pulmonary edema can develop after removal of longstanding pleural space air or fluid. These patients may develop hypotension or oliguria with pulmonary edema resulting from rapid fluid shifts into the lung. Diuretics and preload reduction are contraindicated, and intravascular volume repletion often is needed while supporting oxygenation and gas exchange.

High-altitude pulmonary edema often can be prevented by use of dexamethasone, calcium channel–blocking drugs, or long-acting inhaled β_2-adrenergic agonists. Treatment includes descent from altitude, bed rest, oxygen, and, if feasible, inhaled NO; nifedipine may also be effective.

For pulmonary edema resulting from upper airway obstruction, recognition of the obstructing cause is key because treatment then is to relieve or bypass the obstruction.

FURTHER READING

Alviar CL et al: For the ACC Critical Care Cardiology Working Group. Positive pressure ventilation in the cardiac intensive care unit. J Am Coll Cardiol 72:1532, 2018.

Amin AP et al: The evolving landscape of Impella use in the United States among patients undergoing percutaneous coronary intervention with mechanical circulatory support. Circulation 141:273, 2020.

Baran DA et al: SCAI clinical expert consensus statement on the classification of cardiogenic shock. Cathet Cardiovasc Interv 94:29, 2019.

Damluji AA et al: Mechanical complications of acute myocardial infarction. A scientific statement from the American Heart Association. Circulation 144:e16, 2021.

Dhruva SS et al: Association of use of intravascular microaxial left ventricular assist device vs intra-aortic balloon pump on in-hospital mortality and major bleeding among patients with acute myocardial infarction complicated by cardiogenic shock. JAMA 323:734, 2020.

Heidenreich PA et al: 2022 AHA/ACC/HFS guideline for the management of heart failure. Circulation 145:895, 2022.

Ingbar DH: Cardiogenic pulmonary edema: Mechanisms and treatment—An intensivists view. Curr Opin Crit Care 25:371, 2019.

Møller JE et al: Microaxial flow pump or standard care in infarct-related cardiogenic shock. N Engl J Med 390:1382, 2024.

Naidu SS et al: SCAI SHOCK stage classification expert consensus update: A review and incorporation of validation studies. J Am Coll Cardiol 79:933, 2022.

Thiele H et al: Intraaortic balloon support for myocardial infarction with cardiogenic shock. N Engl J Med 367:1287, 2012.

Thiele H et al: PCI strategies in patients with acute myocardial infarction and cardiogenic shock. N Engl J Med 377:2419, 2017.

Thiele H et al: Percutaneous short-term active mechanical support devices in cardiogenic shock: A systematic review and collaborative meta-analysis of randomized trials. Eur Heart J 38:3523, 2017.

Thiele H et al: One-year outcomes after PCI strategies in cardiogenic shock. N Engl J Med 379:1699, 2018.

Thiele H et al: Management of cardiogenic shock complicating myocardial infarction: An update 2019. Eur Heart J 40:2671, 2019.

Thiele H et al: Extracorporeal life support in infarct-related cardiogenic shock. N Engl J Med 389:1286, 2023.

van Diepen S et al: Contemporary management of cardiogenic shock: A scientific statement. Circulation 136:e232, 2017.

Zeymer U et al: Veno-arterial extracorporeal membrane oxygenation in patients with infarct-related cardiogenic shock: An individual patient data meta-analysis of randomised trials. Lancet 402:1338, 2023.

317

Cardiovascular Collapse, Cardiac Arrest, and Sudden Cardiac Death

Christine M. Albert, William H. Sauer

OVERVIEW AND DEFINITIONS (SEE TABLE 317-1)

Cardiovascular collapse is severe hypotension from acute cardiac dysfunction or loss of peripheral vasculature resistance resulting in cerebral hypoperfusion and loss of consciousness. This condition can be the result of a cardiac arrhythmia, severe myocardial or valvular dysfunction, loss of vascular tone, and/or acute disruption of venous return. When effective circulation is restored spontaneously, patients present with syncope **(see Chap. 23)**. In the absence of spontaneous resolution, cardiac arrest occurs, ultimately resulting in death if resuscitation attempts are unsuccessful or not initiated. Underlying etiologies for cardiovascular collapse include benign conditions, such as neurocardiogenic syncope, but also life-threatening conditions including ventricular tachyarrhythmias, severe bradycardia, severely depressed myocardial contractility, as with massive acute myocardial infarction (MI) or pulmonary embolus, and other catastrophic events interfering with cardiac function such as myocardial rupture with cardiac tamponade or papillary muscle rupture with torrential mitral regurgitation.

Sudden cardiac arrest (SCA) refers to an *abrupt* loss of cardiac function resulting in complete cardiovascular collapse due either to an acute life-threatening cardiac arrhythmia or abrupt loss of myocardial pump function that requires emergency medical intervention for restoration of effective circulation. Most SCAs occur outside the hospital, and fewer than 10% of these victims survive to be discharged from the hospital despite undergoing attempted resuscitation by emergency medical services (EMS). For those who die prior to hospital admission, a cardiovascular cause for the arrest is often presumed based upon the absence of evidence for a traumatic or other noncardiac cause at the time of the arrest. If the patient does not survive a SCA, the death is classified as a sudden cardiac death (SCD). Deaths that occur during hospitalization or within 30 days after resuscitated cardiac arrest are usually counted as SCDs in epidemiologic studies.

SCD also includes a broader category of unexplained rapid deaths thought to be due to cardiac causes where resuscitation was not attempted. In epidemiologic studies, SCD is usually defined as an unexpected death without obvious extracardiac cause that occurs in association with a witnessed rapid collapse or within 1 hour of symptom onset. This definition is based on the presumption that rapid deaths are often due to an arrhythmia, an assumption that cannot always be validated. Approximately half of all SCDs are not witnessed. In the United States, few deaths undergo autopsies, and noncardiac conditions that evolve rapidly such as acute cerebral hemorrhage, aortic rupture, and pulmonary embolism cannot be excluded without an autopsy. Therefore, definitive information necessary to establish the cause of death is usually not available. In unwitnessed cases, the definition is often further expanded to include unexpected deaths where the subject was documented to be well when last observed within the preceding 24 hours. This expanded definition further decreases the certainty that the death was due to an arrhythmia or other cardiac causes, and recent data suggest that noncardiac causes may comprise a larger than expected percentage of these unwitnessed sudden deaths. Most countries, including the United States, do not have national surveillance systems or reporting requirements for SCD; thus, the true incidence and frequency of SCD and its different mechanisms can only be estimated.

EPIDEMIOLOGY

DEMOGRAPHICS

SCA and SCD are major public health problems that account for 15% of all deaths and comprise 50% of all cardiac deaths. In the United States alone, there are an estimated 350,000 EMS-attended out-of-hospital cardiac arrests and 210,000 SCDs in the adult population annually. The estimated societal burden of premature death due to SCD is 2 million years of potential life lost for men and 1.3 million years of potential life lost for women, which is greater than most other leading causes of death. Although cardiac pathology, particularly coronary heart disease (CHD), underlies the majority of SCDs, up to two-thirds of all SCDs occur as the first clinical expression of previously undiagnosed heart disease. SCD rates have declined but not as steeply as rates for CHD in general. Age, gender, race, and geographic region all influence the incidence of SCD. Rates of out-of-hospital cardiac arrest are lower in Asia (52.5 per 100,000 person-years) than Europe (86.4 per 100,000 person-years), North America (98.1 per 100,000 person-years), and Australia (111.9 per 100,000 person-years), and also vary within geographic regions of the United States. SCD is rare in individuals younger than 35 years of age (1–3 per 100,000 per year) and increases markedly with age as the incidence of coronary artery disease (CAD), heart failure (HF),

TABLE 317-1 Distinction Between Cardiovascular Collapse, Cardiac Arrest, and Death

TERM	DEFINITION	QUALIFIERS	MECHANISMS
Cardiovascular collapse	Sudden loss of effective circulation due to cardiac and/or peripheral vascular factors that may reverse spontaneously (e.g., neurocardiogenic syncope transient hypotension) or require interventions (e.g., hypovolemia, tamponade, ventricular fibrillation).	Broad term that includes cardiac arrest as well as transient events that characteristically revert spontaneously presenting as syncope.	Same as cardiac arrest, plus neurocardiogenic syncope or other causes of transient loss of blood flow.
Cardiac arrest	Abrupt cessation of cardiac function resulting in loss of effective circulation that may be reversible by prompt emergency medical intervention.	Rare spontaneous reversions; likelihood of successful intervention relates to mechanism of arrest, clinical setting, availability of emergency medical services, and prompt return of circulation.	Ventricular fibrillation, ventricular tachycardia, asystole, bradycardia, pulseless electrical activity, noncardiac mechanical factors (e.g., pulmonary embolism).
Sudden cardiac death	Sudden unexpected death attributed to cardiac arrest, which occurs within 1 hour of symptom onset.	In unwitnessed cases, the definition is often expanded to include unexpected deaths where the subject was documented to be well within the preceding 24 hours.	Same as cardiac arrest.

and other predisposing conditions also increases. Although absolute SCD rates increase with age, the proportion of deaths that are due to SCD decreases as other causes of death increase.

Women have a lower incidence of SCD and SCA than men, and women are more likely to present with pulseless electrical activity (PEA) and to have their SCD occur at home as compared with men. Possibly related to these factors, the SCD rate has not declined as much for younger women compared to men in recent years. Black as opposed to white Americans have higher rates of SCD, are more likely to have unwitnessed arrests and to be found with PEA, and have lower rates of survival. Socioeconomic disparities, with resuscitation being less likely in low-income neighborhoods, are contributing factors but do not appear to account for the entirety of the elevated SCD rate in blacks. Alternatively, individuals of Hispanic ethnicity appear to have lower rates of SCD, despite having a higher prevalence of cardiac risk factors. It appears that the incidence of SCD may be relatively low among Asian populations as well, both within the United States and globally. These gender and racial differences in SCD/SCA incidence and survival are poorly understood and warrant further research.

■ RISK FACTORS (SEE FIG. 317-1)

The presence of overt structural heart disease and/or certain types of inherited arrhythmia syndromes markedly elevates SCD risk (see Chaps. 261 and 262). Preexisting CHD and HF are the most prevalent predisposing cardiac conditions and are associated with a four- to tenfold increase in SCD risk. Correspondingly, SCD shares many of the same risk factors with CHD and heart failure (HF), including hypertension, diabetes, hypercholesterolemia, obesity, and smoking. Diabetes is a particularly strong risk factor for SCD even in patients with established CHD. Hypertension and resultant left ventricular hypertrophy (LVH) appear to be particularly important markers of SCD risk in blacks, in whom the prevalence of these conditions is greater. Smoking markedly elevates risk, and smoking cessation lowers risk particularly among individuals who have not yet developed overt CHD. Serum cholesterol appears to be more strongly related to SCD at younger ages, and the benefits of cholesterol lowering on SCD incidence have not been firmly established. There also appears to be a genetic component to SCD risk that is distinct from that associated with other manifestations of atherosclerosis. A history of SCD in a first-degree relative is associated with an increased risk for SCD, and with the occurrence of ventricular fibrillation (VF) during acute MI, but is not associated with an increased risk for acute MI. These data suggest that genetic factors may predispose to fatal ventricular arrhythmia in the setting of ischemia, rather than to CHD in general.

Obstructive sleep apnea and seizure disorders are also associated with increased SCD risk; the underlying mechanism is not clear but may be due to hypoxia-induced cardiac arrest. Atrial fibrillation also appears to be associated with an increased risk of SCD, which is partly, but not entirely, accounted for by its association with underlying heart disease. Patients with chronic kidney disease are also at higher SCD risk with annualized SCD rates approaching 5.5% in patients undergoing dialysis. Electrolyte shifts and LVH, which are common in this population, have been suggested to play a role. There are also potential dietary influences on SCD risk. Individuals with higher intakes of polyunsaturated fatty acids, particularly n-3 fatty acids, and other components of a Mediterranean-style diet have lower SCD risks in observational studies, possibly due to antiarrhythmic effects of dietary components. Low levels of alcohol intake may be beneficial, but heavy intake (>3 drinks/day) appears to elevate risk.

■ PRECIPITATING FACTORS

SCD/SCA occurs with higher frequency at certain times, locations, and in association with certain activities and exposures. Although not consistently observed across all studies, there do appear to be circadian variations in the incidence of SCD and cardiac arrest, with peaks in incidence in the morning hours and again in the later afternoon. There is also seasonal variability in SCD rates, which may be related to temperature and light exposure. Rates are highest during winter in the northern hemisphere and summer in the southern hemisphere. SCD rates also acutely peak during disasters such as earthquakes and terrorist attacks. SCA arrests are more likely to occur in certain locations as well, with notable clustering around train stations, airports, and other public places where there is significant population transit. SCD rates tend to be higher in urban areas, and individuals who live near major roadways are at elevated SCD risk. There is also a well-recognized acute elevation in SCD risk that occurs during or shortly after bouts of vigorous exertion, and men appear to be more susceptible. Habitual exercise and training lower this acute risk but do not eliminate it entirely. Exertion-associated SCDs are particularly tragic and highly publicized when they occur in highly trained athletes; however, the majority of such deaths actually occur in the general population. The common thread among these precipitating factors is likely heightened autonomic tone, which can promote ischemia and has direct proarrhythmia and electrophysiologic actions that lower the threshold for sustained VF.

CAUSES OF SUDDEN CARDIAC DEATH

■ UNDERLYING HEART DISEASE (FIG. 317-1)

Our understanding regarding the diseases that contribute to SCD is derived primarily from autopsy series and cardiac evaluations in cardiac arrest survivors, which are highly variable in level of detail. Despite the limitations of these data, it is generally accepted that sudden death due to cardiac causes is most commonly due to CAD, although the proportion with CAD varies markedly by age, race, and sex. It is estimated that ~70% of SCDs in white men are due to CAD, as compared with only 40–50% in women and blacks. The proportion of SCDs with underlying CAD may be even lower in Asian ethnicities. Recent data suggest that the proportion of SCDs with CAD on autopsy may be declining in some parts of Europe and the United States, and,

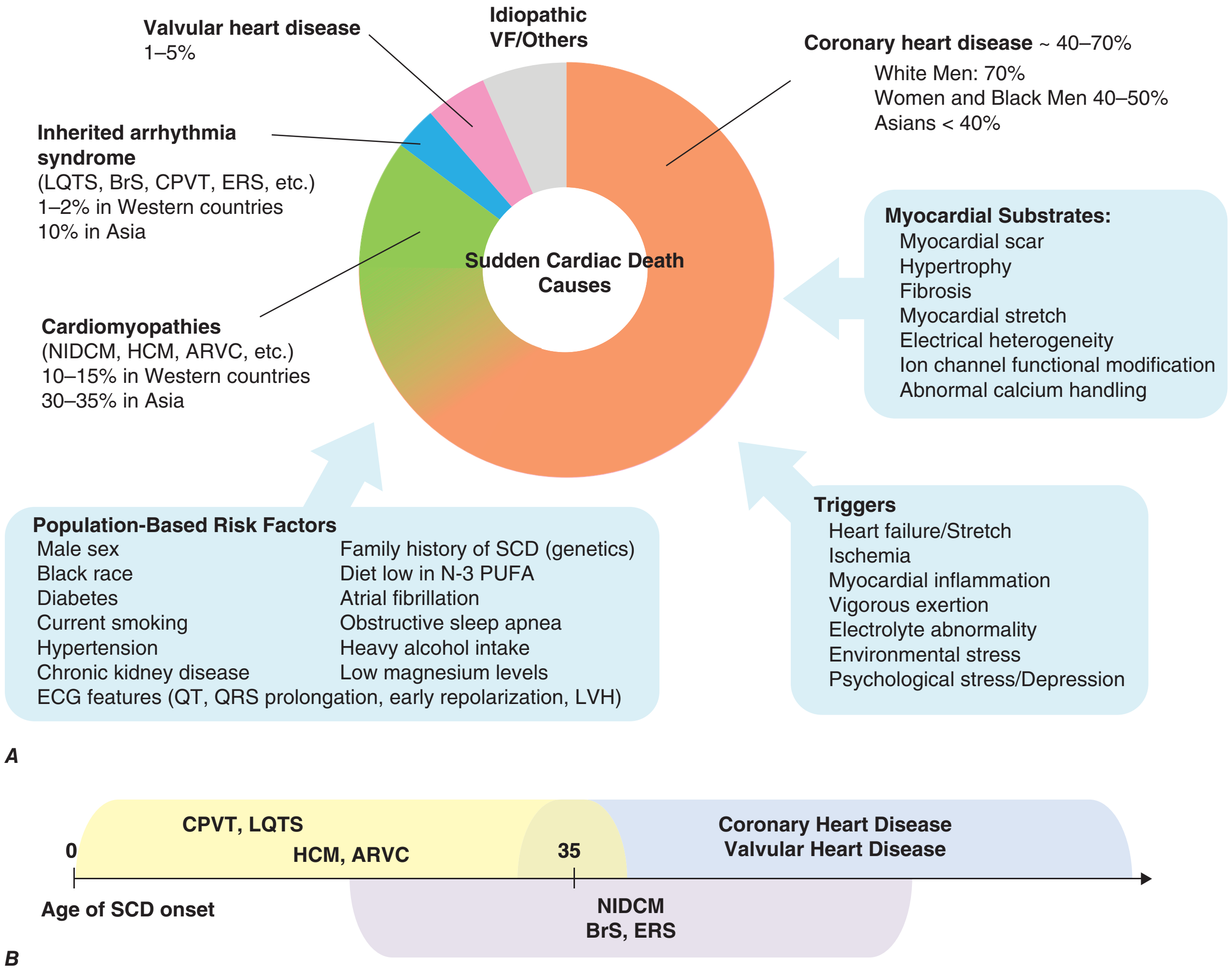
Valvular heart disease
1–5%
Idiopathic
VF/Others
Coronary heart disease ~ 40–70%
White Men: 70%
Women and Black Men 40–50%
Asians < 40%
Inherited arrhythmia syndrome
(LQTS, BrS, CPVT, ERS, etc.)
1–2% in Western countries
10% in Asia
Sudden Cardiac Death
Causes
Myocardial Substrates:
Myocardial scar
Hypertrophy
Fibrosis
Myocardial stretch
Electrical heterogeneity
Ion channel functional modification
Abnormal calcium handling
Cardiomyopathies
(NIDCM, HCM, ARVC, etc.)
10–15% in Western countries
30–35% in Asia
Population-Based Risk Factors
Male sex
Black race
Diabetes
Current smoking
Hypertension
Chronic kidney disease
Family history of SCD (genetics)
Diet low in N-3 PUFA
Atrial fibrillation
Obstructive sleep apnea
Heavy alcohol intake
Low magnesium levels
ECG features (QT, QRS prolongation, early repolarization, LVH)
Triggers
Heart failure/Stretch
Ischemia
Myocardial inflammation
Vigorous exertion
Electrolyte abnormality
Environmental stress
Psychological stress/Depression
A
CPVT, LQTS
HCM, ARVC
0
35
Coronary Heart Disease
Valvular Heart Disease
Age of SCD onset
NIDCM
BrS, ERS
B

FIGURE 317-1 ***A.*** Proportionate causes, substrates, risk factors, and triggers of sudden cardiac death (SCD). ***B.*** Variation of causes by age of onset. ARVC, arrhythmogenic right ventricular cardiomyopathy; BrS, Brugada syndrome; CPVT, catecholaminergic ventricular tachycardia; ECG, electrocardiogram; ERS, early repolarization syndrome; HCM, hypertrophic cardiomyopathy; LQTS, long QT syndrome; LVH, left ventricular hypertrophy; NICDM, nonischemic cardiomyopathy; PUFA, polyunsaturated fatty acid; SCD, sudden cardiac death; VF, ventricular fibrillation. *(Reproduced with permission from M Hayashi et al: The spectrum of epidemiology underlying sudden cardiac death. Circ Res 116:1887, 2015.)*

at the same time, increasing in parts of Japan and other parts of Asia. Beyond CAD, nonischemic cardiomyopathies (hypertrophic, dilated, and infiltrative) are the second most frequent cause of SCD in the United States and European countries. Other less common causes include valvular heart disease, myocarditis, myocardial hypertrophy (often from hypertension), and rare primary electrical heart diseases such as long QT and Brugada syndromes. On average, 5–10% of SCA victims do not have a significant cardiac abnormality at the time of autopsy or after extensive premortem cardiac evaluation, and this also varies by gender and race. Before 35 years of age, atherosclerotic CAD accounts for a much smaller proportion of deaths, with hypertrophic cardiomyopathy (HCM), coronary artery anomalies, myocarditis, arrhythmogenic right ventricular cardiomyopathy, and primary ion channelopathies accounting for a significant number of these deaths.

CARDIAC RHYTHMS AND SUDDEN DEATH

The initial rhythm found when EMS arrive at the scene of an out-of-hospital cardiac arrest is an important indication of the potential cause of the arrest and of the prognosis. In the early days of EMS systems, over half of victims were found in VF, giving rise to the hypothesis that ischemic VF or ventricular tachycardia (VT) degenerating to VF was the most common event. The proportion of cardiac arrests found in VF has decreased markedly since the 1970s, to only 20–25% in more recent studies, and PEA and asystole are now the most common scenarios. However, the vast majority of cardiac arrests are not monitored at the time of collapse, and since arrhythmias are inherently unstable once hemodynamic collapse occurs, the rhythm at the time of EMS arrival may not reflect the rhythm that initially precipitated the SCA because VF and primary bradycardias can degenerate into asystole. Nonetheless, VF as an initial rhythm still predominates in public locations or in other situations when there is a short time frame between witnessed arrest and arrival of EMS, suggesting that VF remains a common initial precipitating rhythm. However, there are also data to support an absolute decrease in VF incidence. Proposed explanations include decreases in underlying CHD incidence, increased use of beta blockers in CHD, and implantable cardioverter defibrillators (ICD) in high-risk patients. There also appears to be an increase in PEA incidence over the past several years, suggesting that the proportion of SCD due to abrupt hemodynamic collapse in the absence of preceding fatal arrhythmia may be increasing. Proposed explanations for these proportional changes in PEA versus VF include the aging of the population and the increased prevalence of end-stage cardiovascular disease and other severe comorbidities. These older, sicker patients may be more likely to have arrests in the home and to have acute precipitants leading to PEA (i.e., respiratory, metabolic, vascular) and/or be less likely to sustain VF up to the point of EMS arrival.

DISEASE-SPECIFIC MECHANISMS

CAD can cause SCD through several mechanisms **(Table 317-2)**. The most common cause is acute MI or transient myocardial ischemia that leads to polymorphic VT and VF **(see Chap. 262)**. Other primary mechanisms include severe bradyarrhythmias such as heart block with a slow escape rhythm, or PEA due to a massive MI or associated myocardial rupture. Areas of ventricular scar from prior infarcts increase

TABLE 317-2 Causes of Cardiovascular Collapse and Sudden Cardiac Arrest

CAUSE	PATHOPHYSIOLOGIC SUBSTRATE	RHYTHM PRESENTATION
Cardiac Causes		
Coronary artery disease Atherosclerotic, coronary spasm, congenital anomalies	Acute myocardial ischemia/infarction, ventricular rupture, tamponade Ventricular scar from healed infarction	Polymorphic VT/VF Bradyarrhythmia Pulseless electrical activity VT VF
Cardiomyopathies Dilated, hypertrophic, ARVC, infiltrative disease, valvular disease with LV failure	Ventricular scar Ventricular hypertrophy Pump failure	VT Polymorphic VT/VF Pulseless electrical activity Bradyarrhythmia
Congenital heart disease (Tetralogy of Fallot, VSD, others)	Ventricular scar from surgical repair Hypertrophy	VT Bradyarrhythmias Polymorphic VT/VF
Aortic stenosis	Obstruction to aortic outflow Ventricular hypertrophy	Bradyarrhythmia Pulseless electrical activity Bradyarrhythmia Polymorphic VT/VF
Mitral valve prolapse/mitral regurgitation	Pump failure Ventricular scar	Polymorphic VT/VF
Arrhythmia syndromes without structural heart disease: Genetic: Long QT Brugada CPVT Idiopathic VF, early repolarization Drug toxicities (acquired long QT, others) Electrolyte abnormalities (severe hypokalemia)	Abnormal cellular electrophysiology	Polymorphic VT/VF
Wolff-Parkinson-White syndrome	Accessory atrioventricular connection	Pre-excited AF/VF
Commotio cordis	Blunt precordial impact	Polymorphic VT/VF
Noncardiac Causes of Cardiovascular Collapse		
Pulmonary embolism		PEA
Stroke		PEA, bradyarrhythmia
Aortic dissection		PEA, VF
Exsanguination/hypovolemia		PEA
Tension pneumothorax		PEA
Sepsis		PEA
Neurogenic		PEA, bradyarrhythmia
Drug overdose		PEA, bradyarrhythmia

Abbreviations: AF, atrial fibrillation; ARVC, arrhythmogenic right ventricular cardiomyopathy; CPVT, catecholaminergic polymorphic ventricular tachycardia; LV, left ventricle; PEA, pulseless electrical activity; VF, ventricular fibrillation; VSD, ventricular septal defect; VT, ventricular tachycardia.

the predisposition to reentrant VT, which often degenerates to VF. Once patients have suffered an MI, their risk of SCD elevates up to 10-fold, with the highest absolute rates in the first 30 days after MI. The mechanisms underlying SCD vary at different time points after MI, with nonarrhythmic causes such as myocardial rupture and/or extensive reinfarction predominating early, within the first 1–2 months, and ischemic polymorphic VT and/or scar-related ventricular arrhythmias prevailing later. VT and sudden death can, and often do, occur years after an initial MI.

Cardiomyopathies and Other Forms of Structural Heart Disease Scar-mediated reentrant VT can also occur in a host of nonischemic cardiomyopathies in which replacement fibrosis and/or inflammatory ventricular infiltrates occur **(Chap. 261)**. In congenital heart disease, surgical scars created during corrective surgery, such as those performed to correct ventricular septal defects in tetralogy of Fallot, can also serve as the substrate for ventricular reentry. Other common predisposing processes such as LVH, ventricular stretch due to fluid overload, and cardiomyocyte dysfunction can result in electrical heterogeneity and other electrophysiologic changes that predispose to ventricular arrhythmias, including ion channel alterations that prolong action potential duration, impair cellular calcium handling, and diminish cellular coupling. These processes occur in a wide variety of diseases associated with depressed ventricular function and/or hypertrophy, including CAD, valvular heart disease, myocarditis, and nonischemic cardiomyopathies.

Absence of Structural Heart Disease In the absence of structural heart disease, VF can be due to an inherited ion channel abnormality, as in long QT and Brugada syndromes **(Chap. 262)**, rapid atrial fibrillation associated with Wolff-Parkinson-White syndrome **(Chap. 256)**, or drug toxicities, such as polymorphic VT due to drugs that prolong the QT interval **(Chap. 263)**. Blunt, nonpenetrating precordial impact over the (left) chest wall can lead to commotio cordis and is a rare cause of SCD in otherwise healthy individuals. PEA can result from pulmonary emboli, exsanguination, or the terminal phase of respiratory arrest.

MANAGEMENT OF CARDIAC ARREST

As the ability to predict SCA in the population is very limited, community approaches to reduce death focus on the rapid identification of victims and implementation of resuscitation measures by those who first encounter the victim, most likely the lay public, who ideally summon EMS and initiate basic life-support measures with chest compressions. The approach is codified in the "out-of-hospital chain of survival," which includes: (1) initial evaluation and recognition of the SCA and activation of the emergency response system; (2) rapid initiation of cardiopulmonary resuscitation (CPR) with an emphasis on chest compressions; (3) defibrillation as quickly as possible usually with an automatic external defibrillation applied by the lay rescuer or emergency medical technician (EMT); (4) advanced life support; (5) postcardiac arrest care; and (6) recovery from cardiac arrest. There have been major advances in each of these areas, and survival rates to hospital discharge for out-of-hospital cardiac arrest have increased, particularly for patients found in VT or VF, where survival rates can approach 30% in some regions. Overall survival rates for out-of-hospital cardiac arrest are also higher for patients receiving CPR, with recent studies in Europe reporting survival rates of 16%. Multiple studies have pointed to socioeconomic disparities in the administration of CPR and application of automatic external defibrillators (AEDs) contributing to reduced survival rates from out-of-hospital cardiac arrest in black and Hispanic populations in the United States.

The initial goal of resuscitation is to achieve the return of spontaneous circulation (ROSC). Success is strongly related to the time between collapse and initiation of resuscitation, decreasing markedly after 5 min, and the rhythm at the time of EMT arrival, being best for VT, worse for VF, and poor for PEA and asystole. Outcomes are also determined by the age, clinical state, and comorbidities of the victim prior to the arrest.

INITIAL EVALUATION AND INITIATION OF CPR

The rescuer should check for a response from the victim, shout for help, and call or ask someone else to call their local emergency number (e.g., 911), ideally on a cell phone that can be placed on speaker mode at the patient's side such that the responding dispatcher can provide instructions and queries to the rescuer. Consideration of aspiration or airway obstruction is important, and if suspected, a Heimlich maneuver may dislodge the obstructing body. A trained health care provider would also check for a pulse (taking no longer than 10 seconds so as not to delay initiation of chest compressions) and assess breathing. Gasping respirations and brief seizure activity are common during SCA and may be misinterpreted as breathing and responsiveness. Chest compressions should be initiated without delay and administered at a rate of 100–120/min depressing the sternum by 5 cm (2 in.) and allowing full chest recoil between compressions. Chest compressions generate forward cardiac output with sequential filling and emptying of the cardiac chambers, with competent valves maintaining forward direction of flow. Interruption of chest compressions should be minimized to reduce end-organ ischemia. Ventilation may be administered with two breaths for every 30 compressions if a trained rescuer is present, but for lay rescuers without training, chest compressions alone ("hands-only CPR") are more likely to be effectively applied and of similar benefit. If a second rescuer is present, they should be sent to seek out an AED, which are now widely available in many public areas.

RHYTHM-BASED MANAGEMENT (SEE FIG. 317-2)

The rapidity with which defibrillation/cardioversion is achieved is an important predictor of outcome. A defibrillator, most often an AED, should be applied as soon as available. AEDs are easily used by lay rescuers and trained first responders, such as police officers and trained security guards. When the arrest is witnessed, the use of AEDs by lay responders can improve cardiac arrest survival rates. Once patches are applied to the chest, a brief pause in chest compressions is required to allow the AED to record the rhythm. An AED will advise delivery of a shock if the recorded rhythm meets criteria for VF or VT. Chest compressions are continued while the defibrillator is being charged. As soon as a diagnosis of VF or VT is established, a 200-J biphasic waveform shock should be delivered. Chest compressions are resumed immediately and continue for 2 min until the next rhythm check. If VT/VF is still present, a second maximal energy shock is delivered. This sequence is continued until personnel to administer advanced life support are available or ROSC is achieved. Electrocardiogram (ECG) rhythm strips produced by the AED should be retrieved, as the initial rhythm can be an important consideration in determining the cause of the arrest and to guide further therapy and evaluation if resuscitation is successful.

When advanced cardiac life support is available, an intravenous line is established for administration of medication and consideration given to placement of an advanced airway (endotracheal tube or supraglottic airway device). Intraosseous access may be considered if attempts at intravenous access are not successful or are not feasible. Epinephrine 1 mg every 3–5 min may be administered intravenously or intraosseously. If circulation is not restored or the patient is less than fully conscious despite return of circulation, confirmation that acidosis and hypoxia are adequately addressed should be assessed with arterial blood gas analysis. If metabolic acidosis persists after successful defibrillation and with adequate ventilation, 1 mEq/kg $NaHCO_3$ may be administered.

The cardiac rhythm guides resuscitation when monitoring is available. VT is treated with external shocks synchronized to the QRS when VT is monomorphic, and asynchronous shocks for polymorphic VT or VF. If VT/VF recurs after one or more shocks, amiodarone 300 mg can be administered as a bolus via intravenous or intraosseous route in the hope that arrhythmia recurrence will be prevented after the next shock, followed by a 150-mg bolus if the arrhythmia recurs. If amiodarone fails, lidocaine can be administered.

Consideration of etiology should also guide therapy **(Chaps. 261 and 262)**. Commonly encountered causes of recurrent VT/VF may be due to ongoing myocardial ischemia or infarction that would benefit from emergent coronary angiography and revascularization, or QT prolongation causing the polymorphic VT torsades des pointes that may respond to administration of magnesium. Hyperkalemia should respond to administration of calcium, while other measures are implemented to reduce serum potassium.

PEA/asystole should be managed with CPR, ventilation, and administration of epinephrine. Causes of PEA/asystole that require specific therapy should be considered including airway obstruction, hypoxia, hypovolemia, acidosis, hyperkalemia, hypothermia, toxins, cardiac tamponade, tension pneumothorax, pulmonary embolism, and MI. Naloxone should be administered if opiate overdose is suspected.

POSTCARDIAC ARREST ACUTE MANAGEMENT

Following restoration of effective circulation, the possibility of acute MI should be immediately assessed. The majority of patients who have ST elevation consistent with acute MI will be found to have a culprit coronary stenosis/occlusion and emergent coronary angiography with percutaneous angioplasty, and stenting is recommended. Emergent angiography may also be considered if cardiogenic shock, electrical instability, signs of significant myocardial damage, or ongoing ischemia is present. Emergent or early angiography has not been found to result in better outcomes compared to delayed angiography in patients presenting with out-of-hospital cardiac arrest due to a VT/VF with no ECG evidence of ST-segment elevation. Thus, decisions regarding which patients without ST-segment elevation should undergo urgent angiography are complex, and factors such as hemodynamic or electrical instability and evidence of ongoing ischemia are taken into consideration.

Hemodynamic instability is often present following resuscitation, and further ischemic end-organ damage is a major consideration. Optimizing ventilation with consideration of acidosis, hypoxemia, and electrolyte abnormalities is important. Maintaining systolic blood pressure at >90 mmHg and mean blood pressure >65 mmHg is desirable and may require administration of vasopressors and adjustment of volume status. Potentially treatable reversible causes, including hyperkalemia, severe hypokalemia, and drug toxicity with QT prolongation causing torsades des pointes, should be identified and treated **(Chap. 262)**.

After stable spontaneous circulation is achieved, brain injury due to ischemia and reperfusion is a major determinant of survival and accounts for over two-thirds of deaths. The probability of successful

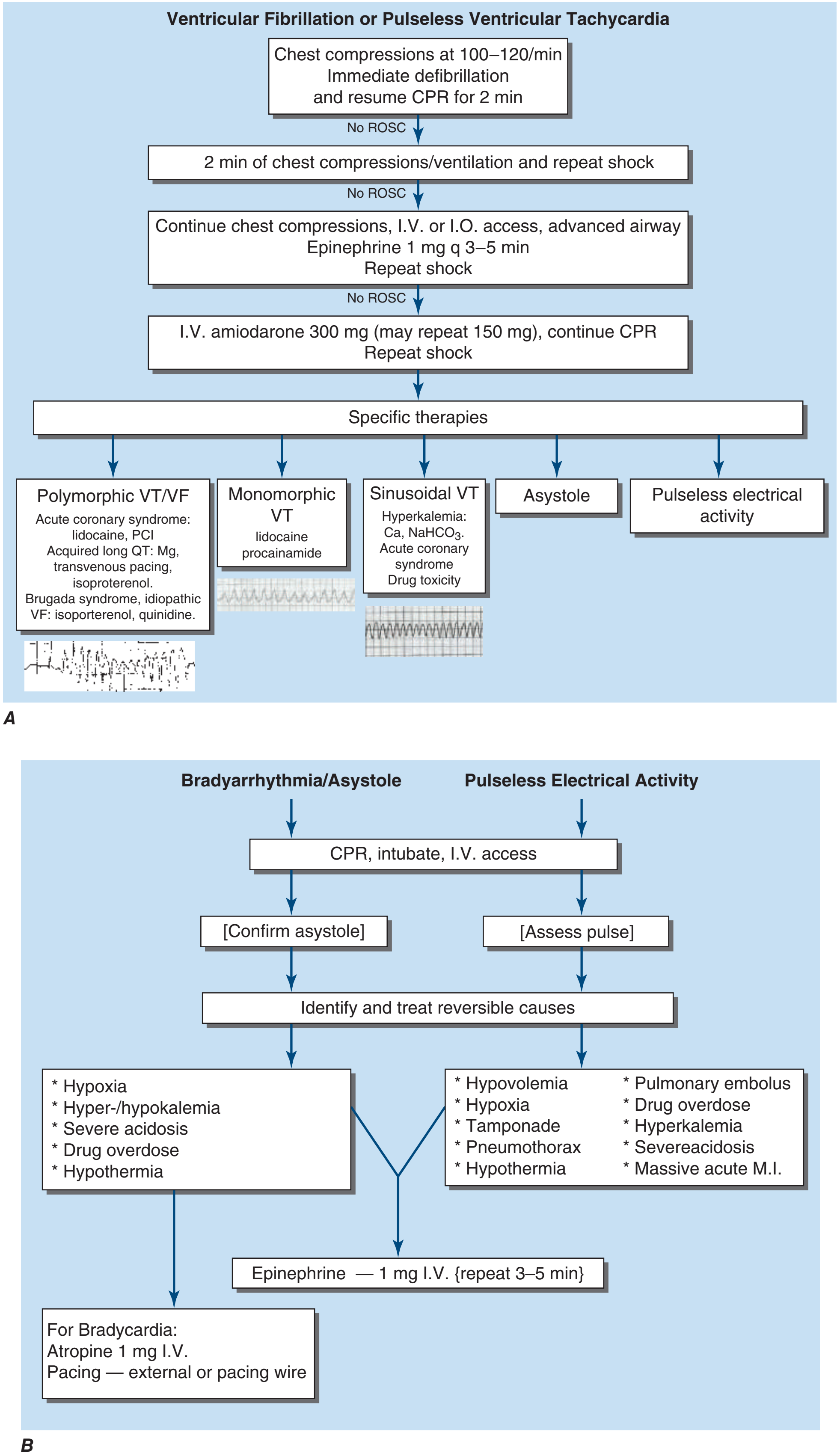
Ventricular Fibrillation or Pulseless Ventricular Tachycardia
Chest compressions at 100–120/min
Immediate defibrillation
and resume CPR for 2 min
No ROSC
2 min of chest compressions/ventilation and repeat shock
No ROSC
Continue chest compressions, I.V. or I.O. access, advanced airway
Epinephrine 1 mg q 3–5 min
Repeat shock
No ROSC
I.V. amiodarone 300 mg (may repeat 150 mg), continue CPR
Repeat shock
Specific therapies
Polymorphic VT/VF
Acute coronary syndrome: lidocaine, PCI
Acquired long QT: Mg, transvenous pacing, isoproterenol.
Brugada syndrome, idiopathic VF: isoporterenol, quinidine.
Monomorphic VT
lidocaine
procainamide
Sinusoidal VT
Hyperkalemia: Ca, NaHCO3.
Acute coronary syndrome
Drug toxicity
Asystole
Pulseless electrical activity
A
Bradyarrhythmia/Asystole
Pulseless Electrical Activity
CPR, intubate, I.V. access
[Confirm asystole]
[Assess pulse]
Identify and treat reversible causes
* Hypoxia
* Hyper-/hypokalemia
* Severe acidosis
* Drug overdose
* Hypothermia
* Hypovolemia
* Hypoxia
* Tamponade
* Pneumothorax
* Hypothermia
* Pulmonary embolus
* Drug overdose
* Hyperkalemia
* Severeacidosis
* Massive acute M.I.
Epinephrine — 1 mg I.V. {repeat 3–5 min}
For Bradycardia:
Atropine 1 mg I.V.
Pacing — external or pacing wire
B

FIGURE 317-2 Algorithm for approach to cardiac arrest due to ventricular tachycardia (VT) or ventricular fibrillation (VF; shockable rhythm). ***A.*** Chest compressions with ventilation and defibrillation or cardioversion should be initiated as soon as possible. Defibrillation should be repeated with minimal interruption of chest compressions. Once an intravenous or intraosseous access is established, administration of epinephrine defibrillation and amiodarone and defibrillation are performed. Further therapy can be guided by possible causes as suggested by the initial or recurrent cardiac rhythm as shown. CPR, cardiopulmonary resuscitation; I.O., intraosseous; I.V., intravenous; PCI, percutaneous coronary intervention; ROSC, return of spontaneous circulation. ***B.*** Algorithm for approach to cardiac arrest due to bradyarrhythmias/asystole and pulseless electrical activity. Chest compressions with ventilation (and intubation) should be initiated as soon as possible, and intravenous access should be obtained. Once an intravenous or intraosseous access is established, administration of epinephrine is performed. At the same time, an investigation for potential reversible causes should be made and any such causes should be treated if present. For bradycardic rhythms, atropine 1 mg administered intravenously and external subcutaneous or transvenous pacing are also performed. Defibrillation should be repeated with minimal interruption of chest compressions. Further therapy can be guided by possible causes. CPR, cardiopulmonary resuscitation; I.O., intraosseous; I.V., intravenous; M.I., myocardial infarction.

neurologic recovery decreases rapidly with time from collapse to ROSC and is <30% at 5 min in the absence of bystander CPR. The time between collapse and restoration of circulation is generally imprecise, and some patients have a period of hypotensive VT prior to complete collapse, such that a reported long period before the arrival of rescuers does not always preclude good neurologic recovery. Therapeutic hypothermia (targeted temperature management) has been shown to improve the likelihood of survival and neurologic recovery in patients who present with shockable (VT or VF) rhythms and is recommended for all cardiac arrest patients who remain comatose, regardless of presenting rhythm, who have lack of purposeful response to verbal commands following ROSC. A constant target temperature of 32–37.5°C for at least 24 h is recommended, although a recent trial failed to demonstrate benefit compared with a strategy of targeted normothermia with early and aggressive treatment of fever. Shivering suppression with analgesics and sedatives may be needed. Induction of hypothermia should be started in the hospital, as no benefit was shown for implementation before hospital arrival, and administration of large volumes of cold saline for this purpose increased the risk of pulmonary edema. Brain injury is often accompanied by seizures and status epilepticus that may have further deleterious effects, warranting periodic or continuous electroencephalography (EEG) monitoring. Treatment of clinically apparent seizures is indicated and may also be reasonable in those with EEG patterns on the ictal-interictal continuum on monitoring. Several other therapies hoped to improve postarrest outcomes have been assessed but have not been shown to be beneficial, including administration of corticosteroids, hemofiltration, and efforts to tightly control blood glucose.

Hypothermia and sedation preclude reliable prognostication for neurologic recovery. Functional neurologic assessment for neurologic recovery is generally deferred for at least 72 hours after return to normothermia, typically 4–5 days after the cardiac arrest. Features that predict poor outcome include the absence of pupillary reflex to light, status myoclonus, absence of EEG reactivity to external stimuli, and persistent burst suppression on EEG.

■ LONG-TERM MANAGEMENT AFTER SURVIVAL OF OUT-OF-HOSPITAL CARDIAC ARREST

For patients who survive cardiac arrest and have neurologic recovery, the likely underlying cause of the arrest guides further treatment. For arrests not due to an obvious noncardiac cause, a full evaluation for the forms of structural heart disease outlined in Fig. 317-1 and Table 317-2 should be performed including an assessment for underlying CAD and ischemia as well as echocardiography and/or cardiac magnetic resonance imaging (MRI) to look for evidence of prior MI, valvular disease, and nonischemic cardiomyopathies, and to provide an assessment of left ventricular ejection fraction (LVEF). If the initial evaluation is not definitive or is suggestive of an inflammatory cardiomyopathy (i.e., sarcoidosis, myocarditis), a cardiac positron emission tomography (PET) scan and/or endomyocardial biopsy may also be performed. Patients without obvious structural abnormalities should undergo an evaluation for primary electrical disease (long QT syndrome [LQTS], Brugada syndrome, early repolarization syndrome, or Wolff-Parkinson-White syndrome). In cases where a heritable syndrome is suspected, further genetic evaluation should be considered. Diagnostic electrophysiology studies are warranted in selected patients to assess inducible arrhythmias, or provocative testing, such as with epinephrine challenge for LQTS, or sodium channel blocker (e.g., procainamide) challenge for Brugada syndrome.

Patients with shockable rhythms at arrest (VF and VT) that are not deemed to have been due to a transient reversible cause and have reasonable life expectancy should undergo insertion of an ICD for secondary prevention of SCA/SCD. Most of these patients will be found to have CAD. Patients with a VF arrest that occurs within the first 48 h of a documented acute MI generally do not require an ICD because they have a similar risk of sudden death over the next 5 years as infarct survivors who did not have a cardiac arrest. However, patients who have a large infarction with acutely depressed LVEF (e.g., <35%) have an increased risk for future development of life-threatening ventricular arrhythmias related to reentry in the infarct scar **(Chap. 259)**. The percentage of patients with such large infarcts has been declining due to improved revascularization strategies for acute MI. Implantation of an ICD early after MI in these patients does not, however, improve overall survival, in part because a significant number of sudden deaths in the first 3 months are due to recurrent myocardial ischemia or myocardial rupture, rather than cardiac arrhythmias. For patients with large infarcts, a wearable defibrillator that will treat VT/VF if it occurs may be used while left ventricular remodeling is taking place, followed by reevaluation of arrhythmia risk after the infarct is healed to determine if an ICD is warranted. Patients who experience VF in the hospital >48 h after MI or in the setting of myocardial ischemia without infarction may be at risk for recurrent VT/VF. These patients should be evaluated and optimally treated for ischemia. If there is evidence that clearly implicates ischemia immediately preceding the onset of VF without evidence of a prior MI, coronary revascularization may be adequate therapy. Others may warrant ICD implantation. When the cardiac arrest is due to sustained monomorphic VT, a prior infarct scar is often present, and the recurrence rate is significant regardless of whether the arrest occurred in association with elevated serum troponin. In this circumstance, even when revascularization is performed for ischemia, an ICD is usually warranted owing to the risk of recurrence of scar-related VT.

Patients who have cardiac arrest due to a treatable reversible cause, such as hyperkalemia or drug toxicity with QT prolongation causing torsades des pointes **(Chap. 262)**, which can be adequately addressed and prevented by other means, do not usually need an ICD. An ICD is usually recommended for cardiac arrest due to VT or VF without a clearly reversible cause, particularly when structural heart disease, such as hypertrophic or dilated cardiomyopathy, arrhythmogenic cardiomyopathy, cardiac sarcoidosis, or a cardiac syndrome associated with sudden death, including Brugada syndrome, or LQTS is present **(Chaps. 261 and 262)**. In patients with structural heart disease, it is important to recognize that life-threatening arrhythmias can be an indication of terminal, end-stage heart disease with minimal prospect for meaningful survival despite successful resuscitation, and ICDs will not alter the course of these patients and should not be implanted in this situation unless there is a prospect for cardiac replacement therapy with future cardiac transplantation or a ventricular assist device.

Finally, the psychological needs of both the SCA survivor and family members need to be assessed and addressed. Comprehensive rehabilitation and treatment plans for physical, neurologic, cardiopulmonary, and cognitive impairments of the SCA survivor should be formulated before hospital discharge.

PREVENTION OF SCD

Although advances in CPR and postresuscitation care have improved survival rates after cardiac arrest, 90% of patients will not survive to be discharged from the hospital. Of those who do survive, a proportion (~20%) are left with severe neurologic and/or physical disability. The majority of cardiac arrests do not occur in public places where AEDs and rapid defibrillation have the greatest impact. Patients who suffer an arrest at home also have longer EMS response times and are much less likely to be found in VF. Finally, 50% of cardiac arrests are not witnessed, precluding effective resuscitation efforts. Thus, preventive efforts are critical to reducing mortality from cardiac arrest.

■ SCD RISK STRATIFICATION

The presence of overt structural heart disease and/or primary electrical heart disease is associated with an increased risk of SCD that varies with the severity and type of disease. For patients with structural heart disease, depressed left ventricular function is the best validated marker for risk, and clinical HF elevates risk further. After MI, SCD risk increases gradually as the LVEF decreases to 40% and then exponentially thereafter. In addition to LVEF and congestive heart failure, other potential markers of increased SCD risk in the setting of structural heart disease include unexplained syncope, sustained VT induced at electrophysiologic study (EP study), left ventricular scar size and heterogeneity on cardiac magnetic resonance, markers of altered autonomic function and altered repolarization, and QRS prolongation. The majority of these tests, with the exception of the EP study in post-MI patients, broadly predict death from cardiovascular causes and are not able to discriminate between patients who will die suddenly from an arrhythmia and those who will die of other cardiac causes. For example, patients with the greatest degree of systolic HF and/or lowest LVEF, although at elevated risk for SCD, are more likely to die from pump failure. Although sustained VT at EP study does identify individuals at a higher risk of SCA versus non-SCA in certain subsets of patients, the sensitivity of the test is generally inadequate when LV function is significantly reduced.

PREVENTIVE THERAPIES FOR SCD IN HIGH-RISK POPULATIONS

Therapy with beta-adrenergic blockers has been demonstrated to reduce SCD risk in a multitude of settings, including after MI, among patients with ischemic and nonischemic cardiomyopathy, and in LQTS. Angiotensin-converting enzyme inhibitors, aldosterone antagonists, and more recently angiotensin receptor/neprilysin inhibitors and sodium-glucose cotransporter 2 inhibitors (SGLT2i) have been associated with reductions in SCD in subsets of patients with structural heart disease, primarily ischemic and nonischemic cardiomyopathy accompanied by HF. Coronary artery bypass grafting has also been associated with reductions in SCD risk, and revascularization may lower SCD risk through reduction in ischemic events and resultant improvements in left ventricular systolic function by reducing areas of hibernating myocardium.

For patients whose disease continues to confer a substantial risk of sustained VT or VF on optimal medical therapy, an ICD is recommended (Table 317-3). The ICD indication in these patients is referred to as “primary prevention of sudden death.” The indications for primary prevention ICDs vary depending on the type of underlying structural heart disease and its severity, and the strength of evidence varies by indication. In some cases, there are slight differences in the

TABLE 317-3 Implantable Cardioverter Defibrillator (ICD) Indications

	ESC GUIDELINES	AHA/ACC/HRS GUIDELINES	LEVEL OF EVIDENCE
Secondary Prevention			
All Disease States with VT or VF			
Cardiac arrest due to VF or hemodynamically unstable sustained VT after evaluation to define the cause of the event and to exclude any completely reversible causes	Class I	Class I	A
Structural heart disease and spontaneous hemodynamically stable sustained VT	Class IIa	Class I	B
Structural heart disease and spontaneous hemodynamically not-tolerated sustained VT	Class I	Class I	B
Sustained VT and normal or near-normal ventricular function	–	Class IIa	C
Syncope			
Patients with syncope and inducible VT or VF at EP study	–	Class I	B
Patients with syncope and structural heart disease in whom invasive and noninvasive studies have failed to determine a cause	–	Class IIb	C
Primary Prevention			
Coronary Artery Disease			
LVEF ≤35% + NYHA functional class II–III	Class I	Class I	A
LVEF ≤35% (ESC) + NYHA functional class I	Class IIa	–	B
LVEF ≤30% (AHA) + NYHA functional class I	–	Class I	A
LVEF ≤40% + NSVT + inducible monomorphic VT	Class IIa	Class I	B
LVEF ≤40% + unexplained syncope + inducible monomorphic VT	Class IIa	Class I	B
Nonischemic Cardiomyopathy			
LVEF ≤35% + NYHA functional class II–II	Class IIa	Class I	A
LVEF ≤35% + NYHA functional class I	–	Class IIb	B
Pathogenic mutation in *LMNA* gene + 2 or more risk factors (NSVT, LVEF <45%, nonmissense mutation, male sex)	–	Class IIa	C
LMNA mutation with estimated 5-year risk of VA ≥10% + NSVT or LVEF <50% or AV conduction delay	Class IIa	–	C
LVEF >35% and ≥ 2 risk factors (syncope, LGE on CMR, inducible VT at PES, pathogenic mutations in *PLN*, *FLNC*, and *RBM20* genes)	Class IIa	–	C
NYHA Functional Class IV Candidates for Advanced Heart Failure Therapy			
Awaiting cardiac transplant	Class IIa	Class IIa	C
With destination LVAD and sustained VT	Class IIa	–	B
With destination LVAD	–	Class IIa	B
Arrhythmogenic Right Ventricular Cardiomyopathy			
Arrhythmic syncope	Class IIa	Class IIa	B
Moderate right (<40%) or left (<45%) ventricular dysfunction and NSVT or inducible monomorphic VT	Class IIa	–	C
Significant right ventricular dysfunction with LVEF ≤35%	–	Class I	B
Significant right ventricular dysfunction with RVEF ≤35%	Class IIa	Class I	C
Hypertrophic Cardiomyopathy			
Maximum left ventricular wall thickness >30 mm	–	Class IIa	B
SCD in first-degree relative presumably due to HCM	–	Class IIa	B
Unexplained syncope	–	Class IIa	B
NSVT or abnormal blood pressure response during exercise + additional SCD risk modifiers or high-risk features	–	Class IIa	C

(*Continued*)

TABLE 317-3 Implantable Cardioverter Defibrillator (ICD) Indications (Continued)

	ESC GUIDELINES	AHA/ACC/HRS GUIDELINES	LEVEL OF EVIDENCE
NSVT or abnormal blood pressure response during exercise without additional SCD risk modifiers or high-risk features	–	Class IIb	C
Estimated 5-year risk of sudden death based on the HCM Risk–SCD Calculator ≥6%	Class IIa	–	B
Estimated 5-year risk of sudden death based on HCM Risk–SCD Calculator (≥4 to <6%) AND Significant LGE on CMR or LVEF <50% or Abnormal blood pressure during exercise test or Left ventricular apical aneurysm or Presence of sarcomeric pathogenic mutation	Class IIa	–	B
Estimated 5-year risk of sudden death based on the HCM Risk–SCD Calculator ≥4 to <6%	Class IIb	–	B
Estimated 5-year risk of sudden death based on the HCM Risk–SCD Calculator <4% AND Significant LGE on CMR or LVEF <50% or Left ventricular apical aneurysm	Class IIb	–	B
Congenital Long QT Syndrome			
Symptomatic high-risk patients + ineffectiveness or intolerance of β-blocker therapy (high risk: QTc >500 ms, genotypes LQTS 2 and LQTS 3, LQTS 2 females, age <40 years, onset of symptoms <10 years, recurrent syncope)	–	Class I	B
Unexplained syncope during β-blocker and genotype-specific therapy	Class I	–	B
Symptomatic patients + intolerance or contraindication of β-blocker and genotype-specific therapy	Class IIa	–	B
Asymptomatic patients with QTc >500 ms during β-blocker treatment	–	Class IIb	B
Asymptomatic patients with high-risk profile according to 1-2-3- LQTS-Risk calculator	Class IIb	–	
Catecholaminergic Polymorphic Ventricular Tachycardia (CPVT)			
Syncope during β-blocker treatment	–	Class I	C
Syncope during combined β-blocker and flecainide treatment	Class IIa	–	C
Brugada Syndrome			
Spontaneous type 1 Brugada ECG + recent history of syncope presumed due to ventricular arrhythmia	Class IIa	Class I	C
Asymptomatic patients with type 1 Brugada ECG and inducible ventricular fibrillation	Class IIb	–	C
Cardiac Sarcoidosis			
Cardiac sarcoidosis with 1 or more risk factors for SCD	–	Class IIa	C
Indication for permanent pacemaker implantation regardless of LVEF	Class IIa	–	C
LVEF >35% and significant LGE on CMR after resolution of acute inflammation	Class IIa	–	C
LVEF 35–50% and inducible VT at PES	Class IIa	–	C
Familial Cardiomyopathy			
Patients with familial cardiomyopathy associated with SCD	–	Class IIb	C
Left Ventricular Noncompaction			
Patients with left ventricular noncompaction	–	Class IIb	C

Abbreviations: ACC, American College of Cardiology; AHA, American Heart Association; AV, atrioventricular; CMR, cardiac magnetic resonance; ECG, electrocardiogram; EP, electrophysiologic; ESC, European Society of Cardiology; HCM, hypertrophic cardiomyopathy; HRS, Heart Rhythm Society; LGE, late gadolinium enhancement; LQTS, long QT syndrome; LVAD, left ventricular assist device; LVEF, left ventricular ejection fraction; NSVT, nonsustained ventricular tachycardia; NYHA, New York Heart Association; PES, programmed electrical stimulation; RVEF, right ventricular ejection fraction; SCD, sudden cardiac death; VA, ventricular arrhythmia; VF, ventricular fibrillation; VT, ventricular tachycardia.

strength of recommendation among professional societies in Europe and America for implantation of an ICD, but overall, there is agreement on ICD implantation for the highest-risk patients. In patients with a history of MI >40 days ago, primary prevention ICDs are indicated for those with class II–III New York Heart Association (NYHA) HF and LVEF <35% and those who are NYHA functional class I with LVEF <30%. Although ICDs have not been found to be beneficial when implanted within 40 days of an MI, those with recent or old MI, nonsustained VT, LVEF <40%, and inducible sustained VT at EP study also warrant an ICD. In general, these criteria are not applied to patients who are within 90 days of myocardial revascularization, since some will experience improvement in ventricular function and older trial data suggested there was no benefit with ICDs in these patients. High-risk patients with low LVEFs may be considered for a wearable defibrillator with later reassessment of ventricular function and ICD placement.

ICDs for primary prevention of sudden death are also recommended for patients with diseases other than CAD that put them at risk for SCD. Primary prevention ICDs are currently indicated in select high-risk patients with HCM, arrhythmogenic right ventricular dysplasia, cardiac sarcoidosis, Brugada syndrome, and congenital LQTS. ICDs are currently also recommended for those with nonischemic dilated cardiomyopathy (DCM) who have an LVEF ≤35% and who have NYHA functional class II or III symptoms on guideline-directed medical therapy. In addition to invasive electrophysiologic testing, new risk stratification methods for certain conditions have emerged, including genetic findings and the presence of late gadolinium enhancement (LGE) on MRI. For example, the European guidelines account for pathogenic mutations associated with a high risk of ventricular arrhythmias in patients with nonischemic cardiomyopathy, hypertrophic cardiomyopathy, and long QT syndrome. The presence of LGE has been shown to be a risk factor for sudden death in patients with cardiac sarcoidosis and hypertrophic cardiomyopathy, and thus, this imaging finding has been incorporated into ICD implantation guidelines.

TABLE 317-4 Implantable Cardioverter Defibrillator (ICD) Not Indicated

Patients who do not have a reasonable expectation of survival with an acceptable functional status for at least 1 year, even if they meet ICD implantation criteria.
Patients with incessant VT or VF.
Patients with significant psychiatric illnesses that may be aggravated by device implantation or that may preclude systematic follow-up.
Patients with drug-refractory New York Heart Association class IV congestive heart failure who are not candidates for cardiac transplantation or cardiac resynchronization therapy.
Syncope of undetermined cause in a patient without inducible ventricular tachyarrhythmias and without structural heart disease.
VF or VT is amenable to surgical or catheter ablation in patients without other disease predisposing to sudden cardiac arrest (e.g., atrial arrhythmias associated with Wolff-Parkinson-White syndrome, RV or LV outflow tract VT, idiopathic VT, or fascicular VT in the absence of structural heart disease).
Patients with ventricular tachyarrhythmias due to a completely reversible disorder in the absence of structural heart disease (e.g., electrolyte imbalance, drugs, or trauma).

Abbreviations: LV, left ventricular; RV, right ventricular; VF, ventricular fibrillation; VT, ventricular tachycardia.

Source: Adapted from H Könemann et al: Management of ventricular arrhythmias worldwide: comparison of the latest ESC, AHA/ACC/HRS, and CCS/CHRS guidelines. JACC Clin Electrophysiol 9:715, 2023.

Although the ICD is very effective for treatment of arrhythmic sudden death, competing causes of mortality must be considered in patients with severe cardiomyopathy. Data from a recent randomized trial, the Danish Study to Assess the Efficacy of ICDs in Patients with Non-ischemic Systolic Heart Failure on Mortality (DANISH), performed in patients with nonischemic DCM and LVEF ≤35%, who also had elevated N-terminal pro-B-type natriuretic peptide levels and NYHA class II–IV HF, have resulted in some debate regarding the utility of ICDs in this population. This trial did not demonstrate an overall mortality benefit of the ICD despite a reduction in the incidence of SCD. In subgroup analyses, mortality benefits were observed in younger patients in whom the competing risk of dying from other causes of death was lower. These data underscore the importance of considering competing risks for other causes of mortality when deciding to implant a primary prevention ICD. Patients who are likely to die from other causes are unlikely to benefit from an ICD. Patients who do not have a reasonable expectation of survival with an acceptable functional status for at least 1 year should not undergo ICD placement. There are also other circumstances where an ICD is not indicated even if there is a significant sudden death risk (Table 317-4).

THE CHALLENGE OF SCD PREVENTION (FIG. 317-3)

The Greatest Number of Sudden Deaths Occur in "Low-Risk" Patients While patients with reduced left ventricular

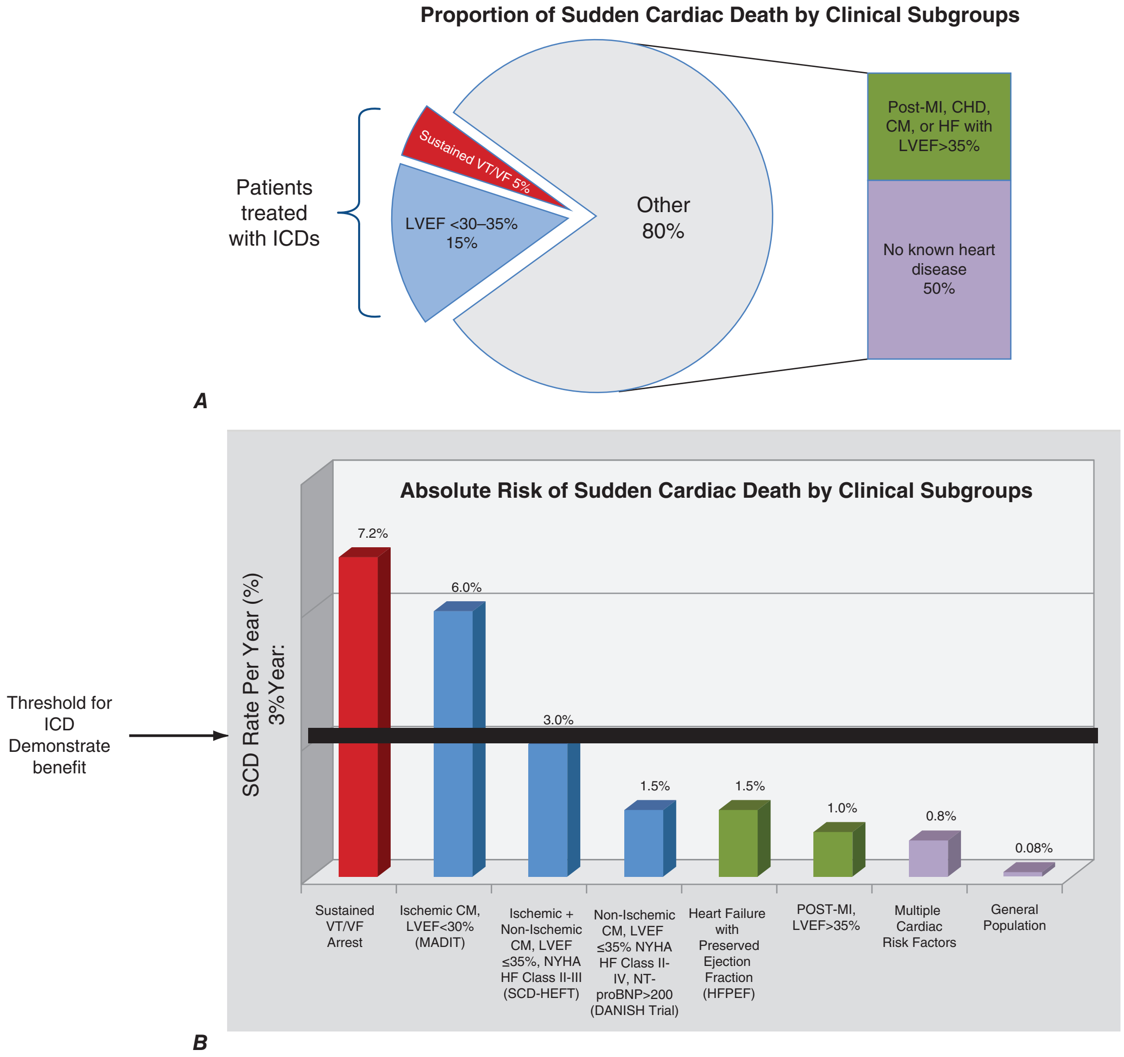

FIGURE 317-3 ***A.*** Proportion of sudden cardiac deaths that occur in clinical subgroups of the population treated and not treated with implantable cardioverter defibrillators (ICDs). ***B.*** Absolute risk of sudden cardiac death within clinical subgroups in comparison to the threshold of risk where ICDs demonstrated benefit.

function and HF are at substantially elevated SCD risk, only ~20% of all SCDs occur in patients with poor left ventricular function. Most SCDs occur in individuals with preserved ventricular function who would not qualify for a primary prevention ICD. Although SCD rates are elevated compared to the general population, the absolute SCD risk in patients with CHD or HF who have an LVEF >35% is not high enough to warrant consideration of ICD therapy. While the incidence of SCD is lower in patients with preserved LVEF, SCD accounts for a greater proportion of cardiac deaths, and active efforts are being made to advance SCD risk stratification in this segment of the population. However, at present, SCD prevention primarily involves cardiac risk factor modification and standard medical therapy for the underlying condition.

Preventing Sudden Death in the General Population Only about one-half of men and one-third of women who suffer SCA are recognized to have heart disease prior to the event, and only half have warning symptoms prior to the event. SCD often occurs without warning as the first manifestation of cardiac disease. In order to prevent these SCDs, preventive interventions would need to be employed broadly to the general population. Although several risk scores have recently been developed with the intent to stratify SCD risk in low-risk populations, the clinical utility to date is limited by the low absolute incidence of SCD, which is estimated to be only 50–90 per 100,000 in the general adult population. Therefore, current efforts aimed at preventing SCD in general populations primarily focus on modification of the SCD risk factors outlined previously. Individuals who adhere to a low-risk, healthy lifestyle that includes avoidance of smoking, maintaining a healthy body weight, participating in moderate exercise, and a Mediterranean-type dietary pattern have markedly lower rates of SCD. A substantial number of SCDs are likely to be preventable through lifestyle modifications and treatment of risk factors.

FURTHER READING

Al-Khatib SM et al: 2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. J Am Coll Cardiol 72:e91, 2018.

Callaway CW et al: Part 8: Post-cardiac arrest care: 2015 American Heart Association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 132:S465, 2015.

Dankiewicz J et al: Hypothermia versus normothermia after out-of-hospital cardiac arrest. N Engl J Med 384:2283, 2021.

Deo R, Albert CM: Epidemiology and genetics of sudden cardiac death. Circulation 125:620, 2012.

Marijon E et al: Lancet Commission to reduce the global burden of sudden cardiac death: A call for multidisciplinary action. Lancet 402:883, 2023.

Merchant RM et al: Part 1: Executive summary: 2020 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 142:S337, 2020.

Myerburg RJ et al: Pulseless electric activity: Definition, causes, mechanisms, management, and research priorities for the next decade: Report from a National Heart, Lung, and Blood Institute workshop. Circulation 128:2532, 2013.

Perman SM et al: 2023 American Heart Association focused update on adult advanced cardiovascular life support: An update to the American Heart Association Guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 149:e254, 2024.

Zeppenfeld K et al: 2022 ESC guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death. Eur Heart J 43:3997, 2022.

Section 3 Neurologic Critical Care

318

Nervous System Disorders in Critical Care

J. Claude Hemphill, III, Wade S. Smith,
S. Andrew Josephson, Daryl R. Gress

Life-threatening neurologic illness may be caused by a primary disorder affecting any region of the neuraxis or may occur as a consequence of a systemic disorder such as hepatic failure, multisystem organ failure, or cardiac arrest **(Table 318-1)**. Neurologic critical care focuses on preservation of neurologic tissue and prevention of secondary brain injury caused by ischemia, hemorrhage, edema, herniation, and elevated intracranial pressure (ICP). Encephalopathy is a general term describing brain dysfunction that is diffuse, global, or multifocal. Severe acute encephalopathies represent a group of various disorders due to different neurologic or systemic etiologies but that share the common themes of primary and secondary brain injury.

PATHOPHYSIOLOGY

Brain Edema Swelling, or edema, of brain tissue occurs with many types of brain injury. The two principal types of edema are vasogenic and cytotoxic. *Vasogenic edema* refers to the influx of fluid and solutes into the brain through an incompetent blood-brain barrier (BBB). In the normal cerebral vasculature, endothelial tight junctions associated with astrocytes create an impermeable barrier (the BBB), through which access into the brain interstitium is dependent upon specific transport mechanisms. The BBB may be compromised in ischemia, trauma, infection, and metabolic derangements, and typically develops rapidly following injury. *Cytotoxic edema* results from cellular swelling, membrane breakdown, and ultimately cell death. Clinically significant brain edema usually represents a combination of vasogenic and cytotoxic components. Edema can lead to increased ICP as well as tissue shifts and brain displacement or herniation from focal processes **(Chap. 30)**. These tissue shifts can cause injury by mechanical distention and compression in addition to the ischemia of impaired perfusion consequent to the elevated ICP.

Ischemic Cascade and Cellular Injury When delivery of substrates, principally oxygen and glucose, is inadequate to sustain cellular function, a series of interrelated biochemical reactions known as the *ischemic cascade* is initiated **(see Fig. 437-2)**. The release of excitatory amino acids, especially glutamate, leads to influx of calcium and sodium ions, which disrupt cellular homeostasis. An increased intracellular calcium concentration may activate proteases and lipases, which then lead to lipid peroxidation and free radical–mediated cell membrane injury. Cytotoxic edema ensues, and ultimately necrotic cell death and tissue infarction occur. This pathway to irreversible cell death is common to ischemic stroke, global cerebral ischemia, and traumatic brain injury.

Penumbra refers to areas of ischemic brain tissue that have not yet undergone irreversible infarction, implying that these regions are potentially salvageable if ischemia can be reversed. Factors that may exacerbate ischemic brain injury include systemic hypotension and hypoxia, which further reduce substrate delivery to vulnerable brain tissue, and fever, seizures, and hyperglycemia, which can increase cellular metabolism, outstripping compensatory processes. Clinically, these events are known as *secondary brain insults* because they lead to exacerbation of the primary brain injury. Prevention, identification, and treatment of secondary brain insults are fundamental goals of management.

An alternative pathway of cellular injury is *apoptosis*. This process implies programmed cell death, which may occur in the setting of ischemic stroke, global cerebral ischemia, traumatic brain injury, and

TABLE 318-1 Neurologic Disorders in Critical Illness

LOCALIZATION ALONG NEUROAXIS	SYNDROME
Central Nervous System	
Brain: Cerebral hemispheres	Global encephalopathy
	Delirium
	Sepsis
	Organ failure—hepatic, renal
	Medication related—sedatives, hypnotics, analgesics, H_2 blockers, antihypertensives
	Drug overdose
	Electrolyte disturbance—hyponatremia, hypoglycemia
	Hypotension/hypoperfusion
	Hypoxia
	Meningitis
	Subarachnoid hemorrhage
	Wernicke's disease
	Seizure—postictal or nonconvulsive status epilepticus
	Hypertensive encephalopathy
	Hypothyroidism—myxedema
	Focal deficits
	Ischemic stroke
	Tumor
	Abscess, subdural empyema
	Intraparenchymal hemorrhage
	Subdural/epidural hematoma
Brainstem/cerebellum	Mass effect and compression
	Basilar artery thrombosis
	Intraparenchymal hemorrhage
	Central pontine myelinolysis
Spinal cord	Mass effect and compression
	Disk herniation
	Epidural hematoma
	Epidural abscess
	Ischemia—hypotension/embolic
	Trauma
	Myelitis
Peripheral Nervous System	
Peripheral nerve	
Axonal	Critical illness polyneuropathy
	Neuromuscular blocking agent complications
	Metabolic disturbances, uremia, hyperglycemia
	Medication effects—chemotherapeutic, antiretroviral
Demyelinating	Guillain-Barré syndrome
	Chronic inflammatory demyelinating polyneuropathy
Neuromuscular junction	Prolonged effect of neuromuscular blockade
	Medication effects—aminoglycosides
	Myasthenia gravis, Lambert-Eaton syndrome, botulism
Muscle	Critical illness myopathy
	Cachectic myopathy
	Acute necrotizing myopathy
	Thick-filament myopathy
	Electrolyte disturbances—hypokalemia/ hyperkalemia, hypophosphatemia
	Rhabdomyolysis

possibly intracerebral hemorrhage. Apoptotic cell death can be distinguished histologically from the necrotic cell death of ischemia and is mediated through a different set of biochemical pathways; apoptotic cell death occurs without cerebral edema and therefore is often not seen on brain imaging. At present, interventions for prevention and treatment of apoptotic cell death remain less well defined than those for ischemia.

Cerebral Perfusion and Autoregulation Brain tissue requires constant perfusion in order to ensure adequate delivery of substrate. The hemodynamic response of the brain has the capacity to preserve perfusion across a wide range of systemic blood pressures. Cerebral perfusion pressure (CPP), defined as the mean systemic arterial pressure (MAP) minus the ICP, provides the driving force for circulation across the capillary beds of the brain. *Autoregulation* refers to the physiologic response whereby cerebral blood flow (CBF) is regulated via alterations in cerebrovascular resistance in order to maintain perfusion over wide physiologic changes such as neuronal activation or changes in hemodynamic function. If systemic blood pressure drops, cerebral perfusion is preserved through vasodilation of arterioles in the brain; likewise, arteriolar vasoconstriction occurs at high systemic pressures to prevent hyperperfusion, resulting in fairly constant perfusion across a wide range of systemic blood pressures **(Fig. 318-1)**. At the extreme limits of MAP or CPP (high or low), flow becomes directly related to perfusion pressure. These autoregulatory changes occur in the microcirculation and are mediated by vessels below the resolution of those seen on angiography. CBF is also strongly influenced by pH and $\mathrm{Paco_2}$. CBF increases with hypercapnia and acidosis and decreases with hypocapnia and alkalosis because of pH-related changes in cerebral vascular resistance. This forms the basis for the use of hyperventilation to lower ICP, and this effect on ICP is mediated through a decrease in both CBF and intracranial blood volume. Cerebral autoregulation is a complex process critical to the normal homeostatic functioning of the brain, and this process may be disordered focally and unpredictably in disease states such as traumatic brain injury and severe focal cerebral ischemia.

Cerebrospinal Fluid (CSF) and ICP The cranial contents consist essentially of brain, CSF, and blood. CSF is produced principally in the choroid plexus of each lateral ventricle, exits the brain via the foramens of Luschka and Magendie, and flows over the cortex to be absorbed into the venous system along the superior sagittal sinus. In adults, ~150 mL of CSF are contained within the ventricles and surrounding the brain and spinal cord; the cerebral blood volume is also ~150 mL. The bony skull offers excellent protection for the brain but allows little tolerance for additional volume. Significant increases in volume eventually result in increased ICP. Obstruction of CSF outflow, edema of cerebral tissue, or increases in volume from tumor or hematoma may increase ICP. Elevated ICP diminishes cerebral perfusion and can lead to tissue ischemia. Ischemia in turn may lead to vasodilation via autoregulatory mechanisms designed to restore cerebral perfusion. However, vasodilation also increases cerebral blood volume, which in turn then increases ICP, lowers CPP, and provokes further ischemia. This vicious cycle is commonly seen in traumatic brain injury, massive intracerebral hemorrhage, and large hemispheric infarcts with significant tissue shifts.

APPROACH TO THE PATIENT

Severe Brain Dysfunction

Critically ill patients with severe central nervous system (CNS) dysfunction require rapid evaluation and intervention in order to limit primary and secondary brain injury. Initial neurologic evaluation should be performed concurrent with stabilization of basic respiratory, cardiac, and hemodynamic parameters. Significant barriers may exist to neurologic assessment in the critical care unit, including endotracheal intubation and the use of sedative or paralytic agents to facilitate procedures.

An impaired level of consciousness is common in critically ill patients. The essential first task in assessment is to determine

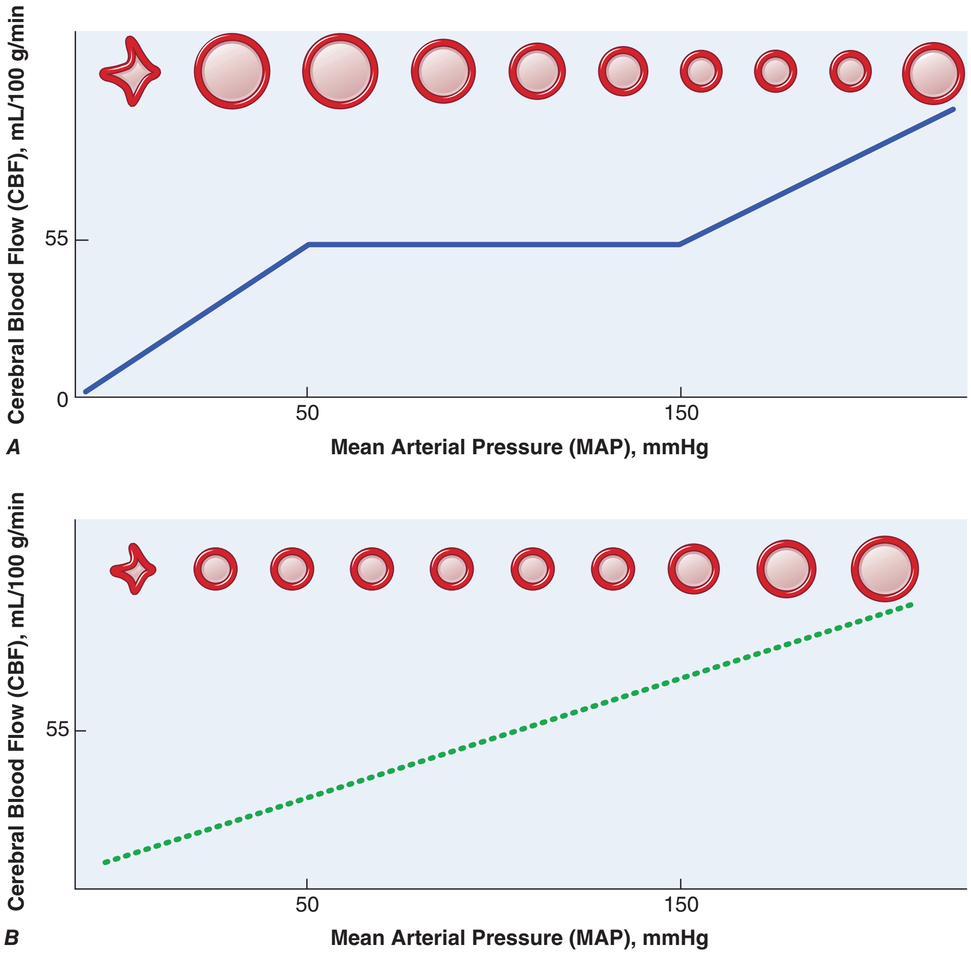

FIGURE 318-1 Pressure autoregulation of cerebral blood flow. In the normal state where autoregulation is intact ***A***, cerebral perfusion is constant over a wide range of systemic blood pressures (BP). This is mediated by dilation and constriction of small cerebral arterioles (*round circles*). Below the BP threshold for maximal dilation, cerebral blood flow becomes pressure-dependent and decreases, whereas above the threshold for maximum constriction, cerebral blood flow increases with increasing systemic BP. In severe brain injury, autoregulatory mechanisms may be impaired and cerebral blood flow becomes pressure-dependent throughout (***B***). At the extremes of BP, there may be vascular collapse (very low BP) or forced vasodilation (very high BP).

whether the cause of dysfunction is related to a diffuse, usually metabolic, process or whether a focal, usually structural, process is implicated. Examples of diffuse processes include metabolic encephalopathies related to organ failure, drug overdose, or hypoxia-ischemia. Focal processes include ischemic and hemorrhagic stroke and traumatic brain injury, especially with intracranial hematomas. Because these two categories of disorders have fundamentally different causes, treatments, and prognoses, the initial focus is on making this distinction rapidly and accurately. **The approach to the comatose patient is discussed in Chap. 30; etiologies are listed in Table 30-1.**

Minor focal deficits may be present on the neurologic examination in patients with metabolic encephalopathies. However, the finding of prominent focal signs such as pupillary asymmetry, hemiparesis, gaze palsy, or visual field deficit should suggest the possibility of a structural lesion. All patients with a decreased level of consciousness associated with focal findings should undergo an urgent neuroimaging procedure, as should all patients with coma of unknown etiology. Computed tomography (CT) scanning is usually the most appropriate initial study because it can be performed quickly in critically ill patients and demonstrates hemorrhage, hydrocephalus, and intracranial tissue shifts well. Magnetic resonance imaging (MRI) may provide more specific information in some situations, such as acute ischemic stroke (diffusion-weighted imaging [DWI]). Any suggestion of trauma from the history or examination should alert the examiner to the possibility of cervical spine injury and prompt an imaging evaluation using CT or MRI. Neurovascular imaging using CT or MRI angiography or venography is increasingly available and may suggest arterial occlusion or cerebral venous thrombosis.

Acute brainstem ischemia due to basilar artery thrombosis may cause brief episodes of spontaneous extensor posturing superficially resembling generalized seizures. Coma of sudden onset, accompanied by these movements and cranial nerve abnormalities, necessitates emergency imaging. A noncontrast CT scan of the brain may reveal a hyperdense basilar artery indicating thrombus in the vessel, and subsequent CT or MR angiography can assess basilar artery patency.

Other diagnostic studies are best used in specific circumstances, usually when neuroimaging studies fail to reveal a structural lesion and the etiology of the altered mental state remains uncertain. Electroencephalography (EEG) can be important in the evaluation of critically ill patients with severe brain dysfunction. The EEG of metabolic encephalopathy typically reveals generalized slowing. One of the most important uses of EEG is to help exclude inapparent seizures, especially nonconvulsive status epilepticus. Untreated continuous or frequently recurrent seizures may cause neuronal injury, making the diagnosis and treatment of seizures crucial in this patient group. Lumbar puncture (LP) may be necessary to exclude infectious or inflammatory processes, and an elevated opening pressure may be an important clue to cerebral venous sinus thrombosis. In patients with coma or profound encephalopathy, it is preferable to perform a neuroimaging study prior to LP. If bacterial meningitis is suspected, an LP may be performed urgently, but most often, it is prudent to administer antibiotics empirically before the diagnostic studies are completed. Standard laboratory evaluation of critically ill patients should include assessment of serum electrolytes (especially sodium and calcium), glucose, renal and hepatic function, complete blood count, and coagulation. Serum or urine toxicology screens should be performed in patients with encephalopathy of unknown cause. EEG and LP are most useful when the mechanism of the altered level of consciousness is uncertain; they are not routinely performed for diagnosis in clear-cut cases of stroke or traumatic brain injury.

Monitoring of ICP can be an important tool in selected patients. In general, patients who should be considered for ICP monitoring are those with primary neurologic disorders, such as stroke or traumatic brain injury, who are at significant risk for secondary brain injury due to elevated ICP and decreased CPP. Included are patients with the following: severe traumatic brain injury (Glasgow Coma Scale [GCS] score ≤8 [see **Table 454-1**]); large tissue shifts from supratentorial ischemic or hemorrhagic stroke; or hydrocephalus from subarachnoid hemorrhage (SAH), intraventricular hemorrhage, or posterior fossa stroke. An additional disorder in which ICP monitoring can add important information is fulminant hepatic failure, in which elevated ICP may be treated with barbiturates or, eventually, liver transplantation. In general, ventriculostomy is preferable to ICP monitoring devices that are placed in the brain parenchyma, because ventriculostomy allows CSF drainage as a method of treating elevated ICP. However, parenchymal ICP monitoring is most appropriate for patients with diffuse edema and small ventricles (which may make ventriculostomy placement more difficult) or any degree of coagulopathy (in which ventriculostomy carries a higher risk of hemorrhagic complications) **(Fig. 318-2)**.

TREATMENT OF ELEVATED ICP

Elevated ICP may occur in a wide range of disorders, including head trauma, intracerebral hemorrhage, SAH with hydrocephalus, and fulminant hepatic failure. Because CSF and blood volume can be redistributed initially, by the time elevated ICP occurs, intracranial compliance is severely impaired. At this point, any small increase in the volume of CSF, intravascular blood, edema, or a mass lesion may result in a significant increase in ICP and a decrease in cerebral perfusion. This is a fundamental mechanism of secondary ischemic brain injury and constitutes an emergency that requires immediate attention. In general, ICP should be maintained at <20 mmHg and CPP should be maintained at ≥60 mmHg.

Interventions to lower ICP are ideally based on the underlying mechanism responsible for the elevated ICP **(Table 318-2)**. For example, in hydrocephalus from SAH, the principal cause of elevated ICP is impairment of CSF drainage. In this setting, ventricular drainage of CSF is likely to be sufficient and most appropriate. In head trauma and stroke, cytotoxic edema may be most responsible, and the use of osmotic agents such as mannitol or hypertonic saline becomes an appropriate early step. As described above, elevated ICP may cause tissue ischemia, and, if cerebral autoregulation is intact, the resulting vasodilation can lead to a cycle of worsening ischemia. Paradoxically, administration of vasopressor agents to increase MAP may actually lower ICP by improving perfusion, thereby allowing autoregulatory vasoconstriction as ischemia is relieved and ultimately decreasing intracranial blood volume.

TABLE 318-2 Stepwise Approach to Treatment of Elevated Intracranial Pressure (ICP)[a]

Insert ICP monitor—ventriculostomy versus parenchymal device

General goals: maintain ICP <20 mmHg and CPP ≥60 mmHg. For ICP >20–25 mmHg for >5 min:

1. Elevate head of the bed; midline head position
2. Drain CSF via ventriculostomy (if in place)
3. Osmotherapy—mannitol 25–100 g q4h as needed (maintain serum osmolality <320 mosmol) or hypertonic saline (30 mL, 23.4% NaCl bolus)
4. Glucocorticoids—dexamethasone 4 mg q6h for vasogenic edema from tumor, abscess (avoid glucocorticoids in head trauma, ischemic and hemorrhagic stroke)
5. Sedation (e.g., morphine, propofol, or midazolam); add neuromuscular paralysis if necessary (patient will require endotracheal intubation and mechanical ventilation at this point, if not before)
6. Hyperventilation—to PaCO_2 30–35 mmHg (short-term use or skip this step)
7. Pressor therapy—phenylephrine, dopamine, or norepinephrine to maintain adequate MAP to ensure CPP ≥60 mmHg (maintain euvolemia to minimize deleterious systemic effects of pressors). May adjust target CPP in individual patients based on autoregulation status.
8. Consider second-tier therapies for refractory elevated ICP
    a. Decompressive craniectomy
    b. High-dose barbiturate therapy ("pentobarb coma")
    c. Hypothermia to 33°C

[a]Throughout ICP treatment algorithm, consider repeat head computed tomography to identify mass lesions amenable to surgical evacuation. May alter order of steps based on directed treatment to specific cause of elevated ICP.

Abbreviations: CPP, cerebral perfusion pressure; CSF, cerebrospinal fluid; MAP, mean arterial pressure; PaCO_2, arterial partial pressure of carbon dioxide.

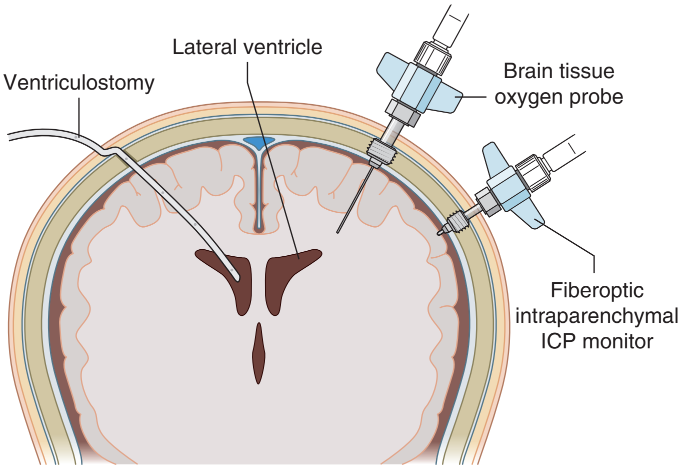
Lateral ventricle
Ventriculostomy
Brain tissue oxygen probe
Fiberoptic intraparenchymal ICP monitor

FIGURE 318-2 Intracranial pressure (ICP) and brain tissue oxygen monitoring. A ventriculostomy allows for drainage of cerebrospinal fluid to treat elevated ICP. Fiberoptic ICP and brain tissue oxygen monitors are usually secured using a screw-like skull bolt. Cerebral blood flow and microdialysis probes (not shown) may be placed in a manner similar to the brain tissue oxygen probe.

Early signs of elevated ICP include drowsiness and a diminished level of consciousness. Neuroimaging studies may reveal evidence of edema and mass effect. Hypotonic IV fluids should be avoided, and elevation of the head of the bed is recommended. Patients must be carefully observed for risk of aspiration and compromise of the airway as the level of alertness declines. Coma and unilateral pupillary changes are late signs and require immediate intervention. Emergent treatment of elevated ICP is most quickly achieved by intubation and hyperventilation, which causes vasoconstriction and reduces cerebral blood volume. To avoid provoking or worsening cerebral ischemia, hyperventilation, if used at all, is best administered only for short periods of time until a more definitive treatment can be instituted. Furthermore, the effects of hyperventilation on ICP are short-lived, often lasting only for several hours because of the buffering capacity of the cerebral interstitium, and rebound elevations of ICP may accompany abrupt discontinuation of hyperventilation. As the level of consciousness declines to coma, the ability to follow the neurologic status of the patient by examination lessens and measurement of ICP assumes greater importance. If a ventriculostomy device is in place, direct drainage of CSF to reduce ICP is possible. Finally, high-dose barbiturates, decompressive hemicraniectomy, and hypothermia are sometimes used for refractory elevations of ICP, although these have significant side effects and only decompressive hemicraniectomy has been shown to improve outcome in select patients.

SECONDARY BRAIN INSULTS

Patients with primary brain injuries, whether due to trauma or stroke, are at risk for ongoing secondary ischemic brain injury.

Because secondary brain injury can be a major determinant of a poor outcome, strategies for minimizing secondary brain insults are an integral part of the critical care of all patients. Although elevated ICP may lead to secondary ischemia, most secondary brain injury is mediated through other clinical events that exacerbate the ischemic cascade already initiated by the primary brain injury. Episodes of secondary brain insults are usually not associated with apparent neurologic worsening. Rather, they lead to cumulative injury limiting eventual recovery, which manifests as a higher mortality rate or worsened long-term functional outcome. Thus, close monitoring of vital signs is important, as is early intervention to prevent secondary ischemia. Avoiding hypotension and hypoxia is critical, as significant hypotensive events (systolic blood pressure <90 mmHg) as short as 10 min in duration have been shown to adversely influence outcome after traumatic brain injury. Even in patients with stroke or head trauma who do not require ICP monitoring, close attention to adequate cerebral perfusion is warranted. Hypoxia (pulse oximetry saturation <90%), particularly in combination with hypotension, also leads to secondary brain injury. Likewise, fever and hyperglycemia both worsen experimental ischemia and have been associated with worsened clinical outcome after stroke and head trauma. Aggressive control of fever with a goal of normothermia is warranted but may be difficult to achieve with antipyretic medications and cooling blankets. The value of newer surface or intravascular temperature control devices for the management of refractory fever is under investigation. The use of IV insulin infusion is encouraged for control of hyperglycemia because this allows better regulation of serum glucose levels than SC insulin. A reasonable goal is to maintain the serum glucose level at <10.0 mmol/L (<180 mg/dL), although episodes of hypoglycemia appear equally detrimental and the optimal targets remain uncertain. New cerebral monitoring tools that allow continuous evaluation of brain tissue oxygen tension, CBF, cortical spreading depolarizations, and cerebral metabolism (via microdialysis) may further improve the management of secondary brain injury.

CRITICAL CARE DISORDERS OF THE CNS

■ HYPOXIC-ISCHEMIC BRAIN INJURY

This occurs from lack of delivery of oxygen to the brain because of extreme hypotension (hypoxia-ischemia) or hypoxia due to respiratory failure. Causes include myocardial infarction, cardiac arrest, shock, asphyxiation, paralysis of respiration, and carbon monoxide or cyanide poisoning. In some circumstances, hypoxia may predominate. Carbon monoxide and cyanide poisoning are sometimes termed *histotoxic hypoxia* because they cause a direct impairment of the respiratory chain.

Clinical Manifestations Mild degrees of pure hypoxia, such as occur at high altitudes, cause impaired judgment, inattentiveness, motor incoordination, and, at times, euphoria. However, with hypoxia-ischemia, such as occurs with circulatory arrest, consciousness is lost within seconds. If circulation is restored within 3–5 min, full recovery may occur, but if hypoxia-ischemia lasts beyond 3–5 min, some degree of permanent cerebral damage often results. Except in extreme cases, it may be difficult to judge the precise degree of hypoxia-ischemia, and some patients make a relatively full recovery after even 10 min or more of global cerebral ischemia. The brain is more tolerant to pure hypoxia than it is to hypoxia-ischemia. For example, a Pao$_2$ as low as 20 mmHg (2.7 kPa) can be well tolerated if it develops gradually and normal blood pressure is maintained, whereas short durations of very low or absent cerebral circulation may result in permanent impairment.

Clinical examination at different time points after a hypoxic-ischemic insult (especially cardiac arrest) is useful in assessing prognosis for long-term neurologic outcome. The prognosis is better for patients with intact brainstem function, as indicated by normal pupillary light responses and intact oculocephalic (doll's eyes), oculovestibular (caloric), and corneal reflexes. Absence of these reflexes and the presence of persistently dilated pupils that do not react to light are concerning prognostic signs. A lower likelihood of a favorable outcome from hypoxic-ischemic brain injury is suggested by an absent pupillary light reflex or extensor or absent motor response to pain 5–7 days following the injury, excluding patients with metabolic disturbances and those treated with high-dose sedative medications or hypothermia, which confound interpretation of these signs. Electrophysiologically, the bilateral absence of the N20 component of the somatosensory evoked potential (SSEP) after several days also conveys a poor prognosis. Also, the presence of a burst-suppression pattern of myoclonic status epilepticus on EEG (Fig. 318-3) or a nonreactive EEG is associated with a lower likelihood of good functional outcome. A very elevated serum level (>60 μg/L) of the biochemical marker neuron-specific enolase (NSE) obtained during first 1–3 days following injury is indicative of brain damage after resuscitation from cardiac arrest and is associated with worse outcome. Current approaches to prognostication after cardiac arrest encourage the use of a multimodal approach that includes these diagnostic tests, along with CT or MRI neuroimaging, in conjunction with clinical neurologic assessment. The administration of mild hypothermia after cardiac arrest (see "Treatment") may affect the time points when these clinical and electrophysiologic predictors become reliable in identifying patients with a very low likelihood of clinically meaningful recovery. For example, the false-positive rate for incorrect prediction of poor neurologic outcome may be as high as 21% (95% confidence interval [CI] 8–43%) for patients treated with mild hypothermia who exhibit 3-day motor function no better than extensor posturing. Thus, sufficient time from injury is important to ensure accuracy of prognostic assessment and to avoid the self-fulfilling prophecy of poor outcome when withdrawal of life-sustaining therapy is undertaken in a patient who has the potential for recovery. The minimum observation period to ensure accuracy of prognostication remains unclarified, but some patients may awaken after a week or longer. Long-term consequences of hypoxic-ischemic encephalopathy include persistent coma or an unresponsive wakeful state (Chap. 30), dementia (Chap. 31), visual agnosia (Chap. 32), parkinsonism, choreoathetosis, cerebellar ataxia, myoclonus, seizures, and an amnestic state, which may be a consequence of selective damage to the hippocampus.

Pathology Principal histologic findings are extensive multifocal or diffuse laminar cortical injury (Fig. 318-4), with frequent involvement of the deep gray nuclei and hippocampus. The hippocampal CA1 neurons are vulnerable to even brief episodes of hypoxia-ischemia, perhaps explaining why selective persistent memory deficits may occur after brief cardiac arrest. Scattered small areas of infarction or neuronal loss may be present in the basal ganglia, hypothalamus, or brainstem. In some cases, extensive bilateral thalamic scarring may affect pathways that mediate arousal, and this pathology may be responsible for the unresponsive wakeful state (previously known as the vegetative state). A specific form of hypoxic-ischemic brain injury, so-called watershed infarcts, occurs at the distal territories between the major cerebral arteries and can cause cognitive deficits, including visual agnosia, and weakness that is greater in proximal than in distal muscle groups.

Diagnosis Diagnosis is based on the history of a hypoxic-ischemic event such as cardiac arrest. Blood pressure <70 mmHg systolic or Pao$_2$ <40 mmHg is usually necessary, although both absolute levels and duration of exposure are important determinants of cellular injury. Carbon monoxide intoxication can be confirmed by measurement of carboxyhemoglobin and is suggested by a cherry red color of the venous blood and skin, although the latter is an inconsistent clinical finding.

TREATMENT

Hypoxic-Ischemic Brain Injury

Treatment should be directed at restoration of normal cardiorespiratory function. This includes securing a clear airway, ensuring adequate oxygenation and ventilation, and restoring cerebral

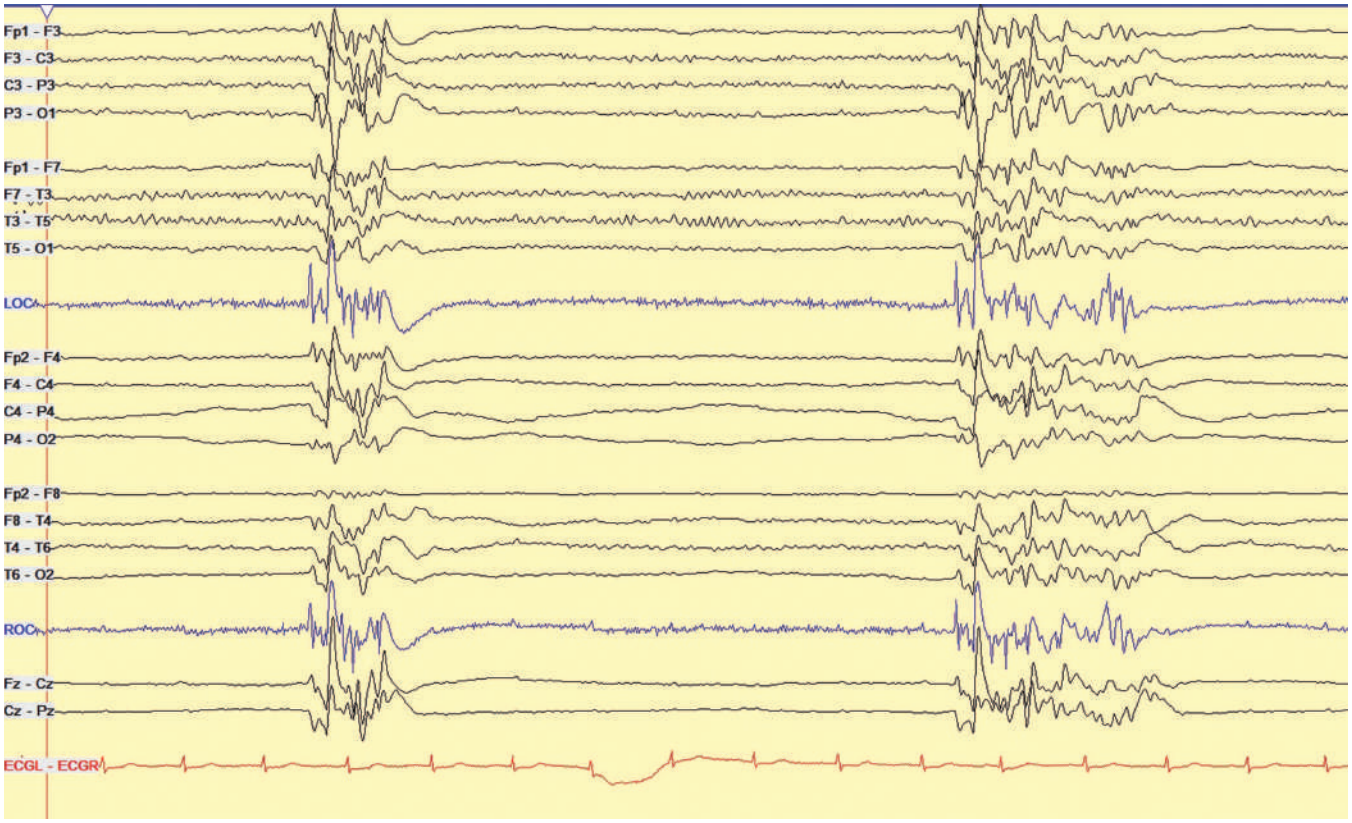

FIGURE 318-3 Electroencephalography (EEG) after cardiac arrest. A burst-suppression pattern is seen in a comatose patient with severe hypoxic-ischemic encephalopathy after cardiac arrest. In this patient, each burst on EEG was associated with a whole-body jerking movement leading to the clinical and electrophysiologic diagnosis of myoclonic status epilepticus.

perfusion, whether by cardiopulmonary resuscitation, fluid, pressors, or cardiac pacing. Hypothermia may target the neuronal cell injury cascade and has substantial neuroprotective properties in experimental models of brain injury. Several clinical trials found that mild hypothermia (33°C) administered for 12–24 h improved functional outcome in patients who remained comatose after resuscitation from cardiac arrest. These trials varied in the patients included, with some involving out-of-hospital arrest with a shockable cardiac rhythm and others including in-hospital arrest and focusing on those with nonshockable rhythms. Two larger subsequent clinical trials that included patients with high rates of bystander cardiopulmonary resuscitation and primary cardiac causes of arrest did not find hypothermia to 33°C as beneficial. In one, targeted temperature management (TTM) to 33 or 36°C resulted in similar outcomes, while in another, early treatment of fever (temperature ≥37.8°C) resulted in similar outcomes to those of patients treated with hypothermia to 33°C. Given these heterogeneous findings from clinical trials, there exists variability in how clinicians treat patients with hypoxic-ischemic brain injury from cardiac arrest. Current guidelines recommend selecting and maintaining a constant temperature between 32 and 37.5°C during postarrest temperature control for patients who have no meaningful response to verbal commands after return of spontaneous circulation (ROSC). It is our current practice to target either 33 or 36°C. Fever should be avoided in all cases. Potential complications of hypothermia include systemic coagulopathy and an increased risk of infection.

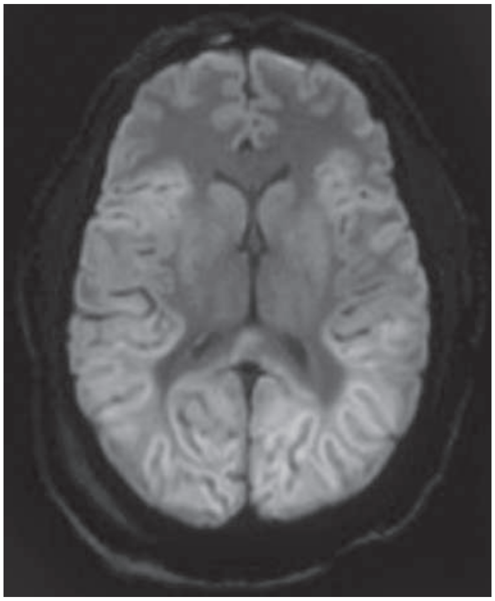

FIGURE 318-4 Hypoxic-ischemic brain injury after cardiac arrest. Diffusion-weighted magnetic resonance imaging shows reduced diffusion (bright signal) throughout the cerebral cortex as well as in the caudate, globus pallidus, and thalamus bilaterally.

Additional clinical trials have been conducted to assess whether targeting specific physiologic parameters that may impact cerebral blood flow and oxygen delivery after ROSC can improve outcome. A factorial design clinical trial tested both a higher mean arterial blood pressure target (77 vs 63 mmHg) and a higher partial pressure of oxygen target (98–105 vs 68–75 mmHg) and found no differences in clinical outcome. Likewise, targeting early mild hypercapnia (Paco_2 50–55 vs 35–45 mmHg) did not improve outcome in a separate clinical trial.

Anticonvulsants may be needed to control seizures, although these are not usually given prophylactically. Myoclonic status epilepticus within 24 h after a primary circulatory arrest generally portends a poor prognosis, even if seizures are controlled. A clinical trial of complete suppression of rhythmic and periodic EEG activity for 48 h did not result in improved patient outcomes compared with standard care that included TTM, and therefore, this strategy should likely not be practiced. Posthypoxic myoclonus may respond to oral administration of clonazepam at doses of 1.5–10 mg daily or valproate at doses of 300–1200 mg daily in divided doses.

Severe acute carbon monoxide intoxication may be treated with hyperbaric oxygen. Carbon monoxide and cyanide intoxication can also cause a delayed encephalopathy. Little clinical impairment is evident when the patient first regains consciousness, but a parkinsonian syndrome characterized by akinesia and rigidity without tremor may develop. Symptoms can worsen over months, accompanied by increasing evidence of damage in the basal ganglia as seen on both CT and MRI.

POSTCARDIAC BYPASS BRAIN INJURY

CNS injuries following open heart or coronary artery bypass grafting (CABG) surgery are common and include acute encephalopathy, stroke, and a chronic syndrome of cognitive impairment. Hypoperfusion and embolic disease are frequently involved in the pathogenesis of these syndromes, although multiple mechanisms may be involved in these critically ill patients who are at risk for various metabolic and polypharmaceutical complications.

The frequency of hypoxic injury secondary to inadequate blood flow intraoperatively has been markedly decreased by modern surgical and anesthetic techniques. Despite these advances, some patients still experience neurologic complications from cerebral hypoperfusion or suffer focal ischemia from carotid or focal intracranial stenoses in the setting of regional hypoperfusion. Postoperative infarcts in the border zones between vascular territories are often attributed to systemic hypotension, although these infarcts can also result from embolic disease.

Embolic disease is likely the predominant mechanism of cerebral injury during cardiac surgery as evidenced by diffusion-weighted MRI and intraoperative transcranial Doppler ultrasound studies. Thrombus in the heart itself as well as atheromas in the aortic arch can become dislodged during cardiac surgeries, releasing a shower of particulate matter into the cerebral circulation. Cross-clamping of the aorta, manipulation of the heart, extracorporeal circulation techniques ("bypass"), arrhythmias such as atrial fibrillation, and introduction of air through suctioning have all been implicated as potential sources of emboli.

This shower of microemboli results in a number of clinical syndromes. Occasionally, a single large embolus leads to an isolated large-vessel stroke that presents with obvious clinical focal deficits. When there is a high burden of very small emboli, an acute encephalopathy can occur postoperatively, presenting as either a hyperactive or hypoactive confusional state, the latter of which is frequently and incorrectly ascribed to depression or a sedative-induced delirium. When the burden of microemboli is lower, no acute syndrome is recognized, but the patient may suffer a chronic cognitive deficit.

METABOLIC ENCEPHALOPATHIES

Altered mental states, variously described as confusion, delirium, disorientation, and encephalopathy, are present in many patients with severe illness in an intensive care unit (ICU). Older patients are particularly vulnerable to delirium **(Chap. 29)**, a confusional state characterized by disordered perception, frequent hallucinations, delusions, and sleep disturbance. This is often attributed to medication effects, sleep deprivation, pain, and anxiety. The presence of delirium is associated with a worse outcome in critically ill patients, even in those without an identifiable CNS pathology such as stroke or brain trauma. In these patients, the cause of delirium is often multifactorial, resulting from organ dysfunction, sepsis, and especially the use of medications given to treat pain, agitation, or anxiety. Critically ill patients are often treated with a variety of sedative and analgesic medications, including opiates, benzodiazepines, neuroleptics, and sedative-anesthetic medications, such as propofol. In critically ill patients requiring sedation, use of the centrally acting α_2 agonist dexmedetomidine may reduce delirium and shorten the duration of mechanical ventilation compared to the use of benzodiazepines such as lorazepam or midazolam. The presence of family members in the ICU may also help to calm and orient agitated patients, and in severe cases, low doses of neuroleptics (e.g., haloperidol 0.5–1 mg) can be useful. Current strategies focus on limiting the use of sedative medications when this can be done safely.

In the ICU setting, several metabolic causes of an altered level of consciousness predominate. Hypercarbic encephalopathy can present with headache, confusion, stupor, or coma. Hypoventilation syndrome occurs most frequently in patients with a history of chronic CO_2 retention who are receiving oxygen therapy for emphysema or chronic pulmonary disease **(Chap. 307)**. The elevated $Pa{CO_2}$ leading to CO_2 narcosis may have a direct anesthetic effect, and cerebral vasodilation from increased $Pa{CO_2}$ can lead to increased ICP. Hepatic encephalopathy is suggested by asterixis and can occur in chronic liver failure or acute fulminant hepatic failure. Both hyperglycemia and hypoglycemia can cause encephalopathy, as can hypernatremia and hyponatremia. Confusion, impairment of eye movements, and gait ataxia are the hallmarks of acute Wernicke's disease (see below).

SEPSIS-ASSOCIATED ENCEPHALOPATHY

Pathogenesis In patients with sepsis, the systemic response to infectious agents leads to the release of circulating inflammatory mediators that appear to contribute to encephalopathy. Critical illness, in association with the systemic inflammatory response syndrome (SIRS), can lead to multisystem organ failure. This syndrome can occur in the setting of apparent sepsis, severe burns, or trauma, even without clear identification of an infectious agent. Many patients with critical illness, sepsis, or SIRS develop encephalopathy without obvious explanation. This condition is broadly termed *sepsis-associated encephalopathy*. Although the specific mediators leading to neurologic dysfunction remain uncertain, it is clear that the encephalopathy is not simply the result of metabolic derangements of multiorgan failure. The cytokines tumor necrosis factor, interleukin (IL) 1, IL-2, and IL-6 are thought to play a role in this syndrome.

Diagnosis Sepsis-associated encephalopathy presents clinically as a diffuse dysfunction of the brain without prominent focal findings. Confusion, disorientation, agitation, and fluctuations in level of alertness are typical. In more profound cases, especially with hemodynamic compromise, the decrease in level of alertness can be more prominent, at times resulting in coma. Hyperreflexia and frontal release signs such as a grasp or snout reflex **(Chap. 32)** can be seen. Abnormal movements such as myoclonus, tremor, or asterixis can occur. Sepsis-associated encephalopathy is quite common, occurring in the majority of patients with sepsis and multisystem organ failure. Diagnosis is often difficult because of the multiple potential causes of neurologic dysfunction in critically ill patients and requires exclusion of structural, metabolic, toxic, and infectious (e.g., meningitis or encephalitis) causes. The mortality rate of patients with sepsis-associated encephalopathy severe enough to produce coma approaches 50%, although this principally reflects the severity of the underlying critical illness and is generally not a direct result of the encephalopathy. Patients dying from severe sepsis or septic shock may have elevated levels of the serum brain injury biomarker S-100β and neuropathologic findings of neuronal apoptosis and cerebral ischemic injury. Successful treatment of the underlying critical illness almost always results in substantial improvement of the encephalopathy. However, although severe disability to the level of chronic unresponsive wakeful or minimally conscious states is uncommon, long-term cognitive dysfunction clinically similar to dementia is being increasingly recognized in some survivors, especially in older patients.

OSMOTIC DEMYELINATION SYNDROME (CENTRAL PONTINE MYELINOLYSIS)

This disorder often presents in a devastating fashion as quadriplegia and pseudobulbar palsy, although less severe presentations may occur. Predisposing factors include severe underlying medical illness or nutritional deficiency; most cases are associated with rapid correction of hyponatremia or with hyperosmolar states, and clinical symptoms are usually identified a few days after sodium correction. Previously termed *central pontine myelinolysis*, the more accurate term *osmotic demyelination syndrome* is now preferred. The pathology consists of

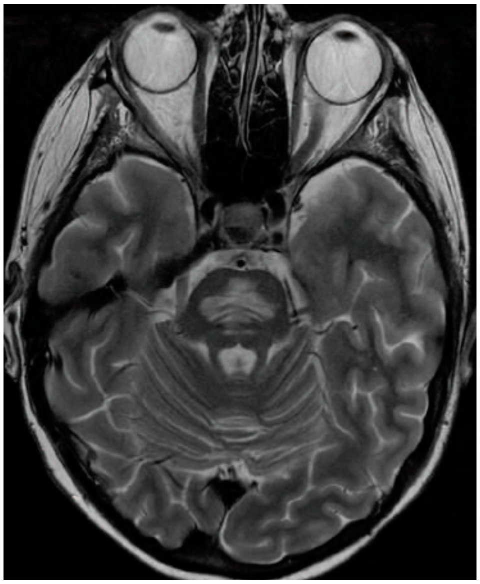

FIGURE 318-5 Osmotic demyelination syndrome. Axial T2-weighted magnetic resonance scan through the pons reveals a symmetric area of abnormal pontine high signal intensity that characteristically involves transverse pontine fibers and spares the descending corticospinal tracts. *(Image courtesy of Dr. Jared Narvid, Department of Radiology & Biomedical Imaging, University of California, San Francisco.)*

demyelination without inflammation in the base of the pons, with relative sparing of axons and nerve cells. MRI is useful in establishing the diagnosis **(Fig. 318-5)** and may also identify partial forms that present as confusion, dysarthria, and/or disturbances of conjugate gaze without quadriplegia. Occasional cases present with lesions outside of the brainstem. Therapy for the restoration of severe hyponatremia should aim for gradual correction, i.e., by ≤8 mmol/L (8 meq/L) within 24 h and 15 mmol/L (15 meq/L) within 48 h.

WERNICKE'S DISEASE

Wernicke's disease is a common and preventable disorder due to a deficiency of thiamine **(Chap. 344)**. In the United States, alcoholics account for most cases, but patients with malnutrition due to hyperemesis, starvation, renal dialysis, cancer, HIV/AIDS, or rarely gastric surgery are also at risk. The characteristic clinical triad is ophthalmoplegia, ataxia, and global confusion. However, only one-third of patients with acute Wernicke's disease present with the classic clinical triad. Most patients are profoundly disoriented, indifferent, and inattentive, although rarely they have an agitated delirium related to ethanol withdrawal. If the disease is not treated, stupor, coma, and death may ensue. Ocular motor abnormalities include horizontal nystagmus on lateral gaze, lateral rectus palsy (usually bilateral), conjugate gaze palsies, and rarely ptosis. Gait ataxia probably results from a combination of polyneuropathy, cerebellar involvement, and vestibular paresis. The pupils are usually spared, but they may become miotic with advanced disease.

Wernicke's disease is usually associated with other manifestations of nutritional disease, such as polyneuropathy. Rarely, amblyopia or myelopathy occurs. Tachycardia and postural hypotension may be related to impaired function of the autonomic nervous system or to the coexistence of cardiovascular beriberi. Patients who recover show improvement in ocular palsies within hours after the administration of thiamine, but horizontal nystagmus may persist. Ataxia improves more slowly than the ocular motor abnormalities. Approximately half recover incompletely and are left with a slow, shuffling, wide-based gait and an inability to tandem walk. Apathy, drowsiness, and confusion improve more gradually. As these symptoms recede, an amnestic state with impairment in recent memory and learning may become more apparent (*Korsakoff's psychosis*). Korsakoff's psychosis is frequently persistent; the residual mental state is characterized by gaps in memory, confabulation, and disordered temporal sequencing.

Pathology Periventricular lesions surround the third ventricle, aqueduct, and fourth ventricle, with petechial hemorrhages in occasional acute cases and atrophy of the mammillary bodies in most chronic cases. There is frequently endothelial proliferation, demyelination, and some neuronal loss. These changes may be detected by MRI **(Fig. 318-6)**. The amnestic defect is related to lesions in the dorsal medial nuclei of the thalamus.

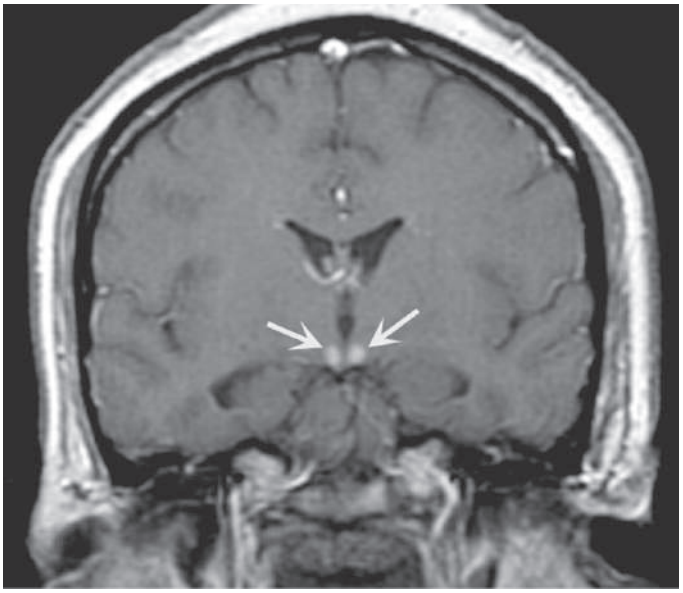

FIGURE 318-6 Wernicke's disease. Coronal T1-weighted postcontrast magnetic resonance imaging reveals abnormal enhancement of the mammillary bodies (*arrows*), typical of acute Wernicke's encephalopathy.

Pathogenesis Thiamine is a cofactor of several enzymes, including transketolase, pyruvate dehydrogenase, and α-ketoglutarate dehydrogenase. Thiamine deficiency produces a diffuse decrease in cerebral glucose utilization and results in mitochondrial damage. Glutamate accumulates due to impairment of α-ketoglutarate dehydrogenase activity and, in combination with the energy deficiency, may result in excitotoxic cell damage.

TREATMENT

Wernicke's Disease

Wernicke's disease is a medical emergency and requires immediate administration of high-dose thiamine, in a dose of 500 mg IV. The dose should be begun prior to treatment with IV glucose solutions and continued three times daily for 2–3 days. Thiamine may then be given in a dose of 250 mg IV or IM daily for 5 more days (in conjunction with other B vitamins), with oral thiamine then continued at 100 mg daily until the patient is no longer considered at risk. Glucose infusions may precipitate Wernicke's disease in a previously unaffected patient or cause a rapid worsening of an early form of the disease. For this reason, thiamine should be administered to all alcoholic patients requiring parenteral glucose.

HYPERPERFUSION DISORDERS (POSTERIOR REVERSIBLE ENCEPHALOPATHY SYNDROME)

Several seemingly diverse syndromes including hypertensive encephalopathy, eclampsia, postcarotid endarterectomy syndrome, and toxicity from calcineurin inhibitor and other medications share the common pathogenesis of hyperperfusion likely due to endothelial dysfunction. Vasogenic edema is typically the primary process leading to neurologic dysfunction, and this is thought to result from one of two mechanisms: exceeding the cerebral autoregulatory threshold leading to increased CBF and capillary leakage into the interstitium, or direct impairment of the BBB itself. The predilection of all of the hyperperfusion disorders to affect the posterior rather than anterior portions of the brain may be due to a lower threshold for autoregulatory breakthrough in the posterior circulation or a vasculopathy that is more common in these blood vessels.

TABLE 318-3 Common Etiologies of Posterior Reversible Encephalopathy Syndrome

Disorders in which increased capillary pressure dominates the pathophysiology
Hypertensive encephalopathy, including secondary causes such as renovascular hypertension, pheochromocytoma, cocaine use, etc.
Postcarotid endarterectomy syndrome
Preeclampsia/eclampsia
Disorders in which endothelial dysfunction dominates the pathophysiology
Calcineurin inhibitor toxicity (e.g., cyclosporine, tacrolimus)
Chemotherapeutic agent toxicity (e.g., cytarabine, azathioprine, 5-fluorouracil, cisplatin, methotrexate, tumor necrosis factor α antagonists)
HELLP syndrome (hemolysis, elevated liver enzyme levels, low platelet count)
Hemolytic-uremic syndrome (HUS)

These disorders of hyperperfusion can be divided into those caused primarily by increased pressure and those due to endothelial dysfunction from a toxic or autoimmune etiology **(Table 318-3)**. In reality, both of these processes likely play some role in each of these disorders. The clinical presentation of all of the hyperperfusion syndromes is similar with prominent headaches, seizures, or focal neurologic deficits. Headaches have no specific characteristics, range from mild to severe, and may be accompanied by alterations in consciousness ranging from confusion to coma. Seizures may be present, and these can be of multiple types depending on the severity and location of the edema. Nonconvulsive seizures have been described; therefore, a low threshold for obtaining an electroencephalogram (EEG) should be maintained. The typical focal deficit in hyperperfusion states is cortical visual loss, given the tendency of the process to involve the occipital lobes. However, any focal deficit can occur depending on the area affected, as evidenced by patients who, after carotid endarterectomy, exhibit neurologic dysfunction referable to the ipsilateral newly reperfused hemisphere. It appears as if the rapidity of rise, rather than the absolute value of pressure, is the most important risk factor.

MRI classically exhibits the high T2 signal of edema primarily in the posterior occipital lobes, not respecting any single vascular territory **(Fig. 318-7)**. CT is less sensitive but may show a pattern of patchy hypodensity in the involved territory. The term *posterior reversible encephalopathy syndrome* (PRES) is often used to describe these conditions; however, the clinical syndrome is not always reversible or limited just to the posterior brain regions. Vessel imaging may demonstrate narrowing of the cerebral vasculature, especially in the posterior circulation; whether this noninflammatory vasculopathy is a primary cause of the edema or occurs as a secondary phenomenon remains unclear. Other ancillary studies such as CSF analysis often yield nonspecific results. Many of the substances that have been implicated, such as cyclosporine, can cause this syndrome even at low doses or after years of treatment. Therefore, normal serum levels of these medications do not exclude them as inciting agents.

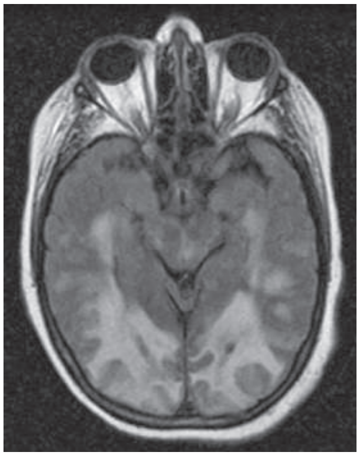

FIGURE 318-7 Axial fluid-attenuated inversion recovery (FLAIR) magnetic resonance imaging (MRI) of the brain in a patient taking cyclosporine after liver transplantation, who presented with seizures, headache, and cortical blindness. Increased signal is seen bilaterally in the occipital lobes predominantly involving the white matter, consistent with a hyperperfusion state secondary to calcineurin inhibitor exposure.

Treatment involves judicious lowering of the blood pressure with IV agents such as labetalol or nicardipine, removal of the offending medication, and treatment of an underlying medical condition such as eclampsia. If the blood pressure is very elevated, it is reasonable to lower the MAP by ~20% initially, as further lowering of the pressure may cause secondary ischemia and possibly infarction as pressure drops below the lower range of the patient's autoregulatory capability. Seizures must be identified and controlled, often necessitating continuous EEG monitoring. Anticonvulsants are effective when seizure activity is identified, but in the special case of eclampsia, there is evidence to support the use of magnesium sulfate for seizure control.

POST–SOLID ORGAN TRANSPLANT BRAIN INJURY

Immunosuppressive medications are administered in high doses to patients after solid organ transplant, and many of these compounds have well-described neurologic complications. In patients with headache, seizures, or focal neurologic deficits taking calcineurin inhibitors, the diagnosis of hyperperfusion syndrome should be considered, as discussed above. This neurotoxicity occurs mainly with cyclosporine and tacrolimus and can present even in the setting of normal serum drug levels. Treatment primarily involves lowering the drug dosage or discontinuing the drug. Sirolimus has very few recorded cases of neurotoxicity and may be a reasonable alternative for some patients. Another example of an immunosuppressive medication with neurologic complications is the leukoencephalopathy seen with methotrexate, especially when it is administered intrathecally or with concurrent radiotherapy. In any solid organ transplant patient with neurologic complaints, a careful examination of the medication list is required to search for these possible drug effects.

Cerebrovascular complications of solid organ transplant are often first recognized in the immediate postoperative period. Border zone territory infarctions can occur, especially in the setting of systemic hypotension during cardiac transplant surgery. Embolic infarctions classically complicate cardiac transplantation, but all solid organ transplant procedures place patients at risk for systemic emboli. When cerebral embolization accompanies renal or liver transplantation surgery, a careful search for right-to-left shunting should include evaluation of the heart with agitated saline echocardiography (i.e., "bubble study"), as well as looking for intrapulmonary shunting. Renal and some cardiac transplant patients often have advanced atherosclerosis, providing a risk for stroke. Imaging with CT or MRI should be done when cerebrovascular complications are suspected to confirm the diagnosis and to exclude intracerebral hemorrhage, which most often occurs in the setting of coagulopathy secondary to liver failure or after cardiac bypass procedures.

Given that patients with solid organ transplants are chronically immunosuppressed, infections are a common concern **(Chap. 148)**. In any transplant patient with new CNS signs or symptoms such as seizure, confusion, or focal deficit, the diagnosis of a CNS infection should be considered and evaluated through imaging (usually MRI) and possibly LP. The most common pathogens responsible for CNS infections in these patients vary based on time since transplant. In the first month posttransplant, common pathogens include the usual bacterial organisms associated with surgical procedures and indwelling catheters. Starting in the second month posttransplant, opportunistic infections of the CNS become more common, including *Nocardia* and *Toxoplasma* species as well as fungal infections such as aspergillosis. Viral infections that can affect the brain of the immunosuppressed patient, such as herpes simplex virus, cytomegalovirus, human herpesvirus type 6 (HHV-6), and varicella, also become more common after the first month posttransplant. Beyond 6 months, immunosuppressed posttransplant patients still remain at risk for these opportunistic bacterial, fungal, and viral infections but can also suffer late CNS infectious complications such as progressive multifocal leukoencephalopathy

(PML) associated with JC virus **(Chap. 142)** and Epstein-Barr virus–driven clonal expansions of B cells resulting in posttransplant lymphoproliferative disorder or CNS lymphoma **(Chap. 95)**.

CNS COMPLICATIONS OF CHECKPOINT INHIBITOR AND CHIMERIC ANTIGEN RECEPTOR T-CELL THERAPY

Cancer immunotherapy is now a widely used treatment for both solid tumors and hematologic malignances. Two types of this immunotherapy, checkpoint inhibitors and chimeric antigen receptor (CAR) T-cell (CAR-T) therapy, can carry significant neurologic toxicity that may manifest as encephalopathy, cerebral edema, or white matter demyelination. These complications may be severe and require neurocritical care evaluation and intervention.

Immune checkpoint inhibitors are monoclonal antibodies that bind to normally occurring checkpoint proteins such as PD-1, PD-L1, and CTLA-4, thereby freeing T cells to attack cancerous cells. Currently available checkpoint inhibitors include pembrolizumab, nivolumab, cemiplimab, atezolizumab, avelumab, durvalumab, and ipilimumab. Common side effects are diarrhea, rash, and pneumonitis. Neurologic side effects occur in ~5% of patients treated with monotherapy and ~10% undergoing combination therapy, presumably as a result of shared antigens between tumor cells and self, leading to an autoimmune process **(Chap. 99)**. CNS adverse events include limbic encephalitis, cerebellitis, and myelitis. A clinical syndrome of encephalopathy, memory disturbances, and seizures may occur. Peripheral nervous system complications such as myasthenia gravis, myositis, and neuropathy have also been described and may be even more common than CNS manifestations.

Patients who develop CNS neurologic symptoms while on checkpoint inhibitor treatment should undergo MRI studies of the brain or spinal cord, based on clinical symptoms. Mesial temporal lobe hyperintensities and a lymphocytic CSF pleocytosis may be present. EEG may be appropriate to evaluate for subclinical seizures. Various autoantibodies such as anti-Ma2, anti-GFAP, anti-Hu, and anti-CASPR2 have been described but are not required for diagnosis. Treatment consists of discontinuing the checkpoint inhibitor and administering high-dose glucocorticoids. Intravenous immunoglobulins and plasmapheresis have been used in severe cases. For mild cases, restarting the checkpoint inhibitor may be considered; however, relapse with fatal necrotizing encephalitis has been described. Given that checkpoint inhibitor–treated patients are immunocompromised, before checkpoint inhibitor–related neurotoxicity is diagnosed, it is imperative to rule out an alternative diagnosis such as cerebral metastases, infection, or stroke.

CAR-T therapy for leukemia or lymphoma involves removing a patient's T cells and genetically engineering them using a disarmed virus to produce surface chimeric antigen receptors that, when given back to the patient, recognize antigens on tumor cells. CAR-T therapy is frequently associated with significant side effects, which usually occur as either cytokine release syndrome (CRS) or neurotoxicity. These two types of CAR-T side effects are distinct but often occur in the same patient, and both occur within days of initiation of CAR-T treatment. CRS is a clinical syndrome of hypotension, fever, and hypoxia, which may have associated multiorgan dysfunction. CRS occurs in 80–100% of CAR-T–treated patients and is due to widespread release of proinflammatory cytokines. Treatment is with the IL-6 receptor pathway blocker tocilizumab, which can alleviate CRS symptoms without impairing the antitumor efficacy of the CAR-T cells; glucocorticoids may also be administered.

CAR-T neurotoxicity is less common but still occurs in more than half of treated patients. Clinical manifestations may include headache, encephalopathy, aphasia, seizures, tremors, and life-threatening cerebral edema. Predictors of occurrence of neurotoxicity include earlier and more severe CRS, fever, elevated C-reactive protein and serum ferritin, and older patient age. Treatment of CAR-T neurotoxicity also involves administration of tocilizumab (as most of these patients also have CRS) and glucocorticoids. In addition to these treatments, patients with CAR-T neurotoxicity should also undergo brain imaging and EEG if indicated based on symptoms, with concurrent treatment of cerebral edema and seizures if present.

CRITICAL CARE DISORDERS OF THE PERIPHERAL NERVOUS SYSTEM

Critical illness with disorders of the peripheral nervous system (PNS) arises in two contexts: (1) primary neurologic diseases that require critical care interventions such as intubation and mechanical ventilation, and (2) secondary PNS manifestations of systemic critical illness, often involving multisystem organ failure. The former include acute polyneuropathies such as Guillain-Barré syndrome **(Chap. 458)**, neuromuscular junction disorders including myasthenia gravis **(Chap. 459)** and botulism **(Chap. 158)**, and primary muscle disorders such as polymyositis **(Chap. 377)**. The latter result either from the systemic disease itself or as a consequence of interventions and as a group are often referred to as ICU-acquired weakness (ICUAW).

General principles of respiratory evaluation in patients with PNS involvement, regardless of cause, include assessment of pulmonary mechanics, such as maximal inspiratory force (MIF) and vital capacity (VC), and evaluation of strength of bulbar muscles. Regardless of the cause of weakness, endotracheal intubation should be considered when the MIF falls to below –25 cmH_2O or the VC is <1 L. Also, patients with severe palatal weakness may require endotracheal intubation in order to prevent acute upper airway obstruction or recurrent aspiration. Arterial blood gases and oxygen saturation from pulse oximetry are used to follow patients with potential respiratory compromise from PNS dysfunction. However, intubation and mechanical ventilation should be undertaken based on clinical assessment rather than waiting until oxygen saturation drops or CO_2 retention develops from hypoventilation. Noninvasive mechanical ventilation may be considered initially in lieu of endotracheal intubation in myasthenia gravis but is generally insufficient in patients with severe bulbar weakness or ventilatory failure with hypercarbia. **Principles of mechanical ventilation are discussed in Chap. 313.**

■ NEUROPATHY

Although encephalopathy may be the most obvious neurologic dysfunction in critically ill patients, dysfunction of the PNS is also quite common. It is typically present in patients with prolonged critical illnesses lasting several weeks and involving sepsis; clinical suspicion is aroused when there is failure to wean from mechanical ventilation despite improvement of the underlying sepsis and critical illness. *Critical illness polyneuropathy* refers to the most common PNS complication related to critical illness; it is seen in the setting of prolonged critical illness, sepsis, and multisystem organ failure. Neurologic findings include diffuse weakness, decreased reflexes, and distal sensory loss. Electrophysiologic studies demonstrate a diffuse, symmetric, distal axonal sensorimotor neuropathy, and pathologic studies have confirmed axonal degeneration. The precise mechanism of critical illness polyneuropathy remains unclear, but circulating factors such as cytokines, which are associated with sepsis and SIRS, are thought to play a role. It has been reported that up to 70% of patients with the sepsis syndrome have some degree of neuropathy, although far fewer have a clinical syndrome profound enough to cause severe respiratory muscle weakness requiring prolonged mechanical ventilation or resulting in failure to wean. Aggressive glycemic control with insulin infusions appears to decrease the risk of critical illness polyneuropathy. Treatment is otherwise supportive, with specific intervention directed at treating the underlying illness. Although spontaneous recovery is usually seen, the time course may extend over weeks to months and necessitate long-term ventilatory support and care even after the underlying critical illness has resolved.

■ DISORDERS OF NEUROMUSCULAR TRANSMISSION

A defect in neuromuscular transmission may be a source of weakness in critically ill patients. Botulism **(Chap. 158)** may be acquired by ingesting botulinum toxin from improperly stored food or may arise from an anaerobic abscess from *Clostridium botulinum* (wound botulism). Infants can present with generalized weakness from gut-derived *Clostridium* infection, especially if they are fed honey. Diplopia and dysphagia are early signs of food-borne botulism. Treatment is mostly supportive, although use of antitoxin early in the course may limit the duration of the neuromuscular blockade. General ICU care is similar to patients with Guillain-Barré syndrome or myasthenia gravis with focused care to avoid ulcer formation at pressure points, deep venous thromboprophylaxis, and infection prevention. Public health officers should be rapidly informed when the diagnosis is made to prevent further exposure to others from the tainted food or source of wound botulism (such as injection drug use).

Undiagnosed myasthenia gravis **(Chap. 459)** may be a consideration in weak ICU patients; however, persistent weakness secondary to impaired neuromuscular junction transmission is almost always due to administration of drugs. A number of medications impair neuromuscular transmission; these include antibiotics, especially aminoglycosides, and beta-blocking agents. In the ICU, the nondepolarizing neuromuscular blocking agents (nd-NMBAs), also known as muscle relaxants, are most commonly responsible. Included in this group of drugs are such agents as pancuronium, vecuronium, rocuronium, and cisatracurium. They are often used to facilitate mechanical ventilation or other critical care procedures, but with prolonged use, persistent neuromuscular blockade may result in weakness even after discontinuation of these agents hours or days earlier. Risk factors for this prolonged action of neuromuscular blocking agents include female sex, metabolic acidosis, and renal failure.

Prolonged neuromuscular blockade does not appear to produce permanent damage to the PNS. Once the offending medications are discontinued, full strength is restored, although this may take days. In general, the lowest dose of neuromuscular blocking agent should be used to achieve the desired result, and when these agents are used in the ICU, a peripheral nerve stimulator should be used to monitor neuromuscular junction function.

MYOPATHY

Critically ill patients, especially those with sepsis, frequently develop muscle weakness and wasting, often in the face of seemingly adequate nutritional support. *Critical illness myopathy* is an overall term that describes several different discrete muscle disorders that may occur in critically ill patients. The assumption has been that a catabolic myopathy may develop as a result of multiple factors, including elevated cortisol and catecholamine release and other circulating factors induced by the SIRS. In this syndrome, known as *cachectic myopathy*, serum creatine kinase levels and electromyography (EMG) are normal. Muscle biopsy shows type II fiber atrophy. Panfascicular muscle fiber necrosis may also occur in the setting of profound sepsis. This less common *acute necrotizing intensive care myopathy* is characterized clinically by weakness progressing to a profound level over just a few days. There may be associated elevations in serum creatine kinase and urine myoglobin. Both EMG and muscle biopsy may be normal initially but eventually show abnormal spontaneous activity and panfascicular necrosis with an accompanying inflammatory reaction. Acute rhabdomyolysis can occur from alcohol ingestion or from compartment syndromes.

A *thick-filament myopathy* may occur in the setting of glucocorticoid and nd-NMBA use. The most frequent scenario in which this is encountered is the asthmatic patient who requires high-dose glucocorticoids and nd-NMBA to facilitate mechanical ventilation. This muscle disorder is not due to prolonged action of nd-NMBAs at the neuromuscular junction but, rather, is an actual myopathy with muscle damage; it has occasionally been described with high-dose glucocorticoid use or sepsis alone. Clinically this syndrome is most often recognized when a patient fails to wean from mechanical ventilation despite resolution of the primary pulmonary process. Pathologically, there may be loss of thick (myosin) filaments. Thick-filament critical illness myopathy has a good prognosis. If patients survive their underlying critical illness, the myopathy invariably improves and most patients return to normal. However, because this syndrome is a result of true muscle damage, not just prolonged blockade at the neuromuscular junction, this process may take weeks or months, and tracheotomy with prolonged ventilatory support may be necessary. Some patients do have residual long-term weakness, with atrophy and fatigue limiting ambulation. At present, it is unclear how to prevent this myopathic complication, except by avoiding use of nd-NMBAs, a strategy not always possible. Monitoring with a peripheral nerve stimulator can help to avoid the overuse of these agents. However, this is more likely to prevent the complication of prolonged neuromuscular junction blockade than it is to prevent this myopathy.

FURTHER READING

Cook AM et al: Guidelines for the acute treatment of cerebral edema in neurocritical care patients. Neurocrit Care 32:647, 2020.

Dhar R: Neurologic complications of transplantation. Handb Clin Neurol 141:545, 2017.

Donnelly J et al: Regulation of the cerebral circulation: Bedside assessment and clinical implications. Crit Care 20:129, 2016.

Nolan JP et al: European Resuscitation Council and European Society of Intensive Care Medicine guidelines 2021: Post-resuscitation care. Intensive Care Med 47:369, 2021.

Perman SM et al: 2023 American Heart Association Focused Update on Adult Advanced Cardiovascular Life Support: An update to the American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation 149:e254, 2024.

Posner JB et al: *Plum and Posner's Diagnosis and Treatment of Stupor and Coma*, 5th ed. New York, Oxford University Press, 2019.

Quillinan N et al: Neuropathophysiology of brain injury. Anesthesiol Clin 34:453, 2016.

Rubin D et al: Clinical predictors of neurotoxicity after chimeric antigen receptor T-cell therapy. JAMA Neurol 77:1, 2020.

Ruijter BJ et al: Treating rhythmic and periodic EEG patterns in comatose survivors of cardiac arrest. N Engl J Med 386:724, 2022.

Toledano M, Fugate JE: Posterior reversible encephalopathy in the intensive care unit. Handb Clin Neurol 141:467, 2017.

Vanhorebeek I et al: ICU-acquired weakness. Intensive Care Med 46:637, 2020.

Disorders of the Kidney and Urinary Tract PART 9

319

Approach to the Patient with Renal Disease or Urinary Tract Disease

Julian L. Seifter

The upper urinary tract consists of the kidneys, their vasculature, the renal parenchyma, and its collecting system. The lower urinary tract is composed of the peristaltic ureters from each kidney; the bladder, which receives and stores urine from each ureter; and the urethra, which, upon bladder contraction, excretes the final urine. The excretory system begins with a plasma ultrafiltrate from renal cortical glomeruli, modified by the renal tubules of the nephron, and ends with the excretion of urinary water and solutes.

Disease at any level of these functional structures may be asymptomatic at patient presentation or may cause nonspecific complaints such as fatigue. There may be signs that are highly specific for a site of dysfunction (such as proteinuria or polyuria), or referable to a structure (but not diagnostic of a particular disease), or highly characteristic of a specific diagnosis (such as polycystic kidney disease). Patients with a disorder of the kidney may present without symptoms but with signs of urinary tract disease (such as painless microhematuria or an incidental mass noted on an abdominal imaging study).

It is especially important to perform renal testing in patients with systemic disease; for example, routine screening for microalbuminuria and an estimated glomerular filtration rate (eGFR) in diabetes mellitus (DM). One should look at renal function when a patient's history includes known risk for renal or urinary tract complications, for example, lithium for psychiatric treatment, occupational use of lead, or, in the case of bladder cancer, exposure to aniline dyes. This chapter will address an approach to patients with and without known history of kidney disease, as well as those who have one of a wide variety of systemic illnesses that involve the kidney or the lower urinary tract.

To begin with the evaluation of renal disorders, note that it is time-consuming and wasteful to do random testing prematurely, which would generate a very long list of potential diagnoses. A better approach is to review the evidence from the history, physical examination, and basic laboratory studies to arrive at a logical basis for targeted tests that would identify a particular disease process (Table 319-1). The findings are influenced by the regions of the kidney that are involved and by factors such as family history, toxic exposures, the patient's birth weight for gestational age, and time course of any findings. As a rule, kidney diseases that begin primarily with glomerular dysfunction have albuminuria as the hallmark, whereas kidney diseases that begin in the tubular structures may present primarily with electrolyte disorders or disorders of dilution and concentration of the urine. In both glomerular and tubular disorders, upon progressing to chronic disease, the distinction is more difficult because glomerular diseases eventually affect the tubular interstitium and tubular diseases progress to glomerular dysfunction and scarring.

The complement cascade (a name suggestive of the torrents of a waterfall) has been implicated in a number of diseases via successive steps of activation and amplification of proinflammatory complement proteins. The complement cascade has three pathways, the classical, the alternative, and the lectin-binding, each of which can cause disease if the pathway is dysregulated. In the alternative pathway, the activating signal for producing complement to augment (i.e., complement) the inflammatory response is the presence of circulating antigen-antibody immune complexes. The sites of inflammation are tissue deposits of these complexes. Though immune complex deposition is not specific to the kidney, as a pathologic finding, its presence has great diagnostic and mechanistic value for many forms of systemic and nonsystemic renal disease, encompassing glomerular, tubular, and vascular structures. In the process of stepwise activation of complement, components are released into the circulation primarily from hepatocytes.

Complement component C3 is an important protein in the classical, alternative, and lectin pathways. In the presence of C3 convertase, C3 yields C3a and C3b, the latter of which has further proinflammatory effects: an amplification loop for C3, production of C4 and C5, activation of antibody production, formation of the membrane attack complex (C5b-9), and augmentation of immune complex clearance. In the alternative complement cascade, the heightened proinflammatory response, as seen in a positive feedback loop, circumvents proinflammatory excess by several inhibitory effects of hereditary factors D, H and I. An unsuppressed proinflammatory cycle could lead to life-threatening endothelial damage, which underlies the damage caused by interaction with the coagulation cascade causing thrombotic microangiopathies (TMAs), including thrombotic thrombocytopenic purpura (TTP). The latter is deficient in AdamTS13, an enzyme that inhibits shortening of Von Willebrand factor (VWF), polymer that in its normal length attracts platelets and thus prevents clotting. Failure to find a low level of AdamTS13 in the blood introduces the possibility of atypical HUS. These disease

TABLE 319-1 Acute Kidney Injury

FEATURES	PRERENAL	RENAL	POSTRENAL
Context	Heart failure, hepatic failure, burns, shock, post-op, dehydration, renovascular disease, vascular, drugs	ATN, bilateral cortical necrosis, AIN, crush injuries, myoglobulinuria, hemoglobinuria, ischemia, sepsis, recent contrast drugs	Bladder outlet obstruction, obstruction of solitary kidney, abdominal mass, bilateral stones, drugs
Urine output	Oliguria (usually <500 mL/d) Anuria	Oliguria Normal Polyuria	Anuria Polyuria Both
Clinical history and exam	Trauma, shock, hypotension, burns, GI, sweat, or renal losses	Rhabdomyolysis, hemolysis, thrombotic microangiopathy, AIN, atheroemboli	Extrinsic ureteral obstruction retroperitoneal disease, bladder outlet or urethral obstruction by prostate or cervical cancer
Urinalysis	Concentrated acid urine, hyaline casts, crystalluria	"Muddy brown" granular casts, hematuria, dysmorphic RBCs, RBC casts in GN, WBC casts in AIN, eosinophils in atheroemboli	Intrarenal obstruction by uric acid or CaPhos in tumor lysis syndrome, blood clots in lower urinary tract bleeding, CaOx after ethylene glycol ingestion
Urine chemistry	Low FENa ≤1%, U/P Cr ≥40, UNa ≤10, U/P Osm ≥1	FENa 1–3%, U/P Cr ≤40, UOsm ~isosmotic	Early looks like prerenal, late looks like renal
Laboratory studies	High BUN/Cr ratio, usually hypercatabolic with increased uric acid; may have hyper- or hyponatremia	May have eosinophilia in allergic AIN, high phosphate, low Ca, high PTH, metabolic acidosis	Hydronephrosis on ultrasound, extrinsic or intrinsic diseases on CT scan, tumors on MRI

Abbreviations: AIN, acute interstitial nephritis; ATN, acute tubular necrosis; BUN, blood urea nitrogen; Ca, calcium; CaPhos, calcium phosphate; Cr, creatinine; CT, computed tomography; FENa, fractional excretion of sodium, GI, gastrointestinal; GN, glomerulonephritis; MRI, magnetic resonance imaging; PTH, parathyroid hormone; RBC, red blood cell; U/P, urine/plasma; WBC, white blood cell.

syndromes consist of the triad of microangiopathic hemolytic anemia (with schistocytes, high serum lactate dehydrogenase, and low haptoglobin), thrombocytopenia, and end-organ dysfunction (skin, gastrointestinal, renal, and central nervous systems). Some hereditary diseases have now been recognized as dysregulation of the complement cascade, with a failure to modulate a complement-mediated inflammatory response. Investigations of immune complex disease such as lupus (systemic lupus erythematosus [SLE]), syphilis, and cryoglobulinemia have shed light on the role of complement in the activity of these disorders. Many inhibitors including C3 and C5 convertase have been investigated because of potential drug therapies to correct dysregulation. Serum complement is usually low in these immune complex diseases.

In the following discussion, the focus is on symptoms and signs that constitute the major syndromes observed in the patient with kidney or lower urinary tract disease. These syndromes are the foundation for the diagnosis of particular diseases. Emphasis will be on new classifications of disease, a better understanding of the role of complement and its dysregulation in nephrologic disorders, use of biomarkers, and the impact of genomic knowledge on the expression, identification, and treatment of disorders.

NEPHRITIC SYNDROMES

Nephritis literally means an inflammatory condition of the kidney. Nephritic inflammation may occur in association with infection, allergic response to medications, systemic autoimmune disorders, or toxic exposures. The kidneys may be one of many organs, or the only organ, involved in the inflammatory disorder. Inflammation with enlargement of the kidneys is often associated with point tenderness over the flank and sometimes exquisite costovertebral angle tenderness, requiring gentle palpation for diagnosis.

Nephritic syndromes can be further subdivided depending on the inflamed structures within the kidney, including glomerular, tubulointerstitial, and vascular; and on the time course of progression of the inflammatory process, which may be acute, subacute, or chronic.

GLOMERULONEPHRITIS

Classification of glomerulonephritis (GN) has been challenging because of overlapping features of infection-related and immunologic-mediated pathology. GN is associated with hypertension, volume expansion, and an abnormal urinalysis. In most cases, the volume expansion manifests as edema and hypertension; in the child, ascites may develop; and in the elderly, restlessness and anxiety may be the first signs of incipient acute pulmonary edema. There may be orthopnea at night or dyspnea on exertion, with or without significant peripheral edema. Acute GN is usually associated with low urine output (oliguria) or, at times, complete anuria, low urine sodium content, and concentrated urine, resulting in the retention of salt and water. It is vital to observe a spun urinary sediment to diagnose active GN.

The disease presentation may be acute or subacute, occurring over days or weeks, as in postinfectious GN, or as a chronic, smoldering process with inexorable progression to chronic kidney failure over several months to years. The pathologic correlate of these presentations—the relative amount of acute inflammatory, proliferative, or necrotizing lesions in renal tissue versus chronic sclerotic and atrophic findings on renal biopsy—reflects the acuity versus chronicity of disease. Therefore, in these progressive destructive conditions, early diagnosis is critical, and treatment for acute reversible disease should be instituted as quickly as possible.

The evaluation of the findings on history, physical examination, and urinalysis in narrowing diagnostic possibilities for a specific etiology of the glomerular syndromes is demonstrated in **Fig. 319-1**. The nephritic urinalysis shows hematuria, proteinuria, cells or clumps of cells, and cellular casts in the spun urinary sediment. The cells in the urine are usually a mixture of red blood cells (hematuria) and inflammatory cells (pyuria), including polymorphonuclear (PMN) leukocytes. Microscopic hematuria is invisible to the naked eye, as opposed to macroscopic hematuria, which is often described as cola or "tea-colored" urine, caused by hemoglobin entering an acid urine, forming hematin. The process of urinary erythrocytes crossing the glomerular membrane barrier causes the red blood cells to become misshapen or dysmorphic. It is unusual in these nephritic syndromes to see blood clots or gross hematuria, a term used for overtly blood-red urine that is often accompanied by dysuria and characterizes trauma and other syndromes of lower urinary tract bleeding. However, in episodes of immunoglobulin A (IgA) nephropathy, glomerular hematuria is often gross as red blood cells traverse diseased glomerular basement membrane (GBM).

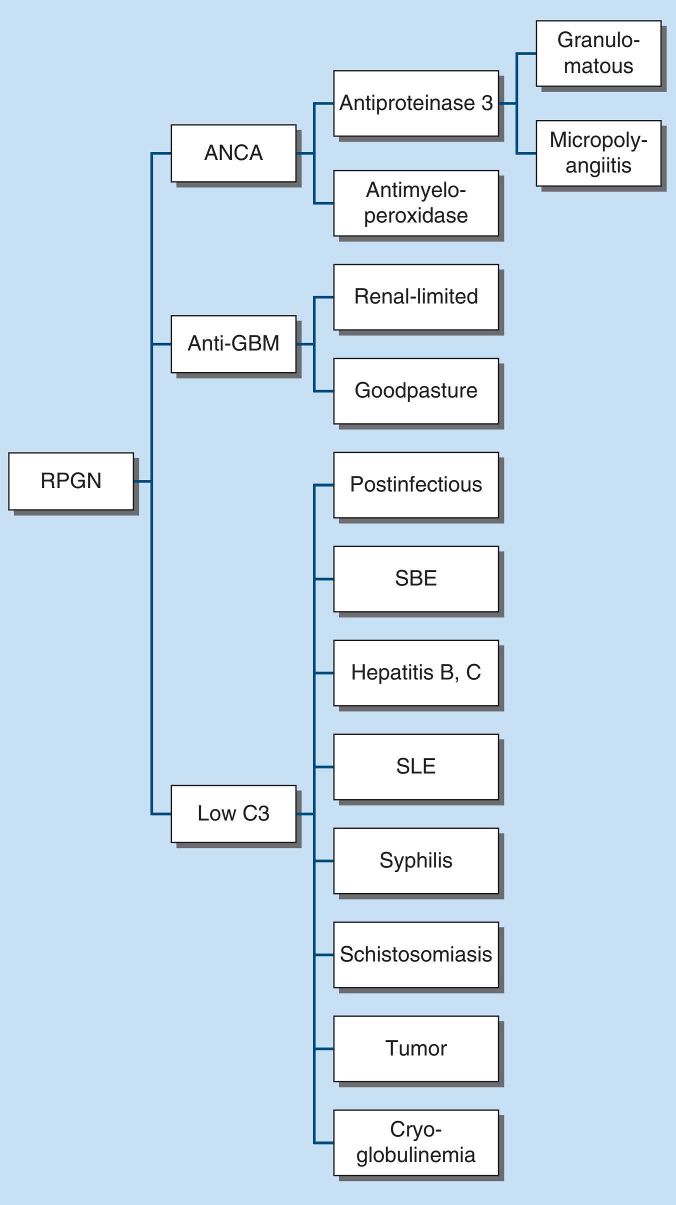

FIGURE 319-1 Rapidly progressive glomerulonephritis (RPGN). The syndrome of RPGN is diagnosed clinically. The three major categories are the antineutrophilic cytoplasmic antibodies (ANCAs; ANCA-positive vasculitides), anti–glomerular basement membrane antibody-positive (anti-GBM), and immune complex disorders with low C3 complement. C3 is synthesized in the liver, bound to circulating infectious or neoplastic antigens with their antibody complexes, deposited in the glomerular subendothelium, and associated with nonrenal manifestations. Syphilis is usually accompanied by vasculitis, and cryoglobulins may be seen in the setting of hepatitis C or myeloma and often lower C4 as well. SBE, subacute bacterial endocarditis; SLE, systemic lupus erythematosus.

Postinfectious Glomerulonephritis In postinfectious glomerulonephritis (PIGN), the classic example of acute GN, the immune response involves formation of antibody-bacterial antigen complexes within the circulation, usually 10 days to 3 weeks after infection of the pharynx or skin (as in impetigo) by specific nephritogenic strains of β-hemolytic group A streptococci. The deposition of these immune complexes amplifies and activates the complement cascade, causing an intense inflammatory response that accounts for the active (nephritic) sediment in the urine and decreased circulating C3 levels in the blood. The cellular proliferation is intense enough to minimize filtering surface area, causing a prerenal retention of Na^+ and water, oliguria, and sometimes anuria. Pathologically, the proliferation disrupts capillary blood flow, hence the description of "bloodless capillaries" for this condition.

In the general population, the incidence of PIGN has decreased given the widespread availability of antibiotics, specifically the penicillins and sulfa drugs. At the time of presentation with renal findings, there may be no ongoing infection, but rather a continued immune response to specific bacterial antigens. Antibiotics may not be necessary if the infection has resolved. Evidence indicating the initial infection may still be present in the form of titers of streptozyme, anti–streptolysin O, and anti–deoxyribonuclease B. Upon deposition, the circulating immune complexes bind and activate C3 complement, which, with other cytokines, causes a proliferative pattern of injury with subendothelial glomerular deposits. Taken together, there is increased clearance of total hemolytic complement and C3, causing hypocomplementemia. Over time, subepithelial deposits known as humps develop and proteinuria increases. As subendothelial deposits clear, nephritic signs disappear. Infective endocarditis, both acute and subacute, may also result in circulating immune complexes even while the patient remains on antibiotic treatment. These glomerulonephritic conditions are also recognized clinically by low C3 complement levels in the blood.

In PIGN, the postinfectious cases should be distinguished from cases of *infection-associated* GN that occur during an ongoing infection, such as a staphylococcal (often methicillin-resistant *Staphylococcus aureus*) abscess, osteomyelitis, or empyema, which are treated with antibiotics and purulent drainage and are characteristically associated with IgA deposition in the glomeruli. PIGN is also to be distinguished from synpharyngitic hematuria, another glomerular syndrome that frequently produces a heavier excretion of red blood (gross hematuria), which may appear to the patient as a frightening hemorrhage. The patient may present with this concern, so it is important for the clinician to recognize that as little as 10–20 mL of blood will turn a liter of urine red. Synpharyngitic hematuria, usually occurring simultaneously with, or 1–3 days into, a viral pharyngitis, is most often related to IgA nephropathy (IgAN) rather than PIGN. Though C3 complement is deposited in both PIGN and IgAN, only PIGN is associated with low complement levels in the blood.

IgAN is one of the leading causes of renal failure worldwide and is a common type of GN that is caused by the mesangial deposition of C3 binding to excess galactose-deficient IgA-1 immunoglobulin. This results in damage to the basement membrane, manifesting as hematuria and renal insufficiency. Although IgAN is a common disease, particularly in Asia, data on prevalence are often based on clinical assumptions rather than kidney biopsy and, therefore, are inaccurate. Treatment is based on each patient's personal presentation, but all patients need management of hypertension and proteinuria, often with anti–angiotensin II inhibitors. Because of its association with viral infections, IgAN has historically been treated with removal of lymphoid tissue with strategies such as tonsillectomy. Other treatments have included fish oils and corticosteroids. However, recent advances in understanding the lectin and alternative pathways of complement have led to investigation of targeted therapies based on kidney biopsies showing IgA-1 and glomerular deposition of mannan-binding lectin, which occurs in 50% of patients with IgAN. Complement blockers, B-cell therapies, and plasma cell therapies are currently either in use or being further investigated to treat patients with a genetic predisposition for high levels of galactose-deficient IgA-1 in order to antagonize development of autoantibodies to the abnormal IgA1 or to interfere with the proinflammatory complement binding to the largely mesangial deposits of the immune complexes. Avacopan, a selective C5a receptor inhibitor effective in treating antineutrophil cytoplasmic antibody-associated (ANCA)-vasculitis, lowers proteinuria in IgAN nephropathy.

The U.S. Food and Drug Administration has recently approved an oral dual endothelin 1 and angiotensin II receptor antagonist for the reduction of proteinuria in adults with primary IgAN at high risk for rapid disease progression. These individuals typically have a urine protein-to-creatinine ratio of >1.5 g/g and decreased GFR. Patients with chronic kidney disease and either micro- or macrohematuria as a sign of disease have been less likely to receive a renal biopsy than patients with azotemia or proteinuria. Increasing appreciation that patients with IgAN have a more rapid GFR decline to end-stage kidney disease has encouraged greater use of renal pathology to guide a more aggressive management strategy. Attention to risk factors and advances in development of novel pharmacologic approaches based on a better understanding of the pathogenesis of IgAN have led to greater judgement to perform a renal biopsy for diagnosis and treatment. Increasing appreciation that patients with IgAN, particularly those of African ancestry with high-risk APOL1 alleles have a more rapid GFR decline, and clinicians are encouraged to make use of renal pathology to guide a more aggressive management strategy.

Rapidly Progressive Glomerulonephritis (RPGN) Rapidly progressive glomerulonephritis (RPGN) is a syndrome arising from a variety of causes (Fig. 319-1). Pathologically, it is associated with a proliferation of glomerular parietal epithelial cells and inflammatory cells, known as cellular crescents, that surround and compress the capillaries, reducing the filtration surface area. Over time, the crescents become fibrotic and atrophic with global loss of the glomerular tuft. When >50% of glomeruli are affected (diffuse), this highly destructive process usually leads to global glomerulosclerosis. When <50% of glomeruli are acutely affected (focal) and only part of each glomerulus (segment) is involved, a condition known as focal and segmental necrotizing GN develops, with or without crescents. When such lesions heal, focal segmental glomerulosclerosis (FSGS) results. This pathological finding on biopsy must be distinguished from the primary podocytopathy (FSGS that causes nephrotic syndrome) and from adaptive changes seen in most progressive chronic kidney diseases. Recognition of RPGN and choice of appropriate therapy are extremely important because, if left unchecked, the capillaritis can lead to complete and irreversible loss of kidney function, as well as fatal pulmonary hemorrhage when associated with lung vasculitis.

In the case of RPGN, one may see the typical signs of the nephritic syndrome, usually with subnephrotic proteinuria, although other clues to diagnosis may be present in pulmonary-renal syndromes, where the patient may present with hemoptysis, interstitial lung disease, epistaxis, or upper airway symptoms of sinusitis or nasal congestion. Blood testing, including at least three necessary immunologic studies, help narrow the differential diagnosis. The antineutrophil cytoplasmic antibody (ANCA) test, particularly anti-proteinase-3 antibodies (cANCA) or anti-myeloperoxidase (pANCA), is diagnostic for the pauci-immune disorders of systemic vasculitis, such as granulomatous vasculitis (formerly known as Wegener's granulomatosis), giant cell arteritis, and microscopic polyangiitis, which are particularly associated with pulmonary and upper respiratory disease. In ANCA-positive diseases, the kidney and peripheral vessels are the target of autoantibodies, which attract complement and result in inflammatory vasculitis. Churg-Strauss syndrome, which is both granulomatous and eosinophilic, is a vasculitis associated with asthma.

By contrast, a positive test for anti-GBM antibodies, with or without a positive ANCA, would be consistent with renal-limited anti-GBM disease. One key distinction between ANCA-positive vasculitis and anti-GBM disease is that the vasculitides may recur, even subclinically historically at different times, whereas onset of anti-GBM is usually a single acute fulminant presentation that must be diagnosed rapidly with the ANCA and anti-GBM studies and renal pathology and treated promptly with plasma exchange followed by anti-inflammatory medication and cyclophosphamide or rituximab as immunotherapy to preserve kidney function. A positive anti-GBM test, when associated with pulmonary hemorrhage, is consistent with Good pasture disease, which is a disorder most often seen in an elderly population or in young males who smoke or have a history of inhaling hydrocarbon solvents. Not infrequently, both anti-GBM and ANCA are present with pathology showing pauci-immune disease, where the proposed role of ANCA is to expose the GBM epitope. This disorder should be treated in the same way as ANCA itself. The goal of plasmapheresis is to remove all of the GBM autoantibody, which may take weeks to months. The circulating IgG anti-GBM, complexed in situ with C3 complement, is directed at an epitope in the α3 chain of the non-collagenous domain of Type 4 collagen in the basement membrane. This occurs at the site absent in hereditary nephritis. People with hereditary nephritis do not develop anti-GBM disease; however, if they receive a kidney in which the donor kidney contains the missing antigen, they may present with fulminant anti-GBM disease.

The immune complex diseases that may also manifest as RPGN may have low circulating levels of the third component of complement (C3) secondary to focal or diffuse immune complex deposits in the glomerular capillary subendothelial space. Certain of these disorders, such as SLE and cryoglobulinemia, can present with pulmonary hemorrhage. Patients with cryoglobulinemia often have low C4 complement and a high rheumatoid factor, and the syndrome may be caused by paraproteinemias or hepatitis C. Another cause of crescentic RPGN but with normal levels of C3 complement is a form of IgA vasculitis (Henoch-Schönlein purpura), a syndrome of skin vasculitis characterized by palpable purpura, gastrointestinal bleeding, arthralgias, and pulmonary hemorrhage. The C3 complement (also an acute-phase reactant) is detected on the kidney biopsy, particularly with galactose-deficient IgA1, in the mesangium rather than the capillary loops, and is associated with mesangial cell proliferation but not with low serum levels of C3 or IgA. The diagnostic tests and distinct diseases associated with these immune complex disorders are shown in Fig. 319-1.

ACUTE TUBULOINTERSTITIAL NEPHRITIS

Tubulointerstitial nephritis (TIN) comprises inflammatory disorders of the renal tubules and interstitium that may be caused by infection, autoimmune disease, or allergic immunologic responses to certain drugs **(Fig. 319-2)** and that have a time course ranging from days to weeks or months.

Acute allergic immune tubulointerstitial nephritis (ATIN) usually develops 1 day to 2 weeks following an exposure, change in dose, or re-exposure to an offending drug (anamnestic response) and may be associated with rapid and potentially reversible loss of kidney function, making early diagnosis imperative. Associated heavy glomerular proteinuria sometimes occurs with the use of nonsteroidal anti-inflammatory drugs (NSAIDs) or ampicillin, whereas protein secreted by the proximal tubule (β_2-microglobulinuria) indicates tubular disease. Clinically, there may be fever, rash, and eosinophilia, any of which is typical with certain types of penicillins, cephalosporins, fluoroquinolones, and some biologic cancer drugs that act as checkpoint inhibitors (CPIs), but are atypical for NSAIDs. Patients who recover from CPI-induced acute kidney injury (AKI) may resume treatment with the inhibitor. The urinalysis usually shows pyuria and at times eosinophiluria, but the most characteristic cell types are activated T lymphocytes, eosinophils, and plasma cells, along with white blood cell casts. A cyto-centrifuged urine sample in a small test tube, when exposed to centrifugal forces, sprays sediment onto a slide mounted perpendicular to the tube's aperture. The slide, then stained with Giemsa, best demonstrates these cell types.

The patient may experience symptoms of polyuria, tender kidneys, and signs of tubular dysfunction, including nephrogenic diabetes insipidus, hypo- or hyperkalemia, and hyperchloremic metabolic acidosis. Common drugs that cause ATIN include proton pump inhibitors (PPIs) and sulfa drugs, especially sulfamethoxazole, as well as sulfa-containing diuretics such as acetazolamide, thiazides, furosemide, and, less frequently, bumetanide. PPIs may increase risk for ATIN in patients on CPIs. Frequently, rifampin and several other drugs cause noncaseating granulomatous interstitial nephritis. In caseating granulomas resulting from hematogenous spread of *Mycobacterium tuberculosis*, the first granulomas appear connected to the glomeruli in the cortex, where oxygen tension is at its highest. Granulomatous tubulointerstitial nephritis may occur in sarcoidosis and histoplasmosis.

Systemic infections including bacterial, viral, fungal, and parasitic may induce tubulointerstitial nephritis and were first described pathologically in the preantibiotic era as Councilman's nephritis in the course of scarlet fever. Autoimmune interstitial nephritis is seen in diseases such as lupus nephritis, ANCA-positive vasculitis, and other rheumatic disorders including Sjögren's syndrome, which may present with dry eyes as part of sicca syndrome. In a patient with photophobia and red painful eyes, the clinician should look at renal function and the urinalysis for evidence of tubulointerstitial nephritis uveitis syndrome.

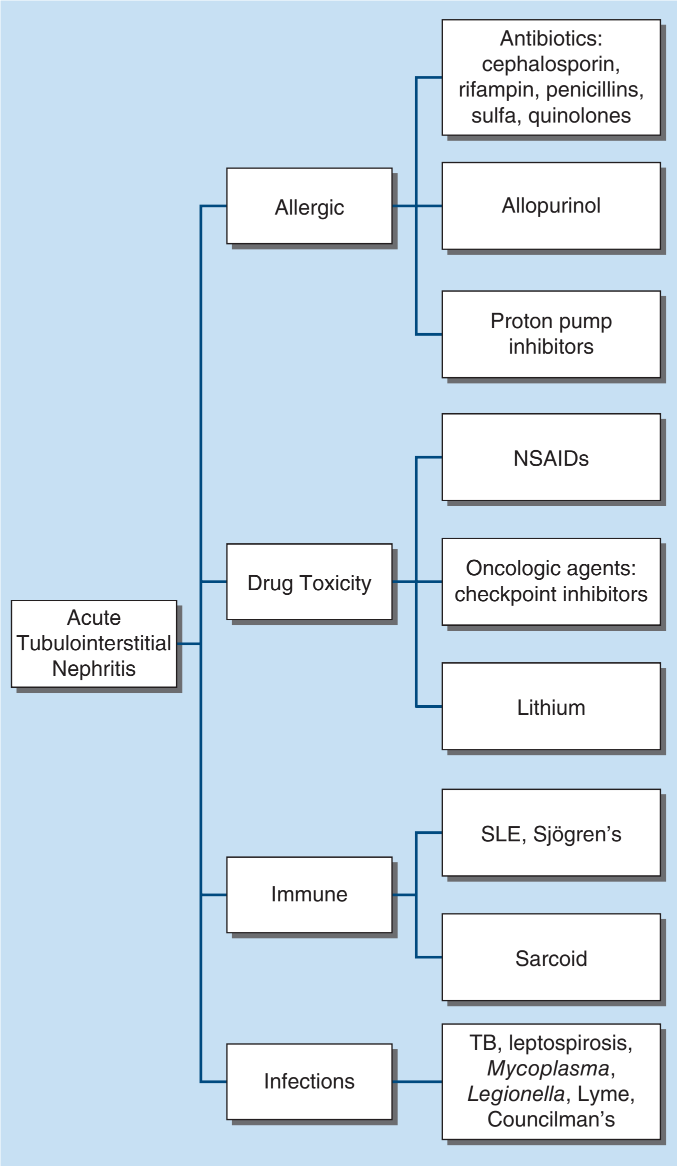
Acute Tubulointerstitial Nephritis
Allergic
Antibiotics: cephalosporin, rifampin, penicillins, sulfa, quinolones
Allopurinol
Proton pump inhibitors
Drug Toxicity
NSAIDs
Oncologic agents: checkpoint inhibitors
Lithium
Immune
SLE, Sjögren's
Sarcoid
Infections
TB, leptospirosis, *Mycoplasma*, *Legionella*, Lyme, Councilman's

FIGURE 319-2 Acute tubulointerstitial nephritis. Diseases resulting from injury to the tubular and interstitial components of the renal cortex and medulla are shown. NSAID, nonsteroidal anti-inflammatory drug; SLE, systemic lupus erythematosus; TB, tuberculosis.

CHRONIC TUBULOINTERSTITIAL NEPHRITIS

Noninflammatory interstitial diseases are often caused by toxic exposures that damage the tubular interstitial structures. For example, a cancer patient on ifosfamide may develop Fanconi's syndrome, indicating proximal tubule damage. Heavy-metal exposures such as cadmium, lead, and mercury from old dental fillings may lead to proximal tubule injury and, once again, Fanconi's syndrome. A patient with hypomagnesemia and AKI may have had past exposure to platinum-containing chemotherapeutic agents, and a patient with nephrogenic diabetes insipidus, unresponsive to antidiuretic hormone, may have had treatment with lithium, analgesics, or chemotherapy. Hypokalemia and interstitial disease are consistent with a past history of exposure to aminoglycosides or amphotericin B. The patient with the former may complain of hearing loss, and the patient with the latter may present with hypokalemia and renal tubular acidosis (RTA).

Chronic interstitial diseases should be suspected in cases of paraproteinemia (light-chain nephropathy or amyloidosis) and in patients who ingest herbal remedies containing aristolochic acid, as occurs commonly in China and has been determined to be the cause of Balkan nephropathy. Perhaps the most common cause of chronic interstitial nephritis is prolonged analgesic use to treat chronic pain (not only NSAIDs but also acetaminophen or combinations of phenacetin, aspirin, and caffeine) as part of the analgesic syndrome. Clinicians should inquire about a prior history of pain and also gastrointestinal symptoms that may precede the kidney disease; the analgesic drugs often go unreported by the patient. It is important to recognize this syndrome because, if the drugs are not discontinued, a later development may be uroepithelial cancers of the distal ureter and urinary bladder. Lithium after prolonged use can also cause chronic interstitial nephritis along with nephrogenic diabetes insipidus and slowly progressive kidney failure.

Patients who have hypercalcemia or hyperoxaluria may develop nephrocalcinosis (a form of interstitial nephritis) that is characterized by calcifications within the renal parenchyma, often at the cortical medullary boundary. When nephrocalcinosis and nephrolithiasis are concurrent, the most common etiologies include hypercalcemic disorders, particularly primary hyperparathyroidism, and congenital medullary sponge disease (tubular ectasia), as well as hereditary distal RTA. In patients with DM, chronic analgesic use, or sickle cell diseases, the phenomenon of papillary necrosis is associated with chronic interstitial damage characterized by nephrogenic diabetes insipidus. For example, phenacetin toxicity is a result of the drug concentrating in the medulla due to the normal urinary concentrating mechanism; ironically, the first function lost because of medullary toxicity is the ability to concentrate the urine. In the ensuing papillary necrosis, the patient may observe solids in the urine, which are sloughed tissue from the ischemic medulla. When a patient with inflammatory bowel disease or one who has had gastric bypass procedures such as the Roux-en-Y develops kidney injury, with or without calcium oxalate kidney stones, a 24-h urine oxalate measurement must be determined to diagnose intestinal hyperoxaluria, which otherwise may lead to nephrocalcinosis. Chronic bacterial pyelonephritis is not a common cause of chronic renal failure, as it rarely affects both kidneys. Acute pyelonephritis in a solitary kidney or transplanted kidney may cause AKI.

PROTEINURIC STATES AND NEPHROTIC SYNDROME

Proteinuric States Some patients notice generalized edema and foamy urine, manifestations of albuminuria. Proteinuria is not synonymous with nephrotic syndrome (NS). In DM, the urinary loss of 30–300 mg/d of microalbuminuria (MALB) defines excessive albuminuria. Glomerular proteinuria, which typically occurs after years of small-vessel disease (explaining the likelihood of coincident small-vessel retinopathy), predicts future development of progressive renal failure and NS. The rate of renal progression is greater with hypertension, obesity, and poor glucose control. Further risks are partial nephrectomy, hepatitis C, and FSGS in patients of African ancestry, possibly related to predisposition by and prevalence of *APOL1* gene mutations. APOL1-mediated kidney disease is caused by two risk variants of the *APOL1* gene. Gene expression occurs only when a second hit of infection or inflammation triggers damage to podocytes and glomerular endothelial cells, creating channels in those cells that result in dysfunction and proteinuria.

Because treatment of patients with MALB is indicated, the clinician caring for diabetic patients should screen regularly for the presence of small amounts of albumin in the urine, which can be detected as microalbuminuria by radioimmunoassay. Subnephrotic albuminuria is characteristic of focal diseases of the kidney with <50% of glomeruli involved, whereas NS is likely when disease is diffuse, involving most glomeruli. Lower levels of albuminuria are characteristic of GN, as discussed above.

Microalbuminuria is beneath detection by urinary dipstick protein analysis. Any detectable albuminuria on dipstick is called overt proteinuria and, when detected at the highest level on the strip, is consistent with nephrotic proteinuria. The dipstick measurement picks up only the most acidic proteins like albumin but not proteins with a higher isoelectric point, most notably kappa and lambda light chains found in multiple myeloma. A patient being evaluated for renal disease should have either a timed direct measure of albumin or protein or, at least, a urinary microalbumin-to-creatinine ratio. The filtration of light chains may cause glomerular damage and albuminuria (kappa light-chain nephropathy or lambda AL amyloid) but does not reliably indicate glomerular damage because their charge and molecular weight allow them to freely cross the glomerular barrier. Such filtration is considered overflow proteinuria.

A laboratory measurement of albumin or total protein concentration should be normalized to urine creatinine concentration as a ratio in order to discount the effects of urinary dilution or concentration. It is also important to note that the ratio of total protein concentration to creatinine concentration is not specific for albumin excretion and may indicate the presence of light chains. The detection of light chains requires serum or urine protein immunoelectrophoresis.

Nephrotic Syndrome NS has three defining features: edema, hypoalbuminemia (<3.5 g/dL), and proteinuria >3.5 g/d. This syndrome is often associated with lipid abnormalities such as high low-density lipoprotein, low high-density lipoprotein, and lipiduria. The urine may contain large tubular epithelial cells that are engulfed with lipids in a recognizable shape under polarized light. These oval fat bodies may also contain cholesterol monohydrate crystals, which appear as Maltese crosses. Under low light, the refractile lipid may be seen as fatty casts.

Clinically, patients present with generalized edema, and in contrast to heart failure with orthopnea, facial, eyelid, and periorbital swelling is observed in NS. Penile and scrotal edema can be severe enough to obstruct urethral urine flow in men. In some cases, edema may form large bullae that may rupture, predisposing to ulceration and cellulitis. The skin becomes smooth and may appear to "weep." Additionally, patients may present with hoarseness caused by vocal cord edema. Occasionally, nephrotic range proteinuria is not associated with edema (i.e., in the case of HIV-associated nephropathy [HIVAN], which is more common in African Americans with HIV infection). In these cases, rapid loss of renal function may occur due to implosion of the glomerular capillary by proliferation of visceral epithelial cells that crush the glomerular loops, decreasing capillary flow and filtering surface area (collapsing glomerulopathy). This glomerulopathy is a severe, treatment-resistant form of FSGS. Very heavy proteinuria is often noted in this syndrome.

Other secondary forms of FSGS, more often not associated with NS, include remnant kidney after partial surgical removal, adaptive injury, chronic hypoxemia from chronic lung or congenital heart disease, sickle cell diseases, recurrent vasculitis with healing of some glomeruli, obesity, and congenital urogenital anomalies. Secondary forms of FSGS are shown in **Fig. 319-3**.

NS is a hypercatabolic state with negative nitrogen balance due to proximal tubule absorption and lysosomal catabolism exceeding hepatic albumin synthesis. A characteristic finding of rapid-onset hypoalbuminemia in NS is horizontal linear white lines in the nail bed, known as Muehrcke's lines. When NS in the adult occurs abruptly, with severe elevation of cholesterol, one must consider glomerular epithelial injury (podocytopathy), which may be idiopathic minimal change disease (MCD), a name historically based on the appearance of renal tissue on light microscopy and which is often preceded by an upper respiratory infection, allergies, or immunization. Secondary causes of MCD are Hodgkin's lymphoma or other lymphoproliferative disorders. MCD can occur at any age and in the adult may present with AKI as well, primarily because of the underlying vascular disease and hypoalbuminemia.

Age is an important aspect in the onset of NS, in that young children <6 years of age frequently have MCD. Several gene mutations encoding for proteins of the podocyte, such as nephrin (*NPHS1*) and podocin (*NPHS2*) in the slit diaphragm, have been shown to be responsible for most cases of hereditary NS **(Table 319-2)**.

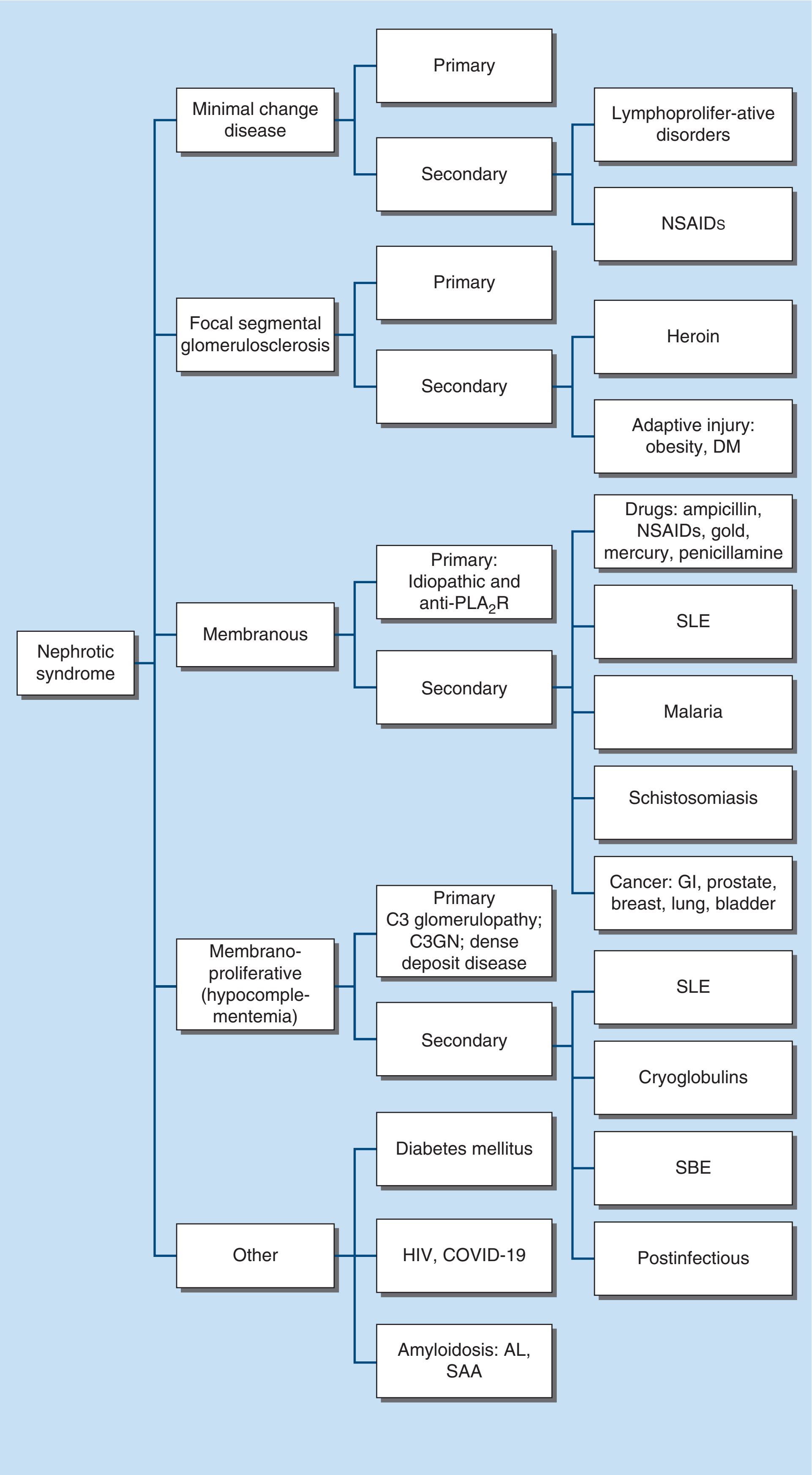
Nephrotic syndrome
Minimal change disease
Primary
Secondary
Lymphoprolifer-ative disorders
NSAIDs
Focal segmental glomerulosclerosis
Primary
Secondary
Heroin
Adaptive injury: obesity, DM
Membranous
Primary: Idiopathic and anti-PLA$_2$R
Secondary
Drugs: ampicillin, NSAIDs, gold, mercury, penicillamine
SLE
Malaria
Schistosomiasis
Cancer: GI, prostate, breast, lung, bladder
Membrano-proliferative (hypocomple-mentemia)
Primary C3 glomerulopathy; C3GN; dense deposit disease
Secondary
SLE
Cryoglobulins
SBE
Postinfectious
Other
Diabetes mellitus
HIV, COVID-19
Amyloidosis: AL, SAA

FIGURE 319-3 Nephrotic syndrome. Different pathologic syndromes are broken down into idiopathic and secondary causes. Seventy percent of primary membranous nephropathy (MN) is associated with antibodies to the integral subepithelial basement membrane antigen phospholipase A$_2$ receptor (PLA$_2$R). The former term for primary membranoproliferative glomerulonephritis has been replaced by pathologic syndromes involving complement deposition (C3 glomerulopathy and dense-deposit disease in children). AL amyloid is the form of amyloidosis secondary to λ light-chain deposition, and serum amyloid A (SAA) is the protein associated with chronic inflammatory diseases such as rheumatoid arthritis, familial Mediterranean fever, and tuberculosis. DM, diabetes mellitus; GI, gastrointestinal; NSAID, nonsteroidal anti-inflammatory drug; SBE, subacute bacterial endocarditis; SLE, systemic lupus erythematosus.

TABLE 319-2 Hereditary and Congenital Diseases

HEREDITARY GLOMERULAR DISEASES: NEPHROTIC SYNDROME OR RENAL FAILURE	HEREDITARY TUBULAR DISEASES	APOL1	CILIOPATHIES	CHANNELOPATHIES	CONGENITAL ANOMALIES
Podocyte mutations; basement membrane mutations; hereditary nephritis; X-linked and somatic mutations; *COL4A5*; Alport's syndrome with hematuria, neurosensory hearing loss, conus lentus; Fabry's disease (X-linked alpha-galactosidase deficiency)	Fanconi's syndrome with renal glycosuria, proximal tubule renal tubular acidosis (RTA), hypophosphatemia, hypouricemia; distal RTA with hearing loss, renal calcification; nephrogenic diabetes insipidus; hyper- or hypotensive disorders with hypo- or hyperkalemia; metabolic alkalosis or acidosis	Mutant allele in African-American population, leading to susceptibility to further injury in many other disorders including diabetes mellitus, IgAN	Autosomal dominant and recessive polycystic kidney disease; medullary cystic kidneys (metabolic acidosis and salt-wasting); other phenotypes	Gitelman's and Bartter's syndromes; disorders of epithelial sodium channel and proton pumps	Unilateral agenesis; dysgenetic disease; medullary sponge kidney (hematuria, stones, or infection); horseshoe kidneys (proteinuria); malposition; ureteropelvic obstruction; cystoureteral reflux (proteinuria)

In children who have a serum albumin concentration <2 g/dL, tissue ischemia may develop, causing a "nephrotic crisis," in which severe abdominal pain may be mistaken for a surgical abdomen. A 20-year-old presenting with NS may have MCD, FSGS, or membranous nephropathy (MN).

Similar to MCD, MN may be of primary or secondary etiology. Whereas NS is a cause of a hypercoagulable state, particularly when it is accompanied by severe hypoalbuminemia (<2 g/dL), MN in NS is most often associated with renal vein thrombosis (RVT) for a number of reasons, among them being urinary loss of antithrombin-3 and plasminogen, as well as hyperfibrinogenemia. RVT may present with back pain, renal failure or hematuria, increased edema in the lower extremities, or pulmonary emboli.

NS with heavy proteinuria or large kidneys on ultrasound in advanced stages of chronic kidney disease (CKD) is often noted in diabetic nephropathy, amyloid, HIVAN, and MN with RVT. MN may be the presenting finding in cases of SLE, whereas diffuse proliferative GN with low C3, when associated with SLE, occurs in patients who also have signs and symptoms of joint, skin, or systemic organ involvement. A cause of very early MN is an idiosyncratic reaction to NSAIDs, even when taken in small amounts. Many other drugs including penicillamine, mercury, and gold can cause MN. Proteinuria from any cause may present earlier, or increase, when DM, obesity, partial nephrectomy, or increased renal venous pressure exists.

Solid tumors, particularly lung, gastric, intestinal, bladder, breast, and prostate, may underlie a paraneoplastic syndrome of MN. While it is important to take a history and examine the patient carefully for signs of cancer, it is not the rule to work up every patient with MN exhaustively for the disease. Neoplastic nephropathy may be a sign of recurrent disease and may even be recognized by tumor antigens. MN also occurs in the context of many infectious diseases, including syphilis, malaria, schistosomiasis, and hepatitis B. Recently, 70% of patients with primary MN have been found to have circulating levels of an autoantibody to the phospholipase A_2 receptor (anti-PLA$_2$R) in the subepithelial region of the GBM. Compared to non-PLA$_2$R patients, fewer patients in this group have drug-induced MN and are more likely to experience spontaneous remission; one of four subclasses of immunoglobulins has been implicated in TIN and sometimes MN as part of a multiorgan disease. IgG4 increases in response to chronic exposure to an antigen; it has been associated with autoimmune diseases such as acute pancreatitis-1 (AP1) and severe and often allergic asthma and is often detected by high levels in the blood. Patients who are thought to have diabetic nephropathy on clinical grounds are often found to have an additional diagnosis, particularly MN, on renal biopsy. The increased frequency of MN in DM may indicate that glomerular hyperfiltration accentuates the proteinuria of MN or is associated with a genetic variant, *COL45A5*. Diabetic nephropathy is perhaps the most common reason for the renal patient to begin dialysis; therefore, much attention has been paid to the progressive nature of chronic failure in these patients. Risk factors include those that affect hemodynamics and blood sugar. In the poorly controlled diabetic patient, increased blood pressure, renal plasma flow, and GFR develop in single nephrons that are spared until glomerular sclerosis in other nephrons has already occurred. This finding is clinically evident by the presence of proteinuria or low GFR. The proteinuria is responsive to angiotensin II–inhibiting drugs such as angiotensin-converting enzyme inhibitors, angiotensin II receptor blockers, and renin inhibitors. Sodium-glucose cotransporter 2 (SGLT2) inhibitors in the proximal tubule of the kidney create glycosuria, improve glucose control, and have been shown to reduce the rate of progression. The more glucose filtered during periods of hyperglycemia, the greater are the renal interstitial glucose and advanced glycation end products that cause fibrosis.

Patients who have nephrotic and nephritic sediments in childhood may have membranoproliferative glomerular nephritis with low C3 as a primary disease (C3 glomerulopathy or dense-deposit disease). The adult presenting with a nephrotic and nephritic profile should be considered to have an immune complex disorder unless proven otherwise. In amyloidosis, there may be a simultaneous and severe tubular syndrome in addition to the albuminuria, consisting of nephrogenic diabetes insipidus and a hyperkalemic RTA. When NS is associated with Fanconi's syndrome (glycosuria, phosphaturia, uricosuria, and aminoaciduria), one must consider multiple myeloma in the adult.

HEMATURIA AND LOWER URINARY TRACT SYNDROMES

Since blood could enter the urinary tract from any structure, identifying the source is the necessary first step. As a rule, hematuria associated with flank pain or ureteral colic is more characteristic of a lower urinary tract source, such as a stone or an obstructing lesion. However, when flank pain without colic is noted, one has to consider swelling of the kidney and stretching of the renal capsule, as might be seen in acute urinary obstruction, kidney infection, interstitial nephritis, and, on occasion, acute glomerular nephritis. Heavy bleeding with clots is more characteristic of lower tract bleeding but could result from trauma to the kidney. In patients who have acute flank pain and hypertension, assuming the pain is due to a kidney stone may miss an occlusive vascular event from an embolus or in situ clot within the arterial blood supply. Patients who engage in strenuous exercise may develop hematuria that can be ruled out by retesting the urine after a week or two of rest. True hematuria can be distinguished from hemoglobinuria and myoglobinuria by urinalysis.

Symptoms of lower urinary tract disorders are dysuria, urinary frequency, urgency, incomplete emptying, hematuria that is most pronounced at the beginning of flow, and a poor urinary stream. Among these disorders are urinary tract infections and prostatic hypertrophy. Hematuria of recent onset noted in the fourth decade of life or beyond should undergo a urologic evaluation and most likely cystoscopic examination of the bladder. However, chronic hematuric conditions may be associated with chronic glomerular disease such as thin basement membrane, hereditary nephritis, or IgAN. This is why it is so important to search for previous examinations such as insurance, military, or employment examinations done in the past to determine if hematuria was present in order to establish chronicity.

ACUTE KIDNEY INJURY

AKI is defined by the retention of nitrogenous solutes such as urea, uric acid, and creatinine. The creatinine concentration goes up primarily related to the amount of water absorbed along the nephron. Because urea but not creatinine has a reabsorptive component, clearance of urea is less than the clearance of creatinine. With a constant catabolic rate, a decrease in glomerular filtration alone will affect the blood urea nitrogen (BUN) and creatinine proportionately. By contrast, in a disorder associated with enhanced urinary concentration (high antidiuretic hormone [ADH]), where urea and water have increased reabsorption in the process of concentrating the urine, the retention of urea nitrogen exceeds that of creatinine, accounting for an elevated BUN-to-creatinine ratio. Oliguria or oligoanuric states are often characterized as prerenal, defined as preglomerular vascular disease or low perfusion states; renal, indicating intrinsic renal disease; or postrenal, indicating obstructive nephropathy or uropathy (see Table 319-1).

Prerenal states are often related to a decreased perfusion pressure to the glomerular capillaries or some other interference with filtration. The differential diagnosis includes proliferative GN (the bloodless capillary); a preglomerular vascular disease such as scleroderma or TMA; the use of pressors for the maintenance of blood pressure; vasoconstrictive substances including cocaine, iodinated IV contrast media, hemoglobin or myoglobin, and certain antibiotics such as vancomycin, cyclosporine, and tacrolimus (calcineurin inhibitors); and other vasoconstrictive drugs, such as NSAIDs and renal nerve stimulators. Hypercalcemia and hypoxia are also vasoconstrictive. Correcting hypoxemia has a diuretic effect.

Intrinsic Renal Disease All of these disorders will decrease perfusion pressure and decrease GFR while stimulating sodium and water reabsorption. Therefore, the characteristic urinary findings are low urinary sodium excretion and concentrated urine. The urine sodium concentration is usually low and the urine to plasma creatinine concentration is usually >40. The same result could occur in bilateral renal artery stenosis, heart failure, and other circulatory collapse. If there is damage done to the renal tubules (so-called intrinsic renal disease), there may or may not be oliguria, and in fact, there could be polyuria (unresponsiveness to ADH) and sodium-wasting, potassium-wasting, and acid-base disorders, each of which may predominate. One of the most frequent causes of oliguric acute renal failure is acute tubular necrosis, caused by gram-negative sepsis, hemodynamic collapse, hemoglobinuria or myoglobinuria, and medication or toxin ingestion. Examples are crush injury, heroin and alcohol stupor, and compartment syndrome. Drug causes include NSAIDs, aminoglycosides, and chemotherapeutic drugs such as cisplatin and methotrexate.

Postrenal States Postrenal states accompanied by anuria suggest complete obstruction to the flow of urine, while polyuria and bland urinary sediment suggest incomplete obstruction combined with ADH unresponsiveness. The fractional excretion of sodium is usually >1–2%, whereas in oliguric prerenal syndromes, the fractional excretion of sodium is usually <1%. Exception to the low sodium excretion would be in the case of hyperchloremic metabolic alkalosis, where chloride is being conserved and sodium excreted with bicarbonate.

Urinary tract obstruction (UTO) is usually characterized by hydronephrosis visualized with ultrasound imaging or large volumes of urine removed by bladder catheterization, indicating bladder outlet obstruction. Exceptions to finding hydronephrosis on an imaging study include the extrinsic obstructive compression of the kidney or ureter in retroperitoneal disease; cases in which imaging was done too early, for example <4 days after obstruction; or when the patient was simultaneously obstructed and volume depleted.

As previously mentioned, acute tubular diseases that are not primarily due to inflammation can be caused by intratubular obstruction, hyperoxaluric states, or uric acid or phosphates in tumor lysis syndrome following chemotherapy for lymphoproliferative disorders. Other drug therapies that cause tubular injury include cisplatin, ifosfamide, methotrexate, aminoglycosides, and amphotericin B, and toxicity through exposure to heavy metals.

Kidney injury in the setting of systemic disease with dysfunction in other organs includes cardiorenal, pulmonary-renal, and hepatorenal syndromes. Other syndromes that typically involve the kidney include febrile illness, systemic infection with septic shock, hemodynamic collapse due to cytokines, or capillary leak syndrome. These syndromes and the endothelial origins of TMAs and hemolytic-uremic syndrome are shown in **Fig. 319-4**. Obstetric patients with abruptio placenta may develop bilateral cortical necrosis and, as a consequence, anuria and, often, irreversible renal failure. In addition, patients with atheroembolic syndrome can present with AKI or chronic progressive damage following an aortic catheterization. If the blood vessels catheterized involve the aortic arch, an encephalopathic syndrome that can mimic uremic encephalopathy may follow. As many as 70% of patients with COVID-19 who required ventilators during the early phase of the pandemic developed multiorgan failure, including AKI. AKI may also be a risk for CKD following a COVID-19 infection, but most patients return to baseline function.

CHRONIC KIDNEY DISEASE AND THE UREMIC SYNDROME

CKD, progressing over months to years, can be the long-term result of any of the aforementioned diseases. If the time course is unknown, other ways to diagnose CKD versus a more recent kidney disorder are useful (i.e., the finding of bilaterally small kidneys on a renal ultrasound). Kidneys imaged by this method are normally 10–12 cm in length. Small kidneys, <8 cm, are likely atrophic with irreversibly low function; but in some cases of CKD, such as in DM, the kidneys may be large despite kidney failure. Remarkably, even atrophic kidneys may still produce renin to sustain blood pressure and erythropoietin to minimize anemia; however, due to low calcitriol in CKD, patients develop secondary hyperparathyroidism. The normal parathyroid gland weighing 25 mg would be limited in raising parathyroid hormone (PTH) levels, such that very elevated PTH likely favors a diagnosis of CKD. Thinning of the renal cortex is also a sign of chronicity.

DM accounts for ~50% of CKD patients who progress to end-stage renal disease (ESRD). Other important causes leading to ESRD are IgAN; ischemic nephropathy without proteinuria manifesting after age 50 with hypertension and signs of other large-vessel disease, such as intermittent claudication from peripheral vascular disease; stroke; and coronary artery disease. By contrast, renal arterial disease due to fibromuscular dysplasia seen predominantly in white women in the fourth decade of life does not progress to renal failure and is a treatable form of hypertension. Other arterial diseases include polyarteritis nodosum and, in medium-sized vessels, a Kawasaki-like illness affecting children, which has had a 30-fold increase since the onset of COVID-19 as part of the multisystem inflammatory syndrome.

Hereditary and congenital diseases of the kidney are shown in Table 319-2. Autosomal dominant polycystic kidney disease (PCKD) occurs in all ethnic populations. The patient may first show signs of cysts in late teen years with hypertension, urinary or cyst infection, or pain and bleeding. Kidney stones may be seen. By the age of 30, individuals with the disorder will have ultrasound-detectable cysts. The clinician should be aware that a diagnosis of PCKD requires multiple cysts in both kidneys. Cysts are also found in the liver. The first manifestation of PCKD may be a cerebral hemorrhage, since brain aneurysms are present in a small subset of patients with this disease, most commonly in the circle of Willis. In PCKD, rapid progression of kidney failure has been observed and previously explained by the crowding out of normal nephrons by the cysts. Cyst enlargement is due to fluid secretion as well as increased growth of the cyst wall. It is helpful to measure the size of cysts and to antagonize growth with tolvaptan, an inhibitor of vasopressin type 2 receptors in the collecting duct. Vasopressin, however, is a growth factor for cysts. Risk factors for more rapid progression include African-American heritage, obesity, liver abnormalities, and hypertension.

Hereditary nephritis with many variations includes Alport syndrome, involving a gene mutation (*COL4A5*) affecting the α3 region of the noncollagenous domain of the alpha 3 chain type IV collagen in GBM. There is an X-linked pattern of genetic transmission, where

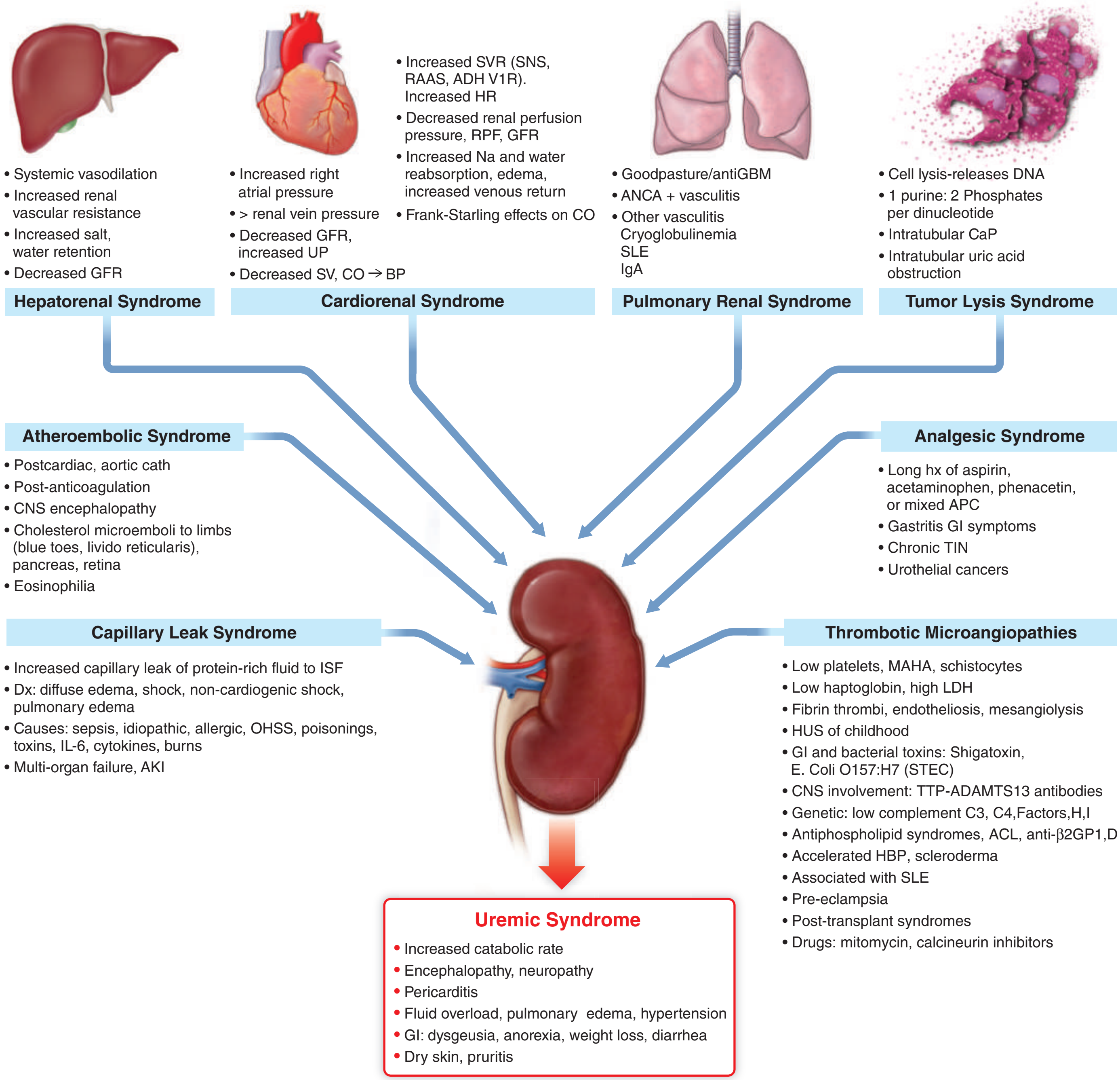
• Systemic vasodilation
• Increased renal vascular resistance
• Increased salt, water retention
• Decreased GFR
Hepatorenal Syndrome
• Increased right atrial pressure
• > renal vein pressure
• Decreased GFR, increased UP
• Decreased SV, CO → BP
• Increased SVR (SNS, RAAS, ADH V1R). Increased HR
• Decreased renal perfusion pressure, RPF, GFR
• Increased Na and water reabsorption, edema, increased venous return
• Frank-Starling effects on CO
Cardiorenal Syndrome
• Goodpasture/antiGBM
• ANCA + vasculitis
• Other vasculitis
Cryoglobulinemia
SLE
IgA
Pulmonary Renal Syndrome
• Cell lysis-releases DNA
• 1 purine: 2 Phosphates per dinucleotide
• Intratubular CaP
• Intratubular uric acid obstruction
Tumor Lysis Syndrome
Atheroembolic Syndrome
• Postcardiac, aortic cath
• Post-anticoagulation
• CNS encephalopathy
• Cholesterol microemboli to limbs (blue toes, livido reticularis), pancreas, retina
• Eosinophilia
Analgesic Syndrome
• Long hx of aspirin, acetaminophen, phenacetin, or mixed APC
• Gastritis GI symptoms
• Chronic TIN
• Urothelial cancers
Capillary Leak Syndrome
• Increased capillary leak of protein-rich fluid to ISF
• Dx: diffuse edema, shock, non-cardiogenic shock, pulmonary edema
• Causes: sepsis, idiopathic, allergic, OHSS, poisonings, toxins, IL-6, cytokines, burns
• Multi-organ failure, AKI
Thrombotic Microangiopathies
• Low platelets, MAHA, schistocytes
• Low haptoglobin, high LDH
• Fibrin thrombi, endotheliosis, mesangiolysis
• HUS of childhood
• GI and bacterial toxins: Shigatoxin, E. Coli O157:H7 (STEC)
• CNS involvement: TTP-ADAMTS13 antibodies
• Genetic: low complement C3, C4,Factors,H,I
• Antiphospholipid syndromes, ACL, anti-β2GP1,D
• Accelerated HBP, scleroderma
• Associated with SLE
• Pre-eclampsia
• Post-transplant syndromes
• Drugs: mitomycin, calcineurin inhibitors
Uremic Syndrome
• Increased catabolic rate
• Encephalopathy, neuropathy
• Pericarditis
• Fluid overload, pulmonary edema, hypertension
• GI: dysgeusia, anorexia, weight loss, diarrhea
• Dry skin, pruritis

FIGURE 319-4 Categories of systemic or non-renal-based syndromes that involve the kidney and in advanced cases can cause the uremic syndrome requiring renal replacement therapy. ACL, anticardiolipin antibodies; ADH V1R, antidiuretic hormone V1 receptor; AKI, acute kidney injury; ANCA, antineutrophil cytoplasmic antibody; anti-2GP1, anti-β-2-glycoprotein-1; APC, aspirin, phenacetin, caffeine; BP, blood pressure; CaP, calcium phosphate crystals; CNS, central nervous system; CO, cardiac output; GBM, glomerular basement membrane; GFR, glomerular filtration rate; GI, gastrointestinal; HBP, high blood pressure; HR, heart rate; HUS, hemolytic-uremic syndrome; hx, history; IgA, immunoglobulin A; IL, interleukin; ISF, interstitial fluid; LDH, lactate dehydrogenase; MAHA, microangiopathic hemolytic anemia; OHSS, ovarian hyperstimulation syndrome; RAAS, renin-angiotensin-aldosterone system; RPF, renal plasma flow; SLE, systemic lupus erythematosus; SV, stroke volume; SNS, sympathetic nervous system; STEC, Shiga toxin–producing *Escherichia coli*; SVR systemic vascular resistance; TIN, tubulointerstitial nephritis; TTP, thrombotic thrombocytopenic purpura; UP, urine protein.

males may be most severely affected by renal failure in middle age. The syndrome involves hematuria, neurosensory hearing loss, and ocular deformities of the lens. A predominant symptom in hereditary nephritis is microscopic hematuria, and the pathological correlate to this is thin basement membranes. It must be distinguished from the less severe and more commonly related condition known as thin basement membrane disease as a cause of hematuria.

Certain congenital abnormalities of the kidney, such as horseshoe and ectopic kidneys, also shown in Table 319-2, consist of low nephron numbers that may result in secondary FSGS. The condition known as ureteral reflux is an abnormality involving the functional insertion of the ureter into the muscular bladder, which normally closes off the ureter upon bladder contraction. In reflux disease, the ureter cannot prevent retrograde flux of urine during micturition. In children, this may cause hypertension and, in some cases, secondary FSGS. Surgical correction of the abnormality may be curative, but the child may outgrow it in adulthood.

The stages of CKD are listed in Table 319-3 and are central in terms of planning for renal replacement therapy (dialysis and transplantation) and for assessing and slowing the rate of progression, defined as the decline in GFR over time. In CKD, the eGFR is determined by the current stable serum creatinine, which is entered into one of several equations derived from clinical trial data that compare the subject's value to actual measurements of GFR. For example, the Modification of Diet in Renal Disease (MDRD) study used radioisotope (I^{125}-iothalamate) clearance to measure true GFR. The equation

TABLE 319-3 Stages of Chronic Kidney Disease (CKD)

STAGES OF CKD, ABNORMALITY >3 MONTHS	RANGE OF eGFR	CLINICAL FEATURES	IMPORTANCE
1	>60 mL/min/1.73 m^2	Abnormal urinalysis, abnormal renal imaging	Risk of progression to later stages increases with increased proteinuria and dependent on cause of disease
2	>60 mL/min/1.73 m^2	Abnormal urinalysis, abnormal renal imaging	Mild risk of progression to later stages increases with increased proteinuria and dependent on cause of disease
3a	45–60 mL/min/1.73 m^2	Cardiovascular disease or other organ damage	Moderate risk of progression of disease. Pay attention to other vascular risk factors, high BP, lipids, smoking, weight
3b	30–45 mL/min/1.73 m^2	Proteinuria	High risk of progression
4	15–30 mL/min/1.73 m^2		High likelihood of progression to ESRD, need preparation and education regarding choices for RRT including transplant and dialysis
5	<15 mL/min/1.73 m^2		Highest risk of requiring RRT

Abbreviations: BP, blood pressure; eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; RRT, renal replacement therapy.

contains additional factors such as age and sex to derive the current eGFR, corrected for 1.73 m^2 body surface area. This value of eGFR is not a creatinine clearance or a measurement of GFR.

There has been debate about the inclusion of corrective factors that would alter the result from that equation based on the patient's ethnicity and race. For example, it has been recommended in some centers to multiply the eGFR value by a factor of 1.2 in African-American patients in order to correct for the perceived underestimate of GFR using the MDRD equation. There is considerable objection to making this correction, which may have the effect of introducing bias based on race. Because the corrected eGFR will give a higher value to the GFR, there could be delays in discussion of renal replacement therapy and in initiation of treatment. Consequently, the correction has fallen out of favor. Recently, an alternative, the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI), has been used that excludes race as a parameter. Another method of estimating GFR uses cystatin-C, a known filtration marker for GFR. Levels of the patient's current cystatin-C are compared to historical values of cystatin-C from populations where the true GFR was the comparator. Both creatinine and cystatin-C methods are susceptible to error because they do not take into account renal and extrarenal clearance or, in the case of creatinine, muscle mass.

During the course of CKD, the clinician must evaluate the patient for symptoms and signs of worsening renal function. When the eGFR is <15–20 mL/min, the patient should already have had conversations about choices of eventual treatment and, if indicated, surgical preparation for peritoneal dialysis or hemodialysis. Attention should be given to the sensitive topic of transition to a lifesaving but mechanical procedure. It is also important to discuss transplantation and issues of timing. The ethical standard is for separate clinicians to evaluate the recipient and donor of a kidney transplant. Confidentiality and privacy are essential. Importantly, access to transplantation has not been offered equally to minority populations, and it is incumbent on clinicians to avoid bias by discussing the option with every patient.

Symptoms of CKD are often nonspecific and include fatigue, weakness, loss of appetite and taste, weight loss, mood changes, and metabolic encephalopathy that may involve cognitive changes such as decline in executive function and the ability to calculate and remember. Other symptoms are peripheral and autonomic neuropathies and sleep and movement disorders such as restless legs and asterixis or myoclonus. Patients with CKD lose their ability to concentrate urine and, therefore, have an obligate urine output of ~2 L. This results in nocturia since the bladder fills during the night. Ironically, the loss of nocturia may be sensed by the patient as an improvement of symptoms but actually indicates worsening oliguria. Pruritus commonly occurs. Clinical issues include changes in medications and doses. The full uremic syndrome is characterized by severe hypertension, the above-mentioned encephalopathic manifestations, gastrointestinal bleeding, pericarditis, severe electrolyte disturbances, particularly hyperkalemia, and secondary hyperparathyroidism. Anemia is frequently present, and although it may be due to iron deficiency or decreased erythropoietin, other causes of anemia should be investigated. Patients respond well to erythropoietic stimulating agents, with the goal of therapy >10 g/dL but <12 g/dL. Notably, not all patients are anemic. For example, some patients with UTO, PCKD, renal vascular disease, and renal cell carcinoma may in fact have erythrocytosis.

In NS, heavy proteinuria may persist throughout the course of CKD and carries a worse prognosis. In all kidney diseases, including the nephritic processes, renal disease may progress to late stages even when the acute inflammation has subsided; however, the underlying mechanisms of progressive renal failure involve inflammatory pathways (reflected in elevated levels of C-reactive protein and erythrocyte sedimentation rate in CKD), emphasizing the importance of managing risk factors for cardiovascular disease, such as hypertension, smoking, hyperlipidemias, diabetes mellitus, and obesity, early in the course of CKD.

Some chronic renal syndromes may present with predominant tubular dysfunction, including genetic diseases known as channelopathies. Examples are the depletion of electrolytes such as phosphorus and potassium and proximal RTA due to Fanconi's syndrome, mentioned above; hypokalemic alkalosis and volume depletion in Bartter's syndrome (mutations resulting in abnormal loop of Henle function); and electrolyte losses in Gitelman's syndrome, a defect in the thiazide-sensitive sodium-chloride transporter. Gitelman's syndrome also causes hypochloremic hypokalemic metabolic alkalosis. Bartter's and Gitelman's syndromes can be distinguished by examining the urine for the usual solutes regulated by these transporters. Thus, Bartter's syndrome is associated with hypercalciuria and the inability to put out a concentrated urine, whereas Gitelman's syndrome is associated with hypocalciuria and preserved ability to concentrate the urine. Fanconi's syndrome results in hypokalemic acidosis and can be distinguished from distal RTA, which is not associated with the loss of glucose, amino acids, and phosphate in the urine.

Disorders involving the collecting duct and its transporters of sodium, potassium, and acid base can result in syndromes that also suggest particular diagnoses. For example, polyuria due to inappropriate free water losses and insensitive to ADH is characteristic of nephrogenic diabetes insipidus, and it is common to see simultaneous defects in potassium secretion and hydrogen ion secretion with hypertension suggesting blockade of one of the sodium regulators of the epithelial sodium channel. These disorders are usually associated with hyporenin and/or hypoaldosterone states, which include UTO, type 4 RTAs in DM, amyloidosis, or Addison's disease. In contrast to this, hypokalemic metabolic alkalosis with volume expansion may suggest an adrenal adenoma, unilateral renal artery stenosis, adrenocorticotropic hormone–secreting tumors, licorice abuse, or potassium-sparing diuretics. Although genomic testing may reveal gene mutations characteristic of a disorder, such testing should not be automatic since many mutations require the presence of other high-risk conditions (a "second hit") or are only indirectly associated with the development of disease.

RENAL MASSES

Patients with renal cell carcinoma may have been diagnosed via an incidental finding in an abdominal imaging study or sometimes by a palpable mass best felt with the patient supine, usually in a thin patient but occasionally due to a very large protruding mass. Very large masses or multiple masses that are easily palpable may represent cystic diseases of the kidney, including PCKD or even a single cyst, versus a congenital ureteral pelvic obstruction. Other times, a renal cell carcinoma can present as anemia, possibly caused by hematuria, or as back pain associated with metastatic lytic vertebral lesions. Metastases may involve the lungs and bone marrow as well.

IMAGING AND RENAL BIOPSY INDICATIONS

For hematuric syndromes, imaging may provide valuable information, particularly in a patient who has heavy bleeding or blood clots in the urine. Renal pathology may be detected as an abdominal mass, as in the setting of renal cell carcinoma, chronic UTO, or cystic diseases of the kidney including PCKD and simple cyst. If the patient has a known history of tubular sclerosis or the finding of skin fibroadenoma, one might identify a renal mass found on CT imaging as an angiomyolipoma.

The renal ultrasound is efficacious in determining the size and symmetry of the kidneys and in excluding urinary obstruction. It is useful in detecting renal cysts or masses but less effective in kidney stone disease. Ultrasound is not as accurate a tool as a computed tomography (CT) scan for angiomyolipomas. The renal-limited noncontrast CT scan is the standard test for nephrolithiasis but carries the risk of accumulative radiation. Magnetic resonance imaging (MRI) is often useful in evaluating and following renal masses, including renal cell carcinoma. The patient with renal disease may develop a toxic complication of systemic sclerosis after receiving multiple gadolinium studies for MRI enhancement; new contrast agents to replace it are emerging. CT scans with high-osmolality iodinated contrast media, administered in large volumes, remain an important cause of AKI (contrast nephropathy) in elderly male patients with vascular disease of the kidney, multiple myeloma, hepatic disease, extracellular fluid depletion, DM, or concurrent use of NSAIDs. The accompanied radiation is a concern for radiation exposure.

Radioisotope scanning is useful in demonstrating the percentage of renal function coming from each kidney. Finally, in many of the diseases discussed above, diagnosis ultimately depends on renal biopsy and pathologic evaluation.

FURTHER READING

Bhosale SJ, Kulkarni AP: Biomarkers in acute kidney injury. Indian J Crit Care Med 24:S90, 2020.

Canki E et al: Urinary biomarkers in kidney disease. Clin Chim Acta 55:117798, 2024.

Glassock RJ: Kidney biopsy is required for nephrotic syndrome with PLA2R$^+$ and normal kidney function: Commentary. Kidney 360:894, 2020.

Levey AS et al: Nomenclature for kidney function and disease: Report of Kidney Disease: Improving Global Outcomes (KDIGO) Consensus Conference. Kidney Int 97:1117, 2020.

Liang Y et al: Serum anti-phospholipase A2 receptor (PLA2R) antibody detected at diagnosis as a predictor for clinical remission in patients with primary membranous nephropathy: A meta-analysis. BMC Nephrol 20:360, 2019.

Pollak MR, Friedman DJ: APOL1 and APOL1-associated kidney disease: A common disease, an unusual disease gene—Proceedings of the Henry Shavelle Professorship. Glomerular Dis 3:75, 2023.

Ponticelli C et al: C3 glomerulopathies: Dense deposit disease and C3 glomerulonephritis. Front Med (Lausanne) 24:1289812, 2023.

Romagnani P et al: The five types of glomerulonephritis classified by pathogenesis, activity and chronicity (GN-AC). Nephrol Dial Transplant 38:ii3, 2023.

Rysz J et al: Novel biomarkers in the diagnosis of chronic kidney disease and the prediction of its outcome. Int J Mol Sci 18:1702, 2017.

Vijay P et al: Cystatin C: Best biomarker for acute kidney injury and estimation of glomerular filtration rate in childhood cirrhosis. Eur J Pediatr 180:3287, 2021.

Vink CH et al: Antibody-guided therapy in phospholipase A-2 receptor associated membranous nephropathy. Clin Res 8:432, 2023.

320

Cell Biology and Physiology of the Kidney

Alfred L. George, Jr., Eric G. Neilson

The kidney is one of the most highly differentiated organs in the body. At the conclusion of embryologic development, nearly 30 different cell types form a multitude of filtering capillaries and segmented nephrons enveloped by a dynamic interstitium. This cellular diversity modulates a variety of complex physiologic processes. Endocrine functions, the regulation of blood pressure and intraglomerular hemodynamics, solute and water transport, acid-base balance, and removal of drug metabolites are all accomplished by intricate mechanisms of renal response. This breadth of physiology hinges on the clever ingenuity of nephron architecture that evolved as complex organisms came out of water to live on land.

EMBRYOLOGIC DEVELOPMENT

Kidneys develop from intermediate mesoderm under the timed or sequential control of a growing number of genes, described in **Fig. 320-1**. The transcription of these genes are determined or guided by morphogenic cues, the orientation of cilia-derived planar cell polarity, and the generational legacy of epigenetic marks that invite two ureteric buds to each penetrate bilateral metanephric blastema, where they induce primary mesenchymal cells to form early nephrons. The two ureteric buds emerge from posterior nephric ducts and mature into separate collecting systems that eventually form a renal pelvis and ureter. Induced mesenchyme undergoes mesenchymal epithelial transitions to form comma-shaped bodies at the proximal end of each ureteric bud leading to the formation of S-shaped nephrons that cleft and enjoin with penetrating endothelial cells derived from sprouting angioblasts. Under the influence of vascular endothelial growth factor A (VEGF-A), these penetrating cells form capillaries with surrounding mesangial cells that differentiate into a glomerular filter for plasma water and solute. The ureteric buds branch, and each branch produces a new set of nephrons. The number of branching events ultimately determines the total number of nephrons in each kidney. There are ~900,000 glomeruli in each kidney in normal-birth-weight adults and as few as 225,000 in low-birth-weight adults, with the latter producing numerous comorbid risks.

Glomeruli evolve as complex capillary filters with fenestrated endothelia under the guiding influence of VEGF-A and angiopoietin-1 secreted by adjacently developing podocytes. Epithelial podocytes facing the urinary space envelop the exterior basement membrane supporting these emerging endothelial capillaries. Podocytes are partially polarized and periodically slough into the urinary space by epithelial-mesenchymal transition and, to a lesser extent, apoptosis, only to be replenished by migrating parietal epithelia from Bowman capsule. Impaired replenishment results in heavy proteinuria. Podocytes attach to the basement membrane by special foot processes and share a slit-pore membrane with their neighbor. The slit-pore membrane forms a filter for plasma water and solute by the synthetic interaction of nephrin, annexin-4, CD2AP, FAT, ZO-1, P-cadherin, podocin, TRPC6, PLCE1, and Neph 1-3 proteins. Mutations in many of these proteins also result in heavy proteinuria. The glomerular capillaries are embedded in a mesangial matrix shrouded by parietal and proximal tubular epithelia forming Bowman capsule. Mesangial cells have an embryonic lineage consistent with arteriolar or juxtaglomerular cells and contain contractile actin-myosin fibers. These mesangial cells make contact with glomerular capillary loops, and their local matrix holds them in condensed arrangement.

Between nephrons lies the renal interstitium. This region forms a functional space surrounding glomeruli and their downstream tubules, which are home to resident and trafficking cells such as fibroblasts, dendritic cells, occasional lymphocytes, and lipid-laden macrophages. The cortical and medullary peritubular capillaries, which siphon off

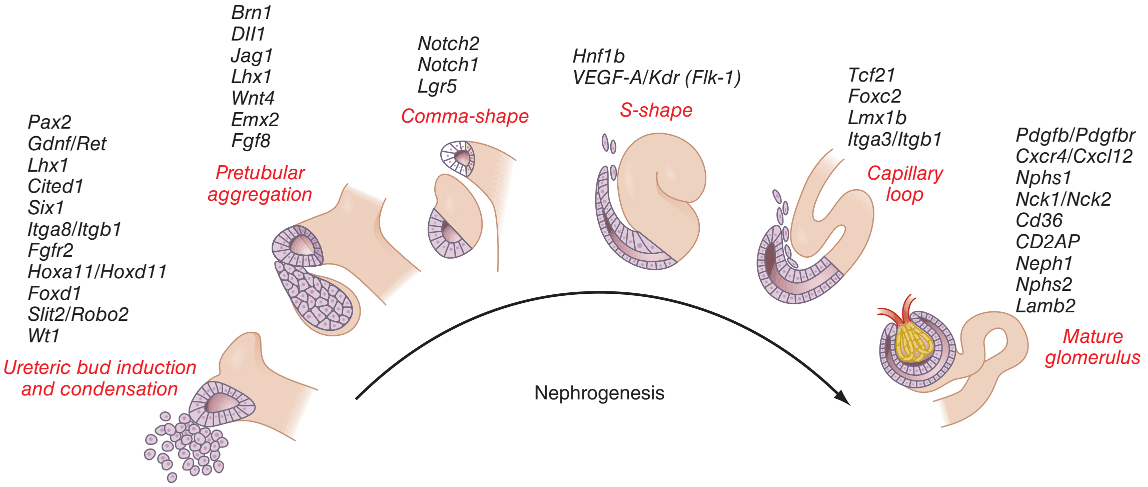

FIGURE 320-1 Genes controlling renal nephrogenesis. A growing number of genes have been identified at various stages of glomerulotubular development in the mammalian kidney. The genes listed have been tested in various genetically modified mice, and their location corresponds to the classical stages of kidney development postulated by Saxen in 1987.

solute and water following tubular reclamation of glomerular filtrate, are also part of the interstitial fabric as well as a web of connective tissue that supports the kidney's emblematic architecture of folding tubules. The relational precision of these structures determines the unique physiology of the kidney.

Each nephron is partitioned during embryologic development into a proximal tubule, descending and ascending limbs of the loop of Henle, distal tubule, and the collecting duct. These classic tubular segments build from subsegments lined by highly unique epithelia serving regional physiology. All nephrons have the same structural components, but there are two types whose structures depend on their location within the kidney. The majority of nephrons are cortical, with glomeruli located in the mid-to-outer cortex. Fewer nephrons are juxtamedullary, with glomeruli at the boundary of the cortex and outer medulla. Cortical nephrons have short loops of Henle, whereas juxtamedullary nephrons have long loops of Henle. There are critical differences in blood supply as well. The peritubular capillaries surrounding cortical nephrons are shared among adjacent nephrons. By contrast, juxtamedullary nephrons depend on individual capillaries called *vasa recta* that run alongside the long loops of Henle. Cortical nephrons perform most of the glomerular filtration because there are more of them and because their afferent arterioles are larger than their respective efferent arterioles. The juxtamedullary nephrons, with longer loops of Henle, create an osmotic gradient for concentrating urine. How developmental instructions specify the differentiation of all these unique epithelia among various tubular segments is still unknown.

DETERMINANTS AND REGULATION OF GLOMERULAR FILTRATION

Renal blood flow normally drains ~20% of the cardiac output, or 1000 mL/min. Blood reaches each nephron through the afferent arteriole leading into a glomerular capillary where ultrafiltration forms the tubular fluid. The distal ends of the glomerular capillaries coalesce to form an efferent arteriole leading to the first segment of a second capillary network (cortical peritubular capillaries or medullary vasa recta) surrounding the tubules (Fig. 320-2*A*). Thus, nephrons have two capillary beds arranged in a series separated by the efferent arteriole that regulates the hydrostatic pressure in both capillary beds. The distal capillaries empty into small venous branches that coalesce into larger veins to eventually form the renal vein.

The hydrostatic pressure gradient across the glomerular capillary wall is the primary driving force for glomerular filtration. Oncotic pressure within the capillary lumen, determined by the concentration of unfiltered plasma proteins, partially offsets the hydrostatic pressure gradient and opposes filtration. As the oncotic pressure rises along the length of the glomerular capillary, the driving force for filtration falls to zero en route to the efferent arteriole. Approximately 20% of the renal plasma flow is filtered into Bowman space, and the ratio of glomerular filtration rate (GFR) to renal plasma flow determines the filtration fraction. Several factors, mostly hemodynamic, contribute to the regulation of filtration under physiologic conditions.

Although glomerular filtration is affected by renal artery pressure, this relationship is not linear across the range of physiologic blood pressures due to autoregulation of GFR. Autoregulation of glomerular filtration is the result of three major factors that modulate either afferent or efferent arteriolar tone; these include an autonomous vasoreactive (myogenic) reflex in the afferent arteriole, *tubuloglomerular feedback* (TGF), and angiotensin II–mediated vasoconstriction of the efferent arteriole. The myogenic reflex is a first line of defense against fluctuations in renal blood flow. Acute changes in renal perfusion pressure evoke reflex constriction or dilatation of the afferent arteriole in response to rising or falling pressure, respectively. This phenomenon helps protect the glomerular capillary from sudden changes in systolic pressure.

TGF changes the rate of filtration and tubular flow by reflex vasoconstriction or dilatation of the afferent arteriole. TGF is mediated by specialized cells in the thick ascending limb of the loop of Henle called the *macula densa* that act as sensors of solute concentration and tubular fluid flow rate. With high tubular flow rates, a proxy for an inappropriately high filtration rate, greater solute delivery to the macula densa (Fig. 320-2*B*) evokes vasoconstriction of the afferent arteriole, causing GFR to return toward normal. One component of the soluble signal from the macula densa is adenosine triphosphate (ATP) released by the cells during increased NaCl reabsorption. ATP is metabolized in the extracellular space to generate adenosine, a potent vasoconstrictor of the afferent arteriole. During conditions associated with a fall in filtration rate, a lower rate of solute delivery to the macula densa attenuates TGF, allowing afferent arteriolar dilatation and restoring GFR to normal levels. Angiotensin II and reactive oxygen species enhance TGF, whereas nitric oxide (NO) blunts TGF. A distinct feedback mechanism may exist between the connecting tubule and GFR in which high Na^+ delivery evokes afferent arteriolar dilation possibly mediated by prostaglandins.

The third component underlying autoregulation of GFR involves angiotensin II. During states of reduced renal blood flow, renin is released from granular cells within the wall of the afferent arteriole near the macula densa in a region called the juxtaglomerular apparatus (Fig. 320-2*B*). Renin, a proteolytic enzyme, catalyzes the conversion of angiotensinogen to angiotensin I, which is subsequently converted to angiotensin II by angiotensin-converting enzyme (ACE) (Fig. 320-2*C*). Angiotensin II evokes vasoconstriction of the efferent arteriole, and the resulting increased glomerular hydrostatic pressure elevates GFR to normal levels.

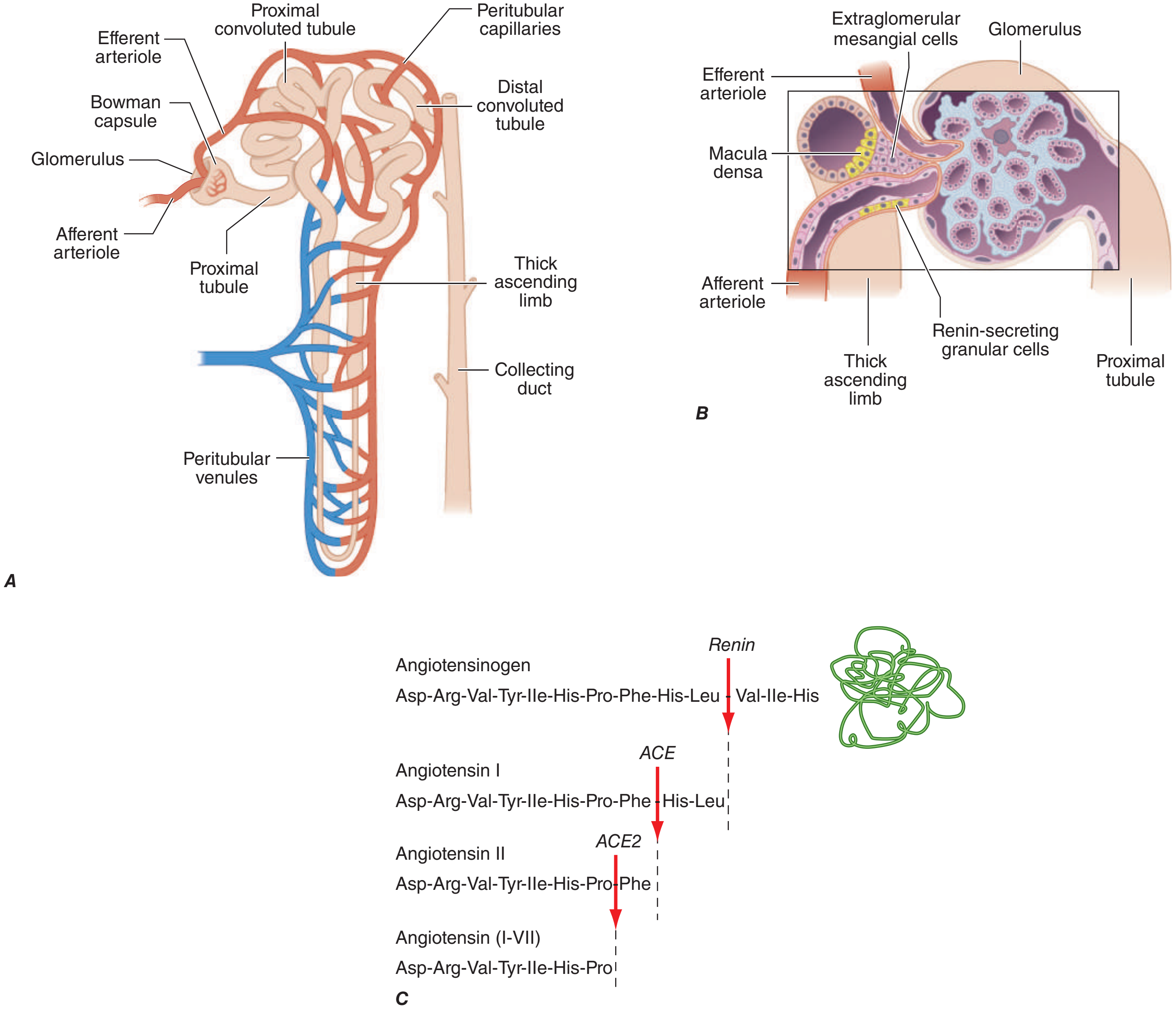
Proximal convoluted tubule
Peritubular capillaries
Efferent arteriole
Bowman capsule
Distal convoluted tubule
Glomerulus
Afferent arteriole
Proximal tubule
Thick ascending limb
Collecting duct
Peritubular venules
A
Extraglomerular mesangial cells
Glomerulus
Efferent arteriole
Macula densa
Afferent arteriole
Thick ascending limb
Renin-secreting granular cells
Proximal tubule
B
Renin
Angiotensinogen
Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu Val-Ile-His
ACE
Angiotensin I
Asp-Arg-Val-Tyr-Ile-His-Pro-Phe His-Leu
ACE2
Angiotensin II
Asp-Arg-Val-Tyr-Ile-His-Pro Phe
Angiotensin (I-VII)
Asp-Arg-Val-Tyr-Ile-His-Pro
C

FIGURE 320-2 Renal microcirculation and the renin-angiotensin system. ***A.*** Diagram illustrating relationships of the nephron with glomerular and peritubular capillaries. ***B.*** Expanded view of the glomerulus with its juxtaglomerular apparatus including the macula densa and adjacent afferent arteriole. ***C.*** Proteolytic processing steps in the generation of angiotensins.

MECHANISMS OF RENAL TUBULAR TRANSPORT

The renal tubules are composed of highly differentiated epithelia that vary in morphology and function along the nephron (Fig. 320-3). The cells lining the various tubular segments form monolayers connected to one another by a specialized region of the adjacent lateral membranes called the *tight junction*. Tight junctions form an occlusive barrier that separates the lumen of the tubule from the interstitial spaces surrounding the tubule and also apportions the cell membrane into discrete domains: the apical membrane facing the tubular lumen and the basolateral membrane facing the interstitium. This regionalization allows cells to allocate membrane proteins and lipids asymmetrically. Owing to this feature, renal epithelial cells are said to be *polarized*. The asymmetric assignment of membrane proteins, especially proteins mediating transport processes, provides the machinery for directional movement of fluid and solutes by the nephron.

■ EPITHELIAL SOLUTE TRANSPORT

There are two types of epithelial transport. Movement of fluid and solutes sequentially across the apical and basolateral cell membranes (or vice versa) mediated by transporters, channels, or pumps is called *cellular transport*. By contrast, movement of fluid and solutes through the narrow passageway between adjacent cells is called *paracellular transport*. Paracellular transport occurs through tight junctions, indicating that they are not completely "tight" or occlusive. Indeed, some epithelial cell layers allow rather robust paracellular transport to occur (*leaky epithelia*), whereas other epithelia have more restrictive tight junctions (*tight epithelia*). In addition, because the ability of ions to flow through the paracellular pathway determines the electrical resistance across the epithelial monolayer, leaky and tight epithelia are also referred to as low- or high-resistance epithelia, respectively. The proximal tubule contains leaky epithelia, whereas distal nephron segments, such as the collecting duct, contain tight epithelia. Leaky epithelia are most well suited for bulk fluid reabsorption, whereas tight epithelia allow for more refined control and regulation of transport.

■ MEMBRANE TRANSPORT

Cell membranes are composed of hydrophobic lipids that repel water and aqueous solutes. The movement of solutes and water across cell membranes is made possible by discrete classes of integral membrane proteins, including channels, pumps, and transporters. These different mechanisms mediate specific types of transport activities, including *active transport* (pumps), *passive transport* (channels), *facilitated diffusion* (transporters), and *secondary active transport* (cotransporters). Active transport requires metabolic energy generated by the hydrolysis of ATP. Active transport pumps are ion-translocating ATPases, including the ubiquitous Na^+/K^+-ATPase, the H^+-ATPases, and Ca^{2+}-ATPases. Active transport creates asymmetric ion concentrations across a cell membrane and can move ions against a chemical gradient. The potential energy stored in a concentration gradient of an ion such as Na^+ can be used to drive transport through other mechanisms (secondary active transport). The movement of solutes through a membrane protein by simple diffusion is called passive transport. This activity is mediated by channels created by selectively permeable membrane proteins, and it allows solute or water to move across a membrane driven by favorable *concentration gradients* or *electrochemical potential*. Facilitated diffusion is a specialized type of passive transport mediated by simple transporters called *carriers* or *uniporters*. For example, hexose transporters such as GLUT2 mediate glucose transport by tubular cells. These transporters are driven by the concentration gradient for glucose that is highest in extracellular fluids and lowest in the cytoplasm due to rapid metabolism. Many other transporters operate by translocating two or more ions/solutes in concert either in the same direction (*symporters* or *cotransporters*) or in opposite directions (*antiporters* or *exchangers*) across the cell membrane. The movement of two or more ions/solutes may produce no net change in the balance of electrostatic charges across the membrane (*electroneutral*), or a transport event may alter the balance of charges (*electrogenic*). Several inherited disorders of renal tubular solute and water transport occur as a consequence of mutations in genes encoding a variety of channels, transporter proteins, and their regulators **(Table 320-1)**.

The rate of tubular fluid flow through the nephron is highly variable, for which reason ion transport needs to be adjusted accordingly. Renal epithelial cells have the ability to monitor tubular fluid flow by mechanisms involving cellular structures (apical microvilli, primary cilium) or mechanosensing ion channels. Tubular transport is regulated in response to changes in fluid flow through changes in intracellular Ca^{2+} or by purinergic receptor-mediated signal transduction.

SEGMENTAL NEPHRON FUNCTIONS

Each anatomic segment of the nephron has unique characteristics and specialized functions enabling selective transport of solutes and water **(Fig. 320-3*A*)**. Through sequential events of reabsorption and secretion along the nephron, tubular fluid is progressively conditioned into urine. Knowledge of the major tubular mechanisms responsible for solute and water transport is critical for understanding hormonal regulation of kidney function and the pharmacologic manipulation of renal excretion.

PROXIMAL TUBULE

The proximal tubule is responsible for reabsorbing ~60% of filtered NaCl and water, as well as ~90% of filtered bicarbonate and most critical nutrients such as glucose and amino acids. The proximal tubule uses both cellular and paracellular transport mechanisms. The apical membrane of proximal tubular cells has an expanded surface area available for reabsorption created by a dense array of microvilli called the *brush border*, and leaky tight junctions enable high-capacity fluid reabsorption.

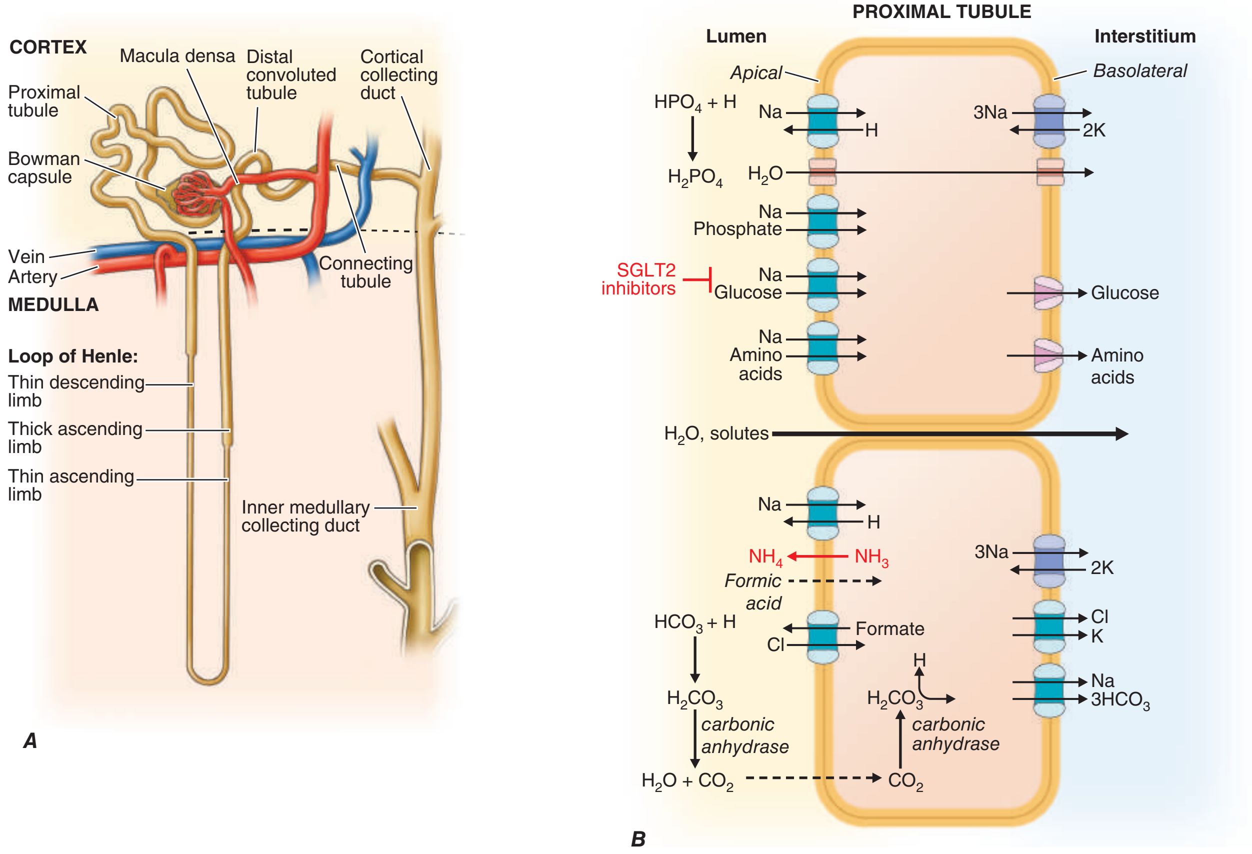
CORTEX
Macula densa
Distal convoluted tubule
Cortical collecting duct
Proximal tubule
Bowman capsule
Vein
Artery
Connecting tubule
MEDULLA
Loop of Henle:
Thin descending limb
Thick ascending limb
Thin ascending limb
Inner medullary collecting duct
A
PROXIMAL TUBULE
Lumen
Interstitium
Apical
Basolateral
HPO_4 + H
H_2PO_4
Na
H
3Na
2K
H_2O
Na
Phosphate
SGLT2 inhibitors
Na
Glucose
Glucose
Na
Amino acids
Amino acids
H_2O, solutes
Na
H
NH_4
NH_3
3Na
2K
Formic acid
HCO_3 + H
Formate
Cl
Cl
K
H
Na
$3HCO_3$
H_2CO_3
carbonic anhydrase
H_2CO_3
carbonic anhydrase
H_2O + CO_2
CO_2
B

FIGURE 320-3 Transport activities of the major nephron segments. Representative cells from five major tubular segments are illustrated with the lumen side (apical membrane) facing left and interstitial side (basolateral membrane) facing right. ***A.*** Overview of entire nephron. ***B.*** Proximal tubular cells. ***C.*** Typical cell in the thick ascending limb of the loop of Henle. ***D.*** Distal convoluted tubular cell. ***E.*** Cortical collecting duct principal cell. ***F.*** Cortical collecting duct type A and type B intercalated cells. ***G.*** Typical cell in the inner medullary collecting duct. The major membrane transporters, channels, and pumps are drawn with *arrows* indicating the direction of solute or water movement. For some events, the stoichiometry of transport is indicated by numerals preceding the solute. Targets for major diuretic agents are labeled. The actions of hormones are illustrated by *arrows with plus signs* for stimulatory effects and *lines with perpendicular ends* for inhibitory events. The *dashed line* indicates water impermeability of cell membranes in the thick ascending limb and distal convoluted tubule.

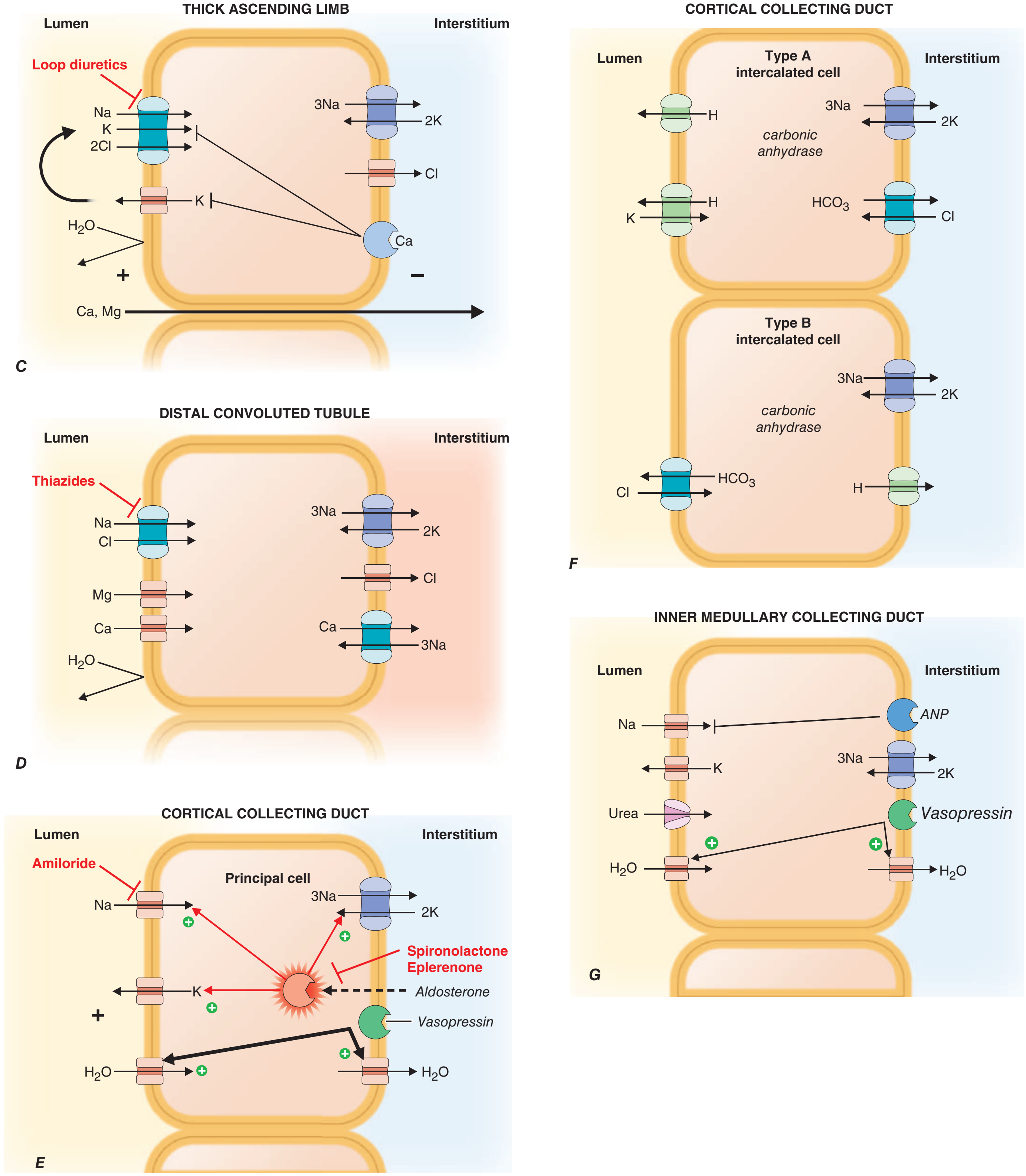
THICK ASCENDING LIMB
Lumen
Interstitium
Loop diuretics
Na
K
2Cl
3Na
2K
Cl
K
Ca
H_2O
+
−
Ca, Mg
C
DISTAL CONVOLUTED TUBULE
Lumen
Interstitium
Thiazides
Na
Cl
3Na
2K
Cl
Mg
Ca
Ca
3Na
H_2O
D
CORTICAL COLLECTING DUCT
Lumen
Interstitium
Amiloride
Principal cell
Na
3Na
2K
Spironolactone
Eplerenone
K
Aldosterone
+
Vasopressin
H_2O
H_2O
E
CORTICAL COLLECTING DUCT
Lumen
Type A intercalated cell
Interstitium
H
3Na
2K
carbonic anhydrase
K
H
HCO_3
Cl
Type B intercalated cell
3Na
2K
carbonic anhydrase
Cl
HCO_3
H
F
INNER MEDULLARY COLLECTING DUCT
Lumen
Interstitium
Na
ANP
K
3Na
2K
Urea
Vasopressin
H_2O
H_2O
G

FIGURE 320-3 (Continued)

Solute and water pass through these tight junctions to enter the lateral intercellular space where absorption by the peritubular capillaries occurs. Bulk fluid reabsorption by the proximal tubule is driven by high oncotic pressure and low hydrostatic pressure within the peritubular capillaries. Cellular transport of most solutes by the proximal tubule is coupled to the Na^+ concentration gradient established by the activity of a basolateral Na^+/K^+-ATPase **(Fig. 320-3*B*)**. This active transport mechanism maintains a steep Na^+ gradient by keeping intracellular Na^+ concentrations low. Solute reabsorption from the tubular lumen is coupled to the Na^+ gradient by Na^+-dependent transporters such as Na^+-glucose and Na^+-phosphate cotransporters present in apical membranes. In addition to the paracellular route, water reabsorption also occurs through the cellular pathway enabled by constitutively active water channels (aquaporin-1) present on both apical and basolateral membranes.

Proximal tubular cells reclaim nearly all filtered bicarbonate by a mechanism dependent on carbonic anhydrases. Filtered bicarbonate is first titrated by protons delivered to the lumen mainly by apical

TABLE 320-1 Inherited Disorders Affecting Renal Tubular Ion and Solute Transport

DISEASE OR SYNDROME	PROTEIN (*GENE*)	OMIM[a]
Disorders Involving the Proximal Tubule		
Proximal renal tubular acidosis	Sodium bicarbonate cotransporter (*SLC4A4*, 4q21)	604278
Fanconi-Bickel syndrome	Glucose transporter, GLUT2 (*SLC2A2*, 3q26.2)	227810
Isolated renal glycosuria	Sodium glucose cotransporter (*SLC5A2*, 16p11.2)	233100
Cystinuria		
Type I	Cystine, dibasic and neutral amino acid transporter (*SLC3A1*, 2p16.3)	220100
Non-type I	Amino acid transporter, light subunit (*SLC7A9*, 19q13.1)	600918
Lysinuric protein intolerance	Amino acid transporter (*SLC7A7*, 4q11.2)	222700
Dicarboxylic aminoaciduria	Glutamate transporter (*SLC1A1*, 9q24.2)	222730
Hartnup disorder	Neutral amino acid transporter (*SLC6A19*, 5p15.33)	34500
Hypophosphatemic nephrolithiasis/osteoporosis 1	Sodium phosphate cotransporter (*SLC34A1, 5q35.3)*	612286
Hereditary hypophosphatemic rickets with hypercalcemia	Sodium phosphate cotransporter (*SLC34A3*, 9q34)	241530
Renal hypouricemia		
Type 1	Urate-anion exchanger (*SLC22A12*, 11q13)	220150
Type 2	Urate transporter, GLUT9 (*SLC2A9*, 4p16.1)	612076
Dent's disease	Chloride channel, ClC-5 (*CLCN5*, Xp11.22)	300009
X-linked recessive nephrolithiasis with renal failure	Chloride channel, ClC-5 (*CLCN5*, Xp11.22)	310468
X-linked recessive hypophosphatemic rickets	Chloride channel, ClC-5 (*CLCN5*, Xp11.22)	307800
Disorders Involving the Loop of Henle		
Bartter's syndrome		
Type 1	Sodium, potassium chloride cotransporter (*SLC12A1*, 15q21.1)	241200
Type 2	Potassium channel, ROMK (*KCNJ1*, 11q24)	601678
Type 3	Chloride channel, ClC-Kb (*CLCNKB*, 1p36)	602023
with sensorineural deafness	Chloride channel accessory subunit, Barttin (*BSND*, 1p31)	602522
Autosomal dominant hypocalcemia with Bartter-like syndrome	Calcium-sensing receptor (*CASR*, 3q13.33)	601199
Familial hypocalciuric hypercalcemia	Calcium-sensing receptor (*CASR*, 3q13.33)	145980
Familial hypomagnesemia type 3	Claudin-16 (*CLDN16*, 3q27)	248250
Familial hypomagnesemia type 5	Claudin-19 (*CLDN19*, 1p34.2)	603959
Isolated renal magnesium loss	Sodium potassium ATPase, γ_1-subunit (*ATP1G1*, 11q23)	154020
Disorders Involving the Distal Tubule and Collecting Duct		
Gitelman syndrome	Sodium chloride cotransporter (*SLC12A3*, 16q13)	263800
Primary hypomagnesemia with secondary hypocalcemia	Melastatin-related transient receptor potential cation channel 6 (*TRPM6*, 9q22)	602014
Pseudoaldosteronism (Liddle's syndrome)	Epithelial sodium channel β and γ subunits (*SCNN1B*, *SCNN1G*, 16p12.1)	177200
Recessive pseudohypoaldosteronism type 1	Epithelial sodium channel, α, β, and γ subunits (*SCNN1A*, 12p13; *SCNN1B*, *SCNN1G*, 16pp12.1)	264350
Pseudohypoaldosteronism type 2 (Gordon's hyperkalemia-hypertension syndrome)	Kinases WNK-1, WNK-4 (*WNK1*, 12p13; *WNK4*, 17q21.31)	145260
X-linked nephrogenic diabetes insipidus	Vasopressin V2 receptor (AVPR2, Xq28)	304800
EAST/SeSAME syndrome	Potassium channel Kir4.1 (*KCNJ10*, 1q23.2)	612780
Nephrogenic diabetes insipidus (autosomal)	Water channel, aquaporin-2 (*AQP2*, 12q13)	125800
Distal renal tubular acidosis		
autosomal dominant	Anion exchanger-1 (*SLC4A1*, 17q21.31)	179800
autosomal recessive	Anion exchanger-1 (*SLC4A1*, 17q21.31)	602722
with neural deafness	Proton ATPase, β1 subunit (*ATP6V1B1*, 2p13.3)	192132
with normal hearing	Proton ATPase, 116-kD subunit (*ATP6V0A4*, 7q34)	602722

[a] Online Mendelian Inheritance in Man database (*https://www.ncbi.nlm.nih.gov/omim*).

membrane Na^+/H^+ exchange. The resulting carbonic acid (H_2CO_3) is metabolized by brush border carbonic anhydrase to water and carbon dioxide. Dissolved carbon dioxide then diffuses into the cell, where it is enzymatically hydrated by cytoplasmic carbonic anhydrase to re-form carbonic acid. Finally, intracellular carbonic acid dissociates into free protons and bicarbonate anions, and bicarbonate exits the cell through a basolateral Na^+/HCO_3^- cotransporter. This process is saturable, which can result in renal bicarbonate excretion when plasma levels exceed the physiologically normal range (24–26 meq/L). Carbonic anhydrase inhibitors such as acetazolamide block proximal tubule bicarbonate reabsorption and are useful for alkalinizing the urine.

The proximal tubule contributes to acid secretion by two mechanisms involving the titration of the urinary buffers ammonia (NH_3) and phosphate. Renal NH_3 is produced by glutamine metabolism in the proximal tubule. Subsequent diffusion of NH_3 out of the proximal tubular cell enables trapping of H^+, which is secreted by apical Na^+/H^+ exchange, in the lumen as ammonium ion (NH_4^+). Cellular K^+ levels inversely modulate proximal tubular ammoniagenesis, and in the setting of high serum

K^+ from hypoaldosteronism, reduced ammoniagenesis promotes type IV renal tubular acidosis. Filtered hydrogen phosphate ion (HPO_4^{2-}) is also titrated in the proximal tubule by secreted H^+ to form $H_2PO_4^-$, and this reaction constitutes a major component of the urinary buffer referred to as titratable acid. Most filtered phosphate ion is reabsorbed by the proximal tubule through a sodium-coupled cotransport process that is regulated by parathyroid hormone (PTH).

Chloride is poorly reabsorbed throughout the first segment of the proximal tubule, and a rise in Cl^- concentration counterbalances the removal of bicarbonate anion from tubular fluid. In later proximal tubular segments, cellular Cl^- reabsorption is initiated by apical exchange of cellular formate for higher luminal concentrations of Cl^-. Once in the lumen, formate anions are titrated by H^+ (provided by Na^+/H^+ exchange) to generate neutral formic acid, which can diffuse passively across the apical membrane back into the cell where it dissociates a proton and is recycled. Basolateral Cl^- exit is mediated by a K^+/Cl^- cotransporter.

Reabsorption of glucose is nearly complete by the end of the proximal tubule. Cellular transport of glucose is mediated by apical Na^+-glucose cotransport coupled with basolateral, facilitated diffusion by a glucose transporter. This process is also saturable, leading to glycosuria when plasma levels exceed 180–200 mg/dL, as seen in untreated diabetes mellitus. Inhibitors of the Na^+-glucose cotransporter SLGT2 in proximal tubules block glucose reabsorption and lower blood glucose, which has therapeutic benefits in diabetes mellitus and chronic diabetic kidney disease.

The proximal tubule possesses specific transporters capable of secreting a variety of organic acids (carboxylate anions) and bases (mostly primary amine cations). Organic anions transported by these systems include several protein-bound drugs not filtered at the glomerulus (penicillins, cephalosporins, salicylates, and others). Organic cations secreted by the proximal tubule include various biogenic amine neurotransmitters (dopamine, acetylcholine, epinephrine, norepinephrine, and histamine) and creatinine. The ATP-dependent transporter P-glycoprotein encoded by *ABCB1* is expressed in brush border membranes and secretes several medically important drugs, including cyclosporine, digoxin, tacrolimus, and various cancer chemotherapeutic agents. Certain drugs such as cimetidine and trimethoprim compete with endogenous compounds for transport by the organic cation pathways. Although these drugs elevate serum creatinine levels, there is no actual change in GFR in this setting.

Calcium and phosphorus homeostasis depends on normal functioning of the proximal tubule. Approximately 60–70% of filtered calcium and ~85% of filtered phosphorus (in the form of inorganic phosphate) are reabsorbed by the proximal tubule. Whereas calcium reabsorption is mostly by passive diffusion through the paracellular route, phosphate reabsorption is mediated by sodium-coupled cotransport. In addition to direct reabsorption, the proximal tubule contributes to systemic mineral balance by participating in specific endocrine pathways. Circulating 25-hydroxy vitamin D (calcidiol) is bioactivated by proximal tubular 1α-hydroxylase to produce 1,25-di-hydroxy vitamin D (calcitriol), the most active form of the hormone, which acts on the small intestine to promote calcium absorption. Phosphate balance is affected by circulating fibroblast growth hormone 23 (FGF23), a bone-derived hormone that interacts with its receptor (FGFR1) and co-receptor (Klotho) on proximal tubular cells to suppress sodium-phosphate cotransport and promote renal phosphate excretion. PTH stimulates proximal tubular 1α-hydroxylation of vitamin D, whereas it suppresses sodium-phosphate cotransport. Derangements in PTH and FGF23 account for abnormal calcium and phosphate balance in chronic kidney disease.

The proximal tubule, through distinct classes of Na^+-dependent and Na^+-independent transport systems, reabsorbs amino acids efficiently. These transporters are specific for different groups of amino acids. For example, cystine, lysine, arginine, and ornithine are transported by a system comprising two proteins encoded by the *SLC3A1* and *SLC7A9* genes. Mutations in either *SLC3A1* or *SLC7A9* impair reabsorption of these amino acids and cause the disease cystinuria. Peptide hormones, such as insulin and growth hormone, β_2-microglobulin, and other small proteins, are taken up by the proximal tubule through a process of absorptive endocytosis and are degraded in acidified endocytic lysosomes. Acidification of these vesicles depends on a multi-subunit vacuolar H^+-ATPase and a separate Cl^-/H^+ exchanger encoded by *CLCN5*. Impaired acidification of endocytic vesicles because of *CLCN5* pathogenic variants causes low-molecular-weight proteinuria in Dent's disease.

LOOP OF HENLE

The loop of Henle consists of three major segments: descending thin limb, ascending thin limb, and ascending thick limb. Approximately 15–25% of filtered NaCl is reabsorbed in the loop of Henle, mainly by the thick ascending limb. The loop of Henle has an important role in urinary concentration by contributing to the generation of a hypertonic medullary interstitium in a process called *countercurrent multiplication*. The loop of Henle is the site of action for the most potent class of diuretic agents (loop diuretics) and also contributes to reabsorption of calcium and magnesium ions.

The descending thin limb is highly water permeable owing to dense expression of constitutively active aquaporin-1 water channels. By contrast, water permeability is negligible in the ascending thin and thick limbs. In the thick ascending limb, there is a high level of secondary active NaCl transport enabled by the $Na^+/K^+/2Cl^-$ cotransporter on the apical membrane in series with basolateral Cl^- channels and Na^+/K^+-ATPase **(Fig. 320-3C)**. The $Na^+/K^+/2Cl^-$ cotransporter is the molecular target for loop diuretics. Tubular fluid K^+ is the limiting substrate for this cotransporter (tubular concentration of K^+ is similar to plasma, ~4 meq/L), but transporter activity is maintained by K^+ recycling through an apical potassium channel. The cotransporter also enables reabsorption of NH_4^+ in lieu of K^+, and this leads to accumulation of both NH_4^+ and NH_3 in the medullary interstitium. An inherited disorder of the thick ascending limb, Bartter's syndrome, is a salt-wasting renal disease associated with hypokalemia and metabolic alkalosis. Loss-of-function mutations in one of five distinct genes encoding components of the $Na^+/K^+/2Cl^-$ cotransporter (*NKCC2*), apical K^+ channel (*KCNJ1*), basolateral Cl^- channel (*CLCNKB*, *BSND*), or calcium-sensing receptor (*CASR*) can cause Bartter's syndrome.

Potassium recycling also contributes to a positive electrostatic charge in the lumen relative to the interstitium that promotes divalent cation (Mg^{2+} and Ca^{2+}) reabsorption through a paracellular pathway. A Ca^{2+}-sensing, G protein–coupled receptor (CaSR) on basolateral membranes regulates NaCl reabsorption in the thick ascending limb through dual signaling mechanisms using either cyclic AMP or eicosanoids. This receptor enables a steep relationship between plasma Ca^{2+} levels and renal Ca^{2+} excretion. Loss-of-function mutations in CaSR cause familial hypercalcemic hypocalciuria because of a blunted response of the thick ascending limb to extracellular Ca^{2+}. Mutations in *CLDN16* encoding paracellin-1, a transmembrane protein located within the tight junction complex, leads to familial hypomagnesemia with hypercalciuria and nephrocalcinosis, suggesting that the ion conductance of the paracellular pathway in the thick limb is regulated.

The loop of Henle contributes to urine-concentrating ability by establishing a *hypertonic medullary interstitium* that promotes water reabsorption by the inner medullary collecting duct located downstream in the nephron. *Countercurrent multiplication* produces a hypertonic medullary interstitium using two countercurrent systems: the loop of Henle (opposing descending and ascending limbs) and the vasa recta (medullary peritubular capillaries enveloping the loop). The countercurrent flow in these two systems helps maintain the hypertonic environment of the inner medulla, but NaCl reabsorption by the thick ascending limb is the primary initiating event. Reabsorption of NaCl without water dilutes the tubular fluid and adds new osmoles to medullary interstitial fluid. Because the descending thin limb is highly water permeable, osmotic equilibrium occurs between the descending limb tubular fluid and the interstitial space, leading to progressive solute trapping in the inner medulla. Maximum medullary interstitial osmolality also requires partial recycling of urea mediated by urea transporters in the inner medullary collecting duct.

DISTAL CONVOLUTED TUBULE

The distal convoluted tubule reabsorbs ~5% of filtered NaCl. This segment is composed of a tight epithelium with little water permeability. The major NaCl-transporting pathway uses an apical membrane, electroneutral thiazide-sensitive Na^+/Cl^- cotransporter in tandem with basolateral Na^+/K^+-ATPase and Cl^- channels **(Fig. 320-3*D*)**. Apical Ca^{2+}-selective channels (TRPV5) and basolateral Na^+/Ca^{2+} exchange mediate calcium reabsorption in the distal convoluted tubule. Ca^{2+} reabsorption is inversely related to Na^+ reabsorption and is stimulated by PTH. Blocking apical Na^+/Cl^- cotransport will reduce intracellular Na^+, favoring increased basolateral Na^+/Ca^{2+} exchange and passive apical Ca^{2+} entry. Loss-of-function mutations of *SLC12A3* encoding the apical Na^+/Cl^- cotransporter cause Gitelman syndrome, a salt-wasting disorder associated with hypokalemic alkalosis and hypocalciuria. Mutations in *TRPM6* encoding Mg^{2+} permeable ion channels also cause familial hypomagnesemia with hypocalcemia. A molecular complex of TRPM6 and TRPM7 proteins is critical for Mg^{2+} reabsorption in the distal convoluted tubule. Basolateral Mg^{2+} exit from these cells is postulated to involve Na^+/Mg^{2+} exchange.

COLLECTING DUCT

The collecting duct modulates the final composition of urine. The two major divisions, the cortical collecting duct and inner medullary collecting duct, contribute to reabsorbing ~4–5% of filtered Na^+ and are important for hormonal regulation of salt and water balance. Cells in both segments of the collecting duct express vasopressin-regulated water channels (aquaporin-2 on the apical membrane, aquaporin-3 and -4 on the basolateral membrane). The antidiuretic hormone vasopressin binds to the V2 receptor on the basolateral membrane and triggers an intracellular signaling cascade through G protein–mediated activation of adenylyl cyclase, which raises intracellular levels of cyclic AMP. This signaling cascade stimulates the insertion of water channels into the apical membrane of collecting duct cells to promote water permeability, water reabsorption, and production of concentrated urine. In the absence of vasopressin, collecting duct cells are water impermeable, and urine remains dilute. Nonpeptide V2 receptor blockers (vaptans) antagonize the antidiuretic effect of vasopressin and produce a water diuresis to treat symptomatic hyponatremia and can slow the progression of polycystic kidney disease.

The cortical collecting duct contains *high-resistance epithelia* with two cell types. Principal cells are the main water-reabsorbing, Na^+-reabsorbing, and K^+-secreting cells, and the site of action of aldosterone, K^+-sparing diuretics, and mineralocorticoid receptor antagonists such as spironolactone and eplerenone. The other cells are type A and B intercalated cells. Type A intercalated cells mediate acid secretion and bicarbonate reabsorption also under the influence of aldosterone. Type B intercalated cells mediate bicarbonate secretion and acid reabsorption.

Virtually all transport is mediated through the cellular pathway for both principal cells and intercalated cells. In principal cells, passive apical Na^+ entry occurs through an amiloride-sensitive, epithelial Na^+ channel (ENaC) with basolateral exit mediated by the Na^+/K^+-ATPase **(Fig. 320-3*E*)**. This Na^+ reabsorptive process is tightly regulated by aldosterone and is physiologically activated by a variety of proteolytic enzymes that cleave extracellular domains of ENaC; plasmin in the tubular fluid of individuals affected by nephrotic syndrome, for example, activates ENaC, leading to Na^+ retention. Aldosterone enters the cell across the basolateral membrane, binds to a cytoplasmic mineralocorticoid receptor, and then translocates into the nucleus, where it modulates gene transcription, which potentiates Na^+ reabsorption and K^+ secretion. Activating mutations in ENaC increase Na^+ reclamation and produce hypokalemia, hypertension, and metabolic alkalosis (Liddle's syndrome). The potassium-sparing diuretics amiloride and triamterene block ENaC, resulting in lower Na^+ reabsorption.

Principal cells secrete K^+ through an apical membrane potassium channel. Several forces govern the secretion of K^+. Most importantly, the high intracellular K^+ concentration generated by Na^+/K^+-ATPase creates a favorable concentration gradient for K^+ secretion into tubular fluid. With reabsorption of Na^+ without an accompanying anion, the tubular lumen becomes negative relative to the cell interior, creating a favorable electrical gradient for secretion of potassium. When Na^+ reabsorption is blocked, the electrical component of the driving force for K^+ secretion is blunted, and this explains lack of excess urinary K^+ loss during treatment with potassium-sparing diuretics or mineralocorticoid receptor antagonists. K^+ secretion is also promoted by aldosterone actions that potentiate regional Na^+ transport, which favor more lumen electronegativity, and by increasing the number and activity of potassium channels. Fast tubular fluid flow rates that occur during volume expansion or diuretics acting "upstream" of the cortical collecting duct also promote K^+ secretion, as does the presence of relatively nonreabsorbable anions (including bicarbonate and semisynthetic penicillins) that contribute to the lumen-negative potential. Off-target effects of certain antibiotics, such as trimethoprim and pentamidine, block ENaCs and predispose to hyperkalemia, especially when renal K^+ handling is impaired for other reasons. Principal cells, as described below, also participate in water reabsorption in response to vasopressin.

Intercalated cells do not participate in Na^+ reabsorption but instead mediate acid-base balance. These cells perform two types of transport: active H^+ transport mediated by H^+-ATPase (proton pump) and Cl^-/HCO_3^- exchange. Intercalated cells arrange the two transport mechanisms on opposite membranes to enable either acid or base secretion. Type A intercalated cells have an apical proton pump that mediates acid secretion and a basolateral Cl^-/HCO_3^- anion exchanger for bicarbonate reabsorption (Fig. 320-3*E*). Aldosterone increases the number of H^+-ATPase pumps, sometimes contributing to the development of metabolic alkalosis. Secreted H^+ is buffered by NH_3 that has diffused into the collecting duct lumen from the surrounding interstitium. By contrast, type B intercalated cells have the Cl^-/HCO_3^- exchanger on the apical membrane to mediate bicarbonate secretion while the proton pump resides on the basolateral membrane to enable H^+ reabsorption. Under conditions of acidemia, the kidney preferentially uses type A intercalated cells to secrete the excess H^+ and generate more HCO_3^-. The opposite is true in states of bicarbonate excess with alkalemia where the type B intercalated cells predominate. An extracellular protein called *hensin* mediates this adaptation.

Inner medullary collecting duct cells share many similarities with principal cells of the cortical collecting duct. They have apical Na^+ and K^+ channels that mediate Na^+ reabsorption and K^+ secretion, respectively **(Fig. 320-3*F*)**. Sodium reabsorption by inner medullary collecting duct cells is also inhibited by the natriuretic peptides *atrial natriuretic peptide* or *renal natriuretic peptide* (urodilatin); the same gene encodes both peptides but uses different posttranslational processing of a common preprohormone to generate different proteins. Atrial natriuretic peptides are secreted by atrial myocytes in response to volume expansion, whereas urodilatin is secreted by renal tubular epithelia. Natriuretic peptides interact with either apical (urodilatin) or basolateral (atrial natriuretic peptides) receptors on inner medullary collecting duct cells to stimulate guanylyl cyclase and raise levels of cytoplasmic cGMP. This effect in turn reduces the activity of the apical Na^+ channel in these cells and attenuates net Na^+ reabsorption, producing natriuresis.

The inner medullary collecting duct transports urea out of the lumen, returning urea to the interstitium, where it contributes to the hypertonicity of the medullary interstitium. Urea is recycled by diffusing from the interstitium into the descending and ascending limbs of the loop of Henle.

HORMONAL REGULATION OF SODIUM AND WATER BALANCE

The balance of solute and water in the body is determined by the amounts ingested, distributed to various fluid compartments, and excreted by skin, bowel, and kidneys. *Tonicity*, the osmolar state determining the volume behavior of cells in a solution, is regulated by water balance **(Fig. 320-4*A*)**, and *extracellular blood volume* is regulated by Na^+ balance **(Fig. 320-4*B*)**. The kidney is a critical modulator of both physiologic processes.

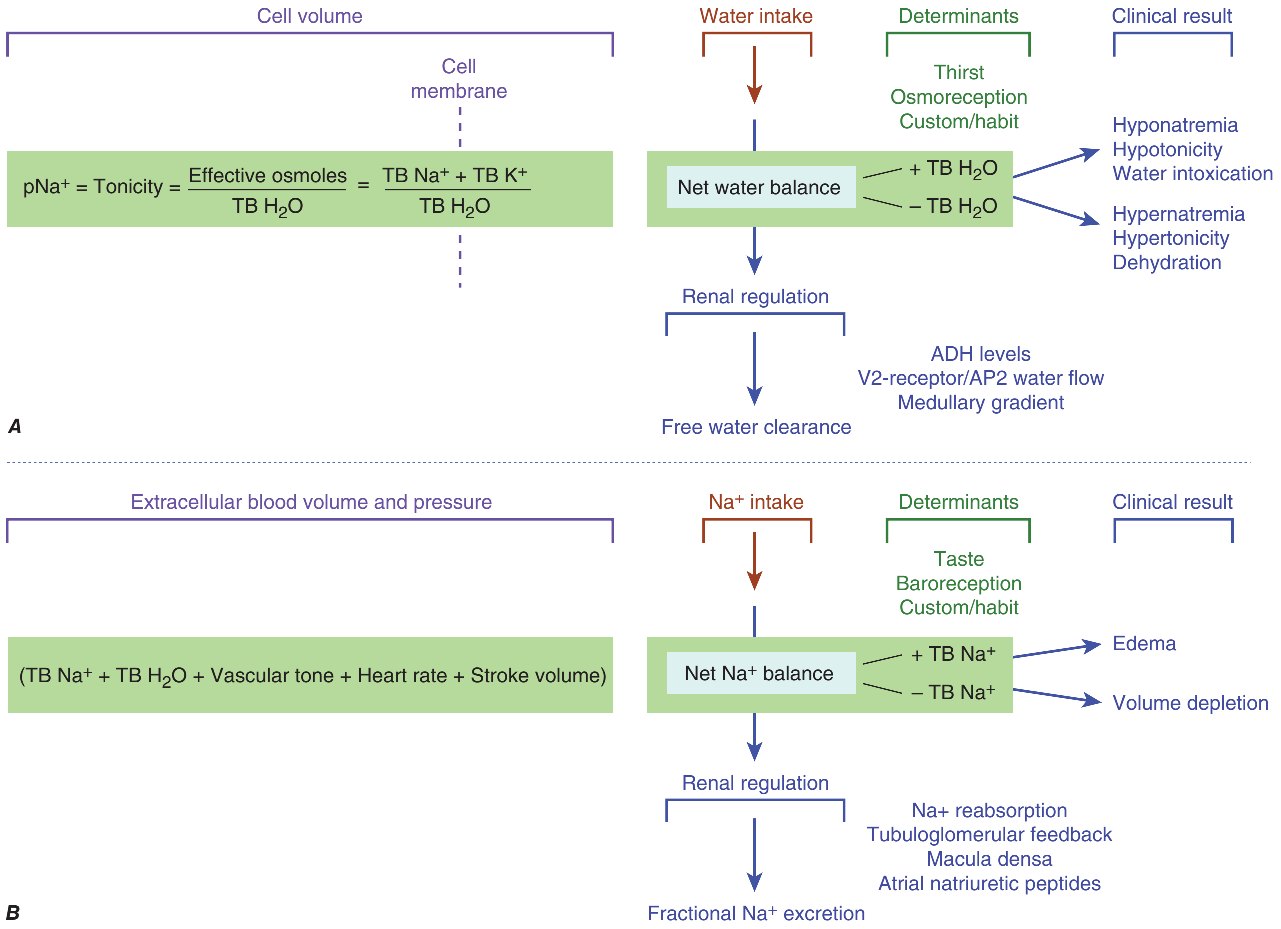
Cell volume
Cell membrane
pNa+ = Tonicity = Effective osmoles / TB H2O = TB Na+ + TB K+ / TB H2O
Water intake
Determinants
Thirst
Osmoreception
Custom/habit
Clinical result
Net water balance
+ TB H2O
– TB H2O
Hyponatremia
Hypotonicity
Water intoxication
Hypernatremia
Hypertonicity
Dehydration
Renal regulation
ADH levels
V2-receptor/AP2 water flow
Medullary gradient
Free water clearance
A
Extracellular blood volume and pressure
Na+ intake
Determinants
Taste
Baroreception
Custom/habit
Clinical result
(TB Na+ + TB H2O + Vascular tone + Heart rate + Stroke volume)
Net Na+ balance
+ TB Na+
– TB Na+
Edema
Volume depletion
Renal regulation
Na+ reabsorption
Tubuloglomerular feedback
Macula densa
Atrial natriuretic peptides
Fractional Na+ excretion
B

FIGURE 320-4 Determinants of sodium and water balance. ***A.*** Plasma Na$^+$ concentration is a surrogate marker for plasma tonicity. Tonicity is determined by the number of effective osmoles in the body divided by the total body H_2O (TB H_2O), which translates simply into the total body Na (TB Na$^+$) and anions outside the cell separated from the total body K (TB K$^+$) inside the cell by the cell membrane. Net water balance is determined by the integrated functions of thirst, osmoreception, Na reabsorption, vasopressin release, and the strength of the medullary gradient in the kidney, keeping tonicity within a narrow range of osmolality (~280 mosmol/L). When water metabolism is disturbed and total body water increases, hyponatremia, hypotonicity, and water intoxication occur; when total body water decreases, hypernatremia, hypertonicity, and dehydration occur. ***B.*** Extracellular blood volume and pressure are an integrated function of total body Na$^+$ (TB Na$^+$), total body H_2O (TB H_2O), vascular tone, heart rate, and stroke volume that modulates volume and pressure in the vascular tree of the body. This extracellular blood volume is determined by net Na balance under the control of taste, baroreception, habit, Na$^+$ reabsorption, macula densa/tubuloglomerular feedback, and natriuretic peptides. When Na$^+$ metabolism is disturbed and total body Na$^+$ increases, edema occurs; when total body Na$^+$ is decreased, volume depletion occurs. ADH, antidiuretic hormone; AQP2, aquaporin-2.

WATER BALANCE

Tonicity depends on the variable concentration of *effective osmoles* inside and outside the cell causing water to move in either direction across its membrane. Classic effective osmoles, like Na$^+$, K$^+$, and their anions, are solutes trapped on either side of a cell membrane, where they collectively partition and obligate water to move and find equilibrium in proportion to retained solute. Normal tonicity (~280 mosmol/L) is rigorously defended by osmoregulatory mechanisms that control water balance to protect tissues from inadvertent *dehydration* (cell shrinkage) or *water intoxication* (cell swelling), both of which impair cell function (Fig. 320-4*A*).

The mechanisms that control osmoregulation are distinct from those governing extracellular volume, although there is some shared physiology in both processes. While cellular concentrations of K$^+$ have a determinant role in any level of tonicity, the routine surrogate marker for assessing clinical tonicity is the concentration of serum Na$^+$. Any reduction in total body water, which raises the Na$^+$ concentration, triggers a brisk sense of thirst and conservation of water by decreasing renal water excretion mediated by release of vasopressin from the posterior pituitary. Conversely, a lower plasma Na$^+$ concentration triggers more renal water excretion by suppressing the secretion of vasopressin. Whereas all cells expressing mechanosensitive TRPV1, 2, or 4 channels, among potentially other sensors, respond to changes in tonicity by altering their volume and Ca^{2+} concentration, only TRPV$^+$ neuronal cells connected to the organum vasculosum of the lamina terminalis are *osmoreceptive*. Only these cells, because of their neural connectivity and adjacency to a minimal blood-brain barrier, modulate the downstream release of vasopressin by the posterior lobe of the pituitary gland. Secretion is stimulated primarily by changing tonicity and secondarily by other nonosmotic signals such as variable blood volume, stress, pain, nausea, and some drugs. The release of vasopressin by the posterior pituitary increases linearly as plasma tonicity rises above normal, although this varies, depending on the perception of extracellular volume (one form of cross-talk between mechanisms that regulate blood volume and osmolality). Changing the intake or excretion of water provides a means for adjusting plasma tonicity; thus, osmoregulation governs water balance.

The kidneys contribute to maintaining water balance through the regulation of renal water excretion. The ability to concentrate urine to an osmolality exceeding that of plasma enables water conservation, whereas the ability to produce urine more dilute than plasma promotes excretion of excess water. For water to enter or exit a cell, the cell membrane must express aquaporins. In the kidney, aquaporin-1 is constitutively active in all water-permeable segments (e.g., proximal tubule, descending thin limb of the loop of Henle), whereas aquaporin-2, -3, and -4 in the collecting duct promote vasopressin-regulated water permeability. Net water reabsorption is ultimately driven by the osmotic gradient between dilute tubular fluid and a hypertonic medullary interstitium.

SODIUM BALANCE

The perception of *extracellular blood volume* is determined, in part, by the integration of arterial tone, cardiac stroke volume, heart rate, and the water and solute content of extracellular fluid. Na^+ and accompanying anions are the most abundant extracellular effective osmoles and together support a blood volume around which pressure is generated. Under normal conditions, this volume is regulated by sodium balance (Fig. 320-4*B*), and the balance between daily Na^+ intake and excretion is under the influence of *baroreceptors* in regional blood vessels and vascular hormone sensors modulated by atrial natriuretic peptides, the renin-angiotensin-aldosterone system, Ca^{2+} signaling, adenosine, vasopressin, and the neural adrenergic axis. If Na^+ intake exceeds Na^+ excretion (positive Na^+ balance), then a rising blood volume will trigger a proportional increase in urinary Na^+ excretion. Conversely, when Na^+ intake is less than urinary excretion (negative Na^+ balance), blood volume will fall and trigger enhanced renal Na^+ reabsorption, leading to decreased urinary Na^+ excretion.

The renin-angiotensin-aldosterone system is the best-understood hormonal system modulating renal Na^+ excretion. Renin is synthesized and secreted by granular cells in the wall of the afferent arteriole. Its secretion is controlled by several factors, including β_1-adrenergic stimulation to the afferent arteriole, input from the macula densa, and prostaglandins. Renin and ACE activity eventually produce angiotensin II that directly and indirectly promotes renal Na^+ and water reabsorption. Stimulation of proximal tubular Na^+/H^+ exchange by angiotensin II directly increases Na^+ reabsorption. Angiotensin II also promotes Na^+ reabsorption along the collecting duct by stimulating aldosterone secretion by the adrenal cortex. Constriction of the efferent glomerular arteriole by angiotensin II indirectly boosts the filtration fraction and raises peritubular capillary oncotic pressure to promote tubular Na^+ reabsorption. Finally, angiotensin II inhibits renin secretion through a negative feedback loop. Alternative metabolism of angiotensin by ACE2 generates the vasodilatory peptide angiotensin 1-7 that acts through Mas receptors to counterbalance several actions of angiotensin II on blood pressure and renal function (Fig. 320-2*C*).

Aldosterone is synthesized and secreted by granulosa cells in the adrenal cortex. It binds to cytoplasmic mineralocorticoid receptors in the collecting duct principal cells and boosts the activity of ENaC, apical membrane K^+ channel, and basolateral Na^+/K^+-ATPase. These effects are mediated in part by aldosterone-stimulated transcription of the gene encoding serum/glucocorticoid-induced kinase 1 (SGK1). The activity of ENaC is increased by SGK1-mediated phosphorylation of the ubiquitin-protein ligase Nedd4-2 that promotes ubiquitination and recycling of the Na^+ channel from the plasma membrane. Phosphorylated Nedd4-2 has impaired interactions with ENaC, leading to higher channel density at the plasma membrane and greater capacity for Na^+ reabsorption by the collecting duct.

Chronic exposure to aldosterone is associated with lower urinary Na^+ excretion lasting only a few days, after which Na^+ excretion returns to previous levels. This phenomenon, called *aldosterone escape*, is explained by lower proximal tubular Na^+ reabsorption following blood volume expansion. Excess Na^+ that is not reabsorbed by the proximal tubule overwhelms the reabsorptive capacity of more distal nephron segments. This escape may be facilitated by atrial natriuretic peptides that lose their effectiveness in the clinical settings of heart failure, nephrotic syndrome, and cirrhosis, leading to severe Na^+ retention and volume overload.

FURTHER READING

Avraham S: The mesangial cell: The glomerular stromal cell. Nature Rev Nephrol 17:855, 2021.

Carrisoza-Gaytan R: PIEZO1 is a distal nephron mechanosensory and is required for flow-induced K^+ secretion. J Clin Invest 134:E174806, 2024.

De Baaij JHF: Magnesium reabsorption in the kidney. Am J Physiol Renal Physiol 324:F227, 2023.

Downie ML et al: Inherited tubulopathies of the kidney. Clin J Am Soc Nephrol 16:620, 2021.

Polidoro JZ et al: Paracrine and endocrine regulation of renal K^+ secretion. Am J Physiol Renal Physiol 322:F360, 2022.

Schnell J: Principles of human and mouse nephron development. Nature Rev Nephrol 18:628, 2022.

Verschuren EHJ et al: Sensing of tubular flow and renal electrolyte transport. Nature Rev Nephrol 16:337, 2020.

321

Acute Kidney Injury

Sushrut S. Waikar, Joseph V. Bonventre

Acute kidney injury (AKI) is defined by the impairment of kidney filtration and excretory function over days to weeks (generally known or expected to have occurred within 7 days), resulting in the retention of nitrogenous and other waste products normally cleared by the kidneys. The generally accepted Kidney Disease: Improving Global Outcomes (KDIGO) definition of AKI is an increase in serum creatinine (SCr) of ≥0.3 mg/dL within 48 h or an increase in SCr ≥1.5 times from baseline over 7 days or urine output <0.5 mL/kg per h for >6 h. AKI is not a single disease but rather a designation for a heterogeneous group of conditions that share common diagnostic features: specifically, an increase in filtration markers (SCr or cystatin C) often associated with a reduction in urine volume. It is important to recognize that AKI is a clinical diagnosis and not a structural one. A patient may have AKI with or without injury to the kidney parenchyma making the term a misnomer. AKI can range in severity from asymptomatic and transient changes in laboratory parameters of glomerular filtration rate (GFR), to overwhelming and rapidly fatal derangements in the ability of the kidney to maintain effective circulating volume regulation, excrete nitrogenous wastes and metabolic toxins, and maintain electrolyte and acid-base composition of the plasma.

EPIDEMIOLOGY

AKI complicates 5–7% of acute-care hospital admissions and up to 30% of admissions to the intensive care unit (ICU). AKI severity is staged based on the magnitude of the rise in SCr and severity and duration of oliguria (**Table 321-1**). The incidence of AKI has grown by more than fourfold in the United States since 1988 and is estimated to have a yearly incidence of 500 per 100,000 population, higher than the yearly incidence of stroke. Morbidity of AKI in those admitted to the ICU exceeds 50% in many studies. AKI also has longer term implications even if the patient survives the hospitalization. AKI increases the risk

TABLE 321-1 Staging of Acute Kidney Injury Severity

STAGE	SERUM CREATININE	URINE OUTPUT
1	1.5–1.9 times baseline OR ≥0.3 mg/dL (≥26.5 μmol/L) increase	<0.5 mL/kg per h for 6–12 h
2	2.0–2.9 times baseline	<0.5 mL/kg per h for ≥12 h
3	3.0 times baseline OR increase in serum creatinine to ≥4.0 mg/dL (≥353.6 μmol/L) OR initiation of renal replacement therapy OR, in patients <18 years of age, decrease in eGFR to <35 mL/min per 1.73 m^2	<0.3 mL/kg per h for ≥24 h OR Anuria for ≥12 h

Abbreviation: eGFR, estimated glomerular filtration rate.

for the development or worsening of chronic kidney disease (CKD) and also increases the risk of future cardiovascular disease. AKI may also occur in the community. Common causes of community-acquired AKI include volume depletion, heart failure, adverse effects of medications, obstruction of the urinary tract, or malignancy. The most common clinical settings for hospital-acquired AKI are sepsis, major surgical procedures, critical illness involving heart or liver failure, and nephrotoxic medication administration. Clinically, AKI more commonly develops when ischemia or other insults occur in the context of limited renal functional reserve. The healthy kidney has the ability to increase its regional or overall function in response to damage to a subset of nephrons or in response to a perceived need to enhance excretion, such as in response to a protein load. With normal aging, there is reduction in this capacity, which is also reduced in individuals with CKD or coexisting insults such as sepsis, vasoactive or nephrotoxic drugs, rhabdomyolysis, or the systemic inflammatory states associated with burns and pancreatitis. When there is reduced renal reserve, any additional impairment in GFR is likely to be reflected by a change in SCr or cystatin C and hence a more likely diagnosis of AKI.

AKI IN THE DEVELOPING WORLD

AKI is also a major medical complication in the developing world, where the epidemiology differs from that in developed countries due to differences in demographics, economics, environmental factors, and comorbid disease burden. While certain features of AKI are common in developed and developing countries—particularly because urban centers of some developing countries increasingly resemble those in the developed world—many etiologies for AKI are region-specific, such as envenomations from snakes, spiders, caterpillars, and bees; infectious causes such as malaria and leptospirosis; and crush injuries and resultant rhabdomyolysis from earthquakes. In developing countries, resources to diagnose and manage AKI are often limited.

ETIOLOGY AND PATHOPHYSIOLOGY

The causes of AKI have traditionally been divided into three broad categories: prerenal azotemia, intrinsic renal parenchymal disease, and postrenal obstruction **(Fig. 321-1)**.

PRERENAL AZOTEMIA

Prerenal azotemia (from "azo," meaning nitrogen, and "-emia," meaning in the blood), the most common form of AKI, results from inadequate renal plasma flow and intraglomerular hydrostatic pressure to support normal glomerular filtration. The most common clinical conditions associated with prerenal azotemia are hypovolemia, decreased cardiac output, and medications that interfere with renal autoregulatory vascular responses such as nonsteroidal anti-inflammatory drugs (NSAIDs) and inhibitors of the renin-angiotensin system **(Fig. 321-2)**. Sodium-glucose cotransporter 2 (SGLT-2) inhibitors used for the treatment of diabetes mellitus and related complications do not appear to increase the risk of AKI despite their effects on lowering intraglomerular pressure and inducing natriuresis; in fact, recent studies have suggested a protective effect of these agents in preventing AKI.

By definition, prerenal azotemia involves no parenchymal damage to the kidney and is rapidly reversible once parenchymal blood flow and intraglomerular hemodynamics are restored. In many cases, however, prerenal azotemia may coexist with other forms of intrinsic AKI associated with processes acting directly on the renal parenchyma. Prolonged periods of prerenal azotemia may lead to ischemic injury to the tubular cells with necrosis, hence termed acute tubular necrosis (ATN).

Normal GFR is maintained in part by renal blood flow and the relative resistances of the afferent and efferent renal arterioles, which determine the glomerular plasma flow rate and the transcapillary hydraulic pressure gradient that drive glomerular ultrafiltration. Mild degrees of hypovolemia and reductions in cardiac output elicit compensatory renal vasoconstriction and enhanced reabsorption of salt and water to maintain blood pressure and increase intravascular volume to sustain perfusion to the cerebral and coronary vessels. Mediators of this response include angiotensin II, norepinephrine, and vasopressin (also termed antidiuretic hormone). Glomerular filtration can be maintained despite reduced renal blood flow by angiotensin II–mediated renal efferent arteriolar vasoconstriction. In addition, a myogenic reflex within the afferent arteriole leads to dilation in the setting of low perfusion pressure, thereby maintaining glomerular perfusion. Intrarenal biosynthesis of vasodilator prostaglandins (prostacyclin, prostaglandin E_2), kallikrein and kinins, and possibly nitric oxide (NO) also increases in response to low renal perfusion pressure. Autoregulation is also accomplished by tubuloglomerular feedback, in which decreases in solute delivery to the macula densa (specialized cells within the distal tubule) elicit dilation of the juxtaposed afferent arteriole in order to maintain glomerular perfusion, a mechanism mediated, in part, by NO. There is a limit, however, to the ability of

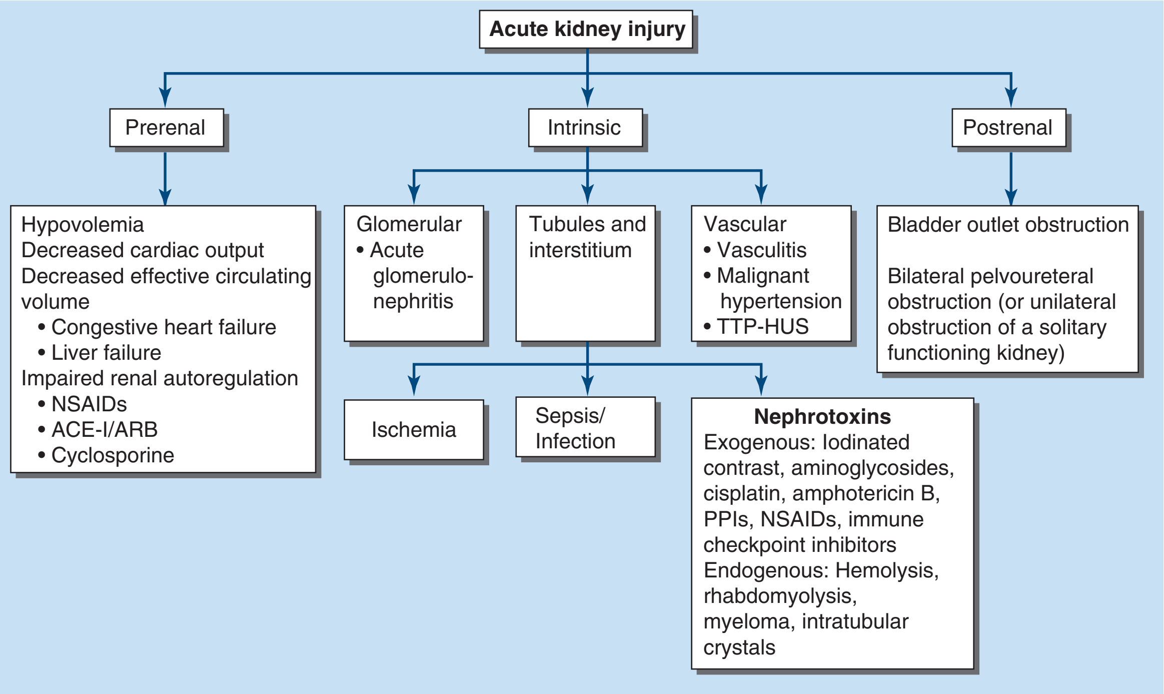
Acute kidney injury
Prerenal
Intrinsic
Postrenal
Hypovolemia
Decreased cardiac output
Decreased effective circulating volume
• Congestive heart failure
• Liver failure
Impaired renal autoregulation
• NSAIDs
• ACE-I/ARB
• Cyclosporine
Glomerular
• Acute glomerulonephritis
Tubules and interstitium
Vascular
• Vasculitis
• Malignant hypertension
• TTP-HUS
Bladder outlet obstruction
Bilateral pelvoureteral obstruction (or unilateral obstruction of a solitary functioning kidney)
Ischemia
Sepsis/ Infection
Nephrotoxins
Exogenous: Iodinated contrast, aminoglycosides, cisplatin, amphotericin B, PPIs, NSAIDs, immune checkpoint inhibitors
Endogenous: Hemolysis, rhabdomyolysis, myeloma, intratubular crystals

FIGURE 321-1 Classification of the major causes of acute kidney injury. ACE-I, angiotensin-converting enzyme inhibitor-I; ARB, angiotensin receptor blocker; NSAIDs, nonsteroidal anti-inflammatory drugs; PPI, proton pump inhibitors; TTP-HUS, thrombotic thrombocytopenic purpura–hemolytic-uremic syndrome.

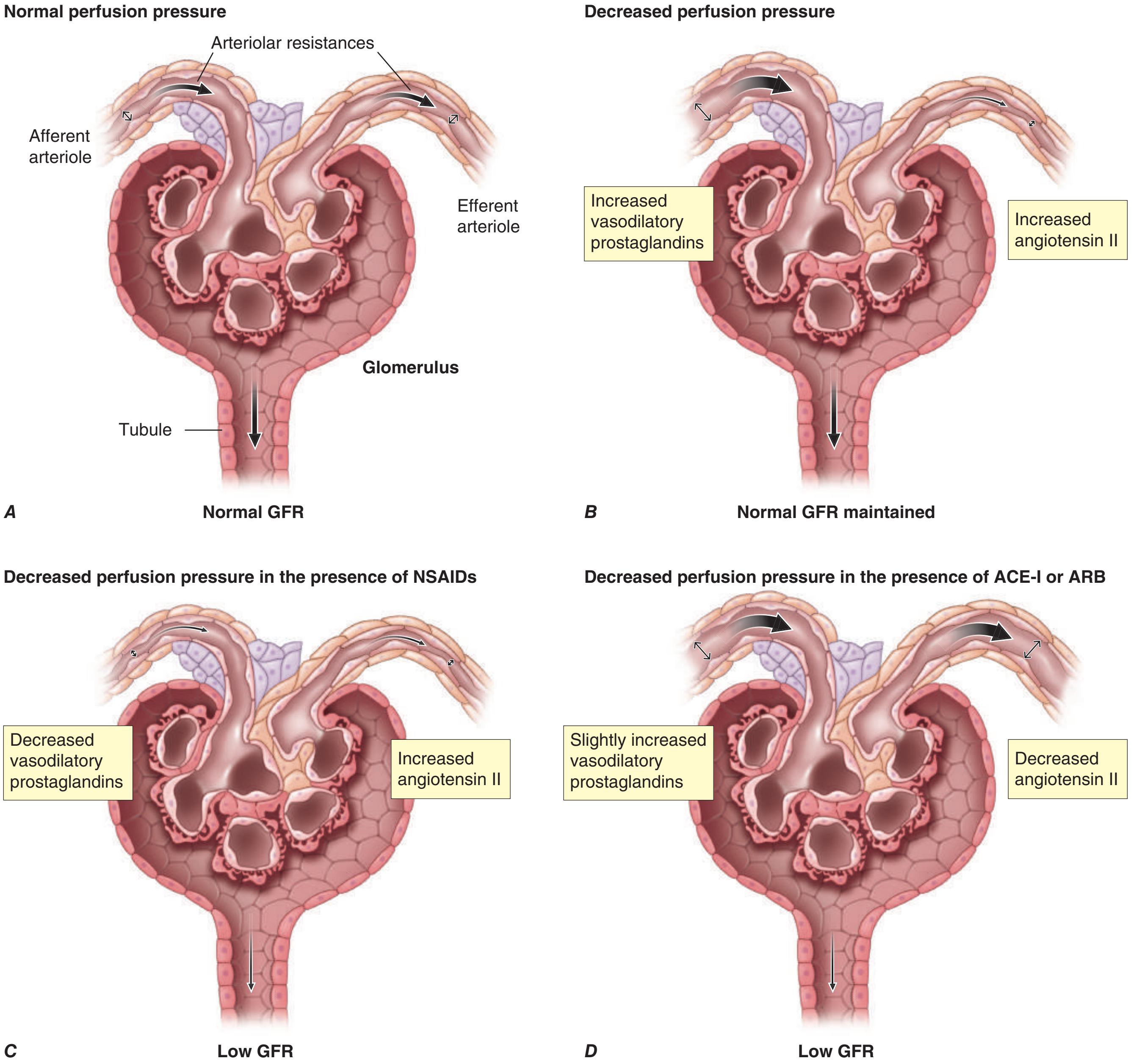
Normal perfusion pressure
Arteriolar resistances
Afferent arteriole
Efferent arteriole
Glomerulus
Tubule
A
Normal GFR
Decreased perfusion pressure
Increased vasodilatory prostaglandins
Increased angiotensin II
B
Normal GFR maintained
Decreased perfusion pressure in the presence of NSAIDs
Decreased vasodilatory prostaglandins
Increased angiotensin II
C
Low GFR
Decreased perfusion pressure in the presence of ACE-I or ARB
Slightly increased vasodilatory prostaglandins
Decreased angiotensin II
D
Low GFR

FIGURE 321-2 Intrarenal mechanisms for autoregulation of the glomerular filtration rate (GFR) under decreased perfusion pressure and reduction of the GFR by drugs. ***A.*** Normal conditions and a normal GFR. ***B.*** Reduced perfusion pressure within the autoregulatory range. Normal glomerular capillary pressure is maintained by afferent vasodilatation and efferent vasoconstriction. Angiotensin constricts afferent (preglomerular) and efferent (postglomerular) arterioles but preferentially increases efferent arteriolar resistance. ***C.*** Reduced perfusion pressure with a nonsteroidal anti-inflammatory drug (NSAID). Loss of vasodilatory prostaglandins increases afferent resistance; this causes the glomerular capillary pressure to drop below normal values and the GFR to decrease. ***D.*** Reduced perfusion pressure with an angiotensin-converting enzyme inhibitor (ACE-I) or an angiotensin receptor blocker (ARB). Loss of angiotensin II action reduces efferent resistance; this causes the glomerular capillary pressure to drop below normal values and the GFR to decrease. *(From JG Abuelo: Normotensive ischemic acute renal failure. N Engl J Med 357:797, 2007. Copyright © 2007, Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

these counterregulatory mechanisms to maintain GFR in the face of systemic hypotension. Even in healthy adults, renal autoregulation usually fails once the systolic blood pressure falls below 80 mmHg.

A number of factors determine the robustness of the autoregulatory response and the risk of prerenal azotemia. Atherosclerosis, long-standing hypertension, and older age can lead to hyalinosis and myointimal hyperplasia, causing structural narrowing of the intrarenal arterioles and impaired capacity for renal afferent vasodilation. In CKD, renal afferent vasodilation may be operating at maximal capacity in order to maximize GFR in response to reduced functional renal mass. Drugs can affect the compensatory changes evoked to maintain GFR. NSAIDs inhibit renal prostaglandin production, limiting renal afferent vasodilation. Angiotensin-converting enzyme (ACE) inhibitors and angiotensin receptor blockers (ARBs) limit renal efferent vasoconstriction; this effect is particularly pronounced in patients with bilateral renal artery stenosis or unilateral renal artery stenosis (in the case of a solitary functioning kidney) because, as indicated above, efferent arteriolar vasoconstriction is needed to maintain GFR due to low renal perfusion. The combined use of NSAIDs with ACE inhibitors or ARBs poses a particularly high risk for developing prerenal azotemia.

Hepatorenal syndrome is a cause of AKI in individuals with multiorgan pathobiology affecting kidney and liver. Many individuals with advanced liver disease exhibit a hemodynamic profile that resembles prerenal azotemia in the setting of total-body volume overload. Systemic vascular resistance is markedly reduced due to primary arterial vasodilation in the splanchnic circulation, resulting ultimately in activation of vasoconstrictor responses similar to those seen in hypovolemia. AKI is a common complication in this setting, and it can be triggered by volume depletion and spontaneous bacterial peritonitis.

The hepatorenal syndrome, which represents the advanced stage of impaired perfusion to the kidneys secondary to advanced liver disease, is difficult to distinguish from prerenal azotemia and is a diagnosis of exclusion. A particularly poor prognosis is seen in the case of type 1 hepatorenal syndrome, in which AKI persists despite volume administration and withholding of diuretics. Type 2 hepatorenal syndrome is a less severe form characterized mainly by refractory ascites.

INTRINSIC AKI

The most common causes of intrinsic AKI are sepsis, ischemia, and nephrotoxins, both endogenous and exogenous **(Fig. 321-3)**. As mentioned previously, in many cases, prerenal azotemia advances to tubular injury. Although often the AKI is attributed to "acute tubular necrosis," human biopsy confirmation of tubular necrosis is, in general, often lacking in cases of sepsis and ischemia; indeed, processes such as inflammation, apoptosis, and altered regional perfusion may be important contributors pathophysiologically without frank necrosis. There are other potential causes of AKI in settings such as sepsis, including drug-induced interstitial nephritis or glomerulonephritis. These and other causes of intrinsic AKI can be catalogued anatomically according to the major site of renal parenchymal damage: glomeruli, tubulointerstitium, and vessels, although there is frequently overlap in tissue compartment involvement. For example, glomerulonephritis can alter efferent arteriolar blood flow, which then reduces capillary perfusion to a region of the nephron leading to cell death, obstruction of the lumen with cellular debris, and impaired tubular function.

SEPSIS-ASSOCIATED AKI

In the United States, >1 million cases of sepsis occur each year. AKI complicates >50% of cases of severe sepsis and greatly increases the risk of death. Sepsis is also a very important cause of AKI in the developing world. AKI also predisposes to sepsis. Decreases in GFR with sepsis can occur even in the absence of overt hypotension, although many cases of severe AKI typically occur in the setting of hemodynamic compromise requiring vasopressor support. Reduced urine output is common in sepsis-induced AKI. While there can be tubular injury associated with AKI in sepsis as manifest by the presence of tubular debris and casts in the urine, postmortem examinations of kidneys from individuals with severe sepsis suggest that other factors, perhaps related to inflammation, mitochondrial dysfunction, and interstitial edema, contribute to the pathophysiology of sepsis-induced AKI.

The hemodynamic effects of sepsis—arising from generalized arterial vasodilation, mediated in part by cytokines that upregulate the expression of inducible NO synthase in the vasculature—can lead to a

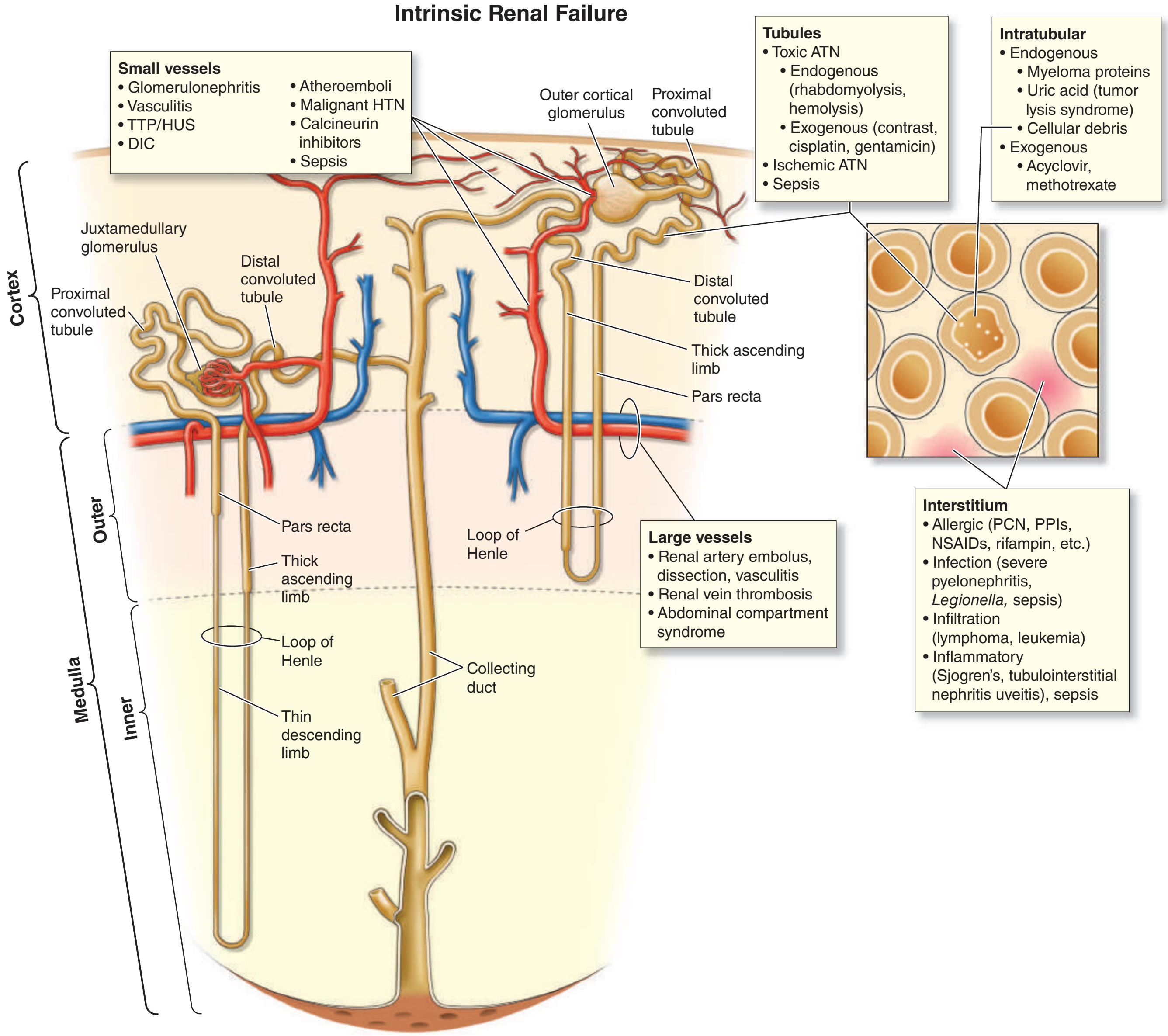
Intrinsic Renal Failure
Small vessels
• Glomerulonephritis
• Vasculitis
• TTP/HUS
• DIC
• Atheroemboli
• Malignant HTN
• Calcineurin inhibitors
• Sepsis
Outer cortical glomerulus
Proximal convoluted tubule
Tubules
• Toxic ATN
• Endogenous (rhabdomyolysis, hemolysis)
• Exogenous (contrast, cisplatin, gentamicin)
• Ischemic ATN
• Sepsis
Intratubular
• Endogenous
• Myeloma proteins
• Uric acid (tumor lysis syndrome)
• Cellular debris
• Exogenous
• Acyclovir, methotrexate
Cortex
Juxtamedullary glomerulus
Proximal convoluted tubule
Distal convoluted tubule
Distal convoluted tubule
Thick ascending limb
Pars recta
Medulla
Outer
Inner
Pars recta
Thick ascending limb
Loop of Henle
Thin descending limb
Loop of Henle
Collecting duct
Large vessels
• Renal artery embolus, dissection, vasculitis
• Renal vein thrombosis
• Abdominal compartment syndrome
Interstitium
• Allergic (PCN, PPIs, NSAIDs, rifampin, etc.)
• Infection (severe pyelonephritis, Legionella, sepsis)
• Infiltration (lymphoma, leukemia)
• Inflammatory (Sjogren's, tubulointerstitial nephritis uveitis), sepsis

FIGURE 321-3 Major causes of intrinsic acute kidney injury. ATN, acute tubular necrosis; DIC, disseminated intravascular coagulation; HTN, hypertension; PCN, penicillin; PPI, proton pump inhibitors; TINU, tubulointerstitial nephritis-uveitis; TTP/HUS, thrombotic thrombocytopenic purpura/hemolytic-uremic syndrome.

reduction in GFR. The operative mechanisms may be excessive efferent arteriole vasodilation, particularly early in the course of sepsis, or renal vasoconstriction from activation of the sympathetic nervous system, the renin-angiotensin-aldosterone system, or increased levels of vasopressin or endothelin. Sepsis may lead to endothelial damage, which results in increased microvascular leukocyte adhesion and migration, thrombosis, permeability, increased interstitial pressure, reduction in local flow to tubules, and activation of reactive oxygen species, all of which may injure renal tubular cells.

AKI can be an important complication of viral infections, such as hantavirus, dengue virus, or SARS-CoV-2. The pathophysiology of AKI due to viral infections remains incompletely understood. As an example, some have reported infection of the kidney with SARS-CoV-2, whereas others have found less direct involvement. SARS-CoV-2 is associated with a large release of cytokines into the circulation ("cytokine storm"), which may cause diffuse intrarenal vasoconstriction. Finally, there is a generalized hypercoagulable state associated with SARS-CoV-2 that may contribute to the impairment of intrarenal blood flow.

■ ISCHEMIA-ASSOCIATED AKI

Healthy kidneys receive 20% of the cardiac output and account for 10% of resting oxygen consumption, despite constituting only 0.5% of the human body mass. The kidneys are also the site of one of the most hypoxic regions in the body, the renal medulla. The outer medulla is particularly vulnerable to ischemic damage because of the architecture of the blood vessels that supply oxygen and nutrients to the tubules. In the outer medulla, enhanced leukocyte-endothelial interactions in the small vessels lead to inflammation and reduced local blood flow to the metabolically very active S3 segment of the proximal tubule, which depends on oxidative metabolism for survival. Mitochondrial dysfunction leads to impaired oxidative phosphorylation with less efficient adenosine triphosphate (ATP) generation and mitochondrial release of reactive oxygen species, both of which play a role in renal tubular injury. Transient ischemia alone in a normal kidney is usually not sufficient to cause severe AKI, as evidenced by the relatively low risk of severe AKI even after total interruption of renal blood flow during suprarenal aortic clamping or cardiac arrest. Prerenal azotemia and ischemia-associated AKI represent a continuum of the manifestations of renal hypoperfusion leading to ATN. Persistent preglomerular vasoconstriction may be a common underlying cause of the reduction in GFR seen in AKI; implicated factors for vasoconstriction include activation of tubuloglomerular feedback from enhanced delivery of solute to the macula densa following proximal tubule injury and reduced reabsorption, increased basal vascular tone and reactivity to vasoconstrictive agents, and decreased vasodilator responsiveness. Other contributors to low GFR include backleak of filtrate across damaged and denuded tubular epithelium and mechanical obstruction of tubules from necrotic debris **(Fig. 321-4)**.

Postoperative AKI Ischemia-associated AKI is a serious complication in the postoperative period, especially after major operations involving significant blood loss and intraoperative hypotension. The procedures most commonly associated with AKI are cardiac surgery with cardiopulmonary bypass (particularly for combined valve and bypass procedures), vascular procedures with aortic cross clamping, and intraperitoneal procedures. Severe AKI requiring dialysis occurs in ~1% of cardiac and vascular surgery procedures. The risk of severe AKI has been less well studied for major intraperitoneal procedures but appears to be of comparable magnitude. Common risk factors for postoperative AKI include underlying CKD, older age, diabetes mellitus, congestive heart failure, and emergency procedures. The pathophysiology of AKI following cardiac surgery is multifactorial. Major AKI risk factors are common in the population undergoing cardiac or vascular surgery. Over time, more of these surgical procedures are performed on older patients with comorbidities that predispose them to AKI and hasten progression of end-stage kidney disease (ESKD) if they develop AKI. Longer duration of cardiopulmonary bypass is a risk factor for AKI. In addition to ischemic injury from sustained hypoperfusion, cardiopulmonary bypass may cause AKI through a number of mechanisms including extracorporeal circuit activation of leukocytes and inflammatory processes, hemolysis with resultant pigment nephropathy (see below), and aortic injury with resultant atheroemboli. AKI from atheroembolic disease, which can also occur following percutaneous catheterization of the aorta, or spontaneously, is due to cholesterol crystal embolization resulting in partial or total occlusion of multiple small arteries within the kidney. Over time, a foreign body reaction can result in intimal proliferation, giant cell formation, and further narrowing of the vascular lumen, accounting for the generally subacute (over a period of weeks rather than days) decline in renal function.

Burns and Acute Pancreatitis Extensive fluid losses into the extravascular compartments of the body frequently accompany severe burns and acute pancreatitis. AKI is an ominous complication of burns, affecting 25% of individuals with >10% total body surface area involvement. In addition to severe hypovolemia resulting in decreased cardiac output and increased neurohormonal activation, burns and acute pancreatitis both lead to dysregulated inflammation and an increased risk of sepsis and acute lung injury, all of which may facilitate the development and progression of AKI. Individuals undergoing massive fluid resuscitation for trauma, burns, and acute pancreatitis can also develop abdominal compartment syndrome, where markedly elevated intraabdominal pressures, usually >20 mmHg, lead to renal vein compression and reduced GFR. Drug nephrotoxicity is also an important contributor to AKI. Mortality is much higher in patients who develop AKI.

Diseases of the Vasculature Leading to Ischemia These diseases can compromise oxygen and metabolic substrate delivery to the tubules and glomeruli. Microvascular causes of AKI include the thrombotic microangiopathies (due to cocaine, certain chemotherapeutic agents, antiphospholipid antibody syndrome, radiation nephritis, malignant hypertensive nephrosclerosis, thrombotic thrombocytopenic purpura/hemolytic-uremic syndrome [TTP-HUS]), scleroderma, some chemotherapeutic agents, and atheroembolic disease. Large-vessel diseases associated with AKI include renal artery dissection,

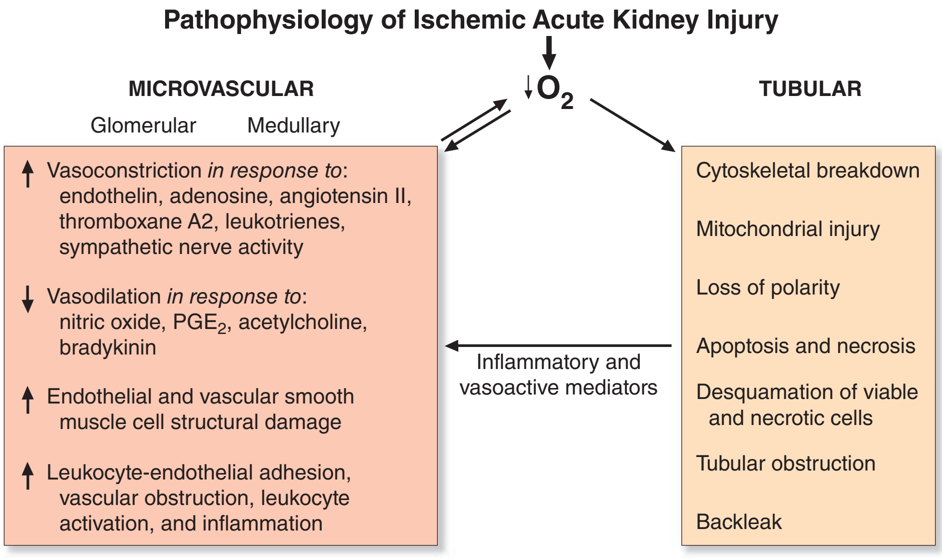
Pathophysiology of Ischemic Acute Kidney Injury
↓O$_2$
MICROVASCULAR
Glomerular
Medullary
TUBULAR
↑ Vasoconstriction *in response to*: endothelin, adenosine, angiotensin II, thromboxane A2, leukotrienes, sympathetic nerve activity
↓ Vasodilation *in response to*: nitric oxide, PGE$_2$, acetylcholine, bradykinin
↑ Endothelial and vascular smooth muscle cell structural damage
↑ Leukocyte-endothelial adhesion, vascular obstruction, leukocyte activation, and inflammation
Inflammatory and vasoactive mediators
Cytoskeletal breakdown
Mitochondrial injury
Loss of polarity
Apoptosis and necrosis
Desquamation of viable and necrotic cells
Tubular obstruction
Backleak

FIGURE 321-4 Interacting microvascular and tubular events contributing to the pathophysiology of ischemic acute kidney injury. PGE$_2$, prostaglandin E$_2$. *(Reproduced with permission from JV Bonventre, JM Weinberg. J Am Soc Nephrol. 14:2199, 2003.)*

thromboembolism, or thrombosis, and renal vein compression or thrombosis.

NEPHROTOXIN-ASSOCIATED AKI

The kidney has very high susceptibility to nephrotoxic agents due to extremely high blood perfusion and concentration of filtered substances along the nephron where filtrate water is reabsorbed. Nephrotoxic injury occurs in response to a number of pharmacologic compounds with diverse structures, endogenous substances, and environmental exposures. All structures of the kidney are vulnerable to toxic injury, including the tubules, interstitium, vasculature, and collecting system. As with other forms of AKI, risk factors for nephrotoxicity include older age, CKD, and prerenal azotemia. Hypoalbuminemia may increase the risk of some forms of nephrotoxin-associated AKI due to increased free circulating drug concentrations.

Contrast Agents Iodinated contrast agents used for cardiovascular and computed tomography (CT) imaging have been implicated as a cause of AKI. Many question whether AKI in response to contrast agents represents an important consequence of contrast studies. It is likely to have been diagnosed too frequently in the past, particularly in individuals who had many risk factors for AKI, making the cause difficult to identify. The terminology has changed so that the former “contrast nephropathy” has been replaced by “contrast-associated AKI” or “contrast-induced AKI” (CI-AKI), with the latter representing a smaller subgroup of the former. The occurrence of CI-AKI is negligible in those with normal renal function but increases in the setting of CKD, particularly in individuals with diabetic kidney disease. The most common clinical course of contrast nephropathy is characterized by a rise in SCr beginning 24–48 h following exposure, peaking within 3–5 days, and resolving within 1 week. More severe, dialysis-requiring AKI is uncommon except in the setting of significant preexisting CKD, often in association with congestive heart failure or other coexisting causes for ischemia-associated AKI. Patients with multiple myeloma and/or renal disease are particularly susceptible. Other diagnostic agents implicated as a cause of AKI are high-dose gadolinium used for magnetic resonance imaging (MRI) and oral sodium phosphate solutions used as bowel purgatives. Gadolinium has been associated with development of nephrogenic systemic fibrosis (NSF) in subjects with advanced kidney disease or AKI, but the majority of these cases were associated with group I gadolinium-based contrast media, which are rarely used now in the United States and have been withdrawn from the market in many other countries. The risk of AKI associated with standard doses of group II gadolinium-based contrast media is very low.

Antibiotics Several antimicrobial agents are commonly associated with AKI. *Vancomycin* may be associated with AKI from tubular injury, particularly when trough levels are high and when used in combination with other nephrotoxic antibiotics. Vancomycin can also crystalize in tubules and cause intratubular obstruction. *Aminoglycosides* and *amphotericin B* both cause tubular necrosis. Nonoliguric AKI (i.e., with a urine volume >400 mL/d) accompanies 10–30% of courses of aminoglycoside antibiotics, even when plasma levels are in the therapeutic range. Aminoglycosides are freely filtered across the glomerulus and then accumulate within the renal cortex, where concentrations can greatly exceed those of the plasma. AKI typically manifests after 5–7 days of therapy and can present even after the drug has been discontinued. Hypomagnesemia is a common finding.

Amphotericin B causes renal vasoconstriction from an increase in tubuloglomerular feedback as well as direct tubular toxicity mediated by reactive oxygen species. Nephrotoxicity from amphotericin B is dose and duration dependent. This drug binds to tubular membrane cholesterol and introduces pores. Clinical features of amphotericin B nephrotoxicity include polyuria, hypomagnesemia, hypocalcemia, and nongap metabolic acidosis.

Acyclovir can precipitate in tubules and cause AKI by tubular obstruction, particularly when given as an intravenous bolus at high doses (500 mg/m^2) or in the setting of hypovolemia. *Foscarnet, pentamidine, tenofovir*, and *cidofovir* are also frequently associated with AKI due to tubular toxicity. AKI secondary to acute interstitial nephritis can occur as a consequence of exposure to many antibiotics, including *penicillins, cephalosporins, quinolones, sulfonamides*, and *rifampin*.

Chemotherapeutic Agents *Cisplatin* and *carboplatin* are accumulated by proximal tubular cells and cause necrosis and apoptosis. Intensive hydration regimens have reduced the incidence of cisplatin nephrotoxicity, but it remains a dose-limiting toxicity. *Ifosfamide* may cause hemorrhagic cystitis and tubular toxicity, manifested as type II renal tubular acidosis (Fanconi syndrome), polyuria, hypokalemia, and a modest decline in GFR. Antiangiogenesis agents, such as *bevacizumab*, can cause proteinuria and hypertension via injury to the glomerular microvasculature (thrombotic microangiopathy). Other antineoplastic agents such as mitomycin C and gemcitabine may cause thrombotic microangiopathy with resultant AKI. Immune checkpoint inhibitors, such as *ipilimumab*, *tremelimumab*, *nivolumab*, and *pembrolizumab* can cause immune-related adverse events, often manifesting in the kidney as acute interstitial nephritis. Lower GFR, proton pump inhibitor use, and extrarenal immune-related adverse events are predisposing risk factors for AKI secondary to immune checkpoint inhibitors. The checkpoint inhibitors result in hyperactivity of the immune system triggered by these agents.

Toxic Ingestions Ethylene glycol, present in automobile antifreeze, is metabolized to oxalic acid, glycolaldehyde, and glyoxylate, which may cause AKI through direct tubular injury and tubular obstruction. Diethylene glycol is an industrial agent that has caused outbreaks of severe AKI around the world due to adulteration of pharmaceutical preparations. The metabolite 2-hydroxyethoxyacetic acid (HEAA) is thought to be responsible for tubular injury. Melamine contamination of foodstuffs has led to nephrolithiasis and AKI, either through intratubular obstruction or possibly direct tubular toxicity. Aristolochic acid was found to be the cause of “Chinese herb nephropathy” and “Balkan nephropathy” due to its contamination of medicinal herbs or farming. The list of environmental toxins is likely to grow and contribute to a better understanding of previously catalogued “idiopathic” chronic tubular interstitial disease, a common diagnosis in both the developed and developing world.

Endogenous Toxins AKI may be caused by a number of endogenous compounds, including myoglobin, hemoglobin, uric acid, and myeloma light chains. Myoglobin can be released by injured muscle cells, and hemoglobin can be released during massive hemolysis leading to pigment nephropathy. Rhabdomyolysis may result from traumatic crush injuries, muscle ischemia during vascular or orthopedic surgery, compression during coma or immobilization, prolonged seizure activity, excessive exercise, heat stroke or malignant hyperthermia, infections, metabolic disorders (e.g., hypophosphatemia, severe hypothyroidism), and myopathies (drug-induced, metabolic, or inflammatory). Pathogenic factors contributing to AKI upon exposure to endogenous toxins include intrarenal vasoconstriction, direct proximal tubular toxicity, and mechanical obstruction of the distal nephron lumen when myoglobin or hemoglobin precipitates with Tamm-Horsfall protein (uromodulin, the most common protein in urine and produced in the thick ascending limb of the loop of Henle), a process favored by acidic urine. Tumor lysis syndrome may follow initiation of cytotoxic therapy in patients with high-grade lymphomas and acute lymphoblastic leukemia; massive release of uric acid (with serum levels often exceeding 15 mg/dL) leads to precipitation of uric acid in the renal tubules and AKI **(Chap. 75)**. Other features of tumor lysis syndrome include hyperkalemia and hyperphosphatemia. The tumor lysis syndrome can also occasionally occur spontaneously or with treatment for solid tumors or multiple myeloma. Myeloma light chains can also cause AKI by glomerular damage and/or direct tubular toxicity and by binding to Tamm-Horsfall protein to form obstructing intratubular casts. Hypercalcemia, which can also be seen in multiple myeloma, may cause AKI by intense renal vasoconstriction and inhibition of sodium and water reabsorption in the nephron with resultant volume depletion.

Other Causes of Acute Tubulointerstitial Disease Leading to AKI While many drugs result in toxin-induced injury to the nephron with subsequent inflammation, drugs can also lead to the development of an allergic response characterized by an inflammatory infiltrate, sometimes associated with blood and urinary eosinophilia. Proton pump inhibitors and NSAIDs are commonly used drugs that have been associated with acute tubulointerstitial nephritis. AKI may be also caused by severe infections and infiltrative malignant or nonmalignant (e.g., sarcoidosis) diseases with tubulointerstitial disease.

Anticoagulant-Related Nephropathy Excessive anticoagulation with warfarin or other classes of anticoagulants has been reported to cause AKI through glomerular hemorrhage resulting in the formation of obstructing red blood cell casts within the kidney tubule and tubular injury.

Glomerulonephritis Diseases involving the glomerular podocytes, mesangial, and/or endothelial cells can lead to AKI by compromising the filtration barrier and blood flow within the renal circulation. Although glomerulonephritis is a less common (~5%) cause of AKI, early recognition is particularly important because the diseases can respond to timely treatment with immunosuppressive agents or therapeutic plasma exchange, and the treatment may reverse the AKI and decrease subsequent longer term injury.

■ POSTRENAL AKI

(See also Chap. 331.) Postrenal AKI occurs when the normally unidirectional flow of urine is acutely blocked either partially or totally, leading to increased retrograde hydrostatic pressure and interference with glomerular filtration. Obstruction to urinary flow may be caused by functional or structural derangements anywhere from the renal pelvis to the tip of the urethra (Fig. 321-5). Normal urinary flow rate does not rule out the presence of partial obstruction, because the GFR is normally two orders of magnitude higher than the urinary flow rate and hence a preservation of urine output may be misleading in hiding the postrenal partial obstruction. For moderate to severe AKI to occur in individuals with two healthy functional kidneys, obstruction must affect both kidneys in order to observe large increases in SCr, unless there is asymmetric kidney function with one chronically diseased. Unilateral obstruction may cause AKI in the setting of significant underlying CKD with loss of renal reserve or, in rare cases, from reflex vasospasm of the contralateral kidney. Bladder neck obstruction is a common cause of postrenal AKI. This can be due to prostate disease (benign prostatic hypertrophy or prostate cancer), neurogenic bladder, or therapy with anticholinergic drugs. Obstructed bladder catheters can cause postrenal AKI if not recognized. Other causes of lower tract obstruction are blood clots, calculi, and urethral strictures. Ureteric obstruction can occur from intraluminal obstruction (e.g., calculi, blood clots, sloughed renal papillae), infiltration of the ureteric wall (e.g., neoplasia), or external compression (e.g., retroperitoneal fibrosis, neoplasia, abscess, or inadvertent surgical damage). The pathophysiology of postrenal AKI involves hemodynamic alterations triggered by an abrupt increase in intratubular pressures. An initial period of hyperemia from afferent arteriolar dilation is followed by intrarenal vasoconstriction from the generation of angiotensin II, thromboxane A2, and vasopressin, and a reduction in NO production. Secondary reductions in glomerular function are due to underperfusion of glomeruli and, possibly, changes in the glomerular ultrafiltration coefficient.

DIAGNOSTIC EVALUATION (TABLE 321-2)

As described previously, AKI is defined by an elevation in the SCr concentration from baseline of at least 0.3 mg/dL within 48 h or at least 50% within 1 week, or a reduction in urine output to <0.5 mL/kg per h for longer than 6 h. Serum cystatin C is increasingly being used to estimate GFR and may have a role in AKI diagnosis; both SCr and cystatin C have distinct non-GFR determinants that can influence their sensitivity and specificity. As indicated previously, some patients with AKI will not have tubular or glomerular damage (e.g., prerenal azotemia). The distinction between AKI and CKD is important for proper diagnosis and treatment. CKD is defined by an estimated GFR <60 mL/min per 1.73 m^2 or an albumin-to-creatinine ratio (ACR) of >30 mg/g for a period of at least 3 months. If the diagnosis of AKI is made and renal dysfunction persists for more than a week but not yet 3 months, then some refer to this renal dysfunction as acute kidney disease. The distinction between AKI and CKD is straightforward when a recent baseline SCr concentration is available, but more difficult in the many instances in which the baseline is unknown. In such cases, clues suggestive of CKD can come from radiologic studies (e.g., small, shrunken kidneys with cortical thinning on renal ultrasound) or laboratory tests such as normocytic anemia in the absence of blood loss or secondary hyperparathyroidism with hyperphosphatemia and hypocalcemia, consistent with CKD. No set of tests, however, can rule out AKI superimposed on CKD because AKI is a frequent complication in patients with CKD, further complicating the distinction. Serial blood tests showing a continued substantial rise of SCr represent clear evidence of AKI. Once the diagnosis of AKI is established, its cause needs to be determined because the elevation of SCr or reduction in urine output can be due to a large number of physiologic and pathophysiologic processes, as described previously. Increasingly, the electronic medical record is being utilized for automated alerts to identify AKI and artificial intelligence approaches for AKI prediction. The role of automated alerts and artificial intelligence to predict and/or identify AKI is an area of active investigation.

■ HISTORY AND PHYSICAL EXAMINATION

The clinical context, careful history taking, and physical examination often narrow the differential diagnosis for the cause of AKI. Prerenal azotemia should be suspected in the setting of vomiting, diarrhea, glycosuria causing polyuria, and several medications

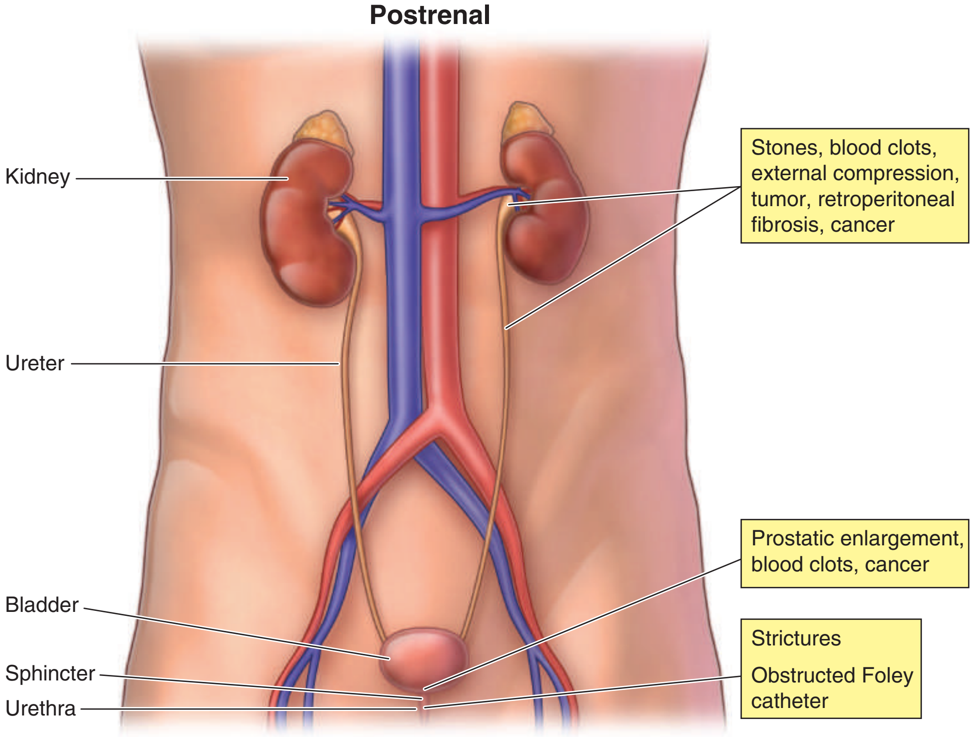
Postrenal
Kidney
Ureter
Bladder
Sphincter
Urethra
Stones, blood clots, external compression, tumor, retroperitoneal fibrosis, cancer
Prostatic enlargement, blood clots, cancer
Strictures
Obstructed Foley catheter

FIGURE 321-5 Anatomic sites and causes of obstruction leading to postrenal acute kidney injury.

TABLE 321-2 Major Causes, Clinical Features, and Diagnostic Studies for Prerenal and Intrinsic Acute Kidney Injury

ETIOLOGY	CLINICAL FEATURES	LABORATORY FEATURES	COMMENTS
Prerenal azotemia	History of poor fluid intake or fluid loss (hemorrhage, diarrhea, vomiting, sequestration into extravascular space); NSAID/ACE-I/ARB; heart failure; evidence of volume depletion (tachycardia, absolute or postural hypotension, low jugular venous pressure, dry mucous membranes), decreased effective circulatory volume (cirrhosis, heart failure)	BUN/creatinine ratio above 20, FeNa <1%, hyaline casts in urine sediment, urine specific gravity >1.018, urine osmolality >500 mOsm/kg	Low FeNa, high specific gravity and osmolality may not be seen in the setting of CKD, diuretic use; BUN elevation out of proportion to creatinine may alternatively indicate upper GI bleed or increased catabolism. Response to restoration of hemodynamics is most diagnostic.
Sepsis-associated AKI	Sepsis, sepsis syndrome, or septic shock; overt hypotension not always seen in mild to moderate AKI	Positive culture from normally sterile body fluid or other test confirming infection; urine sediment often contains granular casts, renal tubular epithelial cell casts	FeNa may be low (<1%), particularly early in the course, but is usually >1% with osmolality <500 mOsm/kg
Ischemia-associated AKI	Systemic hypotension, often superimposed upon sepsis and/or reasons for limited renal reserve such as older age, CKD	Urine sediment often contains granular casts, renal tubular epithelial cell casts; FeNa typically >1%	
Nephrotoxin-Associated AKI: Endogenous			
Rhabdomyolysis	Traumatic crush injuries, seizures, immobilization	Elevated myoglobin, creatine kinase; urine heme positive with few red blood cells	FeNa may be low (<1%)
Hemolysis	Recent blood transfusion with transfusion reaction	Anemia, elevated LDH, low haptoglobin	FeNa may be low (<1%); evaluation for transfusion reaction
Tumor lysis	Recent chemotherapy	Hyperphosphatemia, hypocalcemia, hyperuricemia	
Multiple myeloma	Age >60 years, constitutional symptoms, bone pain	Monoclonal spike in urine or serum electrophoresis; elevated serum free light chains, low anion gap; anemia	Bone marrow or renal biopsy can be diagnostic
Nephrotoxin-Associated AKI: Exogenous			
Contrast nephropathy	Exposure to iodinated contrast	Characteristic course is rise in SCr within 1–2 d, peak within 3–5 d, recovery within 7 d	FeNa may be low (<1%)
Tubular injury	Aminoglycoside antibiotics, cisplatin, tenofovir, vancomycin, zoledronate, ethylene glycol, aristolochic acid, and melamine (to name a few)	Urine sediment often contains granular casts, renal tubular epithelial cell casts. FeNa typically >1%.	Can be oliguric or nonoliguric
Other Causes of Intrinsic AKI			
Glomerulonephritis/vasculitis	Variable **(Chap. 326)** features include skin rash, arthralgias, sinusitis (AGBM disease), lung hemorrhage (AGBM, ANCA, lupus), recent skin infection or pharyngitis (poststreptococcal), thrombotic microangiopathies including those related to drugs, such as cocaine, anti-VEGF agents, genetic abnormalities of the complement pathways	ANA, ANCA, Anti-GBM antibody, hepatitis serologies, cryoglobulins, blood culture, complement abnormalities, ASO titer (abnormalities of these tests depending on etiology)	Kidney biopsy may be necessary
Tubulointerstitial nephritis	Drugs are responsible for about 75% of the biopsy-proven acute interstitial nephritis, which involves tubules in most cases. Examples of causes include antibiotics, PPIs, immune checkpoint inhibitors. Non-drug-related causes include tubulointerstitial nephritis-uveitis (TINU) syndrome, lupus, viral infection (e.g., COVID, HIV, hantavirus), and *Legionella* infection.	Eosinophilia, sterile pyuria; often nonoliguric	Urine eosinophils have limited diagnostic accuracy; kidney biopsy may be necessary
TTP/HUS	Neurologic abnormalities and/or AKI; recent diarrheal illness; use of calcineurin inhibitors; pregnancy or postpartum; spontaneous	Schistocytes on peripheral blood smear, elevated LDH, anemia, thrombocytopenia	"Typical HUS" refers to AKI with a diarrheal prodrome, often due to Shiga toxin released from *Escherichia coli* or other bacteria; "atypical HUS" is due to inherited or acquired complement dysregulation. Diagnosis may involve screening for ADAMTS13 activity, Shiga toxin–producing *E. coli*, genetic evaluation of complement regulatory proteins, and kidney biopsy.
Atheroembolic disease	Recent manipulation of the aorta or other large vessels; may occur spontaneously or after anticoagulation; retinal plaques, palpable purpura, livedo reticularis, GI bleed	Hypocomplementemia, eosinophiliuria (variable), variable amounts of proteinuria	Skin or kidney biopsy can be diagnostic
Postrenal AKI	History of kidney stones, prostate disease, obstructed bladder catheter, retroperitoneal or pelvic neoplasm	No specific findings other than AKI; may have pyuria or hematuria	Imaging with computed tomography or ultrasound

Abbreviations: ACE-I, angiotensin-converting enzyme inhibitor-I; AGBM, antiglomerular basement membrane; AKI, acute kidney injury; ANA, antinuclear antibody; ANCA, antineutrophilic cytoplasmic antibody; ARB, angiotensin receptor blocker; ASO, antistreptolysin O; BUN, blood urea nitrogen; CKD, chronic kidney disease; FeNa, fractional excretion of sodium; GI, gastrointestinal; LDH, lactate dehydrogenase; NSAID, nonsteroidal anti-inflammatory drug; PPI, proton pump inhibitors; TTP/HUS, thrombotic thrombocytopenic purpura/hemolytic-uremic syndrome.

including diuretics, NSAIDs, ACE inhibitors, and ARBs. Physical signs of orthostatic hypotension, tachycardia, reduced jugular venous pressure, decreased skin turgor, and dry mucous membranes are often present in prerenal azotemia. Congestive heart failure, liver disease, and nephrotic syndrome can be associated with reductions in renal blood flow and/or alterations in intrarenal hemodynamics leading to reduced GFR. Extensive vascular disease raises the possibility of renal artery disease, especially if kidneys are known to be asymmetric in size. Atheroembolic disease can be associated with livedo reticularis and other signs of emboli to the legs. The presence of sepsis is an important clue to causation, although, as described above, the detailed pathophysiology may be multifactorial.

A history of prostatic disease, nephrolithiasis, or pelvic or paraaortic malignancy would suggest the possibility of postrenal AKI. Whether or not symptoms are present early during obstruction of the urinary tract depends on the location of obstruction. Colicky flank pain radiating to the groin suggests acute ureteric obstruction. Nocturia and urinary frequency or hesitancy can be seen in prostatic disease. Abdominal fullness and suprapubic pain can accompany bladder enlargement. Definitive diagnosis of obstruction requires radiologic investigations.

A careful review of all medications is imperative in the evaluation of an individual with AKI. Not only are medications frequently a nephrotoxic cause of AKI, but doses of administered medications must be adjusted for reductions in kidney function. In this regard, it is important to recognize that reductions in true GFR are not reflected by equations that estimate GFR because those equations are dependent on SCr and the patient being in a steady state. With AKI, changes in SCr will lag behind changes in filtration rate. Allergic interstitial nephritis may be accompanied by fever, arthralgias, and a pruritic erythematous rash. The absence of systemic features of hypersensitivity, however, does not exclude the diagnosis of interstitial nephritis, and a kidney biopsy should be considered for definitive diagnosis when the cause of AKI is not apparent from the clinical presentation.

AKI accompanied by palpable purpura, pulmonary hemorrhage, or sinusitis raises the possibility of systemic vasculitis with glomerulonephritis. A history of autoimmune disease, such as systemic lupus erythematosus, should lead to consideration of the possibility that the AKI is related to worsening of this underlying disease. Pregnancy should lead to the consideration of preeclampsia as a pathophysiologic contributor to the AKI. A tense abdomen should prompt consideration of acute abdominal compartment syndrome, a diagnosis facilitated by measurement of bladder pressure. Signs and/or symptoms of limb ischemia may be clues to the diagnosis of rhabdomyolysis.

URINE FINDINGS

Complete anuria early in the course of AKI is uncommon except in the following situations: complete urinary tract obstruction, renal artery occlusion, overwhelming septic shock, severe ischemia (often with cortical necrosis), or severe proliferative glomerulonephritis or vasculitis. A reduction in urine output (oliguria, defined as <400 mL/24 h) usually denotes more severe AKI (i.e., lower GFR) than when urine output is preserved. Oliguria is associated with worse clinical outcomes in AKI. Preserved urine output can be seen in nephrogenic diabetes insipidus characteristic of long-standing urinary tract obstruction, tubulointerstitial disease, or nephrotoxicity from cisplatin or aminoglycosides, among other causes. Red or brown urine may be seen with or without gross hematuria; if the color persists in the supernatant after centrifugation, then pigment nephropathy from rhabdomyolysis or hemolysis should be suspected.

The urinalysis and urine sediment examination are invaluable tools, but they require clinical correlation because of generally limited sensitivity and specificity **(see Fig. 321-6 and Chap. A4)**. In the absence of preexisting proteinuria from CKD, AKI secondary to ischemia or nephrotoxins leads to mild proteinuria (<1 g/d). Greater proteinuria in AKI suggests damage to the glomerular ultrafiltration barrier or excretion of small molecular weight proteins such as myeloma light chains; the latter are not detected with conventional urine dipsticks (which detect albumin) and require the sulfosalicylic acid test or immunoelectrophoresis. Atheroemboli can cause a variable degree of proteinuria. Heavy proteinuria ("nephrotic range," >3.5 g/d) can occasionally be seen in glomerulonephritis, vasculitis, or toxins/medications that can affect the glomerulus as well as the tubulointerstitium (e.g., NSAIDs). AKI can also complicate cases of minimal change disease, a cause of the nephrotic syndrome often associated with low serum albumin concentrations **(Chap. 320)**.

Prerenal azotemia may present with hyaline casts or an unremarkable urine sediment examination. Postrenal AKI may also be associated with an unremarkable sediment, but hematuria and pyuria may be seen depending on the cause of obstruction. AKI from ATN due to ischemic injury, sepsis, or certain nephrotoxins has characteristic urine sediment

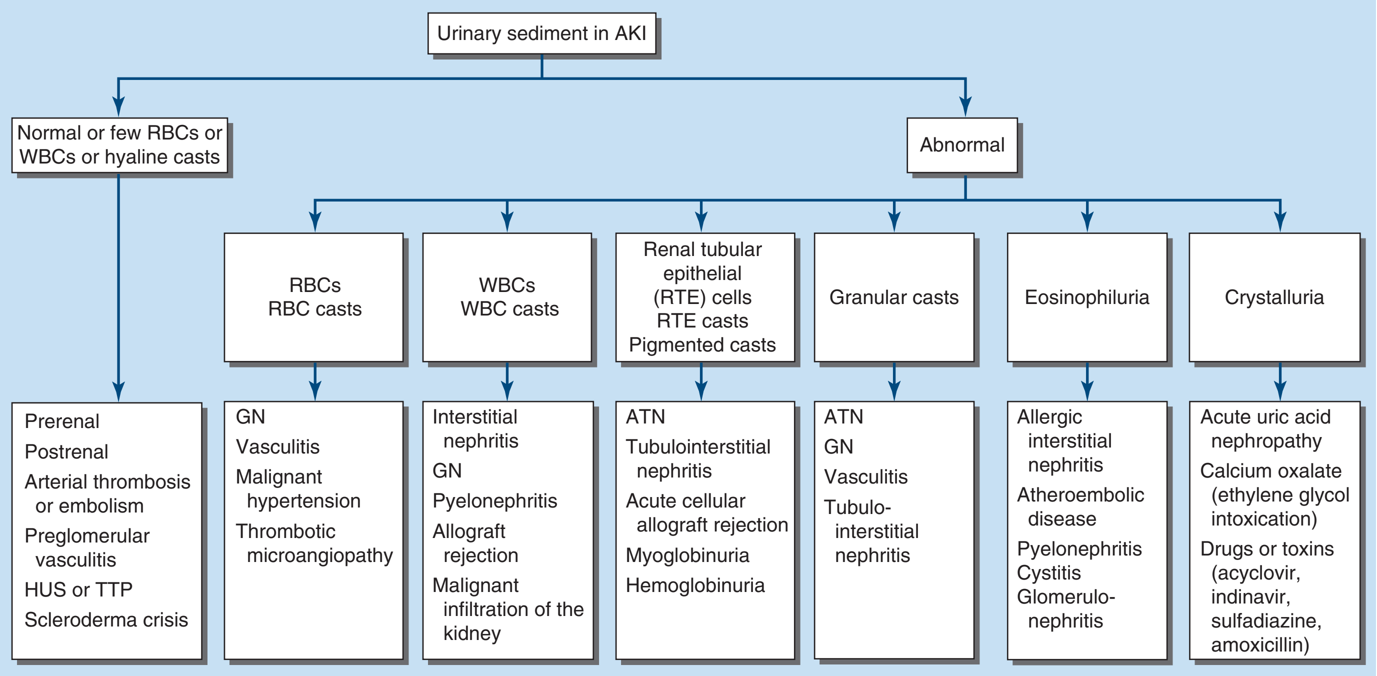
Urinary sediment in AKI
Normal or few RBCs or WBCs or hyaline casts
Abnormal
RBCs RBC casts
WBCs WBC casts
Renal tubular epithelial (RTE) cells RTE casts Pigmented casts
Granular casts
Eosinophiluria
Crystalluria
Prerenal
Postrenal
Arterial thrombosis or embolism
Preglomerular vasculitis
HUS or TTP
Scleroderma crisis
GN
Vasculitis
Malignant hypertension
Thrombotic microangiopathy
Interstitial nephritis
GN
Pyelonephritis
Allograft rejection
Malignant infiltration of the kidney
ATN
Tubulointerstitial nephritis
Acute cellular allograft rejection
Myoglobinuria
Hemoglobinuria
ATN
GN
Vasculitis
Tubulo-interstitial nephritis
Allergic interstitial nephritis
Atheroembolic disease
Pyelonephritis
Cystitis
Glomerulo-nephritis
Acute uric acid nephropathy
Calcium oxalate (ethylene glycol intoxication)
Drugs or toxins (acyclovir, indinavir, sulfadiazine, amoxicillin)

FIGURE 321-6 Interpretation of urinary sediment findings in acute kidney injury (AKI). ATN, acute tubular necrosis; GN, glomerulonephritis; HUS, hemolytic-uremic syndrome; RBCs, red blood cells; RTE, renal tubular epithelial; TTP, thrombotic thrombocytopenic purpura; WBCs, white blood cells. *(Adapted from L Yang, JV Bonventre: Diagnosis and clinical evaluation of acute kidney injury. In Comprehensive Nephrology, 4th ed. J Floege et al [eds]. Philadelphia, Elsevier, 2010.)*

findings: pigmented "muddy brown" granular casts and tubular epithelial cell casts. These findings may be absent in >20% of cases, however. Glomerulonephritis may lead to dysmorphic red blood cells or red blood cell casts. Interstitial nephritis may lead to white blood cell casts. The urine sediment findings overlap somewhat in glomerulonephritis and interstitial nephritis, and a diagnosis is not always possible on the basis of the urine sediment alone. Urine eosinophils have a limited role in differential diagnosis; they can be seen in interstitial nephritis, pyelonephritis, cystitis, atheroembolic disease, or glomerulonephritis. Crystalluria may be important diagnostically. The finding of oxalate crystals in AKI should prompt an evaluation for ethylene glycol toxicity. Abundant uric acid crystals may be seen in tumor lysis syndrome.

BLOOD LABORATORY FINDINGS

Certain forms of AKI are associated with characteristic patterns in the rise and fall of SCr. Prerenal azotemia typically leads to modest rises in SCr that return to baseline with improvement in hemodynamic status. In comparison, atheroembolic disease usually manifests with more subacute rises in SCr, although severe AKI with rapid increases in SCr can occur in this setting. With many of the epithelial cell toxins such as aminoglycoside antibiotics and cisplatin, the rise in SCr is characteristically delayed for 3–5 days to 2 weeks after initial exposure.

A complete blood count may provide diagnostic clues. Anemia is common in AKI and is usually multifactorial in origin. It is not related to an effect of AKI solely on production of red blood cells because this effect in isolation takes longer to manifest. Myeloma can be diagnosed with serum immunoelectrophoresis or free light chain assay, and it can often be suspected if the blood anion gap is low due to unmeasured cationic proteins. Peripheral eosinophilia can accompany interstitial nephritis, atheroembolic disease, polyarteritis nodosa, and Churg-Strauss vasculitis. Severe anemia in the absence of bleeding may reflect hemolysis, multiple myeloma, or thrombotic microangiopathy (e.g., HUS or TTP). Other laboratory findings of thrombotic microangiopathy include thrombocytopenia, schistocytes on peripheral blood smear, elevated lactate dehydrogenase level, and low haptoglobin content. Evaluation of patients suspected of having TTP or HUS includes measurement of levels of the von Willebrand factor cleaving protease (ADAMTS13) and testing for Shiga toxin–producing *Escherichia coli*. "Atypical HUS" constitutes the majority of adult cases of HUS; genetic testing is important because it is estimated that 60–70% of atypical HUS patients have mutations in genes encoding proteins that regulate the alternative complement pathway.

AKI often leads to hyperkalemia, hyperphosphatemia, and hypocalcemia. Marked hyperphosphatemia with accompanying hypocalcemia may suggest rhabdomyolysis or tumor lysis syndrome. Serum creatine kinase and uric acid levels are often elevated in rhabdomyolysis, while tumor lysis syndrome can be associated with normal or marginally elevated creatine kinase and markedly elevated serum uric acid. The anion gap may be increased with any cause of uremia due to retention of anions such as phosphate, hippurate, sulfate, and urate. The co-occurrence of an increased anion gap and an osmolal gap may suggest ethylene glycol poisoning, which may also cause oxalate crystalluria and oxalate deposition in kidney tissue. As discussed previously, low anion gap may provide a clue to the diagnosis of multiple myeloma due to the presence of unmeasured cationic proteins. Laboratory blood tests helpful for the diagnosis of glomerulonephritis and vasculitis include depressed complement levels and high titers of antinuclear antibodies (ANAs), antineutrophil cytoplasmic antibodies (ANCAs), antiglomerular basement membrane (anti-GBM) antibodies, and cryoglobulins. It is important to diagnose glomerulonephritis or myeloma early in the course of AKI since effective therapies (e.g., immunosuppression or chemotherapy) are now available for some of the causes. In general, however, the therapies are less effective when severe kidney injury has progressed.

RENAL FAILURE INDICES

Several indices have been used to help differentiate prerenal azotemia from intrinsic AKI when the tubules are malfunctioning. The low tubular flow rate and increased renal medullary recycling of urea seen in prerenal azotemia may cause a disproportionate elevation of the blood urea nitrogen (BUN) compared to creatinine. Other causes of disproportionate BUN elevation need to be kept in mind, however, including upper gastrointestinal bleeding, hyperalimentation, increased tissue catabolism, and glucocorticoid use.

The FeNa is the fraction of the filtered sodium load that is not reabsorbed by the tubules and is a measure of both the kidney's ability to reabsorb sodium as well as endogenously and exogenously administered factors that affect tubular reabsorption. As such, it depends on sodium intake, effective intravascular volume, GFR, diuretic intake, and intact tubular reabsorptive mechanisms. With prerenal azotemia, the FeNa may be <1%, suggesting avid tubular sodium reabsorption. In patients with CKD, a FeNa significantly >1% can be present due to tubular dysfunction in the reabsorption of sodium, despite a superimposed prerenal state. The FeNa may also be >1% despite hypovolemia due to treatment with diuretics. Low FeNa is often seen early in glomerulonephritis and other disorders and, hence, should not be taken as prima facie evidence of prerenal azotemia. Low FeNa is therefore suggestive of, but not synonymous with, effective intravascular volume depletion, and should not be used as the sole guide for volume management. The response of urine output to crystalloid or colloid fluid administration may be both diagnostic and therapeutic in prerenal azotemia. In ischemic AKI, the FeNa is frequently >1% because of tubular injury and resultant impaired ability to reabsorb sodium. Several causes of ischemia-associated and nephrotoxin-associated AKI can present with FeNa <1%, however, including sepsis (often early in the course), rhabdomyolysis, and contrast nephropathy. FeNa has the most utility in oliguric patients who are not given diuretics and do not have CKD.

The ability of the kidney to produce a concentrated urine is dependent upon many factors and relies on good blood flow and tubular function in multiple regions of the kidney. In the patient not taking diuretics and with good baseline kidney function, urine osmolality may be >500 mOsm/kg in prerenal azotemia, consistent with an intact medullary concentration gradient and elevated serum vasopressin levels causing water reabsorption by passive diffusion from the collecting duct into a concentrated medullary interstitium, resulting in concentrated urine. In elderly patients and those with CKD, however, baseline concentrating defects may exist, making urinary osmolality unreliable in many instances. Concentrating ability may also be maintained early in the course of glomerular disease when the tubules are not yet affected. Loss of concentrating ability (<350 mOsm/kg) is common in most forms of AKI that affect the tubules and interstitium, but this finding is not specific.

RADIOLOGIC EVALUATION

Postrenal AKI should always be considered in the differential diagnosis of AKI because treatment is usually successful if instituted early. Simple bladder catheterization can rule out urethral obstruction. Imaging of the urinary tract with renal ultrasound or CT scan should be undertaken to investigate obstruction in individuals with AKI unless an alternate diagnosis is apparent. Findings of obstruction include dilation of the collecting system and hydroureteronephrosis. Obstruction can be present without radiologic abnormalities in the setting of volume depletion, retroperitoneal fibrosis, encasement with tumor, and also early in the course of obstruction. If a high clinical index of suspicion for obstruction persists despite normal imaging, antegrade or retrograde pyelography should be performed. Imaging may also provide additional helpful information about kidney size and echogenicity to assist in the distinction between AKI versus CKD. In CKD, kidneys are usually smaller unless the patient has diabetic nephropathy, HIV-associated nephropathy, or infiltrative diseases. Normal-sized kidneys are expected in AKI. Enlarged kidneys in a patient with AKI suggest the possibility of acute interstitial nephritis or infiltrative diseases. Vascular imaging may be useful if venous or arterial obstruction is suspected.

KIDNEY BIOPSY

If the cause of AKI is not apparent based on the clinical context, physical examination, laboratory studies, and radiologic evaluation, kidney biopsy should be considered. The kidney biopsy can provide definitive diagnostic and prognostic information about acute kidney disease and CKD. The procedure is most often used in AKI when prerenal azotemia, postrenal AKI, and ischemic or nephrotoxic AKI have been deemed unlikely and other possible diagnoses are being considered such as glomerulonephritis, vasculitis, interstitial nephritis, myeloma kidney, HUS and TTP, and allograft dysfunction. Kidney biopsy is associated with a risk of bleeding, which can be severe and organ- or life-threatening in patients with thrombocytopenia or coagulopathy, but the diagnostic and prognostic information obtained can be invaluable.

NOVEL BIOMARKERS

BUN, SCr, and cystatin C are functional biomarkers of glomerular filtration rather than tissue injury biomarkers and, therefore, may be suboptimal for the diagnosis of actual parenchymal kidney damage. Filtration markers are also relatively slow to rise after kidney injury. Several urine and blood biomarkers have been investigated and show promise for earlier and accurate diagnosis of AKI and for predicting AKI prognosis. In cases of oliguric AKI, the urinary flow rate in response to bolus intravenous furosemide 1.0–1.5 mg/kg can be used as a prognostic test: urine output <200 mL over 2 h after intravenous furosemide may identify patients at higher risk of progression to more severe AKI and the need for renal replacement therapy. The severity or risk of progressive AKI may also be reflected in findings on urine microscopy. In one study involving review of fresh urine sediments by board-certified nephrologists, a greater number of renal tubular epithelial cells and/or granular casts in the urine sediment was associated with both the severity and worsening of AKI. Protein biomarkers of kidney injury have also been identified in animal models of AKI and have been used in humans and found to be particularly useful in toxicity identification. *Kidney injury molecule-1* (KIM-1) is a type 1 transmembrane protein that is abundantly expressed in proximal tubular cells injured by ischemia or multiple, distinct nephrotoxins, such as cisplatin. KIM-1 is not expressed in appreciable quantities in the absence of tubular injury or in extrarenal tissues. KIM-1 can be detected after ischemic or nephrotoxic injury in the urine and plasma. *Neutrophil gelatinase associated lipocalin* (NGAL, also known as lipocalin-2 or siderocalin) is another biomarker of AKI. NGAL was first discovered as a protein in granules of human neutrophils. NGAL can bind to iron siderophore complexes and may have tissue-protective effects in the proximal tubule. NGAL is highly upregulated after inflammation and kidney injury and can be detected in the plasma and urine within 2 h of cardiopulmonary bypass–associated AKI. Soluble urokinase plasminogen activator receptor (suPAR) is a signaling glycoprotein expressed in multiple cell types and thought to be involved in the pathogenesis of certain kidney diseases; suPAR has been measured in the plasma and found to predict the subsequent development of AKI. In 2014, the U.S. Food and Drug Administration (FDA) approved the marketing of a test based on the combination of the urinary concentrations of two cell-cycle arrest biomarkers, insulin-like growth factor binding protein 7 (IGFBP7) and tissue inhibitor of metalloproteinase-2 (TIMP-2) as predictive biomarkers for higher risk of the development of moderate to severe AKI in critically ill patients. In 2023, the FDA also approved the use of NGAL for the early identification of AKI in pediatric patients. Biomarkers may also be helpful in distinguishing tubulointerstitial nephritis where interstitial inflammation plays a dominant role from other causes of AKI that primarily affect the glomeruli or the tubule where inflammation also exists but may be less dominant. One such marker, CXCL9, has recently been reported. The optimal use of AKI biomarkers in clinical settings is an area of ongoing investigation.

COMPLICATIONS OF AKI

The kidney plays a central role in homeostatic control of volume status, blood pressure, plasma electrolyte composition, acid-base balance, and the excretion of nitrogenous and other waste products. Complications associated with AKI are, therefore, protean, and depend on the severity of AKI and other associated conditions. Mild to moderate AKI may be entirely asymptomatic, particularly early in the course.

UREMIA

Buildup of nitrogenous waste products, manifested as an elevated BUN concentration, is a hallmark of AKI. BUN itself poses little direct toxicity at levels <100 mg/dL. In more severe AKI or when, as is often the case, AKI is on the backdrop of CKD, mental status changes and bleeding complications can arise. Other toxins normally cleared by the kidney may be responsible for the symptom complex known as uremia, which literally means "urine in the blood." Urea has direct and indirect toxic effects. There are increased blood levels of parathyroid hormone, advanced glycosylation end products, and many other "middle molecules" that contribute to the uremic syndrome. Few of the many possible uremic toxins have been definitively identified. The correlation of filtration markers or BUN concentrations with uremic symptoms is extremely variable, due in part to differences in generation rates across individuals.

HYPERVOLEMIA AND HYPOVOLEMIA

Expansion of extracellular fluid volume is a major complication of oliguric and anuric AKI, due to impaired salt and water excretion. The result can be weight gain, dependent edema, increased jugular venous pressure, and pulmonary edema. AKI may also induce or exacerbate acute lung injury characterized by increased vascular permeability and inflammatory cell infiltration in lung parenchyma. Recovery from AKI is often heralded by an increase in urine output. This "polyuric" phase of recovery may be due to an osmotic diuresis from retained urea and other waste products as well as delayed recovery of tubular reabsorptive functions.

HYPONATREMIA

Abnormalities in plasma electrolyte composition can be mild or life-threatening. The dysfunctional kidney has limited ability to regulate electrolyte balance. Administration of excessive hypotonic crystalloid or isotonic dextrose solutions can result in hypoosmolality and hyponatremia, which, if severe, can cause neurologic abnormalities, including seizures.

HYPERKALEMIA

An important complication of AKI is hyperkalemia. Marked hyperkalemia is particularly common in rhabdomyolysis, hemolysis, and tumor lysis syndrome due to release of intracellular potassium from damaged cells. Muscle weakness may be a symptom of hyperkalemia. Potassium affects the cellular membrane potential of cardiac and neuromuscular tissues. The more serious complication of hyperkalemia is due to effects on cardiac conduction, leading to potentially fatal arrhythmias.

ACIDOSIS

Metabolic acidosis, usually accompanied by an elevation in the anion gap, is common in AKI, and can further complicate acid-base and potassium balance in individuals with other causes of acidosis, including sepsis, diabetic ketoacidosis, or respiratory acidosis.

HYPERPHOSPHATEMIA AND HYPOCALCEMIA

AKI can lead to hyperphosphatemia, particularly in highly catabolic patients or those with AKI from rhabdomyolysis, hemolysis, and tumor lysis syndrome. Metastatic deposition of calcium phosphate can lead to hypocalcemia. AKI-associated hypocalcemia may also arise from derangements in the vitamin D–parathyroid hormone–fibroblast growth factor-23 axis. Hypocalcemia is often asymptomatic but can lead to perioral paresthesias, muscle cramps, seizures, carpopedal spasms, and prolongation of the QT interval on electrocardiography. Calcium levels should be corrected for the degree of hypoalbuminemia, if present, or ionized calcium levels should be followed. Mild, asymptomatic hypocalcemia does not require treatment.

BLEEDING

Hematologic complications of AKI include anemia and bleeding, both of which are exacerbated by coexisting disease processes such as sepsis, liver disease, and disseminated intravascular coagulation. Direct hematologic effects from AKI-related uremia include decreased erythropoiesis and platelet dysfunction.

INFECTIONS

Infections are a common precipitant of AKI and also a dreaded complication of AKI. Impaired host immunity has been described in ESKD and may be operative in severe AKI.

CARDIAC COMPLICATIONS

The major cardiac complications of AKI are arrhythmias, pericarditis, and pericardial effusion. In addition, volume overload and uremia may lead to cardiac injury and impaired cardiac function. In animal studies, cellular apoptosis and capillary vascular congestion as well as mitochondrial dysfunction have been described in the heart after renal ischemia reperfusion.

MALNUTRITION

AKI is often a severely hypercatabolic state, and therefore, malnutrition is a major complication.

PREVENTION AND TREATMENT OF AKI

The management of individuals with and at risk for AKI varies according to the underlying cause **(Table 321-3)**. Common to all are several principles. Optimization of hemodynamics, correction of fluid and electrolyte imbalances, discontinuation of nephrotoxic medications, and dose adjustment of administered medications are all critical. Common causes of AKI such as sepsis and ischemic ATN do not yet have specific therapies once injury is established, but meticulous clinical attention is needed to support the patient until (if) AKI resolves. The kidney possesses remarkable capacity to repair itself after even severe, dialysis-requiring AKI, when baseline renal function was previously intact. However, many patients with AKI, particularly when superimposed on preexisting CKD, undergo maladaptive repair processes and do not recover fully and may remain dialysis dependent. It has become increasingly apparent that AKI predisposes to accelerated progression of CKD, and CKD is an important risk factor for AKI.

Prerenal Azotemia Prevention and treatment of prerenal azotemia require optimization of renal perfusion. In AKI, oliguria alone is not an indication for fluid administration. Intravascular hypovolemia should be the only indication. The composition of replacement fluids should be targeted to the type of fluid lost. Severe acute blood loss should be treated with packed red blood cells. Crystalloids are in general favored over colloid-containing solutions (e.g., hyperoncotic albumin-containing solutions, which are still commonly used for volume resuscitation in liver failure). The colloidal solution hydroxyethyl starch is no longer available for hospitalized patients, due in part to concerns over increased risk of AKI. The most commonly used crystalloid solution is 0.9% saline. Other options are buffered crystalloid solutions (e.g., Ringer's lactate, Hartmann's solution, Plasma-Lyte). The choice between 0.9% saline or buffered crystalloid solutions can be based on serum electrolyte (e.g., some buffered crystalloid solutions are slightly hypotonic to plasma water and may be preferred for patients with hypernatremia; 0.9% saline can cause or exacerbate hyperchloremic metabolic acidosis and can be used in those with metabolic alkalosis; bicarbonate-containing solutions like dextrose water with 150 mEq sodium bicarbonate can be used in those with metabolic acidosis).

Optimization of cardiac function in AKI may require use of inotropic agents, preload- and afterload-reducing agents, antiarrhythmic drugs, and mechanical aids such as ventricular assist devices. Invasive hemodynamic monitoring to guide therapy may be necessary.

Cirrhosis and Hepatorenal Syndrome Fluid management in individuals with cirrhosis, ascites, and AKI is challenging because of the frequent difficulty in ascertaining intravascular volume status. Administration of intravenous fluids as a volume challenge may be required diagnostically as well as therapeutically. Excessive volume administration may, however, result in worsening ascites and pulmonary compromise in the setting of hepatorenal syndrome or AKI due to superimposed spontaneous bacterial peritonitis. Peritonitis should be ruled out by inflammatory cell count and culture of ascitic fluid. Albumin may prevent AKI in those treated with antibiotics for spontaneous bacterial peritonitis. The definitive treatment of the hepatorenal syndrome is orthotopic liver transplantation. Bridge therapies that have shown promise include norepinephrine, terlipressin (a vasopressin analogue), or combination therapy with octreotide (a somatostatin analogue) and midodrine (an α_1-adrenergic agonist), in combination with intravenous albumin (25–50 g, maximum 100 g/d).

TABLE 321-3 Management of Acute Kidney Injury

General Issues

1. Optimization of systemic and renal hemodynamics through volume resuscitation and judicious use of vasopressors
2. Maintain mean arterial pressure >65 mmHg
3. Elimination of nephrotoxic agents (e.g., ACE inhibitors, ARBs, NSAIDs, aminoglycosides, chemotherapeutic agents, checkpoint inhibitors, antibiotics) if possible
4. Sufficient protein and calorie intake (20–30 kcal/kg per day) to avoid negative nitrogen balance. Nutrition should be provided by the enteral route if oral intake is not possible.
5. Initiation of renal replacement therapy when indicated

Specific Issues

1. Nephrotoxin-specific
   a. Rhabdomyolysis: aggressive intravenous fluids; consider forced alkaline diuresis
   b. Tumor lysis syndrome: aggressive intravenous fluids and allopurinol or rasburicase
2. Volume overload
   a. Salt and water restriction
   b. Diuretics
   c. Ultrafiltration
3. Hyponatremia
   a. Restriction of enteral free water intake, minimization of hypotonic intravenous solutions including those containing dextrose
   b. Hypertonic saline is rarely necessary in AKI. Vasopressin antagonists are generally not needed.
4. Hyperkalemia
   a. Restriction of dietary potassium intake
   b. Discontinuation of potassium-sparing diuretics, ACE inhibitors, ARBs, NSAIDs
   c. Loop diuretics to promote urinary potassium loss
   d. Potassium-binding molecules (sodium zirconium cyclosilicate) or ion-exchange resins (patiromer, sodium or calcium polystyrene sulfonate)
   e. Insulin and glucose to promote entry of potassium intracellularly
   f. Inhaled beta-agonist therapy to promote entry of potassium intracellularly
   g. Calcium gluconate or calcium chloride (1 g) to stabilize the myocardium
5. Metabolic acidosis
   a. Sodium bicarbonate (if pH <7.2 to keep serum bicarbonate >15 mmol/L)
   b. Renal replacement therapy
6. Hyperphosphatemia
   a. Restriction of dietary phosphate intake
   b. Phosphate binding agents (calcium acetate, sevelamer hydrochloride, aluminum hydroxide—taken with meals)
7. Hypocalcemia
   a. Calcium carbonate or calcium gluconate if symptomatic
8. Hypermagnesemia
   a. Discontinue Mg^{2+}-containing antacids
9. Hyperuricemia
   a. Acute treatment is usually not required except in the setting of tumor lysis syndrome (see above)
10. Drug dosing
    a. Careful attention to dosages and frequency of administration of drugs, adjustment for degree of renal failure
    b. Note that serum creatinine concentration may overestimate renal function in the non-steady-state characteristic of patients with AKI

Abbreviations: ACE, angiotensin-converting enzyme; AKI, acute kidney infection; ARBs, angiotensin receptor blockers; NSAIDs, nonsteroidal anti-inflammatory drugs.

Intrinsic AKI Several agents have been tested and have failed to show benefit in the treatment of acute tubular injury. These include atrial natriuretic peptide, low-dose dopamine, endothelin antagonists, erythropoietin, prostaglandin analogues, antioxidants, antibodies against leukocyte adhesion molecules, growth factors, intra-aortic mesenchymal stem cells, and small interfering RNAs to inhibit p53-mediated cell death, among many others. Most studies have used changes in SCr to identify AKI; kidney injury biomarkers described previously may provide an opportunity to test agents with greater sensitivity.

AKI due to acute glomerulonephritis or vasculitis may respond to immunosuppressive agents, anticomplement therapies, and/or plasmapheresis (Chap. 320). Allergic interstitial nephritis due to medications requires discontinuation of the offending agent. Glucocorticoids have been used but not tested in randomized trials, in cases where AKI persists or worsens despite discontinuation of the suspected medication. AKI due to scleroderma (scleroderma renal crisis) should be treated with ACE inhibitors. Idiopathic TTP is a medical emergency and should be treated promptly with plasma exchange. Pharmacologic blockade of complement activation may be effective in atypical HUS.

Early and aggressive volume repletion is mandatory in patients with rhabdomyolysis, who may initially require 10 L of fluid per day. Alkaline fluids (e.g., 75 mmol/L sodium bicarbonate added to 0.45% saline) may be beneficial in preventing tubular injury and cast formation but carry the risk of worsening hypocalcemia. Diuretics may be used if fluid repletion is adequate but unsuccessful in achieving urinary flow rates of 200–300 mL/h. There is no specific therapy for established AKI in rhabdomyolysis, other than dialysis in severe cases or general supportive care to maintain fluid and electrolyte balance and tissue perfusion. Careful attention must be focused on calcium and phosphate status because of precipitation in damaged tissue and release when the tissue heals.

Postrenal AKI Prompt recognition and relief of urinary tract obstruction can forestall the development of permanent structural damage induced by urinary stasis. The site of obstruction defines the treatment approach. Transurethral or suprapubic bladder catheterization may be all that is needed initially for urethral strictures or functional bladder impairment. Ureteric obstruction may be treated by percutaneous nephrostomy tube placement or ureteral stent placement. Relief of obstruction is usually followed by an appropriate diuresis for several days. In rare cases, severe polyuria persists due to tubular dysfunction and may require continued administration of intravenous fluids and electrolytes for a period of time.

SUPPORTIVE MEASURES FOR AKI

Volume Management Hypervolemia in oliguric or anuric AKI may be life threatening due to acute pulmonary edema, especially because many patients have coexisting pulmonary disease, and AKI likely increases pulmonary vascular permeability. Fluid and sodium should be restricted, and diuretics may be used to increase the urinary flow rate. There is no evidence that increasing urine output itself improves the natural history of AKI, but diuretics may help to avoid the need for dialysis in some cases. In severe cases of volume overload, furosemide may be given as a bolus (200 mg) followed by an intravenous drip (10–40 mg/h), with or without a thiazide diuretic. In decompensated heart failure, stepped diuretic therapy was found to be superior to ultrafiltration in preserving renal function. Dopamine in low doses may transiently increase salt and water excretion by the kidney in prerenal states, but clinical trials have failed to show any benefit in patients with intrinsic AKI. Because of the risk of arrhythmias and potential bowel ischemia, the risks of dopamine outweigh the benefits if used specifically for the treatment or prevention of AKI.

Electrolyte and Acid-Base Abnormalities The treatment of dysnatremias and hyperkalemia is described in Chap. 56. Metabolic acidosis associated with AKI is generally not treated unless severe (pH <7.20 and serum bicarbonate <15 mmol/L). Acidosis can be treated with oral or intravenous sodium bicarbonate (Chap. 58), but overcorrection should be avoided because of the possibility of metabolic alkalosis, hypocalcemia, hypokalemia, and volume overload. Hyperphosphatemia is common in AKI and can usually be treated by limiting intestinal absorption of phosphate using phosphate binders (calcium carbonate, calcium acetate, lanthanum, sevelamer, or aluminum hydroxide). Symptomatic hypocalcemia should be treated with calcium gluconate or calcium chloride. Ionized calcium should be monitored rather than total calcium when hypoalbuminemia is present.

Malnutrition Increased catabolism with protein energy wasting is common in severe AKI, particularly in the setting of multisystem organ failure. Inadequate nutrition may lead to starvation ketoacidosis and protein catabolism. Excessive nutrition may increase the generation of nitrogenous waste and lead to worsening azotemia. Total parenteral nutrition requires large volumes of fluid administration and may complicate efforts at volume control. According to the KDIGO guidelines, patients with AKI should achieve a total energy intake of 20–30 kcal/kg per day. Protein intake should vary depending on the severity of AKI: 0.8–1.0 g/kg per day in noncatabolic AKI without the need for dialysis; 1.0–1.5 g/kg per day in patients on dialysis; and up to a maximum of 1.7 g/kg per day if hypercatabolic and receiving continuous renal replacement therapy. Trace elements and water-soluble vitamins should also be supplemented in AKI patients treated with dialysis and continuous renal replacement therapy.

Anemia The anemia seen in AKI is usually multifactorial and is not improved by erythropoiesis-stimulating agents, due to their delayed onset of action and the presence of bone marrow resistance in critically ill patients. Uremic bleeding may respond to desmopressin or estrogens but may require dialysis for treatment in the case of long-standing or severe uremia. Gastrointestinal prophylaxis with proton pump inhibitors or histamine (H_2) receptor blockers is required. It is important to recognize, however, that proton pump inhibitors have been associated with AKI due to interstitial nephritis, a relationship that is increasingly being recognized. Venous thromboembolism prophylaxis is important and should be tailored to the clinical setting; low-molecular-weight heparins and factor Xa inhibitors have unpredictable pharmacokinetics in severe AKI and should generally be avoided if possible.

Dialysis Indications and Modalities (See also Chap. 323.)

Dialysis is indicated when medical management fails to control volume overload, hyperkalemia, or acidosis; in some toxic ingestions; and when there are severe complications of uremia (asterixis, pericardial rub or effusion, encephalopathy, uremic bleeding). Late initiation of dialysis carries the risk of avoidable volume, electrolyte, and metabolic complications of AKI. On the other hand, initiating dialysis too early may unnecessarily expose individuals to intravenous lines and invasive procedures, with the attendant risks of infection, bleeding, procedural complications, and hypotension. In randomized controlled trials, earlier versus later initiation of dialysis has not been demonstrated to improve survival and may increase the risk of adverse events. The initiation of dialysis should not, however, await the development of a life-threatening complication of renal failure. Many nephrologists initiate dialysis for AKI empirically when the BUN exceeds a certain value (e.g., 100 mg/dL) in patients without clinical signs of recovery of kidney function. The available modes for renal replacement therapy in AKI require either access to the peritoneal cavity (for peritoneal dialysis) or the large blood vessels (for hemodialysis, hemofiltration, and other hybrid procedures). Small solutes are removed across a semipermeable membrane down their concentration gradient ("diffusive" clearance) and/or along with the movement of plasma water ("convective" clearance). Hemodialysis can be used intermittently or continuously and can be done through convective clearance, diffusive clearance, or a combination of the two. Vascular access is through the femoral, internal jugular, or subclavian veins. Hemodialysis is an intermittent procedure that removes solutes through diffusive and convective clearance. Hemodialysis is typically performed 3–4 h per day, three to four times per week, and is the most common form of renal replacement therapy for AKI. One of the major complications of hemodialysis is hypotension, particularly in the critically ill, which can perpetuate AKI by causing ischemic injury to the recovering organ.

Continuous intravascular procedures were developed in the early 1980s to treat hemodynamically unstable patients without inducing the rapid shifts of volume, osmolarity, and electrolytes characteristic of intermittent hemodialysis. Continuous renal replacement therapy (CRRT) can be performed by convective clearance (continuous venovenous hemofiltration [CVVH]), in which large volumes of plasma water (and accompanying solutes) are forced across the semipermeable membrane by means of hydrostatic pressure; the plasma water is then replaced by a physiologic crystalloid solution. CRRT can also be performed by diffusive clearance (continuous venovenous hemodialysis [CVVHD]), a technology similar to hemodialysis except at lower blood flow and dialysate flow rates. A hybrid therapy combines both diffusive and convective clearance (continuous venovenous hemodiafiltration [CVVHDF]). To achieve some of the advantages of CRRT without the need for 24-h staffing of the procedure, some physicians favor slow low-efficiency dialysis (SLED) or extended daily dialysis (EDD). In this therapy, blood flow and dialysate flow are higher than in CVVHD, but the treatment time is reduced to ≤12 h. The choice of modality is often dictated by the immediate availability of technology and the expertise of medical staff.

The optimal dose of dialysis for AKI for any particular patient is not clear. Daily intermittent hemodialysis and high-dose CRRT do not confer a demonstrable survival or renal recovery advantage, but care should be taken to avoid undertreatment. Studies have failed to show that continuous therapies are superior to intermittent therapies when measuring survival rates. If available, CRRT is often preferred in patients with severe hemodynamic instability, cerebral edema, or significant volume overload.

Peritoneal dialysis can be performed through a temporary intraperitoneal catheter. It is rarely used in the United States for AKI in adults (although it was "rediscovered" during the COVID-19 pandemic owing to inadequate numbers of continuous and intermittent hemodialysis machines). Peritoneal dialysis has enjoyed widespread use internationally, particularly when hemodialysis technology is not as readily available. Dialysate solution is instilled into and removed from the peritoneal cavity at regular intervals in order to achieve diffusive and convective clearance of solutes across the peritoneal membrane; ultrafiltration of water is achieved by the presence of an osmotic gradient across the peritoneal membrane achieved by high concentrations of dextrose in the dialysate solution. Because of its continuous nature, it is often better tolerated than intermittent procedures like hemodialysis in hypotensive patients. Peritoneal dialysis may not be sufficient for hypercatabolic patients due to inherent limitations in dialysis efficacy.

OUTCOME AND PROGNOSIS

The development of AKI is associated with a significantly increased risk of in-hospital and long-term mortality, longer length of stay, and increased costs. AKI is also associated with an increased risk of later cardiovascular disease events, though the mechanisms are not well understood. Prerenal azotemia, with the exception of the cardiorenal and hepatorenal syndromes, and postrenal azotemia carry a better prognosis than most cases of intrinsic AKI. The kidneys may recover even after severe, dialysis-requiring AKI. Survivors of an episode of AKI requiring temporary dialysis, however, are at extremely high risk for progressive CKD, and up to 10% may develop ESKD requiring dialysis or transplantation. AKI and CKD are increasingly seen as interrelated syndromes: CKD is a major risk factor for the development of AKI, and AKI is a risk factor for the future development of CKD. Measurement of albuminuria after an AKI episode can help predict the risk of kidney disease progression and can serve as a valuable risk-stratification tool. Postdischarge care after AKI under the supervision of a nephrologist for aggressive secondary prevention of kidney disease is prudent.

■ FURTHER READING

Chawla LS et al: Acute kidney injury and chronic kidney disease as interconnected syndromes. N Engl J Med 371:58, 2014.

Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group: KDIGO Clinical Practice Guidelines for Acute Kidney Injury. Kidney Int Supp 2:1, 2012.

Lake BB et al: An atlas of healthy and injured cell states and niches in the human kidney. Nature 610:585, 2023.

Molitoris BA: Low-flow acute kidney injury: The pathophysiology of prerenal azotemia, abdominal compartment syndrome, and obstructive uropathy. Clin J Am Soc Nephrol 17:1039, 2022.

Ronco C et al: Acute kidney injury. Lancet 394: 1949, 2019.

STARRT-AKI Investigators for the Canadian Critical Care Trials Group: Timing of initiation of renal-replacement therapy in acute kidney injury. N Engl J Med 383:240, 2020.

Tomasev N et al: A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 572:116, 2019.

Yu SM, Bonventre JV: Acute kidney injury and maladaptive tubular repair leading to renal fibrosis. Curr Opin Nephrol Hypertens 29:310, 2020.

Wilson FP et al: A randomized clinical trial assessing the effect of automated medication-targeted alerts on acute kidney injury outcomes. Nat Commun 14:2826, 2023.

322

Chronic Kidney Disease

Joanne M. Bargman, Karl L. Skorecki

Chronic kidney disease (CKD) encompasses a spectrum of pathophysiologic processes associated with abnormal kidney function, often with a progressive decline in glomerular filtration rate (GFR). The risk of worsening CKD is closely linked to the GFR, its trajectory over time, and the quantity of urinary albumin excretion (albuminuria). **Figure 322-1** provides a staging of CKD stratified by the estimated risk for further progressive decline of GFR based on these parameters.

The dispiriting term *end-stage renal disease* represents a stage of CKD where the accumulation of toxins, fluid, and electrolytes normally excreted by the kidneys leads to death unless the toxins are removed by renal replacement therapy by means of dialysis or kidney transplantation. These interventions are discussed in **Chaps. 323 and 325**. End-stage renal disease will be supplanted in this chapter by the term *stage 5 CKD*.

■ PATHOPHYSIOLOGY OF CKD

The pathophysiology of CKD involves two broad mechanisms of damage: (1) specific initiating mechanisms particular to the underlying etiology (e.g., genetic abnormalities in development, immune complex deposition, inflammation, metabolic, microvascular perturbation, or toxin exposure affecting vascular, glomerular, or tubulointerstitial compartments of the kidney), and (2) nonspecific mechanisms involving hyperfiltration and hypertrophy of the remaining viable nephrons, which are common consequences of long-term reduction of renal mass, irrespective of underlying etiology. The responses to reduction in nephron number are mediated by vasoactive hormones, cytokines, and growth factors. Eventually, the short-term adaptations of hyperfiltration and hypertrophy to maintain GFR become maladaptive as the increased pressure and flow within the nephron predisposes to distortion of glomerular architecture, abnormal podocyte function, and disruption of the filtration barrier, leading to sclerosis and dropout of the remaining nephrons. Increased intrarenal activity of the renin-angiotensin system (RAS) together with reduced tubuloglomerular feedback appears to contribute both to the initial compensatory hyperfiltration and to the subsequent maladaptive hypertrophy and sclerosis **(Fig. 322-2)**. This process explains why a reduction in functioning nephron number from initial injuries may lead to a progressive decline in kidney function over many years.

■ IDENTIFICATION OF RISK FACTORS AND STAGING OF CKD

There has been significant progress in the identification of risk factors that increase the risk for CKD, even in individuals with normal GFR and often years prior to the development of overt kidney impairment **(Table 322-1)**.

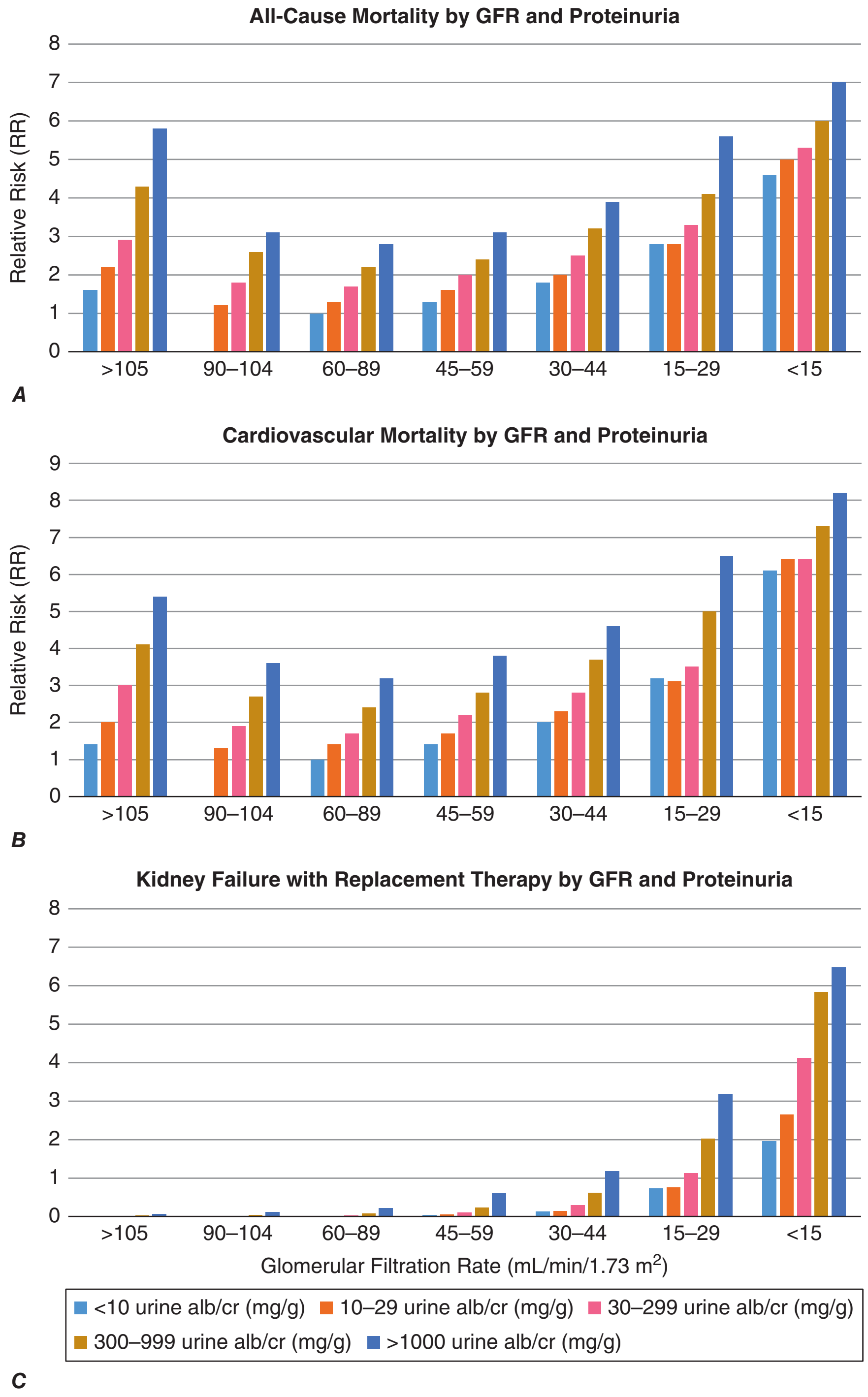

FIGURE 322-1 Kidney Disease Improving Global Outcome (KDIGO) classification of chronic kidney disease (CKD). Glomerular filtration rate (GFR) and Increasing albuminuria correspond to increasing risk of All-Cause Mortality (***A***), Cardiovascular Mortality (***B***) and Progression of CKD (***C***). *(Figure created using data from ME Grams et al: JAMA 330:1266, 2023.)*

Adults with such risk factors should be monitored yearly for urinary albumin excretion level, decline in estimated GFR (eGFR), and blood pressure, so that a clinical reno-protective management pathway can be planned. More recently, identified risk factors for which there is now a consensus include tobacco use, increased body mass index (BMI) and sedentary lifestyle, a past episode of clinically recovered acute kidney injury (AKI), and many forms of apparently recovered childhood and adolescent kidney disease.

There is also an increasing awareness of the role of genetic risk factors, which may account for up to 20% of adult-onset CKD, depending on the clinical phenotype, family background, demographic history, and population ancestry. Many rare inherited forms of CKD follow a Mendelian inheritance pattern, sometimes as part of a systemic syndrome, with the most common in this category being autosomal dominant polycystic kidney disease (ADPKD). In addition, it is now appreciated that many unique, kindred-specific, site-specific copy number variants and microdeletions, as well as functional single nucleotide variants at >300 genetic loci known to harbor systemic and kidney-only disease pathogenic mutations with high penetrance, also contribute to pleiotropic presentations of CKD **(Table 322-2)**. Many of the genes with identified CKD-causing mutations are expressed in the podocytes of the glomeruli or in the glomerular basement membrane, but others are expressed in tubule segments associated with a primary tubulointerstitial process and secondary glomerular injury.

In addition to these high penetrance mutations, DNA sequence variants with partial penetrance for causation or progression of CKD have been identified and often require an acquired second hit for emergence of disease. A striking example is the finding of allelic versions of the *APOL1* gene, of sub-Saharan African population ancestry, which contributes to the several-fold higher frequency of certain common

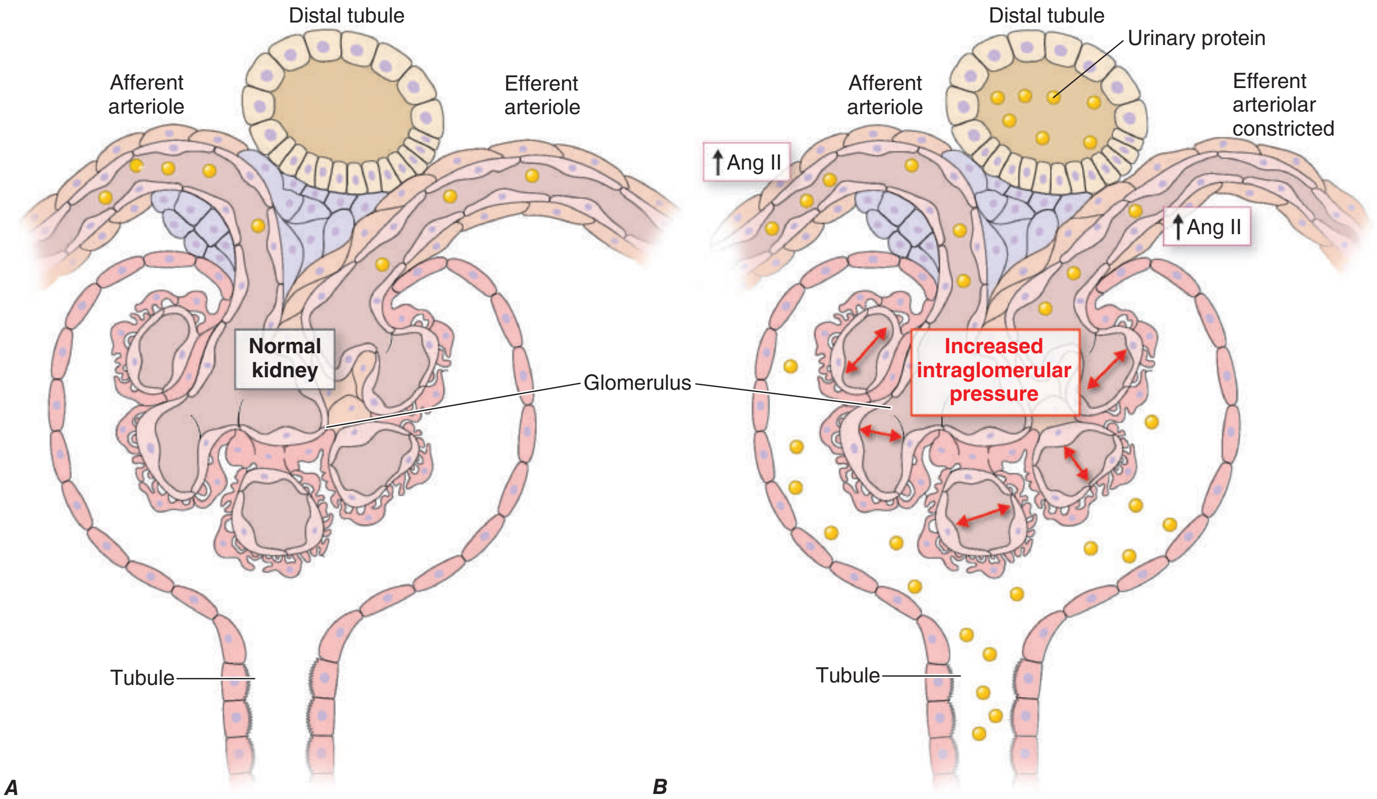

FIGURE 322-2 Schematic representation of the effect of intraglomerular hypertension on nephron survival.

etiologies of nondiabetic CKD (e.g., focal segmental glomerulosclerosis, HIV- and SARS-CoV-2–associated nephropathy, CKD with hypertension, lupus nephritis) observed among African and Hispanic Americans in major regions of continental Africa and the global African diaspora. The prevalence in West African populations seems to have arisen as an evolutionary adaptation conferring protection from tropical pathogens. As in other common diseases with a heritable component, acquired triggers (e.g., increased levels of interferon-γ and other cytokines) can transform genetic risk into disease. In addition to these single-gene loci, recent studies have identified genome-wide patterns of DNA sequence variants that confer increased risk for CKD. These include risk alleles associated with idiopathic IgA and idiopathic membranous glomerulopathy.

Staging of CKD (Fig. 322-1) is based on both GFR and on urinary albumin excretion rate and is key to understanding the emergence of symptoms, for determining the risk and rate of CKD progression and complications, and for determining indications for medical intervention. GFR is estimated (eGFR) rather than directly measured, using equations that include serum creatinine concentration and various other individual patient parameters, in place of timed urine collections, which have proven cumbersome and unreliable. U.S.-based national professional societies have recommended utilizing the equations shown in **Table 322-3**. eGFR determination in CKD is valid only if the patient is in steady state, that is, the serum creatinine is neither rising nor falling over days. The newly recommended equations no longer include a parameter that adjusts for differences in creatinine production based on continental ancestry (race-free eGFR estimation equations). However, in some situations where precise estimation of GFR might affect medication dosing or other individual clinical decisions, it is reasonable to include an estimated adjustment based on various parameters that could modify creatinine production downward (e.g., loss of a limb) or upward (increased muscle mass) or, in some cases, revert to measure timed creatinine production rates or use of the more costly and less widely available marker cystatin-C, which is not affected by variables that influence creatinine production.

The normal annual mean decline in GFR with age from the peak GFR (~120 mL/min per 1.73 m^2) attained during the third decade of life is ~1 mL/min per year per 1.73 m^2, reaching a mean value of 70 mL/min per 1.73 m^2 at age 70, with considerable interindividual variability. Although reduced GFR is expected with aging, the lower GFR signifies a true loss of kidney function with attendant consequences in terms of risk of CKD complications and requirement for dose adjustment of medications. The mean GFR is lower in women than in men, though men are at greater risk for CKD than women. A woman in her eighties with a laboratory report of serum creatinine in the normal range may have a GFR of <50 mL/min per 1.73 m^2. Relatedly, even a mild elevation in serum creatinine concentration often signifies a substantial reduction in GFR in older individuals. It is not entirely clear as yet as to whether advancing age as a risk factor for progressive CKD is independent of such age-related reduction in the measure of eGFR.

Measurement of albuminuria is also helpful for monitoring nephron injury and the response to therapy in many forms of CKD, especially chronic glomerular diseases. The cumbersome 24-h urine collection has been replaced by measurement of urinary albumin-to-creatinine ratio (UACR) in one and preferably several spot first-morning urine samples as a measure pointing to glomerular injury. Even in patients with negative conventional urinary dipstick tests for protein, persistent UACR >2.5 mg/mmol (male) or >3.5 mg/mmol (female) on two to three occasions serves as a marker not only for early detection of primary kidney disease but for systemic microvascular disease as well.

A Kidney Failure Risk (KFR) equation has been devised to predict the risk of progression to stage 5 dialysis-dependent kidney disease. The equation is available on many sites online (for example, *www.kidneyfailurerisk.com*) and uses age, sex, region (North American or non–North American), GFR, and UACR. It has been validated in several cohorts around the world, although the risk for progression appears to be greater in North America, accounting for the regional adjustment in the equation.

Stages 1 and 2 CKD are usually asymptomatic, such that the recognition of CKD occurs more often as a result of laboratory testing in clinical settings other than suspicion of kidney disease. Moreover, in the absence of the risk factors noted above, population-wide screening is not recommended. With progression to CKD stages 3 and 4, clinical and laboratory complications become more prominent. Virtually all organ systems are affected, but the most evident complications include

TABLE 322-1 Risk Factors for Chronic Kidney Disease (CKD)[a]
Noncommunicable Diseases
Diabetes
Increased BMI
Autoinflammatory disease (e.g., lupus, vasculitis, cancer immunotherapy)
Nephrotoxic exposure (including many antineoplastic therapies)
Hypertension (risk, cause, or consequence)
Communicable Diseases
Streptococcal infection
Mycobacterial infection
HIV infection (HIVAN)
SARS-CoV-2
HBV, HCV
Demographic, Anthropomorphic, Ancestry, Geographic
Age
Male sex
Population ancestry
Region-specific CKD risk of uncertain etiology (e.g., Central America, Sri Lanka, and indigenous peoples of Australia and New Zealand)
Family history of kidney disease
Genetic
Monogenic inheritance with (1) high penetrance or (2) low to medium penetrance
Polygenic risk factors
Childhood-Related Risk Factors
Premature and SGA birth
Persistent asymptomatic microscopic hematuria
Childhood kidney disease (even resolved)
Treated childhood cancer
Lifestyle
Tobacco use
Sedentary lifestyle
Other
Prior acute kidney injury
Preeclampsia
Kidney donation (or other acquired nephrectomy)

[a] Not biomarkers.

Abbreviations: BMI, body mass index; HBV, hepatitis B virus; HCV, hepatitis C virus; HIVAN, HIV-associated nephropathy; SGA, small for gestational age.

anemia with easy fatigability; decreased appetite with progressive malnutrition; abnormalities in calcium, phosphorus, and mineral-regulating hormones, such as $1,25(OH)_2D_3$ (calcitriol), parathyroid hormone (PTH), and fibroblast growth factor 23 (FGF-23); and abnormalities in sodium, potassium, water, and acid-base homeostasis. Many patients, especially older individuals, will have eGFR values compatible with stage 2 or 3 CKD. However, the majority of these patients will show no further deterioration of kidney function. In this setting, it is advised to recheck kidney function, and if it is stable and not associated with proteinuria, the patient can usually be followed with interval repeat testing without referral to a nephrologist. If repeat testing shows declining GFR, albuminuria, or uncontrolled hypertension, referral to a nephrologist is appropriate. If the patient progresses to stage 5 CKD (GFR <15 mL/min), toxins accumulate such that patients usually experience a disturbance in their activities of daily living, well-being, nutritional status, and water and electrolyte homeostasis, eventuating in the uremic syndrome.

TABLE 322-2 Monogenic Risk Loci for Chronic Kidney Disease (CKD)
Copy Number Variants Causative of Congenital Renal Anomalies
1q21
4p16.1-p16.3
16p11.2
16p13.11
17q12
22q11.2
Five Most Predominant Causes of CKD with Mendelian Inheritance
Genes for autosomal dominant polycystic kidney disease
ADPKD1
ADPKD2
IFT140
GANAB
DNAJB2
ALG9
Genes for type IV collagen-associated nephropathy
COL4A3
COL4A4
COL4A5
Genes for autosomal dominant tubulointerstitial kidney disease
UMOD
MUC1
HNF1B
Genes for nephronophthisis
NPHP genes
Other
Genes with known common variants that confer increased risk with odds ratio exceeding 2 with non-Mendelian inheritance patterns
APOL1

ETIOLOGY AND EPIDEMIOLOGY

It has been estimated from population data that at least 6% of the adult population in the United States has CKD at stages 1 and 2. An additional 4.5% of the U.S. population is estimated to have stages 3 and 4 CKD. Table 322-4 lists the five most frequent clinical categories of CKD, cumulatively accounting for >90% of the CKD disease burden worldwide. The relative contribution of each category varies among different geographic regions. The most frequent cause of CKD in North America and Europe is diabetic nephropathy, most often secondary to type 2 diabetes mellitus. Patients with newly diagnosed CKD often have hypertension. When no overt evidence for a primary glomerular or tubulointerstitial kidney disease process is present, CKD is frequently attributed to hypertension. However, it is now appreciated that some of these patients may have a subclinical primary glomerulopathy, such as focal segmental or global glomerulosclerosis, and the elevated blood pressure is a consequence of the kidney disease. In other patients, progressive nephrosclerosis and hypertension are the renal correlates of a systemic vascular disease, often also involving large and small vessels elsewhere, such as the heart and brain. This latter combination is especially common in older patients, among whom chronic kidney ischemia as a cause of CKD may be underdiagnosed.

PATHOPHYSIOLOGY AND BIOCHEMISTRY OF UREMIA

Uremia is the syndrome with symptoms, signs, and accompanying disturbances in laboratory measurements that result from reduced kidney

TABLE 322-3 Recommended Equations for Estimation of Glomerular Filtration Rate (GFR) Using Serum Creatinine Concentration (S_{Cr}), Age, Sex, Race, and Body Weight
1. Equation from the Modification of Diet in Renal Disease Study
Estimated GFR (mL/min per 1.73 m^2) = $1.86 \times (S_{Cr})^{-1.154} \times (\text{age})^{-0.203}$
Multiply by 0.742 for women
Multiply by 1.21 for African ancestry (currently under review)
2. CKD-EPI Equation
$GFR = 141 \times \min(S_{Cr}/\text{kappa}, 1)^{\alpha} \times \max(S_{Cr}/\text{kappa}, 1)^{-1.209} \times 0.993^{\text{Age}}$
Multiply by 1.018 for women
Multiply by 1.159 for African ancestry (currently under review)
where S_{Cr} is serum creatinine in mg/dL, kappa is 0.7 for females and 0.9 for males, α is –0.329 for females and –0.411 for males, min indicates the minimum of S_{Cr}/kappa or 1, and max indicates the maximum of S_{Cr}/kappa or 1.

Abbreviation: CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration.

TABLE 322-4 Leading Categories of Etiologies of Chronic Kidney Disease (CKD)[a]

- Diabetic nephropathy
- Glomerulonephritis
- Hypertension-associated CKD (includes vascular and ischemic kidney disease and primary glomerular disease with associated hypertension)
- Autosomal dominant polycystic kidney disease
- Other cystic and tubulointerstitial nephropathy

[a]Relative contribution of each category varies with geographic region and race.

function. Although serum urea and creatinine concentrations rise with reduced excretory capacity of the kidneys, accumulation of these two molecules themselves does not account for the symptoms and signs that characterize the uremic syndrome. Large numbers of solutes that accumulate when GFR declines have been implicated. These include water-soluble, hydrophobic, protein-bound, charged, and uncharged nitrogen-containing nonvolatile products of metabolism. It is thus evident that the serum concentrations of urea and creatinine should be viewed as being readily measured but very incomplete surrogate markers for retained toxins, and monitoring the levels of urea and creatinine in the patient with impaired kidney function represents a vast oversimplification of the uremic state.

The uremic syndrome involves more than renal excretory failure. A host of metabolic and endocrine functions normally performed by the kidneys are also impaired and can result in anemia, malnutrition, and abnormal metabolism of carbohydrates, fats, and proteins. Furthermore, plasma levels of many hormones, including PTH, FGF-23, insulin, glucagon, steroid hormones including vitamin D and sex hormones, and prolactin change with CKD as a result of reduced excretion, decreased degradation, or abnormal regulation. Finally, CKD is associated with increased systemic inflammation. Elevated levels of C-reactive protein are detected along with other acute-phase reactants, whereas levels of so-called negative acute-phase reactants, such as albumin and fetuin, decline. Thus, the inflammation associated with CKD is important in the malnutrition-inflammation-atherosclerosis/calcification syndrome, which contributes in turn to the acceleration of vascular disease and morbidity associated with advanced kidney disease.

In summary, the pathophysiology of the uremic syndrome can be divided into manifestations in three spheres of dysfunction: (1) those consequent to the accumulation of toxins that normally undergo renal excretion; (2) those consequent to the loss of other kidney functions, such as fluid and electrolyte homeostasis and hormone regulation; and (3) progressive systemic inflammation and its vascular and nutritional consequences.

CLINICAL AND LABORATORY MANIFESTATIONS OF CKD AND UREMIA

■ FLUID, ELECTROLYTE, AND ACID-BASE DISORDERS

Sodium and Water Homeostasis With normal kidney function, excretion of filtered sodium and water matches intake. Many forms of kidney disease disrupt this balance such that dietary intake of sodium exceeds its excretion, leading to sodium retention and attendant extracellular fluid volume (ECFV) expansion. This expansion may contribute to hypertension, which itself can accelerate nephron hyperfiltration and injury. As long as water intake does not exceed the capacity for renal water clearance, the ECFV expansion will be isonatric and the patient will have a normal plasma sodium concentration. Hyponatremia is not commonly seen in CKD patients, but when present, often responds to water restriction. The patient with ECFV expansion should be counseled regarding salt restriction. While the thiazide diuretic chlorthalidone alone has been shown to reduce elevated blood pressure in some patients even with stage 4 CKD, administration of loop diuretics, including furosemide, bumetanide, or torsemide, may be needed to manage sodium accumulation. Resistance to loop diuretics in CKD often mandates use of higher doses than those used in patients with normal GFR. The combination of loop diuretics with metolazone may be helpful. Diuretic resistance with intractable edema and hypertension in advanced CKD may serve as an indication to initiate dialysis.

Rarely, patients with CKD may have impaired renal conservation of sodium and water. When an extrarenal cause for fluid loss, such as gastrointestinal (GI) loss, is present, these patients may be prone to ECFV depletion because of the inability of the failing kidney to reclaim filtered sodium adequately. Any depletion of ECFV, whether due to GI losses, renal sodium loss, or overzealous diuretic therapy, can further compromise kidney function through hypoperfusion, or a "prerenal" state, leading to acute-on-chronic kidney failure. In this setting, holding or adjusting the diuretic dose or rarely even cautious volume repletion with normal saline may return the ECFV to normal and restore renal function to baseline. Many patients are given a "sick day" warning, wherein should they experience volume depletion, for example from vomiting or diarrhea, they are told to not take their diuretics or other antihypertensive medications until they resume eating and drinking normally.

Potassium Homeostasis In CKD, the decline in GFR is not necessarily accompanied by a parallel decline in urinary potassium excretion, which is predominantly mediated by aldosterone-dependent secretion in the distal nephron. Another defense against potassium retention in these patients is augmented potassium excretion in the GI tract. Notwithstanding these two homeostatic responses, hyperkalemia may be precipitated in certain settings. These include increased dietary potassium intake, hemolysis, transfusion of stored red blood cells, and metabolic acidosis. Importantly, a host of medications can inhibit renal potassium excretion and lead to hyperkalemia. The most important medications in this respect include the RAS inhibitors and spironolactone and other potassium-sparing diuretics such as amiloride, eplerenone, and triamterene, as well as the new nonsteroidal mineralocorticoid receptor antagonists. As will be outlined below, several of these are major agents in the medical management armamentarium to slow or prevent progression of CKD. The benefits of the RAS inhibitors in ameliorating hyperfiltration and progression of CKD and mitigating cardiovascular complications very often favor their cautious and judicious use with very close monitoring of plasma potassium concentration. Coadministration of potassium-lowering agents may allow for the use of RAS inhibitors with reduced risk of hyperkalemia. Gratifyingly, the gliflozin diuretics, administrated even in advanced stages of CKD, seem to have counterbalancing effects on kidney potassium handling that result in net preservation of potassium homeostasis.

Certain causes of CKD can be associated with earlier and more severe disruption of potassium secretory mechanisms in the distal nephron, out of proportion to the decline in GFR. These include conditions associated with hyporeninemic hypoaldosteronism, such as diabetes, and renal diseases that preferentially affect the distal nephron, such as obstructive uropathy and sickle cell nephropathy.

Hypokalemia is not common in CKD and usually reflects markedly reduced dietary potassium intake, especially in association with excessive diuretic therapy or concurrent GI losses. The use of potassium supplements and potassium-sparing diuretics may be risky in patients with impaired renal function and needs to be monitored closely.

Metabolic Acidosis Metabolic acidosis is a common disturbance in CKD. The majority of patients can still acidify the urine, but they produce less ammonia and, therefore, cannot excrete the quantity of protons required to maintain acid-base balance in most diets. Hyperkalemia, if present, further depresses ammonia production. The combination of hyperkalemia and hyperchloremic metabolic acidosis is often present, even at earlier stages of CKD, in patients with diabetic nephropathy or in those with predominant tubulointerstitial disease including obstructive uropathy. With further declining GFR, the total urinary net daily acid excretion may be severely limited to <30–40 mmol, and the accumulation of anions of retained organic acids can then lead to an anion-gap metabolic acidosis. Thus, the non-anion-gap metabolic acidosis seen in earlier stages of CKD may be complicated by the addition of an anion-gap metabolic acidosis as CKD progresses. In most patients, the metabolic acidosis is mild; the pH is rarely <7.32

and can usually be corrected with oral sodium bicarbonate supplementation. Studies have suggested that even modest degrees of metabolic acidosis may be associated with the development of protein catabolism and progression of CKD.

TREATMENT

Fluid, Electrolyte, and Acid-Base Disorders

Dietary salt restriction and the use of loop diuretics, occasionally in combination with metolazone, may be needed to maintain euvolemia. Water restriction is indicated only if there is hyponatremia.

Hyperkalemia often responds to dietary restriction of potassium, the use of kaliuretic diuretics, and both avoidance of potassium supplements (including occult sources, such as dietary salt substitutes) and monitoring with dose adjustment, or at times avoidance of potassium-retaining medications, which are often prescribed to slow CKD progression or afford cardioprotection (RAS inhibitors, steroidal or nonsteroidal mineralocorticoid antagonists). Kaliuretic diuretics promote urinary potassium excretion, whereas potassium-binding resins, such as calcium resonium, sodium polystyrene, or the newer agents patiromer and calcium zirconium cyclosilicate, promote potassium loss through the GI tract and may reduce the incidence of hyperkalemia. Intractable hyperkalemia is an indication (although uncommon) to consider institution of dialysis in a CKD patient. The renal tubular acidosis and subsequent anion-gap metabolic acidosis in progressive CKD will respond to alkali supplementation. Recent studies suggest that this replacement should be considered when the serum bicarbonate concentration falls below 20–23 mmol/L to avoid the protein catabolic state seen with even mild degrees of metabolic acidosis. Relatedly, a recent study suggested that bicarbonate supplementation in stages 3–5 CKD was also associated with slower progression to dialysis. The sodium load in sodium bicarbonate supplementation needs to be taken into account when ECFV expansion is present.

DISORDERS OF CALCIUM AND PHOSPHATE METABOLISM

The principal complications of abnormalities of calcium and phosphate metabolism in CKD occur in the skeleton and the vascular bed, with occasional involvement of soft tissues. It is likely that disorders of bone turnover and disorders of vascular and soft tissue calcification are related to each other.

Bone Manifestations of CKD The major disorders of bone disease can be classified into those associated with high bone turnover with increased PTH levels (including osteitis fibrosa cystica, the classic lesion of secondary hyperparathyroidism), osteomalacia due to reduced vitamin D effect, and low bone turnover with low or normal PTH levels (adynamic bone disease) or often some combination of the foregoing.

The pathophysiology of secondary hyperparathyroidism and the consequent high-turnover bone disease is related to abnormal mineral metabolism through the following series of interrelated mechanisms: (1) declining GFR leads to reduced excretion of phosphate and, thus, phosphate retention; (2) the retained phosphate stimulates increased synthesis of both FGF-23 by osteocytes and of PTH and also stimulates growth of parathyroid gland mass; and (3) decreased levels of ionized calcium, which results from decreased levels of renal calcitriol production due to phosphate retention and elevated levels of FGF-23, which also increases degradation of calcitriol. Low calcitriol levels contribute to hyperparathyroidism, both by leading to hypocalcemia and also by a direct effect on PTH gene transcription. In addition, the normal inhibitory effect of FGF-23 on PTH production, which is Klotho-dependent, is also attenuated in CKD. These changes start to occur when the GFR falls below 60 mL/min, though some studies point to retention of phosphate as an event antedating measurable reduction in GFR, together with early elevation of FGF-23 as well. FGF-23 is part of a family of phosphatonins that promotes phosphate excretion, and high levels of FGF-23 are an independent risk factor for left ventricular hypertrophy and are associated with increased mortality in CKD, dialysis, and kidney transplant patients.

Hyperparathyroidism stimulates bone turnover and leads to osteitis fibrosa cystica. Bone histology shows abnormal osteoid, bone and bone marrow fibrosis, and, in advanced stages, the formation of bone cysts, sometimes with hemorrhagic elements so that they appear brown in color; hence, the term *brown tumor*. Clinical manifestations of severe hyperparathyroidism include bone pain and fragility, brown tumors with compression syndromes, and resistance to erythropoiesis-stimulating agents (ESA) in part related to the bone marrow fibrosis. Furthermore, PTH itself is considered a uremic toxin, and high levels are associated with muscle weakness, fibrosis of cardiac muscle, and constitutional symptoms.

Adynamic bone disease is increasing in prevalence, especially among diabetics and older patients. It is characterized by reduced bone volume and mineralization and may result from excessive suppression of PTH production, chronic inflammation, or both. Suppression of PTH can result from the use of vitamin D preparations or from excessive calcium exposure in the form of calcium-containing phosphate binders or high-calcium dialysis solutions.

Complications of adynamic bone disease include an increased incidence of fracture and bone pain and an association with increased vascular and cardiac calcification. Occasionally, the calcium will precipitate in the soft tissues into large concretions termed *tumoral calcinosis* **(Fig. 322-3)**. Patients with adynamic bone disease often experience the most severe symptoms of musculoskeletal pain, owing to the inability to repair the microfractures that occur normally as a part of healthy skeletal homeostasis with regular physical activity. Patients with advanced CKD experience more frequent fractures than their age-matched controls. Osteomalacia is a distinct process, consequent to reduced production and action of $1{,}25(OH)_2D_3$, leading to accumulation of nonmineralized osteoid.

Calcium, Phosphorus, and the Cardiovascular System There is a strong association between hyperphosphatemia and increased cardiovascular mortality in patients with CKD. Hyperphosphatemia and hypercalcemia are associated with increased vascular calcification, but it is unclear whether the excessive mortality is mediated by this mechanism. Studies using computed tomography (CT) and electron-beam CT scanning show that CKD patients have calcification in the media of coronary arteries and heart valves that appears to be orders of magnitude greater than that in patients without kidney disease. The magnitude of the calcification is proportional to age and hyperphosphatemia and is also associated with low PTH levels and low bone turnover. It is possible, that in CKD patients, ingested calcium cannot be incorporated into bones with low turnover, and therefore, is deposited at

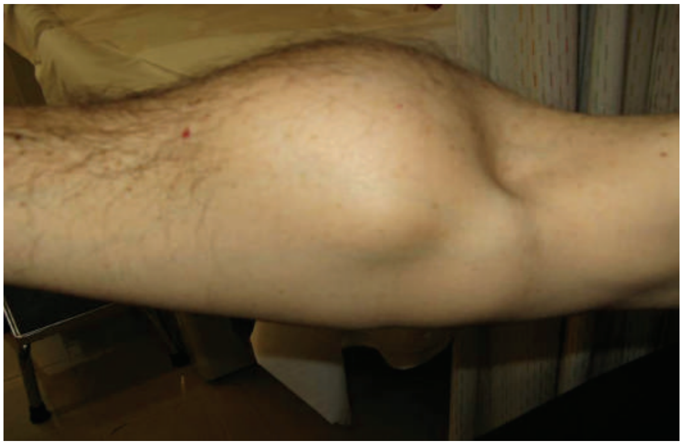

FIGURE 322-3 Tumoral calcinosis. This patient was on hemodialysis for many years and was nonadherent to dietary phosphorus restriction or the use of phosphate binders. He was chronically severely hyperphosphatemic. He developed an enlarging painful mass on his arm that was extensively calcified.

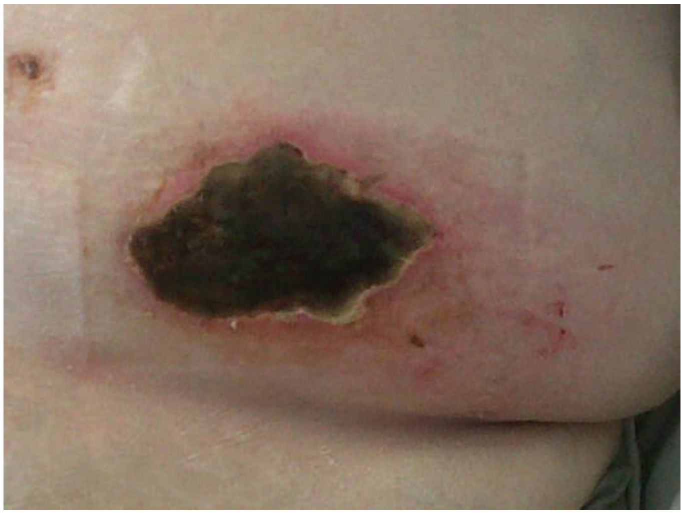

FIGURE 322-4 **Calciphylaxis.** This peritoneal dialysis patient was on chronic warfarin therapy for atrial fibrillation. She noticed a small painful nodule on the abdomen that was followed by progressive skin necrosis and ulceration of the anterior abdominal wall. She was treated with hyperbaric oxygen, intravenous thiosulfate, and discontinuation of warfarin, with slow resolution of the ulceration.

extraosseous sites, such as the vascular bed and soft tissues. There is a similar association between osteoporosis and vascular calcification in the general population. Finally, hyperphosphatemia can induce a change in gene expression in vascular cells to an osteoblast-like profile, leading to vascular calcification and even ossification.

Other Complications of Abnormal Mineral Metabolism Calciphylaxis is a devastating condition seen almost exclusively in patients with advanced CKD. It is heralded by painful livedo reticularis and subcutaneous nodules that advance to patches of ischemic necrosis, especially on the legs, thighs, abdomen, and breasts (Fig. 322-4). Pathologically, there is evidence of vascular occlusion in association with extensive vascular and soft tissue calcification. It appears that this condition is increasing in incidence. Originally it was ascribed to severe abnormalities in calcium and phosphorus control in dialysis patients, usually associated with advanced hyperparathyroidism. However, more recently, calciphylaxis has been seen with increasing frequency in the absence of severe hyperparathyroidism. Warfarin is still used in some CKD patients in whom several members of the direct oral anticoagulant (DOAC) family of drugs are contraindicated, and one of the effects of warfarin therapy is to decrease the vitamin K–dependent activation of matrix GLA protein. This latter protein is important in preventing vascular calcification. Thus, warfarin treatment is considered a risk factor for calciphylaxis, and if a patient develops this syndrome, this medication should be discontinued and alternative means of anticoagulation should be chosen, depending on the specific underlying indication for anticoagulation.

TREATMENT

Disorders of Calcium and Phosphate Metabolism

The optimal management of secondary hyperparathyroidism and osteitis fibrosa is prevention. Once the parathyroid gland mass is very large, it is difficult to control the disease. Careful attention should be paid to the plasma phosphate concentration in CKD patients, who should be counseled on a low-phosphate diet as well as the appropriate use of phosphate-binding agents, which are taken with meals and complex dietary phosphate to limit its GI absorption. Examples of phosphate binders are calcium acetate and calcium carbonate. A major side effect of calcium-based phosphate binders is calcium accumulation and hypercalcemia, especially in patients with low-turnover bone disease. Sevelamer and lanthanum are non-calcium-containing polymers that also function as phosphate binders; they do not predispose CKD patients to hypercalcemia and may attenuate calcium deposition in the vascular bed. Tenapanor is a sodium-proton inhibitor that decreases GI phosphate absorption and may be useful to manage hyperphosphatemia in CKD and dialysis patients.

Calcitriol exerts a direct suppressive effect on PTH secretion and also indirectly suppresses PTH secretion by raising the concentration of ionized calcium. However, calcitriol therapy may result in hypercalcemia and/or hyperphosphatemia through increased GI absorption of these minerals. Certain analogues of calcitriol are available (e.g., paricalcitol) that suppress PTH secretion with less attendant hypercalcemia.

Recognition of the role of the extracellular calcium-sensing receptor has led to the development of calcimimetic agents that enhance the sensitivity of parathyroid cells to the suppressive effect of calcium. This class of drug, which includes cinacalcet and etelcalcetide, produces a dose-dependent reduction in PTH and plasma calcium concentration in some patients.

Current National Kidney Foundation Kidney Disease Outcomes Quality Initiative guidelines recommend a target PTH level between 2 and 9 times the upper limit of normal, recognizing that very low PTH levels are associated with adynamic bone disease and possible consequences of fracture and ectopic calcification.

For CKD patients requiring anticoagulation, careful assessment of the indication and choice and dosing of medication appropriate for reduced renal clearance, with avoidance of warfarin, should be considered to reduce the risk of calciphylaxis.

■ CARDIOVASCULAR ABNORMALITIES

Cardiovascular disease is the leading cause of morbidity and mortality in patients at every stage of CKD. The incremental risk of cardiovascular disease in those with CKD compared to the age- and sex-matched general population ranges from 10- to 200-fold, depending on the stage of CKD. As a result, most patients with CKD succumb to cardiovascular disease before ever reaching stage 5 CKD. Between 30 and 45% of those patients who do reach stage 5 CKD have advanced significant cardiovascular complications.

Vascular Disease The increased prevalence of vascular disease in CKD patients derives from both traditional (“classic”) and nontraditional (CKD-related) risk factors. Traditional risk factors include hypertension, diabetes mellitus, hypervolemia, dyslipidemia, sympathetic overactivity, and hyperhomocysteinemia. The CKD-related risk factors comprise anemia, hyperphosphatemia, hyperparathyroidism, increased FGF-23, sleep apnea, and systemic inflammation. The inflammatory state appears to accelerate vascular occlusive disease, and low levels of fetuin may permit more rapid vascular calcification, especially in the face of hyperphosphatemia. Other abnormalities seen in CKD may augment myocardial ischemia, including left ventricular hypertrophy and microvascular disease. It is noteworthy that both high and low ejection fraction congestive heart failure, left ventricular hypertrophy, systemic hypertension, and pulmonary hypertension are no less prominent than coronary ischemia as causes of cardiovascular mortality in patients with advanced stages of CKD.

Cardiac troponin levels are frequently elevated in CKD without evidence of acute ischemia. The elevation complicates the diagnosis of acute myocardial infarction in this population. Serial measurements may be needed. Therefore, the trend in levels over the hours after presentation may be more informative than a single, elevated level. Interestingly, consistently elevated levels are an independent prognostic factor for adverse cardiovascular events.

Heart Failure Abnormal cardiac function secondary to myocardial ischemia, left ventricular hypertrophy, diastolic dysfunction, and frank cardiomyopathy, in combination with salt and water retention, often results in heart failure or even pulmonary edema. Heart failure can occur with preserved (diastolic dysfunction) or reduced (systolic dysfunction) ejection fraction. A form of “low-pressure” pulmonary edema can also occur in advanced CKD, manifesting as shortness of breath and a “bat wing” distribution of alveolar edema fluid on chest x-ray. This finding can occur even in the absence of ECFV overload and is associated with normal or mildly elevated pulmonary venous pressure. This process has been ascribed to increased permeability of alveolar capillary membranes as a manifestation of the uremic state, and it responds to dialysis. Other CKD-related risk factors, including anemia and sleep apnea, may contribute to the risk of heart failure.

Hypertension and left ventricular hypertrophy are common complications of CKD. Hypertension usually develops early during the course of CKD and is associated with adverse outcomes, including the development of ventricular hypertrophy and a more rapid loss of renal function. Left ventricular hypertrophy and dilated cardiomyopathy are among the strongest risk factors for cardiovascular morbidity and mortality in patients with CKD and are thought to be related primarily, but not exclusively, to prolonged hypertension and ECFV overload. In addition, anemia and the placement of an arteriovenous fistula for hemodialysis can generate a high cardiac output state and consequent high-output heart failure.

It is important to note that in advanced stages of CKD, the presence of an advanced malnutrition-inflammation state can actually reverse the elevation of classic cardiovascular risk factors such as hypertension and hyperlipidemia and is associated in such patients with reduced left ventricular function, loss of body weight, and a poor prognosis.

The use of exogenous ESAs can increase blood pressure and the requirement for antihypertensive drugs. Chronic ECFV overload is also a contributor to hypertension, and improvement in blood pressure can often be seen with the use of dietary sodium restriction and diuretics, and when refractory, hypertension can serve as an indication of initiating renal replacement therapy. Nevertheless, because of activation of the RAS and other disturbances in the balance of vasoconstrictors and vasodilators, some patients remain hypertensive, despite careful attention to ECFV status.

TREATMENT

Cardiovascular Abnormalities

MANAGEMENT OF HYPERTENSION

The overarching goal of hypertension therapy in CKD is to prevent the extrarenal complications of high blood pressure, such as cardiovascular disease and stroke. National guideline panels recommend that in CKD patients with diabetes or proteinuria >1 g per 24 h, blood pressure should be reduced to <130/80 mmHg, if achievable without prohibitive adverse effects. Salt restriction should be the first line of therapy. When volume management alone is not sufficient, the choice of antihypertensive agent is similar to that in the general population. Angiotensin-converting enzyme (ACE) inhibitors and angiotensin receptor blockers (ARBs) appear to slow the rate of decline of kidney function in a manner that extends beyond reduction of systemic arterial pressure and that involves reduction in the intraglomerular hyperfiltration and hypertension by disproportionately reducing glomerular efferent arteriolar vasoconstriction. Occasionally, introduction of ACE inhibitors and ARBs can actually precipitate an episode of AKI, especially when used in combination in patients with ischemic renovascular disease.

More commonly, a slight reduction of GFR (<30% of baseline) may actually signify a salutary reduction in intraglomerular hypertension and hyperfiltration and, if stable over time, can be tolerated with continued monitoring. With progressive CKD, RAS inhibitors can be continued with careful monitoring. Careful clinical studies have shown that even in patients with an eGFR <30 mL/min, continuation of these agents was not associated with any signal for harm compared to those in whom the drugs were stopped and that discontinuing RAS inhibitors in CKD was associated with an increased risk of death or cardiovascular events but a lower incidence of starting kidney replacement therapy. Since the use of ACE inhibitors and ARBs may also be complicated by the development of hyperkalemia, the concomitant use of kaliuretic diuretics (e.g., furosemide with metolazone) or a potassium-lowering GI tract binder, such as patiromer, can improve potassium excretion in addition to improving blood pressure control. Likewise, if gliflozins are implemented in renoprotection, then these may also assist in potassium homeostasis. Potassium-sparing diuretics, such as amiloride and triamterene, should be avoided in most patients, and steroidal (e.g., spironolactone) or nonsteroidal (finerenone) mineralocorticoid receptor blockers should be accompanied by careful monitoring of serum potassium concentration, weighing potential cardiovascular and renoprotective benefits against risk for lethal hyperkalemia.

The recent movement to even lower blood pressure targets in the general population may not be applicable to patients with CKD, who often lack autoregulation to maintain GFR in the face of low perfusion pressure. If a patient experiences sudden decline in kidney function with intensification of antihypertensive therapy, consideration should be given to reducing therapy.

MANAGEMENT OF CARDIOVASCULAR DISEASE

There are many strategies available to treat the traditional and nontraditional risk factors in CKD patients. Although these have proved effective in the general population, there is little evidence for their benefit in patients with the most advanced stages of CKD. Certainly, hypertension and dyslipidemia promote atherosclerotic disease and are treatable complications of CKD. Renal disease complicated by nephrotic syndrome is associated with a very atherogenic lipid profile and hypercoagulability, which increases the risk of occlusive vascular disease. Because diabetes mellitus and hypertension are frequently associated with advanced CKD, it is not surprising that cardiovascular disease is the most frequent cause of death in dialysis patients. The use of the gliflozins (sodium-glucose cotransporter 2 [SGLT2] inhibitors) in patients with and without diabetes mellitus is quickly becoming a mainstay for both kidney protection, with marked amelioration of GFR reduction, and a reduction in cardiovascular events, including heart failure, even in advanced stages of CKD (including stage 4).

Pericardial Disease Chest pain with respiratory accentuation, accompanied by a friction rub, is diagnostic of pericarditis. Classic electrocardiographic abnormalities include PR-interval depression and diffuse ST-segment elevation. Pericarditis can be accompanied by pericardial effusion that is seen on echocardiography and can rarely lead to tamponade. However, the pericardial effusion can be asymptomatic, and pericarditis can be seen without significant effusion.

Pericarditis is observed in advanced uremia and, with the advent of timely initiation of dialysis, is not as common as it once was. It is now more often observed in underdialyzed, nonadherent patients than in those starting dialysis.

TREATMENT

Pericardial Disease

Uremic pericarditis is an absolute indication for the urgent initiation of dialysis or for intensification of the dialysis prescription in those already receiving dialysis. Because of the propensity to hemorrhage in pericardial fluid, hemodialysis should be performed without heparin. A pericardial drainage procedure should be considered in patients with recurrent pericardial effusion, especially with echocardiographic signs of impending tamponade. Nonuremic causes of pericarditis and effusion include viral, malignant, tuberculous, and autoimmune etiologies. It may also be seen after myocardial infarction and as a complication of treatment with the antihypertensive drug minoxidil. Consideration could be given to the use of colchicine or nonsteroidal anti-inflammatory drugs, although the latter agents could adversely affect renal function.

■ HEMATOLOGIC ABNORMALITIES

Anemia A normocytic, normochromic anemia is observed as early as stage 3 CKD and is almost universal by stage 4. The primary cause is insufficient production of erythropoietin (EPO) by the diseased kidneys, together with reduced erythrocyte lifespan and other factors. **(See Table 322-5.)**

TABLE 322-5 Causes of Anemia in Chronic Kidney Disease

Relative deficiency of erythropoietin
Diminished red blood cell survival
Bleeding diathesis
Iron deficiency due to poor dietary absorption and gastrointestinal blood loss
Hyperparathyroidism/bone marrow fibrosis
Chronic inflammation
Folate or vitamin B_{12} deficiency
Hemoglobinopathy
Comorbid conditions: hypo-/hyperthyroidism, pregnancy, HIV-associated disease, autoimmune disease, immunosuppressive drugs

Causes of Anemia in Chronic Kidney Disease The anemia of CKD is associated with a number of adverse pathophysiologic consequences, including decreased tissue oxygen delivery and utilization, increased cardiac output, ventricular dilation, and ventricular hypertrophy. Clinical manifestations include fatigue and diminished exercise tolerance, angina, heart failure, decreased cognition and mental acuity, and impaired host defense against infection. In addition, anemia may play a role in growth restriction in children with CKD. Although many studies in CKD patients have found that anemia and resistance to exogenous ESAs are associated with a poor prognosis, the relative contribution to a poor outcome of the low hematocrit itself, versus inflammation as a cause of the anemia and ESA resistance, remains unclear.

TREATMENT

Anemia

The availability of recombinant human ESA has been one of the most significant advances in the care of renal patients since the introduction of dialysis and renal transplantation. Its routine use has obviated the need for regular blood transfusions in severely anemic CKD patients, thus dramatically reducing the incidence of transfusion-associated infections, iron overload, and the development of alloantibodies that can sensitize patients to donor kidney antigens and render kidney transplantation more problematic.

Adequate bone marrow iron stores should be available before treatment with ESA is initiated. Iron supplementation is usually essential to ensure an optimal response to ESA in patients with CKD because the demand for iron by the marrow frequently exceeds the amount of iron that is immediately available for erythropoiesis (measured by percent transferrin saturation), as well as the amount in iron stores (measured by serum ferritin). For the CKD patient not yet on dialysis or the patient treated with peritoneal dialysis, oral iron supplementation should be attempted. If there is GI intolerance or poor GI absorption, the patient may have to undergo IV iron infusion, keeping in mind that parenteral iron therapy can increase the susceptibility to bacterial infections and that the adverse effects of free serum iron are still under investigation. In addition to iron, an adequate supply of other major substrates and cofactors for red cell production must be ensured, including vitamin B_{12} and folate. Anemia resistant to recommended doses of ESA in the face of adequate iron stores may be due to some combination of the following: acute or chronic inflammation, inadequate dialysis, severe hyperparathyroidism, chronic blood loss or hemolysis, chronic infection, or malignancy.

A new class of agents to treat the anemia of CKD are the prolyl-hydroxylase inhibitors of endogenous hypoxia-inducible factors (HIFs). This inhibition leads to an increase in both endogenous production of EPO and an increase in GI absorption of iron. While studies comparing HIF inhibitors to ESAs have demonstrated similar effectiveness, some safety considerations remain to be clarified in the realms of cardiovascular, retinal vessel, and possibly tumor vasculogenesis.

Randomized, controlled trials of ESAs in CKD have failed to show an improvement in cardiovascular outcomes with this therapy. Indeed, there has been an indication that the use of ESAs in CKD may be associated with an increased risk of stroke in those with type 2 diabetes or an increase in thromboembolic events and perhaps a faster progression of renal decline.

Therefore, any benefit in terms of improvement of anemic symptoms needs to be balanced against such potential cardiovascular risk. Since normalization of the hemoglobin concentration has not been demonstrated to be of benefit to CKD patients, current practice is to target a hemoglobin concentration of 100–115 g/L.

Abnormal Hemostasis Patients with later stages of CKD may have a prolonged bleeding time, decreased activity of platelet factor III, abnormal platelet aggregation and adhesiveness, and impaired prothrombin consumption. Clinical manifestations include an increased tendency to bleeding and bruising, prolonged bleeding from surgical incisions, menorrhagia, and GI bleeding. Interestingly, CKD patients also have a greater susceptibility to thromboembolism, especially if they have renal disease that includes nephrotic-range proteinuria. The latter condition results in hypoalbuminemia and renal loss of anticoagulant factors, which can lead to a thrombophilic state.

TREATMENT

Abnormal Hemostasis

Abnormal bleeding time and coagulopathy in patients with renal failure may be reversed temporarily with desmopressin (DDAVP), cryoprecipitate, IV conjugated estrogens, blood transfusions, and ESA therapy. Rarely with refractory and life-threatening bleeding, tranexamic acid and epsilon aminocaproic acid have also been used for hemostasis control. Optimal dialysis will usually correct a prolonged bleeding time.

Given the coexistence of bleeding disorders and a propensity to thrombosis that is unique in the CKD patient, decisions about anticoagulation that have a favorable risk-benefit profile in the general population may not be applicable to the patient with advanced CKD. One example is warfarin anticoagulation for atrial fibrillation; the decision to anticoagulate should be made on an individual basis in the CKD patient because there appears to be a greater risk of bleeding complications, as well as the consideration of precipitating calciphylaxis.

Certain anticoagulants, such as fractionated low-molecular-weight heparin, may need to be avoided or dose-adjusted in these patients, with monitoring of factor Xa activity where available. It is often more prudent to use conventional unfractionated heparin, titrated to the measured partial thromboplastin time, in hospitalized patients requiring an alternative to warfarin anticoagulation. Most of the new classes of oral anticoagulants are eliminated by the kidneys, although apixaban in either full or reduced dose is being used more often in CKD and dialysis patients **(Chap. 123)**.

NEUROMUSCULAR ABNORMALITIES

Central nervous system (CNS), peripheral, and autonomic neuropathy, as well as abnormalities in muscle structure and function, are all well-recognized complications of CKD. Subtle clinical manifestations of uremic neuromuscular disease usually become evident at stage 3 CKD.

Early manifestations of CNS complications include mild disturbances in memory and disturbances in concentration and sleep. Neuromuscular irritability, including hiccups, cramps, and twitching, becomes evident at later stages. In advanced untreated kidney failure, asterixis, myoclonus, seizures, and coma can be seen.

Peripheral neuropathy usually becomes clinically evident after the patient reaches stage 4 CKD, although electrophysiologic and histologic evidence occurs earlier. Initially, sensory nerves are involved more than motor, lower extremities more than upper, and distal parts of the extremities more than proximal. The "restless leg syndrome" is characterized by ill-defined sensations of sometimes debilitating discomfort in the legs and feet relieved by frequent leg movement. Evidence of peripheral neuropathy without another cause (e.g., diabetes mellitus or iron deficiency) is an indication for starting renal replacement therapy. Many of the complications described above will resolve with dialysis, although some may persist.

■ GASTROINTESTINAL AND NUTRITIONAL ABNORMALITIES

Uremic fetor, a urine-like odor on the breath, derives from the breakdown of urea to ammonia in saliva and is often associated with an unpleasant metallic taste (dysgeusia). Gastritis, peptic disease, and mucosal ulcerations at any level of the GI tract occur in uremic patients and can lead to abdominal pain, nausea, vomiting, and GI bleeding. These patients are also prone to constipation, which can be worsened by the administration of calcium and iron supplements. The retention of uremic toxins also leads to anorexia, nausea, and vomiting.

Protein restriction may be useful to decrease nausea and vomiting; however, it may put the patient at risk for malnutrition and should be carried out, if possible, in consultation with a registered dietitian specializing in the management of CKD patients. Weight loss and protein-energy malnutrition, consequences of low protein and caloric intake, are common in advanced CKD and are often an indication for initiation of renal replacement therapy. Metabolic acidosis and the activation of inflammatory cytokines can promote protein catabolism. A number of indices are useful in nutritional assessment and include dietary history, including food diary, and subjective global assessment; edema-free body weight; and measurement of urinary protein nitrogen appearance. Dual-energy x-ray absorptiometry bioimpedance analysis is now widely used to estimate lean body mass versus fluid weight. Nutritional guidelines for patients with CKD are summarized in the "Treatment" section.

■ ENDOCRINE-METABOLIC DISTURBANCES

Glucose metabolism is impaired in CKD. However, fasting blood glucose is usually normal or only slightly elevated, and mild glucose intolerance does not require specific therapy. Because the kidney contributes to insulin removal from the circulation, plasma levels of insulin are slightly to moderately elevated in most uremic patients, both in the fasting and postprandial states. Because of this diminished renal degradation of insulin, patients on insulin therapy may need progressive reduction in dose as their renal function worsens. Many antihyperglycemic agents, including the gliptins, require dose reduction in renal failure, and some, such as metformin and sulfonylureas, are contraindicated when the GFR is less than half of normal. The gliflozins, discussed above, that inhibit sodium-glucose transport in the proximal tubule result in glucose lowering, accompanied by striking reductions in kidney function decline and in cardiovascular events. The stabilization of GFR in many patients with this therapeutic intervention represents a major, important added beneficial effect of these drugs. Their long-term stabilizing effect on GFR and urine albumin excretion appears to result from correction of hyperfiltration early in type 2 diabetes mellitus via reactivation of the tubuloglomerular feedback loop. This represents a fortunate convergence of pathophysiology of glomerular hyperfiltration in diabetes with drug discovery. A similar effect on hyperfiltration by residual nephrons in certain nondiabetic forms of CKD may explain the salutary role of this class of medications more broadly in CKD. Other studies have also pointed to a more direct effect on proximal tubule metabolic pathways that alleviate cell injury.

In women with CKD, estrogen levels are low, and menstrual abnormalities, infertility, and inability to carry pregnancies to term are common. When the GFR has declined to ~40 mL/min, pregnancy is associated with a high rate of spontaneous abortion, with only ~20% of pregnancies leading to live births, and pregnancy may hasten the progression of the kidney disease itself. Women with CKD who are contemplating pregnancy should consult first with a nephrologist in conjunction with an obstetrician specializing in high-risk pregnancy. Men with CKD have reduced plasma testosterone levels, and sexual dysfunction and oligospermia may supervene. Sexual maturation may be delayed or impaired in adolescent children with CKD, even among those treated with dialysis. Many of these abnormalities improve or reverse with intensive dialysis or with successful renal transplantation.

■ DERMATOLOGIC ABNORMALITIES

Abnormalities of the skin are prevalent in progressive CKD. Pruritus is quite common and one of the most vexing manifestations of the uremic state. In advanced CKD, even on dialysis, patients may become more pigmented, and this is felt to reflect the deposition of retained pigmented metabolites, or urochromes. Although many of the cutaneous abnormalities improve with dialysis, pruritus is often tenacious. The first lines of management are to rule out unrelated skin disorders, such as scabies, and to treat hyperphosphatemia, which can cause itch. Local moisturizers, mild topical glucocorticoids, oral antihistamines, and ultraviolet radiation have been reported to be helpful. Agonists of kappa opioid receptors can reduce pruritus intensity and improve quality of life.

A skin condition unique to CKD patients called nephrogenic fibrosing dermopathy consists of progressive subcutaneous induration, especially on the arms and legs. The condition is seen very rarely in patients with CKD who have been exposed to the magnetic resonance contrast agent gadolinium. Current recommendations are that patients with CKD stage 3 (GFR 30–59 mL/min) should minimize exposure to gadolinium and those with CKD stages 4–5 (GFR <30 mL/min) should avoid the use of gadolinium agents unless it is medically necessary. However, no patient should be denied an imaging investigation that is critical to management, and under such circumstances, rapid removal of gadolinium by hemodialysis (even in patients not yet receiving renal replacement therapy) shortly after the procedure may mitigate this sometimes devastating complication. Newer forms of gadolinium are not associated with this complication, and it remains to be seen if caution about their use will remain relevant.

EVALUATION AND MANAGEMENT OF PATIENTS WITH CKD

■ INITIAL APPROACH

History and Physical Examination Symptoms and overt signs of kidney disease are often subtle or absent until renal failure supervenes. Thus, the diagnosis of kidney disease often surprises patients and may lead to denial, especially given that pain in the region of the kidneys and decrease in urinary volume are not clinical features. Particular aspects of the history that are germane to renal disease include a history of hypertension (which can cause CKD or more commonly be a consequence of CKD), diabetes mellitus, abnormal urinalyses, and preeclampsia or early pregnancy loss. A careful drug history should be elicited. Drugs to consider include nonsteroidal anti-inflammatory agents, cyclooxygenase-2 (COX2) inhibitors, antimicrobials, chemotherapeutic agents, antiretroviral agents, proton pump inhibitors, phosphate-containing bowel cathartics, and lithium. In evaluating the uremic syndrome, questions about appetite, weight loss, nausea, hiccups, peripheral edema, muscle cramps, pruritus, and restless legs are especially helpful. A family history of kidney disease, together with assessment of manifestations in other organ systems such as auditory, visual, and integumentary, may lead to the diagnosis of a heritable form of CKD (e.g., Alport's or Fabry's disease, cystinosis) or shared environmental exposure to nephrotoxic agents (e.g., heavy metals, aristolochic acid). It should be noted that clustering of CKD, sometimes of different etiologies, is often observed within families.

The physical examination should focus on blood pressure and target organ damage from hypertension. Fundoscopy is especially important in the diabetic patient, because it may show evidence of diabetic retinopathy, which is associated with diabetic nephropathy. Other manifestations of CKD include edema and sensory polyneuropathy. The finding of asterixis or a pericardial friction rub not attributable to other causes signifies the presence of the uremic syndrome.

Laboratory Investigation Laboratory studies should focus on a search for clues to an underlying causative or aggravating disease process and on the degree of renal damage and its consequences. Serum and urine protein electrophoresis, looking for multiple myeloma, should be obtained in all patients >35 years old with unexplained CKD, especially if there is associated anemia and elevated, or even inappropriately normal, serum calcium concentration in the face of renal insufficiency. In the presence of hematuria or proteinuria, autoimmune diseases such as lupus and underlying infectious etiologies such as hepatitis B and C and HIV should be tested. Serial measurements of renal function should be obtained to determine the pace of deterioration and ensure that the disease is truly chronic, rather than acute or subacute and hence potentially reversible. Serum concentrations of calcium, phosphorus, and PTH should be measured to evaluate metabolic bone disease. Hemoglobin concentration, iron, vitamin B_{12}, and folate should also be evaluated. A 24-h urine collection may be helpful, because protein excretion >300 mg may be an indication for therapy with ACE inhibitors, ARBs, and SGLT2 inhibitors and also is associated with a higher risk of progression. If the patient has difficulty in carrying out a 24-h urine collection, a random urine albumin/creatinine or protein/creatinine is less accurate than a well-done 24-h collection but less cumbersome.

Imaging Studies The most useful imaging study is a renal ultrasound, which can verify the presence of two kidneys, determine if they are symmetric, provide an estimate of kidney size, and rule out renal masses and evidence of obstruction. Because it takes time for kidneys to shrink as a result of chronic disease, the finding of bilaterally small kidneys supports the diagnosis of CKD of long-standing duration. If the kidney size is normal, it is possible that the kidney disease is acute or subacute. The exceptions are diabetic nephropathy (where kidney size is increased at the onset of diabetic nephropathy before CKD supervenes), amyloidosis, and HIV nephropathy, where kidney size may be normal in the face of CKD. Polycystic kidney disease that has reached some degree of renal failure will almost always present with enlarged kidneys with multiple cysts **(Chap. 327)**. A discrepancy >1 cm in kidney length suggests either a unilateral developmental abnormality or a disease process or renovascular disease with arterial insufficiency affecting one kidney more than the other. The diagnosis of renovascular disease can be undertaken with different techniques, including Doppler sonography, nuclear medicine studies, or CT or magnetic resonance imaging (MRI) studies. If there is a suspicion of reflux nephropathy (recurrent childhood urinary tract infection, asymmetric renal size with scars on the renal poles), a voiding cystogram may be indicated. However, in most cases, by the time the patient has CKD, the reflux has resolved, and even if still present, repair does not improve renal function. Radiographic contrast imaging studies are not particularly helpful in the investigation of CKD. Intravenous or intraarterial dye should be avoided where possible in the CKD patient, especially with diabetic nephropathy, because of the risk of radiographic contrast dye–induced renal failure. When unavoidable, appropriate precautionary measures include avoidance of hypovolemia at the time of contrast exposure, minimization of the dye load, and choice of radiographic contrast preparations with the least nephrotoxic potential. Additional measures thought to attenuate contrast-induced worsening of renal function include judicious administration of sodium bicarbonate–containing solutions and N-acetylcysteine, although these agents may not be as effective as previously thought.

Kidney Biopsy In the patient with bilaterally small kidneys, renal biopsy is not advised because (1) it is technically difficult and has a greater likelihood of causing bleeding and other adverse consequences, (2) there is usually so much scarring that the underlying disease may not be apparent, and (3) the window of opportunity to render disease-specific therapy has passed. Other contraindications to renal biopsy include uncontrolled hypertension, active urinary tract infection, bleeding diathesis (including ongoing anticoagulation), and severe obesity. Ultrasound-guided percutaneous biopsy is the favored approach, but a surgical or laparoscopic approach can be considered, especially in the patient with a single kidney where direct visualization and control of bleeding are crucial. In the CKD patient in whom a kidney biopsy is indicated (e.g., suspicion of a concomitant or superimposed active process or in the face of accelerated loss of GFR), the bleeding time should be measured, and if increased, desmopressin should be administered immediately prior to the procedure.

A brief run of hemodialysis (without heparin) may also be considered prior to renal biopsy to normalize the bleeding time.

ESTABLISHING THE DIAGNOSIS AND ETIOLOGY OF CKD

The most important initial diagnostic step is to distinguish newly diagnosed CKD from acute or subacute renal failure, because the latter two conditions may respond to targeted therapy. Previous measurements of serum creatinine concentration are particularly helpful in this regard. Normal values from recent months or even years suggest that the current extent of renal dysfunction could be more acute, and hence reversible, than might otherwise be appreciated. In contrast, elevated serum creatinine concentration in the past suggests that the renal disease represents a chronic process. Even if there is evidence of chronicity, there is the possibility of a superimposed acute process (e.g., ECFV depletion, urinary infection or obstruction, or nephrotoxin exposure) supervening on the chronic condition. If the history suggests multiple systemic manifestations of recent onset (e.g., fever, polyarthritis, rash), it should be assumed that renal insufficiency is part of an acute systemic illness.

Although kidney biopsy can usually be performed in early CKD (stages 1–3), it is not always indicated. For example, in a patient with a history of type 1 diabetes mellitus for 15–20 years with retinopathy, nephrotic-range proteinuria, and absence of hematuria, the diagnosis of diabetic nephropathy is very likely and biopsy is usually not necessary. However, if there is another finding not typical of diabetic nephropathy, such as hematuria or white blood cell casts, or absence of diabetic retinopathy, some other disease may be present and a biopsy may be indicated.

In the absence of a clinical diagnosis, kidney biopsy may be the only recourse to establish an etiology in early-stage CKD. However, as noted above, once the CKD is advanced and the kidneys are small and scarred, there is little utility and significant risk in attempting to arrive at a specific diagnosis.

Genetic testing is increasingly entering the repertoire of diagnostic tests since the patterns of injury and kidney morphologic abnormalities often reflect overlapping causal mechanisms, whose origins can sometimes be attributed to a genetic predisposition or cause (Table 322-2). The increased application of genetic testing has often yielded surprising diagnoses that deviate from the cause of CKD suggested by clinical or even kidney pathology alone. Given this realization and the cumulative significant contribution of monogenic disease etiologies, consideration is now given to large CKD gene panels, chromosomal microarrays, whole exome or even whole genome sequencing, and other advanced genetic analysis technologies in the evaluation and management of CKD. Suggested indications for such genetic evaluation in patients without a clinically evident etiology and course include family history (family member[s] with unexplained kidney failure, dialysis, or kidney transplant recipient; family member affected with maturity-onset diabetes of the young, kidney disease together with psychiatric morbidity, hearing or vision loss, autism spectrum disorder, developmental delay, and/or intellectual disability; multiple family members affected with similar phenotype); parental consanguinity; extrarenal involvement (dysmorphic facial features, autism spectrum disorder, developmental delay, and/or intellectual disability); multiple congenital structural kidney anomalies; steroid-resistant nephrotic syndrome; persistent active clinical or laboratory manifestations of glomerulonephritis or glomerular endotheliopathy; imaging consistent with cystic kidney diseases/ciliopathies; young adult onset of unexplained progressive CKD; consideration of *APOL1*-mediated kidney disease; and unexplained kidney failure prior to kidney transplantation.

TREATMENT

Chronic Kidney Disease

Therapy for CKD can be divided into interventions that are directed at the specific etiology and those that attenuate progression related to the common pathway of glomerular hyperfiltration, which perpetuates kidney injury following reduced nephron mass, as discussed below. Treatments aimed at causes of CKD related to systemic diseases are discussed in the respective chapters. Recent developments in the etiology-directed therapy of CKD include the emergence of genome-specific therapies for certain patients with ADPKD **(Chap. 327)**, siRNA therapy (lumasiran) for type 1 hyperoxaluria, and highly specific inhibitors for *APOL1*-mediated kidney disease (AMKD). Inaxaplin, a specific inhibitor of APOL1 channel function, has been shown to decrease urine protein excretion in patients with focal segmental glomerulosclerosis caused by the high-risk genotypes at the *APOL1* gene. In the absence of validated and specific biomarkers of imminent or ongoing kidney injury, the optimal timing for such specific therapies is usually well before there has been a measurable decline in GFR and certainly before CKD is established. It is helpful to measure sequentially and plot the rate of decline of GFR in all patients. Any acceleration in the rate of decline should prompt a search for superimposed acute or subacute processes that may be reversible. These include ECFV depletion, uncontrolled hypertension, urinary tract infection, new obstructive uropathy, exposure to nephrotoxic agents (e.g., nonsteroidal anti-inflammatory drugs [NSAIDs] or radiographic dye), and reactivation or flare of the original disease, such as lupus or vasculitis.

SLOWING THE PROGRESSION OF CKD

There is variation in the rate of decline of GFR among patients with CKD. However, the several established and newer available interventions should be strongly considered in an effort to stabilize or slow the decline of renal function in most patients with reduced nephron mass at risk for progression from increased intraglomerular pressure and hyperfiltration as described above. Fortunately, these very same interventions also reduce cardiovascular complications of CKD and are therefore expected to greatly alleviate the global burden of CKD in the coming years to decades.

Reducing Intraglomerular Hypertension and Proteinuria Increased intraglomerular filtration pressures and glomerular hypertrophy develop as a response to loss of nephron number. This response is maladaptive as it promotes the ongoing decline of kidney function even if the inciting process has been treated or spontaneously resolved. Control of glomerular hypertension is important in slowing the progression of CKD. Moreover, elevated blood pressure increases proteinuria by increasing its flux across the glomerular capillaries. Conversely, the renoprotective effect of antihypertensive medications is gauged through the consequent reduction of proteinuria. Thus, the more effective a given treatment is in lowering protein excretion, the greater is the subsequent impact on protection from decline in GFR. This observation is the basis for the treatment guideline establishing 130/80 mmHg as a target blood pressure in proteinuric CKD patients.

Several controlled studies have shown that ACE inhibitors and ARBs are effective in slowing the progression of renal failure in patients with advanced stages of both diabetic and nondiabetic CKD, in large part through effects on efferent vasodilatation and the subsequent decline in glomerular hypertension. The combination of these two classes should be avoided, due to a demonstrated greater incidence of AKI and adverse cardiac events from such combination therapy. While a nonprogressive decrease in eGFR of up to 30% may be a tolerable reflection of effective unloading of glomerular hyperfiltration, a progressive increase in serum creatinine concentration with these agents may suggest the presence of renovascular disease within the large or small arteries.

Together with efferent arteriolar vasodilation, afferent vasoconstriction should reduce intraglomerular pressure. Indeed, gliflozins, which are a family of inhibitors of SGLT2 transporters in the proximal tubule, result in precisely this glomerular response by activating tubuloglomerular feedback **(Fig. 322-5)** and are indicated in slowing the decline of GFR in both diabetic and nondiabetic kidney disease. Other more direct cellular protective mechanisms of the gliflozins have been invoked in kidney and other tissues. This class of agents also has been demonstrated to reduce major cardiovascular events in CKD patients. Recent studies have also shown that the glucagon-like peptide receptor agonists reduce major cardiovascular events and CKD in at-risk patients.

Among the calcium channel blockers, diltiazem and verapamil may exhibit superior antiproteinuric and renoprotective effects compared to the dihydropyridines. At least two different categories of response can be considered: one in which progression is strongly associated with systemic and intraglomerular hypertension and proteinuria (e.g., diabetic nephropathy, glomerular diseases) and in which ACE inhibitors and ARBs are recommended choices, and another in which proteinuria is mild or absent initially (e.g., ADPKD and other tubulointerstitial diseases), where the contribution of intraglomerular hypertension is less prominent and other antihypertensive agents can be useful for control of systemic hypertension.

MANAGING OTHER COMPLICATIONS OF CKD

Medication Dose Adjustment Although the loading dose of most drugs is not affected by CKD because renal elimination is not used in the calculation, the maintenance doses of many drugs will need to

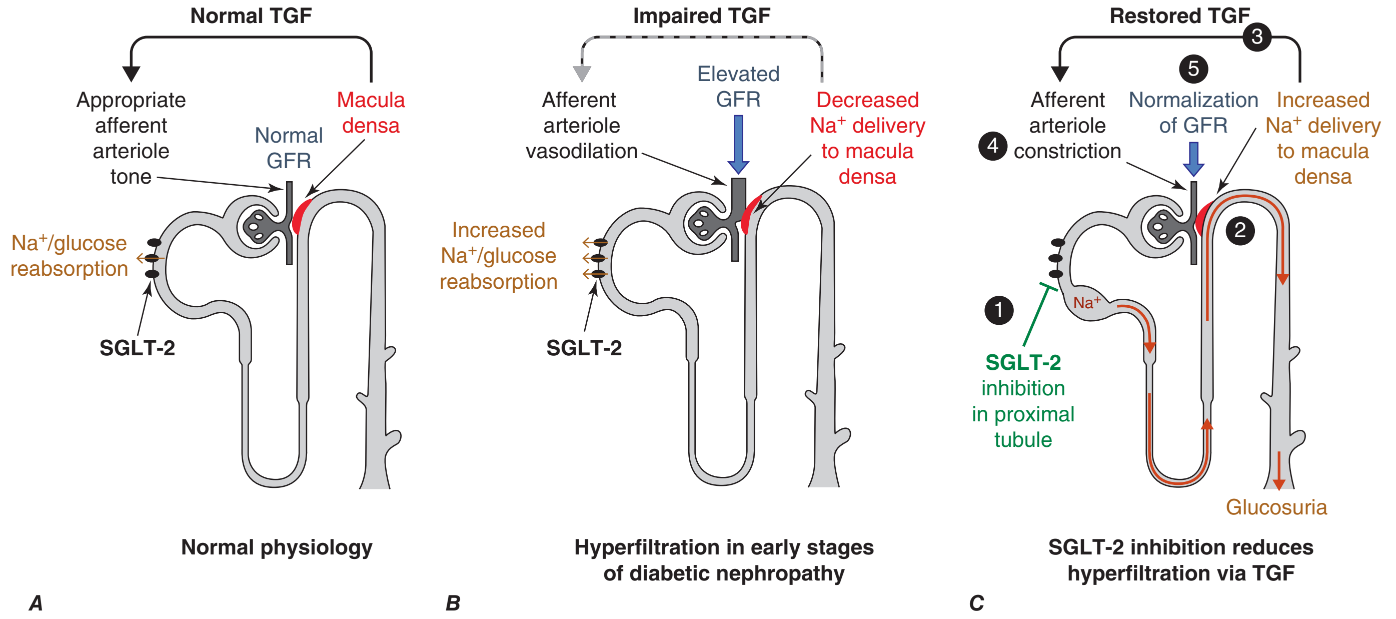
Normal TGF
Appropriate afferent arteriole tone
Macula densa
Normal GFR
Na⁺/glucose reabsorption
SGLT-2
Normal physiology
A
Impaired TGF
Elevated GFR
Afferent arteriole vasodilation
Decreased Na⁺ delivery to macula densa
Increased Na⁺/glucose reabsorption
SGLT-2
Hyperfiltration in early stages of diabetic nephropathy
B
Restored TGF
3
5
Afferent arteriole
4 constriction
Normalization of GFR
Increased Na⁺ delivery to macula densa
2
1
Na⁺
SGLT-2 inhibition in proximal tubule
Glucosuria
SGLT-2 inhibition reduces hyperfiltration via TGF
C

FIGURE 322-5 The postulated role of the gliflozins in generating tubuloglomerular feedback (TGF) to reduce intraglomerular hypertension. *(Reproduced with permission from DZ Cherney et al: Cherney et al: Circulation 129:587, 2014.)*

be adjusted. For those agents in which >70% excretion is by a nonrenal route, such as hepatic elimination, dose adjustment may not be needed. Some drugs that should be avoided include metformin, meperidine, and oral antihyperglycemics that are eliminated by the kidney. NSAIDs including COX2 inhibitors should be avoided because of the risk of further worsening of kidney function. Many antibiotics, antihypertensives, and antiarrhythmics may require a reduction in dosage or change in the dose interval. Several online Web-based databases for dose adjustment of medications according to stage of CKD or estimated GFR are available (e.g., *http://www.globalrph.com/index_renal.htm*). Nephrotoxic radiocontrast agents and gadolinium should be used according to strict guidelines when medically necessary, as discussed above.

PREPARATION FOR RENAL REPLACEMENT THERAPY

(See also Chap. 325.) Temporary relief of symptoms and signs of impending uremia, such as anorexia, nausea, vomiting, lassitude, and pruritus, may sometimes be achieved with dietary protein restriction. However, this diet carries a risk of malnutrition; thus, plans for more long-term management should be in place.

Maintenance dialysis and kidney transplantation have extended the lives of hundreds of thousands of patients with CKD worldwide. Clear indications for initiation of renal replacement therapy for patients with CKD include anorexia and nausea not attributable to reversible causes such as peptic ulcer disease, evidence of malnutrition, and fluid and electrolyte abnormalities, principally hyperkalemia or ECFV overload, that are refractory to other measures. Encephalopathy and pericarditis are very late complications, so it is now rare that they serve as indications for initiation of renal replacement therapy.

Recommendations for the Optimal Time for Initiation of Renal Replacement Therapy Because of the individual variability in the severity of uremic symptoms and renal function, it is ill-advised to assign an arbitrary urea nitrogen or creatinine level to the need to start dialysis. Moreover, patients may become accustomed to chronic uremia and deny symptoms, only to find that they feel better with dialysis and realize in retrospect how poorly they were feeling before its initiation.

Previous studies suggested that starting dialysis before the onset of severe symptoms and signs of uremia was associated with prolongation of survival. This led to the concept of “healthy” start and is congruent with the philosophy that it is better to keep patients feeling well rather than allowing them to become ill with uremia and then attempting to return them to better health with dialysis or transplantation. Although recent studies have not confirmed an association of early-start dialysis with improved patient survival, there may be merit in this approach for some patients. At a practical level, advanced preparation may help to avoid problems with the dialysis process itself (e.g., a poorly functioning fistula for hemodialysis or malfunctioning peritoneal dialysis catheter) and, thus, preempt the morbidity associated with resorting to the insertion of temporary hemodialysis access with its attendant risks of sepsis, bleeding, thrombosis, and association with accelerated mortality.

Patient Education Social, psychological, and physical preparation for the transition to renal replacement therapy and the choice of the optimal initial modality are best accomplished with a gradual approach involving a multidisciplinary team. Along with conservative measures discussed in the sections above, it is important to prepare patients with an intensive educational program, explaining the likelihood and timing of initiation of renal replacement therapy and the various forms of therapy available and the option of nondialytic conservative care. The more knowledgeable that patients are about hemodialysis (both in-center and home-based), peritoneal dialysis, and kidney transplantation, the easier and more appropriate will be their decisions. Patients who are provided with education are more likely to choose home-based dialysis therapy. This approach is of societal benefit because home-based therapy is less expensive in most jurisdictions and is associated with improved quality of life. The educational programs should be commenced no later than stage 4 CKD so that the patient has sufficient time and cognitive function to learn the important concepts, make informed choices, and implement preparatory measures for renal replacement therapy.

Exploration of social support is also important. Early education of family members for selection and preparation of a home dialysis helper or a biologically or emotionally related potential living kidney donor should occur long before the onset of symptomatic renal failure.

Kidney transplantation **(Chap. 325)** offers the best potential for complete rehabilitation because dialysis replaces only a small fraction of the kidneys’ filtration function and none of the other renal functions, including endocrine and anti-inflammatory effects. Generally, kidney transplantation follows a period of dialysis treatment, although preemptive kidney transplantation (usually from a living donor) can be carried out if it is certain that the renal failure is irreversible and, under such circumstances, is preferred to transplanting after a period of dialysis.

IMPLICATIONS FOR GLOBAL HEALTH

In contrast to the natural decline and successful eradication of many devastating infectious diseases, there is rapid growth in the prevalence of metabolic and vascular disease in developing countries. Diabetes mellitus is becoming increasingly prevalent in these countries, perhaps due in part to change in dietary habits, diminished physical activity, and weight gain. Therefore, it follows that there will be a proportionate increase in vascular and renal disease. According to a recent analysis, in the absence of effective implementation of significant preventive and therapeutic inroads, a forecasting analysis for the U.S. population suggests that the number of global years of life lost (YLLs) due to CKD is expected to rise from a prior value of ~26 million in 2016 to 52.5 million in 2040, accompanied by a rise in mortality from 1.2 million in 2016 to 3.1 million in 2040. These increases are predicted to move CKD in the YLL rankings from 16th in 2016 to 5th in 2040. This rise will be disproportionately large in many other regions of the world where CKD prevalence is already rising at alarming rates due to population aging and the rapid increase in diabetes, hypertension, and obesity.

Health care agencies must plan for improved screening of high-risk individuals for early detection, prevention, and treatment plans in these nations and must start considering options for improved availability of renal replacement therapies.

There is also increasing recognition of endemic nephropathies in developing countries that particularly target young males working in agriculture. The extent of morbidity and mortality associated with these nephropathies is only starting to be appreciated. It is unclear what the cause is, but population genetic risk, endemic nephrotoxins, exposure to pesticides, NSAID use, and chronic volume depletion have all been suggested to contribute. Global warming and air pollution have also been implicated in the development of kidney diseases.

FURTHER READING

Bhandari S et al: Renin-angiotensin system inhibition in advanced chronic kidney disease. N Engl J Med 387:2021, 2022.

Chertow G et al: Effects of dapagliflozin in stage 4 chronic kidney disease. J Am Soc Nephrol 32:2352, 2021.

Dahl NK et al: The clinical utility of genetic testing in the diagnosis and management of adults with chronic kidney disease. J Am Soc Nephrol 34:2039, 2023.

Egbuna OL et al: Inaxaplin for proteinruic kidney disease in persons with two APOL1 variants. N Engl J Med 388:969, 2023.

Kovesdy CP: Epidemiology of chronic kidney disease: An update 2022. Kidney Int Suppl (2011) 12:7, 2021.

Vivante A: Genetics of chronic kidney disease. N Engl J Med 391:627, 2024.

Yu JH et al: GLP-1 receptor agonists in diabetic kidney disease: Current evidence and future directions. Kidney Res Clin Pract 41:136, 2022.

323

Dialysis in the Treatment of Kidney Failure

Kathleen D. Liu, Glenn M. Chertow

Dialysis may be required for the treatment of either acute or chronic kidney disease (CKD). The use of continuous renal replacement therapies (CRRTs) and prolonged intermittent renal replacement therapy (PIRRT)/slow low-efficiency dialysis (SLED) is specific to the management of acute kidney injury/acute kidney disease and is discussed in **Chap. 321**. These modalities are performed continuously (CRRT) or over 6–12 h per session (PIRRT/SLED), in contrast to the 3–4 h of an intermittent hemodialysis session. Advantages and disadvantages of CRRT and PIRRT/SLED are discussed in **Chap. 321**.

Peritoneal dialysis is rarely used in developed countries for the treatment of acute kidney injury/acute kidney disease because of the increased risk of infection and (as will be discussed in more detail below) less efficient clearance per unit of time. The focus of this chapter will be on the use of peritoneal and hemodialysis for end-stage kidney disease (ESKD).

With the widespread availability of dialysis, the lives of hundreds of thousands of patients with ESKD have been prolonged. In the United States alone, there are now >800,000 patients with treated ESKD (kidney failure requiring dialysis or transplantation), the vast majority of whom require dialysis. Since 2000, the prevalence of treated ESKD has increased 65%, which predominantly reflects enhanced survival of patients receiving dialysis. The crude incidence rate for treated ESKD in the United States is 363 cases per million population per year; this incidence has slowly fallen over time. ESKD is disproportionately higher in African Americans as compared with white Americans. In the United States, the leading cause of ESKD is diabetes mellitus, currently accounting for ~45% of newly diagnosed cases of ESKD. Approximately 30% of patients have ESKD that has been attributed to hypertension, although it is unclear whether in these cases hypertension is the cause or a consequence of vascular disease or other unknown causes of kidney failure. Other prevalent causes of ESKD include glomerulonephritis, polycystic kidney disease, and obstructive uropathy. A fraction of the excess incidence of ESKD in African Americans is likely related to transmission of high-risk alleles for the *APOL1* gene.

Globally, mortality rates for patients with ESKD are lowest in Europe and Japan but very high in the developing world because of the limited availability of dialysis. In the United States, the mortality rate of patients on dialysis had decreased somewhat but remained extremely high prior to the COVID-19 pandemic. During the pandemic, there was a significant increase in the mortality of patients receiving hemodialysis or peritoneal dialysis, but we remove the point estimates. Deaths are due mainly to cardiovascular diseases and infections. Older age, male sex, nonblack race, diabetes mellitus, malnutrition, and underlying heart disease are important predictors of death.

TREATMENT OPTIONS FOR PATIENTS WITH ESKD

Commonly accepted criteria for initiating patients on maintenance dialysis include the presence of uremic symptoms, the presence of hyperkalemia unresponsive to conservative measures, persistent extracellular volume expansion despite diuretic therapy, acidosis refractory to medical therapy, a bleeding diathesis, and a creatinine clearance or estimated glomerular filtration rate (GFR) <10 mL/min per 1.73 m^2, although we should emphasize that there is no absolute level of serum creatinine, cystatin C, blood urea nitrogen (BUN), or estimated GFR that should be considered an absolute indication (see **Chap. 322** for estimating equations). Timely referral to a nephrologist for advanced planning and creation of a dialysis access, education about ESKD treatment options, and management of the complications of advanced CKD, including hypertension, anemia, metabolic acidosis, and disorders of bone and mineral metabolism, including secondary hyperparathyroidism, is advisable. Recent data have suggested that a sizable fraction of ESKD cases result following episodes of acute kidney injury, particularly among persons with underlying CKD. Furthermore, there is no benefit to initiating dialysis preemptively at a GFR of 10–14 mL/min per 1.73 m^2 compared to initiating dialysis for symptoms of uremia.

In ESKD, treatment options include hemodialysis (in a center or at home); peritoneal dialysis, as either continuous ambulatory peritoneal dialysis (CAPD) or continuous cyclic peritoneal dialysis (CCPD); or transplantation **(Chap. 325)**. Although there are significant geographic variations and differences in practice patterns, in-center hemodialysis remains the most common therapeutic modality for ESKD (>80% of patients) in the United States. In contrast to hemodialysis, peritoneal dialysis is continuous, but much less efficient in terms of solute clearance. While no large-scale clinical trials have been completed comparing outcomes among patients randomized to either hemodialysis or peritoneal dialysis, outcomes associated with both therapies are similar in most reports, and the decision of which modality to select is often based on personal preferences and quality-of-life considerations.

HEMODIALYSIS

Hemodialysis relies on the principles of solute diffusion across a semipermeable membrane. Movement of by-products of metabolism (often referred to as "waste products" or "toxins"—terms that may understandably create anxiety in patients) takes place down a concentration gradient from the circulation into the dialysate. The rate of diffusive transport increases in response to several factors, including the magnitude of the concentration gradient, the membrane surface area, and the mass transfer coefficient of the membrane. The latter is a function of the porosity and thickness of the membrane, the size of the solute molecule, and the conditions of flow on the two sides of the membrane. According to laws of diffusion, the larger the molecule, the slower is its rate of transfer across the membrane. A small molecule, such as urea (60 Da), undergoes substantial clearance, whereas a larger molecule, such as creatinine (113 Da), is cleared less efficiently. In addition to diffusive clearance, movement of waste products from the circulation into the dialysate may occur as a result of ultrafiltration. Convective clearance occurs because of solvent drag, with solutes being swept along with water across the semipermeable dialysis membrane.

■ THE DIALYZER

There are three essential components to hemodialysis: the dialyzer, the composition and delivery of the dialysate, and the blood delivery system **(Fig. 323-1)**. The dialyzer is a plastic chamber with the ability to perfuse blood and dialysate compartments simultaneously at very high flow rates. The hollow-fiber dialyzer is the most common in use in the United States. These dialyzers are composed of bundles of capillary tubes through which blood circulates while dialysate travels on the outside of the fiber bundle. Virtually all dialyzers now manufactured in the United States are "biocompatible" synthetic membranes derived from polysulfone or related compounds (versus older cellulose "bioincompatible" membranes that activated the complement cascade). The frequency of reprocessing and reuse of hemodialyzers and blood lines varies across the world. In general, as the cost of disposable supplies has decreased, the use of single-use dialyzers has increased. In the United States, reprocessing of dialyzers is now extremely rare. Formaldehyde, peracetic acid–hydrogen peroxide, glutaraldehyde, and bleach have all been used as reprocessing agents.

■ DIALYSATE

The potassium concentration of dialysate may be varied from 0–4 mmol/L depending on the predialysis serum potassium concentration. The use of 0- or 1-mmol/L potassium dialysate is becoming much less common owing to data suggesting that patients who undergo treatments with very low potassium dialysate have an increased risk of sudden death, perhaps due to arrhythmias in the setting of potassium shifts. The usual dialysate calcium concentration is 1.25 mmol/L (2.5 mEq/L), although modification may be required in selected settings (e.g., higher dialysate calcium concentrations may be used in patients with hypocalcemia

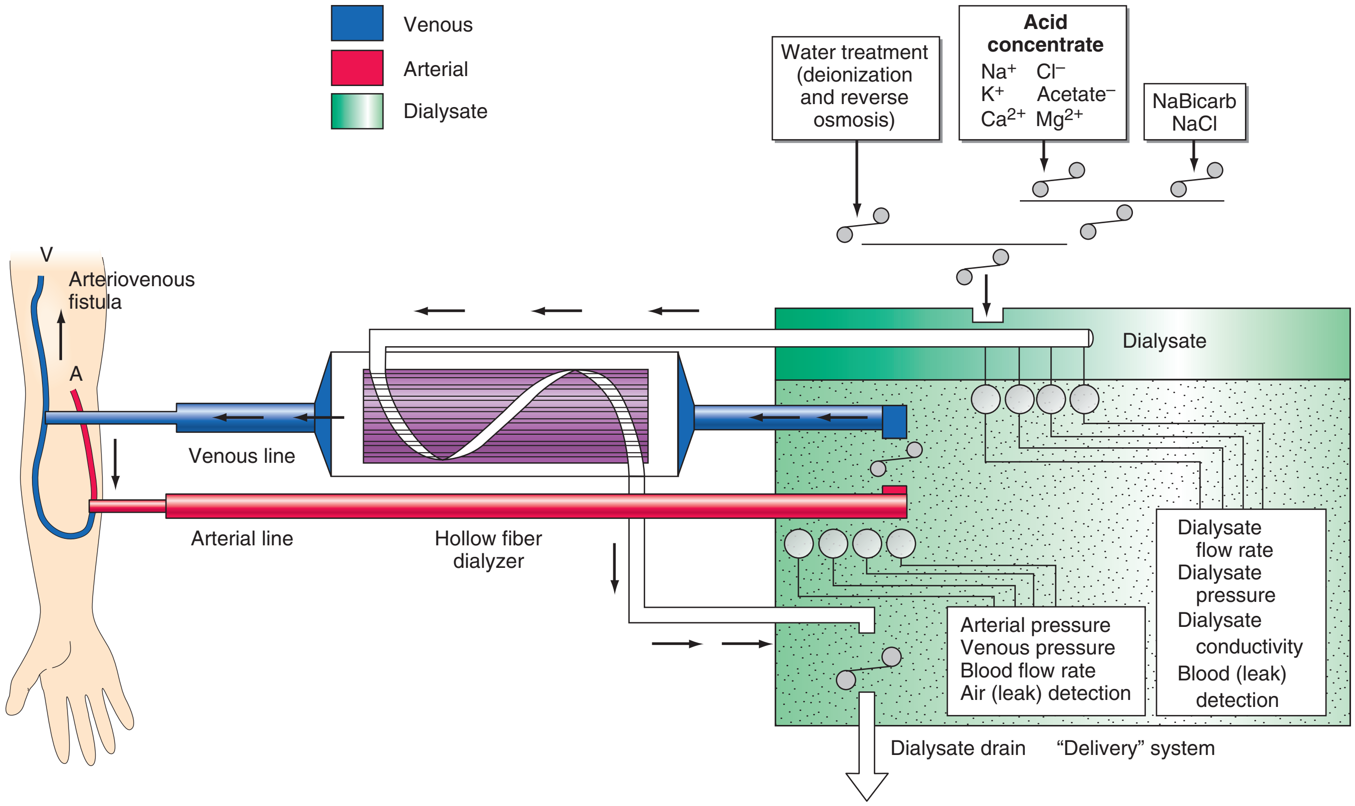

FIGURE 323-1 Schema for hemodialysis.

associated with secondary hyperparathyroidism or with “hungry bone syndrome” following parathyroidectomy). The usual dialysate sodium concentration is 135–140 mmol/L. In patients who frequently develop hypotension during their dialysis run, “sodium modeling” to counterbalance urea-related osmolar gradients may be employed. With sodium modeling, the dialysate sodium concentration is gradually lowered from the range of 145–155 mmol/L to isotonic concentrations (135–140 mmol/L) near the end of the dialysis treatment, typically declining either in steps or in a linear or exponential fashion. However, higher dialysate sodium concentrations and sodium modeling may predispose patients to positive sodium balance and increased thirst; thus, these strategies to ameliorate intradialytic hypotension may be undesirable in patients with hypertension or in patients with large interdialytic weight gains. Because patients are exposed to ~120 L of water during each dialysis treatment, water used for the dialysate is subjected to filtration, softening, deionization, and, ultimately, reverse osmosis to remove microbiologic contaminants and dissolved ions.

BLOOD DELIVERY SYSTEM

The blood delivery system is composed of the extracorporeal circuit and the dialysis access. The dialysis machine consists of a blood pump, dialysis solution delivery system, and various safety monitors. The blood pump moves blood from the access site, through the dialyzer, and back to the patient. The blood flow rate typically ranges from 250–450 mL/min, depending on the type and integrity of the vascular access and needle gauge. Negative hydrostatic pressure on the dialysate side can be manipulated to achieve desirable fluid removal or *ultrafiltration*. Dialysis membranes have different ultrafiltration coefficients (i.e., mL removed/min per mmHg) so that along with hydrostatic changes, fluid removal can be varied. The dialysis solution delivery system dilutes the concentrated dialysate with water and monitors the temperature, conductivity, and flow of dialysate.

DIALYSIS ACCESS

The fistula, graft, or catheter through which blood is obtained for hemodialysis is often referred to as a *hemodialysis (or vascular) access*. A native fistula created by the anastomosis of an artery to a vein (e.g., the Brescia-Cimino fistula, in which the cephalic vein at the wrist is anastomosed end-to-side to the radial artery) results in arterialization of the vein. This facilitates its subsequent use in the placement of large needles (typically 15 or 16 gauge) to access the circulation. Fistulas have the highest long-term patency rate of all hemodialysis access options. For patients in whom fistulas fail to mature or in patients whose vasculature does not allow creation of a successful fistula (i.e., poor arterial inflow or recipient veins of inadequate caliber), patients undergo placement of an arteriovenous graft (i.e., the interposition of prosthetic material, usually polytetrafluoroethylene, between an artery and a vein) or a tunneled hemodialysis catheter. Nephrologists, vascular surgeons, and health care policymakers in the United States have encouraged creation of arteriovenous fistulas in a larger fraction of patients (the “fistula first” initiative). Unfortunately, even when created, arteriovenous fistulas may not mature sufficiently to provide reliable access to the circulation, or they may thrombose early in their development.

The most important complication of arteriovenous grafts is thrombosis of the graft and graft failure, due principally to intimal hyperplasia at the anastomosis between the graft and recipient vein. When grafts (or fistulas) fail, catheter-guided angioplasty can be used to dilate stenoses; monitoring of venous pressures on dialysis and of access flow, although not universally performed, may assist in the early recognition of impending vascular access failure. In addition to increased rates of access failure, grafts and (in particular) catheters are associated with much higher rates of infection than fistulas.

Intravenous large-bore catheters are often used in patients with acute renal failure and CKD. For persons on maintenance hemodialysis, tunneled catheters (either two separate catheters or a single catheter with two lumens) are often used when arteriovenous fistulas and grafts have failed or are not feasible due to anatomic considerations. These catheters are tunneled under the skin; the tunnel reduces bacterial translocation from the skin, resulting in a lower infection rate than with nontunneled temporary catheters. Most tunneled catheters are placed in the internal jugular veins; the external jugular, femoral, and subclavian veins may also be used. Infection, venous thrombosis, and venous stenosis resulting in swelling of the extremity or superior vena cava syndrome are complications best avoided by limiting the time during which catheters are employed.

Nephrologists, interventional radiologists, and vascular surgeons generally prefer to avoid placement of catheters into the subclavian veins; while flow rates are usually excellent, subclavian stenosis is a frequent complication and, if present, will likely prohibit permanent vascular access (i.e., a fistula or graft) in the ipsilateral extremity. Infection rates may be higher with femoral catheters. For patients with multiple vascular access complications and no other options for permanent vascular access, tunneled catheters may be the last "lifeline" for hemodialysis. Translumbar or transhepatic approaches into the inferior vena cava may be required if the superior vena cava or other central veins draining the upper extremities are stenosed or thrombosed.

GOALS OF DIALYSIS

The hemodialysis procedure consists of pumping heparinized blood through the dialyzer at a flow rate of 250–450 mL/min, while dialysate flows in an opposite *counter-current* direction at 500–800 mL/min. The efficiency of dialysis is determined by blood and dialysate flow through the dialyzer as well as dialyzer characteristics (i.e., its efficiency in removing solute). The *dose* of dialysis, which is currently defined as a derivation of the fractional urea clearance during a single treatment, is further governed by patient size (in particular, the total body water or urea volume of distribution, which resides to large extent in the skeletal muscle), residual kidney function, dietary protein intake, the degree of anabolism or catabolism, and the presence of comorbid conditions, including, in particular, heart failure.

Since the landmark studies of Sargent and Gotch relating the measurement of the dose of dialysis using urea concentrations with morbidity in the National Cooperative Dialysis Study, the *delivered* dose of dialysis has been measured and considered as a quality assurance and improvement tool. While the fractional removal of urea nitrogen and derivations thereof are considered to be the standard methods by which "adequacy of dialysis" is measured, a large multicenter randomized clinical trial (the HEMO study) failed to show a difference in mortality associated with a large difference in per-session urea clearance. Current targets include a urea reduction ratio (the fractional reduction in blood urea nitrogen per hemodialysis session) of >65–70% and a total body water–indexed clearance × time product (Kt/V) >1.2 or 1.05, depending on whether urea concentrations are "equilibrated." For the majority of patients with ESKD, 9–12 h of dialysis are required each week, usually divided into three equal sessions. Several studies have suggested that longer hemodialysis session lengths may be beneficial (independent of urea clearance), although these studies are confounded by a variety of patient characteristics, including body size and nutritional status. Hemodialysis "dose" should be individualized, and factors other than the urea nitrogen should be considered, including the adequacy of ultrafiltration or fluid removal and control of hyperkalemia, hyperphosphatemia, and metabolic acidosis. A randomized clinical trial comparing hemodialysis performed 6 versus 3 times per week (the Frequent Hemodialysis Network Daily Trial) demonstrated improved control of hypertension and hyperphosphatemia, reduced left ventricular mass, and improved self-reported physical health with more frequent hemodialysis. Secondary analyses also demonstrated improvements in other metrics of health-related quality of life, including improved self-reported general health and a reduced "time to recovery" (time until usual activities can be resumed) among patients randomized to more frequent hemodialysis. A companion trial in which frequent nocturnal hemodialysis was compared to conventional hemodialysis at home showed no significant effect on left ventricular mass or self-reported physical health. Finally, an evaluation of the U.S. Renal Data System registry showed a significant increase in mortality and hospitalization for heart failure after the longer interdialytic interval that occurs over the dialysis "weekend."

COMPLICATIONS DURING HEMODIALYSIS

Hypotension is the most common acute complication of hemodialysis, particularly among patients with diabetes mellitus. Numerous factors appear to increase the risk of hypotension, including excessive ultrafiltration with inadequate compensatory vascular filling, impaired vasoactive or autonomic responses, osmolar shifts, overzealous use of antihypertensive agents, and reduced cardiac reserve. Patients with arteriovenous fistulas and grafts may develop high-output cardiac failure due to shunting of blood through the dialysis access; on rare occasions, this may necessitate ligation of the fistula or graft. The management of hypotension during dialysis consists of discontinuing ultrafiltration, the administration of 100–250 mL of isotonic saline, or administration of salt-poor albumin, although the latter is generally unavailable in outpatient settings. Hypotension during dialysis can frequently be prevented by careful evaluation of the dry weight and by ultrafiltration modeling, such that more fluid is removed at the beginning rather than the end of the dialysis procedure. Excessively rapid fluid removal (>13 mL/kg per h) should be avoided, as rapid fluid removal has been associated with adverse outcomes, including cardiovascular deaths. Additional maneuvers to prevent intradialytic hypotension include the performance of sequential ultrafiltration followed by dialysis, cooling of the dialysate during dialysis treatment, and avoiding heavy meals during dialysis. Midodrine, an oral selective α_1 adrenergic agent, has been advocated by some practitioners, although there is insufficient evidence of its safety and efficacy to support its routine use.

Muscle cramps during dialysis are also a common complication. The etiology of dialysis-associated cramps remains obscure. Changes in muscle perfusion because of excessively rapid volume removal or targeted removal below the patient's estimated dry weight often precipitate dialysis-associated cramps. Strategies that may be used to prevent cramps include reducing volume removal during dialysis, ultrafiltration profiling, and the use of sodium modeling (see above).

Anaphylactoid reactions to the dialyzer, particularly on its first use, have been reported, most frequently with the bioincompatible cellulosic-containing membranes, which are rarely used today. Dialyzer reactions can be divided into two types, A and B. Type A reactions are attributed to an IgE-mediated intermediate hypersensitivity reaction to ethylene oxide used in the sterilization of new dialyzers. This reaction typically occurs soon after the initiation of a treatment (within the first few minutes) and can progress to full-blown anaphylaxis. The type B reaction consists of a symptom complex of nonspecific chest and back pain, which appears to result from complement activation and cytokine release. These symptoms typically occur several minutes into the dialysis run and typically resolve over time with continued dialysis.

PERITONEAL DIALYSIS

In peritoneal dialysis, 1.5–3 L of a dextrose-containing solution is infused into the peritoneal cavity and allowed to dwell for a set period of time, usually 2–4 h. As with hemodialysis, metabolic by-products are removed through a combination of convective clearance generated through ultrafiltration and diffusive clearance down a concentration gradient. The clearance of solutes and water during a peritoneal dialysis exchange depends on the balance between the movement of solute and water into the peritoneal cavity versus absorption from the peritoneal cavity. The rate of diffusion diminishes with time and eventually stops when equilibration between plasma and dialysate is reached. Absorption of solutes and water from the peritoneal cavity occurs across the peritoneal membrane into the peritoneal capillary circulation and via peritoneal lymphatics into the lymphatic circulation. The rate of peritoneal solute transport varies from patient to patient and may be altered by the presence of infection (peritonitis), drugs, and physical factors such as position and exercise.

FORMS OF PERITONEAL DIALYSIS

Peritoneal dialysis may be carried out as CAPD, CCPD, or a combination of both. In CAPD, dialysate is manually infused into the peritoneal cavity and exchanged three to five times during the day. A nighttime dwell is frequently instilled at bedtime and remains in the peritoneal cavity through the night. In CCPD, exchanges are performed in an automated fashion, usually at night; the patient is connected to an automated cycler that performs a series of exchange cycles while the patient sleeps. The number of exchange cycles required to optimize peritoneal solute clearance varies by the peritoneal membrane characteristics; as with hemodialysis, solute clearance should be tracked to ensure dialysis "adequacy."

Peritoneal dialysis solutions are available in volumes typically ranging from 1.5–3 L. The major difference between the dialysate used for peritoneal dialysis rather than hemodialysis is that the hypertonicity of peritoneal dialysis solutions drives solute and fluid removal, whereas solute removal in hemodialysis depends on concentration gradients, and fluid removal requires transmembrane pressure. Typically, dextrose at varying concentrations contributes to the hypertonicity of peritoneal dialysate. In some cases in which patients have developed severe protein-energy malnutrition or have struggled with insufficient ultrafiltration, amino acid–containing peritoneal dialysis solutions have been added to the CAPD or CCPD regimen. Icodextrin is a nonabsorbable carbohydrate that is frequently used in place of dextrose during a single daily exchange. Studies have demonstrated more efficient ultrafiltration with icodextrin than with dextrose-containing solutions. Icodextrin is typically used as the "last fill" for patients on CCPD or for the longest dwell in patients on CAPD. The most common additives to peritoneal dialysis solutions are heparin to prevent obstruction of the dialysis catheter lumen with fibrin and antibiotics during an episode of acute peritonitis. Insulin may also be added in patients with diabetes mellitus.

ACCESS TO THE PERITONEAL CAVITY

Access to the peritoneal cavity is obtained through a peritoneal catheter. Catheters used for maintenance peritoneal dialysis are flexible, being made of silicone rubber with numerous side holes at the distal end. These catheters usually have two Dacron cuffs. The scarring that occurs around the cuffs anchors the catheter and seals it from bacteria tracking from the skin surface into the peritoneal cavity; it also prevents the external leakage of fluid from the peritoneal cavity. The cuffs are placed in the preperitoneal plane and ~2 cm from the skin surface.

The *peritoneal equilibrium test* is a formal evaluation of peritoneal membrane characteristics that measures the transfer rates of creatinine and glucose across the peritoneal membrane. Patients are classified as low, low–average, high–average, and high transporters. Patients with rapid equilibration (i.e., high transporters) tend to absorb more glucose and lose efficiency of ultrafiltration with long daytime dwells. High transporters also tend to lose larger quantities of albumin and other proteins across the peritoneal membrane. In general, patients with rapid transporting characteristics require more frequent, shorter dwell-time exchanges, nearly always obligating use of a cycler. Slower (low and low–average) transporters tend to do well with fewer exchanges. The efficiency of solute clearance also depends on the volume of dialysate infused. Larger volumes allow for greater solute clearance, particularly with CAPD in patients with low and low–average transport characteristics.

As with hemodialysis, the optimal dose of peritoneal dialysis is unknown. Several observational studies have suggested that higher rates of urea and creatinine clearance (the latter generally measured in L/week) are associated with lower mortality rates and fewer uremic complications. However, a randomized clinical trial (Adequacy of Peritoneal Dialysis in Mexico [ADEMEX]) failed to show a significant reduction in mortality or complications with a relatively large increment in urea clearance. In general, patients on peritoneal dialysis do well when they retain residual kidney function. Rates of technique failure increase with years on dialysis and have been correlated with loss of residual function to a greater extent than loss of peritoneal membrane capacity. For some patients in whom CCPD does not provide sufficient solute clearance, a hybrid approach can be adopted where one or more daytime exchanges are added to the CCPD regimen. While this approach can enhance solute clearance and prolong a patient's capacity to remain on peritoneal dialysis, the burden of the hybrid approach can be overwhelming.

COMPLICATIONS DURING PERITONEAL DIALYSIS

The major complications of peritoneal dialysis are peritonitis, catheter-associated nonperitonitis infections, weight gain and other metabolic disturbances, and residual uremia (especially among patients with little or no residual kidney function).

Peritonitis typically develops when there has been a break in sterile technique during one or more of the exchange procedures. Peritonitis is usually defined by an elevated peritoneal fluid leukocyte count ($100/mm^3$, of which at least 50% are polymorphonuclear neutrophils); these cutoffs are lower than in spontaneous bacterial peritonitis because of the presence of dextrose in peritoneal dialysis solutions and rapid bacterial proliferation in this environment without antibiotic therapy. The clinical presentation typically consists of pain and cloudy dialysate, often with fever and other constitutional symptoms. The most common culprit organisms are gram-positive cocci, including *Staphylococcus*, reflecting the origin from the skin. Gram-negative rod infections are less common; fungal and mycobacterial infections can be seen in selected patients, particularly after antibacterial therapy. Most cases of peritonitis can be managed either with intraperitoneal or oral antibiotics, depending on the organism; many patients with peritonitis do not require hospitalization. In cases where peritonitis is due to hydrophilic gram-negative rods (e.g., *Pseudomonas* spp.) or yeast, antimicrobial therapy is usually not sufficient, and catheter removal is required to ensure complete eradication of infection. Nonperitonitis catheter-associated infections (often termed *tunnel infections*) vary widely in severity. Some cases can be managed with local antibiotic or silver nitrate administration, while others are severe enough to require parenteral antibiotic therapy and catheter removal.

Peritoneal dialysis is associated with a variety of metabolic complications. Albumin and other proteins can be lost across the peritoneal membrane in concert with the loss of metabolic wastes. Hypoproteinemia obligates a higher dietary protein intake in order to maintain nitrogen balance. Hyperglycemia and weight gain are also common complications of peritoneal dialysis. Several hundred calories in the form of dextrose are absorbed each day, depending on the concentration of dextrose employed. Patients receiving peritoneal dialysis, particularly those with diabetes mellitus, are prone to other complications of insulin resistance, including hypertriglyceridemia. On the positive side, the continuous nature of peritoneal dialysis usually allows for a more liberal diet due to continuous removal of potassium and phosphorus—two major dietary components whose accumulation can be hazardous in ESKD.

LONG-TERM OUTCOMES IN ESKD

Cardiovascular disease constitutes the major cause of death in patients with ESKD. Cardiovascular mortality and event rates are higher in patients receiving dialysis than in patients posttransplantation, although rates are extraordinarily high in both populations. The underlying cause of cardiovascular disease is unclear but may be related to shared risk factors (e.g., diabetes mellitus, hypertension, atherosclerotic and arteriosclerotic vascular disease), chronic inflammation, massive changes in extracellular volume (especially with high interdialytic weight gains), inadequate treatment of hypertension, dyslipidemia, anemia, dystrophic (vascular) calcification, and, perhaps, alterations in cardiovascular dynamics during the dialysis treatment. Few studies have targeted cardiovascular risk reduction in patients with ESKD; none has demonstrated consistent benefit. Two clinical trials of statin agents in ESKD demonstrated significant reductions in low-density lipoprotein (LDL) cholesterol concentrations but no significant reductions in death or cardiovascular events (Die Deutsche Diabetes Dialyse Studie [4D] and AURORA studies). The Study of Heart and Renal Protection (SHARP), which included patients on dialysis and others with non-dialysis-requiring CKD, showed a 17% reduction in the rate of major cardiovascular events or cardiovascular death with simvastatin-ezetimibe treatment. Most experts recommend conventional cardioprotective strategies (e.g., lipid-lowering agents, aspirin, inhibitors of the renin-angiotensin-aldosterone system, and β-adrenergic antagonists) in patients receiving dialysis based on the patients' cardiovascular risk profile, which appears to be increased by more than an order of magnitude relative to persons unaffected by kidney disease. To date, sodium-glucose cotransporter 2 (SGLT-2) inhibitors and glucagon-like peptide 1 (GLP-1) receptor agonists have not been formally studied in the context of ESKD for cardiovascular risk reduction. Other complications of ESKD include a high incidence of infection, progressive debility and frailty, protein-energy malnutrition, and impaired cognitive function.

GLOBAL PERSPECTIVE

The incidence of ESKD is increasing worldwide with longer life expectancies and improved care of infectious and cardiovascular diseases. The management of ESKD varies widely by country and within country by region, and it is influenced by economic and other major factors. In general, peritoneal dialysis is more commonly performed in poorer countries owing to its lower expense and the high cost of establishing in-center hemodialysis units.

FURTHER READING

Cooper BA et al: A randomized, controlled trial of early versus late initiation of dialysis. N Engl J Med 363:609, 2010.

Correa-Rotter R et al: Peritoneal dialysis, in *Brenner and Rector's The Kidney*, 11th ed, MW Taal et al (eds). Philadelphia, Elsevier, 2020.

Fellstrom BC et al: Rosuvastatin and cardiovascular events in patients undergoing hemodialysis. N Engl J Med 360:1395, 2009.

Flythe JE et al: Rapid fluid removal during dialysis is associated with cardiovascular morbidity and mortality. Kidney Int 79:250, 2011.

Foley RN et al: Long interdialytic interval and mortality among patients receiving hemodialysis. N Engl J Med 365:1099, 2011.

Frequent Hemodialysis Network Trial Group: In-center hemodialysis six times per week versus three times per week. N Engl J Med 363:2287, 2010.

National Kidney Foundation: Kidney disease quality initiative clinical practice guidelines: Hemodialysis and peritoneal dialysis adequacy, 2006. Available at *http://www.kidney.org/professionals/kdoqi/guidelines.cfm*.

Rocco MV et al: The effects of frequent nocturnal home hemodialysis: The frequent hemodialysis network nocturnal trial. Kidney Int 80:1080, 2011.

U.S. Renal Data System: USRDS 2021 Annual Data Report: Atlas of End-Stage Renal Disease in the United States. Bethesda, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Disease, 2021.

324

Interventional Nephrology

Dirk M. Hentschel

Interventional nephrology is a procedure-oriented subspecialty with a focus on dialysis access for peritoneal and hemodialysis, typically performed under fluoroscopy. Ultrasound (US) evaluation of dialysis access is common, and some practitioners perform renal and renal artery US evaluation as well as renal biopsies. Endovascular creation of arteriovenous fistulas (AVFs) is a recent addition to the procedural spectrum; (open) surgical access creation by nephrologists is limited to very few centers in the United States, while common in China, Germany, India, and Italy.

Interventional nephrologists (INs) usually provide patient care in multidisciplinary teams that include clinical nephrologists; access surgeons with vascular, transplant, or general surgery background; other interventionalists (with radiology or cardiology training); and dialysis unit access coordinators, nurses, and technicians involved in needle placement. Long-term preservation of venous and arterial vascular access options is one tenet of chronic kidney disease (CKD) care, leading INs to advocate for specific vascular access options (tunneled small-diameter catheters over peripherally inserted central catheters [PICCs]) and cardiac devices (epicardial rather than endovascular lead passage).

HISTORY

The history of vascular access for hemodialysis is closely tied to the history of dialysis. The first hemodialysis treatments in humans were performed in 1924 using glass needles to access the radial artery and return blood into the cubital vein. In 1943, a "rotating drum kidney" was used to dialyze a 29-year-old housemaid with CKD by surgical exposure of different arteries until she ran out of access sites after 12 treatments. The challenge of repetitive vascular access prevented dialysis from becoming a routine method for the treatment of CKD until the development of an arteriovenous Teflon shunt and then the development of an autogenous arterial-venous access (arteriovenous fistula, AVF) by side-to-side-anastomoses between the radial artery and the cephalic vein at the wrist (Cimino fistula). Catheter-based approaches for chronic renal replacement therapy (RRT) were designed initially in 1961 for hemodialysis and in 1968 for peritoneal dialysis, both using Dacron felt cuffs to protect against infection.

Material sciences have continued to evolve the development of grafts for use in hemodialysis. A modified bovine carotid artery biological graft was introduced in 1972, followed by the use of expanded polytetrafluoroethylene (ePTFE) grafts in 1976, and most recently in 2016, tissue-engineered blood vessels from human fibroblasts and endothelial cells. Some ePTFE grafts are modified with a silicone layer to allow for early cannulation within days of insertion. Ultra-high-pressure (up to 40 atm) angioplasty balloons are a mainstay of peripheral and central venous therapy, and Nitinol self-expanding stents and stent grafts serve as rescue tools for unsuccessful angioplasty as well as vessel rupture with extravasation. The role of drug-coated balloons in dialysis access care continues to be assessed, although greater cost and, to date, mixed clinical trial results have limited widespread use.

PHYSIOLOGY AND PATHOPHYSIOLOGY OF DIALYSIS ACCESS

Peritoneal Dialysis Peritoneal dialysis (PD) catheters can be placed fluoroscopically, peritoneoscopically, laparoscopically, and open surgically. Procedural success is typically linked to provider experience and procedural planning to optimize positioning of the PD catheter coil as this improves function and decreases drain pain and other complications. The internal cuff is placed within the rectus sheath just laterally to the linea alba, while the external PD catheter cuff should be located 2–4 cm from the skin exit site. Ingrowth of both cuffs ensures secure positioning of the catheter and allows water emersion. Over time, the peritoneal catheter can become encased in a fibrinous sheath, which, if limiting fluid flow during exchanges, can be disrupted by guidewire manipulation. Omental entrapment of the catheter often requires laparoscopic intervention; omentopexy at the time of PD catheter placement can prevent later entrapment. Repeated infections affect the permeability of the peritoneal membrane, as does long-term exposure to glucose-containing exchange solutions. Encapsulating peritoneal sclerosis is a late-stage complication of PD thought to be triggered by repeated peritonitis.

Hemodialysis Catheters Dialysis catheters are typically made of polyurethane that softens at body temperature but is sufficiently strong to allow for blood flow rates of 400–500 mL/min in each of two channels inside a 14.5–16 French design without collapse of the catheter lumen. Tunneled catheters have a cuff that creates a barrier between skin flora at the exit site and the sterile catheter tunnel leading into the fibrinous sheath covering the catheter from the vessel insertion point to its tip. The fibrinous sheath can extend too far, impeding catheter flow and necessitating exchange of the catheter with disruption of the sheath by balloon angioplasty. Catheter-related bacteremia is best treated with exchange of the catheter and disruption of any fibrinous sheath, although removal of the catheter and delayed reinsertion after several days is also successful. Catheter infection-related sepsis unresponsive to antibiotics requires removal of the catheter. Thrombotic occlusion and later sclerotic scarring of the vein at catheter insertion sites is common, however, and removal of a catheter may lead to loss of this access site. Catheter vessel wall contact points are thought to lead to central vein stenosis, which is more commonly observed in patients with catheter contact times of longer duration (>3 months). Catheter tip position in the large central veins instead of the right atrium causes additional injury from dynamic blood movement during dialysis treatments and should be corrected. A thrombus is commonly found attached to the catheter, often tethering the catheter to the vessel wall and right atrium. While some thrombi are mobile and dissolve with anticoagulation, a wall-tethered thrombus is often well organized with cellular components and quite resistant to pharmacologic lysis. Clinically significant pulmonary embolism from catheter-associated thrombus is rare, and it may be that only intra-atrial thrombus >2 cm in diameter deserves active intervention.

Arteriovenous Graft and Fistulas During the first decades of hemodialysis for loss of renal function, U.S. patients were relatively young and without long-term systemic vascular disease. Creation of forearm Cimino AVFs was common, and access failure usually led to creation of a second AVF slightly higher up on the arm. As diabetes and hypertension with associated systemic arterial and venous vascular disease became more prevalent in the CKD population, placement of nonautogenous accesses (arteriovenous grafts [AVGs]) increased. In the mid-1990s, 65% of prevalent dialysis patients used an AVG for access. The United States was an international outlier in this regard, and studies associated increased mortality in U.S. dialysis patients with lower AVF prevalence. In the context of "Fistula First" and then "Fistula First, Catheter Last" campaigns, AVG prevalence decreased to its current value of <20%, while AVFs increased to near 65% prevalence. However, most centers still struggle with the challenging conditions of arteries and veins in these patients requiring that 75% of AVFs are now created in the upper arm, where the veins a priori are larger in diameter, and arteries can deliver higher blood flow rates due to large vessel diameter **(see Fig. 324-1)**.

To provide successful dialysis, an AVF or AVG has to provide at least the desired blood pump speed **(Chap. 323)** plus 100–200 mL/min to minimize recirculation and prevent collapse of the access. In the United States, this usually means flow in the 600–800 mL/min range. A fistula created using an artery with ≥2 mm diameter and a vein ≥3 mm will typically have flow volumes >500 mL/min when the systolic blood pressure is >100 mmHg. After creation of the arterial-venous anastomosis (or insertion of the AVG), blood flow increases significantly: brachial artery flow at rest is typically <50 mL/min, but after access creation, flow volume in AVFs increases within weeks to 800 mL/min, while flow volume in AVGs increases within minutes to 1000 mL/min. The increased flow changes the arterial shear stress profile and leads to enlargement of the artery over time. In AVGs, this process is limited by the graft itself, which typically is 6 mm in diameter and 35–40 cm long, and access flows remain in the 1200–1800 mL/min range. The access vein in AVFs in the right shear stress environment enlarges over time, often to >10 mm in diameter in the upper arm such that the artery continues to enlarge until a narrow segment in the venous conduit becomes flow limiting. Flow volumes in these mature upper arm AVFs are usually 1400–1800 mL/min, but after a few years can be as high as 2000–4000 mL/min. Forearm AVFs usually have lower flow volumes (500–700 mL/min) as the feeding radial artery is of smaller diameter and, in the context of systemic vascular disease in the United States, only increases in diameter over many years.

Increased flows and pressure in the venous segment of the access circuit combine to lead to "chronic dialysis access disease" that manifests differently for each type of the common long-term accesses in predetermined segments particularly prone to shear stress and needle insertion–related injury. AVGs develop venous anastomotic stenoses that recur with very short periodicity in the 3- to 4-month range. Stent grafts can effectively be deployed to extend patency for usually 1 year at the site, after which the buildup of pauci-cellular fibrous depositions at the stent edges requires re-angioplasty one to three times per year. Forearm radial-cephalic autogenous accesses of the Cimino type at the wrist are most prone to low flow due to juxta-anastomotic stenoses. Over time, these stenoses can stabilize, and with enlargement of the inflow artery, they effectively provide protection against excessive flows and their sequelae. "Snuff box" radial-cephalic fistulas require minimal mobilization of the cephalic vein with fewer juxta-anastomotic stenoses. However, the additional side branches and associated venous valves may present as stenosis and require ligation to allow cannulation (see below discussion of augmentation). Upper arm brachial-cephalic autogenous accesses typically develop stenoses in the cephalic arch, which recur in accelerated fashion after each angioplasty. Flexible stent grafts in the cephalic arch extend intraprocedural intervals usually to 9–12 months. Upper arm transposed brachial-basilic autogenous accesses develop stenoses in the swing point where the basilic vein is curved during a mobilization procedure to provide a location more lateral and closer to the skin to facilitate needle cannulation. Angioplasty and stent graft placement approaches extend patency. In both types of upper arm accesses, there are often prolonged periods with increased intra-access pressures due to outflow stenoses, which lead to enlargement of needle insertion site aneurysm as the skin heals in a pressurized, stretched state. Continued use of pressurized accesses leads to enlargement of needle sites, then thinning of the skin, scab formation, and, finally, full-thickness ulceration with often significant bleeding events. Recognizing the occurrence of outflow stenoses early

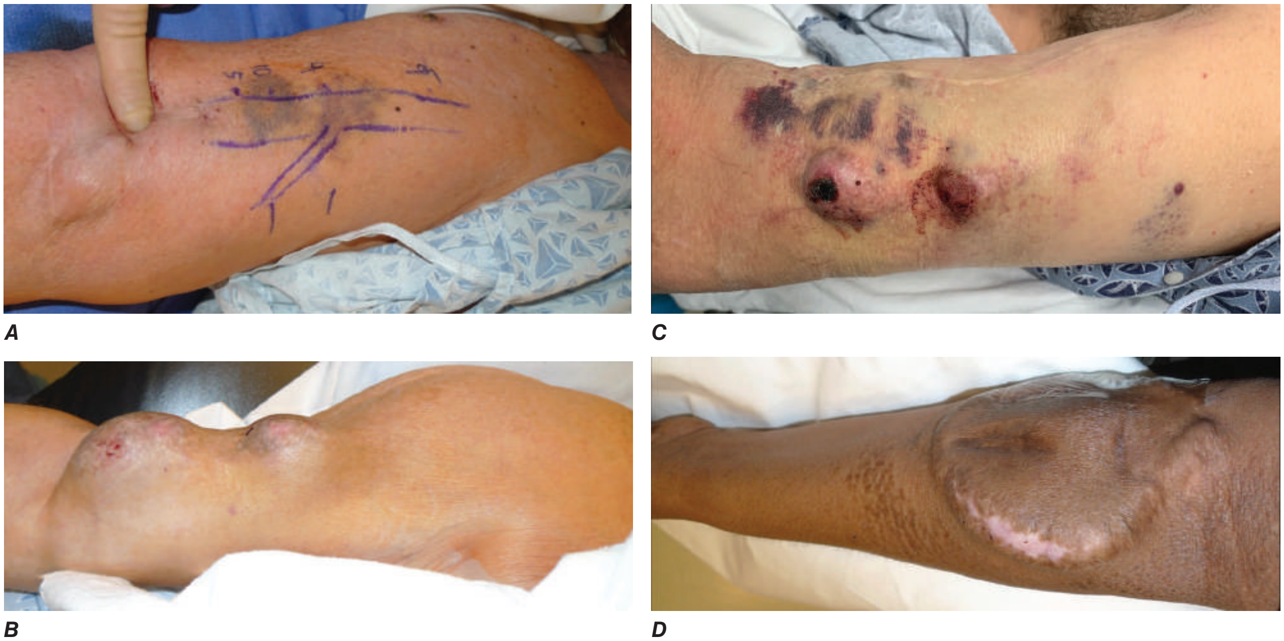

FIGURE 324-1 Dialysis access health depends on intra-access pressures and needle insertions. ***A.*** A right upper arm brachial-cephalic arteriovenous fistula (AVF) with two recurrences of clinically relevant inflow stenosis in 4 years has low-normal intra-access pressure before and after angioplasty; there is only minimal needle insertion site enlargement. ***B.*** In contrast, a right upper arm brachial-cephalic AVF with seven recurrences of cephalic arch outflow stenosis in 4-year cycles between states of high-normal to high intra-access pressures with notable needle insertion site enlargement. ***C.*** Focal needle insertions despite available graft segments led to penetrating skin ulcers over 3 years. ***D.*** Segmental needle rotation preserves skin integrity even after 7 years of arteriovenous graft (AVG) use.

is an important skill for nurses and technologists working in dialysis units to acquire in order to avoid irreversible loss of skin coverage and possible loss of the access.

High-access flow (>1500 mL/min) can lead to systemic complications, such as heart failure and pulmonary hypertension. Fistula inflow higher than outflow capacity leads to accelerated aneurysm formation and breakdown of skin coverage as intra-access pressures are increased over the ideal pressure of 20–35 mmHg. High-access flows are also associated with steal syndrome, typically ischemia of the hand. A variety of procedures have been described to reduce access flows, the most common being “banding,” where typically a 2-0 Prolene suture is guided around the inflow and a 3- or 4-mm spacer and is tied snugly over the spacer to create an inflow stenosis.

APPROACH TO THE PATIENT

Physical Examination of Dialysis Access

The 2019 Kidney Disease Outcomes Quality Initiative (KDOQI) vascular access guidelines were developed under the tenet, “[t]he right access for the right patient at the right time.” Progression of CKD is highly variable, many patients die from other causes before reaching end-stage renal disease (ESRD), and some AVFs require 6–12 months to mature to usability in patients with hypertension and diabetes, leading to uncertainty as to when to create AVFs. The more common need of upper arm accesses for interventions to maintain patency favors the creation of forearm accesses during the pre-ESRD period. The processes of care from vein mapping, surgery, follow-up visits after access creation, to availability and timing of open surgical or endovascular interventions have profound effects on the overall success rate needed to achieve mature and usable accesses and appear to be key factor in the highly variable outcomes across the United States.

A central skill in dialysis access evaluation is the physical examination. Five steps in the access examination capture all aspects of possible pathology: *Pulsatility* reflects the force of access expansion during systole and the degree of softening during diastole. Very high blood pressures will suggest increased pulsatility, but the access softens remarkably during diastole. An outflow stenosis will lead to increased pulsatility and reduced softening during diastole. An inflow stenosis will blunt the systolic component and create the impression of an “empty” access during diastole unless there is a coexisting outflow stenosis. The audible *flow murmur* can be characterized by pitch and continuity **(Video 324-1)**. A change in pitch toward higher frequency is typical at the site of a stenosis due to accelerated flow velocity at this site. A discontinuous flow murmur indicates that during diastole flow is so low that no audible shear force is created; this is the sign of a severe inflow or outflow stenosis. Typically, the stenotic inflow murmur is faint (like a whistle), whereas the stenotic outflow murmur can be coarse and loud (akin to a handsaw) **(Video 324-2)**. A *thrill* is palpable through the skin when the vessel is close enough to the surface and the flow high enough in relation to the diameter of the vessel to create vibration of the vessel wall. A continuous thrill can be a sign of a well-developed access, usually in the inflow segment, dissipating as the access vessel branches and takes a deeper course. In contrast, a discontinuous thrill is found with severe stenosis. An isolated thrill is also found focally immediately after a stenosis. The differentiation from a “healthy” thrill can be made by documenting a change in pulsatility at the site of the focal thrill, increased retrograde (inflow) and decreased antegrade (outflow). *Augmentation* is the engorgement of the body of the access (where needles are inserted) with occlusion of the outflow necessary for safe and successful needle insertions. An inflow stenosis will impair augmentation, as will side branches and collaterals between the occluding finger/tourniquet and the inflow. The location of side branches can be elucidated by moving the occluding finger closer toward the anastomosis until augmentation is achieved. With several collaterals, this may be a staged phenomenon. *Collapse of the access* with arm elevation (against gravity) is a measure of inflow and outflow capacity match or mismatch. A forearm access typically displays complete collapse, while upper arm accesses typically show only partial collapse. An outflow stenosis or very high inflow will decrease the degree of collapse; banding of an upper arm access or a natural flow-limiting stenosis may lead to complete collapse of an upper arm access.

Enlarged needle insertion sites (and any sites of suspected skin thinning) are best examined while occluding inflow: the completely empty access allows palpation of a firm, layered thrombus inside aneurysms as well as a better appreciation of the thickness of the overlying skin by rolling it between thumb and index finger. The chest wall and neck should be inspected for the presence of skin veins and venous distention, which are associated with central venous stenosis or occlusion, as is ipsilateral arm edema.

PRESERVATION OF VENOUS “REAL ESTATE”

Preserving access is a key care component for the patient with advancing CKD. Approximately 8–10% of this population has the need for cardiac rhythm management devices (CRMDs) that can lead to loss of the upper arm cephalic vein as well as central venous stenoses and occlusions around device leads. Planning for which side a future autogenous access is to be placed and where a CRMD is located is recommended in all cases. CKD patients also have an increased frequency of hospitalizations, some of which require intravenous access beyond the hospital stay for antibiotics, nutritional support, or hydration. Avoiding PICCs in a patient with CKD stage 3 or 3b and higher and, instead, using internal (or external) jugular vein tunneled small-diameter catheters preserves arm veins for long-term access creation. Arterial access points for cardiac procedures should be chosen with AVF creation in mind.

Approach to the dialysis access of patients with a kidney transplant depends on the function of the transplanted kidney, the risk of recurrence of kidney disease in the transplant, the ability to limit access flow over time while maintaining patency, and the potential benefit of the AVF for blood pressure control.

■ FURTHER READING

DePietro DM, Trerotola SO: Choosing the right treatment for the right lesion, Part II: A narrative review of drug-coated balloon angioplasty and its evolving role in dialysis access maintenance. Cardiovasc Diagn Ther 13:233, 2023.

Haskal ZJ, Dolmatch BL: Hemodialysis access stent graft trials: Past, present, and future. Cardiovasc Intervent Radiol 46:1154, 2023.

Hentschel DM: Hemodialysis access maintenance and salvage, in *Mastery of Surgery: Vascular Surgery: Hybrid, Venous, Dialysis Access, Thoracic Outlet, and Lower Extremity Procedures*, Philadelphia, Wolters Kluwer, 2015, pp 191–205.

Hentschel DM et al: Hemodialysis access interventions, in *Vascular Imaging and Intervention*, 2nd ed. D Kim et al (eds). India, Jaypee Brothers Medical Publishers, 2020, pp 1655–1686.

Hoggard J et al: Guidelines for venous access in patients with chronic kidney disease. Semin Dial 21:186, 2008.

Lok CK et al: KDOQI clinical practice guideline for vascular access: 2019 update. Am J Kidney Dis 75:S1, 2020.

Ozaki CK et al: Non-maturing autogenous arteriovenous fistula, in *Vascular Decision Making*. Philadelphia, Wolters Kluwer, 2020.

VIDEO 324-1 Flow murmur of an upper arm brachial-cephalic autogenous access (arteriovenous fistula [AVF]) with a juxta-anastomotic stenosis. The sound is discontinuous as the stenosis is severe enough that only during systole is the flow volume high enough to create audible turbulence. There also is a high-pitch component of the murmur due to the high flow velocity during the peak of the flow cycle.

VIDEO 324-2 Flow murmur of an upper arm brachial-cephalic autogenous access (arteriovenous fistula [AVF]) with a juxta-anastomotic stenosis after angioplasty. The sound is now continuous with systolic-diastolic modulation. There is an even pitch, overall lower than the pitch associated with peak flow in the setting of an untreated stenosis.

325 Kidney Transplantation

Jamil R. Azzi, Naoka Murakami,
Anil K. Chandraker

Kidney transplantation is the treatment of choice for patients with end-stage kidney disease (ESKD). Worldwide, tens of thousands of kidney transplants have been performed, and >220,000 patients are living with a functioning kidney transplant in the United States today. The first successful kidney transplant was performed in Boston in 1954 between identical twins. The introduction of immunosuppressive therapies such as azathioprine and prednisone in the 1960s established kidney transplantation across nonidentical individuals (allografts). During the 1970s and 1980s, the success rate at the 1-year mark for deceased-donor allografts markedly improved after the introduction of calcineurin inhibitors. Currently, 1-year survival rates for living-donor and deceased-donor allografts are 98 and 93%, respectively, in the United States. However, long-term survival has not improved as much over time, and the average allograft survival time is 19 and 12 years for living-donor and deceased-donor grafts, respectively.

Age-related mortality rates after transplantation are highest in the first year due to the surgical risks: 2% for ages 18–34 years, 3% for ages 35–49 years, and 6.8% for ages ≥50–60 years. Despite this, the actual survival benefit of transplantation compared to chronic dialysis becomes apparent within days to months following transplantation, even after risk adjustments for age, diabetes, and cardiovascular status. While the loss of kidney transplant due to acute rejection is now a rare event, most allografts eventually succumb at varying rates to a chronic process consisting of interstitial fibrosis, tubular atrophy, vasculopathy, and glomerulopathy, the pathogenesis of which in varying degrees is likely a combination of an alloimmune response, drug toxicity, and the result of a variety of other insults. Overall, transplantation results in an improved life expectancy with a higher quality of life compared to patients who remain on dialysis.

RECENT ACTIVITY AND RESULTS

In 2022, >19,000 deceased-donor kidney transplants and approximately 6000 living-donor transplants were performed in the United States, with the ratio of deceased-donor to living-donor transplants remaining stable over the past few years. As the number of patients with ESKD increases, the number of patients on the transplant waitlist also increases. The donor shortage remains a critical challenge; as of 2022, there were nearly 139,000 adult kidney transplant candidates on the waiting list, with 25,000 patients being transplanted yearly. This imbalance is set to worsen over the coming years with the predicted increased rates of kidney failure associated with obesity and diabetes worldwide. In an attempt to increase utilization of marginal kidneys and allocate organs equitably, a new allocation system within the United States was implemented in 2014. The guiding principles of the changes were to offer an opportunity for transplantation to patients who were highly sensitized and, thus, less likely to find a suitable donor, while at the same time allowing patients expected to survive the longest to receive the best-quality deceased donor organs. The Kidney Donor Profile Index (KDPI) score, which ranges from 0 to 100%, was introduced to estimate the potential risk of graft failure after kidney transplant based on 10 donor factors. The lower KDPI values are associated with higher expected posttransplant survival. Hence, the kidneys with a KDPI <20% are allocated to the 20% of the potential recipients with the highest expected posttransplant survival. Kidneys with a KDPI >85% (previously called expanded criteria donor [ECD] kidneys) are directed toward patients who are expected to fare less well on dialysis and would benefit from being transplanted earlier even if it means accepting a lower-quality organ.

In 2021, a new distance-based kidney allocation policy was introduced. Under this policy, deceased-donor kidneys are to be offered first to candidates listed at transplant hospitals within 250 nautical miles of the donor hospitals. While intended to reduce geographic disparities in access to kidney transplantation, this program has resulted in more complex organ sharing, unintended higher kidney discard rates, and increased cold ischemia time. A variety of other means to increase the donor pool and equity have also become more popular. Kidneys from donors after cardiac death (DCD) are now commonly used to overcome the demand for organs, consisting of 30% of total deceased kidney transplants (Table 325-1). Furthermore, with the advancement of the direct-acting antiviral therapies for hepatitis C virus (HCV), transplantation from HCV-positive donors to HCV-positive or -negative recipients has been performed since 2017 in order to increase the donor pool. Now this practice is becoming more common, consisting of 9% of deceased kidney transplants. Recently, the HOPE (Human Immunodeficiency Virus [HIV] Organ Policy Equity) Act authorized organ donation from HIV-positive candidates, and >100 transplants have been performed. As patients with blood group B wait longer for deceased donor offers, eligible B blood type candidates who have low anti-A titer are eligible for an allograft from A blood type donors. This helps improve access and reduce disparities in wait time for minorities, especially for the African-American ESKD population, in whom blood type B is more common than in other ethnicities. Finally, with recent advances in gene editing technology, xenotransplantation is becoming a more realistic endeavor. Using kidneys from genetically engineered pigs, three experimental kidney transplantations into brain-dead recipients were performed in 2022. There are still multiple issues to overcome, including xenoimmunity, transmission of zoonosis, and ethical challenges, but xenotransplantation has the potential to provide an unlimited source of organs.

The overall results of transplantation are presented in Table 325-2. At the 1-year mark, allograft survival is higher for living-donor recipients. This is most likely related to less ischemic injury of the transplant organ. The introduction of more effective immunosuppression and more accurate matching between recipients and donors has almost equalized the risk of graft rejection in the majority of patients within the first year. At 5- and 10-year follow-up, however, there remains a steeper decline in survival of those with deceased-donor kidneys.

RECIPIENT EVALUATION

Virtually all patients with ESKD benefit from transplantation with a longer life expectancy and a better quality of life. While the mortality rate after transplantation is highest in the first year due to perioperative complications, recipient evaluation is critical in identifying patients at risk. It involves a multidisciplinary approach that requires thorough medical, surgical, social, and psychosocial evaluations to identify the risk factors that prohibit transplantation or mandate treatment before proceeding, as well as ensuring the appropriate use of limited organs.

There are a few absolute contraindications to kidney transplantation: chronic illness that limits predicted survival for <2 years, active malignancy, active infection, psychosocial issues affecting adherence to the medical care, and active substance abuse. Cardiovascular risk assessment is crucial during both the perioperative and postoperative periods. Patients with ESKD are at higher cardiovascular mortality risk,

TABLE 325-1 Definition of a Non-Heart-Beating Donor (Donation After Cardiac Death[a] [DCD])

I: Brought in dead
II: Unsuccessful resuscitation
III: Awaiting cardiac arrest
IV: Cardiac arrest after brainstem death
V: Cardiac arrest in a hospital patient

[a] Kidneys can be used for transplantation from categories II–V but are commonly only used from categories III and IV. The survival of these kidneys has not been shown to be inferior to that of deceased-donor kidneys.

Note: Kidneys can both have a Kidney Donor Profile Index (KDPI) score >85% and be DCD. High KDPI kidneys have been shown to have a poorer survival, and there is a separate shorter waiting list for those kidneys. They are generally utilized for patients for whom the benefits of being transplanted earlier outweigh the associated risks of using a lower-quality kidney.

TABLE 325-2 Mean Rates of Graft and Patient Survival for Kidneys Transplanted in the United States from 1999 to 2018[a]

	1-YEAR FOLLOW-UP		5-YEAR FOLLOW-UP		10-YEAR FOLLOW-UP	
	GRAFTS, %	PATIENTS, %	GRAFTS, %	PATIENTS, %	GRAFTS, %	PATIENTS, %
Deceased donor	94	95	76	77	54	67
Living donor	98	99	87	87	70	83

[a]All patients transplanted are included, and the follow-up unadjusted survival data from the 1-, 5-, and 10-year periods are presented to show the attrition rates over time within the two types of organ donors.

Source: Data from Summary Tables, 2021 Annual Reports, Scientific Registry of Transplant Recipients.

and thorough cardiovascular evaluation for coronary artery diseases, valvular diseases, and heart failure is critical.

At most centers, there is no official age limit for transplantation, with >20% of waitlisted candidates currently being older than 65. However, overall physical and cognitive function of the candidates needs to be fully assessed. While history of malignancy itself is not a contraindication for kidney transplantation, potential recipients should be treated to ensure cancer-free wait time of 2–5 years depending on the type and stages of malignancy to decrease the risk of cancer recurrence. Latent or indolent infection (HIV, hepatitis B or C, tuberculosis) should be a routine part of the candidate workup. While historically transplant centers considered overt AIDS and active hepatitis absolute contraindications to transplantation because of the high risk of opportunistic infection, with the introduction of potent antiviral regimens, many centers are now transplanting individuals with hepatitis and HIV infection under strict protocols.

One of the few "immunologic" contraindications to transplantation is the presence of preformed antibodies against the donor kidney at the time of the anticipated transplant that can cause hyperacute rejection. Those harmful antibodies include natural antibodies against the ABO blood group antigens and antibodies against human leukocyte antigen (HLA) class I (A, B, C) or class II (DR, DQ, DP) antigens. These antibodies are routinely excluded by proper screening of the candidate's ABO compatibility and direct cytotoxic cross-matching of candidate serum with lymphocytes of the donor. Removal of these antibodies directed at donor tissue through a variety of strategies (desensitization) is now routinely performed with varying levels of success.

TISSUE TYPING AND CLINICAL IMMUNOGENETICS

Matching of HLA major histocompatibility complex antigens (Chap. 361) is an important criterion for the selection of donors. Each mammalian species has a single chromosomal region that encodes the major histocompatibility antigens, and this region on the human chromosome 6 codes the HLA genes. HLA is highly polymorphic; therefore, it can be an immunologic target of organ rejection when mismatched between the donor and the recipient. Historically, HLA antigens have been defined by serologic techniques by adding sera of a recipient (potentially containing anti-HLA antibodies) with a "library" of leukocytes with known serotypes. However, currently, molecular typing of HLA by genomic sequencing is almost universally used. Other "minor," non-HLA antigens may also elicit an alloimmune response in addition to the ABH(O) antigens and endothelial antigens that are not expressed on lymphocytes. The number of HLA antigen mismatches in A, B, and DR loci correlates with allograft survival; the more mismatches, the higher is the risk of allograft rejection. Nevertheless, some HLA-identical renal allografts are rejected, often within the first few weeks after transplantation. These observations may represent prior sensitization to non-HLA antigens. Non-HLA minor antigens are relatively weak when initially encountered and are, therefore, suppressible by conventional immunosuppressive therapy. If prior exposure to the antigen and priming of the recipient immune system have occurred, secondary exposure at the time of transplantation may lead to an immune response refractory to treatment. More recently, a genetics study of donor-recipient pairs revealed that non-HLA antigen polymorphisms in the LIMS-1 locus can contribute to the risk of acute rejection; more of these non-HLA antigens are likely to be discovered.

DONOR EVALUATION

LIVING-DONOR EVALUATION

Living kidney donors experience the immediate risk of surgery and the long-term potential risk of developing kidney dysfunction prematurely; thus, the basic principle of "first, do no harm" (Chap. 12) is important. Therefore, donor evaluation must take every effort to exclude any medical conditions that may cause morbidity and mortality after kidney donation, such as hypertension, diabetes, and/or proteinuria. Although studies have shown that the risk of ESKD after kidney donation is not greater than that of the general population, donation is associated with a small but significant potential lifetime risk of ESKD (0.3–0.4%; absolute risk increased by 0.2–0.3% compared to that of healthy nondonors). The mechanism of premature renal failure is thought to be due to increased blood flow and hyperfiltration injury in the remaining kidney. There are a few reports of the development of hypertension, proteinuria, and even lesions of focal segmental sclerosis in donors over long-term follow-up. In family members of type 1 diabetics, anti-insulin and anti-islet cell antibodies should be measured, and a glucose tolerance test should be performed. African-American donors have a higher risk of ESKD after donation (in line with their higher risk of kidney failure in general), and the genetic screening for APOL1 risk alleles may be appropriate (Chap. 326). Additionally, as more robust genetic testing becomes available, predonation genetic testing per protocol is being introduced to stratify the risks of kidney donors. From the surgical perspective, selective renal arteriography is essential to reveal any anatomic anomaly and to assess the size imbalance and laterality of donor kidneys. In most cases, donor nephrectomy is performed laparoscopically to minimize the surgical scar and to enhance a faster postsurgical recovery. Lastly, although financial and nonfinancial conflicts of interest between kidney donors and recipients are strictly prohibited, removing financial disincentives is increasingly accepted to reduce barriers toward living donation, and legislative efforts to protect kidney donors are ongoing (Chap. 12).

DECEASED-DONOR EVALUATION

Deceased donors should be free of malignant neoplastic disease, hepatitis, and HIV owing to possible transmission to the recipient, although under certain circumstances, HCV- and HIV-positive organs may be used. Increased risk of graft failure exists when the donor is elderly or has acute kidney injury or when the kidney experiences a prolonged period of ischemia.

In the United States, there is a national system of regulations, allocation support, and outcomes analysis for kidney transplantation called the Organ Procurement Transplant Network. Studies have shown that deceased-donor kidneys can be maintained for up to 48 h on cold pulsatile perfusion or on ice before being used for transplantation. Normothermic perfusion of donated organs has been studied, but it has not been part of clinical care as of yet. Generally, an ischemic time of <24 h is preferred; this approach permits adequate time for typing, cross-matching, transportation, and selection issues to be resolved.

PRESENSITIZATION

The presence of antibodies against donor antigens, either HLA or non-HLA, can be a potential cause of allograft injury after transplantation, and, hence, it is important to perform crossmatching prior to transplantation. For the purposes of crossmatching, donor T lymphocytes, which express class I but not class II HLA, are used as a surrogate target for detection of circulating anti–class I (HLA-A and -B) antibodies in the recipient. Note that T cells are used as surrogate cells to detect class I HLA as a matter of convenience and this is unrelated to the risk of "T cell–mediated" rejection. A positive cytotoxic crossmatch of recipient serum with donor T lymphocytes indicates the presence of preformed donor-specific anti-HLA class I antibodies and is usually predictive of an acute vasculitic event termed *hyperacute rejection*. This finding represents the only widely accepted absolute immunologic contraindication for kidney transplantation. Recently, an increasing number of tissue-typing laboratories have shifted to a more sensitive flow cytometric crossmatch assay, which detects the presence of anti-HLA IgG antibodies that are not necessarily detected on a cytotoxic crossmatch assay and may not be an absolute contraindication to transplantation. The known sources of sensitization are blood transfusion, a prior transplant, pregnancy, and less commonly, vaccination or infection.

Preformed anti–class II (HLA-DR and -DQ) antibodies against the donor also carry a higher risk of graft loss, particularly in recipients who have suffered early loss of a prior kidney transplant. B lymphocytes (again, used for convenience), which express both class I and class II HLA, are used as a surrogate target in these assays. Some non-HLA antigens restricted in expression to endothelium and monocytes have been described, but their clinical relevance is not well established. A series of minor histocompatibility antigens do not elicit antibodies, and sensitization to these antigens is detectable only by cytotoxic T cells, an assay too cumbersome for routine use. Recent studies revealed the importance of "eplet" matching in antigen recognition. Eplets are short sequences of polymorphic amino acids on the surface of HLA antigens, recognized by HLA antibodies, and can be shared among different HLA antigens. Especially in class II HLA DQ loci, eplet mismatches are shown to be important risk factors for acute rejection.

Desensitization prior to transplantation by reducing the level of anti-donor antibodies utilizing plasmapheresis and/or the administration of pooled immunoglobulin (IV immunoglobulin [IVIG]) has been useful in reducing the risk of hyperacute rejection following transplantation. In addition, kidney paired donation programs where living donor kidneys are swapped so each recipient receives a compatible transplant organ are increasingly popular to transplant presensitized candidates safely.

IMMUNOLOGY OF REJECTION

Both T cell–mediated and antibody-mediated effector mechanisms can play roles in kidney transplant rejection.

T cell–mediated rejection is caused by recipient T lymphocytes that respond to donor HLA antigens expressed within the transplanted organ. CD4+ lymphocytes respond to class II (HLA-DR) incompatibility by proliferating and releasing proinflammatory cytokines that augment the proliferative response of the immune system. CD8+ cytotoxic lymphocytes respond primarily to class I (HLA-A, -B) antigens and mature into cytotoxic effector cells that cause organ damage through direct contact and lysis of donor target cells. Full T-cell activation requires not only T-cell receptor binding to the alloantigens presented by self or donor HLA molecules (known as indirect and direct presentation, respectively) but also engagement of costimulatory molecules such as CD28 on T cells and CD80 and CD86 ligands on antigen-presenting cells **(Fig. 325-1)**. Signaling through both of these pathways induces activation of the kinase activity of calcineurin, which, in turn, activates transcription factors leading to upregulation of multiple genes, including interleukin (IL) 2 and interferon γ. IL-2 signals through the target of rapamycin (TOR) to induce cell proliferation in an autocrine fashion.

Antibody-mediated rejection is caused by circulating antibodies against donor antigens. After transplantation, donor-derived antigens are delivered to the recipient's draining lymph nodes and activate an alloimmune response. A subset of CD4+ T cells called follicular helper T cells (Tfh) are activated and promote differentiation of B cells into antibody-secreting plasma cells. Plasma cells produce donor-targeting antibodies against HLA and non-HLA antigens, which can deposit in allograft kidney and cause injury via complement-dependent and independent mechanisms. C4d deposition in peritubular capillaries and glomerular basement membrane is a footprint of complement activation and is one of the diagnostic criteria of antibody-mediated rejection, together with the presence of circulating donor-specific antibody.

IMMUNOSUPPRESSIVE TREATMENT

Kidney transplant recipients need to take immunosuppressive drugs for life, except identical twins or simultaneous bone marrow–kidney transplant recipients. Currently clinically available immunosuppressive therapies suppress all immune responses nonspecifically, including those to exogenous pathogens (bacteria, viruses, and fungi) and even malignant tumors, and tend to spare memory immune responses. Immunosuppressive agents are divided into induction and maintenance agents. Those currently in clinical use are listed in **Table 325-3**.

INDUCTION THERAPY

Induction therapy is given to most kidney transplant recipients in the United States at the time of transplant to reduce the risk of early acute rejection and to minimize or eliminate the use of either steroids or

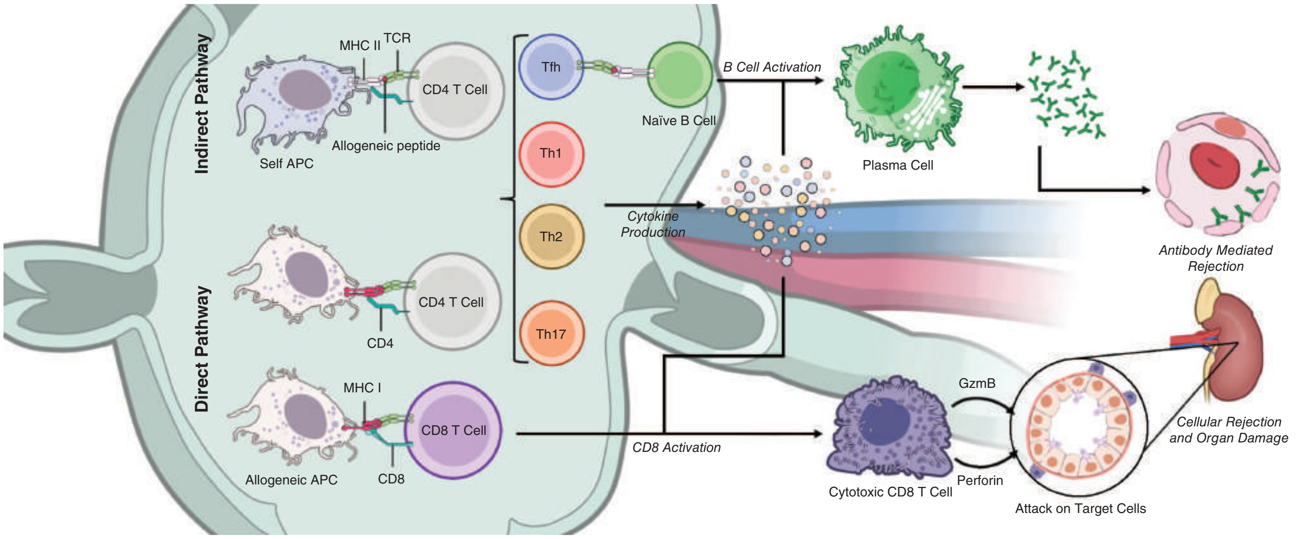

FIGURE 325-1 Recognition pathways for major histocompatibility complex (MHC) antigens. Graft rejection is initiated by CD4 helper T lymphocytes (T_H) having antigen receptors that bind to specific complexes of peptides and MHC class II molecules on antigen-presenting cells (APC). In transplantation, in contrast to other immunologic responses, there are two sets of T-cell clones involved in rejection. In the direct pathway, the class II MHC of donor allogeneic APCs is recognized by CD4 T_H cells that bind to the intact MHC molecule, and class I MHC allogeneic cells are recognized by CD8 T cells. The latter generally proliferate into cytotoxic cells (T_C). In the indirect pathway, the incompatible MHC molecules are processed into peptides that are presented by the self-APCs of the recipient. The indirect, but not the direct, pathway is the normal physiologic process in T-cell recognition of foreign antigens. Once T_H cells are activated, they proliferate and, by secretion of cytokines and direct contact, exert strong helper effects on macrophages, TC, and B cells. *(Courtesy of Andrew Badoui and Nadim Al Rahy.)*

TABLE 325-3 Maintenance Immunosuppressive Drugs

AGENT	PHARMACOLOGY	MECHANISMS	SIDE EFFECTS
Glucocorticoids	Increased bioavailability with hypoalbuminemia and liver disease; prednisone/prednisolone generally used	Binds cytosolic receptors and heat shock proteins. Blocks transcription of IL-1, -2, -3, -6, TNF-α, and IFN-γ	Hypertension, glucose intolerance, dyslipidemia, osteoporosis
Cyclosporine (CsA)	Lipid-soluble polypeptide, variable absorption, microemulsion more predictable	Trimolecular complex with cyclophilin and calcineurin → block in cytokine (e.g., IL-2) production; however, stimulates TGF-β production	Nephrotoxicity, hypertension, dyslipidemia, glucose intolerance, hirsutism/hyperplasia of gums
Tacrolimus	Macrolide, well absorbed	Trimolecular complex with FKBP-12 and calcineurin → block in cytokine (e.g., IL-2) production; may stimulate TGF-β production	Similar to CsA, but hirsutism/hyperplasia of gums unusual, and diabetes more likely
Azathioprine	Mercaptopurine prodrug	Hepatic metabolites inhibit purine synthesis	Marrow suppression (WBC > RBC > platelets)
Mycophenolate mofetil/sodium	Metabolized to mycophenolic acid	Inhibits purine synthesis via inosine monophosphate dehydrogenase	Diarrhea/cramps; dose-related liver and marrow suppression is uncommon
Sirolimus/everolimus	Macrolide, poor oral bioavailability	Complexes with FKBP-12 and then blocks p70 S6 kinase in the IL-2 receptor pathway for proliferation	Hyperlipidemia, thrombocytopenia
Belatacept	Fusion protein, intravenous injections	Binds CD80 and CD86, prevents CD28 binding and T-cell activation	Posttransplant lymphoproliferative disease (PTLD)

Abbreviations: FKBP-12, FK506 binding protein 12; IFN, interferon; IL, interleukin; RBC, red blood cells; TGF, transforming growth factor; TNF, tumor necrosis factor; WBC, white blood cells.

calcineurin inhibitors and their associated toxicities. Induction therapy consists of antibodies that could be depleting or nondepleting.

Depleting Agents Antithymocyte globulin (ATG) is a lymphocyte-depleting agent. Peripheral human lymphocytes, thymocytes, or lymphocytes from spleens or thoracic duct fistulas are injected into horses or rabbits to produce antilymphocyte serum, from which the immunoglobulin fraction is then separated. Those polyclonal antibodies induce lymphocyte depletion, and the immune system may take several months, if not years, to fully recover.

Monoclonal antibodies against defined lymphocyte subsets offer a more precise and standardized form of therapy. Alemtuzumab is directed against CD52, widely expressed on immune cells such as B and T cells, natural killer cells, macrophages, and some granulocytes.

Nondepleting Agents Another more selective approach is to target the 55-kDa alpha chain of the IL-2 receptor, which is expressed only on activated T cells. This approach is used as prophylaxis for (but not treatment of) acute rejection in the immediate posttransplant period and is effective at decreasing the early acute rejection rate with few adverse side effects.

MAINTENANCE THERAPY

The most frequently used combination is a calcineurin inhibitor (CNI), usually tacrolimus, and an antimetabolite, usually mycophenolic acid, with or without early steroid withdrawal. Belatacept is a co-stimulatory bocking antibody, used as an alternative to long-term toxic CNI therapy. The mammalian TOR (mTOR) inhibitors sirolimus and everolimus are infrequently used as first-line maintenance immunosuppression.

Antimetabolites *Azathioprine* is a prodrug that must first be activated to form thioguanine nucleotides. Thiopurine *S*-methyltransferase (TPMT) inactivates azathioprine. Patients with two nonfunctional TPMT alleles experience life-threatening myelosuppression when treated with azathioprine, and those who carry one nonfunctional TPMT allele may also have significant side effects; therefore, the U.S. Food and Drug Administration (FDA) recommends TPMT genotyping or phenotyping before starting treatment with azathioprine. Azathioprine, which inhibits synthesis of DNA and RNA and thereby inhibits T-cell proliferation, was the keystone of immunosuppressive therapy in kidney transplant recipients until the 1990s but has been replaced by more effective agents. Concomitant use of allopurinol is best avoided and, if used, very carefully monitored, owing to inhibition of xanthine oxidase.

Mycophenolate mofetil and *mycophenolate sodium*, both of which are metabolized to mycophenolic acid, are now used in place of azathioprine based on superior efficacy. Mycophenolic acid has a similar mode of action as azathioprine and is associated with a mild degree of gastrointestinal toxicity but less bone marrow suppression.

Steroids *Glucocorticoids* are important adjuncts to immunosuppressive therapy and used as both induction and maintenance therapy. In general, methylprednisolone 250–500 mg is given immediately before or at the time of transplantation, and the dose is tapered to 20 mg within a week. The side effects of the glucocorticoids, particularly impairment of wound healing and predisposition to infection, make it desirable to taper the dose as rapidly as possible in the immediate postoperative period. Early discontinuation or avoidance of steroids is common to avoid long-term adverse effects on bone, skin, and glucose metabolism. Most patients whose renal function is stable after 6 months or a year do not require large doses of prednisone; maintenance doses of 5–10 mg per day are the rule. A major effect of steroids is preventing the release of IL-6 and IL-1 by monocytes-macrophages.

Calcineurin Inhibitors *Cyclosporine* is a fungal peptide with potent immunosuppressive activity. It acts on the calcineurin pathway to inhibit transcription of IL-2 and other proinflammatory cytokines, thereby inhibiting T-cell proliferation. It works synergistically with glucocorticoids and mycophenolate. Among its toxic effects (nephrotoxicity, hepatotoxicity, hirsutism, tremor, gingival hyperplasia, and diabetes), nephrotoxicity presents a serious management problem and is further discussed below.

Tacrolimus (FK506) is a fungal macrolide that has the same mode of action as cyclosporine as well as a similar side effect profile; it does not, however, produce hirsutism or gingival hyperplasia; in contrast, it can be associated with hair loss. Posttransplant diabetes mellitus more commonly occurs with tacrolimus. An extended-release formulation of tacrolimus is now available and is given once daily. Owing to its nephrotoxicity and narrow therapeutic window, the drug level of CNIs should be monitored, and drug–drug interactions should be carefully examined. Antibiotics and antifungals (e.g., erythromycin, ketoconazole, fluconazole) and nondihydropyridine calcium channel blockers (e.g., diltiazem, verapamil) inhibit the activity of cytochrome P450 C3A enzyme and cause elevated levels of CNIs. On the other hand, antiepileptics, such as phenytoin and carbamazepine, increase metabolism, resulting in lower levels.

mTOR Inhibitors *Sirolimus* (previously called rapamycin) is another fungal macrolide but has a different mode of action from tacrolimus; i.e., it inhibits T-cell growth factor signaling pathways, preventing the response to IL-2 and other cytokines. Sirolimus can be used in conjunction with cyclosporine or tacrolimus, or with mycophenolic acid, to avoid the use of CNIs.

Everolimus is another mTOR inhibitor with similar mechanism of action as *sirolimus* but with better bioavailability. mTOR inhibitors are modestly tolerated and are associated with gastrointestinal disturbance, stomatitis, mucositis, and pneumonitis. Poor wound healing associated with mTOR inhibitors makes them less preferable agents during the perisurgical period. While the PI3K-mTOR is the most commonly mutated cellular pathway in malignant cells, mTOR inhibitors have been used more frequently in transplant patients who develop cancers, in particular recurrent skin cancers.

Belatacept *Belatacept* is a fusion protein composed of the Fc fragment of human IgG1 immunoglobulin and the extracellular domain of cytotoxic T-lymphocyte associated protein 4 (CTLA-4). It binds to its costimulatory ligands (CD80 and CD86) on antigen-presenting cells, interrupting their binding to CD28 on T cells. This inhibition leads to T-cell anergy and apoptosis. Belatacept is FDA approved for kidney transplant recipients and is given monthly as an intravenous infusion. The 7-year follow-up of the Belatacept Evaluation of Nephroprotection and Efficacy as First-Line Immunosuppression Trial (BENEFIT) showed improved patient and graft survival for the belatacept-treated group compared to patients treated with cyclosporine, despite short-term risks of higher rates of acute rejection.

CLINICAL COURSE AND MANAGEMENT OF THE RECIPIENT

Adequate hemodialysis should be performed within 48 h prior to the surgery as needed to control serum potassium to prevent cardiac arrhythmias. During the transplantation surgery, the kidney allograft is usually placed in the recipient's iliac fossa using a retroperitoneal approach. An anastomosis is made between donor renal artery and recipient external iliac artery and donor renal vein to recipient external iliac vein. The donor ureter is anastomosed to the recipient bladder mucosa. Native kidney nephrectomy is rarely performed except in the case of an extremely enlarged polycystic kidney or chronic pyelonephritis. In many cases, especially after living kidney transplantation, the allograft starts to produce urine immediately after anastomosis. The allograft often has some degree of acute tubular injury due to ischemia, which accounts for postoperative diuresis. Large amounts of sodium, potassium, and water may be lost postoperatively, which requires close monitoring and replacement. The recipient's serum creatinine should start to fall as the allograft starts to function, and recovery usually occurs within 2 weeks, although periods as long as 6 weeks have been reported. Slow recovery or oliguria should prompt an allograft biopsy because superimposition of rejection on acute tubular injury is common and difficult to distinguish without an allograft biopsy. Induction immunosuppression therapy and maintenance steroids and antimetabolites start on the day of surgery, and it is usually safe to delay introduction of a CNI for a few days if a lymphocyte-depleting induction agent is used. Figure 325-2 illustrates a typical algorithm followed by transplant centers for early posttransplant management of recipients at high or low risk of early renal dysfunction.

■ MANAGEMENT OF REJECTION

Early diagnosis of rejection allows prompt institution of therapy to preserve allograft function and prevent irreversible damage. Clinical

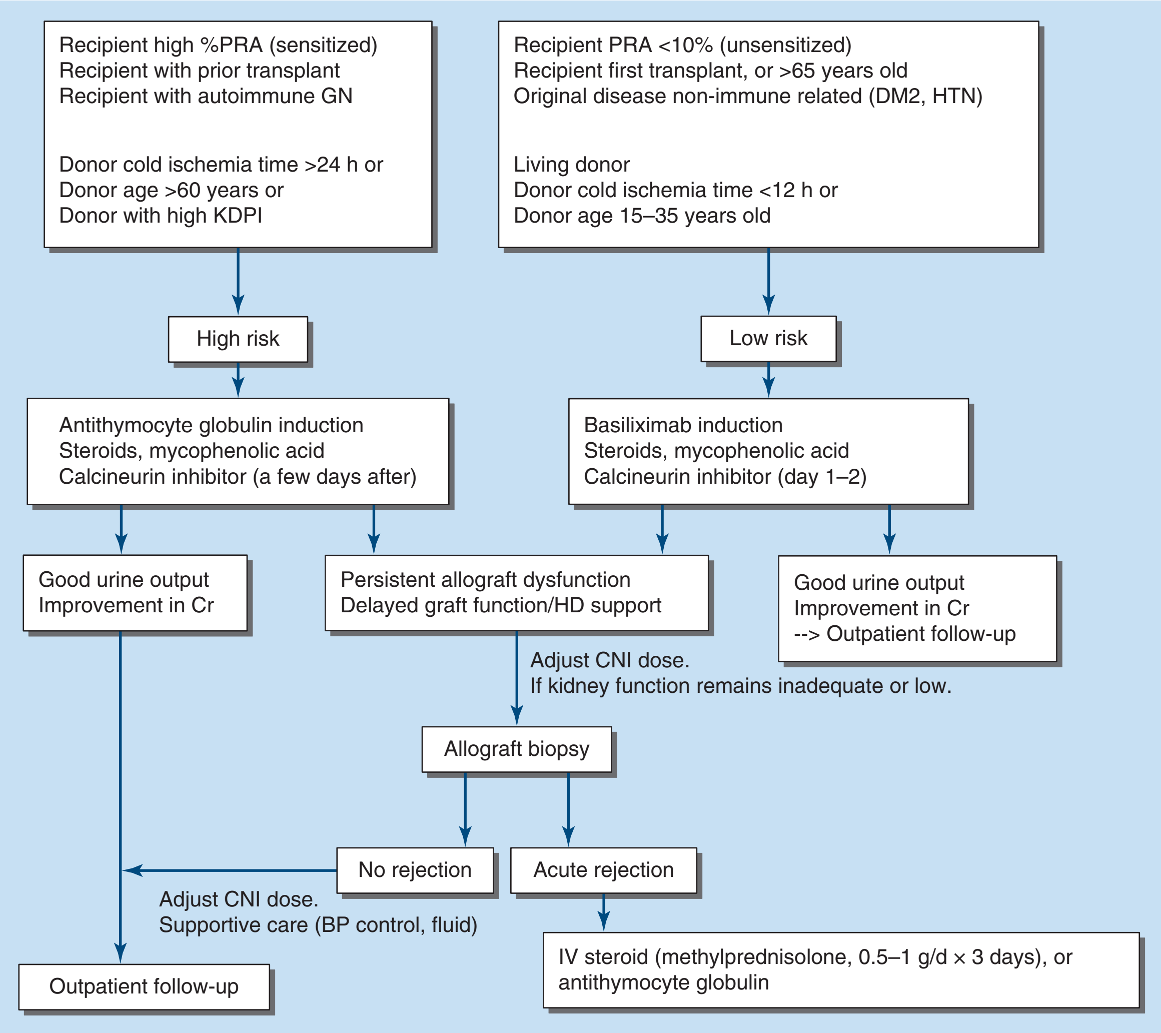
Recipient high %PRA (sensitized)
Recipient with prior transplant
Recipient with autoimmune GN
Donor cold ischemia time >24 h or
Donor age >60 years or
Donor with high KDPI
Recipient PRA <10% (unsensitized)
Recipient first transplant, or >65 years old
Original disease non-immune related (DM2, HTN)
Living donor
Donor cold ischemia time <12 h or
Donor age 15–35 years old
High risk
Low risk
Antithymocyte globulin induction
Steroids, mycophenolic acid
Calcineurin inhibitor (a few days after)
Basiliximab induction
Steroids, mycophenolic acid
Calcineurin inhibitor (day 1–2)
Good urine output
Improvement in Cr
Persistent allograft dysfunction
Delayed graft function/HD support
Good urine output
Improvement in Cr
--> Outpatient follow-up
Adjust CNI dose.
If kidney function remains inadequate or low.
Allograft biopsy
No rejection
Acute rejection
Adjust CNI dose.
Supportive care (BP control, fluid)
Outpatient follow-up
IV steroid (methylprednisolone, 0.5–1 g/d × 3 days), or
antithymocyte globulin

FIGURE 325-2 A typical algorithm for early posttransplant care of a kidney recipient. If any of the recipient or donor "high-risk" factors exist, more aggressive management is called for. Low-risk patients can be treated with a standard immunosuppressive regimen with no or less-potent induction therapy (e.g., basiliximab). Patients at higher risk of rejection or early ischemic transplant dysfunction are often induced with an antithymocyte globulin to provide more potent early immunosuppression or to spare calcineurin use in the immediate posttransplant period. When there is early transplant dysfunction, prerenal, obstructive, and vascular causes must be ruled out by ultrasonographic examination. The panel reactive antibody (PRA) is a quantitation of how much antibody is present in a candidate against a panel of cells representing the distribution of antigens in the donor pool. BP, blood pressure; CNI, calcineurin inhibitor; Cr, creatinine; DM2, type 2 diabetes; GN, glomerulonephritis; HD, hemodialysis; HTN, hypertension; KDPI, Kidney Donor Profile Index.

evidence of rejection is rarely characterized by fever, swelling, and tenderness over the allograft. Rejection may present only with a rise in serum creatinine, with or without a reduction in urine volume. The focus should be on ruling out other causes of functional deterioration, such as acute tubular injury, calcineurin toxicity, BK nephropathy, and recurrent glomerular diseases.

Doppler ultrasonography is useful in ascertaining changes in the allograft vasculature and in blood flow. Thrombosis of the renal vein occurs rarely; it may be reversible if it is caused by technical factors and intervention is prompt. Diagnostic ultrasound is also helpful in identifying urinary obstruction or the presence of perirenal collections of urine (urinoma), blood (hematoma), or lymph (lymphocele).

Allograft biopsy is the gold standard for diagnosis of acute T cell–mediated and antibody-mediated rejection. Acute T cell–mediated rejection is diagnosed by the presence of immune cell infiltration in the interstitial, tubular, or vascular compartments, according to the Banff classification. Treatment of T cell–mediated rejection involves a high-dose steroid, e.g., IV administration of methylprednisolone, 500–1000 mg daily for 3 days. Failure to respond is an indication for antibody therapy, usually with ATG.

Evidence of antibody-mediated rejection is present when endothelial injury and deposition of complement component C4d is detected in peritubular capillaries. This is usually accompanied by detection of the circulating donor-specific antibody in the recipient's blood. Treatment of antibody-mediated rejection remains a challenge, and aggressive use of plasmapheresis, IVIG, anti-CD20 monoclonal antibody (rituximab) to target B lymphocytes, and bortezomib to target antibody-producing plasma cells is indicated. Recently, noninvasive biomarkers such as circulating donor-derived cell-free DNA, urine chemokine markers (e.g., CXCL9), and characterization of the urine exosome have been used as adjunct diagnostic markers for rejection. Future studies to identify prognostic, noninvasive biomarkers that predict response to therapy, that risk stratify, and that provide personalized immunosuppression strategies will be needed.

MANAGEMENT OF CHRONIC COMPLICATIONS

Cardiovascular events (29%), infection (18%), and malignancy (17%) are the major causes of death in kidney transplant recipients. Typical time courses of opportunistic infections after transplantation are shown in **Table 325-4**.

The signs and symptoms of infection may be atypical due to immunosuppression, which makes diagnosis challenging. In addition to commensal infections, opportunistic infections should be considered based on the clinical presentation. Diagnostic measures such as culture (blood, urine, drain fluids), viral load in plasma, and imaging (allograft ultrasound and computed tomography [CT]) should be obtained. Overall therapy involves adequate source control, anti-microorganism therapy, and reduction of immunosuppression.

Pneumocystis jirovecii is a rare but critical opportunistic infection **(Chap. 227)**. Aggressive diagnostic procedures, including transbronchial and open-lung biopsy, are frequently indicated. Trimethoprim-sulfamethoxazole (TMP-SMX) is the treatment of choice; amphotericin B has been used effectively in systemic fungal infections. Prophylaxis against *P. jirovecii* with daily low-dose TMP-SMX for 6 months is effective. Involvement of the oropharynx with *Candida* **(Chap. 222)** may be treated with local nystatin. Tissue-invasive fungal infections require treatment with systemic agents such as fluconazole or one of the newer antifungal agents. *Aspergillus* **(Chap. 223)**, *Nocardia* **(Chap. 179)**, and especially cytomegalovirus (CMV) **(Chap. 200)** infections also occur.

TABLE 325-4 The Most Common Opportunistic Infections in Renal Transplant Recipients

Peritransplant (<1 month)	Late (>6 months)
Wound infections	*Aspergillus*
Herpesvirus	*Nocardia*
Oral candidiasis	BK virus (polyoma)
Urinary tract infection	Herpes zoster
Early (1–6 months)	Hepatitis B
Pneumocystis carinii	Hepatitis C
Cytomegalovirus	
Legionella	
Listeria	
Hepatitis B	
Hepatitis C	

CMV infection is a serious complication after kidney transplantation associated with increased morbidity and mortality. While the seronegative recipients of seropositive donors are at the highest risk, presentation varies from asymptomatic CMV viremia to a systemic syndrome (fever, leukopenia) and tissue-specific manifestation (hepatitis, gastroenteritis, and retinopathy). Plasma viral load and a rise in IgM antibodies to CMV are diagnostic. Valganciclovir has proved effective in both prophylaxis and treatment of CMV disease. Acyclovir is an effective therapy for herpes simplex virus infections.

BK virus is a latent polyomavirus that lies dormant in the kidney and urothelial tract and can be activated in the setting of immunosuppression. Reactivation of BK, if left untreated, will lead to progressive fibrosis and loss of the graft within 1 year in most cases. However, as risk of reactivation of BK infection is associated with the overall degree of immunosuppression, in most cases, BK infection can be managed by regular testing of BK viral load and judicious reduction of maintenance immunosuppression. Renal biopsy can be useful in examining for the presence of interstitial nephritis, tubular cytopathic changes of BK nephropathy, and viral antigens in the allograft. In difficult to treat cases beyond reduction in immunosuppression, a variety of therapies including leflunomide, cidofovir, and quinolone antibiotics (which are effective against polyoma helicase) and IVIG have been tried but with inconsistent results.

CHRONIC LESIONS OF THE TRANSPLANTED KIDNEY

Although current 1-year transplant survival is excellent, most recipients experience a progressive decline in kidney function over time thereafter. Chronic renal transplant dysfunction can be caused by chronic active antibody-mediated rejection, recurrent glomerular disease, hypertension, CNI nephrotoxicity, secondary focal glomerulosclerosis, or a combination of these pathophysiologies. Chronic vascular changes with intimal proliferation and medial hypertrophy are commonly found. Control of systemic and intrarenal hypertension with calcium channel blockers is thought to have a beneficial influence on the rate of progression of chronic allograft dysfunction. Kidney allograft biopsy can distinguish subacute cellular rejection from recurrent disease or secondary focal sclerosis.

MALIGNANCY

The incidence of tumors in patients on immunosuppressive therapy is 5–6%, or ~100 times greater than that in the general population in the same age range. The most common lesions are cancer of the skin and lips. Hence, surveillance for skin cancers and protection from ultraviolet radiation are necessary. Solid organ transplant recipients are at higher risk to develop posttransplant lymphoproliferative disease, most frequently seen early (<1 year) or late (7–10 years) after transplantation. Most cases are associated with Epstein-Barr virus infection, and the prognosis is poor. The overall malignancy risks are increased in proportion to the total immunosuppressive load administered and the time elapsed since transplantation. Treatment of cancer after transplant involves the reduction of immunosuppression, surgery, conventional cytotoxic chemotherapy, and radiotherapy. Cancer immunotherapy is associated with a high risk of allograft rejection (30–40%), and the multidisciplinary risk-benefit discussion should be made before the initiation of therapy.

OTHER COMPLICATIONS

Both chronic dialysis and renal transplant patients have a higher incidence of death from myocardial infarction and stroke than the population at large, and this is particularly true of diabetic patients. Contributing factors are the use of glucocorticoids and sirolimus, as well as hypertension. Recipients of renal transplants have a high prevalence of coronary artery and peripheral vascular diseases. The percentage of deaths from these causes has been slowly rising as the numbers of transplanted diabetic patients and the average age of recipients increase. More than 30% of kidney transplant recipient mortality is attributable to cardiovascular disease. Strict control of blood pressure and blood sugar and lipid levels is essential in this population.

Hypertension may be caused by (1) native kidney disease, (2) rejection activity in the transplant, (3) renal artery stenosis if an end-to-end anastomosis was constructed with an iliac artery branch, and (4) renal CNI toxicity, which may improve with reduction in dose. Calcium channel blockers are shown to improve long-term mortality. Amelioration of hypertension to the range of 120–130/70–80 mmHg should be the goal in all patients.

Hypercalcemia after transplantation may indicate failure of hyperplastic parathyroid glands to regress. Aseptic necrosis of the head of the femur when it occurs is probably due to preexisting hyperparathyroidism, with aggravation by glucocorticoid treatment. With improved management of calcium and phosphorus metabolism during chronic dialysis, the incidence of parathyroid-related complications has fallen dramatically. Persistent hyperparathyroid activity may require subtotal parathyroidectomy.

Although most transplant patients have robust production of erythropoietin and normalization of hemoglobin, *anemia* is commonly seen in the posttransplant period. Often the anemia is attributable to bone marrow–suppressant immunosuppressive medications such as azathioprine, mycophenolic acid, and mTOR inhibitors. Gastrointestinal bleeding is a common side effect of high-dose and long-term steroid administration. Many transplant patients have creatinine clearances of 30–50 mL/min and can be considered to have chronic renal insufficiency for anemia management, including supplemental erythropoietin.

Chronic hepatitis, particularly when due to hepatitis B virus, can be a progressive, fatal disease over a decade or so. Patients who are persistently hepatitis B surface antigen–positive are at higher risk, according to some studies, but the presence of HCV is also a concern when one embarks on a course of immunosuppression in a transplant recipient. However, the introduction of the new highly effective, direct-acting HCV antiviral medications reduced this risk significantly.

In conclusion, while kidney transplantation has progressed significantly toward the goals of longer patient survival and better quality of life, the field still has significant challenges and unmet needs. Advanced immunologic and genetic studies have led and will continue to lead us to detailed understanding of alloimmunity at the molecular level. Noninvasive biomarkers for monitoring and diagnosing rejection and novel therapeutic targets will continue to evolve. Further effort is needed to achieve equity and improve personalized care of kidney transplant recipients.

■ FURTHER READING

Allen PJ et al: Recurrent glomerulonephritis after kidney transplantation: Risk factors and allograft outcomes. Kidney Int 92:461, 2017.

Chadban SJ et al: Summary of the Kidney Disease: Improving Global Outcomes (KDIGO) clinical practice guideline on the evaluation and management of candidates for kidney transplantation. Transplantation 104:708, 2020.

Chapman JR et al: Cancer in the transplant recipient. Cold Spring Harb Perspect Med 3:pii:a015677, 2013.

Euvrard S et al: Sirolimus and secondary skin-cancer prevention in kidney transplantation. N Engl J Med 367:329, 2012.

Grams ME et al: Kidney-failure risk projection for the living kidney-donor candidate. N Engl J Med 374:411, 2016.

Hariharan S et al: Long-term survival after kidney transplantation. N Eng J Med 385:729, 2021.

Hirsh HH et al: BK polyomavirus in solid organ transplantation: Guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. Clin Transplant 33:e13528, 2019.

Kotton CN et al: The third international consensus guidelines on the management of cytomegalovirus in solid-organ transplantation. Transplantation 102:900, 2018.

Lentine KL et al: OPTN/SRTR 2021 Annual data report: Kidney. Am J Transplant 23:21, 2023.

Loupy A et al: Complement-binding anti-HLA antibodies and kidney-allograft survival. N Engl J Med 369:1215, 2013.

Orandi BJ et al: Survival benefit with kidney transplants from HLA-incompatible live donors. N Engl J Med 374:940, 2016.

326

Glomerular Diseases

Julia B. Lewis, Eric G. Neilson

Two human kidneys harbor nearly 1.8 million glomerular capillary tufts. Each glomerular tuft resides within Bowman's space. The capsule circumscribing this space is lined by parietal epithelial cells that transition into tubular epithelia forming the proximal nephron or migrate into the tuft to replenish podocytes. The glomerular capillary tuft derives from an afferent arteriole that forms a branching capillary bed embedded in mesangial matrix **(Fig. 326-1)**. This capillary network funnels into an efferent arteriole, which passes filtered blood into cortical peritubular capillaries or medullary vasa recta that supply and exchange with a folded tubular architecture. Hence, the glomerular capillary tuft, fed and drained by arterioles, represents an arteriolar portal system. Fenestrated endothelial cells resting on a glomerular basement membrane (GBM) line glomerular capillaries. Delicate foot processes extending from epithelial podocytes shroud the outer surface of these capillaries, and adjacent podocytes interconnect to each other by slit-pore membranes forming a selective filtration barrier.

The glomerular capillaries filter 120–180 L/d of plasma water containing various solutes for reclamation or discharge by downstream tubules. Most large proteins and all cells are excluded from filtration by a physicochemical barrier governed by pore size and negative electrostatic charge. The mechanics of filtration and reclamation are quite complicated for many solutes **(Chap. 320)**. For example, in the case of serum albumin, the glomerulus is an imperfect barrier. Although albumin has a negative charge, which would tend to repel the negatively charged GBM, it only has a physical radius of 3.6 nm, while pores in the GBM and slit-pore membranes have a radius of 4 nm. Consequently, variable amounts of albumin inevitably cross the filtration barrier to be reclaimed by megalin and cubilin receptors along the proximal tubule. Remarkably, humans with normal nephrons excrete on average 8–10 mg of albumin in daily voided urine, ~20–60% of total excreted protein. This amount of albumin, and other proteins, can rise to gram quantities following glomerular injury.

The breadth of diseases affecting the glomerulus is expansive because the microenvironment supporting the glomerular capillaries can be injured in a variety of ways, producing many different lesions. Some order to this vast subject is brought by grouping all of these diseases into a smaller number of clinical syndromes.

PATHOGENESIS OF GLOMERULAR DISEASE

There are many forms of glomerular disease with pathogenesis variably linked to the presence of genetic mutations, infection, toxin exposure, autoimmunity, atherosclerosis, hypertension, emboli, thrombosis, or diabetes mellitus. Even after careful study, however, the cause often remains unknown, and the lesion is called *idiopathic*. Specific or unique features of pathogenesis are mentioned with the description of each of the glomerular diseases later in this chapter.

Some glomerular diseases result from genetic mutations producing familial disease or a founder effect: congenital nephrotic syndrome from mutations in *NPHS1* (nephrin) and *NPHS2* (podocin) affects the slit-pore membrane at birth, and *TRPC6* cation channel mutations produce *focal segmental glomerulosclerosis* (*FSGS*) in adulthood;

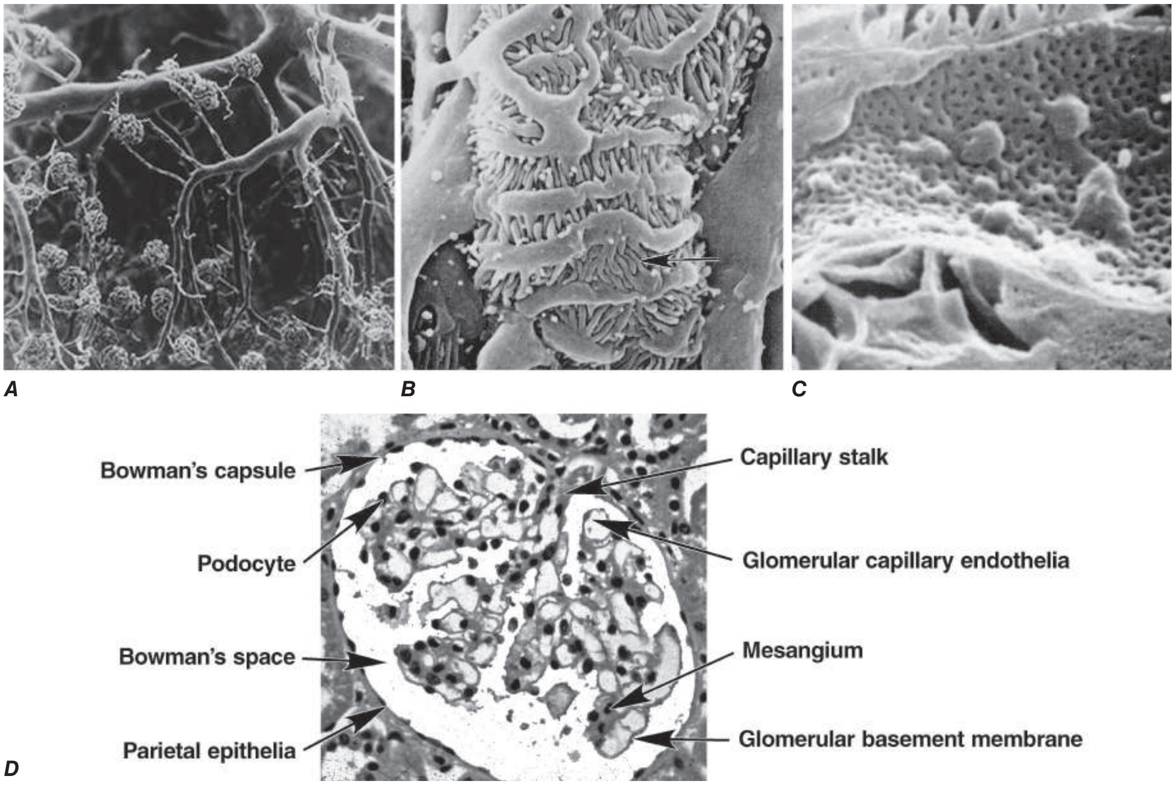

FIGURE 326-1 Glomerular architecture. ***A.*** The glomerular capillaries form from a branching network of renal arteries, arterioles leading to an afferent arteriole, glomerular capillary bed (tuft), and a draining efferent arteriole. *(From VH Gattone II et al: Hypertension 5:8, 1983.)* ***B.*** Scanning electron micrograph of podocytes that line the outer surface of the glomerular capillaries (*arrow* shows foot process). ***C.*** Scanning electron micrograph of the fenestrated endothelia lining the glomerular capillary. ***D.*** The various normal regions of the glomerulus on light microscopy. *(A–C: Courtesy of Dr. Vincent Gattone, Indiana University; with permission.)*

polymorphisms in the gene encoding apolipoprotein L1, *APOL1*, are a major risk for nearly 70% of African Americans with nondiabetic end-stage kidney disease (ESKD), particularly FSGS; monogenetic causes of FSGS are increasingly linked to early age of onset and to genes encoding type IV collagen in older adults, suggesting that much of FSGS may be hereditary; mutations in control of the complement pathway increasingly associate with various forms of *membranoproliferative glomerulonephritis* (*MPGN*) and C_3 glomerulopathies including dense deposit disease, or *atypical hemolytic-uremic syndrome* (*aHUS*); type II partial lipodystrophy from mutations in genes encoding lamin A/C or PPARγ causes a metabolic syndrome associated with MPGN; IgG3 subclass antibodies to antigens expressed on podocytes encoded by *PLAR2gc* indicate a poor prognosis in membranous nephropathy; Alport's syndrome, from mutations in the genes encoding for the α3, α4, or α5 chains of type IV collagen, produces *split basement membranes* with *glomerulosclerosis*; and lysosomal storage diseases, such as α-galactosidase A deficiency causing Fabry's disease and *N*-acetylneuraminic acid hydrolase deficiency causing nephrosialidosis, produce FSGS.

Systemic hypertension and atherosclerosis can produce pressure stress, ischemia, or lipid oxidants that lead to *chronic glomerulosclerosis*. *Malignant hypertension* can quickly complicate glomerulosclerosis with fibrinoid necrosis of arterioles and glomeruli, thrombotic microangiopathy, and acute kidney failure. *Diabetic nephropathy* is an acquired sclerotic injury associated with thickening of the GBM secondary to the long-standing effects of hyperglycemia, advanced glycosylation end products, and reactive oxygen species.

Inflammation of the glomerular capillaries is called *glomerulonephritis*. Most glomerular or mesangial antigens involved in *immune-mediated glomerulonephritis* are unknown **(Fig. 326-2)**. Glomerular epithelial or mesangial cells may shed or express epitopes that mimic other immunogenic proteins made elsewhere in the body. Bacteria, fungi, and viruses can directly infect the kidney, producing their own antigens. Autoimmune diseases such as idiopathic *membranous glomerulonephritis* (*MGN*) or MPGN are confined to the kidney, whereas systemic inflammatory diseases such as *lupus nephritis* or *granulomatosis with polyangiitis* spread to the kidney, causing secondary glomerular injury. *Antiglomerular basement membrane disease* producing Goodpasture's syndrome primarily injures both the lung and kidney because of the narrow distribution of the α3 NC1 domain of type IV collagen that is the target antigen.

Local activation of Toll-like receptors on glomerular cells, deposition of immune complexes, or complement injury to glomerular structures induces mononuclear cell infiltration, which subsequently leads to an adaptive immune response attracted to the kidney by local release of chemokines. Neutrophils, macrophages, and T cells are drawn by chemokines into the glomerular tuft, where they react with antigens and epitopes on or near somatic cells or their structures, producing more cytokines and proteases that damage the mesangium, capillaries, and/or the GBM. While the adaptive immune response is similar to that of other tissues, early T-cell activation plays an important role in the mechanism of glomerulonephritis. Antigens presented by class II major histocompatibility complex (MHC) molecules on macrophages and dendritic cells in conjunction with associative recognition molecules engage the CD4/8 T-cell repertoire.

Mononuclear cells by themselves can injure the kidney, but autoimmune events that damage glomeruli classically produce a humoral immune response. *Poststreptococcal glomerulonephritis, lupus nephritis*, and idiopathic *membranous nephritis* typically are associated with immune deposits along the GBM, while anti-GBM antibodies produce the linear binding of anti-GBM disease. Preformed circulating immune complexes can precipitate along the subendothelial side of the GBM, while other immune deposits form in situ on the subepithelial side. These latter deposits accumulate when circulating autoantibodies find their antigen trapped along the subepithelial edge of the GBM. Immune deposits in the glomerular mesangium may result from the deposition of preformed circulating complexes or in situ antigen-antibody interactions. Immune deposits stimulate the release of local proteases and activate the complement cascade, producing $C_{5–9}$ attack complexes. In addition, local oxidants damage glomerular structures, producing proteinuria and effacement of the podocytes. Overlapping etiologies or pathophysiologic mechanisms can produce similar

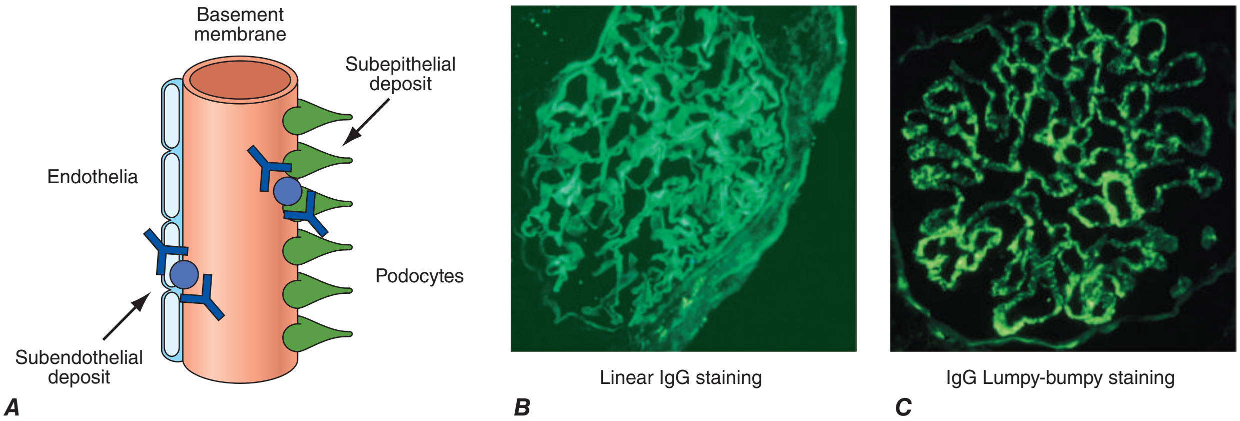

Linear IgG staining

IgG Lumpy-bumpy staining

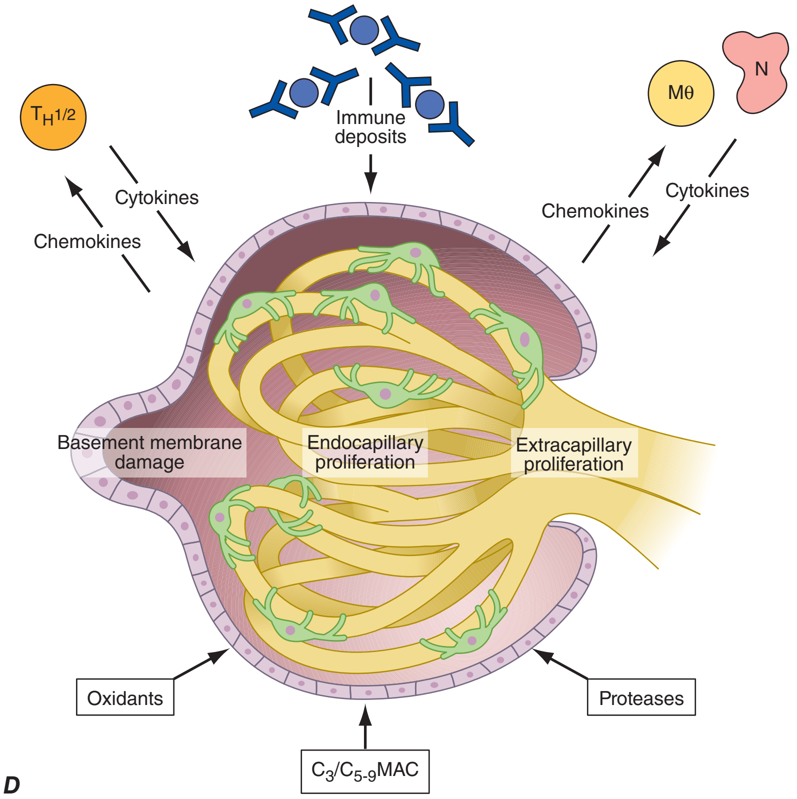

FIGURE 326-2 The glomerulus is injured by a variety of mechanisms. ***A.*** Preformed immune deposits can precipitate from the circulation and collect along the glomerular basement membrane (GBM) in the subendothelial space or can form in situ along the subepithelial space. ***B.*** Immunofluorescent staining of glomeruli with labeled anti-IgG demonstrating linear staining from a patient with anti-GBM disease or immune deposits from a patient with membranous glomerulonephritis. ***C.*** The mechanisms of glomerular injury have a complicated pathogenesis. Immune deposits and complement deposition classically draw macrophages and neutrophils into the glomerulus. T lymphocytes may follow to participate in the injury pattern as well. ***D.*** Amplification mediators as locally derived oxidants and proteases expand this inflammation, and depending on the location of the target antigen and the genetic polymorphisms of the host, basement membranes are damaged with either endocapillary or extracapillary proliferation.

glomerular lesions, suggesting that downstream molecular and cellular responses often converge toward common patterns of injury.

PROGRESSION OF GLOMERULAR DISEASE

Persistent glomerulonephritis that worsens kidney function is always accompanied by interstitial nephritis, renal fibrosis, and tubular atrophy. What is not so obvious, however, is that kidney failure in glomerulonephritis best correlates histologically with the appearance of tubulointerstitial nephritis rather than with the type of inciting glomerular injury.

Loss of kidney function due to interstitial damage is explained hypothetically by several mechanisms. The simplest explanation is that urine flow is impeded by tubular obstruction as a result of interstitial inflammation and fibrosis. Thus, obstruction of the tubules with debris or by extrinsic compression functionally results in aglomerular nephrons. A second mechanism suggests that interstitial changes, including interstitial edema or fibrosis, alter tubular and vascular architecture and thereby compromise the normal tubular transport of solutes and water from tubular lumen to vascular space. This failure increases the solute and water content of the tubule fluid, resulting in isosthenuria and polyuria. Adaptive mechanisms related to tubuloglomerular feedback also fail, resulting in a reduction of renin output from the juxtaglomerular apparatus trapped by interstitial inflammation. Consequently, the local vasoconstrictive influence of angiotensin II on the glomerular arterioles decreases, and filtration drops owing to a generalized decrease in arteriolar tone. A third mechanism involves changes in vascular resistance due to damage of peritubular capillaries. The cross-sectional volume of these capillaries is decreased by interstitial inflammation, edema, or fibrosis. These structural alterations in vascular resistance affect kidney function through two mechanisms. First, tubular cells are very metabolically active, and as a result, decreased perfusion leads to tubular ischemic injury. Second, impairment of glomerular arteriolar outflow leads to increased intravascular hypertension in less-involved glomeruli; this selective intraglomerular hypertension aggravates and extends *mesangial sclerosis* and *glomerulosclerosis* to less-involved glomeruli. Regardless of the exact mechanism, early *acute tubulointerstitial nephritis* **(see Fig. A4-31)** suggests potentially recoverable kidney function, whereas the development of *chronic interstitial fibrosis* prognosticates permanent loss **(see Fig. A4-25)**.

Persistent damage to glomerular capillaries spreads to the tubulointerstitium in association with proteinuria. There is a hypothesis that efferent arterioles leading from inflamed glomeruli carry forward inflammatory mediators, which induces downstream interstitial nephritis, resulting in fibrosis. Glomerular filtrate from injured glomerular capillaries adherent to Bowman's capsule may also be misdirected to the periglomerular interstitium. Most nephrologists believe, however, that proteinuric glomerular filtrate forming tubular fluid is the primary route to downstream tubulointerstitial injury, although none of these hypotheses are mutually exclusive.

The simplest explanation for the effect of proteinuria on the development of interstitial nephritis is that increasingly severe proteinuria, carrying activated cytokines and lipoproteins producing reactive oxygen species, triggers a downstream inflammatory cascade in and around epithelial cells lining the tubular nephron. These effects induce T lymphocyte and macrophage infiltrates in the interstitial spaces along with fibrosis and tubular atrophy.

Tubules disaggregate following direct damage to their basement membranes, leading to more interstitial fibroblasts and fibrosis at the site of injury; recent comprehensive evidence suggests that renal fibroblasts increase through several mechanisms: epithelial or endothelial-mesenchymal transitions (15%), bone marrow–derived fibrocytes (35%), and the proliferation of resident fibroblasts (50%). Most renal myofibroblasts are formed from bone marrow fibrocytes or proliferating fibroblasts. Transforming growth factor β (TGF-β), fibroblast growth factor 2 (FGF-2), hypoxemia-inducible factor 1α (HIF-1α), and platelet-derived growth factor (PDGF) are particularly active in this transition. With persistent nephritis, fibroblasts multiply and lay down tenascin and a fibronectin scaffold for the polymerization of new interstitial collagen types I/III. These events form scar tissue through a process called fibrogenesis. In experimental studies, bone morphogenetic protein 7 and hepatocyte growth factor can reverse early fibrogenesis and preserve tubular architecture. When fibroblasts outdistance their survival factors, apoptosis occurs, and the permanent renal scar becomes acellular, leading to irreversible kidney failure.

APPROACH TO THE PATIENT

Glomerular Disease

HEMATURIA, PROTEINURIA, AND PYURIA

Patients with glomerular disease usually have some hematuria with varying degrees of proteinuria. Hematuria is typically asymptomatic. As few as 3–5 red blood cells in the spun sediment from first-voided morning urine is suspicious. The diagnosis of glomerular injury can be delayed because patients will not realize they have *microscopic hematuria*, and only rarely with the exception of IgA nephropathy and sickle cell disease is *gross hematuria* present. When working up microscopic hematuria, perhaps accompanied by minimal proteinuria (<500 mg/24 h), it is important to exclude anatomic lesions, such as malignancy of the urinary tract, particularly in older men. Microscopic hematuria may also appear with the onset of benign prostatic hypertrophy, interstitial nephritis, papillary necrosis, hypercalciuria, kidney stones, cystic kidney diseases, or renal vascular injury. However, when red blood cell casts **(see Fig. A4-38)** or dysmorphic red blood cells are found in the sediment, glomerulonephritis is likely. A mean of 8–10 mg/24 h of albumin appears in the urine in the absence of kidney disease. In early nephropathy, such as in diabetic nephropathy, proteinuria increases to 30–300 mg/24 h and is called microalbuminuria and represents the presence of kidney disease. Screening spot urine albumin/creatinine ratio (UACR) of >30 mg/g suggests a need for further investigation. Greater than 300 mg/24 h of albuminuria represents frank proteinuria and more advanced kidney disease **(Table 326-1)**.

Sustained proteinuria >1–2 g/24 h is also commonly associated with glomerular disease. Patients often will not know they have proteinuria unless they become edematous or notice foaming urine on voiding. *Sustained proteinuria* has to be distinguished from lesser amounts of so-called *benign proteinuria* in the normal population. (Table 326-1). This latter class of proteinuria is nonsustained, generally <1 g/24 h, and is sometimes called *functional* or *transient proteinuria*. Fever, exercise, obesity, sleep apnea, emotional stress, and congestive heart failure can explain transient proteinuria. Proteinuria only seen with upright posture is called *orthostatic proteinuria* and has a benign prognosis. Isolated proteinuria sustained over multiple clinic visits is found in many glomerular lesions. Proteinuria in most adults with glomerular disease is *nonselective*, containing albumin and a mixture of other serum proteins, whereas in children with *minimal change disease* (MCD), the proteinuria is *selective* and composed largely of albumin.

Some patients with inflammatory glomerular disease, such as acute poststreptococcal glomerulonephritis or MPGN, have *pyuria* characterized by the presence of considerable numbers of leukocytes. This latter finding has to be distinguished from urine infected with bacteria.

CLINICAL SYNDROMES

Various forms of glomerular injury can also be parsed into several distinct syndromes on clinical grounds **(Table 326-2)**. These syndromes, however, are not always mutually exclusive. There is an *acute nephritic syndrome* producing 1–2 g/24 h of proteinuria, hematuria with red blood cell casts, pyuria, hypertension, fluid retention, and a rise in serum creatinine associated with a reduction in glomerular filtration. If glomerular inflammation develops slowly, the serum creatinine will rise gradually over many weeks, but if the serum creatinine rises quickly, particularly over a few days, acute nephritis is sometimes called *rapidly progressive glomerulonephritis* (RPGN); the histopathologic term *crescentic glomerulonephritis* is the pathologic equivalent of the clinical presentation of RPGN. When patients with RPGN present with lung hemorrhage from Goodpasture's syndrome, antineutrophil cytoplasmic antibody (ANCA)–associated small-vessel vasculitis, lupus erythematosus, or cryoglobulinemia, they are often diagnosed as having a *pulmonary-renal syndrome. Nephrotic syndrome* describes the onset of heavy proteinuria (>3.0 g/24 h), hypertension, hypercholesterolemia, hypoalbuminemia, edema/anasarca, and microscopic hematuria; if only large amounts of proteinuria are present without clinical manifestations, the condition is sometimes called *nephrotic-range proteinuria*. The glomerular filtration rate (GFR) in these patients may initially be normal or, rarely, higher than normal, but with persistent hyperfiltration and continued nephron loss, it typically declines over months to years. Patients with a *basement membrane syndrome* either have genetically abnormal basement membranes (Alport's syndrome) or an autoimmune

TABLE 326-1 Urine Assays for Albuminuria/Proteinuria

	24-h ALBUMIN[a] (mg/24 h)	ALBUMIN[a]/CREATININE RATIO (mg/g)	DIPSTICK PROTEINURIA	24-h URINE PROTEIN[b] (mg/24 h)
Normal	8–10	<30	–	<150
Microalbuminuria	30–300	30–300	–/Trace/1+	–/>150
Proteinuria	>300	>300	Trace–3+	>150

[a]Albumin detected by radioimmunoassay. [b]Albumin represents 20–60% of the total protein excreted in the urine.

TABLE 326-2 Patterns of Clinical Glomerulonephritis

GLOMERULAR SYNDROMES	PROTEINURIA	HEMATURIA	VASCULAR INJURY
Acute Nephritic Syndromes			
Poststreptococcal glomerulonephritis[a]	+/++	++/+++	–
Subacute bacterial endocarditis[a]	+/++	++	–
Lupus nephritis[a]	+/++	++/+++	+
Antiglomerular basement membrane disease[a]	++	++/+++	–
IgA nephropathy[a]	+/++	+++[c]	–
ANCA small-vessel vasculitis[a]			
Granulomatosis with polyangiitis (Wegener's)	+/++	++/+++	++++
Microscopic polyangiitis	+/++	++/+++	++++
Churg-Strauss syndrome	+/++	++/+++	++++
Henoch-Schönlein purpura[a]	+/++	++/+++[c]	++++
Cryoglobulinemia[a]	+/++	++/+++	++++
Membranoproliferative glomerulonephritis[a]	++	++/+++	–
C_3 glomerulopathies	++	++/+++	–
Mesangioproliferative glomerulonephritis	+	+/++	–
Pulmonary-Renal Syndromes			
Goodpasture's syndrome[a]	++	++/+++	–
ANCA small-vessel vasculitis[a]			
Granulomatosis with polyangiitis (Wegener's)	+/++	++/+++	++++
Microscopic polyangiitis	+/++	++/+++	++++
Churg-Strauss syndrome	+/++	++/+++	++++
Henoch-Schönlein purpura[a]	+/++	++/+++[c]	++++
Cryoglobulinemia[a]	+/++	++/+++	++++
Nephrotic Syndromes			
Minimal change disease	++++	–	–
Focal segmental glomerulosclerosis	+++/++++	+	–
Membranous glomerulonephritis	++++	+	–
Diabetic nephropathy	++/++++	–/+	–
AL and AA amyloidosis	+++/++++	+	+/++
Light chain deposition disease	+++	+	–
Fibrillary-immunotactoid disease	+++/++++	+	+
Fabry's disease	+	+	–
Basement Membrane Syndromes			
Anti-GBM disease[a]	++	++/+++	–
Alport's syndrome	++	++	–
Thin basement membrane disease	+	++	–
Nail-patella syndrome	++/+++	++	–
Glomerular Vascular Syndromes			
Atherosclerotic nephropathy	+	+	+++
Hypertensive nephropathy[b]	+/++	+/++	++
Cholesterol emboli	+/++	++	+++
Sickle cell disease	+/++	+++[c]	+++
Thrombotic microangiopathies	++	++	+++
Antiphospholipid syndrome	++	++	+++
ANCA small-vessel vasculitis[a]			
Granulomatosis with polyangiitis (Wegener's)	+/++	++/+++	++++
Microscopic polyangiitis	+/++	++/+++	++++
Churg-Strauss syndrome	+++	++/+++	++++
Henoch-Schönlein purpura[a]	+/++	++/+++[c]	++++
Cryoglobulinemia[a]	+/++	++/+++	++++
AL and AA amyloidosis	+++/++++	+	+/++
Infectious Disease–Associated Syndromes			
Poststreptococcal glomerulonephritis[a]	+/++	++/+++	–
Subacute bacterial endocarditis[a]	+/++	++	–
HIV	+++	+/++	–

(Continued)

TABLE 326-2 Patterns of Clinical Glomerulonephritis (Continued)

GLOMERULAR SYNDROMES	PROTEINURIA	HEMATURIA	VASCULAR INJURY
Hepatitis B and C	+++	+/++	–
Syphilis	+++	+	–
Leprosy	+++	+	–
Malaria	+++	+/++	–
Schistosomiasis	+++	+/++	–

[a]Can present as rapidly progressive glomerulonephritis (RPGN); sometimes called crescentic glomerulonephritis. [b]Can present as a malignant hypertensive crisis producing an aggressive fibrinoid necrosis in arterioles and small arteries with microangiopathic hemolytic anemia. [c]Can present with gross hematuria.

Abbreviations: AA, amyloid A; AL, amyloid L; ANCA, antineutrophil cytoplasmic antibodies; GBM, glomerular basement membrane.

response to basement membrane collagen IV (Goodpasture's syndrome) associated with microscopic hematuria, mild to heavy proteinuria, and hypertension with variable elevations in serum creatinine. *Glomerular-vascular syndrome* describes patients with vascular injury producing hematuria and moderate proteinuria. Affected individuals can have vasculitis, thrombotic microangiopathy, antiphospholipid syndrome, or, more commonly, a systemic disease such as atherosclerosis, cholesterol emboli, hypertension, sickle cell anemia, and autoimmunity. *Infectious disease–associated syndrome* is most important if one has a global perspective. Save for subacute bacterial endocarditis (SBE) in the Western Hemisphere, malaria and schistosomiasis may be the most common causes of glomerulonephritis throughout the world, closely followed by HIV and chronic hepatitis B and C. These infectious diseases produce a variety of inflammatory reactions in glomerular capillaries, ranging from nephrotic syndrome to acute nephritic injury, and urinalyses that demonstrate a combination of hematuria and proteinuria.

These six general categories of syndromes are usually determined at the bedside with the help of a history and physical examination, blood chemistries, kidney ultrasound, and urinalysis. These initial studies help frame further diagnostic workup that typically involves testing of the serum for the presence of various proteins (HIV and hepatitis B and C antigens) or antibodies (anti-GBM, antiphospholipid, antistreptolysin O [ASO], PLA2R, THSD7A, anti-DNAse, antihyaluronidase, ANCA, anti-DNA, cryoglobulins, anti-HIV, and anti-hepatitis B and C antibodies) or depletion of complement components (C_3 and C_4). The bedside history and physical examination can also help determine whether the glomerulonephritis is isolated to the kidney (*primary glomerulonephritis*) or is part of a systemic disease (*secondary glomerulonephritis*).

When confronted with an abnormal urinalysis and elevated serum creatinine, with or without edema or congestive heart failure, one must consider whether the glomerulonephritis is *acute* or *chronic*. This assessment is best made by careful history (last known urinalysis or serum creatinine during pregnancy or insurance physical, evidence of infection, or use of medication or recreational drugs), the size of the kidneys on ultrasound examination, and how the patient feels at presentation. Chronic glomerular disease often presents with decreased kidney size. Patients who quickly develop kidney failure are fatigued and weak and often have uremic symptoms associated with nausea, vomiting, fluid retention, and somnolence. Primary glomerulonephritis presenting with kidney failure that has progressed slowly, however, can be remarkably asymptomatic, as are patients with acute glomerulonephritis without much loss in kidney function. Once this initial information is collected, selected patients who are clinically stable, have adequate blood clotting parameters, and are willing and able to receive treatment are encouraged to have a kidney biopsy.

KIDNEY PATHOLOGY

A kidney biopsy in the setting of glomerulonephritis quickly identifies the type of glomerular injury and often suggests a course of treatment. The biopsy is processed for light microscopy using stains for *hematoxylin and eosin* (*H&E*) to assess cellularity and architecture, *periodic acid–Schiff* (*PAS*) to stain carbohydrate moieties in the membranes of the glomerular tuft and tubules, *Jones-methenamine silver* to enhance basement membrane structure, *Congo red* for amyloid deposits, and *Masson's trichrome* to identify collagen deposition and assess the degree of glomerulosclerosis and interstitial fibrosis. Biopsies are also processed for direct immunofluorescence using conjugated antibodies against IgG, IgM, and IgA to detect the presence of "lumpy-bumpy" immune deposits or "linear" IgG or IgA antibodies bound to GBM, antibodies against trapped complement proteins (C_3 and C_4), or specific antibodies against a relevant antigen (PLA2R, THSD7A, and DNAJB9). High-resolution electron microscopy can clarify the principal location of immune deposits and the status of the basement membrane.

Each region of a kidney biopsy is assessed separately. By light microscopy, glomeruli (ideally 20) are reviewed individually for discrete lesions; <50% involvement is considered *focal*, and >50% is *diffuse*. Injury in each glomerular tuft can be *segmental,* involving a portion of the tuft, or *global,* involving most of the glomerulus. Glomeruli having *proliferative* characteristics show increased cellularity. When cells in the capillary tuft proliferate, it is called *endocapillary*, and when cellular proliferation extends into Bowman's space, it is called *extracapillary. Synechiae* are formed when epithelial podocytes attach to Bowman's capsule in the setting of glomerular injury; *crescents*, which in some cases may be the extension of synechiae, develop when fibrocellular/fibrin collections fill all or part of Bowman's space; and *sclerotic* glomeruli show acellular, amorphous accumulations of proteinaceous material throughout the tuft with loss of functional capillaries and normal mesangium. Since *age-related glomerulosclerosis* is common in adults, one can estimate the background percentage of sclerosis by dividing the patient's age in half and subtracting 10. Immunofluorescent and electron microscopy can detect the presence and location of *subepithelial, subendothelial, or mesangial* immune deposits, or *reduplication* or *splitting* of the basement membrane. In the other regions of the biopsy, the vasculature surrounding glomeruli and tubules can show *angiopathy, vasculitis*, the presence of *fibrils*, or *thrombi*. The tubules can be assessed for adjacency to one another; separation can be the result of edema, tubular dropout, or collagen deposition resulting from interstitial fibrosis. Interstitial fibrosis is an ominous sign of irreversibility and progression to kidney failure.

ACUTE NEPHRITIC SYNDROMES

Acute nephritic syndromes classically present with hypertension, hematuria, red blood cell casts, pyuria, and mild to moderate proteinuria. Extensive inflammatory damage to glomeruli causes a fall in GFR and eventually produces uremic symptoms with salt and water retention, leading to edema and hypertension.

POSTSTREPTOCOCCAL GLOMERULONEPHRITIS

Poststreptococcal glomerulonephritis is prototypical for *acute endocapillary proliferative glomerulonephritis*. The incidence of poststreptococcal glomerulonephritis has dramatically decreased in developed countries, and in these locations is typically sporadic. Acute nephritis in developing countries is epidemic and usually affects children between the ages of 2 and 14 years. In developed countries, it is more typical in the elderly, especially in association with debilitating conditions. It is more common in males, and the familial or cohabitant incidence is as high as 40%. Skin and more commonly throat infections with particular M types of streptococci (nephritogenic strains) antedate glomerular disease. Antibiotic therapy does not reduce the occurrence of nephritis. Poststreptococcal glomerulonephritis due to pharyngitis develops 1–3 weeks after infection and 2–6 weeks after skin infection.

The kidney biopsy in poststreptococcal glomerulonephritis demonstrates hypercellularity of mesangial and endothelial cells; glomerular infiltrates of polymorphonuclear leukocytes; granular subendothelial immune deposits of IgG, IgM, C_3, C_4, and C_{5-9}; and subepithelial deposits, which appear as "humps" **(see Fig. A4-6). (See Glomerular Schematic 1.)** Poststreptococcal glomerulonephritis is an immune-mediated disease involving putative streptococcal antigens, circulating immune complexes, and activation of the alternate complement pathway in association with cell-mediated injury. Leading candidate antigens from nephritogenic streptococci are a cationic cysteine proteinase known as streptococcal pyrogenic exotoxin B (SPEB) and NAPlr, the nephritis-associated plasmin receptor. The nephritogenic antigen SPEB has been demonstrated inside the subepithelial deposits.

The classic presentation is an acute nephritic picture with hematuria, pyuria, red blood cell casts, edema, hypertension, and oliguric acute kidney injury (AKI), which may be severe enough to appear as RPGN. Systemic symptoms of headache, malaise, anorexia, and flank pain (due to swelling of the renal capsule) are reported in as many as 50% of cases. Five percent of children and 20% of adults have proteinuria in the nephrotic range. In the first week of symptoms, 90% of patients will have a depressed CH_{50} and decreased levels of C_3 with normal levels of C_4. Positive rheumatoid factor (30–40%), cryoglobulins, circulating immune complexes (60–70%), and ANCA against myeloperoxidase (10%) are also reported. Positive cultures for streptococcal infection are inconsistently present (~25%) but the streptozyme test is positive in 80–95% of patients and includes antibodies to ASO, anti-DNAse, Altase, ASKase, and anti-NAD. Consequently, the diagnosis of poststreptococcal glomerulonephritis rarely requires a kidney biopsy. A subclinical disease is reported in some series to be 4–5 times as common as clinical nephritis, and these latter cases are characterized by asymptomatic microscopic hematuria with low serum C_3 complement levels.

Treatment is supportive, with control of hypertension, edema, and dialysis as needed. Antibiotic treatment for active streptococcal infection should be given to patients and their cohabitants. There is no role for immunosuppressive therapy, even in the setting of crescents. Recurrent poststreptococcal glomerulonephritis is rare despite repeated streptococcal infections. Early death is rare in children but does occur in the elderly. Complete resolution of the azotemia, hematuria, and proteinuria in the majority of children occurs within 3–6 weeks of the onset of nephritis, but 3–10% of children may have persistent microscopic hematuria, non-nephrotic proteinuria, or hypertension. Overall, the prognosis is good, with ESKD being very uncommon in children and adults. The prognosis in elderly patients is worse, with a high incidence of azotemia (up to 60%), nephrotic-range proteinuria, and ESKD.

Glomerular schematic 1

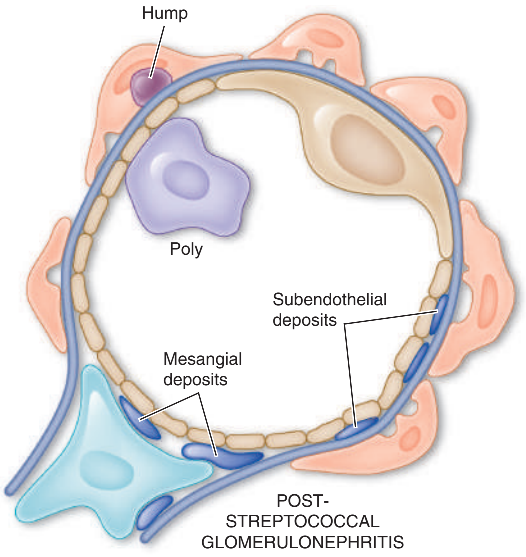

SUBACUTE BACTERIAL ENDOCARDITIS

Endocarditis-associated glomerulonephritis is typically a complication of SBE, particularly in patients who remain untreated for a long time, have negative blood cultures, or have right-sided endocarditis. Common comorbidities are valvular heart disease, intravenous drug use, hepatitis C, and diabetes mellitus. Glomerulonephritis is unusual in acute bacterial endocarditis because it takes 10–14 days to develop immune complex–mediated injury, by which time the patient has been treated, often with emergent surgery. Grossly, the kidneys in SBE have subcapsular hemorrhages with a "flea-bitten" appearance, and kidney biopsy reveals focal or diffuse proliferation with C3 (94%), IgG, and IgA staining, as well as mesangial, subendothelial, and subepithelial immune deposits. Commonly patients present with a clinical picture of RPGN and have crescents on biopsy. Embolic infarcts or septic abscesses may also be present. The pathogenesis hinges on the deposition of circulating immune complexes in the kidney with complement activation. Patients present with gross or microscopic hematuria, pyuria, and mild proteinuria, acute kidney injury, or RPGN with rapid loss of kidney function. A normocytic anemia, elevated erythrocyte sedimentation rate, hypocomplementemia, high titers of rheumatoid factor, type III cryoglobulins, circulating immune complexes, and ANCAs may be present. Levels of serum creatinine may be elevated at diagnosis, but with modern therapy, there is little progression to chronic kidney disease. Primary treatment is eradication of the infection with 4–6 weeks of antibiotics, and if accomplished expeditiously, the prognosis for kidney recovery is good. ANCA-associated vasculitis sometimes accompanies or is confused with SBE and should be ruled out, as the treatment is different.

As variants of persistent bacterial infection in blood-associated glomerulonephritis, infection-associated glomerulonephritis can occur in patients with ventriculoatrial and ventriculoperitoneal shunts; pulmonary, intraabdominal, pelvic, or cutaneous infections; and infected vascular prostheses. In developed countries, a significant proportion of cases afflict adults, especially the immunocompromised, and the predominant organism is *Staphylococcus*. The clinical presentation of these conditions is variable and includes proteinuria, microscopic hematuria, acute kidney injury, and hypertension. Serum complement levels are low, and there may be elevated levels of C-reactive proteins, rheumatoid factor, antinuclear antibodies, and cryoglobulins. Biopsy findings include membranoproliferative glomerulonephritis (MPGN), diffuse proliferative and exudative glomerulonephritis (DPGN), or mesangioproliferative glomerulonephritis, sometimes leading to RPGN. Treatment focuses on eradicating the infection, with most patients treated as if they have endocarditis. The prognosis is guarded.

LUPUS NEPHRITIS

Lupus nephritis is a common and serious complication of systemic lupus erythematosus (SLE). Clinical manifestations of kidney disease are present in 30% of patients at the time of diagnosis, and the majority will develop kidney abnormalities in the course of their disease. Lupus nephritis results from the deposition of circulating immune complexes composed of primarily DNA and anti-DNA, which activate the complement cascade, leading to complement-mediated damage, leukocyte infiltration, activation of procoagulant factors, and release of various cytokines. In situ immune complex formation also plays a role in kidney injury. These immune deposits may occur in the mesangial, subendothelial, and/or subepithelial spaces.

The clinical manifestations, course of disease, and treatment of lupus nephritis are closely linked to kidney pathology. The most common clinical sign of kidney disease is proteinuria, but hematuria, hypertension, varying degrees of kidney injury, and active urine sediment with red blood cell casts can all be present. Anti-dsDNA antibodies that fix complement correlate best with the presence of kidney disease. Hypocomplementemia is common in patients with acute lupus nephritis (70–90%), and declining complement levels may

TABLE 326-3 Classification for Lupus Nephritis

Class I	Minimal mesangial	Normal histology with mesangial deposits
Class II	Mesangial proliferation	Mesangial hypercellularity with expansion of the mesangial matrix
Class III	Focal nephritis	Focal endocapillary ± extracapillary hypercellularity with focal subendothelial immune deposits and mild mesangial expansion ± fibrinoid necrosis
Class IV	Diffuse nephritis	Diffuse endocapillary ± extracapillary hypercellularity with diffuse subendothelial immune deposits and mesangial alterations ± crescents ± fibrinoid necrosis
Class V	Membranous nephritis	Thickened basement membranes with diffuse subepithelial immune deposits; may occur with class III or IV lesions and is sometimes called mixed membranous and proliferative nephritis
Class VI	Sclerotic nephritis	Global sclerosis of nearly all glomerular capillaries

Note: Revised in 2004 by the International Society of Nephrology-Renal Pathology Society Study Group. Recommendation for revision 2018.

herald a flare. A kidney biopsy should be performed in most patients with kidney involvement to establish the histologic subtype, which guides therapy.

The World Health Organization (WHO) workshop in 1974 first outlined several distinct patterns of lupus-related glomerular injury, and this classification was modified in 2004. This version with ongoing refinement **(Table 326-3)** forms the basis for treatment recommendations. Class I nephritis describes normal glomerular histology by normal light microscopy with minimal mesangial deposits on immunofluorescent or electron microscopy. Class II designates mesangial immune complexes with *mesangial proliferation*. Both class I and II lesions are typically associated with minimal kidney manifestation and normal kidney function; nephrotic syndrome is rare. Patients with lesions limited to the renal mesangium have an excellent prognosis and need little or no therapy for their lupus nephritis.

The subject of lupus nephritis is presented under acute nephritic syndromes because of the aggressive and important proliferative lesions seen in class III–V kidney diseases **(see Figs. A4-14 and A4-15)**. Class III describes *focal lesions involving <50% of the glomeruli with proliferation or scarring*, often involving only a segment of the glomerulus. Class III lesions have the most varied course. Hematuria and proteinuria are present, and some patients also have an active urinary sediment, nephrotic syndrome, hypertension, and a decreased GFR. Patients with mild proliferation involving a small percentage of glomeruli respond well to therapy with steroids alone, and <5% progress to kidney failure over 5 years. Patients with more severe proliferation involving a greater percentage of glomeruli or include fibrinoid necrosis have a far worse prognosis and lower remission rates. Treatment of those patients is the same as that for class IV lesions. Class IV describes *diffuse lesions with >50% of the glomeruli involved and proliferative endocapillary lesions with or without extracapillary lesions that may be segmental (IV-S), involving <50% of the glomerular tuft, or global (IV-G), involving >50%.* Patients with class IV lesions commonly have high anti-DNA antibody titers, low serum complement, hematuria, red blood cell casts, proteinuria, hypertension, and decreased kidney function; 50% of patients have nephrotic-range proteinuria. Patients with crescents on biopsy often have a rapidly progressive decline in kidney function. Without treatment, this aggressive lesion has the worst kidney prognosis, with class IV-S worse than class IV-G. However, if a remission—defined as a return to near-normal kidney function and proteinuria ≤330 mg/dL per day—is achieved with treatment, kidney outcomes are excellent. Inducing a remission with administration of high-dose steroids and either cyclophosphamide or mycophenolate mofetil for 2–6 months, followed by maintenance therapy with lower doses of steroids and mycophenolate mofetil or azathioprine, best balances the likelihood of successful remission with the side effects of therapy. Voclosporin may also be used in combination with steroids and mycophenolate mofetil in patients with focal or diffuse lupus nephritis. Belimumab can also be added to standard maintenance therapy. There is no consensus on use of high-dose intravenous methylprednisolone versus oral prednisone, monthly intravenous cyclophosphamide versus daily oral cyclophosphamide, or other immunosuppressants such as cyclosporine, tacrolimus, or rituximab. Prolonged use of cyclophosphamide is avoided in patients of childbearing age.

The class V lesion describes subepithelial immune deposits producing a *membranous pattern*; a subcategory of class V lesions is associated with proliferative lesions and is sometimes called *mixed membranous and proliferative disease* **(see Fig. A4-14)**; this category of injury is treated like class IV glomerulonephritis. Sixty percent of patients present with nephrotic syndrome or lesser amounts of proteinuria. Patients with lupus nephritis class V, like patients with primary membranous nephropathy, are predisposed to renal vein thrombosis and other thrombotic complications. A minority of patients with class V will present with hypertension and kidney dysfunction. There are conflicting data on the clinical course, prognosis, and appropriate therapy for patients with class V disease, which may reflect the heterogeneity of this group of patients. Patients with severe nephrotic syndrome, elevated serum creatinine, and a progressive course will probably benefit from therapy with steroids in combination with other immunosuppressive agents. Therapy with inhibitors of the renin-angiotensin system also may attenuate the proteinuria. Antiphospholipid antibodies present in lupus may result in glomerular microthromboses and a thrombotic microangiopathy. The kidney prognosis is worse despite anticoagulant therapy.

Patients with any of the above lesions also can transform to another lesion; hence, patients often require reevaluation, including repeat kidney biopsy. Lupus patients with class VI lesions have >90% *sclerotic glomeruli* and ESKD with interstitial fibrosis. Up to 20% of patients with lupus nephritis will reach end-stage disease, requiring dialysis or transplantation. Patients with lupus nephritis have a markedly increased mortality compared with the general population. Kidney transplantation usually performed after ~6 months of inactive disease results in allograft survival rates comparable to patients transplanted for other reasons.

■ ANTIGLOMERULAR BASEMENT MEMBRANE DISEASE

Patients who develop autoantibodies directed against glomerular basement antigens frequently develop a glomerulonephritis termed *anti-glomerular basement membrane (anti-GBM) disease*. When they present with lung hemorrhage and glomerulonephritis, they have a pulmonary-renal syndrome called *Goodpasture's syndrome*. The target epitopes for this autoimmune disease lie in the quaternary structure of α3 NC1 domain of collagen IV. Indeed, anti-GBM disease may be considered an autoimmune "conformeropathy" that involves the perturbation of quaternary structure of the α 345NC1 hexamer. MHC-restricted T cells initiate the autoantibody response because humans are not tolerant to the epitopes created by this quaternary structure. The epitopes are normally sequestered in the collagen IV hexamer and can be exposed by infection, smoking, oxidants, or solvents. Goodpasture's syndrome is rare and appears in two age groups: in young men in their late twenties and in men and women in their sixties and seventies. Younger patients are more likely to present with the full Goodpasture's syndrome, with hemoptysis, a sudden fall in hemoglobin, fever, dyspnea, and hematuria, and older patients are more likely to present with isolated glomerulonephritis. Those who present with lung hemorrhage as a group do better than older populations who have prolonged, asymptomatic kidney injury; presentation with oliguria is often associated with a particularly bad outcome. The performance of an urgent kidney biopsy is important in suspected cases of Goodpasture's syndrome to confirm the diagnosis and assess prognosis. Kidney biopsies typically show *focal or segmental necrosis* that later, with aggressive destruction of the capillaries by cellular proliferation, leads to crescent formation in Bowman's space **(see Fig. A4-17)**. As these lesions progress, there is concomitant interstitial nephritis with fibrosis and tubular atrophy.

The presence of anti-GBM antibodies and complement is recognized on biopsy by linear immunofluorescent staining for IgG (rarely IgA). In testing serum for anti-GBM antibodies, it is particularly important that the α3 NC1 domain of collagen IV alone be used as the target. This is because non-nephritic antibodies against the α1 NC1 domain are seen in paraneoplastic syndromes and cannot be discerned from assays that use whole basement membrane fragments as the binding target. Between 10 and 15% of sera from patients with Goodpasture's syndrome also contain ANCA antibodies against myeloperoxidase. Prognosis at presentation is worse if there are >50% crescents on kidney biopsy with advanced fibrosis, if serum creatinine is >5–6 mg/dL, if oliguria is present, or if there is a need for acute dialysis. Patients who present with hemoptysis should be treated for their lung hemorrhage, as it responds to plasmapheresis. Treated patients with less severe disease typically respond to 8–10 treatments of plasmapheresis accompanied by oral prednisone and cyclophosphamide. Maintenance therapy with low-dose immunosuppressants should be considered until antibody titers are negative. There are scarce data alternatively using rituximab or mycophenolate mofetil. Kidney transplantation should wait for 6 months and until serum antibodies are undetectable.

IgA NEPHROPATHY

Berger first described the glomerulonephritis now termed *IgA nephropathy*. It is classically characterized by episodic hematuria associated with the deposition of IgA in the mesangium. IgA nephropathy is one of the most common forms of glomerulonephritis worldwide. There is a male preponderance, a peak incidence in the second and third decades of life, and rare familial clustering. There are geographic differences in the prevalence of IgA nephropathy, with 30-45% prevalence along the Asian and Pacific Rim and 20% in southern Europe, compared to 10% in northern Europe and North America. This may reflect variation in detection or a true variation among racial and ethnic groups.

IgA nephropathy is predominantly a sporadic disease, but susceptibility to it has been shown uncommonly to have a genetic component depending on geography and the existence of "founder effects." No single causal gene has been identified. Clinical and laboratory evidence suggests close similarities between IgA vasculitis (formerly called Henoch-Schonlein purpura) and IgA nephropathy. IgA vasculitis is distinguished clinically from IgA nephropathy by prominent systemic symptoms, a younger age (<20 years old), preceding infection, and abdominal complaints. Deposits of IgA are also found in the glomerular mesangium in a variety of systemic diseases, including chronic liver disease, Crohn's disease, celiac disease, chronic bronchiectasis, idiopathic interstitial pneumonia, dermatitis herpetiformis, mycosis fungoides, ankylosing spondylitis, HIV infection, and Sjögren's syndrome. IgA deposition in these entities is not usually associated with clinically significant kidney disease. IgA-dominant *Staphylococcus*-associated infectious glomerulonephritis is associated with clinically significant kidney disease.

The pathognomonic finding on kidney biopsy is dominant or codominant mesangial IgA deposits, either alone or with IgG, IgM, or complement. **(See Glomerular Schematic 2.)** IgA deposits are typically J-chain containing polymeric IgA1. Abnormalities in the *O*-glycosylation of the hinge region of primarily polymeric IgA1 seem to best account for the pathogenesis. Synthesis of poorly galactosylated IgA1 results in exposure of *N*-acetyl-galactosamine in truncated IgA1 hinge regions, which is recognized by IgG or IgA1 antibodies leading to formation of immune complexes in the circulation or in situ after glomerular deposition of galactose-deficient IgA1 activating the complement system through the alternative and lectin pathways. Viral or other antigen exposure, or hereditary defects in alternative complement pathway proteins may affect the manifestation of disease. Despite the presence of elevated serum IgA levels in 20–50% of patients and IgA deposition in skin biopsies in 15–55% of patients, a kidney biopsy is necessary to confirm the diagnosis. Although the immunofluorescent pattern of IgA on kidney biopsy defines IgA nephropathy in the proper clinical context, a variety of histologic lesions may be seen on light microscopy **(see Fig. A4-8)**, including DPGN; *segmental sclerosis*; and, rarely, *segmental necrosis with cellular crescent formation*, which typically presents as RPGN.

Glomerular schematic 2

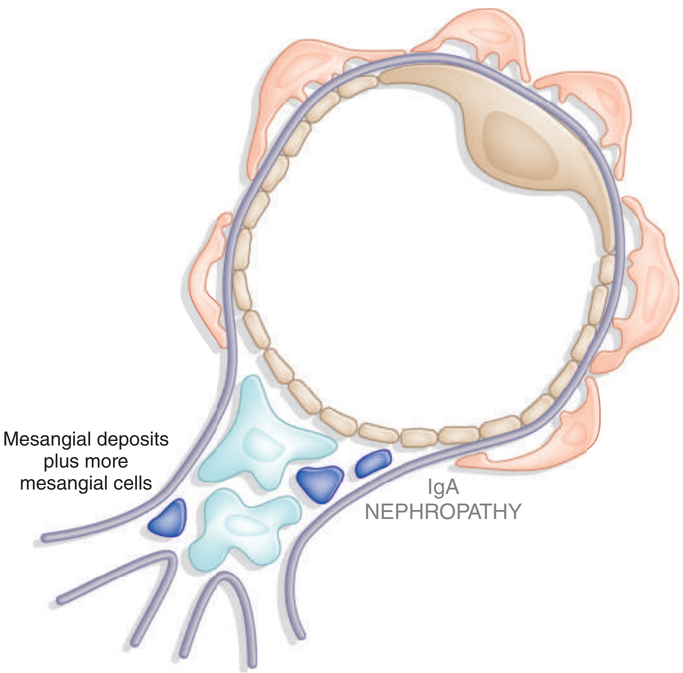

The two most common presentations of IgA nephropathy are persistent asymptomatic microscopic hematuria and recurrent episodes of macroscopic hematuria during or immediately following an upper respiratory infection, often accompanied by proteinuria. Nephrotic syndrome is uncommon. Proteinuria can also first appear late in the course of the disease. Rarely, patients present with acute kidney failure and a rapidly progressive clinical picture. IgA nephropathy can be a benign disease with some patients going into complete remission while others may have ongoing hematuria but well-preserved kidney function. Slow progression to ESKD is seen in only 25–30% of patients over 20–25 years. This risk varies considerably among populations. Cumulatively, risk factors for the loss of kidney function identified thus far account for <50% of the variation in observed outcome but include the presence of hypertension or proteinuria, the absence of episodes of macroscopic hematuria, male sex, and older age of onset. Mesangial hypercellularity (M), endocapillary hypercellularity (E), segmental glomerulosclerosis (S), tubular interstitial fibrosis (T), and crescents (C) have predictive value as established by the Oxford Classification and the MEST-C score. Several analyses in large populations of patients found persistent proteinuria for 6 months or longer to have the greatest predictive power for adverse kidney outcomes.

There is no agreement on optimal treatment. Both large studies that include patients with multiple glomerular diseases and small studies of patients with IgA nephropathy support the use of angiotensin-converting enzyme (ACE) inhibitors in patients with proteinuria or declining kidney function. Steroid treatment or other immunosuppressives have demonstrated conflicting results. An oral targeted-release formulation of budesonide has been shown to reduce proteinuria and improve preserved kidney function in patients with risk of rapid progression. In preliminary studies the sodium-glucose transport-2 inhibitor (SGLT2i) dapagliflozin and the endothelin antagonist sparsentan have reduced adverse kidney outcomes. Tonsillectomy and fish oil have also been suggested in small studies to benefit select patients. When presenting as RPGN, patients typically receive steroids, cytotoxic agents, and plasmapheresis.

ANCA SMALL-VESSEL VASCULITIS

A group of patients with small-vessel vasculitis (arterioles, capillaries, and venules; rarely small arteries) and glomerulonephritis have circulating antineutrophil cytoplasmic autoantibodies (ANCAs); the most common antigen targets are proteinase 3 (PR3) and myeloperoxidase

(MPO) **(Chap. 375)**. ANCA are produced with the help of T cells and activate leukocytes and monocytes, which together damage the walls of small vessels. Endothelial injury also attracts more leukocytes and extends the inflammation. Granulomatosis with polyangiitis (GPA), microscopic polyangiitis (MPA), eosinophilic granulomatosis with polyangiitis (EGPA), and renal-limited vasculitis belong to this group because they are associated with ANCAs and have a *pauci-immune glomerulonephritis* given the absence or paucity of immune complex deposition. Patients with any of these diseases can have any combination of the above serum antibodies, but anti-PR3 antibodies are more common in GPA, and anti-MPO antibodies are more common in MPA or EGPA. Although each of these diseases has some unique clinical features, most features do not predict relapse or progression, and as a group, they are generally treated in the same way. Targeted determination of ANCA levels may be useful in monitoring response to therapy or if a relapse is clinically suspected. Since mortality is high without treatment, virtually all patients receive urgent treatment. Induction therapy usually includes glucocorticoids and either cyclophosphamide or rituximab. Plasmapheresis remains controversial but is generally recommended in rapidly progressive kidney failure or pulmonary hemorrhage, or with concomitantly positive anti-GBM antibodies. Remission is induced in 75–90% of patients, but clinical relapse is common (25–50%). There is evidence to support a reduced dose corticosteroid regimen with discontinuation at 16 weeks recommended. Additionally, novel complement antagonists may have a role as steroid sparing agents. Maintenance therapy is recommended for up to 1–2 years following remission and includes cyclophosphamide or rituximab, or in certain circumstances azathioprine or methotrexate.

Granulomatosis with Polyangiitis Patients with this disease classically present with fever, purulent rhinorrhea, nasal ulcers, sinus pain, polyarthralgias/arthritis, cough, hemoptysis, shortness of breath, hematuria, and subnephrotic proteinuria; occasionally, there may be cutaneous purpura and mononeuritis multiplex. Patients may present without kidney involvement, although most of these patients develop kidney injury later. Chest x-ray often reveals nodules and persistent infiltrates, sometimes with cavities. Biopsy of involved tissue will show a small-vessel vasculitis and adjacent noncaseating granulomas. Kidney biopsies during active disease demonstrate *segmental necrotizing glomerulonephritis* without immune deposits and have been classified as focal, mixed, crescentic, or sclerotic **(see Fig. A4-16)**. The disease is more common in patients exposed to silica dust and those with α_1-antitrypsin deficiency, which is an inhibitor of PR3. Relapse after achieving remission is common and is more common in patients with GPA than the other ANCA-associated vasculitis, necessitating diligent follow-up care. Although associated with an unacceptable high mortality rate without treatment, the greatest threat to patients is from adverse events often secondary to treatment rather than active vasculitis; this is particularly true in elderly patients in the first year of therapy. Patients should also be monitored long term for malignancy after immunosuppressive therapy.

Microscopic Polyangiitis Clinically, these patients present similarly to GPA, except they rarely have significant lung disease or destructive sinusitis. The distinction is made on biopsy, where the vasculitis in MPA is without granulomas. Some patients will also have injury limited to the capillaries and venules.

Eosinophilic Granulomatosis with Polyangiitis When small-vessel vasculitis is associated with peripheral eosinophilia, cutaneous purpura, mononeuritis, asthma, and allergic rhinitis, a diagnosis of EGPA is considered. Hypergammaglobulinemia, elevated levels of serum IgE, or the presence of rheumatoid factor sometimes accompanies the allergic state. Lung inflammation, including fleeting cough and pulmonary infiltrates, often precedes the systemic manifestations of disease by years; lung manifestations are rarely absent. A third of patients may have exudative pleural effusions associated with eosinophils. Small-vessel vasculitis and *focal segmental necrotizing glomerulonephritis* without immune deposits can be seen on kidney biopsy, usually absent of eosinophils or granulomas. The cause of is thought to be abnormal immune function, but the inciting factors are unknown.

C_3 GLOMERULOPATHIES

C_3 glomerulopathy is a recent disease classification that is defined by the glomerular accumulation of C_3 with little or no immunoglobulin and encompasses dense deposit disease (DDD), formerly MPGN type II (see below), and C_3 glomerulonephritis (C_3GN). DDD is defined morphologically by dense deposits forming ribbons in the GBM. In the absence of this specific morphology, the entity is categorized as C_3GN. Both are associated with the presence of a complement mutation believed to cause the kidney pathology, including mutations in the complement factor H regulatory (CFHR) protein genes. DDD is primarily a disease of children and young adults, whereas the other C_3 glomerulopathies are reported to present in an older age group (mean age 30). By definition, kidneys with C_3 glomerulopathy show sole or dominant staining for C_3 but can have variable light microscopy, with mesangial proliferative or membranoproliferative patterns seen most commonly. Morphologically, many cases are not distinguishable from recovering postinfectious glomerulonephritis. Low serum C3 and a dense thickening of the GBM containing ribbons of dense despits and C3 characterize DDD **(see Fig. A4-10)**. Patients with DDD present with proteinuria, which may be nephrotic range, and/or hematuria, which may be macroscopic or microscopic. Partial lipodystrophy and Drusen bodies in the retina may also be present. Prognosis is poor, with 50% of patients progressing to ESKD. C_3GN patients are clinically less well defined, but approximately two-thirds have hematuria and one-third have proteinuria. C_3 levels are low with normal C_4, and C_3 nephritic factor is present in most patients with DDD and less commonly in C_3GN. Abnormalities in factor H, soluble C5b-9, paraprotein detection, and specific CFHR genetic mutations should be assessed as well. Screening family members may be indicated. The optimal therapies remain undefined but include inhibition of the renin-angiotensin system, lipid lowering, steroids, and other immunosuppressants. Evidence suggests a benefit of therapy with eculizumab, a monoclonal antibody directed at C_5, which is activated by C_3.

MEMBRANOPROLIFERATIVE GLOMERULONEPHRITIS

MPGN has been previously identified as a disease with the classification MPGN types I, II, and III; however, it is now considered a pattern of glomerular injury characterized by mesangial and subendothelial immune complexes, complement deposits and/or monoclonal protein deposits, or chronic endothelial cell injury. In addition to increased mesangial and endocapillary hypercellularity, there are thickened GBM's with interposition of cellular elements between the endothelial cell and the GBM creating double contours, sometimes call "tram-tracking" **(see Fig. A4-9). (See Glomerular Schematic 3.)** The MPGN pattern of injury can be seen with immune complex mediated diseases

Glomerular schematic 3

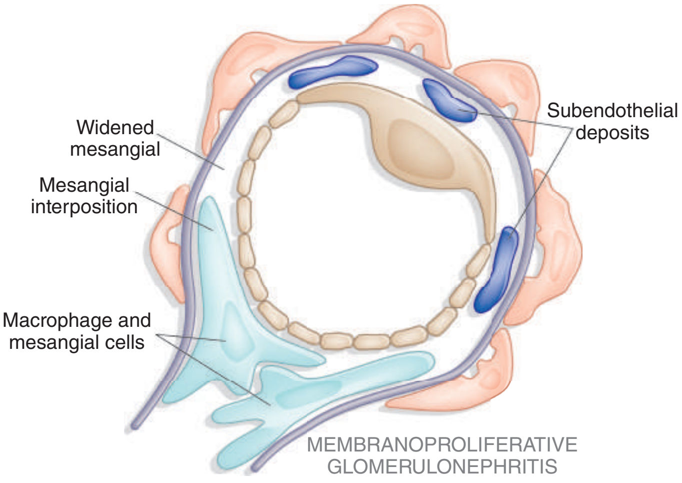

including cryoglobulinemia, infection associated GN, and hepatitis C and B; complement mediated diseases including inherited or acquired dysregulation of complement (see C3G Glomerulopathies); and monoclonal immunoglobulin mediated diseases including light and heavy chain deposition diseases. The MPGN pattern of injury can be seen with unusual deposits as well including fibrillary glomerulonephritis. Each of these diseases have distinct findings on biopsy, natural history and treatments.

MESANGIOPROLIFERATIVE GLOMERULONEPHRITIS

Mesangioproliferative glomerulonephritis is characterized by expansion of the mesangium, sometimes associated with mesangial hypercellularity; thin, single contoured capillary walls; and mesangial immune deposits. Mesangioproliferative pathology may be seen in IgA nephropathy, *Plasmodium falciparum* malaria, resolving postinfectious glomerulonephritis, and class II nephritis from lupus, all of which can have a similar histologic appearance. With these secondary entities excluded, the diagnosis of *primary mesangioproliferative glomerulonephritis* is made in <15% of kidney biopsies.

NEPHROTIC SYNDROME

Nephrotic syndrome classically presents with heavy proteinuria, minimal hematuria, hypoalbuminemia, hypercholesterolemia, edema, and hypertension. If left undiagnosed or untreated, some of these syndromes will progressively damage enough glomeruli to cause a fall in GFR, producing kidney failure. Multiple studies have noted that the higher the 24-h urine protein excretion, the more rapid is the decline in GFR.

Therapies for various causes of nephrotic syndrome are noted under individual disease headings below. In general, all patients with hypercholesterolemia secondary to nephrotic syndrome should be treated with lipid-lowering agents because they are at increased risk for cardiovascular disease. Edema secondary to salt and water retention can be controlled with the use of diuretics, avoiding intravascular volume depletion. Venous complications secondary to the hypercoagulable state associated with nephrotic syndrome can be treated with anticoagulants. The losses of various serum binding proteins, such as thyroid-binding globulin, lead to alterations in functional tests. Lastly, proteinuria itself is hypothesized to be nephrotoxic, and treatment of proteinuria with inhibitors of the renin-angiotensin system and SGLT2i can lower urinary protein excretion and preserve kidney function across a wide range of chonic kidney diseases.

MINIMAL CHANGE DISEASE

MCD, sometimes known as *nil lesion,* causes 70–90% of nephrotic syndrome in childhood but only 10–15% of nephrotic syndrome in adults. MCD usually presents as a primary kidney disease but can be associated with several conditions, including Hodgkin's disease, allergies, use of nonsteroidal anti-inflammatory agents or lithium, infections, and other glomerular diseases. MCD on kidney biopsy shows no glomerular lesion by light microscopy and is negative for deposits by immunofluorescent microscopy or occasionally shows small amounts of IgM in the mesangium **(see Fig. A4-1). (See Glomerular Schematic 4.)** Electron microscopy, however, consistently demonstrates an effacement of the foot processes supporting the epithelial podocytes with weakening of slit-pore membranes. The pathogenesis of this lesion is unclear, although immune dysfunction is likely the initiating factor for MCD. Disturbances in T cell response may result in circulating glomerular permeability factors which directly affect the glomerular capillary wall, postulated to be T cell derived cytokines, such as IL13 and IL4. There is also evidence of increased CD80 on podocytes, promoting dysregulated interactions with T cells via T lymphocyte antigen-4 (CLTA-4). Other evidence supporting cell-mediated immunity includes the presence of preceding allergies, altered cell-mediated immunity during viral infections, and a high frequency of remissions with steroids. Evidence to suggest that there is a role for B cell dysfunction includes reports of patients with anti-nephrin antibodies and the successful treatment with the anti-CD20 monoclonal antibody, rituximab.

Glomerular schematic 4

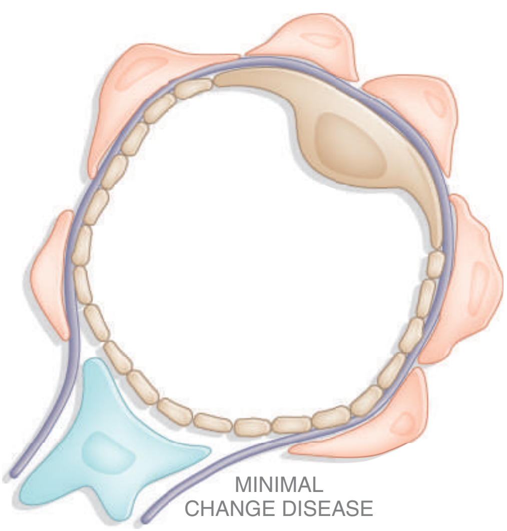

MCD presents clinically with the abrupt onset of edema and nephrotic syndrome accompanied by acellular urinary sediment. Average urine protein excretion reported in 24 h is 10 g with severe hypoalbuminemia. Less common clinical features include hypertension (30% in children, 20–50% in adults), microscopic hematuria (20% in children, 33% in adults), atopy or allergic symptoms (40% in children, 30% in adults), and decreased kidney function (25–40%), which often returns to normal after remission of the nephrotic syndrome. The appearance of acute kidney failure in adults is often seen more commonly in patients with low serum albumin and intrarenal edema (nephrosarca) that is responsive to diuretics. This presentation must be distinguished from acute kidney failure secondary to hypovolemia. Acute tubular necrosis and interstitial inflammation are also reported. In children, the abnormal urine principally contains albumin with minimal amounts of higher-molecular-weight proteins and is sometimes called *selective proteinuria*. Although up to 30% of children have a spontaneous remission, most children today are treated with steroids; only children who are nonresponders are biopsied. Primary responders are patients who have a complete remission (<0.2 mg/24 h of proteinuria), often abruptly after a single course of prednisone; steroid-dependent patients relapse as their steroid dose is tapered. Frequent relapsers have two or more relapses in the 6 months following taper, and steroid-resistant patients fail to respond to steroid therapy. Adults are not considered steroid-resistant until after 4 months of therapy. Ninety to 95% of children will develop a complete remission after 8 weeks of steroid therapy, and 80–90% of adults will achieve complete remission, but the response is more delayed. Patients with steroid resistance may have FSGS on repeat biopsy. If the first kidney biopsy does not have a sample of deeper corticomedullary glomeruli, then the correct diagnosis of FSGS may be missed.

Relapses occur in 70–75% of children after the first remission, and early relapse predicts multiple subsequent relapses, as do high levels of basal proteinuria. The frequency of relapses decreases after puberty. Relapses are less common in adults but are more resistant to subsequent therapy. Prednisone is first-line therapy. Other immunosuppressive drugs, such as cyclophosphamide, mycophenolate mofetil, calcineurin inhibitors (CNIs), and rituximab are reserved for frequent relapsers, steroid-dependent patients, or steroid-resistant patients. CNIs can induce remission, but relapse is also common when withdrawn. The long-term prognosis in adults is less favorable when acute kidney failure or steroid resistance occurs.

FOCAL SEGMENTAL GLOMERULOSCLEROSIS

FSGS refers to a pattern of kidney injury characterized by segmental glomerular scars that involve some but not all glomeruli (focal); the clinical findings of FSGS largely manifest as proteinuria. When the secondary and genetic causes of FSGS are eliminated **(Table 326-4)**, the remaining patients are considered to have primary FSGS. FSGS

TABLE 326-4 Focal Segmental Glomerulosclerosis

Primary focal segmental glomerulosclerosis

- Yet to be identified circulating permeable factor

Secondary focal segmental glomerulosclerosis

- Adaptive response to hyperfiltration/reduced kidney mass, obesity
- Viruses: HIV/hepatitis B/parvovirus/SARS-CoV-2
- Hypertensive nephropathy
- Reflux nephropathy
- Cholesterol emboli
- Drugs: Heroin/analgesics/bisphosphonates/ecstasy/interferon/anabolic steroids
- Oligomeganephronia
- Sickle cell disease
- Radiation nephritis
- Familial podocytopathies
  - *NPHS1* mutation/nephrin
  - *NPHS2* mutation/podocin
  - *PLCE1* mutation/phospholipase Cε1
  - *INF2* mutation/inverted formin 2
  - *WT1* mutation/Wilms tumor
  - *TRPC6* mutation/cation channel
  - *ACTN4* mutation/actinin
  - α-Galactosidase A deficiency/Fabry's disease
  - *N*-Acetylneuraminic acid hydrolase deficiency/nephrosialidosis

Uncertain cause

is now recognized as the most common cause of primary glomerular disease in patients with ESKD in the US. The pathogenesis of FSGS has multiple possible mechanisms including a circulating permeability factor (primary FSGS), an adaptive response to glomerular hypertrophy or hyperfiltration, and podocyte abnormalities associated with direct toxic injury or genetic mutations. Risk polymorphisms at the *APOL1* locus expressed in podocytes are thought to predispose patients to FSGS.

The pathologic changes of FSGS are most prominent in glomeruli located at the corticomedullary junction **(see Fig. A4-2)**, so if the kidney biopsy specimen is from superficial tissue, the lesions can be missed, which sometimes leads to a misdiagnosis of MCD. The most frequent variant is FSGS not otherwise specified (NOS), previously called classic FSGS, which may occur in primary or secondary FSGS. In addition to focal and segmental scarring, other variants have been described, including *cellular* lesions with endocapillary hypercellularity and heavy proteinuria; *collapsing glomerulopathy* **(see Fig. A4-3)** with segmental or global glomerular collapse and a rapid decline in kidney function; a *perihilar* lesion **(see Fig. A4-4)**; or the *glomerular tip* lesion **(see Fig. A4-5)**, which is usually steroid responsive and has a favorable prognosis. **(See Glomerular Schematic 5.)**

FSGS can present with hematuria, hypertension, any level of proteinuria, and kidney injury. Nephrotic-range proteinuria and kidney injury are associated with a poor outcome, with 50% of patients reaching ESKD in 10 years. FSGS rarely remits spontaneously, but treatment-induced remission of proteinuria significantly improves prognosis. Treatment of patients with *FSGS* should include inhibitors of the renin-angiotensin system and SGLT2i. Patients with primary FSGS with nephrotic-range proteinuria can be treated with steroids but respond far less often and after a longer course of therapy than patients with MCD. Proteinuria remits in only 20–45% of patients receiving a course of steroids over 6–12 months. CNIs have been used in patients requiring a steroid sparing regimen. Limited evidence suggests the use of cyclosporine in steroid-responsive patients helps ensure remissions. Relapse frequently occurs after cessation of cyclosporine therapy, and CNIs themselves can lead to a deterioration of kidney function due to their nephrotoxic effects. A role for other agents that suppress the immune system such as rituximab or mycophenolate mofetil has not

Glomerular schematic 5

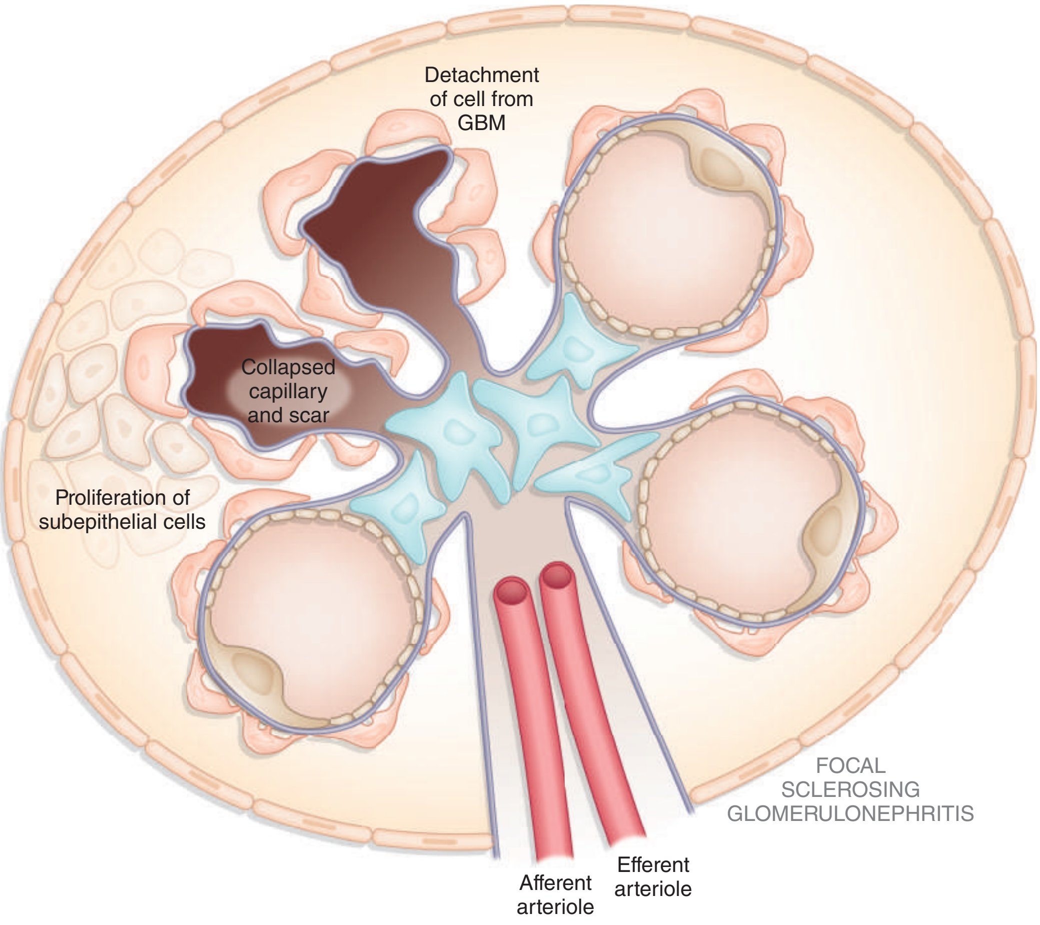
Detachment of cell from GBM
Collapsed capillary and scar
Proliferation of subepithelial cells
FOCAL SCLEROSING GLOMERULONEPHRITIS
Efferent arteriole
Afferent arteriole

TABLE 326-5 Membranous Glomerulonephritis

Primary/antigen-associated membranous glomerulonephritis
PLA2R, NELL1, THSD7A, Sema3B, PCDH7, HTRA1, EXT1, EXT2, NCAM1
Secondary membranous glomerulonephritis
Infection: Hepatitis B and C, syphilis, malaria, schistosomiasis, leprosy, filariasis
Cancer: Breast, colon, lung, stomach, kidney, esophagus, neuroblastoma
Drugs: Gold, mercury, penicillamine, nonsteroidal anti-inflammatory agents, probenecid, antitumor necrosis factor agents
Autoimmune diseases: Systemic lupus erythematosus, rheumatoid arthritis, primary biliary cirrhosis, dermatitis herpetiformis, bullous pemphigoid, myasthenia gravis, Sjögren's syndrome, Hashimoto's thyroiditis, IgG4 disease
Other systemic diseases: Fanconi's syndrome, sickle cell anemia, diabetes, Crohn's disease, sarcoidosis, Guillain-Barré syndrome, Weber-Christian disease, angiofollicular lymph node hyperplasia

Abbreviations: PLA2R, phospholipase A2 receptor; EXT1 and EXT2, exostosin 1 and 2; NCAM1, neural cell adhesion molecule 1; NELL1, neural epidermal growth factor-like 1; PCDH7, protocadherin 7; *Sema3B,* semaphorin 3B; THSD7A, thrombospondin type-1 domain containing 7A.

been established. FSGS recurs in 30% of kidney transplants, more commonly in primary FSGS, less commonly in secondary FSGS, and rarely in genetic FSGS. In recurrent posttransplant FSGS from presumed circulating factor, many patients will achieve a full or partial remission with plasmapheresis. The treatment of *secondary FSGS* typically involves treating the underlying cause and controlling proteinuria. There is no role for steroids or other immunosuppressive agents in secondary or genetic FSGS.

MEMBRANOUS GLOMERULONEPHRITIS

MGN, or *membranous nephropathy* as it is sometimes called, accounts for ~25% of cases of nephrotic syndrome in adults, with a peak incidence between the ages of 30 and 50 years and a male-to-female ratio of 2:1. MGN is most often primary in the absence of an identifiable associated disease, but with the identification of novel associated antigens **(see Table 326-5)**, this nomenclature may change particularly in children in whom these antigens are increasingly found. In 20–30% of cases, MGN is secondary and is associated with a malignancy (solid tumors of the breast, lung, colon), infection (hepatitis B, syphilis, malaria, schistosomiasis), rheumatologic disorders such as lupus or rheumatoid arthritis, IgG4 diseases, or drug exposure (see Table 326-5).

Uniform thickening of the basement membrane along the peripheral capillary loops is seen by light microscopy on kidney biopsy **(see Fig. A4-7)**; this thickening needs to be distinguished from that seen in diabetes and amyloidosis. **(See Glomerular Schematic 6.)** There may be characteristic "spikes" seen on silver stain due to the glomerular basement membrane reactivity to deposits. Immunofluorescence demonstrates diffuse granular deposits of IgG and C_3, and electron microscopy typically reveals electron-dense subepithelial deposits. While different stages (I–V) of progressive membranous lesions have been described, some published analyses indicate the degree of tubular atrophy or interstitial fibrosis is more predictive of progression than is the stage of glomerular disease. The presence of subendothelial deposits or the presence of tubuloreticular inclusions strongly points to a diagnosis of membranous lupus nephritis, which may precede the extrarenal manifestations of lupus. In 70–80% of cases of primary MGN, IgG_4 autoantibodies against the M-type phospholipase A_2 receptor circulate and bind to a conformational epitope present in the PLA2R on human podocytes, producing characteristic in situ deposits. The diagnosis of MGN may no longer require a kidney biopsy in the setting of anti-PLA2R antibody positivity, normal creatinine, and no identifiable other causes of MGN. THSD7A is another antigen localized on the podocyte surface that complexes with IgG_4, while NELL1 is the most common non-PLA2R antigen accounting for up to 16% of primary MGN cases. These novel antigens (see Table 326-5) along with PLA2R account for over 90% of cases of primary MGN. In most cases of secondary membranous nephropathy, autoantibodies to these antigens are absent, with rare reports of autoantibodies to PLA2R associated with hepatitis B, cancer, and sarcoidosis. Circulating deposits and glomerular deposits of these autoantibodies have correlated with the likelihood of a spontaneous remission, severity of primary MGN, and the response to therapy.

Glomerular schematic 6

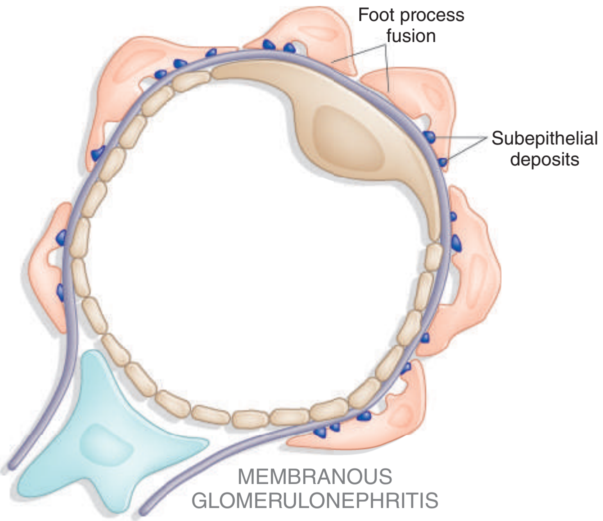

Eighty percent of patients with MGN present with nephrotic syndrome and nonselective proteinuria. Microscopic hematuria is seen but less commonly than in IgA nephropathy or FSGS. Spontaneous remissions occur in 20–33% of patients and often occur late in the course, making treatment decisions difficult. One-third of patients continue to have relapsing nephrotic syndrome but maintain normal kidney function, and approximately another third of patients develop kidney failure or die from the complications of nephrotic syndrome. Male gender, older age, hypertension, and the persistence of nephrotic-range proteinuria are associated with worse prognosis. Although thrombotic complications are a feature of all nephrotic syndromes, MGN has the highest reported incidences of renal vein thrombosis, pulmonary embolism, and deep-vein thrombosis. Prophylactic anticoagulation is controversial but has been recommended in select patients with hypoalbuminemia.

The treatment of edema, dyslipidemia, hypertension, and inhibition of the renin-angiotensin system and the use of SGLT2i is recommended. Therapy with immunosuppressive drugs is recommended for patients with primary MGN and persistent proteinuria. The choice of immunosuppressive drugs for therapy is controversial, however, patient risk stratification based on proteinuria, GFR, and serum albumin can help guide therapy choices with steroids and cyclophosphamide, CNIs or rituximab. Attaining remission is associated with a good long-term prognosis.

DIABETIC NEPHROPATHY

Diabetic nephropathy is the single most common cause of chronic kidney disease in the United States and worldwide. The dramatic increase in the number of patients with diabetic nephropathy reflects the epidemic increase in obesity and type 2 diabetes mellitus. Approximately 40% of patients with diabetes develop nephropathy; the vast majority of patients with diabetic nephropathy will have type 2 diabetes due to the higher prevalence compared to type 1 diabetes.

Within 1–2 years after the onset of clinical diabetes, morphologic changes appear in the kidney. Thickening of the GBM is a sensitive indicator for the presence of diabetes but correlates poorly with the presence of nephropathy. The composition of the GBM is altered notably with a loss of heparan sulfate moieties that form the negatively charged filtration barrier resulting in increased filtration of serum proteins into the urine. The expansion of the mesangium due to the accumulation of extracellular matrix correlates with the clinical manifestations of diabetic nephropathy **(see stages in Fig. A4-24)**. This expansion in mesangial matrix is associated with the development of *mesangial sclerosis.* Some patients also develop eosinophilic, PAS^+ nodules called *nodular glomerulosclerosis* or *Kimmelstiel-Wilson nodules.*

Immunofluorescence microscopy often reveals the nonspecific deposition of IgG (at times in a linear pattern) or complement staining without immune deposits on electron microscopy. Prominent vascular changes are frequently seen with hyaline and hypertensive arteriosclerosis. This is associated with varying degrees of chronic glomerulosclerosis and tubulointerstitial changes. Kidney biopsies from patients with types 1 and 2 diabetic nephropathies with albuminuria are largely indistinguishable. Patients with type 2 diabetes without albuminuria are classified as having diabetic kidney disease as opposed to diabetic nephropathy and may have myriad pathologic findings.

Multiple lines of evidence support an important role for changes in glomerular hemodynamics including increases in glomerular capillary pressure and glomerular hyperfiltration in these pathologic changes. Hyperglycemia activates the renin-angiotensin-aldosterone system and alters insulin-like growth factor, reactive oxygen species, and endothelin 1. Diabetes upregulates the sodium-glucose cotransporters (SGLT1 and SGLT2) in the proximal tubule, resulting in decreased distal delivery of sodium to the macula densa and further glomerular hyperfiltration. Sustained glomerular hypertension increases matrix production and alterations in the GBM with disruption in the filtration barrier. Other factors that alter matrix production include the accumulation of advanced glycosylation end products, circulating factors including growth hormone, connective tissue growth factor, TGF-β, and dyslipidemia.

The natural history of diabetic nephropathy has been historically well characterized in the ~40% of diabetics who develop it as a progression from glomerular hyperfiltration and renal hypertrophy to increasing albuminuria followed by declining GFR and ESKD. However, since the onset of type 1 diabetes is readily identifiable and the onset of type 2 diabetes is not, a patient newly diagnosed with type 2 diabetes may present with *advanced diabetic nephropathy*. Albuminuria and decreased GFR are potent risk factors for cardiovascular disease, with some patients dying before they reach ESKD. Furthermore, contemporary studies reveal that up to 24% of patients with type 1 diabetes and 50% with type 2 diabetes and chronic kidney disease may be normoalbuminuric. It is unknown whether this alteration in the natural history reflects contemporary effective interventions or perhaps other kidney diseases that happen to occur in patients with diabetes. The degree of early glomerular hyperfiltration does correlate with the development of albuminuria and declining GFR. Albuminuria in the range of 30–300 mg/24 h is called *microalbuminuria* (Table 326-1). Microalbuminuria appears 5–10 years after the onset of diabetes. It is currently recommended to test patients with type 1 disease for microalbuminuria 5 years after diagnosis of diabetes and yearly thereafter. Patients with type 2 diabetes should be tested for microalbuminuria at time of diagnosis followed by annual monitoring. Microalbuminuria classically progresses over 5–10 years to proteinuria and declining GFR, but in contemporary studies, greater heterogeneity is reported with regression to normoalbuminuria. Albuminuria remains the single most important predictor of a faster decline in GFR. Regression of albuminuria with a treatment intervention is a good prognostic sign. Proteinuria in diabetic nephropathy can be variable, ranging from 500 mg to 25 g/24 h. More than 90% of patients with type 1 diabetes and nephropathy have diabetic retinopathy, so the absence of retinopathy in type 1 patients with proteinuria should prompt consideration of a diagnosis other than diabetic nephropathy; only 60% of patients with type 2 diabetes with nephropathy have diabetic retinopathy. There is a significant correlation between the presence of retinopathy and the presence of Kimmelstiel-Wilson nodules. Even with advanced chronic kidney disease, patients with diabetic nephropathy will have enlarged kidneys. Using the above data, and in the absence of other clinical or serologic data suggesting another disease, diabetic nephropathy is usually diagnosed without a kidney biopsy. The risk of progression to ESKD is influenced by treatment and other risk factors, and reports vary from a decline of 1.8–14 mL/min per year. Survival on dialysis is worse for patients with diabetes. Kidney transplantation results in better survival.

Good evidence supports the benefits of blood sugar and blood pressure control, inhibitors of the renin-angiotensin-aldosterone system (RAAS), and inhibitors of SGLT2 in slowing the progression of diabetic nephropathy. In patients with type 1 diabetes, intensive control of blood sugar clearly prevents the development or progression of diabetic nephropathy. The evidence for benefit of intensive blood glucose control in patients with type 2 diabetes is less certain.

Controlling systemic blood pressure decreases kidney and cardiovascular adverse events in this high-risk population. The vast majority of patients with diabetic nephropathy require three or more antihypertensive drugs to achieve this goal. Drugs that inhibit the RAAS (ACE inhibitors, angiotensin receptor blockers [ARBs]) have been shown in large clinical trials to slow the progression of diabetic nephropathy at early (microalbuminuria) and late (proteinuria with reduced glomerular filtration) stages, independent of their effects on systemic blood pressure. Finerenone, a nonsteroidal mineralocorticoid receptor antagonist, therapy in patients with diabetes and nephropathy on ACEi or ARB improved cardiovascular and kidney outcomes. In patients with type 2 diabetes and kidney disease, the risk of kidney failure and cardiovascular events was lower in those receiving SGLT2 inhibitors in addition to ACE inhibitors or ARBs.

GLOMERULAR DEPOSITION DISEASES

Plasma cell dyscrasias producing excess light chain (80%), heavy chain (10%) or both immunoglobulins can lead to the formation of glomerular and tubular deposits. The same is true for the accumulation of serum amyloid A protein fragments and monoclonal gammopathy of renal significance (MGRS). This broad group of proteinuric patients has *glomerular deposition disease*.

Light Chain Deposition Disease The biochemical characteristics of nephrotoxic light chains produced in patients with light chain malignancies confer kidney injury; that of either *cast nephropathy* **(see Fig. A4-21)**, which causes kidney injury but not heavy proteinuria or amyloidosis, or light chain deposition disease (LCDD) **(see Fig. A4-20)**, which produces proteinuria with kidney injury. These latter patients produce kappa light chains that do not have the biochemical features necessary to form amyloid fibrils. Instead, they self-aggregate and form granular deposits along the glomerular capillary and mesangium or, more prominently, in the tubular basement membrane and Bowman's capsule. Light chain deposits are not fibrillar and do not stain with Congo red, but they are easily detected with anti–light chain antibody. A combination of the light chain rearrangement, self-aggregating properties at neutral pH, and abnormal metabolism probably contributes to the deposition. Multiple myeloma, Waldenström's macroglobulinemia, or lymphoma may be present, as well as heart, liver, and pulmonary involvement. The monoclonal protein may be found with serum electrophoresis or with serum free light chain analysis. Nephrotic syndrome may develop, and ~70% of patients progress to dialysis. Treatment for LCDD is treatment of the primary disease. Less commonly, the deposits may be composed of heavy chains only (HCDD) or both ligh and heavy chains (LHCDD); both diseases have similar clinical characteristics to LCDD.

Monoclonal Gammopathy of Renal Significance In MGRS, a monoclonal immunoglobulin secreted by a nonmalignant or premalignant B cell or plasma cell clone results in monoclonal deposits in the kidney which can manifest as proteinuria and ESKD. Treatment is controversial but may include chemotherapy.

Renal Amyloidosis Most *renal amyloidosis* is either the result of primary fibrillar deposits of immunoglobulin light chains known as amyloid L (AL) or secondary to fibrillar deposits of serum amyloid A (AA) protein fragments **(Chap. 117)**. Hereditary amyloidosis is rare. Even though AA and AL amyloid occur for different reasons, their clinicopathophysiology is quite similar. Amyloid infiltrates the liver, heart, peripheral nerves, carpal tunnel, upper pharynx, and kidney, ultimately producing restrictive cardiomyopathy, hepatomegaly, macroglossia, and heavy proteinuria sometimes associated with renal vein thrombosis. In contrast to LCDD, amyloid kidney deposits are fibrillar, stain with Congo red, and contain predominantly the variable region of lambda chains **(see Fig. A4-19)**. In systemic AL amyloidosis, also called *primary amyloidosis,* light chains produced in excess by clonal plasma cell dyscrasias are made into fragments by macrophages that aggregate into amyloid fibrils. Approximately 10% of patients have overt myeloma as defined by CRAB (hypercalcemia, renal insufficiency, anemia, or lytic bone lesions). Nephrotic syndrome is common and ~20% of patients progress to dialysis. AA amyloidosis is sometimes called *secondary amyloidosis* and is due to deposition of β-pleated sheets of serum amyloid A protein, an acute phase reactant. Fragments of serum amyloid A protein increase and self-aggregate by attaching to receptors for advanced glycation end products in the extracellular environment. Patients with AA amyloid have associated inflammatory diseases including autoimmune diseases, chronic infections, and genetic autoinflammatory diseases. An increasing proportion of patients have unidentified chronic inflammation; this may reflect better treatments for the previously associated diseases or a rise in chronic inflammation due to obesity. Nephrotic syndrome is common, and ~40–60% of patient's progress to dialysis. Serum-free light chain analysis is useful in the early diagnosis and follow-up of disease progression. Biopsy of involved liver or kidney is diagnostic 90% of the time when the pretest probability is high; abdominal fat pad aspirates are positive ~70% of the time, but apparently less so when looking for AA amyloid. Amyloid deposits are distributed along blood vessels and in the mesangial regions of the kidney. The recommended treatment for primary amyloidosis is melphalan followed by autologous hematopoietic cell transplantation (HCT) which can achieve remission, however relapses are common. Patients who are not candidates for HCT often receive bortezomib-based regimens. Secondary amyloidosis is relentless unless the primary disease can be controlled. Drugs in development that disrupt the formation of fibrils may be available in the future.

Fibrillary and Immunotactoid Glomerulopathies Fibrillary and immunotactoid glomerulopathies are rare (<1.0% of kidney biopsies), morphologically defined diseases characterized by glomerular accumulation of nonbranching randomly arranged fibrils that are Congo red negative **(see Fig. A4-18)**. Fibrillary glomerulopathy accounts for 85–90% of cases and is identified by the presence of the protein DnaJ heat shock protein family member B9 (DNAJB9) in the glomeruli, which is absent in the rarer immunotactoid glomerulopathy. In both, glomerular and mesangial deposits contain oligoclonal or oligotypic immunoglobulins and complement, with 12- to 24-nm fibrils randomly arranged in fibrillary glomerulopathy and 16 to 52-nm fibrils organized into microtubules in immunotactoid glomerulopathy. The cause of this "nonamyloid" glomerulopathy is mostly idiopathic; reports of fibrillary glomerulonephritis describe associations with malignancy, autoimmune disease, and monoclonal gammopathy. Immunotactoid glomerulopathy has been associated with lymphoma or plasma cell disorders. Both disorders appear in adults aged 40–80 years old, with moderate to heavy proteinuria (100%), hematuria (70%), kidney injury (50%), a wide variety of histologic lesions, and ESKD within 2 to 6 years in 50% of patients. Most patients have disease limited to the kidney. Patients should be screened for associated disorders. There is no consensus on treatment of this uncommon disorder, although rituximab has been reported to remit proteinuria. These diseases can recur in the kidney transplant.

FABRY'S DISEASE

Fabry's disease is an X-linked inborn error of globotriaosylceramide metabolism secondary to deficient lysosomal α-galactosidase A (alpha-Gal A) activity, resulting in excessive intracellular storage of globotriaosylceramide. Affected organs include the vascular endothelium, heart, brain, and kidneys. Classically, Fabry's disease presents in childhood in males with acroparesthesias, angiokeratomas commonly in groin and periumbilical areas, abdominal pain, cornea verticillate, and hypohidrosis. Over time, male patients develop cardiac involvement, cerebrovascular disease, and kidney injury, with an average age of death around 50 years of age. Female heterozygotes with unfavorable X inactivation present with mild single-organ involvement or rarely severe manifestations including kidney failure but do so later in life than males. Kidney biopsy reveals enlarged glomerular visceral epithelial cells packed with small clear vacuoles containing globotriaosylceramide; vacuoles may also be found in parietal and tubular epithelia **(see Fig. A4-22)**. These vacuoles of electron-dense materials in parallel arrays (zebra bodies) are easily seen on electron microscopy. Ultimately, kidney biopsies reveal FSGS. The nephropathy of Fabry's disease typically presents in the third decade as mild to moderate proteinuria, sometimes with microscopic hematuria or nephrotic syndrome. Urinalysis may reveal oval fat bodies and birefringent glycolipid globules under polarized light (Maltese cross). Measurement of alpha-Gal A activity and mutational analysis of the gene is diagnostic, with kidney biopsies sometimes helpful. Progression to ESKD occurs by the fourth or fifth decade. Treatment with inhibitors of the renin-angiotensin system is recommended. Treatment with recombinant agalsidase alpha or beta or migalastat, a chaperone that facilitates trafficking of alpha-Gal A, clears microvascular endothelial deposits of globotriaosylceramide from the kidneys, heart, and skin. In patients with advanced organ involvement including chronic kidney disease, progression of disease occurs despite enzyme replacement therapy. Variable responses to enzyme therapy may be due to the occurrence of neutralizing antibodies or differences in uptake of the enzyme. Graft and patient survival following kidney transplantation in patients with Fabry's disease are similar to those of other causes of ESKD.

PULMONARY-RENAL SYNDROMES

Several diseases can present with catastrophic hemoptysis and glomerulonephritis associated with varying degrees of kidney injury. The usual causes include Goodpasture's syndrome, granulomatosis with polyangiitis, microscopic polyangiitis, Churg-Strauss vasculitis, and, rarely, Henoch-Schönlein purpura or cryoglobulinemia. Each of these diseases can also present without hemoptysis and are discussed in detail earlier in "Acute Nephritic Syndromes." **(See Glomerular Schematic 7.)** Pulmonary bleeding in this setting is life-threatening and often results in airway intubation, and acute kidney injury may require dialysis. Diagnosis is difficult initially because biopsies and serologic testing take time. Treatment with plasmapheresis and methylprednisolone is often empirical and temporizing until results of testing are available.

BASEMENT MEMBRANE SYNDROMES

All kidney epithelia, including podocytes, rest on basement membranes assembled into a planar surface through the interweaving of collagen IV with laminins, nidogen, and sulfated proteoglycans. Structural abnormalities in GBM associated with hematuria are characteristic of several familial disorders related to the expression of collagen IV genes. The extended family of collagen IV contains six chains, which are expressed in different tissues at different stages of embryonic development. All epithelial basement membranes early in human development are composed of interconnected triple-helical protomers rich in α1.α1.α2(IV) collagen. Some specialized tissues undergo a developmental switch replacing α1.α1.α2(IV) protomers with an α3.α4.α5(IV) collagen network; this switch occurs in the kidney (glomerular and tubular basement membrane), lung, testis, cochlea, and eye, while an α5.α5.α6(IV) network appears in skin, smooth muscle, and esophagus and along Bowman's capsule in the kidney. This switch probably occurs because the α3.α4.α5(IV) network is more resistant to proteases and ensures the structural longevity of critical tissues. When basement membranes are the target of glomerular disease, they produce moderate proteinuria, some hematuria, and progressive kidney failure.

ANTI-GBM DISEASE

Autoimmune disease where antibodies are directed against the α3 NC1 domain of collagen IV produces an *anti-GBM disease* often associated with RPGN and/or a pulmonary-renal syndrome called *Goodpasture's syndrome*. Discussion of this disease is covered earlier in "Acute Nephritic Syndromes."

ALPORT'S SYNDROME

Classically, patients with Alport's syndrome develop hematuria, and mild proteinuria (<1–2 g/24 h), which appears late in the course,

Glomerular schematic 7

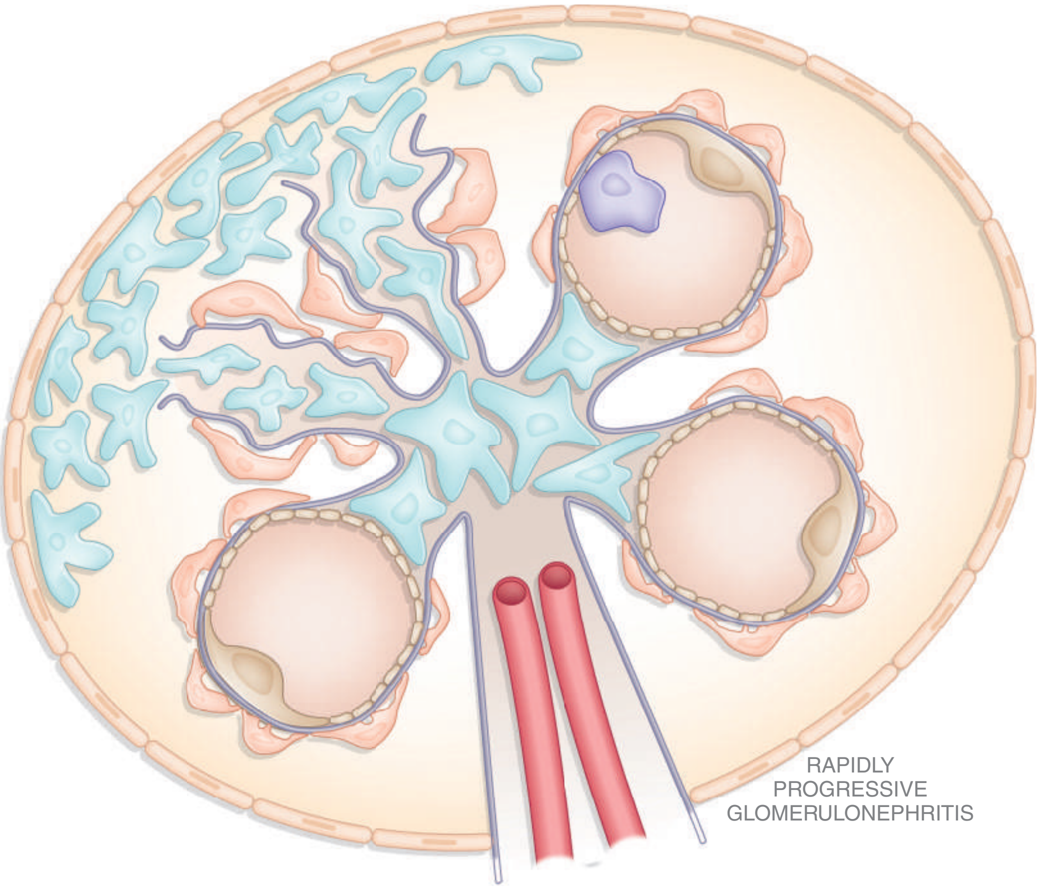

followed by chronic glomerulosclerosis leading to kidney failure in association with sensorineural deafness. Some patients develop lenticonus of the anterior lens capsule, "dot and fleck" retinopathy, and rarely, leiomyomatosis. Approximately 80–85% of patients with Alport's syndrome have an X-linked inheritance of mutations at the *COL4A5* locus affecting in the α5(IV) collagen chain on chromosome Xq22–24. Female carriers have variable penetrance depending on the type of mutation or the degree of mosaicism created by X inactivation. Mutations on chromosome 2q35-37 at the *COL4A3* and *COL4A4* loci of the α3(IV) and α4(IV) chains, respectively, are associated with autosomal recessive (AR) disease and less commonly autosomal dominant (AD) disease. With use of next generation sequencing, AD disease is now noted to likely occur more frequently than previously thought.

Pedigrees with the X-linked syndrome are quite variable in their rate and frequency of tissue damage leading to organ failure. There is strong correlation between genotype and phenotype regarding risk of disease progression in males; truncating variants, including large deletions and nonsense mutations, are associated with rapidly progressive disease with ESKD by age 30 in up to 90% of males. By contrast, those with missense or splice variants may not deteriorate until after the age of 30 with mild or late deafness. Early severe deafness, lenticonus, or proteinuria suggests a poorer prognosis. Usually females from X-linked pedigrees have only microhematuria, but up to 25% of carrier females have been reported to have more severe kidney manifestations. Pedigrees with the AR form of the disease have severe early disease in both females and males with asymptomatic parents.

Clinical evaluation should include a careful eye examination and hearing tests. However, the absence of extrarenal symptoms does not rule out the diagnosis. Genetic testing can be used for the diagnosis of Alport's syndrome and the demonstration of the mode of inheritance. In certain cases with high clinical suspicion or family history of Alport's, genetic testing alone may be sufficient for diagnosis. Since α5(IV) collagen is expressed in the skin, some X-linked Alport's patients can be diagnosed with a skin biopsy revealing the lack of the α5(IV) collagen chain on immunofluorescent analysis. Patients with mutations in α3(IV) or α4(IV) require a kidney biopsy. Early in their disease, Alport's patients typically have thin basement membranes on kidney biopsy **(see Fig. A4-23)**, which thicken over time into multilamellations surrounding lucent areas that often contain granules of varying density—the so-called split basement membrane. In any Alport's kidney, there are areas of thinning mixed with splitting of the GBM. Tubules drop out, glomeruli scar, and the kidney eventually succumbs to interstitial fibrosis. All affected members of a family with X-linked Alport's syndrome should be identified and followed, including mothers of affected males. Primary treatment is control of systemic hypertension and use of ACE inhibitors and possibly SGLT2i to slow kidney disease progression. Although patients who receive kidney allografts usually develop anti-GBM antibodies directed toward the collagen epitopes absent in their native kidney, overt Goodpasture's syndrome is rare and graft survival is good.

■ THIN BASEMENT MEMBRANE DISEASE

Thin basement membrane disease (TBMD), a relatively common disorder characterized by persistent or intermittent hematuria, which is usually microscopic hematuria and rarely macroscopic hematuria with flank pain. It is not typically associated with proteinuria, hypertension, or loss of kidney function or extrarenal disease. TBMD is often familial, with pedigrees exhibiting an autosomal dominant pattern. It usually presents in childhood in multiple family members and has also been called *benign familial hematuria*. Many cases of TBMD have genetic defects in type IV collagen; in contrast to Alport's syndrome, the disease behaves as an autosomal dominant disorder that in ~40% of families segregates with the *COL(IV)* α3/*COL(IV)* α4 loci. Mutations in these loci can result in a spectrum of disease, ranging from TBMD to autosomal dominant or recessive Alport's. The GBM shows diffuse thinning compared to normal values for the patient's age in otherwise normal biopsies **(see Fig. A4-23)**. The vast majority of patients have a benign course.

NAIL-PATELLA SYNDROME

Patients with nail-patella syndrome develop iliac horns on the pelvis and dysplasia of the dorsal limbs involving the patella, elbows, and nails, variably associated with neural-sensory hearing impairment, glaucoma, and abnormalities of the GBM and podocytes, leading to hematuria, proteinuria, and FSGS. The syndrome is autosomal dominant, with haploinsufficiency for the LIM homeodomain transcription factor LMX1B; pedigrees are extremely variable in the penetrance for all features of the disease. LMX1B regulates the expression of genes encoding α3 and α4 chains of collagen IV, interstitial type III collagen, podocin, and CD2AP that help form the slit-pore membranes connecting podocytes. Mutations in the LIM domain region of *LMX1B* associate with glomerulopathy in 30–40% of patients and rarely progress to ESKD. Proteinuria or isolated hematuria is discovered throughout life but usually by the third decade. Genetic testing can confirm the diagnosis. Treatment is nonspecific, but renin-angiotensin system inhibition is recommended. Patients with ESKD do well with transplantation.

GLOMERULAR-VASCULAR SYNDROMES

A variety of diseases result in classic vascular injury to the glomerular capillaries. Most of these processes also damage blood vessels elsewhere in the body. The group of diseases discussed here lead to vasculitis, renal endothelial injury, thrombosis, ischemia, and/or lipid-based occlusions.

ATHEROSCLEROTIC NEPHROPATHY

Aging in the developed world is commonly associated with the occlusion of coronary and systemic blood vessels. When the renal arterial circulation is involved, the glomerular microcirculation is damaged, leading to *chronic nephrosclerosis*. Several aggressive lipid disorders can accelerate this process, but most of the time, atherosclerotic progression to chronic nephrosclerosis is associated with poorly controlled hypertension.

HYPERTENSIVE NEPHROSCLEROSIS

Systemic hypertension causes permanent damage to the kidneys in ~6% of patients with elevated blood pressure. As many as 27% of patients with ESKD have hypertension as a primary cause, and it is the second most common cause of ESKD after diabetic nephropathy in the US. Risk alleles associated with *APOL1,* a functional gene for apolipoprotein L1 expressed in podocytes are associated with an increased risk of ESKD. Other associated risk factors for progression to end-stage kidney disease include increased age, male gender, smoking, hypercholesterolemia, duration of hypertension, low birth weight, and preexisting kidney injury. Kidney biopsies in patients with hypertension, microhematuria, and moderate proteinuria demonstrate arteriolosclerosis, chronic nephrosclerosis, and interstitial fibrosis in the absence of immune deposits **(see Fig. A4-25)**. Based on a careful history, physical examination, urinalysis, and some serologic testing, the diagnosis of chronic nephrosclerosis is usually inferred without a biopsy. Recent studies suggest, in the absence of diabetes, adults with hypertension and cardiovascular risk factors benefit from achieving a systolic blood pressure <120 mmHg, compared to <140 mmHg. In the presence of kidney disease, most patients begin antihypertensive therapy with two drugs, classically a thiazide diuretic and an ACE inhibitor; most will require three drugs. There is strong evidence in a study with self-identified African Americans with hypertensive nephrosclerosis that therapy initiated with an ACE inhibitor can slow the rate of decline in kidney function independent of effects on systemic blood pressure. Malignant acceleration of hypertension complicates the course of chronic nephrosclerosis, particularly in the setting of scleroderma or cocaine use **(see Fig. A4-28)**. The hemodynamic stress of malignant hypertension leads to fibrinoid necrosis of small blood vessels, thrombotic microangiography, a nephritic urinalysis, and AKI. In the setting of kidney injury, chest pain, or papilledema, the condition is treated as a hypertensive emergency.

CHOLESTEROL EMBOLI

Aging patients with clinical complications from atherosclerosis sometimes shower cholesterol crystals into the circulation following an endovascular procedure with manipulation of the aorta or with use of systemic anticoagulation. Less commonly, spontaneous emboli can occur and may shower acutely or shower subacutely which is somewhat more silently. Irregular emboli trapped in the microcirculation produce ischemic damage that induces an inflammatory reaction. Depending on the location of the atherosclerotic plaques releasing these cholesterol fragments, one may see cerebral transient ischemic attacks; livedo reticularis in the lower extremities; Hollenhorst plaques in the retina with visual field cuts; necrosis of the toes; and acute glomerular capillary injury leading to FSGS sometimes associated with hematuria, mild proteinuria, and loss of kidney function, which typically progresses over a few years. Occasional patients have fever, eosinophilia, or eosinophiluria. A skin biopsy of an involved area may be diagnostic. Since tissue fixation dissolves the cholesterol, one typically sees only residual, biconvex clefts in involved vessels **(see Fig. A4-26)**. There is no therapy to reverse embolic occlusions, and steroids do not help. Controlling blood pressure and lipids and cessation of smoking are usually recommended for prevention.

SICKLE CELL DISEASE

Sickle cell disease (SCD) is an autosomal recessive disease that occurs due to a mutation in the hemoglobin β-chain (HβS). Clinical disease occurs in homozygous patients (SS) or heterozygous patients in the presence of an abnormal or missing β-chain (SC, HbSβ). Under certain circumstances (i.e., hypoxia, hypovolemia, acidity, hyperosmolality), HbS polymerizes which causes the red blood cell shape to distort. These cells attach to endothelia, causing obstruction as well as other changes to the vasculature that produce acute and chronic vaso-occlusion disease in many organs, including the kidney. The major site of injury in the kidney is the renal medulla, an area supplied by the vasa recta capillaries that promotes sickling due to the relative hypoxic, acidotic, and hypertonic environment. Repeated injury overtime produces chronic organ damage leading to sickle cell nephropathy (SCN). Early changes of SCN include glomerular hyperfiltration, albuminuria, and both micro- and macro-hematuria. Nearly all SCD patients have concentrating defects which cause hyposthenuria and can lead to severe dehydration. Later changes include papillary necrosis, renal infarction, interstitial nephritis, proteinuria, and FSGS. Rarely patients may present with MPGN. Several genetic risk factors have been identified that may increase risk for progression of CKD, including *APOL1* gene variants. Treatment is directed at reducing the frequency of vaso-occlusive events and administering ACE inhibitors/ARBs and hydroxyurea in the hope of delaying a progressive decline in kidney function. Management of anemia in SCN patients is complex and may require high doses of erythropoiesis-stimulating agents. A number of patients will develop progressive CKD, and one study found that 20% of SCD patients develop ESKD before the age of 30. SCD patients with ESKD have poor prognosis on kidney replacement therapy (KRT), with a mean time to death of 4 years. Prognosis for ESKD patients improves after kidney transplant with 7 year survival of 67%.

Sickle cell trait occurs in patients with one HbS and one normal hemoglobin. Although individuals usually do not experience vaso-occlusive symptoms, most will gradually develop hyposthenuria due to subclinical infarction of the renal medulla and a consequent loss of concentrating ability. They may experience hematuria and are at higher risk for renal medullary carcinoma than SCD patients.

THROMBOTIC MICROANGIOPATHIES

Thrombotic microangiopathy (TMA) refers to a pathologic lesion that causes thrombocytopenia and microangiopathic hemolytic anemia with schistocytes. *Thrombotic thrombocytopenic purpura* (TTP), Shiga toxin–mediated *hemolytic-uremic syndrome* (HUS), and complement-mediated HUS represent a spectrum of primary TMAs that share these features and may have concurrent fever, kidney failure, and neurologic disturbances. HUS is suspected with patients have more severe kidney injury while TTP is suspected in adults with neurologic disease and more severe thrombocytopenia. On examination of kidney tissue, there is evidence of *glomerular capillary endotheliosis* associated with platelet thrombi, damage to the capillary wall, and formation of fibrin

TABLE 326-6 Thrombotic Microangiopathies

- Primary thrombotic microangiopathy
  - TTP
  - Shiga-toxin HUS
  - Complement-mediated HUS
- Secondary thrombotic microangiopathy
  - Pregnancy related: preeclampsia, HELLP (*h*emolysis, *e*levated *l*iver enzymes, and *l*ow *p*latelet count syndrome), postpartum (thought to be complement mediated)
  - Drug induced: oral contraceptives or quinine, calcineurin inhibitors, antiplatelet agents (ticlopidine and clopidogrel), drugs of abuse (cocaine, IV use of oxycodone)
  - Kidney transplant patients given OKT3 for rejection
  - Malignant hypertension
  - Autoimmune: antiphospholipid syndrome, lupus, scleroderma
  - Infections: HIV, pneumococcal, CMV
  - Cobalamin deficiency

Abbreviations: CMV, cytomegalovirus; HUS, hemolytic-uremic syndrome; TTP, thrombotic thrombocytopenic purpura.

material in and around glomeruli **(see Fig. A4-27)**. These tissue findings are similar to what is seen in secondary TMA, which includes a broad group of conditions known to be associated with TMA **(see Table 326-6)**.

Shiga toxin–mediated HUS is caused by a toxin released by *Escherichia coli* 0157:H7 and occasionally by *Shigella dysenteriae*. This Shiga toxin (verotoxin) directly injures endothelia, enterocytes, and kidney cells, causing apoptosis, platelet clumping, and intravascular hemolysis by binding to the glycolipid receptors (Gb3). These receptors are more abundant along endothelia in children compared to adults. Shiga toxin also inhibits the endothelial production of ADAMTS13. In familial cases of adult TTP, there is a genetic deficiency of the ADAMTS13 metalloprotease that cleaves large multimers of von Willebrand's factor (VWF). In the absence of ADAMTS13, these large multimers cause platelet clumping and intravascular hemolysis. An antibody to ADAMTS13 is found in many sporadic cases of adult TTP. Patients can be tested for ADAMTS13 activity, and if low, the presence of antibodies to ADAMTS13 distinguishes the deficiency from the immune-mediated disease. Complement-mediated HUS, previously referred to as atypical HUS, is thought to occur when there is a hereditary deficiency or antibody to a regulatory protein in the alternative complement pathway leading to overactivation. The treatment of adult TTP with ADAMTS13 antibodies is daily plasmapheresis, which can be lifesaving. Plasmapheresis with fresh frozen plasma is given until the platelet count rises, or longer for relapsing patients. There is an anecdotal role in relapsing patients for splenectomy. Refractory or relapsing patients may benefit from steroids, immunosuppressive drugs such as rituximab, or caplacizumab, a monoclonal antibody that blocks interaction between VWF and platelets. In the absence of ADAMTS13 antibodies, patients with a genetic deficiency of ADAMTS13 production can be treated with fresh frozen plasma alone. Patients with Shiga toxin–mediated HUS are treated primarily with supportive care because antibiotics are thought to accelerate the release of the toxin and the diarrhea is usually self-limited. Patients with complement-mediated HUS are treated with anticomplement therapy, such as eculizumab or ravulizumab.

ANTIPHOSPHOLIPID ANTIBODY SYNDROME (SEE CHAP. 369)

GLOBAL CONSIDERATIONS

INFECTIOUS DISEASE–ASSOCIATED SYNDROMES

A number of infectious diseases will injure the glomerular capillaries as part of a systemic reaction producing an immune response or from direct infection of kidney tissue. Evidence of this immune response is collected by glomeruli in the form of immune deposits that damage the kidney, producing moderate proteinuria and hematuria. A high prevalence of many of these infectious diseases in developing countries results in infection-associated kidney disease being the most common cause of glomerulonephritis in many parts of the world.

Poststreptococcal Glomerulonephritis This form of glomerulonephritis is one of the classic complications of streptococcal infection. The discussion of this disease can be found earlier, in the section "Acute Nephritic Syndromes."

Subacute Bacterial Endocarditis Kidney injury from persistent bacteremia absent the continued presence of a foreign body, regardless of cause, is treated presumptively as if the patient has endocarditis. The discussion of this disease can be found earlier, in the section "Acute Nephritic Syndromes."

Human Immunodeficiency Virus Kidney disease is an important complication of HIV disease. About 50% of HIV-infected patients with kidney disease have HIV-associated nephropathy (HIVAN) on biopsy. The lesion in HIVAN is FSGS, characteristically revealing a collapsing glomerulopathy **(see Fig. A4-3)** with visceral epithelial cell swelling, microcystic dilatation of renal tubules, and tubuloreticular inclusion. Renal epithelial cells express replicating HIV virus, but host immune responses also play a role in the pathogenesis. HIVAN is thought to be linked to *APOL1* risk variants. HIV immune complex kidney disease (HIVICK) is a group of immune complex–mediated glomerular lesions seen in HIV patients that, on biopsy, can look like a constellation of other glomerular lesions, including postinfectious glomerulonephritis, MGN, MPGN, DPGN, MCD, and IgA nephropathy. The HIVICK effect is a complication of active HIV viremia.

HIV patients with FSGS typically present with nephrotic-range proteinuria and hypoalbuminemia, but unlike patients with other etiologies for nephrotic syndrome, they do not commonly have hypertension, edema, or hyperlipidemia. Kidney ultrasound also reveals large, echogenic kidneys despite the finding that kidney function in some patients declines rapidly. Treatment with inhibitors of the renin-angiotensin system decreases the proteinuria. Effective antiretroviral therapy benefits both the patient and the kidney and improves survival of HIV-infected patients with HIVAN and, in some cases, HIVICK-associated chronic kidney disease or ESKD. In HIV-infected patients not yet on therapy, the presence of HIVAN is an indication to initiate therapy. Following the introduction of antiretroviral therapy, survival on dialysis for the HIV-infected patient has improved dramatically. Kidney transplantations in HIV-infected patients without detectable viral loads or histories of opportunistic infections provide a better survival benefit over dialysis. Following transplantation, patient and graft survival are similar to the general transplant population despite frequent rejections.

Hepatitis B and C Typically, infected patients present with microscopic hematuria, nonnephrotic or nephrotic-range proteinuria, and hypertension. There is a close association between hepatitis B infection and polyarteritis nodosa, with vasculitis appearing generally in the first 6 months following infection. Kidney manifestations include renal artery aneurysms, renal infarction, and ischemic scars. Alternatively, the hepatitis B carrier state can produce an MGN with predominant IgG1 deposition that is more common in children than adults or MPGN that is more common in adults than in children. Kidney histology is indistinguishable from idiopathic MGN or MPGN. Viral antigens, most commonly HBeAG, are found in the kidney deposits. Cryoglobulinemic glomerulonephritis has also been reported. Treatment is with antiviral agents. Children have a better prognosis than adults.

Up to 30% of patients with chronic hepatitis C infection have some kidney manifestations. Patients often present with type II mixed cryoglobulinemia, nephrotic syndrome, microscopic hematuria, abnormal liver function tests, depressed C_3 levels, anti–hepatitis C virus (HCV) antibodies, and viral RNA in the blood. The lesions most commonly seen, in order of decreasing frequency, are cryoglobulinemic glomerulonephritis, MGN, and MPGN, but polyarteritis nodosa (PAN), IgA nephropathy, and FSGS have been reported. With the availability of direct-acting antivirals, which can achieve a viral remission in >95% of patients, the prevalence of glomerular disease in HCV patients should decline. These drugs are currently the treatment of choice for patients with HCV-related MPGN or PAN.

SARS-CoV-2 The novel coronavirus, SARS-CoV-2, is associated with several complications related to kidney disease including acute kidney injury, chronic kidney disease, and, rarely, glomerular diseases. The glomerular lesions reported in patients with COVID-19 include FSGS, minimal change disease, membranous nephropathy, ANCA-associated vasculitis, anti-GBM disease, IgA nephropathy and thrombotic microangiopathy. The glomerular disease most commonly associated with COVID-19 is collapsing glomerulopathy, a morphologic variant of FSGS, also referred to as COVID-associated nephropathy (COVAN). Patients may present with new onset nephrotic syndrome or nephrotic range proteinuria with AKI. There have been case reports of *de novo* glomerular disease and relapse of pre-existing glomerular disease following administration of COVID-19 mRNA vaccines, although evidence for causal link is not well established. The use of immunosuppression therapy may be considered in patients with severe nephrotic syndrome or persistent disease despite resolution of infection, however there is limited data to guide treatment recommendations.

Other Viruses Other viral infections are occasionally associated with glomerular lesions, but cause and effect are not well established. These viral infections and their respective glomerular lesions include cytomegalovirus producing MPGN or FSGS; influenza and anti-GBM disease; measles-associated endocapillary proliferative glomerulonephritis, with measles antigen in the capillary loops and mesangium; parvovirus causing mild proliferative or mesangioproliferative glomerulonephritis or FSGS; mumps and mesangioproliferative glomerulonephritis; Epstein-Barr virus producing MPGN, diffuse proliferative nephritis, or IgA nephropathy; dengue hemorrhagic fever causing endocapillary proliferative glomerulonephritis; Hanta virus and mesangial proliferative glomerulonephritis; and coxsackievirus producing *focal glomerulonephritis* or DPGN.

Syphilis Secondary syphilis, with rash and constitutional symptoms, develops weeks to months after the chancre first appears and occasionally presents with the nephrotic syndrome from MGN caused by subepithelial immune deposits containing treponemal antigens. Neuron-derived neurotrophic factor has also been identified as an antigenic target. Other lesions have also rarely been described, including interstitial syphilitic nephritis. The diagnosis is confirmed with nontreponemal and treponemal tests for *Treponema pallidum*. The kidney lesion responds to treatment with penicillin or an alternative drug, if allergic. Additional testing for other sexually transmitted diseases is an important part of disease management.

Leprosy Despite aggressive eradication programs, new cases of leprosy appear primarily in developing countries. The diagnosis is best made in patients with multiple skin lesions accompanied by sensory loss in affected areas, using skin smears showing paucibacillary or multibacillary infection (WHO criteria). Leprosy is caused by infection with *Mycobacterium leprae* and can be classified by Ridley-Jopling criteria into various types: tuberculoid, borderline tuberculoid, mid-borderline and borderline lepromatous, and lepromatous. Kidney involvement in leprosy is related to the quantity of bacilli in the body, and the kidney is one of the target organs during splanchnic localization. In some series, all cases with borderline lepromatous and lepromatous types of leprosy have various forms of kidney involvement including FSGS, mesangioproliferative glomerulonephritis, or renal amyloidosis; much less common are DPGN and MPGN. Treatment of the infection with multidrug therapy can reduce the incidence of kidney disease or produce remission of the kidney disease.

Malaria There are 300–500 million incident cases of malaria each year worldwide, and the kidney is commonly involved. Glomerulonephritis is due to immune complexes containing malarial antigens that are implanted in the glomerulus. In malaria from *P. falciparum*, mild proteinuria is associated with subendothelial deposits, mesangial deposits, and mesangioproliferative glomerulonephritis that usually resolve with treatment. In quartan malaria from infection with *Plasmodium malariae*, children are more commonly affected and kidney involvement is more severe. Transient proteinuria and microscopic hematuria can resolve with treatment of the infection. However, resistant nephrotic syndrome with progression to ESKD over 3–5 years does happen, as <50% of patients respond to steroid therapy. Affected patients with nephrotic syndrome have thickening of the glomerular capillary walls, with subendothelial deposits of IgG, IgM, and C_3 associated with a sparse membranoproliferative lesion. The rare mesangioproliferative glomerulonephritis reported with *Plasmodium vivax* or *Plasmodium ovale* typically has a benign course. Acute kidney injury can often complicate these glomerulopathies.

Schistosomiasis Schistosomiasis affects >300 million people worldwide and primarily involves the urinary and gastrointestinal tracts. Glomerular involvement varies with the specific strain of schistosomiasis; *Schistosoma mansoni* is most commonly associated with clinical kidney disease, and the glomerular lesions can be classified as follows: class I is a *mesangioproliferative glomerulonephritis*; class II is an *extracapillary proliferative glomerulonephritis*; class III is a *membranoproliferative glomerulonephritis*; class IV is a *focal segmental glomerulonephritis*; and class V is *amyloidosis*. Classes I–II often remit with treatment of the infection, but class III and IV lesions are associated with IgA immune deposits and progress despite antiparasitic and/or immunosuppressive therapy.

Other Parasites Kidney involvement with toxoplasmosis infections is rare. When it occurs, patients present with nephrotic syndrome and have a histologic picture of MPGN. Fifty percent of patients with leishmaniasis will have mild to moderate proteinuria and microscopic hematuria, but decreased GFR is rare. Acute DPGN, MGN, and mesangioproliferative glomerulonephritis have all been observed on biopsy. Filariasis and trichinosis are caused by nematodes and are sometimes associated with glomerular injury presenting with proteinuria, hematuria, and a variety of histologic lesions that typically resolve with eradication of the infection.

Acknowledgment

The authors wish to thank Nicole Wyatt for her assistance with this chapter.

FURTHER READING

De Vriese AS et al: Differentiating primary, genetic, and secondary FSGS in adults: A clinicopathologic approach. J Am Soc Nephrol 29:759, 2018.

Kupin WL: Viral-associated GN: Hepatitis C and HIV. Clin J Am Soc Nephrol 12:1337, 2017.

Hou J et al: C3 glomerulopathy: a review with emphasis on ultrastructural features. Glomerular Dis 2:107, 2022.

Lebleu VS et al: Origin and functional heterogeneity of fibroblasts. FASEB J 34(3):3519-, 2020.

Miller P, Caza T: The expanding spectrum and utility of antigens in membranous nephropathology. Curr Opin Nephrol Hypertens 32:232, 2023.

Papazachariou L et al: Frequent COL4 mutations in familial microheamaturia accompanied by later-onset/alport nephropathy due to focal segmental glomerulosclerosis. Clin Genet 92:517, 2017.

Ronco P et al: Membranous nephropathy. Nat Rev Dis Primers 7:69, 2021.

Sethi S et al: Mayo Clinic/Renal Pathology Society consensus report on pathologic classification, diagnosis, and reporting of GN. J Am Soc Nephrol 27:1278, 2016.

Smith RJH et al: C3 glomerulopathy - understanding a rare complement-driven renal disease. Nat Rev Nephrol 15:129, 2019.

Yau K et al: Prescribing SGLT2 inhibitors in patients with CKD: Expanding indications and practical considerations. Kidney Int Rep 7:1463, 2022.

327 Polycystic Kidney Disease and Other Inherited Disorders of Tubule Growth and Development

Jing Zhou, Martin R. Pollak

The polycystic kidney diseases are a group of genetically heterogeneous disorders and a leading cause of kidney failure. The autosomal dominant form of polycystic kidney disease (ADPKD) is the most common life-threatening monogenic disease, affecting 12 million people worldwide. The autosomal recessive form of polycystic kidney disease (ARPKD) is rarer but affects the pediatric population. Kidney cysts are often seen in a wide range of syndromic diseases. Recent studies have shown that defects in the structure or function of the primary cilia may underlie this group of genetic diseases collectively termed *ciliopathies* (Table 327-1).

AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE

Etiology and Pathogenesis (Fig. 327-1) ADPKD is characterized by progressive formation of epithelial lined cysts in the kidney. Although cysts occur in only 5% of the tubules in the kidney, the enormous growth of these cysts ultimately leads to the loss of normal surrounding tissues and loss of renal function. The cellular defects in ADPKD that have been known for a long time are increased cell proliferation and fluid secretion, decreased cell differentiation, and abnormal extracellular matrix. ADPKD is caused by mutations in *PKD1* and *PKD2*, which, respectively, code for polycystin-1 (PC1) and polycystin-2 (PC2). PC1 is a large 11-transmembrane protein that functions like a G protein–coupled receptor. PC2 is a calcium-permeable six-transmembrane protein that structurally belongs to the transient receptor potential (TRP) cation channel family. PC1 and PC2 are widely expressed in almost all tissues and organs. PC1 expression is high in development and low in the adult, whereas PC2 expression is relatively constant. PC1/2 are found on the primary cilium, a hair-like structure present on the apical membrane of a cell, in addition to the cell membranes and cell–cell junctions of tubular epithelial cells. Defects in the primary cilia are linked to a wide spectrum of human diseases, collectively termed *ciliopathies*. The most common phenotype shared by many ciliopathies is kidney cysts. PC1 and PC2 bind to each other via their respective C-terminal tails to form a receptor-channel complex and regulate each other's function. Recent evidence suggests a 1:3 stoichiometry for PC1:PC2 in the PC1/2 channel complex. The PC1/2 protein complex serves as a mechanosensor or chemical sensor and regulates calcium and G-protein signaling. The PC1/2 protein complex may also directly regulate a number of cellular functions, including the cell cycle, the actin cytoskeleton, planar cell polarity (PCP), and cell migration. This protein complex has also

TABLE 327-1 Inherited Diseases Commonly Associated with a Cystic Phenotype

DISEASE	MODE OF INHERITANCE	RENAL ABNORMALITIES	OTHER CLINICAL FEATURES	GENES
Autosomal dominant polycystic kidney disease	AD	Bilaterally enlarged kidneys with cortical and medullary cysts	Liver, pancreatic cysts, hypertension, subarachnoid hemorrhage	*PKD1, PKD2*
Autosomal dominant polycystic kidney disease-like	AD	Normal to smaller sized kidneys with fewer cortical and medullary cysts	Liver cysts at variable degree (from absent to severe)	*GANAB, DNAJB11*
Autosomal recessive polycystic kidney disease	AR	Distal and collecting duct cysts	Oligohydramnios if severe, hypertension, ascending cholangitis, liver fibrosis	*PKHD1*
Autosomal dominant tubulointerstitial kidney disease	AD	Small fibrotic kidneys; medullary cysts	In adults, gout	*UMOD, MUC1, REN, HNF1β, SEC61A1*
Renal cysts and diabetes syndrome	AD	Kidney cysts, aberrant nephrogenesis, irregular collecting systems, abnormal renal calyces, hyperuricemic nephropathy. Highly variable.	Diabetes	*HNF1B*
Nephronophthisis	AR	Small fibrotic kidneys; medullary cysts	Growth retardation, anemia (In syndromic forms: visual loss, liver fibrosis, cerebellar ataxia, other)	*NPHP1-20, IQCB1, CEP290, GLIS2, RPGRIP1L, NEK8, SDCCAG8, TMEM67, TTC21B*
Senior-Loken syndrome	AR	Renal cysts	Juvenile nephronophthisis, Leber amaurosis	*NPHP1-6, SDCCAG8*
Leber congenital amaurosis	AR	Renal cysts	Visual impairment in first year of life; pigmentary retinopathy	*GUCY2D, RPE65, LCA3-14*
Meckel-Gruber syndrome	AR	Cortical and medullary cysts	CNS anomalies, polydactyly, congenital heart defects	*MKS1, TMEM216, TMEM67, TMEM231, TMEM107, CEP290, RPGRIP1L, CC2D2A, TCTN2, B9D1, B9D2, NPHP3, KIF14*
Bardet-Biedl syndrome	AR	Renal cysts	Obesity, polydactyly, retinitis pigmentosa, anosmia, congenital heart defects, intellectual disability	*BBS1, 2, ARL6, BBS4,5, MKKS, BBS7, TTC8, BBS9, 10, TRIM32, BBS12, MKS1, CEP290, C2ORF86*
Oral-facial-digital syndrome type I	X-linked dominant	Renal cysts	Oral cavity, face, and digit anomalies; CNS abnormalities; cystic kidney disease; X-linked with male lethality, primary ciliary dyskinesia	*OFD1*
Tuberous sclerosis	AD	Renal cysts	Angiomyolipomas; renal cell carcinoma Facial angiofibromas; CNS hamartomas	*TSC1, TSC2*
Von Hippel-Lindau disease	AD	Renal cysts	Renal cell carcinoma, retinal angiomas; CNS hemangioblastomas; pheochromocytomas	*VHL*

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; CNS, central nervous system.

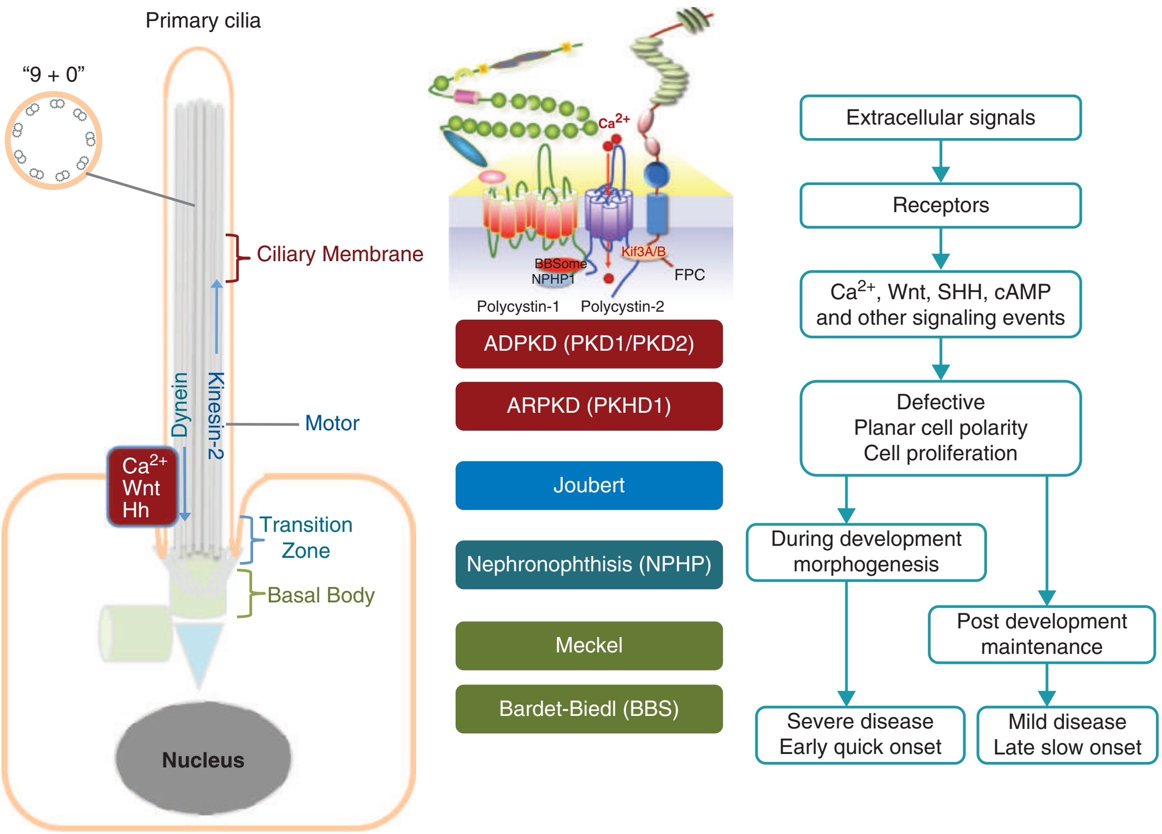

FIGURE 327-1 Scheme of the primary cilium and cystic kidney disease proteins. ***Left:*** A scheme of the primary cilium. Primary cilia share a "9+0" organization of microtubule doublets. Proteins are transported into the cilium by motor protein kinesin 2 and transported out of the cilium by dynein. The cilium is connected to the basal body through the transition zone. ***Middle:*** Topology of ADPKD and ARPKD proteins polycystin 1, polycystin 2, and FPC is shown. Localizations of disease proteins in the cilium, the transition zone, and the basal body are color coded. ***Right:*** Potential disease mechanisms due to cilium-mediated signaling events.

been implicated in regulating a number of signaling pathways, including Wnt, mammalian target of rapamycin (mTOR), STAT3, cMET, phosphoinositide 3-kinase (PI3K/Akt), G protein–coupled receptor (GPCR), and epidermal growth factor receptor (EGFR), as well as in the localization and activity of cystic fibrosis transmembrane conductance regulator (CFTR). One hypothesis is that loss of ciliary function of PC1 and PC2 leads to aberrant calcium signaling and a subsequent increase of adenylyl cyclase activity and decrease of phosphodiesterase activity, which, in turn, causes increased cellular cAMP. Increased cAMP promotes protein kinase A activity, among other effectors, and, in turn, leads to cyst growth by promoting proliferation and fluid secretion of cyst-lining cells through chloride and aquaporin channels in ADPKD kidneys.

ADPKD is inherited as an autosomal dominant trait with complete penetrance but variable expressivity. The disease affects all ethnic groups worldwide with an estimated prevalence of 1:1000 to 1:400. Only half of the patients with ADPKD are clinically diagnosed during their lifetimes. ADPKD is genetically heterogeneous. The first disease gene (*PKD1*) was localized to the region of the alpha-globin gene on chromosome 16p13 in 1985, and a second disease gene (*PKD2*) locus was mapped to chromosome 4q21-q23 in 1993. Mutations of *PKD1* and *PKD2* are responsible for ~85% and ~15% of ADPKD cases, respectively. However, patients with *PKD2* mutations may be >15% because they tend to have milder clinical disease and, as a result, are underdiagnosed. Embryonic lethality of *Pkd1* and *Pkd2* knockout mice suggests human homozygotes may be lethal and thus not clinically recognized.

PKD1 is comprised of 46 exons occupying ~52 kb of genomic DNA. It produces a ~14 kb transcript that encodes polycystin-1, a protein of ~4300 amino acids. A feature of the *PKD1* gene is that the 5′ three-quarters of *PKD1* have been duplicated at six other sites on chromosome 16p, and many of them produce mRNA transcripts, which provides a major challenge for genetic analysis of the duplicated region. *PKD2* is a single-copy gene with 15 exons producing a ~5.3 kb mRNA transcript that encodes polycystin-2, a protein of 968 amino acids. Two additional genes, *GANAB* and *DNAJB11*, have been found in patients with autosomal dominant form of polycystic kidney disease. The *GANAB* gene encodes the glucosidase IIa subunit, and the *DNAJB11* gene produces a cofactor of BiP, a key chaperone in the endoplasmic reticulum controlling folding, trafficking, and degradation of secreted and membrane proteins. Both proteins appear to affect PC1 trafficking. However, these mutations have only been found in a very small number of families.

In ADPKD patients, every cell carries a germline mutant allele of either *PKD1* or *PKD2*. However, cysts develop in only a small fraction of the nephrons. Cysts are thought to originate from clonal growth of single cells that have received a somatic "second hit" mutation in the "normal" allele of the *PKD1* or *PKD2* gene. Accumulating evidence in mouse models now shows that partial loss of function of the second allele of *Pkd1* in a proliferative environment is sufficient for cystogenesis, suggesting that a critical amount of *PKD1* is needed in a cell. Somatic inactivation of the second allele of *Pkd1* in adult mice results in very slow onset of cyst development in the kidney, but a "third hit," such as an additional genetic or epigenetic event, the inactivation of a growth suppressor gene, the activation of a growth promoting gene(s), or an event such as renal injury that activates the developmental program, may promote rapid cyst formation.

Clinical Manifestations ADPKD is characterized by the progressive bilateral formation of renal cysts. Focal renal cysts are typically detected in affected subjects aged <30 years. Hundreds to thousands of cysts are usually present in the kidneys of most patients in the fifth decade (**Fig. 327-2**). Enlarged kidneys can each reach a fourfold increase in length and weigh up to 20 times the normal weight. The clinical presentations of ADPKD are highly variable. While many patients are asymptomatic until the fourth to fifth decade of life and

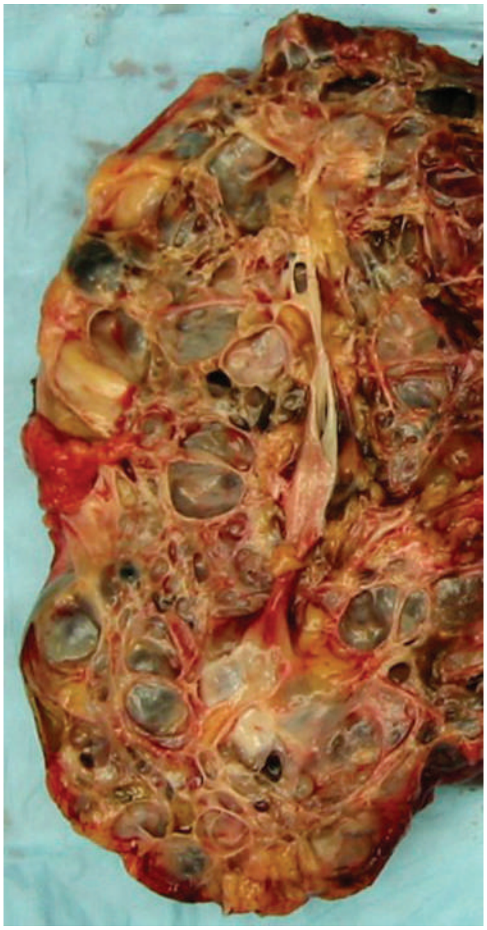

FIGURE 327-2 Photograph showing a kidney from a patient with autosomal dominant polycystic kidney disease. The kidney has been cut open to expose the parenchyma and internal aspects of cysts.

are diagnosed by incidental discoveries of hypertension or abdominal masses, back or flank pain is a frequent symptom in ~60% of patients with ADPKD. The pain may result from renal cyst infection, hemorrhage, or nephrolithiasis. Gross hematuria resulting from cyst rupture occurs in ~40% of patients during the course of their disease, and many of them will have recurrent episodes. Flank pain and hematuria may coexist if the cyst that ruptures is connected with the collecting system. Proteinuria is usually a minor feature of ADPKD. Infection is the second most common cause of death for patients with ADPKD. Up to half of patients with ADPKD will have one or more episodes of renal infection during their lifetimes. An infected cyst and acute pyelonephritis are the most common renal infections often due to gram-negative bacteria, which are associated with fever and flank pain, with or without bacteremia. These complications and renal insufficiency often correlate with structural abnormality of the renal parenchyma. Mutations in *GANAB* and *DNAJB11* genes result in milder cystic kidney disease than that in classic ADPKD with small renal cysts and normal-sized kidneys. Sometimes patients with *GANAB* mutations present with autosomal dominant polycystic liver disease (ADPLD)-like phenotype. Patients with *DNAJB11* develop renal fibrosis, characteristic of autosomal dominant tubulointerstitial disease (ADTKD), discussed below. Kidney stones occur in ~20% of patients with ADPKD. Different from the general population, more than half of the stones in patients with ADPKD are composed of uric acid, with the remainder due to calcium oxalate. Distal acidification defects, abnormal ammonium transport, low urine pH, and hypocitraturia may be important in the pathogenesis of renal stones in ADPKD. Renal cell carcinoma is a rare complication of ADPKD with no apparent increased frequency compared to the general population. However, in ADPKD, these tumors are more often bilateral at presentation, multicentric, and sarcomatoid in type. Radiologic imaging is often not helpful in distinguishing cyst infection and cyst hemorrhage because of their complexity. Computed tomography (CT) scan and magnetic resonance imaging (MRI) are often useful in distinguishing a malignancy from a complex cyst. Cardiovascular complications are the major cause of mortality in patients with ADPKD. Hypertension is common and typically occurs before any reduction in glomerular filtration rate (GFR). Hypertension is a risk factor for both cardiovascular and kidney disease progression in ADPKD. Notably, some normotensive patients with ADPKD may also have left ventricular hypertrophy. Hypertension in ADPKD may result from the increased activation of the renin-angiotensin-aldosterone system, increased sympathetic nerve activity, and impaired endothelial cilium function–dependent relaxation of small resistant blood vessels.

The progression of ADPKD has striking inter- and intrafamilial variability. The disease can present as early as in utero, but end-stage renal disease (ESRD) typically occurs in late middle age. Risk factors include early diagnosis of ADPKD, hypertension, gross hematuria, multiple pregnancies, and large kidney size. Liver cysts derived from the biliary epithelia are the most common extrarenal complication. Polycystic liver disease associated with ADPKD is different from ADPLD, which is caused by mutations in at least two distinct genes (*PRKCSH* and *SEC63*) and does not progress to renal failure. Massive polycystic liver disease occurs almost exclusively in women with ADPKD, particularly those with multiple pregnancies. Heterozygous loss-of-function variants in *PKHD1, ALG8, GANAB*, and *SEC61B* are now found in ADPLD. *ALG8, GANAB,* and *SEC61B* all encode ER proteins that are involved in the same pathway as GIIβ and SEC63, and each appears to affect PC1 biogenesis.

Intracranial aneurysm (ICA) occurs four to five times more frequently in ADPKD patients than in the general population and causes high mortality. The disease gene products PC1 and PC2 may be directly responsible for defects in arterial smooth muscle cells and myofibroblasts. The focal nature and the natural history of ICA in ADPKD remain unclear. A family history of ICA is a risk factor of aneurysm rupture in ADPKD; whether hypertension and cigarette smoking are independent risk factors is not clear. About 20–50% of patients may experience "warning headaches" preceding the index episode of subarachnoid hemorrhage due to ruptured ICA. A CT scan is generally used as the first diagnostic test. A lumbar puncture may be used to confirm the diagnosis. The role of radiologic screening for ICA in asymptomatic patients with ADPKD remains unclear. ADPKD patients with a positive family history of ICAs may undergo presymptomatic screening of ICAs by magnetic resonance angiography. Other vascular abnormalities in ADPKD patients include diffuse arterial dolichoectasias of the anterior and posterior cerebral circulation, which can predispose to arterial dissection and stroke. Mitral valve prolapse occurs in up to 30% of patients with ADPKD, and tricuspid valve prolapse is less common. Other valvular abnormalities occurring with increased frequency in ADPKD patients include insufficiency of the mitral, aortic, and tricuspid valves. Most patients are asymptomatic, but some may progress and require valve replacement. The prevalence of colonic diverticulae and abdominal wall hernias is also increased in ADPKD patients.

Diagnosis A diagnosis is typically made from a positive family history consistent with autosomal dominant inheritance and multiple kidney cysts bilaterally. Renal ultrasonography is often used for presymptomatic screening of at-risk subjects and for evaluation of potential living-related kidney donors from ADPKD families. The presence of *at least two renal cysts (unilateral or bilateral)* is sufficient for diagnosis among at-risk subjects between 15 and 29 years of age with a sensitivity value of 96% and specificity value of 100%. The presence of *at least two cysts in each kidney* and *at least four cysts in each kidney*, respectively, is required for the diagnosis among at-risk subjects aged 30–59 years and aged ≥60 years with a sensitivity value of 100% and specificity value of 100%. This is because there is an increased frequency of developing simple renal cysts with age. Conversely, in subjects aged between 30 and 59 years, the absence of *at least two cysts in each kidney*, which is associated with a false-negative rate of 0%, can be used for disease exclusion. These criteria have a lower sensitivity for patients with a

PKD2 mutation because of a late onset of ADPKD2. CT scan and T2-MRI, with and without contrast enhancement, are more sensitive than ultrasonography and can detect cysts of smaller size. However, a CT scan exposes the patient to radiation and radiocontrast, which may cause serious allergic reactions and nephrotoxicity in patients with renal insufficiency. T2-MRI, with gadolinium as a contrast agent, has minimal renal toxicity and can detect cysts of only 2–3 mm in diameter. However, a large majority of cysts may still be below the detection level. Genetic testing by linkage analyses and mutational analyses is available for ambiguous cases. Because of the large size of the *PKD1* gene and the presence of multiple highly homologous pseudogenes, mutational analysis of the *PKD1* gene is difficult and costly. Application of new technologies such as paired-end next-generation sequencing with multiplexing individually bar-coded long-range polymerase chain reaction (PCR) libraries may reduce the costs and improve the sensitivity for clinical genetic testing.

TREATMENT

Autosomal Dominant Polycystic Kidney Disease

No specific treatment to prevent cyst growth or the decline of renal function has been approved by the U.S. Food and Drug Administration. Blood pressure control to a target of 140/90 mmHg is recommended according to the guidelines from the eighth report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC VIII report) for reducing cardiovascular complications in ADPKD and renal disease progression. More rigorous blood pressure control does not equal greater clinical benefits. Maintaining a target systolic blood pressure to 110 mmHg in patients with moderate or advanced disease may increase the risk of renal disease progression by reducing renal blood flow. Lipid-soluble antibiotics against common gram-negative enteric organisms include trimethoprim-sulfamethoxazole, quinolones, and chloramphenicol, and are preferred for cyst infection because most renal cysts are not connected to glomerular filtration and antibiotics that are capable of penetrating the cyst walls are likely to be more effective. Treatment often requires 4–6 weeks. The treatment of kidney stones in ADPKD includes standard measures such as analgesics for pain relief and hydration to ensure adequate urine flow. Management of chronic flank, back, or abdominal pain due to renal enlargement may include both pharmacologic (nonnarcotic and narcotic analgesics) and nonpharmacologic treatment (transcutaneous electrical nerve stimulation, acupuncture, and biofeedback). Occasionally surgical decompression of cysts may be necessary. More than half of ADPKD patients eventually require peritoneal dialysis, hemodialysis, or kidney transplantation. Peritoneal dialysis may not be suitable for some patients with massively enlarged polycystic kidneys due to the small intraabdominal space for efficient peritoneal exchange of fluid and solutes and increased chance of abdominal hernia and back pain. Patients with very large polycystic kidneys and recurrent renal cyst infection may require pretransplant nephrectomy or bilateral nephrectomy to accommodate the allograft and reduce the pain.

Specific treatment strategies for ADPKD have focused on slowing renal disease progression and lowering cardiovascular risk. For the latter, the main approach is to control blood pressure by inhibiting the renin-angiotensin-aldosterone system. The HALT PKD trial was set to evaluate the impact of intensive blockade of the renin-angiotensin-aldosterone system and levels of blood pressure control on progressive renal disease. This trial found that rigorous blood pressure control could slow cyst growth. Most approaches target the slowing of renal disease progression by inhibiting cell proliferation and fluid secretion. Several clinical trials have been conducted targeting cell proliferation: sirolimus and everolimus, inhibitors of the mTOR pathway; OPC31260 and tolvaptan, which inhibit cyclic adenosine monophosphate (cAMP) pathways by antagonizing the activation of vasopressin V2 receptor (V2R) in collecting ducts and reduces cell proliferation by decreasing renal cAMP levels; and somatostatin analogues, which reduce cAMP levels by binding to several GPCRs. The TAMPO and ALADIN trials showed that V2R antagonists and somatostatin analogues (octreotide-LAR groups), respectively, slowed the decline of renal function. Some side effects, such as liver function impairment, polydipsia, and diarrhea, have been observed for tolvaptan and cholecystitis for octreotide-LAR. A recent report also showed that tolvaptan reduces renal pain. DIPAK, a small multicenter European study, showed that nerve block may be used to relieve pain in ADPKD patients suffering with refractory chronic pain. A combination of different growth inhibitors may enhance efficacy and reduce side effects. Notably, treatments may vary depending on the patient population. For example, the U.S. Food and Drug Administration (FDA) has indicated tolvaptan to be only for patients at risk of rapidly progressing disease. Combining genotypic and imaging information may predict kidney growth rates and help in selecting this patient population. Several other trials are being conducted. For example, compounds targeting microRNAs regulating the production of PKD1 and PKD2 proteins are currently under phase I trials in patients with ADPKD. Additionally, daily caloric restriction and intermittent fasting are also being trialed in overweight and obese adult patients.

Additional preclinical studies in animal models include the use of inhibitors to nonreceptor tyrosine kinase Src, B-raf, cyclin-dependent kinase (CDK), transcription factors STAT3 and STAT6 (pyrimethamine and leflunomide), purinergic receptors, hepatocyte growth factor receptor, glucosylceramide, agonists to peroxisome proliferator-activated receptor-gamma (PPARγ) receptors (thiazolidinediones), and targeting microRNAs. Reprogramming the metabolic pathway through studies of transcription regulator super-enhancer as well as dietary control, including time-restricted feeding, has been shown in murine models to reduce cyst area, kidney fibrosis, inflammation, and injury. Branched-chain amino acids appear to enhance cyst development in a mouse model.

■ AUTOSOMAL RECESSIVE POLYCYSTIC KIDNEY DISEASE

Genetic Considerations ARPKD is a significant hereditary renal disease in childhood, with an estimated prevalence of 1 in 20,000 live births. A carrier frequency of up to 1:70 has been reported. Mutations in a single gene, *PKHD1*, are responsible for all the clinical presentations of ARPKD. *PKHD1*, localized on human chromosome region 6p21.1–6p12.2, is one of the largest genes in the genome, occupies ~450 kb of DNA, and contains at least 86 exons. It produces multiple alternatively spliced transcripts. The largest transcript encodes fibrocystin/polyductin (FPC), which is a large receptor-like integral membrane protein of 4074 amino acids. FPC has a single transmembrane, a large N-terminal extracellular region, and a short intracellular cytoplasmic domain. FPC is localized on the primary cilia of epithelia cells of cortical and medullary collecting ducts and cholangiocytes of bile ducts, similar to polycystins and several other ciliopathy proteins. FPC is also expressed on the basal body and plasma membrane. The large extracellular domain of FPC is presumed to bind to an as yet unknown ligand(s) and is involved in cell-cell and cell-matrix interactions. FPC interacts with ADPKD protein PC2 and may also participate in regulation of the mechanosensory function of the primary cilia, calcium signaling, and PCP, suggesting a common mechanism underlying cystogenesis between ADPKD and ARPKD. FPC is also found on the centrosomes and mitotic spindle and may regulate centrosome duplication and mitotic spindle assembly during cell division. A large number of various mutations have been found throughout *PKHD1* and are unique to individual families. Most patients are compound heterozygotes for *PKHD1* mutations. Patients with two truncation mutations appear to have an earlier onset of the disease.

Clinical Features Classic ARPKD is generally diagnosed in utero or within the neonatal period and characterized by greatly enlarged echogenic kidneys in diseased fetuses. Reduced fetal urine production may contribute to oligohydramnios and pulmonary hypoplasia. About 30% of affected neonates die shortly after birth due to respiratory insufficiency. Close to 60% of mortality occurs within the first month of life. In the classic group, most patients are born with renal insufficiency and ESRD. However, infants often have a transient improvement in their GFR; death from renal insufficiency at this stage is rare. Some patients are diagnosed after the neonatal stage, which form the older group. Morbidity and mortality in this group often involve systemic hypertension, progressive renal insufficiency, and liver manifestations. The hallmarks of ARPKD liver disease are biliary dysgenesis due to a primary ductal plate malformation with associated periportal fibrosis, namely congenital hepatic fibrosis (CHF) and dilatation of intrahepatic bile ducts (Caroli disease). CHF and Caroli disease can then lead to portal hypertension exhibiting hepatosplenomegaly, variceal bleeding, and cholangitis. Some patients with the diagnosis of ARPKD at 1 year of age with nephromegaly exhibit slowly declining renal function over 20 years with only minimally enlarged kidneys at ESRD and markedly atrophic kidneys following renal transplantation. The slow progression of renal disease is likely due to increasing fibrosis rather than the development of cysts. Systemic hypertension is common in all ARPKD patients, even those with normal renal function.

Diagnosis Ultrasonography, CT, and MRI all can be used for diagnosis. Ultrasonography reveals large, echogenic kidneys with poor corticomedullary differentiation. The diagnosis can be made in utero after 24 weeks of gestation in severe cases. Macrocysts generally are not common at birth in ARPKD patients. The absence of renal cysts in either parent, particularly if they are >40 years of age, on ultrasonography helps distinguish ARPKD from ADPKD in older patients. Clinical, laboratory, or radiographic evidence of hepatic fibrosis, hepatic pathology demonstrating characteristic ductal plate abnormalities, family history of affected siblings, or parental consanguinity suggestive of autosomal recessive inheritance is helpful. The lack of mutational hot spots and the large and complex genomic structure of *PKHD1* make molecular diagnosis difficult; however, presymptomatic screen of other at-risk members in a family with already identified ARPKD mutations is straightforward and inexpensive.

TREATMENT

Autosomal Recessive Polycystic Kidney Disease

There is no specific therapy for ARPKD. Appropriate neonatal intensive care, blood pressure control, dialysis, and kidney transplantation increase survival into adulthood. Complications of hepatic fibrosis may necessitate liver transplantation. Patients with severe Caroli disease may need portosystemic shunting. Tolvaptan, an approved drug, functions by inhibiting the vasopressin pathway through the blockade of the vasopressin receptor 2 (V2). Preclinical studies conducted on a rat model of PKD, wherein the causative gene of ARPKD, *PKHD1*, is mutated, revealed promising results. These studies demonstrated a reduction in kidney cyst formation, decreased cyst cell growth, and preserved kidney function following tolvaptan administration. These findings suggest that the vasopressin pathway may contribute to the progression of renal manifestations in ARPKD. A multinational, multicenter phase 3 clinical trial is evaluating the safety and efficacy of the drug tolvaptan in treating infants and children with ARPKD. Upcoming therapies may target abnormal cell signaling mechanisms, as described above for ADPKD.

OTHER DISEASES CHARACTERIZED BY LARGE KIDNEY CYSTS

TUBEROUS SCLEROSIS

Tuberous sclerosis (TS) is a rare autosomal dominant syndrome caused by mutations in one of two genes, *TSC1*, encoding hamartin, or *TSC2*, encoding tuberin. Published estimates of prevalence vary widely, but it certainly occurs in <1:5000 births. Kidney cysts are a frequent feature of this condition, as are two other abnormalities of kidney growth, renal cell carcinoma and renal angiomyolipomas. TS is a syndrome affecting multiple organ systems. Other features of TS include benign growths in the nervous system, eyes, heart, lung, liver, and the skin. Essentially all TS patients have such skin lesions, and a large proportion of patients have neurologic and cognitive manifestations. Kidney cysts and their complications should be managed similarly to the management of these problems in the general population. The neurologic features of TS can be a source of significant morbidity.

The *TSC2* gene is adjacent to *PKD1* in the human genome. Some patients have deletions in their genomic DNA that inactivate these two genes. Such individuals may have manifestations of both ADPKD and TS. The majority of TS-causing mutations are found in TSC2.

Renal cysts are observed in about 20–30% of people with TS. The most common kidney finding in TS is the presence of angiomyolipomas. These growths tend to be multiple and bilateral. While they are usually benign, they may bleed. Surgical removal is often recommended as a prophylactic measure in people with angiomyolipomas >4 cm in diameter. The cysts in TS are radiographically similar to those seen in ADPKD. In contrast to ADPKD, there is a clearly increased risk of renal cell carcinoma in TS patients. Advances in imaging technology have improved the ability to detect TSC-related abnormalities, aiding in early diagnosis and treatment planning.

Regular periodic imaging is recommended in TS patients with kidney involvement to screen for the development of renal cell carcinoma. These cysts may rarely become large and hemorrhagic, occasionally requiring nephrectomy when nephron-sparing surgery is not possible. Although a rare problem, TS may lead to significant chronic kidney disease (CKD) and progress to end-stage kidney failure. Patients with TS and CKD typically have an unremarkable urine sediment and only small amounts of proteinuria.

Mechanistically, the *TSC1* and *TSC2* gene products tuberin and hamartin interact physically. This protein complex is localized to the base of the cilia and inhibits intracellular signaling processes mediated by mTOR, leading to abnormal growth in a number of tissues. Everolimus, an mTOR inhibitor, has been approved in the United States for treatment of TSC-associated kidney tumors as well as nonkidney manifestations of TS. mTOR inhibitors are now considered first-line treatment for TS.

VON HIPPEL-LINDAU DISEASE

Von Hippel-Lindau disease (VHL) is an inherited cancer syndrome with renal manifestations. VHL is an autosomal dominant condition caused by mutations in the VHL tumor-suppressor gene. The VHL protein plays a critical role in the regulation of hypoxia pathways and oxygen sensing via the transcription factor hypoxia-inducible factor (HIF). Like many other autosomal dominant cancer syndromes, VHL is recessive at the cellular level: a somatic mutation in the second VHL allele leads to loss of VHL in the cell and abnormal growth. Kidney manifestations of VHL include multiple bilateral kidney cysts and renal cell carcinomas. Kidney cysts and carcinoma affect the majority of VHL patients. Nonrenal features of VHL include pheochromocytomas, cerebellar hemangioblastomas, and retinal hemangiomas. While much rarer than ADPKD, VHL is an entity that should be considered in the differential diagnosis of an individual with newly recognized kidney cysts.

In these patients, annual screening of the kidneys by imaging with CT or MRI scanning is recommended for early detection of renal cell carcinomas. Increasingly, nephron-sparing surgical approaches are being used for removal of cancerous lesions in order to preserve kidney function. The HIF inhibitor belzutifan has been approved for VHL-associated renal cell carcinoma.

OTHER INHERITED DISEASES OF TUBULE GROWTH AND DEVELOPMENT

ADPKD is by far the most common adult-onset single-gene form of adult-onset kidney disease. The large cysts that are sometimes seen in VHL and TS are similar in appearance to the cysts seen in ADPKD. A variety of other inherited disorders affecting primarily tubule and renal interstitial function can lead to CKD and eventual end-stage kidney disease in the absence of large tubule-derived cysts.

Inherited diseases affecting the tubulointerstitial compartment of the kidney can lead to secondary glomerular stress and glomerulosclerosis with some degree of concomitant proteinuria. Similarly, disorders of glomerular function will typically lead to secondary interstitial fibrosis and tubule atrophy. From a clinical perspective, therefore, distinguishing between a genetic disease of the renal tubules and a disease of the glomerulus may not be easy, particularly in the absence of a gross phenotype such as large kidney cysts. In recent years, DNA sequencing panels have greatly improved the ability to make rapid and accurate genetic diagnoses of the underlying cause of many kidney disease, including diseases characterized by cysts, tumors, and other growth abnormalities.

AUTOSOMAL DOMINANT TUBULOINTERSTITIAL KIDNEY DISEASE (MEDULLARY CYSTIC KIDNEY DISEASE)

The term *autosomal dominant tubulointerstitial kidney disease* (ADTKD) has replaced the phrase *medullary cystic kidney disease* (MCKD) as the preferred designation for a group of autosomal dominant disorders characterized by progressive kidney failure and a benign urine sediment. Despite the old nosology, kidney cysts are not invariably present. Older literature often grouped MCKD together with the childhood-onset disorders known as the nephronophthises, but these are distinct clinical and genetic entities.

ADTKD-MUC1 Patients with medullary cystic kidney disease type I (MCKD I) have mutations in the mucin 1 gene *MUC1*. In contrast to MCKD II patients, individuals with MCKD I do not have elevated uric acid levels. The disease-causing *MUC1* mutations that have been reported all alter a highly repetitive region within the *MUC1* gene. This leads to the production of a large "neoprotein" fragment that has toxic effects on the kidney tubule.

Clinically, patients with MCKD I exhibit slowly progressive CKD in adulthood, with only minimal amounts of increased urine protein and occasional renal cysts seen on ultrasound examination. Kidney histology shows tubulointerstitial fibrosis and tubular atrophy. Disease does not recur in transplanted kidneys.

ADTKD-UMOD ADTKD-UMOD (also called MCKD II) is caused by mutations in the *UMOD* gene, which encodes the protein uromodulin, also known as Tamm-Horsfall protein. Uromodulin is also found on the centrosome, the mitotic spindle, and the primary cilia; it colocalizes with nephrocystin-1 and KIF3A on the cilia. *UMOD* mutations also cause the conditions that have been referred to as familial juvenile hyperuricemic nephropathy (HNFJ1) and glomerulocystic kidney disease (GCKD), although it does not seem that these different names represent clearly distinct disorders. The term *uromodulin-associated kidney disease* (UAKD) has been suggested as a better name for MCKD II and the various other related *UMOD*-associated diseases. Despite the name, kidney cysts are not a common feature of MCKD II. MCKD II should be suspected clinically in patients with a family history of late-onset kidney disease, benign urine sediments, absence of significant proteinuria, and hyperuricemia. Large genome-wide association studies have suggested that certain common noncoding sequence variants in *UMOD* are associated with a moderately increased risk of CKD in the general population. *UMOD*-associated disease is often associated with gout. The features of UMOD-associated kidney disease are fairly nonspecific, and the diagnosis should be made by genetic testing.

Other Forms of Familial Tubulointerstitial Kidney Disease A small number of families have been identified with ADTKD and hyperuricemia who lack *UMOD* mutations. Some of these families carry disease-segregating mutations in the renin gene *REN* (disease designation ADTKD-REN). ADKTKD-REN patients demonstrate hyporeninemia with mild hyperkalemia and often have hyperuricemia and gout. Mutations in *HNF1β* and *SEC61A1* are even rarer causes of ADTKD.

Kidney biopsies in patients with any of the various forms of ADTKD typically show interstitial fibrosis. These histologic features are not diagnostic of any particular genetic entity, and the specific diagnosis must be made by other means. Genetic tests for alterations in specific genes and in large panels of kidney disease genes are available in the clinical setting. High cost and complexity in interpretation are the major barriers to the use of such testing.

Patients with autosomal dominant interstitial kidney disease, *UMOD* or *REN* mutations, and hyperuricemia and gout should be treated similarly to others with these findings, with uric acid–lowering agents, such as allopurinol or febuxostat.

NEPHRONOPHTHISIS

A large and growing number of genetically distinct but related sets of autosomal recessive disorders are referred to as nephronophthises, or nephronophthisis-related ciliopathies. These entities should not be confused with the adult-onset autosomal dominant MCKD discussed above, despite the often-confusing nomenclature seen in older medical literature. Each of the individual forms of nephronophthisis is quite rare, but together this category constitutes the most common inherited childhood form of kidney failure requiring kidney replacement therapy. Most individuals with nephronophthisis develop end-stage kidney disease.

Like ADPKD and ARPKD, the various genetically heterogeneous entities that fall under the category of nephronophthisis (NPHP) are disorders of ciliary function. Mutations in >90 genes have been identified that lead to NPHP under an autosomal recessive pattern of inheritance. Some of these gene defects cause limited kidney disease, while many cause ciliopathies characterized by multiple organ involvement. The various forms of NPHP share common features, including tubulointerstitial fibrosis, corticomedullary cysts, and progressive CKD, leading to renal failure. Proteinuria is absent or mild, and the urine sediment is not active.

NPHP is often divided into infantile, juvenile, and adolescent forms. The juvenile form is the most frequent, and usually caused by mutations in the *NPHP2* gene. The infantile form, usually caused by *NPHP2* mutations, is associated with end-stage kidney failure in early childhood. Patients with the adolescent form of NPHP typically develop end-stage kidney failure in early adulthood. Hypertension, if present, tends to be a late finding in the course of the NPHPs. The products of the NPHP genes are referred to as nephrocystins. *NPHP1* through *NPHP20* have been reported; some are referred to by other names as well.

NPHP can present as an isolated finding or be part of several multiorgan syndromes. Neurologic abnormalities are present in a significant number of patients. Bone and liver abnormalities are seen in some NPHP patients. Senior-Loken syndrome is defined by the presence of NPHP with retinitis pigmentosa. Joubert syndrome is defined by multiple neurologic findings, including hypoplasia of the cerebellar vermis. Some forms of this genetically heterogeneous syndrome include NPHP as a component.

The multisystem disease Bardet-Biedl syndrome (BBS) is a subset of the ciliopathies. BBS is defined clinically by a spectrum of features, including truncal obesity, cognitive impairment, retinal dystrophy, polydactyly, developmental urogenital abnormalities, and kidney cysts. The kidney phenotype is NPHP-like, with small cysts deriving from the tubules, tubulointerstitial and often secondary glomerular disease, and urine concentrating defects. To date, 26 BBS genes have been identified. BBS follows autosomal recessive inheritance.

The multiple genes and gene products (nephrocystins) that are responsible for NPHP are expressed in cilia, basal bodies, and the centrosomes of kidney tubule cells. It has been hypothesized that all of the NPHP gene defects lead to a clinical phenotype by interfering with the regulation of PCP.

There are no specific clinical tests that define NPHP. Genetic diagnosis has become fairly straightforward due to new DNA sequencing technologies and should be pursued as part of the diagnostic workup. There are no specific therapies for NPHP. Rather, therapy is aimed at treating signs of these diseases as well as those systemic abnormalities seen with all CKDs. Chronic dialysis or kidney transplantation is eventually required for most NPHP-affected individuals.

KARYOMEGALIC TUBULOINTERSTITIAL NEPHRITIS

Karyomegalic tubulointerstitial nephritis is an exceptionally rare form of kidney disease with adult-onset progressive kidney failure. Kidney biopsy shows chronic tubulointerstitial nephritis, as well as interstitial fibrosis. This is a recessive disorder caused by inheritance of two mutant copies of the *FAN1* gene. *FAN1* encodes a component of a DNA repair machinery complex. Individuals with two mutant *FAN1* genes are genetically sensitized to the effect of DNA damage. Kidney histology shows karyomegaly in addition to the nonspecific findings of interstitial fibrosis and tubular atrophy.

MEDULLARY SPONGE KIDNEY

Medullary sponge kidney (MSK) is often grouped together with inherited disorders of the kidney affecting tubule growth and development, although it is usually a sporadic finding rather than an inherited phenotype. MSK is caused by developmental malformation and cystic dilatation of the renal collecting ducts. The medullary cysts seen in this entity can be quite variable in size.

MSK is usually a benign entity. The diagnosis of MSK is often made incidentally. In the past, the diagnosis of MSK was often made by intravenous pyelography (IVP). CT urography, which has replaced IVPs for much routine kidney imaging, is not as sensitive in detecting MSK.

MSK is associated with an increased frequency of calcium phosphate and calcium oxalate kidney stones. Altered flow characteristics in the kidney tubules may lead to the development of formation of a nidus for stone formation. Kidney stones in this group are treated the same as are kidney stones in the general population. MSK patients also often exhibit reduced kidney concentrating ability and an increased frequency of urinary tract infections.

CONGENITAL ABNORMALITIES OF THE KIDNEY AND URINARY TRACT

The structural abnormalities known as CAKUT (congenital abnormalities of the kidney and urinary tract) are a group of etiologically and phenotypically heterogeneous disorders. Some form of CAKUT is estimated to occur in up to 1 in 500 live births. Specific abnormalities classified as part of the CAKUT spectrum include kidney hypoplasia, kidney agenesis, ureteropelvic junction obstruction, and vesicoureteral reflux.

CAKUT can be the cause of clinically significant problems in both adults and children. However, it is a major contributor to kidney failure in children, accounting for more than one-third of end-stage kidney disease in this group.

CAKUT is typically a sporadic finding but can also cluster in families. Familial forms can be observed as parts of multisystem developmental syndromes. A growing list of specific genes have been identified that when mutated lead to both nonsyndromic and syndromic forms of CAKUT. For example, the branchio-oto-renal syndrome, characterized by developmental abnormalities in the neck, ears, and kidney, can be caused by mutations in the *EYA1* and *SIX1* genes. Mutations in the *PAX2* transcription factor gene can cause the autosomal dominant renal coloboma syndrome, characterized by optic nerve malformations and hypoplastic kidneys. A nontrivial fraction of children with CKD have an unsuspected genomic imbalance, often disrupting genes known to be relevant to CAKUT and kidney development. It is not uncommon for such genetic lesions to affect both kidney and neurocognitive function.

CAKUT can be caused by both environmental influences and genetic alterations. For example, renal tubular dysgenesis, defined by altered tubule development, can be caused by prenatal exposure of angiotensin-converting enzyme inhibitors or angiotensin receptor blockers.

MITOCHONDRIAL DISEASE

Inherited disorders of the mitochondrial genome (discussed elsewhere in this text **[see also Chap. 481]**) commonly affect kidney function. Thirteen of the genes involved in encoding components of the mitochondrial respiratory chain are located on the mitochondrial genome that is inherited maternally. The remainder of these components are encoded by the nuclear genome. These defects of oxidative phosphorylation may affect multiple organs and tissues.

Neuromuscular disease is the best recognized part of this complex phenotype. Kidney disease is now recognized as a common component, as well. Tubulointerstitial disease may be seen on kidney biopsy, and progression to kidney failure may occur. Glomerular involvement, manifest as proteinuria and glomerulosclerosis, can also develop. Changes in proximal tubule activity are the most common renal phenotype. Patients may have several defects in proximal tubule transport, including the Fanconi syndrome. Some patients may also have acidosis, hypophosphatemic rickets, hypercalciuria, glycosuria, and tubular proteinuria. Decreased urine concentrating ability is common.

DIAGNOSTIC CONSIDERTIONS

Recent studies using new DNA sequencing technologies suggest that variants in Mendelian kidney disease genes contribute to a nontrivial fraction of CKD cases, even when a clear Mendelian disease phenotype or family history of disease is lacking. Many studies also lead to the conclusion that various rare genetically mediated kidney diseases are difficult to categorize by phenotype alone. These diseases may mimic each other, an argument for using fairly large panels (or the entire genome) in genetic testing in the setting of kidney disease. The old and complicated nomenclature used to describe many human kidney diseases has largely been replaced by newer, genetically defined, categories.

■ GLOBAL CONSIDERATIONS

The disorders discussed above are all seen worldwide. In addition, a previously unrecognized epidemic of kidney disease is leading to very high rates of kidney failure in and near the western coast of Central America. This Mesoamerican nephropathy is particularly common in Nicaragua and El Salvador. Mesoamerican nephropathy patients do not have significant proteinuria, suggesting that this is a disease of the kidney tubules and interstitium. The causes are unknown, but it is thought that a combination of toxic environmental factors and heat stress underlie the development of this kidney disease, which has a striking male predominance. The fact that in many families a large fraction of the men have kidney disease has suggested that a strong genetic component is involved as well.

■ FURTHER READING

Cornec-Le Gall E et al: Autosomal dominant polycystic kidney disease. Lancet 393:919, 2019.

Devuyst O et al: Autosomal dominant tubulointerstitial kidney disease. Nat Rev Dis Primers 5:60, 2019.

Hays T et al: Genetic testing for kidney disease of unknown etiology. Kidney Int 98:590, 2020.

LaRiviere WB et al: Novel therapeutic approaches to autosomal dominant polycystic kidney disease. Transl Res 165:488, 2015.

Mekahli D et al: Clinical practice recommendations for kidney involvement in tuberous sclerosis complex: A consensus statement by the ERKNet Working Group for Autosomal Dominant Structural Kidney Disorders and the ERA Genes & Kidney Working Group. Nat Rev Nephrol 20:402, 2024.

Murugapoopathy V, Gupta IR: A primer on congenital anomalies of the kidneys and urinary tracts (CAKUT). Clin J Am Soc Nephrol 15:723, 2020.

Reddy BV, Chapman AB: The spectrum of autosomal dominant polycystic kidney disease in children and adolescents. Pediatr Nephrol 32:31, 2017.

Vivante A, Hildebrandt F: Exploring the genetic basis of early-onset chronic kidney disease. Nat Rev Nephrol 12:133, 2016.

Živná M et al: Autosomal dominant tubulointerstitial kidney disease: A review. Am J Med Genet C Semin Med Genet 190:309, 2022.

328

Tubulointerstitial Diseases of the Kidney

Laurence H. Beck, Jr., David J. Salant

Inflammation or fibrosis of the renal interstitium and atrophy of the tubular compartment are common consequences of diseases that target the glomeruli or vasculature. Distinct from these secondary phenomena, however, are a group of disorders that primarily affect the tubules and interstitium, with relative sparing of the glomeruli and renal vessels. Such disorders are conveniently divided into acute and chronic tubulointerstitial nephritis (TIN) **(Table 328-1)**.

Acute TIN most often presents with acute kidney injury **(Chap. 321)**. The acute nature of this group of disorders may be caused by aggressive inflammatory infiltrates that lead to tissue edema, tubular cell injury, and compromised tubular flow, or by frank obstruction of the tubules with casts, cellular debris, or crystals. There is sometimes flank pain due to distention of the renal capsule. Urinary sediment is often active with leukocytes and cellular casts but depends on the exact nature of the disorder in question.

The clinical features of chronic TIN are more indolent and may manifest with disorders of tubular function, including polyuria from impaired concentrating ability (nephrogenic diabetes insipidus), defective proximal tubular reabsorption leading to features of Fanconi's syndrome (glycosuria, phosphaturia, aminoaciduria, hypokalemia, and type II renal tubular acidosis [RTA] from bicarbonaturia), or non-anion-gap metabolic acidosis and hyperkalemia (type IV RTA) due to impaired ammoniagenesis, as well as progressive azotemia (rising creatinine and blood urea nitrogen [BUN]). There is often modest proteinuria (rarely >2 g/d) attributable to decreased tubular reabsorption of filtered proteins; however, nephrotic-range albuminuria may occur in some conditions due to the development of secondary focal segmental glomerulosclerosis (FSGS). Renal ultrasonography may reveal changes of "medical renal disease," such as increased echogenicity of the renal parenchyma with loss of corticomedullary differentiation, prominence of the renal pyramids, and cortical scarring in some conditions. The predominant pathology in chronic TIN is interstitial fibrosis with patchy mononuclear cell infiltration and widespread tubular atrophy, luminal dilation, and thickening of tubular basement membranes. Because of the nonspecific nature of the histopathology, biopsy specimens rarely provide a specific diagnosis. Thus, diagnosis relies on careful analysis of history, drug or toxin exposure, associated symptoms, and imaging studies.

ACUTE INTERSTITIAL NEPHRITIS

In 1897, Councilman reported eight cases of acute interstitial nephritis (AIN) in the Medical and Surgical Reports of the Boston City Hospital—three as a postinfectious complication of scarlet fever and two from diphtheria. Later, he described the lesion as "an acute inflammation of the kidney characterized by cellular and fluid exudation in the interstitial tissue, accompanied by, but not dependent on, degeneration of the epithelium; the exudation is not purulent in character, and the lesions may be both diffuse and focal." Today AIN is far more often encountered as an allergic reaction to a drug (Table 328-1). Immune-mediated AIN may also occur as part of a known autoimmune syndrome, but in some cases, there is no identifiable cause despite features suggestive of an immunologic etiology (Table 328-1).

ALLERGIC INTERSTITIAL NEPHRITIS

Although biopsy-proven AIN accounts for no more than ~15% of cases of unexplained acute kidney injury, this is likely a substantial underestimate of the true incidence. This is because potentially offending medications are more often identified and empirically discontinued in a patient noted to have a rising serum creatinine, without the benefit of a kidney biopsy to establish the diagnosis of AIN.

TABLE 328-1 Classification of the Causes of Tubulointerstitial Diseases of the Kidney

Acute Tubulointerstitial Disorders

Acute Interstitial Nephritis

Therapeutic agents

- Antibiotics (β-lactams, sulfonamides, quinolones, vancomycin, erythromycin, linezolid, minocycline, rifampin, ethambutol, acyclovir)
- Nonsteroidal anti-inflammatory drugs, COX-2 inhibitors
- Diuretics (rarely thiazides, loop diuretics, triamterene)
- Anticonvulsants (phenytoin, valproate, carbamazepine, phenobarbital)
- Immune modulators (immune checkpoint inhibitors, vedolizumab, lenalidomide)
- Miscellaneous (proton pump inhibitors, H_2 blockers, captopril, mesalazine, indinavir, allopurinol)

Infection

- Bacteria (*Streptococcus, Staphylococcus, Legionella, Salmonella, Brucella, Yersinia, Corynebacterium diphtheriae*)
- Viruses (EBV, CMV, hantavirus, polyomavirus, HIV)
- Miscellaneous (*Leptospira, Rickettsia, Mycoplasma, Histoplasma*)

Autoimmune

- Tubulointerstitial nephritis with uveitis (TINU)
- Sjögren's syndrome
- Systemic lupus erythematosus
- Granulomatous interstitial nephritis
- IgG4-related systemic disease
- Tubulointerstitial disease related to checkpoint inhibitors
- Anti-brush border disease (anti-LRP2 nephropathy)
- Idiopathic autoimmune interstitial nephritis

Acute Obstructive Disorders

- Light chain cast nephropathy ("myeloma kidney")
- Acute phosphate nephropathy
- Acute urate nephropathy

Chronic Tubulointerstitial Disorders

- Vesicoureteral reflux/reflux nephropathy
- Sickle cell disease
- Chronic exposure to toxins or therapeutic agents
- Analgesics, especially those containing phenacetin
- Lithium
- Heavy metals (lead, cadmium)
- Aristolochic acid (Chinese herbal and Balkan endemic nephropathies)
- Calcineurin inhibitors (cyclosporine, tacrolimus)
- Chronic interstitial nephritis in agricultural communities

Metabolic Disturbances

- Hypercalcemia and/or nephrocalcinosis
- Hyperuricemia
- Prolonged hypokalemia
- Hyperoxaluria
- Cystinosis **(see Chap. 327)**

Cystic and Hereditary Disorders (see Chap. 327)

- Polycystic kidney disease
- Nephronophthisis
- Autosomal dominant tubulointerstitial kidney disease (medullary cystic kidney disease)
- Medullary sponge kidney

Miscellaneous

- Aging
- Chronic glomerulonephritis
- Chronic urinary tract obstruction
- Ischemia and vascular disease
- Radiation nephritis (rare)

Abbreviations: CMV, cytomegalovirus; COX, cyclooxygenase; EBV, Epstein-Barr virus.

Clinical Features The classic presentation of AIN, namely, fever, rash, peripheral eosinophilia, and oliguric kidney injury occurring after 7–10 days of treatment with methicillin or another β-lactam antibiotic, is the exception rather than the rule. More often, patients are found incidentally to have a rising serum creatinine or present with symptoms attributable to acute kidney injury **(Chap. 321)**. Atypical reactions can occur, most notably with nonsteroidal anti-inflammatory drug (NSAID)–induced AIN, in which fever, rash, and eosinophilia are rare, but acute kidney injury with heavy proteinuria is common. A particularly severe and rapid-onset AIN may occur upon reintroduction of rifampin after a drug-free period. More insidious reactions to the agents listed in Table 328-1 may lead to progressive tubulointerstitial damage. Examples include proton pump inhibitors and, rarely, sulfonamide and 5-aminosalicylate (mesalazine and sulfasalazine) derivatives and antiretrovirals. It is not clear if the association of proton pump inhibitors with incident chronic kidney disease involves an intermediate step of prolonged, subclinical interstitial nephritis.

Diagnosis The finding of otherwise unexplained kidney injury with or without oliguria and exposure to a potentially offending agent usually points to the diagnosis. Peripheral blood eosinophilia adds supporting evidence but is present in only a minority of patients. Urinalysis reveals pyuria with white blood cell casts and hematuria. Urinary eosinophils are neither sensitive nor specific for AIN; therefore, testing is not recommended. Kidney biopsy is generally not required for diagnosis but reveals extensive interstitial and tubular infiltration of leukocytes, including eosinophils.

TREATMENT

Allergic Interstitial Nephritis

Discontinuation of the offending agent often leads to reversal of the kidney injury. However, depending on the duration of exposure and degree of tubular atrophy and interstitial fibrosis that has occurred, the kidney damage may not be completely reversible. Glucocorticoid therapy may accelerate kidney recovery but is not required in most cases. It is best reserved for those cases with severe kidney injury in which dialysis is imminent and should be started promptly if kidney function continues to deteriorate despite stopping the offending drug; delay in initiating dialysis leads to worse long-term outcomes **(Fig. 328-1 and Table 328-2)**.

TABLE 328-2 Indications for Corticosteroids and Immunosuppressives in Interstitial Nephritis

Absolute Indications

- Sjögren's syndrome
- Sarcoidosis
- SLE interstitial nephritis
- Adults with TINU
- Interstitial nephritis from IgG4-related disease
- Idiopathic and other granulomatous interstitial nephritis

Relative Indications

- Drug-induced or idiopathic AIN with:
  - Rapid progression of renal failure
  - Diffuse infiltrates on biopsy
  - Impending need for dialysis
  - Delayed recovery
- Children with TINU
- Postinfectious AIN with delayed recovery (?)

Abbreviations: AIN, acute interstitial nephritis; SLE, systemic lupus erythematosus; TINU, tubulointerstitial nephritis with uveitis.

Source: From Treatment of acute interstitial nephritis, S Reddy & DJ Salant: Renal Failure, 07 Jul 2009, Taylor and Francis. Reprinted by permission of the publisher: Taylor and Francis Ltd, *http://www.tandfonline.com*.

SJÖGREN'S SYNDROME

Sjögren's syndrome is a systemic autoimmune disorder that primarily targets the exocrine glands, especially the lacrimal and salivary glands, and thus results in symptoms, such as dry eyes and mouth, which constitute the "sicca syndrome" **(Chap. 373)**. TIN with a predominant lymphocytic infiltrate is the most common renal manifestation of Sjögren's syndrome and can be associated with impaired kidney function, distal RTA, and nephrogenic diabetes insipidus. Diagnosis is strongly supported by positive serologic testing for anti-Ro (SS-A) and anti-La (SS-B) antibodies. A large proportion of patients with Sjögren's syndrome also have polyclonal hypergammaglobulinemia. Treatment

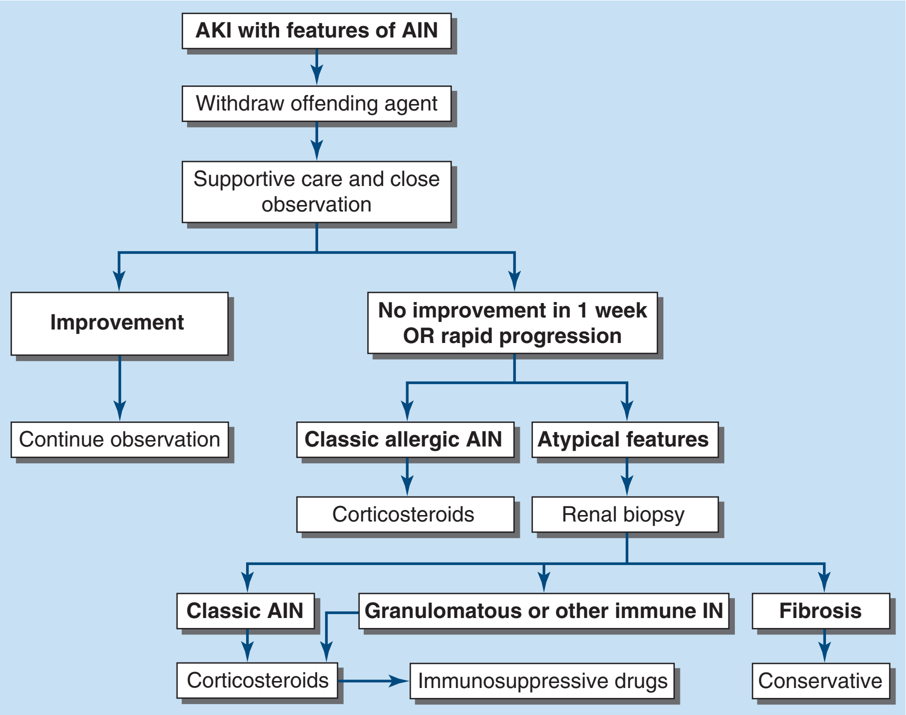

FIGURE 328-1 Algorithm for the treatment of allergic and other immune-mediated acute interstitial nephritis (AIN). AKI, acute kidney injury; IN, interstitial nephritis. See text for immunosuppressive drugs used for refractory or relapsing AIN. *(From Treatment of acute interstitial nephritis, S Reddy & DJ Salant: Renal Failure, 07 Jul 2009, Taylor and Francis. Reprinted by permission of the publisher: Taylor and Francis Ltd, http://www.tandfonline.com.)*

is initially with glucocorticoids, although patients may require maintenance therapy with azathioprine or mycophenolate mofetil to prevent relapse (Fig. 328-1 and Table 328-2).

■ TUBULOINTERSTITIAL NEPHRITIS WITH UVEITIS

Tubulointerstitial nephritis with uveitis (TINU) is a systemic autoimmune disease of unknown etiology. It accounts for <5% of all cases of AIN, affects females three times more often than males, and has a median age of onset of 15 years. Its hallmark feature, in addition to a lymphocyte-predominant interstitial nephritis **(Fig. 328-2)**, is a painful anterior uveitis, often bilateral and accompanied by blurred vision and photophobia. Diagnosis is often confounded by the fact that the ocular symptoms precede or accompany the kidney disease in only one-third of cases. Additional extrarenal features include fever, anorexia, weight loss, abdominal pain, and arthralgia. The presence of such symptoms as well as elevated creatinine, sterile pyuria, mild proteinuria, features of Fanconi's syndrome, and elevated erythrocyte sedimentation rate should raise suspicion for this disorder. Serologies suggestive of the more common autoimmune diseases are usually negative, and TINU is often a diagnosis of exclusion after other causes of uveitis and kidney disease, such as Sjögren's syndrome, Behçet's disease, sarcoidosis, and systemic lupus erythematosus, have been considered. Clinical symptoms are typically self-limited in children but are more apt to follow a relapsing course in adults. The renal and ocular manifestations generally respond well to oral glucocorticoids, although maintenance therapy with agents such as methotrexate, azathioprine, or mycophenolate may be necessary to prevent relapses (Fig. 328-1 and Table 328-2).

■ SYSTEMIC LUPUS ERYTHEMATOSUS

An interstitial mononuclear cell inflammatory reaction accompanies the glomerular lesion in most cases of class III or IV lupus nephritis **(Chap. 326)**, and deposits of immune complexes can be identified in tubular basement membranes in ~50% of cases. Occasionally, however, the tubulointerstitial inflammation predominates and may manifest with azotemia and type IV RTA rather than features of glomerulonephritis.

■ GRANULOMATOUS INTERSTITIAL NEPHRITIS

Some patients may present with features of AIN but follow a protracted and relapsing course. Kidney biopsy in such patients reveals a more chronic inflammatory infiltrate with granulomas and multinucleated giant cells. Most often, no associated disease or cause is found; however, some of these cases may have or subsequently develop the pulmonary, cutaneous, or other systemic manifestations of *sarcoidosis* such as hypercalcemia. Most patients experience some improvement in kidney function if treated early with glucocorticoids before the development of significant interstitial fibrosis and tubular atrophy (Table 328-2). Other immunosuppressive agents may be required for those who relapse frequently upon steroid withdrawal (Fig. 328-1). Tuberculosis should be ruled out before starting treatment because this too is a rare cause of granulomatous interstitial nephritis.

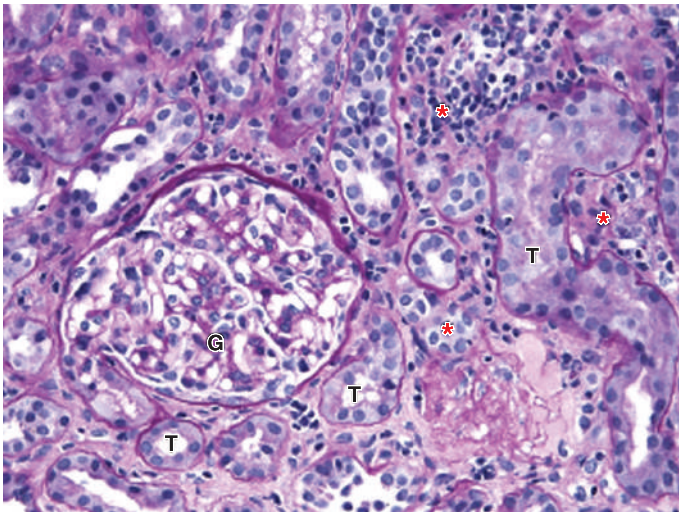

FIGURE 328-2 Acute interstitial nephritis (AIN) in a patient who presented with acute iritis, low-grade fever, erythrocyte sedimentation rate of 103, pyuria and cellular casts on urinalysis, and a newly elevated serum creatinine of 2.4 mg/dL. Both the iritis and AIN improved after intravenous methylprednisolone. This PAS-stained kidney biopsy shows a mononuclear cell interstitial infiltrate (*asterisks*) and edema separating the tubules (T) and a normal glomerulus (G). Some of the tubules contain cellular debris and infiltrating inflammatory cells. The findings in this biopsy are indistinguishable from those that would be seen in a case of drug-induced AIN. PAS, periodic acid–Schiff.

■ IgG4-RELATED SYSTEMIC DISEASE

A form of AIN characterized by a dense inflammatory infiltrate containing IgG4-expressing plasma cells can occur as a part of a syndrome known as IgG4-related systemic disease **(Chap. 380)**. Autoimmune pancreatitis, sclerosing cholangitis, retroperitoneal fibrosis, and a chronic sclerosing sialadenitis (mimicking Sjögren's syndrome) may variably be present as well. Fibrotic lesions that form pseudotumors in the affected organs soon replace the initial inflammatory infiltrates and often lead to biopsy or excision for fear of true malignancy. Although the involvement of IgG4 in the pathogenesis is not understood, glucocorticoids have been successfully used as first-line treatment in this group of disorders, once they are correctly diagnosed.

■ AIN ASSOCIATED WITH THE USE OF IMMUNE CHECKPOINT INHIBITORS

The use of immune checkpoint inhibitors has had a major impact in cancer care by disrupting mechanisms by which tumor cells elude the body's immune surveillance systems. However, such success comes at the cost of increasing the incidence of autoimmune phenomena. While dermatologic, gastrointestinal, and endocrine manifestations prevail, the kidney is impacted in 2% of cases with monotherapy and up to 5% when dual checkpoint inhibitor therapy is used. An acute rise in serum creatinine is typically noted within 15 weeks after starting therapy, although it can occur later during therapy or up to 2 months following the final dose. Biopsy, when performed, typically shows acute interstitial inflammation, although glomerular pathologies may also be found. Patients are often taking medications commonly known to cause acute drug-associated TIN such as proton pump inhibitors or NSAIDs. Treatment for severe acute kidney injury includes corticosteroids, discontinuation of potential inciting medications, and avoidance of further checkpoint inhibitor doses until the kidney function has recovered. Reinitiation of these agents for continued treatment of the underlying malignancy may be possible in some cases.

■ IDIOPATHIC AIN

Some patients present with typical clinical and histologic features of AIN but have no evidence of drug exposure or clinical or serologic features of an autoimmune disease. The presence in some cases of autoantibodies to a tubular antigen, similar to that identified in rats with an induced form of interstitial nephritis, suggests that an autoimmune response may be involved. Like TINU and granulomatous interstitial nephritis, idiopathic AIN is responsive to glucocorticoid therapy but may follow a relapsing course requiring maintenance treatment with another immunosuppressive agent (Fig. 328-1 and Table 328-2). Recently, cases have been identified in which autoantibodies that may be important in disease pathogenesis were seen to target antigens expressed by the collecting duct or proximal tubular brush border.

■ INFECTION-ASSOCIATED AIN

AIN may also occur as a local inflammatory reaction to microbial infection (Table 328-1) and should be distinguished from acute bacterial pyelonephritis **(Chap. 140)**. Acute bacterial pyelonephritis does not generally cause acute kidney injury unless it affects both kidneys or causes septic shock. Presently, infection-associated AIN is most often seen in immunocompromised patients, particularly kidney transplant recipients with reactivation of polyomavirus BK **(Chaps. 148 and 325)**. A significant increase in cases of acute TIN (without uveitis) or TINU has been noted in children after infection with SARS-CoV-2.

CRYSTAL DEPOSITION DISORDERS AND OBSTRUCTIVE TUBULOPATHIES

Acute kidney injury may occur when crystals of various types are deposited in tubular cells and interstitium or when they obstruct tubules. Impaired kidney function, often accompanied by flank pain from tubular obstruction, may occur in patients treated with sulfadiazine for toxoplasmosis, indinavir and atazanavir for HIV, and intravenous acyclovir for severe herpesvirus infections. Urinalysis reveals "sheaf of wheat" sulfonamide crystals, individual or parallel clusters of needle-shaped indinavir crystals, or red-green birefringent needle-shaped crystals of acyclovir. This adverse effect is generally precipitated by hypovolemia and is reversible with saline volume repletion and drug withdrawal. Distinct from the obstructive disease, a frank AIN from indinavir crystal deposition has also been reported.

Acute tubular obstruction is also the cause of oliguric kidney injury in patients with *acute urate nephropathy*. It typically results from severe hyperuricemia from tumor lysis syndrome in patients with lympho- or myeloproliferative disorders treated with cytotoxic agents but also may occur spontaneously before the treatment has been initiated **(Chap. 80)**. Uric acid crystallization in the tubules and collecting system leads to partial or complete obstruction of the collecting ducts, renal pelvis, or ureter. A dense precipitate of birefringent uric acid crystals is found in the urine, usually in association with microscopic or gross hematuria. Prophylactic allopurinol reduces the risk of uric acid nephropathy but is of no benefit once tumor lysis has occurred. Once oliguria has developed, attempts to increase tubular flow and solubility of uric acid with alkaline diuresis may be of some benefit; however, emergent treatment with hemodialysis or rasburicase, a recombinant urate oxidase, is usually required to rapidly lower uric acid levels and restore kidney function.

Calcium oxalate crystal deposition in tubular cells and interstitium may lead to permanent kidney dysfunction in patients who survive ethylene glycol intoxication, in patients with enteric hyperoxaluria from ileal resection or small-bowel bypass surgery, and in patients with hereditary hyperoxaluria **(Chap. 330)**. *Acute phosphate nephropathy* is an uncommon but serious complication of oral Phosphosoda used as a laxative or for bowel preparation for colonoscopy. It results from calcium phosphate crystal deposition in tubules and interstitium and occurs especially in subjects with underlying kidney disease and hypovolemia. Consequently, Phosphosoda should be avoided in patients with chronic kidney disease, and appropriately, the product is no longer widely available in the United States.

LIGHT CHAIN CAST NEPHROPATHY

Patients with multiple myeloma may develop acute kidney injury in the setting of hypovolemia, infection, or hypercalcemia or after exposure to NSAIDs or radiographic contrast media. The diagnosis of light chain cast nephropathy (LCCN)—commonly known as *myeloma kidney*—should be considered in patients who fail to recover when the precipitating factor is corrected or in any elderly patient with otherwise unexplained acute kidney injury.

In this disorder, filtered monoclonal immunoglobulin light chains (Bence-Jones proteins) form intratubular aggregates with secreted Tamm-Horsfall protein in the distal tubule. Casts, in addition to obstructing the tubular flow in affected nephrons, incite a giant cell or foreign-body reaction and can lead to tubular rupture, resulting in interstitial fibrosis **(Fig. 328-3)**. Although LCCN generally occurs in patients with known multiple myeloma and a large plasma cell burden, the disorder should also be considered as a possible diagnosis in patients who have known monoclonal gammopathy even in the absence of frank myeloma. Filtered monoclonal light chains may also cause less pronounced renal manifestations in the absence of obstruction, due to direct toxicity to proximal tubular cells and intracellular crystal formation. This may result in isolated tubular disorders such as RTA or full Fanconi's syndrome.

Diagnosis Clinical clues to the diagnosis include anemia, bone pain, hypercalcemia, and an abnormally narrow anion gap due to hypoalbuminemia and hypergammaglobulinemia. Urinary dipsticks detect albumin but not immunoglobulin light chains; however, laboratory detection of increased amounts of protein in a spot urine specimen and a negative dipstick result are highly suggestive that the urine contains Bence-Jones protein. Serum and urine should both be sent for protein electrophoresis and for immunofixation for the detection and identification of a potential monoclonal band. A sensitive method is available to detect urine and serum free light chains.

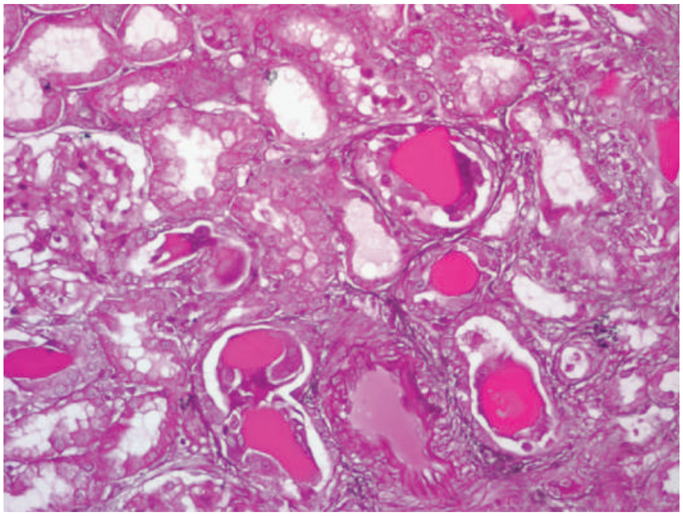

FIGURE 328-3 Histologic appearance of myeloma cast nephropathy. A hematoxylin-eosin–stained kidney biopsy shows many atrophic tubules filled with eosinophilic casts (consisting of Bence-Jones protein), which are surrounded by giant cell reactions. *(Courtesy of Dr. Michael N. Koss, University of Southern California Keck School of Medicine; with permission.)*

TREATMENT

Light Chain Cast Nephropathy

The goals of treatment are to correct precipitating factors such as hypovolemia and hypercalcemia, discontinue potential nephrotoxic agents, and treat the underlying plasma cell dyscrasia **(Chap. 116)**; plasmapheresis to remove light chains is of questionable value for LCCN.

LYMPHOMATOUS INFILTRATION OF THE KIDNEY

Interstitial infiltration by malignant B lymphocytes is a common autopsy finding in patients dying of chronic lymphocytic leukemia and non-Hodgkin's lymphoma; however, this is usually an incidental finding. Rarely, such infiltrates may cause massive enlargement of the kidneys and oliguric acute kidney injury. Although high-dose glucocorticoids and subsequent chemotherapy often result in recovery of kidney function, the prognosis in such cases is generally poor.

CHRONIC TUBULOINTERSTITIAL DISEASES

Improved occupational and public health measures, together with the banning of over-the-counter phenacetin-containing analgesics, has led to a dramatic decline in the incidence of chronic interstitial nephritis (CIN) from heavy metal—particularly lead and cadmium—exposure and analgesic nephropathy in North America. Today, CIN is most often the result of renal ischemia or secondary to a primary glomerular disease **(Chap. 326)**. Other important forms of CIN are the result of developmental anomalies or inherited diseases such as reflux nephropathy or sickle cell nephropathy and may not be recognized until adolescence or adulthood. Although it is impossible to reverse damage that has already occurred, further deterioration may be prevented or at least slowed in such cases by treating glomerular hypertension, a common denominator in the development of secondary FSGS and progressive loss of functioning nephrons. Therefore, awareness and early detection of patients at risk may prevent them from developing end-stage renal disease (ESRD).

VESICOURETERAL REFLUX AND REFLUX NEPHROPATHY

Reflux nephropathy is the consequence of vesicoureteral reflux (VUR) or other urologic anomalies in early childhood. It was previously called *chronic pyelonephritis* because it was believed to result from recurrent urinary tract infections (UTIs) in childhood. VUR stems from abnormal retrograde urine flow from the bladder into one or both ureters and kidneys because of mislocated and incompetent ureterovesical valves **(Fig. 328-4)**. Although high-pressure sterile reflux may impair normal growth of the kidneys, when coupled with recurrent UTIs in early childhood, the result is patchy interstitial scarring and tubular atrophy. Loss of functioning nephrons leads to hypertrophy of the remnant glomeruli and eventual secondary FSGS. Reflux nephropathy often goes unnoticed until early adulthood when chronic kidney disease is detected during routine evaluation or during pregnancy. Affected adults are frequently asymptomatic but may give a history of prolonged bed-wetting or recurrent UTIs during childhood and may exhibit variable degrees of kidney injury as well as hypertension, mild to moderate proteinuria, and an unremarkable urine sediment. When both kidneys are affected, the disease often progresses inexorably over several years to ESRD, despite the absence of ongoing urinary infections or reflux. A single affected kidney may go undetected, except for the presence of hypertension. Kidney ultrasound in adults characteristically shows asymmetric small kidneys with irregular outlines, thinned cortices, and regions of compensatory hypertrophy (Fig. 328-4).

TREATMENT

Vesicoureteral Reflux and Reflux Nephropathy

While it was previously believed that maintenance of sterile urine in childhood limits scarring of the kidneys, a recent study found

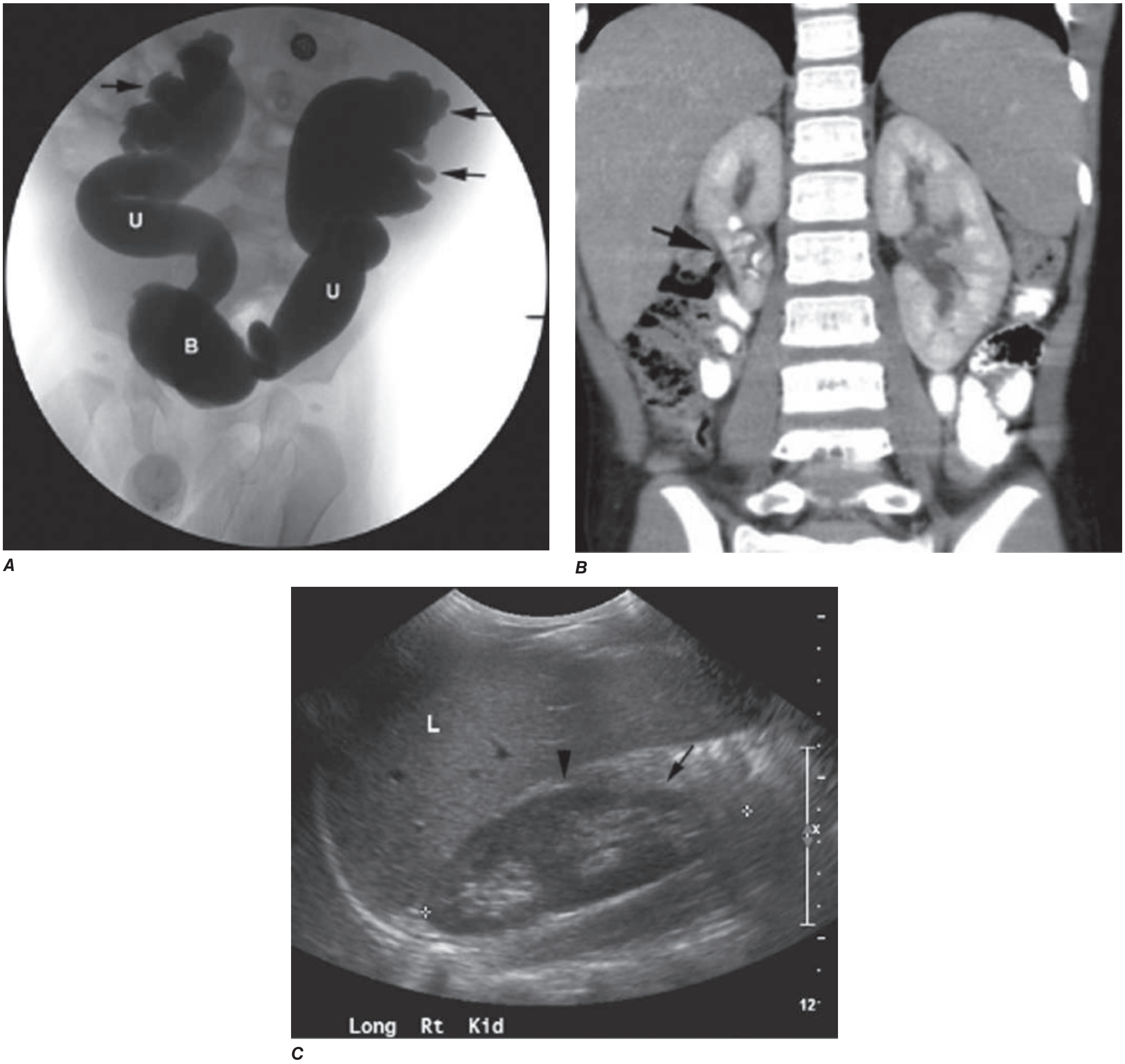

FIGURE 328-4 Radiographs of vesicoureteral reflux (VUR) and reflux nephropathy. ***A.*** Voiding cystourethrogram in a 7-month-old baby with bilateral high-grade VUR evidenced by clubbed calyces (*arrows*) and dilated tortuous ureters (U) entering the bladder (B). ***B.*** Abdominal computed tomography scan (coronal plane reconstruction) in a child showing severe scarring of the lower portion of the right kidney (*arrow*). ***C.*** Sonogram of the right kidney showing loss of parenchyma at the lower pole due to scarring (*arrow*) and hypertrophy of the mid-region (*arrowhead*). *(Courtesy of Dr. George Gross, University of Maryland Medical Center; with permission.)*

no such benefit, at least for 2 years. Surgical reimplantation of the ureters into the bladder to restore competency is indicated in young children with persistent high-grade reflux but is ineffective and is not indicated in adolescents or adults after scarring has occurred. Aggressive control of blood pressure and/or glomerular hyperfiltration with an angiotensin-converting enzyme inhibitor (ACEI) or angiotensin receptor blocker (ARB) and other agents (e.g., sodium-glucose cotransporter 2 [SGLT2] inhibitors) is effective in reducing proteinuria and may significantly forestall further deterioration of kidney function.

SICKLE CELL NEPHROPATHY

The pathogenesis and clinical manifestations of sickle cell nephropathy are described in **Chap. 329**. Evidence of tubular injury may be evident in childhood and early adolescence in the form of polyuria due to decreased concentrating ability or type IV RTA years before there is significant nephron loss and proteinuria from secondary FSGS. Early recognition of these subtle renal abnormalities or development of microalbuminuria in a child with sickle cell disease may warrant consultation with a nephrologist and/or therapy with low-dose ACEIs. Papillary necrosis may result from ischemia due to sickling of red cells in the relatively hypoxemic and hypertonic medullary vasculature and present with gross hematuria and ureteric obstruction by sloughed ischemic papillae **(Table 328-3)**.

TUBULOINTERSTITIAL ABNORMALITIES ASSOCIATED WITH GLOMERULONEPHRITIS

Primary glomerulopathies are often associated with damage to tubules and interstitium. This may occasionally be due to the same pathologic process affecting the glomerulus and tubulointerstitium, as is the case with immune-complex deposition in lupus nephritis. More often, however, chronic tubulointerstitial changes occur as a secondary consequence of prolonged glomerular dysfunction. Potential mechanisms by which glomerular disease might cause tubulointerstitial injury include proteinuria-mediated damage to the epithelial cells, activation of tubular cells by cytokines and complement, or reduced peritubular blood flow leading to downstream tubulointerstitial ischemia, especially in the case of glomeruli that are globally obsolescent due to severe glomerulonephritis. It is often difficult to discern the initial cause of injury by kidney biopsy in a patient who presents with advanced kidney disease in this setting.

ANALGESIC NEPHROPATHY

Analgesic nephropathy results from the long-term use of compound analgesic preparations containing phenacetin (banned in the United States since 1983), aspirin, and caffeine. In its classic form, analgesic nephropathy is characterized by impaired kidney function, papillary necrosis (Table 328-3) attributable to the presumed concentration of the drug to toxic levels in the inner medulla, and a radiographic constellation of small, scarred kidneys with papillary calcifications. Patients may also have polyuria due to impaired concentrating ability and non-anion-gap metabolic acidosis from tubular damage. Shedding of a sloughed necrotic papilla can cause gross hematuria and ureteric colic due to ureteral obstruction. Individuals with ESRD as a result of analgesic nephropathy are at increased risk of a urothelial malignancy compared to patients with other causes of kidney failure.

ARISTOLOCHIC ACID NEPHROPATHY

Two seemingly unrelated forms of CIN, Chinese herbal nephropathy and Balkan endemic nephropathy, have recently been linked by the underlying etiologic agent aristolochic acid and are now collectively termed aristolochic acid nephropathy (AAN). In Chinese herbal nephropathy, first described in the early 1990s in young women taking traditional Chinese herbal preparations as part of a weight-loss regimen, one of the offending agents has been identified as aristolochic acid, a known carcinogen from the plant *Aristolochia*. Multiple *Aristolochia* species have been used in traditional herbal remedies for centuries and continue to be available despite official bans on their use in many countries. Molecular evidence has also implicated aristolochic acid in Balkan endemic nephropathy, a chronic TIN found primarily in towns along the tributaries of the Danube River and first described in the 1950s. Although the exact route of exposure is not known with certainty, contamination of local grain preparations with the seeds of *Aristolochia* species seems most likely. Aristolochic acid, after prolonged exposure, produces renal interstitial fibrosis with a relative paucity of cellular infiltrates. The urine sediment is bland, with rare leukocytes and only mild proteinuria. Anemia may be disproportionately severe relative to the level of kidney dysfunction. Definitive diagnosis of AAN requires two of the following three features: characteristic histology on kidney biopsy; confirmation of aristolochic acid ingestion; and detection of aristolactam-DNA adducts in kidney or urinary tract tissue. These latter lesions represent a molecular signature of aristolochic acid–derived DNA damage and often consist of characteristic A:T-to-T:A transversions. Due to this mutagenic activity, AAN is associated with a very high incidence of upper urinary tract urothelial cancers, with risk related to cumulative dose. Surveillance with computed tomography, ureteroscopy, and urine cytology is warranted, and consideration should be given to bilateral nephroureterectomy once a patient has reached ESRD.

TABLE 328-3 Major Causes of Papillary Necrosis

Analgesic nephropathy
Sickle cell nephropathy
Diabetes with urinary tract infection
Prolonged NSAID use (rare)

Abbreviation: NSAID, nonsteroidal anti-inflammatory drug.

KARYOMEGALIC INTERSTITIAL NEPHRITIS

Karyomegalic interstitial nephritis is an unusual form of slowly progressive chronic kidney disease with mild proteinuria, interstitial fibrosis, tubular atrophy, and oddly enlarged nuclei of proximal tubular epithelial cells. It has been linked to mutations in *FAN1*, a nuclease involved in DNA repair, which may render carriers of the mutation susceptible to environmental DNA-damaging agents.

LITHIUM-ASSOCIATED NEPHROPATHY

The use of lithium salts for the treatment of manic-depressive illness may have several renal sequelae, the most common of which is nephrogenic diabetes insipidus manifesting as polyuria and polydipsia. Lithium accumulates in principal cells of the collecting duct by entering through the epithelial sodium channel (ENaC), where it inhibits glycogen synthase kinase 3β and downregulates vasopressin-regulated aquaporin water channels. Less frequently, chronic TIN develops after prolonged (>10–20 years) lithium use and is most likely to occur in patients who have experienced repeated episodes of toxic lithium levels. Findings on kidney biopsy include interstitial fibrosis and tubular atrophy that are out of proportion to the degree of glomerulosclerosis or vascular disease, a sparse lymphocytic infiltrate, and small cysts or dilation of the distal tubule and collecting duct that are highly characteristic of this disorder. The degree of interstitial fibrosis correlates with both duration and cumulative dose of lithium. Individuals with lithium-associated nephropathy are typically asymptomatic, with minimal proteinuria, few urinary leukocytes, and normal blood pressure. Some patients develop more severe proteinuria due to secondary FSGS, which may contribute to further loss of kidney function.

TREATMENT

Lithium-Associated Nephropathy

Kidney function should be followed regularly in patients taking lithium, and caution should be exercised in patients with underlying kidney disease. The use of amiloride to inhibit lithium entry via ENaC has been effective to prevent and treat lithium-induced nephrogenic diabetes insipidus, but it is not clear if it will prevent lithium-induced CIN. Once lithium-associated nephropathy is detected, the discontinuation of lithium in attempt to forestall further deterioration of kidney function can be problematic, as lithium is an effective mood stabilizer that is often incompletely substituted by other agents. Furthermore, despite discontinuation of lithium, chronic kidney disease in such patients is often irreversible and can slowly progress to ESRD. The most prudent approach is to monitor lithium levels frequently and adjust dosing to avoid toxic levels (preferably <1 meq/L). This is especially important because lithium is cleared less effectively as kidney function declines.

CALCINEURIN INHIBITOR NEPHROTOXICITY

The calcineurin inhibitor (CNI) immunosuppressive agents cyclosporine and tacrolimus can cause both acute and chronic kidney injury. Acute forms can result from vascular causes such as vasoconstriction or the development of thrombotic microangiopathy or can be due to a toxic tubulopathy. Chronic CNI-induced kidney injury is typically seen in solid organ (including heart-lung and liver) transplant recipients and manifests with a slow but irreversible reduction of glomerular filtration rate, with mild proteinuria and arterial hypertension. Hyperkalemia is a relatively common complication and is caused, in part, by tubular resistance to aldosterone. The histologic changes in kidney tissue include patchy interstitial fibrosis and tubular atrophy, often in a "striped" pattern. In addition, the intrarenal vasculature often demonstrates hyalinosis, and focal glomerulosclerosis can be present as well. Similar changes may occur in patients receiving CNIs for autoimmune diseases, although the doses are generally lower than those used for organ transplantation. Dose reduction or CNI avoidance appears to mitigate the chronic tubulointerstitial changes but may increase the risk of rejection and graft loss.

HEAVY METAL (LEAD) NEPHROPATHY

Heavy metals, such as lead or cadmium, can lead to a chronic tubulointerstitial process after prolonged exposure. The disease entity is no longer commonly diagnosed, because such heavy metal exposure has been greatly reduced due to the known health risks from lead and the consequent removal of lead from most commercial products and fuels. Nonetheless, occupational exposure is possible in workers involved in the manufacture or destruction of batteries, removal of lead paint, or manufacture of alloys and electrical equipment (cadmium) in countries where industrial regulation is less stringent. In addition, ingestion of moonshine whiskey distilled in lead-tainted containers has been one of the more frequent sources of lead exposure.

Early signs of chronic lead intoxication are attributable to proximal tubule dysfunction, particularly hyperuricemia as a result of diminished urate secretion. The triad of "saturnine gout," hypertension, and impaired kidney function should prompt a practitioner to ask specifically about lead exposure. Unfortunately, evaluating lead burden is not as straightforward as ordering a blood test; the preferred methods involve measuring urinary lead after infusion of a chelating agent or by radiographic fluoroscopy of bone. Several recent studies have shown an association between chronic low-level lead exposure and decreased kidney function, although either of these two factors may have been the primary event. In patients who have CIN of unclear origin and an elevated total body lead burden, repeated treatments of lead chelation therapy have been shown to slow the decline in kidney function.

METABOLIC DISORDERS

Disorders leading to excessively high or low levels of certain electrolytes and products of metabolism can also lead to chronic kidney disease if untreated.

CHRONIC URIC ACID NEPHROPATHY

The constellation of pathologic findings that represent *gouty nephropathy* is very uncommon nowadays and is more of historical interest than clinical importance, as gout is typically well managed with allopurinol and other agents. However, there is emerging evidence that hyperuricemia is an independent risk factor for the development of chronic kidney disease, perhaps through endothelial damage. The complex interactions of hyperuricemia, hypertension, and kidney failure are still incompletely understood.

Presently, gouty nephropathy is most likely to be encountered in patients with severe tophaceous gout and prolonged hyperuricemia from a hereditary disorder of purine metabolism. This should be distinguished from juvenile hyperuricemic nephropathy, a form of medullary cystic kidney disease caused by mutations in uromodulin (UMOD) **(Chap. 327)** and now grouped into the larger category of autosomal dominant tubulointerstitial kidney disease. Histologically, the distinctive feature of gouty nephropathy is the presence of crystalline deposits of uric acid and monosodium urate salts in the kidney parenchyma. These deposits not only cause intrarenal obstruction but also incite an inflammatory response, leading to lymphocytic infiltration, foreign-body giant cell reaction, and eventual fibrosis, especially in the medullary and papillary regions of the kidney. Since patients with gout frequently suffer from hypertension and hyperlipidemia, degenerative changes of the renal arterioles may constitute a striking feature of the histologic abnormality, out of proportion to the other morphologic defects. Clinically, gouty nephropathy is an insidious cause of chronic kidney disease. Early in its course, glomerular filtration rate may be near normal, often despite morphologic changes in medullary and cortical interstitium, proteinuria, and diminished urinary concentrating ability. Treatment with allopurinol and urine alkalinization is generally effective in preventing uric acid nephrolithiasis and the consequences of recurrent kidney stones; however, gouty nephropathy may be intractable to such measures. Furthermore, the use of allopurinol in asymptomatic hyperuricemia has not been consistently shown to improve kidney function.

HYPERCALCEMIC NEPHROPATHY

(See also Chap. 422) Chronic hypercalcemia, as occurs in primary hyperparathyroidism, sarcoidosis, multiple myeloma, vitamin D intoxication, or metastatic bone disease, can cause tubulointerstitial disease and progressive kidney injury. The earliest lesion is a focal degenerative change in renal epithelia, primarily in collecting ducts, distal tubules, and loops of Henle. Tubular cell necrosis leads to nephron obstruction and stasis of intrarenal urine, favoring local precipitation of calcium salts and infection. Dilation and atrophy of tubules eventually occur, as do interstitial fibrosis, mononuclear leukocyte infiltration, and interstitial calcium deposition (nephrocalcinosis). Calcium deposition may also occur in glomeruli and the walls of renal arterioles.

Clinically, the most striking defect is an inability to maximally concentrate the urine, due to reduced collecting duct responsiveness to arginine vasopressin and defective transport of sodium and chloride in the loop of Henle. Reductions in both glomerular filtration rate and renal blood flow can occur, both in acute and in prolonged hypercalcemia. Eventually, uncontrolled hypercalcemia leads to severe tubulointerstitial damage and overt kidney injury. Abdominal x-rays may demonstrate nephrocalcinosis as well as nephrolithiasis, the latter due to the hypercalciuria that often accompanies hypercalcemia.

Treatment consists of reducing the serum calcium concentration toward normal and correcting the primary abnormality of calcium metabolism **(Chap. 422)**. Acute kidney injury from acute hypercalcemia may be completely reversible. Gradual progressive kidney dysfunction related to chronic hypercalcemia, however, may not improve even with correction of the calcium disorder.

HYPOKALEMIC NEPHROPATHY

Patients with prolonged and severe hypokalemia from chronic laxative or diuretic abuse, surreptitious vomiting, or primary aldosteronism may develop a reversible tubular lesion characterized by vacuolar degeneration of proximal and distal tubular cells. Eventually, tubular atrophy and cystic dilation accompanied by interstitial fibrosis may ensue, leading to irreversible chronic kidney disease. Timely correction of the hypokalemia will prevent further progression, but persistent hypokalemia can cause ESRD.

GLOBAL PERSPECTIVE

The causes of AIN and CIN vary widely across the globe. Analgesic nephropathy continues to be seen in countries where phenacetin-containing compound analgesic preparations are readily available.

Adulterants in unregulated herbal and traditional medicaments pose a threat of toxic interstitial nephritis, as exemplified by aristolochic acid contamination of herbal slimming preparations. Contamination of food sources with toxins, such as an outbreak of nephrolithiasis and acute kidney injury from melamine contamination of infant milk formula, poses a continuing risk. Large-scale exposure to aristolochic acid remains prevalent in many Asian countries where traditional herbal medicine use is common. Although industrial exposure to lead and cadmium has largely disappeared as a cause of CIN in developed nations, it remains a risk for nephrotoxicity in countries where such exposure is less well controlled.

New endemic forms of chronic kidney disease continue to be described. In particular, nephropathies with features of CIN have been increasing in prevalence among Pacific coastal plantation workers in Central America (Mesoamerican nephropathy), Sri Lanka (Sri Lankan nephropathy), and southern India (Uddanam nephropathy). Together, these disorders have been called chronic interstitial nephritis of agricultural communities (CINAC) or chronic kidney disease of unknown etiology (CKDu) and may be related to repetitive episodes of heat exposure, dehydration, and volume depletion in the field workers. However, toxins, pesticides, and infective agents also remain as possible etiologic agents. Global warming and regional temperature variability have been proposed as contributors to these newly described forms of kidney disease, and tens of thousands of lives have been lost due to ESRD in these resource-poor areas in which renal replacement therapy is often not an option.

FURTHER READING

Eckardt KU et al: Autosomal dominant tubulointerstitial kidney disease: Diagnosis, classification, and management: A KDIGO consensus report. Kidney Int 88:676, 2015.

Johnson RJ et al: Chronic kidney disease of unknown cause in agricultural communities. N Engl J Med 380:1843, 2019.

Perazella MA, Rosner MH: Drug-induced acute kidney injury. Clin J Am Soc Nephrol 17:1220, 2022.

Praga M et al: Changes in the aetiology, clinical presentation and management of acute interstitial nephritis, an increasingly common cause of acute kidney injury. Nephrol Dial Transplant 30:1472, 2015.

Seethapathy H et al: The incidence, causes, and risk factors of acute kidney injury in patients receiving immune checkpoint inhibitors. Clin J Am Soc Nephrol 14:1692, 2019.

329

Thrombotic Renovascular Disorders

Elisabeth M. Battinelli, Rebecca L. Zon

The renal circulation is complex and is characterized by a highly perfused arteriolar network, reaching cortical glomerular structures adjacent to lower-flow vasa recta that descend into medullary segments. This chapter examines primary disorders of the microvessels, many of which are associated with thrombosis and hemolysis. Disorders of the larger vessels, including renal artery stenosis and atheroembolic disease, are discussed elsewhere **(Chap. 289)**.

THROMBOTIC MICROANGIOPATHY

Thrombotic microangiopathy (TMA) is a pathologic lesion characterized by endothelial cell injury in the terminal arterioles and capillaries. Platelet and hyaline thrombi causing partial or complete occlusion are integral to the histopathology of TMA. TMA is usually accompanied by microangiopathic hemolytic anemia (MAHA) with its typical features of thrombocytopenia and schistocytes, but not always. In the kidney, TMA is characterized by swollen endocapillary cells (endotheliosis), fibrin thrombi, platelet plugs, arterial intimal fibrosis, and a membranoproliferative pattern in the glomerulus. Fibrin thrombi may extend into the arteriolar vascular pole, producing glomerular collapse and at times cortical necrosis. In kidneys that recover from acute TMA, secondary focal segmental glomerulosclerosis may develop. Thrombotic vascular diseases include thrombotic thrombocytopenic purpura (TTP), hemolytic-uremic syndrome (HUS), malignant hypertension, scleroderma renal crisis, antiphospholipid syndrome, preeclampsia/HELLP (*h*emolysis, *e*levated *l*iver enzymes, *l*ow *p*latelet count) syndrome, HIV infection, cancer-associated TMA, and microvascular disease associated with COVID-19.

HEMOLYTIC-UREMIC SYNDROME/THROMBOTIC THROMBOCYTOPENIC PURPURA

HUS and TTP are the prototypes for MAHA. Historically, HUS and TTP were distinguished mainly by their clinical and epidemiologic differences. TTP develops more commonly in adults and was thought to have more neurologic complications, while HUS occurs more frequently in children, particularly when associated with hemorrhagic diarrhea. However, atypical HUS (aHUS) can have its first appearance in adulthood, and neurologic involvement can be as common in HUS as in TTP. Currently, HUS and TTP can be differentiated etiologically and treated according to their specific pathophysiologic features.

Hemolytic-Uremic Syndrome HUS is loosely defined by the presence of MAHA and renal impairment. At least four variants are recognized. The most common is Shiga toxin–producing *Escherichia coli* (STEC) HUS, which is also known as D^+ (diarrhea-associated) HUS or enterohemorrhagic *E. coli* (EHEC) HUS. Most cases involve children <5 years of age, but adults also are susceptible, as evidenced by a 2011 outbreak in northern Europe. Diarrhea, often bloody, precedes MAHA within 1 week in >80% of cases. Abdominal pain, cramping, and vomiting are frequent, whereas fever is typically absent. Neurologic symptoms, including dysphasia, hyperreflexia, blurred vision, memory deficits, encephalopathy, perseveration, and agraphia, often develop, especially in adults. Seizures and cerebral infarction can occur in severe cases. STEC HUS is caused by the Shiga toxins (Stx1 and Stx2), which are also referred to as *verotoxins.* These toxins are produced by certain strains of *E. coli* and *Shigella dysenteriae.* In the United States and Europe, the most common STEC strain is O157:H7, but HUS has been reported with other strains (O157/H^-, O111:H^-, O26:H11/H^-, O145:H28, and O104:H4). After entry into the circulation, Shiga toxin binds to the glycolipid surface receptor globotriaosylceramide (Gb3), which is richly expressed on cells of the renal microvasculature. Upon binding, the toxin enters the cells, inducing inflammatory cytokines (interleukin 8 [IL-8], monocyte chemotactic protein 1 [MCP-1], and stromal cell–derived factor 1 [SDF-1]) and chemokine receptors (CXCR4 and CXCR7); this action results in platelet aggregation and the microangiopathic process. *Streptococcus pneumoniae* can also cause HUS. Certain strains produce a neuraminidase that cleaves the *N*-acetylneuraminic acid moieties normally covering the Thomsen-Friedenreich antigen on platelets and endothelial cells. Exposure of this cryptic antigen to preformed IgM results in severe MAHA.

aHUS or complement-mediated HUS is the result of complement dysregulation. The complement dysregulation can be congenital or acquired. The affected patients often have low C3 and normal C4 levels characteristic of alternative pathway activation. Factor H deficiency, the most common defect, has been linked to families with aHUS. Factor H competes with factor B to prevent the formation of C3bBb and acts as a cofactor for factor I, which proteolytically degrades C3b. More than 70 mutations of the factor H gene have been identified. Most are missense mutations that produce abnormalities in the C-terminus region, affecting its binding to C3b but not its concentration. Other mutations result in low levels or the complete absence of the protein. Deficiencies in other complement-regulatory proteins, such as factor I, factor B, membrane cofactor protein (CD46), C3, complement factor H (CFH)–related protein 1 (CFHR1), CFHR3, CFHR5, and thrombomodulin, have also been reported. Finally, an autoimmune variant of aHUS, DEAP (deficiency of CFHR plasma proteins and

CFH autoantibody positive) HUS, occurs when an autoantibody to factor H is formed. DEAP HUS is often associated with a deletion of an 84-kb fragment of the chromosome that encodes for CFHR1 and CFHR3. The autoantibody blocks the binding of factor H to C3b and surface-bound C3 convertase. Renal injury is often severe, resulting in end-stage renal disease. The severity of the renal injury and recurrence after kidney transplant depend on the complement regulatory protein.

Thrombotic Thrombocytopenic Purpura Traditionally, TTP is characterized by the pentad of MAHA, thrombocytopenia, neurologic symptoms, fever, and renal failure; however, <5% of individuals with immune-mediated TTP will have the full pentad. The pathophysiology of TTP involves the accumulation of ultra-large multimers of von Willebrand factor as a result of the absence or markedly decreased activity of the plasma protease ADAMTS13, a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13. TTP is now defined as MAHA associated with ADAMTS13 activity of <5–10%. These ultra-large multimers form clots and shear erythrocytes, resulting in MAHA; however, the absence of ADAMTS13 alone may not by itself produce TTP. Often, an additional inflammatory trigger (such as infection, surgery, pancreatitis, or pregnancy) is required to initiate clinical TTP. This may be mediated by human neutrophil peptides that inhibit cleavage of von Willebrand factor by ADAMTS13. TTP can be congenital from ADAMTS13 mutation (cTTP) or acquired from autoantibody against ADAMTS13 protein (iTTP).

cTTP, also known as Upshaw-Schülman syndrome, is characterized by congenital deficiency of ADAMTS13. cTTP can start within the first weeks of life but, in some instances, may not present until adulthood, especially during pregnancy. Both environmental and genetic factors are thought to influence the development of cTTP. Plasma transfusion is an effective strategy for prevention and treatment. In iTTP, autoantibody to ADAMTS13 (IgG or IgM) either increases its clearance or inhibits its activity. Data from the Oklahoma TTP/HUS Registry suggest an iTTP incidence rate of 2.9 cases/10^6 patients in the United States. The median age of onset is 40 years. The incidence is more than nine times higher among blacks than nonblacks. Like that of systemic lupus erythematosus, the incidence of iTTP is nearly three times higher among women than among men. If untreated, iTTP has a mortality rate exceeding 90%. Even with modern therapy, 20% of patients die within the first month from complications of microvascular thrombosis.

Drug-induced TMA is a recognized complication of treatment with some chemotherapeutic agents, immunosuppressive agents, and quinine. Two different mechanisms are now recognized. Toxic or endothelial damage (pathologically similar to that of HUS) is the main cause of the TMA that develops in association with chemotherapeutic agents (e.g., proteasome inhibitors [bortezomib, carfilzomib, and ixazomib], mitomycin C, and gemcitabine) and immunosuppressive agents (cyclosporine, interferon, sirolimus, and tacrolimus). This process is usually dose-dependent. Alternatively, TMA may develop as a result of drug-induced autoantibodies. This form is less likely to be dose-dependent and can, in fact, occur after a single dose in patients with previous exposure (quinine). ADAMTS13 deficiency is found in fewer than half of patients with clopidogrel-associated TTP. Quinine appears to induce autoantibodies to granulocytes, lymphocytes, endothelial cells, and platelet glycoprotein Ib/IX or IIb/IIIa complexes, but not to ADAMTS13. Quinine-associated TTP is more common among women. TMA has also been reported with drugs that inhibit vascular endothelial growth factor, such as bevacizumab; the mechanism is not completely understood.

TREATMENT

Hemolytic-Uremic Syndrome/Thrombotic Thrombocytopenic Purpura

Treatment should be based on pathophysiology. iTTP and DEAP HUS respond to the combination of plasma exchange and prednisone. In addition to removing the autoantibodies, plasma exchange with fresh-frozen plasma replaces ADAMTS13. Twice-daily plasma exchanges with administration of rituximab may be effective in refractory cases. The use of caplacizumab, a monoclonal antibody fragment that binds to the A1 domain of von Willebrand factor, blocking its interaction with platelets, was recently shown to improve platelet count recovery and reduce the composite risk of death, disease exacerbation, and thromboembolic events. It is now approved for use in iTTP in conjunction with plasma exchange and immunosuppressive therapy. Plasma infusion is usually sufficient to replace the ADAMTS13 in cTTP. Plasma exchange should be considered if larger volumes are necessary. Additionally, newer medications are being studied and utilized in TTP, such as caplacizumab, a humanized monoclonal antibody fragment that binds to von Willebrand factor and blocks its interaction with platelet glycoproteins. Caplacizumab can be considered as an additive medication, especially in those with severe features of TTP or in cases with continued thrombocytopenia that does not respond to initial therapy.

Plasma infusion/exchange is effective in certain types of aHUS because it replaces complement-regulatory proteins. Eculizumab and ravulizumab, anti-C5 monoclonal antibodies, are approved for use in aHUS and have been shown to abort MAHA and improve renal function. Antibiotics and washed red cells should be given in neuraminidase-associated HUS, and plasmapheresis may be helpful; however, plasma and whole-blood transfusion should be avoided since these products contain IgM, which may exacerbate MAHA. Combined factor H and ADAMTS13 deficiency has been reported. The affected patients are generally less responsive to plasma infusion, an outcome that illustrates the complexity of the management of these cases.

Drug-induced TMA secondary to endothelial damage typically does not respond to plasma exchange and is treated primarily by discontinuing the use of the agent and, if refractory, a trial of C5 inhibitors. Similarly, STEC HUS should be treated with supportive measures as plasma exchange has not been found to be effective. Antimotility agents and antibiotics increase the incidence of HUS among children, but azithromycin may decrease the duration of bacterial shedding in adults.

HEMATOPOIETIC STEM CELL TRANSPLANTATION–ASSOCIATED THROMBOTIC MICROANGIOPATHY

Hematopoietic stem cell transplantation (HSCT)–associated TMA develops after allogeneic HSCT, with an incidence of ~8%. Etiologic factors include conditioning regimens, immunosuppression, infections, and graft-versus-host disease. Other risk factors include female sex and human leukocyte antigen (HLA)–mismatched donor grafts. HSCT-TMA usually occurs within the first 100 days of HSCT. Table 329-1 lists definitions of HSCT-TMA currently used for clinical trials. Diagnosis may be difficult since thrombocytopenia, anemia, and renal insufficiency are common after HSCT. HSCT-TMA carries a high mortality rate (75% within 3 months). The majority of patients have >10% ADAMTS13 activity, and plasma exchange is beneficial in <25%

TABLE 329-1 Criteria for Establishing Microangiopathic Kidney Injury Associated with Hematopoietic Stem Cell Transplantation

INTERNATIONAL WORKING GROUP	BLOOD AND MARROW TRANSPLANT CLINICAL TRIALS NETWORK TOXICITY COMMITTEE
>4% schistocytes in the blood	RBC fragmentation and at least 2 schistocytes per high-power field
De novo, prolonged, or progressive thrombocytopenia	Concurrent increase in LDH above baseline
A sudden and persistent increase in LDH	Negative direct and indirect Coombs test
Decrease in hemoglobin or increased RBC transfusion requirement	Concurrent renal and/or neurologic dysfunction without other explanations
Decrease in haptoglobin concentration	

Abbreviations: LDH, lactate dehydrogenase; RBC, red blood cell.

of patients. Discontinuation of calcineurin inhibitors and treatment of infections or sinusoidal obstruction syndrome (if present) are recommended. There are increasing reports of successful use of eculizumab, but clinical trial data are lacking.

CANCER-ASSOCIATED TMA

When MAHA and thrombocytopenia are present, one should assess for evidence of malignancy and consider the diagnosis of cancer-associated TMA. The mechanism of TMA in cancer is thought to be tumor cell obstruction in the microvasculature leading to cell fragmentation and platelet consumption: TMAs in cancer can occur on initial diagnosis of the cancer or when the cancer becomes refractory, and have been identified in both solid malignancies and hematologic malignancies. Clinical features, such as bone pain and respiratory symptoms, have been found to occur more in cancer-associated TMA than in TTP. Additionally, compared to immune-mediated TTP, there is no role for plasma exchange, steroids, or other immunosuppression in cancer-associated TMA; instead, the management is treatment of the underlying malignancy.

HIV-RELATED THROMBOTIC MICROANGIOPATHY

HIV-related TMA is a complication encountered mainly before widespread use of highly active antiretroviral therapy. It is seen in patients with advanced AIDS and low CD4+ T-cell counts, although it can be the first manifestation of HIV infection. The presence of MAHA, thrombocytopenia, and renal failure is suggestive, but renal biopsy is required for diagnosis since other renal diseases are also associated with HIV infection. Thrombocytopenia may prohibit renal biopsy in some patients. The mechanism of injury is unclear, although HIV can induce apoptosis in endothelial cells. ADAMTS13 activity is not reduced in these patients. Cytomegalovirus co-infection may also be a risk factor. Effective antiviral therapy is key, while plasma exchange should be limited to patients who have evidence of TTP.

PROGRESSIVE SYSTEMIC SCLEROSIS (SCLERODERMA)

MAHA can also be present in scleroderma renal crisis, which is also characterized by acute kidney injury, abrupt onset of hypertension, and a normal urine sediment, although each of these aspects is not required for the diagnosis. Scleroderma renal crisis occurs in 12% of patients with diffuse systemic sclerosis but in only 2% of those with limited systemic sclerosis. Although MAHA is present in more than half of patients, coagulopathy is rare. Retinopathy and encephalopathy may accompany the hypertension. Salt and water retention with microvascular injury can lead to pulmonary edema. Cardiac manifestations, including myocarditis, pericarditis, and arrhythmias, denote an especially poor prognosis.

The renal lesion in scleroderma renal crisis is characterized by arcuate artery intimal and medial proliferation with luminal narrowing. This lesion is described as "onion-skinning" and can be accompanied by glomerular collapse due to reduced blood flow. Histologically, scleroderma renal crisis is indistinguishable from malignant hypertension, with which it can coexist. Fibrinoid necrosis and thrombosis are common. Before the availability of angiotensin-converting enzyme (ACE) inhibitors, the mortality rate for scleroderma renal crisis was >90% at 1 month. Introduction of renin-angiotensin system blockade has lowered the mortality rate to 30% at 3 years. Nearly two-thirds of patients with scleroderma renal crisis may require dialysis support, with recovery of renal function in 50% (median time, 1 year). Glomerulonephritis and vasculitis associated with antineutrophil cytoplasmic antibodies and systemic lupus erythematosus have been described in patients with scleroderma. An association has been found with a speckled pattern of antinuclear antibodies and with antibodies to RNA polymerases I and III. Anti-U3-RNP may identify young patients at risk for scleroderma renal crisis. Anticentromere antibody, in contrast, is a negative predictor of this disorder. Because of the overlap between scleroderma renal crisis and other autoimmune disorders, a renal biopsy is recommended for patients with atypical renal involvement, especially if hypertension is absent.

Treatment with ACE inhibition is the first-line therapy unless contraindicated. The goal of therapy is to reduce systolic and diastolic blood pressure by 20 mmHg and 10 mmHg, respectively, every 24 h until blood pressure is normal. Additional antihypertensive therapy may be given once the dose of drug for ACE inhibition is maximized. Angiotensin II receptor antagonists are less effective at preventing renal failure; thus, they are only recommended if the patient is intolerant of ACE inhibitors. ACE inhibition alone does not prevent scleroderma renal crisis, but it does reduce the impact of hypertension. In addition, it has been observed that patients on ACE inhibitors have a higher renal recovery rate after initiation of dialysis, and thus, ACE inhibitors are continued even after starting dialysis. Intravenous iloprost has been used in Europe for blood pressure management and improvement of renal perfusion. Kidney transplantation is not recommended for 2 years after the start of dialysis since delayed recovery may occur. Bosentan (endothelin-1 antagonist) and eculizumab have both been investigated for use in this disease.

ANTIPHOSPHOLIPID SYNDROME

Antiphospholipid syndrome (APLS) **(Chap. 369)** can be either primary or secondary, with the most common rheumatologic association being with systemic lupus erythematosus. It is characterized by a predisposition to systemic thrombosis (arterial and venous) and/or recurrent fetal loss mediated by antiphospholipid antibodies: anticardiolipin antibodies, lupus anticoagulant, and/or anti-β-2 glycoprotein I antibodies (antiβ2GPI). In addition to the antibody laboratory findings, classification criteria based on the 2023 American College of Rheumatology and European Alliance of Associations for Rheumatology include clinical features, such as macrovascular venous and arterial thromboses, microvascular thrombotic events, specific obstetric complications, cardiac valve abnormalities, and thrombocytopenia. Being positive for lupus anticoagulant, antiβ2GPI, and anticardiolipin antibodies ("triple positive") is associated with the highest thrombosis risk. APLS is the leading cause of stroke in patients <45 years old.

The vascular compartment within the kidney is the main site of renal involvement. Arteriosclerosis is commonly present in the arcuate and intralobular arteries. In the intralobular arteries, fibrous intimal hyperplasia characterized by intimal thickening secondary to intense myofibroblastic intimal cellular proliferation with extracellular matrix deposition is frequently seen along with onion-skinning. Arterial and arteriolar fibrous and fibrocellular occlusions are present in more than two-thirds of biopsy samples. Cortical necrosis and focal cortical atrophy may result from vascular occlusion. TMA is commonly present in renal biopsies, although signs of MAHA and platelet consumption are usually absent. TMA is especially common in the catastrophic variant of antiphospholipid syndrome, which is a life-threatening syndrome characterized by rapid onset of symptoms, multiorgan failure, and severe thrombotic events.

Treatment entails lifelong anticoagulation. For most individuals with APLS, warfarin is the preferred anticoagulant over direct oral anticoagulants (DOACs), with an international normalized ratio (INR) goal of 2–3. One can consider DOACs in a few, specific instances, such as a single venous thrombosis event or in those intolerant to warfarin, but never in those with arterial thrombotic events. Providers should have clear risk-benefit discussions with the patient if choosing a DOAC over warfarin given inferiority in many instances. In those with arterial thrombotic events, antiplatelet therapy is often a recommended addition to warfarin. Catastrophic APLS management includes anticoagulation, glucocorticoids, plasma exchange, and/or IV immunoglobulin (IVIG). Addition of rituximab and eculizumab can be considered, often in the refractory setting, although ongoing studies are evaluating earlier use of these agents.

HELLP SYNDROME

HELLP (*h*emolysis, *e*levated *l*iver enzymes, *l*ow *p*latelets) syndrome is a dangerous complication of pregnancy associated with microvascular injury. Occurring in 0.2–0.9% of all pregnancies and in 10–20% of women with severe preeclampsia, this syndrome carries a mortality rate of 7.4–34%. Most commonly developing in the third trimester, 20% of cases occur before week 28, and 30% occur postpartum. Although a strong association exists between HELLP syndrome and preeclampsia, nearly 20% of cases are not preceded by recognized preeclampsia. Risk factors include abnormal placentation, family history, and elevated levels of fetal mRNA for FLT1 (vascular endothelial growth factor receptor 1) and endoglin. Patients with HELLP syndrome have higher levels of inflammatory markers (C-reactive protein, IL-1Ra, and IL-6) and soluble HLA-DR than do those with preeclampsia alone.

Renal failure occurs in half of patients with HELLP syndrome, although the etiology is not well understood. Limited data suggest that renal failure is the result of both preeclampsia and acute tubular necrosis. Proteinuria is present in 86–100% of cases. Renal histologic findings are those of TMA with endothelial cell swelling and occlusion of the capillary lumens, but luminal thrombi are typically absent. However, thrombi become more common in severe eclampsia and HELLP syndrome. Although renal failure is common, the organ that defines this syndrome is the liver. Subcapsular hepatic hematomas sometimes produce spontaneous rupture of the liver and can be life-threatening. Neurologic complications such as cerebral infarction, hemorrhage, and edema are other potentially life-threatening complications. Nonfatal complications include placental abruption, permanent vision loss due to Purtscher-like (hemorrhagic and vaso-occlusive vasculopathy) retinopathy, pulmonary edema, bleeding, and fetal demise.

Many features are shared by HELLP syndrome and MAHA. Diagnosis of HELLP syndrome is complicated by the fact that aHUS and TTP also can be triggered by pregnancy; in addition, complement gene mutations and complement pathway dysfunction are common (30–40%) among patients with HELLP syndrome. Patients with antiphospholipid syndrome also have an elevated risk of HELLP syndrome. A history of MAHA before pregnancy is of diagnostic value. Serum levels of ADAMTS13 activity are reduced (by 30–60%) in HELLP syndrome but not to the levels seen in TTP (<10%). Determination of the ratio of lactate dehydrogenase to aspartate aminotransferase may be helpful. This ratio is 13:1 in patients with HELLP syndrome and preeclampsia as opposed to 29:1 in patients without preeclampsia. Other markers, such as antithrombin III (decreased in HELLP syndrome but not in TTP) and D-dimer (elevated in HELLP syndrome but not in TTP), may also be useful. HELLP syndrome usually resolves spontaneously after delivery. Management includes administering magnesium sulfate for seizure prophylaxis and treating hypertension if present. A key component of treatment is delivery; timing of delivery is based on severity of symptoms and age of the fetus. In pregnancies with serious maternal or fetal complications, prompt delivery is recommended. In those without serious complications, the age of the fetus and fetal maturity can help distinguish between prompt delivery or a course of antenatal steroids. Plasma exchange has no benefit in HELLP but may be indicated if TTP has not yet been ruled out. Eculizumab has been reported to be effective in a small number of cases, but dosing, efficacy, and indications remain undetermined.

■ POEMS SYNDROME

POEMS syndrome is a systemic disease characterized by polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, and skin changes, with polyneuropathy and monoclonal gammopathy being mandatory major criteria. Peripheral neuropathy often presents with severe motor-sensory deficit. Three other major criteria, at least one of which is required, are elevated vascular endothelial growth factor (VEGF) levels, evidence of Castleman disease, and sclerotic bone lesions. Patients also commonly have elevated IL-6. Another characteristic is that >95% of monoclonal light chain is of the lambda isotype. IgA also makes up ~50% of the monoclonal proteins involved. Organomegaly can involve any organ and often presents as lymphadenopathy. Other findings can include skin changes, thrombocytosis, polycythemia, papilledema, and volume overload. POEMS syndrome can occur with Castleman disease. In the kidney, the hypertrophy frequently is unilateral. One study suggests the difference in kidney size is due to unilateral contraction; however, a volumetric study showed that enlargement is responsible for the difference in kidney size in some patients. Glomerulomegaly is not uncommon. Lobular appearance, endothelial cell swelling, hypercellularity, mesangiolysis, microaneurysm, and glomerular enlargement are reminiscent of membranoproliferative glomerulonephritis. Most patients present with mild to moderate renal impairment and low-grade proteinuria. Progression to end-stage renal disease is rare.

■ SICKLE CELL NEPHROPATHY

Renal complications in sickle cell disease result from occlusion of the vasa recta in the renal medulla. The low partial pressure of oxygen and high osmolarity predispose to hemoglobin S polymerization and erythrocyte sickling. Sequelae include hyposthenuria, hematuria, and papillary necrosis (which can also occur in sickle trait). The kidney responds by increasing blood flow and glomerular filtration rate mediated by prostaglandins. This dependence on prostaglandins may explain the greater reduction of glomerular filtration rate by nonsteroidal anti-inflammatory drugs in these patients than in others. The glomeruli are typically enlarged. Intracapillary fragmentation and phagocytosis of sickled erythrocytes are thought to be responsible for the membranoproliferative glomerulonephritis–like lesion, and focal segmental glomerulosclerosis is seen in more advanced cases. Screening for sickle cell nephropathy should include blood pressure monitoring, urine studies for albumin-to-creatinine ratio, and metabolic panel. Proteinuria is present in 20–30%, and nephrotic-range proteinuria is associated with progression to renal failure. ACE inhibitors reduce proteinuria, although data are lacking on prevention of renal failure. Patients with sickle cell disease are also more prone to acute renal failure. The cause is thought to reflect microvascular occlusion associated with nontraumatic rhabdomyolysis, high fever, infection, and generalized sickling. Chronic kidney disease from sickle cell nephropathy is present in 25–33% of patients. Despite the frequency of renal disease, hypertension is significantly lower in patients with sickle cell disease compared to the general population of black individuals. Treatment can include medications, such as hydroxyurea, which has been shown to decrease vaso-occlusive events and lessen albuminuria. Two gene therapies for patients with sickle cell who have recurrent vaso-occlusive episodes and are >12 years old were approved in December 2023. Exagamglogene autotemcel uses CRISPR/Cas9 gene editing to modify *BCL11A* and increased fetal hemoglobin, whereas lovotibeglogene autotemcel uses lentiviral technology to produce a modified hemoglobin similar to hemoglobin A. Evaluation of outcomes in these individuals is ongoing.

RENAL VEIN THROMBOSIS

Renal vein thrombosis either can present with flank pain, tenderness, hematuria, rapid decline in renal function, and proteinuria or can be silent. Occasionally, renal vein thrombosis is identified during a workup for pulmonary embolism. The left renal vein is more commonly involved, and two-thirds of cases are bilateral. Etiologies can be divided into three broad categories: endothelial damage, venous stasis, and hypercoagulability. Homocystinuria, endovascular intervention, and surgery can produce vascular endothelial damage. Dehydration, which is more common among male patients, is a common cause of stasis in the pediatric population. Stasis also can result from compression and kinking of the renal veins from retroperitoneal processes such as retroperitoneal fibrosis and abdominal neoplasms. Thrombosis can occur throughout the renal circulation, including the renal veins, with antiphospholipid syndrome. Renal vein thrombosis can also be secondary to nephrotic syndrome, particularly membranous nephropathy. Other hypercoagulable states less commonly associated with renal vein thrombosis include proteins C and S, antithrombin deficiency, factor V Leiden, disseminated malignancy, and oral contraceptives. Severe nephrotic syndrome may also predispose patients to renal vein thrombosis.

Diagnostic screening can be performed with Doppler ultrasonography, which is more sensitive than ultrasonography alone. Computed tomography angiography is almost 100% sensitive. Magnetic resonance angiography is another option but is more expensive. Treatment for renal vein thrombosis consists of anticoagulation and therapy for the underlying cause. Endovascular thrombolysis may be considered in severe cases. Occasionally, nephrectomy may be undertaken for life-threatening complications. In patients who cannot receive anticoagulation, suprarenal inferior vena cava filters can be considered.

COVID-19 AND MICROVASCULAR DISEASE

The COVID-19 pandemic, cause by SARS-CoV-2, has led to many complications, including both macrovascular and microvascular thromboses. SARS-CoV-2 attaches to the cell membrane via interaction of the viral spike protein with the ACE2 protein and then enters the cell via endocytosis, leading to increased cytokine release, inflammatory cell activation, and cellular and tissue damage. Endothelial injury and dysfunction in the setting of COVID-19 have led to many microvascular phenotypes including renal injury, heart disease, liver complications, strokes, and respiratory distress syndrome. In response to SARS-CoV-2, endothelial exocytosis occurs, releasing von Willebrand factor and P-selectin from granules. Autopsy specimens have shown evidence of small-vessel occlusions in the heart, kidneys, and lungs. Certain factors related to microvascular damage, such as ADAMTS13, VEGF-A, and Angpt-2, have been shown to predict mortality. In the renal system, autopsies showed diffuse proximal tubule injury, which is thought to be due to the immune dysregulation and inflammatory cytokines, in addition to poor renal perfusion. In certain instances, viral particles have been observed in the tubular epithelium, although direct infection of SARS-CoV-2 in the kidney remains controversial. Having previous chronic kidney disease or diabetic nephropathy is a risk factor for developing acute kidney injury with COVID-19. Treatment includes antiviral medications and remdesivir, with these medications needing to be adjusted or avoided based on kidney function.

Acknowledgment

Ronald S. Go and Nelson Leung contributed to this chapter in the 21st edition and some material from that chapter has been retained here.

FURTHER READING

Al-Nouri ZL et al: Drug-induced thrombotic microangiopathy: A systematic review of published reports. Blood 125:616, 2015.

Barbhaiya M et al: ACR/EULAR APS Classification Criteria Collaborators. The 2023 ACR/EULAR antiphospholipid syndrome classification criteria. Arthritis Rheumatol 75:1687, 2023.

Brocklebank V et al: Thrombotic microangiopathy and the kidney. Clin J Am Soc Nephrol 13:300, 2018.

Fakhouri F et al: How I diagnose and treat atypical hemolytic uremic syndrome. Blood 141:984, 2023.

George JN, Nester CM: Syndromes of thrombotic microangiopathy. N Engl J Med 371:1847, 2014.

Go RS et al: Thrombotic microangiopathy care pathway: A consensus statement for the Mayo Clinic Complement Alternative Pathway-Thrombotic Microangiopathy (CAP-TMA) Disease-Oriented Group. Mayo Clin Proc 91:1189, 2016.

Liu N et al: Direct promoter repression by BCL11A controls the fetal to adult hemoglobin switch. Cell 2018173:430, 2018.

Zabatta E et al: Therapy of scleroderma renal crisis: State of the art. Autoimmun Rev 17:882, 2018.

330

Nephrolithiasis

Gary C. Curhan

Nephrolithiasis, or kidney stone disease, is a common, painful, and costly condition. Each year, billions of dollars are spent on nephrolithiasis-related activity, with the majority of expenditures on surgical treatment of existing stones. While a stone may form due to crystallization of lithogenic factors in the upper urinary tract, it can subsequently move into the ureter and cause renal colic. Although nephrolithiasis is rarely fatal, patients who have had renal colic report that it is the worst pain they have ever experienced. The evidence on which to base clinical recommendations is not as strong as desired; nonetheless, most experts agree that the recurrence of most, if not all, types of stones can be prevented with careful evaluation and targeted recommendations. Preventive treatment may be lifelong; therefore, an in-depth understanding of this condition must inform the implementation of tailored interventions that are most appropriate for and acceptable to the patient.

There are several types of kidney stones. It is clinically important to identify the stone type, which informs prognosis and selection of the optimal preventive regimen. Calcium oxalate stones are most common (~75%); next, in order, are calcium phosphate (~15%), uric acid (~8%), struvite (~1%), and cystine (<1%) stones. Many stones are a mixture of crystal types (e.g., calcium oxalate and calcium phosphate) and also contain protein in the stone matrix. Rarely, stones are composed of medications, such as acyclovir, atazanavir, and triamterene. Stones that form as a result of an upper tract infection, if not appropriately treated, can have devastating consequences and lead to end-stage renal disease. Consideration should be given to teaching practitioners strategies to prevent recurrence of all stone types and the related morbidity.

EPIDEMIOLOGY

Nephrolithiasis is a global disease. Data suggest an increasing prevalence, likely due to Westernization of lifestyle habits (e.g., dietary changes, increasing body mass index). National Health and Nutrition Examination Survey data for 2007–2016 indicate that up to 20% of men and 10% of women will develop at least one stone during their lifetime. The prevalence is ~50% lower among black individuals than among whites. The incidence of nephrolithiasis (i.e., the rate at which previously unaffected individuals develop their first stone) also varies by age, sex, and race. Among white men, the peak annual incidence is ~3.5 cases/1000 at age 40 and declines to ~2 cases/1000 by age 70. Among white women in their thirties, the annual incidence is ~2.5 cases/1000; the figure decreases to ~1.5/1000 at age 50 and beyond. In addition to the medical costs associated with nephrolithiasis, this condition also has a substantial economic impact, as those affected are often of working age. Once an individual has had a stone, the prevention of a recurrence is essential. Published recurrence rates vary by the definitions and diagnostic methods used. Some reports have relied on symptomatic events, while others have been based on imaging. Most experts agree that radiographic evidence of a second stone should be considered to represent a recurrence, even if the stone has not yet caused symptoms.

ASSOCIATED MEDICAL CONDITIONS

Nephrolithiasis is a systemic disorder. Several conditions predispose to stone formation, including gastrointestinal malabsorption (e.g., Crohn's disease, gastric bypass surgery), primary hyperparathyroidism, obesity, type 2 diabetes mellitus, and distal renal tubular acidosis. A number of other medical conditions are more likely to be present in individuals with a history of nephrolithiasis, including hypertension, gout, cardiovascular disease, cholelithiasis, reduced bone mineral density, and chronic kidney disease.

Although nephrolithiasis does not directly cause upper urinary tract infections (UTIs), a UTI in the setting of an obstructing stone is a urologic emergency ("pus under pressure") and requires urgent intervention to reestablish drainage.

PATHOGENESIS

In the consideration of the processes involved in crystal formation, it is helpful to view urine as a complex solution. A clinically useful concept is *supersaturation* (the point at which the concentration product exceeds the solubility product). However, even though the urine in most individuals is supersaturated with respect to one or more types of crystals, the presence of inhibitors of crystallization prevents the majority of the population from continuously forming stones. The most clinically important inhibitor of calcium-containing stones is urine citrate. While the calculated supersaturation value does not perfectly predict stone formation, it is a useful guide as it integrates the multiple factors that are measured in a 24-h urine collection.

Recent studies have changed the paradigm for the site of initiation of stone formation. Renal biopsies of stone formers have revealed calcium phosphate in the renal interstitium. It is hypothesized that this calcium phosphate deposits at the thin limb of the loop of Henle and then extends down to the papilla and erodes through the papillary epithelium, where it provides a site for deposition of calcium oxalate and calcium phosphate crystals. The majority of calcium oxalate stones grow on calcium phosphate at the tip of the renal papilla (*Randall's plaque*). Tubular plugs of calcium phosphate may be the initiating event in calcium phosphate stone development. Thus, the process of stone formation may begin years before a clinically detectable stone is identified. The processes involved in interstitial deposition are under active investigation.

RISK FACTORS

Risk factors for nephrolithiasis can be categorized as dietary, nondietary, or urinary. These risk factors vary by stone type and clinical characteristics.

Dietary Risk Factors Patients who develop stones often change their diet; therefore, studies that retrospectively assess diet may be hampered by recall bias. Some studies have examined the relation between diet and changes in the lithogenic composition of the urine, often using calculated supersaturation. However, the composition of the urine does not perfectly predict risk, and not all components that modify risk are included in the calculation of supersaturation. Thus, dietary associations are best investigated by prospective studies that examine actual stone formation as the outcome. Dietary factors that are associated with an increased risk of nephrolithiasis include animal protein, oxalate, sodium, sucrose, and fructose. Dietary factors associated with a lower risk include calcium, potassium, phytate, and higher fluid intake.

CALCIUM The role of dietary calcium deserves special attention. Although in the distant past dietary calcium had been suspected of increasing the risk of stone disease, several prospective observational studies and a randomized controlled trial have demonstrated that higher dietary calcium intake is related to a *lower* risk of stone formation. The reduction in risk associated with higher calcium intake may be due to a reduction in intestinal absorption of dietary oxalate that results in lower urine oxalate. Low calcium intake is contraindicated as it increases the risk of stone formation and may contribute to lower bone density in stone formers.

Despite similar bioavailability, supplemental calcium may increase the risk of stone formation. The discrepancy between the risks from dietary calcium and calcium supplements may be due to the timing of supplemental calcium intake or to higher total calcium consumption leading to higher urinary calcium excretion.

OXALATE Urinary oxalate is derived from both endogenous production and absorption of dietary oxalate. Owing to its low and often variable bioavailability, much of the oxalate in food may not be readily absorbed. However, absorption may be higher in stone formers. Although observational studies demonstrate that dietary oxalate is only a weak risk factor for stone formation, urinary oxalate is a strong risk factor for calcium oxalate stone formation, and efforts to avoid high oxalate intake should thus be beneficial.

OTHER NUTRIENTS Several other nutrients have been studied and implicated in stone formation. Higher intake of animal protein may lead to increased excretion of calcium and uric acid as well as to decreased urinary excretion of citrate, all of which increase the risk of stone formation. Higher sodium and sucrose intake increases calcium excretion independent of calcium intake. Higher potassium intake decreases calcium excretion, and many potassium-rich foods increase urinary citrate excretion due to their alkali content. Other dietary factors that have been inconsistently associated with lower stone risk include magnesium and phytate.

Vitamin C supplements are associated with an increased risk of calcium oxalate stone formation in men, possibly owing to raised levels of oxalate in urine. Thus, male calcium oxalate stone formers should be advised to avoid high-dose vitamin C supplements. Although high doses of supplemental vitamin B_6 may be beneficial in selected patients with type 1 primary hyperoxaluria, the risk is not reduced in other patients.

FLUIDS AND BEVERAGES The risk of stone formation increases as urine volume decreases. When the urine output is <1 L/d, the risk of stone formation more than doubles. Fluid intake is the main determinant of urine volume, and the importance of fluid intake in preventing stone formation has been demonstrated in observational studies and in a randomized controlled trial. Observational studies of individual beverages have found that coffee, tea, beer, wine, and orange juice are associated with a reduced risk of stone formation. Sugar-sweetened beverage consumption may increase risk.

Nondietary Risk Factors Age, race, body size, and environment are important risk factors for nephrolithiasis. The incidence of stone disease is highest in middle-aged white men, but stones can form in infants as well as in the elderly. There is geographic variability, with the highest prevalence in the southeastern United States. Weight gain increases the risk of stone formation, and the increasing prevalence of nephrolithiasis in the United States may be due in part to the increasing prevalence of obesity. Environmental and occupational influences that may lead to lower urine volume, such as working in a hot environment or lack of ready access to water or a bathroom, are important considerations. Several observational studies, but not all, have reported that antibiotic use in early adulthood and middle age is associated with higher risk for nephrolithiasis in later life. The largest of these studies found an increased risk for five different classes of antibiotics.

Urinary Risk Factors

URINE VOLUME Lower urine volume results in higher concentrations of lithogenic factors and is a common and readily modifiable risk factor. A randomized trial has demonstrated the effectiveness of higher fluid intake in increasing urine volume and reducing the risk of stone recurrence.

URINE CALCIUM Higher urine calcium excretion increases the likelihood of formation of calcium oxalate and calcium phosphate stones. While the term *hypercalciuria* is often used, there is no cutoff that distinguishes between normal and abnormal urine calcium excretion. In fact, the relation between urine calcium and stone risk appears to be continuous; thus, the use of an arbitrary threshold should be avoided. Levels of urine calcium excretion are higher in individuals with a history of nephrolithiasis; however, the mechanisms remain poorly understood. Greater gastrointestinal calcium absorption is one important contributor, and greater bone turnover (with a resultant reduction in bone mineral density) may be another. Primary renal calcium loss, with lower serum calcium concentrations and elevated serum levels of parathyroid hormone (PTH) (and a normal 25-hydroxy vitamin D level), is rare.

URINE OXALATE Higher urine oxalate excretion increases the likelihood of calcium oxalate stone formation. As for urine calcium, no definition for "abnormal" urine oxalate excretion is widely accepted. Given that the relation between urine oxalate and stone risk is continuous, simple dichotomization of urine oxalate excretion is not helpful in assessing risk. The two sources of urine oxalate are endogenous generation and dietary intake. Dietary oxalate is the major contributor and also the source that can be modified. Notably, higher dietary calcium intake reduces gastrointestinal oxalate absorption and thereby reduces urine oxalate.

URINE CITRATE Urine citrate is a natural inhibitor of calcium-containing stones; thus, lower urine citrate excretion increases the risk of stone formation. Citrate reabsorption is influenced by the intracellular pH of proximal tubular cells. Metabolic acidosis, including that due to higher animal flesh intake, will lead to a reduction in citrate excretion by increasing reabsorption of filtered citrate. However, a notable proportion of patients have lower urine citrate for reasons that remain unclear.

URINE URIC ACID Higher urine levels of uric acid—a risk factor for uric acid stone formation—are found in individuals with excess purine consumption and rare genetic conditions that lead to overproduction of uric acid. This characteristic is not associated with the risk of calcium oxalate stone formation.

URINE pH Urine pH influences the solubility of some crystal types. Uric acid stones form when the urine pH is consistently ≤5.5, whereas calcium phosphate stones are more likely to form when the urine pH is ≥6.5. Cystine is more soluble at higher urine pH. Calcium oxalate stones are not influenced by urine pH.

Genetic Risk Factors The risk of nephrolithiasis is more than twofold greater in individuals with a family history of stone disease. This association is likely due to a combination of genetic predisposition and similar environmental exposures. While a number of rare monogenic disorders cause nephrolithiasis, genome-wide association studies have begun to reveal genetic contributors to common forms of stone disease.

The two most common and well-characterized rare monogenic disorders that lead to stone formation are primary hyperoxaluria and cystinuria. *Primary hyperoxaluria* is an autosomal recessive disorder that causes excessive endogenous oxalate generation by the liver, with consequent calcium oxalate stone formation and crystal deposition in organs. Intraparenchymal calcium oxalate deposition in the kidney can eventually lead to renal failure. *Cystinuria* is an autosomal recessive disorder that causes abnormal reabsorption of filtered basic amino acids. The excessive urinary excretion of cystine, which is poorly soluble, leads to cystine stone formation. Cystine stones are visible on plain radiographs and often manifest as staghorn calculi or multiple bilateral stones. Repeat episodes of obstruction and instrumentation can cause a reduction in the glomerular filtration rate (GFR).

APPROACH TO THE PATIENT

Nephrolithiasis

Evidence-based guidelines for the evaluation and treatment of nephrolithiasis have been published. Although there is limited evidence for several aspects, there are standard approaches to patients with acute and chronic presentations that can reasonably guide the clinical evaluation.

It typically requires weeks to months (and often much longer) for a kidney stone to grow to a clinically detectable size. Although the passage of a stone is a dramatic event, stone formation and growth are characteristically clinically silent. A stone can remain asymptomatic in the kidney for years or even decades before signs (e.g., hematuria) or symptoms (e.g., pain) become apparent. Thus, it is important to remember that the onset of symptoms, typically attributable to a stone moving into the ureter, does not provide insight into when the stone actually formed. The factors that induce stone movement are unknown.

CLINICAL PRESENTATION AND DIFFERENTIAL DIAGNOSIS

There are two common presentations for individuals with an acute stone event: renal colic and painless gross hematuria. *Renal colic* is a misnomer because pain typically does not subside completely; rather, it varies in intensity. When a stone moves into the ureter, the discomfort often begins with a sudden onset of unilateral flank pain.

The intensity of the pain can increase rapidly, and there are no alleviating factors. This pain, which is accompanied often by nausea and occasionally by vomiting, may radiate, depending on the location of the stone. If the stone lodges in the upper part of the ureter, pain may radiate anteriorly; if the stone is in the lower part of the ureter, pain can radiate to the ipsilateral testicle in men or the ipsilateral labium in women. Occasionally, a patient has gross hematuria without pain.

Other diagnoses may be confused with acute renal colic. If the stone is lodged at the right ureteropelvic junction, symptoms may mimic those of acute cholecystitis. If the stone blocks the ureter as it crosses over the right pelvic brim, symptoms may mimic acute appendicitis, whereas blockage at the left pelvic brim may be confused with acute diverticulitis. If the stone lodges in the ureter at the ureterovesical junction, the patient may experience urinary urgency and frequency. In female patients, the latter symptoms may lead to an incorrect diagnosis of bacterial cystitis; the urine will contain red and white blood cells, but the urine culture will be negative. An obstructing stone with proximal infection may present as acute pyelonephritis. A UTI in the setting of ureteral obstruction is a medical emergency that requires immediate restoration of drainage by placement of either a ureteral stent or a percutaneous nephrostomy tube. Other conditions to consider in the differential diagnosis include muscular or skeletal pain, herpes zoster, duodenal ulcer, abdominal aortic aneurysm, gynecologic conditions, ureteral stricture, and ureteral obstruction by materials other than a stone, such as a blood clot or sloughed papilla. Extraluminal processes can lead to ureteral compression and obstruction; however, because of the gradual onset, these conditions do not typically present with renal colic.

DIAGNOSIS AND INTERVENTION

Serum chemistry findings are typically normal, but the white blood cell count may be elevated. Examination of the urine sediment will usually reveal red and white blood cells and occasionally crystals **(Fig. 330-1)**. The absence of hematuria does not exclude a stone, particularly when urine flow is completely obstructed by a stone.

The diagnosis is often made on the basis of the history, physical examination, and urinalysis. Thus, it may not be necessary to wait for radiographic confirmation before treating the symptoms. The diagnosis is confirmed by an appropriate imaging study—preferably noncontrast helical computed tomography (CT), which is highly sensitive and allows visualization of uric acid stones (traditionally considered "radiolucent") **(Fig. 330-2)**. Helical CT detects stones as small as 1 mm that may be missed by other imaging modalities.

Typically, helical CT reveals a ureteral stone or evidence of recent passage (e.g., perinephric stranding or hydronephrosis), whereas a plain abdominal radiograph (kidney/ureter/bladder [KUB]) can miss a stone in the ureter or kidney, even if it is radiopaque, and does not provide information on obstruction. Abdominal ultrasound offers the advantage of avoiding radiation and provides information on hydronephrosis, but it is not as sensitive as CT and images only the kidney and possibly the proximal segment of the ureter; thus, most ureteral stones are not detectable by ultrasound.

Many patients who experience their first episode of colic seek emergent medical care. Randomized trials have demonstrated that parenterally administered nonsteroidal anti-inflammatory drugs (e.g., ketorolac) are just as effective as opioids in relieving symptoms and have fewer side effects. Excessive fluid administration has not been shown to be beneficial; therefore, the goal should be to maintain euvolemia. If the pain can be adequately controlled and the patient is able to take fluids orally, hospitalization can be avoided. Use of an alpha blocker may increase the rate of spontaneous stone passage.

Urologic intervention should be postponed unless there is evidence of UTI, a low probability of spontaneous stone passage (e.g., a stone measuring ≥6 mm or an anatomic abnormality), or intractable pain. A ureteral stent may be placed cystoscopically, but this procedure typically requires general anesthesia, and the stent can be quite uncomfortable, may cause gross hematuria, and may increase the risk of UTI.

If an intervention is indicated, the selection of the most appropriate intervention is determined by the size, location, and composition of the stone; the urinary tract anatomy; and the experience of the urologist. Extracorporeal shockwave lithotripsy (ESWL), the least invasive option, uses shockwaves generated outside the body to fragment the stone, but is being used less frequently. An endourologic approach, now more frequently used than ESWL,

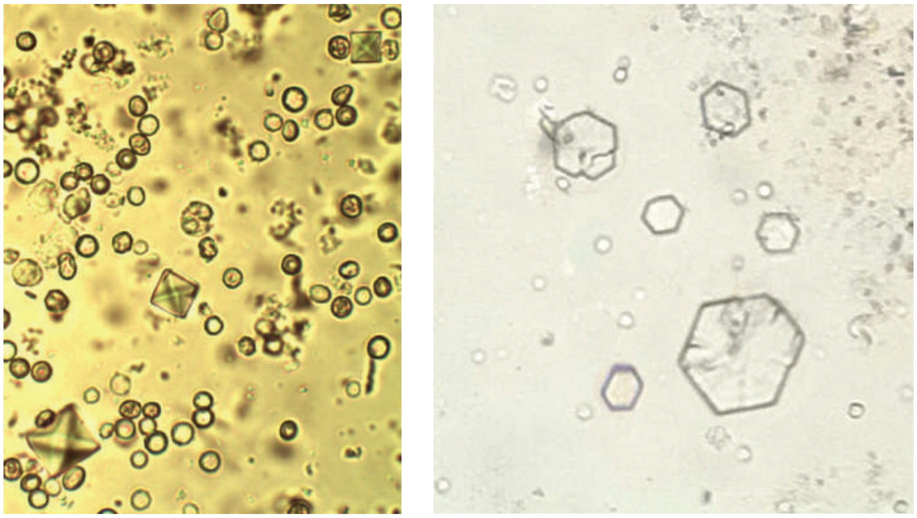

FIGURE 330-1 **Urine sediment from a patient with calcium oxalate stones (*left*) and a patient with cystine stones (*right*).** Calcium oxalate dihydrate crystals are bipyramidally shaped, and cystine crystals are hexagonal. *(Left panel image courtesy of Dr. Mark Perazella, Yale School of Medicine. Right panel image courtesy Dr. John Lieske, Mayo Clinic.)*

can remove a stone by basket extraction or laser fragmentation. For large upper-tract stones, percutaneous nephrostolithotomy has the highest likelihood of rendering the patient stone-free. Advances in urologic approaches and instruments have nearly eliminated the need for open surgical procedures such as ureterolithotomy or pyelolithotomy.

EVALUATION FOR STONE PREVENTION

More than half of first-time stone formers will have a recurrence within 10 years. A careful evaluation is indicated to identify predisposing factors, which can then be modified to reduce the risk of new stone formation or growth of existing renal stones. It is appropriate to proceed with an evaluation even after the first stone if the patient is interested because recurrences are common and are usually preventable with inexpensive lifestyle modifications or other treatments.

HISTORY

A detailed history, obtained from the patient and from a thorough review of medical records, should include the number and frequency of episodes (distinguishing stone passage from stone formation) and previous imaging studies, interventions, evaluations, and treatments. Inquiries about the patient's medical history should cover UTIs, gastric bypass surgery and other malabsorptive conditions, gout, hypertension, and diabetes mellitus. A family history of stone disease may reveal a genetic predisposition. A complete list of current prescription and over-the-counter medications as well as vitamin and mineral supplements is essential. The review of systems should focus on identifying possible etiologic factors related to low urine volume (e.g., high insensible losses) and gastrointestinal malabsorption as well as on ascertaining how frequently the patient voids during the day and overnight.

A large body of compelling evidence has demonstrated the important role of diet in stone disease. Thus, the dietary history should encompass information on usual dietary habits (meals and snacks), calcium intake, consumption of high-oxalate foods (spinach, rhubarb, potatoes), and fluid intake (including amount of specific beverages typically consumed). Amount and frequency of use of vitamin and mineral supplements should be carefully assessed.

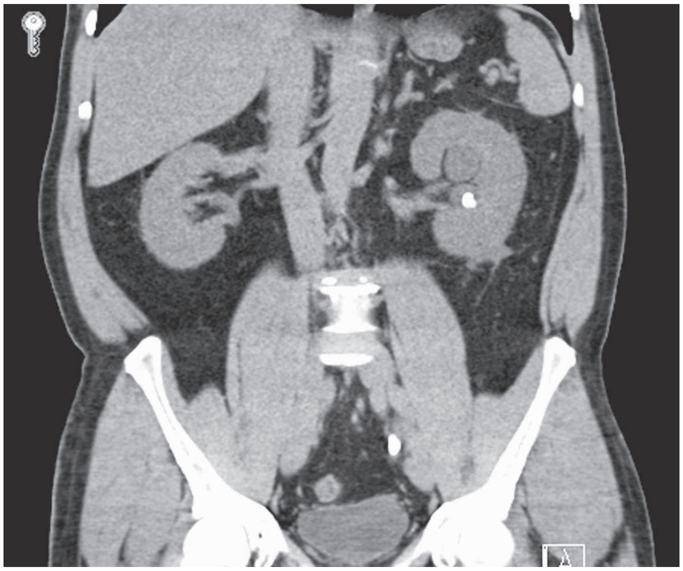

FIGURE 330-2 **Coronal noncontrast CT image from a patient who presented with left-sided renal colic**. An obstructing calculus, present in the distal left ureter at the level of S1, measures 10 mm in maximal dimension. There is severe left hydroureteronephrosis and associated left perinephric fat stranding. In addition, there is a nonobstructing 6-mm left renal calculus in the interpolar region. *(Image courtesy of Dr. Stuart Silverman, Brigham and Women's Hospital.)*

PHYSICAL EXAMINATION

The physical examination should assess weight, blood pressure, costovertebral angle tenderness, and lower-extremity edema as well as signs of other systemic conditions such as primary hyperparathyroidism and gout.

LABORATORY EVALUATION

If not recently measured, the following serum levels should be determined: electrolytes (to uncover hypokalemia or renal tubular acidosis), creatinine, calcium, phosphorus, and uric acid. The PTH level should be measured if indicated by elevated, high-normal, or low-normal serum and urine calcium concentrations. 25-Hydroxy vitamin D should be measured in concert with PTH to investigate the possible role of secondarily elevated PTH levels in the setting of vitamin D insufficiency.

The urinalysis, including examination of the sediment, can provide useful information. In individuals with asymptomatic residual renal stones, red and white blood cells are frequently present in urine. If there is concern about the possibility of an infection, a urine culture should be performed. The sediment may also reveal crystals (Fig. 330-1), which may help identify the stone type and also provide prognostic information as crystalluria is a strong risk factor for new stone formation.

The results from 24-h urine collections serve as the cornerstone on which therapeutic recommendations are based. Recommendations on lifestyle modification should be deferred until urine collection is complete. As a baseline assessment, patients should collect at least two 24-h urine samples while consuming their usual diet and usual volume of fluid. The following factors should be measured: total volume, calcium, oxalate, citrate, uric acid, sodium, potassium, phosphorus, magnesium, pH, and creatinine. When available, the calculated supersaturation is also informative. There is substantial day-to-day variability in the 24-h excretion of many relevant factors; therefore, obtaining values from two collections is important before committing a patient to long-term lifestyle changes or medication. The interpretation of the 24-h urine results should take into account that the collections are usually performed on a weekend day when the patient is staying at home; an individual's habits may differ dramatically (beneficially or detrimentally) at work or outside the home. Specialized testing, such as calcium loading or restriction, is not recommended as it does not influence clinical recommendations.

Stone composition analysis is essential if a stone or fragment is available; patients should be encouraged to retrieve passed stones. The stone type cannot be determined with certainty from 24-h urine results, but pure uric acid stones can be identified by low Hounsfield units on CT.

IMAGING

The "gold standard" diagnostic test is helical CT without contrast. If not already performed during an acute episode, a low-dose renal-limited noncontrast CT should be considered to establish definitively the baseline stone burden. A suboptimal imaging study may not detect a residual stone that, if subsequently passed, would be mistaken for a new stone. In this instance, the preventive medical regimen might be unnecessarily changed as the result of a preexisting stone.

Recommendations for follow-up imaging should be tailored to the individual patient. While CT provides the best information, the radiation dose is higher than from other modalities; therefore, CT should be performed only if the results will lead to a change in clinical recommendations. Although less sensitive, renal ultrasound is typically used to minimize radiation exposure, with recognition of the limitations.

PREVENTION OF NEW STONE FORMATION

Recommendations for preventing stone formation depend on the stone type and the results of metabolic evaluation. After remediable secondary causes of stone formation (e.g., primary hyperparathyroidism) are excluded, the focus should turn to modification of the urine composition to reduce the risk of new stone formation. The urinary constituents and calculated urine supersaturation are continuous variables, and the associated risk is continuous; thus, there are no definitive thresholds. Dichotomization into "normal" and "abnormal" can be misleading and should be avoided.

For all stone types, consistently diluted urine reduces the likelihood of crystal formation. The urine volume should be at least 2 L/d. Because of differences in insensible fluid losses and fluid intake from food sources, the required total fluid intake will vary from person to person. Rather than specify how much to drink, it is more helpful to educate patients about how much *more* they need to drink in light of their 24-h urine volume. For example, if the daily urine volume is 1.5 L, then the patient should be advised to drink consistently at least 0.5 L more per day in order to increase the urine volume to the goal of 2 L/d.

RECOMMENDATIONS FOR SPECIFIC STONE TYPES

Calcium Oxalate Risk factors for calcium oxalate stones include higher urine calcium, higher urine oxalate, and lower urine citrate. This stone type is insensitive to urine pH in the physiologic range.

Individuals with higher urine calcium excretion tend to absorb a higher percentage of ingested calcium. Nevertheless, dietary calcium restriction is not beneficial and, in fact, is likely to be harmful (see "Dietary Risk Factors," above). In a randomized trial in men with high urine calcium and recurrent calcium oxalate stones, a diet containing 1200 mg of calcium and a low intake of sodium and animal protein significantly reduced subsequent stone formation compared with a low-calcium diet (400 mg/d). Excessive calcium intake (>1200 mg/d) should be avoided.

A thiazide diuretic, in doses higher than those used to treat hypertension, can substantially lower urine calcium excretion. Several randomized controlled trials, but not all, have demonstrated that thiazide diuretics can reduce calcium oxalate stone recurrence by ~50%. When a thiazide is prescribed, dietary sodium restriction is essential to obtain the desired reduction in urinary calcium excretion and minimize urinary potassium losses. While bisphosphonates may reduce urine calcium excretion in some individuals, there are only observational data to suggest whether this class of medication can reduce stone formation; therefore, bisphosphonates cannot be recommended solely for stone prevention at present, but they can be used to treat those individuals with low bone density.

A reduction in urine oxalate will, in turn, reduce the supersaturation of calcium oxalate. In patients with the common form of nephrolithiasis, avoiding high-dose vitamin C supplements is the only known strategy that reduces endogenous oxalate production. In patients with primary hyperoxaluria (type I), an siRNA-based therapy (lumasiran, nedosiran) can reduce oxalate generation in the liver.

Oxalate is a metabolic end product; therefore, any dietary oxalate that is absorbed will be excreted in the urine. Reducing absorption of exogenous oxalate involves two approaches. First, the avoidance of foods that contain high amounts of oxalate, such as spinach, rhubarb, almonds, and potatoes, is prudent. However, extreme oxalate restriction has not been demonstrated to reduce stone recurrence and could be harmful to overall health, given other health benefits of many foods that are erroneously considered to be high in oxalate. Second, the absorption of oxalate is reduced by higher calcium intake; therefore, individuals with higher-than-desired urinary oxalate should be counseled to consume adequate calcium. Oxalate absorption can be influenced by the intestinal microbiota, depending on the presence of oxalate-degrading bacteria. Currently, however, there are no available therapies to alter the microbiota that beneficially affect urinary oxalate excretion over the long term.

Citrate is a natural inhibitor of calcium oxalate and calcium phosphate stones. Increased consumption of foods rich in alkali (i.e., fruits and vegetables) can increase urine citrate. For patients with lower urine citrate in whom dietary modification does not adequately increase urine citrate, the addition of supplemental alkali (typically potassium citrate or bicarbonate) will lead to an increase in urinary citrate excretion. Sodium salts, such as sodium bicarbonate, while successful in raising urine citrate, are typically avoided due to the adverse effects of sodium on urine calcium excretion.

Past reports suggested that higher levels of urine uric acid may increase the risk of calcium oxalate stones, but more recent studies do not support this association. However, allopurinol reduced stone recurrence in one randomized controlled trial in patients with calcium oxalate stones and high urine uric acid levels. The lack of association between urine uric acid level and calcium oxalate stones suggests that a different mechanism underlies the observed beneficial effect of allopurinol in this setting.

Additional dietary modifications may be beneficial in reducing stone recurrence. Restriction of nondairy animal protein (e.g., meat, chicken, seafood) is a reasonable approach and may result in higher excretion of citrate and lower excretion of calcium. In addition, reducing sodium intake to <2.5 g/d may decrease urinary excretion of calcium. Sucrose and fructose intake should be minimized.

For adherence to a dietary pattern that is more manageable for patients than manipulating individual nutrients, the Mediterranean diet or the DASH (Dietary Approaches to Stop Hypertension) diet provides an appropriate and readily available option. Randomized trials have conclusively shown the DASH diet to reduce blood pressure. At present, only data from observational studies are available, but these demonstrate a strong and consistent inverse association between these healthy dietary patterns and risk of stone formation.

Calcium Phosphate Calcium phosphate stones share risk factors with calcium oxalate stones, including higher concentrations of urine calcium and lower concentrations of urine citrate, but additional factors deserve attention. Higher urine phosphate levels and higher urine pH (typically ≥6.5) are associated with an increased likelihood of calcium phosphate stone formation. Calcium phosphate stones are more common in patients with distal renal tubular acidosis and primary hyperparathyroidism.

There are no randomized trials on which to base preventive recommendations for calcium phosphate stone formers, so the interventions are focused on modification of the recognized risk factors. Thiazide diuretics (with sodium restriction) may be used to reduce urine calcium, as described above for calcium oxalate stones. In patients with low urine citrate levels, alkali supplements (e.g., potassium citrate or bicarbonate) may be used to increase urine citrate. However, the urine pH of these patients should be monitored initially because supplemental alkali can raise urine pH, thereby potentially increasing the risk of stone formation. Because these patients tend to have a urinary acidification defect, reducing the urine pH is not an option. Reduction of dietary phosphate may be beneficial by reducing urine phosphate excretion.

Uric Acid The two main risk factors for uric acid stones are persistently low urine pH and higher uric acid excretion. Urine pH is the predominant influence on uric acid solubility; therefore, the mainstay of prevention of uric acid stone formation entails increasing urine pH. Alkalinizing the urine can be achieved by increasing the intake of foods rich in alkali (e.g., fruits and vegetables) and reducing the intake of foods that produce acid (e.g., animal flesh). If necessary, supplementation with bicarbonate or citrate salts (preferably potassium-based) can be used to reach the recommended pH goal of 6.5 throughout the day and night.

Urine uric acid excretion is determined by uric acid generation. Uric acid is the end product of purine metabolism; thus, reduced consumption of purine-containing foods can lower urine uric acid excretion. It is noteworthy that the *serum* uric acid level is dependent on the fractional excretion of uric acid and, therefore, does not provide information on *urine* uric acid excretion. For example, an individual with high uric acid generation and concurrent high fractional excretion of uric acid will have high urine uric acid excretion with a normal (or even low) serum uric acid level. If alkalinization of the urine alone is not successful and if dietary modifications do not reduce urine uric acid sufficiently, then the addition of a xanthine oxidase inhibitor, such as allopurinol or febuxostat, can reduce urine uric acid excretion by 40–50%.

Cystine Cystine excretion is not easily modified. Long-term dietary cystine restriction is not feasible and is unlikely to be successful; thus, the focus for cystine stone prevention is on increasing cystine solubility. This goal may be achieved by treatment with medication that covalently binds to cystine (tiopronin or penicillamine) and a medication that raises urine pH. Tiopronin is the preferred choice owing to its better adverse event profile. The preferred alkalinizing agent to achieve a urine pH of 7.5 is potassium citrate or bicarbonate as sodium salts may increase cystine excretion. As with all stone types, and especially in patients with cystinuria, maintaining a high urine volume is an essential component of the preventive regimen.

Struvite Struvite stones, also known as *infection stones* or *triple-phosphate stones*, form only when the upper urinary tract is infected with urease-producing bacteria such as *Proteus mirabilis, Klebsiella pneumoniae*, or *Providencia* species. Urease produced by these bacteria hydrolyzes urea and may elevate the urine pH to a supraphysiologic level (>8.0). Struvite stones may grow quickly and fill the renal pelvis (*staghorn calculi*).

Struvite stones require complete removal by a urologist. New stone formation can be avoided by the prevention of UTIs. In patients with recurrent upper UTIs (e.g., some individuals with surgically altered urinary drainage or spinal cord injury), the urease inhibitor acetohydroxamic acid can be considered; however, this agent should be used with caution because of potential side effects.

LONG-TERM FOLLOW-UP

In general, the preventive regimens described above do not cure the underlying pathophysiologic process. Thus, these recommendations typically need to be followed for the patient's lifetime, and it is essential to tailor recommendations in a way that is acceptable to the patient. Because the memory of the acute stone event fades and patients often return to old habits (e.g., insufficient fluid intake), long-term follow-up, including repeat 24-h urine collections typically annually, is important to ensure that the preventive regimen has been implemented and has resulted in the desired reduction in the risk of new stone formation.

Follow-up imaging should be planned thoughtfully. Many patients with recurrent episodes of renal colic that lead to emergency room visits often undergo repeat CT studies. While CT does provide the best information, the radiation dose is higher than that with plain abdominal radiography (KUB). Small stones may be missed by KUB, and ultrasound has a limited ability to determine the size and number of stones. Minimizing radiation exposure should be a goal of the long-term follow-up plan and must be balanced against the gain in diagnostic information.

FURTHER READING

Coe FL: The physiology of kidney stone prevention: A narrative for patients and physicians. Available at *https://bpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/c/3637/files/2022/10/KIDNEY-STONES_-Pathogenesis-and-Prevention.pdf*. Accessed November 2, 2023.

Coe FL, Worcester EM, Evan AP, Lingeman JE (eds): *Kidney Stones: Medical and Surgical Management*, 2nd ed. New Dehli, Jaypee Brothers Medical Publishing, 2019.

European Association of Urology Guidelines on Urolithiasis. Available at *https://d56bochluxqnz.cloudfront.net/documents/full-guideline/EAU-Guidelines-on-Urolithiasis-2023.pdf*. Accessed November 22, 2023.

Pearle MS et al: Medical management of kidney stones: AUA guideline. J Urol 192:316, 2014.

331

Urinary Tract Obstruction

Julian L. Seifter

Obstruction to the flow of urine, with attendant stasis and elevation in urinary tract pressure, impairs renal and urinary conduit functions and is a common cause of acute and chronic kidney disease (obstructive nephropathy). Early recognition and prompt treatment of urinary tract obstruction (UTO) can prevent or reverse devastating effects on kidney structure and function, and decrease susceptibility to hypertension, infection, and stone formation. Chronic obstruction may lead to permanent loss of renal mass (renal atrophy) and excretory capability. Because obstructive disease may be secondary to serious underlying inflammatory, vascular, or malignant disease, familiarity with clinical findings, appropriate diagnostic testing, and the therapeutic approach is of great importance to the clinician.

Recent developments have advanced our knowledge of fetal and genetic causes of lower urinary tract obstruction (LUTO) and chronic anomalies of the kidneys and urinary tracts (CAKUT), the rubric for a cluster of developmental uropathies associated with cysts and obstruction. In the last few years, studies in mice, zebrafish, and pluripotent stem cells (kidney organoids) have identified possible genetic mutations implicated in cases of CAKUT producing low nephron number, including duplications and deletions (copy number variants) and familial candidate genes (*HFN1β*, *PAX2*, *SIX1*, *GATA3*, and many others). Given that there is no nephrogenesis in humans after birth, obstructive disease manifesting in early or middle adulthood may in fact have origins in a decreased number of nephrons during fetal development. Fetal LUTO, if untreated, often results in pulmonary hypoplasia and death in infancy; kidney failure may be a more long-term outcome. Since amniotic fluid is produced by fetal ingestion of urine, LUTO should be suspected when the mother develops oligohydramnios or anhydramnios. Improvements in second-trimester diagnosis and treatment with vesicoamniotic shunting have improved neonatal survival.

One of the most significant concerns in renal medicine is progression to a chronically fibrotic failing kidney. To date, there has been no demonstration of reversal of scarring, but many investigations are underway to find antifibrosis drugs, including those that address fibrosis in genetic diseases that lead to obstructive nephropathy. Many mechanisms contribute to this progression, including inflammatory responses, hemodynamic consequences of the declining number of healthy nephrons, and compression of normal nephrons by cysts in genetic diseases such as autosomal dominant polycystic kidney disease (ADPKD), where a gene mutation in polycystin causes proliferation of cystic masses. In ADPKD, primary cilia of tubule epithelial cells have been implicated in the development of cysts, resulting in inflammatory processes and cell death. Polycystic disease is one of several such ciliopathies that ultimately lead to fibrosis.

ETIOLOGY

Obstruction to urine flow can result from *intrinsic* or *extrinsic mechanical blockade* as well as from *functional defects* not associated with fixed occlusion of the urinary drainage system. Mechanical obstruction can occur at any level of the urinary tract, from within the renal tubules or the renal calyces to the external urethral meatus (obstructive uropathy). Normal points of narrowing, such as the ureteropelvic and ureterovesical junctions, bladder neck, and urethral meatus, are common sites of obstruction. When lower UTO is above the level of the bladder, unilateral dilatation of the ureter (*hydroureter*) and renal pyelocalyceal system (*hydronephrosis*) occurs; lesions at or below the level of the bladder cause bilateral involvement.

Common forms of obstruction are listed in Table 331-1. Childhood causes include *congenital malformations*, such as narrowing of the ureteropelvic junction (UPJ) and abnormal insertion of the ureter into the bladder, the most common cause. Vesicoureteral reflux in the absence of urinary tract infection or bladder neck obstruction often resolves with age. Reinsertion of the ureter into the bladder is indicated if reflux is severe and unlikely to improve spontaneously, if renal function deteriorates, or if urinary tract infections recur despite chronic antimicrobial therapy. Vesicoureteral reflux may cause prenatal hydronephrosis and, if severe, can lead to recurrent urinary infections, hypertension, and renal scarring in childhood. Posterior urethral valves are the most common cause of bilateral hydronephrosis in boys. In adults, UTO has usually been attributed to *acquired defects*. Pelvic tumors, calculi, and urethral stricture predominate. Ligation of, or injury to, the ureter during pelvic or colonic surgery can lead to hydronephrosis, which, if unilateral, may remain undetected. Obstructive uropathy may also result from extrinsic neoplastic (carcinoma of cervix or colon) or inflammatory disorders. Lymphomas, particularly follicular, and pelvic or colonic neoplasms with retroperitoneal involvement are causes of ureteral obstruction. As many as 50% of men aged >40 years may have lower urinary tract symptoms associated with benign prostatic hypertrophy, but these symptoms may occur without bladder outlet obstruction. Regardless of the primary cause, the chronic kidney disease that results may be exacerbated by a secondary contributor like diabetes mellitus, hypertension, kidney stones, infection, and metabolic disturbances.

Functional impairment of urine flow occurs when voiding is altered by abnormal pontine or sacral centers of micturition control. It may be asymptomatic or associated with lower urinary tract symptoms such as frequency, urgency, and postmicturition incontinence, nocturia, straining to void, slow stream, hesitancy, or a feeling of incomplete emptying. A history should be sought for trauma, back injury, surgery, diabetes mellitus, neurologic or psychiatric conditions, and medications. Causes include neurogenic bladder, often with adynamic ureter, and vesicoureteral reflux. Reflux in children may result in severe unilateral or bilateral hydroureter and hydronephrosis. Overflow urinary incontinence combined with sudden-onset fecal incontinence, severe lower back pain, and saddle anesthesia, requires emergency evaluation for possible cauda equina syndrome. Urinary retention may be the consequence of α-adrenergic and anticholinergic agents, as well as opiates. Hydronephrosis in pregnancy is due to relaxational effects of progesterone on smooth muscle of the renal pelvis, as well as ureteral compression by the enlarged uterus, more often on the right side.

TABLE 331-1 Common Mechanical Causes of Urinary Tract Obstruction

URETER	BLADDER OUTLET	URETHRA
Congenital		
Ureteropelvic junction narrowing or obstruction Ureterovesical junction narrowing or obstruction and reflux Ureterocele Retrocaval ureter	Bladder neck obstruction Ureterocele	Posterior urethral valves Anterior urethral valves Stricture Meatal stenosis Phimosis
Acquired Intrinsic Defects		
Calculi Inflammation Infection Trauma Sloughed papillae Tumor Blood clots	Benign prostatic hyperplasia Cancer of prostate Cancer of bladder Calculi Diabetic neuropathy Spinal cord disease Anticholinergic drugs and α-adrenergic agonists	Stricture Tumor Calculi Trauma Phimosis
Acquired Extrinsic Defects		
Pregnant uterus Retroperitoneal fibrosis Aortic aneurysm Uterine leiomyomata Carcinoma of uterus, prostate, bladder, colon, rectum Lymphoma Pelvic inflammatory disease, endometriosis Accidental surgical ligation	Carcinoma of cervix, colon Trauma	Trauma

Diagnostic tools to identify anatomic obstruction include urinary flow measurements and a postvoid residual measurement. Bladder volume may be readily assessed by bedside ultrasound. Cystourethroscopy and urodynamic studies may be reserved for the symptomatic patient to assess the filling phase (cystometry), pressure-volume relationship of the bladder, bladder compliance, and capacity. Pressure-flow analysis evaluates bladder contractility and bladder outlet resistance during voiding. Bladder obstruction is characterized by high pressures in women, whereas in men, a diagnosis of bladder outlet obstruction is based on flow rate and voiding pressures. A voiding cystourethrogram may be useful in evaluating incomplete emptying and bladder neck and urethral pathology.

CLINICAL FEATURES AND PATHOPHYSIOLOGY

The pathophysiology and clinical features of UTO are summarized in Table 331-2. Flank pain, the symptom that most commonly leads to medical attention, is due to distention of the collecting system or renal capsule. Pain severity is influenced more by the rate at which distention develops than by the degree of distention. Acute supravesical

TABLE 331-2 Pathophysiology of Bilateral Ureteral Obstruction

HEMODYNAMIC EFFECTS	TUBULE EFFECTS	CLINICAL FEATURES
Acute		
↑ Renal blood flow ↓ GFR ↓ Medullary blood flow ↑ Vasodilator prostaglandins, nitric oxide	↑ Ureteral and tubule pressures ↑ Reabsorption of Na^+, urea, water	Pain (capsule distention) Azotemia, oliguria, or anuria
Chronic		
↓ Renal blood flow ↓↓ GFR ↑ Vasoconstrictor prostaglandins ↑ Renin-angiotensin production	↓ Medullary osmolarity ↓ Concentrating ability Structural damage; parenchymal atrophy ↓ Transport functions for Na^+, K^+, H^+	Azotemia Hypertension AVP-insensitive polyuria Natriuresis Hyperkalemic, hyperchloremic acidosis
Release of Obstruction		
Slow ↑ in GFR (variable)	↓ Tubule pressure ↑ Solute load per nephron (urea, NaCl) Natriuretic factors present	Postobstructive diuresis Potential for volume depletion and electrolyte imbalance due to losses of Na^+, K^+, PO_4^{2-}, Mg^{2+}, and water

Abbreviations: AVP, arginine vasopressin; GFR, glomerular filtration rate.

obstruction, as from a stone lodged in a ureter **(Chap. 330)**, is associated with excruciating, sometimes intermittent, pain, known as *renal colic*. This pain often radiates to the lower abdomen, testes, or labia. By contrast, more insidious causes of obstruction, such as chronic narrowing of the UPJ, may produce little or no pain and yet result in total destruction of the affected kidney. Flank pain that occurs only with micturition is pathognomonic of vesicoureteral reflux.

Obstruction of urine flow results in an increase in hydrostatic pressures proximal to the site of obstruction. It is this buildup of pressure that leads to the accompanying pain, the distention of the collecting system in the kidney, and elevated intratubular pressures that initiate tubular dysfunction. In the first days of obstruction, the dilatation of the poorly compliant collecting system may be minimal. As the increased hydrostatic pressure is expressed in the urinary space of the glomeruli, further filtration decreases or stops completely.

Azotemia develops when overall excretory function is impaired, often in the setting of bladder outlet obstruction, bilateral renal pelvic or ureteric obstruction, or unilateral disease in a patient with a solitary functioning kidney. Complete bilateral obstruction should be suspected when acute renal failure is accompanied by anuria. Any patient with renal failure otherwise unexplained, or with a history of nephrolithiasis, hematuria, diabetes mellitus, prostatic enlargement, pelvic surgery, trauma, or tumor should be evaluated for UTO.

In the acute setting, partial, bilateral obstruction may mimic prerenal azotemia with a high blood urea nitrogen–to–creatinine ratio, concentrated urine, and sodium retention. Renal vascular resistance may be increased. However, with more prolonged obstruction, symptoms of *polyuria* and *nocturia* commonly accompany partial UTO and result from loss of medullary hypertonicity with diminished renal concentrating ability. Failure to produce urine free of salt (natriuresis) is due to downregulation of salt reabsorption in the proximal tubule and of transport proteins including the Na^+, K^+ adenosine triphosphatase (ATPase), Na:K:2Cl cotransporter (NKCC2) in the thick ascending limb, and the epithelial Na^+ channel (ENaC) in collecting duct cells. In addition to direct effects on renal transport mechanisms, increased prostaglandin E_2 (PGE_2) (due to induction of cyclooxygenase-2 [COX-2]), angiotensin II (with its downregulation of Na^+ transporters), and atrial or B-type natriuretic peptides (ANP or BNP) due to volume expansion in the azotemic patient contribute to decreased salt reabsorption along the nephron. Nitric oxide synthases (NOS) in ureteral smooth muscle and urothelial tissues have been found to oppose the high ureteral pressure in unilateral obstruction.

Dysregulation of aquaporin-2 water channels in the collecting duct contributes to the polyuria. The defect usually does not improve with administration of vasopressin and is, therefore, a form of acquired nephrogenic diabetes insipidus.

Wide fluctuations in urine output in a patient with azotemia should always raise the possibility of intermittent or partial UTO. If fluid intake is inadequate, severe dehydration and hypernatremia may develop. However, as with other causes of poor renal function, excesses of salt and water intake may result in edema and hyponatremia.

Partial bilateral UTO often results in *acquired distal renal tubular acidosis, hyperkalemia*, and *renal salt wasting*. The H^+-ATPase, situated on the apical membrane of the α-intercalated cells of the collecting duct, is critical for distal H^+ secretion. The trafficking of intracellular H^+ pumps from the cytoplasm to the cell membrane is disrupted in UTO. The decreased function of the ENaC, in the apical membrane of neighboring collecting duct principal cells, contributes to decreased Na^+ reabsorption (salt wasting) and, therefore, decreased K^+ secretion via K^+ channels. Ammonium (NH^+) excretion important to the elimination of H^+ is impaired. These defects in tubule function are often accompanied by renal tubulointerstitial damage. Azotemia with hyperkalemia and metabolic acidosis should prompt consideration of UTO.

The renal interstitium becomes edematous and infiltrated with mononuclear inflammatory cells early in UTO. Later, interstitial fibrosis and atrophy of the papillae and medulla occur and precede these processes in the cortex. The increase in angiotensin II noted in UTO contributes to the inflammatory response and fibroblast accumulation through mechanisms involving profibrotic cytokines. With time, this process leads to chronic kidney damage.

UTO must always be considered in patients with urinary tract infections or urolithiasis. Urinary stasis encourages the growth of organisms. Urea-splitting bacteria are associated with magnesium ammonium phosphate (struvite) calculi that may take on a staghorn appearance. *Hypertension* is frequent in acute and subacute unilateral obstruction and is usually a consequence of increased release of renin by the involved kidney. Chronic kidney disease from bilateral UTO, often associated with extracellular volume expansion, may result in significant hypertension. *Erythrocytosis*, an infrequent complication of obstructive uropathy, is secondary to increased erythropoietin production.

DIAGNOSIS

A history of difficulty in voiding, pain, infection, or change in urinary volume is common. Evidence for distention of the kidney or urinary bladder can often be obtained by palpation and percussion of the abdomen. A careful rectal and genital examination may reveal enlargement or nodularity of the prostate, abnormal rectal sphincter tone, or a rectal or pelvic mass.

Urinalysis may reveal hematuria, pyuria, and bacteriuria. The urine sediment is often normal, even when obstruction leads to marked azotemia and extensive structural damage. An abdominal scout film, although insensitive, may detect nephrocalcinosis or a radiopaque stone. As indicated in **Fig. 331-1**, if UTO is suspected, a bladder catheter should be inserted. Abdominal ultrasonography should be performed to evaluate renal and bladder size, as well as pyelocalyceal contour. Ultrasonography is ~90% specific and sensitive for detection of hydronephrosis. False-positive results are associated with diuresis, renal cysts, or the presence of an extrarenal pelvis, a normal congenital variant. Congenital UPJ obstruction may be mistaken for renal cystic disease. Hydronephrosis may be absent on ultrasound when obstruction is <48 h in duration or associated with volume contraction, staghorn calculi, retroperitoneal fibrosis, or infiltrative renal disease. Duplex Doppler ultrasonography may detect an increased resistive index in urinary obstruction. The radiodensity of renal stones is measured in Hounsfield units (HUs). Calcium oxalate and calcium phosphate stones may be distinguished from uric acid calculi by their HU radiodensities on noncontrast computed tomography (CT) scan, the test of choice for acute flank pain.

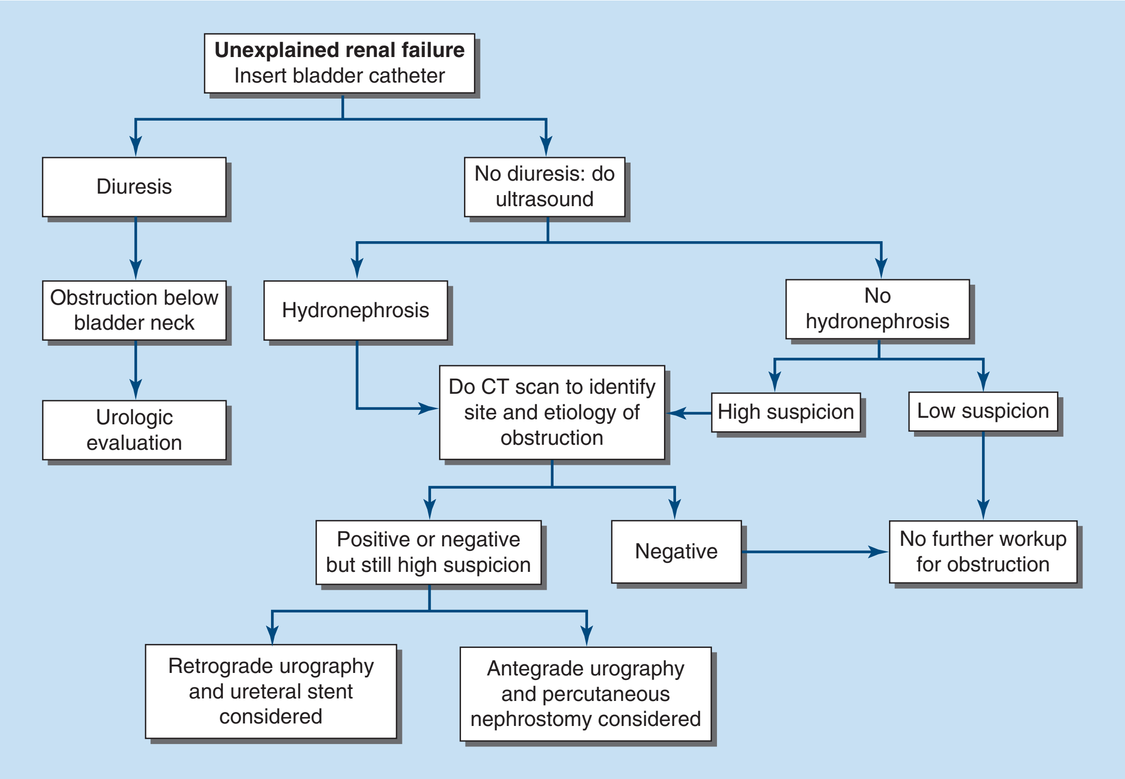
Unexplained renal failure
Insert bladder catheter
Diuresis
No diuresis: do ultrasound
Obstruction below bladder neck
Urologic evaluation
Hydronephrosis
No hydronephrosis
Do CT scan to identify site and etiology of obstruction
High suspicion
Low suspicion
Positive or negative but still high suspicion
Negative
No further workup for obstruction
Retrograde urography and ureteral stent considered
Antegrade urography and percutaneous nephrostomy considered

FIGURE 331-1 Diagnostic approach for urinary tract obstruction in unexplained renal failure. CT, computed tomography.

Recent advances in technology have led to alternatives and have replaced the once standard intravenous urogram in the further evaluation of UTO. The high-resolution multidetector row CT scan, in particular, has the advantages of visualizing the retroperitoneum, as well as identifying both intrinsic and extrinsic sites of obstruction. Noncontrast CT scans improve visualization of the urinary tract in the patient with renal impairment and are safer for patients at risk for contrast nephropathy. Magnetic resonance urography is not at this time superior to the CT scan, and certain gadolinium agents carry a risk of systemic sclerosis in patients with renal insufficiency. Recently, promising alternatives to gadolinium have emerged, including iron-based contrast and inhalation of xenon gas, although thus far, gadolinium produces the clearest images and most-established safety. CT scanning may define the site of obstruction, identify and characterize kidney stones, and demonstrate dilatation of the calyces, renal pelvis, and ureter above the obstruction. The ureter may be tortuous in chronic obstruction. Though radionuclide scans give less anatomic detail than CT scans, they are able to give differential renal function. In the case of asymmetric renal function, the clinician may decide on a preferable kidney to decompress in the case of bilateral obstruction. Furosemide is sometimes given to increase detection with imaging and to distinguish functional from anatomic obstruction. The increase in urinary flow may bring out the pain of an acute obstructive process.

To facilitate visualization of a suspected lesion in a ureter or renal pelvis, *retrograde* or *antegrade urography* should be attempted. These procedures do not carry risk of contrast-induced acute kidney injury in patients with renal insufficiency. The retrograde approach involves catheterization of the involved ureter under cystoscopic control, whereas the antegrade technique necessitates percutaneous placement of a catheter into the renal pelvis. Although the antegrade approach may provide immediate decompression of a unilateral obstructing lesion, many urologists initially attempt the retrograde approach unless the catheterization is unsuccessful.

Voiding cystourethrography is of value in the diagnosis of vesicoureteral reflux and bladder neck and urethral obstructions. Postvoiding films reveal residual urine. Endoscopic visualization by the urologist often permits precise identification of lesions involving the urethra, prostate, bladder, and ureteral orifices.

TREATMENT

Urinary Tract Obstruction

UTO complicated by infection requires immediate relief of obstruction to prevent development of generalized sepsis and progressive renal damage. Sepsis necessitates prompt urologic intervention. Drainage may be achieved by nephrostomy, ureterostomy, or ureteral, urethral, or suprapubic catheterization. Prolonged antibiotic treatment may be necessary. Chronic or recurrent infections in a poorly functioning obstructed kidney may necessitate nephrectomy. When infection is not present, surgery is often delayed until acid-base, fluid, and electrolyte status is restored. Nevertheless, the site of obstruction should be ascertained as soon as feasible. Elective relief of obstruction is usually recommended in patients with urinary retention, recurrent urinary tract infections, persistent pain, or progressive loss of renal function. Benign prostatic hypertrophy may be treated medically with α-adrenergic blockers and 5α-reductase inhibitors. Renal colic may be treated with anti-inflammatory medication as edema often contributes to an obstructing ureteral stone, and α-adrenergic blockers may also be of benefit. The clinician should be aware of the risk of intraoperative floppy iris syndrome associated with cataract surgery in patients taking α-adrenergic blockers. Use of nonsteroidal anti-inflammatory medication must take into account the potential for renal harm, and opiates in patients with decreased renal function may be dangerous and should be used with caution. Functional obstruction secondary to neurogenic bladder may be decreased with the combination of frequent voiding and cholinergic drugs.

PROGNOSIS

With relief of obstruction, the prognosis regarding return of renal function depends largely on whether irreversible renal damage has occurred. When obstruction is not relieved, the course will depend mainly on whether the obstruction is complete or incomplete and bilateral or unilateral, as well as whether or not urinary tract infection is also present. Complete obstruction with infection can lead to total destruction of the kidney within days. Partial return of glomerular filtration rate may follow relief of complete obstruction of 1 and 2 weeks' duration, but after 8 weeks of obstruction, recovery is unlikely. In the absence of definitive evidence of irreversibility, every effort should be made to decompress the obstruction in the hope of restoring renal function at least partially. A renal radionuclide scan, performed after a prolonged period of decompression, may be used to predict the reversibility of renal dysfunction.

POSTOBSTRUCTIVE DIURESIS

Relief of bilateral, but not unilateral, complete obstruction commonly results in polyuria, which may be massive. The urine is usually hypotonic and may contain large amounts of sodium chloride, potassium, phosphate, and magnesium. The natriuresis is due in part to the correction of extracellular volume expansion, the increase in natriuretic factors accumulated during the period of renal failure, and depressed salt and water reabsorption when urine flow is reestablished. The retained urea is excreted with improved glomerular filtration rate, resulting in an osmotic diuresis that increases the urine volume of electrolyte-free water. Electrolyte-free water excretion (hypotonic urine) is recognized as being present when the sum of the urinary concentrations of sodium and potassium is lower than the serum sodium concentration. Precipitating factors include suppression of antidiuretic hormone at arterial baroreceptor sites, elevation of natriuretic peptides, or nephrogenic diabetes insipidus due to obstructive tubular injury. In the majority of patients, this diuresis results in the *appropriate* excretion of the excesses of retained salt and water. When extracellular volume and composition return to normal, the diuresis usually abates spontaneously. Occasionally, iatrogenic expansion of extracellular volume is responsible for, or sustains, the diuresis observed in the postobstructive period. Replacement with intravenous fluids in amounts less than urinary losses usually prevents this complication. More aggressive fluid management is required in the setting of hypovolemia, hypotension, or disturbances in serum electrolyte concentrations.

The loss of electrolyte-free water with urea may result in hypernatremia. Measured urinary output and serum and urine sodium, potassium, and osmolal concentrations should guide the use of appropriate intravenous replacement. Often replacement with 0.45% saline is required because of the likelihood of appropriately matching urine electrolyte concentrations. Relief of obstruction may be followed by urinary salt and water losses severe enough to provoke profound dehydration and vascular collapse. In these patients, decreased tubule reabsorptive capacity is probably responsible for the marked diuresis. Appropriate therapy in such patients includes intravenous administration of salt-containing solutions to replace sodium and volume deficits.

FURTHER READING

Capone V et al: Definition, diagnosis and management of fetal lower urinary tract obstruction: Consensus of the Erknet Cakut-Obstructive Uropathy Work Group. Nature Rev Urol 19:295, 2022.

Frokiaer J: Urinary tract obstruction, in *Brenner and Rector's The Kidney*, 10th ed, Skorecki K et al (eds). Philadelphia, W.B. Saunders & Company, 2016, pp 1257–1282.

McConnachie DJ et al: Ciliopathies and the kidney: A review. Am J Kidney Dis 77: 410, 2021.

Meldrum KK: Pathophysiology of urinary tract obstruction, in *Campbell Walsh Wein Urology*, Partin AW et al (eds). Philadelphia, Elsevier, 2020, Chapter 48.

Murugapoopathy V, Gupta IR: A primer on congenital anomalies of the kidneys and urinary tracts (CAKUT). Clin J Am Soc Nephrol 15:723, 2020.

Smith-Bindman R et al: Ultrasonography versus computed tomography for suspected nephrolithiasis. N Engl J Med 371:1100, 2014.

Stoller ML: Urinary obstruction and stasis, in *Smith and Tanagho's General Urology*, 18th ed. JW McAninch, TF Lue (eds). New York, McGraw-Hill, 2013, pp 170–182.

Tanagho EA, Nguyen HT: Vesicoureteral reflux, in *Smith and Tanagho's General Urology*, 18th ed. WJ McAninch, TF Lue (eds). New York, McGraw-Hill, 2013, pp 182–197.

Vollman DE et al: Intraoperative floppy iris and prevalence of intraoperative complications: Results from ophthalmic surgery outcomes database. Am J Ophthalmol 157:1130, 2014.

This page intentionally left blank

Disorders of the Gastrointestinal System PART 10

Section 1 Disorders of the Alimentary Tract

332

Approach to the Patient with Gastrointestinal Disease

William L. Hasler, Chung Owyang

ANATOMIC CONSIDERATIONS

The gastrointestinal (GI) tract extends from the mouth to the anus and is composed of organs with distinct functions. Sphincters assist in gut compartmentalization to separate the organs. Distinct gut wall layers contribute to regional activities. The mucosa regulates fluid and nutrient transfer but is also a barrier to prevent entry of some luminal contents. Smooth muscle and enteric nerves mediate propulsion. Many organs possess a serosal layer to provide a supportive foundation and permit external input. The gut interacts with many other systems. Pancreaticobiliary conduits deliver bile and enzymes into the duodenum. Lymphatic channels assist in gut immune activities. Intrinsic nerves control propulsion and fluid regulation. Extrinsic neural input provides volitional or involuntary control specific for each gut region.

FUNCTIONS OF THE GI TRACT

The GI tract serves two main functions—assimilating nutrients and eliminating waste. In the mouth, food is processed, mixed with salivary amylase, and delivered to the pharynx. Aborally propagating esophageal contractions propel the ingested bolus into the stomach in coordination with relaxation of the upper and lower esophageal sphincters. Basal tone in the lower esophageal sphincter prevents gastroesophageal reflux.

The stomach triturates and mixes food with pepsin and acid. The proximal stomach serves a storage function by relaxing to accommodate the meal. Phasic contractions in the distal stomach propel food residue against the pylorus, where it is ground and thrust proximally for further mixing prior to emptying into the duodenum.

Most nutrient absorption occurs in the small intestine. Mucosal villi provide maximal surface area and possess specialized enzymes and transporters. Triturated food from the stomach mixes with pancreatic secretions and bile in the duodenum. Pancreatic juice contains both enzymes for digestion and bicarbonate to optimize luminal pH for enzyme activation. Bile secreted by the liver and stored in the gallbladder is needed for lipid digestion. Most nutrients and minerals are rapidly absorbed in the proximal intestine, while bile acids and vitamin B_{12} are absorbed in the ileum. Bile contains byproducts of erythrocyte degradation, toxins, medications, and cholesterol that are emptied into the colon along with indigestible residue. The ileocecal junction is a sphincter that limits coloileal reflux of colonic microbes.

The colon prepares waste for evacuation by to-and-fro contractions that dehydrate the stool. As a consequence, daily luminal volumes decrease from 1000–1500 mL in the ileum to 100–200 mL that are expelled from the rectum. Esophageal transit takes seconds, stomach and small intestine emptying times range from minutes to a few hours, but colon propagation requires >1 day in most individuals. A dense colonic microbiome ferments undigested carbohydrates and short-chain fatty acids and modulates immune activity. The colon terminates in the anus, which possesses volitional and involuntary controls to retain stool in the rectum until it can be released in a convenient setting.

EXTRINSIC MODULATION OF GUT FUNCTION

GI function is modified by influences outside the gut. Unlike other systems, the gut is in continuity with the outside environment. Several mechanisms protect against injury from foods, medications, toxins, and microbes. Mucosal immune pathways include epithelial and lamina propria lymphocytes and plasma cells and lymph node chains to prevent noxious agents from entering the circulation. Antimicrobial peptides secreted by Paneth cells defend against pathogens. Drugs and toxins absorbed into the portal venous circulation are filtered and detoxified in the liver. Many GI reflexes involve extrinsic vagus or splanchnic nerve pathways. The brain-gut axis alters function in gut regions not under volitional regulation. Stress disrupts normal gut motor, secretory, and sensory activities.

OVERVIEW OF GI DISEASES

GI diseases range in severity from conditions that produce mild symptoms and no long-term morbidity to those with intractable symptoms or adverse outcomes. Diseases may be localized to one gut organ or exhibit diffuse involvement at many sites.

■ CLASSIFICATION OF GI DISEASES

GI diseases are manifestations of alterations in nutrient and fluid assimilation, transit and sensation, immune function, and vascular supply. Neoplastic degeneration may affect any GI site. Genetic factors underlie some disorders.

Impaired Digestion and Absorption Diseases of the stomach, small intestine, biliary tree, and pancreas can disrupt digestion and absorption. The most common maldigestion syndrome, lactase deficiency, produces gas and diarrhea after ingesting dairy products but has no adverse outcomes. Intestinal sucrase-isomaltase deficiency produces similar symptoms after consuming sucrose. Disorders that affect digestion and/or absorption more diffusely, including celiac disease, bacterial overgrowth, infectious enteritis, Crohn's ileitis, and radiation enteritis produce anemia, dehydration, electrolyte disorders, or malnutrition. Gastric acid hypersecretory conditions impair digestion by damaging the intestinal mucosa, impairing pancreatic enzyme activation, and accelerating transit. Benign or neoplastic biliary obstruction impairs fat digestion. Impaired pancreatic enzyme release in chronic pancreatitis or pancreatic cancer impairs digestion and leads to malnutrition.

Altered Secretion Some GI diseases result from alteration of gut secretion. Gastric acid hypersecretion occurs in gastrinoma, G-cell hyperplasia, retained antrum, and some patients with duodenal ulcers. Gastric acid is reduced in atrophic gastritis and pernicious anemia. Intestinal or colonic hypersecretion and/or impaired absorption causes diarrhea with acute bacterial or viral infection, chronic *Giardia* or cryptosporidia infections, small-intestinal bacterial overgrowth, bile salt diarrhea, microscopic colitis, diabetic diarrhea, and endocrine neoplasias with tumor overproduction of secretagogue transmitters such as vasoactive intestinal polypeptide.

Altered Gut Transit Mechanical obstruction can block transit of gut contents. Esophageal occlusion most often is due to stricture (from acid exposure or eosinophilic esophagitis) or neoplasm. Gastric obstruction develops from ulcer disease or cancer. Small-intestinal obstruction commonly results from adhesions but also occurs with Crohn's disease, radiation- or drug-induced strictures, and malignancy. The most common cause of colonic obstruction is colon cancer, although inflammatory strictures develop with inflammatory bowel disease (IBD), after diverticulitis, or with some medications.

Delayed propulsion can develop in any gut region. Achalasia is characterized by impaired esophageal body peristalsis and incomplete lower esophageal sphincter relaxation. Gastroparesis is the delay in gastric emptying resulting from impaired antral motility or spastic pyloric contraction. Intestinal pseudoobstruction is the disruption of small-bowel contractility due to enteric nerve or smooth-muscle injury. Slow-transit constipation results from diffusely impaired colon propulsion. Constipation also occurs with anorectal outlet abnormalities such as rectal prolapse, intussusception, or dyssynergia—a failure of anal or puborectalis relaxation with defecation.

Disorders of rapid propulsion are less common. Rapid gastric emptying occurs with postvagotomy dumping syndrome and some cases of functional dyspepsia and cyclic vomiting syndrome. Exaggerated intestinal or colonic motor patterns may contribute to diarrhea in irritable bowel syndrome (IBS). Accelerated transit with hyperdefecation is noted in hyperthyroidism.

Immune Dysregulation Many inflammatory GI conditions are consequences of altered immune function. Mucosal inflammation in celiac disease results from dietary ingestion of gluten-containing grains. Eosinophilic esophagitis and eosinophilic gastroenteritis exhibit prominent mucosal eosinophil infiltration. Ulcerative colitis and Crohn's disease are types of IBD that produce mucosal injury primarily in the lower gut. The microscopic colitides, lymphocytic and collagenous colitis, exhibit subepithelial infiltrates without visible mucosal damage. Abnormal mast cell function underlies mastocytosis and may play a role in some IBS cases. Bacterial, viral, and protozoal organisms produce ileitis or colitis in selected patients. GI manifestations of COVID-19 include nausea, vomiting, diarrhea, abdominal pain, and hepatic abnormalities. Gut microbiome alterations (termed dysbiosis) may trigger IBD, celiac disease, and IBS flares and may be factors in oncogenesis in some cases of pancreatic cancer.

Impaired Gut Blood Flow Ischemic damage occurs as a consequence of impaired blood flow. Rare cases of gastroparesis result from celiac artery blockage. Intestinal and colonic ischemia are more common and can be consequences of arterial embolus or thrombosis, venous thrombosis, or hypoperfusion from dehydration, sepsis, hemorrhage, or reduced cardiac output. These conditions produce mucosal injury, hemorrhage, perforation, or stricturing with chronic ischemia. Radiation enterocolitis may lead to reduced mucosal blood flow without large vessel occlusion.

Neoplastic Degeneration All GI regions are susceptible to malignant degeneration. In the United States, colorectal cancer is most common and usually presents after age 45. Gastric cancer is prevalent, especially in certain Asian populations, and may be a consequence of *Helicobacter pylori* infection. Esophageal cancer develops with chronic acid reflux or after excessive alcohol or tobacco use. Small-intestinal neoplasms are rare but occur with underlying inflammatory diseases. Most GI cancers exhibit carcinomatous histology, but lymphomas and other cell types can be seen. Anal cancers arise after prior anal infection or inflammation. Pancreatic and biliary cancers elicit severe pain, weight loss, and jaundice and have poor prognoses. Hepatocellular carcinoma usually arises in the setting of chronic viral hepatitis or cirrhosis secondary to other causes.

Disorders without Obvious Organic Abnormalities The most prevalent GI disorders show no abnormalities on biochemical or structural testing and include IBS, functional dyspepsia, and functional heartburn. These disorders have been given a new nomenclature and are termed disorders of gut-brain interaction (DGBI). DGBIs exhibit complex pathophysiology including altered gut motor function, exaggerated visceral sensory response to noxious stimulation, abnormal central nervous system processing of visceral pain sensation, and mucosal immune defects. Many patients with DGBIs have significant emotional disturbances on psychometric testing.

Genetic Influences Some GI diseases exhibit hereditary components. Family members of IBD patients show genetic predisposition to disease development. Colonic, esophageal, and pancreatic malignancies arise in certain inherited disorders. Lynch syndrome is an autosomal dominant condition with increased risk for developing colon and gynecologic cancers. Genetic dysmotility syndromes are described. Familial clustering is observed in DGBIs, although this may be secondary learned familial illness behavior rather than a true genetic factor.

SYMPTOMS OF GI DISEASE

Symptoms of GI disease include abdominal pain, heartburn, nausea and vomiting, altered bowel habits, GI bleeding, jaundice, and other manifestations (Table 332-1).

Abdominal Pain Abdominal pain results from GI disease and extraintestinal conditions of the genitourinary tract, abdominal wall, thorax, or spine. Visceral pain generally is midline in location and vague in character. Parietal pain is localized and precisely described. Painful inflammatory diseases include peptic ulcer, appendicitis, diverticulitis, IBD, pancreatitis, cholecystitis, and infectious enterocolitis. Noninflammatory sources include biliary colic, mesenteric ischemia, and neoplasia. The most common causes of abdominal pain are IBS and functional dyspepsia.

Heartburn Heartburn, a burning substernal sensation, is experienced by 40% of the population. Heartburn usually results from excess gastroesophageal acid reflux, but some cases exhibit normal acid exposure and are caused by nonacidic reflux or heightened esophageal nerve sensitivity.

Nausea and Vomiting Nausea and vomiting are caused by GI diseases, medications, toxins, infection, endocrine disorders, labyrinthine conditions, and central nervous system disease. Mechanical obstruction should be excluded as a cause of chronic nausea and vomiting. Disorders of propulsion including gastroparesis and intestinal pseudoobstruction elicit similar symptoms. Nausea and vomiting also are reported with IBS or gastroduodenal DGBIs (chronic nausea vomiting syndrome, cyclic vomiting syndrome, cannabinoid hyperemesis syndrome).

Altered Bowel Habits Altered bowel habits are common complaints. Constipation may be reported as infrequent defecation,

TABLE 332-1 Common Causes of Common Gastrointestinal (GI) Symptoms

ABDOMINAL PAIN	NAUSEA AND VOMITING	DIARRHEA	GI BLEEDING	OBSTRUCTIVE JAUNDICE
Appendicitis	Medications	Infection	Ulcer disease	Bile duct stones
Gallstone disease	GI obstruction	Poorly absorbed sugars	Esophagitis	Cholangiocarcinoma
Pancreatitis	Motor disorders	Inflammatory bowel disease	Varices	Cholangitis
Diverticulitis	Functional bowel disorder	Microscopic colitis	Vascular lesions	Sclerosing cholangitis
Ulcer disease	Cyclic vomiting syndrome	Functional bowel disorder	Neoplasm	Ampullary stenosis
Esophagitis	Cannabinoid hyperemesis syndrome	Celiac disease	Diverticula	Ampullary carcinoma
GI obstruction	Enteric infection	Pancreatic insufficiency	Hemorrhoids	Pancreatitis
Inflammatory bowel disease	Pregnancy	Hyperthyroidism	Fissures	Pancreatic tumor
Functional bowel disorder	Endocrine disease	Ischemia	Inflammatory bowel disease	
Vascular disease	Motion sickness	Endocrine tumor	Infectious colitis	
Gynecologic causes	Central nervous system disease			
Renal stone				

straining with defecation, passage of hard stools, or a sense of incomplete fecal evacuation and is caused by obstruction, motor disorders, medications, and endocrine diseases (hypothyroidism, hyperparathyroidism). Diarrhea may be reported as frequent defecation, passage of loose/watery stools, fecal urgency, or a similar sense of incomplete evacuation and can result from infections, inflammatory causes, malabsorption, and medications. IBS produces constipation, diarrhea, or an alternating bowel pattern. Fecal mucus is common in IBS, whereas pus and blood characterize IBD. Malabsorption can cause steatorrhea. A history of contaminated water exposure raises concern for protozoal causes (*Giardia, Cryptosporidium*).

GI Bleeding Upper GI bleeding presents with melena or hematemesis, whereas lower GI bleeding produces bright red or maroon stools. However, briskly bleeding upper sites can elicit red rectal bleeding, while slowly bleeding ascending colon sites may cause melena. Chronic occult GI bleeding may present with iron deficiency anemia. Causes of upper GI bleeding include ulcers, gastroduodenitis, esophagitis, portal hypertensive etiologies, malignancy, gastroesophageal junction tears, and vascular lesions. Lower GI sources include hemorrhoids, anal fissures, diverticula, ischemic colitis, neoplasm, IBD, infectious colitis, drug-induced colitis, arteriovenous malformations, and other vascular lesions.

Jaundice Jaundice results from prehepatic, intrahepatic, or posthepatic disease. Posthepatic causes of jaundice include biliary (choledocholithiasis, acute cholangitis, primary sclerosing cholangitis, strictures, neoplasm) and pancreatic (pancreatitis, stricture, malignancy) disorders.

Other Symptoms Dysphagia, odynophagia, and unexplained chest pain suggest esophageal disease. Globus sensation is reported with acid reflux but also occurs with upper esophageal sphincter dysfunction, hypersensitivity, or affective disorders. Weight loss, anorexia, and fatigue are reported with neoplastic, inflammatory, motility, pancreatic, and psychiatric conditions. IBD is associated with hepatobiliary dysfunction, skin and eye lesions, and arthritis. Celiac disease may present with dermatitis herpetiformis. Jaundice produces pruritus. Conversely, systemic diseases have GI consequences. Systemic lupus may cause gut ischemia, presenting with pain or bleeding.

EVALUATION OF THE PATIENT WITH GI DISEASE

Evaluation of suspected GI disease begins with a careful history and examination and may be followed by investigation with tools to test gut structure or function and luminal contents in selected cases. In patients with normal diagnostic assessment, validated symptom profiles can be characterized to confidently define the presence of a DGBI.

HISTORY

The history in suspected GI disease has several components. Symptom timing, profiles, and duration suggest specific etiologies. Short-duration symptoms commonly result from acute infection or inflammation, toxin exposure, or ischemia. Long-standing symptoms point to chronic inflammation, neoplasia, or DGBIs. Luminal obstruction presents with dysphagia, nausea and vomiting, bloating and distention, or constipation depending on the site of blockage. Symptoms from obstruction, ischemia, IBD, and DGBIs are worsened by meals, while ulcer symptoms may be relieved by eating. Ulcer pain occurs intermittently over weeks to months, whereas biliary colic has sudden onset and lasts up to several hours. Acute pancreatitis pain is severe and persists for days to weeks. Meals elicit diarrhea while defecation relieves discomfort in some IBD and IBS cases. DGBIs are exacerbated by stress. Sudden awakening by pain suggests organic rather than functional causes. Diarrhea from malabsorption usually improves with fasting, whereas secretory diarrhea persists without oral intake.

Symptom relation to other factors offers insight into underlying diagnostic possibilities. Obstructive symptoms with a history of abdominal surgery raise concern for adhesions. Loose stools after gastrectomy or cholecystectomy suggest dumping syndrome or postcholecystectomy diarrhea. Symptom onset after travel prompts consideration of infection. Medications produce pain, altered bowel habits, or GI bleeding. Celiac disease is prevalent in people of northern European descent, whereas IBD is more common in Jewish populations. A sexual history may raise concern for infection or immunodeficiency.

Symptom criteria devised by working groups improve diagnosis of DGIBs and reduce unnecessary diagnostic testing. The best accepted symptom-based criteria are the Rome criteria, which exhibit sensitivities and specificities of 55–75% when tested against structural findings in IBS and functional dyspepsia, indicating the need for careful test selection in patients with high risk of organic disease.

PHYSICAL EXAMINATION

The physical examination complements information from the history. Fever suggests inflammation or neoplasm. Orthostasis is produced by significant blood loss, dehydration, sepsis, or autonomic neuropathy. Skin, eye, or joint findings may point to specific diagnoses such as IBD. Neck examination with swallowing assessment evaluates dysphagia. Lung and cardiac examinations can suggest cardiopulmonary disease as causes of abdominal pain or nausea, while pelvic examination tests for gynecologic sources of pain. Rectal examination may detect blood from mucosal injury or neoplasm or a palpable inflammatory mass in appendicitis. Metabolic conditions and gut motor disorders can have associated peripheral neuropathy.

Abdominal inspection may reveal distention from obstruction, neoplasm, or ascites or vascular abnormalities with liver disease. Auscultation detects bruits or friction rubs from vascular disease or hepatic tumors. Loss of bowel sounds signifies ileus, whereas high-pitched, hyperactive sounds characterize intestinal obstruction. Percussion assesses liver size and detects shifting dullness from ascites. Palpation assesses for hepatosplenomegaly and neoplastic or inflammatory masses. Intestinal ischemia elicits severe pain but little tenderness. Patients with visceral pain may exhibit generalized discomfort, whereas those with parietal pain or peritonitis have localized pain with involuntary guarding, rigidity, or rebound. Patients with musculoskeletal abdominal wall pain may note tenderness exacerbated by Valsalva or leg lift maneuvers.

TOOLS FOR PATIENT EVALUATION

Laboratory, radiographic, and functional tests help to diagnose suspected GI disease. The GI tract is amenable to internal evaluation using endoscopy and to examination of luminal contents. Histopathologic examination of GI tissues complements this testing.

Laboratory Laboratory tests can provide useful information. Iron-deficiency anemia suggests mucosal blood loss, whereas vitamin B_{12} deficiency results from intestinal, gastric, or pancreatic disease. Both conditions also result from inadequate oral intake. Leukocytosis and increased sedimentation rates and C-reactive proteins are found in inflammation, whereas leukopenia is seen in viral illness. Severe vomiting or diarrhea elicits electrolyte disturbances, acid-base abnormalities, and elevated blood urea nitrogen. Pancreaticobiliary or liver disease produces elevated pancreatic or liver chemistries. Thyroid chemistries, cortisol, and calcium levels evaluate for endocrinologic causes of symptoms. Pregnancy testing is considered for women with unexplained nausea. Serologic tests screen for atrophic gastritis, celiac disease, connective tissue diseases, and paraneoplastic dysmotility syndromes. Hormone and neurotransmitter levels are obtained for suspected endocrine neoplasia. Measurement of serum gastrin after secretin administration can facilitate gastrinoma diagnosis. Intraabdominal malignancies produce tumor markers including carcinoembryonic antigen, CA 19-9, CA 125, and α-fetoprotein. Blood testing monitors thiopurine metabolites in IBD patients on immunosuppressant agents. Research into novel biomarkers is being conducted to predict longitudinal course and treatment response in IBD and other conditions. Ascitic fluid is analyzed for infection, malignancy, or findings of portal hypertension. Urine samples screen for carcinoid, porphyria, and heavy metal intoxication. Genetic testing is indicated in some patients with hereditary predisposition to selected GI diseases. Pharmacogenetic methods can determine which therapies should not be given to specific patients.

Luminal Contents Luminal contents are analyzed in selected settings. Fecal *Helicobacter pylori* antigen testing can detect active infection with this agent in ulcer disease or dyspepsia. Stool samples are cultured for bacterial pathogens, examined for leukocytes and parasites, or tested for *Giardia* antigen. Fecal fat is increased with malabsorption. Fecal elastase is decreased with exocrine pancreatic insufficiency. Inflammatory conditions such as IBD produce fecal calprotectin or lactoferrin elevations. Stool electrolytes can be measured in diarrheal conditions. Laxative screens are performed for suspected laxative abuse. Fecal immunochemical and multitarget fecal DNA tests can be used for colon cancer screening in low-risk populations. Gastric acid is quantified to exclude gastrinoma. Esophageal pH/impedance testing is done for refractory gastroesophageal reflux and can characterize mechanisms underlying unexplained belching.

Endoscopy Indications for endoscopic procedures are listed in Table 332-2, which can diagnose causes of bleeding, pain, nausea and vomiting, weight loss, altered bowel function, and fever. Upper endoscopy evaluates the esophagus, stomach, and duodenum and is advocated as the initial test for suspected ulcer disease, esophagitis, neoplasm, malabsorption, and Barrett's metaplasia and dysplasia because of its abilities to visualize and biopsy abnormalities. Endoscopically obtained duodenal aspirates can be examined for parasites or cultured for bacterial overgrowth. It should be recognized that upper endoscopy is normal in 85% of patients with unexplained dyspepsia. Colonoscopy is the preferred procedure for colon cancer screening and surveillance with possible excision of noncancerous polyps. Colonoscopy also can obtain biopsies for colitis or ileitis secondary to IBD, microscopic colitis, immune checkpoint inhibitor use, infection, ischemia, and radiation. Sigmoidoscopy examines the colon to the splenic flexure and excludes distal causes of bleeding, inflammation, or obstruction in young patients not at significant risk for colon cancer. For elusive GI bleeding from arteriovenous malformations or superficial ulcers, small-intestinal examination is performed with push enteroscopy, capsule endoscopy, or double-balloon enteroscopy. Capsule endoscopy also visualizes small-intestinal Crohn's disease in individuals with negative radiography. Endoscopic ultrasound (EUS) diagnoses and stages GI malignancy, excludes choledocholithiasis, evaluates pancreatitis, and assesses anal sphincter integrity. Endoscopic retrograde cholangiopancreatography (ERCP) provides diagnoses and affords therapy of pancreatic and biliary disease.

Radiography/Nuclear Medicine Radiographic tests evaluate visceral and extraluminal structures. Esophagography is the initial procedure to exclude subtle rings, strictures, or achalasia as causes of dysphagia, whereas fluoroscopic triple-phase swallow testing evaluates oropharyngeal dysphagia. Small-bowel contrast radiology detects intestinal tumors, strictures, and fistulae and can estimate intestinal transit. Contrast enemas are performed when colonoscopy is unsuccessful or contraindicated. Defecography characterizes anorectal dysfunction as a cause of constipation. Ultrasound and computed tomography (CT) evaluate regions not accessible by endoscopy or contrast studies, including the liver, pancreas, gallbladder, kidneys, and retroperitoneum, and can diagnose appendicitis, mass lesions, fluid collections, organ enlargement, and (in the case of ultrasound) gallstones. CT and magnetic resonance (MR) colonography are alternatives to colonoscopy for colon cancer screening. MR methods image pancreaticobiliary ducts to exclude neoplasm, stones, and sclerosing cholangitis and the liver to characterize benign and malignant tumors. CT or MR enterography assesses small-bowel thickening, stricturing, masses, and fistulae in IBD and other inflammatory or neoplastic conditions. Percutaneous transhepatic cholangiography accesses the biliary tree in obstructive jaundice. CT and MR angiographic techniques can define mesenteric occlusion and detect acute intraluminal hemorrhage. Ultrasound or MR elastography quantifies fibrosis in chronic liver disease. Positron emission tomography (PET) distinguishes malignant from benign disease in several organ systems. Imaging with combined PET-CT techniques using DOTA-octreotate and related agents has improved detection of neuroendocrine tumors.

Scintigraphy evaluates structural abnormalities and quantifies luminal transit. Radionuclide scans localize bleeding sites in patients with brisk hemorrhage to direct therapy with endoscopy, angiography, or surgery. Radiolabeled leukocyte scans search for intraabdominal abscesses not visualized on CT. Biliary scintigraphy complements ultrasound in assessing cholecystitis. Scintigraphy to quantify esophageal and gastric emptying is well established. Scintigraphic measures of small-intestinal or colonic transit are less widely used.

TABLE 332-2 Common Indications for Endoscopy

UPPER ENDOSCOPY	COLONOSCOPY	ENDOSCOPIC RETROGRADE CHOLANGIOPANCREATOGRAPHY	ENDOSCOPIC ULTRASOUND	CAPSULE ENDOSCOPY	DOUBLE-BALLOON ENDOSCOPY
Dyspepsia despite treatment	Cancer screening	Jaundice	Staging of malignancy	Obscure bleeding	Ablation of small-intestinal bleeding sources
Dyspepsia with signs of organic disease	Lower gastrointestinal (GI) bleeding	Postbiliary surgery complaints	Characterize and biopsy submucosal mass	Suspected Crohn's disease of the small intestine	Biopsy of suspicious small-intestinal masses/ulcers
Refractory vomiting	Anemia	Cholangitis	Bile duct stones		
Dysphagia	Diarrhea	Gallstone pancreatitis	Chronic pancreatitis		
Upper GI bleeding	Polypectomy	Pancreatic/biliary/ampullary tumor	Drain pseudocyst		
Anemia	Obstruction	Unexplained pancreatitis	Anal continuity		
Weight loss	Biopsy radiologic abnormality	Pancreatitis with unrelenting pain	Direct stent placement		
Malabsorption	Cancer surveillance: family history prior polyp/cancer, colitis	Fistulas			
Biopsy radiologic abnormality	Palliate neoplasm	Biopsy radiologic abnormality			
Polypectomy	Remove foreign body	Pancreaticobiliary drainage			
Place gastrostomy	Place stent across stenosis	Sample bile			
Barrett's surveillance		Sphincter of Oddi manometry			
Palliate neoplasm					
Sample duodenal tissue/fluid					
Remove foreign body					
Endoscopic mucosal resection or endoscopic submucosal dissection for dysplasia or early cancer					
Place stent across stenosis					
Endoscopic myotomy for achalasia or gastroparesis					
Endoscopic bariatric procedures					

Histopathologic Assessment Endoscopic mucosal biopsies evaluate for inflammatory (including celiac disease, IBD, microscopic colitis), infectious, dysplastic, and neoplastic disease. Gastric biopsies can detect *H. pylori* on immunohistochemical staining and determine antibiotic sensitivity from biopsy cultures. Deep rectal biopsies facilitate diagnosis of Hirschsprung's disease or amyloid. Liver biopsy is performed for abnormal liver chemistries, unexplained jaundice, some cases of viral hepatitis, and following liver transplants to exclude rejection. Biopsies obtained during CT or ultrasound evaluate for intraabdominal conditions not accessible by endoscopy.

Methods to visually scrutinize histologic features and detect dysplasia during endoscopy prior to targeted biopsy in colitis, Barrett's esophagus, and gastric malignancy include narrow-band imaging, chromoendoscopy, confocal laser endomicroscopy, and optical coherence tomography. Artificial intelligence techniques using machine learning are in development for detecting dysplasia and early cancer in images from biopsies.

Functional Testing Tests of gut function provide important data when structural testing is nondiagnostic. High-resolution esophageal manometry defines esophageal motility disorders including achalasia. Small-intestinal manometry tests for pseudoobstruction and colon manometry evaluate for colonic inertia. Anorectal manometry with balloon expulsion testing characterizes unexplained incontinence or constipation from outlet dysfunction. The functional lumen imaging probe measures reduced distensibility in the lower esophageal sphincter in achalasia, the upper esophageal sphincter with Zenker's diverticula, and the pylorus in gastroparesis. Measurement of breath hydrogen and methane while fasting and after oral mono- or oligosaccharide challenge screen for carbohydrate intolerance and small-intestinal bacterial overgrowth. Urea breath testing assesses for persistent *H. pylori* infection. Gastric emptying breath testing is an alternative to scintigraphy for gastroparesis diagnosis.

TREATMENT

Gastrointestinal Disease

Management of GI diseases depends on the cause of symptoms. Treatment options include dietary modifications, medications, treatment of gut dysbiosis, therapeutic endoscopy, interventional radiology, surgery, psychological approaches, and physical therapy. Smartphone applications have been developed for diverse purposes ranging from providing instructions for endoscopy preparation to educating and promoting adherence to diet restrictions.

NUTRITIONAL MODIFICATION

Nutritional modifications for GI disease can reduce symptoms, correct pathologic defects, or replace normal food intake with enteral or parenteral formulations. Changes that improve symptoms but do not reverse organic abnormalities include lactose restriction for lactase deficiency, liquid meals in gastroparesis, carbohydrate restrictions with dumping syndrome, avoidance of selected meats with alpha gal syndrome, and low-FODMAP (fermentable oligo-, di-, monosaccharides and polyols) diets in IBS. The gluten-free diet for celiac disease exemplifies a primary therapy to reduce mucosal inflammation. Likewise, elimination diets may improve histology and symptoms in eosinophilic esophagitis. Medium-chain triglycerides replace normal fats in short-gut syndrome or severe ileal disease. Perfusing liquid meals through a gastrostomy is performed in patients who cannot swallow safely. Enteral jejunostomy feedings are considered for gastric dysmotility syndromes that preclude feeding into the stomach. Total parenteral hyperalimentation is used for generalized gut failure in patients for whom enteral nutrition cannot be sustained.

PHARMACOTHERAPY

Medications in several categories treat GI diseases. Considerable resources are expended on over-the-counter remedies. Prescription drugs are offered as short-term or continuous therapy to reduce GI symptoms. Other prescription agents modify disease activity with the intent to alter longitudinal disease course. Alternative treatments are popular in conditions for which traditional therapies provide incomplete relief.

Over-the-Counter Agents Over-the-counter agents are reserved for mild GI symptoms. Antacids, histamine H_2 antagonists, alginates, and proton pump inhibitors (PPIs) decrease symptoms in gastroesophageal reflux disease (GERD) and dyspepsia. Fiber supplements, stool softeners, enemas, and laxatives are used for constipation. Laxatives are categorized as stimulants, osmotic agents (including isotonic preparations containing polyethylene glycol), and poorly absorbed sugars. Nonprescription antidiarrheal agents include bismuth subsalicylate, kaolin-pectin combinations, and loperamide. Lactase pills are used for lactose intolerance. Gaseous symptoms may be reduced by bacterial α-galactosidase, antiflatulents, and adsorbents. Using nonprescription preparations for chronic persistent symptoms should be supervised by a health care provider to ensure that an underlying severe medical illness is not missed.

Prescription Drugs for Symptom Control Prescription drugs are approved to treat symptoms in many GI diseases. Higher-dose PPIs are advocated for GERD when over-the-counter preparations are inadequate. Given acutely, PPIs slow acute upper GI bleeding, reducing requirements for endoscopic therapy. PPIs slow progression of Barrett's metaplasia and can manage eosinophilic esophagitis. Vonoprazan is the first drug in a newer acid suppressant class of potassium-competitive acid blockers for severe acid-related disorders and as an adjunct to *H. pylori* eradication. Baclofen reduces gastroesophageal reflux by increasing lower esophageal sphincter barrier function. Other therapies of eosinophilic esophagitis include swallowed fluticasone and budesonide. Cytoprotective agents like sucralfate are available for upper gut ulcers. Prokinetic drugs stimulate GI propulsion in gastroparesis, chronic intestinal pseudoobstruction, acute colonic pseudoobstruction, and slow-transit constipation. Secretagogue drugs are prescribed for refractory constipation, whereas peripheral opioid antagonists are offered for opioid-induced constipation. Prescription antidiarrheals include opioid drugs, anticholinergic antispasmodics, tricyclics, bile acid binders, and serotonin antagonists. Opioids may be needed for pain control in disseminated malignancy and chronic pancreatitis. Antiemetics reduce nausea and vomiting. Pancreatic enzymes decrease malabsorption and pain from pancreatic disease. The antisecretory somatostatin analogue octreotide treats hypersecretory states. Some DGBIs require neuromodulators, including tricyclic agents, for control of pain, diarrhea, or nausea. Nonselective beta blockers can reduce the risk of variceal hemorrhage with portal hypertension.

Prescription Drugs for Disease Modification Other prescription medication classes modify disease activity by actions to reduce inflammation, modulate immune function, and target neoplasm. Immunomodulatory drugs are used in IBD, microscopic colitis, immune checkpoint inhibitor colitis, refractory celiac disease, autoimmune pancreatitis, and gut vasculitis. Biologic agents that inhibit tumor necrosis activity and antibodies targeting integrins, proinflammatory cytokines, Janus kinase inhibitors, and cell adhesion molecules have benefits in IBD. Agents effective for acute severe ulcerative colitis requiring hospitalization include corticosteroids, infliximab, and cyclosporine. Biologics that deplete eosinophils or inhibit mast cells show promise in eosinophilic disorders. *H. pylori* eradication treats some gastric malignancies. Most GI carcinomas respond poorly to chemotherapy with or without radiotherapy, whereas lymphomas may be cured with these interventions. Cancer treatment programs including immunotherapies, immune checkpoint inhibitors, and tumor vaccines are in development.

Complementary and Alternative Medicines Several alternative treatments are available to manage GI symptoms. Ginger, acupressure, and acustimulation are used for nausea. Pyridoxine is given for nausea in first-trimester pregnancy. Peppermint oil and caraway seed oil products and herbal preparations such as STW 5 (a nine-herb mixture) are useful for functional dyspepsia and IBS. Low-potency pancreatic enzyme preparations are sold as digestive aids but have little evidence to support their efficacy.

THERAPIES TARGETING GUT DYSBIOSIS

Antibiotics are prescribed to treat *H. pylori*–induced ulcers, infectious diarrhea, complicated diverticulitis, and small intestinal bacterial overgrowth. Some cases of diarrhea-predominant IBS respond to the nonabsorbable antibiotic rifaximin. Probiotics containing bacterial cultures and prebiotics that selectively nourish nonnoxious commensal bacteria have been given as adjunctive therapy for infectious diarrhea and IBS, with limited evidence of efficacy. Postbiotics are metabolites made by probiotic organisms that inhibit pathogenic luminal bacteria. Transplanting donor feces into the colon by colonoscopy or enema is effective treatment for recurrent *Clostridioides difficile* colitis. Commercial oral and rectal microbiota-based products have been approved or are in development for this indication.

THERAPEUTIC ENDOSCOPY

In addition to its diagnostic role, endoscopy has numerous therapeutic capabilities. Cautery techniques and injection of vasoconstrictor substances can stop hemorrhage from ulcers and vascular malformations. Endoscopically placed clips can occlude arterial bleeding sites, while hemostatic powder sprays can stop brisk persistent GI bleeding. Endoscopic encirclement of esophageal varices and hemorrhoids with constricting bands stops hemorrhage from these sites. Bleeding gastric varices can be injected with thrombin or cyanoacrylate. Endoscopy can remove polyps in the stomach, small bowel, or colon. Decompressive colonoscopy withdraws excess gas in some cases of acute colonic pseudoobstruction. Endoscopic mucosal resection, submucosal dissection, and radiofrequency techniques ablate some cases of Barrett's esophagus with dysplasia and resect superficial cancers or subepithelial tumors elsewhere in the gut. Obstructions of the GI lumen and pancreaticobiliary tree are relieved by endoscopic dilation or placing plastic or expandable metal stents. Endoscopic sphincterotomy of the ampulla of Vater treats choledocholithiasis. Cholangioscopy can facilitate stone lithotripsy in the common bile duct, ablation of small ductal tumors, and placement of gallbladder stents to facilitate drainage in nonoperative candidates. Interventional EUS is used for pancreatic cyst gastrostomy using lumen-apposing metal stents, pancreatic necrosectomy, and placement of fiducial markers to direct pancreatic and rectal radiotherapy. EUS also can facilitate endoscopic access to the excluded distal stomach in patients who have undergone bariatric gastric bypass surgery using similar stents so that ERCP can be done for pancreaticobiliary conditions. EUS-directed stent placement can manage postsurgical stenoses after pancreatic resection. Endoscopy is used to insert gastric feeding tubes. Peroral endoscopic myotomy is performed on the lower esophageal sphincter in achalasia, the pylorus in gastroparesis, and the upper esophageal sphincter for Zenker's diverticulum. Endoscopic treatments for acid reflux including transoral incisionless fundoplication have been developed as potential alternatives to surgery. Endoscopic bariatric methods, including intragastric balloons, sleeve gastroplasty, and duodenal resurfacing and diversion, have been devised.

INTERVENTIONAL RADIOLOGY

Interventional radiology techniques offer benefits in selected settings. Angiographic embolization or vasoconstriction decreases bleeding from gut sites not amenable to endoscopic intervention. Angiographic embolectomy, stent placement, and thrombolysis also manage mesenteric ischemia. Dilatation or stenting under fluoroscopic guidance relieves luminal strictures. Contrast enemas can reduce colon volvulus. CT- and ultrasound-directed drainage of abdominal fluid collections can obviate the need for surgery. Percutaneous transhepatic cholangiography relieves biliary obstruction when ERCP is contraindicated. Percutaneous cholecystostomy treats acute cholecystitis in patients unable to undergo cholecystectomy. Transjugular intrahepatic portosystemic shunts are performed for variceal hemorrhage not amenable to endoscopic therapy. Lithotripsy is rarely performed to fragment gallstones in patients who are not surgical candidates. Radiologic approaches are often chosen over endoscopy for gastroenterostomy placement. Radiographic assistance is sometimes needed for placement of central venous catheters for parenteral nutrition.

SURGERY

Roles of surgery in GI conditions include disease cure, symptom control, maintenance of nutrition, and palliation of unresectable neoplasm. Surgery cures medication-unresponsive ulcerative colitis, diverticulitis, cholecystitis, appendicitis, and intraabdominal abscess, but only reduces symptoms and treats complications in Crohn's disease. Surgery is performed for ulcer complications like bleeding, obstruction, or perforation and intestinal obstructions that persist after conservative care. Gastroesophageal fundoplication is performed for refractory acid reflux. Acid exposure time on pH testing helps select candidates for fundoplication. Achalasia responds to operations to reduce lower esophageal sphincter tone. Operations for motor disorders include implanted electrical stimulators for gastroparesis and electrical devices and artificial sphincters for fecal incontinence. Surgery can place a jejunostomy for long-term enteral feedings. Other common indications for surgery include hernias, hemorrhoids, and nonhealing anal fissures.

PSYCHOLOGICAL APPROACHES AND PHYSICAL THERAPY

Psychological therapies, including psychotherapy, cognitive behavioral therapy, and hypnosis, show efficacy in DGBIs and are most beneficial for patients with significant psychological dysfunction. Behavioral therapists provide instruction in diaphragmatic breathing for belching or rumination. Biofeedback methods administered by physical therapists can treat refractory fecal incontinence or constipation secondary to dyssynergia.

■ FURTHER READING

Benech N et al: Update on microbiota-derived therapies for recurrent *Clostridioides difficile* infections. Clin Microbiol Infect 30:462, 2024.

Gergely M et al: Management of refractory inflammatory bowel disease. Curr Opin Gastroenterol 38:347, 2022.

Hossain B et al: Prevalence and impact of gastrointestinal manifestations in COVID-19 patients: A systematic review. J Community Hosp Intern Med Perspect 13:39, 2023.

Jain S et al: Optimal strategies for colorectal cancer screening. Curr Treat Options Oncol 23:474, 2022.

Orpen-Palmer J et al: Update on the management of upper gastrointestinal bleeding. BMJ Med 1:e000202, 2022.

Shakir SM et al: Updates to the diagnosis and clinical management of *Helicobacter pylori* infections. Clin Chem 69:869, 2023.

333

Gastrointestinal Endoscopy

Louis Michel Wong Kee Song, Vinay Chandrasekhara, Mark Topazian

Gastrointestinal endoscopy has been attempted for over 200 years, but the introduction of semirigid and flexible gastroscopes in the mid-twentieth century marked the dawn of the modern endoscopic era. Since then, rapid advances in endoscopic technology have led to

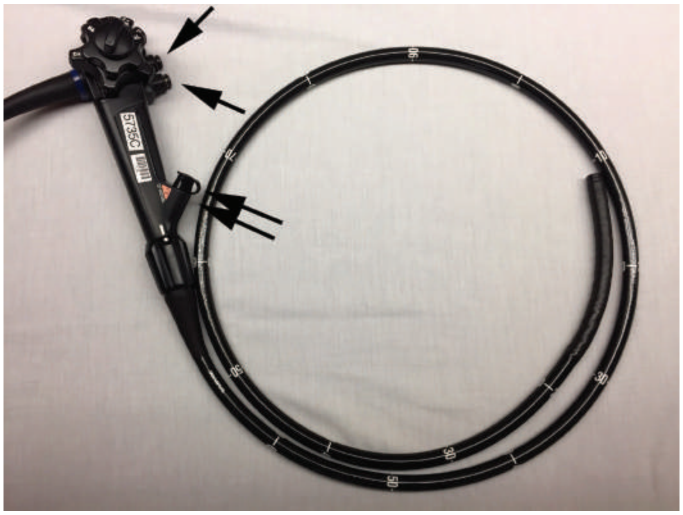

FIGURE 333-1 Gastrointestinal endoscope. Shown here is a conventional colonoscope with control knobs for tip deflection, push buttons for suction and air insufflation (*single arrows*), and a working channel for passage of accessories (*double arrows*).

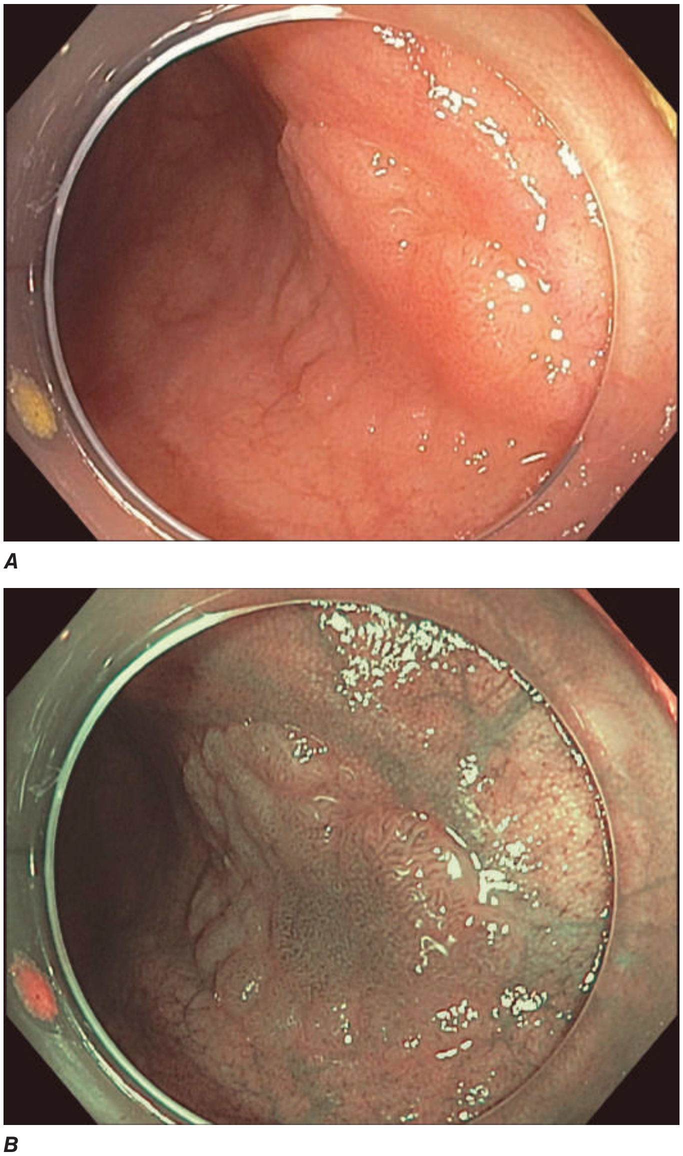

FIGURE 333-2 Flat colon polyp. ***A.*** White-light imaging. ***B.*** Corresponding narrow-band imaging enhances mucosal features and lesion delineation.

dramatic changes in the diagnosis and treatment of many digestive diseases. Innovative endoscopic devices and new endoscopic treatment modalities continue to expand the use of endoscopy in patient care.

Flexible endoscopes provide an electronic video image generated by a charge-coupled device (CCD) or a complementary metal oxide semiconductor (CMOS) chip in the tip of the endoscope. Operator controls permit deflection of the endoscope tip; fiberoptic bundles or light-emitting diodes provide light at the tip of the endoscope; and working channels allow washing, suctioning, and the passage of instruments **(Fig. 333-1)**. Progressive changes in the diameter and stiffness of endoscopes have improved the ease and patient tolerance of endoscopy. High-resolution and high-definition endoscopes equipped with electronic and optical magnification capabilities enable acquisition of images with a high level of detail. Advanced imaging techniques, including narrow-band imaging **(Fig. 333-2)** and real-time image-processing enhancement algorithms, aid in tissue characterization or differentiation.

ENDOSCOPIC PROCEDURES

■ UPPER ENDOSCOPY

Upper gastrointestinal endoscopy, also referred to as esophagogastroduodenoscopy (EGD), is performed by passing a flexible endoscope through the mouth into the esophagus, stomach, and duodenum. The procedure is the best method for examining the upper gastrointestinal mucosa **(Fig. 333-3)**. While the upper gastrointestinal radiographic series has similar accuracy for diagnosis of duodenal ulcer **(Fig. 333-4)**, EGD is superior for detection of gastric ulcers **(Fig. 333-5)** and flat mucosal lesions, such as Barrett's esophagus **(Fig. 333-6)**, and it permits directed biopsy and endoscopic therapy. Intravenous sedation is given to most patients in the United States to ease the anxiety and discomfort of the procedure, although in many countries, EGD is routinely performed with topical pharyngeal anesthesia only. Patient tolerance of unsedated EGD is improved by the use of an ultrathin, 5-mm diameter endoscope that can be passed transorally or transnasally.

■ COLONOSCOPY

Colonoscopy is performed by passing a flexible colonoscope through the anal canal into the rectum and colon. The cecum is reached in >95% of cases, and the terminal ileum **(Fig. 333-7)** can usually be examined. Colonoscopy is the gold standard for imaging the colonic mucosa **(Fig. 333-8)**. Colonoscopy has greater sensitivity than barium enema for the detection of colitis **(Fig. 333-9)**, colon polyps **(Fig. 333-10)**, and colorectal cancer **(Fig. 333-11)**. Computed tomography (CT) colonography rivals the accuracy of colonoscopy for the detection of some polyps and cancer, although it is not as sensitive for the detection of flat lesions, such as serrated polyps **(Fig. 333-12)**. Intravenous sedation is usually given before colonoscopy in the United States, although a willing patient and a skilled examiner can complete the procedure without sedation in many cases.

■ FLEXIBLE SIGMOIDOSCOPY

Flexible sigmoidoscopy is akin to colonoscopy, but it visualizes only the rectum and a variable portion of the left colon, typically to 60 cm from the anal verge. This procedure may result in abdominal cramping and discomfort, but it is brief and thus can be performed without sedation. Flexible sigmoidoscopy is primarily used for evaluation of diarrhea and rectal outlet bleeding.

■ SMALL-BOWEL ENDOSCOPY

Three endoscopic techniques are currently used to evaluate the small intestine, most often in patients presenting with presumed small-bowel bleeding. For *capsule endoscopy*, the patient swallows a disposable capsule that contains a CMOS chip camera. Color still images **(Fig. 333-13)** are transmitted wirelessly to an external receiver at several frames per second until the capsule's battery is exhausted or it is passed into the toilet. Capsule endoscopy enables visualization of the small-bowel mucosa beyond the reach of a conventional endoscope,

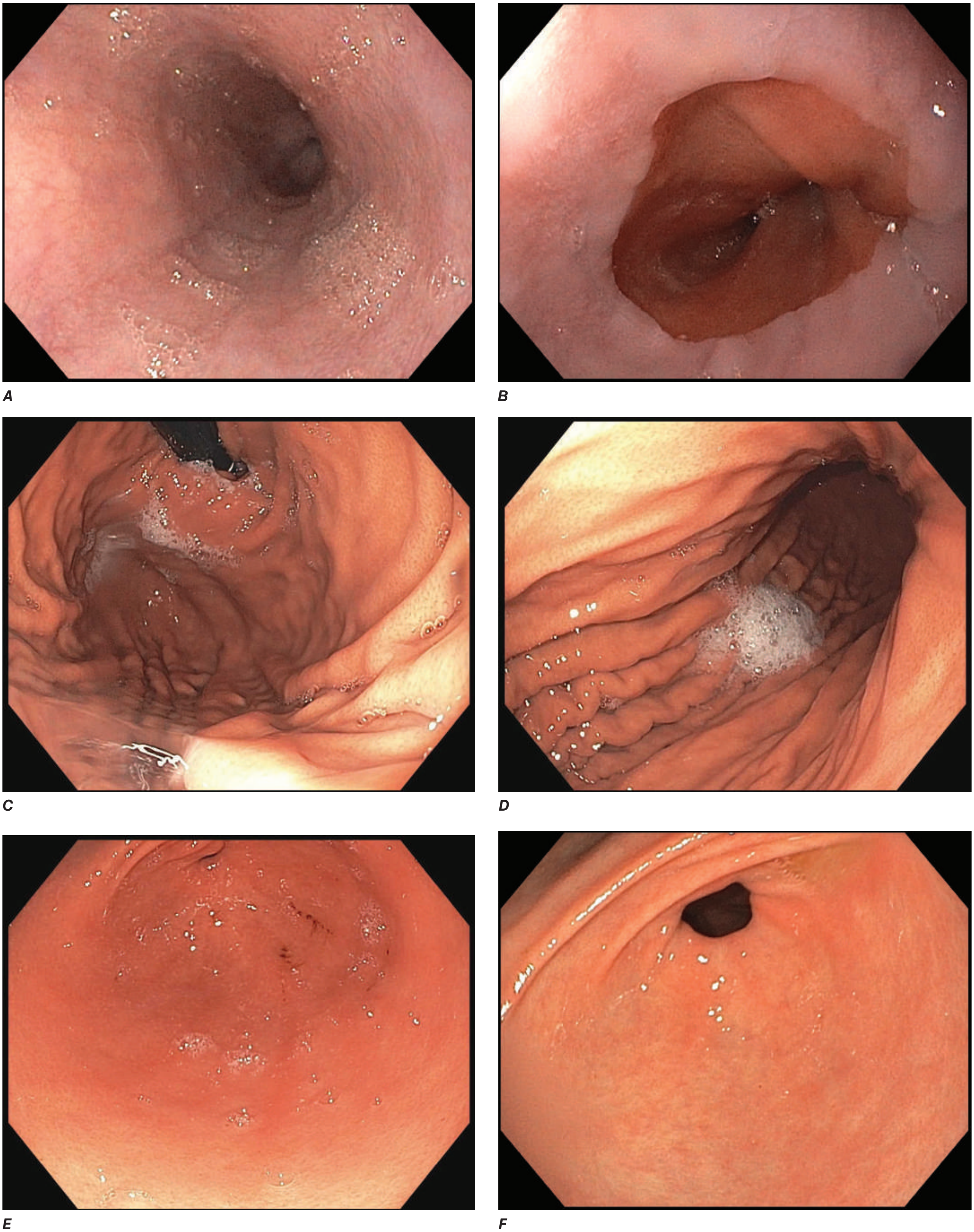

FIGURE 333-3 Normal upper endoscopic examination. ***A.*** Esophagus. ***B.*** Gastroesophageal junction. ***C.*** Gastric fundus. ***D.*** Gastric body. ***E.*** Gastric antrum. ***F.*** Pylorus. ***G.*** Duodenal bulb. ***H.*** Second portion of the duodenum.

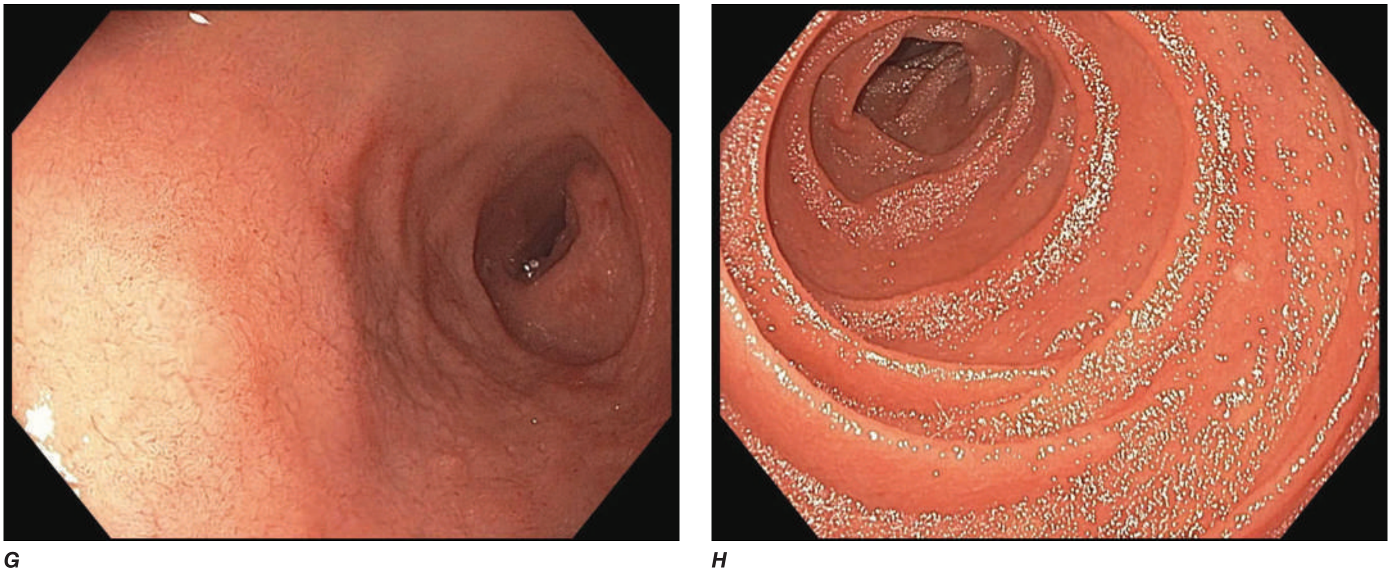

FIGURE 333-3 (*Continued*)

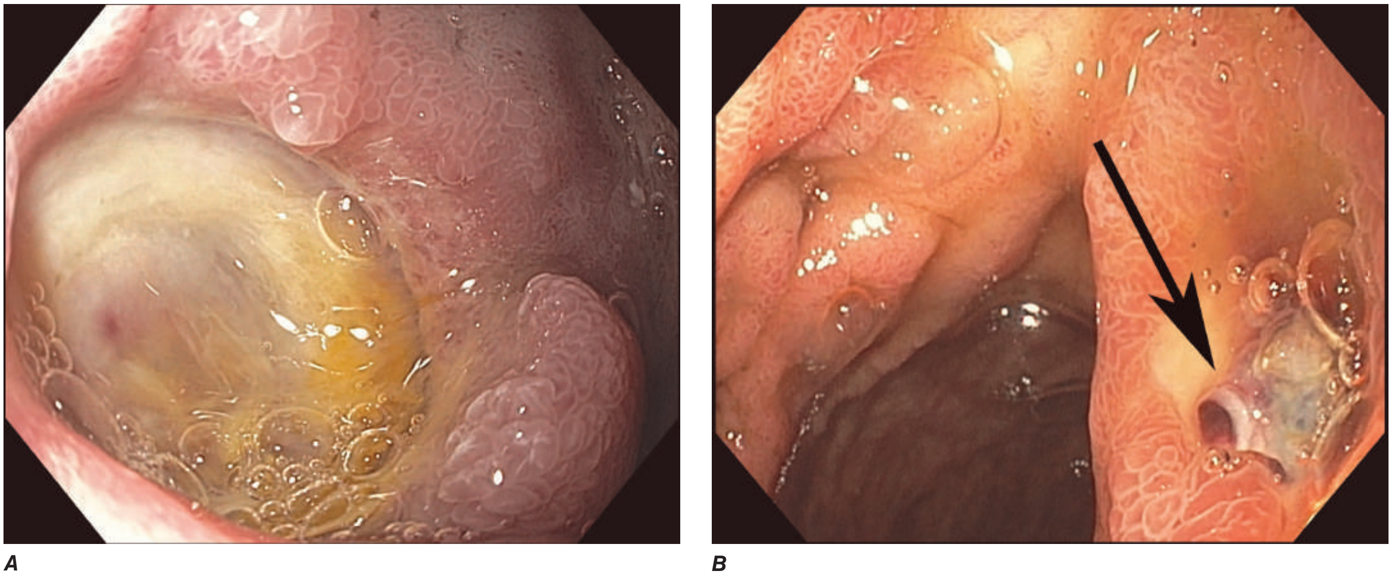

FIGURE 333-4 Duodenal ulcers. ***A.*** Ulcer with a small, flat, pigmented spot in its base. ***B.*** Ulcer with a visible vessel (*arrow*) in a patient with recent hemorrhage.

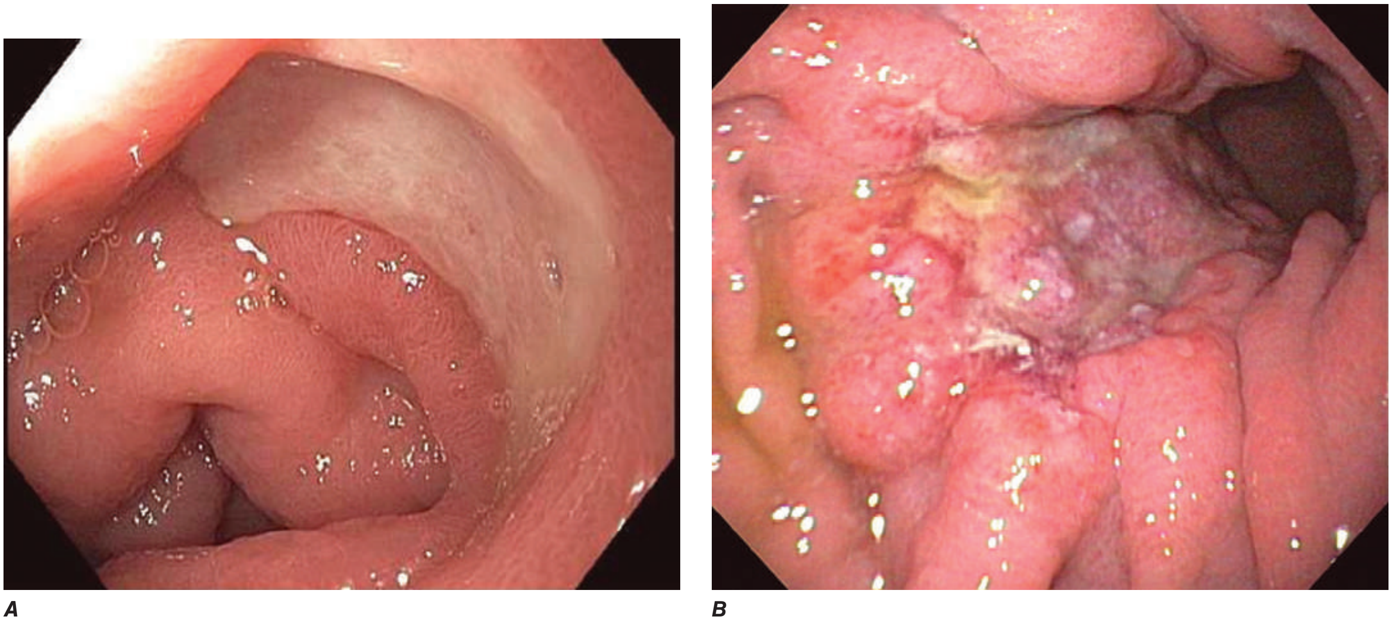

FIGURE 333-5 Gastric ulcers. ***A.*** Benign gastric ulcer in the antrum. ***B.*** Malignant gastric ulcer involving greater curvature of stomach.

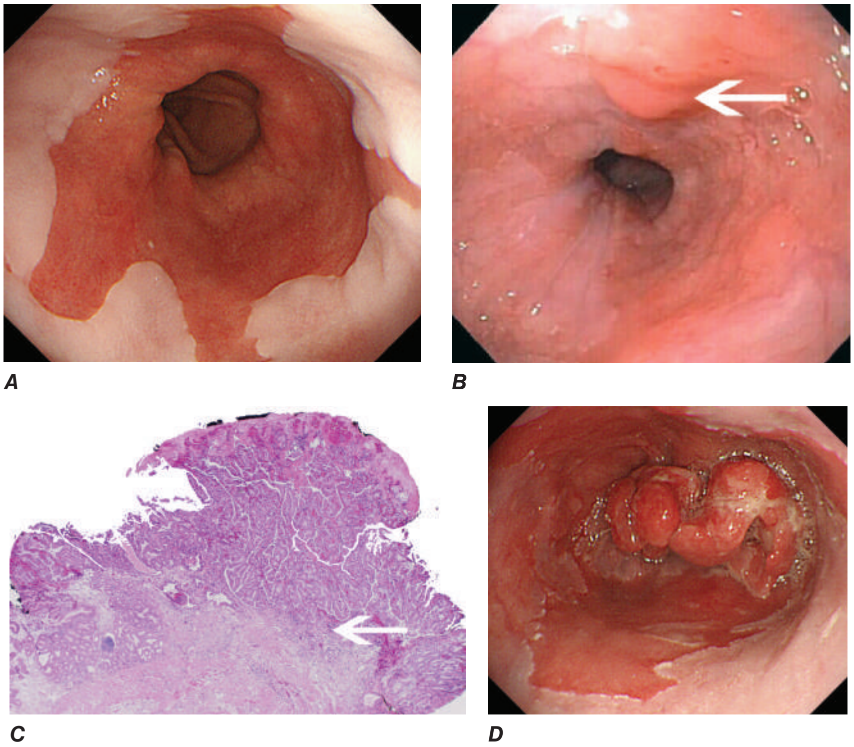

FIGURE 333-6 Barrett's esophagus. ***A.*** Salmon-colored Barrett's mucosa extending proximally from the gastroesophageal junction. ***B.*** Barrett's esophagus with a suspicious nodule (*arrow*) identified during endoscopic surveillance. ***C.*** Histologic finding of intramucosal adenocarcinoma in the endoscopically resected nodule. Tumor extends into the esophageal submucosa (*arrow*). ***D.*** Barrett's esophagus with locally advanced adenocarcinoma.

and at present, it is solely a diagnostic procedure. Patients with a history of prior intestinal surgery or Crohn's disease are at risk for capsule retention at the site of a clinically unsuspected small-bowel stricture, and ingestion of a "patency capsule" composed of radiologically opaque biodegradable material may be indicated before capsule endoscopy in such patients.

Push enteroscopy is generally performed using a variable-stiffness pediatric or adult colonoscope, or a dedicated enteroscope with or without the assistance of a stiffening overtube that extends from the mouth to the small intestine. The proximal to mid-jejunum is usually reached, and the instrument channel of the endoscope allows for biopsy or endoscopic therapy.

Deeper intubation of the small bowel can be accomplished by *single- or double-balloon enteroscopy*, which enables pleating of the small intestine onto an overtube **(Fig. 333-14, Video V5-1)**. With balloon-assisted enteroscopy, the entire small intestine can be visualized in some patients when both the oral and anal routes of insertion are used. Biopsies and endoscopic therapy, such as thermal ablation of vascular ectasias and polypectomy, can be performed throughout the visualized small bowel **(Fig. 333-15)**.

■ ENDOSCOPIC RETROGRADE CHOLANGIOPANCREATOGRAPHY

During endoscopic retrograde cholangiopancreatography (ERCP), a side-viewing endoscope is passed through the mouth to the duodenum,

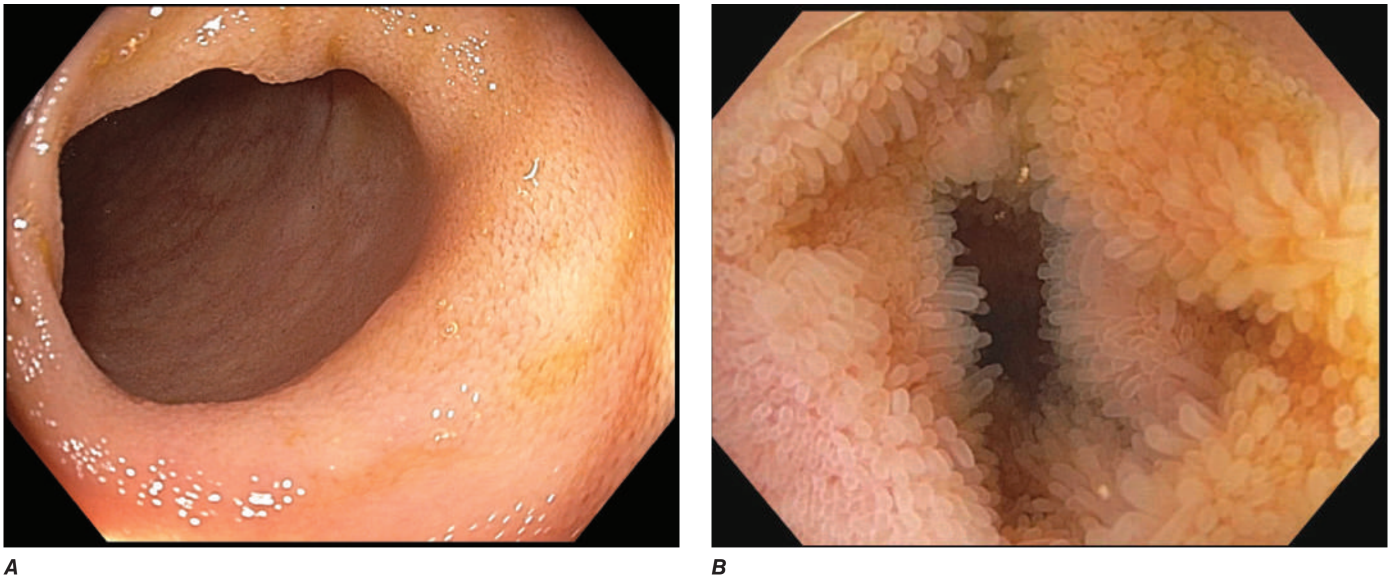

FIGURE 333-7 Colonoscopic view of terminal ileum. ***A.*** Normal-appearing terminal ileum (TI). ***B.*** View of normal villi of TI enhanced by examination under water immersion.

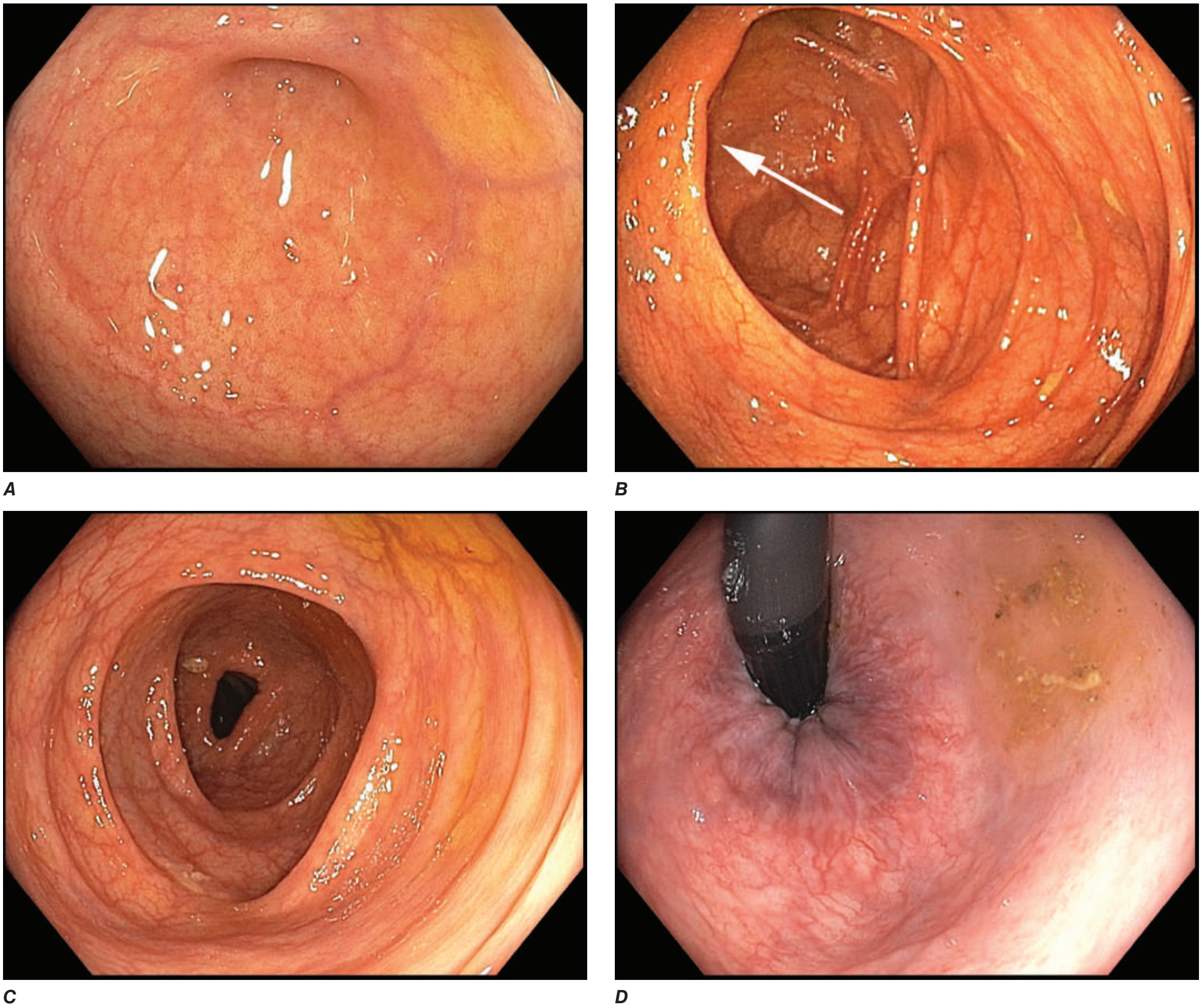

FIGURE 333-8 Normal colonoscopic examination. ***A.*** Cecum with view of appendiceal orifice. ***B.*** Ileocecal valve. ***C.*** Normal-appearing colon. ***D.*** Rectum (retroflexed view).

the bile duct and/or pancreatic duct is cannulated with a thin plastic catheter, and radiographic contrast material is injected under fluoroscopic guidance **(Fig. 333-16)**. When indicated, the major papilla can be incised using the technique of endoscopic sphincterotomy **(Fig. 333-17)**. Stones can be retrieved from the ducts, biopsies can be performed, strictures can be dilated and/or stented **(Fig. 333-18)**, and ductal leaks can be treated **(Fig. 333-19)**. ERCP is usually performed for therapy but is also important diagnostically as it facilitates tissue sampling of biliary or pancreatic ductal strictures.

■ ENDOSCOPIC ULTRASOUND

Endoscopic ultrasound (EUS) utilizes ultrasound transducers incorporated into the tip of a flexible endoscope. Ultrasound images are obtained of the gut wall and adjacent organs, vessels, lymph nodes, and other structures. High-resolution images are obtained by bringing a high-frequency ultrasound transducer close to the area of interest via endoscopy. EUS provides the most accurate preoperative local staging of esophageal, pancreatic, and rectal malignancies **(Fig. 333-20)**, but it does not detect distant metastases that are beyond its imaging range. EUS is also useful for diagnosis of bile duct stones, gallbladder disease, subepithelial gastrointestinal lesions, and chronic pancreatitis. Fine-needle aspirates and core biopsies of organs, masses, and lymph nodes in the posterior mediastinum, abdomen, retroperitoneum, and pelvis can be obtained under EUS guidance **(Fig. 333-21)**. EUS-guided therapeutic procedures are increasingly performed, including drainage of abscesses, pseudocysts, and pancreatic necrosis into the gut lumen **(Video V5-2)**; celiac plexus neurolysis for treatment of pancreatic pain; ethanol ablation of pancreatic neuroendocrine tumors; treatment of gastric and intestinal varices; biliary or pancreatic drainage; and endoscopic gastrojejunostomy for palliation of malignant gastric outlet obstruction **(Video V5-3)**.

■ NATURAL ORIFICE TRANSLUMINAL ENDOSCOPIC SURGERY

Natural orifice transluminal endoscopic surgery (NOTES) represents a collection of endoscopic methods that entail passage of an endoscope or its accessories into or through the wall of the gastrointestinal tract to perform diagnostic or therapeutic interventions. Some NOTES procedures, such as percutaneous endoscopic gastrostomy (PEG) or endoscopic necrosectomy of pancreatic necrosis, are well-established clinical procedures **(Video V5-2)**; others such as peroral endoscopic myotomy (POEM) for achalasia **(Fig. 333-22)** and gastroparesis, peroral endoscopic tumorectomy (POET) **(Fig. 333-23)**, and endoscopic full-thickness resection (EFTR) of gastrointestinal mural lesions **(Fig. 333-24, Video V5-4)**, are newer minimally invasive therapeutic options.

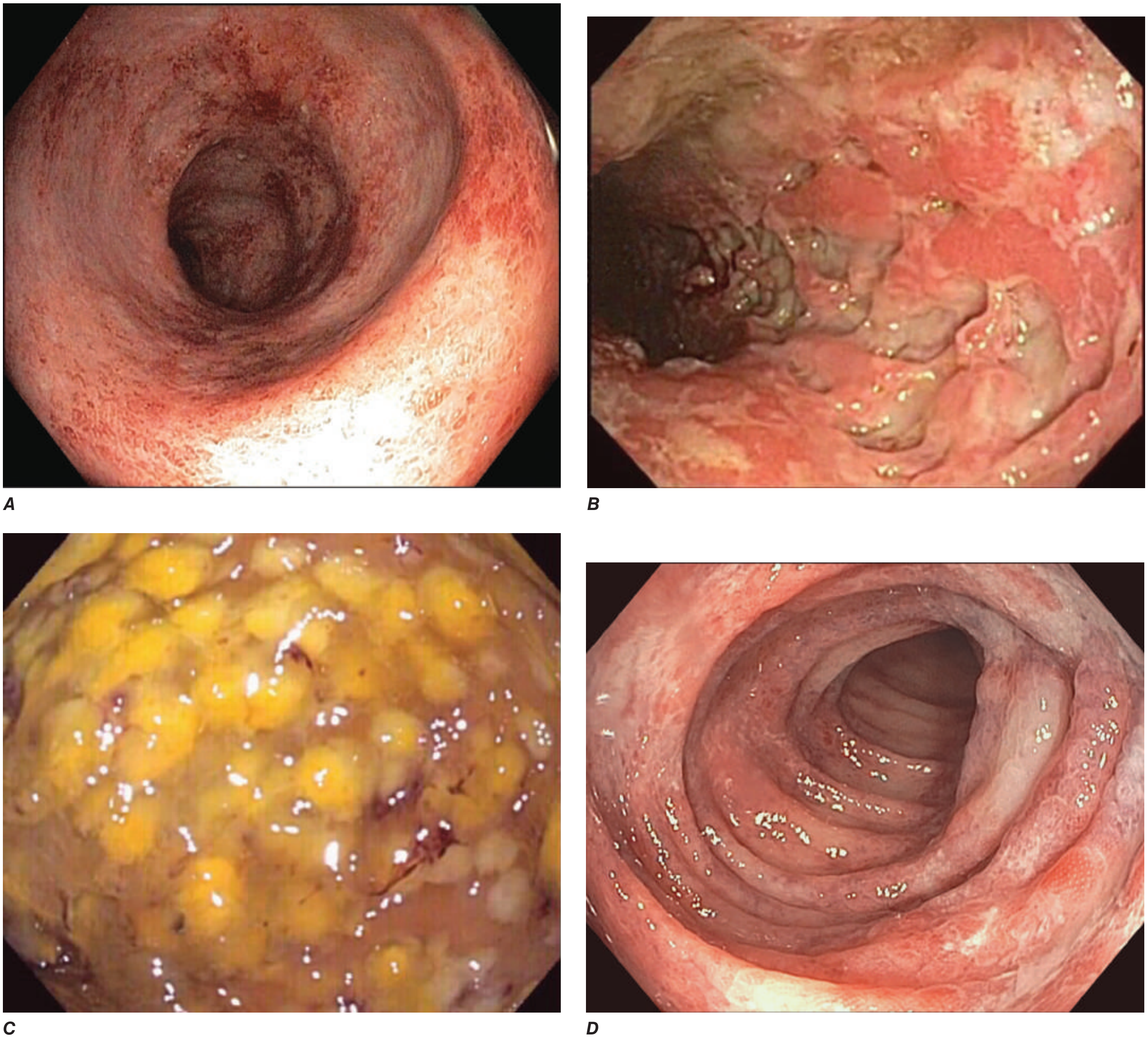

FIGURE 333-9 Causes of colitis. ***A.*** Chronic ulcerative colitis with diffuse inflammation. ***B.*** Severe Crohn's colitis with deep ulcers. ***C.*** Pseudomembranous colitis with yellow, adherent pseudomembranes. ***D.*** Ischemic colitis with patchy mucosal edema, subepithelial hemorrhage, superficial ulcerations, and cyanosis.

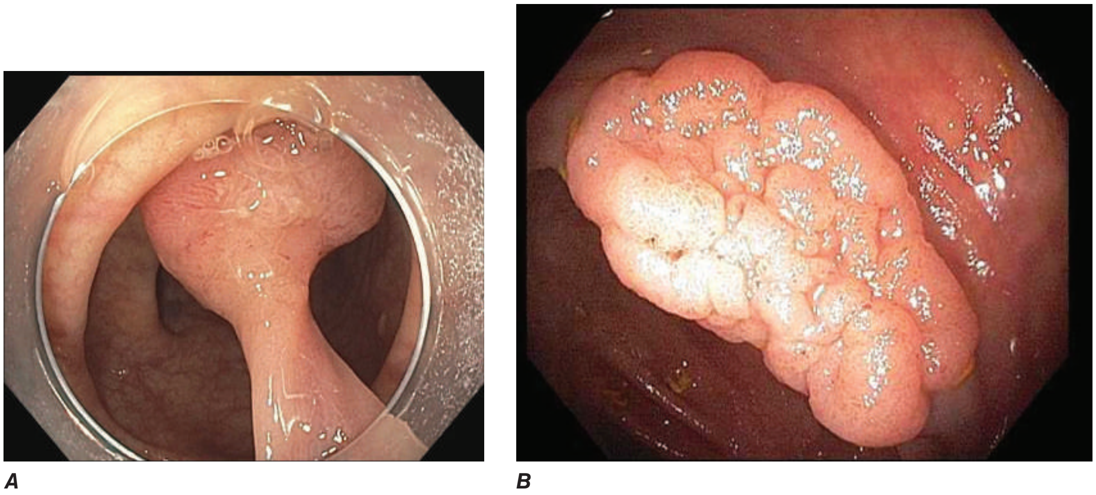

FIGURE 333-10 Colonic polyps. ***A.*** Pedunculated polyp on a stalk. ***B.*** Sessile polyp.

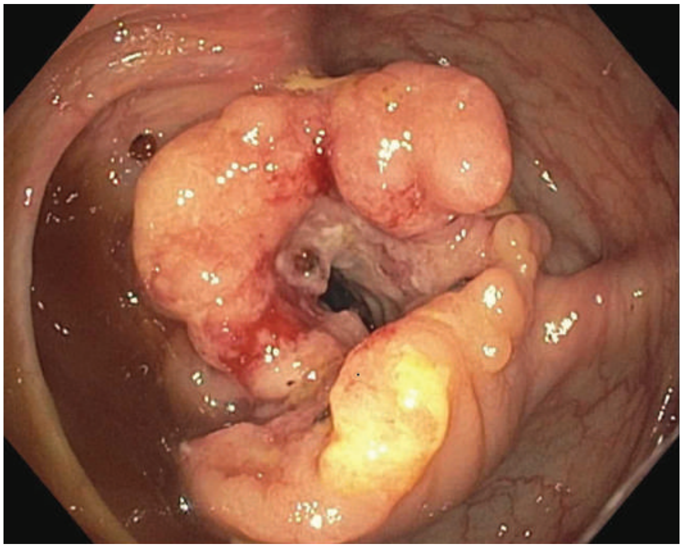

FIGURE 333-11 Ulcerated colon adenocarcinoma narrowing the colonic lumen.

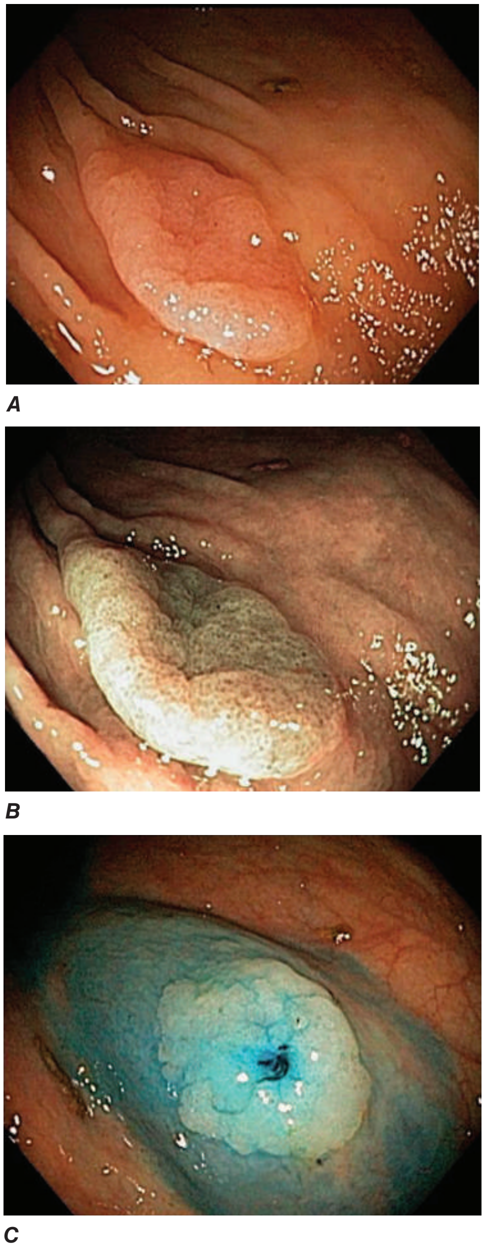

FIGURE 333-12 Flat serrated polyp in the cecum. ***A.*** Appearance of the lesion under conventional white-light imaging. ***B.*** Mucosal patterns and boundary of the lesion enhanced with narrow-band imaging. ***C.*** Submucosal lifting of the lesion with dye (methylene blue) injection before resection.

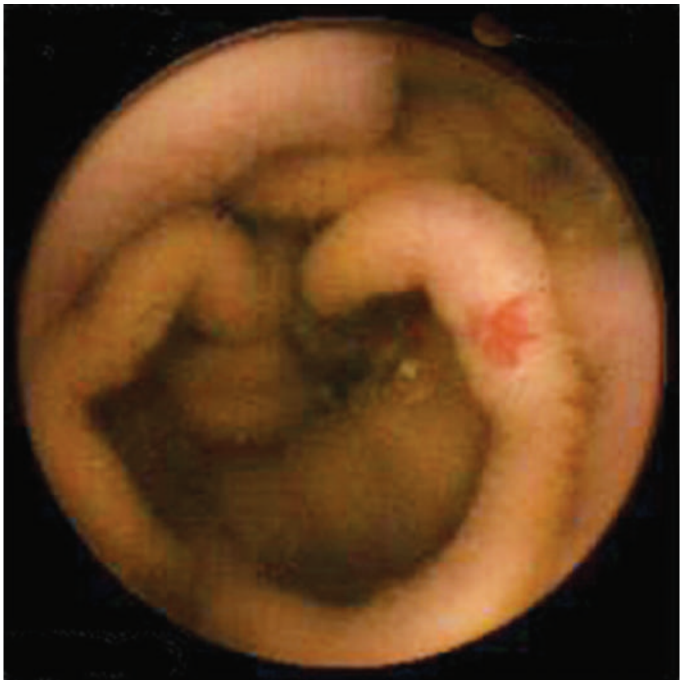

FIGURE 333-13 Capsule endoscopy. Image of a jejunal vascular ectasia.

ENDOSCOPIC RESECTION AND CLOSURE TECHNIQUES

Endoscopic mucosal resection (EMR) **(Fig. 333-25, Video V5-5)** and endoscopic submucosal dissection (ESD) **(Fig. 333-26, Video V5-6)** are the two commonly used techniques for the resection of benign and early-stage malignant gastrointestinal neoplasms. In addition to providing larger specimens for more accurate histopathologic assessment and diagnosis, these techniques may be curative for some dysplastic lesions and superficial carcinomas involving the esophagus, stomach, and colon.

Several devices are available for closure of mucosal defects created by EMR and ESD, as well as gastrointestinal fistulas and perforations. Endoscopic clips deployed through the working channel of an

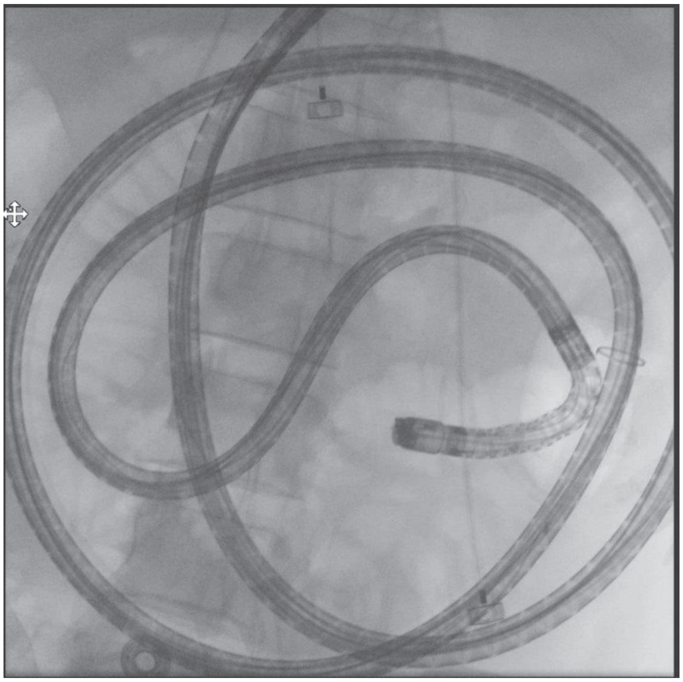

FIGURE 333-14 Double-balloon enteroscopy. Radiograph of the orally inserted instrument deep in the small intestine.

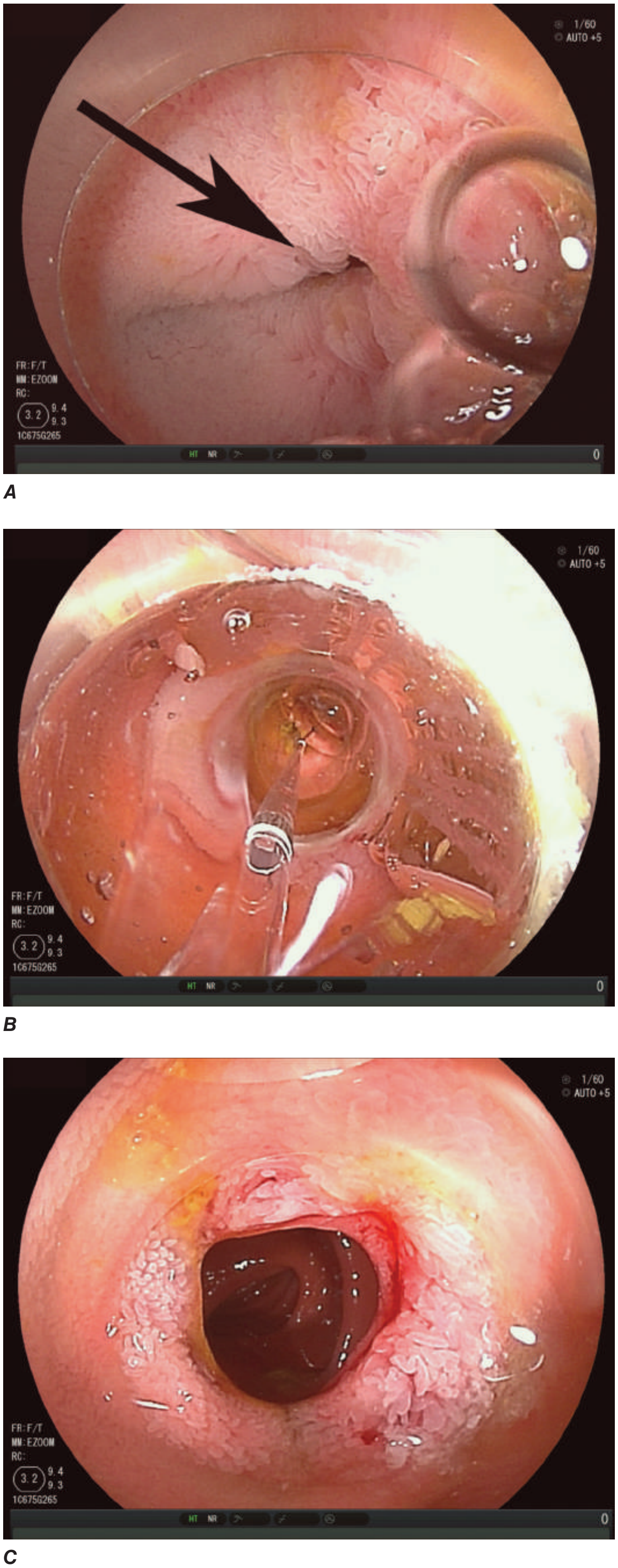
1/60
AUTO +5
FR:F/T
MM:EZOOM
RC:
3.2
9.4
9.3
1C675G265
HT
NR
0
A
B
C

FIGURE 333-15 Nonsteroidal anti-inflammatory drug (NSAID)–induced proximal ileal stricture managed via double-balloon enteroscopy. ***A.*** High-grade ileal stricture causing obstructive symptoms. ***B.*** Balloon dilation of the ileal stricture. ***C.*** Appearance of the stricture after dilation.

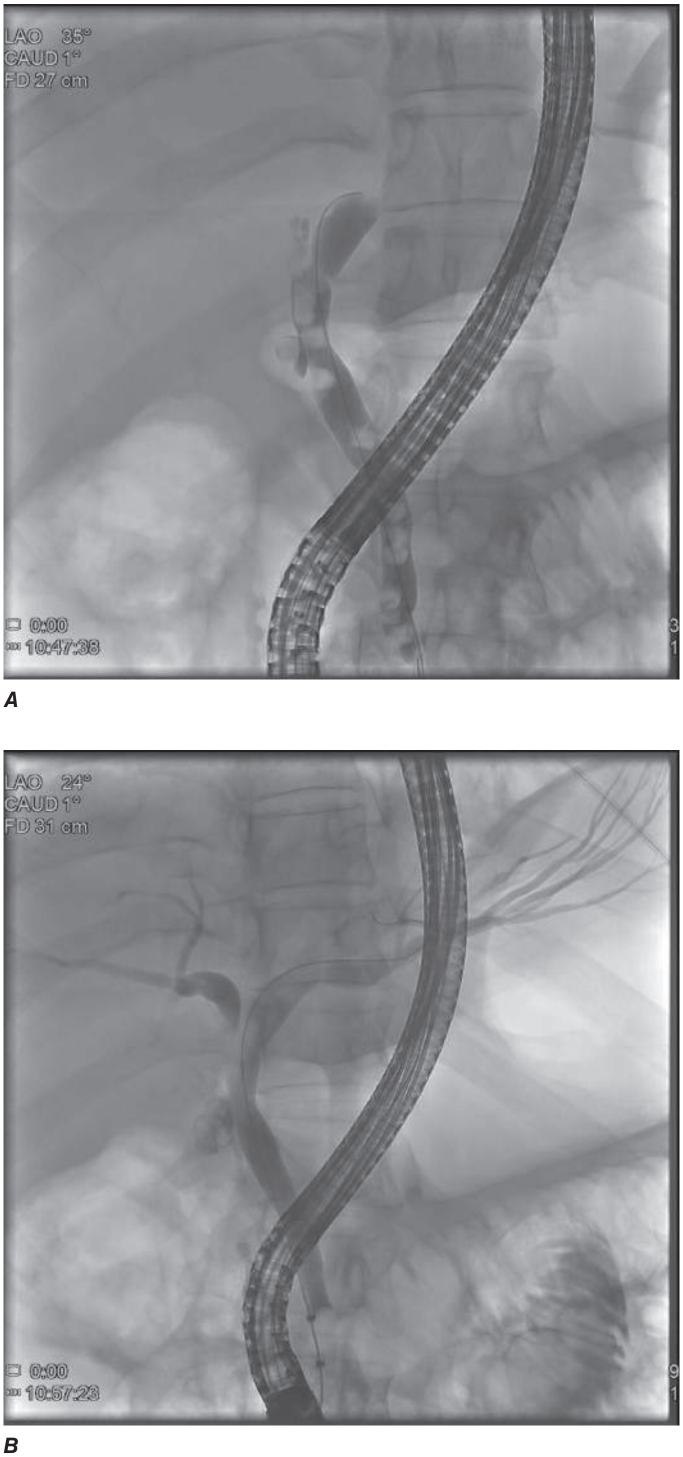
LAO 35°
CAUD 1°
FD 27 cm
0:00
10:47:38
A
LAO 24°
CAUD 1°
FD 31 cm
0:00
10:57:23
B

FIGURE 333-16 Endoscopic retrograde cholangiopancreatography (ERCP) for bile duct stones. ***A.*** Faceted bile duct stones are demonstrated in the common bile duct and common hepatic duct. ***B.*** After endoscopic sphincterotomy, the stones are extracted with a stone extraction balloon.

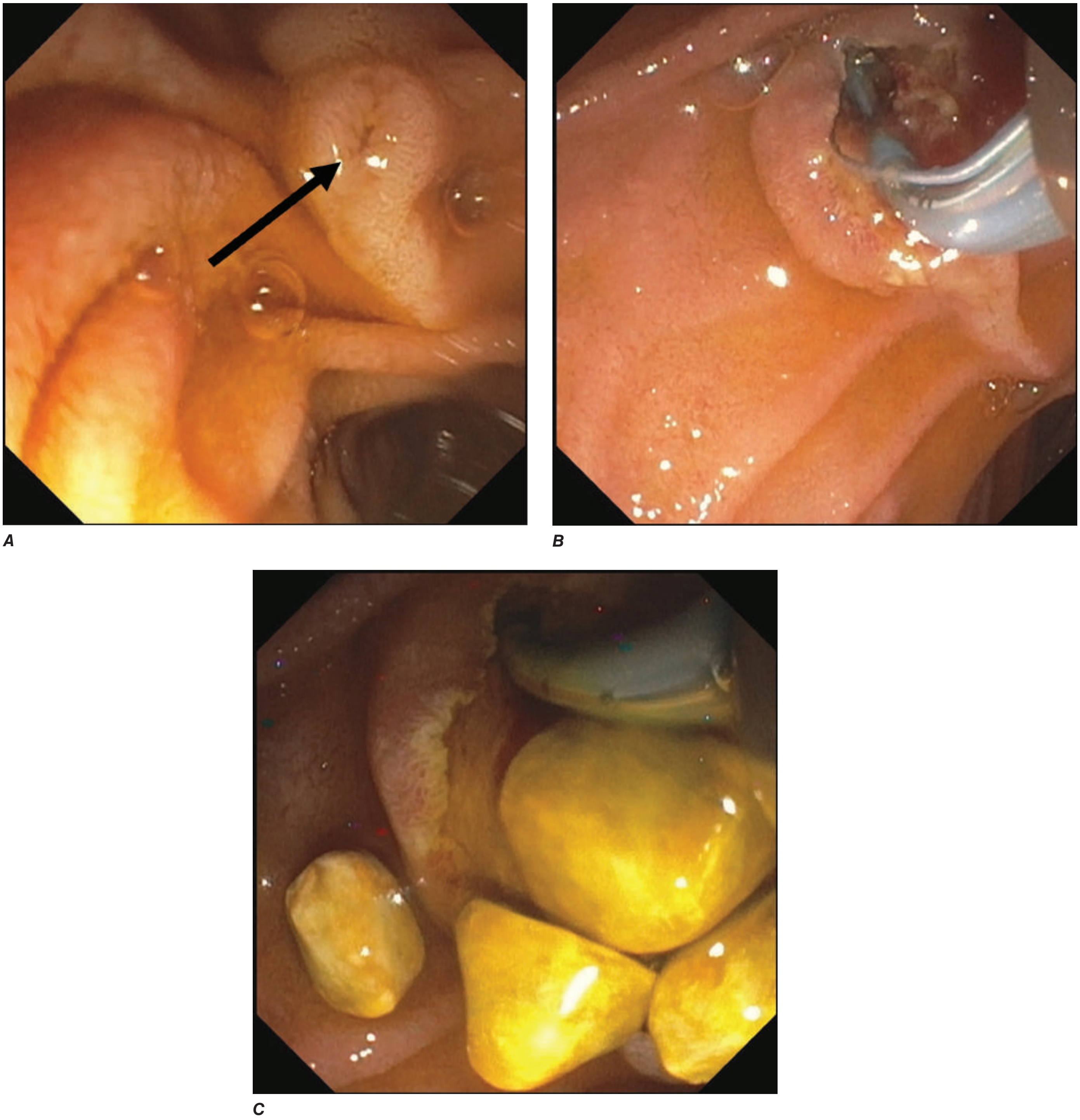

FIGURE 333-17 Endoscopic sphincterotomy. ***A.*** A normal-appearing ampulla of Vater (*arrow*). ***B.*** Biliary endoscopic sphincterotomy is performed with electrosurgery. ***C.*** Bile duct stones are extracted with a balloon catheter.

endoscope have been used for many years to treat bleeding lesions, and the development of larger over-the-scope clips has facilitated endoscopic closure of gastrointestinal fistulas and perforations not previously amenable to endoscopic therapy (Video V5-7). Endoscopic suturing can be used to close some perforations and large defects (Fig. 333-27), anastomotic leaks, and fistulas. Endoscopic suturing may also be used to prevent stent migration (Fig. 333-28, Video V5-8) and to perform endoscopic bariatric procedures. These technologies are playing an expanding role in patient care.

RISKS OF ENDOSCOPY

Medications used during sedation may cause respiratory depression or allergic reactions. All endoscopic procedures carry some risk of bleeding and gastrointestinal perforation. The risk is small with diagnostic upper endoscopy, flexible sigmoidoscopy, and colonoscopy (<1:1000 procedures), but it ranges from 0.5 to 5% when therapeutic maneuvers such as polypectomy, EMR, ESD, control of hemorrhage, or stricture dilation are performed. The risk of adverse events for diagnostic EUS (without needle aspiration) is similar to that for diagnostic upper endoscopy.

Infectious complications are uncommon with most endoscopic procedures. Some procedures carry a higher incidence of postprocedural bacteremia, and prophylactic antibiotics may be indicated (Table 333-1). Management of antithrombotic agents before endoscopic procedures should take into account the procedural risk of hemorrhage, the agent, and the patient condition, as summarized in Table 333-2.

ERCP carries additional risks. Pancreatitis occurs in ~5% of patients undergoing the procedure, and young, anicteric patients with normal

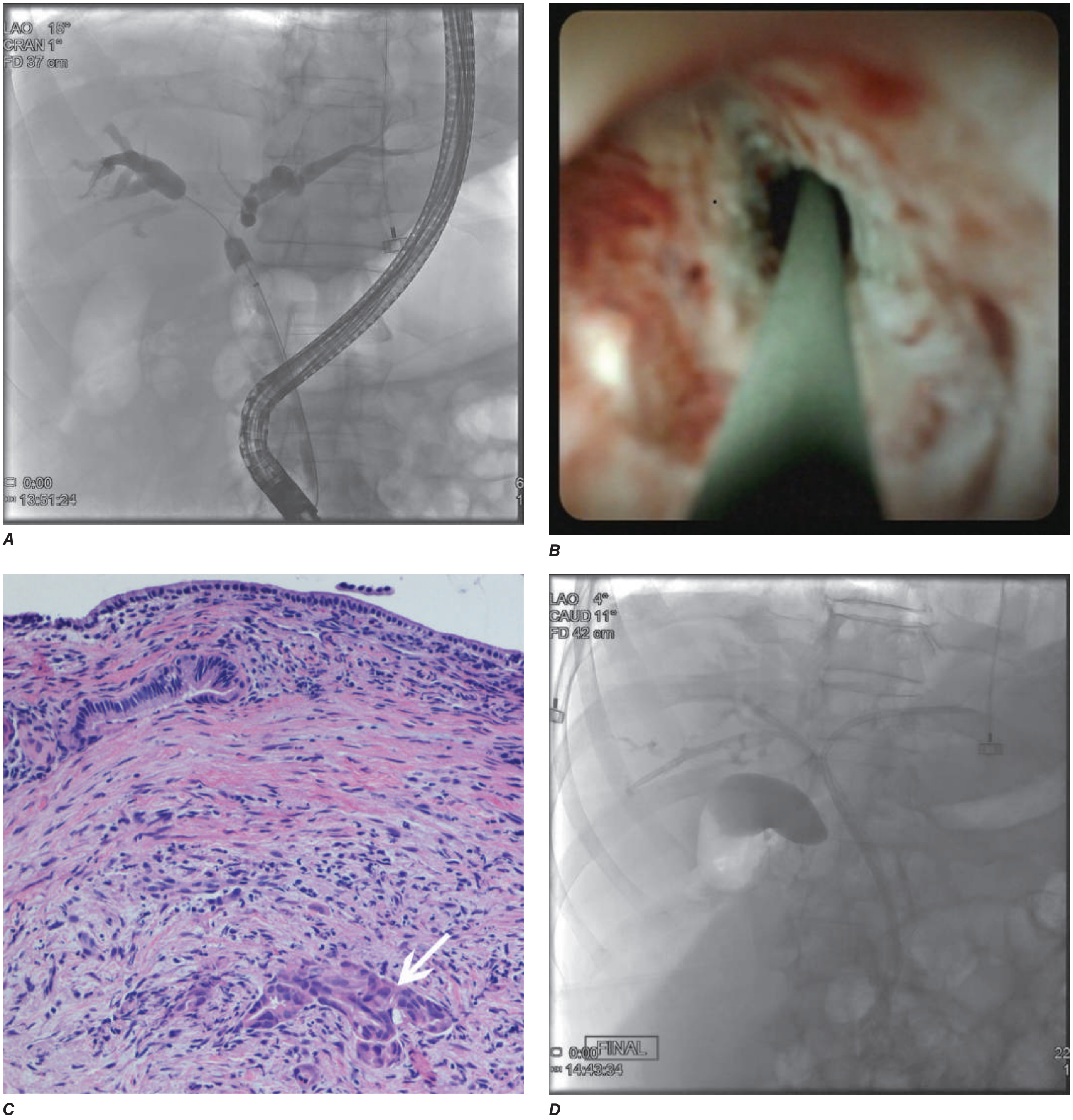

FIGURE 333-18 Endoscopic diagnosis, staging, and palliation of hilar cholangiocarcinoma. ***A.*** Endoscopic retrograde cholangiopancreatography (ERCP) in a patient with obstructive jaundice demonstrates a malignant-appearing stricture of the biliary confluence extending into the left and right intrahepatic ducts. ***B.*** Per oral cholangioscopy demonstrating a stricture with dilated, tortuous vessels with a malignant appearance. ***C.*** Intraductal biopsy obtained during ERCP demonstrates malignant cells infiltrating the submucosa of the bile duct wall (*arrow*). *(Image courtesy of Dr. Thomas Smyrk.)* ***D.*** Endoscopic placement of multiple plastic stents draining the right anterior, right posterior, and left systems relieves the biliary obstruction.

ducts are at increased risk (up to 25%). Post-ERCP pancreatitis is usually mild and self-limited, but it may result in prolonged hospitalization, surgery, diabetes, or death when severe. Significant bleeding occurs after endoscopic sphincterotomy in ~1% of cases. Ascending cholangitis, pseudocyst infection, duodenal perforation, and abscess formation may occur as a result of ERCP.

PEG tube placement during EGD is associated with a 10–15% incidence of adverse events, most often wound infections. Fasciitis, pneumonia, bleeding **(Fig. 333-29)**, buried bumper syndrome **(Fig. 333-30)**, and colonic injury may result from PEG tube placement.

URGENT ENDOSCOPY

■ ACUTE GASTROINTESTINAL HEMORRHAGE

Endoscopy is the primary diagnostic and therapeutic procedure for patients with acute gastrointestinal hemorrhage. Although gastrointestinal bleeding stops spontaneously in most cases, some patients will have persistent or recurrent hemorrhage that may be life-threatening. Clinical predictors of rebleeding help identify patients most likely to benefit from urgent endoscopy and endoscopic, angiographic, or surgical hemostasis.

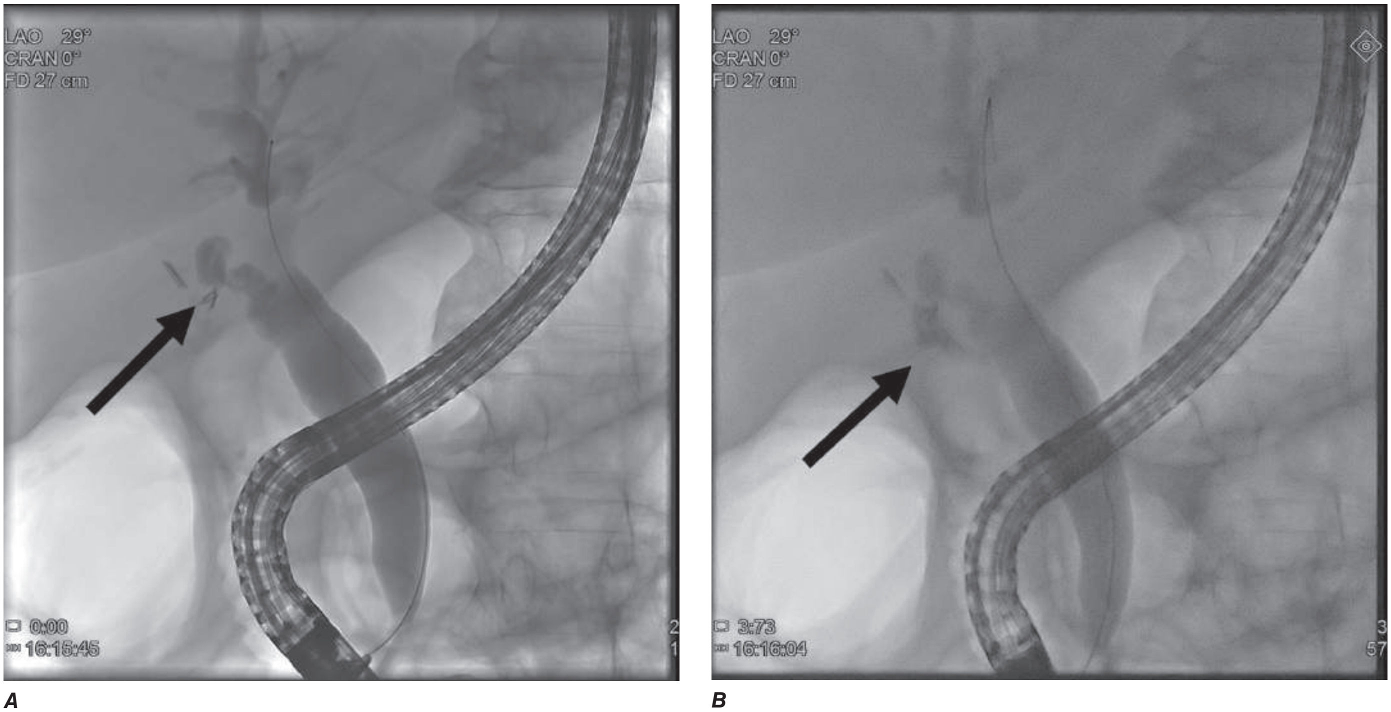

FIGURE 333-19 Bile leak. ***A.*** Site of leak (*arrow*) from the cystic duct after laparoscopic cholecystectomy. ***B.*** Contrast leaks from the cystic duct stump across surgical clips into the gallbladder fossa (*arrow*).

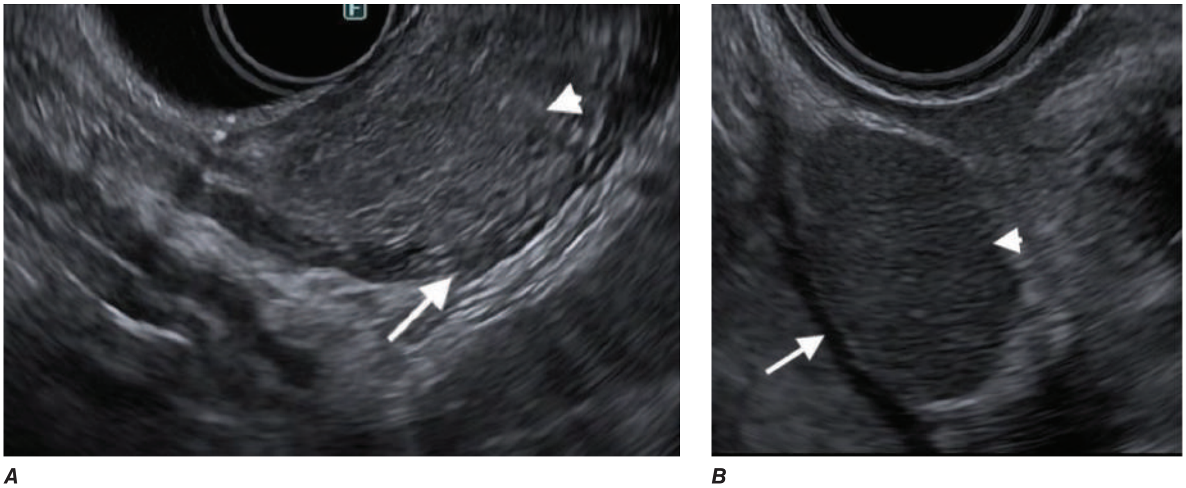

FIGURE 333-20 Local staging of gastrointestinal cancers with endoscopic ultrasound. In each example, the *arrowhead* marks the primary tumor and the *arrow* indicates the muscularis propria (MP) of the intestinal wall. ***A.*** T2 gastric cancer. The tumor invades the MP. ***B.*** Submucosal gastric tumor. The tumor is confined to the submucosal space without invasion into the MP.

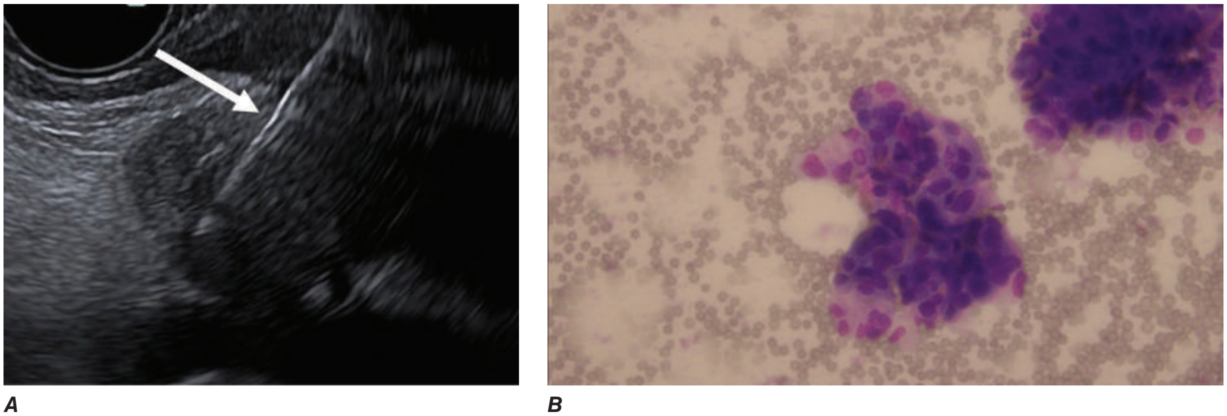

FIGURE 333-21 Endoscopic ultrasound (EUS)–guided tissue sampling. ***A.*** Ultrasound image of a 22-gauge needle (*arrow*) passed through the gastric wall and positioned in a hypoechoic pancreatic neck mass. ***B.*** Touch preparation demonstrating aspirated malignant cells.

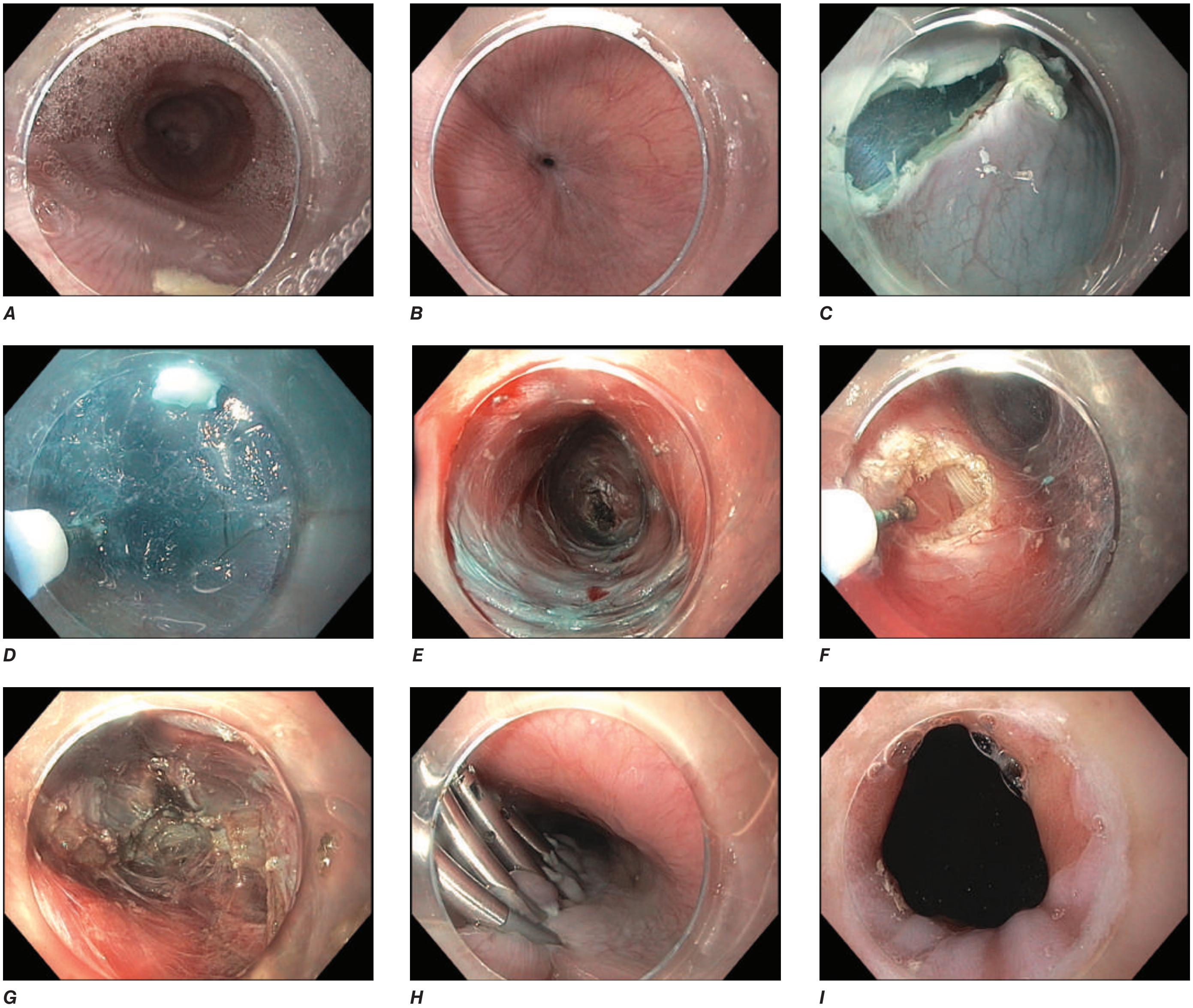

FIGURE 333-22 Peroral endoscopic myotomy (POEM) for achalasia. ***A.*** Dilated aperistaltic esophagus with retained secretions. ***B.*** Hypertonic lower esophageal sphincter (LES) region. ***C.*** Mucosal incision (mucosotomy) 10 cm proximal to the LES. ***D.*** Submucosal dissection using an electrosurgical knife following endoscope entry through the mucosotomy site into the submucosal space. ***E.*** Completion of submucosal tunnel to the cardia. ***F.*** Initiation of myotomy of the muscularis propria distal to the mucosotomy site. ***G.*** Completion of myotomy to the cardia. ***H.*** Closure of mucosotomy site with clips. ***I.*** Patulous gastroesophageal junction following myotomy.

Initial Evaluation The initial evaluation of the bleeding patient focuses on the severity of hemorrhage as reflected by the presence of supine hypotension or tachycardia, postural vital sign changes, and the frequency of hematemesis or melena. Decreases in hematocrit and hemoglobin lag behind the clinical course and are not reliable gauges of the magnitude of acute bleeding. Nasogastric tube aspiration and lavage can also be used to judge the severity of bleeding, but these are no longer routinely performed for this purpose. The bedside initial evaluation, completed well before the bleeding source is confidently identified, guides immediate supportive care of the patient; triage to outpatient follow-up, a hospital ward, or an intensive care unit; and timing of endoscopy. The severity of the initial hemorrhage is the most important indication for urgent endoscopy, since a large initial bleed increases the likelihood of ongoing or recurrent bleeding. Patients with resting hypotension or orthostatic change in vital signs, repeated hematemesis, bloody nasogastric aspirate that does not clear with large-volume lavage, or those requiring blood transfusions should be considered for urgent endoscopy within 12–24 h of presentation. In addition, patients with cirrhosis, coagulopathy, or respiratory or renal failure and those >70 years old are more likely to have significant rebleeding and to benefit from prompt evaluation and treatment.

Bedside evaluation also suggests an upper or lower gastrointestinal source of bleeding in most patients. Over 90% of patients with melena are bleeding proximal to the ligament of Treitz, and ~85% of patients with hematochezia are bleeding from the colon. Melena can result from bleeding in the small bowel or right colon, especially in older patients with slow colonic transit. Conversely, some patients with massive hematochezia may be bleeding from an upper gastrointestinal source, with rapid intestinal transit. An urgent upper endoscopy should be considered in such patients.

Endoscopy should be performed after the patient has been resuscitated with intravenous fluids and transfusions, as necessary. Marked coagulopathy or thrombocytopenia is usually treated before endoscopy, since correction of these abnormalities may lead to resolution of bleeding, and techniques for endoscopic hemostasis are limited in such patients. Metabolic derangements should also be addressed. Tracheal intubation for airway protection should be considered before upper endoscopy in patients with repeated recent hematemesis, particularly in those with suspected variceal hemorrhage. A single dose of erythromycin (3–4 mg/kg or 250 mg) administered intravenously 30–90 min before upper endoscopy increases gastric emptying and may clear blood and clots from the stomach to improve endoscopic visualization.

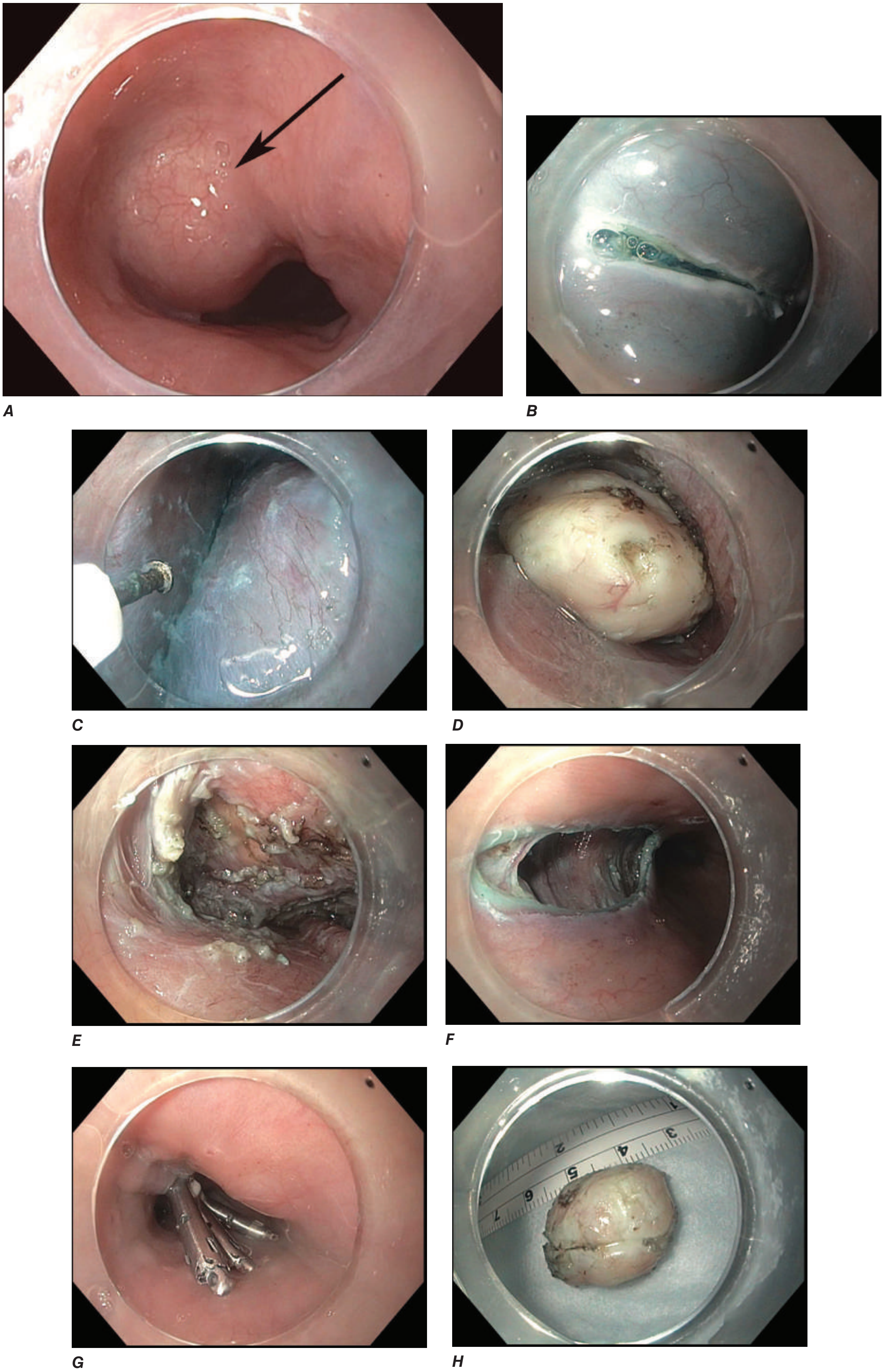

FIGURE 333-23 Peroral endoscopic tumorectomy (POET). ***A.*** Mid-esophageal subepithelial lesion (*arrow*). ***B.*** Mucosal incision (mucosotomy) 5 cm proximal to the lesion. ***C.*** Submucosal dissection and tunneling to the site of the lesion. ***D.*** Dissection of the lesion from its attachment to the muscularis propria. ***E.*** Postresection defect through the muscularis propria. ***F.*** Mucosotomy site. ***G.*** Closure of mucosotomy site with clips. ***H.*** Resected specimen (leiomyoma).

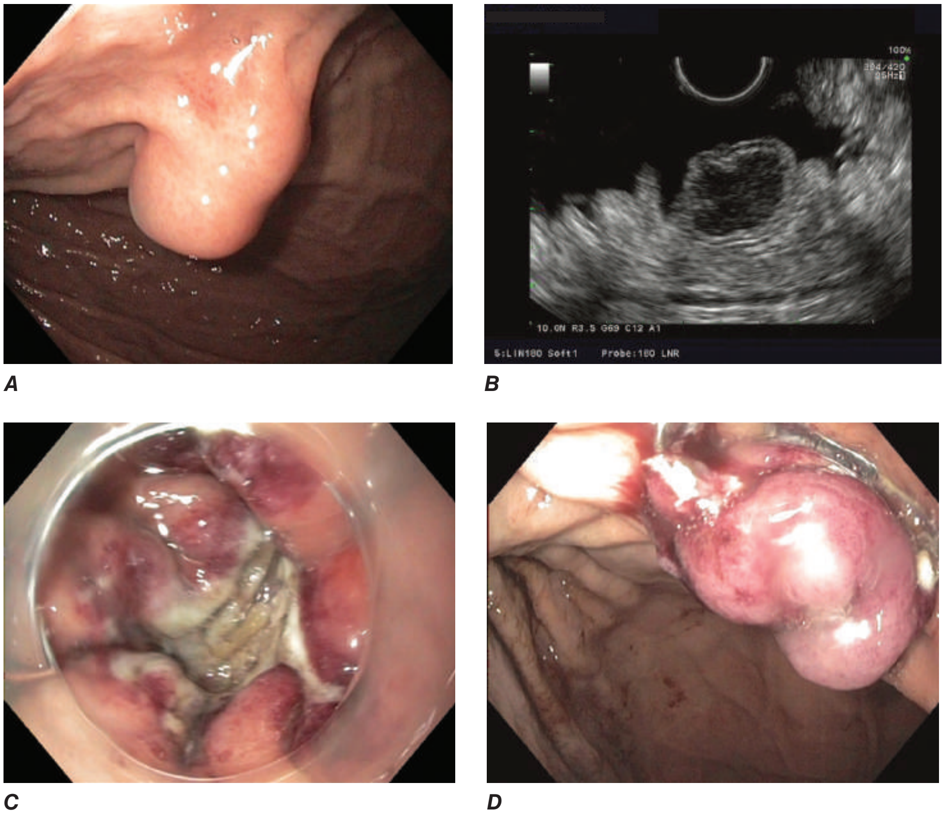

FIGURE 333-24 Endoscopic full-thickness resection (EFTR) of a gastrointestinal stromal tumor. ***A.*** Subepithelial lesion in the proximal stomach. ***B.*** Hypoechoic lesion arising from the fourth layer (muscularis propria) at endoscopic ultrasound. ***C.*** Full-thickness resection defect. ***D.*** Closure of defect using an over-the-scope clip.

Most patients with hematochezia who are otherwise stable can undergo semi-elective colonoscopy. Controlled trials have not shown a benefit to urgent colonoscopy (within 24 h of presentation) in patients hospitalized with hematochezia, although selected patients with massive or recurrent large-volume episodes of hematochezia should probably undergo urgent colonoscopy after both upper endoscopy and a rapid colonic purge with an oral polyethylene glycol solution. Colonoscopy has a higher diagnostic yield than radionuclide bleeding scans or catheter-based angiography in lower gastrointestinal bleeding, and endoscopic therapy can be applied in some cases. Urgent colonoscopy can be hindered by poor visualization due to persistent vigorous bleeding with recurrent hemodynamic instability, and other techniques (e.g., angiography or even emergent subtotal colectomy) must be employed. The anal and rectal mucosa should also be visualized endoscopically early in the course of massive rectal bleeding, as bleeding lesions in or close to the anal canal may be identified that are amenable to endoscopic or surgical transanal hemostatic techniques.

Peptic Ulcer The endoscopic appearance of peptic ulcers provides useful prognostic information and guides the need for endoscopic

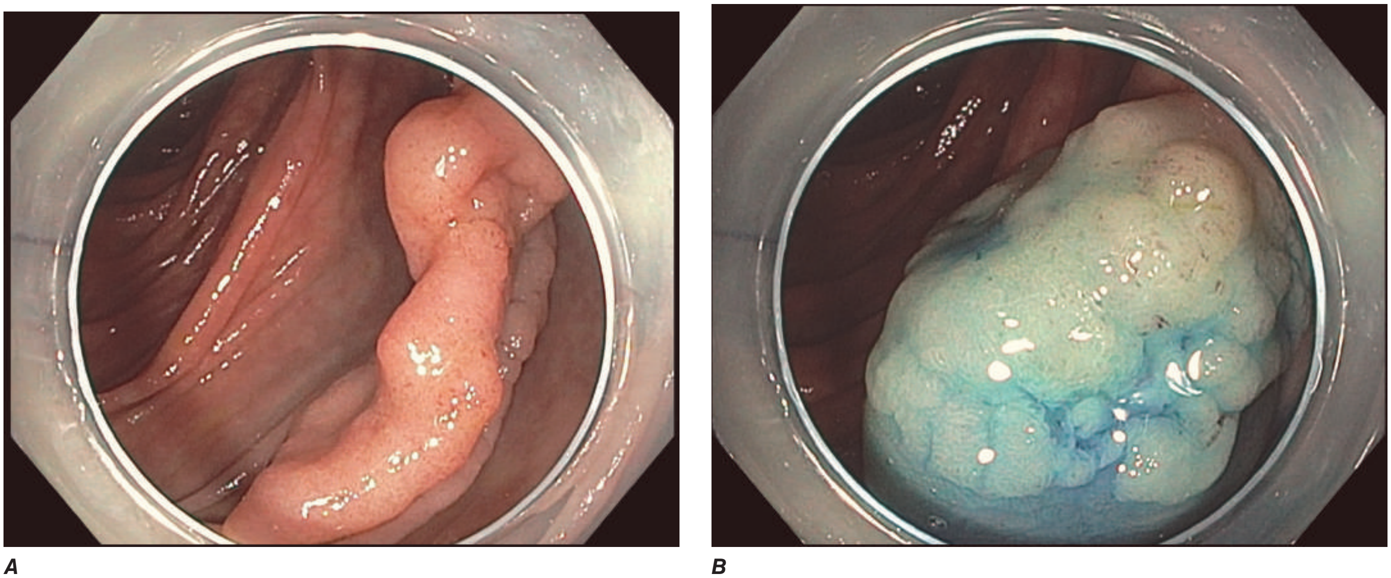

FIGURE 333-25 Endoscopic mucosal resection (EMR). ***A.*** Large sessile polypoid fold in the transverse colon. ***B.*** Lifting of lesion following submucosal fluid injection. ***C.*** Piecemeal hot snare resection. ***D.*** Initial resection site. ***E.*** Resection defect following completion of piecemeal EMR.

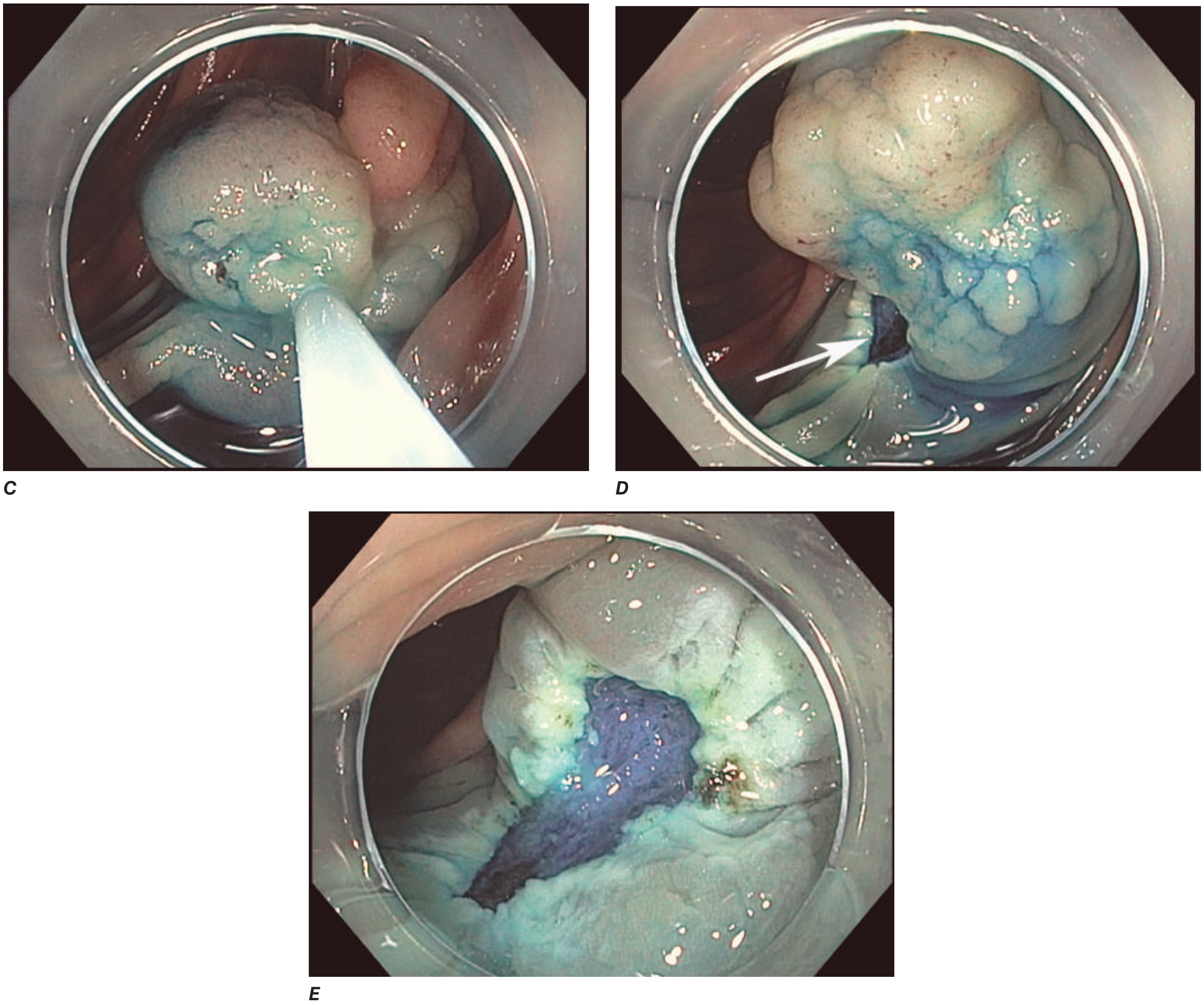

FIGURE 333-25 (*Continued*)

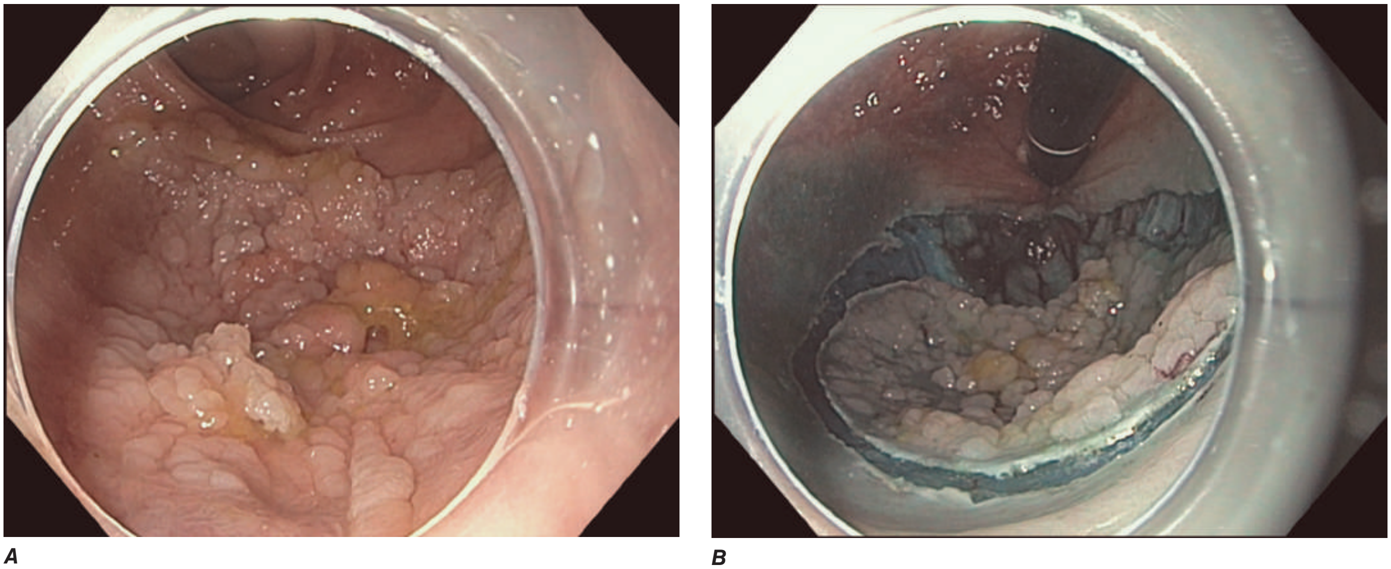

FIGURE 333-26 Endoscopic submucosal dissection (ESD). ***A.*** Large, flat, distal rectal adenoma. ***B.*** Circumferential incision following submucosal fluid injection at the periphery of the lesion. ***C.*** ESD using an electrosurgical knife. ***D.*** Rectal defect following ESD. ***E.*** Specimen resected en bloc.

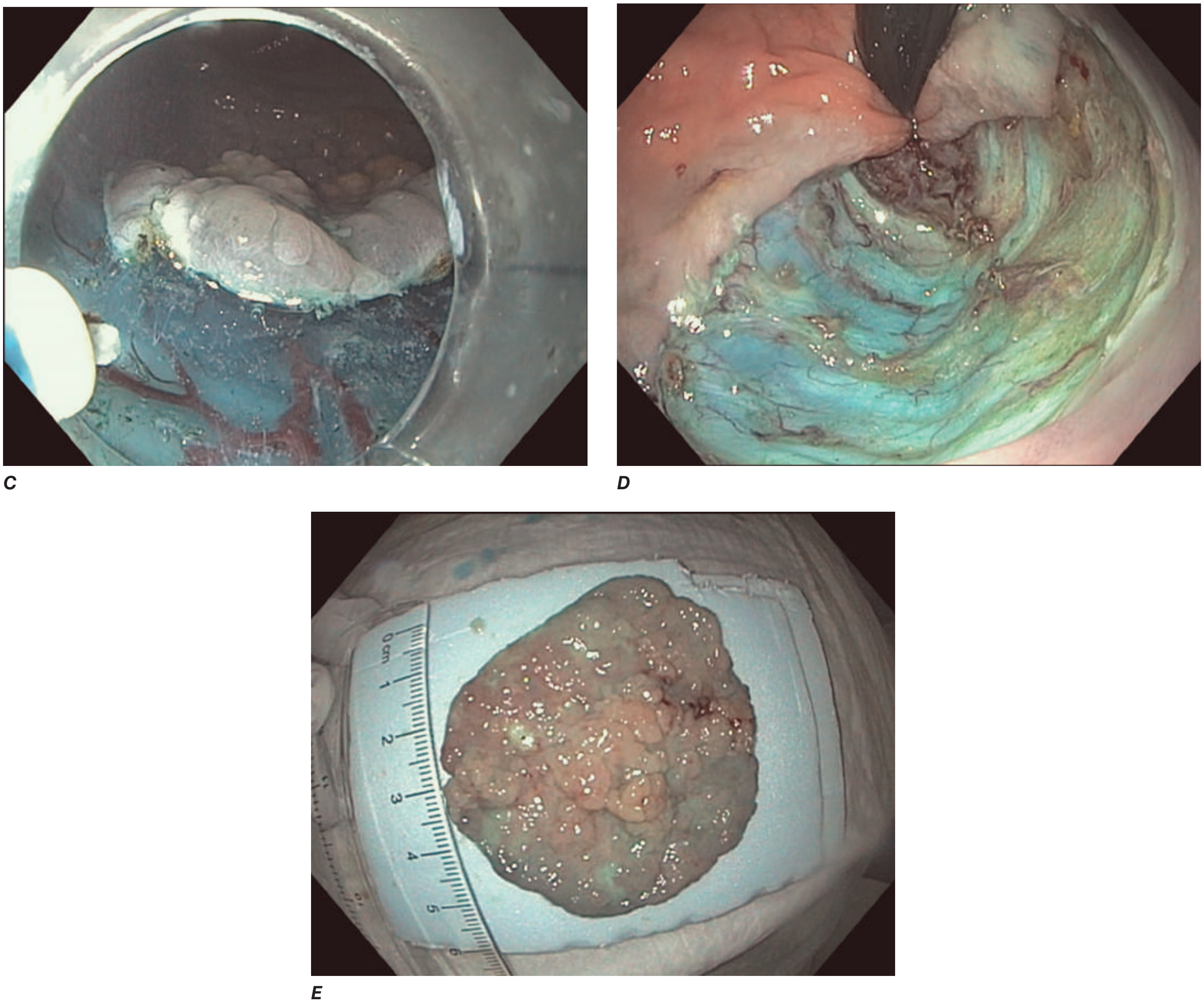

FIGURE 333-26 (*Continued*)

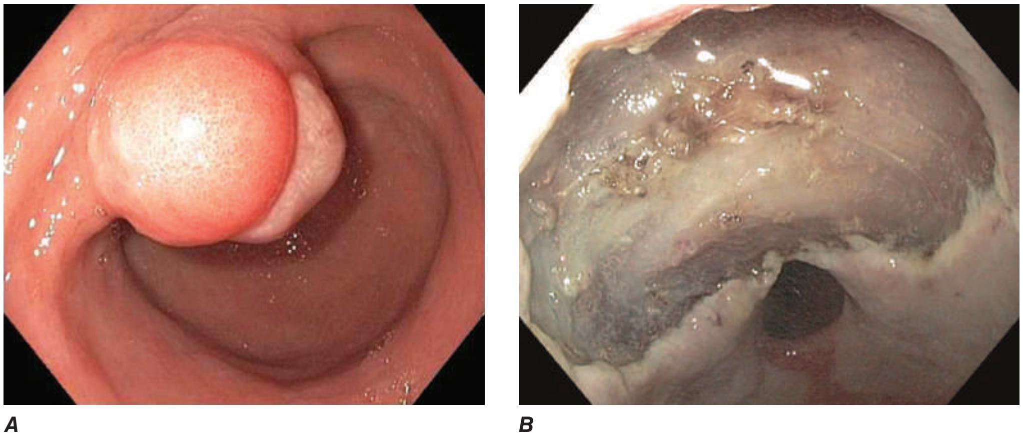

FIGURE 333-27 Closure of large defect using an endoscopic suturing device. ***A.*** Ulcerated inflammatory fibroid polyp in the antrum. ***B.*** Large defect following endoscopic submucosal dissection of the lesion. ***C.*** Closure of the defect using endoscopic sutures (*arrows*). ***D.*** Resected specimen.

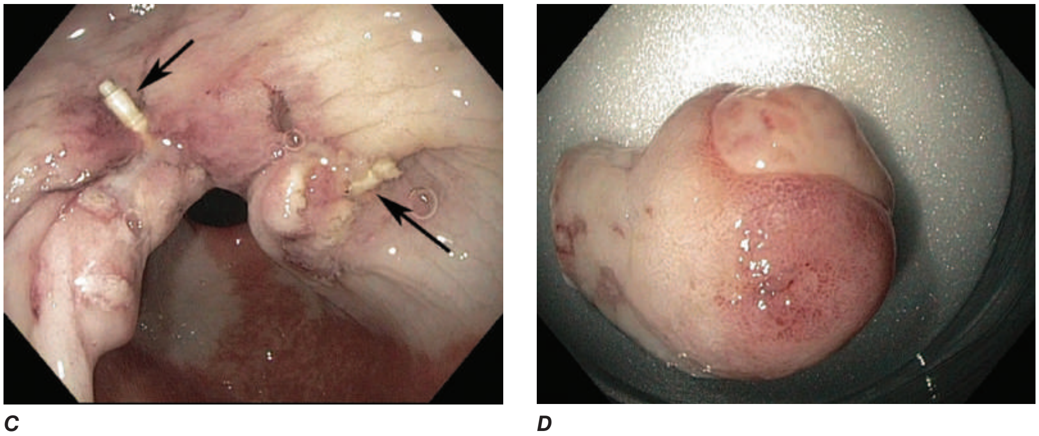

FIGURE 333-27 (*Continued*)

therapy in patients with acute hemorrhage **(Fig. 333-31)**. A clean-based ulcer is associated with a low risk (3–5%) of rebleeding; patients with melena and a clean-based ulcer may be discharged home from the emergency room or endoscopy suite if they are young, reliable, otherwise healthy, and able to return as needed. Flat pigmented spots and adherent clots covering the ulcer base have a 10% and 20% risk of rebleeding, respectively. Flat pigmented spots do not require treatment, but endoscopic therapy is often applied to an ulcer with an adherent clot. When a fibrin plug is seen protruding from a vessel wall in the base of an ulcer (so-called sentinel clot or visible vessel), the risk of rebleeding from the ulcer approximates 40%. This finding typically leads to endoscopic therapy to decrease the rebleeding rate. When active spurting from an ulcer is seen, there is a 90% risk of ongoing bleeding without endoscopic or surgical therapy.

Endoscopic therapy of ulcers with high-risk stigmata typically lowers the rebleeding rate to 5–10%. Several hemostatic techniques are available, including injection of epinephrine or a sclerosant into and around the vessel **(Fig. 333-32)**, "coaptive coagulation" of the vessel in the base of the ulcer using a thermal probe that is pressed against the site of bleeding **(Fig. 333-33)**, placement of through-the-scope clips **(Fig. 333-34)** or an over-the-scope clip **(Fig. 333-35)**, application of a hemostatic powder or gel, or a combination of these modalities **(Video V5-9)**. Epinephrine injection can slow or stop active bleeding, but it is not a stand-alone technique for definitive

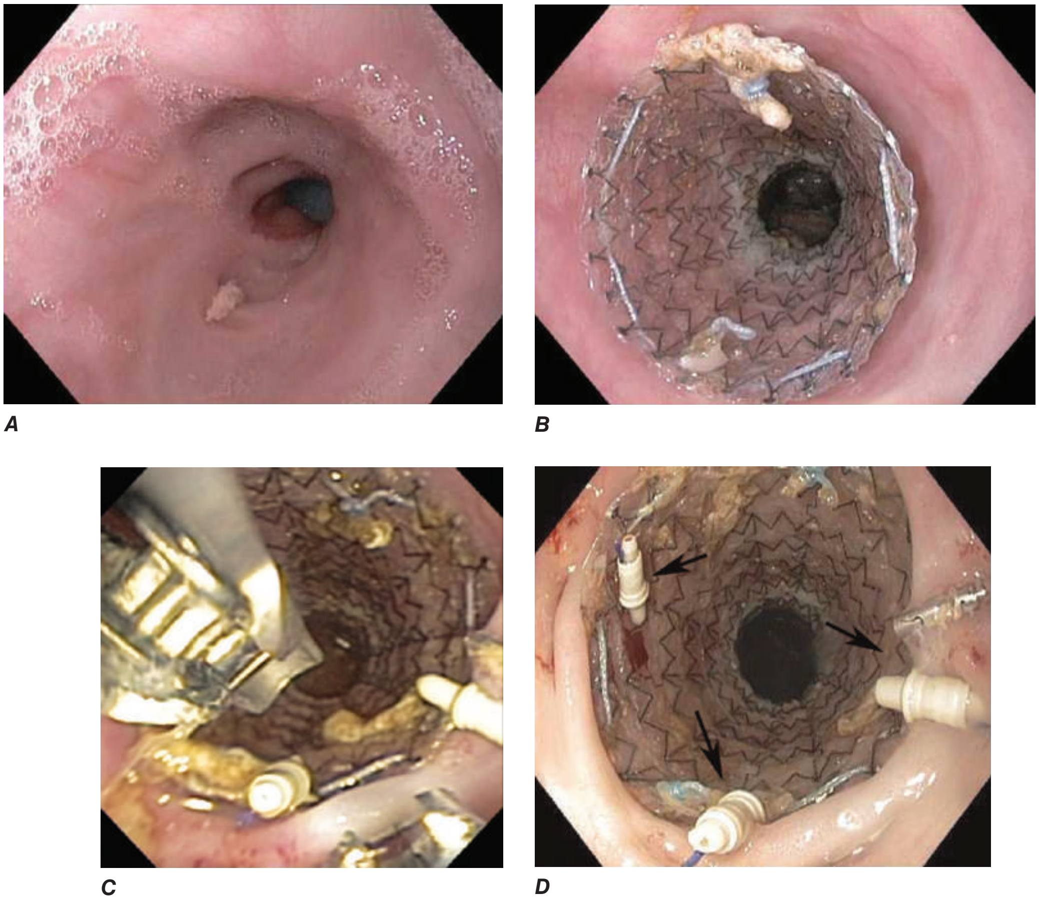

FIGURE 333-28 Prevention of stent migration using endoscopic sutures. ***A.*** Esophagogastric anastomotic stricture refractory to balloon dilation. ***B.*** Temporary placement of a covered esophageal stent. ***C.*** Endoscopic suturing device to anchor the stent to the esophageal wall. ***D.*** Stent fixation with endoscopic sutures (*arrows*).

TABLE 333-1 Antibiotic Prophylaxis for Endoscopic Procedures

PATIENT CONDITION	PROCEDURE CONTEMPLATED	GOAL OF PROPHYLAXIS	PERIPROCEDURAL ANTIBIOTIC PROPHYLAXIS
All cardiac conditions	Any endoscopic procedure	Prevention of infective endocarditis	Not recommended
Bile duct obstruction in the absence of cholangitis	ERCP with complete drainage	Prevention of cholangitis	Not recommended
Bile duct obstruction in the absence of cholangitis	ERCP with anticipated incomplete drainage (e.g., sclerosing cholangitis, hilar strictures)	Prevention of cholangitis	Recommended; continue antibiotics after the procedure
Sterile pancreatic fluid collection (e.g., pseudocyst, necrosis), which communicates with pancreatic duct	ERCP	Prevention of cyst infection	Recommended; continue antibiotics after the procedure
Sterile pancreatic fluid collection	Transmural drainage	Prevention of cyst infection	Recommended
Solid lesion along upper GI tract	EUS-FNA or FNB	Prevention of local infection	Not recommended[a]
Solid lesion along lower GI tract	EUS-FNA or FNB	Prevention of local infection	Not recommended[a]
Cystic lesions along GI tract (including mediastinum and pancreas)	EUS-FNA or FNB	Prevention of cyst infection	Recommended
All patients	Percutaneous endoscopic feeding tube placement	Prevention of peristomal infection	Recommended[b]
Cirrhosis with acute GI bleeding	Recommended for all such patients, regardless of endoscopic procedures	Prevention of infectious complications and reduction of mortality	Recommended, upon admission[c]
Continuous peritoneal dialysis	Lower GI tract endoscopy	Prevention of bacterial peritonitis	Recommended
Synthetic vascular graft and other nonvalvular cardiovascular devices	Any endoscopic procedure	Prevention of graft and device infection	Not recommended[d]
Prosthetic joints	Any endoscopic procedure	Prevention of septic arthritis	Not recommended[d]

[a]Low rates of bacteremia and local infection. [b]Cefazolin or an antibiotic with equivalent coverage of oral and skin flora. [c]Risk for bacterial infection associated with cirrhosis and GI bleeding is well established; ceftriaxone or a quinolone antibiotic recommended. [d]Very low risk of infection.

Abbreviations: ERCP, endoscopic retrograde cholangiopancreatography; EUS-FNA, endoscopic ultrasound–fine-needle aspiration; EUS-FNB, endoscopic ultrasound–fine needle biopsy; GI, gastrointestinal.

Source: Reproduced with permission from MA Kashab et al: Antibiotic prophylaxis for GI endoscopy. Gastrointest Endosc 81:81, 2015.

hemostasis. In conjunction with endoscopic therapy, the administration of a proton pump inhibitor decreases the risk of rebleeding and improves patient outcome.

Varices Two complementary strategies guide therapy of bleeding varices: local treatment of the bleeding varices and treatment of the underlying portal hypertension. Local therapies, including endoscopic variceal band ligation, endoscopic variceal sclerotherapy, stent placement, and tamponade with a stent or Sengstaken-Blakemore tube, effectively control acute hemorrhage in most patients, although therapies that decrease portal pressure (pharmacologic treatment, surgical shunts, or radiologically placed transjugular intrahepatic portosystemic shunts) also play an important role.

Endoscopic variceal ligation (EVL) is indicated for the prevention of a first bleed (primary prophylaxis) from large esophageal varices **(Fig. 333-36)**, particularly in patients in whom nonselective beta blockers are contraindicated or not tolerated. EVL is also the preferred endoscopic therapy for control of active esophageal variceal bleeding and for subsequent eradication of esophageal varices (secondary prophylaxis). During EVL, a varix is suctioned into a cap fitted at the tip of the endoscope, and a rubber band is released from the cap, ligating the varix **(Fig. 333-37, Video V5-10)**. EVL controls acute hemorrhage in up to 90% of patients. Complications of EVL, such as postligation ulcer bleeding and esophageal stenosis, are uncommon. Endoscopic variceal sclerotherapy (EVS) involves the injection of a sclerosing, thrombogenic solution into or next to esophageal varices. EVS also controls acute hemorrhage in most patients, but due to its higher complication rate, it is generally used as salvage therapy when band ligation fails. Bleeding from large gastric fundal varices **(Fig. 333-38)** is best treated with endoscopic cyanoacrylate ("glue") injection **(Video V5-11)** or EUS-guided coil placement and cyanoacrylate injection, since EVL or EVS of these varices is associated with a high rebleeding rate. Complications of cyanoacrylate injection include infection and glue embolization to other organs, such as the lungs, brain, and spleen.

After treatment of the acute hemorrhage, an elective course of endoscopic therapy can be undertaken with the goal of eradicating esophageal varices and preventing rebleeding months to years later. However, this chronic therapy is less successful, preventing long-term rebleeding in ~50% of patients. Pharmacologic therapies that decrease portal pressure have similar efficacy. The preferred strategy, however, for secondary prophylaxis of variceal bleeding is the combination of EVL with a nonselective beta blocker.

Dieulafoy's Lesion This lesion, also called persistent caliber artery, is a large-caliber arteriole that runs immediately beneath the gastrointestinal epithelium and bleeds through a focal mucosal erosion **(Fig. 333-39)**. Dieulafoy's lesion commonly involves the lesser curvature of the proximal stomach, causes impressive arterial hemorrhage, and may be difficult to diagnose when not actively bleeding; it is often recognized only after repeated endoscopy for recurrent bleeding. Endoscopic therapy, such as thermal coagulation, band ligation, clip placement, or endoscopic suturing, is typically effective for control of bleeding and sealing of the underlying vessel once the lesion has been identified **(Video V5-12)**. Rescue therapies, such as angiographic embolization or surgical oversewing, are considered in situations where endoscopic therapy has failed.

Mallory-Weiss Tear A Mallory-Weiss tear is a linear mucosal rent near or across the gastroesophageal junction that is often associated with retching or vomiting **(Fig. 333-40)**. When the tear disrupts a submucosal arteriole, brisk hemorrhage may result. Endoscopy is the best method for diagnosis, and an actively bleeding tear can be treated endoscopically with coaptive coagulation, band ligation, or clip placement, with or without epinephrine injection **(Video V5-13)**. Unlike peptic ulcer, a Mallory-Weiss tear with a nonbleeding sentinel clot in its base rarely rebleeds and thus does not necessitate endoscopic therapy.

Vascular Ectasias Vascular ectasias are flat mucosal vascular anomalies that are best diagnosed by endoscopy. They usually cause slow intestinal blood loss and occur either in a sporadic fashion or in a well-defined pattern of distribution (e.g., gastric antral vascular ectasia [GAVE] or "watermelon stomach") **(Fig. 333-41)**. Cecal vascular ectasias, GAVE, and radiation-induced rectal ectasias are often responsive

TABLE 333-2 Management of Antithrombotic Drugs Prior to Endoscopic Procedures

DRUG	BLEEDING RISK OF PROCEDURE	MANAGEMENT	INTERVAL BETWEEN LAST DOSE AND PROCEDURE	COMMENTS
Warfarin	Low[a]	Continue	N/A	Ensure that INR is not supratherapeutic
	High[b]	Discontinue	3–7 days (usually 5), INR should be ≤1.5 for procedure	Consider bridging therapy with low-molecular-weight heparin for patients at high risk of thrombosis[c]; usually safe to resume warfarin on the same or next day For life-threatening GI hemorrhage, consider reversal with unactivated prothrombin complex concentrate
Dabigatran, rivaroxaban, apixaban, edoxaban	Low[a]	Hold morning dose on day of procedure	N/A	
Dabigatran	High[b]	Discontinue	2–3 days if GFR is ≥50 mL/min, 4–5 days if GFR is 30–49 mL/min	Bridging therapy not recommended; resume drug when bleeding risk is low For life-threatening GI hemorrhage, consider use of a reversal agent
Rivaroxaban, apixaban, edoxaban	High[a]	Discontinue	2 days if GFR is ≥60 mL/min, 3 days if GFR is 30–59 mL/min, 4 days if GFR is <30 mL/min	Bridging therapy not recommended; resume drug when bleeding risk is low For life-threatening GI hemorrhage, consider use of a reversal agent
Heparin	Low[a]	Continue	N/A	
	High[b]	Discontinue	4–6 h for unfractionated heparin	Skip one dose if using low-molecular-weight heparin
Aspirin	Any	Continue	N/A	Low-dose aspirin does not substantially increase the risk of endoscopic procedures
Aspirin with dipyridamole	Low[a]	Continue	N/A	
	High[b]	Discontinue	2–7 days	Consider continuing aspirin monotherapy
$P2Y_{12}$ receptor antagonists (clopidogrel, prasugrel, ticlopidine, ticagrelor, cangrelor)	Low[a]	Continue	N/A	
	High[b]	Coronary stent in place: discuss with cardiologist No coronary stent: discontinue, consider substituting aspirin	5 days (clopidogrel or ticagrelor), 7 days (prasugrel), 10–14 days (ticlopidine)	Risk of stent thrombosis for at least 12 months after insertion of drug-eluting coronary stent or 1 month after insertion of bare metal coronary stent

[a] Low-risk endoscopic procedures include esophagogastroduodenoscopy (EGD) or colonoscopy with or without biopsy, endoscopic ultrasound (EUS) without fine-needle aspiration (FNA), and endoscopic retrograde cholangiopancreatography (ERCP) with stent exchange. [b] High-risk endoscopic procedures include EGD or colonoscopy with dilation, polypectomy, or thermal ablation; percutaneous endoscopic gastrostomy (PEG); EUS with FNA; and ERCP with sphincterotomy or pseudocyst drainage. [c] Bridging therapy with low-molecular-weight heparin should be considered for patients discontinuing warfarin who are at high risk for thromboembolism, including those with (1) prosthetic metal heart valve, (2) atrial fibrillation with a CHA_2DS_2-VASc score ≥3, mitral stenosis, prosthetic valve or history of stroke or transient ischemic attack; (3) mechanical mitral valve; (4) mechanical aortic valve with other thromboembolic risk factors or older-generation mechanical aortic valve; or (5) venous thromboembolism (VTE) within the past 3 months.

Abbreviations: GFR, glomerular filtration rate; GI, gastrointestinal; INR, international normalized ratio; N/A, not applicable.

Source: Adapted from RD Acosta et al: Gastrointest Endosc 83:3, 2016; and AM Veitch et al: Gut 70:1611, 2021.

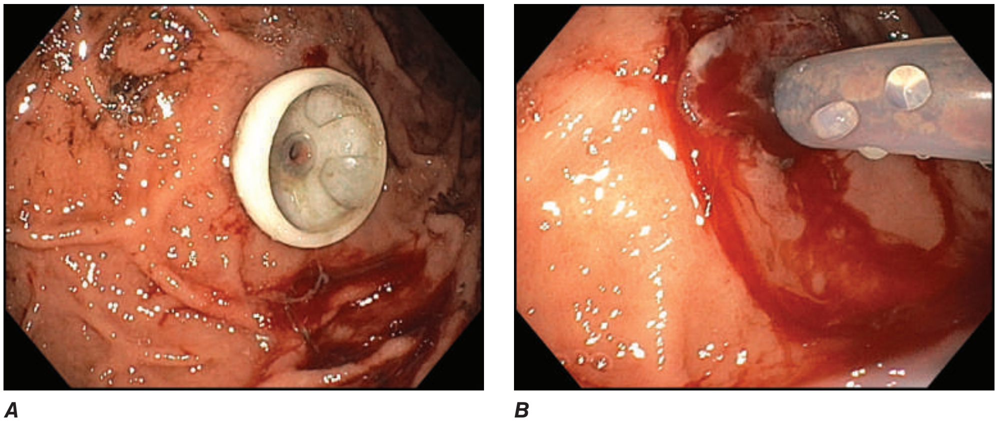

FIGURE 333-29 Bleeding from percutaneous endoscopic gastrostomy (PEG) tube placement. ***A.*** Patient with melena from a recently placed PEG tube. ***B.*** Loosening of the internal disk bumper of the PEG tube revealed active bleeding from within the PEG tract.

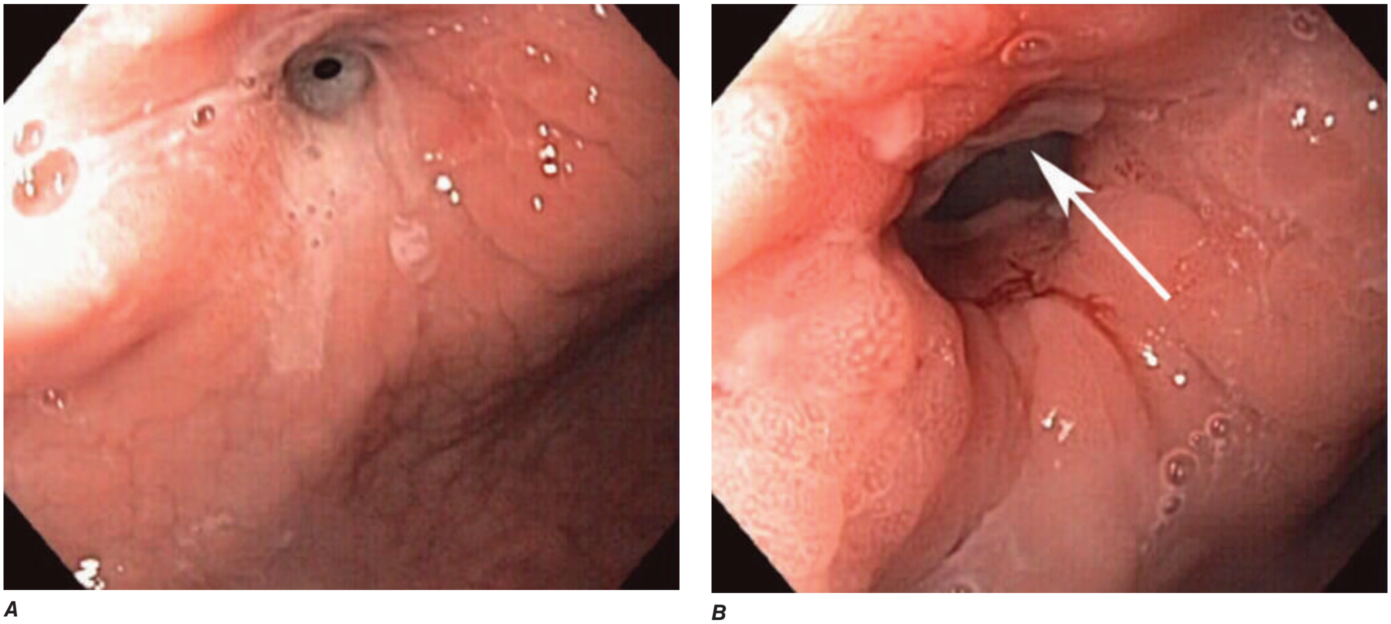

FIGURE 333-30 Buried bumper syndrome. ***A.*** Migration of the internal disk bumper of a percutaneous endoscopic gastrostomy (PEG) tube through the gastric wall. ***B.*** Close-up view of the disk bumper (*arrow*) buried in the gastric wall.

to local endoscopic ablative therapy, such as argon plasma coagulation (Video V5-14). Patients with diffuse small-bowel vascular ectasias (associated with chronic renal failure and with hereditary hemorrhagic telangiectasia) may continue to bleed despite endoscopic treatment of easily accessible lesions by conventional endoscopy. These patients may benefit from deep enteroscopy with endoscopic hemostasis, or pharmacologic therapy, such as octreotide or low-dose thalidomide, in those who continue to bleed despite endoscopic therapy.

Colonic Diverticula Diverticula form where nutrient arteries penetrate the muscular wall of the colon en route to the colonic mucosa (Fig. 333-42). The artery found in the base of a diverticulum may bleed, causing painless and impressive hematochezia. Colonoscopy is indicated in patients with hematochezia and suspected diverticular hemorrhage, since other causes of bleeding (e.g., vascular ectasias, colitis, and colon cancer) must be excluded. In addition, an actively bleeding diverticulum may be seen and treated during colonoscopy (Fig. 333-43, Video V5-15).

GASTROINTESTINAL OBSTRUCTION AND PSEUDOOBSTRUCTION

Endoscopy is useful for evaluation and treatment of some forms of gastrointestinal obstruction. An important exception is small-bowel obstruction due to surgical adhesions, which is generally not diagnosed or treated endoscopically. Esophageal, gastroduodenal, and colonic obstruction or pseudoobstruction can all be diagnosed and often managed endoscopically.

Acute Esophageal Obstruction Esophageal obstruction by impacted food (Fig. 333-44) or an ingested foreign body (Fig. 333-45) is a potentially life-threatening event and represents an endoscopic emergency. Left untreated, the patient may develop esophageal ulceration,

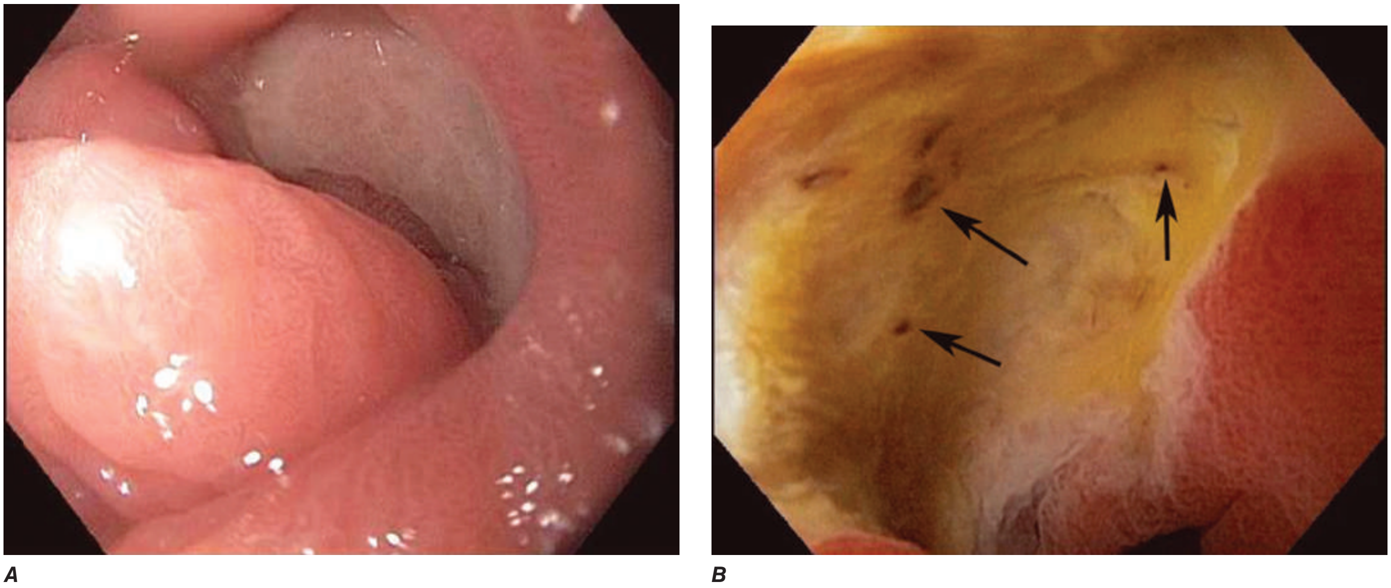

FIGURE 333-31 Stigmata of hemorrhage in peptic ulcers. ***A.*** Gastric antral ulcer with a clean base. ***B.*** Duodenal ulcer with flat pigmented spots (*arrows*). ***C.*** Duodenal ulcer with a dense adherent clot. ***D.*** Duodenal ulcer with a pigmented protuberance/visible vessel (*arrow*). ***E.*** Duodenal ulcer with active spurting (*arrow*).

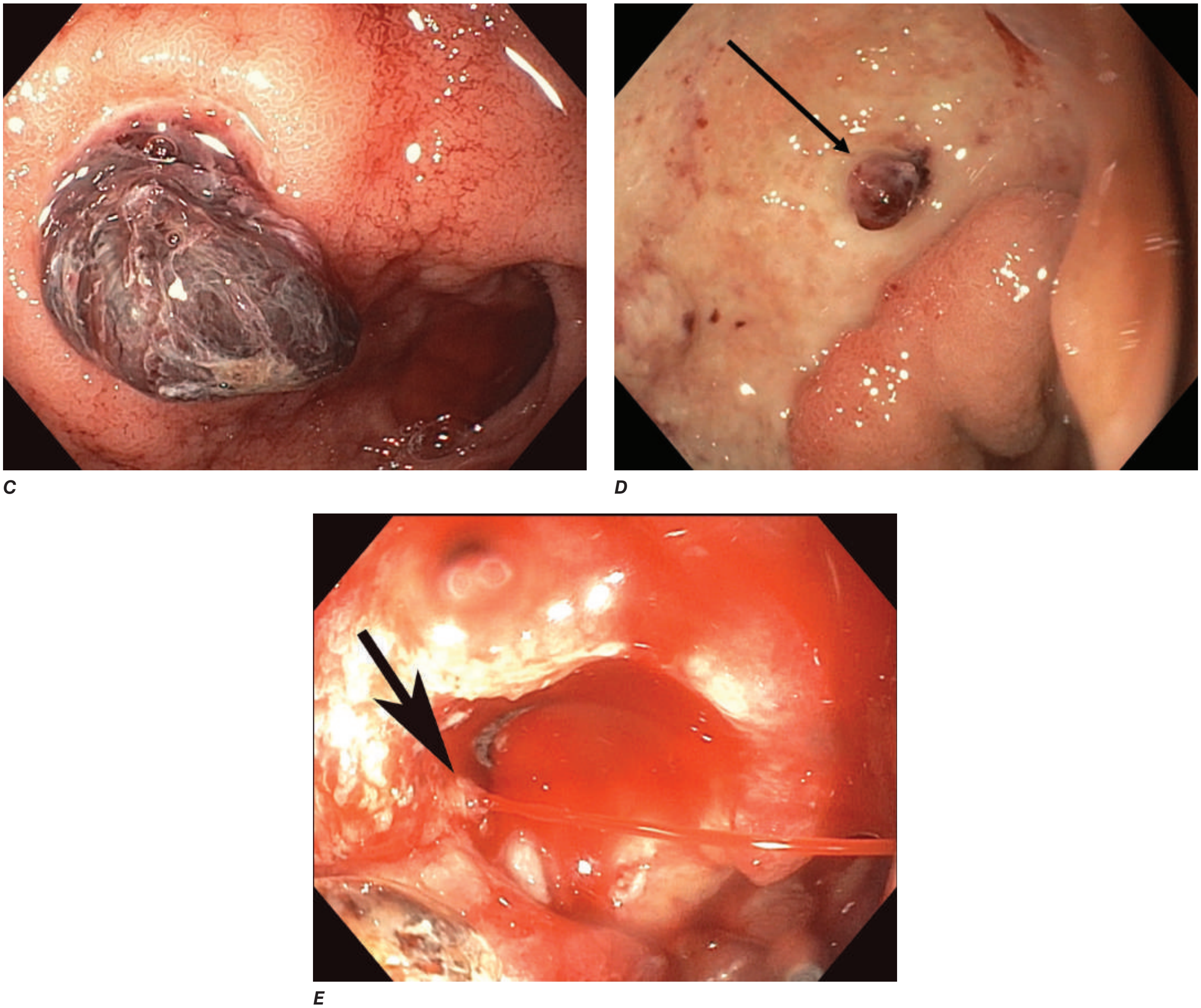

FIGURE 333-31 (*Continued*)

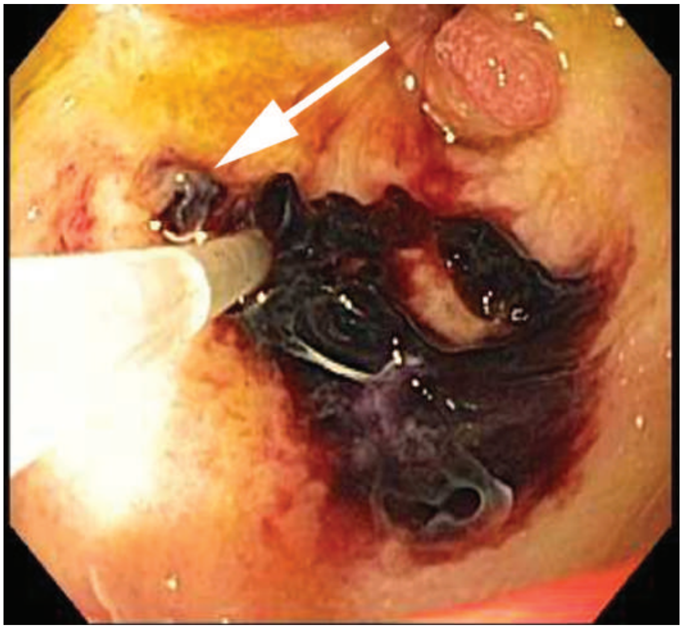

FIGURE 333-32 Injection therapy for ulcer hemostasis. Epinephrine injection into a duodenal ulcer with visible vessel (*arrow*) and adherent clot.

ischemia, and perforation. Patients with persistent esophageal obstruction often have hypersalivation and are usually unable to swallow water. Sips of a carbonated beverage, sublingual nifedipine or nitrates, or intravenous glucagon may resolve an esophageal food impaction, but in many patients, an underlying web, ring, or stricture is present, and endoscopic removal of the obstructing food bolus is necessary. Endoscopy is generally the best initial test in such patients since endoscopic removal of the obstructing material is usually possible, and the presence of an underlying esophageal pathology can often be determined. Radiographs of the chest and neck should be considered before endoscopy in patients with fever, obstruction for ≥24 h, or ingestion of a sharp object, such as a fishbone. Radiographic contrast studies interfere with subsequent endoscopy and are not advisable in most patients with a clinical picture of esophageal obstruction.

Gastric Outlet Obstruction Obstruction of the gastric outlet is commonly caused by gastric, duodenal, or pancreatic malignancy or chronic peptic ulceration with stenosis of the pylorus (**Fig. 333-46**). Patients vomit partially digested food many hours after eating. Gastric decompression with a nasogastric tube and subsequent lavage for removal of retained material is the first step in treatment. Endoscopy is useful for diagnosis and treatment. Patients with benign pyloric stenosis may be treated with endoscopic balloon dilation of the pylorus, and a course of endoscopic dilation results in long-term relief of symptoms

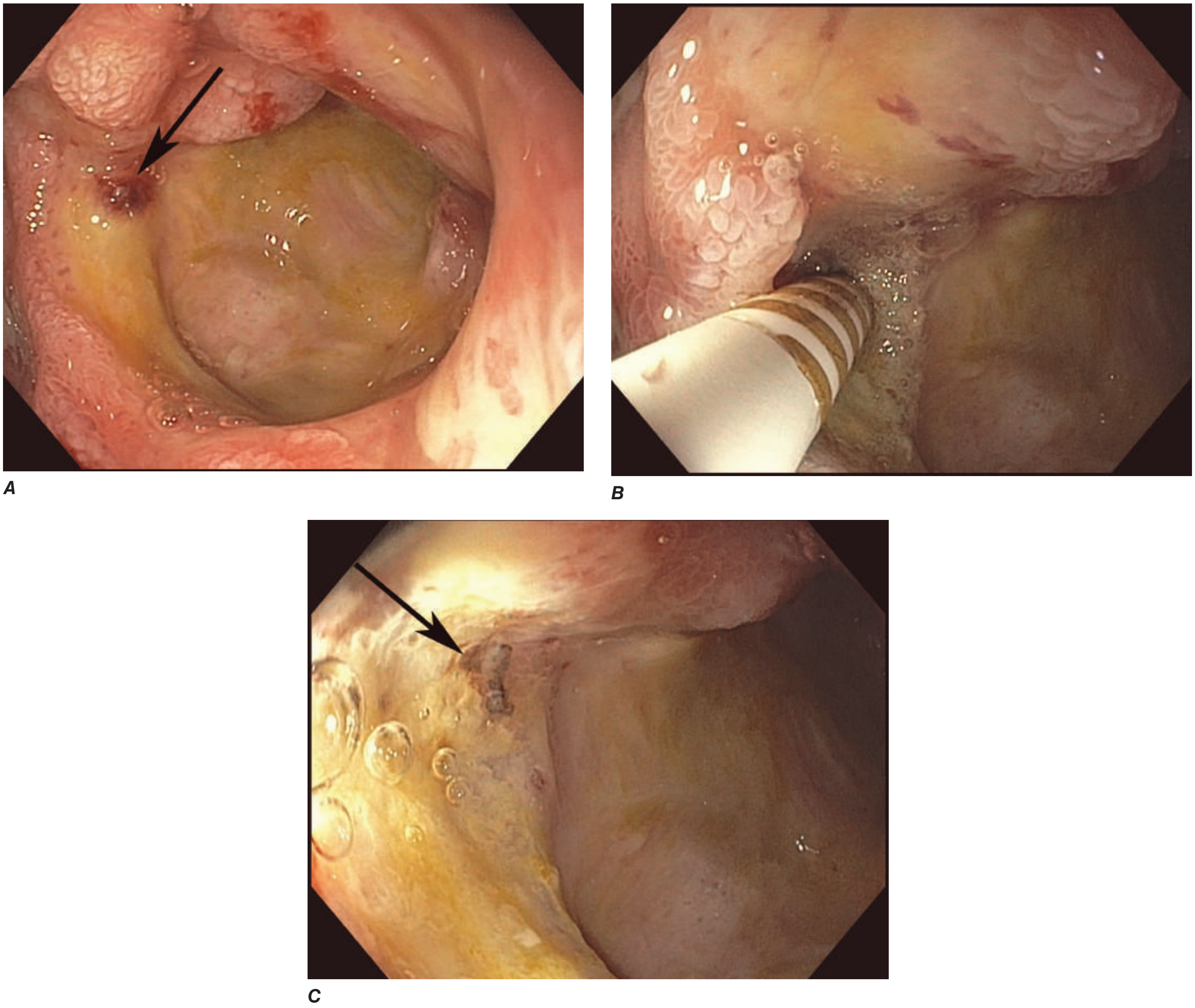

FIGURE 333-33 Contact coagulation for ulcer hemostasis. ***A.*** Duodenal ulcer with a visible vessel (*arrow*). ***B.*** Coagulation of the vessel with a contact thermal probe. ***C.*** Obliteration of the treated vessel (*arrow*).

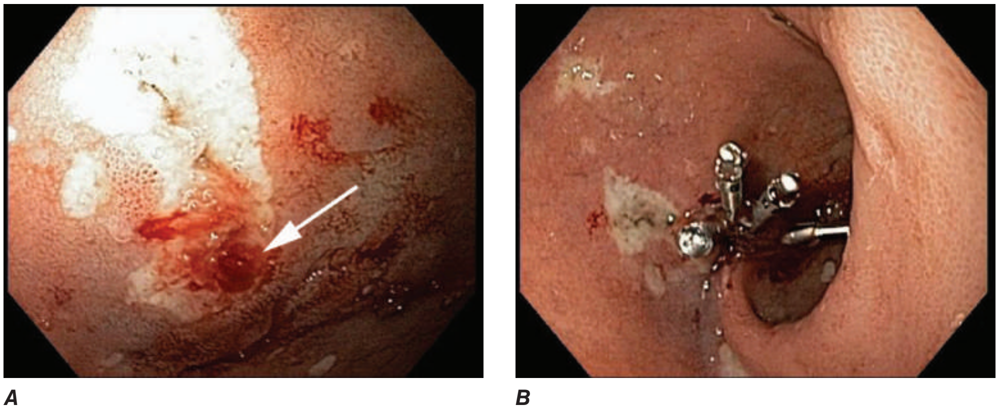

FIGURE 333-34 Through-the-scope clip placement for ulcer hemostasis. ***A.*** Superficial duodenal ulcer with visible vessel (*arrow*). ***B.*** Hemostasis secured following placement of multiple through-the-scope clips.

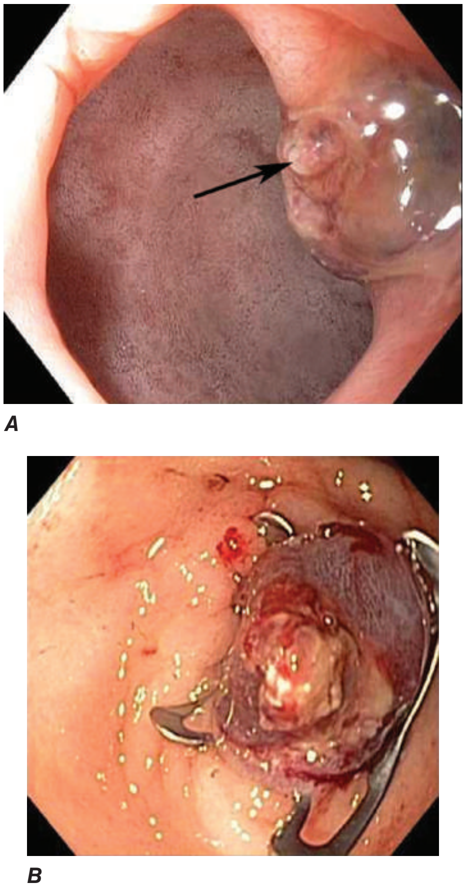

FIGURE 333-35 Over-the-scope clip placement for ulcer hemostasis. ***A.*** Pyloric channel ulcer with visible vessel (*arrow*). ***B.*** Hemostasis secured following placement of an over-the-scope clip.

in ~50% of patients. Removable, fully covered lumen-apposing metal stents (LAMS) may also be used to treat benign pyloric stenosis **(Video V5-16)**. Malignant gastric outlet obstruction can be relieved with endoscopically placed expandable stents across the obstruction in patients with inoperable malignancy **(Video V5-17)** or by EUS-guided gastroenterostomy to bypass the obstruction **(Video V5-3)**.

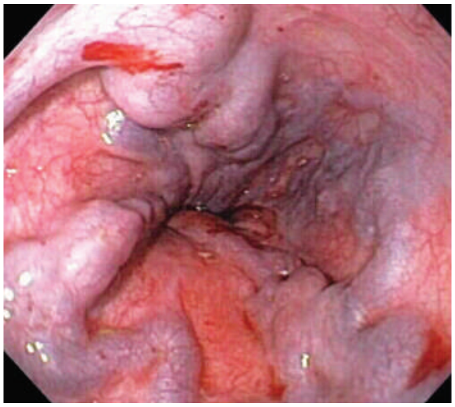

FIGURE 333-36 Esophageal varices.

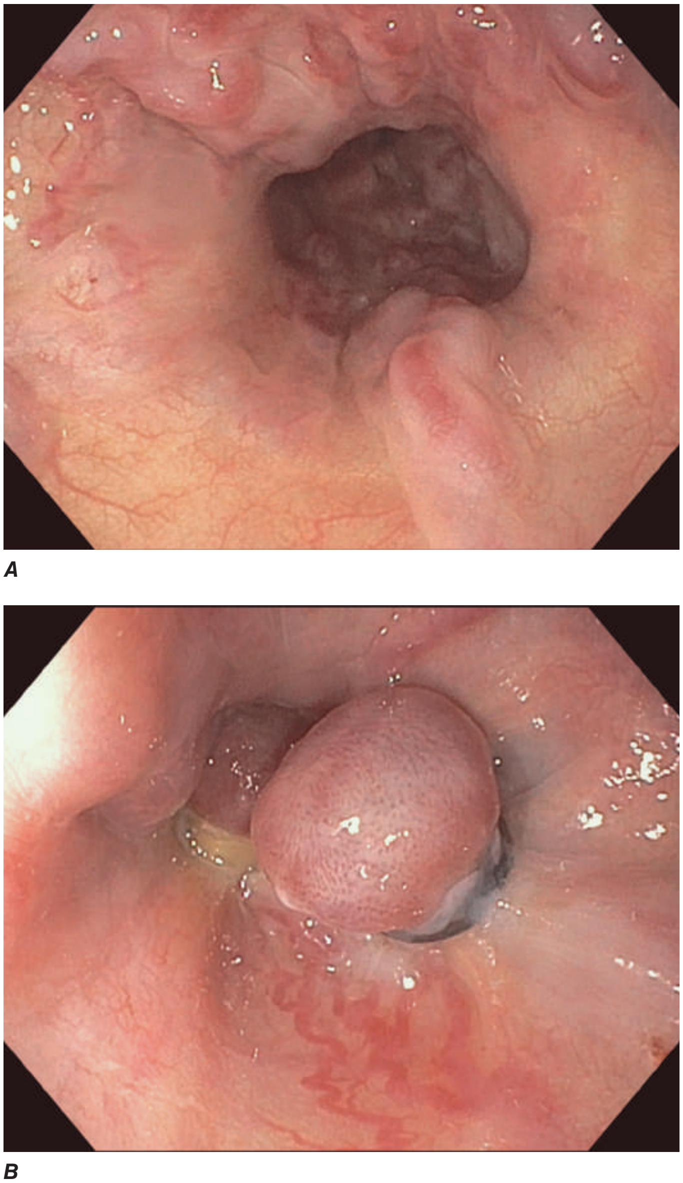

FIGURE 333-37 Endoscopic variceal ligation. ***A.*** Esophageal varices with red wale marks. ***B.*** Band ligation of varices.

Colonic Obstruction and Pseudoobstruction

These conditions both present with abdominal distention and discomfort, tympany, and a dilated colon on plain abdominal radiography. The radiographic appearance may be characteristic of a particular condition, such as sigmoid volvulus **(Fig. 333-47)**. Both obstruction and pseudoobstruction may lead to colonic perforation if left untreated. Acute colonic pseudoobstruction is a form of colonic ileus that is usually attributable to electrolyte disorders, narcotic and anticholinergic medications, immobility (as after surgery), or retroperitoneal hemorrhage or mass. Multiple causative factors are often present. Colonoscopy, water-soluble contrast enema, or CT may be used to assess for an obstructing lesion and differentiate obstruction from pseudoobstruction. One of these diagnostic studies should be strongly considered if the patient does not have clear risk factors for pseudoobstruction, if radiographs do not show air in the rectum, or if the patient fails to improve when underlying causes of pseudoobstruction have been addressed. The risk of cecal perforation in pseudoobstruction rises when the cecal diameter exceeds 12 cm, and decompression of the colon may be achieved using intravenous neostigmine or via colonoscopic decompression **(Fig. 333-48)**. Most patients should receive a trial of conservative therapy (with correction of electrolyte disorders, discontinuation of offending medications, and increased mobilization) before undergoing an invasive decompressive procedure for colonic pseudoobstruction.

Acute colonic obstruction is an indication for urgent intervention. In the past, emergent diverting colostomy was usually performed with a subsequent second operation after bowel preparation to treat the underlying cause of obstruction. Colonoscopic placement of an

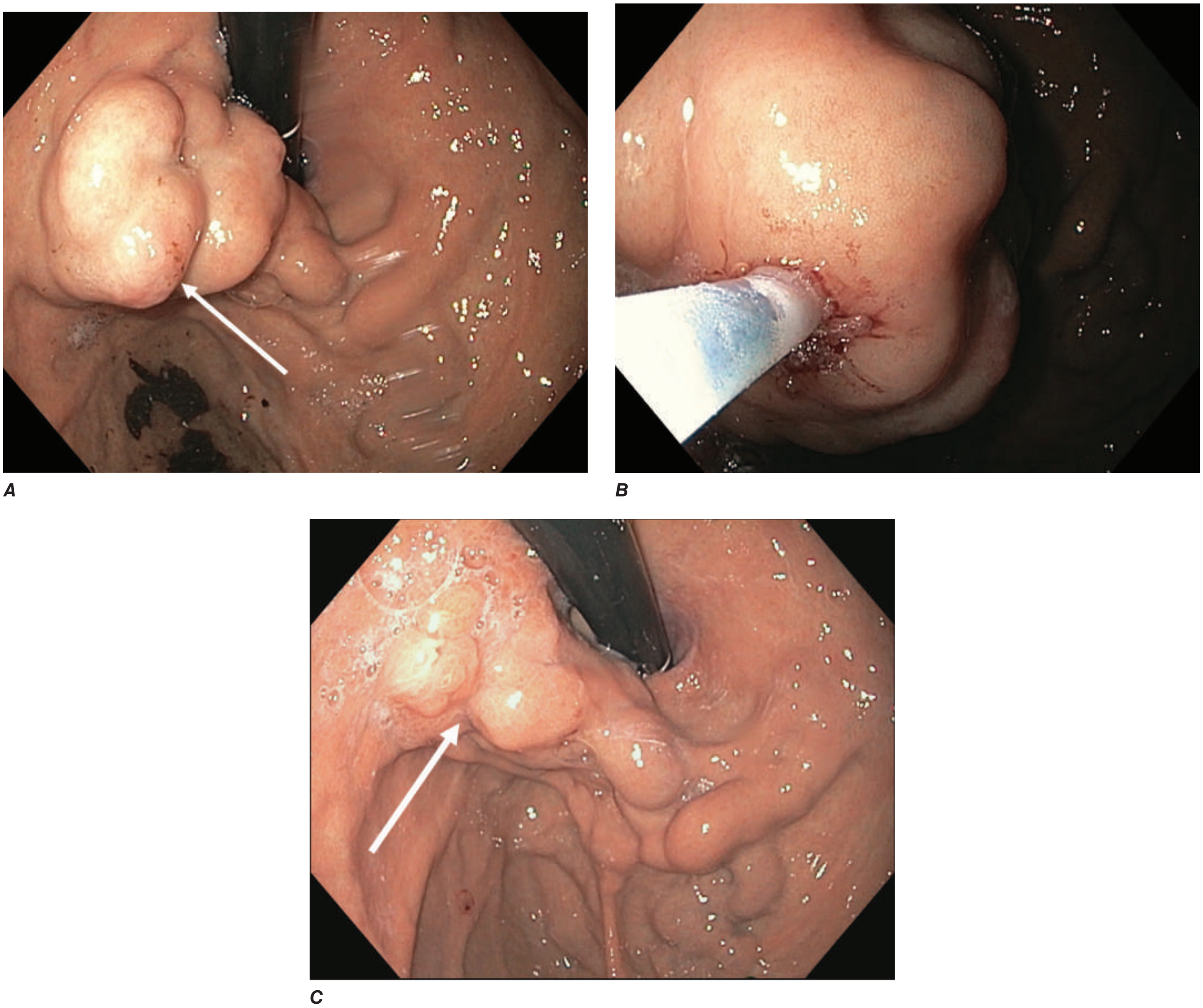

FIGURE 333-38 Gastric varices. ***A.*** Large gastric fundal varices with stigmata of recent bleeding (*arrow*). ***B.*** Injection of cyanoacrylate (glue) into the culprit gastric varix. ***C.*** Obliterated varix following glue injection on endoscopic follow-up at 1 month (*arrow*).

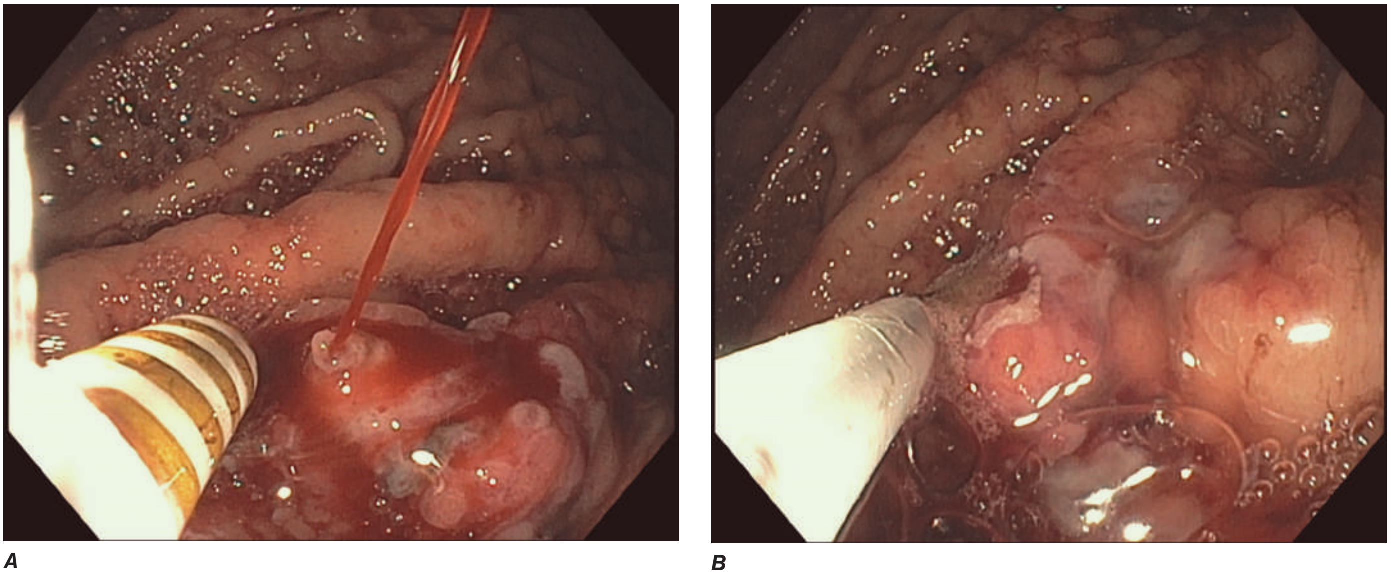

FIGURE 333-39 Dieulafoy's lesion. ***A.*** Actively spurting gastric Dieulafoy's lesion. ***B.*** Coagulation of the lesion using a contact thermal probe. ***C.*** Hemostasis secured following contact coagulation (*arrow*). ***D.*** Histology of a gastric Dieulafoy's lesion. A persistent caliber artery (*arrows*) is present in the gastric submucosa, immediately beneath the mucosa.

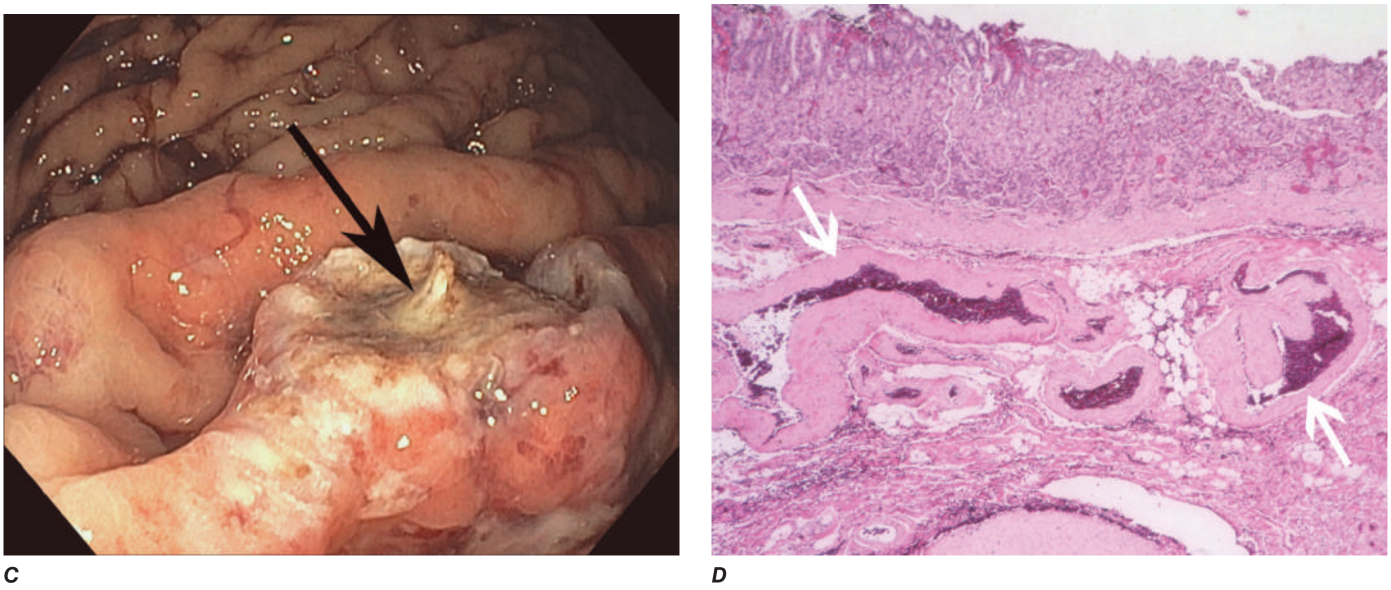

FIGURE 333-39 (*Continued*)

expandable stent is an alternative treatment option that can relieve malignant colonic obstruction without emergency surgery and permit bowel preparation for an elective one-stage operation **(Fig. 333-49, Video V5-18)**.

■ ACUTE BILIARY OBSTRUCTION

The steady, severe pain that occurs when a gallstone acutely obstructs the common bile duct often brings patients to seek medical attention. The diagnosis of a ductal stone is suspected when the patient is jaundiced or when serum liver tests or pancreatic enzyme levels are elevated; it is confirmed by transabdominal ultrasound, EUS, magnetic resonance cholangiopancreatography (MRCP), or direct cholangiography (performed endoscopically, percutaneously, or during surgery). ERCP is the primary means of treating common bile duct stones (Figs. 333-16 and 333-17), although they can also be removed by bile duct exploration at the time of cholecystectomy. Radiologic percutaneous biliary drainage may be required in some cases.

Bile Duct Imaging While transabdominal ultrasound diagnoses only a minority of bile duct stones, MRCP and EUS are >90% accurate and have an important role in diagnosis. Examples of these modalities are shown in **Fig. 333-50**.

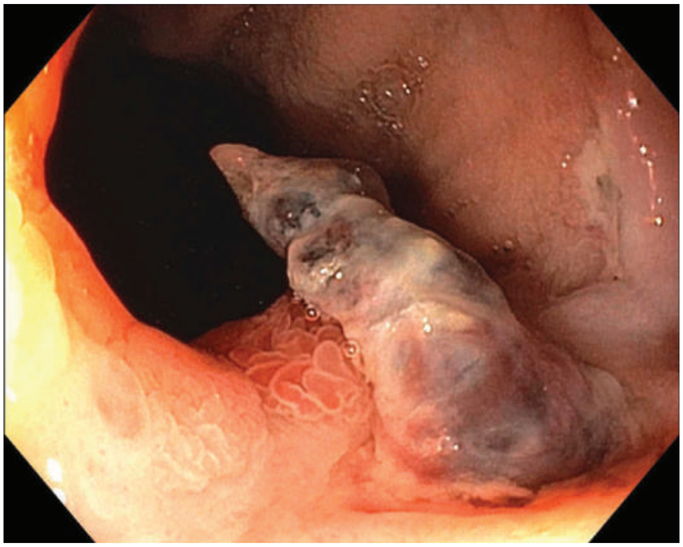

FIGURE 333-40 Mallory-Weiss tear with an adherent clot at the gastroesophageal junction following forceful retching and vomiting.

If the suspicion for a bile duct stone is high and urgent treatment is required (as in a patient with obstructive jaundice and biliary sepsis), ERCP is the procedure of choice since it remains the gold standard for diagnosis and allows for immediate treatment **(Video V5-19)**. If a persistent bile duct stone is relatively unlikely (as in a patient with gallstone pancreatitis), ERCP may be supplanted by less invasive imaging techniques, such as EUS, MRCP, or intraoperative cholangiography performed during cholecystectomy, sparing some patients the risk and discomfort of ERCP.

Ascending Cholangitis Charcot's triad of jaundice, abdominal pain, and fever is present in ~70% of patients with ascending cholangitis and biliary sepsis. These patients are managed initially with fluid resuscitation and intravenous antibiotics. Abdominal ultrasound is often performed to assess for gallbladder stones and bile duct dilation. However, the bile duct may not be dilated early in the course of acute biliary obstruction. Medical management usually improves the patient's clinical status, providing a window of ~24 h during which biliary drainage should be established, typically by ERCP. Undue delay can result in recrudescence of overt sepsis and increased morbidity and mortality rates. In addition to Charcot's triad, the presence of shock and confusion (Reynolds's pentad) is associated with a high mortality rate and should prompt urgent intervention to restore biliary drainage.

Gallstone Pancreatitis Gallstones may cause acute pancreatitis as they pass through the ampulla of Vater. The occurrence of gallstone pancreatitis usually implies passage of a stone into the duodenum, and only ~20% of patients harbor a persistent stone in the ampulla or the common bile duct. Retained stones are more common in patients with jaundice, rising serum liver tests following hospitalization, severe pancreatitis, or superimposed ascending cholangitis.

Urgent ERCP decreases the morbidity rate of gallstone pancreatitis in a subset of patients with retained bile duct stones. It is unclear whether the benefit of ERCP is mainly attributable to treatment and prevention of ascending cholangitis or to relief of pancreatic ductal obstruction. ERCP is warranted early in the course of gallstone pancreatitis if ascending cholangitis is suspected, especially in a jaundiced patient. Urgent ERCP may also benefit patients predicted to have severe pancreatitis using a clinical index of severity, such as the Glasgow or Ranson score. Since the benefit of ERCP is limited to patients with a retained bile duct stone, a strategy of initial MRCP or EUS for diagnosis decreases the utilization of ERCP in gallstone pancreatitis and improves clinical outcomes by limiting the occurrence of ERCP-related adverse events.

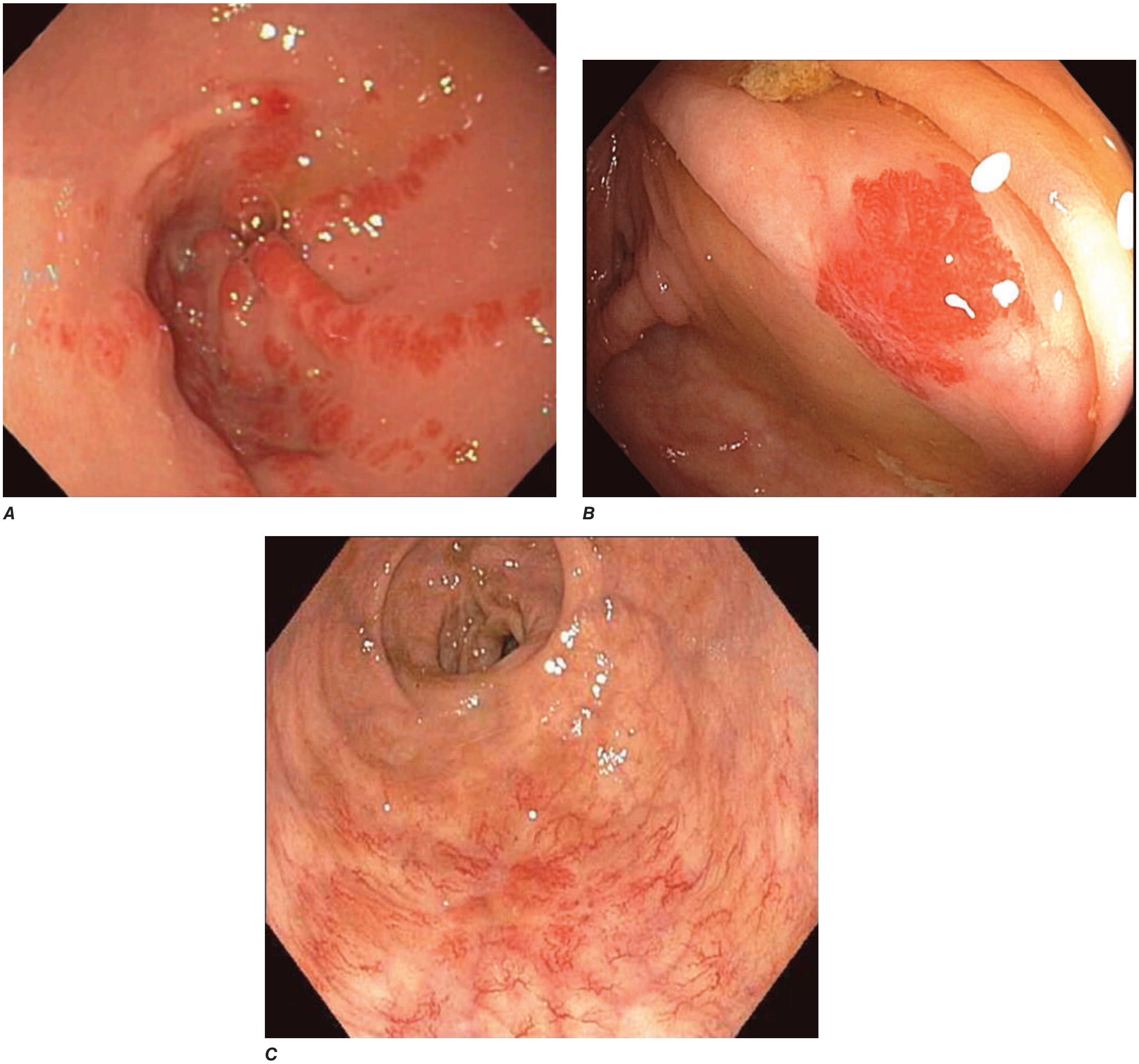

FIGURE 333-41 Gastrointestinal vascular ectasias. ***A.*** Gastric antral vascular ectasia ("watermelon stomach") characterized by stripes of prominent flat or raised vascular ectasias. ***B.*** Cecal vascular ectasia. ***C.*** Radiation-induced vascular ectasias of the rectum in a patient previously treated for prostate cancer.

ELECTIVE ENDOSCOPY

■ DYSPEPSIA

Dyspepsia is a chronic or recurrent burning discomfort or pain in the upper abdomen that may be caused by diverse processes, such as gastroesophageal reflux, peptic ulcer disease, and "nonulcer dyspepsia," a heterogeneous category that includes disorders of motility, sensation, and somatization. Gastric and esophageal malignancies are less common causes of dyspepsia. Careful history-taking allows accurate differential diagnosis of dyspepsia in only about half of patients. In the remainder, endoscopy can be a useful diagnostic tool, especially in patients whose symptoms are not resolved by *Helicobacter pylori* treatment or an empirical trial of acid-reducing therapy. Endoscopy should be performed at the outset in patients with dyspepsia and alarm features, such as weight loss, obstructive symptoms, or iron-deficiency anemia.

■ GASTROESOPHAGEAL REFLUX DISEASE

When classic symptoms of gastroesophageal reflux are present, such as water brash and substernal heartburn, presumptive diagnosis and empirical treatment are often sufficient. Endoscopy is a sensitive test for diagnosis of esophagitis **(Fig. 333-51)**, but it will miss nonerosive reflux disease (NERD) since some patients have symptomatic reflux without esophagitis. The most sensitive test for diagnosis of gastroesophageal reflux disease (GERD) is 24-h ambulatory pH and impedance monitoring. Endoscopy is indicated in patients with reflux symptoms refractory to antisecretory therapy; in those with alarm symptoms, such as dysphagia, weight loss, or gastrointestinal bleeding; and in those with recurrent dyspepsia after treatment that is not clearly due to reflux on clinical grounds alone. Endoscopy should be considered in patients with long-standing GERD, as they have a sixfold increased risk of harboring Barrett's esophagus compared to patients with <1 year of reflux symptoms.

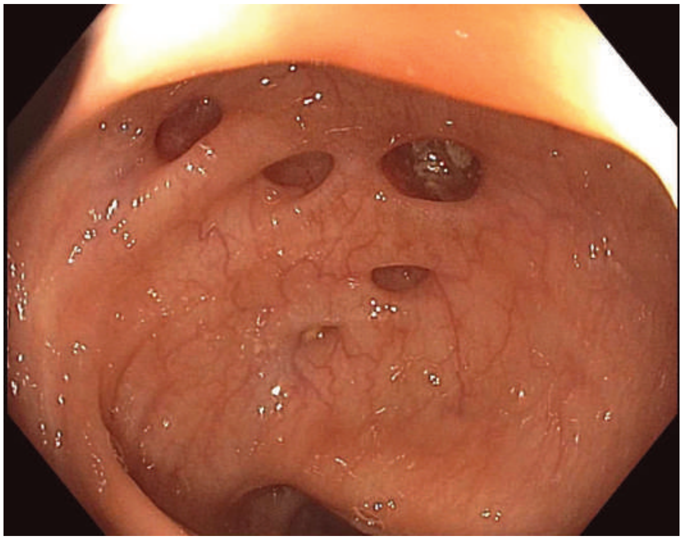

FIGURE 333-42 Colonic diverticula.

Barrett's Esophagus and Esophageal Squamous Dysplasia

Barrett's esophagus is specialized columnar metaplasia that replaces the normal squamous mucosa of the distal esophagus in some persons with GERD. Barrett's epithelium is a major risk factor for adenocarcinoma of the esophagus and is readily detected endoscopically, due to proximal displacement of the squamocolumnar junction (Fig. 333-6). A screening EGD for Barrett's esophagus should be considered in patients with a chronic (≥10 year) history of GERD symptoms, even if their symptoms have been mild. Endoscopic biopsy is the gold standard for confirmation of Barrett's esophagus and for dysplasia or cancer arising in Barrett's mucosa.

Periodic EGD with biopsies is recommended for surveillance of patients with Barrett's esophagus. Endoscopic resection (EMR or ESD) and/or ablation are treatment options when high-grade dysplasia or intramucosal cancer are found in the Barrett's mucosa. Both endoscopic therapy and periodic surveillance are acceptable options in patients with Barrett's esophagus and low-grade dysplasia. Radiofrequency ablation (RFA) is the most common ablative modality used for endoscopic treatment of Barrett's esophagus, and other modalities, such as cryotherapy, are also available.

Esophageal squamous dysplasia is the precursor lesion of esophageal squamous cell cancer (ESCC), the most common type of esophageal malignancy worldwide. Endoscopic detection of esophageal squamous dysplasia often requires specialized imaging methods, such as chromoendoscopy with Lugol's iodine solution. Once detected, it can be treated endoscopically with EMR, ESD, or RFA **(Fig. 333-52)**. Population-based screening for esophageal squamous dysplasia has been shown to decrease the occurrence of ESCC in high-incidence regions.

■ PEPTIC ULCER

Peptic ulcer classically causes epigastric gnawing or burning, often occurring nocturnally and promptly relieved by food or antacids. Although endoscopy is the most sensitive diagnostic test for peptic ulcer, it is not a cost-effective strategy in young patients with ulcer-like dyspeptic symptoms unless endoscopy is available at low cost. Patients with suspected peptic ulcer should be evaluated for *H. pylori* infection. Serology (which diagnoses past or current infection), urea breath testing (current infection), and stool tests (current infection) are noninvasive and less costly than endoscopy with biopsy. Patients aged >50 and those with alarm symptoms or persistent symptoms despite treatment should undergo endoscopy to exclude malignancy.

■ NONULCER DYSPEPSIA

Nonulcer dyspepsia may be associated with bloating and, unlike peptic ulcer, tends not to remit and recur. Most patients describe persistent symptoms despite acid-reducing, prokinetic, or anti-*Helicobacter*

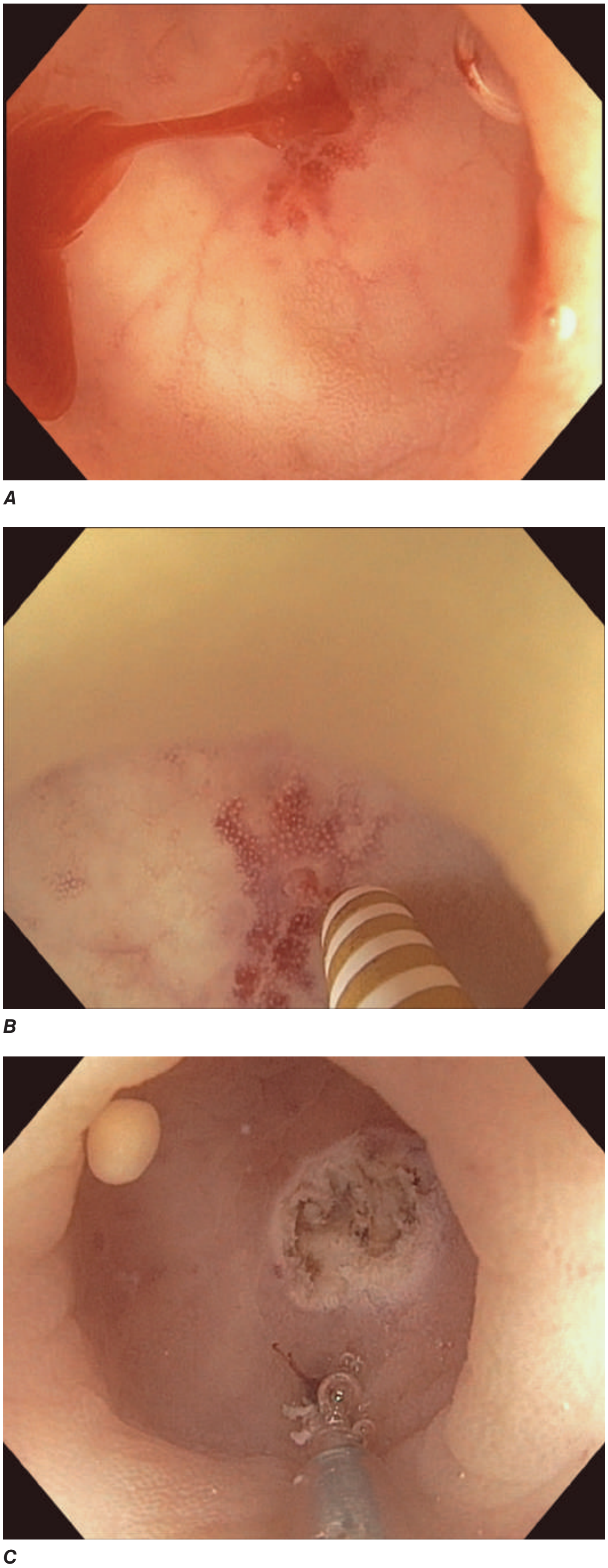

FIGURE 333-43 Diverticular hemorrhage. ***A.*** Actively bleeding sigmoid diverticulum. ***B.*** Treatment of the bleeding vessel at the dome of the diverticulum with a contact thermal probe. ***C.*** Hemostasis secured following contact coagulation with tattoo injection to aid future localization.

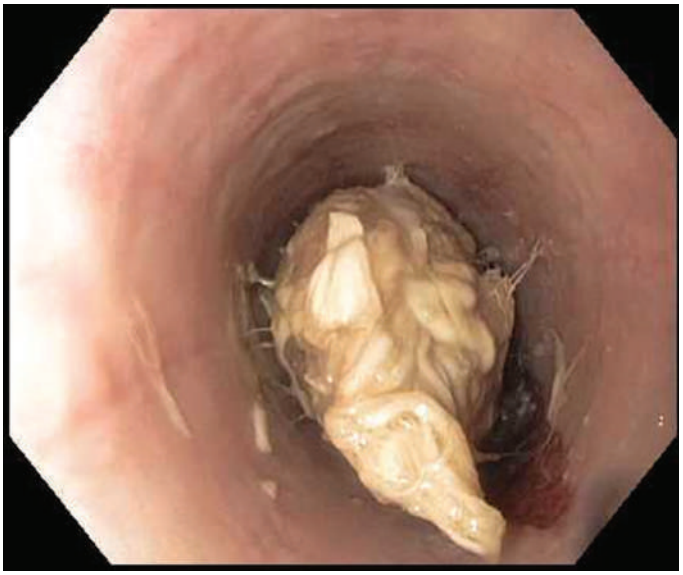

FIGURE 333-44 Esophageal food impaction. Meat bolus impacted in the distal esophagus.

therapy and are referred for endoscopy to exclude a refractory ulcer and assess for other causes. Although endoscopy is useful for excluding other diagnoses, its impact on the treatment of patients with nonulcer dyspepsia is limited.

DYSPHAGIA

About 50% of patients presenting with difficulty swallowing have a mechanical obstruction; the remainder have a motility disorder, such as achalasia or diffuse esophageal spasm, or an inflammatory disorder, such as eosinophilic esophagitis. Careful history-taking often points to a presumptive diagnosis and leads to the appropriate use of diagnostic tests. Esophageal strictures **(Fig. 333-53)** typically cause progressive dysphagia, first for solids, then for liquids; motility disorders often cause intermittent dysphagia for both solids and liquids. Some underlying disorders have characteristic historic features: Schatzki's ring **(Fig. 333-54)** causes episodic dysphagia for solids, typically at the beginning of a meal; oropharyngeal motor disorders typically present with difficulty initiating deglutition (*transfer dysphagia*) and nasal reflux or coughing with swallowing; and achalasia may cause nocturnal regurgitation of undigested food from the esophagus.

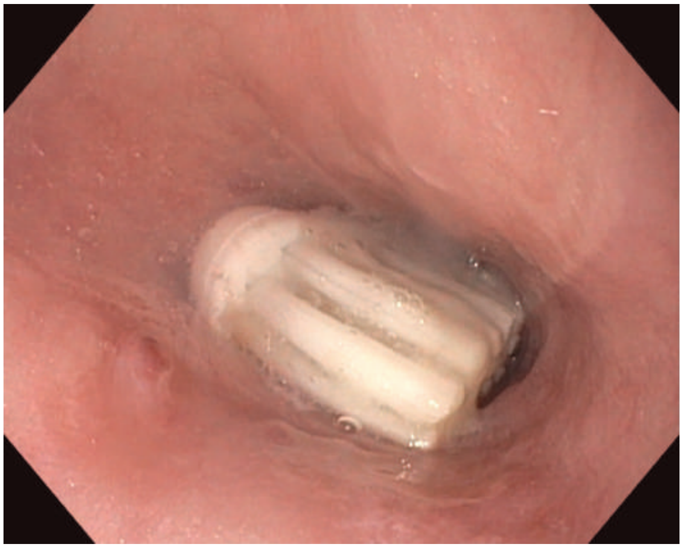

FIGURE 333-45 Esophageal foreign body. Intentionally ingested toothbrush impacted in the esophageal lumen.

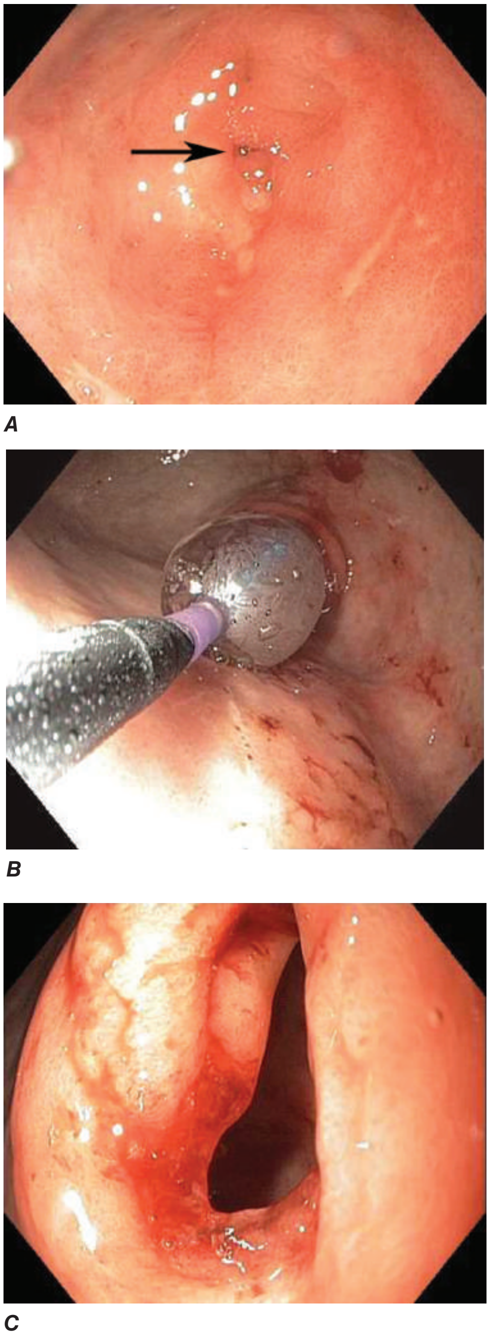

FIGURE 333-46 Gastric outlet obstruction due to pyloric stenosis. ***A.*** Nonsteroidal anti-inflammatory agent–induced ulcer disease with severe stenosis of the pylorus (*arrow*). ***B.*** Balloon dilation of the stenosis. ***C.*** Appearance of pyloric ring after dilation.

When mechanical obstruction is suspected, endoscopy is a useful initial diagnostic test, since it permits immediate biopsy and/or dilation of strictures, masses, or rings. The presence of linear furrows and multiple corrugated rings throughout a narrowed esophagus should raise suspicion for eosinophilic esophagitis, an increasingly recognized cause of recurrent dysphagia and food impaction **(Fig. 333-55)**. Blind or forceful passage of an endoscope may lead to perforation in a patient with stenosis of the cervical esophagus or a Zenker's diverticulum **(Fig. 333-56)**, but gentle passage of an endoscope under direct visual guidance is reasonably safe. Endoscopy can miss a subtle stricture or ring in some patients.

When transfer dysphagia is evident or an esophageal motility disorder is suspected, esophageal radiography and/or a video-swallow study are the best initial diagnostic tests. The oropharyngeal swallowing mechanism, esophageal peristalsis, and the lower esophageal sphincter

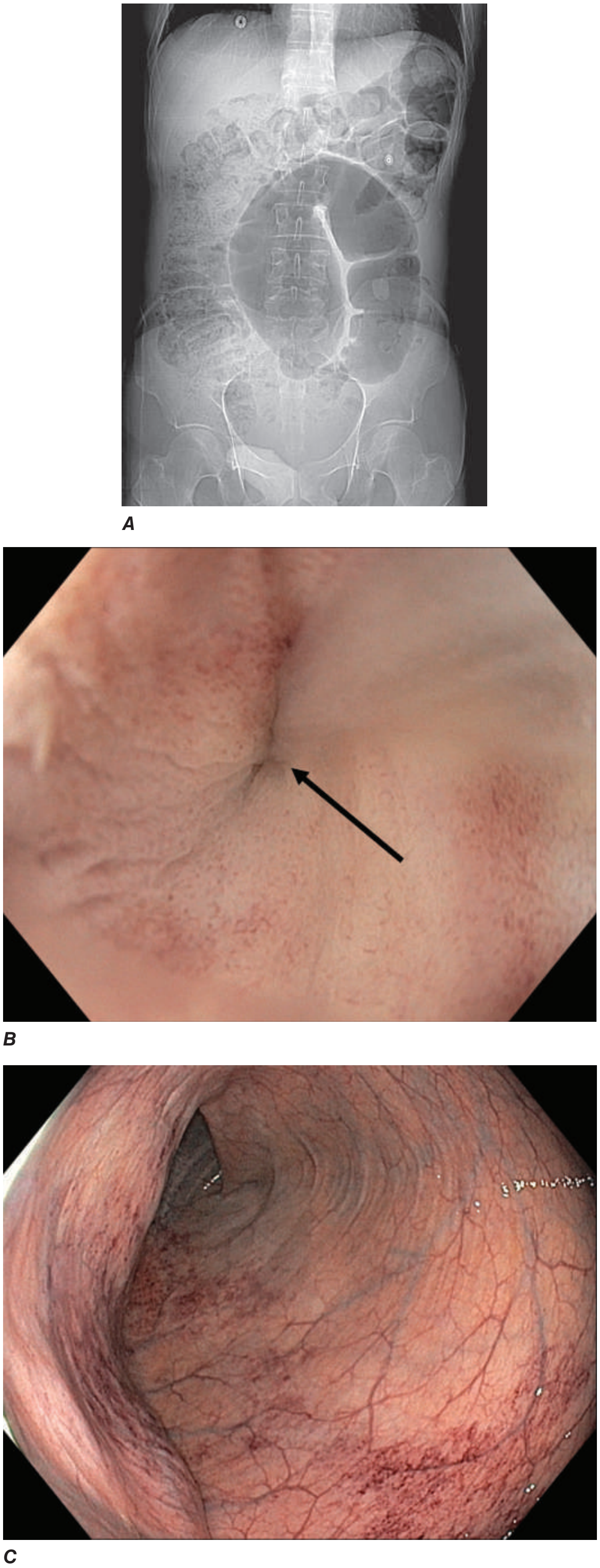

FIGURE 333-47 Sigmoid volvulus. ***A.*** Abdominal x-ray showing characteristic radiologic appearance of a "bent inner tube." ***B.*** Site of sigmoid torsion identified at colonoscopy ***C.*** Significantly dilated colonic lumen proximal to the sigmoid twist with mild ischemic-appearing mucosal changes.

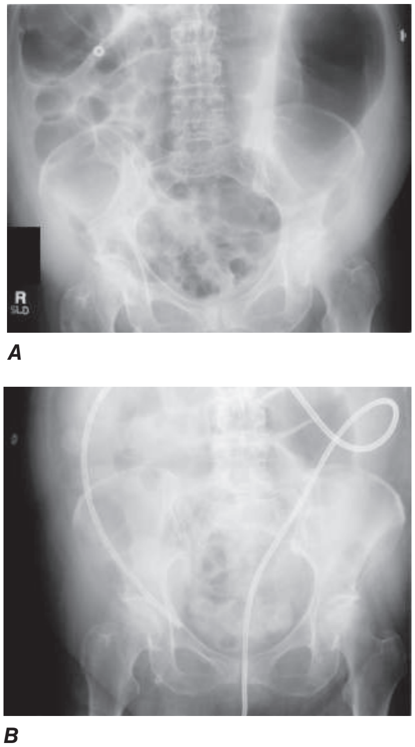

FIGURE 333-48 Acute colonic pseudoobstruction. ***A.*** Acute colonic dilation occurring in a patient soon after knee surgery. ***B.*** Colonoscopic placement of decompression tube with marked improvement in colonic dilation.

can all be assessed. In some disorders, subsequent esophageal manometry is required for diagnosis.

Various causes of dysphagia are amenable to endoscopic therapy. Benign strictures, rings, and webs can be dilated using a through-the-scope balloon (Fig. 333-57) or a tapered polyvinyl dilator passed over a guide wire. In some instances, fibrotic strictures may respond to needle-knife electroincision (Fig. 333-58) when they prove refractory to dilation. Self-expanding esophageal stents can be used to palliate dysphagia from malignant obstruction (Fig. 333-59), and flexible endoscopic cricopharyngeal myotomy is an option for Zenker's diverticulum (Video V5-20). Recent advances in third-space (submucosal) endoscopy have enabled the development of procedures, such as POEM (Video V5-21) and POET (Video V5-22), for the management of achalasia and select subepithelial esophageal tumors, respectively.

■ ENDOSCOPIC TREATMENT OF OBESITY

A significant proportion of Americans are overweight or obese, and obesity-associated diabetes has become a major public health problem. Bariatric surgery is the most effective weight-loss intervention, decreasing long-term mortality in obese persons, but many patients choose not to undergo surgery. Endoscopic treatments for obesity have been developed and include insertion of an intragastric balloon or duodenojejunal bypass liner, placement of a percutaneous gastric tube for aspiration of gastric contents after meals, duodenal mucosal electroporation, or endoscopic sleeve gastroplasty, which utilizes endoscopic suturing to narrow the lumen of the gastric body (Video V5-23). Prospective trials show that these treatments induce total-body weight loss of 7–20% and provide varying degrees of glycemic control and improvement in other diseases linked to obesity, such as metabolic-associated fatty liver disease (MAFLD). Additional endoscopic modalities are undergoing clinical trials. The efficacy of bariatric endoscopy may approach that of bariatric surgery in the short term. Long-term outcomes are still under evaluation.

■ TREATMENT OF MALIGNANCIES

Endoscopy plays an important role in the treatment of gastrointestinal malignancies. Early-stage malignancies limited to the mucosal and superficial submucosal layers may be resected using the techniques

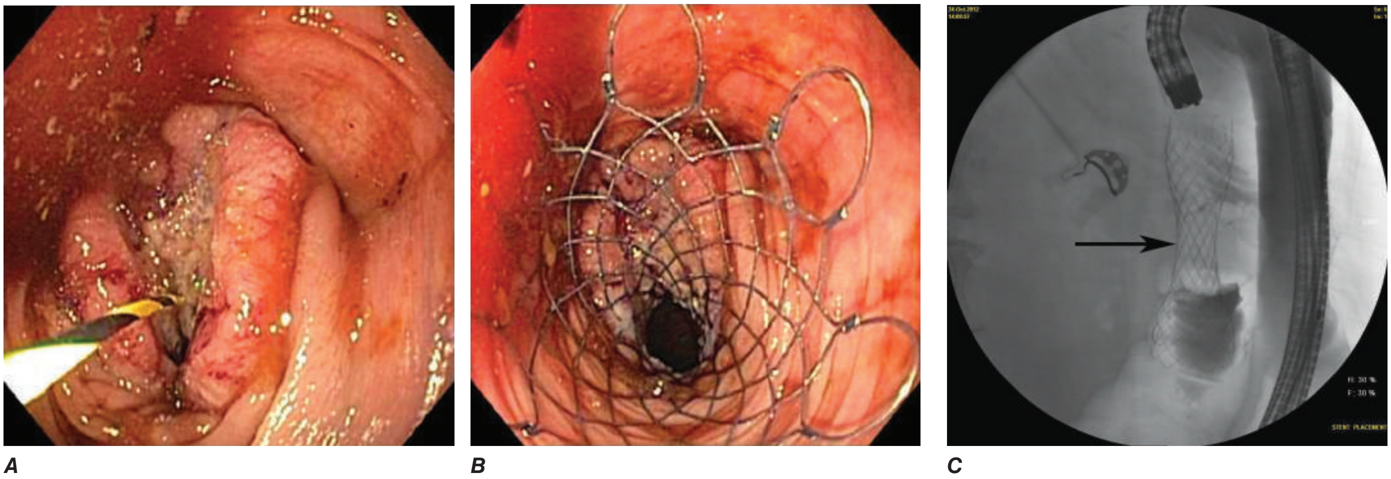

FIGURE 333-49 Obstructing colonic carcinoma. ***A.*** Colonic adenocarcinoma causing marked luminal narrowing of the distal transverse colon. ***B.*** Endoscopic placement of a self-expandable metal stent. ***C.*** Radiograph of expanded stent across the obstructing tumor with a residual waist (*arrow*).

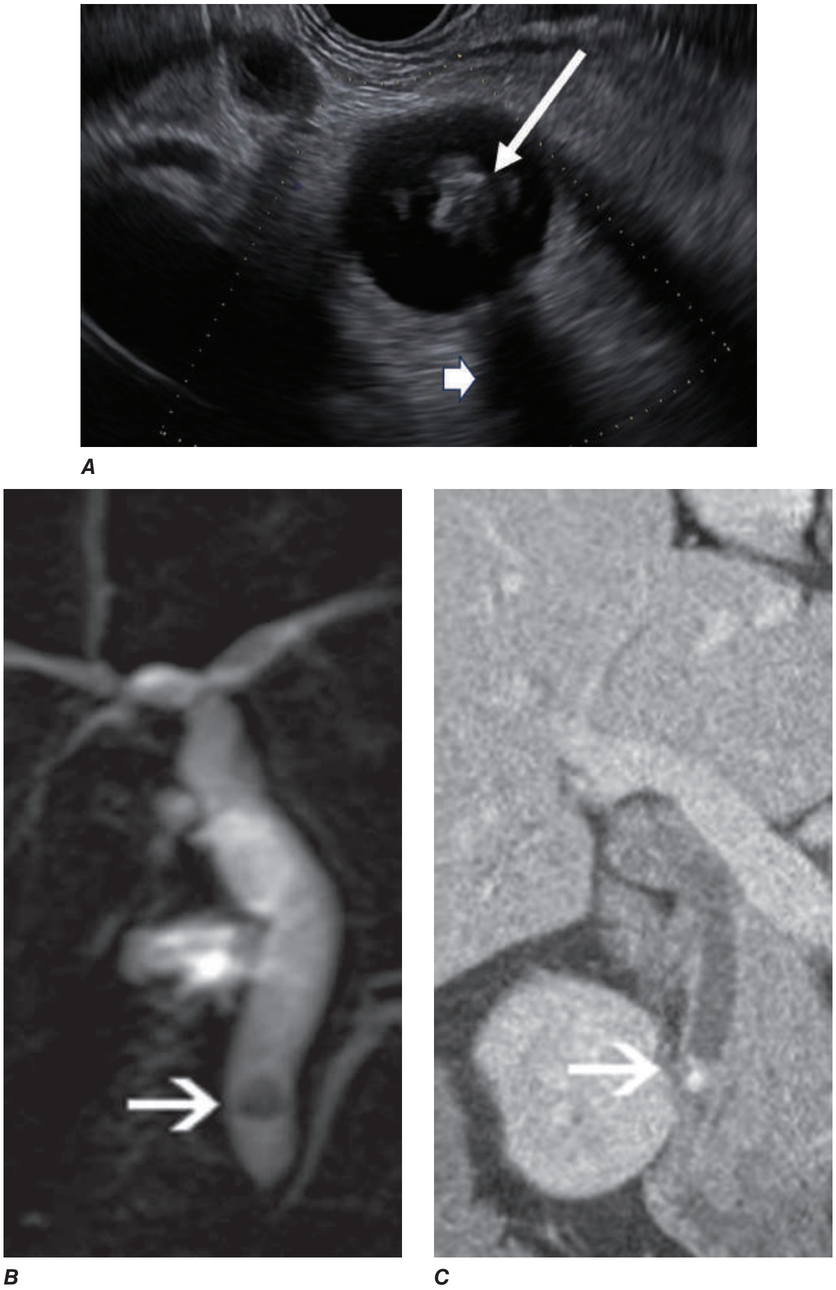

FIGURE 333-50 Methods of bile duct imaging. *Arrows* mark bile duct stones. ***A.*** Endoscopic ultrasound (EUS) demonstrating a hyperechoic stone with acoustic shadowing (*arrowhead*). ***B.*** Magnetic resonance cholangiopancreatography (MRCP). ***C.*** Helical computed tomography (CT).

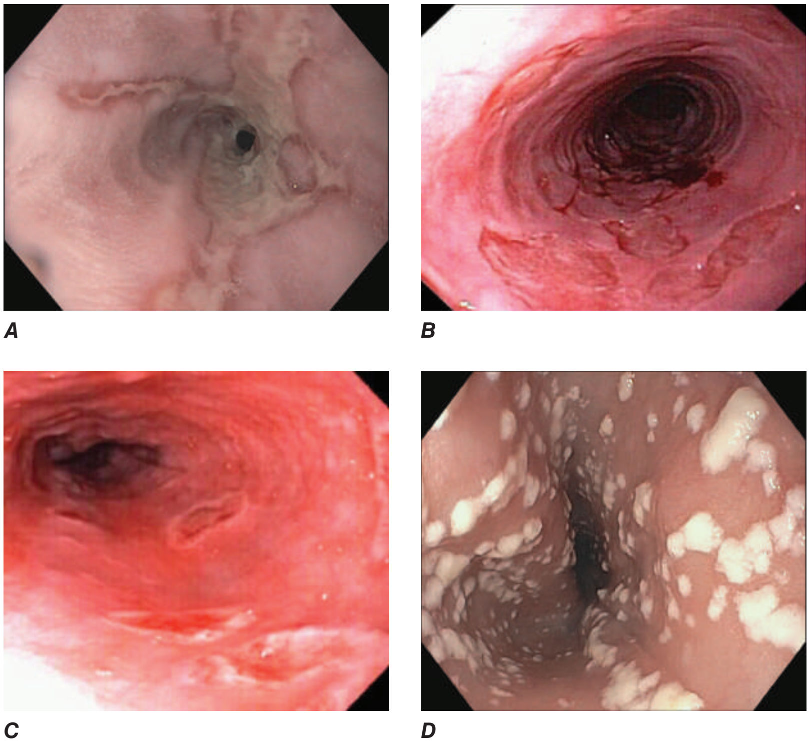

FIGURE 333-51 Causes of esophagitis. ***A.*** Severe reflux esophagitis with mucosal ulceration and friability. ***B.*** Cytomegalovirus esophagitis. ***C.*** Herpes simplex virus esophagitis with target-type shallow ulcerations. ***D.*** *Candida* esophagitis with white plaques adherent to the esophageal mucosa.

of EMR (Video V5-5) or ESD (Video V5-6). RFA and cryotherapy are effective modalities for ablative treatment of high-grade dysplasia and intramucosal cancer in Barrett's esophagus **(Video V5-24)**. Gastrointestinal stromal tumors can be removed en bloc by endoscopic full-thickness resection (Video V5-4). In general, endoscopic techniques offer the advantage of a minimally invasive approach to treatment but rely on other imaging techniques (such as CT, magnetic resonance imaging [MRI], positron emission tomography [PET], and EUS) to exclude distant metastases or locally advanced disease better treated by surgery or other modalities. The decision to treat an early-stage gastrointestinal malignancy endoscopically is often made in collaboration with a surgeon and/or oncologist.

Endoscopic palliation of gastrointestinal malignancies relieves symptoms and, in many cases, prolongs survival. Malignant obstruction can be relieved by endoscopic stent placement (Figs. 333-18, 333-49, 333-59, and **333-60**; Videos V5-17 and V5-18), and malignant gastrointestinal bleeding can be palliated endoscopically as well. EUS-guided celiac plexus neurolysis may relieve pancreatic cancer pain.

ANEMIA AND OCCULT BLOOD IN THE STOOL

Iron-deficiency anemia may be attributed to poor iron absorption (as in celiac sprue) or, more commonly, chronic blood loss. Intestinal bleeding should be strongly suspected in men and postmenopausal women with iron-deficiency anemia, and colonoscopy is indicated in

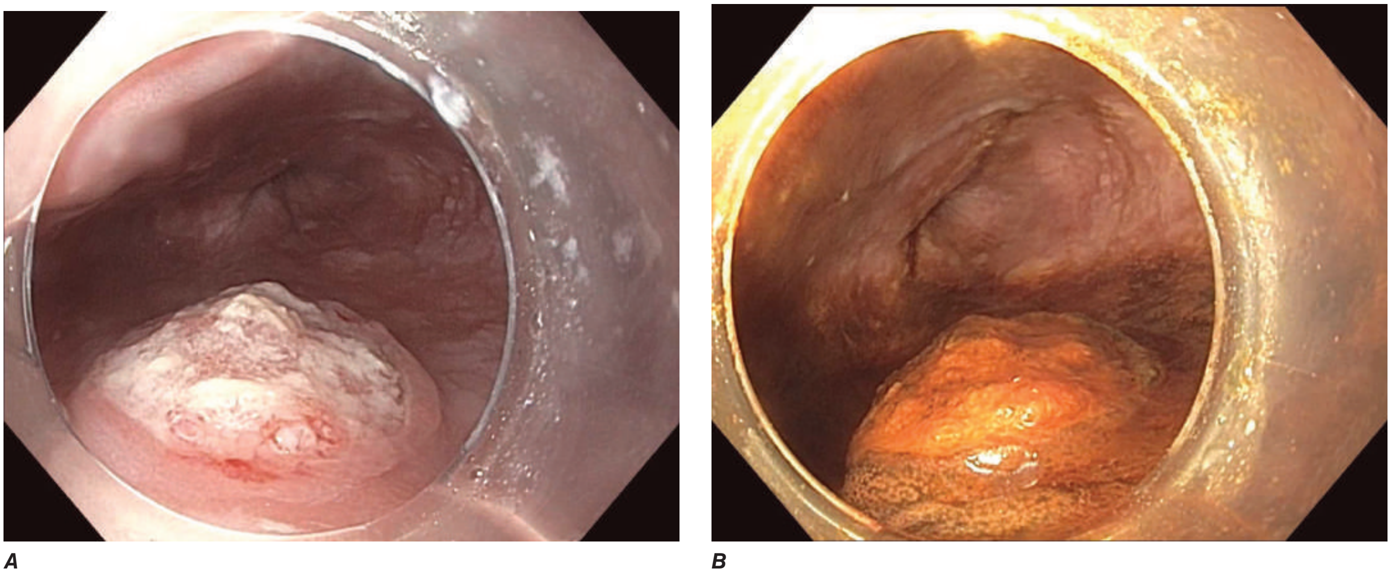

FIGURE 333-52 Early squamous cell cancer. ***A.*** Nodularity in the distal esophagus due to T1 esophageal squamous cell cancer. ***B.*** The nodular lesion remains unstained following chromoendoscopy with Lugol's solution without additional unstained areas. ***C.*** Circumferential mucosal incision around the lesion. ***D.*** Resection defect following en bloc removal of the lesion via endoscopic submucosal dissection.

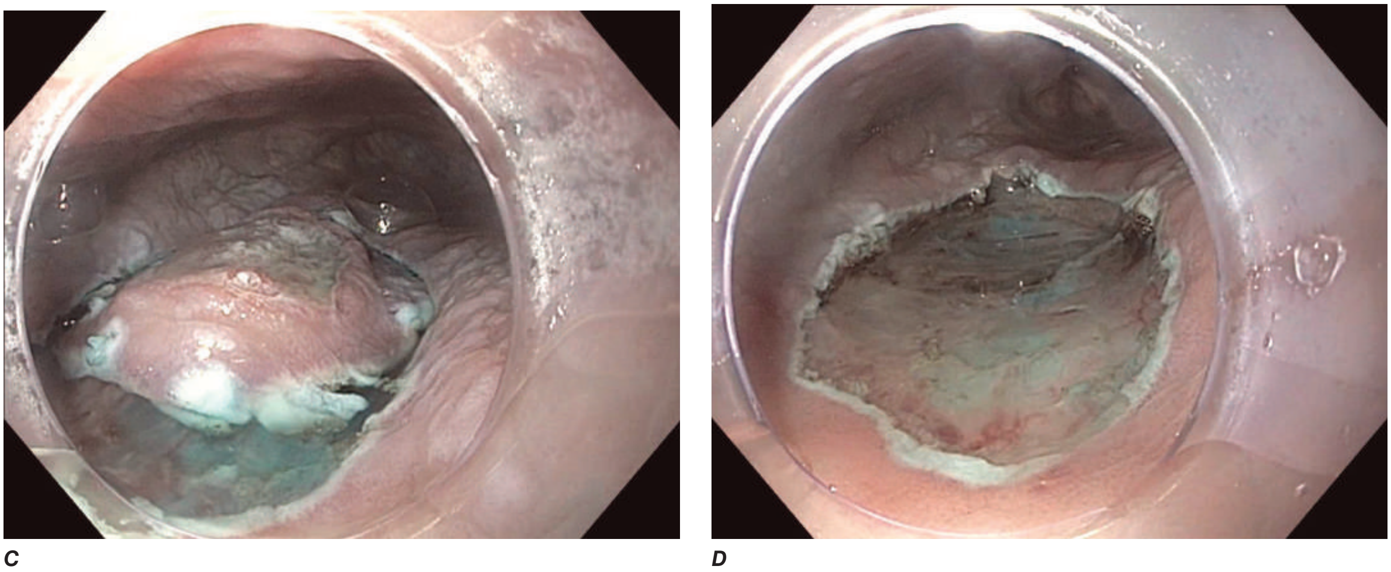

FIGURE 333-52 (*Continued*)

such patients, even in the absence of detectable occult blood in the stool. Approximately 30% will have large colonic polyps or colorectal cancer, and a few patients will have colonic vascular lesions. When a convincing source of blood loss is not found in the colon, upper gastrointestinal endoscopy should be considered; if no lesion is found, duodenal biopsies should be obtained to exclude sprue **(Fig. 333-61)**. Small-bowel evaluation with capsule endoscopy **(Fig. 333-62)**, CT or magnetic resonance (MR) enterography, or deep enteroscopy may be appropriate if both EGD and colonoscopy are unrevealing.

Tests for occult blood in the stool detect hemoglobin or the heme moiety and are most sensitive for colonic blood loss, although they will also detect larger amounts of upper gastrointestinal bleeding. Patients with occult blood in the stool should undergo colonoscopy to diagnose or exclude colorectal neoplasia, especially if they are >50 years old or have a family history of colonic neoplasia. Whether upper endoscopy is also indicated depends on the patient's symptoms.

The small intestine may be the source of chronic intestinal bleeding, especially if colonoscopy and upper endoscopy are not diagnostic. The utility of small-bowel evaluation varies with the clinical setting and is most important in patients in whom bleeding causes chronic or recurrent anemia. In contrast to the low diagnostic yield of small-bowel radiography, positive findings on capsule endoscopy are seen in 50–70% of patients with suspected small intestinal bleeding. The most common finding is mucosal vascular ectasia. CT and MR enterography accurately detect small-bowel masses and Crohn's disease and are also useful for initial small-bowel evaluation. Deep enteroscopy may follow capsule endoscopy for biopsy of lesions or to provide specific therapy, such as argon plasma coagulation of vascular ectasias **(Fig. 333-63)**.

COLORECTAL CANCER SCREENING

The majority of colon cancers develop from preexisting colonic adenomas, and colorectal cancer can be largely prevented by the detection and removal of adenomatous polyps **(Video V5-25)**. The choice of screening strategy for an asymptomatic person depends on personal and family history. Individuals with inflammatory bowel disease, a history of colorectal polyps or cancer, family members with adenomatous polyps or cancer, or certain familial cancer syndromes **(Fig. 333-64)** are at increased risk for colorectal cancer. An individual without these factors is generally considered at average risk.

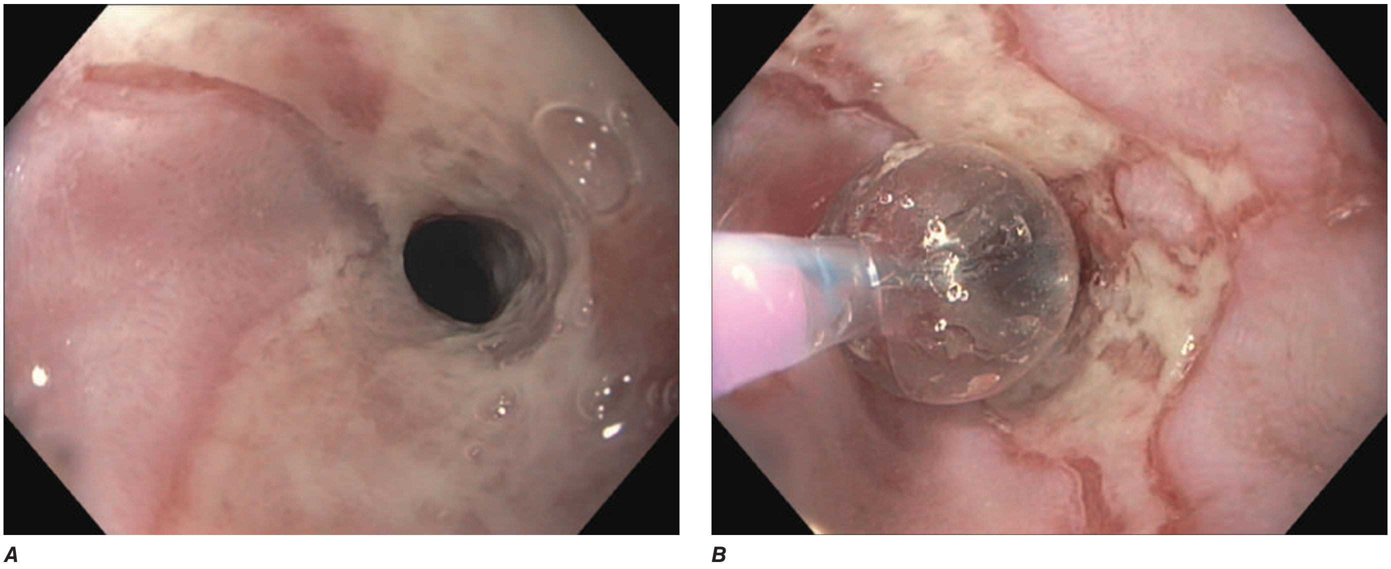

FIGURE 333-53 Esophageal stricture. ***A.*** Peptic stricture associated with esophagitis. ***B.*** Balloon dilation of peptic stricture.

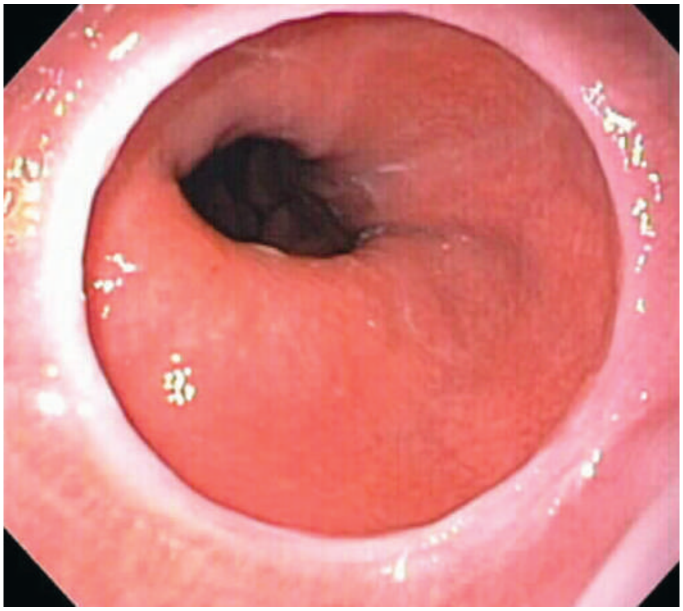

FIGURE 333-54 Schatzki's ring at the gastroesophageal junction.

Screening strategies are summarized in **Table 333-3**. While fecal immunochemical tests (FITs) for heme or stool tests for occult blood have been shown to decrease the mortality rate from colorectal cancer, they do not detect some cancers and many polyps. FIT-DNA multitargeted stool DNA tests appear to be more sensitive, but direct visualization of the colon is the gold standard method for detection of polyps and cancers and remains a preferred screening strategy. Sigmoidoscopy is also used for colorectal cancer screening. However, the distribution of colon cancers has changed in the United States over time, with proportionally fewer rectal and left-sided cancers than in the past. Large American studies of colonoscopy for screening of average-risk individuals show that cancers are roughly equally distributed between the left and right colon and half of patients with right-sided lesions have no polyps in the left colon. Visualization of the entire colon thus appears to be the optimal strategy for colorectal cancer screening and prevention.

Computed tomography colonography (CTC) is a radiologic technique that images the colon with CT following rectal insufflation of the colonic lumen. Computer rendering of CT images generates an electronic display of a virtual "flight" along the colonic lumen, simulating colonoscopy **(Fig. 333-65)**. Findings detected during CTC often require subsequent conventional colonoscopy for confirmation and treatment.

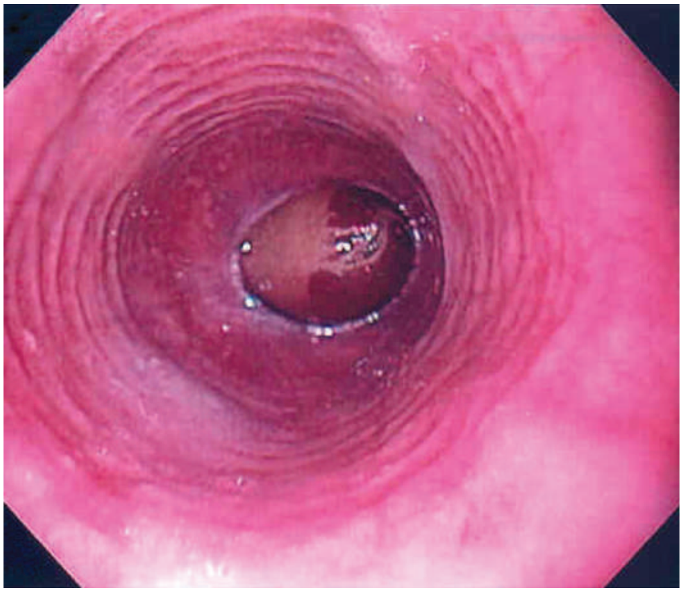

FIGURE 333-55 Eosinophilic esophagitis. Multiple circular rings of the esophagus creating a corrugated appearance and an impacted grape at the narrowed esophagogastric junction. The diagnosis requires biopsy with histologic finding of ≥15 eosinophils/high-power field.

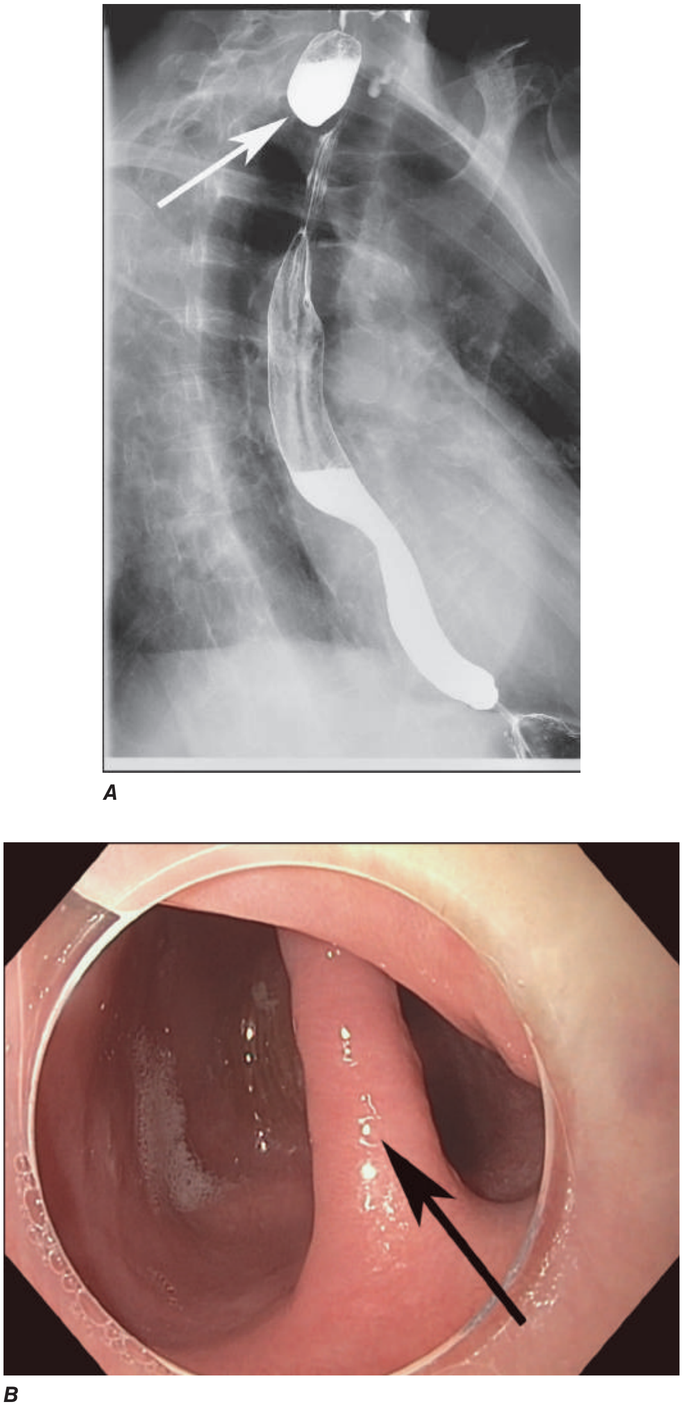

FIGURE 333-56 Zenker's diverticulum. ***A.*** Contrast esophagography demonstrates a moderate-sized Zenker's diverticulum. ***B.*** Endoscopic view of the Zenker's diverticulum (*left*) relative to the true esophageal lumen (*right*) separated by the diverticular septum. ***C.*** Flexible endoscopic diverticulotomy using an electrosurgical knife. ***D.*** Appearance after diverticulotomy.

■ DIARRHEA

Most cases of diarrhea are acute, self-limited, and due to infections or medication. Chronic diarrhea (lasting >4–6 weeks) is more often due to a primary inflammatory, malabsorptive, or motility disorder; is less likely to resolve spontaneously; and generally, requires diagnostic evaluation. Patients with chronic diarrhea or severe, unexplained acute diarrhea often undergo endoscopy if stool tests for pathogens are

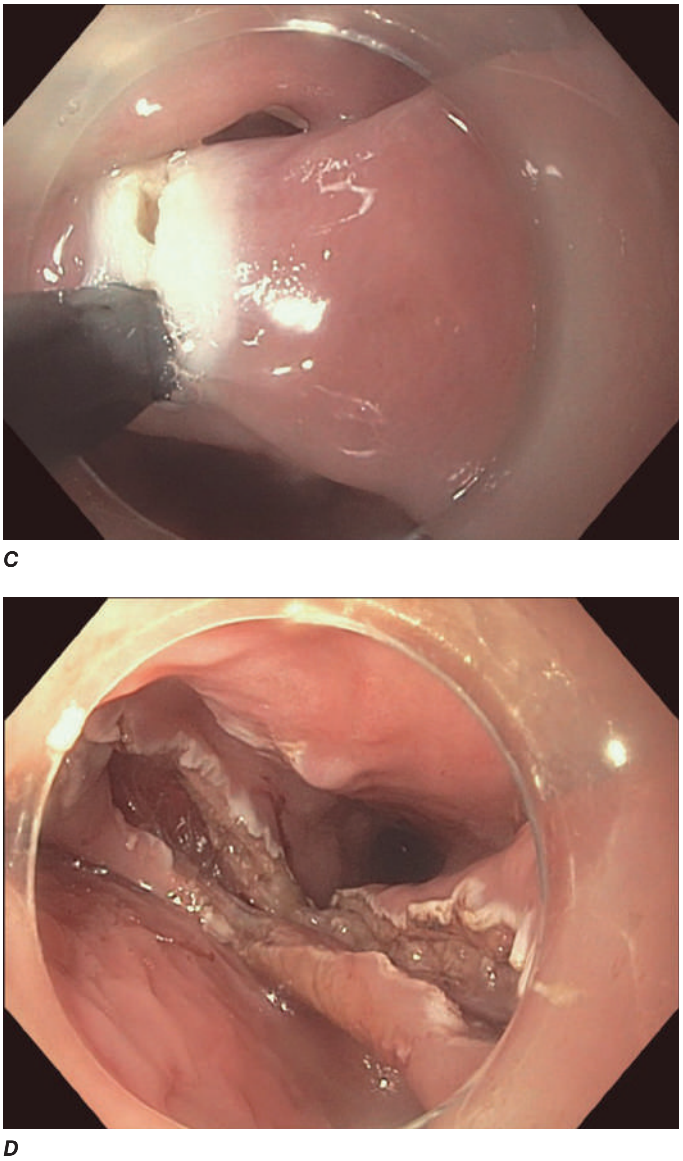

FIGURE 333-56 (*Continued*)

unrevealing. The choice of endoscopic testing depends on the clinical setting.

Patients with colonic symptoms and findings such as bloody diarrhea, tenesmus, fever, or leukocytes in stool generally undergo sigmoidoscopy or colonoscopy to assess for colitis (Fig. 333-9). Sigmoidoscopy is an appropriate initial test in most patients. Conversely, patients with symptoms and findings suggesting small-bowel disease, such as large-volume watery stools, substantial weight loss, and malabsorption of iron, calcium, or fat, may undergo upper endoscopy with duodenal aspirates for assessment of bacterial overgrowth and biopsies for assessment of mucosal diseases, such as celiac sprue.

Many patients with chronic diarrhea do not fit either of these patterns. In the setting of a long-standing history of altered bowel habits dating to early adulthood, without findings such as blood in the stool or anemia, a diagnosis of irritable bowel syndrome may be made without direct visualization of the bowel and by relying on appropriate blood tests (including complete blood count, C-reactive protein and antibody tests for celiac disease) to screen for other diagnoses. Steatorrhea and upper abdominal pain may prompt evaluation of the pancreas rather than the gut. Patients whose chronic diarrhea is not easily categorized often undergo initial colonoscopy to examine the entire colon and terminal ileum for inflammatory or neoplastic disease (Fig. 333-66).

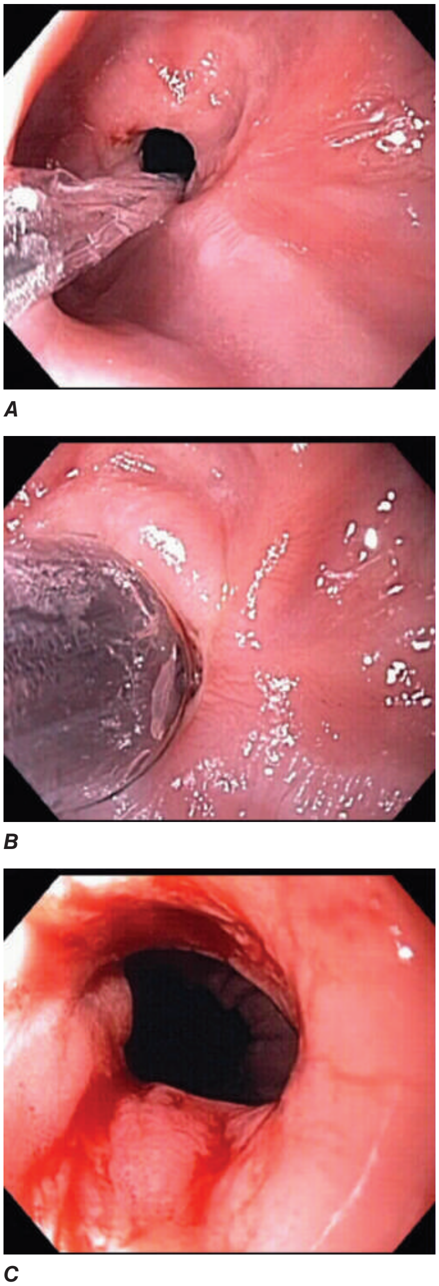

FIGURE 333-57 Endoscopic management of peptic stricture. ***A.*** Peptic stricture. ***B.*** Through-the-scope balloon dilation of stricture. ***C.*** Improvement in luminal diameter after dilation.

MINOR HEMATOCHEZIA

Bright red blood passed with or on formed brown stool usually has an anal, rectal, or sigmoid source (Fig. 333-67). Even trivial amounts of hematochezia should be investigated with colonoscopy and/or flexible sigmoidoscopy together with anoscopy to exclude polyps or cancers, especially in patients >40 years old and those with a personal or family history of colorectal polyps or cancer. Patients reporting red blood on the toilet tissue only, without blood in the toilet or on the stool, are generally bleeding from a lesion in the anal canal; careful external inspection, digital examination, and sigmoidoscopy with anoscopy may be sufficient for diagnosis in such cases.

PANCREATITIS

About 20% of patients with pancreatitis have no identified cause after routine clinical investigation (including a review of medication and alcohol use; measurement of serum triglyceride, calcium, and immunoglobulin G subclass 4 [IgG4] levels; abdominal ultrasonography; and CT or MRI). Endoscopic assessment leads to a specific diagnosis

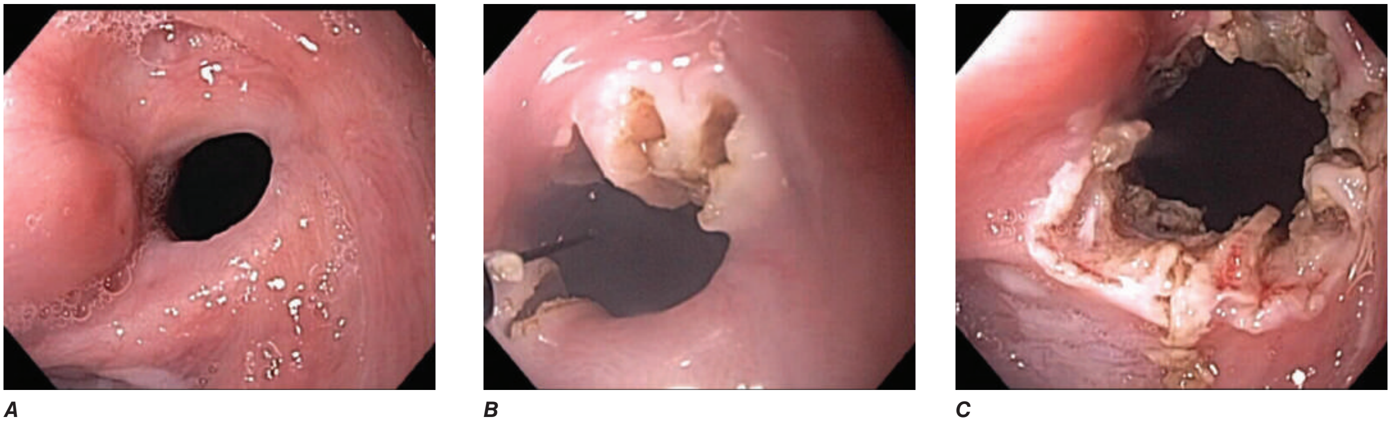

FIGURE 333-58 Endoscopic management of an esophagogastric anastomotic stricture. ***A.*** Recurrent anastomotic stricture despite periodic balloon dilation. ***B.*** Needle-knife electroincision of stricture. ***C.*** Improvement in luminal opening after therapy.

in the majority of such patients, often altering clinical management. Endoscopic investigation is particularly appropriate if the patient has had more than one episode of pancreatitis.

Microlithiasis, or the presence of microscopic crystals in bile, is a leading cause of previously unexplained acute pancreatitis and is sometimes seen during abdominal ultrasonography as layering sludge or flecks of floating, echogenic material in the gallbladder. EUS may identify microlithiasis or gallstones not seen on transabdominal ultrasound.

Previously undetected chronic pancreatitis, pancreatic malignancy, or pancreas divisum may be diagnosed by either ERCP or EUS. Autoimmune pancreatitis is often suspected based on CT, MRI, or serologic findings, but it may first become apparent during EUS and may require EUS-guided pancreatic biopsy for histologic diagnosis.

Severe pancreatitis often results in pancreatic fluid collections. Symptomatic pseudocysts and areas of walled-off pancreatic necrosis can be drained into the stomach or duodenum endoscopically, using transpapillary and transmural endoscopic techniques. Pancreatic necrosis can be debrided by direct endoscopic necrosectomy (Video V5-2) via an endoscopically created transmural drainage site.

■ CANCER STAGING

Local staging of esophageal, gastric, pancreatic, bile duct, and rectal cancers can be obtained with EUS (Fig. 333-20). EUS with fine-needle aspiration (Fig. 333-21) or biopsy currently provides the most accurate preoperative assessment of local tumor and nodal staging, but it does not detect many distant metastases. Details of the local tumor stage can guide treatment decisions, including resectability and need for neoadjuvant therapy. EUS with transesophageal needle biopsy may also be used to assess the presence of non-small-cell lung cancer in mediastinal nodes.

OPEN-ACCESS ENDOSCOPY

Direct scheduling of endoscopic procedures by providers without preceding gastroenterology consultation, or *open-access endoscopy*, is common. When the indications for endoscopy are clear-cut and appropriate, the procedural risks are low, and the patient understands what to expect, open-access endoscopy streamlines patient care and decreases costs.

Patients referred for open-access endoscopy should have a recent history, physical examination, and medication list that are available for review when the patient comes to the endoscopy suite. Patients with unstable or symptomatic cardiovascular or respiratory conditions should not be referred directly for open-access endoscopy. Those with particular conditions who are undergoing certain procedures should be prescribed prophylactic antibiotics before endoscopy (Table 333-1). In addition, patients taking anticoagulants and/or antiplatelet drugs may require adjustment of these agents before endoscopy based on the procedural risk for bleeding and their underlying risk for a thromboembolic event (Table 333-2).

Common indications for open-access EGD include dyspepsia resistant to a trial of appropriate therapy, dysphagia, gastrointestinal bleeding, and persistent anorexia or early satiety. Open-access colonoscopy is often requested in men or postmenopausal women with iron-deficiency anemia, in patients with hematochezia or occult blood in the stool, in patients with a previous history of colorectal adenomatous polyps or cancer, and for colorectal cancer screening. Flexible sigmoidoscopy is commonly performed as an open-access procedure.

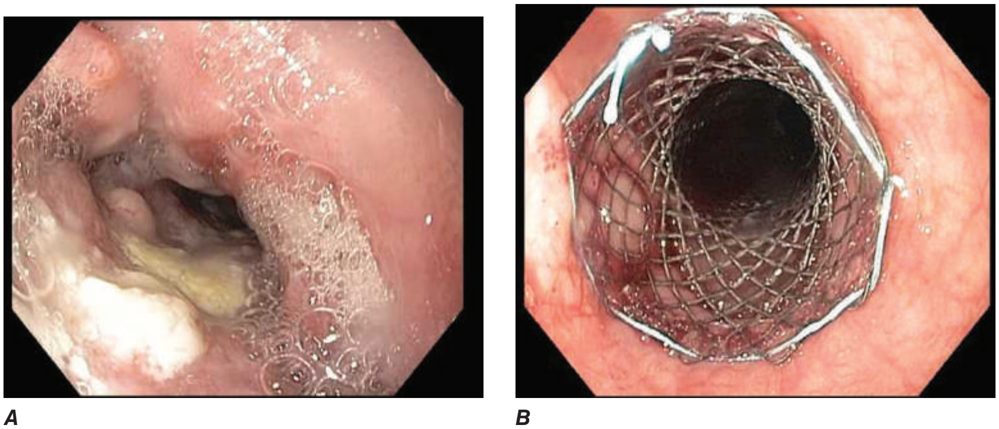

FIGURE 333-59 Palliation of malignant dysphagia. ***A.*** Obstructing distal esophageal cancer. ***B.*** Palliative stent placement.

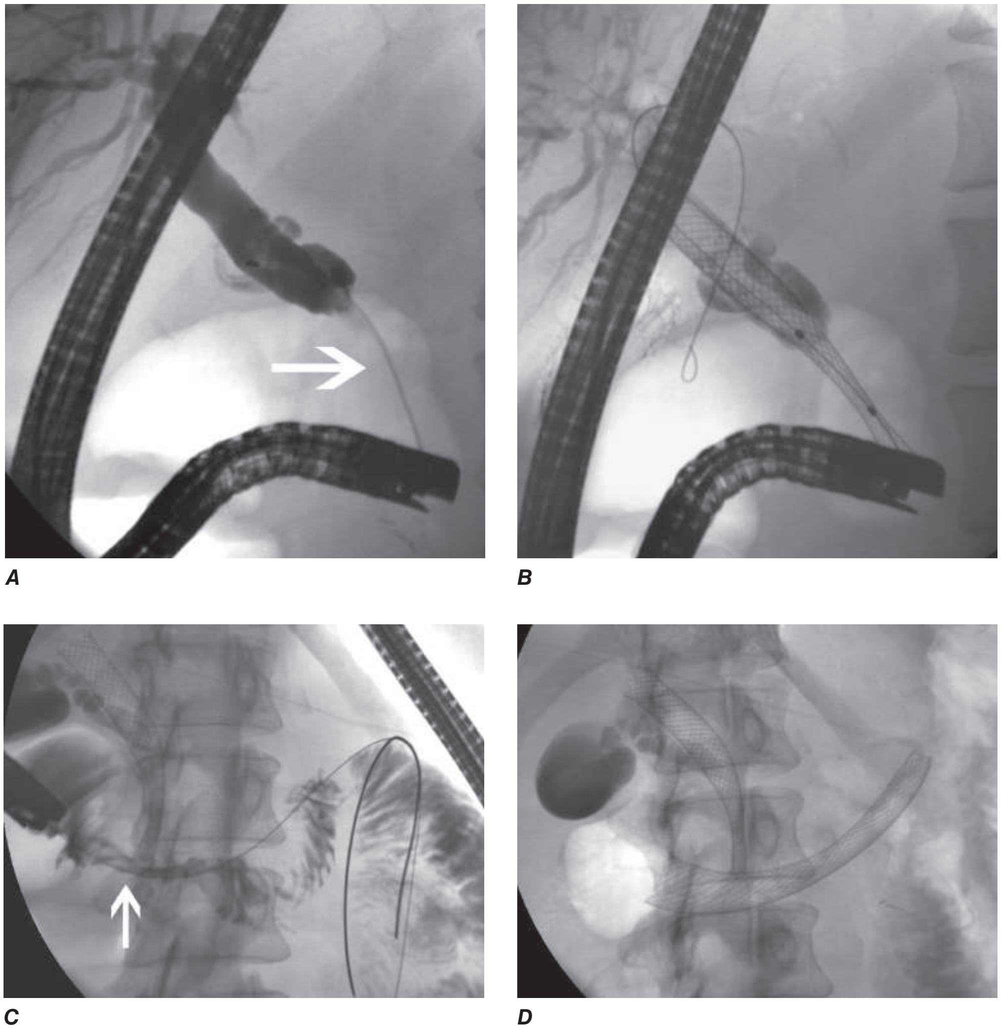

FIGURE 333-60 Placement of biliary and duodenal self-expanding metal stents (SEMS) for obstruction caused by pancreatic cancer. ***A.*** Endoscopic retrograde cholangiopancreatography (ERCP) demonstrates a distal bile duct stricture (*arrow*). ***B.*** A biliary SEMS is placed. ***C.*** Contrast injection demonstrates a duodenal stricture (*arrow*). ***D.*** Biliary and duodenal SEMS in place.

When patients are referred for open-access colonoscopy, the primary care provider may need to choose a colonic preparation. Commonly used oral preparations include polyethylene glycol lavage solution, with or without citric acid. A "split-dose" regimen improves the quality of colonic preparation. Osmotic purgative preparations (such as sodium phosphate or magnesium citrate) are also effective but may cause fluid and electrolyte abnormalities and renal toxicity, especially in patients with renal failure or congestive heart failure and those >70 years of age.

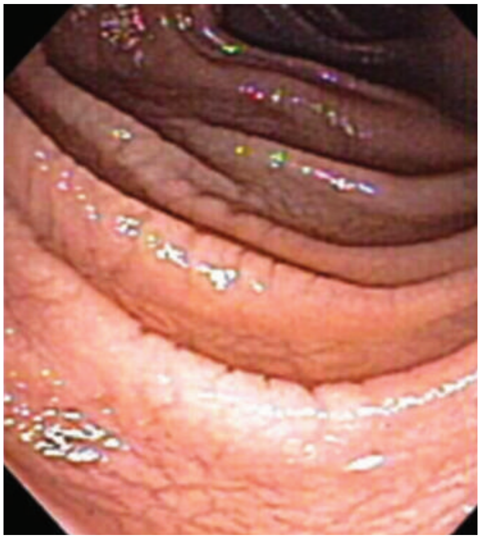

FIGURE 333-61 Celiac sprue. Scalloped duodenal folds in a patient with celiac sprue.

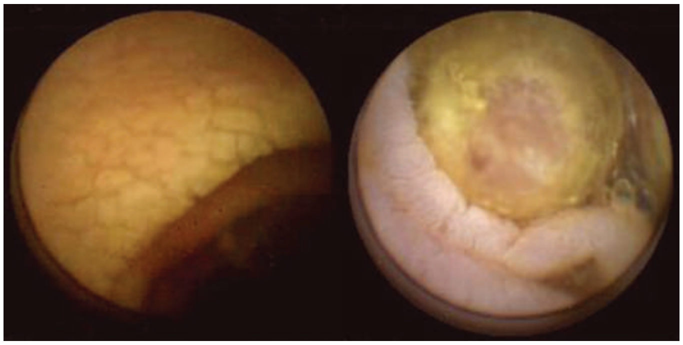

FIGURE 333-62 Capsule endoscopy. Images of a mildly scalloped jejunal fold (*left*) and an ileal tumor (*right*) in a patient with celiac sprue. *(Images courtesy of Dr. Elizabeth Rajan; with permission.)*

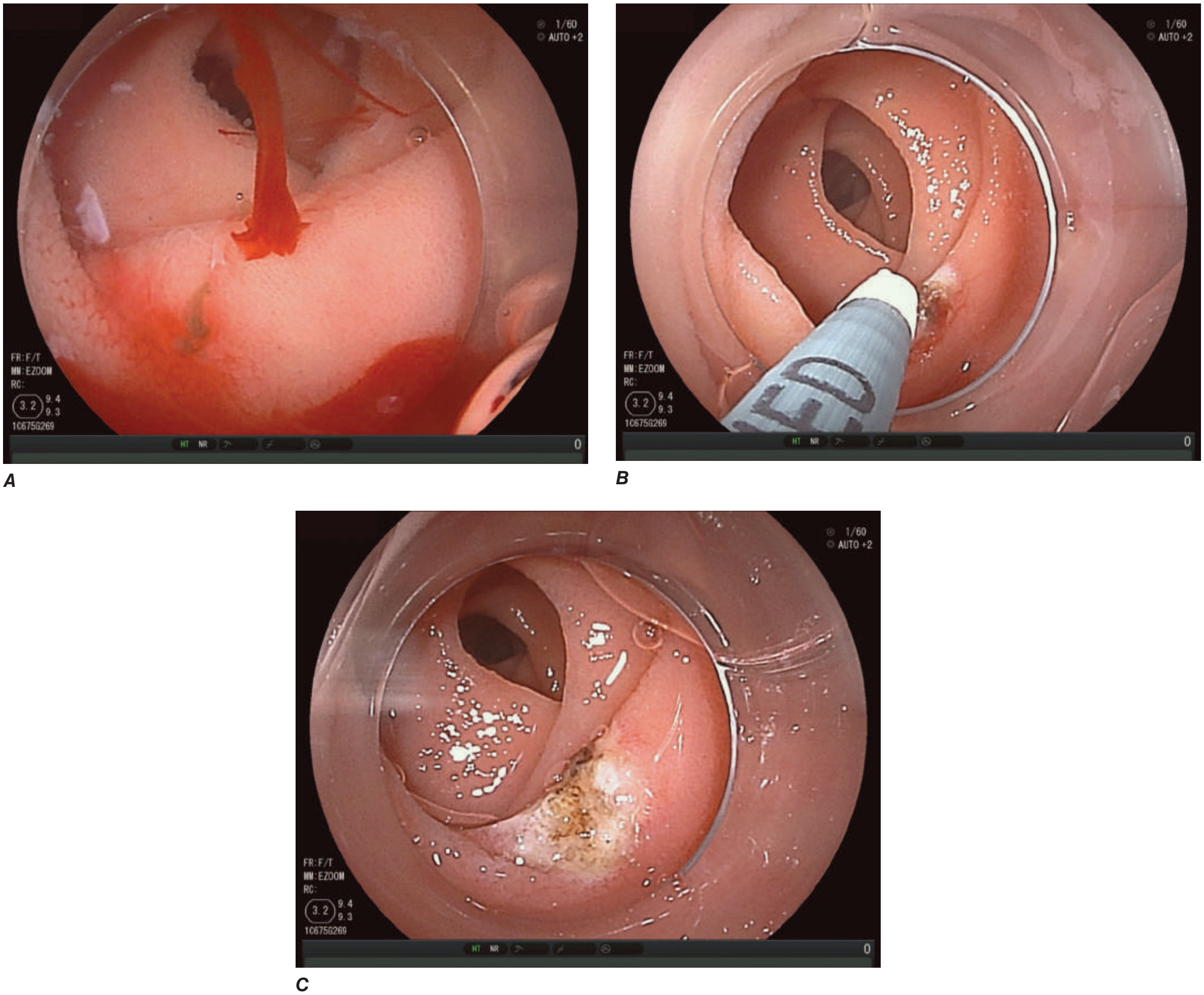

FIGURE 333-63 Small-bowel vascular ectasia. ***A.*** Actively bleeding mid-jejunal vascular ectasia identified by double-balloon enteroscopy. ***B.*** Ablation of vascular ectasia with argon plasma coagulation (APC). ***C.*** Hemostasis secured following APC.

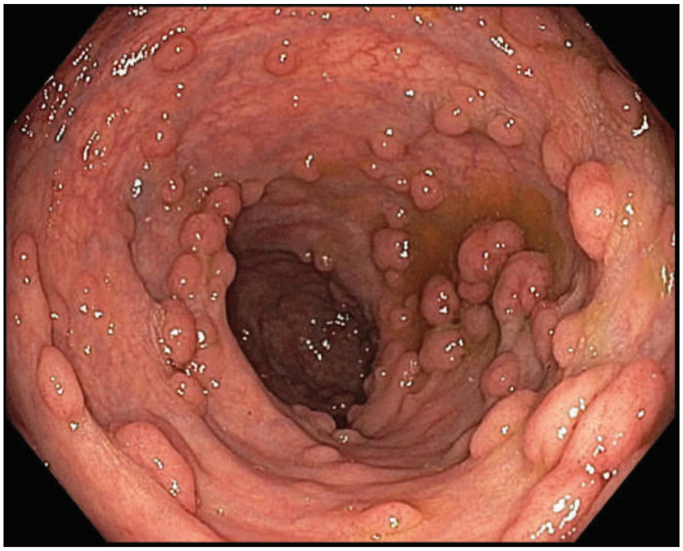

FIGURE 333-64 Familial adenomatous polyposis. Numerous colon polyps in a patient with familial adenomatous polyposis syndrome.

TABLE 333-3 Colorectal Cancer Screening Strategies

	CHOICES/RECOMMENDATIONS	COMMENTS
Average-Risk Patients		
Asymptomatic individuals between 45 and 75 years of age	Colonoscopy every 10 years[a]	Gold standard cancer prevention strategy
	Multitargeted stool DNA test every 1–3 years	Less sensitive than colonoscopy; colonoscopy if results are positive
	FIT or HSgFOBT every year, with or without flexible sigmoidoscopy every 10 years	Less sensitive than colonoscopy; colonoscopy if results are positive
	CT colonography every 5 years	Colonoscopy if results are positive
	Flexible sigmoidoscopy every 5 years	Does not detect proximal colon polyps and cancers; colonoscopy if an adenomatous polyp is found
Asymptomatic individuals > 75 years of age	Selective screening	Consider patient's overall health, results of previous screening exams, and preferences
Personal History of Polyps or CRC		
1–2 small (<10 mm) tubular adenomas	Repeat colonoscopy in 7–10 years[a]	Assuming complete polyp resection. Interval may vary based on prior personal history and family history
3–4 tubular adenomas <10 mm	Repeat colonoscopy in 3–5 years[a]; subsequent colonoscopy based on findings	Assuming complete polyp resection
5–10 tubular adenomas <10 mm	Repeat colonoscopy in 3 years[a]	Assuming complete polyp resection
>10 adenomas on a single exam	Repeat colonoscopy in 1 year[a]	Consider evaluation for FAP or HNPCC; see recommendations below
Adenoma ≥10 mm, or adenoma with tubulovillous or villous histology, or high-grade dysplasia	Repeat colonoscopy in 3 years[a]	Assuming complete polyp resection
Piecemeal removal of any sessile polyp	Exam in 6 months to verify complete removal	
Small (<1 cm) hyperplastic polyps of sigmoid and rectum	Repeat colonoscopy in 10 years[a]	Those with hyperplastic polyposis syndrome merit more frequent follow-up
Hyperplastic polyp ≥10 mm	Repeat colonoscopy in 3–5 years[a]	
1–2 SSPs <10 mm	Repeat colonoscopy in 5–10 years[a]	Assuming complete polyp resection
3–4 SSPs <10 mm	Repeat colonoscopy in 3–5 years[a]	Assuming complete polyp resection
5–10 SSPs <10 mm, or any single SSP ≥10 mm, or any SSP with dysplasia	Repeat colonoscopy in 3 years[a]	Assuming complete polyp resection. Serrated polyposis syndrome merits more frequent follow-up
Piecemeal removal of serrated polyp ≥1 cm	Exam in 2–6 months to verify complete removal	
Colon cancer	Evaluate entire colon around the time of resection, then repeat colonoscopy in 1 year[a]	Subsequent colonoscopy in 3 years if the 1-year examination is normal
Inflammatory Bowel Disease		
Long-standing (>8 years) ulcerative pancolitis or Crohn's colitis, or left-sided ulcerative colitis of >15 years' duration	Colonoscopy with biopsies every 1–2 years	Consider chromoendoscopy or other advanced imaging techniques for detection of flat dysplasia during colonoscopy
Family History of Polyps or CRC		
First-degree relatives with only small tubular adenomas	Same as average risk	
One first-degree relative with CRC or advanced adenoma at age ≥60 years	Begin screening starting at age 40, tests and intervals per average-risk recommendations	
One first-degree relative with CRC or advanced adenoma at age <60 years, *or* two first-degree relatives with CRC or advanced adenomas at any age	Colonoscopy every 5 years beginning at age 40 years or 10 years before the age at diagnosis of the youngest affected relative, whichever is earlier	
Familial adenomatous polyposis (FAP)	Sigmoidoscopy or colonoscopy annually, beginning at age 10–12 years	Consider genetic counseling and testing; consider screening family members
Hereditary nonpolyposis colorectal cancer (HNPCC; Lynch syndrome)	Colonoscopy every 2 years beginning at age 20–25 years (or 10 years younger than the youngest first-degree relative was when diagnosed with CRC) until age 40, then annually thereafter	Consider histologic evaluation for microsatellite instability in tumor specimens of patients who meet modified Bethesda criteria; consider genetic counseling and testing, consider screening family members
Serrated polyposis syndrome (SPS)	Colonoscopy at age 40 (or the same age at which the youngest first-degree relative was when diagnosed with SPS, or 10 years younger than the youngest first-degree relative was when diagnosed with CRC), then every 1–2 years thereafter	Consider screening family members, even of patients with multiple serrated polyps who do not meet SPS criteria
Family colon cancer syndrome X	Colonoscopy every 3–5 years beginning 10 years before the age at diagnosis of the youngest affected relative	

[a]Assumes good colonic preparation and complete examination to cecum. [b]High-risk adenoma: any adenoma ≥1 cm in size or containing high-grade dysplasia or villous features.

Abbreviations: CRC, colorectal cancer; CT, computed tomography; FIT, fecal immunochemical test; HSgFOBT, high-sensitivity guaiac fecal occult blood test; SSP, sessile serrated polyp.

Sources: Adapted from U.S. Preventative Services Task Force Draft Guidelines finalized May 18, 2021 (*https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/colorectal-cancer-screening*) and American Cancer Society Guidelines (*https://www.cancer.org/cancer/types/colon-rectal-cancer/detection-diagnosis-staging/acs-recommendations.html*), both accessed on September 15, 2023, as well as the following articles: DK Rex et al: Colorectal cancer screening: Recommendations for physicians and patients from the U.S. Multi-Society Task Force on Colorectal Cancer. Gastroenterology 153:307, 2017; S Gupta et al: Recommendations for follow-up after colonoscopy and polypectomy: A consensus update by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 158:1131, 2020; G Mankaney et al: Serrated polyposis syndrome. Clin Gastroenterol Hepatol 18:777, 2020.

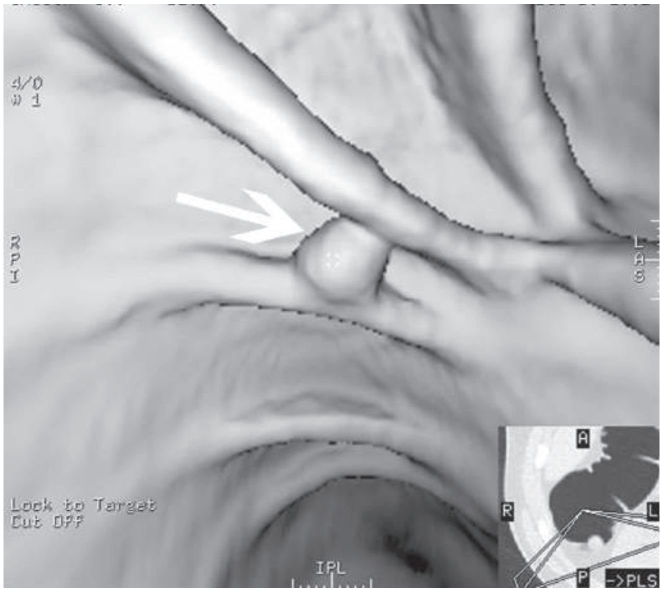

FIGURE 333-65 Virtual colonoscopy image of a colon polyp (*arrow*). *(Image courtesy of Dr. Jeff Fidler; with permission.)*

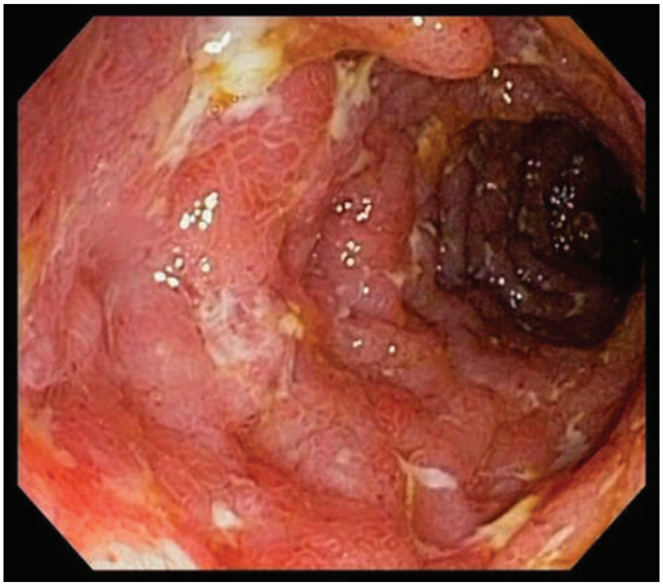

FIGURE 333-66 Crohn's ileitis. Edema, erythema, ulcers, and exudates involving the terminal ileum.

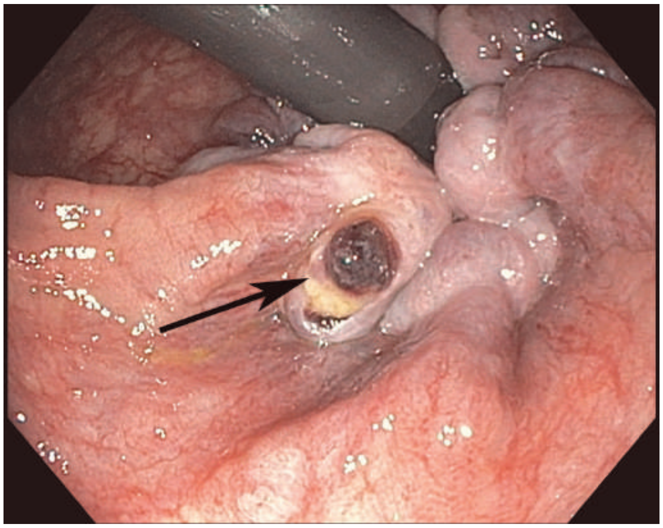

FIGURE 333-67 Internal hemorrhoids with bleeding stigmata (*arrow*) as seen on retroflexed view of the rectum.

FURTHER READING

Ahmed O et al: AGA clinical practice update on the optimal management of the malignant alimentary tract obstruction: expert review. Clin Gastroenterol Hepatol 19:1780, 2021.

ASGE Standards of Practice Committee et al: Antibiotic prophylaxis for GI endoscopy. Gastrointest Endosc 81:81, 2015.

ASGE Standards of Practice Committee et al: Open-access endoscopy. Gastrointest Endosc 81:1326, 2015.

Kaplan DE et al: AASLD practice guidance on risk stratification and management of portal hypertension and varices in cirrhosis. Hepatology 79:1180, 2024.

Laine L et al: ACG clinical guideline: Upper gastrointestinal and ulcer bleeding. Am J Gastroenterol 116:899, 2021.

Sengupta N et al: Management of patients with acute lower gastrointestinal bleeding: An updated ACG guideline. Am J Gastroenterol 118:208, 2023.

Shaheen NJ et al: Guideline to practice: diagnosis and management of Barrett's esophagus: An updated ACG guideline. Am J Gastroenterol 117:1177, 2022.

Shaukat A et al: ACG clinical guidelines: colorectal cancer screening 2021. Am J Gastroenterol 116:458, 2021.

334

Diseases of the Esophagus

Peter J. Kahrilas, Ikuo Hirano*

ESOPHAGEAL STRUCTURE AND FUNCTION

The esophagus is a hollow, muscular tube coursing through the posterior mediastinum joining the hypopharynx to the stomach with a sphincter at each end. It functions to transport food and fluid between these ends, otherwise remaining empty. The physiology of swallowing, esophageal motility, and oral and pharyngeal dysphagia are described in Chap. 47. Esophageal diseases can be manifested by impaired function or pain. Key functional impairments are swallowing disorders and excessive gastroesophageal reflux. Pain, sometimes indistinguishable from cardiac chest pain, can result from inflammation, infection, dysmotility, or neoplasm.

SYMPTOMS OF ESOPHAGEAL DISEASE

The clinical history remains central to the evaluation of esophageal symptoms. A thoughtfully obtained history will often expedite management. Important details include weight gain or loss, gastrointestinal bleeding, dietary habits including the timing of meals, smoking, and alcohol consumption. The major esophageal symptoms are heartburn, regurgitation, chest pain, dysphagia, odynophagia, and globus sensation.

Heartburn (pyrosis), the most common esophageal symptom, is characterized by a discomfort or burning sensation behind the sternum that arises from the epigastrium and may radiate toward the neck. Heartburn is usually an intermittent symptom, most commonly experienced after eating, during exercise, and while lying recumbent. The discomfort is relieved with drinking water or taking an antacid but can occur frequently, interfering with normal activities including sleep. The association between heartburn and gastroesophageal reflux disease (GERD) is so strong that empirical therapy for GERD has become accepted management. However, the term *heartburn* is often misused and/or referred to using other terms such as *indigestion* or *repeating*, making it important to clarify the intended meaning.

---

*Deceased.

Regurgitation is the effortless return of food or fluid into the pharynx without nausea or retching. Patients report a sour or burning fluid in the throat or mouth that may also contain undigested food particles. Bending, belching, or maneuvers that increase intraabdominal pressure can provoke regurgitation. A clinician needs to discriminate among regurgitation, vomiting, and rumination. *Vomiting* is preceded by nausea and accompanied by retching. *Rumination* is a behavior in which recently swallowed food is regurgitated and then reswallowed repetitively for up to an hour. Although there is some linkage between rumination and cognitive deficiency, the behavior is also exhibited by unimpaired individuals.

Chest pain is a common esophageal symptom with characteristics similar to cardiac pain, sometimes making this distinction difficult. Esophageal pain is usually experienced as a pressure-type sensation in the mid chest, radiating to the mid back, arms, or jaws. The similarity to cardiac pain is likely because the two organs share a nerve plexus and the nerve endings in the esophageal wall have poor discriminative ability among stimuli. Esophageal distention or even chemostimulation (e.g., with acid) will often be perceived as chest pain. Gastroesophageal reflux is the most common cause of esophageal chest pain.

Esophageal *dysphagia* **(Chap. 47)** is often described as a feeling of food "sticking" or even lodging in the chest. Important distinctions are between uniquely solid food dysphagia as opposed to liquid and solid, episodic versus constant dysphagia, and progressive versus static dysphagia. If the dysphagia is for liquids as well as solid food, it suggests a motility disorder such as achalasia. Conversely, uniquely solid food dysphagia is suggestive of a stricture, web, ring, or tumor. Of note, a patient's localization of food hang-up in the esophagus is notoriously imprecise. Approximately 30% of distal esophageal obstructions are perceived as cervical dysphagia. In such instances, the absence of concomitant symptoms generally associated with oropharyngeal dysphagia such as aspiration, nasopharyngeal regurgitation, cough, drooling, or obvious neuromuscular compromise should suggest an esophageal etiology.

Odynophagia is pain either caused by or exacerbated by swallowing. Although typically considered distinct from dysphagia, odynophagia may manifest concurrently with dysphagia. Odynophagia is more common with pill or infectious esophagitis than with reflux esophagitis and should prompt a search for these entities. When odynophagia does occur in GERD, it is likely related to an esophageal ulcer or extensive erosions.

Globus sensation, also known as globus pharyngeus, is the perception of a lump or fullness in the throat that is felt irrespective of swallowing. Although such patients are frequently referred for an evaluation of dysphagia, globus sensation is often relieved by the act of swallowing. As implied by its alternative name, "globus hystericus," globus sensation often occurs in the setting of anxiety or obsessive-compulsive disorders. Clinical experience teaches that it is often attributable to GERD.

Water brash is excessive salivation resulting from a vagal reflex triggered by acidification of the esophageal mucosa. This is not a common symptom. Afflicted individuals will describe the unpleasant sensation of the mouth rapidly filling with salty thin fluid, often in the setting of concomitant heartburn.

DIAGNOSTIC STUDIES

ENDOSCOPY

Endoscopy, also known as esophagogastroduodenoscopy (EGD), is the most useful test for the evaluation of the proximal gastrointestinal tract. Modern instruments produce high-quality, color images of the esophageal, gastric, and duodenal lumen. Endoscopes also have an instrumentation channel through which biopsy forceps, injection catheters for focal delivery of therapeutic agents, balloon dilators, or devices for hemostasis or removal of mucosal lesions can be used. The key advantages of endoscopy over barium radiography are as follows: (1) increased sensitivity for the detection of mucosal lesions; (2) vastly increased sensitivity for the detection of abnormalities mainly identifiable by color, such as Barrett's metaplasia or vascular lesions; (3) the ability to obtain biopsy specimens for histologic examination of suspected abnormalities; and (4) the ability to dilate strictures during the examination. Submucosal ("third space") endoscopy has emerged as a diagnostic modality for assessment of subepithelial lesions, resection of superficial dysplastic areas, and therapy of esophageal motility disorders. The main disadvantages of endoscopy are low sensitivity for detection of very proximal esophageal strictures or narrow caliber esophagus without focal stricturing, cost, and the need for sedatives or anesthetics.

RADIOGRAPHY

Contrast radiography of the esophagus, stomach, and duodenum can demonstrate reflux of the contrast media, hiatal hernia, mucosal granularity, erosions, ulcerations, and strictures. The sensitivity of radiography compared with endoscopy for detecting reflux esophagitis reportedly ranges from 22 to 95%, with higher grades of esophagitis (i.e., ulceration or stricture) exhibiting greater detection rates. Conversely, the sensitivity of barium radiography for detecting esophageal strictures is greater than that of endoscopy, especially when the study is done in conjunction with a 13-mm barium tablet. Barium studies also provide an assessment of esophageal function and morphology that may be undetected on endoscopy. Tracheoesophageal fistula, altered postsurgical anatomy, and extrinsic esophageal compression are conditions where radiographic imaging complements endoscopic assessment. Hypopharyngeal pathology and disorders of the cricopharyngeus muscle are better appreciated on radiographic examination than with endoscopy, particularly with rapid sequence or video fluoroscopic recording. The major shortcoming of barium radiography is that it rarely obviates the need for endoscopy. Either a positive or a negative study is usually followed by an endoscopic evaluation to obtain biopsies, provide therapy, clarify findings in the case of a positive examination, or add a level of certainty in the case of a negative examination.

ENDOSCOPIC ULTRASOUND

Endoscopic ultrasound (EUS) instruments combine an endoscope with an ultrasound transducer to create a transmural image of the tissue surrounding the endoscope tip. The key advantage of EUS over alternative radiologic imaging techniques is much greater resolution attributable to the proximity of the ultrasound transducer to the area being examined. Available devices can provide either radial imaging (360-degree, cross-sectional) or a curved linear image that can guide fine-needle aspiration of imaged structures such as lymph nodes or tumors. Major esophageal applications of EUS are to stage esophageal cancer, to evaluate dysplasia in Barrett's esophagus, and to assess submucosal lesions.

ESOPHAGEAL MANOMETRY

Esophageal manometry, or motility testing, entails positioning a catheter with multiple pressure sensors within the esophagus and then observing the contractility following test swallows. The upper esophageal sphincter and lower esophageal sphincter (LES) appear as zones of high pressure that relax on swallowing, whereas the intersphincteric esophagus exhibits peristaltic contractions. Manometry is used to diagnose motility disorders (achalasia, diffuse esophageal spasm [DES]) and to assess peristaltic integrity prior to the surgery for reflux disease. Technologic advances have enhanced esophageal manometry as high-resolution esophageal pressure topography **(Fig. 334-1)**. Manometry can also be combined with intraluminal impedance monitoring. Impedance recordings use a series of paired electrodes added to the manometry catheter. Esophageal luminal contents in contact with the electrodes decrease (liquid) or increase (air) the impedance signal, allowing detection of anterograde or retrograde esophageal bolus transit.

FUNCTIONAL LUMEN IMAGING PROBE

The functional lumen imaging probe (FLIP) is a catheter-based technology that utilizes high-resolution impedance planimetry during volume-controlled esophageal distension to measure esophageal cross-sectional area and distensibility (i.e., cross-sectional area in relation to distension pressure) along a 16-cm length of the esophagus. These data

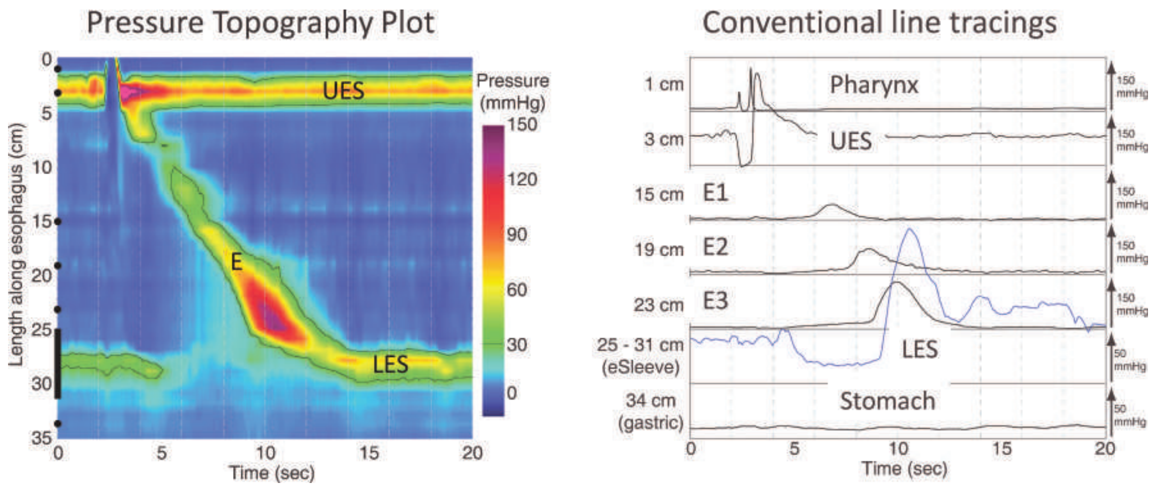

FIGURE 334-1 High-resolution esophageal pressure topography (*left*) and conventional manometry (*right*) of a normal swallow. E, esophageal body; LES, lower esophageal sphincter; UES, upper esophageal sphincter.

are displayed in real time on a computer monitor both as a cylinder of varied diameter along the 16-cm length of the probe or as real-time diameter topography akin to high-resolution manometry. FLIP studies are performed with sedation in conjunction with upper endoscopy or intraoperatively; the probe is passed transorally. Assessing esophageal motility with FLIP is based on quantifying esophagogastric junction opening during volumetric distention and characterizing distension-induced esophageal contractility, which is, in essence, secondary peristalsis. These findings are classified according to a scheme known as FLIP panometry. There is excellent correlation between FLIP panometry and high-resolution manometry in the detection of both achalasia and normal contractility, the two most important potential findings. Hence, FLIP can be used to evaluate patients with esophageal dysphagia, often in lieu of manometry.

REFLUX TESTING

GERD is often diagnosed in the absence of endoscopic signs of esophagitis, which would otherwise define the disease. This occurs in the settings of partially treated disease, an abnormally sensitive esophageal mucosa, or, most commonly, in nonerosive reflux disease. In such instances, reflux testing can demonstrate excessive esophageal exposure to refluxed gastric fluid, the physiologic abnormality of GERD. This can be done by ambulatory 24- to 96-h esophageal pH recording using either a wireless pH-sensitive transmitter that is affixed to the esophageal mucosa or a transnasally positioned wire electrode with the tip stationed in the distal esophagus. Either way, the outcome is expressed as the percentage of the day that the pH was <4 (indicative of recent acid reflux), with values exceeding 6% indicative of GERD. Reflux testing is useful in the evaluation of patients presenting with atypical symptoms or an inexplicably poor response to therapy. Intraluminal impedance monitoring can be added to pH monitoring to detect reflux events irrespective of whether or not they are acidic, potentially increasing the sensitivity of the study.

STRUCTURAL DISORDERS

■ HIATAL HERNIA

Hiatal hernia is a herniation of viscera, most commonly the stomach, into the mediastinum through the esophageal hiatus of the diaphragm. Four types of hiatal hernia are distinguished, with type I, or sliding hiatal hernia, comprising about 95% of the overall total. A sliding hiatal hernia is one in which the gastroesophageal junction and gastric cardia translocate cephalad as a result of weakening of the phrenoesophageal ligament attaching the gastroesophageal junction to the diaphragm at the hiatus and dilatation of the diaphragmatic hiatus. The incidence of sliding hernia increases with age. True to its name, sliding hernias enlarge with increased intraabdominal pressure, swallowing, and respiration. Conceptually, sliding hernias are the result of wear and tear: increased intraabdominal pressure from abdominal obesity, pregnancy, etc., along with hereditary factors predisposing to the condition. The main significance of sliding hernias is the propensity of affected individuals to develop GERD.

Type II, III, and IV hiatal hernias are all subtypes of paraesophageal hernia in which there is herniation into the mediastinum of viscera above the gastric cardia. With type II and III paraesophageal hernias, the gastric fundus herniates, with the distinction being that in type II, the gastroesophageal junction remains fixed at the hiatus, whereas type III is a combined sliding and paraesophageal hernia. With type IV hiatal hernias, viscera other than the stomach herniate into the mediastinum, most commonly the colon. With type II and III paraesophageal hernias, the stomach progressively inverts as it herniates, and large paraesophageal hernias can lead to an "upside down stomach," gastric volvulus, and even strangulation of the stomach. Because of this risk, surgical repair is often advocated for large paraesophageal hernias, particularly when they are symptomatic.

■ RINGS AND WEBS

A lower esophageal mucosal ring, also called a *B ring*, is a thin membranous narrowing at the squamocolumnar mucosal junction (Fig. 334-2).

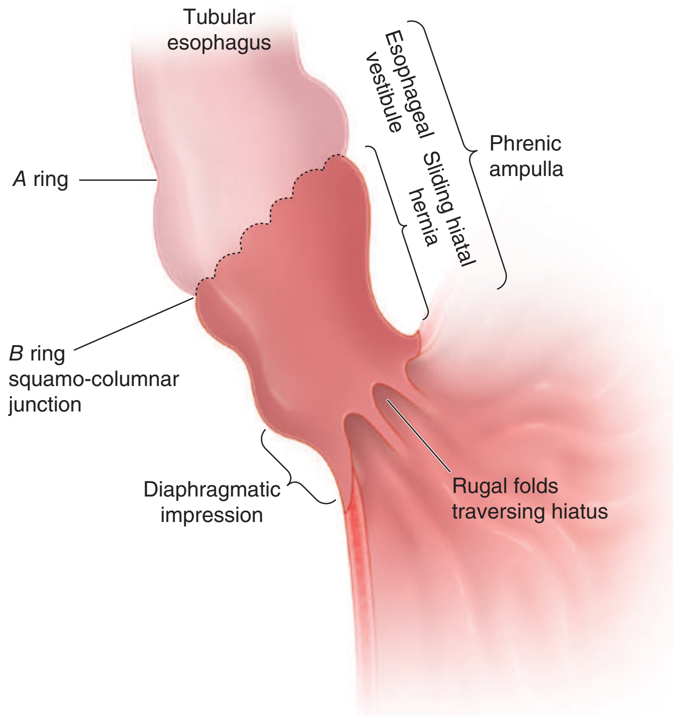

FIGURE 334-2 Radiographic anatomy of the gastroesophageal junction.

Its origin is unknown, but B rings are demonstrable in ~10–15% of the general population and are usually asymptomatic. When the lumen diameter is <13 mm, distal rings are usually associated with episodic solid food dysphagia and are called *Schatzki rings*. Patients typically present older than 40 years, consistent with an acquired rather than congenital origin. Schatzki ring is a common cause of intermittent food impaction, also known as "steakhouse syndrome" because meat is a typical instigator. Symptomatic rings are readily treated by dilation.

Web-like constrictions higher in the esophagus can be of congenital or inflammatory origin. Asymptomatic cervical esophageal webs are demonstrated in ~10% of people and typically originate along the anterior aspect of the esophagus. Depending on the degree of impingement, they can cause intermittent dysphagia to solids similar to Schatzki rings and are similarly treated with dilation. The combination of symptomatic proximal esophageal webs and iron-deficiency anemia in middle-aged women constitutes Plummer-Vinson or Paterson-Kelly syndrome.

■ DIVERTICULA

Esophageal diverticula are categorized by location, with the most common being epiphrenic, hypopharyngeal (Zenker's), and midesophageal. Epiphrenic and Zenker's diverticula are false diverticula involving herniation of the mucosa and submucosa through the muscular layer of the esophagus. These lesions result from increased intraluminal pressure associated with distal obstruction. In the case of Zenker's, the obstruction is a stenotic cricopharyngeus muscle (upper esophageal sphincter), and the hypopharyngeal herniation occurs in an area of natural weakness proximal to the cricopharyngeus known as *Killian's triangle* **(Fig. 334-3)**. Small Zenker's diverticula are usually asymptomatic, but when they enlarge sufficiently to retain food and saliva, they can be associated with dysphagia, halitosis, and aspiration. Treatment is by surgical diverticulectomy and cricopharyngeal myotomy or transoral, endoscopic myotomy.

Epiphrenic diverticula are often associated with achalasia or esophageal hypercontractile disorders. Midesophageal diverticula may be caused by traction from adjacent inflammation (tuberculosis, histoplasmosis), in which case they are true diverticula involving all layers of the esophageal wall, or by pulsion associated with esophageal motility disorders. Midesophageal and epiphrenic diverticula are often asymptomatic; symptoms tend to correlate more with the underlying esophageal disorder. Large epiphrenic diverticula can be removed surgically, usually in conjunction with a myotomy. Diffuse intramural esophageal pseudodiverticulosis is a rare entity resulting from dilatation of the excretory ducts of submucosal esophageal glands **(Fig. 334-4)**. Esophageal candidiasis or strictures are commonly found in association with pseudodiverticulosis.

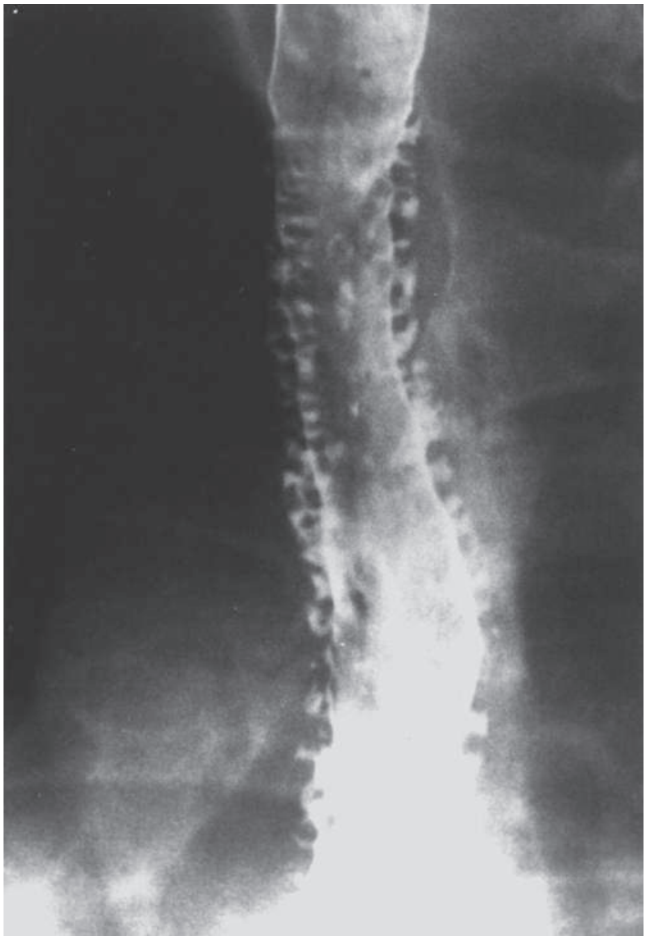

FIGURE 334-4 Intramural esophageal pseudodiverticulosis associated with chronic obstruction. Invaginations of contrast into the esophageal wall outline deep esophageal glands.

■ TUMORS

The American Cancer Society's 2023 estimates for esophageal cancer in the United States are about 21,560 new cases (79% men) and

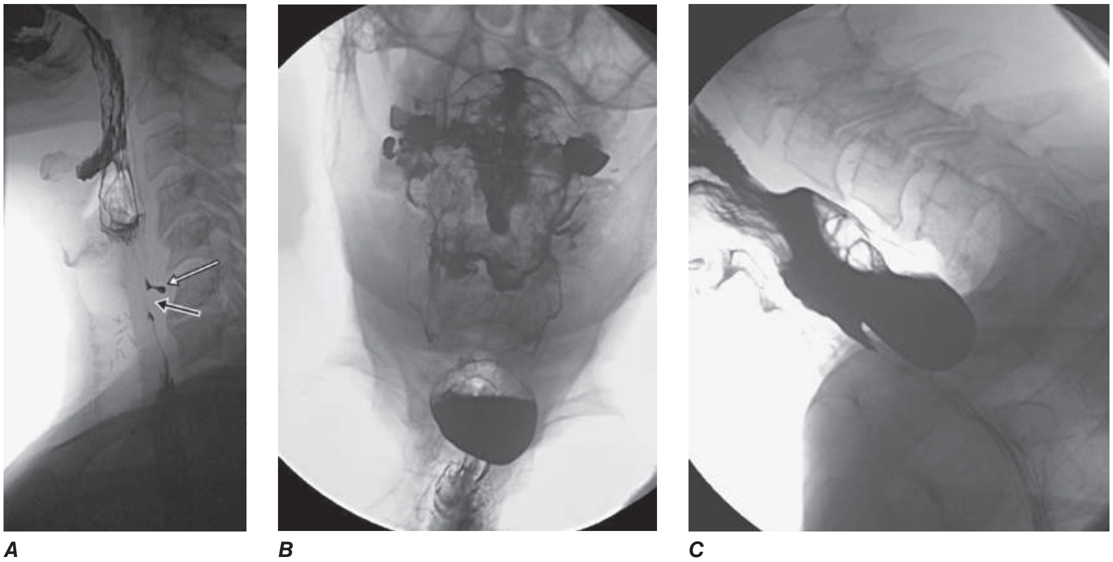

FIGURE 334-3 Examples of small (*A*) and large (*B, C*) Zenker's diverticula arising from Killian's triangle in the distal hypopharynx. Smaller diverticula are evident only during the swallow, whereas larger ones retain food and fluid.

16,120 deaths. It is ~7 times less common than colorectal cancer but kills about a third as many patients, emphasizing both the rarity and lethality of esophageal cancer. One notable trend is the shift of dominant esophageal cancer type from squamous cell to adenocarcinoma, strongly linked to reflux disease and Barrett's metaplasia. Other distinctions between cell types are the predilection for adenocarcinoma to affect the distal esophagus in white males and for squamous cell carcinoma to affect the more proximal esophagus in black males with the added risk factors of smoking, alcohol consumption, caustic injury, and human papillomavirus infection **(Chap. 85)**.

The typical presentation of esophageal cancer is of progressive solid food dysphagia and weight loss. Associated symptoms may include odynophagia, iron deficiency, cough from tracheoesophageal fistula, and hoarseness from left recurrent laryngeal nerve injury. Generally, respiratory symptoms are manifestations of locally invasive or even metastatic disease. Even when detected as a small lesion, esophageal cancer has poor survival because of the abundant esophageal lymphatics leading to early regional lymph node metastases.

Benign esophageal tumors are uncommon and usually discovered incidentally. They include gastrointestinal stromal tumors, leiomyoma, fibrovascular polyps, squamous papilloma, granular cell tumors, lipomas, mesenchymal neoplasms, and inflammatory fibroid polyps.

CONGENITAL ANOMALIES

The most common congenital esophageal anomaly is esophageal atresia, occurring in ~2.5 per 10,000 live births. Atresia can occur in several permutations, the common denominator being developmental failure of fusion between the proximal and distal esophagus associated with a tracheoesophageal fistula, most commonly with the distal segment excluded. Alternatively, there can be an H-type configuration in which esophageal fusion has occurred, but with a tracheoesophageal fistula. Esophageal atresia is usually recognized and corrected surgically within the first few days of life. Later life complications include dysphagia from anastomotic strictures or absent peristalsis and reflux, which can be severe. Less common developmental anomalies include congenital esophageal stenosis, webs, and duplications.

Dysphagia can also result from vascular congenital abnormalities that cause extrinsic compression of the esophagus, a condition called dysphagia lusoria. Most commonly, the esophagus is compressed by an aberrant right subclavian artery arising from the descending aorta and passing behind the esophagus. Alternatively, vascular rings may surround and constrict the esophagus.

Heterotopic gastric mucosa, also known as an esophageal inlet patch, is a focus of gastric-type epithelium in the proximal cervical esophagus; the estimated prevalence is 4–5%. The inlet patch is thought to result from incomplete replacement of embryonic columnar epithelium with squamous epithelium. The majority of inlet patches are asymptomatic, but acid production along with associated symptoms and even ulceration can occur as most contain fundic-type gastric epithelium with parietal cells.

ESOPHAGEAL MOTILITY DISORDERS

Esophageal motility disorders are diseases attributable to abnormal esophageal neuromuscular dysfunction commonly associated with dysphagia, chest pain, or heartburn. The major entities are achalasia, distal esophageal spasm (DES), and GERD. Motility disorders can also be secondary to systemic disease processes, as is the case with pseudoachalasia, Chagas' disease, and scleroderma. Not included in this discussion are diseases affecting the pharynx and proximal esophagus, the impairment of which is almost always part of a more global neuromuscular disease process.

■ ACHALASIA

Achalasia is a rare disease with a population incidence estimated to be 1–3 per 100,000 and presentation usually occurring between age 25 and 60 years. Although there is some degree of heterogeneity in its pathogenesis, the classic cases are caused by autoimmune-mediated death of ganglion cells within the esophageal myenteric plexus; with long-standing disease, aganglionosis is noted. The disease involves both excitatory (cholinergic) and inhibitory (nitric oxide) ganglionic neurons, which mediate deglutitive LES relaxation and the sequential propagation of peristalsis. Their absence leads to impaired deglutitive LES relaxation and absent peristalsis. Increasing evidence suggests that the ultimate cause of ganglion cell degeneration in achalasia is an autoimmune process attributable to a latent infection with human herpes simplex virus 1 combined with genetic susceptibility.

Long-standing achalasia is characterized by progressive dilatation and sigmoid deformity of the esophagus, sometimes associated with hypertrophy of the LES. Clinical manifestations may include dysphagia, regurgitation, chest pain, and weight loss. Most patients report solid and liquid food dysphagia. Regurgitation occurs when food, fluid, and secretions are retained in the dilated esophagus. Patients with advanced achalasia are at risk for bronchitis, pneumonia, or lung abscess from chronic regurgitation and aspiration. Chest pain may manifest early in the course of achalasia. Patients describe a squeezing, pressure-like retrosternal pain, sometimes radiating to the neck, arms, jaw, and back. Paradoxically, some patients complain of heartburn that may be a chest pain equivalent. Treatment of achalasia is less effective at alleviating chest pain than it is in relieving dysphagia or regurgitation.

The differential diagnosis of achalasia includes DES, Chagas' disease, opioid-induced esophageal dysmotility, and pseudoachalasia. Chagas' disease is endemic in areas of central Brazil, Venezuela, and northern Argentina and spread by the bite of the reduviid (kissing) bug that transmits the protozoan *Trypanosoma cruzi.* The chronic phase of the disease develops years after infection and results from destruction of autonomic ganglion cells throughout the body, including the heart, gut, urinary tract, and respiratory tract. Manometric features of achalasia have been described in patients on chronic opioids and may be confused with primary achalasia. Tumor infiltration, most commonly seen with carcinoma in the gastric cardia or distal esophagus, can also mimic primary achalasia. The resultant "pseudoachalasia" accounts for up to 5% of suspected cases and is more likely with advanced age, abrupt onset of symptoms (<1 year), and weight loss. Hence, endoscopy is a necessary part of the evaluation of achalasia. When the clinical suspicion for pseudoachalasia is high and endoscopy is nondiagnostic, computed tomography (CT) scanning or EUS may be of value. Rarely, pseudoachalasia can result from a paraneoplastic syndrome with circulating antineuronal antibodies.

Achalasia is diagnosed by barium swallow x-ray and/or esophageal manometry. Endoscopy excludes tumors or benign mechanical strictures of the esophagogastric junction. The barium swallow x-ray appearance is of a dilated esophagus with poor emptying, an air-fluid level, and tapering at the LES giving it a beak-like appearance **(Fig. 334-5)**. Occasionally, an epiphrenic diverticulum is observed. In long-standing achalasia, the esophagus may assume a sigmoid configuration. The diagnostic criteria for achalasia with esophageal manometry

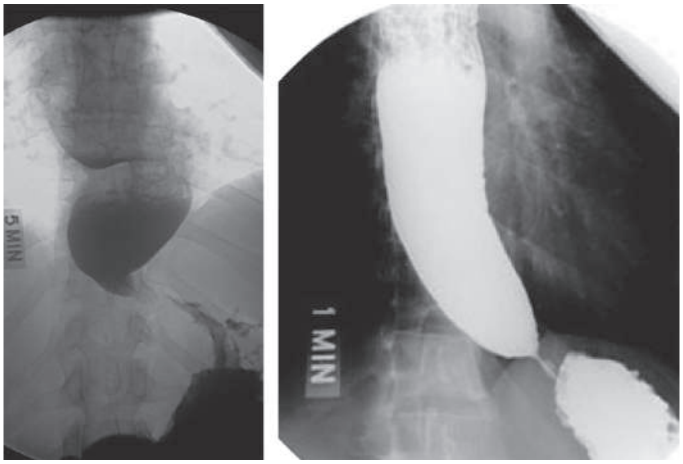

FIGURE 334-5 Achalasia with esophageal dilatation, tapering at the gastroesophageal junction, and an air-fluid level within the esophagus. The example on the *left* shows sigmoid deformity with very advanced disease.

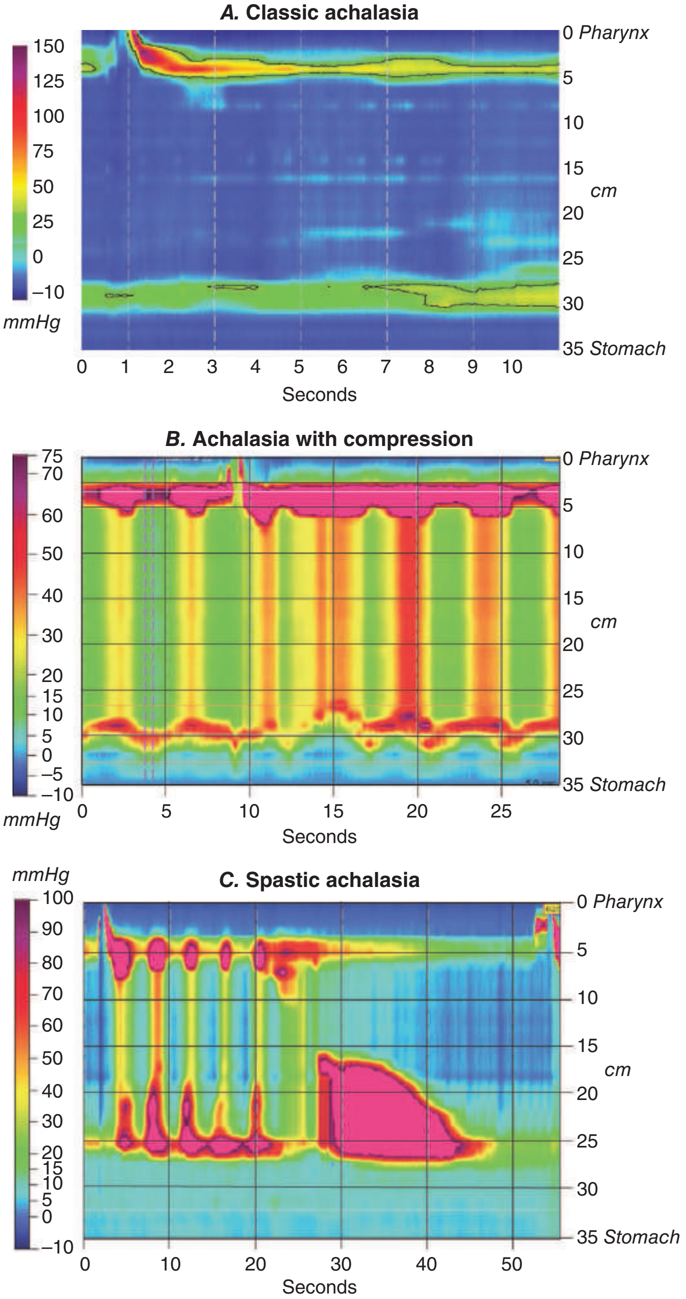

FIGURE 334-6 Three subtypes of achalasia: classic (*A*), with esophageal compression (*B*), and spastic achalasia (*C*) imaged with high-resolution manometry. All are characterized by impaired lower esophageal sphincter (LES) relaxation and absent peristalsis. However, classic achalasia has minimal pressurization of the esophageal body, whereas substantial fluid pressurization is observed in achalasia with esophageal compression, and spastic esophageal contractions are observed with spastic achalasia.

are impaired LES relaxation and absent peristalsis. High-resolution manometry has somewhat advanced this diagnosis; three subtypes of achalasia are differentiated based on the pattern of pressurization in the nonperistaltic esophagus **(Fig. 334-6)**. Because manometry identifies early disease before esophageal dilatation and food retention, it is the most sensitive diagnostic test.

No method of preventing or "curing" achalasia is known. Therapy is thus directed at reducing LES pressure so that gravity and esophageal pressurization permit esophageal emptying. While peristalsis does not recover, remnants of peristalsis masked by esophageal pressurization and dilatation prior to therapy may be demonstrable following effective treatment. LES pressure can be reduced by pharmacologic therapy, pneumatic balloon dilation, or LES myotomy by means of submucosal (third space) endoscopy or laparoscopic surgery. Pharmacologic therapies are relatively ineffective but can be offered as temporizing therapies. Nitrates or calcium channel blockers are administered before eating but should be used with caution because of their effects on blood pressure. Botulinum toxin, injected into the LES under endoscopic guidance, inhibits acetylcholine release from nerve endings and improves dysphagia in about two-thirds of cases for at least 6 months. Sildenafil and alternative phosphodiesterase inhibitors effectively decrease LES pressure, but practicalities limit their clinical use in achalasia.

The only durable therapies for achalasia are pneumatic dilation and LES myotomy. Pneumatic dilation, with a reported efficacy ranging widely from 60 to 90%, is an endoscopic technique using a noncompliant, cylindrical balloon dilator positioned across the LES and inflated to a diameter of 3–4 cm. The major complication is perforation, with a reported incidence of 0.5–5%. The most common surgical procedure for achalasia is laparoscopic Heller myotomy, usually performed in conjunction with an antireflux procedure (partial fundoplication); good to excellent results are reported in 62–90% of cases. A European randomized controlled trial demonstrated an equivalent response rate of ~90% for both pneumatic dilation and laparoscopic Heller myotomy at 5-year follow-up. Occasionally, patients with advanced disease fail to respond to pneumatic dilation or Heller myotomy or relapse years after response to primary therapy. In such refractory cases, esophageal resection with gastric pull-up or interposition of a segment of transverse colon may be the only option other than gastrostomy feeding.

An endoscopic approach to LES myotomy is increasingly available, referred to as peroral endoscopic (or esophageal) myotomy (POEM). This technique involves endoscopically incising the esophageal mucosa and creating a tunnel in the submucosa of the esophageal wall through which the circular muscle of the LES and a calibrated length of distal esophagus are transected with electrocautery. GERD is common after POEM but managed effectively with medications. Potential advantages over the conventional laparoscopic approach include avoidance of surgical disruption of the diaphragmatic hiatus and more rapid recovery. An international, multicenter, randomized trial of POEM and pneumatic dilation demonstrated greater symptom relief with POEM compared to dilation at 2 and 5 years. A European, multicenter, randomized trial of POEM and Heller myotomy reported similar efficacy for symptom relief, exceeding 80% with either.

In untreated or inadequately treated achalasia, esophageal dilatation predisposes to stasis esophagitis. Prolonged stasis esophagitis is the likely explanation for the association between achalasia and esophageal squamous cell cancer. Tumors develop after years of achalasia, usually in the setting of extreme esophageal dilatation, with the overall squamous cell cancer risk increased 17-fold compared to controls.

DISTAL ESOPHAGEAL SPASM

DES is manifest by episodes of dysphagia and chest pain attributable to abnormal esophageal contractions with normal deglutitive LES relaxation. The pathophysiology and natural history of DES are poorly defined. Radiographically, DES has been characterized by tertiary contractions or a "corkscrew esophagus" **(Fig. 334-7)**, but in many instances, these abnormalities are indicative of achalasia. Manometrically, a variety of defining features have been proposed including uncoordinated ("spastic") activity in the distal esophagus, spontaneous and repetitive contractions, or high-amplitude and prolonged contractions. High-resolution manometry has defined DES by the occurrence of contractions in the distal esophagus with short latency relative to the time of the pharyngeal contraction, a dysfunction indicative of impairment of inhibitory myenteric plexus neurons. When defined with this restrictive criterion **(Fig. 334-8)**, DES is substantially less common than achalasia.

Esophageal chest pain closely mimics angina pectoris. Features suggesting esophageal pain include pain that is nonexertional, prolonged, meal-related, relieved with antacids, accompanied by heartburn, dysphagia, or regurgitation, and interrupts sleep. However, all of these features exhibit overlap with cardiac pain, which still must be the primary consideration. Furthermore, even within the spectrum of esophageal diseases, both chest pain and dysphagia are also characteristic of peptic

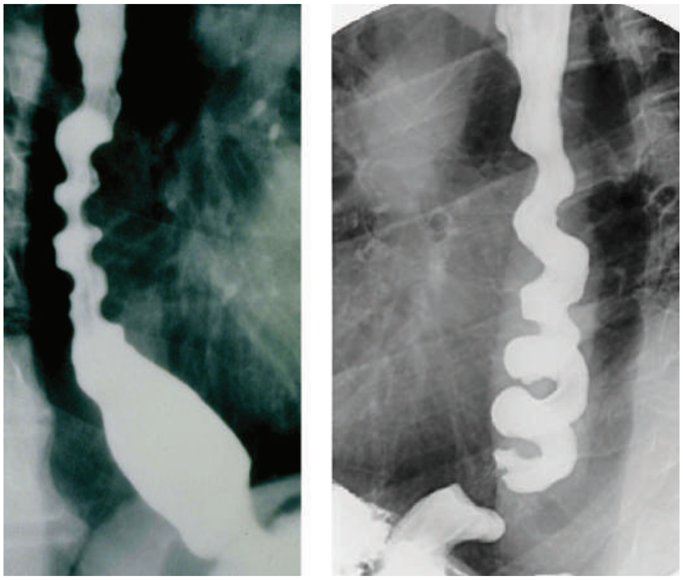

FIGURE 334-7 Distal esophageal spasm. The characteristic "corkscrew" esophagus results from spastic contraction of the circular muscle in the esophageal wall; more precisely, this is actually a helical array of muscle. These findings are also seen with spastic achalasia.

or infectious esophagitis. Only after these more common entities have been excluded by evaluation and/or treatment should a diagnosis of DES be pursued.

Although DES is diagnosed by manometry, endoscopy is useful to identify alternative structural and inflammatory lesions that may cause chest pain. Radiographically, a "corkscrew esophagus," "rosary bead esophagus," pseudodiverticula, or curling can be indicative of DES, but these are also found with spastic achalasia. Given these vagaries of defining DES and the resultant heterogeneity of patients identified for inclusion in therapeutic trials, it is not surprising that trial results have been disappointing. Only small, uncontrolled trials exist, reporting response to nitrates, calcium channel blockers, hydralazine, botulinum toxin, and anxiolytics. POEM with distal esophageal myotomy or surgical myotomy should be considered only with severe weight loss or intractable pain. These indications are extremely rare.

NONSPECIFIC MANOMETRIC FINDINGS

Manometric studies done to evaluate chest pain and/or dysphagia often report minor abnormalities (e.g., hypertensive or hypotensive peristalsis, hypertensive LES) that are insufficient to diagnose either achalasia or DES. These findings are of unclear significance. Reflux and psychiatric diagnoses, particularly anxiety and depression, are common among such individuals. A lower visceral pain threshold and symptoms of irritable bowel syndrome are noted in more than half of such patients. Consequently, therapy for these individuals should target either the most common esophageal disorder, GERD, or cognitive disorders that may be present.

GASTROESOPHAGEAL REFLUX DISEASE

The current concept of GERD is that it encompasses a family of conditions with the commonality that they are caused by gastroesophageal reflux resulting in either troublesome symptoms or an array of potential esophageal and extraesophageal manifestations. It is estimated that 10–15% of adults in the United States are affected by GERD, although such estimates are based on population studies of self-reported chronic heartburn. With respect to the esophagus, the spectrum of injury includes esophagitis, stricture, Barrett's esophagus, and adenocarcinoma **(Fig. 334-9)**. Of particular concern is the rising incidence of esophageal adenocarcinoma, an epidemiologic trend that parallels the increasing incidence of GERD. About 9200 incident cases of esophageal adenocarcinoma were noted in the United States in 2020 (estimated as half of all esophageal cancers); this disease burden has increased sixfold in the past 20 years.

PATHOPHYSIOLOGY

The best-defined subset of GERD patients, albeit a minority overall, have esophagitis. Esophagitis occurs when refluxed gastric acid and pepsin induce inflammation of the esophageal mucosa that leads to microscopic injury and macroscopic erosions or ulcers. Experimental evidence supports a cytokine-mediated inflammatory pathway rather than direct caustic injury to the esophageal epithelium. Note that some degree of gastroesophageal reflux is normal, physiologically intertwined with the mechanism of belching (transient LES relaxation), but esophagitis results from excessive reflux, often accompanied by impaired clearance/neutralization of the refluxed gastric juice. Restricting reflux to that which is physiological levels depends on the anatomic and physiologic integrity of the esophagogastric junction, a complex valvular mechanism functionally dependent on the LES, the surrounding crural diaphragm, and the native subdiaphragmatic end-to-side architecture of the gastroesophageal junction. Pathologic reflux is a consequence of the interplay between progressive anatomic distortion of the native architecture of the gastroesophageal junction and physiology. Relevant factors include (1) widening of the diaphragmatic hiatus and diminished crural diaphragm sphincteric function; (2) loss of an intraabdominal esophageal segment with complete disabling of the native flap valve architecture; (3) axial hiatus hernia, which causes LES hypotension and loss of the ability to prevent reflux during swallow-induced LES relaxation, inspiration, or straining; and (4) increased compliance of gastroesophageal junction leading to wider sphincter opening during transient LES relaxations (the belch reflex) with loss of the ability to prevent liquid escaping along with the gas. Factors tending to exacerbate reflux regardless of mechanism are abdominal obesity, pregnancy, gastric hypersecretory states, delayed gastric emptying, disruption of esophageal peristalsis, and gluttony.

After acid reflux, peristalsis returns the refluxed fluid to the stomach, and acid clearance is completed by titration of the residual acid by bicarbonate contained in swallowed saliva. Consequently, two causes of prolonged acid clearance are impaired peristalsis and reduced salivation. Impaired peristaltic emptying can be attributable to disrupted peristalsis or superimposed reflux associated with a hiatal hernia. With

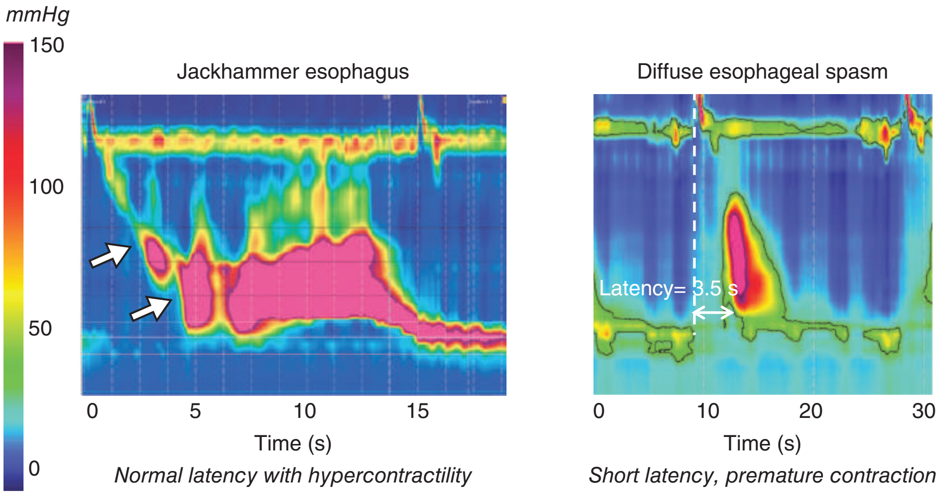

Normal latency with hypercontractility

Short latency, premature contraction

FIGURE 334-8 Esophageal pressure topography of the two major variants of esophageal spasm: hypercontractile esophagus (*left*) and distal esophageal spasm (*right*). Hypercontractile esophagus is defined by the extraordinarily vigorous, sometimes repetitive contractions with normal peristaltic onset and normal latency of the contraction. Distal esophageal spasm is similar but primarily defined by a short latency (premature) contraction.

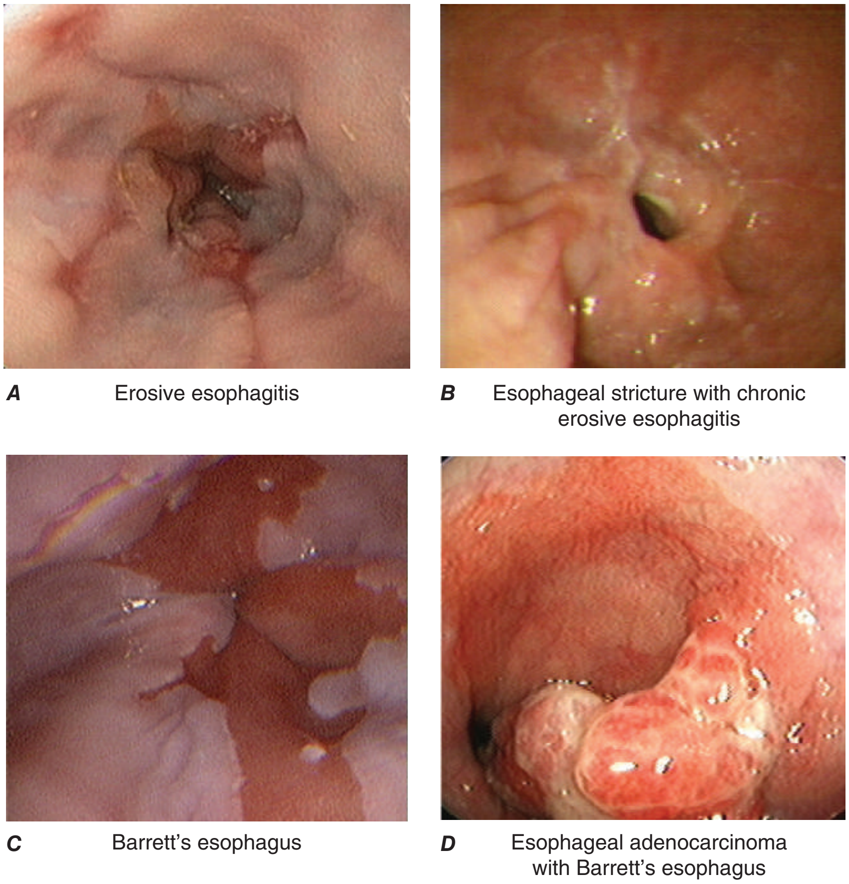

A Erosive esophagitis

B Esophageal stricture with chronic erosive esophagitis

C Barrett's esophagus

D Esophageal adenocarcinoma with Barrett's esophagus

FIGURE 334-9 Endoscopic appearance of (***A***) peptic esophagitis, (***B***) a peptic stricture, (***C***) Barrett's metaplasia, and (***D***) adenocarcinoma developing within an area of Barrett's esophagus.

superimposed reflux, fluid retained within a sliding hiatal hernia refluxes back into the esophagus during swallow-related LES relaxation, a phenomenon that does not normally occur.

Inherent in the pathophysiologic model of GERD is that gastric juice is harmful to the esophageal epithelium. However, gastric acid hypersecretion is usually not a dominant factor in the development of esophagitis. An obvious exception is with Zollinger-Ellison syndrome, which is associated with severe esophagitis in ~50% of patients. Another caveat is with chronic *Helicobacter pylori* gastritis, which may have a protective effect by inducing atrophic gastritis with concomitant hypoacidity. Pepsin, bile, and pancreatic enzymes within gastric secretions can also injure the esophageal epithelium, but their noxious properties are either lessened without an acidic environment or dependent on acidity for activation. Bile warrants attention because it persists in refluxate despite acid-suppressing medications. Bile can traverse the cell membrane, imparting severe cellular injury in a weakly acidic environment, and has also been invoked as a cofactor in the pathogenesis of Barrett's metaplasia and adenocarcinoma. Hence, the causticity of gastric refluxate extends beyond hydrochloric acid.

SYMPTOMS

Heartburn and regurgitation are the typical symptoms of GERD. Somewhat less common are dysphagia and chest pain. In each case, multiple potential mechanisms for symptom genesis operate that extend beyond the basic concepts of mucosal erosion and activation of afferent sensory nerves. Specifically, visceral hypersensitivity is increasingly recognized as a cofactor. Nonetheless, the dominant clinical strategy is empirical treatment with acid inhibitors, reserving further evaluation for those who fail to respond. Important exceptions to this are patients with chest pain or persistent dysphagia, each of which may be indicative of GERD or alternative diagnoses. With chest pain, cardiac disease must be considered. With dysphagia, eosinophilic esophagitis (EoE) should be considered as an alternative diagnosis along with peptic stricture, motility disorders, or esophageal cancer.

Extraesophageal syndromes associated with GERD include chronic cough, laryngitis, asthma, and dental erosions. Other conditions including pharyngitis, chronic bronchitis, pulmonary fibrosis, chronic sinusitis, cardiac arrhythmias, sleep apnea, and recurrent aspiration pneumonia have proposed associations with GERD. However, in both cases, it is important to emphasize the word *association* as opposed to *causation*. In many instances, the disorders likely coexist without causality. Potential mechanisms for extraesophageal GERD manifestations are either regurgitation with direct contact between the refluxate and supraesophageal structures or via a vagovagal reflex wherein reflux activation of esophageal afferent nerves triggers efferent vagal reflexes such as bronchospasm, cough, or arrhythmias.

DIFFERENTIAL DIAGNOSIS

Although generally quite characteristic, symptoms from GERD need to be distinguished from symptoms related to infectious or pill esophagitis, EoE, peptic ulcer disease, dyspepsia, biliary colic, coronary artery disease, and esophageal motility disorders. It is especially important that coronary artery disease be given early consideration because of its potentially lethal implications. The remaining elements of the differential diagnosis can be addressed by endoscopy, upper gastrointestinal series, or esophageal manometry as appropriate. Erosive esophagitis at the esophagogastric junction is the endoscopic hallmark of GERD, but it is identified in only about one-third of GERD patients. The distinction among etiologies of esophagitis is readily made by endoscopic appearance, but mucosal biopsies are necessary to evaluate for infectious or eosinophilic inflammation. In terms of endoscopic appearance, the ulcerations seen in peptic esophagitis are usually few and distal, whereas infectious ulcerations are numerous, punctate, and diffuse. EoE characteristically exhibits multiple esophageal edema, rings, linear furrows, white punctate exudate, and strictures. Esophageal ulcerations from pill esophagitis are usually singular and deep at points of luminal narrowing, especially near the carina, with sparing of the distal esophagus.

COMPLICATIONS

The complications of GERD are related to chronic esophagitis (bleeding and stricture) and the relationship between GERD and esophageal adenocarcinoma. However, both erosive esophagitis and peptic strictures have become increasingly rare in the era of potent antisecretory medications. Conversely, the most severe histologic consequence of GERD is Barrett's metaplasia with the associated risk of esophageal adenocarcinoma, and the incidence of these lesions has increased, not decreased, in the era of potent acid suppression. Barrett's metaplasia, recognized endoscopically by salmon-colored mucosa extending proximally from the gastroesophageal junction (Fig. 334-9) or histopathologically by the finding of intestinal metaplasia, is associated with a significantly increased risk for development of esophageal adenocarcinoma.

Barrett's metaplasia can progress to adenocarcinoma through the intermediate stages of low- and high-grade dysplasia **(Fig. 334-10)**. Owing to this risk, areas of Barrett's metaplasia and especially any included areas of mucosal irregularity should be carefully inspected and extensively biopsied. The rate of cancer development is estimated at 0.1–0.3% per year, but vagaries in definitional criteria and of the extent of Barrett's metaplasia requisite to establish the diagnosis have contributed to variability and inconsistency in this risk assessment. The group at greatest risk is obese white males in their sixth decade of life. However, despite common practice, the utility of endoscopic screening and surveillance programs intended to control the adenocarcinoma risk remains an open question. Also of note, although in a large

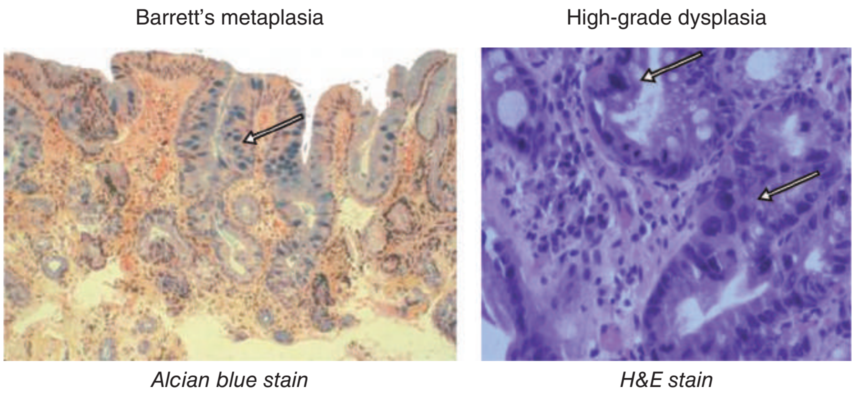
Barrett's metaplasia

High-grade dysplasia

Alcian blue stain

H&E stain

FIGURE 334-10 Histopathology of Barrett's metaplasia and Barrett's metaplasia with high-grade dysplasia (arrows). H&E, hematoxylin and eosin.

randomized, controlled trial of chemoprevention in Barrett's patients, high-dose proton pump inhibitor therapy along with aspirin did significantly better at achieving the composite endpoint of delaying all-cause mortality or Barrett's progression, the effect was driven mainly by improved overall survival rather than reduced Barrett's dysplasia or adenocarcinoma development.

Although the management of Barrett's esophagus remains controversial, the finding of dysplasia in Barrett's, particularly high-grade dysplasia, mandates further intervention. In addition to the high rate of progression to adenocarcinoma, there is also a high prevalence of unrecognized coexisting cancer with high-grade dysplasia. Treatment recommendations for Barrett's esophagus with high-grade dysplasia have evolved over the past several years. Historically, esophagectomy was the gold standard treatment for high-grade dysplasia. However, esophagectomy has a mortality ranging from 3 to 10%, along with substantial morbidity. Prospective studies have demonstrated the efficacy of endoscopic mucosal ablation therapy with substantially less morbidity and essentially no mortality. Consequently, current societal guidelines endorse endoscopic mucosal ablation therapies for the management of high-grade dysplasia.

TREATMENT

Gastroesophageal Reflux Disease

Lifestyle modifications are routinely advocated as GERD therapy. Broadly speaking, these fall into three categories: (1) avoidance of foods that reduce LES pressure, making them "refluxogenic" (these commonly include fatty foods, alcohol, spearmint, peppermint, and possibly coffee and tea); (2) avoidance of acidic foods that are inherently irritating (citrus fruits, tomato-based foods); and (3) adoption of behaviors to minimize reflux and/or heartburn. In general, minimal evidence supports the efficacy of these measures. However, clinical experience dictates that subsets of patients benefit from specific recommendations based on their individual history and symptom profile. A patient with sleep disturbance from nighttime heartburn is more likely to benefit from elevation of the head of the bed and avoidance of eating before retiring. The most broadly applicable recommendation is for weight reduction. Even though the benefit with respect to reflux cannot be assured, the strong epidemiologic relationship between body mass index and GERD and the secondary health gains of weight reduction is beyond dispute.

The dominant pharmacologic approach to GERD management is with inhibitors of gastric acid secretion, and abundant data support the effectiveness of this approach. Pharmacologically reducing the acidity of gastric juice does not prevent reflux, but it ameliorates reflux symptoms and allows esophagitis to heal. The hierarchy of effectiveness among pharmaceuticals for healing esophagitis parallels their antisecretory potency. Potassium competitive acid blockers (PCABs) are more efficacious than proton pump inhibitors (PPIs), which are more efficacious than histamine-2 receptor antagonists (H_2RAs). The differences are most evident with severe esophagitis, much less so with mild disease.

Paradoxically, the perceived frequency and severity of heartburn correlate poorly with the presence or severity of esophagitis. When GERD treatments are assessed in terms of resolving heartburn, both efficacy and differences among pharmaceuticals are less clear-cut than with the objective of healing esophagitis. Although the same overall hierarchy of effectiveness exists between PPIs and H_2RAs (minimal data are available for PCABs), observed efficacy rates are lower and vary widely, likely reflecting patient heterogeneity.

Reflux symptoms tend to be chronic, irrespective of esophagitis. Thus, a common management strategy is indefinite treatment with PPIs or H_2RAs as necessary for symptom control. The side effects of PPI therapy are generally minimal. Rare cases of interstitial nephritis and severe, reversible hypomagnesemia have been reported. Vitamin B_{12} and iron absorption may be compromised and susceptibility to enteric infections, particularly *Clostridioides difficile* colitis, increased with treatment. Observational data have also noted an association between PPI exposure and renal disease, dementia, and cardiovascular disease, but the hazard ratios reported in these studies were small, and the potential for unrecognized residual confounding bias was substantial. Population studies have also suggested a slight increased risk of bone fracture with chronic PPI use suggesting an impairment of calcium absorption, but prospective studies have failed to corroborate this. Nonetheless, as with any medication, PPI dosage should be minimized to that necessary for the clinical indication.

Laparoscopic fundoplication, wherein the proximal stomach is wrapped around the distal esophagus to create an antireflux barrier, is a surgical alternative to the management of chronic GERD. Just as with PPI therapy, evidence on the utility of fundoplication is strongest for treating esophagitis, and controlled trials suggest similar efficacy to PPI therapy. However, the benefits of fundoplication must be weighed against potential deleterious effects, including surgical morbidity and mortality, postoperative dysphagia, failure or breakdown requiring reoperation, an inability to belch, and increased bloating, flatulence, and bowel symptoms after surgery.

■ EOSINOPHILIC ESOPHAGITIS

EoE is increasingly recognized in adults and children around the world. Current prevalence estimates in the United States identified 4–8 cases per 10,000 with a predilection for white males between 30 and 40 years of age. The increasing prevalence of EoE is attributable to a combination of an increasing incidence and a growing recognition of the condition. There is also an incompletely understood, but important, interaction between EoE and GERD that may confound the diagnosis of the disease. Genome-wide analysis studies demonstrated susceptibility elements at 5q22 (thymic stromal lymphopoietin) and 2p23 (CAPN14) in EoE.

EoE is diagnosed by the combination of esophageal symptoms and esophageal mucosal biopsies demonstrating eosinophil-predominant inflammation. Alternative etiologies of esophageal eosinophilia include GERD, drug hypersensitivity, connective tissue disorders, hypereosinophilic syndrome, Crohn's disease, eosinophilic gastroenteritis, and infection. EoE is an immunologic disorder induced by antigen sensitization in susceptible individuals. Food allergens are the dominant triggers, although aeroallergens may also contribute. The natural history of EoE is incompletely understood, but an increased risk of esophageal stricture development paralleling the duration of untreated disease has been noted.

EoE should be strongly considered in children and adults with dysphagia and esophageal food impactions. In preadolescent children, symptom presentations of EoE include chest or abdominal pain, nausea, vomiting, and food aversion. Other symptoms in adults may include atypical chest pain and heartburn. An atopic history of IgE-mediated food allergy, asthma, eczema, and/or allergic rhinitis is present in the majority of patients. Peripheral blood eosinophilia is demonstrable in 25–50% of patients, but the specificity of this finding is problematic in the setting of concomitant atopy. The characteristic endoscopic esophageal findings include loss of vascular markings (edema), multiple esophageal rings, longitudinally oriented furrows, whitish exudate, and strictures (Fig. 334-11). Histologic confirmation is made with the demonstration of esophageal mucosal eosinophilia (peak density ≥15 eosinophils per high-power field) (Fig. 334-12). Complications of EoE in adolescents and adults include food impaction, esophageal stricture, narrow-caliber esophagus, and (rarely) esophageal perforation. In children, complications include feeding difficulties and weight loss.

Since the complications of EoE are a consequence of ongoing eosinophilic inflammation, preventing them is dependent on follow up endoscopy with biopsies to verify a selected therapy's effectiveness in controlling eosinophilic inflammation. Primary therapy often starts with a PPI, which is effective at improving eosinophilic inflammation in 30–50% of patients. Additional first-line therapies include elimination diets or swallowed topical glucocorticoids. Elemental formula diets devoid of allergenic protein are a highly effective therapy but are limited by palatability. Notably, allergy testing by means of either serum IgE or skin prick testing has demonstrated poor sensitivity and specificity in the identification of foods responsible for EoE in an individual patient. Empiric elimination of common food allergies (milk, wheat, egg, soy, nuts, and seafood) followed by systematic reintroduction has been an effective diet therapy in both children and adults. The intent of the elimination diet approach is the identification and long-term avoidance of specific food trigger(s). Swallowed, topical glucocorticoids (e.g., fluticasone propionate or budesonide) are effective in 50–80% of patients. Systemic glucocorticoids are not generally recommended due to side effects and lack of proven benefit beyond that achieved with topical glucocorticoids. The first U.S. Food and Drug Administration–approved treatment for EoE is the interleukin 4 and 13–blocking biologic dupilumab, injected subcutaneously once a week. Additional therapies targeting other immune pathways and allergic cytokine mediators have also shown promise in initial clinical trials. Esophageal dilation is highly effective at relieving dysphagia in patients with fibrostenosis but does not address the underlying inflammatory process. Dilation should be approached cautiously because of the risk of deep, esophageal mural laceration or perforation in the characteristic stiff-walled esophagus. Once an effective therapy is identified, maintenance therapy is advocated due to the chronicity of EoE and nearly universal disease recurrence upon cessation of therapy.

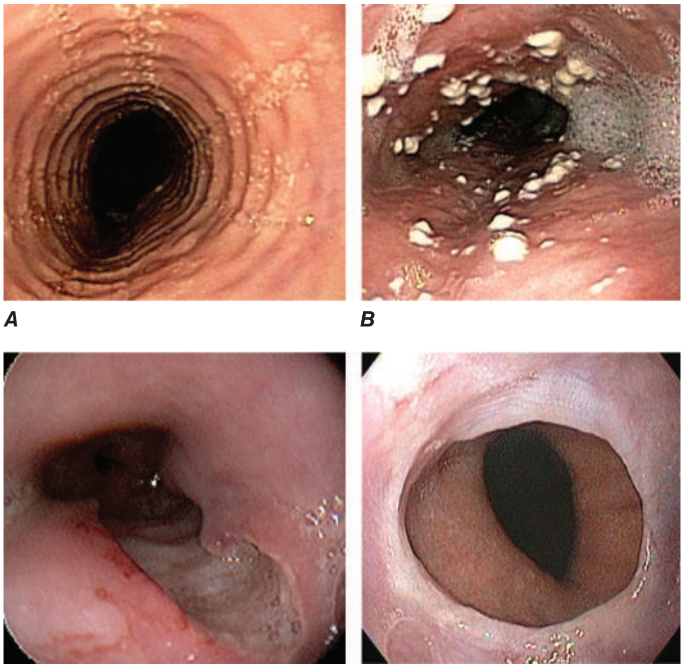

FIGURE 334-11 Endoscopic features of (*A*) eosinophilic esophagitis (EoE), (***B***) *Candida* esophagitis, (***C***) giant ulcer associated with HIV, and (***D***) a Schatzki ring.

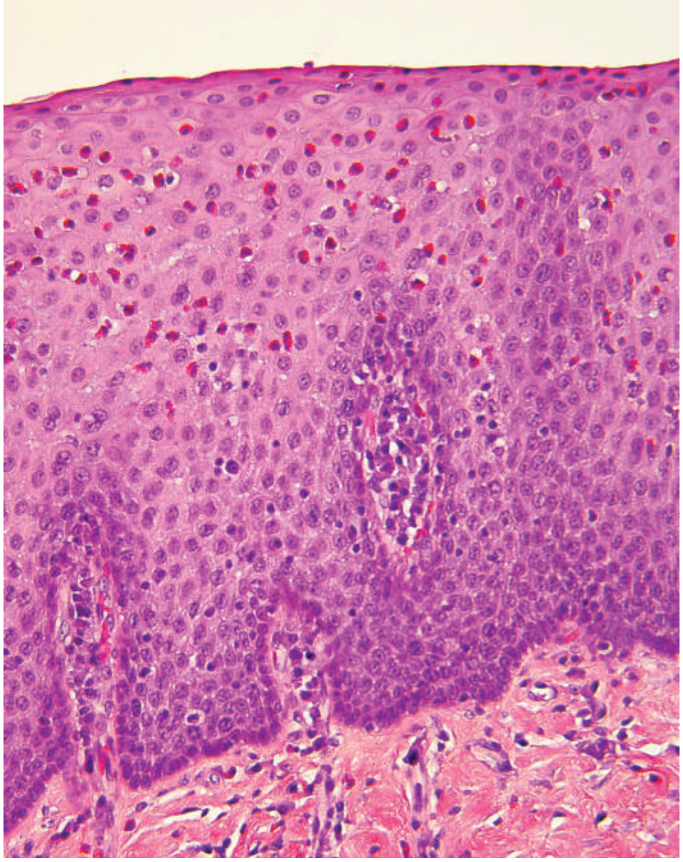

FIGURE 334-12 Histopathology of eosinophilic esophagitis (EoE) showing infiltration of the esophageal squamous epithelium with eosinophils. Additional features of basal cell hyperplasia and lamina propria fibrosis are present. Eosinophilic inflammation can also be seen with gastroesophageal reflux disease.

INFECTIOUS ESOPHAGITIS

With the increased use of immunosuppression for organ transplantation and chronic inflammatory diseases, use of chemotherapy, and the AIDS epidemic, infections with *Candida* species, herpesvirus, and cytomegalovirus (CMV) have become relatively common. Although rare, infectious esophagitis also occurs among the non-immunocompromised, with herpes simplex and *Candida albicans* being the most common pathogens. Among AIDS patients, infectious esophagitis becomes more common as the CD4 count declines; cases are rare with a CD4 count >200 and common when <100. HIV itself may also be associated with a self-limited syndrome of acute esophageal ulceration with oral ulcers and a maculopapular skin rash at the time of seroconversion. Additionally, some patients with advanced disease have deep, persistent esophageal ulcers treated with oral glucocorticoids or thalidomide. However, with the widespread use of highly effective antiviral therapies, these HIV complications have become less common.

Regardless of the infectious agent, odynophagia is a characteristic symptom of infectious esophagitis; dysphagia, chest pain, and hemorrhage are also common. Odynophagia is uncommon with reflux esophagitis, so its presence should always raise suspicion of an alternative etiology.

■ *CANDIDA* ESOPHAGITIS

Candida is normally found in the throat but can become pathogenic and produce esophagitis in a compromised host; *C. albicans* is most common. *Candida* esophagitis also occurs with esophageal stasis secondary to esophageal motor disorders and diverticula. Patients complain of odynophagia and dysphagia. If oral thrush is present, empirical therapy is appropriate, but co-infection is common, and persistent symptoms should lead to prompt endoscopy with biopsy, which is the most useful diagnostic evaluation. *Candida* esophagitis has a characteristic appearance of white plaques or exudate with friability. Oral fluconazole (400 mg on the first day, followed by 200 mg daily) for 7–14 days is the preferred initial treatment. Patients refractory to that may respond to higher dose fluconazole, voriconazole, itraconazole, or posaconazole. Alternatively, poorly responsive patients or those who cannot swallow medications can be treated with an intravenous echinocandin or amphotericin B.

■ HERPETIC ESOPHAGITIS

Herpes simplex virus type 1 or 2 may cause esophagitis. Vesicles on the nose and lips may coexist and are suggestive of a herpetic etiology. Varicella-zoster virus can also cause esophagitis in children with chickenpox or adults with zoster. The characteristic endoscopic findings are vesicles and small, superficial ulcerations. Because herpes simplex infections are limited to squamous epithelium, biopsies from the ulcer margins are most likely to reveal the characteristic ground-glass nuclei, eosinophilic Cowdry's type A inclusion bodies, and giant cells. Culture or polymerase chain reaction (PCR) assays are helpful to identify acyclovir-resistant strains. Acyclovir (200 mg orally five times a day for 7–10 days) can be used for immunocompetent hosts, although the disease is typically self-limited after a 1- to 2-week period in such patients. Immunocompromised patients are treated with acyclovir (400 mg orally five times a day for 14–21 days), famciclovir (500 mg orally three times a day), or valacyclovir (1 g orally three times a day). In patients with severe odynophagia, intravenous acyclovir, 5 mg/kg every 8 h for 7–14 days, reduces this morbidity.

■ CYTOMEGALOVIRUS

CMV esophagitis occurs primarily in immunocompromised patients, particularly those with HIV, malignancy, and recipients of bone marrow or organ transplants. CMV is usually reactivated from a latent infection. Endoscopically, CMV lesions appear as large serpiginous ulcers in an otherwise normal mucosa, particularly in the distal esophagus. Biopsies from the ulcer bases have the greatest diagnostic yield for finding the pathognomonic large nuclear or cytoplasmic inclusion bodies. Immunohistology with monoclonal antibodies to CMV and in situ hybridization tests are useful for early diagnosis. Data on therapy for CMV esophagitis are limited. Treatment studies of CMV gastrointestinal disease have demonstrated effectiveness of both ganciclovir (5 mg/kg every 12 h IV) and valganciclovir (900 mg orally every 12 h). Therapy is continued until healing, which may take 3–6 weeks. Maintenance therapy may be needed for patients with relapsing disease.

MECHANICAL TRAUMA AND IATROGENIC INJURY

■ ESOPHAGEAL PERFORATION

Most cases of esophageal perforation are from instrumentation of the esophagus or trauma. Alternatively, forceful vomiting or retching can lead to spontaneous rupture at the gastroesophageal junction (Boerhaave's syndrome). More rarely, corrosive esophagitis or neoplasms lead to perforation. Instrument perforation from endoscopy or nasogastric tube placement typically occurs in the hypopharynx or at the gastroesophageal junction. Perforation may also occur at the site of a stricture in the setting of endoscopic food disimpaction or therapeutic esophageal dilation. Esophageal perforation causes pleuritic retrosternal pain often associated with pneumomediastinum and subcutaneous emphysema. Mediastinitis is a major complication of esophageal perforation, and prompt recognition is key to optimizing outcome. CT of the chest is most sensitive in detecting mediastinal air. Esophageal perforation is confirmed by a contrast swallow, usually Gastrografin followed by thin barium. Treatment includes nasogastric suction and parenteral broad-spectrum antibiotics with prompt surgical drainage and repair in noncontained leaks. Conservative therapy with nasogastric suction, NPO status, and antibiotics without surgery may be appropriate in cases of contained perforation that are detected early. Endoscopic clipping or stent placement may be indicated in nonoperated iatrogenic perforations or nonoperable cases such as perforated tumors.

■ MALLORY-WEISS TEAR

Vomiting, retching, or vigorous coughing can cause a nontransmural tear at the gastroesophageal junction that is a common cause of upper gastrointestinal bleeding. Most patients present with hematemesis. Antecedent vomiting is the norm but not always evident. Bleeding usually abates spontaneously, but protracted bleeding may respond to local epinephrine or cauterization therapy, endoscopic clipping, or angiographic embolization. Surgery is rarely needed.

■ RADIATION ESOPHAGITIS

Radiation esophagitis can complicate treatment for thoracic cancers, especially breast and lung cancers, with the risk proportional to radiation dosage. Radiosensitizing drugs such as doxorubicin, bleomycin, cyclophosphamide, and cisplatin increase the risk. Dysphagia and odynophagia may last weeks to months after therapy. The esophageal mucosa becomes erythematous, edematous, and friable. Submucosal fibrosis and degenerative tissue changes and stricturing may occur years after the radiation exposure. Radiation exposure in excess of 5000 cGy has been associated with increased risk of esophageal stricture. Treatment for acute radiation esophagitis is supportive. Chronic strictures are managed with esophageal dilation.

■ CORROSIVE ESOPHAGITIS

Caustic esophageal injury from ingestion of alkali or, less commonly, acid can be accidental or from attempted suicide. Absence of oral injury does not exclude possible esophageal involvement. Thus, early endoscopic evaluation is recommended to assess and grade injury to the esophageal mucosa. Severe corrosive injury may lead to esophageal perforation, bleeding, stricture, and death. Glucocorticoids have not been shown to improve the clinical outcome of acute corrosive esophagitis and are not recommended. Healing of more severe grades of caustic injury is commonly associated with severe stricture formation and often requires life-long repeated dilation.

■ PILL ESOPHAGITIS

Pill-induced esophagitis occurs when a swallowed pill fails to traverse the entire esophagus and lodges within the lumen. Generally, this is attributed to poor "pill-taking habits": inadequate liquid with the pill or lying down immediately after taking a pill. The most common location for the pill to lodge is in the mid-esophagus near the crossing of the aorta or carina. Extrinsic compression from these structures halts the movement of the pill or capsule. Since initially reported in 1970, thousands of cases of pill esophagitis have been reported, suggesting that this is not an unusual occurrence. A wide variety of medications are implicated, with the most common being doxycycline, tetracycline, quinidine, phenytoin, potassium chloride, ferrous sulfate, nonsteroidal anti-inflammatory drugs (NSAIDs), and bisphosphonates.

Typical symptoms of pill esophagitis are the sudden onset of chest pain and odynophagia. Characteristically, the pain will develop over a period of hours or will awaken the individual from sleep. A classic history in the setting of ingestion of a recognized pill offender obviates the need for diagnostic testing in most patients. When endoscopy is performed, localized ulceration or inflammation is evident. Histologically, acute inflammation is typical. Chest CT imaging will sometimes reveal esophageal thickening consistent with transmural inflammation. Although the condition usually resolves within days to weeks, symptoms may persist for months and stricture can develop in severe cases. No specific therapy is known to hasten the healing process, but antisecretory medications are frequently prescribed to remove concomitant reflux as an aggravating factor. When healing results in stricture formation, dilation is indicated.

FOREIGN BODIES AND FOOD IMPACTION

Food or foreign bodies may lodge in the esophagus, causing complete obstruction, which in turn can cause an inability to handle secretions (foaming at the mouth) and severe chest pain. Food impaction may occur due to peptic stricture, carcinoma, Schatzki ring, EoE, achalasia, or simply inattentive eating. If it does not resolve spontaneously, impacted food should be removed endoscopically. Use of meat tenderizer enzymes to facilitate passage of a meat bolus is discouraged because of potential esophageal injury. Glucagon (1 mg IV) is sometimes tried before endoscopic dislodgement. After emergent treatment, patients should be evaluated for potential causes of the impaction with treatment rendered as indicated.

ESOPHAGEAL MANIFESTATIONS OF SYSTEMIC DISEASE

SCLERODERMA AND CONNECTIVE TISSUE DISORDERS

Scleroderma esophagus (hypotensive LES and absent esophageal contractility) was initially described as a manifestation of scleroderma or other collagen vascular diseases and thought to be specific for these disorders. However, this nomenclature subsequently has been discarded because an estimated half of qualifying patients do not have an identifiable rheumatologic disease, and reflux disease is often the only identifiable association. When scleroderma esophagus occurs as a manifestation of a connective tissue disorder, the histopathologic findings are of infiltration and destruction of the esophageal muscularis propria with collagen deposition and fibrosis and reduction in the number of interstitial cells of Cajal. The pathogenesis of absent peristalsis and LES hypotension in the absence of a connective tissue disorder is unknown. Regardless of the underlying cause, the manometric abnormalities predispose patients to severe GERD due to inadequate LES barrier function combined with poor esophageal clearance of refluxed acid. Dysphagia may also be manifest but is generally mild and alleviated by eating in an upright position and using liquids to facilitate solid transit.

DERMATOLOGIC DISEASES

A host of dermatologic disorders (lichen planus, pemphigus vulgaris, bullous pemphigoid, cicatricial pemphigoid, Behçet's syndrome, and epidermolysis bullosa) can affect the oropharynx and esophagus, particularly the proximal esophagus, with blisters, bullae, ulceration, webs, and strictures. Topical or systemic anti-inflammatory therapy is effective for mucosal healing. Stevens-Johnson syndrome and graft-versus-host disease can also involve the esophagus. Esophageal dilation may be necessary to treat strictures.

FURTHER READING

Hirano I et al: American Gastroenterological Institute and the joint task force on allergy-immunology practice parameters clinical guidelines for the management of eosinophilic esophagitis. Gastroenterology 158:1776, 2020.

Kahrilas PJ et al: Advances in the management of oesophageal motility disorders in the era of high-resolution manometry: A focus on achalasia syndromes. Nat Rev Gastroenterol Hepatol 15:323, 2018.

Katzka DA, Kahrilas PJ: Advances in the diagnosis and management of gastroesophageal reflux disease. BMJ 23:371, 2020.

Katzka DA et al: Phenotypes of gastroesophageal reflux disease: Where Rome, Lyon, and Montreal meet. Clin Gastroenterol Hepatol 18:767, 2020.

Shaheen NJ et al: Diagnosis and management of Barrett's esophagus: An updated ACG guideline. Am J Gastroenterol 117:559, 2022.

Straumann A, Katzka DA: Diagnosis and treatment of eosinophilic esophagitis. Gastroenterology 154:346, 2018.

Von Arnim U et al: Monitoring patients with eosinophilic esophagitis in routine clinical practice – international expert recommendations. Clin Gastroenterol Hepatol 21:2526, 2023.

335

Peptic Ulcer Disease and Related Disorders

John Del Valle

PEPTIC ULCER DISEASE

A *peptic ulcer* is defined as disruption of the mucosal integrity of the stomach and/or duodenum leading to a local defect or excavation due to active inflammation. Although burning epigastric pain exacerbated by fasting and improved with meals is a symptom complex associated with peptic ulcer disease (PUD), it is now clear that >90% patients with this symptom complex (dyspepsia) do not have ulcers and that the majority of patients with peptic ulcers may be asymptomatic. Ulcers occur within the stomach and/or duodenum and are often chronic in nature. Acid peptic disorders are very common in the United States, with 4 million individuals (new cases and recurrences) affected per year. Lifetime overall prevalence of PUD in the United States is ~8.4% with a slightly higher prevalence in men. PUD significantly affects quality of life by impairing overall patient well-being and contributing substantially to work absenteeism. Moreover, an estimated 15,000 deaths per year occur as a consequence of complicated PUD. The financial impact of these common disorders has been substantial, with an estimated burden on direct and indirect health care costs of ~$6 billion per year in the United States, with $3 billion spent on hospitalizations, $2 billion on physician office visits, and $1 billion in decreased productivity and days lost from work.

GASTRIC PHYSIOLOGY

Gastric Anatomy The gastric epithelial lining consists of rugae that contain microscopic gastric pits, each branching into four or five gastric glands made up of highly specialized epithelial cells. The makeup of gastric glands varies with their anatomic location. Glands within the gastric cardia comprise <5% of the gastric gland area and contain mucous and endocrine cells. The 75% of gastric glands are found within the oxyntic mucosa and contain mucous neck, parietal, chief, endocrine, enterochromaffin, and enterochromaffin-like (ECL) cells (Fig. 335-1). Highly specialized tuft cells are located in the neck region of the gastric gland. These specialized cells are thought to sample luminal contents, which in turn may be important in regulating gastric acid secretion. Pyloric glands contain mucous and endocrine cells (including gastrin cells) and are found in the antrum.

The parietal cell, also known as the oxyntic cell, is usually found in the neck or isthmus or in the oxyntic gland. The resting, or unstimulated, parietal cell has prominent cytoplasmic tubulovesicles and intracellular canaliculi containing short microvilli along its apical surface (Fig. 335-2). H^+,K^+-adenosine triphosphatase (ATPase) is expressed in the tubulovesicle membrane; upon cell stimulation, this membrane, along with apical membranes, transforms into a dense network of apical intracellular canaliculi containing long microvilli. Acid secretion, a process requiring high energy, occurs at the apical canalicular surface. Numerous mitochondria (30–40% of total cell volume) generate the energy required for secretion.

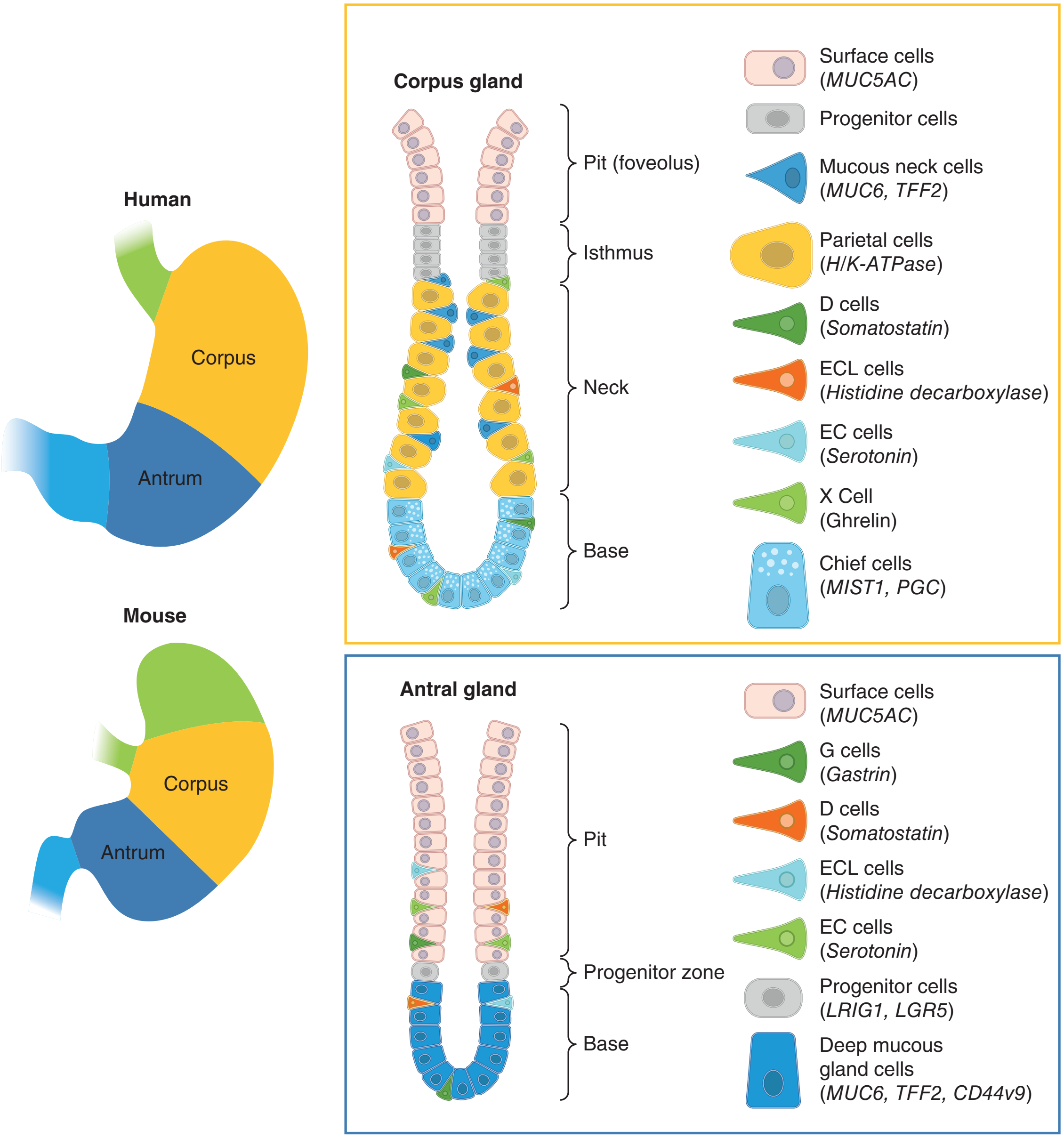
Human
Corpus
Antrum
Mouse
Corpus
Antrum
Corpus gland
Pit (foveolus)
Isthmus
Neck
Base
Surface cells (MUC5AC)
Progenitor cells
Mucous neck cells (MUC6, TFF2)
Parietal cells (H/K-ATPase)
D cells (Somatostatin)
ECL cells (Histidine decarboxylase)
EC cells (Serotonin)
X Cell (Ghrelin)
Chief cells (MIST1, PGC)
Antral gland
Pit
Progenitor zone
Base
Surface cells (MUC5AC)
G cells (Gastrin)
D cells (Somatostatin)
ECL cells (Histidine decarboxylase)
EC cells (Serotonin)
Progenitor cells (LRIG1, LGR5)
Deep mucous gland cells (MUC6, TFF2, CD44v9)

FIGURE 335-1 Diagrammatic representation of the oxyntic gastric gland. Cellular constituents of the stomach. The human stomach is divided into three components the cardia, the corpus and the antrum. The oxyntic gland which contains the majority of the acid producing cells are located in the corpus. Glands located in the antrum secrete predominantly mucus and also contain the important gastrin producing cells. It is also important to note that key progenitor cells are located within both the gastric corpus and antral glands. *(Reproduced with permission from AC Engevik et al: The physiology of the gastric parietal cell. Physiol Rev 100:573, 2020, Figure 1.)*

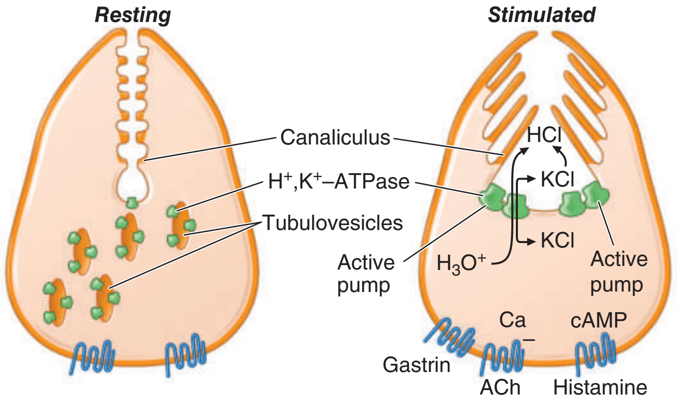
Resting
Stimulated
Canaliculus
H+,K+–ATPase
Tubulovesicles
Active pump
HCl
KCl
KCl
H3O+
Active pump
Ca
cAMP
Gastrin
ACh
Histamine

FIGURE 335-2 Gastric parietal cell undergoing transformation after secretagogue-mediated stimulation. cAMP, cyclic adenosine monophosphate. *(Reproduced with permission from SJ Hersey, G Sachs: Gastric acid secretion. Am Physiol Soc 75:155, 1995.)*

Gastroduodenal Mucosal Defense The gastric epithelium is under constant assault by a series of endogenous noxious factors, including hydrochloric acid (HCl), pepsinogen/pepsin, and bile salts. In addition, a steady flow of exogenous substances such as medications, alcohol, and bacteria encounter the gastric mucosa. A highly intricate biologic system is in place to provide defense from mucosal injury and to repair any injury that may occur.

The mucosal defense system can be envisioned as a three-level barrier, composed of preepithelial, epithelial, and subepithelial elements (**Fig. 335-3**). The first line of defense is a mucus-bicarbonate-phospholipid layer, which serves as a physicochemical barrier to multiple molecules, including hydrogen ions. Mucus is secreted in a regulated fashion by gastroduodenal surface epithelial cells. It consists primarily of water (95%) and a mixture of phospholipids and glycoproteins (mucin). The mucous gel functions as a nonstirred water layer impeding diffusion of ions and molecules such as pepsin. Bicarbonate, secreted in a regulated manner by surface epithelial cells of the gastroduodenal mucosa into the mucous gel, forms a pH gradient ranging

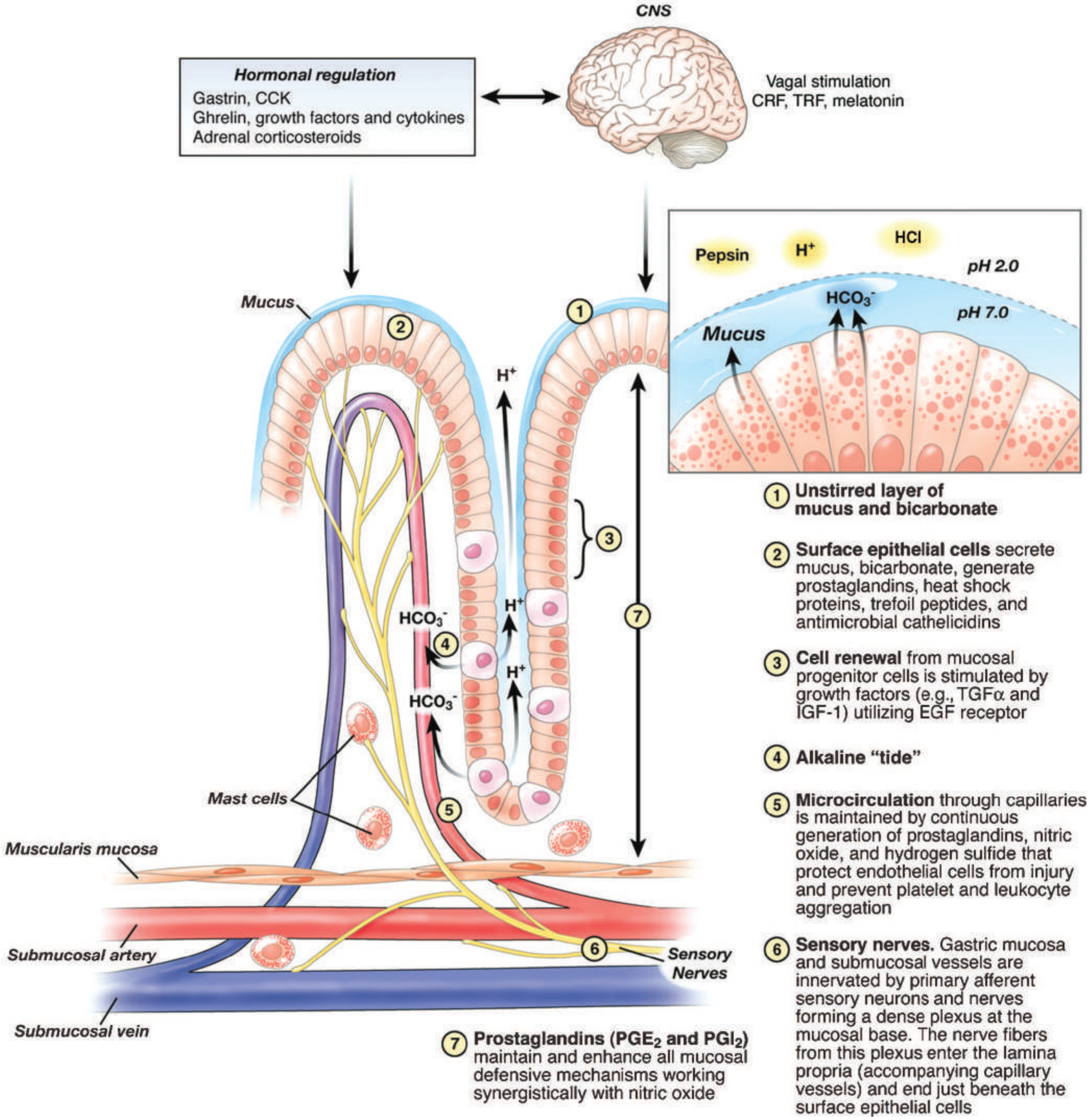
CNS
Hormonal regulation
Gastrin, CCK
Ghrelin, growth factors and cytokines
Adrenal corticosteroids
Vagal stimulation
CRF, TRF, melatonin
Pepsin
H+
HCl
pH 2.0
HCO3-
pH 7.0
Mucus
Mucus
① Unstirred layer of mucus and bicarbonate
② Surface epithelial cells secrete mucus, bicarbonate, generate prostaglandins, heat shock proteins, trefoil peptides, and antimicrobial cathelicidins
③ Cell renewal from mucosal progenitor cells is stimulated by growth factors (e.g., TGFα and IGF-1) utilizing EGF receptor
④ Alkaline "tide"
⑤ Microcirculation through capillaries is maintained by continuous generation of prostaglandins, nitric oxide, and hydrogen sulfide that protect endothelial cells from injury and prevent platelet and leukocyte aggregation
⑥ Sensory nerves. Gastric mucosa and submucosal vessels are innervated by primary afferent sensory neurons and nerves forming a dense plexus at the mucosal base. The nerve fibers from this plexus enter the lamina propria (accompanying capillary vessels) and end just beneath the surface epithelial cells
⑦ Prostaglandins (PGE2 and PGI2) maintain and enhance all mucosal defensive mechanisms working synergistically with nitric oxide
H+
HCO3-
H+
H+
HCO3-
Mast cells
Muscularis mucosa
Submucosal artery
Submucosal vein
Sensory Nerves

FIGURE 335-3 Components involved in providing gastroduodenal mucosal defense and repair. CCK, cholecystokinin; CRF, corticotropin-releasing factor; EGF, epidermal growth factor; HCl, hydrochloride; IGF, insulin-like growth factor; TGFα, transforming growth factor α; TRF, thyrotropin releasing factor. *(Republished with permission of John Wiley and Son's Inc, from Bioregulation and Its Disorders in the Gastrointestinal Tract, T Yoshikawa, T Arakawa [eds]: 1998; permission conveyed through Copyright Clearance Center, Inc.)*

from 1 to 2 at the gastric luminal surface and reaching 6–7 along the epithelial cell surface.

Surface epithelial cells provide the next line of defense through several factors, including mucus production, epithelial cell ionic transporters that maintain intracellular pH and bicarbonate production, and intracellular tight junctions. Surface epithelial cells generate heat shock proteins that prevent protein denaturation and protect cells from certain factors such as increased temperature, cytotoxic agents, or oxidative stress. Epithelial cells also generate trefoil factor family peptides and cathelicidins, which also play a role in surface cell protection and regeneration. If the preepithelial barrier is breached, gastric epithelial cells bordering a site of injury can migrate to restore a damaged region (*restitution*). This process occurs independent of cell division and requires uninterrupted blood flow and an alkaline pH in the surrounding environment. Several growth factors, including epidermal growth factor (EGF), transforming growth factor (TGF) α, and basic fibroblast growth factor (FGF), modulate the process of restitution. Larger defects that are not effectively repaired by restitution require cell proliferation. Epithelial cell regeneration is regulated by prostaglandins and growth factors such as EGF and TGF-α. In tandem with epithelial cell renewal, formation of new vessels (*angiogenesis*) within the injured microvascular bed occurs. Both FGF and vascular endothelial growth factor (VEGF) are important in regulating angiogenesis in the gastric mucosa. In addition, the gastric peptide gastrin (see below) has been found to stimulate cell proliferation, migration, invasion, angiogenesis, and autophagy. Finally, gastric parietal cells (see below) express sonic hedgehog, a family of proteins important in regulating cell lineage in multiple organs. This latter finding suggests that parietal cells may also have the ability to regulate gastric stem cells.

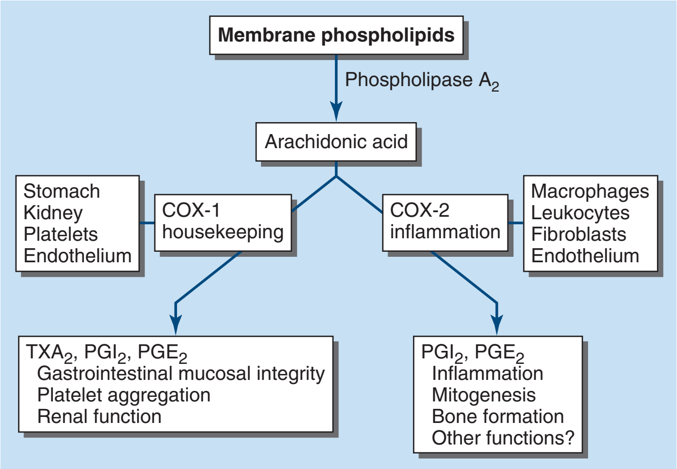

FIGURE 335-4 Schematic representation of the steps involved in synthesis of prostaglandin E_2 (PGE_2) and prostacyclin (PGI_2). Characteristics and distribution of the cyclooxygenase (COX) enzymes 1 and 2 are also shown. TXA_2, thromboxane A_2.

An elaborate microvascular system within the gastric submucosal layer is the key component of the subepithelial defense/repair system, providing HCO_3^-, which neutralizes the acid generated by the parietal cell. Moreover, this microcirculatory bed provides an adequate supply of micronutrients and oxygen while removing toxic metabolic byproducts. Several locally produced factors including nitric oxide (NO) (see below), hydrogen sulfide, and prostacyclin contribute to the vascular protective pathway through vasodilation of the microcirculation.

Prostaglandins play a central role in gastric epithelial defense/repair **(Fig. 335-4)**. The gastric mucosa contains abundant levels of prostaglandins that regulate the release of mucosal bicarbonate and mucus, inhibit parietal cell secretion, and are important in maintaining mucosal blood flow and epithelial cell restitution. Prostaglandins are derived from esterified arachidonic acid, which is formed from phospholipids (cell membrane) by the action of phospholipase A_2. A key enzyme that controls the rate-limiting step in prostaglandin synthesis is cyclooxygenase (COX), which is present in two isoforms (COX-1, COX-2), each having distinct characteristics regarding structure, tissue distribution, and expression. COX-1 is expressed in a host of tissues, including the stomach, platelets, kidneys, and endothelial cells. This isoform is expressed in a constitutive manner and plays an important role in maintaining the integrity of renal function, platelet aggregation, and gastrointestinal (GI) mucosal integrity. In contrast, the expression of COX-2 is inducible by inflammatory stimuli, and it is expressed in macrophages, leukocytes, fibroblasts, and synovial cells. The beneficial effects of nonsteroidal anti-inflammatory drugs (NSAIDs) on tissue inflammation are due to inhibition of COX-2; the toxicity of these drugs (e.g., GI mucosal ulceration and renal dysfunction) is related to inhibition of the COX-1 isoform. The highly COX-2–selective NSAIDs have the potential to provide the beneficial effect of decreasing tissue inflammation while minimizing toxicity in the GI tract. Selective COX-2 inhibitors have had adverse effects on the cardiovascular (CV) system, leading to increased risk of myocardial infarction. Therefore, the U.S. Food and Drug Administration (FDA) has removed two of these agents (valdecoxib and rofecoxib) from the market (see below).

NO is important in the maintenance of gastric mucosal integrity. The key enzyme NO synthase is constitutively expressed in the mucosa and contributes to cytoprotection by stimulating gastric mucus, increasing mucosal blood flow, and maintaining epithelial cell barrier function. The central nervous system (CNS) and hormonal factors also play a role in regulating mucosal defense through multiple pathways (Fig. 335-3).

Since the discovery of *Helicobacter pylori* and its impact on gastric pathology, it has become clear that the stomach has an elaborate and complex inherent immunologic system in place. Although a detailed description of the gastric immune system is beyond the scope of this chapter, several features are worth highlighting. The gastric immune response to certain pathogens such as *H. pylori* (see below) involves extensive interplay between innate (dendritic cells, epithelial cells, neutrophils, and macrophages) and adaptive (B and T cells) components. Helper T cells (T_H and T_H regulatory cells) have been extensively studied and appear to play an important role in a broad array of gastric physiology extending from gastric secretion to epithelial cell turnover via production of a number of cytokines.

The discovery of *H. pylori* has also led to the understanding that the stomach, once thought to be devoid of microorganisms due to its highly adverse environment (acid and pepsin), can serve as host for bacterial communities consisting of hundreds of phylotypes, otherwise known as its microbiota. The conceptual framework of the microbiome has been receiving extensive attention in light of its importance in human health and disease. The overall relevance of the gastric microbiome and its impact on gastric pathology remain to be established, but it is likely that alteration of microorganism homeostasis will play a role in aspects of certain disorders such as PUD, gastritis, and gastric cancer.

Physiology of Gastric Secretion Understanding of the physiology of gastric secretion is important when considering the pathophysiology of PUD and the therapeutic options available. A detailed review of the many components of gastric acid secretion is beyond the scope of this chapter, and the reader is referred to Engevik et al (see Further Reading) for a comprehensive discussion on this very important topic. For the purposes of this chapter, a brief review of acid secretion will follow. HCl and pepsinogen are the two principal gastric secretory products capable of inducing mucosal injury. Gastric acid and pepsinogen play a physiologic role in protein digestion; absorption of iron, calcium, magnesium, and vitamin B_{12}; and killing ingested bacteria. Acid secretion should be viewed as occurring under basal and stimulated conditions. Basal acid production occurs in a circadian pattern, with the highest levels occurring during the night and lowest levels during the morning hours. Cholinergic input via the vagus nerve and histaminergic input from local gastric sources are the principal contributors to basal acid secretion. Stimulated gastric acid secretion occurs primarily in three phases based on the site where the signal originates (cephalic, gastric, and intestinal). Sight, smell, and taste of food are the components of the cephalic phase, which stimulates gastric secretion via the vagus nerve. The gastric phase is activated once food enters the stomach. This component of secretion is driven by nutrients (amino acids and amines) that directly (via peptone and amino acid receptors) and indirectly (via stimulation of intramural gastrin-releasing peptide neurons) stimulate the G cell to release gastrin, which in turn activates the parietal cell via direct and indirect mechanisms. Distention of the stomach wall also leads to gastrin release and acid production. The last phase of gastric acid secretion is initiated as food enters the intestine and is mediated by luminal distention and nutrient assimilation. A series of pathways that inhibit gastric acid production are also set into motion during these phases. The GI hormone somatostatin is released from endocrine cells found in the gastric mucosa (D cells) in response to HCl. Somatostatin can inhibit acid production by both direct (parietal cell) and indirect mechanisms (decreased histamine release from ECL cells, ghrelin release from Gr cells, and gastrin release from G cells). Additional neural (central and peripheral) and humoral (amylin, atrial natriuretic peptide [ANP], cholecystokinin, ghrelin, interleukin 11 [IL-11], obestatin, secretin, and serotonin) factors play a role in counterbalancing acid secretion. Under physiologic circumstances, these phases occur simultaneously. Ghrelin, the appetite-regulating hormone expressed in Gr cells in the stomach, and its related peptide motilin (released from the duodenum) may increase gastric acid secretion through stimulation of histamine release from ECL cells, but this remains to be confirmed.

The acid-secreting parietal cell is located in the oxyntic gland, adjacent to other cellular elements (ECL cell, D cell) important in the gastric secretory process **(Fig. 335-5)**. This unique cell also secretes intrinsic factor (IF) and IL-11. The parietal cell expresses receptors for several stimulants of acid secretion, including histamine (H_2), gastrin (cholecystokinin 2/gastrin receptor), and acetylcholine (muscarinic, M_3). Binding of histamine to the H_2 receptor and activation of the

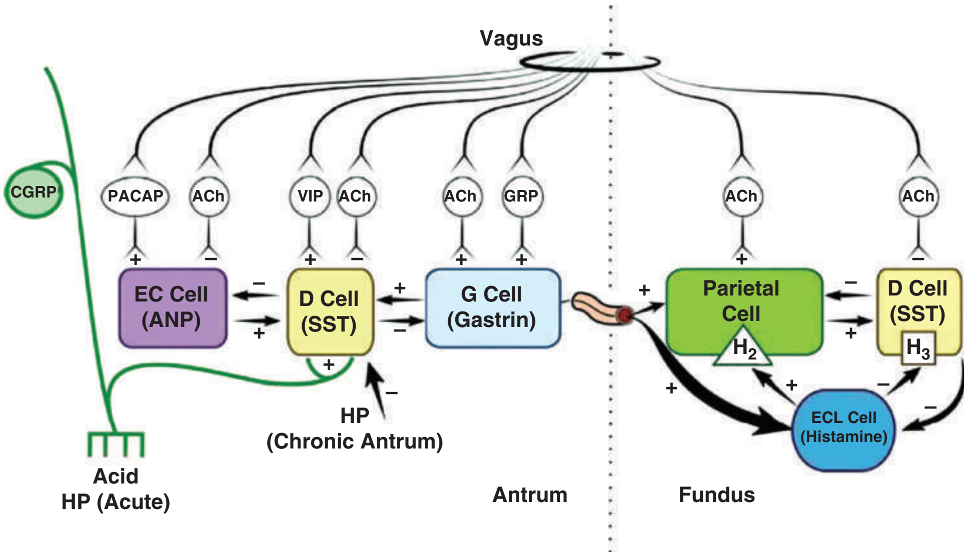

FIGURE 335-5 Regulation of gastric acid secretion at the cellular level. ACh, acetylcholine; ANP, atrial natriuretic peptide; CGRP, calcitonin gene-related peptide; EC, enterochromaffin; ECL, enterochromaffin-like; GRP, gastrin-releasing peptide; PACAP, pituitary adenylate-cyclase activating peptide; SST, somatostatin; VIP, vasoactive intestinal peptide.

gastrin and muscarinic receptors result in stimulation of downstream signaling pathways, which in turn regulates the acid-secreting pump, H^+,K^+-ATPase. Parietal cells also express receptors for ligands that inhibit acid production (glucagon-like peptide 1, prostaglandins, somatostatin, EGF, neurotensin, and urocortin). Histamine also stimulates gastric acid secretion indirectly by activating the histamine H_3 receptor on D cells, which inhibits somatostatin release.

The enzyme H^+,K^+-ATPase is responsible for generating the large concentration of H^+ ions. This enzyme uses the chemical energy of adenosine triphosphate (ATP) to transfer H^+ ions from parietal cell cytoplasm to the secretory canaliculi in exchange for K^+. The H^+,K^+-ATPase is located within the secretory canaliculus and in nonsecretory cytoplasmic tubulovesicles. The tubulovesicles are impermeable to K^+, which leads to an inactive pump in this location. The distribution of pumps between the nonsecretory vesicles and the secretory canaliculus varies according to parietal cell activity (Fig. 335-2). Proton pumps are recycled back to the inactive state in cytoplasmic vesicles once parietal cell activation ceases. In addition, acid secretion requires a number of apical and basolateral parietal cell membrane chloride and potassium channels. Parietal cells also express members of the sonic hedgehog (Shh) family proteins, which play an important role in regulating cell types in multiple organs. This family of proteins may also regulate cell differentiation as well as restitution of mucosal defense in the gastric epithelium.

The chief cell, found primarily in the gastric fundus, synthesizes and secretes pepsinogen, the inactive precursor of the proteolytic enzyme pepsin. The acid environment within the stomach leads to cleavage of the inactive precursor to pepsin and provides the low pH (<2) required for pepsin activity. Pepsin activity is significantly diminished at a pH of 4 and irreversibly inactivated and denatured at a pH of ≥7. Many of the secretagogues that stimulate acid secretion also stimulate pepsinogen release. The precise role of pepsin in the pathogenesis of PUD remains to be established.

■ PATHOPHYSIOLOGIC BASIS OF PUD

PUD encompasses both gastric ulcers (GUs) and duodenal ulcers (DUs). *Ulcers* are defined as breaks in the mucosal surface >5 mm in size, with depth to the submucosa. DUs and GUs share many common features in terms of pathogenesis, diagnosis, and treatment, but several factors distinguish them from one another. *H. pylori* and NSAIDs are the most common risk factors for PUD, with estimated odds ratios in the United States of 3.7 and 3.3, respectively. Additional risk factors (odds ratio) include chronic obstructive lung disease (2.34), chronic renal insufficiency (2.29), current tobacco use (1.99), former tobacco use (1.55), older age (1.67), three or more doctor visits in a year (1.49), coronary heart disease (1.46), former alcohol use (1.29), African-American race (1.20), obesity (1.18), and diabetes (1.13). Selective serotonin reuptake inhibitors (SSRIs) and gastric bypass surgery are also associated with an increased incidence of PUD. A rise in idiopathic PUD has also been noted. The mechanisms by which some of these risk factors lead to ulcer disease are highlighted below.

Epidemiology • **DUODENAL ULCERS** DUs are estimated to occur in 6–15% of the Western population. The incidence of DUs declined steadily from 1960 to 1980 and has remained stable since then. The death rates, need for surgery, and physician visits have decreased by >50% over the past 30 years. The reason for the reduction in the frequency of DUs is likely related to the decreasing frequency of *H. pylori* in turn associated with overall improved sanitary conditions across the world. Before the discovery of *H. pylori*, the natural history of DUs was typified by frequent recurrences after initial therapy. Eradication of *H. pylori* has reduced these recurrence rates by >80%.

GASTRIC ULCERS GUs tend to occur later in life than duodenal lesions, with a peak incidence reported in the sixth decade. More than one-half of GUs occur in males and are less common than DUs, perhaps due to the higher likelihood of GUs being silent and presenting only after a complication develops. Autopsy studies suggest a similar incidence of DUs and GUs.

Pathology • **DUODENAL ULCERS** DUs occur most often in the first portion of the duodenum (>95%), with ~90% located within 3 cm of the pylorus. They are usually ≤1 cm in diameter but can occasionally reach 3–6 cm (giant ulcer). Ulcers are sharply demarcated, with depth at times reaching the muscularis propria. The base of the ulcer often consists of a zone of eosinophilic necrosis with surrounding fibrosis. Malignant DUs are extremely rare.

GASTRIC ULCERS In contrast to DUs, GUs can represent a malignancy and should be biopsied upon discovery. Benign GUs are most often found distal to the junction between the antrum and the acid secretory mucosa. Benign GUs are quite rare in the gastric fundus and are histologically similar to DUs. Benign GUs associated with *H. pylori* are also associated with antral gastritis. In contrast, NSAID-related GUs are not accompanied by chronic active gastritis but may instead have evidence of a chemical gastropathy, typified by foveolar hyperplasia, edema of the lamina propria, and epithelial regeneration in the absence of *H. pylori*. Extension of smooth-muscle fibers into the upper portions of the mucosa, where they are not typically found, may also occur.

Pathophysiology • **DUODENAL ULCERS** *H. pylori* and NSAID-induced injuries account for the majority of DUs. Many acid secretory abnormalities have been described in DU patients. Of these, average basal and nocturnal gastric acid secretion appears to be increased in DU patients as compared to controls; however, the level of overlap between DU patients and control subjects is substantial. The reason for this altered secretory process is unclear, but *H. pylori* infection may contribute. Bicarbonate secretion is significantly decreased in the duodenal bulb of patients with an active DU as compared to control subjects. *H. pylori* infection may also play a role in this process (see below).

GASTRIC ULCERS As in DUs, the majority of GUs can be attributed to either *H. pylori* or NSAID-induced mucosal damage. Prepyloric GUs or those in the body associated with a DU or a duodenal scar are similar in pathogenesis to DUs. Gastric acid output (basal and stimulated) tends to be normal or decreased in GU patients. When GUs develop in the presence of minimal acid levels, impairment of mucosal defense factors may be present.

***H. PYLORI* AND ACID PEPTIC DISORDERS** Although gastric infection with the bacterium *H. pylori* was once thought to account for the majority of PUD **(Chap. 168)**, more recent studies suggest that only one-fourth of DUs and one-sixth of GUs were associated with *H. pylori* infection. These changes are likely due to successful rates of eradication of the organism coupled with improved sanitary conditions. It appears that NSAIDs have become the most common cause of PUD (see below). This organism also plays a role in the development of gastric mucosa-associated lymphoid tissue (MALT) lymphoma and gastric adenocarcinoma. Although the entire genome of *H. pylori* has been sequenced, it is still not clear how this organism, which resides in the stomach, causes ulceration in the duodenum. *H. pylori* eradication efforts may lead to a decrease in gastric cancer in high-risk populations, particularly in individuals who have not developed chronic atrophic gastritis and gastric metaplasia.

The Bacterium The bacterium, initially named *Campylobacter pyloridis*, is a gram-negative microaerophilic rod found most commonly in the deeper portions of the mucous gel coating the gastric mucosa or between the mucous layer and the gastric epithelium. It may attach to gastric epithelium but under normal circumstances does not appear to invade cells. It is strategically designed to live within the aggressive environment of the stomach. It is S-shaped (~0.5–3 μm in size) and contains multiple sheathed flagella. Initially, *H. pylori* resides in the antrum but, over time, migrates toward the more proximal segments of the stomach. The organism is capable of transforming into a coccoid form, which represents a dormant state that may facilitate survival in adverse conditions. The genome of *H. pylori* (1.65 million base pairs) encodes ~1500 proteins. Among this multitude of proteins, some factors are essential determinants of *H. pylori*–mediated pathogenesis and colonization such as the outer membrane protein (Hop proteins), urease, and the vacuolating cytotoxin (Vac A). Moreover, the majority of *H. pylori* strains contain a genomic fragment that encodes the cag pathogenicity island (cag-PAI). Several of the genes that make up cag-PAI encode components of a type IV secretion island that translocates Cag A into host cells. Once in the cell, Cag A activates a series of cellular events important in cell growth and cytokine production. *H. pylori* also has extensive genetic diversity that in turn enhances its ability to promote disease. The first step in infection by *H. pylori* is dependent on the bacteria's motility and its ability to produce urease. Urease produces ammonia from urea, an essential step in alkalinizing the surrounding pH. Additional bacterial factors include but are not limited to, catalase, lipase, adhesins, platelet-activating factor, and pic B (induces cytokines). Multiple strains of *H. pylori* exist and are characterized by their ability to express several of these factors (Cag A, Vac A, etc.). It is possible that the different diseases related to *H. pylori* infection can be attributed to different strains of the organism with distinct pathogenic features.

Epidemiology The prevalence of *H. pylori* varies throughout the world and depends largely on the overall standard of living in the region, with overall global prevalence decreasing. The global prevalence of *H. pylori* infection in adults has declined from 50–55% to 43% during 2014 to 2020. Contributing factors to these changes likely include improvement of socioeconomic status and living standards as well as enhanced hygiene conditions. It is also possible that the increased use of antibiotics specifically with eradication therapies in individuals with infection could be a contributing factor to the overall decrease in global prevalence with a high variability in *H. pylori* infection throughout the world. In developing parts of the world, 80% of the population may be infected by the age of 20, whereas the prevalence is 20–50% in industrialized countries. In contrast, in the United States, this organism is rare in childhood. The overall prevalence of *H. pylori* in the United States is ~30%, with individuals born before 1950 having a higher rate of infection than those born later. About 10% of Americans <30 years of age are colonized with the bacteria. The rate of infection with *H. pylori* in industrialized countries has decreased substantially in recent decades. The steady increase in the prevalence of *H. pylori* noted with increasing age is due primarily to a cohort effect, reflecting higher transmission during a period in which the earlier cohorts were children. It has been calculated through mathematical models that improved sanitation during the latter half of the nineteenth century dramatically decreased transmission of *H. pylori*. Moreover, with the present rate of intervention, the organism ultimately will be eliminated from the United States. Two factors that predispose to higher colonization rates include poor socioeconomic status and less education about the organism. These factors, not race, are responsible for the rate of *H. pylori* infection in blacks and Hispanic Americans being double the rate seen in whites of comparable age. Other risk factors for *H. pylori* infection are (1) birth or residence in a developing country, (2) domestic crowding, (3) unsanitary living conditions, (4) unclean food or water, and (5) exposure to gastric contents of an infected individual.

Transmission of *H. pylori* occurs from person to person, following an oral-oral or fecal-oral route. The risk of *H. pylori* infection is declining in developing countries. The rate of infection in the United States has fallen by >50% when compared to 30 years ago.

Pathophysiology *H. pylori* infection is virtually always associated with a chronic active gastritis, but only 10–15% of infected individuals develop frank peptic ulceration. The basis for this difference is unknown but is likely due to a combination of host and bacterial factors, some of which are outlined below. Initial studies suggested that >90% of all DUs were associated with *H. pylori*, but *H. pylori* is present in only one-sixth of individuals with GUs and one-fourth of patients with DUs. The pathophysiology of ulcers not associated with *H. pylori* or NSAID ingestion (or the rare Zollinger-Ellison syndrome [ZES]) is becoming more relevant as the incidence of *H. pylori* is dropping, particularly in the Western world (see below).

The particular end result of *H. pylori* infection (gastritis, PUD, gastric MALT lymphoma, gastric cancer) is determined by a complex interplay between bacterial and host factors **(Fig. 335-6)**.

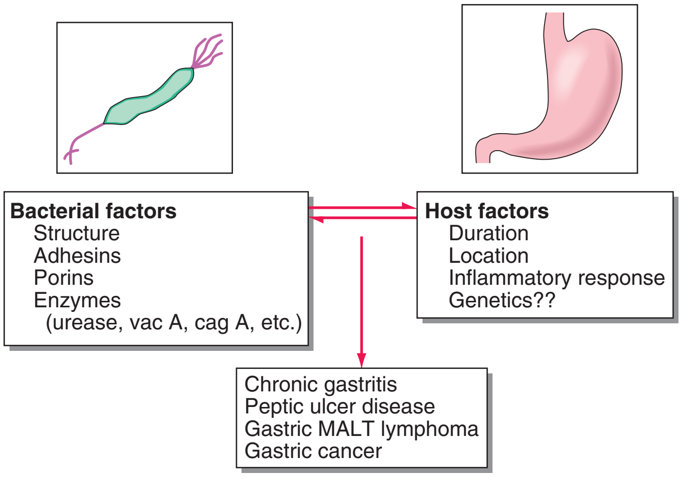

FIGURE 335-6 Outline of the bacterial and host factors important in determining *H. pylori*–induced gastrointestinal disease. MALT, mucosal-associated lymphoid tissue.

Bacterial factors: H. pylori is able to facilitate gastric residence, induce mucosal injury, and avoid host defense. Different strains of *H. pylori* produce different virulence factors including γ-glutamyl transpeptidase (GGT), cytotoxin-associated gene A (Cag A) product, and virulence components vacuolating toxin (Vac A), in addition to pathogen-associated molecular patterns (PAMPs) such as flagella and lipopolysaccharide (LPS). A specific region of the bacterial genome, the pathogenicity island (cag-PAI), encodes the virulence factors Cag A and pic B. Vac A also contributes to pathogenicity, although it is not encoded within the pathogenicity island. These virulence factors, in conjunction with additional bacterial constituents, can cause mucosal damage, in part through their ability to target the host immune cells. For example, Vac A targets human CD4 T cells, inhibiting their proliferation, and in addition can disrupt normal function of B cells, CD8 T cells, macrophages, and mast cells. Multiple studies have demonstrated that *H. pylori* strains that are cag-PAI positive are associated with a higher risk of PUD, premalignant gastric lesions, and gastric cancer than are strains that lack the cag-PAI. In addition, *H. pylori* may directly inhibit parietal cell H^+,K^+-ATPase activity through a Cag A–dependent mechanism, leading in part to the low acid production observed after acute infection with the organism. Urease, which allows the bacteria to reside in the acidic stomach, generates NH_3, which can damage epithelial cells. The bacteria produce surface factors that are chemotactic for neutrophils and monocytes, which in turn contribute to epithelial cell injury (see below). *H. pylori* makes proteases and phospholipases that break down the glycoprotein lipid complex of the mucous gel, thus reducing the efficacy of this first line of mucosal defense. *H. pylori* expresses adhesins (outer membrane proteins like BabA), which facilitate attachment of the bacteria to gastric epithelial cells. Although LPS of gram-negative bacteria often plays an important role in the infection, *H. pylori* LPS has low immunologic activity compared to that of other organisms. It may promote a smoldering chronic inflammation.

Host factors: Studies in twins suggest that there may be genetic predisposition to acquire *H. pylori.* The inflammatory response to *H. pylori* includes recruitment of neutrophils, lymphocytes (T and B), macrophages, and plasma cells. The pathogen leads to local injury by binding to class II major histocompatibility complex (MHC) molecules expressed on gastric epithelial cells, leading to cell death (*apoptosis*). Moreover, bacterial strains that encode cag-PAI can introduce Cag A into the host cells, leading to further cell injury and activation of cellular pathways involved in cytokine production and repression of tumor-suppressor genes. Elevated concentrations of multiple cytokines are found in the gastric epithelium of *H. pylori*–infected individuals, including interleukin (IL) 1α/β, IL-2, IL-6, IL-8, tumor necrosis factor (TNF) α, and interferon (IFN) γ. *H. pylori* infection also leads to both a mucosal and a systemic humoral response, which does not lead to eradication of the bacteria but further compounds epithelial cell injury. Additional mechanisms by which *H. pylori* may cause epithelial cell injury include (1) activated neutrophil-mediated production of reactive oxygen or nitrogen species and enhanced epithelial cell turnover and (2) apoptosis related to interaction with T cells (T helper 1 [T_H1] cells) and IFN-γ. Finally, the human stomach is colonized by a host of commensal organisms that may affect the likelihood of *H. pylori* infection and subsequent mucosal injury. Moreover, colonization of the stomach with *H. pylori* likely alters the composition of the gastric microbiota. The impact of the latter on gastric pathophysiology remains unknown but some studies suggest a potential increase in the development of gastric cancer. *H. pylori* also appears to regulate NO formation via different mechanisms that in turn may contribute to the organism's cytotoxic effects. Specifically, *H. pylori*–derived factors, such as urease, or the bacterium itself, stimulate NO synthase (NOS2) expression in macrophages and in gastric epithelial cells leading to NO release and subsequent cytotoxic effect on surrounding cells. *H. pylori* also leads to the formation of 8-nitroguanine (8-NO2-Gua), which in conjunction with oncoprotein Cag A, may contribute to the development of gastric cancer.

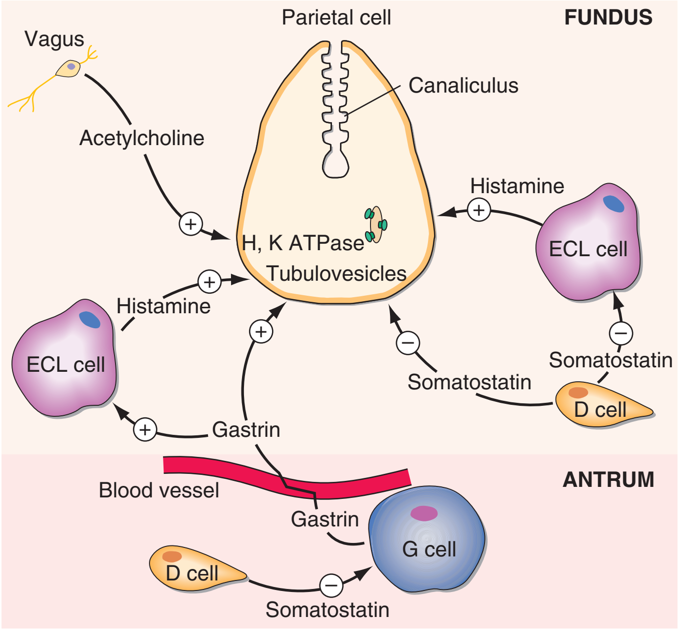

FIGURE 335-7 Summary of potential mechanisms by which *H. pylori* may lead to gastric secretory abnormalities. D, somatostatin cell; ECL, enterochromaffin-like cell; G, G cell. *(Reproduced with permission from J Calam et al: How does Helicobacter pylori cause mucosal damage? Its effect on acid and gastrin physiology. Gastroenterology 113:543, 1997.)*

The basis for *H. pylori*–mediated duodenal ulceration remains unclear. Studies suggest that *H. pylori* associated with duodenal ulceration may be more virulent. In addition, certain specific bacterial factors such as the DU-promoting gene A (*dupA*) may be associated with the development of DUs. Another potential contributing factor is that gastric metaplasia in the duodenum of DU patients, which may be due to high acid exposure (see below), permits *H. pylori* to bind to it and produce local injury secondary to the host response. Another hypothesis is that *H. pylori* antral infection could lead to increased acid production, increased duodenal acid, and mucosal injury. Basal and stimulated (meal, gastrin-releasing peptide [GRP]) gastrin release is increased in *H. pylori*–infected individuals, and somatostatin-secreting D cells may be decreased. *H. pylori* infection might induce increased acid secretion through both direct and indirect actions of *H. pylori* and proinflammatory cytokines (IL-8, TNF, and IL-1) on G, D, and parietal cells **(Fig. 335-7)**. GUs, in contrast, are associated with *H. pylori*–induced pangastritis and normal or low gastric acid secretion. The *H. pylori*–mediated decrease in gastric acid secretion after long-term infection may be due to the bacterium's ability to inhibit H^+,K^+-ATPase expression. *H. pylori* infection has also been associated with decreased duodenal mucosal bicarbonate production. Data supporting and contradicting each of these interesting theories have been demonstrated. Thus, the mechanism by which *H. pylori* infection of the stomach leads to duodenal ulceration remains to be established. The development of in vitro organoids, a unique tool that replicates in part the multicellular structure of the intact organ, provides a more physiologic model for experimentation in an in vitro system. Moreover, the development of advanced microscopic optical imaging techniques will lead to increased understanding of parietal cell adaptation to *H. pylori* infection.

In summary, the final effect of *H. pylori* on the GI tract is variable and determined by microbial and host factors. The type and distribution of gastritis correlate with the ultimate gastric and duodenal pathology observed. Specifically, the presence of antral-predominant gastritis is associated with DU formation; gastritis involving primarily the corpus predisposes to the development of GUs, gastric atrophy, and ultimately gastric carcinoma **(Fig. 335-8)**.

NSAID-INDUCED DISEASE

Epidemiology NSAIDs represent a group of the most commonly used medications in the world and the United States. It is estimated that 7 billion dollars per year are spent on NSAIDs worldwide, with >30 billion over-the-counter tablets sold. More than 30 million individuals take NSAIDs, with >100 million prescriptions sold yearly in the United States

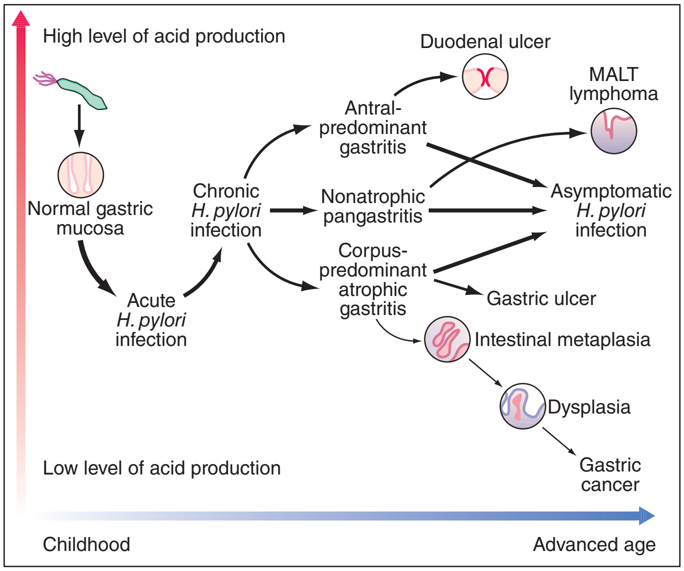
High level of acid production
Duodenal ulcer
MALT lymphoma
Antral-predominant gastritis
Normal gastric mucosa
Chronic H. pylori infection
Nonatrophic pangastritis
Asymptomatic H. pylori infection
Corpus-predominant atrophic gastritis
Acute H. pylori infection
Gastric ulcer
Intestinal metaplasia
Dysplasia
Low level of acid production
Gastric cancer
Childhood
Advanced age

FIGURE 335-8 Natural history of *H. pylori* infection. MALT, mucosal-associated lymphoid tissue. *(From The New England Journal of Medicine, Medical progress: Helicobacter pylori infection, S Suerbaum, P Michetti: 347:1175-1186. Copyright @2002 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

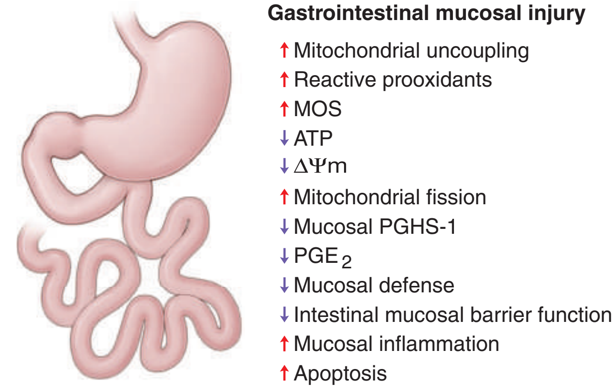

FIGURE 335-9 Effect of nonsteroidal anti-inflammatory drugs (NSAIDs) on different target organs. The action of NSAIDs on major organs including stomach, small intestine, heart, liver, kidney, respiratory tract, and brain is mainly mediated through prostaglandin endoperoxide synthase (PGHS)–dependent prostanoid modulation and alteration of mitochondrial functional integrity leading to mitochondrial oxidative stress (MOS) generation, depolarization of mitochondrial transmembrane potential (ΔΨm), and consequent cell death. However, in heart, low-dose aspirin actually offers cardioprotection through antithrombotic effect. *Upward arrows* indicate upregulation/elevation; *downward arrows* indicate downregulation/depletion. *(From S Bindu et al: Non-steroidal anti-inflammatory drugs (NSAIDs) and organ damage: A current perspective. Biochem Pharmacol 180:114147, 2020.)*

alone. In fact, after the introduction of COX-2 inhibitors in the year 2000, the number of prescriptions written for NSAIDs was >111 million at a cost of $4.8 billion. Side effects and complications due to NSAIDs are considered the most common drug-related toxicities in the United States. The spectrum of NSAID-induced morbidity ranges from nausea and dyspepsia (prevalence reported as high as 50–60%) to a serious GI complication such as endoscopy-documented peptic ulceration (15–30% of individuals taking NSAIDs regularly), which is complicated by bleeding or perforation in as many as 1.5% of users per year. It is estimated that NSAID-induced GI bleeding accounts for 60,000–120,000 hospital admissions per year, and deaths related to NSAID-induced toxicity may be as high as 16,000 per year in the United States. Approximately 4–5% of patients develop symptomatic ulcers within 1 year. Unfortunately, dyspeptic symptoms do not correlate with NSAID-induced pathology. Over 80% of patients with serious NSAID-related complications did not have preceding dyspepsia. In view of the lack of warning signs, it is important to identify patients who are at increased risk for morbidity and mortality related to NSAID usage. Even 75 mg/d of aspirin may lead to serious GI ulceration; thus, no dose of NSAID is completely safe. In fact, the incidence of mucosal injury (ulcers and erosions) in patients taking low-dose aspirin (75–325 mg) has been estimated to range from as low as 8 to as high as 60%. It appears that *H. pylori* infection increases the risk of PUD-associated GI bleeding in chronic users of low-dose aspirin. Established risk factors include advanced age, history of ulcer, concomitant use of glucocorticoids, high-dose NSAIDs, multiple NSAIDs, concomitant use of anticoagulants or clopidogrel, and serious or multisystem disease. Possible risk factors include concomitant infection with *H. pylori*, cigarette smoking, and alcohol consumption. SSRIs have a synergistic effect on the induction of GI bleeding believed to be due in part to this agent's ability to decrease platelet aggregation by decreasing serotonin content in platelets.

Pathophysiology Prostaglandins play a critical role in maintaining gastroduodenal mucosal integrity and repair. It therefore follows that interruption of prostaglandin synthesis can impair mucosal defense and repair, thus facilitating mucosal injury via a systemic mechanism. Animal studies have demonstrated that neutrophil adherence to the gastric microcirculation plays an essential role in the initiation of NSAID-induced mucosal injury. A summary of the pathogenetic pathways by which systemically administered NSAIDs may lead to mucosal injury is shown in **Fig. 335-9**. Single nucleotide polymorphisms (SNPs) have been found in several genes, including those encoding certain subtypes of cytochrome P450 (see below), IL-1β (*IL-1β*), angiotensinogen (*AGT*), and an organic ion transporting polypeptide (*SLCO1B1*), but these findings need confirmation in larger-scale studies.

Injury to the mucosa also occurs as a result of the topical use of NSAIDs, leading to increased epithelial surface permeability. Aspirin and many NSAIDs are weak acids that remain in a nonionized lipophilic form when found within the acid environment of the stomach. Under these conditions, NSAIDs migrate across lipid membranes of epithelial cells, leading to cell injury once trapped intracellularly in an ionized form. Topical NSAIDs can also alter the surface mucous layer, permitting back diffusion of H^+ and pepsin, leading to further epithelial cell damage. Moreover, enteric-coated or buffered preparations are also associated with risk of peptic ulceration. NSAIDs can also lead to mucosal injury via production of additional proinflammatory mediators such as TNF and leukotrienes through simultaneous activation of the lipoxygenase pathway.

The interplay between *H. pylori* and NSAIDs in the pathogenesis of PUD is complex. Meta-analysis supports the conclusion that each of these aggressive factors is an independent and synergistic risk factor for PUD and its complications such as GI bleeding. For example, eradication of *H. pylori* reduces the likelihood of GI complications in high-risk individuals to levels observed in individuals with average risk of NSAID-induced complications.

In summary, NSAID-induced mucosal injury is a multifaceted process involving the interaction of multiple, often synergistic pathophysiologic processes at the epithelium and surrounding interfaces.

PATHOGENETIC FACTORS UNRELATED TO *H. PYLORI* AND NSAIDS IN ACID PEPTIC DISEASE Cigarette smoking has been implicated in the pathogenesis of PUD. Not only have smokers been found to have ulcers more frequently than do nonsmokers, but smoking appears to decrease healing rates, impair response to therapy, and increase ulcer-related complications such as perforation. The mechanism responsible for increased ulcer diathesis in smokers is unknown. Theories have included altered gastric emptying, decreased proximal duodenal bicarbonate production, increased risk for *H. pylori* infection, and cigarette-induced generation of noxious mucosal free radicals. Genetic predisposition may play a role in ulcer development. First-degree relatives of DU patients are three times as likely to develop an ulcer; however, the potential role of *H. pylori* infection in contacts is a major consideration. Increased frequencies of blood group O and of the nonsecretor status have also been implicated as genetic risk factors for peptic diathesis. However, *H. pylori* preferentially binds to group O antigens. Additional genetic factors have been postulated to predispose certain individuals to developing PUD and/or upper GI bleeding. Specifically, genes encoding the NSAID-metabolizing enzymes cytochrome P450 2C9 and 2C8 (CYP2C9 and CYP2C8) are potential susceptibility genes for NSAID-induced PUD, but unfortunately, the studies have not been consistent in demonstrating this association. In a United Kingdom study, the $CYP2C19^{*}17$ gain-of-function polymorphism was associated with PUD in a Caucasian cohort, irrespective of ulcer etiology. These findings need to be confirmed in broader studies. Psychological stress has been thought to contribute to PUD, but studies examining the role of psychological factors in its pathogenesis have generated conflicting results. Although PUD is associated with certain personality traits (neuroticism), these same traits are also present in individuals with nonulcer dyspepsia (NUD) and other functional and organic disorders.

Diet has also been thought to play a role in peptic diseases. Certain foods and beverages can cause dyspepsia, but no convincing studies indicate an association between ulcer formation and a specific diet. Specific chronic disorders have been shown to have a strong association with PUD: (1) advanced age, (2) chronic pulmonary disease, (3) chronic renal failure, (4) cirrhosis, (5) nephrolithiasis, (6) α_1-antitrypsin deficiency, and (7) systemic mastocytosis. Disorders with a possible association are (1) hyperparathyroidism, (2) coronary artery disease, (3) polycythemia vera, (4) chronic pancreatitis, (5) former alcohol use, (6) obesity, (7) African-American race, and (8) three or more doctor visits in a year.

Multiple factors play a role in the pathogenesis of PUD. The two predominant causes are *H. pylori* infection and NSAID ingestion. PUD not related to *H. pylori* or NSAIDs is increasing. Other less common causes of PUD are shown in **Table 335-1**. These etiologic agents should be considered as the incidence of *H. pylori* is decreasing. Independent of the inciting or injurious agent, peptic ulcers develop as a result of an imbalance between mucosal protection/repair and aggressive factors. Gastric acid plays an important role in mucosal injury.

CLINICAL FEATURES

History Abdominal pain is common to many GI disorders, including DU and GU, but it has a poor predictive value for the presence of either DU or GU. Approximately two-thirds of patients with PUD do not have abdominal pain, and up to 87% of patients with NSAID-induced mucosal disease can present with a complication (bleeding, perforation, and obstruction) without antecedent symptoms. Despite this poor correlation, a careful history and physical examination are essential components of the approach to a patient suspected of having peptic ulcers.

TABLE 335-1 Causes of Ulcers Not Caused by *Helicobacter pylori* and NSAIDs

Pathogenesis of Non-Hp and Non-NSAID Ulcer Disease
Infection
Cytomegalovirus
Herpes simplex virus
Helicobacter heilmannii
Drug/Toxin
Bisphosphonates
Checkpoint inhibitor
Chemotherapy
Clopidogrel
Crack cocaine
Glucocorticoids (when combined with NSAIDs)
Mycophenolate mofetil
Potassium chloride
Miscellaneous
Basophilia in myeloproliferative disease
Duodenal obstruction (e.g., annular pancreas)
Infiltrating disease
Ischemia
Radiation therapy
Eosinophilic infiltration
Sarcoidosis
Crohn's disease
Idiopathic hypersecretory state

Abbreviations: Hp, *H. pylori*; NSAIDs, nonsteroidal anti-inflammatory drugs.

Epigastric pain described as a burning or gnawing discomfort can be present in both DU and GU. The discomfort is also described as an ill-defined, aching sensation or as hunger pain. The typical pain pattern in DU occurs 90 min to 3 h after a meal and is frequently relieved by antacids or food. Pain that awakes the patient from sleep (between midnight and 3 A.M.) is the most discriminating symptom, with two-thirds of DU patients describing this complaint. Unfortunately, this symptom is also present in one-third of patients with NUD (see below). Elderly patients are less likely to have abdominal pain as a manifestation of PUD and may instead present with a complication such as ulcer bleeding or perforation. The pain pattern in GU patients may be different from that in DU patients, where discomfort may actually be precipitated by food. Nausea and weight loss occur more commonly in GU patients. Endoscopy detects ulcers in <30% of patients who have dyspepsia.

The mechanism for development of abdominal pain in ulcer patients is unknown. Several possible explanations include acid-induced activation of chemical receptors in the duodenum, enhanced duodenal sensitivity to bile acids and pepsin, and altered gastroduodenal motility.

Variation in the intensity or distribution of the abdominal pain, as well as the onset of associated symptoms such as nausea and/or vomiting, may be indicative of an ulcer complication. Dyspepsia that becomes constant, is no longer relieved by food or antacids, or radiates to the back may indicate a penetrating ulcer (pancreas). Sudden onset of severe, generalized abdominal pain may indicate perforation. Pain worsening with meals, nausea, and vomiting of undigested food suggest gastric outlet obstruction. Tarry stools or coffee-ground emesis indicate bleeding.

Physical Examination Epigastric tenderness is the most frequent finding in patients with GU or DU. Pain may be found to the right of the midline in 20% of patients. Unfortunately, the predictive value of this finding is low. Physical examination is critically important for discovering evidence of ulcer complication. Tachycardia and orthostasis suggest dehydration secondary to vomiting or active GI blood loss. A severely tender, board-like abdomen suggests a perforation. Presence of a succussion splash indicates retained fluid in the stomach, suggesting gastric outlet obstruction.

PUD-Related Complications • GASTROINTESTINAL BLEEDING GI bleeding is the most common complication observed in PUD. Bleeding is estimated to occur in 19.4–57 per 100,000 individuals in a general population or in ~15% of patients. Bleeding and complications of ulcer disease occur more often in individuals >60 years of age. The 30-day mortality rate is as high as 2.5–10%. The higher incidence in the elderly is likely due to the increased use of NSAIDs in this group. In addition, up to 80% of the mortality in PUD-related bleeding is due to nonbleeding causes such as multiorgan failure (24%), pulmonary complications (24%), and malignancy (34%).

Greater than 50% of patients with ulcer-related hemorrhage bleed without any preceding warning signs or symptoms.

PERFORATION The second most common ulcer-related complication is perforation, being reported in as many as 6–7% of PUD patients with an estimated 30-day mortality of >20%. Acute abdominal pain, tachycardia, and abdominal rigidity compose the classic triad associated with this complication. It is essential to remember that elderly patients or individuals who are immunosuppressed may not have this classic presentation. As in the case of bleeding, the incidence of perforation in the elderly appears to be increasing secondary to increased use of NSAIDs. Perforation of DUs has become less common in light of the increased rates of *H. pylori* eradication, with NSAID-induced GUs leading to perforation occurring more commonly. *Penetration* is a form of perforation in which the ulcer bed tunnels into an adjacent organ. DUs tend to penetrate posteriorly into the pancreas, leading to pancreatitis, whereas GUs tend to penetrate into the left hepatic lobe. Gastrocolic fistulas associated with GUs have also been described. Mortality for this complication can be >20% within 30 days.

GASTRIC OUTLET OBSTRUCTION Gastric outlet obstruction is the least common ulcer-related complication, occurring in 1–2% of patients. A patient may have relative obstruction secondary to ulcer-related inflammation and edema in the peripyloric and duodenal region. This process often resolves with ulcer healing. A fixed, mechanical obstruction secondary to scar formation in the peripyloric areas is also possible. The latter requires endoscopic (balloon dilation with or without placement of a biodegradable stent) or surgical intervention with a stricturoplasty or gastrojejunostomy. Signs and symptoms relative to mechanical obstruction may develop insidiously. New onset of early satiety, nausea, vomiting, increase of postprandial abdominal pain, and weight loss should make gastric outlet obstruction a possible diagnosis.

Differential Diagnosis The list of GI and non-GI disorders that can mimic ulceration of the stomach or duodenum is quite extensive. The most commonly encountered diagnosis among patients seen for upper abdominal discomfort is functional dyspepsia (FD) or *essential dyspepsia*, which refers to a group of heterogeneous disorders typified by upper abdominal pain without the presence of an ulcer. The symptoms can range from postprandial fullness and early satiety to epigastric burning pain. The dichotomy of this symptom complex has led to the identification of two subcategories of FD including postprandial distress syndrome (PDS) and epigastric pain syndrome (EPS). Dyspepsia has been reported to occur in up to 30% of the U.S. population. Up to 80% of patients seeking medical care for dyspepsia have a negative diagnostic evaluation. The etiology of FD is not established, but postinfectious states, certain foods, and *H. pylori* infection may contribute to the pathogenesis of this common disorder.

Several additional disease processes that may present with "ulcer-like" symptoms include proximal GI tumors, gastroesophageal reflux, vascular disease, pancreaticobiliary disease (biliary colic, chronic pancreatitis), and gastroduodenal Crohn's disease.

Diagnostic Evaluation In view of the poor predictive value of abdominal pain for the presence of a gastroduodenal ulcer and the multiple disease processes that can mimic this disease, the clinician is often confronted with having to establish the presence of an ulcer. Documentation of an ulcer requires either a radiographic (barium study, rarely done in today's environment) or an endoscopic procedure. However, a large percentage of patients with symptoms suggestive of an ulcer have NUD; testing for *H. pylori* and antibiotic therapy (see below) are appropriate for individuals who are otherwise healthy, without red flag signs such as nausea, vomiting, weight loss, or evidence of GI bleeding and <60 years of age, before embarking on a diagnostic evaluation **(Chap. 48)**.

Barium studies of the proximal GI tract are rarely used as a first test for documenting an ulcer **(Fig. 335-10)**. Endoscopy provides the most sensitive and specific approach for examining the upper GI tract **(Fig. 335-11)**. In addition to permitting direct visualization of the mucosa, endoscopy facilitates photographic documentation of a mucosal defect and tissue biopsy to rule out malignancy (GU) or *H. pylori*. Endoscopic examination is particularly helpful in identifying lesions too small to detect by radiographic examination, for evaluation of atypical radiographic abnormalities, or to determine if an ulcer is a source of blood loss.

Although the methods for diagnosing *H. pylori* are outlined in **Chap. 168**, a brief summary will be included here **(Table 335-2)**. Several biopsy urease tests have been developed (PyloriTek, CLOtest, Hpfast, Pronto Dry) that have a sensitivity and specificity of >90–95%. Several noninvasive methods for detecting this organism have been developed. Three types of studies routinely used include serologic testing, the ^{13}C-urea breath test, and the fecal *H. pylori* (Hp) antigen test (monoclonal antibody test). A urinary Hp antigen test and a home breath test appear promising.

Occasionally, specialized testing such as serum gastrin and gastric acid analysis may be needed in individuals with complicated or refractory PUD (see "Zollinger-Ellison Syndrome," below). Screening for aspirin or NSAIDs (blood or urine) may also be necessary in refractory *H. pylori*–negative PUD patients.

TREATMENT

Peptic Ulcer Disease

Before the discovery of *H. pylori*, the therapy of PUD was centered on the old dictum by Schwartz of "no acid, no ulcer." Although acid secretion is still important in the pathogenesis of PUD, eradication of *H. pylori* and therapy/prevention of NSAID-induced disease is the mainstay of treatment. A summary of commonly used drugs for treatment of acid peptic disorders is shown in **Table 335-3**.

ACID-NEUTRALIZING/INHIBITORY DRUGS

Antacids Before we understand the important role of histamine in stimulating parietal cell activity, neutralization of secreted acid with antacids constituted the main form of therapy for peptic ulcers. They are now rarely, if ever, used as the primary therapeutic agent

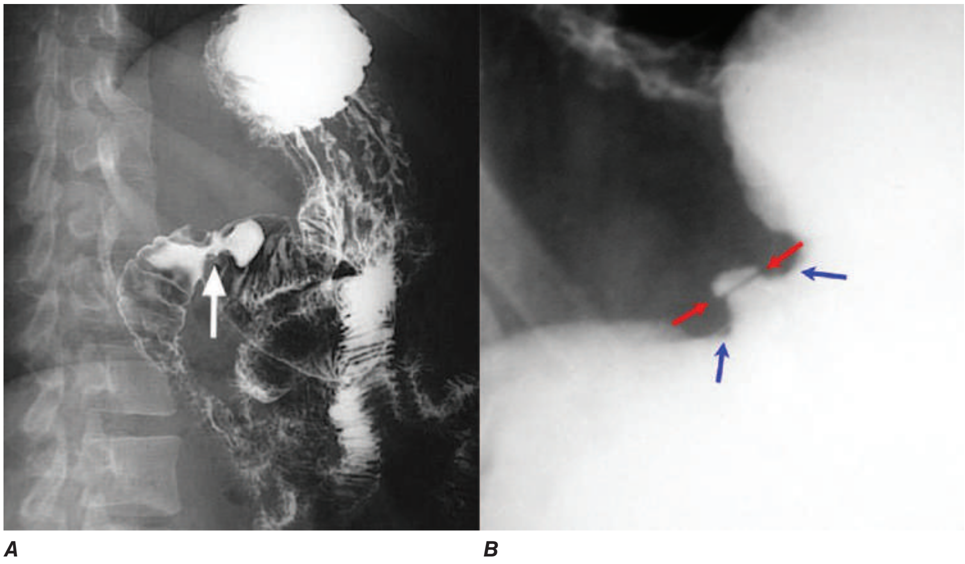

FIGURE 335-10 Barium study demonstrating (*A*) a benign duodenal ulcer and (***B***) a benign gastric ulcer.

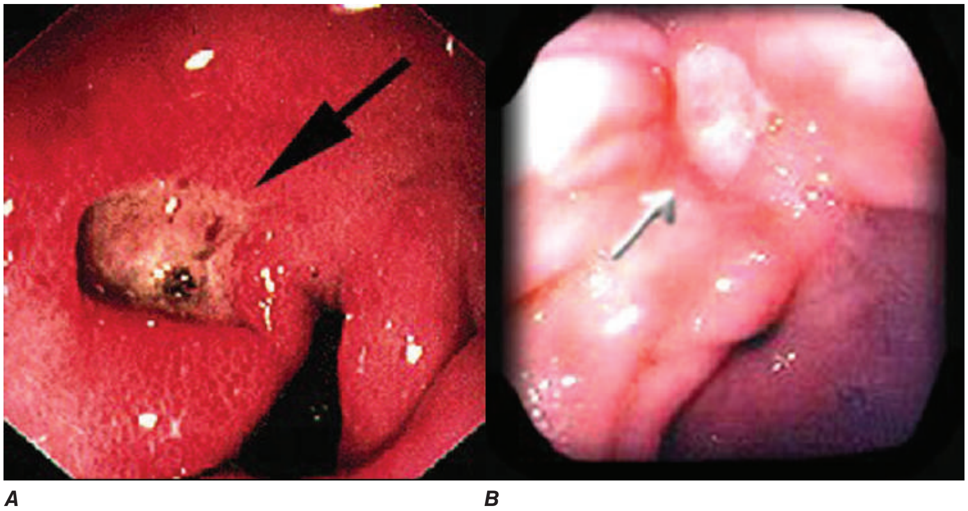

FIGURE 335-11 Endoscopy demonstrating (***A***) a benign duodenal ulcer and (***B***) a benign gastric ulcer.

but instead are often used by patients for symptomatic relief of dyspepsia. The most commonly used agents are mixtures of aluminum hydroxide and magnesium hydroxide. Aluminum hydroxide can produce constipation and phosphate depletion; magnesium hydroxide may cause loose stools. Many of the commonly used antacids (e.g., Maalox, Mylanta) have a combination of both aluminum and magnesium hydroxide in order to avoid these side effects. The magnesium-containing preparation should not be used in chronic renal failure patients because of possible hypermagnesemia, and aluminum may cause chronic neurotoxicity in these patients.

Calcium carbonate and sodium bicarbonate are potent antacids with varying levels of potential problems. The long-term use of calcium carbonate (converts to calcium chloride in the stomach) can lead to milk-alkali syndrome (hypercalcemia and hyperphosphatemia with possible renal calcinosis and progression to renal insufficiency). Sodium bicarbonate may induce systemic alkalosis.

H_2 Receptor Antagonists Four of these agents are presently available (cimetidine, ranitidine, famotidine, and nizatidine), and their structures share homology with histamine. Although each has different potency, all will significantly inhibit basal and stimulated acid secretion to comparable levels when used at therapeutic doses. Moreover, similar ulcer-healing rates are achieved with each drug when used at the correct dosage. Presently, this class of drug is often used for treatment of active ulcers (4–6 weeks) in combination with antibiotics directed at eradicating *H. pylori* (see below).

TABLE 335-2 Tests for Detection of *Helicobacter pylori*

TEST	SENSITIVITY/ SPECIFICITY, %	COMMENTS
Invasive (Endoscopy/Biopsy Required)		
Rapid urease	80–95/95–100	Simple, false negative with recent use of PPIs, antibiotics, or bismuth compounds
Histology	60–90/>95	Requires pathology processing and staining; provides histologic information
Culture	76-90/100	Time-consuming, expensive, dependent on experience; allows determination of antibiotic susceptibility
Noninvasive		
Serology	74.4/59	Inexpensive, convenient; not useful for early follow-up; cannot distinguish active and prior infection
Urea breath test	>95/>95	Simple, rapid; useful for early follow-up; false negatives with recent therapy (see rapid urease test)
Stool antigen	>95/>95	Inexpensive, convenient

Abbreviation: PPIs, proton pump inhibitors.

Cimetidine was the first H_2 receptor antagonist used for the treatment of acid peptic disorders. Cimetidine may have weak antiandrogenic side effects resulting in reversible gynecomastia and impotence, primarily in patients receiving high doses for prolonged periods of time (months to years). In view of cimetidine's ability to inhibit cytochrome P450, careful monitoring of drugs such as warfarin, phenytoin, and theophylline is indicated with long-term usage. Other rare reversible adverse effects reported with cimetidine include confusion and elevated levels of serum aminotransferases, creatinine, and serum prolactin. Famotidine is a more potent H_2 receptor antagonists than cimetidine. It can be used once a day at bedtime for ulcer prevention, which was commonly done before the discovery of *H. pylori* and the development of proton pump inhibitors (PPIs). Patients may develop tolerance to H_2 blockers, a rare event with PPIs (see below). Comparable nighttime dosing regimens are cimetidine 800 mg and famotidine 40 mg.

Additional rare, reversible systemic toxicities reported with H_2 receptor antagonists include pancytopenia, neutropenia, anemia, and thrombocytopenia, with a prevalence rate varying from 0.01 to 0.2%. Cimetidine can bind to hepatic cytochrome P450; famotidine does not.

Proton Pump (H^+,K^+-ATPase) Inhibitors Omeprazole, esomeprazole, lansoprazole, rabeprazole, and pantoprazole are substituted

TABLE 335-3 Drugs Used in the Treatment of Peptic Ulcer Disease

DRUG TYPE/ MECHANISM	EXAMPLES	DOSE
Acid-Suppressing Drugs		
Antacids	Mylanta, Maalox, Tums, Gaviscon	100–140 meq/L 1 and 3 h after meals and hs
H_2 receptor antagonists	Cimetidine	400 mg bid
	Famotidine	40 mg hs
Proton pump inhibitors	Omeprazole	20 mg/d
	Lansoprazole	30 mg/d
	Rabeprazole	20 mg/d
	Pantoprazole	40 mg/d
	Esomeprazole	20 mg/d
	Dexlansoprazole	30 mg/d
Mucosal Protective Agents		
Sucralfate	Sucralfate	1 g qid
Prostaglandin analogue	Misoprostol	200 μg qid
Bismuth-containing compounds	Bismuth subsalicylate (BSS)	See anti–*H. pylori* regimens (Table 335-4)

Abbreviation: hs, at bedtime (*hora somni*).

benzimidazole derivatives that covalently bind and irreversibly inhibit H^+,K^+-ATPase. Esomeprazole is the S-enantiomer of omeprazole, which is a racemic mixture of both S- and R-optical isomers. The R-isomer of lansoprazole, dexlansoprazole, is the most recent PPI approved for clinical use. Its reported advantage is a dual delayed-release system aimed at improving treatment of gastroesophageal reflux disease (GERD). These are the most potent acid inhibitory agents available. Omeprazole and lansoprazole are the PPIs that have been used for the longest time. Both are acid-labile and are administered as enteric-coated granules in a sustained-release capsule that dissolves within the small intestine at a pH of 6. Lansoprazole is available in an orally disintegrating tablet that can be taken with or without water, an advantage for individuals who have significant dysphagia. Absorption kinetics are similar to the capsule. In addition, a lansoprazole-naproxen combination preparation that has been made available is targeted at decreasing NSAID-related GI injury (see below). Omeprazole is available as non–enteric-coated granules mixed with sodium bicarbonate in a powder form that can be administered orally or via gastric tube. The sodium bicarbonate has two purposes: to protect the omeprazole from acid degradation and to promote rapid gastric alkalinization and subsequent proton pump activation, which facilitates rapid action of the PPI. Pantoprazole and rabeprazole are available as enteric-coated tablets. Pantoprazole is also available as a parenteral formulation for intravenous use. These agents are lipophilic compounds; upon entering the parietal cell, they are protonated and trapped within the acid environment of the tubulovesicular and canalicular system. These agents potently inhibit all phases of gastric acid secretion. Onset of action is rapid, with a maximum acid inhibitory effect between 2 and 6 h after administration and duration of inhibition lasting up to 72–96 h. With repeated daily dosing, progressive acid inhibitory effects are observed, with basal and secretagogue-stimulated acid production being inhibited by >95% after 1 week of therapy. The half-life of PPIs is ~18 h; thus, it can take between 2 and 5 days for gastric acid secretion to return to normal levels once these drugs have been discontinued. Because the pumps need to be activated for these agents to be effective, their efficacy is maximized if they are administered before a meal (except for the immediate-release formulation of omeprazole) (e.g., in the morning before breakfast). Mild to moderate hypergastrinemia has been observed in patients taking these drugs. Carcinoid tumors developed in some animals given the drugs chronically; however, extensive experience has failed to demonstrate gastric carcinoid tumor development in humans. Serum gastrin levels return to normal levels within 1–2 weeks after drug cessation. Rebound gastric acid hypersecretion has been described in *H. pylori*–negative individuals after discontinuation of PPIs. It occurs even after relatively short-term usage (2 months) and may last for up to 2 months after the PPI has been discontinued. The mechanism involves gastrin-induced hyperplasia and hypertrophy of histamine-secreting ECL cells. The clinical relevance of this observation is that individuals may have worsening symptoms of GERD or dyspepsia upon stopping the PPI. Gradual tapering of the PPI and switching to an H_2 receptor antagonist may prevent this from occurring. *H. pylori*–induced inflammation and concomitant decrease in acid production may explain why this does not occur in *H. pylori*–positive patients. IF production is also inhibited, but vitamin B_{12} deficiency anemia is uncommon, probably because of the large stores of the vitamin. As with any agent that leads to significant hypochlorhydria, PPIs may interfere with absorption of drugs such as ketoconazole, ampicillin, iron, and digoxin. Hepatic cytochrome P450 can be inhibited by the earlier PPIs (omeprazole, lansoprazole). Rabeprazole, pantoprazole, and esomeprazole do not appear to interact significantly with drugs metabolized by the cytochrome P450 system. The overall clinical significance of this observation is not definitely established. Caution should be taken when using theophylline, warfarin, diazepam, atazanavir, and phenytoin concomitantly with PPIs.

The list of potential side effects with long-term PPI use has steadily grown over the years. These agents are commonly used since several formulations have become available as over-the-counter medications. Moreover, up to 70% of current prescriptions for long-term PPIs may be unwarranted and between 35 and 60% of in-hospital use of PPIs may be inappropriate. Interpretation of the multiple studies should take into consideration that the vast majority were retrospective observational studies in which confounding factors could not be accounted for entirely.

Long-term acid suppression, especially with PPIs, has been associated with a higher incidence of community-acquired pneumonia as well as community- and hospital-acquired *Clostridioides difficile*–associated disease. A meta-analysis showed a 74% increased risk of *C. difficile* infection and a 2.5-fold higher risk of reinfection as compared to nonusers. In light of these concerns, the FDA published a safety alert regarding the association between *C. difficile* infection and PPI use. Although the risk of spontaneous bacterial peritonitis in cirrhotics was thought to be increased, the data here are less supportive. The impact of PPI-induced changes in the host microbiome is postulated to play a role in the increased risk of infection, but this theory needs to be confirmed. These observations require confirmation but should alert the practitioner to take caution when recommending these agents for long-term use, especially in elderly patients at risk for developing pneumonia or *C. difficile* infection.

Diarrhea is also associated with PPI use, which in some cases has been associated with the development of collagenous colitis (hazard ratio of 4.5), particularly with lansoprazole. The mechanism for PPI-induced collagenous colitis is unclear, but in vitro studies demonstrate that PPIs may induce collagen gene expression. The colitis usually resolves with cessation of the PPI.

A population-based study revealed that long-term use of PPIs was associated with the development of hip fractures in older women. The absolute risk of fracture remained low and may be zero despite an observed increase associated with the dose and duration of acid suppression. The mechanism for this observation is not clear, and this finding must be confirmed before making broad recommendations regarding the discontinuation of these agents in patients who benefit from them. Long-term use of PPIs has also been implicated in the development of iron, vitamin B_{12}, and magnesium deficiency. A meta-analysis of nine observational studies found a 40% increase in hypomagnesemia in PPI users as compared to nonusers. One approach to consider in patients needing to take PPIs long term is to check a complete blood count looking for evidence of anemia due to iron or vitamin B_{12} deficiency, vitamin B_{12} level, and a magnesium level after 1–2 years of PPI use, but these recommendations are not evidence based or recommended by expert opinion. PPIs may exert a negative effect on the antiplatelet effect of clopidogrel. Although the evidence is mixed and inconclusive, a small increase in mortality and readmission rate for coronary events was seen in patients receiving a PPI while on clopidogrel in earlier studies. Subsequently, three meta-analyses reported an inverse correlation between clopidogrel and PPI use; therefore, the influence of this drug interaction on mortality is not clearly established. The mechanism involves the competition of the PPI and clopidogrel with the same cytochrome P450 (CYP2C19). Whether this is a class effect of PPIs is unclear; there appears to be at least a theoretical advantage of pantoprazole over the other PPIs, but this has not been confirmed. This drug interaction is particularly relevant in light of the common use of aspirin and clopidogrel for prevention of coronary events and the efficacy of PPIs in preventing GI bleeding in these patients. The FDA has made several recommendations while awaiting further evidence to clarify the impact of PPI therapy on clopidogrel use. Health care providers should continue to prescribe clopidogrel to patients who require it and should reevaluate the need for starting or continuing treatment with a PPI. From a practical standpoint, additional recommendations to consider include the following: Patients taking clopidogrel with aspirin, especially with other GI risk factors for bleeding, should receive GI protective therapy. Although high-dose H_2 blockers have been considered an option, these do not appear to be as effective as PPIs. If PPIs are to be given, some have recommended that there be a 12-h separation between

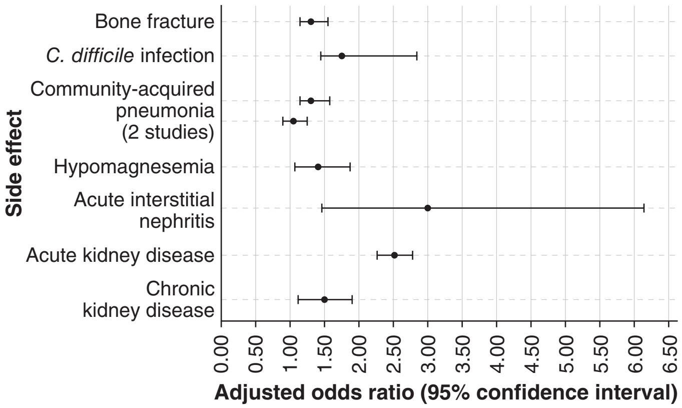

FIGURE 335-12. Evidence supporting the potential adverse effects of proton pump inhibitor drugs. *(Adapted from AJ Schoenfeld, D Grady: Adverse effects associated with proton pump inhibitors. JAMA Intern Med 176:172, 2016.)*

administration of the PPI and clopidogrel to minimize competition of the two agents with the involved cytochrome P450. One option is to give the PPI 30 min before breakfast and the clopidogrel at bedtime. Insufficient data are available to firmly recommend one PPI over another. Additional concerning side effects with long-term PPI use include increased cardiac risks independent of clopidogrel use, dementia, and acute and chronic kidney injury. Again, the data are often retrospective and confounding variables were not consistently eliminated, thus making it difficult to establish a definitive association between PPIs and the toxicities outlined. A summary of the side effects with the corresponding relative risks is shown in **Fig. 335-12**. Ultimately, heightened awareness of inappropriate long-term use of PPIs is paramount. Patients aged ≥65 years of age have a higher risk for some of the long-term side effects of PPIs highlighted above, in part due to the higher prevalence of concomitant chronic diseases. It is therefore essential to carefully select individuals, especially among the elderly, who need long-term PPI therapy and discontinue it in those individuals who do not need it. Abrupt withdrawal of a PPI in a long-term user may result in a component of rebound hyperacidity; thus, this agent should be tapered gradually over the course of 1–2 weeks with possible transition to an H_2 blocker for a short period of time.

Development of novel acid inhibitory agents continues in an attempt to primarily address the need for better agents to treat GERD. For example, modified H_2 blockers with greater potency and duration as well as novel PPIs with longer half-life and potency are under study. For example, tenatoprazole is a PPI containing an imidazopyridine ring instead of a benzimidazole ring, which promotes irreversible proton pump inhibition. This agent has a longer half-life than the other PPIs and may be beneficial for inhibiting nocturnal acid secretion, which has significant relevance in GERD. Additional PPIs with longer half-life and combined with other agents are being studied, but the details are beyond the scope of this chapter. A second new class of agents is the potassium-competitive acid pump antagonists (P-CAPs). These compounds inhibit gastric acid secretion via potassium competitive binding of the H^+,K^+-ATPase. Revaprazan, vonoprazan and tegoprazan are agents approved for use in Korea and Japan, and vonoprazan has been approved by the FDA for use in the United States. Vonoprazan may be superior to PPIs when combined with antibiotics for the treatment of *H. pylori*, and this novel agent was awarded Fast Track status by the FDA for the treatment of *H. pylori* in combination with both amoxicillin and clarithromycin and with amoxicillin alone.

CYTOPROTECTIVE AGENTS

Sucralfate Sucralfate is a complex sucrose salt in which the hydroxyl groups have been substituted by aluminum hydroxide and sulfate. This compound is insoluble in water and becomes a viscous paste within the stomach and duodenum, binding primarily to sites of active ulceration. Sucralfate may act by several mechanisms: serving as a physicochemical barrier, promoting a trophic action by binding growth factors such as EGF, enhancing prostaglandin synthesis, stimulating mucus and bicarbonate secretion, and enhancing mucosal defense and repair. Toxicity from this drug is rare, with constipation being most common (2–3%). It should be avoided in patients with chronic renal insufficiency to prevent aluminum-induced neurotoxicity. Hypophosphatemia and gastric bezoar formation have also been reported rarely. Standard dosing of sucralfate is 1 g qid.

Bismuth-Containing Preparations Sir William Osler considered bismuth-containing compounds the drug of choice for treating PUD. The resurgence in the use of these agents is due to their effect against *H. pylori*. Colloidal bismuth subcitrate (CBS) and bismuth subsalicylate (BSS; Pepto-Bismol) are the most widely used preparations. The mechanism by which these agents induce ulcer healing is unclear. Adverse effects with short-term use include black stools, constipation, and darkening of the tongue. Long-term use with high doses, especially with the avidly absorbed CBS, may lead to neurotoxicity. These compounds are commonly used as one of the agents in an anti–*H. pylori* regimen (see below).

Prostaglandin Analogues In view of their central role in maintaining mucosal integrity and repair, stable prostaglandin analogues were developed for the treatment of PUD. The mechanism by which this rapidly absorbed drug provides its therapeutic effect is through enhancement of mucosal defense and repair. The most common toxicity noted with this drug is diarrhea (10–30% incidence). Other major toxicities include uterine bleeding and contractions; misoprostol is contraindicated in women who may be pregnant, and women of childbearing age must be made clearly aware of this potential drug toxicity. The standard therapeutic dose is 200 μg qid.

Miscellaneous Drugs A number of drugs including anticholinergic agents and tricyclic antidepressants were used for treating acid peptic disorders, but in light of their toxicity and the development of potent antisecretory agents, these are rarely, if ever, used today. Newer agents such as teprenone, an acyclic polyisoprenoid compound used as a gastric mucosal protector that is employed to treat gastritis and GUs outside of the United States; plant-based therapies; and CCK2 receptor antagonists are intriguing therapies but require further evaluation.

THERAPY OF *H. PYLORI*

The physician's goal in treating PUD is to provide relief of symptoms (pain or dyspepsia), promote ulcer healing, and ultimately prevent ulcer recurrence and complications. The greatest influence of understanding the role of *H. pylori* in peptic disease has been the ability to prevent recurrence. Documented eradication of *H. pylori* in patients with PUD is associated with a dramatic decrease in ulcer recurrence to <10–20% as compared to 59% in GU patients and 67% in DU patients when the organism is not eliminated. Eradication of the organism may lead to diminished recurrent ulcer bleeding. The effect of its eradication on ulcer perforation is unclear.

Extensive effort has been made in determining who of the many individuals with *H. pylori* infection should be treated. The common conclusion arrived at by multiple consensus conferences around the world is that *H. pylori* should be eradicated in patients with documented PUD. This holds true independent of time of presentation (first episode or not), severity of symptoms, presence of confounding factors such as ingestion of NSAIDs, or whether the ulcer is in remission. Some have advocated treating patients with a history of documented PUD who are found to be *H. pylori* positive by stool antigen or breath testing. Between 60 and 90% of patients with gastric MALT lymphoma experience complete remission of the tumor in response to *H. pylori* eradication. The Maastricht VI/Florence Consensus Report recommends a test-and-treat approach for patients with uninvestigated dyspepsia if the local incidence of *H. pylori* is >20%. The American College of Gastroenterology (ACG) clinical guidelines (developed for North America) recommend that individuals aged <60 years with uninvestigated dyspepsia should be tested and treated for *H. pylori*. In addition, recommendations from this consensus report and the ACG clinical guidelines include testing and offering eradication of *H. pylori* in patients who will be using NSAIDs (including low-dose aspirin) on a long-term basis, especially if there is a prior history of PUD. These individuals will require continued PPI treatment as well as eradication treatment, because eradication of the organism alone does not eliminate the risk of gastroduodenal ulcers in patients already receiving long-term NSAIDs. Moreover, it appears that eradication of *H pylori* decreases NSAID-induced GI bleeding. Treating patients with NUD to prevent gastric cancer or patients with GERD requiring long-term acid suppression remains controversial. Guidelines from the ACG suggest eradication of *H. pylori* in patients who have undergone resection of early gastric cancer. The Maastricht VI/Florence Consensus Report also evaluated *H. pylori* treatment in gastric cancer prevention and recommends that eradication should be considered in the following situations: first-degree relatives of family members with gastric cancer; patients with previous gastric neoplasm treated by endoscopic or subtotal resection; individuals with a risk of gastritis (severe pangastritis or body-predominant gastritis) or severe atrophy; patients with gastric acid inhibition for >1 year; individuals with strong environmental risk factors for gastric cancer (heavy smoking; high exposure to dust, coal, quartz, or cement; and/or work in quarries); and *H. pylori*–positive patients with a fear of gastric cancer. Finally, the ACG clinical guidelines recommend testing and offering *H. pylori* eradication to patients with unexplained iron-deficiency anemia and idiopathic thrombocytopenic purpura. The Maastricht VI/Florence Consensus Report concurs with the ACG with the aforementioned recommendations and in addition recommends eradicating *H. pylori* in patients with unexplained vitamin B_{12} deficiency. Despite this, concerns have been raised about the widespread use of antibiotics for the therapy of all cases of *H. pylori* positivity, including the potential for increased bacterial resistance rates, reported weight gain, and alteration of the microbiome. In fact, the increasing incidence of bacterial resistance to antibiotics coupled with variability in general prevalence of the organism worldwide and limitations in developing personalized *H. pylori*–driven sensitivity therapies have led to increasing challenges in effectively treating this organism.

Multiple drugs have been evaluated in the therapy of *H. pylori.* No single agent is effective in eradicating the organism. Combination therapy for 14 days provides the greatest efficacy, although regimens based on sequential administration of antibiotics also appear promising (see below). A shorter administration course (7–10 days), although attractive, has not proved as successful as the 14-day regimens. The agents used with the greatest frequency include amoxicillin, metronidazole, tetracycline, clarithromycin, and bismuth compounds. Moreover, the steady increase in *H pylori* resistance to antibiotics, in particular clarithromycin and metronidazole, has significantly impacted the approach to eradication of this organism (see below). In addition, increasing resistance to antibiotics has led to a higher incidence of *H. pylori* infections that are refractory to first-line therapies.

Suggested treatment regimens for *H. pylori* are outlined in **Table 335-4**. Choice of a particular regimen will be influenced by several factors, including efficacy, patient tolerance, existing antibiotic resistance, prior antibiotic use, and cost of the drugs. The aim for initial eradication rates should be 85–90%. The combination of bismuth, metronidazole, and tetracycline was the first triple regimen found effective against *H. pylori.* The combination of two

TABLE 335-4 Recommended First-Line Therapies for *H. pylori* Infection

REGIMEN	DRUGS (DOSES)	DOSING FREQUENCY	DURATION (DAYS)	FDA APPROVAL
Clarithromycin triple (Only in patients without prior exposure to macrolides, incidence of clarithromycin <15%, or either of the above unknown)	PPI (standard or double dose)	bid	14	Yes[a]
	Clarithromycin (500 mg)			
	Amoxicillin (1 g) or metronidazole (500 mg tid)			
Bismuth quadruple (Now often recommended as first-line therapy)	PPI (standard dose)	bid	10–14	No[b]
	Bismuth subcitrate (120–300 mg) or subsalicylate (300 mg)	qid		
	Tetracycline (500 mg)	qid		
	Metronidazole (250–500 mg)	qid (250 mg) tid to qid (500 mg)		
Concomitant	PPI (standard dose)	bid	10–14	No
	Clarithromycin (500 mg)			
	Amoxicillin (1 g)			
	Nitroimidazole (500 mg)[c]			
Sequential	PPI (standard dose)	bid	5–7	No
	PPI, clarithromycin (500 mg) + nitroimidazole (500 mg)[c]	bid	5–7	
Hybrid	PPI (standard dose) + amoxicillin (1 g)	bid	7	No
	PPI, amoxicillin, clarithromycin (500 mg), nitroimidazole (500 mg)[c]	bid	7	
Levofloxacin triple	PPI (standard or double dose) + amoxicillin (1 g)	bid	5–7	No
	Levofloxacin (500 mg)	qd		
	Amoxicillin (1 g)	bid		
Levofloxacin sequential	PPI (standard or double dose) + amoxicillin (1 g)	bid	5–7	No
	PPI, amoxicillin, levofloxacin (500 mg qd), nitroimidazole (500 mg)[c]	bid	5–7	
LOAD	Levofloxacin (250 mg)	qd	7–10	No
	PPI (double dose)	qd		
	Nitazoxanide (500 mg)	bid		
	Doxycycline (100 mg)	qd		

[a] Several PPI, clarithromycin, and amoxicillin combinations have achieved FDA approval. The regimen of a PPI, clarithromycin, and metronidazole is not an FDA-approved treatment regimen. [b] The regimen of a PPI, bismuth, tetracycline, and metronidazole combined with a PPI for 10 days is an FDA-approved treatment regimen. [c] Metronidazole or tinidazole.

Abbreviations: bid, twice daily; FDA, Food and Drug Administration; PPI, proton pump inhibitor; tid, three times daily; qd, once daily; qid, four times daily.

Source: Reproduced with permission from WD Chey et al: ACG clinical guideline: Treatment of *Helicobacter pylori* infection. Am J Gastroenterol 112:212, 2017.

antibiotics plus either a PPI, H_2 blocker, or bismuth compound has comparable success rates. Addition of acid suppression assists in providing early symptom relief and enhances bacterial eradication.

Triple therapy, although effective, has several drawbacks, including the potential for poor patient compliance and drug-induced side effects. Compliance is being addressed by simplifying the regimens so that patients can take the medications twice a day. Simpler (dual therapy) and shorter regimens (7 and 10 days) are not as effective as triple therapy for 14 days. One anti–*H. pylori* regimen is available in prepackaged formulation: Prevpac (lansoprazole, clarithromycin, and amoxicillin).The contents of the Prevpac are to be taken twice per day for 14 days, understanding that clarithromycin-based triple therapy should be avoided in settings where *H. pylori* resistance to this agent exceeds 15% or the patient has been recently exposed to macrolide antibiotics or if neither of these factors is known (see below).

Side effects have been reported in up to 20–30% of patients on triple therapy. Bismuth may cause black stools, constipation, or darkening of the tongue. The most feared complication with amoxicillin is pseudomembranous colitis, but this occurs in <1–2% of patients. Amoxicillin can also lead to antibiotic-associated diarrhea, nausea, vomiting, skin rash, and allergic reaction. Concomitant use of probiotics may ameliorate some of the antibiotic side effects (see below). Tetracycline has been reported to cause rashes and, very rarely, hepatotoxicity and anaphylaxis.

One important concern with treating patients who may not need therapy is the potential for development of antibiotic-resistant strains. The incidence and type of antibiotic-resistant *H. pylori* strains vary worldwide. Strains resistant to metronidazole, clarithromycin, amoxicillin, and tetracycline have been described, with the latter two being uncommon. Antibiotic-resistant strains are the most common cause for treatment failure in compliant patients. Unfortunately, in vitro resistance does not predict outcome in patients. Culture and sensitivity testing of *H. pylori* is not performed routinely. In light of this, culture-independent methods are needed to determine antibiotic resistance. Detection of resistance toward several antibiotics can be achieved by detecting different *H. pylori* mutations or other genetic changes, but the accuracy of the molecular tools available vary widely between different antibiotics. Therefore, a standardized approach to determine antibiotic resistance that is easy to use is still not readily available. In addition, the standard method of obtaining tissue through endoscopy is cost prohibitive when considering the widespread presence of this organism in certain parts of the world; thus, new efforts are aimed at developing techniques that capitalize on DNA extraction and performing polymerase chain reaction (PCR) on stool samples from patients infected with *H. pylori*. These diagnostic modalities are still in early phases of development. In light of this, most recommendations for treatment of *H. pylori* outline that if individual susceptibility testing is not available, first-line therapy in areas of a high incidence of clarithromycin resistance (>15%) or in situations where resistance to clarithromycin is not known that bismuth quadruple therapy should be the first line of treatment. If not available, nonbismuth concomitant quadruple therapy may be a consideration. An algorithmic approach to the therapy of *H. pylori* as outlined in the Maastricht VI/Florence Consensus Report is provided in **Fig. 335-13**. Although resistance to metronidazole has been found in as many as 30% of isolates in North America and 80% in developing countries, triple therapy is effective in eradicating the organism in >50% of patients infected with a resistant strain. Clarithromycin resistance is seen in 13–16% of individuals in the United States, with resistance to amoxicillin being <1% and resistance to both metronidazole and clarithromycin in the 5% range. Resistance to tetracycline and rifabutin (see below) is reported to be <2% in the United States. In light of the paucity of *H. pylori* antibiotic real-time resistance data, asking the patient about prior antibiotic exposure should be included in the decision-making and used as a surrogate for potential antibiotic resistance, especially when it comes to prior macrolide use. Clarithromycin use should be excluded in patients with prior macrolide usage.

Failure of *H. pylori* eradication with triple therapy in a compliant patient is usually due to infection with a resistant organism. Additional important factors in treatment failure include inadequate acid suppression and inadequate adherence to the recommended regimen. Addressing both of these latter issues is critical before embarking on selecting alternative antibiotics for eradication of *H. pylori*. An in-depth review of the approach to patients with refractory *H. pylori* is beyond the scope of this chapter, and readers are referred to the American Gastroenterological Association 2021 guidelines (see Further Reading) for in-depth discussion of this important topic. A series of salvage therapies for *H. pylori* are shown in **Table 335-5**. The combination of PPI, amoxicillin, and rifabutin for 10 days has also been used successfully (86% cure rate) in patients infected with resistant strains. Additional regimens considered for second-line therapy include levofloxacin-based triple therapy (levofloxacin, amoxicillin, PPI) for 10 days and furazolidone-based triple therapy (furazolidone, amoxicillin, PPI) for 14 days. Unfortunately, no treatment regimen is universally accepted for patients in whom two courses of antibiotics have failed. If eradication is still not achieved in a compliant patient, then culture and sensitivity of the organism should be considered. One challenge with this approach is that culture and sensitivity testing is cumbersome and not widely available; thus, *H. pylori* resistance data within specific communities are often not available. Non-culture-based approaches using molecular markers to determine potential resistance through stool testing are being developed but are not widely available. Additional factors that may lower eradication rates include the patient's country of origin (higher in Northeast Asia than other parts of Asia or Europe) and cigarette smoking. In addition, meta-analysis suggests that even the most effective regimens (quadruple therapy including PPI, bismuth, tetracycline, and metronidazole and triple therapy including PPI, clarithromycin, and amoxicillin) may have suboptimal eradication rates (<80%), thus demonstrating the need for the development of more efficacious treatments.

In view of the observation that 15–25% of patients treated with first-line therapy may still remain infected with the organism, new approaches to treatment have been explored. One promising approach is sequential therapy. Regimens examined consist of 5 days of amoxicillin and a PPI, followed by an additional 5 days of PPI plus tinidazole and clarithromycin or levofloxacin. One promising regimen that has the benefit of being shorter in duration, easier to take, and less expensive is 5 days of concomitant therapy (PPI twice daily, amoxicillin 1 g twice daily, levofloxacin 500 mg twice daily, and tinidazole 500 mg twice daily). Initial studies have demonstrated eradication rates of >90% with good patient tolerance. Confirmation of these findings and applicability of this approach in the United States are needed, although some experts are recommending abandoning clarithromycin-based triple therapy in the United States for the concomitant therapy or the alternative sequential therapies highlighted above.

Innovative non-antibiotic-mediated approaches have been explored in an effort to improve eradication rates of *H. pylori*. Pretreatment of patients with *N*-acetylcysteine as a mucolytic agent to destroy the *H. pylori* biofilm and therefore impair antibiotic resistance has been examined, but more studies are needed to confirm the applicability of this approach. In vitro studies suggest that certain probiotics like *Lactobacillus* or its metabolites can inhibit *H. pylori.* Administration of probiotics has been attempted in several clinical studies in an effort to maximize antibiotic-mediated eradication with varying results. Overall, it appears that the use of certain probiotics, such as *Lactobacillus* spp., *Saccharomyces* spp., *Bifidobacterium* spp., and *Bacillus clausii*, did not alter eradication rates but importantly decreased antibiotic-associated side effects including nausea, dysgeusia, diarrhea, and abdominal discomfort/pain, resulting in enhanced tolerability of *H. pylori* therapies. Additional studies are needed to confirm the potential benefits of probiotics in this setting. Statins, specifically atorvastatin, have been used with some success as an adjunct to quadruple therapy in patients with NUD. Additional nonpharmacologic therapies are

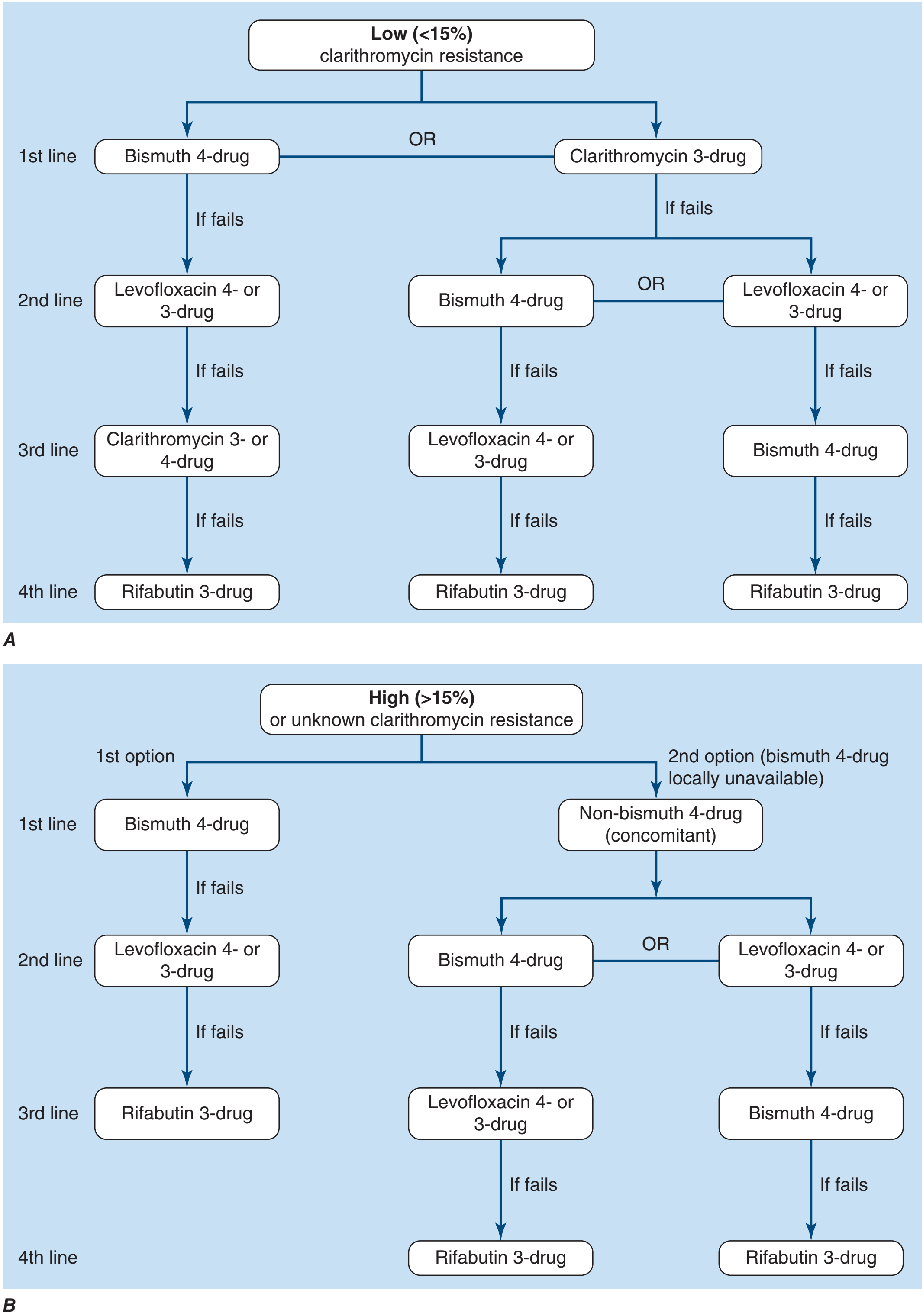
Low (<15%)
clarithromycin resistance
OR
1st line
Bismuth 4-drug
Clarithromycin 3-drug
If fails
If fails
2nd line
Levofloxacin 4- or 3-drug
Bismuth 4-drug
OR
Levofloxacin 4- or 3-drug
If fails
If fails
If fails
3rd line
Clarithromycin 3- or 4-drug
Levofloxacin 4- or 3-drug
Bismuth 4-drug
If fails
If fails
If fails
4th line
Rifabutin 3-drug
Rifabutin 3-drug
Rifabutin 3-drug
A
High (>15%)
or unknown clarithromycin resistance
1st option
2nd option (bismuth 4-drug locally unavailable)
1st line
Bismuth 4-drug
Non-bismuth 4-drug (concomitant)
If fails
2nd line
Levofloxacin 4- or 3-drug
Bismuth 4-drug
OR
Levofloxacin 4- or 3-drug
If fails
If fails
If fails
3rd line
Rifabutin 3-drug
Levofloxacin 4- or 3-drug
Bismuth 4-drug
If fails
If fails
4th line
Rifabutin 3-drug
Rifabutin 3-drug
B

FIGURE 335-13 Empirical approach to *Helicobacter pylori* eradication if individual antibiotic susceptibility testing is not avaliable. A. Low <15% clarithromycin resistance; B. High >15% or unknown clarithromycin resistance. Treatment regimens: Bismuth 4-drug—proton pump inhibitor (PPI), bismuth, tetracycline, metronidazole; Clarithromycin 3-drug—PPI, clarithromycin, amoxicillin (use only if clarithromycin sensitivity is known); Nonbismuth 4-drug—PPI, clarithromycin, amoxicillin, metronidazole; Levofloxacin 4-drug—PPI, levofloxacin, amoxicillin, bismuth; Levofloxacin 3-drug—same as levofloxacin 4-drug omitting bismuth; Rifabutin 3-drug in the setting of high fluoroquinolone resistance—rifabutin, high dose PPI, amoxicillin. (modified from Malfetheiner P, et al: Gut 71:9, 2022). *(Adapted from permission from P Malfertheiner et al: Management of Helicobacter pylori infection-the Maastricht V/Florence Consensus Report. Gut 71:9, 2022.)*

also being explored including oxygen-enriched environment or hyperbaric oxygen therapy, antimicrobial photodynamic therapy, nanomaterials, antimicrobial peptides, phase therapy, and modified lysins. Readers are referred to a recent comprehensive review on this topic (Luo et al in Further Reading).

Reinfection after successful eradication of *H. pylori* is rare in the United States (<1% per year). If recurrent infection occurs within the first 6 months after completing therapy, the most likely explanation is recrudescence as opposed to reinfection.

THERAPY OF NSAID-RELATED GASTRIC OR DUODENAL INJURY

Medical intervention for NSAID-related mucosal injury includes treatment of an active ulcer and primary prevention of future

TABLE 335-5 Salvage Therapies for *H. pylori* Infection

REGIMEN	DRUGS (DOSES)	DOSING FREQUENCY	DURATION (DAYS)	FDA APPROVAL
Bismuth quadruple	PPI (standard dose)	bid	14	No[a]
	Bismuth subcitrate (120–300 mg) or subsalicylate (300 mg)	qid		
	Tetracycline (500 mg)	qid		
	Metronidazole (500 mg)	tid or qid		
Levofloxacin triple	PPI (standard dose)	bid	14	No
	Levofloxacin (500 mg)	qd		
	Amoxicillin (1 g)	bid		
Concomitant	PPI (standard dose)	bid	10–14	No
	Clarithromycin (500 mg)	bid		
	Amoxicillin (1 g)	bid		
	Nitroimidazole (500 mg)	bid or tid		
Rifabutin triple	PPI (standard dose)	bid	10	No
	Rifabutin (300 mg)	qd		
	Amoxicillin (1 g)	bid		
High-dose dual	PPI (standard to double dose)	tid or qid	14	No
	Amoxicillin (1 g tid or 750 mg qid)	tid or qid		

[a] PPI, bismuth, tetracycline, and metronidazole prescribed separately is not an FDA-approved treatment regimen. However, Pylera, a combination product containing bismuth subcitrate, tetracycline, and metronidazole, combined with a PPI for 10 days is an FDA-approved treatment regimen.

Abbreviations: bid, twice daily; FDA, Food and Drug Administration; PPI, proton pump inhibitor; tid, three times daily; qd, once daily; qid, four times daily.

Source: Reproduced with permission from WD Chey et al: ACG clinical guideline: Treatment of *Helicobacter pylori* infection. Am J Gastroenterol 112:212, 2017.

injury. Recommendations for the treatment and primary prevention of NSAID-related mucosal injury are listed in Table 335-6. Ideally, the injurious agent should be stopped as the first step in the therapy of an active NSAID-induced ulcer. If that is possible, then treatment with one of the acid inhibitory agents (H_2 blockers, PPIs) is indicated. Cessation of NSAIDs is not always possible because of the patient's severe underlying disease. Only PPIs can heal GUs or DUs, independent of whether NSAIDs are discontinued.

The widespread use of NSAIDs has created some concern due to the increasing likelihood of GI and CV side effects associated with these agents. The approach to primary prevention has included avoiding the agent, using the lowest possible dose of the agent for the shortest period of time possible, using NSAIDs that are theoretically less injurious, using newer topical NSAID preparations, and/or using concomitant medical therapy to prevent NSAID-induced injury. Several nonselective NSAIDs that are associated with a lower likelihood of GI and CV toxicity include naproxen and ibuprofen, although the beneficial effect may be eliminated if higher dosages of the agents are used. Primary prevention of NSAID-induced ulceration can be accomplished by a PPI and, if not tolerated, misoprostol (200 μg qid). High-dose H_2 blockers (famotidine 40 mg bid) have also shown some promise in preventing endoscopically documented ulcers, although PPIs are superior. The highly selective COX-2 inhibitors, celecoxib and rofecoxib, are 100 times more selective inhibitors of COX-2 than standard NSAIDs, leading to gastric or duodenal mucosal injury that is comparable to placebo. Their utilization led to an increase in CV events initially leading to both celecoxib and rofecoxib being withdrawn from the market. After continued analysis of available data, the FDA concluded that celecoxib is no less safe than several nonopioid pain killers and has been allowed to remain on the market with a warning label that it was associated with increased risk of heart attacks and strokes. Additional caution was engendered when the CLASS study demonstrated that the advantage of celecoxib in preventing GI complications was offset when low-dose aspirin was used simultaneously. Therefore, gastric protection therapy is required in individuals taking COX-2 inhibitors and aspirin prophylaxis. Finally, much of the work demonstrating the benefit of COX-2 inhibitors and PPIs on GI injury has been performed in individuals of average risk; it is unclear if the same level of benefit will be achieved in high-risk patients. For example, concomitant use of warfarin and a COX-2 inhibitor was associated with rates of GI bleeding similar to those observed in patients taking nonselective NSAIDs. A combination of factors, including withdrawal of the majority of COX-2 inhibitors from the market, the observation that low-dose aspirin appears to diminish the beneficial effect of COX-2–selective inhibitors, and the growing use of aspirin for prophylaxis of CV events, has significantly altered the approach to gastric protective therapy during the use of NSAIDs. A set of guidelines for the approach to the use of NSAIDs was published by the ACG and is shown in Table 335-7.

TABLE 335-6 Recommendations for Treatment of NSAID-Related Mucosal Injury

CLINICAL SETTING	RECOMMENDATION
Active ulcer	
NSAID discontinued	H_2 receptor antagonist or PPI
NSAID continued	PPI
Prophylactic therapy	Misoprostol PPI Selective COX-2 inhibitor
H. pylori infection	Eradication if tests positive for *H. pylori*

Abbreviations: COX-2, isoenzyme of cyclooxygenase; NSAID, nonsteroidal anti-inflammatory drug; PPI, proton pump inhibitor.

TABLE 335-7 Guide to NSAID Therapy

	NO/LOW NSAID GI RISK	NSAID GI RISK
No CV risk (no aspirin)	Traditional NSAID	Coxib *or* Traditional NSAID + PPI *or* misoprostol Consider non-NSAID therapy
CV risk (consider aspirin)	Traditional NSAID + PPI *or* misoprostol if GI risk warrants gastroprotection Consider non-NSAID therapy	A gastroprotective agent must be added if a traditional NSAID is prescribed Consider non-NSAID therapy

Abbreviations: CV, cardiovascular; GI, gastrointestinal; NSAID, nonsteroidal anti-inflammatory drug; PPI, proton pump inhibitor.

Source: Reproduced with permission from AM Fendrick: Am J Manag Care 10:740, 2004.

Individuals who are not at risk for CV events, do not use aspirin, and are without risk for GI complications can receive nonselective NSAIDs without gastric protection. In those without CV risk factors but with a high potential risk (prior GI bleeding or multiple GI risk factors) for NSAID-induced GI toxicity, cautious use of a selective COX-2 inhibitor and co-therapy with high-dose PPI or misoprostol are recommended. Individuals at moderate GI risk without cardiac risk factors can be treated with a COX-2 inhibitor alone or with a nonselective NSAID with PPI or misoprostol. Individuals with CV risk factors, who require low-dose aspirin and have low potential for NSAID-induced toxicity, should be considered for a non-NSAID agent or use of a traditional NSAID such as naproxen (lower CV side effects) in combination with gastric protection, if warranted. Finally, individuals with CV and GI risks who require aspirin must be considered for non-NSAID therapy, but if that is not an option, then gastric protection with any type of NSAID must be considered. Any patient, regardless of risk status, who is being considered for long-term traditional NSAID therapy should also be considered for *H. pylori* testing and treatment if positive. Assuring the use of GI protective agents with NSAIDs is difficult, even in high-risk patients. This is in part due to under-prescribing of the appropriate protective agent; other times, the difficulty is related to patient compliance. The latter may be due to patients forgetting to take multiple pills or preferring not to take the extra pill, especially if they have no GI symptoms. Several NSAID gastroprotective-containing combination pills are now commercially available, including double-dose famotidine with ibuprofen, diclofenac with misoprostol, and naproxen with esomeprazole. Although initial studies suggested improved compliance and a cost advantage when taking these combination drugs, their clinical benefit over the use of separate pills has not been established. One additional concern with NSAID-induced GI complications is the relatively low rate of primary care provider compliance with established guidelines outlining preventative measures. An intervention including professional education, informatics to facilitate review, and financial incentives for practices to review patients' charts to assess appropriateness showed a reduced rate of high-risk prescribing of antiplatelet medications and NSAIDs with a tendency toward improved clinical outcomes. Efforts continue toward developing safer NSAIDs, including topical NSAIDs, NSAID formulations that are rapidly absorbed (diclofenac potassium powder mixed with a buffering agent, Prosorb and SoluMatrix technology), NO-releasing NSAIDs, hydrogen sulfide–releasing NSAIDs, dual COX/5-LOX inhibitors, NSAID prodrugs, and agents that can effectively sequester unbound NSAIDs without interfering with their efficacy.

APPROACH AND THERAPY: SUMMARY

Controversy continues regarding the best approach to the patient who presents with dyspepsia **(Chap. 48)**. The discovery of *H. pylori* and its role in pathogenesis of ulcers has added a new variable to the equation. Previously, if a patient <60 years of age presented with dyspepsia and without alarming signs or symptoms suggestive of an ulcer complication or malignancy, an empirical therapeutic trial with acid suppression was commonly recommended; today, however, a test-and-treat approach with a noninvasive diagnostic tool for *H. pylori* and eradication of the organism, if positive, is recommended **(Fig. 335-14)**.

Once an ulcer (GU or DU) is documented, the main issue to consider is whether *H. pylori* or an NSAID is involved. With *H. pylori* present, independent of the NSAID status, triple or quadruple therapy is recommended for 14 days, followed by continued acid-suppressing drugs (H_2 receptor antagonist or PPIs) for a total of 4–6 weeks. *H. pylori* eradication should be documented 4 weeks after completing antibiotics. The test of choice for documenting eradication is the laboratory-based validated monoclonal stool antigen test or a urea breath test (UBT). The patient must be off antisecretory agents for at least 7 days when being tested for eradication of *H. pylori* with UBT or stool antigen. Serologic testing is not useful for the purpose of documenting eradication because antibody titers fall slowly and often do not become undetectable. Some recommend that patients with complicated ulcer disease or who are frail should be treated with long-term acid suppression, thus making documentation of *H. pylori* eradication a moot point. In view of this discrepancy in practice, it would be best to discuss with the patient the different options available.

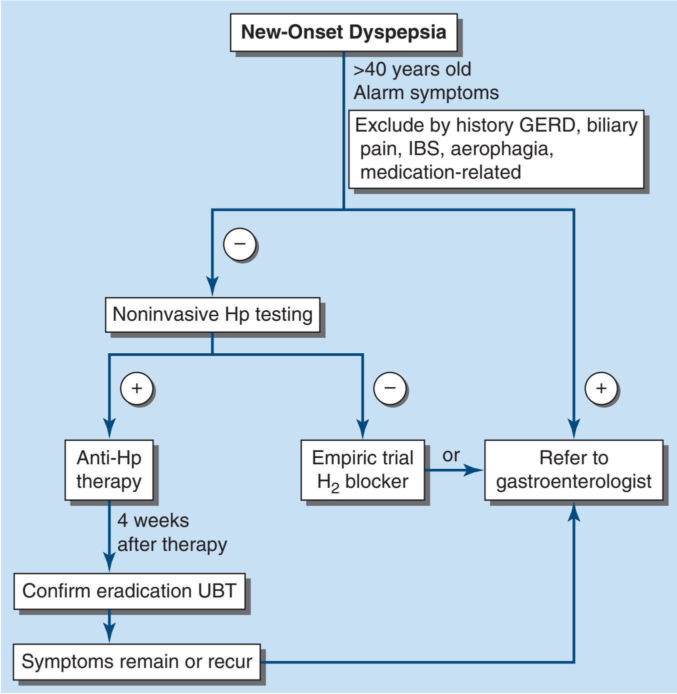

FIGURE 335-14 Overview of new-onset dyspepsia. GERD, gastroesophageal reflux disease; Hp, *Helicobacter pylori*; IBS, irritable bowel syndrome; UBT, urea breath test. *(Reproduced with permission from BS Anand, DY Graham: State-of-the-Art: Ulcer and Gastritis, Endoscopy 31:215, 1999. © Georg Thieme Verlag KG.)*

Several issues differentiate the approach to a GU versus a DU. GUs, especially of the body and fundus, have the potential of being malignant. Multiple biopsies of a GU should be taken initially; even if these are negative for neoplasm, repeat endoscopy to document healing at 8–12 weeks should be performed, with biopsy if the ulcer is still present. About 70% of GUs eventually found to be malignant undergo significant (usually incomplete) healing. Repeat endoscopy is warranted in patients with DU if symptoms persist despite medical therapy or a complication is suspected.

The majority (>90%) of GUs and DUs heal with the conventional therapy outlined above. A GU that fails to heal after 12 weeks and a DU that does not heal after 8 weeks of therapy should be considered refractory. Once poor compliance and persistent *H. pylori* infection have been excluded, NSAID use, either inadvertent or surreptitious, must be excluded. In addition, cigarette smoking must be eliminated. For a GU, malignancy must be meticulously excluded. Next, consideration should be given to a gastric acid hypersecretory state such as ZES (see "Zollinger-Ellison Syndrome," below) or the idiopathic form, which can be excluded with gastric acid analysis. Although a subset of patients has gastric acid hypersecretion of unclear etiology as a contributing factor to refractory ulcers, ZES should be excluded with a fasting gastrin or secretin stimulation test (see below). More than 90% of refractory ulcers (either DUs or GUs) heal after 8 weeks of treatment with higher doses of PPI (omeprazole 40 mg/d; lansoprazole 30–60 mg/d). This higher dose is also effective in maintaining remission. Surgical intervention may be a consideration at this point; however, other rare causes of refractory ulcers must be excluded before recommending surgery. Rare etiologies of refractory ulcers that may be diagnosed by gastric or duodenal biopsies include ischemia, Crohn's disease, amyloidosis, sarcoidosis, lymphoma, eosinophilic gastroenteritis, smoking crack cocaine, or infection (cytomegalovirus [CMV], tuberculosis, or syphilis).

SURGICAL THERAPY

Surgical intervention in PUD can be viewed as being either elective, for treatment of medically refractory disease, or as urgent/emergent, for the treatment of an ulcer-related complication. The development of pharmacologic and endoscopic approaches for the treatment of peptic disease and its complications has led to a substantial decrease in the number of operations needed for this disorder with a drop of >90% for elective ulcer surgery over the past four decades. Refractory ulcers are an exceedingly rare occurrence. Surgery is more often required for treatment of an ulcer-related complication.

Hemorrhage is the most common ulcer-related complication, occurring in ~15–25% of patients. Bleeding may occur in any age group but is most often seen in older patients (sixth decade or beyond). The majority of patients stop bleeding spontaneously, but endoscopic therapy **(Chap. 333)** is necessary in some. Parenterally and orally administered PPIs also decrease ulcer rebleeding in patients who have undergone endoscopic therapy. Patients unresponsive or refractory to endoscopic intervention will require angiographic intervention or surgery (~5% of transfusion-requiring patients).

Free peritoneal perforation occurs in ~2–3% of DU patients, with NSAID-induced GU perforations occurring more commonly. Sudden onset of severe abdominal pain with peritoneal signs and evidence of pneumoperitoneum on abdominal imaging is the classic presentation of a perforated viscous, but this presentation occurs in only two-thirds of patients. The latter is especially true in elderly patients (>70 years old), obese individuals, and immunocompromised patients. It is important to keep in mind that, as in the case of bleeding, up to 10% of these patients will not have antecedent ulcer symptoms. Delay in diagnosis clearly leads to higher mortality; thus, early suspicion and intervention with nasogastric suction, intravenous PPI, antibiotics, and surgical consultation are essential. Concomitant bleeding may occur in up to 10% of patients with perforation, with mortality being increased substantially. Peptic ulcer can also penetrate into adjacent organs, especially with a posterior DU, which can penetrate into the pancreas, colon, liver, or biliary tree.

Pyloric channel ulcers or DUs can lead to gastric outlet obstruction in ~2–3% of patients. This can result from chronic scarring or from impaired motility due to inflammation and/or edema with pylorospasm. Patients may present with early satiety, nausea, vomiting of undigested food, and weight loss. Conservative management with nasogastric suction, intravenous hydration/nutrition, and antisecretory agents is indicated for 7–10 days with the hope that a functional obstruction will reverse. If a mechanical obstruction persists, endoscopic intervention with balloon dilation may be effective. Surgery should be considered if all else fails.

Specific Operations for Duodenal Ulcers Surgical treatment was originally designed to decrease gastric acid secretion. Operations most commonly performed include (1) vagotomy and drainage (by pyloroplasty, gastroduodenostomy, or gastrojejunostomy), (2) highly selective vagotomy (which does not require a drainage procedure), and (3) vagotomy with antrectomy. The specific procedure performed is dictated by the underlying circumstances: elective versus emergency, the degree and extent of duodenal ulceration, the etiology of the ulcer (*H. pylori*, NSAIDs, malignancy), and the expertise of the surgeon. Moreover, the trend has been toward a dramatic decrease in the need for surgery for treatment of refractory PUD, and when needed, minimally invasive and anatomy-preserving operations are preferred.

Vagotomy is a component of each of these procedures and is aimed at decreasing acid secretion through ablating cholinergic input to the stomach. Unfortunately, both truncal and selective vagotomy (preserves the celiac and hepatic branches) result in gastric atony despite successful reduction of both basal acid output (BAO; decreased by 85%) and maximal acid output (MAO; decreased by 50%). Drainage through pyloroplasty or gastroduodenostomy is required in an effort to compensate for the vagotomy-induced gastric motility disorder. This procedure has an intermediate complication rate and a 10% ulcer recurrence rate. To minimize gastric dysmotility, highly selective vagotomy (also known as parietal cell, super-selective, or proximal vagotomy) was developed. Only the vagal fibers innervating the portion of the stomach that contains parietal cells are transected, thus leaving fibers important for regulating gastric motility intact. Although this procedure leads to an immediate decrease in both BAO and stimulated acid output, acid secretion recovers over time. By the end of the first postoperative year, basal and stimulated acid output are ~30 and 50%, respectively, of preoperative levels. Ulcer recurrence rates are higher with highly selective vagotomy (≥10%), although the overall complication rates are the lowest of the three procedures.

The procedure that provides the lowest rates of ulcer recurrence (1%) but has the highest complication rate is vagotomy (truncal or selective) in combination with antrectomy. Antrectomy is aimed at eliminating an additional stimulant of gastric acid secretion, gastrin. Two principal types of reanastomoses are used after antrectomy: gastroduodenostomy (Billroth I) or gastrojejunostomy (Billroth II) **(Fig. 335-15)**. Although Billroth I is often preferred over II, severe duodenal inflammation or scarring may preclude its performance. Prospective, randomized studies confirm that partial gastrectomy followed by Roux-en-Y reconstruction leads to a significantly better clinical, endoscopic, and histologic outcome than Billroth II reconstruction.

Of these procedures, highly selective vagotomy may be the procedure of choice in the elective setting, except in situations where ulcer recurrence rates are high (prepyloric ulcers and those refractory to medical therapy). Selection of vagotomy and antrectomy may be more appropriate in these circumstances.

These procedures have been traditionally performed by standard laparotomy. The advent of laparoscopic surgery has led several surgical teams to successfully perform highly selective vagotomy, truncal vagotomy/pyloroplasty, and truncal vagotomy/antrectomy

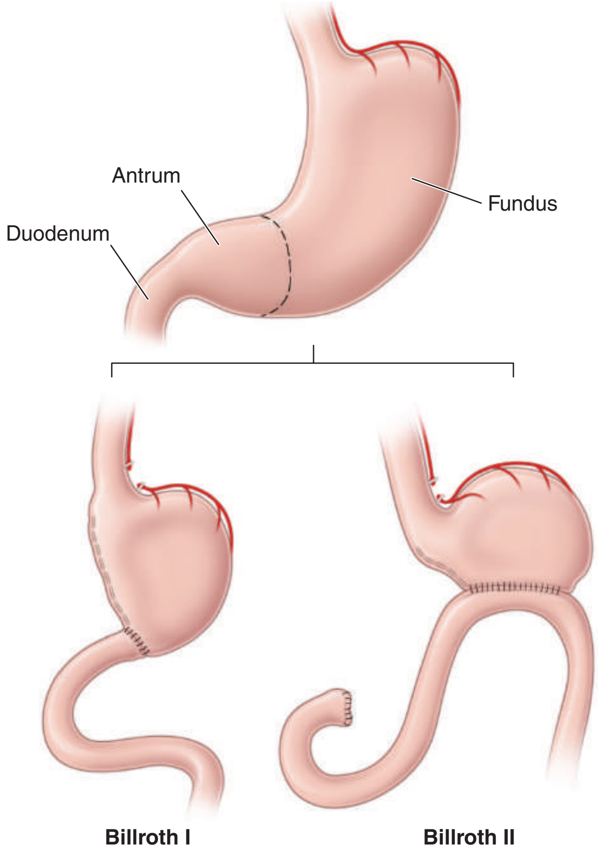

FIGURE 335-15 Schematic representation of Billroth I and II procedures.

through this approach. An increase in the number of laparoscopic procedures for treatment of PUD has occurred. Laparoscopic repair of perforated peptic ulcers is safe, feasible for the experienced surgeon, and associated with decreased postoperative pain, although it does take longer than an open approach. Moreover, no difference between the two approaches is noted in postoperative complications or length of hospital stay.

Specific Operations for GUs The location and presence of a concomitant DU dictate the operative procedure performed for a GU. Antrectomy (including the ulcer) with a Billroth I anastomosis is the treatment of choice for an antral ulcer. Vagotomy is performed only if a DU is present. Although ulcer excision with vagotomy and drainage procedure has been proposed, the higher incidence of ulcer recurrence makes this a less desirable approach. Ulcers located near the esophagogastric junction may require a more radical approach, a subtotal gastrectomy with a Roux-en-Y esophagogastrojejunostomy (Csendes' procedure). A less aggressive approach, including antrectomy, intraoperative ulcer biopsy, and vagotomy (Kelling-Madlener procedure), may be indicated in fragile patients with a high GU. Ulcer recurrence approaches 30% with this procedure.

Surgery-Related Complications Complications seen after surgery for PUD are related primarily to the extent of the anatomic modification performed. Minimal alteration (highly selective vagotomy) is associated with higher rates of ulcer recurrence and less GI disturbance. More aggressive surgical procedures have a lower rate of ulcer recurrence but a greater incidence of GI dysfunction. Overall, morbidity and mortality related to these procedures are quite low. Morbidity associated with vagotomy and antrectomy or pyloroplasty is ≤5%, with mortality ~1%. Highly selective vagotomy has lower morbidity and mortality rates of 1 and 0.3%, respectively.

In addition to the potential early consequences of any intraabdominal procedure (bleeding, infection, thromboembolism), gastroparesis, duodenal stump leak, and efferent loop obstruction can be observed.

Recurrent Ulceration The risk of ulcer recurrence is directly related to the procedure performed. Ulcers that recur after partial gastric resection tend to develop at the anastomosis (stomal or marginal ulcer). Epigastric abdominal pain is the most frequent presenting complaint (>90%). Severity and duration of pain tend to be more progressive than observed with DUs before surgery.

Ulcers may recur for several reasons, including incomplete vagotomy, inadequate drainage, retained antrum, and, less likely, persistent or recurrent *H. pylori* infection. ZES should have been excluded preoperatively. Surreptitious use of NSAIDs is an important reason for recurrent ulcers after surgery, especially if the initial procedure was done for an NSAID-induced ulcer. Once *H. pylori* and NSAIDs have been excluded as etiologic factors, the question of incomplete vagotomy or retained gastric antrum should be explored. For the latter, fasting plasma gastrin levels should be determined. If elevated, retained antrum or ZES (see below) should be considered. Incomplete vagotomy can be ruled out by gastric acid analysis coupled with sham feeding. In this test, gastric acid output is measured while the patient sees, smells, and chews a meal (without swallowing). The cephalic phase of gastric secretion, which is mediated by the vagus, is being assessed with this study. An increase in gastric acid output in response to sham feeding is evidence that the vagus nerve is intact. A rise in serum pancreatic polypeptide >50% within 30 min of sham feeding is also suggestive of an intact vagus nerve.

Medical therapy with H_2 blockers will heal postoperative ulceration in 70–90% of patients. The efficacy of PPIs has not been fully assessed in this group, but one may anticipate greater rates of ulcer healing compared to those obtained with H_2 blockers. Repeat operation (complete vagotomy, partial gastrectomy) may be required in a small subgroup of patients who have not responded to aggressive medical management.

Afferent Loop Syndromes Although rarely seen today as a result of the decrease in the performance of Billroth II anastomosis, two types of afferent loop syndrome can occur in patients who have undergone this type of partial gastric resection. The more common of the two is bacterial overgrowth in the afferent limb secondary to stasis. Patients may experience postprandial abdominal pain, bloating, and diarrhea with concomitant malabsorption of fats and vitamin B_{12}. Cases refractory to antibiotics may require surgical revision of the loop. The less common afferent loop syndrome can present with severe abdominal pain and bloating that occur 20–60 min after meals. Pain is often followed by nausea and vomiting of bile-containing material. The pain and bloating may improve after emesis. The cause of this clinical picture is theorized to be incomplete drainage of bile and pancreatic secretions from an afferent loop that is partially obstructed. Cases refractory to dietary measures may need surgical revision or conversion of the Billroth II anastomosis to a Roux-en-Y gastrojejunostomy.

Dumping Syndrome Dumping syndrome consists of a series of vasomotor and GI signs and symptoms and occurs in patients who have undergone vagotomy and drainage (especially Billroth procedures). Two phases of dumping, early and late, can occur. Early dumping takes place 15–30 min after meals and consists of crampy abdominal discomfort, nausea, diarrhea, belching, tachycardia, palpitations, diaphoresis, light-headedness, and, rarely, syncope. These signs and symptoms arise from the rapid emptying of hyperosmolar gastric contents into the small intestine, resulting in a fluid shift into the gut lumen with plasma volume contraction and acute intestinal distention. Release of vasoactive GI hormones (vasoactive intestinal polypeptide, neurotensin, motilin) is also theorized to play a role in early dumping.

The late phase of dumping typically occurs 90 min to 3 h after meals. Vasomotor symptoms (light-headedness, diaphoresis, palpitations, tachycardia, and syncope) predominate during this phase. This component of dumping is thought to be secondary to hypoglycemia from excessive insulin release.

Dumping syndrome is most noticeable after meals rich in simple carbohydrates (especially sucrose) and high osmolarity. Ingestion of large amounts of fluids may also contribute. After vagotomy and drainage, up to 50% of patients will experience dumping syndrome to some degree early on. Signs and symptoms often improve with time, but a severe protracted picture can occur in up to 1% of patients.

Dietary modification is the cornerstone of therapy for patients with dumping syndrome. Small, multiple (six) meals devoid of simple carbohydrates coupled with elimination of liquids during meals is important. Antidiarrheals and anticholinergic agents are complementary to diet. Guar and pectin, which increase the viscosity of intraluminal contents, may be beneficial in more symptomatic individuals. Acarbose, an α-glucosidase inhibitor that delays digestion of ingested carbohydrates, has also been shown to be beneficial in the treatment of the late phases of dumping. The somatostatin analogue octreotide has been successful in diet-refractory cases. This drug is administered subcutaneously (50 μg tid), titrated according to clinical response. A long-acting depot formulation of octreotide can be administered once every 28 days and provides symptom relief comparable to the short-acting agent. In addition, patient weight gain and quality of life appear to be superior with the long-acting form.

Postvagotomy Diarrhea Up to 10% of patients may seek medical attention for the treatment of postvagotomy diarrhea. This complication is most commonly observed after truncal vagotomy, which is rarely performed today. Patients may complain of intermittent diarrhea that occurs typically 1–2 h after meals. Occasionally, the symptoms may be severe and relentless. This is due to a motility disorder from interruption of the vagal fibers supplying the luminal gut. Other contributing factors may include decreased absorption of nutrients (see below), increased excretion of bile acids, and release of luminal factors that promote secretion. Diphenoxylate or loperamide is often useful in symptom control. The bile salt–binding agent cholestyramine may be helpful in severe cases. Surgical reversal of a 10-cm segment of jejunum may yield a substantial improvement in bowel frequency in a subset of patients.

Bile Reflux Gastropathy A subset of post–partial gastrectomy patients who present with abdominal pain, early satiety, nausea, and vomiting will have mucosal erythema of the gastric remnant as the only finding. Histologic examination of the gastric mucosa reveals minimal inflammation but the presence of epithelial cell injury. This clinical picture is categorized as bile or alkaline reflux gastropathy/gastritis. Although reflux of bile is implicated as the reason for this disorder, the mechanism is unknown. Prokinetic agents, cholestyramine, and sucralfate have been somewhat effective treatments. Severe refractory symptoms may require using either nuclear scanning with ^{99m}Tc-HIDA to document reflux. Surgical diversion of pancreaticobiliary secretions away from the gastric remnant with a Roux-en-Y gastrojejunostomy consisting of a long (50–60 cm) Roux limb has been used in severe cases. Bilious vomiting improves, but early satiety and bloating may persist in up to 50% of patients.

Maldigestion and Malabsorption Weight loss can be observed in up to 60% of patients after partial gastric resection. Patients can experience a 10% loss of body weight, which stabilizes 3 months postoperatively. A significant component of this weight reduction is due to decreased oral intake. However, mild steatorrhea can also develop. Reasons for maldigestion/malabsorption include decreased gastric acid production, rapid gastric emptying, decreased food dispersion in the stomach, reduced luminal bile concentration, reduced pancreatic secretory response to feeding, and rapid intestinal transit.

Decreased serum vitamin B_{12} levels can be observed after partial gastrectomy. This is usually not due to deficiency of IF, since a minimal amount of parietal cells (source of IF) is removed during antrectomy. Reduced vitamin B_{12} may be due to competition for the vitamin by bacterial overgrowth or inability to split the vitamin from its protein-bound source due to hypochlorhydria.

Iron-deficiency anemia may be a consequence of impaired absorption of dietary iron in patients with a Billroth II gastrojejunostomy. Absorption of iron salts is normal in these individuals; thus, a favorable response to oral iron supplementation can be anticipated. Folate deficiency with concomitant anemia can also develop in these patients. This deficiency may be secondary to decreased absorption or diminished oral intake.

Malabsorption of vitamin D and calcium resulting in osteoporosis and osteomalacia is common after partial gastrectomy and gastrojejunostomy (Billroth II). Osteomalacia can occur as a late complication in up to 25% of post–partial gastrectomy patients. Bone fractures occur twice as commonly in men after gastric surgery as in a control population. It may take years before x-ray findings demonstrate diminished bone density. Elevated alkaline phosphatase, reduced serum calcium, bone pain, and pathologic fractures may be seen in patients with osteomalacia. The high incidence of these abnormalities in this subgroup of patients justifies treating them with vitamin D and calcium supplementation indefinitely. Therapy is especially important in females. Copper deficiency has also been reported in patients undergoing surgeries that bypass the duodenum, where copper is primarily absorbed. Patients may present with a rare syndrome that includes ataxia, myelopathy, and peripheral neuropathy.

Gastric Adenocarcinoma The incidence of adenocarcinoma in the gastric stump is increased 15 years after resection. Some have reported a four- to fivefold increase in gastric cancer 20–25 years after resection. The pathogenesis is unclear but may involve alkaline reflux, bacterial proliferation, or hypochlorhydria. The role of endoscopic screening is not clear, and most guidelines do not support its use.

Additional Complications Reflux esophagitis and a higher incidence of gallstones and cholecystitis have been reported in patients undergoing subtotal gastrectomy. The latter is thought to be due to decreased gallbladder contractility associated with vagotomy and bypass of the duodenum, leading to decreased postprandial release of cholecystokinin.

RELATED CONDITIONS

ZOLLINGER-ELLISON SYNDROME

Severe peptic ulcer diathesis secondary to gastric acid hypersecretion due to unregulated gastrin release from a non–β-cell, often well-differentiated neuroendocrine tumor (NET; gastrinoma) defines the components of ZES. Initially, ZES was typified by aggressive and refractory ulceration in which total gastrectomy provided the only chance for enhancing survival. Today, it can be cured by surgical resection in up to 40% of patients with the sporadic form of the disease (see below).

Epidemiology The true incidence of ZES is unknown, but estimates suggest that it varies from 0.1 to 1% of individuals presenting with PUD, with 0.1–3 individuals per year having this rare diagnosis. Others have estimated an incidence of 0.5–3 per million population. Females are slightly more commonly affected than males, and the majority of patients are diagnosed between ages 30 and 65. In addition, the time of diagnosis is estimated to be between 4 and 7 years after the onset of symptoms. Gastrinomas are classified into sporadic tumors (80%) and those associated with multiple endocrine neoplasia (MEN) type 1 (see below). The widespread availability and use of PPIs have led to a decreased patient referral for gastrinoma evaluation, delay in diagnosis, and an increase in false-positive diagnoses of ZES. In fact, diagnosis may be delayed for ≥6 years after symptoms consistent with ZES are displayed.

Pathophysiology Hypergastrinemia originating from an autonomous neoplasm is the driving force responsible for the clinical manifestations in ZES. Gastrin stimulates acid secretion through gastrin receptors on parietal cells and by inducing histamine release from ECL cells. Gastrin also has a trophic action on gastric epithelial cells. Longstanding hypergastrinemia leads to markedly increased gastric acid secretion through both parietal cell stimulation and increased parietal cell mass. The increased gastric acid output leads to peptic ulcer diathesis, erosive esophagitis, and diarrhea.

Tumor Distribution Although early studies suggested that the vast majority of gastrinomas occurred within the pancreas, a significant number of these lesions are extrapancreatic. Between 60 and 90% of these tumors are found within the hypothetical gastrinoma triangle (confluence of the cystic and common bile ducts superiorly, junction of the second and third portions of the duodenum inferiorly, and junction of the neck and body of the pancreas medially). Duodenal tumors constitute the most common nonpancreatic lesion; between 60 and 100% of gastrinomas are found here. Duodenal tumors are smaller, slower growing, and less likely to metastasize than pancreatic lesions and occur primarily in the first and second portion of the duodenum (90%). Less common extrapancreatic sites include stomach, bones, ovaries, heart, liver, and lymph nodes. More than 60% of tumors are considered malignant (determined by local invasion and/or evidence of metastasis), with up to 30–50% of patients having multiple lesions or metastatic disease at presentation. Histologically, gastrin-producing cells appear well-differentiated (grade 1 or 2 histologically), expressing markers typically found in endocrine neoplasms (chromogranin, neuron-specific enolase). Although not clearly established in gastrinomas, histologic grade in pancreatic NETs generally is an important predictor of survival in these rare neoplasms **(Chap. 89)**.

Clinical Manifestations Gastric acid hypersecretion is responsible for the signs and symptoms observed in patients with ZES. The most common clinical presentation for gastrinoma patients is abdominal pain in the presence of acid peptic disorders. Peptic ulcer is the most common clinical manifestation, occurring in >90% of gastrinoma patients. Initial presentation and ulcer location (duodenal bulb) may be indistinguishable from common PUD. Clinical situations that should create suspicion of gastrinoma are ulcers in unusual locations (second part of the duodenum and beyond), ulcers refractory to standard medical therapy, ulcer recurrence after acid-reducing surgery, ulcers presenting with frank complications (bleeding, obstruction, and perforation), or ulcers in the absence of *H. pylori* or NSAID ingestion. Symptoms of esophageal origin are present in up to two-thirds of patients with ZES, with a spectrum ranging from mild esophagitis to frank ulceration with stricture and Barrett's mucosa.

Diarrhea, the next most common clinical manifestation, is found in up to 70% of patients. Although diarrhea often occurs concomitantly with acid peptic disease, it may also occur independent of an ulcer and classically will abate with PPI therapy. Etiology of the diarrhea is multifactorial, resulting from marked volume overload to the small bowel, pancreatic enzyme inactivation by acid, and damage of the intestinal epithelial surface by acid. The epithelial damage can lead to a mild degree of maldigestion and malabsorption of nutrients. The diarrhea may also have a secretory component due to the direct stimulatory effect of gastrin on enterocytes or the co-secretion of additional hormones from the tumor such as vasoactive intestinal peptide.

Gastrinomas can develop in the presence of MEN 1 syndrome **(Chaps. 89 and 400)** in ~25% of patients. This autosomal dominant disorder involves primarily three organ sites: the parathyroid glands (80–90%), pancreas (40–80%), and pituitary gland (30–60%). The syndrome is caused by inactivating mutations of the *MEN1* tumor-suppressor gene found on the long arm of chromosome 11q13. The gene encodes for menin, which has an important role in DNA replication and transcriptional regulation. A genetic diagnosis is obtained by sequencing of the *MEN1* gene, which can reveal mutations in 70–90% of typical MEN 1 cases. A family may have an unknown mutation, making a genetic diagnosis impossible, and therefore, certain individuals will require a clinical diagnosis, which is determined by whether a patient has tumors in two of the three endocrine organs (parathyroid, pancreas/duodenum, or pituitary) or has a family history of MEN 1 and one of the endocrine organ tumors. In view of the stimulatory effect of calcium on gastric secretion, the hyperparathyroidism and hypercalcemia seen in MEN 1 patients may have a direct effect on ulcer disease. Resolution of hypercalcemia by parathyroidectomy reduces gastrin and gastric acid output in gastrinoma patients. An additional distinguishing feature in ZES patients with MEN 1 is the higher incidence of gastric carcinoid tumor development (as compared to patients with sporadic gastrinomas). ZES presents and is diagnosed earlier in MEN 1 patients, and they have a more indolent course as compared to patients with sporadic gastrinoma. Gastrinomas tend to be smaller, multiple, and located in the duodenal wall more often than is seen in patients with sporadic ZES. Establishing the diagnosis of MEN 1 is critical in order to provide genetic counseling to the patient and their family and also to determine the recommended surgical approach. Therefore, gastrinoma patients should be screened for MEN 1 by performing a detailed family history and obtaining several serum markers including calcium, parathyroid, prolactin, and pancreatic polypeptide levels.

TABLE 335-8 When to Obtain a Fasting Serum Gastrin Level

Multiple ulcers
Ulcers in unusual locations; associated with severe esophagitis; resistant to therapy with frequent recurrences; in the absence of nonsteroidal anti-inflammatory drug ingestion or *Helicobacter pylori* infection
Ulcer patients awaiting surgery
Severe or refractory GERD
GERD associated with diarrhea
Extensive family history for peptic ulcer disease
Postoperative ulcer recurrence
Basal hyperchlorhydria
Unexplained diarrhea or steatorrhea
Diarrhea improved with PPI
Hypercalcemia
Family history of pancreatic islet, pituitary, or parathyroid tumor
Prominent gastric or duodenal folds

Diagnosis Establishing an early diagnosis is important in order to minimize the long-term sequelae of gastric acid hypersecretion, prevent metastatic disease, and counsel family members if a diagnosis of MEN 1 is established. Biochemical measurements of gastrin and acid secretion in patients suspected of having ZES play an important role is establishing this rare diagnosis. Often, patients suspected of having ZES will be treated with a PPI in an effort to ameliorate symptoms and decrease the likelihood of possible acid-related complications. The presence of the PPI, which will lower acid secretion and potentially elevate fasting gastrin levels in normal individuals, will make the diagnostic approach in these individuals somewhat difficult. Significant morbidity related to peptic diathesis has been described when stopping PPIs in gastrinoma patients; therefore, a systematic approach in stopping these agents is warranted (see below). The first step in the evaluation of a patient suspected of having ZES is to obtain a fasting gastrin level. A list of clinical scenarios that should arouse suspicion regarding this diagnosis is shown in **Table 335-8**. Fasting gastrin levels obtained using a dependable assay are usually <150 pg/mL. A normal fasting gastrin, on two separate occasions, especially if the patient is on a PPI, virtually excludes this diagnosis. Virtually all gastrinoma patients will have a gastrin level >150–200 pg/mL. Measurement of fasting gastrin should be repeated to confirm the clinical suspicion. Some of the commercial biochemical assays used for measuring serum gastrin may be inaccurate. Variable specificity of the antibodies used have led to both false-positive and false-negative fasting gastrin levels, placing in jeopardy the ability to make an accurate diagnosis of ZES.

Multiple processes can lead to an elevated fasting gastrin level, the most frequent of which are gastric hypochlorhydria and achlorhydria, with or without pernicious anemia. Gastric acid induces feedback inhibition of gastrin release. A decrease in acid production will subsequently lead to failure of the feedback inhibitory pathway, resulting in net hypergastrinemia. Gastrin levels will thus be high in patients using antisecretory agents for the treatment of acid peptic disorders and dyspepsia. *H. pylori* infection can also cause hypergastrinemia. Additional causes of elevated gastrin include retained gastric antrum; G-cell hyperplasia; gastric outlet obstruction; renal insufficiency; massive small-bowel obstruction; and conditions such as rheumatoid arthritis, vitiligo, diabetes mellitus, and pheochromocytoma. Although a fasting gastrin >10 times normal is highly suggestive of ZES, two-thirds of patients will have fasting gastrin levels that overlap with levels found in the more common disorders outlined above, especially if a PPI is being taken by the patient. The effect of the PPI on gastrin levels and acid secretion will linger several days after stopping the PPI; therefore, it should be stopped for a minimum of 7 days before testing. During this period, the patient should be placed on a histamine H_2 antagonist, such as famotidine, twice to three times per day. Although this type of agent has a short-term effect on gastrin and acid secretion, it needs to be stopped 24 h before repeating fasting gastrin levels or performing some of the tests highlighted below. The patient may take antacids for the final day, stopping them ~12 h before testing is performed. Heightened awareness of complications related to gastric acid hypersecretion during the period of PPI cessation is critical.

Historically the next study typically recommended for establishing a biochemical diagnosis of gastrinoma is to assess acid secretion. Unfortunately, few centers perform this type of testing; thus, only a brief summary will be provided here. Typically, nothing further needs to be done if decreased acid output in the absence of a PPI is observed. A pH can be measured on gastric fluid obtained either during endoscopy or through nasogastric aspiration; a pH <3 is suggestive of a gastrinoma, but a pH >3 is not helpful in excluding the diagnosis. In those situations where the pH is >3, formal gastric acid analysis should be performed if available. Normal BAO in nongastric surgery patients is typically <5 meq/h. A BAO >15 meq/h in the presence of hypergastrinemia is considered pathognomonic of ZES, but up to 12% of patients with common PUD may have elevated BAO to a lesser degree that can overlap with levels seen in ZES patients. In an effort to improve the sensitivity and specificity of gastric secretory studies, a BAO/MAO ratio was established using pentagastrin infusion as a way to maximally stimulate acid production, with a BAO/MAO ratio >0.6 being highly suggestive of ZES. Pentagastrin is no longer available in the United States, making measurement of MAO virtually impossible. An endoscopic method for measuring gastric acid output has been developed but requires further validation.

Gastrin provocative tests have been developed in an effort to differentiate between the causes of hypergastrinemia and are especially helpful in patients with indeterminate acid secretory studies. The tests are the secretin stimulation test and the calcium infusion study; the latter is rarely, if ever, utilized in our current environment due to the cumbersome nature of the test and its lower sensitivity and specificity than secretin stimulation. The most sensitive and specific gastrin provocative test for the diagnosis of gastrinoma is the secretin study. An increase in gastrin of ≥120 pg within 15 min of secretin injection has a sensitivity and specificity of >90% for ZES. PPI-induced hypochlorhydria or achlorhydria may lead to a false-positive secretin test; thus, this agent must be stopped for 1 week before testing.

In light of the limited availability of the biochemical studies outlined above, more studies make a diagnosis of gastrinoma based on the presence of elevated gastrin and low gastric pH in the right clinical setting coupled with tumor localization tests outlined below and positive histology by biopsy (difficult to obtain). Revised guidelines for the best approach to establishing a diagnosis of gastrinoma taking into consideration the above outlined limitations are being considered, but none have replaced the established guidelines outlined earlier in this section (see Jensen et al in Further Reading).

Tumor Localization Once the biochemical diagnosis of gastrinoma has been confirmed (if possible), the tumor must be located. Multiple imaging studies have been used in an effort to enhance tumor localization **(Table 335-9)**. The broad range of sensitivity is due to the variable success rates achieved by the different investigative groups. Endoscopic ultrasound (EUS) permits imaging of the pancreas with a high degree of resolution (<5 mm). This modality is particularly helpful in excluding small neoplasms within the pancreas and in assessing the presence of surrounding lymph nodes and vascular involvement, but it is not very sensitive (43%) for finding duodenal lesions. This latter observation has led some to not include EUS in the routine preoperative evaluation of a patient suspected of having a gastrinoma. Several types of endocrine tumors express cell-surface receptors for somatostatin, in particular the subtype 2 (SSTR2). This permits the localization, staging, and prediction of therapeutic response to somatostatin analogues (see below) by gastrinomas. The original functional scinitigraphic tool developed measuring the uptake of the stable somatostatin analogue ^{111}In-pentetreotide (OctreoScan) has demonstrated sensitivity and specificity rates of >80%. Positron emission tomography (PET)–computed tomography (CT) with ^{68}Ga-DOTATATE has been developed and is superior than OctreoScan for assessing tumor presence in patients with well-differentiated NETs such as gastrinomas, with sensitivity and specificity of >90%, making it the functional imaging study of choice when available. ^{18}F-Fluordeoxyglucose (^{18}F-FDG) PET imaging has been found to be useful in pancreatic NETs, including gastrinomas, particularly as a prognostic marker.

TABLE 335-9 Sensitivity of Imaging Studies in Zollinger-Ellison Syndrome

	SENSITIVITY, %	
STUDY	PRIMARY GASTRINOMA	HEPATIC METASTATIC GASTRINOMA
Ultrasound	0–28	15–77
CT scan	0–59	99–100
Selective angiography	35–68	96–100
Portal venous sampling	70–90	N/A
SASI	55–78	N/A
MRI	20–25	88–100
OctreoScan	55–77	90–100
EUS	28–86	N/A

Abbreviations: CT, computed tomography; EUS, endoscopic ultrasonography; MRI, magnetic resonance imaging; N/A, not applicable; OctreoScan, imaging with ^{111}In-pentetreotide; SASI, selective arterial secretin injection.

Up to 50% of patients have metastatic disease at diagnosis. Success in controlling gastric acid hypersecretion has shifted the emphasis of therapy toward providing a surgical cure. Detecting the primary tumor and excluding metastatic disease are critical in view of this paradigm shift. Once a biochemical diagnosis has been confirmed, the patient should first undergo an abdominal CT scan, magnetic resonance imaging (MRI), or OctreoScan/PET-CT with ^{68}Ga-DOTATATE (depending on availability) to exclude metastatic disease. Once metastatic disease has been excluded, an experienced endocrine surgeon may opt for exploratory laparotomy with intraoperative ultrasound or transillumination. In other centers, careful examination of the peripancreatic area with EUS, accompanied by endoscopic exploration of the duodenum for primary tumors, will be performed before surgery. Selective arterial secretin injection may be a useful adjuvant for localizing tumors in a subset of patients. The extent of the diagnostic and surgical approach must be carefully balanced with the patient's overall physiologic condition and the natural history of a slow-growing gastrinoma.

TREATMENT

Zollinger-Ellison Syndrome

Treatment of functional endocrine tumors is directed at ameliorating the signs and symptoms related to hormone overproduction, curative resection of the neoplasm, and attempts to control tumor growth in metastatic disease.

PPIs are the treatment of choice and have decreased the need for total gastrectomy. Initial PPI doses tend to be higher than those used for treatment of GERD or PUD. The initial dose of omeprazole, lansoprazole, rabeprazole, or esomeprazole should be in the range of 60 mg in divided doses in a 24-h period. When gastric acid analysis was more widely available, dosing was adjusted to achieve a BAO <10 meq/h (at the drug trough) in surgery-naive patients and to <5 meq/h in individuals who have previously undergone an acid-reducing operation. Close monitoring of clinical symptoms when starting PPIs and increasing the dose accordingly are paramount. Although the somatostatin analogue has inhibitory effects on gastrin release from receptor-bearing tumors and inhibits gastric acid secretion to some extent, PPIs have the advantage of reducing parietal cell activity to a greater degree. Despite this, octreotide or lanreotide may be considered as adjunctive therapy to the PPI in patients with tumors that express somatostatin receptors and have peptic symptoms that are difficult to control with high-dose PPI.

The ultimate goal of surgery would be to provide a definitive cure. Improved understanding of tumor distribution has led to immediate cure rates as high as 33% with 10-year disease-free intervals as high as 95% in sporadic gastrinoma patients undergoing surgery. A positive outcome is highly dependent on the experience of the surgical team treating these rare tumors. Surgical therapy of gastrinoma patients with MEN 1 remains controversial because of the difficulty in rendering these patients disease-free with surgery. In contrast to the encouraging postoperative results observed in patients with sporadic disease, <5% of MEN 1 patients are disease-free 5 years after an operation. Moreover, in contrast to patients with sporadic ZES, the clinical course of MEN 1 patients tends to be benign and rarely leads to disease-related mortality, recommending that early surgery be deferred. Some groups suggest surgery only if a clearly identifiable, nonmetastatic lesion is documented by structural studies. Others advocate a more aggressive approach, where all patients free of hepatic metastasis are explored and all detected tumors in the duodenum are resected; this is followed by enucleation of lesions in the pancreatic head, with a distal pancreatectomy to follow. The outcome of the two approaches has not been clearly defined. Laparoscopic surgical interventions may provide attractive approaches in the future but currently seem to be of some limited benefit in patients with gastrinoma because a significant percentage of the tumors may be extrapancreatic and difficult to localize with a laparoscopic approach. Finally, patients selected for surgery should be individuals whose health status would lead them to tolerate a more aggressive operation and obtain the long-term benefits from such aggressive surgery, which are often witnessed after 10 years.

Therapy of metastatic endocrine tumors in general remains suboptimal; gastrinomas are no exception. In light of the observation that in many instances tumor growth is indolent and that many individuals with metastatic disease remain relatively stable for significant periods of time, many advocate not instituting systemic tumor-targeted therapy until evidence of tumor progression or refractory symptoms not controlled with PPIs are noted. Medical approaches, including biologic therapy (IFN-α, long-acting somatostatin analogues, and peptide receptor radionuclides), systemic chemotherapy (streptozotocin, 5-fluorouracil, and doxorubicin), and hepatic artery embolization, may lead to significant toxicity without a substantial improvement in overall survival. Use of temozolomide with capecitabine has demonstrated radiographic regression and progression-free survival in patients with well-differentiated NETs in the range of 70% and 18 months, respectively. Systemic therapy with radiolabeled somatostatin analogues (peptide receptor radiotherapy [PRRT]) has been used in the therapy of metastatic NETs and appears to be very promising in terms of radiographic regression, symptoms, and progression-free survival, but additional studies are warranted. Several promising therapies are being explored, including radiofrequency ablation or cryoablation of liver lesions and use of agents that block the VEGF receptor pathway (sunitinib, surufatinib), the mammalian target of rapamycin, and immune checkpoint inhibitors **(Chap. 89)**.

Surgical approaches, including debulking surgery and liver transplantation for hepatic metastasis, have also produced limited benefit.

The overall 5- and 10-year survival rates for gastrinoma patients are 62–75% and 47–53%, respectively. Individuals with the entire tumor resected or those with a negative laparotomy have 5- and 10-year survival rates >90%. Patients with incompletely resected tumors have 5- and 10-year survival rates of 43 and 25%, respectively. Patients with hepatic metastasis have <20% survival at 5 years. Favorable prognostic indicators include primary duodenal wall tumors, isolated lymph node tumor, the presence of MEN 1, and undetectable tumor upon surgical exploration. Poor outcome is seen in patients with shorter disease duration; female sex; older age at diagnosis; higher gastrin levels (>10,000 pg/mL); poor histologic differentiation; high proliferative index; certain tumor molecular changes including chromosome 1qLOH and chromosome XLOH; large pancreatic primary tumors (>2–3 cm); metastatic disease to lymph nodes, liver, and bone; and Cushing's syndrome. Rapid growth of hepatic metastases is also predictive of poor outcome.

STRESS-RELATED MUCOSAL INJURY

Patients suffering from shock, sepsis, massive burns, severe trauma, or head injury can develop acute erosive gastric mucosal changes or frank ulceration with bleeding. Classified as stress-induced gastritis or ulcers, injury is most commonly observed in the acid-producing (fundus and body) portions of the stomach. The most common presentation is GI bleeding, which is usually minimal but can occasionally be life-threatening. Respiratory failure requiring mechanical ventilation and underlying coagulopathy are risk factors for bleeding, which tends to occur 48–72 h after the acute injury or insult.

Histologically, stress injury does not contain inflammation or *H. pylori*; thus, "gastritis" is a misnomer. Although elevated gastric acid secretion may be noted in patients with stress ulceration after head trauma (Cushing's ulcer) and severe burns (Curling's ulcer), mucosal ischemia, breakdown of the normal protective barriers of the stomach, systemic release of cytokines, poor GI motility, and oxidative stress also play an important role in the pathogenesis. Acid must contribute to injury in view of the significant drop in bleeding noted when acid inhibitors are used as prophylaxis for stress gastritis.

Improvement in the general management of intensive care unit patients has led to a significant decrease in the incidence of GI bleeding due to stress ulceration. The estimated decrease in bleeding is from 20–30% to <5%. This improvement has led to some debate regarding the need for prophylactic therapy. The high mortality associated with stress-induced clinically important GI bleeding (>40%) and the limited benefit of medical (endoscopic, angiographic) and surgical therapy in a patient with hemodynamically compromising bleeding associated with stress ulcer/gastritis support the use of preventive measures in high-risk patients (mechanically ventilated, coagulopathy, multiorgan failure, or severe burns). Meta-analysis comparing H_2 blockers with PPIs for the prevention of stress-associated clinically important and overt GI bleeding demonstrates superiority of the latter without increasing the risk of nosocomial infections, increasing mortality, or prolonging intensive care unit length of stay. Therefore, PPIs are the treatment of choice for stress prophylaxis. Oral PPI is the best option if the patient can tolerate enteral administration. Pantoprazole is available as an intravenous formulation for individuals in whom enteral administration is not possible. If bleeding occurs despite these measures, endoscopy, intra-arterial vasopressin, and embolization are options. If all else fails, then surgery should be considered. Although vagotomy and antrectomy may be used, the better approach would be a total gastrectomy, which has an exceedingly high mortality rate in this setting. Concerns with the effect of PPIs on the immune system coupled with the high cost of this agent have led to several comparative studies of PPIs and H_2 receptor antagonists for stress prophylaxis in patients requiring mechanical ventilation. Although the PEPTIC trial demonstrated comparative efficacy between the two agents regarding mortality, technical aspects of the study led to some limitation in the final interpretation of the results. A meta-analysis performed in neurocritical patients did not reach strong clinical recommendations about the utility of ulcer prophylaxis. Reasons for this conclusion included the overall high or unclear risk of bias of individual trials, the low event rates, and the modest sample size examined. Currently, a trial to determine safety and efficacy of pantoprazole compared to placebo for preventing stress-induced erosions is ongoing in mechanically ventilated patients.

GASTRITIS

The term *gastritis* should be reserved for histologically documented inflammation of the gastric mucosa. Gastritis is not the mucosal erythema seen during endoscopy and is not interchangeable with "dyspepsia." The etiologic factors leading to gastritis are broad and heterogeneous. Gastritis has been classified based on time course (acute vs chronic), histologic features, and anatomic distribution or proposed pathogenic mechanism **(Table 335-10)**.

The correlation between the histologic findings of gastritis, the clinical picture of abdominal pain or dyspepsia, and endoscopic findings noted on gross inspection of the gastric mucosa is poor. Therefore, there is no typical clinical manifestation of gastritis.

Acute Gastritis The most common causes of acute gastritis are infectious. Acute infection with *H. pylori* induces gastritis. However, *H. pylori* acute gastritis has not been extensively studied. It is reported as presenting with sudden onset of epigastric pain, nausea, and vomiting, and limited mucosal histologic studies demonstrate a marked infiltrate of neutrophils with edema and hyperemia. If not treated, this picture will evolve into one of chronic gastritis. Hypochlorhydria lasting for up to 1 year may follow acute *H. pylori* infection.

Bacterial infection of the stomach or phlegmonous gastritis is a rare, potentially life-threatening disorder characterized by marked and diffuse acute inflammatory infiltrates of the entire gastric wall, at times accompanied by necrosis. Elderly individuals, alcoholics, and AIDS

TABLE 335-10 Classification of Gastritis

- I. Acute gastritis
  - A. Acute *Helicobacter pylori* infection
  - B. Other acute infectious gastritides
    - 1. Bacterial (other than *H. pylori*)
    - 2. *Helicobacter heilmannii*
    - 3. Phlegmonous
    - 4. Mycobacterial
    - 5. Syphilitic
    - 6. Viral
    - 7. Parasitic
    - 8. Fungal
- II. Chronic atrophic gastritis
  - A. Type A: Autoimmune, body-predominant
  - B. Type B: *H. pylori*–related, antral-predominant
  - C. Indeterminate
- III. Uncommon forms of gastritis
  - A. Lymphocytic
  - B. Eosinophilic
  - C. Crohn's disease
  - D. Sarcoidosis
  - E. Isolated granulomatous gastritis
  - F. Russell body gastritis

patients may be affected. Potential iatrogenic causes include polypectomy and mucosal injection with India ink. Organisms associated with this entity include streptococci, staphylococci, *Escherichia coli*, *Proteus*, and *Haemophilus* species. Failure of supportive measures and antibiotics may result in gastrectomy.

Other types of infectious gastritis may occur in immunocompromised individuals such as AIDS patients. Examples include herpetic (herpes simplex) or CMV gastritis. The histologic finding of intranuclear inclusions would be observed in the latter.

Chronic Gastritis Chronic gastritis is identified histologically by an inflammatory cell infiltrate consisting primarily of lymphocytes and plasma cells, with very scant neutrophil involvement. Distribution of the inflammation may be patchy, initially involving superficial and glandular portions of the gastric mucosa. This picture may progress to more severe glandular destruction, with atrophy and metaplasia. Chronic gastritis has been classified according to histologic characteristics. These include superficial atrophic changes and gastric atrophy. The association of atrophic gastritis with the development of gastric cancer has led to the development of endoscopic and serologic markers of severity. Some of these include gross inspection and classification of mucosal abnormalities during standard endoscopy, magnification endoscopy, endoscopy with narrow band imaging and/or autofluorescence imaging, and measurement of several serum biomarkers including pepsinogen I and II levels, gastrin-17, and anti–*H. pylori* serologies. The clinical utility of these tools is currently being explored.

The early phase of chronic gastritis is *superficial gastritis*. The inflammatory changes are limited to the lamina propria of the surface mucosa, with edema and cellular infiltrates separating intact gastric glands. The next stage is *atrophic gastritis*. The inflammatory infiltrate extends deeper into the mucosa, with progressive distortion and destruction of the glands. The final stage of chronic gastritis is *gastric atrophy*. Glandular structures are lost, and there is a paucity of inflammatory infiltrates. Endoscopically, the mucosa may be substantially thin, permitting clear visualization of the underlying blood vessels.

Gastric glands may undergo morphologic transformation in chronic gastritis. Intestinal metaplasia denotes the conversion of gastric glands to a small intestinal phenotype with small-bowel mucosal glands containing goblet cells. The metaplastic changes may vary in distribution from patchy to fairly extensive gastric involvement. Intestinal metaplasia is an important predisposing factor for gastric cancer **(Chap. 85)**.

Chronic gastritis has also been classified according to the predominant site of involvement. Type A refers to the body-predominant form (autoimmune), and type B is the antral-predominant form (*H. pylori*–related). This classification is artificial in view of the difficulty in distinguishing between these two entities. The term *AB gastritis* has been used to refer to a mixed antral/body picture. Histologic classification is the most accurate way to approach chronic gastritis.

TYPE A GASTRITIS The less common of the two forms involves primarily the fundus and body, with antral sparing. Traditionally, this form of gastritis has been associated with pernicious anemia **(Chap. 104)** in the presence of circulating antibodies against parietal cells and IF; thus, it is also called *autoimmune gastritis. H. pylori* infection can lead to a similar distribution of gastritis.

Antibodies to parietal cells have been detected in >90% of patients with pernicious anemia and in up to 50% of patients with type A gastritis. The parietal cell antibody is directed against H^+,K^+-ATPase. T cells are also implicated in the injury pattern of this form of gastritis. A subset of patients infected with *H. pylori* develop antibodies against H^+,K^+-ATPase, potentially leading to the atrophic gastritis pattern seen in some patients infected with this organism. The mechanism is thought to involve molecular mimicry between *H. pylori* LPS and H^+,K^+-ATPase.

Parietal cell antibodies and atrophic gastritis are observed in family members of patients with pernicious anemia. These antibodies are observed in up to 20% of individuals aged >60 and in ~20% of patients with vitiligo and Addison's disease. About one-half of patients with pernicious anemia have antibodies to thyroid antigens, and ~30% of patients with thyroid disease have circulating anti–parietal cell antibodies. Anti-IF antibodies are more specific than parietal cell antibodies for type A gastritis, being present in ~40% of patients with pernicious anemia. Another parameter consistent with this form of gastritis being autoimmune in origin is the higher incidence of specific familial histocompatibility haplotypes such as HLA-B8 and HLA-DR3. Low pepsinogen levels have also been observed; thus, this marker has been used as an additional diagnostic tool in autoimmune gastritis.

The parietal cell–containing gastric gland is preferentially targeted in this form of gastritis, and achlorhydria results. Parietal cells are the source of IF, the lack of which will lead to vitamin B_{12} deficiency and its sequelae (megaloblastic anemia, neurologic dysfunction).

Gastric acid plays an important role in feedback inhibition of gastrin release from G cells. Achlorhydria, coupled with relative sparing of the antral mucosa (site of G cells), leads to hypergastrinemia. Gastrin levels can be markedly elevated (>500 pg/mL) in patients with pernicious anemia. ECL cell hyperplasia with frank development of gastric carcinoid tumors may result from gastrin trophic effects. Hypergastrinemia and achlorhydria may also be seen in nonpernicious anemia–associated type A gastritis.

TYPE B GASTRITIS Type B, or antral-predominant, gastritis is the more common form of chronic gastritis. *H. pylori* infection is the cause of this entity. Although described as "antral-predominant," this is likely a misnomer in view of studies documenting the progression of the inflammatory process toward the body and fundus of infected individuals. The conversion to a pangastritis is time dependent and estimated to require 15–20 years. This form of gastritis increases with age, being present in up to 100% of persons aged >70. Histology improves after *H. pylori* eradication. The number of *H. pylori* organisms decreases dramatically with progression to gastric atrophy, and the degree of inflammation correlates with the level of these organisms. Early on, with antral-predominant findings, the quantity of *H. pylori* is highest and a dense chronic inflammatory infiltrate of the lamina propria is noted, accompanied by epithelial cell infiltration with polymorphonuclear leukocytes **(Fig. 335-16)**.

Multifocal atrophic gastritis, gastric atrophy with subsequent metaplasia, has been observed in chronic *H. pylori*–induced gastritis. This may ultimately lead to development of gastric adenocarcinoma (Fig. 335-8; **Chap. 85**). *H. pylori* infection is now considered an independent risk factor for gastric cancer. Worldwide epidemiologic studies have documented a higher incidence of *H. pylori* infection in

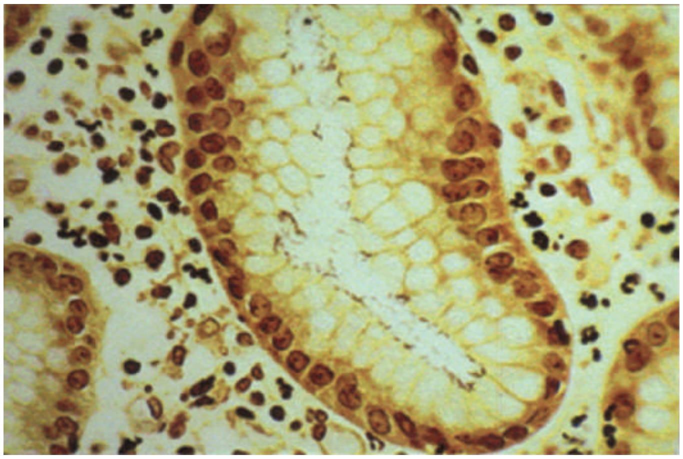

FIGURE 335-16 Chronic gastritis and *H. pylori* organisms. Steiner silver stain of superficial gastric mucosa showing abundant darkly stained microorganisms layered over the apical portion of the surface epithelium. Note that there is no tissue invasion.

patients with adenocarcinoma of the stomach as compared to control subjects. Seropositivity for *H. pylori* is associated with a three- to sixfold increased risk of gastric cancer. This risk may be as high as ninefold after adjusting for the inaccuracy of serologic testing in the elderly. The mechanism by which *H. pylori* infection leads to cancer is unknown, but it appears to be related to the chronic inflammation induced by the organism. Eradication of *H. pylori* as a general preventative measure for gastric cancer is being evaluated but is not yet recommended.

Infection with *H. pylori* is also associated with development of a low-grade B-cell lymphoma, gastric MALT lymphoma **(Chap. 113)**. The chronic T-cell stimulation caused by the infection leads to production of cytokines that promote the B-cell tumor. The tumor should be initially staged with a CT scan of the abdomen and EUS. Tumor growth remains dependent on the presence of *H. pylori*, and its eradication is often associated with complete regression of the tumor. The tumor may take more than a year to regress after treating the infection. Such patients should be followed by EUS every 2–3 months. If the tumor is stable or decreasing in size, no other therapy is necessary. If the tumor grows, it may have become a high-grade B-cell lymphoma. When the tumor becomes a high-grade aggressive lymphoma histologically, it loses responsiveness to *H. pylori* eradication.

TREATMENT

Chronic Gastritis

Treatment in chronic gastritis is aimed at the sequelae and not the underlying inflammation. Patients with pernicious anemia will require parenteral vitamin B_{12} supplementation on a long-term basis. Eradication of *H. pylori* is often recommended even if PUD or a low-grade MALT lymphoma is not present. Expert opinion suggests that patients with atrophic gastritis complicated by intestinal metaplasia without dysplasia should undergo surveillance endoscopy every 3 years.

Miscellaneous Forms of Gastritis *Lymphocytic gastritis* is characterized histologically by intense infiltration of the surface epithelium with lymphocytes. The infiltrative process is primarily in the body of the stomach and consists of mature T cells and plasmacytes. The etiology of this form of chronic gastritis is unknown. It has been described in patients with celiac sprue, but whether there is a common factor associating these two entities is unknown. No specific symptoms suggest lymphocytic gastritis. A subgroup of patients has thickened folds noted on endoscopy. These folds are often capped by small nodules that contain a central depression or erosion; this form of the disease is called *varioliform gastritis*. *H. pylori* probably plays no significant role in lymphocytic gastritis. Therapy with glucocorticoids or sodium cromoglycate has obtained unclear results.

Marked eosinophilic infiltration involving any layer of the stomach (mucosa, muscularis propria, and serosa) is characteristic of *eosinophilic gastritis*. Affected individuals will often have circulating eosinophilia with clinical manifestation of systemic allergy. Involvement may range from isolated gastric disease to diffuse eosinophilic gastroenteritis. Antral involvement predominates, with prominent edematous folds being observed on endoscopy. These prominent antral folds can lead to outlet obstruction. Patients can present with epigastric discomfort, nausea, and vomiting. Treatment with glucocorticoids has been successful.

Several systemic disorders may be associated with *granulomatous gastritis*. Gastric involvement has been observed in Crohn's disease. Involvement may range from granulomatous infiltrates noted only on gastric biopsies to frank ulceration and stricture formation. Gastric Crohn's disease usually occurs in the presence of small-intestinal disease. Several rare infectious processes can lead to granulomatous gastritis, including histoplasmosis, candidiasis, syphilis, and tuberculosis. Other unusual causes of this form of gastritis include sarcoidosis, idiopathic granulomatous gastritis, and eosinophilic granulomas involving the stomach. Establishing the specific etiologic agent in this form of gastritis can be difficult, at times requiring repeat endoscopy with biopsy and cytology. Occasionally, a surgically obtained full-thickness biopsy of the stomach may be required to exclude malignancy.

Russell body gastritis (RBG) is a mucosal lesion of unknown etiology that has a pseudotumoral endoscopic appearance. Histologically, it is defined by the presence of numerous plasma cells containing Russell bodies (RBs) that express kappa and lambda light chains. Only 10 cases have been reported, and 7 of these have been associated with *H. pylori* infection. The lesion can be confused with a neoplastic process, but it is benign in nature, and the natural history of the lesion is not known. There have been cases of resolution of the lesion when *H. pylori* was eradicated.

Immune checkpoint inhibitor–induced enterocolitis and gastritis are recognized sequelae of these oncologic therapies. The gastritis typically occurs later in the course of therapy. The diagnosis is made by the histologic findings on gastric mucosal biopsies obtained endoscopically. This is an important diagnosis to make since therapy with glucocorticoids and potentially IL-6 receptor blockers will be required. Moreover, this side effect will have an effect on the oncologic therapy prescribed.

■ MÉNÉTRIER'S DISEASE

Ménétrier's disease (MD) is a very rare gastropathy characterized by large, tortuous mucosal folds. MD has an average age of onset of 40–60 years with a male predominance. The differential diagnosis of large gastric folds includes ZES, malignancy (lymphoma, infiltrating carcinoma), infectious etiologies (CMV, histoplasmosis, syphilis, tuberculosis), gastritis polyposa profunda, and infiltrative disorders such as sarcoidosis. MD is most commonly confused with large or multiple gastric polyps (prolonged PPI use) or familial polyposis syndromes. The mucosal folds in MD are often most prominent in the body and fundus, sparing the antrum. Histologically, massive foveolar hyperplasia (hyperplasia of surface and glandular mucous cells) and a marked reduction in oxyntic glands and parietal cells and chief cells are noted. This hyperplasia produces the prominent folds observed. The pits of the gastric glands elongate and may become extremely dilated and tortuous. Although the lamina propria may contain a mild chronic inflammatory infiltrate including eosinophils and plasma cells, MD is not considered a form of gastritis. The etiology of this unusual clinical picture in children is often CMV, but the etiology in adults is unknown. Overexpression of the growth factor TGF-α has been demonstrated in patients with MD. The overexpression of TGF-α in turn results in overstimulation of the epidermal growth factor receptor (EGFR) pathway and increased proliferation of mucus cells, resulting in the observed foveolar hyperplasia.

The clinical presentation in adults is usually insidious and progressive. Epigastric pain, nausea, vomiting, anorexia, peripheral edema, and weight loss are signs and symptoms in patients with MD. Occult GI bleeding may occur, but overt bleeding is unusual and, when present, is due to superficial mucosal erosions. In fact, bleeding is more often seen in one of the common mimics of MD, gastric polyposis. Twenty to 100% of patients (depending on time of presentation) develop a protein-losing gastropathy due to hypersecretion of gastric mucus accompanied by hypoalbuminemia and edema. Gastric acid secretion is usually reduced or absent because of the decreased parietal cells. Large gastric folds are readily detectable by either radiographic (barium meal) or endoscopic methods. Endoscopy with deep mucosal biopsy, preferably full thickness with a snare technique, is required to establish the diagnosis and exclude other entities that may present similarly. A nondiagnostic biopsy may lead to a surgically obtained full-thickness biopsy to exclude malignancy. Although MD is considered premalignant by some, the risk of neoplastic progression is not defined. Complete blood count, serum gastrin, serum albumin, CMV and *H. pylori* serology, and pH testing of gastric aspirate during endoscopy should be included as part of the initial evaluation of patients with large gastric folds. A retrospective case-control study suggested an increased risk of developing gastric adenocarcinoma and decreased 5- and 10-year survival compared to a control cohort. Additional studies will be required to confirm these initial findings, but this may be difficult to achieve in light of the rarity of this unusual diagnosis.

TREATMENT

Ménétrier's Disease

Medical therapy with anticholinergic agents, prostaglandins, PPIs, prednisone, somatostatin analogues (octreotide), and H_2 receptor antagonists yields varying results. Ulcers should be treated with a standard approach. The discovery that MD is associated with overstimulation of the EGFR pathway has led to the successful use of the EGF inhibitory antibody, cetuximab, in these patients. Specifically, four of seven patients who completed a 1-month trial with this agent demonstrated near-complete histologic remission and improvement in symptoms. Cetuximab is now considered the first-line treatment for MD, leaving partial or total gastrectomy for severe disease with persistent and substantial protein loss despite therapy with this agent.

■ FURTHER READING

Abrignani MG et al: Proton pump inhibitors and gastroprotection in patients treated with antithrombotic drugs: A cardiologic point of view. World J Cardiology 15:8, 2023.

Bindu S et al: Non-steroidal anti-inflammatory drugs (NSAIDs) and organ damage: A current perspective. Biochem Pharmacol 180:114147, 2020.

Bjarnason I et al: Mechanisms of damage to the gastrointestinal tract from nonsteroidal anti-inflammatory drugs. Gastroenterology 154:500, 2018.

Brandi ML et al: Multiple endocrine neoplasia type 1: Latest insights. Endocr Rev 42:133, 2021.

Chey WD et al: ACG clinical guideline: Treatment of *Helicobacter pylori* infection. Am J Gastroenterol 112:212, 2017.

Engevik A et al: The physiology of the gastric parietal cell. Physiol Rev 100:2, 2020.

Jensen RT, Ito T: Gastrinoma, in *Endotext* [Internet]. Feingold KR, Anawalt B, Blackman MR, et al (eds). South Dartmouth (MA) 2000-. Updated Sept. 2023. Available from MDText.com *https://www.ncbi.nlm.nih.gov/books/NBK279075/*.

Kavitt RT et al: Diagnosis and treatment of peptic ulcer disease. Am J Med 132:447, 2019.

Luo Q et al: A review on the research progress on non-pharmacological therapy of *Helicobacter pylori*. Front Microbiol 14:1134254, 2023.

Malfertheiner P et al: Management of *Helicobacter pylori* infection: The Maastricht V/Florence Consensus Report. Gut 71:9, 2022.

Malfertheiner P et al: *Helicobacter pylori* infection. Nat Rev 9:19, 2023.

Pennelli G et al: Gastritis: Update on etiological features and histological practice approach. Pathologica 112:153, 2020.

Rao S et al: Survival times of patients with Menetrier's disease and risk of gastric cancer. Clin Gastroenterol Hepatol 19:10, 2021.

Savarino V et al: Proton pump inhibitors: Use and misuse in the clinical setting. Expert Rev Clin Pharmacol 11:1123, 2018.

Shah SC et al: AGA Clinical Practice Update on the management of refractory *Helicobacter pylori* infection: Expert review. Gastroenterology 160:5, 2021.

Yao X, Smolka AJ: Gastric parietal cell physiology and *Helicobacter pylori*-induced disease. Gastroenterology 156:2158, 2019.

336 Disorders of Absorption

Deborah C. Rubin

A wide range of diseases affect gastrointestinal (GI) absorptive function and may result in malabsorption syndromes. These disorders affect one or more of the three phases of enteral nutrient processing. *Luminal digestion* is initiated by lingual and gastric lipase and gastric pepsin and continues in the small bowel by the actions of pancreatic enzymes and bile salts. *Small-intestinal mucosal digestion and absorption* are mediated by enterocyte brush border enzymes including disaccharidases, enterokinases, and peptidases, which digest nutrients upon contact, and by mixed micelles containing lipids and bile salts. Protein and carbohydrate digestive products are transported into the enterocyte by carriers and transporters, and lipids enter by diffusion mediated by micelles. Once in the enterocyte, nutrients may be reprocessed for *postmucosal absorption* and entry into lymphatics (long-chain triglycerides as part of chylomicrons) or are transported into the bloodstream. Malabsorptive diseases or syndromes can be classified by their effects on one or more of these three phases of absorption **(Table 336-1)**.

Disorders of absorption also have diverse clinical presentations. For example, the deficiency of a single brush border membrane protein such as lactase causes symptoms of diarrhea by affecting the absorption of one nutrient, lactose. Celiac sprue may be localized to the duodenum and present with isolated iron deficiency or may cause diffuse intestinal mucosal disease, affecting the absorption of multiple nutrients and causing a constellation of symptoms and clinical presentations.

Definition of Diarrhea Diarrhea is the most common symptom associated with disorders of absorption. For most patients, diarrhea as a symptom is defined as an increase in stool number or frequency or a change in consistency. Because normal bowel patterns may vary from as many as two to four bowel movements per day to one stool per week, it is critical to use an objective measure of diarrhea to help direct evaluation. In health, stool volume or weight is <200 mL or <200 g, respectively, in 24 h. Collection of stool for weight/volume determination is one of the most useful tools for an evaluation of diarrhea. In particular, a 72-h collection for weight/volume and fecal fat determination is the gold standard for documenting the presence of steatorrhea, or fatty stool. Steatorrhea, defined as increased stool fat excretion to >7% of dietary fat, is a common manifestation of malabsorption. Steatorrhea often results in large, bulky, and malodorous stools. Malabsorption of single nutrients like lactose may result in an osmotic diarrhea, in which the osmotically active unabsorbed nutrient causes fluid to be drawn into the GI tract lumen. Malabsorptive diarrhea frequently is precipitated by eating and resolves or significantly decreases at night,

TABLE 336-1 Classification of Malabsorption Syndromes

- Inadequate digestion
  - Postgastrectomy[a]
  - Deficiency or inactivation of pancreatic lipase
    - Exocrine pancreatic insufficiency
      - Chronic pancreatitis
      - Pancreatic carcinoma
      - Cystic fibrosis
      - Pancreatic insufficiency—congenital or acquired
  - Gastrinoma—acid inactivation of lipase
  - Drugs—orlistat
- Reduced intraduodenal bile-acid concentration/impaired micelle formation
  - Liver disease
    - Parenchymal liver disease
    - Cholestatic liver disease
  - Bacterial overgrowth in small intestine:
    - Anatomic stasis
      - Afferent loop
      - Stasis/blind
      - Loop/strictures/fistulae
    - Functional stasis
      - Diabetes[a]
      - Scleroderma[a]
      - Intestinal pseudo-obstruction
  - Interrupted enterohepatic circulation of bile salts
    - Ileal resection
    - Crohn's disease
  - Drugs (binding or precipitating bile salts)—neomycin, cholestyramine, calcium carbonate
  - Impaired mucosal absorption/mucosal loss or defect
    - Intestinal resection or bypass[a]
    - Inflammation, infiltration, or infection:
      - Crohn's disease[a]
      - Amyloidosis
      - Scleroderma[a]
      - Lymphoma[a]
      - Eosinophilic enteritis
      - Mastocytosis
      - Tropical sprue
      - Celiac disease
      - Collagenous sprue
      - Whipple's disease[a]
      - Radiation enteritis[a]
      - Folate and vitamin B_{12} deficiency
      - Infections—giardiasis
      - Graft versus host disease
  - Genetic disorders
    - Disaccharidase deficiency
    - Agammaglobulinemia
    - Abetalipoproteinemia
    - Hartnup's disease
    - Cystinuria
- Impaired nutrient delivery to and/or from intestine:
  - Lymphatic obstruction
    - Lymphoma[a]
    - Lymphangiectasia
  - Circulatory disorders
    - Congestive heart failure
    - Constrictive pericarditis
    - Mesenteric artery atherosclerosis
    - Vasculitis
- Endocrine and metabolic disorders
  - Diabetes[a]
  - Hypoparathyroidism
  - Adrenal insufficiency
  - Hyperthyroidism
  - Carcinoid syndrome

[a]Malabsorption caused by more than one mechanism.

with fasting, and thus can frequently be distinguished from secretory diarrheas, for example from infectious causes such as bacterial enterotoxigenic *Escherichia coli*. In this circumstance, intestinal fluid and electrolyte secretion is stimulated by enterotoxin and will continue even during fasting.

OVERVIEW: NUTRIENT DIGESTION AND ABSORPTION

Luminal digestive processes begin in the mouth and proceed throughout the GI tract, mediated by salivary amylase, lingual and gastric lipases, gastric acid, pancreatic enzymes, and bile salts. As nutrients are digested in the lumen of the proximal GI tract, they are further processed by enterocyte brush border enzymes including disaccharidases such as lactase and sucrase-isomaltase, which produce monosaccharides, and peptidases, which hydrolyze polypeptides into tripeptides and dipeptides and amino acids. Lipids in mixed micelles are then absorbed into enterocytes.

The surface area of the small bowel, which is normally 600 cm long, is further enhanced by circular folds, villi, and microvilli. Following uptake into enterocytes, nutrients are further processed and transported into the lymphatics or into the portal circulation for use by other cells throughout the body. The intestine is also presented with 7–9 L of fluid daily, a volume comprising dietary fluid intake (1–2 L/day) and salivary, gastric, pancreatic, biliary, and intestinal fluid (6–7 L/day). In health, almost all of this fluid is reabsorbed by the small bowel and colon, resulting in a normal stool volume of <200 mL or stool weight of <200 g.

SPECIFIC NUTRIENTS

Lipids Lipid absorption is a complex process that requires hydrolysis by pancreatic enzymes and bile salts for physiochemical dispersion of fats, followed by absorption of processed lipid nutrients dispersed in bile salt–mixed micelles across the intestinal epithelium. Bile acids are synthesized in the liver, secreted into the intestinal lumen, and constantly recirculated by absorption in the ileum. The ileum expresses fibroblast growth factor 19 (FGF19), which is a physiologic bile acid sensor. FGF19 is secreted from the ileum into the bloodstream in response to bile acid flux and negatively regulates hepatic bile acid synthesis by affecting the transcription of hepatic CYP7A1.

Thus, assimilation of dietary lipid requires three integrated processes: an intraluminal or digestive phase, a mucosal or absorptive phase, and a delivery or postabsorptive phase (Table 336-2).

Gastric lipases begin the lipolytic process. Following entry into the small bowel, long-chain triglycerides, with carbon lengths >12 and that are the major component of dietary lipid, are hydrolyzed by pancreatic lipases into fatty acids and monoglyceride during a process called *lipolysis* (Fig. 336-1). Long-chain free fatty acids are dispersed by bile salts into mixed micelles, which contact the brush border and permit fatty acid absorption into enterocytes across this specialized apical membrane. The other two types of fatty acids that compose fats, medium-chain and short-chain fatty acids, are soluble in the unstirred water layer. Medium-chain triglycerides with carbon chain lengths of 8–12 are found in coconut oil. Long-chain fatty acids are re-esterified to triglycerides in enterocytes, packaged into chylomicrons that contain apolipoproteins on the surface, which are subsequently secreted into the extracellular space, and because of their size, are excluded

TABLE 336-2 Defects in Lipid Digestion and Absorption in Steatorrhea

PHASE, PROCESS	PATHOPHYSIOLOGIC DEFECT	DISEASE EXAMPLE
Digestive		
Lipolysis formation	Decreased lipase secretion	Chronic pancreatitis
Micelle formation	Decreased intraduodenal bile acids	
Absorptive		
Mucosal uptake and re-esterification	Mucosal dysfunction	Celiac disease
Postabsorptive		
Chylomicron formation	Absent β-lipoproteins	Abetalipoproteinemia
Delivery from intestine	Abnormal lymphatics	Intestinal lymphangiectasia

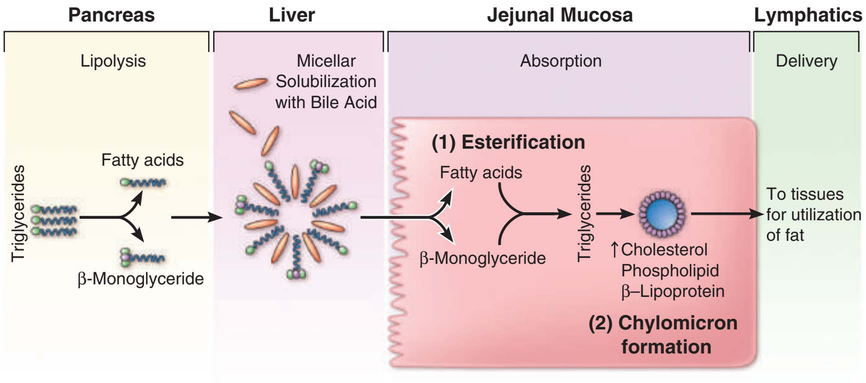

FIGURE 336-1 Schematic representation of lipid digestion and absorption. Dietary lipid is in the form of long-chain triglycerides. The overall process can be divided into (1) a digestive phase that includes both lipolysis and micelle formation requiring pancreatic lipase and conjugated bile acids, respectively, in the duodenum; (2) an absorptive phase for mucosal uptake and re-esterification; and (3) a postabsorptive phase that includes chylomicron formation and exit from the intestinal epithelial cell via lymphatics. *(Courtesy of John M. Dietschy, MD; with permission.)*

from capillaries and enter the lymphatics. Medium-chain triglycerides do not require micelle formation or pancreatic lipolysis as they are directly absorbed intact from the small bowel into the bloodstream, and short-chain fatty acids (carbon length <8) are produced by and absorbed in the colon.

Carbohydrates Dietary carbohydrate consists of starch, sucrose, lactose, maltose, and monosaccharides such as glucose and fructose. Starch is digested by salivary α-amylase in the mouth, followed by pancreatic amylase. The main products include maltotriose, maltose, and α-dextrins. These are further digested on the brush border membrane by disaccharidases such as glucoamylase and sucrase-isomaltase. Dietary lactose is digested by brush border lactase, sucrose by sucrase, and trehalose by trehalase. The final digested products are glucose, fructose, and galactose, which are transported into the enterocyte by transporters such as SLCA5 (formerly SGLT-1), which transports glucose or galactose in a sodium-dependent manner, and GLUT-5, which transports fructose by facilitated diffusion. Glucose, galactose, and fructose exit the cell via GLUT-2.

Proteins Dietary protein digestion begins in the stomach by pepsin. Pancreatic proteases including endopeptidases, exopeptidases, and trypsin are activated in the small-bowel lumen. Trypsinogen is activated by brush border enterokinase to generate active trypsin. Trypsin in turn activates chymotrypsinogen to chymotrypsin, proelastase to elastase, and procarboxypeptidases to carboxypeptidases A and B. These enzymes digest protein into dipeptides, tripeptides, larger polypeptides, or free amino acids. At the brush border, peptidases digest larger peptides into dipeptides and tripeptides or free amino acids, which enter the enterocyte via specialized carriers. Most dipeptides and tripeptides are further metabolized intracellularly by cytoplasmic peptidase into amino acids, which directly enter the bloodstream via carriers in the basolateral membrane. Small amounts of dipeptides and tripeptides may also enter the bloodstream.

■ LUMINAL PHASE OF DIGESTION

The luminal phase of digestion begins in the mouth, starting with mastication and lipase secretion by the tongue and salivary glands. The stomach continues the luminal digestive process, via gastric acid, gastric lipase, and pepsin secretion as well as mechanical trituration of contents. In the small-bowel lumen, pancreatic enzymes (amylase, lipases, carboxypeptidase, trypsin, and other endopeptidases) contribute to carbohydrate, lipid, and protein digestion, respectively. Bile salts produced by the liver are secreted into the intestinal lumen (and reabsorbed in the ileum via the enterohepatic circulation) and are required for efficient lipid absorption.

Disorders That Affect the Luminal Phase of Digestion The luminal phase may be disrupted by *disorders of gastric and intestinal motility* including the sequelae of gastric surgery, systemic diseases such as scleroderma, or endocrine disorders such as diabetes mellitus; *pancreatic diseases* leading to pancreatic insufficiency with reduced pancreatic enzyme secretion; or *luminal bile salt deficiency* caused by hepatobiliary disease, ileal disease, or small-bowel bacterial overgrowth.

Gastric Resection Surgical procedures that remove or bypass part of the stomach and duodenal bulb such as Roux-en-Y gastric bypass for weight loss, or resection of the gastric antrum and duodenal bulb with creation of a Billroth II anastomosis for treatment of peptic ulcer disease, result in rapid gastric emptying into the jejunum, which leads to diarrhea and weight loss due to inadequate mixing of luminal nutrients with bile and pancreatic secretions.

Disordered Intestinal Motility Hyperthyroidism may cause diarrhea and malabsorption due to increased intestinal motility with rapid transit, also resulting in inadequate nutrient mixing with pancreaticobiliary secretions. Long-standing diabetes mellitus may result in damage to the enteric nervous system resulting in increased motility and diarrhea, or reduced motility and constipation. Disorders that affect the intestinal smooth muscle such as connective tissue disorders including scleroderma may have profound effects on GI motility.

Pancreatic Disorders Chronic pancreatitis (see Chap. 359) may result in a marked reduction in pancreatic enzyme secretion and pancreatic insufficiency, with subsequent fat, protein, and carbohydrate malabsorption. Patients with chronic pancreatitis present with *steatorrhea*, or fatty stools, which are often voluminous, bulky, and malodorous. Patients with steatorrhea also develop deficiency of fat-soluble vitamins including vitamins A, E, and most commonly, vitamins D and K, which depend on the same lipid absorption mechanisms and thus are malabsorbed along with dietary fat. Weight loss is common. For a discussion of causes of acute and chronic pancreatitis, please see Chap. 359.

Disorders That Result in Luminal Bile Salt Deficiency Bile acid synthesis and the enterohepatic circulation are shown in Fig. 336-2.

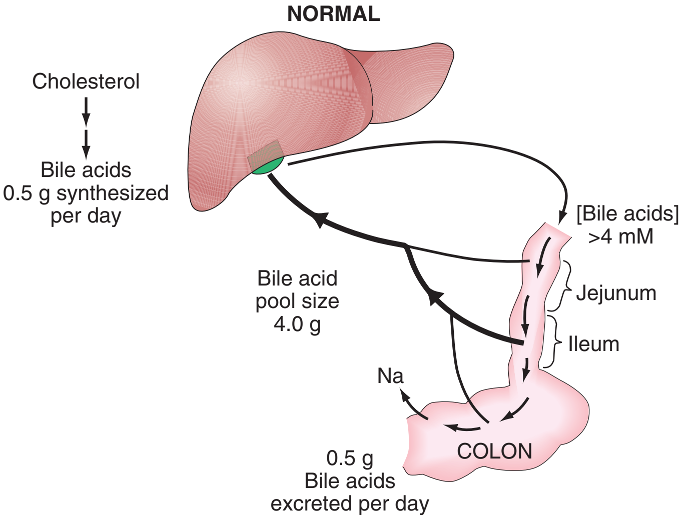

FIGURE 336-2 Schematic representation of the enterohepatic circulation of bile acids. Bile-acid synthesis is cholesterol catabolism and occurs in the liver. Bile acids are secreted in bile and are stored in the gallbladder between meals and at night. Food in the duodenum induces the release of cholecystokinin, a potent stimulus for gallbladder contraction resulting in bile-acid entry into the duodenum. Bile acids are primarily absorbed via an Na-dependent transport process that is located only in the ileum. A relatively small quantity of bile acids (~500 mg) is not absorbed in a 24-h period and is lost in stool. Fecal bile acid losses are matched by bile-acid synthesis. The bile acid pool (the total amount of bile acids in the body) is ~4 g and is circulated twice during each meal or six to eight times in a 24-h period.

TABLE 336-3 Defects in Enterohepatic Circulation of Bile Acids

PROCESS	PATHOPHYSIOLOGIC DEFECT	DISEASE EXAMPLE
Synthesis	Decreased hepatic function	Cirrhosis
Biliary secretion	Altered canalicular function	Primary biliary cholangitis
Maintenance of conjugated bile acids	Bacterial overgrowth	Jejunal diverticulosis
Reabsorption	Abnormal ileal function	Crohn's disease

Bile acids are synthesized from cholesterol in the liver. The two primary bile acids are cholic acid and chenodeoxycholic acid. These are conjugated in the liver to taurine and glycine and are secreted into bile ducts, stored in the gallbladder, and then delivered to the intestinal lumen. Conjugation prevents bile acids from passive diffusion in the small-bowel lumen, retaining bile acid concentrations required for lipid absorption. Bile acids emulsify fats and fat-soluble vitamins to facilitate their absorption. Bile acids are efficiently reabsorbed in the ileum into the portal circulation and are extracted by the liver in a process called *enterohepatic circulation* (Fig. 336-2). Small amounts are deconjugated in the ileum by bacteria or pass into the colon and are deconjugated and metabolized by colonic bacteria to become secondary bile acids. The two major secondary bile acids are lithocholic acid and deoxycholic acid.

Processes that affect any of the above pathways may result in luminal bile salt deficiency and malabsorption. Thus, hepatobiliary diseases, intestinal ileal resection, extensive disease such as Crohn's disease, and small-bowel bacterial overgrowth may result in luminal bile salt deficiency and malabsorption (Table 336-3).

Hepatobiliary Disease Hepatic disorders that result in decreased bile acid synthesis due to hepatocyte dysfunction or reduced secretion of bile into the gut lumen caused by diseases of the bile ducts such as primary sclerosing cholangitis or primary biliary cholangitis may result in luminal bile salt deficiency and fat malabsorption. These are discussed in Chap. 357.

Ileal Resection or Ileal Disease Diseases that involve the ileal mucosa or that result in ileal resection may lead to reduced recycling of bile acids by the enterohepatic circulation and increased entry into and concentration of bile acids in the colon, which produces a secretory diarrhea, or may result in malabsorption due to inadequate bile acid concentrations in the small-bowel lumen. In general, resection or disease involving <100 cm of ileum results in bile acid spillage into the colon; resections of >100 cm result in loss of bile acids that exceed liver synthetic capacity, and malabsorption becomes the dominant pathophysiologic mechanism for diarrhea, due to bile acid deficiency (Table 336-4). The most common disorder of the GI tract that targets the ileum is Crohn's disease (Chap. 337), which is a chronic inflammatory disorder that may involve the entire GI tract, but most commonly the ileum and colon. If severe or refractory to treatment, Crohn's disease may lead to chronic inflammation, marked epithelial dysfunction, and stricturing and fibrosis, and surgical resection may be required to treat small-bowel obstruction or refractory disease.

TABLE 336-4 Comparison of Bile Acid and Fatty Acid Diarrhea

	BILE-ACID DIARRHEA	FATTY ACID DIARRHEA
Extent of ileal disease	Limited	Extensive
Ileal bile acid absorption	Reduced	Reduced
Fecal bile acid excretion	Increased	Increased
Fecal bile acid loss compensated by hepatic synthesis	Yes	No
Bile acid pool size	Normal	Reduced
Intraduodenal (bile acid)	Normal	Reduced
Steatorrhea	None or mild	>20 g
Response to cholestyramine	Yes	No
Response to low fat diet	No	Yes

Primary Bile Acid Diarrhea A subset of patients with functional diarrhea or irritable bowel syndrome with diarrhea have been shown to have bile acid malabsorption. Although the mechanisms are still being elucidated, reduced FGF19 secretion by ileal enterocytes has been observed. FGF19 regulates serum 7-alpha-hydroxy-4-cholesten-3-one (C4) levels; reductions in circulating FGF19 lead to increased hepatic bile acid synthesis via increased C4 expression. Chronic diarrhea results from increased bile acid spillage into the colon, which induces a secretory diarrhea and exerts prokinetic effects via TGR5.

Treatment Bile acid sequestrants are effective in reducing diarrhea by binding bile acids to prevent spillage into the colon. Hepatic synthesis of bile acids is sufficient to maintain intraluminal concentrations that are adequate for fat absorption.

Small-Intestinal Bacterial Overgrowth The intestine contains a rich microbiome. Bacterial titers increase along the horizontal axis of the gut from duodenum to ileum. However, intestinal disorders affecting motility or causing stasis of bowel contents may lead to small-intestinal bacterial overgrowth. These include scleroderma bowel, chronic intestinal pseudo-obstruction, the creation of blind surgical loops such as Roux-en-Y gastric bypass, Billroth II anastomosis, small-bowel strictures, or fibrosis from inflammatory disorders such as Crohn's disease and diffuse diverticulosis (Fig. 336-3). Surgical resection of the ileocecal valve increases ileal bacterial counts from the colon. Hypochlorhydria from chronic proton pump inhibitor use also increases the risk of small-intestinal bacterial overgrowth. Bacterial overgrowth causes deconjugation of bile acids, which facilitates their absorption in the proximal bowel and results in luminal bile acid deficiency, which in turn causes malabsorptive diarrhea with steatorrhea. Bacterial overgrowth may also damage the brush border and result in carbohydrate maldigestion and short-chain fatty acid production in the colon, with diarrhea and gas. These patients are also at risk for B_{12} deficiency due to bacterial metabolism of B_{12} resulting in macrocytic anemia and peripheral neuropathy. In contrast, elevated serum folate levels may also be observed, derived from bacterial synthesis of folate.

Small-intestinal bacterial overgrowth has also been observed in patients with diarrhea-predominant irritable bowel syndrome. The underlying mechanisms are unclear, but treatment of bacterial overgrowth leads to resolution of symptoms in a subset of irritable bowel syndrome patients.

Diagnosis Duodenal aspirate for bacterial titers is the gold standard but is not generally available to most practitioners. Breath hydrogen testing with administration of lactulose, a nondigestible disaccharide, or glucose is widely available but must be interpreted carefully to avoid false-positive results. Many clinicians choose to treat empirically with antibiotics (see "Treatment") and observe for resolution of symptoms.

Treatment When possible, surgical correction of blind loops, endoscopic or surgical treatment of strictures, and removal of large diverticula can be pursued for definitive therapy, in addition to treatment of underlying disorders such as Crohn's disease to avoid recurrent stricture formation or fibrosis. Other disorders such as scleroderma or other diffuse motility disorders may not be easily treated. In these circumstances, treatment with the nonabsorbable antibiotic, rifaximin, or with other antibiotics such as metronidazole, doxycycline, amoxicillin-clavulanic acid, or cephalosporins for several weeks is often pursued. Patients may require retreatment or even chronic therapy with rotating antibiotics depending on the severity of symptoms.

■ MUCOSAL PHASE OF DIGESTION AND ABSORPTION

The intestinal epithelium (also known as the mucosa) plays a critical role in continued digestion of nutrients and absorption from the intestinal lumen into the bloodstream and lymphatics.

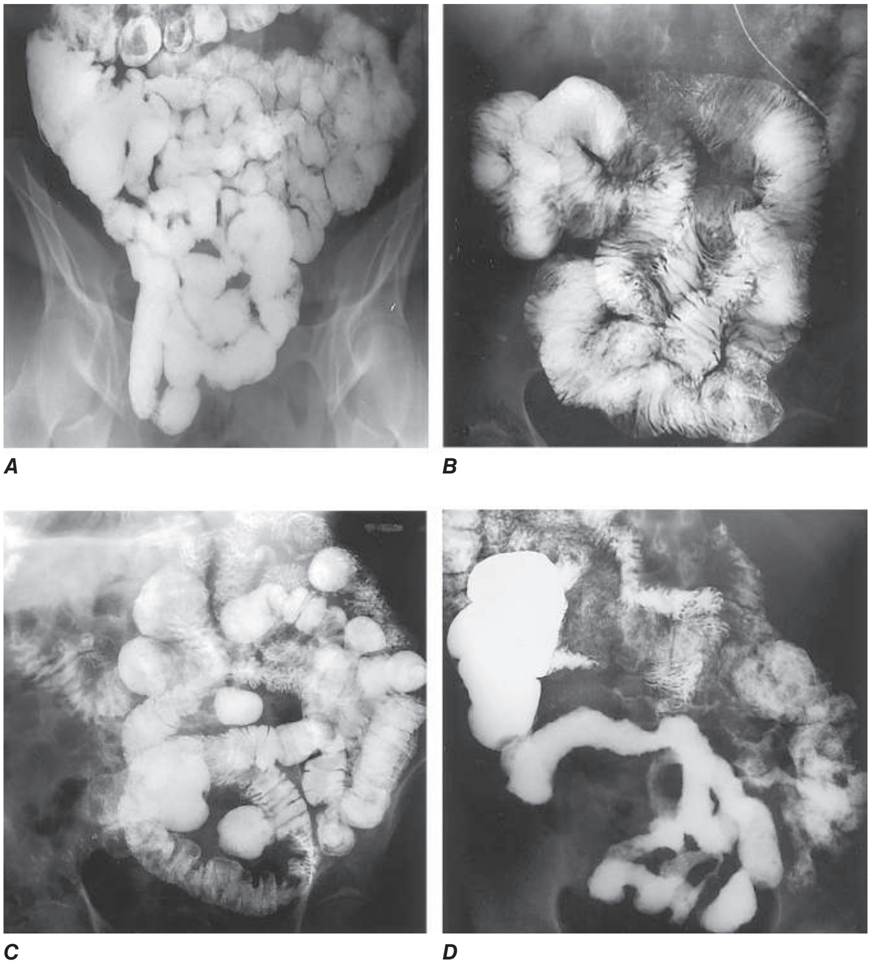

FIGURE 336-3 Barium contrast small-intestinal radiologic examinations. ***A.*** Normal individual. ***B.*** Celiac disease. ***C.*** Jejunal diverticulosis. ***D.*** Crohn's disease. *(Courtesy of Morton Burrell, MD, Yale University; with permission.)*

The small-bowel epithelial or mucosal digestive and absorptive phase is mediated by enterocytic brush border enzymes, including peptidases and hydrolases. Brush border enterokinase is required for the conversion of pancreatic trypsinogen to trypsin, which further activates trypsinogen and other pancreatic protease proenzymes. The brush border membrane of the small-bowel epithelium expresses a wide variety of disaccharidases, peptidases, and other hydrolases that continue the digestive process for carbohydrates and proteins, with enzymatic digestion of disaccharides to monosaccharides and dipeptidases to amino acids, which are then absorbed by specific transporters. Long-chain fatty acids are re-esterified to triglycerides in enterocytes, packaged into chylomicrons with apolipoproteins on the surface, which are subsequently secreted into the extracellular space, and because of their size, are excluded from capillaries and enter the lymphatics.

INTESTINAL MUCOSAL DISORDERS

■ DISORDERS OF ENTEROCYTE CARBOHYDRATE TRANSPORTERS AND ENZYME DEFICIENCIES

Lactose Intolerance Due to Lactase Deficiency This is the most common brush border disaccharidase deficiency and is a frequent cause of diarrhea, abdominal pain, gassiness, and bloating. Lactose is present in many dairy products but is also a "hidden" component of a vast number of processed foods.

Lactose malabsorption can result from lactase deficiency, which is regulated by primary genetic mechanisms (adult-type hypolactasia) or secondary due to damage to the epithelial (mucosal) lining of the gut, from infections (viral, bacterial, or parasitic) or from intestinal mucosal diseases. Congenital lactase deficiency is very rare and is an autosomal recessive disorder. Hypolactasia in adulthood is very common throughout the world and is considered to be the genetic wild-type; lactase persistence results from a C to T mutation (LACTASE LCT-13910CT and LCT-13910TT), and adults with hypolactasia have absence of this "persistence" allele. Lactose is metabolized by lactase into glucose and galactose, which are both absorbed by transporters at the enterocyte surface. Patients who are lactase deficient have elevated luminal lactose levels upon ingestion of lactose. The mechanism for diarrhea in lactase deficiency is complex. Undigested lactose acts as an osmotic substance to draw fluid into the small-bowel lumen. In addition, when unabsorbed lactose enters the colon, luminal bacteria ferment lactose, producing intestinal gas (hydrogen, carbon dioxide, and methane), bloating, and abdominal pain. Luminal lactose is metabolized by bacteria into short-chain fatty acids that can be absorbed by the colon, but watery diarrhea may occur when a large lactose load exceeds the colon's absorptive capacity.

Diagnosis When lactose intolerance is suspected, a common initial approach is to institute a lactose-exclusion diet and assess for resolution of symptoms. This is a rapid and generally effective diagnostic and therapeutic method. Patients are provided with a list of lactose-containing foods and lactose-free alternatives. Patients are also counseled on alternative calcium sources, because dairy-containing foods are a major source of dietary calcium, which is important for osteoporosis prevention.

Should the results of dietary exclusion be ambiguous, a lactose-tolerance test or breath hydrogen test may prove useful. For the lactose-tolerance test, patients ingest a standardized liquid lactose solution (usually 50 g of lactose) followed by timed measurements of serum glucose for 90 min. If lactose digestion is normal, glucose levels should rise by >20 mg/L. Serum glucose rise <20 mg/L plus the presence of symptoms of lactose intolerance (abdominal discomfort, gassiness, and diarrhea) is considered a positive test. A breath hydrogen test is performed by measuring breath hydrogen levels following ingestion of a standardized lactose load. Breath hydrogen levels should not exceed >20 ppm above the fasting baseline. Generally, the peak occurs between 2 and 4 h. Both methods may be inaccurate if the patient has abnormal gastric emptying or abnormal intestinal transit. Breath hydrogen measurements may be abnormal in the setting of bacterial overgrowth, which may cause very similar symptoms.

Treatment Patients may elect to completely eliminate lactose from their diets. It is very important to consider calcium and vitamin D supplementation because elimination of milk and soft cheeses removes important dietary sources. They also may need to consult a dietitian for guidance about hidden lactose in prepared or other foods. An alternative is to consider using lactase supplementation, which is available over the counter but which may need to be titrated to avoid symptoms.

Glucose-Galactose Malabsorption This rare congenital disorder is an autosomal recessive disease in which mutations occur in the *SLC5A1* gene (also known as *SGLT1*). SLC5A1 is a brush border protein and member of the sodium-dependent glucose transporter family; mutations in this gene result in malabsorption of glucose and galactose. Gene sequencing has shown that most patients have loss-of-function single-nucleotide variations. SLC5A1 actively transports glucose or galactose coupled to sodium cotransport; patients who are homozygous for these loss-of-function variants have severe congenital diarrhea and death if unrecognized. Treatment focuses on eliminating glucose- and galactose-containing foods and substituting fructose-containing foods. Fructose is absorbed by the brush border transporter GLUT5 by facilitated diffusion and is not dependent on SLC5A1.

Sucrase-Isomaltase Deficiency This is another rare congenital autosomal recessive disorder, caused by mutations in the sucrase-isomaltase gene. Infants present with diarrhea that begins with ingestion of sucrose coincident with the introduction of table foods. Sucrase-isomaltase gene variants have been identified that cause symptoms later in life and may be associated with increased risk of irritable bowel syndrome.

Abetalipoproteinemia Abetalipoproteinemia is a rare disorder of lipid metabolism associated with abnormal erythrocytes (acanthocytes), neurologic symptoms, and steatorrhea **(see Chap. 419)**. Lipolysis, micelle formation, and lipid uptake are all normal in patients with abetalipoproteinemia, but the re-esterified triglyceride cannot exit the epithelial cell because of the failure to produce chylomicrons. This disorder results from mutation of microsomal triglyceride transfer protein (MTP), which catalyzes the transfer of triglyceride onto nascent apolipoprotein B–containing particles. Mutations in MTP decrease this transfer and decrease formation of chylomicrons. Small-intestinal biopsy samples obtained from these rare patients in the postprandial state reveal lipid-laden small-intestinal epithelial cells that become normal in appearance after a 72- to 96-h fast.

INTESTINAL MUCOSAL DISORDERS THAT RESULT IN MALABSORPTION OF MULTIPLE NUTRIENTS

■ CELIAC DISEASE

Celiac disease, also known as celiac sprue or gluten-sensitive enteropathy, is a small-intestinal enteropathy that results from an immune response to gluten ingestion and is characterized by autoantibodies to tissue transglutaminase. Gluten is found in foods produced from wheat, rye, barley, and some varieties of oats, and it is a common additive to prepared foods and pharmaceuticals. Tissue transglutaminase is involved in the pathogenesis of this disorder, as it deamidates glutamine residues of gluten-derived peptides, facilitating their presentation by antigen-presenting cells.

Epidemiology and Genetics The incidence and prevalence of celiac disease have been increasing worldwide. Increased awareness among clinicians and patients has led to increases in detection, but there is evidence that the true incidence appears to be increasing as well. Global prevalence has been measured at 1.4%. In the United States, data from the National Health and Nutrition Examination survey showed seroprevalence of 0.2% in non-Hispanic black populations, 0.3% in Hispanic individuals, and 1.0% in white populations.

The prevalence of celiac disease is 10–15% in first-degree relatives. Host genetic factors include histocompatibility locus antigens HLADQ2 and DQ8; the presence of one of the two haplotypes is necessary but not sufficient for developing celiac disease. HLADQ2 and DQ8 are found in 25–35% of the general population; because most carriers never develop celiac disease, detection of these alleles is not useful for diagnosis. However, a negative test is very useful for ruling out celiac disease, with a negative predictive value of >99%. This is particularly helpful in patients who self-discontinued gluten ingestion prior to serologic or endoscopic testing.

Presentation Patients with celiac disease have a wide variety of disease manifestations, ranging from being asymptomatic, to having isolated iron-deficiency anemia due to duodenal disease, to severe diarrhea, weight loss, and malabsorption of multiple nutrients with more diffuse disease. Celiac disease primarily affects the proximal small intestine; it may involve the duodenum only or may cause widespread jejunal disease resulting in severe symptoms.

Diarrhea, weight loss, and growth failure in children are common presenting complaints, but additional signs and symptoms have become increasingly recognized to be associated with celiac disease, including bloating and irregular bowel habits, migraine headaches, and ataxia. In addition, patients may be identified after presenting with osteoporosis, iron-deficiency anemia, or detection of abnormal liver enzymes.

Mechanism of Diarrhea Patients with celiac disease have villus atrophy in the proximal small intestine and thus develop steatorrhea from mucosal malabsorption and may have lactase deficiency. However, they also develop a secretory component due to crypt hyperplasia and fluid hypersecretion from the crypt epithelium.

Associated Diseases Patients with celiac disease have a higher incidence of other autoimmune disorders such as type 1 diabetes mellitus and autoimmune thyroid disease. Dermatitis herpetiformis is a skin disorder that is highly associated with celiac disease, characterized by a vesicular rash mediated by IgA deposits in the skin. Down syndrome and Turner syndrome patients also have an increased risk of celiac disease.

Diagnosis Patients are screened for celiac disease first by testing for serum antibodies, including the screening test of choice, which is tissue transglutaminase IgA (TTG-IgA). Serum IgA levels should also be measured to detect false-negative results from IgA deficiency. If a patient is IgA deficient, IgG-based testing including TTG-IgG or anti- deamidated gliadin peptide (DGP-IgG) antibody testing can be performed. The diagnosis in adults with positive antibody levels is confirmed by endoscopy with small-intestinal biopsy; biopsy is required in most patients to confirm the diagnosis. Biopsies typically show characteristic villus blunting, crypt hyperplasia, and inflammation, including increased intraepithelial lymphocytes. The Marsh classification categorizes different types of celiac disease–related lesions and is currently used to quantify severity of disease involvement.

Family members of patients with celiac disease are screened if symptomatic; recommendations regarding screening asymptomatic family members are still controversial.

Complications Complications of celiac disease include refractory celiac disease, enteropathy-associated T-cell lymphoma, hyposplenism, and small-bowel adenocarcinoma.

Refractory Celiac Disease This complication is most common in patients with ongoing active celiac disease, found in about 10% of patients with persistent active disease. Patients have ongoing diarrhea and weight loss with persistent villus atrophy on biopsy after 1 year of following a strict gluten-free diet. These patients also have negative celiac serology, confirming their adherence to the gluten-free diet. Type 1 refractory celiac disease has a normal intraepithelial lymphocyte population, whereas type 2 disease has clonal expansion of CD3+ intraepithelial lymphocytes that also contain a monoclonal rearrangement of the gamma chain of the T-cell receptor. Type 2 refractory celiac disease has a worse prognosis due to its association with T-cell lymphoma, which occurs in 33–50% of cases after 5 years. The therapy for celiac disease–related lymphoma is intense and includes high-dose chemotherapy and sometimes stem cell transplantation.

Small-bowel adenocarcinoma is a very rare cancer in the general population but is increased in celiac disease patients.

Therapy and Follow-Up The mainstay of celiac disease treatment is institution of a strict gluten-free diet. This is challenging for patients because of the widespread presence of gluten in both raw and prepared foods, inaccurate food labeling, and cross-contamination during food preparation. Patients must receive rigorous dietary instruction from a dietitian and adhere lifelong to a gluten-free diet.

For patients whose symptoms resolve, serologic follow-up is generally recommended to confirm compliance with a gluten-free diet. A follow-up biopsy to document complete healing of villus atrophy is also generally recommended. However, subsequent biopsies are not recommended unless symptoms recur. For patients without symptom resolution, a biopsy is required to determine the degree of disease activity and to rule out other causes of persistent diarrhea and complications such as refractory celiac disease or T-cell lymphoma. The most common cause of residual disease activity is dietary nonadherence or inadvertent gluten exposure. These patients pursue repeat consultation with a dietitian and efforts to reduce restaurant or other out-of-the-home exposure or cross-contamination at home. If biopsies are negative but symptoms persist, other causes of abdominal pain and diarrhea that are associated with celiac disease are considered, including irritable bowel syndrome, microscopic colitis, small-bowel bacterial overgrowth, and lactose or fructose intolerance.

Nonceliac Gluten Sensitivity These patients have symptoms consistent with celiac disease but have negative serology and negative biopsies and no evidence for wheat allergy. Upon discontinuation of gluten, they have relief of abdominal pain, diarrhea, headaches/migraines, and other celiac disease–type symptoms. The etiology of this disorder remains unknown.

WHIPPLE'S DISEASE

Whipple's disease is a chronic, multiorgan disease caused by *Tropheryma whipplei*, a gram-positive non-acid-fast, periodic acid–Schiff (PAS)–positive rod, which is ubiquitous in the environment. Whipple's disease most commonly occurs in middle-aged men. Classic Whipple's disease is defined by the presence of arthralgias, weight loss, diarrhea, and abdominal pain. Other manifestations including central nervous system (CNS) and cardiac involvement are common and occur later in the disease. *T. whipplei* can be detected by polymerase chain reaction on involved tissue and is difficult to detect in the bloodstream. The intestinal lesion is also characterized by PAS-positive macrophages.

Clinical Presentation Arthralgias and arthritis are present for an average of 6 years before the GI symptoms begin, consistent with a persistent and substantial lag in diagnosis, which is still a problem today. Joint disease is present in >80% of patients. GI manifestations include diarrhea, abdominal pain, and weight loss from malabsorption. CNS involvement is common and may include symptoms such as psychiatric manifestations or memory problems. Dementia and encephalitis may occur in later stages. Cardiac involvement may include endocarditis, pericarditis, and myocarditis.

Diagnosis For patients with GI manifestations, endoscopy with biopsies is performed and tissue is tested for *T. whipplei* by polymerase chain reaction. Tissue is also stained for PAS-positive macrophages, and immunohistochemistry may also be performed to detect *T. whipplei.*

Treatment Prolonged antibiotics are recommended, although the optimal regimen is still uncertain. Relapses are common and often associated with the first manifestations of CNS involvement.

TROPICAL SPRUE

Tropical sprue is a poorly understood syndrome that is manifested by chronic diarrhea, steatorrhea, weight loss, and nutritional deficiencies, including both folate and vitamin B_{12}. Malabsorption of two unrelated substances is required for diagnosis. This disease occurs in 8–20% of people who have had an attack of infectious gastroenteritis in India and is considered by some to be a postinfectious complication. It is prevalent in some but not all tropical areas, including southern India, Pakistan, the Philippines, Puerto Rico, Haiti, and Cuba. It occurs in residents of as well as visitors to these areas.

Chronic diarrhea in a tropical environment is most often caused by infectious agents, including *Giardia lamblia*, *Yersinia enterocolitica*, *Entamoeba histolytica, Clostridioides difficile*, *Cryptosporidium parvum*, *Isospora belli*, *Strongyloides stercoralis*, and *Cyclospora cayetanensis*. Tropical sprue should not be entertained as a possible diagnosis until the presence of cysts and trophozoites has been excluded in three stool samples. **Infections of the GI tract and diarrhea are discussed in Chaps. 49, 138, 139, 169–173, and 230.**

Etiology Because tropical sprue responds to antibiotics, the consensus is that it may be caused by one or more infectious agents. Nonetheless, the etiology and pathogenesis of tropical sprue are uncertain. Its occurrence is not evenly distributed in all tropical areas; it is rarely observed in Africa, Jamaica, or Southeast Asia. The incidence of tropical sprue appears to have decreased substantially during the past two or three decades, perhaps in relation to improved sanitation in many tropical countries during this time. Some have speculated that the reduced occurrence is attributable to the wider use of antibiotics in acute diarrhea, especially in travelers to tropical areas from temperate countries. Folic acid is absorbed exclusively in the duodenum and proximal jejunum, and most patients with tropical sprue have evidence of folate malabsorption and depletion.

The clinical pattern of tropical sprue varies in different areas of the world (e.g., India vs Puerto Rico). Not infrequently, individuals in southern India initially report the occurrence of acute enteritis before the development of steatorrhea and malabsorption. In contrast, in Puerto Rico, a more insidious onset of symptoms and a more dramatic response to antibiotics are seen compared with some other locations. Tropical sprue in different areas of the world may not be the same disease, and similar clinical entities may have different etiologies.

Diagnosis The diagnosis of tropical sprue is based on an abnormal small-intestinal mucosal biopsy in an individual with chronic diarrhea and evidence of malabsorption who is either residing or has recently lived in a tropical country. The small-intestinal biopsy in tropical sprue does not reveal pathognomonic features but resembles, and can often be indistinguishable from, that seen in celiac disease **(Fig. 336-4)**. The biopsy sample in tropical sprue has less villous architectural alteration and more mononuclear cell infiltrate in the lamina propria. In contrast to those of celiac disease, the histologic features of tropical sprue manifest with a similar degree of severity throughout the small intestine, and a gluten-free diet does not result in either clinical or histologic improvement in tropical sprue.

Treatment Broad-spectrum antibiotics and folic acid are most often curative, especially if the patient leaves the tropical area and does not return. Tetracycline should be used for up to 6 months and may be associated with improvement within 1–2 weeks. Folic acid alone induces hematologic remission as well as improvement in appetite, weight gain, and some morphologic changes in small-intestinal biopsy. Because of marked folate deficiency, folic acid is most often given together with antibiotics.

ENVIRONMENTAL ENTERIC DYSFUNCTION

Environmental enteric dysfunction (EED; formerly known as tropical enteropathy) occurs in low- and middle-income countries and is characterized by altered crypt villus morphology with crypt atrophy and mucosal inflammation and barrier dysfunction, resulting in chronic low-grade malabsorption. It is associated with childhood stunting, a major worldwide health problem. Gut microbiota and adverse environmental and economic conditions contribute to its pathogenesis. Although the gold standard for diagnosis is small-bowel biopsy, this is often limited by lack of resources. Therapies for EED are the subject of intense investigation.

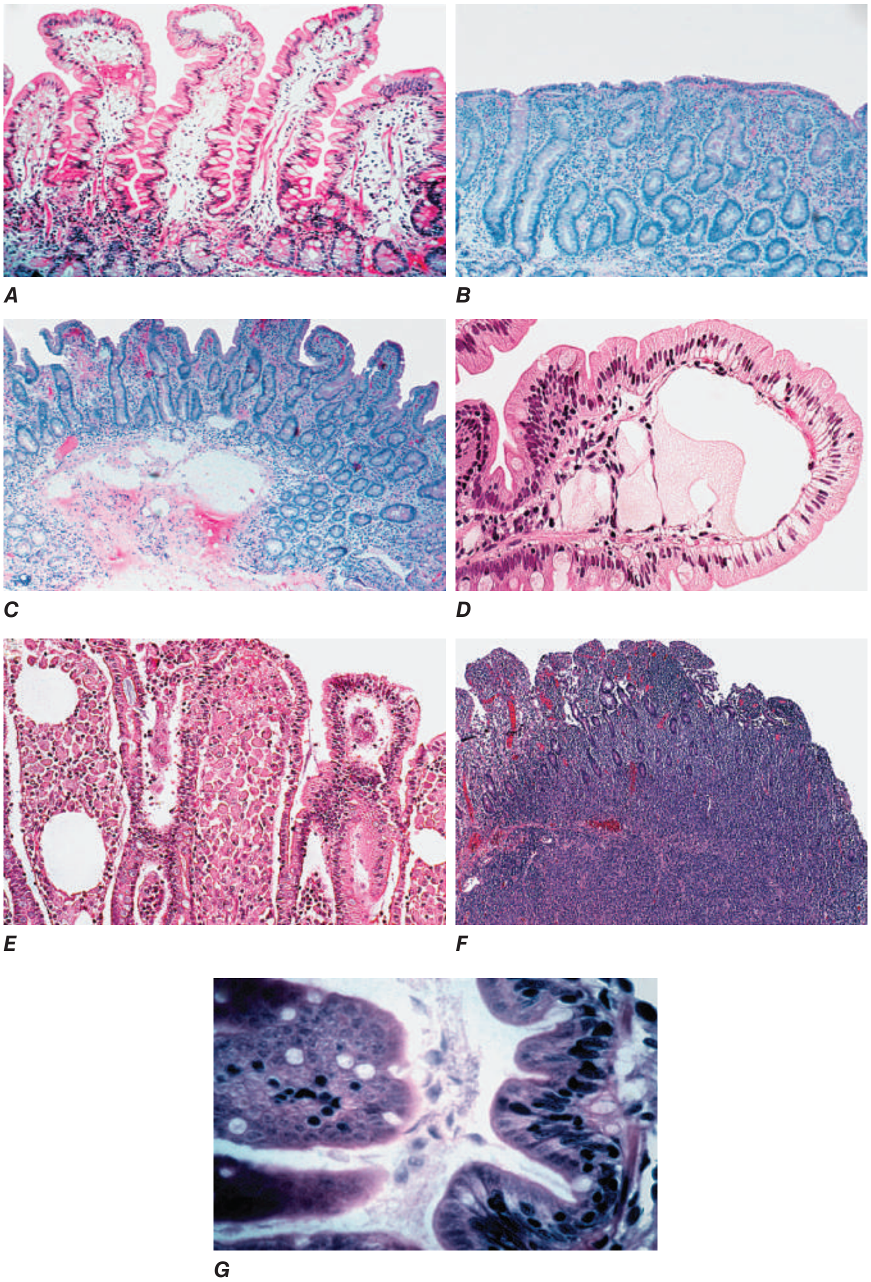

FIGURE 336-4 Small-intestinal mucosal biopsies. ***A.*** Normal individual. ***B.*** Untreated celiac disease. ***C.*** Treated celiac disease. ***D.*** Intestinal lymphangiectasia. ***E.*** Whipple's disease. ***F.*** Lymphoma. ***G.*** Giardiasis. *(Courtesy of Marie Robert, MD, Yale University; with permission.)*

SHORT-BOWEL SYNDROME

OVERVIEW

Short-bowel syndrome results from intestinal resection to treat a multitude of disorders including Crohn's disease, vascular diseases such as mesenteric arterial or venous thrombosis resulting in intestinal ischemia, volvulus, trauma, internal herniation, radiation enteritis, and diffuse carcinoma, among others. In children, the most common causes of short-bowel syndrome are necrotizing enterocolitis, intestinal atresias, volvulus, and malrotation. Short-bowel syndrome is defined as extensive removal of small intestine resulting in <200 cm of remaining small bowel. Intestinal failure is functionally defined as persistent parenteral nutrition dependence, generally found in patients who have <100 cm of remaining small bowel and no residual colon in continuity.

Clinical Features Loss of small-bowel surface area in short-bowel syndrome results in severe diarrhea, weight loss, and malabsorption of multiple nutrients, including fat, protein, and carbohydrate. The severity of symptoms and ultimate dependence on parenteral nutrition are generally related to the extent of resection, presence or absence of residual colon in continuity, retention of the ileocecal valve, and severity of the underlying disease. The intestine has a remarkable capacity to adapt to loss of small-bowel surface area, but this adaptive process is variable from patient to patient. Following resection, the adapting residual intestine exhibits an increase in crypt cell proliferation resulting in epithelial hyperplasia. The adaptive process generally continues for up to 2 years after resection, but improvements in nutrient, fluid, and electrolyte absorptive capacity have been reported even as late as 3–5 years after surgery. Massive diarrhea generally occurs in the first three postoperative months, associated with increased gastric acid secretion and malabsorption. Gradually, patients show enhanced functional capacity and reduced diarrhea. Specific nutrient deficiencies are dependent upon which segment of gut has been removed. For example, resection of the ileum results in loss of B_{12} absorptive and bile salt reabsorptive capacity. Malabsorbed bile salts reach the colon and cause a secretory diarrhea. In addition, resection of >100 cm of ileum results in such severe bile salt malabsorption that the liver cannot compensate by increased synthesis, thus precipitating fat malabsorption due to bile salt insufficiency/deficiency. Substantial resection of the colon also results in fluid and electrolyte loss and imbalance. The colon also plays a role in nutrient absorption because it metabolizes malabsorbed carbohydrate into short-chain fatty acids that can be absorbed by the colon and can contribute several hundred additional calories per day.

Long-Term Complications Because massive resection often leads to severe fat malabsorption, fat-soluble vitamin deficiency is common, and vitamin D deficiency can be very difficult to treat even with high-dose oral vitamin D supplementation, resulting in an increased risk of osteoporosis. Patients with a history of multiple surgeries often have extensive adhesive disease, and the residual intestine may have markedly abnormal motility or areas of stricturing and narrowing, resulting in recurrent bacterial overgrowth. The frequency of renal calcium oxalate stones increases in patients with a shortened small bowel with an intact colon in continuity; calcium is saponified in the intestinal luminal contents that contain fatty acids, freeing oxalate to be absorbed in the colon resulting in hyperoxaluria.

Treatment The major focus of treatment for short-bowel syndrome is to control diarrhea and normalize nutrient, fluid, and electrolyte absorption so that patients can maintain their weight and have a healthy nutritional status without the support of parenteral nutrition. Medications include opiates and derivatives including loperamide and diphenoxylate-atropine, which slow intestinal motility to allow for more contact time between luminal nutrients and the small-bowel mucosal surface. In the first year following resection, acid-blocking medications are used to treat gastric hypersecretion, including proton pump inhibitors or histamine 2 antagonists. Small-bowel bacterial overgrowth is common and is treated with antibiotics if suspected. The only medication that is specific for short-bowel syndrome but limited for use in parenteral nutrition or intravenous fluid–dependent patients is teduglutide, a glucagon-like peptide 2 (GLP-2) analogue that enhances crypt cell proliferation and villus hyperplasia and increases nutrient and fluid and electrolyte absorption. Patients treated with teduglutide have an average reduction of 20% of their parenteral nutrition requirements. Greater efficacy has been noted for patients without a residual colon, likely due to lower circulating endogenous GLP-2 levels compared to those with a colon in continuity.

Dietary Therapy Patients with short-bowel syndrome must consume three to four times their normal caloric intake to maintain their weight. The presence of luminal nutrients is required for the adaptive process to occur, so early feeding is recommended, even if parenteral nutrition is also required. These effects are most likely mediated by direct contact with the mucosa as well as stimulation of secretion of gut hormones such as GLP-2.

If the patient has all or part of their colon remaining in continuity, a low-fat diet is instituted to reduce the concentration of malabsorbed fatty acids that induce a secretory diarrhea. High complex carbohydrates are encouraged because when malabsorbed and present in the colon, they are converted to short-chain fatty acids and are absorbed, contributing several hundred additional kilocalories per day. All patients are asked to take a high-potency multivitamin on a daily basis. Patients in whom oral nutrition fails are fed with parenteral nutrition.

Monitoring Patients with short-bowel syndrome are at high risk for osteoporosis due to dietary calcium and vitamin D malabsorption, so they are periodically monitored for vitamin D deficiency and calcium levels and with dual x-ray absorptiometry (DEXA) studies to assess bone density. Malabsorption of vitamins and minerals is common; therefore, fat-soluble vitamins, vitamin B_{12}, folic acid, iron, magnesium, and zinc are monitored periodically. More unusual deficiencies include copper, selenium, and chromium, but these are usually seen in parenteral nutrition–dependent patients and can be corrected by adjusting daily intravenous dosages. Signs and symptoms of vitamin and mineral deficiency are also carefully monitored (e.g., hair loss, skin and nail changes, neurologic symptoms such as peripheral neuropathy).

DISORDERS OF POSTMUCOSAL ABSORPTION

Following uptake into enterocytes, nutrients are further processed and transported into the lymphatics or into the portal circulation for use by other cells throughout the body. Primary or secondary disorders of the lymphatics may result in significant diarrhea and malabsorption. Primary disorders of the intestinal lymphatics include intestinal lymphangiectasia, which may be congenital or acquired. Secondary causes of intestinal lymphatic damage or blockage include retroperitoneal fibrosis, fibrosing mesenteritis, and lymphoma. Circulatory causes of impaired delivery of nutrients from the intestine include Fontan physiology, congestive heart failure, and constrictive pericarditis. The end result of damage to lymphatic channels is malabsorption and diarrhea with concomitant protein-losing enteropathy.

PROTEIN-LOSING ENTEROPATHY

Protein-losing enteropathy refers to a large group of GI and non-GI disorders characterized by hypoproteinemia and edema in the absence of liver disease with reduced protein synthesis, or kidney disease with proteinuria. These diseases are characterized by excess protein loss in the GI tract. Diseases that may result in increased protein loss into the GI tract can be classified into three groups: (1) mucosal ulceration, such that the protein loss primarily represents exudation across damaged mucosa (e.g., ulcerative colitis, GI carcinomas); (2) nonulcerated mucosa, but with evidence of mucosal damage so that the protein loss represents loss across epithelia with altered permeability (e.g., celiac disease and Ménétrier's disease [hypertrophic gastropathy] in the small intestine and stomach, respectively); and (3) lymphatic dysfunction, representing either primary lymphatic disease or lymphatic disease secondary to partial lymphatic obstruction that may occur as a result of enlarged lymph nodes or cardiac disease. These result in increased lymphatic pressure, causing exudation of protein into the GI tract lumen.

Diagnosis The diagnosis of protein-losing enteropathy is suggested by diarrhea, peripheral edema, and low serum albumin and globulin levels in the absence of renal and hepatic disease. An individual with protein-losing enteropathy rarely has selective loss of *only* albumin or *only* globulins. Therefore, marked reduction of serum albumin with normal serum globulins should suggest renal and/or hepatic disease. Likewise, reduced serum globulins with normal serum albumin levels are more likely a result of reduced globulin synthesis rather than enhanced globulin loss into the intestine. Alpha-1 antitrypsin, a protein that accounts for ~4% of total serum proteins and is resistant to proteolysis, can be used to detect enhanced rates of serum protein loss into the intestinal tract but cannot be used to assess gastric protein loss because of its degradation in an acid milieu. Alpha-1 antitrypsin can be measured in a spot or 24-h stool collection and, if elevated, is diagnostic. A more accurate determination is alpha-1 antitrypsin clearance, measured by determining stool volume as well as both stool and plasma alpha-1 antitrypsin concentrations. In addition to the loss of protein via abnormal and distended lymphatics, peripheral lymphocytes may be lost via lymphatics, with consequent relative lymphopenia and specifically loss of CD3+ T cells. Thus, lymphopenia in a patient with hypoproteinemia indicates increased loss of protein into the GI tract.

Patients with increased protein loss into the GI tract from lymphatic obstruction often have steatorrhea and diarrhea. The steatorrhea is a result of altered lymphatic flow as lipid-containing chylomicrons exit from intestinal epithelial cells via intestinal lymphatics (Table 336-2; Fig. 336-4). In the absence of mechanical or anatomic lymphatic obstruction, intrinsic intestinal lymphatic dysfunction—with or without lymphatic dysfunction in the peripheral extremities—has been designated *intestinal lymphangiectasia*. Similarly, ~50% of individuals with intrinsic peripheral lymphatic disease (Milroy's disease) also have intestinal lymphangiectasia and hypoproteinemia. Other than steatorrhea and enhanced protein loss into the GI tract, all other aspects of intestinal absorptive function are normal in intestinal lymphangiectasia.

Endoscopy and Imaging Endoscopy (including double-balloon enteroscopy) with biopsy and video capsule endoscopy may be performed to rule out mucosal disease. Magnetic resonance enterography or computed axial tomography may provide additional insight into the underlying cause and can detect specific lesions, and lymphangiography can be used for diagnosis and potential therapeutic intervention.

Other Causes Ménétrier's disease (also called *hypertrophic gastropathy*) is an uncommon entity that involves the body and fundus of the stomach and is characterized by large gastric folds, reduced gastric acid secretion, and, at times, enhanced protein loss into the stomach.

Patients who have idiopathic protein-losing enteropathy without evidence of GI disease should be examined for cardiac disease. As more patients with congenital heart disease reach adulthood, Fontan physiology has become a more common cause of protein-losing enteropathy. Other cardiac causes include right-sided valvular disease and chronic pericarditis **(Chaps. 277, 278, 279 and 281)**.

Treatment As excess protein loss into the GI tract is most often secondary to a specific disease, treatment should be directed primarily to the underlying disease process and not to the hypoproteinemia. When enhanced protein loss is secondary to lymphatic obstruction, it is critical to establish the nature of this obstruction. Identification of mesenteric nodes or lymphoma may be possible by imaging studies. Similarly, it is important to exclude cardiac disease as a cause of protein-losing enteropathy. Patients with congenital heart disease may be examined by intranodal lymphangiography or noncontrast magnetic resonance lymphangiography and may undergo embolization of the site of lymphatic leakage or surgical lymphatic interventions to decompress the lymphatic system or target exclusion of abnormal lymphatic channels.

The increased protein loss that occurs in intestinal lymphangiectasia is a result of distended lymphatics associated with lipid malabsorption. The hypoproteinemia is treated with a low-fat, high-protein diet and the administration of medium-chain triglycerides, which do not exit from the intestinal epithelial cells via lymphatics but are delivered to the body via the portal vein. Other medical therapies including octreotide, a somatostatin analogue, intravenous heparin, budesonide and other glucocorticoids, and sirolimus have been studied but have generally been ineffective. Supportive care includes repletion of nutritional deficiencies and support stockings for leg edema control.

APPROACH TO THE PATIENT

Evaluation of the Patient with Suspected Malabsorption

The evaluation of patients with malabsorption is often challenging due to the large number of underlying disorders and the wide array of available tests. Thus, an extensive history and careful physical examination are essential to develop a more limited differential diagnosis and thereby avoid extensive and unnecessary testing.

HISTORY

A careful history should include questions about symptoms including abdominal pain, diarrhea, weight loss, bloating, symptoms or signs of selective nutrient deficiency including iron-deficiency anemia, bone fracture, or osteoporosis suggesting vitamin D and/or calcium deficiency, peripheral neuropathy resulting from vitamin B_{12} deficiency, hair loss that may result from generalized protein deficiency, predisposing disorders such as chronic pancreatitis or liver disease particularly involving the bile ducts such as primary biliary cholangitis or primary sclerosing cholangitis, history of small-bowel resection (due to Crohn's disease, trauma, ischemic bowel disease, etc.), and travel history. A multitude of nonspecific symptoms such as fatigue and weakness may also be reported. The protean manifestations of malabsorption and the underlying pathophysiology of clinical manifestations are summarized in **Table 336-5**.

PHYSICAL EXAMINATION

A careful physical examination may provide clues to underlying nutrient deficiencies and help assess severity of the malabsorptive process. For example, evidence of significant weight loss may be detected by bitemporal wasting and reduced arm circumference, iron deficiency may cause nail spooning, and vitamin B_{12} deficiency may result in significant peripheral neuropathy resulting in sensory reduction with tingling or numbness.

TABLE 336-5 Pathophysiology of Clinical Manifestations of Malabsorption Disorders

SYMPTOM OR SIGN	MECHANISM
Weight loss/malnutrition	Anorexia, malabsorption of nutrients
Diarrhea	Impaired absorption or secretion of water and electrolytes; colonic fluid secretion secondary to unabsorbed dihydroxy bile acids and fatty acids
Flatus	Bacterial fermentation of unabsorbed carbohydrate
Glossitis, cheilosis, stomatitis	Deficiency of iron, vitamin B_{12}, folate, and vitamin A
Abdominal pain	Bowel distention or inflammation, pancreatitis
Bone pain	Calcium, vitamin D malabsorption, protein deficiency, osteoporosis
Tetany, paresthesia	Calcium and magnesium malabsorption
Weakness	Anemia, electrolyte depletion (particularly K^+)
Azotemia, hypotension	Fluid and electrolyte depletion
Amenorrhea, decreased libido	Protein depletion, decreased calories, secondary hypopituitarism
Anemia	Impaired absorption of iron, folate, vitamin B_{12}
Bleeding	Vitamin K malabsorption, hypoprothrombinemia
Night blindness/ xerophthalmia	Vitamin A malabsorption
Peripheral neuropathy	Vitamin B_{12}, thiamine, vitamin E, pyridoxine, and niacin deficiency
Dermatitis	Deficiency of vitamin A, zinc, and essential fatty acid

LABORATORY EXAMINATION

Diseases that exclusively affect the proximal small intestine (e.g., celiac disease limited to the duodenum) may result in iron-deficiency anemia. Resection or disease of the terminal ileum frequently results in B_{12} deficiency since B_{12} absorption occurs exclusively in the ileum, causing a macrocytic anemia. Disorders that cause steatorrhea are almost invariably associated with fat-soluble vitamin deficiency, specifically vitamin D (very common), vitamin E, vitamin A, and vitamin K. The functional result of vitamin K deficiency is an elevated prothrombin time/international normalized ratio (INR), so this blood test is frequently measured instead of vitamin K levels. Serum carotene levels can suggest fat malabsorption but may decrease simply due to poor dietary consumption of leafy vegetables.

To diagnose steatorrhea, a spot stool can be submitted for Sudan III staining, which is specific for fecal fat. This is a useful qualitative but not quantitative test. Stool for elastase is helpful for diagnosing pancreatic insufficiency. A 24-h assessment of stool volume/weight is useful to establish the presence of clinically significant absorptive or secretory diarrhea versus diarrhea from other causes such as proctitis, which causes frequent, small, low-volume stools. The gold standard for documenting steatorrhea is the 72-h fecal fat collection, which is performed in concert with the patient's consumption of a 100-g fat diet. This test is highly accurate but difficult to obtain due to patient reluctance to collect stool. Also patients with fat malabsorption may poorly tolerate a 100-g fat diet. A diet with strictly quantified albeit reduced fat calories may be substituted. Finally, the calculation of the stool osmotic gap is a very useful and easy way to diagnose an osmotic diarrhea. A spot stool sample is sent to the lab for quantitation of fecal sodium and potassium concentration. Although stool osmolality can also be measured in the lab, measurements are often inaccurate due to bacterial degradation of nonabsorbed carbohydrate as the stool sits prior to examination.

Because normal stool osmolality reflects serum osmolality at 290 mOsm/kg H_2O, the osmotic gap may be calculated as follows:

$$290 - 2\,(\text{stool }[\mathrm{Na}+] + \text{stool }[\mathrm{K}+]).$$

If >50–100, a stool osmotic gap is present indicating the presence of unmeasured osmoles (e.g., malabsorbed lactose), and osmotic diarrhea can be diagnosed. If <50, one can presume a secretory component. Of note, malabsorbed fatty acids may also cause a secretory diarrhea by inducing secretion in the colon, so a malabsorptive diarrhea may have both an osmotic and secretory component. Extensive celiac disease may cause both osmotic diarrhea due to malabsorbed carbohydrate and also secretory diarrhea due to crypt hyperplasia.

Urinary D-xylose Test The urinary D-xylose test for carbohydrate absorption provides a measure of proximal small-bowel absorptive function. D-Xylose, a pentose, is absorbed almost exclusively in the proximal small intestine and is excreted in the urine. The D-xylose test is usually performed by administering 25 g of D-xylose and collecting urine for 5 h. An abnormal test (excretion of <4.5 g) primarily reflects duodenal/jejunal mucosal disease. The ease of obtaining a mucosal biopsy of the small intestine by endoscopy and the false-negative rate of the D-xylose test have led to its diminished use. When small-intestinal mucosal disease is suspected, a small-intestinal mucosal biopsy should be performed. The D-xylose test can also be abnormal in patients with delayed gastric emptying, impaired renal function, and sequestration in patients with large collections of fluid in a third space (i.e., ascites, pleural fluid).

Radiologic Examination A small-bowel follow-through barium examination may be very useful for detecting evidence of small-bowel diseases such as celiac disease, jejunal diverticulosis that predisposes to small-bowel bacterial overgrowth, or Crohn's disease (Fig. 336-3). Magnetic resonance enterography and computed tomography enterography are commonly used for diagnosis and management of inflammatory, stricturing disorders such as Crohn's disease and as an initial assessment of malabsorption, providing a means to visualize the entire luminal GI tract as well as the hepatobiliary tree and pancreas.

Endoscopic Evaluation and Small-Bowel Biopsies Endoscopy with small-bowel biopsy is essential in the evaluation of patients with documented steatorrhea or chronic diarrhea, as well as to evaluate abnormalities detected by radiologic imaging or by capsule endoscopy. In patients with documented steatorrhea and no evidence of pancreatic or hepatobiliary disease, an upper endoscopy and possible small-bowel enteroscopy are required to examine the small-bowel mucosa and to take biopsies for analysis. An upper endoscopy will visualize the stomach and duodenum; the maximum reach of the typical upper endoscopy scope is the ligament of Treitz. Small-bowel enteroscopy using a longer scope such as a pediatric colonoscope can be used to visualize the jejunum. Single- and double-balloon enteroscopy provide a means for examining much more of the jejunum and, if successful, will reach the ileum. Capsule endoscopy provides another means for visualizing the entire small bowel. Colonoscopy can be used for a retrograde view and biopsy of the terminal ileum.

Biopsy Analysis Small-bowel pathology may be divided into the three groups **(Table 336-6)** described below.

TABLE 336-6 Diseases That Can Be Diagnosed by Small-Intestinal Mucosal Biopsies

LESIONS	PATHOLOGIC FINDINGS
Diffuse, Specific	
Agammaglobulinemia	No plasma cells; either normal or absent villi ("flat mucosa")
Abetalipoproteinemia	Normal villi; epithelial cells vacuolated with fat postprandially
Patchy, Specific	
Intestinal lymphoma	Malignant cells in lamina propria and submucosa
Intestinal lymphangiectasia	Dilated lymphatics; clubbed villi
Eosinophilic gastroenteritis	Eosinophil infiltration of lamina propria and mucosa
Amyloidosis	Amyloid deposits
Crohn's disease	Noncaseating granulomas
Infection by one or more microorganisms (see text)	Specific organisms
Mastocytosis	Mast cell infiltration of lamina propria
Whipple's disease	Lamina propria includes macrophages containing material positive on periodic acid–Schiff staining; can be diffuse or patchy.
Diffuse, Nonspecific	
Celiac disease	Short or absent villi; mononuclear infiltrate; epithelial cell damage; hypertrophy of crypts
Tropical sprue	Similar to celiac disease
Bacterial overgrowth	Patchy damage to villi; lymphocyte infiltration
Folate deficiency	Short villi; decreased mitosis in crypts; megalocytosis
Vitamin B_{12} deficiency	Similar to folate deficiency
Radiation enteritis	Similar to folate deficiency
Zollinger-Ellison syndrome	Mucosal ulceration and erosion from acid
Protein-calorie malnutrition	Villous atrophy; secondary bacterial overgrowth
Drug-induced enteritis	Variable histology

1. *Diffuse histopathologic findings* involving the entire or majority of the mucosa that are specific for a particular disease entity; these include agammaglobulinemia (e.g., combined variable immunodeficiency) and abetalipoproteinemia. Immune globulin deficiency is associated with a variety of histopathologic findings on small-intestinal mucosal biopsy. The characteristic feature is the absence of or substantial reduction in the number of plasma cells in the lamina propria; the mucosal architecture may be either perfectly normal or flat (i.e., villous atrophy). Abetalipoproteinemia is characterized by a normal mucosal appearance except for the presence of mucosal absorptive cells that contain lipid postprandially and disappear after a prolonged period of either fat-free intake or fasting.
2. *Patchy lesions that are specific for a disease entity* include, for example, intestinal lymphoma or intestinal lymphangiectasia. Several diseases feature an abnormal small-intestinal mucosa with a patchy distribution. As a result, biopsy samples obtained randomly or in the absence of endoscopically visualized abnormalities may not reveal diagnostic features. Intestinal lymphoma can at times be diagnosed on mucosal biopsy by the identification of malignant lymphoma cells in the lamina propria and submucosa **(Chap. 113)**. Dilated lymphatics in the submucosa and sometimes in the lamina propria indicate lymphangiectasia associated with hypoproteinemia secondary to protein loss into the intestine. Eosinophilic gastroenteritis comprises a heterogeneous group of disorders with a spectrum of presentations and symptoms, with an eosinophilic infiltrate of the lamina propria, and with or without peripheral eosinophilia. The patchy nature of the infiltrate and its presence in the submucosa often lead to an absence of histopathologic findings on mucosal biopsy. Amyloid deposition can be identified by Congo red staining in some patients with amyloidosis involving the duodenum **(Chap. 117)**. Whipple's disease exhibits PAS-positive macrophages, and immunohistochemical analysis can detect the pathogenic organism. Although in severe cases the lesions are diffuse, patchy disease may also occur, requiring enteroscopy and sampling of multiple mucosal sites.
3. *Diffuse nonspecific lesions* may be found in more than one disorder. For example, villus atrophy/absence may be found in celiac disease, tropical sprue, or bacterial overgrowth, among other disorders. Several microorganisms can be identified in small-intestinal biopsy samples, establishing a correct diagnosis. At times, the biopsy is

performed specifically to diagnose infection (e.g., Whipple's disease or giardiasis). In most other instances, the infection is detected incidentally during the workup for diarrhea or other abdominal symptoms. Many of these infections occur in immunocompromised patients with diarrhea; the etiologic agents include *Cryptosporidium*, *Isospora belli*, microsporidia, *Cyclospora*, *Toxoplasma*, cytomegalovirus, adenovirus, *Mycobacterium avium-intracellulare*, and *G. lamblia*. In immunocompromised patients, when *Candida, Aspergillus, Cryptococcus*, or *Histoplasma* organisms are seen on duodenal biopsy, their presence generally reflects systemic infection. Apart from Whipple's disease and infections in the immunocompromised host, small-bowel biopsy is seldom used as the primary mode of diagnosis of infection. Even giardiasis is more easily diagnosed by stool antigen studies and/or duodenal aspiration than by duodenal biopsy.

SUMMARY

The evaluation and management of patients with disorders of absorption are challenging due to the complexity of the underlying pathophysiology and the large number of associated diseases. A diagnostic approach based on the information summarized in Tables 336-1 and 336-5 should prove useful for guiding the care of these challenging patients.

Acknowledgments

Henry Binder wrote this chapter in prior editions and some material from his chapter has been retained.

FURTHER READING

Bering J, DiBaise JK: Short bowel syndrome: Complications and management. Nutr Clin Pract 38:S46, 2023.

Boumaza A et al: Whipple's disease and *Tropheryma whipplei* infections: From bench to bedside. Lancet Infect Dis 22:e280, 2022.

Bousaba J et al: Bile acid diarrhea: As bad as it gets? Curr Opin Gastroenterol 39:184, 2023.

Cowardin CA et al: Environmental enteric dysfunction: Gut and microbiota adaptation in pregnancy and infancy. Nat Rev Gastroenterol Hepatol 20:223, 2023.

Fernández-Bañares F et al: Carbohydrate maldigestion and intolerance. Nutrients 14:1923, 2022.

Gregory M, Rubin DC: Small bowel disorders, in *The Washington Manual Gastroenterology Subspecialty Consult*, 4th ed, Gyawali CP et al (eds). Walters Kluwer, 2021, pp.142–155.

Jansson-Knodell CL, Rubio-Tapia A: Gluten-related disorders from bench to bedside. Clin Gastroenterol Hepatol 22:693, 2024.

Johnson LR: Digestion and absorption of nutrients, in *Gastrointestinal Physiology*, 9th ed. Philadelphia, Elsevier, 2019, pp 102–120.

Ozen A, Lenardo MJ: Protein-losing enteropathy. N Engl J Med 389:733, 2023.

Zafar H et al: Small intestinal bacterial overgrowth: current update. Curr Opin Gastroenterol 39:522, 2023.

337

Inflammatory Bowel Disease

Sonia Friedman, Richard S. Blumberg

Inflammatory bowel disease (IBD) is a chronic idiopathic inflammatory disease of the gastrointestinal tract. Ulcerative colitis (UC) and Crohn's disease (CD) are the two major types of IBD.

GLOBAL CONSIDERATIONS: EPIDEMIOLOGY

UC and CD have emerged as global diseases in the twenty-first century. The disease burden is high, with a prevalence of >0.3% in North America, Oceania, and most countries in Europe. In newly industrialized countries in Africa, Asia, and South America where there is increased urbanization and Westernization, the incidence of IBD has been rising and mirrors the prior increase of IBD in the Western world in the twentieth century. For example, in Brazil, the annual percent change is +11.1% (95% confidence interval [CI], 4.8–17.8%) for CD and +14.9% (95% CI, 10.4–19.6%) for UC, whereas in Taiwan, the annual percent change is +4.0% (95% CI, 1.0–7.1%) for CD and +4.8% (95% CI, 1.8–8.0%) for UC. In a study of newly diagnosed IBD cases between 2011 and 2013 from 13 countries or regions in the Asia Pacific, the mean annual IBD incidence per 100,000 was 1.50 (95% CI, 1.43–1.57). India (9.31; 95% CI, 8.38–10.31) and China (3.64; 95% CI, 2.97–4.42) had the highest IBD incidences in Asia. The highest reported prevalence values were in Europe (UC, 505 per 100,000 in Norway; CD, 322 per 100,000 in Germany) and North America (UC, 286 per 100,000 in the United States; CD, 319 per 100,000 in Canada). The most recent U.S. study showed that 2.39 million Americans are diagnosed with IBD. The most likely factors that explain the geographic variability of IBD rates, especially the rising incidence in developing countries and urban areas, are environmental variables including changes in diet (with downstream effects on the intestinal microbiota), exposure to sunlight or temperature differences, and socioeconomic status and hygiene **(Table 337-1)**.

Peak incidence of UC and CD is in the second to fourth decades, with 78% of CD studies and 51% of UC studies reporting the highest incidence among those aged 20–29 years old. A second modest rise in incidence occurs between the seventh and ninth decades of life. The female-to-male ratio ranges from 0.51 to 1.58 for UC studies and 0.34 to 1.65 for CD studies, suggesting that the diagnosis of IBD is not gender-specific. Pediatric IBD (patients <17 years old) composes ~20–25% of all IBD patients, and ~5% of all IBD patients are <10 years old. Children with IBD are also grouped as those with early-onset (EO) IBD (patients <10 years old), very-early-onset (VEO) IBD (patients <6 years old), and infantile IBD (patients <2 years old). VEOIBD and infantile IBD mainly affect the colon and are resistant to standard medications, and patients often have a strong family history of IBD, with at least one first-degree relative affected. In infantile IBD or VEOIBD, a number of rare, single genetic mutations have been identified as the basis for this susceptibility in up to 10% of patients, suggesting a simple Mendelian origin of the disease in these cases.

Although the greatest prevalence of IBD is among Whites, the prevalence of IBD in Latinx, Black, and Asian people is increasing. Urban areas have a higher prevalence of IBD than rural areas, and high socioeconomic classes have a higher prevalence than lower socioeconomic classes.

Epidemiologic studies have identified a number of potential environmental factors that are associated with disease risk **(Fig. 337-1)**.

TABLE 337-1 Epidemiology of IBD

	ULCERATIVE COLITIS	CROHN'S DISEASE
Age of onset	Second to fourth decades and seventh to ninth decades	Second to fourth decades and seventh to ninth decades
Ethnicity	White > Black > Latinx > Asian	
Female-to-male ratio	0.51–1.58	0.34–1.65
Smoking	May prevent disease (odds ratio 0.58)	May cause disease (odds ratio 1.76)
Oral contraceptives	No increased risk	Hazard ratio 2.82
Appendectomy	Protective (risk reduction 13–26%)	Not protective
Monozygotic twins	6–18% concordance	38–58% concordance
Dizygotic twins	0–2% concordance	4% concordance
Infections in the first year of life	1.6 and 3 times the risk of developing IBD by age 10 and 20 years	

Abbreviation: IBD, inflammatory bowel disease.

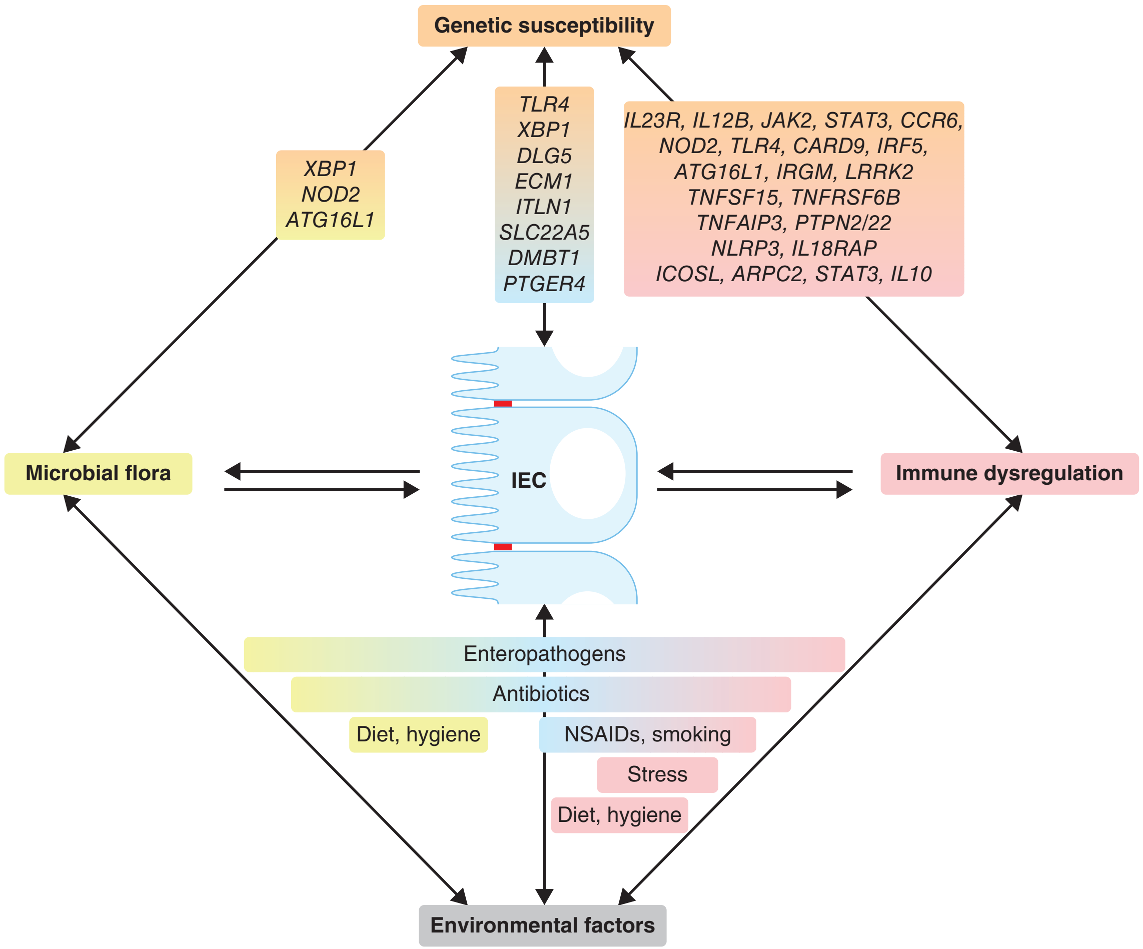
Genetic susceptibility
XBP1
NOD2
ATG16L1
TLR4
XBP1
DLG5
ECM1
ITLN1
SLC22A5
DMBT1
PTGER4
IL23R, IL12B, JAK2, STAT3, CCR6,
NOD2, TLR4, CARD9, IRF5,
ATG16L1, IRGM, LRRK2
TNFSF15, TNFRSF6B
TNFAIP3, PTPN2/22
NLRP3, IL18RAP
ICOSL, ARPC2, STAT3, IL10
Microbial flora
IEC
Immune dysregulation
Enteropathogens
Antibiotics
Diet, hygiene
NSAIDs, smoking
Stress
Diet, hygiene
Environmental factors

FIGURE 337-1 Pathogenesis of inflammatory bowel disease (IBD). In IBD, the tridirectional relationship between the commensal flora (microbiota), intestinal epithelial cells (IECs), and mucosal immune system is dysregulated, leading to chronic inflammation. Each of these three factors is affected by genetic and environmental factors that determine risk for the disease. NSAIDs, nonsteroidal anti-inflammatory drugs. (*Republished with permission Annual Review of Immunology from Inflammatory Bowel Disease, A Kaser et al: 28:573, 2010. Permission conveyed through Copyright Clearance Center, Inc.*)

Smoking is an important risk factor in IBD with opposite effects on UC (odds ratio [OR] 0.58) and CD (OR 1.76) that may be influenced by genetic risk factors and ethnic origin. Previous appendectomy with confirmed appendicitis (risk reduction of 13–26%), particularly at a young age, has a protective effect on the development of UC across different geographical regions and populations. Appendectomy is modestly associated with the development of CD, but this may be due to diagnostic bias. Oral contraceptive use is associated with an increased risk of CD, with a reported hazard ratio as high as 2.82 among current users and 1.39 among past users. The association between oral contraceptive use and UC is limited to women with a history of smoking. Breast-feeding may protect against the development of IBD, which supports the potential importance of early-life exposures to later IBD development. Infections in the first year of life are associated with development of IBD, especially before the ages of 10 and 20 years. Infectious gastroenteritis with pathogens (e.g., *Salmonella*, *Shigella*, *Campylobacter* spp., *Clostridioides difficile*) increases IBD risk by two- to threefold. Diets high in animal protein, sugars, sweets, oils, fish and shellfish, and dietary fat, especially ω-6 fatty acids, and low in ω-3 fatty acids have been implicated in increasing the risk of IBD. A protective effect of vitamin D on the risk of CD has been reported.

IBD is a familial disease in 5–10% of patients (Fig. 337-2), and the strongest risk factor for the development of IBD is a first-degree relative with the disease. The children of mothers and fathers with UC have an approximately fourfold increased risk of UC, and the children of mothers and fathers with CD have an almost eightfold increased risk of CD. Some of these patients may exhibit early-onset disease during the first decade of life and, in CD, a concordance of anatomic site and clinical type within families. In twin studies, 38–58% of monozygotic twins are concordant for CD, and 6–18% are concordant for UC, whereas 4% of dizygotic twins are concordant for CD, and 0–2% are concordant for UC in Swedish and Danish cohorts. In the remainder of patients, IBD is observed in the absence of a family history (i.e., sporadic disease).

GLOBAL CONSIDERATIONS: IBD PHENOTYPES

IBD location and behavior show racial differences that may reflect underlying genetic variations and have important implications for diagnosis and management of disease. Blacks and Latinxs tend to have an ileocolonic CD distribution. Data from East Asia show that

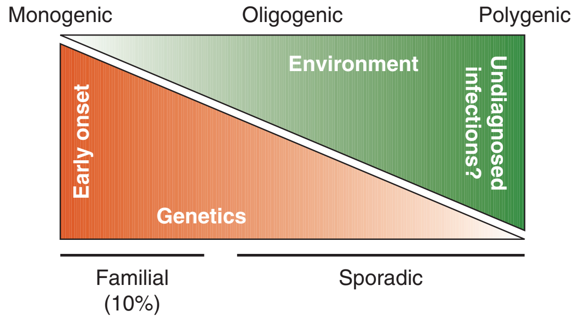
Monogenic
Oligogenic
Polygenic
Environment
Undiagnosed infections?
Early onset
Genetics
Familial (10%)
Sporadic

FIGURE 337-2 A model for the syndromic nature of inflammatory bowel disease (IBD). Genetic and environmental factors variably influence the development and phenotypic manifestations of IBD. At the one extreme, IBD is exemplified as a simple Mendelian disorder as observed in early-onset IBD due to single-gene defects such as *IL10*, *IL10RA*, and *IL10RB*; and at the other extreme, it may be exemplified by as yet to be described emerging infectious diseases. (*Reproduced with permission from A Kaser et al: Genes and environment: How will our concepts on the pathophysiology of IBD develop in the future? Dig Dis 28:395, 2010.*)

ileocolonic CD is the most common CD phenotype (50.5–71%) and perianal disease is more common in East Asian patients (30.3–58.8%) than Whites (25.1–29.6%). Pancolonic disease is more common than left-sided colitis or proctitis among Black, Latinx, and Asian patients with UC. Older Asian patients with UC (age >60) tend to have a more aggressive disease course. Among Blacks, joint involvement is the predominant extraintestinal manifestation (EIM) reported and ranges from 15.7 to 29.6%. Ocular involvement is also common in African Americans and ranges from 7.1 to 13%. Dermatologic manifestations are the most common EIMs reported in Latinxs (10–13%). These ethnic variations indicate the importance of different genetic and/or environmental factors in the pathogenesis of this disorder.

ETIOLOGY AND PATHOGENESIS

Under physiologic conditions, homeostasis normally exists between the commensal microbiota, epithelial cells that line the interior of the intestines (intestinal epithelial cells [IECs]), and immune cells within the tissues (Fig. 337-1). A consensus hypothesis is that each of these three major host compartments that function together as an integrated "supraorganism" (microbiota, IECs, and immune cells) are affected by specific environmental (e.g., smoking, antibiotics, enteropathogens) and genetic factors that, in a susceptible host, cumulatively and interactively disrupt homeostasis during the course of one's life and, in so doing, culminate in a chronic state of dysregulated inflammation; i.e., IBD. Although chronic activation of the mucosal immune system may represent an appropriate response to an infectious agent, a search for such an agent has thus far been unrewarding in IBD. As such, IBD is currently considered an inappropriate immune response to the endogenous (autochthonous) commensal microbiota within the intestines, with or without some component of autoimmunity. Importantly, the normal, uninflamed intestines contain a large number of immune cells that are in a unique state of activation, in which the gut is restrained from full immunologic responses to the commensal microbiota and dietary antigens by very powerful regulatory pathways that function within the immune system (e.g., T regulatory cells that express the FoxP3 transcription factor and suppress inflammation). Maintenance of homeostasis also involves oversight from local parenchymal cells including nerve, endothelial, and stromal cells, as well as the commensal microbiota that provide essential remedial factors necessary for health and serve as a target of the immune response. During the course of infections or other environmental stimuli in the normal host, full activation of the lymphoid tissues in the intestines occurs but is rapidly superseded by dampening of the immune response and tissue repair. In IBD, such processes may not be regulated normally.

GENETIC CONSIDERATIONS

The genetic underpinning of IBD is known from its concordance in identical twins, its occurrence in the context of several genetic syndromes, and the development of severe, refractory IBD in early life in association with single-gene defects that affect the immune system **(Table 337-2)**. More than 60 different gene defects have been identified in patients with VEOIBD and infantile IBD by whole-exome sequencing (WES), in whom most of the monogenic mutations have been discovered. These include mutations in genes encoding, for example, interleukin (IL) 10, the IL-10 receptor (IL-10R), cytotoxic T-lymphocyte-associated protein-4 (*CTLA4*), neutrophil cytosolic factor 2 protein (*NCF2*), X-linked inhibitor of apoptosis protein (*XIAP*), lipopolysaccharide responsive and beige-like anchor protein (*LRBA*), and tetratricopeptide repeat domain 7A protein (*TTC7*), among many other genes that are involved in host-commensal interactions. A monogenic etiology may also be possible in a small subset of adult patients with IBD. In addition, IBD has a familial origin in at least 10% of afflicted adults, consistent with an inherited basis for this disease (Fig. 337-2). However, the majority of pediatric and adult IBD cases are multigenic (or polygenic) in origin, suggesting a syndromic nature of this disease that gives rise to multiple clinical subgroups beyond the simple classification as UC and CD. The polygenic nature of the disease has been elucidated through a variety of genetic approaches, including candidate gene studies, linkage analysis, and genome-wide association studies (GWAS) that focus on the identification of disease-associated single nucleotide polymorphisms (SNPs) within the human genome and WES and whole-genome sequencing to elucidate the specific mutations potentially involved. GWAS have identified ~240 genetic loci with multiple potential candidate genes; two-thirds of these loci are associated with both disease phenotypes, with the remainder being specific for either CD or UC **(Table 337-3)**. These genetic similarities account for the overlapping immunopathogenesis and consequently epidemiologic observations of both diseases in the same families and similarities in response to therapies. Because the specific causal variants for each identified gene or locus are mostly unknown as most risk loci are contained within regulatory (noncoding) regions of the associated genes, it is not clear whether the similarities in the genetic risk factors associated with CD and UC are shared at a structural or functional level. The risk conferred by each identified gene or locus is unequal and generally small, such that only ~20% of the disease risk is considered to be explained by the current genetic information. Further, many of the genetic risk factors identified are also observed to be associated with risk for other immune-mediated diseases, suggesting that related immunogenetic pathways are involved in the pathogenesis of multiple different disorders, accounting for the common responsiveness to similar types of biologic therapies (e.g., anti–tumor necrosis factor [TNF] therapies) and possibly the simultaneous occurrence of these disorders. The diseases and the genetic risk factors that are shared with IBD include, for example, rheumatoid arthritis (*TNFAIP3*), psoriasis (*IL23R*, *IL12B*), ankylosing spondylitis (*IL23R*), type 1 diabetes mellitus (*IL10*, *PTPN2*), asthma (*ORMDL3*), and systemic lupus erythematosus (*TNFAIP3*, *IL10*), among others.

TABLE 337-2 Primary Genetic Disorders Associated with IBD

NAME	GENETIC ASSOCIATION	PHENOTYPE
Turner's syndrome	Loss of part or all of X chromosome	Associated with UC and colonic CD
Hermansky-Pudlak syndrome	Autosomal recessive disorder genes involved in the biogenesis of lysosome-related organelles or adaptor protein-3 complex	Granulomatous colitis, oculocutaneous albinism, platelet dysfunction, and pulmonary fibrosis
Wiskott-Aldrich syndrome (WAS)	X-linked recessive disorder, loss of WAS protein function	Colitis, immunodeficiency, severely dysfunctional platelets, and thrombocytopenia
Glycogen storage disease type 1b	Autosomal recessive disorder of *SLC37A4* resulting in deficiency of the glucose-6-phosphate translocase	Granulomatous colitis, presents in infancy with hypoglycemia, growth failure, hepatomegaly, and neutropenia
Immune dysregulation polyendocrinopathy, enteropathy X-linked (IPEX)	Loss of FoxP3 transcription factor and T regulatory cell function	UC-like autoimmune enteropathy, with endocrinopathy (neonatal type 1 diabetes or thyroiditis), dermatitis
Early-onset IBD	Deficient IL-10 and IL-10 receptor function	Severe, refractory IBD in early life

Abbreviations: CD, Crohn's disease; IBD, inflammatory bowel disease; IL, interleukin; UC, ulcerative colitis.

The genetic factors that are recognized to mediate risk for IBD have highlighted the importance of shared mechanisms of disease that variably affect CD and/or UC (Table 337-3). These include the following: those genes that are associated with fundamental cell biologic processes such as the unfolded protein response due to endoplasmic reticulum stress, autophagy, and metabolism that regulate the ability of cells to manage the physiologic needs of the intestinal environment; those associated with innate immunity associated with nonlymphoid cells that function in responses to and control of microbes; those associated with the regulation of adaptive immunity that control the balance between inflammatory and anti-inflammatory cellular pathways associated with lymphocytes; and, finally, those that are involved in the development and resolution of inflammation associated with

TABLE 337-3 Some Genetic Loci Associated with Crohn's Disease and/or Ulcerative Colitis

CHROMOSOME	PUTATIVE GENE	GENE NAME	PROTEIN FUNCTION	CD	UC
Unfolded Protein Response, Autophagy, and Metabolism					
2q37	*ATG16L1*	ATG16 autophagy related 16-like 1	Autophagy	+	
5q31	*SLC22A5*	Solute carrier family 22, member 5	β-Carnitine transporter	+	
5q33	*IRGM*	Immunity-related GTPase family, M	Autophagy	+	
7p21	*AGR2*	Anterior gradient 2	Unfolded protein response	+	+
12q12	*LRRK2*	Leucine-rich repeat kinase 2	Autophagy	+	
13q14	*LACC1*	Laccase domain containing 1 (FAMIN)	Immunometabolic regulator	+	
17q21	*ORMDL3*	Orosomucoid related member 1-like 3	Unfolded protein response and lipid synthesis	+	+
22q12	*XBP1*	X-box binding protein 1	Unfolded protein response	+	+
Innate Immunity					
1q23	*ITLN1*	Intelectin 1	Bacterial binding	+	
16q12	*NOD2*	Nucleotide-binding oligomerization domain containing 2	Bacterial sensing and autophagy activation	+	
Adaptive Immunity					
1p31	*IL23R*	Interleukin 23 receptor	T_H17 cell stimulation	+	+
1q32	*IL10*	Interleukin 10	Treg-associated cytokine		+
5q33	*IL12B*	Interleukin 12 subunit beta	IL-12 p40 chain of IL-12/IL-23	+	+
18p11	*PTPN2*	Protein tyrosine phosphatase, nonreceptor type 2	T-cell regulation	+	
Inflammation and Healing					
3p21	*MST1*	Macrophage Stimulating 1	Macrophage activation	+	+
5p13	*PTGER4*	Prostaglandin E receptor 4	PGE_2 receptor	+	+
6q23	*TNFAIP3*	Tumor necrosis factor, alpha-induced protein 3 (A20)	Toll-like receptor regulation	+	
6q27	*CCR6*	Chemokine (C-C motif) receptor 6	Dendritic cell migration	+	
9p24	*JAK2*	Janus kinase 2	IL-6R and IL-23R signaling	+	+
9q32	*TNFSF15*	Tumor necrosis factor–like cytokine 1A (TL1A)	Promotes inflammation and fibrosis	+	+
17q21	*STAT3*	Signal transducer and activator of transcription 3	IL-6R, IL-23R, and IL-10R signaling	+	+

Abbreviations: CD, Crohn's disease; GTPase, guanosine triphosphatase; IL, interleukin; PGE_2, prostaglandin E_2; Treg, T regulatory cell; UC, ulcerative colitis.

Source: Adapted from A Kaser et al: Ann Rev Immunol 28:573, 2010; Graham DB, Xavier RJ: Nature 578:527, 2020.

healing that control leukocyte recruitment and inflammatory mediator production or the development of fibrosis. Each of these genetic susceptibilities contributes in an incremental manner to IBD risk, variably affects the activities of virtually all subtypes of immune and nonimmune cells within the intestines, and encodes mutations (polymorphisms) that promote or protect from IBD. Some of these loci are associated with specific subtypes of disease such as the association between *NOD2* polymorphisms and fibrostenosing CD or *ATG16L1* and fistulizing disease, especially within the ileum. However, the clinical utility of these genetic risk factors for the diagnosis or determination of prognosis and therapeutic responses remains to be defined.

COMMENSAL MICROBIOTA AND IBD

The endogenous commensal microbiota within the intestines plays a central role in the pathogenesis of IBD. Humans are born with sterile guts and acquire their commensal microbiota initially from the mother during egress through the birth canal and subsequently from environmental sources. A stable configuration of up to 1000 species of bacteria that achieves a biomass of ~10^{12} colony-forming units per gram of feces is achieved by 3 years of age, which likely persists into adult life, with each individual human possessing a unique combination of species. In addition, the intestines contain other microbial life forms including fungi, archaea, viruses, and protists. The microbiota is thus considered as a critical and sustaining component of the human organism. The establishment and maintenance of the intestinal microbiota composition and function are under the control of host (e.g., immune and epithelial responses), environmental (e.g., diet and antibiotics), and likely genetic (e.g., *NOD2*) factors (Fig. 337-1). In turn, the microbiota, through its structural components and metabolic activity, has major influences on the epithelial and immune function of the host, which, through epigenetic effects, may have durable consequences. During early life when the commensal microbiota is being established, these microbial effects on the host may be particularly important in determining later life risk for IBD. Specific components of the microbiota can promote or protect from disease. The commensal microbiota in patients with both UC and CD is demonstrably different from that of nonafflicted individuals, a state of dysbiosis suggesting the presence of microorganisms that drive disease (e.g., Proteobacteria such as enteroinvasive and adherent *Escherichia coli*) and to which the immune response is directed and/or the loss of microorganisms that hinder inflammation (e.g., Firmicutes such as *Faecalibacterium prausnitzii*). Many of the changes in the commensal microbiota occur as a consequence of the inflammation and are thus potential secondary drivers of disease. In addition, agents that alter the intestinal microbiota such as metronidazole, ciprofloxacin, and elemental diets, may improve CD. CD may also respond to fecal diversion, demonstrating the ability of luminal contents to exacerbate disease.

DEFECTIVE IMMUNE REGULATION IN IBD

The mucosal immune system does not normally elicit an inflammatory immune response to luminal contents due to oral (mucosal) tolerance. Administration of soluble antigens orally, rather than subcutaneously or intramuscularly, leads to antigen-specific control of the response and the host's ability to tolerate the antigen. Multiple mechanisms are involved in the induction of oral tolerance and include deletion or anergy (nonresponsiveness) of antigen-reactive T cells or induction of CD4+ T cells that suppress gut inflammation (e.g., T regulatory cells expressing the FoxP3 transcription factor) and that secrete anti-inflammatory cytokines such as IL-10, IL-35, and transforming growth factor β (TGF-β). Oral tolerance may be responsible for the lack of immune responsiveness to dietary antigens and the commensal microbiota in the intestinal lumen. In IBD, this suppression of inflammation is altered, leading to uncontrolled inflammation. The mechanisms of this regulated immune suppression are incompletely known.

Gene knockout ($^{-/-}$) or transgenic (Tg) mouse models of IBD, including those that are directed at genes associated with risk for the human disease, have revealed that deleting specific cytokines (e.g., IL-2, IL-10, TGF-β) or their receptors, deleting molecules associated with T-cell antigen recognition (e.g., T-cell antigen receptors), or interfering with IEC barrier function and the regulation of responses to commensal bacteria (e.g., mucus glycoproteins or nuclear factor-κB [NF-κB]) leads to spontaneous colitis or enteritis. In the majority of circumstances, intestinal inflammation in these animal models requires the presence of the commensal microbiota. However, in some cases, activation of certain elements of the intestinal immune system may be exacerbated by the absence of bacteria, resulting in severe colitis and emphasizing the presence of protective properties of the commensal microbiota. Thus, a variety of specific alterations in either the microbiota or host can lead to uncontrolled immune activation and inflammation directed at the intestines in mice. How these relate to human IBD remains to be defined, but they are consistent with inappropriate responses of the genetically susceptible host to the commensal microbiota.

THE INFLAMMATORY CASCADE IN IBD

In both UC and CD, inflammation likely emerges from the genetic predisposition of the host in the context of yet-to-be-defined environmental factors. Once initiated in IBD by abnormal innate immune sensing of bacteria by parenchymal cells (e.g., IECs) and hematopoietic cells (e.g., macrophages, dendritic cells), the immune inflammatory response is perpetuated by T-cell and B-cell activation when coupled together with inadequate regulatory pathways. A sequential cascade of inflammatory mediators extends the response, making each step a potential target for therapy. Inflammatory cytokines from innate immune cells such as IL-1, IL-6, IL-12, IL-23, and TNF have diverse effects on tissues. They promote fibrogenesis, collagen production, activation of tissue metalloproteinases, and/or the production of other inflammatory mediators; they also activate the coagulation cascade in local blood vessels (e.g., increased production of von Willebrand factor). These cytokines are normally produced in response to infection but are usually turned off or inhibited by cytokines such as IL-10 and TGF-β at the appropriate time to limit tissue damage. In IBD, their activity is not regulated, resulting in an imbalance between the proinflammatory and anti-inflammatory mediators. Some of these cytokines activate other inflammatory cells (macrophages and B cells), and others such as chemokines act indirectly to recruit other lymphocytes, inflammatory leukocytes, and mononuclear cells from the bloodstream into the gut through interactions between homing receptors on leukocytes (e.g., α4β7 integrin) and addressins on vascular endothelium (e.g., MadCAM1). Lymphocytes such as CD4+ T helper (T_H) cells emerge from the lymph nodes under the influence of sphingosine-1-phosphate (S1P) gradients that act on S1P receptors (S1PR) expressed on the lymphocyte and endothelium. CD4+ T_H cells that promote inflammation are of four major types, all of which may be associated with colitis in animal models and perhaps humans: T_H1 cells (secrete IL-2, interferon [IFN] γ), T_H2 cells (secrete IL-4, IL-5, IL-13), T_H9 cells (secrete IL-9), and T_H17 cells (secrete IL-17, IL-21, IL-22). T_H17 cells may also provide protective functions. Innate immune-like cells (ILCs) that lack T-cell receptors are also present in intestines, polarize to the same functional fates, and may similarly participate in IBD. T_H1 cells induce transmural granulomatous inflammation that resembles CD; T_H2 cells and related natural killer T cells that secrete IL-4, IL-5, and IL-13 induce superficial mucosal inflammation resembling UC in animal models; T_H9 cells promote allergic-like inflammation; and T_H17 cells may be responsible for neutrophilic recruitment. Each of these T-cell subsets cross-regulates each other. The T_H1 cytokine pathway is initiated by IL-12, a key cytokine in the pathogenesis of experimental models of mucosal inflammation. IL-4 and IL-23, together with IL-6 and TGF-β, induce T_H2 and T_H17 cells, respectively, and IL-23 inhibits the suppressive function of regulatory T cells. Activated macrophages secrete TNF and IL-6. Cytokines produced by innate immune cells and lymphocytes exert their effects on target cells through cytokine-specific receptors, which precisely engage intracellular Janus kinases (JAK) 1, 2, and/or 3 for transmission of intracellular activating signals.

These characteristics of the immune response in IBD explain the beneficial therapeutic effects of antibodies to block proinflammatory cytokines or the signaling by their receptors (e.g., anti-TNF, anti-IL-12, anti-IL-23, JAK inhibitors), molecules that antagonize the activity of S1PR-induced emigration of lymphocytes from the lymph nodes (ozanimod), or antibodies that impede leukocyte recruitment into the intestines (e.g., anti-α4β7). They also highlight the potential usefulness of cytokines that inhibit inflammation and promote regulatory T cells or promote intestinal barrier function (e.g., IL-10) in the treatment of IBD. Therapies such as the 5-aminosalicylic acid (5-ASA) compounds and glucocorticoids are also potent inhibitors of these inflammatory mediators through inhibition of transcription factors such as NF-κB that regulate their expression.

PATHOLOGY

ULCERATIVE COLITIS: MACROSCOPIC FEATURES

UC is a mucosal disease that usually involves the rectum and extends proximally to involve all or part of the colon. About 40–50% of patients have disease limited to the rectum and rectosigmoid, 30–40% have disease extending beyond the sigmoid but not involving the whole colon, and 20% have a pancolitis. Proximal spread occurs in continuity without areas of uninvolved mucosa. When the whole colon is involved, the inflammation extends 2–3 cm into the terminal ileum in 10–20% of patients. The endoscopic changes of *backwash ileitis* are superficial and mild and are of little clinical significance. Although variations in macroscopic activity may suggest skip areas, biopsies from normal-appearing mucosa are usually abnormal. Thus, it is important to obtain multiple biopsies from apparently uninvolved mucosa, whether proximal or distal, during endoscopy. One caveat is that effective medical therapy can change the appearance of the mucosa such that either skip areas or the entire colon can be microscopically normal.

With mild inflammation, the mucosa is erythematous and has a fine granular surface that resembles sandpaper. In more severe disease, the mucosa is hemorrhagic, edematous, and ulcerated **(Fig. 337-3)**. In long-standing disease, inflammatory polyps (pseudopolyps) may be present as a result of epithelial regeneration. The mucosa may appear normal in remission, but in patients with many years of disease, it appears atrophic and featureless, and the entire colon becomes narrowed and shortened. Patients with fulminant disease can develop a toxic colitis or megacolon where the bowel wall becomes thin and the mucosa is severely ulcerated; this may lead to perforation.

ULCERATIVE COLITIS: MICROSCOPIC FEATURES

Histologic findings correlate well with the endoscopic appearance and clinical course of UC. The process is limited to the mucosa and

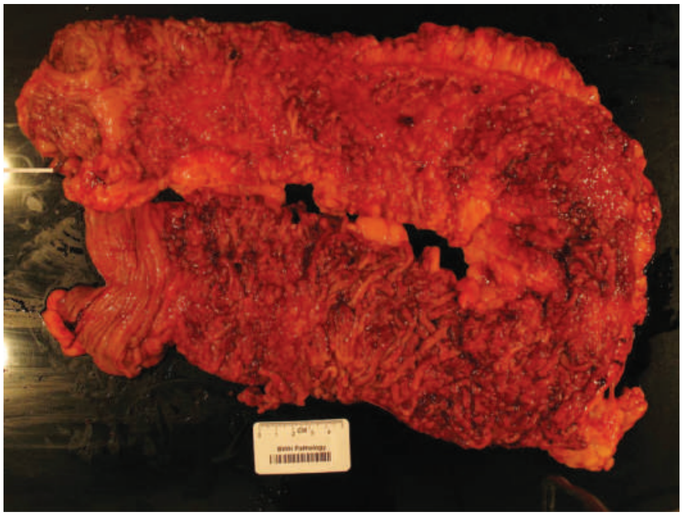

FIGURE 337-3 Ulcerative colitis. Diffuse (nonsegmental) mucosal disease, with broad areas of ulceration. The bowel wall is not thickened, and there is no cobblestoning. (*Courtesy of Dr. R. Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.*)

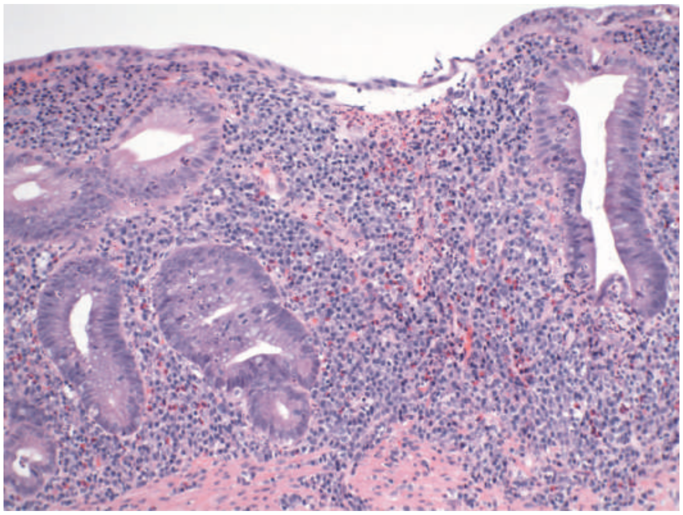

FIGURE 337-4 Medium-power view of colonic mucosa in ulcerative colitis showing diffuse mixed inflammation, basal lymphoplasmacytosis, crypt atrophy and irregularity, and superficial erosion. These features are typical of chronic active ulcerative colitis. (*Courtesy of Dr. R. Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.*)

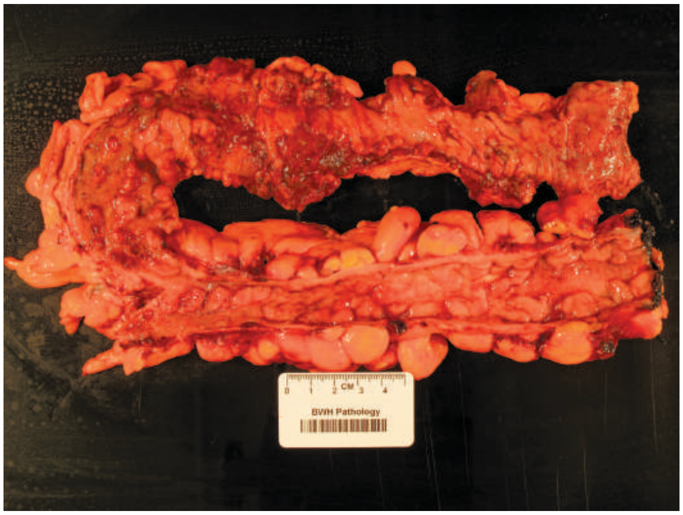

FIGURE 337-5 Crohn's disease of the colon showing thickening of the wall, with stenosis, linear serpiginous ulcers, and cobblestoning of the mucosa. (*Courtesy of Dr. R. Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.*)

superficial submucosa, with deeper layers unaffected except in fulminant disease. In UC, two major histologic features suggest chronicity and help distinguish it from infectious or acute self-limited colitis. First, the crypt architecture of the colon is distorted; crypts may be bifid and reduced in number, often with a gap between the crypt bases and the muscularis mucosae. Second, some patients have basal plasma cells and multiple basal lymphoid aggregates. Mucosal vascular congestion, with edema and focal hemorrhage, and an inflammatory cell infiltrate of neutrophils, lymphocytes, plasma cells, and macrophages may be present. The neutrophils invade the epithelium, usually in the crypts, giving rise to cryptitis and, ultimately, to crypt abscesses **(Fig. 337-4)**. Ileal changes in patients with backwash ileitis include villous atrophy and crypt regeneration with increased inflammation, increased neutrophil and mononuclear inflammation in the lamina propria, and patchy cryptitis and crypt abscesses.

CROHN'S DISEASE: MACROSCOPIC FEATURES

CD can affect any part of the gastrointestinal (GI) tract from the mouth to the anus. Some 30–40% of patients have small-bowel disease alone, 40–55% have disease involving both the small and large intestines, and 15–25% have colitis alone. In the 75% of patients with small-intestinal disease, the terminal ileum is involved in 90%. Unlike UC, which almost always involves the rectum, the rectum is often spared in CD. CD is often segmental with skip areas throughout the diseased intestine **(Fig. 337-5)**. Perianal disease, manifesting as perirectal fistulas, fissures, abscesses, and anal stenosis, is present in one-third of patients with CD, particularly those with colonic involvement. Rarely, CD may also involve the liver and the pancreas.

Unlike UC, CD is a transmural process. Endoscopically, aphthous or small superficial ulcerations characterize mild disease; in more active disease, stellate ulcerations fuse longitudinally and transversely to demarcate islands of mucosa that frequently are histologically normal. This "cobblestone" appearance is characteristic of CD, both endoscopically and by barium radiography. As in UC, pseudopolyps can form in CD.

Active CD is characterized by focal inflammation and formation of fistula tracts, which resolve by fibrosis and stricturing of the bowel. The bowel wall thickens and becomes narrowed and fibrotic, leading to chronic, recurrent bowel obstructions. Projections of thickened mesentery known as "creeping fat" encase the bowel, and serosal and mesenteric inflammation promotes adhesions and fistula formation.

CROHN'S DISEASE: MICROSCOPIC FEATURES

The earliest lesions are aphthoid ulcerations and focal crypt abscesses with loose aggregations of macrophages, which form noncaseating granulomas in all layers of the bowel wall **(Fig. 337-6)**. Granulomas are a characteristic feature of CD and are less commonly found on mucosal biopsies than on surgical resection specimens. Other histologic features of CD include submucosal or subserosal lymphoid aggregates, particularly away from areas of ulceration, gross and microscopic skip areas, and transmural inflammation that is accompanied by fissures that penetrate deeply into the bowel wall and sometimes form fistulous tracts or local abscesses.

CLINICAL PRESENTATION

ULCERATIVE COLITIS

Signs and Symptoms The major symptoms of UC are diarrhea, rectal bleeding, tenesmus, passage of mucus, and crampy abdominal pain. The severity of symptoms correlates with the extent of disease. Although UC can present acutely, symptoms usually have been present for weeks to months.

Patients with proctitis usually pass fresh blood or blood-stained mucus, either mixed with stool or streaked onto the surface of a normal or hard stool. They also have tenesmus, or urgency with a feeling of incomplete evacuation, but rarely have abdominal pain. With proctitis

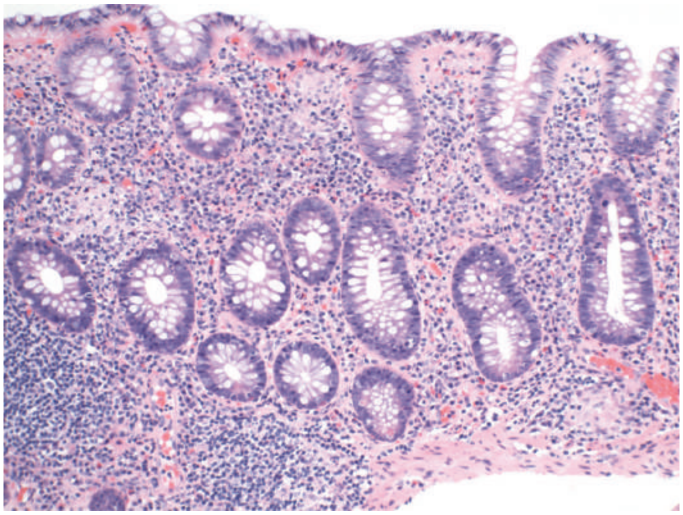

FIGURE 337-6 Medium-power view of Crohn's colitis showing mixed acute and chronic inflammation, crypt atrophy, and multiple small epithelioid granulomas in the mucosa. (*Courtesy of Dr. R. Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.*)

TABLE 337-4 Montreal Classification of Extent and Severity of Ulcerative Colitis (UC)

EXTENT	ANATOMY
E1: Ulcerative proctitis	Involvement limited to the rectum
E2: Left-sided UC (distal UC)	Involvement limited to the colorectum distal to the splenic flexure
E3: Extensive UC (pancolitis)	Involvement extends proximal to the splenic flexure
SEVERITY	**DEFINITION**
S0: Clinical remission	Absence of symptoms
S1: Mild disease activity	≤4 stools/d (with or without blood), absence of systemic illness, normal inflammatory markers (ESR)
S2: Moderate disease activity	≥4 stools/d but minimal signs of systemic toxicity
S3: Severe disease activity	≥6 bloody stools/d, pulse ≥90 beats/min, temperature ≥37.5°C, hemoglobin <10.5 g/100 mL, and ESR ≥30 mm/h

Abbreviation: ESR, erythrocyte sedimentation rate.

Source: C Gasche et al: A simple classification of Crohn's disease: Report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. Inflamm Bowel Dis 6:8, 2000; and J Satsangi et al: The Montreal classification of inflammatory bowel disease: Controversies, consensus, and implications. Gut 55:749, 2006.

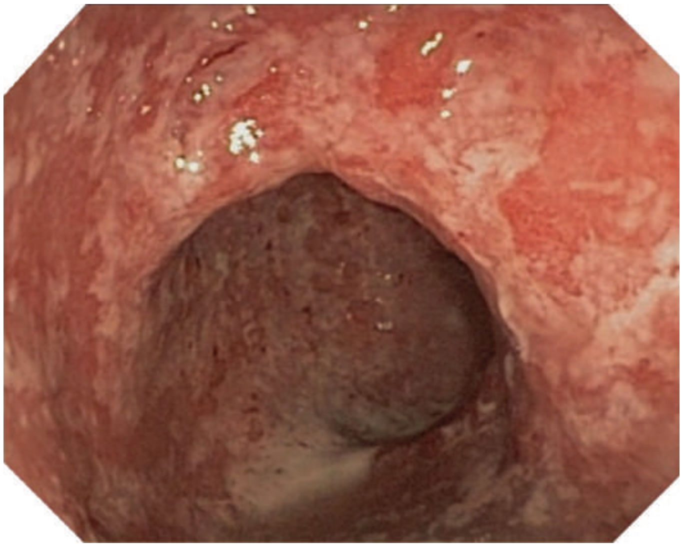

FIGURE 337-7 Colonoscopy with acute ulcerative colitis: severe colon inflammation with erythema, friability, and exudates. (*Courtesy of Dr. M. Hamilton, Gastroenterology Division, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts; with permission.*)

or proctosigmoiditis, proximal transit slows, which may account for the constipation commonly seen in patients with distal disease.

When the disease extends beyond the rectum, blood is usually mixed with stool or grossly bloody diarrhea may be noted. Colonic motility is altered by inflammation with rapid transit through the inflamed intestine. When the disease is severe, patients pass a liquid stool containing blood, pus, and fecal matter. Diarrhea is often nocturnal and/or postprandial. Although severe pain is not a prominent symptom, some patients with active disease may experience lower abdominal discomfort or mild central abdominal cramping. Severe cramping and abdominal pain can occur with severe attacks of the disease. Other symptoms in moderate to severe disease include anorexia, nausea, vomiting, fever, and weight loss.

Physical signs of proctitis include a tender anal canal and blood on rectal examination. With more extensive disease, patients have tenderness to palpation directly over the colon. Patients with a toxic colitis have severe pain and bleeding, and those with megacolon have hepatic tympany. Both may have signs of peritonitis if a perforation has occurred. The classification of disease activity is shown in **Table 337-4**.

Laboratory, Endoscopic, and Radiographic Features Active disease can be associated with a rise in acute-phase reactants (C-reactive protein [CRP]), platelet count, and erythrocyte sedimentation rate (ESR) and a decrease in hemoglobin. Fecal lactoferrin, a glycoprotein present in activated neutrophils, is a highly sensitive and specific marker for detecting intestinal inflammation. Fecal calprotectin is present in neutrophils and monocytes, and the levels correlate well with histologic inflammation, predict relapses, and detect pouchitis. Both fecal lactoferrin and calprotectin are an integral part of IBD management and are used frequently to rule out active inflammation versus symptoms of irritable bowel or bacterial overgrowth. In severely ill patients, the serum albumin level will fall rather quickly. Leukocytosis may be present but is not a specific indicator of disease activity. Diagnosis relies on the patient's history, clinical symptoms, negative stool and/or tissue examination for bacteria, *C. difficile* toxin, ova and parasites, and viruses depending on epidemiologic considerations and clinical presentation; sigmoidoscopic appearance **(see Fig. 333-4*A*)**; and histology of rectal or colonic biopsy specimens.

Sigmoidoscopy is used to assess disease activity and is usually performed before treatment. If the patient is not having an acute flare, colonoscopy is used to assess disease extent and activity **(Fig. 337-7)**. Endoscopically mild disease is characterized by erythema, decreased vascular pattern, and mild friability. Moderate disease is characterized by marked erythema, absent vascular pattern, friability, and erosions, and severe disease is characterized by spontaneous bleeding and ulcerations. Histologic features change more slowly than clinical features but can also be used to grade disease activity. Intestinal ultrasound is a newer tool to assess UC activity and mucosal healing and correlates well with endoscopy and inflammatory markers.

Complications Only 15% of patients with UC present initially with severe disease. Massive hemorrhage occurs in 1% of patients, and treatment for the disease usually stops the bleeding. *Toxic megacolon* is defined as a transverse or right colon with a diameter of >6 cm, with loss of haustration in patients with severe attacks of UC. It occurs rarely and can be triggered by electrolyte abnormalities and narcotics. About 50% of acute dilations will resolve with conservative management alone, but urgent colectomy is required for those who do not improve. Perforation is the most dangerous of the local complications, and the physical signs of peritonitis may not be obvious, especially if the patient is receiving glucocorticoids. Although perforation is rare, the mortality rate for perforation complicating a toxic megacolon is ~15%. In addition, patients can develop a toxic colitis and such severe ulcerations that the bowel may perforate without first dilating.

Strictures occur in ~5% of patients and are always a concern in UC because of the possibility of underlying neoplasia. Although benign strictures can form from the inflammation and fibrosis of UC, strictures that are impassable with the colonoscope should be presumed malignant until proven otherwise. A stricture that prevents passage of the colonoscope is an indication for surgery. UC patients occasionally develop anal fissures, perianal abscesses, or hemorrhoids, but the occurrence of extensive perianal lesions should suggest CD.

■ CROHN'S DISEASE

Signs and Symptoms Although CD usually presents as acute or chronic bowel inflammation, the inflammatory process evolves toward one of two patterns of disease: a fibrostenotic obstructing pattern or a penetrating fistulous pattern, each with different treatments and prognoses. The site of disease influences the clinical manifestations **(Table 337-5)**.

ILEOCOLITIS Because the most common site of inflammation is the terminal ileum, the usual presentation of ileocolitis is a chronic history of recurrent episodes of right lower quadrant pain and diarrhea. Sometimes the initial presentation mimics acute appendicitis with pronounced right lower quadrant pain, a palpable mass, fever, and leukocytosis. Pain is usually colicky; it precedes and is relieved by defecation. A low-grade fever is usually noted. High-spiking fever suggests intraabdominal abscess formation. Weight loss is common—typically

TABLE 337-5 Vienna and Montreal Classifications of Crohn's Disease

	VIENNA	MONTREAL
Age at diagnosis	A1: <40 years A2: >40 years	A1: <16 years A2: Between 17 and 40 years A3: >40 years
Location	L1: Ileal L2: Colonic L3: Ileocolonic L4: Upper	L1: Ileal L2: Colonic L3: Ileocolonic L4: Isolated upper disease[a]
Behavior	B1: Nonstricturing, nonpenetrating B2: Stricturing B3: Penetrating	B1: Nonstricturing, nonpenetrating B2: Stricturing B3: Penetrating p: Perianal disease modifier[b]

[a] L4 is a modifier and can be added to L1–L3 when there is concomitant foregut disease. [b] p is added to B1–B3 when there is concomitant perianal disease.

10–20% of body weight—and develops as a consequence of diarrhea, anorexia, and fear of eating.

An inflammatory mass may be palpated in the right lower quadrant of the abdomen. The mass is composed of inflamed bowel, induration of the mesentery, and enlarged abdominal lymph nodes. The "string sign" on radiographic studies results from a severely narrowed loop of bowel, which makes the lumen resemble a frayed cotton string. It is caused by incomplete filling of the lumen as the result of edema, irritability, and spasms associated with inflammation and ulcerations. The sign may be seen in both nonstenotic and stenotic phases of the disease.

Bowel obstruction may take several forms. In the early stages of disease, bowel wall edema and spasm produce intermittent obstructive manifestations and increasing symptoms of postprandial pain. Over several years, persistent inflammation gradually progresses to fibrostenotic narrowing and stricture. Diarrhea will decrease and be replaced by chronic bowel obstruction. Acute episodes of obstruction occur as well, precipitated by bowel inflammation and spasm or sometimes by impaction of undigested food or medication. These episodes usually resolve with intravenous fluids and gastric decompression.

Severe inflammation of the ileocecal region may lead to localized wall thinning, with microperforation and fistula formation to the adjacent bowel, the skin, or the urinary bladder, or to an abscess cavity in the mesentery. Enterovesical fistulas typically present as dysuria or recurrent bladder infections or, less commonly, as pneumaturia or fecaluria. Enterocutaneous fistulas follow tissue planes of least resistance, usually draining through abdominal surgical scars. Enterovaginal fistulas are rare and present as dyspareunia or as a feculent or foul-smelling, often painful vaginal discharge. They are unlikely to develop without a prior hysterectomy.

JEJUNOILEITIS Extensive inflammatory disease is associated with a loss of digestive and absorptive surface, resulting in malabsorption and steatorrhea. Nutritional deficiencies can also result from poor intake and enteric losses of protein and other nutrients. Intestinal malabsorption can cause anemia, hypoalbuminemia, hypocalcemia, hypomagnesemia, coagulopathy, and hyperoxaluria with nephrolithiasis in patients with an intact colon. Many patients need to take intravenous iron since oral iron is poorly tolerated and often ineffective. Vertebral fractures are caused by a combination of vitamin D deficiency, hypocalcemia, and prolonged glucocorticoid use. Pellagra from niacin deficiency can occur in extensive small-bowel disease, and malabsorption of vitamin B_{12} can lead to megaloblastic anemia and neurologic symptoms. Other important nutrients to measure and replete if low are folate and vitamins A, E, and K. Levels of minerals such as zinc, selenium, copper, and magnesium are often low in patients with extensive small-bowel inflammation or resections, and these should be repleted as well. Most patients should take daily multivitamin, calcium, and vitamin D supplements.

Diarrhea is characteristic of active disease; its causes include (1) bacterial overgrowth in obstructive stasis or fistulization, (2) bile acid malabsorption due to a diseased or resected terminal ileum, (3) intestinal inflammation with decreased water absorption and increased secretion of electrolytes, and (4) enteroenteric fistula(e).

COLITIS AND PERIANAL DISEASE Patients with colitis present with low-grade fevers, malaise, diarrhea, crampy abdominal pain, and sometimes hematochezia. Gross bleeding is not as common as in UC and appears in about one-half of patients with exclusively colonic disease. Only 1–2% exhibit massive bleeding. Pain is caused by passage of fecal material through narrowed and inflamed segments of the large bowel. Decreased rectal compliance is another cause for diarrhea in Crohn's colitis patients.

Stricturing can occur in the colon in 4–16% of patients and produce symptoms of bowel obstruction. If the endoscopist is unable to traverse a stricture in Crohn's colitis, surgical resection should be considered, especially if the patient has symptoms of chronic obstruction. Colonic disease may fistulize into the stomach or duodenum, causing feculent vomiting, or to the proximal or mid-small bowel, causing malabsorption by "short circuiting" the absorptive surface and bacterial overgrowth. Ten percent of women with Crohn's colitis will develop a rectovaginal fistula.

Perianal disease affects about one-third of patients with Crohn's colitis and is manifested by incontinence, large hemorrhoidal tags, anal strictures, anorectal fistulas, and perirectal abscesses. Not all patients with perianal fistula will have endoscopic evidence of colonic inflammation.

GASTRODUODENAL DISEASE Symptoms and signs of upper GI tract disease include nausea, vomiting, and epigastric pain. Patients usually have a *Helicobacter pylori*–negative gastritis. The second portion of the duodenum is more commonly involved than the bulb. Fistulas involving the stomach or duodenum arise from the small or large bowel and do not necessarily signify the presence of upper GI tract involvement. Patients with advanced gastroduodenal CD may develop a chronic gastric outlet obstruction. About 30% of children diagnosed with CD have esophagogastroduodenal involvement. The classification of disease activity is shown in Table 337-5.

Laboratory, Endoscopic, and Radiographic Features Laboratory abnormalities include elevated ESR and CRP. In more severe disease, findings include hypoalbuminemia, anemia, and leukocytosis. Fecal calprotectin and lactoferrin levels have been used to distinguish IBD from irritable bowel syndrome (IBS), to assess whether CD is active, and to detect postoperative recurrence of CD. Fecal calprotectin is a more sensitive marker of ileocolonic or colonic inflammation rather than isolated ileal inflammation.

Endoscopic features of CD include rectal sparing, aphthous ulcerations, fistulas, and skip lesions. Colonoscopy allows examination and biopsy of mass lesions or strictures and biopsy of the terminal ileum. Upper endoscopy is useful in diagnosing gastroduodenal involvement in patients with upper tract symptoms. An ileal or colonic stricture may be dilated with a balloon introduced through the colonoscope. Strictures ≤4 cm in length and those at anastomotic sites respond better to endoscopic dilation. The perforation rate is as high as 10%. Most endoscopists dilate only fibrotic strictures and not those associated with active inflammation. Wireless capsule endoscopy (WCE) allows direct visualization of the entire small-bowel mucosa **(Fig. 337-8)**. The diagnostic yield of detecting lesions suggestive of active CD is higher with WCE than computed tomography (CT) or magnetic resonance (MR) enterography. WCE should not be used in the setting of a small-bowel stricture. Capsule retention occurs in <1% of patients with suspected CD, but retention rates of 4–6% are seen in patients with established CD. It is helpful to give the patient with CD a patency capsule, which is made of barium and starts to dissolve 30 h after ingestion. An abdominal x-ray can be taken at around 30 h after ingestion to see if the capsule is still present in the small bowel, which would indicate a stricture.

In CD, early radiographic findings in the small bowel include thickened folds and aphthous ulcerations. "Cobblestoning" from longitudinal and transverse ulcerations most frequently involves the small bowel. In more advanced disease, strictures, fistulas, inflammatory masses, and abscesses may be detected. The earliest macroscopic findings of colonic CD are aphthous ulcers. These small ulcers are often

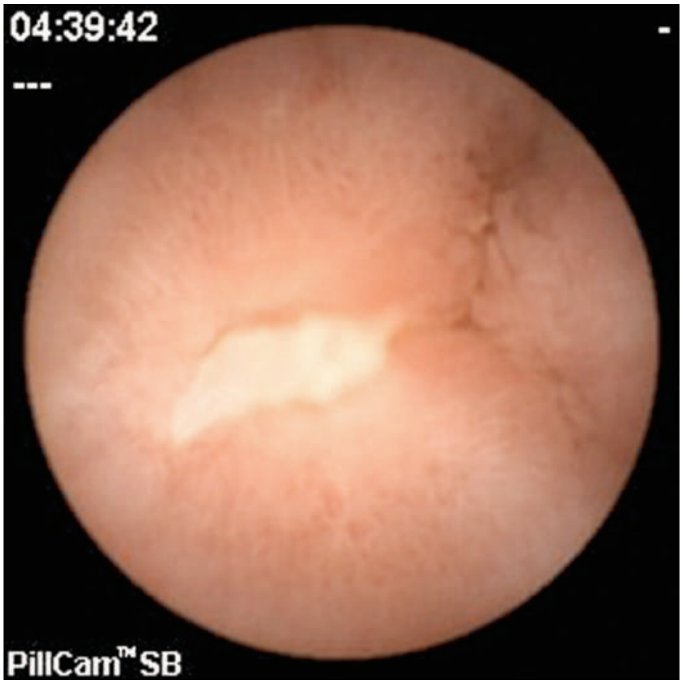

FIGURE 337-8 Wireless capsule endoscopy image in a patient with Crohn's disease of the ileum shows ulcerations and narrowing of the intestinal lumen. (*Courtesy of Dr. S. Reddy, Gastroenterology Division, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts; with permission.*)

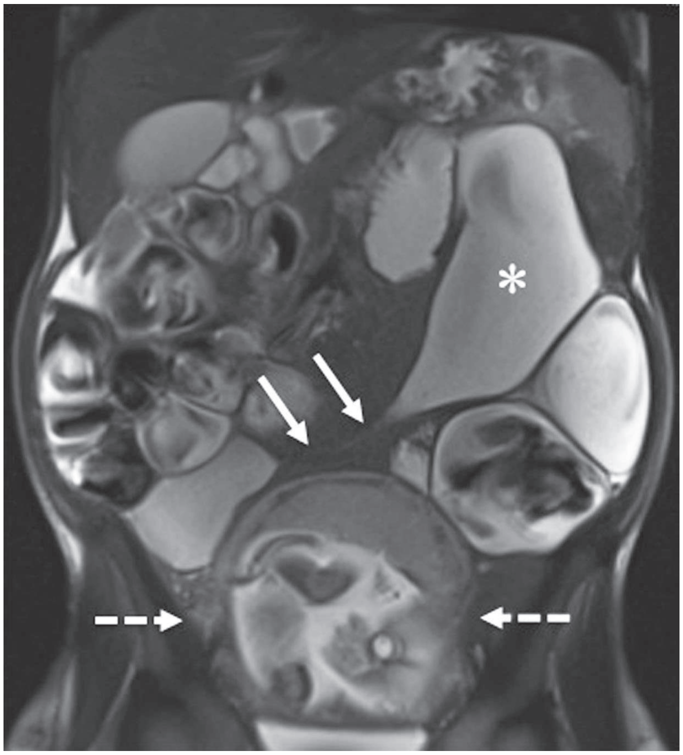

FIGURE 337-9 A coronal magnetic resonance image was obtained using a half Fourier single-shot T2-weighted acquisition with fat saturation in a 27-year-old pregnant (23 weeks' gestation) woman. The patient had Crohn's disease and was maintained on mercaptopurine and prednisone. She presented with abdominal pain, distension, vomiting, and small-bowel obstruction. The image reveals a 7- to 10-cm long stricture at the terminal ileum (*white arrows*) causing obstruction and significant dilatation of the proximal small bowel (*white asterisk*). A fetus is seen in the uterus (*dashed white arrows*). (*Courtesy of Drs. J. F. B. Chick and P. B. Shyn, Abdominal Imaging and Intervention, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; with permission.*)

multiple and separated by normal intervening mucosa. As the disease progresses, aphthous ulcers become enlarged, deeper, and occasionally connected to one another, forming longitudinal stellate, serpiginous, and linear ulcers **(see Fig. 333-4*B*)**.

The transmural inflammation of CD leads to decreased luminal diameter and limited distensibility. As ulcers progress deeper, they can lead to fistula formation. The segmental nature of CD results in wide gaps of normal or dilated bowel between involved segments.

CT enterography and MR enterography have been shown to be equally accurate in the identification of active small-bowel inflammation. MRI is thought to offer superior soft tissue contrast and has the added advantage of avoiding radiation exposure changes **(Figs. 337-9** and **337-10)**. The lack of ionizing radiation is particularly appealing in younger patients and when monitoring response to therapy where serial images will be obtained. Pelvic MRI is superior to pelvic CT for demonstrating pelvic lesions such as ischiorectal abscesses and perianal fistulas **(Fig. 337-11)**. An underutilized resource for assessing small-bowel CD is small-bowel ultrasound (SBUS). SBUS is at least as sensitive as MR enterography and CT enterography for detecting small-bowel CD, with a sensitivity of 94%, specificity of 97%, positive predictive value of 97%, and negative predictive value of 94%. Use of oral contrast medium can increase the sensitivity and specificity to detect small-bowel lesions to 100%. SBUS is best suited for distal small-bowel assessment, as the sensitivity of detecting lesions within the duodenum and proximal jejunum may be lower due to anatomic position. The limitations of SBUS include availability and operator dependence.

Complications Because CD is a transmural process, serosal adhesions develop that provide direct pathways for fistula formation and reduce the incidence of free perforation. Perforation occurs in 1–2% of patients, usually in the ileum but occasionally in the jejunum or as a complication of toxic megacolon. The peritonitis of free perforation, especially colonic, may be fatal. Intraabdominal and pelvic abscesses occur in 10–30% of patients with CD at some time in the course of their illness. CT-guided percutaneous drainage of the abscess is standard therapy. Despite adequate drainage, most patients need resection of the offending bowel segment. Percutaneous drainage has an especially high failure rate in abdominal wall abscesses. Systemic glucocorticoid therapy increases the risk of intraabdominal and pelvic abscesses in CD patients who have never had an operation. Other complications include intestinal obstruction in 40%, massive hemorrhage, malabsorption, and severe perianal disease.

Serologic Markers Clinical factors described at diagnosis are more helpful than serologies at predicting the natural history of IBD. The positive and negative predictive values of serologic tests such as anti-*Saccharomyces cerevisiae* antibody (ASCA) and perinuclear antineutrophil cytoplasmic antibody (pANCA) vary depending on the prevalence of IBD in different populations, such that their clinical usefulness is unclear.

DIFFERENTIAL DIAGNOSIS OF UC AND CD

Once a diagnosis of IBD is made, distinguishing between UC and CD is impossible initially in up to 15% of cases. These are termed *indeterminate colitis.* Fortunately, in most cases, the true nature of the underlying colitis becomes evident later in the course of the patient's disease. Approximately 5% of colon resection specimens are difficult to classify as either UC or CD because they exhibit overlapping histologic features.

■ INFECTIOUS DISEASES

Infections of the small intestines and colon can mimic CD or UC. They may be bacterial, fungal, viral, or protozoal in origin **(Table 337-6)**. *Campylobacter* colitis can mimic the endoscopic appearance of severe UC and can cause a relapse of established UC. *Salmonella* can cause watery or bloody diarrhea, nausea, and vomiting. Shigellosis causes watery diarrhea, abdominal pain, and fever followed by rectal tenesmus and by the passage of blood and mucus per rectum. All three are usually self-limited, but 1% of patients infected with *Salmonella* become asymptomatic carriers. *Yersinia enterocolitica* infection occurs mainly in the terminal ileum and causes mucosal ulceration, neutrophil

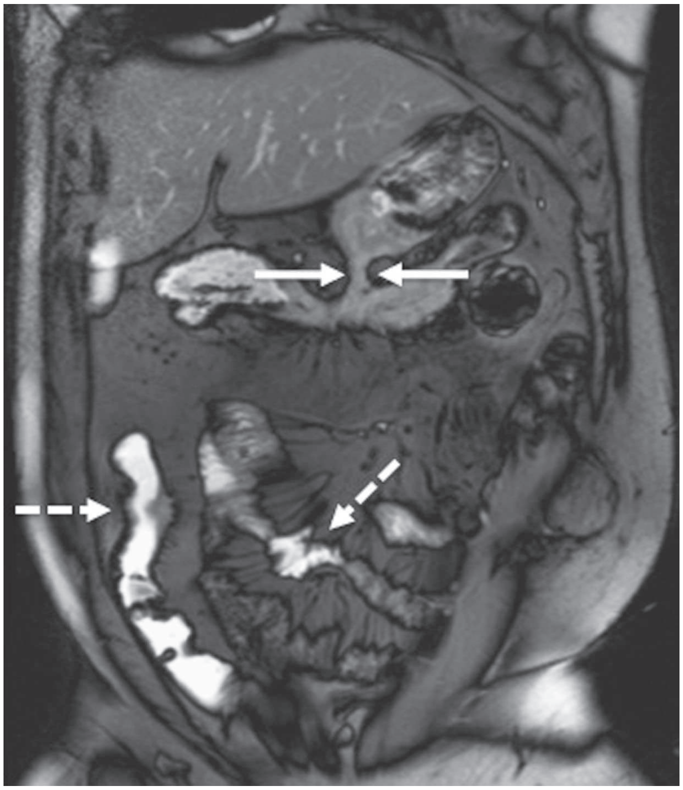

FIGURE 337-10 A coronal balanced, steady-state, free precession, T2-weighted image with fat saturation was obtained in a 32-year-old man with Crohn's disease and prior episodes of bowel obstruction, fistulas, and abscesses. He was being treated with mercaptopurine and presented with abdominal distention and diarrhea. The image demonstrates a new gastrocolic fistula (*solid white arrows*). Multifocal involvement of the small bowel and terminal ileum is also present (*dashed white arrows*). (*Courtesy of Drs. J. F. B. Chick and P. B. Shyn, Abdominal Imaging and Intervention, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; with permission.*)

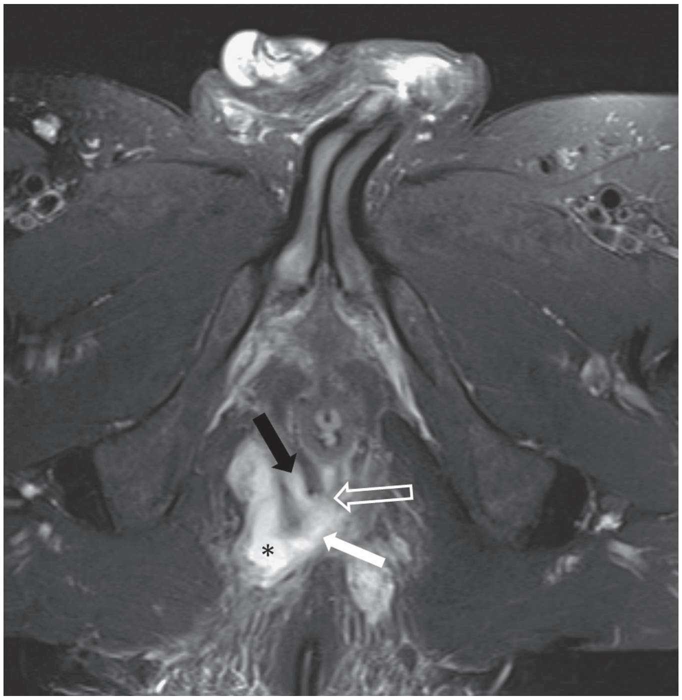

FIGURE 337-11 Axial T2-weighted fat-saturated image obtained in a 39-year-old male with Crohn's disease shows a defect in the internal sphincter at the 6 o'clock position of the mid anal canal (*open white arrow*) communicating with a 1.1-cm intersphincteric collection (*black arrow*). Wide defect in the external sphincter at the 7 o'clock position (*solid white arrow*) leads to a moderate-sized perianal abscess in the ischioanal fossa (*asterisk*). (*Courtesy of Drs. J.S. Quon and P.B. Shyn, Abdominal Imaging and Intervention, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; with permission.*)

TABLE 337-6 Diseases That Mimic IBD

Infectious Etiologies		
Bacterial	**Mycobacterial**	**Viral**
Salmonella	Tuberculosis	Cytomegalovirus
Shigella	*Mycobacterium avium*	Herpes simplex
Toxigenic	**Parasitic**	HIV
Escherichia coli	Amebiasis	**Fungal**
Campylobacter	*Isospora*	Histoplasmosis
Yersinia	*Trichuris trichiura*	*Candida*
Clostridioides *difficile*	Hookworm	*Aspergillus*
Gonorrhea	*Strongyloides*	
Chlamydia trachomatis		
Noninfectious Etiologies		
Inflammatory	**Neoplastic**	**Drugs and Chemicals**
Appendicitis	Lymphoma	NSAIDs
Diverticulitis	Metastatic	Phosphosoda
Diversion colitis	Carcinoma	Cathartic colon
Collagenous/lymphocytic colitis	Carcinoma of the ileum	Gold
Ischemic colitis	Carcinoid	Oral contraceptives
Radiation colitis/enteritis	Familial polyposis	Cocaine
Solitary rectal ulcer syndrome		Immune checkpoint inhibitor colitis
Eosinophilic gastroenteritis		Mycophenolate mofetil
Neutropenic colitis		
Behçet's syndrome		
Graft-versus-host disease		

Abbreviations: IBD, inflammatory bowel disease; NSAIDs, nonsteroidal anti-inflammatory drugs.

invasion, and thickening of the ileal wall. Other bacterial infections that may mimic IBD include *C. difficile*, which presents with watery diarrhea, tenesmus, nausea, and vomiting; and *E. coli*, three categories of which can cause colitis. These are enterohemorrhagic, enteroinvasive, and enteroadherent *E. coli*, all of which can cause bloody diarrhea and abdominal tenderness. Gonorrhea, *Chlamydia*, and syphilis can also cause proctitis.

GI involvement with mycobacterial infection occurs primarily in the immunosuppressed patient but may occur in patients with normal immunity. Distal ileal and cecal involvement predominates, and patients present with symptoms of small-bowel obstruction and a tender abdominal mass. The diagnosis is made most directly by colonoscopy with biopsy and culture. Although most of the patients with viral colitis are immunosuppressed, cytomegalovirus (CMV) and herpes simplex proctitis may occur in immunocompetent individuals. CMV occurs most commonly in the esophagus, colon, and rectum but may also involve the small intestine. Symptoms include abdominal pain, bloody diarrhea, fever, and weight loss. With severe disease, necrosis and perforation can occur. Diagnosis is made by identification of characteristic intranuclear inclusions in mucosal cells on biopsy. Herpes simplex infection of the GI tract is limited to the oropharynx, anorectum, and perianal areas. Symptoms include anorectal pain, tenesmus, constipation, inguinal adenopathy, difficulty with urinary voiding, and sacral paresthesias. Diagnosis is made by rectal biopsy with identification of characteristic cellular inclusions and viral culture. HIV itself can cause diarrhea, nausea, vomiting, and anorexia. Small-intestinal biopsies show partial villous atrophy; small-bowel bacterial overgrowth and fat malabsorption may also be noted.

Protozoan parasites include *Isospora belli*, which can cause a self-limited infection in healthy hosts but causes a chronic profuse, watery diarrhea and weight loss in AIDS patients. *Entamoeba histolytica* or related species infect ~10% of the world's population; symptoms include abdominal pain, tenesmus, frequent loose stools containing blood and mucus, and abdominal tenderness. Colonoscopy reveals focal punctate ulcers with normal intervening mucosa; diagnosis is made by biopsy or serum amebic antibodies. Fulminant amebic colitis is rare but has a mortality rate of >50%.

Other parasitic infections that may mimic IBD include hookworm (*Necator americanus*), whipworm (*Trichuris trichiura*), and *Strongyloides stercoralis*. In severely immunocompromised patients, *Candida* or *Aspergillus* can be identified in the submucosa. Disseminated histoplasmosis can involve the ileocecal area.

NONINFECTIOUS DISEASES

Diverticulitis can be confused with CD clinically and radiographically. Both diseases cause fever, abdominal pain, tender abdominal mass, leukocytosis, elevated ESR, partial obstruction, and fistulas. Perianal disease or ileitis on small-bowel series favors the diagnosis of CD. Significant endoscopic mucosal abnormalities are more likely in CD than in diverticulitis. Endoscopic or clinical recurrence following segmental resection favors CD. Diverticular-associated colitis is similar to CD, but mucosal abnormalities are limited to the sigmoid and descending colon.

Ischemic colitis is commonly confused with IBD. The ischemic process can be chronic and diffuse, as in UC, or segmental, as in CD. Colonic inflammation due to ischemia may resolve quickly or may persist and result in transmural scarring and stricture formation. Ischemic bowel disease should be considered in the elderly following abdominal aortic aneurysm repair or when a patient has a hypercoagulable state or a severe cardiac or peripheral vascular disorder. Patients usually present with sudden onset of left lower quadrant pain, urgency to defecate, and the passage of bright red blood per rectum. Endoscopic examination often demonstrates a normal-appearing rectum and a sharp transition to an area of inflammation in the descending colon and splenic flexure.

The effects of radiotherapy on the GI tract can be difficult to distinguish from IBD. Acute symptoms can occur within 1–2 weeks of starting radiotherapy. When the rectum and sigmoid are irradiated, patients develop bloody, mucoid diarrhea and tenesmus, as in distal UC. With small-bowel involvement, diarrhea is common. Late symptoms include malabsorption and weight loss. Stricturing with obstruction and bacterial overgrowth may occur. Fistulas can penetrate the bladder, vagina, or abdominal wall. Flexible sigmoidoscopy reveals mucosal granularity, friability, numerous telangiectasias, and occasionally discrete ulcerations. Biopsy can be diagnostic.

Solitary rectal ulcer syndrome is uncommon and can be confused with IBD. It occurs in persons of all ages and may be caused by impaired evacuation and failure of relaxation of the puborectalis muscle. Single or multiple ulcerations may arise from anal sphincter overactivity, higher intrarectal pressures during defecation, and digital removal of stool. Patients complain of constipation with straining and pass blood and mucus per rectum. Other symptoms include abdominal pain, diarrhea, tenesmus, and perineal pain. Ulceration, which may be as large as 5 cm in diameter, is usually observed anteriorly or anterolaterally 3–15 cm from the anal verge. Biopsies can be diagnostic.

Several types of colitis are associated with nonsteroidal anti-inflammatory drugs (NSAIDs), including de novo colitis, reactivation of IBD, and proctitis caused by use of suppositories. Most patients with NSAID-related colitis present with diarrhea and abdominal pain, and complications include stricture, bleeding, obstruction, perforation, and fistulization. Withdrawal of these agents is crucial, and in cases of reactivated IBD, standard therapies are indicated.

Colitis secondary to immune checkpoint inhibitors (ICIs), termed ICI-related colitis, has emerged as these agents have found use in a wide variety of cancers. Immune checkpoint proteins such as CTLA-4 and programmed cell death protein 1 (PD-1) are receptors expressed on the surface of effector T cells that interact with their ligands CD80/CD86 (CTLA-4) and programmed death ligand 1 (PD-L1) on antigen-presenting cells and normally function as inhibitors of immune responses. ICIs block these inhibitory pathways and promote the activation and proliferation of the native adaptive T-cell response against malignant cells as their mechanism of antitumor activity. While very effective at enhancing antitumor T-cell activity, ICIs also activate global T-cell responses that induce several autoimmune-related adverse events. Although immune-related adverse events of ICIs occur in multiple organ systems, the GI tract is affected in 21–44% of patients. The most common clinical presentation is self-limited diarrhea that can be associated with frank colitis and can lead to significant morbidity and mortality if not managed appropriately. Treatment is generally based on symptom severity. Moderate to severe symptoms usually require glucocorticoids, whereas biologics such as anti-TNF agents and integrin inhibitors are used in steroid-refractory cases.

THE ATYPICAL COLITIDES

Two atypical colitides—collagenous colitis and lymphocytic colitis—have normal endoscopic appearances. Collagenous colitis has two main histologic components: increased subepithelial collagen deposition and colitis with increased intraepithelial lymphocytes. The female-to-male ratio is 9:1, and most patients present in the sixth or seventh decade of life. The main symptom is chronic watery diarrhea. Risk factors include smoking; use of NSAIDs, proton pump inhibitors, or beta blockers; and a history of autoimmune disease.

Lymphocytic colitis has features similar to collagenous colitis, including age at onset and clinical presentation, but it has almost equal incidence in men and women and no subepithelial collagen deposition on pathologic section. However, intraepithelial lymphocytes are increased. Use of sertraline (but not beta blockers) is an additional risk factor. The frequency of celiac disease is increased in lymphocytic colitis and ranges from 9 to 27%. Celiac disease should be excluded in all patients with lymphocytic colitis, particularly if diarrhea does not respond to conventional therapy. Treatments for both microscopic colitides vary depending on symptom severity and include, antidiarrheals (e.g., loperamide and diphenoxylate), bismuth, aminosalicylates, budesonide, systemic glucocorticoids, and biologics for refractory disease.

Diversion colitis is an inflammatory process that arises in segments of the large intestine that are not continuous with the fecal stream. It usually occurs in patients with ileostomy or colostomy when a mucus fistula or a Hartmann's pouch has been created. Clinically, patients have mucus or bloody discharge from the rectum. Erythema, granularity, friability, and, in more severe cases, ulceration can be seen on endoscopy. Histopathology shows areas of active inflammation with foci of cryptitis and crypt abscesses. Crypt architecture is normal, which differentiates it from UC but not necessarily CD. Short-chain fatty acid enemas may help in diversion colitis, but the definitive therapy is surgical reanastomosis.

EXTRAINTESTINAL MANIFESTATIONS

Up to one-third of IBD patients have at least one extraintestinal disease manifestation. Please see **Table 337-7** for a summary of IBD EIMs.

DERMATOLOGIC

Erythema nodosum (EN) occurs in up to 15% of CD patients and 10% of UC patients. Attacks usually correlate with bowel activity; skin lesions develop after the onset of bowel symptoms, and patients frequently have concomitant active peripheral arthritis. The lesions of EN are hot, red, tender nodules measuring 1–5 cm in diameter and are found on the anterior surface of the lower legs, ankles, calves, thighs, and arms. Therapy is directed toward the underlying bowel disease.

Pyoderma gangrenosum (PG) is seen in 1–12% of UC patients and less commonly in Crohn's colitis. Although it usually presents after the diagnosis of IBD, PG may occur years before the onset of bowel symptoms, run a course independent of the bowel disease, respond poorly to colectomy, and even develop years after proctocolectomy. It is usually associated with severe disease. Lesions are commonly found on the dorsal surface of the feet and legs but may occur on the arms, chest, stoma, and even the face. PG usually begins as a pustule and then spreads concentrically to rapidly undermine healthy skin. Lesions then ulcerate, with violaceous edges surrounded by a margin of erythema. Centrally, they contain necrotic tissue with blood and exudates. Lesions may be single or multiple and grow as large as 30 cm. They are sometimes very difficult to treat and often require IV antibiotics, IV glucocorticoids, dapsone, azathioprine, thalidomide, IV cyclosporine (CSA), infliximab, or adalimumab.

TABLE 337-7 Extraintestinal Manifestations

CATEGORY	CLINICAL COURSE	TREATMENT
Rheumatologic Disorders (5–20%)		
Peripheral arthritis	Asymmetric, migratory Parallels bowel activity	Reduce bowel inflammation
Sacroiliitis	Symmetric: spine and hip joints Independent of bowel activity	Steroids, injections, methotrexate, anti-TNF
Ankylosing spondylitis	Gradual fusion of spine Independent of bowel activity Two-thirds have HLA-B27 antigen	Physical therapy, steroids, injections, methotrexate, anti-TNF, IL-17 inhibitors, JAK inhibitors
Metabolic Bone Disorders (up to 40% of patients)		
Osteoporosis	Risk increased by glucocorticoids, cyclosporine, methotrexate, total parenteral nutrition, malabsorption, and inflammation Fracture rates highest in the elderly (age >60)	Screening with DEXA scan, check vitamin D levels, treat if osteoporosis or osteopenia on long-term corticosteroids
Osteonecrosis	Death of osteocytes and adipocytes and eventual bone collapse; affects hips more than knees or shoulders; risk factor is steroid use	Pain control, injections, joint replacement
Dermatologic Disorders (10–20%)		
Erythema nodosum	Hot, red, tender, nodules/extremities Parallels bowel activity	Reduce bowel inflammation
Pyoderma gangrenosum	Ulcerating, necrotic lesions on extremities, trunk, face, stoma Independent of bowel activity	Antibiotics, steroids, cyclosporine, infliximab, dapsone, azathioprine, intralesional steroids; not debridement or colectomy
Psoriasis	Unrelated to bowel activity	Topical steroids, light therapy, methotrexate, infliximab, adalimumab, ustekinumab, risankizumab, JAK inhibitors
Pyoderma vegetans	Intertriginous areas Parallels bowel activity	Evanescent; resolves without progression
Pyostomatitis vegetans	Mucous membranes Parallels bowel activity	Evanescent; resolves without progression
Metastatic Crohn's disease (CD)	CD of the skin Parallels bowel activity	Reduce bowel inflammation
Sweet syndrome	Neutrophilic dermatosis Parallels bowel activity	Reduce bowel inflammation
Aphthous stomatitis	Oral ulcerations Parallels bowel activity	Reduce bowel inflammation/topical therapy
Ocular Disorders (1–11%)		
Uveitis	Ocular pain, photophobia, blurred vision, headache Independent of bowel activity	Topical or systemic steroids
Episcleritis	Mild ocular burning Parallels bowel activity	Topical corticosteroids
Hepatobiliary Disorders (10–35%)		
Fatty liver	Secondary to chronic illness, malnutrition, steroid therapy	Improve nutrition, reduce steroids
Cholelithiasis	Patients with ileitis or ileal resection Malabsorption of bile acids, depletion of bile salt pool, secretion of lithogenic bile	Reduce bowel inflammation; cholecystectomy in symptomatic patients
Primary sclerosing cholangitis (PSC)	Intrahepatic and extrahepatic Inflammation and fibrosis leading to biliary cirrhosis and hepatic failure 7–10% cholangiocarcinoma Small-duct PSC involves small-caliber bile ducts and has a better prognosis	Cholecystectomy in patients with gallbladder polyps due to the high incidence of malignancy
Urologic		
Nephrolithiasis (10–20%)	CD patients following small-bowel resection; calcium oxalate stones most common	Low-oxalate diet; control of bowel inflammation; surgical intervention
Less Common Extraintestinal Manifestations		
Thromboembolic disorders	Increased risk of venous and arterial thrombosis; factors responsible include abnormalities of the platelet-endothelial interaction, hyperhomocysteinemia, alterations in the coagulation cascade, impaired fibrinolysis, involvement of tissue factor–bearing microvesicles, disruption of the normal coagulation system by autoantibodies, and a genetic predisposition	Anticoagulation; control of inflammation
Cardiopulmonary	Endocarditis, myocarditis, pleuropericarditis, interstitial lung disease	Treatment is varied; stop 5-ASA agents as they can rarely cause interstitial lung disease
Systemic amyloidosis	Secondary (reactive) in long-standing IBD, especially CD	Colchicine and referral to specialty center
Pancreatitis	Duodenal fistulas, ampullary CD, gallstones, PSC, drugs (MP, azathioprine, 5-ASAs), autoimmune, primary CD of the pancreas	Treatment is varied; stop offending medication; diagnose and treat with ERCP and/or cholecystectomy

Abbreviations: 5-ASA, 5-aminosalicylic acid; DEXA, dual-energy x-ray absorptiometry; ERCP, endoscopic retrograde cholangiopancreatography; IBD, inflammatory bowel disease; IL, interleukin; MP, mercaptopurine; TNF, tumor necrosis factor.

Other dermatologic manifestations include pyoderma vegetans, which occurs in intertriginous areas; pyostomatitis vegetans, which involves the mucous membranes; Sweet syndrome, a neutrophilic dermatosis; and metastatic CD, a rare disorder defined by cutaneous granuloma formation. Psoriasis affects 5–10% of patients with IBD and is unrelated to bowel activity, consistent with the potential shared immunogenetic basis of these diseases. Perianal skin tags are found in 75–80% of patients with CD, especially those with colon involvement. Oral mucosal lesions, seen often in CD and rarely in UC, include aphthous stomatitis and "cobblestone" lesions of the buccal mucosa.

■ RHEUMATOLOGIC

Peripheral arthritis develops in 15–20% of IBD patients, is more common in CD, and worsens with exacerbations of bowel activity. It is asymmetric, polyarticular, and migratory and most often affects large joints of the upper and lower extremities. Treatment is directed at reducing bowel inflammation. In severe UC, colectomy frequently cures the arthritis.

Ankylosing spondylitis (AS) occurs in ~10% of IBD patients and is more common in CD than UC. About two-thirds of IBD patients with AS express the HLA-B27 antigen. The AS activity is not related to bowel activity and does not remit with glucocorticoids or colectomy. It most often affects the spine and pelvis, producing symptoms of diffuse low-back pain, buttock pain, and morning stiffness. The course is continuous and progressive, leading to permanent skeletal damage and deformity. Anti-TNF therapy reduces spinal inflammation and improves functional status and quality of life.

Sacroiliitis is symmetric, occurs equally in UC and CD, is often asymptomatic, does not correlate with bowel activity, and does not always progress to AS. Other rheumatic manifestations include hypertrophic osteoarthropathy, pelvic/femoral osteomyelitis, and relapsing polychondritis.

■ OCULAR

The incidence of ocular complications in IBD patients is 1–10%. The most common are conjunctivitis, anterior uveitis/iritis, and episcleritis. Uveitis is associated with both UC and Crohn's colitis, may be found during periods of remission, and may develop in patients following bowel resection. Symptoms include ocular pain, photophobia, blurred vision, and headache. Prompt intervention, sometimes with systemic glucocorticoids, is required to prevent scarring and visual impairment. Episcleritis is a benign disorder that presents with symptoms of mild ocular burning. It occurs in 3–4% of IBD patients, more commonly in Crohn's colitis, and is treated with topical glucocorticoids.

■ HEPATOBILIARY

Hepatic steatosis is detectable in about one-half of the abnormal liver biopsies from patients with CD and UC; patients usually present with hepatomegaly. Fatty liver usually results from a combination of chronic debilitating illness, malnutrition, and glucocorticoid therapy. Cholelithiasis occurs in 10–35% of CD patients with ileitis or ileal resection. Gallstone formation is caused by malabsorption of bile acids, resulting in depletion of the bile salt pool and the secretion of lithogenic bile.

Primary sclerosing cholangitis (PSC) is a disorder characterized by both intrahepatic and extrahepatic bile duct inflammation and fibrosis, frequently leading to biliary cirrhosis and hepatic failure; ~5% of patients with UC have PSC, but 50–75% of patients with PSC have IBD. PSC occurs less often in patients with CD. Although it can be recognized after the diagnosis of IBD, PSC can be detected earlier or even years after proctocolectomy. Consistent with this, the immunogenetic basis for PSC appears to be overlapping but distinct from UC based on GWAS, although both IBD and PSC are commonly pANCA positive. Most patients have no symptoms at the time of diagnosis; when symptoms are present, they consist of fatigue, jaundice, abdominal pain, fever, anorexia, and malaise. The traditional gold standard diagnostic test is endoscopic retrograde cholangiopancreatography (ERCP), but magnetic resonance cholangiopancreatography (MRCP) is sensitive, specific, and safer. MRCP is reasonable as an initial diagnostic test in children and adults and can visualize irregularities, multifocal strictures, and dilatations of all levels of the biliary tree. In patients with PSC, both ERCP and MRCP demonstrate multiple bile duct strictures alternating with relatively normal segments.

Gallbladder polyps in patients with PSC have a high incidence of malignancy, and cholecystectomy is recommended, even if a mass lesion is <1 cm in diameter. Gallbladder surveillance with ultrasound should be performed annually. Endoscopic stenting may be palliative for cholestasis secondary to bile duct obstruction. Patients with symptomatic disease develop cirrhosis and liver failure over 5–10 years and eventually require liver transplantation. PSC patients have a 10–15% lifetime risk of developing cholangiocarcinoma and then cannot be transplanted. Patients with IBD and PSC are at increased risk of colon cancer and should be surveyed yearly by colonoscopy and biopsy.

In addition, cholangiography is normal in a small percentage of patients who have a variant of PSC known as *small duct primary sclerosing cholangitis*. This variant (sometimes referred to as "pericholangitis") is probably a form of PSC involving small-caliber bile ducts. It has similar biochemical and histologic features to classic PSC. It has a significantly better prognosis than classic PSC, although it may evolve into classic PSC. Granulomatous hepatitis and hepatic amyloidosis are much rarer EIMs of IBD.

■ UROLOGIC

The most frequent genitourinary complications are calculi, ureteral obstruction, and ileal bladder fistulas. The highest frequency of nephrolithiasis (10–20%) occurs in patients with CD following small-bowel resection. Calcium oxalate stones develop secondary to hyperoxaluria, which results from increased absorption of dietary oxalate. Normally, dietary calcium combines with luminal oxalate to form insoluble calcium oxalate, which is eliminated in the stool. In patients with ileal dysfunction, however, nonabsorbed fatty acids bind calcium and leave oxalate unbound. The unbound oxalate is then delivered to the colon, where it is readily absorbed, especially in the presence of inflammation.

■ METABOLIC BONE DISORDERS

Low bone mass occurs in 14–42% of IBD patients. The risk is increased by glucocorticoids, CSA, methotrexate (MTX), and total parenteral nutrition (TPN). Malabsorption and inflammation mediated by IL-1, IL-6, TNF, and other inflammatory mediators also contribute to low bone density. An increased incidence of hip, spine, wrist, and rib fractures has been noted: 36% in CD and 45% in UC. The absolute risk of an osteoporotic fracture is ~1% per person per year. Fracture rates, particularly in the spine and hip, are highest among the elderly (age >60). One study noted an OR of 1.72 for vertebral fracture and an OR of 1.59 for hip fracture. The disease severity predicted the risk of a fracture. Only 13% of IBD patients who had a fracture were on any kind of antifracture treatment. Up to 20% of bone mass can be lost per year with chronic glucocorticoid use. The effect is dose-dependent. Budesonide may also suppress the pituitary-adrenal axis and thus carries a risk of causing osteoporosis.

Osteonecrosis is characterized by death of osteocytes and adipocytes and eventual bone collapse. The pain is aggravated by motion and swelling of the joints. It affects the hips more often than knees and shoulders, and in one series, 4.3% of patients developed osteonecrosis within 6 months of starting glucocorticoids. Diagnosis is made by bone scan or magnetic resonance imaging (MRI), and treatment consists of pain control, cord decompression, osteotomy, and joint replacement.

■ THROMBOEMBOLIC DISORDERS

Patients with IBD have an increased risk of both venous and arterial thrombosis even if the disease is not active. Factors responsible for the hypercoagulable state have included abnormalities of the platelet-endothelial interaction, hyperhomocysteinemia, alterations in the coagulation cascade, impaired fibrinolysis, involvement of tissue factor–bearing microvesicles, disruption of the normal coagulation system by autoantibodies, and a genetic predisposition. A spectrum of vasculitides involving small, medium, and large vessels has also been observed.

More common cardiopulmonary manifestations include endocarditis, myocarditis, pleuropericarditis, and interstitial lung disease. A secondary or reactive amyloidosis can occur in patients with long-standing IBD, especially in patients with CD. Amyloid material is deposited systemically and can cause diarrhea, constipation, and renal failure. The renal disease can be successfully treated with colchicine. Pancreatitis is a rare EIM of IBD and results from duodenal fistulas; ampullary CD; gallstones; PSC; drugs such as mercaptopurine, azathioprine, or, very rarely, 5-ASA agents; autoimmune pancreatitis; and primary CD of the pancreas.

TREATMENT

Inflammatory Bowel Disease

5-ASA AGENTS

These agents are effective at inducing and maintaining remission in UC. Peroxisome proliferator–activated receptor γ (PPAR-γ) may mediate 5-ASA therapeutic action by decreasing nuclear localization of NF-κB. Sulfa-free aminosalicylate formulations include alternative azo-bonded carriers, 5-ASA dimers, and delayed-release and controlled-release preparations. Each has the same efficacy as sulfasalazine when equimolar concentrations are used.

Sulfasalazine is effective treatment for mild to moderate UC, but its high rate of side effects limits its use. Although sulfasalazine is more effective at higher doses, at 6 or 8 g/d, up to 30% of patients experience allergic reactions or intolerable side effects such as headache, anorexia, nausea, and vomiting that are attributable to the sulfapyridine moiety that is attached to 5-ASA. Hypersensitivity reactions, independent of sulfapyridine levels, include rash, fever, hepatitis, agranulocytosis, hypersensitivity pneumonitis, pancreatitis, worsening of colitis, and reversible sperm abnormalities. Sulfasalazine can also impair folate absorption, and patients should be given folic acid supplements.

Balsalazide contains an azo bond binding mesalamine to the carrier molecule 4-aminobenzoyl-β-alanine; it is effective in the colon.

Delzicol and Asacol HD (high dose) are enteric-coated forms of mesalamine with the 5-ASA being released at pH >7. They disintegrate with complete breakup of the tablet occurring in many different parts of the gut ranging from the small intestine to the splenic flexure; they have increased gastric residence when taken with a meal. *Lialda* is a once-a-day formulation of mesalamine (Multi-Matrix System [MMX]) designed to release mesalamine in the colon. The MMX technology incorporates mesalamine into a lipophilic matrix within a hydrophilic matrix encapsulated in a polymer resistant to degradation at a low pH (<7) to delay release throughout the colon. The safety profile appears to be comparable to other 5-ASA formulations.

Apriso is a formulation containing encapsulated mesalamine granules that delivers mesalamine to the terminal ileum and colon via a proprietary extended-release mechanism (Intellicor). The outer coating of this agent (Eudragit L) dissolves at a pH >6. In addition, there is a polymer matrix core that aids in sustained release throughout the colon. Because Lialda and Apriso are given once daily, an anticipated benefit is improved compliance compared with two to four daily doses required for other mesalamine preparations.

Pentasa is another mesalamine formulation that uses an ethylcellulose coating to allow water absorption into small beads containing the mesalamine. Water dissolves the 5-ASA, which then diffuses out of the bead into the lumen. Disintegration of the capsule occurs in the stomach. The microspheres then disperse throughout the entire GI tract from the small intestine through the distal colon in both fasted and fed conditions.

Salofalk Granu-Stix, an unencapsulated version of mesalamine, has been in use in Europe for induction and maintenance of remission for several years.

Appropriate doses of the 5-ASA compounds are shown in **Table 337-8**. Some 50–75% of patients with mild to moderate UC improve when treated with 5-ASA doses equivalent to 2 g/d of mesalamine; the dose response continues up to at least 4.8 g/d.

More common side effects of the 5-ASA medications include headaches, nausea, hair loss, and abdominal pain. Rare side effects of the 5-ASA medications include renal impairment, hematuria, pancreatitis, and paradoxical worsening of colitis. Renal function tests and urinalysis should be checked yearly.

Topical *Rowasa* enemas are composed of mesalamine and are effective in mild-to-moderate distal UC. Combination therapy with mesalamine in both oral and enema form is more effective than either treatment alone for both distal and extensive UC.

Canasa suppositories composed of mesalamine are effective in treating proctitis.

GLUCOCORTICOIDS

The majority of patients with moderate to severe UC benefit from oral or parenteral glucocorticoids. Prednisone is usually started at doses of 40–60 mg/d for active UC that is unresponsive to 5-ASA therapy. Parenteral glucocorticoids may be administered as hydrocortisone, 300 mg/d, or methylprednisolone, 40–60 mg/d. A newer glucocorticoid for UC, budesonide (Uceris), is released entirely in the colon and has minimal to no glucocorticoid side effects. The dose is 9 mg/d for 8 weeks, and no taper is required. Topically applied glucocorticoids (hydrocortisone enemas or budesonide foam) are also beneficial for distal colitis and may serve as an adjunct in those who have rectal involvement plus more proximal

TABLE 337-8 Oral 5-Aminosalicylic Acid (5-ASA) Preparations

PREPARATION	FORMULATION	DELIVERY	DOSING PER DAY
Azo-Bond			
Sulfasalazine (500 mg) (Azulfidine)	Sulfapyridine-5-ASA	Colon	3–6 g (acute) 2–4 g (maintenance)
Balsalazide (750 mg) (Colazal)	Aminobenzoyl-alanine–5-ASA	Colon	6.75–9 g
Delayed-Release			
Mesalamine (400, 800 mg) (Delzicol, Asacol HD)	Eudragit S (pH 7)	Distal ileum-colon	2.4–4.8 g (acute) 1.6–4.8 g (maintenance)
Mesalamine (1.2 g) (Lialda)	MMX mesalamine (SPD476)	Ileum-colon	2.4–4.8 g
Controlled-Release			
Mesalamine (250, 500, 1000 mg) (Pentasa)	Ethylcellulose microgranules	Stomach-colon	2–4 g (acute) 1.5–4 g (maintenance)
Delayed- and Extended-Release			
Mesalamine (0.375 g) (Apriso)	Intellicor extended-release mechanism	Ileum-colon	1.5 g (maintenance)

Abbreviation: MMX, Multi-Matrix System.

disease. Hydrocortisone enemas are significantly absorbed from the rectum and can lead to adrenal suppression with prolonged administration. Topical 5-ASA therapy is more effective than topical steroid therapy in the treatment of distal UC.

Glucocorticoids are also effective for treatment of moderate to severe CD. Controlled-ileal-release budesonide has been nearly equal to prednisone for ileocolonic CD with fewer glucocorticoid side effects. Budesonide is used for 2–3 months at a dose of 9 mg/d and then tapered. Glucocorticoids play no role in maintenance therapy in either UC or CD. Once clinical remission has been induced, they should be tapered according to the clinical activity, normally at a rate of no more than 5–10 mg/week. The side effects are numerous, including fluid retention, abdominal striae, fat redistribution, hyperglycemia, subcapsular cataracts, osteonecrosis, osteoporosis, myopathy, emotional disturbances, and withdrawal symptoms. Most of these side effects, aside from osteonecrosis, are related to the dose and duration of therapy.

ANTIBIOTICS

Antibiotics have no role in the treatment of active or quiescent UC. However, pouchitis, which occurs in ~30–50% of UC patients after colectomy and ileal pouch anal anastomosis (IPAA), usually responds to treatment with a variety of antibiotics including metronidazole and ciprofloxacin. Some patients require long-term treatment with antibiotics for chronic pouchitis.

AZATHIOPRINE AND MERCAPTOPURINE

Azathioprine and mercaptopurine (MP) are purine analogues used concomitantly with biologic therapy or, much less often, as the sole immunosuppressants. Azathioprine is rapidly absorbed and converted to MP, which is then metabolized to the active end product, thioinosinic acid, an inhibitor of purine ribonucleotide synthesis and cell proliferation. Efficacy can be seen as early as 3–4 weeks but can take up to 4–6 months. Adherence can be monitored by measuring the levels of 6-thioguanine and 6-methylmercaptopurine, end products of MP metabolism. The doses used range from 2 to 3 mg/kg per day for azathioprine and 1 to 1.5 mg/kg per day for MP.

Although azathioprine and MP are usually safe, pancreatitis occurs in 3–4% of patients, typically presents within the first few weeks of therapy, and is completely reversible when the drug is stopped. Other side effects include nausea, fever, rash, and hepatitis. Bone marrow suppression (particularly leukopenia) is dose-related and often delayed, necessitating regular monitoring of the complete blood cell count (CBC). Additionally, 1 in 300 individuals lacks thiopurine methyltransferase, the enzyme responsible for drug metabolism to inactive end products (6-methylmercaptopurine); an additional 11% of the population are heterozygotes with intermediate enzyme activity. Both are at increased risk of toxicity because of increased accumulation of active 6-thioguanine metabolites. Although 6-thioguanine and 6-methylmercaptopurine levels can be followed to determine correct drug dosing and reduce toxicity, weight-based dosing is an acceptable alternative. CBCs and liver function tests should be monitored frequently regardless of dosing strategy.

One meta-analysis demonstrated a fourfold risk of lymphoma in IBD patients on azathioprine and MP. The highest risk for thiopurine-associated lymphoma is in patients >65 years old actively using thiopurines (yearly incidence rate per 1000 patient-years of 5.41), with a moderate risk in those between the ages of 50 and 65 (incidence rate of 2.58 compared to an incidence rate of 0.37 in patients <50 years old). Patients using thiopurines also have a two- to threefold increased risk of nonmelanoma skin cancers.

METHOTREXATE

MTX inhibits dihydrofolate reductase, resulting in impaired DNA synthesis. Additional anti-inflammatory properties may be related to decreases in the production of IL-1. It is used most often concomitantly with biologic therapy to decrease antibody formation and improve disease response. Intramuscular (IM) or subcutaneous (SC) doses range from 15 to 25 mg/week. Potential toxicities include leukopenia and hepatic fibrosis, necessitating periodic evaluation of CBCs and liver enzymes. The role of liver biopsy in patients on long-term MTX is uncertain but is probably limited to those with increased liver enzymes. Hypersensitivity pneumonitis is a rare but serious complication of therapy.

CYCLOSPORINE

CSA is a lipophilic peptide with inhibitory effects on both the cellular and humoral immune systems. CSA blocks the production of IL-2 by T helper lymphocytes. CSA binds to cyclophilin, and this complex inhibits calcineurin, a cytoplasmic phosphatase enzyme involved in the activation of T cells. CSA also indirectly inhibits B-cell function by blocking helper T cells. CSA has a more rapid onset of action than MP and azathioprine.

CSA is most effective when given at 2–4 mg/kg per day IV in severe UC that is refractory to IV glucocorticoids, with 82% of patients responding. CSA can be an alternative to infliximab for steroid-refractory UC before colectomy. Levels as measured by monoclonal radioimmunoassay or by the high-performance liquid chromatography assay should be maintained between 150 and 350 ng/mL.

CSA may cause significant toxicity; renal function should be monitored frequently. Hypertension, gingival hyperplasia, hypertrichosis, paresthesias, tremors, headaches, and electrolyte abnormalities are common side effects. Creatinine elevation calls for dose reduction or discontinuation. Seizures may also complicate therapy, especially if the patient is hypomagnesemic or if serum cholesterol levels are <3.1 mmol/L (<120 mg/dL). Opportunistic infections, most notably *Pneumocystis jirovecii* pneumonia, may occur with combination immunosuppressive treatment; antibiotic prophylaxis with trimethoprim-sulfamethoxazole should be given.

TACROLIMUS

Tacrolimus is a macrolide antibiotic with immunomodulatory properties similar to CSA but 100 times as potent and not dependent on bile or mucosal integrity for absorption. Thus, tacrolimus has good oral absorption despite proximal small-bowel Crohn's involvement. Tacrolimus is effective in children with refractory IBD and in adults with extensive involvement of the small bowel. Tacrolimus use is decreasing due to side effects and the expanding choice of biologic and small-molecule therapies.

BIOLOGIC THERAPIES

Biologic therapy is now commonly given as an initial therapy for patients with moderate to severe CD and UC to prevent future complications of IBD. High-risk patients with UC who are more likely to require biologics include those with moderate to severe disease, steroid-dependent or steroid-refractory disease, and refractory pouchitis. High-risk patients with CD who are more likely to require biologics include those who are <30 years old, with extensive disease, perianal or severe rectal disease and/or deep ulcerations in the colon, and stricturing or penetrating disease behavior. The current goal of IBD treatment with biologics is to treat early in the disease course, treat aggressively with appropriate therapies, check drug and drug metabolite levels, administer dual therapy with immunomodulators and biologics in appropriate patients, and aim for deep remission (endoscopic and histologic remission). Patients who respond to biologic therapies enjoy an improvement in clinical symptoms; a better quality of life; less disability, fatigue, and depression; and fewer surgeries and hospitalizations.

Anti-TNF Therapies TNF is a proinflammatory cytokine that regulates immune cells to coordinate a systemic immune response. Dysregulation of TNF production has been associated with immune-mediated disorders including IBD, and inhibition of TNF signaling is used in the treatment of IBD. Four TNF inhibitors are currently approved for the treatment of IBD: infliximab, adalimumab, certolizumab pegol, and golimumab. Infliximab, a chimeric IgG1 antibody against TNF-α, was the first biologic therapy approved for moderately to severely active inflammatory and fistulizing CD and UC.

Randomized trials support combination therapy with infliximab and azathioprine for moderate to severe CD and UC. For moderate to severe CD, combination therapy has been shown to be more effective than either infliximab or azathioprine alone. Similarly, combination therapy has been shown to be more effective for moderate to severe UC than either infliximab or azathioprine alone.

Hospitalized patients with acute severe glucocorticoid-refractory UC have a high inflammatory burden and may develop a protein-losing enteropathy, leading to an accelerated consumption, excessive fecal wasting, and low serum concentrations of infliximab. Given a clear exposure–response relationship for infliximab in patients with IBD, intensive infliximab dosing regimens have been used in these patients.

Adalimumab (ADA) is a recombinant human monoclonal IgG1 antibody containing only human peptide sequences and is injected subcutaneously. ADA binds TNF and neutralizes its function by blocking the interaction between TNF and its cell-surface receptor. Therefore, it seems to have a similar mechanism of action to infliximab but with less immunogenicity. ADA is approved for treatment of moderate to severe CD and UC. For CD and UC, results with ADA are better in patients who are naïve to anti-TNF than in patients who have previously been treated with infliximab. In clinical practice, the remission rate in both CD and UC patients taking ADA increases with a dose increase to 40 mg weekly instead of every other week.

Certolizumab pegol is a pegylated form of an anti-TNF Fab portion of an antibody administered SC once monthly. SC certolizumab pegol is effective for induction of clinical response in patients with active inflammatory CD.

Golimumab is another fully human IgG1 antibody against TNF-α and is currently approved for the treatment of moderately to severely active UC. Like ADA and certolizumab, golimumab is injected SC.

Side Effects of Anti-TNF Therapies

Development of Antibodies and Drug Levels The development of antibodies to infliximab is associated with an increased risk of infusion reactions and a decreased response to treatment. Current practice does not include giving on-demand or episodic infusions in contrast to scheduled periodic infusions because patients are most likely to develop antibodies. Anti-infliximab antibodies are generally present when the quality of response or the response duration to infliximab infusion decreases. Commercial assays can detect both infliximab and ADA antibodies and measure trough levels to determine optimal dosing. If a patient has high anti-infliximab antibodies and a low trough level of infliximab, it is best to switch to another anti-TNF therapy. If a patient has a therapeutic anti-TNF level and active inflammatory symptoms, the drug should be switched to a different class of biologic. Most acute infusion reactions and serum sickness can be managed with glucocorticoids and antihistamines. Some reactions can be serious and would necessitate a change in therapy, especially if a patient has anti-infliximab antibodies. It is now common practice to add an immunomodulator such as azathioprine, MP, or MTX to anti-TNF therapy to help prevent antibody formation.

Non-Hodgkin's Lymphoma (NHL) The baseline risk of NHL in CD patients is 2 in 10,000, slightly higher than in the general population. Azathioprine and/or MP therapy increases the risk to ~4 in 10,000. It is difficult to assess whether anti-TNF medications are associated with lymphoma because many patients are also receiving thiopurines. After adjustment for co-treatments, no excess risk of lymphoma was found in a Danish study of a cohort of IBD patients exposed to anti-TNF medications.

Hepatosplenic T-Cell Lymphoma (HSTCL) HSTCL is a nearly universally fatal lymphoma in patients with or without CD. In patients with CD, a total of 37 unique cases have been reported. Eighty-six percent of the patients were male, and the median age was 26 years. Patients had CD for a mean of 10 years before the diagnosis of HSTCL. Thirty-six patients had used either MP or azathioprine, and 28 patients had used infliximab.

Skin Lesions New-onset psoriasiform skin lesions develop in nearly 5% of IBD patients treated with anti-TNF therapy. Most often, these can be treated topically, and occasionally, anti-TNF therapy must be decreased, switched, or stopped. Patients with IBD may have a slight, unexplained, intrinsic higher risk of developing melanoma. The risk of melanoma is increased almost twofold with anti-TNF and not thiopurine use. The risk of nonmelanoma skin cancer is increased with thiopurines and biologics, especially with ≥1 year of follow-up. Patients on these medications should have a skin check at least once a year.

Infections All of the anti-TNF drugs are associated with an increased risk of infections, particularly reactivation of latent tuberculosis and opportunistic fungal infections including disseminated histoplasmosis and coccidioidomycosis. Patients should have a purified protein derivative (PPD) or a QuantiFERON-TB Gold test before initiation of anti-TNF therapy. Patients >65 years old have a higher rate of infections and death on infliximab or ADA than those <65 years old.

Other Acute liver injury due to reactivation of hepatitis B virus and to autoimmune effects and cholestasis has been reported. Rarely, infliximab and the other anti-TNF drugs have been associated with optic neuritis, seizures, new onset or exacerbation of clinical symptoms, and radiographic evidence of central nervous system demyelinating disorders, including multiple sclerosis. They may exacerbate symptoms in patients with New York Heart Association functional class III/IV heart failure.

ANTI-INTEGRINS

Integrins are expressed on the cell surface of leukocytes and serve as mediators of leukocyte adhesion to vascular endothelium and entry into the intestines. α4-Integrin along with its β1 or β7 subunit interact with endothelial ligands termed adhesion molecules. Interaction between α4β7 and mucosal addressin cellular adhesion molecule (MAdCAM-1) is important in lymphocyte trafficking to gut mucosa.

Natalizumab is a recombinant humanized IgG4 antibody against α4-integrin and is effective in induction and maintenance of patients with CD. However, natalizumab is no longer used for CD due to the risk of progressive multifocal leukoencephalopathy (PML) and the development of alternative biologic and small-molecule therapies.

Vedolizumab (VDZ), another leukocyte trafficking inhibitor, is a monoclonal antibody directed against α4β7-integrin specifically and has the ability to convey gut-selective immunosuppression. Unlike natalizumab, it inhibits adhesion of a discrete gut-homing subset of T lymphocytes to MAdCAM-1, but not to vascular adhesion molecule 1. VDZ decreases GI inflammation without inhibiting systemic immune responses or affecting T-cell trafficking to the central nervous system. It may be prescribed as a first-line biologic or after failure of a TNF antagonist in patients with CD or UC. Vedolizumab is more effective than adalimumab for first-line biologic therapy in moderate to severe UC.

Ustekinumab, a fully human IgG1 monoclonal antibody, blocks the biologic activity of IL-12 and IL-23 through their common p40 subunit by inhibiting the interaction of these cytokines with their receptors on T cells, natural killer cells, and antigen-presenting cells. It is as equally effective as adalimumab for first-line biologic therapy for moderate to severe CD and is another option for treatment of moderate to severe UC. It is particularly appealing for use in patients with concomitant psoriatic arthritis.

Risankizumab is a monoclonal antibody that selectively binds to the IL-23 p19 subunit, inhibiting its interaction with the IL-23R complex It is the first selective IL-23 inhibitor approved for moderate to severe CD and provides an additional therapeutic option for patients, particularly those who have been previously treated with other advanced IBD therapies.

SMALL MOLECULES

Small molecules (drugs with molecular weight <1 kDa) are a new class of orally administered medications developed for IBD that lack the immunogenicity associated with monoclonal antibodies. The advantage of small molecules is their ability to diffuse through cell membranes into the intracellular space and inhibit cytokine signaling pathways. This mechanism of action may be more efficacious compared to monoclonal antibodies that inhibit specific targets because several cytokine pathways are involved in IBD pathogenesis and inhibiting numerous cytokines may be synergistic. A key regulatory pathway of many cytokines associated with IBD is the JAK/STAT pathway that activates transcription and translation of proteins that mediate the immune response. Janus kinase (JAK) is a family of intracellular, nonreceptor tyrosine kinases that regulate cytokine signaling via the JAK/STAT pathway, ultimately suppressing the immune response and inflammation. The JAK family members include JAK1, JAK2, JAK3, and tyrosine kinase 2 (TYK2).

Tofacitinib is a reversible and competitive JAK inhibitor used for the treatment of moderate to severe UC refractory to conventional therapy. It competes with ATP to bind to the ATP-docking site of the kinase domain of JAK. By competing with ATP, tofacitinib inhibits phosphorylation and activation of JAK, leading to downstream reduction of cytokine production and alteration of the immune response. Although tofacitinib is a pan-JAK inhibitor, it has higher specificity for JAK1 and JAK3 than for JAK2 and TYK2. The pan-JAK inhibition is concerning for adverse events and overall safety.

Upadacitinib is an oral, selective, and reversible JAK inhibitor that is approved for both UC and CD. It potently inhibits JAK1 and is less potent against the other isoforms, JAK2, JAK3, and TYK2. Upadacitinib demonstrates rapid clinical and endoscopic improvement at the end of induction, which is sustained to the end of maintenance, with a positive benefit/risk profile.

The FDA review concluded that there is an increased risk of serious adverse events including heart attack, stroke, cancer, blood clots, and death in patients who are prescribed JAK inhibitors. Patients who are at risk for cardiovascular disease, have a history of blood clots, are current or past smokers, and/or are over the age of 50 should consider alternative therapies. The risk of herpes zoster is higher with JAK inhibitors, and patients should receive the shingles vaccine (Shingrix). Tofacitinib and upadacitinib are FDA approved for use as second-line agents after failure of anti-TNF therapy.

Ozanimod is a potent sphingosine-1-phosphate (S1P1) receptor modulator that binds selectively with high affinity to the S1P receptor subtypes S1P1 and S1P5, both of which are involved in immune regulation. By preventing trafficking of disease-exacerbating lymphocytes from the lymph nodes to the gut, ozanimod may provide immunomodulatory effects and moderate disease processes.

Ozanimod is approved for the treatment of moderate to severe ulcerative colitis. It is administered as a daily capsule.

The biologic and small-molecule therapies used in daily practice are detailed in **Table 337-9**.

NUTRITIONAL THERAPIES

Diet has long been thought to contribute to the pathogenesis of IBD and may also be an avenue for managing disease activity. Diet plays a significant role in shaping the gut microbiome, and dietary components may interact with the microbiome and stimulate a mucosal immune response. In fact, active CD responds to exclusive enteral nutrition (EEN) or bowel rest with TPN, interventions as effective as glucocorticoids in inducing remission but not as effective for maintenance therapy. In contrast to CD, active UC is not effectively treated by elemental diets or TPN.

Dietary approaches to maintenance therapy in CD have largely been adapted from epidemiologic studies; however, significant heterogeneity is noted among research study outcomes. In general, low fiber, refined carbohydrates (especially sweetened beverages), animal fats, red meat, and processed meat have been associated with onset of IBD. Therefore, the overall dietary approach is to maximize fiber intake, particularly from fruits and vegetables, and to limit consumption of higher-risk foods. Several defined diets adhere to these principles with some variation, including the Mediterranean diet pattern, specific carbohydrate diet, semi-vegetarian diet, and IBD anti-inflammatory diet (IBD-AID). However, it remains unclear whether diet studies will eventually lead to evidence-based nutrition guidelines.

Standard medical management of UC and CD is shown in **Fig. 337-12**.

SURGICAL THERAPY

Ulcerative Colitis Nearly one-half of patients with extensive chronic UC undergo surgery within the first 10 years of their illness. The indications for surgery are listed in **Table 337-10**. Morbidity is ~20% for elective, 30% for urgent, and 40% for emergency proctocolectomy. The risks are primarily hemorrhage, contamination and sepsis, and neural injury. The operation of choice is an IPAA.

Because UC is a mucosal disease, the rectal mucosa can be dissected and removed down to the dentate line of the anus or ~2 cm proximal to this landmark. The ileum is fashioned into a pouch that serves as a neorectum. This ileal pouch is then sutured circumferentially to the anus in an end-to-end fashion. If performed carefully, this operation preserves the anal sphincter and maintains continence. The overall operative morbidity is 10%, with the major complication being bowel obstruction. Pouch failure necessitating conversion to permanent ileostomy occurs in 5–10% of patients. Some inflamed rectal mucosa is usually left behind, and thus, endoscopic surveillance is necessary. Primary dysplasia of the ileal mucosa of the pouch has occurred rarely.

Patients with IPAA usually have ~6–10 bowel movements a day. On validated quality-of-life indices, they report better performance in sports and sexual activities than ileostomy patients. The most frequent complication of IPAA is pouchitis in ~30–50% of patients with UC. This syndrome consists of increased stool frequency, watery stools, cramping, urgency, nocturnal leakage of stool, arthralgias, malaise, and fever. Pouch biopsies may distinguish true pouchitis from underlying CD. Although pouchitis usually responds to antibiotics, 3–5% of patients remain refractory and may require chronic antibiotic therapy, biologics, or even pouch removal.

Crohn's Disease The majority of patients with CD will require at least one operation in their lifetime. The need for surgery is related to duration of disease and the site of involvement. Patients with small-bowel disease have an 80% chance of requiring surgery. Those with colitis alone have a 50% chance. Surgery is an option only when medical treatment has failed or complications dictate its necessity. The indications for surgery are shown in Table 337-10.

Small-Intestinal Disease Because CD is chronic and recurrent, with no clear surgical cure, as little intestine as possible is resected. Current surgical alternatives for treatment of obstructing CD include resection of the diseased segment and strictureplasty. Surgical resection of the diseased segment is the most frequently performed operation, and in most cases, primary anastomosis can be done to restore continuity. An end-to-end anastomosis may provide the best opportunity for an optimal functional outcome, compared to an antiperistaltic side-to-side anastomosis, which creates a functional block to motility leading to distention and pain at the anastomotic site in a subgroup of patients. If much of the small bowel has already been resected and the strictures are short, with intervening areas of normal mucosa, strictureplasties should be done to avoid a functionally insufficient length of bowel. The strictured area of intestine is incised longitudinally and the incision sutured transversely, thus widening the narrowed area. Complications of strictureplasty include prolonged ileus, hemorrhage, fistula, abscess, leak, and restricture.

Risk factors for early recurrence of disease include cigarette smoking, penetrating disease (internal fistulas, abscesses, or other evidence of penetration through the wall of the bowel), early recurrence since a previous surgery, multiple surgeries, and a young age at the time of the first surgery. Aggressive postoperative treatment with biologics should be considered for this group of patients. It is also recommended to evaluate for endoscopic recurrence of CD via a colonoscopy, if possible, 3–6 months after surgery.

TABLE 337-9 Biologic and Small-Molecule Agents in the Treatment of Inflammatory Bowel Disease

MEDICATION	DOSAGE	INDICATION	SERIOUS TOXICITIES	OTHER COMMON SIDE EFFECTS	TESTING
Infliximab	5 mg/kg at 0, 2, and 6 weeks, then every 8 weeks; may increase dose to 10 mg/kg every 4 weeks depending on trough levels Intensive dosing for hospitalized corticosteroid-refractory patients	Moderate to severe Crohn's disease and ulcerative colitis Fistulizing Crohn's disease	Increased risk of infections (bacterial and fungal), tuberculosis (TB) reactivation, hepatitis B reactivation, lymphoma (controversial), psoriasis, melanoma and nonmelanoma skin cancers, drug-induced lupus Contraindicated in multiple sclerosis, class III/IV congestive heart failure	Infusion reactions	Prior to infusion: TB testing Hepatitis B testing (HBsAb, HBsAg, HBcAb) Maintenance: Skin check yearly Influenza, pneumococcal vaccinations Hepatitis B vaccine if not immune
Adalimumab	160 mg day 0, 80 mg day 14 and then 40 mg every 14 days; may increase to 40 mg every 7 days depending on trough levels	Moderate to severe Crohn's disease and ulcerative colitis Fistulizing Crohn's disease	As above	Injection site reactions (better with citrate-free preparation)	As above
Certolizumab	400 mg on days 0 and 14, then 400 mg every 28 days	Moderate to severe Crohn's disease	As above	As above	As above
Golimumab	200 mg on day 0, 100 mg on day 14, then 100 mg every 28 days	Moderate to severe ulcerative colitis	As above	As above	As above
Vedolizumab	300 mg at 0, 2, and 6 weeks, then every 8 weeks; may increase dose to 300 mg every 4 weeks	Moderate to severe ulcerative colitis (more effective than adalimumab as first-line therapy in one study) Moderate to severe Crohn's disease.	No increased risk of serious systemic or opportunistic infections No increased risk of malignancy	Nasopharyngitis, headache, arthralgias, nausea	Prior to infusion: TB testing Hepatitis B testing (HBsAb, HBsAg, HBcAb) Maintenance: Influenza, pneumococcal vaccinations Hepatitis B vaccine if not immune
Ustekinumab	6 mg/kg IV, then 90 mg every 8 weeks; may increase dose to 90 mg every 4 weeks	Moderate to severe Crohn's disease and ulcerative colitis	Reversible posterior leukoencephalopathy syndrome (presents with headaches, seizures, confusion, and visual disturbances), anaphylaxis, and angioedema	Nasopharyngitis, upper respiratory tract infection, fatigue, headache	Prior to infusion: TB testing Hepatitis B testing (HBsAb, HBsAg, HBcAb) Maintenance: Influenza, pneumococcal vaccinations Hepatitis B vaccine if not immune
Risankizumab	600 mg IV every 4 weeks × 3 doses then 180 mg or 360 mg SC every 8 weeks	Moderate to severe Crohn's disease	Reversible posterior leukoencephalopathy syndrome (presents with headaches, seizures, confusion, and visual disturbances), anaphylaxis, and angioedema	Nasopharyngitis, upper respiratory tract infection, fatigue, headache	TB testing Hepatitis B testing (HBsAb, HBsAg, HBcAb) Maintenance: Influenza, pneumococcal vaccinations Hepatitis B vaccine if not immune
Tofacitinib	10 mg bid; can decrease to 5 mg bid when patient in remission	Moderate to severe ulcerative colitis	Increased risk of heart attack, stroke, cancer, blood clots, and death. Patients who are at risk for cardiovascular disease, are current or past smokers, and/or are over the age of 50 should consider alternative therapies. Increased risk of viral infections, including herpes zoster, and bacterial and invasive fungal infections	Elevated lipids, neutropenia, anemia, elevated liver enzymes	Prior to infusion: First dose of Shingrix recommended, TB testing Hepatitis B testing (HBsAb, HBsAg, HBcAb) Maintenance: Influenza, pneumococcal vaccinations Hepatitis B vaccine if not immune
Upadacitinib	45 mg PO qd × 8 weeks for ulcerative colitis; 45 mg PO qd × 12 weeks for Crohn's disease, then 15 or 30 mg qd	Moderate to severe ulcerative colitis or Crohn's disease	Increased risk of heart attack, stroke, cancer, blood clots, and death. Patients who are at risk for cardiovascular disease, are current or past smokers, and/or are over the age of 50 should consider alternative therapies. Increased risk of viral infections, including herpes zoster, and bacterial and invasive fungal infections	Elevated lipids, neutropenia, anemia, elevated liver enzymes	Prior to infusion: First dose of Shingrix recommended, TB testing Hepatitis B testing (HBsAb, HBsAg, HBcAb) Maintenance: Influenza, pneumococcal vaccinations Hepatitis B vaccine if not immune
Ozanimod	0.23 mg PO days 1–4, 0.46 mg PO days 5–7, 0.92 mg PO thereafter	Moderate to severe UC	Bradycardia, atrioventricular conduction delay, increased blood pressure, macular edema, posterior reversible encephalopathy syndrome	Infections, elevated liver enzymes	Prior to infusion: First dose of Shingrix recommended, TB testing Hepatitis B testing (HBsAb, HBsAg, HBcAb) Maintenance: Influenza, pneumococcal vaccinations Hepatitis B vaccine if not immune

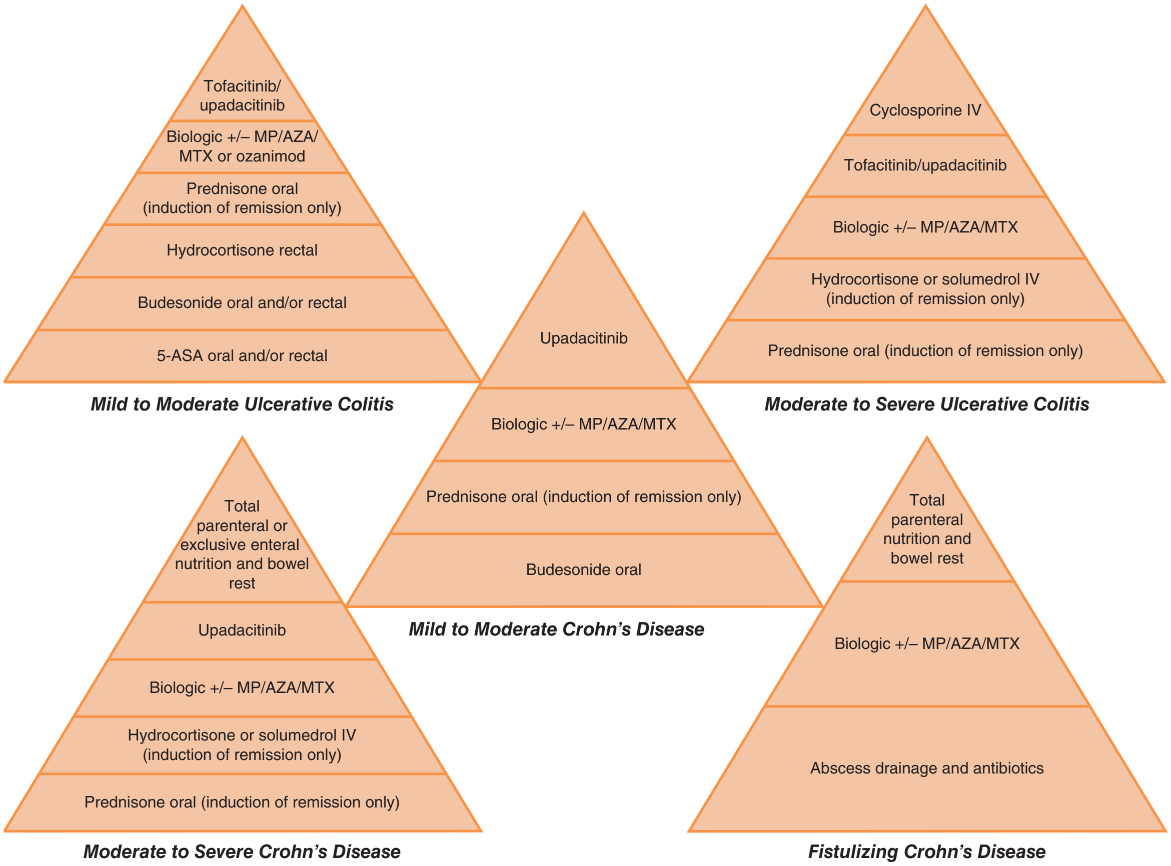
Tofacitinib/ upadacitinib
Biologic +/– MP/AZA/ MTX or ozanimod
Prednisone oral (induction of remission only)
Hydrocortisone rectal
Budesonide oral and/or rectal
5-ASA oral and/or rectal
Upadacitinib
Biologic +/– MP/AZA/MTX
Prednisone oral (induction of remission only)
Budesonide oral
Cyclosporine IV
Tofacitinib/upadacitinib
Biologic +/– MP/AZA/MTX
Hydrocortisone or solumedrol IV (induction of remission only)
Prednisone oral (induction of remission only)
Total parenteral or exclusive enteral nutrition and bowel rest
Upadacitinib
Biologic +/– MP/AZA/MTX
Hydrocortisone or solumedrol IV (induction of remission only)
Prednisone oral (induction of remission only)
Total parenteral nutrition and bowel rest
Biologic +/– MP/AZA/MTX
Abscess drainage and antibiotics

Mild to Moderate Ulcerative Colitis

Mild to Moderate Crohn's Disease

Moderate to Severe Ulcerative Colitis

Moderate to Severe Crohn's Disease

Fistulizing Crohn's Disease

FIGURE 337-12 Medical management of inflammatory bowel disease. 5-ASA, 5-aminosalicylic acid; AZA, azathioprine; MP, mercaptopurine; MTX, methotrexate.

Colorectal Disease A greater percentage of patients with Crohn's colitis require surgery for intractability, fulminant disease, and anorectal disease. Several alternatives are available, ranging from the use of a temporary loop ileostomy to resection of segments of diseased colon or even the entire colon and rectum. For patients with segmental involvement, segmental colon resection with primary anastomosis can be performed. In 20–25% of patients with extensive colitis, the rectum is spared sufficiently to consider rectal preservation. Most surgeons believe that an IPAA is contraindicated in CD due to the high incidence of pouch failure. A diverting colostomy may help heal severe perianal disease or rectovaginal fistulas, but disease almost always recurs with reanastomosis. These patients often require a total proctocolectomy and ileostomy.

TABLE 337-10 Indications for Surgery

ULCERATIVE COLITIS	CROHN'S DISEASE
Intractable disease	Small Intestine
Fulminant disease	Stricture and obstruction
Toxic megacolon	unresponsive to medical therapy
Colonic perforation	Massive hemorrhage
Massive colonic hemorrhage	Refractory fistula
Extracolonic disease	Abscess
Colonic obstruction	Colon and rectum
Colon cancer prophylaxis	Intractable disease
Colon dysplasia or cancer	Fulminant disease
	Perianal disease unresponsive to medical therapy
	Refractory fistula
	Colonic obstruction
	Cancer prophylaxis
	Colon dysplasia or cancer

IBD AND PREGNANCY

Patients with quiescent UC and CD have normal fertility rates; the fallopian tubes can be scarred by the inflammatory process of CD, especially on the right side because of the proximity of the terminal ileum. In addition, perirectal, perineal, and rectovaginal abscesses and fistulas as well as pelvic surgery can result in dyspareunia. Infertility in men can be caused by sulfasalazine but reverses when treatment is stopped. Women with an IPAA have decreased fertility due to scarring or occlusion of the fallopian tubes secondary to pelvic inflammation and adhesions, although studies have shown that fertility is improved with laparoscopic versus open IPAA.

Mild or quiescent UC or CD has no effect on birth outcomes. The courses of CD and UC during pregnancy mostly correlate with disease activity at the time of conception. Patients should be in remission for 3–6 months before conceiving. Most CD patients can deliver vaginally, but cesarean delivery may be the preferred route of delivery for patients with anorectal and perirectal abscesses and fistulas to reduce the likelihood of fistulas developing or extending into the episiotomy scar. Unless they desire multiple children, UC patients with an IPAA may consider a cesarean delivery due to an increased risk of future fecal incontinence.

Sulfasalazine and all mesalamines are safe for use in pregnancy and nursing with the caveat that additional folate supplementation must be given with sulfasalazine. Topical 5-ASA agents are safe during pregnancy and nursing. Glucocorticoids are generally safe for use during pregnancy and are indicated for patients with moderate to severe disease activity. The amount of glucocorticoids received by the nursing infant is minimal. The safest antibiotics to use for CD in pregnancy for short periods of time (weeks, not months) are ampicillin and cephalosporins. Metronidazole can be used in the second or third trimester. Ciprofloxacin causes cartilage lesions in immature animals and should be avoided because of the absence of data on its effects on growth and development in humans.

MP and azathioprine pose minimal or no risk during pregnancy. Breast milk has been shown to contain negligible levels of MP/azathioprine when measured in a limited number of patients.

MTX is teratogenic and should be discontinued at least 3 months before conception.

In a large prospective and multiple retrospective studies, no increased risk of stillbirths, miscarriages, or spontaneous abortions was seen with infliximab, ADA, certolizumab, vedolizumab, or ustekinumab. These biologics, with the exception of certolizumab, are IgG1 antibodies and are actively transported across the placenta in the late second and third trimesters. Infants can have serum levels of infliximab, ADA, vedolizumab, and ustekinumab well into the first year of life, and live vaccines should be avoided until 12 months of age. Certolizumab crosses the placenta by passive diffusion, and infant serum and cord blood levels are minimal. The biologics are safe in nursing. Miniscule levels of infliximab, ADA, certolizumab, vedolizumab, and ustekinumab have been reported in breast milk. These levels are of no clinical significance. It is recommended that drugs should not be switched during pregnancy unless necessitated by the medical condition of the IBD. Tofacitinib and upadacitinib should not be used during pregnancy. Animal studies show teratogenic effects with both of these drugs, and data in humans are limited. A washout period of at least 4 weeks is recommended before conception. Surgery in UC should be performed only for emergency indications, including severe hemorrhage, perforation, and megacolon refractory to medical therapy. Total colectomy and ileostomy carry a 50% risk of postoperative spontaneous abortion. The best time to perform surgery is in the second trimester if necessary. Patients with IPAAs have increased nighttime stool frequency during pregnancy that resolves postpartum. Transient small-bowel obstruction or ileus has been noted in up to 8% of patients with ileostomies.

CANCER IN IBD

ULCERATIVE COLITIS

Patients with long-standing UC are at increased risk for developing colonic epithelial dysplasia and carcinoma (Fig. 337-13).

The risk of neoplasia in chronic UC increases with duration and extent of disease. In contrast to the relatively high risk in one large meta-analysis (2% after 10 years, 8% after 20 years, and 18% after 30 years of disease), a decrease in the risk of colorectal cancer has been noted over time potentially due to better control of inflammation and better colonoscopic surveillance. The rates of colon cancer are still about 1.5–2 times higher than in the general population, and colonoscopic surveillance is the standard of care.

Annual or biennial colonoscopy with multiple biopsies is recommended for patients with >8–10 years of extensive colitis (greater than one-third of the colon involved) or 12–15 years of proctosigmoiditis (less than one-third but more than just the rectum) and has been widely used to screen and survey for subsequent dysplasia and carcinoma. International guideline societies have recommended chromoendoscopy for dysplasia surveillance in IBD. Chromoendoscopy enhances the visualization of the surface and pit pattern of the mucosa, as well as borders of lesions, in order to better define areas of dysplasia compared to standard-definition white light endoscopy. In real-life settings, the practice has been to use standard-definition white light endoscopy with surveillance biopsies in patients with chronic colitis at average risk and chromoendoscopy in higher-risk patients including those with a history of dysplasia, PSC, or family history of colorectal cancer.

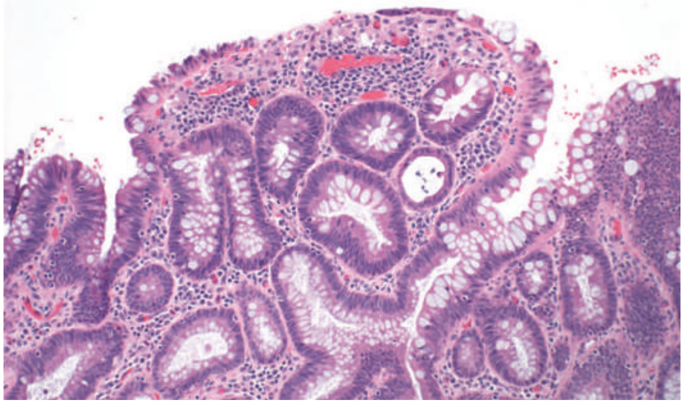

FIGURE 337-13 Medium-power view of low-grade dysplasia in a patient with chronic ulcerative colitis. Low-grade dysplastic crypts are interspersed among regenerating crypts. (*Courtesy of Dr. R. Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.*)

Risk factors for cancer in UC include long-duration disease, extensive disease, family history of colon cancer, PSC, a colon stricture, and the presence of postinflammatory pseudopolyps on colonoscopy.

CROHN'S DISEASE

Risk factors for developing cancer in Crohn's colitis are long-duration and extensive disease, bypassed colon segments, colon strictures, PSC, and family history of colon cancer. The cancer risks in CD and UC are probably equivalent for similar extent and duration of disease. In the CESAME study, a prospective observational cohort of IBD patients in France, the standardized incidence ratios of colorectal cancer were 2.2 for all IBD patients (95% CI, 1.5–3.0; $p < .001$) and 7.0 for patients with long-standing extensive colitis (both Crohn's and UC) (95% CI, 4.4–10.5; $p < .001$). Thus, the same endoscopic surveillance strategy used for UC is recommended for patients with chronic Crohn's colitis. A pediatric colonoscope can be used to pass narrow strictures in CD patients, but surgery should be considered in symptomatic patients with impassable strictures.

MANAGEMENT OF DYSPLASIA AND CANCER

Dysplasia can be flat or polypoid. If flat high-grade dysplasia is encountered on colonoscopic surveillance, the usual treatment is colectomy for UC and either colectomy or segmental resection for CD. If flat low-grade dysplasia is found (Fig. 337-13), most investigators recommend immediate colectomy. Adenomas may occur coincidently in UC and CD patients with chronic colitis and can be removed endoscopically provided that biopsies of the surrounding mucosa are free of dysplasia.

IBD patients are also at greater risk for other malignancies. Patients with CD may have an increased risk of NHL, leukemia, and myelodysplastic syndromes. Severe, chronic, complicated perianal disease in CD patients may be associated with an increased risk of cancer in the lower rectum and anal canal (squamous cell cancers). Although the absolute risk of small-bowel adenocarcinoma in CD is low (2.2% at 25 years in one study), patients with long-standing, extensive, small-bowel disease should be considered for screening.

COVID-19 AND IBD

COVID-19, caused by SARS-CoV-2, was first reported in December 2019 and has rapidly spread throughout the world, leading to an international pandemic. Glucocorticoids, immunomodulators (thiopurines, MTX), biologics, and JAK inhibitors, commonly used to treat IBD, are associated with higher rates of serious viral and bacterial infections, and patients with IBD using these medications are potentially at increased risk of a serious COVID-19 infection. Yet, it is also possible that some forms of immune suppression may blunt the excessive immune response/cytokine storm characteristic of severe COVID-19 infection and consequently reduce mortality. Using data from the Surveillance Epidemiology of Coronavirus Under Research Exclusion for Inflammatory Bowel Disease, it was found that increasing age (adjusted OR 1.04; 95% CI, 1.01–1.02), two or more comorbidities (adjusted OR 2.9; 95% CI, 1.1–7.8), and systemic glucocorticoids (adjusted OR 6.9; 95% CI, 2.3–20.5) are associated with severe COVID-19 in IBD patients. Anti-TNF treatment was not associated with severe COVID-19 (adjusted OR 0.9; 95% CI, 0.4–2.2).

FURTHER READING

Barberio B et al: Efficacy of biological therapies and small molecules in induction and maintenance of remission in luminal Crohn's disease: Systematic review and network meta-analysis. Gut 72:264, 2023.

Chang JT: Pathophysiology of inflammatory bowel disease. N Eng J Med 383:2652, 2020.

D'Haens G et al: Risankizumab as induction therapy for Crohn's disease: Results from the phase 3 ADVANCE and MOTIVATE induction trials. Lancet 399:2015, 2022.

Graham DB, Xavier RJ: Pathway paradigms revealed from the genetics of inflammatory bowel disease. Nature 578:527, 2020.

Levine A et al: Crohn's disease exclusion diet plus partial enteral nutrition induces sustained remission in a randomized controlled trial. Gastroenterology 157:440, 2019.

Lewis JD et al: A randomized trial comparing the specific carbohydrate diet to a Mediterranean diet in adults with Crohn's disease. Gastroenterology 161:837, 2021.

Lewis JD et al: Incidence, prevalence, and racial and ethnic distribution of inflammatory bowel disease in the United States. Gastroenterology 165:1197, 2023.

Mahadevan U et al: Pregnancy and neonatal outcomes after fetal exposure to biologics and thiopurines among women with inflammatory bowel disease. Gastroenterology 160:1131, 2021.

Ng SC et al: Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: A systematic review of population-based studies. Lancet 390:2769, 2017.

Sands BE: Ustekinumab versus adalimumab for induction and maintenance therapy in biologic-naive patients with moderately to severely active Crohn's disease: A multicentre, randomised, double-blind, parallel-group, phase 3b trial. Lancet 399:2200, 2022.

Sands BE et al: VARSITY Study Group. Vedolizumab versus adalimumab for moderate-to-severe ulcerative colitis. N Engl J Med 381:1215, 2019.

338

Irritable Bowel Syndrome

Chung Owyang, Prashant Singh

Irritable bowel syndrome (IBS) is a functional bowel disorder characterized by abdominal pain or discomfort and altered bowel habits in the absence of detectable structural abnormalities. No clear diagnostic markers exist for IBS; thus, the diagnosis of the disorder is based on clinical presentation. In 2016, the Rome III criteria for the diagnosis of IBS were updated to Rome IV **(Table 338-1)**. Throughout the world, ~10% of adults and adolescents have symptoms consistent with IBS. IBS symptoms fluctuate over time and often overlap with other functional disorders such as fibromyalgia, headache, backache, and genitourinary symptoms. Severity of symptoms varies and can significantly impair quality of life, resulting in high direct and indirect health care costs. Altered gastrointestinal (GI) motility, visceral hyperalgesia, disturbance of brain–gut interaction, abnormal central processing, autonomic and hormonal events, microbial dysbiosis, genetic and environmental factors, and psychosocial disturbances are variably involved, depending on the individual.

TABLE 338-1 Rome IV Diagnostic Criteria for Irritable Bowel Syndrome[a]

Recurrent abdominal pain, on average, at least 1 day per week in the last 3 months, associated with ≥2 of the following criteria:

1. Related to defecation
2. Associated with a change in frequency of stool
3. Associated with a change in form (appearance) of stool

[a] Criteria fulfilled for the last 3 months with symptom onset at least 6 months prior to diagnosis.

CLINICAL FEATURES

IBS is a disorder that affects all ages, although most patients have their first symptoms before age 45. Women are diagnosed with IBS two to three times as often as men and make up 80% of the population with severe IBS. As indicated in Table 338-1, abdominal pain is essential for diagnosing IBS. This should be associated with defecation and/or have its onset associated with a change in frequency or form of stool. In comparison to Rome III, the Rome IV criteria are more stringent, requiring abdominal pain to occur at a minimum of once a week and eliminating "discomfort" as one of the criteria. Painless diarrhea or constipation does not fulfill the diagnostic criteria to be classified as IBS. Supportive symptoms that are not part of the diagnostic criteria include straining, urgency or a feeling of incomplete defecation, passing mucus, and bloating.

Abdominal Pain According to the current IBS diagnostic criteria, abdominal pain is a prerequisite clinical feature of IBS. Abdominal pain in IBS is highly variable in intensity and location and can occur anywhere in the abdomen. It is frequently episodic and crampy, but it may be superimposed on a background of constant ache. Pain may be mild enough to be ignored, or it may interfere with daily activities. Despite this, malnutrition due to inadequate caloric intake is exceedingly rare with IBS. Sleep deprivation is also unusual because abdominal pain is almost uniformly present only during waking hours. Pain is often exacerbated by eating or emotional stress and improved by the passage of flatus or stools. In addition, female patients with IBS commonly report worsening symptoms during the premenstrual and menstrual phases.

Altered Bowel Habits Alteration in bowel habits is the most consistent clinical feature in IBS. Based on the predominant bowel type, IBS can be classified into IBS with predominant constipation (IBS-C), IBS with predominant diarrhea (IBS-D), and IBS with both mixed constipation and diarrhea (IBS-M). However, bowel pattern subtypes are highly unstable. In a patient population with ~33% prevalence rates of IBS-D, IBS-C, and IBS-M forms, 75% of patients change subtypes and 29% switch between IBS-C and IBS-D over 1 year.

Patients with IBS-C have symptoms of hard stools (often with narrow caliber), infrequent stools (less than three bowel movements per week), incomplete evacuation, straining, sensation of anal blockage, and digitalization. On the other hand, those with IBS-D usually have small volumes of loose stools (generally <200 mL). This is often associated with fecal urgency, the passage of mucus, and in some cases incontinence. Diarrhea may be aggravated by emotional stress or eating. Weight loss, nocturnal symptoms, and bleeding per rectum are not common in patients with IBS and, when present, should warrant evaluation for other causes.

Gas and Flatulence Patients with IBS frequently complain of bloating, abdominal distention and increased belching or flatulence, all of which they attribute to increased gas. Although some patients with these symptoms actually may have a larger amount of gas, quantitative measurements reveal that most patients who complain of increased gas generate no more than a normal amount of intestinal gas. Most patients with IBS have impaired transit and intolerance of intestinal gas loads. In addition, patients with IBS tend to reflux gas from the distal to the more proximal intestine, which may explain the belching. Some patients with bloating may also experience visible distention with increase in abdominal girth.

Upper GI Symptoms Between 25 and 50% of patients with IBS complain of dyspepsia, heartburn, nausea, and vomiting. This suggests that other areas of the gut apart from the colon may be involved. Prolonged ambulant recordings of small-bowel motility in patients with IBS show a high incidence of abnormalities in the small bowel during the diurnal (waking) period; nocturnal motor patterns are not different from those of healthy controls. The overlap between dyspepsia and IBS is high. The prevalence of IBS is higher among patients with dyspepsia (31.7%) than among those who reported no symptoms of dyspepsia (7.9%). Conversely, among patients with IBS, 55.6% reported symptoms of dyspepsia. In addition, the functional abdominal symptoms can change over time. Those with predominant dyspepsia or IBS can flux between the two. Thus, it is conceivable that functional dyspepsia and IBS are two manifestations of a single, more extensive digestive system disorder. Furthermore, IBS symptoms are prevalent in other functional GI disorders such as noncardiac chest pain patients and esophageal hypersensitivity.

■ PATHOPHYSIOLOGY

The pathogenesis of IBS is poorly understood, although roles of abnormal gut motor and sensory activity, central neural dysfunction, psychological disturbances, mucosal inflammation, stress, and luminal factors such as bile acid malabsorption and gut dysbiosis have been proposed **(Fig. 338-1)**.

GI Motor Abnormalities Studies of colonic myoelectrical and motor activity under unstimulated conditions have not shown consistent abnormalities in IBS. In contrast, colonic motor abnormalities are more prominent under stimulated conditions in IBS. IBS patients may exhibit increased rectosigmoid motor activity for up to 3 h after eating. Similarly, inflation of rectal balloons both in IBS-D and IBS-C patients leads to marked and prolonged distention-evoked contractile activity. Recordings from the transverse, descending, and sigmoid colon showed that the motility index and peak amplitude of high-amplitude propagating contractions (HAPCs) in IBS-D patients were greatly increased compared to those in healthy subjects and were associated with rapid colonic transit and accompanied by abdominal pain.

Visceral Hypersensitivity IBS patients frequently exhibit exaggerated sensory responses to visceral stimulation. The frequency of perceptions of food intolerance is at least twofold more common than in the general population. On the other hand, prolonged fasting in IBS patients is often associated with significant improvement in symptoms. Rectal balloon inflation produces nonpainful and painful sensations at lower volumes in IBS patients than in healthy controls without altering rectal tension, suggestive of visceral afferent dysfunction in IBS. Similar studies show gastric and esophageal hypersensitivity in patients with nonulcer dyspepsia and noncardiac chest pain, raising the possibility that these conditions have a similar pathophysiologic basis. Lipids lower the thresholds for the first sensation of gas, discomfort, and pain in IBS patients. Hence, postprandial symptoms in IBS patients may be explained in part by a nutrient-dependent exaggerated sensory component of the gastrocolonic response. In contrast to enhanced gut sensitivity, IBS patients do not exhibit heightened sensitivity elsewhere in the body. Thus, the afferent pathway disturbances in IBS appear to be selective for visceral innervation with sparing of somatic pathways. The mechanisms responsible for visceral hypersensitivity are still under investigation.

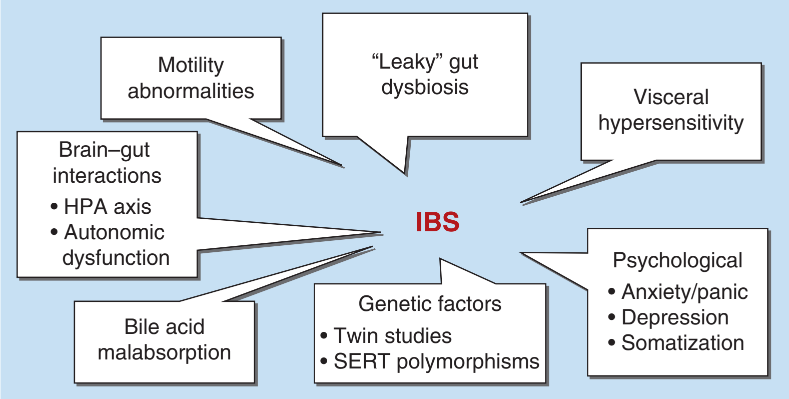

FIGURE 338-1 Pathophysiology of irritable bowel syndrome (IBS). The cause of IBS is likely to be multifactorial. Patients often show evidence of visceral hypersensitivity and motility abnormalities. Many IBS patients have increased anxiety and/or depression, and their symptoms are often exacerbated by mental or physical stress, suggesting abnormal brain–gut interaction. Genetic studies suggest a few IBS patients may have genetic abnormalities affecting the serotonin transport system in the enteric nerves. Up to 30% of IBS patients may have bile acid malabsorption. Gut dysbiosis and impaired mucosa permeability also have been reported in many IBS patients. This may lead to subclinical mucosa inflammation. HPA, hypothalamic-pituitary-adrenal; SERT, serotonin reuptake transporter.

Central Neural Dysregulation The role of central nervous system (CNS) factors in the pathogenesis of IBS is strongly suggested by the clinical association of emotional disorders and stress with symptom exacerbation and the therapeutic response to therapies that act on cerebral cortical sites. Functional brain imaging studies such as magnetic resonance imaging (MRI) have shown that in response to distal colonic stimulation, the mid-cingulate cortex—a brain region concerned with attention processes and response selection—shows greater activation in IBS patients. Modulation of this region is associated with changes in the subjective unpleasantness of pain. In addition, IBS patients also show preferential activation of the prefrontal lobe, which contains a vigilance network within the brain that increases alertness. These may represent a form of cerebral dysfunction, leading to hypervigilance of visceral pain.

Abnormal Psychological Features Abnormal psychiatric features are recorded in up to 80% of IBS patients, especially in referral centers; however, no single psychiatric diagnosis predominates. Most of these patients demonstrated exaggerated symptoms in response to visceral distention, and this abnormality persists even after exclusion of psychological factors.

Psychological factors influence pain thresholds in IBS patients, as stress alters sensory thresholds. An association between prior sexual or physical abuse and development of IBS has been reported. Clinical studies suggest that IBS has a strong developmental component that involves interactions of genetic and epigenetic factors early in life. These may modulate brain networks related to emotional arousal and/or central autonomic control, salience, and somatosensory integration. Abuse is associated with greater pain reporting, psychological distress, and poor health outcome. Brain functional MRI studies show greater activation of the posterior and middle dorsal cingulate cortex, which is implicated in affect processing in IBS patients with a past history of sexual abuse.

Postinfectious IBS A systematic review found that >10% of patients with infectious enteritis develop IBS later and the risk was four-fold higher compared to individuals without infectious enteritis. Conversely, about a third of IBS patients experienced an acute "gastroenteritis-like" illness at the onset of their chronic IBS symptomatology. This group of "postinfective" IBS occurs more commonly in females and affects younger rather than older patients. Other risk factors include clinical severity of infection (e.g., prolonged duration, requiring antibiotics), type of infection (risk is higher for parasitic and bacterial infection compared to viral), and presence of psychological comorbidities. Increased rectal mucosal enteroendocrine cells and T lymphocytes following *Campylobacter* enteritis may contribute to postinfective IBS. Persistent bacterial dysbiosis following an infection causing abnormal suppression of host proteolytic activity leading to altered intestinal barrier has also been implicated in a subset of these patients.

Immune Activation and Mucosal Inflammation Some patients with IBS display persistent signs of low-grade mucosal inflammation with activated lymphocytes, mast cells, and enhanced expression of proinflammatory cytokines. Other studies also indicate that peripheral blood mononuclear cells (PBMCs) from IBS patients show abnormal release of proinflammatory cytokines such as interleukin (IL) 6, IL-1β, and tumor necrosis factor (TNF). These abnormalities may contribute to abnormal epithelial secretion, increased permeability, and visceral hypersensitivity. Located at the host-environment interface, mast cells are in close proximity to sensory nerves. Electromicroscopic evidence of mast cell activation is commonly observed in the colonic mucosa of IBS patients. The proximity of activated mast cells to submucosal nerve fibers correlates with the frequency and severity of abdominal pain in patients with IBS. Other studies report that the colonic mucosa of IBS patients releases increased amounts of mast cell mediators, including histamine, proteases, and prostaglandin E_2.

These findings, together with the observation that marked excitation of visceral sensory nerves innervating the colon occurs after exposure to IBS mucosal supernatant, support a prominent role for mast cells in the pathogenesis of visceral hypersensitivity. Increasing evidence suggests that some members of the superfamily of transient receptor potential (TRP) cation channels such as TRPV1 (vanilloid) channels are central to the initiation and persistence of visceral hypersensitivity. Mucosal inflammation can lead to increased expression and/or sensitization of TRPV1. Enhanced expression of TRPV1 channels in the sensory neurons of the gut has been observed in IBS, and such expression appears to correlate with visceral hypersensitivity and abdominal pain.

Altered Intestinal Permeability The gut epithelial lining represents a semipermeable barrier that allows nutrient and water absorption while being restrictive against pathogenic molecules, toxins, and bacteria. The intestinal barrier is impaired in a subset of patients with IBS. This is primarily described in patients with IBS-D and postinfectious IBS. It is unclear whether or not those with IBS-C and IBS-M also have increased intestinal permeability. There appears to be a correlation between altered intestinal permeability and symptom severity in IBS. Studies have not only shown structural changes in barrier function, such as a decrease in gene and protein expression of tight junction proteins, but also functional consequences, such as the increased passage of macromolecules including bacteria across the epithelial barrier in Ussing chamber studies. Although the exact etiology of the altered mucosal barrier is not well understood, several factors such as stress, diet, and microbial dysbiosis appear to play a key role. Several studies have shown a key role of mast cell activation in mediating barrier dysfunction due to stress and diet-related dysbiosis.

Altered Gut Flora IBS patients have fourfold increased odds of having small-intestinal bacterial overgrowth compared to healthy controls based on positive lactulose or glucose hydrogen breath test. This finding, however, has been challenged by a number of other studies that found no increased incidence of bacterial overgrowth based on jejunal aspirate culture. An abnormal H_2 breath test can occur because of rapid small-bowel transit and may lead to erroneous interpretation. It is unclear if an abnormal breath test can predict the response to available treatment options such as nonselective antibiotics, and therefore, the role of testing for small-intestinal bacterial overgrowth in IBS patients remains unclear.

Many, but not all, studies using culture-independent approaches such as 16S rRNA gene-based analysis found significant differences between the molecular profile of the fecal microbiota of IBS patients and that of healthy subjects. Although an IBS-specific microbial signature has not yet been identified, but overall, the diversity of fecal microbiome appears to be significantly lower in IBS patients compared to healthy controls. However, the functional significance of this finding is not well understood. Despite a lack of consensus on the exact microbial difference between IBS patients and controls, IBS patients generally had decreased proportions of the genera *Bifidobacterium* and *Faecalibacterium* and increased abundance of family Enterobacteriaceae (phylum Proteobacteria), family Lactobacillaceae, and genus *Bacteroides* (phylum Bacteroidetes) **(Fig. 338-2)**. Experiments with animal models have shown that colonization of germ-free animals with microbiota from patients with IBS can induce visceral hypersensitivity, alter gut-brain axis, impair intestinal permeability, and alter GI transit time, indicating the importance and possible etiologic role of the microbiota in IBS. Gut dysbiosis acting in concert with genetic susceptibility and environmental insults may alter mucosal permeability and increase antigen presentation to the immune cells in the lamina propria. This may result in mast cell activation and altered enteric neuronal and smooth-muscle function causing IBS symptoms. In addition, release of cytokines and chemokines from mucosal inflammation may generate extra-GI symptoms such as chronic fatigue, muscle pain, and anxiety **(Fig. 338-3)**.

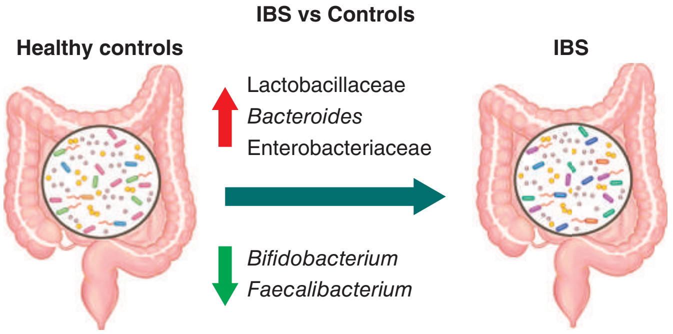

FIGURE 338-2 Changes in gut microbiota among patients with irritable bowel syndrome (IBS). (*Adapted from R Pittayanon et al: Gastroenterology 157:97, 2019.*)

Abnormal Serotonin Pathways The serotonin-containing enterochromaffin cells in the colon are increased in a subset of IBS-D patients compared to healthy individuals. Furthermore, postprandial plasma serotonin levels were significantly higher in this group of patients compared to healthy controls. Tryptophan hydroxylase 1 (TPH1) is the rate-limiting enzyme in enterochromaffin cell serotonin biosynthesis, and functional TPH1 polymorphism has been shown to be associated with IBS subtypes. In addition, gut microbes promote colonic serotonin production through an effect of short-chain fatty acids on enterochromaffin cells. In IBS patients, the expression of mucosal serotonin reuptake transporter (SERT) is downregulated due to gram-negative gut dysbiosis, leading to abnormal mucosal serotonin levels in IBS. Because serotonin plays an important role in the regulation of GI motility and visceral perception, the increased release of serotonin may contribute to the postprandial symptoms of these patients.

Bile Acids Up to 25% of patients with IBS-D are found to have evidence of idiopathic bile acid diarrhea. Mechanisms leading to increased hepatic synthesis and colonic bile acid levels in idiopathic bile acid diarrhea include reduced synthesis of fibroblast growth factor 19 (FGF-19) by the ileal mucosa or genetic variation in proteins involved in negative feedback of bile acid synthesis. Increased colonic bile acid levels in turn influence bowel habit by accelerating colonic transit, intestinal permeability, and visceral hypersensitivity.

APPROACH TO THE PATIENT

Irritable Bowel Syndrome

Because IBS is a disorder for which no pathognomonic abnormalities have been identified, its diagnosis relies on recognition of positive clinical features and elimination of other organic diseases. Symptom-based criteria have been developed for the purpose of differentiating patients with IBS from those with organic diseases. These include the Manning, Rome I, Rome II, Rome III, and Rome IV criteria. Rome IV criteria for the diagnosis of IBS were published in 2016 (Table 338-1) and defined IBS on the basis of abdominal pain and altered bowel habits that occur with sufficient frequency in affected patients. A careful history and physical examination are frequently helpful in establishing the diagnosis. Clinical features suggestive of IBS include recurrence of lower abdominal pain with altered bowel habits over a period of time without progressive deterioration, onset of symptoms during periods of stress or emotional upset, absence of other systemic symptoms such as fever and weight loss, and small-volume stool without any evidence of blood.

On the other hand, the appearance of the disorder for the first time in old age, progressive course from time of onset, rectal bleeding, significant weight loss persistent diarrhea after a 48-h fast, and presence of nocturnal diarrhea or steatorrheal stools argue against the diagnosis of IBS.

Because the major symptoms of IBS—abdominal pain, abdominal bloating, and alteration in bowel habits—are common complaints of many GI organic disorders, the list of differential diagnoses is a long one. The quality, location, and timing of pain may be

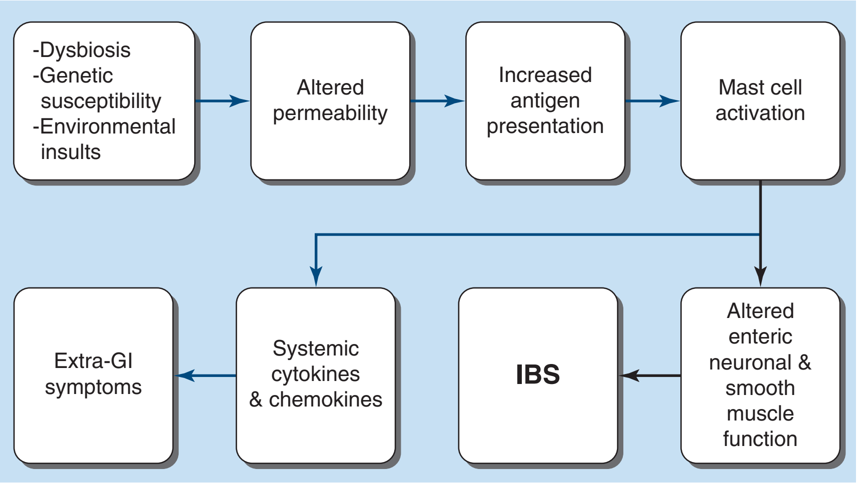

FIGURE 338-3 Gut dysbiosis and irritable bowel syndrome (IBS). Gut dysbiosis acting in concert with genetic and environmental factors may alter intestinal permeability and increase antigen presentation, resulting in mast cell activation. Products of mast cell degranulation may alter neuronal and smooth-muscle function causing IBS symptoms. The cytokines and chemokines generated from mucosal inflammation may cause symptoms such as fibromyalgia, chronic fatigue, and mood changes. GI, gastrointestinal. (*Adapted from NJ Talley, AA Fodor: Gastroenterology 141:1555, 2011.*)

helpful to suggest specific disorders. Pain due to IBS that occurs in the epigastric or periumbilical area must be differentiated from biliary tract disease, peptic ulcer disorders, intestinal ischemia, and carcinoma of the stomach and pancreas. If pain occurs mainly in the lower abdomen, the possibility of diverticular disease of the colon, inflammatory bowel disease (including ulcerative colitis and Crohn's disease), and carcinoma of the colon must be considered. Postprandial pain accompanied by bloating, nausea, and vomiting suggests gastroparesis or partial intestinal obstruction. Patients with small intestinal bacteria overgrowth can present with abdominal pain, nausea, and bloating. Intestinal infestation with *Giardia lamblia* or other parasites may cause similar symptoms. When diarrhea is the major complaint, the possibility of lactase deficiency, laxative abuse, malabsorption, celiac sprue, hyperthyroidism, inflammatory bowel disease (IBD), and infectious diarrhea must be ruled out. On the other hand, constipation may be a side effect of many different drugs, such as anticholinergic, antihypertensive, and antidepressant medications. Endocrinopathies such as hypothyroidism and hypoparathyroidism must also be considered in the differential diagnosis of constipation. In addition, acute intermittent porphyria and lead poisoning may present in a fashion similar to IBS, with painful constipation as the major complaint. These possibilities are suspected on the basis of their clinical presentations and are confirmed by appropriate serum and urine tests.

Few tests are required for patients who have typical IBS symptoms and no alarm features. Unnecessary investigations may be costly and even harmful. The American Gastroenterological Association has delineated factors to be considered when determining the aggressiveness of the diagnostic evaluation. These include the duration of symptoms, the change in symptoms over time, the age and sex of the patient, the referral status of the patient, prior diagnostic studies, a family history of colorectal malignancy, and the degree of psychosocial dysfunction. Thus, a younger individual with mild symptoms requires a minimal diagnostic evaluation, while an older person or an individual with rapidly progressive symptoms should undergo a more thorough exclusion of organic disease. Most patients should have a complete blood count and fecal calprotectin (or fecal lactoferrin), and/or C-reactive protein should be checked in patients with IBS-D even in the absence of alarm features to rule out IBD. Serologic testing for celiac disease should be performed in patients with symptoms of IBS-D if the prevalence of celiac disease in the general population is 1% or higher. Stool examination for ova and parasites is not necessary for all patients with IBS-D but could be helpful in regions where parasitic infections are endemic/common or if history suggests a possibility of parasitic infection (e.g., drinking untreated water). In the presence of alarm features or symptom onset after age 40, or in the absence of response to traditional therapies such as antidiarrheals, colonoscopy is warranted. Random biopsies to rule out microscopic colitis should be performed if the patient has persistent or chronic diarrhea. If the main symptoms are diarrhea and increased gas, the possibility of lactase deficiency should be ruled out with a hydrogen breath test or with evaluation after a 3-week lactose-free diet. Excessive gas with bloating also raises the possibility of small-bowel bacteria overgrowth, and this should be ruled out with a glucose hydrogen breath test. In patients with concurrent symptoms of dyspepsia, upper GI radiographs or esophagogastroduodenoscopy may be advisable. In patients with postprandial right upper quadrant pain, an ultrasonogram of the gallbladder should be obtained. Laboratory features that argue against IBS include evidence of anemia, elevated sedimentation rate, presence of leukocytes or blood in stool, and stool volume >200–300 mL/d. These findings would necessitate other diagnostic considerations.

TREATMENT

Irritable Bowel Syndrome

PATIENT COUNSELING AND DIETARY ALTERATIONS

Reassurance and careful explanation of the functional nature of the disorder are of primary importance, along with counseling on how to identify triggers for the patient's IBS symptoms. Stress management may be helpful if stress is a trigger, and identifying obvious food precipitants is an important first step in patient counseling and dietary change. As a therapeutic trial, patients should be encouraged to eliminate any foodstuffs that appear to produce symptoms. However, patients should avoid nutritionally depleted diets. It is unclear if exercise improves overall IBS symptoms, but given its low risk and overall health benefits, it should be encouraged in all IBS patients.

STOOL-BULKING AGENTS

High-fiber diets and fiber supplementation are frequently used in treating IBS. The water-holding action of fibers may contribute to increased stool bulk, and fiber also speeds up colonic transit in most people. Furthermore, because of its hydrophilic properties, soluble fiber binds water by soaking up extra water in the colon and therefore can be helpful with diarrhea in some IBS patients. Furthermore, fiber supplementation with psyllium has been shown to reduce perception of rectal distention, indicating that fiber may have a positive effect on visceral afferent function.

Controlled trials of dietary fiber in IBS patients have produced variable results. This is not surprising since IBS is a heterogeneous disorder, with some patients being constipated and other having predominant diarrhea. Most investigations report increases in stool weight, decreases in colonic transit times, and improvement in constipation. Others have noted benefits in patients with alternating diarrhea and constipation, pain, and bloating. However, most studies observe no response in patients with diarrhea- or pain-predominant IBS. While studies do not support the use of insoluble dietary fiber such as wheat bran in IBS, soluble fibers such as psyllium preparations are often found to be effective in managing IBS symptoms. Fiber should be started at a nominal dose and slowly titrated up as tolerated over the course of several weeks to a targeted dose of 20–30 g of total dietary and supplementary fiber per day. Even when used judiciously, fiber can exacerbate bloating, flatulence, constipation, and diarrhea.

ANTISPASMODICS

Clinicians have observed that anticholinergic drugs may provide temporary relief for symptoms such as painful cramps related to intestinal spasm. Although controlled clinical trials have produced mixed results, evidence generally supports the beneficial effects of anticholinergic drugs for pain. Physiologic studies demonstrate that anticholinergic drugs inhibit the gastrocolic reflex; hence, postprandial pain is best managed by giving antispasmodics 30 min before meals so that effective blood levels are achieved shortly before the anticipated onset of pain. Most anticholinergics contain natural belladonna alkaloids, which may cause xerostomia, urinary hesitancy and retention, blurred vision, and drowsiness. They should be used in the elderly with caution. Some physicians prefer to use synthetic anticholinergics such as dicyclomine that have less effect on mucous membrane secretion and produce fewer undesirable side effects. Peppermint oil appears to reduce abdominal cramps by some undefined mechanism. In a meta-analysis of 10 double-blind, randomized controlled trials evaluating 1030 IBS patients, peppermint oil was found to be significantly superior to placebo for global improvement of IBS symptoms and reduction in abdominal pain. The most commonly reported adverse event was heartburn, which was mild and transient.

ANTIDIARRHEAL AGENTS

Peripherally acting opiate-based agents are the initial therapy of choice for IBS-D. Physiologic studies demonstrate increases in segmenting colonic contractions, delays in fecal transit, increases in anal pressures, and reductions in rectal perception with these drugs. When diarrhea is severe or the predominant symptom, small doses of loperamide, 2–4 mg every 4–6 h up to a maximum of 12 mg/d, can be prescribed. These agents are less addictive than paregoric, codeine, or tincture of opium. In general, the intestines do not become tolerant of the antidiarrheal effect of opiates, and increasing doses are not required to maintain antidiarrheal potency. These agents are most useful if taken before anticipated stressful events that are known to cause diarrhea. However, not infrequently, a high dose of loperamide may cause cramping because of increases in colonic contractions. Eluxadoline, a locally acting, mixed μ-opioid receptor agonist, κ-opioid receptor agonist, and δ-opioid receptor antagonist with minimal systemic bioavailability, has been approved for treating abdominal pain and diarrhea by the U.S. Food and Drug administration (FDA) for patients with IBS-D. Given a small risk of pancreatitis associated with eluxadoline, it should not be prescribed for patients with prior cholecystectomy or significant alcohol use or prior history of pancreatitis. Other antidiarrheal agents that may be used in IBS patients are the bile acid binders cholestyramine or colesevelam because up to 30% of IBS-D patients may have bile acid malabsorption.

OSMOTIC LAXATIVES

Osmotic agents such as magnesium citrate–based products, sodium phosphate–based products, and nonabsorbable carbohydrates are hypertonic products that, through osmosis, extract fluid into the intestinal lumen to soften stool and enhance colonic transit. In contrast, polyethylene glycol (PEG)–based solution is iso-osmotic and induces bowel movement by high-volume lavage. The osmotic laxatives were better than placebo in improving symptoms of chronic constipation in clinical trials. However, chronic use of magnesium hydroxide may result in severe hypermagnesemia in patients with renal impairment. Frequent sodium phosphate–based bowel cleansing should be avoided as this is associated with hyperphosphatasemia, hypocalcemia, and hypokalemia. In 19 trials, PEG consistently induced more bowel movements than placebo. A Cochrane review of 10 randomized trials showed that PEG was superior to lactulose for improving stool frequency and abdominal pain. Among the nonabsorbable carbohydrates, lactulose and sorbitol had similar laxative effects. However, bacterial metabolism of unabsorbed carbohydrates often leads to gas production and abdominal pain, which can limit long-term use.

ANTIFLATULENCE THERAPY

The management of excessive gas is seldom satisfactory, except when there is obvious aerophagia or disaccharidase deficiency. Patients should be advised to eat slowly and not chew gum or drink carbonated beverages. Bloating may decrease if an associated gut syndrome such as IBS or constipation is improved. If bloating is accompanied by diarrhea and worsens after ingesting dairy products, fresh fruits, vegetables, or juices, further investigation or a dietary exclusion trial may be worthwhile. Avoiding gas forming foods, exercising, losing excess weight, and taking activated charcoal are safe but unproven remedies. A low FODMAP (fermentable oligosaccharides, disaccharides, monosaccharides, and polyols) diet has been shown to be quite effective in reducing gas and bloating (see “Low FODMAP Diet”) **(Table 338-2)**. Data regarding the use of surfactants such as simethicone are conflicting. Antibiotics may help in a subgroup of IBS patients with predominant symptoms of bloating. Beano, an over-the-counter oral β-glycosidase solution, may reduce rectal passage of gas without decreasing bloating and pain.

ANTIDEPRESSANT DRUGS

In addition to their mood-elevating effects, antidepressant medications have several physiologic effects that suggest they may be beneficial in IBS. In IBS-D patients, the tricyclic antidepressant imipramine slows jejunal migrating motor complex transit propagation and delays orocecal and whole-gut transit, indicative of a motor inhibitory effect. Some studies also suggest that tricyclic agents may alter visceral afferent neural function.

Tricyclic antidepressants is effective in IBS patients of all subtypes. The beneficial effects of the tricyclic compounds in the treatment of IBS appear to be independent of their effects on depression. The therapeutic benefits for the bowel symptoms occur faster and at a lower dosage. The efficacy of antidepressant agents from other chemical classes in the management of IBS is less well evaluated. Selective serotonin reuptake inhibitors (SSRIs) have also been studied in patients with IBS with conflicting results and significant heterogeneity among individual trials.

SEROTONIN RECEPTOR MODULATORS

Serotonin 5-HT$_3$ and 5-HT$_4$ receptors are found throughout the GI tract. Tegaserod, a 5-HT$_4$ receptor agonist, exhibits prokinetic activity by stimulating peristalsis. Clinical studies involving >4000 IBS-C patients reported reduction in abdominal discomfort and improvements in constipation and bloating compared to placebo. Diarrhea is the only major side effect. In 2007, the drug was voluntarily withdrawn from the market after a greater number of cardiovascular complications were observed in a database of 18,000 patients receiving tegaserod (0.11 vs 0.01% in placebo). In 2019, the FDA reviewed additional data and approved the use of tegaserod in women younger than 65 years old who do not have a history of ischemic cardiovascular disease and who have no more than one risk factor for cardiovascular disease.

TABLE 338-2 Some Common Food Sources of FODMAPs

FOOD TYPE	FREE FRUCTOSE	LACTOSE	FRUCTANS	GALACTO-OLIGOSACCHARIDES	POLYOLS
Fruits	Apple, cherry, mango, pear, watermelon		Peach, persimmon, watermelon		Apple, apricot, pear, avocado, blackberries, cherry, nectarine, plum, prune
Vegetables	Asparagus, artichokes, sugar snap peas		Artichokes, beetroot, Brussels sprout, chicory, fennel, garlic, leek, onion, peas		Cauliflower, mushroom, snow peas
Grains and cereals			Wheat, rye, barley		
Nuts and seeds			Pistachios		
Milk and milk products		Milk, yogurt, ice cream, custard, soft cheeses			
Legumes			Legumes, lentils, chickpeas	Legumes, chickpeas, lentils	
Other	Honey, high-fructose corn syrup		Chicory drinks		
Food additives			Inulin, FOS		Sorbitol, mannitol, maltitol, xylitol, isomalt

Abbreviations: FODMAPs, fermentable oligosaccharides, disaccharides, monosaccharides, and polyols; FOS, fructo-oligosaccharides.

Source: Reproduced with permission from PR Gibson et al: Food choice as a key management strategy for functional gastrointestinal symptoms. Am J Gastroenterol 107:657, 2012.

SECRETAGOGUES

Lubiprostone, linaclotide, and plecanatide are secretagogues that stimulate net efflux of ions and water into the intestinal lumen and thus enhance transit and facilitate ease of defecation. By activating channels on the apical (luminal) enterocyte surface, these secretagogues increase intestinal chloride secretion. Other ion channels and transporters secrete sodium into the intestine to maintain electroneutrality, followed by the secretion of water. Lubiprostone is a bicyclic fatty acid derived from prostaglandin E_1 that activates type 3 chloride channels in the apical membrane of intestinal epithelial cells. Oral lubiprostone was effective in the treatment of patients with IBS-C in large phase 3, randomized, double-blind, placebo-controlled multicenter trials. The recommended daily dose is 24 mcg twice daily. In general, the drug is well tolerated. The major side effects are nausea and diarrhea. Linaclotide and plecanatide are minimally absorbed 14-amino-acid peptide guanylate cyclase-C (GC-C) agonists that bind to and activate GC-C on the luminal surface of intestinal epithelium. The subsequent increase in cyclic guanosine monophosphate activates the cyclic fibrosis transmembrane regulator and induces fluid secretion into the GI tract. These drugs are similar to endogenous peptides secreted by the small intestine (uroguanylin) or colon (guanylin). In two 12-week, double-blind, randomized, controlled trials, linaclotide (290 or 145 μg, once daily) reduced constipation and pain. A lower dose (72 μg once daily) was also more effective than placebo. Linaclotide has been approved by the FDA for treatment of IBS-C. Plecanatide (3- and 6-mg doses) also has been shown to be more effective than placebo in two phase 3 trials for IBS-C. The 3-mg once-daily dose has been approved by the FDA. The only significant side effect was diarrhea, which occurred in <5% of patients. Tenapanor, a small-molecule inhibitor of GI sodium-hydrogen exchange-3, is more effective than placebo when given at 50 mg twice daily in patients with IBS-C and has been approved by the FDA for treatment of IBS-C.

MODULATION OF GUT FLORA

Because altered colonic flora (gut dysbiosis) may contribute to the pathogenesis of IBS, this has led to great interest in using antibiotics, prebiotics, probiotics, and dietary measures to treat IBS.

Antibiotics Antibiotic treatment benefits a subset of IBS patients. In a double-blind, randomized, placebo-controlled study, neomycin dosed at 500 mg twice daily for 10 days was more effective than placebo at improving symptom scores among IBS patients. The nonabsorbed oral antibiotic rifaximin is the most thoroughly studied antibiotic for the treatment of IBS. In two identical double-blind, placebo-controlled studies with >1200 nonconstipated IBS patients, those receiving rifaximin at a dose of 550 mg three times daily for 2 weeks were significantly more likely to report adequate relief of IBS symptoms compared with those taking placebo. Currently, rifaximin is approved by the FDA for treatment of IBS-D. Moreover, a placebo-controlled randomized clinical trial has shown that patients with IBS-D who respond to rifaximin are more likely to respond to subsequent courses of rifaximin compared to placebo if they experience relapse of symptoms.

Prebiotics These are nondigestible food ingredients that stimulate growth and/or activity of bacteria in the GI tract. There have been four randomized trials to examine the effects of prebiotics. Three of the four studies reported that prebiotics worsened or did not improve IBS symptoms. This is not surprising given the adverse effects of a high-carbohydrate diet on IBS symptoms.

Probiotics These are defined as live microorganisms that when administered in adequate amounts confer a health benefit on the host. Multiple meta-analyses have suggested that probiotics might be an effective strategy in improving global IBS symptoms including abdominal pain and bloating; however, given the limitations of existing studies (small sample size, short duration, lack of rigorous endpoints, single-centered), it is unclear which probiotics to recommend for patients with IBS. Given the lack of high-quality data, the American Gastroenterological Association makes no recommendations for use of probiotics in patients with IBS. Large-scale studies of well-phenotyped IBS patients are needed to establish the efficacy of probiotics in IBS.

Low FODMAP Diet A diet rich in FODMAPs often triggers symptoms in IBS patients. FODMAPs are poorly absorbed by the small intestine and fermented by bacteria in the colon to produce gas and osmotically active carbohydrates **(Fig. 338-4)**. At the same time, on entering the colon, FODMAPs may serve as nutrients for the colonic bacteria and promote the growth of gram-negative commensal bacteria, which may induce barrier dysfunction and subclinical mucosal inflammation. Fructose and fructans induce IBS symptoms in a dose-dependent manner. In contrast, a low FODMAP diet reduces IBS symptoms in 50–80% of patients with IBS. A network meta-analysis of 13 randomized controlled trials of IBS patients found that a low FODMAP diet was superior to all other dietary interventions in reducing global IBS symptoms and primarily improves abdominal pain and bloating in patients with IBS. The current approach to low FODMAP diet implementation

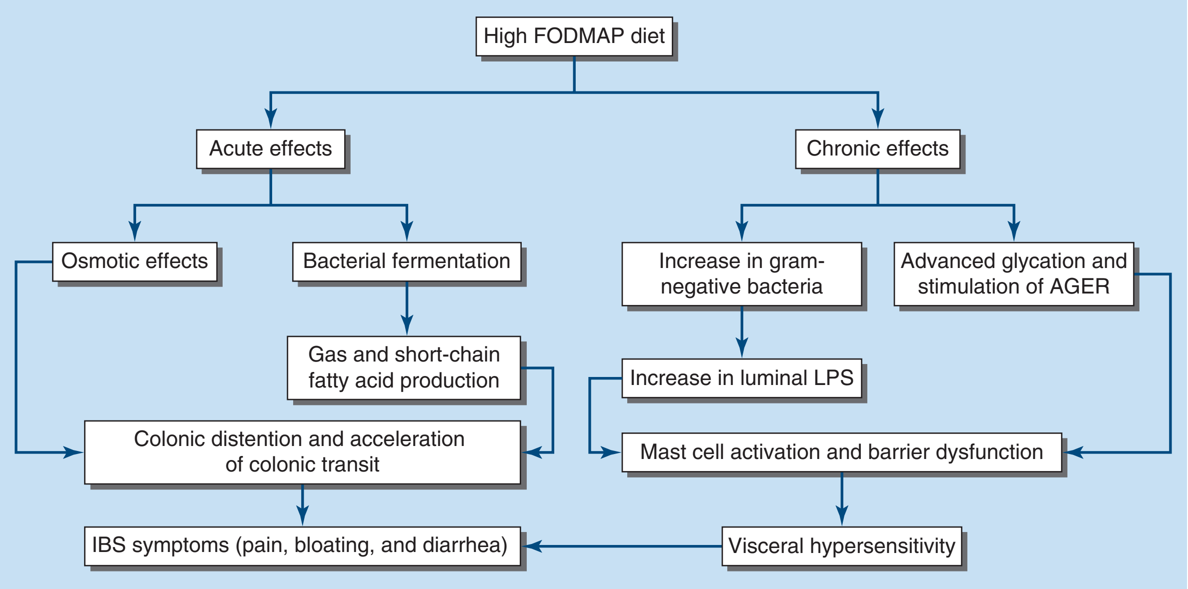
High FODMAP diet
Acute effects
Chronic effects
Osmotic effects
Bacterial fermentation
Increase in gram-negative bacteria
Advanced glycation and stimulation of AGER
Gas and short-chain fatty acid production
Increase in luminal LPS
Colonic distention and acceleration of colonic transit
Mast cell activation and barrier dysfunction
IBS symptoms (pain, bloating, and diarrhea)
Visceral hypersensitivity

FIGURE 338-4 Pathogenesis of FODMAP-related symptoms. FODMAPs are poorly absorbed by the small intestine and fermented by gut bacteria to produce gas and osmotically active carbohydrates. These events act in concert to cause bloating, flatulence, and diarrhea. FODMAPs may also serve as nutrients for colonic bacteria, which may induce mucosal changes such as mast cell activation and barrier dysfunction. AGER, advanced glycosylation end product receptor; FODMAP, fermentable oligosaccharides, disaccharides, monosaccharides, and polyols; LPS, lipopolysaccharide. (*Adapted from P Singh et al: Am J Gastroenterol 117:947, 2022.*)

is a three-step "step-down" process with an initial 2- to 6-week restriction phase (which restricts all five FODMAP subgroups [fructose, lactose, fructo- and galacto-oligosaccharides, and polyols]), followed by a reintroduction phase in which food challenges with individual FODMAPs are done to identify an individual's dietary triggers, and finally, a long-term maintenance phase in which the diet is personalized based on the outcome of the reintroduction phase. Physicians should be careful in not implementing a low FODMAP diet among patients with a history of eating disorders and those showing signs or symptoms of avoidant restricting eating behavior. Given that between 20 and 50% of patients do not respond to a low FODMAP diet, the identification of patients who are more likely to respond to a low FODMAP diet at baseline would be highly beneficial.

SUMMARY

The treatment strategy of IBS depends on the severity of the disorder **(Table 338-3)**. Most IBS patients have mild symptoms. They are usually cared for in primary care practices, have little or no psychosocial difficulties, and do not seek health care often. Treatment usually involves education, reassurance, and dietary/lifestyle changes. A smaller portion have moderate symptoms that are usually intermittent and correlate with altered gut physiology, e.g., worsened with eating or stress and relieved by defecation. For IBS-D patients, treatments include gut-acting pharmacologic agents such as antispasmodics, antidiarrheals, and bile acid binders **(Table 338-4)**. In IBS-C

TABLE 338-3 Spectrum of Severity in IBS

	MILD	MODERATE	SEVERE
Clinical Features			
Prevalence	70%	25%	5%
Correlations with gut physiology	+++	++	+
Symptoms constant	0	+	+++
Psychosocial difficulties	0	+	+++
Health care issues	+	++	+++
Practice type	Primary	Specialty	Referral

TABLE 338-4 Possible Drugs for a Dominant Symptom in IBS

SYMPTOM	DRUG	DOSE
Diarrhea	Loperamide	2–4 mg when necessary/maximum 12 g/d
	Diphenoxylate hydrochloride and atropine sulfate (Lomotil)	1–2 tabs as needed or daily (can be taken up to 4 times daily)
	Cholestyramine resin	4 g with meals, can be started qd and increased to tid
	Eluxadoline	100 mg bid
	Alosetron[a]	0.5–1 mg bid (for severe IBS, women)
Constipation	Psyllium husk	3–4 g bid with meals, then adjust
	Methylcellulose	2 g bid with meals, then adjust
	Calcium polycarbophil	1 g qd to qid
	Lactulose syrup	10–20 g bid
	70% sorbitol	15 mL bid
	Polyethylene glycol 3350	17 g in 250 mL water qd, can be titrated to tid
	Lubiprostone	8 or 24 μg bid
	Magnesium hydroxide	15–60 mL qd
	Linaclotide	72, 145, and 290 μg qd
	Plecanatide	3 mg qd
	Tenapanor	50 mg bid
Abdominal pain	Smooth-muscle relaxant	qd to qid ac
	Tricyclic antidepressants	Start 25–50 mg hs, then adjust
	Low FODMAP diet	
Gas and bloating	Low FODMAP diet	
	Probiotics	qd
	Rifaximin	550 mg bid

[a] Available only in the United States.

Abbreviation: FODMAP, fermentable oligosaccharides, disaccharides, monosaccharides, and polyols; IBS, irritable bowel syndrome.

Source: Reproduced with permission from GF Longstreth et al: Functional bowel disorders. Gastroenterology 130:1480, 2006.

patients, increased fiber intake and the use of osmotic agents such as PEG may achieve satisfactory results. For patients with more severe constipation, a chloride channel opener (lubiprostone) or GC-C agonist (linaclotide or plecanatide) or NHE3 inhibitor (tenapanor) may be considered. For IBS patients with predominant gas and bloating, a low FODMAP diet may provide significant relief. Some patients may benefit from probiotics and rifaximin treatment. A small proportion of IBS patients have severe and refractory symptoms, are usually seen in referral centers, and frequently have constant pain and psychosocial difficulties. This group of patients is best managed with neuromodulators and other psychological treatments (Table 338-4). Clinical trials demonstrating success of a low FODMAP diet in improving IBS symptoms and quality of life provide strong evidence supporting the use of this dietary approach in the treatment of IBS. These observations, if confirmed, may lead to the use of the low FODMAP diet as the first line of treatment of IBS patients with moderate to severe symptoms.

FURTHER READING

Chang L et al: AGA Clinical Practice Guideline on the pharmacological management of irritable bowel syndrome with constipation. Gastroenterology 163:118, 2022.

Dionne J et al: A systematic review and meta-analysis evaluating the efficacy of a gluten-free diet and a low FODMAP diet in treating symptoms of irritable bowel syndrome. Am J Gastroenterol 113:1290, 2018.

Drossman DA: Functional gastrointestinal disorders: History, pathophysiology, clinical features, and Rome IV. Gastroenterology 150:1262, 2016.

Lembo A et al: AGA Clinical Practice Guideline on the pharmacological management of irritable bowel syndrome with diarrhea. Gastroenterology 163:137, 2022.

Mayer EA et al: Brain-gut microbiome interactions and functional bowel disorders. Gastroenterology 146:1500; 2014.

Pittayanon R et al: Gut microbiota in patients with irritable bowel syndrome: A systematic review. Gastroenterology 157:97, 2019.

Zhou SY et al: FODMAP diet modulates visceral nociception by lipopolysaccharide-mediated intestinal barrier dysfunction and intestinal inflammation. J Clin Invest 128:267, 2018.

339

Diverticular Disease and Common Anorectal Disorders

Susan L. Gearhart

DIVERTICULAR DISEASE

Incidence and Epidemiology In the United States, diverticulosis affects one-half of the population aged >60 years, and the majority of affected individuals will have no associated symptoms. However, studies have shown that ~5% of individuals with diverticulosis will develop acute diverticular disease. In addition, 10–25% of individuals with diverticular disease will experience recurrent symptoms, and up to 10% will develop complications leading to surgery. Diverticular disease has become the fifth most costly gastrointestinal disorder in the United States and is the leading indication for elective colon resection. The incidence of diverticular disease is on the rise and most prevalent among middle-aged individuals. The majority of patients with diverticular disease report a lower health-related quality of life and more depression as compared to matched controls, thus adding to health care costs. Formerly, diverticular disease was confined to developed countries; however, with the adoption of westernized diets in developing countries, diverticulosis is on the rise across the globe. Immigrants to the United States develop diverticular disease at the same rate as U.S. natives. Although the prevalence among females and males is similar, males tend to present at a younger age. Known risk factors for the development of diverticular disease include the use of nonsteroidal anti-inflammatory drugs (NSAIDS), aspirin, steroids, opioids, smoking, and sedentary lifestyle.

Anatomy and Pathophysiology Two types of diverticula occur in the intestine: true and false (or pseudodiverticula). A true diverticulum is a saclike herniation of the entire bowel wall, whereas a pseudodiverticulum involves only a protrusion of the mucosa and submucosa through the muscularis propria of the colon **(Fig. 339-1)**. The type of diverticulum most commonly seen in the colon is the pseudodiverticulum. The diverticula occur at the point where the nutrient artery, or vasa recta, penetrates through the muscularis propria, resulting in a break in the integrity of the colonic wall. Diverticula are most commonly encountered in the sigmoid colon. This anatomic restriction may be a result of the relative high-pressure zone within the muscular sigmoid colon. Higher-amplitude contractions combined with constipated, high-fat-content stool within the sigmoid lumen in an area of weakness in the colonic wall result in the creation of these diverticula.

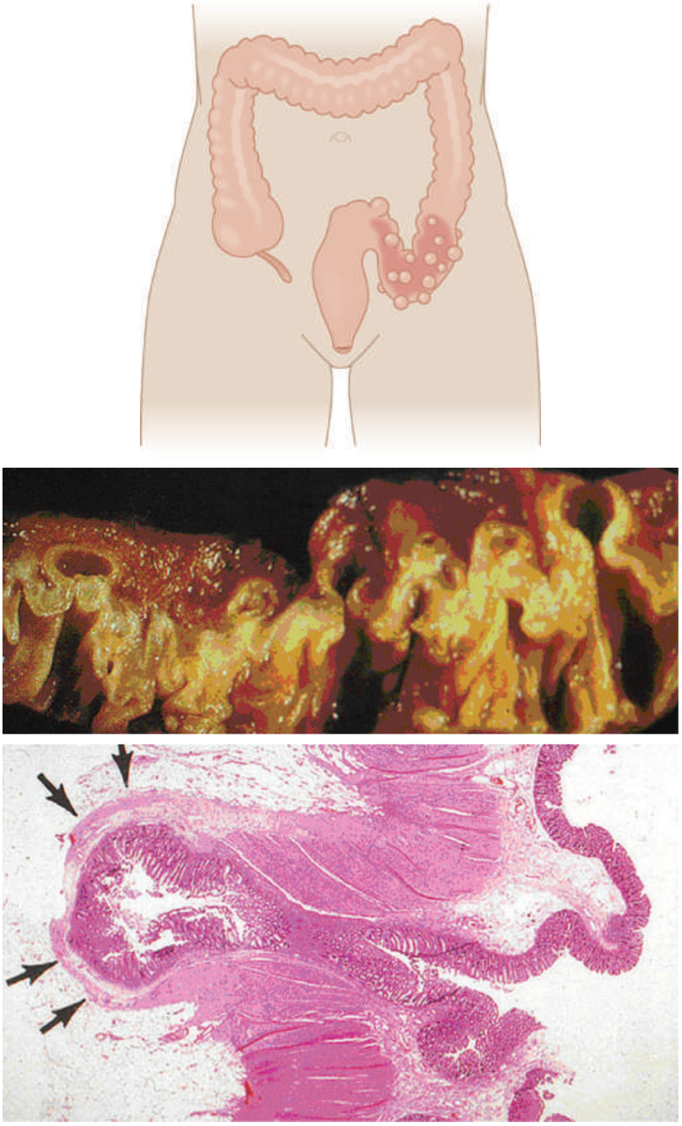

FIGURE 339-1 Gross and microscopic view of sigmoid diverticular disease. *Arrows* mark an inflamed diverticulum with the diverticular wall made up only of mucosa.

Consequently, the vasa recta is either compressed or eroded, leading to either perforation or bleeding.

As mentioned above, diverticula commonly affect the left and sigmoid colon; the rectum is always spared. However, in Asian populations, 70% of diverticula are seen in the right colon and cecum. *Diverticulitis* is inflammation of a diverticulum. Previous understanding of the pathogenesis of diverticulosis attributed the disease to poor dietary choices, and the onset of diverticulitis would occur acutely when these diverticula become obstructed. However, evidence now suggests that the pathogenesis is more complex and multifactorial. Better understanding of the gut microbiota suggests that dysbiosis is an important aspect of disease. Chronic low-grade inflammation is thought to play a key role in neuronal degeneration, leading to dysmotility and high intraluminal pressure. As a consequence, pockets or outpouchings develop in the colonic wall where it is weakest. Studies have also shown that it is more common for patients with diverticular disease to have abnormal collagen cross-linking. Researchers hypothesize that abnormal collagen cross-linking leads to loss of intestinal compliance and, therefore, higher intraluminal pressure, leading to pockets or outpouchings.

A newer approach to examine causality in diseases such as diverticulitis is to examine genome-wide associations (GWAs) using a variety of biobanks. Biobanks from Iceland, Denmark, and the United Kingdom have been examined for GWAs in diverticular disease and identified *LAMB4* and *TNFSF15* variants to be associate with early-onset, severe diverticular disease in otherwise healthy families. Although the exact function these genes play in the development of diverticular disease is unknown, the protein encoded by the gene *LAMB4* is part of the extracellular matrix lamin family, whereas the protein encoded by the gene *TNFSF15* is a member of the tumor necrosis family, which has implications in inflammation.

Presentation, Evaluation, and Management of Diverticular Bleeding Hemorrhage from a colonic diverticulum is the most common cause of hematochezia in patients >60 years, yet only 20% of patients with diverticulosis will have gastrointestinal bleeding. Patients at increased risk for bleeding tend to be hypertensive, have atherosclerosis, and regularly use anticoagulants and NSAIDs. Additional risk factors include obesity and a history of diabetes mellitus. Of note, due to increased use of anticoagulants in our aging population, there has been a rise in the incidence of diverticular bleeding. Most bleeds are self-limited and stop spontaneously with bowel rest. The lifetime risk of rebleeding is 25%.

Initial localization of diverticular bleeding may include colonoscopy, multiplanar computed tomography (CT) angiogram, or nuclear medicine tagged red cell scan. If the patient is stable, ongoing bleeding is best managed by angiography. If mesenteric angiography can localize the bleeding site, the vessel can be occluded successfully with a coil in 80% of cases. The patient can then be followed closely with repetitive colonoscopy, if necessary, looking for evidence of colonic ischemia. However, with highly selective coil embolization, the rate of colonic ischemia is <10%, and the risk of acute rebleeding is <25%. Long-term results (40 months) indicate that >50% of patients with acute diverticular bleeds treated with highly selective angiography have had definitive treatment. Alternatively, colonoscopic ligation with banding, placement of a detachable snare, and over-the-scope hemoclip have been shown to be effective methods to obtain hemostasis if the bleeding site can be localized. These approaches have been shown to prevent rebleeding or the requirement of emergent surgery. In the event that these measures fail to achieve hemostasis, a segmental resection of the colon may be undertaken. This may be advantageous in patients on chronic anticoagulation and immunosuppression as delayed bleeding and perforation have been reported in this subpopulation.

If the patient is refractory to angiographic or endoscopic treatments, unstable, or has required a large-volume transfusion, current recommendations are that surgery should be performed. If the bleeding has been localized, a segmental resection can be performed. If the site of bleeding has not been definitively identified, a subtotal colectomy may be required. In patients without severe comorbidities, surgical resection can be performed with a primary anastomosis. A higher anastomotic leak rate has been reported in patients who received >10 units of blood.

TABLE 339-1 Presentation of Diverticular Disease

Uncomplicated Diverticular Disease—75%
Abdominal pain
Fever
Leukocytosis
Anorexia/obstipation
Complicated Diverticular Disease—25%
Abscess 16%
Perforation 10%
Stricture 5%
Fistula 2%

Presentation, Evaluation, and Staging of Diverticulitis Acute uncomplicated diverticulitis characteristically presents with fever, anorexia, left lower quadrant abdominal pain, and obstipation (Table 339-1). The diagnosis of diverticulitis is best made on a contrast-enhanced abdominal and pelvic CT scan demonstrating the following findings: sigmoid diverticula, thickened colonic wall >4 mm, and inflammation within the pericolic fat without the collection of contrast material or fluid. Additional syndromes have been identified that relate to the presence of diverticulosis or diverticular disease. Symptomatic colitis-associated diverticulosis (SCAD) is uncommon (<1% of patients with diverticulosis) and occurs when inflammation is identified between the diverticula on endoscopic assessment. Symptoms of SCAD mimic irritable bowel disease. Symptomatic uncomplicated diverticular disease (SUDD) is a condition seen in patients with known diverticular disease and ongoing abdominal pain without evidence of overt inflammation on radiographic imaging.

Complicated diverticular disease is defined as diverticular disease associated with an abscess or perforation and less commonly with a fistula (Table 339-1). Symptoms of complicated diverticular disease may be similar to uncomplicated disease, or patients may exhibit signs of peritonitis indicating the presence of a diverticular perforation. If a pericolonic abscess has formed, the patient may have abdominal distention and signs of localized peritonitis. Laboratory investigations often demonstrate a leukocytosis. Rarely, a patient may present with an air-fluid level in the left lower quadrant on plain abdominal film. This should raise a concern for a giant diverticulum of the sigmoid colon and is managed with resection to avoid impending perforation.

Perforated diverticular disease is staged using the Hinchey classification system (Fig. 339-2). This staging system was developed to predict outcomes following the surgical management of complicated diverticular disease. The Hinchey staging system has been modified to include the development of a phlegmon or early abscess (Hinchey stage Ia). In complicated diverticular disease with fistula formation, common locations include cutaneous, vaginal, or vesicle fistulas. These conditions present with either passage of stool through the skin or vagina or the presence of air in the urinary stream (pneumaturia). Colovaginal fistulas are more common in women who have undergone a hysterectomy.

TREATMENT

Medical Management of Diverticular Disease

Asymptomatic diverticular disease discovered on imaging studies or at the time of colonoscopy is best managed by lifestyle changes. Although the data regarding dietary risks and symptomatic diverticular disease are limited (Table 339-2), patients may benefit from a fiber-enriched diet or supplements that include 30 g of fiber each day. The use of fiber decreases colonic transit time and, therefore, prevents increased intraluminal pressure, leading to the

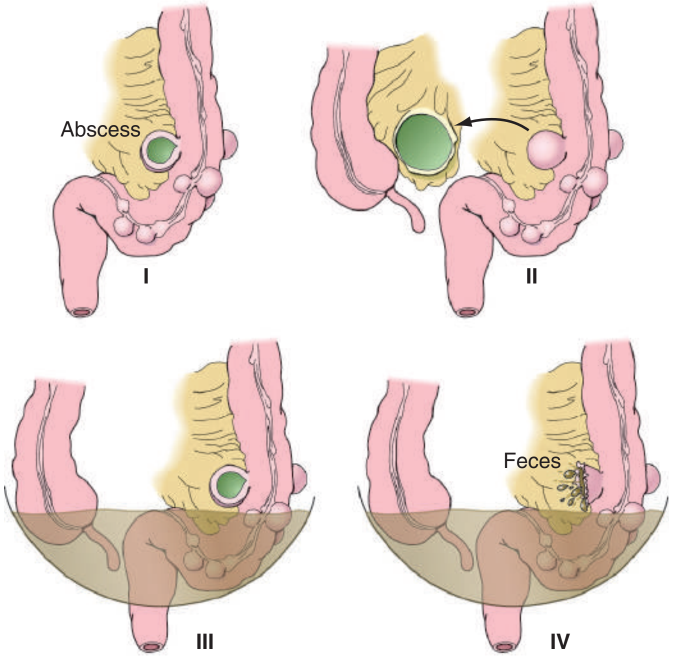

FIGURE 339-2 Hinchey classification of diverticulitis. Stage I: Perforated diverticulitis with a confined paracolic abscess. **Stage II:** Perforated diverticulitis that has closed spontaneously with distant abscess formation. **Stage III:** Noncommunicating perforated diverticulitis with fecal peritonitis (the diverticular neck is closed off, and therefore, contrast will not freely expel on radiographic images). **Stage IV:** Perforation and free communication with the peritoneum, resulting in fecal peritonitis.

development of diverticulosis. The incidence of complicated diverticular disease appears to also be increased in patients who smoke and are obese. Therefore, patients should be encouraged to refrain from smoking and to join a weight loss program. The historical recommendation to avoid eating nuts is based on no more than anecdotal data.

ANTIBIOTICS

The routine use of antibiotics in uncomplicated diverticular disease does not appear to reduce time to symptom resolution or reduce the risk of complications or recurrence. Two large, randomized trials (the AVOD trial and the Diabolo trial) and a large meta-analysis demonstrated that immunocompetent patients with uncomplicated diverticular disease had no difference in time to symptom resolution, recurrence rates, development of complicated diverticular disease, or surgery if treated with or without antibiotics. Currently, patients who are immunocompromised, have findings of extensive inflammation on radiographic studies, are at risk for disease progression, or have computed tomography (CT) findings of complicated diverticular disease should be treated with antibiotics. Known risk factors for disease progression included the American Society of Anesthesiologists (ASA) classification of III or IV, >5 days of symptoms, and elevation of C-reactive protein (CRP) or white blood cell (WBC) count.

If the use of antimicrobial therapy is desired, the current recommended antimicrobial coverage for uncomplicated acute diverticulitis is a third-generation cephalosporin (or ciprofloxacin if there is a known allergy to cephalosporins) and metronidazole targeting a mixed flora. Alternatively, single-agent therapy with a third-generation penicillin such as IV piperacillin or oral penicillin/clavulanic acid may be effective. The usual course of antibiotics is 5 days. A study compared the use of intravenous (IV) versus oral antibiotics in uncomplicated diverticular disease and noted no difference in recovery time or progression of the disease and recommended that safe home treatment on oral antibiotics after a 6-h observation in the emergency department is reasonable. Exclusion criteria for home treatment included complicated diverticular disease, immunocompromised patient, significant active comorbidities, poor social support, or a decline during observation.

DIET AND OTHER MEDICAL THERAPIES

Patients should remain on a limited diet until their pain resolves. The use of anti-inflammatory medications (mesalazine) in randomized clinical trials has shown them to be beneficial at reducing symptoms and disease recurrence in patients with SUDD. However, when objective signs of inflammation such as CRP and computerized imaging are taken into consideration, no benefit for the use of mesalazine has been shown.

Probiotics are increasingly used by gastroenterologists for multiple bowel disorders and may prevent recurrence of diverticulitis. Specifically, probiotics containing *Lactobacillus acidophilus* and *Bifidobacterium* strains may be beneficial; however, a systematic review was unable to show any benefit to the use of probiotics alone. The addition of fiber or mesalazine with probiotics has shown some promise in maintaining remission. Rifaximin (a poorly absorbed broad-spectrum antibiotic), when compared to fiber alone for the treatment of SUDD, is associated with 30% less frequent recurrent symptoms from uncomplicated diverticular disease.

COLONOSCOPY

Colonoscopy should be considered ~6 weeks after the first episode of uncomplicated diverticular disease and after the development of complicated diverticular disease as the overall prevalence of colon cancer is low in these patients (<2%) but higher incidence has been seen with complicated diverticulitis (6–8%). The parallel epidemiology of colorectal cancer and diverticular disease provides enough concern for an endoscopic evaluation before operative management.

SURGICAL MANAGEMENT OF DIVERTICULAR DISEASE

Preoperative risk factors influencing postoperative mortality rates include higher ASA physical status class **(Table 339-3)** and preexisting organ failure. In patients who are low risk (ASA P1 and P2), surgical therapy can be offered to those who do not rapidly improve on medical therapy. For uncomplicated diverticular disease, medical therapy can be continued beyond two attacks without an increased risk of perforation requiring a colostomy. However,

TABLE 339-2 The Use of Fiber in the Management of Diverticular Disease (DD)

JOURNAL, STUDY YEAR	PATIENTS (N)	INTERVENTION	STUDY LENGTH	FINDINGS
Lancet, 1977	18	Wheat or bran crisp bread	3 months	Significant reduction of symptoms score
BMJ, 1981	58	Bran, ispaghula, placebo	16 weeks	No difference
J Gastroenterol, 1977	30	Methylcellulose	3 months	Significant reduction in symptoms
BMJ, 2011	47,033	Vegetarian vs nonvegetarian	11.6 years	Vegetarians had a 31% lower risk of DD
Gastroenterology, 2012	2104	Fiber consumption	12 years	Fiber associated with great risk of DD
JAMA, 2008	47,288	Nut, corn, popcorn consumption	18 years	Higher nut, corn, and popcorn had lower risk of recurrence
Ann R Coll Surg Engl, 1985	56	Fiber consumption	66 months	Higher fiber associated with 19% reduction in symptom recurrence

Source: Modified from A Turis et al: Review article: The pathophysiology and medical management of diverticulosis and diverticular disease of the colon. Aliment Parmacol Ther 42:664, 2015.

TABLE 339-3 American Society of Anesthesiologists Physical Status Classification System

P1	A normal healthy patient
P2	A patient with mild systemic disease
P3	A patient with severe systemic disease
P4	A patient with severe systemic disease that is a constant threat to life
P5	A moribund patient who is not expected to survive without the operation
P6	A declared brain-dead patient whose organs are being removed for donor purposes

patients on immunosuppressive therapy, in chronic renal failure, or with a collagen-vascular disease have a fivefold greater risk of perforation during recurrent attacks. A multicentered randomized clinical trial (DIRECT trial) comparing surgery with conservative management for recurrent SUDD demonstrated that elective surgical resection was associated with an improved quality of life and was more cost-effective at 5 years following resection as compared to conservative management. Surgical therapy is generally indicated in all low-surgical-risk patients with complicated diverticular disease. A randomized trial (LASER trial) examined the use of laparoscopic surgery for refractory or complicated diverticular disease and demonstrated improved quality of life and a reduction in recurrent symptoms by 50% when compared to conservative treatment.

The goals of surgical management of diverticular disease include controlling sepsis, eliminating complications such as fistula or obstruction, removing the diseased colonic segment, and restoring intestinal continuity. **Table 339-4** lists the operations most commonly indicated based on the Hinchey classification and the predicted postoperative outcomes. The current options for uncomplicated diverticular disease include an open or a minimally invasive resection of the diseased area with reanastomosis to the rectosigmoid. Preservation of portions of the sigmoid colon may lead to early recurrence of the disease. The benefits of minimally invasive resection over open surgical techniques include early discharge (by at least 1 day), less narcotic use, less postoperative complications, and an earlier return to work.

The options for the surgical management of complicated diverticular disease **(Fig. 339-3)** include the following open or minimally invasive procedures: (1) proximal diversion of the fecal stream with an ileostomy or colostomy and sutured omental patch with drainage, (2) resection with colostomy and mucous fistula or closure of distal bowel with formation of a Hartmann's pouch (Hartmann's procedure), (3) resection with anastomosis (coloproctostomy), or (4) resection with anastomosis and diversion (coloproctostomy with loop ileostomy or colostomy). (5) Laparoscopic technique of washout and drainage without diversion has been described for Hinchey III patients; however, a threefold increased risk of recurrent peritonitis requiring reoperation with washout alone has been reported. Robotic surgery resection for complicated diverticular disease is associated with a lower rate of conversion to an open procedure.

Patients with Hinchey stage Ia may be managed with antibiotic therapy only or followed by resection with anastomosis following further workup and symptom resolution. Patients with Hinchey stages Ib and II disease are managed with percutaneous drainage followed by resection with anastomosis following further evaluation and symptom resolution. Current guidelines put forth by the American Society of Colon and Rectal Surgeons suggest, in addition to antibiotic therapy, CT-guided percutaneous drainage of diverticular abscesses that are >3 cm and have a well-defined wall. Abscesses that are <5 cm may resolve with antibiotic therapy alone. Contraindications to percutaneous drainage are no percutaneous access route, pneumoperitoneum, and fecal peritonitis. Drainage of a diverticular abscess is associated with a 20–25% failure rate. Urgent operative intervention is undertaken if percutaneous drainage fails and patients develop generalized peritonitis, and most will need to be managed with a Hartmann's procedure (resection of the sigmoid colon with end colostomy and stapling of the rectosigmoid distal to the diseased segment). In selected cases, nonoperative therapy may be considered.

The management of Hinchey stage III disease is under debate. In this population of patients, no fecal peritonitis is present, and it is presumed that the perforation has sealed. Historically, Hinchey stage III has been managed with a Hartmann's procedure or with primary anastomosis and proximal diversion. Several studies have examined short- and long-term outcomes for laparoscopic peritoneal lavage to remove the peritoneal contamination and place drainage catheters should a communication to the bowel still exist. However, this procedure has been associated with an increased risk of requiring reoperation for ongoing peritonitis. Overall, ostomy rates are lower with the use of laparoscopic peritoneal lavage. No anastomosis of any type should be attempted in Hinchey stage IV disease or in the presence of fecal peritonitis. A limited approach to these patients is associated with a decreased mortality rate.

Recurrent Symptoms Recurrent abdominal symptoms following surgical resection for diverticular disease occur in 10% of patients. Recurrent diverticular disease develops in patients following inadequate surgical resection. A retained segment of diseased rectosigmoid colon is associated with twice the incidence of recurrence. The presence of irritable bowel syndrome may also cause recurrence of initial symptoms. Patients undergoing surgical resection for presumed diverticulitis and symptoms of chronic abdominal cramping and irregular loose bowel movements consistent with irritable bowel syndrome have poorer functional outcomes.

TABLE 339-4 Outcome Following Surgical Therapy for Complicated Diverticular Disease Based on Modified Hinchey Staging

HINCHEY STAGE	OPERATIVE PROCEDURE	ANASTOMOTIC LEAK RATE, %	OVERALL MORBIDITY RATE, %
Ia (pericolic phlegmon)	Laparoscopic or open colon resection	3	15
Ib (pericolic abscess)	Percutaneous drainage followed by laparoscopic or open colon resection	3	15
II	Percutaneous drainage followed by laparoscopic or open colon resection +/– proximal diversion with an ostomy	3	15
III	Laparoscopic washout and drainage *or* Laparoscopic or open resection with proximal diversion (ostomy) *or* Hartmann's procedure	3	30% risk of peritonitis requiring reoperation if no resection is performed. Overall morbidity 50% Overall mortality 15%
IV	Hartmann's procedure *or* Washout with proximal diversion	—	Overall morbidity 50% Overall mortality 15%

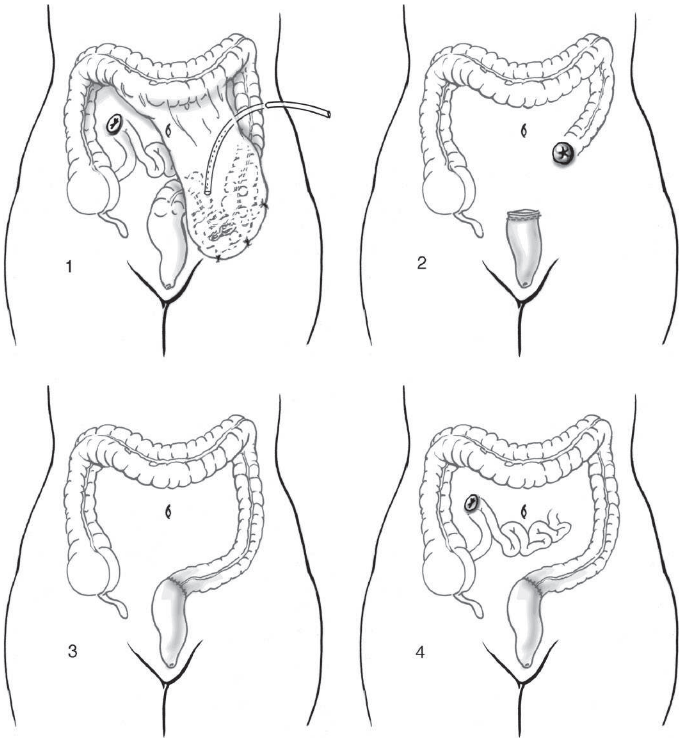

FIGURE 339-3 Methods of surgical management of complicated diverticular disease. ***1.*** Drainage, omental pedicle graft, and proximal diversion. ***2.*** Hartmann's procedure. ***3.*** Sigmoid resection with coloproctostomy. ***4.*** Sigmoid resection with coloproctostomy and proximal diversion.

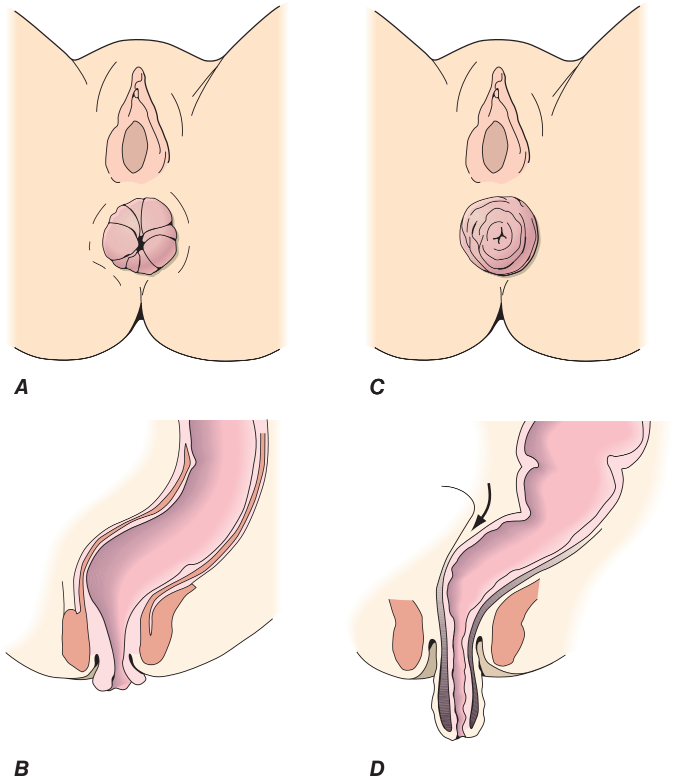

FIGURE 339-4 Degrees of rectal prolapse. Mucosal prolapse only (***A, B,*** sagittal view). Full-thickness prolapse associated with redundant rectosigmoid and deep pouch of Douglas (***C, D,*** sagittal view).

COMMON DISEASES OF THE ANORECTUM

RECTAL PROLAPSE (PROCIDENTIA)

Incidence and Epidemiology Rectal prolapse is six times more common in women than in men. The incidence of rectal prolapse peaks in women >60 years. Women with rectal prolapse have a higher incidence of associated pelvic floor disorders including urinary incontinence, rectocele, cystocele, and enterocele. About 20% of children with rectal prolapse will have cystic fibrosis. All children presenting with prolapse should undergo a sweat chloride test. Less common associations include Ehlers-Danlos syndrome, solitary rectal ulcer syndrome, congenital hypothyroidism, Hirschsprung's disease, dementia, cognitively impaired, and schizophrenia.

Anatomy and Pathophysiology Rectal prolapse (procidentia) is a circumferential, full-thickness protrusion of the rectal wall through the anal orifice. It is often associated with a redundant sigmoid colon, pelvic laxity, and a deep rectovaginal septum (pouch of Douglas). Initially, rectal prolapse was believed to be the result of early internal rectal intussusception, which occurs in the upper to mid rectum. This was considered to be the first step in an inevitable progression to full-thickness external prolapse. However, only 1 of 38 patients with internal prolapse followed for >5 years developed full-thickness prolapse. Others have suggested that full-thickness prolapse is the result of damage to the nerve supply to the pelvic floor muscles or pudendal nerves from repeated stretching with straining to defecate. Damage to the pudendal nerves would weaken the pelvic floor muscles, including the external anal sphincter muscles. Bilateral pudendal nerve injury is identified significantly more with full-thickness prolapse and fecal incontinence than unilateral injury.

Presentation and Evaluation In external prolapse, patient complaints include a palpable anal mass, bleeding per rectum, leakage of blood and mucus, and poor perianal hygiene. Prolapse of the rectum usually occurs following defecation and will spontaneously reduce or require the patient to manually reduce the prolapse. Constipation occurs in ~30–67% of patients with rectal prolapse. Differing degrees of fecal incontinence occur in 50–70% of patients. Patients with internal rectal prolapse will present with symptoms of both constipation and incontinence. Other associated findings include outlet obstruction (anismus) in 30%, colonic inertia in 10%, and solitary rectal ulcer syndrome in 12%.

Office evaluation is best performed after the patient has been given an enema, which enables the prolapse to protrude. An important distinction should be made between full-thickness rectal prolapse and isolated mucosal prolapse associated with hemorrhoidal disease **(Fig. 339-4)**. Mucosal prolapse is known for radial grooves rather than circumferential folds around the anus and is due to increased laxity of the connective tissue between the submucosa and underlying muscle of the anal canal. The evaluation of prolapse should also include cystoproctography and colonoscopy. These examinations evaluate for associated pelvic floor disorders and rule out a malignancy or a polyp as the lead point for prolapse.

TREATMENT

Rectal Prolapse

The medical approach to the management of rectal prolapse is limited and includes stool-bulking agents or fiber supplementation to ease the process of evacuation. Surgical correction of rectal prolapse is the mainstay of therapy. Previously, the presence of internal rectal prolapse identified on imaging studies has been considered a nonsurgical disorder, and biofeedback was recommended. However, only one-third of patients will have successful resolution of symptoms from biofeedback. Two approaches are commonly considered: transabdominal and transperineal. Transabdominal approaches have been associated with lower recurrence rates, but some patients with significant comorbidities are better served by a transperineal approach.

Common transperineal approaches include a transanal proctectomy (Altmeier procedure), mucosal proctectomy (Delorme procedure), or placement of a Tirsch wire encircling the anus. The goal of the transperineal approach is to remove the redundant rectosigmoid colon. Common transabdominal approaches include presacral suture or mesh rectopexy (Ripstein) with (Frykman-Goldberg) or without resection of the redundant sigmoid. Colon resection, in general, is reserved for patients with constipation and outlet obstruction. Ventral rectopexy is an effective method of abdominal repair of internal and full-thickness prolapse that does not require sigmoid resection. This repair has been shown to have improved pelvic floor functional results over other abdominal repairs. Transabdominal procedures can be performed effectively with laparoscopic and robotic techniques. Short- and long-term recurrence rates are low (<10%), and symptoms improved in more than three-fourths of patients.

FECAL INCONTINENCE

Incidence and Epidemiology Fecal incontinence is the involuntary passage of fecal material or the inability to control the initiation of defecation. The prevalence of fecal incontinence in adults in the United States approaches 15% and is expected to increase given our aging population. A higher incidence of incontinence is seen among older parous women. One-half of patients with fecal incontinence also suffer from urinary incontinence. The cause of fecal incontinence is often multifactorial; however, the majority of women with fecal incontinence are parous and may have experienced obstetrical injury to the pelvic floor, either while carrying a fetus or during the delivery. An anatomic sphincter defect may occur in up to 32% of women following childbirth regardless of visible damage to the perineum. Risk factors at the time of delivery include prolonged labor, the use of forceps, and the need for an episiotomy. Symptoms of incontinence can present two or more decades after obstetric injury. Medical conditions known to contribute to the development of fecal incontinence are listed in Table 339-5.

Anatomy and Pathophysiology The anal sphincter complex is made up of the internal and external anal sphincter. The internal sphincter is smooth muscle and a continuation of the circular fibers of the rectal wall. It is innervated by the intestinal myenteric plexus and is therefore not under voluntary control. The external anal sphincter is formed in continuation with the levator ani muscles and is under voluntary control. The pudendal nerve supplies motor innervation to the external anal sphincter. Obstetric injury may result in tearing of the muscle fibers anteriorly at the time of the delivery. This results in an obvious anterior defect on endoanal ultrasound. Injury may also be the result of stretching of the pudendal nerves during pregnancy or delivery of the fetus through the birth canal.

Presentation and Evaluation Patients may suffer with varying degrees of fecal incontinence. Fecal incontinence is classified into three categories based on the clinical presentation. These categories include passive incontinence (passage of stool without awareness), urge incontinence (leakage of stool despite attempts at holding), and fecal seepage (seepage of residue generally following defecation associated with normal continence). Beyond the immediate problems associated with fecal incontinence, these patients are often withdrawn and suffer from depression. For this reason, quality-of-life measures are an important component in the evaluation of patients with fecal incontinence.

The evaluation of fecal incontinence should include a thorough history and physical examination including digital rectal examination (DRE). Weak sphincter tone on DRE and loss of the "anal wink" reflex (S1-level control) may indicate a neurogenic dysfunction. Perianal scars may represent surgical injury. Other studies helpful in the diagnosis of fecal incontinence include anal manometry, pudendal nerve terminal motor latency (PNTML), and endoanal ultrasound. Centers that care for patients with fecal incontinence will have an anorectal physiology laboratory that uses standardized methods of evaluating anorectal physiology. Pudendal nerve studies evaluate the function of the nerves innervating the anal canal using a finger electrode placed in the anal canal. Stretch injuries to these nerves will result in a delayed response of the sphincter muscle to a stimulus, indicating a prolonged latency. Finally, endoanal ultrasound will evaluate the extent of the injury to the sphincter muscles before surgical repair. Unfortunately, all of these investigations are user-dependent, and very few studies demonstrate that these studies predict outcome following an intervention. Magnetic resonance imaging (MRI) has also been utilized in the evaluation of the sphincter muscle complex in treatment planning for fecal incontinence, but its role has not been well-established.

Rarely does a pelvic floor disorder exist alone. The majority of patients with fecal incontinence will have some degree of urinary incontinence. Similarly, fecal incontinence is a part of the spectrum of pelvic organ prolapse. For this reason, patients may present with symptoms of obstructed defecation as well as fecal incontinence. Careful evaluation including dynamic MRI or cinedefecography should be performed to search for other associated defects. Surgical repair of incontinence without attention to other associated defects may decrease the success of the repair.

TABLE 339-5 Medical Conditions That Contribute to Symptoms of Fecal Incontinence

Neurologic Disorders

- Dementia
- Brain tumor
- Stroke
- Multiple sclerosis
- Tabes dorsalis
- Cauda equina lesions

Skeletal Muscle Disorders

- Myasthenia gravis
- Myopathies, muscular dystrophy

Miscellaneous

- Hypothyroidism
- Irritable bowel syndrome
- Diabetes
- Severe diarrhea
- Scleroderma

TREATMENT

Fecal Incontinence

Medical management of fecal incontinence includes strategies to bulk up the stool, which help in increasing fecal sensation and complete evacuation. Stool bulking agents include fiber supplementation, loperamide, diphenoxylate, and bile acid binders. These agents help to bind the stool, resulting in more complete evacuation and decreasing the frequency of bowel movements. This can be particularly helpful in patients with mild symptoms of fecal incontinence. Patients may be offered a form of physical therapy called biofeedback. This therapy helps strengthen the external sphincter muscle while training the patient to relax with defecation to avoid unnecessary straining and further injury to the sphincter muscles. At a minimum, biofeedback is risk-free and most patients will have some improvement. For this reason, it should be incorporated into the initial recommendation to all patients with fecal incontinence. It is important to note that there is no medical or surgical therapy to reduce incontinence of flatus. Dietary alterations to reduce gas production and the use of probiotics are the only antidotal remedies.

Historically, the "gold standard" for the treatment of fecal incontinence with an isolated sphincter defect has been the overlapping sphincteroplasty. The external anal sphincter muscle and scar tissue, as well as any identifiable internal sphincter muscle, are dissected free from the surrounding adipose and connective tissue. An overlapping sphincter repair is performed in an attempt to rebuild the muscular ring and restore its function. However, long-term results following overlapping sphincteroplasty have been poor, with a 50% failure rate over 5 years.

Alternative therapies for the treatment of fecal incontinence included sacral nerve modulation, collagen-enhancing injectables, and anal implantable self-expanding prostheses (THD Gatekeeper and Sphinkeeper). Sacral nerve stimulation (SNS) was the first implantable pacemaker and is a U.S. Food and Drug Administration (FDA)-approved adaptation of a procedure developed for the management of urinary incontinence. SNS is indicated in patients with two or more episodes of frank fecal incontinence per week. Long-term results for SNS have been promising, with nearly 80% of patients having a reduction in incontinence episodes by at least 50%. There are inherent challenges with the SNS device, and to avoid these challenges, percutaneous or transcutaneous tibial nerve stimulation (PTNS or TTNS) is being offered to patients with fecal incontinence. PTNS is offered in a clinical setting, while TTNS can be performed at home. Both require weekly sessions of stimulation of the tibial nerve, which serves to feedback to the third sacral nerve root. Collagen-enhancing injectables and anal self-expanding implants have been around for several years. Results from randomized trials have been variable but support the use of injectables/implantables in patients who are not candidates for sacral nerve modulation.

Finally, the use of stem cells to increase the bulk of the sphincter muscles is currently being tested. Stem cells can be harvested from the patient's own muscle, grown, and then implanted into their sphincter complex. Concern for cost and the need for an additional procedure have dampened enthusiasm.

MIXED CONSTIPATION AND FECAL INCONTINENCE

It is estimated that up to 20% of patients presenting for evaluation of fecal incontinence have mixed constipation and fecal incontinence. Fecal incontinence, in this instance, is related to the large number of laxatives required to produce stool evacuation. This condition is more difficult to manage and often requires close collaboration with different specialties. One promising intervention is the use of a transanal irrigation system (Peristeen by Coloplast). This system provides daily irrigations of the left side of the colon to assist with evacuations and has shown promising results. The daily assistance with evacuation leads to less need for oral laxatives and reduced episodes of incontinence. One drawback is long-term adherence to the irrigations.

HEMORRHOIDAL DISEASE

Incidence and Epidemiology Symptomatic hemorrhoids affect >1 million individuals in the Western world per year. Hemorrhoidal disease can occur at any age and affects females and males at similar rates. The prevalence of hemorrhoidal disease is less in developing countries. The typical low-fiber, high-fat diet prominent in developed countries is associated with constipation and straining and the development of symptomatic hemorrhoids. Other risk factors for hemorrhoids include pregnancy and high-impact activities (bike riding).

Anatomy and Pathophysiology Hemorrhoidal cushions are a normal part of the anal canal. The vascular structures contained within this tissue aid in continence by preventing damage to the sphincter muscle. Three main hemorrhoidal complexes traversing the anal canal include the left lateral, the right anterior, and the right posterior. Prolonged engorgement with straining leads to prolapse of this tissue into the anal canal. Over time, the anatomic support system of the hemorrhoidal complex weakens, exposing this tissue to the outside of the anal canal where it is susceptible to injury. Hemorrhoids are commonly classified as external or internal. External hemorrhoids originate below the dentate line and are covered with squamous epithelium and are not associated with an internal component. Internal hemorrhoids originate above the dentate line and are covered with mucosa and transitional zone epithelium and represent the majority of hemorrhoids.

Presentation and Evaluation Patients commonly present to a physician for two reasons: bleeding and protrusion. Pain is less common than with fissures and, if present, is described as a dull ache from engorgement of the hemorrhoidal tissue. Severe pain may indicate a thrombosed external hemorrhoid. Hemorrhoidal bleeding is described as painless bright red blood seen either in the toilet or upon wiping. Occasional patients can present with significant bleeding, which may be a cause of anemia; however, the presence of a colonic neoplasm must be ruled out in anemic patients, especially given the noted rise in young patients with colorectal cancer. Patients who present with protruding tissue complain about inability to maintain perianal hygiene and anal itching (pruritis) and are often concerned about the presence of a malignancy.

The diagnosis of hemorrhoidal disease is made on physical examination. Inspection of the perianal region for evidence of thrombosis or excoriation is performed, followed by a careful digital examination. Anoscopy is performed paying particular attention to the known position of hemorrhoidal disease. The patient is asked to strain. It is important to differentiate the circumferential appearance of a full-thickness rectal prolapse from the radial nature of prolapsing hemorrhoids (see "Rectal Prolapse," above). The stage and location of the hemorrhoidal complexes are defined.

TREATMENT

Hemorrhoidal Disease

The treatment for bleeding hemorrhoids is based on the stage of the disease **(Table 339-6)**. In all patients with bleeding, the possibility of other causes must be considered, and further studies should include endoscopic assessments.

With rare exceptions, the acutely thrombosed hemorrhoid can be excised within the first 48 h by performing an elliptical excision and clot extraction. This can often be performed at the bedside with local anesthesia; sitz baths, fiber, and stool softeners are prescribed. Additional nonoperative therapies for bleeding hemorrhoids include rubber band ligation, infrared coagulation, and sclerotherapy. Sensation begins at the dentate line; therefore, all procedures can be performed without discomfort either endoscopically or in the office. Bands are placed around the engorged tissue, causing ischemia and fibrosis. This aids in fixing the tissue proximally in the anal canal. Patients may complain of a dull ache for 24 h following band application. During sclerotherapy, 1–2 mL of a sclerosant (usually sodium tetradecyl sulfate) is injected using a 25-gauge needle into the submucosa of the hemorrhoidal complex. Care must be taken not to inject the anal canal circumferentially, or stenosis may occur.

For surgical management of hemorrhoidal disease, excisional hemorrhoidectomy with sharp dissection (Milligan-Morgan hemorrhoidectomy), bipolar electrocautery ligation (LigaSure), transhemorrhoidal dearterialization (THD), and stapled hemorrhoidectomy ("the procedure for prolapse or hemorrhoids" [PPH]) are the procedures of choice. All surgical methods of management

TABLE 339-6 The Staging and Treatment of Hemorrhoids

STAGE	DESCRIPTION OF CLASSIFICATION	TREATMENT
I	Enlargement with bleeding	Fiber supplementation Short course of cortisone suppository Sclerotherapy Infrared coagulation
II	Protrusion with spontaneous reduction	Fiber supplementation Short course of cortisone suppository Sclerotherapy Infrared coagulation
III	Protrusion requiring manual reduction	Fiber supplementation Short course of cortisone suppository Rubber band ligation Operative hemorrhoidectomy
IV	Irreducible protrusion	Fiber supplementation Cortisone suppository Operative hemorrhoidectomy

are equally effective in the treatment of symptomatic third- and fourth-degree hemorrhoids. Current postprocedure outcomes following hemorrhoidectomy target improvements in bleeding, pain, and recurrence rates. The sutured hemorrhoidectomy involves the removal of redundant tissue down to the anal verge, and unpleasant anal skin tags are removed as well. However, pain and postoperative bleeding are more common. The stapled hemorrhoidectomy is associated with less discomfort; however, this procedure does not remove anal skin tags, and an increased number of complications are associated with use of the stapling device. THD uses ultrasound guidance to ligate the blood supply to the anal tissue, hence reducing hemorrhoidal engorgement. Ligation with electrocautery uses a bipolar device (LigaSure) to remove the unwanted prolapsed tissue. Both THD and ligation with electrocautery have demonstrated similar short-term and long-term results; however, ligation with electrocautery requires less operative time. No procedures on hemorrhoids should be done in patients who are immunocompromised or who have active proctitis. Emergent hemorrhoidectomy for bleeding hemorrhoids is associated with a higher complication rate.

Acute complications associated with the treatment of hemorrhoids include pain, infection, recurrent bleeding, and urinary retention. Care should be taken to place bands properly and to avoid overhydration in patients undergoing operative hemorrhoidectomy. Late complications include fecal incontinence as a result of injury to the sphincter during the dissection. Anal stenosis may develop from overzealous excision, with loss of mucosal skin bridges for reepithelialization. Finally, an *ectropion* (prolapse of rectal mucosa from the anal canal) may develop. Patients with an ectropion complain of a "wet" anus as a result of inability to prevent soiling once the rectal mucosa is exposed below the dentate line.

ANORECTAL ABSCESS

Incidence and Epidemiology The development of a perianal abscess is more common in men than women by a ratio of 3:1. The peak incidence is in the third to fifth decade of life. Perianal pain associated with the presence of an abscess accounts for 15% of office visits to a colorectal surgeon. The disease is more prevalent in immunocompromised patients such as those with diabetes, hematologic disorders, or inflammatory bowel disease (IBD) and persons who are HIV positive. These disorders should be considered in patients with recurrent perianal infections.

Anatomy and Pathophysiology An anorectal abscess is an abnormal fluid-containing cavity in the anorectal region. Anorectal abscess results from an infection involving the glands surrounding the anal canal. Normally, these glands release mucus into the anal canal, which aids in defecation. When stool accidentally enters the anal glands, the glands become infected, and an abscess develops. Anorectal abscesses are perianal in 40–50% of patients, ischiorectal in 20–25%, intersphincteric in 2–5%, and supralevator in 2.5% **(Fig. 339-5)**.

Presentation and Evaluation Perianal pain and fever are the hallmarks of an abscess. Patients may have difficulty voiding and have blood in the stool. On physical examination, a large fluctuant area is usually readily visible. Routine laboratory evaluation shows an elevated WBC count. Diagnostic procedures are rarely necessary unless evaluating a recurrent abscess. A CT scan or MRI has an accuracy of 80% in determining incomplete drainage. If there is a concern about the presence of IBD, a rigid or flexible sigmoidoscopic examination may be done at the time of drainage to evaluate for inflammation within the rectosigmoid region. A more complete evaluation for Crohn's disease would include a full colonoscopy and small-bowel enterography.

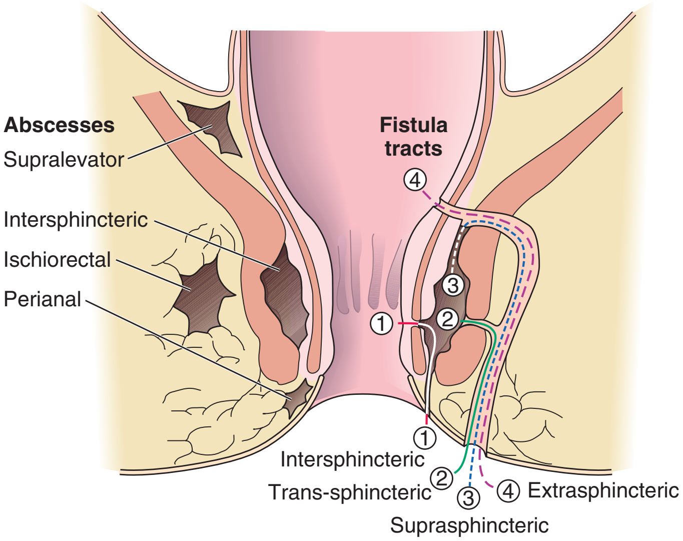

FIGURE 339-5 Common locations of anorectal abscess (*left*) and fistula in ano (*right*).

TREATMENT

Anorectal Abscess

As with all abscesses, the "gold standard" is drainage. Office drainage of an uncomplicated anorectal abscess may suffice. A small incision close to the anal verge is made, and a Mallenkot drain is advanced into the abscess cavity. The Mallenkot catheter (a mushroom-shaped catheter with grooves) can be maintained by the patient for up to 1 week before being removed to avoid tissue ingrowth. For patients with a complicated abscess or who are diabetic or immunocompromised, drainage should be performed in an operating room under anesthesia. These patients are at greater risk for developing necrotizing fasciitis. The role of antibiotics in the management of anorectal abscesses is limited. Antibiotics are only warranted in patients who are immunocompromised or have obvious cellulitis on physical exam.

FISTULA IN ANO

Incidence and Epidemiology The incidence of fistulizing perianal disease parallels the incidence of anorectal abscess and is estimated to be 1 in 10,000 individuals. Some 30–40% of abscesses will give rise to an anal fistula. Although the majority of the fistulas are cryptoglandular in origin, 10% are associated with IBD, tuberculosis, malignancy, and radiation.

Anatomy and Pathophysiology An anal fistula is defined as a communication of an abscess cavity with an identifiable internal opening within the anal canal. This identifiable opening is commonly located at the dentate line where the anal glands enter the anal canal. Patients experiencing continuous drainage at 1 month following the treatment of a perianal abscess likely have an anal fistula. These fistulas are classified by their relationship to the anal sphincter muscles, with 70% being intersphincteric, 23% transsphincteric, 5% suprasphincteric, and 2% extrasphincteric (Fig. 339-5).

Presentation and Evaluation A patient with an anal fistula will complain of constant drainage from the perianal region associated with tenderness. Examination under anesthesia is the best way to evaluate a fistula. However, in cases with a complex fistula, a preoperative MRI will identify tracts with an accuracy of 80%. During surgery, hydrogen peroxide injected through the external opening will aid in identifying the internal opening. Goodsall's rule states that a posterior external fistula will enter the anal canal in the posterior midline, whereas an anterior fistula will enter at the nearest crypt. A fistula exiting >3 cm from the anal verge may have a complicated upward extension and may not obey Goodsall's rule.

TREATMENT

Fistula in Ano

A newly diagnosed draining fistula is best managed with an operative intervention including initial placement of a seton catheter (a vessel loop or silk tie placed through the fistula tract), which maintains the tract patency and reduces the likelihood of recurrent infection and inflammation. Once the inflammation is reduced (4–6 weeks), the exact relationship of the fistula tract to the anal sphincters can be ascertained. Two general approaches are common: sphincter-cutting procedures and sphincter-sparing procedures. A simple fistulotomy or fistulectomy is a sphincter-cutting procedure and is usually performed for intersphincteric and low (less than one-third of the muscle) transsphincteric fistulas. In a systematic review, sphincter-cutting procedures were associated with a 94% success rate and a 13% rate of mild fecal incontinence. Sphincter-sparing surgery is usually performed for a higher transsphincteric fistula. These procedures include an anorectal advancement flap, ligation of the internal fistula tract (LIFT procedure), or insertion of fibrin or cyanoacrylate glue. Sphincter-preserving procedures are associated with up to 78% success rates with no incidence of fecal incontinence.

Fistulizing disease of the anus is common in Crohn's disease, and the use of mesenchymal stem cell therapy may improve healing rates of fistula associated with Crohn's disease. The ADMIRE study examined the use of allogeneic expanded adipose-derived mesenchymal stem cells in the treatment of complex perianal fistula in Crohn's disease. The study included 212 patients randomized to stem cell therapy or placebo. Fistula remission rates at 52 weeks were significantly higher with the use of stem cell therapy over placebo (59 vs 42%, respectively). Further studies are currently being performed to elucidate the benefits of stem cell therapy in perianal fistula. As with any stem cell therapy, the cost can be prohibitive to its standard use.

■ ANAL FISSURE

Incidence and Epidemiology Anal fissures occur at all ages but are more common in the third through the fifth decades. The prevalence is equal in males and females. It is associated with constipation, diarrhea, infectious etiologies, perianal trauma, and Crohn's disease.

Anatomy and Pathophysiology Trauma to the anal canal occurs following defecation. This injury occurs in the anterior or, more commonly, posterior anal canal. Irritation caused by the trauma to the anal canal results in an increased resting pressure of the internal sphincter. The blood supply to the sphincter and anal mucosa enters laterally. Therefore, increased anal sphincter tone results in a relative ischemia in the region of the fissure and leads to poor healing of the anal injury. A fissure that is not in the posterior or anterior position should raise suspicion for other causes, including tuberculosis, syphilis, Crohn's disease, and malignancy.

Presentation and Evaluation A fissure can be easily diagnosed on history alone. The classic complaint is pain, which is strongly associated with defecation and is relentless. The bright red bleeding that can be associated with a fissure is less extensive than that associated with hemorrhoids and is usually only noted on wiping. On examination, most fissures are located in either the posterior or anterior position. A lateral fissure is worrisome because it may have a less benign nature, and systemic disorders should be ruled out. A chronic fissure is indicated by the presence of a hypertrophied anal papilla at the proximal end of the fissure and a sentinel pile or skin tag at the distal end. Often the circular fibers of the hypertrophied internal sphincter are visible within the base of the fissure.

TREATMENT

Anal Fissure

The management of the acute fissure is conservative. Stool softeners for those with constipation, increased dietary fiber, topical anesthetics, glucocorticoids, and sitz baths are prescribed and will heal 60–90% of fissures. Chronic fissures are those present for >6 weeks. These can be treated with modalities aimed at decreasing the anal canal resting pressure including nifedipine ointment applied three times a day and botulinum toxin type A, up to 20 units, injected into the internal sphincter on each side of the fissure. Both treatments are associated with a fissure healing rate of >80%. Surgical management includes anal dilatation and lateral internal sphincterotomy. Usually, one-third of the internal sphincter muscle is divided; it is easily identified because it is hypertrophied. Recurrence rates from medical therapy are higher, but this is offset by a risk of incontinence following sphincterotomy. Lateral internal sphincterotomy may lead to incontinence more commonly in women.

Acknowledgment

The author thanks Cory Sandore for providing some illustrations for this chapter.

■ FURTHER READING

Bharucha AE et al: Surgical interventions and the use of device-aided therapy for the treatment of fecal incontinence and defecatory disorders. Clin Gastroenterol Hepatol 15:1844, 2017.

Daniels L et al: Randomized clinical trial of observation versus antibiotic treatment for a first episode of CT-proven uncomplicated acute diverticulitis (DIABOLO trial). BJS 104:52, 2017.

Guttenplan M: The evaluation and office management of hemorrhoids for the gastroenterologist. Curr Gastroenterol Rep 19:30, 2017.

Hall J et al: The American Society of Colon and Rectal Surgeons Clinical Practice Guidelines for the treatment of left-sided colonic diverticulitis. Dis Colon Rectum 63:728, 2020.

Panes J et al: Long-term efficacy and safety of stem cell therapy (Cx601) for complex perianal fistulas in patients with Crohn's disease. Gastroenterology 154:1334, 2018.

Prichard D, Bharucha AE: Management of pelvic floor disorders: Biofeedback and more. Curr Treat Options Gastroenterol 12:456, 2014.

Salfity HV et al: Minimally invasive incision and drainage technique in the treatment of simple subcutaneous abscess in adults. Am Surg 83:699, 2017.

Strate LL, Morris AM: Epidemiology, pathophysiology, and treatment of diverticulitis. Gastroetenterology 156:1282, 2019.

Sugrue J et al: Sphincter-sparing anal fistula repair: Are we getting better? Dis Colon Rectum 60:1071, 2017.

Tursi A: Dietary pattern and colonic diverticulosis. Curr Opin Clin Nutr Metab Care 20:409, 2017.

340

Mesenteric Vascular Insufficiency

Daniel Willie-Permor, Mahmoud Malas

INTESTINAL ISCHEMIA

■ INCIDENCE AND EPIDEMIOLOGY

Intestinal ischemia occurs when splanchnic perfusion fails to meet the metabolic demands of the intestines, resulting in ischemic tissue injury. Mesenteric ischemia affects 2–3 people per 100,000 with an increasing incidence in the aging population. Mortality with acute presentation remains high (between 50 and 80%), and early diagnosis and prompt intervention are crucial in improving clinical outcomes. Intestinal ischemia is further classified into chronic mesenteric ischemia (CMI) and acute mesenteric ischemia (AMI). CMI is secondary to multiple major visceral arterio-occlusive disease with involvement of the superior mesenteric artery (SMA) most worrisome. AMI is most associated with

(1) arterio-occlusive mesenteric ischemia, (2) nonocclusive mesenteric ischemia, and (3) mesenteric venous thrombosis.

CMI is the failure to achieve normal postprandial hyperemic intestinal blood flow. This occurs due to an imbalance between supply and demand of oxygen metabolites to the intestinal tract similar to cardiac angina. CMI occurs due to significant atherosclerotic disease leading to the narrowing of the SMA and/or celiac artery origins.

AMI is the occurrence of an abrupt cessation of mesenteric blood flow, usually embolic or thrombotic in nature. Approximately 50% of AMI is due to embolus to the mid to distal SMA. The embolus etiology includes atrial fibrillation, recent myocardial infarction, soft atherosclerotic plaque, infective endocarditis, valvular heart disease, and recent cardiac or vascular catheterization. Approximately 25–30% of the cases are characterized by an acute-on-chronic thrombosis in patients with preexisting mesenteric atherosclerosis. Thrombotic occlusion most commonly occurs at areas of severe atherosclerotic narrowing at the SMA and the celiac artery.

Nonocclusive mesenteric ischemia represents 20% of the cases and is secondary to intestinal ischemia when subjected to acute hemodynamic instability. Patients above the age of 50, especially those with coexisting conditions like myocardial infarction, congestive heart failure, aortic insufficiency, and renal or liver disease, who are also undergoing cardiovascular surgery, face the highest risk. Hypovolemia, shock, and use of vasoconstrictive agents (e.g., digoxin, α-adrenergic agonists, cocaine) can precipitate ischemia in these patients. It is the most prevalent gastrointestinal disease complicating cardiovascular surgery. The incidence of ischemic colitis following elective aortic repair is 5–9%, and the incidence triples in patients following emergent repair.

Mesenteric venous thrombosis accounts for <10% of cases and is generally precipitated by a hypercoagulable state due to an underlying inherited disorder such as factor V Leiden, prothrombin mutation, protein S deficiency, protein C deficiency, antithrombin deficiency, and antiphospholipid syndrome. It may also occur as a result of acquired thrombophilia in malignancies, hematologic disorders, and use of oral contraceptives.

ANATOMY AND PATHOPHYSIOLOGY

The blood supply to the intestines is supplied by the celiac artery, SMA, and inferior mesenteric artery (IMA) **(Fig. 340-1)**. Extensive collateralization occurs between major mesenteric trunks and branches of the mesenteric arcades. Collateral vessels within the small bowel are numerous and meet within the duodenum and the bed of the pancreas. Collateral vessels within the colon meet at the splenic flexure and descending/sigmoid colon. These areas, which are inherently at risk for decreased blood flow, are known as Griffiths' point and Sudeck's point, respectively, and are the most common locations for colonic ischemia (Fig. 340-1, shaded areas). The splanchnic circulation can receive up to 30% of the cardiac output. Protective responses to prevent intestinal ischemia include abundant collateralization, autoregulation of blood flow, and the ability to increase oxygen extraction from the blood.

Occlusive ischemia is a result of disruption of blood flow by an embolus or progressive thrombosis in a major artery supplying the intestine. In >75% of cases, emboli originate from the heart and preferentially lodge in the SMA just distal to the origin of the middle colic artery. Progressive stenosis of typically two of the three major vessels supplying the intestine is required for the development of chronic intestinal angina. Involvement of the SMA is most worrisome. Nonocclusive ischemia is disproportionate mesenteric vasoconstriction (arteriolar vasospasm) in response to a severe physiologic stress such as shock. If left untreated, early mucosal stress ulceration will progress to full-thickness injury.

PRESENTATION, EVALUATION, AND MANAGEMENT

Patients with CMI typically present with insidious onset of symptoms and classically present with recurrent episodes of acute dull, crampy, postprandial epigastric pain, which has also been referred to as "intestinal angina." Patients also describe fear of eating resulting in weight loss. Chronic diarrhea may also be noted. Duration of symptoms is typically 6–12 months. Physical examination will often reveal a malnourished patient with other manifestations of atherosclerosis. Prompt diagnostic investigation is necessary to rule out gastrointestinal malignancies and other possible causes, and may include accelerated evaluation via esophagogastroduodenoscopy, colonoscopy, abdominal computed tomography (CT) scan, and abdominal ultrasound examination.

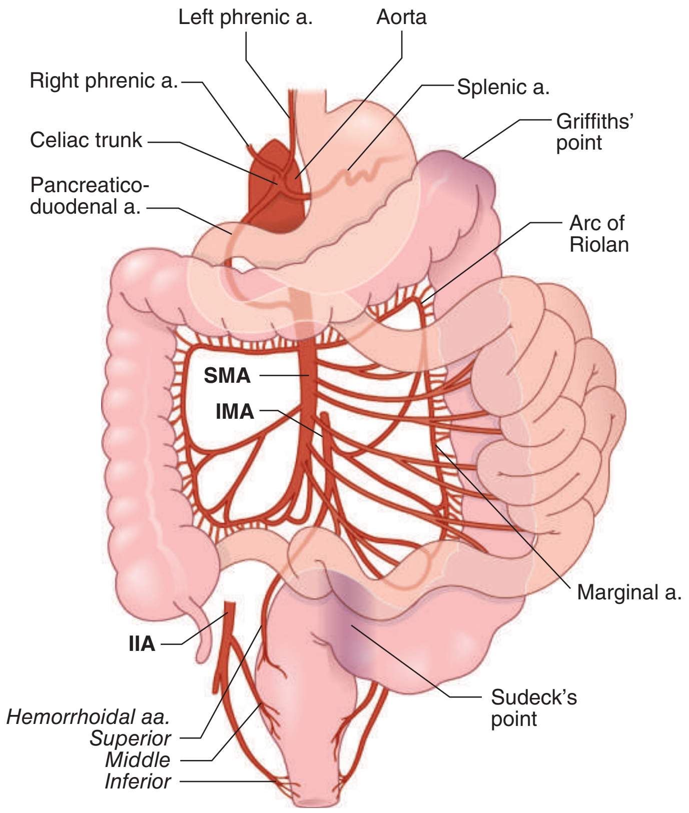

FIGURE 340-1 Blood supply to the intestines includes the celiac artery, superior mesenteric artery (SMA), inferior mesenteric artery (IMA), and branches of the internal iliac artery (IIA). Griffiths' and Sudeck's points, indicated by shaded areas, are watershed areas within the colonic blood supply and common locations for ischemia.

Duplex ultrasound has gained popularity as a screening tool for evaluation of the mesenteric vessels due to high sensitivity and specificity. Mesenteric duplex scan demonstrating a high peak velocity of flow in the SMA is associated with an ~80% positive predictive value of mesenteric ischemia. More significantly, a negative duplex scan virtually precludes the diagnosis of mesenteric ischemia. It is important to perform the test while the patient is fasting because the presence of increased bowel gas prevents adequate visualization of flow disturbances within the vessels or the lack of a vasodilation response to feeding during the test. Thin-sliced CT angiography is the gold standard diagnostic tool in assessing the degree of atherosclerotic disease of the aortic and visceral vessels as well as evaluating the bowels. Venous phase can also help diagnose mesenteric vein thrombosis.

The management of CMI includes aggressive medical therapy of atherosclerotic disease including cessation of smoking and antiplatelet and lipid-lowering medications. A full cardiac and vascular evaluation should be performed before intervention on CMI.

Treatment, involving either endovascular, open surgical, or hybrid revascularization, should be individualized based on the patient's comorbidities and anatomy. Endovascular revascularization involves targeted vessel treatment with visceral stents with the SMA anatomy being the key determinant. The revascularization of the celiac axis and IMA represents secondary focal points, offering potential therapeutic benefits. This approach becomes particularly relevant when the SMA is deemed unsuitable for intervention or when technical outcomes are deemed suboptimal. Open revascularization involves antegrade bypass from the supraceliac aorta or retrograde bypass typically from the common or external iliac arteries with a synthetic or autogenous graft to the targeted vessels, usually the SMA and/or celiac artery. In patients with suitable lesions requiring revascularization, an endovascular approach is recommended as the first-line therapy. It is especially favorable for short segment stenosis with minimal to moderate calcification or thrombus. Angioplasty with endovascular stenting in the treatment of CMI is associated with an 80% long-term success rate. Open revascularization should be considered in patients with lesions not amenable to endovascular treatment, such as severe calcification, longer lesions, small vessel diameter, or failed endovascular interventions, or in a specific subset of younger, healthier patients, in whom the potential long-term advantages may outweigh the heightened perioperative risks. Retrograde open mesenteric stenting (ROMS) is a hybrid approach, combining aspects of both traditional open surgical bypass and percutaneous endovascular therapy. ROMS is primarily indicated for treating mesenteric ischemia, both acute and chronic, particularly in cases where conventional endovascular or open surgical approaches are not feasible or have been unsuccessful. This includes scenarios where there is significant stenosis or occlusion in the mesenteric arteries that cannot be adequately addressed through less invasive percutaneous methods from the aorta or in situations where immediate direct visualization and potential resection of the bowel are necessary due to the presence of necrosis or perforation. The technique involves a laparotomy that allows for direct bowel assessment and the exposure and stenting of the SMA using a retrograde approach through the midsegment of the SMA, thus allowing for immediate revascularization, assessment of bowel viability, and, if necessary, bowel resection. ROMS also offers the advantage of faster operative times compared to traditional bypass and avoids the placement of prosthetic material in potentially contaminated peritoneal cavities, which is a significant concern in the setting of bowel necrosis and peritonitis.

Acute intestinal ischemia remains one of the most challenging diagnoses. The mortality rate of AMI is >50%. The most significant indicator of survival is the timeliness of diagnosis and treatment. An overview of diagnosis and management of each form of intestinal ischemia is given in **Table 340-1**.

AMI resulting from arterial embolus or thrombosis presentation is nonspecific and requires a high index of suspicion for the diagnosis. The most common complaint, occurring in 95% of cases, is severe, acute, nonremitting abdominal pain that is strikingly out of proportion to the physical findings. The reason behind the pain being disproportionate to the clinical findings is that ischemia initiates from the mucosal layer and progresses toward the serosal layer. This may be associated with nausea (44%), vomiting (35%), diarrhea (35%), and blood per rectum (16%). Later findings will demonstrate peritonitis and cardiovascular collapse. Specific clinical features can help differentiate the underlying etiology, whether embolic or thrombotic. Patients with embolic ischemia are typically older adults with underlying conditions that predispose to embolism such as atrial fibrillation, prior embolic event, or recent infective endocarditis. Thrombotic ischemia typically presents as an acute occlusion in patients with underlying atherosclerotic disease who may have been previously diagnosed with CMI.

AMI is a surgical emergency requiring emergent admission to a monitored bed or intensive care unit for resuscitation with fluids and broad-spectrum antibiotics in addition to further evaluation. If the diagnosis of intestinal ischemia is being considered, consultation with a surgical service is necessary. Often the decision to operate is made on a high index of suspicion from the history and physical exam despite normal laboratory findings. In patients with suspected AMI, a CT angiography with 1-mm or thinner cuts should be used to detect mesenteric arterial occlusive disease most likely from embolic or thrombotic etiology and is the gold standard. Additional diagnostic modalities that may be useful in diagnosis, but should not delay surgical therapy, include electrocardiogram (ECG) and echocardiogram. Patients with AMI should be given a heparin bolus immediately and started on a therapeutic heparin drip. Correction of electrolyte abnormalities and empiric broad-spectrum antibiotic therapy should also be initiated instantly.

If CT angiography verifies acute embolic occlusion of the SMA, surgical exploration should not be delayed. The goal of operative exploration is to resect compromised bowel, restore blood supply, and preserve all viable bowel. The entire length of the small and large bowel beginning at the ligament of Treitz should be evaluated. The SMA artery should be localized, typically at the mesocolon of the transverse colon. A transverse arteriotomy of the SMA should be made with removal of embolus with an embolectomy Fogarty catheter passed in a retrograde and antegrade manner to restore blood flow. In the case of SMA occlusion where the embolus usually lies just proximal to the origin of the middle colic artery, the proximal jejunum is often spared while the remainder of the small bowel up to the transverse colon may become ischemic. Nonviable bowel should be resected. Questionable bowel

TABLE 340-1 Overview of the Management of Acute Intestinal Ischemia

CONDITION	KEY TO EARLY DIAGNOSIS	TREATMENT OF UNDERLYING CAUSE	TREATMENT OF SPECIFIC LESION	TREATMENT OF SYSTEMIC CONSEQUENCE
Arterio-occlusive mesenteric ischemia 1. Arterial embolus	Computed tomography angiography (CTA) Early laparotomy	Anticoagulation Cardioversion Thrombectomy Broad-spectrum antibiotics	Laparotomy Embolectomy Assess viability and resect nonviable bowel	Anticoagulation Resuscitation Broad-spectrum antibiotics Emergent surgical intervention Assessment of bowel
2. Arterial thrombosis	Duplex ultrasound CTA	Anticoagulation Broad-spectrum antibiotics Resuscitation	Endovascular approach: thrombolysis, angioplasty, and stenting Endarterectomy/thrombectomy or vascular bypass Assess viability and resect nonviable bowel	Anticoagulation Resuscitation Broad-spectrum antibiotics Emergent surgical intervention Assessment of bowel
Mesenteric venous thrombosis Venous thrombosis	CTA with venous phase	Anticoagulation Resuscitation	Anticoagulation Hypercoagulable workup	Anticoagulation Resuscitation Broad-spectrum antibiotics Support cardiac output Avoid vasoconstrictors
Nonocclusive mesenteric ischemia	Vasospasm: Hypoperfusion: CT	Resuscitation Support cardiac output Avoid vasoconstrictors Broad-spectrum antibiotics	Vasospasm Intraarterial vasodilators Hypoperfusion Assess viability and resect dead bowel	Resuscitation Broad-spectrum antibiotics Support cardiac output Avoid vasoconstrictors

Source: Modified from GB Bulkley, in JL Cameron (ed): *Current Surgical Therapy*, 2nd ed. Toronto, BC Decker, 1986.

should undergo a second-look laparotomy in a 24- to 48-h period. After revascularization, peristalsis and return of a pink color of the bowel wall should be observed. Palpation of major arterial mesenteric vessels can be performed, as well as applying a Doppler flowmeter to the antimesenteric border of the bowel wall, but neither is a definitive indicator of viability.

Acute-on-chronic mesenteric ischemia typically involves the orifice of the SMA. Therefore, the entire small bowel is compromised. Revascularization with an endovascular, open, and/or a hybrid approach should be individualized based the patient's critical status, comorbidities, and anatomy. Endovascular stenting, suction thrombectomy, and/or a thrombolysis catheter should be considered for intervention. The bowel should be evaluated for viability typically via a laparoscopic or exploratory laparotomy.

Nonocclusive or vasospastic mesenteric ischemia presents with generalized abdominal pain, anorexia, bloody stools, and abdominal distention. Often these patients are obtunded, and physical findings may not assist in the diagnosis or be obscured by the underlying etiology. The presence of leukocytosis, metabolic acidosis, and/or lactic acidosis is useful in support of the diagnosis of advanced intestinal ischemia; however, these markers may not be indicative of either reversible ischemia or frank necrosis.

Emergent admission to a monitored bed or intensive care unit is recommended for resuscitation and further evaluation, and the patient should be started on broad-spectrum antibiotics. Anticoagulation is not recommended as the goal is resuscitation to maintain hemodynamics. For select patients, intramesenteric infusion of vasodilators such as papaverine, prostaglandins, or nitroglycerin can be used for the reversal of spasm and mesenteric ischemia, but the priority should be resuscitation and treatment of the underlying pathology. If ischemic colitis is a concern, colonoscopy should be considered to assess the integrity of the colonic mucosa. Ischemia of the colonic mucosa is graded as mild with minimal mucosal erythema or as moderate with pale mucosal ulcerations and evidence of extension to the muscular layer of the bowel wall. Severe ischemic colitis presents with severe ulcerations resulting in black or green discoloration of the mucosa, consistent with full-thickness bowel-wall necrosis. Ischemic colitis is optimally treated with resection of the ischemic bowel and formation of a proximal stoma.

Onset of mesenteric venous thrombosis can be acute or subacute based on the location of thrombosis in the splanchnic circulation. Patients often present with vague abdominal pain associated with nausea and vomiting. Physical examination findings include abdominal distention with mild to moderate tenderness and signs of dehydration. Findings on CT delayed venous phase include diffuse bowel-wall thickening and thrombus within the splanchnic system. IV therapeutic anticoagulation, broad-spectrum antibiotics, and correction of electrolyte abnormalities should be performed. Surgical intervention is not performed unless there is evidence of peritonitis and/or bowel perforation. If there is evidence of bowel compromise, an exploratory laparotomy should be performed, with resection of the compromised bowel. Second-look laparotomy after 24–48 h should be attempted because anticoagulation can help prevent resection of viable bowel. Hypercoagulability testing should be performed, and if underlying inherited disorders are diagnosed, life-long anticoagulation is recommended.

Acknowledgments

Rizwan Ahmed contributed to the 19th edition and Maryam Khan and Jaideep Das Gupta contributed to the 21st edition.

FURTHER READING

Bala M et al: Acute mesenteric ischemia: Updated guidelines of the World Society of Emergency Surgery. World J Emerg Surg 17:54, 2022.

Cirillo-Penn NC et al: Midterm clinical outcomes of retrograde open mesenteric stenting for mesenteric ischemia. Ann Vasc Surg 89:20, 2023.

Deng QW et al: Risk factors for postoperative acute mesenteric ischemia among adult patients undergoing cardiac surgery: A systematic review and meta-analysis. J Crit Care 42:294, 2017.

Oderich GS et al: Multicenter study of retrograde open mesenteric artery stenting through laparotomy for treatment of acute and chronic mesenteric ischemia. J Vasc Surg 68:470, 2018.

Salsano G et al: What is the best revascularization strategy for acute occlusive arterial mesenteric ischemia: Systematic review and meta-analysis. Cardiovasc Intervent Radiol 41:27, 2018.

Sise MJ: Acute mesenteric ischemia. Surg Clin North Am 94:165, 2014.

341

Acute Intestinal Obstruction

Danny O. Jacobs

EPIDEMIOLOGY

Globally, although the incidence and prevalence of acute intestinal obstruction have increased over the past two decades, morbidity and mortality appear to be decreasing. Diagnosis remains challenging. The extent of mechanical obstruction is typically described as partial, high grade, or complete—generally correlating with the risk of complications and the urgency with which the underlying disease process must be addressed. Obstruction is also commonly described as being either "simple" or, alternatively, "strangulated" if vascular insufficiency and intestinal ischemia are evident.

Acute intestinal obstruction occurs either *mechanically* from blockage or from intestinal dysmotility when there is no blockage. In the latter instance, the abnormality is *functional*. Mechanical bowel obstruction may be caused by extrinsic processes, intrinsic abnormalities of the bowel wall, or intraluminal abnormalities **(Table 341-1)**. Within each of these broad categories are many diseases that can impede intestinal propulsion. Intrinsic diseases that can cause intestinal obstruction are usually congenital, inflammatory, neoplastic, or traumatic in origin, although intussusception and radiation injury can also be etiologic.

Acute intestinal obstruction accounts for ~1–3% of all hospitalizations and a quarter of all urgent or emergent general surgery admissions. Approximately 80% of cases involve the small bowel, and about one-third of these patients show evidence of significant ischemia. The mortality rate for patients with strangulation who are operated on within 24–30 h of the onset of symptoms is ~8% but triples shortly thereafter.

Extrinsic diseases most commonly cause mechanical obstruction of the small intestine. In the United States and Europe, almost all cases are caused by postoperative adhesions, carcinomatosis, or herniation of the anterior abdominal wall. Carcinomatosis most often originates from the ovary, but can originate from the pancreas, stomach, or colon. Rarely, metastasis from distant organs like the breast and skin can also occur.

Adhesions are responsible for the majority of cases of early postoperative obstruction that require intervention. Approximately 20% of patients who were treated conservatively and between 5 and 30% of patients who were managed operatively will require readmission within 10 years for recurrence.

Open operations of the lower abdomen, including appendectomy and colorectal and gynecologic procedures, are especially likely to create adhesions that can cause bowel obstruction **(Table 341-2)**. Although laparoscopic procedures may generate fewer postoperative adhesions compared with open surgery, the risk of obstructive adhesion formation is not eliminated. The risk of internal herniation is increased by abdominal procedures such as laparoscopic or open Roux-en-Y gastric bypass.

Volvulus, which occurs when bowel twists on its mesenteric axis, can cause partial or complete obstruction and vascular insufficiency

TABLE 341-1 Most Common Causes of Acute Intestinal Obstruction

Extrinsic Disease
Adhesions (especially due to previous abdominal surgery), internal or external hernias, neoplasms (including carcinomatosis and extraintestinal malignancies, mostly commonly ovarian), endometriosis or intraperitoneal abscesses, and idiopathic sclerosis
Intrinsic Disease
Congenital (e.g., malrotation, atresia, stenosis, intestinal duplication, cyst formation, and congenital bands—the latter rarely in adults)
Inflammation (e.g., inflammatory bowel disease, especially Crohn's disease, but also diverticulitis, radiation, tuberculosis, lymphogranuloma venereum, and schistosomiasis)
Neoplasia (note: primary small-bowel cancer is rare; obstructive colon cancer may mimic small-bowel obstruction if the ileocecal valve is incompetent)
Traumatic (e.g., hematoma formation, anastomotic strictures)
Other, including intussusception (where the lead point is typically a polyp or tumor in adults), volvulus, obstruction of duodenum by superior mesenteric artery, radiation or ischemic injury, and aganglionosis, which is Hirschsprung's disease
Intraluminal Abnormalities
Bezoars, feces, foreign bodies including inspissated barium, gallstones (entering the lumen via a cholecystoenteric fistula), enteroliths

affecting the sigmoid colon most commonly comprising approximately two-thirds of all cases of volvulus and 4% of all cases of large-bowel obstruction. The cecum and terminal ileum can also volvulize, or the cecum alone may be involved as a cecal bascule. Risk factors include institutionalization, the presence of neuropsychiatric conditions requiring psychotropic medication, chronic constipation, and aging; patients typically present in their seventies or eighties. Colonic volvulus is more common in Eastern Europe, Russia, and Africa than it is the United States. It is rare for adhesions or hernias to obstruct the colon. Cancer of the descending colon and rectum is responsible for approximately two-thirds of all cases, followed by diverticulitis. Functional obstruction, also known as *ileus* and *pseudo-obstruction*, is present when dysmotility prevents intestinal contents from being propelled distally and no mechanical blockage exists. Ileus that occurs after intraabdominal surgery is the most commonly known form of functional bowel obstruction, but there are numerous other causes **(Table 341-3)**. Although postoperative ileus is most often transient, it is a common reason why hospital discharge is delayed. Pseudo-obstruction of the colon, also known as Ogilvie's syndrome, is a relatively rare disease. Some patients with Ogilvie's syndrome have colonic dysmotility due to abnormalities of their autonomic nervous system that may be inherited.

■ PATHOPHYSIOLOGY

The manifestations of acute intestinal obstruction depend on the nature of the underlying disease process, its location, and changes in blood flow **(Fig. 341-1)**. Increased intestinal contractility, which occurs proximally *and* distal to the obstruction, is a characteristic response. Subsequently, intestinal peristalsis slows as the intestine or stomach proximal to the point of obstruction dilates and fills with gastrointestinal secretions and swallowed air. Although swallowed air is a primary source of intestinal distension, intraluminal air may also accumulate from fermentation, local carbon dioxide production, and altered gaseous diffusion.

TABLE 341-2 Acute Small-Intestinal and Colonic Obstruction Incidences

CAUSE	INCIDENCE
Postoperative adhesions	>50% overall
Neoplasms	~20%
Hernias (especially ventral or internal types, where the risk of strangulation is increased)	~10%
Inflammatory bowel disease, other inflammation (obstruction may resolve if acute inflammation and edema subside)	~5%
Intussusception, volvulus, other miscellaneous diseases	<15%

TABLE 341-3 Most Common Causes of Ileus (Functional or Pseudo-Obstruction of the Intestine)

Intraabdominal procedures, lumbar spinal injuries, or surgical procedures on the lumbar spine and pelvis
Metabolic or electrolyte abnormalities, especially hypokalemia and hypomagnesemia, but also hyponatremia, uremia, and severe hyperglycemia
Drugs such as opiates, antihistamines, and some psychotropic (e.g., haloperidol, tricyclic antidepressants) and anticholinergic agents
Intestinal ischemia
Intraabdominal or retroperitoneal inflammation or hemorrhage
Lower lobe pneumonias
Intraoperative radiation (likely due to smooth muscle damage)
Systemic sepsis
Hyperparathyroidism
Pseudo-obstruction (Ogilvie's syndrome)
Ileus secondary to hereditary or acquired visceral myopathies and neuropathies that disrupt myocellular neural coordination
Some collagen vascular diseases such as lupus erythematosus or scleroderma

Intraluminal dilation also increases intraluminal pressure. When luminal pressure exceeds venous pressure, venous and lymphatic drainage is impeded. Edema ensues, and the bowel wall proximal to the site of blockage may become hypoxemic. Epithelial necrosis can be identified within 12 h of obstruction. Ultimately, arterial blood supply may become so compromised that full-thickness ischemia, necrosis, and perforation result. Stasis increases bacteria counts within the jejunum and ileum. Bacteria, such as *Escherichia coli*, *Streptococcus faecalis*, and *Klebsiella*, and other pathogens may be recovered from intestinal cultures, mesenteric lymph nodes, the bloodstream, and other sites.

Other manifestations depend on the degree of hypovolemia, the patient's metabolic response, and the presence or absence of associated intestinal ischemia. Inflammatory edema eventually increases the production of reactive oxygen species and activates neutrophils and macrophages, which accumulate within the bowel wall. Their accumulation, along with changes in innate immunity, disrupts secretory and neuromotor processes. Dehydration is caused by loss of the normal intestinal absorptive capacity as well as fluid accumulation in the gastric or intestinal wall and intraperitoneally.

Anorexia and emesis tend to exacerbate intravascular volume depletion. In the worst-case scenario after high-grade distal obstruction, emesis leads to losses of gastric potassium, hydrogen, and chloride, while dehydration stimulates proximal renal tubule bicarbonate reabsorption. Intraperitoneal fluid accumulation, especially in patients with severe distal bowel obstruction, may increase intraabdominal pressure enough to elevate the diaphragm, inhibit respiration, impede systemic venous return, and promote vascular instability. Severe hemodynamic compromise may elicit a systemic inflammatory response and generalized microvascular leakage.

Closed-loop obstruction results when the proximal and distal openings of a given bowel segment are both occluded, for example, due to volvulus or a hernia. It is the most common precursor for strangulation, but not every closed loop strangulates. The risk of vascular insufficiency, systemic inflammation, hemodynamic compromise, and irreversible intestinal ischemia is much greater in patients with closed-loop obstruction. Pathologic changes may occur rapidly, such that emergent intervention is indicated. Irreversible bowel ischemia may progress to transmural necrosis even if obstruction is relieved. The provider should remember that patients with high-grade distal colonic obstruction who have competent ileocecal valves may present with closed-loop obstruction. In this instance, the cecum may progressively dilate such that ischemic necrosis results in perforation, especially when the cecal diameter exceeds 10–12 cm. Patients with distal colonic obstruction whose ileocecal valves are incompetent tend

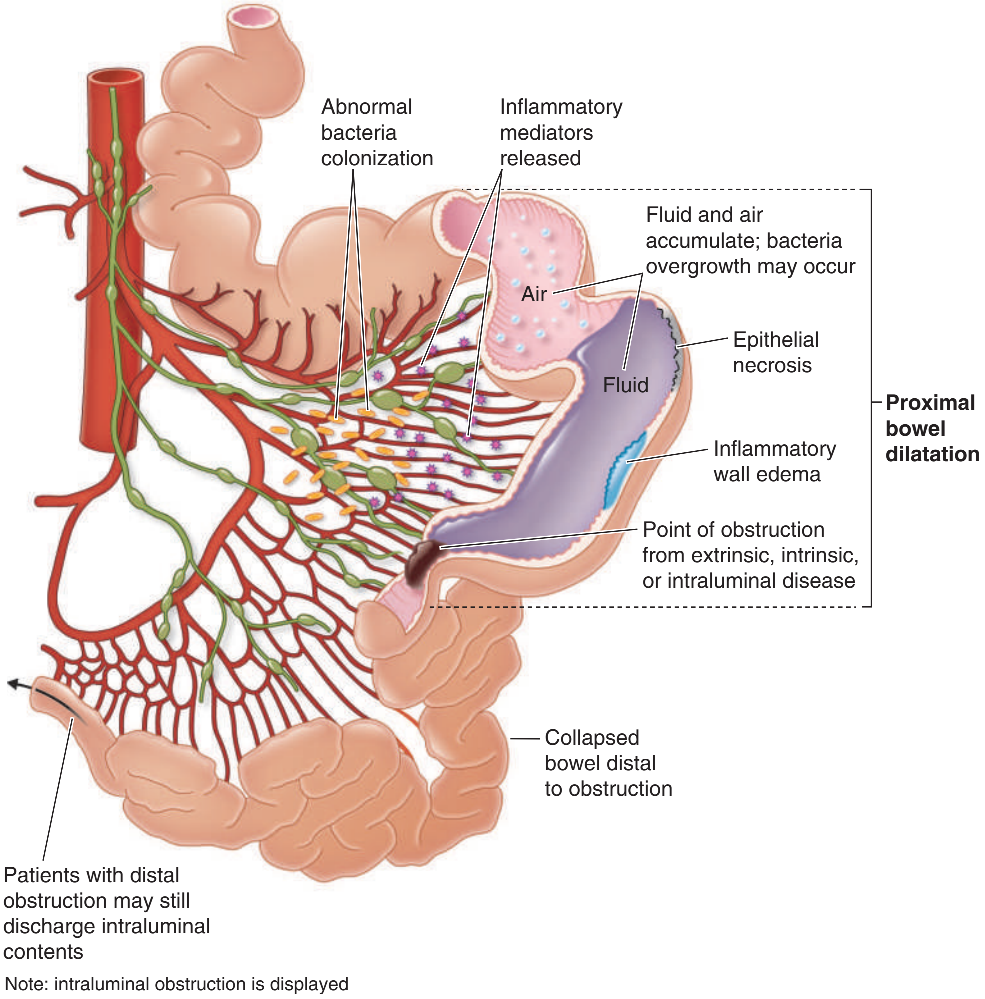
Abnormal bacteria colonization
Inflammatory mediators released
Fluid and air accumulate; bacteria overgrowth may occur
Air
Fluid
Epithelial necrosis
Proximal bowel dilatation
Inflammatory wall edema
Point of obstruction from extrinsic, intrinsic, or intraluminal disease
Collapsed bowel distal to obstruction
Patients with distal obstruction may still discharge intraluminal contents
Note: intraluminal obstruction is displayed

FIGURE 341-1 Pathophysiologic changes of small-bowel obstruction.

to present later in the course of disease and mimic patients with distal small-bowel obstruction.

HISTORY AND PHYSICAL FINDINGS

Even though the presenting signs and symptoms can be misleading, many patients with acute obstruction can be accurately diagnosed after a thorough history and physical examination is performed before imaging. Even though small-bowel obstruction with strangulation can be especially difficult to diagnosis promptly, early recognition allows earlier treatment, which decreases the risk of morbidity and mortality.

The cardinal signs are colicky abdominal pain, abdominal distention, emesis, and obstipation. More intraluminal fluid accumulates in patients with distal obstruction, which typically leads to greater distention, more discomfort, and delayed emesis. This emesis is feculent when there is bacterial overgrowth. Patients with more proximal obstruction commonly present with less abdominal distention but more pronounced vomiting. Elements of the history that might be helpful include any prior history of surgery, including herniorrhaphy, as well as any history of cancer or inflammatory bowel disease.

Most patients, even those with simple obstruction, appear to be critically ill. Many may be oliguric, hypotensive, and tachycardic because of severe intravascular volume depletion. Fever is worrisome for strangulation or systemic inflammation. Bowel sounds and bowel functional activity are notoriously difficult to interpret. Classically, many patients with early small-bowel obstruction will have high-pitched, "musical" tinkling bowel sounds and peristaltic "rushes" known as borborygmi. Later in the course of disease, the bowel sounds may be absent or hypoactive as peristaltic activity decreases. This contrasts with the common findings in patients with ileus or pseudo-obstruction where bowel sounds are typically absent or hypoactive from the beginning. Lastly, patients with partial blockage may continue to pass flatus and stool, and those with complete blockage may even evacuate bowel contents present downstream beyond their obstruction.

All surgical incisions should be examined, and the presence of a tender abdominal or groin mass strongly suggests that an incarcerated hernia may be the cause of obstruction. The presence of tenderness should increase the concern about the presence of complications such as ischemia, necrosis, or localized peritonitis. Severe pain with localization or signs of peritoneal irritation is suspicious for strangulated or closed-loop obstruction. It is important to remember that the discomfort may be out of proportion to physical findings mimicking the complaints of patients with acute mesenteric ischemia. Patients with colonic volvulus present with the classic manifestations of closed-loop obstruction: severe abdominal pain, vomiting, and obstipation. Asymmetrical abdominal distension and a tympanic mass may be evident.

Patients with ileus or pseudo-obstruction may have signs and symptoms like those with bowel obstruction. Although abdominal distention is present, colicky abdominal pain is typically absent, and patients may not have nausea or emesis. Ongoing, regular discharge of stool or flatus can sometimes help distinguish patients with ileus from those with complete mechanical bowel obstruction. The overall risk of ileus appears to be less in patients who undergo laparoscopic procedures.

LABORATORY AND IMAGING STUDIES

Laboratory testing should include a complete blood count and serum electrolyte and creatinine measurements. Serial assessments are often useful. Mild hemoconcentration and slight elevation of the white blood cell count commonly occur after simple bowel obstruction. Emesis and dehydration may cause hypokalemia, hypochloremia, elevated blood urea nitrogen–to–creatinine ratios, and metabolic alkalosis. Patients may be hyponatremic on admission because many have attempted to rehydrate themselves with hypotonic fluids. The presence of guaiac-positive stools and iron-deficiency anemia are strongly suggestive of malignancy.

Higher white blood cell counts with the presence of immature forms and the presence of metabolic acidosis are worrisome for severe volume depletion or ischemic necrosis and sepsis. Presently, no laboratory tests are especially useful for identifying the presence of simple or strangulated obstruction, although increases in serum D-lactate, creatine kinase BB isoenzymes, or intestinal fatty acid binding protein levels may be suggestive of the latter.

Recommendations for diagnostic imaging continue to evolve. In all cases, the key is not to delay operative intervention unnecessarily when the patient's signs or symptoms strongly suggest that high-grade or complete obstruction or bowel compromise is present. Abdominal radiography, which must include upright or cross-table lateral views, can be completed quickly and may indicate the need for emergency surgical intervention in patients who are not in the immediate postoperative period. A "staircasing" pattern of dilated air and fluid-filled small-bowel loops >2.5 cm in diameter with little or no air seen in the colon are classical findings in patients with small-bowel obstruction. Little bowel gas appears in patients with proximal bowel obstruction or in patients whose intestinal lumens are filled with fluid.

Upright plain films of the abdomen of patients with large-bowel obstruction typically show colon dilatation. Small-bowel air-fluid levels may not be obvious if the ileocecal valve is incompetent. Although it can be difficult to distinguish from ileus, small-bowel obstruction is more likely when air-fluid levels are seen without significant colonic distension. Free air suggests that perforation has occurred in patients who have not recently undergone surgical procedures. A gas-filled, "coffee bean"–shaped dilated shadow may be seen in patients with volvulus.

More sophisticated imaging can be beneficial when the diagnosis is unclear. Computed tomography (CT) is the most frequently used imaging modality. Its sensitivity for detecting bowel obstruction is ~95% (78–100%) in patients with high-grade obstruction, with a specificity of 96% and an accuracy of ≥95%. Its accuracy in diagnosing closed-loop obstruction is much lower (60%). CT may also provide useful information regarding location or to identify circumstances where surgical intervention is urgently needed. Patients who have evidence of contrast appearing within the cecum within 4–24 h of oral administration of water-soluble contrast can be expected to improve with high sensitivity and specificity (~95% each).

Contrast studies may demonstrate a "bird's beak," a "c-loop," or "whorl" deformity on CT imaging at the site where twisting obstructs the lumen when a colonic volvulus is present. Abdominal radiography, unlike CT imaging, may not accurately distinguish obstruction from other causes of colonic dysmotility. Examples of some CT images are provided in **Fig. 341-2**.

Ultrasonographic evaluations are especially difficult to interpret but may be sensitive and appropriate studies to evaluate patients who are pregnant or for whom x-ray exposure is otherwise contraindicated or inappropriate.

CT imaging with enteral and IV contrast can also identify ischemia. Altered bowel wall enhancement is the most specific early finding, but its sensitivity is low. Mesenteric venous gas, pneumoperitoneum, and pneumatosis intestinalis are late findings indicating the presence of bowel necrosis. CT scanning after a water-soluble contrast enema may help distinguish ileus or pseudo-obstruction from distal large-bowel obstruction in patients who present with evidence of small-bowel and colonic distention. CT enteroclysis, though now rarely performed, can accurately identify neoplasia as a cause of bowel obstruction. Contrast enemas or colonoscopies are almost always needed to identify causes of acute colonic obstruction.

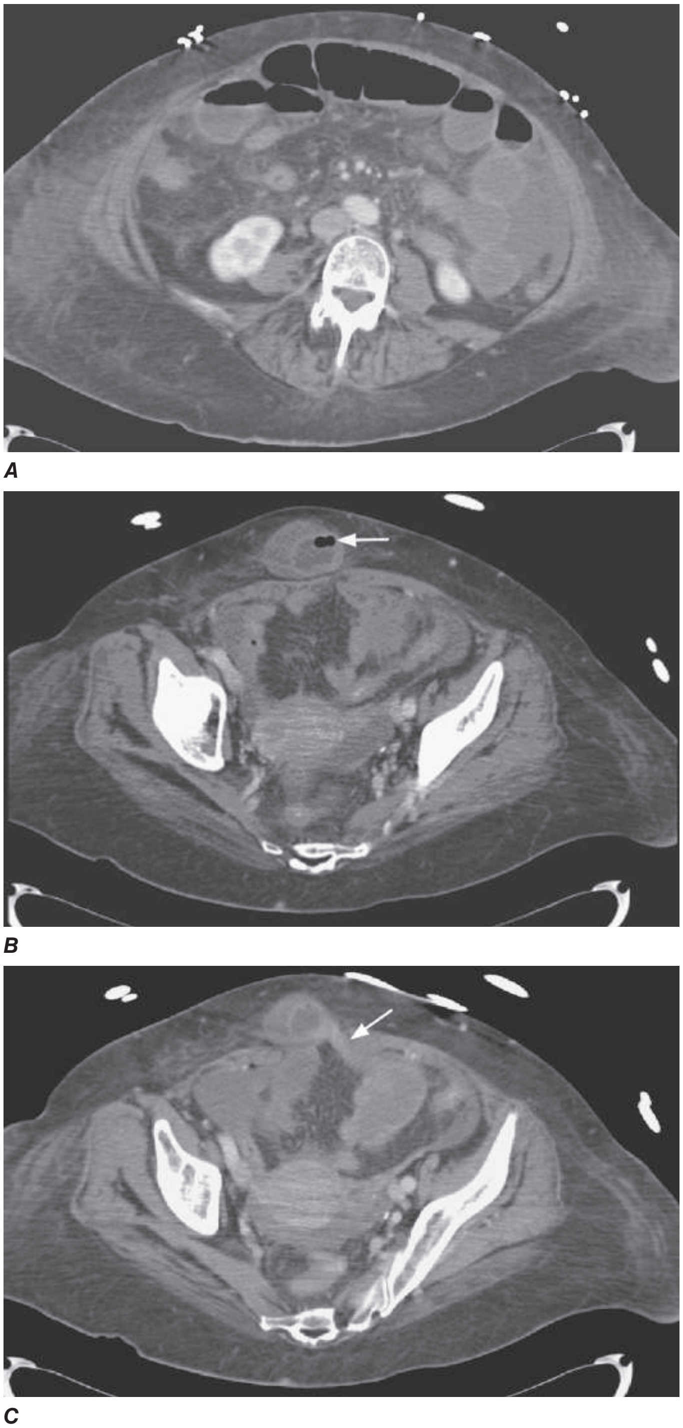

FIGURE 341-2 Computed tomography with oral and intravenous contrast demonstrating (***A***) evidence of small-bowel dilatation with air-fluid levels consistent with a small-bowel obstruction; (***B***) a partial small-bowel obstruction from an incarcerated ventral hernia (*arrow*); and (***C***) decompressed bowel seen distal to the hernia (*arrow*). (*Reproduced with permission from D Longo et al: Harrison's Principles of Internal Medicine, 18th ed. New York: McGraw-Hill; 2012.*)

Barium studies are generally contraindicated in patients with firm evidence of complete or high-grade bowel obstruction, especially when they present acutely. Barium should never be given orally to a patient with possible obstruction until that diagnosis has been excluded. In every other instance, such investigations should only be performed in exceptional circumstances and with great caution because patients with significant obstruction may develop barium concretions as an additional source of blockage and some who would have otherwise recovered will require operative intervention. Barium opacification also renders cross-sectional imaging studies or angiography uninterpretable.

TREATMENT

Acute Intestinal Obstruction

An improved understanding of the pathophysiology of bowel obstruction and the importance of fluid resuscitation, electrolyte repletion, intestinal decompression, and the selected use of antibiotics has likely contributed to a reduction in mortality from acute bowel obstruction. Patients should be stabilized as quickly as possible. Nasogastric tube suction decompresses the stomach, minimizes further distention from swallowed air, improves patient comfort, and reduces the risk of aspiration. Urine output should be assessed using a Foley catheter. In some cases, for example, in patients with cardiac disease, central venous pressures should be monitored. The use of antibiotics is controversial, although prophylactic administration may be warranted if operation is anticipated.

Complete bowel obstruction is an indication for intervention. Stenting may be possible and warranted for some patients with high-grade obstruction due to unresectable stage IV malignancies. Stenting may also allow elective mechanical bowel preparation before operation. Because treatment options are so variable, it is helpful to make as precise a diagnosis as possible preoperatively.

ILEUS

Patients with ileus are treated supportively with IV fluids and nasogastric decompression while any underlying pathology is treated, taking care to optimize the use of narcotics. Pharmacologic treatments continue to be evaluated with some studies showing that treatment with peripherally active μ-opioid receptor antagonists (e.g., alvimopan and methylnaltrexone) or 5-HT$_4$ agonists that stimulate the release of acetylcholine from enteric neurons (e.g., mosapride and prucalopride) may accelerate gastrointestinal recovery in some patients who have undergone abdominal surgery.

COLONIC PSEUDO-OBSTRUCTION (OGILVIE'S DISEASE)

Neostigmine is an acetylcholinesterase inhibitor that increases cholinergic (parasympathetic) activity, which can stimulate colonic motility. Some studies have shown it to be moderately effective in alleviating acute colonic pseudo-obstruction. It is the most common therapeutic approach and can be used once it is certain that there is no mechanical obstruction. Cardiac monitoring is required, and atropine should be immediately available. Intravenous administration induces defecation and flatus within 10 min in the majority of patients who will respond. Sympathetic blockade by epidural anesthesia can successfully ameliorate pseudo-obstruction in some patients.

VOLVULUS

Patients with sigmoid volvulus can often be decompressed using a flexible tube inserted through a rigid proctoscope or using a flexible sigmoidoscope. Successful decompression results in sudden release of gas and fluid with evidence of decreased abdominal distension and allows definitive correction to be scheduled electively. Cecal volvulus most often requires laparotomy or laparoscopic correction.

INTRAOPERATIVE STRATEGIES

Approximately 60–80% of selected patients with mechanical bowel obstruction can be successfully treated conservatively. Most cases of radiation-induced obstruction should be managed nonoperatively if possible. Early consultation with a surgeon is prudent when there is concern about strangulation obstruction or other abnormality that needs to be addressed urgently. Deterioration signifies a need for intervention. At this time, the decision as to whether the patient can continue to be treated nonoperatively can only be based on clinical judgment, although, as described earlier, imaging studies can sometimes be helpful. The frequency of major complications after operation ranges from 12 to 47%, with greater risk being attributed to resection therapies and the patient's overall health. Risk is increased for patients with American Society of Anesthesiologists (ASA) physical status of class III or higher.

At operation, dilation proximal to the site of blockage with distal collapse is a defining feature of bowel obstruction. Intraoperative strategies depend on the underlying problem and range from lysis of adhesions to resection with or without diverting ostomy to primary resection with anastomosis. Resection is warranted when there is concern about the bowel's viability after the obstructive process is relieved. Laparoscopic approaches can be useful for patients with early obstruction when extensive adhesions are not expected to be present. Some patients with high-grade obstruction secondary to malignant disease that is not amendable to resection will benefit from bypass procedures.

ADULT INTUSSUSCEPTION AND GALLSTONE ILEUS

Primary resection is prudent. Careful manual reduction of any involved bowel may limit the amount of intestine that needs to be removed. A proximal ostomy may be required if unprepped colon is involved. The most common site of intestinal obstruction in patients with gallstone "ileus" is the ileum (60% of patients). The gallstone enters the intestinal tract most often via a cholecystoduodenal fistula. It can usually be removed by operative enterolithotomy. Addressing gallbladder disease during urgent or emergent surgery is not recommended.

POSTOPERATIVE BOWEL OBSTRUCTION

Early postoperative mechanical bowel obstruction is that which occurs within the first 6 weeks of operation. Although it tends to respond and behave differently from classic mechanical bowel obstruction and may be very difficult to distinguish from postoperative ileus, most are partial and can be expected to resolve spontaneously. A higher index of suspicion for a definitive site of obstruction is warranted for patients who undergo laparoscopic surgical procedures. Patients who first had ileus and then subsequently develop obstructive symptoms after an initial return of normal bowel function are more likely to have true postoperative small-bowel obstruction. The longer it takes for a patient's obstructive symptoms to resolve after hospitalization, the more likely the patient is to require surgical intervention.

Acknowledgment

The wisdom and expertise of Dr. William Silen are gratefully acknowledged.

FURTHER READING

Catena F et al: Adhesive small bowel adhesions obstruction: Evolutions in diagnosis, management and prevention. World J Gastrointest Surg 27:222, 2016.

Ferrada P et al: Surgery or stenting for colonic obstruction: A practice management guideline from the Eastern Association for the Surgery of Trauma. J Trauma Acute Care Surg 80:659, 2016.

Griffiths S, Glancy DG: Intestinal obstruction. Surgery (Oxford) 41:47, 2023.

Jaffe T, Thompson WM: Large-bowel obstruction in the adults: Classic radiographic and CT findings, etiology and mimics. Radiology 275:651, 2015.

Long B et al: Emergency medicine evaluation and management of small bowel obstruction: Evidence-based recommendations. J Emerg Med 56:166, 2019.

Paulson EK, Thompson WM: Review of small-bowel obstruction: The diagnosis and when to worry. Radiology 275:332, 2015.

Perry H et al: Relative accuracy of emergency CT in adults with non-traumatic abdominal pain. Brit Inst Rad 89:20150416, 2016.

Taylor MR, Lalani N: Adult small bowel obstruction. Acad Emerg Med 20:528, 2013.

342

Acute Appendicitis and Peritonitis

Danny O. Jacobs

ACUTE APPENDICITIS

INCIDENCE AND EPIDEMIOLOGY

Acute appendicitis is the most common acute general surgery emergency affecting the abdomen, with a rate of ~10–11 cases per 10,000 people annually over the past several decades affecting biological males and females equivalently. Appendicitis still occurs most commonly in 10- to 19-year-olds, although the worldwide incidence and average age at diagnosis appear to be increasing gradually, whereas disability from complications and mortality appear to be decreasing. Significant differences in geographic risk variation are noted in the United States and worldwide.

The risk of perforation is ~10–20% and is a significant cause of excess morbidity for all patients, but the risk appears to be significantly higher in patients <5 or >65 years of age. Patients with higher socioeconomic status appear to have lower risk of appendicitis with or without perforation. It is also important to note that the incidence of appendiceal tumors also appears to be increasing.

How best to manage patients who present with an appendiceal mass and small localized abscess without perforation is controversial, especially whether interval or immediate appendectomy should be performed when both can be appropriate depending on the clinical details.

PATHOGENESIS OF APPENDICITIS AND APPENDICEAL PERFORATION

Appendicitis was first described in 1886 by Reginald Fitz. Its etiology is still not completely understood. Fecaliths, incompletely digested food residue, lymphoid hyperplasia, intraluminal scarring, tumors, bacteria, viruses, and inflammatory bowel disease have all been associated with inflammation of the appendix with potentially different outcomes depending on pathogenesis.

Although not proven, obstruction of the appendiceal lumen is believed to be an important step in the development of appendicitis—at least in some cases. Here, obstruction leads to bacterial overgrowth and luminal distension, with an increase in intraluminal pressure that can inhibit the flow of lymph and blood. Then, vascular thrombosis and ischemic necrosis with perforation of the distal appendix may occur. For this reason, perforation that occurs near the base of the appendix should raise concerns about another disease process. Most patients who will perforate do so before they are evaluated by surgeons.

Appendiceal fecaliths (or appendicoliths) are found in ~50% of patients with gangrenous appendicitis who perforate but are rarely identified in those who have simpler disease. This suggests that the underlying pathophysiologic processes are different. Uncomplicated appendicitis (e.g., without gangrenous necrosis, appendicoliths, perforation, or tumors) does not always progress to perforation. It appears that at least some cases of simple acute appendicitis may temporarily resolve or be managed with antibiotic therapy, although at least one-third will require subsequent appendectomy. Nevertheless, appendicitis does not invariably progress to perforation, and the use of nonoperative therapies to treat uncomplicated appendicitis continues to be studied intensively.

These findings highlight the importance of clinical decision-making and risk assessment when deciding and discussing treatment options with patients who presumably have simple disease, for example, deciding who is an appropriate candidate for nonoperative management and who is not without bias. The latter is especially pertinent given the difficulty in assessing which patients might progress to perforation and which will not.

When perforation occurs, the resultant leak may be contained by the omentum or other surrounding tissues to form an abscess. Free perforation normally causes severe peritonitis. These patients may also develop infective suppurative thrombosis of the portal vein and its tributaries along with intrahepatic abscesses. The prognosis of the very unfortunate patients who develop this rare but dreaded complication is very poor.

TABLE 342-1 Some Conditions That Mimic Appendicitis

- Crohn's disease
- Cholecystitis or other gallbladder disease
- COVID-19 infection occasionally co-incident
- Diverticulitis
- Ectopic pregnancy
- Endometriosis
- Gastroenteritis or colitis
- Gastric or duodenal ulceration
- Hepatitis
- Kidney disease, including nephrolithiasis
- Liver abscess
- Meckel's diverticulitis
- Mittelschmerz
- Mesenteric adenitis
- Omental torsion
- Pancreatitis
- Lower lobe pneumonia
- Pelvic inflammatory disease
- Ruptured ovarian cyst or other cystic disease of the ovaries
- Small-bowel obstruction
- Urinary tract infection

CLINICAL MANIFESTATIONS

Improved diagnosis, supportive care, and surgical interventions are likely responsible for the remarkable decrease in the risk of mortality from simple appendicitis to currently <1%. Nevertheless, it is still important to identify patients who might have appendicitis as early as possible. Patients who have persistent symptoms that have not improved over 48 h may be more likely to perforate or develop other complications.

Appendicitis should be included in the differential diagnosis of abdominal pain for every patient in any age group unless it is certain that the organ has been previously removed (Table 342-1).

The appendix's anatomic location, which varies, may directly influence how the patient presents. Where the appendix can be "found" ranges from local differences in how the appendiceal body and tip lie relative to its attachment to the cecum (Figs. 342-1 and 342-2), to where the appendix is actually situated in the peritoneal cavity—for example, from its typical location in the right lower quadrant, to the pelvis, right flank, right upper quadrant (as may be observed during pregnancy), or even the left side of the abdomen for patients with malrotation or who have severely redundant colons.

Because the differential diagnosis of appendicitis is so extensive, deciding if a patient has appendicitis can be difficult (Table 342-2). Many patients may not present with the classically described history or physical findings, and some may not have any abdominal discomfort early in the disease process. Soliciting an appropriate history requires detecting and evaluating symptoms that might suggest alternative diagnoses.

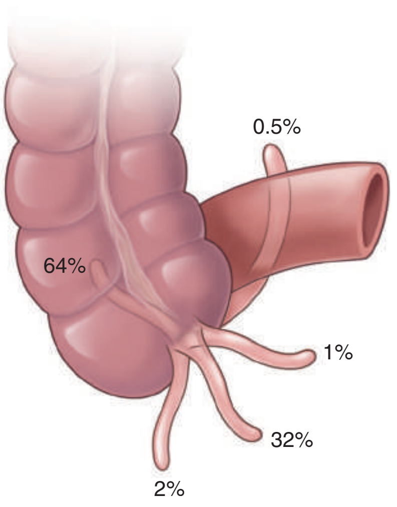

FIGURE 342-1 Regional anatomic variations of the appendix.

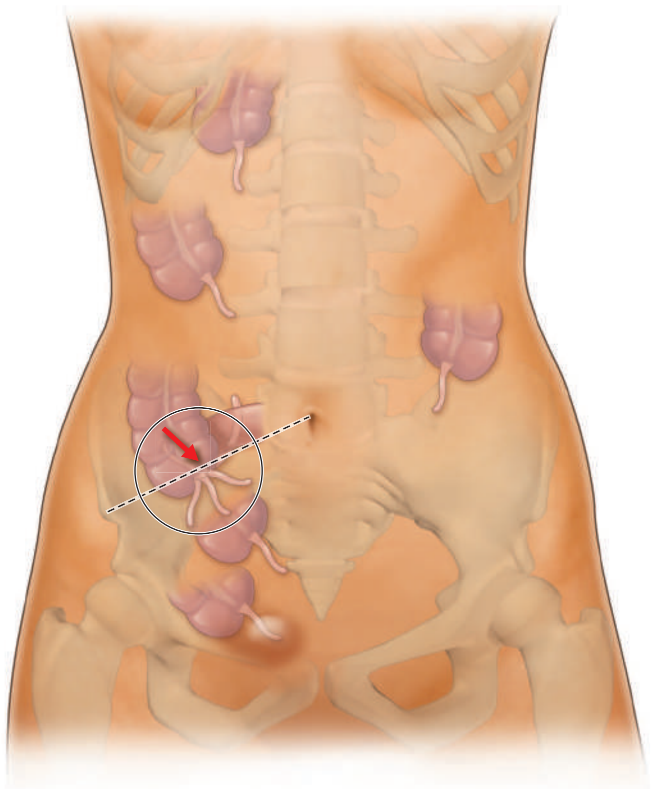

FIGURE 342-2 Locations of the appendix and cecum.

What is the classic history? Nonspecific complaints occur first. Patients may notice changes in bowel habits or malaise and vague, perhaps intermittent, crampy abdominal pain in the epigastric or periumbilical region. The pain subsequently migrates to the right lower quadrant over 12–24 h, where it is sharper and can be definitively localized as transmural inflammation when the appendix irritates the parietal peritoneum. Parietal peritoneal irritation may be associated with local muscle rigidity and stiffness. Patients with appendicitis will most often observe that their nausea, if present, followed the development of abdominal pain, which can help distinguish them from patients with gastroenteritis, for example, in whom nausea occurs first. Emesis, if present, also occurs after the onset of pain and is typically mild and scant. Thus, timing of the onset of symptoms and the characteristics of the patient's pain and any associated findings must be rigorously assessed. Anorexia is so common that the diagnosis of appendicitis should be questioned in its absence.

Arriving at the correct diagnosis is even more challenging when the appendix is not located in the right lower quadrant, in women of childbearing age, and in the very young or elderly. Because the differential diagnosis of appendicitis is so broad, often the key question to answer expeditiously is whether the patient has appendicitis or some other condition that requires immediate operative intervention. A major concern is that the likelihood of a delay in diagnosis is greater if the appendix is unusually positioned. All patients should undergo a rectal examination. An inflamed appendix located behind the cecum or below the pelvic brim may prompt very little tenderness of the anterior abdominal wall.

TABLE 342-2 Relative Frequency of Common Presenting Symptoms

SYMPTOMS	FREQUENCY
Abdominal pain	>95%
Anorexia	>70%
Constipation	4–16%
Diarrhea	4–16%
Fever	10–20%
Migration of pain to right lower quadrant	50–60%
Nausea	>65%
Vomiting	50–75%

TABLE 342-3 Relative Frequency of Some Presenting Signs

SIGNS	FREQUENCY
Abdominal tenderness	>95%
Right lower quadrant tenderness	>90%
Rebound tenderness	30–70%
Rectal tenderness	30–40%
Cervical motion tenderness	30%
Rigidity	~10%
Psoas sign	3–5%
Obturator sign	5–10%
Rovsing's sign	5%
Palpable mass	<5%

Patients with pelvic appendicitis are more likely to present with dysuria, urinary frequency, diarrhea, or tenesmus. They may only experience pain in the suprapubic region on palpation or on rectal or pelvic examination. A pelvic examination in women is mandatory to rule out conditions affecting urogynecologic organs that can cause abdominal pain and mimic appendicitis such as pelvic inflammatory disease, ectopic pregnancy, and ovarian torsion. None of the currently available decision tools yet appear to be able to circumvent or obviate the need for expert clinical opinion. The relative frequencies of some presenting signs are displayed in Table 342-3.

Patients with simple appendicitis normally only appear mildly ill with a pulse and temperature that are usually only slightly above normal. The provider should be concerned about other disease processes beside appendicitis or the presence of complications such as perforation, phlegmon, or abscess formation if the temperature is >38.3°C (~101°F) and if there are rigors.

Patients with appendicitis will be found to lie quite still to avoid peritoneal irritation caused by movement, and some will report discomfort caused by a bumpy car ride on the way to the hospital or clinic, coughing, sneezing, or other actions that replicate a Valsalva maneuver. The entire abdomen should be examined systematically starting in an area where the patient does not report discomfort if possible. Classically, maximal tenderness is identified where the appendix is most often located—in the right lower quadrant at or near McBurney's point, which is approximately one-third of the way along a line originating at the anterior iliac spine and running to the umbilicus. Gentle pressure in the left lower quadrant may elicit pain in the right lower quadrant if the appendix is located there. This is Rovsing's sign (Table 342-4). Evidence of parietal peritoneal irritation is often best elicited by gentle abdominal percussion, jiggling the patient's gurney or bed, or mildly bumping the feet.

Atypical presentation and pain patterns are common, especially in the very old or the very young. Diagnosing appendicitis in children can be especially challenging because they tend to respond so dramatically to stimulation and obtaining an accurate history may be difficult. In addition, it is important to remember that the smaller omentum found in children may be less likely to wall off an appendiceal perforation. Observing the child in a quiet surrounding may be helpful.

TABLE 342-4 Classic Signs of Appendicitis in Patients with Abdominal Pain

MANEUVER	FINDINGS
Rovsing's sign	Palpating in the left lower quadrant causes pain in the right lower quadrant
Obturator sign	Internal rotation of the hip causes pain, suggesting the possibility of an inflamed appendix located in the pelvis
Iliopsoas sign	Extending the right hip causes pain along posterolateral back and hip, suggesting retrocecal appendicitis

Signs and symptoms of appendicitis can be subtle in the elderly who may not react as vigorously to appendicitis as younger people. Pain, if noticed, may be minimal and have originated in the right lower quadrant or, otherwise, where the appendix is located. It may never have been noticed to be intermittent, or there may only be significant discomfort with deep palpation. Nausea, anorexia, and emesis may be the predominant complaints. The rare patient may even present with signs and symptoms of distal bowel obstruction secondary to appendiceal inflammation and phlegmon or abscess formation.

■ LABORATORY TESTING

Laboratory testing does not identify patients with appendicitis. The white blood cell count is only mildly to moderately elevated in ~70% of patients with simple appendicitis (with a leukocytosis of 10,000–18,000 cells/μL). A "left shift" toward immature polymorphonuclear leukocytes is present in >95% of cases. A sickle cell preparation may be prudent to obtain in those of African, Spanish, Mediterranean, or Indian ancestry. Serum amylase and lipase levels should be measured.

Urinalysis is indicated to help exclude genitourinary conditions that may mimic acute appendicitis, but a few red or white blood cells may be present as a nonspecific finding. An inflamed appendix that abuts the ureter or bladder may cause sterile pyuria or hematuria. Every woman of childbearing age should have a pregnancy test. Cervical cultures are indicated if pelvic inflammatory disease is suspected. Anemia and guaiac-positive stools should raise concern about the presence of other diseases or complications such as cancer.

■ IMAGING

Plain films of the abdomen are rarely helpful and so are not routinely obtained unless the clinician is worried about other conditions such as intestinal obstruction, perforated viscus, or ureterolithiasis. Less than 5% of patients will present with an opaque fecalith in the right lower quadrant. The presence of a fecalith is not diagnostic of appendicitis, although its presence in an appropriate location where the patient complains of pain is suggestive and is associated with a greater likelihood of complications.

The effectiveness of ultrasonography as a tool to diagnosis appendicitis is highly operator dependent. Even in very skilled hands, the appendix may not be visualized. Its overall sensitivity is ~0.86, with a specificity of 0.81. Ultrasonography, especially intravaginal techniques, appears to be most useful for identifying pelvic pathology in women. Ultrasonographic findings suggesting the presence of appendicitis include wall thickening, an increased appendiceal diameter, and the presence of free fluid. Current practice in some institutions is to first perform ultrasonography and progress to other imaging studies only if the findings are equivocal or complications are suspected.

The sensitivity and specificity of computed tomography (CT) are at least 0.94 and 0.95, respectively. Thus, CT imaging, given its high negative predictive value, especially with the safe use of oral and intravenous contrast, may be helpful if the diagnosis is in doubt, although studies performed early in the course of disease may not have any typical radiographic findings. Overall, in patients in whom the diagnosis is uncertain, delaying operation at the time of presentation to obtain CT does not appear to increase the risk of perforation. CT scanning is a superior method for assessing the severity of acute appendicitis in the absence of peritoneal findings indicative of perforation, abscess, or suspicion of an associated malignancy.

Suggestive findings on CT examination include dilatation >6 mm with wall thickening, a lumen that does not fill with enteric contrast, and fatty tissue stranding or air surrounding the appendix, which suggests inflammation **(Figs. 342-3 and 342-4)**. The presence of luminal air or contrast is not consistent with a diagnosis of appendicitis. Furthermore, nonvisualization of the appendix is a nonspecific finding that should not be used to rule out the presence of appendiceal or periappendiceal inflammation.

■ SPECIAL PATIENT POPULATIONS

Appendicitis is the most common extrauterine general surgical emergency observed during pregnancy. Early symptoms of appendicitis such as nausea and anorexia may be overlooked. Diagnosing appendicitis in pregnant patients may be especially difficult because as the uterus enlarges the appendix may be pushed higher along the right flank even to the right upper quadrant or because the gravid uterus may obscure typical physical findings. Ultrasonography may facilitate early diagnosis. A high index of suspicion is required because of the effects of unrecognized and untreated appendicitis on the fetus. For example, the fetal mortality rate is four times greater (from 5 to 20%) in patients with perforation.

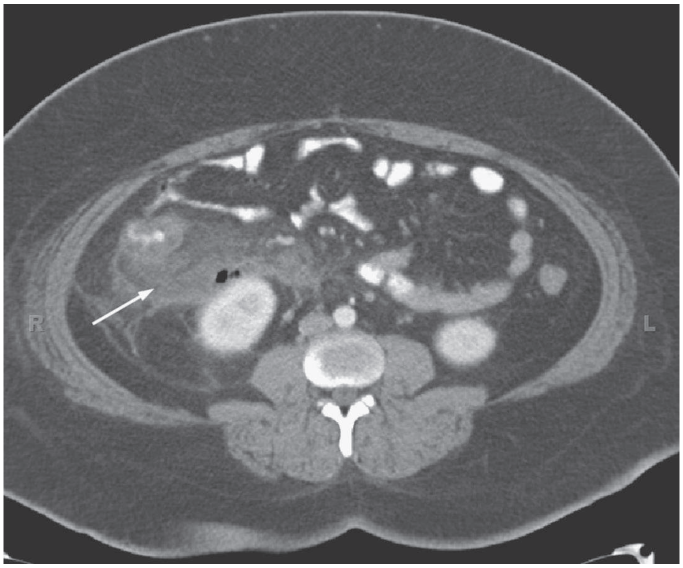

FIGURE 342-3 Computed tomography with oral and intravenous contrast of acute appendicitis. There is thickening of the wall of the appendix and periappendiceal stranding (*arrow*).

Immunocompromised patients may present with only mild tenderness and may have many other disease processes in their differential diagnosis, including atypical infections from mycobacteria, *Cytomegalovirus*, or other fungi. Enterocolitis is a concern and may be present in patients who present with abdominal pain, fever, and neutropenia due to chemotherapy. CT imaging may be very helpful, although it is important not to be overly cautious and delay operative intervention for those patients who are believed to have appendicitis.

TREATMENT

Acute Appendicitis

In the absence of contraindications, most patients who have strongly suggestive medical histories and physical examinations

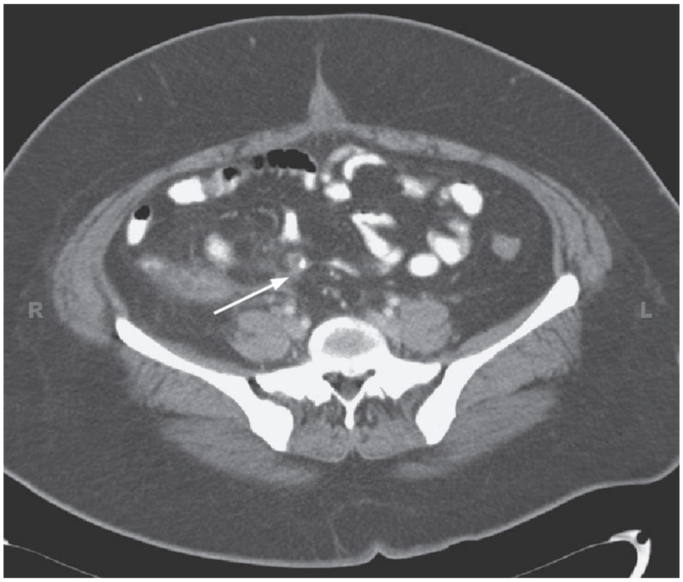

FIGURE 342-4 Appendiceal fecalith (*arrow*).

with supportive laboratory findings are candidates for appendectomy. Certainly, in some instances, imaging studies are not mandatory but are often obtained before surgical consultation is requested. Imaging and close observation are appropriate in patients whose evaluations are suggestive but not convincing.

Of course, CT may accurately indicate the other intraabdominal processes that warrant intervention. Whenever the diagnosis is uncertain, it is prudent to observe the patient and repeat the abdominal examination over 6–8 h. Any evidence of progression is an indication for operation. Narcotics can be given to patients with severe discomfort after an initial, thorough examination.

All patients should be fully prepared for surgery and have any fluid and electrolyte abnormalities corrected. Either laparoscopic or open appendectomy is a satisfactory choice for patients with uncomplicated appendicitis, although most procedures are now performed in a minimally invasive fashion to the patient's benefit in terms of recovery time and fewer overall potential complications. Endoscopic treatment for patients with uncomplicated appendicitis is being evaluated for efficacy.

Management of those who present with a mass representing a phlegmon or abscess can be more difficult. Such patients are most commonly treated with broad-spectrum antibiotics, percutaneous drainage especially if an abscess is noted >3 cm in diameter, parenteral fluids, and bowel rest. If they appear to respond to conservative management, the appendix can then be more safely removed 6–12 weeks later when inflammation has diminished.

A laparoscopic approach may also be useful when the exact diagnosis is uncertain. A laparoscopic approach may also facilitate exposure in those who are very obese. Absent complications, most patients can be discharged within 24–40 h of operation. The most common postoperative complications are fever and leukocytosis. Continuation of these findings beyond 5 days should raise concern for the presence of an intraabdominal abscess. The mortality rate for uncomplicated, nonperforated appendicitis is 0.1–0.5%, which approximates the overall risk of general anesthesia. The mortality rate for perforated appendicitis or other complicated disease is much higher, ranging from 3% overall to as high as 15% in the elderly.

ACUTE PERITONITIS

Acute peritonitis, or inflammation of the visceral and parietal peritoneum, is most often but not always infectious in origin, resulting from perforation of a hollow viscus. This is called *secondary peritonitis*, as opposed to *primary* or *spontaneous peritonitis*, when a specific intraabdominal source cannot be identified. In either instance, the inflammation can be localized or diffuse.

ETIOLOGY

Infective organisms may contaminate the peritoneal cavity after spillage from a hollow viscus, because of a penetrating wound of the abdominal wall, or because of the introduction of a foreign object like a peritoneal dialysis catheter or port that becomes infected. Secondary peritonitis most commonly results from perforation of the appendix, colonic diverticula, or the stomach and duodenum. It may also occur as a complication of bowel infarction or incarceration, cancer, inflammatory bowel disease, and intestinal obstruction or volvulus. Conditions that may cause secondary bacterial peritonitis and their mechanisms are listed in **Table 342-5**. Over 90% of the cases of primary or spontaneous bacterial peritonitis occur in patients with ascites or hypoproteinemia (<1 g/L).

Aseptic peritonitis is most commonly caused by the abnormal presence of physiologic fluids such as gastric juice, bile, pancreatic enzymes, blood, or urine. It can also be caused by the effects of normally sterile foreign bodies such as surgical sponges or instruments. More rarely, it occurs as a complication of systemic diseases such as lupus erythematosus, porphyria, and familial Mediterranean fever. The chemical irritation caused by stomach acid and activated pancreatic enzymes is extreme, and secondary bacterial infection may occur.

TABLE 342-5 Conditions Leading to Secondary Bacterial Peritonitis

Bowel perforation	**Perforation or leakage of other organs**
Appendicitis	Biliary leakage (e.g., after liver biopsy)
Anastomotic leakage	Cholecystitis
Adhesion	Intraperitoneal bleeding
Diverticulitis	Pancreatitis
Iatrogenic (including endoscopic perforation)	Salpingitis
Ingested foreign body	Urinary bladder
Inflammation	**Loss of peritoneal integrity**
Intussusception	Intraperitoneal chemotherapy
Neoplasms	Iatrogenic (e.g., postoperative foreign body)
Obstruction	Perinephric abscess
Peptic ulcer disease	Peritoneal dialysis or other indwelling devices
Strangulated hernia	Trauma
Vascular (including ischemia or embolus)	
Trauma (blunt or penetrating)	

CLINICAL FEATURES

The cardinal signs and symptoms of peritonitis are acute, typically severe, abdominal pain with tenderness and fever. How patients' complaints of pain are manifested depends on their overall physical health and whether the inflammation is diffuse or localized. Elderly and immunosuppressed patients may not respond as aggressively to the irritation. Diffuse, generalized peritonitis is most often recognized as diffuse abdominal tenderness with guarding, rigidity, and other evidence of parietal peritoneal irritation. Physical findings may only be identified in a specific region of the abdomen if the intraperitoneal inflammatory process is limited or otherwise contained as may occur in patients with uncomplicated appendicitis or diverticulitis. Bowel sounds are usually absent to hypoactive but are not reliable as a physical finding.

Most patients present with tachycardia and signs of volume depletion with hypotension. Laboratory testing typically reveals a significant leukocytosis, and patients may be severely acidotic. Radiographic studies may show dilatation of the bowel and associated bowel wall edema. Free air or other evidence of leakage requires attention and could represent a surgical emergency. In stable patients in whom ascites is present, diagnostic paracentesis is indicated, where the fluid is tested for protein and lactate dehydrogenase and the cell count is measured.

THERAPY AND PROGNOSIS

Whereas mortality rates can be <10% for reasonably healthy patients with relatively uncomplicated, localized peritonitis, mortality rates >40% have been reported for the elderly or immunocompromised. Successful treatment depends on correcting any electrolyte abnormalities, restoration of fluid volume and stabilization of the cardiovascular system, appropriate antibiotic therapy, and surgical correction of any underlying abnormalities.

ACKNOWLEDGMENT

The wisdom and expertise of Dr. William Silen are gratefully acknowledged in this updated chapter on acute appendicitis and peritonitis.

FURTHER READING

Andersson RE: Short-term complications and long-term morbidity of laparoscopic and open appendicectomy in a national cohort. Br J Surg 101:1135, 2014.

Buckius MT et al: Changing epidemiology of acute appendicitis in the United States: Study period 1993–2008. J Surg Res 175:185, 2012.

CODA Collaborative: A randomized trial comparing antibiotics with appendectomy for appendicitis. N Engl J Med 383:1907, 2020.

Di Saverio S et al: Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. World J Emerg Surg 15:27, 2020.

Drake FT et al: Time to appendectomy and risk of perforation in acute appendicitis. JAMA Surg 149:837, 2014.

Flum DR: Acute appendicitis—appendectomy of the “antibiotics first” strategy. N Engl J Med 372:1937, 2015.

Khan S et al: Endoscopic retrograde appendicitis therapy: Is it really a need of the hour. Ann Surg 277:e1, 2023.

Moris D et al: Diagnosis and management of acute appendicitis adults. JAMA 326:2299, 2021.

Ohle R et al: The Alvarado score for predicting acute appendicitis: A systematic review. BMC Med 9:139, 2011.

Talan DA, DiSaverio S: Treatment of acute uncomplicated appendicitis. N Engl J Med 385:1116, 2021.

Vons C et al: Amoxicillin plus clavulanic acid versus appendicectomy for treatment of acute uncomplicated appendicitis: An open-label, non-inferiority, randomised controlled trial. Lancet 377:1573, 2011.

Section 2 Nutrition

343

Nutrient Requirements and Dietary Assessment

Johanna T. Dwyer

Nutrients are substances that are essential but not synthesized in sufficient amounts in the body and therefore must be supplied by the diet. Nutrient requirements for groups of healthy persons have been determined experimentally. The absence of essential nutrients leads to growth impairment, organ and metabolic dysfunction, and failure to maintain nitrogen balance or adequate status of protein and other nutrients. For good health, we require energy-providing nutrients (protein, fat, and carbohydrate), vitamins, minerals, and water. Requirements for organic nutrients include 9 essential amino acids, several fatty acids, glucose, 4 fat-soluble vitamins, 10 water-soluble vitamins, dietary fiber, and choline. The inorganic nutrients include 4 minerals, 7 trace minerals, 3 electrolytes, and the ultra trace elements that must also be supplied by diet. Individuals of different ages and physiologic states differ in the amounts of nutrients they require.

Conditionally essential nutrients are not required in the diet but must be supplied to certain individuals such as those with genetic defects; pathologies such as infection, disease, or trauma with nutritional implications; and some developmentally immature infants who need them because they do not synthesize inositol, taurine, arginine, and/or glutamine in adequate amounts. Many other organic and inorganic bioactive compounds that are present in foods and dietary supplements may also have health effects, including pesticides, heavy metals like lead, phytochemicals, zoochemicals, and microbial products.

■ ESSENTIAL NUTRIENT REQUIREMENTS

Energy The estimated energy requirement (EER) is the predicted average requirement to maintain energy balance and health in an adult of a defined age, sex, weight, height, level of physical activity, and life stage. For weight to remain stable, energy intake must match total energy expenditure (TEE). The major components of TEE are resting energy expenditure (REE) and the physical activity level (PAL), and minor components include the energy cost of metabolizing food (thermic effect of food, or specific dynamic action) and cold-induced (shivering) thermogenesis. The average energy intake is ~2600 kcal/d for American men and ~1800 kcal/d for American women, but individuals vary with body size and activity level. Although estimates introduce considerable error, energy intakes are usually calculated from formulas rather than measuring TEE directly. In individuals whose weights are stable, for males, REE = 900 + 10m, and for females, REE = 700 + 7m, where m is mass in kilograms. The calculated REE is then multiplied by the appropriate PAL to account for physical activity that ranges from 1.4 (inactive) for individuals who perform only essential activities of daily living (e.g., 30 minutes walking and 90 minutes light to moderate activity) to 1.6 (low active), 1.75 (active), and 2.05 (very active). The result is the estimated energy (EER) or intake. When describing their PALs, sedentary, inactive, and low active individuals tend to overestimate, and so in practice, use of PALs of 1.2 to 1.4 may be closer to their actual activity levels. The EER provides a rough target for planning caloric needs for a person of a certain age, sex, weight, height, and physical activity level who is neither gaining nor losing weight. **For further discussion of energy balance in health and disease, see Chap. 345.**

Protein Dietary protein consists of both “essential” (e.g., not synthesized endogenously and must be supplied by diet) and “nonessential” (e.g., synthesized endogenously or obtained from diet) amino acids that are all required for protein synthesis. The nine essential amino acids are histidine, isoleucine, leucine, lysine, methionine/cystine, phenylalanine/tyrosine, threonine, tryptophan, and valine. Several amino acids, such as alanine, arginine, aspartic acid, glutamic acid, glutamine, and glycine can also be converted to glucose and used for energy and gluconeogenesis. When energy intake is inadequate, protein intake must be increased, because ingested amino acids are diverted into pathways of glucose synthesis and oxidation. In extreme energy deprivation, protein-calorie malnutrition may ensue **(Chap. 345)**.

For adults, the recommended dietary allowance (RDA) for protein is ~0.8 g/kg desirable body mass per day, assuming that energy needs are met and that the protein is of relatively high biological value. Current recommendations for a healthy diet call for at least 10–14% of calories from protein. Most American diets provide at least those amounts. Biologic value tends to be highest for animal proteins, followed by proteins from legumes (beans), cereals (rice, wheat, corn), and roots. Combinations of plant proteins that complement one another in their essential amino acid profiles or combinations of animal and plant proteins can increase biologic value and lower total protein intakes necessary to meet requirements. In healthy people with adequate diets, the timing of protein intake over the course of the day has little effect.

Protein needs increase during growth, pregnancy, lactation, and rehabilitation after injury or undernutrition. Tolerance to dietary protein is decreased in renal insufficiency (with consequent uremia) and in liver failure. Usual protein intakes can precipitate encephalopathy in patients with cirrhosis of the liver.

Fat and Carbohydrate Carbohydrate (4 kcal/g), fat (9 kcal/g), and protein (4 kcal/g) all provide energy. So does alcohol (ethanol) (7 kcal/g), but it is not a nutrient. Fats are a concentrated source of energy and constitute, on average, 34% of calories in U.S. diets. However, for optimal health, fat intake should total no more than 30% of calories. Saturated fat and trans-fat should be limited to <10% of calories and polyunsaturated fats to <10% of calories, with monounsaturated fats accounting for the remainder of fat intake. At least 45–55% of total calories should be derived from carbohydrates, with <10% and preferably <6% from added sugars. The brain requires ~100 g of glucose per day for fuel; other tissues use ~50 g/d. Some tissues (e.g., brain and red blood cells) rely on glucose supplied either exogenously or from muscle proteolysis. Over time, during hypocaloric states, adaptations that lower carbohydrate needs are possible.

Water For adults, 1–1.5 mL of water per kilocalorie of energy expenditure is sufficient under usual conditions to allow for normal variations in physical activity, sweating, and the diet’s solute load. Water losses include 50–100 mL/d in the feces; 500–1000 mL/d by evaporation or exhalation; and, depending on the renal solute load, ≥1000 mL/d in the urine. If external losses increase, intakes must increase accordingly to avoid underhydration. Fever increases water losses by ~200 mL/d per °C; diarrheal losses vary but may be as great as 5 L/d in severe diarrhea. Heavy sweating, vigorous exercise, and vomiting also increase water losses. When renal function is normal and solute intakes are adequate, the kidneys can adjust to increased water intake by excreting up to 18 L of excess water per day **(Chap. 393)**. However, obligatory urine outputs can compromise hydration status when there is inadequate water intake or when losses increase in disease or kidney damage.

Infants have high requirements for water because of their large surface area to volume ratios, their inability to communicate their thirst, and the limited capacity of the immature kidney to handle high renal solute loads. Increased water needs during pregnancy are ~30 mL/d. During lactation, milk production increases daily water requirements so that ~1000 mL of additional water is needed, or 1 mL for each milliliter of milk produced. Special attention must also be paid to the water needs of the elderly, who have reduced total-body water, blunted thirst sensation, and frequently use medications such as diuretics.

Other Nutrients **See Chap. 344 for detailed descriptions of vitamins and minerals.**

DIETARY REFERENCE INTAKES AND RDAS

Fortunately, human life and well-being can be maintained within a fairly wide range of most nutrient intakes. However, the capacity for adaptation is not infinite—too much, as well as too little, intake of a nutrient can have adverse effects itself or decrease health benefits conferred by another nutrient. Therefore, benchmark recommendations regarding nutrient intakes have been developed to guide menu development and clinical practice. They are referred to in the United States and Canada as the *dietary reference intakes* (DRIs) and are set by National Academy of Medicine's (formerly Institute of Medicine [IOM]) Food and Nutrition Board and Health Canada. The recommendations apply to the general population, including those with or at risk for chronic diseases, overweight, and obesity. Exceptions may be needed for individuals with severe comorbidities, diseases, disabilities, or use of medications known to alter energy or some nutrient requirements. The DRIs supplanted the recommended dietary allowances (RDAs)—the single reference values used in the United States until the early 1990s. DRIs now include an *estimated average requirement* (EAR) as well as other reference values used for dietary planning: the RDA, the *adequate intake* (AI), the *chronic disease risk reduction intake* (CDRR), and the tolerable *upper level* (UL). The DRIs also include acceptable macronutrient distribution ranges (AMDRs) for protein, fat, and carbohydrate. The types of evidence and criteria used to establish nutrient requirements vary by nutrient, age, and physiologic group. The current DRIs for vitamins and elements are provided in **Tables 343-1** and **343-2**, respectively. **Table 343-3** provides DRIs for water and macronutrients. Energy has neither an RDA nor a UL because intakes above or below requirements lead to weight gain or loss; **EERs are used instead and are discussed in Chap. 345 on energy balance in health and disease.**

Estimated Average Requirement (EAR) When florid manifestations of the classic dietary-deficiency diseases such as rickets (deficiency of vitamin D and calcium), scurvy (deficiency of vitamin C), xerophthalmia (deficiency of vitamin A), and protein-calorie malnutrition were common, nutrient adequacy was inferred from the absence of their clinical deficiency signs. Later, biochemical and other earlier changes were used that appeared long before the deficiency was clinically apparent. Today, criteria of adequacy are based on biologic markers when they are available, and to sensitive biochemical, physiologic, or behavioral tests that reflect early changes in regulatory processes; maintenance of body stores of nutrients; or, if available, the amount of a nutrient that minimizes the risk of chronic degenerative disease. However, relevant biomarkers of disease risk are often not available, and the long time lags between intake and disease outcomes further complicate the picture.

The EAR is the amount of a nutrient estimated to be adequate for half of the healthy individuals of a specific age and sex. It is not an effective estimate of nutrient adequacy in individuals because it is a median requirement for a group, with 50% of individuals in a group falling below and 50% falling above the requirement. Thus, a person with a usual intake at the EAR has a 50% risk of inadequate intake. For these reasons, the other standards described below are more useful for clinical purposes.

Recommended Dietary Allowances The RDA is the nutrient intake goal for planning diets of individuals. It is the average daily dietary intake level that meets the nutrient requirements of nearly all healthy persons of a specific sex, age, life stage, or physiologic condition (e.g., pregnancy or lactation). The RDA is set two standard deviations above the EAR to ensure that the needs of virtually any individual are met. The risk of dietary inadequacy increases as one's intake falls below the RDA. However, the RDA is an overly generous criterion for evaluating nutrient adequacy because, by definition, the RDA exceeds the actual requirements of all but ~2–3% of the population. Therefore, many people whose intake falls below the RDA are still getting enough of the nutrient.

The RDAs are used to formulate food guides such as the U.S. Department of Agriculture's (USDA) MyPlate Plan for individuals (*https://www.choosemyplate.gov/resources/MyPlatePlan*), as well as to create food-exchange lists for therapeutic diet planning. An online tool, available at *https://www.nal.usda.gov/fnic/dri-calculator/*, allows health professionals to calculate individualized daily nutrient recommendations for dietary planning based on the DRIs.

The RDAs are also used as a standard for labels describing the nutritional content of foods and nutrient-containing dietary supplements. On food labels, the nutrient content in a food is stated by weight and/or as a percentage of the daily value (DV), a variant of the RDA used on the nutrition facts panel that, for an adult, represents the highest RDA for an adult consuming 2000 kcal.

Adequate Intake (AI) Some nutrients lack an EAR and so an RDA cannot be set. In this circumstance, the AI is used, based on observed or experimentally determined approximations of nutrient intakes in healthy people. In the DRIs, AIs rather than RDAs are proposed for nutrients consumed by infants (up to age 1 year) as well as for chromium, fluoride, manganese, sodium, potassium, pantothenic acid, biotin, choline, fiber, and water consumed by persons of all ages.

Tolerable Upper Levels (UL) Healthy individuals derive no established benefit from consuming nutrient levels above the RDA or AI. In fact, excessive nutrient intake can disturb body functions and cause acute, progressive, or permanent disabilities. The tolerable UL is the highest level of chronic (usually daily) nutrient intake that is unlikely to pose a risk of adverse health effects for most of the population. Data on the adverse effects of large amounts of many nutrients are unavailable or too limited to establish a UL and so the lack of a UL does *not* mean that the risk of adverse effects from high intake is nonexistent. Nutrient levels in commonly eaten foods rarely exceed the UL, although very highly fortified foods and dietary supplements provide more concentrated amounts of nutrients per serving and thus pose a potential risk of toxicity. Dietary supplements have Supplement Facts panels on labels that express the amount of nutrients present in absolute units or as the percentage of the DV provided per recommended serving size. Total nutrient intakes, including that from foods, supplements, and over-the-counter medications (e.g., antacids), should not exceed RDA levels.

Chronic Disease Risk Reduction Intake (CDRR) The CDRR is the level above which a reduction in intake is expected to lower chronic disease risk is the CDRR. For example, the sodium CDRR for adults is 2300 mg/d; this is the lowest level of intake for which there is sufficiently strong evidence to characterize a CDRR. CDRRs are not available for potassium or other nutrients, but note that the AI for potassium was recently reduced to 2500 mg/d.

Acceptable Macronutrient Distribution Ranges (AMDRs) AMDRs are estimated ranges from epidemiologic data of energy-providing macronutrient intakes (protein, carbohydrate, and fat) considered to be healthful. These ranges are 10–35% of calories for protein, 20–35% of calories for fat, and 45–65% of calories for carbohydrate. Alcohol, which also provides energy, is not a nutrient; therefore, no recommendations for it are provided. The latest DRI committee

TABLE 343-1 Dietary Reference Intakes (DRIs): Recommended Dietary Allowances and Adequate Intakes for Vitamins

LIFE-STAGE GROUP	VITAMIN A (μg/d)[a]	VITAMIN C (mg/d)	VITAMIN D (μg/d)[b,c]	VITAMIN E (mg/d)[d]	VITAMIN K (μg/d)	THIAMIN (mg/d)	RIBOFLAVIN (mg/d)	NIACIN (mg/d)[e]	VITAMIN B_6 (mg/d)	FOLATE (μg/d)[f]	VITAMIN B_{12} (μg/d)	PANTOTHENIC ACID (mg/d)	BIOTIN (μg/d)	CHOLINE (mg/d)[g]
Infants														
Birth to 6 mo	400*	40*	10	4*	2.0*	0.2*	0.3*	2*	0.1*	65*	0.4*	1.7*	5*	125*
6–12 mo	500*	50*	10	5*	2.5*	0.3*	0.4*	4*	0.3*	80*	0.5*	1.8*	6*	150*
Children														
1–3 y	**300**	**15**	**15**	**6**	30*	**0.5**	**0.5**	**6**	**0.5**	**150**	**0.9**	2*	8*	200*
4–8 y	**400**	**25**	**15**	**7**	55*	**0.6**	**0.6**	**8**	**0.6**	**200**	**1.2**	3*	12*	250*
Males														
9–13 y	**600**	**45**	**15**	**11**	60*	**0.9**	**0.9**	**12**	**1.0**	**300**	**1.8**	4*	20*	375*
14–18 y	**900**	**75**	**15**	**15**	75*	**1.2**	**1.3**	**16**	**1.3**	**400**	**2.4**	5*	25*	550*
19–30 y	**900**	**90**	**15**	**15**	120*	**1.2**	**1.3**	**16**	**1.3**	**400**	**2.4**	5*	30*	550*
31–50 y	**900**	**90**	**15**	**15**	120*	**1.2**	**1.3**	**16**	**1.3**	**400**	**2.4**	5*	30*	550*
51–70 y	**900**	**90**	**15**	**15**	120*	**1.2**	**1.3**	**16**	**1.7**	**400**	**2.4**[h]	5*	30*	550*
>70 y	**900**	**90**	**20**	**15**	120*	**1.2**	**1.3**	**16**	**1.7**	**400**	**2.4**[h]	5*	30*	550*
Females														
9–13 y	**600**	**45**	**15**	**11**	60*	**0.9**	**0.9**	**12**	**1.0**	**300**	**1.8**	4*	20*	375*
14–18 y	**700**	**65**	**15**	**15**	75*	**1.0**	**1.0**	**14**	**1.2**	**400**[i]	**2.4**	5*	25*	400*
19–30 y	**700**	**75**	**15**	**15**	90*	**1.1**	**1.1**	**14**	**1.3**	**400**[i]	**2.4**	5*	30*	425*
31–50 y	**700**	**75**	**15**	**15**	90*	**1.1**	**1.1**	**14**	**1.3**	**400**[i]	**2.4**	5*	30*	425*
51–70 y	**700**	**75**	**15**	**15**	90*	**1.1**	**1.1**	**14**	**1.5**	**400**	**2.4**[h]	5*	30*	425*
>70 y	**700**	**75**	**20**	**15**	90*	**1.1**	**1.1**	**14**	**1.5**	**400**	**2.4**[h]	5*	30*	425*
Pregnant Women														
14–18 y	**750**	**80**	**15**	**15**	75*	**1.4**	**1.4**	**18**	**1.9**	**600**[j]	**2.6**	6*	30*	450*
19–30 y	**770**	**85**	**15**	**15**	90*	**1.4**	**1.4**	**18**	**1.9**	**600**[j]	**2.6**	6*	30*	450*
31–50 y	**770**	**85**	**15**	**15**	90*	**1.4**	**1.4**	**18**	**1.9**	**600**[j]	**2.6**	6*	30*	450*
Lactating Women														
14–18 y	**1200**	**115**	**15**	**19**	75*	**1.4**	**1.6**	**17**	**2.0**	**500**	**2.8**	7*	35*	550*
19–30 y	**1300**	**120**	**15**	**19**	90*	**1.4**	**1.6**	**17**	**2.0**	**500**	**2.8**	7*	35*	550*
31–50 y	**1300**	**120**	**15**	**19**	90*	**1.4**	**1.6**	**17**	**2.0**	**500**	**2.8**	7*	35*	550*

Note: This table (taken from the DRI reports; see *www.nap.edu*) presents recommended dietary allowances (RDAs) in **bold type** and adequate intakes (AIs) in ordinary type followed by an asterisk (*). An RDA is the average daily dietary intake level sufficient to meet the nutrient requirements of nearly all healthy individuals (97–98%) in a group. The RDA is calculated from an estimated average requirement (EAR). If sufficient scientific evidence is not available to establish an EAR and thus to calculate an RDA, an AI is usually developed. For healthy breast-fed infants, an AI is the mean intake. The AI for other life-stage and sex-specific groups is believed to cover the needs of all healthy individuals in those groups, but lack of data or uncertainty in the data makes it impossible to specify with confidence the percentage of individuals covered by this intake.

[a] As retinol activity equivalents (RAEs). 1 RAE = 1 μg retinol, 12 μg β-carotene, 24 μg α-carotene, or 24 μg β-cryptoxanthin. The RAE for dietary provitamin A carotenoids is twofold greater than the retinol equivalent (RE), whereas the RAE for preformed vitamin A is the same as the RE. [b] As cholecalciferol. 1 μg cholecalciferol = 40 IU vitamin D. [c] Under the assumption of minimal sunlight. [d] As α-tocopherol. α-Tocopherol includes *RRR*-α-tocopherol, the only form of α-tocopherol that occurs naturally in foods, and the *2R*-stereoisomeric forms of α-tocopherol (*RRR-*, *RSR-*, *RRS-*, and *RSS*-α-tocopherol) that occur in fortified foods and supplements. It does not include the *2S*-stereoisomeric forms of α-tocopherol (*SRR-*, *SSR-*, *SRS-*, and *SSS*-α-tocopherol) also found in fortified foods and supplements. [e] As niacin equivalents (NEs). 1 mg of niacin = 60 mg of tryptophan; 0–6 months = preformed niacin (not NE). [f] As dietary folate equivalents (DFEs). 1 DFE = 1 μg food folate = 0.6 μg of folic acid from fortified food or as a supplement consumed with food = 0.5 μg of a supplement taken on an empty stomach. [g] Although AIs have been set for choline, there are few data to assess whether a dietary supply of choline is needed at all stages of the life cycle, and it may be that the choline requirement can be met by endogenous synthesis at some of these stages. [h] Because 10–30% of older people may malabsorb food-bound B_{12}, it is advisable for those >50 years of age to meet their RDA mainly by consuming foods fortified with B_{12} or a supplement containing B_{12}. [i] In view of evidence linking inadequate folate intake with neural tube defects in the fetus, it is recommended that all women capable of becoming pregnant consume 400 μg of folate from supplements or fortified foods in addition to intake of food folate from a varied diet. [j] It is assumed that women will continue consuming 400 μg from supplements or fortified food until their pregnancy is confirmed and they enter prenatal care, which ordinarily occurs after the end of the periconceptional period—the critical time for formation of the neural tube.

Source: National Academies of Sciences, Engineering, and Medicine. 2019. Dietary Reference Intakes for Sodium and Potassium. *https://doi.org/10.17226/25353*. Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies.

TABLE 343-2 Dietary Reference Intakes (DRIs): Recommended Dietary Allowances and Adequate Intakes for Elements

LIFE-STAGE GROUP	CALCIUM (mg/d)	CHROMIUM (μg/d)	COPPER (μg/d)	FLUORIDE (mg/d)	IODINE (μg/d)	IRON (mg/d)	MAGNESIUM (mg/d)	MANGANESE (mg/d)	MOLYBDENUM (μg/d)	PHOSPHORUS (mg/d)	SELENIUM (μg/d)	ZINC (mg/d)	POTASSIUM (g/d)	SODIUM (g/d)	CHLORIDE (g/d)
Infants															
Birth to 6 mo	200*	0.2*	200*	0.01*	110*	0.27*	30*	0.003*	2*	100*	15*	2*	0.4*	0.12*	0.18*
6–12 mo	260*	5.5*	220*	0.5*	130*	**11**	75*	0.6*	3*	275*	20*	**3**	0.7*	0.37*	0.57*
Children															
1–3 y	**700**	11*	**340**	0.7*	**90**	**7**	**80**	1.2*	**17**	**460**	**20**	**3**	3.0*	1.0*	1.5*
4–8 y	**1000**	15*	**440**	1*	**90**	**10**	**130**	1.5*	**22**	**500**	**30**	**5**	3.8*	1.2*	1.9*
Males															
9–13 y	**1300**	25*	**700**	2*	**120**	**8**	**240**	1.9*	**34**	**1250**	**40**	**8**	4.5*	1.5*	2.3*
14–18 y	**1300**	35*	**890**	3*	**150**	**11**	**410**	2.2*	**43**	**1250**	**55**	**11**	4.7*	1.5*	2.3*
19–30 y	**1000**	35*	**900**	4*	**150**	**8**	**400**	2.3*	**45**	**700**	**55**	**11**	4.7*	1.5*	2.3*
31–50 y	**1000**	35*	**900**	4*	**150**	**8**	**420**	2.3*	**45**	**700**	**55**	**11**	4.7*	1.5*	2.3*
51–70 y	**1000**	30*	**900**	4*	**150**	**8**	**420**	2.3*	**45**	**700**	**55**	**11**	4.7*	1.3*	2.0*
>70 y	**1200**	30*	**900**	4*	**150**	**8**	**420**	2.3*	**45**	**700**	**55**	**11**	4.7*	1.2*	1.8*
Females															
9–13 y	**1300**	21*	**700**	2*	**120**	**8**	**240**	1.6*	**34**	**1250**	**40**	**8**	4.5*	1.5*	2.3*
14–18 y	**1300**	24*	**890**	3*	**150**	**15**	**360**	1.6*	**43**	**1250**	**55**	**9**	4.7*	1.5*	2.3*
19–30 y	**1000**	25*	**900**	3*	**150**	**18**	**310**	1.8*	**45**	**700**	**55**	**8**	4.7*	1.5*	2.3*
31–50 y	**1000**	25*	**900**	3*	**150**	**18**	320	1.8*	**45**	**700**	**55**	**8**	4.7*	1.5*	2.3*
51–70 y	**1200**	20*	**900**	3*	**150**	**8**	**320**	1.8*	**45**	**700**	**55**	**8**	4.7*	1.3*	2.0*
>70 y	**1200**	20*	**900**	3*	**150**	**8**	**320**	1.8*	**45**	**700**	**55**	**8**	4.7*	1.2*	1.8*
Pregnant Women															
14–18 y	**1300**	29*	**1000**	3*	**220**	**27**	**400**	2.0*	**50**	**1250**	**60**	**12**	4.7*	1.5*	2.3*
19–30 y	**1000**	30*	**1000**	3*	**220**	**27**	**350**	2.0*	**50**	**700**	**60**	**11**	4.7*	1.5*	2.3*
31–50 y	**1000**	30*	**1000**	3*	**220**	**27**	**360**	2.0*	**50**	**700**	**60**	**11**	4.7*	1.5*	2.3*
Lactating Women															
14–18 y	**1300**	44*	**1300**	3*	**290**	**10**	**360**	2.6*	**50**	**1250**	**70**	**13**	5.1*	1.5*	2.3*
19–30 y	**1000**	45*	**1300**	3*	**290**	**9**	**310**	2.6*	**50**	**700**	**70**	**12**	5.1*	1.5*	2.3*
31–50 y	**1000**	45*	**1300**	3*	**290**	**9**	**320**	2.6*	**50**	**700**	**70**	**12**	5.1*	1.5*	2.3*

Note: This table (taken from the DRI reports; see *www.nap.edu*) presents recommended dietary allowances (RDAs) in **bold type** and adequate intakes (AIs) in ordinary type followed by an asterisk (*). An RDA is the average daily dietary intake level sufficient to meet the nutrient requirements of nearly all healthy individuals (97–98%) in a group. The RDA is calculated from an estimated average requirement (EAR). If sufficient scientific evidence is not available to establish an EAR and thus to calculate an RDA, an AI is usually developed. For healthy breast-fed infants, an AI is the mean intake. The AI for other life-stage and sex-specific groups is believed to cover the needs of all healthy individuals in those groups, but lack of data or uncertainty in the data makes it impossible to specify with confidence the percentage of individuals covered by this intake.

Source: National Academies of Sciences, Engineering, and Medicine. 2019. Dietary Reference Intakes for Sodium and Potassium. *https://doi.org/10.17226/25353*. Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies.

TABLE 343-3 Dietary Reference Intakes (DRIs): Recommended Dietary Allowances and Adequate Intakes for Total Water and Macronutrients

LIFE-STAGE GROUP	TOTAL WATER[a] (L/d)	CARBOHYDRATE (g/d)	TOTAL FIBER (g/d)	FAT (g/d)	LINOLEIC ACID (g/d)	α-LINOLENIC ACID (g/d)	PROTEIN[b] (g/d)
Infants							
Birth to 6 mo	0.7*	60*	ND[c]	31*	4.4*	0.5*	9.1*
6–12 mo	0.8*	95*	ND	30*	4.6*	0.5*	**11.0**
Children							
1–3 y	1.3*	**130**	19*	ND	7*	0.7*	**13**
4–8 y	1.7*	**130**	25*	ND	10*	0.9*	**19**
Males							
9–13 y	2.4*	**130**	31*	ND	12*	1.2*	**34**
14–18 y	3.3*	**130**	38*	ND	16*	1.6*	**52**
19–30 y	3.7*	**130**	38*	ND	17*	1.6*	**56**
31–50 y	3.7*	**130**	38*	ND	17*	1.6*	**56**
51–70 y	3.7*	**130**	30*	ND	14*	1.6*	**56**
>70 y	3.7*	**130**	30*	ND	14*	1.6*	**56**
Females							
9–13 y	2.1*	**130**	26*	ND	10*	1.0*	**34**
14–18 y	2.3*	**130**	26*	ND	11*	1.1*	**46**
19–30 y	2.7*	**130**	25*	ND	12*	1.1*	**46**
31–50 y	2.7*	**130**	25*	ND	12*	1.1*	**46**
51–70 y	2.7*	**130**	21*	ND	11*	1.1*	**46**
>70 y	2.7*	**130**	21*	ND	11*	1.1*	**46**
Pregnant Women							
14–18 y	3.0*	**175**	28*	ND	13*	1.4*	**71**
19–30 y	3.0*	**175**	28*	ND	13*	1.4*	**71**
31–50 y	3.0*	**175**	28*	ND	13*	1.4*	**71**
Lactating Women							
14–18	3.8*	**210**	29*	ND	13*	1.3*	**71**
19–30 y	3.8*	**210**	29*	ND	13*	1.3*	**71**
31–50 y	3.8*	**210**	29*	ND	13*	1.3*	**71**

Note: This table (taken from the DRI reports; see *www.nap.edu*) presents recommended dietary allowances (RDAs) in **bold type** and adequate intakes (AIs) in ordinary type followed by an asterisk (*). An RDA is the average daily dietary intake level sufficient to meet the nutrient requirements of nearly all healthy individuals (97–98%) in a group. The RDA is calculated from an estimated average requirement (EAR). If sufficient scientific evidence is not available to establish an EAR and thus to calculate an RDA, an AI is usually developed. For healthy breast-fed infants, an AI is the mean intake. The AI for other life-stage and sex-specific groups is believed to cover the needs of all healthy individuals in those groups, but lack of data or uncertainty in the data make it impossible to specify with confidence the percentage of individuals covered by this intake.

[a]Total water includes all water contained in food, beverages, and drinking water. [b]Based on grams of protein per kilogram of body weight for the reference body weight (e.g., for adults: 0.8 g/kg body weight for the reference body weight). [c]Not determined.

Source: National Academies of Sciences, Engineering, and Medicine. 2019. Dietary Reference Intakes for Sodium and Potassium. *https://doi.org/10.17226/25353.* Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies.

recommended that the CDRR method be used instead of the AMDR for establishing DRIs for macronutrients in the future.

FACTORS ALTERING NUTRIENT NEEDS

The DRIs are affected by age, sex, growth rate, pregnancy, lactation, physical activity level, concomitant diseases, drugs, and dietary composition.

Physiologic Factors Growth, strenuous physical activity, pregnancy, and lactation all increase needs for energy and several essential nutrients. Energy needs rise during pregnancy due to fetal growth demands and increased energy required for milk production during lactation. Energy needs decrease with loss of lean body mass, the major determinant of REE. The energy needs of older persons, especially those aged >70 years, tend to be lower than those of younger persons because lean tissue, physical activity, and health often decline with age.

Dietary Composition Dietary composition affects the biologic availability and use of nutrients. For example, iron absorption may be impaired by large amounts of calcium or lead; likewise, non-heme iron uptake may be impaired by a lack of ascorbic acid and amino acids in the meal. Bodily protein may be decreased when essential amino acids are not present in sufficient amounts—a rare scenario in U.S. diets. Animal foods, such as milk, eggs, and meat, have high biologic values, with most of the needed amino acids present in adequate amounts. Plant proteins in corn (maize), soy, rice, and wheat have lower biologic values and must be combined with other plant or animal proteins or fortified with the amino acids that are deficient to achieve optimal use by the body.

Route of Intake The RDAs apply only to oral intakes. When nutrients are administered parenterally, similar values can sometimes be used for amino acids, glucose (carbohydrate), fats, sodium, chloride, potassium, and most vitamins because their intestinal absorption rate is nearly 100%. However, the oral bioavailability of most mineral elements may be only half that obtained by parenteral administration. For some nutrients that are not readily stored in the body or that cannot be stored in large amounts, timing of administration may also be important. For example, amino acids cannot be used for protein synthesis if they are not supplied together intravenously; instead, they will be used for energy production. In contrast, among healthy individuals eating adequate diets, the distribution of protein intake over the course of the day has little effect on health.

Disease Dietary deficiency diseases include protein-calorie malnutrition, iron-deficiency anemia, goiter (due to iodine deficiency), rickets and osteomalacia (vitamin D deficiency), xeropthalmia (vitamin A deficiency), megaloblastic anemia (vitamin B_{12} or folic acid deficiency), scurvy (vitamin C/ascorbic acid deficiency), beriberi (thiamin deficiency), and pellagra (niacin and tryptophan deficiency) **(Chaps. 344 and 345)**. Each deficiency disease is characterized by imbalances at the cellular level between the supply of nutrients or energy and the body's nutritional needs for growth, maintenance, and other functions. Imbalances and excesses in nutrient intakes are recognized as risk factors for certain chronic degenerative diseases, such as saturated/trans-fat and cholesterol in coronary artery disease; sodium in hypertension; obesity in hormone-dependent cancers (endometrial and breast); and ethanol in alcohol addiction. These disorders are multifactorial in their etiology and pathogenesis, and diet is only one of the many risk factors for their occurrence. Osteoporosis, for example, is sometimes associated with calcium deficiency secondary to vitamin D deficiency, as well as with environment-related risk factors (e.g., smoking, sedentary lifestyle), physiology (e.g., estrogen deficiency), genetic determinants (e.g., defects in collagen metabolism), and drug use (chronic steroids and aromatase inhibitors) **(Chap. 423)**.

DIETARY ASSESSMENT

Nutrition assessment in clinical situations is an iterative process involving screening for malnutrition risk; assessing the diet and other biomarkers to determine if malnutrition is present and if so its possible causes; planning and implementing appropriate dietary intakes and medical nutrition therapy; and reassessing to make sure that intakes have been consumed. Disease states affecting the bioavailability, requirements, use, or excretion of specific nutrients may require specific measurements of certain nutrients or their biomarkers to assess status and ensure that nutrient replacement is adequate **(Chap. 344)**.

The Joint Commission, which accredits and certifies health care organizations in the United States, requires most health care facilities to perform nutrition screening for identifying possible malnutrition risk within 24 h after hospital admission. However, a single universally recognized or validated standard is not used yet. The Global Leadership Initiative on Malnutrition (GLIM) proposes that in acute care settings a tool be used that evaluates weight loss, low body mass index (BMI), reduced muscle mass, reduced food intake or absorption, and increased disease burden or inflammation. Other screens usually include these and also abnormal weight for height or body mass index (e.g., BMI <19 or >25); reported weight change (involuntary loss or gain of >5 kg in the past 6 months) **(Chap. 50)**; diagnoses with known nutritional implications (e.g., metabolic disease, any disease affecting the gastrointestinal tract, alcohol abuse); present therapeutic dietary prescription; chronic poor appetite; presence of dental, chewing, and swallowing problems; major food allergies or intolerances; need for assistance with preparing or shopping for food, eating, or other aspects of self-care; and social isolation. The nutritional status of hospitalized patients should be reassessed periodically—at least once every week.

Patients who exhibit frank malnutrition or a high risk of it on nutritional screening need more complete dietary assessment that varies with the clinical setting, the severity of the patient's illness, and the stability of the patient's condition.

Acute-Care Settings The goal is to identify and avoid inadequate intake, assure intake is appropriate, and gather enough information to determine if malnutrition is due to poor dietary intake or other causes and whether nutritional therapy is indicated **(Chap. 346)**. In acute-care settings, anorexia, various diseases, test procedures, and medications can further compromise dietary intake. Dietary assessment focuses on what patients are currently eating, whether or not they are able and willing to eat, and whether or not they experience any problems with eating. Simple observations may suffice to suggest inadequate oral intake. These include dietitians' and nurses' notes; observation of a patient's frequent refusal to eat or the amount of food served versus that eaten on trays; the frequent performance of tests and procedures that are likely to cause meals to be skipped; consumption limited to nutritionally inadequate diet orders (e.g., clear liquids or full liquids) for more than a few days; the occurrence of fever, dehydration, gastrointestinal distress, vomiting, diarrhea, or a comatose state; and the presence of diseases or use of treatments that involve any part of the alimentary tract. Acutely ill patients with diet-related diseases such as diabetes and chronic renal failure are in particular need of assessment because an inappropriate diet may exacerbate these conditions and adversely affect other therapies. Abnormal biochemical values (serum albumin levels <35 g/L [<3.5 mg/dL]; serum cholesterol levels <3.9 mmol/L [<150 mg/dL]) are nonspecific but may indicate a need for further nutritional assessment.

Most therapeutic diets offered in hospitals are calculated to meet individual nutrient requirements and the RDA *if they are eaten*. Exceptions include clear liquids, some full-liquid diets, and test diets (such as those adhered to in preparation for gastrointestinal procedures), which are inadequate in several nutrients and should not be used, if possible, for >24 h. However, because as much as half of the food served to hospitalized patients is not eaten, it cannot be assumed that the intakes of hospitalized patients are adequate. Dietary assessment should compare how much and what kinds of food the patient has consumed with the diet that has been provided. Major deviations in intakes of energy, protein, fluids, or other nutrients of special concern for the patient's illness should be noted and corrected, especially for long-staying patients.

Nutritional monitoring is especially important for patients who are very ill, who have extended lengths of hospital stay, or who are discharged home with diagnoses involving the gastrointestinal tract, end-stage renal disease, and insulin-dependent diabetes. Patients who are fed by enteral and parenteral routes and those receiving home chemotherapy infusions also require special nutritional assessment and monitoring by physicians and/or dietitians with certification in nutritional support **(Chap. 346)**.

Ambulatory Settings The aim of dietary assessment in the outpatient setting is to determine whether or not the patient's usual diet is a health risk in itself, if it contributes to existing chronic disease–related problems, and if medical nutrition therapy is required. The information is used to plan an eating pattern that fulfills therapeutic goals while ensuring patient adherence. Dietary assessment should review the adequacy of present and usual food intakes, including vitamin and mineral supplements, oral nutritional supplements, medical foods, other dietary supplements, medications, and alcohol, because all of these may affect the patient's nutritional status. Assessment should focus on the dietary constituents that are most likely to be involved or compromised by the specific diagnosis as well as on any comorbidities that are present. A diet history of several days of intake or a food-frequency questionnaire can be used and reviewed to provide a representation of the usual diet and assess the adequacy of the patient's habitual diet. The Automated Self-Administered 24-Hour Dietary Assessment Tool (ASA24) (which is a web-based, self-administered, 24-h recall tool), a diet history questionnaire (DHQ), a daily food checklist, and validated short dietary instruments are all available at the National Institute of Health's National Cancer Institute webpage at *https://epi.grants.cancer.gov/dietary-assessment/resources.html*. The Nutritools interactive website (*www.nutritools.org*) guides clinicians and researchers to choose among suitable validated dietary assessment tools used in North America and elsewhere. Assessment findings are used to craft dietary recommendations for changes in healthier directions. These may be as simple as a food guide, a food-exchange list, or more personalized and detailed recommendations by a dietitian that are tailored to the patient's food preferences and medical nutrition therapy needs. Therapeutic dietary prescriptions and menu plans for most diseases are available in the diet manuals of most hospitals and from the Academy of Nutrition and Dietetics for many other diseases. For persons on therapeutic diets, counseling using food-exchange lists such as those available from the American Diabetes Association for diabetes and from the Academy of Nutrition and Dietetics or the National Kidney Foundation's Council on Renal Nutrition for renal disease is helpful. Patients who follow ethnic or unusual dietary

TABLE 343-4 Choose My Plate: A Guide to Individualized Dietary Planning

DIETARY FACTOR, UNIT OF MEASURE (ADVICE)	EXAMPLES OF STANDARD PORTION SIZES AT INDICATED ENERGY LEVEL		
	LOWER: 1600 kcal	MODERATE: 2200 kcal	HIGHER: 2800 kcal
Fruits, cups (Focus on fruits.)	1.5	2	2.5
Vegetables, cups (Vary vegetables.)	2	3	3.5
Grains, oz eq (Make at least half of grains whole.)[a]	5	7	10
Protein foods, oz eq (Go lean with protein.)[b]	5	6	7
Dairy, cups or oz[c] (Choose calcium-rich foods.)	3	3	3
"Empty" calories, kcal[d]	120	260	400
Sodium, mg	<2300 at all energy levels		
Physical activity, min	At least 150 min vigorous physical activity per week at all energy levels		

Note: Oils (formerly listed with portions of 5, 6, and 8 teaspoons for the lower, moderate, and higher energy levels, respectively) are no longer singled out in Choose My Plate, but rather are included in the empty calories/added sugar category with SOFAS (calories from solid fats and added sugars). The limit is the remaining number of calories in each food pattern above after intake of the recommended amounts of the nutrient-dense foods.

[a]For example, 1 serving equals 1 slice bread, 1 cup ready-to-eat cereal, or 0.5 cup cooked rice, pasta, or cooked cereal. [b]For example, 1 serving equals 1 oz lean meat, poultry, or fish; 1 egg; 1 tablespoon peanut butter; 0.25 cup cooked dry beans; or 0.5 oz nuts or seeds. [c]For example, 1 serving equals 1 cup milk or yogurt, 1.5 oz natural cheese, or 2 oz processed cheese. [d]Formerly called "discretionary calorie allowance." Portions are calculated as the number of calories remaining after all of the above allotments are accounted for.

Abbreviation: oz eq, ounce equivalent.

Source: Data from U.S. Department of Agriculture (*http://www.Choosemyplate.gov*).

patterns need extra instruction on how to categorize their foods and on the appropriate portion sizes that constitute a serving.

For healthy persons who do not require medical nutrition therapy, a rough guide for avoiding inadequate nutrient intakes as well as excessive intakes of fat (especially saturated and trans fats), sodium, sugar, and alcohol is the USDA's Choose My Plate **(Table 343-4)**. The Choose My Plate graphic emphasizes a balance between calories and nutritional needs, encouraging increased intake of fruits and vegetables, whole grains, and low-fat milk in conjunction with reduced intake of sodium and high-calorie sugary drinks. The guide provides a web-based calculator that tailors the number of servings suggested for healthy people of different weights, sexes, ages, and life-cycle stages to meet their needs while avoiding excess (*https://www.myplate.gov/myplate-plan* and *www.ChooseMyPlate.gov*). Reviewing the guide with patients helps them identify food groups eaten in insufficient amounts or in excess and aids the transition to healthier dietary patterns.

NUTRITIONAL STATUS ASSESSMENT

Full nutritional status assessment is reserved for seriously ill patients and those at very high nutritional risk when the cause of malnutrition is still uncertain after initial clinical evaluation and dietary assessment. It involves multiple dimensions, including documentation of dietary intake, health history, anthropometric measurements, biochemical measurements of blood and urine, clinical examination, and functional status evaluation. **For further discussion of nutritional assessment, see Chap. 345.**

GLOBAL CONSIDERATIONS

Several DRIs (e.g., the EAR, the UL, and energy needs) are estimates of physiologic requirements based on experimental evidence that apply globally to healthy individuals when adjusted for age, sex, body size, and physical activity level. However, the AIs, AMDRs, and CDRRs are not exportable and should be used with caution in other countries. AIs for micronutrients are based on epidemiologic observations of customary and adequate intakes in healthy U.S. and Canadian populations. The AMDRs are not based on direct experimental evidence but represent expert opinion on potentially healthy intakes of energy-providing nutrients in North American populations. The CDRR may also vary in other populations. Nutrient-based standards like the DRIs have also been developed by the World Health Organization/Food and Agricultural Organization of the United Nations. They are available at *https://www.who.int/activities/establishing-global-nutrient-requirements*. The European Food Safety Authority (EFSA) Panel on Dietetic Products, Nutrition, and Allergies periodically publishes its recommendations in the *EFSA Journal* (*https://efsa.onlinelibrary.wiley.com/journal/18314732*). Other countries have also promulgated recommendations similar to the DRI. The different standards have many similarities in their basic concepts, definitions, and nutrient recommendation levels, but some differ from the DRIs as a result of the functional criteria chosen for nutrient requirements, environmental differences, the timeliness of the evidence reviewed, and expert judgment. There is a growing trend toward global harmonization of these recommendations.

DISCLAIMER

The opinions expressed in this article are the author's own and do not reflect the views of the National Institutes of Health, the U.S. Department of Health and Human Services, or the U.S. government.

FURTHER READING

Alves LF et al: GLIM criteria to identify malnutrition in patients in hospital settings: A systematic review. J Parenteral Enteral Nutr 47:702, 2023.

Lewis JL, Dwyer JT: Establishing nutrient intake values, in *Present Knowledge in Nutrition*, Vol 2. BP Marriott, DF Birt, VA Stallings, AA Yates, eds. London, Academic Press, 2020, pp 267–289.

National Academy of Sciences, Engineering, and Medicine: *Guiding Principles for Developing Dietary Reference Intakes Based on Chronic Disease*. Washington, DC, National Academies Press, 2017.

National Academy of Sciences, Engineering, and Medicine: *Dietary Reference Intakes for Sodium and Potassium*. Washington, DC, National Academies Press, 2019.

National Academy of Sciences, Engineering, and Medicine: *Harmonizing the Process for Establishing Nutrient Reference Values: A Tool Kit*. Washington, DC, National Academies Press, 2020.

National Academy of Sciences, Engineering, and Medicine: *Nutrition During Pregnancy and Lactation: Exploring New Evidence: Proceedings of a Workshop—in Brief*. Washington, DC, National Academies Press, 2020.

National Academy of Sciences, Engineering, and Medicine: Approaches to assessing intake of food and dietary supplements in pregnant women and children 2 to 11 years of age: Proceedings of a workshop series. Washington, DC, National Academies Press, 2022.

National Academy of Sciences, Engineering, and Medicine: Defining populations for Dietary Reference Intake recommendations: A letter report. Washington, DC, National Academies Press, 2022.

National Academy of Sciences, Engineering, and Medicine: Assessing intake of food and dietary supplements in older adults: Proceedings of a workshop series. Washington, DC, National Academies Press, 2023.

National Academy of Sciences, Engineering, and Medicine: *Dietary Reference Intakes for Energy*. Washington, DC, National Academies Press, 2023.

National Academy of Sciences, Engineering, and Medicine: Using systematic reviews to support future Dietary Reference Intakes: A letter report. Washington, DC, National Academies Press, 2023.

US Department of Agriculture and U.S. Department of Health and Human Services: Dietary Guidelines for Americans, 2020-2025, 9th ed. Washington, DC, US Government Printing Office, 2020.

344 Vitamin and Trace Mineral Deficiency and Excess

Paolo M. Suter

Vitamins are required constituents of the human diet because they are synthesized inadequately or not at all in the human body. Only small amounts of these substances are needed to carry out essential biochemical reactions (e.g., by acting as coenzymes or prosthetic groups). Overt vitamin or trace mineral deficiencies are rare in Western countries because of a plentiful, varied, and inexpensive food supply; food fortification; and use of supplements. However, multiple nutrient deficiencies may appear together in persons who are chronically ill, alcoholic, on specific medications (see below), or living in poverty. After bariatric surgery, patients are at high risk for multiple nutrient deficiencies. Moreover, subclinical vitamin and trace mineral deficiencies (often designated as "hidden hunger"), as diagnosed by laboratory testing, are quite common in the normal population, especially in the geriatric age group and socioeconomically deprived individuals due to the lack of nutrient-dense foods. Conversely, because of the widespread use of nutrient supplements and food fortification, nutrient toxicities are gaining pathophysiologic and clinical importance.

Victims of famine, emergency-affected and displaced populations, refugees, and camp populations are at increased risk for protein-energy malnutrition and classic micronutrient deficiencies (vitamin A, iron, zinc, iodine) as well as for overt deficiencies in thiamine (beriberi), riboflavin, vitamin C (scurvy), and niacin (pellagra).

Body stores of vitamins and minerals vary tremendously. For example, stores of vitamins B_{12} and A are large, and an adult may not become deficient until ≥1 year after beginning to eat a deficient diet. However, folate and thiamine may become depleted within weeks among those eating a deficient diet. Therapeutic modalities can deplete essential nutrients from the body; for example, hemodialysis or diuretics remove water-soluble vitamins, which must be replaced by supplementation.

Vitamins and trace minerals play several roles in diseases: (1) deficiencies of vitamins and minerals may be caused by disease states such as malabsorption; (2) either deficiency or excess of vitamins and minerals can cause disease in and of itself (e.g., vitamin A intoxication and liver disease); and (3) vitamins and minerals in high doses may be used as drugs (e.g., niacin for hypercholesterolemia). Since they are covered elsewhere, the hematologic-related vitamins and minerals **(Chaps. 102 and 104)** either are not considered or are considered only briefly in this chapter, as are the bone-related vitamins and minerals (vitamin D, calcium, phosphorus, magnesium; **Chap. 421**).

VITAMINS

See also Table 344-1 and Fig. 344-1.

THIAMINE (VITAMIN B_1)

Thiamine was the first B vitamin to be identified and therefore is referred to as vitamin B_1. Thiamine functions in the decarboxylation of α-ketoacids (e.g., pyruvate α-ketoglutarate) and branched-chain amino acids and thus is essential for energy generation. In addition, thiamine pyrophosphate acts as a coenzyme for a transketolase reaction that mediates the conversion of hexose and pentose phosphates. It has been postulated that thiamine plays a role in peripheral nerve conduction, although the exact chemical reactions underlying this function are not known.

Food Sources The median intake of thiamine in the United States from food alone is ~2 mg/d. Primary food sources for thiamine include yeast, organ meat, pork, legumes, beef, whole grains, and nuts. Milled rice and grains contain little thiamine. Thiamine deficiency is therefore more common in cultures that rely heavily on a milled polished rice-based diet. Certain foods contain antithiamine factors such as heat-labile thiaminases (raw fish, shellfish), which destroy the vitamin, or heat-stable polyhydroxyphenols (tannins; in coffee, tea, Brussels sprouts, or betel nuts), which inactivate the vitamin. Thus, drinking large amounts of tea or coffee could theoretically lower thiamine body stores.

TABLE 344-1 Principal Clinical Findings of Vitamin Malnutrition

NUTRIENT	CLINICAL FINDING	DIETARY LEVEL PER DAY ASSOCIATED WITH OVERT DEFICIENCY IN ADULTS	CONTRIBUTING FACTORS TO DEFICIENCY
Thiamine	Beriberi: neuropathy, muscle weakness and wasting, cardiomegaly, edema, ophthalmoplegia, confabulation	<0.3 mg/1000 kcal	Alcoholism, chronic diuretic use, bariatric surgery, hyperemesis, thiaminases in food
Riboflavin	Magenta tongue, angular stomatitis, seborrhea, cheilosis, ocular symptoms, corneal vascularization	<0.4 mg	Alcoholism, individuals with poor diets and low intake of milk products
Niacin	Pellagra: pigmented rash of sun-exposed areas, bright red tongue, diarrhea, apathy, memory loss, disorientation	<9.0 niacin equivalents	Alcoholism, vitamin B_6 deficiency, riboflavin deficiency, tryptophan deficiency
Vitamin B_6	Seborrhea, glossitis, convulsions, neuropathy, depression, confusion, microcytic anemia	<0.2 mg	Alcoholism, isoniazid
Folate	Megaloblastic anemia, atrophic glossitis, depression, ↑ homocysteine	<100 μg/d	Alcoholism, sulfasalazine, pyrimethamine, triamterene
Vitamin B_{12}	Megaloblastic anemia, loss of vibratory and position sense, abnormal gait, dementia, impotence, loss of bladder and bowel control, ↑ homocysteine, ↑ methylmalonic acid	<1.0 μg/d	Gastric atrophy (pernicious anemia), terminal ileal disease, strict vegetarianism, acid-reducing drugs (e.g., H_2 blockers), metformin
Vitamin C	Scurvy: petechiae, ecchymosis, coiled hairs, inflamed and bleeding gums, joint effusion, poor wound healing, fatigue	<10 mg/d	Smoking, alcoholism
Vitamin A	Xerophthalmia, night blindness, Bitot's spots, follicular hyperkeratosis, impaired embryonic development, immune dysfunction	<300 μg/d	Fat malabsorption, infection, measles, alcoholism, protein-energy malnutrition
Vitamin D	Rickets: skeletal deformation, rachitic rosary, bowed legs; osteomalacia	<2.0 μg/d	Aging, lack of sunlight exposure, fat malabsorption, deeply pigmented skin
Vitamin E	Peripheral neuropathy, spinocerebellar ataxia, skeletal muscle atrophy, retinopathy	Not described unless underlying contributing factor is present	Occurs only with fat malabsorption or genetic abnormalities of vitamin E metabolism/transport
Vitamin K	Elevated prothrombin time, bleeding	<10 μg/d	Fat malabsorption, liver disease, antibiotic use

Deficiency Most dietary deficiency of thiamine worldwide is the result of poor dietary intake due to the lack of food or disproportionate reliance on highly processed staple crops. Food processing removes thiamine, and high-heat or long-duration cooking destroys it. In Western countries, the primary causes of thiamine deficiency are alcoholism and chronic illnesses such as cancer. Alcohol interferes directly with the absorption of thiamine and with the synthesis of thiamine pyrophosphate, and it increases urinary excretion. Thiamine should always be replenished when a patient with alcoholism is being refed, as carbohydrate repletion without adequate thiamine can precipitate acute thiamine deficiency with lactic acidosis. Other at-risk populations are women with prolonged hyperemesis gravidarum, anorexia (including eating disorders), patients with overall poor nutritional status who are receiving parenteral glucose, patients who have had

Vitamin	Active derivative or cofactor form	Principal function
Thiamine (B_1) 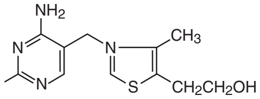	Thiamine pyrophosphate	Coenzyme for cleavage of carbon-carbon bonds; amino acid and carbohydrate metabolism
Riboflavin (B_2) 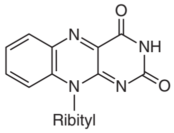	Flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD)	Cofactor for oxidation, reduction reactions, and covalently attached prosthetic groups for some enzymes
Niacin 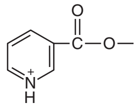	Nicotinamide adenine dinucleotide phosphate (NADP) and nicotinamide adenine dinucleotide (NAD)	Coenzymes for oxidation and reduction reactions
Vitamin B_6 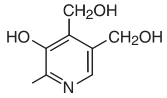	Pyridoxal phosphate	Cofactor for enzymes of amino acid metabolism
Folate 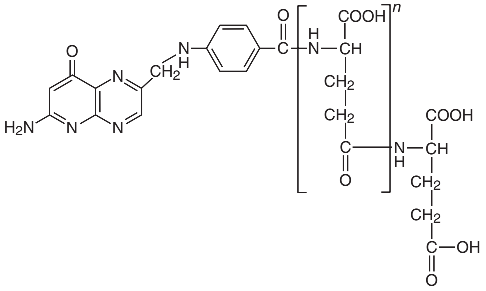	Polyglutamate forms of (5, 6, 7, 8) tetrahydrofolate with carbon unit attachments	Coenzyme for one carbon transfer in nucleic acid and amino acid metabolism
Vitamin B_{12} 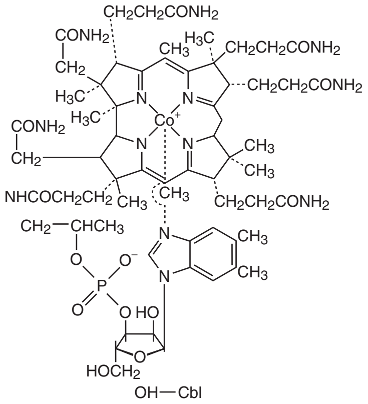	Methylcobalamine Adenosylcobalamin	Coenzyme for methionine synthase and L-methylmalonyl-CoA mutase

FIGURE 344-1 Structures and principal functions of vitamins associated with human disorders.

bariatric/metabolic surgery (*bariatric Wernicke*), and patients receiving chronic diuretic therapy (e.g., in hypertension or systolic heart failure) due to increased urinary thiamine losses. Different drugs (e.g., metformin, verapamil) could inhibit intestinal thiamine transporters (ThTR-2), thereby increasing the risk of deficiency for this vitamin. Maternal thiamine deficiency can lead to infantile beriberi in breast-fed children. Thiamine deficiency could be an underlying factor in motor vehicle accidents and could be overlooked in the setting of head injury.

Thiamine deficiency in its early stage is challenging to identify. It causes anorexia and nonspecific symptoms (e.g., irritability, decrease in short-term memory). Prolonged thiamine deficiency causes *beriberi*, which is classically categorized as wet or dry, although there is considerable overlap between the two categories. In either form of beriberi, patients may complain of pain and paresthesia. *Wet beriberi* presents primarily with cardiovascular symptoms that are due to impaired myocardial energy metabolism and dysautonomia; it can occur after 3 months of a thiamine-deficient diet. Patients present with an enlarged heart, tachycardia, high-output congestive heart failure, peripheral edema, and peripheral neuritis. Patients with *dry beriberi* present with a symmetric peripheral neuropathy of the motor and sensory systems, with diminished reflexes. The neuropathy affects the legs most markedly, and patients have difficulty rising from a squatting position.

Alcoholic patients with chronic thiamine deficiency also may have central nervous system (CNS) manifestations known as *Wernicke's encephalopathy*, which consists of horizontal nystagmus, ophthalmoplegia (due to weakness of one or more extraocular muscles), cerebellar ataxia, and mental impairment **(Chap. 464)**. When there is an additional loss of memory and a confabulatory psychosis, the syndrome is known as *Wernicke-Korsakoff syndrome*. Despite the typical clinical picture and history, Wernicke-Korsakoff syndrome is underdiagnosed.

The laboratory diagnosis of thiamine deficiency usually is made by a functional enzymatic assay of transketolase activity measured before and after the addition of thiamine pyrophosphate. A >25% stimulation in response to the addition of thiamine pyrophosphate (i.e., an activity coefficient of 1.25) is interpreted as abnormal. Thiamine or the phosphorylated esters of thiamine in serum or blood also can be measured by high-performance liquid chromatography to detect deficiency.

TREATMENT

Thiamine Deficiency

In acute thiamine deficiency with either cardiovascular or neurologic signs, 200 mg of thiamine three times daily should be given intravenously until there is no further improvement in acute symptoms; oral thiamine (10 mg/d) should subsequently be given until recovery is complete. Cardiovascular and ophthalmoplegic improvement occurs within 24 h. Other manifestations gradually clear, although psychosis in Wernicke-Korsakoff syndrome may be permanent or may persist for several months. Other nutrient deficiencies should be corrected concomitantly (e.g., magnesium, required for thiamine activation). In view of the widespread, often unrecognized (subclinical) deficiency, a more generous

Vitamin	Active derivative or cofactor form	Principal function
Vitamin C 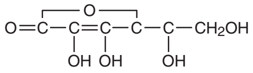	Ascorbic acid and dehydroascorbic acid	Participation as a redox ion in many biologic oxidation and hydrogen transfer reactions
Vitamin A 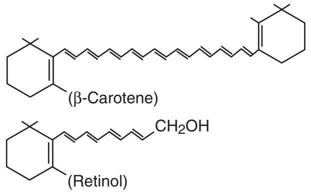 	Retinol, retinaldehyde, and retinoic acid	Formation of rhodopsin (vision) and glycoproteins (epithelial cell function); also regulates gene transcription
Vitamin D 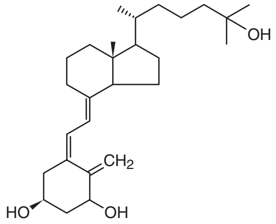	1,25-Dihydroxyvitamin D	Maintenance of blood calcium and phosphorus levels; antiproliferative hormone
Vitamin E 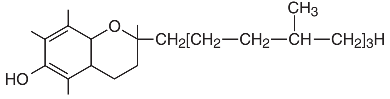	Tocopherols and tocotrienols	Antioxidants
Vitamin K 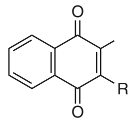	Vitamin K hydroquinone	Cofactor for posttranslation carboxylation of many proteins including essential clotting factors

FIGURE 344-1 (*Continued*)

supplementation of this vitamin in the emergency care setting and in all patients with any alcohol use disorder is warranted.

Toxicity Although hypersensitivity/anaphylaxis has been reported after high intravenous doses of thiamine, no adverse effects have been recorded from either food or supplements at high doses.

RIBOFLAVIN (VITAMIN B$_2$)

Riboflavin is important for the metabolism of fat, carbohydrate, and protein, acting as a respiratory coenzyme and an electron donor in energy production. Enzymes that contain flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN) as prosthetic groups are known as *flavoenzymes* (e.g., succinic acid dehydrogenase, monoamine oxidase, glutathione reductase). FAD is a cofactor for methyltetrahydrofolate reductase and therefore modulates homocysteine metabolism. The vitamin also plays a role in drug and steroid metabolism, including detoxification reactions.

Although much is known about the chemical and enzymatic reactions of riboflavin, the clinical manifestations of riboflavin deficiency are nonspecific and are similar to those of other deficiencies of B vitamins. Riboflavin deficiency is manifested principally by lesions of the mucocutaneous surfaces of the mouth and skin. In addition, corneal vascularization, anemia, and personality changes have been described with riboflavin deficiency.

Deficiency and Excess Riboflavin deficiency is rare and almost always due to dietary deficiency. Milk, other dairy products, and enriched breads and cereals are the most important dietary sources of riboflavin in the United States, although lean meat, fish, eggs, broccoli, and legumes are also good sources. Riboflavin is extremely sensitive to light, and milk should be stored in containers that protect against photodegradation. Laboratory diagnosis of riboflavin deficiency can be made by determination of red blood cell or urinary riboflavin concentrations or by measurement of erythrocyte glutathione reductase activity, with and without added FAD. Because of the limited capacity of the gastrointestinal tract to absorb riboflavin (~27 mg after one oral dose or meal) as well as the instantaneous urinary excretion, riboflavin toxicity has not been described.

NIACIN (VITAMIN B$_3$)

The term *niacin* refers to nicotinic acid and nicotinamide and their biologically active derivatives. Nicotinic acid and nicotinamide serve as precursors of two coenzymes, nicotinamide adenine dinucleotide (NAD) and NAD phosphate (NADP), which are important in numerous oxidation and reduction reactions in the body. In addition, NAD and NADP are active in adenine diphosphate–ribose transfer reactions involved in DNA repair and calcium mobilization.

Metabolism and Requirements Nicotinic acid and nicotinamide are absorbed well from the stomach and small intestine. The bioavailability of niacin from beans, milk, meat, and eggs is high; bioavailability from cereal grains is lower. Since flour is enriched with "free" niacin (i.e., the non-coenzyme form), bioavailability is excellent. Median intakes of niacin in the United States considerably exceed the recommended dietary allowance (RDA).

The amino acid tryptophan can be converted to niacin with an efficiency of 60:1 by weight. Thus, the RDA for niacin is expressed in niacin equivalents. A lower-level conversion of tryptophan to niacin occurs in vitamin B$_6$, riboflavin, and/or iron deficiencies, and in the presence of isoniazid. The urinary excretion products of niacin include 2-pyridone and 2-methyl nicotinamide, measurements of which are used in the diagnosis of niacin deficiency.

Deficiency Niacin deficiency causes *pellagra*, which is found mostly among people eating corn-based diets in parts of China, Africa, and India. Pellagra in North America is found mainly among alcoholics; among patients with congenital defects of intestinal and kidney absorption of tryptophan (Hartnup disease); and among patients with carcinoid syndrome, in which there is increased conversion of tryptophan to serotonin. The antituberculosis drug isoniazid is a structural analogue of niacin and can precipitate pellagra. This also occurs in the setting of mass scale-up of isoniazid tuberculosis preventive therapy in low-income countries. In the setting of famine or population displacement, pellagra results from the absolute lack of niacin but also from the deficiency of micronutrients required for the conversion of tryptophan to niacin (e.g., iron, riboflavin, and pyridoxine). The early symptoms of pellagra include loss of appetite, generalized weakness and irritability, abdominal pain, and vomiting. Bright red glossitis then ensues and is followed by a characteristic skin rash that is pigmented and scaling, particularly in skin areas exposed to sunlight. This rash is known as *Casal's necklace* because it forms a ring around the neck; it is seen in advanced cases. Vaginitis and esophagitis also may occur. Diarrhea (due in part to proctitis and in part to malabsorption), depression, seizures, and dementia are also part of the pellagra syndrome. The primary manifestations of this syndrome are sometimes referred to as "the four Ds": *d*ermatitis, *d*iarrhea, and *d*ementia leading to *d*eath. Aging is characterized by a decline in cellular NAD$^+$, and it seems plausible that maintaining and/or reestablishing cellular NAD$^+$, might have beneficial effects (e.g., metabolic disorders).

TREATMENT

Pellagra

Treatment of pellagra consists of oral supplementation with 100–200 mg of nicotinamide or nicotinic acid three times daily for up to 4 weeks. High doses of nicotinic acid (2 g/d in a time-release form) may be used for the treatment of elevated cholesterol and triglyceride levels and/or low high-density lipoprotein cholesterol levels, but without proven evidence to prevent cardiovascular disease. Nevertheless, nicotinic acid may be useful in patients with statin intolerance or severe hypertriglyceridemia **(Chap. 419)**.

Toxicity Prostaglandin-mediated flushing due to binding of the vitamin to a G protein–coupled receptor has been observed at daily nicotinic acid doses as low as 30 mg taken as a supplement or as therapy for dyslipidemia. There is no evidence of toxicity from niacin that is derived from food sources. Flushing always starts in the face and may be accompanied by skin dryness, itching, paresthesia, and headache. Flushing is subject to tachyphylaxis and often improves with time; premedication with aspirin may alleviate these symptoms. Nausea, vomiting, and abdominal pain also occur at similar doses of niacin. Hepatic toxicity is the most serious toxic reaction caused by sustained-release niacin and may present as jaundice with elevated aspartate aminotransferase (AST) and alanine aminotransferase (ALT) levels. A few cases of fulminant hepatitis requiring liver transplantation have been reported at doses of 3–9 g/d. Other toxic reactions include glucose intolerance, hyperuricemia, macular edema, and macular cysts. The combination of nicotinic acid preparations for dyslipidemia plus 3-hydroxy-3-methylglutaryl–coenzyme A (HMG-CoA) reductase inhibitors may increase the risk of rhabdomyolysis. The upper limit for daily (nontherapeutic) niacin intake has been set at 35 mg.

PYRIDOXINE (VITAMIN B$_6$)

Vitamin B$_6$ refers to a family of compounds that includes pyridoxine, pyridoxal, pyridoxamine, and their 5′-phosphate derivatives. 5′-Pyridoxal phosphate (PLP) is a cofactor for >100 enzymes involved in amino acid metabolism. Vitamin B$_6$ also is involved in heme and neurotransmitter synthesis and in the metabolism of glycogen, lipids, steroids, sphingoid bases, and several vitamins, including the conversion of tryptophan to niacin.

Dietary Sources Plants contain vitamin B$_6$ in the form of pyridoxine, whereas animal tissues contain PLP and pyridoxamine phosphate. The vitamin B$_6$ contained in plants is less bioavailable than that in animal tissues. Rich food sources of vitamin B$_6$ include legumes, nuts, wheat bran, and meat, although it is present in all food groups.

Deficiency Symptoms of vitamin B$_6$ deficiency include epithelial changes, as seen frequently with other B vitamin deficiencies. In addition, severe vitamin B_6 deficiency can lead to peripheral neuropathy, abnormal electroencephalograms, and personality changes that include depression and confusion. In infants, diarrhea, seizures, and anemia have been reported. Microcytic hypochromic anemia is due to diminished hemoglobin synthesis, since the first enzyme involved in heme biosynthesis (aminolevulinate synthase) requires PLP as a cofactor **(Chap. 102)**. In some case reports, platelet dysfunction has been reported. Since vitamin B_6 is necessary for the conversion of homocysteine to cystathionine, it is possible that chronic low-grade vitamin B_6 deficiency may result in hyperhomocysteinemia, which has been associated with vascular dysfunction and an increased risk of cardiovascular disease; however, so far, there is only limited randomized controlled trial evidence **(Chap. 431)**. Independent of homocysteine, low levels of circulating vitamin B_6 have been associated with inflammation and elevated levels of C-reactive protein.

Certain medications, such as isoniazid, L-dopa, penicillamine, and cycloserine, interact with PLP due to a reaction with carbonyl groups. Pyridoxine should be given concurrently with isoniazid to avoid neuropathy. The increased ratio of AST to ALT seen in alcoholic liver disease reflects the relative vitamin B_6 dependence of ALT. Vitamin B_6 dependency syndromes that require pharmacologic doses of vitamin B_6 are rare; they include cystathionine β-synthase deficiency, pyridoxine-responsive (primarily sideroblastic) anemias, and gyrate atrophy with chorioretinal degeneration due to decreased activity of the mitochondrial enzyme ornithine aminotransferase. In these situations, 100–200 mg/d of oral vitamin B_6 is required for treatment.

Severe nausea and vomiting in pregnancy might respond to pyridoxine combined with doxylamine. High doses of vitamin B_6 have been used to treat carpal tunnel syndrome, premenstrual syndrome, schizophrenia, autism, and diabetic neuropathy but have not been found to be effective.

The laboratory diagnosis of vitamin B_6 deficiency is generally based on low plasma PLP values (<20 nmol/L). Vitamin B_6 deficiency is treated with 50 mg/d; higher doses of 100–200 mg/d are given if the deficiency is related to medication use. Vitamin B_6 should not be given with L-dopa, since the vitamin interferes with the action of this drug.

Toxicity The safe upper limit for vitamin B_6 has been set at 100 mg/d, although no adverse effects have been associated with high intakes of vitamin B_6 from food sources only. When toxicity occurs, it causes severe sensory neuropathy, leaving patients unable to walk; however, in most cases, this is reversible upon cessation of the high intake. Medication safety monitoring suggests a rather high prevalence of vitamin B_6–induced neuropathy. Accordingly, long-term high-dose vitamin B_6 supplementation should be discouraged. Some cases of photosensitivity and dermatitis have been reported.

FOLATE (VITAMIN B_{12})

See Chap. 104.

VITAMIN C

Both ascorbic acid (only the L-isomer) and its oxidized product dehydroascorbic acid are biologically active. Actions of vitamin C include antioxidant activity, promotion of nonheme iron absorption, carnitine biosynthesis, conversion of dopamine to norepinephrine, tyrosine catabolism, histone and DNA demethylation, and synthesis of many peptide hormones. Vitamin C is also important for connective tissue metabolism and cross-linking (proline hydroxylation), and it is a component of many drug-metabolizing enzyme systems, particularly the mixed-function oxidase systems.

Absorption and Dietary Sources Vitamin C is almost completely absorbed if <100 mg is administered in a single dose; however, only ≤50% is absorbed at doses >1 g. Enhanced degradation and fecal and urinary excretion of vitamin C occur at higher intake levels.

Good dietary sources of vitamin C include citrus fruits, green vegetables (especially broccoli), tomatoes, and potatoes. Consumption of five servings of fruits and vegetables a day provides vitamin C in excess of the RDA of 90 mg/d for men and 75 mg/d for women. In addition, ~40% of the U.S. population consumes vitamin C as a dietary supplement in which "natural forms" of the vitamin are no more bioavailable than synthetic forms. Smoking (including "passive" smoking), hemodialysis, pregnancy, lactation, and stress (e.g., infection, trauma) appear to increase vitamin C requirements.

Deficiency Vitamin C deficiency causes scurvy. In the United States, this condition is seen primarily among the poor and the elderly, in alcoholics who consume <10 mg/d of vitamin C, and in young adults who eat severely unbalanced diets. In addition to generalized fatigue, symptoms of scurvy primarily reflect impaired formation of mature connective tissue and include bleeding into the skin (petechiae, ecchymoses, perifollicular hemorrhages); inflamed and bleeding gums; and manifestations of bleeding into joints, the peritoneal cavity, the pericardium, and the adrenal glands. In children, vitamin C deficiency may cause impaired bone growth. Laboratory diagnosis of vitamin C deficiency is based on low plasma or leukocyte levels.

Administration of vitamin C (200 mg/d) improves the symptoms of scurvy within several days. High-dose vitamin C supplementation (e.g., 0.2 g up to several grams per day) may slightly decrease the symptoms and duration of upper respiratory tract infections. Vitamin C supplementation has also been reported to be useful in Chédiak-Higashi syndrome **(Chap. 67)** and osteogenesis imperfecta **(Chap. 425)**. Diets high in vitamin C have been claimed to lower the incidence of certain cancers, particularly esophageal and gastric cancers. If proven, this effect may be because vitamin C can prevent the conversion of nitrites and secondary amines to carcinogenic nitrosamines. Emerging evidence suggests a therapeutic effect of intravenous parenteral (*not oral*) pharmacologic doses of up to 1 g/kg body weight of ascorbic acid in the treatment of cancers (e.g., metastatic pancreatic, ovarian, glioblastoma, and non-small-cell lung cancers). The mechanism of pharmacologic ascorbate in cancer treatment (as a stand-alone agent or with other therapeutic agents) appears to be pro-oxidative, either synergistic (e.g., gemcitabine, programmed cell death protein 1 [PD-1] inhibitors, radiation) or additive with other agents.

Toxicity Taking >2 g of vitamin C in a single dose may result in abdominal pain, diarrhea, and nausea. Since vitamin C may be metabolized to oxalate, it is feared that chronic high-dose vitamin C supplementation could result in an increased prevalence of kidney stones. However, except in patients with preexisting renal disease, this association has not been borne out in several trials. Nevertheless, it is reasonable to advise patients with a history of kidney stones (especially oxalate renal stones) and renal insufficiency not to take large doses of vitamin C. There is also an unproven but possible risk that chronic high doses of vitamin C could promote iron overload and iron toxicity (e.g., in patients with hemochromatosis or thalassemia major). High doses of vitamin C can induce hemolysis in patients with glucose-6-phosphate dehydrogenase deficiency, and doses >1 g/d can cause false-negative guaiac reactions and interfere with tests for urinary glucose. High doses may interfere with the activity of certain drugs and diagnostic tests (e.g., false-negative results of guaiac-based fecal occult blood tests). Parenteral high-dose vitamin C in patients with severe infections or sepsis has no beneficial effects and may increase mortality.

BIOTIN

Biotin (also known as vitamin B_7 or vitamin H) is a water-soluble vitamin that plays a role in gene expression, gluconeogenesis, and fatty acid synthesis and serves as a carbon dioxide (CO_2) carrier on the surface of both cytosolic and mitochondrial carboxylase enzymes. The vitamin also functions in the catabolism of specific amino acids (e.g., leucine) and in gene regulation by histone biotinylation. Excellent food sources of biotin include organ meat such as liver or kidney, soy and other beans, yeast, and egg yolks; however, egg white contains the protein avidin, which strongly binds the vitamin and reduces its bioavailability.

Biotin deficiency due to low dietary intake is rare; rather, deficiency is due to inborn errors of metabolism (e.g., biotinidase deficiency). Biotin deficiency has been induced by experimental feeding of egg white diets and by biotin-free parenteral nutrition in patients with short bowels. In adults, biotin deficiency results in mental changes (depression, hallucinations), paresthesia, anorexia, and nausea. A scaling, seborrheic, and erythematous rash may occur around the eyes, nose, and mouth as well as on the extremities. In infants, biotin deficiency presents as hypotonia, lethargy, and apathy. In addition, infants may develop alopecia and a characteristic rash that includes the ears. At present, evidence does not support a therapeutic role of high-dose biotin in multiple sclerosis. The laboratory diagnosis of biotin deficiency can be established on the basis of a decreased concentration of urinary biotin (or its major metabolites), increased urinary excretion of 3-hydroxyisovaleric acid after a leucine challenge, or decreased activity of biotin-dependent enzymes in lymphocytes (e.g., propionyl-CoA carboxylase). Treatment requires pharmacologic doses of biotin, that is, up to 10 mg/d. No toxicity is known. High-dose biotin supplements could interfere with different immunoassay platforms based on streptavidin-biotin technology (e.g., biotinylated antibodies), resulting in false-positive (e.g., free T_4 or T_3) or false-negative tests (e.g., thyroid-stimulating hormone, troponin, β-human chorionic gonadotropin pregnancy test).

PANTOTHENIC ACID (VITAMIN B$_5$)

Pantothenic acid is a component of coenzyme A and phosphopantetheine, which are involved in fatty acid metabolism and the synthesis of cholesterol, steroid hormones, and all compounds formed from isoprenoid units. In addition, pantothenic acid is involved in the acetylation of proteins. The vitamin is excreted in the urine, and the laboratory diagnosis of deficiency is based on low urinary vitamin levels.

The vitamin is ubiquitous in the food supply. Liver, yeast, egg yolks, whole grains, and vegetables are particularly good sources. Human pantothenic acid deficiency has been demonstrated only by experimental feeding of diets low in pantothenic acid or by administration of a specific pantothenic acid antagonist. The symptoms of pantothenic acid deficiency are nonspecific and include gastrointestinal disturbance, depression, muscle cramps, paresthesia, ataxia, and hypoglycemia. Pantothenic acid deficiency is believed to have caused the "burning feet syndrome" seen in prisoners of war during World War II. No toxicity of this vitamin has been reported.

CHOLINE

Choline is a precursor for acetylcholine, phospholipids, and betaine. Choline is necessary for the structural integrity of cell membranes, cholinergic neurotransmission, lipid and cholesterol metabolism, methyl-group metabolism, and transmembrane signaling. Recently, a recommended adequate intake was set at 550 mg/d for men and 425 mg/d for women, although certain genetic polymorphisms can increase an individual's requirement. Choline is thought to be a "conditionally essential" nutrient in that its de novo synthesis occurs in the liver and results in lesser-than-used amounts only under certain stress conditions (e.g., alcoholic liver disease). The dietary requirement for choline depends on the status of other nutrients involved in methyl-group metabolism (folate, vitamin B$_{12}$, vitamin B$_6$, and methionine) and thus varies widely. Choline is widely distributed in food (e.g., egg yolks, wheat germ, organ meat, milk) in the form of lecithin (phosphatidylcholine). Choline deficiency has occurred only in experimental conditions or in patients receiving parenteral nutrition devoid of choline and rarely in specific inborn errors of choline metabolism. Deficiency results in fatty liver, elevated aminotransferase levels, and skeletal muscle damage with high creatine phosphokinase values. The diagnosis of choline deficiency is currently based on low plasma levels, although nonspecific conditions (e.g., heavy exercise) may also suppress plasma levels.

Toxicity from choline results in hypotension, increased sweating, diarrhea, salivation, and a fishy body odor. The upper limit for choline intake has been set at 3.5 g/d. Because of its ability to lower cholesterol and homocysteine levels, choline treatment has been suggested for patients with dementia and patients at high risk of cardiovascular disease. However, the benefits of such treatment have not been firmly documented; recently, signals for an increased cardiovascular risk have been reported. Choline- and betaine-restricted diets are of therapeutic value in trimethylaminuria ("fish odor syndrome") or in decreasing the production of the gut microbiome-derived trimethylamine *N*-oxide (TMAO) as a potential cardiovascular risk modulator.

FLAVONOIDS

Flavonoids constitute a large family of polyphenolic phytochemicals that contribute to the aroma, taste, and color of fruits and vegetables. Major groups of dietary flavonoids include anthocyanidins in berries; flavanols (catechins) in green tea and chocolate; flavonols (e.g., quercetin) in broccoli, kale, leeks, onions, and the skins of grapes and apples; and isoflavones (e.g., genistein) in legumes. Isoflavones have a low bioavailability and are partially metabolized by the intestinal flora. The dietary intake of flavonoids is highly variable and estimated at 10–400 mg/d; this figure is almost certainly an underestimate attributable to a lack of information on their concentrations in many foods. Several flavonoids have antioxidant activity and affect cell signaling. From observational epidemiologic studies and limited clinical (human and animal) studies, flavonoids have been postulated to play a role in the prevention of several chronic diseases, including neurodegenerative disease, diabetes, and osteoporosis. The ultimate importance and usefulness of these compounds against human disease have not been consistently demonstrated. Nevertheless, a dietary pattern with high intake of fruits, vegetables, and legumes should be encouraged to assure a higher intake of these and others nonnutritive bioactives.

VITAMIN A

Vitamin A, in the strictest sense, refers to retinol and retinyl esters. However, the oxidized metabolites retinaldehyde and retinoic acid are also biologically active compounds. The term *retinoids* includes all molecules (including synthetic molecules) that are chemically related to retinol. Retinaldehyde (11-*cis*) is the form of vitamin A that is required for normal vision, whereas retinoic acid is necessary for normal morphogenesis, growth, and cell differentiation. Retinoic acid does not function directly in vision and, in contrast to retinol, is not involved in reproduction. Vitamin A also plays a role in iron utilization, humoral immunity, T cell–mediated immunity, natural killer cell activity, and phagocytosis.

Vitamin A is found in the human food supply in two forms: preformed as retinyl esters and provitamin A carotenoids. There are >700 carotenoids in nature, ~50 of which can be metabolized to vitamin A. β-Carotene is the most prevalent carotenoid with provitamin A activity in the food supply. In humans, significant fractions of carotenoids are absorbed intact and are stored in liver and fat. It is estimated that in healthy humans ≥12 μg (range, 4–27 μg) of dietary all-*trans* β-carotene is equivalent to 1 μg of retinol activity, whereas the figure is ≥24 μg for other dietary provitamin A carotenoids (e.g., β-cryptoxanthin, α-carotene). The vitamin A equivalency for a β-carotene supplement in an oily solution is 2:1.

Metabolism The liver contains ~90% of the vitamin A reserves in healthy individuals and secretes vitamin A in the form of retinol, which is bound in the circulation to retinol-binding protein. Once binding has occurred, the retinol-binding protein complex interacts with a second protein, transthyretin. This trimolecular complex functions to prevent vitamin A from being filtered by the kidney glomerulus, thus protecting the body against the toxicity of retinol and allowing retinol to be taken up by specific cell-surface receptors that recognize retinol-binding protein. A certain amount of vitamin A enters peripheral cells even if it is not bound to retinol-binding protein. After retinol is internalized by the cell, it becomes bound to a series of cellular retinol-binding proteins, which function as sequestering and transporting agents as well as co-ligands for enzymatic reactions. Certain cells also contain retinoic acid–binding proteins, which have sequestering functions but also shuttle retinoic acid to the nucleus and enable its metabolism.

Vitamin A metabolites (retinoids) such as retinoic acid are potent regulators of gene transcription through nuclear receptor signaling, thus playing a key role in many cellular and metabolic pathways. Two families of receptors (retinoic acid receptors [RARs] and retinoid X receptors [RXRs]) are active in retinoid-mediated gene transcription. Retinoid receptors regulate transcription by binding as dimeric complexes to specific DNA sites—the retinoic acid response elements—in target genes **(Chap. 389)**. The receptors can either stimulate or repress gene expression in response to their ligands. RARs bind all-*trans* retinoic acid and 9-*cis*-retinoic acid, whereas RXRs bind only 9-*cis*-retinoic acid.

The retinoid receptors play an important role in controlling cell proliferation and differentiation. RXRs dimerize with other nuclear receptors to function as coregulators of genes responsive to retinoids, but also to thyroid hormone and calcitriol. RXR agonists induce insulin sensitivity experimentally, perhaps because RXRs are cofactors for the peroxisome proliferator-activated receptors, which also mediate fatty acid and carbohydrate metabolism and are targets for different drugs including thiazolidinedione drugs (e.g., rosiglitazone and pioglitazone) **(Chap. 416)**.

Dietary Sources The retinol activity equivalent (RAE) is used to express the vitamin A value of food: 1 RAE is defined as 1 μg of retinol (0.003491 mmol), 12 μg of β-carotene, and 24 μg of other provitamin A carotenoids. In older literature, vitamin A often was expressed in international units (IUs), with 1 μg of retinol equal to 3.33 IU of retinol and 20 IU of β-carotene. Although these IUs are no longer in scientific use, they can still be found in reports of the food industry and in public health interventions in low-income countries.

Liver, fish, and eggs are excellent food sources for preformed vitamin A; vegetable sources of provitamin A carotenoids include dark green and deeply colored fruits and vegetables. Moderate cooking of vegetables enhances carotenoid release for uptake in the gut. Carotenoid absorption is also aided by some fat in a meal. Exclusive breast-feeding can cover the vitamin A needs of infants if the mother has an adequate vitamin A status and a large enough volume of milk. If the nursing mother has inadequate vitamin A intake or concomitant diseases or her infant was a preterm delivery, breast milk probably will not supply enough vitamin A to prevent deficiency. In developing countries, chronic dietary deficiency is the main cause of vitamin A deficiency and is exacerbated by infection. In early childhood, low vitamin A status results from inadequate intakes of animal food sources and edible oils, both of which are expensive, coupled with seasonal unavailability of vegetables and fruits and lack of marketed fortified food products. Factors that interfere with vitamin A metabolism may also affect status or function. For example, concurrent zinc deficiency can interfere with the mobilization of vitamin A from liver stores. Alcohol interferes with the conversion of retinol to retinaldehyde in the eye by competing for alcohol (retinol) dehydrogenase. Drugs that interfere with the absorption of vitamin A include mineral oil, neomycin, and bile acid sequestrants (e.g., cholestyramine).

Deficiency Vitamin A deficiency is endemic in areas where diets are chronically poor, especially in southern Asia, sub-Saharan Africa, some parts of Latin America, and the western Pacific, including parts of China. Vitamin A status is usually assessed by measuring serum retinol (normal range, 1.05–3.50 μmol/L [30–100 μg/dL]) or via dose-response tests or tests of dark adaptation. To assure a correct biochemical assessment of vitamin A status, a simultaneous assessment of the inflammatory status is needed (in analogy to the assessment of iron status); not doing so may result in an overestimation of vitamin A deficiency. Correction factors to adjust the measured plasma vitamin A levels to account for the influence of C-reactive protein and α_1-acid glycoprotein are available. Stable isotopic or invasive liver biopsy methods are available to estimate total-body stores of vitamin A. As judged by deficient serum retinol (<0.70 μmol/L [20 μg/dL]), vitamin A deficiency worldwide is present in 190 million preschool-age children, among whom >5 million have an ocular manifestation of deficiency termed *xerophthalmia*. This condition includes milder stages of night blindness and conjunctival *xerosis* (dryness) with *Bitot's spots* (white patches of keratinized epithelium appearing on the sclera) that may affect 1–5% of children in deficient populations as well as rare, potentially blinding corneal ulceration and necrosis. *Keratomalacia* (softening of the cornea) leads to corneal scarring that blinds an estimated quarter of a million children each year and is associated with fatality rates of 4–25%. However, vitamin A deficiency severe enough to cause any clinical stage poses an increased risk of death from diarrhea, dysentery, measles, malaria, or respiratory disease. This is because vitamin A deficiency can compromise barrier, innate, and acquired immune defenses to infection. In areas where deficiency is widely prevalent, vitamin A supplementation can markedly reduce the risk of childhood mortality (by 23–34%, on average). About 10% of pregnant women in undernourished settings also develop night blindness (assessed by history) during the latter half of pregnancy; this level of moderate to severe vitamin A deficiency is associated with an increased risk of maternal infection and death. Maternal vitamin A deficiency may also exacerbate already low vitamin A nutrition and associated risks for the newborn. In South Asia, where maternal deficiency is prominent, giving infants a single oral dose (50,000 IU) of vitamin A shortly after birth has reduced infant mortality by ≥10%, whereas in African settings less affected by maternal vitamin A deficiency, no effect has been noted, revealing differences in risk of deficiency and benefit of supplementation across regions. However, the World Health Organization does not recommend high-dose supplementation to newborns.

TREATMENT

Vitamin A Deficiency

Vitamin A is commercially available for treatment and prevention in esterified forms (e.g., acetate, palmitate), which are more stable than other forms. Any stage of xerophthalmia should be treated with 60 mg (or RAE) or 200,000 IU of vitamin A in oily solution, usually contained in a soft-gel capsule. The same dose is repeated 1 and 14 days later. Doses should be reduced by half for patients 6–11 months of age. Mothers with night blindness or Bitot's spots should be given vitamin A orally 3 mg daily for at least 3 months. These regimens are efficacious, and they are far less expensive and more widely available than injectable water-miscible vitamin A. A common approach to prevention is to provide vitamin A supplementation every 4–6 months to young children 6 months to 5 years of age (both HIV-positive and HIV-negative) in high-risk areas. For prevention, infants 6–11 months of age should receive 30 mg of vitamin A; children 12–59 months of age should receive 60 mg. For reasons that are not clear, although early neonatal vitamin A may reduce infant mortality, vitamin A given between 1 and 5 months of age has not proven effective in improving survival in high-risk settings.

Uncomplicated vitamin A deficiency is rare in industrialized countries. One high-risk group—extremely low-birth-weight (<1000-g) infants—is likely to be vitamin A deficient and should receive a supplement of 1500 μg (or RAE) three times a week for 4 weeks. Severe measles in any society can lead to secondary vitamin A deficiency. Children hospitalized with measles should receive two 60-mg doses of vitamin A on 2 consecutive days. Vitamin A deficiency most often occurs in patients with malabsorptive diseases (e.g., celiac sprue, short-bowel syndrome) who have abnormal dark adaptation or symptoms of night blindness without other ocular changes. Typically, such patients are diagnosed in advanced care settings where they are treated for 1 month with 15 mg/d of a water-miscible preparation of vitamin A. This treatment is followed by a lower maintenance dose, with the exact amount determined by monitoring serum retinol. Finding application elsewhere in medicine, retinoic acid is useful in the treatment of promyelocytic leukemia **(Chap. 109)** and also is used in the treatment of cystic acne because it inhibits keratinization, decreases sebum secretion, and possibly alters the inflammatory reaction **(Chap. 60)**.

No specific signs or symptoms result from carotenoid deficiency. It was postulated that β-carotene would be an effective chemopreventive agent for cancer because numerous epidemiologic studies had shown that diets high in β-carotene were associated with lower incidences of cancers of the respiratory and digestive systems. However, intervention studies in smokers found that treatment with high doses of β-carotene actually resulted in more lung cancers than did treatment with placebo. Non–provitamin A carotenoids such as lutein and zeaxanthin have been suggested to confer protection against macular degeneration, and one large-scale intervention study did not show a beneficial effect except in those with a low lutein status. The use of the non–provitamin A carotenoid lycopene to protect against prostate cancer has been proposed. However, the effectiveness of these agents has not been proved by intervention studies, and the mechanisms underlying these purported biologic actions are unknown.

Selective plant-breeding techniques that lead to a higher provitamin A carotenoid content in staple foods may decrease vitamin A malnutrition in low-income countries. Moreover, a recently developed genetically modified food (Golden Rice) had a β-carotene–to–vitamin A conversion ratio of ~3:1 in children.

Toxicity The acute toxicity of vitamin A was first noted in Arctic explorers who ate polar bear liver and has also been seen after administration of 150 mg to adults or 100 mg to children. Acute toxicity is manifested by increased intracranial pressure, vertigo, diplopia, bulging fontanels (in children), seizures, and exfoliative dermatitis; it may result in death. Among children being treated for vitamin A deficiency according to the protocols outlined above, transient bulging of fontanels occurs in 2% of infants, and transient nausea, vomiting, and headache occur in 5% of preschoolers. Chronic vitamin A intoxication is largely a concern in industrialized countries and has been seen in otherwise healthy adults who ingest 15 mg/d and children who ingest 6 mg/d over a period of several months. Manifestations include dry skin, cheilosis, glossitis, vomiting, alopecia, bone demineralization and pain, hypercalcemia, lymph node enlargement, hyperlipidemia, amenorrhea, and features of pseudotumor cerebri with increased intracranial pressure and papilledema. Liver fibrosis with portal hypertension may also result from chronic vitamin A intoxication. Provision of vitamin A in excess to pregnant women has resulted in spontaneous abortion and in congenital malformations, including craniofacial abnormalities and valvular heart disease. In pregnancy, the daily dose of vitamin A should not exceed 3 mg. Also, topical retinoids should be avoided during pregnancy. Commercially available retinoid derivatives are also toxic, including 13-*cis*-retinoic acid, which has been associated with birth defects. Thus, contraception should be continued for at least 1 year and possibly longer in women who have taken 13-*cis*-retinoic acid.

In malnourished children, vitamin A supplements (30–60 mg), in amounts calculated as a function of age and given in several rounds over 2 years, are considered to amplify nonspecific effects of vaccines. However, for unclear reasons, in one African setting, there has been a negative effect on mortality rates in incompletely vaccinated girls.

High doses of supplemental carotenoids do not result in toxic symptoms but should be avoided in smokers due to an increased risk of lung cancer. Very high doses of β-carotene (~200 mg/d) have been used to treat or prevent the skin rashes of erythropoietic protoporphyria. Carotenemia, which is characterized by a yellowing of the skin (in creases of the palms and soles) but not the sclerae, may follow ingestion of >30 mg of β-carotene daily. Hypothyroid patients are particularly susceptible to the development of carotenemia due to impaired breakdown of carotene to vitamin A. Reduction of carotenes in the diet results in the disappearance of skin yellowing and carotenemia over a period of 30–60 days.

VITAMIN D

The metabolism of the fat-soluble vitamin D is described in detail in **Chap. 421.** The biologic effects of this vitamin are mediated by vitamin D receptors, which are found in most tissues; binding with these receptors potentially expands vitamin D actions to many different cell systems and organs (e.g., immune cells, brain, breast, colon, and prostate) in addition to the classic endocrine effects on calcium and phosphate metabolism and bone health. Vitamin D is thought to be important for maintaining normal function of many nonskeletal tissues such as muscle (including heart muscle), for immune function, and for inflammation as well as for cell proliferation and differentiation. Older studies have shown that vitamin D may be useful as adjunctive treatment for tuberculosis, psoriasis, and multiple sclerosis or for the prevention of certain cancers. Vitamin D insufficiency may increase the risk of type 1 diabetes mellitus, cardiovascular disease (insulin resistance, hypertension, or low-grade inflammation), or brain dysfunction (e.g., depression). However, the exact physiologic roles of vitamin D in these nonskeletal diseases and the importance of these roles have so far not been clarified. Recent placebo-controlled studies did not show a therapeutic benefit of vitamin D for cancer prevention, control of cardiovascular disease, or risk of type 2 diabetes, depression, tuberculosis infection, or other respiratory infections. Presently, it is not known whether these effects of vitamin D supplements (with or without calcium) might be different according to the baseline status (normal vs severely deficient) of patients.

The skin is a major source of vitamin D, which is synthesized upon skin exposure to ultraviolet B radiation (UV-B; wavelength, 290–320 nm). Except for fish, food (unless fortified) contains only limited amounts of vitamin D. Vitamin D_2 (ergocalciferol) is obtained from plant sources and is the chemical form found in some supplements.

Deficiency Vitamin D status is assessed by measuring serum levels of 25-dihydroxyvitamin D (25[OH] vitamin D); however, there is no consensus on a uniform assay, on optimal serum levels, or on the real benefit of biochemical screening in asymptomatic adults. The optimal level might, in fact, differ according to the targeted disease entity. Epidemiologic and experimental data indicate that a 25(OH) vitamin D level of >20 ng/mL (≥50 nmol/L; to convert ng/mL to nmol/L, multiply by 2.496) is sufficient for good bone health. The latter 25(OH) vitamin D plasma concentration would cover the requirements of 97.5% of the population. Some experts, however, advocate higher serum levels (e.g., >30 ng/mL) for other desirable endpoints of vitamin D action. There is insufficient evidence to recommend combined vitamin D and calcium supplementation as a primary preventive strategy (as opposed to secondary prevention) for reduction of the incidence of fractures in healthy men and premenopausal women.

Risk factors for vitamin D deficiency are old age, lack of sun exposure, dark skin (especially among residents of northern latitudes), fat malabsorption, and obesity; deficiency can also occur after gastric bypass surgery. In addition, in African populations, the prevalence of vitamin D deficiency might be high (especially in women, newborn babies, urban populations, and those living in northern African countries). *Rickets* represents the classic disease of vitamin D deficiency. Signs of deficiency are muscle soreness, weakness, and bone pain. Some of these effects are independent of calcium intake. To *prevent* glucocorticoid-induced osteoporosis, treatment with calcium (1000–1200 mg/d) and vitamin D (600–800 IU/d) through diet and/or supplements in combination with weight-bearing exercise is recommended.

The U.S. National Academy of Sciences recently advised that the majority of adult North Americans should receive 600 IU/d of vitamin D (RDA = 15 μg/d or 600 IU/d; **Chap. 343**). However, for people aged >70 years, the RDA is set at 20 μg/d (800 IU/d). The consumption of fortified or enriched foods as well as suberythemal sun exposure should be encouraged for people at risk for vitamin D deficiency. If adequate intake is impossible, vitamin D supplements should be taken, especially during the winter months. Vitamin D deficiency can be treated by oral administration of 50,000 IU/week for 6–8 weeks followed by a maintenance dose of 800 IU/d (20 μg/d) from food and supplements once normal plasma levels have been attained. There is still uncertainty regarding the optimal therapeutic dosage (high vs low) for elderly at risk of falls. The physiologic effects of vitamin D_2 and vitamin D_3 are similar when these vitamins are ingested over long periods.

Toxicity The upper limit of intake has been set at 4000 IU/d. Contrary to earlier beliefs, acute vitamin D intoxication is rare and usually is caused by the uncontrolled and excessive ingestion of supplements or by faulty food fortification practices. High plasma levels of $1,25(OH)_2$ vitamin D and calcium are central features of toxicity and mandate discontinuation of vitamin D and calcium supplements; in addition, treatment of hypercalcemia may be required.

VITAMIN E

Vitamin E is the collective designation for all stereoisomers of tocopherols and tocotrienols, although only the α-tocopherols meet human requirements. Vitamin E acts as a chain-breaking antioxidant and is an efficient peroxyl radical scavenger that protects low-density lipoproteins and polyunsaturated fats in membranes from oxidation. A network of other antioxidants (e.g., vitamin C, glutathione) and enzymes maintains vitamin E in a reduced state. Vitamin E also inhibits prostaglandin synthesis and the activities of protein kinase C and phospholipase A_2.

Absorption and Metabolism After absorption, vitamin E is taken up from chylomicrons by the liver, and a hepatic α-tocopherol transport protein mediates intracellular vitamin E transport and incorporation into very-low-density lipoprotein. The transport protein has a particular affinity for the RRR isomeric form of α-tocopherol; thus, this natural isomer has the most biologic activity.

Requirement Vitamin E is widely distributed in the food supply, with particularly high levels in sunflower oil, safflower oil, and wheat germ oil; γ-tocotrienols are notably present in soybean and corn oils. Vitamin E is also found in meats, nuts, and cereal grains, and small amounts are present in fruits and vegetables. Vitamin E pills containing doses of 50–1000 mg are ingested by ~10% of the U.S. population. The RDA for vitamin E is 15 mg/d (34.9 μmol or 22.5 IU) for all adults. Diets high in polyunsaturated fats may necessitate a slightly higher intake of vitamin E.

Dietary deficiency of vitamin E does not exist in developed countries but can occur in developing countries due to inadequate intake. Vitamin E deficiency is seen only in severe and prolonged malabsorptive diseases, such as celiac disease, chronic cholestatic liver disease, or after small-intestinal resection or bariatric surgery. Children with cystic fibrosis or prolonged cholestasis may develop vitamin E deficiency characterized by areflexia and hemolytic anemia. Children with abetalipoproteinemia cannot absorb or transport vitamin E and become deficient quite rapidly. A familial form of isolated vitamin E deficiency also exists; it is due to a defect in the α-tocopherol transport protein. Vitamin E deficiency causes axonal degeneration of the large myelinated axons and results in posterior column and spinocerebellar symptoms. Peripheral neuropathy is initially characterized by areflexia, with progression to an ataxic gait, and by decreased vibration and position sensations. Ophthalmoplegia, skeletal myopathy, and pigmented retinopathy may also be features of vitamin E deficiency. A deficiency of either vitamin E or selenium in the host has been shown to increase certain viral mutations and, therefore, virulence. The laboratory diagnosis of vitamin E deficiency is based on low blood levels of α-tocopherol (<5 μg/mL, or <0.8 mg of α-tocopherol per gram of total lipids).

TREATMENT

Vitamin E Deficiency

Symptomatic vitamin E deficiency should be treated with 800–1200 mg of α-tocopherol per day. Patients with abetalipoproteinemia may need as much as 5000–7000 mg/d. Children with symptomatic vitamin E deficiency should be treated orally with water-miscible esters (400 mg/d); alternatively, 2 mg/kg per d may be administered intramuscularly. Vitamin E in high doses may protect against oxygen-induced retrolental fibroplasia and bronchopulmonary dysplasia as well as intraventricular hemorrhage of prematurity. Vitamin E has been suggested to increase sexual performance, treat intermittent claudication, and slow the aging process, but convincing evidence for these properties is lacking. When given in combination with other antioxidants, vitamin E may help prevent macular degeneration. Vitamin E may have favorable therapeutic effects in noncirrhotic nondiabetic patients with nonalcoholic steatohepatitis. High doses (60–800 mg/d) of vitamin E have been shown in controlled trials to improve parameters of immune function and reduce colds in nursing home residents, but intervention studies using vitamin E to prevent cardiovascular disease or cancer have not shown efficacy, and at doses >400 mg/d, vitamin E may even increase all-cause mortality rates and prostate cancer risk (especially in combination with selenium supplements).

Toxicity All forms of vitamin E are absorbed and could contribute to toxicity; however, the toxicity risk seems to be rather low as long as liver function is normal. High doses of vitamin E (>800 mg/d) may reduce platelet aggregation and interfere with vitamin K metabolism and are therefore contraindicated in patients taking warfarin and antiplatelet agents (such as aspirin or clopidogrel). Nausea, flatulence, and diarrhea have been reported at doses >1 g/d.

■ VITAMIN K

There are two natural forms of vitamin K: vitamin K_1, also known as *phylloquinone*, from vegetable sources, and vitamin K_2, or *menaquinones*, which are synthesized by bacterial flora and found in hepatic tissue. Phylloquinone can be converted to menaquinone in some organs.

Vitamin K is required for the posttranslational carboxylation of glutamic acid, which is necessary for calcium binding to γ-carboxylated proteins such as prothrombin (factor II); factors VII, IX, and X; protein C; protein S; and proteins found in bone (osteocalcin) and vascular smooth muscle (e.g., matrix Gla protein). However, the importance of vitamin K for bone mineralization and prevention of vascular calcification in different patient groups (including chronic kidney disease) is unclear. Warfarin-type drugs inhibit γ-carboxylation by preventing the conversion of vitamin K to its active hydroquinone form.

Dietary Sources Vitamin K is found in green leafy vegetables such as kale and spinach, and appreciable amounts are also present in margarine and liver. Vitamin K is present in vegetable oils; olive, canola, and soybean oils are particularly rich sources. The average daily intake by Americans is estimated to be ~100 μg/d.

Deficiency The symptoms of vitamin K deficiency are due to hemorrhage; newborns are particularly susceptible because of low fat stores, low breast milk levels of vitamin K, relative sterility of the infantile intestinal tract, liver immaturity, and poor placental transport. Intracranial bleeding as well as gastrointestinal and skin bleeding can occur in vitamin K–deficient infants 1–7 days after birth. Thus, vitamin K (0.5–1 mg IM) is given prophylactically at delivery.

Vitamin K deficiency in adults may be seen in patients with chronic small-intestinal disease (e.g., celiac disease, Crohn's disease), in those with obstructed biliary tracts, or after small-bowel resection. Broad-spectrum antibiotic treatment can precipitate vitamin K deficiency by reducing numbers of gut bacteria, which synthesize menaquinones, and by inhibiting the metabolism of vitamin K. In patients with warfarin therapy, the antiobesity drug orlistat can lead to changes in international normalized ratio due to vitamin K malabsorption. The assessment of the vitamin K status can be done by measurement of phylloquinone (vitamin K_1) concentration in serum (deficiency <0.15 μg/L); the cellular utilization of vitamin K can be assessed by the serum or plasma concentration of undercarboxylated prothrombin (protein induced by vitamin K absence/antagonism [PIVKA-II]). An elevated prothrombin time or activated partial thromboplastin time or reduced clotting factors are useful markers in severe deficiency but are otherwise nonspecific and lack sensitivity. Vitamin K deficiency is treated with a parenteral dose of 10 mg. For patients with chronic malabsorption, 1–2 mg/d should be given orally or 1–2 mg per week can be taken parenterally. Patients with liver disease may have an elevated prothrombin time because of liver cell destruction as well as vitamin K deficiency. If an elevated prothrombin time does not improve during vitamin K therapy, it can be deduced that this abnormality is not the result of vitamin K deficiency.

Toxicity Toxicity from dietary phylloquinones and menaquinones has not been described. High doses of vitamin K can impair the actions of oral vitamin K antagonist anticoagulants.

MINERALS

See also Table 344-2.

■ CALCIUM

See Chap. 421.

TABLE 344-2 Deficiencies and Toxicities of Metals

ELEMENT	DEFICIENCY	TOXICITY	TOLERABLE UPPER (DIETARY) INTAKE LEVEL
Boron	No biologic function determined	Developmental defects, male sterility, testicular atrophy	20 mg/d (extrapolated from animal data)
Calcium	Reduced bone mass, osteoporosis	Renal insufficiency (milk-alkali syndrome), nephrolithiasis, impaired iron absorption, thiazide diuretics	2500 mg/d (milk-alkali)
Copper	Anemia, growth retardation, defective keratinization and pigmentation of hair, hypothermia, degenerative changes in aortic elastin, osteopenia, mental deterioration	Nausea, vomiting, diarrhea, hepatic failure, tremor, mental deterioration, hemolytic anemia, renal dysfunction	10 mg/d (liver toxicity)
Chromium	Impaired glucose tolerance	*Occupational:* Renal failure, dermatitis, pulmonary cancer	Not determined
Fluoride	↑ Dental caries	Dental and skeletal fluorosis, osteosclerosis	10 mg/d (fluorosis)
Iodine	Thyroid enlargement, ↓ T_4, cretinism	Thyroid dysfunction, acne-like eruptions	1100 μg/d (thyroid dysfunction)
Iron	Muscle abnormalities, koilonychia, pica, anemia, ↓ work performance, impaired cognitive development, premature labor, ↑ perinatal maternal death	Gastrointestinal effects (nausea, vomiting, diarrhea, constipation), iron overload with organ damage, acute and chronic systemic toxicity, increased susceptibility to malaria, increased risk association with certain chronic diseases (e.g., diabetes)	45 mg/d of elemental iron (gastrointestinal side effects)
Manganese	Impaired growth and skeletal development, reproduction, lipid and carbohydrate metabolism; upper body rash	*General:* Neurotoxicity, Parkinson-like symptoms *Occupational:* Encephalitis-like syndrome, Parkinson-like syndrome, psychosis, pneumoconiosis	11 mg/d (neurotoxicity)
Molybdenum	Severe neurologic abnormalities	Reproductive and fetal abnormalities	2 mg/d (extrapolated from animal data)
Selenium	Cardiomyopathy, heart failure, striated muscle degeneration	*General:* Alopecia, nausea, vomiting, abnormal nails, emotional lability, peripheral neuropathy, lassitude, garlic odor to breath, dermatitis *Occupational:* Lung and nasal carcinomas, liver necrosis, pulmonary inflammation	400 μg/d (hair, nail changes)
Phosphorus	Rickets (osteomalacia), proximal muscle weakness, rhabdomyolysis, paresthesia, ataxia, seizure, confusion, heart failure, hemolysis, acidosis	Hyperphosphatemia	4000 mg/d
Zinc	Growth retardation, ↓ taste and smell, alopecia, dermatitis, diarrhea, immune dysfunction, failure to thrive, gonadal atrophy, congenital malformations	*General:* Reduced copper absorption, gastritis, sweating, fever, nausea, vomiting *Occupational:* Respiratory distress, pulmonary fibrosis	40 mg/d (impaired copper metabolism)

■ ZINC

Zinc is an integral component of many metalloenzymes in the body; it is involved in the synthesis and stabilization of proteins, DNA, and RNA and plays a structural role in ribosomes and membranes. Zinc is necessary for the binding of steroid hormone receptors and several other transcription factors to DNA. Zinc is essential for normal spermatogenesis, fetal growth, and embryonic development.

Absorption The absorption of zinc from the diet is inhibited by dietary phytate, fiber, oxalate, iron, and copper as well as by certain drugs, including penicillamine, sodium valproate, and ethambutol. Protein-containing foods, i.e., meat, shellfish, nuts, and legumes, are good sources of bioavailable zinc, whereas zinc in grains and legumes is less available for absorption. Grains and legumes contain phytate that binds zinc in the intestine and reduces its availability for absorption.

Deficiency Mild zinc deficiency has been described in many diseases, including diabetes mellitus, HIV/AIDS, cirrhosis, alcoholism, inflammatory bowel disease, malabsorption syndromes, and sickle cell disease. In these diseases, mild chronic zinc deficiency can cause stunted growth in children, decreased taste sensation (*hypogeusia*), and impaired immune function. Severe chronic zinc deficiency has been described as a cause of hypogonadism and dwarfism in several Middle Eastern countries. In these children, hypopigmented hair is also part of the syndrome. Acrodermatitis enteropathica is a rare autosomal recessive disorder characterized by abnormalities in zinc absorption. Clinical manifestations include diarrhea, alopecia, muscle wasting, depression, irritability, and a rash involving the extremities, face, and perineum. The rash is characterized by vesicular and pustular crusting with scaling and erythema. Occasional patients with Wilson's disease have developed zinc deficiency as a consequence of penicillamine therapy (Chap. 427).

Zinc deficiency is prevalent in many developing countries and usually coexists with other micronutrient deficiencies (especially iron deficiency). Zinc (20 mg/d until recovery) may be an effective adjunctive therapeutic strategy for diarrheal disease and pneumonia in children ≥6 months of age.

The diagnosis of zinc deficiency is usually based on a serum zinc level <12 μmol/L (<70 μg/dL). Pregnancy and birth control pills may cause a slight depression in serum zinc levels, and hypoalbuminemia from any cause can result in hypozincemia. In acute stress situations (illness, but also postexercise recovery), zinc may be redistributed from serum into tissues. Zinc deficiency may be treated with 60 mg of elemental zinc taken by mouth twice a day. Zinc gluconate lozenges (13 mg of elemental zinc every 2 h while awake) have been reported to reduce the duration and symptoms of the common cold in adults, but study results are conflicting.

Toxicity Acute zinc toxicity after oral ingestion causes nausea, vomiting, and fever. Zinc fumes from welding may also be toxic and cause fever, respiratory distress, excessive salivation, sweating, and headache. Chronic large doses of zinc (ranging from 150 to 450 mg/d) may depress immune function and cause hypochromic anemia as a result of a secondary copper deficiency. Intranasal zinc preparations should be avoided because they may lead to irreversible damage of the nasal mucosa and anosmia.

■ COPPER

Copper is an integral part of numerous enzyme systems, including amine oxidases, ferroxidase (ceruloplasmin), cytochrome c oxidase, superoxide dismutase, and dopamine hydroxylase. Copper is also a component of ferroprotein, a transport protein involved in the basolateral transfer of iron during absorption from the enterocyte. As such, copper plays a role in iron metabolism, melanin synthesis, energy production, neurotransmitter synthesis, and CNS function; the synthesis and cross-linking of elastin and collagen; and the scavenging of superoxide radicals. Dietary sources of copper include shellfish, liver, nuts, legumes, bran, and organ meats.

Deficiency Dietary copper deficiency is relatively rare, although it has been described in premature infants who are fed milk diets and in infants with malabsorption (Table 344-2). Copper-deficiency anemia (refractory to therapeutic iron) has been reported in patients with malabsorptive diseases and nephrotic syndrome and in patients treated for Wilson's disease with chronic high doses of oral zinc, which can interfere with copper absorption. *Menkes kinky hair syndrome* is an X-linked metabolic disturbance of copper metabolism characterized by intellectual disability, hypocupremia, and decreased circulating ceruloplasmin **(Chap. 425)**. This syndrome is caused by mutations in the copper-transporting *ATP7A* gene. Children with this disease often die within 5 years because of dissecting aneurysms or cardiac rupture. Aceruloplasminemia is a rare autosomal recessive disease characterized by tissue iron overload, mental deterioration, microcytic anemia, and low serum iron and copper concentrations.

The diagnosis of copper deficiency is usually based on low serum levels of copper (<65 μg/dL) and low ceruloplasmin levels (<20 mg/dL). Serum levels of copper may be elevated in pregnancy or stress conditions since ceruloplasmin is an acute-phase reactant and 90% of circulating copper is bound to ceruloplasmin. It has been suggested that mild or subclinical copper deficiency is more common than expected; at-risk individuals include patients with cholestasis or chronic diarrheal diseases, dialysis patients, and people on long-term zinc supplements. The role of copper in cardiovascular disease, immune function, bone health, or neurodegenerative diseases is still unclear.

Toxicity Copper toxicity is usually accidental (Table 344-2). In severe cases, kidney failure, liver failure, and coma may ensue. In Wilson's disease, mutations in the copper-transporting *ATP7B* gene lead to accumulation of copper in the liver and brain, with low blood levels due to decreased ceruloplasmin **(Chap. 427)**. A potential negative role of copper in the pathogenesis of Alzheimer's disease has been reported.

■ SELENIUM

Selenium, in the form of selenocysteine, is a component of the enzyme glutathione peroxidase, which serves to protect proteins, cell membranes, lipids, and nucleic acids from oxidant molecules. As such, selenium is being actively studied as a chemopreventive agent against certain cancers, such as prostate cancer. However, it remains unclear whether selenium is effective as a chemopreventive agent or whether it increases cancer risk (e.g., prostate cancer). Convincing evidence for a protective effect of selenium on cognitive decline or cardiovascular disease risk is presently lacking. Selenocysteine is also found in the deiodinase enzymes, which mediate the deiodination of thyroxine to triiodothyronine **(Chap. 394)**. Rich dietary sources of selenium include seafood, muscle meat, and cereals, although the selenium content of cereal is determined by the soil concentration. Countries with low soil concentrations include parts of Scandinavia, China, and New Zealand. *Keshan disease* is an endemic cardiomyopathy found in children and young women residing in regions of China where dietary intake of selenium is low (<20 μg/d). Concomitant deficiencies of iodine and selenium may worsen the clinical manifestations of cretinism. Chronic ingestion of large amounts of selenium leads to selenosis, characterized by hair and nail brittleness and loss, garlic breath odor, skin rash, myopathy, irritability, and other abnormalities of the nervous system.

■ CHROMIUM

Chromium potentiates the action of insulin in patients with impaired glucose tolerance, presumably by increasing insulin receptor–mediated signaling, although its usefulness in treating type 2 diabetes is uncertain. In addition, improvement in blood lipid profiles has been reported in some patients. The usefulness of chromium supplements in muscle building has not been substantiated. Rich food sources of chromium include yeast, meat, and grain products. Chromium in the trivalent state is found in supplements and is largely nontoxic; however, chromium-6 is a product of stainless steel welding and is a known pulmonary carcinogen as well as a cause of liver, kidney, and CNS damage.

■ MAGNESIUM

See Chap. 421.

■ FLUORIDE, MANGANESE, AND ULTRATRACE ELEMENTS

An essential function for fluoride in humans has not been described, although it is useful for the maintenance of structure in teeth and bones. Adult fluorosis results in mottled and pitted defects in tooth enamel as well as brittle bone (skeletal fluorosis).

Manganese and molybdenum deficiencies have been reported in patients with rare genetic abnormalities and in a few patients receiving prolonged total parenteral nutrition. Several manganese-specific enzymes have been identified (e.g., manganese superoxide dismutase). Deficiencies of manganese have been reported to result in bone demineralization, poor growth, ataxia, disturbances in carbohydrate and lipid metabolism, and convulsions.

Ultratrace elements are defined as those needed in amounts <1 mg/d. Essentiality has not been established for most ultratrace elements, although selenium, chromium, and iodine are clearly essential **(Chap. 394)**. Molybdenum is necessary for the activity of sulfite and xanthine oxidase, and molybdenum deficiency may result in skeletal and brain lesions.

■ FURTHER READING

Combs GF Jr, McClung JP: *The Vitamins: Fundamental Aspects in Nutrition and Health*, 6th ed. London, Academic Press, 2022.

DeAngelo SL et al: Selenoproteins and tRNA-Sec: Regulators of cancer redox homeostasis. Trends Cancer 9:1006, 2023.

Fan D et al: Cell signaling pathways based on vitamin C and their application in cancer therapy. Biomed Pharmacother 162:114695, 2023.

Imdad A et al: Vitamin A supplementation for preventing morbidity and mortality in children from six months to five years of age. Cochrane Database Syst Rev 3:CD008524, 2022.

Lassi ZS et al: Zinc supplementation for the promotion of growth and prevention of infections in infants less than six months of age. Cochrane Database Syst Rev 4:CD010205, 2020.

Mechanick JI et al: Clinical practice guidelines for the perioperative nutrition, metabolic, and nonsurgical support of patients undergoing bariatric procedures–2019 update. Surg Obes Relat Dis 16:175, 2020.

Ota Y et al: Comprehensive review of Wernicke encephalopathy: Pathophysiology, clinical symptoms and imaging findings. Jpn J Radiol 38:809, 2020.

Staub E et al: Enteral zinc supplementation for prevention of morbidity and mortality in preterm neonates. Cochrane Database Syst Rev 3:CD012797, 2021.

345

Malnutrition and Nutritional Assessment

Charles Chin Han Lew, Charlene W. Compher

Malnutrition occurs in 30–50% of hospitalized patients depending on the setting, the patient's diagnosis, and the criteria that are used to diagnose malnutrition. Notable adverse outcomes associated with malnutrition include poor wound healing, compromised immune status, and impaired organ function. These can lead to increased length of hospital stay, readmissions, higher mortality, and increased health care costs. It is now widely appreciated that acute or chronic inflammation

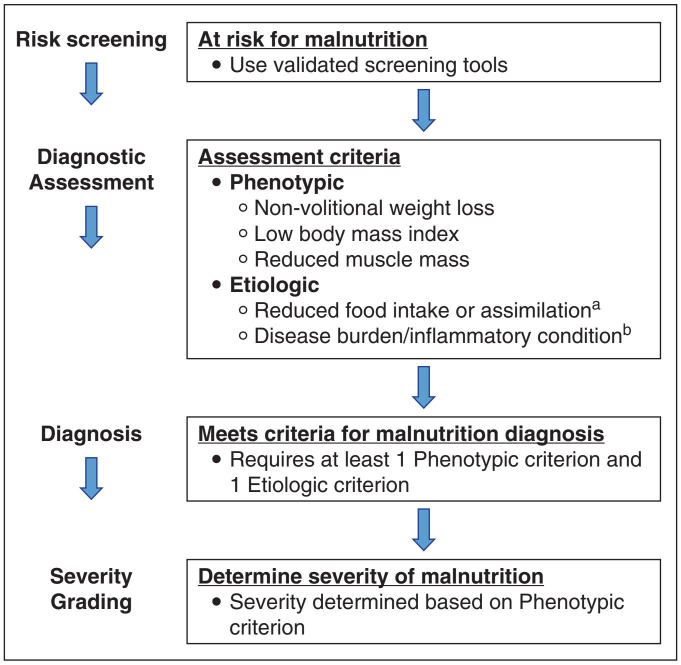
Risk screening
At risk for malnutrition
• Use validated screening tools
Diagnostic Assessment
Assessment criteria
• Phenotypic
○ Non-volitional weight loss
○ Low body mass index
○ Reduced muscle mass
• Etiologic
○ Reduced food intake or assimilation[a]
○ Disease burden/inflammatory condition[b]
Diagnosis
Meets criteria for malnutrition diagnosis
• Requires at least 1 Phenotypic criterion and 1 Etiologic criterion
Severity Grading
Determine severity of malnutrition
• Severity determined based on Phenotypic criterion

FIGURE 345-1 Global Leadership Initiative on Malnutrition (GLIM) diagnostic scheme for screening, assessment, diagnosis, and grading of malnutrition. [a]≤50% of energy requirement >1 week, or any reduction for >2 weeks, or any chronic gastrointestinal condition that adversely impacts food assimilation or absorption. [b]Acute disease/injury (e.g., major infection, burns, trauma or closed head injury) or chronic disease (e.g., malignant disease, chronic obstructive pulmonary disease, congestive heart failure, chronic renal disease or any disease with chronic or recurrent inflammation) or C-reactive protein may be used as a supportive laboratory measure (mild inflammation: 3.0-9.9 mg/L, moderate inflammation: 10-50 mg/L, severe inflammation: >50 mg/L). *(Reproduced with permission from Jensen GL et al: GLIM criteria for the diagnosis of malnutrition: A consensus report from the global clinical nutrition community. JPEN J Parenter Enteral Nutr 43:32, 2019, Figure 1.)*

contributes to the pathophysiology of malnutrition. The presence of inflammation can also render historic nutrition parameters, like albumin and prealbumin, unreliable. In patients with high levels of inflammation, nutrition care is supportive. At moderate or low levels of inflammation, nutrition care may be therapeutic in reducing the nutritional deficits and improving clinical outcomes.

MALNUTRITION DIAGNOSIS IN CLINICAL SETTINGS

Several tools have been used to diagnose malnutrition, such as the Subjective Global Assessment, Mini Nutrition Assessment, and the Academy of Nutrition and Dietetics and American Society for Parenteral and Enteral Nutrition Indicators of Malnutrition. These tools use common parameters such as dietary intake and weight history, muscle and fat loss, body mass index (BMI), and/or physical function to diagnose malnutrition. However, nuances between them lead to incomparable malnutrition prevalence rates in the literature.

In 2019, four international nutrition societies introduced the Global Leadership Initiative on Malnutrition (GLIM) consensus criteria for diagnosing malnutrition in adults **(Fig. 345-1)**. GLIM incorporates broadly available nutrition parameters that are commonly used in other diagnostic tools to enable data from these earlier assessment methods to be described in the GLIM framework. To diagnose malnutrition and its severity, GLIM requires one etiologic criterion (inflammation/disease burden or reduced food intake/assimilation) and one phenotypic criterion (weight loss, low BMI, or reduced muscle mass) **(Table 345-1)**. To facilitate the diagnosis of malnutrition in settings with variable availability and expertise of personnel and equipment across clinical settings, several options are suggested for the phenotypic criterion. The phenotypic criterion is used to define the severity of malnutrition. Recent publications highlight GLIM's widespread adoption in hospitals and clinical settings, showing predictive value for adverse outcomes.

NUTRITION ASSESSMENT FOR PROTEIN-CALORIE MALNUTRITION

The Nutrition Care Process in the hospital setting involves screening for malnutrition risk, comprehensive nutrition assessment of patients at malnutrition risk, development of a nutrition care plan, and monitoring for needed adjustments to the plan. Screening for the risk of malnutrition is typically undertaken by the bedside nurse or by a clinical dietitian using validated tools such as the Malnutrition Screening Tool, Malnutrition Universal Screening Tool, Nutritional Risk Screening 2002, or other in-house tools. A comprehensive nutrition assessment includes a review of medical, surgical, and social history; dietary intake and weight history; medication profile; available laboratory, diagnostic tests, body composition, and anthropometric measures; and nutrition-focused physical examination. The clinical dietitian designs an individualized intervention and monitoring plan for the patient. The nutrition assessment findings and care plan are shared with the physician who may make a malnutrition diagnosis and prescribe the recommended nutrition care plan. The response to nutrition therapy is monitored with adjustments to the care plan as needed. Micronutrient deficiencies of clinical relevance may be detected in association with malnutrition, but a detailed discussion of their assessment is beyond the scope of this chapter **(see Chap. 344)**.

In ambulatory clinic settings, physicians may have less access to dietitian support than in inpatient settings. However, the GLIM criteria are designed to be feasible even in an outpatient setting. Nurses can obtain measures of height and weight (for BMI) and calf or arm circumference to assess muscle mass. A diagnosis of malnutrition can be made in the clinic with referral of the patient to a dietitian for a detailed nutrition assessment and development of a care plan.

COMPONENTS OF A DETAILED NUTRITIONAL ASSESSMENT

Medical/Surgical History and Clinical Diagnosis The risk of malnutrition varies widely across medical/surgical diagnoses and clinical settings. During the early stages of diseases, malnutrition may be less likely than with disease progression. Knowledge of a patient's

TABLE 345-1 Thresholds for Severity Grading of Malnutrition Into Stage 1 (Moderate) and Stage 2 (Severe) Malnutrition

	PHENOTYPIC CRITERIA[a]		
	WEIGHT LOSS (%)	LOW BODY MASS INDEX (kg/m^2)[b]	REDUCED MUSCLE MASS[c]
Stage 1/moderate malnutrition (requires 1 phenotypic criterion that meets this grade)	5%–10% within the past 6 months, or 10%–20% beyond 6 months	<20 if <70 years, <22 if ≥70 years	Mild-to-moderate deficit (per validated assessment methods; see below)
Stage 2/severe malnutrition (requires 1 phenotypic criterion that meets this grade)	>10% within the past 6 months, or >20% beyond 6 months	<18.5 if <70 years, <20 if ≥70 years	Severe deficit (per validated assessment methods; see below)

[a]Severity grading is based on the noted phenotypic criteria, whereas the etiologic criteria described in the text and Figure 345-1 are used to provide the context to guide intervention and anticipated outcomes. [b]Further research is needed to secure consensus reference body mass index for Asian populations in clinical settings. [c]For example, appendicular lean mass index (kg/m^2) by dual-energy absorptiometry or corresponding standards using other body composition methods such as bioelectrical impedance analysis, computed tomography, or magnetic resonance imaging. When not available or by regional preference, physical examination or standard anthropometric measures such as mid-arm muscle or calf circumferences may be used. Functional assessments such as hand-grip strength may be used as a supportive measure.

Source: Reproduced with permission from Jensen GL et al: GLIM criteria for the diagnosis of malnutrition: A consensus report from the global clinical nutrition community JPEN J Parenter Enteral Nutr 43:32, 2019, Table 4.

medical/surgical history and associated clinical diagnoses is especially helpful in discerning the likelihood of malnutrition and inflammation. Several conditions or diseases are characterized by severe acute inflammatory response (e.g., critical illness), whereas others are more typically associated with a chronic inflammatory response that is mild to moderate in severity (e.g., chronic cardiometabolic, oncologic, or gastrointestinal disease, or organ failure) and may be relapsing and remitting. It is common for acute inflammatory events to be superimposed on patients with chronic conditions; for example, a patient with chronic renal disease may be admitted to the hospital with sepsis, such that the acute increase in inflammation should be time-limited while the chronic inflammation will continue.

The inflammatory milieu, especially when severe, may modify nutrient requirements by elevating resting energy expenditure and promoting anabolic resistance and catabolism in muscle. Inflammation also promotes anorexia, decreases food intake, and further compromises nutritional status. A deteriorating course may result because severe inflammation may reduce the benefit of nutritional interventions to being supportive rather than therapeutic, and the associated malnutrition may diminish the effectiveness of medical therapies.

It is also imperative to recognize medical/surgical conditions that place patients at increased risk of becoming malnourished because they have increased nutritional requirements or compromised intake or assimilation. Chronic gastrointestinal disease, even if relapsing and remitting, may limit nutrient assimilation. Therefore, patients with seemingly adequate oral intake may still be at risk of malnutrition due to compromised nutrient assimilation.

Clinical Signs and Physical Examination Nonspecific clinical indicators of inflammation include fever, hypothermia, and tachycardia. The nutrition-focused physical examination should identify edema as well as physical signs of weight gain/loss. Physical findings of weight loss associated with decreased muscle and subcutaneous fat mass should be noted. If edema is present, an estimation of dry weight may be needed to identify actual weight loss.

Anthropometric Data Nonvolitional weight loss is a well-validated nutrition parameter associated with underlying disease or inflammation. The degree and duration of nonvolitional weight loss determine its clinical significance. A 5–10% loss of body weight over 6 months is clinically relevant, whereas >10% loss over the same duration is severe.

The trend in body weight may be obtained from patients, their family or caregivers, and the medical record. Obtaining body weight at each clinic, emergency department, and hospital visit will support efforts to detect changes in this important parameter. Training staff members to weigh patients consistently without shoes and heavy outer garments and maintaining scales and stadiometers in good calibration facilitate acquisition of trustworthy weight data. Clinical sites may maintain chair or bed scales for patients who cannot stand. Height may be estimated by doubling the arm span measurement from the sternal notch to the end of the longest finger or by measurement of knee height using a caliper.

BMI, defined as weight (kg)/height (m^2), is used primarily to screen for overweight or obesity. The National Institutes of Health/World Health Organization BMI categories for adults are as follows: BMI <18.5 kg/m^2 = underweight, BMI 18.5–24.9 kg/m^2 = desirable, BMI 25.0–29.9 kg/m^2 = overweight, and BMI ≥30 kg/m^2 = obese. Note that patients of any BMI status can be malnourished. Overweight or obese patients can be malnourished due to weight loss or loss of muscle mass (e.g., sarcopenic obesity). Similarly, being underweight does not equate to malnutrition. BMI <18.5 kg/m^2 is infrequent in economically developed countries.

Classical anthropometric measurements such as arm and calf circumference can be helpful. However, measuring the calf circumference may be more accessible as it requires less specialized training in comparison to measuring the arm circumference. Additionally, calf circumference has established cutoff values validated for various BMI ranges and ethnicities. Clinically available body composition measures include bioelectrical impedance analysis (BIA), dual-energy x-ray absorptiometry (DXA), and air displacement plethysmography (**Table 345-2**). BIA has the advantage of being portable and is used for patients in diverse clinical settings. Abdominal computed tomography (CT) and magnetic resonance imaging (MRI) taken for disease diagnostic purposes can be interpreted to measure muscle mass, though they are not routinely used to aid in the diagnosis of malnutrition.

Laboratory Indicators There is, unfortunately, no biomarker descriptive of comprehensive nutritional status. Laboratory findings (Table 345-2) should be evaluated in combination with other assessment parameters to appropriately diagnose malnutrition. Albumin and prealbumin are often available in patients with suspected malnutrition; however, these measures have poor sensitivity and specificity as indicators of malnutrition. Albumin and prealbumin are negative acute-phase reactants with levels that are reduced by the systemic response to injury, disease, or inflammation. C-reactive protein, a positive acute-phase reactant, may be useful to identify the presence of inflammation, particularly if the clinical context is unclear about the extent of inflammation. Trends in repeated levels of C-reactive protein over time may help to clarify the extent and trajectory of the inflammatory challenge. Research suggests that interleukin 6, and perhaps other cytokines, may offer promise as indicators of inflammatory status. Nonspecific laboratory indicators that are often associated with inflammatory response include leukocytosis and hyperglycemia. Additional tests that may be obtained to help confirm the presence of inflammatory response include 24-h urine urea nitrogen and indirect calorimetry to measure energy expenditure. In the setting of a severe acute systemic inflammatory response, negative nitrogen balance and elevated resting energy expenditure are anticipated.

Diet History Brief diet surveys are available to target specific aspects of the diet, such as food patterns, to prevent disease or track intake of nutrients of key importance such as calcium, vitamin D, and phosphorus for bone disease management. However, more specific information about the adequacy of intake of key nutrients or food patterns relative to estimated requirements is most often obtained by consultation with a clinical dietitian. Since patients often present to health care practitioners with acute medical events superimposed upon chronic health conditions, it is common for patients to have had decreased food intake and progressive malnutrition for extended periods prior to assessment. Therefore, compromised dietary intake must not be overlooked so that appropriate intervention may be undertaken.

Assessment of the adequacy of nutritional intake must include oral nutrition supplements, enteral tube feedings, and parenteral nutrition both during hospitalization and as a component of home nutrition support. Ongoing reassessment of actual intake received by the patient is undertaken in hospital settings, most commonly by the clinical dietitian, the bedside nurse, and the nutrition support team if available. Changes in the adequacy of oral intake, body weight, or tolerance to enteral or parenteral feedings would all require adjustment of the feedings to optimize patient care.

A comprehensive assessment should also include a review of medications, vitamin/mineral supplements, or herbal remedies with attention to undesirable nutrition-sensitive side effects such as anorexia, dysgeusia, oral mucositis, nausea, vomiting, diarrhea, and constipation. Potential drug–nutrient interactions should also be identified to enable alternative therapeutic options.

Functional Outcomes Advanced malnutrition is accompanied by declines in muscle mass and function that can be detected by various functional test outlined in Table 345-2. Among them, handgrip strength may have better clinical utility since it is easily assessed and helps to identify which patients benefit most from individualized nutritional support.

In summary, malnutrition puts patients at risk of adverse outcomes. Therefore, timely diagnosis and treatment of malnutrition should be provided to improve patient outcome. Robust evidence exists demonstrating that a multipronged approach to malnutrition intervention reduces mortality and holds promise to improve the quality of life of patients.

TABLE 345-2 Common Body Composition Studies, Laboratories, and Other Studies Used in Nutrition Assessment

TEST	NOTES
Body Composition Studies (Recommended)	
Air plethysmography	May be used to assess body composition. It comprises a dual-chamber, sealed compartment containing an oscillating diaphragm, which allows it to measure body volume using Poisson's law, and compute fat mass and fat-free mass. However, its validity in a racially diverse population needs further studies, and it is not readily available in most hospitals.
Bioelectrical impedance analysis (BIA)	A promising method to assess body composition as the equipment is easily portable. It uses the opposition of an electrical current through body tissues (i.e., impedance) and population-specific equations to estimate total body water and body composition. BIA equations are highly device and population specific. Fever, some medications, and fluid/electrolyte disturbances can influence its precision and accuracy.
Calf circumference	A practical and valid clinical estimate of skeletal muscle mass, in which low calf circumference indicates muscle loss. Ethnic specific reference ranges are available, and measurements should be adjusted in persons with body mass index different from the normal range (underweight, overweight, or obese). The technique requires less training than mid-upper arm circumference.
Dual energy x-ray absorptiometry (DXA)	May be used to assess regional and total fat mass in a diverse group of adult clinical patients. However, its validity in assessing muscle mass is unclear.
Imaging with computed tomography (CT) or magnetic resonance imaging (MRI)	May be used to quantify fat and fat-free mass when scans are taken for other diagnostic purposes. Both are costly, and CT entails x-ray exposure.
Mid-upper arm circumference	Mid-upper arm circumference is the circumference of the upper arm measured at the midpoint between the olecranon process and the acromion process. It estimates both muscle and subcutaneous fat stores, but references are population specific. Take note that this measurement technique requires reliability training.
Ultrasound	A promising method to assess body composition as the equipment is easily portable. It uses high-frequency sound waves to capture live images of muscle tissues. Since most studies examined a single muscle, it is unclear if results can be extrapolated to reflect overall nutritional status. More research is needed to establish standardization of measurement protocol.
Laboratory Tests and Other Studies (Recommended)	
Complete blood count with differential	May be used to screen for nutritional anemias (iron, B_{12}, and folate deficiencies) and thrombocytopenia (vitamin C and folate deficiencies).
C-reactive protein (CRP)	May be used to confirm systemic inflammation. While inflammation may be associated with malnutrition, CRP lacks specificity as a biomarker for diagnosing malnutrition. Nevertheless, when combined with other nutrition assessment methods, CRP can complement the diagnostic process for malnutrition. In addition, elevated levels may be associated with reduced food intake and a lack of response to nutritional interventions.
Indirect calorimetry	May be used to determine resting energy expenditure (REE) to aid in determining caloric intake goal. Predictive equations that were based on REE measures in specific populations are also used to estimate REE, especially when indirect calorimetry is not available.
Nitrogen balance (NB)	May be used to reflect the degree of catabolism and adequacy of protein replacement delivery in patients with normal renal and liver function. Method requires collection of 24-h urine. Nitrogen balance = (protein delivery [g]/6.25) – (urinary urea nitrogen + 4 insensible losses).
Specific micronutrients	May be used to validate clinical symptoms of micronutrient deficiencies. Exercise caution with interpretation since levels can be influenced by acute illness.
Laboratory Tests and Other Studies (Not Recommended)	
Blood urea nitrogen (BUN), serum creatinine	Not recommended as malnutrition indicators. Although low BUN and serum creatinine may reflect reduced muscle mass, their levels are influenced by factors unrelated to malnutrition (e.g., renal and hepatic insufficiency).
Cholesterol	Not recommended as a malnutrition indicator. While becoming hypocholesterolemic (<120 mg/dL) during hospitalization may be linked to poorer clinical outcomes, insufficient evidence supports its use for diagnosing malnutrition and evaluating nutritional interventions.
Creatinine height index (CHI)	Not recommended as a malnutrition indicator. CHI = (24-h urinary creatinine excretion/ideal urinary creatinine for gender and height) × 100. Although urinary creatinine excretion reflects muscle mass, it can be influenced by renal insufficiency, meat consumption, physical activity, and trauma. Requires accurate 24-h urine collection.
Cytokines	Not recommended as malnutrition indicators. Their cutoff values and role as an indicator of inflammatory status are still being studied.
Electrocardiogram	Not recommended as a malnutrition indicator. Although prolonged QT interval may be present in severely malnourished patients, the former can be influenced by factors unrelated to malnutrition.
Fecal fat test	May be used to validate clinical symptoms of fat malabsorption. It reflects the percentage of dietary fat that the body does not absorb. Requires stool samples collected over 72 h during ingestion of high-fat diet.
Serum proteins (albumin, prealbumin, transferrin, and retinol-binding protein)	Not recommended as a malnutrition indicator. Their levels are influenced by factors other than malnutrition (e.g., systemic inflammation, hepatic and renal insufficiency, protein-losing enteropathies, corticosteroids, hydration, and iron status), reducing their specificity in diagnosing malnutrition and evaluating nutritional interventions. However, they remain valuable for predicting clinical outcomes.
Skin testing—recall antigens	Not recommended as a malnutrition indicator. Although delayed hypersensitivity is associated with malnutrition, the former can be influenced by factors unrelated to malnutrition.
Total lymphocyte count	Not recommended as malnutrition indicator. Low levels are influenced by factors unrelated to malnutrition.
Urine 3-methylhistidine	Not recommended as a malnutrition indicator. Although it reflects muscle mass, its level is influenced by meat intake and poorly reflects changes in body protein stores. Requires accurate 24-h urine collection

(*Continued*)

TABLE 345-2 Common Body Composition Studies, Laboratories, and Other Studies Used in Nutrition Assessment (Continued)

TEST	NOTES
Functional tests (Recommended)	
Clinical Frailty Scale	A tool that measures frailty, which is defined as a state of increased vulnerability to adverse health outcomes due to reduced physiologic reserves. It uses various self-reported factors (e.g., physical mobility, cognitive function, and the ability to perform daily activities) to classify frailty into nine classifications, ranging from very fit (level 1) to severely frail (level 9). Patients who score ≥5 are considered frail.
Sit-to-Stand Test	Assesses lower limb strength, balance control, and functional mobility by measuring the time required for a person to transfer from a seated to a standing position and back to sitting five times.
Handgrip strength	A measure of muscular strength generated by one's forearm muscles to screen for upper body strength. Low grip strength is associated with poorer clinical outcomes. Patients with low grip strength may benefit most from personalized nutritional support.
SARC-F	A screening tool that determines the risk of sarcopenia, which is defined as a progressive and generalized loss of skeletal muscle mass and strength. It uses self-reported deficiencies in strength, walking, rising from a chair, climbing stairs, and experiencing falls to provide a score from 0 to 10, and a score ≥4 is considered at risk of sarcopenia.
6-Meter Walk Test	Assesses walking speed (meter/sec) over a 6-m distance to determine functional mobility, gait, and vestibular function.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Gordon L. Jensen for contributions to this chapter in previous editions of Harrison's.

FURTHER READING

Cederholm T, Bosaeus I: Malnutrition in adults. N Engl J Med 391:155, 2024.

Gonzalez CC et al: Calf circumference: Cutoff values from NHANES 1999-2006. Am J Clin Nutr 113:1679, 2021.

Guerra RS et al: Usefulness of six diagnostic and screening measures for undernutrition in predicting length of hospital stay: A comparative analysis. J Acad Nutr Diet 115:927, 2015.

Jensen GL et al: GLIM criteria for the diagnosis of malnutrition: A consensus report from the global clinical nutrition community. J Parenter Enteral Nutr 43:32, 2019.

Lew CCL et al: Association between malnutrition and clinical outcomes in the intensive care unit: A systematic review. J Parenter Enteral Nutr 41:744, 2017.

Schuetz P et al: Individualized nutritional support in medical inpatients at nutritional risk: A randomized clinical trial. Lancet 393:2312, 2019.

Uhl S et al: Interventions for malnutrition in hospitalized adults. J Hosp Med 17:556, 2022.

White JV et al: Consensus statement: Academy of Nutrition and Dietetics and American Society for Parenteral and Enteral Nutrition. Characteristics recommended for the identification and documentation of adult malnutrition (under-nutrition). J Parenter Enteral Nutr 36:275, 2012.

Wong A et al: An umbrella review and meta-analysis of interventions, excluding enteral and parenteral nutrition, initiated in the hospital for adults with or at risk of malnutrition. Am J Clin Nutr 118:672, 2023.

346

Enteral and Parenteral Nutrition

L. John Hoffer, Bruce R. Bistrian

There are three kinds of specialized nutritional support (SNS): (1) optimized voluntary nutritional support, which is indicated when a patient's barriers to adequate nutrition can be overcome by special attention to the details of how their food is constituted, prepared, and served and its consumption optimized; (2) instrumental enteral nutrition (EN), in which a liquid nutrient formula is delivered through a tube placed in the stomach or small intestine; and (3) parenteral nutrition (PN), in which a nutritionally complete mixture of crystalline amino acids, glucose, lipid emulsions, minerals, electrolytes, and micronutrients is infused directly into the bloodstream.

When does a hospitalized adult patient need SNS, and when it's indicated, how should it be provided? This chapter summarizes the physiologic principles that guide the correct use of SNS and offers practical information about the diagnosis and management of nutritional disorders in adult hospitalized patients.

The management of in-hospital nutritional disorders follows three steps: (1) screening and diagnosis; (2) determination of the severity and urgency of treating a diagnosed nutritional disorder in its overall clinical context; and (3) selection of the kind of SNS to provide, its composition, and the details of providing it. To follow these steps properly, physicians require a general understanding of nutritional physiology, nutrient requirements, the pathophysiology and diagnosis of the nutritional disorders, and familiarity with the indications, advantages, risks, and administration of the different kinds of SNS. Because most physicians lack formal clinical nutrition training, they should, in general, collaborate with clinical dietitians and specialized pharmacists in this process.

NUTRITIONAL PHYSIOLOGY

(See Chaps. 343–345.)

Energy The total daily energy expenditure of a healthy sedentary adult is ~36 kcal/kg. Resting energy expenditure accounts for ~75% of this total. Resting energy expenditure can be measured by indirect calorimetry or estimated using a variety of predictive equations that input weight, height, age, sex, and sometimes, disease-related factors. Fever and some diseases increase resting energy expenditure; prolonged semi-starvation triggers an adaptive reduction in resting and total energy expenditure. Total energy expenditure determines how much dietary energy must be consumed to maintain an existing store of body fat and protein. The amount of energy a patient actually *requires* may be less than their total energy expenditure, as in obesity therapy or during temporary periods of exogenous energy intolerance, or greater than total energy expenditure during recovery from starvation disease.

Protein and Amino Acids Dietary protein provides the 23 essential and nonessential amino acids required for protein synthesis. Protein must be consumed throughout life, because endogenous protein turnover entails a minimum obligatory rate of amino acid catabolism. Amino acid catabolism is regulated, appropriately increasing and decreasing in response to variations in protein intake, but it cannot fall below a certain minimum rate—the "protein minimum"—that determines an individual's minimum dietary protein requirement. The average daily minimum protein requirement for a healthy adult consuming an otherwise adequate diet is 0.65 g/kg; the "safe" or "recommended" intake is 0.80 g/kg. Average protein consumption in wealthy societies is approximately twice the average minimum requirement.

Energy balance affects the minimum protein requirement. Positive energy balance, when achieved by greater carbohydrate consumption, can improve body protein retention. Negative energy balance, such as during therapeutic weight reduction, reduces the efficiency of protein turnover and increases the dietary protein requirement. Some diseases (or their treatments) increase the protein requirement, by (1) increasing amino acid loss from the body (as in malabsorption and protein loss via wound exudates, fistulas, or inflammatory diarrhea), removing amino acids from the circulation (renal replacement therapy), or (2) increased muscle protein catabolism, which occurs as an adverse effect of high-dose glucocorticoid therapy and, most commonly, as part of the metabolic response to the systemic inflammation that accompanies major injury, serious infections, and intense immune system activation. A highly protein-catabolic patient may excrete 15 g N (nitrogen)/d or more in their urine in the absence of dietary protein provision: this is more than three times faster than occurs with simple fasting. Since 1 g N lost from the body reflects the loss of 6.25 g formed protein, 15 g N loss/d implies the loss of 15 × 6.25 = 94 g protein per day. Since the body's hydrated metabolically active tissue mass (also called body cell mass) is ~20% protein, one may calculate that 94 g protein loss/d indicates a loss from the body of ~470 g (1 lb) of active tissue mass per day. Sufficiently generous protein provision can mitigate this kind of active tissue loss. The extent to which protein-catabolic disease increases the protein requirement is debated, but a common recommendation for critically ill patients is 1.2–2.0 g protein/kg of normal body weight per day—close to the habitual protein intake of healthy people in wealthy societies.

Protein-Energy Interactions Both energy deficiency and systemic inflammation increase the dietary protein requirement; how do these factors interact? The loss of body protein caused by energy deficiency can be prevented or minimized by increased protein consumption. Systemic inflammation reduces the benefit of increased protein provision under hypocaloric conditions but does not eliminate it. Some evidence suggests that a minimum energy dose, such as 50% of total energy expenditure, may help preserve the active tissue store during severe systemic inflammation. It is clear, however, that energy doses >70% of total energy expenditure do not have any greater protein-sparing effect, and the additional amounts of glucose and fluid volume required to deliver them can have adverse effects that include hyperglycemia, dyslipidemia, and extracellular volume expansion.

Permissive Underfeeding and Hypocaloric Nutrition These terms have different meanings, and they should not be conflated. Permissive underfeeding refers to the deliberate underprovision of all nutrients including protein, whereas hypocaloric nutrition refers to reduced energy provision combined with increased protein to compensate for the known protein-wasting effect of negative energy balance.

Micronutrients Minimum amounts of the nine water-soluble vitamins (the B vitamins and vitamin C), four fat-soluble vitamins (A, D, E, and K), eight minerals (calcium, phosphorus, potassium, sodium, chloride, magnesium, zinc, and iron), essential fatty acids, and several essential trace elements are required to avoid deficiency diseases. Overt deficiencies of potassium, sodium, magnesium, and phosphorus are so common in hospitalized patients that it is standard practice to monitor for and correct them. Some drugs cause renal potassium, magnesium, or zinc losses that require increased provision. Gastrointestinal losses from nasogastric drainage tubes or intestinal losses from fistulas or diarrhea incur losses of potassium, sodium, calcium, magnesium, and zinc.

Less investigated, but common, are subclinical deficiencies of zinc, vitamin C, vitamin D, and possibly other micronutrients. Physicians often assume that consumption of a regular hospital diet protects patients from micronutrient deficiencies. This assumption is unwarranted when the patient's nutritional status is deficient upon admission and remains so during their hospital stay.

Systemic inflammation can complicate the clinical laboratory evaluation of micronutrient status. It increases the requirements for certain micronutrients, partially redistributing them from the extracellular compartment into tissue stores, as occurs with iron, selenium, zinc, and perhaps other nutrients. This effect reduces the reliability of many standard laboratory tests of micronutrient status. The evaluation of a patient's micronutrient status should begin with a determination of their previous and current intake and health status, including the identification of potential routes of excessive micronutrient loss. In acute-care clinical settings, the intensity of systemic inflammation should be determined using clinical criteria and by measuring serum concentrations of albumin or C-reactive protein, an acute-phase protein whose wide dynamic range and short half-life make it particularly useful in this situation,

MACRONUTRIENT MALNUTRITION SYNDROMES

The decision to embark on SNS must be justified by a well-formulated nutritional diagnosis and a clearly defined therapeutic goal. This chapter focuses on the diagnosis, treatment, and prevention of in-hospital adult starvation-related malnutrition (SRM) and two related conditions: chronic disease–related malnutrition (CDM) and acute disease–related malnutrition (ADM). As explained in **Chap. 345**, SRM is the disease brought about by prolonged semi-starvation. Synonyms for this disease are "starvation-induced protein-energy malnutrition" and "starvation disease." CDM can be understood as SRM (simple starvation disease) that is combined and compounded by moderately severe systemic inflammation. SRM and CDM are anatomically similar but etiologically and metabolically different, as elaborated upon below. ADM refers to an injury-induced metabolic condition that creates a high risk of severe body protein deficiency, rather than to an already-existing anatomic starvation disease.

Starvation-Related Malnutrition The pathologic features that make SRM a disease, and distinguish it from the semi-starvation that precedes it, emerge when the patient's active tissue mass has been depleted enough to impair specific physiologic functions.

The body normally adapts to sustained inadequate energy provision by reducing energy expenditure and net protein catabolism, partly through hormone- and nervous system–regulated alterations in cellular metabolism, but mostly from loss of muscle, which is the major generator of energy expenditure. (Visceral organ mass and function have a higher priority for protection during starvation.) These adaptations permit prolonged survival, although at a cost that includes muscle loss (including cardiac and respiratory muscle), skin thinning, lethargy, a tendency to hypothermia, and functional disability. The cardinal anatomic features of SRM—generalized muscle loss and subcutaneous adipose tissue depletion—are easy to detect by simple physical examination.

SRM always manifests as weight loss, but weight loss alone may not reveal its full severity. Semistarvation increases the extracellular fluid (ECF) volume, sometimes seriously enough to cause edema. Starvation-associated edema may occur even in the absence of its known common causes (it is termed "starvation edema"). In adults with initially normal body composition, starvation-induced weight loss linearly tracks the loss of active tissue mass; this occurs because the weight loss caused by the depletion of adipose tissue is counterbalanced by increased ECF volume. A 25% reduction in body weight (and active tissue mass) significantly compromises physiologic function; a 50% reduction places otherwise uncompromised young adults on the cusp of thermodynamic survival; older people with comorbidities are at even greater risk. People with SRM are frail, feel unwell, lack strength and emotional range, and are at risk of hypothermia.

The main cause of SRM worldwide is involuntary food deprivation. Its causes in hospitalized patients are many: inadvertent or physician-ordered food deprivation; psychological depression or distress; anorexia nervosa; poorly controlled pain or nausea; badly presented unappealing food; communication barriers; physical or sensory disability; dysphagia and other mechanical difficulties ingesting food; partial obstruction of the esophagus, stomach, or intestines; thrush; intestinal angina; and most commonly, combinations of these causes.

Chronic Disease–Related Malnutrition and Cachexia These terms refer to SRM combined with and complicated by chronic systemic inflammation. CDM is common among patients suffering from chronic infection; inflammatory autoimmune disease; chronic severe hepatic, renal, cardiac, or pulmonary disease; and neoplastic diseases that induce a systemic inflammatory response or cause tissue injury. CDM induces anorexia—a strong disinclination to eat even when there is no physical barrier to it—and is worsened by it. Its features include muscle loss, weakness, fatigue, and impaired adaption to starvation, all of which contribute to a vicious cycle of worsening starvation disease. Fortunately, the nutritional deficit on the input side of this equation (anorexia-driven inadequate food consumption) is usually the strongest driver of the patient's CDM, making it amenable to a well-organized nutritional intervention while an effective treatment plan for the primary disease is identified and implemented.

Cachexia is an older term now used with varying meanings. It generally refers to a metabolic syndrome of moderate systemic inflammation, unrelenting generalized muscle loss, and their associated symptoms, including anorexia, fatigue, frailty, and mood disturbance. For the purpose of this chapter, it may be considered as roughly synonymous with CDM.

Acute Disease–Related Malnutrition Other terms for ADM are "injury-induced malnutrition" and "protein-catabolic critical illness." The metabolic-inflammatory response to severe tissue injury and sepsis mobilizes muscle amino acids and leads to rapid and severe generalized muscle loss and increased resting energy expenditure under conditions in which voluntary food intake is usually impossible. SRM or CDM may be absent at the onset of a patient's critical illness, but muscle loss will rapidly develop unless the inciting medical or surgical disease is rapidly and effectively treated and SNS provided. Death from simple starvation of nonobese adults typically occurs within ~8 weeks, whereas death from untreated starvation in patients with sustained ADM will occur much sooner.

■ NUTRITIONAL DIAGNOSIS

The cardinal anatomic features of starvation disease (either SRM or CDM) are generalized muscle loss and diminished body fat. Routine physical examination will immediately reveal these features, but, perhaps surprisingly, what ought to be an easy diagnosis is often overlooked. This section explains the details and pitfalls of diagnosing SRM and CDM.

Muscle Mass Muscle tissue normally accounts for ~80% of an adult's active tissue mass. It is easy to detect loss of muscle mass, and its severity is determinable almost at a glance. But generalized muscle loss has many causes: (1) old age–related muscle loss (sarcopenia); (2) disuse muscle atrophy; (3) high-dose glucocorticoid therapy and certain endocrine diseases (uncontrolled diabetes mellitus, adrenocortical insufficiency, hyperthyroidism, androgen deficiency, hypopituitarism); and (4) primary muscle or neuromuscular diseases. The guiding principle is that SRM and CDM are very common treatable causes of muscle loss. Whenever generalized muscle loss is identified, its potential causes should be considered and the treatable ones addressed. Old age is irreversible, but adequate protein and energy provision, combined with physical rehabilitation, can be lifesaving.

Generalized muscle loss of any cause is especially dangerous in ADM, because the combination of ADM and preexisting muscle loss places patients at the cliff edge of lethal depletion of their active tissue store. A diminished muscle mass is less able to release adequate amounts of amino acids into the circulation for protein synthesis at sites of tissue injury and healing and to the central protein pool to regulate the immunoinflammatory response.

Subcutaneous Adipose Tissue Severe adipose tissue depletion almost always indicates starvation disease, but it is not required to make the diagnosis. The current obesity epidemic has created a population of obese patients with SRM or CDM in whom muscle loss has outpaced fat loss. A conscientious physical examination will easily identify the patient's diminished muscle mass beneath their residual subcutaneous fat.

ECF Volume The ECF compartment normally represents ~20% of body weight. SRM moderately increases ECF volume. CDM is associated with additional edema-promoting conditions, especially the hypoalbuminemia associated with systemic inflammation. The effect of increased ECF volume to increase body weight and distort physical appearance should be borne in mind, for it may conceal the severity of muscle loss.

Body Mass Index Body mass index (BMI) is defined as body weight (kg) divided by the square of height (m^2). The normal reference range is 20–25 kg/m^2. BMI <19–20 usually indicates nonpathologically reduced muscle and fat mass. BMI <15 indicates severe starvation disease, and BMI <13 is usually thermodynamically incompatible with life, especially in older patients with comorbidities. Some guidelines and clinical trial enrollment criteria have defined what they term as "malnutrition"—meant as starvation disease—as a BMI <16 or 17. But it is wrong to rely on BMI as the sole basis for diagnosing starvation disease. A BMI <17 certainly implies starvation disease, for the body composition required for such a BMI can only develop by jettisoning a large fraction of the body's active tissue and adipose tissue store. But a BMI >17 does not rule out starvation disease. Many patients with starvation disease have a normal or above-normal BMI, despite important muscle loss, because of residual obesity or an expanded ECF volume.

Visual BMI With some practice, clinicians can learn to accurately predict the BMI of nonobese, nonedematous patients by systematically examining their muscle groups. This skill enables them to evaluate the severity of starvation disease even in obese or edematous patients—for whom measured BMI is unreliable—simply by evaluating their muscle groups while consciously discounting their subcutaneous fat and edema. Visual BMI can be used to estimate a patient's "normalized dry body weight"—what their body weight would be without obesity, edema, or ascites. For example, the normalized dry body weight of a 1.75-m adult with a visual BMI of 17 is $1.75^2 \times 17 = 52$ kg.

Laboratory and Technical Assessment Clinical laboratory measurements have three main purposes in the evaluation and management of starvation disease.

MUSCLE MASS Bedside ultrasound is proving to be a valuable technique for quantifying muscle mass at specific body sites. However, it need not and should not replace the immediate information provided by the eyes, hands, and mind of an astute clinician.

SYSTEMIC INFLAMMATION The absence or presence of systemic inflammation is what fundamentally distinguishes SRM from CDM. The most useful laboratory indicators of systemic inflammation are a reduced serum albumin concentration and increased serum C-reactive protein concentration. Systemic inflammation increases the permeability of capillary walls to large molecules, and the resulting osmotic shift increases ECF volume. As intravascular albumin redistributes in this larger volume, its serum concentration decreases. Variably increased albumin catabolism and reduced liver albumin synthesis also contribute. Dietary protein deficiency and muscle loss perpetuate inflammation-induced hypoalbuminemia, because the amino acids required for hepatic albumin synthesis come from the diet and endogenous muscle protein.

Hypoalbuminemia and a reduced serum prealbumin concentration are often considered as indicators of "malnutrition." This is incorrect. These serum proteins are negative acute-phase reactants that indicate the presence of systemic inflammation. Systemic inflammation induces anorexia and increases muscle catabolism, effects that increase the risk of CDM, but the disease itself may or not exist at the time, and it may never develop. Serum acute-phase reactants will remain abnormal as long as systemic inflammation persists. Appropriate nutrition can mildly improve serum albumin concentration, but it will not fully normalize until after the systemic inflammation subsides. Even then, full normalization may require several weeks because of albumin's serum half-life of at least 2 weeks.

PROTEIN-CATABOLIC INTENSITY The defining feature of protein-catabolic disease is generalized muscle loss. Situations of greatly increased protein loss can be identified by measuring body N loss. Most N leaves the body in the urine (almost all of it in urea, ammonium, and creatinine). Total N is not usually measured in hospital laboratories, but urinary urea N (which normally accounts for ~85% of urinary N) can be measured and is routinely available. A recent, validated formula estimates daily total N loss as follows: N loss (g) = g N in urinary urea/0.85 + 2.

Net muscle protein catabolism follows approximately first-order kinetics; this means that the rate of N loss from muscle is proportional to the amount of N available to be lost. Protein-catabolic muscle-depleted patients will lose less body N per day than equally catabolic patients with a preserved muscle store, but they are at greater risk of succumbing to their disease because they are closer to lethal depletion. Rate of N loss always has to be interpreted in the context of existing muscle mass.

Instrumental Nutritional Assessment Many nutritional assessment instruments claim to identify "malnutrition" by enumerating and summing a list of risk factors, laboratory results, and physical findings. These tools are often hindered by ambiguity about their definition of malnutrition and by failure to distinguish between screening and diagnosis. Correct clinical diagnosis is the identification of a known pathologic entity—SRM or CDM, for example—by considering the patient's medical history, pertinent findings on physical examination, and laboratory or imaging reports. Diagnosis also involves an estimation of the probability that the diagnosis is correct and a judgment of its severity. By contrast, screening is the application of a simple test that identifies people at sufficiently high risk of a certain disease to warrant definitive procedures to establish the diagnosis or rule it out or that identifies people at sufficiently high risk of developing it to warrant specific preventive interventions. Screening tools and risk predictors are extremely useful, but they should not be confused with clinical diagnosis.

SPECIALIZED NUTRITIONAL SUPPORT

Optimized Voluntary Nutritional Support When feasible, this is the approach of choice because it engages and empowers the patient, encourages mobilization and reconditioning, is consistent with the objectives of patient-centered medicine, and is risk-free. Its disadvantage is that it is time-consuming and labor-intensive, and it demands astute attention to the specific needs of individual patients.

Enteral Nutrition This is nutrition provided through a feeding tube placed through the nose into the stomach or beyond it into the duodenum, by insertion of a tube through the abdominal wall into the stomach or beyond it into the jejunum, or by a surgical approach to access the stomach or small intestine. EN is usually the treatment of choice when optimized voluntary nutritional support is impossible or has failed. It is relatively simple, safe, and inexpensive, and it maintains the digestive, absorptive, and immunologic barrier functions of the gastrointestinal tract. EN is appropriate when optimized voluntary nutrition is not feasible or has failed and the patient's gastrointestinal tract is functioning and can be accessed.

EN Products The most common forms of EN are commercially manufactured formulas.

STANDARD POLYMERIC FORMULAS These are the most widely used EN products. They are available in a wide variety of formats designed to meet the nutritional requirements of a normal, healthy person. Carbohydrates provide most of the energy. The proteins are intact and of high biologic value (casein, whey, or soy) and require adequate pancreatic and intestinal enzyme function for digestion and absorption. They are isotonic or nearly so and provide 1000–2000 kcal and 50–70 g protein/L.

POLYMERIC FORMULAS WITH FIBER The addition of dietary fiber to formulas sometimes improves bowel function and feeding tolerance. Fermentable (soluble) fibers such as pectin and guar are metabolized by colonic bacteria, yielding short-chain fatty acids that fuel colonocytes. Nonfermentable (insoluble) fibers increase fecal bulk, improve peristalsis, and may improve diarrhea.

ELEMENTAL AND SEMI-ELEMENTAL FORMULAS The macronutrients in these formulas are partially or completely hydrolyzed. They were designed for patients with known maldigestion and malabsorption but are sometimes used empirically for patients intolerant of a standard polymeric formula, especially when acutely ill, although without strong evidence of superiority.

IMMUNE-ENHANCING FORMULAS In addition to providing macronutrients and conventional amounts of micronutrients, these EN products contain large amounts of certain nutrients designed to favorably modulate the immune response: arginine and n-3 fatty acids especially have some clinical experimental support, but also various combinations of glutamine, nucleotides, and antioxidants for which evidence of benefit is limited or lacking.

PROTEIN-ENRICHED FORMULAS Most EN products provide energy and protein in a ratio appropriate for a healthy person. Protein-enriched formulas provide ~90 g protein and 1000 kcal/L. Originally marketed to meet the increased protein requirement of weight-reducing obese patients, these products are increasingly used to provide protein-catabolic patients with a more generous amount of protein without energy overfeeding. EN can be further protein-enriched by adding flushes of water-soluble powdered protein supplements.

OTHER FORMULAS Various disease-specific EN products are available for patients with diabetes or hepatic, renal, or pulmonary disease. Their use can improve some metabolic endpoints, without clear evidence that they improve clinical outcomes.

Parenteral Nutrition PN delivers a complete nutritional regimen directly into the bloodstream in the form of crystalline amino acids, glucose, triglyceride emulsions, minerals (calcium, phosphate, magnesium, and zinc), electrolytes, and micronutrients. Because of its high osmolarity (>1200 mOsm/L) and often large volume, PN is infused into a central vein in adults. Ready-to-use PN admixtures typically containing 4–7% hydrated amino acids and 20–25% glucose (with or without electrolytes) are available in two-chamber (amino acids and glucose) or three-chamber (amino acids, glucose, and lipid) bags that are intermixed with vitamins, trace minerals, and additional electrolytes then added just prior to infusion. Although convenient and cost-effective, these products have fixed nutrient compositions and are dosed according to the volume required to meet a patient's energy requirement but not necessarily their protein requirement. In some situations—especially ADM—a more sophisticated approach is justified that uses a computer-controlled sterile compounder to create combinations of amino acids and glucose that precisely meet the protein, energy, fluid, and acid-base regulatory requirements of individual patients.

Amino Acids The amino acids in PN admixtures provide sufficient amounts of the essential amino acids and total nonessential amino acid N. Unlike protein-bound amino acids, which are connected by peptide bonds, free amino acids are hydrated. This reduces their energy density from 4.0 (in formed protein) to 3.3 kcal/g and reduces the amount of protein substrate they provide by ~17%. Thus, 100 g of free mixed amino acids provides 83 g protein substrate.

Carbohydrate and Lipids The glucose in PN is dextrose monohydrate; this hydration state reduces its energy concentration from 4.0 kcal/g (as in formed carbohydrate) to 3.4 kcal/g. Lipid emulsions provide energy (~10 kcal/g) and the essential n-6 and n-3 fatty acids. Traditional lipid emulsions based solely on soybean oil are giving way to mixed emulsions that include medium-chain triglycerides, n-9 monounsaturated fatty acids, and n-3 fatty acids. Emulsions of pure soybean oil; or a mixture of 80% olive oil and 20% soybean oil; or a mixture of 30% soybean oil, 30% medium-chain triglycerides, 25% olive oil, and 15% fish oil are available in the United States.

Fish oil (either as a component of a mixed emulsion or administered separately) may reduce the risk of infections and length of stay in critically ill patients. Some complex lipid emulsions are more highly enriched in n-3 fatty acids and/or contain fewer n-6 polyunsaturated fatty acids than soybean lipid, which is more prone to lipid peroxidation and could promote the formation of the proinflammatory n-6 derivatives. Standard lipid infusion rates should not exceed 8 g/h, equivalent to 175 g (1925 kcal)/d for a 70-kg patient. Because of their longer chain length and slower tissue uptake, pure fish oil emulsions are infused at lower rates.

Minerals, Micronutrients, and Trace Elements The default concentrations of electrolytes, minerals, and micronutrients in PN solutions are designed to meet the requirements of a healthy adult. These starting doses must be adjusted to meet the frequently abnormal and often-changing requirements of individual patients. Being unstable, multivitamin mixtures are injected into PN bags just prior to their delivery to the medical unit. Parenteral water-soluble vitamin requirements are greater than standard oral requirements, because hospitalized patients often have vitamin deficiencies or increased requirements and because intravenous administration of vitamins increases their loss in the urine. Ascorbic acid degrades spontaneously in PN solutions, even when protected from light.

APPROACH TO THE PATIENT

Indications, Selection, and Provision of Specialized Nutritional Support

Most hospitalized patients do not require SNS because they can and will eat enough with appropriate management of their primary disease. Others may have a terminal disease whose downward course will not be mitigated by SNS. Patients who are unable to eat and digest enough of the standard hospital diet and have or are at high risk of developing SRM or CDM are candidates for optimized voluntary nutritional support. Instrumental SNS is indicated when optimized voluntary nutritional support is inappropriate, impractical, or has failed. The decision to provide EN or PN is based on a combination of four factors: (1) the determination that nutrient ingestion will likely continue to be inadequate for many days; (2) there is important muscle loss (of any cause); (3) the patient's nutrient requirements are increased (as from inflammatory diarrhea, enterocutaneous fistulas or exudates, or an inflammatory protein-catabolic state); and (4) the reasoned judgment that SNS has a reasonable prospect of improving the patient's clinical outcome or quality of life.

EN THERAPY

EN is indicated when the patient is unable to eat enough food and is unlikely to do so for a long time, their gastrointestinal tract is functional and accessible, and optimized voluntary nutrition is impossible or inappropriate. EN is commonly used for patients with impaired consciousness, severe dysphagia, or severe upper gastrointestinal tract dysfunction or obstruction, or who need mechanical ventilation. Equally commonly, situations arise in which a patient's voluntary food intake is seriously curtailed by anorexia, unappealing food, nausea, vomiting, pain, distress, delirium, depression, chewing difficulties, mild dysphagia, physical and sensory disability (including dysgeusia), or undiagnosed thrush. In these complicated and difficult situations, the clinical diagnosis of SRM or CDM will tip the decision toward EN or PN.

EN is contraindicated by intestinal ischemia, mechanical obstruction, peritonitis, and gastrointestinal hemorrhage. High-dose pressor therapy is a relative contraindication because of the rare but lethal risk of intestinal ischemic injury. Severe coagulopathy, esophageal varices, absent gag reflex, hypotension, paralytic ileus, diarrhea, and nausea and vomiting are not absolute contraindications, but they increase the risk of complications and make it less likely that EN will succeed in achieving its nutritional goal. In general, intensive EN is best avoided under conditions of very grave and unstable illness.

Initiation, Progression, and Monitoring Nasogastric tube feeding may proceed when the patient's gastrointestinal function is adequate with respect to gastric contractility (e.g., nasogastric tube output <1200 mL/d) and intestinal function (no known or suspected intraabdominal pathologic process and a nondistended abdomen—mere absence of bowel sounds is not, by itself, a contraindication). After consent has been obtained and the appropriate feeding tube (usually a nasogastric tube for short-term feeding) has been placed and its position verified, the head of the patient's bed is raised to at least 30° and kept there to reduce the risk of regurgitation. The clinical dietitian ordinarily orders the formula and adjusts its rate. When a standard polymeric formula is infused, it normally commences at 50 mL/h and is advanced by 25 mL/h every 4–8 h until the goal rate is attained. Elemental formulas begin at a slower rate and progress more slowly. Intragastric bolus feeding is an option (200–400 mL feeding solution infused over 15–60 min at regular intervals with verification of residual gastric contents every 4 h).

Complications The common complications of EN are aspiration of regurgitated or vomited formula, diarrhea, fluid volume and electrolyte derangements, hyperglycemia, nausea, abdominal pain, constipation, and failure to achieve the nutritional goal.

Aspiration Patients with delayed gastric emptying, impaired gag reflex, and ineffective cough are at high risk of aspiration pneumonia. Ventilator-associated pneumonia is most often caused by aspiration of microbial pathogens in the mouth and throat past the cuff of an endotracheal or tracheostomy tube. Tracheal suctioning can induce coughing and gastric regurgitation. Ventilator-associated pneumonia can be prevented by elevating the head of the bed, good mouth hygiene and gastrointestinal decontamination, nurse-directed algorithms for formula advancement, and sometimes postpyloric feeding. EN need not be suspended for gastric residual volumes <300–400 mL in the absence of other signs of gastrointestinal intolerance (nausea, vomiting, abdominal pain, abdominal distention). Continuous EN is often tolerated better than bolus feeding and is the only option during jejunal feeding.

Diarrhea Diarrhea commonly occurs when the patient's bowel function is compromised by disease or drugs (most often, broad-spectrum antibiotics). Once infection and inflammatory causes have been ruled out, EN-associated diarrhea may be controlled by using a fiber-containing formula or adding an antidiarrheal agent to it. H_2 blockers or proton pump inhibitors may help reduce the net volume of fluid presented to the colon. Luminal nutrients exert trophic effects on intestinal mucosa, so it is often appropriate to continue tube feeding despite moderate but tolerable diarrhea, even if it necessitates supplemental parenteral fluid support. Except in cases of impaired small-intestinal absorptive function, there are no well-established indications for elemental formulas; nevertheless, they may be tested empirically when diarrhea persists despite the use of fiber-enriched formulas and antidiarrheal drugs.

Gastrointestinal Intolerance Abnormally high gastric residual volumes, abdominal distention, pain, and nausea are greatly distressing for patients, increase nursing workload, and slow the progression of EN. These problems can be avoided or minimized by ensuring that the patient is in normal fluid and electrolyte balance, by preventing severe hyperglycemia, and, when nausea, vomiting, or abdominal distention develop, by the judicious prescription of antiemetic and prokinetic drugs (and sometimes proton pump inhibitors) on a regular—rather than as-needed—basis. Patients with gastroparesis require postpyloric feeding.

Fluid Volume, Electrolyte, and Blood Glucose Abnormalities EN's essential purpose is to provide macronutrients at an appropriate rate. EN also provides standard amounts of fluid, electrolytes, minerals, and micronutrients. Standard EN products are not designed to manage abnormal fluid volume, electrolyte, and mineral requirements, which vary considerably and can change rapidly. Blood glucose concentrations should be monitored regularly, and additional measures—including intravenous fluid, electrolyte, and insulin therapy—may be required to maintain homeostasis.

Failure to Reach the Nutritional Goal EN is frequently delayed or interrupted by intolerance to the administered formula, by diagnostic tests and procedures (including dialysis), physical or occupational therapy, or a clogged or pulled out tube. The result can be a long delay in the progression of EN and ultimate failure to meet the patient's nutrient requirements.

EN in the Intensive Care Unit Most critically ill patients cannot eat anything and hence depend entirely on SNS. EN serves two purposes in this setting. The first is to meet the patient's macronutrient requirements, especially their increased protein requirement. The second purpose is to infuse sufficient nutrients into the intestines to maintain their barrier and immunologic functions in the face of a systemic inflammatory response that threatens them. Current guidelines recommend starting EN soon after a critically ill patient has been fluid resuscitated and stabilized. Once EN is underway, its rate of delivery is gradually increased toward the nutritional goal. EN frequently falls far short of its protein provision target, even after a week or longer in the intensive care unit. Newer, high-protein EN products (and sometimes the addition of powdered protein supplements) can correct this protein shortfall. There is uncertainty about the timing and macronutrient composition of EN early in ADM. We continue to recommend early EN as a platform and preparation for later, more generous nutrient provision and to maintain intestinal function.

PN THERAPY

PN is more resource-intensive and requires more expertise than EN. It is used when invasive SNS is indicated and EN is impossible, inappropriate, or insufficient to meet the patient's nutritional needs. The risks of PN are those of inserting and maintaining a central venous catheter (traumatic injury from the insertion, serious infection, and venous thrombosis); allergy to some of its components; glucose, electrolyte, magnesium, phosphate, and acid-base balance abnormalities; and the adverse effects of the large intravenous fluid volumes. When prolonged for many weeks—and especially when it delivers excess energy—PN may cause or contribute to hepatic dysfunction.

Initiation, Progression, Monitoring, and Discontinuation When indicated, PN should begin soon after the patient has been hemodynamically stabilized; glucose, electrolyte, and acid-base homeostasis has been established; and they are able to tolerate the fluid volume required to deliver it. Newer evidence indicates that the provision of energy equal to total energy expenditure in the first week of PN is not, in general, beneficial and can be harmful under certain conditions. The high osmolarity of adult PN solutions and the need for strict sterility require that they be infused through a dedicated port in a central venous catheter. Jugular or femoral vein catheters should not be used because of the difficulty maintaining a dry, sterile dressing over the insertion site and the higher rate of complications. The initial dose of glucose should not exceed 200 g/d to avoid hyperglycemia (and—in susceptible patients with adapted SRM—the refeeding syndrome).

Most non–critically ill patients do not require >500 g glucose (1700 kcal)/d—assuming in this example a normalized dry body weight of 70 kg. Patients with ADM should, in general, not receive >350 g glucose (1200 kcal)/d during the intense phase of their critical illness. A glucose infusion rate of ~200 g/d is physiologic and rarely needs to be exceeded. When it eventually becomes appropriate to provide energy equal to total energy expenditure, this may be achieved by infusing a lipid emulsion.

We recommend hypocaloric nutrition (limited in glucose, lipid, and fluid volume but generous in amino acids) for the first 2 weeks of SNS in fat-sufficient or obese patients with ADM. Energy provision may increase, if indicated, after the catabolic storm has abated. Lipids are commonly introduced after the first week of PN to make up energy shortfalls. Serum triglyceride concentrations are measured before commencing lipid infusions to detect preexisting hypertriglyceridemia (>400 mg/dL), a relative contraindication. Lipids may be infused daily or two or three times weekly. Lipid infusions are not necessary to prevent essential fatty acid deficiency during hypocaloric nutrition of patients with a normal or excessive fat store, because the endogenous fat mobilized during energy deficiency provides essential fatty acids to the rest of their body.

Capillary blood glucose concentrations are monitored several times daily. Subcutaneous regular insulin is added to the PN admixture as required to maintain average serum glucose concentrations <140 mg/dL and >80 mg/dL. (Upper and lower limits of 180 and 100 mg/dL appear to be appropriate for critically ill patients with diabetes mellitus.) The dose of regular insulin required on a given day can be added to the following day's PN solution. The insulin dose increases roughly proportionately to the glucose dose. Certain benchmarks are useful. Basal endogenous insulin secretion is ~30 units/d in normal people. When insulin is required for nondiabetic, noncatabolic patients, 10 units of regular insulin roughly cover 100 g infused glucose. Patients with non-insulin-dependent diabetes require ~20 units/100 g glucose. Noncatabolic patients with insulin-dependent diabetes usually require approximately twice their at-home insulin dose, because parenteral glucose stimulates insulin release more potently than oral carbohydrate and because some insulin adheres to the infusion bag.

Biochemical Monitoring Serum urea, creatinine, electrolytes, glucose, magnesium, phosphate, calcium, and albumin concentrations are determined prior to commencing PN and most of them (not albumin) monitored for the first few days, then twice weekly or as required. Serum triglycerides and liver function indicators are measured at baseline and after PN is underway to verify tolerance to lipid infusions. An N balance determination may be useful at the outset to evaluate the severity of protein catabolism in patients with CDM or ADM, to identify patients who will benefit from more generous amino acid provision, and during the course of PN to determine whether N balance is improving with therapy.

Discontinuation PN is tapered and discontinued when the patient can be adequately nourished by the enteral route. The PN dose is proportionately reduced as food intake increases. Once a patient can tolerate one-half to two-thirds of their requirement by the enteral route and there is no mechanical or other barrier to further increases, PN should be terminated. The transition to oral nutrition can be difficult for some patients. In this situation, optimized voluntary nutrition, although labor-intensive, is much preferred to replacing PN with EN, for it is safe, effective, fosters well-being, and prepares the patient for discharge home. It is tempting to discontinue PN in hopes of stimulating the patient's appetite, but this impulse should be resisted. PN does not cause anorexia nor does discontinuing it stimulate appetite. Too-early discontinuation of PN may delay a patient's progression to full voluntary food consumption by inducing anxiety and recreating starvation conditions. A patient is most successfully weaned from PN by optimizing their voluntary nutrition (including food brought from home), providing emotional support, encouraging physical activity, and being patient. Some patients on the cusp of adequate oral nutrition will benefit from discharge to the security and pleasure of home life and homemade food; these patients are identified by observing, asking, and listening.

Complications Patients receiving PN are at higher risk of bloodstream infections than other patients with central venous catheters. Rigorously aseptic insertion technique, meticulous dressing care, one port dedicated solely to PN, and careful glycemic control reduce this risk. EN and PN carry equivalent overall risks and complication rates, apart from gastrointestinal side effects of EN. PN is much more prone to the error of inadvertent toxic overfeeding.

Hyperglycemia The most frequent metabolic complication of PN is hyperglycemia in patients with insulin resistance due to non-insulin-dependent diabetes mellitus, high-dose glucocorticoid therapy, or severe systemic inflammation; the problem is exacerbated by excessive rates of glucose provision. Glucose concentrations are most easily kept <140 mg/dL and with the least risk of hypoglycemia by infusing hypocaloric amounts of glucose and, when necessary, meeting the patient's energy requirement with intravenous lipid. In ADM, the benefits of using the lowest possible insulin dose—minimal hyperinsulinemia and a reduced risk of hypoglycemia—almost always outweigh the doubtful goal of rapidly matching energy provision to energy expenditure.

Hypoglycemia Reactive hypoglycemia is uncommon but can occur immediately after high-glucose, non-insulin-containing PN is abruptly discontinued. It is prevented by slowing the PN infusion rate to 50 mL/h for 1 or 2 h prior to stopping, replacing it with 10% glucose, or when the oral route is available, providing a snack. More often, hypoglycemia occurs when the intensity of the patient's metabolic stress (or their glucocorticoid dose) decreases without an appropriate downward adjustment of the insulin dose. This problem is avoided by frequent capillary glucose determinations and careful attention to medication doses and the patient's general condition.

Artefactual Hyperglycemia and Hyperkalemia Blood samples must be meticulously collected from a dual-port central venous catheter. The intermixing of sample with even a tiny volume of PN solution will falsely indicate hyperglycemia and hyperkalemia and can result in treatment error. This error is identified when the patient's apparent serum glucose (and potassium) concentration abruptly increases without reason, and the high glucose concentration is contradicted by a concurrent capillary glucose reading.

Volume Overload Volume overload causes substantial morbidity and increases mortality. Total fluid provision should not normally exceed 2 L/d in the absence of abnormally high losses. Hypertonic intravenous glucose triggers a more intense insulin response than oral glucose, and it can increase urinary sodium and water retention. Volume overload can be prevented by using a compounder to prepare PN solutions, infusing glucose at the lowest appropriate rate, avoiding energy overfeeding, and monitoring the patient's volume balance.

Hypertriglyceridemia This complication occurs when the rate of lipid infusion exceeds plasma triglyceride clearance capacity. Sepsis, renal failure, diabetes mellitus, high-dose glucocorticoid therapy, and multiple-organ failure reduce triglyceride clearance. An impaired immune response, increased risk of acute pancreatitis, and altered pulmonary hemodynamics are potential, but not well documented, complications of PN-induced severe hypertriglyceridemia. Lipid infusion rates should not usually exceed ~50 g (500 kcal)/d in ADM.

Liver Disease Mild elevations of serum liver enzyme concentrations may occur within 2–4 weeks of initiating PN, but in most cases, they will return to normal even when PN is continued. Clinically important hepatic dysfunction, although common in children, is uncommon in adults for whom energy overfeeding with resultant fatty liver is avoided. Intrahepatic cholestasis occasionally develops after many weeks of continuous PN; it is most often multifactorial in origin. Cyclic PN—in which PN is infused for only 12 h of the day—may prevent or reduce the severity of this complication.

PN in the Intensive Care Unit Current guidelines recommend starting EN soon after a critically ill patient has been resuscitated and stabilized and enteral access to an adequately functioning gastrointestinal tract has been established. If the protein goal has not been achieved after 7–10 days, PN may be added, especially if the protein-catabolic state has not yet abated. Soy-based lipid emulsions should be avoided during the first week of PN; alternative lipid emulsions may prove to be safe and beneficial.

SPECIAL CLINICAL SITUATIONS

The Critical Illness–Nutrition Paradox There is now good clinical trial evidence that confirms what was long indicated by a combination of physiologic reasoning, biologic plausibility, and formal clinical observation—namely, that personalized nutritional interventions improve the clinical outcomes of starving, non–critically ill patients. The case for SNS would seem to be much stronger for patients with ADM, which induces rapid muscle loss and maintains or increases energy expenditure in people rendered incapable of eating voluntarily. Yet well-designed large-enrollment clinical trials have repeatedly failed to demonstrate short-term clinical benefits from nutritional interventions in critical illness. In fact, there is good evidence that, unlike in noncritical illness, early energy provision at a rate near or above the rate of energy expenditure not only fails to improve clinical outcomes but may be deleterious. The failure of formal clinical trials to show benefits from SNS in critical illness has several possible explanations: severe prolonged starvation is so obviously harmful that it is unethical to intentionally include such a treatment arm in a randomized clinical trial; critical illness is extremely heterogeneous, and not every critically ill patient is, or remains, severely protein-catabolic for long; and owing to generous admission criteria and thanks to the high quality of modern intensive care, many patients admitted to intensive care units rapidly improve and are discharged in a handful of days, while others are so mortally ill when admitted to the intensive care unit that their outcome is virtually predetermined. For these reasons, clinical trials that enroll and report the outcomes of all patients admitted to an intensive care unit could well fail to demonstrate an average benefit from SNS. More than that, in current practice, standard EN regimens prescribed for most critically ill patients commonly fail to deliver more than one-half the currently recommended amount of protein, and the low protein-to-energy ratio of most standard EN and PN products can make it impossible to provide critically ill patients sufficiently generous amounts of protein or amino acids without subjecting them to harmful energy overfeeding. We await the publication of large-enrollment clinical trials in which critical illness is diagnosed, not by where the patient is being treated, but using anatomic and pathophysiologic criteria and in which current standard care is compared with treatment arms that provide personalized SNS appropriate to patients' anatomy and pathophysiology. For now, along with other experts, we continue to recommend EN and PN for critically ill patients. Attempts to match energy provision to energy expenditure should be avoided as long as systemic inflammation remains severe. The dose of protein or amino acids should be guided by physiologic reasoning and a personalized evaluation of each patient's anatomic and metabolic condition. We recommend bearing in mind that people who survive critical illness become non–critically ill people for whom the benefit of appropriate nutrition, together with other components of medical and surgical care, is not in question. Accordingly, we recommend a long-view approach to the design of nutritional support regimens for critically ill patients that aims both to increase their likelihood of short-term survival and anticipates their post–critical illness phase, which will be improved by measures that mitigate muscle loss, avoid toxic energy overfeeding, and prevent or correct micronutrient deficiencies.

Iron and PN Iron deficiency is common in hospitalized patients. It has many causes: preexisting deficiency, insufficient food, macro- or microscopic gastrointestinal blood loss, and repeated blood sampling. Even though it is common, the diagnosis is often missed because the anemia of systemic inflammation is even more common, and it confounds interpretation of the usual indicator of iron deficiency, serum ferritin. Serum ferritin and fasting serum transferrin saturation should be part of the patient's initial nutritional evaluation. Current evidence suggests that a fasting transferrin saturation <20% is the most reliable indicator of functional iron deficiency, even in the absence of anemia. In principle, all micronutrient deficiency states should be prevented and corrected. In-hospital iron deficiency causes and prevents recovery from anemia and increases postoperative blood transfusion requirements. Subclinical iron deficiency could contribute to cognitive and immune dysfunction.

Iron is not added to PN mixtures. Iron dextran is incompatible with lipid emulsions, and although it appears to be chemically compatible with aqueous solutions of amino acids and glucose, there is realistic concern that interactions between iron molecules and certain vitamins and amino acids in PN solutions could catalyze the formation of free radicals that degrade vitamins and exert subtle adverse systemic effects.

Intravenous iron is usually necessary to treat in-hospital iron deficiency, but it should be delayed if there is known or suspected bacterial infection, and possibly not during the intense phase of ADM. A substantial rise in the serum iron concentration could increase susceptibility to gram-negative (and possibly other microbial) infections, as well as catalyze the formation of free radicals that increase the intensity of the catabolic response to major tissue injury.

Zinc One liter of secretory diarrhea contains ~12 mg of zinc. Patients with intestinal fistulas or high-volume chronic diarrhea require this amount of zinc in addition to their daily requirement of 15 mg to avoid zinc deficiency. Zinc may be provided parenterally or enterally. Because of its low bioavailability, 12 mg of parenteral zinc is equivalent to 30 mg of oral zinc.

Old Age In addition to their other frailties, elderly people commonly suffer from age-related muscle loss (sarcopenia) compounded by disuse muscle atrophy. These factors place them at high risk of the consequences of starvation disease and make them candidates for earlier SNS.

Inactivity Physical activity and adequate nutrition interact intimately. Reduced physical activity reduces appetite, and physical rehabilitation and its associated emotional benefits restore optimism and appetite. Full nutrient provision will maintain or normalize many physiologic functions in bedridden patients, but without exercise, they will not increase their muscle mass.

Renal Failure Protein provision should not be reduced in patients with renal failure unless renal replacement therapy is unavailable. Renal replacement therapy removes large amounts of amino acids, vitamins, and trace elements from the circulation, so protein and micronutrient provision should be increased to compensate for these losses.

Liver Failure Patients with severe hepatic disease are relatively intolerant to starvation and commonly have CDM when admitted to hospital, so they are prime candidates for SNS. Their SNS should be generous both in energy and protein, despite an increased risk of hepatic encephalopathy. The risk of encephalopathy can be reduced by meticulous attention to fluid balance, acid-base balance, and electrolyte status and by spreading protein provision over the day to accommodate the liver's reduced capacity to clear amino acid–derived ammonia.

Perioperative SNS Patients with SRM or CDM awaiting elective major surgery will benefit from 7–10 days of preoperative SNS. Optimized voluntary nutrition is always preferred, but when a patient has been admitted to hospital in a semi-urgent condition, EN or PN is usually required to quickly meet the patient's nutritional goal. Preoperative SNS improves immunity and reduces postoperative complications, but it will not increase serum albumin concentrations, and it should not be provided for >7–10 days with the expectation that it will do so. More prolonged preoperative EN or PN may confer slight additional nutritional benefits, but they are counterbalanced by their risks and the consequences of prolonged hospitalization and delayed surgery. Surgery should not be delayed for starving patients whose muscle mass is normal or only mildly depleted and who are not experiencing systemic inflammation, for they are well able to withstand the trauma of even major uncomplicated surgery. The need for urgent surgery often precludes otherwise indicated preoperative SNS. Early postoperative PN is usually indicated for these patients, for they are at increased risk of postoperative complications and are unlikely to consume much food over the next many days. Patients with only mild muscle loss, no systemic inflammation, and no postoperative complications do not require postoperative PN unless (1) adequate feeding by mouth has not been achieved by day 5–7 after surgery or (2) there are indications that voluntary feeding will be further delayed. There is evidence that perioperative immune-enhancing EN reduces morbidity in patients undergoing major elective gastrointestinal surgery.

Cancer SNS plays a crucial role in cancer therapy. Many malignant neoplasms (especially those that involve the gastrointestinal tract or induce systemic inflammation) and their cytotoxic therapies create the conditions for starvation and commonly lead to SRM or CDM. The prevention or treatment of these starvation diseases will improve patients' quality of life and their tolerance to anticancer therapy. EN and PN are generally not prescribed to patients with advanced cancer for which effective anticancer therapy is unavailable, because the side effects and complications of instrumental SNS are not counterbalanced by an improved disease trajectory. In some cases, the disease progresses so slowly that the patient will die of starvation long before they succumb to their primary disease. EN or PN is obviously appropriate in such cases.

Advanced Dementia Optimized voluntary nutrition is the key approach in this situation, and it can be used to deal with problems such as disability and dysphagia in patients who get pleasure from eating. There is no evidence that EN or PN improves quality or length of life in patients with advanced dementia who show little or no interest in food, and the side effects and complications are unpleasant and sometimes dangerous.

REFEEDING SYNDROME

The refeeding syndrome can occur in patients with successfully adapted SRM during the first week of nutritional repletion when carbohydrate and sodium are introduced too rapidly. Carbohydrate stimulates insulin secretion, which, owing to its antinatriuretic effect, expands the ECF volume, especially when excessive sodium is provided. Refeeding edema can be minimized by severely limiting sodium provision and increasing carbohydrate provision slowly. Carbohydrate refeeding may stimulate intracellular glucose-6-phosphate and glycogen synthesis sufficiently to seriously lower serum phosphate concentrations, impairing skeletal and cardiac muscle function, and can trigger hemolysis. Abrupt carbohydrate provision increases the downregulated metabolic rate of patients with adapted SRM and stimulates N retention, new cell synthesis, and cellular rehydration. Because phosphorus, potassium, and magnesium deficiencies occur and are dangerous during refeeding, their serum concentrations should be measured frequently and appropriate supplements provided. Left heart failure may occur in predisposed patients; it has three causes: (1) an abrupt increase of intravascular volume due to the administration of fluids and of glucose, which stimulates insulin-mediated renal sodium retention; (2) increased cardiac demand on an atrophic left ventricle created by an insulin-mediated increase of resting energy expenditure; and (3) myocardial deficiencies of potassium, phosphorus, or magnesium. Cardiac arrhythmias may occur. Acute thiamine deficiency encephalopathy is a devastating preventable complication of refeeding, even with simple glucose infusions.

FURTHER READING

Berger MM et al: ESPEN micronutrient guideline. Clin Nutr 41:1357, 2022.

Berger MM, Pichard C: When is parenteral nutrition indicated? J Intensive Med 2:22, 2022.

Compher C et al: Guidelines for the provision of nutrition support therapy in the adult critically ill patient: The American Society for Parenteral and Enteral Nutrition. JPEN J Parenter Enteral Nutr 46:12, 2022.

Jensen GL et al: GLIM criteria for the diagnosis of malnutrition: A consensus report from the global clinical nutrition community. JPEN J Parenter Enteral Nutr 43:32, 2019.

Singer P et al: ESPEN practical and partially revised guideline: Clinical nutrition in the intensive care unit. Clin Nutr 42:1671, 2023.

Wischmeyer PE et al: Personalized nutrition therapy in critical care: 10 expert recommendations. Crit Care 27:261, 2023.

Section 3 Liver and Biliary Tract Disease

347

Approach to the Patient with Liver Disease

Marc G. Ghany, Jay H. Hoofnagle

A diagnosis of liver disease usually can be made accurately by careful elicitation of the patient's history, physical examination, and application of a few laboratory tests. In some circumstances, radiologic examinations are helpful or, indeed, diagnostic. Liver biopsy is considered the criterion standard in evaluation of liver disease but is now needed less for diagnosis than for grading (activity) and staging (fibrosis) of disease. Noninvasive means of assessing fibrosis stage have become increasingly helpful and may allow for avoidance of biopsy in an increasing proportion of patients. This chapter provides an introduction to diagnosis and management of liver disease, briefly reviewing the structure and function of the liver; the major clinical manifestations of liver disease; and the use of clinical history, physical examination, laboratory tests, imaging studies, and liver biopsy.

LIVER STRUCTURE AND FUNCTION

The liver is the largest organ of the body, weighing 1–1.5 kg and representing 1.5–2.5% of the lean body mass. The size and shape of the liver vary and generally match the general body shape—long and lean or squat and square. This organ is located in the right upper quadrant of the abdomen under the right lower rib cage against the diaphragm and projects for a variable extent into the left upper quadrant. It is held in place by ligamentous attachments to the diaphragm, peritoneum, great vessels, and upper gastrointestinal organs. The liver receives a dual blood supply; ~20% of the blood flow is oxygen-rich blood from the hepatic artery, and 80% is nutrient-rich blood from the portal vein arising from the stomach, intestines, pancreas, and spleen.

The majority of cells in the liver are hepatocytes, which constitute two-thirds of the organ's mass. The remaining cell types are Kupffer cells (members of the reticuloendothelial system), stellate (Ito or fat-storing) cells, endothelial and blood vessel cells, bile ductular cells, and cells of supporting structures. Viewed by light microscopy, the liver appears to be organized in lobules, with portal areas at the periphery and central veins in the center of each lobule. However, from a functional point of view, the liver is organized into acini, with both hepatic arterial and portal venous blood entering the acinus from the portal areas (zone 1) and then flowing through the sinusoids to the terminal hepatic veins (zone 3); the intervening hepatocytes constitute zone 2. The advantage of viewing the acinus as the physiologic unit of the liver is that this perspective helps to explain the morphologic patterns and zonality of many vascular and biliary diseases not explained by the lobular arrangement.

Portal areas of the liver consist of small veins, arteries, bile ducts, and lymphatics organized in a loose stroma of supporting matrix and small amounts of collagen. Blood flowing into the portal areas is distributed through the sinusoids, passing from zone 1 to zone 3 of the acinus and draining into the terminal hepatic veins ("central veins"). Secreted bile flows in the opposite direction—that is, in a countercurrent pattern from zone 3 to zone 1. The sinusoids are lined by unique endothelial cells that have prominent fenestrae of variable sizes, allowing the free flow of plasma but not of cellular elements. The plasma is thus in direct contact with hepatocytes in the subendothelial space of Disse.

Hepatocytes have distinct polarity. The basolateral side of the hepatocyte lines the space of Disse and is richly lined with microvilli; it exhibits endocytotic and pinocytotic activity, with passive and active uptake of nutrients, proteins, and other molecules. The apical pole of the hepatocyte forms the canalicular membranes through which bile components are secreted. The canaliculi of hepatocytes form a fine network, which fuses into the bile ductular elements near the portal areas. Kupffer cells usually lie within the sinusoidal vascular space and represent the largest group of fixed macrophages in the body. The stellate cells are located in the space of Disse but are not usually prominent unless activated, whereupon they produce collagen and matrix. Red blood cells stay in the sinusoidal space as blood flows through the lobules, but white blood cells can migrate through or around endothelial cells into the space of Disse and from there to portal areas, where they can return to the circulation through lymphatics.

Hepatocytes perform numerous and vital roles in maintaining homeostasis and health. These functions include the synthesis of most essential serum proteins (albumin, carrier proteins, coagulation factors, many hormonal and growth factors), the production of bile and its carriers (bile acids, cholesterol, lecithin, phospholipids), the regulation of nutrients (glucose, glycogen, lipids, cholesterol, amino acids), and the metabolism and conjugation of lipophilic compounds (bilirubin, anions, cations, drugs) for excretion in the bile or urine. Measurement of these activities to assess liver function is complicated by the multiplicity and variability of these functions. Commonly used liver "function" tests are measurements of serum bilirubin, serum albumin, and prothrombin time. The serum bilirubin level is a measure of hepatic conjugation and excretion; the serum albumin level and prothrombin time are measures of protein synthesis. Abnormalities of bilirubin, albumin, and prothrombin time are typical of hepatic dysfunction. Frank liver failure is incompatible with life, and the functions of the liver are too complex and diverse to be subserved by a mechanical pump; a dialysis membrane; or a concoction of infused hormones, proteins, and growth factors.

LIVER DISEASES

While there are many causes of liver disease **(Table 347-1)**, these disorders generally present clinically in a few distinct patterns and are usually classified as hepatocellular, cholestatic (obstructive), or mixed. In hepatocellular diseases (such as viral hepatitis and alcoholic liver disease), features of liver injury, inflammation, and necrosis predominate. In cholestatic diseases, such as gallstone or malignant obstruction, primary biliary cholangitis (previously referred to as primary biliary cirrhosis), and some drug-induced liver diseases, features of inhibition of bile flow predominate. In a mixed pattern, features of both hepatocellular and cholestatic injury are present (such as in cholestatic forms of viral hepatitis and many drug-induced liver diseases). The pattern of onset and prominence of symptoms can rapidly suggest a diagnosis, particularly if major risk factors are considered, such as the age and sex of the patient and a history of exposure or risk behaviors.

Typical presenting symptoms of liver disease include jaundice, fatigue, itching, right-upper-quadrant pain, nausea, poor appetite, abdominal distention, and intestinal bleeding. At present, however, many patients are diagnosed with liver disease who have no symptoms and who have been found to have abnormalities in biochemical liver tests as a part of a routine physical examination or screening for blood donation or for insurance or employment. The wide availability of batteries of liver tests makes it relatively simple to demonstrate the presence of liver injury as well as to rule it out in someone in whom liver disease is suspected.

Evaluation of patients with liver disease should be directed at determining (1) the etiologic diagnosis, (2) disease severity (grading), and (3) disease stage (staging). Diagnosis should focus on the pattern of

TABLE 347-1 Liver Diseases

- **Inherited hyperbilirubinemia**
  - Gilbert syndrome
  - Crigler-Najjar syndrome, types I and II
  - Dubin-Johnson syndrome
  - Rotor syndrome
- **Viral hepatitis**
  - Hepatitis A
  - Hepatitis B
  - Hepatitis C
  - Hepatitis D
  - Hepatitis E
  - Others (Epstein-Barr virus [mononucleosis] herpesvirus, cytomegalovirus, adenovirus hepatitis)
  - Cryptogenic hepatitis
- **Immune and autoimmune liver diseases**
  - Primary biliary cholangitis
  - Autoimmune hepatitis
  - Sclerosing cholangitis
  - Overlap syndromes
  - Graft-versus-host disease
  - Allograft rejection
- **Genetic liver diseases**
  - α_1 Antitrypsin deficiency
  - Hemochromatosis
  - Wilson disease
  - Benign recurrent intrahepatic cholestasis
  - Progressive familial intrahepatic cholestasis, types I–III
  - Others (galactosemia, tyrosinemia, cystic fibrosis, Niemann-Pick-disease, Gaucher's disease)
- **Alcohol-related liver disease**
  - Acute fatty liver
  - Acute alcoholic hepatitis
  - Laënnec cirrhosis
- Nonalcoholic fatty liver[a]
  - Steatosis
  - Steatohepatitis
- **Acute fatty liver of pregnancy**
- **Liver involvement in systemic diseases**
  - Sarcoidosis
  - Amyloidosis
  - Glycogen storage diseases
  - Celiac disease
  - Tuberculosis
  - *Mycobacterium avium-intracellulare* infection
- **Cholestatic syndromes**
  - Benign postoperative cholestasis
  - Jaundice of sepsis
  - Total parenteral nutrition–induced jaundice
  - Cholestasis of pregnancy
  - Cholangitis and cholecystitis
  - Extrahepatic biliary obstruction (stone, stricture, cancer)
  - Biliary atresia
  - Caroli disease
  - Cryptosporidiosis
- **Drug-induced liver disease**
  - Hepatocellular patterns (isoniazid, acetaminophen)
  - Cholestatic patterns (methyltestosterone)
  - Mixed patterns (sulfonamides, phenytoin)
  - Micro- and macrovesicular steatosis (methotrexate, fialuridine)
- **Vascular injury**
  - Sinusoidal obstruction syndrome
  - Budd-Chiari syndrome
  - Ischemic hepatitis
  - Passive congestion
  - Portal vein thrombosis
  - Nodular regenerative hyperplasia
- **Mass lesions**
  - Hepatocellular carcinoma
  - Cholangiocarcinoma
  - Adenoma
  - Focal nodular hyperplasia
  - Metastatic tumors
  - Abscess
  - Cysts
  - Hemangioma

[a] Also known as metabolic dysfunction–associated steatotic liver disease (MASLD).

disease presentation (hepatocellular, cholestatic, or mixed injury) as well as on the specific etiologic diagnosis. Grading refers to assessment of the severity or activity of disease—active or inactive as well as mild, moderate, or severe. Staging refers to estimation of the point in the course of the natural history of the disease, whether early or late; or precirrhotic, cirrhotic, or end-stage. This chapter introduces general, salient concepts in the evaluation of patients with liver disease that help lead to the diagnoses discussed in subsequent chapters.

CLINICAL HISTORY

The clinical history should focus on the symptoms of liver disease—their nature, patterns of onset, and progression—and on potential risk factors for liver disease. The manifestations of liver disease include constitutional symptoms such as fatigue, weakness, nausea, poor appetite, and malaise and the more liver-specific symptoms of jaundice, dark urine, light stools, itching, abdominal pain, and bloating. Symptoms can also suggest the presence of cirrhosis, end-stage liver disease, or complications of cirrhosis such as portal hypertension. Generally, the constellation of symptoms and their patterns of onset, rather than a specific symptom, point to an etiology.

Fatigue is the most common and most characteristic symptom of liver disease. It is variously described as lethargy, weakness, listlessness, malaise, increased need for sleep, lack of stamina, and poor energy. The fatigue of liver disease typically arises after activity or exercise and is rarely present or severe after adequate rest; that is, it is "afternoon" rather than "morning" fatigue. Fatigue in liver disease is often intermittent and variable in severity from hour to hour and day to day. In some patients, it may not be clear whether fatigue is due to the liver disease or due to other problems such as stress, anxiety, sleep disturbance, or a concurrent illness.

Nausea occurs with more severe liver disease and may accompany fatigue or be provoked by smelling food odors or eating fatty foods. Vomiting can occur but is rarely persistent or prominent. Poor appetite with weight loss occurs frequently in acute liver disease but is rare in chronic disease except when cirrhosis is present and advanced. Diarrhea is uncommon in liver disease except with severe jaundice, in which a lack of bile acids reaching the intestine can lead to steatorrhea.

Right-upper-quadrant discomfort or ache ("liver pain") occurs in many liver diseases and is usually marked by tenderness over the liver area. The pain arises from stretching or irritation of Glisson's capsule, which surrounds the liver and is rich in nerve endings. Severe pain is most typical of gallbladder disease, liver abscess, and severe sinusoidal obstruction syndrome (previously known as venoocclusive disease) but is also an occasional accompaniment of acute hepatitis.

Itching occurs with acute liver disease, appearing early in obstructive jaundice (from biliary obstruction) or drug-induced cholestasis and somewhat later in hepatocellular disease (acute hepatitis). Itching also occurs in chronic liver diseases—typically the cholestatic forms such as primary biliary cholangitis and sclerosing cholangitis, in which it is often the presenting symptom, preceding the onset of jaundice. However, itching can occur in any liver disease, particularly once cirrhosis develops.

Jaundice is the hallmark symptom of liver disease and perhaps the most reliable marker of severity. Patients usually report darkening of the urine before they notice scleral icterus. Jaundice is rarely detectable with a bilirubin level <43 μmol/L (2.5 mg/dL). With severe cholestasis, there will also be lightening of the color of the stools and steatorrhea. Jaundice without dark urine usually indicates indirect (unconjugated) hyperbilirubinemia and is typical of hemolytic anemia and the genetic disorders of bilirubin conjugation, the common and benign form being Gilbert syndrome and the rare and severe form being Crigler-Najjar syndrome. Gilbert syndrome affects up to 5% of the general population; the jaundice in this condition is more noticeable after fasting and with stress.

Major risk factors for liver disease that should be sought in the clinical history include details of alcohol use, medication use (including herbal compounds, birth control pills, and over-the-counter medications), personal habits, sexual activity, travel, exposure to jaundiced or other high-risk persons, injection drug use, recent surgery, remote or recent transfusion of blood or blood products, occupation, accidental exposure to blood or needlestick, and familial history of liver disease.

For assessing the risk of viral hepatitis, a careful history of sexual activity is of particular importance and should include the number of lifetime sexual partners and, for men, a history of having sex with men. Sexual exposure is a common mode of spread of hepatitis B and D but is uncommon for hepatitis C. A family history of hepatitis, liver disease, and liver cancer is also important. Maternal-infant transmission occurs with both hepatitis B and C. Vertical spread of hepatitis B can now be prevented by passive and active immunization of the infant at birth. Additionally, antiviral therapy administered in the second or third trimester of pregnancy is now recommended for mothers with levels of hepatitis B virus DNA >200,000 IU/mL. Vertical spread of hepatitis C is uncommon, but there are no reliable means of prevention. Transmission is more common among HIV-co-infected mothers and is also linked to prolonged and difficult labor and delivery, early rupture of membranes, internal fetal monitoring, and a high maternal viral load. A history of injection drug use, even in the remote past, is of great importance in assessing the risk for hepatitis B, C, and D. Injection drug use is the single most common risk factor for hepatitis C. Transfusion with blood or blood products is no longer an important risk factor for acute viral hepatitis due to screening of blood products. However, blood transfusions received before the introduction of sensitive enzyme immunoassays for antibody to hepatitis C virus in 1992 is an important risk factor for chronic hepatitis C. Blood transfusion before 1986, when screening for antibody to hepatitis B core antigen was introduced, is also a risk factor for hepatitis B. Travel to a developing area of the world, exposure to persons with jaundice, and exposure to young children in day-care centers are risk factors for hepatitis A. Tattooing and body piercing (for hepatitis B, C, and D) and eating shellfish (for hepatitis A) are frequently mentioned but are actually types of exposure that rarely lead to infection.

Hepatitis E is one of the more common causes of jaundice in Asia and Africa but is uncommon in developed nations. In endemic areas, transmission is usually through exposure to fecally contaminated water. Non-travel-related (autochthonous) cases of hepatitis E have been described in developed countries, including the United States. These cases appear to be due to strains of hepatitis E virus that are endemic in swine and some wild animals (genotypes 3 and 4). While occasional cases are associated with eating raw or undercooked pork or game (deer and wild boars), most cases of hepatitis E occur without known exposure, predominantly in elderly men without typical risk factors for viral hepatitis. Hepatitis E infection can become chronic in immunosuppressed individuals (such as transplant recipients, patients receiving chemotherapy, or patients with HIV infection), in whom it presents with abnormal serum enzymes in the absence of markers of hepatitis B or C.

A history of alcohol intake is important in assessing the cause of liver disease and in planning management and recommendations. In the United States, for example, at least 70% of adults drink alcohol to some degree, but significant alcohol intake is less common; in population-based surveys, only 5% of individuals have more than two drinks per day, the average drink representing 11–15 g of alcohol. Alcohol consumption associated with an increased rate of alcoholic liver disease is probably more than two drinks (22–30 g) per day in women and three drinks (33–45 g) in men. Most patients with alcoholic cirrhosis have a much higher daily intake and have drunk excessively for ≥10 years before onset of liver disease. In assessing alcohol intake, the history should also focus on whether alcohol abuse or dependence is present. Alcoholism is usually defined by the behavioral patterns and consequences of alcohol intake, not by the amount. Abuse is defined by a repetitive pattern of drinking alcohol that has adverse effects on social, family, occupational, or health status. Dependence is defined by alcohol-seeking behavior, despite its adverse effects. Many alcoholics demonstrate both dependence and abuse, and dependence is considered the more serious and advanced form of alcoholism. A clinically helpful approach to diagnosis of alcohol dependence and abuse is the use of the CAGE questionnaire **(Table 347-2)**, which is recommended for all medical history-taking.

Family history can be helpful in assessing liver disease. Familial causes of liver disease include Wilson disease; hemochromatosis and α_1 antitrypsin deficiency; and the less common inherited pediatric liver diseases—that is, familial intrahepatic cholestasis, benign recurrent intrahepatic cholestasis, and Alagille syndrome. Onset of severe liver disease in childhood or adolescence in conjunction with a family history of liver disease or neuropsychiatric disturbance should lead to investigation for Wilson disease. A family history of cirrhosis, diabetes, or endocrine failure and the appearance of liver disease in adulthood suggest hemochromatosis and should prompt investigation of iron status. Abnormal iron studies in adult patients warrant genotyping of the *HFE* gene for the C282Y and H63D mutations typical of genetic hemochromatosis. In children and adolescents with iron overload, other non-*HFE* causes of hemochromatosis should be sought. A family history of emphysema should lead to investigation of α_1 antitrypsin levels and, if low, for protease inhibitor (Pi) genotype.

TABLE 347-2 CAGE Questions[a]

ACRONYM	QUESTION
C	Have you ever felt you ought to *c*ut down on your drinking?
A	Have people *a*nnoyed you by criticizing your drinking?
G	Have you ever felt *g*uilty or bad about your drinking?
E	Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover (*e*ye-opener)?

[a] One "yes" response should raise suspicion of an alcohol use problem, and more than one is a strong indication of abuse or dependence.

PHYSICAL EXAMINATION

The physical examination rarely uncovers evidence of liver dysfunction in a patient without symptoms or abnormal laboratory findings, nor are most signs of liver disease specific to one diagnosis. Thus, the physical examination complements rather than replaces the need for other diagnostic approaches. In many patients, the physical examination is normal unless the disease is acute or severe and advanced. Nevertheless, the physical examination is important in that it can yield the first evidence of hepatic failure, portal hypertension, and liver decompensation. In addition, the physical examination can reveal signs—related either to risk factors or to associated diseases or findings—that point to a specific diagnosis.

Typical physical findings in liver disease are icterus, hepatomegaly, hepatic tenderness, splenomegaly, spider angiomata, palmar erythema, and skin excoriations. Signs of advanced disease include muscle wasting, ascites, edema, dilated abdominal veins, hepatic fetor, asterixis, mental confusion, stupor, and coma. In male patients with cirrhosis, particularly that related to alcohol use, signs of hyperestrogenemia such as gynecomastia, testicular atrophy, and loss of male-pattern hair distribution may be found.

Icterus is best appreciated when the sclera is inspected under natural light. In fair-skinned individuals, a yellow tinge to the skin may be obvious. In dark-skinned individuals, examination of the mucous membranes below the tongue can demonstrate jaundice. Jaundice is rarely detectable if the serum bilirubin level is <43 μmol/L (2.5 mg/dL) but may remain detectable below this level during recovery from jaundice (because of protein and tissue binding of conjugated bilirubin).

Spider angiomata and palmar erythema occur in both acute and chronic liver disease; these manifestations may be especially prominent in persons with cirrhosis but can develop in normal individuals and are frequently found during pregnancy. Spider angiomata are superficial, tortuous arterioles, and—unlike simple telangiectasias—typically fill from the center outward. Spider angiomata occur only on the arms, face, and upper torso; they can be pulsatile and may be difficult to detect in dark-skinned individuals.

Hepatomegaly is not a very reliable sign of liver disease because of variability in the liver's size and shape and the physical impediments to assessment of liver size by percussion and palpation. Marked hepatomegaly is typical of cirrhosis, sinusoidal obstruction syndrome, infiltrative disorders such as amyloidosis, metastatic or primary cancers of the liver, and alcoholic hepatitis. Careful assessment of the liver edge may also reveal unusual firmness, irregularity of the surface, or frank nodules. Perhaps the most reliable physical finding in the liver examination is hepatic tenderness. Discomfort when the liver is touched or pressed upon should be carefully sought with percussive comparison of the right and left upper quadrants.

Splenomegaly, which occurs in many medical conditions, can be a subtle but significant physical finding in chronic liver disease and suggests underlying cirrhosis. The availability of ultrasound (US) methods for assessment of the spleen allows confirmation of the physical finding.

Signs of advanced liver disease include muscle wasting and weight loss as well as hepatomegaly, bruising, ascites, and edema. Ascites is best appreciated by attempts to detect shifting dullness by careful percussion. US examination will confirm the finding of ascites in equivocal cases. Peripheral edema can occur with or without ascites. In patients with advanced liver disease, other factors frequently contribute to edema formation, including hypoalbuminemia, venous insufficiency, heart failure, and medications.

Hepatic failure is defined as the occurrence of signs or symptoms of hepatic encephalopathy in a person with severe acute or chronic liver disease. The first signs of hepatic encephalopathy can be subtle and nonspecific—change in sleep patterns, change in personality, irritability, and mental dullness. Thereafter, confusion, disorientation, stupor, and eventually coma supervene. In acute liver failure, excitability and mania may be present. Physical findings include asterixis and flapping tremors of the body and tongue. Fetor hepaticus refers to the slightly sweet, ammoniacal odor that can develop in patients with liver failure, particularly if there is portal-venous shunting of blood around the liver. Other causes of coma and disorientation should be excluded, mainly electrolyte imbalances, sedative use, and renal or respiratory failure. The appearance of hepatic encephalopathy during acute hepatitis is the major criterion for diagnosis of acute liver failure and indicates a poor prognosis. In chronic liver disease, encephalopathy is usually triggered by a medical complication such as gastrointestinal bleeding, overdiuresis, uremia, dehydration, electrolyte imbalance, infection, constipation, or use of narcotic analgesics.

A helpful measure of hepatic encephalopathy is a careful mental status examination and use of the trail-making test, which consists of a series of 25 numbered circles that the patient is asked to connect as rapidly as possible using a pencil. The normal range for the connect-the-dot test is 15–30 s; it is considerably longer in patients with early hepatic encephalopathy. Other tests include drawing of abstract objects or comparison of a signature to previous examples. More sophisticated testing—for example, with electroencephalography and visual evoked potentials—can detect mild forms of encephalopathy but are rarely clinically useful.

Other signs of advanced liver disease include umbilical hernia from ascites, hydrothorax, prominent veins over the abdomen, and caput medusa, a condition that consists of collateral veins radiating from the umbilicus and results from recanulation of the umbilical vein. Widened pulse pressure and signs of a hyperdynamic circulation can occur in patients with cirrhosis as a result of fluid and sodium retention, increased cardiac output, and reduced peripheral resistance. Patients with long-standing cirrhosis and portal hypertension are prone to develop the hepatopulmonary syndrome, which is defined by the triad of liver disease, hypoxemia, and pulmonary arteriovenous shunting. The hepatopulmonary syndrome is characterized by platypnea and orthodeoxia: shortness of breath and oxygen desaturation that occur paradoxically upon the assumption of an upright position. Measurement of oxygen saturation by pulse oximetry is a reliable screening test for hepatopulmonary syndrome.

Several skin disorders and changes are common in liver disease. Hyperpigmentation is typical of advanced chronic cholestatic diseases such as primary biliary cholangitis and sclerosing cholangitis. In these same conditions, xanthelasma and tendon xanthomata occur as a result of retention and high serum levels of lipids and cholesterol. Slate-gray pigmentation of the skin is also seen with hemochromatosis if iron levels are high for a prolonged period. Mucocutaneous vasculitis with palpable purpura, especially on the lower extremities, is typical of cryoglobulinemia of chronic hepatitis C but can also occur in chronic hepatitis B.

Some physical signs point to specific liver diseases. Kayser-Fleischer rings occur in Wilson disease and consist of a golden-brown copper pigment deposited in Descemet's membrane at the periphery of the cornea; they are best seen by slit-lamp examination. Dupuytren contracture and parotid enlargement are suggestive of alcohol use disorder and alcohol-related liver disease. In metastatic liver disease or primary hepatocellular carcinoma (HCC), signs of cachexia and wasting as well as firm hepatomegaly and a hepatic bruit may be prominent.

■ DIAGNOSIS OF LIVER DISEASE

The key diagnostic tests of major causes of acute and chronic liver disease are outlined in **Table 347-3**, and an algorithm for evaluation of the patient with suspected liver disease is shown in **Fig. 347-1**. Specifics of diagnosis are discussed in later chapters. The most common causes of acute liver disease are viral hepatitis (particularly hepatitis A, B, and C), drug-induced liver injury, cholangitis, and alcohol-related liver disease. Liver biopsy usually is not needed for the diagnosis and management of acute liver disease, exceptions being situations where the diagnosis remains unclear despite thorough clinical and laboratory investigation. Liver biopsy can be helpful in diagnosing drug-induced liver disease and acute alcoholic hepatitis.

TABLE 347-3 Important Diagnostic Tests in Common Liver Diseases

DISEASE	DIAGNOSTIC TEST
Hepatitis A	Anti-HAV IgM
Hepatitis B	
Acute	HBsAg and anti-HBc IgM
Chronic	HBsAg and HBeAg and/or HBV DNA
Hepatitis C	Anti-HCV and HCV RNA
Hepatitis D (delta)	HBsAg and anti-HDV
Hepatitis E	Anti-HEV IgM and HEV RNA
Autoimmune hepatitis	ANA or SMA, elevated IgG levels, and compatible histology
Primary biliary cholangitis	Mitochondrial antibody, elevated IgM levels, and compatible histology
Primary sclerosing cholangitis	P-ANCA, cholangiography
Drug-induced liver disease	History of drug ingestion
Alcohol-related liver disease	History of excessive alcohol intake and compatible histology
Nonalcoholic steatohepatitis[a]	Ultrasound or CT evidence of fatty liver and compatible histology
α_1 Antitrypsin disease	Reduced α_1 antitrypsin levels, phenotype PiZZ or PiSZ
Wilson disease	Decreased serum ceruloplasmin and increased urinary copper; increased hepatic copper level
Hemochromatosis	Elevated iron saturation and serum ferritin; genetic testing for *HFE* gene mutations
Hepatocellular cancer	Elevated α-fetoprotein level (to >500 ng/mL); ultrasound or CT image of mass

[a] Also known as metabolic dysfunction–associated steatohepatitis (MASH).

Abbreviations: ANA, antinuclear antibody; anti-HBc, antibody to hepatitis B core (antigen); CT, computed tomography; HAV, HBV, HCV, HDV, HEV, hepatitis A, B, C, D, E virus; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; P-ANCA, peripheral antineutrophil cytoplasmic antibody; SMA, smooth-muscle antibody.

The most common causes of chronic liver disease, in general order of frequency, are chronic hepatitis C, alcoholic liver disease, nonalcoholic steatohepatitis (also called metabolic dysfunction–associated steatohepatitis), chronic hepatitis B, autoimmune hepatitis, sclerosing cholangitis, primary biliary cholangitis, hemochromatosis, and Wilson disease. Hepatitis E virus is a rare cause of chronic hepatitis, with cases occurring mostly in persons who are immunosuppressed or immunodeficient. Strict diagnostic criteria have not been developed for most liver diseases, but liver biopsy plays an important role in the diagnosis of autoimmune hepatitis, primary biliary cholangitis, nonalcoholic and alcoholic steatohepatitis, and Wilson disease (with a quantitative hepatic copper level in the last instance).

Laboratory Testing Diagnosis of liver disease is greatly aided by the availability of reliable and sensitive tests of liver injury and function. A typical battery of blood tests used for initial assessment of liver disease includes measurement of levels of serum alanine (ALT) and aspartate (AST) aminotransferases, alkaline phosphatase (AlkP), direct and total serum bilirubin and albumin, and prothrombin time. The pattern of abnormalities generally points to hepatocellular versus cholestatic liver disease and helps determine whether the disease is acute or chronic and whether cirrhosis and hepatic failure are present. Based on these results, further testing over time may be necessary. Other laboratory tests may be helpful, such as γ-glutamyl transpeptidase (γGT) to define whether AlkP elevations are due to liver disease; hepatitis serology to define the type of viral hepatitis; and autoimmune markers to diagnose primary biliary cholangitis (antimitochondrial

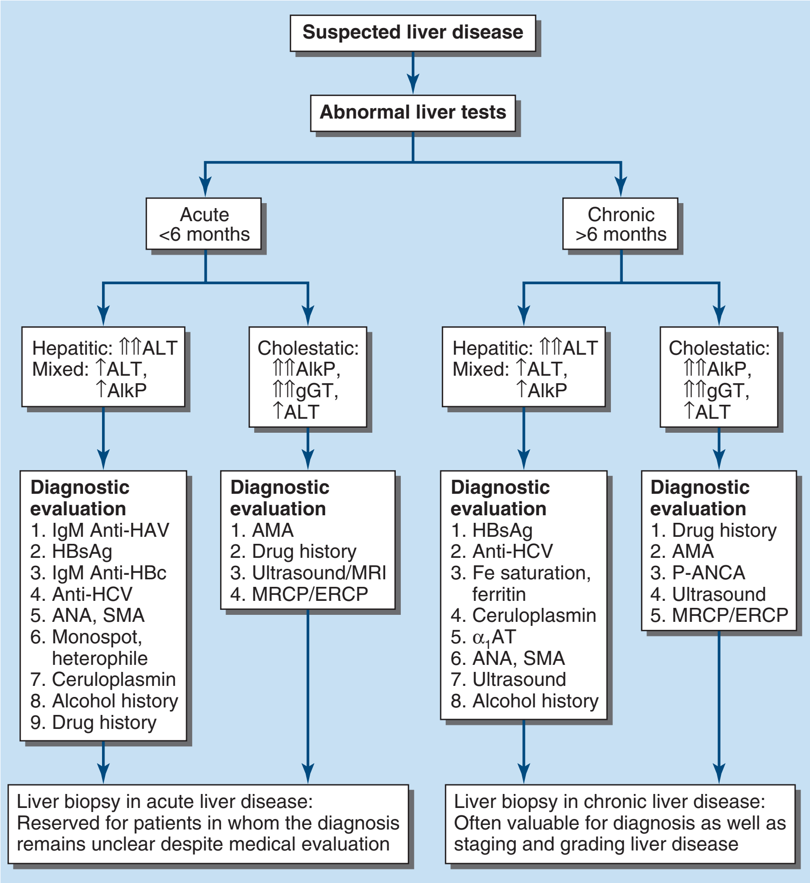
Suspected liver disease
Abnormal liver tests
Acute <6 months
Chronic >6 months
Hepatic: ⇑⇑ALT Mixed: ⇑ALT, ⇑AlkP
Cholestatic: ⇑⇑AlkP, ⇑⇑gGT, ⇑ALT
Hepatic: ⇑⇑ALT Mixed: ⇑ALT, ⇑AlkP
Cholestatic: ⇑⇑AlkP, ⇑⇑gGT, ⇑ALT
Diagnostic evaluation
1. IgM Anti-HAV
2. HBsAg
3. IgM Anti-HBc
4. Anti-HCV
5. ANA, SMA
6. Monospot, heterophile
7. Ceruloplasmin
8. Alcohol history
9. Drug history
Diagnostic evaluation
1. AMA
2. Drug history
3. Ultrasound/MRI
4. MRCP/ERCP
Diagnostic evaluation
1. HBsAg
2. Anti-HCV
3. Fe saturation, ferritin
4. Ceruloplasmin
5. α1AT
6. ANA, SMA
7. Ultrasound
8. Alcohol history
Diagnostic evaluation
1. Drug history
2. AMA
3. P-ANCA
4. Ultrasound
5. MRCP/ERCP
Liver biopsy in acute liver disease: Reserved for patients in whom the diagnosis remains unclear despite medical evaluation
Liver biopsy in chronic liver disease: Often valuable for diagnosis as well as staging and grading liver disease

FIGURE 347-1 Algorithm for evaluation of abnormal liver tests. For patients with suspected liver disease, an appropriate approach to evaluation is initial routine liver testing—for example, measurement of serum bilirubin, albumin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), and alkaline phosphatase (AlkP). These results (sometimes complemented by testing of γ-glutamyl transpeptidase [γGT]) will establish whether the pattern of abnormalities is hepatic, cholestatic, or mixed. In addition, the duration of symptoms or abnormalities will indicate whether the disease is acute or chronic. If the disease is acute and if history, laboratory tests, and imaging studies do not reveal a diagnosis, liver biopsy is appropriate to help establish the diagnosis. If the disease is chronic, liver biopsy can be helpful not only for diagnosis but also for grading of the activity and staging the progression of disease. This approach is generally applicable to patients without immune deficiency. In patients with HIV infection or recipients of bone marrow or solid organ transplants, the diagnostic evaluation should also include evaluation for opportunistic infections (e.g., with adenovirus, cytomegalovirus, *Coccidioides*, hepatitis E virus) as well as for vascular and immunologic conditions (venoocclusive disease, graft-versus-host disease). α_1AT, α_1 antitrypsin; AMA, antimitochondrial antibody; ANA, antinuclear antibody; anti-HBc, antibody to hepatitis B core (antigen); ERCP, endoscopic retrograde cholangiopancreatography; HAV, hepatitis A virus; HBsAg, hepatitis B surface antigen; HCV, hepatitis C virus; MRCP, magnetic resonance cholangiopancreatography; P-ANCA, peripheral antineutrophil cytoplasmic antibody; SMA, smooth-muscle antibody.

antibody), sclerosing cholangitis (peripheral antineutrophil cytoplasmic antibody), and autoimmune hepatitis (antinuclear, smooth-muscle, and liver-kidney microsomal antibody). A simple delineation of laboratory abnormalities and common liver diseases is given in Table 347-3.

The use and interpretation of liver function tests are summarized in Chap. 348.

Diagnostic Imaging Great advances have been made in hepatobiliary imaging, although no method is adequately accurate in demonstrating underlying cirrhosis in its early stages. Of the many modalities available for imaging the liver, US, computed tomography (CT), and magnetic resonance imaging (MRI) are the most commonly employed and are complementary to one another. In general, US and CT are highly sensitive for detecting biliary duct dilation and are the first-line options for investigating cases of suspected obstructive jaundice. All three modalities can detect a fatty liver, which appears bright on imaging studies. Modifications of CT and MRI can be used to quantify liver fat, and this information may ultimately be valuable in monitoring response to therapy in patients with fatty liver disease. Advantages, disadvantages, and clinical utility of each modality are presented in Table 347-4. Magnetic resonance cholangiopancreatography (MRCP) and endoscopic retrograde cholangiopancreatography (ERCP) are the procedures of choice for visualization of the biliary tree. MRCP offers several advantages over ERCP: there is no need for contrast media or ionizing radiation, images can be acquired faster, the procedure is less operator dependent, and it carries no risk of pancreatitis. MRCP is superior to US and CT for detecting choledocholithiasis but is less specific. MRCP is useful in the diagnosis of bile duct obstruction and congenital biliary abnormalities, but ERCP is considered more valuable in evaluating ampullary lesions and primary sclerosing cholangitis. ERCP permits biopsy, direct visualization of the ampulla and common bile duct, and intraductal ultrasonography and brushings for cytologic evaluation of malignancy. It also provides several therapeutic options in patients with obstructive jaundice, such as sphincterotomy, stone extraction, and placement of nasobiliary catheters and biliary stents.

Doppler US and MRI are used to assess hepatic vasculature and hemodynamics and to monitor surgically or radiologically placed vascular shunts, including

TABLE 347-4 Diagnostic Tests to Assess Liver Fat

IMAGING MODALITY	ADVANTAGES	DISADVANTAGES	CLINICAL UTILITY
Ultrasound	No radiation Widely available	Operator dependent Imprecise qualitative assessment of fat severity, particularly mild steatosis	Initial screening test for suspected liver fat
Transient elastography with controlled attenuation parameter	No radiation Point-of-care assessment of liver fat Provides semiquantitative assessment of fat severity	Requires special software No reliable cutoff for diagnosis of liver fat Imprecise qualitative assessment of fat severity	Alternate screening test for suspected liver fat if available
Computed tomography	Rapid assessment Non–operator dependent Quantitative assessment of fat severity	Requires radiation Quantification of fat requires specific protocols Imprecise quantitative assessment of fat severity, particularly mild steatosis	Not recommended for clinical assessment of liver fat due to need for radiation exposure and low sensitivity for mild fat
Magnetic resonance imaging–proton density fat fraction	Direct assessment of liver fat Highly sensitive and specific	Relatively limited accessibility	Test of choice for quantitative assessment of liver fat if available

transjugular intrahepatic portosystemic shunts. Multidetector or spiral CT and MRI with contrast enhancement are the procedures of choice for the identification and evaluation of hepatic masses, the staging of liver tumors, and preoperative assessment. With regard to mass lesions, the sensitivity of hepatic imaging continues to increase; unfortunately, specificity remains a problem, and often two and sometimes three studies are needed before a diagnosis can be reached. An emerging imaging modality for the investigation of hepatic lesions is contrast-enhanced US. This procedure permits enhancement of liver lesions in a similar fashion as contrast-enhanced, cross-sectional CT or MRI. Major advantages are real-time assessment of liver perfusion throughout the vascular phases without risk of nephrotoxicity and radiation exposure. Other advantages are its widespread availability and lower cost. Limitations include body habitus of the patient and skill of the operator.

US is the recommended modality for HCC screening. Contrast-enhanced US, CT, and MRI are appropriate for further investigation of lesions detected on screening US. The American College of Radiologists has developed a Liver Imaging Reporting and Data System (LI-RADS) to standardize the reporting and data collection of CT, MRI, and contrast-enhanced US imaging for HCC. This system allows for more consistent reporting and reduces imaging interpretation variability and errors.

Recently, several US-based elastographic techniques have been developed and approved for the measurement of hepatic stiffness, providing an indirect assessment of fibrosis and cirrhosis. The most commonly used approaches in clinical practice include transient elastography, acoustic radiation force impulse imaging, shear-wave elasticity imaging, and supersonic shear imaging. These techniques can eliminate the need for liver biopsy if the only indication for the test is the assessment of disease stage. Magnetic resonance elastography is more sensitive than US elastography but is also more expensive and requires advanced scheduling and special equipment. Finally, interventional radiologic techniques allow for the biopsy of solitary lesions, the radiofrequency ablation and chemoembolization of cancerous lesions, the insertion of drains into hepatic abscesses, the measurement of portal pressure, and the creation of vascular shunts in patients with portal hypertension. Which modality to use depends on factors such as availability, cost, and experience of the radiologist with each technique.

Liver Biopsy Liver biopsy remains the gold standard in the evaluation of patients with liver disease, particularly chronic liver disease. Liver biopsy is necessary for diagnosis in selected instances but is more often useful for assessment of the severity (grade) and stage of liver damage, prediction of prognosis, and monitoring of the response to treatment. The size of the liver biopsy sample is an important determinant of reliability; a length of 1.5–2 cm with 10 portal tracts is necessary for accurate assessment of fibrosis. Because liver biopsy is an invasive procedure and not without complications, it should be used only when it will contribute materially to decisions about management and therapy. In the future, noninvasive means of assessing disease activity (batteries of blood tests) and fibrosis (elastography and fibrosis markers) may replace liver biopsy for the staging and grading of disease.

■ GRADING AND STAGING OF LIVER DISEASE

Grading refers to an assessment of the severity or activity of liver disease, whether acute or chronic; active or inactive; and mild, moderate, or severe. Liver biopsy is the most accurate means of assessing severity, particularly in chronic liver disease. Serum aminotransferase levels serve as convenient and noninvasive markers for disease activity but do not always reliably reflect disease severity. Thus, normal serum aminotransferase levels in patients with hepatitis B surface antigen in serum may indicate the inactive carrier state or may reflect mild chronic hepatitis B or hepatitis B with fluctuating disease activity. Serum testing for hepatitis B e antigen and hepatitis B virus DNA can help sort out these different patterns, but these markers can also fluctuate and change over time. Similarly, in chronic hepatitis C, serum aminotransferase levels can be normal despite moderate disease activity. Finally, in both alcoholic and nonalcoholic steatohepatitis, aminotransferase levels are quite unreliable in reflecting severity. In these conditions, liver biopsy is helpful in guiding management and identifying appropriate therapy, particularly if treatment is difficult, prolonged, and expensive. Of the several well-verified numerical scales for histologic grading of chronic liver disease, the most commonly used are the METAVIR and the histology activity index.

Liver biopsy is also the most accurate means of assessing stage of disease as early or advanced and as precirrhotic or cirrhotic. Staging of disease pertains largely to chronic liver diseases in which progression to cirrhosis and end-stage disease can occur but may require years or decades. Clinical features, biochemical tests, and hepatic imaging studies are helpful in assessing stage but generally become abnormal only in the middle to late stages of cirrhosis. Noninvasive tests that suggest advanced fibrosis include mild elevations of bilirubin, prolongation of prothrombin time, slight decreases in serum albumin, and mild thrombocytopenia (which is often the first indication of worsening fibrosis). Combinations of blood test results that include clinical features, routine laboratory tests, and special laboratory tests such as serum proteins or small molecules that are affected by or involved with fibrogenesis have been used to create models for predicting advanced liver disease, but these models are not reliable enough to use on a regular basis or for repeated measures and only separate advanced from early disease **(Table 347-5)**. Recently, elastography and noninvasive breath tests using ^{13}C-labeled compounds have been proposed as a means of detecting early stages of fibrosis and liver dysfunction, but their reliability and reproducibility remain to be proven. A major limitation of noninvasive markers is that they can be affected by disease activity. Even elastography is limited in this regard, in that it measures liver stiffness, not fibrosis per se, and can be affected by inflammation, edema, hepatocyte necrosis, and intrasinusoidal cellularity (inflammatory, malignant, or sickled cells). Thus, at present, mild to moderate stages of hepatic fibrosis are detectable only by liver biopsy. In the assessment of stage, the degree of fibrosis is usually used as the quantitative measure. The amount of fibrosis is generally staged on a scale of 0 to 4+ (METAVIR scale) or 0 to 6+ (Ishak scale). The importance of staging relates primarily to prognosis, recommendation of therapy, and optimal management to prevent complications of chronic liver disease. Patients with cirrhosis are candidates for screening and surveillance for esophageal varices and HCC. Patients without advanced fibrosis need not undergo screening.

Once cirrhosis develops, other scoring systems are employed to assess compensated versus decompensated disease and prognosis. The first staging system used for this purpose was the modified Child-Pugh

TABLE 347-5 Selected Noninvasive Methods of Assessing Hepatic Fibrosis and Cirrhosis

METHOD	PARAMETERS	ADVANCED FIBROSIS	CIRRHOSIS
APRI	AST, platelet count	>1	>1.5 (1–2)
ELF	Age, hyaluronic acid, MMP-3, TIMP-1	>7.7	>9.3
FIB-4	Age, AST, ALT, platelet count	>1.45	>3.25
Fibro test[a]	Haptoglobin, α_2-macroglobulin, apolipoprotein A1, γGT, total bilirubin	>0.45	>0.63
TE	Measures speed of a shear wave generated by vibration through liver tissue	>7.3 kPa	>15 kPa (9–26.5 kPa)
ARFI	Measures speed of shear wave generated by acoustic radiation force through liver tissue	>1.3 m/s	>1.87 m/s

[a]Patented models.

Note: The cut points presented in the table were mostly derived from patients with chronic hepatitis C. The cut points for the noninvasive models and tests presented in the table vary among different liver diseases and among patients with the same disease among different populations.

Abbreviations: ALT, alanine aminotransferase; APRI, AST-to-platelet ratio; ARFI, acoustic radiation force imaging; AST, aspartate aminotransferase; ELF, enhanced liver fibrosis panel; γGT, γ-glutamyl transpeptidase; MMP-3, metalloproteinase-3; TIMP-1, tissue inhibitor of metalloproteinase-1; TE, transient elastography.

TABLE 347-6 Child-Pugh Classification of Cirrhosis

	POINTS TOWARD TOTAL SCORE			
FACTOR	**UNITS**	**1**	**2**	**3**
Serum bilirubin	μmol/L	<34	34–51	>51
	mg/dL	<2.0	2.0–3.0	>3.0
Serum albumin	g/L	>35	30–35	<30
	g/dL	>3.5	3.0–3.5	<3.0
Prothrombin time	Seconds prolonged	<4	4–6	>6
	INR[a]	<1.7	1.7–2.3	>2.3
Ascites		None	Easily controlled	Poorly controlled
Hepatic encephalopathy		None	Minimal	Advanced

[a] International normalized ratio.

Note: The Child-Pugh score is calculated by adding the scores for the five factors and can range from 5 to 15. The resulting Child-Pugh class can be A (a score of 5–6), B (7–9), or C (≥10). Decompensation indicates cirrhosis, with a Child-Pugh score of ≥7 (class B). This level has been the accepted criterion for listing a patient for liver transplantation.

classification, with a scoring system of 5–15: scores of 5 and 6 represent Child-Pugh class A (consistent with "compensated cirrhosis"), scores of 7–9 represent class B, and scores of 10–15 represent class C **(Table 347-6)**. This scoring system was initially devised to stratify patients with cirrhosis into risk groups before portal decompressive surgery. The Child-Pugh score is a reasonably reliable predictor of survival in many liver diseases and predicts the likelihood of major complications of cirrhosis, such as bleeding from varices and spontaneous bacterial peritonitis. This classification scheme was used to assess prognosis in cirrhosis and to provide standard criteria for listing a patient as a candidate for liver transplantation (Child-Pugh class B). The Child-Pugh system has been replaced by the Model for End-Stage Liver Disease (MELD) system for the latter purpose. The MELD score is a prospectively derived system designed to predict the short-term mortality of patients awaiting orthotopic liver transplant and following a transjugular intrahepatic portosystemic shunt (TIPS). This score is calculated using three readily available objective variables: the prothrombin time expressed as the international normalized ratio (INR), the serum bilirubin level, and the serum creatinine concentration. The ability of the MELD score to predict outcome after liver transplantation is regularly monitored and was modified to increase its accuracy and improve allocation of donated livers. These modifications include serum sodium concentration as a factor in the model and a reweighting of the MELD components. A separate scoring system, the Pediatric End-Stage Liver Disease (PELD) score, is used for children (<12 years old). Transient elastography has also been used to stage cirrhosis and has been shown to be useful in predicting complications such as variceal hemorrhage, ascites development, and liver-related death.

The MELD system provides a more objective means of assessing disease severity and has less center-to-center variation than the Child-Pugh score as well as a wider range of values. The MELD and PELD systems are currently used to establish priority listing for liver transplantation in the United States. Convenient MELD and PELD calculators are available via the Internet (*https://optn.transplant.hrsa.gov*).

NONSPECIFIC ISSUES IN THE MANAGEMENT OF PATIENTS WITH LIVER DISEASE

Specifics on the management of different forms of acute or chronic liver disease are supplied in subsequent chapters, but certain issues are applicable to any patient with liver disease. These issues include advice regarding alcohol use, medication use, vaccination, and surveillance for certain liver diseases and complications of liver disease. Alcohol should be used sparingly, if at all, by patients with liver disease. Abstinence from alcohol should be encouraged for all patients with alcohol-related liver disease, patients with cirrhosis, and patients receiving interferon-based therapy for hepatitis B or D and during antiviral therapy of hepatitis C. With regard to vaccinations, all patients with liver disease should receive hepatitis A and hepatitis B vaccine, if not previously vaccinated, and screened for hepatitis C if not already done. Adherence to the recommendations of the Centers for Disease Control and Prevention (CDC) on other adult vaccinations should also be encouraged. Patients with liver disease should exercise caution in using any medications other than those that are most necessary. Drug-induced hepatotoxicity can mimic many forms of liver disease and can cause exacerbations of chronic hepatitis and cirrhosis; drugs should be suspected in any situation in which the cause of exacerbation is unknown. Finally, consideration should be given to surveillance for complications of chronic liver disease such as variceal hemorrhage and HCC. Cirrhosis warrants upper endoscopy to assess the presence of varices, and the patient should receive chronic therapy with beta blockers or should be offered endoscopic obliteration if large varices are found. Moreover, cirrhosis warrants screening and long-term surveillance for development of HCC. While the optimal regimen for such surveillance has not been established, an appropriate approach is US of the liver at 6- to 12-month intervals.

FURTHER READING

Jeng W-J et al: Hepatitis B. Lancet 401:1039, 2023.

Manns MP et al: Breakthroughs in hepatitis C research: From discovery to cure. Gastroenterol Hepatol 19:533, 2022.

Powell EE et al: Non-alcoholic fatty liver disease. Lancet 397:2212, 2021.

Tapper EB, Lok AS: Use of liver imaging and biopsy in clinical practice. N Engl J Med 377:756, 2017.

348

Evaluation of Liver Function

Emily D. Bethea, Daniel S. Pratt

There are a number of tests that can be used to evaluate liver function. These tests include biochemical tests, radiologic tests, and pathologic tests.

Serum biochemical tests, also commonly referred to as "liver function tests," can be used to (1) detect the presence of liver disease, (2) distinguish among different types of liver disorders, (3) gauge the extent of liver damage, and (4) follow the response to treatment. However, serum biochemical tests have shortcomings. They lack sensitivity and specificity; they can be normal in patients with serious liver disease and abnormal in patients with diseases that do not affect the liver. Liver tests rarely suggest a specific diagnosis; rather, they suggest a general category of liver disease, such as hepatocellular or cholestatic, which then further directs the evaluation. The liver carries out thousands of biochemical functions, most of which cannot be easily measured by blood tests. Laboratory tests measure only a limited number of these functions. In fact, many tests, such as the aminotransferases and alkaline phosphatase, do not measure liver function at all. Rather, they detect liver cell damage or interference with bile flow. Thus, no one biochemical test enables the clinician to accurately assess the liver's total functional capacity.

To increase the sensitivity and the specificity of biochemical tests in the detection of liver disease, it is best to use them as a battery. Tests usually employed in clinical practice include the bilirubin, aminotransferases, alkaline phosphatase, albumin, and prothrombin time tests. When more than one of these tests provide abnormal findings or the findings are persistently abnormal on serial determinations, the probability of liver disease is high. When all test results are normal, the probability of missing occult liver disease is low.

Serum Bilirubin (See also Chap. 52) Bilirubin, a breakdown product of the porphyrin ring of heme-containing proteins, is found in the blood in two fractions—conjugated and unconjugated. The unconjugated fraction, also termed the *indirect fraction*, is insoluble in water and is bound to albumin in the blood. The conjugated (direct) bilirubin fraction is water-soluble and can therefore be excreted by the kidney. Normal values of total serum bilirubin are reported between 1 and 1.5 mg/dL with 95% of a normal population falling between 0.2 and 0.9 mg/dL. If the direct-acting fraction is <15% of the total, the bilirubin can be considered to all be indirect. The most frequently reported upper limit of normal for conjugated bilirubin is 0.3 mg/dL.

Elevation of the unconjugated fraction of bilirubin is rarely due to liver disease. An isolated elevation of unconjugated bilirubin is seen primarily in hemolytic disorders and in a number of genetic conditions such as Crigler-Najjar and Gilbert's syndromes (Chap. 52). Isolated unconjugated hyperbilirubinemia (bilirubin elevated but <15% direct) should prompt a workup for hemolysis (Fig. 348-1). In the absence of hemolysis, an isolated, unconjugated hyperbilirubinemia in an otherwise healthy patient can be attributed to Gilbert's syndrome, and no further evaluation is required.

In contrast, conjugated hyperbilirubinemia almost always implies liver or biliary tract disease. The rate-limiting step in bilirubin metabolism is not conjugation of bilirubin, but rather the transport of conjugated bilirubin into the bile canaliculi. Thus, elevation of the conjugated fraction may be seen in any type of liver disease including fulminant liver failure. In most liver diseases, both conjugated and unconjugated fractions of the bilirubin tend to be elevated. Except in the presence of an isolated hyperbilirubinemia, fractionation of the bilirubin is rarely helpful in determining the cause of jaundice.

Although the degree of elevation of the serum bilirubin has not been critically assessed as a prognostic marker, it is important in a number of conditions. In viral hepatitis, the higher the serum bilirubin, the greater is the hepatocellular damage. Total serum bilirubin correlates with poor outcomes in alcoholic hepatitis. It is also a critical component of the Model for End-Stage Liver Disease (MELD) score, a tool used to estimate survival of patients with end-stage liver disease, prioritize patients awaiting liver transplantation, and assess operative risk of patients with cirrhosis. An elevated total serum bilirubin in patients with drug-induced liver disease indicates more severe injury. Total bilirubin has also been shown to be of prognostic value in primary biliary cholangitis.

Unconjugated bilirubin always binds to albumin in the serum and is not filtered by the kidney. Therefore, any bilirubin found in the urine is conjugated bilirubin; the presence of bilirubinuria implies the presence of liver disease or obstructive jaundice. A urine dipstick test can theoretically give the same information as fractionation of the serum bilirubin. This test is almost 100% accurate. Phenothiazines may give a false-positive reading with the Ictotest tablet. In patients recovering from jaundice, the urine bilirubin clears prior to the serum bilirubin.

Serum Enzymes The liver contains thousands of enzymes, some of which are also present in the serum in very low concentrations. These enzymes have no known function in the serum and behave like other

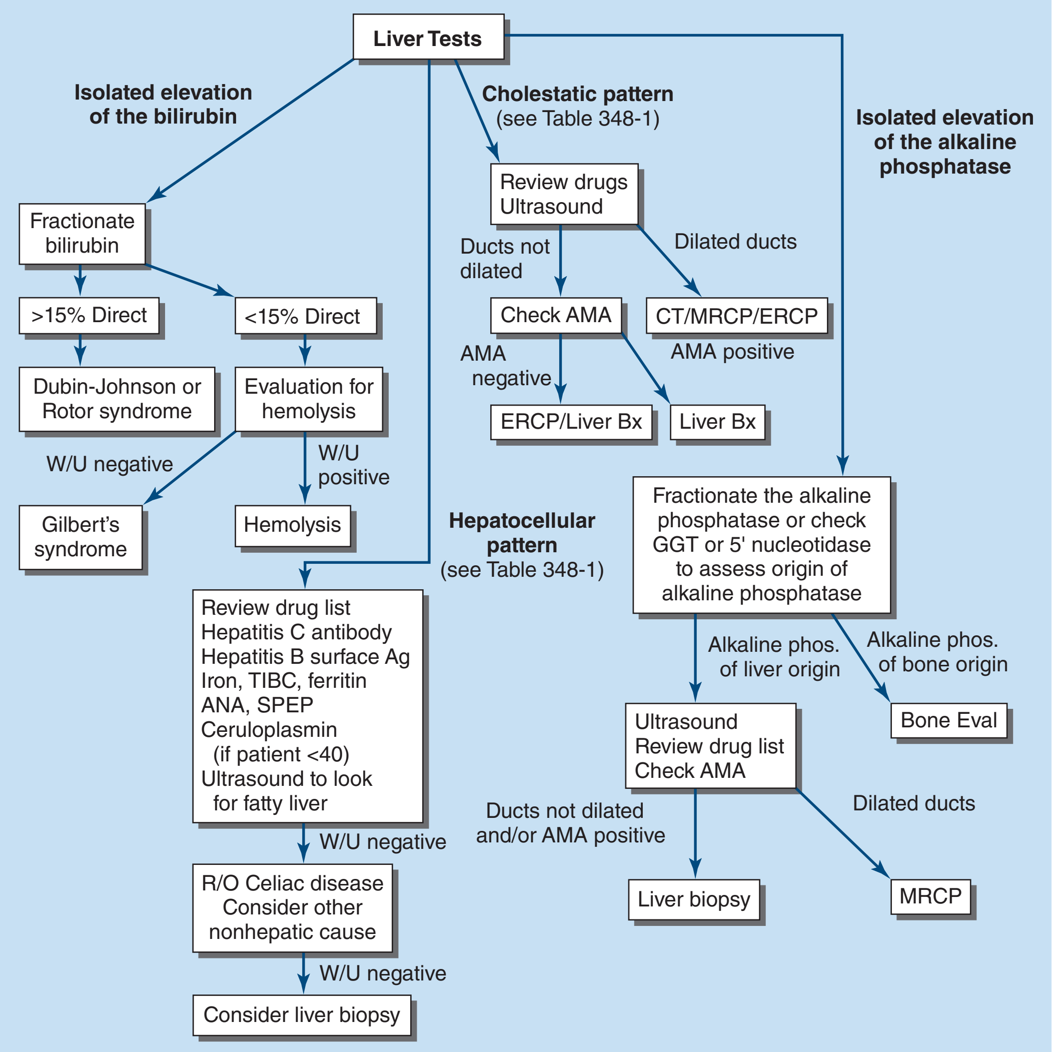
Liver Tests
Isolated elevation of the bilirubin
Cholestatic pattern (see Table 348-1)
Isolated elevation of the alkaline phosphatase
Fractionate bilirubin
>15% Direct
<15% Direct
Dubin-Johnson or Rotor syndrome
Evaluation for hemolysis
W/U negative
W/U positive
Gilbert's syndrome
Hemolysis
Review drugs Ultrasound
Ducts not dilated
Dilated ducts
Check AMA
CT/MRCP/ERCP
AMA negative
AMA positive
ERCP/Liver Bx
Liver Bx
Hepatocellular pattern (see Table 348-1)
Review drug list
Hepatitis C antibody
Hepatitis B surface Ag
Iron, TIBC, ferritin
ANA, SPEP
Ceruloplasmin (if patient <40)
Ultrasound to look for fatty liver
W/U negative
R/O Celiac disease
Consider other nonhepatic cause
W/U negative
Consider liver biopsy
Fractionate the alkaline phosphatase or check GGT or 5' nucleotidase to assess origin of alkaline phosphatase
Alkaline phos. of liver origin
Alkaline phos. of bone origin
Bone Eval
Ultrasound
Review drug list
Check AMA
Ducts not dilated and/or AMA positive
Dilated ducts
Liver biopsy
MRCP

FIGURE 348-1 Algorithm for the evaluation of chronically abnormal liver tests. Ag, antigen; AMA, antimitochondrial antibody; ANA, antinuclear antibody; Bx, biopsy; CT, computed tomography; ERCP, endoscopic retrograde cholangiopancreatography; GGT, γ-glutamyl transpeptidase; MRCP, magnetic resonance cholangiopancreatography; R/O, rule out; SPEP, serum protein electrophoresis; TIBC, total iron-binding capacity; W/U, workup.

serum proteins. They are distributed in the plasma and in interstitial fluid and have characteristic half-lives, which are usually measured in days. Very little is known about the catabolism of serum enzymes, although they are probably cleared by cells in the reticuloendothelial system. The elevation of a given enzyme activity in the serum is thought to primarily reflect its increased rate of entrance into serum from damaged liver cells.

Serum enzyme tests can be grouped into two categories: (1) enzymes whose elevation in serum reflects damage to hepatocytes and (2) enzymes whose elevation in serum reflects cholestasis.

ENZYMES THAT REFLECT DAMAGE TO HEPATOCYTES The aminotransferases (transaminases) are sensitive indicators of liver cell injury and are most helpful in recognizing acute hepatocellular diseases such as hepatitis. They include aspartate aminotransferase (AST) and alanine aminotransferase (ALT). AST is found in the liver, cardiac muscle, skeletal muscle, kidneys, brain, pancreas, lungs, leukocytes, and erythrocytes in decreasing order of concentration. ALT is found primarily in the liver and is therefore a more specific indicator of liver injury. The aminotransferases are normally present in the serum in low concentrations. These enzymes are released into the blood in greater amounts when there is damage to the liver cell membrane, resulting in increased permeability. Liver cell necrosis is not required for the release of the aminotransferases, and there is a poor correlation between the degree of liver cell damage and the level of the aminotransferases. Thus, the absolute elevation of the aminotransferases is of no prognostic significance in acute hepatocellular disorders.

The normal range for aminotransferases varies widely among laboratories, but generally ranges from 10 to 40 IU/L. The interlaboratory variation in normal range is due to technical reasons; no reference standards exist to establish upper limits of normal for ALT and AST. Some have recommended revisions of normal limits of the aminotransferases to adjust for sex and body mass index, but others have noted the potential costs and unclear benefits of implementing this change.

Any type of liver cell injury can cause modest elevations in the serum aminotransferases. Levels of up to 300 IU/L are nonspecific and may be found in any type of liver disorder. Minimal ALT elevations in asymptomatic blood donors rarely indicate severe liver disease; studies have shown that fatty liver disease is the most likely explanation. Striking elevations—that is, aminotransferases >1000 IU/L—occur almost exclusively in disorders associated with extensive hepatocellular injury such as (1) viral hepatitis, (2) ischemic liver injury (prolonged hypotension or acute heart failure), or (3) toxin- or drug-induced liver injury.

The pattern of the aminotransferase elevation can be helpful diagnostically. In most acute hepatocellular disorders, the ALT is higher than or equal to the AST. Whereas the AST:ALT ratio is typically <1 in patients with chronic viral hepatitis and nonalcoholic fatty liver disease, a number of groups have noted that as cirrhosis develops, this ratio rises to >1. An AST:ALT ratio >2:1 is suggestive, whereas a ratio >3:1 is highly suggestive, of alcohol-related liver disease. The AST in alcohol-related liver disease is rarely >300 IU/L, and the ALT is often normal. A low level of ALT in the serum is due to an alcohol-induced deficiency of pyridoxal phosphate.

The aminotransferases are usually not greatly elevated in obstructive jaundice. One notable exception occurs during the acute phase of biliary obstruction caused by the passage of a gallstone into the common bile duct. In this setting, the aminotransferases can briefly be in the 1000–2000 IU/L range. However, aminotransferase levels decrease quickly, and the biochemical tests rapidly evolve into those typical of cholestasis.

ENZYMES THAT REFLECT CHOLESTASIS The activities of three enzymes—alkaline phosphatase, 5′-nucleotidase, and γ-glutamyl transpeptidase (GGT)—are usually elevated in cholestasis. Alkaline phosphatase and 5′-nucleotidase are found in or near the bile canalicular membrane of hepatocytes, whereas GGT is located in the endoplasmic reticulum and in bile duct epithelial cells. Reflecting its more diffuse localization in the liver, GGT elevation in serum is less specific for cholestasis than are elevations of alkaline phosphatase or 5′-nucleotidase. Some have advocated the use of GGT to identify patients with occult alcohol use. Its lack of specificity makes its use in this setting questionable.

The normal serum alkaline phosphatase consists of many distinct isoenzymes found in the liver, bone, placenta, and, less commonly, the small intestine. Patients over age 60 can have a mildly elevated alkaline phosphatase (1–1.5 times normal), whereas individuals with blood types O and B can have an elevation of the serum alkaline phosphatase after eating a fatty meal due to the influx of intestinal alkaline phosphatase into the blood. It is also elevated in children and adolescents undergoing rapid bone growth because of bone alkaline phosphatase and late in normal pregnancies due to the influx of placental alkaline phosphatase.

Elevation of liver-derived alkaline phosphatase is not totally specific for cholestasis, and a less than threefold elevation can be seen in almost any type of liver disease. Alkaline phosphatase elevations greater than four times normal occur primarily in patients with cholestatic liver disorders, infiltrative liver diseases such as cancer and amyloidosis, and bone conditions characterized by rapid bone turnover (e.g., Paget's disease). In bone diseases, the elevation is due to increased amounts of the bone isoenzymes. In liver diseases, the elevation is almost always due to increased amounts of the liver isoenzyme.

If an elevated serum alkaline phosphatase is the only abnormal finding in an apparently healthy person or if the degree of elevation is higher than expected in the clinical setting, identification of the source of elevated isoenzymes is helpful (Fig. 348-1). This problem can be approached in two ways. First, and most precise, is the fractionation of the alkaline phosphatase by electrophoresis. The second, best substantiated, and most available approach involves the measurement of serum 5′-nucleotidase or GGT.

In the absence of jaundice or elevated aminotransferases, an elevated alkaline phosphatase of liver origin often, but not always, suggests early cholestasis and, less often, hepatic infiltration by tumor or granulomata. Other conditions that cause isolated elevations of the alkaline phosphatase include primary biliary cholangitis, primary or secondary sclerosing cholangitis, Hodgkin's disease, diabetes, hyperthyroidism, congestive heart failure, and amyloidosis.

The level of serum alkaline phosphatase elevation is not helpful in distinguishing between intrahepatic and extrahepatic cholestasis. There is essentially no difference among the values found in obstructive jaundice due to cancer, common duct stone, sclerosing cholangitis, or bile duct stricture. Values are similarly increased in patients with intrahepatic cholestasis due to drug-induced hepatitis, primary biliary cholangitis, sepsis, rejection of transplanted livers, and, rarely, alcohol-induced steatohepatitis. Values are also greatly elevated in hepatobiliary disorders seen in patients with AIDS (e.g., AIDS cholangiopathy due to cytomegalovirus or cryptosporidial infection and tuberculosis with hepatic involvement).

■ TESTS THAT MEASURE BIOSYNTHETIC FUNCTION OF THE LIVER

Serum Albumin Serum albumin is synthesized exclusively by hepatocytes. Serum albumin has a long half-life: 18–20 days, with ~4% degraded per day. Because of this slow turnover, the serum albumin is not a good indicator of acute or mild hepatic dysfunction; only minimal changes in the serum albumin are seen in acute liver conditions such as viral hepatitis, drug-related hepatotoxicity, and obstructive jaundice. In hepatitis, albumin levels <3 g/dL should raise the possibility of chronic liver disease. Hypoalbuminemia is more common in chronic liver disorders such as cirrhosis and usually reflects severe liver damage and decreased albumin synthesis. However, hypoalbuminemia is not specific for liver disease and may occur in protein malnutrition of any cause, as well as protein-losing enteropathies, nephrotic syndrome, and chronic infections that are associated with prolonged increases in levels of cytokines that inhibit albumin synthesis, such as serum interleukin 1 and/or tumor necrosis factor. Serum albumin should not be measured to screen patients in whom there is no suspicion of liver disease. A general medical clinic study of consecutive patients in whom no indications were present for albumin measurement showed that although

12% of patients had abnormal test results, the finding was of clinical importance in only 0.4%.

Serum Globulins Serum globulins are a group of proteins made up of γ globulins (immunoglobulins) produced by B lymphocytes and α and β globulins produced primarily in hepatocytes. γ Globulins are increased in chronic liver disease, such as chronic hepatitis and cirrhosis. In cirrhosis, the increased serum γ globulin concentration is due to the increased synthesis of antibodies, some of which are directed against intestinal bacteria. This occurs because the cirrhotic liver fails to clear bacterial antigens that normally reach the liver through the hepatic circulation.

Increases in the concentration of specific isotypes of γ globulins are often helpful in the recognition of certain chronic liver diseases. Diffuse polyclonal increases in IgG levels are common in autoimmune hepatitis; increases >100% should alert the clinician to this possibility. Increases in the IgM levels are common in primary biliary cholangitis, whereas increases in the IgA levels occur in alcoholic liver disease.

COAGULATION FACTORS

With the exception of factor VIII, which is produced by vascular endothelial cells, the blood clotting factors are made exclusively in hepatocytes. Their serum half-lives are much shorter than albumin, ranging from 6 h for factor VII to 5 days for fibrinogen. Because of their rapid turnover, measurement of the clotting factors is the single best acute measure of hepatic synthetic function and helpful in both diagnosis and assessing the prognosis of acute parenchymal liver disease. Useful for this purpose is the *serum prothrombin time*, which collectively measures factors II, V, VII, and X. Biosynthesis of factors II, VII, IX, and X depends on vitamin K. The international normalized ratio (INR) is used to express the degree of anticoagulation on warfarin therapy. The INR standardizes prothrombin time measurement according to the characteristics of the thromboplastin reagent used in a particular lab, which is expressed as an International Sensitivity Index (ISI); the ISI is then used in calculating the INR.

The prothrombin time may be elevated in hepatitis and cirrhosis as well as in disorders that lead to vitamin K deficiency such as obstructive jaundice or fat malabsorption of any kind. Marked prolongation of the prothrombin time, >5 s above control and not corrected by parenteral vitamin K administration, is a poor prognostic sign in acute viral hepatitis and other acute and chronic liver diseases. The INR, along with the total serum bilirubin, creatinine, albumin, and sodium, are components of the MELD 3.0 score, which is used as a measure of hepatic decompensation and to allocate organs for liver transplantation.

OTHER DIAGNOSTIC TESTS

Although tests may direct the physician to a category of liver disease, additional biochemical testing, radiologic testing, and procedures are often necessary to make the proper diagnosis, as shown in Fig. 348-1. The most commonly used ancillary tests are reviewed here, as are the noninvasive tests available for assessing hepatic fibrosis.

Ammonia Ammonia is produced in the body during normal protein metabolism and by intestinal bacteria, primarily those in the colon. The liver plays a role in the detoxification of ammonia by converting it to urea, which is excreted by the kidneys. Striated muscle also plays a role in detoxification of ammonia, where it is combined with glutamic acid to form glutamine. Patients with advanced liver disease typically have significant muscle wasting, which likely contributes to hyperammonemia. Some physicians use the blood ammonia for detecting encephalopathy or for monitoring hepatic synthetic function, although its use for either of these indications has problems. There is very poor correlation between either the presence or the severity of acute encephalopathy and elevation of blood ammonia; it can be occasionally useful for identifying occult liver disease in patients with mental status changes. There is also a poor correlation of the blood serum ammonia and hepatic function. The ammonia can be elevated in patients with severe portal hypertension and portal blood shunting around the liver even in the presence of normal or near-normal hepatic function. Elevated arterial ammonia levels have been shown to correlate with outcome in fulminant hepatic failure.

Liver Biopsy Percutaneous biopsy of the liver is a safe procedure that is easily performed with local anesthesia and ultrasound guidance. Liver biopsy is of proven value in the following situations: (1) hepatocellular disease of uncertain cause, (2) prolonged hepatitis with the possibility of autoimmune hepatitis, (3) unexplained hepatomegaly, (4) unexplained splenomegaly, (5) hepatic lesions uncharacterized by radiologic imaging, (6) fever of unknown origin, and (7) staging of malignant lymphoma. Liver biopsy is most accurate in disorders causing diffuse changes throughout the liver and is subject to sampling error in focal disorders. Liver biopsy should not be the initial procedure in the diagnosis of cholestasis. The biliary tree should first be assessed for signs of obstruction. Contraindications to performing a percutaneous liver biopsy include significant ascites and prolonged INR. Under these circumstances, the biopsy can be performed via the transjugular approach.

Noninvasive Tests to Detect Hepatic Fibrosis Although liver biopsy is the standard for the assessment of hepatic fibrosis, noninvasive measures of hepatic fibrosis have gained favor. These measures include multiparameter tests aimed at detecting and staging the degree of hepatic fibrosis and imaging techniques. FibroTest (marketed as FibroSure in the United States) is the best evaluated of the multiparameter blood tests. The test incorporates haptoglobin, bilirubin, GGT, apolipoprotein A-I, and α_2-macroglobulin and has been found to have high positive and negative predictive values for diagnosing advanced fibrosis in patients with chronic hepatitis C, chronic hepatitis B, alcoholic liver disease, or metabolic dysfunction–associated steatotic liver disease (MASLD) and patients taking methotrexate for psoriasis. The Enhanced Liver Fibrosis (ELF) test is a multiparameter blood test that uses markers of fibrogenesis and cytolysis to estimate hepatic fibrosis: type III procollagen peptide, hyaluronic acid, and tissue inhibitor of metalloproteinase-1. It has been validated in patients with MASLD and is gaining favor in cholestatic liver disease. The FIB4 index is a liver fibrosis biomarker that is easily calculated using age and three routine laboratory parameters: ALT, AST, and platelet count. While originally proposed to help assess hepatic fibrosis in patients with human immunodeficiency virus and hepatitis C virus coinfection, it has been validated for hepatitis C virus, hepatitis B virus, and MASLD. Transient elastography (TE), marketed as FibroScan, and magnetic resonance elastography (MRE) both have gained U.S. Food and Drug Administration approval for use in the management of patients with liver disease. TE uses ultrasound waves to measure hepatic stiffness noninvasively. TE has been shown to be accurate for identifying advanced fibrosis in patients with chronic hepatitis C, primary biliary cholangitis, hemochromatosis, nonalcoholic fatty liver disease, and recurrent chronic hepatitis after liver transplantation. MRE has been found to be superior to TE for staging liver fibrosis in patients with a variety of chronic liver diseases but requires access to a magnetic resonance imaging scanner and is more expensive.

Ultrasonography Ultrasonography is the first diagnostic test to use in patients whose liver tests suggest cholestasis, to look for the presence of a dilated intrahepatic or extrahepatic biliary tree or to identify gallstones. In addition, it shows space-occupying lesions within the liver, enables the clinician to distinguish between cystic and solid masses, and helps direct percutaneous biopsies. Ultrasound with Doppler imaging can detect the patency of the portal vein, hepatic artery, and hepatic veins and determine the direction of blood flow. This is the first test ordered in patients suspected of having Budd-Chiari syndrome.

USE OF LIVER TESTS

As previously noted, the best way to increase the sensitivity and specificity of laboratory tests in the detection of liver disease is to employ a battery of tests that includes the aminotransferases, alkaline phosphatase, bilirubin, albumin, and prothrombin time along with the

TABLE 348-1 Liver Test Patterns in Hepatobiliary Disorders

TYPE OF DISORDER	BILIRUBIN	AMINOTRANSFERASES	ALKALINE PHOSPHATASE	ALBUMIN	PROTHROMBIN TIME
Hemolysis/Gilbert's syndrome	Normal to 86 μmol/L (5 mg/dL) 85% due to indirect fractions No bilirubinuria	Normal	Normal	Normal	Normal
Acute hepatocellular necrosis (viral, ischemic, and drug- or toxin-induced hepatitis)	Both fractions may be elevated Peak usually follows aminotransferases Bilirubinuria	Elevated, often >500 IU, ALT > AST	Normal to <3× normal elevation	Normal	Usually normal. If >5× above control and not corrected by parenteral vitamin K, suggests poor prognosis
Chronic hepatocellular disorders	Both fractions may be elevated Bilirubinuria	Elevated, but usually <300 IU	Normal to <3× normal elevation	Often decreased	Often prolonged Fails to correct with parenteral vitamin K
Alcoholic hepatitis, cirrhosis	Both fractions may be elevated Bilirubinuria	AST:ALT >2 suggests alcoholic hepatitis or cirrhosis	Normal to <3× normal elevation	Often decreased	Often prolonged Fails to correct with parenteral vitamin K
Intra- and extrahepatic cholestasis (obstructive jaundice)	Both fractions may be elevated Bilirubinuria	Normal to moderate elevation Rarely >500 IU	Elevated, often >4× normal elevation	Normal, unless chronic	Normal If prolonged, will correct with parenteral vitamin K
Infiltrative diseases (tumor, granulomata)	Usually normal	Normal to slight elevation	Elevated, often >4× normal elevation Fractionate, or confirm liver origin with 5′-nucleotidase or γ-glutamyl transpeptidase	Normal	Normal

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase.

judicious use of the other tests described in this chapter. **Table 348-1** shows how patterns of liver tests can lead the clinician to a category of disease that will direct further evaluation. However, it is important to remember that no single set of liver tests will necessarily provide a diagnosis. It is often necessary to repeat these tests on several occasions over days to weeks for a diagnostic pattern to emerge. Figure 348-1 is an algorithm for the evaluation of chronically abnormal liver tests.

GLOBAL CONSIDERATIONS

The tests and principles presented in this chapter are applicable worldwide. The causes of liver test abnormalities vary according to region. In developing nations, infectious diseases are more commonly the etiology of abnormal serum liver tests than in developed nations.

FURTHER READING

Berzigotti A et al: EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis: 2021 update. J Hepatology 75:659, 2021.

Kaplan M: Alkaline phosphatase. Gastroenterology 62:452, 1972.

Kim WR et al: MELD 3.0: The model for end-stage liver disease updated for the modern era. Gastroenterology 161:1887, 2021.

Prati D et al: Updated definitions of healthy ranges for serum alanine aminotransferase levels. Ann Intern Med 137:1, 2002.

349

The Hyperbilirubinemias

Allan W. Wolkoff

BILIRUBIN METABOLISM

The details of bilirubin metabolism are presented in **Chap. 52**. However, the hyperbilirubinemias are best understood in terms of perturbations of specific aspects of bilirubin metabolism and transport, and these will be briefly reviewed here as depicted in **Fig. 349-1**.

Bilirubin is the end product of heme degradation. Some 70–90% of bilirubin is derived from degradation of the hemoglobin of senescent red blood cells. Bilirubin produced in the periphery is transported to the liver within the plasma, where, due to its insolubility in aqueous solutions, it is tightly bound to albumin. Under normal circumstances, bilirubin is removed from the circulation rapidly and efficiently by hepatocytes. Transfer of bilirubin from blood to bile involves four distinct but interrelated steps (Fig. 349-1).

1. *Hepatocellular uptake:* Uptake of bilirubin by the hepatocyte has carrier-mediated kinetics. Although a number of candidate bilirubin transporters have been proposed, the identity of the actual transporter remains elusive.
2. *Intracellular binding:* Within the hepatocyte, bilirubin is kept in solution by binding as a nonsubstrate ligand to several of the glutathione-*S*-transferases, formerly called ligandins.
3. *Conjugation:* Bilirubin is conjugated with one or two glucuronic acid moieties by a specific UDP-glucuronosyltransferase to form

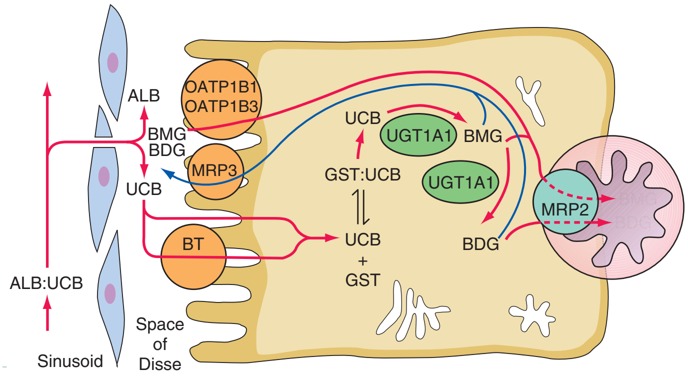

FIGURE 349-1 Hepatocellular bilirubin transport. Albumin-bound bilirubin in sinusoidal blood passes through endothelial cell fenestrae to reach the hepatocyte surface, entering the cell by both facilitated and simple diffusional processes. Within the cell, it is bound to glutathione-*S*-transferases and conjugated by bilirubin-UDP-glucuronosyltransferase (UGT1A1) to mono- and diglucuronides, which are actively transported across the canalicular membrane into the bile. In addition to this direct excretion of bilirubin glucuronides, a portion are transported into the portal circulation by MRP3 and subjected to reuptake into the hepatocyte by OATP1B1 and OATP1B3. ALB, albumin; BDG, bilirubin diglucuronide; BMG, bilirubin monoglucuronide; BT, proposed bilirubin transporter; GST, glutathione-*S*-transferase; MRP2 and MRP3, multidrug resistance–associated proteins 2 and 3; OATP1B1 and OATP1B3, organic anion transport proteins 1B1 and 1B3; UCB, unconjugated bilirubin; UGT1A1, bilirubin-UDP-glucuronosyltransferase.

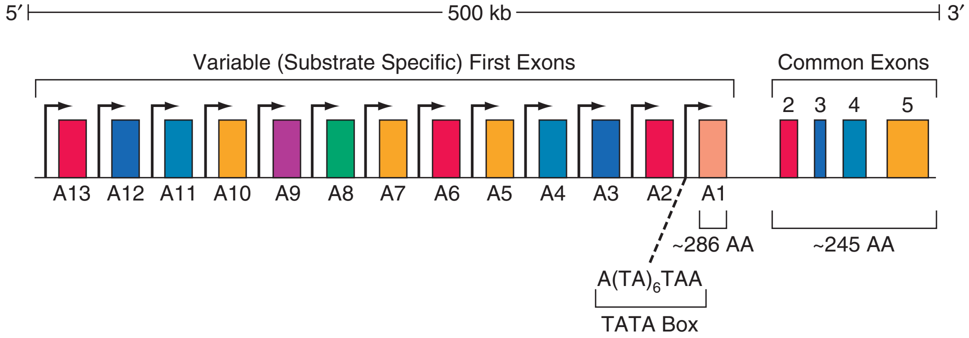
5′
500 kb
3′
Variable (Substrate Specific) First Exons
Common Exons
2 3 4 5
A13 A12 A11 A10 A9 A8 A7 A6 A5 A4 A3 A2 A1
~286 AA
~245 AA
A(TA)6TAA
TATA Box

FIGURE 349-2 Structural organization of the human *UGT1* gene complex. This large complex on chromosome 2 contains at least 13 substrate-specific first exons (A1, A2, etc.). Since four of these are pseudogenes, nine UGT1 isoforms with differing substrate specificities are expressed. Each exon 1 has its own promoter and encodes the amino-terminal substrate-specific ~286 amino acids of the various *UGT1*-encoded isoforms, and common exons 2–5 encode the 245 carboxyl-terminal amino acids common to all of the isoforms. mRNAs for specific isoforms are assembled by splicing a particular first exon such as the bilirubin-specific exon A1 to exons 2 to 5. The resulting message encodes a complete enzyme, in this particular case, bilirubin-UDP-glucuronosyltransferase (UGT1A1). Mutations in a first exon affect only a single isoform. Those in exons 2–5 affect all enzymes encoded by the UGT1 complex.

bilirubin mono- and diglucuronide, respectively. Conjugation disrupts the internal hydrogen bonding that limits aqueous solubility of bilirubin, and the resulting glucuronide conjugates are highly soluble in water. Conjugation is obligatory for excretion of bilirubin across the bile canalicular membrane into bile. The UDP-glucuronosyltransferases have been classified into gene families based on the degree of homology among the mRNAs for the various isoforms. Those that conjugate bilirubin and certain other substrates have been designated the *UGT1* family. These are expressed from a single gene complex by alternative promoter usage. This gene complex contains multiple substrate-specific first exons, designated A1, A2, etc. **(Fig. 349-2)**, each with its own promoter and each encoding the amino-terminal half of a specific isoform. In addition, there are four common exons (exons 2–5) that encode the shared carboxyl-terminal half of all of the *UGT1* isoforms. The various first exons encode the specific aglycone substrate binding sites for each isoform, while the shared exons encode the binding site for the sugar donor, UDP-glucuronic acid, and the transmembrane domain. Exon A1 and the four common exons, collectively designated as the *UGT1A1* gene (Fig. 349-2), encode the physiologically critical enzyme bilirubin-UDP-glucuronosyltransferase (UGT1A1). A functional corollary of the organization of the *UGT1* gene is that a mutation in one of the first exons will affect only a single enzyme isoform. By contrast, a mutation in exons 2–5 will alter all isoforms encoded by the *UGT1* gene complex.

4. *Biliary excretion:* It has been thought until recently that bilirubin mono- and diglucuronides are excreted directly across the canalicular plasma membrane into the bile canaliculus by an ATP-dependent transport process mediated by a canalicular membrane protein called *multidrug resistance–associated protein 2* (MRP2, ABCC2). Mutations of MRP2 result in the Dubin-Johnson syndrome (see below). However, studies in patients with Rotor syndrome (see below) indicate that after formation, a portion of the glucuronides is transported into the portal circulation by a sinusoidal membrane protein called *multidrug resistance–associated protein 3* (MRP3, ABCC3) and is subjected to reuptake into the hepatocyte by the sinusoidal membrane uptake transporters *organic anion transport protein 1B1* (OATP1B1, SLCO1B1) and OATP1B3 (SLCO1B3).

■ EXTRAHEPATIC ASPECTS OF BILIRUBIN DISPOSITION

Bilirubin in the Gut Following secretion into bile, conjugated bilirubin reaches the duodenum and passes down the gastrointestinal tract without reabsorption by the intestinal mucosa. An appreciable fraction is converted by bacterial metabolism in the gut to the water-soluble colorless compound urobilinogen. Urobilinogen undergoes enterohepatic cycling. Urobilinogen not taken up by the liver reaches the systemic circulation, from which some is cleared by the kidneys. Unconjugated bilirubin ordinarily does not reach the gut except in neonates or, by ill-defined alternative pathways, in the presence of severe unconjugated hyperbilirubinemia (e.g., Crigler-Najjar syndrome, type I [CN-I]). Unconjugated bilirubin that reaches the gut is partly reabsorbed, amplifying any underlying hyperbilirubinemia.

Renal Excretion of Bilirubin Conjugates Unconjugated bilirubin is not excreted in urine, as it is too tightly bound to albumin for effective glomerular filtration and there is no tubular mechanism for its renal secretion. In contrast, the bilirubin conjugates are readily filtered at the glomerulus and can appear in urine in disorders characterized by increased bilirubin conjugates in the circulation. It should be kept in mind that the kidney can serve as an "overflow valve" for conjugated bilirubin. Consequently, the level of jaundice in individuals with conjugated hyperbilirubinemia can be amplified in the presence of renal failure.

DISORDERS OF BILIRUBIN METABOLISM LEADING TO UNCONJUGATED HYPERBILIRUBINEMIA

■ INCREASED BILIRUBIN PRODUCTION

Hemolysis Increased destruction of erythrocytes leads to increased bilirubin turnover and unconjugated hyperbilirubinemia; the hyperbilirubinemia is usually modest in the presence of normal liver function. In particular, the bone marrow is only capable of a sustained eightfold increase in erythrocyte production in response to a hemolytic stress. Therefore, hemolysis alone cannot result in a sustained hyperbilirubinemia of more than ~68 μmol/L (4 mg/dL). Higher values imply concomitant hepatic dysfunction. When hemolysis is the only abnormality in an otherwise healthy individual, the result is a purely unconjugated hyperbilirubinemia, with the direct-reacting fraction as measured in a typical clinical laboratory being ≤15% of the total serum bilirubin. In the presence of systemic disease, which may include a degree of hepatic dysfunction, hemolysis may produce a component of conjugated hyperbilirubinemia in addition to an elevated unconjugated bilirubin concentration. Prolonged hemolysis may lead to the precipitation of bilirubin salts within the gallbladder or biliary tree, resulting in the formation of gallstones in which bilirubin, rather than cholesterol, is the major component. Such pigment stones may lead to acute or chronic cholecystitis, biliary obstruction, or any other biliary tract consequence of calculous disease.

Ineffective Erythropoiesis During erythroid maturation, small amounts of hemoglobin may be lost at the time of nuclear extrusion, and a fraction of developing erythroid cells is destroyed within the marrow. These processes normally account for a small proportion of bilirubin that is produced. In various disorders, including thalassemia major, megaloblastic anemias due to folate or vitamin B_{12} deficiency, congenital erythropoietic porphyria, lead poisoning, and various congenital and acquired dyserythropoietic anemias, the fraction of total bilirubin production derived from ineffective erythropoiesis is increased, reaching as much as 70% of the total. This may be sufficient to produce modest degrees of unconjugated hyperbilirubinemia.

Miscellaneous Degradation of the hemoglobin of extravascular collections of erythrocytes, such as those seen in massive tissue infarctions or large hematomas, may lead transiently to unconjugated hyperbilirubinemia.

DECREASED HEPATIC BILIRUBIN CLEARANCE

Decreased Hepatic Uptake Decreased hepatic bilirubin uptake is believed to contribute to the unconjugated hyperbilirubinemia of Gilbert syndrome (GS), although the molecular basis for this finding remains unclear (see below). Several drugs, including flavaspidic acid, novobiocin, and rifampin, as well as various cholecystographic contrast agents, have been reported to inhibit bilirubin uptake. The resulting unconjugated hyperbilirubinemia resolves with cessation of the medication.

Impaired Conjugation • PHYSIOLOGIC NEONATAL JAUNDICE Bilirubin produced by the fetus is cleared by the placenta and eliminated by the maternal liver. Immediately after birth, the neonatal liver must assume responsibility for bilirubin clearance and excretion. However, many hepatic physiologic processes are incompletely developed at birth. Levels of UGT1A1 are low, and alternative excretory pathways allow passage of unconjugated bilirubin into the gut. Since the intestinal flora that convert bilirubin to urobilinogen are also undeveloped, an enterohepatic circulation of unconjugated bilirubin ensues. As a consequence, most neonates develop mild unconjugated hyperbilirubinemia between days 2 and 5 after birth. Peak levels are typically <85–170 μmol/L (5–10 mg/dL) and decline to normal adult concentrations within 2 weeks, as mechanisms required for bilirubin disposition mature. Prematurity, often associated with more profound immaturity of hepatic function and hemolysis, can result in higher levels of unconjugated hyperbilirubinemia. A rapidly rising unconjugated bilirubin concentration, or absolute levels >340 μmol/L (20 mg/dL), puts the infant at risk for bilirubin encephalopathy, or kernicterus. Under these circumstances, bilirubin crosses an immature blood-brain barrier and precipitates in the basal ganglia and other areas of the brain. The consequences range from appreciable neurologic deficits to death. Treatment options include phototherapy, which converts bilirubin into water-soluble photoisomers that are excreted directly into bile, and exchange transfusion. The canalicular mechanisms responsible for bilirubin excretion are also immature at birth, and their maturation may lag behind that of UGT1A1; this can lead to transient conjugated neonatal hyperbilirubinemia, especially in infants with hemolysis.

ACQUIRED CONJUGATION DEFECTS A modest reduction in bilirubin conjugating capacity may be observed in advanced hepatitis or cirrhosis. However, in this setting, conjugation is better preserved than other aspects of bilirubin disposition, such as canalicular excretion. Various drugs, including pregnanediol, novobiocin, chloramphenicol, gentamicin, and atazanavir, may produce unconjugated hyperbilirubinemia by inhibiting UGT1A1 activity. Bilirubin conjugation may be inhibited by certain fatty acids that are present in breast milk, but not serum, of mothers whose infants have excessive neonatal hyperbilirubinemia (*breast milk jaundice*). Alternatively, there may be increased enterohepatic circulation of bilirubin in these infants. The pathogenesis of breast milk jaundice appears to differ from that of transient familial neonatal hyperbilirubinemia (Lucey-Driscoll syndrome), in which there may be a UGT1A1 inhibitor in maternal serum.

HEREDITARY DEFECTS IN BILIRUBIN CONJUGATION

Three familial disorders characterized by differing degrees of unconjugated hyperbilirubinemia have long been recognized. The defining clinical features of each are described below (Table 349-1). While these disorders have been recognized for decades to reflect differing degrees of deficiency in the ability to conjugate bilirubin, recent advances in the molecular biology of the *UGT1* gene complex have elucidated their interrelationships and clarified previously puzzling features.

Crigler-Najjar Syndrome, Type I CN-I is characterized by striking unconjugated hyperbilirubinemia of ~340–765 μmol/L (20–45 mg/dL) that appears in the neonatal period and persists for life. Other conventional hepatic biochemical tests such as serum aminotransferases and alkaline phosphatase are normal, and there is no evidence of hemolysis. Hepatic histology is also essentially normal except for the occasional presence of bile plugs within canaliculi. Bilirubin glucuronides are virtually absent from the bile, and there is no detectable constitutive expression of UGT1A1 activity in hepatic tissue. Neither UGT1A1 activity nor the serum bilirubin concentration responds to administration of phenobarbital or other enzyme inducers. Unconjugated bilirubin accumulates in plasma, from which it is eliminated very slowly by alternative pathways that include direct passage into the bile and small intestine, possibly via bilirubin photoisomers. This accounts for the small amount of urobilinogen found in feces. No bilirubin is found in the urine. First described in 1952, the disorder is rare (estimated prevalence, 0.6–1.0 per million). Many patients are from geographically or socially isolated communities in which consanguinity is common, and pedigree analyses show an autosomal recessive pattern of inheritance. The majority of patients (type IA) exhibit defects in the glucuronide conjugation of a spectrum of substrates in addition to bilirubin, including various drugs and other xenobiotics. These individuals have mutations in one of the common exons (2–5) of the *UGT1* gene (Fig. 349-2). In a smaller subset (type IB), the defect is limited largely to bilirubin conjugation, and the causative mutation is in the bilirubin-specific exon A1. Estrogen glucuronidation is mediated by UGT1A1 and is defective in all CN-I patients. More than 30 different genetic lesions of *UGT1A1* responsible for CN-I have been identified, including deletions, insertions, alterations in intron splice donor and acceptor sites, exon skipping,

TABLE 349-1 Principal Differential Characteristics of Gilbert and Crigler-Najjar Syndromes

	CRIGLER-NAJJAR SYNDROME		
FEATURE	**TYPE I**	**TYPE II**	**GILBERT SYNDROME**
Total serum bilirubin, μmol/L (mg/dL)	310–755 (usually >345) (18–45 [usually >20])	100–430 (usually ≤345) (6–25 [usually ≤20])	Typically ≤70 μmol/L (≤4 mg/dL) in absence of fasting or hemolysis
Routine liver tests	Normal	Normal	Normal
Response to phenobarbital	None	Decreases bilirubin by >25%	Decreases bilirubin to normal
Kernicterus	Usual	Rare	No
Hepatic histology	Normal	Normal	Usually normal; increased lipofuscin pigment in some
Bile characteristics			
Color	Pale or colorless	Pigmented	Normal dark color
Bilirubin fractions	>90% unconjugated	Largest fraction (mean: 57%) monoconjugates	Mainly diconjugates but monoconjugates increased (mean: 23%)
Bilirubin UDP-glucuronosyltransferase activity	Typically absent; traces in some patients	Markedly reduced: 0–10% of normal	Reduced: typically 10–33% of normal
Inheritance (all autosomal)	Recessive	Predominantly recessive	Promoter mutation: recessive Missense mutations: 7 of 8 dominant; 1 reportedly recessive

and point mutations that introduce premature stop codons or alter critical amino acids. Their common feature is that they all encode proteins with absent or, at most, traces of bilirubin-UDP-glucuronosyltransferase enzymatic activity.

Prior to the use of phototherapy, most patients with CN-I died of bilirubin encephalopathy (*kernicterus*) in infancy or early childhood. A few lived as long as early adult life without overt neurologic damage, although more subtle testing usually indicated mild but progressive brain damage, now termed bilirubin-induced neurologic dysfunction (BIND). In the absence of liver transplantation, death eventually supervened from late-onset bilirubin encephalopathy, which often followed a nonspecific febrile illness. Although isolated hepatocyte transplantation has been used in a small number of cases of CN-I, early liver transplantation **(Chap. 356)** remains the best hope to prevent brain injury and death at present. It is anticipated that gene replacement therapy may be an option in the future.

Crigler-Najjar Syndrome, Type II (CN-II) This condition was recognized as a distinct entity in 1962 and is characterized by marked unconjugated hyperbilirubinemia in the absence of abnormalities of other conventional hepatic biochemical tests, hepatic histology, or hemolysis. It differs from CN-I in several specific ways (Table 349-1): (1) although there is considerable overlap, average bilirubin concentrations are lower in CN-II; (2) accordingly, CN-II is only infrequently associated with kernicterus; (3) bile is deeply colored, and bilirubin glucuronides are present, with a striking, characteristic increase in the proportion of monoglucuronides; (4) UGT1A1 in liver is usually present at reduced levels (typically ≤10% of normal); and (5) while typically detected in infancy, hyperbilirubinemia was not recognized in some cases until later in life and, in one instance, at age 34. As with CN-I, most CN-II cases exhibit abnormalities in the conjugation of other compounds, such as salicylamide and menthol, but in some instances, the defect appears limited to bilirubin. Reduction of serum bilirubin concentrations by >25% in response to enzyme inducers such as phenobarbital distinguishes CN-II from CN-I, although this response may not be elicited in early infancy and often is not accompanied by measurable UGT1A1 induction. Bilirubin concentrations during phenobarbital administration do not return to normal but are typically in the range of 51–86 μmol/L (3–5 mg/dL). Although the incidence of kernicterus in CN-II is low, instances have occurred, not only in infants but also in adolescents and adults, often in the setting of an intercurrent illness, fasting, or another factor that temporarily raises the serum bilirubin concentration above baseline and reduces serum albumin levels. For this reason, phenobarbital therapy is widely recommended, a single bedtime dose often sufficing to maintain clinically safe serum bilirubin concentrations.

Over 100 different mutations in the *UGT1* gene have been identified as causing CN-I or CN-II. It was found that missense mutations are more common in CN-II patients, as would be expected in this less severe phenotype. Their common feature is that they encode for a bilirubin-UDP-glucuronosyltransferase with markedly reduced, but detectable, enzymatic activity. The spectrum of residual enzyme activity explains the spectrum of phenotypic severity of the resulting hyperbilirubinemia. Molecular analysis has established that a large majority of CN-II patients are either homozygotes or compound heterozygotes for CN-II mutations and that individuals carrying one mutated and one entirely normal allele have normal bilirubin concentrations.

Gilbert Syndrome This syndrome is characterized by mild unconjugated hyperbilirubinemia, normal values for standard hepatic biochemical tests, and normal hepatic histology other than a modest increase of lipofuscin pigment in some patients. Serum bilirubin concentrations are most often <51 μmol/L (<3 mg/dL), although both higher and lower values are frequent. The clinical spectrum of hyperbilirubinemia fades into that of CN-II at serum bilirubin concentrations of 86–136 μmol/L (5–8 mg/dL). At the other end of the scale, the distinction between mild cases of GS and a normal state is often blurred. Bilirubin concentrations may fluctuate substantially in any given individual, and at least 25% of patients will exhibit temporarily normal values during prolonged follow-up. More elevated values are associated with stress, fatigue, alcohol use, reduced caloric intake, and intercurrent illness, while increased caloric intake or administration of enzyme-inducing agents produces lower bilirubin levels. GS is most often diagnosed at or shortly after puberty or in adult life during routine examinations that include multichannel biochemical analyses. UGT1A1 activity is typically reduced to 10–35% of normal, and bile pigments exhibit a characteristic increase in bilirubin monoglucuronides. Studies of radiobilirubin kinetics indicate that hepatic bilirubin clearance is reduced to an average of one-third of normal. Administration of phenobarbital normalizes both the serum bilirubin concentration and hepatic bilirubin clearance; however, failure of UGT1A1 activity to improve in many such instances suggests the possible coexistence of an additional defect. Compartmental analysis of bilirubin kinetic data suggests that GS patients may have a defect in bilirubin uptake as well as in conjugation, although this has not been shown directly. Defects in the hepatic uptake of other organic anions that at least partially share an uptake mechanism with bilirubin, such as sulfobromophthalein and indocyanine green (ICG), are observed in a minority of patients. The metabolism and transport of bile acids that do not utilize the bilirubin uptake mechanism are normal. The magnitude of changes in the serum bilirubin concentration induced by provocation tests such as 48 h of fasting or the IV administration of nicotinic acid has been reported to be of help in separating GS patients from normal individuals. Other studies dispute this assertion. Moreover, on theoretical grounds, the results of such studies should provide no more information than simple measurements of the baseline serum bilirubin concentration. Family studies indicate that GS and hereditary hemolytic anemias such as hereditary spherocytosis, glucose-6-phosphate dehydrogenase deficiency, and β-thalassemia trait sort independently. Reports of hemolysis in up to 50% of GS patients are believed to reflect better case finding, since patients with both GS and hemolysis have higher bilirubin concentrations and are more likely to be jaundiced than patients with either defect alone.

GS is common, with many series placing its prevalence as high as 8%. Males predominate over females by reported ratios ranging from 1.5:1 to >7:1. However, these ratios may have a large artifactual component since normal males have higher mean bilirubin levels than normal females, but the diagnosis of GS is often based on comparison to normal ranges established in men. The high prevalence of GS in the general population may explain the reported frequency of mild unconjugated hyperbilirubinemia in liver transplant recipients. The disposition of most xenobiotics metabolized by glucuronidation appears to be normal in GS, as is oxidative drug metabolism in the majority of reported studies. The principal exception is the metabolism of the antitumor agent irinotecan (CPT-11), whose active metabolite (SN-38) is glucuronidated specifically by bilirubin-UDP-glucuronosyltransferase. Administration of CPT-11 to patients with GS has resulted in several toxicities, including intractable diarrhea and myelosuppression. Some reports also suggest abnormal disposition of menthol, estradiol benzoate, acetaminophen, tolbutamide, and rifamycin SV. Although some of these studies have been disputed, and there have been no reports of clinical complications from use of these agents in GS, prudence should be exercised in prescribing them or any agents metabolized primarily by glucuronidation in this condition. It should also be noted that the HIV protease inhibitors indinavir and atazanavir **(Chap. 208)** can inhibit UGT1A1, resulting in hyperbilirubinemia that is most pronounced in patients with preexisting GS.

Most older pedigree studies of GS were consistent with autosomal dominant inheritance with variable expressivity. However, studies of the *UGT1* gene in GS have indicated a variety of molecular genetic bases for the phenotypic picture and several different patterns of inheritance. Studies in Europe and the United States found that nearly all patients had normal coding regions for UGT1A1 but were homozygous for the insertion of an extra TA (i.e., A[TA]$_7$TAA rather than A[TA]$_6$TAA) in the promoter region of the first exon. This appeared to be necessary, but not sufficient, for clinically expressed GS, since 15% of normal controls were also homozygous for this variant. While normal by standard criteria, these individuals had somewhat higher bilirubin concentrations than the rest of the controls studied. Heterozygotes for this abnormality had bilirubin concentrations identical to those homozygous for the normal A[TA]$_6$TAA allele. The prevalence of the A[TA]$_7$TAA allele in a general Western population is 30%, in which case 9% would be homozygotes. This is slightly higher than the prevalence of GS based on purely phenotypic parameters. It was suggested that additional variables, such as mild hemolysis or a defect in bilirubin uptake, might be among the factors enhancing phenotypic expression of the defect.

Phenotypic expression of GS due solely to the A[TA]$_7$TAA promoter abnormality is inherited as an autosomal recessive trait. A number of CN-II kindreds have been identified in whom there is also an allele containing a normal coding region but the A[TA]$_7$TAA promoter abnormality. CN-II heterozygotes, who have the A[TA]$_6$TAA promoter, are phenotypically normal, whereas those with the A[TA]$_7$TAA promoter express the phenotypic picture of GS. GS in such kindreds may also result from homozygosity for the A[TA]$_7$TAA promoter abnormality. Seven different missense mutations in the *UGT1* gene that reportedly cause GS with dominant inheritance have been found in Japanese individuals. Another Japanese patient with mild unconjugated hyperbilirubinemia was homozygous for a missense mutation in exon 5. GS in her family appeared to be recessive.

DISORDERS OF BILIRUBIN METABOLISM LEADING TO MIXED OR PREDOMINANTLY CONJUGATED HYPERBILIRUBINEMIA

In hyperbilirubinemia due to acquired liver disease (e.g., acute hepatitis, common bile duct stone), there are usually elevations in the serum concentrations of both conjugated and unconjugated bilirubin. Although biliary tract obstruction or hepatocellular cholestatic injury may present on occasion with a predominantly conjugated hyperbilirubinemia, it is generally not possible to differentiate intrahepatic from extrahepatic causes of jaundice based on the serum levels or relative proportions of unconjugated and conjugated bilirubin. The major reason for determining the amounts of conjugated and unconjugated bilirubin in the serum is for the initial differentiation of hepatic parenchymal and obstructive disorders (mixed conjugated and unconjugated hyperbilirubinemia) from the inheritable and hemolytic disorders discussed above that are associated with unconjugated hyperbilirubinemia.

■ FAMILIAL DEFECTS IN HEPATIC EXCRETORY FUNCTION

Dubin-Johnson Syndrome (DJS) This benign, relatively rare disorder is characterized by low-grade, predominantly conjugated hyperbilirubinemia (Table 349-2). Total bilirubin concentrations are typically between 34 and 85 μmol/L (2 and 5 mg/dL) but on occasion can be in the normal range or as high as 340–430 μmol/L (20–25 mg/dL) and can fluctuate widely in any given patient. The degree of hyperbilirubinemia may be increased by intercurrent illness, oral contraceptive use, and pregnancy. Because the hyperbilirubinemia is due to a predominant rise in conjugated bilirubin, bilirubinuria is characteristically present. Aside from elevated serum bilirubin levels, other routine laboratory tests are normal. Physical examination is usually normal except for jaundice, although an occasional patient may have hepatosplenomegaly.

Patients with DJS are usually asymptomatic, although some may have vague constitutional symptoms. These latter patients have usually undergone extensive diagnostic examinations for unexplained jaundice and have high levels of anxiety. In women, the condition may be subclinical until the patient becomes pregnant or receives oral contraceptives, at which time chemical hyperbilirubinemia becomes frank jaundice. Even in these situations, other routine liver function tests, including serum alkaline phosphatase and transaminase activities, are normal.

A cardinal feature of DJS is the accumulation of dark, coarsely granular pigment in the lysosomes of centrilobular hepatocytes. As a result, the liver may be grossly black in appearance. This pigment is thought to be derived from epinephrine metabolites that are not excreted normally. The pigment may disappear during bouts of viral hepatitis, only to reaccumulate slowly after recovery.

Biliary excretion of a number of anionic compounds is compromised in DJS. These include various cholecystographic agents, as well as sulfobromophthalein (Bromsulphalein [BSP]), a synthetic dye formerly used in a test of liver function. In this test, the rate of disappearance of BSP from plasma was determined following bolus IV administration. BSP is conjugated with glutathione in the hepatocyte; the resulting conjugate is normally excreted rapidly into the bile canaliculus. Patients with DJS exhibit characteristic rises in plasma concentrations at 90 min after injection, due to reflux of conjugated BSP into the circulation from the hepatocyte. Dyes such as ICG that are taken up by hepatocytes but are not further metabolized prior to biliary excretion do not show this reflux phenomenon. Continuous BSP infusion studies suggest a reduction in the time to maximum plasma concentration (t_{max}) for biliary excretion. Bile acid disposition, including hepatocellular uptake and biliary excretion, is normal in DJS. These patients have normal serum and biliary bile acid concentrations and do not have pruritus.

By analogy with findings in several mutant rat strains, the selective defect in biliary excretion of bilirubin conjugates and certain other classes of organic compounds, but not of bile acids, that characterizes DJS in humans was found to reflect defective expression of MRP2 (ABCC2), an ATP-dependent canalicular membrane transporter. Several different mutations in the *ABCC2* gene produce the Dubin-Johnson phenotype, which has an autosomal recessive pattern of inheritance. Although MRP2 is undoubtedly important in the biliary excretion of

TABLE 349-2 Principal Differential Characteristics of Inheritable Disorders of Bile Canalicular Function

	DJS	ROTOR	PFIC1	BRIC1	PFIC2	BRIC2	PFIC3
Gene	*ABCCA*	*SLCO1B1/SLCO1B3*	*ATP8B1*	*ATP8B1*	*ABCB11*	*ABCB11*	*ABCB4*
Protein	MRP2	OATP1B1/1B3	FIC1	FIC1	BSEP	BSEP	MDR3
Cholestasis	No	No	Yes	Episodic	Yes	Episodic	Yes
Serum GGT	Normal	Normal	Normal	Normal	Normal	Normal	↑↑
Serum bile acids	Normal	Normal	↑↑	↑↑ during episodes	↑↑	↑↑ during episodes	↑↑
Clinical features	Mild conjugated hyperbilirubinemia; otherwise, normal liver function; dark pigment in liver; characteristic pattern of urinary coproporphyrins	Mild conjugated hyperbilirubinemia; otherwise, normal liver function; liver without abnormal pigmentation	Severe cholestasis beginning in childhood	Recurrent episodes of cholestasis beginning at any age	Severe cholestasis beginning in childhood	Recurrent episodes of cholestasis beginning at any age	Severe cholestasis beginning in childhood; decreased phospholipids in bile

Abbreviations: BRIC, benign recurrent intrahepatic cholestasis; BSEP, bile salt excretory protein; DJS, Dubin-Johnson syndrome; GGT, γ-glutamyl transferase; MRP2, multidrug resistance–associated protein 2; OATP1A/1B, organic anion transport proteins 1B1 and 1B3; PFIC, progressive familial intrahepatic cholestasis; ↑↑, increased.

conjugated bilirubin, the fact that this pigment is still excreted in the absence of MRP2 suggests that other, as yet uncharacterized, transport proteins may serve in a secondary role in this process.

Patients with DJS also have a diagnostic abnormality in urinary coproporphyrin excretion. There are two naturally occurring coproporphyrin isomers, I and III. Normally, ~75% of the coproporphyrin in urine is isomer III. In urine from DJS patients, total coproporphyrin content is normal, but ≥80% is isomer I. Heterozygotes for the syndrome show an intermediate pattern. The molecular basis for this phenomenon remains unclear.

Rotor Syndrome (RS) This benign, autosomal recessive disorder is clinically similar to DJS (Table 349-2), although it is seen even less frequently. A major phenotypic difference is that the liver in patients with RS has no increased pigmentation and appears totally normal. The only abnormality in routine laboratory tests is an elevation of total serum bilirubin, due to a predominant rise in conjugated bilirubin. This is accompanied by bilirubinuria. Several additional features differentiate RS from DJS. In RS, the gallbladder is usually visualized on oral cholecystography, in contrast to the nonvisualization that is typical of DJS. The pattern of urinary coproporphyrin excretion also differs. The pattern in RS resembles that of many acquired disorders of hepatobiliary function, in which coproporphyrin I, the major coproporphyrin isomer in bile, refluxes from the hepatocyte back into the circulation and is excreted in urine. Thus, total urinary coproporphyrin excretion is substantially increased in RS, in contrast to the normal levels seen in DJS. Although the fraction of coproporphyrin I in urine is elevated, it is usually <70% of the total, compared with ≥80% in DJS. The disorders also can be distinguished by their patterns of BSP excretion. Although clearance of BSP from plasma is delayed in RS, there is no reflux of conjugated BSP back into the circulation as seen in DJS. Kinetic analysis of plasma BSP infusion studies suggests the presence of a defect in intrahepatocellular storage of this compound. This has never been demonstrated directly. Recent studies indicate that the molecular basis of RS results from simultaneous deficiency of the hepatocyte plasma membrane transporters OATP1B1 (SLCO1B1) and OATP1B3 (SLCO1B3). This results in reduced reuptake by these transporters of conjugated bilirubin that has been pumped out of the hepatocyte into the portal circulation by MRP3 (ABCC3) (Fig. 349-1). It should be noted that these transporters normally mediate uptake into the liver of a variety of drugs such as statins, and the possibility of drug toxicity resulting from their deficiency in RS must be considered.

Benign Recurrent Intrahepatic Cholestasis (BRIC) This rare disorder is characterized by recurrent attacks of pruritus and jaundice. The typical episode begins with mild malaise and elevations in serum aminotransferase levels, followed rapidly by rises in alkaline phosphatase and conjugated bilirubin and onset of jaundice and itching. The first one or two episodes may be misdiagnosed as acute viral hepatitis. The cholestatic episodes, which may begin in childhood or adulthood, can vary in duration from several weeks to months, followed by a complete clinical and biochemical resolution. Intervals between attacks may vary from several months to years. Between episodes, physical examination is normal, as are serum levels of bile acids, bilirubin, transaminases, and alkaline phosphatase. The disorder is familial and has an autosomal recessive pattern of inheritance. BRIC is considered a benign disorder in that it does not lead to cirrhosis or end-stage liver disease. However, the episodes of jaundice and pruritus can be prolonged and debilitating, and some patients have undergone liver transplantation to relieve the intractable and disabling symptoms. Treatment during the cholestatic episodes is symptomatic; there is no specific treatment to prevent or shorten the occurrence of episodes.

A gene termed *FIC1* was recently identified and found to be mutated in patients with BRIC. Curiously, this gene is expressed strongly in the small intestine but only weakly in the liver. The protein encoded by *FIC1* shows little similarity to those that have been shown to play a role in bile canalicular excretion of various compounds. Rather, it appears to be a member of a P-type ATPase family that transports aminophospholipids from the outer to the inner leaflet of a variety of cell membranes. Its relationship to the pathobiology of this disorder remains unclear. A second phenotypically identical form of BRIC, termed BRIC type 2, has been described resulting from mutations in the bile salt excretory protein (BSEP), the protein that is defective in progressive familial intrahepatic cholestasis (PFIC) type 2 (Table 349-2). How some mutations in this protein result in the episodic BRIC phenotype is unknown.

Progressive Familial Intrahepatic Cholestasis This name is applied to three phenotypically related syndromes (Table 349-2). PFIC type 1 (Byler's disease) presents in early infancy as cholestasis that may be initially episodic. However, in contrast to BRIC, Byler's disease progresses to malnutrition, growth retardation, and end-stage liver disease during childhood. This disorder is also a consequence of an *FIC1* mutation. The functional relationship of the FIC1 protein to the pathogenesis of cholestasis in these disorders is unknown. Two other types of PFIC (types 2 and 3) have been described. PFIC type 2 is associated with a mutation in the protein originally named *sister of P-glycoprotein*, now known as *bile salt excretory protein* (BSEP, ABCB11), which is the major bile canalicular exporter of bile acids. As noted above, some mutations of this protein are associated with BRIC type 2, rather than the PFIC type 2 phenotype. PFIC type 3 has been associated with a mutation of MDR3 (ABCB4), a protein that is essential for normal hepatocellular excretion of phospholipids across the bile canaliculus. Although all three types of PFIC have similar clinical phenotypes, only type 3 is associated with high serum levels of γ-glutamyl transferase (GGT) activity. In contrast, activity of this enzyme is normal or only mildly elevated in symptomatic BRIC and PFIC types 1 and 2. Interestingly, mutations in FIC1 or BSEP are not found in approximately one-third of patients with clinical PFIC and normal GGT. Recent studies have shown that patients with mutations in *NR1H4*, the gene encoding the farnesoid X receptor (FXR), a nuclear hormone receptor activated by bile acids, have a syndrome identical to PFIC2 with absent expression of BSEP. Mutations in tight junction protein 2 (TJP2) as well as ubiquitin-specific protease 53 (USP53), a protein that interacts with TJP2, have also been associated with severe cholestasis with normal GGT levels, likely due to disruption of tight junctions at the bile canaliculus.

FURTHER READING

Bull LN, Thompson RJ: Progressive familial intrahepatic cholestasis. Clin Liver Dis 22:657, 2018.

Canu G et al: Gilbert and Crigler Najjar syndromes: An update of the UDP-glucuronosyltransferase 1A1 (UGT1A1) gene mutation database. Blood Cells Mol Dis 50:273, 2013.

Hansen TW: Biology of bilirubin photoisomers. Clin Perinatol 43:277, 2016.

Hassan S, Hertel P: Overview of progressive familial intrahepatic cholestasis. Clin Liver Dis 26:371, 2022.

Kavallar AM et al: Management and outcomes after liver transplantation for progressive familial intrahepatic cholestasis: A systematic review and meta-analysis. Hepatology Communications 7:e0286, 2023.

Lamola AA: A pharmacologic view of phototherapy. Clin Perinatol 43:259, 2016.

Memon N et al: Inherited disorders of bilirubin clearance. Pediatr Res 79:378, 2016.

Soroka CJ, Boyer JL: Biosynthesis and trafficking of the bile salt export pump, BSEP: Therapeutic implications of BSEP mutations. Mol Aspects Med 37:3, 2014.

van de Steeg E et al: Complete OATP1B1 and OATP1B3 deficiency causes human Rotor syndrome by interrupting conjugated bilirubin reuptake into the liver. J Clin Invest 122:519, 2012.

van Wessel DBE et al: Genotype correlates with the natural history of severe bile salt export pump deficiency. J Hepatol 73:84, 2020.

Wolkoff AW: Organic anion uptake by hepatocytes. Compr Physiol 4:1715, 2014.

350 Acute Viral Hepatitis

Esperance A. K. Schaefer, Raymond T. Chung, Jules L. Dienstag

Acute viral hepatitis is a systemic infection affecting the liver predominantly. Almost all cases of acute viral hepatitis are caused by one of five viral agents: hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), the HBV-associated delta agent or hepatitis D virus (HDV), and hepatitis E virus (HEV). All these human hepatitis viruses are RNA viruses, except for hepatitis B, which is a DNA virus but replicates like a retrovirus. Although these agents can be distinguished by their molecular and antigenic properties, all types of viral hepatitis produce clinically similar acute illnesses. These range from asymptomatic and inapparent to fulminant and fatal acute infections, which can be observed in all types. On the other hand, the bloodborne types (HBV, HCV, HDV) may also manifest a spectrum of chronic disease, from subclinical, persistent infections to rapidly progressive chronic liver disease with cirrhosis and even hepatocellular carcinoma. Rarely, infections with other viruses (e.g., cytomegalovirus, Epstein-Barr virus, other herpes viruses, SARS-CoV-2) can be associated with mild or severe, even fulminant, hepatitis, more often in immunocompromised hosts but also in otherwise healthy persons. In 2021 and 2022, after the peak years of the COVID-19 pandemic, multiple reports appeared globally of severe acute, often fulminant, hepatitis in previously healthy children. While these instances of severe acute hepatitis remain unexplained, links have been reported of infection with adenovirus type 41 and adeno-associated virus 2, ordinarily not causes of liver injury. A reduction in circulating respiratory virus infections during the isolation of the COVID-19 pandemic has been postulated to have accounted for the increased susceptibility to and severity of the acute hepatitis associated with these nonhepatotropic viruses.

VIROLOGY AND ETIOLOGY

Hepatitis A HAV is a nonenveloped 27-nm, heat-, acid-, and ether-resistant, single-stranded, positive-sense RNA virus in the *Hepatovirus* genus of the picornavirus family (Fig. 350-1). Quasi-enveloped virus particles encased in host plasma membrane–derived membranous vesicles circulate in the bloodstream. The virion contains four structural capsid polypeptides, designated VP1–VP4, as well as six nonstructural proteins, which are cleaved posttranslationally from the polyprotein product of a 7500-nucleotide genome. Inactivation of viral activity can be achieved by boiling for 1 min, by contact with formaldehyde and chlorine, or by ultraviolet irradiation. Despite nucleotide sequence variation of up to 20% among isolates of HAV and despite the recognition of six genotypes (three of which affect humans), all strains of this virus are immunologically indistinguishable and belong to one serotype. Human HAV can infect and cause hepatitis in chimpanzees, tamarins (marmosets), and several monkey species. HAV-like hepatoviruses have also been identified in small mammals, including bats and rodents.

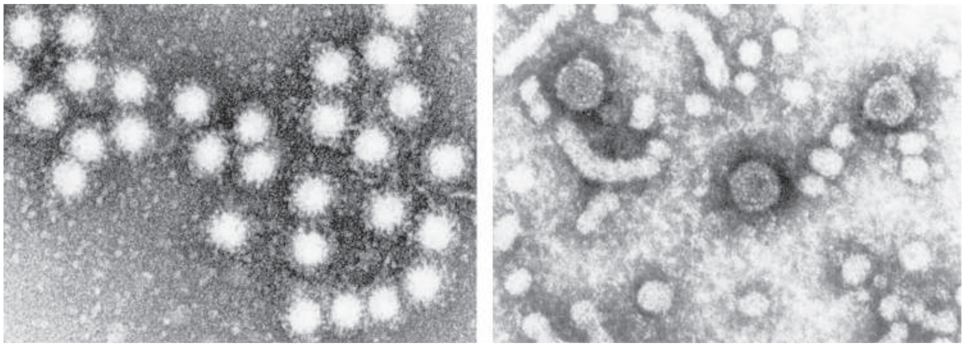

FIGURE 350-1 Electron micrographs of hepatitis A virus particles and serum from a patient with hepatitis B. ***Left:*** 27-nm hepatitis A virus particles purified from stool of a patient with acute hepatitis A and aggregated by antibody to hepatitis A virus. ***Right:*** Concentrated serum from a patient with hepatitis B, demonstrating the 42-nm virions, tubular forms, and spherical 22-nm particles of hepatitis B surface antigen. 132,000×. (Hepatitis D resembles 42-nm virions of hepatitis B but is smaller, 35–37 nm; hepatitis E resembles hepatitis A virus but is slightly larger, 32–34 nm; hepatitis C has been visualized as a 55-nm particle.)

Hepatitis A has an incubation period of ~3–4 weeks. Its replication is limited to the liver, but the virus is present in the liver, bile, stools, and blood during the late incubation period and acute preicteric/presymptomatic phase of illness. Despite slightly longer persistence of virus in the liver, fecal shedding, viremia, and infectivity diminish rapidly once jaundice becomes apparent. Detection of HAV RNA by sensitive reverse transcription polymerase chain reaction assays has been reported to persist at low levels in stool, the liver, and serum for up to several months after acute illness; however, this does not correlate with persistent infectivity, probably because of the presence of neutralizing antibody. HAV can be cultivated reproducibly in vitro and in primate models.

Antibodies to HAV (anti-HAV) can be detected during acute illness when serum aminotransferase activity is elevated and fecal HAV shedding is still occurring. This early antibody response is predominantly of the IgM class and persists for several (~3) months, rarely for 6–12 months. During convalescence, however, anti-HAV of the IgG class becomes the predominant antibody (Fig. 350-2). Therefore, the diagnosis of hepatitis A is made during acute illness by demonstrating anti-HAV of the IgM class. After acute illness, anti-HAV of the IgG class remains detectable indefinitely, and patients with serum anti-HAV are immune to reinfection. Neutralizing antibody activity parallels the appearance of anti-HAV, and the IgG anti-HAV present in immune globulin accounts for the protection it affords against HAV infection.

Hepatitis B HBV is a DNA virus with a remarkably compact genomic structure; despite its small, circular, 3200-bp size, HBV DNA codes for four sets of viral products with a complex, multiparticle structure. HBV achieves its genomic economy by relying on an efficient strategy of encoding proteins from four overlapping genes: S, C, P, and X (Fig. 350-3), as detailed below. Once thought to be unique among viruses, HBV is now recognized as one of a family of animal viruses, hepadnaviruses (hepatotropic DNA viruses), and is classified as hepadnavirus type 1. Similar viruses infect certain species of woodchucks, ground and tree squirrels, and Pekin ducks, to mention the most carefully characterized; genetic evidence of ancient HBV-like virus forbears has been found in fossils of ancient birds, and an HBV-like virus has been identified in contemporary fish. Studies of ancient HBV genomes date an association between HBV and human beings back as long as 21,000 years ago; primate HBV-like viruses date back millions of years, suggesting that HBV predated the emergence of modern humans. Like HBV, all have the same distinctive three morphologic forms, have counterparts to the envelope and nucleocapsid virus antigens of HBV, replicate in the liver but exist in extrahepatic sites, contain their own endogenous DNA polymerase, have partially double-strand and partially single-strand genomes, are associated with acute and chronic hepatitis and hepatocellular carcinoma, and rely on a replicative strategy unique among DNA viruses but typical of retroviruses. Entry of HBV into hepatocytes is mediated by binding to the sodium taurocholate cotransporting polypeptide (NTCP) receptor. Instead of DNA replication directly from a DNA template, hepadnaviruses rely on reverse transcription (effected by the DNA polymerase) of minus-strand DNA from a "pregenomic" RNA intermediate. Then, plus-strand DNA is transcribed from the minus-strand DNA template by the DNA-dependent DNA polymerase and converted in the hepatocyte nucleus to a covalently closed circular DNA, which serves as a template for messenger RNA and pregenomic RNA. Viral proteins

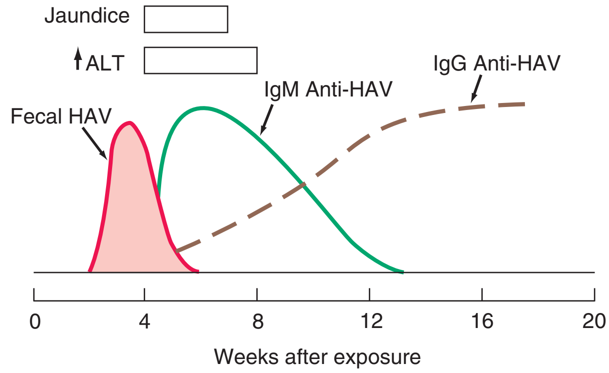

FIGURE 350-2 Scheme of typical clinical and laboratory features of hepatitis A virus (HAV). ALT, alanine aminotransferase.

are translated by the messenger RNA, and the proteins and genome are packaged into virions and secreted from the hepatocyte. Hepatitis B virus has been difficult to cultivate in vitro from clinical materials; therefore, a model that recapitulates the entire HBV life cycle has been particularly elusive. In recent decades, however, advances in molecular virology have permitted the comprehensive study of HBV replication and its viral genes and proteins.

VIRAL PROTEINS AND PARTICLES Three particulate forms of HBV appear in the circulation (Table 350-1), the complete virion and two incomplete or subviral particles. Of these, the most numerous are the 22-nm particles, which appear as spherical or long filamentous forms; these are antigenically indistinguishable from the outer surface or envelope protein of HBV and are thought to represent excess viral envelope protein. Outnumbered in serum by a factor of 100 or 1000 to 1 compared with the spheres and tubules are large, 42-nm, double-shelled spherical particles, which represent the intact hepatitis B virion (Fig. 350-1). The envelope protein expressed on the outer surface of the virion and on the smaller spherical and tubular structures is referred to as *hepatitis B surface antigen* (HBsAg). The concentration of HBsAg and virus particles in the blood may reach 500 μg/mL and 10 trillion particles per milliliter, respectively; HBsAg assays in common clinical use detect all forms of HBsAg (virion and subviral particles). The envelope protein, HBsAg, is the product of the S gene of HBV.

HBV isolates fall into 10 different genotypes (A–J) and multiple subtypes, which are determined by unique gene sequences of the envelope HBsAg protein. Envelope HBsAg subdeterminants include a common group-reactive antigen, *a*, shared by all HBsAg isolates and one of several subtype-specific antigens—*d* or *y*, *w* or *r*. Geographic distribution of genotypes and subtypes varies; genotypes A (corresponding to subtype *adw*) and D (*ayw*) predominate in the United States and Europe, whereas genotypes B (*adw*) and C (*adr*) predominate in Asia; however, these geographic distinctions have been blunted by recent-decade migration across continents. Clinical course and outcome are independent of subtype, but genotype B appears to be associated with less rapidly progressive liver disease and cirrhosis and a lower likelihood, or delayed appearance, of hepatocellular carcinoma than genotype C, D, or F. In a large Japanese cohort, after acute infection, patients with genotype A were more likely to have persistent infection (23.4% with genotype A vs 8.6% with non-A genotypes). Also, genotype may influence response to treatment with interferon and other antiviral agents; for example, higher rates of HBeAg and HBsAg loss have been observed in patients with genotype A. An additional important consequence of genotype is the propensity for "precore" mutations to emerge (see below).

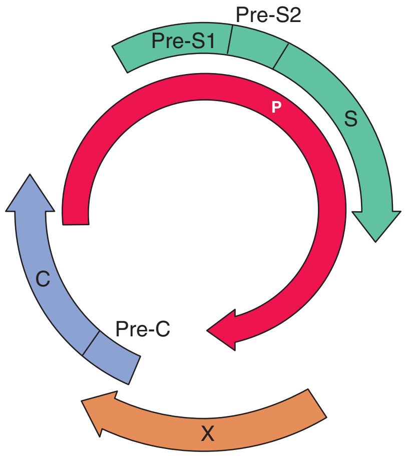

FIGURE 350-3 Compact genomic structure of hepatitis B virus (HBV). This structure, with overlapping genes, permits HBV to code for multiple proteins. The S gene codes for the "major" envelope protein, HBsAg. Pre-S1 and pre-S2, upstream of S, combine with S to code for two larger proteins, "middle" protein, the product of pre-S2 + S, and "large" protein, the product of pre-S1 + pre-S2 + S. The largest gene, P, codes for DNA polymerase. The C gene codes for two nucleocapsid proteins, HBeAg, a soluble, secreted protein (initiation from the pre-C region of the gene), and HBcAg, the intracellular core protein (initiation after pre-C). The X gene codes for HBxAg, which can transactivate the transcription of cellular and viral genes; its clinical relevance is not known, but it may contribute to carcinogenesis by binding to p53.

Upstream of the S gene are the pre-S genes (Fig. 350-3), which code for pre-S gene products, including receptors on the HBV surface for polymerized human serum albumin and for hepatocyte membrane proteins. The pre-S region actually consists of both pre-S1 and pre-S2. Depending on where translation is initiated, three potential HBsAg gene products are synthesized. The protein product of the S gene is HBsAg (*major protein*), the product of the S region plus the adjacent pre-S2 region is the *middle protein*, and the product of the pre-S1 plus pre-S2 plus S regions is the *large protein*. Compared with the smaller spherical and tubular particles of HBV, complete 42-nm virions are enriched in the large protein. Both pre-S proteins and their respective antibodies can be detected during HBV infection, and the period of pre-S antigenemia appears to coincide with other markers of virus replication, as detailed below; however, pre-S proteins have little clinical relevance and are not included in routine serologic testing repertoires.

The intact 42-nm virion contains a 27-nm nucleocapsid core particle. Nucleocapsid proteins are coded for by the C gene. The antigen expressed on the surface of the nucleocapsid core is *hepatitis B core antigen* (HBcAg), and its corresponding antibody is anti-HBc. A third HBV antigen is HBeAg, a soluble, nonparticulate, nucleocapsid protein that is immunologically distinct from intact HBcAg but is a product of the same C gene. The C gene has two initiation codons: a precore and a core region (Fig. 350-3). If translation is initiated at the precore region, the protein product is HBeAg, which has a signal peptide that binds it to the smooth endoplasmic reticulum, the secretory apparatus of the cell, leading to its secretion into the circulation. If translation begins at the core region, HBcAg is the protein product; it has no signal peptide and is not secreted, but it assembles into nucleocapsid particles, which bind to and incorporate RNA and which, ultimately, contain HBV DNA. Also packaged within the nucleocapsid core is a DNA polymerase, which directs replication and repair of HBV DNA. When packaging within viral proteins is complete, synthesis of the incomplete plus strand stops; this accounts for the single-strand gap and for differences in the size of the gap. HBcAg particles remain in the hepatocyte, where they are readily detectable by immunohistochemical staining and are exported after encapsidation by an envelope of HBsAg. Therefore, naked core particles do not circulate in the serum. The secreted nucleocapsid protein, HBeAg, provides a convenient, readily detectable, qualitative marker of HBV replication and relative infectivity.

HBsAg-positive serum containing HBeAg is more likely to be highly infectious and to be associated with the presence of hepatitis B virions (and detectable HBV DNA, see below) than HBeAg-negative or anti-HBe-positive serum. For example, HBsAg-positive mothers who are HBeAg positive almost invariably (>90%) transmit hepatitis B infection to their offspring, whereas HBsAg-positive mothers with anti-HBe rarely (10–15%) infect their offspring.

Early during the course of acute hepatitis B, HBeAg appears transiently; its disappearance may be a harbinger of clinical improvement and resolution of infection. Persistence of HBeAg in serum beyond the first 3 months of acute infection may be predictive of the development

TABLE 350-1 Nomenclature and Features of Hepatitis Viruses

HEPATITIS TYPE	VIRUS PARTICLE, nm	MORPHOLOGY	GENOME[a]	CLASSIFICATION	ANTIGEN(S)	ANTIBODIES	REMARKS
HAV	27	Icosahedral nonenveloped	7.5-kb RNA, linear, ss, +	*Hepatovirus*	HAV	Anti-HAV	Early fecal shedding Diagnosis: IgM anti-HAV Previous infection: IgG anti-HAV
HBV	42	Double-shelled virion (surface and core) spherical	3.2-kb DNA, circular, ss/ds	Hepadnavirus	HBsAg HBcAg HBeAg	Anti-HBs Anti-HBc Anti-HBe	Bloodborne virus; carrier state Acute diagnosis: HBsAg, IgM anti-HBc Chronic diagnosis: IgG anti-HBc, HBsAg Markers of replication: HBeAg, HBV DNA
	27	Nucleocapsid core			HBcAg HBeAg	Anti-HBc Anti-HBe	Liver, lymphocytes, other organs Nucleocapsid contains DNA and DNA polymerase; present in hepatocyte nucleus; HBcAg does not circulate; HBeAg (soluble, nonparticulate) and HBV DNA circulate—correlate with infectivity and complete virions
	22	Spherical and filamentous; represents excess virus coat material			HBsAg	Anti-HBs	HBsAg detectable in >95% of patients with acute hepatitis B; found in serum, body fluids, hepatocyte cytoplasm; anti-HBs appears following infection—protective antibody
HCV	55	Enveloped	9.4-kb RNA, linear, ss, +	*Hepacivirus*	HCV core antigen	Anti-HCV	Bloodborne agent, formerly labeled non-A, non-B hepatitis Acute diagnosis: anti-HCV, HCV RNA Chronic diagnosis: anti-HCV, HCV RNA; cytoplasmic location in hepatocytes
HDV	35–37	Enveloped hybrid particle with HBsAg coat and HDV core	1.7-kb RNA, circular, ss, –	Resembles viroids and plant satellite viruses (genus *Deltavirus*)	HBsAg HDAg	Anti-HBs Anti-HDV	Defective RNA virus, requires helper function of HBV (hepadnaviruses); HDV antigen (HDAg) present in hepatocyte nucleus Diagnosis: anti-HDV, HDV RNA; HBV/HDV co-infection—IgM anti-HBc and anti-HDV; HDV superinfection—IgG anti-HBc and anti-HDV
HEV	32–34	Nonenveloped icosahedral	7.6-kb RNA, linear, ss, +	*Orthohepevirus*	HEV antigen	Anti-HEV	Agent of enterically transmitted hepatitis; rare in the United States; occurs in Asia, Mediterranean countries, Central America Diagnosis: IgM/IgG anti-HEV (assays not routinely available); virus in stool, bile, hepatocyte cytoplasm

[a] ss, single-strand; ss/ds, partially single-strand, partially double-strand; –, minus-strand; +, plus-strand.

Note: See text for abbreviations.

of chronic infection, and the presence of HBeAg during chronic hepatitis B tends to be associated with ongoing viral replication, infectivity, and inflammatory liver injury (except during the early decades after perinatally acquired HBV infection; see below).

The third and largest of the HBV genes, the P gene (Fig. 350-3), codes for HBV DNA polymerase; as noted above, this enzyme has both DNA-dependent DNA polymerase and RNA-dependent reverse transcriptase activities. The fourth gene, X, codes for a small, nonparticulate protein, *hepatitis B x antigen* (HBxAg), which is capable of transactivating the transcription of both viral and cellular genes (Fig. 350-3). This function underpins its role in promoting both viral replication and HBV-related hepatocellular carcinoma. In the cytoplasm, HBxAg effects calcium release (possibly from mitochondria), which activates signal-transduction pathways that lead to stimulation of HBV reverse transcription and HBV DNA replication. Such transactivation may enhance the replication of HBV; the transactivating activity can enhance the transcription and replication of other viruses besides HBV, such as HIV. The clinical relevance of HBxAg is limited, however, and testing for it is not part of routine clinical practice.

SEROLOGIC AND VIROLOGIC MARKERS After a person is infected with HBV, the first virologic marker detectable in serum within 1–12 weeks, usually between 8 and 12 weeks, is HBsAg **(Fig. 350-4)**. Circulating HBsAg precedes elevations of serum aminotransferase activity and clinical symptoms by 2–6 weeks and remains detectable during the entire icteric or symptomatic phase of acute hepatitis B and beyond. In typical self-limited cases, HBsAg becomes undetectable 1–2 months after the onset of jaundice and rarely persists beyond 6 months. After HBsAg disappears, antibody to HBsAg (anti-HBs) becomes detectable in serum and remains detectable indefinitely thereafter. Because HBcAg is intracellular and, when in the serum, sequestered within an HBsAg coat, naked core particles do not circulate in serum, and therefore, assays for HBcAg are not included in routine clinical testing. By contrast, anti-HBc is readily demonstrable in serum, beginning within the first 1–2 weeks after the appearance of HBsAg and preceding detectable levels of anti-HBs by weeks to months. Because variability exists in the time of appearance of anti-HBs after HBV

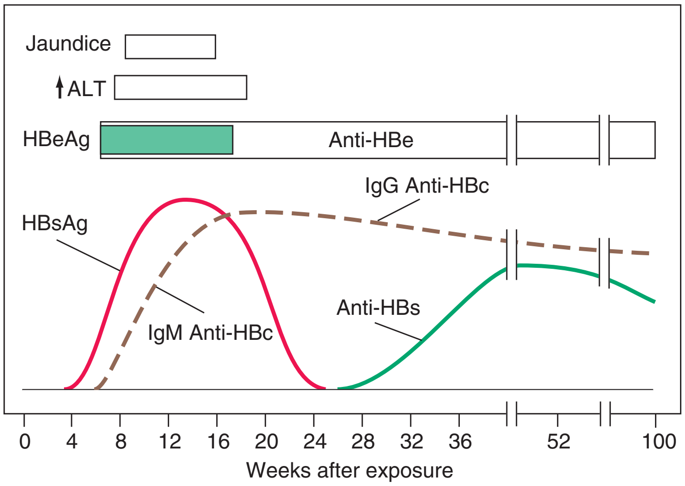

FIGURE 350-4 Scheme of typical clinical and laboratory features of acute hepatitis B. ALT, alanine aminotransferase.

infection, occasionally a gap of several weeks or longer may separate the disappearance of HBsAg and the appearance of anti-HBs. During this "gap" or "window" period, IgM anti-HBc may represent the only serologic evidence of current or recent HBV infection, and blood containing anti-HBc in the absence of HBsAg and anti-HBs had been implicated in the past in transfusion-associated hepatitis B. In part because the sensitivity of immunoassays for HBsAg and anti-HBs has increased, however, this window period is rarely encountered. In some persons, years after HBV infection, anti-HBc may persist in the circulation longer than anti-HBs. Therefore, isolated anti-HBc does not necessarily indicate active virus replication; most instances of isolated anti-HBc represent hepatitis B infection in the remote past. Rarely, however, isolated anti-HBc represents low-level hepatitis B viremia, with HBsAg below the detection threshold, and occasionally, isolated anti-HBc represents a cross-reacting or false-positive immunologic specificity.

Recent and remote HBV infections can be distinguished by determination of the immunoglobulin class of anti-HBc. Anti-HBc of the IgM class (IgM anti-HBc) predominates during the first 6 months after acute infection, whereas IgG anti-HBc is the predominant class of anti-HBc beyond 6 months. Therefore, patients with current or recent acute hepatitis B, including those in the anti-HBc window, have IgM anti-HBc in their serum. In patients who have recovered from hepatitis B in the remote past as well as those with chronic HBV infection, anti-HBc is predominantly of the IgG class. Infrequently, in ≤1–5% of patients with acute HBV infection, levels of HBsAg are too low to be detected; in such cases, the presence of IgM anti-HBc establishes the diagnosis of acute hepatitis B. When isolated anti-HBc occurs in the rare patient with chronic hepatitis B whose HBsAg level is below the sensitivity threshold of contemporary immunoassays (a low-level carrier), anti-HBc is of the IgG class. Generally, in persons who have recovered from hepatitis B, anti-HBs and anti-HBc persist indefinitely.

The temporal association between the appearance of anti-HBs and resolution of HBV infection as well as the observation that persons with anti-HBs in serum are protected against reinfection with HBV suggests that *anti-HBs is the protective antibody*. Therefore, strategies for prevention of HBV infection are based on providing susceptible persons with circulating anti-HBs (see below). Occasionally, in ~5–10% of patients with chronic hepatitis B, low-level, low-affinity anti-HBs can be detected. Although an unusual serologic pattern in chronic hepatitis B, coexisting HBsAg and anti-HBs positivity does not signal imminent clearance of hepatitis B. Recently, the presence of coexisting HBsAg and anti-HBs in 4.2% of 6534 Chinese patients with chronic hepatitis B was reported (in HBeAg-negative but not HBeAg-reactive patients) to be associated with a higher risk of advanced fibrosis and cirrhosis. The fact that the coexistent HBsAg–anti-HBs subset was significantly older suggests that the link with advanced disease was predominantly the result of longer-duration disease. These patients with HBsAg and such nonneutralizing anti-HBs should be categorized as having chronic HBV infection and, because of the reported association with advanced fibrosis, evaluated for a higher risk of advanced disease.

The other readily detectable serologic marker of HBV infection, HBeAg, appears concurrently with or shortly after HBsAg. Its appearance coincides temporally with high levels of virus replication and reflects the presence of circulating intact virions and detectable HBV DNA (with the notable exception of patients with precore mutations who cannot synthesize HBeAg—see "Molecular Variants"). Pre-S1 and pre-S2 proteins are also expressed during periods of peak replication, but assays for these gene products are not routinely available. In self-limited HBV infections, HBeAg becomes undetectable shortly after peak elevations in aminotransferase activity, before the disappearance of HBsAg, and anti-HBe then becomes detectable, coinciding with a period of relatively lower infectivity (Fig. 350-4). Because HBeAg appears transiently (and HBV DNA, see below, is always present) during typical cases of acute infection, testing for HBeAg and HBV DNA is of limited clinical utility; in contrast, testing for the presence of these markers is of substantial importance in patients with chronic infection.

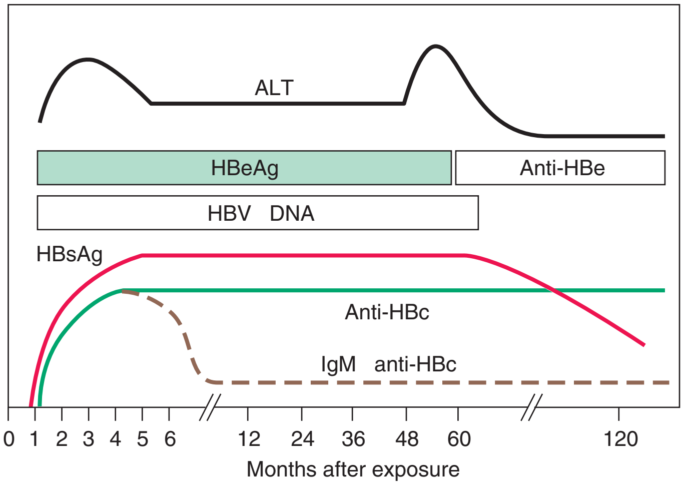

FIGURE 350-5 Scheme of typical laboratory features of wild-type chronic hepatitis B. HBeAg and hepatitis B virus (HBV) DNA can be detected in serum during the relatively *replicative phase* of chronic infection, which is associated with infectivity and liver injury. Seroconversion from the replicative phase to the relatively *nonreplicative phase* occurs at a rate of ~10% per year and is heralded by an acute hepatitis–like elevation of alanine aminotransferase (ALT) activity; during the nonreplicative phase, infectivity and liver injury are limited. In HBeAg-negative chronic hepatitis B associated with mutations in the precore region of the HBV genome, replicative chronic hepatitis B occurs in the absence of HBeAg.

Departing from the pattern typical of acute HBV infections, in chronic HBV infection, HBsAg remains detectable beyond 6 months, anti-HBc is primarily of the IgG class, and anti-HBs is either undetectable or detectable at low levels (see "Laboratory Features") **(Fig. 350-5)**. During early chronic HBV infection, HBV DNA can be detected both in serum and in hepatocyte nuclei, where it is present in free or episomal form. This relatively highly *replicative stage* of HBV infection is the time of maximal infectivity and liver injury; HBeAg is a qualitative marker and HBV DNA a quantitative marker of this replicative phase, during which all three forms of HBV circulate, including intact virions. Over time, the relatively replicative phase of chronic HBV infection gives way to a relatively *nonreplicative phase*. This occurs at a rate of ~10% per year and is accompanied by seroconversion from HBeAg to anti-HBe. In many cases, this seroconversion coincides with a transient, usually mild, acute hepatitis-like elevation in aminotransferase activity, believed to reflect cell-mediated immune clearance of virus-infected hepatocytes. In this relatively nonreplicative or low-replicative phase of chronic infection, when HBV DNA is demonstrable in hepatocyte nuclei, it tends to be integrated into the host genome. In this phase, only spherical and tubular forms of HBV, *not intact virions*, circulate, and liver injury tends to subside. In this phase, clinical measures of liver injury (specifically alanine aminotransferase [ALT]) are normal. Most such patients would be characterized as *inactive HBV carriers*. In reality, the designations *replicative* and *nonreplicative* are only relative; even in the so-called nonreplicative phase, circulating HBV DNA can be detected at levels of approximately ≤10^3 virions/mL with highly sensitive amplification probes such as the polymerase chain reaction (PCR); below this replication threshold, liver injury and infectivity of HBV are limited to negligible. Still, the distinctions are pathophysiologically and clinically meaningful. While true inactive carriers rarely progress to an immune active state, in 5% of cases, relatively nonreplicative HBeAg-negative chronic HBV infection converts back to replicative infection with re-expression of HBeAg. Such spontaneous reactivations may also be accompanied by expression of IgM anti-HBc, as well as by exacerbations of liver injury. Because high-titer IgM anti-HBc can reappear during acute exacerbations of chronic hepatitis B, relying on IgM anti-HBc versus IgG anti-HBc to distinguish between acute and chronic hepatitis B infection, respectively, may not always be reliable; in such cases, patient history and additional follow-up monitoring over time are invaluable in helping to distinguish de novo acute hepatitis B infection from acute exacerbation of chronic hepatitis B infection.

MOLECULAR VARIANTS Variation occurs throughout the HBV genome, and clinical isolates of HBV that do not express typical viral proteins have been attributed to mutations in individual or even multiple gene locations. For example, variants have been described that lack nucleocapsid proteins (commonly), envelope proteins (very rarely), or both. Two categories of naturally occurring HBV variants have attracted the most attention: precore mutations and escape mutations.

One precore mutation was identified initially in Mediterranean countries among patients with severe chronic HBV infection and detectable HBV DNA, but with anti-HBe instead of HBeAg. These patients were found to be infected with an HBV mutant that contained an alteration in the precore region, rendering the virus incapable of encoding HBeAg. Although several potential mutation sites exist in the pre-C region, the region of the C gene necessary for the expression of HBeAg (see "Virology and Etiology"), the most commonly encountered in such patients is a single base substitution, from G to A in the second to last codon of the pre-C gene at nucleotide 1896. This substitution results in the replacement of the TGG tryptophan codon by a stop codon (TAG), which prevents the translation of HBeAg. Another mutation, in the core-promoter region, prevents transcription of the coding region for HBeAg and yields an HBeAg-negative phenotype. Patients with such mutations in the precore region and who are unable to secrete HBeAg may have severe liver disease that progresses more rapidly to cirrhosis, or alternatively, they are identified clinically later in the course of the natural history of chronic hepatitis B, when the disease is more advanced. Both "wild-type" HBV and precore-mutant HBV can coexist in the same patient, or mutant HBV may arise late during wild-type HBV infection. Clusters of fulminant hepatitis B have been observed in Japan and Israel among patients with precore mutant mutations, although, typically, in Europe and North America, most cases of fulminant hepatitis B occur in patients with wild-type virus. HBeAg-negative chronic hepatitis with mutations in the precore region is now the most frequently encountered form of hepatitis B in Mediterranean countries and in Europe. In the United States, where HBV genotype A (less prone to G1896A mutation) is prevalent, precore-mutant HBV had been much less common; however, as a result of immigration from Asia and Europe, the proportion of HBeAg-negative hepatitis B–infected persons has increased in the United States, now accounting for ~30–40% of patients with chronic hepatitis B. Characteristic of such HBeAg-negative chronic hepatitis B are lower levels of HBV DNA (usually $\leq 10^5$ IU/mL) and one of several patterns of aminotransferase activity—persistent elevations, periodic fluctuations above the normal range, and periodic fluctuations between the normal and elevated range.

The second important category of HBV mutants consists of *escape mutants*, in which a single amino acid substitution, from glycine to arginine, occurs at position 145 of the immunodominant *a* determinant common to all HBsAg subtypes. This HBsAg alteration leads to a critical conformational change that results in a loss of neutralizing activity by anti-HBs. This specific HBV/*a* mutant has been observed in two situations, active and passive immunization, in which humoral immunologic pressure may favor evolutionary change ("escape") in the virus—in a small number of hepatitis B vaccine recipients who acquired HBV infection despite the prior appearance of neutralizing anti-HBs and in HBV-infected liver transplant recipients treated with a high-potency human monoclonal anti-HBs preparation. Although such mutants have not been recognized frequently, their existence raises a concern that may complicate vaccination strategies and serologic diagnosis.

Different types of mutations emerge during antiviral therapy of chronic hepatitis B with nucleoside analogues; such YMDD and similar mutations in the polymerase motif of HBV are described in Chap. 352.

EXTRAHEPATIC SITES Hepatitis B antigens and HBV DNA have been identified in extrahepatic sites, including the lymph nodes, bone marrow, circulating lymphocytes, spleen, and pancreas. Although the virus does not appear to be associated with tissue injury in any of these extrahepatic sites, its presence in these "remote" reservoirs has been invoked (but is not necessary) to explain the recurrence of HBV infection after orthotopic liver transplantation. The clinical relevance of such extrahepatic HBV is limited.

Hepatitis D The delta hepatitis agent, or HDV, the only member of the genus *Deltavirus*, is a defective RNA virus that co-infects with and requires the helper function of HBV (or other hepadnaviruses) for its replication and expression. Slightly smaller than HBV, HDV is a formalin-sensitive, 35- to 37-nm virus with a hybrid structure. Its nucleocapsid expresses HDV antigen (HDAg), which bears no antigenic homology with any of the HBV antigens, and contains the virus genome. The HDV core is "encapsidated" by an outer envelope of HBsAg, indistinguishable from that of HBV except in its relative compositions of major, middle, and large HBsAg component proteins. The genome is a small, 1700-nucleotide, circular, single-strand RNA of negative polarity that is nonhomologous with HBV DNA (except for a small area of the polymerase gene) but that has features and the rolling circle model of replication common to genomes of plant satellite viruses or viroids. HDV RNA contains many areas of internal complementarity; therefore, it can fold on itself by internal base pairing to form an unusual, very stable, rod-like structure that contains a very stable, self-cleaving and self-ligating ribozyme. HDV RNA requires host RNA polymerase II for its replication in the hepatocyte nucleus via RNA-directed RNA synthesis by transcription of genomic RNA to a complementary antigenomic (plus strand) RNA; the antigenomic RNA, in turn, serves as a template for subsequent genomic RNA synthesis effected by host RNA polymerase I. HDV RNA has only one open reading frame, and HDAg, a product of the antigenomic strand, is the only known HDV protein; HDAg exists in two forms: a small, 195-amino-acid species, which plays a role in facilitating HDV RNA replication, and a large, 214-amino-acid species, which appears to suppress replication but is required for assembly of the antigen into virions. HDV antigens have been shown to bind directly to RNA polymerase II, resulting in stimulation of transcription. Viral assembly requires farnesylation of the large HDAg for ribonucleoprotein anchoring to HBsAg. Both HBV and HDV enter hepatocytes via the NTCP receptor. Although complete hepatitis D virions and liver injury require the cooperative helper function of HBV, intracellular replication of HDV RNA can occur without HBV. Genomic heterogeneity among HDV isolates has been described. Although pathophysiologic and clinical consequences of this genetic diversity have not been established definitively, preliminarily, genotype 2 has been linked to milder disease and genotype 3 to severe acute disease. The clinical spectrum of hepatitis D is common to all eight genotypes identified, the predominant of which is genotype 1.

HDV can either infect a person simultaneously with HBV (*co-infection*) or superinfect a person already infected with HBV (*superinfection*). In instances of superinfection, when HDV infection is transmitted from a donor with one HBsAg subtype to an HBsAg-positive recipient with a different subtype, HDV assumes the HBsAg subtype of the recipient, rather than the donor. Because HDV relies absolutely on HBV for its replication, the duration of HDV infection is determined by the duration of (and cannot outlast) HBV infection. HDV replication tends to suppress HBV replication; therefore, patients with hepatitis D tend to have lower levels of HBV replication. HDV antigen is expressed primarily in hepatocyte nuclei and is occasionally detectable in serum. During acute HDV infection, anti-HDV of the IgM class predominates, and 30–40 days may elapse after symptoms appear before anti-HDV can be detected. In self-limited infection, anti-HDV is low-titer and transient, rarely remaining detectable beyond the clearance of HBsAg and HDV antigen. In chronic HDV infection, anti-HDV circulates in high titer, and both IgM and IgG anti-HDV can be detected. HDV antigen in the liver and HDV RNA in serum and liver can be detected during HDV replication.

Hepatitis C Hepatitis C virus, which, before its identification, was labeled "non-A, non-B hepatitis," is a linear, single-strand, positive-sense, 9600-nucleotide RNA virus, the genome of which is similar in organization to that of flaviviruses and pestiviruses; HCV is the only member of the genus *Hepacivirus* in the family Flaviviridae. The HCV genome contains a single, large open reading frame (ORF) (gene) that codes for a virus polyprotein of ~3000 amino acids, which is cleaved after translation to yield 10 viral proteins. The 5′ end of the genome consists of an untranslated region (containing an internal ribosomal entry site [IRES]) adjacent to the genes for three structural proteins, the nucleocapsid core protein, C, and two envelope glycoproteins, E1 and E2. The 5′ untranslated region and core gene are highly conserved among genotypes, but the envelope proteins are coded for by the hypervariable region, which varies from isolate to isolate and may allow the virus to evade host immunologic containment directed at accessible virus-envelope proteins. The 3′ end of the genome also includes an untranslated region and contains the genes for seven nonstructural (NS) proteins: p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B. p7 is a membrane ion channel protein necessary for efficient assembly and release of HCV. The NS2 cysteine protease cleaves NS3 from NS2, and the NS3-4A serine protease cleaves all the downstream proteins from the polyprotein. Important NS proteins involved in virus replication include the NS3 helicase; NS3-4A serine protease; the multifunctional membrane-associated phosphoprotein NS5A, an essential component of the viral replication membranous web (along with NS4B); and the NS5B RNA-dependent RNA polymerase (**Fig. 350-6**). Because HCV does not replicate via a DNA intermediate, it does not integrate into the host genome. Because HCV tends to circulate in relatively low titer, 10^3–10^7 virions/mL, visualization of the 50- to 80-nm virus particles remains difficult. Still, the replication rate of HCV is very high, 10^{12} virions per day; its half-life is 2.7 h. Historically, the study of HCV had been hampered by few adequate in vitro and animal models. In 2005, however, complete replication of HCV and intact 55-nm virions was described in cell culture systems. These and other subsequent models were fundamental toward the development of targeted, effective antiviral therapies for HCV infection. Still, the continued absence of robust animal models poses a barrier to much-needed vaccine development. Albeit a helpful animal model, the chimpanzee is cumbersome to study, and access to chimpanzees for medical research has been curtailed substantially. In addition, while HCV replication has been documented in a xenograft immunodeficient mouse model containing explants of human liver and in transgenic mouse and rat models, a tractable immunocompetent murine model has remained elusive.

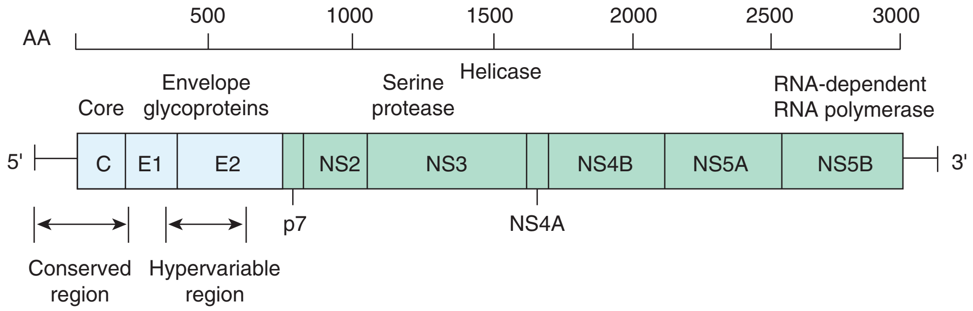

FIGURE 350-6 Organization of the hepatitis C virus genome and its associated, 3000-amino-acid (AA) proteins. The three structural genes at the 5′ end are the core region, C, which codes for the nucleocapsid, and the envelope regions, E1 and E2, which code for envelope glycoproteins. The 5′ untranslated region and the C region are highly conserved among isolates, whereas the envelope domain E2 contains the hypervariable region. At the 3′ end are seven nonstructural (NS) regions—p7, a membrane protein adjacent to the structural proteins that appears to function as an ion channel; NS2, which codes for a cysteine protease; NS3, which codes for a serine protease and an RNA helicase; NS4 and NS4B; NS5A, a multifunctional membrane-associated phosphoprotein, an essential component of the viral replication membranous web; and NS5B, which codes for an RNA-dependent RNA polymerase. After translation of the entire polyprotein, individual proteins are cleaved by both host and viral proteases.

In vitro study has also elucidated important cofactors for the hepatitis C viral life cycle. For example, HCV entry into the hepatocyte occurs via the non-liver-specific CD81 receptor and the liver-specific tight junction protein claudin-1. A growing list of additional host receptors to which HCV binds upon cell entry includes occludin, low-density lipoprotein receptors, glycosaminoglycans, scavenger receptor B1, and epidermal growth factor receptor, among others. After viral entry and uncoating, translation is initiated by the IRES on the endoplasmic reticulum membrane, and the HCV polyprotein is cleaved during translation and posttranslationally by host cellular proteases as well as HCV NS2-3 and NS3-4A proteases. Host cofactors involved in HCV replication include cyclophilin A, which binds to NS5A and yields conformational changes required for viral replication, and the liver-specific host microRNA miR-122. Relying on the same assembly and secretion pathway as low-density and very-low-density lipoproteins, HCV is a lipoviroparticle and masquerades as a lipoprotein, which may limit its visibility to the adaptive immune system, explain its ability to evade immune containment and clearance, and account for the lower circulating LDL levels observed in patients with chronic hepatitis C.

At least six distinct major genotypes (and a minor genotype 7), as well as >50 subtypes within genotypes, of HCV have been identified by nucleotide sequencing. Some HCV genotypes are distributed worldwide, whereas others are more confined geographically (see "Epidemiology and Global Features"). Genotype differences can affect choice of direct-acting antiviral treatment and can affect clinical course (e.g., hepatic steatosis and clinical progression are more likely in genotype 3). Genotypes differ from one another in sequence homology by ≥30%, and subtypes differ by ~20%. Because divergence of HCV isolates within a genotype or subtype and within the same host may vary insufficiently to define a distinct genotype, these intragenotypic differences are referred to as *quasispecies* and differ in sequence homology by only a few percent. The genotypic and quasispecies diversity of HCV, resulting from its high mutation rate, interferes with effective humoral immunity. Neutralizing antibodies to HCV have been demonstrated, but they tend to be short-lived, and HCV infection does not induce lasting immunity against reinfection with different virus isolates or even the same virus isolate. Indeed, the presence of detectable anti-HCV does not indicate protection against re-infection. Thus, neither *heterologous* nor *homologous* immunity appears to develop commonly after acute HCV infection

Currently available, third-generation immunoassays, which incorporate proteins from the core, NS3, and NS5 regions, detect anti-HCV antibodies during acute infection. The most sensitive indicator of HCV infection is the presence of HCV RNA, which requires molecular amplification, for example, by PCR (**Fig. 350-7**). To allow standardization of the quantification of HCV RNA among laboratories and commercial assays, HCV RNA is reported as international units (IUs) per milliliter; quantitative assays with a broad dynamic range are available that allow detection of HCV RNA with a sensitivity as low as 5 IU/mL. HCV RNA can be detected within a few days of exposure to HCV—well before the appearance of anti-HCV—and tends to persist for the duration of HCV infection. Application of sensitive molecular probes for HCV RNA has revealed the presence of replicative HCV in peripheral blood lymphocytes of infected persons; however, as is the case for HBV in lymphocytes, the clinical relevance of HCV lymphocyte infection is not known.

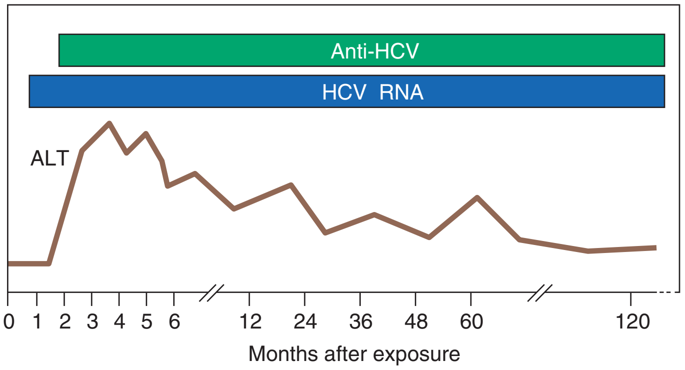

FIGURE 350-7 Scheme of typical laboratory features during acute hepatitis C progressing to chronicity. Hepatitis C virus (HCV) RNA is the first detectable event, preceding alanine aminotransferase (ALT) elevation and the appearance of anti-HCV.

Hepatitis E Previously labeled *epidemic* or *enterically transmitted non-A, non-B hepatitis*, HEV is an enterically transmitted virus that causes clinically apparent hepatitis primarily in India, Asia, Africa, and Central America. In those geographic areas, HEV is the most common cause of acute hepatitis; one-third of the global population appears to have been infected. This agent, with epidemiologic features resembling those of hepatitis A, is a 27- to 34-nm, nonenveloped, heat-stable, HAV-like virus with a 7200-nucleotide, single-strand, positive-sense RNA genome. Like HAV, HEV also exists in a quasi-enveloped form enclosed within host-cell-derived membranes. HEV has three overlapping ORFs (genes), the largest of which, *ORF1*, encodes nonstructural proteins involved in virus replication (the viral replicase, which includes a protease, polymerase, and helicase). A middle-sized gene, *ORF2*, encodes the nucleocapsid protein, the major structural protein, and the smallest, *ORF3*, encodes a small structural phosphoprotein involved in virus particle secretion. All HEV isolates appear to belong to a single serotype, despite genomic heterogeneity of up to 25% and the existence of four species (A–D) and eight genotypes, only four of which, all within species A, have been detected in humans. Genotype associations are important in HEV infection. Genotypes 1 and 2 (common in developing countries) appear to be more virulent anthrotropic variants. In contrast, genotypes 3 (the most common in the United States and Europe) and 4 (seen in China), are endemic in animal species (enzootic variants), represent a zoonotic reservoir for human infections, and are associated with more attenuated or subclinical infection in healthy hosts but with chronic infection in immunocompromised hosts. Contributing to the perpetuation of this virus are the animal reservoirs described above, most notably in swine but also in camels, deer, rats, and rabbits, among others. No genomic or antigenic homology, however, exists between HEV and HAV or other picornaviruses; and HEV, although resembling caliciviruses, is sufficiently distinct from any known agent to merit its own classification as a unique genus, *Orthohepevirus*, within the family Hepeviridae (which includes similar viruses infecting mammals, birds, and fish). The virus has been detected in stool, bile, and liver and is excreted in the stool during the late incubation period. Both IgM anti-HEV during early acute infection and IgG anti-HEV predominating after the first 3 months can be detected (a serologic pattern similar to that of hepatitis A, see above). The presence of HEV RNA in serum and stool accompanies acute infection; viremia resolves as clinical-biochemical recovery ensues, while HEV RNA in stool may outlast viremia by several weeks. Currently, serologic/virologic testing for HEV infection—not approved or licensed by the U.S. Food and Drug Administration (FDA)—can be done in specialized laboratories (e.g., the Centers for Disease Control and Prevention [CDC]) and some commercial laboratories.

■ PATHOGENESIS

Under ordinary circumstances, none of the hepatitis viruses are known to be directly cytopathic to hepatocytes. Evidence suggests that the clinical manifestations and outcomes after acute liver injury associated with viral hepatitis are determined by the immunologic responses of the host. Among the viral hepatitides, the immunopathogenesis of hepatitis B and C has been studied most extensively.

Hepatitis B For HBV, the existence of inactive hepatitis B carriers or “immune tolerant” (see below) patients with normal liver histology and function suggests that the virus is not directly cytopathic. The fact that patients with defects in cellular immune competence are more likely to remain chronically infected rather than to clear HBV supports the role of cellular immune responses in the pathogenesis of hepatitis B–related liver injury. The model that has the most experimental support involves cytolytic T cells sensitized specifically to recognize host and hepatitis B viral antigens on the liver cell surface. Nucleocapsid proteins (HBcAg and possibly HBeAg), present on the cell membrane in minute quantities, are the viral target antigens that, with host antigens, invite cytolytic T cells to destroy HBV-infected hepatocytes. Differences in the robustness and broad polyclonality of CD8+ cytolytic T-cell responsiveness; in the level of HBV-specific helper CD4+ T cells; in attenuation, depletion, and exhaustion of virus-specific T cells; in viral T-cell epitope escape mutations that allow the virus to evade T-cell containment; and in the elaboration of antiviral cytokines by T cells have been invoked to explain differences in outcomes between those who recover after acute hepatitis and those who progress to chronic hepatitis or between those with mild and those with severe (fulminant) acute HBV infection. Ultimately, a robust CD8+ T-cell response is central both to viral clearance and to liver injury.

Although a robust cytolytic T-cell response occurs and eliminates virus-infected liver cells during acute hepatitis B, >90% of HBV DNA has been found in experimentally infected chimpanzees to disappear from the liver and blood before maximal T-cell infiltration of the liver and before most of the biochemical and histologic evidence of liver injury. This observation suggests that components of the innate immune system and inflammatory cytokines, independent of cytopathic antiviral mechanisms, participate in the early immune response to HBV infection; this effect has been shown to represent elimination of HBV replicative intermediates from the cytoplasm and covalently closed circular viral DNA from the nucleus of infected hepatocytes. In turn, the innate immune response to HBV infection is mediated largely by natural killer (NK) cell cytotoxicity, activated by immunosuppressive cytokines (e.g., interleukin [IL] 10 and transforming growth factor [TGF] β), reduced signals from inhibitory receptor expression (e.g., major histocompatibility complex), or increased signals from activating receptor expression on infected hepatocytes. In addition, NK cells reduce helper CD4+ cells, which results in reduced CD8+ cells and exhaustion of the virus-specific T-cell response to HBV infection. Adding to the evidence supporting the role of these immunologic perturbations in the pathogenesis of HBV-associated liver injury are the observations that many of these departures from normal immune function are restored after successful antiviral therapy. Ultimately, HBV-HLA–specific cytolytic T-cell responses of the adaptive immune system are felt to be responsible for recovery from HBV infection.

Debate continues over the relative importance of viral and host factors in the pathogenesis of HBV-associated liver injury and its outcome. As noted above, precore genetic mutants of HBV have been associated with the more severe outcomes of HBV infection (severe chronic and fulminant hepatitis), suggesting that, under certain circumstances, relative pathogenicity is a property of the virus, not the host. The facts that concomitant HDV and HBV infections are associated with more severe liver injury than HBV infection alone and that cells transfected in vitro with the gene for HDV antigen express HDV antigen and then become necrotic in the absence of any immunologic influences are also consistent with a viral effect on pathogenicity. Similarly, in patients who undergo liver transplantation for end-stage chronic hepatitis B, occasionally, rapidly progressive liver injury appears in the new liver. This clinical presentation is associated with an unusual histologic pattern in the new liver, *fibrosing cholestatic hepatitis*, which, ultrastructurally, appears to represent a choking of the cell with overwhelming quantities of HBsAg. This observation suggests that, under the influence of the potent immunosuppressive agents required to prevent allograft rejection, HBV may have a direct cytopathic effect on liver cells, independent of the immune system.

Although the precise mechanism of liver injury in HBV infection remains elusive, studies of nucleocapsid proteins have shed light on the profound immunologic tolerance to HBV of babies born to mothers with highly replicative (HBeAg-positive), chronic HBV infection. In HBeAg-expressing transgenic mice, in utero exposure to HBeAg, which is sufficiently small to traverse the placenta, induces T-cell tolerance to both nucleocapsid proteins. This, in turn, may explain why, when infection occurs so early in life, immunologic clearance does not occur, and protracted, lifelong infection ensues. An alternative explanation proposed to explain why robust liver injury does not accompany neonatal HBV infection but predisposes to chronic infection is defective priming of HBV-specific T cells during in utero exposure to HBV.

“IMMUNE TOLERANT” VERSUS “IMMUNE ACTIVE” CHRONIC HEPATITIS B An important distinction should be drawn between HBV infection acquired at birth, common in endemic areas, such as East Asia, and infection acquired in adulthood, common in the West.

Infection in the neonatal period is associated with the acquisition of what appears to be a high level of immunologic tolerance to HBV and absence of an acute hepatitis illness but the almost invariable establishment of chronic, often lifelong infection. Neonatally acquired HBV infection can culminate decades later in cirrhosis and hepatocellular carcinoma (see "Complications and Sequelae"). In contrast, when HBV infection is acquired during adolescence or early adulthood, the host immune response to HBV-infected hepatocytes tends to be robust, an acute hepatitis-like illness is the rule, and failure to recover is the exception. After adulthood-acquired infection, chronicity is uncommon, and the risk of hepatocellular carcinoma is very low. Based on these observations, some authorities categorize chronic HBV infection into an "immune tolerant" phase, an "immune active" phase, and an "inactive" phase. This somewhat simplistic formulation, however, does not apply at all to typical adult infection with self-limited acute hepatitis B. Moreover, even in early-life-acquired chronic infection, this duality is limited by nuance and imprecision in its distinctions. For example, even among those with neonatally acquired HBV infection, in whom immunologic tolerance appears to be established, immunologic responses to HBV infection have been demonstrated (albeit typically at reduced levels), and intermittent bursts of hepatic necroinflammatory activity punctuate the early decades of life during which liver injury appears to be quiescent. Thus, while labeled by some as the "immune tolerant" phase, it more accurately is a period of dissociation between high-level HBV replication and a paucity of inflammatory liver injury. In addition, even when clinically apparent liver injury and progressive fibrosis emerge and predominate during later decades (the so-called immune active phase), the level of immunologic tolerance to HBV and the corresponding failure of immunologic containment and clearance of HBV infection remain substantial. More accurately, in patients with neonatally acquired HBV infection, a dynamic equilibrium exists between tolerance and intolerance, the outcome of which determines the clinical expression of chronic infection. Persons infected as neonates tend to have a relatively higher level of immunologic tolerance (high replication, low necroinflammatory activity) during the early decades of life and a relatively lower level (but only rarely a loss) of tolerance (and necroinflammatory activity reflecting the level of virus replication) in the later decades of life. Adding to the confusion is an indeterminate category of patients in a gray zone that borders both relative immune activity and relative immune tolerance.

Hepatitis C Cell-mediated immune responses and elaboration by T cells of antiviral cytokines contribute to the multicellular innate and adaptive immune responses involved in the containment of infection and pathogenesis of liver injury associated with hepatitis C. The fact that HCV is so efficient in evading these immune mechanisms is a testament to its highly evolved ability to disrupt host immune responses at multiple levels. After exposure to HCV, the host cell identifies viral product motifs (pattern recognition receptors) that distinguish the virus from "self," resulting in the elaboration of interferons and other cytokines that result in activation of innate and adaptive immune responses. Intrahepatic human leukocyte antigen (HLA) class 1–restricted cytolytic T cells directed at nucleocapsid, envelope, and nonstructural viral protein antigens have been demonstrated in patients with chronic hepatitis C; however, such virus-specific cytolytic T-cell responses do not correlate adequately with the degree of liver injury or with recovery. Yet a consensus has emerged supporting a role in the pathogenesis of HCV-associated liver injury of virus-activated CD4+ helper T cells that stimulate, via the cytokines they elaborate, HCV-specific CD8+ cytotoxic T cells. These responses appear to be more robust (higher in number, more diverse in viral antigen specificity, more functionally effective, and longer lasting) in those who recover from HCV infection than in those who have chronic infection. Contributing to chronic infection are a CD4+ proliferative defect that results in rapid contraction of CD4+ responses, mutations in CD8+ T cell–targeted viral epitopes that allow HCV to escape immune-mediated clearance, and upregulation of inhibitory receptors on functionally impaired, exhausted T cells. Although attention has focused on adaptive immunity, HCV proteins have been shown to interfere with innate immunity by resulting in blocking of type 1 interferon responses and inhibition of interferon signaling and effector molecules in the interferon signaling cascade.

Several single nucleotide polymorphisms have been linked with self-limited hepatitis C, the most convincing of which is the CC haplotype of the *IL28B* gene, which codes for interferon λ3, a component of innate immune antiviral defense. The link between non-CC *IL28B* polymorphisms and failure to clear HCV infection has been explained by a chromosome 19q13.13 frameshift variant upstream of *IL28B*, the ΔG polymorphism of which creates an ORF in a novel interferon gene (*IFN-λ4*) associated with impaired HCV clearance. Also shown to contribute to limiting HCV infection are NK cells of the innate immune system; in spontaneous clearance, NK cells show enhanced IFN-γ elaboration and expression of activating receptors. Conversely, in persistent infection, both peripheral cytotoxicity and intrahepatic NK cell cytotoxicity are dysfunctional. In addition, an important intersection between *IL28B* genotype and NK function can be inferred from the fact that patients with hepatitis C and unfavorable (non-CC, associated with reduced HCV clearance) *IL28B* alleles have been shown to have depressed NK cell/innate immune function. Adding to the complexity of the immune response, HCV core, NS4B, and NS5B have been shown to suppress the immunoregulatory nuclear factor (NF)-κB pathway, resulting in reduced antiapoptotic proteins and a resultant increased vulnerability to tumor necrosis factor (TNF) α-mediated cell death. Finally, the rapid mutation rate of the virus impairs the host immune response; the emergence of substantial viral quasispecies diversity and HCV sequence variation allows the virus to evade attempts by the host to contain HCV infection by both humoral and cellular immune mechanisms.

Hepatitis A and E Viral shedding in these acute hepatitides predates clinical evidence of liver injury, consistent with the absence of a relationship between viral replication and target-organ injury. Instead, as shown for hepatitis B and C, in hepatitis A and E, experimental evidence supports a cytolytic CD8+ T-cell response as the instrument of liver cell injury, in concert with or dwarfed by CD4+ helper T cells or CD4+ interferon γ–secreting cells. Furthermore, in acute HAV infection, expansion of CD8+ T cells active against other, non-HAV, viruses has been demonstrated, and this nonspecific CD8+ T-cell activation correlated with liver injury. HEV, similar to the other hepatitides, has also been shown to interfere with host antiviral defenses, such as interferon signaling and effector function, and to downregulate interferon-stimulated genes. The demonstration of an activated innate immune response in patients with these hepatitides argues for a multitude of immunologic mechanisms in the pathogenesis of the acute liver injury resulting from HAV and HEV infection.

EXTRAHEPATIC MANIFESTATIONS

Immune complex–mediated tissue damage appears to play a pathogenetic role in the extrahepatic manifestations of acute hepatitis B. The occasional prodromal serum sickness–like syndrome observed in acute hepatitis B appears to be related to the deposition in tissue blood vessel walls of HBsAg–anti-HBs circulating immune complexes, leading to activation of the complement system and depressed serum complement levels. Other extrahepatic manifestations of HBV infection may include immune complex–mediated diseases, such as polyarteritis nodosa (generalized vasculitis affecting small- and medium-sized arterioles), glomerulonephritis with nephrotic syndrome, and cryoglobulinemia, which are very rare in acute hepatitis B but far more common in chronic disease **(Chaps. 326 and 375)**.

Immune complex disorders have been linked, albeit rarely, with both hepatitis A and E. In hepatitis E, rare neurologic (including Guillain-Barré syndrome), renal, pancreatic, and hematologic complications have been postulated to result from both immunologic mechanisms and/or direct extrahepatic-site infection with the virus.

PATHOLOGY

Liver biopsy is rarely needed for the diagnosis or management of acute viral hepatitis, for which clinical features and serologic testing remain the cornerstone of diagnosis. The typical morphologic lesions of all types of viral hepatitis are similar and consist of panlobular infiltration with mononuclear cells, hepatic cell necrosis, hyperplasia of Kupffer cells, and variable degrees of cholestasis. Hepatic cell regeneration is present, as evidenced by numerous mitotic figures, multinucleated cells, and "rosette" or "pseudoacinar" formation. The mononuclear infiltration consists primarily of small lymphocytes, although plasma cells and eosinophils occasionally are present. Liver cell damage consists of hepatic cell degeneration and necrosis, cell dropout, ballooning of cells, and acidophilic degeneration of hepatocytes (forming so-called Councilman or apoptotic bodies). Large hepatocytes with a ground-glass appearance of the cytoplasm may be seen in chronic but not in acute HBV infection; these cells contain HBsAg and can be identified histochemically with orcein or aldehyde fuchsin. In uncomplicated viral hepatitis, the reticulin framework is preserved.

In hepatitis C, the histologic lesion is often remarkable for a relative paucity of inflammation, a marked increase in activation of sinusoidal lining cells, lymphoid aggregates, the presence of fat (more frequent in genotype 3 and linked to increased fibrosis), and, occasionally, bile duct lesions in which biliary epithelial cells appear to be piled up without interruption of the basement membrane. Occasionally, microvesicular steatosis occurs in hepatitis D. In hepatitis E, a common histologic feature is marked cholestasis. A cholestatic variant of slowly resolving acute hepatitis A has been described as well.

In the setting of severe acute liver injury, less-specific histologic findings, including portal and lobular inflammation and confluent necrosis, may be present. A more severe histologic lesion, *bridging hepatic necrosis*, also termed *subacute* or *confluent necrosis* or *interface hepatitis*, is observed occasionally in acute hepatitis. "Bridging" between lobules results from large areas of hepatic cell dropout, with collapse of the reticulin framework. Characteristically, the bridge consists of condensed reticulum, inflammatory debris, and degenerating liver cells that span adjacent portal areas, portal to central veins, or central vein to central vein. This lesion had been thought to have prognostic significance; in many of the originally described patients with this lesion, a subacute course terminated in death within several weeks to months, or severe chronic hepatitis and cirrhosis developed; however, the association between bridging necrosis and a poor prognosis in patients with acute hepatitis has not been upheld. Therefore, although demonstration of this lesion in patients with chronic hepatitis has prognostic significance **(Chap. 352)**, its demonstration during acute hepatitis is less meaningful, and liver biopsies to identify this lesion are no longer undertaken routinely in patients with acute hepatitis. In *massive hepatic necrosis* (fulminant hepatitis, "acute yellow atrophy"), the striking feature at postmortem examination is the finding of a small, shrunken, soft liver. Histologic examination reveals massive necrosis and dropout of liver cells of most lobules with extensive collapse and condensation of the reticulin framework. When histologic documentation is required in the management of fulminant or very severe hepatitis, a biopsy can be done by the angiographically guided transjugular route, which permits the performance of this invasive procedure in the presence of severe coagulopathy.

Immunohistochemical and electron-microscopic studies have localized HBsAg to the cytoplasm and plasma membrane of infected liver cells. In contrast, HBcAg predominates in the nucleus, but, occasionally, scant amounts are also seen in the cytoplasm and on the cell membrane. HDV antigen is localized to the hepatocyte nucleus, whereas HAV and HCV antigens are localized to the cytoplasm. Hepatitis E ORF-2 protein staining is distributed in both a cytoplasmic and nuclear pattern.

EPIDEMIOLOGY AND GLOBAL FEATURES

Before the availability of serologic tests for hepatitis viruses, all viral hepatitis cases were labeled either as "infectious" or "serum" hepatitis. Modes of transmission overlap, however, and *a clear distinction among the different types of viral hepatitis cannot be made solely based on clinical or epidemiologic features* **(Table 350-2)**. The most accurate means to distinguish the various types of viral hepatitis involves specific serologic testing.

Hepatitis A *This agent is transmitted almost exclusively by the fecal-oral route*. Person-to-person spread of HAV is enhanced by poor personal hygiene and overcrowding; large outbreaks as well as sporadic cases have been traced to contaminated food, water, milk, frozen raspberries and strawberries, green onions imported from Mexico, and shellfish (e.g., scallops imported from the Philippines used to make sushi, the culprit identified in a 2016 Hawaiian outbreak). Intrafamily and intrainstitutional spreads are also common. Early epidemiologic observations supported a predilection for hepatitis A to occur in late fall and early winter. In temperate zones, epidemic waves have been recorded every 5–20 years as new segments of nonimmune population appeared; however, in developed countries, the incidence of hepatitis A has been declining, presumably as a function of improved sanitation and environmental hygiene, and these cyclic patterns are no longer observed; still, episodic outbreaks among certain high-risk populations continue to be reported (see below). No HAV carrier state has been identified after acute hepatitis A; perpetuation of the virus in nature depends presumably on nonepidemic, inapparent subclinical infection, ingestion of contaminated food or water in, or imported from, endemic areas, and/or contamination linked to environmental reservoirs.

In the general population, anti-HAV, a marker for previous HAV infection, increases in prevalence as a function of increasing age and of decreasing socioeconomic status. In the 1970s, serologic evidence of prior hepatitis A infection occurred in ~40% of urban populations in the United States, most of whose members never recalled having had a symptomatic case of hepatitis. In subsequent decades, however, the prevalence of anti-HAV declined in the United States. In developing countries, exposure, infection, and subsequent immunity are almost universal in childhood. As the frequency of subclinical childhood infections declines in developed countries, a susceptible cohort of adults emerges. Hepatitis A tends to be more symptomatic in adults; therefore, paradoxically, as the frequency of HAV infection declines, the likelihood of clinically apparent, even severe, HAV illnesses increases in the susceptible adult population. Travel to endemic areas is a common source of infection for adults from nonendemic areas. Important recognized epidemiologic foci of HAV infection include childcare centers, neonatal intensive care units, promiscuous men who have sex with men, injection drug users, and unvaccinated close contacts of newly arrived international adopted children, most of whom emanate from countries with intermediate-to-high hepatitis A endemicity. Although hepatitis A is rarely bloodborne, several outbreaks have been recognized in recipients of clotting-factor concentrates. In the United States, the introduction of hepatitis A vaccination programs among children from high-incidence states has resulted in a >70% reduction in the annual incidence of new HAV infections and has shifted the burden of new infections from children to adults. In the 2007–2012 U.S. Public Health Service National Health and Nutrition Examination Survey (NHANES), the prevalence of anti-HAV in the U.S. population aged ≥20 years had declined to 24.2% from the 29.5% measured in NHANES 1999–2006. From 2007 to 2016, the overall anti-HAV prevalence rose, from 30% in 2007–2010 to 40% in 2011–2016 in U.S.-born persons over the age of 2; the largest increases were observed in children (2–11 years) and teenagers (12–19 years), with a slight increase among young adults (age 20–29). In all other age cohorts among U.S.-born adults, however, the prevalence of anti-HAV remained unchanged between 2007 and 2017; no more than 25% of any age group had protective antibodies. The lowest age-specific prevalence of anti-HAV (16.1–17.6%) occurred in adults in the fourth and fifth decades (aged 30–49 years). This is a subgroup of the population who remain susceptible to acute hepatitis A acquired during travel to endemic areas and from contaminated foods, especially those imported from endemic countries. Recognized initially in San Diego, California, in 2016, widespread person-to-person outbreaks, attributed to fecally contaminated environments, of acute hepatitis A occurred primarily among homeless persons and persons who were using injection drugs. Ultimately, this outbreak extended to at least 32 states (highest number of cases in Kentucky), and by March 2020, 31,950 cases were reported, resulting in 19,548 hospitalizations (61% of cases) and 322 deaths (1% of reported cases, 1.6% of hospitalized cases). By 2022, 315 hepatitis A

TABLE 350-2 Clinical and Epidemiologic Features of Viral Hepatitis

FEATURE	HAV	HBV	HCV	HDV	HEV
Incubation (days)	15–45, mean 30	30–180, mean 60–90	15–160, mean 50	30–180, mean 60–90	14–60, mean 40
Onset	Acute	Insidious or acute	Insidious or acute	Insidious or acute	Acute
Age preference	Children, young adults	Young adults (sexual and percutaneous), babies, toddlers	Any age, but more common in adults	Any age (similar to HBV)	Epidemic cases: young adults (20–40 years); sporadic cases: older adults (>60)
Transmission					
Fecal-oral	+++	–	–	–	+++
Percutaneous	Unusual	+++	+++	+++	–
Perinatal	–	+++	±[a]	+	–
Sexual	±	++	±[a]	++	–
Clinical					
Severity	Mild	Occasionally severe	Moderate	Occasionally severe	Mild
Fulminant	0.1%	0.1–1%	0.1%	5–20%[b]	1–2%[c]
Progression to chronicity	None	Occasional (1–10%)	Common (85%)	Common[d]	None[e]
Carrier	None	(90% of neonates)	1.5–3.2%	Variable[g]	None
Cancer	None	0.1–30%[f] + (neonatal infection)	+	±	None
Prognosis	Excellent	Worse with age, debility	Moderate	Acute, good; chronic, poor	Good
Prophylaxis	Ig, inactivated vaccine	HBIG, recombinant vaccine	None	HBV vaccine (none for HBV carriers)	Vaccine
Therapy	None	Interferon[h] Lamivudine[h] Adefovir[h] Pegylated interferon[i] Entecavir[i] Telbivudine[i] Tenofovir disoproxil fumarate[i] Tenofovir alafenamide[i]	Pegylated interferon ribavirin,[h] telaprevir,[h] boceprevir,[h] simeprevir,[h] sofosbuvir, ledipasvir, paritaprevir/ritonavir,[h] ombitasvir,[h] dasabuvir,[h] daclatasvir,[h] velpatasvir, grazoprevir, elbasvir, glecaprevir, pibrentasvir, voxilaprevir	Pegylated interferon ±	None[j]

[a] Primarily with HIV co-infection and high-level viremia in index case; more likely in persons with multiple sex partners or sexually transmitted diseases; risk ~5%. [b] Up to 5% in acute HBV/HDV co-infection; up to 20% in HDV superinfection of chronic HBV infection. [c] 10–20% in pregnant women. [d] In acute HBV/HDV co-infection, the frequency of chronicity is the same as that for HBV; in HDV superinfection, chronicity is invariable. [e] Except as observed in immunosuppressed liver allograft recipients or other immunosuppressed hosts. [f] Varies considerably throughout the world and in subpopulations within countries; see text. [g] Common in Mediterranean countries; rare in North America and western Europe. [h] No longer recommended or not included in first-line therapy. [i] First-line agents. [j] Anecdotal reports and retrospective studies suggest that pegylated interferon and/or ribavirin are effective in treating chronic hepatitis E, observed in immunocompromised persons; ribavirin monotherapy has been used successfully in acute, severe hepatitis E.

Abbreviation: HBIG, hepatitis B immunoglobulin. See text for other abbreviations.

outbreak–related deaths were reported across 37 states. The increased clinical severity, rate of hospitalization, and death in these outbreaks can be attributed to their involving an older population (mean age ranging from 36 to 42 years; median age of death was 55), born before the introduction of universal childhood hepatitis A vaccination and in whom clinical severity, as noted above, is higher than in children. Moreover, the affected homeless and drug-using populations suffer from multiple comorbidities (including HBV or HCV co-infection) and disparities in access to health care. Addressing this multistate outbreak has required a vigorous hepatitis A vaccination effort (still falling short of adequate coverage) as well as environmental sanitation/hygiene and education among these susceptible populations.

Hepatitis B Percutaneous inoculation has long been recognized as a major route of hepatitis B transmission, but the outmoded designation "serum hepatitis" is an inaccurate label for the epidemiologic spectrum of HBV infection. As detailed below, most of the hepatitis transmitted by blood transfusion is not caused by HBV; moreover, in approximately two-thirds of patients with acute type B hepatitis, no history of an identifiable percutaneous exposure can be elicited. We now recognize that many cases of hepatitis B result from less obvious modes of nonpercutaneous or covert percutaneous transmission. HBsAg has been identified in almost every body fluid from infected persons, and at least some of these body fluids—most notably semen and saliva—are infectious, albeit less so than serum, when administered percutaneously or nonpercutaneously to experimental animals. Among the nonpercutaneous modes of HBV transmission, oral ingestion has been documented as a potential but inefficient route of exposure. By contrast, the two nonpercutaneous routes considered to have the greatest impact are intimate (especially sexual) contact and perinatal transmission.

In sub-Saharan Africa, intimate contact among toddlers is considered instrumental in contributing to the maintenance of the high frequency of hepatitis B in the population. Perinatal transmission occurs primarily in infants born to mothers with chronic hepatitis B or (rarely) mothers with acute hepatitis B during the third trimester of pregnancy or during the early postpartum period. Perinatal transmission is uncommon in North America and western Europe but occurs with great frequency and is the most important mode of HBV perpetuation in East Asia and developing countries. Although the precise mode of perinatal transmission is unknown, and although ~10% of infections may be acquired in utero, epidemiologic evidence suggests that most infections occur approximately at the time of delivery and are not related to breast-feeding (which is not contraindicated in women with hepatitis B). The likelihood of perinatal transmission of HBV correlates with the presence of HBeAg and high-level viral replication; 90% of HBeAg-positive mothers but only 10–15% of anti-HBe-positive mothers transmit HBV infection to their offspring. In most cases, acute infection in the neonate is clinically asymptomatic, but the child is very likely to remain chronically infected.

The over 290 million persons with chronic HBV infection in the world constitute the main reservoir of hepatitis B in human beings. Whereas serum HBsAg is infrequent (0.1–0.5%) in normal populations in the United States and western Europe, the global prevalence is estimated at 3.5%, and a prevalence of up to 5–10% has been found in East Asia, sub-Saharan Africa, and tropical countries. The prevalence can be even higher in certain high-risk groups, including persons with Down syndrome, lepromatous leprosy, leukemia, Hodgkin disease, polyarteritis nodosa, and chronic renal disease on hemodialysis, as well as in injection drug users.

Other groups with high rates of HBV infection include spouses of acutely infected persons; sexually promiscuous persons (especially promiscuous men who have sex with men); health care workers and first responders exposed to blood; persons who require repeated transfusions especially with pooled blood-product concentrates (e.g., hemophiliacs); residents and staff of custodial institutions for the developmentally handicapped; prisoners; and, to a lesser extent, family members of chronically infected patients. Because of highly sensitive virologic screening (antigen, antibody, and nucleic acid testing) of donor blood, the risk of acquiring HBV infection from a blood transfusion is estimated to be 1 in 360,000.

Prevalence of infection, modes of transmission, and human behavior conspire to mold geographically different epidemiologic patterns of HBV infection. In East Asia and Africa, hepatitis B, a disease of the newborn and young children, is perpetuated by a cycle of maternal-neonatal spread. In North America and western Europe, hepatitis B is primarily a disease of adolescence and early adulthood, the time of life when intimate sexual contact and recreational and occupational percutaneous exposures tend to occur. To some degree, however, this dichotomy between high-prevalence and low-prevalence geographic regions has been minimized by immigration from high-prevalence to low-prevalence areas. For example, in the United States, NHANES data from 2007 to 2012 revealed an overall prevalence of current HBV infection (detectable HBsAg) of 0.3%; however, the prevalence in Asian persons, 93% of whom were foreign-born, was tenfold higher, 3.1%, representing 50% of the U.S. national disease burden. As a result of adoption of safe behaviors in high-risk groups as well as screening and vaccination programs, the incidence of newly reported HBV infections fell by >80% in the United States during the 1990s and has remained low in the 21st century (with a low of 0.6 cases per 100,000 population in 2021). Paralleling that trend, the imbalance between cases in U.S.-born and foreign-born persons widened; currently, imported cases in non-U.S.-born persons outnumber domestic cases by manyfold; in NHANES 2017–2020, the prevalence of HBV infection was 1.0% in foreign-born versus 0.2% in the entire cohort; non-U.S.-born persons accounted for 73.6% of all infections. The introduction of hepatitis B vaccine in the early 1980s and adoption of universal childhood vaccination policies in many countries resulted in a dramatic, ~90% decline in the incidence of new HBV infections in those countries as well as in the dire consequences of chronic infection, including hepatocellular carcinoma. In the United States, as demonstrated in NHANES 2007–2012, following the 1991 implementation of universal childhood vaccination, HBsAg seropositivity had declined in children aged 6–19 years to as low as 0.03%, an ~85% reduction.

UNIVERSAL SCREENING FOR HEPATITIS B Populations and groups at high risk for hepatitis B are listed in Table 350-3. Screening for HBV infection used to be recommended in these high-risk populations; however, because risk-group screening had not been shown to be effective in reducing the prevalence of HBV infection in the population and the majority of acute hepatitis B cases were found to occur in persons who were not in these high-risk groups, *universal screening of all adults* (≥18 years) for hepatitis B was recommended in 2023 by the CDC. Because current antiviral therapy has been highly effective in reducing the clinical progression and consequences of HBV infection, such universal screening was found to be cost-effective in reducing the morbidity and mortality associated with chronic hepatitis B.

TABLE 350-3 Populations with a High Risk for HBV Infection[a]

Persons born in countries/regions with a high (≥8%) and intermediate (≥2%) prevalence of HBV infection including immigrants and adopted children and including persons born in the United States who were not vaccinated as infants and whose parents emigrated from areas of high HBV endemicity
Household and sexual contacts of, or needle sharing with, persons who have hepatitis B
Babies born to HBsAg-positive mothers
Persons who have used injection drugs
Persons with multiple sexual contacts or a history of sexually transmitted disease
Men who have sex with men
Persons incarcerated in a correctional facility or other detention settings
Persons with elevated alanine or aspartate aminotransferase levels
Persons with HCV or HIV infection
Hemodialysis patients
Health care and laboratory workers and first responders exposed to blood

[a] Screening for hepatitis B had been recommended in the past for persons in these high-risk groups; however, in 2023, the Centers for Disease Control and Prevention recommended universal screening of all adults (>18 years of age) for hepatitis B. Other groups who should be tested for hepatitis B include pregnant women; persons who are the source of blood or body fluids that would be an indication for postexposure prophylaxis (e.g., needlestick, mucosal exposure sexual assault); and persons who require immunosuppressive or cytotoxic therapy (including anti–tumor necrosis factor α therapy for rheumatologic or inflammatory bowel disorders).

Hepatitis D Infection with HDV has a worldwide distribution, but two epidemiologic patterns exist. In Mediterranean countries (northern Africa, southern Europe, the Middle East), HDV infection is endemic among those with hepatitis B, and the disease is transmitted predominantly by nonpercutaneous means, especially close personal contact. In nonendemic areas, such as the United States (where hepatitis D is rare among persons with chronic hepatitis B) and northern Europe, HDV infection is confined to persons exposed frequently to blood and blood products, primarily injection drug users (especially in HIV-infected injection drug users) and hemophiliacs. In the United States, the prevalence of HDV infection in the national population was 0.02% in NHANES 1999–2012 and 0.11% in NHANES 2011–2016; however, among HBsAg-positive persons, the prevalence of HDV infection is highest in injection drug users (11–36%) and hemophiliacs (19%). In one study of persons who inject drugs (PWID) in San Francisco, the overall prevalence of HDV infection was 1.1% but as high as 34.6% in PWID who had chronic HBV infection. HDV infection can be introduced into a population through drug users or by migration of persons from endemic to nonendemic areas. Thus, patterns of population migration and human behavior facilitating percutaneous contact play important roles in the introduction and amplification of HDV infection. Occasionally, the migrating epidemiology of hepatitis D is expressed in explosive outbreaks of severe hepatitis, such as those that have occurred in remote South American villages (e.g., "Lábrea fever" in the Amazon basin) as well as in urban centers in the United States. Ultimately, such outbreaks of hepatitis D—either of co-infections with acute hepatitis B or of superinfections in those already infected with HBV—may blur the distinctions between endemic and nonendemic areas. On a global scale, HDV infection declined at the end of the 1990s. Even in Italy, an HDV-endemic area, public health measures introduced to control HBV infection (e.g., mass hepatitis B vaccination) resulted during the 1990s in a 1.5%/year reduction in the prevalence of HDV infection. Still, the frequency of HDV infection during the first decade of the twenty-first century has not fallen below levels reached during the 1990s; the reservoir has been sustained by survivors infected during 1970–1980 and recent immigrants from still-endemic (e.g., eastern Europe and Central Asia) to less-endemic countries. The current global prevalence of HDV infection has been estimated at 12 million people, albeit with significant regional variation. Of the eight HDV genotypes, genotype 1 is distributed worldwide, while the others are more geographically confined (e.g., genotypes 2 and 4 in the Far East, 3 in South America, and 5–8 in Africa).

Hepatitis C Routine screening of blood donors for HBsAg and the elimination of commercial blood sources in the early 1970s reduced the frequency of, but did not eliminate, transfusion-associated hepatitis. During the 1970s, the likelihood of acquiring hepatitis after transfusion of voluntarily donated, HBsAg-screened blood was

~10% per patient (up to 0.9% per unit transfused); 90–95% of these cases were classified, based on serologic exclusion of hepatitis A and B, as "non-A, non-B" hepatitis. For patients requiring transfusion of pooled products, such as clotting factor concentrates, the risk was even higher, up to 20–30%.

During the 1980s, voluntary self-exclusion of blood donors with risk factors for AIDS and then the introduction of donor screening for anti-HIV reduced further the likelihood of transfusion-associated hepatitis to <5%. During the late 1980s and early 1990s, the introduction first of "surrogate" screening tests for non-A, non-B hepatitis (ALT and anti-HBc, both shown to identify blood donors with a higher likelihood of transmitting non-A, non-B hepatitis to recipients) and, subsequently, after the discovery of HCV, progressively more sensitive immunoassays for anti-HCV and then the application of automated PCR testing of donated blood for HCV RNA reduced the risk of transfusion-associated hepatitis C even further, to almost imperceptible levels ranging between 1 in 2.3 million transfusions to 1 in 4.7 million transfusions.

In addition to being transmitted by transfusion, hepatitis C can be transmitted by other percutaneous routes, such as injection drug use. This virus can be transmitted by occupational exposure to blood, and the likelihood of infection is increased in hemodialysis units. Although the frequency of transfusion-associated hepatitis C fell as a result of blood-donor screening, the *overall* frequency of reported hepatitis C cases did not change until the 1990s, when the overall frequency of reported cases fell by 80%, in parallel with a reduction in the number of new cases in injection drug users, the source of most of the HCV reservoir. After the exclusion of anti-HCV-positive plasma units from the donor pool, rare, sporadic instances occurred of hepatitis C among recipients of immunoglobulin preparations for intravenous (but not intramuscular) use.

Serologic evidence for HCV infection occurs in 90% of patients with a history of transfusion-associated hepatitis (almost all occurring before 1992, when second-generation HCV screening tests were introduced); hemophiliacs and others treated with clotting factors; injection drug users; 60–70% of patients with sporadic "non-A, non-B" hepatitis who lack identifiable risk factors; 0.5% of volunteer blood donors; and, in the NHANES survey conducted in the United States between 1999 and 2002, 1.6% of the general population in the United States, which translated into 4.1 million persons (3.2 million with viremia), the majority of whom were unaware of their infections. Moreover, such population surveys do not include higher-risk groups such as incarcerated persons, homeless persons, and active injection drug users, indicating that the actual prevalence is even higher (estimated to add an additional 1 million with anti-HCV antibody and 0.8 million with HCV RNA in a later cohort assessed in 2003–2010). Comparable frequencies of HCV infection occur in most countries around the world, with 71 million persons infected worldwide, but extraordinarily high prevalences of HCV infection occur in certain countries such as Egypt, where >20% of the population (as high as 50% in persons born prior to 1960) in some cities is infected. The high frequency in Egypt is attributable to contaminated equipment used for medical procedures and unsafe injection practices in the 1950s to 1980s (during a campaign to eradicate schistosomiasis with intravenous tartar emetic). Thanks to a 2018–2019 Egyptian government program to screen its entire adult population (79% participation among >60 million people) for hepatitis C and treat infected persons (2.2 million, 4.6% of those screened; of the 83% with a documented outcome, 99% were cured; the cost to identify and cure a person was $130) with generic versions of direct-acting antiviral (DAA) therapy **(Chap. 352)**, hepatitis C has been nearly eliminated there.

In the United States, African Americans and Mexican Americans have higher frequencies of HCV infection than whites. Data from NHANES showed that between 1988 and 1994, 30- to 40-year-old men had the highest prevalence of HCV infection; however, in the NHANES survey conducted between 1999 and 2002, the peak age decile had shifted to those aged 40–49 years; an increase in hepatitis C–related mortality has paralleled this secular trend, increasing since 1995 predominantly in the 45- to 65-year age group. Thus, despite an 80% reduction in new reported HCV infections during the 1990s, the prevalence of HCV infection in the population was sustained by an aging cohort that had acquired their infections three to four decades earlier, during the 1960s and 1970s, as a result predominantly of self-inoculation with recreational drugs. Retrospective phylogenetic mapping of >45,000 HCV genotype 1a isolates revealed that the hepatitis C epidemic emerged in the United States between 1940 and 1965, peaking in 1950 and aligning temporally with the post–World War II expansion of medical procedures (including reuse of glass syringes). Thus, HCV was amplified iatrogenically not only in Egypt but also in the United States; in the United States, the seeds sewn by medical procedures in the 1950s were reaped in the 1960s and 1970s among transfusion recipients and injection drug users, even those whose drug use was confined to brief adolescent experimentation.

In NHANES 2003–2010, the prevalence of HCV infection (HCV RNA reactivity) in the United States had actually fallen to 1% (2.7 million persons) from 1.3% (3.2 million) the decade before (NHANES 1999–2002), attributable to deaths among the HCV-infected population. In NHANES data from 2013–2016, the prevalence of current HCV infection (HCV RNA reactivity) had fallen lower, to 0.9%. Worth emphasizing, however, is that NHANES datasets account for only the domiciled and noninstitutionalized U.S. civilian population. In the same time frame, the prevalence among incarcerated and homeless persons was estimated to be 10.7%.

As deaths resulting from HIV infection fell after 1999, age-adjusted mortality associated with HCV infection surpassed that of HIV infection in 2007; >70% of HCV-associated deaths occurred in the "baby boomer" cohort born between 1945 and 1965. By 2012, HCV mortality had surpassed deaths from HIV, tuberculosis, hepatitis B, and 57 other notifiable infectious diseases (i.e., *all* infectious diseases) reported to the CDC. In NHANES 1999–2002, compared to the 1.6% prevalence of HCV infection in the population at large, the prevalence in the 1945–1965 birth cohort was 3.2%, representing three-quarters of all infected persons. Therefore, in 2012, the CDC and, in 2013, the U.S. Preventive Services Task Force (USPSTF) recommended that all persons born between 1945 and 1965 be screened for hepatitis C, without ascertainment of risk, a recommendation shown to be cost-effective and predicted to identify 800,000 infected persons. Because of the availability of highly effective antiviral therapy, such screening would have the potential to avert 200,000 cases of cirrhosis and 47,000 cases of hepatocellular carcinoma and to prevent 120,000 hepatitis-related deaths; with the availability of the new generation of DAAs (efficacy >95%, **see Chap. 352)**, screening baby boomers and treating those with hepatitis C have been predicted to reduce the HCV-associated disease burden by 50–70% through 2050.

Still, persons with chronic hepatitis C identified by 1945–1965 birth-cohort screening are older than 50, and by the time they are identified, >20% already have advanced liver disease. In 2020, based on (1) the 95–99% efficacy of all-oral, well-tolerated, highly effective DAAs; (2) the demonstration that the endpoint of DAA therapy (sustained virologic response) was associated with a marked decrease in liver and all-cause mortality, cirrhosis, and hepatocellular carcinoma **(Chap. 352)**; (3) a reduction in the initially high cost of DAA therapy; (4) the demonstration of higher cost-effectiveness of screening all adults rather than birth-cohort screening; and (5) the shifting demographics of HCV infection (see below), especially since 2010, toward a younger population exposed through injection drug use, the American Association for the Study of Liver Diseases and the Infectious Diseases Society of America as well as the USPSTF and CDC expanded recommended *hepatitis C screening to all adolescents and adults* aged 18–79 (and because of the substantial increase in HCV infections among women of child-bearing age [age 20–39], expanded such screening to pregnant women).

Hepatitis C accounts for 40% of chronic liver disease and, before the introduction of high-efficacy DAA therapy, was the most frequent indication for liver transplantation; hepatitis C is estimated to account for 8000–10,000 deaths per year in the United States. The distribution of HCV genotypes varies in different parts of the world. Worldwide, genotype 1 is the most common. In the United States, genotype 1 accounts for 70% of HCV infections, whereas genotypes 2 and 3 account for the remaining 30%; among African Americans, the frequency of genotype 1 is even higher (i.e., 90%). Genotype 4 predominates in Egypt; genotype 5 is localized to South Africa, genotype 6 to Hong Kong, and genotype 7 to Central Africa.

As a bloodborne infection, HCV potentially can be transmitted sexually and perinatally; however, both modes of transmission are inefficient for hepatitis C. Although 10–15% of patients with acute hepatitis C report having potential sexual sources of infection, most studies have failed to identify sexual transmission of this agent. The chances of sexual and perinatal transmission have been estimated to be ~5% but have shown in a prospective study to be only 1% between monogamous sexual partners, well below comparable rates for HIV and HBV infections. An important exception to the low rate of perinatal transmission is the high risk—typically in excess of 10%—in babies born to HIV-HCV coinfected mothers. Moreover, sexual transmission appears to be confined to such subgroups as persons with multiple sexual partners and sexually transmitted diseases; for example, isolated clusters of sexually transmitted HCV infection have been reported in HIV-infected men who have sex with men. Breast-feeding does not increase the risk of HCV infection between an infected mother and her infant. Infection of health workers is not dramatically higher than among the general population; however, health workers are more likely to acquire HCV infection through accidental needle punctures, the efficiency of which is ~3%. Infection of household contacts is rare as well. Most asymptomatic blood donors found to have anti-HCV and ~20–30% of persons with reported cases of acute hepatitis C do not fall into a recognized risk group; however, many such blood donors do recall long forgotten, risk-associated behaviors when questioned carefully.

Besides persons born between 1945 and 1965, other groups with an increased frequency of HCV infection are listed in Table 350-4. In immunosuppressed individuals, levels of anti-HCV may be undetectable, and a diagnosis may require testing for HCV RNA. Although new acute cases of hepatitis C are rare outside of the injection drug–using community, newly diagnosed cases are common among otherwise healthy persons who experimented briefly with injection drugs, as noted above, four or five decades earlier. Such instances usually remain unrecognized for years, until unearthed by laboratory screening for routine medical examinations, insurance applications, and attempted blood donation. Although, overall, the annual incidence of new HCV infections has continued to fall, the rate of new infections has been increasing since 2002, has accelerated since 2010 (tripling from 0.3/100,000 to 1.2/100,000 between 2009 and 2018), and has been amplified by the recent epidemic of opioid use in a new cohort of young injection drug users aged 20–39 years (accounting for a 3.8-fold increase in cases between 2010 and 2017 and for more than two-thirds of all acute cases), who, unlike older cohorts, had not learned to take precautions to prevent bloodborne infections. Reflecting this emerging development, the prevalence of current HCV infection (HCV RNA reactivity) in the United States rose from 0.65% (1.7 million persons) in a 2010–2014 NHANES analysis to 0.9% (2.04 million persons) in a 2013–2016 NHANES analysis. Moreover, based on an estimate of populations excluded from this NHANES analysis, the prevalence would be even higher, 0.93% (2.27 million persons). This late temporal trend was attributed to the increase of acute cases in injections drug users, driven by increases in states most affected by the opioid/injection drug use epidemic. Also, in parallel with this trend, the prevalence of HCV infection in women aged 15–44 years (of child-bearing age) doubled between 2016 and 2014; accordingly, screening of pregnant women for HCV infection is now recommended as well.

TABLE 350-4 Populations with a High Risk for HCV Infection[a]
All adults aged 18–79 should be screened, a recommendation that supplants the earlier focus on persons born between 1945 and 1965[a]
Persons who have ever used injection drugs
Persons with HIV infection
Hemophiliacs treated with clotting factor concentrates prior to 1987
Persons who have ever undergone long-term hemodialysis
Persons with unexplained elevations of aminotransferase levels
Transfusion or transplantation recipients prior to July 1992
Recipients of blood or organs from a donor found to be positive for hepatitis C
Children born to women with hepatitis C
Health care, public safety, and emergency medical personnel following needle injury or mucosal exposure to HCV-contaminated blood
Sexual partners of persons with hepatitis C infection
Pregnant women

[a] Screening for hepatitis C had been recommended in the past for persons in these high-risk groups; however, in 2020, the Centers for Disease Control and Prevention recommended universal screening of all adults (>18 years of age) for hepatitis C.

Hepatitis E This type of hepatitis, identified in India, Asia, Africa, the Middle East, and Central America (endemic areas), resembles hepatitis A in its primarily enteric mode of spread. The commonly recognized cases occur after contamination of water supplies such as after monsoon flooding, but sporadic, isolated cases occur. An epidemiologic feature that distinguishes HEV from other enteric agents is the rarity of secondary person-to-person spread from infected persons to their close contacts. Large waterborne outbreaks in endemic areas are linked to genotypes 1 and 2, arise in populations that are immune to HAV, favor young adults, and account for antibody prevalences of 30–80%. The worldwide annual incidence of acute HEV infections has been estimated conservatively to be at least 20 million (of which 3.3 million are symptomatic), rendering HEV infection as the most common cause of acute viral hepatitis. In nonendemic areas of the world, such as the United States, clinically apparent acute hepatitis E is extremely rare; however, during the 1988–1994 NHANES III survey conducted by the U.S. Public Health Service, the prevalence of anti-HEV was 21%, reflecting subclinical infections, infection with genotypes 3 and 4, predominantly in older males (>60 years). A repeat NHANES study in 2009–2010, however, showed a substantial 70% two-decade reduction in anti-HEV to only 6%, more consistent with the rarity of acute hepatitis E in the United States than the previous NHANES result would suggest and perhaps a reflection of a more specific anti-HEV assay used in the second time period. A subsequent examination of the 2009–2016 NHANES dataset demonstrated a seroprevalence of HEV infection of 6.1%. Moreover, supporting the concern for influence of assay differences on population estimates, a re-interrogation of the NHANES III cohort based on assays of the highest sensitivity identified varying prevalence estimates ranging from 16–21%. Again, in the 2009–2016 cohort, older age was associated with anti-HEV seropositivity, as were non-Hispanic Asian ethnicity and birth outside the United States.

In nonendemic areas, HEV accounts for only a small proportion of cases of sporadic (labeled "autochthonous" or indigenous) hepatitis; however, cases imported from endemic areas have been found in the United States. Evidence supports a zoonotic reservoir for HEV primarily in swine (but also in deer, camels, and rabbits), which may account for the mostly subclinical infections primarily of genotypes 3 and 4 in nonendemic areas. Globally, ~60% of swine have evidence of prior HEV infection, and 13% have active HEV infection. A previously unrecognized high distribution of HEV infection, linked to uncooked or undercooked pork-product ingestion, has been discovered in western Europe (e.g., in Germany, an estimated annual incidence of 300,000 cases and a 17% prevalence of anti-HEV among adults; in France, a 22% prevalence of anti-HEV in healthy blood donors).

CLINICAL AND LABORATORY FEATURES

Symptoms and Signs Acute viral hepatitis occurs after an incubation period that varies according to the responsible agent. Generally, incubation periods for hepatitis A range from 15 to 45 days (mean, 4 weeks), for hepatitis B and D from 30 to 180 days (mean, 8–12 weeks), for hepatitis C from 15 to 160 days (mean, 7 weeks), and for hepatitis E from 14 to 60 days (mean, 5–6 weeks). The *prodromal symptoms* of acute viral hepatitis are systemic and quite variable. Constitutional symptoms—which may include anorexia, nausea and vomiting, diarrhea, fatigue, malaise, arthralgias, myalgias, headache, and photophobia—tend to precede the onset of jaundice (if jaundice occurs). The nausea, vomiting, and anorexia are frequently associated with alterations in olfaction and taste. A low-grade fever between 38° and 39°C (100°–102°F) is more often present in hepatitis A and E than in hepatitis B or C, except when hepatitis B is heralded by a serum sickness–like syndrome; rarely, a fever of 39.5°–40°C (103°–104°F) may accompany the constitutional symptoms. Dark urine and clay-colored stools may be noticed by the patient from 1–5 days before the onset of clinical jaundice.

Many patients will not become jaundiced (anicteric hepatitis), and the likelihood of jaundice varies by virus (e.g., substantially more common in acute hepatitis B than in acute hepatitis C). With the onset of *clinical jaundice*, the constitutional prodromal symptoms usually diminish, but in some patients, mild weight loss (2.5–5 kg) is common and may continue during the entire icteric phase. The liver becomes enlarged and tender and may be associated with right upper quadrant pain and discomfort. Infrequently, patients present with a cholestatic picture, suggesting extrahepatic biliary obstruction. Splenomegaly and cervical adenopathy are present in 10–20% of patients with acute hepatitis. Rarely, a few spider angiomas appear during the icteric phase and disappear during convalescence. During the *recovery phase*, constitutional symptoms disappear, but usually some liver enlargement and abnormalities in liver biochemical tests are still evident. The duration of the posticteric phase is variable, ranging from 2 to 12 weeks, and is usually more prolonged in acute hepatitis B and C. Complete clinical and biochemical recovery is to be expected 1–2 months after all cases of hepatitis A and E and 3–4 months after the onset of jaundice in three-quarters of uncomplicated, self-limited cases of hepatitis B and C (among healthy adults, acute hepatitis B is self-limited in 95–99%, whereas hepatitis C is self-limited in only ~15–20%). In the remainder, biochemical recovery may be delayed.

Infection with HDV can occur in the presence of acute (coinfection) or chronic (superinfection) HBV infection; the duration of HBV infection determines the duration of HDV infection. When acute HDV and HBV infections occur simultaneously (coinfection), clinical and biochemical features may be indistinguishable from those of HBV infection alone, and the clinical severity can range from mild to fulminant disease. When acute HDV infection occurs in a patient with preexisting *chronic* HBV infection (superinfection), the clinical presentation is more commonly severe, and the HDV superinfection mirrors the chronic HBV infection in becoming persistent. In such cases, the HDV superinfection appears as a clinical exacerbation or an episode resembling acute viral hepatitis in someone already chronically infected with HBV. Superinfection with HDV in a patient with chronic hepatitis B often leads to clinical deterioration (see below).

In addition to superinfections with other hepatitis agents, acute hepatitis-like clinical events in persons with chronic hepatitis B may accompany spontaneous HBeAg to anti-HBe seroconversion or spontaneous reactivation (i.e., reversion from relatively nonreplicative to replicative infection). Occasionally, acute clinical exacerbations of chronic hepatitis B may represent the emergence of a precore mutant (see "Virology and Etiology"), and the subsequent course in such patients may be characterized by periodic exacerbations. Reactivations can occur as well in therapeutically immunosuppressed patients with chronic HBV infection when cytotoxic/immunosuppressive drugs are withdrawn; in these cases, restoration of immune competence is thought to allow resumption of previously checked cell-mediated immune cytolysis of HBV-infected hepatocytes. Cancer-directed therapies can also result in HBV or HCV reactivation, with the greatest risk from the B-cell (CD20)-depleting antibody rituximab in HBV. Immune checkpoint inhibitors (ICIs) have been an important new class of antitumor therapies, and the impact of ICI therapy on hepatitis caused by chronic HBV or HCV infection is not yet well characterized, but acute exacerbations of chronic viral hepatitis have been reported and may be related to restoring function of "exhausted" T-cell populations. Of interest, however, in pilot studies of ICIs given for chronic hepatitis B, restoration of T-cell function has also been associated with functional cure in a limited subset.

Laboratory Features The serum aminotransferases aspartate aminotransferase (AST) and ALT (previously designated SGOT and SGPT) increase to a variable degree during the prodromal phase of acute viral hepatitis and precede the rise in bilirubin level (Figs. 350-2 and 350-4). The level of these enzymes, however, does not correlate well with the degree of liver cell damage. Peak levels vary from ~400 to ~4000 IU or more; these levels are usually reached at the time the patient is clinically icteric and diminish progressively during the recovery phase of acute hepatitis. The diagnosis of anicteric hepatitis is based on clinical features and on aminotransferase elevations.

Jaundice is usually visible in the sclera or skin when the serum bilirubin value is >43 μmol/L (2.5 mg/dL). When jaundice appears, the serum bilirubin typically rises to levels ranging from 85 to 340 μmol/L (5–20 mg/dL). The serum bilirubin may continue to rise despite falling serum aminotransferase levels. In most instances, the total bilirubin is equally divided between the conjugated and unconjugated fractions. Bilirubin levels >340 μmol/L (20 mg/dL) extending and persisting late into the course of viral hepatitis are more likely to be associated with severe disease. In certain patients with underlying hemolytic anemia, however, such as glucose-6-phosphate dehydrogenase deficiency and sickle cell anemia, a high serum bilirubin level is common, resulting from superimposed hemolysis. In such patients, bilirubin levels >513 μmol/L (30 mg/dL) have been observed and are not necessarily associated with a poor prognosis.

Neutropenia and lymphopenia are transient and are followed by a relative lymphocytosis. Atypical lymphocytes (varying between 2 and 20%) are common during the acute phase. Measurement of the prothrombin time (PT) is important in patients with acute viral hepatitis, because a prolonged value may reflect a severe hepatic synthetic defect, signify extensive hepatocellular necrosis, and indicate a worse prognosis. Occasionally, a prolonged PT may occur with only mild increases in the serum bilirubin and aminotransferase levels. Prolonged nausea and vomiting, inadequate carbohydrate intake, and poor hepatic glycogen reserves may contribute to hypoglycemia noted occasionally in patients with severe viral hepatitis. Serum alkaline phosphatase may be normal or only mildly elevated, whereas a fall in serum albumin is uncommon in uncomplicated acute viral hepatitis. In some patients, mild and transient steatorrhea has been noted, as well as slight microscopic hematuria and minimal proteinuria.

A diffuse but mild elevation of the γ globulin fraction is common during acute viral hepatitis. Serum IgG and IgM levels are elevated in about one-third of patients during the acute phase of viral hepatitis, but the serum IgM level is elevated more characteristically during acute hepatitis A. During the acute phase of viral hepatitis, antibodies to smooth muscle and other cell constituents may be present, and low titers of rheumatoid factor, nuclear antibody, and heterophile antibody can also be found occasionally. In hepatitis C and D, antibodies to LKM may occur; however, the species of LKM antibodies in the two types of hepatitis are different from each other as well as from the LKM antibody species characteristic of autoimmune hepatitis type 2 **(Chap. 352)**. The autoantibodies in viral hepatitis are nonspecific and can also be associated with other viral and systemic diseases. In contrast, virus-specific antibodies, which appear during and after hepatitis virus infection, are serologic markers of diagnostic importance.

As described above, serologic tests are available routinely with which to establish a diagnosis of hepatitis A, B, D, and C. Tests for fecal or serum HAV are not routinely available. Therefore, a diagnosis of hepatitis A is based on detection of IgM anti-HAV during acute illness (Fig. 350-2). Worth considering, false-positive IgM anti-HAV results have been observed in rheumatologic diseases (especially those with rheumatoid factor, which can mediate nonspecific binding in solid-phase immunoassays) and, very rarely, in other acute viral infections, such as with Epstein-Barr virus (EBV).

A diagnosis of HBV infection can usually be made by detection of HBsAg in serum. Infrequently, levels of HBsAg are too low to be detected during acute HBV infection, even with contemporary, highly sensitive immunoassays. In such cases, the diagnosis can be established by the presence of IgM anti-HBc.

The titer of HBsAg bears little relation to the severity of clinical disease. Indeed, an inverse correlation exists between the serum concentration of HBsAg and the degree of liver cell damage. For example, titers are highest in immunosuppressed patients, lower in patients with chronic liver disease (but higher in mild chronic than in severe chronic hepatitis), and very low in patients with acute fulminant hepatitis. These observations suggest that in hepatitis B the degree of liver cell damage and the clinical course are related to variations in the patient's immune response to HBV rather than to the amount of circulating HBsAg. In immunocompetent persons, however, a correlation exists between markers of HBV *replication* and liver injury (see below).

Another important serologic marker in patients with hepatitis B is HBeAg. Its principal clinical usefulness is as an indicator of relative infectivity. Because HBeAg is invariably present during early acute hepatitis B, HBeAg testing is indicated primarily in chronic infection.

In patients with hepatitis B surface antigenemia of unknown duration (e.g., blood donors found to be HBsAg positive), testing for IgM anti-HBc may be useful to distinguish between acute or recent infection (IgM anti-HBc positive) and chronic HBV infection (IgM anti-HBc negative, IgG anti-HBc positive). A false-positive test for IgM anti-HBc may be encountered in patients with high-titer rheumatoid factor. Also, IgM anti-HBc may be reexpressed during acute reactivation of chronic hepatitis B.

Anti-HBs is rarely detectable in the presence of HBsAg in patients with *acute* hepatitis B, but up to 10% of persons with *chronic* HBV infection may harbor anti-HBs. Variability of the *a* determinant of the S protein is thought to be central to this anti-HBs. The antibody is directed not against the common group determinant, *a*, but against the relatively closely related heterotypic subtype determinant (e.g., HBsAg of subtype *ad* with anti-HBs of subtype *y*). Clinically, patients with both HBsAg and anti-HBs tend to have higher aminotransferase levels, are older, and are more likely to have advanced fibrosis or cirrhosis (primarily in HBeAg-negative patients). This serologic pattern in patients with chronic hepatitis B is not a harbinger of imminent HBsAg clearance; however, in patients clearing HBsAg as acute hepatitis B resolves, infrequently, transiently, both HBsAg and anti-HBs may be detected simultaneously.

After immunization with hepatitis B vaccine, which consists of HBsAg alone, anti-HBs is the only serologic marker to appear. The commonly encountered serologic patterns of hepatitis B and their interpretations are summarized in **Table 350-5**. Tests for the detection of HBV DNA in liver and serum are now available. Like HBeAg, serum HBV DNA is an indicator of HBV replication, but tests for HBV DNA are more sensitive and quantitative. First-generation hybridization assays for HBV DNA had a sensitivity of 10^5–10^6 virions/mL, a relative threshold below which infectivity and liver injury are limited and HBeAg is usually undetectable. Currently, testing for HBV DNA has shifted from insensitive hybridization assays to amplification assays (e.g., the PCR-based assay, which can detect as few as 10 or 100 virions/mL); among the commercially available PCR assays, the most useful are those with the highest sensitivity (5–10 IU/mL) and the largest dynamic range (100–10^9 IU/mL). With increased sensitivity, amplification assays remain reactive well below the current 10^3–10^4 IU/mL threshold for infectivity and liver injury. These markers are useful in following the course of HBV replication in patients with chronic hepatitis B receiving antiviral chemotherapy **(Chap. 352)**. Except for the early decades of life after perinatally acquired HBV infection (see above), in immunocompetent adults with chronic hepatitis B, a general correlation exists between the level of HBV replication, as reflected by the level of serum HBV DNA, and the degree of liver injury. High-serum HBV DNA levels, increased expression of viral antigens, and necroinflammatory activity in the liver go hand in hand unless immunosuppression interferes with cytolytic T-cell responses to virus-infected cells; reduction of HBV replication with antiviral drugs tends to be accompanied by an improvement in liver histology. Among patients with chronic hepatitis B, high levels of HBV DNA increase the risk of cirrhosis, hepatic decompensation, and hepatocellular carcinoma (see "Complications and Sequelae").

A diagnosis of acute hepatitis C hinges upon the presence of anti-HCV and/or HCV RNA with or without the presence of anti-HCV. Compared with the other viral hepatitides, assays specific for IgM anti-HCV, never found to be of clinical value, are lacking; available highly sensitive assays detect IgG anti-HCV. While, in the past, with early-generation immunoassays, up to 12 weeks could elapse before anti-HCV became detectable during acute hepatitis C, with contemporary, late-generation assays, anti-HCV is readily detectable during early acute hepatitis C. Thus, when contemporary immunoassays are used, anti-HCV can be detected in acute hepatitis C during the initial phase of elevated aminotransferase activity and remains detectable both after recovery from acute infection as well as during chronic infection. As alluded to above, nonspecificity can confound immunoassays for anti-HCV, especially in persons with a low prior probability of infection, such as volunteer blood donors, or in persons with circulating rheumatoid factor, which can bind nonspecifically to assay reagents; testing for HCV RNA can be used in such settings to distinguish between true-positive and false-positive anti-HCV determinations. While spontaneous recovery from acute HCV infection occurs no more often than 15–20% of the time, persons who do recover will have

TABLE 350-5 Commonly Encountered Serologic Patterns of Hepatitis B Infection

HBsAg	ANTI-HBs	ANTI-HBc	HBeAg	ANTI-HBe	INTERPRETATION
+	–	IgM	+	–	Acute hepatitis B, high infectivity[a]
+	–	IgG	+	–	Chronic hepatitis B, high infectivity
+	–	IgG	–	+	1. Late acute or chronic hepatitis B, low infectivity 2. HBeAg-negative ("precore-mutant") hepatitis B (chronic or, rarely, acute)
+	+	+	+/–	+/–	1. HBsAg of one subtype and heterotypic anti-HBs (common) 2. Process of seroconversion from HBsAg to anti-HBs (rare)
–	–	IgM	+/–	+/–	1. Acute hepatitis B[a] 2. Anti-HBc "window"
–	–	IgG	–	+/–	1. Low-level hepatitis B carrier 2. Hepatitis B in remote past
–	+	IgG	–	+/–	Recovery from hepatitis B
–	+	–	–	–	1. Immunization with HBsAg (after vaccination) 2. Hepatitis B in the remote past (?) 3. False-positive

[a] IgM anti-HBc may reappear during acute reactivation of chronic hepatitis B.

Note: See text for abbreviations.

a persistent anti-HCV but are not protected against reinfection. In high-risk groups, notably PWID, a pattern of *positive* anti-HCV and HCV RNA may reflect acute reinfection rather than chronic infection. In this setting, prior laboratory tests and clinical history are critical in differentiating between chronic and repeat acute infections.

Assays for HCV RNA are the most sensitive tests for HCV infection and represent the "gold standard" in establishing a diagnosis of hepatitis C. HCV RNA can be detected even before acute elevation of aminotransferase activity and before the appearance of anti-HCV in patients with acute hepatitis C. In addition, HCV RNA remains detectable indefinitely, although at varying levels, in patients with chronic hepatitis C. In the very small minority of patients with hepatitis C who lack anti-HCV, a diagnosis can be supported by detection of HCV RNA. In the rare instance when all these tests are negative and the patient has a well-characterized case of hepatitis after percutaneous exposure to blood or blood products, a diagnosis of hepatitis caused by an unidentified agent can be entertained.

Amplification techniques are required to detect HCV RNA. Currently, such target amplification (i.e., synthesis of multiple copies of the viral genome) is achieved by PCR, in which the viral RNA is reverse transcribed to complementary DNA and then amplified by repeated cycles of DNA synthesis. Quantitative PCR assays provide a measurement of relative "viral load"; current PCR assays have a sensitivity of 10 (lower limit of detection) to 25 (lower limit of quantitation) IU/mL and a wide dynamic range ($10–10^7$ IU/mL). Determination of HCV RNA level is not a reliable marker of disease severity or prognosis but is helpful in predicting relative responsiveness to antiviral therapy. The same is true for determinations of HCV genotype **(Chap. 352)**. Of course, HCV RNA monitoring during and after antiviral therapy is the *sine qua non* for determining on-treatment and durable responsiveness.

A proportion of patients with hepatitis C have isolated anti-HBc in their blood, a reflection of a common risk in certain populations of exposure to multiple bloodborne hepatitis agents. The anti-HBc in such cases is almost invariably of the IgG class and usually represents HBV infection in the remote past (HBV DNA undetectable); it rarely represents current HBV infection with low-level virus carriage. Detectable anti-HCV in the absence of HCV RNA signifies spontaneous or therapeutically induced recovery from ("cured") hepatitis C.

The presence of HDV infection can be identified by demonstrating intrahepatic HDV antigen or, more practically, an anti-HDV (IgM or IgG) seroconversion (a rise in titer of anti-HDV or de novo appearance of anti-HDV). Serum HDV RNA or antigen can be detected early during acute coinfection but is often transient and detectable only briefly, if at all. Because anti-HDV is often undetectable once HBsAg disappears, retrospective serodiagnosis of acute self-limited, simultaneous HBV and HDV infection is difficult. Early diagnosis of acute infection may be hampered by a delay of up to 30–40 days in the appearance of IgG anti-HDV.

When a patient presents with acute hepatitis and has HBsAg and anti-HDV in serum, determination of the class of anti-HBc is helpful in establishing the relationship between infection with HBV and HDV. Although IgM anti-HBc does not distinguish *absolutely* between acute and chronic HBV infection, its presence is a reliable indicator of recent infection and its absence a reliable indicator of infection in the remote past. In simultaneous acute HBV and HDV infections, IgM anti-HBc will be detectable, whereas in acute HDV infection superimposed on chronic HBV infection, anti-HBc will be of the IgG class. Assays for HDV RNA, available in specialized laboratories and yet to be standardized, can be used to confirm HDV infection and to monitor treatment during chronic infection.

The serologic/virologic course of events during acute hepatitis E is entirely analogous to that of acute hepatitis A, with brief fecal shedding of virus and viremia and an early IgM anti-HEV response that predominates during approximately the first 3 months but is eclipsed thereafter by long-lasting IgG anti-HEV. Diagnostic tests of varying reliability for hepatitis E are commercially available outside the United States; in the United States, although tests for HEV infection are not approved by the FDA, reliable diagnostic serologic/virologic assays can be performed at the CDC or other commercial or academic laboratories.

Liver biopsy is rarely necessary or indicated in acute viral hepatitis, except when the diagnosis is questionable or when clinical evidence suggests an alternate diagnosis of chronic hepatitis (e.g., autoimmune hepatitis).

A diagnostic algorithm can be applied in the evaluation of cases of acute viral hepatitis. A patient with acute hepatitis should undergo five serologic tests: HBsAg, IgM anti-HAV, IgM anti-HBc, anti-HCV, and HCV RNA **(Table 350-6)**. The presence of HBsAg, with or without IgM anti-HBc, represents HBV infection. If IgM anti-HBc is present, the HBV infection is considered acute (or, less commonly, an acute reactivation); if IgM anti-HBc is absent, the HBV infection is considered chronic. A diagnosis of acute hepatitis B can be made in the absence of HBsAg when IgM anti-HBc is detectable. A diagnosis of acute hepatitis A is based on the presence of IgM anti-HAV. If IgM anti-HAV coexists with HBsAg, a diagnosis of simultaneous HAV and HBV infections can be made; if IgM anti-HBc (with or without HBsAg) is detectable, the patient has simultaneous acute hepatitis A and B, and if IgM anti-HBc is undetectable, the patient has acute hepatitis A superimposed on chronic HBV infection. In a patient with an acute hepatitis illness, the presence of anti-HCV supports a diagnosis of acute hepatitis C. In the past, sensitivity of anti-HCV assays was limited during acute hepatitis, leading to a delay in the appearance anti-HCV; however, contemporary assays detect anti-HCV reliably during acute hepatitis. If acute hepatitis C is suspected but anti-HCV is undetectable, testing for HCV RNA helps to establish the diagnosis. The serologic diagnosis of acute hepatitis C or B may be challenging, for example, when an acute hepatitis-like illness occurs in someone already infected chronically with HBV or HCV. Still, when encountering a person with acute hepatitis, a series of standard serologic assays for hepatitis A, B, and C is a reliable first step. If these assays are all negative, testing for hepatitis E, especially in the presence of risk factors, is available from specialty laboratories or the CDC. Acute hepatitis D belongs in the differential diagnosis of acute hepatitis, but negative serologic testing for hepatitis B excludes hepatitis D infection. In the setting of serologically documented acute

TABLE 350-6 Simplified Diagnostic Approach in Patients Presenting with Acute Hepatitis

SEROLOGIC TESTS OF PATIENT'S SERUM				
HBsAg	**IgM ANTI-HAV**	**IgM ANTI-HBc**	**ANTI-HCV HCV RNA**	**DIAGNOSTIC INTERPRETATION**
+	–	+	–	Acute hepatitis B
+	–	–	–	Chronic hepatitis B
+	+	–	–	Acute hepatitis A superimposed on chronic hepatitis B
+	+	+	–	Acute hepatitis A and B
–	+	–	–	Acute hepatitis A
–	+	+	–	Acute hepatitis A and B (HBsAg below detection threshold)
–	–	+	–	Acute hepatitis B (HBsAg below detection threshold)
–	–	–	Either +	Acute hepatitis C (vs. chronic HCV)

Note: See text for abbreviations.

hepatitis B, patients who are at high risk for hepatitis D are candidates to be tested for anti-HDV and/or HDV RNA. Absence of all serologic markers is consistent with a diagnosis of "non-A-E," hepatitis (no other proven human hepatitis viruses have been identified). As noted above, an outbreak of severe acute liver injury and liver failure in children was found to be associated with adenovirus infection, and elevations in aminotransferase levels are observed as well in a variety of other acute viral illnesses.

In patients with chronic hepatitis, initial testing should consist of HBsAg and anti-HCV. Anti-HCV supports and HCV RNA testing establishes the diagnosis of chronic hepatitis C. If a serologic diagnosis of chronic hepatitis B is made, testing for HBeAg and anti-HBe is indicated to evaluate relative infectivity. Testing for HBV DNA in such patients provides a more quantitative and sensitive measure of the level of virus replication and therefore is very helpful during antiviral therapy **(Chap. 352)**. In patients with established chronic hepatitis B and normal aminotransferase activity in the absence of HBeAg, serial testing over time is often required to distinguish between the inactive carrier state and HBeAg-negative chronic hepatitis B with fluctuating virologic and necroinflammatory activity. In persons with hepatitis B, testing for anti-HDV is useful in those with severe and fulminant disease, with severe chronic disease, with chronic hepatitis B and acute hepatitis-like exacerbations, with frequent percutaneous exposures (or known HBV acquisition from percutaneous exposure), and from areas where HDV infection is endemic.

PROGNOSIS

Virtually all previously healthy patients with hepatitis A recover completely with no clinical sequelae. Similarly, in acute hepatitis B, 95–99% of previously healthy adults have a favorable course and recover completely. Certain clinical and laboratory features, however, suggest a more complicated and protracted course. Patients of advanced age and with serious underlying medical disorders may have a prolonged course and are more likely to experience severe hepatitis. Initial presenting features such as ascites, peripheral edema, and symptoms of hepatic encephalopathy suggest a poorer prognosis. In addition, a prolonged PT, low serum albumin level, hypoglycemia, and very high serum bilirubin values suggest severe hepatocellular disease. Patients with these clinical and laboratory features deserve prompt hospital admission. The case-fatality rate in hepatitis A and B is very low (~0.1%) but is increased by advanced age and underlying debilitating disorders. Among patients ill enough to be hospitalized for acute hepatitis B, the fatality rate is 1%. Hepatitis C is less severe during the acute phase than hepatitis B and is more likely to be anicteric; fatalities are rare, but the precise case-fatality rate is not known. Interestingly, while spontaneous resolution of HCV is the exception rather than the rule, symptomatic infection (and icteric hepatitis in particular) is associated with a higher likelihood of viral containment. In outbreaks of waterborne hepatitis E (genotypes 1 and 2) in India and Asia, the case-fatality rate is 1–2% and up to 10–20% in pregnant women. Contributing to fulminant hepatitis E in endemic countries (but only very rarely or not at all in nonendemic countries) are instances of acute hepatitis E superimposed on underlying chronic liver disease ("acute-on-chronic" liver disease). Patients with simultaneous acute hepatitis B and D coinfection do not necessarily experience a higher mortality rate than do patients with acute hepatitis B alone; however, in several outbreaks of acute simultaneous HBV and HDV infection among injection drug users, the case-fatality rate was ~5%. When HDV superinfection occurs in a person with chronic hepatitis B, the likelihood of fulminant hepatitis and death is increased substantially. Although the case-fatality rate for hepatitis D is not known definitively, in outbreaks of severe HDV superinfection in isolated populations with a high hepatitis B carrier rate ("Lábrea fever"), a mortality rate >20% has been recorded.

COMPLICATIONS AND SEQUELAE

A small proportion of patients with hepatitis A experience *relapsing hepatitis* weeks to months after apparent recovery from acute hepatitis. Relapses are characterized by recurrence of symptoms, aminotransferase elevations, occasional jaundice, and fecal excretion of HAV. Another unusual variant of acute hepatitis A is *cholestatic hepatitis*, characterized by protracted cholestatic jaundice and pruritus. Rarely, liver test abnormalities persist for many months, even up to 1 year. Even when these complications occur, hepatitis A remains self-limited and does not progress to chronic liver disease.

During the prodromal phase of acute hepatitis B, a serum sickness–like syndrome characterized by arthralgia or arthritis, rash, angioedema, and, rarely, hematuria and proteinuria may develop in 5–10% of patients. This syndrome occurs before the onset of clinical jaundice, and these patients are often diagnosed erroneously as having rheumatologic diseases. The diagnosis can be established by measuring serum aminotransferase levels, which are almost invariably elevated, and serum HBsAg. In patients with chronic hepatitis B or C, immune complex and other extrahepatic manifestations may occur, including metabolic disorders (including insulin resistance and other manifestations of the metabolic syndrome), dermatologic disorders (e.g., porphyria cutanea tarda and lichen planus in chronic hepatitis C), lymphoproliferative and rheumatologic disorders, and, in addition to hepatocellular carcinoma, nonliver malignancies **(Chap. 352)**. In addition, population studies of patients with chronic hepatitis C have shown a convincing increase in cardiovascular and cerebrovascular disease, renal disease, and mental health and cognitive disorders.

The most feared complication of acute viral hepatitis is *fulminant hepatitis* (massive hepatic necrosis); fortunately, this is a rare event. Fulminant hepatitis is seen primarily in hepatitis B, D, and E, but rare fulminant cases of hepatitis A occur primarily in older adults and in persons with underlying chronic liver disease, including, according to some reports, chronic hepatitis B and C. Hepatitis B accounts for >50% of fulminant cases of viral hepatitis, a sizable proportion of which are associated with HDV infection and another proportion with underlying chronic hepatitis C. Fulminant hepatitis is hardly ever seen in hepatitis C, but hepatitis E, as noted above, can be complicated by fatal fulminant hepatitis in 1–2% of all cases and in up to 20% of cases in pregnant women. Patients usually present with signs and symptoms of encephalopathy that may evolve to deep coma. The liver is usually small and the PT excessively prolonged. The combination of rapidly shrinking liver size, rapidly rising bilirubin level, and marked prolongation of the PT, even as aminotransferase levels fall, together with clinical signs of confusion, disorientation, somnolence, ascites, and edema, indicates that the patient has hepatic failure with encephalopathy. Cerebral edema is common; brainstem compression, gastrointestinal bleeding, sepsis, respiratory failure, cardiovascular collapse, and renal failure are terminal events. The mortality rate can be exceedingly high (>80% in patients with deep coma) without liver transplantation, but patients who survive may have a complete biochemical and histologic recovery **(Chap. 356)**. Documenting the disappearance of HBsAg after apparent clinical recovery from acute hepatitis B is particularly important. Before laboratory methods were available to distinguish between acute hepatitis and acute hepatitis–like exacerbations (*spontaneous reactivations*) of chronic hepatitis B, observations suggested that ~10% of previously healthy patients remained HBsAg positive for >6 months after the onset of clinically apparent acute hepatitis B. One-half of these persons cleared the antigen from their circulations during the next several years, but the other 5% remained chronically HBsAg positive. More recent observations suggest that the true rate of chronic infection after clinically apparent acute hepatitis B is as low as 1% in normal, immunocompetent, young adults. Earlier, higher estimates may have been confounded by inadvertent inclusion of acute exacerbations in chronically infected patients; these patients, chronically HBsAg positive before exacerbation, were unlikely to seroconvert to HBsAg negative thereafter. Whether the rate of chronicity is 10% or 1%, such patients have IgG anti-HBc in serum; anti-HBs is either undetected or detected at low titer against the opposite subtype specificity of the antigen (see "Laboratory Features"). These patients may (1) be inactive carriers; (2) have low-grade, mild chronic hepatitis; or (3) have moderate to severe chronic hepatitis with or without cirrhosis. The likelihood of remaining chronically infected after acute HBV infection is especially high among neonates, persons with Down syndrome, chronically hemodialyzed patients, and immunosuppressed patients, including persons with HIV infection.

Chronic hepatitis is an important late complication of acute hepatitis B occurring in a small proportion of patients with acute disease but more common in those who present with chronic infection without having experienced an acute illness, as occurs typically after neonatal infection or after infection in an immunosuppressed host **(Chap. 352)**. The following clinical and laboratory features suggest progression of acute hepatitis to chronic hepatitis: (1) lack of complete resolution of clinical symptoms of anorexia, weight loss, fatigue, and the persistence of hepatomegaly; (2) the presence of bridging/interface or multilobular hepatic necrosis on liver biopsy during protracted, severe acute viral hepatitis; (3) failure of the serum aminotransferase, bilirubin, and globulin levels to return to normal within 6–12 months after the acute illness; and (4) the persistence of HBeAg for >3 months or HBsAg for >6 months after acute hepatitis.

Although acute hepatitis D infection does not increase the likelihood of chronicity of simultaneous acute hepatitis B, hepatitis D has the potential for contributing to the severity of chronic hepatitis B. Hepatitis D superinfection can transform inactive or mild chronic hepatitis B into severe, progressive chronic hepatitis and cirrhosis; it also can accelerate the course of chronic hepatitis B and accelerate the risk of hepatocellular carcinoma. Some HDV superinfections in patients with chronic hepatitis B lead to fulminant hepatitis. As defined in longitudinal studies over three decades, the annual rate of cirrhosis in patients with chronic hepatitis D is 4%. Although HDV and HBV infections are associated with severe liver disease, mild hepatitis and even inactive carriage have been identified in some patients, and the disease may become indolent beyond the early years of infection.

After acute HCV infection, the likelihood of remaining chronically *infected* approaches 85–90%. Although many patients with chronic hepatitis C have no symptoms, cirrhosis may develop in as many as 20% within 10–20 years of acute illness; in prior series of cases reported by referral centers, cirrhosis has been reported in as many as 50% of patients with chronic hepatitis C. Among cirrhotic patients with chronic hepatitis C, the annual risk of hepatic decompensation is ~4%; the annual risk of hepatocellular carcinoma in cirrhotic patients with chronic hepatitis C has been reported in a range between 1 and 4% and, in a recent definitive report, was 3.36%. With the advent of highly effective DAA therapy during the second decade of the 21st century, a decline in HCV-related cirrhosis and hepatocellular carcinoma has been observed. Progression of chronic hepatitis C may be influenced by advanced age of acquisition, long duration of infection, immunosuppression, coexisting excessive alcohol use, concomitant hepatic steatosis, other hepatitis virus infection, or HIV co-infection. In fact, instances of severe and rapidly progressive chronic hepatitis B and C are recognized with increasing frequency in patients with HIV infection **(Chap. 208)**. In contrast, neither HAV nor HEV causes chronic liver disease in immunocompetent hosts; however, cases of chronic hepatitis E (including cirrhosis and end-stage liver disease and even hepatocellular carcinoma) have been observed in immunosuppressed organ-transplant recipients, persons receiving cytotoxic chemotherapy, and persons with HIV infection. Among patients with chronic hepatitis (e.g., caused by hepatitis B or C, alcohol, etc.) in endemic countries, hepatitis E has been reported as the cause of acute-on-chronic liver failure; however, in most experiences among patients from nonendemic countries, HEV has not been found to contribute commonly to hepatic decompensation in patients with chronic hepatitis.

Persons with chronic hepatitis B, particularly those infected in infancy or early childhood and especially those with HBeAg and/or high-level HBV DNA, have an enhanced risk of hepatocellular carcinoma. The risks of cirrhosis and hepatocellular carcinoma increase with the level of HBV replication. The annual rate of hepatocellular carcinoma in patients with chronic hepatitis D and cirrhosis is ~3%.

Rare complications of acute viral hepatitis include pancreatitis, myocarditis, atypical pneumonia, aplastic anemia, transverse myelitis, peripheral neuropathy, and Guillain-Barré syndrome. In children, hepatitis B may present rarely with anicteric hepatitis, a nonpruritic papular rash of the face, buttocks, and limbs, and lymphadenopathy (papular acrodermatitis of childhood or Gianotti-Crosti syndrome).

Rarely, autoimmune hepatitis **(Chap. 352)** can be triggered by a bout of otherwise self-limited acute hepatitis, as reported after acute hepatitis A, B, and C.

DIFFERENTIAL DIAGNOSIS

Viral diseases such as infectious mononucleosis (EBV); those due to cytomegalovirus, herpes simplex virus, adenovirus, and coxsackieviruses; and toxoplasmosis may share certain clinical features with viral hepatitis and cause elevations in serum aminotransferase and, less commonly, in serum bilirubin levels. Tests such as the differential heterophile and serologic tests for these agents may be helpful in the differential diagnosis if HBsAg, anti-HBc, IgM anti-HAV, and anti-HCV (and anti-HEV) determinations are negative. Aminotransferase elevations can accompany almost any systemic viral infection, including the coronavirus SARS-CoV-2 (~10% of all cases and up to half of severe cases); other rare causes of liver injury confused with viral hepatitis are infections with *Leptospira*, *Candida*, *Brucella*, *Mycobacteria*, and *Pneumocystis*. A complete drug history is particularly important, because many drugs and certain anesthetic agents can produce a picture of either acute hepatitis or cholestasis **(Chap. 351)**. Equally important is a history of unexplained "repeated episodes" of acute hepatitis. This history should alert the physician to the possibility that the underlying disorder is chronic hepatitis, for example, autoimmune hepatitis **(Chap. 352)**. Alcohol-associated hepatitis (AH), which can cause an acute icteric hepatitis, must also be considered, but usually the serum aminotransferase levels are not as markedly elevated (generally not much above 200 IU/mL, almost always below 400 IU/mL) in AH, and a heavy drinking history must be present (at least 40 g/d in women and 60 g/d in men for at least 6 months). The finding on liver biopsy of fatty infiltration, a neutrophilic inflammatory reaction, and "alcoholic hyaline" would be consistent with alcohol-induced rather than viral liver injury. Because acute hepatitis may present with right upper quadrant abdominal pain, nausea and vomiting, fever, and icterus, it is often confused with acute cholecystitis, common duct stone, or ascending cholangitis. Patients with acute viral hepatitis may tolerate surgery poorly, and surgical risk is particularly high in the elderly and those with underlying chronic liver disease; therefore, excluding biliary disease clinically and with imaging is important. In confusing cases, a percutaneous liver biopsy may be helpful. Viral hepatitis in the elderly is often misdiagnosed as obstructive jaundice resulting from a common duct stone or carcinoma of the pancreas. Another clinical constellation that may mimic acute hepatitis is right ventricular failure with passive hepatic congestion or hypoperfusion syndromes, such as those associated with shock, severe hypotension, and severe left ventricular failure. Also included in this general category is any disorder that interferes with venous return to the heart, such as right atrial myxoma, constrictive pericarditis, hepatic vein occlusion (Budd-Chiari syndrome), or venoocclusive disease. Clinical features are usually sufficient to distinguish among these vascular disorders and viral hepatitis. Acute fatty liver of pregnancy, cholestasis of pregnancy, eclampsia, and the HELLP (*h*emolysis, *e*levated *l*iver tests, and *l*ow *p*latelets) syndrome can be confused with viral hepatitis during pregnancy. Very rarely, malignancies metastatic to the liver can mimic acute or even fulminant viral hepatitis. Occasionally, genetic or metabolic liver disorders (e.g., Wilson's disease, α_1 antitrypsin deficiency) and nonalcoholic fatty liver disease (NAFLD; now labeled metabolic dysfunction–associated steatotic liver disease [MASLD]; what used to be called nonalcoholic steatohepatitis [NASH] is now labeled metabolic dysfunction–associated steatohepatitis [MASH]) are confused with acute viral hepatitis **(Chap. 354)**. Among patients with biochemical evidence for severe liver injury, i.e., aminotransferase levels of ≥1000 IU/L, the most common causes are ischemic liver injury, drug-induced liver injury (especially caused by acetaminophen), acute viral hepatitis, and pancreaticobiliary disorders.

TREATMENT

Acute Viral Hepatitis

Most persons with acute hepatitis (especially hepatitis A, B, and E) recover spontaneously and do not require specific antiviral therapy. In hepatitis B, among previously healthy adults who present with clinically apparent acute hepatitis, recovery occurs in ~99%; therefore, antiviral therapy is not likely to improve the rate of recovery and is not required. In rare instances of severe acute hepatitis B, treatment with a nucleoside analogue at oral doses used to treat chronic hepatitis B **(Chap. 352)** is included in treatment guidelines by the American Association for the Study of Liver Diseases (AASLD) and the European Association for the Study of the Liver (EASL). Indications include fulminant liver failure (along with such therapy, referral to a liver transplantation center is advised) and a prolonged course (>4 weeks) of acute hepatitis B. Although clinical trials have not been done to establish the efficacy or duration of this approach, most authorities would recommend institution of antiviral therapy with a nucleoside analogue (entecavir or tenofovir, the most potent and least resistance-prone agents) for severe, but not mild-moderate, acute hepatitis B. Treatment should continue until 3 months after HBsAg seroconversion or 6 months after HBeAg seroconversion.

In typical cases of acute hepatitis C, recovery is rare (~15–20% in most experiences), and progression to chronic hepatitis is the rule. Patients with jaundice, those with HCV genotype 1, women, and those with earlier age of infection, lower level of HCV RNA, and HBV co-infection, are more likely to recover from acute hepatitis C, as are persons who have genetic markers associated with spontaneous recovery (favorable *IL28B* CC haplotype).

Because spontaneous recovery can occur and because most cases of acute hepatitis C are not clinically severe or rapidly progressive, during the era of interferon-based therapy, delaying antiviral therapy of acute hepatitis C for 3–6 months was recommended; however, in the current era of highly effective (95–100%) oral DAA therapy, waiting for potential spontaneous recovery is no longer advised. Instead, early treatment with a standard, full 8- to 12-week course of one of the first-line drug combinations approved for treatment of chronic hepatitis C **(Chap. 352)** is recommended for treatment of patients with acute hepatitis C.

Because of the vast reservoir of acute HCV infections acquired four to five decades ago in the 1945–1965 birth cohort, most newly recognized HCV infections are chronic. Opportunities to identify and treat patients with acute hepatitis C occur in two population subsets: (1) in health care workers who sustain hepatitis C–contaminated needle sticks (occupational accidents), monitoring for ALT elevations and the presence of HCV RNA identify acute hepatitis C in ~3%, and this group should be treated; (2) in persons who use injection drugs, the risk of acute hepatitis C has been on the rise during the previous two decades, and the epidemic of opioid use has contributed to an amplification of HCV infection among drug users. Such persons are candidates for antiviral therapy, and efforts to combine antiviral therapy with drug rehabilitation therapy have been very successful.

Notwithstanding these specific therapeutic considerations, in most cases of typical acute viral hepatitis, specific treatment generally is not necessary. Although hospitalization may be required for clinically severe illness, most patients do not require hospital care. Forced and prolonged bed rest is not essential for full recovery, but many patients will feel better with restricted physical activity. A high-calorie diet is desirable, and because many patients may experience nausea late in the day, the major caloric intake is best tolerated in the morning. Intravenous feeding is necessary in the acute stage if the patient has persistent vomiting and cannot maintain oral intake. Drugs capable of producing adverse reactions such as cholestasis and drugs metabolized by the liver should be avoided. If severe pruritus is present, the use of the bile salt–sequestering resin cholestyramine is helpful. Glucocorticoid therapy has no value in acute viral hepatitis, even in severe cases, and may be deleterious, even increasing the risk of chronicity (e.g., of acute hepatitis B).

Physical isolation of patients with hepatitis to a single room and bathroom is rarely necessary except in the case of fecal incontinence for hepatitis A and E or uncontrolled, voluminous bleeding for hepatitis B (with or without concomitant hepatitis D) and C. Because most patients hospitalized with hepatitis A excrete little, if any, HAV, the likelihood of HAV transmission from these patients during their hospitalization is low. Therefore, burdensome *enteric precautions are no longer recommended*. Although gloves should be worn when the bed pans or fecal material of patients with hepatitis A are handled, these precautions do not represent a departure from sensible procedure and contemporary universal precautions for all hospitalized patients. For patients with hepatitis B and C, emphasis should be placed on blood precautions (i.e., avoiding direct, ungloved hand contact with blood and other body fluids). Enteric precautions are unnecessary. The importance of simple hygienic precautions such as hand washing cannot be overemphasized. Universal precautions that have been adopted for all patients apply to patients with viral hepatitis. Hospitalized patients may be discharged following substantial symptomatic improvement, a significant downward trend in the serum aminotransferase and bilirubin values, and a return to normal of the PT. Mild aminotransferase elevations should not be considered contraindications to the gradual resumption of normal activity.

In *fulminant hepatitis*, the goal of therapy is to support the patient by maintenance of fluid balance, nutrition, support of circulation and respiration, control of bleeding, correction of hypoglycemia, early identification and treatment of infection, and treatment of other complications of the comatose state in anticipation of liver regeneration and repair. Glucocorticoid therapy has been shown in controlled trials to be ineffective. Likewise, exchange transfusion, plasmapheresis, human cross-circulation, porcine liver cross-perfusion, hemoperfusion, and extracorporeal liver-assist devices have not been proven to enhance survival. The cornerstone of management is meticulous intensive care with vigilant monitoring for infection (and low threshold for initiation of antibiotics and antifungals), careful mental status/cerebral edema monitoring, and management of bleeding complications. Liver transplantation is resorted to with increasing frequency, with excellent results, in patients with fulminant hepatitis **(Chap. 356)**. Fulminant hepatitis C is very rare; however, in fulminant hepatitis B, oral antiviral therapy has been used successfully, although the benefit may be limited once late-stage acute liver failure is present. In clinically severe acute hepatitis E or acute-on-chronic liver failure, successful therapy with ribavirin (600 mg twice daily, 15 mg/kg) has been reported; however, the precise dose and duration of therapy have not been established, and conflicting reports have appeared documenting the futility of ribavirin in this setting. Unfortunately, when fulminant hepatitis E occurs in pregnant women (as it does in up to 20% of pregnant women with acute hepatitis E), ribavirin, which is teratogenic, is contraindicated. In cases of hepatitis E in organ-transplant recipients, reduction in overall immunosuppressive drug doses and switching from tacrolimus to cyclosporine A have been shown to be effective, often without antiviral therapy, in achieving eradication of HEV. If a change in immunosuppression is inadequate, ribavirin treatment for 3 months has been observed to achieve a sustained virologic response in 78% of treated patients; again however, the optimal dose, duration, and efficacy of ribavirin therapy remain to be determined.

■ PROPHYLAXIS

Because application of therapy for acute viral hepatitis is limited and because chronic viral hepatitis requires prolonged and costly courses of antiviral therapy **(Chap. 352)**, emphasis is placed on prevention through immunization. The prophylactic approach differs for each of the types of viral hepatitis. In the past, immunoprophylaxis relied exclusively on passive immunization with antibody-containing globulin preparations purified by cold ethanol fractionation from the plasma of hundreds of normal donors. Currently, for hepatitis A, B, and E, active immunization with vaccines is the preferable approach to prevention.

Hepatitis A Both passive immunization with immunoglobulin (IG) and active immunization with killed vaccines are available. Passive immunization, which can be considered in the setting of recent exposure, is 80–90% effective at preventing infection and provides 12–20 weeks of protection; however, its utility is limited by availability and cost (and the competing appeal of active immunization with vaccine). For postexposure prophylaxis of intimate contacts (household, sexual, institutional) of persons with hepatitis A, the administration of 0.1 mL/kg is recommended as early after exposure as possible but must be within 14 days. Prophylaxis is not necessary for those who have already received hepatitis A vaccine, for casual contacts (office, factory, school, or hospital), for most elderly persons, who are very likely to be immune, or for those known to have anti-HAV in their serum. By the time most common-source outbreaks of hepatitis A are recognized, it is usually too late in the incubation period for IG to be effective; however, prophylaxis may have limited the frequency of secondary cases. For travelers to tropical countries, developing countries, and other areas outside standard tourist routes, IG prophylaxis had been recommended before a vaccine became available.

Recommendations for postexposure prophylaxis and for preexposure prophylaxis for international travel were updated in 2020. Currently, hepatitis A vaccine, not IG, is recommended for all persons aged ≥12 months for postexposure prophylaxis and for preexposure prophylaxis prior to international travel to HAV-endemic areas. Even though hepatitis A vaccine is indicated for children ≥12 months of age, when infants aged 6–11 months travel internationally to areas with a risk of HAV infection, they should receive the vaccine for preexposure prophylaxis; however, this travel-related dose should not be counted toward the universal childhood two-dose hepatitis A vaccine recommendation, which begins at age 12 months. For postexposure prophylaxis of persons with contraindications to hepatitis A vaccination and infants aged <12 months, the use of IG (0.1 mL/kg) should be retained. In addition, for postexposure prophylaxis in immunocompromised adults and persons with chronic liver disease, *both* hepatitis A vaccination and IG administration (0.1 mL/kg), at different IM sites, are recommended. Finally, for infants aged <6 months and for persons with contraindications to hepatitis A vaccination, preexposure prophylaxis for travel consists of IG at doses of 0.1 mg/kg for travel durations up to 1 month, 0.2 mg/kg for travel up to 2 months, and repeat 0.2 mg/kg every 2 months thereafter for the remainder of travel. Thus, except for these limited considerations, hepatitis A vaccine has supplanted IG in almost all cases for both postexposure prophylaxis and preexposure prophylaxis for travel. In general, practical use of IG at this point tends to be confined to outbreaks, in which early containment is critical. Unlike IG prophylaxis, the protection afforded by active immunization with vaccine is durable and simpler to administer.

Formalin-inactivated vaccines made from strains of HAV attenuated in tissue culture have been shown to be safe, immunogenic, and effective in preventing hepatitis A. Hepatitis A vaccines are approved for use in persons who are at least 1 year old and appear to provide adequate protection beginning 4 weeks after a primary inoculation. As noted above, for travel to an endemic area, hepatitis A vaccine is the preferred approach to *preexposure* immunoprophylaxis and provides long-lasting protection (protective levels of anti-HAV should last at least 20 years after vaccination). Shortly after its introduction, hepatitis A vaccine was recommended for children living in communities with a high incidence of HAV infection; in 1999, this recommendation was extended to include all children living in states, counties, and communities with high rates of HAV infection. As of 2006, the Advisory Committee on Immunization Practices of the U.S. Public Health Service recommended *routine hepatitis A vaccination of all children*. Other groups considered being at increased risk for HAV infection and who are candidates for hepatitis A vaccination include military personnel, populations with cyclic outbreaks of hepatitis A (e.g., Alaskan natives), employees of day-care centers and persons working in facilities for the developmentally delayed, primate handlers, laboratory workers exposed to hepatitis A or fecal specimens, and patients with chronic liver disease (including persons with aminotransferase elevations ≥2 times the upper limit of normal). Because of an increased risk of fulminant hepatitis A—observed in some experiences but not confirmed in others—among patients with chronic hepatitis C, patients with chronic hepatitis C are candidates for hepatitis A vaccination, as are persons with chronic hepatitis B, persons with HIV infection, and the expanding population of persons with nonalcoholic liver disease (MASLD). Other populations whose recognized risk of hepatitis A is increased should be vaccinated, including men who have sex with men, persons who use injection or noninjection drugs, persons experiencing homelessness, persons traveling from the United States to countries with high or intermediate hepatitis A endemicity, postexposure prophylaxis for contacts of persons with hepatitis A, and household members and other close contacts of (or anyone who anticipates close personal contact with) adopted children arriving from countries with high and moderate hepatitis A endemicity. Hepatitis A vaccine is now recommended as well for pregnant women at risk of infection or severe outcomes from infection during pregnancy. Recommendations for dose and frequency differ for the two approved vaccine preparations in the United States and the combination vaccines that include hepatitis A **(Table 350-7)**; all injections are IM. Hepatitis A vaccine has been reported to be effective in preventing secondary household and day-care center–associated cases of acute hepatitis A. In the United States, reported mortality resulting from hepatitis A declined in parallel with hepatitis A vaccine–associated reductions in the annual incidence of new infections.

Hepatitis B Until 1982, prevention of hepatitis B was based on *passive* immunoprophylaxis either with standard IG, containing modest levels of anti-HBs, or hepatitis B immunoglobulin (HBIG), containing high-titer anti-HBs. The efficacy of standard IG has never been established and remains questionable; even the efficacy of HBIG, demonstrated in several clinical trials, has been challenged, and its contribution appears to be in reducing the frequency of clinical *illness*, not in preventing *infection*. The first vaccine for *active* immunization, introduced in 1982, was prepared from purified, noninfectious, 22-nm spherical HBsAg particles derived from the plasma of healthy HBsAg carriers. In 1987, the plasma-derived vaccine was supplanted by a genetically engineered vaccine derived from recombinant yeast. The latter vaccine consists of HBsAg particles that are nonglycosylated but are otherwise indistinguishable from natural HBsAg; two recombinant vaccines were licensed for use in the United States in the 1980s (Recombivax-HB 1986; Engerix-B 1989), a third (Heplisav-B) was licensed in 2017, and a fourth recombinant vaccine (PreHevbrio, VBI Vaccines) was licensed in 2022. In the United States,

TABLE 350-7 Hepatitis A Vaccination Schedules

AGE, YEARS	NO. OF DOSES	DOSE	SCHEDULE, MONTHS
HAVRIX (GlaxoSmithKline)[a]			
1–18	2	720 ELU[b] (0.5 mL)	0, 6–12
≥19	2	1440 ELU (1 mL)	0, 6–12
VAQTA (Merck)			
1–18	2	25 units (0.5 mL)	0, 6–18
≥19	2	50 units (1 mL)	0, 6–18

[a]A combination of this hepatitis A vaccine and hepatitis B vaccine, TWINRIX, is licensed for simultaneous protection against both of these viruses among adults (age ≥18 years). Each 1-mL dose contains 720 ELU of hepatitis A vaccine and 20 µg of hepatitis B vaccine. These doses are recommended at months 0, 1, and 6.
[b]Enzyme-linked immunoassay units. [c]Combination hepatitis A and typhoid vaccines, Hepatyrix (GlaxoSmithKline) and Viatim (Sanofi Pasteur), are available, targeted primarily for travelers to endemic areas. Please consult product insert for doses and schedules

TABLE 350-8 Preexposure Hepatitis B Vaccinations

NAME	VACCINE TYPE	AGE GROUP	NO. OF DOSES	SCHEDULE, MONTH	NOTES
Engerix-B (GlaxoSmithKline)	Single antigen recombinant	From birth, not on dialysis	3	0, 1, 6	0.5-mL dose for 0–19 years 1-mL dose for ≥20 years Contraindicated with severe yeast allergy
		Adults on dialysis	4	0, 1, 2, 6	2-mL dose for adults
Recombivax HB (Merck)	Single antigen recombinant	From birth, not on dialysis	3	0, 1, 6	0.5-mL dose for 0–19 years 1-mL dose for ≥20 years Contraindicated with severe yeast allergy
Recombivax HBV dialysis formulation (Merck)	Single antigen recombinant	Adults on dialysis	3	0, 1, 6	1-mL dose for adults
Heplisav-B (Dynavax)	Single antigen adjuvanted recombinant	≥18 years	2	0, 1	0.5-mL dose Higher levels of protective antibodies, particularly in adults ≥40 years Not for use in severe yeast allergy
PreHevbrio (VBI)*	Three antigen recombinant	≥18 years	3	0, 1, 6	1-mL dose Higher rates of seroprotection in adults ≥45 years Not for use in pregnant women
Twinrix (GlaxoSmithKline)	Combined single antigen recombinant hepatitis A and hepatitis B	≥18 years	3	0, 1, 6 *Note: accelerated 4-dose regimen, days 0, 7, and 21–30, 12 months*	1-mL dose Severe allergy to yeast or any hepatitis A vaccine

*In November 2024, distribution of this vaccine in the United States was discontinued by the manufacturer, because of bankruptcy and termination of operations.

universal birth vaccination against HBV infection has been the standard of care since 2002, with more recent recommendations suggesting further that the birth dose be administered within 12 hours of delivery. For unvaccinated persons, current recommendations can be divided into those for preexposure and postexposure prophylaxis.

For *preexposure* prophylaxis against hepatitis B in persons with a high risk of exposure, three IM (deltoid, not gluteal) injections of hepatitis B vaccine are recommended at 0, 1, and 6 months (other, optional schedules are summarized in **Table 350-8**). In 2022, the Advisory Committee on Immunization Practices (ACIP) recommended universal vaccination for *all unvaccinated persons aged 19–59 years*. Among persons 60 years and older, vaccination is recommended for those at high risk as well as any interested person. Groups considered at high risk include (1) those with sexual exposure (sex workers, those with HBsAg partners, patients at sexually transmitted disease clinics), (2) those with frequent blood exposure (health care workers, first responders, persons who use injection drugs, hemophiliacs, dialysis patients, persons with end-stage renal disease), (3) those who frequently travel to, or first-generation immigrants from, moderate- to high-endemicity areas (>2% prevalence of HBV infection), (4) those with chronic liver disease (HCV infection, steatotic liver disease, autoimmune hepatitis, alcohol-associated liver disease, or aminotransferases >2 times the upper limit of normal), and (5) several other groups (e.g., incarcerated persons and those with HIV infection and diabetes mellitus). Pregnancy is *not* a contraindication to vaccination details of the use of Heplisav-B, a two-injection course a month apart, appear in Table 350-8. In areas of low HBV endemicity such as the United States, despite the availability of safe and effective hepatitis B vaccines, a strategy of vaccinating persons in high-risk groups was not effective. The incidence of new hepatitis B cases continued to increase in the United States after the introduction of vaccines; <10% of all targeted persons in high-risk groups were actually vaccinated, and ~30% of persons with sporadic acute hepatitis B did not fall into any high-risk group category. Therefore, to have an impact on the frequency of HBV infection in an area of low endemicity such as the United States, universal hepatitis B vaccination is now recommended. In HBV-hyperendemic areas (e.g., Asia), universal vaccination of children has resulted in a marked (~70–90%) 40-year decline in complications of hepatitis B, including liver-related mortality and hepatocellular carcinoma.

The original two available aluminum-adjuvanted recombinant hepatitis B vaccines are comparable, one containing 10 μg of HBsAg (Recombivax-HB) and the other containing 20 μg of HBsAg (Engerix-B), and recommended doses for each injection vary for the two preparations (Table 350-8). Combinations of hepatitis B vaccine with other childhood vaccines are available as well (Table 350-8).

More recently, two new recombinant hepatitis B vaccines have been approved for use. One is a recombinant vaccine with a novel adjuvant that activates toll-like 9 receptors (Heplisav-B) and is approved for adults aged 18 or older (although it should be avoided in pregnancy). In a series of prospective trials, compared to three Engerix-B injections, two IM doses a month apart yielded higher proportions with protective levels of anti-HBs (≥10 mIU/mL): 95% of adults aged 18–55 or 18–70 (vs 81% for Engerix-B), 90% of older adults aged 40–70 (vs 71% for Engerix-B), and 90% of adults aged 18–70 with type 2 diabetes (vs 65% for Engerix-B). This two-injection regimen may be useful for revaccination of persons who failed to respond to the original vaccines. The second novel recombinant vaccine (PreHevbrio, VBI Vaccines), which contains all three hepatitis B surface antigen proteins, S, pre-S1, and pre-S2, has been shown in clinical trials (three IM doses at 0, 1, and 6 months) to achieve higher proportions with protective anti-HBs and higher antibody levels than Engerix-B (which contains S antigen only). In addition, PreHevbrio was shown in clinical trials to be associated with higher rates of seroprotection, e.g., in persons ≥45 years (89% compared to 73.1% with an old standard vaccine). This vaccine was approved by the FDA on December 1, 2021, for adults age ≥18 years and recommended for use by the ACIP of the CDC in March 2022. It is the only vaccine not contraindicated in the presence of a yeast allergy.*

For unvaccinated persons sustaining an exposure to HBV, *postexposure* prophylaxis with a combination of HBIG (for rapid achievement of high-titer circulating anti-HBs) and hepatitis B vaccine (for achievement of long-lasting immunity as well as its apparent efficacy in attenuating clinical illness after exposure) is recommended. For *perinatal* exposure of infants born to HBsAg-positive mothers, a single dose of HBIG, 0.5 mL, should be administered IM in the thigh *immediately after birth*, followed by a complete course of three injections of recombinant hepatitis B vaccines approved for children (see doses above) to be started within the first 12 h of life. For those experiencing a direct percutaneous inoculation or transmucosal exposure to HBsAg-positive blood or body fluids (e.g., accidental needle stick, other mucosal penetration, or ingestion), a single IM dose of HBIG, 0.06 mL/kg, administered as soon after exposure as possible, is followed by a complete course of hepatitis B vaccine to begin within the first week. For pregnant mothers with high-level HBV DNA ($>2 \times 10^5$ IU/mL), adding antiviral nucleoside analogues (e.g., pregnancy class B tenofovir, **see Chap. 352**) during the third trimester of pregnancy reduces perinatal transmission even further. For persons exposed by *sexual* contact to a patient with acute hepatitis B, a single IM dose of HBIG, 0.06 mL/kg, should be given within 14 days of exposure, to be followed by a complete course of hepatitis B vaccine. When both HBIG and hepatitis B vaccine are recommended, they may be given at the same time but at separate sites. Testing adults for anti-HBs after a course of vaccine is advisable to document the acquisition of immunity (and can guide management in nonresponders, e.g., whether HBIG and/or repeat vaccination with a more immunogenic vaccine preparation is warranted), but because hepatitis B vaccine immunogenicity is nearly universal in infants, postvaccination anti-HBs testing of children is not recommended. Similarly, testing for the presence of anti-HBs in exposed adults can help guide management (e.g., HBIG, repeat vaccination).

---

*In November 2024, distribution of this vaccine in the United States was discontinued by the manufacturer, because of bankruptcy and termination of operations.

The precise duration of protection afforded by hepatitis B vaccine is unknown; however, ~80–90% of immunocompetent adult vaccinees retain protective levels of anti-HBs for at least 5 years, and 60–80% for 10 years, and protective antibody has been documented to last for at least two decades after vaccination in infancy. Thereafter and even after anti-HBs becomes undetectable, long-term protection (via an anamnestic immune response) persists against clinical hepatitis B, hepatitis B surface antigenemia, and chronic HBV infection. Currently, *booster* immunizations are not recommended routinely, except in immunosuppressed persons who have lost detectable anti-HBs or immunocompetent persons who sustain percutaneous HBsAg-positive inoculations after losing detectable antibody. Specifically, for hemodialysis patients, annual anti-HBs testing is recommended after vaccination; booster doses are recommended when anti-HBs levels fall to <10 mIU/mL. As noted above, for persons at risk of both hepatitis A and B, a combined vaccine is available containing 720 enzyme-linked immunoassay units (ELUs) of inactivated HAV and 20 μg of recombinant HBsAg (at 0, 1, and 6 months).

Hepatitis D Infection with hepatitis D can be prevented by vaccinating susceptible persons with hepatitis B vaccine. No product is available for immunoprophylaxis to prevent HDV superinfection in persons with chronic HBV infection; for these patients, avoidance of percutaneous exposures and limitation of intimate contact with persons who have HDV infection are recommended.

Hepatitis E For prevention of hepatitis E, IG derived from HEV-endemic populations does not appear to be effective. Two safe and effective three-dose (0, 1, and 6 months), recombinant genotype 1 capsid protein vaccines have been shown in randomized, placebo-controlled trials to be highly protective (including against other genotypes) against symptomatic acute hepatitis E. A Chinese vaccine, Hecolin, achieved 100% 12-month efficacy and was licensed in China in 2011; its long-lasting protection (87% efficacy) was documented for up to 4.5 years. A second vaccine developed by GlaxoSmithKline and the U.S. Army achieved a 12-month 96% efficacy. The second vaccine was never developed commercially. The Chinese vaccine is available in China but is not FDA approved or available in the United States.

Hepatitis C IG is ineffective in preventing hepatitis C and is no longer recommended for postexposure prophylaxis in cases of perinatal, needle stick, or sexual exposure. Although prototype vaccines that induce antibodies to HCV envelope proteins have been developed, currently, hepatitis C vaccination is not feasible practically. Genotype and quasispecies viral heterogeneity, as well as rapid evasion of neutralizing antibodies by this rapidly mutating virus, conspire to render HCV a difficult target for immunoprophylaxis with a vaccine. Prevention of transfusion-associated hepatitis C has been accomplished successfully by implementation of highly sensitive virologic screening tests, and the current estimated risk is less than 1 in 10 million.

In the absence of active or passive immunization, prevention of hepatitis C includes behavior changes and precautions to limit exposures to infected persons. Recommendations designed to identify patients with clinically inapparent hepatitis as candidates for medical management have as a secondary benefit the identification of persons whose contacts could be at risk of becoming infected. A so-called look-back program has been recommended to identify persons who were transfused before 1992 with blood from a donor found subsequently to have hepatitis C. In addition, anti-HCV testing, once recommended for persons born between 1945 and 1965, has now been expanded to include all persons 18 years or older, independent of risk factors. For stable, monogamous sexual partners, sexual transmission of hepatitis C is unlikely, and sexual barrier precautions are not recommended. For persons with multiple sexual partners or with sexually transmitted diseases, the risk of sexual transmission of hepatitis C is increased, and barrier precautions (latex condoms) are recommended. A person with hepatitis C should avoid sharing such items as razors, toothbrushes, and nail clippers with sexual partners and family members. No special precautions are recommended for babies born to mothers with hepatitis C, and breast-feeding does not have to be restricted.

A more recent, ambitious goal toward HCV prevention steers focus toward its global eradication by 2030, which would require higher rates of and investment in global surveillance, immediate reflex treatment in persons who test positive for HCV infection, and undelayed, streamlined initiation of treatment (including, and perhaps most importantly, in people who inject drugs to prevent secondary spread) (Table 350-4).

FURTHER READING

Asselah T, Rizzetto M: Hepatitis D virus infection. N Engl J Med 389:58, 2023.

Conners EE, et al: Screening and testing for hepatitis B virus infection: CDC recommendations—United States, 2023. MMWR Morb Mortal Wkly Rep 72:1, 2023.

Doshani M et al: Recommendations of the Advisory Committee on Immunization Practices for use of hepatitis A vaccine for persons experiencing homelessness. MMWR Morb Mortal Wkly Rep 68:153, 2019.

European Association for the Study of the Liver: EASL clinical practice guidelines on hepatitis E virus infection. J Hepatol 68:1256, 2018.

Freedman M et al: Advisory Committee on Immunization Practices. Recommended adult immunization schedule, United States, 2020. Ann Intern Med 172:337, 2020.

Goldberg D et al: Changes in the prevalence of hepatitis C virus infection, nonalcoholic steatohepatitis, and alcoholic liver disease among patients with cirrhosis and liver failure on the waitlist for liver transplantation. Gastroenterology 152:1090, 2017.

Hofmeister MG et al: Estimating prevalence of hepatitis C virus infection in the United States, 2013-2016. Hepatology 69:1020, 2019.

Jeng W-J et al: Hepatitis B. Lancet 401:1039, 2023.

Kelgeri C et al: Clinical spectrum of children with acute hepatitis of unknown cause. N Engl J Med 387:611, 2022.

Koh C et al: Pathogenesis of and new therapies for hepatitis D. Gastroenterology 156:461, 2019.

Le MH et al: Chronic hepatitis B prevalence among foreign-born and U.S.-born adults in the United States, 1999-2016. Hepatology 71:431, 2020.

Lemon SM et al: Type A viral hepatitis: A summary and update on the molecular virology, epidemiology, pathogenesis, and prevention. J Hepatol 68:167, 2018.

Lewis KC et al: Estimated prevalence and awareness of hepatitis C virus infection among US adults: National Health and Nutrition Examination Survey, January 2017–March 2020. Clin Infect Dis 77:1413, 2023.

Nelson NP et al: Prevention of hepatitis A virus infection in the United States; recommendations of the Advisory Committee on Immunization Practices, 2020. MMWR Morb Mortal Wkly Rep 69:1, 2020.

Roberts H et al: Prevalence of chronic hepatitis B virus (HBV) infection in U.S. households: National Health and Nutrition Examination Survey (NHANES), 1988–2012. Hepatology 63:388, 2016.

Schillie S et al: CDC recommendations for hepatitis C screening among adults—United States, 2020. MMWR Recommend Rep 69:1, 2020.

Schweitzer A et al: Estimations of worldwide prevalence of chronic hepatitis B virus infection: A systematic review of data published between 1965 and 2013. Lancet 386:1546, 2015.

Stockdale AJ et al: The global prevalence of hepatitis D virus infection: Systematic review and meta-analysis. J Hepatol 73:523, 2020.

Sureau C et al: The hepatitis delta virus: replication and pathogenesis. J Hepatol 64:S102, 2016.

Van Damme P et al: Hepatitis A virus infection. Nat Rev Dis Primers 9:51, 2023.

Waked I et al: Screening and treatment program to eliminate hepatitis C in Egypt. N Engl J Med 382:1166, 2020.

Weng MK et al: Universal hepatitis B vaccination in adults aged 19–59 years: Updated recommendations of the Advisory Committee on Immunization Practices—United States, 2022. MMWR Morb Mortal Wkly Rep 71:477, 2022.

351

Toxic and Drug-Induced Hepatitis

William M. Lee, Jules L. Dienstag

Liver injury is a possible consequence of ingestion of any xenobiotic, including industrial toxins, pharmacologic agents, and complementary and alternative medications (CAMs). Among patients with acute liver failure, drug-induced liver injury (DILI) is the most common cause, and evidence for hepatotoxicity detected during clinical trials for drug development is the most common reason for failure of compounds to reach approval status. DILI requires careful history-taking to identify unrecognized exposure to chemicals used in work or at home, drugs taken by prescription or bought over the counter, and herbal or dietary supplement medicines. Hepatotoxic drugs can injure the hepatocyte directly, for example, via a free-radical or metabolic intermediate that causes peroxidation of membrane lipids and that results in liver cell injury. Alternatively, a drug or its metabolite may activate components of the innate or adaptive immune system, stimulate apoptotic pathways, or initiate damage to bile excretory pathways **(Fig. 351-1)**. Interference with bile canalicular pumps can allow endogenous bile acids, which can injure the liver, to accumulate. Such secondary injury, in turn, may lead to necrosis of hepatocytes; injure bile ducts, producing cholestasis; or block pathways of lipid movement, inhibit protein synthesis, or impair mitochondrial oxidation of fatty acids, resulting in lactic acidosis and intracellular triglyceride accumulation (expressed histologically as microvesicular steatosis). In other instances, drug metabolites sensitize hepatocytes to toxic cytokines. The differences observed between susceptible and nonsusceptible drug recipients may be attributable to human leukocyte antigen (HLA) haplotypes that determine binding of drug-related haptens on the cell surface as well as to polymorphisms in elaboration of competing, protective cytokines, as has been suggested for acetaminophen hepatotoxicity (see below). Immune mechanisms may include cytotoxic lymphocytes or antibody-mediated cellular cytotoxicity. In addition, a role has been shown for activation of nuclear transporters, such as the constitutive androstane receptor (CAR) or, more recently, the pregnane X receptor (PXR), in the induction of drug hepatotoxicity.

DRUG METABOLISM

Most drugs, which are water-insoluble, undergo a series of metabolic steps, culminating in a water-soluble form appropriate for renal or biliary excretion. This process begins with oxidation or methylation mediated initially by the microsomal mixed function oxygenases, cytochrome P450 (phase I reaction), followed by glucuronidation or sulfation (phase II reaction) or inactivation by glutathione. Most drug hepatotoxicity is the result of formation of a phase I toxic metabolite, but glutathione depletion, precluding inactivation of harmful compounds by glutathione *S*-transferase, can contribute as well by ensuring that the toxic compound is not abrogated.

LIVER INJURY CAUSED BY DRUGS

In general, two major types of chemical hepatotoxicity have been recognized: (1) direct toxic and (2) idiosyncratic. As shown in **Table 351-1**, direct toxic hepatitis occurs with predictable regularity in individuals exposed to the offending agent and is dose-dependent. The latent period between exposure and liver injury is usually short (often several hours), although clinical manifestations may be delayed for 24–48 h. Agents producing toxic hepatitis are generally systemic poisons or are converted in the liver to toxic metabolites. The direct hepatotoxins result in morphologic abnormalities that are reasonably characteristic and reproducible for each toxin. Examples of rare toxins currently include carbon tetrachloride and trichloroethylene that characteristically produce a centrilobular zonal necrosis. The hepatotoxic octapeptides of *Amanita phalloides* usually produce massive hepatic necrosis; the lethal dose of the toxin is ~10 mg, the amount found in a single deathcap mushroom. Acetaminophen, the prime example of a direct toxin, is discussed below.

In idiosyncratic drug reactions, the occurrence of liver injury is infrequent (1 in 10^3–10^5 patients) and unpredictable; the response is not as clearly dose-dependent as is injury associated with direct hepatotoxins, and liver injury may occur at any time after exposure to the drug but typically between 5 and 90 days following its initiation. Although regarded as not dose-related in the fashion of direct toxins, most agents causing idiosyncratic toxicity are given at relatively high daily doses, typically exceeding 100 mg, suggesting a role for dose—drugs with low potency must be given in higher doses that engender greater chances for "off-target" effects. Likewise, drugs given in milligram amounts are of high potency and rarely cause liver or other off-target effects. Adding to the difficulty of predicting or identifying idiosyncratic drug hepatotoxicity is the occurrence of mild, transient, nonprogressive serum aminotransferase elevations that resolve with continued drug use. Such "adaptation," the mechanism of which is unknown, is well recognized for drugs such as isoniazid (INH), valproate, phenytoin, and HMG-CoA reductase inhibitors (statins). Extrahepatic manifestations of hypersensitivity, such as rash, arthralgias, fever, leukocytosis, and eosinophilia, occur in a small fraction of patients with idiosyncratic hepatotoxic drug reactions but are characteristic for certain drugs (phenytoin, trimethoprim-sulfamethoxazole) and not others. Both primary immunologic injury and direct hepatotoxicity related to idiosyncratic differences in generation of toxic metabolites have been invoked to explain idiosyncratic drug reactions. The most current data implicate the adaptive immune system responding to the formation of immune stimulatory compounds resulting from phase I metabolic activation of the offending drug. Differences in host susceptibility may result from varying kinetics of toxic metabolite generation and genetic polymorphisms in downstream drug-metabolizing

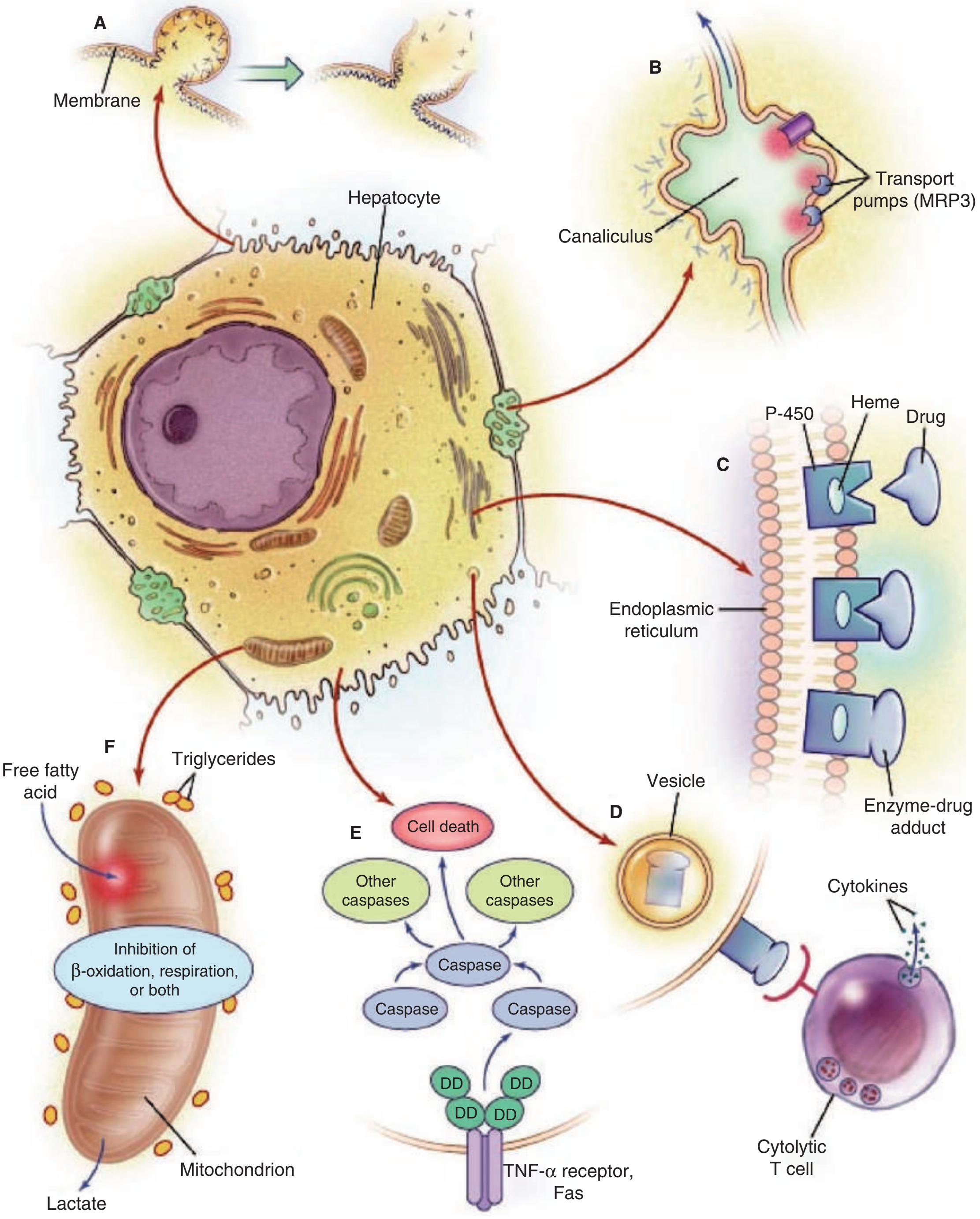
A
Membrane
B
Transport pumps (MRP3)
Canaliculus
Hepatocyte
C
P-450
Heme
Drug
Endoplasmic reticulum
Enzyme-drug adduct
Vesicle
D
Cytokines
Cytolytic T cell
F
Free fatty acid
Triglycerides
Inhibition of β-oxidation, respiration, or both
Mitochondrion
Lactate
E
Cell death
Other caspases
Other caspases
Caspase
Caspase
Caspase
DD
DD
DD
DD
TNF-α receptor, Fas

A. Rupture of cell membrane.
B. Injury of bile canaliculus (disruption of transport pumps).
C. P-450-drug covalent binding (drug adducts).
D. Drug adducts targeted by CTLs/cytokines.
E. Activation of apoptotic pathway by TNFα/Fas.
F. Inhibition of mitochondrial function.

FIGURE 351-1 Potential mechanisms of drug-induced liver injury. The normal hepatocyte may be affected adversely by drugs through (***A***) disruption of intracellular calcium homeostasis that leads to the disassembly of actin fibrils at the surface of the hepatocyte, resulting in blebbing of the cell membrane, rupture, and cell lysis; (***B***) disruption of actin filaments next to the canaliculus (the specialized portion of the cell responsible for bile excretion), leading to loss of villous processes and interruption of transport pumps such as multidrug resistance–associated protein 3 (MRP3), which, in turn, prevents the excretion of bilirubin and other organic compounds; (***C***) covalent binding of the heme-containing cytochrome P450 enzyme to the drug, thus creating nonfunctioning adducts; (***D***) migration of these enzyme-drug adducts to the cell surface in vesicles to serve as target immunogens for cytolytic attack by T cells, stimulating an immune response involving cytolytic T cells and cytokines; (***E***) activation of apoptotic pathways by tumor necrosis factor α (TNF-α) receptor or Fas (DD denotes death domain), triggering the cascade of intercellular caspases, resulting in programmed cell death; or (***F***) inhibition of mitochondrial function by a dual effect on both β-oxidation and the respiratory-chain enzymes, leading to failure of free fatty acid metabolism, a lack of aerobic respiration, and accumulation of lactate and reactive oxygen species (which may disrupt mitochondrial DNA). Toxic metabolites excreted in bile may damage bile-duct epithelium (not shown). CTLs, cytolytic T lymphocytes. *(From WM Lee: Drug-induced hepatotoxicity. N Engl J Med 349:474, 2003. Copyright © 2003, Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

TABLE 351-1 Some Features of Toxic and Drug-Induced Hepatic Injury

	DIRECT TOXIC EFFECT[a]		IDIOSYNCRATIC[a]			OTHER[a]
FEATURES	CARBON TETRACHLORIDE	ACETAMINOPHEN	AMOXICILLIN-CLAVULANATE	ISONIAZID	CIPROFLOXACIN	ESTROGENS/ ANDROGENIC STEROIDS
Predictable and dose-related toxicity	+	+	0	0	0	+
Latent period	Short	Short	Delayed onset	Variable	May be short	Variable
Arthralgia, fever, rash, eosinophilia	0	0	0	0	0	0
Liver morphology	Necrosis, fatty infiltration	Centrilobular necrosis	Mixed hepatocellular/ cholestatic	Hepatocellular injury resembling viral hepatitis	Hepatocellular injury resembling viral hepatitis	Cholestasis *without* portal inflammation

[a]The drugs listed are typical examples.

pathways or cytokine activation; in addition, certain HLA haplotypes have been associated with hepatotoxicity of certain drugs such as amoxicillin-clavulanate and flucloxacillin. Occasionally, however, the clinical features of an allergic reaction (e.g., prominent tissue eosinophilia, autoantibodies) are difficult to ignore and suggest activation of IgE pathways. A few instances of drug hepatotoxicity are observed to be associated with autoantibodies, including a class of antibodies to liver-kidney microsomes, anti-LKM2, directed against a cytochrome P450 enzyme. Four agents that specifically have a phenotype of autoimmune hepatitis with a high likelihood of positive antinuclear antibodies (ANAs) include nitrofurantoin, minocycline, hydralazine, and α-methyldopa.

Idiosyncratic reactions lead to a morphologic pattern that is more variable than those produced by direct toxins; a single agent is often capable of causing a variety of lesions, although certain patterns tend to predominate. Depending on the agent involved, idiosyncratic hepatitis may result in a clinical and morphologic picture indistinguishable from that of viral hepatitis (e.g., INH or ciprofloxacin). So-called hepatocellular injury is the most common form, featuring spotty necrosis in the liver lobule with a predominantly lymphocytic infiltrate resembling that observed in acute hepatitis A, B, or C. Drug-induced cholestasis ranges from mild to increasingly severe: (1) bland cholestasis with limited hepatocellular injury (e.g., estrogens, 17,α-substituted androgens); (2) inflammatory cholestasis (e.g., amoxicillin-clavulanic acid [the most frequently implicated antibiotic among cases of DILI], oxacillin, erythromycin estolate); (3) sclerosing cholangitis (e.g., after intrahepatic infusion of the chemotherapeutic agent floxuridine for hepatic metastases from a primary colonic carcinoma); and (4) disappearance of bile ducts, "ductopenic" cholestasis or vanishing bile duct syndrome, similar to that observed in chronic rejection **(Chap. 357)** following liver transplantation (e.g., carbamazepine, levofloxacin). More recently, a picture resembling primary sclerosing cholangitis has been observed secondary to chemotherapy and after long-term ketamine use. Cholestasis may result from binding of drugs to canalicular membrane transporters, accumulation of toxic bile acids resulting from canalicular pump failure, or genetic defects in canalicular transporter proteins. Clinically, the distinction between a hepatocellular and a cholestatic reaction is indicated by the R value, the ratio of alanine aminotransferase (ALT) to alkaline phosphatase values, both expressed as multiples of the upper limit of normal. An R value of >5.0 is associated with hepatocellular injury, R <2.0 with cholestatic injury, and R between 2.0 and 5.0 with mixed hepatocellular-cholestatic injury.

Morphologic alterations may also include hepatic granulomas (e.g., sulfonamides) or macrovesicular or microvesicular steatosis or steatohepatitis. Severe hepatotoxicity associated with steatohepatitis, most likely a result of mitochondrial toxicity, was recognized with certain antiretroviral therapies, although most of these drugs have been withdrawn **(Chap. 208)**. Another potential target for idiosyncratic drug hepatotoxicity is sinusoidal lining cells; when these are injured, such as by high-dose chemotherapeutic agents (e.g., cyclophosphamide, melphalan, busulfan) administered prior to bone marrow transplantation, venoocclusive disease can result. Nodular regenerative hyperplasia, a subtle form of portal hypertension, may also result from vascular injury to portal or hepatic venous endothelium following systemic chemotherapy, such as with oxaliplatin, as part of adjuvant treatment for colon cancer.

Not all adverse hepatic drug reactions can be classified as either toxic or idiosyncratic. For example, oral contraceptives, which combine estrogenic and progestational compounds, may result in impairment of liver tests and, occasionally, jaundice; however, they do not produce necrosis or fatty change, manifestations of hypersensitivity are generally absent, and susceptibility to the development of oral contraceptive–induced cholestasis appears to be genetically determined. Such estrogen-induced cholestasis is more common in women with cholestasis of pregnancy, a disorder linked to genetic defects in multidrug resistance–associated canalicular transporter proteins.

Any idiosyncratic reaction that occurs in <1:10,000 recipients will go unrecognized in most clinical trials, which involve at most several thousand subjects. The U.S. Food and Drug Administration (FDA) and pharmaceutical companies have learned to look for even subtle indications of serious toxicity and monitor regularly the number of trial subjects in whom any aminotransferase elevations develop, as a possible surrogate for more serious toxicity. Even more valid as a predictor of severe hepatotoxicity is the occurrence of jaundice in patients enrolled in a clinical drug trial, so-called "Hy's Law," named after Dr. Hyman Zimmerman, one of the pioneers of the field of drug hepatotoxicity. He recognized that, if jaundice occurred during a phase 3 trial, more serious liver injury was likely, with a 10:1 ratio between cases of jaundice and liver failure (i.e., 10 patients with jaundice would result in 1 patient with acute liver failure). Thus, the finding of such Hy's Law (jaundiced) cases during drug development often portends failure of approval, particularly if any of the subjects sustains a bad outcome. Troglitazone, a peroxisome proliferator–activated receptor γ agonist, was the first in its class of thiazolidinedione insulin-sensitizing agents. Although in retrospect, Hy's Law cases of jaundice had occurred during phase 3 trials, no instances of liver failure were recognized until well after the drug was introduced, emphasizing the importance of postmarketing surveillance in identifying toxic drugs and in leading to their withdrawal from use. Fortunately, such hepatotoxicity is not characteristic of the second-generation thiazolidinediones rosiglitazone and pioglitazone; in clinical trials, the frequency of aminotransferase elevations in patients treated with these medications did not differ from that in placebo recipients, and isolated reports of liver injury among recipients are extremely rare. Since troglitazone was withdrawn from the market in 2001, no fully approved drugs have had to be withdrawn from the market by the FDA. Several agents have received black box warnings indicating that caution is needed; overall, the industry and FDA in concert have been able to avert severe toxicity in approved agents over the past 25 years.

Proving that an episode of liver injury is caused by a drug (causality) is difficult in many cases. DILI is nearly always a presumptive diagnosis, and many other disorders produce a similar clinicopathologic picture. Thus, causality may be difficult to establish and requires several separate supportive assessment variables to lead to a high level of certainty, including temporal association (time of onset, time to resolution), clinical-biochemical features, type of injury (hepatocellular vs cholestatic), extrahepatic features, likelihood that a given agent is to blame based on its past record, and exclusion of other potential causes. Scoring systems such as the Roussel-Uclaf Causality Assessment Method (RUCAM) yield residual uncertainty and have not been adopted widely. Currently, the U.S. Drug-Induced Liver Injury Network (DILIN) relies on a structured expert opinion process requiring detailed data on each case and a comprehensive review by three experts who arrive at a consensus on a five-degree scale of likelihood (definite, highly likely, probable, possible, unlikely); however, this approach is not practical for routine clinical application. A new Revised Electronic Causality Assessment Method (RECAM) has now been developed to improve on the other two methods but, to date, is not yet in widespread use. The RECAM website can be accessed via the link *dilirecam.com*.

Generally, drug hepatotoxicity is not more frequent in persons with underlying chronic liver disease, although the severity of the outcome may be amplified. Reported exceptions include hepatotoxicity of aspirin, methotrexate, INH (only in certain experiences), antiretroviral therapy for HIV infection, and certain drugs such as conditioning regimens for bone marrow transplantation in the presence of hepatitis C.

TREATMENT

Toxic and Drug-Induced Hepatic Disease

Treatment is largely supportive, except in acetaminophen hepatotoxicity (for which *N*-acetylcysteine is effective, see below). Acute liver failure develops in 10% of patients with DILI; spontaneous recovery, once that threshold is reached, occurs in <30%, and liver transplantation is performed in >40% of those who reach the level of severity of acute liver failure (coagulopathy and hepatic encephalopathy) **(Chap. 356)**. Withdrawal of the suspected agent is indicated at the first sign of an adverse reaction or when aminotransferase levels reach five times the upper limit of normal. A number of studies have suggested that lethal outcomes follow continued use of an agent in the face of symptoms and signs of liver injury. In the case of the direct toxins, liver involvement should not divert attention from renal or other organ involvement, which may also threaten survival. Agents used occasionally but of questionable value include glucocorticoids for DILI with allergic features, silibinin for mushroom poisoning, and ursodeoxycholic acid for cholestatic drug hepatotoxicity. While these medications have been shown to be effective and to have reasonable safety profiles, they are of uncertain value. A double-blind, randomized controlled trial of the use of *N*-acetylcysteine for nonacetaminophen acute liver failure, including cases of DILI, demonstrated benefit, particularly for patients with early-stage hepatic encephalopathy; however, the drug has not been approved by the FDA for this indication, pending "further studies," which, however, are unlikely to be undertaken.

In **Table 351-2**, several classes of chemical agents are listed together with examples of the pattern of liver injury they produce. Certain drugs appear to be responsible for the development of chronic as well as acute hepatic injury. For example, nitrofurantoin, minocycline, hydralazine, and methyldopa have been associated with moderate to severe chronic hepatitis with autoimmune features; corticosteroids may be used and can virtually always be discontinued after 6 months of therapy. Methotrexate, tamoxifen, and amiodarone have been implicated in the development of cirrhosis. Portal hypertension in the absence of cirrhosis, termed *nodular regenerative hyperplasia*, may result from alterations in hepatic architecture produced by excessive intake of vitamin A or following chemotherapy with oxaliplatin. Oral contraceptives have been implicated in the development of focal nodular hyperplasia or hepatic adenoma (both benign lesions) and, rarely, hepatocellular carcinoma and hepatic vein occlusion (Budd-Chiari syndrome). Another unusual lesion, peliosis hepatis (blood cysts of the liver), has been observed in some patients treated with anabolic or contraceptive steroids. The existence of these hepatic disorders expands the spectrum of liver injury induced by chemical agents and emphasizes the need for a thorough drug history in all patients with liver dysfunction. The comprehensive, authoritative LiverTox website, which contains up-to-date information on DILI, is available as a valuable reference through the National Institutes of Health and the National Library of Medicine (*livertox.nih.gov*).

The following are patterns of adverse hepatic reactions for some prototypic agents.

ACETAMINOPHEN HEPATOTOXICITY (DIRECT TOXIN)

Acetaminophen represents the most prevalent cause of acute liver failure in the Western world; up to 72% of patients with acetaminophen hepatotoxicity in Scandinavia—somewhat lower frequencies in the United Kingdom and the United States—progress to encephalopathy and coagulopathy. Acetaminophen causes dose-related centrilobular hepatic necrosis after single-time-point ingestions, as intentional self-harm, or over extended periods, as unintentional overdoses, when multiple drug preparations or inappropriate drug amounts are used daily for several days, for example, for relief of pain or fever. In these instances, 8 g/d, twice the daily recommended maximum dose, over several days can readily lead to liver failure. Use of opioid-acetaminophen combinations appears to be particularly harmful, because habituation to the opioid may occur with a gradual increase in opioid-acetaminophen combination dosing over days or weeks. A single dose of 10–15 g, occasionally less, may produce clinical evidence of liver injury. Fatal fulminant disease is usually (although not invariably) associated with ingestion of ≥25 g. Blood levels of acetaminophen correlate with severity of hepatic injury (levels >300 μg/mL 4 h after ingestion are predictive of the development of severe damage; levels <150 μg/mL suggest that hepatic injury is highly unlikely). Nausea, vomiting, diarrhea, abdominal pain, and shock are early manifestations occurring 4–12 h after ingestion. Then 24–48 h later, when these features are abating, hepatic injury becomes apparent. Maximal abnormalities and hepatic failure are evident 3–5 days after ingestion, and aminotransferase levels exceeding 10,000 IU/L are not uncommon (i.e., levels far exceeding those in patients with viral hepatitis). Renal failure and myocardial injury may be present. Whether or not a clear history of overdose can be elicited, clinical suspicion of acetaminophen hepatotoxicity should be raised by the presence of the extremely high aminotransferase levels in association with low bilirubin levels that are characteristic of this hyperacute injury. This biochemical signature should trigger further questioning of the subject if possible; however, outright denial (or denial of high doses) or altered mentation may confound diagnostic efforts. In this setting, a presumptive diagnosis is reasonable, and the proven antidote, *N*-acetylcysteine, is both safe and will be effective if given early (within 12 h) but is also used even when injury has evolved.

Acetaminophen is metabolized predominantly by a phase II reaction to innocuous sulfate and glucuronide metabolites; however, a small proportion is metabolized by a phase I reaction to a hepatotoxic metabolite formed from the parent compound by cytochrome P450 CYP2E1. This metabolite, *N*-acetyl-p-benzoquinone-imine (NAPQI), is detoxified by binding to "hepatoprotective" glutathione to become harmless, water-soluble mercapturic acid, which undergoes renal excretion. When excessive amounts of NAPQI are formed, or when glutathione levels are low, glutathione levels are depleted and overwhelmed, permitting covalent binding to nucleophilic hepatocyte macromolecules forming acetaminophen-protein "adducts." These adducts, which can be measured in serum by high-performance liquid chromatography, hold promise as diagnostic markers of acetaminophen hepatotoxicity, and a point-of-care assay for acetaminophen-Cys adducts is under development. The binding of acetaminophen to hepatocyte macromolecules is believed to lead to hepatocyte necrosis; the precise sequence and mechanism are unknown. Hepatic injury may be potentiated by prior administration of alcohol, phenobarbital, INH, or other drugs; by conditions that stimulate the mixed-function oxidase system; or by conditions such as starvation (including inability to maintain oral intake during severe febrile illnesses) that reduce hepatic glutathione levels. Alcohol induces cytochrome P450 CYP2E1; consequently, increased levels of the toxic metabolite NAPQI may be produced in chronic alcoholics after acetaminophen ingestion, but the role of alcohol in potentiating acute acetaminophen injury is still debated. Alcohol

TABLE 351-2 Principal Alterations of Hepatic Morphology Produced by Some Commonly Used Drugs and Chemicals[a]

PRINCIPAL MORPHOLOGIC CHANGE	CLASS OF AGENT	EXAMPLE
Cholestasis	Anabolic steroid	Methyl testosterone, many other body-building supplements
	Antibiotic	Erythromycin estolate, nitrofurantoin, rifampin, amoxicillin-clavulanic acid, oxacillin
	Anticonvulsant	Carbamazepine
	Antidepressant	Duloxetine, mirtazapine, tricyclic antidepressants
	Anti-inflammatory	Sulindac
	Antiplatelet	Clopidogrel
	Antihypertensive	Irbesartan, fosinopril
	Antithyroid	Methimazole
	Calcium channel blocker	Nifedipine, verapamil
	Immunosuppressive	Cyclosporine
	Lipid-lowering	Ezetimibe
	Oncotherapeutic	Anabolic steroids, busulfan, tamoxifen, irinotecan, cytarabine, temozolomide
	Oral contraceptive	Norethynodrel with mestranol
	Oral hypoglycemic	Chlorpropamide
	Tranquilizer	Chlorpromazine[b]
Fatty liver	Antiarrhythmic	Amiodarone
	Antibiotic	Tetracycline (high-dose, IV)
	Anticonvulsant	Valproic acid
	Antiviral	Dideoxynucleosides (e.g., zidovudine), protease inhibitors (e.g., indinavir, ritonavir)
	Oncotherapeutic	Asparaginase, methotrexate, tamoxifen
Hepatitis	Anesthetic	Halothane, fluothane
	Antiandrogen	Flutamide
	Antibiotic	Isoniazid,[c] rifampicin, nitrofurantoin, telithromycin, minocycline,[d] pyrazinamide, trovafloxacin[e]
	Anticonvulsant	Phenytoin, carbamazepine, valproic acid, phenobarbital
	Antidepressant	Iproniazid, amitriptyline, trazodone, venlafaxine, fluoxetine, paroxetine, duloxetine, sertraline, nefazodone[e]
	Antifungal	Ketoconazole, fluconazole, itraconazole
	Antihypertensive	Methyldopa,[c] captopril, enalapril, lisinopril, losartan
	Anti-inflammatory	Ibuprofen, indomethacin, diclofenac, sulindac, bromfenac
	Antipsychotic	Risperidone
	Antiviral	Zidovudine, didanosine, stavudine, nevirapine, ritonavir, indinavir, tipranavir, zalcitabine
	Calcium channel blocker	Nifedipine, verapamil, diltiazem
	Cholinesterase inhibitor	Tacrine
	Diuretic	Chlorothiazide
	Laxative	Oxyphenisatin[c,e]
	Norepinephrine reuptake inhibitor	Atomoxetine
	Oral hypoglycemic	Troglitazone,[e] acarbose
	Immune checkpoint inhibitor	Ipilimumab, pembrolizumab, nivolumab
Mixed hepatitis/cholestatic	Antibiotic	Amoxicillin-clavulanic acid, trimethoprim-sulfamethoxazole
	Antibacterial	Clindamycin
	Antifungal	Terbinafine
	Antihistamine	Cyproheptadine
	Immunosuppressive	Azathioprine
	Lipid-lowering	Nicotinic acid, lovastatin, ezetimibe
Toxic (necrosis)	Analgesic	Acetaminophen
	Hydrocarbon	Carbon tetrachloride
	Metal	Yellow phosphorus
	Mushroom	*Amanita phalloides*
	Solvent	Dimethylformamide
Granulomas	Antiarrhythmic	Quinidine, diltiazem
	Antibiotic	Sulfonamides
	Anticonvulsant	Carbamazepine
	Anti-inflammatory	Phenylbutazone
	Xanthine oxidase inhibitor	Allopurinol
Vascular injury	Chemotherapeutic	Oxaliplatin, melphalan

[a] Several agents cause more than one type of liver lesion and appear under more than one category. [b] Rarely associated with primary biliary cholangitis-like lesion. [c] Occasionally associated with chronic hepatitis or bridging hepatic necrosis or cirrhosis. [d] Associated with an autoimmune hepatitis-like syndrome. [e] Withdrawn from use because of severe hepatotoxicity.

also suppresses hepatic glutathione production. Therefore, in chronic alcoholics, the toxic dose of acetaminophen may be as low as 2 g, and alcoholic patients should be warned specifically about the dangers of even standard doses of this commonly used drug. In a 2006 study, aminotransferase elevations were identified in 31–44% of normal subjects treated for 14 days with the maximal recommended dose of acetaminophen, 4 g daily (administered alone or as part of an acetaminophen-opioid combination); because these changes were transient and never associated with bilirubin elevation, the clinical relevance of these findings remains to be determined. Although underlying hepatitis C virus (HCV) infection was found to be associated with an increased risk of acute liver injury in patients hospitalized for acetaminophen overdose, generally, in patients with nonalcoholic liver disease, acetaminophen taken in recommended doses is well tolerated. Acetaminophen use in cirrhotic patients has not been associated with hepatic decompensation. On the other hand, because of the link between acetaminophen use and liver injury and because of the limited safety margin between safe and toxic doses, the FDA has recommended that the daily dose of acetaminophen be reduced from 4 g to 3 g (even lower for persons with chronic alcohol use), that all acetaminophen-containing products be labeled prominently as containing acetaminophen, and that the potential for liver injury be prominent in the packaging of acetaminophen and acetaminophen-containing products. Within opioid combination products, the limit for the acetaminophen component has been lowered to 325 mg per tablet. Adoption of this limit appears to have had a salutary effect in decreasing the number of hospital admissions and instances of liver failure associated with acetaminophen-opioid combinations.

TREATMENT

Acetaminophen Overdosage

Treatment includes gastric lavage, supportive measures, and oral administration of activated charcoal or cholestyramine to prevent absorption of residual drug. Neither charcoal nor cholestyramine appears to be effective if given >30 min after acetaminophen ingestion; if they are used, the stomach lavage should be done before other agents are administered orally. The chances of possible, probable, and high-risk hepatotoxicity can be derived from a nomogram plot, readily available in emergency departments, as a function of measuring acetaminophen plasma levels 4–8 h after ingestion. In patients with high acetaminophen blood levels (>200 μg/mL measured at 4 h or >100 μg/mL at 8 h after ingestion), the administration of *N*-acetylcysteine reduces markedly the severity of hepatic necrosis. This agent provides sulfhydryl donor groups to replete glutathione, which is required to render harmless toxic metabolites that would otherwise bind covalently via sulfhydryl linkages to cell proteins, resulting in the formation of drug metabolite-protein adducts. Therapy should be begun within 8 h of ingestion but may be at least partially effective when given as late as 24–36 h after overdose. Routine use of *N*-acetylcysteine has substantially reduced the occurrence of fatal acetaminophen hepatotoxicity. *N*-acetylcysteine may be given orally but is more commonly used as an IV solution, with a loading dose of 140 mg/kg over 1 h, followed by 70 mg/kg every 4 h for 15–20 doses. Whenever a patient with potential acetaminophen hepatotoxicity is encountered, a local poison control center should be contacted. Treatment can be stopped when plasma acetaminophen levels indicate that the risk of liver damage is low. If signs of hepatic failure (e.g., progressive jaundice, coagulopathy, confusion) occur despite *N*-acetylcysteine therapy for acetaminophen hepatotoxicity, liver transplantation may be the only option. Early arterial blood lactate levels among such patients with acute liver failure may distinguish patients highly likely to require liver transplantation (lactate levels >3.5 mmol/L) from those likely to survive without liver replacement. Acute renal injury occurs in nearly 75% of patients with severe acetaminophen injury but is virtually always self-limited.

Survivors of acute acetaminophen overdose rarely, if ever, have ongoing liver injury or sequela but may be subject to repeat overdosing.

ISONIAZID HEPATOTOXICITY (IDIOSYNCRATIC REACTION)

INH remains central to most antituberculous prophylactic and therapeutic regimens, despite its long-standing recognition as a hepatotoxin. In 10% of patients treated with INH, elevated serum aminotransferase levels develop during the first few weeks of therapy; however, these elevations in most cases are self-limited, are mild (values for ALT <200 IU/L), and resolve despite continued drug use. This adaptive response allows continuation of the agent if symptoms and progressive enzyme elevations do not follow the initial elevations. Acute hepatocellular DILI secondary to INH is evident with a variable latency period up to 6 months and is more frequent in alcoholics and patients taking certain other medications, such as barbiturates, rifampin, and pyrazinamide. If the clinical threshold of encephalopathy is reached, severe hepatic injury is likely to be fatal or to require liver transplantation. Liver biopsy reveals morphologic changes similar to those of viral hepatitis or bridging hepatic necrosis. Substantial liver injury appears to be age-related, increasing substantially after age 35; the highest frequency is in patients over age 50, and the lowest is in patients under the age of 20. Even for patients >50 years of age monitored carefully during therapy, hepatotoxicity occurs in only ~2%, well below the risk estimate derived from earlier experiences. Fever, rash, eosinophilia, and other manifestations of drug allergy are distinctly unusual. Antibodies to INH have been detected in INH recipients, but a link to causality of liver injury remains unclear. A clinical picture resembling chronic hepatitis has been observed in a few patients. Many public health programs that require INH prophylaxis for a positive tuberculin skin test or blood test (Quantiferon or T-Spot) include monthly monitoring of aminotransferase levels, although this practice has been called into question. Even more effective in limiting serious outcomes may be encouraging patients to be alert for symptoms such as nausea, fatigue, or jaundice, because most fatalities occur in the setting of continued INH use despite clinically apparent illness. The incidence of severe INH toxicity may be declining as a result of less frequent use and/or better management.

SODIUM VALPROATE HEPATOTOXICITY (TOXIC AND IDIOSYNCRATIC REACTION)

Sodium valproate, an anticonvulsant useful in the treatment of petit mal and other seizure disorders, has been associated with the development of severe hepatic toxicity and, rarely, fatalities, predominantly in children but also in adults. Among children listed as candidates for liver transplantation, valproate is the most common antiepileptic drug implicated. Asymptomatic elevations of serum aminotransferase levels have been recognized in as many as 45% of treated patients. These "adaptive" changes, however, appear to have no clinical importance, because major hepatotoxicity is not seen in the majority of patients despite continuation of drug therapy. In the rare patients in whom jaundice, encephalopathy, and evidence of hepatic failure are found, examination of liver tissue reveals microvesicular fat and bridging hepatic necrosis, predominantly in the centrilobular zone. Bile duct injury may also be apparent. Most likely, sodium valproate is not directly hepatotoxic, but its metabolite, 4-pentenoic acid, may be responsible for hepatic injury. Valproate hepatotoxicity is more common in persons with mitochondrial enzyme deficiencies and may be ameliorated by IV administration of carnitine, which valproate therapy can deplete. Valproate toxicity has been linked to HLA haplotypes (*DR4* and *B***1502*) and to mutations in mitochondrial DNA polymerase gamma 1.

NITROFURANTOIN HEPATOTOXICITY (IDIOSYNCRATIC REACTION)

This commonly used antibiotic for urinary tract infections may cause an acute hepatitis leading to fatal outcome or, more frequently, chronic hepatitis of varying severity but indistinguishable from autoimmune hepatitis. These two scenarios may reflect the frequent use and reuse of the drug for treatment of recurrent cystitis in women. Although most toxic agents manifest injury within 6 months of first ingestion, nitrofurantoin may have a longer latency period, in part perhaps because of its intermittent, recurrent use. Autoantibodies to nuclear components, smooth muscle, and mitochondria are seen and may subside after resolution of injury; however, glucocorticoid or other immunosuppressive medication may be necessary to resolve the autoimmune injury, and cirrhosis may be seen in cases that are not recognized quickly. Interstitial pulmonary fibrosis presenting as chronic cough and dyspnea may be present and resolve slowly with medication withdrawal. Histologic findings are identical to those of autoimmune hepatitis. A similar disease pattern can be observed with minocycline, which is used repeatedly for the treatment of acne in teenagers, as well as with hydralazine and α-methyldopa.

■ AMOXICILLIN-CLAVULANATE HEPATOTOXICITY (IDIOSYNCRATIC MIXED REACTION)

Currently, the most common agent implicated as causing DILI in the United States and in Europe is amoxicillin-clavulanate (most frequent brand name: Augmentin). This medication causes a very specific syndrome of mixed or primarily cholestatic injury. Because hepatotoxicity may follow amoxicillin-clavulanate therapy after a relatively long latency period, the liver injury may begin to manifest after the drug has been withdrawn. The high prevalence of hepatotoxicity reflects in part the very frequent use of this drug for respiratory tract infections, including community-acquired pneumonia. The mechanism of hepatotoxicity is unclear, but the liver injury is thought to be caused by amoxicillin toxicity that is potentiated in some way by clavulanate, which itself appears not to be toxic. Symptoms include nausea, anorexia, fatigue, and jaundice—which may be prolonged—with pruritus. Rash is quite uncommon. On occasion, amoxicillin-clavulanate, like other cholestatic hepatotoxic drugs, causes permanent injury to small bile ducts, leading to the so-called "vanishing bile duct syndrome." In vanishing bile duct syndrome, initially, liver injury is minimal except for severe cholestasis; however, over time, histologic evidence of bile duct abnormalities is replaced by a paucity and eventual absence of discernible ducts on subsequent biopsies.

■ AMIODARONE HEPATOTOXICITY (TOXIC AND IDIOSYNCRATIC REACTION)

Therapy with this potent antiarrhythmic drug is accompanied in 15–50% of patients by modest elevations of serum aminotransferase levels that may remain stable or diminish despite continuation of the drug. Such abnormalities may appear days to many months after beginning therapy. A proportion of those with elevated aminotransferase levels have detectable hepatomegaly, and clinically important liver disease develops in <5% of patients. Features that represent a direct effect of the drug on the liver and that are common to the majority of long-term recipients are ultrastructural phospholipidosis, unaccompanied by clinical liver disease, and interference with hepatic mixed-function oxidase metabolism of other drugs. The cationic amphiphilic drug and its major metabolite desethylamiodarone accumulate in hepatocyte lysosomes and mitochondria and in bile duct epithelium. The relatively common elevations in aminotransferase levels are also considered a predictable, dose-dependent, direct hepatotoxic effect. On the other hand, in the rare patient with clinically apparent, symptomatic liver disease, liver injury resembling that seen in alcoholic liver disease is observed. The so-called pseudoalcoholic liver injury can range from steatosis, to alcoholic hepatitis–like neutrophilic infiltration and Mallory's hyaline, to cirrhosis. Electron-microscopic demonstration of phospholipid-laden lysosomal lamellar bodies can help to distinguish amiodarone hepatotoxicity from typical alcoholic hepatitis. This category of liver injury appears to be a metabolic idiosyncrasy that allows hepatotoxic metabolites to be generated. Rarely, an acute idiosyncratic hepatocellular injury resembling viral hepatitis or cholestatic hepatitis occurs. Hepatic granulomas have occasionally been observed. Because amiodarone has a long half-life, liver injury may persist for months after the drug is stopped.

■ ANABOLIC STEROIDS (CHOLESTATIC REACTION)

The most common form of liver injury caused by CAMs is the profound cholestasis associated with anabolic steroids used by body builders. Unregulated agents sold in gyms and health food stores as diet supplements, which are taken by athletes to improve their performance, may contain anabolic steroids. In a young male, jaundice that is accompanied by a cholestatic, rather than a hepatitic, laboratory profile almost invariably will turn out to be caused by the use of one of a variety of androgen congeners. Such agents have the potential to injure bile transport pumps and to cause intense cholestasis; the time to onset is variable, and resolution, which is the rule, may require many weeks to months. Initially, anorexia, nausea, and malaise may occur, followed by pruritus in some but not all patients. Serum aminotransferase levels are usually <100 IU/L, and serum alkaline phosphatase levels are generally moderately elevated with bilirubin levels frequently exceeding 342 μmol/L (20 mg/dL). Examination of liver tissue reveals cholestasis without substantial inflammation or necrosis. Anabolic steroids have also been used by prescription to treat bone marrow failure. In this setting, hepatic centrizonal sinusoidal dilatation and peliosis hepatis have been reported in rare patients, as have hepatic adenomas and hepatocellular carcinoma. Recently, a large series of cases with a uniform phenotype has been described. Unfortunately, no genomic signature has become evident despite the unique features of the injury. No permanent sequelae are evident besides prolonged jaundice, lasting frequently 10 weeks or more.

■ TRIMETHOPRIM-SULFAMETHOXAZOLE HEPATOTOXICITY (IDIOSYNCRATIC REACTION)

This antibiotic combination is used routinely for urinary tract infections in immunocompetent persons and for prophylaxis against and therapy of *Pneumocystis jirovecii* pneumonia in immunosuppressed persons (transplant recipients, patients with AIDS). With its increasing use, its occasional hepatotoxicity is being recognized with growing frequency. Its likelihood is unpredictable, but when it occurs, trimethoprim-sulfamethoxazole hepatotoxicity follows a relatively uniform latency period of several weeks and is often accompanied by eosinophilia, rash, and other features of a hypersensitivity reaction, including the drug reaction with eosinophilia and systemic symptoms (DRESS) syndrome. Biochemically and histologically, acute hepatocellular necrosis predominates, but cholestatic features are quite frequent. Occasionally, cholestasis without necrosis occurs, and very rarely, a severe cholangiolytic pattern of liver injury is observed. In most cases, liver injury is self-limited, but rare fatalities have been recorded. The hepatotoxicity is attributable to the sulfamethoxazole component of the drug and is similar in features to that seen with other sulfonamides; tissue eosinophilia and granulomas may be seen. The risk of trimethoprim-sulfamethoxazole hepatotoxicity is increased in persons with HIV infection. In a recent study, unique HLA associations in European Americans and in African Americans have been identified.

■ HMG-COA REDUCTASE INHIBITORS (STATINS) (IDIOSYNCRATIC MIXED HEPATOCELLULAR AND CHOLESTATIC REACTION)

Between 1 and 2% of patients taking lovastatin, simvastatin, pravastatin, fluvastatin, or one of the newer statin drugs for the treatment of hypercholesterolemia experience asymptomatic, reversible elevations (greater than threefold) of aminotransferase activity. Acute hepatitis-like histologic changes, centrilobular necrosis, and centrilobular cholestasis have been described in a very small number of cases. In a larger proportion, minor aminotransferase elevations appear during the first several weeks of therapy. Careful laboratory monitoring can distinguish between patients with minor, transitory changes, who may continue therapy, and those with more profound and sustained abnormalities, who should discontinue therapy. Because clinically meaningful aminotransferase elevations are so rare after statin use and do not differ in meta-analyses from the frequency of such laboratory abnormalities in placebo recipients, a panel of liver experts recommended to the National Lipid Association's Safety Task Force that liver test monitoring was not necessary in patients treated with statins and that statin therapy need not be discontinued in patients found to have asymptomatic isolated aminotransferase elevations during therapy. Statin hepatotoxicity is not increased in patients with chronic hepatitis C, hepatic steatosis, or other underlying liver diseases, and statins can be used safely in these patients.

ALTERNATIVE AND COMPLEMENTARY MEDICINES (IDIOSYNCRATIC HEPATITIS, STEATOSIS)

Herbal medications that are of scientifically unproven efficacy and that lack prospective safety oversight by regulatory agencies account currently for >20% of DILI in the United States. Besides anabolic steroids, the most common category of dietary or herbal products is weight loss agents. Included among the herbal remedies associated with toxic hepatitis are Jin Bu Huan, xiao-chai-hu-tang, germander, chaparral, senna, mistletoe, skullcap, gentian, comfrey (containing pyrrolizidine alkaloids), ma huang, bee pollen, valerian root, pennyroyal oil, kava, celandine, Impila (*Callilepis laureola*), LipoKinetix, Hydroxycut, OxyElite Pro, Herbalife, herbal nutritional supplements, and herbal teas containing *Camellia sinensis* (green tea extract). Well characterized are the acute hepatitis-like histologic lesions following Jin Bu Huan use: focal hepatocellular necrosis, mixed mononuclear portal tract infiltration, coagulative necrosis, apoptotic hepatocyte degeneration, tissue eosinophilia, and microvesicular steatosis. Megadoses of vitamin A can injure the liver, as can pyrrolizidine alkaloids, which often contaminate Chinese herbal preparations and can cause a venoocclusive injury leading to sinusoidal hepatic vein obstruction. Because some alternative medicines induce toxicity via active metabolites, alcohol and drugs that stimulate cytochrome P450 enzymes may enhance the toxicity of some of these products. Conversely, some alternative medicines also stimulate cytochrome P450 and may result in or amplify the toxicity of recognized drug hepatotoxins. In many instances, herbal and dietary supplements actually contain chemicals rather than only leaves, roots, and bark. Antirheumatic "herbs" have been found to contain a nonsteroidal anti-inflammatory drug (NSAID) such as diclofenac, for example. Given the widespread use of such poorly defined herbal preparations, hepatotoxicity is likely to be encountered with increasing frequency; therefore, a drug history in patients with acute and chronic liver disease should include use of "alternative medicines" and other nonprescription preparations sold in so-called health food stores.

■ CHECKPOINT INHIBITOR AND OTHER IMMUNOTHERAPIES FOR CANCER

The introduction of a new class of immunotherapeutic agents for melanoma and other cancers has ushered in a new kind of hepatotoxicity, that associated with activation of the immune response. The three classes of immune-active molecules are cytotoxic T-lymphocyte-associated antigen-4 (CTLA-4), programmed cell death receptor 1 (PD-1), and programmed cell death receptor ligand 1 (PD-L1). Within weeks of beginning treatment with any one of several agents, including ipilimumab (CTLA-4), pembrolizumab (PD-1), or nivolumab (PD-1), an active hepatitis evolves that is associated with positive ANAs and appears to respond to glucocorticoid therapy. Liver histology may share some of the features of, but does not resemble the chronic changes observed in, autoimmune hepatitis. Instead, histologic findings are compatible with a nonspecific acute hepatic injury, assumed to result from the release of host modulation of anti-self-immune responses and mediated predominantly by CD8+ lymphocytes. Immune-mediated injury to thyroid, muscle, and colon may be seen as well. Few deaths have been reported related to these immunotherapies; while these novel agents may need to be halted temporarily, in many cases, they can be restarted (and are tolerated better on retreatment) if patients are showing a favorable antitumor response.

■ HIGHLY ACTIVE ANTIRETROVIRAL THERAPY FOR HIV INFECTION (MITOCHONDRIAL TOXIC, IDIOSYNCRATIC, STEATOSIS; HEPATOCELLULAR, CHOLESTATIC, AND MIXED)

The recognition of drug hepatotoxicity in persons with HIV infection is complicated in this population by the many alternative causes of liver injury (chronic viral hepatitis, fatty infiltration, infiltrative disorders, mycobacterial infection, etc.), but drug hepatotoxicity associated with highly active antiretroviral therapy (HAART) was a common type of liver injury in HIV-infected persons in the early days of HIV therapy; however, it is less frequent now **(Chap. 208)**. Implicated most frequently are combinations including the nucleoside analogue reverse transcriptase inhibitors zidovudine, didanosine, and, to a lesser extent, stavudine; the protease inhibitors ritonavir and indinavir (and amprenavir when used together with ritonavir), as well as tipranavir; and the nonnucleoside reverse transcriptase inhibitors nevirapine and, to a lesser extent, efavirenz. Distinguishing the impact of HAART hepatotoxicity in patients with HIV and hepatitis virus co-infection is made challenging by the following: (1) both chronic hepatitis B and hepatitis C can affect the natural history of HIV infection and the response to HAART, and (2) HAART can have an impact on chronic viral hepatitis. For example, immunologic reconstitution with HAART can result in immunologically mediated liver-cell injury in patients with chronic hepatitis B co-infection if treatment with an antiviral agent for hepatitis B (e.g., nucleoside analogues such as tenofovir) is withdrawn. Infection with HIV, especially with low CD4+ T-cell counts, has been reported to increase the rate of hepatic fibrosis associated with chronic hepatitis C, and HAART therapy can increase levels of serum aminotransferases and HCV RNA in patients with hepatitis C co-infection. Didanosine or stavudine should not be used with ribavirin in patients with HIV/HCV co-infection because of an increased risk of severe mitochondrial toxicity and lactic acidosis.

■ FURTHER READING

Ahmad J et al: Sclerosing cholangitis-like changes on magnetic resonance cholangiography in patients with drug-induced liver injury. Clin Gastroenterol Hepatol 17:789, 2019.

Björnsson ES, Hoofnagle JL: Categorization of drugs implicated in causing liver injury: Critical assessment based upon published case reports. Hepatology 63:590, 2016.

Chalasani NP et al: ACG clinical guideline: Diagnosis and management of idiosyncratic drug-induced liver injury. Am J Gastroenterol 116:878, 2021.

Chalasani N et al: Drug Induced Liver Injury Network. Clinical features, outcomes, and HLA risk factors associated with nitrofurantoin-induced liver injury. J Hepatol 78:293, 2023.

Cirulli ET et al: A missense variant in PTPN22 is a risk factor for drug-induced liver injury. Gastroenterology 156:1707, 2019.

de Boer YS et al: Features of autoimmune hepatitis in patients with drug-induced liver injury. Clin Gastroenterol Hepatol 15:103, 2017.

European Association for the Study of the Liver: EASL clinical practice guidelines: Drug-induced liver injury. J Hepatol 70:1222, 2019.

Fontana RJ et al: AASLD practice guidance on drug, herbal, and dietary supplement-induced liver injury. Hepatology 77:1036, 2023.

Fontana RJ et al: The evolving profile of idiosyncratic drug-induced liver injury. Clin Gastroenterol Hepatol 21:2088, 2023.

Hayashi PH et al: RECAM: A revised electronic version of RUCAM for diagnosis of drug induced liver injury. Hepatology 76:18, 2022.

Hoofnagle JH et al: HLA-B*35:01 and green tea induced liver injury. Hepatology 73:2484, 2021.

Orandi BJ et al: Association of FDA mandate limiting acetaminophen (paracetamol) in prescription combination opioid products and subsequent hospitalizations and acute liver failure. JAMA 329:735, 2023.

Peeraphatdit TB et al: Hepatotoxicity from immune checkpoint inhibitors: A systematic review and management recommendation. Hepatology 72:315, 2020.

Stolz A et al: Severe and protracted cholestasis in 44 young men taking bodybuilding supplements: Assessment of genetic, clinical and chemical risk factors. Aliment Pharmacol Ther 49:1195, 2019.

352 Chronic Hepatitis

Esperance A. K. Schaefer,
Raymond T. Chung, Jules L. Dienstag

Chronic hepatitis represents a series of liver disorders of varying causes and severity in which hepatic inflammation and necrosis continue for at least 6 months. Milder forms are nonprogressive or only slowly progressive, while more severe forms may be associated with scarring and architectural reorganization, which, when advanced, lead ultimately to cirrhosis. Several broad categories of chronic hepatitis have been recognized. These include chronic viral hepatitis, drug-or toxin-induced chronic hepatitis **(Chap. 351)**, alcohol-associated liver disease **(Chap. 353)**, metabolic dysfunction-related chronic hepatitis **(Chap 354)**, and autoimmune chronic hepatitis. In many cases, clinical and laboratory features are insufficient to allow assignment into one of these distinct categories; these "idiopathic" cases are also believed to represent autoimmune chronic hepatitis. Moreover, because of the high prevalence of steatotic liver diseases (metabolic dysfunction-associated steatotic liver disease and alcohol-associated liver disease), patients with more than one, overlapping type of hepatitis are encountered not infrequently (e.g., viral and steatotic liver injury, alcohol-related and nonalcohol-related steatotic liver injury). Finally, clinical and laboratory features of chronic hepatitis are observed occasionally in patients with other hereditary/metabolic disorders, such as Wilson disease (copper overload) and α_1 antitrypsin deficiency **(Chaps. 355 and 427)**. Although all types of chronic hepatitis share certain clinical, laboratory, and histopathologic features, chronic viral and chronic autoimmune hepatitis are sufficiently distinct to merit separate discussions. **For discussion of acute hepatitis, see Chap. 350.**

CLASSIFICATION OF CHRONIC HEPATITIS

Common to all forms of chronic hepatitis are histopathologic distinctions based on localization and extent of liver injury. These vary from the milder forms, previously labeled *chronic persistent hepatitis* and *chronic lobular hepatitis*, to the more severe form, formerly called *chronic active hepatitis*. When first defined, these designations were believed to have prognostic implications, which were not corroborated by subsequent observations. Categorization of chronic hepatitis based primarily on histopathologic features has been replaced by a more informative classification based on a combination of clinical, serologic, and histologic variables. Classification of chronic hepatitis is based on (1) its *cause*; (2) its histologic activity, or *grade*; and (3) its degree of progression based on level of fibrosis, or *stage*. Thus, neither clinical features alone nor histologic features—requiring liver biopsy or noninvasive markers of fibrosis—alone are sufficient to characterize and distinguish among the several categories of chronic hepatitis.

CLASSIFICATION BY CAUSE

Clinical and serologic features allow the establishment of a diagnosis of *chronic viral hepatitis*, caused by hepatitis B, hepatitis B plus D, or hepatitis C; *autoimmune hepatitis*, including several subcategories, I and II, based on serologic distinctions; *drug-associated chronic hepatitis*; and a category of unknown cause, or *cryptogenic chronic hepatitis* **(Table 352-1)**. These are addressed in more detail below.

CLASSIFICATION BY GRADE

Grade, a histologic assessment of necroinflammatory activity, is based on examination of the liver biopsy. An assessment of important histologic features includes the degree of *periportal necrosis* and the disruption of the limiting plate of periportal hepatocytes by inflammatory cells (so-called *piecemeal necrosis* or *interface hepatitis*); the degree of confluent necrosis that links or forms bridges between vascular structures—between portal tract and portal tract or even more important bridges between portal tract and central vein—referred to as *bridging necrosis*; the degree of hepatocyte degeneration and focal necrosis within the lobule; and the degree of *portal inflammation*. Several scoring systems for viral hepatitis that take these histologic features into account have been devised, and the most popular are the histologic activity index (HAI), used commonly in the United States, and the METAVIR score, used in Europe **(Table 352-2)**. Based on the presence and degree of these features of histologic activity, chronic hepatitis can be graded as mild, moderate, or severe.

TABLE 352-1 Clinical and Laboratory Features of Chronic Hepatitis

TYPE OF HEPATITIS	DIAGNOSTIC TEST(S)	THERAPY
Chronic hepatitis B	HBsAg, IgG anti-HBc, HBeAg, HBV DNA	IFN-α, PEG IFN-α Oral agents: First-line: entecavir, tenofovir
Chronic hepatitis C	Anti-HCV, HCV RNA	Sofosbuvir-ledipasvir, glecaprevir-pibrentasvir, sofosbuvir-velpatasvir, elbasvir-grazoprevir, sofosbuvir-velpatasvir-voxilaprevir[a]
Chronic hepatitis D	Anti-HDV, HDV RNA, HBsAg, IgG anti-HBc	IFN-α, PEG IFN-α[c], bulevirtideb
Autoimmune hepatitis	ANA (homogeneous), anti-SMA, anti-SLA, anti-LKM1 (±) Hyperglobulinemia	First line: prednisone, azathioprine Second line: mycophenolate mofetil, tacrolimus, rituximab

See *www.hcvguidelines.org*.

[a] Detailed treatment recommendations can be found at *www.hcvguidelines.org*.
[b] Currently approved by European Medicines Agency, under review at the U.S. Food and Drug Administration.

Abbreviations: HBc, hepatitis B core; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HCV, hepatitis C virus; HDV, hepatitis D virus; IFN-α, interferon α; IgG, immunoglobulin G; LKM, liver-kidney microsome; PEG IFN-α, pegylated interferon α; SLA, soluble liver antigen; SMA, smooth muscle antibody.

CLASSIFICATION BY STAGE

The stage of chronic hepatitis, which reflects the level of progression of the disease, is based on the degree of hepatic fibrosis. When fibrosis is so extensive that fibrous septa surround parenchymal nodules and alter the normal architecture of the liver lobule, the histologic lesion is defined as *cirrhosis*. Staging is based on the degree of fibrosis as categorized on a numerical scale 0–6 (HAI) or 0–4 (METAVIR) (Table 352-2). Several noninvasive approaches have been introduced to provide approximations of hepatic histologic stage, including serum biomarkers of fibrosis; fibrosis scores such as FIB-4, a validated algorithm based on such routine lab tests as aspartate and alanine aminotransferase (AST and ALT) levels and platelet counts (PLT) (age [years] × AST divided by PLT ×); the Enhanced Liver Fibrosis (ELF) score, a proprietary score (applied more frequently in steatotic liver disease than viral liver disease) based on a combination of three nonroutine blood tests (hyaluronic acid, procollagen III N-terminal peptide, and tissue inhibitor of metalloproteinase 1); and imaging determinations of liver elasticity/stiffness.

CHRONIC VIRAL HEPATITIS

Both of the enterically transmitted forms of viral hepatitis, hepatitis A and E, are self-limited and do not cause chronic hepatitis (rare reports notwithstanding in which acute hepatitis A serves as a trigger for the onset of autoimmune hepatitis in genetically susceptible patients or in which hepatitis E **[Chap. 350]** can cause chronic liver disease in immunosuppressed hosts, for example, after liver transplantation). In contrast, the entire clinicopathologic spectrum of chronic hepatitis occurs in patients with chronic viral hepatitis B and C as well as in patients with chronic hepatitis D superimposed on chronic hepatitis B.

CHRONIC HEPATITIS B

The likelihood of chronicity after acute hepatitis B varies as a function of age. Infection at birth is associated with clinically silent acute infection but a 90% chance of chronic infection, whereas infection in young adulthood in immunocompetent persons is typically associated with clinically apparent acute hepatitis but a risk of chronicity of only ~1%. Most cases of chronic hepatitis B among adults, however, are recognized

TABLE 352-2 Histologic Grading and Staging of Chronic Hepatitis

	HISTOLOGIC ACTIVITY INDEX (HAI)[a]		METAVIR[b]	
HISTOLOGIC FEATURE	**SEVERITY**	**SCORE**	**SEVERITY**	**SCORE**
Necroinflammatory Activity (grade)				
Periportal necrosis, including piecemeal necrosis and/or bridging necrosis (BN)	None	0	None	0
	Mild	1	Mild	1
	Mild/moderate	2	Moderate	2
	Moderate	3	Severe	3
	Severe	4	Bridging necrosis	Yes No
Intralobular necrosis Confluent	—None	0	None or mild	0
	—Focal	1	Moderate	1
	—Zone 3 some	2	Severe	2
	—Zone 3 most	3		
	—Zone 3 + BN few	4		
	—Zone 3 + BN multiple	5		
	—Panacinar/multiacinar	6		
Focal	—None	0		
	—≤1 focus/10× field	1		
	—2–4 foci/10× field	2		
	—5–10 foci/10× field	3		
	—>10 foci/10× field	4		
Portal inflammation	None	0		
	Mild	1		
	Moderate	2		
	Moderate/marked	3		
	Marked	4		
	Total	0–18	A0–A3[c]	
Fibrosis (stage)				
None		0	F0	
Portal fibrosis—some		1	F1	
Portal fibrosis—most		2	F1	
Bridging fibrosis—few		3	F2	
Bridging fibrosis—many		4	F3	
Incomplete cirrhosis		5	F4	
Cirrhosis		6	F4	
	Total	6	4	

[a]Ishak K, Baptista A, Bianchi L, et al: Histologic grading and staging of chronic hepatitis. J Hepatol 22:696, 1995. [b]Bedossa P, Poynard T, French METAVIR Cooperative Study Group: An algorithm for grading activity in chronic hepatitis C. Hepatology 24:289, 1996. [c]Necroinflammatory grade: A0 = none; A1 = mild; A2 = moderate; A3 = severe.

in patients who never had a recognized episode of clinically apparent acute viral hepatitis. The degree of liver injury (grade) in patients with chronic hepatitis B is variable, ranging from none in inactive carriers to mild to moderate to severe. Among adults with chronic hepatitis B, histologic features are of prognostic importance. In one long-term study of 379 patients with untreated chronic hepatitis B in the United States, investigators found a 5-year survival rate of 97% for patients with mild chronic hepatitis, 86% for patients with moderate to severe chronic hepatitis, and only 55% for patients with chronic hepatitis and cirrhosis. The 15-year survival in these cohorts was 77%, 66%, and 40%, respectively. On the other hand, more recent observations do not allow us to be so sanguine about the prognosis in patients with mild chronic hepatitis; among such patients followed for 1–13 years, progression to more severe chronic hepatitis and cirrhosis has been observed in more than a quarter of cases. Co-infection with HIV has long been appreciated to accelerate the progression of hepatitis B; early studies documented a nearly 20-fold increase in liver-related mortality between persons with hepatitis B virus (HBV)-HIV co-infection compared to those with HBV infection alone. Similarly, coexisting steatohepatitis has also been found to accelerate disease progression; in one study, patients with HBV infection and steatohepatitis were found to have both higher-grade inflammation and more advanced-stage fibrosis.

More important to consider than histology alone in patients with chronic hepatitis B is the degree of HBV replication. As reviewed in **Chap. 350**, chronic HBV infection can occur in the presence or absence of serum hepatitis B e antigen (HBeAg), and generally, for both HBeAg-reactive and HBeAg-negative chronic hepatitis B, the level of HBV DNA correlates with the level of liver injury and risk of progression. In otherwise healthy adults with *HBeAg-positive chronic hepatitis B*, two phases have been recognized based on the relative level of HBV replication. The relatively *replicative phase* is characterized by the presence in the serum of HBeAg and HBV DNA levels well in excess of 10^3–10^4 IU/mL, sometimes exceeding 10^9 IU/mL; by the presence in the liver of detectable intrahepatocyte nucleocapsid antigens (primarily hepatitis B core antigen [HBcAg]); by high infectivity; and by accompanying liver injury (see below for an important exception during the early decades of life). In contrast, the relatively *nonreplicative phase* is characterized by the *absence* of the conventional serum marker of HBV replication (HBeAg), the appearance of anti-HBe, levels of HBV DNA below a threshold of ~10^3 IU/mL, the absence of intrahepatocytic HBcAg, limited infectivity, and minimal liver injury. Patients in the relatively replicative phase tend to have more severe chronic hepatitis and more rapid progression, whereas those in the relatively nonreplicative phase tend to have minimal or mild chronic hepatitis or to be inactive hepatitis B carriers. Thus, the transition between HBeAg-positive chronic hepatitis B to HBeAg-negative disease is of clinical importance, signaling a slowing of disease progression. In the past, the likelihood of spontaneous HBeAg loss was estimated to be ~10% per year; however, a 2020 systematic review of 26 studies involving 7550 persons with HBeAg-positive chronic hepatitis B revealed a spontaneous HBeAg clearance rate of 6.46 per 100 person-years. Such spontaneous HBeAg loss was found to occur more frequently with older age and in European populations but less frequently in patients with genotype C.

Distinctions in HBV replication and in histologic category, however, do not always coincide. In patients with *HBeAg-positive chronic HBV infection,* especially when acquired at birth or in early childhood, as recognized commonly in Asian countries, a dichotomy is common between very high levels of HBV replication during the early decades of life (when the level of apparent host immunologic tolerance of HBV is relatively high) and negligible levels of liver injury; during this phase of chronic hepatitis B (labeled by some as "HBeAg-positive chronic infection" or "immune tolerant," see **Chap. 350**, "Complications and Sequelae"), the level of viral replication does not correlate with liver injury or late complications. Yet despite the relatively immediate, apparently benign nature of liver disease for many decades in this population, in the middle decades, activation of liver injury emerges as what appears to be the relative tolerance of the host to HBV declines, and these patients with childhood-acquired HBV infection are ultimately at increased risk later in life for cirrhosis, hepatocellular carcinoma (HCC) **(Chap. 87)**, and liver-related death; the link between high-level HBV replication and these late liver complications has been demonstrated convincingly in, and confined mostly to, persons in their middle decades, especially age ≥40. **A discussion of the pathogenesis of liver injury in patients with chronic hepatitis B appears in Chap. 350.**

HBeAg-negative chronic hepatitis B (i.e., chronic HBV infection with active virus replication and readily detectable HBV DNA but without HBeAg [anti-HBe-reactive]) is more common than HBeAg-positive chronic hepatitis B in Mediterranean and European countries and in Asia. Compared to patients with HBeAg-reactive chronic hepatitis B, patients with HBeAg-negative chronic hepatitis B have HBV

DNA levels several orders of magnitude lower (usually no more than 10^5–10^6 IU/mL) than those observed in the HBeAg-reactive subset. Most such HBeAg-negative cases represent precore or core-promoter mutations acquired late in the natural history of mostly early-life-onset disease; these mutations prevent translation of HBeAg from the precore component of the HBV genome (precore mutants) or are characterized by downregulated transcription of precore mRNA (core-promoter mutants; **Chap. 350**). Although their levels of HBV DNA tend to be lower than among patients with HBeAg-reactive chronic hepatitis B, patients with HBeAg-negative chronic hepatitis B can have progressive liver injury (complicated by cirrhosis and HCC) and experience episodic reactivation of liver disease reflected in fluctuating levels of aminotransferase activity ("flares"). The biochemical and histologic activity of HBeAg-negative disease tends to correlate closely with levels of HBV replication. Worth reiterating, the level of HBV replication is the most important risk factor for the ultimate development of cirrhosis and HCC in both HBeAg-reactive and HBeAg-negative patients. Although levels of HBV DNA are lower and more readily suppressed by therapy to undetectable levels in HBeAg-negative (compared to HBeAg-reactive) chronic hepatitis B, achieving sustained responses that permit discontinuation of antiviral therapy is less likely in HBeAg-negative patients (see below). Inactive carriers are patients with circulating hepatitis B surface antigen (HBsAg), normal serum aminotransferase levels, minimal or no histologic evidence of liver injury, undetectable HBeAg, and levels of HBV DNA that are either undetectable or present at a threshold of $\leq 10^3$ IU/mL. This serologic profile occurs not only in inactive carriers but also in patients with HBeAg-negative chronic hepatitis B during periods of relative inactivity; distinguishing between the two requires sequential biochemical and virologic monitoring over many months.

The spectrum of *clinical features* of chronic hepatitis B is broad, ranging from asymptomatic infection to debilitating disease or even end-stage, fatal hepatic failure. As noted above, the onset of the disease tends to be insidious in most patients, apart from the very few in whom chronic disease follows failure of resolution of clinically apparent acute hepatitis B. **The clinical and laboratory features associated with progression from acute to chronic hepatitis B are discussed in Chap. 350.**

Fatigue is a common symptom, and persistent or intermittent *jaundice* is a common feature in severe or advanced cases. Intermittent deepening of jaundice and recurrence of malaise and anorexia, as well as worsening fatigue, are reminiscent of acute hepatitis; such exacerbations may occur spontaneously, often coinciding with evidence of virologic reactivation; may lead to progressive liver injury; and, when superimposed on well-established cirrhosis, may cause hepatic decompensation. Complications of cirrhosis occur in end-stage chronic hepatitis and include ascites, edema, bleeding gastroesophageal varices, hepatic encephalopathy, coagulopathy, and hypersplenism. Occasionally, these complications bring the patient to initial clinical attention. Extrahepatic complications of chronic hepatitis B, similar to those seen during the prodromal phase of acute hepatitis B, are associated with tissue deposition of circulating hepatitis B antigen–antibody immune complexes. These include arthralgias and arthritis, which are common, and the rarer purpuric cutaneous lesions (leukocytoclastic vasculitis), immune-complex glomerulonephritis, and generalized vasculitis (polyarteritis nodosa) **(Chap. 375)**.

Laboratory features of chronic hepatitis B do not always distinguish adequately between histologically mild and severe hepatitis. Aminotransferase elevations tend to be modest for chronic hepatitis B but may fluctuate in the range of 100 to 1000 units. As is true for acute viral hepatitis B, ALT tends to be more elevated than AST; however, once cirrhosis is established, AST tends to exceed ALT. Levels of alkaline phosphatase activity tend to be normal or only marginally elevated. In severe cases, moderate elevations in serum bilirubin (51.3–171 μmol/L [3–10 mg/dL]) occur. Hypoalbuminemia and prolongation of the prothrombin time occur in severe or end-stage cases. Hyperglobulinemia and detectable circulating autoantibodies are distinctly absent in chronic hepatitis B (in contrast to autoimmune hepatitis). **Viral markers of chronic HBV infection are discussed in Chap. 350.**

TREATMENT

Chronic Hepatitis B

Although progression to cirrhosis is more likely in severe than in mild or moderate chronic hepatitis B, all forms of chronic hepatitis B can be progressive, and progression occurs primarily in patients with active HBV replication. Moreover, in populations of patients with chronic hepatitis B who are at risk for HCC **(Chap. 87)**, the risk is highest for those with continued, high-level HBV replication and lower for persons in whom initially high-level HBV DNA falls spontaneously over time. Therefore, management of chronic hepatitis B is directed at suppressing the level of virus replication. Although clinical trials tend to focus on clinical endpoints achieved over 1–2 years (e.g., suppression of HBV DNA to undetectable levels, loss of HBeAg/HBsAg, improvement in histology, normalization of ALT), these short-term gains translate into reductions in the risk of clinical progression, hepatic decompensation, HCC, liver transplantation, and death; regression of cirrhosis and of esophageal varices has been documented to follow long-term pharmacologic suppression of HBV replication. In addition, restoration of impaired HBV-specific T-cell function has been shown following successful suppression of HBV replication with antiviral therapy. To date, eight drugs have been approved for treatment of chronic hepatitis B: injectable interferon (IFN) α and pegylated interferon (long-acting IFN bound to polyethylene glycol, PEG [PEG IFN]) and the oral agents lamivudine, adefovir dipivoxil, entecavir, telbivudine, tenofovir disoproxil fumarate (TDF), and tenofovir alafenamide (TAF).

Antiviral therapy for hepatitis B has evolved rapidly since the mid-1990s, as has the sensitivity of tests for HBV DNA. When IFN and the first oral antiviral lamivudine were evaluated in clinical trials, HBV DNA was measured by insensitive hybridization assays with detection thresholds of 10^5–10^6 virions/mL; when subsequent treatments were studied in clinical trials, HBV DNA was measured by sensitive amplification assays (polymerase chain reaction [PCR]) with detection thresholds of 10^1–10^3 viral copies/mL or IU/mL. Recognition of these distinctions is helpful when comparing results of clinical trials that established the efficacy of these therapies (reviewed below). Of the eight approved treatments, PEG IFN, entecavir, and the two tenofovir preparations (TDF and TAF) are recommended as first-line agents, and generally, the oral agents are favored over injectable PEG IFN.

HISTORICAL, EARLY-GENERATION ANTIVIRAL AGENTS

Interferon IFN-α was the first approved therapy (1992) for chronic hepatitis B. Although it is no longer used to treat hepatitis B, standard IFN is important historically, having provided important lessons about antiviral therapy in general. For immunocompetent adults with *HBeAg-reactive chronic hepatitis B* (who tend to have high-level HBV DNA [>10^5–10^6 virions/mL] and histologic evidence of chronic hepatitis on liver biopsy), a 16-week course of subcutaneous IFN, 5 million units daily or 10 million units thrice weekly, resulted in a loss of HBeAg and hybridization-detectable HBV DNA (i.e., a reduction to levels below 10^5–10^6 virions/mL) in ~30% of patients, with a concomitant improvement in liver histology. Seroconversion from HBeAg to anti-HBe occurred in ~20%, and, in early trials, ~8% lost HBsAg. Successful IFN therapy and seroconversion were often accompanied by an acute hepatitis-like elevation in aminotransferase activity, postulated to result from enhanced cytolytic T-cell clearance of HBV-infected hepatocytes. Relapse after successful therapy was rare (1 or 2%). Responsiveness to IFN was higher in patients with low-level HBV DNA and substantial ALT elevations. Therapy with IFN was not effective in immunosuppressed persons, persons with neonatal acquisition of infection and minimal-to-mild ALT elevations, or patients with decompensated chronic hepatitis B (in whom such therapy was actually detrimental, sometimes precipitating decompensation, often associated with severe adverse effects). After HBeAg loss during IFN therapy, 80% experienced eventual loss of HBsAg and ALT normalization over the ensuing decade. In addition, improved long-term and complication-free survival as well as a reduction in the frequency of HCC were documented among IFN responders, supporting the conclusion that successful antiviral therapy improves the natural history of chronic hepatitis B.

Brief-duration IFN therapy in patients with *HBeAg-negative chronic hepatitis B* was disappointing, suppressing HBV replication transiently during therapy but resulting only rarely in sustained antiviral responses.

Complications of IFN therapy include systemic "flu-like" symptoms; marrow suppression; emotional lability (irritability, depression, anxiety); autoimmune reactions (especially autoimmune thyroiditis); and miscellaneous side effects such as alopecia, rashes, diarrhea, and numbness and tingling of the extremities. With the possible exception of autoimmune thyroiditis, all these side effects are reversible upon dose lowering or cessation of therapy.

Although no longer competitive with the newer generation of antivirals, IFN did represent the first successful antiviral approach, set a standard against which to measure efficacy of subsequent drugs, and demonstrated the benefit of antiviral therapy on the natural history of chronic hepatitis B. Standard IFN has been supplanted by long-acting PEG IFN (see below).

Lamivudine The first of the nucleoside analogues to be approved (in 1998) for hepatitis B, the dideoxynucleoside lamivudine inhibits reverse transcriptase activity of both HIV and HBV and is an effective agent for chronic hepatitis B; however, it is now superseded by newer, more potent, less resistance-prone agents. For a summary of its virologic, serologic, biochemical, and histologic efficacy, as well as its resistance profile, please refer to Table 352-3. In clinical

TABLE 352-3 Comparison of Pegylated Interferon (PEG IFN), Lamivudine, Adefovir, Entecavir, Telbivudine, and Tenofovir Therapy for Chronic Hepatitis B[a]

FEATURE	PEG IFN[b]	LAMIVUDINE	ADEFOVIR	ENTECAVIR	TELBIVUDINE	TENOFOVIR (TDF)	TENOFOVIR (TAF)
Route of administration	Subcutaneous injection (180 μg/week)	Oral (100 mg/d)	Oral (10 mg/d)	Oral (0.5 mg/d)	Oral (600 mg/d)	Oral (300 mg/d)	Oral 25 mg/d)
Status	First-line	No longer preferred	No longer preferred	First-line	No longer preferred, withdrawn	First-line	First-line
Duration of therapy[c]	48–52 weeks	≥52 weeks	≥48 weeks	≥48 weeks	≥52 weeks	≥48 weeks	48 weeks
Tolerability	Poorly tolerated	Well tolerated	Well tolerated; creatinine monitoring recommended	Well tolerated	Well tolerated	Well tolerated; creatinine monitoring recommended	Well tolerated
HBeAg seroconversion							
1 yr Rx	18–20%	16–21%	12%	21%	22%	21%	10% (14% HBeAg loss)
>1 yr Rx	NA	up to 50% at 5 yrs	43% at 3 yrs[d]	31% at 2 yrs 44% at 6 yrs	30% at 2 yrs	40% at 5 yrs	18% at yr 2 (HBeAg loss 22%)
Log_{10} HBV DNA reduction (mean copies/mL)							Not reported in clinical trials, likely same as TDF
HBeAg-reactive	4.5	5.5	Median 3.5–5	6.9	6.4	6.2	
HBeAg-negative	4.1	4.4–4.7	Median 3.5–3.9	5.0	5.2	4.6	
HBV DNA PCR negative (at then current PCR sensitivity[a]) at end of yr 1							
HBeAg-reactive	10–25%	36–44%	13–21%	67% (91% at 4 yrs)	60%	76%	64%
HBeAg-negative	63%	60–73%	48–77%	90%	88%	93%	94%
ALT normalization at end of yr 1							
HBeAg-reactive	39%	41–75%	48–61%	68%	77%	68%	72%
HBeAg-negative	34–38%	62–79%	48–77%	78%	74%	76%	83%
HBsAg loss, yr 1	3–4%	≤1%	0%	2%	<1%	3%	1%
>yr 1	12% 5 yr after 1 yr of Rx	No data	5% at yr 5	6% at yr 6	No data	10% at yr 5	1%
Histologic improvement (≥2 point reduction in HAI) at yr 1							Not included in clinical trials
HBeAg-reactive	38% 6 months after	49–62%	53–68%	72%	65%	74%	
HBeAg-negative	48% 6 months after	61–66%	64%	70%	67%	72%	
Viral resistance	None	15–30% at 1 yr 70% at 5 yrs	None at 1 yr 29% at 5 yrs	≤1% at 1 yr[e] 1.2% at 6 yrs[e]	Up to 5% at yr 1 Up to 22% at yr 2	0% at yr 1 0% through yr 8	0% at yr 1 0% through yr 2
Pregnancy category	C	C[f]	C	C	B	B	B

[a]Generally, these comparisons are based on data on each drug tested individually versus placebo in registration clinical trials; with rare exception, these comparisons are not based on head-to-head testing of these drugs. In addition, the sensitivity of HBV DNA assays increased in sensitivity over the two decades between the introduction of the earliest and latest of these approved drugs. Therefore, relative advantages and disadvantages should be interpreted cautiously. [b]Although standard interferon α administered daily or three times a week is approved as therapy for chronic hepatitis B, it has been supplanted by PEG IFN, which is administered once a week and is more effective. Standard interferon has no advantages over PEG IFN. [c]Duration of therapy in clinical efficacy trials; use in clinical practice may vary. [d]Because of a computer-generated randomization error that resulted in misallocation of drug versus placebo during the second year of clinical trial treatment, the frequency of HBeAg seroconversion beyond the first year is an estimate (Kaplan-Meier analysis) based on the small subset in whom adefovir was administered correctly. [e]7% during a year of therapy (43% at year 4) in lamivudine-resistant patients. [f]Despite its category C designation, lamivudine has an extensive pregnancy safety record in women with HIV/AIDS.

Abbreviations: ALT, alanine aminotransferase; HAI, histologic activity index; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; NA, not applicable; PEG IFN, pegylated interferon; PCR, polymerase chain reaction; Rx, therapy; yr, year; TAF, tenofovir alafenamide; TDF, tenofovir disoproxil fumarate.

trials, lamivudine therapy at daily doses of 100 mg for 48–52 weeks suppressed HBV DNA, as measured by sensitive PCR amplification assays, by a median of ~5.5 $\log_{10}$ copies/mL in HBeAg-positive chronic hepatitis B and ~4.5 $\log_{10}$ copies/mL in HBeAg-negative chronic hepatitis B (baseline HBV DNA levels are lower in HBeAg-negative than in HBeAg-positive chronic hepatitis B) and to undetectable levels in ~40% and ~70%, respectively. Lamivudine was shown to improve histology, retard hepatic fibrosis, prevent progression to cirrhosis, and, in patients with advanced fibrosis, reduce decompensation. Lamivudine was also effective in patients unlikely to respond to IFN (e.g., those with high-level HBV DNA) or in prior IFN nonresponders.

In HBeAg-positive chronic hepatitis B, lamivudine-associated HBeAg seroconversion occurred in ~20%, as was true for IFN therapy. Also similar to IFN therapy, patients with near-normal ALT activity tended not to experience HBeAg responses (despite suppression of HBV DNA), while patients with ALT levels ≥5× the upper limit of normal could expect 1-year HBeAg seroconversion rates of 50–60%. Generally, HBeAg seroconversions were confined to patients who achieved suppression of HBV DNA to $<10^4$ copies/mL (equivalent to ~10^3 IU/mL). Lamivudine-associated HBeAg responses were accompanied by a delayed posttreatment HBsAg seroconversion rate comparable to that seen after IFN. Among Western patients who experienced HBeAg responses during a year-long course of therapy and in whom the response was sustained for 4–6 months after cessation of therapy, the response was durable thereafter in the vast majority (>80%); therefore, *the achievement of an HBeAg response represented a viable stopping point in therapy for HBeAg-positive patients*. Reduced durability, however, was reported in Asian patients; therefore, to support the durability of HBeAg responses, a period of consolidation therapy of ≥6 months in Western patients and ≥1 year in Asian patients was recommended after HBeAg seroconversion (see treatment guidelines below; a full 12-month consolidation period is recommended currently for treatment extension after oral agent–induced HBeAg seroconversion). Close posttreatment monitoring was recommended to identify HBV reactivation promptly and to resume therapy. If HBeAg was unaffected by lamivudine therapy, lamivudine was continued until an HBeAg response occurred, but long-term therapy was required to suppress HBV replication and, in turn, limit liver injury; HBeAg seroconversions increased to a level of 50% after 5 years of therapy. After a cumulative course of 3 years of lamivudine therapy, necroinflammatory activity was reduced in the majority of patients, and even cirrhosis was shown to regress to precirrhotic stages in as many as three-quarters of patients.

Losses of HBsAg were few during the first year of lamivudine therapy, and this observation had been cited as an advantage of IFN-based therapy over lamivudine therapy; however, in head-to-head comparisons between standard IFN and lamivudine monotherapy, HBsAg losses were rare in both groups. Trials in which lamivudine and IFN were administered in combination failed to show a benefit of combination therapy over lamivudine monotherapy for either treatment-naïve patients or prior IFN nonresponders.

Patients with *HBeAg-negative chronic hepatitis B* (i.e., in those with precore or core-promoter HBV mutations and who lack HBeAg) cannot achieve an HBeAg response to nucleoside analogue therapy—a stopping point in HBeAg-reactive patients; almost invariably, when therapy was discontinued, reactivation was the rule. Therefore, these patients required long-term lamivudine therapy.

Clinical and laboratory side effects of lamivudine were negligible and indistinguishable from those observed in placebo recipients; however, lamivudine doses were reduced in patients with reduced creatinine clearance. During lamivudine therapy, transient ALT elevations, resembling those seen during IFN therapy and during spontaneous HBeAg-to-anti-HBe seroconversions, occurred in one-fourth of patients. These ALT elevations may result from restored cytolytic T-cell activation permitted by suppression of HBV replication. Similar ALT elevations, however, occurred at an identical frequency in placebo recipients; however, ALT elevations associated temporally with HBeAg seroconversion in clinical trials were confined to lamivudine-treated patients. When therapy was stopped after a year, two- to threefold ALT elevations occurred in 20–30% of lamivudine-treated patients, representing renewed liver-cell injury as HBV replication returned. Although these posttreatment flares were almost always transient and mild, rare severe exacerbations, especially in cirrhotic patients, were observed, mandating close and careful clinical and virologic monitoring after discontinuation of treatment. Many authorities cautioned against discontinuing therapy in patients with cirrhosis, in whom posttreatment flares could precipitate decompensation.

Long-term monotherapy with lamivudine was associated with methionine-to-valine (M204V) or methionine-to-isoleucine (M204I) mutations, primarily at amino acid 204 in the tyrosine-methionine-aspartate-aspartate (YMDD) motif of the C domain of HBV DNA polymerase, analogous to mutations that occur in HIV-infected patients treated with this drug. During a year of therapy, YMDD mutations occurred in 15–30% of patients; the frequency increased with each year of therapy, reaching 70% at year 5. Ultimately, patients with YMDD mutants experienced degradation of clinical, biochemical, and histologic responses; therefore, if treatment was begun with lamivudine monotherapy, the emergence of lamivudine resistance, reflected clinically by a breakthrough from suppressed levels of HBV DNA and ALT, was managed by adding another antiviral to which YMDD variants are sensitive (e.g., adefovir, tenofovir; see below).

Currently, lamivudine has been eclipsed by more potent antivirals that have superior resistance profiles (see below); it is no longer recommended as first-line therapy. Still, as the first successful oral antiviral agent for use in hepatitis B, lamivudine provided proof of principle that polymerase inhibitors can achieve virologic, serologic, biochemical, and histologic benefits, including retardation and reversal of fibrosis and even of cirrhosis.

Cirrhosis In addition, lamivudine was shown to be effective in the treatment of patients with decompensated hepatitis B (for whom IFN is contraindicated), in some of whom decompensation can be reversed. Moreover, among patients with cirrhosis or advanced fibrosis, lamivudine was shown to be effective in reducing the risk of progression to hepatic decompensation and, based on subsequent population studies, the risk of HCC. In the half decade following the introduction in the United States of lamivudine therapy for hepatitis B, referral of patients with HBV-associated end-stage liver disease for liver transplantation fell by ~30%, supporting further the beneficial impact of oral antiviral therapy on the natural history of chronic hepatitis B.

HIV Co-infection Because lamivudine monotherapy in persons with HIV infection can result universally in the rapid emergence of YMDD variants, testing for HIV infection was recommended for all patients with chronic hepatitis B prior to lamivudine therapy; if HIV infection was identified, lamivudine monotherapy at the HBV daily dose of 100 mg was contraindicated. These patients require treatment for both HIV and HBV with an HIV drug regimen that includes or is supplemented by at least two drugs active against HBV; antiretroviral therapy (ART) often contains two drugs with antiviral activity against HBV (e.g., tenofovir and emtricitabine), but if lamivudine was part of the regimen, the 300-mg daily dose was required **(Chap. 208)**. The safety of lamivudine during pregnancy has not been established; however, the drug is not teratogenic in rodents and has been used safely in pregnant women with HIV infection and with HBV infection. As shown for subsequent nucleoside analogues, administration of lamivudine during the last months of pregnancy to mothers with high-level hepatitis B viremia reduced the likelihood of perinatal transmission of hepatitis B.

Adefovir Dipivoxil At an oral daily dose of 10 mg, the acyclic nucleotide analogue adefovir dipivoxil, the prodrug of adefovir (approved for hepatitis B in 2002), reduces HBV DNA by ~3.5–4 $\log_{10}$

copies/mL; i.e., it is less potent than lamivudine or any of the newer antiviral agents. For a summary of its virologic, serologic, biochemical, and histologic efficacy, as well as its resistance profile, please refer to Table 352-3. Like IFN and lamivudine, adefovir dipivoxil is more likely to achieve an HBeAg response in patients with high baseline ALT; HBeAg responses to it are highly durable and can be relied upon as a treatment stopping point, after a period of consolidation therapy; and biochemical, serologic, and virologic outcomes improve over time with continued therapy.

In HBeAg-negative chronic hepatitis B, as was true for lamivudine, because HBeAg responses—a potential stopping point—cannot be achieved, reactivation is the rule when adefovir therapy is discontinued, and indefinite, long-term therapy is required. Reported attempts to stop adefovir after 5 years were followed by a period of maintained suppression of HBV DNA and ALT; however, most such patients had persistent hepatitis B viremia, and most HBeAg-negative patients were treated indefinitely unless HBsAg loss, albeit very rare, was achieved.

Adefovir was less prone to resistance than lamivudine, and thus the primary contribution of adefovir, its effectiveness in lamivudine-resistant, YMDD-mutant HBV, led to its adoption for lamivudine-resistant hepatitis B. When lamivudine resistance occurred, adding adefovir (i.e., maintaining lamivudine to preempt the emergence of adefovir resistance) was superior to switching to adefovir.

An advantage of adefovir was its relatively favorable resistance profile; however, it was not as potent as the other approved oral agents, it did not suppress HBV DNA as rapidly or as uniformly as the others, it was the least likely of all agents to result in HBeAg seroconversion, and 20–50% of patients failed to suppress HBV DNA by 2 $\log_{10}$ (“primary nonresponders”). For these reasons, adefovir, which has been supplanted in both treatment-naïve and lamivudine-resistant patients by the more potent, less resistance-prone tenofovir (see below), is no longer recommended as first-line therapy.

Telbivudine Telbivudine, a cytosine analogue (approved in 2006), is similar in efficacy to entecavir (see below) but slightly less potent in suppressing HBV DNA (median 6.4 $\log_{10}$ reduction in HBeAg-reactive disease and a 5.2 $\log_{10}$ reduction in HBeAg-negative disease). In its registration trial, telbivudine at an oral daily dose of 600 mg suppressed HBV DNA to <300 copies/mL in 60% of HBeAg-positive and 88% of HBeAg-negative patients, reduced ALT to normal in 77% of HBeAg-positive and 74% of HBeAg-negative patients, and improved histology in 65% of HBeAg-positive and 67% of HBeAg-negative patients. Although resistance to telbivudine (M204I, not M204V, mutations) was less frequent than resistance to lamivudine at the end of 1 year, resistance mutations after 2 years of treatment occurred in up to 22%. Generally well tolerated, telbivudine was associated with a low frequency of asymptomatic creatine kinase elevations and with a very low frequency of myopathy and peripheral neuropathy; frequency of administration had to be reduced for patients with impaired creatinine clearance. Its excellent potency notwithstanding, the inferior resistance and safety profile of telbivudine limited its appeal; telbivudine is neither recommended as first-line therapy nor used widely.

CURRENT FIRST-LINE TREATMENT OPTIONS

PEGYLATED IFN After long-acting PEG IFN was shown to be effective in the treatment of hepatitis C (see below), this more convenient IFN preparation was evaluated in the treatment of chronic hepatitis B. Once-a-week PEG IFN is more effective than the more frequently administered, standard IFN, and several large-scale trials of PEG IFN versus oral lamivudine were conducted in patients with chronic hepatitis B.

In HBeAg-reactive chronic hepatitis B, two large-scale studies were done. In one study, PEG IFN-α2b (100 μg weekly for 32 weeks, then 50 μg weekly for another 20 weeks for a total of 52 weeks) was evaluated against a comparison arm of combination PEG IFN with oral lamivudine in 307 subjects. The other study involved PEG IFN-α2a (180 μg weekly for 48 weeks) in 814 primarily Asian patients, three-fourths of whom had ALT ≥2× the upper limit of normal, with comparison arms of lamivudine monotherapy and combination PEG IFN plus lamivudine. At the end of therapy (48–52 weeks) in the PEG IFN monotherapy arms, HBeAg loss occurred in ~30%, HBeAg seroconversion in 22–27%, undetectable HBV DNA (<400 copies/mL by PCR) in 10–25%, and normal ALT in 34–39%, and a mean reduction in HBV DNA of 2 $\log_{10}$ copies/mL (PEG IFN-α2b) to 4.5 $\log_{10}$ copies/mL (PEG IFN-α2a) was seen. Six months after completing PEG IFN monotherapy in these trials, HBeAg losses were present in ~35%, HBeAg seroconversion in ~30%, undetectable HBV DNA in 7–14%, and normal ALT in 32–41%, and the mean reduction in HBV DNA was 2–2.4 $\log_{10}$ copies/mL. Although the combination of PEG IFN and lamivudine was superior at the end of therapy in one or more serologic, virologic, or biochemical outcomes, neither the combination arm (in both studies) nor the lamivudine monotherapy arm (in the PEG IFN-α2a trial) demonstrated any benefit compared to the PEG IFN monotherapy arms 6 months after therapy. Moreover, HBsAg seroconversion occurred in 3–7% of PEG IFN recipients (with or without lamivudine); some of these seroconversions were identified by the end of therapy, but many were identified during the posttreatment follow-up period. The likelihood of HBeAg loss in PEG IFN–treated HBeAg-reactive patients was associated with HBV genotype A > B > C > D (shown for PEG IFN-α2b but not for PEG IFN-α2a). PEG IFN-α2a was approved in the United States for hepatitis B in 2005; PEG IFN-α2b, which is not approved for hepatitis B in the United States, is used in other countries.

Based on these results, some authorities concluded that PEG IFN monotherapy should be the first-line therapy of choice in *HBeAg-reactive chronic hepatitis B*; however, this conclusion has been challenged. Although a finite, 1-year course of PEG IFN results in a higher rate of sustained response (6 months after treatment) than is achieved with oral nucleoside/nucleotide analogue therapy, the comparison is confounded by the fact that oral agents are not discontinued at the end of 1 year. Instead, taken orally and free of side effects, therapy with oral agents is extended indefinitely or until after the occurrence of an HBeAg response. The rate of HBeAg responses after 2 years of oral-agent nucleoside analogue therapy is at least as high as, if not higher than, that achieved with PEG IFN after 1 year; favoring oral agents is the absence of injections, difficult-to-tolerate side effects, and laboratory monitoring as well as lower direct and indirect medical care costs and inconvenience. HBsAg loss with PEG IFN therapy occurs in such a small proportion of patients that subjecting everyone to PEG IFN for the marginal gain of HBsAg responses during or immediately after therapy in such a very small minority is questionable. Moreover, with the newer, more potent nucleoside analogues, the frequency of HBsAg loss during the first year of therapy nearly equals that of PEG IFN. Of course, resistance is not an issue during PEG IFN therapy, but the risk of resistance is much lower with new agents (≤1% up to 3–8 years in previously treatment-naïve, entecavir-treated patients and 0% in tenofovir-treated patients; see below). Finally, the level of HBV DNA inhibition that can be achieved with the newer agents, and even with lamivudine, exceeds that achieved with PEG IFN, in some cases by several orders of magnitude.

In *HBeAg-negative chronic hepatitis B*, the role of PEG IFN is uncertain. One trial of PEG IFN-α2a (180 μg weekly for 48 weeks vs comparison arms of lamivudine monotherapy and of combination therapy) in 564 patients showed that PEG IFN monotherapy resulted at the end of therapy in suppression of HBV DNA by a mean of 4.1 $\log_{10}$ copies/mL, undetectable HBV DNA (<400 copies/mL by PCR) in 63%, normal ALT in 38%, and loss of HBsAg in 4%. Although lamivudine monotherapy and combination lamivudine–PEG IFN therapy were both superior to PEG IFN at the end of therapy, no advantage of lamivudine monotherapy or combination therapy was apparent over PEG IFN monotherapy 6 months after therapy—suppression of HBV DNA by a mean of 2.3 $\log_{10}$ copies/mL, undetectable HBV DNA in 19%, and normal ALT in 59%. In patients involved in this trial followed for up to 5 years, among the two-thirds followed who had been treated initially with PEG IFN, 17% maintained HBV DNA suppression to <400 copies/mL, but ALT remained normal in only 22%; HBsAg loss increased gradually to 12%. As was the case for standard IFN therapy in HBeAg-negative patients, only a small proportion maintained responsiveness after completion of PEG IFN therapy, raising questions about the relative value of a finite period of PEG IFN versus a longer course with a potent, low-resistance oral nucleoside analogue in these patients. Moreover, the value of PEG IFN for HBeAg-negative chronic hepatitis B has not been confirmed. In the only other controlled clinical trial of PEG IFN for HBeAg-negative chronic hepatitis B, the hepatitis C regimen of PEG IFN plus ribavirin was compared to PEG IFN monotherapy. In this trial, HBV DNA suppression (<400 copies/mL) occurred in only 7.5% of the two groups combined, and no study subject lost HBsAg.

Reductions in quantitative HBsAg levels have been shown to correlate with and to be predictive of responsiveness to PEG IFN in chronic hepatitis B. If HBsAg levels fail to fall within the first 12–24 weeks or to reach <20,000 IU/mL by week 24, PEG IFN therapy is unlikely to be effective and should be discontinued. (Similar observations of HBsAg levels in oral agent–treated patients are of interest but of limited clinical relevance, given the very high likelihood of virologic responses during such therapy.) While PEG IFN remains one of the recommended first-line agents for hepatitis B, subsequent-generation, injection-free, very-well-tolerated, high-barrier-to-resistance, oral agents are used much more widely.

Entecavir Entecavir, an oral cyclopentyl guanosine analogue polymerase inhibitor (approved in 2005), is a highly potent HBV antiviral and is just as well tolerated as lamivudine. In *HBeAg-positive patients*, one 709-subject clinical trial compared oral entecavir, 0.5 mg daily, with lamivudine, 100 mg daily. At 48 weeks, entecavir was superior to lamivudine in suppression of HBV DNA (mean 6.9 vs 5.5 $\log_{10}$ copies/mL), percentage with undetectable HBV DNA (<300 copies/mL by PCR; 67 vs 36%), histologic improvement (≥2-point improvement in necroinflammatory HAI score; 72 vs 62%), and normal ALT (68 vs 60%). The two treatments were indistinguishable in percentage with HBeAg loss (22 vs 20%) and seroconversion (21 vs 18%). Among patients treated with entecavir for 96 weeks, HBV DNA was undetectable cumulatively in 80% (vs 39% for lamivudine), and HBeAg seroconversions had occurred in 31% (vs 26% for lamivudine). After 3–6 years of entecavir, HBeAg seroconversions were observed in 39–44% and HBsAg loss in 5–6%. Similarly, in a 638-subject clinical trial among *HBeAg-negative patients*, at week 48, oral entecavir, 0.5 mg daily, was superior to lamivudine, 100 mg daily, in suppression of HBV DNA (mean 5.0 vs 4.5 $\log_{10}$ copies/mL) and in percentage with undetectable HBV DNA (90 vs 72%), histologic improvement (70 vs 61%), and normal ALT (78 vs 71%). No resistance mutations were encountered in previously treatment-naïve, entecavir-treated patients during 96 weeks of therapy, and in a cohort of subjects treated for up to 6 years, resistance emerged in only 1.2%. Entecavir-induced HBeAg seroconversions are as durable as those achieved with other antivirals. Its high barrier to resistance coupled with its high potency renders entecavir a first-line drug for patients with chronic hepatitis B.

Entecavir is also effective against lamivudine-resistant HBV infection. In a trial of 286 lamivudine-resistant patients, entecavir, at a higher daily dose of 1 mg, was superior to lamivudine, as measured at week 48, in achieving suppression of HBV DNA (mean 5.1 vs 0.48 $\log_{10}$ copies/mL), undetectable HBV DNA (72 vs 19%), normal ALT (61 vs 15%), HBeAg loss (10 vs 3%), and HBeAg seroconversion (8 vs 3%). In this population of lamivudine-experienced patients, however, entecavir resistance emerged in 7% at 48 weeks. Although entecavir resistance requires both a YMDD mutation and a second mutation at one of several other sites (e.g., T184A, S202G/I, or M250V), resistance to entecavir in lamivudine-resistant chronic hepatitis B was reported to increase progressively to 43% at 4 years and 57% at 6 years; therefore, entecavir is not as attractive a choice (and is not recommended, despite its approval for this indication) as adefovir was or as tenofovir is for patients with lamivudine-resistant hepatitis B.

In clinical trials, entecavir had an excellent safety profile. In addition, on-treatment and posttreatment ALT flares are relatively uncommon and relatively mild in entecavir-treated patients. Doses should be reduced for patients with reduced creatinine clearance. Entecavir does have low-level antiviral activity against HIV and, therefore, cannot be used as monotherapy to treat HBV infection in HIV/HBV co-infected persons.

Tenofovir Tenofovir disoproxil fumarate (TDF), an acyclic nucleotide analogue and potent antiretroviral agent used to treat HIV infection (approved for hepatitis B in 2008), is similar to adefovir but more potent in suppressing HBV DNA and inducing HBeAg responses; it is highly active against both wild-type and lamivudine-resistant HBV and active in patients whose response to adefovir is slow and/or limited. At an oral once-daily dose of 300 mg for 48 weeks, tenofovir suppressed HBV DNA by 6.2 $\log_{10}$ (to undetectable levels [<400 copies/mL] in 76%) in HBeAg-positive patients and by 4.6 $\log_{10}$ (to undetectable levels in 93%) in HBeAg-negative patients; reduced ALT to normal in 68% of HBeAg-positive and 76% of HBeAg-negative patients; and improved histology in 74% of HBeAg-positive and 72% of HBeAg-negative patients. In HBeAg-positive patients, HBeAg seroconversions occurred in 21% by the end of year 1, 27% by year 2, 34% by year 3, and 40% by year 5 of tenofovir treatment; HBsAg loss occurred in 3% by the end of year 1, 6% at year 2, and 10% by year 5. After 5 years of tenofovir therapy, 87% of patients experienced histologic improvement, including reduction in fibrosis score (51%) and regression of cirrhosis (74%). The 5-year safety (negligible renal toxicity, in 1%, and mild reduction in bone density, in ~0.5%) and resistance profiles (none recorded through 8 years) of tenofovir are very favorable as well; therefore, tenofovir has supplanted adefovir both as first-line therapy for chronic hepatitis B and as rescue therapy for lamivudine-resistant chronic hepatitis B. Studies of tenofovir and entecavir reviewed in 2015 showed no difference in long-term risks of renal and bone toxicity; however, among patients treated with tenofovir, instances of acute renal failure and of low blood phosphate levels have been reported. Thus, in patients receiving tenofovir, monitoring bone density is not recommended, but periodic (at least annual) monitoring for renal injury is recommended (serum creatinine and phosphate, urine glucose and protein). Frequency of tenofovir administration should be reduced for patients with impaired creatinine clearance.

Tenofovir alafenamide (TAF), a second-generation tenofovir (approved for hepatitis B in 2016), is a prodrug of tenofovir that requires activation to tenofovir in hepatocytes. This targeted delivery to hepatocytes allows a lower dose to suffice and reduces systemic exposure by 90%, thereby minimizing TDF-associated proximal tubular renal injury, its associated phosphate wasting, and the potential consequent loss of bone mineral density. The dose of TAF is 25 mg, which is equivalent in antiviral potency to 300 mg of TDF; both formulations have the same high barrier to resistance, and clinical resistance has not been encountered. Randomized, controlled, double-blind, phase 3 noninferiority trials, one in HBeAg-positive patients and the other in HBeAg-negative patients, provided the safety and efficacy data to support TAF approval.

In 873 *HBeAg-positive* patients treated for 48 weeks, TAF versus TDF achieved (1) HBV DNA reductions to <29 IU/mL in 64% versus 67%; (2) ALT normalization in 72% versus 67% (an unexplained TAF biochemical advantage confirmed in other trials); (3) HBeAg loss in 14% versus 12%; (4) HBeAg seroconversion in 10% versus 8%; and (5) a negligible loss of HBsAg in 1% versus 0.3%. Compared to TDF, TAF was associated with reduced impairment of renal function (median reduction in estimated glomerular filtration rate of −0.6 mL/min for TAF vs −5.4 mL/min for TDF) and of bone density (in hip measurements, mean reduction of −0.10% for TAF vs −1.72% for TDF; adjusted difference, 1.62%).

In the parallel trial among 426 *HBeAg-negative* patients treated for 48 weeks, reductions in HBV DNA to <29 IU/mL occurred in 94% versus 93% of individuals treated with TAF versus TDF, respectively; normalization of ALT occurred in 83% versus 75%, but no HBsAg loss occurred in either group. Similar TAF advantages in maintaining renal function and bone density were reported: reduction in median estimated glomerular filtration rate (−1.8 mL/min for TAF vs −4.8 mL/min for TDF) and in median bone density (in hip measurements, mean reduction of −0.29% for TAF vs −2.16% for TDF; adjusted percentage difference, 1.87%).

At week 96, TAF and TDF HBV DNA and ALT reductions (including the TAF advantage observed at 48 weeks) were maintained. In the original TDF group, when TDF was switched to TAF after week 96, all differences observed during the first 96 weeks (in normalization of ALT and reductions in renal function and bone density) had resolved at week 120. Resistance did not emerge to either TAF or TDF throughout the trial.

Based on these trial outcomes, TAF joined the list of recommended first-line antiviral agents for chronic hepatitis B. This drug is recommended over TDF by the American Association for the Study of Liver Diseases (AASLD) and the European Association for the Study of the Liver (EASL) for patients with reduced renal function (creatinine clearance <50 mL/min), reduced bone density, and risk factors for renal injury (including, according to EASL guidelines, decompensated cirrhosis, creatinine clearance <60 mL/min, poorly controlled hypertension or diabetes, proteinuria, active glomerulonephritis, concomitant nephrotoxic medications, or solid-organ transplantation); the EASL recommendation extends to persons >60 years, who are at increased risk of TDF nephrotoxicity. In patients with creatinine clearances <15 mL/min, neither TDF nor TAF is recommended.

A comparison of antiviral therapies for chronic hepatitis B appears in Table 352-3; their relative potencies in suppressing HBV DNA are shown in Fig. 352-1.

Combination Therapies Adding a nucleos(t)ide inhibitor to PEG IFN does not appear to improve the frequency of HBeAg response or other treatment endpoints. Although the combination of lamivudine and PEG IFN suppresses HBV DNA more profoundly during therapy than does monotherapy with either drug alone (and is much less likely to be associated with lamivudine resistance), this combination used for a year is no better than a year of PEG IFN in achieving sustained responses. To date, combinations of oral nucleoside/nucleotide agents have not achieved an enhancement in virologic, serologic, or biochemical efficacy over that achieved by the more potent of the combined drugs given individually. In a 2-year trial of combination entecavir and tenofovir versus entecavir monotherapy, for a small subgroup of patients with very high HBV DNA levels ($\geq 10^8$ IU/mL), a reduction in HBV DNA to <50 IU/mL was higher in the combination group (79 vs 62%); however, no differences in HBeAg responses or any other endpoint were observed between the combination-therapy and monotherapy groups, even in the high-HBV DNA subgroup.

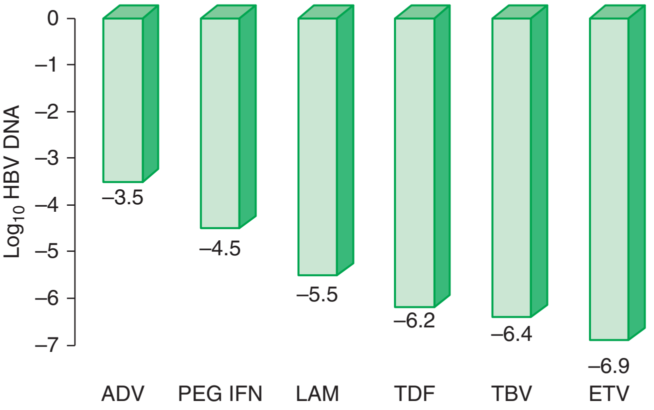

FIGURE 352-1 Relative potency of antiviral drugs for hepatitis B, as reflected by median log$_{10}$ hepatitis B virus (HBV) DNA reduction in HBeAg-positive chronic hepatitis B. These data are from individual reports of large, randomized controlled registration trials that were the basis for approval of the drugs. In most instances, these data do not represent direct comparisons among the drugs, because study populations were different, baseline patient variables were not always uniform, and the sensitivity and dynamic range of the HBV DNA assays used in the trials varied. ADV, adefovir dipivoxil; ETV, entecavir; LAM, lamivudine; PEG IFN, pegylated interferon α2a; TBV, telbivudine; TDF, tenofovir disoproxil fumarate. Because of potency and a high barrier to resistance, ETV and tenofovir (either TDF or the second-generation tenofovir alafenamide) are recommended as first-line therapy. While PEG IFN remains a first-line agent, the oral agents developed earlier, LAM, ADV, and TBV, are no longer preferred agents.

In the setting of resistant virus, combinations of nucleos(t)ide inhibitors can be effective. For resistance to lamivudine or adefovir, adding a second, non-cross-resistant agent was previously the chosen approach. Whereas, initially, in clinical studies of adefovir as rescue therapy for lamivudine resistance, *adding* adefovir to lamivudine (combination therapy) was considered a better strategy than replacing lamivudine with adefovir monotherapy (to minimize ALT flares and to avoid adefovir resistance), newer antiviral therapies have altered the paradigm. According to current treatment recommendations of the AASLD and the EASL, *switching* from the resistant drug to the new drug is preferred. Because the current generation of antivirals is so potent and has such a high barrier to resistance, monotherapy with the rescue drug (e.g., tenofovir for lamivudine resistance) is as effective (as demonstrated in observational reports for up to 5 years) in maintaining viral suppression without the emergence of resistance as combination therapy with the resistant drug and the rescue drug. Generally, in patients treated initially with entecavir and tenofovir preparations, antiviral drug resistance is no longer encountered. For currently rare patients who already have acquired multidrug resistance (to both nucleoside analogues [lamivudine, entecavir, telbivudine] and nucleotide analogues [adefovir, tenofovir]), treatment with a combination of entecavir and tenofovir has been shown to be highly effective in suppressing HBV DNA and overcoming drug resistance.

In addition to the eight approved antiviral drugs for hepatitis B, emtricitabine, a fluorinated cytosine analogue very similar to lamivudine in structure, efficacy, and resistance profile, offers no advantage over lamivudine. A combination of emtricitabine and tenofovir is approved for the treatment of HIV infection and is an appealing combination therapy for hepatitis B, especially for lamivudine-resistant disease; however, neither emtricitabine nor the combination is approved for hepatitis B. Worth considering, however, in patients with chronic hepatitis B who are at high risk of human immunodeficiency virus (HIV) exposure, the combination of TDF and emtricitabine (FTC) may have a unique but complicated role. Although clinical trial data are limited, combined TDF/FTC is used often as preexposure prophylaxis (PrEP) to prevent HIV infection, with a 94% reduction in new HIV infection. Thus, in patients with chronic hepatitis B who require treatment for their hepatitis B but who also desire PrEP, TDF/FTC may be an appealing treatment option. Adherence is essential, however, because virologic and clinical reactivation of hepatitis B is common if TDF/FTC is discontinued.

Novel Antivirals and Strategies The current generation of well-tolerated oral antivirals has been very successful in the management of chronic hepatitis B and in altering its natural history favorably; however, most patients require long-duration, usually indefinite, therapy. Therefore, the quest continues for novel antiviral approaches that can achieve "cure" but are finite in duration. The cure of HBV infection is defined as either "functional cure," with loss of HBsAg and undetectable DNA (that are durable for at least 6 months after treatment), with or without the emergence of anti-HBs, or as "complete (sterilizing) cure," in which covalently closed circular DNA (cccDNA) is eradicated as well. Even an approach to achieving "functional cure" with finite-duration therapy, which is elusive in patients treated with nucleoside analogues, would be welcome. Currently, innovative approaches being investigated focus on viral-targeting strategies or immunomodulatory strategies. Current virologic approaches in clinical testing include HBV entry blockers, therapies to block viral RNA translation, and HBV capsid assembly modulators. Bulevirtide, an entry inhibitor blocking the receptor for HBV on cell membranes, sodium taurocholate co-transporting polypeptide (NTCP), that decreases new cccDNA production, has been approved in the European Union for treatment of HBV/HDV co-infection (see HDV, below) but has not met with success in treating HBV mono-infection. Several therapies targeting viral RNA are under investigation. These include RNA interference (RNAi) therapies and antisense oligonucleotides (ASO). Experimental RNAi therapies tend to reduce HBsAg by ~2 $\log_{10}$ IU/mL during therapy, but ultimately, HBsAg reappears over time after therapy. An early signal of efficacy was reported for the ASO bepirovirsen, which achieved functional cure in ~10% of patients treated with six monthly injections (as monotherapy or in combination with nucleoside analogues); however, the durability of this response was degraded over ensuing follow-up, yielding a very small residual rate of functional cure. Drugs to inhibit capsid assembly have the potential to modulate cccDNA synthesis and contribute to functional cure; however, while this class of drugs reduces HBV DNA by ~2–3 $\log_{10}$ IU/mL, HBsAg levels remain unaffected. Moreover, at least one experimental drug in each of these categories of virologic inhibitors had to be abandoned for toxicity. Immunomodulators being studied have included toll-like receptor agonists (TLR7, TLR8), immune checkpoint inhibitors (programmed cell death 1 [PD-1] and programmed cell death ligand 1 [PD-L1] blockade), T-cell vaccines, and broadly neutralizing antibody approaches. While some immunomodulatory strategies have shown promise, such as one small phase 2 clinical trial in which treatment with the TLR8 agonist selgantolimod resulted in HBsAg loss in a small number, most immunomodulatory approaches have been abandoned for lack of efficacy or for toxicity (e.g., upregulation of RNA sensor retinoic acid-inducible gene I [RIG-I]). Even after more than a decade of early clinical trials, none of these treatments has yet been shown to "cure" hepatitis B, and none is likely to be competitive, unless it can be shown to go beyond current antivirals in achieving recovery (sustained HBsAg loss) from HBV infection. Whether combinations of immune modulators and direct antiviral agents will improve prospects for curative therapies remains to be seen.

TREATMENT RECOMMENDATIONS

Several learned societies and groups of expert physicians have issued treatment recommendations for patients with chronic hepatitis B; the most authoritative and updated are those of the AASLD and the EASL. Although the recommendations differ slightly, a consensus has emerged on most of the important points **(Table 352-4)**. No treatment is recommended or available for inactive "nonreplicative" hepatitis B carriers (*undetectable HBeAg* with normal ALT and HBV DNA $\leq 10^3$ IU/mL documented serially over time). Nor is treatment recommended currently for *HBeAg-positive* chronic hepatitis B with normal ALT (labeled by some as reflecting a relatively "immune-tolerant" phase but more accurately a period of dissociation between high-level HBV replication and a paradoxical paucity of inflammatory liver injury; see **Chap. 350**, "Complications and Sequelae"); however, such patients merit careful monitoring with HBV DNA and ALT at least every 6 months. Conversely, treatment is universally recommended for any patient with compensated cirrhosis and detectable HBV DNA; in decompensated cirrhosis, treatment should be carefully considered. Treatment considerations otherwise are based on the presence or absence of HBeAg, HBV DNA levels and presence or absence of active necroinflammatory activity (as based on ALT or liver biopsy).

HBeAg Positive In patients with detectable HBeAg and HBV DNA levels $>2 \times 10^4$ IU/mL, treatment is recommended by the AASLD for those with ALT levels >2× the upper limit of normal. (The EASL recommends treatment in HBeAg-positive patients for HBV DNA levels $>2 \times 10^3$ IU/mL and ALT above the upper limit of normal.) For HBeAg-positive patients with ALT ≤2× the upper limit of normal, in whom sustained responses are not likely

TABLE 352-4 Recommendations for Treatment of Chronic Hepatitis B[a]

HBeAg STATUS	CLINICAL	HBV DNA (IU/mL)	ALT	RECOMMENDATION
HBeAg-reactive	[b]	$>2 \times 10^4$	≤2 × ULN[c]	No treatment; monitor, *except* in patients >40, with family history of cirrhosis or hepatocellular carcinoma, with extrahepatic manifestations, with a history of previous treatment, and/or with liver biopsy (or noninvasive fibrosis determination) evidence for moderate to severe inflammation or fibrosis
	Chronic hepatitis	$>2 \times 10^4$[d]	>2 × ULN[d]	Treat[d]
	Cirrhosis:			
	Cirrhosis compensated	$>2 \times 10^3$	< or > ULN	Treat[d] with oral agents, not PEG IFN
		$<2 \times 10^3$	>ULN	Treat[d] with oral agents, not PEG IFN; refer for liver transplantation
	Cirrhosis decompensated	Detectable	< or > ULN	Observe; refer for liver transplantation
		Undetectable	< or > ULN	
HBeAg-negative	[b]	$\leq 2 \times 10^3$	≤ULN	Inactive carrier; treatment not necessary
	Chronic hepatitis	$>2 \times 10^3$	1 to >2 × ULN[d]	No treatment; monitor, except in patients >40, with family history of cirrhosis or hepatocellular carcinoma, with extrahepatic manifestations, with a history of previous treatment, and/or with liver biopsy (or noninvasive fibrosis determination) evidence for moderate to severe inflammation or fibrosis
	Chronic hepatitis	$>2 \times 10^3$	>2 × ULN[d]	Treat[e,f]
	Cirrhosis:			
	Cirrhosis compensated	$>2 \times 10^3$	< or > ULN	Treat[d] with oral agents, not PEG IFN
		$<2 \times 10^3$	>ULN	Treatment suggested[f]
	Cirrhosis decompensated	Detectable	< or > ULN	Treat[d] with oral agents[g], not PEG IFN; refer for liver transplantation
		Undetectable	< or > ULN	Observe; refer for liver transplantation

[a]Based on practice guidelines of the American Association for the Study of Liver Diseases (AASLD). Except as indicated in footnotes, these guidelines are similar to those issued by the European Association for the Study of the Liver (EASL). [b]Liver disease tends to be mild or inactive clinically; most such patients do not undergo liver biopsy. [c]According to the EASL guidelines, treat if HBV DNA is $>2 \times 10^3$ IU/mL and ALT >ULN. [d]One of the potent oral drugs with a high barrier to resistance (entecavir or tenofovir) or PEG IFN can be used as first-line therapy (see text for further detail). [e]Because HBeAg seroconversion is not an option, the goal of therapy is to suppress HBV DNA and maintain a normal ALT. PEG IFN is administered by subcutaneous injection weekly for a year; caution is warranted in relying on a 6-month posttreatment interval to define a sustained response, because the majority of such responses are lost thereafter. Oral agents, entecavir or tenofovir, are administered daily, usually indefinitely or until, as very rarely occurs, virologic and biochemical responses are accompanied by HBsAg seroconversion. [f]For older patients and those with advanced fibrosis, consider lowering the HBV DNA threshold to $>2 \times 10^3$ IU/mL.

Abbreviations: ALT, alanine aminotransferase; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; PEG IFN, pegylated interferon; ULN, upper limit of normal.

and who would require multiyear therapy, antiviral therapy is not recommended currently. As noted above, these untreated patients (as well as patients with HBV DNA between 2×10^3 and 2×10^4 IU/mL) merit continued monitoring with HBV DNA and ALT every 3–6 months. Even in these groups (in the "gray zone"), however, therapy would be considered for patients >40 years of age, with extrahepatic manifestations of HBV infection, or with a family history of cirrhosis or HCC; patients whose liver biopsy or noninvasive testing shows moderate to severe necroinflammatory activity or fibrosis; or patients with a history of previous treatment. In HBeAg-positive patients, treatment may be discontinued after HBeAg loss. In a 2016 systematic review of 1716 patients involved in 25 clinical trials, responses after oral-agent therapy were found to be durable. Among patients with HBeAg-reactive chronic hepatitis B, the pooled rates of durable HBeAg seroconversions maintained after cessation of nucleoside/nucleotide analogue therapy (including all the oral agents) were 92 and 88% at posttreatment months 12 and 24, respectively, unaffected by the duration of post-HBeAg-response consolidation therapy (>6 months in all studies evaluated); the pooled rate of durable biochemical remission after therapy in this population was 76%. Thus, per current AASLD recommendations, antiviral treatment with oral agents can be stopped after HBeAg loss/seroconversion in noncirrhotics, and the suggested period of consolidation therapy is 12 months with close monitoring for recurrent viremia (monthly × 6, then every 3 months for the rest of a year) after cessation of therapy.

HBeAg Negative For patients with HBeAg-negative chronic hepatitis B, antiviral therapy is recommended if ALT is >2× the upper limit of normal (above the upper limit of normal according to EASL) and HBV DNA is $>2 \times 10^3$ IU/mL. If HBV DNA is $>2 \times 10^3$ IU/mL and ALT is 1 to >2× the upper limit of normal, the same considerations apply as for HBeAg-positive patients with borderline ALT levels (the "gray zone") for patients >40 years of age, with extrahepatic manifestations of HBV infection, or with a family history of cirrhosis or HCC; patients whose liver biopsy or noninvasive testing shows moderate to severe necroinflammatory activity or fibrosis; or patients with a history of previous treatment. This group also merits continued careful monitoring with HBV DNA and ALT every 3–6 months. For patients with HBeAg-negative chronic hepatitis, the current recommendation with oral agents is for indefinite therapy; stopping therapy in this group can be considered after HBsAg loss, which remains exceedingly rare.

The potential for stopping antiviral therapy in noncirrhotic HBeAg-negative patients after protracted (≥2–5 years) antiviral therapy has been the subject of several studies After such prolonged courses of entecavir or tenofovir, in one small, 57-patient study (DARING-B), 18-month virologic relapse rates (HBV DNA >2000 IU/mL) were high (in 72%); however, 25% of study subjects underwent HBsAg loss after 18 months of follow-up. In a similar study (FINITE), virologic relapse rates were high, but 62% did not meet criteria for retreatment, and 19% lost HBsAg. In contrast, in a study among Asian patients, only ~30% had sustained responses for which resumption of therapy was not introduced, and HBsAg responses were negligible. In the only randomized, controlled trial of stopping therapy versus continuing therapy in HBeAg-negative patients after prolonged antiviral therapy (Toronto STOP study), only 33% had sustained responses after cessation of therapy, and HBsAg loss occurred with equal, small frequencies in both the stop-treatment group (4%) and the continue-treatment group (5%). Most recently, in a large 1552-subject, international cohort study (RETRACT B), treatment cessation in HBeAg-negative patients without cirrhosis was followed by virologic relapse rates of 68.9% at 12 months and 83.4% at 48 months. Relapse requiring retreatment occurred in 29.8% of subjects at 12 months and 54.7% at 48 months. Similar to prior studies, HBsAg loss was observed in 3.2% of subjects at 12 months and increased to 13% at 48 months. Predictors of HBsAg loss were identified, which may offer guidance to determine which patients may benefit from stopping nucleos(t)ide analogue therapy. In this study, the most important predictor of HBsAg loss was quantitative HBsAg level at the time therapy was discontinued (at posttreatment year 4, in Caucasian patients with HBsAg <1000 IU/mL, 41% had lost HBsAg, and in Asian patients with HBsAg <100 IU/mL, 33% had lost HBsAg). At present, however, quantitative HBsAg determination is not routinely available in clinical care. Generally, then, although HBsAg loss can be achieved in a small fraction, and although a subgroup may not require reintroduction of therapy in the short run, enthusiasm for this approach is limited, and for most HBeAg-negative patients, recommendations support indefinite treatment, unless they experience HBsAg loss.

For patients with compensated cirrhosis, because antiviral therapy has been shown to retard clinical progression, treatment is recommended regardless of HBeAg status and ALT as long as HBV DNA is detectable. In addition, stopping therapy in patients with cirrhosis comes with the risk of decompensation. For patients with decompensated cirrhosis, treatment is recommended regardless of serologic and biochemical status, as long as HBV DNA is detectable. In addition, patients with decompensated cirrhosis should be evaluated as candidates for liver transplantation. Cirrhotics should be treated indefinitely (see considerations for stopping antiviral therapy in noncirrhotics, above).

Choice of Treatment Agent Among the eight available drugs for hepatitis B, PEG IFN has supplanted standard IFN, entecavir has supplanted lamivudine, and tenofovir has supplanted adefovir **(Table 352-5)**. PEG IFN, entecavir, or tenofovir (TDF or TAF) is recommended as first-line therapy (Table 352-3). PEG IFN requires finite-duration therapy, achieves the highest rate of HBeAg

TABLE 352-5 Pegylated Interferon Versus Oral Nucleoside Analogues for the Treatment of Chronic Hepatitis B

	PEG IFN	NUCLEOSIDE ANALOGUES
Administration	Weekly injection	Daily, orally
Tolerability	Poorly tolerated, intensive monitoring	Well tolerated, limited monitoring
Duration of therapy	Finite, 48 weeks	≥1 year, indefinite in most patients
Maximum mean HBV DNA suppression	4.5 $\log_{10}$	6.9 $\log_{10}$
Effective in high-level HBV DNA (≥10^9 IU/mL)	No	Yes
HBeAg seroconversion		
During 1 year of therapy	~30%	~20%
During >1 year of therapy	Not applicable	30% (year 2) to up to 50% (year 5)
HBeAg-negative posttreatment HBV DNA suppression	17% at 5 years	7% at 4 years (lamivudine)
HBsAg loss		
During 1 year of therapy	3–4%	0–3%
During >1 year of therapy	Not applicable	3–8% at 5 years of therapy
After 1 year of therapy–HBeAg-negative	12% at 5 years	3.5% at 5 years
Antiviral resistance	None	Lamivudine: ~30% year 1, ~70% year 5 Adefovir: 0% year 1, ~30% year 5 Telbivudine: up to 4% year 1, 22% year 2 Entecavir: ≤1.2% through year 6 Tenofovir: 0% through year 8
Use in cirrhosis, transplantation, immunosuppressed	No	Yes
Cost, 1 year of therapy	++++	+ to ++

Abbreviations: HBV, hepatitis B virus; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; PEG IFN, pegylated interferon.

responses after a year of therapy, and does not support viral mutations, but it requires subcutaneous injections and is associated with inconvenience, more intensive clinical and laboratory monitoring, and intolerability. Oral agents do not require injections or cumbersome laboratory monitoring, are very well tolerated, lead to improved histology in 50–90% of patients, suppress HBV DNA more profoundly than PEG IFN, and are effective even in patients who fail to respond to IFN-based therapy. Although oral agents are less likely to result in HBeAg responses during the first year of therapy, as compared to PEG IFN, treatment with oral agents is usually extended beyond the first year and, by the end of the second year (of the current-generation, potent, high-barrier-to-resistance agents entecavir and tenofovir), yields HBeAg responses (and even HBsAg responses) comparable in frequency to those achieved after 1 year of PEG IFN (and without the associated side effects). Choice of agent should be based on clinical features, any history of prior treatment, and patient and physician preference; availability, relative cost, and insurance coverage often also play a role. As noted above, some physicians prefer to begin with PEG IFN, while most physicians and patients prefer oral agents as first-line therapy. Between the two first-line oral agents, entecavir and tenofovir, data to support the choice of one compared to the other are limited. Several observational studies have suggested that TDF is superior to entecavir in reducing the risk of HCC. Such studies, however, sophisticated statistical analyses notwithstanding, are subject to confounding influences that could favor TDF (e.g., physician preference for TDF in milder disease and for entecavir in more severe disease, including in older patients and those with impaired renal function). In addition, while several studies confirm a differential effect of TDF on long-term HCC risk, many others do not, and, in most cases, adjusting for such factors as age, sex, country, albumin, platelets, AFP, cirrhosis, and diabetes, eliminated the benefit of TDF over entecavir. Therefore, currently, the preponderance of data are insufficient to support this benefit of TDF over entecavir. Between TDF and TAF, TAF is preferred in patients with creatinine clearance <50 mL/min, reduced bone density, and risk factors for renal injury (e.g., decompensated cirrhosis, creatinine clearance <60 mL/min, poorly controlled hypertension or diabetes, proteinuria, active glomerulonephritis, concomitant nephrotoxic medications, or solid-organ transplantation). As noted above, the EASL recommendation for TAF extends to persons >60 years, who are at increased risk of TDF nephrotoxicity.

POPULATIONS WITH UNIQUE TREATMENT CONSIDERATIONS

Pregnancy Substantial experience with lamivudine during pregnancy (see above) has identified no teratogenicity; although widely used during pregnancy, lamivudine remains classified as pregnancy category C. Although IFNs do not appear to cause congenital anomalies, these have antiproliferative properties and should be avoided during pregnancy. Adefovir during pregnancy has not been associated with birth defects; however, the risk of spontaneous abortion may be increased, and adefovir is categorized as pregnancy category C. Data on the safety of entecavir during pregnancy have not been published (pregnancy category C). Sufficient data in animals and limited data in humans suggest that telbivudine and tenofovir (both pregnancy category B) can be used safely during pregnancy; however, telbivudine is not an acceptable first-line drug. In general, then, except for lamivudine and tenofovir, and until additional data become available, the other antivirals for hepatitis B should be avoided or used with extreme caution during pregnancy. Tenofovir is the current drug of choice in pregnancy.

Children For children aged 2 to <18 years old with HBeAg-positive hepatitis B (most children will be HBeAg-positive; children with HBeAg-negative chronic hepatitis B have not been well represented in treatment studies), treatment is recommended if HBV DNA is detectable and ALT levels are elevated, but not if ALT levels are normal. Each of the available first-line drugs is approved for different childhood age groups (PEG IFN α2a age ≥5 years [approved for hepatitis C, not B, but can be used in hepatitis B]; entecavir and tenofovir age ≥2 years). Package inserts should be consulted for childhood doses.

HIV Co-infection Patients with HBV-HIV co-infection can have progressive HBV-associated liver disease and, occasionally, a severe exacerbation of hepatitis B resulting from immunologic reconstitution following antiretroviral therapy. Tenofovir and the combination of tenofovir and emtricitabine in one pill are approved therapies for HIV and represent excellent choices for treating HBV infection in HBV-HIV co-infected patients. Generally, even for HBV-HIV co-infected patients who do not yet meet treatment criteria for HIV infection, treating for both HBV and HIV is recommended. In HIV-HBV co-infection, TAF, because of its better safety profile, is preferable to TDF.

Cirrhosis For patients with compensated or decompensated cirrhosis, PEG IFN should not be used; the emergence of such life-threatening side effects as sepsis can result in further deterioration and loss of antiviral effectiveness. Therefore, in this patient subset, therapy with a very favorable resistance profile (e.g., entecavir or tenofovir) is the preferred treatment. For patients with end-stage chronic hepatitis B who undergo liver transplantation, reinfection of the new liver is almost universal in the absence of antiviral therapy. The majority of patients become high-level viremic carriers with minimal liver injury. Before the availability of antiviral therapy, an unpredictable proportion experienced severe hepatitis B-related liver injury, sometimes a fulminant-like hepatitis and sometimes a rapid recapitulation of the original severe chronic hepatitis B **(Chap. 350)**. Currently, however, prevention of recurrent hepatitis B after liver transplantation has been achieved definitively by *combining* short-term (5–7 days) intravenous hepatitis B immune globulin (HBIG) with lifelong low-resistance oral entecavir or TDF or TAF **(Chap. 356)**; in some patients, especially those with a low risk for recurrence, the newer, more potent, and less resistance-prone oral agents may be used instead of HBIG for posttransplantation therapy. For patients at high risk for recurrence and progressive disease (e.g., patients with HDV-HBV or HIV-HBV co-infections) as well as for nonadherent patients, lifelong combination HBIG-oral agent therapy should be considered. For patients receiving livers from anti-HBc-positive donors, lifelong oral-agent therapy is recommended (without HBIG).

Immunosuppression Patients with chronic hepatitis B who undergo cytotoxic chemotherapy for treatment of malignancies as well as patients treated with immunosuppressive, anticytokine, or anti–tumor necrosis factor (TNF) therapies experience enhanced HBV replication and viral expression on hepatocyte membranes during chemotherapy coupled with suppression of cellular immunity. When chemotherapy is withdrawn, such patients are at risk for reactivation of hepatitis B, often severe and occasionally fatal. Such rebound reactivation represents restoration of cytolytic T-cell function against a target organ enriched in HBV expression. Preemptive treatment with the first of the oral HBV antivirals, lamivudine, prior to the initiation of chemotherapy was shown to reduce the risk of such reactivation substantially; treating *after* reactivation has occurred is less effective. The newer, more potent oral antiviral agents, entecavir and tenofovir, which are even more effective in preventing hepatitis B reactivation and with a lower risk of antiviral drug resistance, are preferred.

The risk of HBV reactivation and liver injury varies from highest (e.g., B cell–depleting agents, anthracycline derivatives, moderate-/high-dose corticosteroids for ≥4 weeks) to moderate (e.g., TNF-α inhibitors, cytokine or integrin inhibitors, tyrosine kinase inhibitors, low-dose corticosteroids for ≥4 weeks) to lowest (e.g., immunosuppressive agents like methotrexate and azathioprine, intraarticular corticosteroids, any dose of corticosteroids for ≤1 week). Most authorities (e.g., Centers for Disease Control and Prevention, AASLD, American Gastroenterological Association, EASL) recommend HBsAg and anti-HBc (± anti-HBs) screening of all patients prior to initiation of chemotherapy or other long-term immunosuppressive regimens. Preemptive antiviral prophylaxis is recommended for any patient with detectable HBsAg, given the high risk of reactivation (which can occur in ~25% of cases but potentially in up to 68%). While, generally, reactivation in HBsAg-negative but anti-HBc-reactive patients is rare, in this population, careful monitoring is recommended, except in the setting of B cell–depleting therapies, which are associated with a high risk for reactivation. Thus, antiviral prophylaxis is recommended in HBsAg-negative/anti-HBc-positive patients undergoing B-cell depletion therapy or stem cell transplantation.

The optimal duration of antiviral therapy after completion of chemotherapy is not known, but a suggested approach is 6 months (12 months for B cell–depleting agents) for inactive hepatitis B carriers and longer-duration therapy in patients with baseline HBV DNA levels >2 × 10^3 IU/mL, until standard clinical endpoints are met (Table 352-4).

■ CHRONIC HEPATITIS D (DELTA HEPATITIS)

Chronic hepatitis D virus (HDV) infection is most commonly observed in patients who were exposed to HDV while already harboring chronic HBV infection (so-called superinfection). Chronic HDV infection may also follow acute co-infection with HBV and HDV but at a rate no higher than the rate of chronicity of acute hepatitis B. That is, although HDV co-infection can increase the severity of acute hepatitis B, HDV does not increase the likelihood of progression to chronic hepatitis B. When, however, HDV superinfection occurs in a person who is already chronically infected with HBV, long-term HDV infection is the rule, and a worsening of the liver disease is the expected consequence. Except for severity, chronic hepatitis B plus D has similar clinical and laboratory features to those seen in chronic hepatitis B alone. Relatively severe and progressive chronic hepatitis, with or without cirrhosis, is the rule, and mild chronic hepatitis is the exception. Occasionally, however, mild hepatitis or even, rarely, inactive carriage occurs in patients with chronic hepatitis B plus D, and the disease may become indolent after several years of infection. **The clinical and laboratory features of chronic HDV infection are summarized in Chap. 350.**

TREATMENT

Chronic Hepatitis D

Management is not well defined, and the host cellular RNA polymerase upon which HDV replication depends cannot be targeted by conventional antiviral agents. Treatment with type I interferons is the only approved treatment for chronic HDV infection, although, per expert opinion, treating underlying HBV infection is suggested as well if the HBV DNA level is >2 × 10^3 IU/mL. Preliminary experimental trials of IFN-α suggested that conventional doses and durations of therapy lower levels of HDV RNA and aminotransferase activity only transiently during treatment but have no impact on the natural history of the disease. In contrast, *high-dose* IFN-α (9 million units three times a week) for 12 months was reported to be associated with a sustained loss of HDV replication and clinical improvement in up to 50% of patients. Furthermore, a substantial improvement in 12-year survival after liver transplantation was observed in patients with hepatitis D who received high-dose IFN-α (86 vs 31% with standard-dose IFN-α). A suggested approach to therapy has been high-dose, long-term IFN for at least a year and, in responders, extension of therapy until HDV RNA and HBsAg clearance; however, extension of therapy to a second year provided no advantage, and sustained responses after completion of therapy have been rare. Compared to standard IFN-α, PEG IFN has been shown to be more effective in chronic hepatitis D but still of limited therapeutic value; after 48 weeks of therapy, durable undetectable HDV RNA for 24 posttreatment weeks has been reported in a quarter to just over a half of patients. Disappointingly, loss of virologic responses (reappearance of HDV RNA) was observed during long-term (median 4.5 years) monitoring in a majority of 24-week-posttreatment responders, with durable HDV RNA suppression to undetectable in only 12%. None of the nucleoside analogue antiviral agents for hepatitis B are effective in hepatitis D, and adding ribavirin or oral nucleoside agents (lamivudine and, more recently, adefovir and tenofovir) to PEG IFN was not found to be more effective than PEG IFN monotherapy. While recommended, 12 months of PEG IFN therapy is far from satisfactory. The lackluster response to type I interferon treatment with or without nucleoside agents highlights the need for novel therapies to target HDV infection.

The approach farthest along in clinical testing is the entry inhibitor bulevirtide, a lipopeptide that blocks the NTCP receptor, preventing hepatocyte entry of both HBV and HDV. In phase 2 studies of subcutaneous bulevirtide alone or combined with PEG IFN for varying treatment durations, a significant reduction occurred in HDV RNA in >71% of subjects, with a more pronounced effect in combination therapy. Similarly, a phase 2 study of bulevirtide combined with tenofovir (vs tenofovir monotherapy) for 24 weeks demonstrated significant reductions in HDV RNA at the end of treatment (50–77% vs 4% with tenofovir alone), but rebound of viremia was universal at follow-up week 48. Most recently, in an ongoing phase 3 clinical trial, ALT normalized in 51–56% of patients receiving bulevirtide after 48 weeks of treatment, and 45–48% of patients had a ≥2 $\log_{10}$ reduction in or undetectable HDV RNA; this treatment effect was maintained though 96 weeks of treatment. In 2020, bulevirtide was authorized conditionally by the European Medicines Agency for the treatment of HDV with or without PEG IFN or nucleoside analogues. Review by the U.S. Food and Drug Administration (FDA) is anticipated.

Preliminary trials have also been performed with lonafarnib, a prenylation inhibitor, which is combined with ritonavir to improve bioavailability. Prenylation, the posttranslational covalent addition of the prenyl lipid farnesyl to large HDV antigen, is required for this HDV protein to interact and form secreted viral particles with HBsAg. Phase 2 studies of lonafarnib/ritonavir have demonstrated significant, on-treatment reductions in HDV RNA, and in one study, the addition of IFNα increased the likelihood of achieving a viral response to 89%. Additional experimental approaches to the treatment of hepatitis D include nucleic acid polymer therapy, often combined with PEG IFNα. Nucleic acid polymer therapy is thought to inhibit HBsAg release and has been administered alone or with PEG IFN; to date, these studies have been done in one Eastern European site. While promising reductions in HDV RNA and HBsAg have been reported, these studies have been plagued by adverse effects, and confirmatory studies have not materialized.

In patients with end-stage liver disease secondary to chronic hepatitis D, liver transplantation has been effective. If hepatitis D recurs in the new liver without the expression of hepatitis B (an unusual serologic profile in immunocompetent persons but common in transplant patients), liver injury is limited. In fact, the outcome of transplantation for chronic hepatitis D is superior to that for chronic hepatitis B; in such patients, combination HBIG and nucleoside analogue therapy for hepatitis B is indicated **(Chap. 356)**.

■ CHRONIC HEPATITIS C

Regardless of the epidemiologic mode of acquisition of hepatitis C virus (HCV) infection, chronic hepatitis follows acute hepatitis C in 50–70% of cases; chronic infection is common even in those with a return to normal in aminotransferase levels after acute hepatitis C, adding up to an 85% likelihood of chronic HCV infection after acute hepatitis C. Few clues had emerged to explain host differences associated with chronic infection until recently, when variation in a single nucleotide polymorphism (SNP) on chromosome 19, *IL28B* (which codes for IFN-λ3, now renamed *IFNL3*), was identified that distinguished between responders and nonresponders to IFN-based antiviral therapy (see below). The same variants correlated with spontaneous resolution after acute infection: 53% in genotype C/C, 30% in genotype C/T, but only 23% in genotype T/T. The association with HCV clearance after acute infection is even stronger when *IL28B* (*IFNL3*) haplotype is combined with haplotype G/G of an SNP near human leukocyte antigen (HLA) class II *DBQ1*03:01*.

In patients with chronic hepatitis C followed for 20 years, progression to cirrhosis occurs in ~20–25%. Such is the case even for patients with relatively clinically mild chronic hepatitis, including those without symptoms, with only modest elevations of aminotransferase activity, and with mild chronic hepatitis on liver biopsy. Even in cohorts of well-compensated patients with chronic hepatitis C referred for clinical research trials (no complications of chronic liver disease and with normal hepatic synthetic function), the prevalence of cirrhosis may be as high as 50%. Most cases of hepatitis C are identified initially in asymptomatic patients who have no history of acute hepatitis C (e.g., those discovered while attempting to donate blood, while undergoing lab testing as part of an application for life insurance, or as a result of routine laboratory tests). The source of HCV infection in many of these cases is not defined, although a long-forgotten percutaneous exposure (e.g., injection drug use) in the remote past can be elicited in a substantial proportion and probably accounts for most infections; most of these infections were acquired in the 1960s and 1970s among persons in the 1945–1965 birth cohort **(Chap. 350)**, coming to clinical attention decades later.

Approximately one-third of patients with chronic hepatitis C have normal or near-normal aminotransferase activity; although one-third to one-half of these patients have chronic hepatitis on liver biopsy, the grade of liver injury and stage of fibrosis tend to be mild in the vast majority. In some cases, more severe liver injury has been reported—even, rarely, cirrhosis, most likely the result of previous histologic activity. Among patients with persistent normal aminotransferase activity sustained over ≥5–10 years, histologic progression has been shown to be rare; however, approximately one-fourth of patients with normal aminotransferase activity experience subsequent aminotransferase elevations, and histologic injury can be progressive once abnormal biochemical activity resumes. Therefore, continued clinical monitoring and antiviral therapy are indicated, even for patients with normal aminotransferase activity.

Despite this substantial rate of progression of chronic hepatitis C and even though liver failure can result from end-stage chronic hepatitis C, the long-term prognosis over 1–2 decades for chronic hepatitis C in most patients is relatively benign. Mortality over 10–20 years among patients with transfusion-associated chronic hepatitis C has been shown not to differ from mortality in a matched population of transfused patients in whom hepatitis C did not develop. Although death in the hepatitis group is more likely to result from liver failure and although hepatic decompensation may occur in ~15% of such patients over the course of a decade, the majority (almost 60%) of patients remain asymptomatic and well compensated, with no clinical sequelae of chronic liver disease. Overall, chronic hepatitis C tends to be very slowly and insidiously progressive, if at all, in most patients, whereas in approximately one-fourth of cases, chronic hepatitis C will progress eventually to end-stage cirrhosis. In fact, because HCV infection is so prevalent, and because a proportion of patients progress inexorably to end-stage liver disease, hepatitis C was the most frequent indication for liver transplantation **(Chap. 356)** in the era prior to the availability of direct-acting antiviral (DAA) therapy (see below). In the United States, hepatitis C accounts for up to 40% of all chronic liver disease; as of 2007, mortality caused by hepatitis C surpassed that associated with HIV/AIDS, and as of 2012, reported deaths caused by hepatitis C surpassed those associated with all other notifiable infectious diseases (HIV, tuberculosis, hepatitis B, and 57 other infectious diseases). Moreover, because the prevalence of HCV infection is so much higher in the "baby boomer" cohort born between 1945 and 1965, three-quarters of the mortality associated with hepatitis C occurs in this age cohort. In a 2010 modeling study based on the prevalence of HCV infection and its known natural history, liver disease mortality associated with hepatitis C was *projected* to peak in the year 2022; however, the *observed* HCV-related mortality in the United States peaked in 2013 and has fallen since, paralleling the introduction in 2013 of highly effective DAA agents (see below). The declining death rates related to HCV infection following the wide use of DAAs have been observed across most of the United States.

Progression of liver disease in patients with chronic hepatitis C has been reported to be more likely in patients with older age, longer duration of infection, advanced histologic stage and grade, more complex HCV quasispecies diversity, increased hepatic iron, concomitant other liver disorders (alcoholic liver disease, chronic hepatitis B, hemochromatosis, α_1 antitrypsin deficiency, and steatohepatitis), HIV infection, and obesity. Among these variables, however, duration of infection appears to be one of the most important, and some of the others probably reflect disease duration to some extent (e.g., quasispecies diversity, hepatic iron accumulation). No other epidemiologic or clinical features of chronic hepatitis C (e.g., severity of acute hepatitis, level of aminotransferase activity, level of HCV RNA, presence or absence of jaundice during acute hepatitis) are predictive of eventual outcome. Despite the relatively benign nature of chronic hepatitis C over time in many patients, cirrhosis following chronic hepatitis C has been associated with the late development, after several decades, of HCC **(Chap. 87)**; the annual rate of HCC in cirrhotic patients with hepatitis C is 1–4%, occurring primarily in patients who have had HCV infection for 30 years or more.

Referral bias may account for the more severe outcomes described in cohorts of patients reported from tertiary care centers (20-year progression of ≥20%) versus the more benign outcomes in cohorts of patients monitored from initial blood-product-associated acute hepatitis or identified in community settings (20-year progression of only 4–7%). Still unexplained, however, are the wide ranges in reported progression to cirrhosis, from 2% over 17 years (eventually 19% over 36 years) in a population of Irish women with hepatitis C infection acquired from contaminated anti-D immune globulin to 30% over ≤11 years in recipients of contaminated intravenous immune globulin.

Perhaps the best prognostic indicator in chronic hepatitis C is liver histology; the rate of hepatic fibrosis may be slow, moderate, or rapid. Patients with mild necrosis and inflammation as well as those with limited fibrosis have an excellent prognosis and limited progression to cirrhosis. In contrast, among patients with moderate to severe necroinflammatory activity or fibrosis, including septal or bridging fibrosis, progression to cirrhosis is highly likely over the course of 10–20 years. Among patients with compensated cirrhosis associated with hepatitis C, the 10-year survival rate is close to 80%; mortality occurs at a rate of 2–6% per year; decompensation at a rate of 4–5% per year; and, as noted above, HCC at a rate of 1–4% per year. Estimates of the natural history of chronic hepatitis C have been made, based on data available on the prevalence of HCV infection in the U.S. population and on the rate of disease progression. Weighted primarily by the concentration of chronic hepatitis C in the baby boomer generation, the peak prevalence was estimated to have occurred in 2015. The calculated frequency of cirrhosis in U.S. patients with hepatitis C was 5% in 1990 and 25% in 2010 and was projected to be 37% in 2020. **A discussion of the pathogenesis of liver injury in patients with chronic hepatitis C appears in Chap. 350.**

Clinical features of chronic hepatitis C are similar to those described above for chronic hepatitis B. Generally, fatigue is the most common symptom; jaundice is rare. Immune complex–mediated extrahepatic complications of chronic hepatitis C are less common than in chronic hepatitis B (despite the fact that assays for immune complexes are often positive in patients with chronic hepatitis C), with the exception of essential mixed cryoglobulinemia **(Chap. 350)**, which is linked to cutaneous vasculitis and membranoproliferative glomerulonephritis as well as lymphoproliferative disorders such as B-cell lymphoma and unexplained monoclonal gammopathy. In addition, chronic hepatitis C has been associated with extrahepatic complications unrelated to immune-complex injury. These include Sjögren's syndrome, lichen planus, porphyria cutanea tarda, renal injury, type 2 diabetes mellitus, and the metabolic syndrome (including insulin resistance and steatohepatitis). In addition, a link has been observed between HCV infection and cardiovascular/cerebrovascular disease, rheumatologic/immunologic disorders, mental health and cognitive disorders, and nonliver malignancies.

Laboratory features of chronic hepatitis C are similar to those in patients with chronic hepatitis B, but aminotransferase levels tend to fluctuate more (the characteristic episodic pattern of aminotransferase activity) and to be lower, especially in patients with long-standing disease. An interesting and occasionally confusing finding in patients with chronic hepatitis C is the presence of autoantibodies. Rarely, patients with autoimmune hepatitis (see below) and hyperglobulinemia have false-positive immunoassays for anti-HCV. On the other hand, some patients with serologically confirmable chronic hepatitis C have circulating anti-LKM. These antibodies are anti-LKM1, as seen in patients with autoimmune hepatitis type 2 (see below), and are directed against a 33-amino-acid sequence of cytochrome P450 IID6.

Histopathologic features of chronic hepatitis C, especially those that distinguish hepatitis C from hepatitis B, are described in Chap. 350.

TREATMENT

Chronic Hepatitis C

Therapy for chronic hepatitis C has evolved substantially in the 30 years since IFN-α was introduced for this indication in 1991. The therapeutic armamentarium grew to include PEG IFN with ribavirin and, then, in 2011, the introduction of the first protease inhibitors, telaprevir and boceprevir, used in combination with PEG IFN and ribavirin in patients with HCV genotype 1. The field of antiviral therapy for hepatitis C was transformed beginning in 2013, with the approval of the first nucleoside analogue polymerase inhibitor sofosbuvir. Although several of these combination regimens have been supplanted by better, later-generation drugs, as of 2020, five all-oral, highly effective (>95%), low-resistance, pangenotypic, well-tolerated, short-duration (primarily 8–12 weeks) combination regimens of DAA drugs are recommended. The remarkable historical evolution of antiviral therapy for hepatitis C is instructive.

THE INTERFERON ERA (1991–2011)

IFN-based therapy has been supplanted by DAA agents introduced in the second decade of the twenty-first century; however, many important lessons about antiviral therapy for chronic hepatitis C were learned from the experience with IFN-based treatment, and many of the limitations of—and disparities in responsiveness to—IFN-based therapy have been overcome by current-generation DAA treatments. Mechanistically, HCV proteins inhibit several steps of the JAK-STAT signal transduction pathway, and by activation of JAK-STAT signaling, exogenous IFN culminates in and restores intracellular expression of IFN-stimulated genes and their protein products that have antiviral properties.

When first approved, subcutaneous IFN-α three times a week for 6 months achieved a sustained virologic response (SVR; defined then as a reduction of HCV RNA to PCR-undetectable levels ≥24 weeks after completion of therapy) in <10%. Doubling the duration of therapy increased the SVR rate to ~20%, and addition to the regimen of daily ribavirin (ineffective when used alone), an oral guanosine nucleoside, increased the SVR rate further to 40% by reducing the likelihood of virologic relapse after completion of treatment. Although its mechanism of action remains poorly understood, ribavirin retains a limited role in supporting DAA agents in several subgroups of otherwise refractory patients (see below).

Beginning in 2011, for the treatment of hepatitis C, standard IFNs were supplanted by PEG IFNs, which have substantially longer half-lives, permitting administration once (rather than three times) a week. Treatment with the combination of PEG IFN and ribavirin increased SVR rates to 55% overall—to >40% in genotypes 1 and 4, requiring 48 weeks of therapy, and to >80% in genotypes 2 and 3, requiring only 24 weeks of therapy—and resulted in histologic improvement in approximately three-fourths of patients. After initiation of IFN treatment, ALT levels fell precipitously, and up to 90% of virologic responses were achieved within the first 12 weeks of therapy. Failure to achieve an early virologic response (EVR), a ≥2 $\log_{10}$ reduction in HCV RNA by week 12, predicted failure to experience a subsequent SVR. Similarly, patients in whom HCV RNA became undetectable within 4 weeks (i.e., who achieve a rapid virologic response [RVR]) had a very high likelihood of achieving an SVR **(Fig. 352-2)**.

Most relapses occurred within the first 12 weeks after treatment, and absence of HCV RNA 12 weeks after completion of therapy has become the current standard for SVR (SVR_{12}); relapses are very rare 6 months to a year after SVR and almost unheard of after 2 years. Of documented durability decades after successful therapy, an SVR to antiviral therapy for chronic hepatitis C is tantamount to a cure and is followed by marked improvements in liver disease outcomes (see below).

Patient variables that correlated with IFN-based SVRs included favorable genotype (genotypes 2 and 3 as opposed to genotypes 1 and 4); low baseline HCV RNA level (<800,000 IU/mL); histologically mild hepatitis and minimal fibrosis, especially absence of cirrhosis; immunocompetence; age <40; female gender; and absence of comorbid metabolic disease. High levels of HCV RNA, more histologically advanced liver disease, and high HCV quasispecies diversity all went hand in hand with advanced duration of infection and reduced IFN responsiveness. Ironically, patients whose disease was least likely to progress were the ones *most* likely to respond to IFN and vice versa.

As described above in the discussion of spontaneous recovery from acute hepatitis C, IFN gene variants discovered in genome-wide association studies were shown to have a substantial impact on IFN responsiveness of patients with genotype 1 to antiviral therapy. In studies of patients treated with PEG IFN and ribavirin, variants

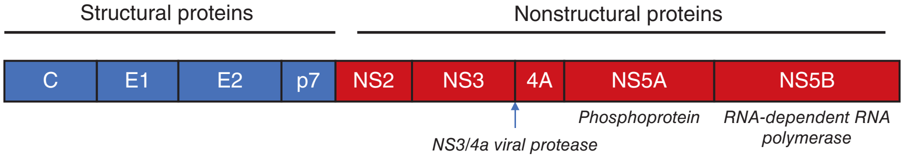

Target	NS3-4A protease inhibitors	NS5A inhibitors	NS5B inhibitors
Function	Block N3-4A cleavage	Inhibit NS5A protein, essential for replication	Block the RNA-dependent RNA polymerase
Suffix	-previr	-asvir	-buvir
Drugs	Glecaprevir, voxilaprevir	Velpatasvir, ledipasvir, pibrentasvir	Sofosbuvir

FIGURE 352-2 Hepatitis C virus (HCV) viral genome and drug targets. The HCV viral genome encodes three structural proteins (c-nucleocapsid and e1-2 envelope proteins) and seven nonstructural proteins. Current, first-line direct-acting antivirals have three viral targets: (1) the NS3/4A viral protease; (2) the NS5A protein, which is a critical component of the viral replication apparatus; and (3) NS5B RNA-dependent RNA polymerase.

of the *IL28B* (now renamed *IFNL3*) SNP that code for IFN-λ3 (a type III IFN, the receptors for which are more discretely distributed than IFN-α receptors and more concentrated in hepatocytes) correlated significantly with responsiveness. Homozygotes for the C allele at this locus (C/C) achieved SVRs of ~80%, heterozygotes (C/T) SVRs of ~35%, and homozygotes for the T allele (T/T) SVRs of ~25%.

Side effects of IFN therapy are described in the section on treatment of chronic hepatitis B (see above). Besides ribavirin-associated nasal and chest congestion, pruritus, and precipitation of gout, the most pronounced ribavirin side effect is hemolysis, often requiring dose reduction or addition of erythropoietin therapy (not shown, however, to increase the likelihood of an SVR); therefore, close monitoring of blood counts is crucial, and ribavirin should be avoided in patients with anemia, hemoglobinopathies, coronary artery disease or cerebrovascular disease, or renal insufficiency (the drug is excreted renally) and in pregnancy (the drug is teratogenic, mandating contraception during, and for several months after, therapy in women of child-bearing age [because of their antiproliferative properties, IFNs also are contraindicated during pregnancy]).

For most of the decade prior to 2011, when protease inhibitors were introduced for HCV genotype 1 (see below), the standard of care was a combination of PEG IFN plus ribavirin (unless ribavirin was contraindicated) for all HCV genotypes. Even after the introduction of protease inhibitors for genotypes 1 and 4, however, PEG IFN–ribavirin remained the standard of care for patients with genotypes 2 and 3 until late 2013. Responsiveness to IFN-ribavirin–based therapy was diminished in immunocompromised patients and in patients with HIV-HCV co-infection and contraindicated in patients with decompensated liver disease or end-stage renal disease. The cumbersome nature of IFN-ribavirin–based therapy (injections, complicated laboratory monitoring, side effects and poor tolerability, modest efficacy, variables and patient subsets associated with poor responsiveness, tailored therapy, futility rules, etc.) was supplanted eventually (in 2016) by DAAs for all genotypes (see below). Most of the variables associated with poor responsiveness to IFN-based therapy became irrelevant, and difficult-to-treat patient subpopulations began to experience responses to DAAs that were indistinguishable from responses in standard patients (see below).

Persons with chronic HCV infection suffer increased liver-related mortality, all-cause mortality, and multiple extrahepatic disorders (see above). On the other hand, successful antiviral therapy of chronic hepatitis C resulting in an SVR was shown to improve survival (and to reduce the need for liver transplantation); to lower the risk of liver failure, liver-related death, and all-cause mortality; to slow the progression of chronic hepatitis C; to reverse fibrosis and even cirrhosis; and to improve such HCV-associated extrahepatic disorders as type 2 diabetes and renal disease. Whereas the 10- and 20-year survival in the absence of an SVR is reduced in cirrhotic patients with chronic hepatitis C, survival at these intervals after an SVR has been found to be indistinguishable from that of the general population. In cirrhotic patients (and in those with advanced fibrosis), although successful treatment reduces mortality and liver failure (three- to fourfold 10-year reduction) and reduces the need for liver transplantation and the likelihood of HCC (14-fold 10-year reduction), the risk of liver-related death and HCC persists, albeit at a much reduced level, necessitating continued clinical monitoring and cancer surveillance after SVR in patients with advanced fibrosis and cirrhosis. On the other hand, in the absence of an SVR, IFN-based therapy does not reduce the risk of HCC. Fortunately, PEG IFN–ribavirin nonresponders can now be retreated with DAAs and experience SVR rates comparable to those in treatment-naïve persons (see below).

FIRST-GENERATION PROTEASE INHIBITORS (2011–2013)

The HCV RNA genome encodes a single polyprotein, which is cleaved during and after translation by host and viral-encoded proteases (Fig. 352-2). One protease involved in the cleavage of the viral polyprotein is an NS3/4A viral protein that has serine protease activity. Telaprevir and boceprevir were first-generation serine protease inhibitors that target the genotype 1 NS3/4A. In 2011, telaprevir and boceprevir used in combination with PEG IFN and ribavirin were approved by the FDA as the first oral DAA agents for the treatment of hepatitis C genotype 1 (not other genotypes) in adults with stable liver disease. Although now replaced by more effective, all-oral regimens, these first-in-class agents represented a breakthrough in the treatment of chronic hepatitis C and established milestones against which subsequent therapies could be measured.

Because selection of resistant variants developed rapidly during monotherapy with telaprevir and boceprevir, these drugs had to be used in combination with PEG IFN and ribavirin. Ribavirin, in particular, appeared to reduce relapse rates significantly in protease-inhibitor-based regimens, such that those who could not take or were intolerant to ribavirin were unlikely to benefit from the addition of these agents. Telaprevir and boceprevir regimens consisted of periods of triple therapy (protease inhibitor plus PEG IFN plus ribavirin) and periods of dual therapy (PEG IFN plus ribavirin). Cumbersome periods of "run-in" therapy or late-treatment consolidation therapy with PEG IFN-ribavirin, reliance on early virologic responses to determine duration of therapy ("response-guided therapy"), and futility rules (for stopping therapy based on poor early responsiveness) rendered first-generation DAA treatment regimens complicated and unwieldy.

For patients with HCV genotype 1, protease inhibitors improved the frequency of RVRs and SVRs significantly as compared to PEG IFN plus ribavirin alone. In treatment-naïve patients, telaprevir-based SVRs were achieved in up to 79% of patients who received 12 weeks of triple therapy followed by 12–36 weeks of dual therapy, with higher rates of response (83–92%) observed in those with undetectable RNA at weeks 4 and 12. In studies with boceprevir in treatment-naïve patients, SVRs occurred in 59–66% of patients, and among those with undetectable HCV RNA at 8 weeks, the SVR rate increased to 86–88%. Adding to the complexity of treatment with these protease inhibitors were absolute stopping rules for futility, that is, absence of HCV RNA reductions at critical treatment milestones, which were shown to be invariably predictive of nonresponse.

The first-generation protease inhibitors were successful as well among patients in whom prior PEG IFN plus ribavirin had failed, even in "null responders," who had never experienced even a 2 $\log_{10}$ reduction in HCV RNA while on treatment. In this hardest-to-treat group, telaprevir-based therapy achieved an SVR in 29–33% and boceprevir achieved an SVR in 30–40%.

In a substantial proportion of protease inhibitor nonresponders, resistance-associated substitutions (RASs; previously referred to as resistance-associated variants [RAVs]) could be identified, but these variants were not archived, and wild-type HCV reemerged in almost all cases within 1.5–2 years. SVRs to these protease inhibitors were highest in prior relapsers and treatment-naïve patients (with SVR rates of >75%) and lowest in cirrhotic prior null responders, for whom no benefit accrued over PEG IFN–ribavirin treatment.

Both protease inhibitors had substantial toxicities. Telaprevir was associated with a severe, generalized (trunk and extremities), often confluent, maculopapular, pruritic rash in ~6% of treated patients (that required careful dermatologic monitoring in all patients and systemic corticosteroid therapy in the most severely affected). Other common side effects included pruritus, rectal burning, nausea, diarrhea, fatigue, dysgeusia (altered or unpleasant taste), and anemia, which required close monitoring, could be relatively refractory, and occasionally required transfusion and even hospitalization (especially in cirrhotic prior nonresponders). Anemia occurred in half of boceprevir-treated patients. The other most common side effects of boceprevir were fatigue, nausea, headache, dysgeusia, dry mouth, vomiting, and diarrhea.

Both drugs came with an inconveniently high pill burden and had to be administered every 8 h with food (telaprevir with a 20-g fat meal). Use of protease inhibitors was further complicated by numerous drug-drug interactions. As telaprevir and boceprevir are both eliminated by and inhibit CYP3A4, these agents could not be administered with other medications that induce CYP3A4 or are dependent on CYP3A4 for elimination. Care had to be taken to examine for any potential interactions between these protease inhibitors and other medications the patient was taking, and a convenient website became available to check for such drug-drug interactions (*www.hep-druginteractions.org*). This website remains updated and of important clinical utility for checking drug-drug interactions for current HCV DAA regimens.

Despite the improvement in SVRs with protease-inhibitor-based regimens for genotype 1 compared to PEG IFN–ribavirin (e.g., in treatment-naïve patients 66–79% vs 38–44%), triple-drug protease inhibitor therapy was hampered by amplified intolerability, the complexity of response-guided regimens and futility stopping rules, the inconvenience of thrice-daily dosing with meals and a high pill burden, the need for PEG IFN injections and ribavirin with all their intolerability, and multiple drug-drug interactions. Moreover, side effects appeared to be more severe and burdensome once these drugs entered practice, especially in cirrhotic nonresponders, in whom studies reported from Europe showed serious adverse events in up to 45% and deaths in up to 3%. All these issues, as well as rapidly accelerating progress on next-generation and all-oral DAA therapy (see below), conspired to temper enthusiasm for these new antivirals; after a brief stint as recommended therapy (2011–2013), these drugs became obsolete and are no longer recommended or available.

DIRECT-ACTING ANTIVIRAL COMBINATIONS OF SECOND-GENERATION PROTEASE INHIBITORS, FIRST-GENERATION POLYMERASE INHIBITORS, AND FIRST-GENERATION NS5A INHIBITORS (2014–2015)

Since late 2013, the number of new antiviral agents for hepatitis C has expanded substantially, and currently, PEG IFN–based treatments have been supplanted by five remaining therapeutic regimens, which are all oral, IFN-free, highly efficacious (>95% SVR), and well tolerated, with high barriers to resistance, simple dosing, low pill burdens, treatment durations as brief as 8–12 weeks, and pangenotypic efficacy (**Table 352-6**). These drugs are distributed among three classes of DAAs: NS3/4 protease inhibitors (which cleave the single HCV polyprotein into constituent structural and nonstructural proteins [drug name ending in "-previr"]), NS5B nucleoside and nonnucleoside polymerase inhibitors (which interfere with the RNA-dependent RNA polymerase [a replicase] involved in synthesis of viral RNA [drug name ending in "-buvir"]), and NS5A inhibitors (which interfere

TABLE 352-6 Indications and Recommendations for Antiviral Therapy of Chronic Hepatitis C[a]

Standard Indications for Therapy	Treatment-Naïve	Previously Treated with DAA
All patients with chronic HCV infection (detectable HCV RNA, with or without elevated ALT) except for those with short life expectancies owing to comorbid conditions. Any stage of fibrosis; pretreatment biopsy is no longer embraced and has been supplanted by noninvasive measures of fibrosis, e.g., imaging to determine liver elasticity. **Retreatment Recommended** Relapsers, partial responders, or nonresponders after a previous course of interferon-based therapy or prior direct-acting antiviral therapy **Antiviral Therapy Not Recommended** Pregnancy: No sizable clinical studies of direct-acting antivirals during pregnancy are available. Ribavirin is contraindicated during pregnancy; therefore, any regimen including ribavirin should not be used. Sofosbuvir and sofosbuvir plus ledipasvir are classified as pregnancy category B, but the other direct-acting antivirals do not have a pregnancy classification. Therefore, these therapies are not indicated routinely in pregnancy and should be used, with caution, only if the benefit of treatment outweighs the potential for fetal risk. **Therapeutic Regimens (based on AASLD-IDSA recommendations, *www.hcvguidelines.org*)[b]**	**Simplified treatment regimen[b]** *All genotypes* Sofosbuvir-velpatasvir × 12 weeks Glecaprevir-pibrentasvir × 8 weeks	**Sofosbuvir failure** *All genotypes* Sofosbuvir-velpatasvir-voxilaprevir × 12 weeks *Genotypes 1, 2, 4, 5, or 6 (alternative)* Glecaprevir-pibrentasvir × 16 weeks[e]
	Cirrhosis – compensated[c] *Genotypes 1–6* Glecaprevir-pibrentasvir × 8 weeks *Genotypes 1, 2, 4, 5, or 6* Sofosbuvir-velpatasvir × 12 weeks	**Glecaprevir-pibrentasvir failure** *All genotypes* Glecaprevir-pibrentasvir plus sofosbuvir and weight-based ribavirin × 16 weeks Sofosbuvir-velpatasvir-voxilaprevir × 12 weeks[f]
	Cirrhosis – decompensated Ribavirin-eligible *Genotypes 1–6* Sofosbuvir-velpatasvir with weight-based ribavirin[d] × 12 weeks *Genotypes 1,4, 5, or 6* Ledipasvir-sofosbuvir with ribavirin[e] × 12 weeks Ribavirin-ineligible *Genotypes 1–6* Sofosbuvir-velpatasvir × 24 weeks *Genotypes 1, 4, 5, or 6* Ledipasvir-sofosbuvir × 24 weeks	**Multiple direct-acting antiviral failures** *All genotypes* Glecaprevir-pibrentasvir plus sofosbuvir and weight-based ribavirin × 16 weeks[g] Sofosbuvir-velpatasvir-voxilaprevir plus weight-based ribavirin × 24 weeks **Cirrhosis – decompensated** *All genotypes* Sofosbuvir-velpatasvir plus weight-based ribavirin × 24 weeks *Genotypes 1, 4, 5, or 6* Ledipasvir-sofosbuvir plus ribavirin[e] × 24 weeks
FEATURES ASSOCIATED WITH REDUCED RESPONSIVENESS TO DIRECT-ACTING ANTIVIRAL COMBINATION THERAPY Genotype and subtype (genotype 1a less responsive than genotype 1b for several drugs) Treatment experience Advanced fibrosis (bridging fibrosis, cirrhosis) Reduced adherence		**DECOMPENSATED CIRRHOSIS** **Any protease-inhibitor-containing regimen is CONTRAINDICATED in decompensated cirrhosis (glecaprevir, grazoprevir, voxilaprevir).**

[a]Rapidly evolving new recommendations continue to be issued; for up-to-date treatment recommendations, please see *www.hcvguidelines.org.* [b]Excludes: cirrhosis, prior treatment, pregnancy, hepatitis B surface antigen positive, hepatocellular carcinoma (HCC), or liver transplant. No adjustment required for end-stage renal disease. [c]Child-Pugh class A cirrhosis only; excludes: prior treatment, pregnancy, hepatitis B surface antigen positive, end-stage renal disease, HCC, or liver transplant. After treatment, ongoing surveillance for HCC is required. [d]For Child-Pugh class C, initial low dose of ribavirin 600 mg is recommended with increase as tolerated. [e]Initial dose of ribavirin 600 mg with increase as tolerated. [e]Not recommended for prior NS3/4 protease inhibitor failure or genotype 3. [f]In compensated cirrhosis, the addition of weight-based ribavirin is recommended. [g]Consider extending treatment to 24 weeks if high-risk features (e.g., genotype 3, cirrhosis, or failure following sofosbuvir-glecaprevir-pibrentasvir).

Abbreviations: ALT, alanine aminotransferase; HCV, hepatitis C virus.

with a membrane-associated phosphoprotein essential to the HCV RNA replication complex [drug name ending in "-asvir"]).

The first of the new DAA agents (approved in November 2013) was simeprevir, a second-generation protease inhibitor for genotype 1, followed shortly thereafter (December 2013) by sofosbuvir, a pangenotypic nucleoside polymerase inhibitor. For genotype 1, both of these agents had to be combined with PEG IFN and ribavirin; for genotypes 2 and 3, sofosbuvir was administered with ribavirin, without PEG IFN; however, these treatment regimens have been supplanted by combinations of all-oral, IFN-free DAAs, and ribavirin is rarely needed and retained only for very limited indications.

Simeprevir When simeprevir was used with PEG IFN, its efficacy (genotype 1b > 1a) was similar to that of first-generation protease inhibitors but required only once-a-day dosing without the complexity of response-guided therapy. Similar to first-generation protease inhibitors, however, simeprevir was hampered by many drug-drug interactions, restriction to genotype 1 HCV, side effects, and limited efficacy in patients with a highly prevalent HCV NS3 polymorphism. Little about simeprevir supported its adoption in combination with PEG IFN and ribavirin. On the other hand, the combination of simeprevir (150 mg) along with sofosbuvir (400 mg) for 12 weeks was found to be effective in treatment-naïve (97% SVR_{12}) or treatment-experienced (95% SVR_{12}) patients without cirrhosis and in treatment-naïve (88% SVR_{12}) or treatment-refractory (79% SVR_{12}) patients with cirrhosis. While this all-oral, simeprevir-sofosbuvir regimen was a landmark in the treatment of chronic hepatitis C, simeprevir was eclipsed by more effective DAAs and is no longer in clinical use.

Sofosbuvir Sofosbuvir, which inhibits the NS5B RNA-dependent RNA polymerase, was the first non–protease inhibitor DAA to be approved and had a paradigm-shifting excellent profile—high potency, high barrier to resistance, pangenotypic activity, very well tolerated with limited adverse effects (most commonly mild fatigue, insomnia, headache, and nausea), once-daily oral administration, and relative freedom from major drug-drug interactions. Sofosbuvir has efficacy in all genotypes (1–6); in treatment-naïve subjects and prior nonresponders to PEG IFN–based and protease-inhibitor-based therapy; with PEG IFN–ribavirin or in IFN-free regimens; in combination with ribavirin or with NS5A inhibitors (see below); and for treatment periods as brief as 8–12 weeks. Currently, sofosbuvir is used in combination with one of two NS5A inhibitors and is a component of three of the five currently recommended DAA regimens (Table 352-6).

Paritaprevir-Ritonavir, Ombitasvir, and Dasabuvir The combination of ritonavir (100 mg)–boosted paritaprevir (150 mg), a protease inhibitor; ombitasvir (25 mg), an NS5A inhibitor; and dasabuvir (250 mg), a nonnucleoside NS5B polymerase inhibitor, with or without weight-based ribavirin (total of five drugs), was approved in December 2014 for genotypes 1 and 4. In clinical trials, this combination achieved SVR_{12} rates of 87–100% in treatment-naïve and treatment-experienced patients with genotype 1; without ribavirin, this combination in genotype 1a was ~7% less responsive than in genotype 1b. Therefore, in treatment-naïve patients with genotype 1a, this combination was administered *with* ribavirin for 12 weeks in the absence of cirrhosis (95–97% SVR_{12}) or for 24 weeks in the presence of compensated cirrhosis (94% SVR_{12}), whereas in patients with genotype 1b, the combination did not require ribavirin, and the duration of therapy was 12 weeks for both noncirrhotic and cirrhotic patients (99–100% SVR_{12}). In July 2016, the FDA approved a long-acting formulation of dasabuvir, allowing once-a-day instead of twice-a-day treatment; for genotype 1a, twice-daily ribavirin dosing remained.

This combination was well tolerated with generally mild side effects. Hyperbilirubinemia (primarily unconjugated) and elevations in ALT levels could occur but resolved during or shortly after treatment. Because of occasional hyperbilirubinemia and potential hepatotoxicity (FDA warning letter issued October 2015 regarding hepatic failure/decompensation reported in treated cirrhotic patients), this combination (and all subsequently introduced protease-inhibitor-containing combinations) was contraindicated in patients with decompensated cirrhosis, and treated cirrhotic patients had be monitored closely for decompensation.

Compared to superior regimens that came later (e.g., sofosbuvir-ledipasvir, see below), this regimen had the disadvantage of requiring twice-a-day ribavirin therapy for genotype 1a and of being contraindicated in decompensated cirrhosis; however, it had the advantage of offering a 12-week, ribavirin-free regimen for prior null responders with cirrhosis and providing an option for patients with renal failure. Based on regimen simplicity and superiority, subsequent-generation, ribavirin-free combination DAAs have supplanted paritaprevir-ritonavir, ombitasvir, and dasabuvir; this regimen is no longer recommended by the AASLD; however, it is retained in EASL recommendations as an alternative regimen for genotype 1b only.

Sofosbuvir and Daclatasvir Daclatasvir, an NS5A inhibitor, along with the polymerase inhibitor sofosbuvir, was approved by the FDA in July 2015 for genotype 3 and in February 2016 for genotype 1. At the time of its approval for genotype 3, daclatasvir filled a need inadequately met by other available combination DAAs; however, eventually, recommendation of this combination regimen was extended to genotypes 1–4 in the United States and to all genotypes (1–6) in Europe. Daclatasvir, a 60-mg tablet, and sofosbuvir, a separate 400-mg tablet, were taken once a day for 12–24 weeks.

In clinical trials among treatment-naïve or treatment-experienced patients, SVR_{12} rates for 12 weeks of daclatasvir plus sofosbuvir were 98% for genotype 1 (comparable results in genotypes 1a and 1b), 92% for genotype 2, and 89% for genotype 3. In patients with cirrhosis, SVR_{12} was achieved in 93% with Child-Pugh class A and B but in only 56% with class C decompensated cirrhosis. For patients with genotype 3 and cirrhosis, the combination was effective in treatment-naïve patients (94% SVR_{12}) but less so in prior nonresponders (69% SVR_{12}).

Like other sofosbuvir–NS5A inhibitor combinations, daclatasvir plus sofosbuvir was well tolerated (mild fatigue, headache, nausea, or diarrhea in 5–14%) but could cause severe bradycardia when administered with amiodarone (contraindicated). Responsiveness to daclatasvir-containing drug combination therapy was reduced in cirrhotic patients with genotype 1a and in both cirrhotic and noncirrhotic patients with genotype 3 who had baseline daclatasvir-associated NS5A RASs.

As new combination DAAs were introduced, however, daclatasvir-sofosbuvir was less competitive and no longer filled a niche. This regimen has been supplanted by better, later-generation combination DAAs and is no longer recommended.

CURRENT FIRST-LINE COMBINATIONS OF DIRECT-ACTING ANTIVIRALS

Sofosbuvir-Ledipasvir The DAA combination that has had an early and dominant role in the treatment of hepatitis C is sofosbuvir (400 mg) plus the NS5A inhibitor ledipasvir (90 mg) in a once-a-day, fixed-dose, single pill, approved in October 2014 for genotype 1 and in November 2015 for genotypes 4, 5, and 6. Phase 3 trials were conducted in treatment-naïve noncirrhotic patients, in treatment-naïve cirrhotic and noncirrhotic patients, and in treatment-experienced cirrhotic and noncirrhotic patients treated for 8, 12, or 24 weeks, both with and without ribavirin. In treatment-naïve noncirrhotics, an SVR_{12} was achieved in 97–99% of subjects, and no benefit was observed by extending therapy from 12 to 24 weeks or by adding ribavirin. Moreover, for treatment-naïve, noncirrhotic patients with baseline HCV RNA $<6 \times 10^6$ IU/mL, a treatment duration of 8 weeks was as effective as one of 12 weeks (94–95% SVR_{12}), which may be a consideration for a proportion of patients. In cirrhotic patients, SVR_{12} was achieved in 97–100% of treatment-naïve subjects (no advantage of extending therapy from 12 to 24 weeks or of adding ribavirin); however, for cirrhotic prior nonresponders to IFN-based therapy, 12 weeks of therapy was inferior (86% SVR_{12})

to 24 weeks of therapy (100% SVR_{12}). This combination, which is equally effective in patients with HIV-HCV co-infection and in African-American patients, has been shown to be highly effective in patients with decompensated cirrhosis and in patients with hepatitis C after liver transplantation and after kidney transplantation. Initially, sofosbuvir-ledipasvir was not recommended in patients with advanced renal failure; however, subsequently, the safety and efficacy of sofosbuvir-ledipasvir in patients with advanced renal failure were established, and the combination was approved for this indication (November 2019). All sofosbuvir-containing regimens can be associated with severe bradycardia in patients taking the antiarrhythmic agent amiodarone, especially along with beta blockers; sofosbuvir-containing combinations are contraindicated with amiodarone. Drug-drug interactions are few, but P-glycoprotein inducers, such as St. John's wort and rifampin, and proton pump gastric acid inhibitors, such as omeprazole, may reduce sofosbuvir-ledipasvir concentrations. Generally, responsiveness to sofosbuvir-ledipasvir is not reduced in patients with baseline RASs to these agents, with the exception of treatment-experienced patients who have baseline NS5A RASs (see Table 352-6).

Elbasvir-Grazoprevir Elbasvir (50 mg), an NS5A inhibitor, combined in a single, fixed-dose pill with grazoprevir (100 mg), an NS3/4 protease inhibitor, was approved in January 2016 as a once-a-day (with or without food) treatment for genotypes 1 and 4. In clinical trials, a 12-week course was effective in treatment-naïve and treatment-experienced patients without cirrhosis or with compensated cirrhosis. In treatment-naïve patients, this combination yielded an SVR_{12} in 92% of patients with genotype 1a, 99% with genotype 1b, and 100% with genotype 4 (very small numbers, however); 10 patients with genotype 6 were included, but only 80% achieved SVR_{12}. Cirrhotic and noncirrhotic patients had comparable rates of SVR_{12}, 97% and 94%, respectively. For this drug combination, however, ~11% of patients with genotype 1a harbor NS5A polymorphisms, which may include RASs, at baseline. If present, these NS5A RASs reduce efficacy of elbasvir-grazoprevir (unlike baseline RASs to most of the other combination DAA regimens described above and below) from 99% to 58% in treatment-naïve patients. Therefore, all patients with genotype 1a required baseline RAS testing prior to starting this regimen; when these RASs were present, treatment extension to 16 weeks and the addition of weight-based ribavirin were documented to bring SVR_{12} rates up to expected levels of close to 100%. In treatment-experienced patients, both extending treatment to 16 weeks and adding ribavirin were studied; however, generally, in the absence of baseline NS5A RASs, SVR_{12} rates were not increased over those without ribavirin for 12 weeks (94–97%). Among nonresponders to prior protease inhibitor therapy, even in the absence of baseline NS5A RASs, ribavirin had to be added to a 12-week regimen; in the presence of baseline NS5A RASs, treatment was extended to 16 weeks and ribavirin added. For genotype 1b, NS5A RASs are not an issue, and the only subgroup requiring modification of a 12-week course of therapy were prior nonresponders to protease inhibitor regimens, for whom ribavirin was added. For genotype 4, the recommended regimen for all prior nonresponders (whether to PEG IFN–ribavirin or protease inhibitor regimens) was 16 weeks of elbasvir-grazoprevir plus ribavirin. Now that simpler, improved combination regimens are available, for patients with NS5A RASs, extending the duration of elbasvir-grazoprevir and adding ribavirin have been abandoned (Table 352-6); however, *elbasvir-grazoprevir is one of the currently recommended DAA combinations* (Table 352-6).

This combination is just as effective in patients with HIV-HCV co-infection and in patients with advanced renal failure (including those requiring hemodialysis), but like all protease-inhibitor-including DAA combinations, it is contraindicated in decompensated cirrhosis. In this vein, like other protease inhibitor regimens, elbasvir-grazoprevir can be associated with aminotransferase elevations and potential hepatotoxicity; because these drugs are excreted by the liver, plasma drug concentrations may become elevated substantially in the presence of impaired hepatic function. Therefore, all treated patients should have ALT screening periodically during therapy, and the drug should be stopped for elevations exceeding 10-fold or for elevations of conjugated bilirubin, alkaline phosphatase, or prothrombin time.

Elbasvir-grazoprevir is well tolerated, with only low levels of mild adverse effects (fatigue, headache, or nausea in 5–11%) seen just as frequently in placebo recipients. Both elbasvir and grazoprevir are substrates for CYP3A and are subject to multiple potential drug-drug interactions. Therefore, this combination should not be used with potent CYP3A inducers; conversely, CYP3A and OATP1B1 inhibitors can lead to untoward elevations of plasma elbasvir-grazoprevir concentrations. Checking for potential drug-drug interactions is advisable prior to initiating therapy (*www.hep-druginteractions.org*).

Compared to other available regimens for genotypes 1 and 4, elbasvir-grazoprevir has the disadvantage/inconvenience of requiring baseline NS5A RAS testing but the advantages of a comparable regimen for cirrhotic and noncirrhotic patients, for treatment-naïve and treatment-experienced patients, and for patients with normal renal function and with renal failure. As noted above, its limitations and complexities notwithstanding, elbasvir-grazoprevir remains one of the recommended DAA combinations (albeit not one of the "simplified" treatment algorithms, see below).

Sofosbuvir-Velpatasvir The combination in a single, fixed-dose pill of velpatasvir (100 mg), a highly potent, pangenotypic NS5A inhibitor, and the polymerase inhibitor sofosbuvir (400 mg) was approved in June 2016 for genotypes 1–6 in treatment-naïve and treatment-experienced noncirrhotic and cirrhotic patients. In August 2017, approval was extended to include patients with HCV-HIV co-infection. Ribavirin is not required, including in patients with genotypes 2 and 3, except in patients with decompensated cirrhosis.

In a series of clinical trials, this combination for 12 weeks in the absence of ribavirin was shown to yield a 99% SVR_{12} (range 97–100%) in genotypes 1, 2, 4, 5, and 6 and 95% in genotype 3. Baseline NS5A RASs had no impact on responsiveness.

Prior to the availability of this drug combination, patients with genotype 3, especially those with cirrhosis and prior null response to other therapies, proved to be the most refractory subset of patients. In treatment-naïve patients with genotype 3, 12 weeks of sofosbuvir-velpatasvir (95% SVR_{12}) was superior to 24 weeks of sofosbuvir plus ribavirin (80% SVR_{12}). In patients with genotype 3, the combination of sofosbuvir-velpatasvir for 12 weeks was comparable in noncirrhotic (97% SVR_{12}) and cirrhotic patients (91% SVR_{12}) and in treatment-naïve (97% SVR_{12}) and treatment-experienced (90% SVR_{12}) patients and was superior in all these categories to 24 weeks of sofosbuvir plus ribavirin (87, 66, 86, and 63%, respectively). In cirrhotic null responder patients, most available IFN-free regimens for genotype 3 (including daclatasvir plus sofosbuvir, which had been approved specifically for this genotype) achieved SVR_{12} rates in the range of ~60–75%, while the combination of PEG IFN, ribavirin, and sofosbuvir could boost SVR_{12} to the mid-80% range. For treatment-experienced patients with genotype 3, sofosbuvir-velpatasvir in noncirrhotic and cirrhotic patients had similarly high efficacy (91 and 89% SVR_{12}, respectively); this was the highest recorded SVR_{12} for genotype 3 cirrhotic null responders treated with IFN-free DAA regimens. Finally, in patients with genotypes 1–4 and 6 and with decompensated, class B cirrhosis (55% treatment-experienced), sofosbuvir-velpatasvir plus ribavirin for 12 weeks yielded an SVR_{12} in 94%; this result was better than sofosbuvir-velpatasvir without ribavirin for 12 weeks (83% SVR_{12}) or 24 weeks (86% SVR_{12}).

Like other all-oral DAAs, sofosbuvir-velpatasvir was very well tolerated; in noncirrhotic and compensated cirrhotic patients, mild headache and fatigue were seen in >10% (this occurred in a comparable proportion of placebo recipients), and in patients with decompensated cirrhosis, mild fatigue, headache, nausea, insomnia, diarrhea, and anemia (ribavirin was part of the regimen) were seen in >10%. Like other sofosbuvir-containing regimens, sofosbuvir-velpatasvir should not be administered along with amiodarone (potential serious bradycardia); in addition, P-glycoprotein inducers and moderate-to-potent CYP3A inducers can reduce plasma levels of sofosbuvir and/or velpatasvir. Checking for drug-drug interactions prior to therapy is advisable (*www.hep-druginteractions.org*). Baseline RASs do not influence responsiveness to this combination. Sofosbuvir-velpatasvir is one of the currently recommended DAA combinations for hepatitis C (Table 352-6). *Because it is so simple and broadly effective across patient subgroups, sofosbuvir-velpatasvir is one of the two combination DAA regimens recommended by the AASLD and EASL as a preferred, simplified treatment algorithm* (Table 352-6).

Sofosbuvir-Velpatasvir-Voxilaprevir Approved in July 2017, the pangenotypic, high-barrier-to-resistance protease inhibitor voxilaprevir (100 mg) added to the polymerase inhibitor–NS5A inhibitor combination of sofosbuvir-velpatasvir yields a very well tolerated *triple-drug* combination with ~97% SVR_{12} across all HCV genotypes and patient subgroups. These include the small percentage of patients with genotype 1 and genotype 3 refractory to previously approved DAA combinations as well as noncirrhotic/cirrhotic, treatment-naïve/treatment-experienced groups, including those who had or who had not received prior NS5A treatment. Efficacy was independent of the number of prior DAA drug classes received, and no effects of baseline NS5A RASs were noted.

The potential for abbreviated (8-week) treatment with this triple combination was explored in a clinical trial involving treatment-naïve patients; however, the shortened duration was inferior to a full 12-week course. The side effect profile for sofosbuvir-velpatasvir-voxilaprevir was similar to that in the placebo arm of clinical trial patients and included mild and uncommon headache, fatigue, nausea, and diarrhea.

Because other DAA regimens are so effective in most patients with chronic hepatitis C, recommendations for sofosbuvir-velpatasvir-voxilaprevir are limited to a small subset of otherwise refractory patients: for treatment-naïve cirrhotic patients with genotype 3 and baseline NS5A velpatasvir RAS Y93H, for treatment-naïve (according to AASLD, not EASL) or IFN-ribavirin–experienced noncirrhotic or cirrhotic patients with genotype 3 (Table 352-6), and for patients with or without compensated cirrhosis and prior, failed NS5A inhibitor–containing therapy (consult *www.hcvguidelines.org*).

This triple-drug combination, like all sofosbuvir-containing combinations, is contraindicated in patients taking amiodarone and, like all protease inhibitor–containing combinations, in patients with decompensated cirrhosis. Concomitant omeprazole, 20 mg, can be taken with this sofosbuvir-containing regimen. Prior to initiating therapy, checking for drug-drug interactions is recommended.

Glecaprevir-Pibrentasvir A regimen of 8 weeks of this single-pill, fixed-dose combination of the protease inhibitor glecaprevir (300 mg) and NS5A inhibitor pibrentasvir (120 mg), two pangenotypic, high-potency DAAs with high barriers to resistance (approved in August 2017), achieves SVR_{12} in close to 100% of treatment-naïve patients with all genotypes, with or without cirrhosis: SVR_{12} of ~99% for genotypes 1, 2, and 4–6 and of 95–98% for genotype 3. Extended treatment for 12 weeks did not increase efficacy. In trials among treatment-experienced patients, treatment with 12 weeks of this DAA combination was just as effective as 16 weeks for all genotypes except genotype 3; however, with increasing numbers of prior treatment courses, SVR_{12} rates fell—100% for patients treated with a protease inhibitor only, 88% for patients treated with an NS5A inhibitor only, and 79% for patients treated previously with both a protease inhibitor and an NS5A inhibitor. Similarly, baseline RASs reduced SVR_{12} rates—from 100% without RASs (or with RASs limited to those reflecting protease inhibitor resistance) to 89% for baseline NS5A RASs.

For retreatment of patients with prior glecaprevir-pibrentasvir failure, 16 weeks of glecaprevir-pibrentasvir plus sofosbuvir are recommended (alternatively, sofosbuvir-velpatasvir-voxilaprevir for 12 weeks [plus ribavirin in cirrhotic patients]). Glecaprevir-pibrentasvir for 16 weeks is recommended as well after failure to respond to the triple-drug combination of sofosbuvir-velpatasvir-voxilaprevir (see below). For retreatment of patients with sofosbuvir-velpatasvir-voxilaprevir failure, 16 weeks of glecaprevir-pibrentasvir plus ribavirin is recommended, as is a repeat course of sofosbuvir-velpatasvir-voxilaprevir plus ribavirin for 24 weeks. For navigating the complexities of treating DAA-refractory patients, checking up-to-date guidelines (*www.hcvguidlines.org*) and referral to a specialized center are advisable.

As is the case for any DAA combination containing a protease inhibitor, glecaprevir-pibrentasvir is contraindicated in decompensated cirrhosis; it has been shown to achieve an SVR_{12} in 98% of patients with stage 4 or 5 renal disease (in treatment-naïve or experienced, cirrhotic or noncirrhotic patients) and is a preferred treatment for patients with severe renal impairment. This DAA combination should be taken with food, and drug-drug interactions should be considered prior to initiating treatment. *Because it is so simple and broadly effective across patient subgroups (8 weeks for all noncirrhotic treatment-naïve patients [including with HIV co-infection]; 12 weeks for all treatment-experienced cirrhotics [including with HIV co-infection] and treatment-naïve cirrhotics with genotype 3 [except treatment-experienced cirrhotic or noncirrhotic genotype 3 {16 weeks}]), glecaprevir-pibrentasvir is one of the two combination DAA regimens recommended by the AASLD and EASL as a preferred, simplified treatment algorithm* (Table 352-6).

Emerging data on the impact of DAAs on the natural history of chronic hepatitis C indicated that, as was documented for IFN-based therapy, successful DAA therapy is associated with a gradual reduction in fibrosis progression and a regression of advanced fibrosis (cirrhosis), improvement in survival among patients with decompensated cirrhosis, a reduction in HCC, and a decline in the number of patients with hepatitis C being referred for liver transplantation. Indeed, early and broad implementation of DAAs in the treatment of chronic hepatitis C has been instrumental in enabling the decline of HCV-related deaths since 2013. Modeling estimates suggest that application of DAA therapies has the potential to result in a 50–70% reduction in the hepatitis C–associated disease burden by 2050.

TREATMENT RECOMMENDATIONS

Because the pace of new drug development and approval has been so rapid, the AASLD and the Infectious Diseases Society of America (IDSA) have been providing a consensus of updated treatment recommendations for patients with hepatitis C; these recommendations, which continue to be revised regularly based on new data, are available online at *www.hcvguidelines.org* and should be consulted before initiating therapy. The EASL issues similar (but not identical) treatment recommendations annually for hepatitis C (*www.easl.eu*), most recently in November 2020.

Prior to therapy, all patients should undergo pretreatment evaluation, which includes HCV RNA, HCV genotype, renal function, hepatitis B status (testing should include HBsAg and anti-HBc), HIV status, and fibrosis stage. Despite the presence of pangenotypic regimens, HCV genotype should still be determined, because the genotype may contribute to decisions about which treatment regimens are indicated and can guide insurance coverage (Table 352-6).

Fibrosis assessment is an especially important element of the pretreatment evaluation: the absence or presence of cirrhosis or advanced fibrosis determines the treatment options from which to select, including the antiviral agents to be used, the duration of therapy, and the now rare need for ribavirin (Table 352-6). In the past, a pretreatment liver biopsy was relied upon to assess histologic grade and stage as well as to identify such histologic factors as steatosis, which can influence adversely responsiveness to therapy. As therapy has improved for patients with a broad range of histologic severity, and as noninvasive measures of the stage of fibrosis (e.g., assessment of liver stiffness by imaging, FIB-4 score [see above]) have advanced in accuracy and popularity, noninvasive approaches have supplanted histology in almost most cases. As noted above, if cirrhosis or advanced fibrosis is present prior to therapy, the risk of HCC, although reduced substantially by successful therapy, is not eliminated, and twice-yearly posttreatment imaging for HCC surveillance remains indicated even after an SVR. In patients with low-level fibrosis at baseline, achievement of SVR allows cessation of such surveillance.

Several reports have appeared describing hepatitis B reactivation, often severe, during and after DAA therapy in patients co-infected with HCV and HBV who were not being treated for their HBV infections. Therefore, screening for HBV infection is recommended prior to initiating DAA therapy for hepatitis C (which should have been done to determine HBV immunity status as a prelude to recommended hepatitis B vaccination in patients with chronic hepatitis C), and therapy for HBV infection (for those meeting HBV treatment criteria, see above) should be initiated prior to or simultaneously with HCV therapy.

Monitoring of serum HCV RNA levels before, during, and after treatment is crucial in assessing the response to therapy; moreover, the baseline level may contribute to determining the duration of therapy (e.g., in noncirrhotic patients with genotype 1 and HCV RNA $<6 \times 10^6$ IU/mL, 8 [instead of the usual 12] weeks of sofosbuvir-ledipasvir may be a consideration). *The goal of treatment is to eradicate HCV RNA during therapy and to document that the virus remains undetectable for at least 12 weeks after completion of therapy (*SVR_{12}*).*

Most recent guidelines support a simplified treatment algorithm for patients with HCV infection, leveraging the well tolerated and highly effective pangenotypic regimens. Patients who are eligible for this simplified algorithm must *not* have any of the following clinical characteristics: prior treatment failure, cirrhosis, presence of HBsAg, pregnancy, HCC, or history of liver transplantation. In the absence of any of these features, the patient may be treated with *either* sofosbuvir/velpatasvir for 12 weeks or glecaprevir/pibrentasvir for 8 weeks.

INDICATIONS FOR ANTIVIRAL THERAPY

Patients with chronic hepatitis C who have detectable HCV RNA in serum, whether or not aminotransferase levels are increased, and chronic hepatitis of any grade and stage are candidates for antiviral therapy with DAA agents. The only exception would be patients with short life expectancies, for whom treating hepatitis C would have no influence on longevity. Certainly, for patients with advanced liver disease, early treatment merits a high priority. Although patients with persistently normal aminotransferase activity tend to progress histologically very slowly or not at all, they respond to antiviral therapy just as well as do patients with elevated aminotransferase levels; therefore, such patients are candidates for antiviral therapy. As noted above, antiviral therapy has been shown to improve survival and complication-free survival and to slow progression of and to reverse fibrosis.

Patients who have relapsed after, or failed to respond to, a course of IFN-based (currently no longer encountered commonly) or DAA agent–based therapy are candidates for retreatment with a DAA therapy regimen (Table 352-6). For patients who have failed to respond to a DAA combination, options include increasing the duration of therapy with the failed regimen, adding ribavirin, or changing the drug class (e.g., after failed protease and polymerase inhibitors, switching to an NS5A-containing combination). In the presence of cirrhosis or a need for urgent retreatment, patients who have failed protease inhibitor plus polymerase inhibitor combination therapy or who have failed an NS5A combination are candidates for RAS testing and tailored therapy based on such resistance testing. If reliable RAS testing is not available, adding ribavirin or extending the duration of therapy are options. For prior nonresponders to IFN-based therapy, NS5A inhibitor–containing regimens are highly effective; however, reduced responsiveness can be encountered, especially in cirrhotic patients. The simplest current approach for patients who have not responded to prior DAA therapy containing sofosbuvir is to re-treat with sofosbuvir-velpatasvir-voxilaprevir for 12 weeks or glecaprevir-pibrentasvir for 16 weeks. Additional details for treatment of such patient subgroups can be found at www.hcvguidelines.org; in the small proportion of patients in whom initial DAA courses do not achieve SVR_{12}, referral to an experienced treatment center is advisable. Worth reiterating, protease inhibitors are contraindicated for patients with decompensated cirrhosis, and sofosbuvir-containing regimens are not recommended for patients taking amiodarone (especially with beta blockers) for treatment of cardiac arrhythmias. While sofosbuvir-containing DAA combinations were not recommended initially for patients with advanced renal failure, subsequent studies demonstrated safety and efficacy in this subgroup, and sofosbuvir-containing DAA combinations are now approved for advanced renal failure.

Persons with *acute* hepatitis C are also candidates for antiviral therapy **(Chap. 350)** with the same pangenotypic combination DAA agents and the same duration of treatment approved for chronic hepatitis C. For example, in a trial of a shortened, 6-week course of sofosbuvir-velpatasvir for acute hepatitis C, a 6-week course was less effective than a full 12-week course (81.7 vs 90.5%, respectively). Delaying the initiation of therapy in acute hepatitis C to allow for spontaneous recovery is no longer recommended. While acute hepatitis C should be treated, postexposure prophylaxis after HCV exposure is not recommended currently

In patients with biochemically and histologically mild chronic hepatitis C, the rate of progression is slow; however, such patients respond just as well to antiviral therapy as those with elevated aminotransferase levels and more histologically severe hepatitis. Because of the high cost of DAA treatments, in the past, a higher priority was assigned to patients with advanced fibrosis/cirrhosis; however, this controversial approach was relied upon by some medical insurers and pharmacy benefit management organizations to withhold therapy from patients with low-level fibrosis. Unfortunately, delaying therapy until fibrosis becomes advanced misses the opportunity to prevent all the dire consequences of chronic hepatitis C (liver failure, death/transplantation, HCC), which can be reduced, but not eliminated completely, once advanced fibrosis is established. Therefore, therapy for patients with mild disease is justified as well as cost-effective.

TREATMENT OF UNIQUE POPULATIONS

Decompensated Cirrhosis Patients with compensated cirrhosis can respond to therapy, and their likelihood of a sustained response with DAAs is comparable to that in the absence of cirrhosis. Patients with decompensated cirrhosis, who were not candidates for IFN-based antiviral therapy, respond well to non–protease-inhibitor-containing regimens, and specifically sofosbuvir combined with an NS5A inhibitor, velpatasvir or ledipasvir, with or without ribavirin. As noted above, protease-inhibitor-containing combinations have been associated with potential hepatotoxicity and hepatic decompensation and are contraindicated in this patient subset. Treatment decisions are informed by whether the patient is eligible for ribavirin: in ribavirin-eligible patients, ribavirin is added to a 12-week course of sofosbuvir-NS5A therapy; however, in cases of ribavirin ineligibility, the duration of therapy should be extended to 24 weeks. In cases of prior failure to respond to sofosbuvir-NS5A therapy, the sofosbuvir-NS5A regimen should be repeated but supplemented with ribavirin and extended to 24 weeks (*www.hcvguidelines.org*). Also, patients with decompensated cirrhosis should be referred to a liver transplantation center.

DAAs are highly effective not only for patients with end-stage liver disease awaiting liver transplantation but also for patients with recurrent hepatitis C after liver transplantation. Ideally, patients should be treated prior to liver transplantation. Given the effectiveness of treatment immediately after transplantation (either for donor or recipient HCV infection), however, the utility of treatment may be limited if the patient has a high Model for End-Stage Liver Disease (MELD) score and may soon receive a donor liver for transplantation. Therefore, advocacy has been expressed for postponing DAA therapy in patients with high-MELD-score (≥18–20), HCV-associated, end-stage liver disease until after liver transplantation; for patients with MELD scores <18–20, pretransplantation DAA therapy is advised. Still, the decision whether to treat pretransplantation or posttransplantation should be individualized thoughtfully for each patient, based on such factors as MELD score, time anticipated prior to availability of a donor organ, relative clinical stability, and comorbidities **(Chap. 356)**. Because DAA therapy is so effective, most transplantation centers are accepting organs from HCV-infected donors, transplanting them into HCV-uninfected recipients, and treating recipients with sofosbuvir-velpatasvir for 12 weeks or glecaprevir-pibrentasvir for 8 weeks after transplantation—with excellent results.

Extrahepatic Manifestations The cutaneous and renal vasculitis of HCV-associated essential mixed cryoglobulinemia **(Chap. 350)** may respond to antiviral therapy, but sustained responses were rare after discontinuation of therapy in the IFN era, and prolonged, potentially indefinite, therapy was recommended. Now that more effective DAAs are available, a 12-week course of sofosbuvir-based combination therapy has been shown to yield an SVR_{12} rate exceeding 80% in cryoglobulinemic vasculitis. Anecdotal reports suggest that IFN-based antiviral therapy may be effective in porphyria cutanea tarda or lichen planus associated with hepatitis C; whether the more appealing DAAs are effective in these groups remains to be documented.

HIV Co-infection In patients with HCV/HIV co-infection, hepatitis C is more progressive and severe than in HCV mono-infected patients. Although patients with HCV/HIV co-infection responded less well to IFN-based antiviral therapy for hepatitis C, they respond as well as patients with HCV infection alone to DAA combination regimens and, generally, should be treated with the same DAA combinations and for the same duration as those for patients (in the absence or presence of cirrhosis) with HCV mono-infection. For patients with HCV/HIV co-infection, an abbreviated, 8-week course of sofosbuvir-ledipasvir for low-level HCV RNA is not recommended, and a full 12 weeks should be given; similarly, for patients with genotype 4, a 12-week course of glecaprevir-pibrentasvir is recommended instead of an 8-week course for treatment-naïve or -experienced patients with or without cirrhosis (Table 352-6). In HCV/HIV-infected patients, ribavirin can potentiate the toxicity of didanosine (e.g., lactic acidosis) and the lipoatrophy of stavudine, and zidovudine can exacerbate ribavirin-associated hemolytic anemia; therefore, these drug combinations should be avoided. In HCV/HIV co-infected persons, the list of potential drug-drug interactions is extensive and should be consulted carefully before beginning DAA treatment (*www.hcvguidelines.org*).

Substance Use Disorder Patients with a history of injection drug use and alcoholism can be treated successfully for chronic hepatitis C, preferably in conjunction with drug and alcohol treatment programs. Moreover, because persons who use injection drugs (PWID), as a source of transmission to others, account disproportionately for perpetuating the spread of HCV infection in the population, the impact of treating people who are actively using injection drugs is amplified by reducing such transmission. In a randomized, controlled clinical trial, treatment with sofosbuvir and velpatasvir for 12 weeks was highly successful in PWID, with SVR_{12} rates of 74–76%. Based on this study, and others, active injection drug use is no longer considered a contraindication to treatment, and indeed, mechanisms to facilitate care of PWID are being sought to address the U.S. and global burden of HCV infection and move toward the World Health Organization goal of HCV elimination by 2030 (see below).

Renal Impairment The approved oral DAA combinations are effective in patients with mild-modest renal impairment and require no dose adjustments. For patients with severe renal impairment (creatinine clearances <30 mL/min), including those undergoing hemodialysis, recommended combinations are 12 weeks of elbasvir-grazoprevir for genotypes 1 and 4 or 12 weeks of glecaprevir-pibrentasvir for all genotypes. Both in severe renal impairment and after kidney transplantation, levels of SVR_{12} in patients treated with these oral DAA combinations have approached 100%. Initially, in patients with severe renal impairment, sofosbuvir-containing combinations were not recommended. Subsequently, however, based on efficacy and safety in a series of clinical trials, sofosbuvir-containing regimens were approved by the FDA in November 2019 for patients with severe renal impairment.

Pregnancy No large clinical studies of the use of DAAs during pregnancy are available. One small study demonstrated the safety and efficacy of sofosbuvir and ledipasvir given for 12 weeks at week 23–24 of gestation. Ribavirin is contraindicated during pregnancy; therefore, any regimen including ribavirin should not be used. Sofosbuvir; sofosbuvir-ledipasvir; and paritaprevir-ritonavir, ombitasvir, and dasabuvir are classified as pregnancy category B; the other DAAs do not have a pregnancy classification. Therefore, these therapies are not indicated routinely in pregnancy and should be used, with caution, only if the benefit of treatment is compelling and justified compared to the potential for fetal risk. Currently, screening all pregnant women for HCV infection is recommended. Breast feeding is not contraindicated in women with HCV infection (unless the mother has a break in the integrity of the nipples or is co-infected with HIV).

CHOOSING AMONG AVAILABLE TREATMENT OPTIONS

Choosing among the number of all-oral DAA combinations approved since 2013 was daunting to treating clinicians. Currently, however, the number of recommended DAA combinations has narrowed to a very manageable few. The most popular of the regimens have been fixed-dose, single-pill, pangenotypic combinations. For simplicity, two "one-size-fits-all" pangenotypic regimens—sofosbuvir-velpatasvir and glecaprevir-pibrentasvir—can be used, for 8–12 weeks in almost all treatment-naïve, noncirrhotic and cirrhotic patients, including those with advanced renal failure and HCV-HIV co-infection. Applicability of the triple-drug combination sofosbuvir-velpatasvir-voxilaprevir is quite limited in treatment-naïve patients, reserved primarily for cirrhotic patients with genotype 3. As noted above, protease-inhibitor-containing DAA regimens (elbasvir-grazoprevir, glecaprevir-pibrentasvir, and sofosbuvir-velpatasvir-voxilaprevir) are contraindicated in decompensated cirrhosis.

APPROACHES TO STEAMLINING DAA TREATMENT

The World Health Organization has set an elimination goal of a 90% reduction in the global prevalence of HCV infection by 2030, and the Viral Hepatitis National Strategic Plan includes a goal of an 80% reduction in the U.S. prevalence of HCV infection by 2030; however, at the current pace of identification and treatment, this ambitious goal is unlikely to be met. To address this shortfall, efforts have been devoted to improving the frequency of screening and diagnosis and to streamlining the process of enrolling persons with HCV infection into treatment programs. Now that DAA regimens have been simplified so dramatically, in-person specialist care is usually not needed (except in complex patients such as those with advanced fibrosis, comorbid conditions, and prior nonresponse to DAA treatment). Programs with an impressive record of success in streamlining the treatment of persons with hepatitis C include those based on telehealth, one-stop screening and treatment initiation, minimal monitoring (dispensing the full course of pills at the beginning of therapy, relying on a limited number of in-person or even all-remote patient contacts during therapy, and involving no HCV RNA or liver test monitoring during treatment and not until 12 weeks after completion of treatment), and collaborative participation on the treatment team of clinicians experienced in managing substance use disorder. The expansion of such programs is anticipated eagerly.

AUTOIMMUNE HEPATITIS

DEFINITION

Autoimmune hepatitis (AIH) is a chronic disorder characterized by continuing hepatocellular necrosis and inflammation, usually with fibrosis, which can progress to cirrhosis and liver failure. The presentation of AIH varies, from asymptomatic liver enzyme abnormalities to severe or even, in its most severe form, marked hepatocellular dropout, multilobular collapse, and clinical features of fulminant hepatitis with high mortality. Based on contemporary estimates of the natural history of AIH, the 10-year survival is 80–98% for treated and 67% for untreated patients. The prominence of extrahepatic features of autoimmunity and seroimmunologic abnormalities in this disorder supports an autoimmune process in its pathogenesis; this concept is reflected in the prior labels *lupoid* and *plasma cell hepatitis*. Similar to other autoimmune diseases, AIH has a strong female predominance, with women accounting for ~75% of all cases. The age distribution is bimodal, with one peak in the early decades (age <30) and the second in middle to late years (often >60). Autoantibodies and other typical features of autoimmunity, however, do not occur in all cases; among the broader categories of "idiopathic" or cryptogenic chronic hepatitis, many, perhaps the majority, are probably autoimmune in origin. Cases in which hepatotropic viruses, metabolic/genetic derangements (including nonalcoholic fatty liver disease), and hepatotoxic drugs have been excluded represent a spectrum of heterogeneous liver disorders of unknown cause, a proportion of which are most likely AIH.

IMMUNOPATHOGENESIS

The weight of evidence suggests that the progressive liver injury in patients with AIH is the result of a cell-mediated immunologic attack directed against liver cells in the setting of a loss of, or failed, immunologic tolerance for self-liver antigens. In all likelihood, predisposition to autoimmunity is inherited, whereas the liver specificity of this injury is triggered by environmental (e.g., chemical, drug [e.g., minocycline], or viral) factors. For example, patients have been described in whom apparently self-limited cases of acute hepatitis A, B, or C led to AIH, presumably because of genetic susceptibility or predisposition. More recently, the emergence of AIH has been reported following—presumably triggered by—COVID-19 and after COVID-19 vaccination, postulated mechanistically to result from the vigorous immune response and loss of immune tolerance in susceptible persons.

Evidence to support an autoimmune pathogenesis in this type of hepatitis includes the following: (1) in the liver, the histopathologic lesions are composed predominantly of cytotoxic T cells and plasma cells; (2) circulating autoantibodies (nuclear, smooth muscle, and others, see below), rheumatoid factor, and hyperglobulinemia are common; (3) other autoimmune disorders—such as autoimmune thyroiditis, rheumatoid arthritis, autoimmune hemolytic anemia, ulcerative colitis, membranoproliferative glomerulonephritis, juvenile diabetes mellitus, vitiligo, celiac disease, and Sjögren's syndrome—occur with increased frequency in patients who have AIH and in their relatives; (4) histocompatibility haplotypes associated with autoimmune diseases, such as HLA-B1, B8, DR3, and DR4 as well as extended haplotype *DRB1*0301* and *DRB1*0401* alleles, are common in patients with AIH; and (5) this type of chronic hepatitis is responsive to glucocorticoid/immunosuppressive therapy, effective in a variety of autoimmune disorders.

Loss of immune tolerance as well as cytotoxic cellular immune responses appear to be important in the pathogenesis of AIH. In vitro studies have suggested that in patients with this disorder, CD4+ T lymphocytes are capable of becoming sensitized to hepatocyte membrane proteins and of destroying liver cells. Molecular mimicry by cross-reacting antigens that contain epitopes similar to liver antigens is postulated to activate these T cells, which infiltrate, and result in injury to, the liver. Abnormalities of immunoregulatory control over cytotoxic lymphocytes (impaired regulatory CD4+CD25+ T-cell influences) may play a role as well. Studies of genetic predisposition to AIH demonstrate that certain haplotypes are associated with the disorder, as enumerated above, as are polymorphisms in cytotoxic T lymphocyte antigens (*CTLA-4*) and tumor necrosis factor α (*TNFA*2*). The precise triggering factors, genetic influences, and cytotoxic and immunoregulatory mechanisms involved in this type of liver injury remain incompletely defined.

Intriguing clues into the pathogenesis of AIH come from the observation that circulating autoantibodies are prevalent in patients with this disorder. Among the autoantibodies described in these patients are antibodies to nuclei (so-called antinuclear antibodies [ANAs], primarily in a homogeneous pattern) and smooth muscle (so-called anti-smooth-muscle antibodies [ASMAs], directed at actin, vimentin, and skeletin), antibodies to F-actin and anti-LKM (see below), antibodies to soluble liver antigen (directed against a uracil-guanine-adenine transfer RNA suppressor protein), antibodies to α-actinin, and antibodies to the liver-specific asialoglycoprotein receptor (or "hepatic lectin") and other hepatocyte membrane proteins. Although some of these provide helpful diagnostic markers, their involvement in the pathogenesis of AIH has not been established.

Humoral immune mechanisms have been shown to play a role in the extrahepatic manifestations of autoimmune and idiopathic hepatitis. Arthralgias, arthritis, cutaneous vasculitis, and glomerulonephritis occurring in patients with AIH appear to be mediated by the deposition of circulating immune complexes in affected tissue vessels, followed by complement activation, inflammation, and tissue injury. While specific viral antigen-antibody complexes can be identified in acute and chronic viral hepatitis, the nature of the immune complexes in AIH has not been defined.

CLINICAL FEATURES

Many of the *clinical features* of AIH are similar to those described for chronic viral hepatitis. The onset of disease may be insidious or abrupt; the disease may present initially like, and be confused with, acute viral hepatitis. The most common symptom, occurring in ~85% of patients, is fatigue. In severe disease, which accounts for 20–25% of cases, jaundice may be present as well. In approximately a quarter of patients, the diagnosis is made in the absence of symptoms, based on abnormal liver laboratory tests. A subset of patients with AIH has distinct features. Such patients are predominantly young to middle-aged women with marked hyperglobulinemia and high-titer circulating ANAs. This is the group with positive lupus erythematosus (LE) preparations (initially labeled *lupoid* hepatitis) in whom other autoimmune features are common. Fatigue, malaise, anorexia, amenorrhea, acne, arthralgias, and jaundice are common. Occasionally, arthritis, maculopapular eruptions (including cutaneous vasculitis), erythema nodosum, colitis, pleuritis, pericarditis, anemia, azotemia, and sicca syndrome (keratoconjunctivitis, xerostomia) occur. In some patients, complications of cirrhosis, such as ascites and edema (associated with portal hypertension and hypoalbuminemia), encephalopathy, hypersplenism, coagulopathy, or variceal bleeding may bring the patient to initial medical attention.

The course of AIH may be variable. In patients with mild disease or limited histologic lesions (e.g., interface hepatitis, labeled in the past as piecemeal necrosis [inflammation and erosion of the limiting place of periportal hepatocytes] without bridging necrosis and fibrosis), progression to cirrhosis may be limited, but, even in this subset, treatment and careful monitoring remain warranted, because, if untreated, up to half can progress to cirrhosis over 15 years. Also, asymptomatic patients, if left untreated, have a lower 10-year overall survival. In North America, cirrhosis at presentation is more common in African Americans than in whites. In those with severe symptomatic AIH (aminotransferase levels >10 times normal, marked hyperglobulinemia, "aggressive" histologic lesions—bridging necrosis or multilobular collapse, cirrhosis), the 6-month mortality without therapy may be as high as 40%. Such severe disease accounts for only 20% of cases; the natural history of milder disease is variable, often accentuated by spontaneous remissions and exacerbations. In a 10-year (2006–2016) national Dutch study, mortality in patients with AIH was higher than that of the general population only in patients with cirrhosis; for patients without cirrhosis, survival was comparable to that of the general population. Especially poor prognostic signs include the presence histologically of multilobular collapse at the time of initial presentation and failure of serum bilirubin to improve after 2 weeks of therapy. Death may result from hepatic failure, hepatic coma, other complications of cirrhosis (e.g., variceal hemorrhage), and intercurrent infection. In patients with established cirrhosis, HCC may be a late complication **(Chap. 87)** but occurs less frequently than in cirrhosis associated with viral hepatitis.

Laboratory features of AIH are similar to those seen in chronic viral hepatitis. Liver biochemical tests are invariably abnormal but may not correlate with the clinical severity or histopathologic features in individual cases. Many patients with AIH have normal serum bilirubin, alkaline phosphatase, and globulin levels with only minimal aminotransferase elevations. Most commonly, serum AST and ALT levels are increased and fluctuate in the range of 100–1000 units. In severe cases, the serum bilirubin level is moderately elevated (51–171 μmol/L [3–10 mg/dL]). Hypoalbuminemia occurs in patients with very active or advanced disease. Serum alkaline phosphatase levels may be moderately elevated or near normal. In a small proportion of patients, marked elevations of alkaline phosphatase activity occur; in such patients, clinical and laboratory features overlap with those of primary biliary cholangitis **(Chap. 355)**. The prothrombin time is often prolonged, particularly late in the disease or during active phases.

Polyclonal hypergammaglobulinemia (>2.5 g/dL) is common in AIH, as is the presence of rheumatoid factor. As noted above, circulating autoantibodies are also prevalent, most characteristically ANAs in a homogeneous staining pattern. Smooth-muscle antibodies are detected in 63% of cases and, when combined with a positive ANA, have high specificity. Of important note, low-titer ANA or ASMA is observed often in other liver diseases, including metabolic dysfunction–associated steatotic liver disease (in one series, 20% of patients had a positive ANA and 14% had a positive ASMA) and viral hepatitis. Studies of autoantibodies in AIH have led to the recognition of two categories of AIH. *Type I AIH* is the classic syndrome prevalent in North America and northern Europe occurring in young women, associated with marked hyperglobulinemia, lupoid features, circulating ANAs, and HLA-DR3 or HLA-DR4 (especially *B8-DRB1*03*). Also associated with type I AIH are autoantibodies against F-actin and atypical perinuclear antineutrophilic cytoplasmic antibodies (pANCA). Included in the spectrum of type I AIH is a subset of patients who lack ANA and anti-LKM1 but who have circulating antibodies to soluble liver antigen (SLA). Most of these patients are women and have clinical features similar to, or perhaps more severe than, those of other patients with type I AIH. The presence of SLA autoantibodies has been reported to be associated with a worse prognosis.

Type II AIH, often seen in children, more common in Mediterranean populations, and linked to HLA-DRB1 and HLA-DQB1 haplotypes, is associated not with ANA but with anti-LKM. Actually, anti-LKMs represent a heterogeneous group of antibodies. In type II AIH, the antibody is anti-LKM1, directed against cytochrome P450 2D6. This is the same anti-LKM seen in some patients with chronic hepatitis C. Anti-LKM2 is seen in drug-induced hepatitis, and anti-LKM3 (directed against uridine diphosphate glucuronyltransferases) is seen in patients with chronic hepatitis D but has also been observed in type II AIH. Another autoantibody observed in type II AIH is directed against liver cytosol formiminotransferase cyclodeaminase (anti-liver cytosol 1 [LC-1]).

Liver biopsy is imperative to make the diagnosis of AIH. Abnormalities are similar to those described for chronic viral hepatitis, but classic features of AIH include inflammation involving not only the portal tracts but extending beyond the portal tracts (*interface hepatitis*), the presence of plasma cells in the inflammatory infiltrate, and central vein perivenulitis. Necroinflammatory activity characterizes the lobular parenchyma, and evidence of hepatocellular regeneration is reflected by "rosette" formation, the occurrence of thickened liver cell plates, and regenerative "pseudolobules." Septal fibrosis, bridging fibrosis, and cirrhosis are frequent. In patients with early AIH presenting as an acute-hepatitis-like illness, lobular and centrilobular (as opposed to the more common periportal) necrosis has been reported. Bile duct injury and granulomas are uncommon; however, a subgroup of patients with AIH has histologic, biochemical, and serologic features overlapping those of primary biliary cholangitis **(Chap. 355)**.

DIAGNOSTIC CRITERIA

An international group has suggested a set of criteria for establishing a diagnosis of AIH. Exclusion of liver disease caused by genetic disorders, viral hepatitis, drug hepatotoxicity, and alcohol is linked with such inclusive diagnostic criteria as hyperglobulinemia, autoantibodies, and characteristic histologic features. This international group has also suggested a comprehensive diagnostic scoring system that, rarely required for typical cases, may be helpful when typical features are not present. Factors that weigh in favor of the diagnosis include female gender; predominant aminotransferase elevation; presence and level of globulin elevation; presence of nuclear, smooth-muscle, LKM1, and other autoantibodies; concurrent other autoimmune diseases; characteristic histologic features (interface hepatitis, plasma cells, rosettes); HLA-DR3 or DR4 markers; and response to treatment (see below). A more simplified, more specific scoring system relies on four variables: presence of autoantibodies (ANA, ASMA, anti-SLA or LKM), elevated serum IgG level, typical or compatible histologic features, and absence of viral hepatitis markers. Elevated aminotransferase levels are also required for the diagnosis of AIH. Weighing against the diagnosis are predominant alkaline phosphatase elevation, mitochondrial antibodies, markers of viral hepatitis, history of hepatotoxic drugs or excessive alcohol, histologic evidence of bile duct injury, or such atypical histologic features as fatty infiltration, iron overload, and viral inclusions.

DIFFERENTIAL DIAGNOSIS

Early during the course of chronic hepatitis, AIH may resemble typical *acute viral hepatitis or drug-induced liver injury* **(Chap. 350)**. Exclusion of acute viral hepatitis with serologic testing is critical. In adolescence, *Wilson disease* **(Chaps. 355 and 427)** may present with features of chronic hepatitis long before neurologic manifestations become apparent and before the formation of Kayser-Fleischer rings (copper deposition in Descemet's membrane in the periphery of the cornea). In this age group, serum ceruloplasmin and serum and urinary copper determinations plus measurement of liver copper levels establish the correct diagnosis. *Cryptogenic cirrhosis* and *primary biliary cholangitis* **(Chap. 355)** share clinical features with AIH, and both alcohol-associated hepatitis **(Chap. 353)** and metabolic dysfunction–associated steatohepatitis (MASH; formerly nonalcoholic steatohepatitis [NASH], **Chap. 354**) may present with many features common to AIH; historic, biochemical, serologic, and histologic assessments are usually sufficient to allow these entities to be distinguished from AIH. The distinction between autoimmune and chronic viral hepatitis may not be straightforward, especially when viral antibodies occur in patients with autoimmune disease or when autoantibodies occur in patients with viral disease. Furthermore, the presence of extrahepatic features such as arthritis, cutaneous vasculitis, or pleuritis—not to mention the presence of circulating autoantibodies—may cause confusion with *rheumatologic disorders* such as rheumatoid arthritis and systemic LE. The existence of clinical and biochemical features of progressive necroinflammatory liver disease distinguishes chronic hepatitis from these other disorders, which are not associated with severe liver disease. Rarely, hepatic venous outflow obstruction (Budd-Chiari syndrome) may present with features suggestive of AIH, but painful hepatomegaly, ascites, and vascular imaging provide distinguishing diagnostic clues. Other diagnostic considerations would include celiac disease and ischemic liver disease, which would be readily distinguishable by clinical and laboratory features from AIH.

In patients treated with immune checkpoint inhibitors for malignancy, the liver may be one of the autoimmune targets of therapy; the syndrome of immune checkpoint inhibitor (ICI) hepatitis resembles AIH in clinical features (without the female preponderance, however) and response to glucocorticoid-based treatment; however, patients with ICI are less likely to have ANA, which, if present, is low-titer and unlikely to have plasma cell inflammatory infiltrates histologically. Similar to de novo AIH, ICI hepatitis encompasses a spectrum of severity, and some cases may be refractory to standard therapeutics. Finally, occasionally, features of AIH overlap with features of autoimmune biliary disorders such as primary biliary cholangitis, primary sclerosing cholangitis **(Chaps. 355 and 357)**, or, even more rarely, mitochondrial antibody-negative autoimmune cholangitis. Such overlap syndromes are difficult to categorize, and often response to therapy may be the distinguishing factor that establishes the diagnosis.

TREATMENT

Autoimmune Hepatitis

The mainstay of management in AIH is glucocorticoid therapy, which is highly effective for the induction of remission. Several controlled clinical trials have documented that such therapy leads to symptomatic, clinical, biochemical, and histologic improvement as well as increased survival. A therapeutic response to glucocorticoids can be expected in up to 80% of patients, and recent consensus opinion defines response as a >50% reduction in ALT by 4 weeks of therapy; however, a dramatic reduction within days is uncommon after the initiation of corticosteroids. Although some advocate the use of prednisolone (the hepatic metabolite of prednisone), prednisone is just as effective and is favored by most authorities. A popular regimen in the United States relies on an initiation dose of 60 mg/d. This high dose is tapered successively over the course of a month down to a goal of 10–20 mg/d. In the European literature, successful treatment for AIH has been achieved with a prednisone starting dose of 20 mg. For patients with a mild presentation, budesonide, while inferior to prednisone in AIH, is an alternative first-line corticosteroid with limited systemic effects related to its extensive hepatic first-pass metabolism that achieves similar rates of biochemical response in patients with aminotransferase levels <2× the upper limit of normal.

After an early response to corticosteroids is observed, a steroid-sparing agent should be initiated to maintain remission. Azathioprine, considered first-line treatment for *nonicteric* or severe AIH, should be introduced 7–14 days after corticosteroids are started and a response is observed. Testing for thiopurine *S*-methyltransferase activity is recommended during this period prior to initiation of azathioprine. Dosing of azathioprine may be between 50 and 150 mg, and some experts titrate the dose upward after starting at 50 mg. Once on maintenance azathioprine (typically 1–2 mg/kg per day), prednisone is tapered further to a maintenance dose of 5–10 mg. If the patient achieves full biochemical remission, an attempt to taper prednisone completely is reasonable.

The advantage of this combination approach is a reduction, over the span of an 18-month course of therapy, in serious, life-threatening complications of steroid therapy (e.g., cushingoid features, hypertension, diabetes, osteoporosis) from 66% down to <20%.

Improvement of fatigue, anorexia, malaise, and jaundice tends to occur within days to several weeks; biochemical improvement occurs over the course of several weeks to months, with a fall in serum bilirubin and globulin levels and an increase in serum albumin. Serum aminotransferase levels usually drop promptly, but improvements in AST and ALT alone do not appear to be reliable markers of recovery in individual patients; histologic improvement, characterized by a decrease in mononuclear infiltration and in hepatocellular necrosis, may be delayed for 6–24 months. Still, if interpreted cautiously, aminotransferase levels, and immunoglobulin levels in some, are valuable indicators of relative disease activity, and although recommended, many authorities do *not* advocate for serial liver biopsies to assess therapeutic success or to guide decisions to alter or stop therapy. Therapy should continue for at least 24 months after biochemical remission. After tapering and cessation of therapy, the likelihood of relapse is at least 50%, and in one multicenter study, the rate of relapse was 81% after 3 years. In fact, most (95% of) patients will require treatment at maintenance doses indefinitely. Patients who relapse after completing one course of therapy almost always relapse after subsequent courses; because each relapse increases the risk of cirrhosis, death, and need for liver transplantation, ideally, patients who relapse once should be treated indefinitely.

If azathioprine is not tolerated because of adverse effects (such as marrow suppression, fatigue, or gastrointestinal symptoms), both 6-mercaptopurine (6MP) and mycophenolate mofetil (MMF) are considered second-line alternative agents; if azathioprine is ineffective, MMF (1.5–2.0 g/d in divided doses) is the next drug of choice. While MMF has yet to be compared to azathioprine in a randomized controlled clinical trial, in a recent prospective study, MMF was at least as effective, if not more so, than azathioprine in achieving biochemical remission and may be better tolerated. A downside of MMF that merits consideration is the risk of teratogenicity.

In cases of AIH refractory to standard treatment (azathioprine, 6-MP, or MMF, with or without prednisone) and in whom coexisting other causes for liver disease have been excluded, several additional treatments are options. Tacrolimus has been shown to be effective. Small studies of tacrolimus have shown a decline in aminotransferase levels in up to 80% after 3 months of therapy; however, tacrolimus requires careful trough monitoring and comes with a risk of nephrotoxicity. Treatment with monoclonal anti-CD20 (rituximab), described initially in a small pediatric case series, is another promising option that has been used with increasing frequency in refractory AIH. In one cohort study, rituximab given to patients with refractory AIH (1000-mg IV infusion given twice, 2 weeks apart) achieved remission rates of 95% at 1 year and 71% at 2 years and permitted substantial corticosteroid dose reduction.

When medical therapy fails, including in severe acute AIH refractory to high-dose corticosteroids or when AIH progresses to cirrhosis and end-stage liver disease, liver transplantation is the only recourse **(Chap. 356)**; in patients with severe AIH, failure of the bilirubin to improve after 2 weeks of therapy should prompt early consideration of the patient for liver transplantation. Recurrence of AIH in the new liver occurs rarely in most experiences but in as many as 35–40% of cases in others; nonetheless, 5-year patient and graft survival exceed 80%.

Like all patients with chronic liver disease, patients with AIH should be vaccinated against hepatitis A and B, ideally before immunosuppressive therapy is begun, if practical. Similarly, age- and gender-appropriate vaccinations, such as human papillomavirus, varicella-zoster virus, pneumococcal, and COVID-19, should be administered prior to immunosuppressive therapy. Patients with AIH and cirrhosis should be screened for HCC with ultrasound at 6-month intervals and may merit screening for varices with upper gastrointestinal endoscopy, based on the severity of liver disease **(Chap. 355)**.

■ FURTHER READING

AASLD/IDSA HCV Guidance Panel: Hepatitis C guidance 2019 update: American Association for the Study of Liver Diseases–Infectious Diseases Society of America recommendations for testing, managing, and treating hepatitis C virus infection. Hepatology 71:686, 2020. Updated regularly and available at *http://www.hcvguidelines.org*. Accessed April 20, 2020.

Asselah T, Rizzetto M: Hepatitis D virus infection. N Engl J Med 389:58, 2023.

Bhattacharya D et al: Hepatitis C guidance 2023 update: American Association for the Study of Liver Diseases-Infectious Diseases Society Recommendations for testing, managing, and treating hepatitis C virus infection. Clin Infect Dis 2023.

Buti M et al: Tenofovir alafenamide versus tenofovir disproxil fumarate for the treatment of HBeAg-negative chronic hepatitis B virus infection: A randomized, double-blind, phase 3 non-inferiority trial. Lance Gastroenterol Hepatol 1:196, 2017.

Carrat F et al: Clinical outcomes in patients with chronic hepatitis C after direct-acting antiviral treatments: A prospective cohort study. Lancet 393:1453, 2019.

Chan HLY et al: Tenofovir alafenamide versus tenofovir disproxil fumarate for the treatment of HBeAg-positive chronic hepatitis B

virus infection: A randomized, double-blind, phase 3 non-inferiority trial. Lancet Gastroenterol Hepatol 1:185, 2017.
Dalekos GN et al: Long-term results of mycophenolate mofeteil vs. azathioprine use in individuals in autoimmune hepatitis. J Hep Rep 4:100601, 2022.
Dougan M et al: AGA clinical practice update on diagnosis and management of immune checkpoint inhibitor colitis and hepatitis: Expert review. Gastroenterology 160:1384, 2021.
Dusheiko G et al: New approaches to chronic hepatitis B. N Engl J Med 388:55, 2023.
European Association for the Study of the Liver: EASL 2017 clinical practice guidelines on the management of hepatitis B virus infection. J Hepatol 67:370, 2017.
European Association for the Study of the Liver: EASL recommendations on treatment of hepatitis C: Final update of the series. J Hepatol 73:1170, 2020.
Forns X et al: Glecaprevir plus pibrentasvir for chronic hepatitis C virus genotype 1, 2, 4, 5, or 6 infection in adults with compensated cirrhosis (EXPEDITION-1): A single-arm, open-label, multicentre phase 3 trial. Lancet Infect Dis 17:1062, 2017.
Heller T et al: Hepatitis D: Looking back, looking forward, seeing the reward and the promise. Clin Gastroenterol Hepatol 21:2051, 2023.
Hirode G et al: Off-therapy response after nucleos(t)ide analogue withdrawal in patients with chronic hepatitis B: An international, multicenter, multiethnic cohort (RETRACT-B Study). Gastroenterology 162:757, 2022.
Jacobson IM et al: American Gastroenterological Association Institute clinical practice update-expert review: Care of patients who have achieved a sustained virologic response after antiviral therapy for chronic hepatitis C infection. Gastroenterology 152:1578, 2017.
Jeng W-J et al: Hepatitis B. Lancet 401:1039, 2023.
Kwo PY et al: Glecaprevir and pibrentasvir yield high response rates in patients with HCV genotype 1-6 without cirrhosis. J Hepatol 67:263, 2017.
Lampertico P et al: Hepatitis D virus infection: Pathophysiology, epidemiology and treatment. Report from the first international delta cure meeting 2022. JHEP Rep 5:100818, 2023.
Liem KS et al: Limited sustained response after stopping nucleos(t)ide analogues in patients with chronic hepatitis B: Results from a randomized controlled trial (Toronto STOP study). Gut 68:2206, 2019.
Litwin AH et al: Patient-centred models of hepatitis C treatment for people who inject drugs: a multicentre, pragmatic randomized trial. Lancet Gastroenterol Hepatol 7:1112, 2022.
Lok ASG et al: Antiviral therapy for chronic hepatitis B viral infection in adults: A systematic review and meta-analysis. Hepatology 63:284, 2016.
Mack CL et al: Diagnosis and management of autoimmune hepatitis in adults and children: 2019 practice guidance and guidelines from the American Association for the Study of Liver Diseases. Hepatology 72:671, 2020.
Martinello M et al: Hepatitis C. Lancet 402:1085, 2023.
Mohareb AM et al: Clearance of hepatitis B e antigen in untreated chronic hepatitis B virus infection: A systematic review and meta-analysis. J Infect Dis 11:761, 2022.
Negro F, Lok AS: Hepatitis D: A review. JAMA 330:2376, 2023.
Papatheodoridis GV et al: DARING-B: Discontinuation of effective entecavir or tenofovir disoporxil fumarate long-term therapy before HBsAg loss in non-cirrhotic HBeAg-negative chronic hepatitis B. Antivir Ther 23:677, 2018.
Papatheodoridis GV et al: Eight-year survival in chronic hepatitis B patients under long-term entecavir or tenofovir is similar to the general population. J Hepatol 68:1129, 2018.
Pawlotsky J-M et al: From non-A, non-B hepatitis to hepatitis C virus cure. J Hepatol 62:S87, 2015.
Perrillo RP et al: American Gastroenterological Association Institute technical review on prevention and treatment of hepatitis B virus reactivation during immunosuppressive drug therapy. Gastroenterology 148:221, 2015.
Reddy KR et al: American Gastroenterological Association Institute guideline on the prevention and treatment of hepatitis B virus reactivation during immunosuppressive drug therapy. Gastroenterology 148:215, 2015.
Rizzetto M et al: The changing context of hepatitis D. J Hepatol 74:1200, 2021.
Romée JALM et al: An open-label randomised-controlled trial of azathioprine vs. mycophenolate mofetil for the induction of remission in treatment-naive autoimmune hepatitis. J Hepatol 80:576, 2024.
Rossi C et al: Sustained virologic response from interferon-based hepatitis C regimens is associated with reduced risk of extrahepatic manifestations. J Hepatol 71:1116, 2019.
Singal AG et al: AGA clinical practice update on interaction between oral direct-acting antivirals for chronic hepatitis C infection and hepatocellular carcinoma: Expert review. Gastroenterology 156:2149, 2019.
Sgamato C et al: Autoimmune liver disease and SARS-CoV-2. World J Gastroenterol 29:1838, 2023.
Terrault N et al: Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. Hepatology 67:1560, 2018.
Thomas DL: Global elimination of chronic hepatitis. N Engl J Med 380:2041, 2019.
Van Den Brand FF et al: Increased mortality among patients with vs without cirrhosis and autoimmune hepatitis. Clin Gastroenterol Hepatol 17:940, 2019.
Wedemeyer H et al: A phase 3, randomized trial of bulevirtide in chronic hepatitis D. N Engl J Med 389:22, 2023.
Yardeni D et al: Current best practice in hepatitis B management and understanding long-term prospects for cure. Gastroenterology 164:42, 2023.

353 Alcohol-Associated Liver Disease

Bernd Schnabl

Alcohol-associated liver diseases (ALD) comprise a spectrum of diseases associated with chronic alcohol consumption ranging from alcohol-associated fatty liver disease and steatohepatitis to more advanced liver disease including fibrosis and cirrhosis. Acute alcohol-associated hepatitis is an acute-on-chronic form of ALD that is associated with liver failure and high mortality.

EPIDEMIOLOGY

Approximately 7% of adults in the United States meet criteria for unhealthy drinking, defined as ≥2 drinks per day in women and ≥3 drinks per day in men, or partake in heavy episodic drinking, defined as ≥4 drinks for women and ≥5 drinks for men on a single occasion (1 drink equals ~14 g of ethanol, which is 1 beer, 4 oz of wine, or 1 oz of 80% spirits). The COVID-19 pandemic increased excessive alcohol intake, resulting in a rise in morbidity and mortality linked to ALD. Prevalence of ALD correlates with the amount of alcohol consumption in different regions. Prevalence of alcohol-associated fatty liver disease is 4.7% of the general population in the United States, and 1.5% has stage 2 or greater fibrosis. Liver cirrhosis is the eleventh leading cause of mortality worldwide, causing 1.16 million deaths annually; 48% of cases of cirrhosis can be attributed to alcohol. Among patients with alcohol misuse, 18% had fibrosis, 26% had cirrhosis, and 7% had acute alcohol-associated hepatitis without underlying cirrhosis. In the European population, the annual incidence rate for acute alcohol-associated hepatitis is between 24 and 27 per million persons in women and between 46 and 65 per million persons in men.

PATHOGENESIS

Alcohol in the form of ethanol is rapidly absorbed in the upper gastrointestinal tract and predominantly metabolized in the liver. Ethanol reaches the liver through the portal vein, and the majority of ethanol is oxidized via alcohol dehydrogenase 1 (ADH1) into acetaldehyde in hepatocytes. Chronic alcohol consumption induces the expression of a second ethanol-metabolizing enzyme, cytochrome P450 family 2 subfamily E member 1 (CYP2E1), which also converts ethanol into acetaldehyde. In addition to the direct cellular toxic effects of acetaldehyde, metabolism of ethanol into acetaldehyde causes the generation of reactive oxygen species (ROS), resulting in further injury of hepatocytes via lipid peroxidation and DNA damage. Acetaldehyde is then oxidized into acetate via acetaldehyde dehydrogenase (ALDH). Inherited deficiency of ALDH2 is common in Asian countries and leads to acetaldehyde accumulation after alcohol consumption. These individuals develop nausea and cutaneous flushing. Several mechanisms contribute to the development of hepatic steatosis related to alcohol consumption. Acetate is converted into acetyl-coenzyme A (CoA), which contributes to fatty acid and triglyceride synthesis. Alcohol, in part through epigenetic changes, increases the expression of genes involved in lipogenesis, while genes involved in fatty acid transport and oxidation are suppressed. Alcohol also increases the ratio of reduced nicotinamide adenine dinucleotide (NAD)/oxidized NAD ($NADH/NAD^+$) in hepatocytes, which further reduces mitochondrial β-oxidation. Alcohol can increase fatty acid mobilization in adipose tissue and the intestine, which will lead to hepatic accumulation of fatty acids and increased hepatic steatosis. Overall, the net effect of these processes contributes to accumulation of lipids in the liver.

RISK FACTORS FOR PROGRESSION OF ALD

Daily alcohol consumption or heavy drinking results in hepatic steatosis, but only 10–20% of such individuals will develop progressive liver disease and cirrhosis. Therefore, other cofactors such as behavioral, environmental, and genetic factors play important roles in progression of ALD **(Table 353-1)**. There is a dose-dependent increase, with regard to the amount of alcohol consumed, in the likelihood of developing liver cirrhosis. Women develop ALD at a lower daily alcohol intake. Cigarette smoking is an independent risk factor for alcohol-associated cirrhosis. The drinking pattern, in particular binge drinking and excessive alcohol drinking outside meals, increases the risk of developing progressive ALD. Obesity and metabolic dysfunction–associated steatotic liver disease (MASLD) are frequent cofactors contributing to progression of ALD. A distinct subset of patients within ALD, termed as metabolic ALD (MetALD), now describes patients with MASLD and increased alcohol consumption. Other chronic liver diseases such as viral hepatitis and hemochromatosis can have synergistic effects on ALD. Twin studies demonstrated a genetic predisposition to alcohol-associated liver cirrhosis that is independent from the genetic predisposition to alcohol use disorder. Gene polymorphisms conferring increased risk of alcohol-associated liver cirrhosis have been found in three genes, patatin-like phospholipase domain-containing 3 (*PNPLA3*), membrane bound O-acyltransferase domain-containing 7 (*MBOAT7*), and transmembrane 6 superfamily member 2 (*TM6SF2*), although the molecular mechanism is not well understood. Genetic variants in hydroxysteroid 17-beta dehydrogenase 13 (*HSD17B13*) and Fas-associated factor family member 2 (*FAF2*) are associated with a reduced risk of developing ALD. A subset of patients with alcohol use disorder develop changes in the gut microbiome and increased intestinal permeability. This leads to the passage of microbial components, like bacterial lipopolysaccharide (LPS), through the portal vein into the liver resulting in hepatic inflammation, hepatocyte death, and activation of fibrotic pathways. Ongoing fibrosis resulting from sustained alcohol consumption progresses to cirrhosis accompanied by portal hypertension. Impaired liver regeneration might contribute to the acute onset of alcohol-associated-hepatitis in patients with underlying cirrhosis **(Chap. 355)**.

TABLE 353-1 Factors for Progression of Alcohol-Associated Liver Disease

- Alcohol amount and duration
- Drinking pattern (drinking without meal, binge drinking)
- Genetic factors, especially *PNPLA3* polymorphism
- Female sex
- Smoking
- Obesity and metabolic dysfunction–associated steatotic liver disease (MASLD)
- Chronic liver diseases such as viral hepatitis and hemochromatosis
- Intestinal microbiota

TABLE 353-2 Symptoms and Signs Associated with Alcohol-Associated Cirrhosis and Alcohol-Associated Hepatitis

- Tiredness
- Malnutrition and sarcopenia
- Abdomen: abdominal discomfort, hepatomegaly, splenomegaly, caput medusae, ascites with weight gain, abdominal pain, and shortness of breath
- Skin: spider angioma, palmar erythema, jaundice, ecchymoses
- Eyes: icteric sclerae
- Hands: Dupuytren contracture
- Extremities: edema
- Face: rhinophyma
- Reproductive system: gynecomastia, gonadal atrophy, loss of libido, amenorrhea
- Neurologic:
  - Peripheral neuropathy
  - Alcohol withdrawal: tachycardia, agitation, tremor, seizures, delirium
  - Hepatic encephalopathy: asterixis (flapping tremor), forgetfulness, inversion of sleep/wake pattern, altered consciousness, confusion, lethargy, coma
  - Wernicke-Korsakoff syndrome

CLINICAL FEATURES

The development of alcohol-associated steatosis, steatohepatitis, and cirrhosis is most often clinically silent. Symptoms arise once the patient with alcohol-associated liver cirrhosis decompensates or develops alcohol-associated hepatitis **(Table 353-2)**. Patients with alcohol-associated hepatitis have been drinking heavily (>40 g/d for women and >50–60 g/d for men) for >6 months with <60 days of abstinence before onset of symptoms. They present with rapid onset of jaundice (serum bilirubin >3 mg/dL), often accompanied by fever, malaise, tender hepatomegaly, and clinical signs of hepatic decompensation, such as ascites, bacterial infection, variceal bleeding, and hepatic encephalopathy. Infections occur in 12–26% of patients with severe alcohol-associated hepatitis at the time of admission. Alcohol-associated hepatitis is often accompanied by systemic inflammatory response syndrome (SIRS) and acute kidney injury (AKI) secondary to hepatorenal syndrome.

LABORATORY FINDINGS

Patients with simple hepatic steatosis can present with normal liver function tests. Steatohepatitis is characterized by elevated levels of aspartate aminotransferase (AST) and γ-glutamyl transferase (GGT). Characteristic laboratory parameters for ALD include a ratio of AST to alanine aminotransferase (ALT) of >1, and serum AST is rarely >300 IU/L. Serum bilirubin and international normalized ratio (INR) are typically normal. Elevated bilirubin and INR and low serum albumin and platelet count are common laboratory finings in patients with cirrhosis. Patients with alcohol-associated hepatitis have AST and ALT elevations that do not exceed 400 IU/L, with AST/ALT ratio of >1.5 and serum bilirubin >3 mg/dL.

DIAGNOSIS

The Alcohol Use Disorders Inventory Test (AUDIT) is a validated tool for identifying patients with alcohol use disorder **(Chap. 464)**. Diagnosis of ALD requires exclusion of other liver diseases in heavy drinkers. Alcohol-associated steatosis can be diagnosed by simple ultrasound, magnetic resonance imaging (MRI), or computed tomography (CT). Noninvasive quantification of hepatic fat can be achieved with the ultrasound technique of controlled attenuation parameter (CAP) or with magnetic resonance proton density fat fraction (MR-PDFF). Liver biopsy is rarely indicated for diagnosing alcohol-associated hepatic steatosis or steatohepatitis. Liver biopsy typically shows hepatocytes with large lipid droplets (macrovesicular steatosis) around pericentral veins (zone 3). Morphologic features of alcohol-associated steatohepatitis include hepatocyte injury and ballooning with Mallory-Denk bodies, necrosis, and lobular inflammation with mononuclear and neutrophilic granulocytes.

Progression of alcohol-associated steatohepatitis to fibrosis can be diagnosed using liver stiffness measurement by techniques such as transient elastography (e.g., FibroScan). Liver stiffness <6 kPa indicates normal liver, whereas cutoffs for each stage of alcohol-associated liver fibrosis have been validated (>8 kPa indicates ≥F3 advanced fibrosis; >12.5 kPa indicates F4 cirrhosis). Fibrosis-4 (Fib-4) score is a serum marker test based on age, AST, ALT, and platelets and can be used to exclude advanced fibrosis or cirrhosis at a threshold value <3.25. Histology shows initially perivenular fibrosis with subsequent extension of collagen fibers into hepatic lobules, described as septal fibrosis. Patients with cirrhosis show liver nodularity on imaging with ultrasound, MRI, or CT scan. Radiologic signs of portal hypertension include ascites, splenomegaly, and portal-systemic collateral vessels. Prognosis and risk of mortality are assessed using Child-Pugh-Turcotte (CPT) or Model for End-Stage Liver Disease (MELD; or sodium-MELD) scores **(Chap. 355)**.

In patients presenting with features suggestive of alcohol-associated hepatitis, imaging is obtained to exclude biliary obstruction and hepatocellular carcinoma (HCC). In addition, other causes of liver disease such as viral hepatitis, Wilson's disease, and severe autoimmune liver disease should be ruled out. Histology shows macrovesicular steatosis, hepatocyte ballooning with Mallory-Denk bodies, megamitochondria, neutrophil infiltration, ductular reaction, bilirubinostasis, and chicken wire fibrosis. The majority of patients with alcohol-associated hepatitis have underlying cirrhosis (80%) **(Chap. 355)**, and 10–20% of patients with a clinical diagnosis of alcohol-associated hepatitis will have other liver diseases on biopsy. Therefore, in the presence of potential confounding factors, including possible ischemic hepatitis (in the setting of, e.g., hypotension, massive gastrointestinal bleeding, recent cocaine use, septic shock), drug-induced liver injury (DILI), autoimmune liver disease, uncertain alcohol use assessment, or atypical laboratory tests (AST <50 IU/L or >400 IU/L, AST/ALT ratio <1.5), a transjugular liver biopsy is recommended to confirm the diagnosis of alcohol-associated hepatitis. Infections need to be assessed routinely with chest x-ray and blood, urine, and ascites cultures in patients presenting with alcohol-associated hepatitis.

TREATMENT

Alcohol-Associated Liver Disease (Fig. 353-1)

To date, the most effective therapy to reduce the progression of and reverse ALD is prolonged alcohol abstinence. In particular, alcohol-associated hepatic steatosis and steatohepatitis are reversible with cessation of alcohol consumption. Thus, treatment of the underlying alcohol use disorder is an integral part for therapy of ALD. There are currently no approved drugs for treatment of alcohol-associated steatosis and steatohepatitis with or without fibrosis.

Patients with alcohol-associated cirrhosis and ongoing alcohol consumption are at risk for decompensation and development of hepatic encephalopathy, ascites, variceal bleeding, hepatorenal syndrome, and HCC **(Chap. 355)**. Patients with cirrhosis should undergo an upper gastrointestinal endoscopy to screen for varices. HCC screening is recommended using liver ultrasonography and serum α-fetoprotein (AFP) every 6 months in patients with cirrhosis. Management of complications of cirrhosis such as variceal bleeding, ascites, hepatic encephalopathy, and HCC does not differ from patients with cirrhosis due to a different etiology **(Chap. 355)**. Liver transplantation for patients with alcohol-associated decompensated cirrhosis or HCC is a definitive therapy and is currently the leading indication for liver transplantation in the United States. Liver transplantation evaluation should be taken into consideration for patients with end-stage liver disease **(Chap. 356)**.

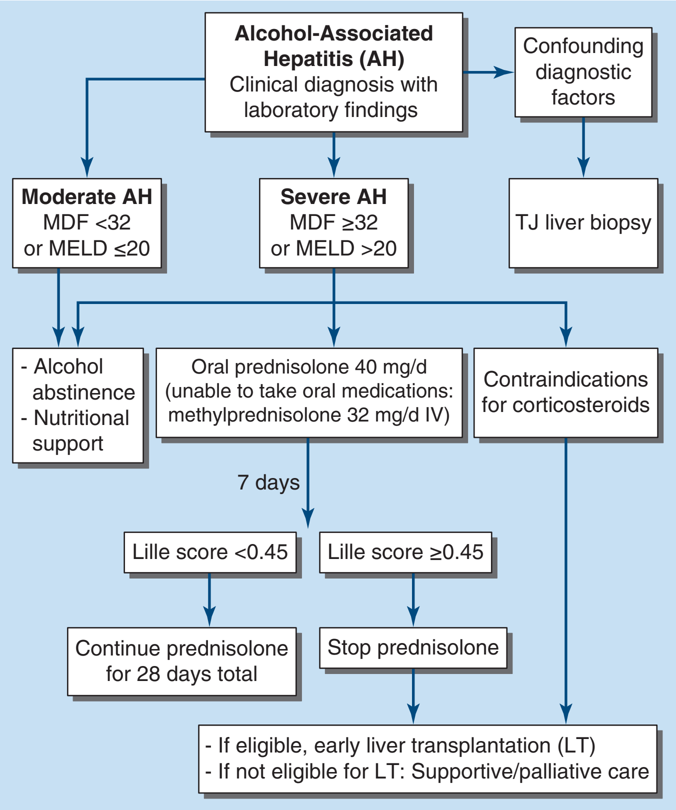
Alcohol-Associated Hepatitis (AH)
Clinical diagnosis with laboratory findings
Confounding diagnostic factors
Moderate AH
MDF <32
or MELD ≤20
Severe AH
MDF ≥32
or MELD >20
TJ liver biopsy
- Alcohol abstinence
- Nutritional support
Oral prednisolone 40 mg/d
(unable to take oral medications: methylprednisolone 32 mg/d IV)
Contraindications for corticosteroids
7 days
Lille score <0.45
Lille score ≥0.45
Continue prednisolone for 28 days total
Stop prednisolone
- If eligible, early liver transplantation (LT)
- If not eligible for LT: Supportive/palliative care

FIGURE 353-1 Treatment algorithm for alcohol-associated hepatitis. In patients with a clinical diagnosis of alcohol-associated hepatitis, confounding factors (see text) need to be ruled out, if necessary, by transjugular (TJ) liver biopsy. Patients with severe alcohol-associated hepatitis (AH), defined as Maddrey discriminant function (MDF) ≥32 or Model for End-Stage Liver Disease (MELD) score >20, without contraindications for glucocorticoids (see text) are candidates for such treatment. Nonresponders or patients with contraindications for treatment should be considered for early liver transplantation (LT) or supportive or palliative care, as clinically appropriate.

In patients diagnosed with alcohol-associated hepatitis (Fig. 353-1), short-term mortality can be predicted using the Maddrey discriminant function (MDF; calculated as 4.6 × [the prolongation of the prothrombin time above control {seconds}] + serum bilirubin [mg/dL]), MELD score **(Chap. 355)**, or age-bilirubin-INR-creatinine (ABIC) score. Patients with MDF <32 or MELD ≤20 are defined as having moderate alcohol-associated hepatitis. Currently, patients with moderate alcohol-associated hepatitis are treated under a multidisciplinary team including an alcohol use disorder specialist, dietitian for nutritional supplementation for patients with markedly reduced intake, and hepatologist for managing liver disease complications. Enteral nutrition with a goal of 35–40 kcal/kg and supplementation of micronutrients (in particular zinc) and vitamin supplementation (in particular vitamin B_1 and K) are recommended for patients with alcohol-associated hepatitis. Intravenous albumin is preferred for volume expansion. Patients with moderate alcohol-associated hepatitis have a 20% risk of 1-year mortality. MDF ≥32 or MELD >20 identifies patients with severe alcohol-associated hepatitis and high short-term mortality of ~30% at 3 months, who will have a survival benefit with glucocorticoid treatment. Contraindications for glucocorticoid treatment include uncontrolled infections or sepsis, AKI and hepatorenal syndrome, uncontrolled upper gastrointestinal bleeding, concomitant diseases (including viral hepatitis, HCC, pancreatitis, DILI, active tuberculosis, and HIV), multiorgan failure, and shock. Glucocorticoids can be used once infection, sepsis, and gastrointestinal bleeding are adequately controlled.

Glucocorticoid use reduces the risk of death in patients with severe alcohol-associated hepatitis within 28 days of treatment but not in the following 6 months. Oral prednisolone, 40 mg/d for a total duration of 4 weeks, is preferred. For patients unable to take oral medications, methylprednisolone, 32 mg/d IV, is used. The combination of glucocorticoids with *N*-acetylcysteine infusion might add short-term survival benefit at 1 month. Failure of improvement of Lille score (≥0.45) after 7 days of glucocorticoid treatment will determine patients with severe alcohol-associated hepatitis who will unlikely benefit from continued treatment with glucocorticoids. Glucocorticoids should be stopped in nonresponders, and early liver transplantation should be considered. Although short-term prognosis is dependent on liver disease severity at the time of presentation, long-term prognosis (>1 year) largely depends on alcohol abstinence and underlying cirrhosis. Patients with severe alcohol-associated hepatitis that is nonresponsive to medical therapy have high 30-day mortality and are therefore unable to fulfill a minimum of 6 months of alcohol abstinence, which is required in many centers for liver transplantation evaluation. Early liver transplantation can be successfully performed in highly selected patients with an excellent psychosocial profile **(Chap. 356)**. If a nonresponder is ineligible for early liver transplantation, supportive or palliative care should be considered for patients with multiple-organ failure.

FURTHER READING

Bataller R et al: Alcohol-associated hepatitis. N Engl J Med 387:2436, 2022.

Crabb DW et al: Diagnosis and treatment of alcohol-associated liver diseases: 2019 practice guidance from the American Association for the Study of Liver Diseases. Hepatology 71:306, 2020.

Louvet A et al: Corticosteroids reduce risk of death within 28 days for patients with severe alcoholic hepatitis, compared with pentoxifylline or placebo-a meta-analysis of individual data from controlled trials. Gastroenterology 155:458, 2018.

Seitz HK et al: Alcoholic liver disease. Nat Rev Dis Primers 4:16, 2018.

Singal AK, Mathurin P: Diagnosis and treatment of alcohol-associated liver disease: A review. JAMA 326:165, 2021.

354

Metabolic Dysfunction–Associated Steatotic Liver Disease and Steatohepatitis

Manal F. Abdelmalek, Anna Mae Diehl

DISEASE NOMENCLATURE

Under the umbrella term of steatotic liver disease (SLD), metabolic dysfunction–associated steatotic liver disease (MASLD) and metabolic dysfunction–associated steatohepatitis (MASH) now replace the historical terms nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH). The new disease nomenclature emphasizes that SLD in people who do not drink alcohol excessively is strongly associated with metabolic dysfunction while minimizing stigma associated with the terms *nonalcoholic* and *fatty*. Accordingly, people who fulfill the new diagnostic criteria for MASLD/MASH have hepatic steatosis, drink little alcohol (<20 g/d for women and <30 g/d for men) and have one or more cardiometabolic risk factors (e.g., overweight/obesity, hypertension, dyslipidemia, insulin resistance/type 2 diabetes). In addition, a new subcategory of SLD has been generated, termed metabolic and alcohol-associated liver disease (MetALD), for SLD patients who have evidence of metabolic dysfunction but also consume social amounts of alcohol (20–50 g/d for woman; 30–60 g/d for men). Higher amounts of alcohol use, irrespective of the presence of metabolic risk factors, is defined as alcohol-associated liver disease (ALD) **(Chap. 353)**. Epidemiologic data demonstrate almost complete (at least 98%) overlap between the population historically defined as NAFLD/NASH and the population categorized as MASLD/MASH using the new diagnostic criteria.

INCIDENCE, PREVALENCE, AND NATURAL HISTORY

SLD is the most common cause of chronic liver disease in the United States and worldwide. The main causes of SLD are ALD and MASLD. The prevalence of MASLD is rising in parallel with increases in the prevalence of obesity and conditions associated with metabolic syndrome (i.e., insulin resistance, dyslipidemia, central obesity, and hypertension). Hence, MASLD is currently a leading indication for liver transplantation in the United States. While prevalence estimates of MASLD vary based on clinical setting, race/ethnicity, and geographic region, 25–30% of the general U.S. adult population is currently estimated to have MASLD, and at least 3–6% of the group with MASLD is thought to have MASH, i.e., a type of inflammatory liver cell injury that can develop in individuals with metabolic dysfunction–associated steatosis (MASL). Because MASH increases the rate of liver cell death, it can trigger liver scarring (i.e., fibrosis) and, over time, progress to cirrhosis. The prevalence of MASH is increasing as the population ages. For example, a recent study identified MASH in 14% of asymptomatic patients who were ≥50 years of age and undergoing colon cancer screening. With the rise in MASH, the prevalence of clinically significant hepatic fibrosis (hepatic fibrosis ≥ stage 2 [F2]) has increased more than twofold in the past two decades alone. Mirroring the projected rise of MASH, the proportion of patients with advanced hepatic fibrosis (bridging fibrosis [F3] or compensated cirrhosis [F4]) will also increase. Similar to cirrhosis caused by other liver diseases, cirrhosis caused by MASLD increases the risk of hepatic decompensation, hepatocellular carcinoma (HCC), and liver-related morbidity and mortality. In addition, MASLD may be a premalignant condition, increasing the risk for both primary liver cancers and extrahepatic malignancies, even in noncirrhotic livers. MASL, MASH, and MASH-related cirrhosis are not limited to adults. Indeed, these diseases are increasingly documented in children and adolescents. As in adults, weight gain and related insulin resistance are the main risk factors for pediatric MASLD. Thus, the rising incidence and prevalence of childhood obesity is a major contributor to projected future public health burden of liver disease.

Although MASLD typically occurs in the presence of overweight/obese status, it can occur in lean individuals and is particularly common in those with truncal obesity and a paucity of peripheral adipose depots (i.e., lipodystrophy). Ethnic and racial factors influence liver fat accumulation, with documented prevalence of MASLD being lowest in African Americans (~25%), highest in Hispanic and Asian-Indians (~50%), and intermediate in whites (~30%).

MASLD encompasses a spectrum of liver pathology **(Fig. 354-1)** and, thus, has a variable clinical prognosis. The accumulation of triglyceride within hepatocytes in the absence of overt liver injury or inflammation on liver biopsy (i.e., hepatic steatosis) is the most clinically benign extreme of the MASLD spectrum. Although steatotic livers can remain stable for years, they can also reversibly transition to a more clinically dangerous state that is characterized by accumulation of inflammatory cells and injured/dying steatotic hepatocytes (i.e., steatohepatitis, MASH). Like steatosis, MASH can remain stable, wax and wane, or intensify over time. Repeated or protracted bouts of MASH trigger regenerative responses that can become maladaptive, and this increases the risk for progressive liver fibrosis, as well as primary liver cancer. Hence, chronic steatohepatitis is the primary predictor for hepatic fibrosis in MASLD. Fibrosis is a biomarker for defective liver regeneration. As functional hepatic parenchyma is replaced by scar, fibrosis severity increases. Livers with MASLD variably progress from a state of no fibrosis (F0) through stages of minimal fibrosis (F1), significant fibrosis (F2), advanced fibrosis (F3), and ultimately, severe fibrosis/

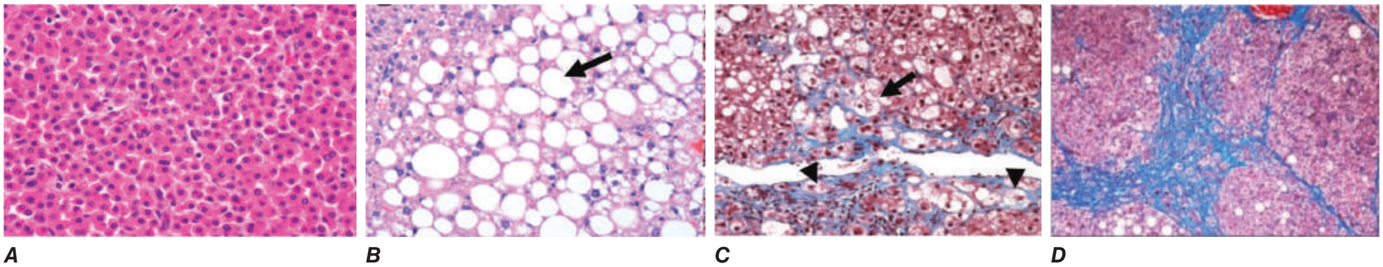

FIGURE 354-1 Histopathologic spectrum of metabolic dysfunction–associated steatotic liver disease (MASLD). MASLD encompasses a dynamic spectrum of liver pathology. ***A.*** Healthy liver. ***B.*** Simple steatosis (*arrow* shows fatty hepatocyte). ***C.*** Metabolic dysfunction–associated steatohepatitis (MASH); note ballooned hepatocyte (*arrow*) near central vein with adjacent blue-stained pericellular fibrosis (arrowheads). ***D.*** Cirrhosis with blue-stained bridging fibrosis surrounding micronodules of liver parenchyma.

cirrhosis (F4). Although cirrhosis is on the opposite end of the clinical spectrum of MASLD, it is important to emphasize that it can regress. Moreover, MASLD patients generally do not manifest symptoms or signs specific for liver disease until cirrhosis has been present for years. Thus, cirrhosis itself is stratified into compensated cirrhosis and decompensated cirrhosis. Both the risk for developing symptoms and signs of liver disease (i.e., decompensating) and the risk for primary liver cancer increase with the duration of cirrhosis.

Fibrosis severity is the primary predictor of liver-related morbidity and mortality in MASLD. As noted next, MASH increases the risk for fibrosis progression. However, the rate at which fibrosis progresses is influenced by many factors, including presence and severity of comorbid diseases (e.g., diabetes mellitus), genetics (e.g., polymorphisms of certain lipid-regulating genes), and environmental factors (e.g., high-fat, high-sugar diets). Meta-analyses of cohorts with longitudinal paired liver biopsies demonstrate an average fibrosis progression rate of one stage every 14 years in people with steatosis, as compared to one stage every 7 years for patients with MASH. Once fibrosis has progressed to stage F2 or higher, patients with MASH are at increased risk for further fibrosis progression and, thus, eventual bad clinical liver outcomes. Therefore, patients with MASH and fibrosis stage ≥F2 have been designated as having "at-risk MASH," a diagnosis that triggers consideration of more regular and intensive monitoring and therapeutic interventions to improve/resolve MASH and hepatic fibrosis. The primary goal of treatment is to prevent fibrosis progression and thereby decrease liver-related and all-cause morbidity and mortality. Among patients with cirrhosis, progression to clinical decompensation ranges from 3 to 20% per year. In addition, once MASLD-related cirrhosis develops, the annual incidence of primary liver cancer can be as high as 1–2% per year. Therefore, the diagnosis of cirrhosis changes clinical management to incorporate biannual screening for primary liver cancer, screening for esophageal varices in accordance to the Baveno criteria, and monitoring for signs or symptoms of hepatic decompensation **(Chap. 355)**.

■ MOLECULAR AND CELLUAR PATHOGENESIS

The presence and severity of MASL and MASH are governed by imbalances between energy intake and utilization. Excessive energy (i.e., food) intake, coupled with reduced energy expenditure (i.e., physical activity) and a limited capacity for expansion of peripheral adipose depots to accommodate energy surplus, contribute to development of visceral adiposity, insulin resistance, and metabolic disease.

Steatosis Steatotic hepatocytes are characterized by accumulation of triglyceride. Hepatocytes synthesize triglyceride to store fatty acids, and thus, triglycerides accumulate with increases in hepatocyte exposure to free fatty acids that derive from (1) uptake of dietary fatty acids, (2) release of fatty acids from adipose depots, and/or (3) de novo lipogenesis (DNL) within hepatocytes themselves. Fatty acids can also accumulate in hepatocytes when mechanisms that convert/consume fatty acids are impaired because this increases the pool of fatty acids for triglyceride synthesis. Finally, triglycerides accumulate in hepatocytes when export of triglyceride is impaired. Triglycerides provide a relatively safe way for hepatocytes to store lipids, and thus, triglyceride accumulation per se (i.e., steatosis) does not damage the liver. However, steatosis is a biomarker of metabolic dysfunction, and left unchecked, the latter can drive other events that promote lipotoxicity.

Steatohepatitis Lipotoxicity (i.e., lipid-associated injury) distinguishes steatohepatitis from steatosis. Lipotoxicity results from dysregulated processing of fatty acids and/or other precursors/products of lipids that are stored with triglyceride in lipid droplets, leading to accumulation of toxic intermediates and/or by-products. The onset of lipotoxic stress has variable outcomes among individuals and within a given individual over the life span because multiple dynamic factors interact to determine hepatocyte resilience (i.e., the capacity to adapt metabolism in order to prevent, withstand, or repair damage to critical cellular components). Efficient resolution of lipotoxic stress permits recovery from steatohepatitis. However, adaptations that merely enable hepatocytes to survive despite ongoing lipotoxic stress change both the signals they release (e.g., hepatokines, metabolites) and how they respond to signals from other cells (e.g., adipokines, myokines, hormones, neurotransmitters), eventually altering liver morphology and function, as well as systemic metabolic homeostasis. Hepatocytes that succumb to lipotoxicity during steatohepatitis must be replaced, and this triggers regenerative responses that impose additional challenges to energy/metabolic homeostasis. When regenerative responses are defective, functional hepatic parenchyma is progressively replaced by fibrous scar, leading to cirrhosis and eventual hepatic decompensation. Another potentially lethal outcome of maladaptive repair is liver cancer. The risk for cirrhosis and liver cancer increases with the severity and duration of steatohepatitis (i.e., hepatic lipotoxicity).

Dietary sugars, such as fructose, drive DNL and, thus, increase the risk for steatosis, as do diets enriched with saturated versus unsaturated fat. Insulin resistance in peripheral tissues promotes the release of free fatty acids from adipocytes (lipolysis), increasing the substrate for hepatic triglyceride production. Factors that govern energy balance (i.e., fasting and/or frequency and intensity of exercise, activation of brown adipose tissue to energy-consuming thermogenic phenotype, and counterregulatory mechanisms that diminish energy disposal in response to decreased calorie intake) also determine susceptibility to hepatic steatosis. Genetic factors that impact hepatocyte lipid handling (i.e., I148M polymorphism of *PNPLA3*, which impairs lipolysis of triglyceride lipid droplets) and polymorphisms in other genes that play a role in hepatocyte fat metabolism (*TM6SF2*, which influences cholesterol metabolism, and *MBOAT7*, which influences phospholipid metabolism) also influence risk for hepatic steatosis. Recently, loss-of-function variants in *HSD17B13*, a gene that encodes an enzyme that localizes to lipid droplets in hepatocytes, appear to be protective against MASH, progressive fibrosis, and liver cancer. Additional factors including hepatocyte uric acid production, oxidative stress, and cytokine-mediated inflammatory response contribute to the development of MASH. Alterations in gut microbiota that accompany obesity enhance both energy harvest from dietary sources and increase intestinal permeability. Reduced intestinal barrier function increases hepatic exposure to gut-derived products, which stimulate liver cells to generate inflammatory mediators that worsen insulin resistance. Obese adipose depots also produce cytokines (adipokines) that inhibit tissue insulin sensitivity. Insulin resistance promotes hyperglycemia, which drives the pancreas to produce more insulin to maintain glucose homeostasis. However, hyperinsulinemia promotes lipid uptake, fat synthesis, and fat storage.

Because the severity and duration of lipotoxic liver injury dictate the intensity and duration of repair, the histologic features and outcomes of MASH are variable. Cirrhosis results from futile repair (i.e., progressive accumulation of wound healing cells, fibrous matrix, and abnormal vasculature [scarring]) rather than efficient reconstruction or regeneration of healthy hepatic parenchyma. Primary liver cancers develop when malignantly transformed liver cells escape mechanisms that normally control regenerative growth. The mechanisms responsible for futile repair (cirrhosis) and liver carcinogenesis are not well understood. Because normal liver regeneration is a very complex process, there are multiple opportunities for deregulation and, thus, pathogenic heterogeneity. To date, this heterogeneity has confounded development of both diagnostic tests and treatments for defective/deregulated liver repair (i.e., cirrhosis and cancer). Hence, current strategies focus on circumventing misrepair by preventing and/or reducing lipotoxic liver injury. Human epidemiologic studies have identified factors that modify susceptibility to steatosis, steatohepatitis, and determinants of regenerative capacity and, thus, cirrhosis and liver cancer (see "Disease Modifiers and Risk Factors").

DISEASE MODIFIERS AND RISK FACTORS

Alcohol A diagnosis of MASLD requires demonstration of increased liver fat in an individual who drinks little to no alcohol (i.e., ≤20 g/d for women; ≤30 g/d for men). Epidemiologic studies demonstrate that serum aminotransferase elevations increase as alcohol consumption increases beyond 10 g/d (equivalent to one can of beer, 4 oz of wine, or 1.5 oz [one shot] of distilled spirits). Therefore, thresholds for potentially dangerous alcohol ingestion have been set at more than two drinks (20 g) per day in women and three drinks (30 g) per day in men. Because the combination of metabolic risk factors and higher levels of alcohol consumption may confer higher risk for liver damage and hepatic fibrosis than MASLD alone, a new diagnostic category, MetALD, was recently created to enable more personalized management of these individuals. Hazardous alcohol consumption (≥50 g/d for women; ≥60 g/d for men) is defined as ALD irrespective of the presence of hepatic steatosis and/or metabolic syndrome.

Overweight/Obesity Although MASLD is commonly associated with overweight/obese status, it can occur in lean (body mass index [BMI] <25 kg/m^2 or <23 kg/m^2) individuals. This is particularly common in individuals of Asian ethnicity. Thus, the prevalence of lean MASLD varies geographically (e.g., as low 4% in the United States to as high as 19% in Asia). Regardless of ethnicity, patients with lean MASLD have increased visceral adiposity, insulin resistance, metabolic comorbidities, alterations in gut microbiota contributing to metabolic dysfunction, and/or decreased muscle mass due to sedentary lifestyle.

Metabolic Syndrome The diagnosis of MASLD is closely linked to insulin resistance and the metabolic syndrome **(Chap. 420)**, but MASLD may precede the development of metabolic syndrome components (i.e., truncal obesity, type 2 diabetes mellitus [T2DM], dyslipidemia, hypertension). Android body fat distribution, characterized by increased truncal subcutaneous and visceral fat depots (as opposed to gynoid body fat distribution predominantly in the hips or buttocks), infers an increased risk for MASLD/MASH. T2DM is the strongest risk factor for MASLD acquisition, fibrosis progression, and HCC. The prevalence of MASLD in patients with T2DM ranges from 30 to 75% and conveys a nearly twofold increased risk for NASH with fibrosis compared to patients without impaired glycemic control. Further, the probability of advanced hepatic fibrosis increases with the duration and severity of T2DM. The relationship between MASLD/MASH and T2DM is bidirectional. MASLD is associated with a reduction in insulin sensitivity, increasing the risk for incident T2DM by two- to fivefold. Hence, patients with MASLD should be screened for T2DM. As liver disease progresses, insulin resistance worsens, making control of T2DM more challenging. Effective management of insulin resistance and hyperglycemia may decrease risk for MASH and progressive hepatic fibrosis. In patients without T2DM, MASLD is related to coexistent metabolic risk factors (i.e., increased BMI, hypertriglyceridemia, truncal obesity, hyperinsulinemia). Hypertension (another feature of metabolic syndrome) adds to the risk MASH and fibrosis progression. Whether hypertension per se promotes the development of MASH (or the inverse) or is a consequence of underlying metabolic drivers for MASLD has not been well established. Plasma lipid abnormalities are twice as prevalent in patients with MASLD than in those without MASLD. In addition, the serum lipid subfractions are more atherogenic, further increasing the cardiovascular risk for patients with MASLD/MASH.

Cardiovascular Disease and Obstructive Sleep Apnea Both cardiovascular disease (CVD) and obstructive sleep apnea (OSA) are associated with MASLD, and several studies suggest that the presence of these conditions is associated with MASH and hepatic fibrosis. Given the strong association between MASLD, CVD, and OSA, patients who are known to have these conditions should be screened for MASLD and those who are overweight or obese with metabolic syndrome should be considered at high risk for OSA and CVD.

Endocrine Dysfunction In addition to its strong association with obesity and other metabolic risk factors, higher rates of MASLD have been reported in patients with hypothyroidism, hypogonadism, growth hormone (GH) deficiency, and polycystic ovarian syndrome (PCOS). The presence of MASLD is higher in postmenopausal women compared with premenopausal women, suggesting a role of higher free testosterone levels and/or lower estradiol level in MASH. Likewise, in PCOS, hyperinsulinemia and excess androgen production contribute to a two- to fourfold increase in prevalence of MASLD, MASH with hepatic fibrosis, and T2DM.

Heritable Factors • **GENETICS** Heritable factors clearly impact susceptibility to hepatic steatosis, MASH, liver fibrosis, and liver cancer. For example, genetic variants on or near *TM6SF2* or *MBOAT7* (genes involved in lipid homeostasis) and palatin-like phospholipase domain-containing 3 gene (*PNPLA3*; a gene that encodes an enzyme involved in intracellular trafficking of lipids) increase the heritability of MASLD. A recent meta-analysis showed that *PNPLA3* exerts a strong influence not only on hepatic fat accumulation but also on the severity of MASH and liver fibrosis. Indeed, twin studies suggest that inheritance accounts for about half the risk for developing cirrhosis.

EPIGENETIC FACTORS Heritable traits that do not result from direct changes in DNA may also influence MASLD pathogenesis and/or progression based on evidence that intrauterine exposures influence susceptibility to obesity and the metabolic syndrome in adolescence. Studies of families with adult-onset obesity have identified genome-wide epigenetic alterations that dysregulate metabolic pathways controlling adiposity, insulin sensitivity, and tissue generation or regeneration. First-degree relatives of probands with MASH cirrhosis have a 12-fold higher risk of advanced fibrosis. Furthermore, the risk of MASLD and advanced fibrosis may be increased among nonrelated household members, likely because of related similar environmental risk factors, lifestyle patterns, and gut microbiota. Whether such epigenetic mechanisms influence susceptibility to MASH and cirrhosis is being investigated.

DIAGNOSIS

Patients with MASLD usually come to medical attention because of incidentally detected abnormalities on abdominal imaging (e.g., steatosis on sonogram or cross-sectional imaging) and/or lab tests (e.g., increased alanine aminotransferase [ALT]). MASLD should be suspected when steatosis is present in the context of metabolic dysfunction (i.e., at least one out of five criteria of metabolic syndrome is present) and the patient drinks little to no alcohol. Confirming MASLD as the diagnosis requires further testing to exclude alternative causes of chronic hepatitis (e.g., viral hepatitis, autoimmune liver disease, iron or copper overload, α_1-antitrypsin deficiency). It is important to note that normal liver enzymes do not exclude MASLD. Also, so-called "normal" ranges for aminotransferases are inflated because they include values

TABLE 354-1 Alternative Causes of Hepatic Steatosis

- Alcohol-related liver disease
- Hepatitis C (particularly genotype 3)
- Inborn errors of metabolism
  - Abetalipoproteinemia
  - Cholesterol ester storage disease
  - Galactosemia
  - Glycogen storage disease
  - Hereditary fructose intolerance
  - Homocystinuria
  - Systemic carnitine deficiency
  - Tyrosinemia
  - Weber-Christian syndrome
  - Wilson's disease
  - Wolman's disease
- Medications **(see Table 354-2)**
- Miscellaneous
  - Industrial exposure to petrochemical
  - Inflammatory bowel disease
  - Lipodystrophy
  - Bacterial overgrowth
  - Starvation
  - Parenteral nutrition
- Surgical procedures
  - Bilopancreatic diversion
  - Extensive small-bowel resection
  - Gastric bypass
  - Jejunoileal bypass
- Reye's syndrome
- Acute fatty liver of pregnancy
- HELLP syndrome (*h*emolytic anemia, *e*levated *l*iver enzymes, *l*ow *p*latelet count)

in people with undiagnosed liver disease. A true normal ALT ranges from 29 to 33 U/L in men and from 19 to 25 U/L in women. Initial evaluation of elevated ALT should include screening for metabolic risk factors, assessment of alcohol intake, and exclusion of other causes of liver disease, as well as a physical examination to identify signs of insulin resistance, metabolic syndrome, and advanced liver disease. In the absence of metabolic syndrome or in the presence of additional signs or symptoms (e.g., diarrhea, iron deficiency anemia, marked elevation of lipoproteins), less common causes of hepatic steatosis or steatohepatitis should be considered in specific clinical contexts **(Table 354-1)**. Medication- or drug-associated steatosis or steatohepatitis **(Table 354-2)** should be excluded with a detailed history.

It is important to emphasize that the liver may not be enlarged and symptoms/signs of liver disease may be absent in patients with MASLD. Further, like serum aminotransferases, other liver tests (e.g., alkaline phosphatase, bilirubin, albumin, prothrombin time) may be completely normal. Because there is yet no one specific blood test for MASLD/MASH, confidence in the diagnosis of MASLD is rendered by identification of a metabolic risk factor(s) in the context of hepatic steatosis. The latter include increased BMI, insulin resistance/T2DM, and other parameters indicative of the metabolic syndrome (e.g., systemic hypertension, dyslipidemia, hyperuricemia/gout, cardiovascular disease) in the patient or first-degree family members. Individuals who have, or have had, pituitary or hypothalamic neoplasms and women with PCOS are also at increased risk for MASLD. Hypothyroidism and OSA may also increase MASLD, presumably by promoting obesity and/or exacerbating the metabolic syndrome.

CLINICAL FEATURES OF MASLD

Most subjects with MASLD are asymptomatic. The diagnosis is often made when abnormal liver aminotransferases or features of fatty liver are noted during an evaluation performed for other reasons. MASLD may also be diagnosed during the workup of vague right upper quadrant abdominal pain, hepatomegaly, or an abnormal-appearing liver at time of abdominal surgery. Obesity is present in 50–90% of subjects. Most patients with MASLD also have other features of the metabolic syndrome. Some have subtle stigmata of chronic liver disease, such as spider angiomata, palmer erythema, or splenomegaly. In a small minority of patients with advanced MASLD, complications of end-stage liver disease (e.g., jaundice, features of portal hypertension such as ascites or variceal hemorrhage) may be the initial findings.

The association of MASLD with obesity, T2DM, hypertriglyceridemia, hypertension, and CVD is well known. Other associations include chronic fatigue, mood alterations, OSA, thyroid dysfunction,

TABLE 354-2 Medications Associated with Hepatic Steatosis and Steatohepatitis

5-Fluorouracil	Macrovesicular steatosis	Unknown
Androgens	Macrovesicular steatosis	Increased hepatic lipid accumulation, impaired glycemic control
Amiodarone	Macrovesicular steatosis, steatohepatitis, and fibrosis	Decreases β-oxidation of fatty acids; increases lipid peroxidation, ROS generation, and lipogenesis
Aspirin (Reye's syndrome) Arrest B oxidation by consuming CoA; increase mi-tochondrial permeability by MPT pore formation	Microvesicular steatosis	Competitively inhibit the enzyme long chain 3-hydroxyacyl-CoA dehydrogenase (LCHAD) and β-oxidation of fatty acids; ATP depletion and loss of mitochrondrial membrane potential.
Cocaine	Microvesicular steatosis	Undergoes n-oxidation to produce hepatotoxic products and inhibit β-oxidation
-cycline antibiotics (i.e., doxycycline, minocycline tetracycline	Microvesicular steatosis	Inhibit β-oxidation of fatty acids and VLDL secretion
Fluoxetine	Macrovesicular steatosis	Increased lipid accumulation and intracellular serotonin content
Glucocorticoids	Macrovesicular steatosis	Glucose intolerance and weight gain, inhibit mitochondrial β-oxidation, decrease hepatic triglyceride secretion, and induce lipid peroxidation
L-Asparaginase	Microvesicular steatosis	Inhibition of protein synthesis and export of lipoproteins and lipids
NSAIDs	Microvesicular steatosis	Inhibit β-oxidation of short- and medium-chain fatty acids
Irinotecan	Macrovesicular steatosis	Unknown
Methotrexate	Macrovesicular steatosis, steatohepatitis, fibrosis	Inhibits mitochondrial ETC
NRTIs (i.e., zidovudine, didanosine)	Microvesicular steatosis	Mitochondrial dysfunction by mtDNA depletion and stimulation of autophagy
Pentochemical (i.e., vinyl chloride)	Microvesicular steatosis and steatohepatitis	Unknown
Tamoxifen	Macrovesicular steatosis and steatohepatitis	Promotes de novo fatty acid synthesis and inhibits fatty acid β-oxidation
Valproic acid	Microvesicular steatosis	Inhibit CPT-1 activity, sequestering CoA, arresting ETC and ATP synthesis

Abbreviations: ATP, adenosine triphosphate; CoA, coenzyme A; CPT, carnitine palmitoyl shuttle; ETC, electron transport chain; NRTIs, nucleoside reverse transcriptase inhibitors; NSAIDs, nonsteroidal anti-inflammatory drugs; ROS, reactive oxygen species; VLDL, very-low-density lipoprotein.

PCOS, and chronic pain syndrome. MASLD is an independent risk factor for metabolic syndrome **(Chap. 420)** and T2DM **(Chap. 415)**. Longitudinal studies suggest that patients with MASH are at two- to threefold increased risk for the development of metabolic syndrome. Similarly, studies have shown that patients with MASH have a higher risk for the development of hypertension and T2DM. The presence of MASLD is also independently associated with endothelial dysfunction, increased carotid intimal thickness, and the number of plaques in carotid and coronary arteries. Such data indicate that MASLD has many deleterious effects on health in general, including patient-reported quality of life.

■ STAGING AND RISK STRATIFICATION

Establishing the severity of MASLD-related liver injury and related scarring (i.e., staging MASLD) is more difficult than simply diagnosing MASLD. Staging is critically important because staging is necessary to define prognosis and thereby determine treatment recommendations. The goal of staging is to distinguish patients with steatohepatitis from those with steatosis and to identify which of the MASH patients have advanced fibrosis. The 10-year probability of developing liver-related morbidity or mortality is negligible in individuals with steatosis alone, and hence, this subgroup of MASLD patients tends to be managed conservatively (see below). In contrast, more intensive follow-up and therapy are justified in MASH patients, and the subgroup with advanced fibrosis merits the most intensive scrutiny and intervention because their 10-year risk of liver-related morbidity and mortality is increased. Staging approaches can be separated into noninvasive testing (i.e., physical examination, blood testing, and imaging) and invasive approaches (i.e., liver biopsy). Blood test evidence of liver inflammation (e.g., aspartate aminotransferase [AST], ALT) and hepatic dysfunction (e.g., hyperbilirubinemia, hypoalbuminemia, prothrombin time prolongation) or portal hypertension (e.g., thrombocytopenia) and stigmata of portal hypertension on physical examination or imaging studies (e.g., spider angiomata, palmer erythema, splenomegaly, ascites, clubbing, encephalopathy) suggest the diagnosis of advanced MASLD. While liver biopsy has historically been the "gold standard" for establishing the severity of liver injury and fibrosis, the increased sensitivity of new noninvasive approaches for staging liver disease severity has decreased the frequency with which liver biopsy is utilized to guide patient management. Although liver biopsy is safe and seldom complicated by serious adverse sequelae (i.e., bleeding, pain, or inadvertent puncture of other organs), it is impractical for population-based risk assessment. Further, liver biopsy suffers from potential sampling and inter- and intraobserver variability, particularly when tissue cores are suboptimal (<2 cm in length) in quality. Also, examination of tissue at a single point in time is a static depiction of disease activity that cannot otherwise define pathologic process of disease progression or regression. Serial liver biopsies within short intervals of time are generally deemed unacceptable outside of research studies for which MASH resolution and/or fibrosis regression are surrogate endpoints to clinically meaningful outcomes.

Although general population screening for MASLD is not recommended, targeted screening and risk stratification for high-risk groups are advised. Such high-risk groups include those with T2DM, obesity with metabolic complications, a family history of cirrhosis, or significant alcohol use. Clinical practice has been advanced by the emergence of an armamentarium of noninvasive tests (NITs) as surrogates to liver biopsy for screening and risk stratification of high-risk cohorts of patients. Screening and risk stratification in these high-risk groups are a cost-effective approach to identifying those with asymptomatic but clinically significant fibrosis. Early detection allows for interventions (i.e., lifestyle modification, pharmacologic and/or endobariatric or surgical interventions) that may prevent fibrosis progression and, thereby, potentially decrease both liver-related and all-cause mortality. Since the majority of patients with MASLD/MASH are asymptomatic, the primary goal of risk stratification using NITs is to identify patients who are not likely to have advanced fibrosis. NITs can be used singularly or sequentially to optimize predictive value and consist of a simple calculator-based risk score or blood-based or imaging biomarkers **(Fig. 354-2)**. Screening with a Fibrosis-4 (FIB-4) score, a score derived from available clinical and laboratory data, is cost-effective. FIB-4 is calculated using a simple algorithm based on age, ALT, AST, and platelet count and outperforms other calculations (i.e., AST to platelet ratio index, BARD score, NAFLD fibrosis score) in its ability to identify patients with a low probability of advanced fibrosis.

The FIB-4 score is well validated and has an excellent negative predictive value for exclusion of advanced hepatic fibrosis (i.e., low risk,

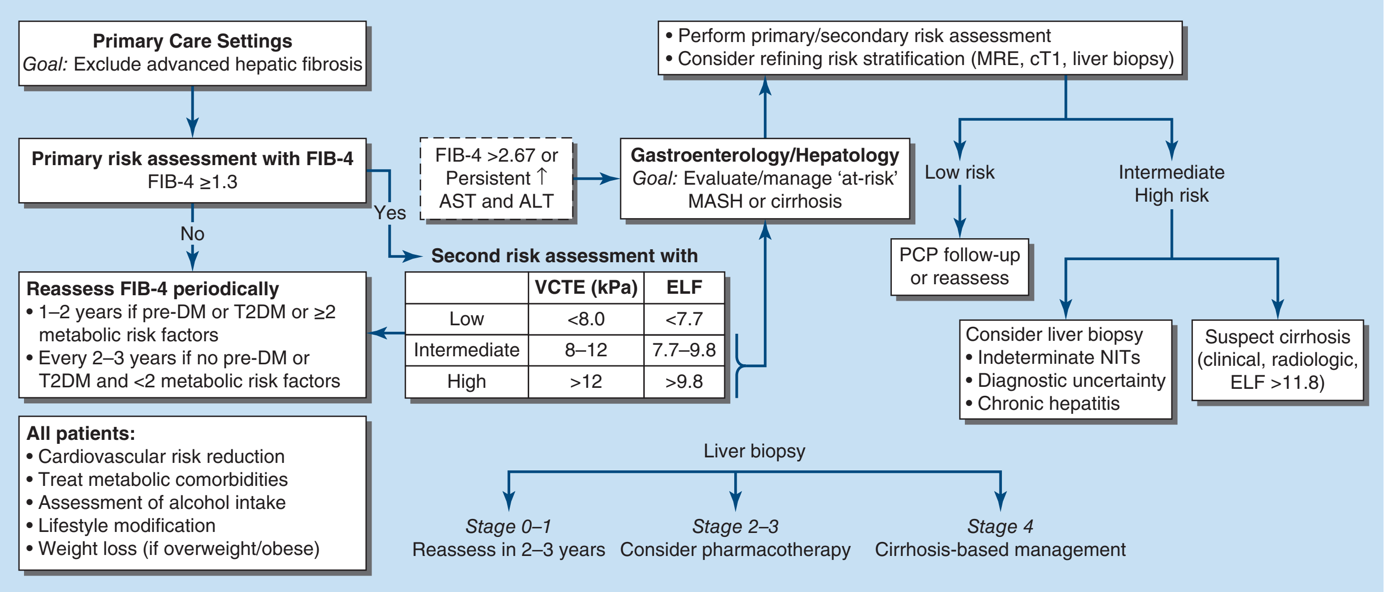
Primary Care Settings
Goal: Exclude advanced hepatic fibrosis
Primary risk assessment with FIB-4
FIB-4 ≥1.3
No
Yes
FIB-4 >2.67 or
Persistent ↑
AST and ALT
Gastroenterology/Hepatology
Goal: Evaluate/manage 'at-risk' MASH or cirrhosis
• Perform primary/secondary risk assessment
• Consider refining risk stratification (MRE, cT1, liver biopsy)
Low risk
Intermediate
High risk
PCP follow-up
or reassess
Second risk assessment with
VCTE (kPa)
ELF
Low
<8.0
<7.7
Intermediate
8–12
7.7–9.8
High
>12
>9.8
Reassess FIB-4 periodically
• 1–2 years if pre-DM or T2DM or ≥2 metabolic risk factors
• Every 2–3 years if no pre-DM or T2DM and <2 metabolic risk factors
Consider liver biopsy
• Indeterminate NITs
• Diagnostic uncertainty
• Chronic hepatitis
Suspect cirrhosis
(clinical, radiologic, ELF >11.8)
All patients:
• Cardiovascular risk reduction
• Treat metabolic comorbidities
• Assessment of alcohol intake
• Lifestyle modification
• Weight loss (if overweight/obese)
Liver biopsy
Stage 0–1
Reassess in 2–3 years
Stage 2–3
Consider pharmacotherapy
Stage 4
Cirrhosis-based management

FIGURE 354-2 Algorithm for the evaluation of patients at risk for or with established metabolic dysfunction–associated steatotic liver disease (MASLD) across practice settings. Patients with suspected MASLD based on imaging, metabolic risk factors, or evaluation of liver enzymes should be risk stratified using noninvasive tests for advanced hepatic fibrosis. Empasis in the primary care setting is the exclusion of advanced hepatic fibrosis using a test with high negative predictive value. When the Fibrosis-4 index (FIB-4) is <1.3, patients can be followed in the primary care setting and reassessed every 2–3 years. Patients with metabolic syndrome, prediabetes (pre-DM), or diabetes (T2DM) are at higher risk for fibrosis progression and can be followed with FIB-4 monitoring every 1–2 years. Due to low accuracy in patients older than 65 years and younger than 35 years, a secondary assessment should be done (i.e., elastography for liver stiffness measure or Enhanced Liver Fibrosis ELF]). Referral to a specialist should be considered for persistent elevation of liver enzymes, clinical suspicion of more advanced liver disease, and/or indeterminate or discordant noninvasive tests (NITs). ALT, alanine aminotransferase; AST, aspartate aminotransferase; MASH, metabolic dysfunction–associated steatohepatitis; MRE, magnetic resonance elastography; PCP, primary care physician; VCTE, vibration-controlled elastography.

FIB-4 <1.3). Patients with a low FIB-4 score of <1.3 can be managed in primary care; however, patients with two or more metabolic risk factors, particularly those with prediabetes or T2DM, are at increased risk for fibrosis progression with time and should have repeat assessment of FIB-4 score every 1–2 years with general health assessments. A change in FIB-4 score over time may allow for the prediction of progression to cirrhosis or hepatic decompensation, although the strength as a predictive biomarker is less robust in patients with T2DM.

Patients with a FIB-4 score ≥2.67 may have moderate or high risk for advanced hepatic fibrosis and should undergo further assessment with a secondary NIT, either blood- or imaging-based biomarker. In the primary care setting, as initial secondary tests, either the Enhanced Liver Fibrosis (ELF) panel, vibration-controlled elastography (VCTE), or ultrasound-based methods such as acoustic radiation force impulse (as available) are favored over magnetic resonance elastography (MRE) due to cost considerations. A VCTE-derived liver stiffness measurement (LSM) <8 kPa can be used to rule out advanced fibrosis, especially if used sequentially after FIB-4. LSMs by VCTE between 8 and 12 kPa may be associated with fibrotic MASH, and LSM >12 kPa is associated with a high likelihood of advanced fibrosis, although the positive predictive value is low (range 0.34–0.71). LSM by VCTE may be suggestive of portal hypertension. Changes in LSMs may also be useful in identifying disease progression. An increase in liver stiffness of 20% on either VCTE or MRE may be associated with disease progression and long-term clinical outcomes. Low LSM is most useful for excluding advanced hepatic fibrosis as cirrhosis is extremely unlikely when LSM is low. However, higher LSM must be interpreted with caution since several factors (obesity, nonfasting state, severe hepatic inflammation, hepatic congestion, and/or iron overload) decrease the specificity of the test.

The ELF panel is a proprietary blood test consisting of three elements involved in matrix turnover: hyaluronic acid, tissue inhibitor of metalloproteinase-1, and N-terminal procollagen III peptide. The ELF test is approved by the U.S. Food and Drug Administration (FDA) for prognostication (ELF ≥9.8) of MASLD patients at increased risk of progression to cirrhosis and liver-related clinical events. The ELF test can be ordered, particularly when access to VCTE is limited. If secondary risk assessment with blood- or imaging-based NITs is consistent with an intermediate or high risk of advanced disease, patients should be referred to specialty care for further evaluation and potential consideration of magnetic resonance imaging (MRI)-based tools such as MRE or MRI-corrected T1 (cT1) and/or liver biopsy to further risk stratify patients in whom other NITs have been indeterminate or not consistent with clinical suspicion of liver disease severity. A liver biopsy should be considered when there is diagnostic uncertainty, discordant or indeterminate NITs, discordance between NITs and clinical, radiologic, and/or laboratory features (i.e., thrombocytopenia) suggesting a diagnosis of advanced fibrosis, competing or concomitant diagnoses (e.g., iron overload, autoimmune hepatitis, drug-induced liver injury), or persistent elevation (>6 months) in liver chemistries in the absence of hepatic steatosis or metabolic risk factors. Knowledge of the presence of advanced hepatic fibrosis or cirrhosis changes clinical care and management (i.e., primary and secondary prevention of complications of portal hypertension and screening for HCC), which can decrease risk for liver-related outcomes.

Although abdominal ultrasound is routinely used in clinical practice, ultrasound lacks sufficient sensitivity for lesser degrees (<30%) of steatosis, particularly in those with concomitant obesity. The semiquantitative assessment of steatosis on ultrasound does not exclude the presence of MASH or hepatic fibrosis, although nodular-appearing morphology, splenomegaly, and/or portosystemic collateral vessels may suggest presence of cirrhosis. The controlled attenuation parameter (CAP) typically measured in conjunction with VCTE provides a point-of-care semiquantitative assessment of hepatic steatosis but does not accurately monitor for changes in liver fat. MRI–proton density fat fraction (PDFF) is routinely use in clinical research and increasingly used in tertiary care settings as an accurate, reproducible, and precise NIT for quantification of liver fat content. Although MRI-PDFF is superior to CAP in the diagnosis and quantification of liver fat, it is not a point-of-care test, and enthusiasm for its more routine use is tempered by cost, patient acceptance, and the fact that quantification of hepatic steatosis alone (as opposed to MASH or fibrosis) has not been associated with a clinically meaningful outcome. Other serum and imaging biomarkers are under study for the detection of "at-risk" MASH (steatohepatitis with ≥F2 fibrosis); however, these newer MASH biomarkers are not yet ready for routine clinical practice and require further validation. NIS4, a panel of four serum biomarkers (miR-34a-5p, α_2-macroglobulin, YKL-40, and glycated hemoglobin) and other blood-based markers utilizing lipidomics, metabolomics, or proteomic biomarkers for at-risk MASH are in development.

Although NITs are being validated against liver histology and clinically meaningful outcomes, data regarding the dynamic context to ascertain a treatment response or prognosticate time to outcomes of progression to cirrhosis or hepatic decompensation are still evolving. ALT normalization in response to any treatment intervention (discussed below) may predict MASH resolution. A decrease of ≥17 IU/L in ALT has been associated with higher odds of histologic response in clinical trials. Likewise, improvements in FIB-4, ELF, VCTE, and MRE were correlated with reduction in liver injury and fibrosis. Several studies and a meta-analysis have shown that ≥30% reduction in liver fat content by MRI-PDFF is associated with a fivefold improved likelihood of MASH resolution; however, thresholds for changes in LSM that correlate with improvements in fibrosis are not well established. Although more data and broader validation are needed, emerging data suggest that improvement in NITs may be a reliable surrogate for predicting histologic improvement in response to a therapeutic intervention and associated clinical outcome.

■ TREATMENT OF MASLD

Treatment approaches are focused on modification of identifiable risk factors and targeted approaches to management of metabolic comorbidities that contribute pathophysiologically to the acquisition and progression of MASLD. For the overwhelming majority of patients, healthy diet and routine physical activity are foundational in the care plan, irrespective of weight loss goals. Thereafter, treatment for MASLD can be divided into the following: (1) therapies that improve cardiometabolic risk and associated comorbidities (e.g., obesity, insulin resistance, T2DM, dyslipidemia); (2) liver-directed therapies, which are typically reserved for those with clinically significant (i.e., at-risk) MASH with ≥F2 fibrosis; and (3) treatment of complications of advanced MASLD/MASH. The subsequent discussion highlights specific therapies for MASH and their impact on MASH disease activity. Generally, treatment(s) of complications of advanced MASLD/MASH, including cirrhosis, portal hypertension, and primary liver cancer, requires tertiary care and multidisciplinary approaches. Approaches to accomplish these objectives are similar to those used in other chronic liver diseases and are covered elsewhere in the textbook.

Lifestyle Modification Lifestyle modification with healthy diet and exercise is foundational in the treatment of MASLD/MASH. For patients with overweight/obese status, even modest amounts of weight loss can be impactful in improving disease activity and clinical outcomes. Weight loss of 3–5% has been demonstrated to improve hepatic steatosis, whereas greater weight loss (>10%) improves MASH and fibrosis. In those with advanced hepatic fibrosis or cirrhosis, moderate- to high-intensity exercise may decrease portal hypertension, sarcopenia, and risk for HCC. Unfortunately, achieving and sustaining weight loss long term is challenging for most patients with insulin resistance, with <10% achieving effective weight loss despite structured interventions at 1 year. However, sustained weight loss improves peripheral insulin sensitivity, cytokine-medicated inflammatory response, both hepatic and nonhepatic tissue stress, and even alterations in gut microbiota that contribute to metabolic dysfunction. A multidisciplinary approach, inclusive of patient and family engagement strategies, behavioral medicine specialists, dieticians, nutritionists, and a multidisciplinary team approach to management of complications of metabolic syndrome, may optimize results.

Modification of dietary macronutrient composition with avoidance of excess saturated fats, refined carbohydrates, and sugar-sweetened beverages (i.e., Western obesogenic diet) is recommended. Excess fructose consumption increases the risk for MASLD, MASH, and hepatic fibrosis, independent of calorie intake. Various approaches to dietary modification (e.g., low-fat vs low-carbohydrate, saturated vs unsaturated fats, intermittent fasting, Mediterranean diet, Mediterranean green diet, or different intensities of calorie restriction) appear relatively comparable in their ability to improve MASLD/MASH. The Mediterranean diet is often recommended based on long-term adherence, ability to be tailored across different cultures and ethnicities, and benefits with cardiovascular risk reduction. Coffee consumption (at least 3 cups daily) may also be beneficial based on epidemiologic studies and meta-analyses with reduced risk for hepatic fibrosis and HCC. Increased physical activity (i.e., energy expenditure), independent of weight loss, compliments dietary modification. Exercise improves insulin sensitivity, decreases cardiometabolic risk, and should be tailored to a patient's preference and physical ability. Regular moderate exercise at least five times per week for a total of 150 min/week or more intensive exercise for at least 60 min/week can prevent or improve MASLD and associated histologic features of liver injury. A dose-dependent response of exercise in improvement of MASLD/MASH has been observed, although the optimal duration, intensity, and type (aerobic vs resistance) needs to be individualized to patient ability and capacity. For patients with advance liver disease due to MASLD, exercise can improve frailty, sarcopenia, and quality of life in patients with cirrhosis. Changes in dietary composition, when coupled with increased physical activity, may be sufficient to improve MASLD/MASH and/or result in improvements in patients with milder disease activity.

Pharmacologic Therapies *Resmetirom*, a selective thyroid hormone receptor beta (THR-B) agonist, was recently granted accelerated approval by the FDA to treat patients with 'at-risk' MASH (i.e. F2-F3 fibrosis) without cirrhosis. Resmetirom works by mimicking the effects of thyroid hormone in the liver, promoting lipid metabolism, reducing liver fat, and improving liver function. In a 52 week phase randomized controlled trial of resmetirom (80 mg or 100 mg) or placebo in 966 patients with biopsy-proven NASH with fibrosis (MAESTRO-NASH trial) which had not yet advanced to cirrhosis, resmetirom 100 mg and 80 mg were superior to placebo with respect to NASH resolution (30%, 26% and 10%, respectively; p<0.0001) and improvement in liver fibrosis (26%, 24% and 14%, respectively, p<0.0001). Resmetirom was generally well tolerated with the predominant treatment emergent side effects being early treatment transient gastrointestinal symptoms of nausea and loose stools. The MEASTRO-NASH study is ongoing to accrue data on long-term safety and liver-related outcomes including progression to cirrhosis for full FDA approval. Pharmacotherapy for liver disease per se is not recommended for patients with hepatic steatosis without steatohepatitis or fibrosis, but these patients should receive counseling for healthy diet and physical activity. Pharmacotherapy to treat metabolic comorbidities may improve MASLD.

The pipeline for emerging therapies for MASH is robust, with a large number of clinical trials designed to identify effective and safe treatment in progress. For the purpose of reporting trial results acquired prior to change in disease definition and nomenclature, the terms NAFLD and NASH will be utilized with the understanding that overall concordance between NAFLD and MASLD and between NASH and MASH is high.

Because MASLD is strongly associated with the metabolic syndrome **(Chap. 420)** and T2DM **(Chap. 415)**, the efficacy of various insulin-sensitizing agents has been examined.

Metformin, a biguanide anti-hyperglycemic agent, improves hepatic insulin sensitivity and has been evaluated in several small, open-label studies in adults and a larger, prospectively randomized trial in children (dubbed the TONIC study). Although several of the adult NASH studies suggested improvements in aminotransferases and, less consistently, liver histology, metformin did not improve liver histology in the TONIC study of children with NASH. Thus, it is not currently recommended as a treatment for NASH. However, meta-analyses suggest metformin may decrease risk of HCC in patients with T2DM.

Thiazolidinediones, ligands for peroxisome proliferator-activated receptor (PPAR) γ, improve insulin resistance and serum lipids and are approved for the treatment of T2DM. In a randomized controlled trial sponsored by the National Institutes of Health of pioglitazone versus vitamin E versus placebo (PIVENS trial) for the treatment of adults (n = 247) with NASH without prediabetes or T2DM, pioglitazone (30 mg/d) for 18 months resolved NASH compared to placebo (47 vs 21%, respectively; p <.001). In an 18-month study in patients with NASH and prediabetes or T2DM, pioglitazone improved the nonalcoholic activity score (NAS) by ≥2 points with a trend toward fibrosis improvement. The effect of pioglitazone on NASH resolution and fibrosis improvement was subsequently demonstrated in a meta-analysis. The side effects of pioglitazone include weight gain, bone loss in postmenopausal women, potential risk for worsening heart failure in patients with preexisting cardiac dysfunction, and a debated risk of bladder cancer. The use of pioglitazone has been largely usurped in clinical practice by the increasing use of incretin mimetics (e.g., glucagon-like peptide 1 receptor agonists [GLP-1RAs]) and sodium-glucose cotransporter 2 inhibitors (SGLT-2i) because these agents are reasonably well tolerated and have multiple benefits (i.e., improve glycemic control, induce weight loss, and reduce cardiovascular mortality).

*Incretin mimetic*s, drugs that act on the pancreas to optimize insulin and glucagon release, have a beneficial effect on MASH. GLP-1RAs are currently approved for the indication of obesity and T2DM. GLP-1RAs improve lipids, glucose, and weight, and decrease cardiovascular risk, making them attractive agents for treatment of MASH. In a small study of patients with NASH, liraglutide improved steatosis, resolved NASH, and reduced fibrosis progression compared to placebo. In a large randomized, double-blind, placebo-controlled study of daily subcutaneous semaglutide in 320 patients with biopsy-proven NASH (F1–F3) randomized to 0.1, 0.2, or 0.4 mg or placebo for 72 weeks, a dose-dependent response on NASH resolution without worsening of fibrosis (primary endpoint) was observed (59% in 0.4-mg treatment group vs 17% in placebo group, p <.001). Although there was no statistically significant improvement in fibrosis, a protective effect against fibrosis progression was observed. A global phase 3 study of semaglutide for the treatment of NASH-related fibrosis is currently underway. Tirzepatide, a GLP-1/glucose-dependent insulintopic polypeptide receptor agonist approved for the treatment of T2DM, demonstrated significant weight loss (up to 20.9% change in body weight in obese patients without diabetes mellitus compared with 3.1% in the placebo group at week 72). The weight loss was associated with significant reductions in liver fat content, identifying this agent as another potentially effective therapy for MASH.

SGLT-2is increase glucose excretion from the glomerular filtrate and are approved for the treatment of T2DM. SGLT-2is induce mild weight loss (2–3%) and exert cardioprotective benefits. Studies evaluating the efficacy of SGLT-2i in MASH are limited by small sample size and lack of histologic endpoints; however, available data suggest that SGLT-2is improve liver aminotransferases and quantitative measures of hepatic steatosis. Whether this translates into resolution of MASH or improvement in hepatic fibrosis needs to be defined.

Antioxidants have also been evaluated for the treatment of MASLD because oxidant stress is thought to contribute to the pathogenesis of MASH. Vitamin E, an inexpensive yet potent antioxidant, has been examined in several small pediatric and adult studies with varying results. In all of those studies, vitamin E was well tolerated, and most studies showed modest improvements in aminotransferase levels, radiographic features of hepatic steatosis, and/or histologic features of NASH. In the PIVENS study, treatment with RRR-α-tocopherol (the natural form of vitamin E) 800 IU daily for 96 weeks improved histologic features of steatohepatitis without worsening of fibrosis (43% in the vitamin E group vs 19% in the placebo group, p <.04), a finding supported by a meta-analysis showing that vitamin E improves serum aminotransferases and features of NASH on biopsy. Although no study has demonstrated an antifibrotic effect with vitamin E, a retrospective study of patients with NASH and advanced hepatic fibrosis (n = 236) showed that vitamin E use, independent of diabetes status, was associated with lower rate of hepatic decompensation (37 vs 62%, p = .04) and higher transplant-free survival (78 vs 49%, p <.01). Enthusiasm for use of vitamin E for treatment of NASH has been dampened by concerns of the risk-to-benefit ratio for safety (i.e., risk of perioperative bleeding, hemorrhagic stroke, increased cardiac mortality, and prostate cancer) with long-term efficacy.

Statins are an important class of agents to treat dyslipidemia and decrease cardiovascular risk, even in patients with MASLD/MASH. Statins are safe across the spectrum of MASLD, including in patients with well-compensated cirrhosis. Statins decrease cardiovascular morbidity and mortality; however, they remain underutilized in clinical practice despite the extensive evidence demonstrating safety in patients with chronic liver disease. Statins have beneficial effects on decreasing portal pressure and, thus, decrease risk for hepatic decompensation. Meta-analyses also suggest that long-term statin use may decrease risk for cirrhosis and HCC.

The incidence of liver enzyme elevations in MASLD patients taking statins is also no different than that of healthy controls or patients with other chronic liver diseases. Moreover, several studies have suggested that statins may improve aminotransferases and histology in patients with MASH. Yet, there is continued reluctance to use statins in patients with MASLD. The lack of evidence that statins harm the liver in MASLD patients, combined with the increase risk for cardiovascular morbidity and mortality in MASLD patients, justifies the use of statins to treat dyslipidemia in patients with MASLD, either as monotherapy or in combination with other lipid-lowering agents to achieve therapeutic goals. Limited data exist on the safety and efficacy of statins in patients with decompensated MASLD-related cirrhosis.

Drugs Shown Not to Benefit MASH Ursodeoxycholic acid (a bile acid conjugate used to treat certain cholestatic liver disease), betaine (a metabolite of choline that raises *S*-adenosylmethioine levels and decreases cellular oxidative damage), dipeptidylpeptidase-4 (an insulin-sensitizing agent), and silymarin (a putative antioxidant) are well studied in NASH and should not be used as a treatment for NASH/MASH because they do not offer a meaningful histologic benefit.

Emerging Therapies Many other pharmacotherapies that target dysregulated energy homeostasis, lipotoxicity, cell death, and liver inflammation, processes that are critically involved in the pathogenesis and/or progression of MASH and liver fibrosis, are currently in clinical trials (e.g., probiotics, farnesoid X receptor agonists, fibroblast growth factor agonists, antiapoptotic agents, anticytokine agents, PPAR modulators, thyroid hormone receptor β-selective agonists, stearyl-CoA desaturase-1 inhibitors, DGAT inhibitors, acyl-CoA carboxylase inhibitors, and direct modulators of liver fibrosis). Sufficient data do not yet exist to justify the use of these emerging therapies as MASH treatments in clinical practice. Given that liver disease outcomes in MASH patients are highly heterogeneous, optimal treatment of MASH may need to be individualized by tailoring therapy based on clinical or histologic phenotypes of MASH and/or genetic susceptibility for disease progression.

Bariatric Surgery Obesity, BMI ≥40 kg/m^2 irrespective of metabolic comorbid conditions, or BMI ≥35 kg/m^2 with comorbidities (e.g., T2DM, hypertension, OSA, hypertension) are accepted indications for bariatric surgery. MASLD is increasingly being accepted as the hepatic manifestation of metabolic syndrome and a complication that improves with effective and sustained weight loss. The overwhelming majority of patients undergoing bariatric surgery have MASLD; however, the prevalence of advanced hepatic fibrosis and cirrhosis is low (likely due to presurgical screening of those with chronic liver disease/cirrhosis or variable risk of visceral fat accumulation in patients with larger capacity for peripheral adipose depot stores). Bariatric surgery cures diabetes in a majority of patients, improves MASH and hepatic fibrosis, and decreases both liver-related and all-cause morbidity and mortality. In a prospective study, bariatric surgery resolved the histologic features of MASH and improved hepatic fibrosis in 80% of patients at 1 year; this improvement was maintained in patients who sustained weight loss at 5 years after surgery. Failure to sustain weight, or the regain of weight, following bariatric surgery is associated with persistent MASH. The impact of bariatric surgery is concordant with the degree of weight loss, with malabsorptive procedures being more effective than restrictive surgical procedures that result in substantially less weight loss. While endoscopic bariatric and metabolic procedures are promising as less invasive options, long-term safety and efficacy data are needed. The risk of hepatic decompensation increases in the presence of cirrhosis. At present, bariatric surgery cannot be considered as a primary salvage therapy for patients with compensated MASH cirrhosis. However, it may be safely accomplished at select tertiary care centers for selected patients. Bariatric surgery in the setting of decompensated cirrhosis and/or portal hypertension is associated with increased mortality and should only be consider at high-volume centers with expertise with combined liver transplant and bariatric surgery.

Liver Transplantation Patients with MASLD in whom end-stage liver disease develops should be evaluated for liver transplantation **(Chap. 356)**. The outcomes of liver transplantation in well-selected patients with MASLD are generally good, but comorbid medical conditions associated with MASLD, such as diabetes mellitus, obesity, and CVD, often limit transplant candidacy. MASLD may recur after liver transplantation. The risk factors for recurrent or de novo MASLD after liver transplantation are multifactorial and include hypertriglyceridemia, obesity, diabetes mellitus, and immunosuppressive therapies, particularly glucocorticoids.

GLOBAL HEALTH CONSIDERATIONS

MASLD, the hepatic manifestation of metabolic syndrome, is a potentially progressive liver disease. The global prevalence is 30% with the highest MASLD prevalence in Latin America (44%), Middle East and North Africa (36.5%), and South Asia (33.8%). The rising prevalence globally requires urgent and comprehensive strategies to raise awareness and address all aspects of MASLD on local, regional, and global levels. Although a minority of patients with MASLD will progress to cirrhosis or HCC, given the vast population affected, the total number of patients who are at risk for adverse clinical outcomes is already huge and will increase. Of concern, individuals are increasingly being affected by MASLD at an earlier age, meaning there is more time for them to develop severe complications. The epidemic of chronic liver disease and MASLD-related cirrhosis is anticipated to rise in parallel with diabetes, obesity, and other complications of metabolic dysfunction. MASLD is a huge burden on public health that has received little attention from the global public health community and policymakers. The economic burden directly attributable to MASLD is already enormous (estimated direct medical costs of ~$103 billion/year in the United States and €35 billion/year in the Europe-4 countries: Germany, France, Italy, and United Kingdom) with projections for continued increases in economic burden in the upcoming years. The global impact of MAFLD is projected to increase, with a 178% increase in liver-related deaths by 2030.

A multistakeholder approach for policymakers, health care practitioners, civil society groups, research institutions, and patient advocacy groups could provide a strong foundation for a comprehensive global public health response to MASLD. Despite the global burden, there are large geographical discrepancies in the prevalence of MASLD, which could be attributable to differing socioeconomic milieus, degree of urbanization, and genetic predisposition (i.e., higher prevalence in certain ethnic groups such as Asians, Hispanics, and Native Americans, and lower in others such as African Americans compared to whites). Although all of these variables confound efforts to predict the net impact of this obesity-related liver disease on global health, it seems likely that MASLD will remain a major cause of chronic liver disease worldwide for the foreseeable future. The landscape of emerging therapies for MASLD could potentially turn the tide on the growing epidemic of MASLD-related chronic liver disease once the first liver-directed pharmacotherapy for MASLD achieves FDA approval.

FURTHER READING

Canivet CM et al: Validation of the new 2021 EASL algorithm for the noninvasive diagnosis of advanced fibrosis in NAFLD. J Hepatology 77:920, 2023.

Chalasani N et al: The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases Hepatology 67:328, 2018.

Cusi K et al: American Association of Clinical Endocrinology clinical practice guideline for the diagnosis and management of nonalcoholic fatty liver disease in primary care and endocrinology clinical settings: Co-sponsored by the American Association for the Study of Liver Diseases (AASLD). Endocr Pract 28:528, 2022.

Diehl AM, Day CSC: Cause, pathogenesis, and treatment of nonalcoholic steatohepatitis. N Engl J Med 377:2063, 2017.

European Association for the Study of the Liver (EASL) et al: EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. J Hepatol 64:1388, 2016.

Harrison SA et al: A phase 3, randomized, controlled trial of resmetirom in NASH with liver fibrosis. N Engl J Med 390:497, 2024.

Kanwal F et al: Clinical care pathway for the risk stratification and management of patients with nonalcoholic fatty liver disease. Gastroenterology 161:1657, 2021.

Rinella ME et al: AASLD practice guidance on the clinical assessment and management of nonalcoholic fatty liver disease. Hepatology 77:1797, 2023.

Vos MB et al: NASPGHAN clinical practice guideline for the diagnosis and treatment of nonalcoholic fatty liver disease in children: Recommendations from the Expert Committee on NAFLD (ECON) and the North American Society of Pediatric Gastroenterology, Hepatology and Nutrition (NASPGHAN). J Pediatr Gastroenterol Nutr 64:319, 2017.

Younossi ZM et al: The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): A systematic review. Hepatology 77:1335, 2023.

355

Cirrhosis and Its Complications

Alex S. Befeler, Bruce R. Bacon

Cirrhosis is a condition that is defined histopathologically and has a variety of clinical manifestations and complications, some of which can be life-threatening. In the past, it has been thought that cirrhosis was never reversible; however, it has become apparent that when the underlying insult that has caused the cirrhosis has been removed, there can be reversal of fibrosis. This is most apparent with the successful treatment of chronic hepatitis C; however, reversal of fibrosis is also seen in patients with hemochromatosis after iron removal and in patients with alcohol-associated liver disease who have discontinued alcohol use.

Regardless of the cause of cirrhosis, the pathologic features consist of the development of fibrosis to the point that there is architectural distortion with the formation of regenerative nodules. This results in a decrease in hepatocellular mass, and thus function, and an alteration of blood flow. The induction of fibrosis occurs with activation of hepatic stellate cells, resulting in the formation of increased amounts of collagen and other components of the extracellular matrix.

Clinical features of cirrhosis are the result of pathologic changes and mirror the severity of the liver disease. Most hepatic pathologists provide an assessment of grading and staging when evaluating liver biopsy samples. These schemes vary between disease states and have been developed for most conditions, including chronic viral hepatitis, nonalcoholic fatty liver disease, and primary biliary cholangitis. Advanced fibrosis usually includes bridging fibrosis with nodularity designated as stage 3 and cirrhosis designated as stage 4. Patients who have cirrhosis have varying degrees of liver function, and clinicians need to differentiate between those who have stable, compensated cirrhosis and those who have decompensated cirrhosis. Patients who have developed ascites, hepatic encephalopathy, or variceal bleeding are classified as decompensated. They should be considered for liver transplantation, particularly if the decompensations are poorly controlled. Many of the complications of cirrhosis will require specific therapy. *Portal hypertension* is a significant complicating feature of decompensated cirrhosis and is responsible for the development of ascites and bleeding from esophagogastric varices. Loss of hepatocellular function results in jaundice, coagulation disorders, and hypoalbuminemia and contributes to the causes of portosystemic encephalopathy. The complications of cirrhosis are basically the same regardless of the etiology. Nonetheless, it is useful to classify patients by the cause of their liver disease **(Table 355-1)**; patients can be divided into broad groups, including those with alcohol-associated cirrhosis, cirrhosis due to chronic viral hepatitis, biliary cirrhosis, metabolic dysfunction–associated steatotic liver disease (the new consensus term for nonalcoholic fatty liver disease), and other, less common causes, such as cardiac cirrhosis, cryptogenic cirrhosis, and other miscellaneous causes.

TABLE 355-1 Causes of Cirrhosis

- Alcohol
- Chronic viral hepatitis
    - Hepatitis B
    - Hepatitis C
- Autoimmune hepatitis
- Metabolic dysfunction-associated steatohepatitis
- Biliary cirrhosis
    - Primary biliary cholangitis
    - Primary sclerosing cholangitis
    - Autoimmune cholangiopathy
- Cardiac cirrhosis
- Inherited metabolic liver disease
    - Hemochromatosis
    - Wilson's disease
    - α_1 Antitrypsin deficiency
    - Cystic fibrosis
- Cryptogenic cirrhosis

ALCOHOL-ASSOCIATED CIRRHOSIS

Excessive chronic alcohol use can cause several different types of chronic liver disease, including alcohol-associated fatty liver, alcoholic hepatitis, and alcohol-associated cirrhosis. Furthermore, use of excessive alcohol can contribute to liver damage in patients with other liver diseases, such as hepatitis C, hemochromatosis, and metabolic dysfunction–associated steatotic liver disease. Chronic alcohol use can produce fibrosis in the absence of accompanying inflammation and/or necrosis. Fibrosis can be centrilobular, pericellular, or periportal. When fibrosis reaches a certain degree, there is disruption of the normal liver architecture and replacement of liver cells by regenerative nodules. In alcohol-associated cirrhosis, the nodules are usually <3 mm in diameter; this form of cirrhosis is referred to as *micronodular*. With cessation of alcohol use, larger nodules may form, resulting in a mixed micronodular and macronodular cirrhosis.

Pathogenesis Alcohol is the most commonly used drug in the United States, and >70% of adults drink alcohol each year. Twenty percent have had a binge within the past month, and >7% of adults regularly consume more than four or five drinks five or more times a month. Unfortunately, >14 million adults in the United States meet the diagnostic criteria for alcohol use disorder. In the United States, chronic liver disease is the tenth most common cause of death in adults, and alcohol-associated cirrhosis accounts for ~48% of deaths due to cirrhosis.

Ethanol is mainly absorbed by the small intestine and, to a lesser degree, through the stomach. Gastric alcohol dehydrogenase (ADH) initiates alcohol metabolism. Three enzyme systems account for metabolism of alcohol in the liver. These include cytosolic ADH, the microsomal ethanol oxidizing system (MEOS) utilizing the inducible cytochrome P450 CYP2E1, and peroxisomal catalase. Normally the majority of ethanol oxidation occurs via ADH to form acetaldehyde, which is a highly reactive molecule that may have multiple effects. The MEOS pathway in chronic alcohol use causes induction of CYP2E1, which leads to generation of reactive oxygen species and produces more acetaldehyde. Ultimately, acetaldehyde is metabolized to acetate by aldehyde dehydrogenase (ALDH). Intake of ethanol increases intracellular accumulation of triglycerides by increasing fatty acid uptake and by reducing fatty acid oxidation and lipoprotein secretion. Protein synthesis, glycosylation, and secretion are impaired. Oxidative damage to hepatocyte membranes occurs due to the formation of reactive oxygen species; acetaldehyde is a highly reactive molecule that combines with proteins and nucleic acids to form acetaldehyde adducts. These adducts may interfere with specific enzyme activities, including microtubular formation and hepatic protein trafficking. With acetaldehyde-mediated hepatocyte damage, certain reactive oxygen species can result in Kupffer cell activation. As a result, profibrogenic cytokines are produced that initiate and perpetuate stellate cell activation, with the resultant production of excess collagen and extracellular matrix. Connective tissue appears in both periportal and pericentral zones and eventually connects portal triads with central veins forming regenerative nodules. Hepatocyte loss occurs, and with increased collagen production and deposition, together with continuing hepatocyte destruction, the liver contracts and shrinks in size. This process generally takes from years to decades to occur and requires repeated insults.

Clinical Features The diagnosis of alcohol-associated liver disease requires an accurate history regarding both amount (>2 drinks per day in women and >3 drinks per day in men) and duration (>12 months) of alcohol consumption. Patients with alcohol-associated liver disease can present with nonspecific symptoms such as vague right upper quadrant abdominal pain, fever, nausea and vomiting, diarrhea, anorexia, and malaise. Alternatively, they may present with complications of chronic liver disease, including ascites, edema, upper gastrointestinal (GI) hemorrhage, jaundice, or encephalopathy. Many cases present incidentally at the time of autopsy or elective surgery. Other patients may be identified during an evaluation of routine laboratory studies that are found to be abnormal. On physical examination, the liver and spleen may be enlarged, with the liver edge being firm and nodular. Other frequent findings include scleral icterus, palmar erythema **(Fig. 355-1)**, spider angiomas **(Fig. 355-2)**, parotid gland enlargement, digital clubbing, muscle wasting, edema, and ascites. Men may have decreased body hair, gynecomastia, and testicular atrophy, which may be a consequence of hormonal abnormalities or a direct toxic effect of alcohol on the testes. In women with advanced disease, menstrual irregularities usually occur including amenorrhea. These changes are often reversible following cessation of alcohol ingestion.

Laboratory tests may be completely normal in patients with early compensated alcohol-associated cirrhosis. Alternatively, in advanced liver disease, many abnormalities usually are present. Patients may be anemic from chronic GI blood loss, nutritional deficiencies, or hypersplenism or as a direct suppressive effect of alcohol on the bone marrow. A unique form of hemolytic anemia (with spur cells and acanthocytes) called *Zieve's syndrome* can occur in patients with severe alcoholic hepatitis. Platelet counts are often reduced early in the disease, reflective of portal hypertension with hypersplenism. Serum total bilirubin can be normal or elevated with advanced disease. Prothrombin times are often prolonged and usually do not respond to administration of parenteral vitamin K. Serum sodium levels are usually normal unless patients have ascites and then can be depressed, largely due to ingestion of excess free water. Serum alanine and aspartate aminotransferases (ALT, AST) are typically elevated but <400 IU/mL, with AST levels being higher than ALT levels, usually by a 2:1 ratio, particularly in patients who continue to drink.

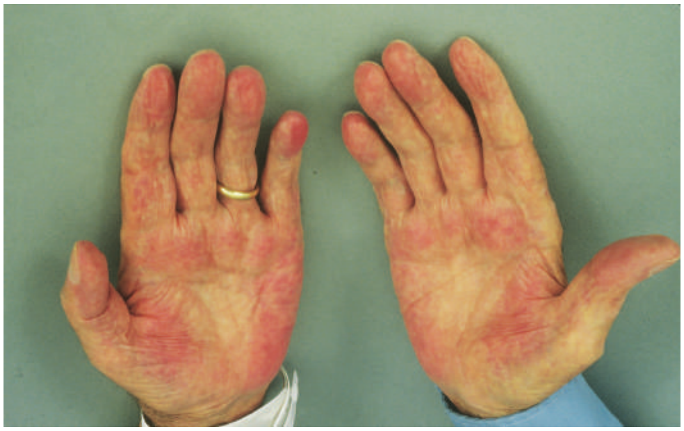

FIGURE 355-1 Palmar erythema. This figure shows palmar erythema in a patient with alcohol-associated cirrhosis. The erythema is peripheral over the palm with central pallor.

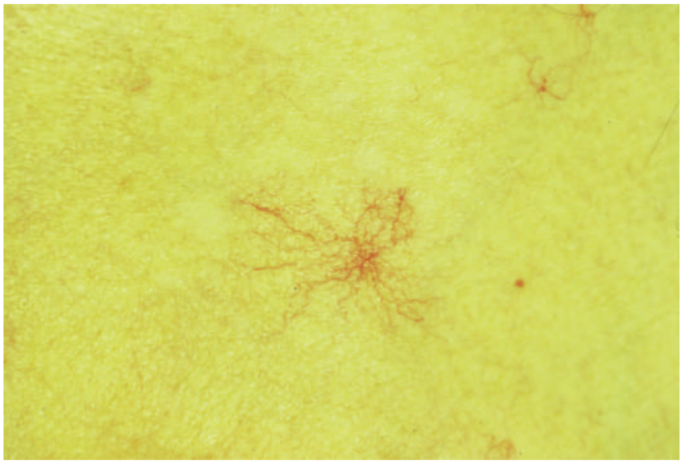

FIGURE 355-2 Spider angioma. This figure shows a spider angioma in a patient with hepatitis C cirrhosis. With release of central compression, the arteriole fills from the center and spreads out peripherally.

Diagnosis Patients who have any of the above-mentioned clinical features, physical examination findings, or laboratory studies in the setting of chronic alcohol consumption should be considered to have alcohol-associated liver disease. Furthermore, other forms of chronic liver disease (e.g., chronic viral hepatitis or metabolic or autoimmune liver diseases) must be considered or ruled out, or if present, an estimate of relative causality along with the alcohol use should be determined. Liver biopsy can be helpful to confirm a diagnosis but generally is not performed unless there is a suspicion of an alternative diagnosis.

In patients who have had complications of cirrhosis and who continue to drink, there is a <50% 5-year survival. In contrast, in patients who remain abstinent, the prognosis is significantly improved, particularly when they have resolution of liver complications; however, some individuals who remain abstinent do not improve and liver transplantation is a viable option.

TREATMENT

Alcohol-Associated Cirrhosis and Alcoholic Hepatitis

Abstinence is the cornerstone of therapy for patients with alcohol-associated liver disease. New clinically available biomarkers of recent alcohol consumption such as phosphatidylethanol (Peth) can be helpful in evaluating abstinence. In addition, patients require good nutrition and long-term medical supervision to manage underlying complications that may develop. Complications such as the development of ascites and edema, variceal hemorrhage, or portosystemic encephalopathy all require specific management and treatment. Liver transplantation can be an effective long-term treatment in those who have been deemed a low enough risk for alcohol relapse and do not respond to other treatments.

Glucocorticoids are occasionally used in patients with severe alcoholic hepatitis in the absence of infection. Short-term survival has been shown to be improved in certain studies and meta-analysis, although 6-month survival is more dependent on abstinence. Treatment is restricted to patients with a discriminant function (DF) value of >32. The DF is calculated as the serum total bilirubin plus the difference in the patient's prothrombin time compared to upper limit of control (in seconds) multiplied by 4.6. Failure to improve total bilirubin after 7 days predicts treatment failure, and glucocorticoids can be stopped; otherwise, they are continued for 28 days.

There is modest evidence that intravenous *N*-acetylcysteine plus glucocorticoids may have a short-term survival benefit in alcoholic hepatitis if the DF is >32. Other therapies including oral pentoxifylline, inhibitors of tumor necrosis factor (TNF) α, anabolic steroids, propylthiouracil, antioxidants, colchicine, and penicillamine have not shown clear-cut benefits and are not recommended. A variety of nutritional therapies have been tried, both parenteral and enteral feedings; however, there is no clear evidence of improved survival. There is evidence that persons who consume >21.5 kcal/kg body weight per day have better survival, so achieving better caloric intake is recommended. Finally, in highly selected patients with good social support structure who fail other treatments for alcoholic hepatitis, early liver transplant can be an effective treatment.

Cessation of alcohol use is key. Recent experience with medications that reduce craving for alcohol, such as acamprosate calcium and baclofen, have been favorable. Patients may take other necessary medications even in the presence of cirrhosis. Acetaminophen use is often discouraged in patients with liver disease; however, if no more than 2 g of acetaminophen per day are consumed, there generally are no problems unless there is active alcohol use.

■ CIRRHOSIS DUE TO CHRONIC VIRAL HEPATITIS B OR C

Of patients exposed to the hepatitis C virus (HCV), ~80% develop chronic hepatitis C, and of those, ~20–30% will develop cirrhosis over 20–30 years. Many of these patients have had concomitant alcohol use, and the true incidence of cirrhosis due to hepatitis C alone is unknown. It is expected that an even higher percentage will go on to develop cirrhosis over longer periods of time. In the United States, ~5–6 million people have been exposed to HCV, and ~4–5 million are chronically viremic. Worldwide, ~170 million individuals have hepatitis C, with some areas of the world (e.g., Egypt) having up to 15% of the population infected. HCV is a noncytopathic virus, and liver damage is probably immune-mediated. Progression of liver disease due to chronic hepatitis C is characterized by portal-based fibrosis with bridging fibrosis and nodularity developing, ultimately culminating in the development of cirrhosis. In cirrhosis due to chronic hepatitis C, the liver is small and shrunken with characteristic features of a mixed micro- and macronodular cirrhosis seen on liver biopsy. In addition, an inflammatory infiltrate is found in portal areas with interface hepatitis and occasionally some lobular hepatocellular injury and inflammation. In patients with HCV genotype 3, steatosis is often present.

Similar findings are seen in patients with cirrhosis due to chronic hepatitis B. Of adult patients exposed to hepatitis B, ~5% develop chronic hepatitis B, and ~20% of those patients will go on to develop cirrhosis. Special stains for hepatitis B core (HBc) and hepatitis B surface (HBs) antigen will be positive, and ground-glass hepatocytes signify HBs antigen (HBsAg) may be present. In the United States, there are ~2 million carriers of hepatitis B, whereas in other parts of the world where hepatitis B virus (HBV) is endemic (i.e., Asia, Southeast Asia, sub-Saharan Africa), up to 15% of the population may be infected, having acquired the infection vertically at the time of birth. Thus, >300–400 million individuals are thought to have hepatitis B worldwide. Approximately 25% of these individuals may ultimately develop cirrhosis.

Clinical Features and Diagnosis Patients with cirrhosis due to either chronic hepatitis C or B can present with the usual symptoms and signs of chronic liver disease. Fatigue, malaise, vague right upper quadrant pain, and laboratory abnormalities are frequent presenting features. Diagnosis requires a thorough laboratory evaluation, including quantitative HCV RNA testing and analysis for HCV genotype, or hepatitis B serologies to include HBsAg, HBeAg (hepatitis B e antigen), anti-HBe, and quantitative HBV DNA levels.

TREATMENT

Cirrhosis due to Chronic Viral Hepatitis B or C

Management of complications of cirrhosis revolves around specific therapy for treatment of whatever complications occur (e.g., esophageal variceal hemorrhage, development of ascites and edema, or encephalopathy). In patients with chronic hepatitis B, numerous studies have shown beneficial effects of antiviral therapy, which is effective at viral suppression, as evidenced by reducing aminotransferase levels and HBV DNA levels and improving histology by reducing inflammation and fibrosis. Several clinical trials and case series have demonstrated that patients with decompensated liver disease can become compensated with the use of antiviral therapy directed against hepatitis B. Currently available therapy includes lamivudine, adefovir, telbivudine, entecavir, and tenofovir, with the latter two strongly preferred because of reduced risk of viral resistance. Interferon α can also be used for treating hepatitis B, but it should not be used in cirrhotics **(see Chap. 352)**.

Treatment of patients with cirrhosis due to hepatitis C with direct-acting antiviral protocols is highly successful (>95% cure rate), well tolerated, and usually of short duration (8–12 weeks) **(see Chap. 352)**.

CIRRHOSIS FROM AUTOIMMUNE HEPATITIS AND METABOLIC DYSFUNCTION–ASSOCIATED STEATOTIC LIVER DISEASE

Other causes of posthepatic cirrhosis include autoimmune hepatitis (AIH) and metabolic dysfunction–associated steatohepatitis (MASH), which was previously called nonalcoholic steatohepatitis. Many patients with AIH present with cirrhosis that is already established. Typically, these patients will not benefit from immunosuppressive therapy with glucocorticoids or azathioprine because the AIH is "burned out." In this situation, liver biopsy does not show a significant inflammatory infiltrate. Diagnosis in this setting requires positive autoimmune markers such as antinuclear antibody (ANA) or anti-smooth-muscle antibody (ASMA). When patients with AIH present with cirrhosis and active inflammation accompanied by elevated liver enzymes, there can be considerable benefit from the use of immunosuppressive therapy.

Patients with MASH are increasingly being found to have progressed to cirrhosis. With the epidemic of obesity that continues in Western countries, more and more patients are identified with metabolic dysfunction–associated steatotic liver disease **(Chap. 354)**. Of these, a significant subset has MASH and can progress to increased fibrosis and cirrhosis. Over the past several years, it has been increasingly recognized that many patients who were thought to have cryptogenic cirrhosis in fact have MASH. As their cirrhosis progresses, they become catabolic and then lose the telltale signs of steatosis seen on biopsy. Management of complications of cirrhosis due to either AIH or MASH is similar to that for other forms of cirrhosis.

■ BILIARY CIRRHOSIS

Biliary cirrhosis has pathologic features that are different from either alcohol-associated cirrhosis or posthepatic cirrhosis, yet the manifestations of end-stage liver disease are the same. Cholestatic liver disease may result from necroinflammatory lesions, congenital or metabolic processes, or external bile duct compression. Thus, two broad categories reflect the anatomic sites of abnormal bile retention: *intrahepatic* and *extrahepatic*. The distinction is important for obvious therapeutic reasons. Extrahepatic obstruction may benefit from surgical or endoscopic biliary tract decompression, whereas intrahepatic cholestatic processes will not improve with such interventions and require a different approach.

The major causes of chronic cholestatic syndromes are primary biliary cholangitis (PBC), autoimmune cholangitis (AIC), primary sclerosing cholangitis (PSC), and idiopathic adulthood ductopenia. These syndromes are usually clinically distinguished from each other by antibody testing, cholangiographic findings, and clinical presentation. However, they all share the histopathologic features of chronic cholestasis, including cholate stasis; copper deposition; xanthomatous transformation of hepatocytes; and irregular, so-called biliary fibrosis. In addition, there may be chronic portal inflammation, interface activity, and chronic lobular inflammation. Ductopenia is a result of progressive disease as patients develop cirrhosis.

PRIMARY BILIARY CHOLANGITIS

PBC is seen in about 100–200 individuals per million, with a strong female preponderance and a median age of ~50 years at the time of diagnosis. The cause of PBC is unknown; it is characterized by portal inflammation and necrosis of cholangiocytes in small- and medium-sized bile ducts. Cholestatic features prevail, and biliary cirrhosis is characterized by an elevated bilirubin level and progressive liver failure. Liver transplantation is the treatment of choice for patients with decompensated cirrhosis due to PBC.

Antimitochondrial antibodies (AMAs) are present in ~95% of patients with PBC. These autoantibodies recognize lipoic acid on the inner mitochondrial membrane proteins that are enzymes of the pyruvate dehydrogenase complex (PDC), the branched-chain 2-oxoacid dehydrogenase complex, and the 2-oxogluterate dehydrogenase complex. These autoantibodies are not pathogenic, but rather are useful markers for making a diagnosis.

Pathology Histopathologic analysis identifies four distinct stages of the disease as it progresses. The earliest lesion is termed *chronic nonsuppurative destructive cholangitis* and is a necrotizing inflammatory process of the portal tracts. Medium and small bile ducts are infiltrated with lymphocytes and undergo duct destruction. Mild fibrosis and sometimes bile stasis can occur. With progression, the inflammatory infiltrate becomes less prominent, but the number of bile ducts is reduced and there is proliferation of smaller bile ductules. Increased fibrosis ensues with the expansion of periportal fibrosis to bridging fibrosis. Finally, cirrhosis, which may be micronodular or macronodular, develops.

Clinical Features Currently, most patients with PBC are middle-aged women diagnosed well before the end-stage manifestations of the disease are present, and as such, most patients are asymptomatic. When symptoms are present, they most prominently include a significant degree of fatigue out of proportion to either the severity of the liver disease or the age of the patient. Pruritus is seen in ~50% of patients at the time of diagnosis, and it can be debilitating. It might be intermittent and usually is most bothersome in the evening. In some patients, pruritus can develop toward the end of pregnancy and can be mistaken for cholestasis of pregnancy. Pruritus that presents prior to the development of jaundice indicates severe disease and a poor prognosis.

Physical examination can show jaundice and other complications of chronic liver disease including hepatomegaly, splenomegaly, ascites, and edema. Other features that are unique to PBC include hyperpigmentation, xanthelasma, and xanthomata, which are related to altered cholesterol metabolism. Hyperpigmentation is evident on the trunk and the arms and is seen in areas of exfoliation and lichenification associated with progressive scratching related to the pruritus. Bone pain resulting from osteopenia or osteoporosis is occasionally seen at diagnosis.

Laboratory Findings Laboratory findings in PBC show cholestatic liver enzyme abnormalities with an elevation in γ-glutamyl transpeptidase and alkaline phosphatase (ALP) along with mild elevations in aminotransferases (ALT and AST). Immunoglobulins, particularly IgM, are typically increased. Hyperbilirubinemia usually is seen once cirrhosis has developed. Thrombocytopenia, leukopenia, and anemia may be seen in patients with portal hypertension and hypersplenism. Liver biopsy shows characteristic features as described above and should be evident to any experienced hepatopathologist. Up to 10% of patients with characteristic PBC will have features of AIH (moderate to severe interphase hepatitis on biopsy, elevated ALT >5× the upper limit of normal, and elevated IgG levels) as well and are defined as having "overlap" syndrome. These patients are usually treated as PBC patients and progress to cirrhosis with the same frequency as typical PBC patients. Some patients require immunosuppressive medications as well.

Diagnosis PBC should be considered in patients with chronic cholestatic liver enzyme abnormalities. AMA testing may be negative in as many as 5–10% of patients with PBC. These patients usually are positive for other PBC-specific autoantibodies including sp100 or gp210, although these tests are not universally available. Liver biopsy is most important in this setting of AMA-negative PBC. In patients who are AMA negative with cholestatic liver enzymes, PSC should be ruled out by way of cholangiography.

TREATMENT

Primary Biliary Cholangitis

Treatment of the typical manifestations of cirrhosis is no different for PBC than for other forms of cirrhosis. Ursodeoxycholic acid (UDCA) has been shown to improve both biochemical and histologic features of the disease, thus slowing but not reversing or curing the disease. Improvement is greatest when therapy is initiated early; the likelihood of significant improvement with UDCA is low in patients with PBC who present with manifestations of cirrhosis. UDCA is given in doses of 13–15 mg/kg per d; the medication is usually well tolerated, although some patients have worsening pruritus with initiation of therapy. A small proportion of patients may have diarrhea or headache as a side effect of the drug. About 30–40% of patients with PBC do not have a satisfactory response to UDCA; more than half of these patients will have significant improvement with obeticholic acid, elafibranor or seladelpar though these medication should be avoided in the setting of cirrhosis with portal hypertension or decompensation. Patients with PBC require long-term follow-up by a physician experienced with the disease. Certain patients may need to be considered for liver transplantation should their liver disease decompensate.

The main symptoms of PBC are fatigue and pruritus. Several therapies have been tried for treatment of fatigue, but none of them has been successful; frequent naps should be encouraged. Pruritus is treated with antihistamines, narcotic receptor antagonists (naltrexone), selective serotonin reuptake inhibitors, and rifampin. Cholestyramine, a bile salt–sequestering agent, has been helpful in some patients but is somewhat tedious and difficult to take. Plasmapheresis has been used rarely in patients with severe intractable pruritus. There is an increased incidence of osteopenia and osteoporosis in patients with cholestatic liver disease, and bone density testing should be performed. Oral calcium and vitamin D are also recommended. Treatment with a bisphosphonate should be instituted when bone disease is identified. Screening for fat-soluble vitamin deficiency (A, D, E, K) should done if total bilirubin is >2 mg/dL.

PRIMARY SCLEROSING CHOLANGITIS

As in PBC, the cause of PSC remains unknown. PSC is a chronic cholestatic syndrome that is characterized by diffuse inflammation and fibrosis involving the entire biliary tree, resulting in chronic cholestasis. This pathologic process ultimately results in obliteration of both the intra- and extrahepatic biliary tree, leading to biliary cirrhosis, portal hypertension, and liver failure. The cause of PSC remains unknown despite extensive investigation into various mechanisms related to bacterial and viral infections, toxins, genetic predisposition, and immunologic mechanisms, all of which have been postulated to contribute to the pathogenesis and progression of this syndrome.

Liver biopsy changes in PSC are not pathognomonic, and establishing the diagnosis of PSC must involve imaging of the biliary tree. Pathologic changes occurring in PSC show bile duct proliferation as well as ductopenia and fibrous cholangitis (pericholangitis). Periductal fibrosis is occasionally seen on biopsy specimens and can be quite helpful in making the diagnosis. As the disease progresses, biliary cirrhosis is the end-stage manifestation of PSC.

Clinical Features The usual clinical features of PSC are those found in cholestatic liver disease, with fatigue, pruritus, steatorrhea, deficiencies of fat-soluble vitamins, and the associated consequences. As in PBC, the fatigue is profound and nonspecific. Pruritus can often be debilitating and is related to the cholestasis. The severity of pruritus does not correlate with the severity of the disease. Metabolic bone disease, as seen in PBC, can occur with PSC and should be treated (see above).

Laboratory Findings Patients with PSC typically are identified during an evaluation of abnormal liver enzymes. Most patients have at least a twofold increase in ALP and may have elevated aminotransferases as well. Albumin levels may be decreased, and prothrombin times can be prolonged at the time of diagnosis. Some degree of correction of a prolonged prothrombin time may occur with parenteral vitamin K. A small subset of patients has aminotransferase elevations >5× the upper limit of normal and may have features of AIH on biopsy indicating an overlap syndrome between PSC and AIH. Autoantibodies are frequently positive in patients with the overlap syndrome but are typically negative in patients who only have PSC. One autoantibody, the perinuclear antineutrophil cytoplasmic antibody (pANCA), is positive in ~65% of patients with PSC. Sixty to eighty percent of patients with PSC have inflammatory bowel disease, predominately ulcerative colitis (UC); thus, a colonoscopy is recommended at diagnosis.

Diagnosis The definitive diagnosis of PSC requires cholangiographic imaging. Magnetic resonance imaging (MRI) with magnetic resonance cholangiopancreatography (MRCP) is the imaging technique of choice for initial evaluation. Endoscopic retrograde cholangiopancreatography (ERCP) can be considered if the MRCP provided suboptimal images or if there is newly elevated total bilirubin or MRCP evidence of a dominant stricture. Typical cholangiographic findings in PSC are multifocal stricturing and beading involving both the intrahepatic and extrahepatic biliary tree. These strictures are typically short and with intervening segments of normal or slightly dilated bile ducts that are distributed diffusely, producing the classic beaded appearance. The gallbladder and cystic duct can be involved in up to 15% of cases. Gradually, biliary cirrhosis develops, and patients will progress to decompensated liver disease with manifestations of ascites, esophageal variceal hemorrhage, and encephalopathy.

TREATMENT

Primary Sclerosing Cholangitis

There is no specific proven treatment for PSC. Some clinicians use UDCA at "PBC dosages" of 13–15 mg/kg per d with anecdotal improvement, although no study has shown convincing evidence of clinical benefit. A study of high-dose (28–30 mg/kg per d) UDCA found it to be harmful. Endoscopic dilatation of dominant strictures can be helpful, but the ultimate treatment is liver transplantation when decompensated cirrhosis develops. Episodes of cholangitis should be treated with antibiotics and can be an indication for liver transplantation. A dreaded complication of PSC is the development of cholangiocarcinoma, which is a relative contraindication to liver transplantation.

CARDIAC CIRRHOSIS

Definition Patients with long-standing right-sided congestive heart failure may develop chronic liver injury from congestive hepatopathy sometimes resulting in cardiac cirrhosis. This is an increasingly uncommon, if not rare, cause of chronic liver disease given the advances made in the care of patients with heart failure, particularly valvular heart disease, but there has been an increase in patients with congenital heart disease particularly after the Fontan operation.

Etiology and Pathology In the case of long-term right-sided heart failure, there is an elevated venous pressure transmitted via the inferior vena cava and hepatic veins to the sinusoids of the liver, which become dilated and engorged with blood. The liver becomes enlarged and swollen, and with long-term passive congestion and relative ischemia due to poor circulation, centrilobular hepatocytes can become necrotic, leading to pericentral fibrosis. This fibrotic pattern can extend to the periphery of the lobule outward until a unique pattern of fibrosis causing cirrhosis can occur.

Clinical Features Patients typically have signs of congestive heart failure and will manifest an enlarged firm liver on physical examination. ALP levels are characteristically elevated, and aminotransferases may be normal or slightly increased, with AST usually higher than ALT. It is unlikely that patients will develop variceal hemorrhage or encephalopathy.

Diagnosis The diagnosis is usually made in someone with clear-cut cardiac disease who has an elevated ALP and an enlarged liver. Liver biopsy shows a pattern of fibrosis that can be recognized by an experienced hepatopathologist. Differentiation from Budd-Chiari syndrome (BCS) can be made by seeing extravasation of red blood cells in BCS, but not in cardiac hepatopathy. Venoocclusive disease, now termed sinusoidal obstructive syndrome (SOS), can also affect hepatic outflow and has characteristic features on liver biopsy. SOS can be seen under the circumstances of conditioning for bone marrow transplant with radiation and chemotherapy; it can also be seen with the ingestion of certain herbal teas as well as pyrrolizidine alkaloids. This is typically seen in Caribbean countries and rarely in the United States. Treatment is based on management of the underlying cardiac disease.

OTHER TYPES OF CIRRHOSIS

There are several other less common causes of chronic liver disease that can progress to cirrhosis. These include inherited metabolic liver diseases such as hemochromatosis, Wilson's disease, α_1 antitrypsin (α_1AT) deficiency, and cystic fibrosis. For these disorders, the manifestations of cirrhosis are similar, with some minor variations, to those seen in other patients with other causes of cirrhosis.

Hemochromatosis is an inherited disorder of iron metabolism that results in a progressive increase in hepatic iron deposition, which, over time, can lead to a portal-based fibrosis progressing to cirrhosis, liver failure, and hepatocellular cancer. While the frequency of hemochromatosis is relatively common, with genetic susceptibility occurring in 1 in 250 individuals, the frequency of end-stage manifestations due to the disease is relatively low, and <5% of those patients who are genotypically susceptible will go on to develop severe liver disease from hemochromatosis. Diagnosis is made with serum iron studies showing an elevated transferrin saturation and an elevated ferritin level, along with abnormalities identified by *HFE* mutation analysis. Treatment is straightforward, with regular therapeutic phlebotomy.

Wilson's disease is an inherited disorder of copper homeostasis with failure to excrete excess amounts of copper, leading to an accumulation in the liver. This disorder is relatively uncommon, affecting 1 in 30,000 individuals. Wilson's disease typically affects adolescents and young adults. Prompt diagnosis before end-stage manifestations become irreversible can lead to significant clinical improvement. Diagnosis requires determination of ceruloplasmin levels, which are low; 24-h urine copper levels, which are elevated; typical physical examination findings, including Kayser-Fleischer corneal rings; and characteristic liver biopsy findings. Treatment consists of copper-chelating medications.

α_1AT deficiency results from an inherited disorder that causes abnormal folding of the α_1AT protein, resulting in failure of secretion of that protein from the liver. It is unknown how the retained protein leads to liver disease. Patients with α_1AT deficiency at greatest risk for developing chronic liver disease have the ZZ phenotype, but only ~10–20% of such individuals will develop chronic liver disease. Diagnosis is made by determining α_1AT levels and phenotype. Characteristic periodic acid–Schiff (PAS)–positive, diastase-resistant globules are seen on liver biopsy. The only effective treatment is liver transplantation, which is curative.

TABLE 355-2 Complications of Cirrhosis

Portal hypertension	Coagulopathy
Gastroesophageal varices	Factor deficiency
Portal hypertensive gastropathy	Fibrinolysis
Splenomegaly, hypersplenism	Thrombocytopenia
Ascites	Bone disease
Spontaneous bacterial peritonitis	Osteopenia
Acute kidney injury-hepatorenal syndrome	Osteoporosis
Chronic kidney disease-hepatorenal syndrome	Osteomalacia
	Hematologic abnormalities
Hepatic encephalopathy	Anemia
Hepatopulmonary syndrome	Hemolysis
Portopulmonary hypertension	Thrombocytopenia
Malnutrition	Neutropenia

TABLE 355-3 Classification of Portal Hypertension

- Prehepatic
  - Portal vein thrombosis
  - Splenic vein thrombosis
  - Massive splenomegaly (Banti's syndrome)
- Hepatic
  - Presinusoidal
    - Schistosomiasis
    - Congenital hepatic fibrosis
  - Sinusoidal
    - Cirrhosis—many causes
    - Alcoholic hepatitis
  - Postsinusoidal
    - Hepatic sinusoidal obstruction (venoocclusive syndrome)
- Posthepatic
  - Budd-Chiari syndrome
  - Inferior vena caval webs
  - Cardiac causes
    - Restrictive cardiomyopathy
    - Constrictive pericarditis
    - Severe congestive heart failure

Cystic fibrosis is an uncommon inherited disorder affecting whites of northern European descent. A biliary-type cirrhosis can occur, and some patients derive benefit from the chronic use of UDCA.

MAJOR COMPLICATIONS OF CIRRHOSIS

These include gastroesophageal variceal hemorrhage, splenomegaly, ascites, hepatic encephalopathy, spontaneous bacterial peritonitis (SBP), hepatorenal syndrome (HRS), and hepatocellular carcinoma **(Table 355-2)**. There are also more rare complications in the pulmonary system including hepatopulmonary syndrome and portopulmonary hypertension.

■ PORTAL HYPERTENSION

Portal hypertension is defined as the elevation of the hepatic venous pressure gradient (HVPG) to >5 mmHg. The portal venous system normally drains blood from most of the GI tract including the stomach, small and large intestines, spleen, pancreas, and gallbladder. Portal hypertension is caused by a combination of two simultaneously occurring hemodynamic processes: (1) increased intrahepatic resistance to the passage of blood flow through the liver due to cirrhosis, regenerative nodules, and microthrombi, and (2) increased splanchnic blood flow secondary to vasodilation within the splanchnic vascular bed. In more advanced stages, there is also activation of neurohumoral responses and vasoconstrictive systems resulting in sodium and water retention, increased blood volume, and hyperdynamic circulatory system producing more portal hypertension. There is usually an initial stage of compensated cirrhosis with HVPG between 5 and 10 mmHg that can be asymptomatic and last for ≥10 years, but when clinically significant portal hypertension (CSPH) develops (defined as a HVPG ≥10 mmHg), there is substantial risk of decompensation with variceal bleeding, ascites, or hepatic encephalopathy. With decompensation, median mortality is <2 years.

The causes of portal hypertension are usually subcategorized as prehepatic, intrahepatic, and posthepatic **(Table 355-3)**. Prehepatic causes of portal hypertension are those affecting the portal venous system before it enters the liver; they include portal vein thrombosis and splenic vein thrombosis. Posthepatic causes encompass those affecting the hepatic veins and venous drainage to the heart; they include BCS and chronic right-sided cardiac congestion. Intrahepatic causes account for >95% of cases of portal hypertension and are represented by the major forms of cirrhosis. Intrahepatic causes of portal hypertension can be further subdivided into presinusoidal, sinusoidal, and postsinusoidal causes. Postsinusoidal causes include venoocclusive disease, whereas presinusoidal causes include congenital hepatic fibrosis and schistosomiasis. Sinusoidal causes are related to cirrhosis from various causes.

Cirrhosis is overwhelmingly the most common cause of portal hypertension in the United States. Portal vein thrombosis may contribute to portal hypertension and is most often associated with cirrhosis but may be idiopathic or can occur in association with infection, pancreatitis, or abdominal trauma. Coagulation disorders that can lead to the development of portal vein thrombosis include polycythemia vera; essential thrombocytosis; deficiencies in protein C, protein S, antithrombin III, and factor V Leiden; and abnormalities in the gene-regulating prothrombin production. Some patients may have a subclinical myeloproliferative disorder.

Clinical Features The three primary complications of portal hypertension are gastroesophageal varices with hemorrhage, ascites, and hypersplenism. Thus, patients may present with upper GI bleeding, which, on endoscopy, is found to be due to esophageal or gastric varices; with the development of ascites along with peripheral edema; or with an enlarged spleen with associated reduction in platelets and white blood cells on routine laboratory testing.

ESOPHAGEAL VARICES *Variceal hemorrhage* is an immediate life-threatening problem with a 20–30% mortality rate associated with each episode of bleeding. Over the past decade, it has become common practice to screen known cirrhotics with endoscopy to look for esophageal varices. Such screening studies have shown that approximately one-third of patients with histologically confirmed cirrhosis have varices. Approximately 5–15% of cirrhotics per year develop varices, and it is estimated that the majority of patients with cirrhosis will develop varices over their lifetimes. Furthermore, it is anticipated that roughly one-third of patients with varices will develop bleeding. Several factors predict the risk of bleeding, including the severity of cirrhosis (Child-Pugh class, Model for End-Stage Liver Disease [MELD] score); the height of wedged-hepatic vein pressure; the size of the varix; the location of the varix; and certain endoscopic stigmata, including red wale signs, hematocystic spots, diffuse erythema, bluish color, cherry red spots, or white-nipple spots. Patients with tense ascites are also at increased risk for bleeding from varices.

Diagnosis In patients with cirrhosis who are being followed chronically, the development of portal hypertension is usually revealed by the presence of thrombocytopenia; the appearance of an enlarged spleen; or the development of ascites, encephalopathy, and/or esophageal varices with or without bleeding. In previously undiagnosed patients, any of these features should prompt further evaluation to determine the presence of portal hypertension and liver disease. Varices should be identified by endoscopy. Contrasted-enhanced abdominal imaging, either by computed tomography (CT) or MRI, can be helpful in demonstrating a nodular liver and in finding changes of portal hypertension with intraabdominal collateral circulation. Rarely, the HVPG is measured by interventional radiology. Patients with a gradient >12 mmHg are at risk for variceal hemorrhage.

TREATMENT

Variceal Hemorrhage

Treatment for esophageal varices as a complication of portal hypertension is divided into two main categories: (1) primary prophylaxis and (2) prevention of rebleeding once there has been an initial variceal hemorrhage. Primary prophylaxis requires routine surveillance by endoscopy. Upper endoscopies are recommended at diagnosis of compensated cirrhosis and then every 2 years if the liver disease is active or every 3 years if inactive (alcohol cessation, viral hepatitis eradication). In the absence of thrombocytopenia and with a liver stiffness by transient elastography <20 kPa, high-risk varices are rare and thus screening is not needed. Endoscopy is also recommended at the time of hepatic decompensation. Once varices that are at increased risk for bleeding are identified, usually defined as medium or large varices or small varices with high-risk stigmata or in decompensated cirrhosis, primary prophylaxis can be achieved either through traditional nonselective beta blockade (NSBB) (propranolol or nadolol) titrated with a goal heart rate of 55–60 beats/min with systolic blood pressure >90 mmHg or by variceal band ligation. Carvedilol is becoming the NSBB of choice. Given its additional anti-α$_1$-adrenergic vasodilating properties, it more effectively lowers portal pressure, lacks the need for heart rate goals, and has emerging data that suggest it can prevent hepatic decompensation and improve survival in persons with CSPH.

Endoscopic variceal ligation (EVL) has been compared to traditional NSBB and carvedilol for primary prophylaxis against variceal bleeding, and EVL appears to have equivalent efficacy at preventing bleeding. NSBBs, especially carvedilol, are generally recommended as first-line treatment for primary prophylaxis of bleeding if tolerated, given their additional benefits. Once primary prophylaxis has been initiated, repeat endoscopy for surveillance of varices is unnecessary.

The approach to patients once they have had a variceal bleed is first to treat the acute bleed, which can be life-threatening, and then to prevent further bleeding. Treatment of acute bleeding requires both fluid and red blood cell replacement to stabilize hemodynamics. A randomized trial of restricted transfusion starting when hemoglobin is <7 g/dL with a goal hemoglobin of 7–9 g/dL, compared to a more liberal strategy, resulted in reduced early rebleeding and mortality. This strategy is recommended, although adjustments should be made based on cardiac risks and hemodynamics. Correcting an elevated prothrombin time with fresh frozen plasma is not recommended unless there is evidence of coagulopathy (bleeding at other sites such as IV lines). The use of vasoconstricting agents, usually somatostatin or octreotide, has been shown to improve initial bleeding control and reduce transfusion requirements and all-cause mortality. Prophylactic antibiotics, usually with ceftriaxone, started prior to endoscopy result in reduced infections, recurrent bleeding, and mortality. Balloon tamponade (Sengstaken-Blakemore tube or Minnesota tube) or placement of self-expandable metal stents can be used in patients who need stabilization prior to endoscopic therapy or as a bridge to transjugular intrahepatic portosystemic shunt (TIPS) after endoscopic failure. Control of bleeding can be achieved in the vast majority of cases; however, bleeding recurs in the majority of patients if definitive endoscopic therapy has not been instituted. Upper endoscopy is used as first-line treatment to diagnose the cause of the bleeding and to control bleeding acutely with EVL. When esophageal varices extend into the proximal stomach or the bleeding varices are entirely within the stomach, band ligation is often unsuccessful. In these situations, consideration for a TIPS should be made. This technique creates a portosystemic shunt by a percutaneous approach using an expandable metal stent, which is advanced under angiographic guidance to the hepatic veins and then through the substance of the liver to create a direct portocaval shunt. Encephalopathy can occur in as many as 20% of patients after TIPS and is particularly problematic in elderly patients and in patients with preexisting encephalopathy. TIPS is usually reserved for individuals who fail or are unable to receive endoscopic therapy, although there is emerging evidence that patients who are highly selected to be at high risk for rebleeding may also benefit. TIPS can sometimes be used as a bridge to transplantation, and all patients requiring TIPS should be considered for transplant evaluation. Some gastric varices are associated with a splenorenal shunt and can be effectively treated with a balloon-occluded retrograde transvenous obliteration (BRTO) of varices sometimes in combination with a TIPS. Prevention of further bleeding is usually accomplished with repeated variceal band ligation until varices are obliterated in combination with NSBB. If recurrent variceal bleeding occurs, then TIPS should be performed for long-term prevention of bleeding. Once a TIPS has been performed, there is no need for further endoscopies for variceal surveillance; however, the TIPS should be periodically monitored with Doppler ultrasound for stenosis **(Fig. 355-3)**.

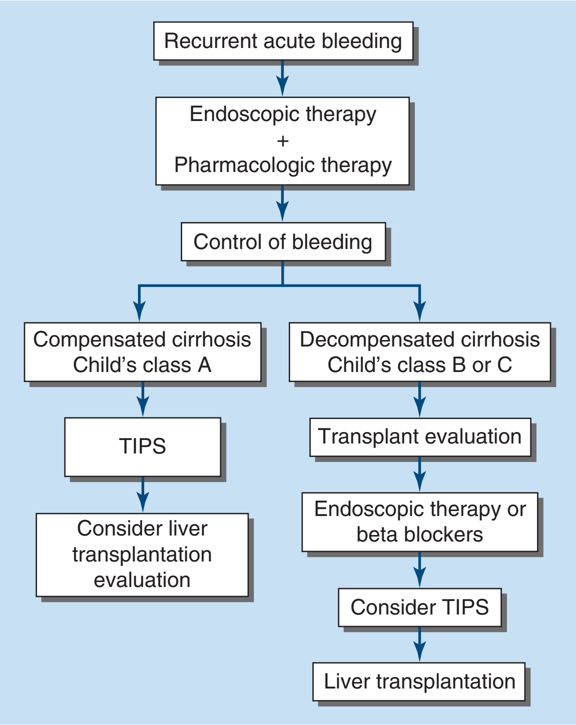

FIGURE 355-3 Management of recurrent variceal hemorrhage. This algorithm describes an approach to management of patients who have recurrent bleeding from esophageal varices. Initial therapy is generally with endoscopic therapy often supplemented by pharmacologic therapy. With control of bleeding, a decision needs to be made as to whether patients should go on to transjugular intrahepatic portosystemic shunt (TIPS; if they are Child's class A) or if they should have TIPS and be considered for transplant (if they are Child's class B or C).

■ PORTAL HYPERTENSIVE GASTROPATHY

Portal hypertensive gastropathy can cause both acute clinical GI bleeding and chronic bleeding resulting in iron-deficiency anemia. It is associated with all causes of portal hypertension and is diagnosed by characteristic endoscopy findings showing a snakeskin-like mosaic pattern of gastric mucosa often with central red or brown spots. When there is bleeding, treatment is with NSBB and iron repletion. Refractory bleeding may respond to TIPS.

■ SPLENOMEGALY AND HYPERSPLENISM

Congestive splenomegaly with hypersplenism is common in patients with portal hypertension and is usually the first indication of portal hypertension. Clinical features include the presence of an enlarged spleen on physical examination and the development of thrombocytopenia and leukopenia in patients who have cirrhosis. Some patients will have significant left-sided and left upper quadrant abdominal pain related to an enlarged spleen. Splenomegaly itself usually requires no specific treatment.

ASCITES

Definition Ascites is the accumulation of fluid within the peritoneal cavity. Overwhelmingly, the most common cause of ascites is portal hypertension related to cirrhosis; however, clinicians should remember that malignant, infectious, and cardiac causes of ascites can be present as well, and careful differentiation of these other causes is obviously important for patient care.

Pathogenesis The presence of portal hypertension contributes to the development of ascites in patients who have cirrhosis **(Fig. 355-4)**. There is an increase in intrahepatic resistance, causing increased portal pressure, but there is also vasodilation of the splanchnic arterial system, which, in turn, results in an increase in portal venous inflow. Both abnormalities result in increased production of splanchnic lymph. Vasodilating factors such as nitric oxide are responsible for the vasodilatory effect. There is activation of the renin-angiotensin-aldosterone system with the development of hyperaldosteronism and activation of the sympathetic nervous system as a consequence of a homeostatic response caused by underfilling of the arterial circulation secondary to arterial vasodilation in the splanchnic vascular bed. The renal effects of increased aldosterone and activation of the sympathetic nervous system lead to sodium retention causing fluid accumulation and expansion of the extracellular fluid volume, resulting in peripheral edema and ascites. Because the retained fluid is constantly leaking out of the intravascular compartment into the peritoneal cavity, the sensation of vascular filling is not achieved, and the process continues. Hypoalbuminemia from decreased synthetic function in a cirrhotic liver results in reduced plasma oncotic pressure and contributes to the loss of fluid from the vascular compartment into the peritoneal cavity.

Clinical Features Patients typically note an increase in abdominal girth that is often accompanied by peripheral edema. The development of ascites is often insidious, and it is surprising that some patients wait so long and become so distended before seeking medical attention. Patients usually have at least 1–2 L of fluid in the abdomen before they are aware that there is an increase. If ascitic fluid is massive, respiratory function can be compromised, causing dyspnea. Hepatic hydrothorax may also contribute to respiratory symptoms. Patients with massive ascites are often malnourished and have muscle wasting and excessive fatigue and weakness.

Diagnosis Diagnosis of ascites is by physical examination and is often aided by abdominal imaging. Patients will have bulging flanks and may have a fluid wave or the presence of shifting dullness. This is determined by taking patients from a supine position to lying on either their left or right side and noting the movement of the dullness to percussion. Subtle amounts of ascites can be detected by ultrasound or CT scanning. Hepatic hydrothorax is more common on the right side and implicates a rent in the diaphragm with free flow of ascitic fluid into the thoracic cavity.

When patients present with ascites for the first time, it is recommended that a diagnostic paracentesis be performed to characterize the fluid. This should include the determination of total protein and albumin content, blood cell counts with differential, and cultures. In the appropriate setting, amylase may be measured and cytology performed. In patients with cirrhosis, the protein concentration of the ascitic fluid is low, usually <2.5 g/dL. The serum ascites-to-albumin gradient (SAAG), calculated by subtracting the fluid albumin level from the serum albumin level, has replaced the description of exudative or transudative fluid. When the SAAG is >1.1 g/dL, the cause of the ascites is most likely due to portal hypertension; this is usually in the setting of cirrhosis. Cardiac ascites can be identified by SAAG >1.1 g/dL and ascites protein >2.5g/dL. When the SAAG is <1.1 g/dL, infectious or malignant causes of ascites should be considered. When ascitic fluid protein is very low, <1.5 g/dL, patients are at increased risk for developing SBP. A high level of red blood cells in the ascitic fluid usually signifies a traumatic tap but can also rarely occur with hepatocellular cancer or a ruptured omental varix. When the absolute level of polymorphonuclear leukocytes is >250/μL, infection is likely.

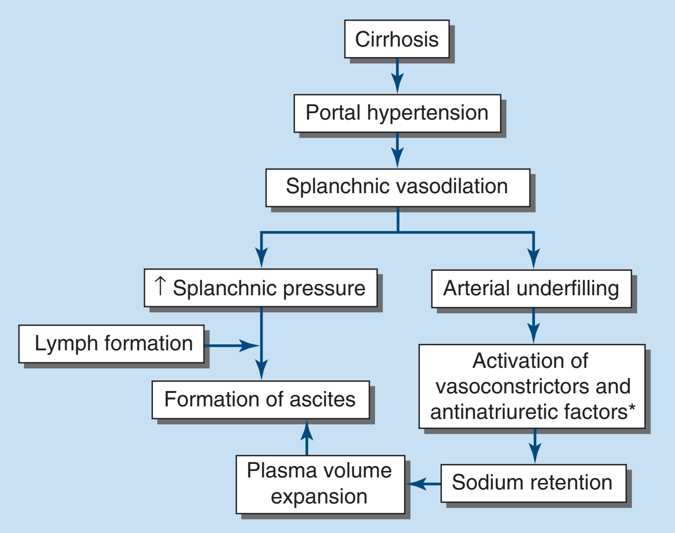

FIGURE 355-4 Development of ascites in cirrhosis. This flow diagram illustrates the importance of portal hypertension with splanchnic vasodilation in the development of ascites. *Antinatriuretic factors include the renin-angiotensin-aldosterone system and the sympathetic nervous system.

TREATMENT

Ascites

Patients with small amounts of ascites can usually be managed with dietary sodium restriction alone. Most average diets in the United States contain 6–8 g of sodium per day, and if patients eat at restaurants or fast-food outlets, the amount of sodium in their diet can exceed this amount. Thus, it is often extremely difficult to get patients to change their dietary habits to ingest 2 g of sodium per day, equivalent to slightly more than three-quarters of a teaspoon of salt, which is the recommended amount. Sodium educational pamphlets are helpful. Often, a simple recommendation is to eat fresh or frozen foods, avoiding canned or processed foods. When a moderate amount of ascites is present, diuretic therapy is usually necessary. Traditionally, spironolactone at 100 mg/d as a single dose is started, and furosemide may be added at 40 mg/d, particularly in patients who have peripheral edema. Failure of the diuretics suggests that patients may not be compliant with a low-sodium diet. If compliance is confirmed and ascitic fluid is not being mobilized, there should be incremental increases in spironolactone to a maximum of 400 mg/d and furosemide to 160 mg/d. If a large amount of ascites is still present on diuretics in patients who are compliant with a low-sodium diet, then they are defined as having *refractory ascites*, and alternative treatment modalities including repeated large-volume paracentesis (LVP) or a TIPS procedure should be considered **(Fig. 355-5)**. After LVP of ≥5 L, IV 25% albumin at a dose of 6–8 g/L of removed ascites should be given to prevent circulatory dysfunction. Multiple studies have shown that TIPS, although effective at managing the ascites, does not improve survival. Unfortunately, TIPS is often associated with an increased frequency of hepatic encephalopathy and must be considered carefully on a case-by-case basis. The prognosis for patients with cirrhosis with ascites is poor, and some studies have shown that <50% of patients survive 2 years after the onset of ascites. Thus, there should be consideration for liver transplantation in patients with ascites. Patients with cirrhosis and ascites are at increased risk for renal failure from certain medications including nonsteroidal anti-inflammatory drugs and aminoglycosides; therefore, these medications should generally be avoided. Angiotensin-converting enzyme inhibitors and angiotensin receptor blockers should be used cautiously with close monitoring of blood pressure and renal function.

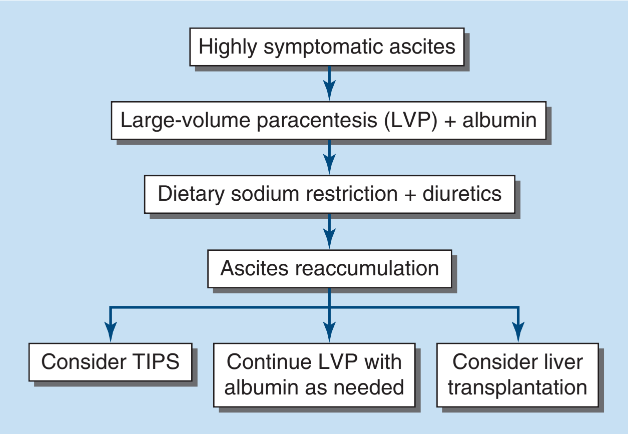
Highly symptomatic ascites
Large-volume paracentesis (LVP) + albumin
Dietary sodium restriction + diuretics
Ascites reaccumulation
Consider TIPS
Continue LVP with albumin as needed
Consider liver transplantation

FIGURE 355-5 Treatment of refractory ascites. In patients who develop azotemia in the course of receiving diuretics in the management of their ascites, some will require repeated large-volume paracentesis (LVP), some may be considered for transjugular intrahepatic portosystemic shunt (TIPS), and some would be good candidates for liver transplantation. These decisions are all individualized.

SPONTANEOUS BACTERIAL PERITONITIS

SBP is a common and severe complication of ascites characterized by spontaneous infection of the ascitic fluid without an intraabdominal source. In hospitalized patients with cirrhosis and ascites, SBP can occur in up to 30% of individuals and can have a 25% in-hospital mortality rate. Bacterial translocation is the presumed mechanism for development of SBP, with gut flora traversing the intestine into mesenteric lymph nodes, leading to bacteremia and seeding of the ascitic fluid. The most common organisms are *Escherichia coli* and other gut bacteria; however, gram-positive bacteria, including *Streptococcus viridans*, *Staphylococcus aureus*, and *Enterococcus* spp., can also be found. If more than two organisms are identified, secondary bacterial peritonitis due to a perforated viscus should be considered. The diagnosis of SBP is made when the fluid sample has an absolute neutrophil count >250/μL. Bedside cultures should be obtained by direct injection of ascitic fluid into blood culture bottles. Patients with ascites may present with fever, altered mental status, elevated white blood cell count, abdominal pain or discomfort, and acute kidney injury, or they may present without any of these features. Therefore, it is necessary to have a high degree of clinical suspicion, and peritoneal taps are recommended for cirrhosis patients hospitalized with ascites and cirrhosis complications or signs of infection. Treatment is commonly with intravenous third-generation cephalosporin for 5 days. In addition, intravenous albumin (1.5 g/kg body weight on day 1 and 1.0 g/kg on day 3) has been shown to reduce the risk of renal failure and to improve survival. In patients with variceal hemorrhage, the frequency of SBP is significantly increased, and prophylaxis against SBP is recommended at presentation with upper GI bleeding. Furthermore, in patients who have had an episode of SBP and recovered, quinolone antibiotic prophylaxis should be given to prevent recurrent SBP.

HEPATORENAL SYNDROME

HRS is a form of functional renal failure without renal pathology that occurs in ~10% of patients with advanced cirrhosis or acute liver failure. There are marked disturbances in the arterial renal circulation in patients with HRS; these include an increase in vascular resistance accompanied by a reduction in systemic vascular resistance. The reason for renal vasoconstriction is most likely multifactorial and is poorly understood. The diagnosis is made usually in the presence of a large amount of ascites in patients who have a stepwise progressive increase in creatinine. Acute kidney injury (AKI) is defined by a 0.3-mg/dL rise in creatinine over 48 h or a 50% rise in creatinine from baseline. HRS-AKI, the new term for type 1 HRS, is characterized by a progressive rapid impairment in renal function. Type 2 HRS, now termed HRS–chronic kidney disease (CKD), is characterized by a reduction in glomerular filtration rate with an elevation of serum creatinine level, but it is stable and is associated with a better outcome than that of type 1 HRS.

HRS-AKI requires exclusion of other causes of acute renal failure, most notably volume depletion. Diuretics should be stopped, and infusion of albumin 1 g/kg per d for 48 h without significant improvements is required. Other causes of AKI, including intrinsic (acute tubular necrosis, acute interstitial nephritis, and glomerulonephritis) and obstructive kidney disease, should be excluded. Treatment is with vasoconstrictors such as terlipressin or, if not available, low-dose norepinephrine, which requires intensive care unit monitoring. Midodrine, an α-agonist, along with octreotide and intravenous albumin are also commonly used in the United States but are now third line. If treatment fails, then renal replacement therapy can be initiated. The best therapy for HRS is liver transplantation; recovery of renal function is typical in this setting. In patients with either AKI-HRS or CKD-HRS, the prognosis is poor unless transplant can be achieved.

HEPATIC ENCEPHALOPATHY

Hepatic encephalopathy is a serious complication of chronic liver disease and is broadly defined as an alteration in mental status and cognitive function occurring as a consequence of liver failure. In severe acute liver injury, the development of encephalopathy is a requirement for a diagnosis of acute liver failure and can be seen in association with life-threatening brain edema, which is not a feature in chronic liver disease. Hepatic encephalopathy is much more commonly seen in patients with chronic liver disease. Gut-derived neurotoxins that are not removed by the liver because of vascular shunting and decreased hepatic mass reach the brain and cause the symptoms known as hepatic encephalopathy. Ammonia levels are typically elevated, but the correlation between severity of liver disease and height of ammonia levels is often poor, and most hepatologists do not rely on ammonia levels to make a diagnosis or follow clinical progress. Other compounds and metabolites that may contribute to the development of encephalopathy include certain false neurotransmitters and mercaptans.

Clinical Features In acute liver failure, changes in mental status can occur rapidly. Brain edema can be seen in these patients, with severe encephalopathy associated with swelling of the gray matter. Cerebral herniation is a feared complication of brain edema in acute liver failure, and treatment to decrease edema is with hypertonic saline or mannitol.

In patients with cirrhosis, encephalopathy is often found as a result of precipitating events such as volume depletion, gastrointestinal bleeding, hyponatremia, infection, or constipation. Patients may be confused or exhibit a change in personality. They may be quite violent and difficult to manage; alternatively, patients may be very sleepy and difficult to rouse. If patients have ascites, this should be tapped to rule out infection. Evidence of GI bleeding should be sought, and patients should be appropriately hydrated. Electrolytes should be measured, and abnormalities corrected. In patients presenting with encephalopathy, asterixis is often present. Asterixis can be elicited by having patients extend their arms and bend their wrists back. Patients who are encephalopathic have a "liver flap"—that is, a sudden forward movement of the wrist. This requires patients to be able to cooperate with the examiner. Alternative causes for altered mental status should also be considered.

The diagnosis of hepatic encephalopathy is clinical and requires an experienced clinician to recognize and put together all the various features. Often when patients have encephalopathy for the first time, they (and/or their caregivers) are unaware of what is transpiring, but once they have been through the experience, they can identify when this is developing in subsequent situations and can often self-medicate to prevent the development or worsening of encephalopathy.

TREATMENT

Hepatic Encephalopathy

Treatment is multifactorial and includes management of the abovementioned precipitating factors. Sometimes hydration and correction of electrolyte imbalance are all that is necessary. In the past, restriction of dietary protein was used; however, the negative impact of that maneuver on overall nutrition is thought to outweigh the benefit, and it is thus strongly discouraged. The mainstay of treatment for encephalopathy is to use lactulose, a nonabsorbable disaccharide, which results in colonic acidification. Catharsis ensues, contributing to the elimination of nitrogenous products in the gut that are responsible for the development of encephalopathy. The goal of lactulose therapy is to promote two to three soft stools per day. Patients are asked to titrate their amount of ingested lactulose to achieve the desired effect. Lactulose is usually continued after the initial episode of encephalopathy. Poorly absorbed antibiotics are often used as adjunctive therapies for patients who have a difficult time with lactulose or in those with recurrent episodes. Rifaximin has replaced neomycin and metronidazole (because of their significant toxicity). The dose is 550 mg twice daily and is very effective in preventing recurrent encephalopathy. In patients with recurrent symptoms despite treatment, closure of large portosystemic shunts can be helpful. Zinc supplementation is sometimes helpful and is relatively harmless. The development of encephalopathy in patients with chronic liver disease is a poor prognostic sign, but this complication can be managed in the vast majority of patients.

ACUTE-ON-CHRONIC LIVER FAILURE

Acute-on-chronic liver failure (ACLF) is a recently recognized clinical syndrome that is characterized by acutely decompensating cirrhosis with associated failure of one or more organ systems, including liver, kidneys, brain, lung, circulatory system, and coagulation. It occurs in the setting of chronic liver disease almost always with cirrhosis and is a major cause of mortality in persons with cirrhosis. ACLF is precipitated by either direct injury to the liver (most commonly from alcoholic hepatitis and less often from new or flaring viral hepatitis, autoimmune hepatitis, or drug-induced liver injury) or systemic effects (most commonly bacterial or fungal infection and less often GI bleeding or the postoperative state), resulting in a marked systemic inflammatory response followed by organ failure and is analogous to sepsis syndrome. Mortality at 28 days ranges from 20 to 70% and increases with the number of organ failures. Clinicians should search for and treat precipitating causes of ACLF, determine if intensive care unit (ICU) care is needed, and consider immediate referral for liver transplantation. Complications of cirrhosis should be treated as described elsewhere in this chapter. If after 3–7 days of ICU support, there continue to be four or more organ failures and liver transplantation is not an option, consideration for a transition to palliative care is recommended.

LIVER-LUNG SYNDROMES

Hepatopulmonary syndrome (HPS) is characterized by arterial hypoxemia in a patient with cirrhosis without significant lung disease. The liver disease causes intrapulmonary vascular dilations, resulting in blood shunting past alveoli and significant ventilation-perfusion mismatch. Clinical symptoms include dyspnea and platypnea. HPS is common, occurring in 4–32% of patients with cirrhosis; however, it is often mild. Diagnosis involves demonstrating hypoxemia, without evidence of significant lung disease, and shunt on bubble echocardiography. Treatment is with oxygen supplementation and liver transplantation.

Portopulmonary hypertension (PPHT) is pulmonary hypertension in a patient with portal hypertension. The portal hypertension results in the production of vasoconstrictor substances that affect the pulmonary artery. Many patients are asymptomatic, especially early in the disease; however, they later can develop dyspnea on exertion and fatigue. PPHT is rare, occurring in ~5% of patients with advanced cirrhosis. Diagnosis includes initial identification on echocardiogram and confirmation on right heart catheterization showing elevated mean pulmonary artery pressure, elevated pulmonary vascular resistance, and normal pulmonary capillary wedge pressure. Prognosis is poor, although liver transplantation after effective reduction in pulmonary artery pressure with vasodilatory medications can be effective.

MALNUTRITION IN CIRRHOSIS

Because the liver is principally involved in the regulation of protein and energy metabolism in the body, patients with decompensated cirrhosis often develop malnutrition, insufficient intake of nutrients, which manifests clinically as sarcopenia, a loss in muscle mass and function resulting in frailty, decreased physical reserve, and increased susceptibility to health stressors. In the setting of cirrhosis and subsequent hepatic decompensation, patients become more catabolic with increased energy needs but often have decreased caloric intake due to anorexia, early satiety, ascites, restrictive diets (low sodium, potassium, and fluid) and hepatic encephalopathy. Obesity and edema may mask underlying sarcopenia. Lack of physical activity can further exacerbate the sarcopenia and resulting frailty. Sarcopenia and frailty have been associated with increased resource utilization and risk of death and poorer quality of life. Guidelines recommend (1) education for patients and families both with asymptomatic compensated and symptomatic decompensated cirrhosis on appropriate nutritional intake (both calories and protein) and the benefits of regular exercise; (2) periodic clinical assessments for malnutrition, sarcopenia, and frailty with increasing frequency as the patient decompensates; and (3) early referrals to registered dietician and physical therapy to develop a personalized therapeutic plan. Generally, patients with cirrhosis should consume 35 kcal/kg per d based on actual weight with estimated discounting for edema and ascites and 1.2–1.5 g/kg per d of protein based upon ideal body weight. Caloric intake goals should be reduced for patients with body mass index over 35 kg/m^2. Multiple small meals are recommended throughout waking hours to minimize muscle wasting during the fasted state. Relaxation in sodium restriction should be considered in those who are not meeting caloric intake goals. Optimization of care for hepatic encephalopathy and ascites is also beneficial. General exercise recommendations include 150–300 min of aerobic exercise a week and 2 days of muscle-strengthening exercises per week, though these recommendations should be tailored, particularly in decompensated cirrhosis.

ABNORMALITIES IN COAGULATION

Coagulation disorders in cirrhosis are poorly understood, and typical clinical measures of coagulation, such as the prothrombin time and international normalized ratio, are not reliable measures of clotting ability. There is decreased synthesis of both pro- and anticoagulant factors and thus some rebalancing in coagulation; however, the coagulation cascade can easily tip toward thrombosis or bleeding. In addition, patients may have thrombocytopenia from hypersplenism due to portal hypertension and some platelet dysfunction, which is counterbalanced with increased von Willebrand factor. Adequate thrombin formation can occur with platelet levels from cirrhosis patients >50,000–60,000/L. Synthesis of vitamin K–dependent clotting factors II, VII, IX, and X is diminished in patients with chronic cholestatic syndromes because absorption of vitamin K requires good bile flow. Intravenous or intramuscular vitamin K can quickly correct this abnormality. Overall, the status of coagulation in a cirrhotic patient needs to be judged clinically rather than relying on current laboratory tests. Standard prophylaxis for deep venous thrombosis when hospitalized is generally recommended. Routine correction of international normalized ratio before procedures or with variceal bleeding is generally not required.

BONE DISEASE IN CIRRHOSIS

Osteoporosis is common in patients with chronic cholestatic liver disease because of malabsorption of vitamin D and decreased calcium ingestion. The rate of bone resorption exceeds that of new bone formation in patients with cirrhosis, resulting in bone loss. Dual-energy x-ray absorptiometry (DEXA) is a useful method for determining osteoporosis or osteopenia. When a DEXA scan shows osteoporosis, treatment with bisphosphonates is effective.

HEMATOLOGIC ABNORMALITIES IN CIRRHOSIS

Numerous hematologic manifestations of cirrhosis are present, including anemia from a variety of causes including hypersplenism, hemolysis, iron deficiency, and perhaps folate deficiency from malnutrition. Macrocytosis is a common abnormality in red blood cell morphology seen in patients with chronic liver disease, and neutropenia may be a result of hypersplenism.

FURTHER READING

Bhattacharya D et al: Hepatitis C guidance update 2023 update: AASLD-IDSA recommendations for testing, managing, and treating hepatitis C virus infection. Clin Infect Dis Ciad319:1, 2023.

Biggins SW et al: Diagnosis, evaluation and management of ascites, spontaneous bacterial peritonitis and hepatorenal syndrome: 2021 practice guidance by the American Association for the Study of Liver Diseases. Hepatology 74:1014, 2021.

EASL: Clinical practice guidelines on acute-on-chronic liver failure. J Hepatol 79:461, 2023.

Kaplan D et al: AASLD practice guidance of risk stratification and management of portal hypertension and varices in cirrhosis. Hepatology 79:1180, 2024.

Lai JC et al: Malnutrition, frailty, and sarcopenia in patients with cirrhosis: 2021 practice guidance by the American Association for the Study of Liver Diseases. Hepatology 74:1611, 2021.

Terrault NA et al: Update on prevention, diagnosis and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. Hepatology 67:1560, 2018.

Vilstrup H et al: Hepatic encephalopathy in chronic liver disease: 2014 Practice Guideline by the American Association for the Study of Liver Diseases and the European Association for the Study of the Liver. Hepatology 60:715, 2014.

356 Liver Transplantation

Emily D. Bethea, Raymond T. Chung, Jules L. Dienstag

Liver transplantation—the replacement of the native, diseased liver by a normal organ (allograft)—has matured from an experimental procedure reserved for desperately ill patients to an accepted, lifesaving operation applied more optimally in the natural history of end-stage liver disease. The preferred and technically most advanced approach is *orthotopic transplantation*, in which the native organ is removed and the donor organ is inserted in the same anatomic location. Pioneered in the 1960s by Thomas Starzl at the University of Colorado and, later, at the University of Pittsburgh and by Roy Calne in Cambridge, England, liver transplantation is now performed routinely worldwide. Success measured as 1-year survival has improved from ~30% in the 1970s to >90% today. These improved prospects for prolonged survival resulted from refinements in operative technique, improvements in organ procurement and preservation, advances in immunosuppressive therapy, and, perhaps most influentially, more enlightened patient selection and timing. Despite the perioperative morbidity and mortality, the technical and management challenges of the procedure, and its costs, liver transplantation has become the approach of choice for selected patients whose chronic or acute liver disease is progressive, life-threatening, and unresponsive to medical therapy. Based on the current level of success, the number of liver transplants has continued to grow each year; in 2022, 9528 patients received liver allografts in the United States—8925 deceased donor (94%) and 603 living donor (6%). Still, the demand for new livers continues to outpace availability; as of September 2023, 10,285 patients in the United States were on a waiting list for a donor liver.

INDICATIONS

Potential candidates for liver transplantation are children and adults who, in the absence of contraindications (see below), suffer from severe, irreversible liver disease for which alternative medical or surgical treatments have been exhausted or are unavailable. *Timing of the operation is of critical importance*. Indeed, improved timing and better patient selection are felt to have contributed more to the increased success of liver transplantation in the 1980s and beyond than all the impressive technical and immunologic advances combined. Although the disease should be advanced, and although opportunities for spontaneous or medically induced stabilization or recovery should be allowed, the procedure should be done sufficiently early to give the surgical procedure a fair chance for success. Ideally, transplantation should be considered in patients with end-stage liver disease who are experiencing or have experienced a life-threatening complication of hepatic decompensation or whose quality of life has deteriorated to unacceptable levels. Although patients with well-compensated cirrhosis can survive for many years, many patients with quasi-stable chronic liver disease have much more advanced disease than may be apparent. As discussed below, the better the status of the patient prior to transplantation, the higher will be its anticipated success rate. The decision about *when* to transplant is complex and requires the combined judgment of an experienced team of hepatologists, transplant surgeons, anesthesiologists, and specialists in support services, not to mention the well-informed consent of the patient and the patient's family.

TRANSPLANTATION IN CHILDREN

Indications for transplantation in children are listed in Table 356-1. The most common is *biliary atresia*. *Inherited or genetic disorders of metabolism* associated with liver failure constitute another major indication for transplantation in children and adolescents. In Crigler-Najjar disease type I and in certain hereditary disorders of the urea cycle and of amino acid or lactate-pyruvate metabolism, transplantation may be the only way to prevent impending deterioration of central nervous system function, despite the fact that the native liver is structurally normal. Combined heart and liver transplantation has yielded dramatic improvement in cardiac function and in cholesterol levels in children with homozygous familial hypercholesterolemia; combined liver and

TABLE 356-1 Indications for Liver Transplantation

CHILDREN	ADULTS
Biliary atresia	Primary biliary cholangitis
Neonatal hepatitis	Primary sclerosing cholangitis
Congenital hepatic fibrosis	Caroli's disease[a]
Alagille's syndrome[b]	Secondary biliary cirrhosis
Byler's disease[c]	Autoimmune hepatitis
Inherited disorders of metabolism	Hemochromatosis-associated cirrhosis
Wilson's disease	α_1 Antitrypsin deficiency
Tyrosinemia	Metabolic dysfunction–associated steatohepatitis (MASH)[d]
Glycogen storage diseases	Alcohol-associated cirrhosis
Lysosomal storage diseases	Severe alcohol-associated hepatitis
Protoporphyria	Cryptogenic cirrhosis
Crigler-Najjar disease type I	Chronic viral hepatitis with cirrhosis
Familial hypercholesterolemia	Hepatic venous outflow obstruction (Budd-Chiari syndrome)
Primary hyperoxaluria type I	
Hemophilia	Acute liver failure (ALF)
	Hepatocellular carcinoma
	Select cases for the following indications:
	Hepatic adenomas
	Familial amyloidosis
	Hepatic epithelioid hemangioendothelioma (HEHE)
	Erythropoietic protoporphyria (EPP)
	Metastatic neuroendocrine tumors
	Polycystic liver disease

[a]Multiple cystic dilatations of the intrahepatic biliary tree. [b]Arteriohepatic dysplasia, with paucity of bile ducts, and congenital malformations, including pulmonary stenosis. [c]Intrahepatic cholestasis, progressive liver failure, and mental and growth retardation. [d]Formerly nonalcoholic steatohepatitis (NASH).

kidney transplantation has been successful in patients with primary hyperoxaluria type I. In hemophiliacs with transfusion-associated hepatitis and liver failure, liver transplantation has been associated with recovery of normal factor VIII synthesis.

TRANSPLANTATION IN ADULTS

Liver transplantation is indicated for end-stage *cirrhosis* of all causes (Table 356-1). In *sclerosing cholangitis* and *Caroli's disease* (multiple cystic dilatations of the intrahepatic biliary tree), recurrent infections and sepsis associated with inflammatory and fibrotic obstruction of the biliary tree may be an indication for transplantation. Because prior biliary surgery complicates and is a relative contraindication for liver transplantation, surgical diversion of the biliary tree has been all but abandoned for patients with sclerosing cholangitis. Currently, the most common indication for liver transplantation is steatotic liver disease (SLD), which encompasses alcohol-associated liver disease (ALD) and metabolic dysfunction–associated steatotic liver disease (MASLD), formerly known as nonalcoholic fatty liver disease (NAFLD). Patients with alcohol-associated cirrhosis can be considered as candidates for transplantation if they meet strict criteria for abstinence and reform; however, these criteria still do not prevent return to alcohol use in up to a quarter of cases. In highly selected cases in a limited but growing number of centers, transplantation for severe *acute* alcohol-associated hepatitis has been performed with success; however, because patients with acute alcohol-associated hepatitis are still actively using alcohol and because continued alcohol use remains a concern, acute alcohol-associated hepatitis is not a routine indication for liver transplantation. Patients with chronic hepatitis C have early allograft and patient survival comparable to those of other subsets of patients after transplantation; however, they have historically experienced a higher frequency of allograft failure beyond 5 years due to universal reinfection and progressive fibrosis in cases of recurrent hepatitis C is insidiously progressive, with allograft cirrhosis and failure occurring at a higher frequency beyond 5 years. Fortunately, with the introduction of highly effective direct-acting antiviral (DAA) agents targeting hepatitis C virus (HCV), allograft outcomes have improved substantially. In patients with chronic hepatitis B, in the absence of measures to prevent recurrent hepatitis B, survival after transplantation is reduced by ~10–20%; however, prophylactic use of hepatitis B immune globulin (HBIg) during and after transplantation increases the success of transplantation to a level comparable to that seen in patients with nonviral causes of liver decompensation. Specific oral antiviral drugs (e.g., entecavir, tenofovir disoproxil fumarate, tenofovir alafenamide) **(Chap. 352)** can be used both for prophylaxis against and for treatment of recurrent hepatitis B, facilitating further the management of patients undergoing liver transplantation for end-stage hepatitis B; most transplantation centers rely on antiviral drugs with or without HBIg to manage patients with hepatitis B. Issues of disease recurrence are discussed in more detail below. In patients who undergo transplantation for *hepatic vein thrombosis* (*Budd-Chiari syndrome*), postoperative anticoagulation is essential, as is the treatment of any myeloproliferative disorder. Patients with acute liver failure are candidates for liver transplantation, provided a donor organ can be located quickly, before life-threatening complications—including cerebral edema—set in. Patients with nonmetastatic primary hepatobiliary tumors—primary hepatocellular carcinoma (HCC), cholangiocarcinoma, hepatoblastoma, angiosarcoma, epithelioid hemangioendothelioma, and multiple or massive hepatic adenomata—have undergone liver transplantation; however, for some hepatobiliary malignancies, overall survival is significantly lower than that for other categories of liver disease. Most transplantation centers have reported 5-year recurrence-free survival rates in patients with unresectable HCC for single tumors <5 cm in diameter or for three or fewer lesions all <3 cm comparable to those seen in patients undergoing transplantation for nonmalignant indications. Consequently, liver transplantation is currently restricted to patients whose hepatic malignancies meet these criteria. Expanded criteria for patients with HCC continue to be evaluated in exploratory clinical trials. Because the likelihood of recurrent cholangiocarcinoma is very high, only highly selected patients with limited disease are being evaluated for transplantation after intensive chemotherapy and radiation.

CONTRAINDICATIONS

Absolute contraindications for transplantation include life-threatening systemic diseases, uncontrolled extrahepatic bacterial or fungal infections, preexisting advanced cardiovascular or pulmonary disease, multiple uncorrectable life-threatening congenital anomalies, metastatic malignancy, and active drug or alcohol use disorders **(Table 356-2)**. Because carefully selected patients in their sixties and even seventies have undergone transplantation successfully, advanced age per se is no longer considered an absolute contraindication; however, in older patients, a more thorough preoperative evaluation should be undertaken to exclude ischemic cardiac disease and other comorbid conditions. Still, advanced age (>70 years) should be considered a *relative contraindication*—that is, a factor to be considered with other relative contraindications. Other relative contraindications include extensive portal vein thrombosis, preexisting renal disease not associated with liver disease (which may prompt consideration of combined liver and kidney transplantation), intrahepatic or biliary sepsis, severe hypoxemia (PO_2 <50 mmHg) resulting from right-to-left intrapulmonary shunts, portopulmonary hypertension with high mean pulmonary artery pressures (>35 mmHg), uncontrolled psychiatric conditions, and lack of sufficient social supports. Any one of these relative contraindications may be insufficient in and of itself to preclude transplantation. For example, the problem of portal vein thrombosis may be overcome in certain cases by constructing a graft from the donor liver portal vein to the recipient's superior mesenteric vein. Now that combination antiretroviral therapy has dramatically improved the survival of persons with HIV infection **(Chap. 208)** and because end-stage liver disease caused by chronic hepatitis C and B has emerged as a serious source of morbidity and mortality in the HIV-infected population, liver transplantation has now been performed successfully in selected HIV-positive persons who have excellent control of HIV infection. Selected patients with CD4+ T-cell counts >100/μL and with pharmacologic suppression of HIV viremia have undergone transplantation for end-stage liver disease. HIV-infected persons who have received liver allografts for end-stage liver disease resulting from chronic hepatitis B have experienced survival rates comparable to those of HIV-negative persons undergoing transplantation for the same indication. In contrast, until recently, recurrent HCV infection in the allograft has limited long-term success in persons with HCV-related end-stage

TABLE 356-2 Contraindications to Liver Transplantation

ABSOLUTE	RELATIVE
Uncontrolled extrahepatobiliary infection	Advanced age[a]
Active, untreated sepsis	Prior extensive hepatobiliary surgery
Life-limiting congenital anomalies	Extensive portal vein thrombosis
Cholangiocarcinoma (except those tumors that fit into protocols)	Renal failure not attributable to liver disease (consider dual organ transplantation)
Advanced cardiopulmonary disease	Extrahepatobiliary malignancy Severe obesity Severe malnutrition/wasting Medical noncompliance
AIDS	HIV seropositivity with failure to control HIV viremia or CD4 <100/μL
Life-threatening systemic diseases	Severe hypoxemia secondary to right-to-left intrapulmonary shunts (PO_2 <50 mmHg) Severe pulmonary hypertension (mean pulmonary artery pressure >35 mmHg)
Active substance use disorder	Uncontrolled psychiatric disorder

[a]For patients 70 years and older, comprehensive assessment to exclude concurrent comorbidities, in particular cardiovascular compromise, is indicated.

liver disease. Again, the availability of DAA agents targeting HCV (see below and **Chap. 352**) has improved allograft outcomes substantially.

TECHNICAL CONSIDERATIONS

DECEASED-DONOR SELECTION

Deceased-donor livers for transplantation are procured primarily from victims of head trauma, termed donation after brain death (DBD). Cardiovascular and respiratory functions in these donors are maintained artificially until the liver can be removed. Organs from brain-dead donors up to age 60 are acceptable if the following criteria are met: hemodynamic stability, adequate oxygenation, absence of bacterial or fungal infection, absence of abdominal trauma, and absence of hepatic dysfunction. Historically serologic exclusion of infections with hepatitis B virus (HBV), HCV, and HIV was required, but organs from donors infected with HBV, HCV, and HIV are now used in select cases when matched appropriately with the recipient after thorough informed consent and antiviral medication planning. Transplantation of organs procured from deceased donors who have succumbed to cardiac death, termed donation after circulatory death (DCD), can be performed successfully under selected circumstances, when ischemic time is minimized and liver histology preserved. Encouraging improvements in normothermic ex vivo liver perfusion techniques may make broader use of these organs possible. Compatibility in ABO blood group and organ size between donor and recipient are important considerations in donor selection; however, ABO-incompatible, split-liver, or reduced-donor-organ allografts can be performed in emergencies or marked donor scarcity. Tissue typing for human leukocyte antigen (HLA) matching is not required, and preformed cytotoxic HLA antibodies do not preclude liver transplantation. Following perfusion with cold electrolyte solution, the donor liver is removed and packed in ice. The use of University of Wisconsin (UW) solution, rich in lactobionate and raffinose, has permitted the extension of cold ischemic time up to 20 h; however, 12 h may be a more reasonable limit. Improved techniques for harvesting multiple organs from the same donor have increased the availability of donor livers, but the availability of donor livers is far outstripped by the demand. Currently in the United States, all donor livers are distributed through a nationwide organ-sharing network (United Network for Organ Sharing [UNOS]) designed to allocate available organs based on regional considerations and recipient acuity. Recipients who have the highest disease severity generally have the highest priority, but allocation strategies that balance highest urgency against best outcomes continue to evolve to distribute deceased-donor organs most effectively. Allocation based on the Child-Turcotte-Pugh (CTP) score, which uses five clinical variables (encephalopathy stage, ascites, bilirubin, albumin, and prothrombin time) and waiting time, has been replaced by allocation based on urgency alone, calculated using the Model for End-Stage Liver Disease (MELD) score. The MELD score is based on a mathematical model that includes serum bilirubin, albumin, sodium, creatinine, and prothrombin time expressed as international normalized ratio (INR); gender assigned at birth is included in the MELD calculation for females **(Table 356-3)**. The MELD scale is continuous with scores ranging between 6 and 40; higher values correlate with increased illness severity. On the lower end of this scale, the value of proceeding to transplantation is limited, because liver recipients with MELD scores <15 experience higher posttransplantation mortality rates than similarly classified patients who remain on the waiting list. The MELD score has undergone modifications over time, including the development of MELD-Na in 2016 and MELD-3.0 in 2023 to improve model accuracy in predicting short-term survival and prioritizing allocation. Neither waiting time (except as a tie breaker between two potential recipients with the same MELD scores) nor posttransplantation outcome is taken into account, but use of the MELD score has been shown to reduce waiting list mortality, to reduce waiting time prior to transplantation, to be the best predictor of pretransplantation mortality, to satisfy the prevailing view that medical need should be the decisive determinant, and to eliminate both the subjectivity inherent in the CTP scoring system (presence and degree of ascites and hepatic encephalopathy) and the differences in waiting times among different regions of the country.

TABLE 356-3 United Network for Organ Sharing (UNOS) Liver Transplantation Waiting List Criteria

Status 1	Acute liver failure (including primary graft nonfunction and hepatic artery thrombosis)[a]
The Model for End-Stage Liver Disease (MELD) score, on a continuous scale, determines allocation of the remainder of donor organs[b].	
Candidates who are **at least 18 years old** at the time of registration receive an initial MELD score equal to: $1.33 \text{ (if female)} + [4.56 \times \log_e(\text{bilirubin})] + [0.82 \times (137 - \text{sodium})] - [0.24 \times (137 - \text{sodium}) \times \log_e(\text{bilirubin})] + [9.09 \times \log_e(\text{INR})] + [11.14 \times \log_e(\text{creatinine})] + [1.85 \times (3.5 - \text{albumin})] - [1.83 \times (3.5 - \text{albumin}) \times \log_e(\text{creatinine})] + 6$ Candidates who are currently **at least 12 years** old and were less than 18 years old at the time of registration receive a MELD score equal to: $[4.56 \times \log_e(\text{bilirubin})] + [0.82 \times (137 - \text{sodium})] - [0.24 \times 137 - \text{sodium}) \times \log_e(\text{bilirubin})] + [9.09 \times \log_e(\text{INR})] + 11.14 \times \log_e(\text{creatinine})] + [1.85 \times (3.5 - \text{albumin})] - [1.83 \times (3.5 - \text{albumin})] - [1.83 \times (3.5 - \text{albumin}) \times \log_e(\text{creatinine})] + 7.33$ The **PELD calculator is for candidates under the age of 12.**	
Online calculators to determine MELD scores are available, such as the following: *https://optn.transplant.hrsa.gov/resources/allocation-calculators/meld-calculator/*	

[a]For children <18 years of age, status 1 includes acute or chronic liver failure plus hospitalization in an intensive care unit or inborn errors of metabolism. Status 1 is retained for those persons with acute liver failure and supersedes the MELD score. MELD scale is continuous, with 34 levels ranging between 6 and 40 (scores above 40 are categorized as 40). [b]In certain cases, the natural MELD score may not represent the severity of illness or need for transplantation; in such cases, an exception MELD score may be granted if select criteria are met upon review by the National Liver Review Board (NLRB): *https://optn.transplant.hrsa.gov/professionals/by-topic/guidance/liver-review-board-guidance/*.

The highest priority (status 1) continues to be reserved for patients with acute liver failure (ALF) or complications following transplantation such as primary graft nonfunction and hepatic artery thrombosis. Because candidates for liver transplantation who have HCC may not be sufficiently decompensated to compete for donor organs based on urgency criteria alone, and because protracted waiting for deceased-donor organs often results in tumor growth beyond acceptable limits for transplantation, such patients are assigned disease-specific MELD exception points. In addition to HCC, other disease-specific MELD exceptions are evaluated on a regular basis and updated by the National Liver Review Board (NLRB) (Table 356-3).

LIVING-DONOR TRANSPLANTATION

Occasionally, especially for liver transplantation in children, one deceased-donor organ can be split between two recipients (one adult and one child). A more viable alternative, transplantation of the right lobe of the liver from a healthy adult donor into an adult recipient, has gained increased popularity. Living-donor transplantation of the left lobe (left lateral segment), introduced in the early 1990s to alleviate the extreme shortage of donor organs for small children, accounts currently for approximately one-third of all liver transplantation procedures in children. Driven by the shortage of deceased-donor organs, living-donor transplantation involving the more sizable right lobe is being considered with increasing frequency in adults; however, living-donor liver transplantation cannot be expected to solve the donor organ shortage; 603 such procedures were done in 2023, representing only ~6% of all liver transplant operations done in the United States.

Living-donor transplantation can reduce waiting time and cold ischemia time; is done under elective, rather than emergency, circumstances; and is lifesaving in recipients who cannot afford to wait for a deceased donor. The downside, of course, is the risk to the healthy donor (a mean of 10 weeks of medical disability; biliary complications in ~5%; postoperative complications such as wound infection, small-bowel obstruction, and incisional hernias in 9–19%; and death in 0.2–0.4%) as well as the increased frequency of biliary (15–32%) and vascular (10%) complications in the recipient. Potential donors must participate voluntarily without coercion, and transplantation teams should go to great lengths to exclude subtle coercive or inappropriate psychological factors as well as outline carefully to both donor and recipient the potential benefits and risks of the procedure. Donors for the procedure should be 18–65 years old; have a compatible blood type with the recipient; have no serious chronic medical problems or history of major abdominal surgery; and pass an exhaustive series of clinical, biochemical, and serologic evaluations to unearth disqualifying medical disorders. The recipient should meet the same UNOS criteria for liver transplantation as recipients of a deceased donor allograft.

■ SURGICAL TECHNIQUE

Removal of the recipient's native liver is technically difficult, particularly in the presence of portal hypertension with its associated collateral circulation and extensive varices and especially in the presence of scarring from previous abdominal operations. The combination of portal hypertension and coagulopathy (elevated prothrombin time and thrombocytopenia) may translate into large blood-product transfusion requirements. After the portal vein and infrahepatic and suprahepatic inferior vena cava are dissected, the hepatic artery and common bile duct are dissected. Then the native liver is removed and the donor organ inserted. During the anhepatic phase, coagulopathy, hypoglycemia, hypocalcemia, and hypothermia are encountered and must be managed by the anesthesiology team. Caval, portal vein, hepatic artery, and bile duct anastomoses are performed in succession, the last by end-to-end suturing of the donor and recipient common bile ducts **(Fig. 356-1)** or by choledochojejunostomy to a Roux-en-Y loop if the recipient common bile duct cannot be used for reconstruction (e.g., in sclerosing cholangitis). A typical transplant operation lasts 8 h, with a range of 6–18 h. Because of excessive bleeding, large volumes of blood, blood products, and volume expanders may be required during surgery; however, blood requirements have fallen sharply with improvements in surgical technique, blood-salvage interventions, and experience.

As noted above, emerging alternatives to orthotopic liver transplantation include split-liver grafts, in which one donor organ is divided and inserted into two recipients, and living-donor procedures, in which part of the left (for children), the left (for children or small adults), or the right (for adults) lobe of the liver is harvested from a living donor for transplantation into the recipient. In the adult procedure, once the right lobe is removed from the donor, the donor right hepatic vein is anastomosed to the recipient right hepatic vein remnant, followed by donor-to-recipient anastomoses of the portal vein and then the hepatic artery. Finally, the biliary anastomosis is performed, duct-to-duct if practical or via Roux-en-Y anastomosis. Heterotopic liver transplantation, in which the donor liver is inserted without removal of the native liver, has met with very limited success and acceptance, except in a very small number of centers. Areas of research with the potential to overcome the shortage of donor organs include hepatocyte transplantation and xenotransplantation with genetically modified organs of nonhuman origin (e.g., swine).

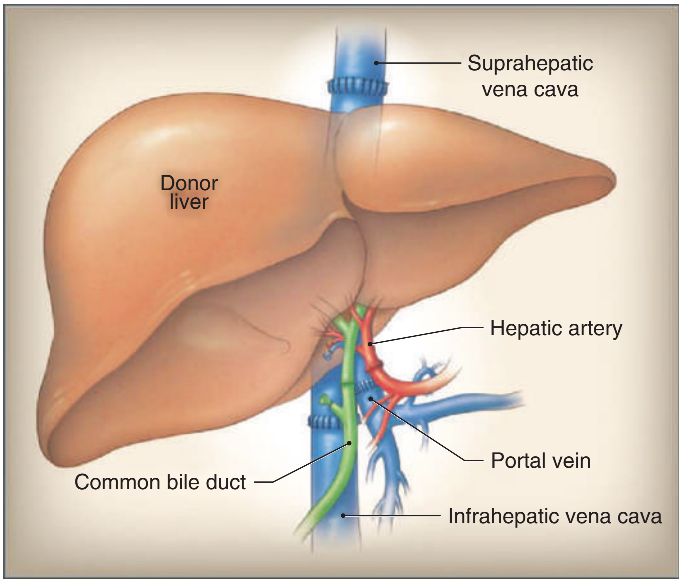

FIGURE 356-1 The anastomoses in orthotopic liver transplantation. The anastomoses are performed in the following sequence: (1) suprahepatic and infrahepatic vena cava, (2) portal vein, (3) hepatic artery, and (4) common bile duct-to-duct anastomosis. *(Reproduced with permission from S Vilarinho, RP Lifton: Liver transplantation: From inception to clinical practice. 150:1096, 2012.)*

POSTOPERATIVE COURSE AND MANAGEMENT

■ IMMUNOSUPPRESSIVE THERAPY

The introduction in 1980 of cyclosporine as an immunosuppressive agent contributed substantially to the improvement in survival after liver transplantation. Cyclosporine, a calcineurin inhibitor, blocks early activation of T cells and is specific for T-cell functions that result from the interaction of the T cell with its receptor and that involve the calcium-dependent signal transduction pathway. As a result, the activity of cyclosporine leads to inhibition of lymphokine gene activation, blocking interleukins 2, 3, and 4, tumor necrosis factor α, and other lymphokines. Cyclosporine also inhibits B-cell functions. This process occurs without affecting rapidly dividing cells in the bone marrow, which may account for the reduced frequency of posttransplantation systemic infections. The most common and important side effect of cyclosporine therapy is nephrotoxicity. Cyclosporine causes dose-dependent renal tubular injury and direct renal artery vasospasm. Following renal function is therefore important in monitoring cyclosporine therapy and is perhaps even a more reliable indicator than blood levels of the drug. Nephrotoxicity is reversible and can be managed by dose reduction. Other adverse effects of cyclosporine therapy include hypertension, hyperkalemia, tremor, hirsutism, glucose intolerance, and gingival hyperplasia.

Tacrolimus, a macrolide lactone antibiotic isolated from a Japanese soil fungus, *Streptomyces tsukubaensis*, has the same mechanism of action as cyclosporine but is 10–100 times more potent. Initially applied as "rescue" therapy for patients in whom rejection occurred despite the use of cyclosporine, tacrolimus was shown to be associated with a reduced frequency of acute, refractory, and chronic rejection. Although patient and graft survival are the same with these two drugs, the advantage of tacrolimus in minimizing episodes of rejection, reducing the need for additional glucocorticoid doses, and lowering the likelihood of bacterial and cytomegalovirus (CMV) infection has simplified the management of patients undergoing liver transplantation. In addition, the oral absorption of tacrolimus is more predictable than that of cyclosporine. As a result, tacrolimus has now supplanted cyclosporine for primary immunosuppression, and most centers rely on oral rather than IV administration from the outset.

Although more potent than cyclosporine, tacrolimus is also more toxic and more likely to be discontinued for adverse events. The toxicity of tacrolimus is similar to that of cyclosporine; nephrotoxicity and neurotoxicity are the most commonly encountered adverse effects, and neurotoxicity (tremor, seizures, hallucinations, psychoses, coma) is more likely and more severe in tacrolimus-treated patients. Both drugs can cause diabetes mellitus, but tacrolimus does not cause hirsutism or gingival hyperplasia. Because of overlapping toxicity between cyclosporine and tacrolimus, especially nephrotoxicity, and because tacrolimus reduces cyclosporine clearance, these two drugs should not be used together. Because 99% of tacrolimus is metabolized by the liver, hepatic dysfunction reduces its clearance; in primary graft nonfunction (when, for technical reasons or because of ischemic damage prior to its insertion, the allograft is defective and does not function normally from the outset), tacrolimus doses have to be reduced substantially, especially in children. Both cyclosporine and tacrolimus are metabolized by the cytochrome P450 IIIA system, and therefore, drugs that induce cytochrome P450 (e.g., phenytoin, phenobarbital, carbamazepine, rifampin) reduce available levels of cyclosporine and tacrolimus, and drugs that inhibit cytochrome P450 (e.g., erythromycin, fluconazole, ketoconazole, clotrimazole, itraconazole, verapamil, diltiazem, danazol, metoclopramide, the HIV protease inhibitor ritonavir, and the HCV protease inhibitors glecaprevir [cyclosporine only] and grazoprevir) increase cyclosporine and tacrolimus blood levels. Indeed, itraconazole is used occasionally to help boost tacrolimus levels. Like azathioprine, cyclosporine and tacrolimus appear to be associated with a risk of lymphoproliferative malignancies (see below), which may occur earlier after cyclosporine or tacrolimus than after azathioprine therapy. Because of these side effects, combinations of cyclosporine or tacrolimus with prednisone and an antimetabolite (azathioprine or mycophenolic acid, see below)—all at reduced doses—are preferable regimens for immunosuppressive therapy.

Mycophenolic acid, a nonnucleoside purine metabolism inhibitor derived as a fermentation product from several *Penicillium* species, is another immunosuppressive drug being used for patients undergoing liver transplantation. Mycophenolate has been shown to be better than azathioprine, when used with other standard immunosuppressive drugs, in preventing rejection after renal transplantation and has been adopted widely as well for use in liver transplantation. The most common adverse effects of mycophenolate are bone marrow suppression and gastrointestinal complaints.

In patients with pretransplantation renal dysfunction or renal deterioration that occurs intraoperatively or immediately postoperatively, tacrolimus or cyclosporine therapy may not be practical; under these circumstances, induction or maintenance of immunosuppression with antithymocyte globulin (ATG; thymoglobulin) or monoclonal antibodies to T cells (basiliximab, daclizumab) may be appropriate. Therapy with some of these agents (ATG) has also been effective in reversing acute rejection in the posttransplantation period and is the standard treatment for acute rejection that fails to respond to methylprednisolone boluses. Available data support the use of thymoglobulin induction to delay calcineurin inhibitor use and its attendant nephrotoxicity. IV infusions of thymoglobulin may be complicated by fever and chills, which can be ameliorated by premedication with antipyretics and a low dose of glucocorticoids.

Sirolimus, an inhibitor of the mammalian target of rapamycin (mTOR), blocks later events in T-cell activation and is another agent approved for use in solid-organ transplantation. Sirolimus, however, is not used early in the posttransplantation course, because it impairs wound healing and is associated with an increased frequency of hepatic artery thrombosis in the first 30 days after transplantation. In patients with calcineurin inhibitor–related nephrotoxicity, conversion to sirolimus has been demonstrated to be effective in preventing rejection with accompanying improvements in renal function. Because of its profound antiproliferative effects, sirolimus has also been suggested to be a useful immunosuppressive agent in patients with a prior or current history of malignancy, such as HCC. Side effects include hyperlipidemia, peripheral edema, oral ulcers, and interstitial pneumonitis. Everolimus is a hydroxyethyl derivative of sirolimus that, when used in conjunction with low-dose tacrolimus, also provides successful protection against acute rejection, with decreased renal impairment compared to that associated with standard tacrolimus dosing. Everolimus and sirolimus share a similar adverse event profile. The most important principle of immunosuppression is that the ideal approach strikes a balance between immunosuppression and immunologic competence. In general, given sufficient immunosuppression, acute liver allograft rejection is nearly always reversible. On one hand, incompletely treated acute rejection predisposes to the development of chronic rejection, which can threaten graft survival. On the other hand, if the cumulative dose of immunosuppressive therapy is too large, the patient may succumb to opportunistic infection. Data show a benefit to minimizing the use of glucocorticoids, a mainstay of immunosuppressive regimens, as cases allow. In select conditions, such as hepatitis C, early steroid withdrawal and initiation of DAAs can avert recurrent allograft hepatitis successfully. Patients who undergo liver transplantation for autoimmune diseases such as autoimmune hepatitis are less likely to achieve freedom from glucocorticoids, although, in many cases, immunosuppression can be narrowed to a dual-agent regimen of a calcineurin inhibitor and an antimetabolite.

■ POSTOPERATIVE COMPLICATIONS

Complications of liver transplantation can be divided into nonhepatic and hepatic categories (**Tables 356-4 and 356-5**). In addition, both immediate postoperative and late complications are encountered. As a rule, patients who undergo liver transplantation have been chronically ill for protracted periods and may be malnourished and wasted. The impact of such chronic illness and the multisystem failure that

TABLE 356-4 Nonhepatic Complications of Liver Transplantation

CATEGORY	COMPLICATION
Cardiovascular instability	Arrhythmias Congestive heart failure Cardiomyopathy
Pulmonary compromise	Pneumonia Pulmonary capillary vascular permeability Fluid overload
Renal dysfunction	Prerenal azotemia Hypoperfusion injury (acute tubular necrosis) Drug nephrotoxicity ↓ Renal blood flow secondary to ↑ intraabdominal pressure
Hematologic	Anemia secondary to gastrointestinal and/or intraabdominal bleeding Hemolytic anemia, aplastic anemia Thrombocytopenia
Infection	Bacterial: early, common postoperative infections Fungal/parasitic: late, opportunistic infections Viral: late, opportunistic infections, recurrent hepatitis
Neuropsychiatric	Seizures Metabolic encephalopathy Depression Difficult psychosocial adjustment
Diseases of donor	Infectious Malignant
Malignancy	B-cell lymphoma (posttransplantation lymphoproliferative disorders) De novo neoplasms (particularly squamous cell skin carcinoma)

TABLE 356-5 Hepatic Complications of Liver Transplantation

Hepatic Dysfunction Common after Major Surgery	
Prehepatic	Pigment load Hemolysis Blood collections (hematomas, abdominal collections)
Intrahepatic	
Early	Hepatotoxic drugs and anesthesia Hypoperfusion (hypotension, shock, sepsis) Benign postoperative cholestasis
Late	Transfusion-associated hepatitis Exacerbation of primary hepatic disease
Posthepatic	Biliary obstruction ↓ Renal clearance of conjugated bilirubin (renal dysfunction)
Hepatic Dysfunction Unique to Liver Transplantation	
Primary graft nonfunction	
Vascular compromise	Portal vein obstruction Hepatic artery thrombosis Anastomotic leak with intraabdominal bleeding
Bile duct disorder	Stenosis, obstruction, leak
Rejection	
Recurrent primary hepatic disease	

accompanies liver failure continue to require attention in the postoperative period. Because of the massive fluid losses and fluid shifts that occur during the operation, patients may remain fluid overloaded during the immediate postoperative period, straining cardiovascular reserve; this effect can be amplified in the face of transient renal dysfunction and pulmonary capillary vascular permeability. Continuous monitoring of cardiovascular and pulmonary function, measures to maintain the integrity of the intravascular compartment and to treat extravascular volume overload, and scrupulous attention to potential sources and sites of infection are of paramount importance. Cardiovascular instability may also result from the electrolyte imbalance that may accompany reperfusion of the donor liver as well as from restoration of systemic vascular resistance following implantation. Pulmonary function may be compromised further by paralysis of the right hemidiaphragm associated with phrenic nerve injury. The hyperdynamic state with increased cardiac output that is characteristic of patients with liver failure reverses rapidly after successful liver transplantation.

Other immediate management issues include renal dysfunction. Prerenal azotemia, acute kidney injury associated with hypoperfusion (acute tubular necrosis), and renal toxicity caused by antibiotics, tacrolimus, or cyclosporine are encountered frequently in the postoperative period, sometimes necessitating dialysis. Hemolytic-uremic syndrome can be associated with cyclosporine and tacrolimus. Occasionally, postoperative intraperitoneal bleeding may be sufficient to increase intraabdominal pressure, which, in turn, may reduce renal blood flow; this effect is rapidly reversible when abdominal distention is relieved by exploratory laparotomy to identify and ligate the bleeding site and to remove intraperitoneal clot.

Anemia may also result from acute upper gastrointestinal bleeding or from transient hemolytic anemia, which may be autoimmune, especially when blood group O livers are transplanted into blood group A or B recipients. This autoimmune hemolytic anemia is mediated by donor intrahepatic lymphocytes that recognize red blood cell A or B antigens on recipient erythrocytes. Transient in nature, this process resolves once the donor liver is repopulated by recipient bone marrow–derived lymphocytes; the hemolysis can be treated by transfusing blood group O red blood cells and/or by administering higher doses of glucocorticoids. Transient thrombocytopenia is also encountered commonly. Aplastic anemia, a late occurrence, is rare but has been reported in almost 30% of patients who underwent liver transplantation for acute, severe hepatitis of unknown cause.

Bacterial, fungal, or viral infections are common and may be life-threatening postoperatively. Early after transplant surgery, common postoperative infections predominate—pneumonia, wound infections, infected intraabdominal collections, urinary tract infections, and IV line infections—rather than opportunistic infections; these infections may involve the biliary tree and liver as well. Beyond the first postoperative month, the toll of immunosuppression becomes evident, and opportunistic infections—CMV, herpes viruses, fungal infections (*Aspergillus*, *Candida*, cryptococcal disease), mycobacterial infections, parasitic infections (*Pneumocystis*, *Toxoplasma*), and bacterial infections (*Nocardia*, *Legionella*, *Listeria*)—predominate. Rarely, early infections represent those transmitted with the donor liver, either infections present in the donor or infections acquired during procurement processing. De novo viral hepatitis infections acquired from the donor organ or, almost unheard of now, from transfused blood products occur after typical incubation periods for these agents (well beyond the first month). Obviously, infections in an immunosuppressed host demand early recognition and prompt management; prophylactic antibiotic therapy is administered routinely in the immediate postoperative period. Use of sulfamethoxazole with trimethoprim reduces the incidence of postoperative *Pneumocystis jirovecii* pneumonia. Antiviral prophylaxis for CMV with ganciclovir should be administered in patients at high risk (e.g., when a CMV-seropositive donor organ is implanted into a CMV-seronegative recipient).

Neuropsychiatric complications include seizures (commonly associated with cyclosporine and tacrolimus toxicity), metabolic encephalopathy, depression, and difficult psychosocial adjustment. Rarely, diseases are transmitted by the allograft from the donor to the recipient. In addition to viral and bacterial infections, malignancies of donor origin have occurred. Posttransplantation lymphoproliferative disorders, especially B-cell lymphoma, are a recognized complication associated with immunosuppressive drugs such as azathioprine, tacrolimus, and cyclosporine (see above). Epstein-Barr virus has been shown to play a contributory role in some of these tumors, which may regress when immunosuppressive therapy is reduced. De novo neoplasms appear at increased frequency after liver transplantation, particularly squamous cell carcinomas of the skin. Routine screening should be performed.

Long-term complications after liver transplantation attributable primarily to immunosuppressive medications include diabetes mellitus and osteoporosis (associated with glucocorticoids and calcineurin inhibitors) as well as hypertension, hyperlipidemia, and chronic renal insufficiency (associated with cyclosporine and tacrolimus). Monitoring and treating these disorders are routine components of posttransplantation care; in some cases, they respond to changes in immunosuppressive regimen, while in others, specific treatment of the disorder is introduced. Data from a large U.S. database showed that the prevalence of renal failure was 18% at year 5 and 25% at year 10 after liver transplantation. Similarly, the high frequency of diabetes, hypertension, hyperlipidemia, obesity, and the metabolic syndrome renders patients susceptible to cardiovascular disease after liver transplantation; although hepatic complications account for most of the mortality after liver transplantation, renal failure and cardiovascular disease are the other leading causes of late mortality after liver transplantation.

■ HEPATIC COMPLICATIONS

Hepatic dysfunction after liver transplantation is similar to the hepatic complications encountered after major abdominal and cardiothoracic surgery; however, in addition, hepatic complications include primary graft failure, vascular compromise, failure or stricture of the biliary anastomoses, and rejection. As in nontransplantation surgery, postoperative jaundice may result from prehepatic, intrahepatic, and posthepatic sources. *Prehepatic* sources represent the massive hemoglobin pigment load from transfusions, hemolysis, hematomas, ecchymoses, and other collections of blood. *Early intrahepatic* liver injury includes effects of hepatotoxic drugs and anesthesia; hypoperfusion injury associated with hypotension, sepsis, and shock; and benign postoperative cholestasis. *Late intrahepatic* sources of liver injury include exacerbation of primary disease. *Posthepatic* sources of hepatic dysfunction include biliary obstruction and reduced renal clearance of conjugated bilirubin. Hepatic complications unique to liver transplantation include primary graft failure associated with ischemic injury to the organ during harvesting; vascular compromise associated with thrombosis or stenosis of the portal vein or hepatic artery anastomoses; vascular anastomotic leak; stenosis, obstruction, or leakage of the anastomosed common bile duct; recurrence of primary hepatic disorder (see below); and rejection.

■ ALLOGRAFT REJECTION

Despite the use of immunosuppressive drugs, rejection of the transplanted liver still occurs in a proportion of patients, beginning 1–2 weeks after surgery. Clinical signs suggesting rejection may include fever, right upper quadrant pain, and reduced bile pigment and volume. Leukocytosis may occur, but the most reliable indicators are increases in serum bilirubin and aminotransferase levels. Because these tests lack specificity, and because patients can be asymptomatic, distinguishing among rejection, biliary obstruction, primary graft nonfunction, vascular compromise, viral hepatitis, CMV infection, drug hepatotoxicity, and recurrent primary disease may prove difficult. Radiographic visualization of the biliary tree and/or percutaneous liver biopsy often help to establish the correct diagnosis. Morphologic features of acute rejection include a mixed portal cellular infiltrate, bile duct injury, and/or endothelial inflammation ("endothelialitis"); some of these findings are reminiscent of graft-versus-host disease, primary biliary cholangitis, or recurrent allograft hepatitis C. As soon as allograft rejection is suspected, treatment consists of increased immunosuppression, most commonly IV methylprednisolone in repeated boluses; if this fails to abort rejection, consideration is given to thymoglobulin. Caution should be exercised when managing acute rejection with pulse glucocorticoids in patients with HCV infection, because of the high risk of triggering recurrent allograft hepatitis C; however, the availability of DAAs for HCV can obviate this concern effectively.

Chronic rejection is a relatively rare outcome that can follow repeated bouts of acute rejection or that occurs unrelated to preceding rejection episodes. Morphologically, chronic rejection is characterized by progressive cholestasis, focal parenchymal necrosis, mononuclear infiltration, vascular lesions (intimal fibrosis, subintimal foam cells, fibrinoid necrosis), and fibrosis. This process may be reflected as ductopenia—the vanishing bile duct syndrome, which is more common in patients undergoing liver transplantation for autoimmune liver disease. Reversibility of chronic rejection is limited; in patients with therapy-resistant chronic rejection, retransplantation has yielded encouraging results.

OUTCOME

SURVIVAL

The survival rate for patients undergoing liver transplantation has improved steadily since 1983. One-year survival rates have increased from ~70% in the early 1980s to 85–90% from 2003 to the present time. Currently, the 5-year survival rate exceeds 60%. An important observation is the relationship between clinical status before transplantation and outcome. For patients who undergo liver transplantation when their level of compensation is high (e.g., still working or only partially disabled), a 1-year survival rate of >85% is common. For those whose level of decompensation mandates continuous in-hospital care prior to transplantation, the 1-year survival rate is ~70%, whereas for those who are so decompensated that they require life support in an intensive care unit, the 1-year survival rate is ~50%. Since the adoption by UNOS in 2002 of the MELD system for organ allocation, posttransplantation survival has been found to be affected adversely for candidates with MELD scores >25, considered high disease severity. Thus, irrespective of allocation scheme, high disease severity before transplantation corresponds to diminished posttransplantation survival. Another important distinction in survival has been drawn between high- and low-risk patient categories. For patients who do not fit any "high-risk" designations, 1- and 5-year survival rates of 85% and 80%, respectively, have been recorded. In contrast, among patients in high-risk categories—cancer, fulminant hepatitis, age >65, concurrent renal failure, respirator dependence, portal vein thrombosis, and history of a portacaval shunt or multiple right upper quadrant operations—survival statistics fall into the range of 60% at 1 year and 35% at 5 years. Survival after retransplantation for primary graft nonfunction is ~50%. Causes of failure of liver transplantation vary with time. Failures within the first 3 months result primarily from technical complications, postoperative infections, and hemorrhage. Transplant failures after the first 3 months are more likely to result from infection, rejection, or recurrent disease (such as malignancy, viral hepatitis, or return to alcohol use).

RECURRENCE OF PRIMARY DISEASE

Features of autoimmune hepatitis, primary sclerosing cholangitis, and primary biliary cholangitis overlap with those of rejection or posttransplantation bile duct injury. Autoimmune hepatitis and sclerosing cholangitis can recur after liver transplantation. Similarly, reports of recurrent primary biliary cholangitis after liver transplantation have appeared; however, the histologic features of primary biliary cholangitis and chronic rejection are virtually indistinguishable and occur as frequently in patients with primary biliary cholangitis as in patients undergoing transplantation for other reasons. The presence of a florid inflammatory bile duct lesion is highly suggestive of the recurrence of primary biliary cholangitis, but even this lesion can be observed in acute rejection. Hereditary disorders such as Wilson's disease and α_1 antitrypsin deficiency have not recurred after liver transplantation; however, recurrence of disordered iron metabolism has been observed in some patients with hemochromatosis. Hepatic vein thrombosis (Budd-Chiari syndrome) may recur; this can be minimized by treating underlying myeloproliferative disorders and by anticoagulation. Because cholangiocarcinoma recurs almost invariably, few centers now offer transplantation to such patients; however, a few highly selected patients with operatively confirmed stage I or II cholangiocarcinoma who undergo liver transplantation combined with neoadjuvant chemoradiation may experience excellent outcomes. In patients with intrahepatic HCC who meet criteria for transplantation, 1- and 5-year survivals are similar to those observed in patients undergoing liver transplantation for nonmalignant disease. Finally, metabolic disorders such as MASLD are seen de novo and recur frequently, especially if the underlying metabolic predisposition is not altered.

Hepatitis A can recur after transplantation for fulminant hepatitis A, but such acute reinfection has no serious clinical sequelae. In fulminant hepatitis B, recurrence is not the rule; however, in the absence of any prophylactic measures, hepatitis B usually recurs after transplantation for end-stage chronic hepatitis B. Before the introduction of prophylactic antiviral therapy, immunosuppressive therapy sufficient to prevent allograft rejection led inevitably to marked increases in hepatitis B viremia, regardless of pretransplantation levels. Overall graft and patient survival were poor, and some patients experienced a rapid recapitulation of severe injury—severe chronic hepatitis or even fulminant hepatitis—after transplantation. Also recognized in the era before availability of antiviral regimens was *fibrosing cholestatic hepatitis*, rapidly progressive liver injury associated with marked hyperbilirubinemia, substantial prolongation of the prothrombin time (both out of proportion to relatively modest elevations of aminotransferase activity), and rapidly progressive liver failure. This lesion has been suggested to represent a "choking off" of the hepatocyte by an overwhelming density of HBV proteins. Complications such as sepsis and pancreatitis were also observed more frequently in patients undergoing liver transplantation for hepatitis B prior to the introduction of antiviral therapy. The introduction of long-term prophylaxis with HBIg revolutionized liver transplantation for chronic hepatitis B. Preoperative hepatitis B vaccination, preoperative or postoperative interferon (IFN) therapy, or short-term (≤2 months) HBIg prophylaxis has not been shown to be effective, but a retrospective analysis of data from several hundred European patients followed for 3 years after transplantation has shown that long-term (≥6 months) prophylaxis with HBIg is associated with a lowering of the risk of HBV reinfection from ~75 to 35% and a reduction in mortality from ~50 to 20%.

As a result of long-term HBIg use following liver transplantation for chronic hepatitis B, similar improvements in outcome have been observed in the United States, with 1-year survival rates between 75 and 90%. Currently, with HBIg prophylaxis, the outcome of liver transplantation for chronic hepatitis B is indistinguishable from that for chronic liver disease unassociated with chronic hepatitis B; essentially, medical concerns regarding liver transplantation for chronic hepatitis B have been eliminated. Passive immunoprophylaxis with HBIg is begun during the anhepatic stage of surgery, repeated daily in the postoperative days for high-risk patients (active viremia at the time of transplantation and/or co-infection with hepatitis D or HIV), and then continued with infusions that are given either at regular intervals of 4–6 weeks or, alternatively, when antibody to hepatitis B surface (anti-HBs) levels fall below a threshold of 100 mIU/mL. The current approach in most centers is to continue HBIg indefinitely in select high-risk cases, which can add ~$20,000 per year to the cost of care, and to transition to maintenance with antiviral alone in low-risk-hepatitis B liver allograft recipients. Still, "breakthrough" HBV infection occurs occasionally.

Further improving the outcome of liver transplantation for chronic hepatitis B is the current availability of such antiviral drugs as entecavir, tenofovir disoproxil fumarate, and tenofovir alafenamide **(Chap. 352)**. When these drugs are administered to patients with decompensated liver disease, a proportion improves sufficiently to postpone imminent liver transplantation. In addition, antiviral therapy can be used to prevent recurrence of HBV infection when administered *prior to* transplantation; to treat hepatitis B that recurs *after* transplantation, including in patients who break through HBIg prophylaxis; and to reverse the course of otherwise fatal fibrosing cholestatic hepatitis. Clinical trials have shown that entecavir or tenofovir antiviral therapy reduces the level of HBV replication substantially, sometimes even resulting in clearance of hepatitis B surface antigen (HBsAg); reduces alanine aminotransferase (ALT) levels; and improves histologic features of necrosis and inflammation. Currently, most liver transplantation centers combine HBIg plus one of the high-barrier-to-resistance oral nucleoside (entecavir) or nucleotide analogues (tenofovir). In low-risk patients with no detectable hepatitis B viremia at the time of transplantation, results from a number of clinical trials have suggested that antiviral prophylaxis can suffice, without HBIg or with a finite duration of HBIg, to prevent recurrent HBV infection of the allograft. In patients documented at the time of liver transplantation to have undetectable HBV DNA in serum and covalently closed circular DNA in the liver (i.e., with low risk for recurrence of HBV infection), results of a clinical trial suggested that, after receipt of 5 years of combined therapy, both HBIg and oral-agent therapy can be withdrawn sequentially (over two 6-month periods) with a success rate, as monitored over a median of 6 years after withdrawal, of 90% and an anti-HBs seroconversion rate of 60% (despite transient reappearance of HBV DNA and/or HBsAg in some of these patients).

Antiviral prophylactic approaches applied to patients undergoing liver transplantation for chronic hepatitis B are being used as well for patients without hepatitis B who receive organs from donors with antibody to hepatitis B core antigen (anti-HBc) but who do not have detectable HBsAg. Patients who undergo liver transplantation for chronic hepatitis B plus D are less likely to experience recurrent liver injury than patients undergoing liver transplantation for hepatitis B alone; still, such co-infected patients would also be offered standard posttransplantation prophylactic antiviral therapy for hepatitis B.

Until recently, the most common indication for liver transplantation was end-stage liver disease resulting from chronic hepatitis C **(Fig. 356-2)**. For patients undergoing liver transplantation for hepatitis C, because of an aggressive natural history of recurrent allograft hepatitis C, graft and patient survival were diminished substantially compared to other indications for transplantation.

The approval over the last decade of several DAA agents and of IFN-free DAA regimens against HCV has had a major impact on the management and outcome of both pretransplantation and posttransplantation HCV infection. Such therapeutic approaches (1) permit the clearance of viremia in a substantial proportion of decompensated cirrhotics, thereby preventing recurrent allograft infection and even improving the clinical status of most of these patients, delaying or obviating the need for liver replacement; and (2) achieve sustained virologic responses in a much higher proportion of persons with allograft HCV infection, because of improvements in antiviral treatment efficacy and tolerability. Ideally, patients should be treated prior to liver transplantation. This approach has already reduced the numbers of patients with HCV infection referred for liver transplantation and led to delisting of others. A concern, however, is that eradication of HCV infection will reduce the MELD score and lower the priority for a donor organ in some patients who still require transplantation because of continued hepatic decompensation and profound reduction in quality of life (a situation referred to as MELD purgatory). In addition, elimination of HCV infection prior to transplantation, historically, would have disqualified such patients as recipients of donor livers from persons with HCV infection, contracting the potential donor pool and limiting accessibility to donor organs and timely transplantation. With the advent of highly effective DAAs, however, allografts from HCV-infected donors are being transplanted into both HCV-infected and HCV-uninfected recipients, who can be treated posttransplantation with DAAs. Thus, earlier concerns about the potential impact of pretransplantation DAA treatment on limiting the donor pool are no longer an issue. The approach to HCV treatment should be individualized thoughtfully for each patient, based on such factors as MELD score, time anticipated prior to availability of a donor organ, relative clinical stability, and comorbidities.

DAA combinations that are being used successfully against allograft HCV include ledipasvir plus sofosbuvir; velpatasvir plus sofosbuvir; and grazoprevir plus pibrentasvir. (For updated guidelines, see *www.hcvguidelines.org*.) In patients with recurrent HCV infection after

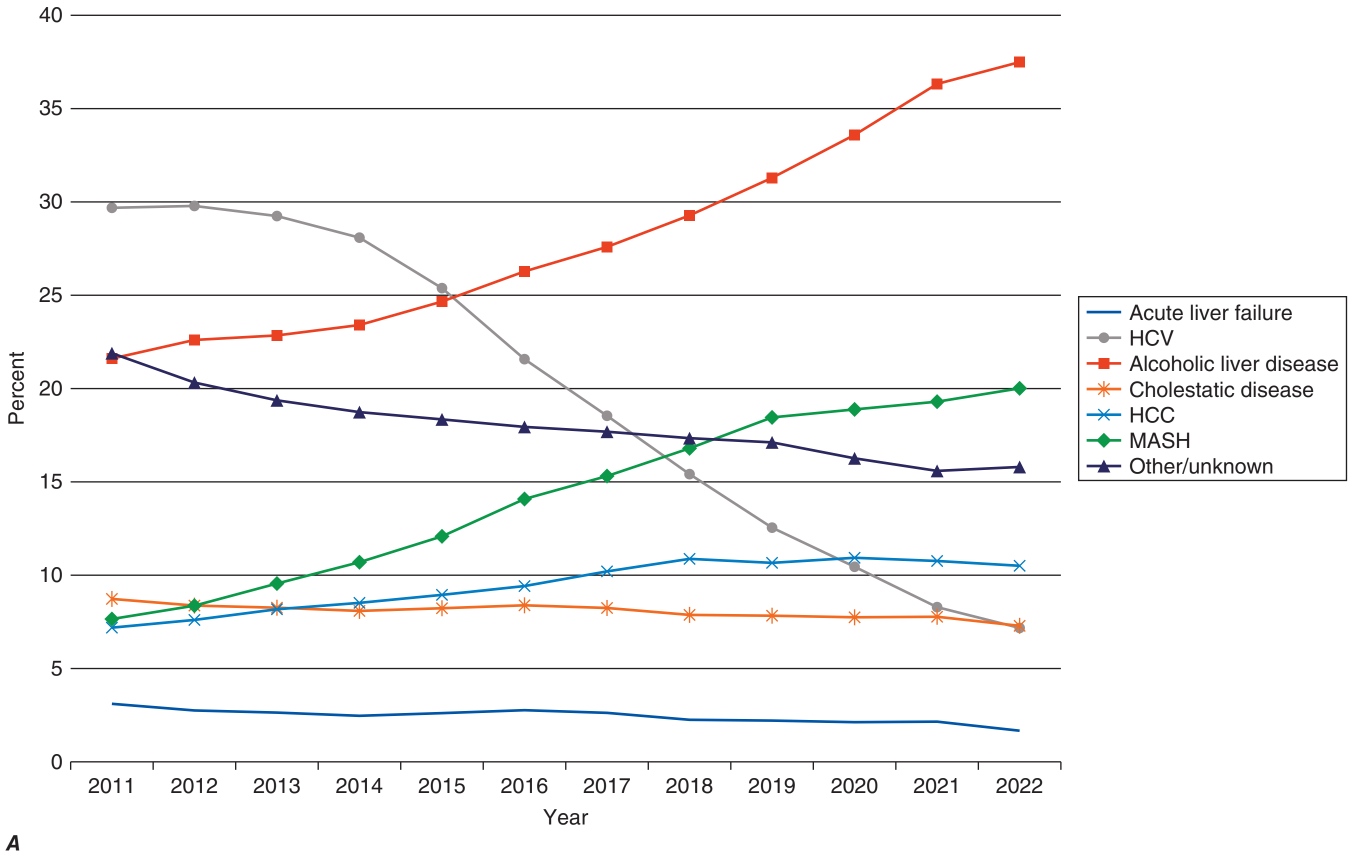

FIGURE 356-2 Trends in liver transplantation. *A.* Candidates waiting for liver transplantation by indication at any time in the given year. Candidates listed at more than one center are counted once per listing. Active and inactive patients are included. ***B.*** All liver allograft recipients by indication, including adult and pediatric, retransplantation, and multiorgan recipients. HCC, hepatocellular carcinoma; HCV, hepatitis C virus; MASH, metabolic dysfunction–associated steatohepatitis. *(From the Scientific Registry of Transplant Recipients (SRTR.org).)*

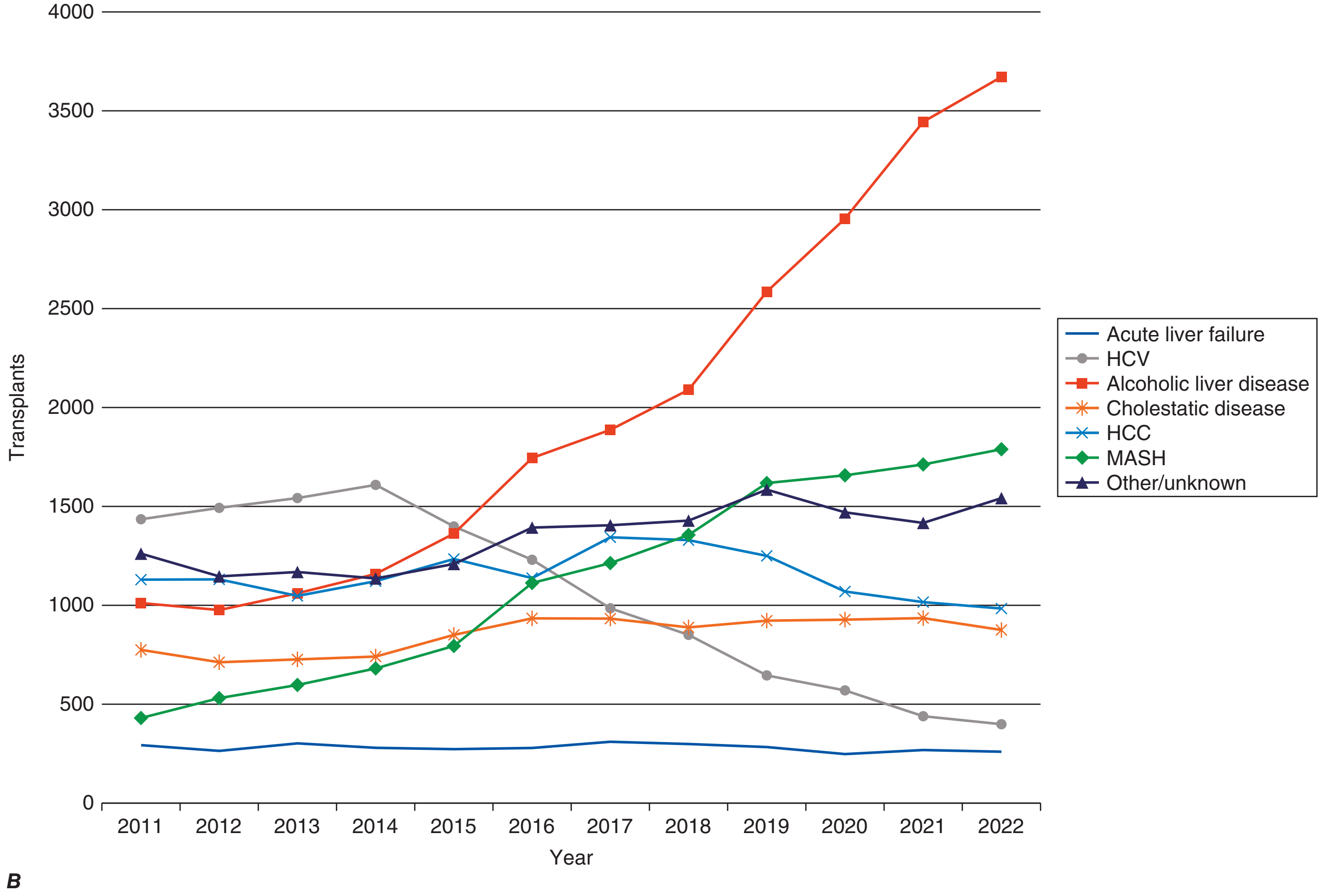

FIGURE 356-2 (*Continued*)

liver transplantation, each of these regimens has yielded response rates approaching those seen in compensated nontransplant patient populations.

Historically, small numbers of allograft recipients have succumbed to early HCV-associated liver injury, and a syndrome reminiscent of fibrosing cholestatic hepatitis (see above) has been observed rarely. Currently, however, the routine use of DAA regimens early after transplantation, before the onset of these variant presentations, has already had a profound impact on the frequency of severe recurrent allograft hepatitis C.

Patients who undergo liver transplantation for end-stage alcohol-associated cirrhosis are at risk of a return to drinking again after transplantation, a potential source of recurrent alcohol-associated liver injury. Currently, alcohol-associated liver disease is the most common indication for liver transplantation, accounting for 40% of all liver transplantation procedures, and most transplantation centers screen candidates carefully for predictors of continued abstinence. Return to alcohol use is more likely in patients whose sobriety prior to transplantation was <6 months. For abstinent patients with alcohol-associated cirrhosis, liver transplantation can be undertaken successfully, with outcomes comparable to those for other categories of patients with chronic liver disease, when coordinated by a team approach that includes substance use counseling.

The syndemic of hazardous alcohol consumption, opioid use, and obesity continues to influence important changes in liver disease epidemiology, health disparities, indications for transplant (Fig. 356-2), and disease recurrence.

POSTTRANSPLANTATION QUALITY OF LIFE

Full rehabilitation is achieved in most patients who survive the early postoperative months and escape chronic rejection or unmanageable infection. Psychosocial maladjustment interferes with medical compliance in a small number of patients, but most manage to adhere to immunosuppressive regimens, which must be continued indefinitely. In one study, 85% of patients who survived their transplant operations returned to gainful activities. In fact, some women have conceived and carried pregnancies to term after transplantation without demonstrable injury to their infants.

FURTHER READING

Bethea E et al: Immediate administration of antiviral therapy after transplantation of hepatitis C-infected livers into uninfected recipients: Implications for therapeutic planning. Am J Transplant 20:1619, 2020.

Bhattacharya D et al: Hepatitis C Guidance 2023 Update: American Association for the Study of Liver Diseases-Infectious Diseases Society of America recommendations for testing, managing, and treating hepatitis C virus infection. Clin Infect Dis 2023. (Updated regularly, available at *http://www.hcvguidelines.org.* Accessed September 23, 2023.)

Brustia R et al: Guidelines for perioperative care for liver transplantation: Enhanced Recovery After Surgery (ERAS) recommendations. Transplantation 106:552, 2022.

European Association for the Study of the Liver: EASL Clinical Practice Guidelines on acute-on-chronic liver failure. J Hepatol 79:461, 2023.

Goldberg D et al: Changes in the prevalence of hepatitis C virus infection, nonalcoholic steatohepatitis, and alcoholic liver disease among patients with cirrhosis or liver failure on the waitlist for liver transplantation. Gastroenterology 152:1090, 2017.

Kim WR et al: MELD 3.0: The Model for End-State Liver Disease updated for the modern era. Gastroenterology 161:1887, 2021.

Louvet A et al: Early liver transplantation for severe alcohol-related hepatitis not responding to medical treatment: A prospective controlled study. Lancet Gastroenterol Hepatol 7:416, 2022.

Lucey MR et al: Liver transplantation. N Engl J Med 389:1888, 2023.

Nephew LD, Serper M: Racial, gender, and socioeconomic disparities in liver transplantation. Liver Transpl 27:900, 2021.

Terrault NA et al: Liver transplantation 2023: Status report, current and future challenges. Clin Gastroeterol Hepatol 21:2150, 2023.

357

Diseases of the Gallbladder and Bile Ducts

Norton J. Greenberger*, **Gustav Paumgartner***, **Daniel S. Pratt**

PHYSIOLOGY OF BILE PRODUCTION AND FLOW

BILE SECRETION AND COMPOSITION

Bile formed in hepatocytes is secreted into a complex network of canaliculi lined by hepatocytes that flow into the canals of Hering lined partially by hepatocytes and partially by cholangiocytes, leading to small bile ductules and larger bile ducts lined by cholangiocytes that run with lymphatics and branches of the portal vein and hepatic artery in portal tracts situated between hepatic lobules. These interlobular bile ducts coalesce to form larger septal bile ducts that join to form the right and left hepatic ducts, which in turn, unite to form the common hepatic duct. The common hepatic duct is joined by the cystic duct of the gallbladder to form the common bile duct (CBD), which enters the duodenum (often after joining the main pancreatic duct) through the ampulla of Vater.

Hepatic bile is an isotonic fluid with an electrolyte composition resembling blood plasma. The electrolyte composition of gallbladder bile differs from that of hepatic bile because most of the inorganic anions, chloride, and bicarbonate have been removed by reabsorption across the gallbladder epithelium. As a result of water reabsorption, total solute concentration of bile increases from 3–4 g/dL in hepatic bile to 10–15 g/dL in gallbladder bile.

Major solute components of bile by moles percent include bile acids (80%), phospholipids (lecithins, cephalins, and sphingomyelin) (16%), and unesterified cholesterol (4.0%). In the lithogenic state, the cholesterol value can be as high as 8–10%. Other constituents include conjugated bilirubin; proteins (all immunoglobulins, albumin, metabolites of hormones, and other proteins metabolized in the liver); electrolytes; mucus; heavy metals; and, often, drugs and their metabolites.

The total daily basal secretion of hepatic bile is ~500–600 mL. Many substances taken up or synthesized by the hepatocyte are secreted into the bile canaliculi. The canalicular membrane forms microvilli and is associated with microfilaments of actin, microtubules, and other contractile elements. Prior to their secretion into the bile, many substances are taken up into the hepatocyte, while others, such as phospholipids, a portion of primary bile acids, and some cholesterol, are synthesized de novo in the hepatocyte. Three mechanisms are important in regulating bile flow: (1) active transport of bile acids from hepatocytes into the bile canaliculi, (2) active transport of other organic anions, and (3) cholangiocellular secretion. The last is a secretin-mediated and cyclic AMP–dependent mechanism that results in the secretion of a bicarbonate-rich fluid into the bile ducts.

Active vectorial trans-hepatocellular movement of bile acids from the portal blood into the bile canaliculi is driven by a set of transport systems at the basolateral (sinusoidal) and the canalicular apical plasma membrane domains of the hepatocyte. Two sinusoidal bile salt uptake systems have been cloned in humans, the Na^+/taurocholate cotransporter (NTCP, SLC10A1) and the organic anion–transporting proteins (OATP1B1/1B3), which also transport a large variety of non–bile salt organic anions. Several ATP-dependent canalicular transport systems, "export pumps" (ATP-binding cassette transport proteins, also known as ABC transporters), have been identified, the most important of which are the bile salt export pump (BSEP, ABCB11); the anionic conjugate export pump (MRP2, ABCC2), which mediates the canalicular excretion of various amphiphilic conjugates formed by phase II conjugation (e.g., bilirubin mono- and diglucuronides and drugs); the multidrug export pump (MDR1, ABCB1) for hydrophobic cationic compounds; and the phospholipid export pump (MDR3, ABCB4). Two hemitransporters, ABCG5/G8, functioning as a couple, constitute the canalicular cholesterol and phytosterol transporter. F1C1 (ATP8B1) is an aminophospholipid transferase ("flippase") essential for maintaining the lipid asymmetry of the canalicular membrane. The canalicular membrane also contains ATP-independent transport systems such as the Cl/HCO_3 anion exchanger isoform 2 (AE2, SLC4A2) for canalicular bicarbonate secretion. For most of these transporters, genetic defects have been identified that are associated with various forms of cholestasis or defects of biliary excretion. F1C1 (ATP8B1) is defective in progressive familial intrahepatic cholestasis type 1 (PFIC1) and benign recurrent intrahepatic cholestasis type 1 (BRIC1) and results in ablation of all other ATP-dependent transporter functions. BSEP (ABCB11) is defective in PFIC2 and BRIC2. Mutations of MRP2 (ABCC2) cause the Dubin-Johnson syndrome, an inherited form of conjugated hyperbilirubinemia **(Chap. 349)**. A defective MDR3 (ABCB4) results in PFIC3. ABCG5/G8, the canalicular half transporters for cholesterol and other neutral sterols, are defective in sitosterolemia. The cystic fibrosis transmembrane regulator (CFTR, ABCC7), located on bile duct epithelial cells but not on canalicular membranes, is defective in cystic fibrosis, which is associated with impaired cholangiocellular pH regulation during ductular bile formation and chronic cholestatic liver disease, occasionally resulting in biliary cirrhosis.

THE BILE ACIDS

The primary bile acids, cholic acid and chenodeoxycholic acid (CDCA), are synthesized in hepatocytes from cholesterol, conjugated with glycine or taurine, and secreted into the bile canaliculus. Secondary bile acids, including deoxycholate and lithocholate, are formed in the colon as bacterial metabolites of the primary bile acids. However, lithocholic acid is much less efficiently absorbed from the colon than deoxycholic acid. Another secondary bile acid, found in low concentration, is ursodeoxycholic acid (UDCA), a stereoisomer of CDCA. In healthy subjects, the ratio of glycine to taurine conjugates in bile is ~3:1.

Bile acids are detergent-like molecules that in aqueous solutions and above a critical concentration of ~2 m*M* form molecular aggregates called *micelles*. Cholesterol alone is sparingly soluble in aqueous environments, and its solubility in bile depends on both the total lipid concentration and the relative molar percentages of bile acids and lecithin. Normal ratios of these constituents favor the formation of solubilizing *mixed micelles*, while abnormal ratios promote the precipitation of cholesterol crystals in bile via an intermediate liquid crystal phase.

In addition to facilitating the biliary excretion of cholesterol, bile acids facilitate the normal intestinal absorption of dietary fats, mainly cholesterol, and fat-soluble vitamins, via a micellar transport mechanism **(Chap. 336)**. Bile acids also serve as a major physiologic driving force for hepatic bile flow and aid in water and electrolyte transport in the small bowel and colon.

Bile acids also function as hormones binding to nuclear (farnesoid X receptor [FXR]) and G protein–coupled (TGR5) receptors that regulate bile acid metabolism and their enterohepatic circulation.

ENTEROHEPATIC CIRCULATION

Bile acids are efficiently conserved under normal conditions. Unconjugated, and to a lesser degree also conjugated, bile acids are absorbed by *passive diffusion* along the entire gut. Quantitatively much more important for bile salt recirculation, however, is the *active transport* mechanism for conjugated bile acids in the distal ileum **(Chap. 336)**. The reabsorbed bile acids enter the portal bloodstream and are taken up rapidly by hepatocytes, reconjugated, and resecreted into bile (enterohepatic circulation).

The normal bile acid pool size is ~2–4 g. During digestion of a meal, the bile acid pool undergoes at least one or more enterohepatic cycles, depending on the size and composition of the meal. Normally, the bile acid pool circulates ~5–10 times daily. Intestinal reabsorption of the pool is ~95% efficient; therefore, daily fecal loss of bile acids is in the range of 0.2–0.4 g. In the steady state, this fecal loss is compensated by an equal daily synthesis of bile acids by the liver, and thus, the size

---

*Deceased.

of the bile acid pool is maintained. Bile acids in the intestine stimulate the release of fibroblast growth factor 19 (FGF19), which suppresses the hepatic synthesis of bile acids from cholesterol by inhibiting the rate-limiting enzyme cytochrome P450 7A1 (CYP7A1). FGF19 also promotes gallbladder relaxation. While the loss of bile salts in stool is usually matched by increased hepatic synthesis, the maximum rate of synthesis is ~5 g/d, which may be insufficient to replete the bile acid pool size when there is pronounced impairment of intestinal bile salt reabsorption.

The expression of ABC transporters in the enterohepatic circulation and of the rate-limiting enzymes of bile acid and cholesterol synthesis are regulated in a coordinated fashion by nuclear receptors, which are ligand-activated transcription factors. The hepatic BSEP (ABCB11) is upregulated by the FXR that also represses bile acid synthesis. The expression of the cholesterol transporter, ABCG5/G8, is upregulated by the liver X receptor (LXR), which is an oxysterol sensor.

GALLBLADDER AND SPHINCTERIC FUNCTIONS

In the fasting state, the sphincter of Oddi (SOD) offers a high-pressure zone of resistance to bile flow from the CBD into the duodenum. Its tonic contraction serves to (1) prevent reflux of duodenal contents into the pancreatic and bile ducts and (2) promote filling of the gallbladder. The major factor controlling the evacuation of the gallbladder is the peptide hormone cholecystokinin (CCK), which is released from the duodenal mucosa in response to the ingestion of fats and amino acids. CCK produces (1) powerful contraction of the gallbladder, (2) decreased resistance of the SOD, and (3) enhanced flow of biliary contents into the duodenum.

Hepatic bile is "concentrated" within the gallbladder by energy-dependent transmucosal absorption of water and electrolytes. Almost the entire bile acid pool may be sequestered in the gallbladder following an overnight fast for delivery into the duodenum with the first meal of the day. The normal capacity of the gallbladder is ~30 mL.

DISEASES OF THE GALLBLADDER

CONGENITAL ANOMALIES

Anomalies of the biliary tract are not uncommon and include abnormalities in number, size, and shape (e.g., agenesis of the gallbladder, duplications, rudimentary or oversized "giant" gallbladders, and diverticula). *Phrygian cap* is a clinically innocuous entity in which a partial or complete septum (or fold) separates the fundus from the body. Anomalies of position or suspension are not uncommon and include left-sided gallbladder, intrahepatic gallbladder, retrodisplacement of the gallbladder, and "floating" gallbladder. The latter condition predisposes to acute torsion, volvulus, or herniation of the gallbladder.

GALLSTONES

Epidemiology and Pathogenesis Gallstones are quite prevalent in most Western countries. Gallstone formation increases after age 50. In the United States, the prevalence is highest in Native Americans followed by Hispanics, non-Hispanic whites, and black Americans. The prevalence is higher in women than men across all ages.

Gallstones form because of abnormal bile composition. They are divided into two major types: cholesterol stones and pigment stones. Cholesterol stones account for >90% of all gallstones in Western industrialized countries. Cholesterol gallstones usually contain >50% cholesterol monohydrate plus an admixture of calcium salts, bile pigments, proteins, and fatty acids. Pigment stones are composed primarily of calcium bilirubinate; they contain <20% cholesterol and are classified into "black" and "brown" types, the latter forming secondary to chronic biliary infection.

CHOLESTEROL STONES AND BILIARY SLUDGE Cholesterol is essentially water-insoluble and requires aqueous dispersion into either micelles or vesicles, both of which require the presence of a second lipid to solubilize the cholesterol. Cholesterol and phospholipids are secreted into bile as unilamellar bilayered vesicles, which are converted into mixed micelles consisting of bile acids, phospholipids, and cholesterol by the action of bile acids. If there is an excess of cholesterol in relation to phospholipids and bile acids, unstable, cholesterol-rich vesicles remain, which aggregate into large multilamellar vesicles from which cholesterol crystals precipitate (Fig. 357-1).

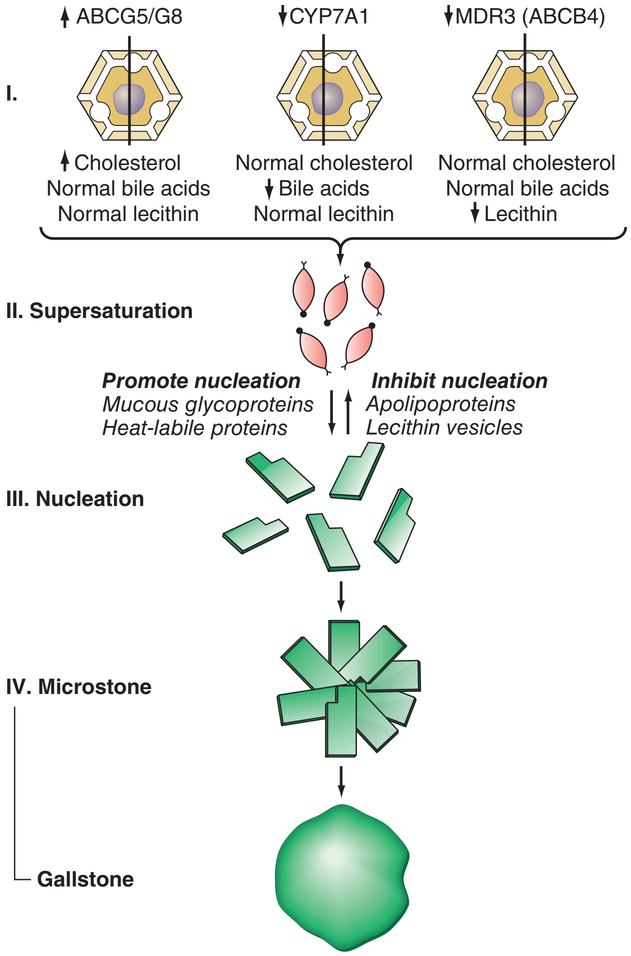

FIGURE 357-1 Scheme showing pathogenesis of cholesterol gallstone formation. Conditions or factors that increase the ratio of cholesterol to bile acids and phospholipids (lecithin) favor gallstone formation. ABCB4, ATP-binding cassette transporter; ABCG5/8, ATP-binding cassette (ABC) transporter G5/G8; CYP7A1, cytochrome P450 7A1; MDR3, multidrug resistance protein 3, also called phospholipid export pump.

There are several important mechanisms in the formation of lithogenic (stone-forming) bile. The most important is increased biliary secretion of cholesterol. This may occur in association with obesity, the metabolic syndrome, high-caloric and cholesterol-rich diets, or drugs (e.g., clofibrate) and may result from increased activity of hydroxymethylglutaryl-coenzyme A (HMG-CoA) reductase, the rate-limiting enzyme of hepatic cholesterol synthesis, and increased hepatic uptake of cholesterol from blood. In patients with gallstones, dietary cholesterol *increases* biliary cholesterol secretion. This does not occur in nongallstone patients on high-cholesterol diets. In addition to environmental factors such as high-caloric and cholesterol-rich diets, genetic factors play an important role in gallstone disease. A large study of symptomatic gallstones in Swedish twins provided strong evidence for a role of genetic factors in gallstone pathogenesis. Genetic factors accounted for 25%, shared environmental factors for 13%, and individual environmental factors for 62% of the phenotypic variation among monozygotic twins. A single nucleotide polymorphism of the gene encoding the hepatic cholesterol transporter ABCG5/G8 has been found in 21% of patients with gallstones, but only in 9% of the general population. It is thought to cause a gain of function of the cholesterol transporter and to contribute to cholesterol hypersecretion. A high prevalence of gallstones is found among first-degree relatives of gallstone carriers and in certain ethnic populations such as American Indians, Chilean Indians, and Chilean Hispanics. A common genetic trait has been identified for some of these populations by mitochondrial DNA analysis. In some patients, impaired hepatic conversion of cholesterol to bile acids may also occur, resulting in an increase of the lithogenic cholesterol/bile acid ratio. Although most cholesterol stones have a polygenic basis, there are rare monogenic (Mendelian) causes. Mutations in the *CYP7A1* gene have been described that result in a deficiency of the enzyme cholesterol 7-hydroxylase, which catalyzes the initial step in cholesterol catabolism and bile acid synthesis. The homozygous state is associated with hypercholesterolemia and gallstones. Because the phenotype is expressed in the heterozygote state, mutations in the *CYP7A1* gene may contribute to the susceptibility to cholesterol gallstone disease in the population. Mutations in the *MDR3* (ABCB4) gene, which encodes the phospholipid export pump in the canalicular membrane of the hepatocyte, may cause defective phospholipid secretion into bile, resulting in cholesterol supersaturation of bile and formation of cholesterol gallstones in the gallbladder and in the bile ducts. Thus, an excess of biliary cholesterol in relation to bile acids and phospholipids is primarily due to hypersecretion of cholesterol, but hyposecretion of bile acids or phospholipids may contribute. An additional disturbance of bile acid metabolism that is likely to contribute to supersaturation of bile with cholesterol is enhanced conversion of cholic acid to deoxycholic acid, with replacement of the cholic acid pool by an expanded deoxycholic acid pool. It may result from enhanced dehydroxylation of cholic acid and increased absorption of newly formed deoxycholic acid. An increased deoxycholate secretion is associated with hypersecretion of cholesterol into bile. A recent study suggested a role for the gut microbiota as an environmental factor contributing to gallstone formation through an influence on cholesterol and bile acid metabolism.

While supersaturation of bile with cholesterol is an important prerequisite for gallstone formation, it is generally not sufficient by itself to produce cholesterol precipitation in vivo. Most individuals with supersaturated bile do not develop stones because the time required for cholesterol crystals to nucleate and grow is longer than the time bile remains in the gallbladder.

An important mechanism is *nucleation* of cholesterol monohydrate crystals, which is greatly accelerated in human lithogenic bile. Accelerated nucleation of cholesterol monohydrate in bile may be due to either an *excess of pronucleating factors* or a *deficiency of antinucleating factors*. Mucin and certain nonmucin glycoproteins, principally immunoglobulins, appear to be pronucleating factors, while apolipoproteins A-I and A-II and other glycoproteins appear to be antinucleating factors. Pigment particles may possibly play a role as nucleating factors. In a genome-wide analysis of serum bilirubin levels, the uridine diphosphate-glucuronyltransferase 1A1 (*UGT1A1*) Gilbert's syndrome gene variant was associated with the presence of gallstone disease. Because most gallstones associated with the *UGT1A1* variant were cholesterol stones, this finding points to the role of pigment particles in the pathogenesis of gallbladder stones. Cholesterol monohydrate crystal nucleation and crystal growth probably occur within the mucin gel layer. Vesicle fusion leads to liquid crystals, which, in turn, nucleate into solid cholesterol monohydrate crystals. Continued growth of the crystals occurs by direct nucleation of cholesterol molecules from supersaturated unilamellar or multilamellar biliary vesicles.

A third important mechanism in cholesterol gallstone formation is *gallbladder hypomotility*. If the gallbladder emptied all supersaturated or crystal-containing bile completely, stones would not be able to grow. A high percentage of patients with gallstones exhibits abnormalities of gallbladder emptying. Ultrasonographic studies show that gallstone patients display an increased gallbladder volume during fasting and after a test meal (residual volume) and that fractional emptying after gallbladder stimulation is decreased. The incidence of gallstones is increased in conditions associated with infrequent or impaired gallbladder emptying such as fasting, parenteral nutrition, or pregnancy and in patients using drugs that inhibit gallbladder motility.

Biliary sludge is a thick, mucous material that, upon microscopic examination, reveals lecithin-cholesterol liquid crystals, cholesterol monohydrate crystals, calcium bilirubinate, and mucin gels. Biliary sludge typically forms a crescent-like layer in the most dependent portion of the gallbladder and is recognized by characteristic echoes on ultrasonography (see below). The presence of biliary sludge implies two abnormalities: (1) the normal balance between gallbladder mucin secretion and elimination has become deranged, and (2) nucleation of biliary solutes has occurred. That biliary sludge may be a precursor form of gallstone disease is evident from several observations. In one study, 96 patients with gallbladder sludge were followed prospectively by serial ultrasound studies. In 18%, biliary sludge disappeared and did not recur for at least 2 years. In 60%, biliary sludge disappeared and reappeared; in 14%, gallstones (8% asymptomatic, 6% symptomatic) developed; and in 6%, severe biliary pain with or without acute pancreatitis occurred. In 12 patients, cholecystectomies were performed, 6 for gallstone-associated biliary pain and 3 in symptomatic patients with sludge but without gallstones who had prior attacks of pancreatitis; the latter did not recur after cholecystectomy. It should be emphasized that biliary sludge can develop with disorders that cause gallbladder hypomotility; that is, surgery, burns, total parenteral nutrition, pregnancy, and oral contraceptives—all of which are associated with gallstone formation. However, the presence of biliary sludge implies supersaturation of bile with either cholesterol or calcium bilirubinate.

Two other conditions are associated with cholesterol-stone or biliary-sludge formation: pregnancy and rapid weight reduction through a very-low-calorie diet. There appear to be two key changes during pregnancy that contribute to a "cholelithogenic state": (1) a marked increase in cholesterol saturation of bile during the third trimester and (2) sluggish gallbladder contraction in response to a standard meal, resulting in impaired gallbladder emptying. That these changes are related to pregnancy per se is supported by several studies that show reversal of these abnormalities quite rapidly after delivery. During pregnancy, gallbladder sludge develops in 20–30% of women and gallstones in 5–12%. Although biliary sludge is a common finding during pregnancy, it is usually asymptomatic and often resolves spontaneously after delivery. Gallstones, which are less common than sludge and frequently associated with biliary colic, may also disappear after delivery because of spontaneous dissolution related to bile becoming unsaturated with cholesterol postpartum.

Approximately 10–20% of persons with rapid weight reduction achieved through very-low-calorie dieting develop gallstones. In a study involving 600 patients who completed a 3-month, 520-kcal/d diet, UDCA in a dosage of 600 mg/d proved highly effective in preventing gallstone formation; gallstones developed in only 3% of UDCA recipients, compared to 28% of placebo-treated patients. In obese patients treated by gastric banding, 500 mg/d of UDCA reduced the risk of gallstone formation from 30 to 8% within a follow-up of 6 months.

To summarize, cholesterol gallstone disease occurs because of several defects, which include (1) bile supersaturation with cholesterol, (2) nucleation of cholesterol monohydrate with subsequent crystal retention and stone growth, and (3) abnormal gallbladder motor function with delayed emptying and stasis. Other important factors known to predispose to cholesterol-stone formation are summarized in **Table 357-1**.

PIGMENT STONES Black pigment stones are composed of either pure calcium bilirubinate or polymer-like complexes with calcium and mucin glycoproteins. They are more common in patients who have chronic hemolytic states (with increased conjugated bilirubin in bile); cirrhosis, especially related to alcohol; Gilbert's syndrome; or cystic fibrosis. Gallbladder stones in patients with ileal diseases, ileal resection, or ileal bypass generally are also black pigment stones. Enterohepatic recycling of bilirubin in ileal disease states contributes to their pathogenesis. Brown pigment stones are composed of calcium salts of unconjugated bilirubin with varying amounts of cholesterol and protein. They are caused by the presence of increased amounts of unconjugated, insoluble bilirubin in bile that precipitates to form stones. Deconjugation of an excess of soluble bilirubin mono- and diglucuronides may be mediated by endogenous β-glucuronidase but may also occur by spontaneous hydrolysis. Sometimes, the enzyme is also produced when bile is chronically infected by bacteria, and such

TABLE 357-1 Predisposing Factors for Cholesterol and Pigment Gallstone Formation

Cholesterol Stones

1. Demographic/genetic factors: Prevalence highest in North American Indians, Chilean Indians, and Chilean Hispanics, greater in Northern Europe and North America than in Asia, lowest in Japan; familial disposition; hereditary aspects
2. Obesity, metabolic syndrome: Normal bile acid pool and secretion but increased biliary secretion of cholesterol
3. Rapid weight loss: Mobilization of tissue cholesterol leads to increased biliary cholesterol secretion while enterohepatic circulation of bile acids is decreased
4. Female sex hormones
    a. Estrogens stimulate hepatic lipoprotein receptors, increase uptake of dietary cholesterol, and increase biliary cholesterol secretion
    b. Natural estrogens, other estrogens, and oral contraceptives lead to decreased bile salt secretion and decreased conversion of cholesterol to cholesteryl esters
5. Pregnancy: Impaired gallbladder emptying caused by progesterone combined with the influence of estrogens, which increase biliary cholesterol secretion
6. Increasing age: Increased biliary secretion of cholesterol, decreased size of bile acid pool, decreased secretion of bile salts
7. Gallbladder hypomotility leading to stasis and formation of sludge
    a. Total parenteral nutrition
    b. Fasting
    c. Pregnancy
    d. Drugs such as octreotide
8. Clofibrate therapy: Increased biliary secretion of cholesterol
9. Decreased bile acid secretion
    a. Genetic defect of the *CYP7A1* gene
10. Decreased phospholipid secretion: Genetic defect of the *MDR3* gene
11. Miscellaneous
    a. High-calorie, high-fat diet
    b. Spinal cord injury

Pigment Stones

1. Demographic/genetic factors: Asia, rural setting (presumed due to increased prevalence of parasitic biliary infections; the incidence has been dropping with time)
2. Chronic hemolysis (example: sickle cell disease)
3. Alcohol related liver cirrhosis
4. Ineffective erythropoiesis (example: pernicious anemia)
5. Cystic fibrosis
6. Chronic biliary tract infection, parasite infections
7. Increasing age
8. Ileal disease, ileal resection or bypass

stones are brown. Pigment stone formation is frequent in Asia and is often associated with parasitic infections in the gallbladder and biliary tree (Table 357-1).

Diagnosis Procedures of potential use in the diagnosis of cholelithiasis and other diseases of the gallbladder are detailed in Table 357-2. Ultrasonography of the gallbladder is very accurate in the identification of cholelithiasis and has replaced oral cholecystography (OCG) **(Fig. 357-2A)**. Stones as small as 1.5 mm in diameter may be confidently identified provided that firm criteria are used (e.g., acoustic "shadowing" of opacities that are within the gallbladder lumen and that change with the patient's position [by gravity]). In major medical centers, the false-negative and false-positive rates for ultrasound in gallstone patients are ~2–4%. Biliary sludge is material of low echogenic activity that typically forms a layer in the most dependent position of the gallbladder. This layer shifts with postural changes but fails to produce acoustic shadowing; these two characteristics distinguish sludges from gallstones. Ultrasound can also be used to assess the emptying function of the gallbladder.

The plain abdominal film may detect gallstones containing sufficient calcium to be radiopaque (10–15% of cholesterol and ~50% of pigment stones). Plain radiography may also be of use in the diagnosis of emphysematous cholecystitis, porcelain gallbladder, limey bile, and gallstone ileus.

TABLE 357-2 Diagnostic Evaluation of the Gallbladder

DIAGNOSTIC ADVANTAGES	DIAGNOSTIC LIMITATIONS	COMMENT
Ultrasound		
Rapid Accurate identification of gallstones (>95%) Simultaneous scanning of GB, liver, bile ducts, pancreas "Real-time" scanning allows assessment of GB volume, contractility Not limited by jaundice, pregnancy May detect very small stones	Bowel gas Massive obesity Ascites	Procedure of choice for detection of stones
Plain Abdominal X-Ray		
Low cost Readily available	Relatively low yield Contraindicated in pregnancy	Pathognomonic findings in: calcified gallstones, limey bile, porcelain GB, emphysematous cholecystitis, gallstone ileus
Cholescintigraphy (HIDA, DISIDA, etc.)		
Accurate identification of cystic duct obstruction Simultaneous assessment of bile ducts	Contraindicated in pregnancy Serum bilirubin >103–205 μmol/L (6–12 mg/dL) Cholecystogram of low resolution	Indicated for confirmation of suspected acute cholecystitis; less sensitive and less specific in chronic cholecystitis; useful in the diagnosis of acalculous cholecystopathy, especially if given with CCK to assess GB emptying

Abbreviations: CCK, cholecystokinin; DISIDA, diisopropyl iminodiacetic acid; GB, gallbladder; HIDA, hydroxyl iminodiacetic acid.

OCG was historically a useful procedure for the diagnosis of gallstones but has been replaced by ultrasound and is now regarded as obsolete. It may be used to assess the patency of the cystic duct and gallbladder emptying function. Further, OCG can also delineate the size and number of gallstones and determine whether they are calcified, useful information if medical dissolution is being considered.

Radiopharmaceuticals such as ^{99m}Tc-labeled *N*-substituted iminodiacetic acids (HIDA and DISIDA) are rapidly extracted from the blood and are excreted into the biliary tree in high concentration even in the presence of mild to moderate serum bilirubin elevations. Failure to image the gallbladder in the presence of biliary ductal visualization may indicate cystic duct obstruction, acute or chronic cholecystitis, or surgical absence of the organ. Such scans have application in the diagnosis of acute cholecystitis and may play a role in the detection of a postcholecystectomy bile leak.

Symptoms of Gallstone Disease Gallstones usually produce symptoms by causing inflammation or obstruction following their migration into the cystic duct or CBD. The most specific and characteristic symptom of gallstone disease is biliary colic that is a constant and often long-lasting pain (see below). Obstruction of the cystic duct or CBD by a stone produces increased intraluminal pressure and distention of the viscus that cannot be relieved by repetitive biliary contractions. The resultant visceral pain is characteristically a severe,

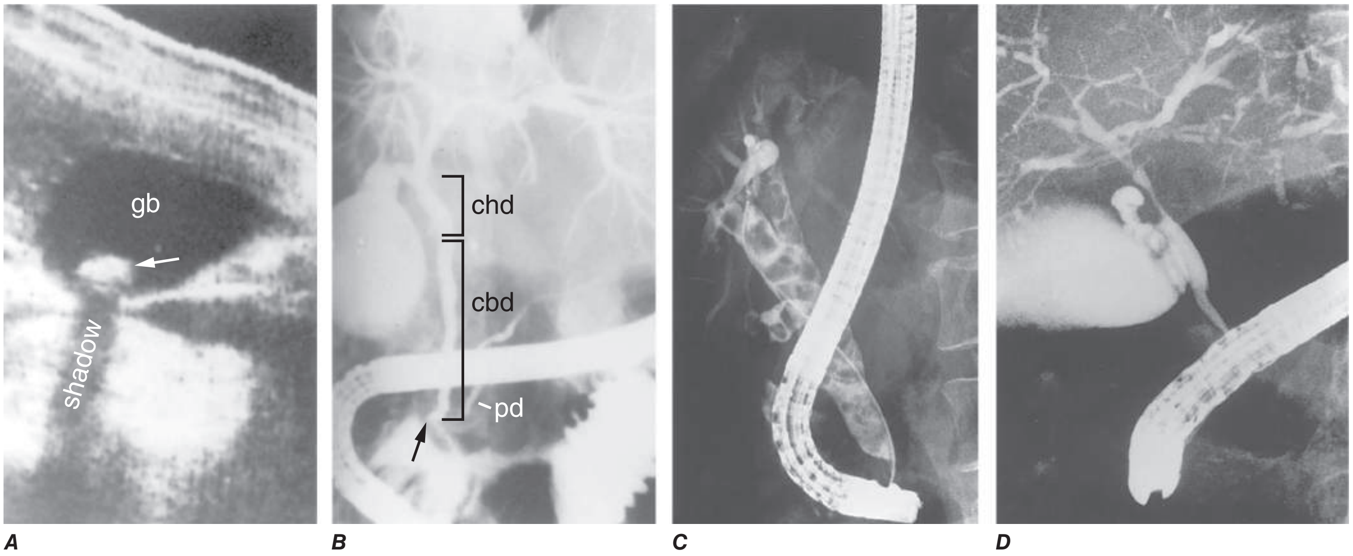

FIGURE 357-2 Examples of ultrasound and radiologic studies of the biliary tract. ***A.*** An ultrasound study showing a distended gallbladder (GB) containing a single large stone (*arrow*), which casts an acoustic shadow. ***B.*** Endoscopic retrograde cholangiopancreatogram (ERCP) showing normal biliary tract anatomy. In addition to the endoscope and large vertical gallbladder filled with contrast dye, the common hepatic duct (CHD), common bile duct (CBD), and pancreatic duct (PD) are shown. The *arrow* points to the ampulla of Vater. ***C.*** Endoscopic retrograde cholangiogram (ERC) showing choledocholithiasis. The biliary tract is dilated and contains multiple radiolucent calculi. ***D.*** ERCP showing sclerosing cholangitis. The CBD shows areas that are strictured and narrowed.

steady ache or fullness in the epigastrium or right upper quadrant (RUQ) of the abdomen with frequent radiation to the interscapular area, right scapula, or shoulder.

Biliary colic begins quite suddenly and may persist with severe intensity for 30 min to 5 h, subsiding gradually or rapidly. It is steady rather than intermittent, as would be suggested by the word *colic*, which must be regarded as a misnomer, although it is in widespread use. An episode of biliary pain persisting beyond 5 h should raise the suspicion of acute cholecystitis (see below). Nausea and vomiting frequently accompany episodes of biliary pain. An elevated level of serum bilirubin and/or alkaline phosphatase suggests a common duct stone. Fever or chills (rigors) with biliary pain usually imply a complication, that is, cholecystitis, pancreatitis, or cholangitis. Complaints of short-lasting, vague epigastric fullness, dyspepsia, eructation, or flatulence, especially following a fatty meal, should not be confused with biliary pain. Such symptoms are frequently elicited from patients with or without gallstone disease but are not specific for biliary calculi. Biliary colic may be precipitated by eating a fatty meal, by consumption of a large meal following a period of prolonged fasting, or by eating a normal meal; it is frequently nocturnal, occurring within a few hours of retiring.

Natural History Gallstone disease discovered in an asymptomatic patient or in a patient whose symptoms are not referable to cholelithiasis is a common clinical problem. Sixty to 80% of persons with asymptomatic gallstones remain asymptomatic over follow-up periods of up to 25 years. The probability of developing symptoms within 5 years after diagnosis is 2–4% per year and decreases in the years thereafter to 1–2%. The yearly incidence of complications is about 0.1–0.3%. Patients remaining asymptomatic for 15 years were found to be unlikely to develop symptoms during further follow-up, and most patients who did develop complications from their gallstones experienced *prior* warning symptoms. Similar conclusions apply to diabetic patients with silent gallstones. Decision analysis has suggested that (1) the cumulative risk of death due to gallstone disease while on expectant management is small, and (2) prophylactic cholecystectomy is not warranted.

Complications requiring cholecystectomy are much more common in gallstone patients who have developed symptoms of biliary pain. Patients found to have gallstones at a young age are more likely to develop symptoms from cholelithiasis than are patients >60 years at the time of initial diagnosis. Patients with diabetes mellitus and gallstones may be somewhat more susceptible to septic complications, but the magnitude of risk of septic biliary complications in diabetic patients is incompletely defined.

TREATMENT

Gallstones

SURGICAL THERAPY

In asymptomatic gallstone patients, the risk of developing symptoms or complications requiring surgery is quite small (see above). Thus, a recommendation for cholecystectomy in a patient with gallstones should probably be based on assessment of three factors: (1) the presence of symptoms that are frequent enough or severe enough to interfere with the patient's general routine; (2) the presence of a prior complication of gallstone disease, that is, history of acute cholecystitis, pancreatitis, gallstone fistula, etc.; or (3) the presence of an underlying condition predisposing the patient to increased risk of gallstone complications (e.g., a previous attack of acute cholecystitis regardless of current symptomatic status). Patients with very large gallstones (>3 cm in diameter) and patients harboring gallstones in a congenitally anomalous gallbladder might also be considered for prophylactic cholecystectomy. Although young age is a worrisome factor in asymptomatic gallstone patients, few authorities would now recommend routine cholecystectomy in young patients with silent stones. Laparoscopic cholecystectomy is a minimal-access approach for the removal of the gallbladder together with its stones. Its advantages include a markedly shortened hospital stay, minimal disability, and decreased cost, and it is the procedure of choice for most patients referred for elective cholecystectomy.

From several studies involving >4000 patients undergoing laparoscopic cholecystectomy, the following key points emerge: (1) complications develop in ~4% of patients, (2) conversion to laparotomy occurs in 5%, (3) the death rate is remarkably low (i.e., <0.1%), and (4) the rate of bile duct injuries is low (i.e., 0.2–0.6%) and comparable with open cholecystectomy. These data indicate why laparoscopic cholecystectomy has become the "gold standard" for treating symptomatic cholelithiasis.

MEDICAL THERAPY—GALLSTONE DISSOLUTION

In carefully selected patients with a functioning gallbladder and with radiolucent stones <10 mm in diameter, complete dissolution can be achieved in ~50% of patients within 6 months to 2 years. For good results within a reasonable time period, this therapy should be limited to radiolucent stones <5 mm in diameter. The dose of UDCA should be 10–15 mg/kg per d. Stones >10 mm in size rarely dissolve. Pigment stones are not responsive to UDCA therapy. Probably ≤10% of patients with *symptomatic* cholelithiasis are candidates for such treatment. However, in addition to the vexing problem of recurrent stones (30–50% over 3–5 years of follow-up), there is also the factor of taking a drug for up to 2 years and perhaps indefinitely. The advantages and success of laparoscopic cholecystectomy have largely reduced the role of gallstone dissolution to patients who wish to avoid or are not candidates for elective cholecystectomy. However, patients with cholesterol gallstone disease who develop recurrent choledocholithiasis after cholecystectomy should be on long-term treatment with UDCA.

ACUTE AND CHRONIC CHOLECYSTITIS

Acute Cholecystitis Acute inflammation of the gallbladder wall usually follows obstruction of the cystic duct by a stone. Inflammatory response can be evoked by three factors: (1) *mechanical inflammation* produced by increased intraluminal pressure and distention with resulting ischemia of the gallbladder mucosa and wall, (2) *chemical inflammation* caused by the release of lysolecithin (due to the action of phospholipase on lecithin in bile) and other local tissue factors, and (3) *bacterial inflammation*, which may play a role in 50–85% of patients with acute cholecystitis. The organisms most frequently isolated by culture of gallbladder bile in these patients include *Escherichia coli*, *Klebsiella* spp., *Streptococcus* spp., and *Clostridium* spp.

Acute cholecystitis often begins as an attack of biliary pain that progressively worsens. Approximately 60–70% of patients report having experienced prior attacks that resolved spontaneously. As the episode progresses, however, the pain of acute cholecystitis becomes more generalized in the right upper abdomen. As with biliary colic, the pain of cholecystitis may radiate to the interscapular area, right scapula, or shoulder. Peritoneal signs of inflammation such as increased pain with jarring or on deep respiration may be apparent. The patient is anorectic and often nauseated. Vomiting is relatively common and may produce symptoms and signs of vascular and extracellular volume depletion. Jaundice is unusual early in the course of acute cholecystitis but may occur when edematous inflammatory changes involve the bile ducts and surrounding lymph nodes.

A low-grade fever is characteristically present, but shaking chills or rigors are not uncommon. The RUQ of the abdomen is almost invariably tender to palpation. An enlarged, tense gallbladder is palpable in 25–50% of patients. Deep inspiration or cough during subcostal palpation of the RUQ usually produces increased pain and inspiratory arrest (Murphy's sign). Localized rebound tenderness in the RUQ is common, as are abdominal distention and hypoactive bowel sounds from paralytic ileus, but generalized peritoneal signs and abdominal rigidity are usually lacking, in the absence of perforation.

The diagnosis of acute cholecystitis is usually made on the basis of a characteristic history and physical examination. The triad of sudden onset of RUQ tenderness, fever, and leukocytosis is highly suggestive. Typically, leukocytosis in the range of 10,000–15,000 cells per microliter with a left shift on differential count is found. The serum bilirubin is mildly elevated (<85.5 μmol/L [5 mg/dL]) in fewer than half of patients, whereas about one-fourth have modest elevations in serum aminotransferases (usually less than a fivefold elevation). Ultrasound will demonstrate calculi in 90–95% of cases and is useful for detection of signs of gallbladder inflammation including thickening of the wall, pericholecystic fluid, and dilatation of the bile duct. The radionuclide (e.g., HIDA) biliary scan may be confirmatory if bile duct imaging is seen without visualization of the gallbladder.

Approximately 75% of patients treated medically have remission of acute symptoms within 2–7 days following hospitalization. In 25%, however, a complication of acute cholecystitis will occur despite conservative treatment (see below). In this setting, prompt surgical intervention is required. Of the 75% of patients with acute cholecystitis who undergo remission of symptoms, ~25% will experience a recurrence of cholecystitis within 1 year, and 60% will have at least one recurrent bout within 6 years. In view of the natural history of the disease, acute cholecystitis is best treated by early surgery whenever possible. Mirizzi's syndrome is a rare complication in which a gallstone becomes impacted in the cystic duct or neck of the gallbladder causing compression of the CBD, resulting in CBD obstruction and jaundice. Ultrasound shows gallstone(s) lying outside the hepatic duct. Endoscopic retrograde cholangiopancreatography (ERCP) **(Fig. 357-2*B*)**, percutaneous transhepatic cholangiography (PTC), or magnetic resonance cholangiopancreatography (MRCP) will usually demonstrate the characteristic extrinsic compression of the CBD. Surgery consists of removing the cystic duct, diseased gallbladder, and impacted stone. The preoperative diagnosis of Mirizzi's syndrome is important to avoid CBD injury.

ACALCULOUS CHOLECYSTITIS In 5–14% of patients with acute cholecystitis, calculi obstructing the cystic duct are not found at surgery. In >50% of such cases, an underlying explanation for acalculous inflammation is not found. An increased risk for the development of acalculous cholecystitis is especially associated with prolonged fasting, with serious trauma or burns, in the postpartum period following prolonged labor, and with orthopedic and other nonbiliary major surgical operations in the postoperative period. It may possibly complicate periods of prolonged parenteral hyperalimentation. For some of these cases, biliary sludge in the cystic duct may be responsible. Other precipitating factors include vasculitis, obstructing adenocarcinoma of the gallbladder, diabetes mellitus, torsion of the gallbladder, "unusual" bacterial infections of the gallbladder (e.g., *Leptospira*, *Streptococcus*, *Salmonella*, or *Vibrio cholerae*), and parasitic infestation of the gallbladder. Acalculous cholecystitis may also be seen with a variety of other systemic disease processes (e.g., sarcoidosis, cardiovascular disease, tuberculosis, syphilis, actinomycosis).

Although the clinical manifestations of acalculous cholecystitis are indistinguishable from those of calculous cholecystitis, the setting of acute gallbladder inflammation complicating severe underlying illness is characteristic of acalculous disease. Ultrasound or computed tomography (CT) examinations demonstrating a large, tense, static gallbladder without stones and with evidence of poor emptying over a prolonged period may be diagnostically useful in some cases. The complication rate for acalculous cholecystitis exceeds that for calculous cholecystitis. Successful management of acute acalculous cholecystitis appears to depend primarily on early diagnosis and surgical intervention, with meticulous attention to postoperative care.

ACALCULOUS CHOLECYSTOPATHY Disordered motility of the gallbladder can produce recurrent biliary pain in patients without gallstones. Infusion of an octapeptide of CCK can be used to measure the gallbladder ejection fraction during cholescintigraphy. The surgical findings have included abnormalities such as chronic cholecystitis, gallbladder muscle hypertrophy, and/or a markedly narrowed cystic duct. Some of these patients may well have had antecedent gallbladder disease. The following criteria can be used to identify patients with acalculous cholecystopathy: (1) recurrent episodes of typical RUQ pain characteristic of biliary tract pain, (2) abnormal CCK cholescintigraphy demonstrating a gallbladder ejection fraction of <40%, and (3) infusion of CCK reproducing the patient's pain. An additional clue would be the identification of a large gallbladder on ultrasound examination. Importantly, it should be noted that SOD dysfunction can also give rise to recurrent RUQ pain and CCK-scintigraphic abnormalities.

EMPHYSEMATOUS CHOLECYSTITIS So-called emphysematous cholecystitis is thought to begin with acute cholecystitis (calculous or acalculous) followed by ischemia or gangrene of the gallbladder wall and infection by gas-producing organisms. Bacteria most frequently cultured in this setting include anaerobes, such as *Clostridium welchii* or *C. perfringens*, and aerobes, such as *E. coli*. This condition occurs most frequently in elderly men and in patients with diabetes mellitus. The clinical manifestations are essentially indistinguishable from those of nongaseous cholecystitis. The diagnosis is usually made on plain abdominal film by finding gas within the gallbladder lumen, dissecting within the gallbladder wall to form a gaseous ring, or in the pericholecystic tissues. The morbidity and mortality rates with emphysematous cholecystitis are considerable. Prompt surgical intervention coupled with appropriate antibiotics is mandatory.

Chronic Cholecystitis Chronic inflammation of the gallbladder wall is almost always associated with the presence of gallstones and is thought to result from repeated bouts of subacute or acute cholecystitis or from persistent mechanical irritation of the gallbladder wall by gallstones. The presence of bacteria in the bile occurs in >25% of patients with chronic cholecystitis. The presence of infected bile in a patient with *chronic* cholecystitis undergoing elective cholecystectomy probably adds little to the operative risk. Chronic cholecystitis may be asymptomatic for years, which may progress to symptomatic gallbladder disease or to acute cholecystitis or may present with complications (see below).

Complications of Cholecystitis • **EMPYEMA AND HYDROPS** Empyema of the gallbladder usually results from progression of acute cholecystitis with persistent cystic duct obstruction to superinfection of the stagnant bile with a pus-forming bacterial organism. The clinical picture resembles that of cholangitis with high fever; severe RUQ pain; marked leukocytosis; and often, prostration. Empyema of the gallbladder carries a high risk of gram-negative sepsis and/or perforation. Emergency surgical intervention with proper antibiotic coverage is required as soon as the diagnosis is suspected.

Hydrops or mucocele of the gallbladder may also result from prolonged obstruction of the cystic duct, usually by a large solitary calculus. In this instance, the obstructed gallbladder lumen is progressively distended, over a period of time, by mucus (mucocele) or by a clear transudate (hydrops) produced by mucosal epithelial cells. A visible, easily palpable, nontender mass sometimes extending from the RUQ into the right iliac fossa may be found on physical examination. The patient with hydrops of the gallbladder frequently remains asymptomatic, although chronic RUQ pain may also occur. Cholecystectomy is indicated, because empyema, perforation, or gangrene may complicate the condition.

GANGRENE AND PERFORATION Gangrene of the gallbladder results from ischemia of the wall and patchy or complete tissue necrosis. Underlying conditions often include marked distention of the gallbladder, vasculitis, diabetes mellitus, empyema, or torsion resulting in arterial occlusion. Gangrene usually predisposes to perforation of the gallbladder, but perforation may also occur in chronic cholecystitis without premonitory warning symptoms. *Localized perforations* are usually contained by the omentum or by adhesions produced by recurrent inflammation of the gallbladder. Bacterial superinfection of the walled-off gallbladder contents results in abscess formation. Most patients are best treated with cholecystectomy, but some seriously ill patients may be managed with cholecystostomy and drainage of the abscess. *Free perforation* is less common but is associated with a mortality rate of ~30%. Such patients may experience a sudden transient relief of RUQ pain as the distended gallbladder decompresses; this is followed by signs of generalized peritonitis.

FISTULA FORMATION AND GALLSTONE ILEUS Fistula formation into an adjacent organ adherent to the gallbladder wall may result from inflammation and adhesion formation. Fistulas into the duodenum are most common, followed in frequency by those involving the hepatic flexure of the colon, stomach or jejunum, abdominal wall, and renal pelvis. Clinically "silent" biliary-enteric fistulas occurring as a complication of acute cholecystitis have been found in up to 5% of patients undergoing cholecystectomy. Asymptomatic cholecystoenteric fistulas may sometimes be diagnosed by finding gas in the biliary tree on plain abdominal films. Barium contrast studies or endoscopy of the upper gastrointestinal tract or colon may demonstrate the fistula. Treatment in the symptomatic patient usually consists of cholecystectomy, CBD exploration, and closure of the fistulous tract.

Gallstone ileus refers to mechanical intestinal obstruction resulting from the passage of a large gallstone into the bowel lumen. The stone customarily enters the duodenum through a cholecystoenteric fistula at that level. The site of obstruction by the impacted gallstone is usually at the ileocecal valve, provided that the more proximal small bowel is of normal caliber. Most patients do not give a history of either prior biliary tract symptoms or complaints suggestive of acute cholecystitis or fistula formation. Large stones, >2.5 cm in diameter, are thought to predispose to fistula formation by gradual erosion through the gallbladder fundus. Diagnostic confirmation may occasionally be found on the plain abdominal film (e.g., small-intestinal obstruction with gas in the biliary tree [pneumobilia] and a calcified, ectopic gallstone) or following an upper gastrointestinal series (cholecystoduodenal fistula with small-bowel obstruction at the ileocecal valve). Laparotomy with stone extraction (or propulsion into the colon) remains the procedure of choice to relieve obstruction. Evacuation of large stones within the gallbladder should also be performed. In general, the gallbladder and its attachment to the intestines should be left alone.

LIMEY (MILK OF CALCIUM) BILE AND PORCELAIN GALLBLADDER Calcium salts in the lumen of the gallbladder in sufficient concentration may produce calcium precipitation and diffuse, hazy opacification of bile or a layering effect on plain abdominal roentgenography. This so-called limey bile, or milk of calcium bile, is usually clinically innocuous, but cholecystectomy is often performed, especially when it occurs in a hydropic gallbladder. In the entity called *porcelain gallbladder*, calcium salt deposition within the wall of a chronically inflamed gallbladder may be detected on the plain abdominal film. In the past, cholecystectomy was advised in all patients with porcelain gallbladder because there was felt to be a high incidence of carcinoma of the gallbladder associated with this condition, an association challenged by a number of studies. Two patterns of gallbladder wall calcification have now been appreciated: complete intramural calcification and selective mucosal calcification. The incidence of cancer in those with selective intramural calcification is higher than those with complete mucosal wall calcification, but the risk is very small. As such, the need for cholecystectomy for porcelain gallbladder is not absolute; close surveillance in these patients is also acceptable.

TREATMENT

Acute Cholecystitis

MEDICAL THERAPY

Although surgical intervention remains the mainstay of therapy for acute cholecystitis and its complications, a period of in-hospital stabilization may be required before cholecystectomy. Oral intake is eliminated, nasogastric suction may be indicated, extracellular volume depletion and electrolyte abnormalities are repaired, and analgesia is provided. Intravenous antibiotic therapy is indicated in patients with severe acute cholecystitis, even though bacterial superinfection of bile may not have occurred in the early stages of the inflammatory process. Antibiotic therapy is guided by the most common organisms likely to be present including *E. coli*, *Klebsiella*, *Enterococcus*, *Enterobacter*, and *Streptococcus*. Effective antibiotics include piperacillin plus tazobactam, imipenem, meropenem, ceftriaxone plus metronidazole, and levofloxacin plus metronidazole **(Chap. 166)**. Postoperative complications of wound infection, abscess formation, and sepsis are reduced in antibiotic-treated patients.

SURGICAL THERAPY

The optimal timing of surgical intervention in patients with acute cholecystitis depends on stabilization of the patient. The clear trend is toward earlier surgery, and this is due in part to requirements for shorter hospital stays. Urgent (emergency) cholecystectomy or percutaneous cholecystostomy is probably appropriate in most patients in whom a complication of acute cholecystitis such as empyema, emphysematous cholecystitis, or perforation is suspected or confirmed. Patients with uncomplicated acute cholecystitis should undergo early elective laparoscopic cholecystectomy, ideally within 48–72 h after diagnosis. The complication rate is not increased in patients undergoing early as opposed to delayed (>6 weeks after diagnosis) cholecystectomy. Delayed surgical intervention is probably best reserved for (1) patients in whom the overall medical condition imposes an unacceptable risk for early surgery and (2) patients in whom the diagnosis of acute cholecystitis is in doubt. Thus, early cholecystectomy (within 72 h) is the treatment of choice for most patients with acute cholecystitis. Mortality figures for emergency cholecystectomy in most centers range from 1 to 3%, whereas the mortality risk for early elective cholecystectomy is ~0.5% in patients under age 60. Of course, the operative risks increase with age-related diseases of other organ systems and with the presence of long- or short-term complications of gallbladder disease. Seriously ill or debilitated patients with cholecystitis may be managed with percutaneous drainage (a cholecystostomy tube), transpapillary drainage (an endoscopically placed transpapillary drainage catheter via the cystic duct), or transmural drainage (an endoscopically placed covered, lumen-apposing stent). Elective cholecystectomy may then be done at a later date.

Postcholecystectomy Complications Early complications following cholecystectomy include atelectasis and other pulmonary disorders, abscess formation (often subphrenic), external or internal hemorrhage, biliary-enteric fistula, and bile leaks. Jaundice may indicate absorption of bile from an intraabdominal collection following a biliary leak or mechanical obstruction of the CBD by retained calculi, intraductal blood clots, or extrinsic compression.

Overall, cholecystectomy is a very successful operation that provides total or near-total relief of preoperative symptoms in 75–90% of patients. The most common cause of persistent postcholecystectomy symptoms is an overlooked symptomatic nonbiliary disorder (e.g., reflux esophagitis, peptic ulceration, pancreatitis, or—most often—irritable bowel syndrome). In a small percentage of patients, however, a disorder of the extrahepatic bile ducts may result in persistent symptomatology. These so-called postcholecystectomy syndromes may be due to (1) biliary strictures, (2) retained biliary calculi, (3) cystic duct stump syndrome, (4) stenosis or dyskinesia of the SOD, or (5) bile salt–induced diarrhea or gastritis.

CYSTIC DUCT STUMP SYNDROME In the absence of cholangiographically demonstrable retained stones, symptoms resembling biliary pain or cholecystitis in the postcholecystectomy patient have frequently been attributed to disease in a long (>1 cm) cystic duct remnant (cystic duct stump syndrome). Careful analysis, however, reveals that postcholecystectomy complaints are attributable to other causes in almost all patients in whom the symptom complex was originally thought to result from the existence of a long cystic duct stump. Accordingly, considerable care should be taken to investigate the possible role of other factors in the production of postcholecystectomy symptoms before attributing them to cystic duct stump syndrome.

SOD STENOSIS, SOD DYSKINESIA, AND BILIARY DYSKINESIA Symptoms of biliary colic accompanied by signs of recurrent, intermittent biliary obstruction may be produced by acalculous cholecystopathy, SOD stenosis, or SOD dyskinesia. SOD stenosis is thought to result from acute or chronic inflammation of the papilla of Vater or from glandular hyperplasia of the papillary segment. Five criteria have been used to define SOD stenosis: (1) upper abdominal pain, usually RUQ or epigastric; (2) abnormal liver tests; (3) dilatation of the CBD upon MRCP or ERCP examination; (4) delayed (>45 min) drainage of contrast material from the duct; and (5) increased basal pressure of the SOD. Treatment consists of endoscopic or surgical sphincteroplasty to ensure wide patency of the distal portions of both the bile and pancreatic ducts. The greater the number of the preceding criteria present, the greater is the likelihood that a patient does have a degree of SOD sufficient to justify correction. The factors usually considered as indications for sphincterotomy include (1) prolonged duration of symptoms, (2) lack of response to symptomatic treatment, (3) presence of severe disability, and (4) the patient's choice of sphincterotomy over surgery (given a clear understanding on the patient's part of the risks involved in both procedures).

Biliary SOD disorders are characterized by three criteria: (1) biliary pain, (2) absence of bile duct stones or other abnormalities, and (3) elevated liver enzymes or a dilated CBD, but not both. In this setting, either hepatobiliary scintigraphy or SOD manometry can support the diagnosis. Importantly, the presence of both elevated liver enzymes and a dilated CBD should raise the question of obstruction. Proposed mechanisms to account for SOD dysfunction include spasm of the sphincter, denervation sensitivity resulting in hypertonicity, and abnormalities in the sequencing or frequency rates of the sphincteric-contraction waves. When thorough evaluation has failed to demonstrate another cause for the pain and when cholangiographic and manometric criteria suggest a diagnosis of SOD dyskinesia, medical treatment with nitrites or anticholinergics to attempt pharmacologic relaxation of SOD has been proposed but not evaluated in detailed studies. Endoscopic biliary sphincterotomy (EBS) or surgical sphincterotomy may be indicated in patients who fail to respond to a 2- to 3-month trial of medical therapy, especially if SOD pressures are elevated. Approximately 45% of such patients have long-term pain relief after EBS. EBS has become the procedure of choice for removing bile duct stones and for other biliary and pancreatic problems.

BILE SALT–INDUCED DIARRHEA AND GASTRITIS Postcholecystectomy patients may develop symptoms of dyspepsia, which have been attributed to duodenogastric reflux of bile. However, firm data linking these symptoms to bile gastritis after surgical removal of the gallbladder are lacking. Cholecystectomy induces persistent changes in gut transit, and these changes effect a noticeable modification of bowel habits. Cholecystectomy shortens gut transit time by accelerating passage of the fecal bolus through the colon with marked acceleration in the right colon, thus causing an increase in colonic bile acid output and a shift in bile acid composition toward the more diarrheagenic secondary bile acids, that is, deoxycholic acid. Diarrhea that is severe enough, that is, three or more watery movements per day, can be classified as postcholecystectomy diarrhea, and this occurs in 5–10% of patients undergoing elective cholecystectomy. Treatment with bile acid–sequestering agents such as cholestyramine or colestipol is often effective in ameliorating troublesome diarrhea.

THE HYPERPLASTIC CHOLECYSTOSES

The term *hyperplastic cholecystoses* is used to denote a group of disorders of the gallbladder characterized by excessive proliferation of normal tissue components.

Adenomyomatosis is characterized by a benign proliferation of gallbladder surface epithelium with glandlike formations, extramural sinuses, transverse strictures, and/or fundal nodule ("adenoma" or "adenomyoma") formation.

Cholesterolosis is characterized by abnormal deposition of lipid, especially cholesteryl esters, within macrophages in the lamina propria of the gallbladder wall. In its diffuse form ("strawberry gallbladder"), the gallbladder mucosa is brick red and speckled with bright yellow flecks of lipid. The localized form shows solitary or multiple "cholesterol polyps" studding the gallbladder wall. Cholesterol stones of the gallbladder are found in nearly half the cases. Cholecystectomy is only indicated in adenomyomatosis or cholesterolosis when biliary symptoms are present.

The prevalence of gallbladder polyps in the adult population is 3–6% with a marked male predominance. Types of gallbladder polyps include cholesterol polyps, adenomyomas, inflammatory polyps, and adenomas (rare). No significant changes have been found over a 5-year period in asymptomatic patients with gallbladder polyps <6 mm and few changes in polyps 7–9 mm. Currently, cholecystectomy is recommended in symptomatic patients as well as in asymptomatic patients >50 years whose polyps are >10 mm or associated with gallstones or polyp growth on serial ultrasonography. A recent consensus conference of the Society of Radiologists in Ultrasound recommends that patients with polyps between 10 and 14 mm and low-risk radiologic features could be monitored instead of undergoing cholecystectomy.

DISEASES OF THE BILE DUCTS

CONGENITAL ANOMALIES

Biliary Atresia and Hypoplasia Atretic and hypoplastic lesions of the extrahepatic and large intrahepatic bile ducts are the most common biliary anomalies of clinical relevance encountered in infancy. The clinical picture is one of severe obstructive jaundice during the first month of life, with pale stools. When biliary atresia is suspected on the basis of clinical, laboratory, and imaging findings, the diagnosis is confirmed by surgical exploration and operative cholangiography. Approximately 10% of cases of biliary atresia are treatable with Roux-en-Y choledochojejunostomy, with the Kasai procedure (hepatic portoenterostomy) being attempted in the remainder in an effort to restore some bile flow. Most patients, even those having successful biliary-enteric anastomoses, eventually develop chronic cholangitis, extensive hepatic fibrosis, and portal hypertension.

Choledochal Cysts Cystic dilatation may involve the free portion of the CBD, that is, choledochal cyst, or may present as diverticulum formation in the intraduodenal segment. In the latter situation, chronic reflux of pancreatic juice into the biliary tree can produce inflammation and stenosis of the extrahepatic bile ducts, leading to cholangitis or biliary obstruction. Because the process may be gradual, ~50% of patients present with onset of symptoms after age 10. The diagnosis may be made by ultrasound, abdominal CT, MRCP, or cholangiography. Only one-third of patients show the classic triad of abdominal pain, jaundice, and an abdominal mass. Ultrasonographic detection of a cyst separate from the gallbladder should suggest the diagnosis of choledochal cyst, which can be confirmed by demonstrating the entrance of extrahepatic bile ducts into the cyst. Surgical treatment involves excision of the "cyst" and biliary-enteric anastomosis. Patients with choledochal cysts are at increased risk for cholangiocarcinoma.

Congenital Biliary Ectasia Dilatation of intrahepatic bile ducts may involve either the major intrahepatic radicles (Caroli's disease), the inter- and intralobular ducts (congenital hepatic fibrosis), or both. In Caroli's disease, clinical manifestations include recurrent cholangitis, abscess formation in and around the affected ducts, and, often, brown pigment gallstone formation within portions of ectatic intrahepatic biliary radicles. Ultrasound, MRCP, and CT are of great diagnostic value in demonstrating cystic dilatation of the intrahepatic bile ducts. Treatment with ongoing antibiotic therapy is usually undertaken in an effort to limit the frequency and severity of recurrent bouts of cholangitis. Progression to secondary biliary cirrhosis with portal hypertension, extrahepatic biliary obstruction, cholangiocarcinoma, or recurrent episodes of sepsis with hepatic abscess formation is common.

CHOLEDOCHOLITHIASIS

Pathophysiology and Clinical Manifestations Passage of gallstones into the CBD occurs in ~10–15% of patients with cholelithiasis. The incidence of common duct stones increases with increasing age of the patient, so up to 25% of elderly patients may have calculi in the common duct at the time of cholecystectomy. Undetected duct stones are left behind in ~1–5% of cholecystectomy patients. The overwhelming majority of bile duct stones are cholesterol stones formed in the gallbladder, which then migrate into the extrahepatic biliary tree through the cystic duct. Primary calculi arising de novo in the ducts are usually brown pigment stones developing in patients with (1) hepatobiliary parasitism or chronic, recurrent cholangitis; (2) congenital anomalies of the bile ducts (especially Caroli's disease); (3) dilated, sclerosed, or strictured ducts; or (4) an *MDR3* (ABCB4) gene defect leading to impaired biliary phospholipids secretion (low phospholipid–associated cholesterol cholelithiasis). Common duct stones may remain asymptomatic for years, may pass spontaneously into the duodenum, or (most often) may present with biliary colic or a complication.

Complications • **CHOLANGITIS** Cholangitis may be acute or chronic, and symptoms result from inflammation, which usually is caused by at least partial obstruction to the flow of bile. Bacteria are present on bile culture in ~75% of patients with acute cholangitis early in the symptomatic course. The characteristic presentation of acute cholangitis involves biliary pain, jaundice, and spiking fevers with chills (Charcot's triad). Blood cultures are frequently positive, and leukocytosis is typical. *Nonsuppurative acute cholangitis* is most common and may respond relatively rapidly to supportive measures and to treatment with antibiotics. In *suppurative acute cholangitis*, however, the presence of pus under pressure in a completely obstructed ductal system leads to symptoms of severe toxicity—mental confusion, bacteremia, and septic shock. Response to antibiotics alone in this setting is relatively poor, multiple hepatic abscesses are often present, and the mortality rate approaches 100% unless prompt endoscopic or surgical relief of the obstruction and drainage of infected bile are carried out. Endoscopic management of bacterial cholangitis is as effective as surgical intervention. ERCP with endoscopic sphincterotomy is safe and the preferred initial procedure for both establishing a definitive diagnosis and providing effective therapy.

OBSTRUCTIVE JAUNDICE Gradual obstruction of the CBD over a period of weeks or months usually leads to initial manifestations of jaundice or pruritus without associated symptoms of biliary colic or cholangitis. Painless jaundice may occur in patients with choledocholithiasis but is much more characteristic of biliary obstruction secondary to malignancy of the head of the pancreas, bile ducts, or ampulla of Vater.

In patients whose obstruction is secondary to choledocholithiasis, associated chronic calculous cholecystitis is very common, and the gallbladder in this setting may be unable to distend. The absence of a palpable gallbladder in most patients with biliary obstruction from duct stones is the basis for Courvoisier's law, that is, that the presence of a palpably enlarged gallbladder suggests that the biliary obstruction is secondary to an underlying malignancy rather than to calculous disease. Biliary obstruction causes progressive dilatation of the intrahepatic bile ducts as intrabiliary pressures rise. Hepatic bile flow is suppressed, and reabsorption and regurgitation of conjugated bilirubin into the bloodstream lead to jaundice accompanied by dark urine (bilirubinuria) and light-colored (acholic) stools.

CBD stones should be suspected in any patient with cholecystitis whose serum bilirubin level is >85.5 μmol/L (5 mg/dL). The maximum bilirubin level is seldom >256.5 μmol/L (15.0 mg/dL) in patients with choledocholithiasis unless concomitant hepatic or renal disease or another factor leading to marked hyperbilirubinemia exists. Serum bilirubin levels ≥342.0 μmol/L (20 mg/dL) should suggest the possibility of neoplastic obstruction. The serum alkaline phosphatase level is almost always elevated in biliary obstruction. A rise in alkaline phosphatase often precedes clinical jaundice and may be the only abnormality in routine liver function tests. There may be a two- to tenfold elevation of serum aminotransferases, especially in association with acute obstruction. Following relief of the obstructing process, serum aminotransferase elevations usually return rapidly to normal, while the serum bilirubin level may take 1–2 weeks to return to normal. The alkaline phosphatase level usually falls slowly, lagging behind the decrease in serum bilirubin.

PANCREATITIS The most common associated entity discovered in patients with nonalcoholic acute pancreatitis is biliary tract disease. Biochemical evidence of pancreatic inflammation complicates acute cholecystitis in 15% of cases and choledocholithiasis in >30%, and the common factor appears to be the passage of gallstones through the common duct. Coexisting pancreatitis should be suspected in patients with symptoms of cholecystitis who develop (1) back pain or pain to the left of the abdominal midline, (2) prolonged vomiting with paralytic ileus, or (3) a pleural effusion, especially on the left side. Surgical treatment of gallstone disease is usually associated with resolution of the pancreatitis.

SECONDARY BILIARY CIRRHOSIS Secondary biliary cirrhosis may complicate prolonged or intermittent duct obstruction with or without recurrent cholangitis. Although this complication may be seen in patients with choledocholithiasis, it is more common in cases of prolonged obstruction from stricture. Once established, secondary biliary cirrhosis may be progressive even after correction of the obstructing process, and increasingly severe hepatic cirrhosis may lead to portal hypertension, hepatic failure and death. Prolonged biliary obstruction may also be associated with clinically relevant deficiencies of the fat-soluble vitamins A, D, E, and K.

Diagnosis and Treatment The diagnosis of choledocholithiasis is made by cholangiography **(Table 357-3)**, either preoperatively by endoscopic retrograde cholangiogram (ERC) **(Fig. 357-2C)** or MRCP or intraoperatively at the time of cholecystectomy. As many as 15% of patients undergoing cholecystectomy will prove to have CBD stones. When CBD stones are suspected prior to laparoscopic cholecystectomy, preoperative ERCP with endoscopic papillotomy and stone extraction is the preferred approach. It not only provides stone clearance but also defines the anatomy of the biliary tree in relationship to the cystic duct. CBD stones should be suspected in gallstone patients who have any of the following risk factors: (1) a history of jaundice or pancreatitis, (2) abnormal tests of liver function, and (3) ultrasonographic or MRCP evidence of a dilated CBD or stones in the duct. Alternatively, if intraoperative cholangiography reveals retained stones, postoperative ERCP can be carried out. The need for preoperative ERCP is expected to decrease further as laparoscopic techniques for bile duct exploration improve.

The widespread use of laparoscopic cholecystectomy and ERCP has decreased the incidence of complicated biliary tract disease and the need for choledocholithotomy and T-tube drainage of the bile ducts. EBS followed by spontaneous passage or stone extraction is the treatment of choice in the management of patients with common duct stones, especially in elderly or poor-risk patients.

TRAUMA, STRICTURES, AND HEMOBILIA

Most benign strictures of the extrahepatic bile ducts result from surgical trauma and occur in about 1 in 500 cholecystectomies. Strictures may present with bile leak or abscess formation in the immediate postoperative period or with biliary obstruction or cholangitis as long as

TABLE 357-3 Diagnostic Evaluation of the Bile Ducts

DIAGNOSTIC ADVANTAGES	DIAGNOSTIC LIMITATIONS	CONTRAINDICATIONS	COMPLICATIONS	COMMENT
Ultrasound				
Rapid Simultaneous scanning of GB, liver, bile ducts, pancreas Accurate identification of dilated bile ducts Not limited by jaundice, pregnancy Guidance for fine-needle biopsy	Bowel gas Massive obesity Ascites Barium Partial bile duct obstruction Poor visualization of distal CBD	None	None	Initial procedure of choice in investigating possible biliary tract obstruction
Computed Tomography				
Simultaneous scanning of GB, liver, bile ducts, pancreas Accurate identification of dilated bile ducts, masses Not limited by jaundice, gas, obesity, ascites High-resolution image Guidance for fine-needle biopsy	Extreme cachexia Movement artifact Ileus Partial bile duct obstruction	Pregnancy	Reaction to iodinated contrast, if used	Indicated for evaluation of hepatic or pancreatic masses or for assessing for complications related to gallstones (pancreatitis) Procedure of choice in investigating possible biliary obstruction if diagnostic limitations limit US
Magnetic Resonance Cholangiopancreatography				
Noninvasive modality for visualizing pancreatic and biliary ducts Has excellent sensitivity for bile duct dilatation, biliary stricture, and intraductal abnormalities Can identify pancreatic duct dilatation or stricture, pancreatic duct stenosis, and pancreas divisum	Cannot offer therapeutic intervention High cost	Claustrophobia Certain metals (iron)	None	First choice to assess for choledocholithiasis given comparable sensitivity and specificity to ERCP
Endoscopic Retrograde Cholangiopancreatography				
Simultaneous pancreatography Best visualization of distal biliary tract Bile or pancreatic cytology Endoscopic sphincterotomy and stone removal Biliary manometry	Gastroduodenal obstruction Roux-en-Y biliary-enteric anastomosis	Pregnancy Acute pancreatitis Severe cardiopulmonary disease	Pancreatitis Cholangitis, sepsis Infected pancreatic pseudocyst Perforation (rare) Hypoxemia, aspiration	Cholangiogram of choice if there is believed to be a need for intervention: Diagnosed or high clinical probability of choledocholithiasis Biliary stricture requiring sampling and stenting Need for sphincterotomy such as sphincter of Oddi dysfunction
Percutaneous Transhepatic Cholangiogram				
Best when bile ducts dilated Best visualization of proximal biliary tract May be used to obtain bile cytology/culture Allows for percutaneous transhepatic drainage	Nondilated or sclerosed ducts	Pregnancy Uncorrectable coagulopathy Massive ascites Hepatic abscess	Bleeding Hemobilia Bile peritonitis Bacteremia, sepsis	Indicated for the drainage of obstructed and infected ducts when ERCP is contraindicated or failed
Endoscopic Ultrasound				
Most sensitive method to detect ampullary stones and exclude pathology in the head of the pancreas				Excellent for detecting choledocholithiasis

Abbreviations: CBD, common bile duct; ERCP, endoscopic retrograde cholangiopancreatography; GB, gallbladder; US, hepatobiliary ultrasound.

2 years or more following the inciting trauma. The diagnosis is established by percutaneous or endoscopic cholangiography. Endoscopic brushing of biliary strictures may be helpful in establishing the nature of the lesion and is more accurate than bile cytology alone. When positive exfoliative cytology is obtained, the diagnosis of a neoplastic stricture is established. This procedure is especially important in patients with primary sclerosing cholangitis (PSC) who are predisposed to the development of cholangiocarcinomas. Successful operative correction of non-PSC bile duct strictures by a skillful surgeon with duct-to-bowel anastomosis is usually possible, although mortality rates from surgical complications, recurrent cholangitis, or secondary biliary cirrhosis are high.

Hemobilia may follow traumatic or operative injury to the liver or bile ducts, intraductal rupture of a hepatic abscess or aneurysm of the hepatic artery, biliary or hepatic tumor hemorrhage, or mechanical complications of choledocholithiasis or hepatobiliary parasitism. Diagnostic procedures such as liver biopsy, PTC, and transhepatic biliary drainage catheter placement may also be complicated by hemobilia. Patients often present with a classic triad of biliary pain, obstructive jaundice, and melena or occult blood in the stools. The diagnosis is sometimes made by cholangiographic evidence of blood clot in the biliary tree, but selective angiographic verification may be required. Although minor episodes of hemobilia may resolve without intervention, arteriography and transcatheter embolization or surgical ligation of the bleeding vessel may be required.

■ EXTRINSIC COMPRESSION OF THE BILE DUCTS

Partial or complete biliary obstruction may be produced by extrinsic compression of the ducts. The most common cause of this form of obstructive jaundice is carcinoma of the head of the pancreas. Biliary obstruction may also occur as a complication of either acute or chronic pancreatitis or involvement of lymph nodes in the porta hepatis by lymphoma or metastatic carcinoma. The latter should be distinguished from cholestasis resulting from massive replacement of the liver by tumor.

■ HEPATOBILIARY PARASITISM

Infestation of the biliary tract by adult helminths or their ova may produce a chronic, recurrent pyogenic cholangitis with or without multiple hepatic abscesses, ductal stones, or biliary obstruction. This condition is relatively rare but does occur in inhabitants of southern China and elsewhere in Southeast Asia. The organisms most commonly involved are trematodes or flukes, including *Clonorchis sinensis*, *Opisthorchis viverrini* or *Opisthorchis felineus*, and *Fasciola hepatica*. The biliary tract also may be involved by intraductal migration of adult *Ascaris lumbricoides* from the duodenum or by intrabiliary rupture of hydatid cysts of the liver produced by *Echinococcus* spp. The diagnosis is made by cholangiography and the presence of characteristic ova on stool examination. When obstruction is present, the treatment of choice is laparotomy under antibiotic coverage, with common duct exploration and a biliary drainage procedure.

■ SCLEROSING CHOLANGITIS

PSC is characterized by a progressive, inflammatory, sclerosing, and obliterative process affecting the extrahepatic and/or the intrahepatic bile ducts. PSC is strongly associated with inflammatory bowel disease, especially ulcerative colitis.

Immunoglobulin G4 (IgG4)–associated cholangitis is a biliary disease of unknown etiology that presents with biochemical and cholangiographic features indistinguishable from PSC, is often associated with autoimmune pancreatitis and other fibrosing conditions and is characterized by elevated serum IgG4 and infiltration of IgG4-positive plasma cells in bile ducts and liver tissue. All patients diagnosed with sclerosing cholangitis should have a serum IgG4 level checked to rule out IgG4 disease as a cause of secondary sclerosing cholangitis, particularly if they do not have inflammatory bowel disease. Glucocorticoids are the initial treatment of choice. Relapse is common after steroid withdrawal, especially with proximal strictures. Steroid-sparing agents such as azathioprine or rituximab may be needed after relapse or for inadequate response **(Chap. 359)**.

Patients with PSC often present with signs and symptoms of chronic or intermittent biliary obstruction: RUQ abdominal pain, pruritus, jaundice, or acute cholangitis. Late in the course, complete biliary obstruction, secondary biliary cirrhosis, hepatic failure, and/or portal hypertension with bleeding varices may occur. The diagnosis is established by finding multifocal, diffusely distributed strictures with intervening segments of normal or dilated ducts, producing a beaded appearance on cholangiography **(Fig. 357-2*D*)**. The cholangiographic technique of choice in suspected cases is MRCP. When a diagnosis of sclerosing cholangitis has been established, causes of secondary sclerosing should be considered before reaching a diagnosis of PSC. Patients with PSC should undergo testing for associated diseases, especially inflammatory bowel disease, if that diagnosis has not already been established.

Small duct PSC is defined by the presence of chronic cholestasis and hepatic histology consistent with PSC in a patient with IBD, but with normal findings on cholangiography. Small duct PSC is found in ~5% of patients with PSC and is associated with a significantly better long-term prognosis. However, such patients may progress to classic PSC and/or end-stage liver disease with consequent necessity of liver transplantation.

In patients with AIDS, cholangiopancreatography may demonstrate a broad range of biliary tract changes as well as pancreatic duct obstruction and occasionally pancreatitis **(Chap. 208)**. Further, biliary tract lesions in AIDS include infection and cholangiopancreatographic changes of sclerosing cholangitis. Changes noted include (1) diffuse involvement of intrahepatic bile ducts alone, (2) involvement of both intra- and extrahepatic bile ducts, (3) ampullary stenosis, (4) stricture of the intrapancreatic portion of the CBD, and (5) pancreatic duct involvement. Associated infectious organisms include *Cryptosporidium*, *Mycobacterium avium-intracellulare*, cytomegalovirus, *Microsporidia*, and *Isospora*. ERCP sphincterotomy can provide significant pain reduction in patients with AIDS-associated papillary stenosis.

TREATMENT

Primary Sclerosing Cholangitis

There is no proven medical therapy for PSC. Therapy to treat pruritus associated with PSC includes cholestyramine, rifampin, fenofibrate/bezafibrate, sertraline, and naltrexone. Antibiotics are useful when bacterial cholangitis complicates the clinical picture. Vitamin D and calcium supplementation may be used as initial therapy to help prevent the loss of bone mass frequently seen in patients with chronic cholestasis. In cases where high-grade biliary obstruction (dominant strictures) has occurred, balloon dilatation is preferred over stenting due to the higher complication rate associated with stenting including pancreatitis and cholangitis. Only rarely is surgical intervention indicated. PSC is a progressive disease with a median survival of 12–18 years following the diagnosis, regardless of therapy. Four variables (age, serum bilirubin level, histologic stage, and splenomegaly) predict survival in patients with PSC and serve as the basis for a risk score. PSC is a common indication for liver transplantation.

■ FURTHER READING

Baron TH et al: Interventional approaches to gallbladder disease. N Engl J Med 373:357, 2015.

Hu H et al: Gut microbiota promotes cholesterol gallstone formation by modulating bile acid composition and biliary cholesterol secretion. Nature Comm 13:252, 2022.

Kamaya A et al: Management of incidentally detected gallbladder polyps: Society of Radiologists in Ultrasound consensus conference recommendations. Radiology 305:227, 2022.

Lindor K et al: American College of Gastroenterology (ACG) guidelines: Primary sclerosing cholangitis. Hepatology 51:660, 2010.

Ryl JK et al: Clinical features of acute acalculous cholecystitis. J Clin Gastroenterol 36:166, 2003.

Strasberg S: Clinical practice. Acute calculous cholecystitis. N Engl J Med 358:2804, 2008.

Section 4 Disorders of the Pancreas

358 Approach to the Patient with Pancreatic Disease

Somashekar G. Krishna, Darwin L. Conwell, Phil A. Hart

GENERAL CONSIDERATIONS

Globally, pancreatic disorders, including acute and chronic pancreatitis, pancreatic cysts, and pancreatic cancer, are challenging to manage and associated with a high burden on health care resources. While the relationships between these diseases are multifaceted, there is ongoing scientific progress and a growing understanding in this field. Acute pancreatitis is one of the most common reasons for hospitalizations in gastroenterology, and there is increasing evidence of sequelae including diabetes, exocrine pancreas insufficiency, and chronic pancreatitis. In elderly patients, acute pancreatitis may serve as an early symptom of pancreatic cancer. Chronic pancreatitis, an irreversible disease of the pancreas, associated with poor quality of life due to abdominal pain and associated exocrine insufficiency, is also an established risk factor for pancreatic cancer. Pancreatic cysts, mostly incidental, are increasingly detected on cross-sectional abdominal imaging studies. Although only a small proportion of pancreatic cysts can progress to pancreatic cancer, the diagnostic uncertainty can introduce unwanted anxiety to patients and treating physicians. Meanwhile, with persistently high mortality rates, the incidence of pancreatic adenocarcinoma is increasing and is the seventh leading cause of cancer-related death in the industrialized world and the third most common in the United States.

As emphasized in **Chap. 359**, the etiologies and clinical manifestations of pancreatitis are quite varied. Although it is well-appreciated that acute pancreatitis is frequently secondary to biliary tract disease or alcohol abuse, it can also be caused by medications, genetic mutations, and trauma. In ~30% of patients with acute pancreatitis and 25–40% of patients with chronic pancreatitis, the etiology is initially unexplained.

The global pooled incidence of acute pancreatitis is ~33.7 cases (95% confidence interval [CI], 23.3–48.8) with 1.16 deaths (95% CI, 0.85–1.6) per 100,000 person-years. The global pooled incidence of chronic pancreatitis is ~9.6 cases (95% CI, 7.9–11.8) with 0.09 attributable deaths (95% CI, 0.02–0.5) per 100,000 person-years. In the United States, the number of patients admitted to the hospital with acute pancreatitis is increasing, with estimated rates of almost 300,000 annually, whereas the number of patients hospitalized for chronic pancreatitis is decreasing, with recent estimates of ~13,000 admissions per year. Chronic pancreatitis has an annual prevalence of 42–73 cases per 100,000 adults in the United States, although higher prevalence rates (0.04–5%) have been noted among adults at autopsy. Together, acute and chronic pancreatic disease costs an estimated $3 billion annually in health care expenditures. During the COVID-19 pandemic, it was noted that the infection was associated with elevated pancreas enzyme serum levels and presumed acute pancreatitis, though causal relationships have not been definitively established.

The diagnosis of acute pancreatitis is generally defined based on a combination of laboratory, imaging, and clinical symptoms. The diagnosis of chronic pancreatitis, especially in mild disease, is hampered by the relative inaccessibility of the pancreas to direct examination and the nonspecificity of the associated abdominal pain. Many patients with chronic pancreatitis do not have elevated blood amylase or lipase levels. Some patients with chronic pancreatitis develop signs and symptoms of exocrine pancreatic insufficiency (EPI), and thus, objective evidence for pancreatic disease can be demonstrated. However, there is a large reservoir of pancreatic exocrine function. Maldigestion of fat and protein becomes evident only when more than 90% of the pancreas is functionally damaged or obstructed. Noninvasive, indirect tests of pancreatic exocrine function (e.g., fecal elastase) are much more likely to give abnormal results in patients with obvious advanced pancreatic disease (i.e., pancreatic calcification, steatorrhea, or diabetes mellitus) than in patients with occult disease. Invasive, direct tests of pancreatic secretory function (e.g., secretin stimulation test) are the most sensitive and specific tests to detect early chronic pancreatic disease when imaging is equivocal or normal.

The increasing utilization of cross-sectional imaging modalities with their improved resolution has contributed to a high prevalence (2–5% with computed tomography [CT] scans, 20–30% with magnetic resonance imaging [MRI]) of incidentally detected pancreatic cysts. The most common cyst type encountered is an intraductal papillary mucinous neoplasm (IPMN), which is classified as a precancerous mucinous cyst. In the absence of high-risk features, radiographic surveillance is typically recommended **(Fig. 358-1)**. Mucinous cystic neoplasms (MCNs) are relatively less common and occur almost exclusively in women. Among the neoplastic cysts, serous cystadenomas have a negligible risk of progression to malignancy. Other infrequent neoplastic cysts include neuroendocrine tumors and solid pseudopapillary neoplasms. The most commonly encountered benign cyst is a pseudocyst, which can occur in patients with a history of acute

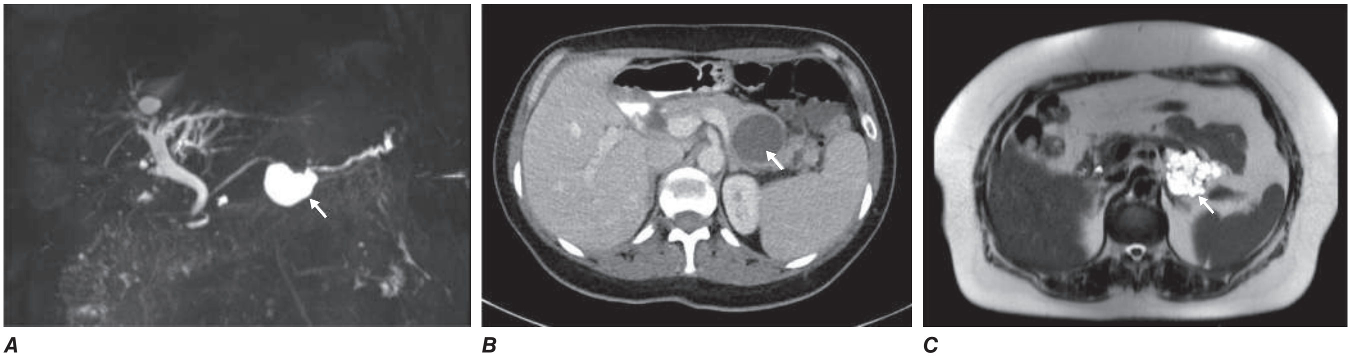

FIGURE 358-1 ***A.*** Side-branch intraductal papillary mucinous neoplasm (magnetic resonance imaging [MRI] with magnetic resonance cholangiopancreatography [MRCP]). T2-weighted MRCP image demonstrates a dominant, lobulated, hyperintense cystic structure (*arrow*) within the posterior body of the pancreas. The pancreatic duct upstream from the cyst is dilated and irregular. Endoscopic ultrasound and fine-needle aspiration of cyst fluid were consistent with a mucinous cyst. Surgical histopathology revealed an infiltrating moderately differentiated adenocarcinoma, 0.3 cm, arising in a background of an intraductal papillary mucinous neoplasm (IPMN). ***B.*** Mucinous cystic neoplasm (computed tomography [CT] scan). In the tail of the pancreas, there is a well-circumscribed hypodense cyst (*arrow*) without any nodular enhancing components. Endoscopic ultrasound and fine-needle aspiration of cyst fluid were suggestive of a mucinous cyst. Surgical histopathology revealed a mucinous cystic neoplasm (3.4 cm) with low-grade dysplasia. The stroma of the cyst demonstrated diffuse positivity for progesterone receptor and focal positivity for CD10 (ovarian stroma), confirming the diagnosis. ***C.*** Serous cystadenoma (MRI). A lobulated microcystic cyst (*arrow*) is observed in the tail of the pancreas. Neither a communication with the main pancreatic duct nor intracystic soft tissue enhancing nodular components were observed. However, the cyst continued to increase in size, and a distal pancreatectomy was performed. Histopathology revealed a serous microcystic adenoma. *(Courtesy of Dr. Z.K. Shah, The Ohio State University Wexner Medical Center; with permission.)*

or chronic pancreatitis. The challenges with accurately predicting the risk of malignant transformation of precancerous pancreatic cysts has contributed to the growing number of patients on imaging surveillance protocols placing a burden on health care systems in industrialized nations.

TESTS USEFUL IN THE DIAGNOSIS OF PANCREATIC DISEASE

Several tests are of value in the evaluation of pancreatic disease. Examples of specific tests and their usefulness in the diagnosis of acute and chronic pancreatitis are summarized in **Table 358-1** and **Fig. 358-2**. At some

TABLE 358-1 Tests Useful in the Diagnosis of Acute and Chronic Pancreatitis and Pancreatic Neoplasms

TEST	PRINCIPLE	COMMENT
Pancreatic Enzymes in Body Fluids		
Serum lipase	Pancreatic inflammation leads to increased serum enzyme levels	Enzyme measurement of choice for the diagnosis of acute pancreatitis; increased specificity if the level is more than three times the upper limit of normal (3× ULN)
Amylase		
1. Serum	Pancreatic inflammation leads to increased serum enzyme levels	Simple; increased specificity if the level is >3× ULN; may be falsely normal in patients with hypertriglyceridemic pancreatitis
2. Urine	Renal clearance of amylase is increased in acute pancreatitis	Infrequently used
3. Ascitic fluid	Disruption of gland or main pancreatic duct leads to increased amylase concentration	Can help establish source of ascites; false positives occur with intestinal obstruction and perforated ulcer; can also measure lipase
4. Pleural fluid	Exudative pleural effusion with pancreatitis	False positives occur with carcinoma of the lung and esophageal perforation
Studies Pertaining to Pancreatic Structure		
Radiologic and radionuclide tests		
1. Plain film of the abdomen or upper gastrointestinal x-rays	Can demonstrate large calcifications in chronic pancreatitis	Infrequently used
1. Ultrasonography (US)	Can provide limited information on edema, inflammation, calcification, pseudocysts, and mass lesions	Simple, noninvasive; sequential studies quite feasible; useful in diagnosis of gallstones; pancreas visualization limited by interference from overlying bowel gas
2. Computed tomography (CT) scan	Permits detailed visualization of pancreas and surrounding structures, pancreatic fluid collection, pseudocyst; assessment of necrosis or interstitial disease	Useful in the diagnosis of pancreatic calcification, dilated pancreatic ducts, and pancreatic tumors; may not be able to distinguish between inflammatory and neoplastic mass lesions; multiphasic CT scans are the preferred imaging modality for staging pancreatic cancer; IV contrast is needed for characterization of most features
3. Magnetic resonance imaging (MRI) and cholangiopancreatography (MRCP)	Permits noninvasive detailed evaluation of the pancreatic parenchyma, biliary and pancreatic ducts, adjacent soft tissues, and vascular structures.	Has mostly replaced ERCP for diagnostic assessment of the pancreatic duct; more sensitive than CT scan for detection of mild pancreatitis, necrosis, choledocholithiasis, pancreatic ductal abnormalities, and cystic neoplasms; no exposure to ionizing radiation
4. Endoscopic ultrasonography (EUS) and fine-needle aspiration/biopsy (FNA/B)	High-frequency transducer used with EUS produces very-high-resolution images permitting focused evaluation of pancreatic parenchyma and biliary and pancreatic ducts, and FNA/B provides targeted tissue acquisition	Can be used to assess gallstones, choledocholithiasis, chronic pancreatitis, pancreatic masses, and cystic neoplasms; FNA/B facilitates diagnostic and therapeutic management of pancreatic diseases
5. Endoscopic retrograde cholangiopancreatography (ERCP)	Cannulation of pancreatic and/or common bile duct permits visualization of pancreaticobiliary ductal system	Primarily a therapeutic procedure; invasive with risks for iatrogenic complications
Tests of Exocrine Pancreatic Function		
Direct stimulation of the pancreas with analysis of duodenal contents		
1. Secretin test	Secretin leads to increased output of pancreatic juice and HCO_3^-; pancreatic secretory response is related to the functional mass of pancreatic tissue; involves duodenal intubation and fluoroscopic placement of gastroduodenal tube	Sensitive to detect occult disease; poorly defined normal enzyme response; large secretory reserve capacity of the pancreas; rarely performed
2. Endoscopic pancreatic function test (ePFT)	Secretin-stimulated collection of pancreatic juice performed during upper endoscopy; replaces need for tube placement in the duodenum	Sensitive to detect occult disease; high negative predictive value for chronic pancreatitis; requires sedation
3. EUS-ePFT	Combines endosonographic evaluation of the pancreas and endoscopic collection of pancreatic juice	Single endoscopic evaluation of pancreatic structure and function
4. Secretin-stimulated MRCP	Combines imaging evaluation of the pancreas and a semiquantitative estimation of pancreatic juice output in the duodenum	Improved visualization of pancreatic ductal anatomy; functional evaluation is less accurate than ePFT; noninvasive
Measurement of intraluminal digestion products		
1. Stool fat determination	Lack of lipolytic enzymes brings about impaired fat digestion; quantitative 72-h stool collection and estimation are more reliable than qualitative analysis of a random stool sample	Reliable reference standard for defining severity of fat malabsorption; does not distinguish between pancreatic and nonpancreatic cause of malabsorption
Measurement of pancreatic enzymes in feces		
1. Fecal elastase	Pancreatic secretion of proteolytic enzymes; not degraded in intestine	Diagnostic accuracy is highest when the pretest probability is high and the value is <100 μg/g; false positives will occur in patients with nonformed stools

	• **Clinical signs and symptoms suggestive of chronic pancreatic disease:** abdominal pain, nausea, weight loss, steatorrhea, malabsorption, history of alcohol abuse, recurrent pancreatitis, fatty-food intolerance • Perform history, physical examination, review of laboratory studies; consider fecal elastase measurement
Step 1	• **Contrast-enhanced CT scan** • *CP diagnostic criteria*: calcifications in combination with atrophy and/or dilated duct • Diagnostic criteria met; no further imaging needed • Inconclusive or nondiagnostic results; continue to step 2
Step 2	• **MRI and MRCP, with or without secretin enhancement (sMRCP)** • *CP diagnostic criteria*: Cambridge class III,[a] dilated duct, atrophy of gland, filling defects in duct suggestive of stones • Diagnostic criteria met; no further imaging needed • Inconclusive or nondiagnostic results; continue to step 3
Step 3	• **EUS with quantification of parenchymal and ductal criteria** • *CP diagnostic criteria*: ≥5 EUS CP criteria • Diagnostic criteria met; no further imaging needed • Inconclusive or nondiagnostic results; continue to step 4
Step 4	• **Pancreas function test (with secretin)—endoscopic (ePFT) collection method preferred; consider combining ePFT with EUS** • *CP diagnostic criteria*: peak [bicarbonate] <80 mEq/L • Diagnostic criteria met; no further imaging needed • Inconclusive or nondiagnostic results require monitoring of signs and symptoms and repeat testing in 6 months–1 year

FIGURE 358-2 A stepwise diagnostic approach to the patient with suspected chronic pancreatitis (CP). Endoscopic ultrasonography (EUS) and magnetic resonance imaging (MRI) with secretin-stimulated magnetic resonance cholangiopancreatography (sMRCP/MRCP) are appropriate diagnostic alternatives to endoscopic retrograde cholangiopancreatography (ERCP). CT, computed tomography; ePFT, endoscopic pancreas function test. [a]Cambridge classification of pancreatic duct findings: class 0: normal—visualization of complete normal ductal anatomy; class I: equivocal—normal main duct, 1–3 abnormal side branches; class II: mild—normal main duct, >3 abnormal side branches; class III—dilated and irregular main duct, >3 abnormal side branches, small (<10 mm) cysts; class IV—irregular main duct, intraductal calculi, strictures, obstruction with dilation, or large (>10 mm) cysts.

institutions, pancreatic function tests are available and performed if the diagnosis of chronic pancreatitis remains a possibility after noninvasive tests (i.e., ultrasound, CT scan, MRI with magnetic resonance cholangiopancreatography [MRCP]) or invasive tests (i.e., endoscopic retrograde cholangiopancreatography [ERCP], endoscopic ultrasound [EUS]) have given normal or inconclusive results. In this regard, tests using *direct* stimulation of the pancreas with secretin are the most sensitive.

Pancreatic Enzymes in Body Fluids The serum amylase and lipase levels are widely used as screening tests for acute pancreatitis in the patient with acute abdominal pain or back pain. Lipase is more specific for the pancreas, and values greater than three times the upper limit of normal (3× ULN) in combination with epigastric pain strongly suggest the diagnosis of acute pancreatitis. In acute pancreatitis, the serum amylase and lipase are usually elevated within 24 h of onset and remain so for 3–7 days. Levels usually return to normal within 7 days unless there is pancreatic ductal disruption, ductal obstruction, or pseudocyst formation. Approximately 85% of patients with acute pancreatitis have threefold or greater elevated serum lipase and amylase levels. The values may be normal if (1) there is a delay (2–5 days) before blood samples are obtained, (2) the underlying disorder is chronic pancreatitis rather than acute pancreatitis, or (3) hypertriglyceridemia is present. Patients with hypertriglyceridemia and acute pancreatitis have been found to have spuriously low levels of amylase and perhaps lipase activity. In the absence of objective evidence of pancreatitis by abdominal ultrasound, contrast-enhanced CT scan, MRI with MRCP, or EUS, mild to moderate elevations of amylase and/or lipase are not helpful in making a diagnosis of chronic pancreatitis.

It should be noted that the serum amylase can be elevated in other conditions **(Table 358-2)**, in part because the enzyme is found in many organs. In addition to the pancreas and salivary glands, small quantities of amylase are found in the tissues of the fallopian tubes, lung, thyroid, and tonsils and can be produced by various tumors (carcinomas of the lung, esophagus, breast, and ovary). Isoamylase determinations do not accurately distinguish elevated blood amylase levels from pancreatic or nonpancreatic sources. In patients with unexplained hyperamylasemia, the measurement of macroamylase can avoid numerous tests in patients with this rare disorder.

Elevation of ascitic fluid amylase occurs in acute pancreatitis as well as in (1) ascites due to disruption of the main pancreatic duct or a leaking pseudocyst and (2) other abdominal disorders that simulate pancreatitis (e.g., intestinal obstruction, intestinal infarction, or perforated peptic ulcer). Elevation of pleural fluid amylase can occur in acute pancreatitis, chronic pancreatitis, carcinoma of the lung, and esophageal perforation. Serum lipase is the single best enzyme to measure for the diagnosis of acute pancreatitis. It is important to acknowledge that levels are often mildly elevated in the setting of renal disease, so determining whether a patient with renal failure and abdominal pain has pancreatitis remains a challenging clinical problem. One study found that serum amylase levels were elevated in patients with renal dysfunction only when creatinine clearance was <0.8 mL/s (<50 mL/min). In such patients, the serum amylase level was invariably <500 IU/L in the absence of objective evidence of acute pancreatitis. In that study, serum lipase and trypsin levels paralleled serum amylase values. With these limitations in mind, the recommended screening test for acute pancreatitis in renal disease is serum lipase, but a high index of clinical suspicion is needed based on symptoms. Elevations in serum lipase >3× ULN due to nonpancreatic etiology can be observed in hepatobiliary or gastrointestinal malignancies, septicemia, liver cirrhosis, systemic lupus erythematosus, severe head injury, chronic alcoholism, diabetes mellitus, and post-ERCP without any associated evidence of pancreatitis.

COVID-19 infection has been associated with asymptomatic elevations of both amylase and lipase. In a systematic review of 21 studies involving 36,496 patients during the COVID-19 pandemic, the pooled prevalence of hyperlipasemia (>3 × ULN) and hyperamylasemia (>3 × ULN) was 5.6% (95% CI, 2.8–9.3%) and 4.0% (95% CI, 0.9–8.7%), respectively. Importantly, the overall prevalence of acute pancreatitis in this cohort was 1.7%. Abnormal levels of amylase and lipase were more closely associated with the severity of COVID-19 than a diagnosis of acute pancreatitis.

Studies Pertaining to Pancreatic Structure • **RADIOLOGIC TESTS** Plain films of the abdomen rarely provide useful information related to pancreatic disease and have been superseded by more detailed imaging studies (ultrasound, EUS, CT, and MRI with MRCP).

TABLE 358-2 Causes of Hyperamylasemia

Pancreatic Disease

I. Pancreatitis
   A. Acute
   B. Chronic: ductal obstruction
   C. Complications of pancreatitis
      1. Pancreatic pseudocyst
      2. Ascites caused by pancreatic duct disruption
      3. Pancreatic necrosis
II. Pancreatic trauma
III. Pancreatic adenocarcinoma

Nonpancreatic Disorders

I. Renal insufficiency
II. Salivary gland lesions
   A. Mumps
   B. Calculus
   C. Irradiation sialadenitis
   D. Maxillofacial surgery
III. "Tumor" hyperamylasemia
   A. Carcinoma of the lung, esophagus, breast, or ovary
IV. Macroamylasemia
V. Burns
VI. Diabetes mellitus, particularly when ketoacidosis is present
VII. Pregnancy
VIII. Renal transplantation
IX. COVID-19 infection
X. Cerebral trauma
XI. Drugs: opiates

Other Abdominal Disorders

I. Biliary tract disease: cholecystitis, choledocholithiasis
II. Intraabdominal disease
   A. Perforated or penetrating peptic ulcer
   B. Intestinal obstruction or inflammation
   C. Ruptured ectopic pregnancy
   D. Peritonitis
   E. Aortic aneurysm
   F. Postoperative hyperamylasemia

Ultrasonography (US) can provide important information in the initial emergency ward evaluation of patients with acute pancreatitis, chronic pancreatitis, pseudocysts, and pancreatic adenocarcinoma. Sonographic changes can indicate the presence of edema, inflammation, and calcification (not obvious on plain films of the abdomen), as well as gallstones, biliary dilation, pseudocysts, and mass lesions. In acute pancreatitis, the pancreas is characteristically enlarged. In pancreatic pseudocyst, the usual appearance is primarily that of a smooth, round fluid collection. Pancreatic adenocarcinoma distorts the usual landmarks, and mass lesions >3.0 cm are usually detected as localized, solid lesions. US is often the initial investigation for patients with suspected pancreatic disease. However, obesity and excess intestinal bowel gas can interfere with pancreatic imaging, limiting its sensitivity.

CT with intravenous contrast is the best imaging study for the assessment of complications of acute and chronic pancreatitis. It is especially useful in the detection of pancreatic and peripancreatic acute fluid collections, fluid-containing lesions such as pseudocysts, walled-off necrosis **(see Chap. 359, Figs. 359-1, 359-2, and 359-4)**, and pancreatic neoplasms. Acute pancreatitis is characterized by (1) enlargement of the pancreas, (2) distortion of the pancreatic contour with peripancreatic stranding of adjacent fat tissue, and/or (3) the presence of pancreatic fluid that has a different attenuation coefficient than normal pancreas. When possible, CT scans should ideally be performed with oral and intravenous contrast to detect areas of pancreatic necrosis. The major benefit of CT scan in acute pancreatitis is the diagnosis of pancreatic necrosis in patients not responding to conservative management within 72 h. It may take 48–72 h to develop perfusion defects indicative of pancreatic necrosis. Therefore, if acute pancreatitis is confirmed with serology and physical examination findings, CT scan in the first 3 days is not recommended to minimize risk of contrast-induced nephropathy and unnecessary health care costs. Improved imaging technology and increased resolution are facilitated by multiphasic CT scans using multidetector technology (MDCT) in which a pancreas protocol consisting of dual-phase scanning with intravenous contrast is utilized for the detection and staging of pancreatic cancers. While the sensitivity of MDCT for detecting smaller (≤2 cm) lesions is lower, the reported overall sensitivity for pancreatic cancers is 76–97%. The contraindications to using intravenous contrast include renal failure (serum creatinine >2 mg/dL) and a history of severe allergic reaction to iodinated contrast agents. In situations where EUS is not available, CT-guided percutaneous aspiration or biopsy of a pancreatic mass can be performed. Prior to the major advance of EUS-guided fine-needle aspiration (FNA), CT-guided biopsy was utilized in the preceding decades and is regarded as a safe procedure.

MRI and MRCP provide excellent imaging of the bile duct, pancreatic duct, and pancreas parenchyma in both acute pancreatitis and chronic pancreatitis. MRI is more sensitive than transabdominal US and CT scans and comparable to EUS for the detection of choledocholithiasis. Similar to CT, MRI can evaluate for the severity of acute pancreatitis. Moreover, T2-weighted MRI of fluid collections can differentiate necrotic debris from fluid in suspected walled-off necrosis, and T1 imaging can diagnose hemorrhage in suspected pseudoaneurysm rupture. In chronic pancreatitis, secretin-enhanced MRCP is a method to enhance the evaluation of major and minor ductal changes. While imaging is comparable to CT for evaluating pancreatic mass lesions, MRI with MRCP is the preferred imaging modality for evaluating pancreatic cystic lesions. Nephrogenic systemic fibrosis has been described in patients with chronic renal failure following exposure to the gadolinium contrast, but incidence rates are extraordinarily low with contemporary contrast agents.

EUS produces high-resolution images of the bile duct, pancreatic parenchyma, and pancreatic duct with a transducer fixed to an endoscope that can be directed onto the surface of the pancreas through the stomach or duodenum. EUS is not beneficial for the evaluation of pancreas during acute pancreatitis. It is preferable to perform EUS after the resolution of acute pancreatitis (~4 weeks) to detect any predisposing factors, including malignancy, choledocholithiasis, pancreatic divisum, or ampullary lesions. EUS can be combined with ERCP in a single session and is increasingly preferred for the diagnosis and management of choledocholithiasis in acute pancreatitis and pancreatic neoplasm with biliary obstruction. EUS has been studied as a diagnostic modality for chronic pancreatitis. Criteria for abnormalities on EUS in severe chronic pancreatic disease have been developed. There is general agreement that the presence of five or more of the nine criteria listed in **Table 358-3** is highly predictive of chronic pancreatitis in the correct clinical context. The sensitivity of EUS (81%; 95% CI, 70–89%) to diagnose chronic pancreatitis is comparable to that of MRI/MRCP (78%; 95% CI, 69–85%) and better than CT (75%; 95% CI, 66–83%); however, nonspecific changes are commonly seen in the pancreas that may be attributable to cigarette smoking, diabetes, or normal aging. EUS also facilitates the delivery of nerve-blocking agents via fine-needle injection in patients suffering from pancreatic pain from chronic pancreatitis (celiac plexus block) or cancer (celiac plexus neurolysis). When

TABLE 358-3 Endoscopic Ultrasonographic Criteria for Chronic Pancreatitis (Total Criteria = 9)

DUCTAL	PARENCHYMAL
Stones	Echogenic strands
Hyperechoic main duct margins	Echogenic foci
Main duct irregularity	Lobular contour
Main duct dilatation	Cysts
Visible side branches	

clinically suspected, EUS imaging is more sensitive than MDCT for the detection of pancreatic malignancy and permits fine-needle aspiration/biopsy (FNA/B). Currently, EUS-guided FNA/B is the diagnostic modality of choice for the acquisition of diagnostic tissue and cyst fluid in patients with pancreatic masses and cystic lesions, respectively.

Although a pancreatogram during ERCP is the most specific and sensitive test for evaluating the ductal anatomy, EUS and MRI/MRCP have largely replaced ERCP in the diagnostic evaluation of pancreatic disease to avoid the risk of complications. Therefore, ERCP is primarily of therapeutic value after CT, EUS, or MRI/MRCP has detected abnormalities requiring endoscopic treatment. ERCP is the most sensitive modality for the detection of bile duct stones. In the management of acute biliary pancreatitis, ERCP should not be unduly delayed in patients with high clinical suspicion of biliary obstruction. In chronic pancreatitis, ERCP abnormalities in the main pancreatic duct and side branches have been outlined by the Cambridge classification (Fig. 358-2). The presence of ductal stenosis and irregularity can make it difficult to distinguish chronic pancreatitis from pancreatic adenocarcinoma. It is important to be aware that ERCP changes interpreted as indicating chronic pancreatitis actually may be due to the effects of aging on the pancreatic duct, sequelae of a recent attack of acute pancreatitis, or changes secondary to placement of pancreatic duct stent. Pancreatic adenocarcinoma is characterized by stenosis or obstruction of either the pancreatic duct or the common bile duct; both ductal systems are often abnormal (double-duct sign). When indicated, ERCP permits acquisition of diagnostic tissue as in biopsy of ampullary lesions or biliary brushings for distal bile duct strictures. Elevated serum amylase levels after ERCP have been reported in the majority of patients, and clinical pancreatitis has been reported in 5–10% of patients. Historically, pancreatic duct stents were commonly placed to prevent post-ERCP pancreatitis. Recent data, however, have revealed that the periprocedural administration of rectal indomethacin can decrease the incidence of post-ERCP pancreatitis. In a randomized multicenter study, it was demonstrated that the combination of rectal indomethacin with prophylactic pancreatic duct stents was more effective than indomethacin alone in preventing post-ERCP pancreatitis in high-risk patients.

TESTS OF EXOCRINE PANCREATIC FUNCTION

Pancreatic function tests (Table 358-1) can be divided into the following:

1. *Direct stimulation of the pancreas* by IV administration of secretin followed by collection and measurement of duodenal contents: The secretin test, used to detect diffuse pancreatic disease, is based on the physiologic principle that the pancreatic secretory response is directly related to the functional mass of pancreatic tissue. In the standard assay, secretin is given IV in a dose of 0.2 μg/kg of synthetic human secretin as a bolus. Normal values for the standard secretin test are (1) volume output >2 mL/kg per h, (2) bicarbonate (HCO_3^-) concentration >80 mmol/L, and (3) HCO_3^- output >10 mmol/L in 1 h. The most reproducible measurement, giving the highest level of discrimination between normal subjects and patients with chronic pancreas dysfunction, appears to be the maximal bicarbonate concentration. A cutoff point <80 mmol/L is considered abnormal and suggestive of reduced secretory function that is most commonly observed in early chronic pancreatitis.
2. There may be a dissociation between the results of the secretin test and other tests of absorptive function. For example, patients with chronic pancreatitis often have abnormally low outputs of HCO_3^- after secretin but have normal fecal fat excretion. The secretin test directly measures the secretory capacity of ductular epithelium, whereas fecal fat excretion indirectly reflects intraluminal lipolytic activity. Steatorrhea does not occur until intraluminal levels of lipase are markedly reduced, underscoring the fact that only small amounts of enzymes are necessary for intraluminal digestive activities. It must be emphasized that an abnormal secretin test result suggests that pancreatic ductal secretory function is abnormal. This is an early abnormality in chronic pancreatitis but should not be considered diagnostic and must be interpreted within the proper clinical context.
3. *Measurement of fecal pancreatic enzymes* such as elastase: Measurement of *intraluminal digestion products* (i.e., undigested muscle fibers, stool fat, and fecal nitrogen) is discussed in **Chap. 336**. The amount of human elastase in stool reflects the pancreatic output of this proteolytic enzyme. Decreased fecal elastase-1 (FE-1) activity in stool is a test to detect severe EPI in patients with chronic pancreatitis and cystic fibrosis. FE-1 levels >200 μg/g are normal, levels of 100–200 μg/g are considered mild-moderate EPI, and levels <100 μg/g are severe EPI. Although the test is simple and noninvasive, it can yield false-positive results if stools are not formed and should not generally be used for the evaluation of a patient with diarrhea. False-positive results have also been observed in diabetes and irritable bowel syndrome.

Tests useful in the diagnosis of EPI and the differential diagnosis of malabsorption are also discussed in Chaps. 336 and 359.

FURTHER READING

Conwell DL et al: American Pancreatic Association practice guidelines in chronic pancreatitis: Evidence-based report on diagnostic guidelines. Pancreas 43:1143, 2014.

Hart PA et al: Endoscopic pancreas fluid collection: Methods and relevance for clinical care and translational science. Am J Gastroenterol 111:1258, 2016.

Petrov MS, Yadav D: Global epidemiology and holistic prevention of pancreatitis. Nat Rev Gastroenterol Hepatol 16:175, 2019.

Singh VK et al: Diagnosis and management of chronic pancreatitis: A review. JAMA 322:2422, 2019.

359

Acute and Chronic Pancreatitis

Phil A. Hart, Darwin L. Conwell,
Somashekar G. Krishna

BIOCHEMISTRY AND PHYSIOLOGY OF PANCREATIC EXOCRINE SECRETION

GENERAL CONSIDERATIONS

The pancreas secretes 1500–3000 mL of isosmotic alkaline (pH >8) fluid per day containing ~20 enzymes. Pancreatic secretions provide the enzymes and bicarbonate needed to perform the major digestive activity of the gastrointestinal tract and produce an optimal pH for the function of these enzymes.

REGULATION OF PANCREATIC SECRETION

Secretions from the exocrine pancreas are highly regulated by neurohormonal systems in a phasic manner (cephalic, gastric, and intestinal phases). *Gastric acid* is the stimulus for the release of secretin from the duodenal mucosa (S cells), which stimulates the secretion of water and electrolytes from pancreatic ductal cells. Release of cholecystokinin (CCK) from the duodenal and proximal jejunal mucosa (Ito cells) is largely triggered by long-chain fatty acids, essential amino acids (tryptophan, phenylalanine, valine, methionine), and gastric acid itself. CCK evokes an enzyme-rich secretion from acinar cells in the pancreas. The *parasympathetic nervous system* (via the vagus nerve) exerts significant control over pancreatic secretion, particularly during the cephalic phase. Secretion evoked by secretin and CCK depends on the permissive roles of vagal afferent and efferent pathways. This is particularly true for enzyme secretion, whereas water and bicarbonate secretions are heavily dependent on the hormonal effects of secretin and to a lesser extent CCK. Also, vagal stimulation affects the release of vasoactive intestinal peptide (VIP), a secretin agonist. Pancreatic exocrine secretion is also influenced by inhibitory neuropeptides including somatostatin, pancreatic polypeptide, peptide YY, neuropeptide Y, enkephalin, pancreastatin, calcitonin gene–related peptides, glucagon, and galanin. Pancreatic polypeptide and peptide YY may act primarily on nerves outside the pancreas, while somatostatin acts at multiple sites.

■ WATER AND ELECTROLYTE SECRETION

Bicarbonate is the ion of primary physiologic importance in pancreatic secretions. The ductal cells secrete bicarbonate predominantly derived from plasma (93%) more so than from intracellular metabolism (7%). Bicarbonate enters the duct lumen through the sodium bicarbonate cotransporter with depolarization caused by chloride efflux through the cystic fibrosis transmembrane conductance regulator (CFTR). Secretin and VIP bind at the basolateral surface and cause an increase in secondary messenger intracellular cyclic AMP and act on the apical surface of the ductal cells opening the CFTR to promote secretion. CCK, acting as a neuromodulator, markedly potentiates the stimulatory effects of secretin. Acetylcholine also plays an important role in ductal cell secretion. Intraluminal bicarbonate secreted from the ductal cells helps neutralize gastric acid, increases the solubility of fatty acids and bile acids, maintains an optimal pH for pancreatic and brush border enzymes, and prevents intestinal mucosal damage.

■ ENZYME SECRETION

The acinar cell is highly compartmentalized for the production and secretion of pancreatic enzymes. Proteins synthesized by the rough endoplasmic reticulum are processed in the Golgi apparatus and then targeted to the appropriate site: zymogen granules, lysosomes, or other cell compartments. The zymogen granules migrate to the apical region of the acinar cell awaiting the appropriate neural or hormonal stimulatory response. The pancreas secretes amylolytic, lipolytic, and proteolytic enzymes into the duct lumen. *Amylolytic enzymes*, such as amylase, hydrolyze starch to oligosaccharides and the disaccharide maltose. The *lipolytic enzymes* include lipase, phospholipase A_2, and cholesterol esterase. Bile salts inhibit lipase in isolation, but colipase, another constituent of pancreatic secretion, binds to lipase and prevents this inhibition. Bile salts activate phospholipase A and cholesterol esterase. *Proteolytic enzymes* include endopeptidases (trypsin, chymotrypsin), which act on internal peptide bonds of proteins and polypeptides; exopeptidases (carboxypeptidases, aminopeptidases), which act on the free carboxyl- and amino-terminal ends of peptides, respectively; and elastase. The proteolytic enzymes are secreted as inactive zymogen precursors. Ribonucleases (deoxyribonucleases, ribonuclease) are also secreted. *Enterokinase*, an enzyme found in the duodenal mucosa ("brush border"), cleaves the lysine-isoleucine bond of trypsinogen to form trypsin. Trypsin then activates the other proteolytic zymogens and phospholipase A_2 in a cascade. The nervous system initiates pancreatic enzyme secretion. The neurologic stimulation is cholinergic, involving extrinsic innervation by the vagus nerve and subsequent innervation by intrapancreatic cholinergic nerves. The stimulatory neurotransmitters are acetylcholine and gastrin-releasing peptides. These neurotransmitters activate calcium-dependent secondary messenger systems, resulting in the release of zymogens into the pancreas duct. VIP is present in intrapancreatic nerves and potentiates the effect of acetylcholine. In contrast to other species, there are no CCK receptors on acinar cells in humans. CCK in physiologic concentrations contributes to pancreatic secretion by stimulating afferent vagal and intrapancreatic nerves.

■ AUTOPROTECTION OF THE PANCREAS

Autodigestion of the pancreas is prevented by (1) the packaging of pancreatic proteases in the precursor (proenzyme) form, (2) intracellular calcium homeostasis (low intracellular calcium in the cytosol of the acinar cell promotes the destruction of spontaneously activated trypsin), (3) acid-base balance, and (4) the synthesis of protective protease inhibitors serine protease inhibitor, Kazal type 1 (SPINK1) which can bind and inactivate ~20% of intracellular trypsin activity. Chymotrypsin C can also lyse and inactivate trypsin. These protease inhibitors are found in acinar cells, pancreatic secretions, and the α_1- and α_2-globulin fractions of plasma. Loss of any of these four protective mechanisms leads to premature enzyme activation, autodigestion, and ultimately acute pancreatitis.

■ ENTEROPANCREATIC AXIS AND FEEDBACK INHIBITION

Pancreatic enzyme secretion is controlled, at least in part, by a negative feedback mechanism induced by the presence of active serine proteases in the duodenum and nutrients in the distal small intestine. For example, perfusion of the duodenal lumen with phenylalanine (stimulates early digestion) causes a prompt increase in plasma CCK levels as well as increased secretion of chymotrypsin and other pancreatic enzymes. However, simultaneous perfusion with trypsin (stimulates late digestion) blunts both responses. Conversely, perfusion of the duodenal lumen with protease inhibitors leads to enzyme hypersecretion. Available evidence supports the concept that the duodenum contains a peptide called *CCK-releasing factor* (CCK-RF) that is involved in stimulating CCK release. It appears that serine proteases inhibit pancreatic secretion by inactivating a CCK-releasing peptide in the lumen of the small intestine. Thus, the integrative result of both bicarbonate and enzyme secretion depends on a feedback process for both bicarbonate and pancreatic enzymes. Acidification of the duodenum releases secretin, which stimulates vagal and other neural pathways to activate pancreatic duct cells, which secrete bicarbonate. This bicarbonate then neutralizes the duodenal acid, and the feedback loop is completed. Dietary proteins bind proteases, thereby leading to an increase in free CCK-RF. CCK is then released into the blood in physiologic concentrations, acting primarily through the neural pathways (vagal-vagal). This leads to acetylcholine-mediated pancreatic enzyme secretion. Proteases continue to be secreted from the pancreas until the protein within the duodenum is digested. At this point, pancreatic protease secretion is reduced to basic levels, thus completing this step in the feedback process. Additional hormonal feedback inhibition of pancreatic enzyme secretion occurs via peptide YY and glucagon-like peptide-1 following lipid or carbohydrate exposure to the ileum.

ACUTE PANCREATITIS

■ GENERAL CONSIDERATIONS

Recent U.S. estimates indicate that acute pancreatitis is one of the most common principal gastrointestinal diagnoses, accounting for almost 300,000 hospitalizations annually. The annual incidence ranges from 15–45/100,000 persons, depending on the distribution of etiologies (e.g., alcohol, gallstones, metabolic factors, drugs; **Table 359-1**) and country of study. The median length of hospital stay is 4 days, with a median hospital cost of ~$6000 and a mortality of ~1%. The estimated annual cost approaches $3 billion. Hospitalization rates increase with age and are higher among African Americans. Globally, there is a steady increase in the incidence of acute pancreatitis, most prominently in North America and Europe. Thus, the incidence of acute pancreatitis continues to rise and is associated with substantial health care costs.

■ ETIOLOGY AND PATHOGENESIS

There are many causes of acute pancreatitis (Table 359-1), and the mechanisms by which each of these conditions triggers pancreatic inflammation have not been fully elucidated. Gallstones and alcohol account for 80–90% of identified cases of acute pancreatitis in the United States. Gallstones are the leading cause of acute pancreatitis in most series (30–60%). The risk of acute pancreatitis in patients with at least one gallstone <5 mm in diameter is fourfold greater than patients with larger stones. Alcohol is the second most common cause, responsible for 15–30% of cases in the United States. The incidence of pancreatitis in individuals with heavy alcohol use is surprisingly low (5/100,000), indicating that in addition to the amount of alcohol

TABLE 359-1 Causes of Acute Pancreatitis

Common Causes
Gallstones (including microlithiasis)
Heavy alcohol use
Severe hypertriglyceridemia
Endoscopic retrograde cholangiopancreatography (ERCP), especially after therapeutic intervention
Idiopathic
Uncommon Causes
Drugs (azathioprine, 6-mercaptopurine, sulfonamides, estrogens, tetracycline, valproic acid, 5-aminosalicylic acid [5-ASA], dipeptidyl peptidase-4 [DPP4] inhibitors)
Connective tissue disorders and thrombotic thrombocytopenic purpura (TTP)
Pancreatic cancer
Hypercalcemia
Periampullary diverticulum
Pancreas divisum[a]
Hereditary pancreatitis
Cystic fibrosis
Renal failure
Infections (mumps, coxsackievirus, cytomegalovirus, echovirus, parasites)
Autoimmune (e.g., type 1 and type 2)
Trauma (especially blunt abdominal trauma)
Postoperative (abdominal and nonabdominal operations)
Causes to Consider in Patients With Recurrent Bouts of Acute Pancreatitis Without an Obvious Etiology
Occult disease of the biliary tree or pancreatic ducts, especially microlithiasis/biliary sludge
Alcohol abuse
Metabolic: Hypertriglyceridemia, hypercalcemia
Anatomic: Pancreas divisum[a]
Pancreatic cancer
Intraductal papillary mucinous neoplasm (IPMN)
Hereditary pancreatitis
Cystic fibrosis
Idiopathic

[a] Pancreas divisum is not believed to cause acute pancreatitis in isolation of another disease precipitant.

ingested, other factors affect a person's susceptibility to pancreatic injury, such as cigarette smoking and genetic predisposition. Acute pancreatitis occurs in 5–10% of patients following endoscopic retrograde cholangiopancreatography (ERCP); however, this risk can be decreased with proper patient selection and the use of a prophylactic pancreatic duct stent and/or rectal nonsteroidal anti-inflammatory drugs (NSAIDs; indomethacin). Risk factors for post-ERCP pancreatitis include minor papilla sphincterotomy, suspected functional sphincter dysfunction (previously termed sphincter of Oddi dysfunction), prior history of post-ERCP pancreatitis, age <60 years, more than two contrast injections into the pancreatic duct, and endoscopist experience.

Hypertriglyceridemia is the cause of acute pancreatitis in 1–4% of cases; serum triglyceride levels are usually >1000 mg/dL. Most patients with hypertriglyceridemic pancreatitis have undiagnosed or uncontrolled diabetes mellitus. An additional subset has an underlying derangement in lipid metabolism, probably unrelated to pancreatitis. Such patients are prone to recurrent episodes of pancreatitis and increased severity. Any factor (e.g., alcohol or medications, such as oral contraceptives) that causes an abrupt increase in serum triglycerides can potentially precipitate a bout of acute pancreatitis. Patients with a deficiency of apolipoprotein CII have an increased incidence of pancreatitis; apolipoprotein CII activates lipoprotein lipase, which is important in clearing chylomicrons from the bloodstream. Although frequently entertained, <2% of cases of acute pancreatitis are drug related. Drugs cause pancreatitis either by a hypersensitivity reaction or by the generation of a toxic metabolite, although in some cases, it is not clear which of these mechanisms is operative (Table 359-1). Increased attention has been directed toward a potential increased risk for acute pancreatitis in patients taking glucagon-like peptide-1 receptor agonists (GLP-1RA) and dipeptidyl peptidase-4 (DPP4) inhibitors. The most recent metanalysis of cardiovascular outcome trials does support an increased risk with DPP4 inhibitors, but not with GLP-1RA medications. When used for treating hyperglycemia, it is important to be conscious that >10% of individuals with diabetes will have elevated amylase or lipase levels irrespective of medication use.

Pathologically, acute pancreatitis ranges from *interstitial pancreatitis* (pancreas blood supply maintained), which is generally self-limited, to *necrotizing pancreatitis* (pancreas blood supply interrupted). Autodigestion is a currently accepted pathogenic theory resulting when proteolytic enzymes (e.g., trypsinogen, chymotrypsinogen, proelastase) and lipolytic enzymes (such as phospholipase A_2) are activated in the pancreas acinar cell compartment rather than the intestinal lumen. Several factors (e.g., endotoxins, exotoxins, viral infections, ischemia, oxidative stress, lysosomal calcium, direct trauma) are believed to facilitate premature activation of trypsin, which can activate other enzymes. Spontaneous activation of trypsin also can occur, resulting in autodigestion.

ACTIVATION OF PANCREATIC ENZYMES IN THE PATHOGENESIS OF ACUTE PANCREATITIS

Several studies have suggested that pancreatitis is a disease that evolves in three phases. The *initial phase* is characterized by intrapancreatic digestive enzyme activation and acinar cell injury. Trypsin activation appears to be mediated by lysosomal hydrolases such as cathepsin B that become colocalized with digestive enzymes in intracellular organelles; it is currently believed that acinar cell injury is the consequence of trypsin activation. The *second phase* of pancreatitis involves the activation, chemoattraction, and sequestration of leukocytes and macrophages in the pancreas, resulting in an enhanced intrapancreatic inflammatory reaction. Neutrophil depletion induced by prior administration of an antineutrophil serum has been shown to reduce the severity of experimentally induced pancreatitis. There is also evidence to support the concept that neutrophils can activate trypsinogen. Thus, intrapancreatic acinar cell activation of trypsinogen could be a two-step process (i.e., an early neutrophil-independent and a later neutrophil-dependent phase). The *third phase* of pancreatitis is due to the effects of activated proteolytic enzymes and cytokines, released by the inflamed pancreas, on distant organs. Activated proteolytic enzymes, especially trypsin, not only digest pancreatic and peripancreatic tissues but also activate other enzymes such as elastase and phospholipase A_2. The active enzymes and cytokines then digest cellular membranes and cause proteolysis, edema, interstitial hemorrhage, vascular damage, coagulation necrosis, fat necrosis, and cellular necrosis in the parenchyma. Cellular injury and death result in the liberation of bradykinin peptides, vasoactive substances, and histamine that can produce vasodilation, increased vascular permeability, and edema with profound effects on other organs. The systemic inflammatory response syndrome (SIRS) and acute respiratory distress syndrome (ARDS), as well as multiorgan failure, may occur because of this cascade producing local and distant effects.

Several genetic factors can increase the susceptibility and/or modify the severity of pancreatic injury in acute pancreatitis, recurrent acute pancreatitis, and chronic pancreatitis. The major genetic susceptibility factors center on the control of trypsin activity within the pancreatic acinar cell, in part because they were identified as candidate genes linked to intrapancreatic trypsin control. Six genetic variants have been identified as being associated with susceptibility to pancreatitis. The genes that have been identified include (1) cationic trypsinogen gene (*PRSS1*), (2) *SPINK1*, (3) the cystic fibrosis transmembrane conductance regulator gene (*CFTR*), (4) the chymotrypsin C gene (*CTRC*), (5) the calcium-sensing receptor (*CASR*), and (6) claudin-2 (*CLDN2*). Among these variants, only *PRSS1* mutations are sufficient to precipitate acute pancreatitis in the absence of other risk factors, whereas the other variants are disease modifiers. Investigations for other genetic variants are in progress, so it is expected this list will be expanded in the future.

APPROACH TO THE PATIENT

Abdominal Pain

Abdominal pain is the major symptom of acute pancreatitis. Pain may vary from mild discomfort to severe, constant, and incapacitating distress. Characteristically, the pain, which is steady and boring in character, is located in the epigastrium region and may radiate to the back, chest, flanks, and lower abdomen. Nausea, vomiting, and abdominal distention due to gastric and intestinal hypomotility are also frequent complaints.

Physical examination frequently reveals a distressed and anxious patient. Low-grade fever, tachycardia, and hypotension are common. Shock is not unusual and may result from (1) hypovolemia secondary to exudation of blood and plasma proteins into the retroperitoneal space; (2) increased formation and release of kinin peptides, which cause vasodilation and increased vascular permeability; and (3) systemic effects of proteolytic and lipolytic enzymes released into the circulation. Jaundice occurs infrequently; when present, it may be a consequence of extrinsic compression due to peripancreatic edema, a pancreatic head mass, or intraductal obstruction from a common bile duct stone or sludge. Erythematous skin nodules ("pancreatic panniculitis") due to subcutaneous fat necrosis are possible, but rarely occur. There are pulmonary findings in 10–20% of patients, including basilar rales, atelectasis, and pleural effusion, the latter most frequently left-sided. Abdominal tenderness and muscle rigidity are present to a variable degree, but compared with the intense pain, these signs may be less impressive. Bowel sounds are usually diminished or absent. An enlarged pancreas from an acute fluid collection, walled-off necrosis, or a pseudocyst may be palpable in the upper abdomen later in the course of the disease (i.e., 4–6 weeks). A faint blue discoloration around the umbilicus (Cullen's sign) may occur as the result of hemoperitoneum, and a blue-red-purple or green-brown discoloration of the flanks (Turner's sign) reflects tissue breakdown of hemoglobin from severe necrotizing pancreatitis with hemorrhage; both findings are rare but reflect an increased clinical severity.

■ LABORATORY DATA

Serum amylase and lipase values greater than three times the upper limit of normal are strongly supportive of the diagnosis if alternate etiologies, including gut perforation, ischemia, and infarction, are excluded. However, it should be noted that there is no correlation between the severity of pancreatitis and the degree of serum lipase and amylase elevations or serial trends. After 3–7 days, even with continuing evidence of pancreatitis, total serum amylase values tend to return to normal. However, pancreatic lipase levels may remain elevated for 7–14 days. It should be recognized that amylase elevations in serum and urine occur in many conditions other than pancreatitis **(see Chap. 358, Table 358-2)**. Importantly, patients with *acidemia* (arterial pH ≤7.32) may have spurious elevations in serum amylase. This finding explains why patients with diabetic ketoacidosis may have marked elevations in serum amylase without any other evidence of acute pancreatitis. On the other hand, serum amylase levels can be spuriously low in severe hypertriglyceridemia. Serum lipase activity increases in parallel with amylase activity and is more specific than amylase, making it the preferred test. A serum lipase measurement can be instrumental in differentiating a pancreatic or nonpancreatic cause for hyperamylasemia.

Leukocytosis (15,000–20,000 leukocytes/μL) occurs frequently. Patients with more severe disease may show hemoconcentration with hematocrit values >44% and prerenal azotemia (elevated blood urea nitrogen [BUN]) resulting from loss of plasma into the retroperitoneal space and peritoneal cavity. In fact, one study reported that a combination of admission BUN and a change in BUN at 48 h was a strong predictor of mortality (area under the curve [AUC] 0.91).

Hemoconcentration may be the harbinger of more severe disease, whereas azotemia is a significant risk factor for mortality. *Hyperglycemia* is common and is due to multiple factors, including decreased insulin release, increased glucagon release, and increased output of adrenal glucocorticoids and catecholamines. *Hypocalcemia* occurs in ~25% of patients, and its pathogenesis is incompletely understood. Although earlier studies suggested that the response of the parathyroid gland to a decrease in serum calcium is impaired, subsequent observations have failed to confirm this phenomenon. Intraperitoneal saponification of calcium by fatty acids in areas of fat necrosis occurs occasionally, with large amounts (up to 6.0 g) dissolved or suspended in ascitic fluid. Such "soap formation" may also be significant in patients with pancreatitis, mild hypocalcemia, and little or no obvious ascites. *Hyperbilirubinemia* (serum bilirubin >4.0 mg/dL) occurs in ~10% of patients. However, jaundice is usually transient, and serum bilirubin levels return to normal in 4–7 days. Serum alkaline phosphatase and transaminase levels may also be transiently elevated and parallel serum bilirubin values. Elevations of alanine aminotransferase (ALT) >3× the upper limit of normal are strongly associated with a gallstone etiology in patients with acute pancreatitis. Approximately 5–10% of patients have *hypoxemia* (arterial PO_2 ≤60 mmHg), which may herald the onset of ARDS. Finally, the electrocardiogram is occasionally abnormal in acute pancreatitis with ST-segment and T-wave abnormalities simulating myocardial ischemia.

An abdominal ultrasound is recommended as the initial diagnostic imaging modality and is most useful for evaluating gallstones and common bile duct dilation.

The Revised Atlanta Criteria have clearly outlined the morphologic features of acute pancreatitis on computed tomography (CT) scan as follows: (1) interstitial pancreatitis, (2) necrotizing pancreatitis, (3) acute pancreatic fluid collection, (4) pancreatic pseudocyst, (5) acute necrotic collection (ANC), and (6) walled-off necrosis (WON) **(Table 359-2** and **Fig. 359-1)**. Radiologic studies useful in the diagnosis of acute pancreatitis are discussed in **Chap. 358** and listed in **Table 358-1**.

■ DIAGNOSIS

Any severe acute pain in the abdomen or back should raise the possibility of acute pancreatitis. The diagnosis is established by two of the following three criteria: (1) typical abdominal pain in the epigastrium that may radiate to the back, (2) threefold or greater elevation in serum lipase and/or amylase, and (3) confirmatory findings of acute pancreatitis on cross-sectional abdominal imaging. Although not required for diagnosis, markers of severity may include hemoconcentration (hematocrit >44%), admission azotemia, SIRS, and signs of organ failure **(Table 359-3)**.

The *differential diagnosis* should include the following disorders: (1) perforated viscus, especially peptic ulcer; (2) acute cholecystitis and biliary colic; (3) acute intestinal obstruction; (4) mesenteric vascular occlusion; (5) renal colic; (6) inferior myocardial infarction; (7) dissecting aortic aneurysm; (8) connective tissue disorders with vasculitis; (9) pneumonia; and (10) diabetic ketoacidosis. It may be difficult to differentiate acute cholecystitis from acute pancreatitis, because an elevated serum amylase may be found in both disorders. Pain of biliary tract origin is more right sided or epigastric than periumbilical or left upper quadrant and can be more severe; ileus is usually absent. Ultrasound is helpful in establishing the diagnosis of cholelithiasis and cholecystitis. Intestinal obstruction due to mechanical factors can be differentiated from pancreatitis by the history of crescendo-decrescendo pain, findings on abdominal examination, and CT of the abdomen showing changes characteristic of anatomic obstruction. Acute mesenteric vascular occlusion is usually suspected in elderly debilitated patients with leukocytosis, abdominal distention, and bloody diarrhea, confirmed by CT or magnetic resonance angiography. Vasculitides secondary to systemic lupus erythematosus and polyarteritis nodosa may be confused with pancreatitis, especially because pancreatitis may develop as a complication of these diseases. Diabetic ketoacidosis is often accompanied

TABLE 359-2 Revised Atlanta Definitions of Morphologic Features of Acute Pancreatitis

	DEFINITION	COMPUTED TOMOGRAPHY FEATURES
Types of Acute Pancreatitis		
Interstitial pancreatitis	Acute inflammation of the pancreatic parenchyma and peripancreatic tissues, but without recognizable tissue necrosis	Pancreatic parenchyma enhancement by IV contrast agent and without peripancreatic necrosis
Necrotizing pancreatitis	Inflammation associated with pancreatic parenchymal and/or peripancreatic necrosis	Lack of pancreatic parenchymal enhancement by IV contrast agent and/or presence of findings of peripancreatic necrosis (see below—ANC and WON)
Morphologic Features		
Acute pancreatic fluid collection	Peripancreatic fluid associated with interstitial edematous pancreatitis with no associated peripancreatic necrosis. This term applies only to areas of peripancreatic fluid seen within the first 4 weeks after onset of interstitial edematous pancreatitis and without the features of a pseudocyst.	Occurs in the setting of interstitial pancreatitis Homogeneous collection with fluid density Confined by normal peripancreatic fascial planes No definable wall encapsulating the collection Adjacent to pancreas (no intrapancreatic extension)
Pancreatic pseudocyst	An encapsulated collection of fluid with a well-defined wall usually outside the pancreas with minimal or no necrosis. This usually occurs >4 weeks after onset of interstitial edematous pancreatitis.	Well circumscribed, usually round or oval Homogeneous fluid density No solid component Well-defined wall; that is, completely encapsulated Maturation usually requires >4 weeks after onset of acute interstitial pancreatitis
Acute necrotic collection (ANC)	A collection containing variable amounts of both fluid and necrosis associated with necrotizing pancreatitis; the necrosis can involve the pancreatic parenchyma and/or the peripancreatic tissues.	Occurs in the setting of acute necrotizing pancreatitis Heterogeneous and nonliquid density of varying degrees in different locations (some appear homogeneous early in their course) No definable wall encapsulating the collection Location—intrapancreatic and/or extrapancreatic
Walled-off necrosis (WON)	A mature, encapsulated collection of pancreatic and/or peripancreatic necrosis that has developed a well-defined inflammatory wall. WON usually occurs >4 weeks after onset of acute necrotizing pancreatitis.	Heterogeneous with liquid and nonliquid density with varying degrees of loculations (some may appear homogeneous) Well-defined wall; that is, completely encapsulated Location—intrapancreatic and/or extrapancreatic Maturation usually requires >4 weeks after onset of acute necrotizing pancreatitis

Source: Data from P Banks et al: Gut 62:102, 2013.

by abdominal pain and elevated total serum amylase levels, thus closely mimicking acute pancreatitis; however, the serum lipase level is often not elevated in diabetic ketoacidosis, and pancreas imaging is normal.

CLINICAL COURSE, DEFINITIONS, AND CLASSIFICATIONS

The Revised Atlanta Criteria define (1) phases of acute pancreatitis, (2) severity of acute pancreatitis, and (3) radiographic definitions, as outlined below.

Phases of Acute Pancreatitis Two clinical phases of acute pancreatitis have been defined, early (<2 weeks) and late (>2 weeks), which primarily describe the hospital course of the disease. In the *early phase* of acute pancreatitis, which lasts 1–2 weeks, severity is defined by clinical parameters rather than morphologic findings. Most patients exhibit SIRS, and if this persists, patients are predisposed to organ failure. Three organ systems should be assessed to define organ failure: respiratory, cardiovascular, and renal. Organ failure is defined as a score of 2 or more for one of these three organ systems using the modified Marshall scoring system. Persistent organ failure (>48 h) is the most important clinical finding regarding the severity of the acute pancreatitis episode. CT imaging is usually not needed or recommended during the first 48 h of admission in acute pancreatitis.

The *late phase* is characterized by a protracted course of illness and may require imaging to evaluate for local complications. The critical clinical parameter of severity, as in the early phase, is persistent organ failure. These patients may require supportive measures such as renal dialysis, ventilator support, or need for supplemental nutrition. The radiographic feature of greatest importance to recognize in this phase is the development of necrotizing pancreatitis on CT imaging. Necrosis is associated with prolonged hospitalization and, if infected, may require intervention (percutaneous, endoscopic, or, rarely, surgical).

Severity of Acute Pancreatitis Three classes of severity have been defined: mild, moderately severe, and severe. *Mild acute pancreatitis* is without local complications or organ failure. Most patients with interstitial acute pancreatitis have mild pancreatitis. In mild acute pancreatitis, the disease is self-limited, usually within 3–7 days after onset. Oral intake can be resumed if the patient is hungry, has normal bowel function, and is without nausea and vomiting. Typically, a clear or full liquid diet has been recommended for the initial meal; however, a low-fat solid diet is advised early in mild acute pancreatitis.

Moderately severe acute pancreatitis is characterized by transient organ failure (i.e., it resolves in <48 h) or local or systemic complications in the absence of persistent organ failure. These patients may or may not have necrosis but may develop a local complication such as a fluid collection, which often requires a prolonged hospitalization >1 week. As with mild acute pancreatitis, the mortality rate for these patients remains low.

Severe acute pancreatitis is characterized by persistent organ failure (>48 h), involving one or more organs. A CT scan or magnetic resonance imaging (MRI) should be obtained to assess for necrosis and/or complications. If a local complication is encountered, management is dictated by clinical symptoms, evidence of infection, the maturity of fluid collection, and clinical stability of the patient. Prophylactic antibiotics are no longer recommended for severe acute pancreatitis.

Imaging in Acute Pancreatitis Two types of pancreatitis are recognized on imaging as *interstitial* or *necrotizing* based on pancreatic perfusion. CT imaging with IV contrast is best evaluated 3–5 days into hospitalization if patients are not responding to supportive care to assess for local complications such as necrosis. Recent studies report the overutilization of CT imaging within 72 h for acute pancreatitis, including those with a mild severity of disease. The Revised Atlanta Criteria outline the terminology for local complications and fluid collections along with a CT imaging template to guide reporting of findings. Local morphologic features are summarized in Table 359-2. *Interstitial pancreatitis* occurs in 90–95% of admissions for acute pancreatitis and is characterized by diffuse gland enlargement, homogenous contrast enhancement, and mild inflammatory changes or peripancreatic

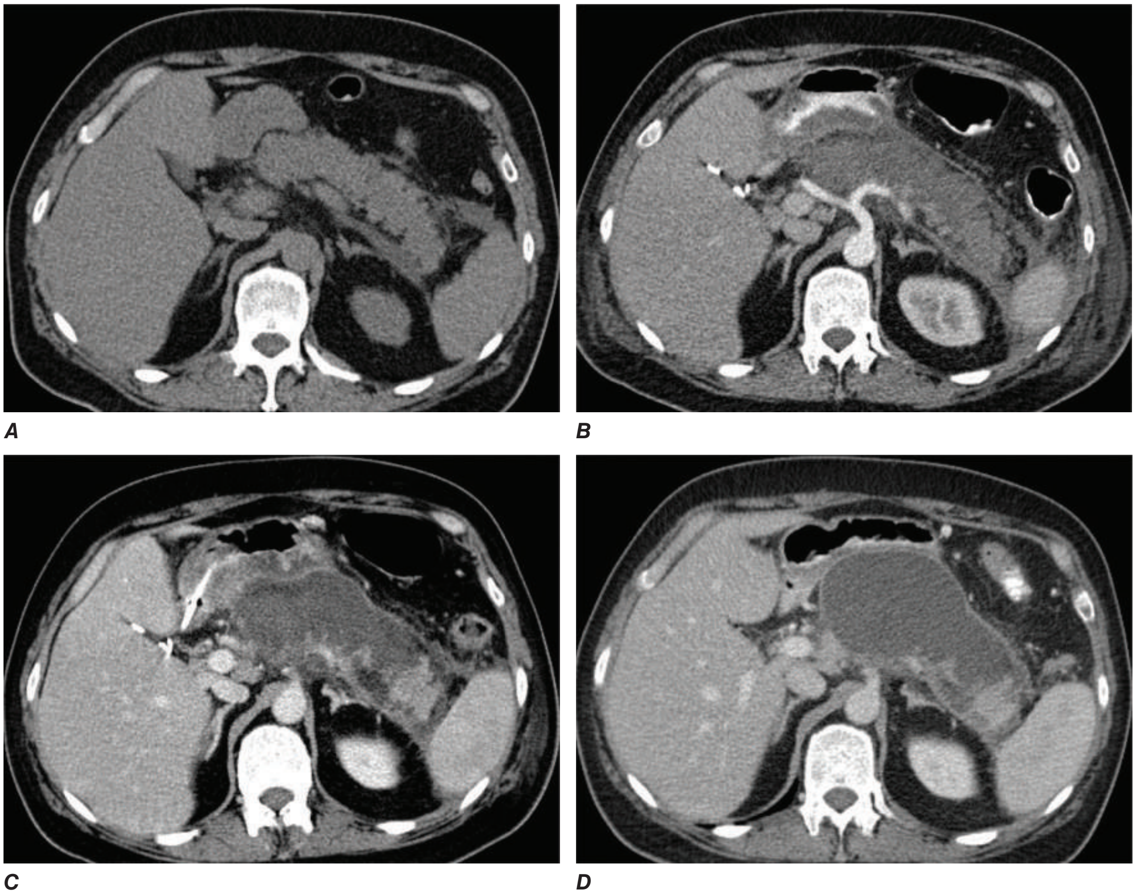

FIGURE 359-1 Evolution of changes of acute necrotizing pancreatitis on computed tomography (CT). ***A.*** CT scan of the abdomen without IV contrast performed on admission for a patient with acute gallstone pancreatitis, showing mild peripancreatic stranding. ***B.*** Contrast-enhanced CT scan of the abdomen performed on the same patient 1 week after admission shows extensive intrapancreatic necrosis, evidenced by the lack of contrast enhancement in the pancreatic body with minimal enhancement noted at the distal most aspect of the pancreatic tail. ***C.*** Contrast-enhanced CT scan of the abdomen performed on the same patient 2 weeks after admission demonstrates a semiorganized, heterogeneous fluid collection, referred to as an acute necrotic collection. On this image, a small area of viable pancreatic parenchyma is seen at the tail of the pancreas. ***D.*** Contrast-enhanced CT scan of the abdomen performed on the same patient 5 weeks after admission demonstrates a well-encapsulated fluid collection, essentially replacing the pancreas, referred to as walled-off necrosis.

stranding. Symptoms generally resolve with a week of hospitalization. *Necrotizing pancreatitis* occurs in 5–10% of acute pancreatitis admissions and may not evolve until after several days of hospitalization. It is characterized by lack of pancreatic parenchymal enhancement by intravenous contrast agent and/or presence of peripancreatic necrosis. The natural history of pancreatic and peripancreatic necrosis is variable because it may remain solid or liquefy, remain sterile or become infected, and persist or resolve over time. Importantly, those with only extrapancreatic necrosis have a more favorable prognosis than patients with pancreatic necrosis (with or without extrapancreatic necrosis). CT identification of local complications, particularly necrosis, is critical in patients who are not responding to therapy because patients with necrosis are at the greatest risk of mortality (Figs. 359-1 and 359-2). The median prevalence of organ failure is >50% in necrotizing pancreatitis and is slightly higher in infected versus sterile necrosis. With single-organ system failure, the mortality is 3–10%, but increases to nearly 50% with multiorgan failure.

ACUTE PANCREATITIS MANAGEMENT

The management of patients with acute pancreatitis from the time of diagnosis in the emergency ward to hospital discharge is briefly reviewed, highlighting salient features based on severity and complications. It is important to recognize that 85–90% of cases of acute pancreatitis are self-limited and subside spontaneously, usually within 3–7 days after onset, and do not exhibit organ failure or local complications.

The management of acute pancreatitis begins in the emergency ward. After a diagnosis has been confirmed, early and aggressive fluid resuscitation is critical. Additionally, intravenous analgesics are administered, severity is assessed, and a search for etiologies that may impact acute care is begun. Patients who do not respond to aggressive fluid resuscitation in the emergency ward should be considered for admission to a step-down or intensive care unit for further intensive monitoring and management, including hemodynamic monitoring and management of organ failure, if present.

Fluid Resuscitation and Monitoring Response to Therapy The most important treatment intervention for acute pancreatitis is early, intravenous fluid resuscitation to prevent systemic complications from the secondary systemic inflammatory response. The patient is initially made NPO to minimize nutrient-induced stimulation of the pancreas and is given intravenous narcotic analgesics to control abdominal pain and supplemental oxygen (as needed).

Lactated Ringer's solution has been shown to decrease systemic inflammation (lower C-reactive protein levels from admission) and may be a better crystalloid than normal saline. Aggressive intravenous fluids have historically been promoted including an initial bolus of approximately 15–20 mL/kg (1050–1400 mL), followed by 2–3 mL/kg per h (200–250 mL/h), to maintain urine output >0.5 mL/kg per hour. However, a recent randomized controlled trial demonstrated that this hydration strategy is associated with an increased risk of fluid overload (based on signs or symptoms) compared to a less aggressive strategy (10 mL/kg bolus, followed by 1.5 mL/kg per h) without an improvement in clinical outcomes. Irrespective of the initial fluid management strategy, serial bedside evaluations are required every 6–8 h to assess vital signs, oxygen saturation, and change in physical examination to optimize fluid resuscitation. A *targeted resuscitation strategy* with measurement of hematocrit and BUN every 8–12 h is recommended to ensure the adequacy of fluid resuscitation and monitor response

TABLE 359-3 Severity Assessment of Acute Pancreatitis

Risk Factors for Increased Severity

- Age >60 years
- Obesity, BMI >30 kg/m^2
- Comorbid disease (based on Charlson comorbidity index)

Markers of Severity at Admission or Within 24 h

- SIRS—defined by presence of 2 or more criteria:
  - Core temperature <36° or >38°C
  - Heart rate >90 beats/min
  - Respirations >20/min or PCO_2 <32 mmHg
  - White blood cell count >12,000/μL, <4000/μL, or 10% bands
- APACHE II (≥8 at 24 h)
- Hemoconcentration (hematocrit >44%)
- Admission BUN (>20 mg/dL)
- BISAP score (≥3 present)
  - (B) BUN >25 mg/dL
  - (I) Impaired mental status
  - (S) SIRS: ≥2 of 4 present
  - (A) Age >60 years
  - (P) Pleural effusion
- Organ failure (Modified Marshall score) (≥1 present):
  - Cardiovascular: systolic BP <90 mmHg, heart rate >130 beats/min
  - Pulmonary: PaO_2 <60 mmHg
  - Renal: serum creatinine >2.0 mg/dL

Markers of Severity during Hospitalization

- Elevated C-reactive protein (CRP) (variable thresholds >100 mg/L reported)
- Persistent organ failure (≥48 h)
- Pancreatic or extrapancreatic necrosis

Abbreviations: APACHE II, Acute Physiology and Chronic Health Evaluation II; BISAP, Bedside Index of Severity in Acute Pancreatitis; BMI, body mass index; BP, blood pressure; BUN, blood urea nitrogen; SIRS, systemic inflammatory response syndrome.

to therapy, noting that a less aggressive resuscitation strategy may be sufficient in milder forms of pancreatitis. Importantly, a rising BUN during hospitalization is associated not only with inadequate hydration but also higher in-hospital mortality.

A decrease in hematocrit and BUN during the first 12–24 h is strong evidence that an adequate volume of fluids is being administered. Adjustments in fluid resuscitation may be required in patients with cardiac, pulmonary, or renal disease.

Assessment of Severity and Hospital Triage Severity of acute pancreatitis should be determined in the emergency ward to assist in patient triage to a regular hospital ward, step-down unit, or intensive care unit. The Bedside Index of Severity in Acute Pancreatitis (BISAP) incorporates five clinical and laboratory parameters obtained within the first 24 h of hospitalization (Table 359-3)—BUN >25 mg/dL, impaired mental status (Glasgow coma scale score <15), SIRS, age >60 years, and pleural effusion on radiography—that can be useful in assessing severity. The presence of three or more of these factors was associated with substantially increased risk for in-hospital mortality among patients with acute pancreatitis. In addition, an elevated hematocrit and admission BUN are also associated with more severe acute pancreatitis. Incorporating these indices with the overall patient response to initial fluid resuscitation in the emergency ward can be useful in triaging patients to the appropriate hospital acute care setting.

In general, patients with lower BISAP scores, hematocrit, and admission BUN levels tend to respond to initial management and can be safely triaged to a regular hospital ward for ongoing care. If SIRS is not present at 24 h, the patient is unlikely to develop organ failure or necrosis. Conversely, patients with persistent SIRS at 24 h or underlying comorbid illnesses (e.g., chronic obstructive pulmonary disease, congestive heart failure) should be considered for a step-down unit setting if available. Patients with higher BISAP scores and elevations in hematocrit and admission BUN who do not respond to initial fluid resuscitation and exhibit evidence of respiratory failure, hypotension, or organ failure should be considered for direct admission to an intensive care unit.

Special Considerations Based on Etiology A careful history, review of medications, selected laboratory studies (liver profile, serum triglycerides, serum calcium), and abdominal ultrasound are recommended in the emergency ward to assess for etiologies that may impact acute management. An abdominal ultrasound is the initial imaging modality of choice and will evaluate the gallbladder, common bile duct, and pancreatic head.

GALLSTONE PANCREATITIS Patients with evidence of ascending cholangitis (including sepsis with evidence of biliary obstruction on laboratory or imaging studies) should undergo ERCP within 24–48 h of admission. Patients with gallstone pancreatitis are at increased risk of recurrence, and consideration should be given to performing a cholecystectomy during the same admission in mild acute pancreatitis. An alternative for patients who are not surgical candidates would be to perform an endoscopic biliary sphincterotomy before discharge.

HYPERTRIGLYCERIDEMIA Serum triglycerides >1000 mg/dL are associated with acute pancreatitis. Initial therapy should focus on the treatment of hyperglycemia with intravenous insulin, which often corrects hypertriglyceridemia. In contrast to the standard nutritional management of acute pancreatitis, patients should remain fasting for the initial 24–36 h to promote the resolution of severe triglyceride elevations.

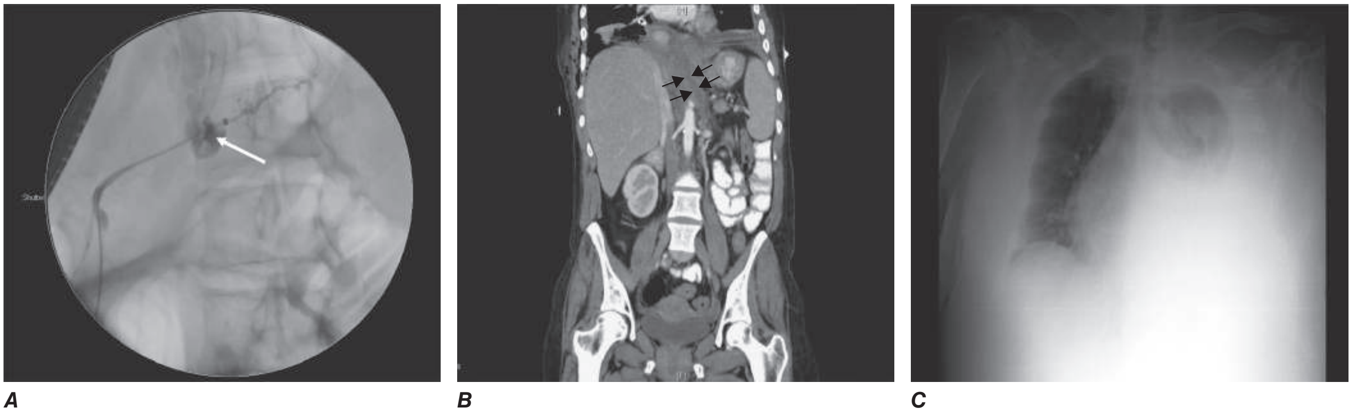

FIGURE 359-2 Imaging features of a pancreaticopleural fistula secondary to acute pancreatitis. ***A.*** Pancreaticopleural fistula: pancreatic duct leak on endoscopic retrograde cholangiopancreatography. Pancreatic duct leak (*arrow*) demonstrated at the time of retrograde pancreatogram in a patient with exacerbation of alcohol-induced acute pancreatitis. ***B.*** Pancreaticopleural fistula: computed tomography (CT) scan. Contrast-enhanced CT scan (coronal view) with *arrows* showing fistula tract from pancreatic duct disruption in the pancreatic pleural fistula. ***C.*** Pancreaticopleural fistula: chest x-ray. Large pleural effusion in the left hemithorax from a disrupted pancreatic duct. Analysis of pleural fluid revealed elevated amylase concentration. *(Courtesy of Dr. K.J. Mortele, Brigham and Women's Hospital; with permission.)*

Adjunct therapies may also include heparin or plasmapheresis, but there is no compelling evidence these measures improve clinical outcomes. Outpatient therapies include control of diabetes if present, administration of lipid-lowering agents, weight loss, and avoidance of drugs that elevate lipid levels.

Other potential etiologies that may impact acute hospital care include *hypercalcemia* and *post-ERCP pancreatitis*. Treatment of hyperparathyroidism or malignancy is effective at reducing serum calcium. Pancreatic duct stenting and/or rectal indomethacin administration are effective at decreasing the risk of pancreatitis after ERCP. Drugs that cause pancreatitis should be discontinued. Multiple drugs have been implicated, but only about 30 have been rechallenged (Class 1A) and found to be causative.

Nutritional Therapy A low-fat solid diet can be administered to subjects with mild acute pancreatitis once they are able to eat. Enteral nutrition should be considered 2–3 days after admission in subjects with more severe pancreatitis instead of total parenteral nutrition (TPN). Enteral feeding maintains gut barrier integrity, limits bacterial translocation, is less expensive, and has fewer complications than TPN. While there may be physiologic advantages of jejunal feeding, gastric feeding is safe, and the benefits of jejunal over gastric feeding remain under investigation.

Management of Local Complications (Table 359-4) Patients exhibiting signs of clinical deterioration despite aggressive fluid resuscitation and hemodynamic monitoring should be assessed for local complications, including necrosis, pseudocyst formation, pancreas duct disruption, peripancreatic vascular complications, and extrapancreatic infections. A multidisciplinary team approach is recommended, including gastroenterology, surgery, interventional radiology, and intensive care specialists; consideration should be made for transfer to a tertiary pancreas center of excellence if these services are not available.

NECROSIS The management of necrosis requires a multidisciplinary team approach. Percutaneous fine-needle aspiration of necrosis with Gram stain and culture was previously performed to evaluate for infected pancreatic necrosis in those with sustained leukocytosis, fever, or organ failure. However, the current use of this technique varies depending on institutional preference, with many abandoning this diagnostic test to avoid potentially contaminating an otherwise sterile collection, particularly when culture results will not lead to a clinical decision to de-escalate antimicrobial therapy. Even though there is currently no role for *prophylactic antibiotics* in necrotizing pancreatitis, *empiric antibiotics* should be considered in those with clinical decompensation. Prophylactic antibiotics do not lead to improved survival and may promote the development of opportunistic fungal infections. Repeated CT or MRI imaging should also be considered with any change in clinical course to monitor for complications (e.g., thromboses, hemorrhage, abdominal compartment syndrome).

In general, *sterile necrosis* is most often managed conservatively unless complications arise. Once a diagnosis of *infected necrosis* is established and an organism identified, targeted antibiotics should be instituted. Pancreatic drainage and/or debridement (necrosectomy) should be considered for definitive management of *infected necrosis*, but management decisions are ultimately influenced by the clinical response since almost two-thirds of patients respond to antibiotic treatment with or without drainage. Symptomatic local complications as outlined in the Revised Atlanta Criteria typically require definitive therapy.

A step-up approach (percutaneous or endoscopic transgastric/transduodenal drainage followed, if necessary, by endoscopic or, rarely, surgical necrosectomy) is the current treatment paradigm. Randomized trials have reported advantages to an initial endoscopic approach compared to an initial surgical necrosectomy approach in patients requiring intervention for symptomatic WON. Taken together, a more conservative approach to the management of infected pancreatic necrosis has evolved under the close supervision of a multidisciplinary team. If conservative therapy can be safely implemented, it is recommended to do so for 4–6 weeks to allow the pancreatic collections to either resolve or evolve to develop a more organized boundary (i.e., to "wall off") since interventions are generally safer and more technically straightforward.

TABLE 359-4 Complications of Acute Pancreatitis

Local
Pancreatic/peripancreatic fluid collections (Table 359-2):
Acute necrotic collection (sterile or infected)
Walled-off necrosis (sterile or infected)
Pancreatic pseudocyst
Disruption of main pancreatic duct or secondary branches
Pancreatic ascites
Chylous ascites (secondary to disruption of lymphatic ducts)
Involvement of contiguous organs by necrotizing pancreatitis (e.g., colon perforation)
Splanchnic thromboses (splenic vein, superior mesenteric vein, and/or portal vein)
Gastric outlet obstruction
Biliary obstruction (jaundice)
Systemic
Pulmonary
Pleural effusion
Atelectasis
Mediastinal fluid
Pneumonitis
Acute respiratory distress syndrome
Cardiovascular
Hypotension
Hypovolemia
Nonspecific ST-T changes in electrocardiogram simulating myocardial infarction
Pericardial effusion
Hematologic
Disseminated intravascular coagulation
Gastrointestinal hemorrhage
Peptic ulcer disease
Erosive gastritis
Hemorrhagic pancreatic necrosis with erosion into major blood vessels
Variceal hemorrhage secondary to splanchnic thrombosis
Renal
Oliguria (<300 mL/d)
Azotemia
Renal artery and/or renal vein thrombosis
Acute tubular necrosis
Metabolic
Hyperglycemia
Hypertriglyceridemia
Hypocalcemia
Encephalopathy
Sudden blindness (Purtscher's retinopathy)
Central nervous system
Psychosis
Fat emboli
Fat necrosis
Subcutaneous tissues (erythematous nodules)
Bone
Miscellaneous (mediastinum, pleura, nervous system)

PSEUDOCYST The incidence of pseudocyst is low, and most acute collections resolve over time. Less than 10% of patients have persistent fluid collections after 4 weeks that would meet the definition of a pseudocyst. Only symptomatic collections require intervention with endoscopic or surgical drainage.

PANCREATIC DUCT DISRUPTION Pancreatic duct disruption may present with symptoms of increasing abdominal pain or shortness of breath in the setting of an enlarging fluid collection resulting in pancreatic ascites (ascitic fluid has high amylase level). Diagnosis can be confirmed on magnetic resonance cholangiopancreatography (MRCP) or ERCP. Placement of a bridging pancreatic stent for at least 6 weeks is >90% effective at resolving the leak with or without parenteral nutrition and octreotide. Nonbridging stents are less effective (25–50%) but may be useful in combination with parenteral nutrition and octreotide prior to surgical evaluation.

PERIVASCULAR COMPLICATIONS Perivascular complications may include *splenic vein thrombosis* with gastric varices and pseudoaneurysms, as well as *portal and superior mesenteric vein thromboses. Gastric varices* rarely bleed but can be life-threatening. Similarly, life-threatening bleeding from a ruptured *pseudoaneurysm* can be diagnosed and treated with mesenteric angiography and embolization.

EXTRAPANCREATIC INFECTIONS Hospital-acquired infections occur in up to 20% of patients with acute pancreatitis. Patients should be continually monitored for the development of pneumonia, urinary tract infection, and line infection. Continued culturing of urine, monitoring of chest x-rays, and routine changing of intravenous lines are important during hospitalization.

Follow-Up Care Hospitalizations for moderately severe and severe acute pancreatitis can be prolonged and last weeks to months and often involve periods of intensive care unit admission and outpatient rehabilitation or subacute nursing care. Follow-up evaluation should assess for development of diabetes, exocrine pancreatic insufficiency, recurrent cholangitis, or infected fluid collections. As mentioned previously, cholecystectomy should be performed during the initial hospitalization for acute gallstone pancreatitis with mild clinical severity. For patients with necrotizing gallstone pancreatitis, the timing of cholecystectomy needs to be individualized.

RECURRENT ACUTE PANCREATITIS

Approximately 25% of patients who have had an attack of acute pancreatitis will experience a recurrence. The two most common etiologic factors are alcohol and cholelithiasis. In patients with recurrent pancreatitis without an obvious cause, the differential diagnosis should consider occult biliary tract disease, including microlithiasis, hypertriglyceridemia, pancreatic cancer, and hereditary pancreatitis (Table 359-1). In one small series, up to two-thirds of patients with recurrent acute pancreatitis without an initially apparent cause had occult gallstone disease attributed to microlithiasis. Genetic defects as in hereditary pancreatitis and cystic fibrosis mutations can result in recurrent pancreatitis. Other diseases of the biliary tree and pancreatic ducts that can cause acute pancreatitis include choledochocele; ampullary tumors; pancreas divisum; and pancreatic duct stones, stricture, and tumors. Approximately 2–4% of patients with pancreatic cancer present with acute pancreatitis.

PANCREATITIS IN PATIENTS WITH AIDS

The incidence of acute pancreatitis is theoretically increased in patients with AIDS for two reasons: (1) the high incidence of infections involving the pancreas such as infections with cytomegalovirus, *Cryptosporidium*, and the *Mycobacterium avium* complex; and (2) the frequent use of medications such as pentamidine, trimethoprim-sulfamethoxazole, and protease inhibitors. The incidence has been markedly reduced due to advances in therapy, including the disuse of didanosine **(Chap. 208)**.

CHRONIC PANCREATITIS AND EXOCRINE PANCREATIC INSUFFICIENCY

PATHOPHYSIOLOGY

Chronic pancreatitis is a disease process characterized by irreversible damage to the pancreas, in contrast to the reversible changes noted in acute pancreatitis (Table 359-4). The events that initiate and then perpetuate the inflammatory process in the pancreas are becoming more clearly understood. Irrespective of the mechanism of injury, it is apparent that stellate cell activation leads to cytokine expression and production of extracellular matrix proteins that contribute to acute and chronic inflammation and collagen deposition in the pancreas. This condition is defined by the presence of histologic abnormalities, including chronic inflammation, fibrosis, and progressive destruction (atrophy) of both exocrine and endocrine tissue. A number of etiologies have been associated with chronic pancreatitis resulting in the cardinal manifestations of the disease such as abdominal pain, steatorrhea, weight loss, diabetes mellitus, and, less commonly, pancreatic cancer **(Table 359-5)**.

Even in individuals in whom alcohol is believed to be the primary cause of chronic pancreatitis, other factors are likely required for the development and progression of disease, which explains why not all heavy consumers of alcohol develop pancreatic disease. Importantly, there is a strong association between smoking and chronic pancreatitis. Cigarette smoke leads to an increased susceptibility to pancreatic autodigestion and predisposes to dysregulation of duct cell CFTR function. Smoking is an independent, dose-dependent risk factor for chronic pancreatitis and recurrent acute pancreatitis. Both continued alcohol and smoking exposure are associated with disease progression, including pancreatic fibrosis and calcifications.

Characterization of pancreatic stellate cells (PSCs) has added insight into the underlying cellular responses behind development of chronic pancreatitis. Specifically, PSCs are believed to play a role in maintaining normal pancreatic architecture that shifts toward fibrogenesis in those who develop chronic pancreatitis. It is believed that alcohol or additional stimuli lead to matrix metalloproteinase–mediated destruction of normal collagen in pancreatic parenchyma, which later allows for pancreatic

TABLE 359-5 Classification of Chronic Pancreatitis: The TIGAR-O System

Toxic-metabolic
Alcoholic
Tobacco smoking
Hypercalcemia
Hyperlipidemia (hypertriglyceridemia)
Chronic renal failure
Idiopathic
Early onset
Late onset
Tropical
Genetic
Cationic trypsinogen (*PRSS1*)
Cystic fibrosis transmembrane conductance regulator gene (*CFTR*)[a]
Calcium-sensing receptor (*CASR*)[a]
Chymotrypsin C gene (*CTRC*)[a]
Pancreatic secretory trypsin inhibitor gene (*SPINK1*)[a]
Autoimmune
Type 1 autoimmune pancreatitis (associated with IgG4-related disease)
Type 2 autoimmune pancreatitis (idiopathic duct-centric chronic pancreatitis)
Recurrent and severe acute pancreatitis
Postnecrotic (severe acute pancreatitis)
Recurrent acute pancreatitis
Vascular diseases/ischemia
Radiation induced
Obstructive
Pancreas divisum[a]
Duct obstruction (e.g., tumor)
Preampullary duodenal wall cysts
Posttraumatic pancreatic duct strictures

[a]These conditions are believed to be disease modifiers that require additional factors to cause chronic pancreatitis.

Abbreviations: TIGAR-O, toxic-metabolic, idiopathic, genetic, autoimmune, recurrent and severe acute pancreatitis, obstructive.

remodeling. Proinflammatory cytokines, tumor necrosis factor α (TNF-α), interleukin 1 (IL-1), and interleukin 6 (IL-6), as well as oxidant complexes, can induce PSC activity with subsequent new collagen synthesis. In addition to being stimulated by cytokines, oxidants, or growth factors, PSCs also possess transforming growth factor β (TGF-β)–mediated self-activating autocrine pathways that may explain disease progression in chronic pancreatitis even after removal of noxious stimuli.

■ ETIOLOGIC CONSIDERATIONS

Among adults in the United States, heavy alcohol use is the most common cause of clinically apparent chronic pancreatitis. As many as 25% of adults in the United States with chronic pancreatitis have the *idiopathic* form, including a subset of patients who do not develop clinical manifestations until later in life (referred to as idiopathic late-onset chronic pancreatitis). Recent investigations have indicated that up to 15% of patients with chronic pancreatitis previously classified as having idiopathic pancreatitis may have an underlying genetic predisposition (Table 359-5).

The prototypical genetic defect was identified in the cationic trypsinogen gene (*PRSS1*) by studying several large families with chronic pancreatitis. Additional pathogenic and nonpathogenic mutations have been identified in this gene. The defect prevents the destruction of prematurely activated trypsin and allows it to be resistant to the intracellular protective effect of trypsin inhibition. It is hypothesized that this continual activation of digestive enzymes within the gland leads to acute injury and, finally, chronic pancreatitis. Since the initial discovery of the *PRSS1* mutation defect, other genetic disease modifiers have been identified (Table 359-5).

The *CFTR* gene functions as a cyclic AMP–regulated chloride channel. In patients with cystic fibrosis, the high concentration of macromolecules can block the pancreatic ducts. It must be appreciated, however, that there is a great deal of heterogeneity in relationship to the *CFTR* gene defect. More than 1700 putative mutations of the *CFTR* gene have been identified. The large number and different classes of *CFTR* mutations have hampered attempts to elucidate the relationship between the genotype and pancreatic manifestations. The ability to detect *CFTR* mutations has led to the recognition that the clinical spectrum of the disease is broader than initially thought. Studies have clarified the association between mutations of the *CFTR* gene and another monosymptomatic form of cystic fibrosis (i.e., chronic pancreatitis). It is estimated that in patients with idiopathic pancreatitis, the frequency of a single *CFTR* mutation is 11 times the expected frequency and the frequency of two mutant alleles is 80 times the expected frequency. In these studies, patients were adults when the diagnosis of pancreatitis was made; none had any clinical evidence of pulmonary disease, and sweat test results were not diagnostic of cystic fibrosis. The prevalence of such mutations is unclear, and further studies are needed. In addition, the therapeutic and prognostic implication of these findings with respect to managing pancreatitis remains to be determined. *CFTR* mutations are common in the general population, so it is unclear whether the *CFTR* mutation alone can lead to pancreatitis as an autosomal recessive disease. A study evaluated 39 patients with idiopathic chronic pancreatitis to assess the risk associated with these mutations. Patients with two *CFTR* mutations (compound heterozygotes) demonstrated *CFTR* function at a level between that seen in typical cystic fibrosis and cystic fibrosis carriers and had a 40-fold increased risk of pancreatitis. The presence of a separate genetic mutation (*N34S SPINK1*) increased the risk 20-fold. A combination of two *CFTR* mutations and an *N34S SPINK1* mutation increased the risk of pancreatitis 900-fold. Knowledge of the genetic defects and downstream alterations in protein expression has led to the development of novel therapies in children with cystic fibrosis that potentiate the CFTR channel, resulting in improvement in lung function, quality of life, and weight gain. Some studies have shown that use of CFTR modulators may reduce the frequency of acute pancreatitis in heterozygous carriers. Table 359-5 lists other recognized causes of chronic pancreatitis.

■ AUTOIMMUNE PANCREATITIS (TABLE 359-6)

Autoimmune pancreatitis (AIP) refers to a form of chronic pancreatitis with distinct histopathology and several unique differences in the clinical phenotype. Currently, two subtypes of AIP are recognized, type 1 AIP and idiopathic duct-centric chronic pancreatitis (IDCP, also referred to as type 2 AIP). Type 1 AIP is identified as the pancreatic manifestation of a multiorgan syndrome currently referred to as IgG4-related disease **(Chap. 380)**. The characteristic histopathologic findings of type 1 AIP include lymphoplasmacytic infiltrate, storiform fibrosis, and abundant IgG4 cells. IDCP is histologically defined by the presence of granulocytic infiltration of the duct wall (termed a granulocytic epithelial lesion [GEL]) but without IgG4-positive cells. Type 1 AIP is often associated with involvement of other organs in the setting of IgG4-related disease, including bilateral submandibular gland enlargement, characteristic renal lesions, retroperitoneal fibrosis, and stricturing of the extrapancreatic biliary tree. In contrast, IDCP is a pancreas-specific disorder that is associated with inflammatory bowel disease in ~10% of patients. AIP is not a common cause of idiopathic recurrent acute pancreatitis.

TABLE 359-6 Comparison of the Autoimmune Pancreatitis (AIP) Subtypes

	TYPE 1 AIP	TYPE 2 AIP
Age at diagnosis, mean	Seventh decade	Fifth decade
Male sex	75%	50%
Serum IgG4 elevation	~66%	~25%
Other organ involvement	50%	No[a]
Histologic findings:		
Lymphoplasmacytic Infiltration	++	++
Periductal inflammation	++	++
Storiform fibrosis	++	+
Obliterative phlebitis	++	+
Granulocytic epithelial lesion (GEL)	–	+++
IgG4 tissue staining	Abundant (≥10 cells/hpf)	Scant (<10 cells/hpf)
Response to steroids	~100%	~100%
Risk for relapse	Moderate to high (20–60%)	Low (<10%)
Associated with IgG4-RD	Yes	No

[a] Inflammatory bowel disease is seen in ~10–20% of patients with idiopathic duct-centric chronic pancreatitis but may also occur in type 1 AIP.

Abbreviations: hpf, high-power field; IgG4-RD, IgG4-related disease.

Source: Reproduced with permission from PA Hart: Reviews in basic and clinical gastroenterology and hepatology. Gastroenterology 149:39, 2015.

Jaundice, weight loss, and new-onset diabetes are the most common presenting symptoms. Elevated serum IgG4 levels are supportive of the diagnosis (elevated in two-thirds of patients with type 1 AIP) but have a low positive predictive value when used in isolation of other clinical findings. CT imaging demonstrates abnormalities in the majority of patients with either diffuse or focal enlargement during active disease, unless the gland is atrophic due to previous disease **(Fig. 359-3)**. The presence of an inflammatory rim, termed a capsule sign, is highly specific (but not sensitive) for AIP. ERCP or MRCP reveals strictures in the bile duct in more than one-third of patients with AIP, including some patients with isolated intrahepatic bile duct strictures (type 1 AIP only), which can mimic primary sclerosing cholangitis, and is referred to as IgG4-related sclerosing cholangitis (previously termed IgG4-associated cholangitis).

The Mayo Clinic HISORt criteria provide a helpful mnemonic to remember the key diagnostic features of this disease, including (1) *h*istology; (2) *i*maging; (3) *s*erology (elevated serum IgG4 levels); (4) *o*ther organ involvement; and (5) *r*esponse to glucocorticoid *t*herapy. These diagnostic criteria have been harmonized with those from other countries to develop the International Consensus Diagnostic Criteria for AIP, which are the most comprehensive criteria. Glucocorticoids have shown efficacy in alleviating symptoms, decreasing the size of

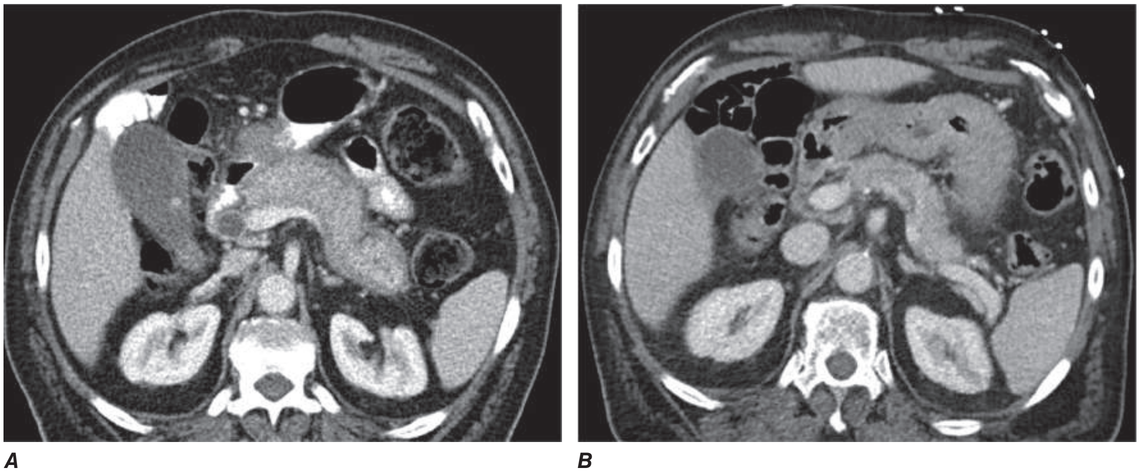

FIGURE 359-3 Imaging features of the pancreatic parenchyma in a patient with type 1 autoimmune pancreatitis on computed tomography (CT). ***A.*** Contrast-enhanced CT scan of the abdomen demonstrates diffuse pancreatic enlargement and a hypoechoic rim (capsule sign) in a patient who presented with jaundice. The serum IgG4 level was elevated to 942 mg/dL (reference range 4–86 mg/dL), so the patient was diagnosed with definitive type 1 autoimmune pancreatitis. ***B.*** Contrast-enhanced CT scan of the abdomen following a treatment course with high-dose steroids demonstrates return to normal size of the pancreas, reappearance of normal lobulations along the margin, and absence of the hypoechoic rim.

the pancreas, and reversing histopathologic features in patients with AIP. Patients typically respond dramatically to glucocorticoid therapy within a 2- to 4-week period. Prednisone is usually administered at an initial dose of 40 mg/d for 4 weeks followed by a taper of the daily dosage by 5 mg per week based on monitoring of clinical parameters. Relief of symptoms, liver biochemistries, and abnormal imaging of the pancreas and bile ducts are followed to assess for treatment response. A poor response to glucocorticoids should raise suspicion of an alternate diagnosis, such as pancreatic cancer. A prior multicenter international study examined >1000 patients with AIP. Clinical remission was achieved in 99% of type 1 AIP and 92% of type 2 AIP patients with steroids. However, disease relapse occurred in 31 and 9% of patients with type 1 and type 2 AIP, respectively. Patients with multiple relapses have been managed with immunomodulators (e.g., azathioprine, 6-mercaptopurine, or mycophenolate mofetil) with variable success. Management with B-cell depletion therapy (e.g., rituximab) is likely the most effective treatment option for patients with recurrent relapses with other therapies under investigation. The appearance of interval cancers following a diagnosis of AIP is uncommon.

Clinical Features of Chronic Pancreatitis Patients with chronic pancreatitis primarily seek medical attention due to abdominal pain or symptoms of maldigestion. The abdominal pain may be variable in location, severity, and frequency. The pain can be constant or intermittent with pain-free intervals. Eating may exacerbate the pain, leading to a fear of eating with consequent weight loss. The spectrum of abdominal pain ranges from mild to quite severe, with narcotic dependence as a frequent consequence. There is often a disparity between the reported severity of abdominal pain and the physical findings, which primarily consist of nonfocal abdominal tenderness. Patients with chronic abdominal pain may or may not experience symptoms of maldigestion, such as chronic diarrhea, steatorrhea, and/or weight loss. Fat-soluble vitamin deficiencies are increasingly recognized. Importantly, there is an exceedingly high prevalence of metabolic bone disease in chronic pancreatitis, with ~65% of patients having either osteopenia or osteoporosis. Patients with chronic pancreatitis have impaired quality of life and develop significant morbidity, requiring frequent health encounters.

The diagnosis of early or mild chronic pancreatitis can be challenging because there is no accurate biomarker for the disease. In contrast to acute pancreatitis, the serum amylase and lipase levels are usually not strikingly elevated in chronic pancreatitis. Rather, low serum pancreatic enzyme levels are moderately specific for a diagnosis of chronic pancreatitis but have poor sensitivity. Elevation of serum bilirubin and alkaline phosphatase may indicate cholestasis secondary to common bile duct stricture caused by chronic inflammation or fibrosis. The cumulative prevalence of exocrine pancreatic insufficiency is >80%. The presence of overt steatorrhea in a patient with chronic pancreatitis is highly suggestive of this complication. However, in those with milder symptoms, additional testing, such as a random fecal elastase-1 level (on a formed stool specimen) may be needed to confirm the diagnosis of exocrine pancreatic insufficiency. The radiographic evaluation of a patient with suspected chronic pancreatitis usually proceeds from a noninvasive to more invasive approach. Abdominal CT imaging **(Fig. 359-4)** is the initial modality of choice, followed by MRI, endoscopic ultrasound, and pancreas function testing. In addition to excluding a pseudocyst and pancreatic cancer, CT imaging may show calcifications, dilated pancreatic or biliary ducts, or an atrophic pancreas. Although abdominal CT scanning and MRI greatly aid in the diagnosis of pancreatic disease, the diagnostic test with the best sensitivity is the direct pancreatic function test using secretin. The secretin test becomes abnormal when ≥60% of the pancreatic exocrine function has been lost and usually correlates well with the onset of chronic abdominal pain. The role of endoscopic ultrasonography (EUS) in diagnosing early chronic pancreatitis is still evolving. A total of nine endosonographic features have been described in chronic pancreatitis. The presence of five or more features is considered diagnostic of chronic pancreatitis. EUS is not a specific enough test for detecting early chronic pancreatitis alone **(Chap. 358)** and may show positive features in patients with diabetes, patients with a history of cigarette smoking, or even in normal aging individuals. Recent data suggest that EUS can be combined with endoscopic pancreatic function testing (EUS-ePFT) during a single endoscopy to screen for chronic pancreatitis in patients with chronic abdominal pain. Diffuse calcifications noted on plain film of the abdomen usually indicate significant damage to the pancreas and are pathognomonic for chronic pancreatitis. While patterns of calcification are not specific for an etiology, patients with tropical pancreatitis (often associated with a *SPINK1* mutation) characteristically have bulky calcifications.

Complications of Chronic Pancreatitis There are a number of disease-related complications from chronic pancreatitis in addition to the aforementioned abdominal pain and exocrine pancreatic insufficiency **(Table 359-7)**. The lifetime prevalence of chronic pancreatitis–related diabetes exceeds 80%. Although most patients develop hyperglycemia due to insulin deficiency caused by loss of islet cells, diabetic ketoacidosis and diabetic coma are uncommon. Likewise, end-organ damage (retinopathy, neuropathy, nephropathy) is also uncommon. Nondiabetic retinopathy may be due to vitamin A and/or zinc deficiency. Osteoporosis and osteopenia are increasingly recognized in chronic pancreatitis and likely related to a combination of shared risk factors (e.g., alcohol use, cigarette smoking), vitamin D deficiency, and detrimental effects on the bone from chronic inflammation. Gastrointestinal bleeding may occur from peptic ulceration, gastritis, a pseudocyst

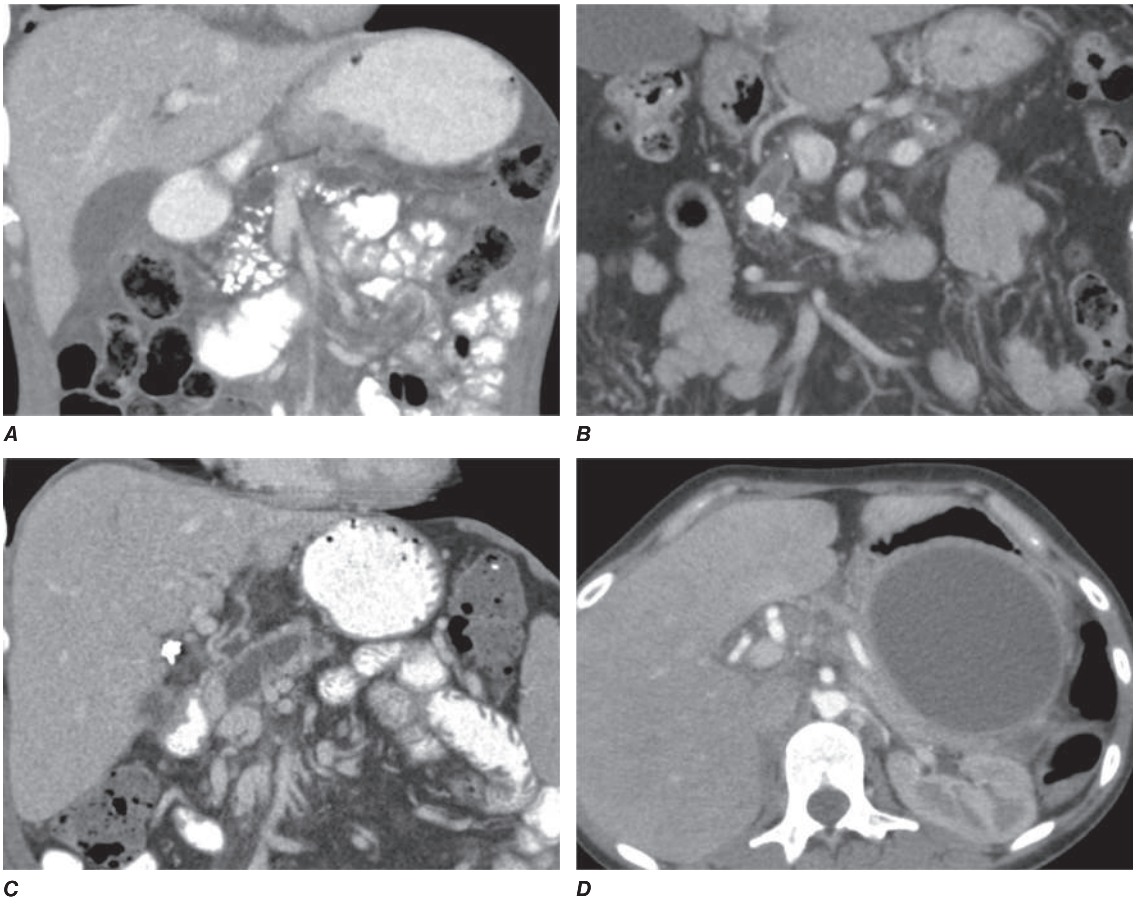

FIGURE 359-4 Distribution of imaging features of chronic pancreatitis on computed tomography (CT). Distinct features of chronic pancreatitis are seen on selected images from contrast-enhanced CT scans of the abdomen from four unique patients, including the following. ***A.*** Numerous punctate calcifications involving the pancreatic parenchyma in the head and body. ***B.*** A moderate-sized calculus visualized in the pancreatic duct with associated ductal dilation. ***C.*** Significant pancreatic duct dilation and adjacent parenchymal atrophy secondary to a pancreatic duct stricture (which is not well seen on this scan). ***D.*** A large unilocular, encapsulated cyst in the tail of the pancreas consistent with a pseudocyst from prior pancreatitis. Note adjacent pancreatic parenchymal atrophy.

eroding into the duodenum, arterial bleeding into the pancreatic duct (hemosuccus pancreaticus), or ruptured varices secondary to splenic vein thrombosis. Jaundice, cholestasis, and biliary cirrhosis may occur from the chronic inflammatory reaction around the intrapancreatic portion of the common bile duct. Twenty years after the diagnosis of chronic calcific pancreatitis, the cumulative risk of pancreatic cancer is 4%. Patients with hereditary *PRSS1* or tropical pancreatitis have an increased risk for pancreatic cancer compared to other forms of chronic pancreatitis.

TREATMENT

Chronic Pancreatitis

There are currently no therapies to reverse or delay the disease progression of chronic pancreatitis, so management is primarily focused on screening for and management of disease-related complications.

STEATORRHEA

The treatment of steatorrhea with pancreatic enzyme replacement therapy is conceptually straightforward, yet complete correction of steatorrhea is uncommon. Enzyme therapy usually brings diarrhea under control and restores absorption of fat to an acceptable level and affects weight gain. Thus, pancreatic enzyme replacement is the cornerstone of therapy. In treating steatorrhea, it is important to use a potent pancreatic formulation that will deliver sufficient lipase into the duodenum to correct maldigestion and decrease steatorrhea. For adult patients with exocrine pancreatic insufficiency, it is generally recommended to start at a dosage of 25,000–50,000 units of lipase taken during each meal; however, the dose may need to be increased up to 100,000 units of lipase depending on the response in symptoms, nutritional parameters, and/or pancreas function test results (although rarely used for this purpose). Additionally, some may require acid suppression with proton pump inhibitors to optimize the response to pancreatic enzymes. Monitoring nutritional parameters such as fat-soluble vitamins, zinc levels, body weight, and periodic bone mineral density measurement should be considered.

TABLE 359-7 Complications of Chronic Pancreatitis

Chronic abdominal pain	Biliary stricture and/or biliary cirrhosis
Exocrine pancreatic insufficiency	Pancreatic duct stricture
Diabetes mellitus	Pseudocyst
Splanchnic venous thrombosis	Pancreatic cancer
Metabolic bone disease (osteoporosis)	Malnutrition, micronutrient deficiencies

ABDOMINAL PAIN

The management of pain in patients with chronic pancreatitis is challenging due to the complex mechanisms of pancreatitis-related pain. Recent meta-analyses have shown no consistent benefit of enzyme therapy at reducing pain in chronic pancreatitis. Pain relief experienced by patients treated with pancreatic enzymes may be due to improvements in dyspepsia from maldigestion. One short-term randomized trial showed that pregabalin could decrease pain in chronic pancreatitis and lower pain medication requirement. Other studies using antioxidants have yielded mixed results.

Endoscopic treatment of chronic pancreatitis pain may involve sphincterotomy, pancreatic duct stenting, stone extraction, and drainage of a pancreatic pseudocyst. Therapy directed to the pancreatic duct would seem to be most appropriate in the setting of a dominant stricture, especially if there is an obstructing intraductal stone. The use of endoscopic stenting for patients with chronic pain has not been investigated in controlled trials. It is now appreciated that significant complications can occur from stenting (e.g., stent migration, stent occlusion, and stent-induced pancreatic duct strictures). Recent guidelines recommend considering celiac plexus block for treatment of pain in chronic pancreatitis, but recommendations were conditional with very low quality of evidence. Celiac plexus block has not been rigorously studied for chronic pancreatitis and does not provide durable pain relief. It can provide relief in some selected patients, but the a priori identification of those who will respond is difficult. In patients with pancreatic duct dilation, ductal decompression with *surgical therapy* has been the therapy of choice. Among such patients, 80% seem to obtain immediate relief; however, at the end of 3 years, one-half of the patients have recurrence of pain. Two randomized prospective trials comparing endoscopic to surgical therapy for chronic pancreatitis demonstrated that surgical therapy was superior to endoscopy at decreasing pain and improving quality of life in selected patients with dilated ducts and abdominal pain. This would suggest that chronic pancreatitis patients with dilated ducts and pain should be considered for surgical intervention. A recent randomized controlled trial suggests that early surgical intervention may provide superior pain relief in the short term compared to a conservative, nonoperative approach. The role of preoperative stenting prior to surgery as a predictor of response has yet to be proven.

Total pancreatectomy with or without autologous islet cell transplantation has been used in highly selected patients with chronic pancreatitis and abdominal pain refractory to conventional therapy. However, some patients will continue to have pain postoperatively, illustrating the complex nature of pain in this patient population. Patients who benefit most from total pancreatectomy have a shorter duration of symptoms and lower pain medication requirements. The role of this procedure remains to be fully defined but may be an option in lieu of ductal decompression surgery or partial pancreatic resection in patients with intractable, painful, small-duct disease or hereditary pancreatitis.

HEREDITARY PANCREATITIS

Hereditary pancreatitis (*PRSS1*) is a rare form of pancreatitis with early age of onset that is typically associated with familial aggregation of cases. A genome-wide search using genetic linkage analysis identified the hereditary pancreatitis gene on chromosome 7. Mutations in ion codons 29 (exon 2) and 122 (exon 3) of the cationic trypsinogen gene (*PRSS1*) cause an autosomal dominant form of pancreatitis. The codon 122 mutations lead to a substitution of the corresponding arginine with another amino acid, usually histidine. This substitution, when it occurs, eliminates a fail-safe trypsin self-destruction site necessary to eliminate trypsin that is prematurely activated within the acinar cell. These patients have recurring episodes of acute pancreatitis. Patients frequently develop pancreatic calcification, diabetes mellitus, and steatorrhea; in addition, they have an increased incidence of pancreatic cancer with a cumulative incidence of ~10%. A previous natural history study of hereditary pancreatitis in >200 patients from France reported that abdominal pain started in childhood at age 10 years, steatorrhea developed at age 29 years, diabetes at age 38 years, and pancreatic cancer at age 55 years. Abdominal complaints in relatives of patients with hereditary pancreatitis should raise the question of pancreatic disease.

PANCREATIC ENDOCRINE TUMORS

Pancreatic endocrine tumors are discussed in Chap. 89.

OTHER CONDITIONS

ANNULAR PANCREAS

When the ventral pancreatic anlage fails to migrate correctly to make contact with the dorsal anlage, the result may be a ring of pancreatic tissue encircling the duodenum. Such an annular pancreas may cause intestinal obstruction in the neonate or adult. Symptoms of postprandial fullness, epigastric pain, nausea, and vomiting may be present for years before the diagnosis is entertained. The radiographic findings are symmetric dilation of the proximal duodenum with bulging of the recesses on either side of the annular band, effacement but not destruction of the duodenal mucosa, accentuation of the findings in the right anterior oblique position, and lack of change on repeated examinations. The differential diagnosis should include duodenal webs, tumors of the pancreas or duodenum, duodenal ulcer, regional enteritis, and adhesions. Patients with annular pancreas have an increased incidence of pancreatitis and peptic ulcer. Because of these and other potential complications, the treatment for refractory symptoms is surgical. Retrocolic duodenojejunostomy is the procedure of choice, although some surgeons advocate Billroth II gastrectomy, gastroenterostomy, and vagotomy.

PANCREAS DIVISUM

Pancreas divisum is present in 7–10% of the population and occurs when the embryologic ventral and dorsal pancreatic anlagen fail to fuse, so that pancreatic drainage is accomplished mainly through the accessory minor papilla. Pancreas divisum is the most common congenital anatomic variant of the human pancreas. Current evidence indicates that this anomaly does not predispose to the development of pancreatitis in the majority of patients. However, the combination of pancreas divisum and a small accessory orifice can result in dorsal duct obstruction. The challenge is to identify this subset of patients with dorsal duct pathology. Cannulation of the dorsal duct by ERCP is technically challenging and associated with a very high risk of post-ERCP pancreatitis, so patients with pancreatitis and pancreas divisum should likely be treated with conservative measures. In many of these patients, pancreatitis is idiopathic and unrelated to the pancreas divisum. Endoscopic or surgical intervention is indicated if pancreatitis recurs and no other cause can be found. It should be stressed that the ERCP/MRCP appearance of pancreas divisum (i.e., a small-caliber ventral duct with an arborizing pattern) may be mistaken as representing an obstructed main pancreatic duct secondary to a mass lesion.

MACROAMYLASEMIA

In macroamylasemia, amylase circulates in the blood in a polymer form too large to be easily excreted by the kidney. Patients with this condition demonstrate an elevated serum amylase value and a low urinary amylase value. The presence of macroamylase can be documented by chromatography of the serum. The prevalence of macroamylasemia is 1.5% of the nonalcoholic general adult hospital population. Usually, macroamylasemia is an incidental finding and is not related to disease of the pancreas or other organs. Macrolipasemia has now been documented in patients with cirrhosis or non-Hodgkin's lymphoma. In these patients, the pancreas appeared normal on ultrasound and CT examination. Lipase was shown to be complexed with immunoglobulin A. Thus, the possibility of *both* macroamylasemia and macrolipasemia should be considered in patients with elevated blood levels of these enzymes.

ACKNOWLEDGMENT

This chapter represents a revised version of chapters by Drs. Norton J. Greenberger (deceased), Phillip P. Toskes (deceased), Peter A. Banks, and Bechien Wu that were in previous editions of Harrison's.

FURTHER READING

Crockett SD et al: American Gastroenterological Association Institute guideline on initial management of acute pancreatitis. Gastroenterology 154:1096, 2018.

De-Maderia E et al: Aggressive or moderate fluid resuscitation in acute pancreatitis. N Engl J Med 387:989, 2022.

Forsmark CE et al: Acute pancreatitis. N Engl J Med 375:1972, 2016.

Gardner TB et al: ACG clinical guideline: Chronic pancreatitis. Am J Gastroenterol 115:322, 2020.

Hart PA et al: Recent advances in autoimmune pancreatitis. Gastroenterology 149:39, 2015.

Petrov MS, Yadav D: Global epidemiology and holistic prevention of pancreatitis. Nat Rev Gastroenterol Hepatol 16:175, 2019.

Vege SS, Chari ST: Chronic pancreatitis. N Engl J Med 386:869, 2022.

Immune-Mediated, Inflammatory, and Rheumatologic Disorders PART 11

Section 1 The Immune System in Health and Disease

360

Introduction to the Immune System

Barton F. Haynes, Kelly A. Cuttle, Anthony S. Fauci

DEFINITIONS

- *Adaptive immune system*—recently evolved system of immune responses mediated by T and B lymphocytes. Immune responses by these cells are based on specific antigen recognition by clonotypic receptors that are products of genes that rearrange during development and throughout the life of the organism. Additional cells of the adaptive immune system include various types of antigen-presenting cells (APCs).
- *Antibody*—B cell–produced molecules encoded by genes that rearrange during B-cell development consisting of immunoglobulin heavy and light chains that together form the central component of the B-cell receptor (BCR) for antigen. Antibody can exist as B cell–surface antigen-recognition molecules or as secreted molecules in plasma and other body fluids.
- *Antigens*—foreign or self-molecules that are recognized by the adaptive and innate immune systems resulting in immune cell triggering, T-cell activation, and/or B-cell antibody production.
- *Antigen-presenting cells (APCs)*—a group of immune cells that process antigens to mediate both adaptive immune responses and the maintenance of peripheral tolerance. Classical APCs include dendritic cells, macrophages, and B cells.
- *Apoptosis*—the process of *programmed cell death* whereby signaling through "death receptors" on the surface of cells (e.g., tumor necrosis factor [TNF] receptors, CD95) leads to signaling cascades that involve activation of the caspase family of molecules and leads to DNA cleavage and cell death. Apoptosis, which does not lead to induction of inordinate inflammation, is to be contrasted with *cell necrosis*, which does lead to induction of inflammatory responses.
- *Autoimmune diseases*—diseases such as systemic lupus erythematosus and rheumatoid arthritis in which cells of the adaptive immune system such as autoreactive T and B cells become overreactive and produce pathogenic T-cell and antibody responses.
- *Autoinflammatory diseases*—hereditary disorders such as hereditary periodic fevers (HPFs) characterized by recurrent episodes of severe inflammation and fever due to mutations in controls of the innate inflammatory response, i.e., the inflammasome (see below and Table 360-5). Patients with HPFs also have rashes and serosal and joint inflammation, and some can have neurologic symptoms. Autoinflammatory diseases are different from autoimmune diseases in that evidence for activation of adaptive immune cells such as autoreactive B cells is not present.
- *Autophagy*—lysosomal degradation pathway mechanism of cells to dispose of intracellular debris and damaged organelles. Autophagy by cells of the innate immune system is used to control intracellular infectious agents such as mycobacteria, in part by initiation of phagosome maturation and enhancing major histocompatibility complex (MHC) class II antigen presentation to CD4 T cells.
- *B-cell receptor (BCR) for antigen*—complex of surface molecules that rearrange during postnatal B-cell development, made up of surface immunoglobulin (Ig) and associated Ig αβ chain molecules that recognize nominal antigen via Ig heavy- and light-chain variable regions, and signal the B cell to terminally differentiate to make antigen-specific antibody.
- *B lymphocytes*—bone marrow–derived lymphocytes that express surface immunoglobulin (the BCR for antigen) and secrete specific antibody after interaction with antigen.
- *B regulatory cells*—a population of suppressive B cells that aid in the inhibition of inflammation through the release of cytokines such as interleukin-(IL) 10.
- *CD classification of human lymphocyte differentiation antigens*—the development of monoclonal antibody technology led to the discovery of a large number of new leukocyte surface molecules. From a series of International Workshop on Leukocyte Differentiation Antigens has come the *cluster of differentiation* (CD) classification of leukocyte antigens.
- *CD4 T cell*—T lymphocyte subset that participates in adaptive immunity and helps B cells make antibody.
- *CD8 T cell*—cytotoxic T lymphocyte subset that kills tumor cells and cells infected with pathogens.
- *Chemokines*—soluble molecules that direct and determine immune cell movement and circulation pathways.
- *Complement*—cascading series of plasma enzymes and effector proteins that function to lyse pathogens and/or target them to be phagocytized by neutrophils and monocyte/macrophage lineage cells of the reticuloendothelial system.
- *Co-stimulatory molecules*—molecules of APCs (such as B7-1, B7-2, or CD40) that lead to T-cell activation when bound by ligands on activated T cells (such as CD28 or CD40 ligand).
- *Crystallopathies*—nanoparticle- or microparticle-sized deposits of crystals, misfolded proteins, or airborne particulate matter that can stimulate the inflammasome and initiate inflammation and tissue damage.
- *Cytokines*—soluble proteins that interact with specific cellular receptors that are involved in the regulation of the growth and activation of immune cells and mediate normal or pathologic inflammatory and immune responses.
- *Dendritic cells*—myeloid and/or lymphoid lineage APCs of the adaptive immune system. Immature dendritic cells (DCs), or DC precursors, are key components of the innate immune system by responding to infections with production of high levels of cytokines. DCs are key initiators of innate immune responses via cytokine production and mediators of adaptive immune responses via presentation of antigen to T lymphocytes.
- *Ig fragment crystallizable (Fc) receptors (Rs)*—receptors found on the surface of certain cells including B cells, natural killer (NK) cells, macrophages, neutrophils, and mast cells. Fc receptors bind to the Fc domains of antibodies that have attached to invading pathogen-infected cells. FcRs stimulate cytotoxic cells to destroy microbe-infected cells through antibody-dependent cell-mediated cytotoxicity (ADCC). Examples of important FcRs include CD16 (FcγRIIIa), CD23 (FcεR), CD32 (FcγRII), CD64 (FcγRI), and CD89 (FcαR).
- *Inflammasome*—large cytoplasmic complexes of intracellular proteins that link the sensing of microbial products and cellular stress to the proteolytic activation of IL-1β and IL-18 inflammatory cytokines. Activation of molecules in the inflammasome is a key step in the response of the innate immune system for intracellular recognition of microbial and other danger signals in both health and pathologic states.
- *Innate immune system*—ancient immune recognition system of host cells bearing germline-encoded pattern recognition receptors (PRRs) that recognize pathogens and trigger a variety of mechanisms of pathogen elimination. Cells of the innate immune system include NK cell lymphocytes, monocytes/macrophages, DCs, neutrophils, basophils, eosinophils, tissue mast cells, and epithelial cells.
- *Innate lymphoid cells (ILCs)*—lymphocytes that do not express the type of diversified antigen receptors on T cell and B cells. ILC1s,

ILC2s, and ILC3s are tissue resident cells and functionally may be analogous to CD4 T_H1, T_H2, and T_H17 cells, respectively.

- *Natural killer (NK) cells*—a type of ILC that kills target cells expressing few or no human leukocyte antigen (HLA) class I molecules, such as malignantly transformed cells and virally infected cells. NK cells express receptors that inhibit killer cell function when self-MHC class I is present. Innate NK cells mirror the cytolytic functions of CD8 cytotoxic T cells of the adaptive immune system.
- *NK T cells*—innate-like lymphocytes that use an invariant T-cell receptor (TCR)-α chain combined with a limited set of TCR-β chains and coexpress receptors commonly found on NK cells. NK T cells recognize lipid antigens of bacterial, viral, fungal, and protozoal infectious agents.
- *Pathogen-associated molecular patterns (PAMPs)*—invariant molecular structures expressed by large groups of microorganisms that are recognized by host cellular PRRs in the mediation of innate immunity.
- *Pattern recognition receptors (PRR)*—germline-encoded receptors expressed by cells of the innate immune system that recognize PAMPs.
- *T lymphocytes*—thymus-derived lymphocytes that mediate adaptive cellular immune responses including T helper, T regulatory, and cytotoxic T lymphocyte effector cell functions.
- *T-cell exhaustion*—state of T cells when the persistence of antigen disrupts memory T-cell function, resulting in defects in memory T-cell responses. Most frequently occurs in malignancies and in chronic viral infections such as HIV-1 and hepatitis C.
- *TCR for antigen*—complex of surface molecules that rearrange during postnatal T-cell development made up of clonotypic TCR-α and -β chains that are associated with the CD3 complex composed of invariant γ, δ, ε, ζ, and η chains. TCR-α and -β chains recognize peptide fragments of protein antigen physically bound in APC MHC class I or II molecules, leading to signaling via the CD3 complex to mediate effector functions.
- *T follicular helper T cells (TFH)*—CD4 T cells regulated by bcl-6 in B-cell follicle germinal centers that produce IL-4 and IL-21 and drive B-cell differentiation and affinity maturation in peripheral lymphoid tissues such as lymph node and spleen.
- *T_H1 T cells*—CD4 helper T-cell subset regulated by transcription factor T-bet that produces interferon (IFN)-γ, IL-2, and TNF-β and participates in cell-mediated immunity.
- *T_H2 T cells*—CD4 helper T-cell subset regulated by transcription factors STAT6 and GATA3 that produces IL-4, IL-5, IL-6, IL-9, IL-10, and IL-13 and regulates antibody and eosinophil responses.
- *T regulatory cells (Treg)*—CD4 or CD8 T cells regulated by the transcription factor forkhead box P3 (FOXP3) that play roles in modulating immune responses to prevent harmful immune activation. Treg cells that prevent autoimmunity can arise in the thymus (thymic Tregs) after exposure to self-antigens on thymic epithelial cells or can arise outside the thymus and are called peripheral Tregs. Intestinal Tregs are essential to preventing pathogenic immune responses to the gut microbiome, thus preventing intestinal inflammation.
- *T_H9 T cells*—CD4 T cells regulated by the transcription factor PU.1 that secrete IL-9 and enhance inflammation in atopic disease and inflammatory bowel disease as well as mediate antitumor immunity.
- *T_H13 T cells*—T follicular helper cells (TFH) regulated by the GATA3 transcription factor that produce IL-4, IL-5, and IL-13. T_H13 TFH induce high-affinity IgE antibody responses that cause anaphylactic reactions to allergens.
- *T_H17 T cells*—CD4 T cells regulated by the transcription factor RORγt that secrete IL-17, IL-22, and IL-26 and play roles in autoimmune inflammatory disorders as well as defend against bacterial and fungal pathogens.
- *Tolerance*—B- and T-cell nonresponsiveness to antigens that results from encounter with foreign or self-antigens by B and T lymphocytes in the absence of expression of APC co-stimulatory molecules. Tolerance to antigens may be induced and maintained by multiple mechanisms either centrally (B-cell deletion in the thymus for T cells or bone marrow for B cells) or peripherally (by cell deletion or anergy at sites throughout the peripheral immune system).
- *Trained immunity*—the epigenetic, transcriptional, and functional reprogramming of innate immune cells to adapt to previous encounters with pathogens and respond to a second challenge in an altered manner.

INTRODUCTION

The human immune system has evolved over millions of years from both invertebrate and vertebrate organisms to develop sophisticated defense mechanisms that protect the host from microbes and their virulence factors. The normal immune system has three key properties: a highly diverse repertoire of antigen receptors that enables recognition of a nearly infinite range of pathogens; immune memory, to mount rapid recall immune responses; and immunologic tolerance, to avoid immune damage to self-tissues.

From invertebrates, humans have inherited the *innate immune system*, an ancient defense system that uses germline-encoded proteins to recognize pathogens. Cells of the innate immune system, such as macrophages, DCs, and NK lymphocytes, recognize PAMPs that are highly conserved among many microbes and use a diverse set of PRR molecules. Important components of the recognition of microbes by the innate immune system include recognition by germline-encoded host molecules, recognition of key microbe virulence factors but not recognition of self-molecules, and nonrecognition of benign foreign molecules or microbes such as are found in mucosal or other barrier microbiomes. Upon contact with pathogens, cells of the innate immune system may kill pathogens directly or, in concert with DCs, activate a series of events that both slow the infection and recruit the more recently evolved arm of the human immune system, the *adaptive immune system*. In addition, innate immune cells undergo epigenetic, transcriptional, and functional changes that allow adapted (either enhanced or reduced) innate cell responses to repeat encounters with pathogens, called *trained immunity*.

Adaptive immunity is found only in vertebrates and is based on the generation of antigen receptors on T and B lymphocytes by gene rearrangements, such that individual T or B cells express unique antigen receptors on their surface capable of specifically recognizing diverse antigens of infectious agents in the environment. Coupled with specific recognition mechanisms that maintain tolerance (nonreactivity) to self-antigens or nonpathogenic microbes **(Chap. 361)**, T and B lymphocytes bring both *specificity* and *immune memory* to vertebrate host defenses.

This chapter describes the cellular components, key molecules **(Table 360-1)**, and mechanisms that make up the innate and adaptive immune systems and describes how adaptive immunity is recruited to the defense of the host by innate immune responses. An appreciation of the cellular and molecular bases of innate and adaptive immune responses is critical to understanding the pathogenesis of inflammatory, autoimmune, infectious, and immunodeficiency diseases, as well as a wide range of diseases associated with inflammation such as atherosclerotic cardiovascular disease and neurodegenerative diseases.

THE INNATE IMMUNE SYSTEM

All multicellular organisms, including humans, have developed the use of surface and intracellular germline-encoded molecules that recognize pathogens. Because of the myriad of human pathogens, host molecules of the human innate immune system sense "danger signals" and either recognize PAMPs, the common molecular structures shared by many pathogens, or recognize host cell molecules produced in response to infection such as heat shock proteins and fragments of the extracellular matrix. PAMPs must be conserved structures vital to pathogen virulence and survival, such as bacterial endotoxin, so that pathogens cannot mutate molecules of PAMPs to evade human innate immune responses. PRRs are host proteins of the innate immune system that recognize PAMPs as host danger signal molecules **(Tables 360-2 and 360-3)**. Thus, recognition of pathogen molecules by hematopoietic and nonhematopoietic cell types leads to activation/production of the complement cascade, cytokines, or antimicrobial peptides as effector molecules. In addition, pathogen PAMPs as host danger signal molecules activate DCs to mature and to express molecules on the DC surface that optimize antigen presentation to respond to foreign antigens.

TABLE 360-1 Human Leukocyte Surface Antigens—The CD Classification of Leukocyte Differentiation Antigens

SURFACE ANTIGEN (OTHER NAMES)	FAMILY	MOLECULAR MASS, kDa	DISTRIBUTION	LIGAND(S)	FUNCTION
CD1a (T6, HTA-1)	Ig	49	CD, cortical thymocytes, Langerhans type of DCs	TCRγδ T cells, NK T cells	CD1 molecules present lipid antigens of intracellular bacteria such as *Mycobacterium leprae* and *M. tuberculosis* to TCRγδT cells or NK T cells
CD1b	Ig	45	CD, cortical thymocytes, Langerhans type of DCs	TCRγδ T cells, NK T cells	
CD1c	Ig	43	DC, cortical thymocytes, subset of B cells, Langerhans type of DCs	TCRγδ T cells, NK T cells	
CD1d	Ig	37	Cortical thymocytes, intestinal epithelium, Langerhans type of DCs	TCRγδ T cells, NK T cells	
CD2 (T12, LFA-2)	Ig	50	T, NK	CD58, CD48, CD59, CD15	Alternative T-cell activation, T-cell anergy, T-cell cytokine production, T- or NK-mediated cytolysis, T-cell apoptosis, cell adhesion
CD3 (T3, Leu-4)	Ig	γ:25–28, δ:21–28, ε:20–25, η:21–22, ζ:16	T, NK T	Associates with the TCR	T-cell activation and function; ζ is the signal transduction component of the CD3 complex
CD4 (T4, Leu-3)	Ig	55	T, myeloid	MHC-II, HIV gp120, IL-16, SABP	T-cell selection, T-cell activation, signal transduction with p56*lck*, primary receptor for HIV-1
CD7 (3A1, Leu-9)	Ig	40	T, NK	K-12 (CD7L)	T- and NK-cell signal transduction and regulation of IFN-γ, TNF-α production
CD8 (T8, Leu-2)	Ig	34	T, subset of NK	MHC-I	T-cell selection, T-cell activation, signal transduction with p56*lck*
CD14 (LPS-receptor)	LRG	53–55	M, G (weak), not by myeloid progenitors	Endotoxin (lipopolysaccharide), lipoteichoic acid, PI	TLR4 mediates with LPS and other PAMP activation of innate immunity
CD16a (FcγRIIIa)	Ig	50–80	NK, macrophages, neutrophils	Fc portion of IgG	Mediates phagocytosis and ADCC
CD19 B4	Ig	95	B (except plasma cells), FDC	Not known	Associates with CD21 and CD81 to form a complex involved in signal transduction in B-cell development, activation, and differentiation
CD20 (B1)	Unassigned	33–37	B (except plasma cells)	Not known	Cell signaling, may be important for B-cell activation and proliferation
CD21 (B2, CR2, EBV-R, C3dR)	RCA	145	Mature B, FDC, subset of thymocytes	C3d, C3dg, iC3b, CD23, EBV	Associates with CD19 and CD81 to form a complex involved in signal transduction in B-cell development, activation, and differentiation; Epstein-Barr virus receptor
CD22 (BL-CAM)	Ig	130–140	Mature B	CDw75	Cell adhesion, signaling through association with p72*sky*, p53/56*lyn*, PI3 kinase, SHP1, fLCγ
CD23 (FcεRII, B6, Leu-20, BLAST-2)	C-type lectin	45	B, M, FDC	IgE, CD21, CD11b, CD11c	Regulates IgE synthesis, cytokine release by monocytes
CD28	Ig	44	T, plasma cells	CD80, CD86	Co-stimulatory for T-cell activation; involved in the decision between T-cell activation and anergy
CD32a (FcγRIIa)	Ig	40	NK, macrophages, neutrophils	Fc portion of IgG	Mediates phagocytosis and ADCC
CD40	TNFR	48–50	B, DC, EC, thymic epithelium, MP, cancers	CD154 (CD40L)	B-cell activation, proliferation, and differentiation; formation of GCs; isotype switching; rescue from apoptosis
CD45 (LCA, T200, B220)	PTP	180, 200, 210, 220	All leukocytes	Galectin-1, CD2, CD3, CD4	T and B activation, thymocyte development, signal transduction, apoptosis
CD45RA	PTP	210, 220	Subset T, medullary thymocytes, "naive" T	Galectin-1, CD2, CD3, CD4	Isoforms of CD45 containing exon 4 (A), restricted to a subset of T cells
CD45RB	PTP	200, 210, 220	All leukocytes	Galectin-1, CD2, CD3, CD4	Isoforms of CD45 containing exon 5 (B)
CD45RC	PTP	210, 220	Subset T, medullary thymocytes, "naive" T	Galectin-1, CD2, CD3, CD4	Isoforms of CD45 containing exon 6 (C), restricted to a subset of T cells
CD45RO	PTP	180	Subset T, cortical thymocytes, "memory" T	Galectin-1, CD2, CD3, CD4	Isoforms of CD45 containing no differentially spliced exons, restricted to a subset of T cells
CD64 (FcγRI)	Ig	45–55	Macrophages and monocytes	Fc portion of IgG	Mediates phagocytosis and ADCC
CD80 (B7-1, BB1)	Ig	60	Activated B and T, MP, DC	CD28, CD152 (CTLA-4)	Co-regulator of T-cell activation; signaling through CD28 stimulates and through CD152 inhibits T-cell activation
CD86 (B7-2, B70)	Ig	80	Subset B, DC, EC, activated T, thymic epithelium	CD28, CD152 (CTLA-4)	Co-regulator of T-cell activation; signaling through CD28 stimulates and through CD152 inhibits T-cell activation
CD89 (FCαR)	Ig	55–100	Neutrophils, eosinophils, monocytes, and MP	Fc portion of IgG	Mediates phagocytosis and ADCC of IgA-coated pathogens

(*Continued*)

TABLE 360-1 Human Leukocyte Surface Antigens—The CD Classification of Leukocyte Differentiation Antigens (Continued)

SURFACE ANTIGEN (OTHER NAMES)	FAMILY	MOLECULAR MASS, kDa	DISTRIBUTION	LIGAND(S)	FUNCTION
CD95 (APO-1, Fas)	TNFR	43	Activated T and B	Fas ligand	Mediates apoptosis
CD112 (nekton-2, PVRL2)	Ig	62	Epithelial cells, endothelial cells, other tissues	DNAM-1 (CD226), TIGIT	T-cell activation (DNAM-1), T-cell inhibition (TIGIT)
CD134 (OX40)	TNFR	48	Activated T	OX40L (CD252)	T-cell survival, cytokine stimulation
CD137 (4-1BB)	TNFR	19	Activated T, DCs, B, NK	CD137L (41BBL)	T-cell co-stimulation
CD155 (PVR)	Ig	50–65	DCs, NK, epithelial cells	TIGIT, CD96, DNAM-1	T-cell inhibition (TIGIT, CD96), T-cell activation (DNAM-1)
CD223 (LAG-3)	Ig	57	NK, B, activated T	MHC class II	T-cell inhibition
CD226 (DNAM-1)	Ig	65	NK, monocytes, T	CD112, CD155	T-cell activation (CD112), T-cell activation (CD155)
CD252 (OX40L)	TNFR	16–25	Antigen-presenting cells, endothelial cells	OX40	T-cell survival, cytokine stimulation
CD272 (BTLA)	Ig	16	Activated T	HVEM	T-cell inhibition
CD274 (PD-L1)	Ig	40	T, NK, myeloid, B, tumor cells	PD-1 (CD279)	Inhibit TCR activation
CD278 (ICOS)	Ig	55–60	Activated T	ICOSL	T-cell activation
CD357 (GTTR)	TNFR	41	Activated T, Tregs	GITRL	T-cell activation
CD152 (CTLA-4)	Ig	30–33	Activated T	CD80, CD86	Inhibits T-cell proliferation
CD154 (CD40L)	TNF	33	Activated CD4+ T, subset CD8+ T, NK, M, basophil	CD40	Co-stimulatory for T-cell activation, B-cell proliferation and differentiation
CD279 (PD-1)	Ig	50–55	B, T, TFH	PD-L1 (CD274), PD-L2 (CD273)	Inhibits T-cell proliferation

Abbreviations: ADCC, antibody-dependent cell-mediated cytotoxicity; BTLA, band T lymphocyte attenuators; CTLA, cytotoxic T lymphocyte–associated protein; DC, dendritic cells; DNAM-1, DNAX accessory molecule-1; EBV, Epstein-Barr virus; EC, endothelial cells; ECM, extracellular matrix; FcγRIII, low-affinity IgG receptor isoform A; FDC, follicular dendritic cells; G, granulocytes; GC, germinal center; GITR, glucocorticoid-induced TNFR-related protein; GPI, glycosyl phosphatidylinositol; HTA, human thymocyte antigen; HVEM, herpesvirus entry mediator; ICOS, inducible T-cell co-stimulator; Ig, immunoglobulin; IgG, immunoglobulin G; LAG-3, lymphocyte-activation gene 3; LCA, leukocyte common antigen; LPS, lipopolysaccharide; MHC-I, major histocompatibility complex class I; MP, macrophages; Mr, relative molecular mass; NK, natural killer cells; P, platelets; PBT, peripheral blood T cells; PD-1, programmed cell death-1; PI, phosphatidylinositol; PI3K, phosphatidylinositol 3-kinase; PLC, phospholipase C; PTP, protein tyrosine phosphatase; PVR, polio virus receptor; PVRL2, polio virus receptor-related 2; RCA, regulators of complement activation; SABP, seminal actin binding protein; TCR, T-cell receptor; TFH, T follicular helper cells; TIGIT, T-cell immunoreceptor with Ig and ITIM domains; TNF, tumor necrosis factor; TNFR, tumor necrosis factor receptor.

Note: For an expanded list of cluster of differentiation (CD) human antigens, see Harrison's Online at *accessmedicine.com*; and for a full list of CD human antigens from the most recent Human Workshop on Leukocyte Differentiation Antigens (VII), D Mason, P Andre, A Bensussan, et al (eds): *Leucocyte Typing VII*. Oxford: Oxford University Press, 2002.

Source: Compiled from T Kishimoto et al (eds): *Leukocyte Typing VI*. New York: Garland Publishing, 1997; R Brines et al: Immunol Today 18S:1, 1997; and D Mason et al: CD antigens 2002. Blood 99:3877, 2002.

PATTERN RECOGNITION

Major PRR families of proteins include transmembrane proteins, such as the Toll-like receptors (TLRs) and C-type lectin receptors (CLRs), and cytoplasmic proteins, such as the retinoic acid–inducible gene (RIG)-1-like receptors (RLRs) and NOD-like receptors (NLRs) (Table 360-4). A major group of PRR collagenous glycoproteins with C-type lectin domains are termed *collectins* and include the serum protein mannose-binding lectin (MBL). MBL and other collectins, as well as two other protein families—the pentraxins (such as C-reactive protein and serum amyloid P) and macrophage scavenger receptors—all have the property of opsonizing (coating) bacteria for phagocytosis by macrophages and can also activate the complement cascade to lyse bacteria. Integrins are cell-surface adhesion molecules that affect attachment between cells and the extracellular matrix and mediate signal transduction that reflects the chemical composition of the cell environment. For example, integrins signal after cells bind bacterial lipopolysaccharide (LPS) and activate phagocytic cells to ingest pathogens.

TABLE 360-2 Major Components of the Innate Immune System

Pattern recognition receptors (PRRs)	Toll-like receptors (TLRs), C-type lectin receptors (CLRs), retinoic acid–inducible gene (RIG)-1-like receptors (RLRs), and NOD-like receptors (NLRs)
Antimicrobial peptides	α-Defensins, β-defensins, cathelin, protegrin, granulysin, histatin, secretory leukoprotease inhibitor, and probiotics
Cells	Macrophages, dendritic cells, innate lymphoid cells (ILC1, ILC2, ILC3, NK cells, lymphoid tissue inducer [LTi] cells), mucosal-associated invariant T (MAIT) cells, NK-T cells, neutrophils, eosinophils, mast cells, basophils, and epithelial cells
Complement components	Classic and alternative complement pathway, and proteins that bind complement components
Cytokines	Autocrine, paracrine, endocrine cytokines that mediate host defense and inflammation, as well as recruit, direct, and regulate adaptive immune responses

Abbreviation: NK, natural killer.

There are multiple connections between the innate and adaptive immune systems; these include (1) a plasma protein, LPS-binding protein, that binds and transfers LPS to the macrophage LPS receptor, CD14; (2) the human family of proteins called *Toll-like receptor proteins* (TLRs), some of which are associated with CD14, bind LPS, and signal epithelial cells, DCs, and macrophages to produce cytokines and upregulate cell-surface molecules that signal the initiation of adaptive immune responses (**Fig. 360-1**, Table 360-3); and (3) families of intracellular microbial sensors called NLRs and RLRs. Proteins in the Toll family can be expressed on macrophages, DCs, and B cells as well as on a variety of nonhematopoietic cell types, including respiratory epithelial cells. Eleven TLRs have been identified in humans (Table 360-3). Upon ligation, TLRs activate a series of intracellular events that lead to the killing of bacteria- and viral-infected cells as well as to the recruitment and ultimate activation of antigen-specific T and B lymphocytes (Fig. 360-1). Importantly, signaling by massive amounts of LPS through TLR4 leads to the release of high levels of cytokines that mediate LPS-induced shock. Mutations in TLR4 proteins in mice protect from LPS shock, and TLR mutations in humans can protect from LPS-induced inflammatory diseases such as LPS-induced asthma. **Table 360-4** lists diseases caused by gene variants in nucleic acid–sensing Toll family and related receptors.

Two other families of cytoplasmic PRRs are the NLRs and the RLRs. These families, unlike the TLRs, are composed primarily of soluble

TABLE 360-3 Pattern Recognition Receptors (PRRs) and Their Ligands

PRR	LOCALIZATION	LIGAND	ORIGIN OF THE LIGAND
TLR			
TLR1	Plasma membrane	Triacyl lipoprotein	Bacteria
TLR2	Plasma membrane	Lipoprotein	Bacteria, viruses, parasite, self
TLR3	Endolysosome	dsRNA	Virus
TLR4	Plasma membrane	LPS	Bacteria, viruses, self
TLR5	Plasma membrane	Flagellin	Bacteria
TLR6	Plasma membrane	Diacyl lipoprotein	Bacteria, viruses
TLR7 (human TLR8)	Endolysosome	ssRNA	Virus, bacteria, self
TLR9	Endolysosome	CpG-DNA	Virus, bacteria, protozoa, self
TLR10	Endolysosome	Unknown	Unknown
TLR11	Plasma membrane	Profilin-like molecule	Protozoa
RLR			
RIG-I	Cytoplasm	Short dsRNA, triphosphate dsRNA	RNA viruses, DNA virus
MDA5	Cytoplasm	Long dsRNA	RNA viruses (Picornaviridae)
LGP2	Cytoplasm	Unknown	RNA viruses
NLR			
NOD1	Cytoplasm	iE-DAP	Bacteria
NOD2	Cytoplasm	MDP	Bacteria
CLR			
Dectin-1	Plasma membrane	β^2-Glucan	Fungi
Dectin-2	Plasma membrane	β^2-Glucan	Fungi
MINCLE	Plasma membrane	SAP130	Self, fungi

Abbreviations: CLR, C-type lectin receptors; dsRNA, double-strand RNA; iE-DAP, D-glutamyl-meso-diaminopimelic acid moiety; LGP2, Laboratory of Genetics and Physiology 2 protein encoded by the gene *DHX58*; MDA5, melanoma differentiation-associated protein 5; MDP, MurNAc-L-Ala-D-isoGln, also known as muramyl dipeptide; MINCLE, macrophage-inducible C-type lectin; NLR, NOD-like receptor; NOD, NOTCH protein domain; RIG, retinoic acid–inducible gene; RLR, RIG-like receptors; SAP130, Sin-3 associated protein 130; TLR, Toll-like receptor.

Source: Reproduced with permission from O Takeuchi: Pattern recognition receptors and inflammation. Cell 140:805, 2010.

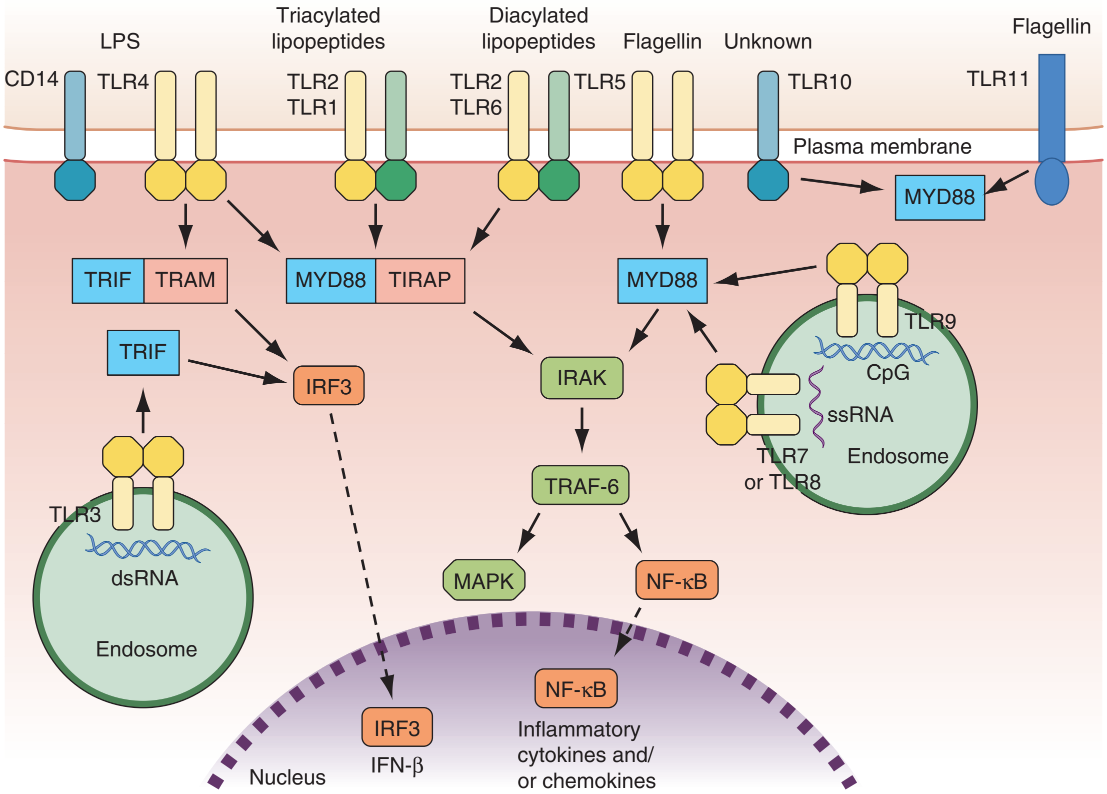
LPS
Triacylated lipopeptides
Diacylated lipopeptides
Flagellin
Unknown
Flagellin
CD14
TLR4
TLR2
TLR1
TLR2
TLR6
TLR5
TLR10
TLR11
Plasma membrane
MYD88
TRIF
TRAM
MYD88
TIRAP
MYD88
TLR9
TRIF
IRF3
IRAK
CpG
ssRNA
TLR7
or TLR8
Endosome
TRAF-6
TLR3
dsRNA
Endosome
MAPK
NF-κB
NF-κB
IRF3
IFN-β
Inflammatory
cytokines and/
or chemokines
Nucleus

FIGURE 360-1 Overview of major TLR signaling pathways. All TLRs signal through MYD88, with the exception of TLR3. TLR4 and the TLR2 subfamily (TLR1, TLR2, TLR6) also engage TIRAP (Toll-interleukin 1 receptor domain-containing adapter protein). TLR3 signals through TRIF (Toll-interleukin 1 receptor domain-containing adapter-inducing interferon-β). TRIF is also used in conjunction with TRAM (TRIF-related adaptor molecule) in the TLR4-MYD88-independent pathway. *Dashed arrows* indicate translocation into the nucleus. dsRNA, double-strand RNA; IFN, interferon; IRF3, interferon regulatory factor 3; LPS, lipopolysaccharide; MAPK, mitogen-activated protein kinases; NF-κB, nuclear factor-κB; ssRNA, single-strand RNA; TLR, Toll-like receptor. *(Reproduced with permission from D Van Duin et al: Triggering TLR signaling in vaccination. Trends Immunol 27:49, 2006.)*

TABLE 360-4 Diseases Caused by Gene Variants in Nucleic Acid-Sensing Receptors and Related Proteins

GENE	PROTEIN	CHANGE IN FUNCTION	LOCATION	LIGAND OR PARTNER	AUTOIMMUNE OR AUTOINFLAMMATORY DISEASE IN OMIM	OMIM NUMBER	LUPUS SUSCEPTIBILITY ALLELES
DNASE1	DNASE1	*LOF*	Extracellular	dsDNA (NETs)	SLE	152700	Yes
DNASE1L3	DNASE1L3	*LOF*	Extracellular	Nucleosomes exposed on microparticles or apoptotic bodies	SLE16	614420	Yes
TLR7	TLR7	*GOF*	Endosomal	ssRNA, 2′3′cGMP	SLE 17	301080	Yes
TLR9	TLR9	–	Endosomal	CpG dsDNA	N/A	–	No
MyD88	MyD88	–	Endosomal	TLRs	N/A (GOF)	–	No
ADA2	ADA2	*LOF*	Endolysosomal	Adenosine, 2′-deoxyadenosine	Sneddon syndrome Vasculitis, autoinflammation, immunodeficiency, and hematologic defects syndrome (DADA2)	182410 615688	No
DNASE2	DNASE2	*LOF*	Lysosomal	Exogenous DNA	Autoinflammatory-pancytopenia syndrome	619858	No
TREX1	TREX1	*LOF*	Cytoplasmic	DNA (retroviral/retrotransposon)	AGS 1 Chilblain lupus Vasculopathy, retinal, with cerebral leukoencephalopathy and systemic manifestations Susceptibility to SLE	225750 610448 192315 152700	Yes
TMEM173	STING	*LOF*	Cytoplasmic	cGAMP and excess DNA	STING-associated vasculopathy, infantile-onset	615934	Yes
CGAS	cGAS	–	Cytoplasmic	Cytosolic DNA	N/A (GOF)	–	No
SAMHD1	SAMHD1	*LOF*	Cytoplasmic	dNTPs	Chilblain lupus? AGS 5	614415 612952	Yes
RNASEH2A	RNASEH2A	*LOF*	Cytoplasmic	RNA in RNA-DNA hybrids	AGS 4	610333	Yes
RNASEH2B	RNASEH2B	*LOF*	Cytoplasmic	RNA in RNA-DNA hybrids	AGS 2	610181	Yes
RNASEH2C	RNASEH2C	*LOF*	Cytoplasmic	RNA in RNA-DNA hybrids	AGS 3	610329	Yes
IFIH1	MDA5	*GOF*	Cytoplasmic	Long dsRNA	AGS 7 Singleton-Merten syndrome 1	615846 182250	Yes
RIGI	DDX58	*GOF*	Cytoplasmic	Short dsRNA	Singleton-Merten syndrome 2	616298	No
MAVS	MAVS	–	Cytoplasmic	–	N/A (GOF)	–	No
AIM2	AIM2	–	Cytoplasmic	Cytosolic dsDNA	N/A	–	No
IFI16	IFI16	–	Nuclear	Viral DNA or damaged self-DNA	N/A (GOF)	–	No

Note: See the online catalogue of human genes and genetic disorders (OMIM) at *https://omim.org*.

Abbreviations: GOF, gain of function; LOF, loss of function; N/A, not applicable; SLE, systemic lupus erythematosus.

Source: Reproduced with permission from CG Vinuesa CG et al: Innate virus-sensing pathways in B cell systemic autoimmunity. Science 380:478, 2023.

intracellular proteins that scan host cell cytoplasm for intracellular pathogens (Tables 360-2 and 360-3).

The intracellular microbial sensors, NLRs, after triggering, form large cytoplasmic complexes termed *inflammasomes*, which are aggregates of molecules including NOD-like receptor pyrin (NLRP) proteins **(Table 360-5)**. Inflammasomes activate inflammatory caspases and IL-1β in the presence of nonbacterial danger signals (cell stress) and bacterial PAMPs. Mutations in inflammasome proteins can lead to chronic inflammation in a group of periodic febrile diseases called *autoinflammatory syndromes*. Polymorphisms in inflammasome components can either protect or enhance risk of infections or autoimmune/autoinflammatory diseases (Table 60-4). Inflammasomes are activated upon sensing of PAMPs. Crystallopathies are diseases caused by tissue crystal deposition such as monosodium urate that can activate the inflammasome and, in the case of urate deposition, can lead to gout with arthritis or renal disease.

■ EFFECTOR CELLS OF INNATE IMMUNITY

Cells of the innate immune system and their roles in the first line of host defense are listed in **Table 360-6**. Equally important as their roles in the mediation of innate immune responses are the roles that each cell type plays in recruiting T and B lymphocytes of the adaptive immune system to engage in specific pathogen responses.

Monocytes-Macrophages Monocytes arise from precursor cells within bone marrow **(Fig. 360-2)** and circulate with a half-life ranging from 1 to 3 days. Monocytes leave the peripheral circulation via capillaries and migration into a vast extravascular cellular pool. Tissue macrophages arise from monocytes that have migrated out of the circulation and by in situ proliferation of macrophage precursors in tissue. Common locations where tissue macrophages (and certain of their specialized forms) are found are lymph node, spleen, bone marrow, perivascular connective tissue, serous cavities such as the peritoneum, pleura, skin connective tissue, lung (alveolar macrophages), liver (Kupffer cells), bone (osteoclasts), central nervous system (microglia cells), and synovium (type A lining cells).

In general, monocytes-macrophages are on the first line of defense associated with innate immunity and ingest and destroy microorganisms through the release of toxic products such as hydrogen peroxide (H_2O_2) and nitric oxide (NO). Inflammatory mediators produced by macrophages attract additional effector cells such as neutrophils to the site of infection. Macrophage mediators include prostaglandins; leukotrienes; platelet activating factor; cytokines such as IL-1, TNF-α, IL-6, and IL-12; and chemokines **(Tables 360-7 and 360-8)**.

Although monocytes-macrophages were originally thought to be the major APCs of the immune system, it is now clear that cell types called *dendritic cells* are the most potent and effective classical APCs in the body (see below). Monocytes-macrophages mediate innate immune effector functions such as destruction of antibody-coated bacteria, tumor cells, or even normal hematopoietic cells in certain types of autoimmune cytopenias. Monocytes-macrophages ingest bacteria

TABLE 360-5 Mutations in Innate Inflammasome Molecules Associated with Clinical Disease

Inherited Inflammasomopathies

MUTATED GENE	DISEASE	INHERITED PATTERN AND EFFECT	PHENOTYPE	PREDOMINANT EFFECTOR CELLS
NLRP1	NLRP1-associated autoinflammation with arthritis and dyskeratosis	Autosomal dominant GoF	Hyperkeratotic ulcerative skin lesions, fever, arthritis, ANA	Keratinocytes
NLRP3	Cryopyrin-associated periodic syndromes (CAPS)	Autosomal dominant GoF	Spectrum from cold-induced urticaria and fever to CNS inflammation and bone overgrowth	Monocytes, granulocytes (neutrophils), chondrocytes
NLRC4	Autoinflammatory infantile fever with enterocolitis (AIFEC)	Autosomal dominant GoF	Recurrent MAS, enterocolitis, cold-induced fever and urticaria, CNS inflammation	Monocytes/macrophages
MEFV	Familial Mediterranean fever (FMF)	Autosomal recessive LoF or gene-dosage-dependent autosomal dominant GoF	Fever, serositis, rash, SAA amyloidosis	Neutrophils, monocytes, serosal and synovial fibroblasts

Genetic Polymorphisms in Inflammasome Components and Human Infectious Diseases

INFECTIOUS AGENT/DISEASE	GENE	VARIANT ID	EFFECT ON INFLAMMASOME ACTIVATION	ASSOCIATION
Candida albicans (recurrent vulvovaginal candidiasis)	*NLRP3*	rs74163773	Increased	Risk
Chlamydia trachomatis	*NLRP3*	rs12065526	Unknown	Risk
HCV	*NLRP3*	rs1539019; rs35829419	Unknown; increased	Protection
HIV-1	*NLRP3*	rs10754558	Increased	Protection
	IFI16	rs1417806	Increased	Protection
HPV	*NLRP1*	rs11651270	Increased	Protection
	NLRP3	rs10754558	Increased	Protection
HSV-2	*IFI16*	rs2276404	Increased	Protection
HTLV	*NLRP3*	rs10754558	Increased	Protection
Microbial infection in lungs	*NLRC4*	rs212704	Decreased	Risk
Mycobacterium leprae	*NLRP1*	rs2670660, rs12150220	Increased	Protection
		rs2137722	(Haplotype)	
Mycobacterium tuberculosis	*NLRP3*	rs10754558	Increased	Protection
		rs10754558	Increased	Risk
	CARD8	rs6509365	Unknown	Risk
	NLRC4	rs385076	Decreased	Protection
Plasmodium vivax	*NLRP1*	rs12150220	Increased	Risk
Renal parenchymal infections	*NLRP3*	rs4612666	Increased	Protection
Streptococcus pneumoniae	*NLRP1*	rs11651270	Increased	Risk
	CARD8	rs2043211	Increased	
Trypanosoma cruzi	*NLRP1*	rs11691270	Increased	Risk
	CASP1	rs501192	Unknown	Risk

Genetic Polymorphisms in Inflammasome Components and Autoimmune in Polygenic Autoinflammatory Diseases

INFECTIOUS AGENT/DISEASE	GENE	VARIANT ID	EFFECT ON INFLAMMASOME ACTIVATION	ASSOCIATION
Addison disease	*NLRP1*	rs12150220	Increased	Risk
Ankylosing spondylitis	*NLRP3*	rs4612666	Increased	Risk
	MEFV	rs224204	Unknown	Risk
	CARD8	rs2043211	Increased	Protection
Autoimmune thyroiditis	*NLRP1*	rs12150220, rs2670660	Increased	Risk
	AIM2	rs855873	Unknown	Risk
Behçet disease	*AIM2*	rs855873	Unknown	Risk
	IFI16	rs6940	Decreased	
Celiac disease	*NLRP3*	rs35829419	Increased	Protection; risk
IBD: Crohn's disease (CD) and ulcerative colitis (UC)	*NLRP3*	rs35829419	Increased	Risk (men)
			Increased	Protection
		rs10754558	Increased	Risk
		rs10925019	Unknown	Risk
		rs4925648	Unknown	Risk
		rs4353135, rs55646866; rs4266924, rs6672995, rs10733113	Decreased; unknown	Risk

(*Continued*)

TABLE 360-5 Mutations in Innate Inflammasome Molecules Associated with Clinical Disease (Continued)

INFECTIOUS AGENT/DISEASE	GENE	VARIANT ID	EFFECT ON INFLAMMASOME ACTIVATION	ASSOCIATION
IBD: Crohn's disease (CD) and ulcerative colitis (UC) (Cont.)	*MEFV*	rs182674, rs224217, rs224225, rs224224, rs224223, rs224222	Unknown	Risk
	CARD8	rs2043211	Increased	Risk
				Protection
		rs1972619	Unknown	Risk
HS purpura	*MEFV*	rs3743930	Unknown	Risk
Kawasaki disease	*NLRP1*	rs11651270, rs8079034, rs3744717, rs11078571, rs16954813, rs8079727	Increased (haplotype)	Risk
Multiple sclerosis	*NLRP3*	rs3806265, rs10754557	Unknown	Risk
		rs35829419	Increased	Risk
	NLRC4	rs479333	Decreased	Protection
PFAPA	*CARD8*	rs140826611	Unknown	Risk
Psoriasis	*NLRP1*	rs8079034	Unknown	Risk
	NLRP3	rs3806265, rs10754557	Unknown	Risk
		rs10733113	Unknown	Risk
	CARD8	rs2043211	Increased	Risk
	AIM2	rs2276405	Unknown	Protection
Psoriatic JIA	*NLRP3*	rs4353135	Decreased	Risk
		rs3806265	Unknown	Risk
	MEFV	rs224204	Unknown	Risk
Rheumatoid arthritis	*NLRP1*	rs878329	Unknown	Risk
	NLRP3	rs35829419	Increased	Risk
		rs10754558	Increased	Risk
		rs10159239, rs4925648, rs4925659	Unknown	Risk
	CASP5	rs9651713	Unknown	Risk
SLE	*NLRP1*	rs12150220, rs2670660	Increased	Risk
Systemic sclerosis	*NLRP1*	rs8182352	Unknown	Risk
Type 1 diabetes	*NLRP1*	rs12150220	Increased	Risk
		rs2670660, rs11651270	Increased	Protection
	NLRP3	rs10754558	Increased	Protection
Vitiligo	*NLRP1*	rs12150220	Increased	Risk
		rs2670660	Increased	Risk
		rs8182352	Unknown	Risk
		rs6502867	Unknown	Risk
		rs1008588	Unknown	Risk

Note: Mutated gene and respective syndrome name are reported for inflammasomopathies, as well as inheritance pattern and effect of mutations, clinical phenotype, and predominant disease effector cells. Inflammasome variants previously associated with infectious agents and/or diseases are briefly resumed from literature (*https://www.ncbi.nlm.nih.gov/pubmed*). Significantly associated polymorphisms were grouped according to the infectious agent/disease. Infectious agent or disease (in alphabetical order), gene name (gene), identification number of polymorphism (ID), resulting effect on inflammasome activation ("increased," "decreased," or "unknown"), cohort origin (cohort) and eventually specifications (severity, etc.), sample size (n), and type (case/control or cases only), association result ("risk" or "protection"), and respective reference are reported.

Abbreviations: ANA, antinuclear antibodies; CNS, central nervous system; GoF, gain-of-function; HCV, hepatitis C virus; HIV, human immunodeficiency virus; HPV, human papillomavirus; HS, Henoch-Schönlein; HSV, herpes simplex virus; HTLV, human T-lymphotropic virus; IBD, inflammatory bowel disease; JIA, juvenile idiopathic arthritis; LoF, loss-of-function; MAS, macrophage activation syndrome; PFAPA, periodic fever with aphthous stomatitis, pharyngitis, and cervical adenitis; SAA, serum amyloid A; SLE, systemic lupus erythematosus.

Source: Reproduced with permission from FP Fernandes et al: Inflammasome genetics and complex diseases: A comprehensive review. Eur J Hum Genet 28:1307, 2020.

or are infected by viruses, and in doing so, they frequently undergo programmed cell death or *apoptosis*. Macrophages that are infected by intracellular infectious agents are recognized by DCs as infected and apoptotic cells and are phagocytosed by DCs. In this manner, DCs "cross-present" infectious agent antigens of macrophages to T cells. Activated macrophages can also mediate antigen-nonspecific lytic activity and eliminate cell types such as tumor cells in the absence of antibody. This activity is largely mediated by cytokines (i.e., TNF-α and IL-1). Monocytes-macrophages express lineage-specific molecules (e.g., the cell-surface LPS receptor, CD14) as well as surface receptors for a number of molecules, including the Fc region of IgG, activated complement components, and various cytokines (Table 360-7).

Dendritic Cells Human DCs contain several subsets, including myeloid DCs and plasmacytoid DCs. Myeloid DCs can differentiate into either macrophages-monocytes or tissue-specific DCs. In contrast to myeloid DCs, plasmacytoid DCs are potent producers of TLR-7 dependent type I IFN (e.g., IFN-α) in response to free virus and virus-infected cells. The maturation of DCs is regulated through cell-to-cell contact and soluble factors, and DCs attract immune effectors through secretion of chemokines. When DCs come in contact with bacterial products, viral proteins, or host proteins released as danger signals from distressed host cells (Fig. 360-2), infectious agent molecules bind to various TLRs and activate DCs to release cytokines and chemokines that drive cells of the innate immune system to become activated to

TABLE 360-6 Cells of the Innate Immune System and Their Major Roles in Triggering Adaptive Immunity

CELL TYPE	MAJOR ROLE IN INNATE IMMUNITY	MAJOR ROLE IN ADAPTIVE IMMUNITY
Macrophages	Phagocytose and kill bacteria; produce antimicrobial peptides; bind LPS; produce inflammatory cytokines	Produce IL-1 and TNF-α to upregulate lymphocyte adhesion molecules and chemokines to attract antigen-specific lymphocyte. Produce IL-12 to recruit T_H1 T helper cell responses; upregulate co-stimulatory and MHC molecules to facilitate T and B lymphocyte recognition and activation. Macrophages and dendritic cells, after LPS signaling, upregulate co-stimulatory molecules B7-1 (CD80) and B7-2 (CD86) that are required for activation of pathogen-specific T cells. There are also Toll-like proteins on B cells and dendritic cells that, after LPS ligation, induce CD80 and CD86 on these cells for T-cell antigen presentation.
Plasmacytoid dendritic cells (DCs) of lymphoid lineage	Produce large amounts of interferon-α (IFN-α), which has antitumor and antiviral activity, and are found in T-cell zones of lymphoid organs; they circulate in blood	IFN-α is a potent activator of macrophage and mature DCs to phagocytose invading pathogens and present pathogen antigens to T and B cells.
Myeloid DCs are of two types: interstitial and Langerhans-derived	Interstitial DCs are strong producers of IL-12 and IL-10 and are located in T-cell zones of lymphoid organs, circulate in blood, and are present in the interstices of the lung, heart, and kidney; Langerhans DCs are strong producers of IL-12; are located in T-cell zones of lymph nodes, skin epithelia, and the thymic medulla; and circulate in blood	Interstitial DCs are potent activators of macrophage and mature DCs to phagocytose invading pathogens and present pathogen antigens to T and B cells.
ILC1 cells	Weakly cytotoxic, dependent on T-bet transcription factor, first line of defense against viruses and bacteria	Produce IFN-γ to recruit CD4 T_H1 T cells
ILC2 cells	Mediate innate responses to parasites/helminths, repair damaged tissues by producing amphiregulin	Produce IL-4, IL-5, IL-13; recruit CD4 T_H2 T cells
ILC3 cells	Innate immune response to extracellular bacteria and gut microbiome	Produce IL-22, IL-17, GM-CSF, lymphotoxin; recruit CD4 T_H17 T cells
Lymphoid tissue inducer (LTi) cells	Critical for formation of secondary lymphoid tissue during embryogenesis	Produce lymphotoxin for lymph node and Peyer's patch development in which adaptive immune responses occur
Natural killer (NK) cells	Kill foreign and host cells that have low levels of MHC+ self-peptides. Express NK receptors that inhibit NK function in the presence of high expression of self-MHC.	Produce TNF-α and IFN-γ, which recruit T_H1 helper T-cell responses
NK-T cells	Lymphocytes with both T-cell and NK surface markers that recognize lipid antigens of intracellular bacteria such as *Mycobacterium tuberculosis* by CD1 molecules and kill host cells infected with intracellular bacteria	Produce IL-4 to recruit T_H2 helper T-cell responses, IgG1 and IgE production
Neutrophils	Phagocytose and kill bacteria, produce antimicrobial peptides	Produce nitric oxide synthase and nitric oxide, which inhibit apoptosis in lymphocytes and can prolong adaptive immune responses
Eosinophils	Kill invading parasites	Produce IL-5, which recruits Ig-specific antibody responses
Mast cells and basophils	Release TNF-α, IL-6, and IFN-γ in response to a variety of bacterial PAMPs	Produce IL-4, which recruits T_H2 helper T cell responses, and recruit IgG1- and IgE-specific antibody responses
Epithelial cells	Produce antimicrobial peptides; tissue-specific epithelia produce mediator of local innate immunity; e.g., lung epithelial cells produce surfactant proteins (proteins within the collectin family) that bind and promote clearance of lung-invading microbes	Produces TGF-β, which triggers IgA-specific antibody responses

Abbreviations: GM-CSF, granulocyte-macrophage colony-stimulating factor; IL-4, IL-5, IL-6, IL-10, and IL-12, interleukin 4, 5, 6, 10, and 12, respectively; ILC, innate lymphoid cell; MHC, major histocompatibility complex: LPS, lipopolysaccharide; PAMP, pathogen-associated molecular patterns; TGF, transforming growth factor; T_H, helper T cell; TNF-α, tumor necrosis factor-alpha.

Source: Reproduced with permission from R Medzhitov, CA Janeway: Curr Opinion Immunol 9:4-9; 1997.

respond to invading organisms, and recruit T and B cells of the adaptive immune system to respond. Plasmacytoid DCs produce antiviral IFN-α that activates NK cell killing of pathogen-infected cells; IFN-α also activates CD8 T cells to mature into antipathogen cytotoxic (killer) T cells. Following contact with pathogens, both plasmacytoid and myeloid DCs produce chemokines that attract helper and cytotoxic T cells, B cells, polymorphonuclear cells, and naïve and memory T cells as well as regulatory T cells to ultimately dampen the immune response once the pathogen is controlled. TLR engagement on DCs upregulates MHC class II, B7-1 (CD80), and B7-2 (CD86), which enhance DC-specific antigen presentation and induce cytokine production. Thus, DCs are important bridges between early (innate) and later (adaptive) immunity. DCs also modulate and determine the types of immune responses induced by pathogens via the TLRs expressed on DCs (TLR7–9 in plasmacytoid DCs, TLR4 on monocytoid DCs) and via the TLR adapter proteins that are induced to associate with TLRs (Fig. 360-1, Table 360-1). In addition, other PRRs, such as C-type lectins, NLRs, and mannose receptors, upon ligation by pathogen products, activate cells of the adaptive immune system and, like TLR stimulation, by a variety of factors, determine the type and quality of the adaptive immune response that is triggered.

Innate Lymphoid Cells ILCs are comprised of ILC1, ILC2, ILC3, lymphoid tissue inducer (LTi), and NK cells. ILC1, ILC2, ILC3, and LTi are primarily tissue resident cells. ILCs develop from a common lymphoid precursor in the bone marrow and then differentiate into one of five ILC types—ILC1, ILC2, ILC3, LTi, or NK cells—based on their development **(Fig. 360-3*A*)** and function **(Fig. 360-3*B*)**. NK cells and ILC1s depend on T-bet transcription factor for their development and function and produce IFN-γ. NK cells are innate analogues to CD8 cytotoxic T cells in that they both mediate granzyme and perforin-based cytotoxic cell activity. ILC1s mirror CD4 T_H1 lymphocytes and react to intracellular pathogens such as viruses and to tumors. ILC2s are the analogues of T_H2 CD4 T cells and are dependent on GATA3 and RORα factors and produce type 2 cytokines, such as IL-5 and IL-13. ILC2s respond to extracellular parasites and allergens. ILC3s and LTi cells are dependent on transcription factor retinoic acid receptor-related orphan receptor γt (RORγt) and produce IL-17. ILC3s are analogues of CD4 T_H17 lymphocytes and attack extracellular pathogens such as bacteria and fungi. LTi cells are critical for the formation of lymph nodes and Peyer's patches in gut during fetal development (Fig. 360-3*B*).

In the intestine, a critical function of the immune system is not only to quickly respond to pathogens but also to ignore benign

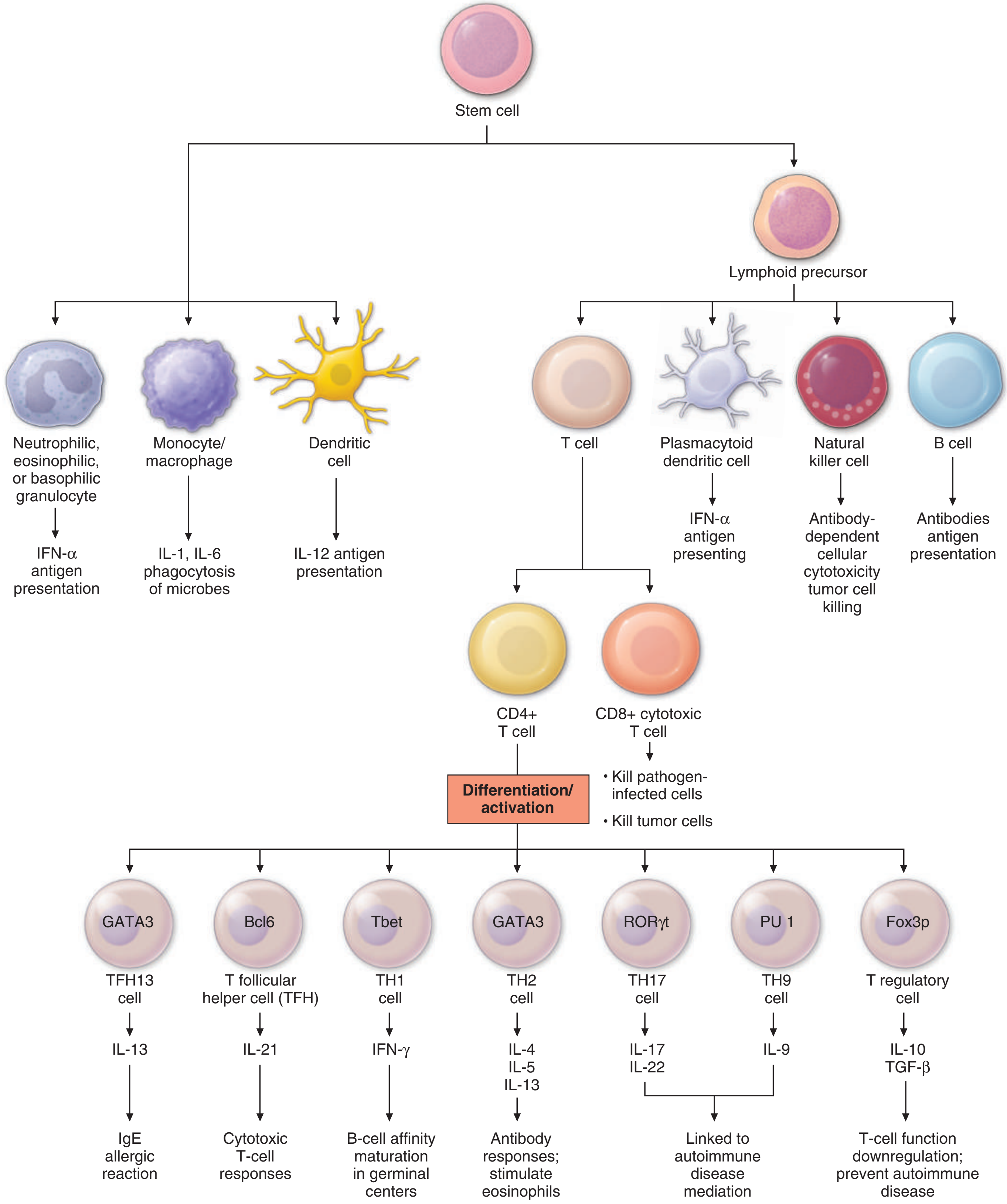
Stem cell
Lymphoid precursor
Neutrophilic, eosinophilic, or basophilic granulocyte
IFN-α antigen presentation
Monocyte/ macrophage
IL-1, IL-6 phagocytosis of microbes
Dendritic cell
IL-12 antigen presentation
T cell
Plasmacytoid dendritic cell
IFN-α antigen presenting
Natural killer cell
Antibody-dependent cellular cytotoxicity tumor cell killing
B cell
Antibodies antigen presentation
CD4+ T cell
CD8+ cytotoxic T cell
• Kill pathogen-infected cells
• Kill tumor cells
Differentiation/ activation
GATA3
TFH13 cell
IL-13
IgE allergic reaction
Bcl6
T follicular helper cell (TFH)
IL-21
Cytotoxic T-cell responses
Tbet
TH1 cell
IFN-γ
B-cell affinity maturation in germinal centers
GATA3
TH2 cell
IL-4
IL-5
IL-13
Antibody responses; stimulate eosinophils
RORγt
TH17 cell
IL-17
IL-22
PU 1
TH9 cell
IL-9
Linked to autoimmune disease mediation
Fox3p
T regulatory cell
IL-10
TGF-β
T-cell function downregulation; prevent autoimmune disease

FIGURE 360-2 Model of immune effector cell development. Hematopoietic stem cells differentiate into T cells, antigen-presenting dendritic cells, natural killer cells, macrophages, granulocytes, or B cells. Foreign antigen is processed by dendritic cells, macrophages, and B cells, and peptide fragments of foreign antigen are presented to CD4+ and/or CD8+ T cells. CD8+ T-cell activation leads to induction of cytotoxic T lymphocyte (CTL) or killer T-cell generation, as well as induction of cytokine-producing CD8+ cytotoxic T cells. Granulocytes (neutrophils, eosinophils, or basophils) are effector cells of the innate immune system and mediate anti-infectious agent activity by cytokine production, infectious agent killing, or both. T_H1 CD4+ T cells play an important role in defense against intracellular microbes and help in the generation of CD8+ cytotoxic T cells. T_H2 CD4+ T cells producing interferon (IFN) γ or interleukin (IL) 4, IL-5, or IL-13 regulate Ig class switching and determine the type of antibody produced. T_H17 cells secrete IL-17 and IL-22, T_H9 cells secrete IL-9, and TFH13 cells secrete IL-4, IL-5, and IL-13. T_H17 and T_H9 CD4 T cells are linked to mediation of autoimmune disease, and TFH13 cells are linked to IgE-mediated anaphylaxis. CD4+ T regulatory cells produce IL-10 and transforming growth factor (TGF)-β and downregulate T- and B-cell responses once the microbe has been eliminated. Each of the types of CD4+ T cells are regulated by different transcription factors, and the key transcription factors are shown in the circles above each CD4+ T-cell type.

TABLE 360-7 Cytokines and Cytokine Receptors

CYTOKINE	RECEPTOR	CELL SOURCE	CELL TARGET	BIOLOGIC ACTIVITY
IL-1α, β	Type I IL-1r, type II IL-1r	Monocytes/macrophages, B cells, fibroblasts, most epithelial cells including thymic epithelium, endothelial cells	All cells	Upregulates adhesion molecule expression, neutrophil and macrophage emigration, mimics shock, fever, upregulates hepatic acute-phase protein production, facilitates hematopoiesis
IL-2	IL-2r α, β, common γ	T cells	T cells, B cells, NK cells, monocytes-macrophages	Promotes T-cell activation and proliferation, B-cell growth, NK-cell proliferation and activation, enhanced monocyte/macrophage cytolytic activity
IL-3	IL-3r, common β	T cells, NK cells, mast cells	Monocytes-macrophages, mast cells, eosinophils, bone marrow progenitors	Stimulates hematopoietic progenitors
IL-4	IL-4r α, common γ	T cells, mast cells, basophils	T cells, B cells, NK cells, monocytes-macrophages, neutrophils, eosinophils, endothelial cells, fibroblasts	Stimulates T_H2 helper T-cell differentiation and proliferation; stimulates B-cell Ig class switch to IgG1 and IgE anti-inflammatory action on T cells, monocytes; produced by T follicular helper cells in B-cell germinal centers that stimulate B-cell maturation
IL-5	IL-5r α, common γ	T cells, mast cells, eosinophils	Eosinophils, basophils, murine B cells	Regulates eosinophil migration and activation
IL-6	IL-6r, gp130	Monocytes-macrophages, B cells, fibroblasts, most epithelium including thymic epithelium, endothelial cells	T cells, B cells, epithelial cells, hepatocytes, monocytes-macrophages	Induces acute-phase protein production, T- and B-cell differentiation and growth, myeloma cell growth, and osteoclast growth and activation
IL-7	IL-7r α, common γ	Bone marrow, thymic epithelial cells	T cells, B cells, bone marrow cells	Differentiates B-, T-, and NK-cell precursors, activates T and NK cells
IL-8	CXCR1, CXCR2	Monocytes-macrophages, T cells, neutrophils, fibroblasts, endothelial cells, epithelial cells	Neutrophils, T cells, monocytes-macrophages, endothelial cells, basophils	Induces neutrophil, monocyte, and T-cell migration, induces neutrophil adherence to endothelial cells and histamine release from basophils, and stimulates angiogenesis; suppresses proliferation of hepatic precursors
IL-9	IL-9r α, common γ	T cells	Bone marrow progenitors, B cells, T cells, mast cells	Induces mast cell proliferation and function, synergizes with IL-4 in IgG and IgE production and T-cell growth, activation, and differentiation
IL-10	IL-10r	Monocytes-macrophages, T cells, B cells, keratinocytes, mast cells	Monocytes-macrophages, T cells, B cells, NK cells, mast cells	Inhibits macrophage proinflammatory cytokine production, downregulates cytokine class II antigen and B7-1 and B7-2 expression, inhibits differentiation of T_H1 helper T cells, inhibits NK cell function, stimulates mast cell proliferation and function, B-cell activation, and differentiation
IL-11	IL-11r α, gp130	Bone marrow stromal cells	Megakaryocytes, B cells, hepatocytes	Induces megakaryocyte colony formation and maturation, enhances antibody responses, stimulates acute-phase protein production
IL-12 (35-kDa and 40-kDa subunits)	IL-12r	Activated macrophages, dendritic cells, neutrophils	T cells, NK cells	Induces T_H1 T helper cell formation and lymphokine-activated killer cell formation; increases CD8+ CTL cytolytic activity; ↓IL-17, ↑IFN-γ
IL-13	IL-13r/IL-4r α	T cells (T_H2)	Monocytes-macrophages, B cells, endothelial cells, keratinocytes	Upregulates VCAM-1 and C-C chemokine expression on endothelial cells and B-cell activation and differentiation, and inhibits macrophage proinflammatory cytokine production
IL-14	Unknown	T cells	Normal and malignant B cells	Induces B-cell proliferation, inhibits antibody secretion, and expands selected B-cell subgroups
IL-15	IL-15r α, common γ, IL2r β	Monocytes-macrophages, epithelial cells, fibroblasts	T cells, NK cells	Promotes T-cell activation and proliferation, angiogenesis, and NK cells
IL-16	CD4	Mast cells, eosinophils, CD8+ T cells, respiratory epithelium	CD4+ T cells, monocytes-macrophages, eosinophils	Promotes chemoattraction of CD4+ T cells, monocytes, and eosinophils; inhibits HIV-1 replication; inhibits T-cell activation through CD3/T-cell receptor
IL-17	IL-17r	CD4+ T cells	Fibroblasts, endothelium, epithelium, macrophages	Enhances cytokine/chemokine secretion; promotes delayed-type reactions
IL-18	IL-18r (IL-1R-related protein)	Keratinocytes, macrophages	T cells, B cells, NK cells	Upregulates IFN-γ production, enhances NK cell cytotoxicity
IL-21	IL-δγ chain/IL-21R	CD4 T cells	NK cells	Downregulates NK cell–activating molecules, NKG2D/DAP10; produced by T follicular helper cells in B-cell germinal centers that stimulate B-cell maturation
IL-22	IL-22 R1/IL-10R2	DC, T cells	Epithelial cells	Innate responses against bacterial pathogens; promotes hepatocyte survival
IL-23	IL-12Rb1/IL23R	Macrophages, other cell types	T cells	Opposite effects of IL-12 (↑IL-17, ↑IFN-γ)
IL-24	IL-20R1/IL-20R2 IL-22R1/IL-20R2	Macrophages, T_H2 cells	Nonhematopoietic cells such as fibroblasts	Promotes wound healing

(*Continued*)

TABLE 360-7 Cytokines and Cytokine Receptors (Continued)

CYTOKINE	RECEPTOR	CELL SOURCE	CELL TARGET	BIOLOGIC ACTIVITY
IL-25 (also called IL-17E)	IL-17RB	CD4 T cells, mast cells	Fibroblasts, endothelium, epithelium, macrophages	Proinflammatory; induces cytokine production
IL-26	IL-20R1/IL-10R2	T_H1, T_H17 T cells, synovial cells	Epithelial cells	Proinflammatory; induces cytokine production
IL-27	gp130t wsx-1	Myeloid cells such as macrophages and DCs	T cells	Collaborates with other cytokines to activate T-cell differentiation
IL-28A (IFN-λ2)	IFN-λ receptor 1, IL-28Rα, IL-10Rβ	Myeloid lineage cells; epithelial cells	Epithelial cells	Enhanced clearance of viral infections
IL-28B (IFN-λ3)	IFN-λ receptor 1, IL-28Rα, IL-10Rβ	Myeloid lineage cells; epithelial cells	Epithelial cells	Enhanced clearance of viral infections
IL-29 (IFN-λ1)	IFN-λ receptor 1, IL-28Rα, IL-10Rβ	Myeloid lineage cells; epithelial cells	Epithelial cells	Enhanced clearance of viral infections
IL-30 (p28 of IL-27)	IL-27Rα; gp130+wsx-1	Activated macrophages and DCs; epithelial malignancies	Monocytes	Anti-inflammatory cytokines; upregulation of breast and prostate cancer metastasis
IL-31	IL-31RA/ oncostatin MRβ	Eosinophils, CD4 T cells	Epithelial cells, monocytes	Pruritis, proinflammatory
IL-32 (NK4)	?	Monocytes, T cells, NK cells, epithelial cells	Monocytes, macrophages, bone marrow stroma	Angiogenesis, IL-2 production in bone marrow, proinflammatory
IL-33 (NF-HEV; IL-1 F11)	ST-2	Endothelial cells, epithelial cells, fibroblasts, mucosal epithelium	T cells, mast cells eosinophils, basophils, ILC2s	Alarmin cytokine, proinflammatory
IL-34 (C16of77)	CSF-1R, PTP-E, CD138	Neurons, Treg, myeloid cells		Anti-inflammatory myeloid cell proliferation
IL-35	IL-12Rβ2/ IL-12Rβ2, gp130/ gp130, IL-12Rb2/ gp130	Tregs, Bregs	Macrophages, T cells	Prevents T_H1 and T_H17 proliferation; induced Treg/Breg proliferation/anti-inflammatory
IL-36α IL36β IL36γ IL36RA (IL-1 F5)	IL-36R	Keratocytes Mucosal epithelial cells Monocytes-macrophages Langerhans cells CD4 T cells	Epithelial cells, macrophages, DCs, T cells, B cells, plasma cells	T_H responses, proinflammatory
IL-38 IL-10 F10	IL-1R, IL-36R, IL-1RA PL1	Epithelial cells, B cells	Epithelial cells, macrophages, DCs, T cells, B cells, plasma cells	Blocks IL-36; anti-inflammatory
IL-39	?	Macrophages, DCs, B cells	Neutrophils	Proinflammatory
IL-40	?	B cells, bone marrow/stroma	B cells	Involved in IgA production, B-cell homeostasis and development
IFN-α	Type I interferon receptor	All cells	All cells	Promotes antiviral activity; stimulates T-cell, macrophage, and NK-cell activity; direct antitumor effects; upregulates MHC class I antigen expression; used therapeutically in viral and autoimmune conditions
IFN-β	Type I interferon receptor	All cells	All cells	Antiviral activity; stimulates T-cell, macrophage, and NK-cell activity; direct antitumor effects; upregulates MHC class I antigen expression; used therapeutically in viral and autoimmune conditions
IFN-γ	Type II interferon receptor	T cells, NK cells	All cells	Regulates macrophage and NK-cell activations; stimulates immunoglobulin secretion by B cells; induction of class II histocompatibility antigens; T_H1 T-cell differentiation
TNF-α	TNFrI, TNFrII	Monocytes-macrophages, mast cells, basophils, eosinophils, NK cells, B cells, T cells, keratinocytes, fibroblasts, thymic epithelial cells	All cells except erythrocytes	Fever, anorexia, shock, capillary leak syndrome, enhanced leukocyte cytotoxicity, enhanced NK-cell function, acute phase protein synthesis, proinflammatory cytokine induction
TNF-β	TNFrI, TNFrII	T cells, B cells	All cells except erythrocytes	Cell cytotoxicity, lymph node and spleen development
LT-β	LTβR	T cells	All cells except erythrocytes	Cell cytotoxicity, normal lymph node development
G-CSF	G-CSFr; gp130	Monocytes-macrophages, fibroblasts, endothelial cells, thymic epithelial cells, stromal cells	Myeloid cells, endothelial cells	Regulates myelopoiesis; enhances survival and function of neutrophils; clinical use in reversing neutropenia after cytotoxic chemotherapy
GM-CSF	GM-CSFr, common β	T cells, monocytes-macrophages, fibroblasts, endothelial cells, thymic epithelial cells	Monocytes-macrophages, neutrophils, eosinophils, fibroblasts, endothelial cells	Regulates myelopoiesis; enhances macrophage bactericidal and tumoricidal activity; mediator of dendritic cell maturation and function; upregulates NK-cell function; clinical use in reversing neutropenia after cytotoxic chemotherapy

(*Continued*)

TABLE 360-7 Cytokines and Cytokine Receptors (Continued)

CYTOKINE	RECEPTOR	CELL SOURCE	CELL TARGET	BIOLOGIC ACTIVITY
M-CSF	M-CSFr (*c-fms* protooncogene)	Fibroblasts, endothelial cells, monocytes-macrophages, T cells, B cells, epithelial cells including thymic epithelium	Monocytes-macrophages	Regulates monocyte-macrophage production and function
LIF	LIFr-α; gp130	Activated T cells, bone marrow stromal cells, thymic epithelium	Megakaryocytes, monocytes, hepatocytes, possibly lymphocyte subpopulations	Induces hepatic acute-phase protein production; stimulates macrophage differentiation; promotes growth of myeloma cells and hematopoietic progenitors; stimulates thrombopoiesis
OSM	OSMr; LIFr; gp130	Activated monocytes-macrophages and T cells, bone marrow stromal cells, some breast carcinoma cell lines, myeloma cells	Neurons, hepatocytes, monocytes-macrophages, adipocytes, alveolar epithelial cells, embryonic stem cells, melanocytes, endothelial cells, fibroblasts, myeloma cells	Induces hepatic acute-phase protein production; stimulates macrophage differentiation; promotes growth of myeloma cells and hematopoietic progenitors; stimulates thrombopoiesis; stimulates growth of Kaposi's sarcoma cells
SCF	SCFr (*c-kit* protooncogene)	Bone marrow stromal cells and fibroblasts	Embryonic stem cells, myeloid and lymphoid precursors, mast cells	Stimulates hematopoietic progenitor cell growth, mast cell growth; promotes embryonic stem cell migration
TGF-β (3 isoforms)	Type I, II, III TGF-β receptor	Most cell types	Most cell types	Downregulates T-cell, macrophage, and granulocyte responses; stimulates synthesis of matrix proteins; stimulates angiogenesis
Lymphotactin/ SCM-1	XCR1	NK cells, mast cells, double-negative thymocytes, activated CD8+ T cells	T cells, NK cells	Chemoattractant for lymphocytes; only known chemokine of C class
MCP-1	CCR2	Fibroblasts, smooth-muscle cells, activated PBMCs	Monocytes-macrophages, NK cells, memory T cells, basophils	Chemoattractant for monocytes, activated memory T cells, and NK cells; induces granule release from CD8+ T cells and NK cells; potent histamine-releasing factor for basophils; suppresses proliferation of hematopoietic precursors; regulates monocyte protease production
MCP-2	CCR1, CCR2	Fibroblasts, activated PBMCs	Monocytes-macrophages, T cells, eosinophils, basophils, NK cells	Chemoattractant for monocytes, memory and naïve T cells, eosinophils,? NK cells; activates basophils and eosinophils; regulates monocyte protease production
MCP-3	CCR1, CCR2	Fibroblasts, activated PBMCs	Monocytes-macrophages, T cells, eosinophils, basophils, NK cells, dendritic cells	Chemoattractant for monocytes, memory and naïve T cells, dendritic cells, eosinophils,? NK cells; activates basophils and eosinophils; regulates monocyte protease production
MCP-4	CCR2, CCR3	Lung, colon, small intestinal epithelial cells, activated endothelial cells	Monocytes-macrophages, T cells, eosinophils, basophils	Chemoattractant for monocytes, T cells, eosinophils, and basophils
Eotaxin	CCR3	Pulmonary epithelial cells, heart	Eosinophils, basophils	Potent chemoattractant for eosinophils and basophils; induces allergic airways disease; acts in concert with IL-5 to activate eosinophils; antibodies to eotaxin inhibit airway inflammation
TARC	CCR4	Thymus, dendritic cells, activated T cells	T cells, NK cells	Chemoattractant for T and NK cells
MDC	CCR4	Monocytes-macrophages, dendritic cells, thymus	Activated T cells	Chemoattractant for activated T cells; inhibits infection with T-cell tropic HIV-1
MIP-1α	CCR1, CCR5	Monocytes-macrophages, T cells	Monocytes-macrophages, T cells, dendritic cells, NK cells, eosinophils, basophils	Chemoattractant for monocytes, T cells, dendritic cells, and NK cells, and weak chemoattractant for eosinophils and basophils; activates NK-cell function; suppresses proliferation of hematopoietic precursors; necessary for myocarditis associated with coxsackievirus infection; inhibits infection with monocytotropic HIV-1
MIP-1β	CCR5	Monocytes-macrophages, T cells	Monocytes-macrophages, T cells, NK cells, dendritic cells	Chemoattractant for monocytes, T cells, and NK cells; activates NK-cell function; inhibits infection with monocytotropic HIV-1
RANTES	CCR1, CCR2, CCR5	Monocytes-macrophages, T cells, fibroblasts, eosinophils	Monocytes-macrophages, T cells, NK cells, dendritic cells, eosinophils, basophils	Chemoattractant for monocytes-macrophages, CD4+, CD45Ro+ T cells, CD8+ T cells, NK cells, eosinophils, and basophils; induces histamine release from basophils; inhibits infections with monocytotropic HIV-1
LARC/MIP-3α/ Exodus-1	CCR6	Dendritic cells, fetal liver cells, activated T cells	T cells, B cells	Chemoattractant for lymphocytes
ELC/MIP-3β	CCR7	Thymus, lymph node, appendix	Activated T cells and B cells	Chemoattractant for B and T cells; receptor upregulated on EBV-infected B cells and HSV-infected T cells
I-309/TCA-3	CCR8	Activated T cells	Monocytes-macrophages, T cells	Chemoattractant for monocytes; prevents glucocorticoid-induced apoptosis in some T-cell lines
SLC/TCA-4/ Exodus-2	CCR7	Thymic epithelial cells, lymph node, appendix, and spleen	T cells	Chemoattractant for T lymphocytes; inhibits hematopoiesis
DC-CK1/PARC	Unknown	Dendritic cells in secondary lymphoid tissues	Naïve T cells	May have a role in induction of immune responses

(*Continued*)

TABLE 360-7 Cytokines and Cytokine Receptors (Continued)

CYTOKINE	RECEPTOR	CELL SOURCE	CELL TARGET	BIOLOGIC ACTIVITY
TECK	CCR9	Dendritic cells, thymus, liver, small intestine	T cells, monocytes-macrophages, dendritic cells	Thymic dendritic cell–derived cytokine, possibly involved in T-cell development
GRO-α/MGSA	CXCR2	Activated granulocytes, monocyte-macrophages, and epithelial cells	Neutrophils, epithelial cells,? endothelial cells	Neutrophil chemoattractant and activator; mitogenic for some melanoma cell lines; suppresses proliferation of hematopoietic precursors; angiogenic activity
GRO-β/MIP-2α	CXCR2	Activated granulocytes and monocyte-macrophages	Neutrophils and? endothelial cells	Neutrophil chemoattractant and activator; angiogenic activity
NAP-2	CXCR2	Platelets	Neutrophils, basophils	Derived from platelet basic protein; neutrophil chemoattractant and activator
IP-10	CXCR3	Monocytes-macrophages, T cells, fibroblasts, endothelial cells, epithelial cells	Activated T cells, tumor-infiltrating lymphocytes,? endothelial cells,? NK cells	IFN-γ-inducible protein that is a chemoattractant for T cells; suppresses proliferation of hematopoietic precursors
MIG	CXCR3	Monocytes-macrophages, T cells, fibroblasts	Activated T cells, tumor-infiltrating lymphocytes	IFN-γ-inducible protein that is a chemoattractant for T cells; suppresses proliferation of hematopoietic precursors
SDF-1	CXCR4	Fibroblasts	T cells, dendritic cells,? basophils,? endothelial cells	Low-potency, high-efficacy T-cell chemoattractant; required for B lymphocyte development; prevents infection of CD4+, CXCR4+ cells by T-cell tropic HIV-1
Fractalkine	CX3CR1	Activated endothelial cells	NK cells, T cells, monocytes-macrophages	Cell-surface chemokine/mucin hybrid molecule that functions as a chemoattractant, leukocyte activator, and cell adhesion molecule
PF-4	Unknown	Platelets, megakaryocytes	Fibroblasts, endothelial cells	Chemoattractant for fibroblasts; suppresses proliferation of hematopoietic precursors; inhibits endothelial cell proliferation and angiogenesis

Abbreviations: B7-1, CD80; B7-2, CD86; Breg, regulatory B cells; CCR, CC-type chemokine receptor; CXCR, CXC-type chemokine receptor; DC, dendritic cell; DC-CK, dendritic cell chemokine; EBV, Epstein-Barr virus; ELC, EB11 ligand chemokine (MIP-1b); G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte-macrophage colony-stimulating factor; GRP, growth-related peptide; HSV, herpes simplex virus; IFN, interferon; Ig, immunoglobulin; IL, interleukin; IP-10, IFN-γ-inducible protein-10; LARC, liver- and activation-regulated chemokine; LIF, leukemia inhibitory factor; MCP, monocyte chemotactic protein; M-CSF, macrophage colony-stimulating factor; MDC, macrophage-derived chemokine; MGSA, melanoma growth-stimulating activity; MHC, major histocompatibility complex; MIG, monokine induced by IFN-γ; MIP, macrophage inflammatory protein; NAP, neutrophil-activating protein; NK, natural killer; OSM, oncostatin M; PARC, pulmonary- and activation-regulated chemokine; PBMC, peripheral blood mononuclear cells; PF, platelet factor; RANTES, regulated on activation, normally T cell–expressed and –secreted; SCF, stem cell factor; SDF, stromal cell–derived factor; SLC, secondary lymphoid tissue chemokine; TARC, thymus- and activation-regulated chemokine; TCA, T-cell activation protein; TECK, thymus-expressed chemokine; TGF, transforming growth factor; T_H1 and T_H2, helper T cell subsets; TNF, tumor necrosis factor; Treg, regulatory T cells; VCAM, vascular cell adhesion molecule.

Sources: Data from JS Sundy et al: Appendix B, in *Inflammation, Basic Principles and Clinical Correlates*, 3rd ed. J Gallin, R Snyderman (eds). Philadelphia, Lippincott Williams and Wilkins, 1999; J Ye et al: Frontiers in Pharmacology 11; HM Lazear et al: Immunity 43:15, 2015; J Catalan-Dibene et al: J Interferon and Cytokine Research 38:423, 2018.

microorganisms or environmental antigens. Inability to prevent immune responses to the gut microbiome can be a factor in the cause of inflammatory bowel diseases **(Chap. 337)**. In healthy individuals, a subset of RORγt+ ILCs, possibly ILC3s, stimulate naïve CD4 T cells to differentiate into peripheral Tregs that suppress immune responses to microbiome organisms.

NK cells express surface receptors for the Fc portion of IgG (FcR) (CD16) and for NCAM-I (CD56), and many NK cells express T lineage markers, particularly CD2, CD7, and CD8, and proliferate in response to IL-2. NK cells arise in both bone marrow and thymic microenvironments. In addition to mediating cytotoxicity to foreign or malignant cells, NK cells also mediate ADCC. ADCC is the binding of an opsonized (antibody-coated) target cell to an Fc receptor-bearing effector cell via the Fc region of antibody, resulting in target cell lysis. NK cell cytotoxicity is the MHC-unrestricted, non-antibody-mediated killing of target cells, which are usually malignant cell types, transplanted foreign cells, or virus-infected cells. Thus, NK cell cytotoxicity may play an important role in immune surveillance and destruction of malignant and virus-infected host cells. NK cell hyporesponsiveness is also observed in patients with *Chédiak-Higashi syndrome*, an autosomal recessive disease associated with fusion of cytoplasmic granules and defective degranulation of neutrophil lysosomes.

NK cells have a variety of surface receptors that have inhibitory or activating functions **(Table 360-9)**. NK immunoglobulin superfamily receptors include the killer cell immunoglobulin-like activating or inhibitory receptors (KIRs), many of which have been shown to have HLA class I ligands. The KIRs are made up proteins with either two (KIR2D) or three (KIR3D) extracellular immunoglobulin domains (D). Moreover, their nomenclature designates their function as either inhibitory KIRs with a long (L) cytoplasmic tail and immunoreceptor tyrosine-based inhibitory motif (ITIM) (KIRDL) or activating KIRs with a short (S) cytoplasmic tail (KIRDS). NK cell inactivation by KIRs is a central mechanism to prevent damage to normal host cells. Genetic studies have demonstrated the association of KIRs with viral infection outcome, or to outcomes in autoimmune or malignant diseases **(Table 360-10)**.

In addition to the KIRs, a second set of immunoglobulin superfamily receptors includes the natural cytotoxicity receptors (NCRs), which include NKp46, NKp30, and NKp44. These receptors help to mediate NK cell activation against target cells. The ligands to which NCRs bind on target cells have been recently recognized to be comprised of molecules of pathogens such as influenza, cytomegalovirus, and malaria, as well as host molecules expressed on tumor cells. Signaling lymphocytic activating molecule (SLAM) family receptors are expressed on hematopoietic cells, with SLAMF2 (CD48), SLAMF4 (2B4), and SLAMF7/CD2-like receptor activating cytotoxic cells (CRACCs) the most prominent on NK cells (Table 360-9).

NK cell signaling is, therefore, a highly coordinated series of inhibiting and activating signals that prevent NK cells from responding to uninfected, nonmalignant self-cells; however, they are activated to attack malignant and virally infected cells **(Fig. 360-4)**. Recent evidence suggests that NK cells, although not possessing rearranging immune recognition genes, may be able to mediate recall for NK cell responses to viruses and for immune responses such as contact hypersensitivity.

Some NK cells express CD3 and invariant TCR-α chains and are termed *NK T cells*. TCRs of NK T cells recognize lipid molecules of intracellular bacteria when presented in the context of CD1 molecules on APCs. Upon activation, NK T cells secrete effector cytokines such as IL-4 and IFN-γ. This mode of recognition of intracellular bacteria such as *Listeria monocytogenes* and *Mycobacterium tuberculosis* by NK T cells leads to induction of activation of DCs and is thought to be an important innate defense mechanism against these organisms.

TABLE 360-8 CC, CXC$_1$, CX$_3$, C$_1$, and XC Families of Chemokines and Chemokine Receptors

CHEMOKINE RECEPTOR	CHEMOKINE LIGANDS	CELL TYPES	DISEASE CONNECTION
CCR1	CCL3 (MIP-1α), CCL5 (RANTES), CCL7 (MCP-3), CCL14 (HCC1)	T cells, monocytes, eosinophils, basophils	Rheumatoid arthritis, multiple sclerosis
CCR2	CCL2 (MCP-1), CCL8 (MCP-2), CCL7 (MCP-3), CCL13 (MCP-4), CCL16 (HCC4)	Monocytes, dendritic cells (immature), memory T cells	Atherosclerosis, rheumatoid arthritis, multiple sclerosis, resistance to intracellular pathogens, type 2 diabetes mellitus
CCR3	CCL11 (eotaxin), CCL13 (eotaxin-2), CCL7 (MCP-3), CCL5 (RANTES), CCL8 (MCP-2), CCL13 (MCP-4)	Eosinophils, basophils, mast cells, T$_H$2, platelets	Allergic asthma and rhinitis
CCR4	CCL17 (TARC), CCL22 (MDC)	T cells (T$_H$2), dendritic cells (mature), basophils, macrophages, platelets	Parasitic infection, graft rejection, T-cell homing to skin
CCR5	CCL3 (MIP-1α), CCL4 (MIP-1α), CCL5 (RANTES), CCL11 (eotaxin), CCL14 (HCC1), CCL16 (HCC4)	T cells, monocytes	HIV-1 co-receptor (T cell–tropic strains), transplant rejection
CCR6	CCL20 (MIP-3α, LARC)	T cells (T regulatory and memory), B cells, dendritic cells	Mucosal humoral immunity, allergic asthma, intestinal T-cell homing
CCR7	CCL19 (ELC), CCL21 (SLC)	T cells, dendritic cells (mature)	Transport of T cells and dendritic cells to lymph nodes, antigen presentation, and cellular immunity
CCR8	CCL1 (1309)	T cells (T$_H$2), monocytes, dendritic cells	Dendritic cell migration to lymph node, type 2 cellular immunity, granuloma formation
CCR9	CCL25 (TECK)	T cells, IgA+ plasma cells	Homing of T cells and IgA+ plasma cells to the intestine, inflammatory bowel disease
CCR10	CCL27 (CTACK), CCL28 (MEC)	T cells	T-cell homing to intestine and skin
CXCR1	CXCL8 (interleukin-8), CXCL6 (GCP2)	Neutrophils, monocytes	Inflammatory lung disease, COPD
CXCR2	CXCL8, CXCL1 (GROα), CXCL2 (GROα), CXCL3 (GROα), CXCL5 (ENA-78), CXCL6	Neutrophils, monocytes, microvascular endothelial cells	Inflammatory lung disease, COPD, angiogenic for tumor growth
CXCR3-A	CXCL9 (MIG), CXCL10 (IP-10), CXCL11 (I-TAC)	Type 1 helper cells, mast cells, mesangial cells	Inflammatory skin disease, multiple sclerosis, transplant rejection
CXCR3-B	CXCL4 (PF4), CXCL9 (MIG), CXCL10 (IP-10), CXCL11 (I-TAC)	Microvascular endothelial cells, neoplastic cells	Angiostatic for tumor growth
CXCR4	CXCL12 (SDF-1)	Widely expressed	HIV-1 co-receptor (T cell–tropic), tumor metastases, hematopoiesis
CXCR5	CXCL13 (BCA-1)	B cells, follicular helper T cells	Formation of B-cell follicles
CXCR6	CXCL16 (SR-PSOX)	CD8+ T cells, natural killer cells, and memory CD4+ T cells	Inflammatory liver disease, atherosclerosis (CXCL16)
CX$_3$CR1	CX3CL1 (fractalkine)	Macrophages, endothelial cells, smooth-muscle cells	Atherosclerosis
XCR1	XCL1 (lymphotactin), XCL2	T cells, natural killer cells	Rheumatoid arthritis, IgA nephropathy, tumor response

Abbreviations: BCA-1, B-cell chemoattractant 1; COPD, chronic obstructive pulmonary disease; CTACK, cutaneous T cell–attracting chemokine; ELC, Epstein-Barr I1-ligand chemokine; ENA, epithelial cell–derived neutrophil-activating peptide; GCP, granulocyte chemotactic protein; GRO, growth-regulated oncogene; HCC, hemofiltrate chemokine; IP-10, interferon inducible 10; I-TAC, interferon-inducible T-cell alpha chemoattractant; LARC, liver- and activation-regulated chemokine; MCP, monocyte chemoattractant protein; MDC, macrophage-derived chemokine; MEC, mammary-enriched chemokine; MIG, monokine induced by interferon-γ; MIP, macrophage inflammatory protein; PF, platelet factor; SDF, stromal cell–derived factor; SLC, secondary lymphoid-tissue chemokine; SR-PSOX, scavenger receptor for phosphatidylserine-containing oxidized lipids; TARC, thymus- and activation-regulated chemokine; TECK, thymus-expressed chemokine; T$_H$2, type 2 helper T cells.

Source: From IF Charo, RM Ranshohoff: The many roles of chemokines and chemokine receptors in inflammation. N Engl J Med 354:610, 2006. Copyright © (2006) Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

The receptors for the Fc portion of IgG (FcγRs) are present on NK cells, B cells, macrophages, neutrophils, and mast cells and mediate interactions of IgG with antibody-coated target cells, such as virally infected cells. Antibody-NK interaction via antibody Fc and NK cell FcR links the adaptive and innate immune systems and regulates the mediation of IgG antibody effector functions such as ADCC. There are both activation and inhibitory FcγRs. Activation FcRs, such as FcγRI (CD64), FcγRIIa (CD32a), and FcγRIIIa (CD16a), are characterized by the presence of an immunoreceptor tyrosine-based activating motif (ITAM) sequence, whereas inhibitory FcRs, such as FcγRIIb (CD32b), contain an ITIM sequence. There is evidence that dysregulation in IgG-FcγR interactions plays roles in arthritis, multiple sclerosis, and systemic lupus erythematosus.

Neutrophils, Eosinophils, and Basophils Granulocytes are present in nearly all forms of inflammation and are amplifiers and effectors of innate immune responses (Fig. 360-2). Unchecked accumulation and activation of granulocytes can lead to host tissue damage, as seen in neutrophil- and eosinophil-mediated *systemic necrotizing vasculitis*. Granulocytes are derived from stem cells in bone marrow. Each type of granulocyte (neutrophil, eosinophil, or basophil) is derived from a different subclass of progenitor cell that is stimulated to proliferate by colony-stimulating factors (Table 360-6). During terminal maturation of granulocytes, class-specific nuclear morphology and cytoplasmic granules appear that allow for histologic identification of granulocyte type.

Neutrophils express Fc receptor IIIa for IgG (CD16a) as well as receptors for activated complement components (C3b or CD35). Upon interaction of neutrophils with antibody-coated (opsonized) bacteria or immune complexes, azurophilic granules (containing myeloperoxidase, lysozyme, elastase, and other enzymes) and specific granules (containing lactoferrin, lysozyme, collagenase, and other enzymes) are released, and microbicidal superoxide radicals (O_2^-) are generated at the neutrophil surface. The generation of superoxide leads to inflammation by direct injury to tissue and by alteration of macromolecules such as collagen and DNA.

Eosinophils are potent cytotoxic effector cells for various parasitic organisms. In *Nippostrongylus brasiliensis* helminth infection, eosinophils are important cytotoxic effector cells for removal of these parasites. Key to regulation of eosinophil cytotoxicity to *N. brasiliensis* worms are antigen-specific T helper cells that produce IL-4, thus providing an example of regulation of innate immune responses by adaptive immunity antigen-specific T cells. Intracytoplasmic contents of eosinophils, such as major basic protein, eosinophil cationic protein, and eosinophil-derived neurotoxin, are capable of directly damaging tissues and may be responsible in part for the organ system dysfunction in the *hypereosinophilic syndromes* **(Chap. 67)**. Because the eosinophil granule contains anti-inflammatory types of enzymes (histaminase, arylsulfatase, phospholipase D), eosinophils may also downregulate or terminate ongoing inflammatory responses.

Basophils and tissue mast cells are potent reservoirs of cytokines such as IL-4 and can respond to bacteria and viruses with cytokine production through multiple TLRs expressed on their surface. Mast cells and basophils can also mediate antipathogen immunity through the binding of antibodies. This is a particularly important host defense mechanism against parasitic diseases. Basophils express high-affinity surface receptors for IgE (FcεRII) (CD23) and, upon cross-linking of basophil-bound IgE by antigen, can release histamine, eosinophil chemotactic factor of anaphylaxis, and neutral proteases—all mediators of allergic immediate (anaphylaxis) hypersensitivity responses.

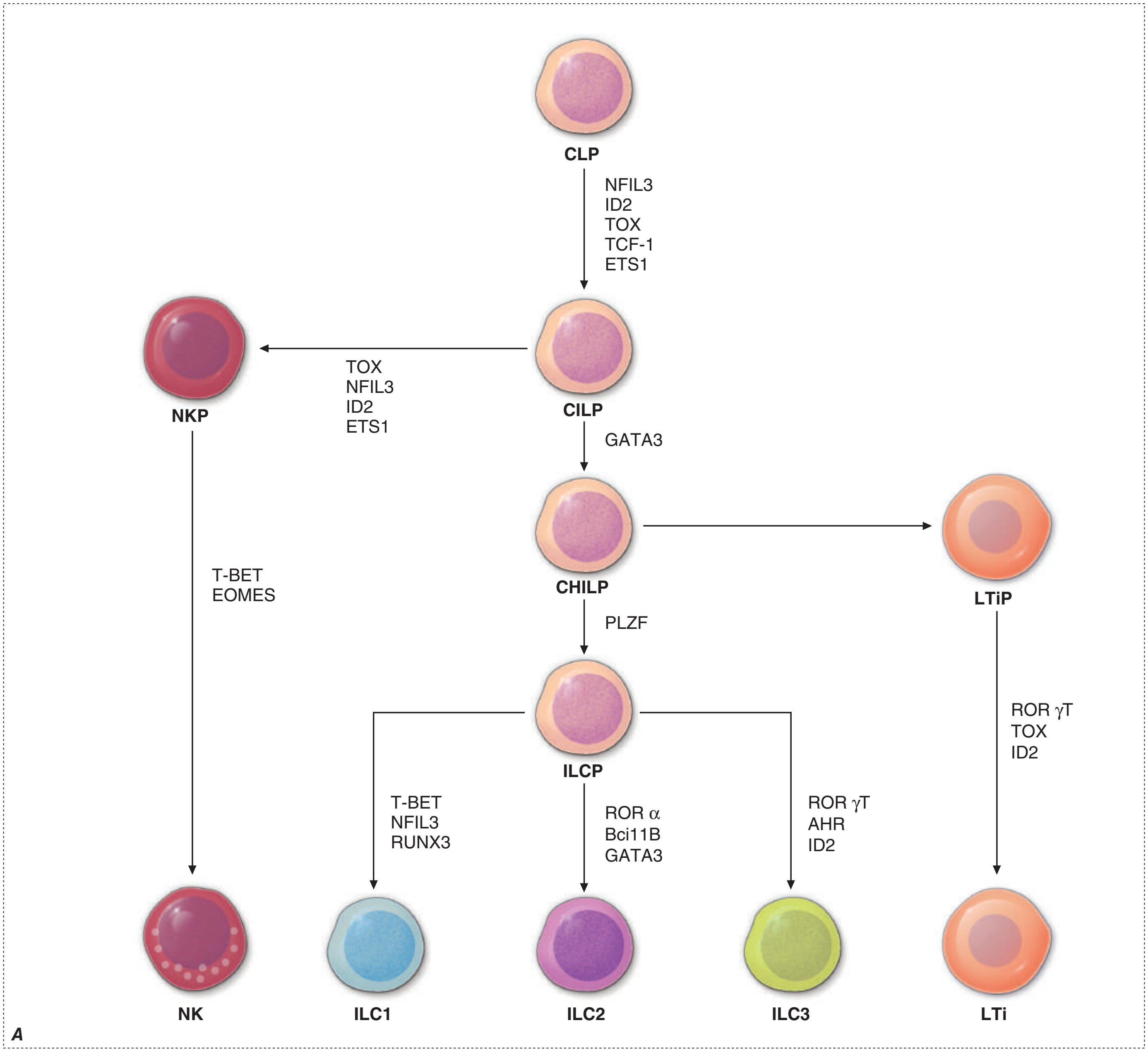

FIGURE 360-3 Development and function of innate lymphoid cells (ILCs). ***A.*** ILC development, mainly based on mouse ILC differentiation paths, is schematized. ILCs develop from common innate lymphoid progenitors (CILPs), which themselves differentiate from common lymphoid progenitors (CLPs). CILPs can differentiate into natural killer (NK) cell precursor (NKP) cells or into common helper innate lymphoid progenitors (CHILPs), which themselves give rise to lymphoid tissue inducer progenitors (LTiPs) and innate lymphoid cell precursors (ILCPs). LTiPs differentiate into lymphoid tissue inducers (LTis) and ILCPs into ILC1, ILC2, or ILC3. Each stage of differentiation is dependent on the expression of the indicated transcription factors: NFIL3 (nuclear factor IL-3 induced), Id2 (inhibitor of DNA binding 2), TOX (thymocyte selection-associated high mobility group box protein), TCF-1 (T-cell factor 1), ETS1 (avian erythroblastosis virus E26 homolog-1), GATA3 (GATA binding protein 3), PLZF (promyelocytic leukemia zinc finger), T-bet (T-box transcription factor), Eomes (eomesodermin), RUNX3 (runt-related transcription factor 3), RORα (RAR-related orphan receptor α), Bcl11b (B cell lymphoma/leukemia 11B), RORγt (RAR-related orphan receptor γt), and AHR (Aryl hydrocarbon receptor). It has been shown in humans that ILC1 subsets may originate from precursors other than ILCPs, but the identity of these precursors remains unknown at this time. ***B.*** Some of the most well-known immune functions of each ILC subset are shown: NK cells and ILC1s react to intracellular pathogens, such as viruses, and to tumors; ILC2s respond to large extracellular parasites and allergens; ILC3s combat extracellular microbes, such as bacteria and fungi; and Lutes are involved in the formation of secondary lymphoid structures. For each ILC subset, effector molecules that can be produced upon activation are indicated AREG, amphiregulin; RANK, receptor activation of nuclear factor κB; RANK-L, RANK-ligand. *(Reproduced with permission from E Vivier et al: Innate lymphoid cells: 10 years on. Cell 174:1054, 2018.)*

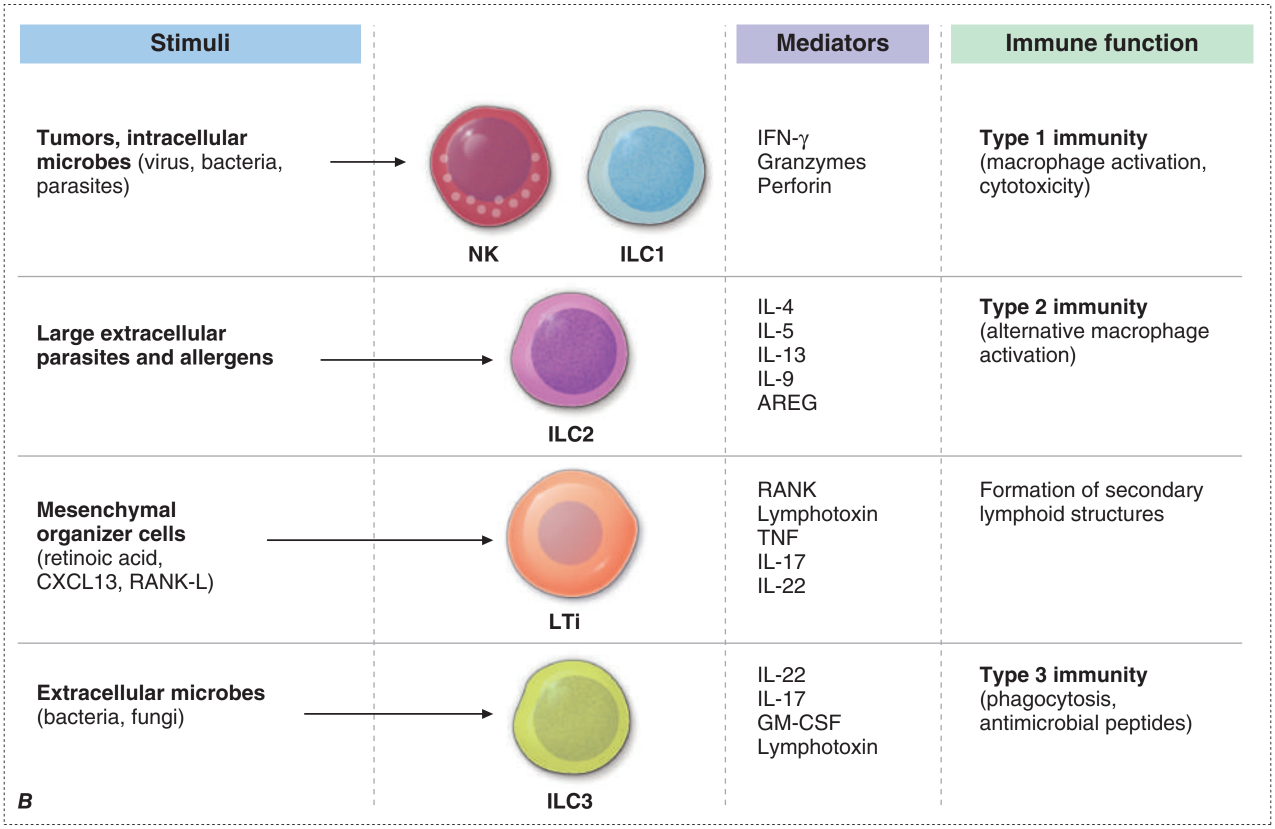
Stimuli
Mediators
Immune function
Tumors, intracellular microbes (virus, bacteria, parasites)
NK
ILC1
IFN-γ
Granzymes
Perforin
Type 1 immunity (macrophage activation, cytotoxicity)
Large extracellular parasites and allergens
ILC2
IL-4
IL-5
IL-13
IL-9
AREG
Type 2 immunity (alternative macrophage activation)
Mesenchymal organizer cells (retinoic acid, CXCL13, RANK-L)
LTi
RANK
Lymphotoxin
TNF
IL-17
IL-22
Formation of secondary lymphoid structures
Extracellular microbes (bacteria, fungi)
ILC3
IL-22
IL-17
GM-CSF
Lymphotoxin
Type 3 immunity (phagocytosis, antimicrobial peptides)
B

FIGURE 360-3 (*Continued*)

In addition, basophils express surface receptors for activated complement components (C3a, C5a), through which mediator release can be directly effected. Thus, basophils, like most cells of the immune system, can be activated in the service of host defense against pathogens, or they can be activated for mediator release and cause pathogenic responses in allergic and inflammatory diseases. **For further discussion of tissue mast cells, see Chap. 366.**

The Complement System The complement system, an important soluble component of the innate immune system, is a series of plasma enzymes, regulatory proteins, and proteins that are activated in a cascading fashion, resulting in cell lysis. There are four pathways of the complement system: the classic activation pathway activated by antigen/antibody immune complexes, the MBL (a serum collectin) activation pathway activated by microbes with terminal mannose groups, the alternative activation pathway activated by microbes or tumor cells, and the terminal pathway that is common to the first three pathways and leads to the membrane attack complex that lyses cells **(Fig. 360-5)**. The series of enzymes of the complement system are serine proteases.

Activation of the classic complement pathway via immune complex binding to C1q links the innate and adaptive immune systems via specific antibody in the immune complex. The alternative complement activation pathway is antibody-independent and is activated by binding of C3 directly to pathogens and "altered self" such as tumor cells. In the renal glomerular inflammatory disease *IgA nephropathy*, IgA activates the alternative complement pathway and causes glomerular damage and decreased renal function. Activation of the classic complement pathway via C1, C4, and C2 and activation of the alternative pathway via factor D, C3, and factor B both lead to cleavage and activation of C3. C3 activation fragments, when bound to target surfaces such as bacteria and other foreign antigens, are critical for opsonization (coating by antibody and complement) in preparation for phagocytosis. The MBL pathway substitutes MBL-associated serine proteases (MASPs) 1 and 2 for C1q, C1r, and C1s to activate C4. The MBL activation pathway is activated by mannose on the surface of bacteria and viruses.

The three pathways of complement activation all converge on the final common terminal pathway. C3 cleavage by each pathway results in activation of C5, C6, C7, C8, and C9, resulting in the membrane attack complex that physically inserts into the membranes of target cells or bacteria and lyses them.

Thus, complement activation is a critical component of innate immunity for responding to microbial infection. The functional consequences of complement activation by the three initiating pathways and the terminal pathway are shown in Fig. 360-5. In general, the cleavage products of complement components facilitate microbe or damaged cell clearance (C1q, C4, C3), promote activation and enhancement of inflammation (anaphylatoxins, C3a, C5a), and promote microbe or opsonized cell lysis (membrane attack complex).

■ CYTOKINES

Cytokines are soluble proteins produced by a wide variety of cell types (Tables 360-7 and 360-8). They are critical for both normal innate and adaptive immune responses, and their expression may be perturbed in most immune, inflammatory, and infectious disease states.

Cytokines are involved in the regulation of the growth, development, and activation of immune system cells and in the mediation of the inflammatory response. In general, cytokines are characterized by considerable redundancy in that different cytokines have similar functions. In addition, many cytokines are pleiotropic in that they are capable of acting on many different cell types. This pleiotropism results from the expression on multiple cell types of receptors for the same cytokine (see below), leading to the formation of "cytokine networks." The action of cytokines may be (1) *autocrine* when the target cell is the same cell that secretes the cytokine, (2) *paracrine* when the target cell is nearby, and (3) *endocrine* when the cytokine is secreted into the circulation and acts distal to the source.

Cytokines have been named based on presumed targets or based on presumed functions. Those cytokines that are thought to primarily target leukocytes have been named IL-1, -2, -3, etc. Many cytokines that were originally described as having a certain function have retained those names (e.g., granulocyte colony-stimulating factor [G-CSF]). Cytokines

TABLE 360-9 NK-Cell Receptors, Cognate Ligands, and Their Known Signaling Domains/Proximal Adapters

RECEPTOR	GENE	LIGAND	SIGNALING DOMAINS AND PROXIMAL ADAPTERS
Inhibitory KIR		Conserved epitopes on HLA I ("KIR ligands")	ITIM, Src phosphatases
KIR2DL1	*KIR2DL1*	HLA-C2 alleles (Lys80)	
KIR2DL2	*KIR2DL2*	HLA-C1 alleles (Asp80); HLA-C2 alleles (Lys80)	
KIR2DL3	*KIR2DL3*	HLA-C1 alleles (Asp80)	
KIR3DL1	*KIR3DL1*	HLA-B alleles carrying the Bw4 motif; HLA-A alleles carrying the Bw4 motif	
KIR3DL2	*KIR3DL2*	HLA-A*03, HLA-A*11 carrying specific peptides	
KIR2DL4	*KIR2DL4*	HLA-G	
KIR2DL5		CD155	
Activating KIR		HLA-C2, HLA-C1, HLA-F, certain configurations of HLA-peptide combinations	ITAM, SFK
KIR2DS1	*KIR2DS1*	HLA-C2 alleles (Lys80) carrying specific peptides	
KIR2DS2	*KIR2DS2*	HLA-C1 (Asp80) HLA-A*11	
KIR2DS3	*KIR2DS3*	Unknown	
KIR3DS1		HLA-F	
KIR2DS5		Unknown	
Natural cytotoxicity receptors (NCRs)			
NKp30	*NCR3*	B7-H6, BAT-3, heparan sulfates	
NKp44	*NCR2*	PDGF, heparan sulfates, PCNA	
NKp46	*NCR1*	Viral hemagglutinins, heparan sulfates, vimentin, ecto-calreticulin	
SLAM family receptors			ITSM, SAP, EAT
SLAMF1 (SLAM, CD150, IPO-3)	*SLAMF1*	SLAMF1	
SLAMF2 (CD48, BLAST-1)	*CD48*	SLAMF4, CD2	
SLAMF3 (CD229, Ly9)	*LY9*	SLAMF3	
SLAMF4 (CD244, 2B4, ERT)	*CD244*	SLAMF2	
SLAMF6 (NTB-A, Ly108, CD352)	*SLAMF6*	SLAMF6	
SLAMF7 (CRACC, CD319)	*SLAMF7*	SLAMF7	
SLAMF8 (BLAME, CD353)	*SLAMF8*	Unknown	

Abbreviations: BAT-3, human leukocyte antigen-B-associated transcript 3; EAT, Ewing sarcoma associated transcript; HLA, human leukocyte antigens; ITAM, immunoreceptor tyrosine-based activation motif; ITIM, immunoreceptor tyrosine-based inhibitory motif; ITSM, immunoreceptor tyrosine-based switch motif; KIR, killer cell immunoglobulin-like receptor; NCRs, natural cytotoxicity receptors; NK, natural killer cell; SAP, SLAM-associated protein; SFK, Src family kinase; SLAM, signalling lymphocyte activating molecule.

Source: Reproduced from S Nersesian et al: Killer instincts: Natural killer cells as multifactorial cancer immunotherapy. Front Immunol. 2023; 14:1269614.

TABLE 360-10 Association of KIRS with Disease

DISEASE	KIR ASSOCIATION	OBSERVATION
Psoriatic arthritis	KIR2DS1/KIR2DS2; HLA-Cw group homozygosity	Susceptibility
Spondylarthritides	Increased KIR3DL2 expression Interaction of HLA-B27 homodimers with KIR3DL1/KIR3DL2; independent of peptide	May contribute to disease pathology May contribute to disease pathogenesis
Ankylosing spondylitis	KIR3DL1/3DS1; HLA-B27 genotypes	Susceptibility
Rheumatoid vasculitis	KIR2DS2; HLA-Cw*03 Increased KIR2L2/2DS2 in patients with extraarticular manifestations	Susceptibility Clinical manifestations may have different genetic backgrounds with respect to KIR genotype
Rheumatoid arthritis	Decreased KIR2DS1/3DS1 in patients without bone erosions KIR2DS4; HLA-Cw4	Susceptibility Susceptibility
Scleroderma	KIR2DS2+/KIR2DL2–	Susceptibility
Behçet's disease	Altered KIR3DL1 expression	Associated with severe eye disease
Psoriasis vulgaris	2DS1; HLA-Cw*06 2DS1; 2DL5; haplotype B	Susceptibility Susceptibility
IDDM	KIR2DS2; HLA-C1	Susceptibility
Type 1 diabetes	KIR2DS2; HLA-C1 and no HLA-C2, no HLA-Bw4	Increased disease progression
Preeclampsia	KIR2DL1 with fewer KIR2DS (mother); HLA-C2 (fetus)	Increased disease progression
AIDS	KIR3DS1; HLA-Bw4Ile80 KIR3DS1 homozygous; no HLA-Bw4Ile80	Decreased disease progression Increased disease progression
HCV infection	KIR2DL3 homozygous; HLA-C1 homozygous	Decreased disease progression
Cervical neoplasia (HPV induced)	KIR3DS1; HLA-C1 homozygous and no HLA-Bw4	Increased disease progression
Malignant melanoma	KIR2DL2 and/or KIR2DL3; HLA-C1	Increased disease progression

Abbreviations: HCV, hepatitis C virus; HLA, human leukocyte antigen; HPV, human papillomavirus; IDDM, insulin-dependent diabetes mellitus; KIR, killer cell immunoglobulin-like receptor.

Source: Reproduced with permission from R Diaz-Pena et al: KIR genes and their role in spondyloarthropathies. Adv Exp Med Biol 649:286, 2009.

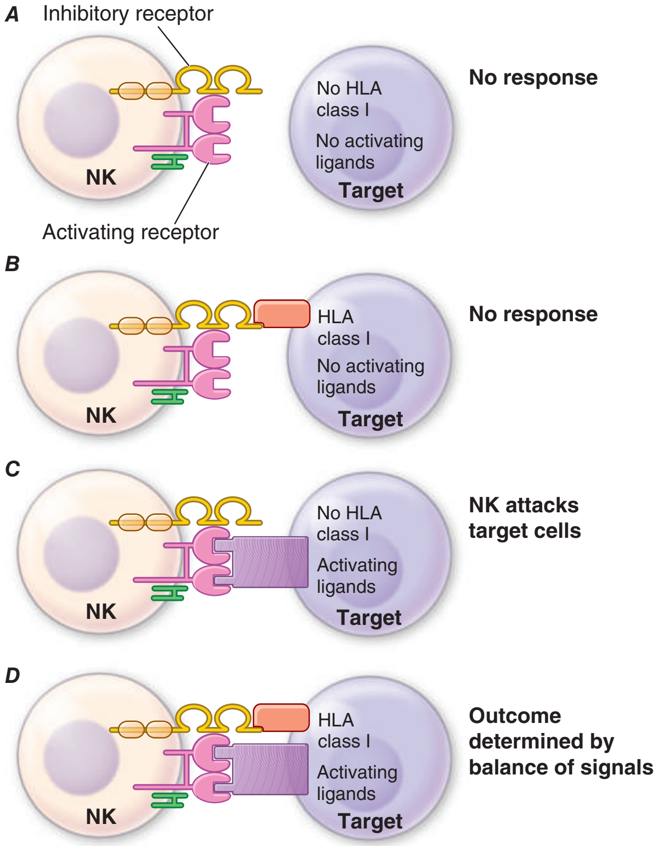
A
Inhibitory receptor
NK
No HLA
class I
No activating
ligands
Target
No response
Activating receptor
B
NK
HLA
class I
No activating
ligands
Target
No response
C
NK
No HLA
class I
Activating
ligands
Target
NK attacks
target cells
D
NK
HLA
class I
Activating
ligands
Target
Outcome
determined by
balance of signals

FIGURE 360-4 Encounters between natural killer (NK) cells: Potential targets and possible outcomes. The amount of activating and inhibitory receptors on the NK cells and the amount of ligands on the target cell, as well as the qualitative differences in the signals transduced, determine the extent of the NK response. ***A.*** When target cells have no HLA class I or activating ligands, NK cells cannot kill target cells. ***B.*** When target cells bear self-HLA, NK cells cannot kill targets. ***C.*** When target cells are pathogen-infected and have downregulated HLA and express activating ligands, NK cells kill target cells. ***D.*** When NK cells encounter targets with both self-HLA and activating receptors, then the level of target killing is determined by the balance of inhibitory and activating signals to the NK cell. HLA, human leukocyte antigen. *(Republished with permission of Annual Review of Immunology, from NK Cell Recognition, L Lanier 23:225,2005: permission conveyed through Copyright Clearance Center, Inc.)*

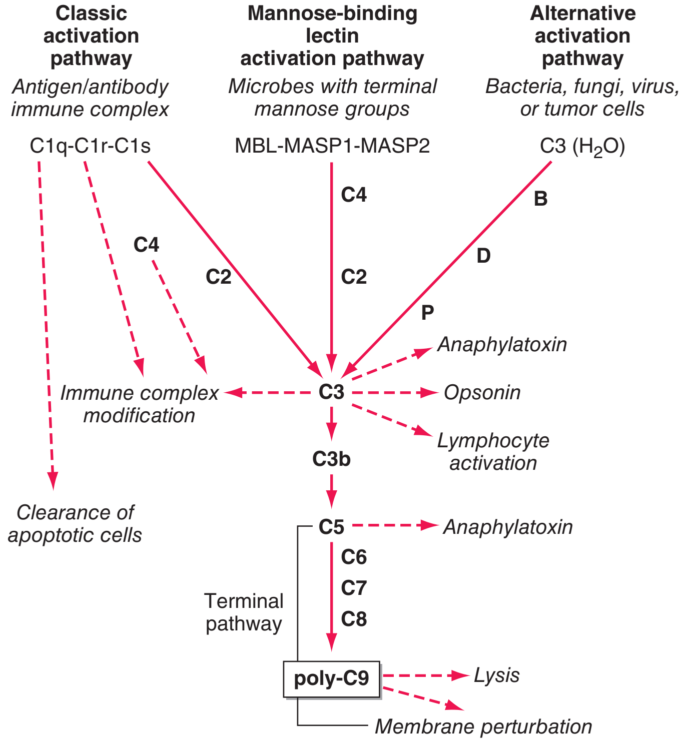
Classic
activation
pathway
Antigen/antibody
immune complex
C1q-C1r-C1s
Mannose-binding
lectin
activation pathway
Microbes with terminal
mannose groups
MBL-MASP1-MASP2
Alternative
activation
pathway
Bacteria, fungi, virus,
or tumor cells
C3 (H2O)
C4
C4
C2
C2
B
D
P
Anaphylatoxin
Immune complex
modification
C3
Opsonin
Lymphocyte
activation
C3b
Clearance of
apoptotic cells
C5
Anaphylatoxin
C6
C7
C8
Terminal
pathway
poly-C9
Lysis
Membrane perturbation

FIGURE 360-5 The four pathways and the effector mechanisms of the complement system. *Dashed arrows* indicate the functions of pathway components. *(Reproduced with permission from BJ Morley, MJ Walport: The Complement Facts Books. London, Academic Press, 2000.)*

belong in general to three major structural families: the hematopoietin family; the TNF, IL-1, platelet-derived growth factor (PDGF), and transforming growth factor (TGF) β families; and the CXC and C-C chemokine families. Chemokines are cytokines that regulate cell movement and trafficking; they act through G protein–coupled receptors and have a distinctive three-dimensional structure (Table 360-7).

In general, cytokines exert their effects by influencing gene activation that results in cellular activation, growth, differentiation, functional cell-surface molecule expression, and cellular effector function. In this regard, cytokines can have dramatic effects on the regulation of immune responses and the pathogenesis of a variety of diseases. Indeed, T cells have been categorized on the basis of the pattern of cytokines that they secrete, which results in either humoral immune response (T_H2) or cell-mediated immune response (T_H1). A third type of T helper cell is the T_H17 cell that contributes to host defense against extracellular bacteria and fungi, particularly at mucosal sites (Fig. 360-2).

Cytokine receptors can be grouped into five general families based on similarities in their extracellular amino acid sequences and conserved structural domains. The *immunoglobulin (Ig) superfamily* represents a large number of cell-surface and secreted proteins. The IL-1 receptors (type 1, type 2) are examples of cytokine receptors with extracellular Ig domains.

The hallmark of the *hematopoietic growth factor (type 1) receptor* family is that the extracellular regions of each receptor contain two conserved motifs. One motif, located at the N terminus, is rich in cysteine residues. The other motif is located at the C terminus proximal to the transmembrane region and comprises five amino acid residues, tryptophan-serine-X-tryptophan-serine (WSXWS). This family can be grouped on the basis of the number of receptor subunits they have and on the utilization of shared subunits. A number of cytokine receptors, i.e., IL-6, IL-11, IL-12, and leukemia inhibitory factor, are paired with gp130. There is also a common 150-kDa subunit shared by IL-3, IL-5, and granulocyte-macrophage colony-stimulating factor (GM-CSF) receptors. The gamma chain (γ_c) of the IL-2 receptor is common to the IL-2, IL-4, IL-7, IL-9, and IL-15 receptors. Thus, the specific cytokine receptor is responsible for ligand-specific binding, whereas the subunits such as gp130, the 150-kDa subunit, and γ_c are important in signal transduction. The γ_c gene is on the X chromosome, and mutations in the γ_c protein result in the *X-linked form of severe combined immune deficiency syndrome (X-SCID)* **(Chap. 362)**.

The members of the *interferon (type II) receptor* family include the receptors for IFN-γ and -β, which share a similar 210-amino-acid binding domain with conserved cysteine pairs at both the amino and carboxy termini. The members of the *TNF (type III) receptor family* share a common binding domain composed of repeated cysteine-rich regions. Members of this family include the p55 and p75 receptors for TNF (TNF-R1 and TNF-R2, respectively); CD40 antigen, which is an important B-cell surface marker involved in immunoglobulin isotype switching; fas/Apo-1, whose triggering induces apoptosis; CD27 and CD30, which are found on activated T cells and B cells; and nerve growth factor receptor.

The common motif for the *seven transmembrane helix family* was originally found in receptors linked to GTP-binding proteins. This family includes receptors for chemokines (Table 360-8), β-adrenergic receptors, and retinal rhodopsin. It is important to note that two members of the chemokine receptor family, CXC chemokine receptor type 4 (CXCR4) and β chemokine receptor type 5 (CCR5), have been found to serve as the two major co-receptors for binding and entry of HIV-1 into CD4-expressing host cells **(Chap. 208)**.

Significant advances have been made in defining the signaling pathways through which cytokines exert their intracellular effects. The Janus family of protein tyrosine kinases (JAK) is a critical element involved in signaling via the hematopoietin receptors. Four JAK kinases, JAK1, JAK2, JAK3, and Tyk2, preferentially bind different cytokine receptor subunits. Cytokine binding to its receptor brings the cytokine receptor subunits into apposition and allows a pair of JAKs to transphosphorylate and activate one another. The JAKs then phosphorylate the receptor on the tyrosine residues and allow signaling molecules to bind to the receptor, whereby the signaling molecules become phosphorylated. Signaling molecules bind the receptor because they have domains (SH2, or src homology 2 domains) that can bind phosphorylated tyrosine residues. There are a number of these important signaling molecules that bind the receptor, such as the adapter molecule SHC, which can couple the receptor to the activation of the mitogen-activated protein kinase pathway. In addition, an important class of substrate of the JAKs is the signal transducers and activators of transcription (STAT) family of transcription factors. STATs have SH2 domains that enable them to bind to phosphorylated receptors, where they are then phosphorylated by the JAKs. It appears that different STATs have specificity for different receptor subunits. The STATs then dissociate from the receptor and translocate to the nucleus, bind to DNA motifs that they recognize, and regulate gene expression. The STATs preferentially bind DNA motifs that are slightly different from one another and thereby control transcription of specific genes. The importance of this pathway is particularly relevant to lymphoid development. Mutations of JAK3 itself also result in a disorder identical to X-SCID; however, because JAK3 is found on chromosome 19 and not on the X chromosome, JAK3 deficiency occurs in boys and girls **(Chap. 362)**. A new class of immunosuppressive drugs of JAK inhibitors has been developed that are approved to treat chronic inflammatory diseases such as rheumatoid arthritis **(Chap. 370)**, psoriatic arthritis **(Chap. 374)**, ulcerative colitis **(Chap. 337)**, atopic dermatitis **(Chap. 59)**, and alopecia areata **(Chap. 61)**.

■ THE ADAPTIVE IMMUNE SYSTEM

Adaptive immunity is characterized by antigen-specific responses to a foreign antigen or pathogen. A key feature of adaptive immunity is that following the initial contact with antigen (*immunologic priming*), subsequent antigen exposure leads to more rapid and vigorous immune responses (*immunologic memory*). The adaptive immune system consists of dual limbs of cellular and humoral immunity. The principal effectors of cellular immunity are T lymphocytes, whereas the principal effectors of humoral immunity are B lymphocytes. Both B and T lymphocytes derive from a common stem cell **(Fig. 360-6)**.

The proportion and distribution of immunocompetent cells in various tissues reflect cell traffic, homing patterns, and functional capabilities. Bone marrow is the major site of maturation of B cells, monocytes-macrophages, DCs, and granulocytes and contains pluripotent stem cells that, under the influence of various colony-stimulating factors, can give rise to all hematopoietic cell types. T-cell precursors also arise from hematopoietic stem cells and home to the thymus for maturation. Mature T lymphocytes, B lymphocytes, monocytes, and DCs enter the circulation and home to peripheral lymphoid organs (lymph nodes, spleen) and mucosal surface-associated lymphoid tissue (gut, genitourinary, and respiratory tracts) as well as the skin and mucous membranes and await activation by foreign antigen.

T Cells The pool of effector T cells is established in the thymus early in life and is maintained throughout life both by new T-cell production in the thymus and by antigen-driven expansion of virgin peripheral T cells into "memory" T cells that reside in peripheral lymphoid organs. The thymus exports ~2% of the total number of thymocytes per day throughout life, with the total number of daily thymic emigrants decreasing by ~3% per year during the first four decades of life.

Mature T lymphocytes constitute 70–80% of normal peripheral blood lymphocytes (only 2% of the total-body lymphocytes are contained in peripheral blood), 90% of thoracic duct lymphocytes, 30–40% of lymph node cells, and 20–30% of spleen lymphoid cells. In lymph nodes, T cells occupy deep paracortical areas around B-cell germinal centers, and in the spleen, they are located in periarteriolar areas of white pulp **(Chap. 70)**. T cells are the primary effectors of cell-mediated immunity, with subsets of T cells maturing into CD8+ cytotoxic T cells capable of lysis of virus-infected or foreign cells (short-lived effector T cells) and CD4+ T cells capable of T-cell help for CD8+ T-cell and B-cell development. Two populations of long-lived memory T cells are triggered by infections: effector memory and central memory T cells. Effector memory T cells reside in nonlymphoid organs and respond rapidly to repeated pathogenic infections with cytokine production and cytotoxic functions to kill virus-infected cells. Central memory T cells home to lymphoid organs where they replenish long- and short-lived and effector memory T cells as needed.

In general, CD4+ T cells are the primary regulatory cells of T and B lymphocyte and monocyte function by the production of cytokines and by direct cell contact (Fig. 360-2). In addition, T cells regulate erythroid cell maturation in bone marrow and, through cell contact (CD40 ligand), have an important role in activation of B cells and induction of Ig isotype switching. Considerable evidence now exists that colonization of the gut by commensal bacteria (the gut microbiome) is responsible for expansion of the peripheral CD4+ T-cell compartment in normal children and adults.

Human T cells express cell-surface proteins that mark stages of intrathymic T-cell maturation or identify specific functional subpopulations of mature T cells. Many of these molecules mediate or participate in important T-cell functions (Table 360-1, Fig. 360-6, **Chap. 361**).

The earliest identifiable T-cell precursors in bone marrow are CD34+ pro-T cells (i.e., cells in which TCR genes are neither rearranged nor expressed). In the thymus, CD34+ T-cell precursors begin cytoplasmic (c) synthesis of components of the CD3 complex of TCR-associated molecules (Fig. 360-6). Within T-cell precursors, TCR for antigen gene rearrangement yields two T-cell lineages, expressing either TCR-αβ chains or TCR-γδ chains. T cells expressing the TCR-αβ chains constitute the majority of peripheral T cells in blood, lymph node, and spleen and terminally differentiate into either CD4+ or CD8+ cells. Cells expressing TCR-γδ chains circulate as a minor population in blood; their functions have been postulated to be those of immune surveillance at epithelial surfaces and cellular defenses against mycobacterial organisms and other intracellular bacteria through recognition of bacterial lipids.

In the thymus, the recognition of self-peptides on thymic epithelial cells, thymic macrophages, and DCs plays an important role in shaping T-cell repertoire. As immature cortical thymocytes begin to express surface TCR for antigen, thymocytes with TCRs capable of interacting with self-peptides in the context of self-MHC antigens with low affinity are activated and survive (positive selection). Thymocytes with TCRs that are incapable of binding to self-MHC antigens or bind with high affinity die of attrition (no selection) or by apoptosis (negative selection). Thymocytes that are positively selected undergo maturation into CD4 or CD8 single positive T cells, and then migrate to the thymus medulla where they interact with self-peptide–self-MHC molecules, where they can again undergo selection. The purpose of negative and positive thymocyte selection is to eliminate potential pathogenic autoreactive T cells, and at the same time, select a repertoire of mature T cells capable of recognizing foreign antigens.

Mature TCRαβ thymocytes that are positively selected are functional MHC class II–restricted CD4+ T cells (Fig. 360-2), or they are CD8+ T cells destined to become CD8+ MHC class I–restricted cytotoxic T cells. *MHC class I* or *class II restriction* means that T cells recognize antigen peptide fragments only when they are presented in the antigen-recognition site of a class I or class II MHC molecule, respectively. After thymocyte maturation and selection, CD4 and CD8 thymocytes leave the thymus and migrate to the peripheral immune system. The thymus can continue to be a contributor to the peripheral immune system well into adult life, both normally and when the peripheral T-cell pool is damaged, such as occurs in AIDS and cancer chemotherapy.

MOLECULAR BASIS OF T-CELL RECOGNITION OF ANTIGEN The TCR for antigen is a complex of molecules consisting of an antigen-binding heterodimer of either αβ or γδ chains noncovalently linked with five CD3 subunits (γ, δ, ε, ζ, and η) **(Fig. 360-7)**. The CD3 ζ chains are either disulfide-linked homodimers (CD3-ζ_2) or disulfide-linked

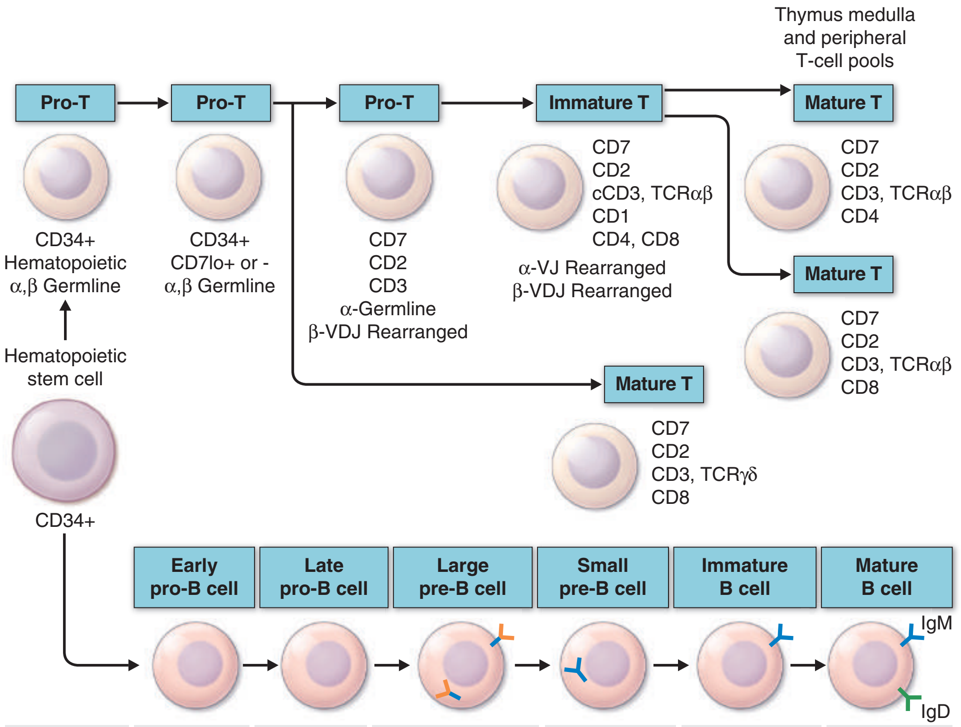

	Early pro-B cell	Late pro-B cell	Large pre-B cell	Small pre-B cell	Immature B cell	Mature B cell
Heavy-chain genes	D-J rearranging	VDJ rearranging	VDJ rearranged	VDJ rearranged	VDJ rearranged	VDJ rearranged
Light-chain genes	Germline	Germline	Germline	VJ rearranging	VJ rearranged	VJ rearranged
Surface Ig	Absent	Absent	μ H-chain at surface as part of the pre-B receptor (orange) containing surrogate light chain (SLC). Receptor is mainly intra-cellular	μ H-chain in cytoplasm	IgM expressed on cell surface	IgD and IgM made from alternatively spliced H-chain transcripts
Surface marker proteins	CD34 CD10 CD38	CD10 CD19 CD38 CD40	CD19 CD20 CD38	CD19 CD20 CD38	CD19 CD20	CD19 CD20 CD21

FIGURE 360-6 Development stages of T and B cells. Elements of the developing T- and B-cell receptor for antigen are shown schematically. The classification into the various stages of B-cell development is primarily defined by rearrangement of the immunoglobulin (Ig) heavy (H) and light (L) chain genes and by the absence or presence of specific surface markers. The classification of stages of T-cell development is primarily defined by cell-surface marker protein expression (sCD3, surface CD3 expression; cCD3, cytoplasmic CD3 expression; TCR, T-cell receptor). For B-cell development, the pre-B-cell receptor is shown as a blue-orange B-cell receptor. *(Adapted from Janeway's Immunobiology, 9th ed by Kenneth Murphy and Casey Weaver. Copyright © 2017 by Garland Science, Taylor & Francis Group, LLC. Used by permission of W. W. Norton & Company, Inc.)*

heterodimers composed of one ζ chain and one η chain. TCR-αβ or TCR-γδ molecules must be associated with CD3 molecules to be inserted into the T-cell surface membrane, TCR-α being paired with TCR-β and TCR-γ being paired with TCR-δ. Molecules of the CD3 complex mediate transduction of T-cell activation signals via TCRs, whereas TCR-α and -β or -γ and -δ molecules combine to form the TCR antigen-binding site.

The α, β, γ, and δ TCR for antigen molecules have amino acid sequence homology and structural similarities to immunoglobulin heavy and light chains and are members of the *immunoglobulin gene superfamily* of molecules. The genes encoding TCR molecules are encoded as clusters of gene segments that rearrange during T-cell maturation. This creates an efficient and compact mechanism for housing the diversity requirements of antigen receptor molecules. The TCR-α chain is on chromosome 14 and consists of a series of V (variable), J (joining), and C (constant) regions. The TCR-β chain is on chromosome 7 and consists of multiple V, D (diversity), J, and C TCR-β loci. The TCR-γ chain is on chromosome 7, and the TCR-δ chain is in the middle of the TCR-α locus on chromosome 14. Thus, molecules of the TCR for antigen have constant (framework) and variable regions, and the gene segments encoding the α, β, γ, and δ chains of these molecules are recombined and selected in the thymus, culminating in synthesis of the completed molecule. In both T- and B-cell precursors (see below), DNA rearrangements of antigen receptor genes involve the same enzymes, recombinase activating gene RAG1 and RAG2, both DNA-dependent protein kinases.

TCR diversity is created by the different V, D, and J segments that are possible for each receptor chain by the many permutations of V, D, and J segment combinations, by "N-region diversification" due to the addition of nucleotides at the junction of rearranged gene segments, and by the pairing of individual chains to form a TCR dimer. As T cells mature in the thymus, the repertoire of antigen-reactive T cells is modified by selection processes that eliminate many autoreactive T cells, enhance the proliferation of cells that function appropriately with self-MHC molecules and antigen, and allow T cells with nonproductive TCR rearrangements to die.

TCR-αβ cells do not recognize native protein or carbohydrate antigens. Instead, T cells recognize only short (~9–13 amino acids) peptide fragments derived from protein antigens taken up or produced in APCs. Foreign antigens may be taken up by endocytosis into acidified

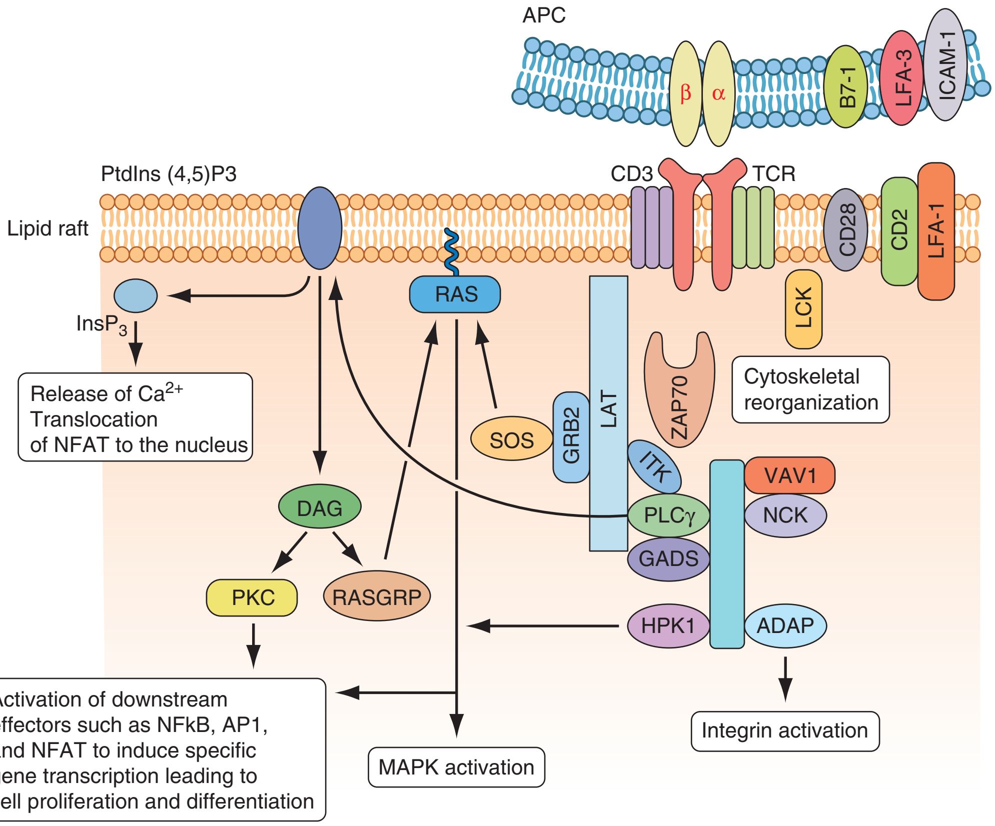

FIGURE 360-7 Signaling through the T-cell receptor. Activation signals are mediated via immunoreceptor tyrosine-based activation (ITAM) sequences in LAT and CD3 chains (*blue bars*) that bind to enzymes and transduce activation signals to the nucleus via the indicated intracellular activation pathways. Ligation of the T-cell receptor (TCR) by MHC complexed with antigen results in sequential activation of LCK and γ-chain-associated protein kinase of 70 kDa (ZAP70). ZAP70 phosphorylates several downstream targets, including LAT (linker for activation of T cells) and SLP76 (SCR homology 2 [SH2] domain-containing leukocyte protein of 76 kDa). SLP76 is recruited to membrane-bound LAT through its constitutive interaction with GADS (GRB2-related adaptor protein). Together, SLP76 and LAT nucleate a multimolecular signaling complex, which induces a host of downstream responses, including calcium flux, mitogen-activated protein kinase (MAPK) activation, integrin activation, and cytoskeletal reorganization. APC, antigen-presenting cell; NFAT, nuclear factor of activated T cells. *(Reproduced with permission from GA Koretzky, F Abtahian, MA Silverman. SLP76 and SLP65: Complex regulation of signalling in lymphocytes and beyond. Nat Rev Immunol 6:67, 2006.)*

intracellular vesicles or by phagocytosis and degraded into small peptides that associate with MHC class II molecules (exogenous antigen-presentation pathway). Other foreign antigens arise endogenously in the cytosol (such as from replicating viruses) and are broken down into small peptides that associate with MHC class I molecules (endogenous antigen-presenting pathway). Thus, APCs proteolytically degrade foreign proteins and display peptide fragments embedded in the MHC class I or II antigen-recognition site on the MHC molecule surface, where foreign peptide fragments are available to bind to TCR-αβ or TCR-γδ chains of reactive T cells. CD4 molecules act as adhesives and, by direct binding to MHC class II (DR, DQ, or DP) molecules, stabilize the interaction of TCR with peptide antigen (Fig. 360-7). Similarly, CD8 molecules also act as adhesives to stabilize the TCR-antigen interaction by direct CD8 molecule binding to MHC class Ia (HLA A, B, or C) molecules or to MHC class Ib (HLA E).

Antigens that arise in the cytosol and are processed via the endogenous antigen-presentation pathway are cleaved into small peptides by a complex of proteases called the *proteasome*. From the proteasome, antigen peptide fragments are transported from the cytosol into the lumen of the endoplasmic reticulum by a heterodimeric complex termed *transporters associated with antigen processing* or TAP proteins. There, MHC class I molecules in the endoplasmic reticulum membrane physically associate with processed cytosolic peptides. Following peptide association with class I molecules, peptide–class I complexes are exported to the Golgi apparatus, and then to the cell surface, for recognition by CD8+ T cells.

Antigens taken up from the extracellular space via endocytosis into intracellular acidified vesicles are degraded by vesicle proteases into peptide fragments. Intracellular vesicles containing MHC class II molecules fuse with peptide-containing vesicles, thus allowing peptide fragments to physically bind to MHC class II molecules. Peptide–MHC class II complexes are then transported to the cell surface for recognition by CD4+ T cells.

Whereas it is generally agreed that the TCR-αβ receptor recognizes peptide antigens in the context of MHC class I or class II molecules, lipids in the cell wall of intracellular bacteria such as *M. tuberculosis* can also be presented to a wide variety of T cells, including subsets of TCR-γδ T cells, and a subset of CD8+ TCR-αβ T cells. Importantly, bacterial lipid antigens are not presented in the context of MHC class I or II molecules, but rather are presented in the context of MHC-related CD1 molecules. Some γδ T cells that recognize lipid antigens via CD1 molecules have very restricted TCR usage, do not need antigen priming to respond to bacterial lipids, and may be a form of innate rather than acquired immunity to intracellular bacteria.

Just as foreign antigens are degraded and their peptide fragments presented in the context of MHC class I or class II molecules on APCs, endogenous self-proteins also are degraded, and self-peptide fragments are presented to T cells in the context of MHC class I or class II molecules on APCs. In peripheral lymphoid organs, there are T cells that are capable of recognizing self-protein fragments but normally are *anergic* or *tolerant*, i.e., nonresponsive to self-antigenic stimulation, due to lack of self-antigen upregulating APC *co-stimulatory molecules* such as B7-1 (CD80) and B7-2 (CD86) (see below and **Chap. 361**).

Once engagement of mature T-cell TCR by foreign peptide occurs in the context of self-MHC class Ia (A, B, or C), class 1b (E), or class II molecules, binding of non-antigen-specific adhesion ligand pairs such as CD54-CD11/CD18 and CD58-CD2 stabilizes MHC peptide-TCR binding, and the expression of these adhesion molecules is upregulated. Once antigen ligation of the TCR occurs, the T-cell membrane is partitioned into *lipid membrane microdomains*, or *lipid rafts*, that coalesce the key signaling molecules TCR/CD3 complex, CD28, CD2, LAT (linker for activation of T cells), intracellular activated (dephosphorylated) src family protein tyrosine kinases (PTKs), and the key CD3ζ-associated protein-70 (ZAP-70) PTK (Fig. 360-7). Importantly, during T-cell activation, the CD45 molecule, with protein tyrosine phosphatase activity, is partitioned away from the TCR complex to allow activating phosphorylation events to occur. The coalescence of signaling molecules of activated T lymphocytes in *microdomains* has suggested that T cell–APC interactions can be considered *immunologic synapses*, analogous in function to neuronal synapses.

After TCR-MHC binding is stabilized, activation signals are transmitted through the cell to the nucleus and lead to the expression of gene products important in mediating the wide diversity of T-cell functions such as the secretion of IL-2. The TCR does not have intrinsic signaling activity but is linked to a variety of signaling pathways via ITAMs expressed on the various CD3 chains that bind to proteins that mediate signal transduction. Each of the pathways results in the activation of particular transcription factors that control the expression of cytokine and cytokine receptor genes. Thus, antigen-MHC binding to the TCR induces the activation of the src family of PTKs, Fyn and Lck (Lck is associated with CD4 or CD8 co-stimulatory molecules); phosphorylation of CD3ζ chain; activation of the related tyrosine kinases ZAP-70 and Syk; and downstream activation of the calcium-dependent calcineurin pathway, the ras pathway, and the protein kinase C pathway. Each of these pathways leads to activation of specific families of transcription factors (including *NF-AT*, *fos* and *jun*, and *rel/NF-κB*) that form heteromultimers capable of inducing expression of IL-2, IL-2 receptor, IL-4, TNF-α, and other T-cell mediators.

In addition to the signals delivered to the T cell from the TCR complex and CD4 and CD8, molecules on the T cell, such as CD28 and inducible co-stimulator (ICOS), and molecules on DCs, such as B7-1 (CD80) and B7-2 (CD86), also deliver important co-stimulatory signals that upregulate T-cell cytokine production and are essential for T-cell activation. If signaling through CD28 or ICOS does not occur, or if CD28 is blocked, the T cell becomes anergic rather than activated (see "Immune Tolerance and Autoimmunity" below and **Chap. 361**). CTLA-4 (CD152) is similar to CD28 in its ability to bind CD80 and CD86. Unlike CD28, CTLA-4 transmits an inhibitory signal to T cells, acting as an off switch.

T-CELL EXHAUSTION IN VIRAL INFECTIONS AND CANCER In chronic viral infections such as HIV-1, hepatitis C virus, and hepatitis B virus and in chronic malignancies, the persistence of antigen disrupts memory T-cell function, resulting in defects in memory T-cell responses. This has been defined as *T-cell exhaustion* and is associated with T-cell programmed cell death protein 1 (PD-1) (CD279) expression. Exhausted T cells have compromised proliferation and lose the ability to produce effector molecules, like IL-2, TNF-α, and IFN-γ. PD-1 and CTLA-4 activation downregulates T-cell responses and is associated with T-cell exhaustion and tumor progression. Inhibition of T-cell PD-1 or CTLA-4 activity to enhance effector T-cell killing of tumor cells has become a critical component of therapy for certain malignancies **(Chap. 361)**.

T-CELL SUPERANTIGENS Conventional antigens bind to MHC class I or II molecules in the groove of the αβ heterodimer and bind to T cells via the V regions of the TCR-α and -β chains. In contrast, superantigens bind directly to the lateral portion of the TCR-β chain and MHC class II β chain and stimulate T cells based solely on the Vβ gene segment used independent of the D, J, and Vα sequences present. *Superantigens* are protein molecules capable of activating up to 20% of the peripheral T-cell pool, whereas conventional antigens activate <1 in 10,000 T cells. T-cell superantigens include staphylococcal enterotoxins and other bacterial products. Superantigen stimulation of human peripheral T cells occurs in the clinical setting of *staphylococcal toxic shock syndrome*, leading to massive overproduction of T-cell cytokines that leads to hypotension and shock **(Chap. 152)**.

B CELLS Mature B cells constitute 5–10% of human peripheral blood lymphocytes, 20–30% of lymph node cells, 50% of splenic lymphocytes, and ~10% of bone marrow lymphocytes. B cells express on their surface intramembrane immunoglobulin (Ig) molecules that function as BCRs for antigen in a complex of Ig-associated α and β signaling molecules with properties similar to those described in T cells **(Fig. 360-8)**. Unlike T cells, which recognize only processed peptide fragments of conventional antigens embedded in the notches of MHC class I and class II antigens of APCs, B cells are capable of recognizing and proliferating to whole unprocessed native antigens via antigen binding to B-cell surface Ig (sIg) receptors. B cells also express surface receptors for the Fc region of IgG molecules (CD32) as well as receptors for activated complement components (C3d or CD21, C3b or CD35). The primary function of B cells is to produce antibodies. B cells also serve as APCs and are highly efficient at antigen processing. Their antigen-presenting function is enhanced by a variety of cytokines. Mature B cells are derived from bone marrow precursor cells that arise continuously throughout life (Fig. 360-6).

B lymphocyte development can be separated into antigen-independent and antigen-dependent phases. Antigen-independent B-cell development occurs in primary lymphoid organs and includes all stages of B-cell maturation up to the sIg+ mature B cell. Antigen-dependent B-cell maturation is driven by the interaction of antigen with the mature B-cell sIg, leading to memory B-cell induction, Ig class switching, and plasma cell formation. Antigen-dependent stages of B-cell maturation occur in secondary lymphoid organs, including lymph node, spleen, and gut Peyer's patches. In contrast to the T-cell repertoire that is generated intrathymically before contact with foreign antigen, the repertoire of B cells expressing diverse antigen-reactive sites is modified by further alteration of Ig genes by the enzyme activation-induced cytidine deaminase after stimulation by antigen—a process called *somatic hypermutation*—that occurs in lymph node germinal centers.

During B-cell development, diversity of the antigen-binding variable region of Ig is generated by an ordered set of Ig gene rearrangements that are similar to the rearrangements undergone by TCR α, β, γ, and δ genes. For the heavy chain, there is first a rearrangement of D segments to J segments, followed by a second rearrangement between a V gene segment and the newly formed D-J sequence; the C segment is aligned to the V-D-J complex to yield a functional Ig heavy chain gene (V-D-J-C). During later stages, a functional κ or γ light chain gene is generated by rearrangement of a V segment to a J segment, ultimately yielding an intact Ig molecule composed of heavy and light chains.

The process of Ig gene rearrangement is regulated and results in a single antibody specificity produced by each B cell, with each Ig molecule comprising one type of heavy chain and one type of light chain. Although each B cell contains two copies of Ig light and heavy chain genes, only one gene of each type is productively rearranged and expressed in each B cell, a process termed *allelic exclusion*.

There are ~300 V_κ genes and 5 J_κ genes, resulting in the pairing of V_κ and J_κ genes to create >1500 different kappa light chain combinations. There are ~70 V_λ genes and 4 J_λ genes for >280 different lambda light chain combinations. The number of distinct light chains that can be generated is increased by somatic mutations within the V and J genes, thus creating large numbers of possible specificities from a limited amount of germline genetic information. As noted above, in heavy chain Ig gene rearrangement, the VH domain is created by the joining of three types of germline genes called V_H, D_H, and J_H, thus allowing for even greater diversity in the variable region of heavy chains than of light chains.

The most immature B-cell precursors (early pro-B cells) lack cytoplasmic Ig (cIg) and sIg (Fig. 360-6). The large pre-B cell is marked by the acquisition of the surface pre-BCR composed of μ heavy (H) chains and a pre-B light chain, termed *V pre-B*. V pre-B is a surrogate light chain receptor encoded by the non-rearranged V pre-B and the γ5 light chain locus (the pre-BCR). Pro- and pre-B cells are driven to proliferate and mature by signals from bone marrow stroma—in particular, IL-7. Light chain rearrangement occurs in the small pre-B-cell stage such that the full BCR is expressed at the immature B-cell stage. Immature B cells have rearranged Ig light chain genes and express sIgM. As immature B cells develop into mature B cells, sIgD is expressed as well as sIgM. At this point, B lineage development in bone marrow is complete, and B cells exit into the peripheral circulation and migrate to secondary lymphoid organs to encounter specific antigens.

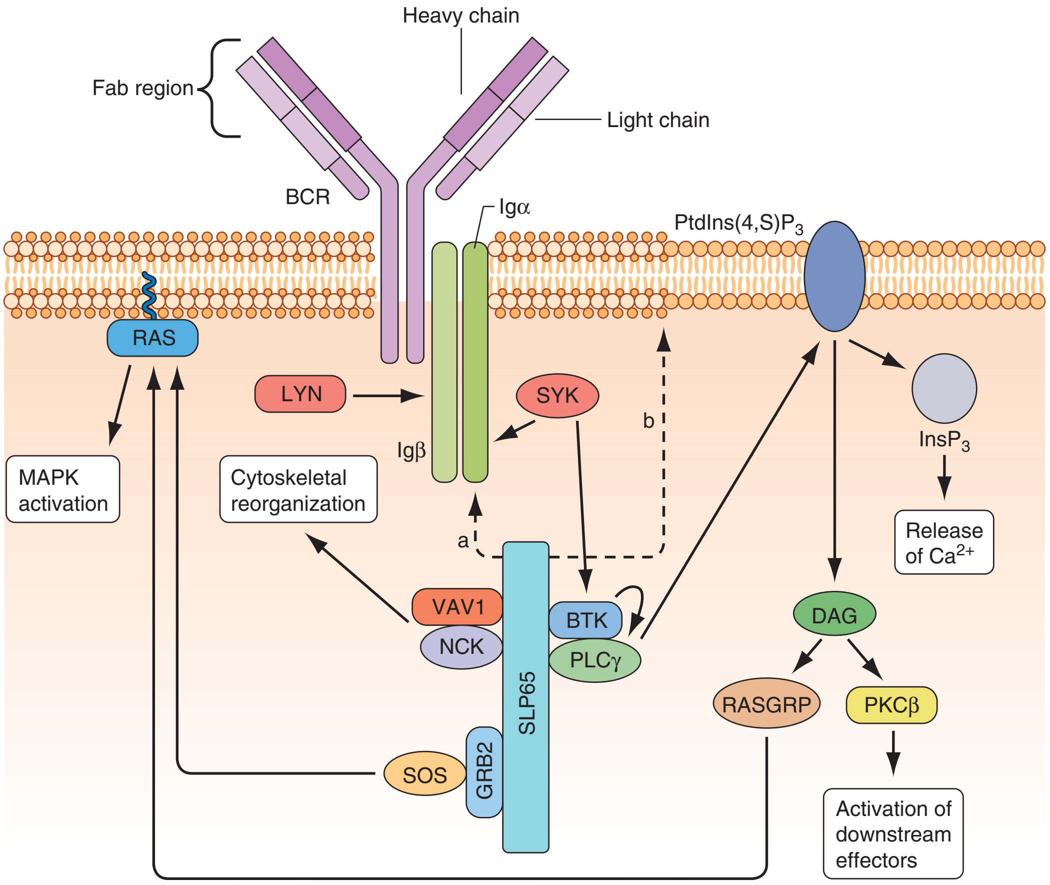

FIGURE 360-8 B-cell receptor (BCR) activation results in the sequential activation of protein tyrosine kinases, which results in the formation of a signaling complex and activation of downstream pathways as shown. Whereas SLP76 is recruited to the membrane through GADS and LAT, the mechanism of SLP65 recruitment is unclear. Studies have indicated two mechanisms: (a) direct binding by the SH2 domain of SLP65 to immunoglobulin (Ig) of the BCR complex or (b) membrane recruitment through a leucine zipper in the amino terminus of SLP65 and an unknown binding partner. ADAP, adhesion- and degranulation-promoting adaptor protein; AP1, activator protein 1; BTK, Bruton's tyrosine kinase; DAG, diacylglycerol; GRB2, growth factor receptor-bound protein 2; HPK1, hematopoietic progenitor kinase 1; $InsP_3$, inositol-1,4,5-trisphosphate; ITK, interleukin-2-inducible T-cell kinase; NCK, noncatalytic region of tyrosine kinase; NF-B, nuclear factor B; PKC, protein kinase C; PLC, phospholipase C; PtdIns(4,5)P_2, phosphatidylinositol-4,5-bisphosphate; RASGRP, RAS guanyl-releasing protein; SOS, son of sevenless homologue; SYK, spleen tyrosine kinase. *(Reproduced with permission from GA Koretzky et al: SLP76 and SLP65: Complex regulation of signalling in lymphocytes and beyond. Nat Rev Immunol 6:67, 2006.)*

Random rearrangements of Ig genes occasionally generate self-reactive antibodies, and mechanisms must be in place to correct these mistakes. One such mechanism is BCR editing, whereby autoreactive BCRs are mutated to not react with self-antigens. If receptor editing is unsuccessful in eliminating autoreactive B cells, then autoreactive B cells may undergo negative selection in the bone marrow through induction of apoptosis after BCR engagement of self-antigen.

After leaving the bone marrow, B cells populate peripheral B-cell sites, such as lymph node and spleen, and await contact with foreign antigens that react with each BCR. Antigen-driven B-cell activation occurs through the BCR, and *somatic hypermutation* takes place whereby point mutations in rearranged H- and L-genes give rise to mutant sIg molecules, some of which bind antigen better than the original sIg molecules. Somatic hypermutation, therefore, is a process whereby memory B cells in peripheral lymph organs have the best binding or the highest-affinity antibodies. This overall process of generating the best antibodies is called *affinity maturation of antibody*.

Lymphocytes that synthesize IgG, IgA, and IgE are derived from sIgM+, sIgD+ mature B cells. Ig class switching occurs in lymph node and other peripheral lymphoid tissue germinal centers. CD40 on B cells and CD40 ligand on T cells constitute a critical co-stimulatory receptor-ligand pair of immune-stimulatory molecules. Pairs of CD40+ B cells and CD40 ligand+ T cells bind and drive B-cell Ig class switching via T cell–produced cytokines such as IL-4 and TGF-β. IL-1, -2, -4, -5, and -6 synergize to drive mature B cells to proliferate and differentiate into Ig-secreting cells.

Humoral Mediators of Adaptive Immunity: Immunoglobulins

Immunoglobulins are the products of differentiated B cells and mediate the humoral arm of the immune response. The primary functions of antibodies are to bind specifically to antigen and bring about the inactivation or removal of the offending toxin, microbe, parasite, or other foreign substance from the body. The structural basis of Ig molecule function and Ig gene organization has provided insight into the role of antibodies in normal protective immunity, pathologic immune-mediated damage by immune complexes, and autoantibody formation against host determinants.

All immunoglobulins have the basic structure of two heavy and two light chains (Fig. 360-8). Immunoglobulin isotype (i.e., G, M, A, D, E) is determined by the type of Ig heavy chain present. IgG and IgA isotypes can be divided further into subclasses (G1, G2, G3, G4, and A1, A2) based on specific antigenic determinants on Ig heavy chains. The characteristics of human immunoglobulins are outlined in Table 360-11. The four chains are covalently linked by disulfide bonds. Each chain is made up of a V region and C regions (also called *domains*), themselves made up of units of ~110 amino acids. Light chains have one variable (V_L) and one constant (C_L) unit; heavy chains have one variable unit (V_H) and three or four constant (C_H) units, depending on isotype. As the name suggests, the constant, or C, regions of Ig molecules are made up of homologous sequences and share the same primary structure as all other Ig chains of the same isotype and subclass. Constant regions are involved in biologic functions of Ig molecules. The C_H2 domain of IgG and the C_H4 units of IgM are involved with the binding of the C1q portion of C1 during complement activation. The C_H region at the carboxy-terminal end of the IgG molecule, the Fc region, binds to surface Fc receptors (CD16, CD32, CD64) of macrophages, DCs, NK cells, B cells, neutrophils, and eosinophils. The Fc of IgA binds to FcαR (CD89), and the Fc of IgE binds to FcεR (CD23).

TABLE 360-11 Physical, Chemical, and Biologic Properties of Human Immunoglobulins

PROPERTY	IgG	IgA	IgM	IgD	IgE
Usual molecular form	Monomer	Monomer, dimer	Pentamer, hexamer	Monomer	Monomer
Other chains	None	J chain, SC	J chain	None	None
Subclasses	G1, G2, G3, G4	A1, A2	None	None	None
Heavy chain allotypes	Gm (=30)	No A1, A2m (2)	None	None	None
Molecular mass, kDa	150	160, 400	950, 1150	175	190
Serum level in average adult, mg/mL	9.5–12.5	1.5–2.6	0.7–1.7	0.04	0.0003
Percentage of total serum Ig	75–85	7–15	5–10	0.3	0.019
Serum half-life, days	23	6	5	3	2.5
Synthesis rate, mg/kg per day	33	65	7	0.4	0.016
Antibody valence	2	2, 4	10, 12	2	2
Classical complement activation	+(G1, 2?, 3)	–	++	–	–
Alternate complement activation	+(G4)	+	–	+	–
Binding cells via Fc	Macrophages, neutrophils, large granular lymphocytes	Lymphocytes	Lymphocytes	None	Mast cells, basophils, B cells
Biologic properties	Placental transfer, secondary antibody for most antipathogen responses	Secretory immunoglobulin	Primary antibody responses	Marker for mature B cells	Allergy, antiparasite responses

Source: Reproduced with permission from L Carayannopoulos, JD Capra, in WE Paul (ed): *Fundamental Immunology,* 3rd ed. New York, Raven, 1993.

Variable regions (V_L and V_H) constitute the antibody-binding (Fab) region of the molecule. Within the V_L and V_H regions are hypervariable regions (extreme sequence variability) that constitute the antigen-binding site unique to each Ig molecule. The idiotype is defined as the specific region of the Fab portion of the Ig molecule to which antigen binds. Antibodies against the idiotype portion of an antibody molecule are called *anti-idiotype antibodies*. The formation of such antibodies in vivo during a normal B-cell antibody response may generate a negative (or "off") signal to B cells to terminate antibody production.

IgG constitutes ~75–85% of total serum immunoglobulin. The four IgG subclasses are numbered in order of their level in serum, IgG1 being found in greatest amounts and IgG4 the least. IgG subclasses have clinical relevance in their varying ability to bind macrophage and neutrophil Fc receptors and to activate complement (Table 360-11). Moreover, selective deficiencies of certain IgG subclasses give rise to clinical syndromes in which the patient is inordinately susceptible to bacterial infections. IgG antibodies are frequently the predominant antibody made after rechallenge of the host with antigen (secondary antibody response).

IgM antibodies normally circulate as a 950-kDa pentamer with 160-kDa bivalent monomers joined by a molecule called the *J chain*, a 15-kDa nonimmunoglobulin molecule that also effects polymerization of IgA molecules. IgM is the first immunoglobulin to appear in the immune response (primary antibody response) and is the initial type of antibody made by neonates. Membrane IgM in the monomeric form also functions as a major antigen receptor on the surface of mature B cells (Table 360-11). IgM is an important component of immune complexes in autoimmune diseases. For example, IgM antibodies against IgG molecules (rheumatoid factors) are present in high titers in *rheumatoid arthritis*, other collagen diseases, and some infectious diseases (*subacute bacterial endocarditis*).

IgA constitutes only 7–15% of total serum immunoglobulin but is the predominant class of immunoglobulin in secretions. IgA in secretions (tears, saliva, nasal secretions, gastrointestinal tract fluid, and human milk) is in the form of secretory IgA (sIgA), a polymer consisting of two IgA monomers, a joining molecule, again termed the J chain, and a glycoprotein called the *secretory protein*. Of the two IgA subclasses, IgA1 is primarily found in serum, whereas IgA2 is more prevalent in secretions. IgA fixes complement via the alternative complement pathway and has potent antiviral activity in humans by prevention of virus binding to respiratory and gastrointestinal epithelial cells.

IgD is found in minute quantities in serum and, together with IgM, is a major receptor for antigen on the naïve B-cell surface. IgE, which is present in serum in very low concentrations, is the major class of immunoglobulin involved in arming mast cells and basophils by binding to these cells via the Fc region. Antigen cross-linking of IgE molecules on basophil and mast cell surfaces results in release of mediators of the immediate hypersensitivity (allergic) response (Table 360-11).

CELLULAR INTERACTIONS IN REGULATION OF NORMAL IMMUNE RESPONSES

The net result of activation of the humoral (B-cell) and cellular (T-cell) arms of the adaptive immune system by foreign antigen is the elimination of antigen directly by specific effector T cells or in concert with specific antibody.

The expression of adaptive immune cell function is the result of a complex series of immunoregulatory events that occur in phases. Both T and B lymphocytes mediate immune functions, and each of these cell types, when given appropriate signals, passes through stages, from activation and induction through proliferation, differentiation, and ultimately effector functions. The effector function expressed may be at the end point of a response, such as secretion of antibody by a differentiated plasma cell, or it might serve a regulatory function that modulates other functions, such as is seen with CD4 and CD8 T lymphocytes that modulate both differentiation of B cells and activation of CD8 cytotoxic T cells.

T_H1 CD4+ T cells, through elaboration of IFN-γ, have a central role in mediating intracellular killing by a variety of pathogens. T_H1 CD4+ T cells also provide T-cell help for generation of cytotoxic T cells and some types of opsonizing antibody, and they generally respond to antigens that lead to delayed hypersensitivity types of immune responses for many intracellular viruses and bacteria (such as HIV-1 or *M. tuberculosis*). In contrast, T_H2 cells have a primary role in regulatory humoral immunity and isotype switching. T_H2 cells, through production of IL-4 and IL-10, have a regulatory role in limiting proinflammatory responses mediated by T_H1 cells (Fig. 360-2). In addition, T_H2 CD4+ T cells provide help to B cells for specific Ig production and respond to antigens that require high antibody levels for foreign antigen elimination (extracellular encapsulated bacteria such as *Streptococcus pneumoniae* and certain parasite infections). T_H17 cells secrete cytokines IL-17, -22, and -26 and have been shown to play a role in autoimmune inflammatory disorders in addition to defense against extracellular bacteria and fungi, particularly at mucosal surfaces. T_H9 cells are defined by their secretion of IL-9 and have been shown to play a role in atopic disease, inflammatory bowel disease, and antitumor immunity. Moreover, the TFH subset of helper T cells secrete IL-21 and is crucial for providing the necessary signals to B cells in germinal centers to undergo affinity maturation. TFH13 cells secrete IL-4, IL-5, and IL-13 in response to allergens and and have been postulated to mediate anaphylaxis reactions (Fig. 360-2). In summary, the type of T-cell response generated in an immune response is determined by the microbe PAMPs presented to the DCs, the TLRs on the DCs that become activated, the types of DCs that are activated, and the cytokines that are produced (Table 360-7). Commonly, myeloid DCs produce IL-12 and activate T_H1 T-cell responses that result in IFN-γ and cytotoxic T-cell induction, and plasmacytoid DCs produce IFN-α and lead to T_H2 responses that result in IL-4 production and enhanced antibody responses.

As shown in Fig. 360-2, upon activation by DCs, T-cell subsets that produce IL-2, IL-3, IFN-γ, and/or IL-4, -5, -6, -10, and -13 are generated and exert positive and negative influences on effector T and B cells. For B cells, trophic effects are mediated by a variety of cytokines, particularly T cell–derived IL-3, -4, -5, and -6, that act at sequential stages of B-cell maturation, resulting in B-cell proliferation, differentiation, and ultimately antibody secretion. For cytotoxic T cells, trophic factors include inducer T-cell secretion of IL-2, IFN-γ, and IL-12.

Important types of immunomodulatory T cells that control immune responses are *CD4 and CD8 Treg cells*. These cells express the α chain of the IL-2 receptor (CD25), produce IL-10, and suppress both T- and B-cell responses. T regulatory cells (Tregs) are induced by immature DCs and play key roles in maintaining tolerance to self-antigens. Loss of Treg cells is the cause of organ-specific autoimmune disease in mice such as autoimmune thyroiditis, adrenalitis, and oophoritis and plays a role in inflammatory bowel disease (see "Immune Tolerance and Autoimmunity" below, **Chap. 361**). Tregs also play key roles in controlling the magnitude and duration of immune responses to microbes. Normally, after the initial immune response to a microbe has eliminated the invader, Tregs are activated to suppress the antimicrobe response and prevent host injury. Some microbes have adapted to induce Treg activation at the site of infection to promote parasite infection and survival. In *Leishmania* infection, the parasite induces Treg accumulation at skin infection sites that dampens anti-*Leishmania* T-cell responses and prevents parasite elimination. Although B cells recognize native antigen via B-cell surface Ig receptors, B cells require T-cell help to produce high-affinity antibody of multiple isotypes that are the most effective in eliminating foreign antigen. In B-cell germinal centers, CD4 T cells that promote B-cell maturation and affinity maturation are termed T follicular helper (TFH) cells. T cell–B cell interactions that lead to high-affinity antibody production require (1) processing of native antigen by B cells and expression of peptide fragments on the B-cell surface for presentation to T_H cells, (2) the ligation of B cells by both the TCR complex and the CD40 ligand, (3) induction of the process termed *antibody isotype switching* in antigen-specific B-cell clones, and (4) induction of the process of affinity maturation of antibody in the germinal centers of B-cell follicles of lymph node and spleen.

Naïve B cells express cell-surface IgD and IgM, and initial contact of naïve B cells with antigen is via binding of native antigen to B-cell surface IgM. T-cell cytokines, released following T_H2 cell contact with B cells or by a "bystander" effect, induce changes in Ig gene conformation that promote recombination of Ig genes. These events then result in the switching of expression of heavy chain exons in a triggered B cell, leading to the secretion of IgG, IgA, or, in some cases, IgE antibody with the same V region antigen specificity as the original IgM antibody, for response to a wide variety of extracellular bacteria, protozoa, and helminths. CD40 ligand expression by activated T cells is critical for induction of B-cell antibody isotype switching and for B-cell responsiveness to cytokines. Patients with mutations in T-cell CD40 ligand have B cells that are unable to undergo isotype switching, resulting in lack of memory B-cell generation and the immunodeficiency syndrome of *X-linked hyper-IgM syndrome* **(Chaps. 361 and 362)**.

IMMUNE TOLERANCE AND AUTOIMMUNITY

Immune tolerance is defined as the absence of activation of pathogenic autoreactivity to self-antigens. Mechanisms of immune tolerance can be classified as *cell intrinsic* or *cell extrinsic*. Cell intrinsic mechanisms of tolerance include apoptosis and induction of cell unresponsiveness (anergy). Mechanisms of cell extrinsic tolerance include suppression of immune responses by immunomodulatory cells such as Tregs. *Autoimmune diseases* are syndromes caused by the activation of T or B cells or both, with no evidence of other causes such as infections or malignancies **(Chaps. 361 and 367)**. Low levels of autoreactivity of T and B cells with self-antigens in the periphery are critical to T- and B-cell survival. Similarly, low levels of autoreactivity and thymocyte recognition of self-antigens in the thymus are the mechanisms whereby normal T cells are positively selected to survive and leave the thymus to respond to foreign microbes in the periphery and T cells highly reactive to self-antigens are negatively selected and die to prevent overly self-reactive T cells from migrating to the periphery (central tolerance). Unlike the presentation of microbial antigens by mature DCs, the presentation of self-antigens by immature DCs neither activates nor matures the DCs to express high levels of co-stimulatory molecules such as B7-1 (CD80) or B7-2 (CD86). When peripheral T cells are stimulated by DCs expressing self-antigens in the context of HLA molecules, sufficient stimulation of T cells occurs to keep them alive, but otherwise, they remain anergic, or nonresponsive, until T cells contact a DC with high levels of co-stimulatory molecules expressing microbial antigens and become activated to respond to the microbe. If B cells have high self-reactive BCRs, they normally undergo either deletion in the bone marrow or receptor editing to express a less autoreactive receptor. Although many autoimmune diseases are characterized by abnormal or pathogenic autoantibody production (**see Chap. 361**, Table 361-4), most autoimmune diseases are caused by a combination of excess T- and B-cell reactivity.

Multiple factors contribute to the genesis of autoimmune disease syndromes, including genetic susceptibility (e.g., HLA-B27 with ankylosing spondylitis), environmental immune stimulants such as drugs (e.g., procainamide and phenytoin [Dilantin] with drug-induced systemic lupus erythematosus), infectious agent triggers (e.g., Epstein-Barr virus and autoantibody production against red blood cells and platelets), and loss of Treg cells (leading to thyroiditis, adrenalitis, and oophoritis).

Immunity at Mucosal Surfaces Mucosa covering the respiratory, digestive, and urogenital tracts; the eye conjunctiva; the inner ear; and the ducts of all exocrine glands contain cells of the innate and adaptive mucosal immune system that protect these surfaces against pathogens. In the healthy adult, mucosa-associated lymphoid tissue (MALT) contains 80% of all immune cells within the body and constitutes the largest mammalian lymphoid organ system.

MALT has three main functions: (1) to protect the mucous membranes from invasive pathogens; (2) to prevent uptake of foreign antigens from food, commensal organisms, and airborne pathogens and particulate matter; and (3) to prevent pathologic immune responses from foreign antigens if they do cross the mucosal barriers of the body.

MALT is a compartmentalized system of immune cells that functions independently from systemic immune organs. Whereas the systemic immune organs are essentially sterile under normal conditions and respond vigorously to pathogens, MALT immune cells are continuously bathed in foreign proteins and commensal bacteria, and they must select those pathogenic antigens that must be eliminated. MALT contains anatomically defined foci of immune cells in the intestine, tonsil, appendix, and peribronchial areas that are inductive sites for mucosal immune responses. From these sites, immune T and B cells migrate to effector sites in mucosal parenchyma and exocrine glands where mucosal immune cells eliminate pathogen-infected cells. In addition to mucosal immune responses, all mucosal sites have strong mechanical and chemical barriers and cleansing functions to repel pathogens.

Key components of MALT include specialized epithelial cells called "membrane" or "M" cells that take up antigens and deliver them to DCs or other APCs. Regulatory cells that maintain gut homeostasis include ILC APCs, likely ILC3, that drive the development of CD4 Tregs that suppress pathogenic immune responses to benign commensal microbiota. Effector cells in MALT include B cells producing antipathogen neutralizing antibodies of secretory IgA as well as IgG isotype, T cells producing similar cytokines as in systemic immune responses, and T helper and cytotoxic T cells that respond to pathogen-infected cells.

Secretory IgA is produced in amounts of >50 mg/kg of body weight per 24 h and functions to inhibit bacterial adhesion, inhibit macromolecule absorption in the gut, neutralize viruses, and enhance antigen elimination in tissue through binding to IgA and receptor-mediated transport of immune complexes through epithelial cells.

Recent studies have demonstrated the importance of commensal gut and other mucosal bacteria to the health of the human immune system. Normal commensal flora induces anti-inflammatory events in the gut and protects epithelial cells from pathogens through TLRs and other PRR signaling. When the gut is depleted of normal commensal flora, the immune system becomes abnormal, with loss of T_H1 T-cell function. Restoration of the normal gut flora can reestablish the balance in Treg and T helper cell ratios characteristic of the normal immune system. Diet also has an impact on the gut microbiome. Altered microbiome composition has been etiologically related to obesity, insulin resistance, inflammatory bowel disease, and diabetes. When the gut barrier is intact, either antigens do not transverse the gut epithelium or, when pathogens are present, a self-limited, protective MALT immune response eliminates the pathogen. However, when the gut barrier breaks down, immune responses to commensal flora antigens can contribute to *Crohn's disease* and, perhaps, *ulcerative colitis* **(Chap. 337)**. Uncontrolled MALT immune responses to food antigens, such as gluten, can cause *celiac disease* **(Chap. 337)**.

THE CELLULAR AND MOLECULAR CONTROL OF PROGRAMMED CELL DEATH

The process of apoptosis (programmed cell death) plays a crucial role in regulating normal immune responses to antigen. In general, a wide variety of stimuli trigger one of several apoptotic pathways to eliminate microbe-infected cells, eliminate cells with damaged DNA, or eliminate activated immune cells that are no longer needed. The largest known family of "death receptors" is the TNF receptor (TNF-R) family (TNF-R1, TNF-R2, Fas [CD95], death receptor 3 [DR3], death receptor 4 [DR4; TNF-related apoptosis-including ligand receptor 1, or TRAIL-R1], and death receptor 5 [DR5, TRAIL-R2]); their ligands are all in the TNF-α family. Binding of ligands to these death receptors leads to a signaling cascade that involves activation of the *caspase* family of molecules that leads to DNA cleavage and cell death. Two other pathways of programmed cell death involve nuclear *p53* in the elimination of cells with abnormal DNA and *mitochondrial cytochrome c* to induce cell death in damaged cells. A number of human diseases have now been described that result from, or are associated with, mutated apoptosis genes. These include mutations in the Fas and Fas ligand genes in autoimmune and lymphoproliferation syndromes, and multiple associations of mutations in genes in the apoptotic pathway with malignant syndromes **(Chap. 361)**.

MECHANISMS OF IMMUNE-MEDIATED DAMAGE TO MICROBES OR HOST TISSUES

Several responses by the host innate and adaptive immune systems to foreign microbes culminate in rapid and efficient elimination of microbes. In these scenarios, the classic weapons of the adaptive immune system (T cells, B cells) interface with cells (macrophages, DCs, NK cells, neutrophils, eosinophils, basophils) and soluble products (microbial peptides, pentraxins, complement and coagulation systems) of the innate immune system **(Chaps. 67 and 363)**.

There are five general phases of host defenses: (1) migration of leukocytes to sites of antigen localization; (2) antigen-nonspecific recognition of pathogens by macrophages and other cells and systems of the innate immune system; (3) specific recognition of foreign antigens mediated by T and B lymphocytes; (4) amplification of the inflammatory response with recruitment of specific and nonspecific effector cells by complement components, cytokines, kinins, arachidonic acid metabolites, and mast cell–basophil products; and (5) macrophage, neutrophil, and lymphocyte participation in destruction of antigen with ultimate removal of antigen particles by phagocytosis (by macrophages or neutrophils) or by direct cytotoxic mechanisms (involving macrophages, neutrophils, DCs, or lymphocytes). Under normal circumstances, orderly progression of host defenses through these phases results in a well-controlled immune and inflammatory response that protects the host from the offending antigen. However, dysfunction of any of the host defense systems can damage host tissue and produce clinical disease. Furthermore, for certain pathogens or antigens, the normal immune response itself might contribute substantially to the tissue damage. For example, the immune and inflammatory response in the brain to certain pathogens such as *M. tuberculosis* may be responsible for much of the morbidity rate of this disease in that organ system **(Chap. 183)**. In addition, the morbidity rate associated with certain pneumonias such as that caused by *Pneumocystis jirovecii* may be associated more with inflammatory infiltrates than with the tissue-destructive effects of the microorganism itself **(Chap. 227)**.

Molecular Basis of Lymphocyte–Endothelial Cell Interactions The control of lymphocyte circulatory patterns between the bloodstream and peripheral lymphoid organs operates at the level of lymphocyte–endothelial cell interactions to control the specificity of lymphocyte subset entry into organs. Similarly, lymphocyte–endothelial cell interactions regulate the entry of lymphocytes into inflamed tissue. Adhesion molecule expression on lymphocytes and endothelial cells regulates the retention and subsequent egress of lymphocytes within tissue sites of antigenic stimulation, delaying cell exit from tissue and preventing reentry into the circulating lymphocyte pool **(Fig. 360-9)**. All types of lymphocyte migration begin with lymphocyte attachment to specialized regions of vessels, termed *high endothelial venules* (HEVs). An important concept is that adhesion molecules do not generally bind their ligand until a conformational change (ligand activation) occurs in the adhesion molecule that allows ligand binding. Induction of a conformation-dependent determinant on an adhesion molecule can be accomplished by cytokines or via ligation of other adhesion molecules on the cell.

The first stage of lymphocyte–endothelial cell interactions, *attachment and rolling*, occurs when lymphocytes leave the stream of flowing blood cells in a postcapillary venule and roll along venule endothelial cells (Fig. 360-9). Lymphocyte rolling is mediated by the L-selectin molecule (LECAM-1, LAM-1, CD62L) and slows cell transit time through venules, allowing time for activation of adherent cells.

The second stage of lymphocyte–endothelial cell interactions, *firm adhesion with activation-dependent stable arrest*, requires stimulation of lymphocytes by chemoattractants or by endothelial cell–derived cytokines. Cytokines thought to participate in adherent cell activation include members of the IL-8 family, platelet-activation factor, leukotriene B_4, and C5a. In addition, HEVs express chemokines, SLC (CCL21) and ELC (CCL19), which participate in this process. Following activation by chemoattractants, lymphocytes shed L-selectin from the cell surface and upregulate cell CD11b/18 (MAC-1) or CD11a/18 (LFA-1) molecules, resulting in firm attachment of lymphocytes to HEVs.

Lymphocyte homing to peripheral lymph nodes involves adhesion of L-selectin to glycoprotein HEV ligands collectively referred to as *peripheral node addressin* (*PNAd*), whereas homing of lymphocytes to intestine Peyer's patches primarily involves adhesion of the α4β7 integrin to mucosal addressin cell adhesion molecule-1 (MAdCAM-1) on the Peyer's patch HEVs. However, for migration to mucosal Peyer's patch lymphoid aggregates, naïve lymphocytes primarily use L-selectin, whereas memory lymphocytes use α4β7 integrin. α4β1 integrin (CD49d/CD29, VLA-4)–VCAM-1 interactions are important in the initial interaction of memory lymphocytes with HEVs of multiple organs in sites of inflammation.

The third stage of leukocyte emigration in HEVs is *sticking and arrest*. Sticking of the lymphocyte to endothelial cells and arrest at the site of sticking are mediated predominantly by ligation of a1β2 integrin LFA-1 to the integrin ligand ICAM-1 on HEVs. Whereas the first three stages of lymphocyte attachment to HEVs take only a few seconds, the fourth stage of lymphocyte emigration, *transendothelial migration*, takes ~10 min. Although the molecular mechanisms that control lymphocyte transendothelial migration are not fully characterized, the HEV CD44 molecule and molecules of the HEV glycocalyx (extracellular matrix) are thought to play important regulatory roles in

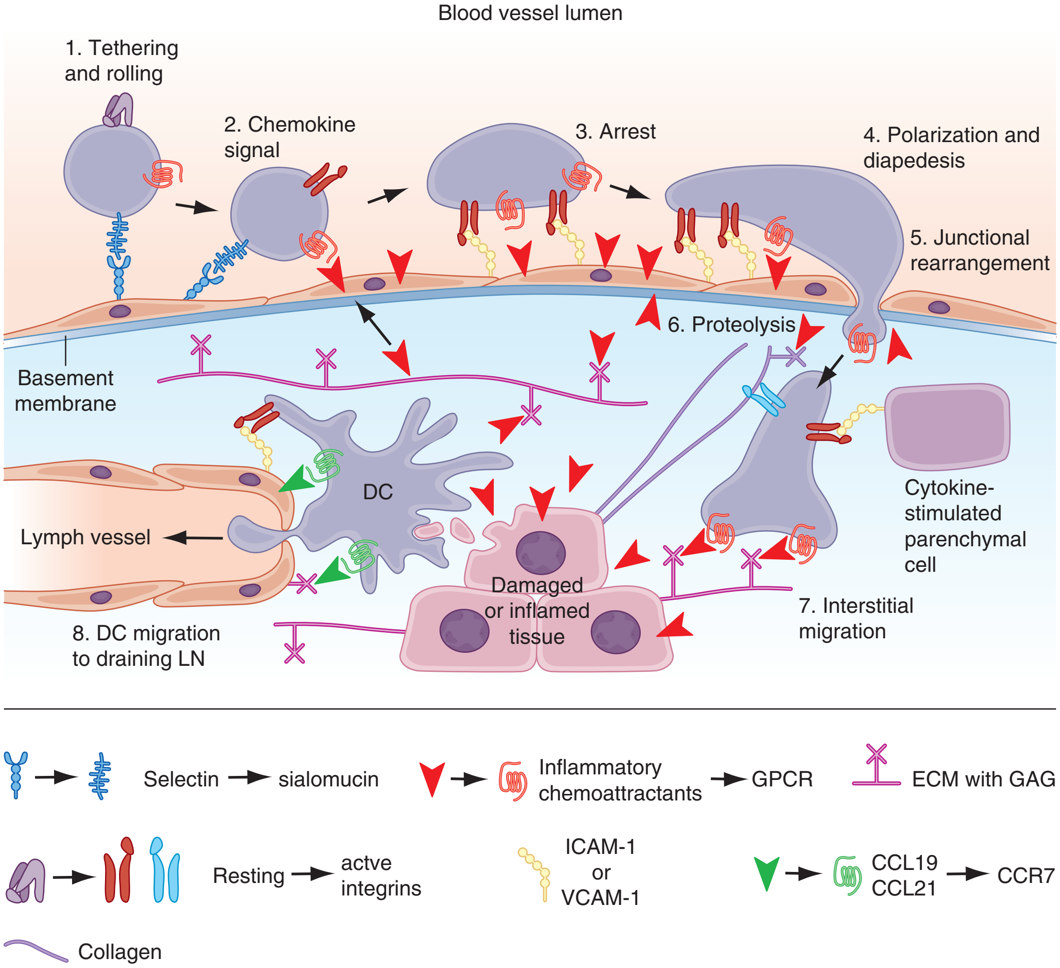

FIGURE 360-9 Key migration steps of immune cells at sites of inflammation. Inflammation due to tissue damage or infection induces the release of cytokines (not shown) and inflammatory chemoattractants (*red arrowheads*) from distressed stromal cells and "professional" sentinels, such as mast cells and macrophages (not shown). The inflammatory signals induce upregulation of endothelial selectins and immunoglobulin "superfamily" members, particularly ICAM-1 and/or VCAM-1. Chemoattractants, particularly chemokines, are produced by or translocated across venular endothelial cells (*red arrow*) and are displayed in the lumen to rolling leukocytes. Those leukocytes that express the appropriate set of trafficking molecules undergo a multistep adhesion cascade (steps 1–3) and then polarize and move by diapedesis across the venular wall (steps 4 and 5). Diapedesis involves transient disassembly of endothelial junctions and penetration through the underlying basement membrane (step 6). Once in the extravascular (interstitial) space, the migrating cell uses different integrins to gain "footholds" on collagen fibers and other ECM molecules, such as laminin and fibronectin, and on inflammation-induced ICAM-1 on the surface of parenchymal cells (step 7). The migrating cell receives guidance cues from distinct sets of chemoattractants, particularly chemokines, which may be immobilized on glycosaminoglycans (GAG) that "decorate" many ECM molecules and stromal cells. Inflammatory signals also induce tissue dendritic cells (DCs) to undergo maturation. Once DCs process material from damaged tissues and invading pathogens, they upregulate CCR7, which allows them to enter draining lymph vessels that express the CCR7 ligand CCL21 (and CCL19). In lymph nodes (LNs), these antigen-loaded mature DCs activate naïve T cells and expand pools of effector lymphocytes, which enter the blood and migrate back to the site of inflammation. T cells in tissue also use this CCR7-dependent route to migrate from peripheral sites to draining lymph nodes through afferent lymphatics. *(Reproduced with permission from AD Luster et al: Immune cell migration in inflammation: present and future therapeutic targets. Nat Immunol 6:1182, 2005.)*

this process **(Fig. 360-10)**. Finally, expression of matrix metalloproteases capable of digesting the subendothelial basement membrane, rich in nonfibrillar collagen, appears to be required for the penetration of lymphoid cells into the extravascular sites.

Abnormal induction of HEV formation and use of the molecules discussed above have been implicated in the induction and maintenance of inflammation in a number of chronic inflammatory diseases. In animal models of type 1 diabetes mellitus, MAdCAM-1 and GlyCAM-1 have been shown to be highly expressed on HEVs in inflamed pancreatic islets, and treatment of these animals with inhibitors of L-selectin and a4 integrin function blocked the development of type 1 diabetes mellitus **(Chap. 415)**. A similar role for abnormal induction of the adhesion molecules of lymphocyte emigration has been suggested in *rheumatoid arthritis* **(Chap. 370)**, *Hashimoto's thyroiditis* **(Chap. 394)**, *Graves' disease* **(Chap. 394)**, *multiple sclerosis* **(Chap. 455)**, *Crohn's disease* **(Chap. 337)**, and *ulcerative colitis* **(Chap. 337)**.

Immune-Complex Formation Clearance of antigen by immune-complex formation between antigen, complement, and antibody is a highly effective mechanism of host defense. However, depending on the level of immune complexes formed and their physicochemical properties, immune complexes may or may not result in host and foreign cell damage. After antigen exposure, certain types of soluble antigen-antibody complexes freely circulate and, if not cleared by the reticuloendothelial system, can be deposited in blood vessel walls and in other tissues such as renal glomeruli and cause *vasculitis* or *glomerulonephritis* syndromes **(Chaps. 326 and 375)**. Deficiencies of early complement components are associated with inefficient clearance of immune complexes and immune complex–mediated autoimmune syndromes or encapsulated bacterial infections such as *S. pneumoniae*, whereas deficiencies of the later complement components are associated with susceptibility to recurrent *Neisseria* infections **(Table 360-12)**.

Immediate-Type Hypersensitivity Helper T cells that drive antiallergen IgE responses are usually T_H2-type inducer T cells that secrete IL-4, IL-5, IL-6, and IL-10. A subset of TFH, TFH13, cells have been identified that produce IL-4, IL-5, and IL-13, which play a key role in responses to allergens that induce IgE and mediate anaphylaxis. Mast cells and basophils have high-affinity receptors for the Fc portion of IgE (FcRI), and cell-bound antiallergen IgE effectively

TABLE 360-12 Complement Deficiencies and Associated Diseases

COMPONENT	ASSOCIATED DISEASES
Classic Pathway	
Clq, Clr, Cls, C4	Immune-complex syndromes,[a] pyogenic infections
C2	Immune-complex syndromes,[a] few with pyogenic infections
C1 inhibitor	Rare immune-complex disease, few with pyogenic infections
C3 and Alternative Pathway C3	
C3	Immune-complex syndromes,[a] pyogenic infections
D	Pyogenic infections
Properdin	*Neisseria* infections
I	Pyogenic infections
H	Hemolytic-uremic syndrome
Membrane Attack Complex	
C5, C6, C7, C8	Recurrent *Neisseria* infections, immune-complex disease
C9	Rare *Neisseria* infections

[a]Immune-complex syndromes include systemic lupus erythematosus (SLE) and SLE-like syndromes, glomerulonephritis, and vasculitis syndromes.

Source: After JA Schifferli, DK Peters: Lancet 322:957, 1983. Copyright 1983.

"arms" basophils and mast cells. Mediator release is triggered by antigen (allergen) interaction with Fc receptor-bound IgE, and the mediators released are responsible for the pathophysiologic changes of *allergic diseases*. Mediators released from mast cells and basophils can be divided into three broad functional types: (1) those that increase vascular permeability and contract smooth muscle (histamine, platelet-activating factor, SRS-A, BK-A), (2) those that are chemotactic for or activate other inflammatory cells (ECF-A, NCF, leukotriene B_4), and (3) those that modulate the release of other mediators (BK-A, platelet-activating factor) **(Chap. 363)**.

Cytotoxic Reactions of Antibody In this type of immunologic injury, complement-fixing (C1-binding) antibodies against normal or foreign cells or tissues (IgM, IgG1, IgG2, IgG3) bind complement via the classic pathway and initiate a sequence of events similar to that initiated by immune-complex deposition, resulting in cell lysis or tissue injury. Examples of antibody-mediated cytotoxic reactions include red cell lysis in *transfusion reactions*, *Goodpasture's syndrome* with anti–glomerular basement membrane antibody formation, and *pemphigus vulgaris* with anti-epidermal antibodies inducing blistering skin disease.

Delayed-Type Hypersensitivity Reactions Inflammatory reactions initiated by mononuclear leukocytes and not by antibody alone have been termed *delayed-type hypersensitivity reactions*. The term *delayed* has been used to contrast a secondary cellular response that appears 48–72 h after antigen exposure with an *immediate* hypersensitivity response generally seen within 12 h of antigen challenge and initiated by basophil mediator release or preformed antibody. For example, in an individual previously infected with *M. tuberculosis* organisms, intradermal placement of tuberculin purified protein derivative as a skin test challenge results in an indurated area of skin at 48–72 h, indicating previous exposure to tuberculosis.

The cellular events that result in classic delayed-type hypersensitivity responses are centered on T cells (predominantly, although not exclusively, IFN-γ, IL-2, and TNF-α-secreting T_H1-type helper T cells) and macrophages. Recently, NK cells have been suggested to play a major role in the form of delayed hypersensitivity that occurs following skin contact with immunogens. First, local immune and inflammatory responses at the site of foreign antigen upregulate endothelial cell adhesion molecule expression, promoting the accumulation of lymphocytes at the tissue site. In the scheme outlined in Fig. 360-2, antigen is processed by DCs and presented to small numbers of CD4+ T cells expressing a TCR specific for the antigen. IL-12 produced by APCs induces T cells to produce IFN-γ (T_H1 response). Macrophages frequently undergo epithelioid cell transformation and fuse to form multinucleated giant cells in response to IFN-γ. This type of mononuclear cell infiltrate is termed *granulomatous inflammation*. Examples of diseases in which delayed-type hypersensitivity plays a major role are fungal infections (*histoplasmosis*; **Chap. 218**), mycobacterial infections (*tuberculosis*, *leprosy*; **Chaps. 183** and **184**), chlamydial infections (*lymphogranuloma venereum*; **Chap. 194**), helminth infections (*schistosomiasis*; **Chap. 241**), reactions to toxins (*berylliosis*; **Chap. 300**), and hypersensitivity reactions to organic dusts (*hypersensitivity pneumonitis*; **Chap. 299**). In addition, delayed-type hypersensitivity responses play important roles in tissue damage in autoimmune diseases such as *rheumatoid arthritis*, *temporal arteritis*, and granulomatosis with polyangiitis **(Chaps. 370 and 375)**.

Autophagy Autophagy is a process that involves a lysosomal degradation pathway mechanism of cells to dispose of intracellular debris and damaged organelles. Autophagy by cells of the innate immune system is used to control intracellular infectious agents such as *M. tuberculosis*, in part by initiation of phagosome maturation and enhancing MHC class II antigen presentation to CD4 T cells.

CLINICAL EVALUATION OF IMMUNE FUNCTION

Clinical assessment of immunity requires investigation of the four major components of the immune system that participate in host defense and in the pathogenesis of autoimmune diseases: (1) humoral immunity (B cells); (2) cell-mediated immunity (T cells, monocytes); (3) phagocytic cells of the reticuloendothelial system (macrophages), as well as polymorphonuclear leukocytes; and (4) complement. Clinical problems that require an evaluation of immunity include chronic infections, recurrent infections, unusual infecting agents, and certain autoimmune syndromes. The type of clinical syndrome under evaluation can provide information regarding possible immune defects **(Chap. 362)**. Defects in cellular immunity generally result in viral, mycobacterial, and fungal infections. An extreme example of deficiency in cellular immunity is AIDS **(Chap. 208)**. Antibody deficiencies result in recurrent bacterial infections, frequently with organisms such as *S. pneumoniae* and *Haemophilus influenzae* **(Chap. 362)**. Disorders of phagocyte function are frequently manifested by recurrent skin infections, often due to *Staphylococcus aureus* **(Chap. 67)**. Finally, deficiencies of early and late complement components are associated with autoimmune phenomena and recurrent *Neisseria* infections (Table 360-12). Artificial intelligence/machine learning algorithms are now being tested for improving the diagnosis of infectious, immune deficiency, and autoimmune diseases. **For further discussion of useful initial screening tests of immune function, see Chap. 362.**

IMMUNOTHERAPY

Many therapies for autoimmune and inflammatory diseases involve the use of nonspecific immune-modulating or immunosuppressive agents such as glucocorticoids or cytotoxic drugs. The goal of development of new treatments for immune-mediated diseases is to design ways to specifically interrupt pathologic immune responses, leaving nonpathologic immune responses intact **(Chap. 361)**. Novel ways to interrupt pathologic immune responses that are under investigation include the use of anti-inflammatory cytokines or specific cytokine inhibitors as anti-inflammatory agents, the use of monoclonal antibodies against T or B lymphocytes as therapeutic agents, the use of intravenous Ig for certain infections and immune complex–mediated diseases, the use of specific cytokines to reconstitute components of the immune system, and bone marrow transplantation to replace the pathogenic immune system with a more normal immune system **(Chaps. 67, 208, 361, and 362)**. CTLA-4 inhibitors such as ipilimumab and tremelimumab and anti-PD-1 antibodies such as nivolumab are *termed checkpoint inhibitors* and have been shown to reverse CD8 T-cell exhaustion in melanoma and other solid tumors and induce immune cell control of tumor growth **(Chap. 361)**. A technique that engineers autologous T cells to express antibody receptors that target leukemic cells, termed *chimeric antigen receptor T cells* (CAR T cells), has been approved by the U.S. Food and Drug Administration (FDA) for the treatment of certain types of leukemias and lymphomas **(Chap. 361)**.

Cell-based therapies have been studied for many years, including ex vivo activation of NK cells for reinfusion into patients with malignancies and DC therapy of ex vivo priming of DCs for enhanced presentation of cancer antigens, with reinfusion of primed DCs into the patient. One such strategy for DC therapy has been approved by the FDA for treatment of advanced prostate cancer.

Cytokines and Cytokine Inhibitors Several TNF inhibitors are used as biological therapies in the treatment of rheumatoid arthritis; these include monoclonal antibodies, TNF-R Fc fusion proteins, and Fab fragments. Use of anti-TNF-α antibody therapies such as adalimumab, infliximab, and golimumab has resulted in clinical improvement in patients with these diseases and has opened the way for targeting TNF-α to treat other severe forms of autoimmune and/or inflammatory disease **(Chap. 361)**. Blockage of TNF-α has been effective in *rheumatoid arthritis*, *psoriasis*, *Crohn's disease*, and *ankylosing spondylitis*. Other cytokine inhibitors are recombinant soluble TNF-α receptor (R) fused to human Ig and anakinra (soluble *IL-1 receptor antagonist*, or IL-1ra).

TNF-αR-Fc fusion protein (etanercept) and IL-1ra act to inhibit the activity of pathogenic cytokines in rheumatoid arthritis, i.e., TNF-α and IL-1, respectively. Similarly, anti-IL-6, IFN-β, and IL-11 act to inhibit pathogenic proinflammatory cytokines. Anti-IL-6 (tocilizumab) inhibits IL-6 activity, whereas IFN-β and IL-11 decrease IL-1 and TNF-α production **(Chap. 361)**. Of particular note has been the successful use of IFN-γ in the treatment of the phagocytic cell defect in *chronic granulomatous disease* **(Chap. 67)**.

T_H17 CD4 T cells have been implicated in the pathogenesis of psoriasis, ulcerative colitis, and other autoimmune diseases. Monoclonal antibodies have now been developed that target cytokines (IL-12, IL-23) that induce T_H17 T-cell differentiation and are licensed by the FDA for treatment of psoriasis and inflammatory bowel disease. Monoclonal antibodies that directly target IL-17 have also recently been licensed for psoriasis and psoriatic arthritis treatment. Monoclonal antibodies against cytokines and immunoregulatory molecules are now mainstays of cancer and autoimmune disease therapy **(Chap. 361)**.

Intravenous Immunoglobulin (IVIg) IVIg has been used successfully to block reticuloendothelial cell function and immune complex clearance in various immune cytopenias such as immune thrombocytopenia **(Chap. 120)**. In addition, IVIg is useful for prevention of tissue damage in certain inflammatory syndromes such as Kawasaki disease **(Chap. 375)** and as Ig replacement therapy for certain types of immunoglobulin deficiencies **(Chap. 362)**. In addition, controlled clinical trials support the use of IVIg in selected patients with graft-versus-host disease, multiple sclerosis, myasthenia gravis, Guillain-Barré syndrome, and chronic demyelinating polyneuropathy.

Stem Cell Transplantation Hematopoietic stem cell transplantation (SCT) is now being studied to treat several autoimmune diseases including systemic lupus erythematosus, multiple sclerosis, and scleroderma. The goal of immune reconstitution in autoimmune disease syndromes is to replace a dysfunctional immune system with a normally reactive immune cell repertoire. Controlled clinical trials in these three diseases are now being launched in the United States and Europe to compare the toxicity and efficacy of conventional immunosuppression therapy with that of myeloablative autologous SCT. Recently, SCT was used in the setting of HIV-1 infection. HIV-1 infection of CD4+ T cells requires the presence of surface CD4 receptor and the chemokine receptor 5 (CCR5) co-receptor. Studies have demonstrated that patients who are homozygous for a 32-bp deletion in the CCR5 allele do not express CD4+ T-cell CCR5 and thus are resistant to HIV-1 infection with HIV-1 strains that use this co-receptor. Stem cells from a homozygous CCR5 delta32 donor were transplanted to an HIV-1-infected individual following standard conditioning for such transplants and resulted in long-term control of the virus without antiretrovirals. Thus, a number of recent insights into immune system function have spawned a new field of interventional immunotherapy and have enhanced the prospect for development of more specific and nontoxic therapies for immune and inflammatory diseases **(Chap. 361)**.

FURTHER READING

Altan-Bonnet G, Mukherjee R: Cytokine-mediated communication: A quantitative appraisal of immune complexity. Nat Rev Immunol 19:205, 2020.

Brown CC, Rudensky AY: Spatiotemporal regulation of peripheral T cell tolerance. Science 380:472, 2023.

McLean KC, Mandal M: It takes three receptors to raise a B cell. Trends Immunol 41:629, 2020.

Nersesian S et al: Killer instincts: Natural killer cells as multifactorial cancer immunotherapy. Front Immunol 14:1269614, 2023.

Netea MG et al: Defining trained immunity and its role in health and disease. Nat Rev Immunol 20:375, 2020.

Pellicci DG et al: Thymic development of unconventional T cells: How NKT, cells MAIT cells and γδ T cells emerge. Nat Rev Immunol 20:756, 2020.

Pulendran B: Immunology taught by vaccines. Science 366:1074, 2019.

Travis MA, Romagnani C: How regulatory T cells are primed to aid tolerance of gut bacteria. Nature 610:638, 2022.

Vinuesa CG et al: Innate virus-sensing pathways in B cell systemic autoimmunity. Science 380:478, 2023.

Vivier E et al: Innate lymphoid cells: 10 years on. Cell 174:1054, 2018.

361

Mechanisms of Regulation and Dysregulation of the Immune System

Barton F. Haynes, Kelly A. Cuttle,
Anthony S. Fauci

DEFINITIONS

Anergy—A reversible tolerance mechanism in which the T or B cell is in an unresponsive state following an antigen encounter but remains alive.

Chimeric antigen receptor T cells (CAR T)—Synthetic hybrid receptors created by recombinant techniques that combine an extracellular domain, usually derived from an antibody single-chain variable fragment (scFv), with intracellular signaling domains from activating co-stimulatory molecules (from endogenous T-cell receptors [TCRs], CD28, or 4-1BB) that allow for retargeting of T cells to antigens on malignant cells.

Checkpoint inhibition therapy—A form of cancer immunotherapy whereby antibodies against T-cell or antigen-presenting cell regulators of immune cell inhibition are used to activate cytotoxic T cells to kill tumor cells.

Co-stimulation of T cells—A secondary signal that T cells require for activation following presentation of peptide antigen by major histocompatibility complex (MHC) molecules to TCRs on either CD4 or CD8 T cells. A prime mediator of co-stimulation is the T cell CD28 molecule binding to B7-1 (CD80, CD86) on antigen-presenting cells.

Cytokines—Soluble proteins that interact with specific cellular receptors that are involved in the regulation of the growth and activation of immune cells and mediate normal and pathologic inflammatory and immune responses.

Immune homeostasis—Balanced protective immunity that does not overreact to pathogens and harm the host, and immunity that is not deficient and does not predispose the host to harmful infections or malignancies.

Immunoediting—Process of immunity selecting clones of cancer cells with reduced immunogenicities resulting in tumor escape.

Natural killer cells—Lymphocytes with cytotoxic potential for host cells with non-self-antigens, such as cells infected with pathogens or tumor cells expressing tumor-specific neoantigens.

T-cell exhaustion—State of T cells when the persistence of antigen disrupts memory T-cell function, resulting in defects in memory T-cell responses. Most frequently occurs in malignancies and in chronic viral infections such as HIV-1 and hepatitis C.

T regulatory cells (Tregs)—CD4 or CD8 T cells regulated by the transcription factor FOXP3 that play roles in downmodulating B- and T-cell responses in peripheral lymphoid tissues to prevent deleterious immune activation that can lead to autoimmune diseases.

Tumor-infiltrating lymphocytes—Lymphocytes that infiltrate tumors that may be in the exhausted state and upon which checkpoint inhibition therapy works.

Tumor neoantigens—Molecules in malignant cells that develop mutations that create non-self-antigens that are recognized by host T cells as non-self and against which tumor-infiltrating CD4 and CD8 T cells respond to reject the tumor.

INTRODUCTION

Immune homeostasis is the maintenance of balanced immunity that protects the host and prevents dysregulated immunity that predisposes the host to immunodeficiency, autoimmunity, malignancy, or harmful immune responses to infectious agents. Overactivity of innate and adaptive immunity leads to autoimmunity and/or inflammatory diseases. Underactivity of immune responses can lead to both immune deficiency and autoimmunity. Disruption of immune homeostasis contributes either directly or indirectly to many forms of disease. Thus, a major goal of immunotherapy is to either maintain immune homeostasis or reestablish immune balance in diseases of immune dysregulation.

This is an important time in the study of the biology of the immune system. The development of high-throughput genome sequencing, full genome transcriptome analysis, proteomics, metabolomics, and the realization of the importance of the human microbiome to immune system homeostasis have provided remarkable opportunities for development of new treatments for immune-mediated and malignant diseases. The emerging concept is that the immune system itself can be manipulated to be used as a therapeutic intervention for cancer and also can be safely manipulated for control of autoimmune disease. Moreover, new data are emerging that immune dysregulation is involved in the pathogenesis of diseases not traditionally thought of as immune-related, such as neurodegenerative diseases and atherosclerotic cardiovascular disease, and that the normal process of aging is associated with inflammatory processes.

Thus, a new frontier of medicine is to understand basic regulatory mechanisms of the innate and adaptive immune system with the goal of learning specific rules of immune system regulation, such that eventually the immune system can be "tuned" to safely stay in the zone of immune homeostasis and, at the same time, effectively protect against emerging and reemerging infectious diseases or eliminate malignancies as they arise. Moreover, when inflammatory or autoimmune diseases arise, effective strategies of immune regulation can be developed to safely treat them.

Thus, this chapter builds on **Chap. 360**, Introduction to the Immune System, to discuss T- and B-cell immunoregulation, to highlight recent successes in translating basic immunologic research into treatments of various hematopoietic and solid tumors, and to discuss mechanisms of immune dysregulation in autoimmunity and aging.

MECHANISMS OF REGULATION OF T-CELL ACTIVATION

Key roles of T cells are to respond to and eliminate cells bearing foreign antigens and to ignore cells expressing self-antigens. To respond to foreign antigens, stimulating cells must deliver an activating co-stimulating signal in addition to TCR ligation. To avoid responding to cells bearing self-antigens, T cells must also maintain immune tolerance. Central T-cell tolerance is maintained by autoreactive T-cell deletion in the thymus, while peripheral tolerance is either by cell apoptosis or anergy (cell intrinsic mechanisms), or maintained by extrinsic mechanism suppression by immune modulating cells such as regulatory T cells (Treg) (see section below, "Mechanisms of Immune Dysregulation in Autoimmune Disease"). Thus, T-cell activation is an integral component of the adaptive immune response to pathogens as well as to tumor neoantigens.

Two stimulatory signals are required for T-cell activation. One T-cell signal is delivered by antigen peptide presented in the context of MHC **(Chap. 360)**. However, in the absence of a co-stimulating signal, the T cell will not be activated by peptide MHC alone but rather will become anergic or unresponsive. A second co-stimulatory signal is needed for T-cell activation, resulting in cytoskeletal remodeling, production of cytokines, and cell survival and differentiation. In addition to the TCR/CD3 complex **(Chap. 360)**, T cells express a complex array of co-stimulatory as well as inhibitory molecules that bind to their respective receptors on the surface of antigen-presenting cells (APCs) and orchestrate both initiation and control of T-cell activation to maintain immune homeostasis **(Fig. 361-1, and see Chap. 360, Table 360-1)**. Of these, CD28, cytotoxic T lymphocyte antigen 4 (CTLA-4), and programmed cell death protein 1 (PD-1) and their ligands were among the first to be recognized to be central to control of T-cell activation.

The CD28 molecule is a member of the immunoglobulin (Ig) superfamily and is the original member of a subfamily of co-stimulatory or inhibitory molecules on the surface of T cells that includes CTLA-4, inducible T-cell costimulator (ICOS), PD-1, T-cell immunoreceptor with Ig and ITIM domains (TIGIT), and B- and T-cell attenuator (BTLA) (Fig. 361-1). CD28 stimulates intracellular signaling through AKT and PI3 kinase, resulting in induction of NF-κB, AP-1, and NFAT, which are all critical for T-cell activation and differentiation **(Chap. 360, Fig. 360-7)**. CTLA-4 and PD-1 are CD28 subfamily members that control T-cell activation by inhibiting the stimulating activity of CD28 and other T-cell co-stimulatory molecules.

CTLA-4 is a key negative regulator of T-cell activation that downmodulates the T-cell response to antigen by interaction with its ligands, B7-1 and B7-2, to control unchecked T-cell proliferation. CTLA-4 is upregulated following TCR engagement with MHC/peptide, thus

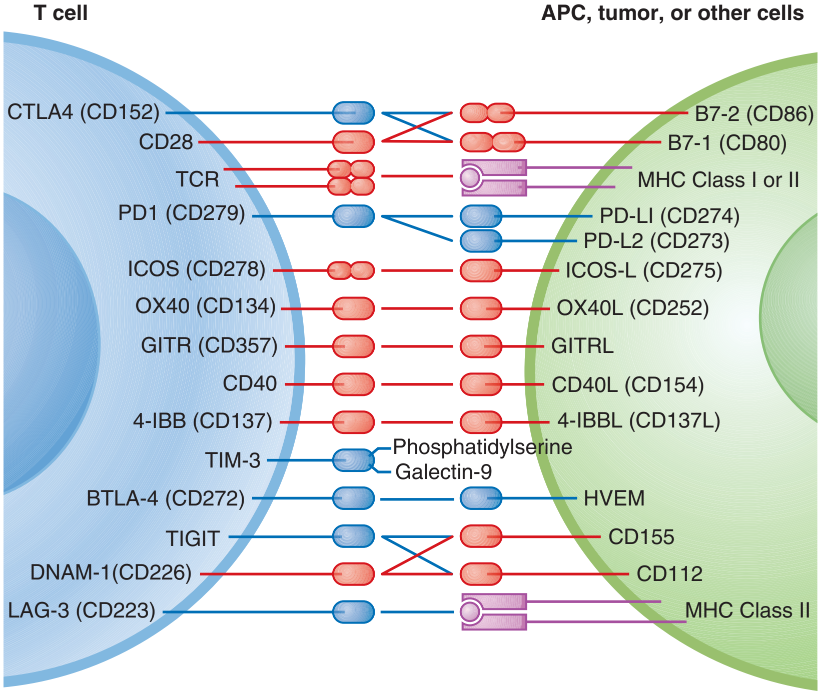

FIGURE 361-1 Regulatory stimulating or inhibiting molecules on T cells or antigen-presenting cells (APCs), tumor cells, or other cells.

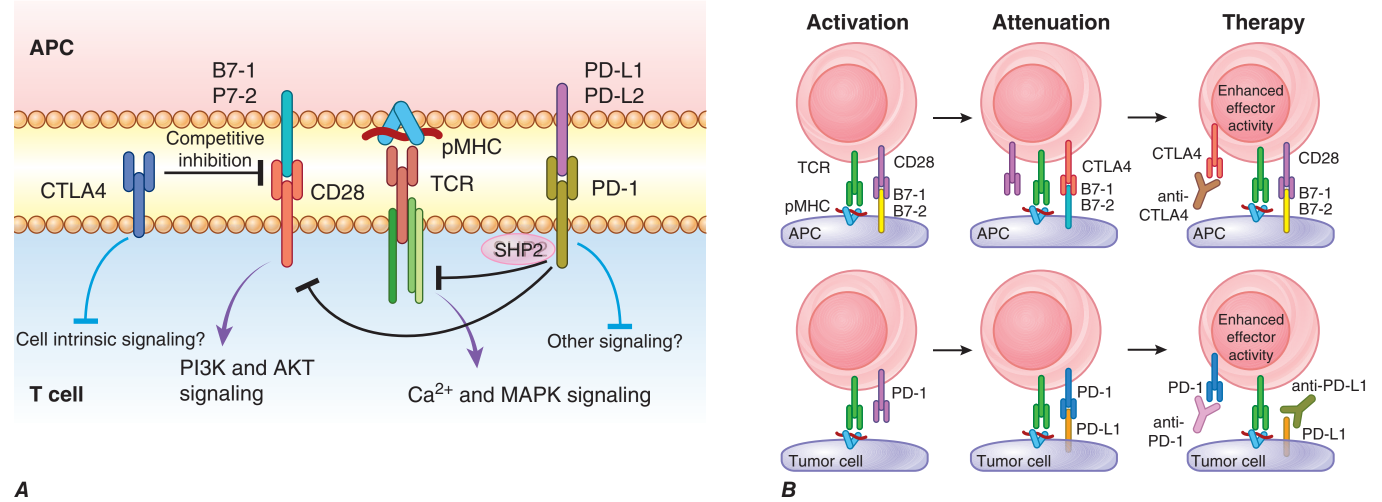

FIGURE 361-2 Molecular mechanisms of CTLA-4 and PD-1 attenuation of T-cell activation and schematic of the molecular mechanisms of action of CTLA-4 and PD-1 blockade. ***A.*** Schematic of the molecular interactions and downstream signaling induced by ligation of CTLA-4 and PD-1 by their respective ligands. The possibility of additional downstream cell-intrinsic signaling mechanisms is highlighted for both CTLA-4 and PD-1. ***B.*** The stepwise progression of T-cell activation, attenuation by normal regulatory mechanisms, and release of such negative regulation by therapeutic intervention using anti-CTLA-4 or anti-PD-1 antibodies is outlined. *(Reprinted from Cancer Discovery, 2018,8/9, 1069–1086, SC Wei et al: Fundamental Mechanisms of Immune Checkpoint Blockade Therapy, with permission from AACR.)*

dampening TCR signaling by competing with CD28 binding to B7 ligands (B7-1 [CD80] and B7-2 [CD86]) on APCs by virtue of higher affinity of CTLA-4 for B7 ligands. In this manner, T cells respond to foreign antigen but are prevented from damaging host tissues due to overexuberant responses **(Fig. 361-2A)**. CTLA-4 thus mediates its dampening effect on T-cell activation by competing with T-cell CD28 binding to B7 receptors, as well as through the suppressive effect of CTLA-4-positive Tregs. The human *CTLA4* gene is just one of several genes in which monogenetic mutations are associated with decreased Treg function and autoimmune syndromes **(Table 361-1)**.

PD-1 is also a major inhibitory molecule of T-cell activation by interaction with its ligands PD-L1 and PD-L2 on APCs (Fig. 361-1). While initially thought to be a cell death receptor, PD-1 is expressed upon activation of T and B cells and is also a marker of T follicular helper CD4+ T cells in B-cell germinal centers. PD-1 acts to dampen T-cell activation by dephosphorylation of CD28 via the Src homology region 2–containing protein tyrosine phosphatase 2 (SHP2).

Although a marker of immune cell activation, PD-1 is also a marker for exhausted T cells. T-cell exhaustion is an important mechanism of maintaining immune homeostasis and preventing host tissue T-cell

TABLE 361-1 Monogenetic Mutations That Lead to Immune Dysregulation and Autoimmunity

MUTATIONS AND FUNCTIONAL DEFICITS	DISEASES OR SYNDROME
RAG1, RAG2; lymphopenia with recombinase deficiency	Severe combined immune deficiency (Omenn's syndrome) with autoreactive T cells
Fas, FasL, CASP10; apoptosis defects	Autoimmune lymphoproliferative disease
AIRE, deletion chromosome 22q11.2; decrease in central tolerance	**Ch. 22q11.2:** DiGeorge's syndrome with autoimmune T cells **AIRE:** autoimmunity, polyneuropathy, candidiasis, ectodermal dysplasia (APECED syndrome)
FOXP3, CD25, CTLA-4, LRBA; decrease in peripheral immune tolerance with decrease in Treg function	**FOXP3:** IPEX syndrome (immune dysregulation, polyendocrinopathy, enteropathy, X-linked) **CD25:** enteropathy, dermatitis, autoimmunity, susceptibility to infections **CTLA-4:** associated with multiple autoimmune syndromes **LRBA:** infant with enteritis; hypogammaglobulinemia, autoimmune cytopenias
STAT-1, STAT-3; modulation of type 1 interferons	**STAT-1 deficiency:** decreased IFN-γ, susceptible to TB **STAT-1 gain of function:** chronic mucocutaneous candidiasis with autoimmune diseases **STAT-3 deficiency:** hyper-IgE syndrome (Job's syndrome) **STAT-3 gain of function:** lymphopenia, autoimmune cytopenias, diabetes, enteropathy
C1q, C1r/s, C2, C4; complement deficiencies	Systemic lupus erythematosus (SLE)
FcγRII, FcγRIII, C-reactive protein, complement receptor for C3bi (ITGAM or compliment receptor 3), COPA, tripeptidyl peptidase; lack of removal of cell debris	**FcγII, FcγIII, CRP, complement receptor for C3bi:** systemic lupus erythematosus **COPA:** autoimmune lung, renal, joint disease **Tripeptidyl peptidase II:** susceptibility to bacterial, viral, and fungal pathogens
Phosphoinositide-3-kinase delta (PI3Kδ), phospho-lipase Cγ2, protein kinase Cδ (PKCδ) deficiency, protein kinase Cδ (PKCδ) deficiency; hyperactivation of lymphocytes	**PI3Kδ:** lymphoproliferation, respiratory infections, hypogammaglobulinemia **Phospholipase Cγ2:** cold urticaria, antibody deficiency, autoimmunity **PKCδ:** early-onset SLE with decreased B-cell apoptosis, autoantibody-mediated renal disease and lymphoproliferation
Activation-induced cytidine deaminase (AID); class-impaired B-cell development	Hyper IgM syndrome type 2, low IgA, IgG, recurrent bacterial switch recombination; infections, autoimmune cytopenias, SLE

Abbreviations: APECED, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy; AIRE, autoimmune regulator; COPA, gene encodes the non-clathrin-coated vesicular coat protein, COP-alpha; CTLA-4, cytotoxic T lymphocyte associated protein-4; FcγR, Fcγ receptor; FOXP3, forkhead box P3; IFN, interferon; ITGAM, integrin alpha M; LRBA, lipopolysaccharide (LPS)-responsive and beige-like anchor protein; RAG, recombinase activating gene; STAT, signal transducer and activator of transcription; TB, tuberculosis; Treg, T regulatory cell.

Source: B Grimbacher et al: The crossroads of autoimmunity and immunodeficiency: Lessons from polygenic traits and monogenic defects. J Allergy Clin Immunol 137:3, 2016.

damage, but also leads to immune dysfunction in the setting of chronic antigenic stimulation such as occurs in chronic viral diseases (HIV-1, hepatitis C) and in cancer. Chronic viral diseases and tumors lead to transcriptional and metabolic signatures that define the exhausted T-cell state. T-cell exhaustion has been a major roadblock to overcome for successful cancer immunotherapy.

CHECKPOINT INHIBITION THERAPY FOR CANCER

The field of immune checkpoint therapy has joined surgery, radiation, chemotherapy, and targeted therapy as a mainstay for cancer therapy. There are now an extensive number of therapeutic monoclonal antibodies approved by the U.S. Food and Drug Administration (FDA), initially starting with melanoma in 2011–2014 and now available for a wide range of malignancies including kidney, lung, liver, head and neck, and gastric tumors **(Table 361-2)**. Both CTLA-4 and PD-1 blockade have proved remarkably successful in treating a number of tumors, but only in a fraction of patients. Because each receptor ligand pair regulates distinct T-cell inhibitory pathways, combination therapy using anti-PD-1 and anti-CTLA-4 antibodies has been especially useful and has induced significant tumor regression in ~50% of melanoma patients. What is different in the use of checkpoint inhibitor antibodies is that the therapy is not targeted to the tumor per se, but rather is targeted to immunoregulatory molecules on host T cells. Moreover, therapy is not targeted to specific molecules on tumors, but rather is targeted to release exhausted tumor-infiltrating T cells (TILs) to be activated to kill tumor cells by removing immune regulatory inhibition.

CTLA-4 induces tumor rejection by a number of mechanisms. First, anti-CTLA-4 antibody mediates direct blockade of CTLA-4 competition with CD28 for B7-1 and B7-2 co-stimulatory ligands, thus allowing CD28-mediated T-cell activation **(Fig. 361-2*B*)**. Tumor cells do not express B7 molecules; thus, CTLA-4 blockade likely occurs in tumor-draining lymph nodes where exhausted T cells interact with APCs presenting tumor neoantigens to T cells. A second mechanism of CTLA-4 blockade–induced tumor rejection is depletion or reduction in suppressive effects of Tregs. Suppressive effect of Tregs include secretion of immunosuppressive cytokines such as transforming growth factor (TGF)-β or interleukin- (IL) 10 or by direct inhibition of T-cell

TABLE 361-2 Summary of the Tumor Types for Which Immune Checkpoint Blockade Therapies Are Approved by the U.S. Food and Drug Administration (FDA)

TUMOR TYPE	THERAPEUTIC AGENT	TARGET	FDA APPROVAL YEAR
Melanoma	Ipilimumab	CTLA-4	2011
Melanoma	Nivolumab	PD-1	2014
Melanoma	Pembrolizumab	PD-1	2014
Non-small-cell lung cancer	Nivolumab	PD-1	2015
Non-small-cell lung cancer	Pembrolizumab	PD-1	2015
Melanoma (*BRAF* wild-type)	Ipilimumab + nivolumab	CTLA-4 + PD-1	2015
Melanoma (adjuvant)	Ipilimumab	CTLA-4	2015
Renal cell carcinoma	Nivolumab	PD-1	2015
Hodgkin's lymphoma	Nivolumab	PD-1	2016
Urothelial carcinoma	Atezolizumab	PD-L1	2016
Head and neck squamous cell carcinoma	Nivolumab	PD-1	2016
Head and neck squamous cell carcinoma	Pembrolizumab	PD-1	2016
Melanoma (any *BRAF* status)	Ipilimumab + nivolumab	CTLA-4 + PD-1	2016
Non-small-cell lung cancer	Atezolizumab	PD-L1	2016
Hodgkin's lymphoma	Pembrolizumab	PD-1	2017
Merkel cell carcinoma	Avelumab	PD-L1	2017
Urothelial carcinoma	Avelumab	PD-L1	2017
Urothelial carcinoma	Durvalumab	PD-L1	2017
Urothelial carcinoma	Nivolumab	PD-1	2017
Urothelial carcinoma	Pembrolizumab	PD-1	2017
MSI-high or MMR-deficient solid tumors of any histology	Pembrolizumab	PD-1	2017
MSI-high, MMR-deficient metastatic colorectal cancer	Nivolumab	PD-1	2017
Pediatric melanoma	Ipilimumab	CTLA-4	2017
Hepatocellular carcinoma	Nivolumab	PD-1	2017
Gastric and gastroesophageal carcinoma	Pembrolizumab	PD-1	2017
Non-small-cell lung cancer	Durvalumab	PD-L1	2018
Renal cell carcinoma	Ipilimumab + nivolumab	CTLA-4 + PD-1	2018
Endometrial cancer	Dostarlimab	PD-1	2021
Melanoma	Relatlimab + nivolumab	LAG-3 + PD-1	2022
Liver cancer	Tremelimumab	CTLA-4	2022
Merkle cell cancer	Retifanlimab	PD-1	2023
Nasopharyngeal carcinoma, esophageal squamous cell carcinoma	Toripalimab	PD-1	2023

Note: A summary of the tumor indications, therapeutic agents, and year of FDA approval for immune checkpoint blockade therapies. FDA approval includes regular approval and accelerated approval granted as of December 2023. Ipilimumab is an anti-CTLA-4 antibody. Nivolumab and pembrolizumab are anti-PD-1 antibodies. Atezolizumab, avelumab, and durvalumab are anti-PD-L1 antibodies. Tumor type reflects the indications for which treatment has been approved. Only the first FDA approval granted for each broad tissue type or indication for each therapeutic agent is noted. In cases where multiple therapies received approval for the same tumor type in the same year, agents are listed alphabetically.

Abbreviations: BRAF, v-raf murine sarcoma viral oncogene, homolog B1; MMR, mismatch repair; MSI, microsatellite instability.

Source: Reprinted from Cancer Discovery, 2018,8/9, 1069–1086, WC Spencer et al: Fundamental mechanisms of immune checkpoint blockade therapy, with permission from AACR.

proliferation and/or cytolytic activity. A third potential mechanism of anti-CTLA-4 blockade is remodeling and broadening of the peripheral TCR repertoire for tumor antigens. Evidence suggests that the effects of anti-CTLA-4 antibody are restricted primarily to tumor neoantigen-specific CD8 T cells within the tumor microenvironment and not to T cells in lymph nodes or spleen.

PD-1 blockade by PD-1 antibody also induces tumor regression by reversing T-cell exhaustion, leading to enhanced T-cell killing of tumor cells. Optimal PD-1 antibody-mediated checkpoint inhibition is seen when infiltrating CD8 T cells are present in the tumor microenvironment and reversal of the exhausted T-cell state can occur in situ. PD-1 blockade can also act by reversal of metabolic reprogramming of exhausted T cells, leading to enhanced cytolytic T-cell effector function. Antibodies to the primary PD-1 receptor, PD-L1, are also sufficient to induce reversal of T-cell exhaustion and induce tumor killing and are approved by the FDA for treatment of non-small-cell lung cancer (Table 361-2). PD-L1 is induced on tumor cells by T_H1 cytokines, which may explain anti-PD-L1 efficacy since T_H1 cytokines drive cytotoxic T-cell responses **(Chap. 360)**. Efficacy of anti-PD-L1 may also be due in part to mediation of antibody-dependent cellular cytotoxicity (ADCC) killing of tumor cells **(Chap. 360)**.

Tumor cells express neoantigens that are targets for T-cell recognition in the tumor microenvironment. Immune pressure within the tumor microenvironment can select for tumor cells that present few or mutated neoantigens and thus escape ongoing antitumor immunity. Moreover, a low number of tumor-infiltrating lymphocytes, tumor microenvironment production of immunosuppressive indoleamine 2,3-dioxygenase (IDO), and tumor infiltration with either myeloid-derived suppressor cells or Tregs can also limit checkpoint blockade therapy.

Other strategies for improving the immunoregulation of antitumor responses include combinations of anti-PD-1, anti-CTLA-4, or anti-PD-L1 antibodies with other checkpoint inhibitors (**see Chap. 360, Table 360-1,** and Fig. 361-1). For example, engagement of the ICOS pathway enhances the efficacy of CTLA-4 blockade in animal models of cancer immunotherapy. T-cell immunoglobulin and mucin-domain containing-3 (TIM-3), TIGIT, or lymphocyte-activation gene 3 (LAG-3) inhibition has been suggested to enhance checkpoint inhibition and augment CD8 tumor cell killing. A new transcription factor, thymocyte-selection-associated high mobility box (TOX), has been defined as a key controller of CD8 T-cell exhaustion. Thus, TOX inhibition could synergize with checkpoint inhibition therapy to reverse the T-cell exhaustion state. The combination of checkpoint inhibition with other cancer treatments including chemotherapy, radiation, tumor signaling pathway inhibitors, and epigenetic modulators is being tested.

Modified mRNAs have revolutionized the field of infectious disease vaccine development, and this technology is now being applied to cancer vaccine development. Tumor neoantigens are being sequenced from individual patients, and mRNAs that encode them are being studied for their ability to boost antitumor CD8 T-cell responses. Messenger RNAs that encode individualized tumor antigens are being tested with checkpoint inhibitors such as pembrolizumab (anti-PD-1) or with traditional chemotherapy.

Anti-PD-L1 antibodies also mediate antitumor effects by ADCC, which utilizes natural killer (NK) effector cells **(Chap. 360)**. Indeed, a number of checkpoint inhibitor molecules have been found to be expressed on NK cells including CTLA-4, PD-1, LAG-3, TIGIT, and TIM-3 (Fig. 361-1; **see also Chap. 360, Table 360-1**). A new field of work is to target NK cells with existing checkpoint inhibitors and with antibodies against an NK-specific inhibitory molecule, NKG2A, that are designed release NK cells to kill tumor cells. One such anti-NKG2A antibody, monalizumab, has entered human clinical trials. NK cells also express natural cytotoxicity-activating receptors including NKp30, NKp44, and NKp46 receptors **(Chap. 360)**. Engagement of natural cytotoxicity-activating receptors in concert with the NK FcγRIII (CD16) and antibody against a tumor antigen also can enhance NK cell targeting of tumor antigens and is in preclinical development.

An important consequence of checkpoint blockade for cancer treatment is the frequent occurrence of immune-related adverse events (irAEs). Clinically, irAEs occur most often at skin, lungs, and the gastrointestinal (GI) tract—barrier sites to invading antigens. Colitis is among the most common irAE. Endocrinopathies with organ-specific autoantibodies such as antithyroid or antipituitary antibodies have also been reported. While the pathophysiology of irAEs is not fully known, it is likely that both checkpoint inhibitor–released autoimmunity with decreased inhibition of Treg cell function and bystander activation of immune cells with nonspecific inflammation contribute to checkpoint inhibitor side effects. The makeup of the gut microbiome can affect the efficacy of checkpoint inhibition therapy, and in addition, certain microbiome bacterial strains are associated with the occurrence of irAEs.

CHIMERIC ANTIGEN RECEPTOR T CELLS

Chimeric antigen receptor (CAR) T cells are synthetic hybrid receptors created by recombinant techniques that combine an extracellular domain, usually derived from an antibody single-chain variable fragment (scFv), with intracellular signaling domains from activating co-stimulatory molecules (from endogenous TCRs, CD28, or 4-1BB) that allow for retargeting of T cells to antigens on malignant cells **(Fig. 361-3)**. A CAR T cell targeting the CD19 molecule on malignant B cells provided the first and most promising therapeutic results in the treatment of B-cell malignancies, with complete response rates of 70–90%. CAR T cells targeting the NY-ESO antigen on sarcoma cells have induced remissions in patients with synovial cell sarcoma. CAR T cells targeting B-cell maturation antigen (BCMA) on myeloma cells have also induced clinical responses. The CAR T-cell strategy is being developed for targeting solid tumors and modified as universal CAR T cells to overcome the need for MHC matching with T CAR recipients. One such strategy is to modify T cells to release cytokine, express co-stimulatory ligands, or secrete checkpoint-blocking single-chain variable fragment (scFvs). The next generation of CAR T cells are known as T cells redirected for universal cytokine-mediated killing (TRUCKs). Cytokine-secreting tumor-specific T cells could harness the adjuvant effect of recombinant cytokines by local delivery to the tumor to decrease cytokine side effects of excess inflammation, termed *cytokine release syndrome*, that can be seen with CAR T-cell therapy. Messenger RNAs that encode for tumor antigens are now being combined with CAR T cells with specificity for the same tumor antigens to expand tumor-specific CAR T cells and amplify their therapeutic effect.

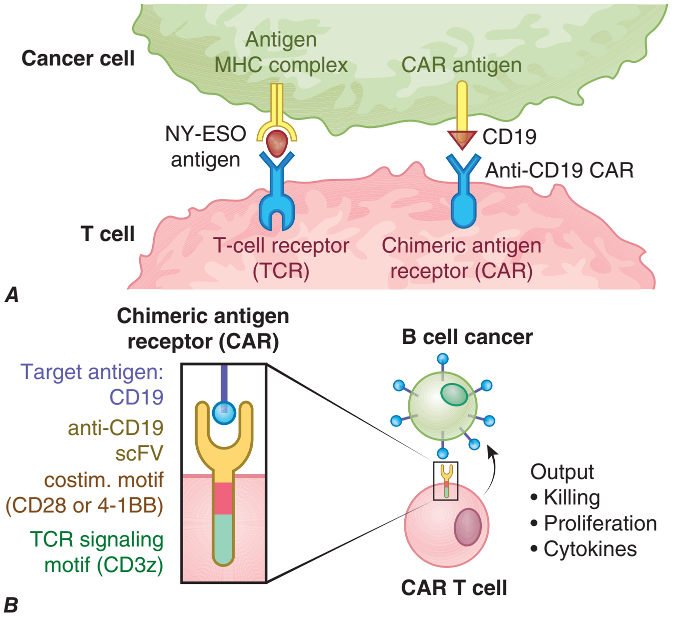

FIGURE 361-3 Platforms for redirecting T cells to cancer. ***A.*** T-cell receptors (TCRs; e.g., anti-NY-ESOI) or chimeric antigen receptors (CARs; e.g., anti-CD19 CAR). ***B.*** CAR structure includes an extracellular antigen recognition domain fused to intracellular TCR signaling domains (CD3z) and co-stimulatory domains (e.g., CD28 or 4-1BB). *(Reproduced with permission from WA Lim, CH June: The principles of engineering immune cells to treat cancer. Cell 168:724, 2017.)*

TABLE 361-3 Immune Tolerance Checkpoints in T- and B-Cell Immunity

CENTRAL TOLERANCE	PERIPHERAL TOLERANCE
THYMUS	**PERIPHERAL LYMPHOID TISSUES**
– TCR editing by V(D)J	– B- and T-cell anergy and inhibitory signaling (CTLA-4, PD-1, and other checkpoint molecules)
– Thymic negative selection	– TCR or BCR induction of BIM
– T-cell anergy and inhibitory signaling	– T-cell competition for IL-2, IL-7, IL-15, and peptide-MHC
– T regulatory cell differentiation	– B-cell competition for survival cytokine BAFF
	– T-cell growth dependence on CD28 ligands and other co-stimulatory molecules
	– Elimination of antigen-bearing dendritic cells by activated T cells producing perforin or FasL
	– Suppression of T- and B-cell responses by T regulatory cells and TGF-β, IL-10
	– T-cell death by FasL
	– Regulation of T follicular helper cell differentiation and function
BONE MARROW	– B-cell growth dependence on BCR ligands
– Immature B cell maturation arrest	– B-cell growth dependence on TCR ligands
– BCR editing by V(D)J recombination	– B-cell death by FasL on T cells
– Immature B cell deletion	– BCR modulation of plasma cell differentiation
	– BCR-induced death of germinal center B cells
	– Germinal center B-cell dependence on T follicular helper cells (CD40L, IL-21)

Abbreviations: BAFF, B cell activating factor; BCR, B-cell receptor; BIM, Bcl-2-like protein 11; FasL, Fas ligand that binds to the death receptor; IL, interleukin; MHC, major histocompatibility complex; TCR, T cell receptor; TGF-β, transforming growth factor beta; V(D)J, variable, diversity, joining regions of antibody V region.

Source: Adapted from CG Goodnow: Multistep pathogenesis of autoimmune disease. Cell 130:25, 2007.

MECHANISMS OF IMMUNE DYSREGULATION IN AUTOIMMUNE DISEASE

Autoimmune diseases occur in ~5% of people and are caused by immune dysregulation from breakdowns in immune tolerance. A complex array of immune checkpoints is involved in the maintenance of immune homeostasis and, when mutated, can result in autoimmune syndromes (Table 361-1 and **Table 361-3**). Central tolerance for deletion of autoreactive T cells or modification of their TCRs occurs in the thymus, and for B cells with self-reactive B-cell receptors, central deletion occurs in bone marrow. Peripheral tolerance occurs in lymph nodes, spleen, and tissue-associated lymphoid tissue such as gastrointestinal tract Peyer's patches. The sites and modes of peripheral tolerance are varied and reflect the complex cellular and cytokine interactions that occur in mediation of T- and B-cell adaptive immunity **(Chap. 360)**. While B- and T-cell deletion by immune cell apoptosis can occur in the periphery, peripheral tolerance can also occur with cell inactivation termed *anergy*, a state of immune unresponsiveness following contact with antigen. T- and B-cell responses are also dampened in the periphery by Tregs producing TGF-β and IL-10. The result of immune dysregulation in autoimmune disease is the production of a myriad of antibodies against self-antigens (autoantibodies), many of which are pathogenic for the clinical manifestations of the autoimmune disease **(Table 361-4)**.

Tregs are CD4 or CD8 T cells that downmodulate B- and T-cell responses in peripheral lymphoid tissues to prevent autoimmune diseases, and the transcriptional regulator FOXP3 is centrally involved in the establishment of the Treg phenotype. Mutations in genes that lead to loss of Tregs or their function result in autoimmune and inflammatory syndromes (Table 361-1). Mutations in FOXP3 lead to an X-linked syndrome characterized by immune dysregulation, polyendocrinopathy, and enteropathy (IPEX). Similarly, mutations in the CD25 (IL-2 receptor α) molecule expressed on Tregs lead to enteropathy, dermatitis, other manifestations of autoimmunity, and susceptibility to infections. Mutations in the checkpoint inhibitor T-cell molecule CTLA-4—also expressed on Tregs—lead to loss of Treg function and result in multiple autoimmune syndromes in humans depending on the CTLA-4 mutation. In mice, knockout of the *ctla4* gene leads to massive uncontrolled lymphoproliferation and early death. Finally, mutations in the lipopolysaccharide (lipopolysaccharide-responsive and beige-like anchor [LRBA]) protein cause a syndrome in infants characterized by enteritis, hypogammaglobulinemia, and autoimmune cytopenias.

Chronic viral infections can perturb Treg number and function. In HIV-1 infection, chronic antigenic stimulation leads to shifts in the B-cell repertoire toward an autoimmune permissive state, with increased numbers of autoreactive B cells and decreased CD4+ Tregs leading to serum autoantibodies or clinical manifestations of autoimmune disease in ~50% of untreated HIV-1-infected individuals.

In addition to checkpoint inhibition for cancer immunotherapy, monoclonal antibodies can be used for immune modulation to correct dysregulated immunity in autoimmune diseases to restore normal levels of immunoregulatory tolerance control. Monoclonal therapies have been developed and successfully used for the treatment of autoimmune and inflammatory diseases **(Table 361-5)**. Some of the monoclonal antibodies such as anti-CD20 (rituximab) have also been used for the treatment of B-cell malignancies. CTLA-4-Fc has been developed to prevent CD28-induced T-cell activation, resulting in immune suppression for rheumatoid arthritis (RA) and transplantation. TNF-α has been shown to play a central role in RA pathogenesis, and anti-TNF-α antibodies have been successful in treatment of RA and are approved for other autoimmune syndromes including other forms of arthritis, inflammatory bowel disease, and psoriasis. Antibodies against α4 integrin block the migration of α4β7+ T cells to the gastrointestinal tract and are used to treat inflammatory bowel disease (IBD). The T_H17 cytokine IL-17 has been found to be overproduced in psoriasis, and monoclonal anti-IL-17 antibody therapy for psoriasis is now approved by the FDA **(Chap. 60)**.

Tregs are therapeutic candidates for restoring immune tolerance in autoimmune and autoinflammatory diseases, with the prospect of reducing or replacing immunosuppressive drugs. Like CAR T cells, Treg therapy involves expanding autologous Treg cells in vitro and reinfusing them into individuals with autoimmune or inflammatory diseases. To make Treg therapy more targeted for suppression of antigen-specific immune responses, CAR T technology is being used to redirect Tregs to pathogenic T and B cells. Treg cellular therapy is in human clinical trials for the treatment of graft-versus-host disease in the setting of transplantation and for prevention of progression of type 1 diabetes mellitus **(Chap. 416)**.

IMMUNE DYSREGULATION IN AGING

Aging of the immune system in humans is characterized by decline in both innate and adaptive immunity. Aging is also paradoxically associated with a state of chronic inflammation, termed "inflammaging,"

TABLE 361-4 Recombinant or Purified Autoantigens Recognized by Autoantibodies Associated with Human Autoimmune Disorders

AUTOANTIGEN	AUTOIMMUNE DISEASES
Cell- or Organ-Specific Autoimmunity	
Acetylcholine receptor	Myasthenia gravis
Actin	Chronic active hepatitis, primary biliary cirrhosis
Adenine nucleotide translator (ANT)	Dilated cardiomyopathy, myocarditis
β-Adrenoreceptor	Dilated cardiomyopathy
Aromatic L-amino acid decarboxylase	Autoimmune polyendocrine syndrome type 1 (APS-1)
Asialoglycoprotein receptor	Autoimmune hepatitis
Bactericidal/permeability-increasing protein (Bpi)	Cystic fibrosis vasculitides
Calcium-sensing receptor	Acquired hypoparathyroidism
Cholesterol side-chain cleavage enzyme (CYP11a)	Autoimmune polyglandular syndrome 1
Citrullinated proteins	Rheumatoid arthritis
Collagen type IV-α3-chain	Goodpasture's syndrome
Cytochrome P450 2D6 (CYP2D6)	Autoimmune hepatitis
Desmin	Crohn's disease, coronary artery disease
Desmoglein 1	Pemphigus foliaceus
Desmoglein 3	Pemphigus vulgaris
F-actin	Autoimmune hepatitis
GM gangliosides	Guillain-Barré syndrome
Glutamate decarboxylase (GAD65)	Type 1 diabetes, stiff-person syndrome
Glutamate receptor (GLUR)	Rasmussen encephalitis
H/K ATPase	Autoimmune gastritis
17-α-Hydroxylase (CYP17)	Autoimmune polyglandular syndrome 1
21-Hydroxylase (CYP21)	Addison's disease
IA-2 (ICA512)	Type 1 diabetes
Insulin	Type 1 diabetes, insulin hypoglycemic syndrome (Hirata's disease)
Insulin receptor	Type B insulin resistance, acanthosis, systemic lupus erythematosus
Intrinsic factor type 1	Pernicious anemia
Leukocyte function-associated antigen (LFA-1)	Treatment-resistant Lyme arthritis
Myelin-associated glycoprotein (MAG)	Polyneuropathy
Myelin-basic protein	Multiple sclerosis, demyelinating diseases
Myelin oligodendrocyte glycoprotein (MOG)	Multiple sclerosis
Myosin	Rheumatic fever
p-80-Collin	Atopic dermatitis
Pyruvate dehydrogenase complex-E2 (PDC-E2)	Primary biliary cirrhosis
Sodium iodide symporter (NIS)	Graves' disease, autoimmune hypothyroidism
SOX-10	Vitiligo
Thyroid and eye muscle shared protein	Thyroid-associated ophthalmopathy
Thyroglobulin	Autoimmune thyroiditis
Thyroid peroxidase	Autoimmune Hashimoto's thyroiditis
Thyrotropin receptor	Graves' disease
Tissue transglutaminase	Celiac disease
Transcription coactivator p75	Atopic dermatitis
Tryptophan hydroxylase	Autoimmune polyglandular syndrome 1
Tyrosinase	Vitiligo, metastatic melanoma
Tyrosine hydroxylase	Autoimmune polyglandular syndrome 1
Systemic Autoimmunity	
Adrenocorticotropic hormone (ACTH)	ACTH deficiency
Aminoacyl-tRNA histidyl synthetase	Myositis, dermatomyositis
Aminoacyl-tRNA synthetase (several)	Polymyositis, dermatomyositis
Cardiolipin	Systemic lupus erythematosus, antiphospholipid syndrome
Carbonic anhydrase II	Systemic lupus erythematosus, Sjögren's syndrome, systemic sclerosis
Collagen (multiple types)	Rheumatoid arthritis, systemic lupus erythematosus, progressive systemic sclerosis
Centromere-associated proteins	Systemic sclerosis
DNA-dependent nucleoside-stimulated ATPase	Dermatomyositis
Fibrillarin	Scleroderma
Fibronectin	Systemic lupus erythematosus, rheumatoid arthritis, morphea
Glucose-6-phosphate isomerase	Rheumatoid arthritis
β2-Glycoprotein I (B2-GPI)	Primary antiphospholipid syndrome
Golgin (95, 97, 160, 180) heat shock protein	Sjögren's syndrome, systemic lupus erythematosus, rheumatoid arthritis, various immune-related disorders
Hemidesmosomal protein 180	Bullous pemphigoid, herpes gestationis, cicatricial pemphigoid
Histone H2A-H2B-DNA	Systemic lupus erythematosus
IgE receptor	Chronic idiopathic urticaria
Keratin-8	Rheumatoid arthritis
Ku-DNA-protein kinase	Systemic lupus erythematosus
Ku-nucleoprotein La phosphoprotein (La 55-B)	Connective tissue syndrome Sjögren's syndrome
Myeloperoxidase	Necrotizing and crescentic glomerulonephritis, systemic vasculitis
Proteinase 3 (PR3)	Granulomatosis with polyangiitis (Wegener's), Churg-Strauss syndrome
RNA polymerase I–III (RNP)	Systemic sclerosis, systemic lupus erythematosus
Signal recognition protein (SRP54)	Polymyositis
Topoisomerase-1 (Scl-70)	Scleroderma, Raynaud's syndrome
Tublin	Chronic liver disease, visceral leishmaniasis
Vimentin	Systemic autoimmune disease
Plasma Protein and Cytokine Autoimmunity	
C1 inhibitor	Autoimmune C1 deficiency
C1q	Systemic lupus erythematosus, membrane proliferative glomerulonephritis
Cytokines (IL-1α, IL-1β, IL-6, TNF-α, IFN-γ IL17A, IL-17F, GM-CSF)	IL-1α, IL-1β: rheumatoid arthritis, systemic sclerosis, systemic lupus erythematosus; IL-6: bacterial infections; IFN-γ: bacterial infections, varicella-zoster virus reactivation; IL-17A, IL-17F: chronic mucocutaneous candidiasis; GM-CSF: pulmonary alveolar proteinosis, fungal infections
Factor II, factor V, factor VII, factor VIII, factor IX, factor X, factor XI, thrombin vWF	Prolonged coagulation time
Glycoprotein IIb/IIIg and Ib/IX	Autoimmune thrombocytopenia purpura
IgA	Immunodeficiency associated with systemic lupus erythematosus, pernicious anemia, thyroiditis, Sjögren's syndrome, and chronic active hepatitis
Oxidized LDL (OxLDL)	Atherosclerosis
Cancer and Paraneoplastic Autoimmunity	
Amphiphysin	Neuropathy, small-cell lung cancer
Cyclin B1	Hepatocellular carcinoma
DNA topoisomerase II	Liver cancer

(*Continued*)

TABLE 361-4 Recombinant or Purified Autoantigens Recognized by Autoantibodies Associated with Human Autoimmune Disorders (Continued)

AUTOANTIGEN	AUTOIMMUNE DISEASES
Desmoplakin	Paraneoplastic pemphigus
Gephyrin	Paraneoplastic stiff-person syndrome
Hu proteins	Paraneoplastic encephalomyelitis
Neuronal nicotinic acetylcholine receptor	Subacute autonomic neuropathy, cancer
p53	Cancer, systemic lupus erythematosus
p62 (IGF-II mRNA-binding protein)	Hepatocellular carcinoma (China)
Recoverin	Cancer-associated retinopathy
Ri protein	Paraneoplastic opsoclonus myoclonus ataxia
βIV spectrin	Lower motor neuron syndrome
Synaptotagmin	Lambert-Eaton myasthenic syndrome
Voltage-gated calcium channels	Lambert-Eaton myasthenic syndrome
Yo protein	Paraneoplastic cerebellar degeneration

Abbreviations: ATPase, enzymes that catalyze hydrolysis of adenosine; Histone 2A and 2B, core histone proteins at the periphery of the nucleosome; Hu proteins, neuronal molecules; IGF-II, insulin-like growth factor-2; Ku, protein complex for DNA repair; La, Sjogren's syndrome antigen B (SSB) involved in RNA processing; RI, ribonuclease inhibitor; SOX-10, SRY-related HMG-box-10 transcription factor; Yo antigen, cellular degeneration-related antigen.

Source: From A Lernmark et al: J Clin Invest 108:1091, 2001; G Ceppelano et al: Am J Clin Exp Immunol 1:136, 2012; C-L Ku et al: Hum Genet 139:783, 2020.

TABLE 361-5 Monoclonal Antibodies Approved for Clinical Use in Autoimmune Disease, Some of Which Are Also Used in Malignancies[a]

TARGET MOLECULE	FUNCTION	FDA-APPROVED mAbs, TRAPS, AND BISPECIFIC mAbs	AUTOIMMUNE/ INFLAMMATORY	MALIGNANCY	OTHER/COMMENTS
CD52	Marker of T and B lymphocyte subsets	Alemtuzumab (Lemtrada)	Multiple sclerosis	Chronic lymphocytic leukemia	Trade name changed from Campath-1H (cancer) to Lemtrada (multiple sclerosis) Humanized IgG1k
CD25	Alpha-chain of the IL-2 receptor	Basiliximab (Simulect) Daclizumab (Zenapax)	Multiple sclerosis Transplant rejection		Basiliximab: chimeric mouse/ human IgG1k Daclizumab: humanized IgG1
CD20	Participates in B-cell differentiation	Obinutuzumab (Gazyva) Ibritumomab tiuxetan (Zevalin) Tositumomab (Bexxar)[a] Ofatumumab (Arzerra) Rituximab (Rituxan)	Rheumatoid arthritis	B-cell malignancies	Obinutuzumab is the first approved glycoengineered IgG1 mAb with enhanced ADCC Rituximab is chimeric mouse/ human IgG1k Ibritumomab tiuxetan and tositumomab are radio-conjugates that can be used when tumors stop responding to the anti-CD20 mAbs Ibritumomab and tositumomab are mouse IgG2a
α4 β7 integrin (LPAM)	Block immune cell migration	Vedolizumab (Entyvio)	Ulcerative colitis, Crohn's disease		Humanized IgG1
CD80/CD86	Provide co-stimulatory signals necessary for T-cell activation and survival; ligand trap prevents activation of CD28 immune checkpoint resulting in immune suppression	Belatacept (Nulojix) Abatacept (Orencia) (both CTLA-4-Fc fusion proteins)	Rheumatoid arthritis Transplant rejection		
IGF-IR	Human IgG1; binds insulin-like growth factor-1 receptor on orbital fibroblasts and decreases inflammatory cytokine production	Teprotumumab (Tepezza)	Thyroid eye disease		
TNF-α	Inflammatory cytokine that drives multiple autoimmune diseases[a]	Adalimumab (Humira) Certolizumab pegol (Cimzia) Golimumab (Simponi) Infliximab (Remicade) Etanercept (Enbrel) There are more than 20 anti-TNF biosimilars in various stages of development. Already approved are infliximab biosimilars (Remsima, Inflectra, Flixabi), etanercept biosimilars (Erelzi, Benepali), and adalimumab biosimilars (Amjevita). This field will change very rapidly as many dossiers are now under regulatory scrutiny. Biosimilars are given a suffix, e.g., etanercept-szzs (Erelzi)	Crohn's disease, ulcerative colitis, RA, juvenile idiopathic arthritis, psoriatic arthritis, ankylosing spondylitis, plaque psoriasis, hidradenitis suppurativa, uveitis		Not all TNF blockers are approved for all indications Drugs in italics are biosimilars: adalimumab, infliximab, infliximab, *Inflectra*, *adalimumab-atto* are IgG1k mAbs; *certolizumab pegol* is a pegylated Fab fragment; etanercept is a soluble Fc:TNF receptor trap that binds TNF (there is one biosimilar). At least four biosimilar TNF blockers have been approved in the EU (two each for infliximab and etanercept)

(*Continued*)

TABLE 361-5 Monoclonal Antibodies Approved for Clinical Use in Autoimmune Disease, Some of Which Are Also Used in Malignancies[a] (Continued)

TARGET MOLECULE	FUNCTION	FDA-APPROVED mAbs, TRAPS, AND BISPECIFIC mAbs	AUTOIMMUNE/ INFLAMMATORY	MALIGNANCY	OTHER/COMMENTS
VEGF	Cytokine that stimulates vasculogenesis and angiogenesis. Overproduced in some inflammatory disorders and tumors to induce increased blood supply.	Bevacizumab (Avastin) Ramucirumab (Cyramza) Aflibercept (Eylea/Zaltrap) Ranibizumab (Lucentis)	Age-related macular degeneration, macular edema, diabetic macular edema, diabetic retinopathy	Colorectal cancer, nonsquamous NSCLC, breast cancer, glioblastoma, renal cell carcinoma, gastric cancer or gastroesophageal junction adenocarcinoma	Bevacizumab and ramucirumab are IgG1 mAbs for cancer therapy (ramucirumab was derived from phage display); ranibizumab is a Fab fragment (single arm binder). It has a short half-life if administered intravenously, but it is stable when locally injected into the eye. Aflibercept is a ligand trap with optical (Eylea) and cancer (Zaltrap) applications.
IL-4 receptor alpha subunit	Receptor that mediates IL-4 and IL-13-induced inflammation	Dupilumab (Dupixent)	Atopic dermatitis (eczema), steroid-dependent asthma		
IgE	Binds to mast cells, basophils, and other cells that express Fc-epsilon receptor and induces release of inflammatory cytokines	Omalizumab (Xolair)	Asthma		IgG1k mAb also used off-label to treat IgE-related conditions (allergic rhinitis, drug allergies, other)
Alpha-4 integrin	Alpha-4 integrin facilitates exit of inflammatory cells from blood into intestine or across the blood-brain barrier	Vedolizumab (Entyvio) Natalizumab (Tysabri)	Multiple sclerosis, Crohn's disease, and ulcerative colitis		IgG4 natalizumab therapy has been associated with PML caused by John Cunningham virus in immunocompromised patients. IgG1k vedolizumab may not be associated with PML.
Complement C5	Inhibits complement cascade	Eculizumab (Solaris)	Prevents destruction of red blood cells by activated complement (paroxysmal nocturnal hemoglobinuria)		IgG2/4 mAb
		Pozelimab (Vepoz)	Treats complement hyperactivation, angiopathic thrombosis, and protein-losing enteropathy (CHAPLE)		Pozelimab, IgG4
FcRn	Delivers IgG antibodies to tissues	Rozanolixizumab (Rustiggo)	Myasthenia gravis		IgG4. Rozanolixizumab binds to human neonatal Fc receptor (FcRn), thus reducing the concentration of pathogenic IgG autoantibodies.
IL-1	Mutations in cryopyrin lead to overproduction of IL-1 and inflammatory disease; IL-1 also drives other inflammatory diseases	Canakinumab (Ilaris) Rilonacept (Arcalyst)[a]	Rare inflammatory syndromes, active juvenile arthritis, gouty arthritis		Canakinumab is an IgG1k mAb; rilonacept is an IL-1 trap designed from the IL-1R fused with human mAb Fc region.
IL-6	Overexpression of IL-6 is associated with multiple malignancies	Siltuximab (Sylvant)		Pseudo-malignancy: Castleman's disease (similar to lymphoma)	Murine/human chimeric IgG1κ
IL-6R (IL-6 receptor)	Current approvals based on role of IL-6 in promoting inflammatory autoimmune disease	Tocilizumab (Actemra)	Rheumatoid arthritis, polyarticular juvenile arthritis, juvenile idiopathic arthritis		Human/mouse chimeric mAb with initial approval for efficacy in RA after failure of TNF blocker
BAFF (tumor necrosis factor superfamily member 13b)	Role in proliferation and differentiation of B cells	Belimumab (Benlysta)	Systemic lupus erythematosus		IgG1-γ/λ
SLAMF7/CD319	SLAMF7 triggers the activation and differentiation of a wide variety of immune cells (innate and adaptive immune response) perhaps primarily mediated by natural killer cells and myeloma cells	Elotuzumab (Empliciti)		Multiple myeloma	This IgG1k mAb is thought to activate SLAMF7 receptor and to have a secondary mechanism of mediating ADCC vs multiple myeloma cells

(*Continued*)

TABLE 361-5 Monoclonal Antibodies Approved for Clinical Use in Autoimmune Disease, Some of Which Are Also Used in Malignancies[a] (Continued)

TARGET MOLECULE	FUNCTION	FDA-APPROVED mAbs, TRAPS, AND BISPECIFIC mAbs	AUTOIMMUNE/ INFLAMMATORY	MALIGNANCY	OTHER/COMMENTS
IL-5	Induces differentiation and survival of eosinophils	Reslizumab (Cinqair) Mepolizumab (Nucala)	Asthma		Both IgG1k mAbs
IL-13	Inflammatory cytokine	Lebrikizumab (Ebglyss) Tralokinumab (Adtralza; Adbry)	Atopic dermatitis		IgG4
IL-17A	Inflammatory cytokine	Ixekizumab (Taltz) Secukinumab (Cosentyx)	Plaque psoriasis, ankylosing spondylitis		Ixekizumab is an IgG4; secukinumab is an IgG1k
IL-23p19		Mirikizumab (Omvoh) Risankizumab (Skyrizi) Guselkumab (Tremfya)	Ulcerative colitis Plaque psoriasis Plaque psoriasis		Mirikizumab humanized IgG4 Risankizumab humanized IgG4 Guselkumab humanized IgG4

[a]Approved for use in United States only.

Note: An actively updated summary of MAb approvals can be found at *www.antibodysociety.org*. mAbs can be murine, chimeric (human Fc region), humanized, or human; traps are derived from receptors and compete with natural receptor for binding target; bispecific mAbs are engineered to bind to two different targets simultaneously (usually to bring immune cell into contact with target cell, thereby triggering target cell killing). Antibody-drug conjugate (X): toxin or radioisotope attached to mAb to increase efficacy. Agents approved for use in the United States only are noted; others are approved in both United States and United Kingdom.

Abbreviations: ADCC, antibody-dependent cell-mediated cytotoxicity; BAFF, B-cell activating factor; CTLA-4, cytotoxic T-cell lymphocyte-associated protein-4; EU, European Union; Ig, immunoglobulin; FDA, U.S. Food and Drug Administration; IL, interleukin; mAb, monoclonal antibody; NSCLC, non-small-cell lung cancer; PML, progressive multifocal leukoencephalopathy; RA, rheumatoid arthritis; SLAMF7, signaling lymphocyte activation molecule family member 7; TNF, tumor necrosis factor; VEGF, vascular endothelial growth factor.

Source: Reproduced with permission from HM Shepard et al: Developments in therapy with monoclonal antibodies and related proteins. Clin Med (Lond) 17:220, 2017.

with an increased risk of autoimmune disease. Aging is associated with reduced NK cell function, reduced monocyte/macrophage immune cell expression of toll-like receptors, reduced chemotaxis and phagocytosis, and reduced MHC expression and signaling. Other phagocytic cells such as polymorphonuclear cells are similarly dysfunctional. Dendritic cells in aged individuals are present in reduced numbers with impaired antigen-presenting function and signaling. Adaptive immunity is similarly impaired with decreased antibody repertoire breadth, decrease in number of B cells, and decrease in B-cell responses to specific antigens. Similarly, T-cell responses to antigen are decreased, such as to seasonal influenza vaccination.

Aging is characterized by accumulation of senescent cells in many tissues that secrete inflammatory cytokines, chemokines, and other inflammatory mediators. The best example of the role of enhanced cytokine production in the "inflammaging" syndrome is in thymic atrophy, which is a major event contributing to age-associated immune system decline. During life, the thymus decreases in size and naïve T-cell output decreases; beginning with puberty, thymocytes progressively decrease in number, such that after ~50 years of age, ~90% of thymocytes have been replaced by adipocytes in the thymus perivascular space. Thymic adipocytes produce leukemia inhibitory factor, oncostatin M, IL-6, and stem cell factor (SCF). Administration of these cytokines to young mice induces thymic atrophy, demonstrating that these thymosuppressive cytokines actively induce thymocyte loss. Adipocytes in other sites also produce inflammatory cytokines, contributing to tissue senescence. Finally, respiratory failure due to SARS-CoV-1, Middle Eastern respiratory syndrome, or SARS-CoV-2 infection is associated with a cytokine release syndrome with IL-6 overproduction, which occurs most frequently in older individuals.

FURTHER READING

Baker D et al: CAR T therapy beyond cancer: The evolution of a living drug. Nature 619:707, 2023.

Burke KP et al: Immune checkpoint receptors in autoimmunity. Curr Opin Immunol 80:102283, 2023.

Ferreira LMR et al: Next-generation regulatory T cell therapy Nat Rev Drug Discov 18:749, 2019.

Kuchroo JR et al: The double-edged sword: Harnessing PD-1 blockade in tumor and autoimmunity. Sci Immunol 6:65, 2021.

Lim WA, June CH: The principles of engineering immune cells to treat cancer. Cell 168:724, 2017.

Schildberg FA et al: Coinhibitory pathways in the B7-CD28 ligand-receptor family. Immunity 44:955, 2016.

Sharma P, Allison JP: Dissecting the mechanisms of immune checkpoint therapy. Nature Rev Immunol 20:75, 2020.

Sharpe A, Pauken KE: The diverse functions of the PD-1 pathway. Nat Rev Immunol 18:153, 2018.

362 Primary Immune Deficiency Diseases

Alain Fischer

Immunity is intrinsic to life and an important tool in the fight for survival against pathogenic microorganisms. The human immune system can be divided into two major components: the innate immune system and the adaptive immune system **(Chap. 360)**. The innate immune system provides the rapid triggering of inflammatory responses based on the recognition (at the cell surface or within cells) of either molecules expressed by microorganisms or molecules that serve as "danger signals" released by cells under attack. These receptor/ligand interactions trigger signaling events that ultimately lead to inflammation. Virtually all cell lineages (not just immune cells) are involved in innate immune responses; however, myeloid cells (i.e., neutrophils and macrophages) play a major role because of their phagocytic capacity. The adaptive immune system operates by clonal recognition of antigens followed by a dramatic expansion of antigen-reactive cells and execution of an immune effector program. Most of the effector cells die off rapidly, whereas memory cells persist. Although both T and B lymphocytes recognize distinct chemical moieties and execute distinct adaptive immune responses, the latter is largely dependent on the former in generating long-lived humoral immunity. Adaptive responses utilize components of the innate immune system; for example, the antigen-presentation capabilities of dendritic cells help to determine the type of effector response. Not surprisingly, immune responses are controlled by a series of regulatory mechanisms.

Hundreds of gene products have been characterized as effectors or mediators of the immune system **(Chap. 360)**. Whenever the expression or function of one of these products is genetically impaired (provided the function is nonredundant), a primary immunodeficiency (PID) occurs.

PIDs are genetic diseases with primarily Mendelian inheritance. More than 550 conditions have now been described, and deleterious mutations in ~500 genes have been identified. The overall prevalence of PIDs has been estimated in various countries at 5–10 per 100,000 individuals; however, given the difficulty in diagnosing these rare and complex diseases, this figure is probably an underestimate. PIDs can involve all possible aspects of immune responses, from innate through adaptive, cell differentiation, and effector function and regulation. For the sake of clarity, PIDs should be classified according to (1) the arm of the immune system that is defective and (2) the mechanism of the defect (when known). **Table 362-1** classifies the most prevalent PIDs according to this manner of classification; however, one should bear in mind that the classification of PIDs sometimes involves arbitrary decisions because of overlap and, in some cases, lack of data.

TABLE 362-1 Classification of Primary Immune Deficiency Diseases

Deficiencies of the Innate Immune System	
• Phagocytic cells:	
• Impaired production: severe congenital neutropenia (SCN)	
• Asplenia	
• Impaired adhesion: leukocyte adhesion deficiency (LAD)	
• Impaired killing: chronic granulomatous disease (CGD)	
• Innate immunity receptors and signal transduction:	
• Defects in Toll-like receptor signaling	
• Mendelian susceptibility to mycobacterial disease	
• Complement deficiencies:	
• Classical, alternative, and lectin pathways	
• Lytic phase	
Deficiencies of the Adaptive Immune System	
• T lymphocytes:	
• Impaired development	Severe combined immune deficiencies (SCIDs)
• Impaired survival, migration, function	DiGeorge's syndrome
	Combined immunodeficiencies
	Hyper-IgE syndrome (autosomal dominant)
	DOCK8 deficiency
	CD40 ligand deficiency
	Wiskott-Aldrich syndrome
	Ataxia-telangiectasia and other DNA repair deficiencies
• B lymphocytes:	
• Impaired development	XL and AR agammaglobulinemia
• Impaired function	Hyper-IgM syndrome
	Common variable immunodeficiency (CVID)
	IgA deficiency
Regulatory Defects	
• Innate immunity	Autoinflammatory syndromes (outside the scope of this chapter)
	Severe colitis
	Hemophagocytic lymphohistiocytosis (HLH)
• Adaptive immunity	Autoimmune lymphoproliferation syndrome (ALPS)
	Autoimmunity and inflammatory diseases (IPEX, APECED)

Abbreviations: APECED, autoimmune polyendocrinopathy candidiasis ectodermal dysplasia; AR, autosomal recessive; IPEX, immunodysregulation polyendocrinopathy enteropathy X-linked syndrome; XL, X-linked.

Today, PIDs tend to be integrated in a larger group of inborn errors of immunity that also include autoinflammatory syndromes.

The consequences of PIDs vary widely as a function of the molecules that are defective. This concept translates into multiple levels of vulnerability to infection by pathogenic and opportunistic microorganisms, ranging from extremely broad (as in severe combined immunodeficiency [SCID]) to narrowly restricted to a single microorganism (as in Mendelian susceptibility to mycobacterial disease [MSMD]). The locations of the sites of infection and the causal microorganisms involved will thus help physicians arrive at proper diagnoses. PIDs can also lead to immunopathologic responses such as allergy (as in Wiskott-Aldrich syndrome [WAS]), lymphoproliferation, and autoimmunity. A combination of recurrent infections, inflammation, and autoimmunity can be observed in a number of PIDs, thus creating obvious therapeutic challenges. Finally, some PIDs increase the risk of cancer, notably but not exclusively lymphocytic cancers, for example, lymphoma.

DIAGNOSIS OF PRIMARY IMMUNODEFICIENCIES

The most frequent symptom prompting the diagnosis of a PID is the presence of recurrent or unusually severe infections. As mentioned above, recurrent allergic or autoimmune manifestations may also alert the physician to a possible diagnosis of PID. In such cases, a detailed account of the subject's personal and family medical history should be obtained. It is of the utmost importance to gather as much medical information as possible on relatives and up to several generations of ancestors. In addition to the obvious focus on primary symptoms, the clinical examination should evaluate the size of lymphoid organs and, when appropriate, look for the characteristic signs of a number of complex syndromes that may be associated with a PID.

The performance of laboratory tests should be guided to some extent by the clinical findings. Infections of the respiratory tract (bronchi, sinuses) mostly suggest a defective antibody response. In general, invasive bacterial infections can result from complement deficiencies, signaling defects of innate immune responses, asplenia, or defective antibody responses. Viral infections, recurrent *Candida* infections, and opportunistic infections are generally suggestive of impaired T-cell immunity. Skin infections and deep-seated abscesses primarily reflect innate immune defects (such as chronic granulomatous disease); however, they may also appear in the autosomal dominant hyper-IgE syndrome. **Table 362-2** summarizes the laboratory tests that are most frequently used to diagnose a PID. More specific tests (notably genetic tests) are then used to make a definitive diagnosis. Genomic tools now allow us to more efficiently track genetic defects through usage of gene panel resequencing and/or whole exome/genome sequencing.

The PIDs discussed below have been grouped together according to the affected cells and the mechanisms involved (Table 362-1, **Fig. 362-1**).

PRIMARY IMMUNODEFICIENCIES OF THE INNATE IMMUNE SYSTEM

PIDs of the innate immune system are relatively rare and account for ~10% of all PIDs.

■ SEVERE CONGENITAL NEUTROPENIA

Severe congenital neutropenia (SCN) consists of a group of inherited diseases that are characterized by severely impaired neutrophil counts (<500 polymorphonuclear leukocytes [PMN]/μL of blood). The condition is usually manifested from birth. SCN may also be cyclic (with a 3-week periodicity), and other neutropenia syndromes can also be intermittent. Although the most frequent inheritance pattern for SCN is autosomal dominant, autosomal recessive and X-linked recessive conditions also exist. Bacterial infections at the interface between the body and the external milieu (e.g., the orifices, wounds, and the respiratory tract) are common manifestations. Bacterial infections can rapidly progress through soft tissue and are followed by dissemination in the bloodstream. Severe visceral fungal infections can also ensue. The absence of pus is a hallmark of this condition.

Diagnosis of SCN requires examination of the bone marrow. Most SCNs are associated with a block in granulopoiesis at the promyelocytic

TABLE 362-2 Tests Most Frequently Used to Diagnose a Primary Immunodeficiency (PID)

TEST	INFORMATION	PID DISEASE
Blood cell counts and cell morphology	Neutrophil counts[a] Lymphocyte counts[a] Eosinophilia Howell-Jolly bodies	↓ Severe congenital neutropenia, ↑↑ LAD T-cell ID WAS, hyper-IgE syndrome Asplenia
Chest x-ray	Thymic shadow Costochondral junctions	SCID, DiGeorge's syndrome Adenosine deaminase deficiency
Bone x-ray	Metaphyseal ends	Cartilage hair hypoplasia
Immunoglobulin serum levels	IgG, IgA, IgM IgE	B-cell ID Hyper-IgE syndrome, WAS, T-cell ID
Lymphocyte phenotype	T, B lymphocyte counts	T-cell ID, agammaglobulinemia
Dihydrorhodamine fluorescence (DHR) assay Nitroblue tetrazolium (NBT) assay	Reactive oxygen species production by PMNs	Chronic granulomatous disease
CH50, AP50	Classic and alternative complement pathways	Complement deficiencies
Ultrasonography of the abdomen	Spleen size	Asplenia

[a]Normal counts vary with age. For example, the lymphocyte count is between 3000 and 9000/μL of blood below the age of 3 months and between 1500 and 2500/μL in adults.

Abbreviations: ID, immunodeficiency; LAD, leukocyte adhesion deficiency; PMNs, polymorphonuclear leukocytes; SCID, severe combined immunodeficiency; WAS, Wiskott-Aldrich syndrome.

stage (Fig. 362-1). SCN has multiple etiologies, and to date, mutations in 21 different genes have been identified. Most of these mutations result in isolated SCN, whereas others are syndromic **(Chap. 67)**. The most frequent forms of SCN are caused by the premature cell death of granulocyte precursors, as observed in deficiencies of GFI1, HAX1, and elastase 2 (*ELANE*), with the latter accounting for 50% of SCN sufferers. Certain *ELANE* mutations cause cyclic neutropenia syndrome. A gain-of-function mutation in the *WASP* gene (see the section "Wiskott-Aldrich Syndrome" below) causes X-linked SCN, which is also associated with monocytopenia.

As mentioned above, SCN exposes the patient to life-threatening, disseminated bacterial and fungal infections. Treatment requires careful hygiene measures, notably in infants. Later in life, special oral and dental care is essential, along with the prevention of bacterial infection by prophylactic administration of trimethoprim/sulfamethoxazole. Subcutaneous injection of the cytokine granulocyte colony-stimulating factor (G-CSF) usually improves neutrophil development and thus prevents infection in most SCN diseases. However, there are two caveats: (1) a few cases of SCN with *ELANE* mutation are refractory to G-CSF and may require curative treatment via allogeneic hematopoietic stem cell transplantation (HSCT); and (2) a subset of G-CSF-treated patients carrying *ELANE* mutations are at a greater risk of developing acute myelogenous leukemia associated (in most cases) with somatic gain-of-function mutations of the G-CSF receptor gene.

A few SCN conditions are associated with additional immune defects involving leukocyte migration as observed in the warts, hypogammaglobulinemia, infections, and myelokathexis (WHIM) syndrome (gain-of-function mutation of the chemokine *CXCR4*) or in moesin deficiency.

■ ASPLENIA

Primary failure of the development of a spleen is an extremely rare disease that can be either syndromic (in Ivemark syndrome) or isolated

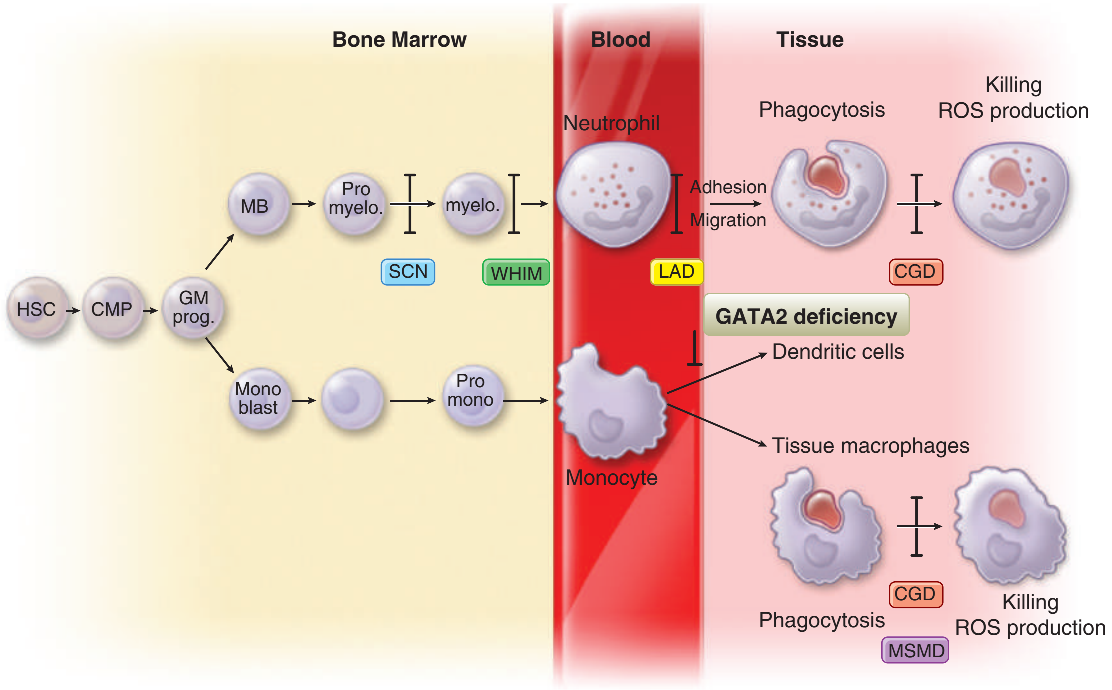

FIGURE 362-1 Differentiation of phagocytic cells and related primary immunodeficiencies (PIDs). Hematopoietic stem cells (HSCs) differentiate into common myeloid progenitors (CMPs) and then granulocyte-monocyte progenitors (GM prog.), which, in turn, differentiate into neutrophils (MB: myeloblasts; Promyelo: promyelocytes; myelo: myelocytes) or monocytes (monoblasts and promonocytes). Upon activation, neutrophils adhere to the vascular endothelium, transmigrate, and phagocytose the targets. Reactive oxygen species (ROS) are delivered to the microorganism-containing phagosomes. Macrophages in tissues kill using the same mechanism. Following activation by interferon γ (not shown here), macrophages can be armed to kill intracellular pathogens such as mycobacteria. For sake of simplicity, not all cell differentiation stages are shown. The abbreviations for PIDs are contained in boxes placed at corresponding stages of the pathway. CGD, chronic granulomatous diseases; GATA2, zinc finger transcription factor; LAD, leukocyte adhesion deficiencies; MSMD, Mendelian susceptibility to mycobacterial disease; SCN, severe congenital neutropenia; WHIM, warts, hypogammaglobulinemia, infections, and myelokathexis.

with an autosomal dominant expression; in the latter case, mutations in the ribosomal protein SA were recently found. Due to the absence of natural filtration of microbes in the blood, asplenia predisposes affected individuals to fulminant infections by encapsulated bacteria. Although most infections occur in the first years of life, cases may also arise in adulthood. The diagnosis is confirmed by abdominal ultrasonography and the detection of Howell-Jolly bodies in red blood cells. Effective prophylactic measures (twice-daily oral penicillin and appropriate vaccination programs) usually prevent fatal outcomes.

■ GATA2 DEFICIENCY

Recently, an immunodeficiency combining monocytopenia and dendritic and lymphoid (B and natural killer [NK]) cell deficiency (DCML), also called monocytopenia with nontuberculous mycobacterial infections (mono-MAC), has been described as a consequence of a dominant mutation in the gene *GATA2*, a transcription factor involved in hematopoiesis. This condition also predisposes to lymphedema, myelodysplasia, and acute myeloid leukemia. Infections (bacterial and viral) are life-threatening, thus indicating, together with the malignant risk, HSCT.

■ LEUKOCYTE ADHESION DEFICIENCY

Leukocyte adhesion deficiency (LAD) consists of three autosomal recessive conditions (LAD I, II, and III) **(Chap. 67)**. The most frequent condition (LAD I) is caused by mutations in the β2 integrin gene; following leukocyte activation, β2 integrins mediate adhesion to inflamed endothelium expressing cognate ligands. LAD III results from a defect in a regulatory protein (kindlin, also known as Fermt 3) involved in activating the ligand affinity of β2 integrins. The extremely rare LAD II condition is the end result of a defect in selectin-mediated leukocyte rolling that occurs prior to β2 integrin binding. There is a primary defect in fucose transporter such that oligosaccharide selectin ligands are missing in this syndromic condition.

Given that neutrophils are not able to reach infected tissues, LAD renders the individual susceptible to bacterial and fungal infections in a way that is similar to that of patients with SCN. LAD also causes impaired wound healing and delayed loss of the umbilical cord. A diagnosis can be suspected in cases of pus-free skin/tissue infections and massive hyperleukocytosis (>30,000/μL) in the blood (mostly granulocytes). Patients with LAD III also develop bleeding because the β2 integrin in platelets is not functional. Use of immunofluorescence and functional assays to detect β2 integrin can help form a diagnosis. Severe forms of LAD may require HSCT, although gene therapy is also now being considered. Inflammatory consequences of LAD can be mitigated by anti-interleukin (IL)-17A antibody therapy. Neutrophil-specific granule deficiency (a very rare condition caused by a mutation in the gene for transcription factor C/EBPα) results in a condition that is clinically similar to LAD. Infrequent additional leukocyte motility defects have also been reported.

■ CHRONIC GRANULOMATOUS DISEASES

Chronic granulomatous diseases (CGDs) are characterized by impaired phagocytic killing of microorganisms by neutrophils and macrophages **(Chap. 67)**. The incidence is ~1 per 200,000 live births. About 70% of cases are associated with X-linked recessive inheritance versus autosomal inheritance in the remaining 30%. CGD causes deep-tissue bacterial and fungal abscesses in macrophage-rich organs such as the lymph nodes, liver, and lungs. Recurrent skin infections (such as folliculitis) are common and can prompt an early diagnosis of CGD. The infectious agents are typically catalase-positive bacteria (such as *Staphylococcus aureus* and *Serratia marcescens*) but also include *Burkholderia cepacia*, pathogenic mycobacteria (in certain regions of the world), and fungi (mainly filamentous molds, such as *Aspergillus*).

CGD is caused by defective production of reactive oxygen species (ROS) in the phagolysosome membrane following phagocytosis of microorganisms. It results from the lack of a component of NADPH oxidase (gp91phox or p22phox) of the associated adapter/activating proteins (p47phox, p67phox, or p40phox) or a regulator of NADPH oxidase expression (CYBC1) that mediates the transport of electrons into the phagolysosome for creating ROS by interaction with O_2. Under normal circumstances, these ROS either directly kill engulfed microorganisms or enable the rise in pH needed to activate the phagosomal proteases that contribute to microbial killing. Diagnosis of CGD is based on assays of ROS production in neutrophils and monocytes (Table 362-2). As its name suggests, CGD is also a granulomatous disease. Macrophage-rich granulomas can often arise in the liver, spleen, and other organs. These are sterile granulomas that cause disease by obstruction (bladder, pylorus, etc.) or protracted inflammation (colitis, restrictive lung disease).

The management of infections in patients with CGD can be a complex process. The treatment of bacterial infections is generally based on combination therapy with antibiotics that are able to penetrate into cells. The treatment of fungal infections requires aggressive, long-term use of antifungals. Inflammatory/granulomatous lesions are usually steroid-sensitive, but often become glucocorticoid-dependent; liver abscesses are best managed by administering antibiotics together with glucocorticoids.

The treatment of CGD mostly relies on preventing infections. It has been unambiguously demonstrated that prophylactic usage of trimethoprim/sulfamethoxazole is both well tolerated and highly effective in reducing the risk of bacterial infection. Daily administration of azole derivatives (notably itraconazole) also reduces the frequency of fungal complications. It has long been suggested that interferon γ administration is helpful, although medical experts continue to disagree over this controversial issue. Patients may do reasonably well for some time with prophylaxis and careful management. However, patients are at high risk lifelong of severe and persistent fungal infections and/or chronic inflammatory complications, leading to consideration of performing HSCT. Due to an increase in reported successes, HSCT is now an established curative approach for CGD; however, the risk-versus-benefit ratio must be carefully assessed on a case-by-case basis. Gene therapy approaches are also being evaluated.

■ MENDELIAN SUSCEPTIBILITY TO MYCOBACTERIAL DISEASE

This group of diseases is characterized by a defect in the IL-12–interferon (IFN) γ axis (including IL-12p40, IL-12 receptor [R] β_1 and β_2, IFN-γ R_1 and R_2, TYK2, STAT1, IRF8, and ISG515 deficiencies), which ultimately leads to impaired IFN-γ-dependent macrophage activation. Both recessive and dominant inheritance modes have been observed. The hallmark of this PID is a specific and relatively narrow vulnerability to tuberculous and nontuberculous mycobacteria. The most severe phenotype (as observed in complete IFN-γ receptor deficiency) is characterized by disseminated infection that can be fatal even when aggressive and appropriate antimycobacterial therapy is applied. In addition to mycobacterial infections, MSMD patients (and particularly those with an IL-12/IL-12R deficiency) are prone to developing *Salmonella* infections. Although MSMDs are very rare, they should be considered in any patient with persistent mycobacterial infection. Treatment with IFN-γ may efficiently bypass an IL-12/IL-12R deficiency. HSCT is a therapeutic option for the most severe cases.

■ TOLL-LIKE RECEPTOR (TLR) PATHWAY DEFICIENCIES

In a certain group of patients with early-onset, invasive *Streptococcus pneumoniae* infections or (less frequently) *Staphylococcus aureus* or other pyogenic infections, conventional screening for PIDs does not identify the cause of the defect in host defense. It has been established that these patients carry recessive mutations in genes that encode essential adaptor molecules (IRAK4 and MYD88) involved in the signaling pathways of the majority of known TLRs **(Chap. 360)**. Remarkably, susceptibility to infection appears to decrease after the first few years of life—perhaps an indication that adaptive immunity (once triggered by an initial microbial challenge) is then able to prevent recurrent infections.

Certain TLRs (TLR-3, -7, -8, and -9) are involved in the recognition of RNA and DNA and usually become engaged during viral infections. Very specific susceptibility to herpes simplex encephalitis has been described in patients with a deficiency in Unc93b (a molecule associated with TLR-3, -7, -8, and -9 required for correct subcellular localization), TLR-3, or associated signaling molecules TRIF, TBK1, and TRAF3, resulting in defective type I IFN production. The fact that no other TLR deficiencies have been found—despite extensive screening of patients with unexplained, recurrent infections—strongly suggests that these receptors are functionally redundant. Hypomorphic mutations in NEMO/IKK-γ (a member of the NF-κB complex, which is activated downstream of TLR receptors) lead to a complex, variable immunodeficiency and a number of associated features. Susceptibility to both invasive, pyogenic infections and mycobacteria may be observed in this particular setting.

Rare cases of predisposition to severe viral infections (influenza, live measles vaccine) have been found in patients with genetic defects in IFN type I receptor and signaling pathways (STAT1, STAT2, KRF9, IRF7, MDA5). TLR7 deficiency has been shown to be a vulnerability factor of severe COVID-19. Autoantibodies to IFN-α or β, as observed in some PIDs (AIRE, NF-κB deficiencies), are also associated with severe COVID-19 disease.

COMPLEMENT DEFICIENCY

The complement system is composed of a complex cascade of plasma proteins (**Chap. 360**) that leads to the deposition of C3b fragments on the surface of particles and the formation of immune complexes that can culminate in the activation of a lytic complex at the bacterial surface. C3 cleavage can be mediated via three pathways: the classic, alternate, and lectin pathways. C3b coats particles as part of the opsonization process that facilitates phagocytosis following binding to cognate receptors. A deficiency in any component of the classic pathway (C1q, C1r, C1s, C4, and C2) can predispose an individual to bacterial infections that are tissue-invasive or that occur in the respiratory tract. Likewise, a C3 deficiency or a deficiency in factor I (a protein that regulates C3 consumption, thus leading to a C3 deficiency due to its absence) also results in the same type of vulnerability to infection. It has recently been reported that a very rare deficiency in ficolin-3 predisposes affected individuals to bacterial infections. Deficiencies in the alternative pathway (factors D and properdin) are associated with the occurrence of invasive *Neisseria* infections.

Lastly, deficiencies of any complement component involved in the lytic phase (C5, C6, C7, C8, and, to a lesser extent, C9) predispose affected individuals to systemic infection by *Neisseria*. This is explained by the critical role of complement in the lysis of the thick cell wall possessed by this class of bacteria.

Diagnosis of a complement deficiency relies primarily on testing the status of the classic and alternate pathway via functional assays, that is, the CH50 and AP50 tests, respectively. When either pathway is profoundly impaired, determination of the status of the relevant components in that pathway enables a precise diagnosis. Appropriate vaccinations and daily administration of oral penicillin are efficient means of preventing recurrent infections. It is noteworthy that several complement deficiencies (in the classic pathway and the lytic phase) may also predispose affected individuals to autoimmune diseases (notably systemic lupus erythematosus; **Chap. 368**).

PRIMARY IMMUNODEFICIENCIES OF THE ADAPTIVE IMMUNE SYSTEM

T LYMPHOCYTE DEFICIENCIES (TABLE 362-1, FIGS. 362-2 AND 362-3)

Given the central role of T lymphocytes in adaptive immune responses (**Chap. 360**), PIDs involving T cells generally have severe pathologic consequences; this explains the poor overall prognosis and the need for early diagnosis and the early intervention with appropriate therapy. Several differentiation pathways of T-cell effectors have been described,

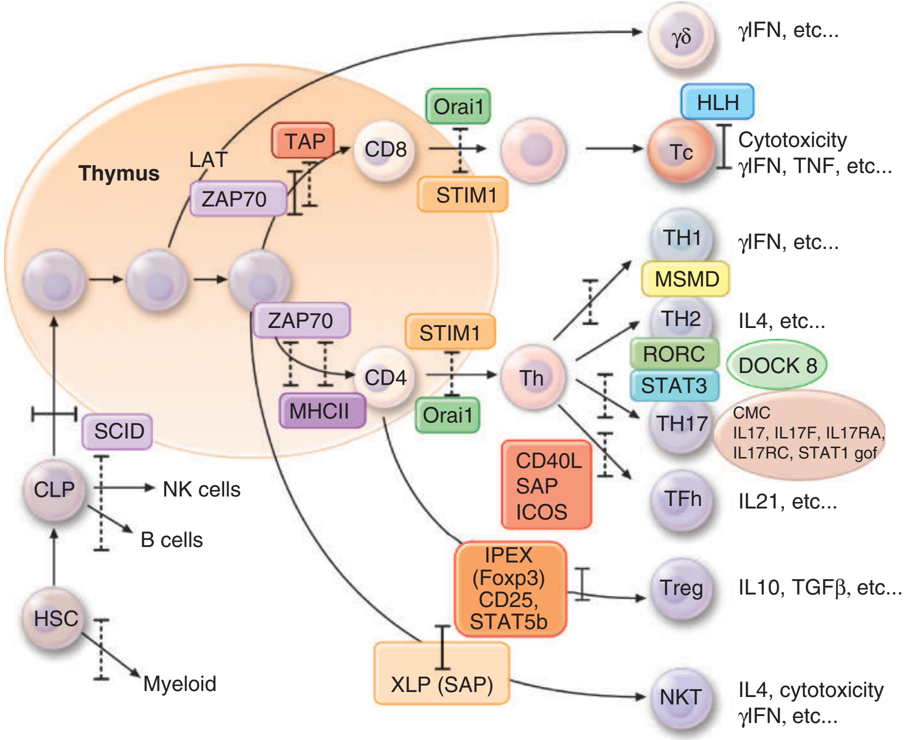

FIGURE 362-2 T-cell differentiation, effector pathways, and related primary immunodeficiencies (PIDs). Hematopoietic stem cells (HSCs) differentiate into common lymphoid progenitors (CLPs), which, in turn, give rise to the T-cell precursors that migrate to the thymus. The development of CD4+ and CD8+ T cells is shown. Known T-cell effector pathways are indicated, that is, γδ cells, cytotoxic T cells (Tc), T_H1, T_H2, T_H17, TFh (follicular helper) CD4 effector T cells, regulatory T cells (Treg), and natural killer T cells (NKTs); abbreviations for PIDs are contained in boxes. Vertical bars indicate a complete deficiency; broken bars a partial deficiency. DOCK8, autosomal recessive form of hyper-IgE syndrome; HLH, hematopoietic lymphohistiocytosis; IL17F, IL17RA, STAT1 (gof: gain of function), CMC (chronic mucocutaneous candidiasis), CD40L, ICOS, SAP deficiencies; IPEX, immunodysregulation polyendocrinopathy enteropathy X-linked syndrome; LAT, linker for activation of T cells; MHCII, major histocompatibility complex class II deficiency; MSMD, Mendelian susceptibility to mycobacterial disease; Orai1, STIM1 deficiencies; RORC, RAR related orphan receptor C; SCID, severe combined immunodeficiency; STAT3, autosomal dominant form of hyper-IgE syndrome; TAP, TAP1 and TAP2 deficiencies; XLP, X-linked proliferative syndromes; ZAP70, zeta-associated protein deficiency.

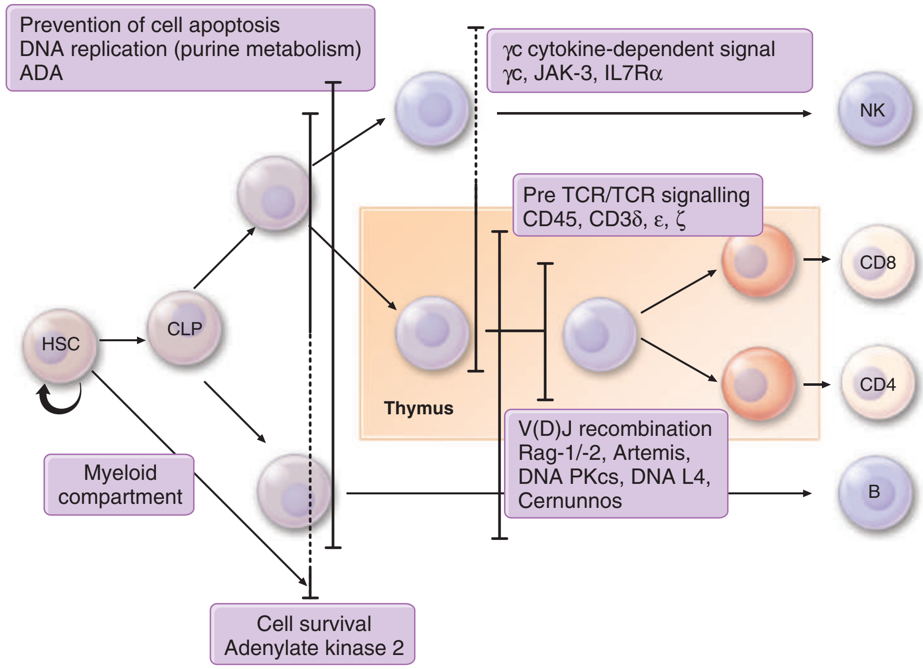

FIGURE 362-3 T-cell differentiation and severe combined immunodeficiencies (SCIDs). The vertical bars indicate the five mechanisms currently known to lead to SCID. The names of deficient proteins are indicated in the boxes adjacent to the vertical bars. A broken line means that deficiency is partial or involves only some of the indicated immunodeficiencies. ADA, adenosine deaminase deficiency; CLPs, common lymphoid progenitors; DNAL4, DNA ligase 4; HSCs, hematopoietic stem cells; NKs, natural killer cells; TCR, T-cell receptor.

one or all of which may be affected by a given PID (Fig. 362-2). Follicular helper CD4+ T cells in germinal centers are required for T-dependent antibody production, including the generation of Ig class-switched, high-affinity antibodies. CD4+ T_H1 cells provide cytokine-dependent (mostly IFN-γ-dependent) help to macrophages for intracellular killing of various microorganisms, including mycobacteria and *Salmonella*. CD4+ T_H2 cells produce IL-4, IL-5, and IL-13 and thus recruit and activate eosinophils and other cells required to fight helminth infections. CD4+ T_H17 cells produce IL-17 and IL-22 cytokines that recruit neutrophils to the skin and lungs to fight bacterial and fungal infections. Cytotoxic CD8+ T cells can kill infected cells, notably in the context of viral infections. In addition, certain T-cell deficiencies predispose affected individuals to *Pneumocystis jirovecii* lung infections early in life and to chronic gut/biliary duct/liver infections by *Cryptosporidium* and related genera later on in life. Lastly, naturally occurring or induced regulatory T cells are essential for controlling inflammation (notably reactivity to commensal bacteria in the gut) and autoimmunity. The role of other T-cell subsets with limited T-cell receptor (TCR) diversity (such as γδTCR T cells or natural killer T [NKT] cells) in PIDs is less well known; however, these subsets can be defective in certain PIDs, and this finding can sometimes contribute to the diagnosis (e.g., NKT-cell deficiency in X-linked proliferative syndrome [XLP]). T-cell deficiencies account for ~20% of all cases of PID.

Severe Combined Immunodeficiencies SCIDs constitute a group of rare PIDs characterized by a profound block in T-cell development and thus the complete absence of these cells. The developmental block is always the consequence of an intrinsic deficiency. The incidence of SCID is estimated to be 1 in 50,000 live births. Given the severity of the T-cell deficiency, clinical consequences occur early in life (usually within 3–6 months of birth). The most frequent clinical manifestations are recurrent oral candidiasis, failure to thrive, and protracted diarrhea and/or acute interstitial pneumonitis caused by *P. jirovecii* (although the latter can also be observed in the first year of life in children with B-cell deficiencies). Severe viral infections or invasive bacterial infections can also occur. Patients may also experience complications related to infections caused by live vaccines (notably Rotavirus causing protracted diarrhea or bacille Calmette-Guérin [BCG]) that may lead not only to local and regional infection but also to disseminated infection manifested by fever, splenomegaly, and skin and lytic bone lesions. A scaly skin eruption can be observed in a context of maternal T-cell engraftment (see below). A diagnosis of SCID can be suspected based on the patient's clinical history and, possibly, a family history of deaths in very young children (suggestive of either X-linked or recessive inheritance). Lymphocytopenia is strongly suggestive of SCID in >90% of cases (Table 362-2). The absence of a thymic shadow on a chest x-ray can also be suggestive of SCID. An accurate diagnosis relies on precise determination of the number of circulating T, B, and NK lymphocytes and their subsets. T-cell lymphopenia may be masked in some patients by the presence of maternal T cells (derived from maternal-fetal blood transfers) that cannot be eliminated. Although counts are usually low (<500/μL of blood), higher maternal T-cell counts may, under some circumstances, initially mask the presence of SCID. Thus, screening for maternal cells by using adequate genetic markers should be performed whenever necessary. Inheritance pattern analysis and lymphocyte phenotyping can discriminate between various forms of SCID and provide guidance in the choice of accurate molecular diagnostic tests (see below). To date, five distinct causative mechanisms for SCID (Fig. 362-3) have been identified. T-cell quantification of receptor excision circles (TREC) by using the Guthrie card is a reliable diagnostic test for newborn screening. It is now operational in the United States and several other countries worldwide. Its more widespread use will lead to the provision of therapy (see below) to uninfected patients resulting in a maximal chance of cure.

SEVERE COMBINED IMMUNODEFICIENCY CAUSED BY A CYTOKINE-SIGNALING DEFICIENCY The most frequent SCID phenotype (accounting for 30–40% of all cases) is the absence of both T and NK cells. This outcome results from a deficiency in either the common γ chain (γc) receptor that is shared by several cytokine receptors (the IL-2, -4, -7, -9, -15, and -21 receptors) or Jak-associated kinase (JAK) 3 that binds to the cytoplasmic portion of the γc chain receptor and induces signal transduction following cytokine binding. The former form of SCID (γc deficiency) has an X-linked inheritance mode, while the second is autosomal recessive. A lack of the IL-7Rα chain (which, together with γc, forms the IL-7 receptor) induces a selective T-cell deficiency.

PURINE METABOLISM DEFICIENCY Ten to 20% of SCID patients exhibit a deficiency in adenosine deaminase (ADA), an enzyme of purine metabolism that deaminates adenosine (ado) and deoxyadenosine (dAdo). An ADA deficiency results in the accumulation of ado and dAdo metabolites that induce premature cell death of lymphocyte progenitors. The condition results in the absence of B and NK lymphocytes as well as T cells. The clinical expression of complete ADA deficiency typically occurs very early in life. Since ADA is a ubiquitous enzyme, its deficiency can also cause bone dysplasia with abnormal costochondral junctions and metaphyses (found in 50% of cases) and neurologic defects. The very rare purine nucleoside phosphorylase (PNP) deficiency causes a profound although incomplete T-cell deficiency that is often associated with severe neurologic impairments.

DEFECTIVE REARRANGEMENTS OF T- AND B-CELL RECEPTORS A series of SCID conditions are characterized by a selective deficiency in T and B lymphocytes with autosomal recessive inheritance. These conditions account for 20–30% of SCID cases and result from mutations in genes encoding proteins that mediate the recombination of V(D)J gene elements in T- and B-cell antigen receptor genes (required for the generation of diversity in antigen recognition). The main deficiencies involve RAG1, RAG2, DNA-dependent protein kinase, and Artemis. A less severe (albeit variable) immunologic phenotype can result from other deficiencies in the same pathway, that is, DNA ligase 4 and Cernunnos deficiencies. Given that these latter factors are involved in DNA repair, these deficiencies may also cause developmental defects.

DEFECTIVE (PRE-)T-CELL RECEPTOR SIGNALING IN THE THYMUS A selective T-cell defect can be caused by a series of rare deficiencies in molecules involved in signaling via the pre-TCR or the TCR. These include deficiencies in CD3 subunits associated with the (pre-)TCR (i.e., CD3δ, ε, and ζ) and CD45.

RETICULAR DYSGENESIS Reticular dysgenesis is an extremely rare form of SCID that causes T and NK deficiencies with severe neutropenia and sensorineural deafness. It results from an adenylate kinase 2 deficiency. RAC-2 gain of function can cause the same immunologic phenotype.

Patients with SCID require appropriate care with aggressive anti-infective therapies, immunoglobulin replacement, and (when necessary) parenteral nutrition support. In most cases, curative treatment relies on HSCT. Today, HSCT provides a very high curative potential for SCID patients who are otherwise in reasonably good condition. Gene therapy has been found to be successful for cases of X-linked SCID (γc deficiency), SCID caused by an ADA deficiency, and SCID caused by Artemis deficiency. Lastly, a third option for the treatment of ADA deficiency consists of enzyme substitution with a pegylated enzyme.

Thymic Defects A profound T-cell defect can also result from faulty development of the thymus, as is most often observed in rare cases of DiGeorge's syndrome—a relatively common condition leading to a constellation of developmental defects. In ~1% of such cases, the thymus is completely absent, leading to virtually no mature T cells. However, expansion of oligoclonal T cells can occur and is associated with skin lesions. Diagnosis (using immunofluorescence in situ hybridization) is based on the identification of a hemizygous deletion in the long arm of chromosome 22. To recover the capability for T-cell differentiation, these cases require a thymic graft. CHARGE (*c*oloboma of the eye, *h*eart anomaly, choanal *a*tresia, *r*etardation, *g*enital, and *e*ar anomalies) syndrome (CHD7 deficiency) is a less frequent cause of impaired thymus development. Lastly, the very rare "nude" defect is characterized by the absence of both hair and the thymus.

Omenn Syndrome *Omenn syndrome* consists of a subset of T-cell deficiencies that present with a unique phenotype, including early-onset erythroderma, alopecia, hepatosplenomegaly, and failure to thrive. These patients usually display T-cell lymphocytosis, eosinophilia, and low B-cell counts. It has been found that the T cells of these patients exhibit a low TCR heterogeneity. This peculiar syndrome is the consequence of hypomorphic mutations in genes usually associated with SCID, that is, *RAG1*, *RAG2*, or (less frequently) *ARTEMIS* or *IL-7Rα*. The impaired homeostasis of differentiating T cells thus causes this immune system–associated disease. These patients are very fragile, requiring simultaneous anti-infective therapy, nutritional support, and immunosuppression. HSCT provides a curative approach.

Functional T-Cell Defects (Fig. 362-2) A subset of T-cell PIDs with autosomal inheritance is characterized by partially preserved T-cell differentiation but defective activation resulting in abnormal effector function. There are many causes of these defects, but all lead to susceptibility to viral and opportunistic infections, chronic diarrhea, and failure to thrive, with onset during childhood often associated with autoimmune manifestations. Careful phenotyping and in vitro functional assays are required to identify these diseases, the best characterized of which are the following.

ZETA-ASSOCIATED PROTEIN 70 (ZAP70) DEFICIENCY Zeta-associated protein 70 (ZAP70) is recruited to the TCR following antigen recognition. A ZAP70 deficiency leads typically to an almost complete absence of CD8+ T cells; CD4+ T cells are present but cannot be activated in vitro by TCR stimulation.

CALCIUM SIGNALING DEFECTS A small number of patients have been reported who exhibit a profound defect in in vitro T- and B-cell activation as a result of defective antigen receptor-mediated Ca^{2+} influx. This defect is caused by a mutation in the calcium channel gene (*ORAI1*) or its activator (*STIM-1*). It is noteworthy that these patients are also prone to autoimmune manifestations (blood cytopenias) and exhibit a nonprogressive muscle disease.

HUMAN LEUKOCYTE ANTIGEN (HLA) CLASS II DEFICIENCY Defective expression of HLA class II molecules is the hallmark of a group of four recessive genetic defects, all of which affect molecules (RFX5, RFXAP, RFXANK, and CIITA) involved in the transactivation of the genes coding for HLA class II. As a result, low but variable CD4+ T-cell counts are observed in addition to defective antigen-specific T- and B-cell responses. These patients are particularly susceptible to herpesvirus, adenovirus, and enterovirus infections and chronic gut/liver *Cryptosporidium* infections.

HLA CLASS I DEFICIENCY Defective expression of molecules involved in antigen presentation by HLA class I molecules (i.e., TAP-1, TAP-2, and tapasin) leads to reduced CD8+ T-cell counts, loss of HLA class I antigen expression, and a particular phenotype consisting of chronic obstructive pulmonary disease and severe vasculitis.

OTHER DEFECTS A variety of other T-cell PIDs have been described, some of which are associated with a precise molecular defect (e.g., IL-2-inducible T-cell kinase [ITK] deficiency; CD27, CD70, IL-21, and IL-21 receptor deficiencies; CARD11 deficiency; MALT1 deficiency; BCL10 deficiency; DOCK8 deficiency, DOCK2 deficiency; RORC deficiency; RLTPR deficiency, IKAROS deficiency). These conditions are also characterized by profound vulnerability to infections, such as severe Epstein-Barr virus (EBV)–induced B-cell proliferation and autoimmune disorders. Milder phenotypes are associated with CD8 and CD3γ deficiencies. Combined immunodeficiency associated with anhidrotic ectodermic dysplasia is the consequence of defects in the NF-κB signaling pathway (X-linked IKKγ deficiency and gain-of-function IKbα). Other deficiencies of the NF-κB complex component also lead to combined immunodeficiency.

HSCT is indicated for most of these diseases, although the prognosis is worse than in SCID because many patients are chronically infected at the time of diagnosis. Fairly aggressive immunosuppression and myeloablation may be necessary to achieve engraftment of allogeneic stem cells.

T-Cell Primary Immunodeficiencies with DNA Repair Defects This is a group of PIDs characterized by a combination of T- and B-cell defects of variable intensity, together with a number of nonimmunologic features resulting from DNA fragility. The autosomal recessive disorder *ataxia-telangiectasia* (AT) is the most frequently encountered condition in this group. It has an incidence of 1:40,000 live births and causes B-cell defects (low IgA, IgG2 deficiency, and low antibody production), which often require immunoglobulin replacement. AT is associated with a progressive T-cell immunodeficiency. As the name suggests, the hallmark features of AT are telangiectasia and cerebellar ataxia. The latter manifestations may not be detectable before the age of 3–4 years, so that AT should be considered in young children with IgA deficiency and recurrent and problematic infections. Diagnosis is based on a cytogenetic analysis showing excessive chromosomal rearrangements (mostly affecting chromosomes 7 and 14) in lymphocytes. AT is caused by a mutation in the gene encoding the ATM protein—a kinase that plays an important role in the detection and repair of DNA lesions (or cell death if the lesions are too numerous) by triggering several different pathways. Overall, AT is a progressive disease that carries a very high risk of lymphoma, leukemia, and (during adulthood) carcinomas. A variant of AT ("AT-like disease") is caused by mutation in the *MRE11* gene.

Nijmegen breakage syndrome (NBS) is a less common condition that also results from chromosome instability (with the same cytogenetic abnormalities as in AT). NBS is characterized by a severe T- and B-cell combined immune deficiency with autosomal recessive inheritance. Individuals with NBS exhibit microcephaly and a bird-like face but have neither ataxia nor telangiectasia. The risk of malignancies is very high. NBS results from a deficiency in nibrin (NBSI, a protein associated with MRE11 and Rad50 that is involved in checking DNA lesions) caused by hypomorphic mutations.

Severe forms of *dyskeratosis congenita* (also known as Hoyeraal-Hreidarsson syndrome) combine a progressive immunodeficiency that can also include an absence of B and NK lymphocytes, progressive bone marrow failure, microcephaly, in utero growth retardation, and gastrointestinal disease. The disease can be X-linked or, more rarely, autosomal recessive. It is caused by the mutation of genes encoding telomere maintenance proteins, including dyskerin (DKC1).

Finally, *immunodeficiency with centromeric and facial anomalies* (ICF) is a complex syndrome of autosomal recessive inheritance that variably combines a mild T-cell immune deficiency with a more severe B-cell immune deficiency, coarse face, digestive disease, and mild intellectual disability. A diagnostic feature is the detection by cytogenetic analysis of multiradial aspects in multiple chromosomes (most frequently 1, 9, and 16) corresponding to an abnormal DNA structure secondary to defective DNA methylation. It is the consequence of a deficiency in most cases in the DNA methyltransferase DNMT3B, ZBTB24, CDCA7, or HELLS.

T-Cell Primary Immunodeficiencies with Hyper-IgE Several T-cell PIDs are associated with elevated serum IgE levels (as in Omenn syndrome). A condition sometimes referred to as *autosomal recessive hyper-IgE syndrome* is notably characterized by recurrent bacterial infections in the skin and respiratory tract and severe skin and mucosal infections by pox viruses and human papillomaviruses, together with severe allergic manifestations. T and B lymphocyte counts are low. Mutations in the *DOCK8* gene have been found in many of these patients. This condition is an indication for HSCT.

A very rare, related condition with autosomal recessive inheritance that causes a similar susceptibility to infection with various microbes (see above), including mycobacteria, reportedly results from a deficiency in Tyk-2, a JAK family kinase involved in the signaling of many different cytokine receptors.

Autosomal Dominant Hyper-IgE Syndrome This unique condition, the *autosomal dominant hyper-IgE syndrome*, is usually diagnosed by the combination of recurrent skin and lung infections that can be complicated by pneumatoceles. Infections are caused by pyogenic bacteria and fungi. Several other manifestations characterize hyper-IgE syndrome, including facial dysmorphy, defective loss of primary teeth, hyperextensibility, scoliosis, and osteoporosis. Elevated serum IgE levels are typical of this syndrome. Defective T_H17 effector responses have been shown to account at least in part for the specific patterns of susceptibility to particular microbes. This condition is caused by a heterozygous (dominant) mutation in the gene encoding the transcription factor STAT3 that is required in a number of signaling pathways following binding of cytokine to cytokine receptors (such as that of IL-6 and the IL-6 receptor). It also results in partially defective antibody production because of defective IL-21 receptor signaling. Hence, immunoglobulin substitution can be considered as prophylaxis of bacterial infections.

Most recently, a recessive condition that mimics immunologic aspects of hyper-IgE syndrome has been ascribed to ZNF341 deficiency.

Cartilage Hair Hypoplasia The autosomal recessive *cartilage hair hypoplasia* (CHH) disease is characterized by short-limb dwarfism, metaphyseal dysostosis, and sparse hair, together with a combined T- and B-cell PID of extremely variable intensity (ranging from quasi-SCID to no clinically significant immune defects). The condition can predispose to erythroblastopenia, autoimmunity, and tumors. It is caused by mutations in the *RMRP* gene for a noncoding ribosome-associated RNA. Schimke immuno-osseous dysplasia is another autosomal recessive condition variably associating combined immunodeficiency, bone disease, and more importantly severe nephropathy. Other extremely rare immuno-osseous dysplasias with combined immunodeficiency have been reported, notably the MOPD1 deficiency.

CD40 Ligand and CD40 Deficiencies *Hyper-IgM syndrome* (HIGM) is a well-known PID that was previously classified as a B-cell immune deficiency (see **Fig. 362-4** and below). It results from defective immunoglobulin class switch recombination (CSR) in germinal centers and leads to profound deficiency in production of IgG, IgA, and IgE (although IgM production is maintained). Approximately half of HIGM sufferers are also prone to opportunistic infections, for example, interstitial pneumonitis caused by *P. jirovecii* (in young children), protracted diarrhea and cholangitis caused by *Cryptosporidium*, and infection of the brain with *Toxoplasma gondii*.

In the majority of cases, this condition has an X-linked inheritance and is caused by a deficiency in CD40 ligand (L). CD40L induces signaling events in B cells that are necessary for both CSR and adequate activation of other CD40-expressing cells that are involved in innate immune responses against the above-mentioned microorganisms. More rarely, the condition is caused by a deficiency in CD40 itself. The poorer prognosis of CD40L and CD40 deficiencies (relative to most other HIGM conditions) implies that (1) thorough investigations have to be performed in all cases of HIGM and (2) potentially curative HSCT should be discussed on a case-by-case basis for this group of patients.

Wiskott-Aldrich Syndrome WAS is a complex, recessive, X-linked disease with an incidence of ~1 in 200,000 live births. It is caused by mutations in the *WASP* gene that affect not only T lymphocytes but also the other lymphocyte subsets, dendritic cells, and platelets. WAS is typically characterized by the following clinical manifestations: recurrent bacterial infections, eczema, and bleeding caused by thrombocytopenia. However, these manifestations are highly variable—mostly as a consequence of the many different *WASP* mutations that have been observed. Null mutations predispose affected individuals to invasive and bronchopulmonary infections, viral infections, severe eczema, and autoimmune manifestations. The latter include autoantibody-mediated blood cytopenia, glomerulonephritis, skin and visceral vasculitis (including brain vasculitis), erythema nodosum, and arthritis. Another possible consequence of WAS is lymphoma, which may be virally induced (e.g., by EBV or Kaposi's sarcoma–associated herpesvirus). Thrombocytopenia can be severe and compounded by the peripheral destruction of platelets associated with autoimmune disorders. Hypomorphic mutations usually lead to milder outcomes that are generally limited to thrombocytopenia. It is noteworthy that even patients with "isolated" X-linked thrombocytopenia can develop severe autoimmune disease and/or lymphoma later in life. The immunologic workup is not very informative; there can be a relative CD8+ T-cell deficiency, frequently accompanied by low serum IgM levels and decreased antigen-specific antibody responses. A typical feature is reduced-sized platelets on a blood smear. Diagnosis is based on intracellular immunofluorescence analysis of WAS protein (WASp) expression in blood cells. WASp regulates the actin cytoskeleton and thus plays an important role in many lymphocyte functions, including cell adhesion and migration and the formation of synapses between antigen-presenting and target cells. Predisposition to autoimmune disorders is in part related to defective regulatory T cells. The treatment

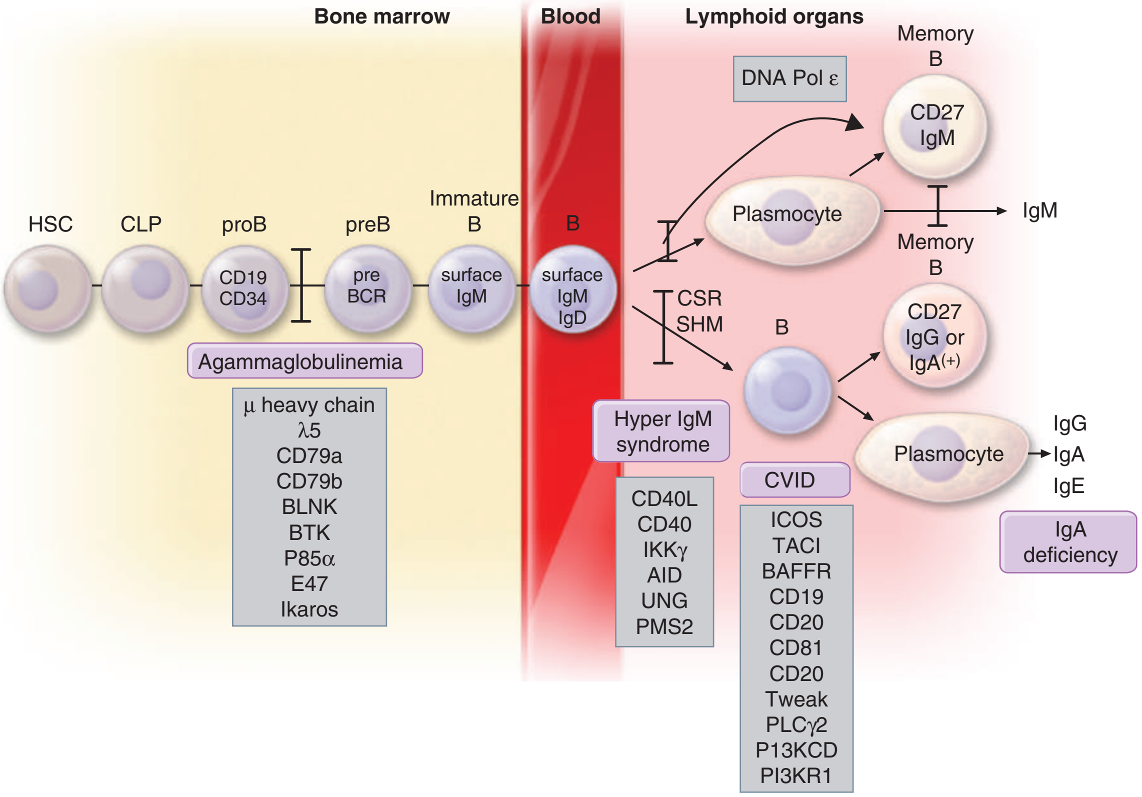

FIGURE 362-4 B-cell differentiation and related primary immunodeficiencies (PIDs). Hematopoietic stem cells (HSCs) differentiate into common lymphoid progenitors (CLPs), which give rise to pre-B cells. The B-cell differentiation pathway goes through the pre–B-cell stage (expression of the μ heavy chain and surrogate light chain), the immature B-cell stage (expression of surface IgM), and the mature B-cell stage (expression of surface IgM and IgD). The main phenotypic characteristics of these cells are indicated. In lymphoid organs, B cells can differentiate into plasma cells and produce IgM or undergo (in germinal centers) Ig class switch recombination (CSR) and somatic mutation of the variable region of V genes (SHM) that enable selection of high-affinity antibodies. These B cells produce antibodies of various isotypes and generate memory B cells. PIDs are indicated in the purple boxes. CVID, common variable immunodeficiency.

of WAS should match the severity of disease expression. Prophylactic antibiotics, immunoglobulin G (IgG) supplementation, and careful topical treatment of eczema are indicated. Although splenectomy improves platelet count in a majority of cases, this intervention is associated with a significant risk of infection and is thus contraindicated (both before and after HSCT). Allogeneic HSCT is curative, with excellent results overall. Gene therapy trials have been performed with success. A similar condition has been reported in a few cases with a deficiency in the Wiskott-Aldrich interacting protein (WIP) with autosomal recessive inheritance.

A few other complex PIDs are worth mentioning. *Sp110 deficiency* causes a T-cell PID with liver venoocclusive disease and hypogammaglobulinemia. *Chronic mucocutaneous candidiasis* (CMC) is a heterogeneous disease, considering the different inheritance patterns that have been observed. In some cases, chronic candidiasis is associated with late-onset bronchopulmonary infections, bronchiectasis, and brain aneurysms. Moderate forms of CMC are related to autoimmunity and AIRE deficiency (see below). In this setting, predisposition to *Candida* infection is associated with the detection of autoantibodies to T_H17 cytokines. Recently, deficiencies in IL-17A, IL-17F, and IL-17 receptor A and C and in the associated protein Act1, and above all, gain-of-function mutations in *STAT1* have been found to be associated with CMC. In all cases, CMC is related to defective T_H17 function. Innate immunodeficiency in CARD9 also predisposes to chronic invasive fungal infection.

■ B LYMPHOCYTE DEFICIENCIES (TABLE 362-1, FIG. 362-4)

Deficiencies that predominantly affect B lymphocytes are the most frequent PIDs and account for 60–70% of all cases. B lymphocytes make antibodies. Pentameric IgMs are found in the vascular compartment and are also secreted at mucosal surfaces. IgG antibodies diffuse freely into extravascular spaces, whereas IgA antibodies are produced and secreted predominantly from mucosa-associated lymphoid tissues. Although Ig isotypes have distinct effector functions, including Fc receptor–mediated and (indirectly) C_3 receptor–dependent phagocytosis of microorganisms, they share the ability to recognize and neutralize a given pathogen. Defective antibody production therefore allows the establishment of invasive, pyogenic bacterial infections as well as recurrent sinus and pulmonary infections (mostly caused by *S. pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, and, less frequently, gram-negative bacteria). If left untreated, recurrent bronchial infections lead to bronchiectasis and, ultimately, cor pulmonale and death. Parasitic infections such as caused by *Giardia lamblia* and bacterial infections caused by *Helicobacter* and *Campylobacter* of the gut are also observed. A complete lack of antibody production (namely agammaglobulinemia) can also predispose affected individuals to severe, chronic, disseminated enteroviral or astrovirus infections causing meningoencephalitis, hepatitis, and a dermatomyositis-like disease. Other enteric viruses can also cause persistent hepatitis.

Even with the most profound of B-cell deficiencies, infections rarely occur before the age of 6 months; this is because of transient protection provided by the transplacental passage of immunoglobulins during the last trimester of pregnancy. Conversely, a genetically nonimmunodeficient child born to a mother with hypogammaglobulinemia is, in the absence of maternal Ig substitution, usually prone to severe bacterial infections in utero and for several months after birth.

Diagnosis of B-cell PIDs relies on the determination of serum Ig levels (Table 362-2). Determination of antibody production following immunization with tetanus toxoid vaccine or nonconjugated pneumococcal polysaccharide antigens can also help diagnose more subtle deficiencies. Another useful test is B-cell phenotype determination in switched μ−δ− CD27+ and nonswitched memory B cells (μ+δ+ CD27+). In agammaglobulinemic patients, examination of bone marrow B-cell precursors (Fig. 362-4) can help obtain a precise diagnosis and guide the choice of genetic tests.

Agammaglobulinemia Agammaglobulinemia is characterized by a profound defect in B-cell development (<1% of the normal B-cell blood count). In most patients, very low residual Ig isotypes can be detected in the serum. In 85% of cases, agammaglobulinemia is caused by a mutation in the *BTK* gene that is located on the X chromosome. The *BTK* gene product is a kinase that participates in (pre) B-cell receptor signaling. When the kinase is defective, there is a block (albeit a leaky one) at the pre-B to B-cell stage (Fig. 362-4). Detection of *BTK* by intracellular immunofluorescence of monocytes, and lack thereof in patients with X-linked agammaglobulinemia (XLA), is a useful diagnostic test. Not all of the mutations in *BTK* result in agammaglobulinemia, since some patients have a milder form of hypogammaglobulinemia and low but detectable B-cell counts. These cases should not be confused with common variable immunodeficiency (CVID; see below). About 10% of agammaglobulinemia cases are caused by alterations in genes encoding elements of the pre-B-cell receptor, i.e., the μ heavy chain, the λ5 surrogate light chain, Igα or Igβ, the scaffold protein BLNK, the p85 α subunit of phosphatidylinositol 3 phosphate kinase (P13K), the E47, and the Ikaros transcription factors. In 5% of cases, the defect is unknown. It is noteworthy that agammaglobulinemia can be observed in patients with ICF syndrome, despite the presence of normal peripheral B-cell counts. Lastly, agammaglobulinemia can be a manifestation of a myelodysplastic syndrome (associated or not with neutropenia). Treatment of agammaglobulinemic patients is based on immunoglobulin replacement (see below). HSCT is considered in special cases. Profound hypogammaglobulinemia is also observed in adults, in association with thymoma.

Hyper-IgM (HIGM) Syndromes *HIGM* is a rare B-cell PID characterized by defective Ig CSR. It results in very low serum levels of IgG and IgA and elevated or normal serum IgM levels. The clinical severity is similar to that seen in agammaglobulinemia, although chronic lung disease and sinusitis are less frequent and enteroviral infections are uncommon. As discussed above, a diagnosis of HIGM involves screening for an X-linked CD40L deficiency and an autosomal recessive CD40 deficiency, which affect both B and T cells. In 50% of cases affecting only B cells, these isolated HIGM syndromes result from mutations in the gene encoding activation-induced deaminase, the protein that induces CSR in B-cell germinal centers. These patients usually have enlarged lymphoid organs. In the other 50% of cases, the etiology is unknown (except for rare UNG and PMS2 deficiencies). Furthermore, IgM-mediated autoimmunity and lymphomas can occur in HIGM syndrome. It is noteworthy that HIGM can result from fetal rubella syndrome or can be a predominant immunologic feature of other PIDs, such as the immunodeficiency associated with ectodermic anhydrotic hypoplasia X-linked NEMO deficiency and the combined T- and B-cell PIDs caused by DNA repair defects such as AT and Cernunnos deficiency.

Common Variable Immunodeficiency CVID is an ill-defined condition characterized by low serum levels of one or more Ig isotypes. Its prevalence is estimated to be 1 in 20,000. The condition is recognized predominantly in adults, although clinical manifestations can occur earlier in life. Hypogammaglobulinemia is associated with at least partially defective antibody production in response to vaccine antigens. B lymphocyte counts are often normal but can be low. Besides infections, CVID patients may develop lymphoproliferation (splenomegaly), granulomatous lesions, colitis, antibody-mediated autoimmune disease, and lymphomas that define disease prognosis. A family history is found in 10% of cases. A clear-cut dominant inheritance pattern is found in some families, whereas recessive inheritance is observed more rarely. In most cases, no molecular cause can be identified. A small number of patients in Germany were found to carry mutations in the *ICOS* gene encoding a T-cell membrane protein that contributes to B-cell activation and survival. In 10% of patients with CVID, monoallelic or biallelic mutations of the gene encoding TACI (a member of the tumor necrosis factor [TNF] receptor family that is expressed on B cells) have been found. In fact, heterozygous TACI mutations correspond to a genetic susceptibility factor, since similar heterozygous mutations are found in 1% of controls. NFkB1 transcription factor mutations have been found in a small fraction of patients with CVID. The B-cell activating factor (BAFF) receptor was found to be defective in a kindred with CVID, although not all individuals carrying the mutation have CVID. A group of patients with hypogammaglobulinemia and lymphoproliferation was shown to exhibit dominant gain-of-function mutations in the *PIK3CD* gene encoding the p110δ form of P13 kinase or in the *PI3KR1* gene encoding the regulatory p85α subunit of PI3 kinase. Rare cases of hypogammaglobulinemia were found to be associated with CD19, CD20, CD21, and CD81 deficiencies. These patients have B cells that can be identified by typing for other B-cell markers.

A diagnosis of CVID should be made after excluding the presence of hypomorphic mutations associated with agammaglobulinemia or more subtle T-cell defects; this is particularly the case in children. It is possible that many cases of CVID result from a constellation of factors, rather than a single genetic defect. Hypogammaglobulinemia can be associated with neutropenia and lymphopenia in the WHIM syndrome caused by a dominant gain-of-function mutation of *CXCR4*, resulting in cell retention in the bone marrow.

Selective Ig Isotype Deficiencies *IgA deficiency* and CVID represent polar ends of a clinical spectrum due to the same underlying gene defect(s) in a large subset of these patients. IgA deficiency is the most common PID; it can be found in 1 in every 600 individuals. It is asymptomatic in most cases; however, individuals may present with increased numbers of acute and chronic respiratory infections that may lead to bronchiectasis. In addition, over their lifetime, these patients experience an increased susceptibility to drug allergies, atopic disorders, and autoimmune diseases. Symptomatic IgA deficiency is probably related to CVID, since it can be found in relatives of patients with CVID. Furthermore, IgA deficiency may progress to CVID. It is thus important to assess serum Ig levels in IgA-deficient patients (especially when infections occur frequently) in order to detect changes that should prompt the initiation of immunoglobulin replacement. Selective IgG2 (+G4) deficiency (which in some cases may be associated with IgA deficiency) can also result in recurrent sinopulmonary infections and should thus be specifically sought in this clinical setting. These conditions are ill-defined and often transient during childhood. A pathophysiologic explanation has not been found.

Selective Antibody Deficiency to Polysaccharide Antigens Some patients with normal serum Ig levels are prone to *S. pneumoniae* and *H. influenzae* infections of the respiratory tract. Defective production of antibodies against polysaccharide antigens (such as those in the *S. pneumoniae* cell wall) can be observed and is probably causative. This condition may correspond to a defect in marginal zone B cells, a B-cell subpopulation involved in T-independent antibody responses.

Immunoglobulin Replacement IgG antibodies have a half-life of 21–28 days. Thus, injection of plasma-derived polyclonal IgG containing a myriad of high-affinity antibodies can provide protection against disease-causing microorganisms in patients with defective IgG antibody production. This form of therapy should not be based on laboratory data alone (i.e., IgG and/or antibody deficiency) but should be guided by the occurrence or not of infections; otherwise, patients might be subjected to unjustified IgG infusions. Immunoglobulin replacement can be performed by IV or subcutaneous routes. In the former case, injections have to be repeated every 3–4 weeks, with a residual target level above 800 mg/mL in patients who had very low IgG levels prior to therapy. Subcutaneous injections are typically performed once a week, although the frequency can be adjusted on a case-by-case basis. New preparations containing hyaluronidase allow reduced frequency of injection by increasing volume injected. A trough level above 800 mg/mL is desirable. Whatever the mode of administration, the main goal is to reduce the frequency of the respiratory tract infections and prevent chronic lung and sinus disease. The two routes appear to be equally safe and efficacious, and so the choice should be left to the preference of the patient.

In patients with chronic lung disease, chest physical therapy with good pulmonary toilet and the use of antibiotics, notably azithromycin, are also needed. Immunoglobulin replacement is well tolerated by most patients, although the selection of the best-tolerated Ig preparation may be necessary in certain cases. Since IgG preparations contain a small proportion of IgAs, caution should be taken in patients with residual antibody production capacity and a complete IgA deficiency, as these subjects may develop anti-IgA antibodies that can trigger anaphylactic shock. These patients should be treated with IgA-free IgG preparations. Immunoglobulin replacement is a lifelong therapy; its rationale and procedures have to be fully understood and mastered by the patient and their family in order to guarantee the strict observance required for efficacy.

PRIMARY IMMUNODEFICIENCIES AFFECTING REGULATORY PATHWAYS (TABLE 362-1)

An increasing number of PIDs have been found to cause homeostatic dysregulation of the immune system, either alone or in association with increased vulnerability to infections. Defects of this type affecting the innate immune system and autoinflammatory syndromes will not be covered in this chapter. However, three specific entities (hemophagocytic lymphohistiocytosis [HLH], lymphoproliferation, and autoimmunity) will be described below.

HEMOPHAGOCYTIC LYMPHOHISTIOCYTOSIS

HLH is characterized by an unremitting activation of CD8+ T lymphocytes and macrophages that leads to organ damage (notably in the liver, bone marrow, and central nervous system). This syndrome results from a broad set of inherited diseases, most of which impair T and NK lymphocyte cytotoxicity. The manifestations of HLH are often induced by a viral infection. EBV is the most frequent trigger. In severe forms of HLH, disease onset may start during the first year of life or even (in rare cases) at birth.

Diagnosis relies on the identification of the characteristic symptoms of HLH (fever, hepatosplenomegaly, edema, neurologic diseases, blood cytopenia, increased liver enzymes, hypofibrinogenemia, high triglyceride [hyperferritinemia] levels, elevated markers of T-cell activation, and hemophagocytic features in the bone marrow or cerebrospinal fluid). Functional assays of postactivation cytotoxic granule exocytosis (CD107 fluorescence at the cell membrane) can suggest genetically determined HLH. The conditions can be classified into three subsets:

1. Familial HLH with autosomal recessive inheritance, including perforin deficiency (30% of cases) that can be recognized by assessing intracellular perforin expression; Munc13-4 deficiency (30% of cases); syntaxin 11 deficiency (10% of cases); Munc18-2 deficiency (20% of cases); rare deficiencies in FAAP24 and RHOG; and a few residual cases that lack a known molecular defect.
2. HLH with partial albinism. Three conditions combine HLH and abnormal pigmentation, where hair examination can help in the diagnosis: Chédiak-Higashi syndrome, Griscelli syndrome, and Hermansky-Pudlak syndrome type II. Chédiak-Higashi syndrome is also characterized by the presence of giant lysosomes within leukocytes **(Chap. 67)**, in addition to a primary neurologic disorder with slow progression of symptoms over time.
3. XLP is characterized in most patients by the induction of HLH following EBV infection, while other patients develop progressive hypogammaglobulinemia similar to what is observed in CVID and/or certain lymphomas. XLP is caused by a mutation in the *SH2DIA* gene that encodes the adaptor protein SAP (associated with a SLAM family receptor). Several immunologic abnormalities have been described, including low 2B4-mediated NK cell cytotoxicity, impaired differentiation of NKT cells, defective antigen-induced T-cell death, and defective T-cell helper activity for B cells. A related disorder (XLP2) has recently been described. It is also X-linked and induces HLH (frequently after EBV infection), although the clinical manifestation may be less pronounced. The condition is associated with a deficiency of the antiapoptotic molecule XIAP. The pathophysiology of XLP2 remains unclear; however, it may be related to control of inflammation in macrophages as there is a functional link between XIAP and NLRC4, an inflammasome component, in which gain of function can also induce HLH. XLP2 is also frequently associated with colitis.

HLH is a life-threatening complication. The treatment of this condition requires aggressive immunosuppression with either the cytotoxic agent etoposide or anti–T-cell antibodies; specific therapy targeting the IFN-γ pathway (JAK1), which is critical in causing HLH, is an additional option to consider. Once remission has been achieved, HSCT should be performed, since it provides the only curative form of therapy. Of note, acquired forms of HLH are more commonly observed in adults as a complication of infection, malignancies, or autoimmune diseases or sometimes on its own.

AUTOIMMUNE LYMPHOPROLIFERATIVE SYNDROME

Autoimmune lymphoproliferative syndrome (ALPS) is characterized by nonmalignant T and B lymphoproliferation causing splenomegaly and enlarged lymph nodes; 70% of patients also display autoimmune manifestations such as autoimmune cytopenias, Guillain-Barré syndrome, uveitis, and hepatitis **(Chaps. 70 and 360)**. A hallmark of ALPS is the presence of CD4–CD8– TCRαβ+ T cells (2–50%) in the blood of affected individuals. Hypergammaglobulinemia involving IgG and IgA is also frequently observed. The syndrome is caused by a defect in Fas-mediated apoptosis of lymphocytes, which can thus accumulate and mediate autoimmunity. Furthermore, ALPS can lead to malignancies.

Most patients carry a heterozygous mutation in the gene encoding Fas that is characterized by dominant inheritance and variable penetrance, depending on the nature of the mutation. A rare and severe form of the disease with early onset can be observed in patients carrying a biallelic mutation of Fas, which profoundly impairs the protein's expression and/or function. Fas-ligand, caspase 10, caspase 8, and somatic neuroblastoma RAS viral oncogene homologue (NRAS) and KRAS mutations have also been reported in a few cases of ALPS. Many cases of ALPS have not been precisely delineated at the molecular level. A B cell–predominant ALPS has recently been found associated with a protein kinase Cδ gene mutation. Treatment of ALPS is essentially based on the use of proapoptotic drugs, which need to be carefully administered in order to avoid toxicity.

COLITIS, AUTOIMMUNITY, AND PRIMARY IMMUNODEFICIENCIES

Several PIDs (most of which are T cell–related) can cause severe gut inflammation. The prototypic example is *immunodysregulation polyendocrinopathy enteropathy X-linked syndrome* (IPEX), characterized by a widespread inflammatory enteropathy, food intolerance, skin rashes, autoimmune cytopenias, and diabetes. The syndrome is caused by loss-of-function mutations in the gene encoding the transcription factor FOXP3, which is required for the acquisition of effector function by regulatory T cells. In most cases of IPEX, CD4+CD25+ regulatory T cells are absent from the blood. This condition has a poor prognosis and requires aggressive immunosuppression. The only possible curative approach is allogeneic HSCT. IPEX-like syndromes that lack a *FOXP3* mutation have also been described. In some cases, CD25 (IL-2 receptor α subunit) and CD122 (IL-2 receptor β subunit) deficiencies have been found. Defective IL-2 receptor expression also impairs regulatory cell expansion/function. This functional T-cell deficiency means that IL-2 receptor–deficient patients are also at increased risk of opportunistic infections. It is noteworthy that abnormalities in regulatory T cells have been described in other PID settings, such as in Omenn syndrome, STAT5b deficiency, STIM1 (Ca flux) deficiency, and WAS; these abnormalities may account (at least in part) for the occurrence of inflammation and autoimmunity. The autoimmune features observed in a small fraction of patients with DiGeorge's syndrome may have the same cause. Severe, early-onset inflammatory gut disease has been described in patients with a deficiency in the IL-10 receptor or IL-10.

Dominant mutations in genes encoding the regulatory molecule CTLA-4, recessive mutations in the gene encoding LRBA (a molecule involved in recycling of CTLA-4), as well as dominant gain-of-function mutation of STAT3 cause a multifaceted lymphoproliferative and autoimmune syndrome, frequently involving inflammatory bowel disease that can be associated with hypogammaglobulinemia. Molecular diagnosis is required before adapted targeted therapies are undertaken. Gain-of-function mutation of *IK2F1* (Ikaros) or BACH2 deficiency can also cause colitis with multiple autoimmune features.

A distinct autoimmune entity is observed in *autoimmune polyendocrinopathy candidiasis ectodermal dysplasia* (APECED) syndrome, which is characterized by autosomal recessive inheritance. It consists of multiple autoimmune manifestations that can affect solid organs in general and endocrine glands in particular. Mild, chronic *Candida* infection is often associated with this syndrome. The condition is due to mutations in the autoimmune regulator (*AIRE*) gene and results in impaired thymic expression of self-antigens by medullary epithelial cells and impaired negative selection of self-reactive T cells that leads to autoimmune manifestations.

A combination of hypogammaglobulinemia, autoantibody production, cold-induced urticaria or skin granulomas, or autoinflammation has been reported and has been termed *PLCγ2-associated antibody deficiency and immune dysregulation* (PLAID or APLAID).

CONCLUSION

The variety and complexity of the clinical manifestations of the many different PIDs strongly indicate that it is important to raise awareness of these diseases. Indeed, early diagnosis is essential for establishing an appropriate therapeutic regimen. Hence, patients with suspected PIDs must always be referred to experienced clinical centers that are able to perform appropriate molecular and genetic tests. A precise molecular diagnosis is not only necessary for initiating the most suitable treatment but also important for genetic counseling and prenatal diagnosis.

One pitfall that may hamper diagnosis is the high variability that is associated with many PIDs. Variable disease expression can result from the differing consequences of various mutations associated with a given condition, as exemplified by WAS and, to a lesser extent, XLA or CTLA4 deficiency. There can also be effects of modifier genes (as also suspected in XLA) and environmental factors such as EBV infection that can be the main trigger of disease in XLP conditions. Furthermore, it has recently been established that somatic mutations in an affected gene can attenuate the phenotype of a number of T-cell PIDs. This has been described for ADA deficiency, X-linked SCID, RAG deficiencies, NF-κB essential modulator (NEMO) deficiency, and, most frequently, WAS. In contrast, somatic mutations can create disease states analogous to PID, as reported for ALPS. The VEXAS syndrome characterized by late-onset inflammation, severe and treatment-resistant vasculitis, and cytopenias is caused by somatic mutation in the X-linked gene *UBA-1*. Lastly, cytokine-neutralizing autoantibodies can mimic a PID, as shown for IFN-γ, IL-6, IL-7, and type I interferon.

Many aspects of the pathophysiology of PIDs are still unknown, and the disease-causing gene mutations have not been identified in all cases (as illustrated by CVID and IgA deficiency). However, our medical understanding of PIDs has now reached the stage where scientifically based approaches to the diagnosis and treatment of these diseases can be implemented. A genetic diagnosis has become a milestone step in the care of PID patients.

FURTHER READING

Abolhassani H et al: Current genetic landscape in common variable immune deficiency. Blood 135:656, 2020.

Casanova JL et al: Guidelines for genetic studies in single patients: Lessons from primary immunodeficiencies. J Exp Med 211:2137, 2014.

Fischer A: Gene therapy for inborn errors of immunity: Past, present and future. Nat Rev Immunol 23:397, 2023.

Holland SM: Chronic granulomatous disease. Hematol Oncol Clin North Am 27:89, viii, 2013.

Kwan A et al: Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. JAMA 312:729, 2014.

Notarangelo LD: Functional T cell immunodeficiencies (with T cells present). Annu Rev Immunol 31:195, 2013.

Ochs HD et al (eds): *Primary Immunodeficiencies: A Molecular and Genetic Approach*. New York, Oxford University Press, 2013.

Picard C, Fischer A: Contribution of high-throughput DNA sequencing to the study of primary immunodeficiencies. Eur J Immunol 44:2854, 2014.

Tangye SG et al: Human inborn errors of immunity: 2022 update on the classification from the International Union of Immunological Societies Expert Committee. J Clin Immunol 42:1473, 2022.

Section 2 Disorders of Immune-Mediated Injury

363

Urticaria, Angioedema, and Allergic Rhinitis

Dinah Foer, Joshua A. Boyce

INTRODUCTION

The term *atopy* implies a tendency to manifest asthma, rhinitis, urticaria, food allergy, and atopic dermatitis alone or in combination, in association with the presence of allergen-specific IgE. However, individuals without an atopic background may also develop hypersensitivity reactions, particularly urticaria and anaphylaxis, associated with the presence of allergen-specific IgE. Since mast cells are key effector cells in allergic rhinitis and asthma, and the dominant effector in urticaria, anaphylaxis, and systemic mastocytosis, mast cell development, activation, product profile, and target tissues will be considered in the introduction to these clinical disorders. Dysregulation of mast cell development seen in mastocytosis will be covered in a separate chapter.

The binding of IgE to human mast cells and basophils, a process termed *sensitization*, prepares these cells for subsequent antigen-specific activation. The high-affinity Fc receptor for IgE, designated FcεRI, is composed of one α, one β, and two disulfide-linked γ chains, which together cross the plasma membrane seven times. The α chain is responsible for IgE binding, and the β and γ chains provide for signal transduction that follows the aggregation of the sensitized tetrameric receptors by polymeric antigen. The binding of IgE stabilizes the α chain at the plasma membrane, thus increasing the density of FcεRI receptors at the cell surface while sensitizing the cell for effector responses. This accounts for the correlation between serum IgE levels and the numbers of FcεRI receptors detected on circulating basophils. Signal transduction is initiated through the action of a Src family–related tyrosine kinase termed Lyn that is constitutively associated with the β chain. Lyn transphosphorylates the canonical immunoreceptor tyrosine-based activation motifs (ITAMs) of the β and γ chains of the receptor, resulting in recruitment of more active Lyn to the β chain and of Syk tyrosine kinase. The phosphorylated tyrosines in the ITAMs function as binding sites for the tandem *src* homology two (SH2) domains within Syk. Syk activates not only phospholipase Cγ, which associates with the linker of activated T cells at the plasma membrane, but also phosphatidylinositol 3-kinase to provide phosphatidylinositol-3,4,5-trisphosphate, which allows membrane targeting of the Tec family kinase Btk and its activation by Lyn. In addition, the Src family tyrosine kinase Fyn becomes activated after aggregation of IgE receptors and phosphorylates the adapter protein Gab2 that enhances activation of phosphatidylinositol 3-kinase. Indeed, this additional input is essential for mast cell activation, but it can be partially inhibited by

Lyn, indicating that the extent of mast cell activation is in part regulated by the interplay between these Src family kinases. Activated phospholipase Cγ cleaves phospholipid membrane substrates to provide inositol-1,4,5-trisphosphate (IP_3) and 1,2-diacylglycerols (1,2-DAGs) to mobilize intracellular calcium and activate protein kinase C, respectively. The subsequent opening of calcium-regulated activated channels provides the sustained elevations of intracellular calcium required to recruit the mitogen-activated protein kinases ERK, JNK, and p38 (serine/threonine kinases), which phosphorylate substrates that permit arachidonic acid release and to mediate nuclear translocation of transcription factors for various cytokines. The calcium ion–dependent activation of phospholipases cleaves membrane phospholipids to generate lysophospholipids, which, like 1,2-DAG, may facilitate the fusion of the secretory granule perigranular membrane with the cell membrane, a step that releases the membrane-free granules containing the preformed mast cell mediators.

The secretory granule of the human mast cell has a crystalline structure. IgE-dependent cell activation results in solubilization and swelling of the granule contents within the first minute of receptor perturbation; this reaction is followed by the ordering of intermediate filaments about the swollen granule, movement of the granule toward the cell surface, and fusion of the perigranular membrane with that of other granules and with the plasmalemma to form extracellular channels for mediator release while maintaining cell viability.

In addition to exocytosis, aggregation of FcεRI initiates three additional pathways for generation of bioactive products, namely, lipid mediators, chemokines, and cytokines. Cytokines elaborated by mast cells include tumor necrosis factor α (TNF-α), interleukin (IL) 1, IL-6, IL-4, IL-5, IL-13, and granulocyte-macrophage colony-stimulating factor (GM-CSF). Chemokines include CCL3, CCL4, CCL8, and CXCL8. These products likely contribute to the recruitment of bloodborne leukocytes to the tissues following local activation of mast cells.

Lipid mediator generation **(Fig. 363-1)** involves translocation of calcium ion–dependent cytosolic phospholipase A_2 to the outer nuclear membrane, with subsequent release of arachidonic acid for metabolic processing by the distinct prostanoid and leukotriene pathways. The constitutive prostaglandin endoperoxide synthase-1 (PGHS-1/cyclooxygenase-1) and the de novo inducible PGHS-2 (cyclooxygenase-2) convert released arachidonic acid to the sequential intermediates, prostaglandins G_2 and H_2. The glutathione-dependent hematopoietic prostaglandin D_2 (PGD_2) synthase then converts PGH_2 to PGD_2, the predominant mast cell prostanoid. The PGD_2 receptor DP_1 is expressed by platelets, natural killer cells, dendritic cells, and epithelial cells, whereas DP_2 is expressed by T_H2 lymphocytes, innate lymphoid type 2 cells, eosinophils, and basophils. DP_2 signaling can promote tissue recruitment and activation of these effector cells. Mast cells also generate thromboxane A_2 (TXA_2), a short lived but powerful mediator that induces bronchoconstriction and platelet activation through the T prostanoid (TP) receptor.

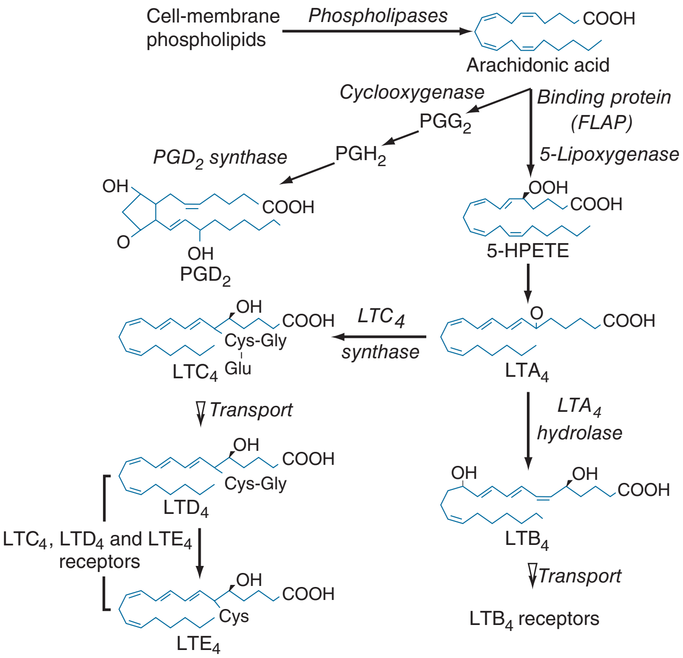

FIGURE 363-1 Pathways for biosynthesis and release of membrane-derived lipid mediators from mast cells. In the 5-lipoxygenase pathway, leukotriene A_4 (LTA_4) is the intermediate from which the terminal-pathway enzymes generate the distinct final products, leukotriene C_4 (LTC_4) and leukotriene B_4 (LTB_4), which leave the cell by separate saturable transport systems. Gamma glutamyl transpeptidase and a dipeptidase then cleave glutamic acid and glycine from LTC_4 to form LTD_4 and LTE_4, respectively. The major mast cell product of the cyclooxygenase system is PGD_2.

For leukotriene biosynthesis, the released arachidonic acid is metabolized by 5-lipoxygenase (5-LO) in the presence of an integral nuclear membrane protein, 5-LO activating protein (FLAP). The calcium ion–dependent translocation of 5-LO to the nuclear membrane converts the arachidonic acid to the sequential intermediates, 5-hydroperoxyeicosatetraenoic acid (5-HPETE) and leukotriene (LT) A_4. LTA_4 is conjugated with reduced glutathione by LTC_4 synthase, an integral nuclear membrane protein homologous to FLAP. Intracellular LTC_4 is released by a carrier-specific export step for extracellular metabolism to the additional cysteinyl leukotrienes, LTD_4 and LTE_4, by the sequential removal of glutamic acid and glycine. Alternatively, cytosolic LTA_4 hydrolase converts some LTA_4 to the dihydroxy leukotriene LTB_4, which also undergoes specific export. Two receptors for LTB_4, BLT_1 and BLT_2, mediate chemotaxis of human neutrophils. Two receptors for the cysteinyl leukotrienes, $CysLT_1$ and $CysLT_2$, are present on smooth muscle of the airways and the microvasculature and on hematopoietic cells such as macrophages, eosinophils, Th2 cells, group 2 innate lymphoid cells, and mast cells. Whereas the $CysLT_1$ receptor has a preference for LTD_4 and is blocked by the receptor antagonists in clinical use, the $CysLT_2$ receptor is equally responsive to LTD_4 and LTC_4, is unaffected by these antagonists, and is a negative regulator of the function of the $CysLT_1$ receptor. LTD_4, acting at $CysLT_1$ receptors, is the most potent known bronchoconstrictor, whereas LTE_4 induces a vascular leak and mediates the recruitment of eosinophils to the bronchial mucosa, as well as the expansion of mucosal tuft/brush cells. Additionally, GPR99, initially identified as a receptor for α-ketoglutarate, was identified as an LTE_4 receptor. The lysophospholipid formed during the release of arachidonic acid from 1-*O*-alkyl-2-acyl-*sn*-glyceryl-3-phosphorylcholine can be acetylated in the second position to form platelet-activating factor (PAF). Serum levels of PAF correlated positively with the severity of anaphylaxis to peanut in a recent study, whereas the levels of PAF acetyl hydrolase (a PAF-degrading enzyme) were inversely related to the same outcome.

Human mast cells express receptors for anaphylatoxin, C5a and C3a, toll-like receptors, receptors for epithelial alarmins thymic stromal lymphopoietin (TSLP) and IL-33, and Mas-related G protein–coupled receptor (MRGPX2), all which activate mast cells in an IgE-dependent manner. MRGPX2 is a target of many small-molecule drugs with a central tetrahydroisoquinoline motif, such as ciprofloxacin and rocuronium, which may explain the observed episodes of anaphylaxis to these medications without evidence of IgE-mediated hypersensitivity. It is also activated by neuropeptides and may participate in neural circuits that activate mast cells.

Unlike most other cells of bone marrow origin, mast cells circulate as committed progenitors lacking their characteristic secretory granules. These committed progenitors lack lineage markers and express c-*kit*, the receptor for stem cell factor (SCF), as well as FcεRI and the β_7 integrin. Unlike most other lineages, they retain and increase c-*kit* expression with maturation. The SCF interaction with c-*kit* is an absolute requirement for the development of both constitutive connective tissue and skin mast cells and for the accumulation of mast cells at mucosal surfaces during T_H2-type immune responses. Several T cell–derived cytokines (IL-3, IL-4, IL-5, and IL-9) can potentiate SCF-dependent mast cell proliferation and/or survival in vitro in mice and humans. Indeed, mast cells are absent from the intestinal mucosa in clinical T-cell deficiencies but are present in the submucosa. Historical mast cell classification has been based on the immunodetection of secretory granule neutral proteases. Mast cells in the lung parenchyma and intestinal mucosa selectively express tryptase, and those in the intestinal and airway submucosa, perivascular spaces, skin, lymph nodes, and breast parenchyma express tryptase, chymase, and carboxypeptidase A (CPA). Selective environmental cues, such as T_H2 inflammation, can lead to different protease expression; in the mucosal epithelium of severe asthmatics and apical epithelium of nasal polyps, mast cells can express tryptase and CPA without chymase. The secretory granules of mast cells selectively positive for tryptase exhibit closed scrolls with a periodicity suggestive of a crystalline structure by electron microscopy, whereas the secretory granules of mast cells with multiple proteases are scroll-poor, with an amorphous or lattice-like appearance. In addition to immunodetection of proteases, expression profiling through single-cell RNA sequencing methods has further elucidated different mast cell populations, indicating that mast cells are far more functionally heterogeneous than previously thought.

Mast cells are distributed at cutaneous and mucosal surfaces and in submucosal tissues about venules and could influence the entry of foreign substances by their rapid response capability **(Fig. 363-2)**. Upon stimulus-specific activation and secretory granule exocytosis, histamine and acid hydrolases are solubilized, whereas the neutral proteases, which are cationic, remain largely bound to the anionic proteoglycans, heparin and chondroitin sulfate E, with which they function as a complex. Histamine and the various lipid mediators (PGD_2, $LTC_4/D_4/E_4$, PAF) alter venular permeability, thereby allowing influx of plasma proteins such as complement and immunoglobulins, whereas LTB_4 and PGD_2 mediate leukocyte–endothelial cell adhesion and subsequent directed migration (chemotaxis). The accumulation of leukocytes and plasma opsonins facilitates defense of the microenvironment. The inflammatory response can also be detrimental, as in asthma, where the smooth-muscle constrictor activity of the cysteinyl leukotrienes is evident and much more potent than that of histamine.

The cellular component of the mast cell–mediated inflammatory response is augmented and sustained by cytokines and chemokines. IgE-dependent activation of human skin mast cells in situ elicits TNF-α production and release, which in turn induces endothelial cell responses favoring leukocyte adhesion. Similarly, activation of purified human lung mast cells or cord blood–derived cultured mast cells in vitro results in substantial production of proinflammatory (TNF-α) and immunomodulatory cytokines (IL-4, IL-5, IL-13) and chemokines. Bronchial biopsy specimens from patients with asthma reveal that mast cells are immunohistochemically positive for IL-4 and IL-5, but that the predominant localization of IL-4, IL-5, and GM-CSF is to T cells, defined as T_H2 by this profile. IL-4 modulates the T-cell phenotype to the T_H2 subtype, determines the isotype switch to IgE (as does IL-13), and upregulates FcεRI-mediated expression of cytokines by mast cells based on in vitro studies.

An immediate and late cellular phase of allergic inflammation can be induced in the skin, nose, or lung of some allergic humans with local allergen challenge. The immediate phase in the nose involves pruritus and watery discharge; in the lung, it involves bronchospasm and mucus secretion; and in the skin, it involves a wheal-and-flare response with pruritus. Diminished nasal patency, reduced pulmonary function, or erythema with swelling at the skin site in a late-phase response at 6–8 h is associated with biopsy findings of infiltrating and activated T_H2 cells, eosinophils, basophils, and some neutrophils. The progression from early mast cell activation to late cellular infiltration has been used as an experimental surrogate of rhinitis or asthma in allergen-challenged subjects. However, in asthma, there is an intrinsic hyperreactivity of the airways independent of the associated inflammation. Moreover, early- and late-phase responses (at least in the lung) are far more sensitive to blockade of IgE-dependent mast cell activation (or actions of histamine and cysteinyl leukotrienes) than are spontaneous or virally induced asthma exacerbations.

Consideration of the mechanism of immediate-type hypersensitivity diseases in the human has focused largely on the IgE-dependent recognition of otherwise innocuous substances. A region of chromosome 5 (5q23-31) contains genes implicated in the control of IgE levels including IL-4 and IL-13, as well as IL-3 and IL-9, which participate in mucosal mast cell hyperplasia, and IL-5 and GM-CSF, which are central to eosinophil development and their enhanced tissue viability. Genes with linkage to the specific IgE response to particular allergens include those encoding the major histocompatibility complex (MHC) and certain chains of the T-cell receptor (TCR). The complexity of atopy and the associated diseases includes susceptibility, severity, and therapeutic responses, each of which is modulated by both innate and adaptive immune stimuli.

The induction of allergic disease requires sensitization of a predisposed individual to a specific allergen. The greatest propensity for the development of atopic allergy occurs in childhood and early adolescence. The allergen is processed by antigen-presenting cells of the monocytic lineage (particularly dendritic cells) located throughout the body at epithelial surfaces that contact the outside environment, such as the nose, lungs, eyes, skin, and intestine. These antigen-presenting cells present the epitope-bearing peptides via their MHC to T helper cells and their subsets. The T-cell response depends both on cognate recognition and on the cytokine microenvironment provided by the antigen-presenting dendritic cells, with IL-4 directing a T_H2 subset, interferon (IFN) γ a T_H1 profile, and IL-6 with transforming growth factor β (TGF-β) a T_H17 subset. Allergens can induce an epithelial

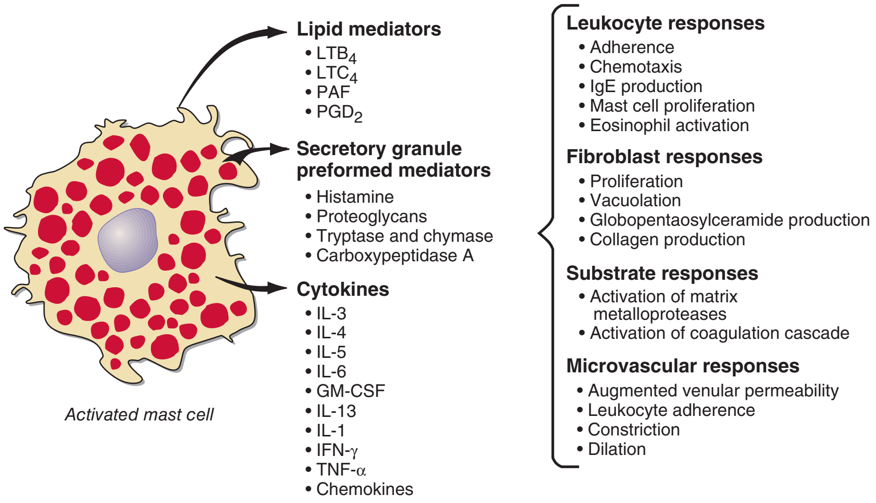

FIGURE 363-2 Bioactive mediators of three categories generated by IgE-dependent activation of murine mast cells can elicit common but sequential target cell effects leading to acute and sustained inflammatory responses. GM-CSF, granulocyte-macrophage colony-stimulating factor; IL, interleukin; IFN, interferon; LT, leukotriene; PAF, platelet-activating factor; PGD_2, prostaglandin D_2; TNF, tumor necrosis factor.

alarmin response, with expression of IL-25, TSLP, and IL-33, which stimulate group 2 innate lymphoid cells, which can generate large quantities of IL-5 and IL-13. Allergens also contain pattern recognition ligands that facilitate the immune response by direct initiation of cytokine generation from innate cell types such as basophils, mast cells, eosinophils, and others. The $\text{T}_\text{H}2$ response is associated with activation of specific B cells that can also present allergens or that transform into plasma cells for antibody production. Synthesis and release into the plasma of allergen-specific IgE results in sensitization of FcεR1-bearing cells such as mast cells and basophils, which become activated on exposure to the specific allergen. In certain diseases, including those associated with atopy, the monocyte and eosinophil populations can express a trimeric FcεR1, which lacks the β chain, and yet respond to its aggregation.

URTICARIA AND ANGIOEDEMA

DEFINITION

Urticaria and angioedema represent similar pathophysiologic processes occurring at different levels of the skin. Urticaria involves dilation of vascular structures in the superficial dermis, while angioedema originates from the deeper dermis and subcutaneous tissues. Not surprisingly, they often appear together, with roughly 40% of patients reporting both, and affect >20% of the population at some time during their life span. Urticaria can occur on any area of the body as well-circumscribed wheals with erythematous raised serpiginous borders and blanched centers that may coalesce to become giant wheals. Urticarial lesions last for <24 h, are intensely pruritic, frequently migrate around the body, and leave no bruising or scarring. Angioedema is marked by dramatic swelling with more pain than pruritus and minimal erythema, which may develop with a pruritic prodrome and takes hours to days to resolve. Acute urticaria and/or angioedema are episodes that occur for <6 weeks' duration, whereas attacks persisting for >6 weeks are designated chronic.

PREDISPOSING FACTORS AND ETIOLOGY

Acute or chronic urticaria and/or angioedema can occur at any point in the life span, with the third to fifth decades the most common for chronic disease. Women are affected more often than men, with a slight predominance for those with a history of atopy. Acute urticaria is most often the result of exposure to a food, environmental, or drug allergen or viral infection, while chronic urticaria is often spontaneous, without an identifiable cause. More than two-thirds of new-onset urticaria cases are ultimately diagnosed as acute.

The classification of urticaria and angioedema, presented in **Table 363-1**, focuses on the different precipitating factors eliciting clinical disease and can be useful for differential diagnosis.

TABLE 363-1 Classification of Urticaria and/or Angioedema

ACUTE	CHRONIC
Drug reactions	Spontaneous/idiopathic—no identifiable trigger
Including (but not limited to): antimicrobials, nonsteroidal anti-inflammatory drugs (NSAIDs), radiocontrast media, opioids, angiotensin-converting enzyme (ACE) inhibitors, dipeptidyl peptidase-4 inhibitors	Autoimmune—autoimmune component Physical stimuli (inducible urticaria) Dermatographism Cholinergic urticaria Vibration, cold, pressure, water (aquagenic) Sun (solar)
Vaccine reactions	Vascular disease—urticarial vasculitis and small vessel vasculitis
Food reactions	Mastocytosis (cutaneous or systemic)
Inhalation or contact with environmental allergens	Hereditary
Transfusion reactions	Hereditary angioedema (HAE)
Stinging and biting insects	C3b inhibitor deficiency
Toxin (scombroid)	CIAS1-associated periodic fever syndromes
Infections—viral, bacterial, parasitic	(familial cold urticaria, Muckle-Wells syndrome)
	Hypereosinophilic syndrome
	Schnitzler's syndrome
	Gleich's syndrome

Additional urticaria syndromes include inducible urticarias from physical stimuli such as cold, heat, solar UV radiation, exercise, and mechanical irritation. These urticarias can be distinguished by the precipitating event and other aspects of the clinical presentation. *Dermatographism*, which occurs in 2–5% of the population, is defined by the appearance of a linear wheal with surrounding erythema at the site of a brisk stroke with a firm object **(Fig. 363-3)**. Dermatographism has a prevalence that peaks in the second to third decades, is not influenced by atopy, and has a duration generally of <5 years. *Pressure urticaria*, which often accompanies chronic spontaneous urticaria, presents in response to a sustained stimulus such as a shoulder strap or belt, running (feet), or manual labor (hands). *Cholinergic urticaria* is distinctive in that the pruritic wheals are small (1–2 mm) and surrounded by a large area of erythema; attacks are precipitated by fever, a hot bath or shower, or exercise and are presumptively attributed to a rise in core body temperature. *Exercise-induced anaphylaxis* can be precipitated by exertion alone or can be dependent on food ingestion prior to exercise. Some cases are associated with the presence of IgE specific for α-5 gliadin, a component of wheat. The clinical presentation can be limited to flushing, erythema, and pruritic urticaria but may progress to angioedema of the face, oropharynx, larynx, or intestine and/or to vascular collapse; the wheals are of conventional size and are not induced with passive heating. *Solar urticaria* is subdivided into six groups by the response to specific portions of the light spectrum. *Cold urticaria* is local at body areas exposed to low ambient temperature or cold objects but can progress to vascular collapse with immersion in cold water (swimming). *Vibratory urticaria and angioedema* may occur after years of occupational exposure or can be idiopathic; it may be accompanied by cholinergic urticaria. In rare cases, variants of cold and vibratory urticaria are inherited and syndromic conditions, with mutations in the NLRP3 component of inflammasome leading to familial cold autoinflammatory syndrome, and mutations in the mast cell mechanoreceptor ADGRE2 associated with familial vibratory urticaria. Other rare forms of physical allergy, always defined by stimulus-specific elicitation, include *local heat urticaria*, *aquagenic urticaria* from contact with water of any temperature (sometimes associated with polycythemia vera), and *contact urticaria* from direct interaction with chemical or protein sensitizer (such as latex).

Isolated Angioedema Angioedema without urticaria can be idiopathic or due to the generation of bradykinin in the setting of C1 inhibitor (C1INH) deficiency that may be inborn as an autosomal dominant mutation or acquired through the appearance of an autoantibody in the setting of malignancy or autoimmune disease. Angiotensin-converting enzyme (ACE) inhibitors can provoke a similar clinical presentation in 0.2–0.7% of exposed patients due to delayed degradation of bradykinin. Black race, organ transplant, female gender, smoking, and increasing age are known risk factors for ACE inhibitor–related angioedema.

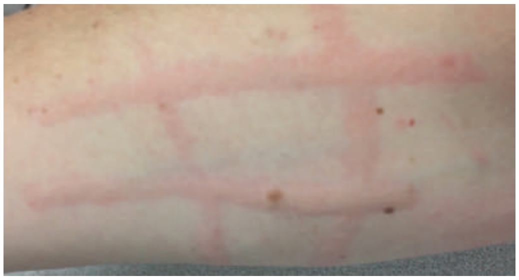

FIGURE 363-3 Dermatographic urticarial lesion induced by stroking the forearm lightly with the edge of a tongue blade. The photograph, taken after 10 min, demonstrates a prominent wheal-and-flare reaction in the shape of a *hashtag*. *(Photograph provided by Katherine N. Cahill, MD, Harvard Medical School.)*

CLINICAL PRESENTATION AND PATHOPHYSIOLOGY

Urticarial eruptions are distinctly pruritic, may involve any area of the body from the scalp to the soles of the feet, and appear in crops of 12- to 36-h duration, with old lesions fading as new ones appear. Most of the inducible urticarias (cold, cholinergic, dermatographism) are an exception, with individual lesions lasting <2 h. Neither urticaria nor angioedema lesions are symmetric or dependent in distribution. The most common sites for angioedema are often periorbital and perioral. Angioedema of the upper respiratory tract may be life-threatening due to transient laryngeal obstruction, whereas gastrointestinal involvement may present with abdominal colic, with or without nausea and vomiting, and can result in unnecessary surgical intervention. No residual scarring occurs with either urticaria or angioedema unless there is an underlying vasculitic process.

The pathology is characterized by edema of the superficial dermis in urticaria and of the subcutaneous tissue and deep dermis in angioedema. Collagen bundles in affected areas are widely separated, and the venules are sometimes dilated. Any perivenular infiltrate consists of lymphocytes, monocytes, eosinophils, and neutrophils that are present in varying combination and numbers.

The best evidence for IgE and mast cell involvement in urticaria and angioedema is *cold urticaria*. Cryoglobulins or cold agglutinins are present in up to 5% of these patients. Ice cube placement on the volar forearm precipitates urticaria or angioedema within minutes of the challenge. Histologic studies reveal marked mast cell degranulation with associated edema of the dermis and subcutaneous tissues. Elevated levels of histamine have been found in the plasma of venous effluent and in the fluid of suction blisters at experimentally induced lesional sites in patients with cold urticaria, dermographism, pressure urticaria, vibratory angioedema, light urticaria, and heat urticaria. By ultrastructural analysis, the pattern of mast cell degranulation in cold urticaria resembles an IgE-mediated response with solubilization of granule contents, fusion of the perigranular and cell membranes, and discharge of granule contents, whereas in a dermatographic lesion, there is additional superimposed zonal (piecemeal) degranulation. Elevations of plasma histamine levels with biopsy-proven mast cell degranulation have also been demonstrated with generalized attacks of *cholinergic urticaria*.

Up to 45% of patients with chronic urticaria have an autoimmune cause for their disease including autoantibodies to IgE or to the α chain of FcεRI. In some patients, autologous serum injected into their own skin can induce a wheal-and-flare reaction involving mast cell activation. The presence of these antibodies can also be recognized by their capacity to release histamine or induce activation markers such as CD63 or CD203 on basophils. An association with antibodies to microsomal peroxidase and/or thyroglobulin has been observed with both clinically significant Hashimoto's thyroiditis as well as a euthyroid state.

The urticaria and angioedema associated with classic serum sickness or with hypocomplementemic cutaneous necrotizing angiitis (urticarial vasculitis) are believed to be immune-complex-mediated diseases.

Isolated Angioedema Hereditary angioedema (HAE) is a fully penetrant, autosomal dominant disease due to a mutation in the *SERPING1* gene leading to a deficiency of C1INH (type 1) in ~85% of patients or to a dysfunctional protein (type 2) in the remainder affecting 1:30,000–80,000 in the general population. Presentation in childhood or adolescence and a family history of angioedema are distinguishing features. C1INH deficiency can also develop in a sporadic acquired form as a result of excessive consumption of C1INH due either to formation of immune complexes or to the generation of an autoantibody directed to C1INH in the setting of lymphoproliferative or autoimmune disease. A third, less common type of HAE has been described in which C1INH function is normal; a causal genetic lesion has been implicated in many of these cases, including a mutant form of factor XII, which leads to generation of excessive bradykinin. C1INH blocks the catalytic function of activated factor XII (Hageman factor) and of kallikrein, as well as the C1r/C1s components of C1, with the common result of degrading bradykinin. During clinical attacks of angioedema, C1INH function or levels fall, patients develop elevated plasma levels of bradykinin leading to angioedema, and excessive activation of C1 lowers C4 and C2 levels.

Bradykinin-mediated angioedema, whether caused by ACE inhibitors or by C1INH deficiency, is noteworthy for the conspicuous absence of concomitant urticaria or pruritus, the frequent involvement of the gastrointestinal tract, and the duration of symptoms >24 h.

DIAGNOSIS

The classification of urticaria and angioedema as presented in Table 363-1 in terms of duration can facilitate identification of possible precipitating factors. History alone of self-limited episodes can be sufficient to make a diagnosis in the setting of acute disease triggered by drug, environmental, or food allergen with history-directed confirmatory skin testing or assay for serum allergen-specific IgE. Provocation and threshold testing for inducible urticaria is particularly valuable because it so often establishes the cause of the lesion. In chronic urticaria with or without angioedema, initial diagnostic testing should be guided by history and physical exam. Practice guidelines provide clinicians two options if history and physical exam are unrevealing: no laboratory testing or limited testing, which includes complete blood count with assessment for eosinophilia, erythrocyte sedimentation rate, or C-reactive protein. Extended testing, such as in thyroid-related tests, varies by guideline. The vast majority of chronic urticaria is associated with no laboratory abnormality. Urticarial lesions that last longer than 36 h, result in scarring, and are reported as painful and not pruritic warrant biopsy to evaluate for cellular infiltration, nuclear debris, and fibrinoid necrosis of the venules consistent with urticarial vasculitis. Isolated angioedema without urticaria additionally warrants assessment of basic chemistries with liver function tests and complement levels. Further testing, including for paraproteinemia, may be guided by patient presentation. Concomitant flushing and hyperpigmented papules that urticate with stroking in the absence of angioedema raise the question of mastocytosis. An appropriate travel history should trigger an evaluation for parasites. Symptom activity and burden can be assessed using validated scoring instruments.

The diagnosis of HAE is suggested not only by family history and age of onset but also by the lack of pruritus and of urticarial lesions, the prominence of recurrent gastrointestinal attacks of colic, episodes of laryngeal edema, and (in contrast to urticaria) a failure to respond to H1 antagonists. Laboratory diagnosis depends on demonstrating a deficiency of C1INH antigen (type 1) or a nonfunctional protein (type 2) by a catalytic inhibition assay. C4 and C2 are chronically depleted and fall further during attacks due to the activation of additional C1. Patients with the acquired forms of C1INH deficiency have the same clinical manifestations but differ in the lack of a familial history. Furthermore, their sera exhibit a reduction of C1 function and C1q protein as well as C1INH, C4, and C2. Lastly, type 3 HAE is associated with normal levels of complement proteins and often an identifiable gene mutation.

TREATMENT

Urticaria and Angioedema

For most forms of urticaria, H_1 antihistamines effectively attenuate both urtication and pruritus; long-acting, nonsedating agents, such as loratadine and fexofenadine, or low-sedating agents, such as cetirizine, generally are used first and can be increased to up to four times daily dosing. Earlier generation antihistamines, such as diphenhydramine, are sedating and induce psychomotor impairment. Their anticholinergic (muscarinic) effects include visual disturbance, urinary retention, and constipation. Some clinical practice guidelines indicate that the addition of an H_2 antagonist such as ranitidine or famotidine in conventional dosages and a $CysLT_1$ receptor antagonist, such as montelukast 10 mg daily, may add benefit when H_1 antihistamines are inadequate. For chronic urticaria that fails to respond to first-line options, a monoclonal anti-IgE antibody (omalizumab) is the next line of therapy. Older agents with antihistamine properties such as doxepin, cyproheptadine, and hydroxyzine have proven effective when H_1 antihistamines fail but are less effective than omalizumab and are sedating. Cyclosporine is efficacious for patients with chronic spontaneous urticaria that is severe and poorly responsive to other modalities. Phase III trial data of a Btk inhibitor, remibrutinib, showed promise in the treatment of chronic spontaneous urticaria that resisted improvement with H1 antagonists.

Topical glucocorticoids are of no value, and systemic glucocorticoids are generally avoided in idiopathic, allergen-induced, or inducible urticarias due to long-term toxicity. Systemic glucocorticoids are useful in the management of patients with pressure urticaria, vasculitic urticaria (especially with eosinophil prominence), idiopathic angioedema with or without urticaria, or chronic urticaria that responds poorly to conventional treatment and should be considered in any patient with debilitating disease. With persistent vasculitic urticaria, hydroxychloroquine, dapsone, or colchicine may be added to the regimen before or along with systemic glucocorticoids. In certain contexts, refractory chronic spontaneous urticaria may be treated with immunosuppressants (i.e., calcineurin inhibitors) and anti-inflammatory agents (i.e., hydroxychloroquine), though evidence is lacking.

BRADYKININ-MEDIATED ANGIOEDEMA

Infusion of plasma-derived C1INH protein, a monoclonal anti-plasma kallikrein antibody (lanadelumab), and a small-molecule plasma kallikrein inhibitor (berotralstat) are approved for prophylaxis of HAE attacks. Administration of plasma-derived or recombinant C1INH protein, a bradykinin 2 receptor antagonist (icatibant), or a kallikrein inhibitor (ecallantide) may be used for treatment of an acute attack of HAE. Older, less expensive preventative options include attenuated androgens, which stimulate production by the normal gene of an amount of functional C1INH. The antifibrinolytic agent ε-aminocaproic acid may be used for preoperative prophylaxis but is contraindicated in patients with thrombotic tendencies or arterial atherosclerosis. Fresh frozen plasma infusion can be used for acute attacks in a setting where newer modalities are not available. Published studies are conflicting on the efficacy of bradykinin 2 receptor antagonists and C1INH protein in the treatment of ACE inhibitor–induced angioedema. Treatment of the underlying autoimmune disease or malignancy is indicated for acquired C1INH deficiency.

ALLERGIC RHINITIS

DEFINITION

Rhinitis is characterized by sneezing; rhinorrhea; obstruction of the nasal passages; conjunctival, nasal, and pharyngeal itching; and lacrimation. It can be classified as allergic or nonallergic based on the presence or absence of IgE specific for relevant antigens. A clinical history of rhinitis symptoms occurring in a temporal relationship to allergen exposure and documentation of sensitization to an environmental allergen are required for diagnosis. Although commonly seasonal due to elicitation by airborne pollens, it can be perennial in an environment of chronic exposure to house dust mites, animal danders, or insect (cockroach) products. The overall prevalence in North America has increased in the past 20 years and is 10–30%, with the peak prevalence (>30%) occurring in the fifth decade.

PREDISPOSING FACTORS AND ETIOLOGY

Allergic rhinitis generally occurs in atopic individuals, often in association with atopic dermatitis, food allergy, urticaria, and/or asthma **(Chap. 298)**. Up to 50% of patients with allergic rhinitis manifest asthma, whereas 70–80% of individuals with asthma and 80% of individuals with chronic bilateral sinusitis experience allergic rhinitis. Female sex, particulate air pollution exposure, and maternal tobacco smoking increase the risk of developing allergic rhinitis.

Wind-pollinated trees, grasses, and weeds produce sufficient quantities of pollen suitable for wide distribution by air currents to elicit seasonal allergic rhinitis. The dates of pollination of these species historically varied little from year to year in a particular locale but may differ in another climate. In the temperate areas of North America, trees typically pollinate from March through May, grasses in June and early July, and weeds from mid-August to early October. Molds, which are widespread in nature because they occur in soil or decaying organic matter, propagate spores in a pattern that depends on climatic conditions. Climate change is impacting these patterns with early tree pollination and prolonged ragweed season with the delay of the first frost. In laboratory studies, exposure to high carbon dioxide concentrations increases pollen production in ragweed and timothy grass. Perennial allergic rhinitis occurs in response to allergens that are present throughout the year, including animal dander, cockroach-derived proteins, mold spores, or dust mites such as *Dermatophagoides farinae* and *Dermatophagoides pteronyssinus*. Dust mites are scavengers of human skin and excrete cysteine protease allergens in their feces. Similarly, cockroach allergens are fecal cysteine proteases.

PATHOPHYSIOLOGY AND MANIFESTATIONS

Episodic rhinorrhea, sneezing, obstruction of the nasal passages with lacrimation, and pruritus of the conjunctiva, nasal mucosa, and oropharynx are the hallmarks of allergic rhinitis. The nasal mucosa is pale and boggy, the conjunctiva congested and edematous, and the pharynx generally unremarkable. Swelling of the turbinates and mucous membranes with obstruction of the sinus ostia and eustachian tubes precipitates secondary infections of the sinuses and middle ear, respectively. A growing number of patients with seasonal allergic rhinitis demonstrate pollen-associated food allergen syndrome characterized by oropharyngeal pruritus and/or mild swelling following the ingestion of plant-based foods in the same plant family as a tree, grass, or weed, which contain cross-reacting allergens.

The nose presents a large mucosal surface area through the folds of the turbinates and serves to adjust the temperature and moisture content of inhaled air and to filter out particulate materials >10 μm in size by impingement in a mucous blanket; ciliary action moves the entrapped particles toward the pharynx. Entrapment of pollen and digestion of the outer coat by mucosal enzymes such as lysozymes release protein allergens. The initial interaction occurs between the allergen and intraepithelial mast cells and then proceeds to involve deeper perivenular mast cells, both of which are sensitized with specific IgE. During the symptomatic season when the mucosae are already swollen and hyperemic, there is enhanced adverse reactivity to the seasonal pollen as well as irritants such as tobacco smoke and fragrances. Biopsy specimens of nasal mucosa during seasonal rhinitis show submucosal edema with infiltration by eosinophils, along with some basophils and neutrophils, and intraepithelial mast cells.

The mucosal surface fluid contains IgA and IgE, which apparently arrives by diffusion from plasma cells in proximity to mucosal surfaces. IgE fixes to mucosal and submucosal mast cells, and the intensity of the clinical response to inhaled allergens is quantitatively related to the naturally occurring pollen dose. In sensitive individuals, the introduction of allergen into the nose is associated with sneezing, nasal obstruction, and discharge, and the fluid contains histamine, PGD_2, and leukotrienes. Thus, the mast cells of the nasal mucosa and submucosa generate and release mediators through IgE-dependent reactions that are capable of producing tissue edema and eosinophilic infiltration.

DIAGNOSIS

The diagnosis of seasonal allergic rhinitis depends largely on an accurate history of symptoms lasting >4 weeks coincident with the pollination of the offending allergen. The continuous character of perennial allergic rhinitis due to allergens found in the home or place of work makes historic analysis difficult, but there may be variability in symptoms that can be related to exposure to animal dander, dust mite and/or cockroach allergens, fungal spores, or work-related allergens such as latex. Patients with perennial rhinitis commonly develop the problem in adult life and manifest nasal congestion and a postnasal discharge, often associated with thickening of the sinus membranes demonstrated by radiography. Perennial nonallergic rhinitis with eosinophilia syndrome (NARES) occurs in the middle decades of life and is characterized by nasal obstruction, anosmia, chronic sinusitis, and prominent eosinophilic nasal discharge in the absence of allergen sensitization. The term *vasomotor rhinitis* or *perennial nonallergic rhinitis* designates a condition of enhanced reactivity of the nasopharynx in which a symptom complex resembling perennial allergic rhinitis occurs with nonspecific stimuli, including chemical odors, temperature and humidity variations, and position changes but occurs without tissue eosinophilia or an allergic etiology. Other entities to be excluded are structural abnormalities of the nasopharynx, including nasal foreign bodies in children, symptoms due to exposure to irritants, gustatory rhinitis associated with cholinergic activation that occurs while eating or ingesting alcohol, hypothyroidism, upper respiratory tract infection, hormonal rhinitis, including pregnancy, atrophic rhinitis in populations outside the United States and Europe, prolonged topical use of α-adrenergic agents in the form of nasal sprays (rhinitis medicamentosa), and the use of certain systemic agents such as β-adrenergic antagonists, ACE inhibitors, direct vasodilators (hydralazine), α_1-adrenergic receptor antagonists, estrogens, progesterone, nonsteroidal anti-inflammatory drugs, gabapentin, phosphodiesterase-5 inhibitors, and psychotropics (risperidone, chlorpromazine, amitriptyline).

The nasal secretions of allergic patients can be rich in eosinophils, and a modest peripheral eosinophilia can be observed. Local neutrophilia implies infection. Total serum IgE may be moderately elevated, but the demonstration of immunologic specificity for IgE is critical to an etiologic diagnosis. A skin test by the intracutaneous route (puncture or prick) with the allergens of interest provides a rapid and reliable approach to identifying allergen-specific IgE that has sensitized cutaneous mast cells. A positive intracutaneous skin test with 1:10–1:20 weight/volume of extract has a high predictive value for the presence of allergy. An intradermal test with a 1:500–1:1000 dilution of 0.05 mL may follow if indicated by history when the intracutaneous test is negative, but while more sensitive, it is less dependable due to the reactivity of some asymptomatic individuals at the test dose.

Enzyme-linked immunosorbent assays (ELISAs) employing anti-IgE bound to either a solid-phase or a liquid-phase particle provide rapid and cost-effective determinations of IgE. Measurements of specific anti-IgE in serum are obtained by its binding to an allergen and quantitation by subsequent uptake of labeled anti-IgE. As compared to the skin test, the assay of specific IgE in serum is less sensitive but has high specificity.

TREATMENT

Allergic Rhinitis

Although allergen avoidance is the most cost-effective means of managing allergic rhinitis, only in the case of animal dander and possibly dust mites is it feasible. Treatment with pharmacologic agents represents the standard initial approach to seasonal or perennial allergic rhinitis. Oral long-acting H_1 antihistamines, such as fexofenadine, loratadine, desloratadine, cetirizine, and levocetirizine, are effective for nasopharyngeal itching, sneezing, and watery rhinorrhea and for such ocular manifestations as itching, tearing, and erythema, but they are less efficacious for the nasal congestion. They reduce nasal and ocular symptoms by about one-third. These antihistamines are less lipophilic and more H_1 selective, thus minimizing their ability to cross the blood-brain barrier and therefore diminishing their sedating and anticholinergic effect; they do not differ appreciably in efficacy for relief of rhinitis and/or sneezing.

Intranasal high-potency glucocorticoids are the most effective drugs available for the relief of persistent rhinitis, seasonal or perennial, and are effective in relieving nasal congestion as well as ocular symptoms. They provide efficacy with substantially reduced side effects as compared with this same class of agent administered orally. Their most frequent side effect is local irritation, with fungal overgrowth being a rare occurrence. The currently available intranasal glucocorticoids are equally effective for nasal symptom relief, including nasal congestion, achieving up to 70% overall symptom relief with some variation in the onset of benefit. The nasal antihistamines azelastine and olopatadine as a first-line option may benefit individuals with nonallergic rhinitis and intermittent or mild allergic rhinitis, as well as have additive benefit to intranasal steroids in persistent allergic rhinitis, but they have an adverse effect of dysgeusia in some patients. Alternative nasal decongestants for significant mucosal edema or unresponsive nasal congestion include α-adrenergic agents such as phenylephrine or oxymetazoline; however, the duration of their efficacy is limited because of rebound rhinitis (i.e., 7- to 14-day use can lead to rhinitis medicamentosa) and such systemic responses as hypertension. Oral α-adrenergic agonist decongestants containing pseudoephedrine can improve management of nasal congestion, generally in combination with an antihistamine. These pseudoephedrine combination products can cause insomnia and are precluded from use in a variety of patient populations, such as those with narrow-angle glaucoma, urinary retention, uncontrolled hypertension, cerebrovascular disease, marked coronary artery disease, or a first-trimester pregnancy. The CysLT_1 antagonist montelukast is approved for treatment of both seasonal and perennial rhinitis. However, it is less effective than H_1 antihistamines and nasal glucocorticoids, and reports of neuropsychiatric events have led to increased U.S. Food and Drug Administration precautions. Cromolyn sodium nasal spray inhibits mast cell degranulation and can be used prophylactically on a continuous basis during allergen season or as needed before a known exposure. Topical ipratropium is an anticholinergic agent effective in reducing rhinorrhea, including that of patients with perennial nonallergic symptoms. For concomitant allergic conjunctivitis, topical treatment with cromolyn sodium is effective in treating mild allergic symptoms, and topical antihistamines such as olopatadine, azelastine, ketotifen, or epinastine administered to the eye provide rapid relief of itching and redness and are more effective than oral antihistamines.

Immunotherapy Immunotherapy consists of repeated exposure to the allergen(s) considered to be specifically responsible for the symptom complex. Two forms of immunotherapy, subcutaneous (SCIT) and sublingual (SLIT), are currently available. Randomized, controlled studies of ragweed, grass, dust mite, and cat dander allergens administered via SCIT for treatment of allergic rhinitis have demonstrated significant improved symptom control over medications alone with the advantage of providing a durable benefit, as well as a reduction in asthma symptoms, medication use, and bronchial hyperreactivity in allergic asthma. Clinical practice guidelines recommend a duration of SCIT of 3–5 years, with discontinuation being based on minimal symptoms over two consecutive seasons of exposure to the allergen. Clinical benefit appears related to the administration of a high dose of relevant allergen, gradually uptitrating concentration and advancing from weekly to monthly intervals. SCIT injections occur in a licensed treatment site; 2–3% of SCIT patients experience a systemic reaction, including anaphylaxis, over a 12-month period. The majority of these reactions occur soon after injection, and thus, patients should remain at the treatment site for at least 30 min after allergen administration. Local reactions with erythema and induration are not uncommon and may persist for 1–3 days. SLIT is prepared as a tablet to be dissolved under the tongue at home after the first dose. The efficacy of SLIT is comparable to SCIT but only for the three allergen formulations currently available: dust mite, timothy/northern grasses, and short ragweed. Systemic

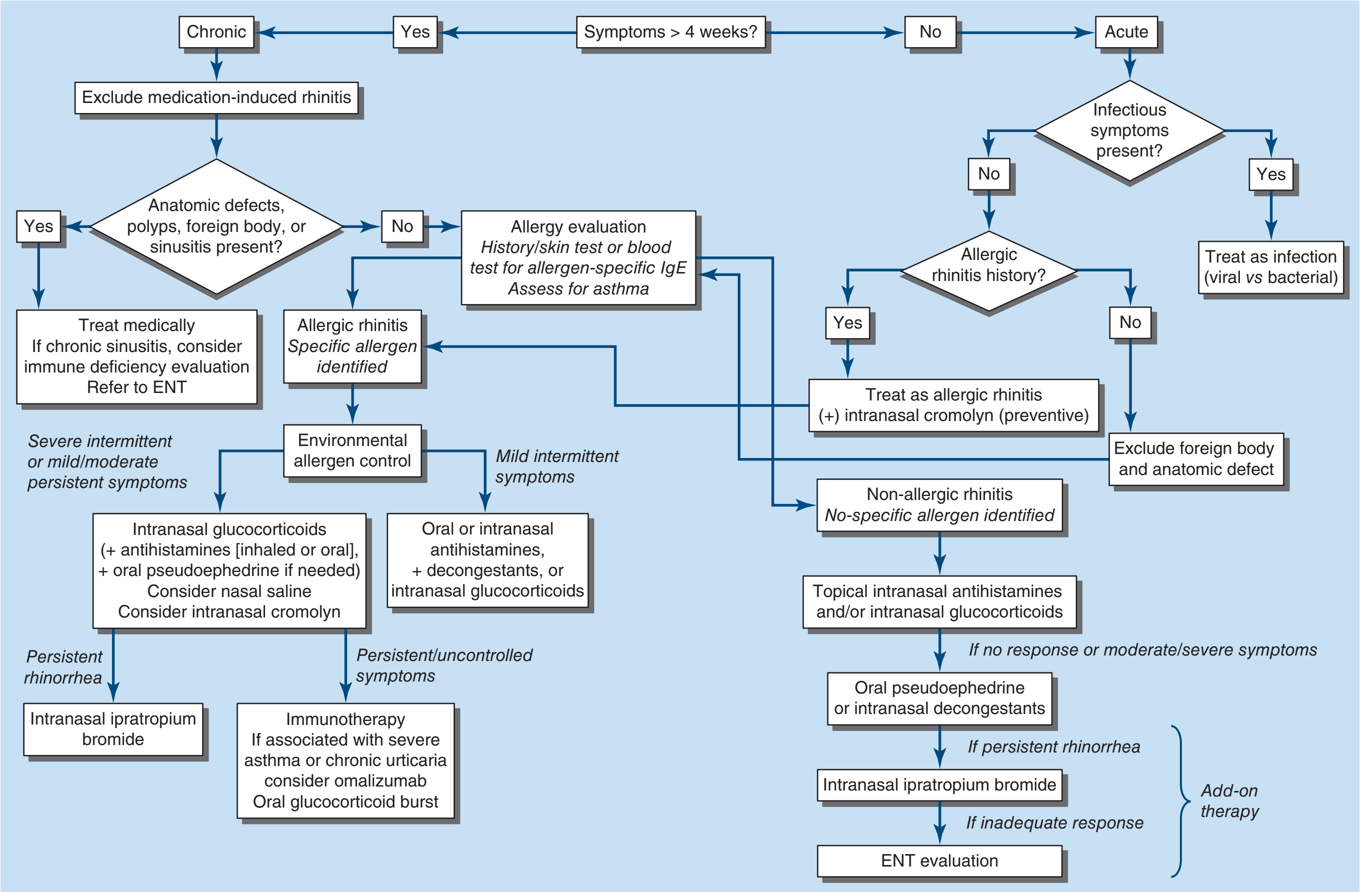
Symptoms > 4 weeks?
Yes
Chronic
No
Acute
Exclude medication-induced rhinitis
Anatomic defects, polyps, foreign body, or sinusitis present?
Yes
No
Treat medically
If chronic sinusitis, consider immune deficiency evaluation
Refer to ENT
Allergy evaluation
History/skin test or blood test for allergen-specific IgE
Assess for asthma
Allergic rhinitis
Specific allergen identified
Environmental allergen control
Severe intermittent or mild/moderate persistent symptoms
Mild intermittent symptoms
Intranasal glucocorticoids (+ antihistamines [inhaled or oral], + oral pseudoephedrine if needed)
Consider nasal saline
Consider intranasal cromolyn
Oral or intranasal antihistamines, + decongestants, or intranasal glucocorticoids
Persistent rhinorrhea
Persistent/uncontrolled symptoms
Intranasal ipratropium bromide
Immunotherapy
If associated with severe asthma or chronic urticaria consider omalizumab
Oral glucocorticoid burst
Infectious symptoms present?
No
Yes
Treat as infection (viral vs bacterial)
Allergic rhinitis history?
Yes
No
Treat as allergic rhinitis (+) intranasal cromolyn (preventive)
Exclude foreign body and anatomic defect
Non-allergic rhinitis
No-specific allergen identified
Topical intranasal antihistamines and/or intranasal glucocorticoids
If no response or moderate/severe symptoms
Oral pseudoephedrine or intranasal decongestants
If persistent rhinorrhea
Intranasal ipratropium bromide
If inadequate response
ENT evaluation
Add-on therapy

FIGURE 363-4 Algorithm for the diagnosis and management of rhinitis. Persistent is defined as >4 days per week for >4 weeks. Moderate/severe is defined as abnormal sleep, impaired daily activities (school, work, sport, leisure), and/or troublesome symptoms. CysLT, cysteinyl leukotriene; ENT, ear, nose, and throat; IgE, immunoglobulin E.

reactions are less frequent with SLIT, but transient oral pruritus is common. Immunotherapy is contraindicated in patients with significant cardiovascular disease or unstable asthma. Severe cases of anaphylaxis have occurred after allergen immunotherapy when patients were taking a β-adrenergic blocking agent. Thus, immunotherapy should be conducted with caution in any patient requiring β-adrenergic blocking therapy.

Immunotherapy should be reserved for clearly documented seasonal or perennial rhinitis that is clinically related to defined allergen exposure with confirmation by the presence of allergen-specific IgE through skin or serum-specific IgE testing. The response to immunotherapy is associated with a complex of cellular and humoral effects that includes a modulation in T lymphocyte cytokine production and allergen-specific IgG_4 expansion. Systemic treatment with omalizumab, an anti-IgE monoclonal antibody, is efficacious for allergic rhinitis and can be used with immunotherapy to enhance safety and efficacy. However, omalizumab is not currently approved for treatment of allergic rhinitis.

A sequence for the management of allergic or perennial rhinitis based on an allergen-specific diagnosis and stepwise management as required for symptom control would include the following: (1) identification of the offending allergen(s) by history with confirmation of the presence of allergen-specific IgE by skin test and/or serum assay; (2) avoidance of the offending allergen if possible; and (3) medical management in a stepwise fashion with treatment stepdown as symptoms permit **(Fig. 363-4)**. Mild symptoms of allergic rhinitis are treated with oral antihistamines, intranasal antihistamines, intranasal corticosteroids, or intranasal cromolyn (for short-term prevention). Moderate to more severe allergic rhinitis is managed with intranasal glucocorticoids and intranasal antihistamines; additional therapies include oral antihistamines, antihistamine-decongestant combinations, and intranasal anticholinergics. Persistent or seasonal moderate to severe allergic rhinitis that remains uncontrolled with maximal medical therapy or coexisting allergic rhinitis with asthma merit consideration of allergen-specific immunotherapy.

FURTHER READING

Bernstein DI et al: Allergic rhinitis: Mechanisms and treatment. Immunol Allergy Clin North Am 36:261, 2016.

Cho SH et al: Chronic rhinosinusitis without nasal polyps. J Allergy Clin Immunol Pract 4:575, 2016.

Cicardi M et al: Classification, diagnosis, and approach to treatment for angioedema: Consensus report from the Hereditary Angioedema International Working Group. Allergy 69:602, 2014.

Corren J et al: Allergic and nonallergic rhinitis, in *Middleton's Allergy: Principles and Practice*, 8th ed. NF Adkinson et al (eds). Philadelphia, Saunders, 2014, pp 664–685.

Dwyer DF et al: Human airway mast cells proliferate and acquire distinct inflammation-driven phenotypes during type 2 inflammation. Sci Immunol 6:56, 2021.

Dykewicz MS et al: Rhinitis 2020: A practice parameter update. J Allergy Clin Immunol 146:4, 2020.

Jutel M et al: International consensus on allergen immunotherapy II: Mechanisms, standardization, and pharmacoeconomics. J Allergy Clin Immunol 137:358, 2016.

Maurer M et al: Omalizumab for the treatment of chronic idiopathic or spontaneous urticaria. N Engl J Med 368:924, 2013.

Saini SS: Urticaria and angioedema, in *Middleton's Allergy: Principles and Practice*, 8th ed. NF Adkinson et al (eds). Philadelphia, Saunders, 2014, pp 575–587.

Zuberbier T et al: The international EAACI/GA^{2}LEN/EuroGuiDerm/APAAACI guideline for the definition, classification, diagnosis and management of urticaria. Allergy 77:734, 2022.

364 Anaphylaxis

David Hong, Joshua A. Boyce

BACKGROUND

Anaphylaxis is a potentially life-threatening systemic allergic reaction involving one or more organ systems that typically occurs within seconds to minutes of exposure to the anaphylactic trigger, most often a drug, food, or Hymenoptera sting. The term *anaphylaxis* was first described in 1902 by Charles Richet and Paul Portier who attempted to immunize dogs against sea anemone toxin in the same way Pasteur was able to vaccinate individuals against the smallpox virus. To their surprise, repeated administration of small, sublethal doses of sea anemone toxin reliably induced acute-onset death when readministered 2–3 weeks after initial "vaccination" to the toxin. The phenomenon was termed ana (anti)-phylaxis ("protection or guarding") because vaccination with anemone toxin resulted in the opposite intended immune effect. Charles Richet was awarded the Nobel Prize in Physiology or Medicine in 1913 for this work which led to further insights into hypersensitivity and mast cell biology.

CLINICAL MANIFESTATIONS

While 80–90% of anaphylactic episodes are uniphasic, about 10–20% of cases are biphasic, in which anaphylactic symptoms return about an hour or longer after resolution of initial symptoms. Anaphylactic reactions are particularly dangerous when hypotension or hypoxia occurs, leading potentially to cardiovascular collapse or respiratory failure, respectively. There may be upper or lower airway obstruction or both. Laryngeal edema may be experienced as a "lump" in the throat, hoarseness, or stridor, whereas bronchial obstruction is associated with a feeling of tightness in the chest and/or audible wheezing. Patients with underlying asthma are predisposed to severe involvement of the lower airways and increased mortality associated with anaphylaxis. In fatal cases with clinical bronchial obstruction, the lungs show marked hyperinflation on gross and microscopic examination. The microscopic findings in the bronchi, however, are limited to luminal secretions, peribronchial congestion, submucosal edema, and eosinophilic infiltration, and the acute emphysema is attributed to intractable bronchospasm that subsides with death. Angioedema resulting in death by mechanical obstruction occurs in the epiglottis and larynx. This process can also be evident in the hypopharynx and trachea. On microscopic examination, there is wide separation of the collagen fibers and the glandular elements. Vascular congestion and eosinophilic infiltration may also be present. Patients dying of vascular collapse without antecedent hypoxia from respiratory insufficiency have visceral congestion with a presumed loss of intravascular fluid volume. The associated electrocardiographic abnormalities in some patients, with or without infarction, may reflect a primary cardiac event mediated by mast cells (which are prominent near the coronary vessels) or may be secondary to a critical reduction in intravascular volume.

Gastrointestinal manifestations represent another severe presentation of anaphylaxis and include nausea, vomiting, crampy abdominal pain, and/or fecal incontinence. Angioedema of the bowel wall may also cause sufficient intravascular volume depletion to precipitate cardiovascular collapse.

Cutaneous manifestations are among the most common presentations of anaphylaxis (>90% of cases). Symptoms include urticarial eruptions, flushing with diffuse erythema, and/or a feeling of generalized warmth. Urticarial eruptions are intensely pruritic and may be localized or disseminated. They may coalesce to form large urticarial plaques but seldom persist beyond 48 h.

PATHOPHYSIOLOGY

Many of the important early mediators of anaphylaxis are derived from mast cells, basophils, and eosinophils. Mast cells contain preformed granules comprised of histamine, proteases (tryptase, chymase), proteoglycans (heparin, chondroitin sulfate), and tumor necrosis factor-α, which are rapidly released into surrounding tissue upon cell activation, a process known as degranulation. Basophils, like mast cells, contain and release histamine. Mast cells, basophils, and eosinophils are also sources of arachidonic acid–derived products, which include cysteinyl leukotrienes, prostaglandins, and platelet-activating factor (PAF). Histamine release results in flushing, urticaria, pruritus, and, in high concentrations, hypotension and tachycardia. Cysteinyl leukotrienes and prostaglandin D_2 cause bronchoconstriction and increased microvascular permeability. Prostaglandin D_2 causes cutaneous flushing and attracts eosinophils and basophils to the site of mast cell activation. Serum PAF levels correlate with anaphylaxis severity and are inversely proportional to the constitutive level of PAF acetylhydrolase, which is necessary for PAF inactivation. Tryptase and chymase can activate complement and coagulation pathways. Activation of these pathways results in production of the anaphylatoxins, C3a and C5a, and activation of the kallikrein-kinin system, which regulates blood pressure and vascular permeability. The actions of these anaphylactic mediators are likely additive or synergistic at the target tissues.

PREDISPOSING FACTORS AND MECHANISMS

Because the most dangerous manifestations of anaphylaxis involve the cardiovascular and/or respiratory systems, preexisting asthma and underlying cardiovascular disease could lead to more rapid decompensation from anaphylaxis. Atopy is not generally thought to be a risk factor for anaphylaxis from drug reactions or Hymenoptera stings, but it is associated with radiocontrast sensitivity, exercise-induced anaphylaxis, idiopathic anaphylaxis, and allergy to foods or latex. Severe Hymenoptera-induced anaphylaxis (generally with prominent hypotension) is the most common initial presentation for patients with underlying systemic mastocytosis. For this reason, it is important to check a baseline tryptase (a reflection of mast cell burden) in patients who present with sting-induced anaphylaxis to screen for this condition. Hymenoptera allergy is also more likely in patients whose occupations (i.e., beekeepers, trash haulers, and landscape workers) place them in regular proximity to stinging insects. Most commonly, allergen-induced cross-linking of IgE-bound FcεRI receptors on mast cells and basophils initiates the signal transduction events leading to hypersensitivity syndromes, including anaphylaxis. The generation of allergen-specific IgE is the end result of sensitization via the adaptive immune system. While the mechanisms underlying sensitization are beyond the scope of this chapter, environmental factors, innate immune responses, and cytokines are among the many variables leading to antigen-specific IgE production by B cells and plasma cells. IgE-mediated drug allergies are most common with antibiotics and certain chemotherapy drugs, though theoretically, they can occur with almost any medication. As is the case with environmental allergies, repeated exposure to the allergy-causing antigen is an important risk factor to keep in mind when evaluating patients with anaphylaxis. In the case of allergy to carboplatin, the incidence of hypersensitivity is 27% in patients who have had ≥7 lifetime infusions and as high as 46% in patients who have had ≥15 lifetime infusions. Similarly, patients with cystic fibrosis have a relatively high incidence of allergic reactions to IV antibiotics, particularly beta-lactams, that they receive periodically for intermittent "clean-outs" to maintain airway clearance. Drugs can also function as haptens that form immunogenic conjugates with host proteins. The conjugating hapten may be the parent compound, a nonenzymatically derived storage product, or a metabolite formed in the host. Recombinant biologics in some cases can also induce the formation of IgE against the proteins or against glycosylated structures that serve as immunogens. Outbreaks of anaphylaxis to the epidermal growth factor receptor (EGFR) antibody, cetuximab, have been reported in association with elevated titers of serum IgE to alpha-1,3-galactose (alpha-gal), an oligosaccharide found in nonprimate mammals. Cetuximab is derived from a mouse cell line expressing a transferase that tags the Fab′ portion of the cetuximab heavy chain with alpha-gal. Interestingly, patients with a history of multiple bites from *Amblyomma americanum* ticks commonly found in the Carolinas, Arkansas, and Tennessee are more likely to have anti-alpha-gal IgE as compared to patients living outside those states. A subset of individuals with anti-alpha-gal IgE can develop episodes of delayed-onset anaphylaxis about 3–6 h after consuming mammalian meat (beef, lamb, or pork) in a condition now known as alpha-gal syndrome (AGS). The mechanisms behind AGS are not yet well understood.

It is also important to remember that allergic reactions caused by the effector cells and mediators described previously can be triggered by a variety of mechanisms that may not require the presence of sensitizing IgE. Non-IgE-mediated mast cell activation secondary to certain drugs is clinically indistinguishable from classical IgE-mediated hypersensitivity reactions but can occur with first known exposure since there is no prior need for mast cell sensitization by IgE. MRGPRX2, a G protein–coupled receptor that is highly expressed in skin mast cells, has been shown in mouse models and in vitro studies using human cells to induce mast cell activation and mediator release secondary to neuromuscular blocking drugs (NMBDs), quinolones, and icatibant. These findings are clinically significant since NMBDs are needed for procedures done under general anesthesia while quinolones are a commonly used antibiotic family. Icatibant, a bradykinin-2 receptor antagonist administered by subcutaneous injection for the treatment of acute attacks of hereditary angioedema, is known to frequently result in local injection site reactions. Another example of non-IgE-mediated anaphylaxis is demonstrated with paclitaxel, a chemotherapy agent most commonly used in combination with carboplatin to treat ovarian cancer. It is derived from yew tree bark and needles that require polyethoxylated castor oil (Cremophor) to be solubilized into aqueous solution. Cremophor has been shown in vitro to activate the complement cascade, resulting in complement-dependent histamine release from mast cells and basophils. A version of paclitaxel that is solubilized by being bound to albumin nanoparticles, Abraxane, has a far lower rate of hypersensitivity, especially for patients who have had infusion reactions to Cremophor-solubilized paclitaxel. Reactions such as flushing, hives, angioedema, and/or anaphylaxis to radiocontrast, opiates, vancomycin, and nonsteroidal anti-inflammatory drugs (NSAIDs) are other examples of non-IgE-mediated hypersensitivity.

An example of non-IgE-mediated food anaphylaxis is scombroid poisoning, caused by the ingestion of a pharmacologically significant dose of histamine from contaminated fish. This is most commonly associated with fish that are rich in histidine, such as mahi-mahi, mackerel, and tuna. If improperly stored, surface contaminant bacteria expressing histidine decarboxylase convert histidine into histamine, which is not broken down by baking or broiling. Scombroid poisoning leads to symptoms including flushing, oropharyngeal pruritis/angioedema, nausea/vomiting, and lightheadedness. Symptoms are often self-limited but still treatable with antihistamines and, if needed, epinephrine. Because the source of histamine is exogenous, tryptase levels are not elevated.

A final example of non-IgE-mediated anaphylaxis occurred in clustered cases of anaphylaxis to IV unfractionated heparin in Europe and North America in 2008. Cases were attributable to specific lots of heparin found to be enriched in a contaminant, oversulfated chondroitin sulfate (OSCS). OSCS is able to activate the contact activation system, leading to the generation of bradykinin and the anaphylatoxins C3a and C5a, which led to the anaphylactic symptoms. Heparin is now routinely screened for this contaminant to prevent further outbreaks.

DIAGNOSIS

The diagnosis of an anaphylactic reaction depends primarily on a history revealing the onset of symptoms and signs within seconds to minutes after the putative trigger is encountered. An exception is delayed anaphylaxis to mammalian meats in patients with AGS. Table 364-1 lists other possibilities to consider on the differential for IgE-mediated anaphylaxis. Every attempt to identify the specific cause or causes should be made to minimize the risk of recurrent anaphylaxis. If a particular drug or food is suspected, skin or serum-specific IgE testing can be useful to confirm clinical suspicions. If a specific trigger cannot be identified by history or testing, a workup of underlying baseline atopic diatheses may be useful to identify risk factors that could play potential contributory roles alone or in concert. In the acute setting, laboratory biomarkers of mast cell degranulation may be useful to document the severity of an anaphylactic episode. The most obvious serum biomarker to assay, histamine, has an extremely short serum half-life with a measurable time-window that expires <1 h from the onset of anaphylaxis. A more practical and reliable biomarker is serum tryptase, which is released from mast cells, peaks 60–90 min after the onset of anaphylaxis, and can be measured as long as 5 h after the onset of anaphylaxis. It may be useful to follow up an elevated tryptase measurement in the acute setting with another measurement when the patient is clinically stable to establish a baseline reference since there is considerable variability of baseline serum tryptase (BST) within the general population. For example, ~6% of Western populations have hereditary alpha-tryptasemia (HAT), a mostly benign condition featuring elevated BST due to expression of an extra allele of TPSAB1, the gene that encodes alpha-tryptase. Some studies suggest that HAT may be a risk factor for severe anaphylaxis and is also overrepresented in patients with indolent mastocytosis. Another cause for temporarily elevated BST is high environmental allergen exposure, something that can occur, for example, in regions with distinct pollen seasons. A working group on mast cell disorders in 2010 came up with a formula to better interpret tryptase levels since (1) the range of "normal" BST is wide within the general population and (2) an acutely elevated tryptase for a given patient may still fall within normal parameters if their BST is low. The "20% + 2" rule stipulates that a tryptase level drawn in the setting of possible anaphylaxis that is 20% above the patient's baseline plus 2 ng/mL is diagnostic for an acute event involving a clinically significant degree of mast cell activation.

TABLE 364-1 Differential Diagnoses for IgE-Mediated Anaphylaxis

CONDITION	DISTINGUISHED BY
Mastocytosis	Elevated baseline tryptase, spindle-shaped mast cells (MCs) on bone marrow
Pheochromocytoma	Elevated urine metanephrines
Carcinoid syndrome	Elevated urine 5-hydroxyindoleacetic acid
Hereditary angioedema	Decreased C4 during attacks
Acquired angioedema	Decreased C1q
Systemic capillary leak syndrome	Severe hypotension on presentation with lack of response to first-line hypersensitivity medications (epinephrine, antihistamines)
Scombroid poisoning	Tryptase not elevated; negative skin test and oral challenge to fish
Drugs (opiates, neuromuscular blocking agents, vancomycin), computed tomography radiocontrast	Direct MC degranulation triggered through MRGPRX2 receptor or other as-yet-undetermined mechanism

TREATMENT

Early recognition of an anaphylactic reaction and appropriate intervention are critically important because severe, even fatal, complications can occur within minutes after symptoms first appear. The treatment of first choice is intramuscular administration of 0.3–0.5 mL of 1:1000 (1 mg/mL) epinephrine, with repeated doses at 5- to 20-min intervals as needed for a severe reaction. The failure to use epinephrine within the first 20 min of symptoms is a risk factor for poor clinical outcomes in various studies of anaphylaxis. IV fluids and vasopressor agents may be administered in the acute medical setting if intractable hypotension occurs. Epinephrine provides both α- and β-adrenergic effects, resulting in vasoconstriction, bronchial smooth-muscle relaxation, and attenuation of enhanced venular permeability. Beta blockers may attenuate this response; therefore, an alternative antihypertensive may be considered in patients at high risk of needing emergency epinephrine. Oxygen alone via a nasal catheter or with nebulized albuterol may be helpful; however, either endotracheal intubation or a tracheostomy is mandatory for oxygen delivery if progressive hypoxia develops. Ancillary agents such as antihistamines, glucocorticoids, and bronchodilators are also useful therapeutics to treat urticaria/angioedema and bronchospasm once the patient is hemodynamically stable.

PREVENTION

Avoidance The simplest and most straightforward approach to the long-term management of a patient with a history of anaphylaxis is strict avoidance of known anaphylactic triggers and education on acute management, specifically, instructing the patient on proper use and indications for use of self-administered epinephrine. Lifelong avoidance is not easy if the trigger is an occupational exposure, Hymenoptera sting, a common food (i.e., peanut), or a drug representing the sole or best therapeutic option for the patient. Special management options may exist for these patients.

Venom Immunotherapy Patients of any age who have had documented anaphylaxis from Hymenoptera sting should be formally evaluated and started on venom immunotherapy (VIT) if skin or serologic IgE testing confirms the history. Immunotherapy is a means of "tolerizing" patients to allergen by means of serial subcutaneous administration of escalating doses of extract containing relevant allergen until a target maintenance dose is achieved. Anaphylaxis can sometimes occur during the course of administering immunotherapy extracts, so formulating extracts and administering them is typically done under the care of a specialist familiar with this type of treatment. In the case of Hymenoptera allergy, patients receive VIT extracts containing actual Hymenoptera venom with a maintenance dose equivalent to 2–5 stings. The recommended duration of treatment is 3–5 years; however, some patients who have experienced severe respiratory or cardiovascular anaphylaxis are put on lifelong therapy. Patients with mastocytosis may also require such lifelong treatment.

Preventative Tolerance Induction to Peanuts IgE sensitization to foods occurs most frequently in infants and young children, especially those with atopic dermatitis, and is a risk factor for anaphylaxis (although detection of specific IgE through skin or serum testing has relatively poor predictive value). While most allergy to egg, milk, soy, and/or wheat resolves spontaneously during childhood, ~80% of children with peanut allergy remain sensitive for life. A sharp rise in the prevalence of peanut allergy was also observed in the late 1990s to early 2000s, especially in countries with Western diets where the average age of peanut introduction was age ≥3 years. Curiously, in cultures where peanut was introduced much earlier into children's diets, the prevalence of peanut allergy remained low. The landmark Learning Early About Peanut Allergy (LEAP) study demonstrated that early introduction of peanut protein to the diet of high-risk infants (4–11 months of age with atopic dermatitis and/or egg allergy) prevented the development of most (80% or more) peanut allergy compared with children who did not consume peanuts (avoidance group), even when IgE sensitization (based on positive skin test) had already developed at the time of study entry. While the induction of tolerance at an early age seems to be key to preventing clinical reactivity later in life, it is not yet clear if this principle holds true for other foods commonly associated with hypersensitivity reactions.

A relatively new treatment for patients who already suffer from severe peanut allergy is peanut oral immunotherapy (OIT) in which patients receive a titrating regimen of precisely dosed peanut protein in the form of an oral capsule. The goal of peanut OIT is to raise a patient's threshold tolerance for accidental peanut exposure before anaphylaxis occurs, and it is not a cure for peanut allergy. At the current time, peanut is the only food for which a U.S. Food and Drug Administration (FDA)-approved treatment exists, but research on developing OIT for other foods, such as milk and egg, remains ongoing.

Desensitization For patients who have experienced anaphylaxis from drug allergy and whose treatment regimen requires the administration of the offending drug, desensitization may be a short-term treatment option to prevent reactions. Desensitization elicits a temporary state of tolerance to the drug in sensitized, clinically reactive patients. While it has been a proven technique for penicillin-allergic patients for decades, desensitization has more recently been proven to be effective for certain chemotherapy agents, especially platin-based chemotherapy agents that can induce IgE-mediated sensitization with repeated exposures. The exact mechanisms underlying desensitization are not fully understood; however, temporary tolerance can be achieved through the serial administration of gradually escalating doses of drug, starting from extremely low doses, over the course of hours. As long as the patient continues to receive the drug in question at regular intervals based on drug half-life, a "desensitized" state can also be maintained until the drug is no longer needed. While drug desensitization certainly works for IgE-mediated reactions, it has been performed in cases of non-IgE-mediated anaphylaxis from Cremophor-solubilized paclitaxel as described earlier in this chapter. Drug desensitization has also been shown by multiple groups to work for non-IgE-mediated reactions from a variety of biologic agents, various chemotherapy drugs, and NSAIDs. Given the complexity and variety of possible drug reactions, the decision to desensitize, challenge, or avoid should be made in conjunction with an allergy specialist for complete evaluation and proper risk stratification of the different possible approaches to take.

FURTHER READING

Brennan PJ et al: Hypersensitivity reactions to mAbs: 105 desensitizations in 23 patients, from evaluation to treatment. J Allergy Clin Immunol 124:1259, 2009.

Castells MC et al: Hypersensitivity reactions to chemotherapy: Outcomes and safety of rapid desensitization in 413 cases. J Allergy Clin Immunol 122:574, 2008.

Chung CH et al: Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. N Engl J Med 358:1109, 2008.

Du Toit G et al: LEAP Study Team. Randomized trial of peanut consumption in infants at risk for peanut allergy. N Engl J Med 373:803, 2015.

Du Toit G et al: Immune Tolerance Network LEAP-On Study Team. Effect of avoidance on peanut allergy after early peanut consumption. N Engl J Med 374:1435, 2016.

Lieberman P et al: Anaphylaxis—A practice parameter update 2015. Ann Allergy Asthma Immunol 115:341, 2015.

McNeil BD et al: Identification of a mast cell-specific receptor crucial for pseudoallergic drug reactions. Nature 519:237, 2015.

Valent P et al: Why the 20% + 2 tryptase formula is a diagnostic gold standard for severe systemic mast cell activation and mast cell activation syndrome. Int Arch Allergy Immunol 180:44, 2019.

365

Desensitization

Mariana Castells

DRUG ALLERGY AND THE NEED FOR DRUG DESENSITIZATION

Drug allergy is a rising problem, paralleling the worldwide increased use of medications, the plethora of new targeted therapies, and the greater longevity of patients. Allergic drug reactions decrease patients' quality of life, limit treatment options, and can lead to vaccine hesitancy, as seen for the COVID-19 mRNA vaccines. The first cases of reported drug-induced anaphylaxis occurred shortly after the discovery and therapeutic applications of penicillin in 1945, and today, 10–20% of the world's population engaged in health care claims to have a penicillin allergy. Truly allergic patients cannot use penicillin and are negatively impacted by the decreased efficacy, increased costs, and limited availability of second-line antibiotics. Similarly, the need to avoid the offending medications in allergic patients with malignancies and chronic inflammatory diseases can increase morbidity and impact life expectancy.

To address the needs of drug-allergic patients, a novel modality of drug delivery aimed at curtailing allergic symptoms has culminated in drug desensitization (DD), a treatment option that takes advantage of immune inhibitory mechanisms. Rapid multistep protocols that deliver sequential small doses of drug until the target therapeutic dose is reached in a few hours have shown outstanding safety, efficacy, and wide applicability in thousands of patients with infections, chronic inflammatory diseases, malignancies, and other conditions.

DRUG ALLERGY DEFINITIONS AND RISK FACTORS

Drug allergy and drug hypersensitivity are interchangeable terms that refer to acute and delayed symptoms occurring after exposure to medications. While there is a need to address drug-allergic patients, those with unconfirmed allergy labels should be delabeled, and single-dose oral challenges have been shown to be safe in children and adults at low risk for penicillin allergy. Drug allergy is multifactorial and associated with female gender, specific human leukocyte antigen (HLA) haplotypes, atopy, polypharmacy, older age, chronic diseases, microbiome changes, and drug-dependent factors such as the route of administration, repeated and high doses, excipients, and glycosylation.

CLASSIFYING HYPERSENSITIVITY REACTIONS

The modern classification of reactions has incorporated timing, severity, biomarkers, and new presentations to the classical Gell and Coombs definitions. Acute or immediate reactions occur during or within 1–6 h of drug exposure, while delayed reactions occur 6 h to several days or weeks after drug exposure. Reactions can be mild (affecting one organ; grade 1), moderate (affecting two or more organs; grade 2), or severe (associated with changes in vital signs; grade 3).

PHENOTYPES, ENDOTYPES, AND BIOMARKERS

The symptom presentation (phenotype) depends on the mechanisms of the reactions (endotype). Acute reactions include infusion reactions, type I IgE-dependent and -independent reactions, cytokine release reactions (CRRs), and mixed reactions (type I and CRR symptoms). Delayed reactions include type II antibody-mediated cytotoxicity; type III immune complex–mediated reactions; and type IV reactions, which include benign reactions, typically skin limited, and severe cutaneous reactions with systemic symptoms (SCARS) such as Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN), drug reaction with eosinophilia and systemic symptoms (DRESS), and acute generalized eosinophilic pustulosis (AGEP) **(Fig. 365-1)**. Symptoms are drug specific, and acute musculoskeletal pain is common in reactions to taxanes, oxaliplatin, doxorubicin, and rituximab.

Type I Reactions Type I reactions result from the activation of mast cells and basophils through IgE-dependent and -independent mechanisms. While IgE sensitization requires repeated exposures and is typical for antibiotics, chemotherapy, and some biologicals, IgE-independent reactions typically occur at first exposure through either complement activation, inhibition of COX-1, or activation of the MRGPRX2 receptor by quinolones, vancomycin, icatibant, and general anesthetics with THIQ binding motif and by basic compounds. Symptoms of type I reactions include itching, flushing, hives, angioedema, dyspnea, wheezing, oxygen desaturation, throat tightening, nausea, vomiting, and hypotension with cardiovascular collapse and are associated with elevated serum tryptase, a specific biomarker of anaphylaxis released from mast cells and basophils granules. Tryptase levels above the normal range (11.4 ng/mL) or above a patient's baseline (× 1.6) obtained within 30 min to 4 h of initial symptoms are diagnostic of type I and mixed reactions. Patients with baseline levels of 7.5 ng/mL or higher may have duplications of the *TPSAB1* tryptase gene located on chromosome 16, known as hereditary alpha tryptasemia (HαT), a familial autosomal dominant trait present in 4–6% of Caucasian populations, which modulates anaphylaxis severity. Other mast cell–derived mediators include urine *N*-methylhistamine, prostaglandins, and leukotrienes.

Cytokine Release Reactions CRRs can occur at first exposure or after several exposures and are thought to be due to activation of T cells and other immune cells. Patients present with fever; chills; back, chest, or pelvic pain; headache; oxygen desaturation; and hyper- or hypotension; these reactions are associated with a transient serum elevation of interleukin (IL) 6.

Mixed Reactions Mixed reactions present with symptoms of type I and CRR reactions and are associated with tryptase and IL-6 elevations (Fig. 365-1).

Delayed Type IV Reactions Delayed type IV reactions amenable to DD present as benign maculopapular rashes without associated systemic symptoms, whereas SCARS, type II, type III, vasculitis, and single-organ toxic reactions are contraindications for DD because minute amounts of medication can trigger severe reactions.

DIAGNOSTIC TESTING

Skin testing (ST) elicits a local wheal-and-flare reaction upon epicutaneous or intradermal injection of drugs inducing IgE-mediated reactions. Its use is limited by the availability of drug components including excipients, skin toxicity, costs, and the time elapsed since the patient's initial reaction (the longer the time, the lower the likelihood that an ST will be positive) and the severity of the reaction (anaphylactic reactions can deplete mediators inducing false-negative ST). Positive and negative predictive values are drug specific, and anaphylaxis to penicillin or carboplatin has not been reported in ST-negative patients.

The basophil activation test (BAT) provides evidence of IgE sensitization by challenging the patient's basophils in vitro with the culprit drug, eliminating the need for direct ST, which has a small risk for anaphylaxis. Serum-specific IgE antibodies to penicillin and platins are found in allergic patients with low sensitivity. Skin patch testing is helpful to evaluate β-lactams, anticonvulsants, corticosteroids, and other drugs that cause delayed type IV reactions with high specific but low sensitivity. Lymphocyte transformation tests (LTTs) measure the activation and/or cytokine release of lymphocytes exposed to culprit drugs and can be used to identify drugs inducing SCARS. Certain HLA haplotypes place patients at risk for SCARS when exposed to abacavir, anticonvulsants, allopurinol, vancomycin, and other drugs, and genotyping is available.

DRUG DESENSITIZATION

DEFINITIONS AND MECHANISMS

DD is a temporary immunotherapy modality, delivered through multistep protocols to safely and timely reintroduce a drug that has induced an acute or delayed allergic reaction. IgE-mediated desensitization takes advantage of inhibitory mast cell/basophil pathways, which are activated by low doses of drug antigens. Rapid delivery of incremental doses recruit phosphatases to IgE receptors, blocking signal transduction and the release of mediators when reaching the target dose and protecting against anaphylaxis. Desensitized patients have transient conversion from positive to negative ST.

PROTOCOLS

Based on inhibitory pathways in mast cell models, human protocols have been generated for type I phenotypes with low starting doses and multiple steps that progress by doubling the dose administered in the previous step at constant time intervals. The first standardized and most used protocol contains three bags of 1/100, 1/10, and undiluted concentrations of the target dose and 12 doubling steps delivered every 15 min so that the target dose is reached in 5.7 h **(Fig. 365-2)**, which has been adapted for IV, PO, SC, IM, intraperitoneal, intraocular, and intrathecal use. Protocols with four bags (starting with a 1/1000 dilution bag) are used for patients with severe initial reactions and associated comorbidities. The recent use of one-bag protocols for type I reactions has produced mixed outcomes, with an increased use of epinephrine during breakthrough reactions (BTRs) and inability to complete treatments in highly sensitized patients.

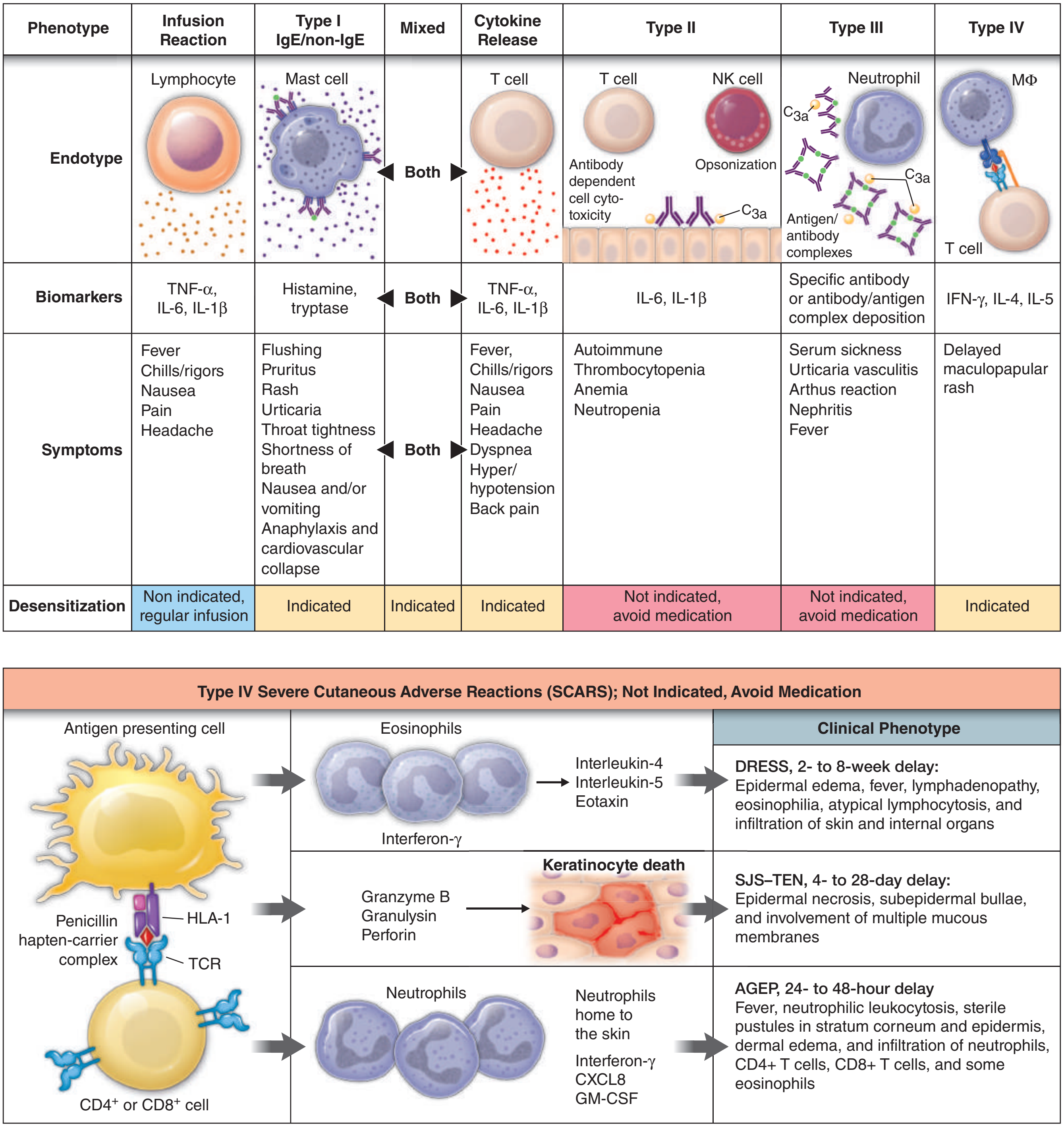

Phenotype	Infusion Reaction	Type I IgE/non-IgE	Mixed	Cytokine Release	Type II	Type III	Type IV
Endotype	Lymphocyte	Mast cell	◀ Both ▶	T cell	T cell NK cell Antibody dependent cell cyto-toxicity Opsonization C3a	Neutrophil C3a C3a Antigen/ antibody complexes	MΦ T cell
Biomarkers	TNF-α, IL-6, IL-1β	Histamine, tryptase	◀ Both ▶	TNF-α, IL-6, IL-1β	IL-6, IL-1β	Specific antibody or antibody/antigen complex deposition	IFN-γ, IL-4, IL-5
Symptoms	Fever Chills/rigors Nausea Pain Headache	Flushing Pruritus Rash Urticaria Throat tightness Shortness of breath Nausea and/or vomiting Anaphylaxis and cardiovascular collapse	◀ Both ▶	Fever, Chills/rigors Nausea Pain Headache Dyspnea Hyper/ hypotension Back pain	Autoimmune Thrombocytopenia Anemia Neutropenia	Serum sickness Urticaria vasculitis Arthus reaction Nephritis Fever	Delayed maculopapular rash
Desensitization	Non indicated, regular infusion	Indicated	Indicated	Indicated	Not indicated, avoid medication	Not indicated, avoid medication	Indicated

Type IV Severe Cutaneous Adverse Reactions (SCARS); Not Indicated, Avoid Medication

Clinical Phenotype

DRESS, 2- to 8-week delay: Epidermal edema, fever, lymphadenopathy, eosinophilia, atypical lymphocytosis, and infiltration of skin and internal organs

SJS–TEN, 4- to 28-day delay: Epidermal necrosis, subepidermal bullae, and involvement of multiple mucous membranes

AGEP, 24- to 48-hour delay Fever, neutrophilic leukocytosis, sterile pustules in stratum corneum and epidermis, dermal edema, and infiltration of neutrophils, CD4+ T cells, CD8+ T cells, and some eosinophils

FIGURE 365-1 Drug allergy phenotypes, endotypes, and biomarkers and indications for desensitization. AGEP, acute generalized eosinophilic pustulosis; DRESS, drug reaction with eosinophilia and systemic symptoms; IL, interleukin; SJS-TEN, Stevens-Johnson syndrome–toxic epidermal necrolysis; TNF, tumor necrosis factor. *(From The New England Journal of Medicine, Penicillin Allergy, M Castells, DA Khan et al: 381: 2338. Copyright @2019 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

The mechanisms of desensitization for other phenotypes are poorly understood, and CRRs have been successfully treated with multistep one-bag protocols, whereas multiple doses over several days have been used for delayed reactions (see specific drugs). Premedications are tailored to the initial symptoms, and steroids do not protect against anaphylaxis. Omalizumab, a monoclonal anti-IgE antibody, has been used as an adjuvant in highly sensitized patients with type I severe IgE-mediated reactions, including life-threatening anaphylaxis. Desensitization is a temporary phenomenon that does not lead to sustained tolerance and must be repeated with each exposure or if there has been a pause between doses equal to two or more half-lives of the drug. However, the desensitized state can be maintained by continued drug exposure as with antibiotics and aspirin/COX-1 inhibitors. Some patients reactive to taxanes and certain biologicals with uneventful DD protocols may eventually return to regular infusions.

INDICATIONS, BREAKTHROUGH REACTIONS, AND OUTCOMES

INDICATIONS (FIG. 365-2)

Qualifications for DD depend on the need of the medication as first-line therapy, the initial reaction phenotype, and its severity. Infusion reactions, mild CRRs, and mild delayed reactions can be addressed by symptoms targeting adjuvant medications. DD is indicated for type I IgE-dependent and IgE-independent, CRR, mixed, and nonsevere delayed type IV phenotypes. Type I and mixed reactions

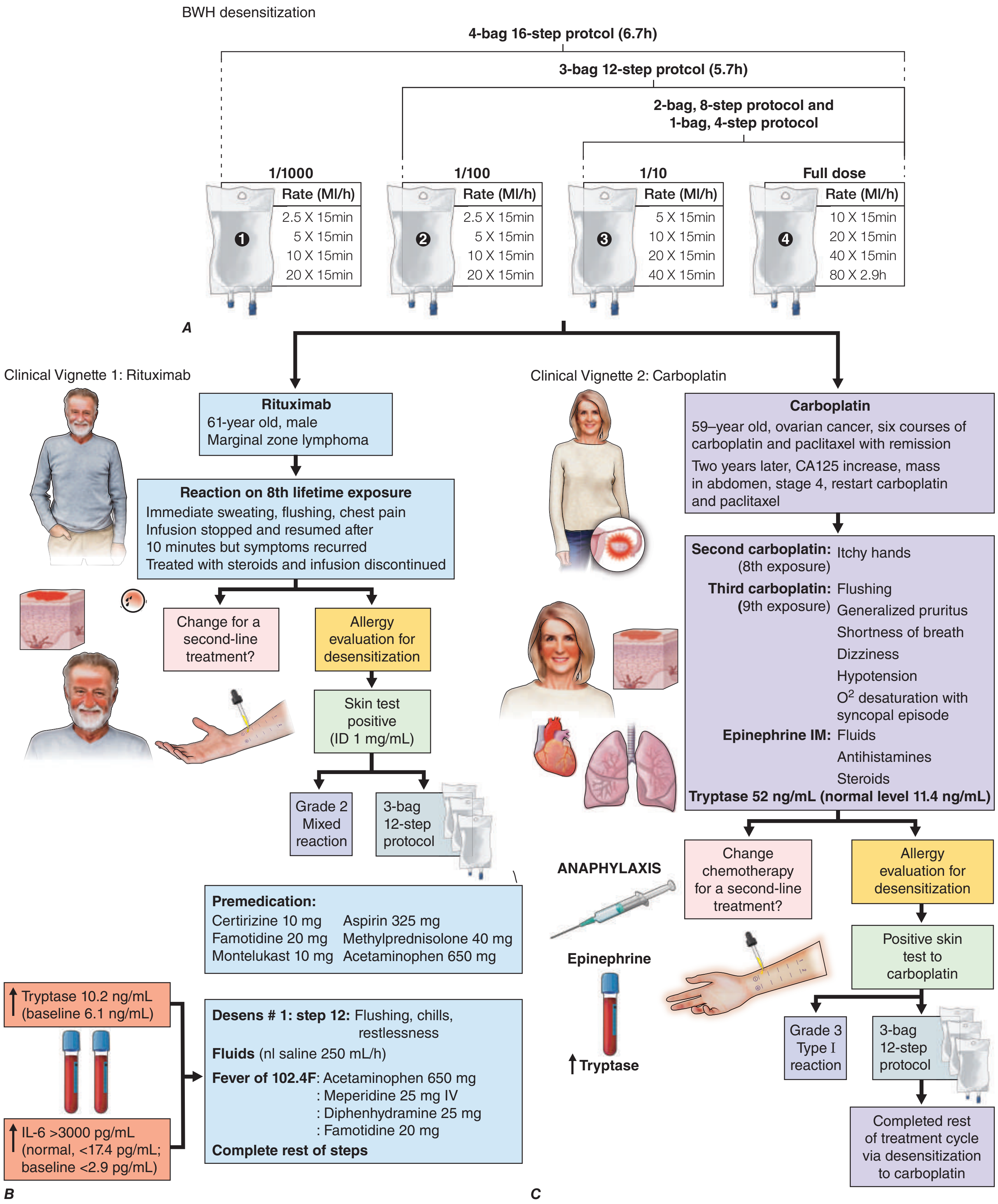
BWH desensitization
4-bag 16-step protcol (6.7h)
3-bag 12-step protcol (5.7h)
2-bag, 8-step protocol and
1-bag, 4-step protocol
1/1000
Rate (Ml/h)
2.5 X 15min
5 X 15min
10 X 15min
20 X 15min
1
1/100
Rate (Ml/h)
2.5 X 15min
5 X 15min
10 X 15min
20 X 15min
2
1/10
Rate (Ml/h)
5 X 15min
10 X 15min
20 X 15min
40 X 15min
3
Full dose
Rate (Ml/h)
10 X 15min
20 X 15min
40 X 15min
80 X 2.9h
4
A
Clinical Vignette 1: Rituximab
Rituximab
61-year old, male
Marginal zone lymphoma
Reaction on 8th lifetime exposure
Immediate sweating, flushing, chest pain
Infusion stopped and resumed after
10 minutes but symptoms recurred
Treated with steroids and infusion discontinued
Change for a second-line treatment?
Allergy evaluation for desensitization
Skin test positive (ID 1 mg/mL)
Grade 2 Mixed reaction
3-bag 12-step protocol
Premedication:
Certirizine 10 mg
Famotidine 20 mg
Montelukast 10 mg
Aspirin 325 mg
Methylprednisolone 40 mg
Acetaminophen 650 mg
↑ Tryptase 10.2 ng/mL (baseline 6.1 ng/mL)
↑ IL-6 >3000 pg/mL (normal, <17.4 pg/mL; baseline <2.9 pg/mL)
Desens # 1: step 12: Flushing, chills, restlessness
Fluids (nl saline 250 mL/h)
Fever of 102.4F: Acetaminophen 650 mg
: Meperidine 25 mg IV
: Diphenhydramine 25 mg
: Famotidine 20 mg
Complete rest of steps
B
Clinical Vignette 2: Carboplatin
Carboplatin
59–year old, ovarian cancer, six courses of carboplatin and paclitaxel with remission
Two years later, CA125 increase, mass in abdomen, stage 4, restart carboplatin and paclitaxel
Second carboplatin: (8th exposure) Itchy hands
Third carboplatin: (9th exposure) Flushing
Generalized pruritus
Shortness of breath
Dizziness
Hypotension
O2 desaturation with syncopal episode
Epinephrine IM: Fluids
Antihistamines
Steroids
Tryptase 52 ng/mL (normal level 11.4 ng/mL)
ANAPHYLAXIS
Epinephrine
↑ Tryptase
Change chemotherapy for a second-line treatment?
Allergy evaluation for desensitization
Positive skin test to carboplatin
Grade 3 Type I reaction
3-bag 12-step protocol
Completed rest of treatment cycle via desensitization to carboplatin
C

FIGURE 365-2 Desensitization: protocols and applications. ***A.*** Multibag, multistep intravenous desensitization protocols. ***B.*** Rituximab allergic reaction with mixed phenotype, positive skin test, and elevated tryptase and interleukin (IL) 6 biomarkers and successful treatment with a three-bag 12-step protocol. ***C.*** Carboplatin allergic reaction with type I phenotype, positive skin test, and elevated tryptase biomarker and successful treatment with a three-bag 12-step protocol. *(Reproduced with permission from L Campos et al: Curr Treat Options Allergy 6:519, 2019.)*

in patients with multiple exposures to antibiotics, chemotherapeutic drugs, and biologicals indicate IgE sensitization. ST-positive patients require DD for reexposure since the risk for anaphylaxis is high. In patients with mild reactions and negative biomarkers, controlled drug challenges are indicated to assess tolerance. Large clinical series support outpatient settings for the majority of DD, which decreases the starting treatment time and costs and improves patients' experiences.

DD BREAKTHROUGH REACTIONS AND OUTCOMES

BTRs are typically mild with symptoms similar to the initial reaction. They occur in 10–30% of protocols and require symptom-specific management including epinephrine for severe reactions. BTRs do not preclude the completion of the DD protocols in 99% of cases, and no deaths due to DD have been reported. Comorbidities such as pregnancy, cystic fibrosis, decreased lung function, cardiac diseases, advanced cancer, beta blocker and/or angiotensin-converting enzyme inhibitor use, and prior severe reactions increase the risk during BTRs (Fig. 365-3). The administration of antibiotics, monoclonal antibodies, and chemotherapy drugs through DD has shown equal efficacy as for standard administration, with the expected clearance of infections, decreased inflammation, and similar cancer responses. Kounis syndrome and takotsubo cardiomyopathy resulting from drug-induced anaphylaxis are contraindications for DD (Fig. 365-3).

APPROACH TO SPECIFIC DRUGS

ANTIBIOTICS

Antibiotics can induce all the reactions in Fig. 365-1, but the most common are type I IgE-dependent and type IV non-SCARS reactions, both of which can be addressed by DD.

β-Lactams The first description of penicillin desensitization was in a World War II allergic soldier who presented with wheezing and hypotension after penicillin injection and received oral incremental doses over 30 days until he was able to tolerate a treatment course. The initial success led to shorter protocols to treat patients with bacterial endocarditis and pregnant women with syphilis who had a history of penicillin-induced anaphylaxis and positive ST. Safe rapid PO and IV desensitization protocols are currently available for all antibiotic classes and can treat high-risk patients, including cystic fibrosis patients with low pulmonary function.

Sulfonamides Delayed type IV rashes to trimethoprim-sulfamethoxazole occurred in up to 40% of patients with <100 total CD4 T cells/μL during the early 1990s HIV/AIDS epidemic, and DD included incremental doses over multiple days. Reactions have decreased since retroviral treatments, and shorter protocols have evolved. One-day PO and IV multidose protocols have shown consistent safety and efficacy for delayed reactions in non-HIV patients.

CHEMOTHERAPY

Taxanes Over 40% of patients exposed to paclitaxel and docetaxel presented with acute reactions, which have been reduced to 1% by antihistamine and corticosteroid premedication. Moderate to severe reactions attributed to lipid excipients (Cremophor and polysorbate 80) and to IgE sensitization through cross-reactive foods and environmental allergens occur typically at first or second exposure with type I, mixed, and delayed phenotypes. Patients with delayed rashes can convert to acute reactions including anaphylaxis upon subsequent exposures. The first DD to paclitaxel was reported in 2002, and an early study of 940 DDs to paclitaxel and docetaxel in 138 patients described BTRs in 20% of cases, and 22% of desensitized patients returning to standard infusions. A review of 25 studies with 976 patients and 2396 DDs to paclitaxel and docetaxel, completed in 95–100% of cases, the majority with three-bag 12-step protocols, showed 32% paclitaxel and 20% docetaxel mild BTRs and no deaths.

Platinums Platinum drug (carboplatin, cisplatin, and oxaliplatin) IgE sensitization requires multiple exposures (typically six or more;

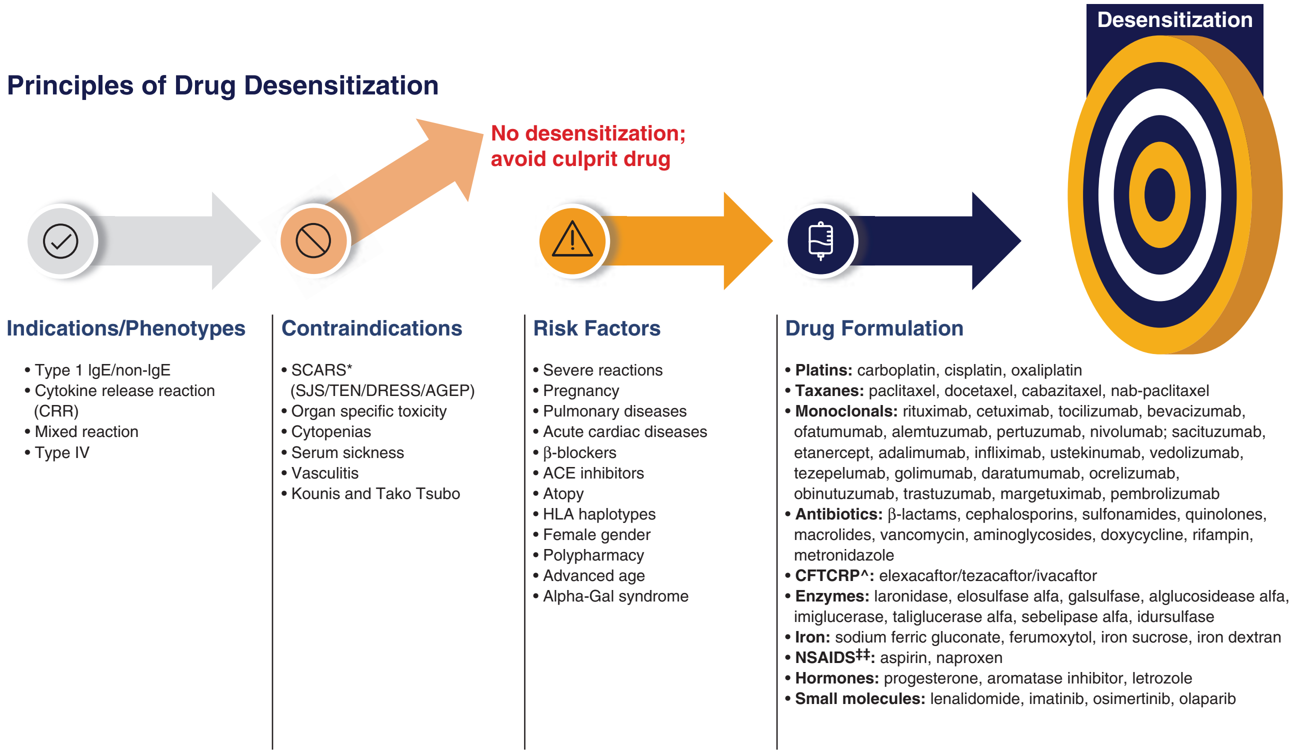
Principles of Drug Desensitization
Desensitization
No desensitization; avoid culprit drug

Indications/Phenotypes

- Type 1 IgE/non-IgE
- Cytokine release reaction (CRR)
- Mixed reaction
- Type IV

Contraindications

- SCARS* (SJS/TEN/DRESS/AGEP)
- Organ specific toxicity
- Cytopenias
- Serum sickness
- Vasculitis
- Kounis and Tako Tsubo

Risk Factors

- Severe reactions
- Pregnancy
- Pulmonary diseases
- Acute cardiac diseases
- β-blockers
- ACE inhibitors
- Atopy
- HLA haplotypes
- Female gender
- Polypharmacy
- Advanced age
- Alpha-Gal syndrome

Drug Formulation

- **Platins:** carboplatin, cisplatin, oxaliplatin
- **Taxanes:** paclitaxel, docetaxel, cabazitaxel, nab-paclitaxel
- **Monoclonals:** rituximab, cetuximab, tocilizumab, bevacizumab, ofatumumab, alemtuzumab, pertuzumab, nivolumab; sacituzumab, etanercept, adalimumab, infliximab, ustekinumab, vedolizumab, tezepelumab, golimumab, daratumumab, ocrelizumab, obinutuzumab, trastuzumab, margetuximab, pembrolizumab
- **Antibiotics:** β-lactams, cephalosporins, sulfonamides, quinolones, macrolides, vancomycin, aminoglycosides, doxycycline, rifampin, metronidazole
- **CFTCRP^:** elexacaftor/tezacaftor/ivacaftor
- **Enzymes:** laronidase, elosulfase alfa, galsulfase, alglucosidease alfa, imiglucerase, taliglucerase alfa, sebelipase alfa, idursulfase
- **Iron:** sodium ferric gluconate, ferumoxytol, iron sucrose, iron dextran
- **NSAIDS‡‡:** aspirin, naproxen
- **Hormones:** progesterone, aromatase inhibitor, letrozole
- **Small molecules:** lenalidomide, imatinib, osimertinib, olaparib

+Severe cutaneous adverse reaction
‡‡Non-steroidal anti-inflammatory drugs
^Cystic fibrosis transmembrane conductance regulator potentiator

FIGURE 365-3 Principles of drug desensitization: Indications, contraindications, risk factors, and drugs with successful desensitizations. ACE, angiotensin-converting enzyme; AGEP, acute generalized eosinophilic pustulosis; DRESS, drug reaction with eosinophilia and systemic symptoms; HLA, human leukocyte antigen; SJS-TEN, Stevens-Johnson syndrome–toxic epidermal necrolysis.

less for *BRCA* mutation carriers), and sensitized patients typically have type I symptoms with pruritus, flushing, urticaria, throat tightening, and hypotension with elevated tryptase (Fig. 365-2). ST is positive in >85% of patients with recent reactions. In a prospective study of 126 women who received six or more carboplatin treatments, repeat ST before each exposure identified reactors, with seven anaphylactic reactions in seven patients who agreed to be reexposed to carboplatin despite positive ST. Oxaliplatin reaction phenotypes include CRR and mixed reactions with conversion from type I to CRR and mixed phenotypes during DD. Oxaliplatin-induced immune-mediated thrombocytopenia (OIIT) with bleeding is a rare complication that precludes further DD.

Carboplatin DD started in early 2000 and a large study in 2016 with 370 patients (212 carboplatin-allergic and 13 cisplatin-allergic patients), who underwent 2177 DDs, mostly in the outpatient setting with three- or four-bag protocols, showed 20% mild BTRs, nonsuperior costs compared to regular infusions, and a life expectancy advantage at 5 years in ovarian cancer patients compared to nonallergic patients. A study of 272 patients allergic to platins, taxanes, and monoclonals who underwent 1471 DDs with three-bag 12-step protocols reported similar safety. Oxaliplatin DD poses challenges due to the heterogeneity of phenotypes and phenotype switching. A study of 48 patients with 200 DDs who underwent three-bag 12-step protocols showed mild to moderate BTRs in 18 patients, 1 patient with OIIT, and 2 patients with hemolytic anemia. A study of patients with type I, CRR, and mixed reactions who underwent 273 DDs indicated a switch from type I to CRR and mixed reactions during DD, with elevated IL-6, and two patients developed OIIT. The survival of platinum-desensitized patients is noninferior to nonallergic patients.

Doxorubicin and Other Chemotherapy Drugs Doxorubicin reactions are rare, occurring at first exposure, and a study of 30 patients and 128 DDs, mostly with three-bag 12-step protocols, showed 16% mild BTRs, and one patient required epinephrine. DDs for etoposide, cyclophosphamide, methotrexate, gemcitabine, and other chemotherapy drugs have shown similar safety.

MONOCLONALS/BIOLOGICALS

Infusion reactions of the type I IgE/non-IgE mediated, CRR, mixed, and delayed-reaction phenotypes have been reported for monoclonal antibodies (MoAbs). Hypersensitivity reactions can occur on first exposure or after multiple exposures (Fig. 365-2). Reactions are more frequent with chimeric MoAbs but also occur with humanized and fully humanized MoAbs. Skin testing is limited by cost. Patients with cetuximab-induced anaphylaxis react to galactose-alpha-1-3-galactose, a nonhuman carbohydrate in the Fab chain of cetuximab, due to preexisting IgE acquired through tick bite sensitization. DDs have been reported for most IV and SC available MoAbs (Fig. 365-3). The first series of MoAb desensitizations included 23 patients with moderate to severe type I and mixed reactions to trastuzumab, infliximab, and rituximab who underwent 105 DDs with three bags and 12 steps and had 29% mild BTRs. A study of 104 patients who reacted to 16 MoAbs with type I, CRR, mixed, and delayed reactions and received 526 DDs, mostly with three bags and 12 steps in the outpatient setting, presented with 23% mild BTRs. Tryptase was elevated in type I and IL-6 in CRR BTRs.

SMALL MOLECULES

Epidermal growth factor receptor and tyrosine kinase inhibitors such as imatinib, osimertinib, and olaparib can induce type I and type IV reaction phenotypes, and successful DDs with multiple oral doubling doses have been reported, allowing for continued long-term maintenance.

NONSTEROIDAL ANTINFLAMMATORY DRUGS (NSAIDS)

Aspirin (ASA) and other NSAIDs can induce type I reactions with urticaria and anaphylaxis thought to be IgE-mediated and, in patients with asthma and dysregulated lipid metabolism, a syndrome termed aspirin-exacerbated respiratory disease (AERD) with bronchospasm. Patients with AERD have universal intolerance to COX-1 inhibitors but tolerate COX-2 inhibitors. The American Academy of Allergy, Asthma, and Immunology Work Group Report in 2020 reviewed 14 series of ASA DDs with intranasal ketorolac or oral aspirin in >900 AERD patients. More than 70% of patients reported improvement of asthma, lack of polyp recurrence, recovery of sense of smell, and less steroid usage, with cross-desensitization and tolerance to all COX-1 inhibitors while maintained on daily ASA. For type I reactions in cardiac and neurologic patients, DD with rapid doubling doses has been proven safe and effective.

OTHER DRUGS

Symptoms of progestogen hypersensitivity (PH) are cyclical and include type I and delayed phenotypes. PH can lead to miscarriage, and oral and vaginal DD protocols have resulted in viable pregnancies. Reactions to iron formulations include type I, CRR, mixed, and delayed phenotypes that are responsive to DD. Allopurinol-induced delayed rashes respond to PO and IV DD protocols. Corticosteroids, glatiramer acetate, vaccines, and other medication DDs have been reported with doubling multistep PO, IV, and SC protocols.

IMPACT

Discontinuing drugs inducing allergic symptoms affects patients for whom second-line therapies may not be available, are more expensive, or do not provide similar therapeutic benefits. DD protocols allow for the safe reintroduction of first-line therapies, and their use should be extended to all patients in need.

FURTHER READING

Adnan A, Acharya S: Multistep IgE mast cell desensitization is a dose- and time-dependent process partially regulated by SHIP-1. J Immunol 210:709, 2023.

Castells M et al: Penicillin allergy. N Engl J Med 381:2338, 2019.

Dhopeshwarkar N: Drug-induced anaphylaxis documented in electronic health records. J Allergy Clin Immunol Pract 7:103, 2019.

Isabwe GAC: Hypersensitivity reactions to therapeutic monoclonal antibodies: Phenotypes and endotypes. J Allergy Clin Immunol 142:159, 2018.

Sloane D: Safety, costs, and efficacy of rapid drug desensitizations to chemotherapy and monoclonal antibodies. J Allergy Clin Immunol Pract 4:497, 2016.

366

Mastocytosis

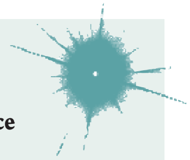

Matthew P. Giannetti, Joshua A. Boyce

DEFINITION AND EPIDEMIOLOGY

Mastocytosis is defined by accumulation of clonally expanded mast cells in tissues such as skin, bone marrow, liver, spleen, and gut. Diagnostically, mast cell expansion is most readily identified in skin and/or bone marrow. Mastocytosis occurs at any age, although it is most commonly diagnosed in infancy and young adulthood. The prevalence of mastocytosis is estimated at ~1 in 10,000 people. Most forms of the disease are characterized by somatic gain-of-function mutations in the stem cell factor receptor (*KIT*) gene, and >90% of patients carry the *KIT* D816V mutation. Familial occurrence is rare, and atopy is not increased compared with the general population.

CLASSIFICATION AND PATHOPHYSIOLOGY

A consensus classification for mastocytosis recognizes cutaneous mastocytosis with variants, systemic mastocytosis with five variants, and mast cell sarcoma (Table 366-1).

Cutaneous mastocytosis is the most common diagnosis in children and indicates disease limited to skin with absence of pathologic infiltrates in internal organs. It is usually diagnosed within the first year

TABLE 366-1 Classification of Mastocytosis

- Cutaneous mastocytosis (CM)
  - Maculopapular cutaneous mastocytosis (MPCM)
  - Solitary mastocytoma of skin
  - Diffuse cutaneous mastocytosis
- Systemic mastocytosis
  - Indolent systemic mastocytosis (ISM)
  - Smoldering systemic mastocytosis
  - Systemic mastocytosis with associated clonal hematologic non–mast cell lineage disease (SM-AHNMD)
  - Aggressive systemic mastocytosis (ASM)
  - Mast cell leukemia (MCL)
- Mast cell sarcoma (MCS)

Source: Modified from DA Arber et al: International consensus classification of myeloid neoplasms and acute leukemias: integrating morphologic, clinical, and genomic data. Blood 140:1200, 2022.

of life with demonstration of fixed, maculopapular, polymorphic, and hyperpigmented lesions (known as maculopapular cutaneous mastocytosis [MPCM]), solitary mastocytoma(s), or diffuse cutaneous mastocytosis. Although mast cell accumulation is limited to the skin, children often have systemic symptoms. Systemic mastocytosis (SM) refers to involvement of an extracutaneous site (most often bone marrow). There are five distinct variants of SM. *Indolent systemic mastocytosis* (ISM) accounts for >70% of adult patients. ISM is diagnosed when there is no evidence of organ dysfunction due to mast cell infiltration, an associated hematologic disorder, or mast cell leukemia. ISM is associated with a normal life expectancy. *Smoldering systemic mastocytosis* (SSM) is characterized by high mast cell burden including bone marrow infiltration of >30% and a baseline serum tryptase >200 ng/mL (B findings), but absence of *systemic mastocytosis associated with clonal hematologic non–mast cell lineage disease* (SM-AHNMD) or *aggressive systemic mastocytosis* (ASM) **(Table 366-2)**. In SM-AHNMD, the prognosis is determined by the nature of the associated disorder, which can range from dysmyelopoiesis to leukemia. In ASM, mast cell infiltration may occur in multiple organs such as liver, spleen, gut, bone, and bone marrow resulting in one or more C findings and a poor prognosis (Table 366-2). *Mast cell leukemia* (MCL) is the rarest form of SM and is invariably fatal at present; the peripheral blood contains circulating, metachromatically staining, atypical mast cells. An aleukemic form of MCL is recognized without circulating mast cells when the percentage of high-grade immature mast cells in bone marrow smears exceeds 20% in a nonspicular area. Mast cell sarcoma is a rare solid mast cell tumor with malignant invasive features.

TABLE 366-2 B and C Findings for Diagnosis of SSM and ASM

B Findings (2 or more in the absence of any C findings are required for a diagnosis of SSM):

1. High MC burden: MC infiltration in bone marrow of >30%, basal serum tryptase level >200 ng/mL, and/or *KIT* D816V >10% VAF in bone marrow
2. Hypercellular bone marrow with signs of dysmyelopoiesis but without cytopenias meeting C criteria or WHO criteria for an MDS or MPN
3. Palpable hepatomegaly, palpable splenomegaly, or lymphadenopathy (on CT or ultrasound: >2 cm) without impaired liver function or hypersplenism

C Findings (1 or more required for a diagnosis of ASM). C findings should be reasonably attributable to high tissue MC infiltration:

1. Cytopenia(s): ANC <1000/μL or Hb <10 g/dL or PLT <100,000/μL
2. Hepatomegaly with ascites and impaired liver function
3. Palpable splenomegaly with associated hypersplenism
4. Malabsorption with hypoalbuminemia and weight loss
5. Skeletal lesions: large area(s) of osteolysis with pathologic fractures (presence of osteoporosis alone without osteolytic lesions does not satisfy this criterion)

Abbreviations: ANC, absolute neutrophil count; ASM, aggressive systemic mastocytosis; CT, computed tomography; Hb, hemoglobin; MC, mast cells; MDS, myelodysplastic syndromes; MPN, myeloproliferative disorders; PLT, platelets; SSM, smoldering systemic mastocytosis; VAF, variant allele frequency; WHO, World Health Organization.

Somatic activating mutations in the *KIT* gene are characteristic of mastocytosis. *KIT* D816V is most common, although other mutations have been reported. *KIT* mutations are found in mast cells and sometimes in other cell lineages in patients with mastocytosis. *KIT* mutations are observed in patients with all forms of SM but are also present in some children with cutaneous mastocytosis in lesional skin. Additional mutations in genes such as *TET2, SRSF2, ASLX1*, and *RUNX1* known to be associated with other hematologic neoplastic disorders can be detected in patients, usually with advanced variants of SM. The life expectancy for patients with cutaneous mastocytosis and for most patients with ISM is normal, whereas that for patients with SM-AHNMD is determined by the non–mast cell component. ASM and MCL have a poor prognosis, while patients with SSM have an intermediate prognosis. Progression from ISM to a more advanced form is uncommon (~5% lifetime risk); however, patients should be monitored for emergence of hematologic disease and end-organ manifestations of ASM. In infants and children with cutaneous manifestations, namely, maculopapular cutaneous mastocytosis, mastocytoma(s), or bullous lesions, visceral involvement is usually lacking, and spontaneous resolution is common prior to adolescence. Polymorphic maculopapular cutaneous mastocytosis usually resolves spontaneously. Progression from cutaneous mastocytosis (CM) to ISM may occur in ~10% of children, especially in those with high mast cell burden (diffuse cutaneous mastocytosis) or hematologic abnormalities and those who present with smaller uniform lesions with diameters measuring <2 cm (monomorphic MPCM).

CLINICAL MANIFESTATIONS

The clinical manifestations of SM are due to the release of bioactive substances acting at both local and distal sites, tissue infiltration by mast cells, and the tissue response to the cellular infiltrate. Clinical manifestations include intermittent urticaria, flushing, tachycardia and vascular collapse, gastric acid hypersecretion, crampy lower abdominal pain, and diarrhea. The increased local mast cell burden in the skin (MPCM), bone marrow, and gastrointestinal tract may be a direct cause of pruritus, bone pain, and malabsorption, respectively. Mast cell–mediated fibrotic changes may occur in liver, spleen, and bone marrow but not in gastrointestinal tissue or skin.

The cutaneous lesions of MPCM are reddish-brown macules, papules, or plaques that respond to trauma with urtication and erythema (Darier's sign). Two distinct forms of MPCM are recognized: polymorphic MPCM and monomorphic MPCM. Children with CM may present with MPCM, mastocytomas, or diffuse cutaneous mastocytosis (DCM). Mastocytomas are generally solitary elevated lesions that are yellow, brown, or red in color. Their size may vary from a few millimeters to several centimeters. Rubbing or irritation of a mastocytoma lesion may lead to systemic symptoms such as flushing and urticaria. Children with DCM present without distinct lesions, but rather a generalized thickening of skin and "peau d'orange" appearance due to diffuse mast cell infiltration. DCM is associated with bullae formation and more severe systemic symptoms, including upper gastrointestinal irritation and vascular collapse in the first few years of life. Maculopapular skin lesions of mastocytosis may be present in patients with adult-onset systemic disease. The apparent incidence of cutaneous lesions is ≥80% in patients with ISM and <50% in those with SM-AHNMD or ASM. In the upper gastrointestinal tract, gastritis and peptic ulcer are significant problems. In the lower intestinal tract, the occurrence of diarrhea and abdominal pain is attributed to increased motility due to mast cell mediators. The periportal fibrosis associated with mast cell infiltration may lead to portal hypertension and ascites. In some patients, anaphylaxis may occur with rapid and life-threatening vascular collapse. Anaphylaxis is most commonly induced by Hymenoptera stings and nonsteroidal anti-inflammatory drugs (NSAIDs). The neuropsychiatric disturbances are clinically most evident as impaired recent memory, decreased attention span, and "migraine-like" headaches. Patients may experience exacerbation of a specific clinical sign or symptom variably with alcohol ingestion, temperature changes, stress, use of mast cell–interactive opioids, or ingestion of NSAIDs.

TABLE 366-3 Diagnostic Criteria for Systemic Mastocytosis[a]

Major:
Multifocal dense infiltrates of mast cells (>15 mast cells per aggregate) in bone marrow or other extracutaneous tissues
Minor:
Abnormal mast cell morphology (spindle shape or atypical immature morphology) in >25% of mast cells
Aberrant mast cell phenotype with expression of CD25, CD2, or CD30
Detection of *KIT* D816V mutation or other activating *KIT* mutation in peripheral blood, bone marrow, or an extracutaneous organ
Total serum tryptase >20 ng/mL

[a] Diagnosis requires either the major criterion and one minor criterion or three minor criteria.

DIAGNOSIS AND DIFFERENTIAL DIAGNOSIS

Cutaneous mastocytosis is diagnosed by observing the characteristic lesions of MPCM or mastocytoma(s). A skin biopsy may be obtained to confirm these subvariants of CM, whereas patients with suspected DCM and bullous mastocytosis require a skin biopsy to confirm the diagnosis. Although the diagnosis of SM is generally suspected based on clinical history, physical examination findings, and laboratory procedures, it can only be confirmed with a tissue diagnosis. The diagnosis of SM necessitates a lesional biopsy to meet the criteria of one major plus one minor or three minor findings **(Table 366-3)**. The major criterion requires mast cell aggregates of >15 mast cells, often in paratrabecular and perivascular locations with associated lymphocytes and eosinophils. Minor criteria include abnormal "spindled" mast cell morphology, aberrant mast cell membrane immunophenotype, or an activating *KIT* mutation (classically *KIT* D816V). A basal serum tryptase level is a noninvasive approach to consider before bone marrow biopsy. A rare histopathologic subvariant called *well-differentiated systemic mastocytosis* (WDSM) is characterized by clusters of mature-appearing, fully granulated and round mast cells, lack of aberrant CD25 and CD2 expression, and lack of D816V *KIT* mutation in most patients. These patients often have a history of childhood-onset cutaneous disease, and their mast cells may display aberrant CD30 expression and other markers of clonality such as atypical (non-D816V) *KIT* mutations. Additional studies directed by the presentation include a bone densitometry, bone scan, or skeletal survey; computed tomography scan or endoscopy; and a neuropsychiatric evaluation. Osteoporosis is increased in mastocytosis and may lead to pathologic fractures.

Some patients presenting with recurrent symptoms referrable to mast cell activation (particularly anaphylaxis with hypotensive syncope) have been found to have underlying mastocytosis. A subset of these patients may be found to have evidence of a clonal hematologic process such as the D816V *KIT* mutation, but lack full diagnostic criteria for SM. Such patients are termed to have *monoclonal mast cell activation syndrome*.

The differential diagnosis requires the exclusion of other disorders. A 24-h urine assessment of 5-hydroxy-indoleacetic acid and metanephrines should exclude a carcinoid tumor and pheochromocytoma, respectively. Hereditary α-tryptasemia may be characterized by symptoms consistent with mast cell activation in addition to multisystem involvement and elevated baseline serum tryptase. These patients have autosomal dominant inheritance of extra-allelic copies of the α-tryptase gene at the TPSAB1 locus. Most patients with recurrent IgE-induced or idiopathic anaphylaxis present with urticaria, angioedema, and/or bronchospasm, which are not typical manifestations of anaphylaxis in SM.

TREATMENT

Mastocytosis

The management of SM is symptom control using a stepwise symptom/sign–directed approach. Medications include an H_1 antihistamine for flushing and pruritus, an H_2 antihistamine or proton pump inhibitor for gastric acid hypersecretion, oral cromolyn sodium for diarrhea and abdominal pain, and occasionally aspirin (in those who are known to be tolerant of NSAIDs) for severe flushing to block biosynthesis of prostaglandin D_2. Systemic glucocorticoids appear to alleviate malabsorption. A self-injectable epinephrine prescription is recommended for most patients due to increased incidence of anaphylaxis. Patients with a history of systemic Hymenoptera venom reaction should be evaluated for venom-specific IgE and placed on lifelong venom immunotherapy if positive.

Mast cell cytoreductive therapy consisting of midostaurin, avapritinib, IFN-α, or cladribine has classically been reserved for advanced variants of SM. Targeted tyrosine kinase inhibitors (such as avapritinib) are now considered first-line therapy for those with advanced SM or refractory symptoms in ISM. In ISM, avapritinib reduces bone marrow mast cell burden and improves symptoms. The efficacy of cytoreductive therapy in mastocytosis is variable, perhaps because of side effects that limit dosage. Imatinib is not effective in most cases as the D816V *KIT* mutation mediates resistance. Combination chemotherapy is appropriate for the frank leukemias. Stem cell transplantation has been shown to be effective in a small subset of patients with advanced mastocytosis.

FURTHER READING

Akin C: *Mastocytosis: A Comprehensive Guide*. New York, Springer International Publishing, 2020.

Arber DA et al: International consensus classification of myeloid neoplasms and acute leukemias: Integrating morphologic, clinical, and genomic data. Blood 140:1200, 2022.

Gotlib J et al: Avapritinib versus placebo in indolent systemic mastocytosis. NEJM Evid 2:6, 2023.

Hartmann K et al: Cutaneous manifestations in patients with mastocytosis: Consensus report of the European Competence Network on Mastocytosis; the American Academy of Allergy, Asthma & Immunology; and the European Academy of Allergology and Clinical Immunology. J Allergy Clin Immunol 137:35, 2016.

Theoharides TC et al: Mast cells, mastocytosis, and related disorders. N Engl J Med 373:163, 2015.

Valent P et al: European Competence Network on Mastocytosis. Proposed diagnostic algorithm for patients with suspected mastocytosis: A proposal of the European Competence Network on Mastocytosis. Allergy 69:1267, 2014.

367

Autoimmunity and Autoimmune Diseases

Mariana J. Kaplan

DEFINITION, INCIDENCE, AND PREVALENCE

A key function of the human immune system is to protect the body from harmful external and internal insults, including infections, various toxins, and cancer cells. In addition to the chemical and structural barriers that protect from noxious agents, the immune system provides two mechanisms of defense: innate and adaptive immune responses. Importantly, an effective immune system should be able to distinguish self from nonself and avoid harmful responses to the body's tissues. As such, while self-recognition is important developmentally to clear dead cells throughout the body and to shape the lymphocytes' immune receptors repertoire, bona fide autoreactive responses that are harmful to the organism are suppressed under normal conditions.

Autoimmunity is defined as the development of adaptive immune responses (autoantibodies [autoAbs] or T lymphocyte responses) that react with autoantigens (autoAgs). This phenomenon is present in all individuals, does not necessarily carry pathogenic consequences, and tends to be more prevalent with age. In contrast, in autoimmune diseases (ADs), an aberrant response of the innate and adaptive immune system develops, leading to a loss of tolerance to self that results in clinically relevant pathogenic responses. In these instances, the immune system attacks various cells, tissues, and organs by mistaking them for external organisms, causing damage and clinical manifestations. Approximately 150 ADs (organ-specific and systemic) have been described (some of them are highlighted in **Table 367-1**). With scientific and technologic advances, a better characterization of the pathogenic pathways leading to clinical manifestations in autoimmunity is being defined, and it is very likely that new definitions of diseases (based on clinical features and/or on pathophysiologic pathways) will continue to emerge. Overall, AD are chronic and incurable conditions that significantly affect the quality of life and, in many cases, are characterized by accelerated mortality.

From various epidemiologic studies, there appears to be an ongoing trend for increased incidence and prevalence of many ADs. It was estimated that, in the United States in 2009, the prevalence of AD was 7.6–9.4% of the general population, or ~25–31 million people, which is likely an underestimation. ADs can occur at any age, and many of them affect predominantly females. As a group, ADs are in the top 10 leading causes of death in women. An individual affected by an AD has a high risk of developing another autoimmune condition, of having other family members who will develop autoimmunity, and of developing certain cancers and premature cardiovascular disease.

TABLE 367-1 Examples of Some Organ-Specific and Systemic Autoimmune Diseases (ADs)

ORGAN-SPECIFIC AD	SYSTEMIC AD
Hashimoto's thyroiditis	Systemic lupus erythematosus
Graves' disease	Rheumatoid arthritis
Type 1 diabetes mellitus	Primary antiphospholipid syndrome
Addison's disease	Ankylosing spondylitis
Pernicious anemia	Idiopathic inflammatory myopathies
Celiac disease	Sjögren syndrome
Autoimmune hepatitis	Progressive systemic sclerosis
Primary biliary cholangitis	Mixed connective tissue disease
Primary sclerosing cholangitis	Giant cell arteritis
Vitiligo	Takayasu arteritis
Pemphigus and other autoimmune bullous diseases	Polyarteritis nodosa
	ANCA-associated vasculitis
Multiple sclerosis	Anti glomerular basement membrane disease
Myasthenia gravis	
Autoimmune polyglandular syndromes	Cryoglobulinemic vasculitis
Autoimmune oophoritis	IgA vasculitis (Henoch-Schönlein)
Autoimmune orchitis	Hypocomplementemic urticarial vasculitis
Primary CNS angiitis	
Isolated aortitis	Kawasaki disease
Cutaneous arteritis	Behcet's disease
Cutaneous leukocytoclastic angiitis	Cogan's syndrome
Guillain-Barré syndrome	Psoriasis/psoriatic arthritis
Immune thrombocytopenic purpura	Inflammatory bowel disease
Autoimmune hemolytic anemia	Rheumatic fever
Autoimmune pancreatitis	Sarcoidosis
Dermatitis herpetiformis	Undifferentiated connective tissue disease
Autoimmune autonomic ganglionopathy	
Autoimmune encephalitis	
Autoimmune inner ear disease	
Autoimmune myocarditis	
Autoimmune myelofibrosis	
Autoimmune retinopathy	
Autoimmune sudden sensorineural hearing loss	
Chronic inflammatory demyelinating polyneuropathy	
Cutaneous lupus	

Abbreviations: ANCA, antineutrophil cytoplasmic antibody; CNS, central nervous system; IgA, immunoglobulin A.

PATHOGENESIS

GENETIC AND EPIGENETIC FACTORS

The development of AD is considered a multistep process, where both genetic and environmental factors play key roles in leading to immune dysregulation and loss of tolerance, which can precede the development of clinical symptoms by many years. A variety of susceptibility genes associated to autoimmunity have been described, and there is familiar aggregation of AD. Furthermore, in type 1 diabetes mellitus (T1DM), rheumatoid arthritis (RA), multiple sclerosis (MS), and systemic lupus erythematosus (SLE), the disease concordance in monozygotic and dizygotic twins is between 15 and 30% and <5%, respectively. This suggests that epigenetic factors also play important roles in AD susceptibility. Many genes associated with enhanced risk for AD are known to regulate immune responses, and some of the identified genes are associated with more than one AD. In most cases, ADs are polygenic, although monogenic disorders that lead to autoimmunity have been described and helped clarify pathways that lead to loss of immunologic tolerance. Among the susceptibility genes that have been identified, specific human leukocyte antigen (HLA)-DR types, deficiency in complement component genes, and genes associated with the type I interferon (IFN) pathway and with B- and T-cell function have all been associated with enhanced autoimmunity risk. HLA haplotype is considered the best predictor of AD risk and has been linked to the development of autoAbs in family members of AD patients. One potential mechanism linked to HLA is that aberrant self-Ags (e.g., misfolded, posttranslationally modified during cell death, citrullinated) that are transported by major histocompatibility complex (MHC) class II can exhibit antigenic ability that is different from normal self-Ags and promote autoimmunity. This phenomenon has been reported, for example, in organ-specific autoimmunity such as Graves' disease and in systemic autoimmunity such as RA. Abs binding to autoAg complexes with MHC class II molecules are associated with enhanced risk of various systemic ADs such as RA and antineutrophil cytoplasmic antigen (ANCA)-associated vasculitis (AAV).

There is significant overlap between specific gene variants and several different ADs, indicating common genetic pathways that may explain risk for more than one disease. For example, a common haplotype of the signal transducer and activation of transcription 4 (*STAT4*) gene has been associated with susceptibility to SLE, RA, and primary Sjögren syndrome, whereas polymorphisms in protein tyrosine phosphatase nonreceptor type 22 (*PTPN22*) are associated with T1DM, SLE, RA, and other ADs. A list of some of the genes associated with AD risk is provided in **Table 367-2**.

Epigenetic factors have been implicated in early onset of some ADs, as well as in the development of lymphocyte autoreactivity and aberrant trained immunity in innate immune cells. Indeed, both metabolic and epigenetic programming characterized by various histone modifications, changes in DNA methylation, and other immunomodulatory alterations have been described in various ADs. Environmental factors that alter genes through epigenetic modifications may have subsequent effects on autoimmunity susceptibility in the offspring of individuals affected by AD.

ENVIRONMENTAL FACTORS

Many studies support the concept that several "hits" from the environment, as well as poorly understood stochastic events in genetically predisposed individuals, will contribute to the development of clinically relevant AD. Many environmental exposures have been linked to an enhanced risk for developing clinical autoimmunity, including ultraviolet light; infections (discussed in detail later in the chapter) and mucosal dysbiosis; tobacco use; exposure to silica, solvents, pesticides, or mercury; certain dietary exposures in childhood (sugar, cow's milk,

TABLE 367-2 Genetic Polymorphisms Associated with Autoimmune Disease Risk

GENE	DISEASE ASSOCIATIONS WITH SOME ADs
HLA-DRB1	MS, RA, T1DM, JIA, IIM, IBD, Graves' disease, Hashimoto's thyroiditis, Addison's disease
TNF	SLE, RA, ankylosing spondylitis
IL10	SLE, RA, MS, psoriasis
IL2/IL2R	MS, RA, celiac disease, T1DM, psoriasis
IL6	RA, SLE, vasculitis, JIA
IL17A	SLE, RA, MS, spondyloarthropathies
IFNG	SLE, RA, MS
FOXP3	T1DM, autoimmune thyroid disease, MS
TLR7	SLE
IL23A	Graves' disease, psoriasis, IBD, MS
MIF	RA, MS, SLE
HLA-B	Ankylosing spondylitis
C4A	SLE
IL33	RA, ankylosing spondylitis,
STAT4	RA, SLE, primary Sjögren syndrome
PTPN22	SLE, RA, JIA, T1DM, vasculitides, autoimmune thyroid diseases, vitiligo

Abbreviations: AD, autoimmune disease; IBD, inflammatory bowel disease; IIM, idiopathic inflammatory myopathies; JIA, juvenile idiopathic arthritis; MS, multiple sclerosis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; T1DM, type 1 diabetes mellitus.

food additives) and adulthood; various medications; and recreational drugs. The mechanisms by which environmental factors promote autoimmunity are complex and incompletely defined and include synthesis of reactive oxygen species (ROS), induction of proinflammatory pathways leading to DNA damage, lipid peroxidation that induces protein modification, and neoAg formation. These inflammatory pathways may promote activation of innate and adaptive immune responses that contribute to AD development.

Cigarette smoking is a known risk factor for some ADs, including RA, Graves' disease, SLE, and MS. It has been hypothesized that smoking promotes citrullination and carbamylation of proteins in the lung, which, in individuals who carry certain genetic risk alleles and synthesize autoAbs characteristic of the disease (anticitrullinated protein peptide Abs [ACPAs]) or rheumatoid factor, can promote the development of full-blown disease. Other described associations include silica and SLE and solvents and systemic sclerosis.

Various proinflammatory dietary factors, including the consumption of sugary beverages and red meat, have been associated with increased risk for T1DM, inflammatory bowel disease (IBD), and RA. Deficiency of vitamin D has also been associated with enhanced risk for AD, including T1DM, Crohn's disease, SLE, and MS. Other dietary factors associated with T1DM include no breast feeding in infancy, higher gluten intake in early childhood, and lower omega-3 fatty acid intake. Overall, more research is needed to better understand the impact of diet and other environmental factors in AD risk in specific populations.

■ SEX

Many ADs show skewing toward enhanced incidence and prevalence in women. Sex hormones, X-chromosome dosing, presence of specific immune-related genes in sex chromosomes, and various environmental effects on sex hormones have been proposed to play important roles in modulating differences in immune responses and in autoimmunity development. Both innate and adaptive immune responses differ between healthy men and women, and this may have a significant effect in modulating autoimmunity. For example, Ag-specific responses to both foreign and self-Ags are enhanced in women compared to men, and young women appear to generate more autoAgs than young men.

Estrogen modulates neutrophil immunometabolism and maturation, activates B- and T-cell responses, and promotes Ab production. Androgens decrease B-cell and neutrophil maturation and can suppress autoAb synthesis in SLE. However, the specific roles of sex hormones in autoimmunity need to be better defined.

A higher incidence of autoimmunity has been associated with the presence of two X chromosomes (in women and in men with Klinefelter syndrome [XXY]) due to enhanced expression of X-linked immunity-related genes encoding for molecules such as Toll-like receptor (TLR)-7, TLR-8, and IRAK-1. Because TLR-7 recognizes endogenous RNA-containing self-Ags and promotes the synthesis of type I IFNs, a group of cytokines with crucial roles in SLE and other systemic ADs, this is an important putative mechanism for augmented autoimmunity risk. Indeed, a breakdown in X chromosome inactivation mechanisms has been implicated in autoimmunity.

■ AGING

Immune aging has been proposed as a mechanism promoting autoimmunity. Aging leads to accumulation of senescent innate and adaptive immune cells with proinflammatory features (inflammaging). In predisposed subjects, it is considered that inflammaging may promote overt AD. Among the observed features, genomic instability, mitochondrial dysfunction, and inefficient clearance of intracellular debris, including cellular stress responses, potentiate aberrant T-cell and myeloid cell dysfunction. Immune senescence has been implicated in ADs such as RA and giant cell arteritis.

THE INNATE IMMUNE SYSTEM AND AUTOIMMUNITY

■ ABERRANT CELL DEATH AND CELL DEATH CLEARANCE

The innate immune system cells express a variety of sensors that can detect nucleic acids derived from exogenous and endogenous sources. Genetic, epigenetic, and environmental factors can lead to various forms of dysregulated/proinflammatory cell death and/or impaired clearance of dead cell debris. Some specific forms of cell death may particularly promote autoAg generation and induction of autoimmune responses. Certain TLRs (7–9) can sense nucleic acids and may not properly distinguish between self- and microorganism-derived nucleic acids. Single-stranded RNA (TLR-7 ligand) and unmethylated CpG motifs (TLR-9 ligand) can be found as self-Ags derived from dead cells. Similarly, cytosolic sensors of DNA and RNA are widely expressed across various cell types and tissues and can detect various RNAs and DNAs (including mitochondrial DNA) that can access the cell to stimulate type I IFN synthesis and/or inflammasome assembly and secretion of proinflammatory cytokines. In the steady state, however, risk of responding to self-nucleic acids is usually contained due to their instability, presence of certain DNA sequences that inhibit TLR signaling, and sequestration of nucleic acid–sensing TLRs in the endosomes. However, in genetically predisposed individuals, these mechanisms may be impaired and tolerance break may occur. Factors such as ultraviolet light, infections, and other insults leading to tissue injury may promote enhanced cell death that can contribute to fueling the fire in individuals at risk. In genetically predisposed individuals, an imbalance in dead cell handling can promote the externalization and modification of DNA, RNA, and other intracellular autoAgs, leading to downstream innate and adaptive immune dysregulation, autoAb development, and the formation of immune complexes (ICs) that deposit in various organs, stimulate immune and stromal cells, and promote damage. Delay in the engulfment of apoptotic cells by phagocytes (through genetic or acquired defects) and/or impaired function of intracellular or extracellular nucleases that can clear the nucleic acids and other cellular debris can lead to prolonged exposure of modified autoAgs. Furthermore, a delayed clearance of dead cells promotes the breakdown of these apoptotic cells with release of proinflammatory components and nucleic acids to the extracellular space through secondary necrosis. Indeed, abnormalities in receptors or ligands that facilitate cell death clearance are associated with increased predisposition to SLE and other ADs, both in humans and in animal models of disease. There is a high frequency of SLE in patients with various deficiencies of complement pathway components, which is considered linked to impaired clearance of dying cells. Dysregulation of other types of cell death such as pyroptosis, necroptosis, and NETosis has been associated with autoimmune and chronic inflammatory conditions. For example, oxidation of nucleic acids by ROS or enzymes during NETosis can enhance the half-life of modified RNA and DNA and prevent their degradation.

■ DYSREGULATION OF INNATE IMMUNE CELLS AND INNATE IMMUNE PATHWAYS

Many innate immune cells and/or innate immune pathways show evidence of dysregulation in patients with AD, but also in those at risk or in relatives of patients with these conditions. Dysregulation in the type I IFN pathway driven by genetic, epigenetic, and environmental factors has been proposed to play a role in the pathogenesis of several ADs and their associated organ damage, including SLE, primary Sjögren syndrome, and idiopathic inflammatory myopathies (IIMs). Type I IFNs play important roles in antiviral defense, but many of the effects of chronic exposure or enhanced cellular response to type I IFNs can promote several features characteristic of immune dysregulation in various systemic ADs. Activation of the inflammasome, complement pathway, and various mechanisms of neuro-immunomodulation in innate immune cells have also been implicated in autoimmunity initiation and perpetuation.

Neutrophils Neutrophils, the first innate immune system cellular responders to noxious stimuli, are implicated in the initiation of, perpetuation of, and organ damage in many systemic ADs. Certain proinflammatory subsets of neutrophils (low-density granulocytes [LDGs]), as well as dysregulation in mechanisms of neutrophil cell death including the formation of neutrophil extracellular traps (NETs; a meshwork composed of various modified nucleic acids and granule proteins), have been proposed to play important roles in disease pathogenesis. Dysregulated NET formation promotes the externalization of modified autoAgs and the induction of aberrant innate and tissue stromal inflammatory responses. Abnormal NET formation and clearance have been implicated in the pathogenesis of SLE, AAV, IIM, primary antiphospholipid syndrome (APS), and RA.

Dendritic Cells (DCs) Both conventional (myeloid) DCs and plasmacytoid DCs (which synthesize large amounts of IFN-α) have been implicated in the pathogenesis of a variety of ADs, particularly SLE. While myeloid DCs can promote tolerogenic responses, such as the generation and maintenance of T regulatory cells (Tregs), their strong antigen-presenting cell (APC) capabilities can lead to priming and differentiation of autoreactive lymphocytes. Myeloid DCs have been reported to have abnormal phenotype and function in many DCs. Plasmacytoid DCs, major producers of type I IFNs in the body, have also been implicated in the pathogenesis of SLE and found to have abnormal phenotype and function in several ADs.

Monocytes, Macrophages, and Natural Killer (NK) Cells A causal link between the dysregulation of monocytes and macrophages and the development of AD has been proposed. In many ADs, these cells infiltrate affected tissues and show evidence of activation and dysregulation. Subsets of inflammatory monocytes are expanded in several ADs. Defective phagocytosis by macrophages, neutrophils, and other phagocytes has been proposed to play an important role in AD by extending the half-life of modified autoAgs. NK cells have also been reported to be dysregulated in ADs and to play a putative pathogenic role in AD.

Trained Immunity This concept describes innate immune responses that are faster and more robust to a second stimulus than to the first one. Trained immunity is acquired through an initial exposure to a stimulus that triggers an immune response that "trains" the stimulated cells. This leads to significant changes in the epigenetic landscape of innate immune cells with modifications in their chromatin structure (decondensation, histone modifications that include methylation and acetylation, and DNA hypomethylation). These changes can occur in hematopoietic cells (and therefore are transmitted to their progeny) or directly in tissue-resident cells. While trained immunity can be beneficial in the case of infections, it can also lead to aberrant inflammation and autoimmunity in maladaptive conditions.

THE ADAPTIVE IMMUNE SYSTEM AND AUTOIMMUNITY

■ LYMPHOCYTE AUTOREACTIVITY

Ag specificities generate randomly, and many lymphocytes can generate autoreactive responses directed at autoAgs. Several regulatory mechanisms are in place to induce self-tolerance, but if they become dysregulated, autoimmune responses can follow. Central and peripheral tolerance mechanisms exist in mammals to avoid autoreactivity. Among central tolerance mechanisms, autoreactive T cells and high-affinity autoreactive B cells are deleted in the thymus and the bone marrow, respectively. Regarding peripheral tolerance, several mechanisms have been proposed, including lymphocyte anergy due to lack of costimulation, cell death and removal of autoreactive lymphocytes, suppression of autoreactive responses by regulatory lymphocytes, and inhibitory costimulatory signals. Other factors include idiotypic network interactions (where Abs are directed against autoAbs), cytokine switching leading to immunosuppression, competition for Ag among different lymphocyte clones, and decreased frequency of lymphocyte precursors. In AD, central and/or peripheral tolerance are impaired due to a variety of putative mechanisms. The break of tolerance against self-Ags can trigger and/or contribute to autoimmunity. These Ags may be found in all cell types (e.g., chromatin, DNA) or be highly specific for a specific cell type (e.g., thyroglobulin, myelin basic protein).

T Lymphocytes T lymphocytes have important pathogenic roles in AD through a variety of mechanisms, including direct cell cytotoxicity, release of proinflammatory cytokines, reactive nitrogen or oxygen species, and promoting B-cell dysregulation. Abnormalities in the number and function of Tregs can have amplifying effects in autoimmunity in SLE, RA, and IBD. A tight balance must exist between T-cell-activating/costimulatory pathways and T-cell inhibitory receptors such as programmed cell death protein 1 (PD-1) and cytotoxic T lymphocyte-associated antigen 4 (CTLA-4) and T-cell-related cytokines (interleukin [IL]-2, IL-6, IL-17, IL-33) to prevent the development of autoimmunity. As such, dysregulation in one or more of these pathways, in a predisposed host, may promote immune dysregulation and autoimmune responses. Polymorphisms in various genes responsible for T-cell regulation and function have been implicated in AD risk. Autoreactive CD4 and/or CD8 T cells have been described in AD, including SLE and RA. A variety of therapies that target aberrant T-cell function have been proven effective in many ADs, strongly supporting the role of T cells in autoimmunity.

B Cells and AutoAbs Abs that react with self-molecules can be detected in healthy individuals. These so-called natural autoAbs are primarily IgM, are encoded by unmutated V(D)J genes, and display a moderate affinity for self-Ags. They are considered to provide a first line of defense against infections as well as contributing to the immune system's homeostasis. On the other hand, high-affinity, somatically mutated IgG autoAbs can perturb various homeostatic pathways, including cell clearance, cell effector functions, and Ag-receptor signaling. These pathogenic high-affinity autoAbs found in AD can form through somatic-hypermutation, class-switch DNA recombination, and Ag-driven clonal selection. AutoAbs are ubiquitous in AD, and they play important roles in disease diagnosis and, in some cases, assessment of disease severity. IgA autoAbs have also been described to play a putative role in AD, particularly in conditions where mucosal immune responses are proposed to play pathogenic roles.

Some autoAbs can be pathogenic in a variety of different ways, including interfering with the function of various cells or soluble molecules, activation of inflammatory cascades through the complement pathway, opsonization of cells or various soluble factors, and activation of cells through the promotion of IC formation by engaging activating

TABLE 367-3 Examples of Autoantibody-Mediated Disease Mechanisms

DISEASE	AUTOANTIBODIES	PATHOGENIC MECHANISM
Rheumatoid arthritis	ANCA, anti-carbamylated Abs, rheumatoid factor	Osteoclast activation promoting bone resorption; activation of macrophages, synovial fibroblasts, neutrophils, mast cells, DCs, platelets; immune complex formation with modified autoAgs with activating properties
Systemic lupus erythematosus	Anti-RNP, anti-dsDNA, etc.	Formation of immune complexes with activation of various immune cells; deposition of immune complexes in tissue promoting tissue damage; induction of type I IFNs through immune complex formation
Goodpasture's disease	Anti-type IV collagen Abs	Abs target basal membrane and promote kidney and lung injury
T1DM	Anti-insulin, glutamic acid decarboxylase, protein tyrosine phosphatase	Association with insulitis and diabetes onset
Pernicious anemia	Anti–parietal cell antibodies	Damage to parietal cells by Abs leads to lack of intrinsic factor
Idiopathic inflammatory myopathies	Anti-aminoacyl-transfer RNA synthetase Abs, anti-Mi2, anti-MDA5, anti-TIF1 gamma, anti-HMGCR, etc.	Activation of inflammatory immune cells, NET formation, induction of muscle atrophy pathways, induction of ROS
Sjögren syndrome	Anti-Ro, anti-La, rheumatoid factor	Anti-Ro Abs can induce neonatal lupus in offspring of SLE and Sjögren patients. IgG anti-Ro can cross the placenta and damage the fetal heart, inducing congenital heart block
ANCA-associated vasculitis	Anti-myeloperoxidase; anti-proteinase 3	Induction of NET, activation of neutrophils, endothelial cell damage
Graves' disease	Anti-TSHR	Anti-TSHR Abs target TSHR on thyroid cells, leading to enhanced thyroid hormones' synthesis
Hashimoto's thyroiditis	Anti-thyroid peroxidase and anti- thyroglobulin	AutoAbs promote complement fixation to thyroid cells, leading to proinflammatory responses and T-cell-mediated cytotoxic injury of thyroid cells
Myasthenia gravis	Anti-acetylcholine receptor	Antagonizes acetylcholine receptor on cells

Abbreviations: Abs, antibodies; Ags, antigens; ANCA, antineutrophil cytoplasmic antibody; DCs, dendritic cells; dsDNA, double-stranded DNA; HMGCR, hydroxymethylglutaryl coenzyme A reductase; IFN, interferon; IgG, immunoglobulin G; MDA5, melanoma differentiation-associated protein 5; NET, neutrophil extracellular trap; RNA, ribonucleic acid; RNP, ribonucleoprotein; ROS, reactive oxygen species; SLE, systemic lupus erythematosus T1DM, type 1 diabetes mellitus; TIF1, transcription intermediary factor 1; TSHR, thyroid-stimulating hormone receptor.

Fc receptors. Of interest, autoAbs not only develop in patients with AD but also are more likely to develop in their family members. In other cases, autoAbs are used as diagnostic tools or as markers of disease activity or worse disease outcomes, but a specific pathogenic role has not been identified. Examples of diseases that are mediated by these autoAb-related mechanisms are listed in Table 367-3.

In several organ-specific and systemic ADs, distinct autoAgs that are targeted by the immune system have been described. For example, in Graves' disease (an autoimmune condition that leads to hyperthyroidism and ophthalmopathy), autoAbs are directed against thyroid-stimulating hormone (TSH) receptor on thyroid cells, and engagement of this receptor leads to enhanced thyroid hormone synthesis. In turn, it has been proposed that TSH receptor and MHC class II complexes present on thyroid epithelial cells can induce autoAb synthesis through presentation of cryptic epitopes. In Hashimoto's thyroiditis, there is formation of autoAbs to thyroid peroxidase and thyroglobulin that promote complement fixation to thyroid cells, leading to a proinflammatory response and T-cell-mediated cytotoxic injury of thyroid cells. In myasthenia gravis, autoAbs can interact and disrupt the function of the acetylcholine receptor, muscle-specific kinase receptor, and other cell surface receptors in the postsynaptic muscle membrane, leading to muscle weakness. Cardiac myosin has been identified as an important autoAg in myocarditis. In T1DM, autoAbs that target β-cell antigens lead to destruction of pancreatic cells that synthesize insulin. Individuals at risk of developing T1DM already can have detectable islet-antigen-specific Abs preceding clinical disease by several years, and their titers may increase as risk of disease progresses. Abs to other autoAgs besides insulin can be detected in this disease and include glutamic acid decarboxylase and insulinoma Ag. Those that develop Abs to multiple targets have a higher risk of evolving into clinically apparent T1DM. In MS, autoAbs to myelin basic protein (MBP) and myelin oligodendrocyte glycoprotein and higher frequencies of T cells that are specific for MBP have been detected and considered pathogenic. In addition, myelin-specific and lipid-specific Abs that can fix complement and induce Ab-dependent cytotoxicity have been described in MS. In pernicious anemia/autoimmune gastritis, the most common cause for vitamin B_{12} deficiency and the end stage of autoimmune gastritis, there is autoAb production and T-cell responses to gastric parietal cells that make intrinsic factor, the protein that promotes absorption of vitamin B_{12} by the intestine. This immune-mediated destruction of gastric parietal cells (targeting the H^+/K^+-ATP in their cell membranes) reduces their ability to secrete intrinsic factor in the stomach. Between 70 and 90% of patients with pernicious anemia have autoAbs to these structures. In primary biliary cholangitis, elevated levels of antimitochondrial (primarily to pyruvate dehydrogenase complex) Abs are characteristic and can precede disease onset by a decade, although whether they have a pathogenic role remains unclear. AutoAbs that recognize citrullinated peptides (ACPA), carbamylated peptides, and IgG (rheumatoid factor) are characteristic of many patients with RA and are associated with worse prognosis. The presence of ACPAs is quite specific for RA, and these autoAbs can be detected in serum of individuals up to a decade before diagnosis. A role for lung and synovial NETs in generating citrullinated and carbamylated autoAgs and promoting autoAb formation has been reported. ACPAs can be directly pathogenic in RA and promote macrophage, synovial fibroblast, and osteoclast activation and NET formation, promoting inflammation and bone damage. Furthermore, anticarbamylated peptide autoAbs are associated with bone damage in RA. Myositis-specific autoantibodies (MSAs) are present in most juvenile and adult cases of IIM. Some interesting examples for associations of autoAbs with disease manifestations arise in these group of muscle diseases. For example, in adult IIM, anti-TIF1γ Abs are associated with myositis-associated cancer, anti-HMGCR Abs with statin-induced myositis, and anti-cN1A Abs with inclusion body myositis. In SLE, Sjögren syndrome, RA, systemic sclerosis, and celiac disease, autoAbs can be detected in serum harvested years before clinical diagnosis of AD is made. It is important to emphasize that the autoAgs associated with many ADs remain unknown.

Age-associated B cells (ABCs) represent a distinct B-cell population that is found expanded in several ADs such as SLE, MS, and RA. ABCs have pathogenic roles in these conditions, through cytokine and autoAb production and T-cell stimulation. Autoreactive long-lived plasma cells, which are more refractory to conventional immunosuppression than other cells, are considered to have an essential role in the maintenance of AD and, potentially, in the development of disease flares. On the other hand, regulatory B cells (Bregs) suppress immune responses and can restrain tissue pathology; as such, their dysfunction may be implicated in autoimmunity predisposition and propagation. Dysregulation of B-cell and plasma cell subsets appears to be a shared feature among many ADs.

Immune Complexes (ICs) ICs have a key role in many ADs, especially those that affect the kidney. They can also modulate immune function in direct and indirect ways, including promoting the uptake of nucleic acids into innate immune cells where they stimulate type I IFN synthesis. ICs can stimulate many immune and stromal cells, producing immune stimulation and tissue damage (Table 367-3).

Inflammatory Cytokines Inflammatory cytokines including TNF, IL-1b, IL-6, IL-17, and type I IFNs are involved in the development and perpetuation of various ADs. Indeed, discoveries related to the role of cytokines in AD have led to major therapeutic advances. For example, blocking TNF is highly effective in significant percentages of patients with RA, psoriasis, seronegative spondyloarthropathies, and IBD. Identifying type I IFNs as major players in autoimmunity has led to the development and use of strategies to downregulate downstream signaling pathways for this group of cytokines. This has included the use of Janus kinase (JAK) inhibitors in several ADs such as RA, psoriatic arthritis, IBD, SLE, and primary Sjögren syndrome, as well as in some organ-specific ADs such as vitiligo and alopecia areata. Drugs that specifically target type I IFN receptor or type I IFNs themselves are being used in the clinic or tested in clinical trials. Other drugs that target aberrant T-cell responses and their downstream effects, including IL-17, IL-23, IL-12/23, and IFN-γ, are also available for use or are being developed. Furthermore, therapies that target abnormal B-cell and plasma cell responses, including anti-CD19, anti-CD20, and anti-BLys, have been developed and approved for SLE, RA, and/or AAV or are being tested in clinical trials.

SOME PROPOSED SPECIFIC MECHANISMS THAT PROMOTE AUTOIMMUNITY

INFECTIONS AND AUTOIMMUNITY

Infections have been extensively described as triggers of AD and as inducers of clinical flares in people with established autoimmunity. Various mechanisms have been proposed, including molecular mimicry, release of hidden self-Ags, epitope spreading, adjuvant effects, epigenetic modifications, and dysregulation of innate immune responses triggered, for example, by direct viral damage.

Molecular Mimicry, Bystander Mechanisms, and Epitope Spreading Molecular mimicry may play a potential role in a variety of ADs; here, self-epitopes that share structural similarities to microbial Ags are presented by APCs. A well-established example of molecular mimicry is rheumatic fever, where Abs to the M-protein of group A *Streptococcus* cross-react with myosin, laminin, and other proteins present in the heart, as well as with neuronal Ags. When these Abs deposit in the heart, they can initiate an inflammatory response, leading to carditis and valvular damage, whereas their deposit in the brain can promote Sydenham's chorea, a specific manifestation of rheumatic fever.

Ab deposition can also lead to enhanced cell death with new autoAg release and a perpetuation of a vicious inflammatory cycle. In this case, during epitope spreading, autoimmune responses develop to self-epitopes that are different from the initiating microbe-specific epitope during an infection. Several viral infections have been hypothesized to trigger autoimmunity through molecular mimicry and/or epitope spreading, including Epstein-Barr virus (EBV; linked to SLE, MS, RA, celiac disease, and Sjögren syndrome), coxsackievirus (linked to myocarditis, Hashimoto's thyroiditis, and T1DM), and influenza A virus (linked to T1DM and IgA vasculitis [Henoch-Schönlein]). Besides molecular mimicry, EBV has been linked to MS through reprogramming of B cells that have been latently infected. Molecular mimicry has also been associated with the etiology of RA, including a potential role for *Porphyromonas gingivalis* through the generation of enzymes that promote Ag citrullination in the gingiva.

Gastrointestinal dysbiosis may also play a role in the development of the early stages of RA. Microbial superAgs, including staphylococcal protein A and staphylococcal enterotoxins, can stimulate a broad number of T and B cells by interacting with immune receptors, independent of Ag specificity. If autoreactive lymphocytes express these receptors, autoimmune responses may be triggered through these bacterial molecules.

Another mechanism by which infections have been implicated in autoimmunity is by enhancing the immunogenicity of autoAgs through innate immune signals triggered by the microbe or its products. This bystander mechanism is many times mediated by cytokine release and activation of autoreactive lymphocytes. Infections can also promote an adjuvant effect, when the microbial Ag activates innate immunity to self-organs, as supported by animal models of RA and MS. It is important to note that, often, the link between infections and AD development is circumstantial and causality has not been proven. Support of causality has come from animal models, such as the case of autoimmune myocarditis triggered by coxsackievirus B3 or cytomegalovirus. More recently, SARS-CoV-2 infection has been linked to myocarditis development, autoAb synthesis, and, in some cases, new-onset autoimmunity including APS, multisystem inflammatory syndrome in children (MIS-C), Kawasaki disease, and Guillain-Barré syndrome.

The "Hygiene Hypothesis" This hypothesis is linked to the role of parasites in preventing autoimmune and allergic responses in developing countries, where the prevalence of parasitic infections is high. This mechanism has been implicated in potential higher risk for certain ADs in developed countries with less prevalence of various infections with parasites, viruses, and bacteria known to dampen immune responses or shift to anti-inflammatory milieus by stimulation of immune regulatory pathways. This may apply to both pathogenic microbes and gut commensal bacteria. This hypothesis may help explain geographical variations in the distribution of some ADs. While the hygiene hypothesis is supported by epidemiologic data, the mechanisms triggering this immunosuppressive effect remain to be better characterized.

Sequestered AutoAg Exposure from Immune-Privileged Sites This phenomenon has been proposed to protect vital tissues and organs (eye, brain, and testis) from injury by inflammatory responses directed at foreign microbial antigens. This special immunologic status is associated with long survival rates of allografts and xenografts, when compared to nonprivileged sites. The eye is considered an immune-privileged site as an evolutionary mechanism to prevent vision loss from inflammatory injury. A variety of mechanisms in the eye confer immune privilege (immunosuppressive cytokines, low MHC expression, lack of lymphatic drainage). Immune privilege in the testis confers protection from systemic immune attack to preserve male fertility. Several mechanisms have been proposed to play a role, including physical structure, local anti-inflammatory milieu (including sex hormones), and systemic immune tolerance.

PARANEOPLASTIC SYNDROMES, CANCER, AND AUTOIMMUNITY

Various tumor Ags can be targets of adaptive autoimmunity; these can include Ags expressed in nonmalignant cells within the tumor. A variety of cancers are associated with autoimmune responses that present as paraneoplastic syndromes and other autoimmune responses. Frequently, these syndromes can present before the diagnosis of the underlying neoplasia. A variety of mechanisms have been implicated, including disruption of central tolerance, peripheral immune dysregulation, and modification of self-Ags with cross-reactivity. Some paraneoplastic autoimmune conditions include Lambert-Eaton myasthenic syndrome, paraneoplastic pemphigus, dermatomyositis, Sweet's syndrome, thymoma-associated multiorgan autoimmunity, autoimmune hemolytic anemia, immune thrombocytopenia, vitiligo, and thyroiditis. Cancers commonly associated with autoimmunity include melanoma, thyroid cancer, and non-small-cell lung cancer. One specific example is the development of encephalitis through generation of Abs to *N*-methyl-D-aspartate glutamate receptor in women affected by ovarian cancer. In some cases, these events may be associated with favorable disease outcomes that may suggest a beneficial effect from the autoimmune response. The autoimmune manifestations can be severe, and management of these patients is difficult given the coexistent malignancy.

IATROGENIC AUTOIMMUNITY

Checkpoint Inhibitors Immune checkpoint inhibitors (ICIs) have revolutionized cancer treatment. They act by blocking inhibitory molecules involved in regulating T lymphocytes, promoting the destruction of tumors by the adaptive immune system. Immune-related adverse events (irAEs) have been described as toxic complications related to tissue-specific autoimmunity from the use of ICIs, and they can persist after withholding therapy and using immunomodulatory therapy to control them. Among these complications, DM, arthritis, thyroiditis, and colitis are prevalent. Up to 85% of individuals treated with an ICI may develop irAEs. Most of these are mild, and overall survival and time to malignancy progression are not influenced by the development of these immune complications.

Drug-Induced Autoimmunity In addition to ICIs, many drugs have been linked to the development of autoimmunity. These include antibiotics, antihypertensives, antiarrhythmics, TNF inhibitors, and antiseizure medications, among many others. Drug-induced lupus is a well-described entity, and patients developing this condition have characteristic detection of antihistone Abs. In many cases, discontinuation of the drug may be sufficient for symptoms to be abrogated or decreased.

PREVENTION AND TREATMENT OF AUTOIMMUNITY

The hope is that better understanding of the pathophysiology of overt development of ADs and their associated organ damage will lead to primary or secondary prevention strategies in conditions such as T1DM, RA, and SLE. One promising example is where infants at risk of developing multiple T1DM autoAbs and clinical disease are being studied to determine if certain interventions (oral insulin) can delay or stop progression to overt disease. Studies have also been performed on people considered at risk for RA to assess delayed disease progression, but more studies are needed to better characterize the utility of these preventive strategies. Overall, opportunities to initiate early therapy depend on the availability of serologic tests that can predict disease, to initiate treatments that have a good efficacy-to-safety ratio.

Specific treatments for ADs are included in chapters focusing on individual diseases and include a variety of immunomodulatory agents (e.g., antimalarials) and immunosuppressive drugs (spanning those drugs with broad effects on the immune system, such as corticosteroids or cyclophosphamide, to highly specific biologic agents targeting specific cytokines or immune cells). It is the hope that the development of personalized medicine approaches will lead in the future to the specific targeting of dysregulated immune cell components while sparing other critical cells and functions of the immune system.

FURTHER READING

Enhancing NIH Research on Autoimmune Diseases. Consensus Study Report. The National Academies of Sciences, Engineering, Medicine, 2022.

Gupta S, Kaplan MJ: Bite of the wolf: Innate immune responses propagate autoimmunity in lupus. J Clin Immunol 131:e144918, 2021.

Johnson D, Jang W: Infectious diseases, autoantibodies and autoimmunity. J Autoimm 137:102962, 2023.

Singh N et al: Immune-related adverse events after immune check point inhibitors: Understanding the intersection with autoimmunity. Immunol Rev 318:81, 2023.

368

Systemic Lupus Erythematosus

Chelsey J.F. Smith, Paul J. Hoover,
Kenneth Kalunian

DEFINITION, PREVALENCE, AND EPIDEMIOLOGY

Systemic lupus erythematosus (SLE) is an autoimmune disease in which overactive innate and adaptive immune systems cause tissue damage through the effects of autoantibodies and immune complexes. Many organ systems can be affected, with cutaneous, musculoskeletal, and renal systems most involved, followed by pulmonary, hematologic, cardiovascular, serosal, and central nervous system involvement. Constitutional symptoms are often present. Autoantibodies can be detected years prior to a clinical diagnosis. Approximately 90% of affected individuals are women, most of childbearing age; however, the disease can also affect neonates and older children, men, and elderly individuals. Certain ethnic and racial groups are at higher risk for more prevalent disease and more severe disease. The Centers for Disease Control and Prevention (CDC) National Lupus Registry estimates the prevalence in the United States to be 204,295 cases, with overall prevalence being nine times higher in women compared to men. The highest prevalence is seen in Black women, followed by Hispanic, White, and Asian/Pacific Islander women. Among men with SLE, Black men have the highest prevalence and White men have the lowest. Prognosis and survival from SLE can vary widely by geographic region, race, ethnicity, and access to both care and medications.

DIAGNOSIS AND CLASSIFICATION

For the diagnosis of SLE, an individualized approach should be taken based on available clinical data and excluding other disease entities. Classification criteria for SLE have been developed for the purposes of enrollment of patients into clinical trials rather than criteria for diagnosis. Currently, the 2019 European League Against Rheumatism (EULAR)/American College of Rheumatology (ACR) classification criteria for SLE and the 2012 Systemic Lupus International Collaborating Clinics (SLICC) SLE classification criteria are both utilized for enrollment of patients into clinical trials. For the EULAR/ACR criteria, an antinuclear antibody (ANA) titer of at least 1:80 is required. Clinical criteria are then weighed by points within several clinical domains, and a patient requires at least 10 points to meet criteria. No single clinical or laboratory value leads to 10 points except for a renal biopsy revealing class III or IV lupus nephritis.

Certain subsets of patients with SLE are currently not included in clinical trials, including patients with incomplete lupus or features of SLE without an established diagnosis and the ~5% of SLE patients who are ANA negative. Researchers have suggested that if predictive or diagnostic biomarkers for progression are developed, then early treatment for patients with incomplete lupus could prevent progression to SLE. Furthermore, because the EULAR/ACR criteria exclude ANA-negative patients from participating in clinical trials, the opportunity to understand whether new drugs have a role in this disease subset is also underserved.

SLE has been classified into two broad categories based on symptoms. Those with predominant classic findings of SLE that have a clear relationship to immune dysregulation, such as nephritis, arthritis, and vasculitis, have been categorized as having type 1 lupus, whereas those with predominant symptoms of fatigue, diffuse body pain, depression, cognitive dysfunction, sleep disturbances, anxiety, and/or brain fog have been categorized as having type 2 lupus. Type 1 symptoms arise from autoimmune inflammation and/or organ damage and often respond to standard immunosuppression, whereas type 2 patients are less responsive to immunosuppressive therapy. This categorization presents an opportunity to explore the genetic and immune association of each category as well as develop biomarkers and treatments for the distinct types.

PATHOGENESIS

■ OVERVIEW

SLE reflects multiple immunoregulatory defects and is characterized by the production of autoantibodies to cellular components, especially nucleic acids and nuclear proteins, that promote inflammation and tissue damage. The interaction of environmental factors with the stochastic dysregulation of genes controlled by common genetic variations underpins the development of SLE **(Fig. 368-1)**.

■ GENETICS

SLE has a strong genetic component as evidenced by familial clustering, a greater incidence in monozygotic twins versus dizygotic (24% vs 2%), and the identification of >150 susceptibility loci from genome-wide association studies (GWAS). Genetic studies have identified groups of immunoregulatory genes associated with SLE, reflecting common pathways that may contribute to pathogenesis. It is possible that each pathway differentially contributes to the expression of disease, accounting for clinical heterogeneity. The most common genetic association is within the major histocompatibility complex that contains genes for antigen presentation molecules (class I and class II), several complement components, and cytokines. Nearly half of the susceptibility loci are associated with type 1 interferon (IFN) production or its downstream signaling. For example, some susceptibility genes lead to increased IFN through defects in nucleic acid sensing and metabolism (*TLR7, ADAR, IFIH1, SAMHD1, RNASEH2B, TREX1*) or nucleic acid degradation (*DNASE1, DNASE1L3*), whereas other IFN-related susceptibility genes amplify the IFN response (*IRF5, IRF7, IRF8, STAT4*). Many predisposing genes are immune cell type specific, affecting activation and survival of T cells (*OX40L*) or B cells (*BANK1, BACH2*). The complement cascade (C1Q, C1r, C1s, C2, C4), which is essential for clearance of apoptotic cellular debris and immune complexes, is often affected in monogenic cases, leading to severe SLE.

■ ENVIRONMENT

About a quarter of monozygotic twins are clinically concordant for SLE, implicating environmental factors in disease expression. Low socioeconomic status is associated with the development of SLE and may include exposures to environmental toxins, including mercury, pesticides, and diet. Viral infections such as Epstein-Barr virus likely contribute to disease development by inducing the IFN response during an acute infection or by expressing viral proteins that activate SLE susceptibility genes. Ultraviolet light can impact DNA methylation, generate self-stimulatory nucleic acids, activate keratinocyte immune responses, and produce autoreactive T cells. Finally, emerging evidence suggests that the gut microbiome may increase gut permeability, promoting the translocation of certain gut microbes into the blood with the development of lupus-specific autoantibodies contributing to SLE pathogenesis.

■ INNATE IMMUNITY

Multiple innate immune cell defects contribute to increased cellular breakdown and reduced cell debris clearance, culminating in the production of autoantibodies. Neutrophils are typically short-lived and abundant immune cells, representing a large burden of cellular debris that is usually cleared without inducing an immune response. Patients with SLE, however, have neutrophils with higher turnover, delivering a large load of stimulatory nucleic acids and autoantigens. Cellular debris and immune complexes are typically cleared by macrophages and dendritic cells through complement-mediated phagocytosis and nucleic acid digestion; these mechanisms are dysregulated in SLE. In turn, the burden of cellular debris exceeds that which can be cleared in SLE, triggering nucleic acid sensors and toll-like receptors (TLRs) in plasmacytoid dendritic cells to express type 1 IFN. Consequently, IFN primes neutrophils for further turnover and sensitizes the adaptive

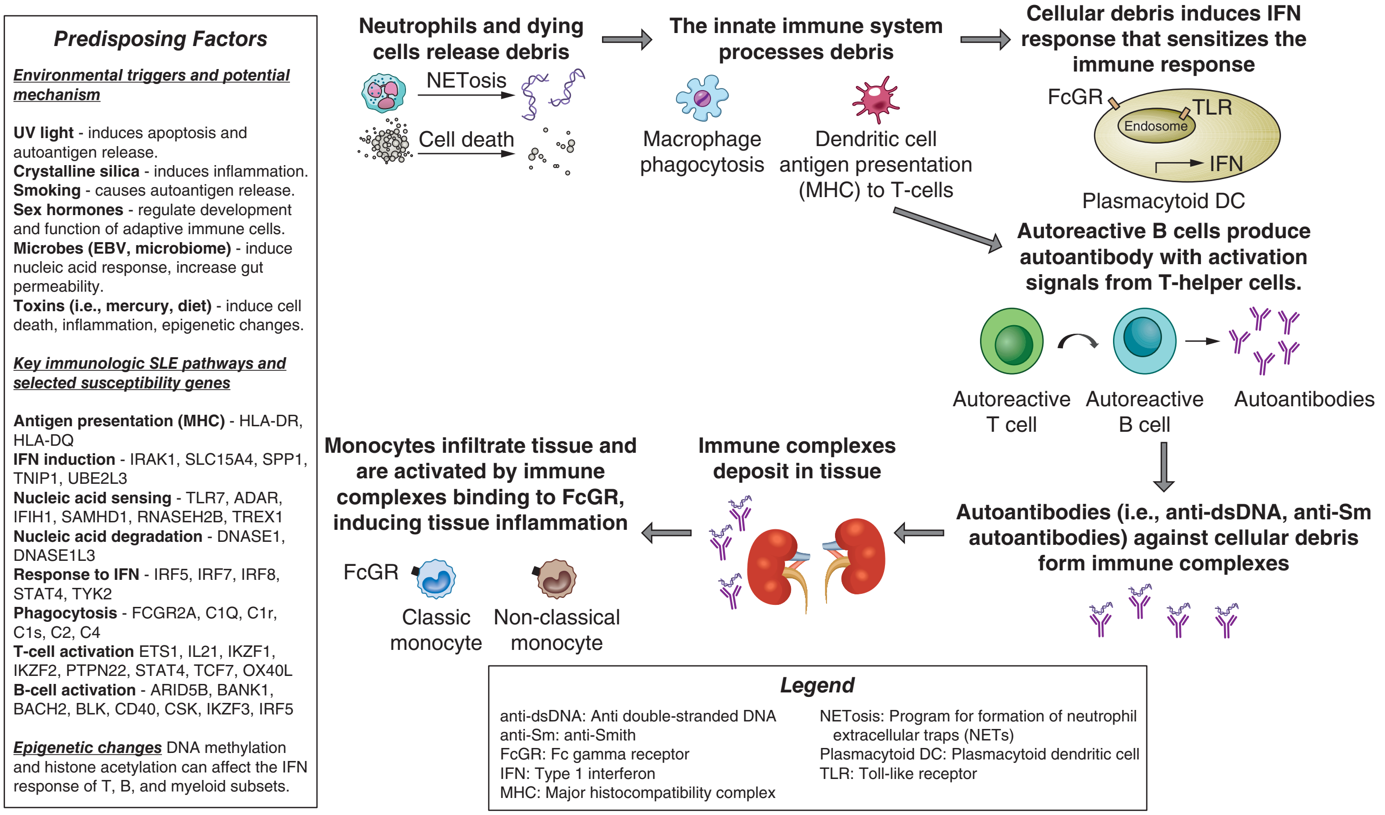
Predisposing Factors
Environmental triggers and potential mechanism
UV light - induces apoptosis and autoantigen release.
Crystalline silica - induces inflammation.
Smoking - causes autoantigen release.
Sex hormones - regulate development and function of adaptive immune cells.
Microbes (EBV, microbiome) - induce nucleic acid response, increase gut permeability.
Toxins (i.e., mercury, diet) - induce cell death, inflammation, epigenetic changes.
Key immunologic SLE pathways and selected susceptibility genes
Antigen presentation (MHC) - HLA-DR, HLA-DQ
IFN induction - IRAK1, SLC15A4, SPP1, TNIP1, UBE2L3
Nucleic acid sensing - TLR7, ADAR, IFIH1, SAMHD1, RNASEH2B, TREX1
Nucleic acid degradation - DNASE1, DNASE1L3
Response to IFN - IRF5, IRF7, IRF8, STAT4, TYK2
Phagocytosis - FCGR2A, C1Q, C1r, C1s, C2, C4
T-cell activation ETS1, IL21, IKZF1, IKZF2, PTPN22, STAT4, TCF7, OX40L
B-cell activation - ARID5B, BANK1, BACH2, BLK, CD40, CSK, IKZF3, IRF5
Epigenetic changes DNA methylation and histone acetylation can affect the IFN response of T, B, and myeloid subsets.
Neutrophils and dying cells release debris
NETosis
Cell death
The innate immune system processes debris
Macrophage phagocytosis
Dendritic cell antigen presentation (MHC) to T-cells
Cellular debris induces IFN response that sensitizes the immune response
FcGR
TLR
Endosome
IFN
Plasmacytoid DC
Autoreactive B cells produce autoantibody with activation signals from T-helper cells.
Autoreactive T cell
Autoreactive B cell
Autoantibodies
Autoantibodies (i.e., anti-dsDNA, anti-Sm autoantibodies) against cellular debris form immune complexes
Immune complexes deposit in tissue
Monocytes infiltrate tissue and are activated by immune complexes binding to FcGR, inducing tissue inflammation
FcGR
Classic monocyte
Non-classical monocyte
Legend
anti-dsDNA: Anti double-stranded DNA
anti-Sm: anti-Smith
FcGR: Fc gamma receptor
IFN: Type 1 interferon
MHC: Major histocompatibility complex
NETosis: Program for formation of neutrophil extracellular traps (NETs)
Plasmacytoid DC: Plasmacytoid dendritic cell
TLR: Toll-like receptor

FIGURE 368-1 Pathogenesis of systemic lupus erythematosus (SLE). Autoantigens from neutrophil NETosis and cellular turnover overwhelm clearance mechanisms regulated by the innate immune system in SLE. Consequently, these autoantigens accumulate and induce an immune response. Nucleic acid activates TLR signaling in plasmacytoid dendritic cells (DCs), producing type 1 IFN, sensitizing the innate and adaptive immune responses. DCs present autoantigens to autoreactive T cells, leading to B-cell stimulation and the production of immune complexes against these autoantigens. Immune complexes deposit in tissue and induce local immune responses that lead to inflammation and repair and, eventually, fibrosis after recurrent SLE flares. EBV, Epstein-Barr virus.

immune system. Other innate immune cells with aberrant function in SLE include classical monocytes that infiltrate and repair injured tissue due to immune complex deposition and nonclassical monocytes that patrol the vascular lumen for injury. Sustained activation of these subsets from SLE leads to tissue inflammation and scarring.

ADAPTIVE IMMUNITY

B cells play a central role in the SLE pathogenesis based on the production of autoantibodies due to their loss of tolerance. Autoantibodies, especially those complexed with nucleic acid, are pathogenic mediators of tissue inflammation and damage in SLE. These autoantibodies induce type 1 IFN and other inflammatory mediators and activate patrolling monocytes in tissue. A subset of circulating age-associated B cells (ABCs) expands in SLE patients and eventually matures into autoantibody-secreting plasma cells, indicating that SLE patients have an increased population of B cells primed to generate pathogenic autoantibodies. Similar to B cells, T cells lose tolerance and play a central role in the autoimmune response. In particular, both T follicular and T peripheral helper cells promote B-cell differentiation into pathogenic, high-affinity, autoantibody-secreting plasma cells. T regulatory cells meant to maintain tolerance in both T and B cells are defective in SLE.

CLINICAL MANIFESTATIONS

OVERVIEW AND SYSTEMIC MANIFESTATIONS

The initial presentation of SLE is variable and may present as nonspecific constitutional symptoms with or without single-organ or multiorgan involvement. Autoantibody presence is helpful in attributing nonspecific symptoms to SLE. Symptoms can be mild to severe at any time over the course of the disease, including at initial presentation. Patients with severe disease often have systemic symptoms such as fever, anorexia, and unintentional weight loss. Approximately 15% of patients have relatively mild disease, which may or may not be accompanied by fatigue, brain fog, and/or mild arthralgia. The majority of SLE patients have active disease despite therapy or have frequent flares of their disease requiring treatment adjustment. Goals for SLE treatment are low-level disease activity or remission on ongoing therapy. (See "Outcomes" section for more details.)

MUSCULOSKELETAL MANIFESTATIONS

Nonspecific arthralgia and myalgia are common in SLE. Lupus arthritis is characterized by a polyarthritis most commonly involving the wrists, the metacarpal-phalangeal and proximal interphalangeal joints of the hands, and the knees. In some cases, damage of the periarticular ligaments may lead to Jaccoud-like changes **(Fig. 368-2*A*)**. Rheumatoid arthritis and SLE can occur simultaneously, and this overlap is denoted by the term *rhupus*. While radiographs rarely demonstrate erosive changes in lupus arthritis, emerging studies with magnetic resonance imaging (MRI) and ultrasound imaging suggest a possible erosive pattern that is less severe than that seen with rheumatoid arthritis. Pain in a single hip, shoulder, or knee out of proportion to other joints should prompt consideration of avascular necrosis of bone, particularly with history of corticosteroid usage. Inflammatory muscle disease is also seen in lupus, characterized by symmetrical proximal weakness, elevated creatine kinase and aldolase levels, and inflammatory changes on muscle biopsy. Corticosteroids used to treat SLE may cause myopathy, as can antimalarials, although this is rare.

CUTANEOUS MANIFESTATIONS

Three major categories of SLE skin manifestations include acute cutaneous lupus erythematosus (ACLE), subacute cutaneous lupus erythematosus (SCLE), and chronic cutaneous lupus **(Fig. 368-2*B*-*F*)**. ACLE occurs with SLE disease activity, whereas SCLE and certain forms of chronic cutaneous lupus may be seen in the absence of systemic disease. The most classic form of ACLE is the malar rash, which appears in a "butterfly" distribution across the cheeks of the face with nasolabial sparing. Due to its photosensitive nature, slightly raised lesions, and location on the face, this may be confused with rosacea, which involves the nasolabial folds. ACLE may also present as a generalized maculopapular rash in sun-exposed areas of the body, bullous lupus, toxic epidermal necrolysis variant, and generalized photosensitivity. SCLE is a photosensitive rash associated with anti-Ro or SSA antibodies that is characterized by flat, red-rimmed annular or psoriasiform lesions. Discoid lesions are rough, circular lesions that are slightly raised, scaly, and hyperpigmented with depigmented, atrophic centers and erythematous rims, and are the most common chronic dermatitis seen in SLE. Other chronic forms of cutaneous lupus include hypertrophic or verrucous lupus, lupus panniculitis, tumid lupus, chilblains lupus, and discoid lupus/lichen planus overlap. Nonspecific cutaneous manifestations that may be seen in SLE include nonscarring alopecia, oral and nasal painful or nonpainful mucosal ulcers, and the less common leukocytoclastic and urticarial vasculitides. Approximately one-third of SLE patients also experience Raynaud's phenomenon, a condition in which small blood vessels undergo vasospasm with exposure to cold environment or stress, causing a classically triphasic color change of the fingers and/or toes, inducing white (blanching), blue (cyanosis), and red (hyperemia) coloring of the extremities.

RENAL MANIFESTATIONS

Nephritis occurs in ~35% of patients with SLE and is more severe in non-White patients, more frequently leading to end-stage renal disease (ESRD) despite therapy. Nephritis can be seen with or without extrarenal SLE disease and may be present in an asymptomatic patient. Patients should therefore be screened at regular 3-month intervals with a random urine protein-to-creatinine ratio (UPCr) or a 24-h urine protein and creatinine determination. Usually, nephritis is seen in patients with elevated serum anti–double-stranded DNA (dsDNA) antibody titers and decreased serum levels of C3 and/or C4, and trends of these biomarkers over time may serve as warning signs of impending active nephritis. Renal biopsy is generally prompted by a UPCr of >0.5 and/or declining kidney function. Classification is based on renal pathology from a renal biopsy and is defined by the International Society of Nephrology/Renal Pathology Society criteria. Aggressive therapy is generally reserved for patients with class III (focal proliferative glomerulonephritis), class IV (diffuse proliferative glomerulonephritis), or a combination of class III or IV with class V disease (membranous) **(Fig. 368-3, Chap. A4)**. Approximately 20% of lupus nephritis patients with significant proteinuria tend to have class V disease and more often have nephrotic syndrome. Patients with class V lupus nephritis have a better overall prognosis than patients with class III or IV lupus nephritis and are treated in the same manner as class III and IV. Patients who do not respond to treatment may require a repeat biopsy to evaluate for a change in class. In the United States, ~20% of SLE patients with class IV lupus nephritis die or develop ESRD within 10 years. Newer approaches are being developed to both increase renal survival and spare patients from accelerated atherosclerosis, hypertension, hyperglycemia, and hyperlipidemia, all of which can be accelerated by renal disease and corticosteroid usage.

NERVOUS SYSTEM MANIFESTATIONS

Lupus patients experience neuropsychiatric manifestations of disease that are characterized by diffuse or focal symptoms. The most common diffuse symptom is "brain fog," a cognitive dysfunction or slowing of thought. The etiology is not well understood but may be related to autoantibodies crossing the blood-brain barrier, leading to local inflammation and neuronal damage. Headache is common and often severe, although it may be difficult to distinguish from migraine or tension headache. Seizures of any type may be seen in neuropsychiatric lupus. Acute psychosis and confusional states are uncommon but can occur in SLE without any structural abnormalities. In these cases, glucocorticoid-induced psychosis and infectious encephalitis should be ruled out. Transverse myelitis is a rare SLE complication that manifests as bilateral neurologic deficits such as symmetric motor weakness and decreased sensation due to spinal cord inflammation.

VASCULAR OCCLUSION: MYOCARDIAL INFARCTIONS AND STROKE

Coronary heart disease from accelerated atherosclerosis in SLE is a significant cause of morbidity and premature death, especially in young patients without other risk factors. Vascular injury due to excessive

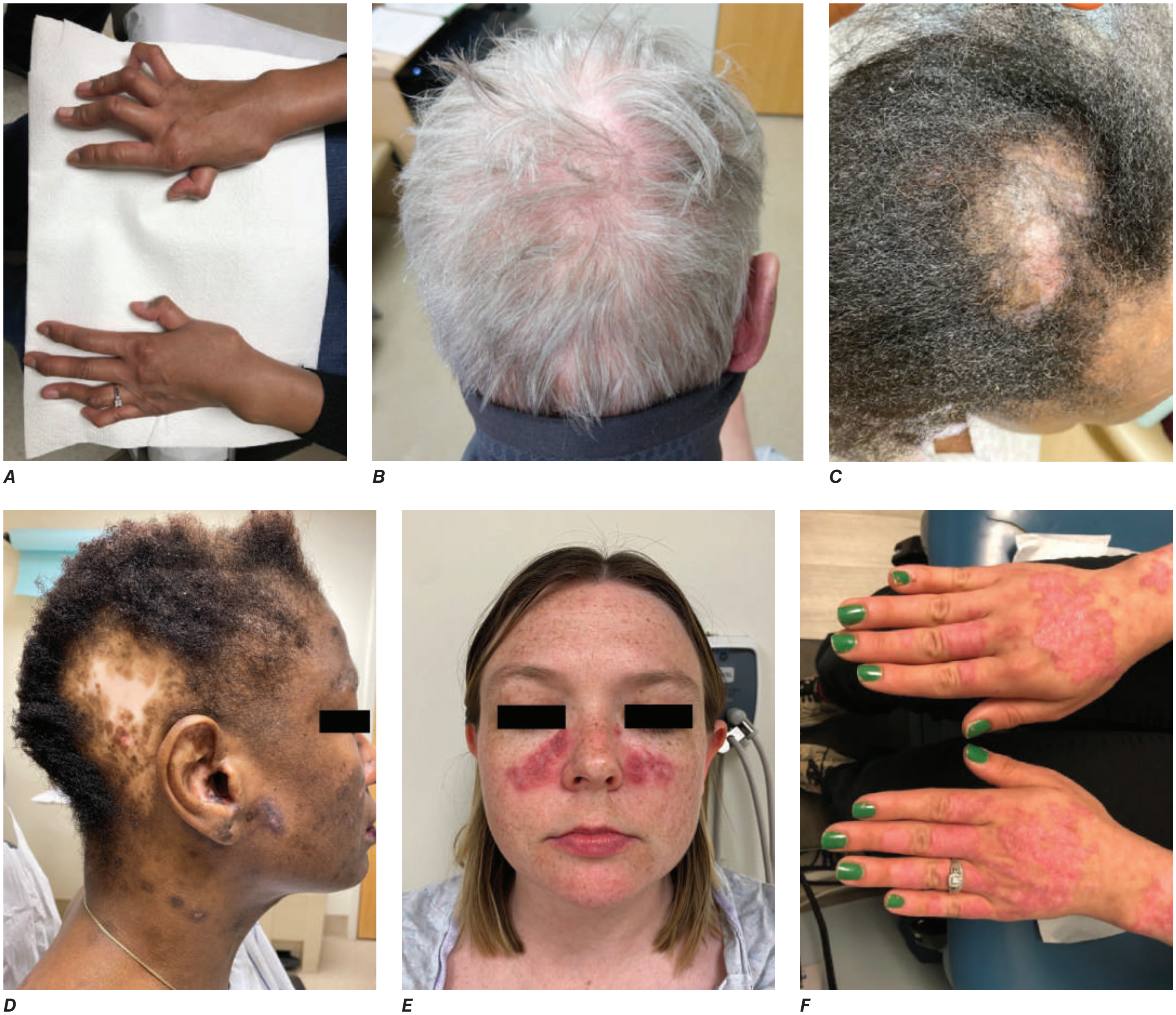

FIGURE 368-2 Characteristic musculoskeletal and cutaneous manifestations seen in systemic lupus erythematosus (SLE). ***A.*** Jaccoud arthropathy, characterized by chronic, nonerosive, and reversible deviations caused by joint capsule inflammation with subsequent fibrotic retraction and metacarpophalangeal joint subluxation. ***B.*** Patchy, nonscarring alopecia. ***C.*** Patchy scarring alopecia. ***D.*** Patchy alopecia associated with chronic, scarring discoid lupus changes. ***E.*** Classic photosensitive malar rash. ***F.*** Subacute cutaneous lupus erythematosus (SCLE) with widespread, nonscarring photosensitive features. *(Panels B–E reproduced with permission from Victoria Werth, MD. Panels A and F reproduced with permission from DoQuyen Huynh, MD.)*

type 1 IFN and cellular death may predispose to atherosclerosis. The risk of coronary artery disease and vascular thrombosis is further increased in SLE patients with antiphospholipid syndrome (APS), an autoimmune syndrome that commonly occurs in SLE and is characterized by arterial, venous, or small vessel thromboembolic events in the presence of antiphospholipid antibodies with prothrombotic properties **(Chap. 369)**. Stroke has been reported in up to 19% of patients with SLE, a result of atherosclerosis and increased risk from APS. In addition to stroke, APS appears to be a significant risk factor for other focal central nervous system manifestations including seizures and transverse myelitis.

■ PULMONARY MANIFESTATIONS

The most common pulmonary manifestation of SLE is pleuritis, which can occur with or without exudative pleural effusions. Acute pneumonitis may present in active SLE, with pulmonary infiltrates often appearing indistinguishable from infection on imaging. Diffuse alveolar hemorrhage with capillaritis and interstitial lung disease may also be seen. Restrictive lung defect from reduced lung volumes, also known as shrinking lung syndrome, is uncommon but may occur. Pulmonary arterial hypertension and pulmonary embolism may be present with or without APS.

■ CARDIAC MANIFESTATIONS

All layers of the heart may be involved in SLE, with pericardial involvement being the most frequent. Pericarditis can typically be managed with anti-inflammatory medications, colchicine, and anti-IL-1–directed therapies, or without glucocorticoids, and rarely leads to tamponade physiology. Myocarditis is rarer and may lead to left-sided heart failure and/or arrhythmia. Libman-Sacks endocarditis, a fibrinous sterile form of endocarditis that may be associated with antiphospholipid antibodies, may increase the risk for embolic events. Right-sided heart failure may be present with pulmonary arterial hypertension or chronic lung disease.

■ HEMATOLOGIC MANIFESTATIONS

Cytopenias in SLE are often multifactorial and may be associated with disease activity, medication use, or infection. Anemia is the most common hematologic manifestation of SLE and is present in >50% of SLE cases, with anemia of chronic disease representing approximately

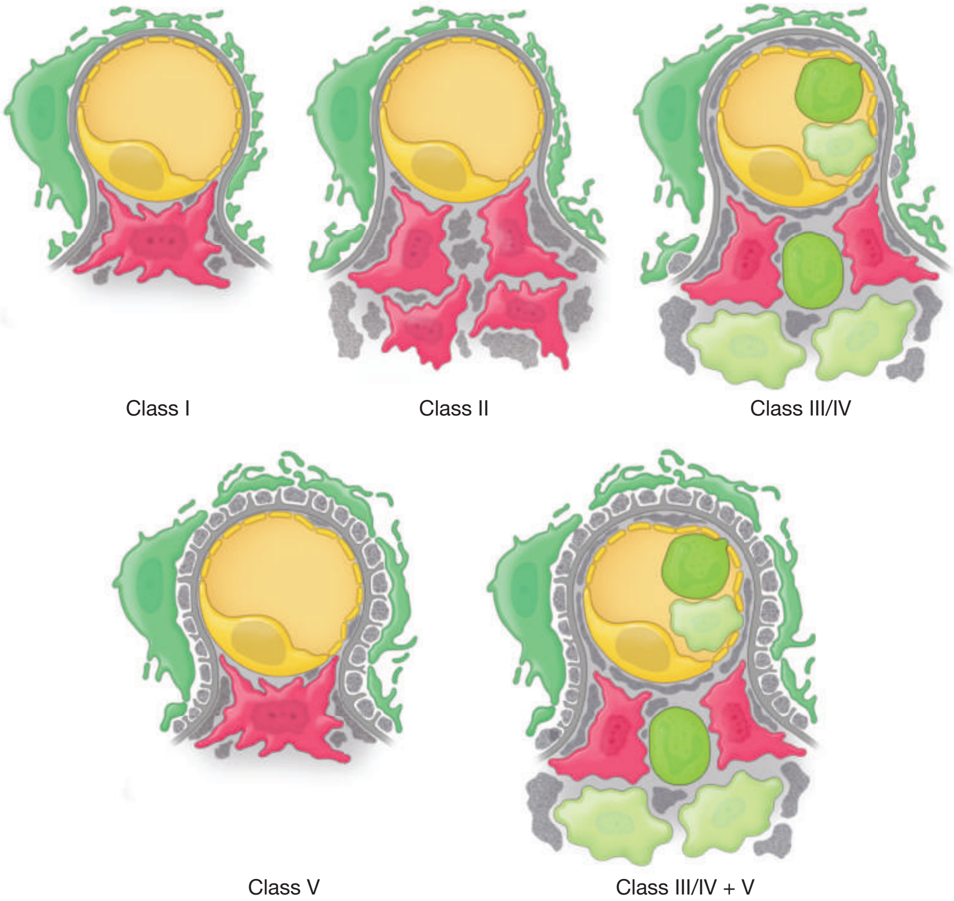

FIGURE 368-3 Ultrastructural features of lupus glomerulonephritis. Class I: Mesangial immune deposits (*black*) without mesangial cell (*red*) hypercellularity or leukocyte infiltration. Class II: Mesangial immune deposits with mesangial cell hypercellularity, no leukocyte infiltration. Class III/IV: Subendothelial immune deposits visible by light microscopy, with infiltration of mesangial and capillary leukocytes (dark green neutrophils and light green monocytes/macrophages). Class III/IV + V: Leukocyte infiltration with subepithelial and subendothelial immune deposits. Class V: Subepithelial immune deposits, no leukocyte infiltration. *(Reproduced with permission from IM Bajema et al: Revision of the International Society of Nephrology/Renal Pathology Society classification for lupus nephritis: Clarification of definitions, and modified National Institutes of Health activity and chronicity indices. Kidney Int 93:789, 2018, Figure 2.)*

one-third of cases; other causes for anemia in SLE include iron deficiency, autoimmune hemolytic anemia, aplastic anemia, microangiopathic hemolytic anemia, and medication effect. Leukopenia (<4000/μL) is common, occurring in about half of SLE patients, and usually consists of lymphopenia (<1500/μL), rather than neutropenia. Thrombocytopenia (<150,000/μL) is usually mild but can also be severe (<50,000/μL) and pose significant bleeding risk. It is most commonly due to immune mechanisms and may be associated with APS or antiphospholipid antibodies against platelet antigens. Splenomegaly, splenic atrophy, and asplenism may all be seen in SLE. Depending on severity of the abnormal blood counts, treatment considerations include glucocorticoids, anti-CD20 biologic agents, platelet growth factors, intravenous immunoglobulin, and/or splenectomy in resistant cases. Rarely, atypical hemolytic-uremic syndrome (formerly designated as thrombotic thrombocytopenic purpura or the syndrome of microvascular thrombotic crisis) may occur. This is a condition that presents with hemolysis, thrombocytopenia, and microvascular thrombosis, as well as brain and other tissue involvement. Laboratory testing shows schistocytes in the peripheral blood smear, low levels of ADAMTS13 activity, and elevated serum lactate dehydrogenase levels.

GASTROINTESTINAL MANIFESTATIONS

Lupus can involve any part of the gastrointestinal tract. Flaring SLE activity has been associated with nonspecific gastrointestinal symptoms including nausea, vomiting, and diarrhea. Abdominal pain may be caused by lupus peritonitis, enteritis, pancreatitis, or vasculitis. Mesenteric vasculitis can lead to intestinal perforation, bleeding, and ischemia and requires high doses of glucocorticoids for treatment. Impaired intestinal motility and protein-losing enteropathy may also occur. Elevated liver enzymes are common and may be associated with flaring disease activity, medications, and/or coexisting autoimmune hepatobiliary disease. Mesenteric thrombosis, Budd-Chiari syndrome, and hepatic venoocclusive disease can be observed with concomitant APS.

OCULAR MANIFESTATIONS

Keratoconjunctivitis sicca is common in SLE, even in the absence of secondary Sjögren's syndrome. Retinal vasculitis, optic neuritis, uveitis, scleritis, peripheral ulcerative keratitis, episcleritis, and nonspecific conjunctivitis may all be seen. Retinal vasculitis is rare but, when present, requires aggressive immunosuppression to prevent blindness. Adverse ocular effects from medications include cataracts and glaucoma from glucocorticoid treatment and hydroxychloroquine (HCQ)-induced maculopathy **(Table 368-1)**.

LABORATORY TESTS

AUTOANTIBODIES

An elevated titer of ANA reflects the underlying immune dysregulation in SLE. Measuring ANA presence and titer is the best screening test for SLE because ANAs are seen in almost all SLE patients. Repeated negative tests by immunofluorescence make a diagnosis unlikely, but rare cases of ANA-negative SLE do occur. Classification of SLE by current EULAR/ACR criteria requires the presence of an ANA titer of 1:80 for enrollment of patients into SLE clinical trials; however, the diagnosis is based on clinical and laboratory features and ultimately, clinical judgement (see "Diagnosis and Classification" section). Numerous ANAs have been identified that target nuclear antigens; in particular, components of the nucleosome (DNA wrapped around a histone

TABLE 368-1 Clinical Manifestations of Systemic Lupus Erythematosus

Constitutional	Fatigue, malaise, fever, weight loss, anorexia
Cutaneous	
Generalized	Photosensitivity, oral and nasal ulcers, alopecia
Acute cutaneous lupus	Malar rash, maculopapular rash, bullous lupus, toxic epidermal necrolysis variant
Subacute cutaneous lupus	
Chronic forms of cutaneous lupus	Discoid rash, panniculitis, tumid lupus, chilblains, verrucous lupus, cutaneous vasculitis
Musculoskeletal	Arthralgia, myalgia, polyarthritis, Jaccoud hand deformity, inflammatory myopathy, avascular necrosis of bone
Hematologic	Anemia, leukopenia, lymphopenia, thrombocytopenia, lymphadenopathy, splenomegaly, venous or arterial thrombosis, atypical hemolytic-uremic syndrome
Cardiopulmonary	Pleurisy, pericarditis, pleural and pericardial effusions, myocarditis, endocarditis, pneumonitis, coronary artery disease, interstitial fibrosis, pulmonary hypertension, diffuse alveolar hemorrhage, shrinking lung syndrome
Vascular	Pulmonary hypertension, stroke, myocardial infarction, peripheral arterial disease, pulmonary embolism, deep vein thrombosis, Raynaud's phenomenon
Neurologic	Cognitive dysfunction, mood disorder, depression, headache, seizures, mononeuropathy, polyneuropathy, stroke, transient ischemic attack, psychosis, aseptic meningitis, transverse myelitis
Renal	Proteinuria ≥500 mg/24 h, cellular casts, nephrotic syndrome, end-stage renal disease
Gastrointestinal	Nausea, abdominal pain, diarrhea, elevated liver enzymes, lupus peritonitis or enteritis, vasculitis, pancreatitis
Ocular	Sicca syndrome, conjunctivitis, episcleritis, scleritis, uveitis, retinal vasculitis

TABLE 368-2 Autoantibodies in Systemic Lupus Erythematosus

ANTIBODY	PREVALENCE, %	CLINICAL ASSOCIATIONS
ANA	95	Titers of 1:80 or higher are considered positive and are sensitive for SLE but not specific.
Anti-dsDNA	70	Higher titers are more specific for SLE. Levels correlate with disease activity and nephritis.
Anti-Sm	30	Most specific autoantibody for SLE and more common in black and Asian patients. May correlate with CNS manifestations and incidence of nephritis.
Anti-RNP	25	Not SLE-specific. Associated with overlap syndromes and increased risk for pulmonary hypertension. Correlates with high interferon gene signature.
Anti-Ro (SS-A)	30	Not SLE-specific. Associated with sicca syndrome, neonatal lupus and congenital heart block, and subacute cutaneous lupus.
Anti-La (SS-B)	20	Associated with anti-Ro.
Antiribosomal P	15	May be associated with CNS involvement. Specific for SLE but not sensitive.
Antihistone	70	Associated with drug-induced lupus.
Anti–U1-RNP	13	Associated with musculoskeletal and lung impairment.
Antiphospholipid *Anticardiolipin* *Anti–β_2-glycoprotein* *Lupus anticoagulant*	50	Arterial or venous thrombosis, pregnancy morbidity and fetal loss, thrombocytopenia. Associated with CNS manifestations, Libman-Sacks endocarditis, hypertension, pulmonary hypertension.

Abbreviations: ANA, antinuclear antibody; CNS, central nervous system; dsDNA, double-stranded DNA; RNP, ribonucleoprotein; SLE, systemic lupus erythematosus; Sm, Smith.

octamer) or RNA-binding proteins (RBPs). In SLE, the most common antinucleosome antibodies are anti-dsDNA antibodies that are present in ~40–60% of patients and highly specific (75–99%). Anti-Smith (anti-Sm) antibodies are the most common anti-RBP, present in ~30% of patients, and are highly specific for SLE (55–100%). Given their high specificity, both anti-dsDNA and anti-Sm antibodies are the only ANA subtypes that are included in the current EULAR/ACR criteria. Anti-Ro (SS-A) antibodies, which recognize a protein complexed primarily to 60-kDa and 52-kDa RNAs, are often seen in ANA-negative SLE patients and are associated with risk for neonatal lupus, sicca symptoms, photosensitivity, and SCLE. Antihistone antibodies may be associated with drug-induced lupus (see "Drug-Induced Lupus" section). Antiphospholipid antibodies, as measured by IgG, IgM, and IgA isotypes of anticardiolipin and anti–beta-2 glycoprotein-1 antibodies (anti-β_2GP-1), and the lupus anticoagulant usually obtained using the dilute Russell venom viper time (dRVVT) are present in ~50% of SLE patients. These antibodies, especially the IgG and IgA isotypes of cardiolipin and β_2GP-1, and the lupus anticoagulant are associated with arterial and venous thrombosis, thrombocytopenia, and recurrent fetal loss **(Table 368-2)**. Many ANA subtypes can be present in SLE and other systemic autoimmune diseases including myositis, rheumatoid arthritis, Sjögren's syndrome, and systemic scleroderma, as well as organ-specific autoimmune diseases such as autoimmune hepatitis, Hashimoto's thyroiditis, and primary biliary cholangitis.

STANDARD DIAGNOSTIC TESTS

Immunologic, renal, and hematologic domains make up the laboratory testing for the diagnosis of SLE. Among the immunologic domain, elevated titers of ANA confer a high likelihood of SLE. Although ANAs are present at low levels in healthy patients, the threshold for a positive test of 1:80 or higher favors SLE. In the setting of ANA negativity with a clinical presentation highly suggestive of SLE, anti-Ro (SSA) confers a high likelihood of SLE. Further immunologic testing for anti-dsDNA and anti-Sm autoantibodies confers specificity. Anti-dsDNA antibodies and reduced levels of complement components C3 and C4 suggest a propensity for lupus nephritis, as these tests indirectly reflect the pathogenic production of autoantibodies and the high load of cellular debris. Laboratory testing is required at initial presentation to screen for hematologic abnormalities including leukopenia, lymphopenia, thrombocytopenia and anemia of chronic inflammation, hemolytic anemia, and atypical hemolytic-uremic syndrome. Initial testing for renal abnormalities is necessary to determine an estimated glomerular filtration rate (eGFR), serum creatinine, and either a random or 24-h UPCr. When commercially available, consideration should be given to measuring urinary biomarkers recently noted to be associated with renal inflammatory and fibrotic disease, including urinary interleukin 16, CD163, type 1 IFN, and transforming growth factor-beta (TGF-β).

TESTS FOR FOLLOWING DISEASE COURSE

Clinical manifestations of SLE vary over time, likely in response to the aberrant immunologic mechanisms that are driving disease. In particular, anti-dsDNA antibody titers are important to track disease activity. These autoantibodies are produced by a short-lived population of B-cell plasmablasts that turn over rapidly, reflecting the presence or absence of immunologic activity. Levels of the complement factors C3 and C4 can drop during episodes of high disease activity, as these are affected by the formation of autoimmune immune complexes and apoptotic cells. In lupus nephritis, the extent of proteinuria and the eGFR are also frequently monitored. Quarterly testing is suggested in asymptomatic patients because hematologic and renal manifestations can be present in the absence of symptoms. Testing should be performed to consider the presence of leukopenia, lymphopenia, thrombocytopenia, anemia (both anemia of chronic inflammation and hemolytic anemia), and rarely, atypical hemolytic-uremic syndrome. Testing for eGFR, serum creatinine, and a random or 24-h UPCr level is essential for routine follow-up care with consideration of a renal biopsy when urinary protein exceeds a new level of UPCr of 0.5 to rule out active lupus nephritis. Consideration should be given to measuring urinary biomarkers on a quarterly basis when available.

MANAGEMENT

EULAR recently updated recommendations for the management of SLE and lupus nephritis. Goals for SLE treatment include improved disease activity with minimal symptoms, prevention of damage, and improved quality of life **(Figs. 368-4 and 368-5)**. Adherence to treatment plans is essential. Achieving these goals with safe treatments and minimal corticosteroids is an integral aspect of achieving low levels of disease activity and remission. If corticosteroids are utilized for active or flaring disease, the goal should be to utilize the lowest possible dosage to suppress disease activity. Aggressive therapy should be reserved for life-threatening manifestations or those that could lead to damage. Consideration should be given to minimize both the effects of active disease and complications from treatment. Renal treatment goals based on EULAR recommendations include (1) a goal that all patients should at least achieve a state of low disease activity, defined by a Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) score of 0–4; or (2) remission, defined as a SLEDAI score of 0, both with stable use of HCQ, immunosuppressives or biologics, and a daily dose of 5 mg of prednisone (or glucocorticoid equivalent) or less (see "Outcomes" section).

■ ESSENTIAL LUPUS THERAPEUTIC APPROACHES AND ANTIMALARIALS

All SLE patients should use daily sunscreen on sun-exposed areas, maintain a healthy diet and weight, avoid smoking, exercise regularly, maintain compliance with vaccination recommendations, and regularly monitor and achieve control of their blood pressure, lipid, and glucose levels. Antimalarials should be used in all patients unless contraindicated, as they have been shown to prevent flares, increase overall survival, and decrease the risk of developing renal disease and accelerated atherosclerosis, among other benefits. HCQ, one of the first drugs approved by the U.S. Food and Drug Administration (FDA) for use in SLE, is the most common antimalarial used and is first-line therapy for patients with skin manifestations and arthritis. It is the most useful drug to improve fatigue in SLE. Use during pregnancy and during breastfeeding has been demonstrated to be safe. Appropriate dosing of HCQ is under investigation, as higher doses may be associated with better disease control. Current recommendations are for dosing up to 5 mg/kg per day (actual body weight) with performance of retinal evaluations at baseline and at yearly intervals to monitor for macular effects, which are associated with prolonged usage. Risk factors for retinal toxicity from HCQ include higher dose and longer duration of use; however, higher doses may also confer better disease control, and overall risk for maculopathy remains low. The most appropriate dose of HCQ for SLE treatment thus remains unclear. Other side effects of HCQ may include rash, nausea, vomiting and diarrhea, and skin dyspigmentation. Tobacco use may interfere with the efficacy of antimalarials. Alternative antimalarials include chloroquine, which has a higher risk of retinal toxicity, and quinacrine, which has a higher

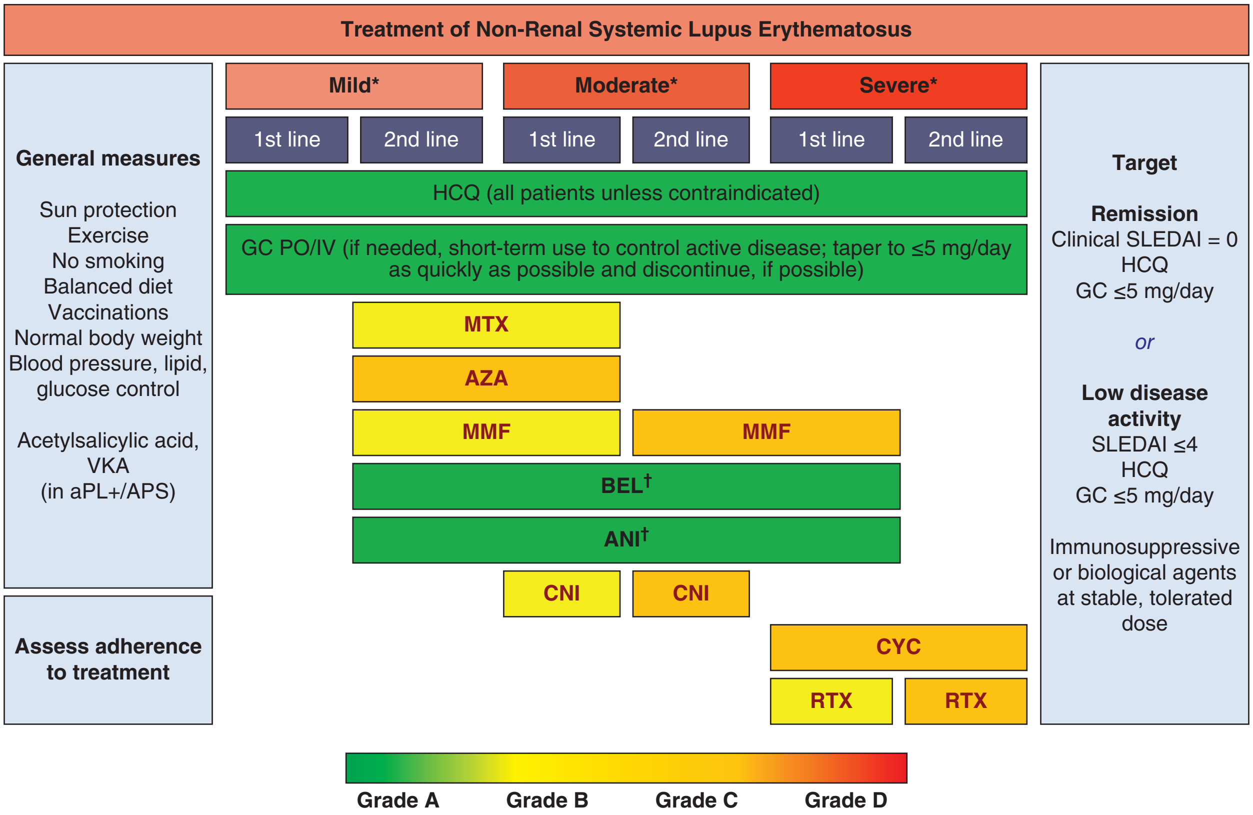
Treatment of Non-Renal Systemic Lupus Erythematosus
General measures
Sun protection
Exercise
No smoking
Balanced diet
Vaccinations
Normal body weight
Blood pressure, lipid, glucose control
Acetylsalicylic acid, VKA (in aPL+/APS)
Assess adherence to treatment
Mild*
Moderate*
Severe*
1st line
2nd line
1st line
2nd line
1st line
2nd line
HCQ (all patients unless contraindicated)
GC PO/IV (if needed, short-term use to control active disease; taper to ≤5 mg/day as quickly as possible and discontinue, if possible)
MTX
AZA
MMF
MMF
BEL†
ANI†
CNI
CNI
CYC
RTX
RTX
Target
Remission
Clinical SLEDAI = 0
HCQ
GC ≤5 mg/day
or
Low disease activity
SLEDAI ≤4
HCQ
GC ≤5 mg/day
Immunosuppressive or biological agents at stable, tolerated dose
Grade A
Grade B
Grade C
Grade D

FIGURE 368-4 Treatment of nonrenal systemic lupus erythematosus (SLE). Notes: Top-to-bottom sequence does not imply order of preference (e.g., MTX, AZA, and MMF are equal options for second-line therapy in mild disease or first-line therapy in moderate disease). *Mild disease: constitutional symptoms; mild arthritis; rash ≤9% body surface area; platelet count (PLTs) 50–100× 10^9/L; SLEDAI ≤6; BILAG C or ≤1 BILAG B manifestation. *Moderate disease: moderate-to-severe arthritis ("rheumatoid arthritis–like"; rash 9–18% body surface area [BSA]; PLTs 20–50 × 10^9/L; serositis; SLEDAI 7–12; ≥2 BILAG B manifestations). *Severe disease: major organ-threatening disease (e.g., cerebritis, myelitis, pneumonitis, mesenteric vasculitis); thrombocytopenia with platelets <20 × 10^9/L; thrombotic thrombocytopenic purpura–like disease or acute hemophagocytic syndrome; rash >18% BSA; SLEDAI >12; ≥1 BILAG A manifestation. †Recommendation of belimumab and anifrolumab as first-line therapy in severe disease refers to cases of extrarenal SLE with nonmajor organ involvement but extensive disease from skin, joints, and so on. The use of anifrolumab as add-on therapy in severe disease refers mainly to severe skin disease. For patients with severe neuropsychiatric disease, anifrolumab and belimumab are not recommended. ANI, anifrolumab; aPL, antiphospholipid antibodies; APS, antiphospholipid syndrome; AZA, azathioprine; BEL, belimumab; BILAG, British Isles Lupus Assessment Group; CNI, calcineurin inhibitor; CYC, cyclophosphamide; GC, glucocorticoids; HCQ, hydroxychloroquine; IV, intravenous; MMF, mycophenolate mofetil; MTX, methotrexate; PO, per os; RTX, rituximab; SLEDAI, SLE Disease Activity Index; VKA, vitamin K antagonists. *(Reproduced with permission from A Fanouriakis et al: Ann Rheum Dis 83:15, 2024.)*

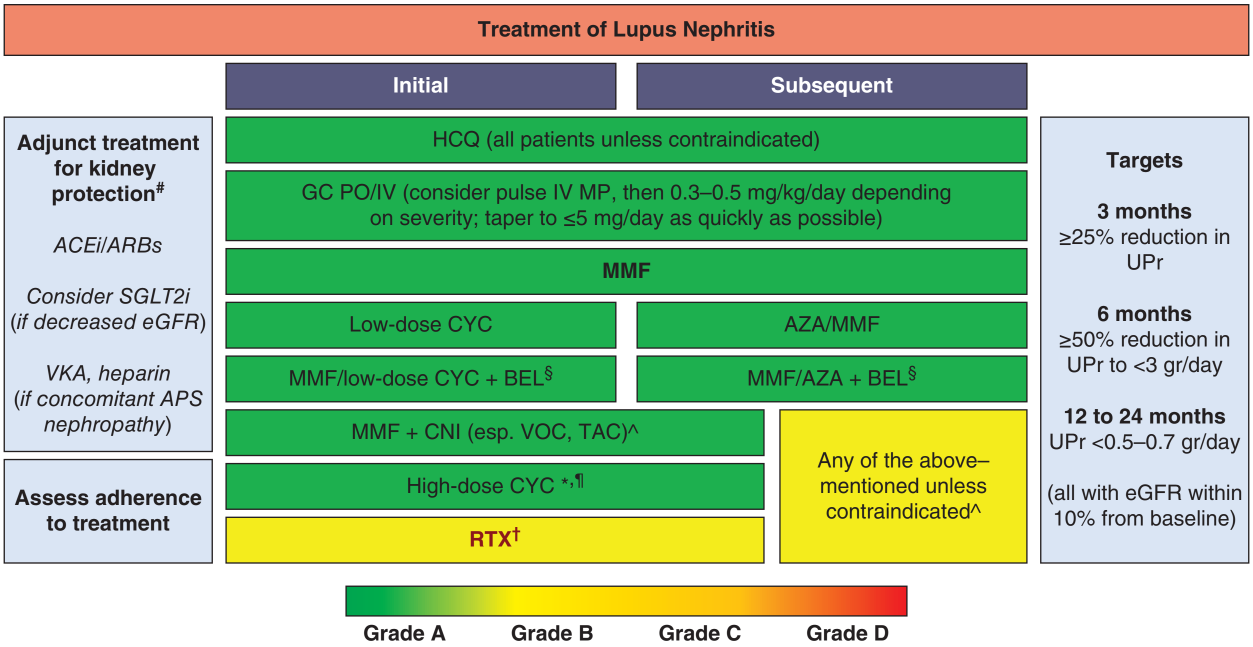
Treatment of Lupus Nephritis
Initial
Subsequent
Adjunct treatment for kidney protection#
ACEi/ARBs
Consider SGLT2i (if decreased eGFR)
VKA, heparin (if concomitant APS nephropathy)
Assess adherence to treatment
HCQ (all patients unless contraindicated)
GC PO/IV (consider pulse IV MP, then 0.3–0.5 mg/kg/day depending on severity; taper to ≤5 mg/day as quickly as possible)
MMF
Low-dose CYC
AZA/MMF
MMF/low-dose CYC + BEL§
MMF/AZA + BEL§
MMF + CNI (esp. VOC, TAC)^
High-dose CYC *,¶
RTX†
Any of the above–mentioned unless contraindicated^
Targets
3 months
≥25% reduction in UPr
6 months
≥50% reduction in UPr to <3 gr/day
12 to 24 months
UPr <0.5–0.7 gr/day
(all with eGFR within 10% from baseline)
Grade A
Grade B
Grade C
Grade D

FIGURE 368-5 Treatment of lupus nephritis. Notes: Top-to-bottom sequence does not imply order of preference. #In addition to general protective measures, as outlined in Figure 368-3. §BEL should always be given in combination with MMF or low-dose CYC as initial therapy and with MMF or AZA as maintenance therapy. ^CNIs should be given in combination with MMF. *Particularly recommended in the presence of poor prognostic factors: reduced eGFR, histologic presence of cellular crescents or fibrinoid necrosis, or severe interstitial inflammation. ¶Extension of high-dose CYC to subsequent phase refers to severe lupus nephritis cases, in which bimonthly or quarterly CYC pulses may be given following 6 monthly pulses. †In relapsing/refractory disease, especially after failure of CYC-based regimens. ACEi, angiotensin-converting enzyme inhibitors; APS, antiphospholipid syndrome; ARB, angiotensin receptor blockers; AZA, azathioprine; BEL, belimumab; CNI, calcineurin inhibitor; CYC, cyclophosphamide; eGFR, estimated glomerular filtration rate; GC, glucocorticoids; HCQ, hydroxychloroquine; IV, intravenous; MMF, mycophenolate mofetil; MP, methylprednisolone; PO, per os; RTX, rituximab; SGLT2i, sodium-glucose cotransporter 2 inhibitors; TAC, tacrolimus; Upr, urine protein; VKA, vitamin K antagonists; VOC, voclosporin. *(Reproduced with permission from A Fanouriakis et al: Ann Rheum Dis 83:15, 2024.)*

risk of skin dyspigmentation. Quinacrine can be used as alternative therapy in cases of toxic retinopathy or as add-on therapy to HCQ for patients with inadequate response. For arthritis and arthralgia, nonsteroidal anti-inflammatory drugs (NSAIDs) can be useful but should be used with caution in SLE patients due to increased risk for aseptic meningitis, hypertension, renal dysfunction, elevated serum transaminases, and increased myocardial infarction risk with cyclooxygenase-2 inhibition.

GLUCOCORTICOIDS

Glucocorticoid dosing is based on the organ-specific needs and severity of disease. While effective at reducing inflammation quickly, their use is associated with dose-dependent side effects including cushingoid metabolic effects, hyperglycemia, hypertension, osteoporosis, and avascular necrosis. If needed, corticosteroids should be tapered as quickly as possible to maintenance doses of 5 mg or less of daily prednisone or its equivalent, and withdrawn as soon as feasible. Use of intravenous methylprednisolone (usually 125–1000 mg daily for 1–3 days) is limited to emergent needs such as active nephritis, severe central nervous system or vascular involvement, and hematologic manifestations such as life-threatening hemolytic anemia or thrombocytopenia. A major goal of novel SLE therapies is the ability to develop treatment algorithms that are less dependent on chronic glucocorticoid use.

SKIN AND MUSCULOSKELETAL MANIFESTATIONS: IMMUNOMODULATING AND IMMUNOSUPPRESSANT DRUGS

All SLE patients and especially those with skin manifestations of disease should be treated with topical sunscreens and antimalarials. If unresponsive to these approaches, topical glucocorticoids and/or calcineurin inhibitors can be utilized. Persistent mucocutaneous disease and/or arthritis may prompt consideration of additional oral immunomodulating drugs such as azathioprine (6-thioguanine purine analogue), leflunomide (lymphocyte-specific pyrimidine synthesis inhibitor), mycophenolate mofetil (MMF), or mycophenolic acid (purine synthesis inhibitors). Biological drugs such as belimumab (anti-BAFF, administered intravenously or subcutaneously) and anifrolumab (anti–type 1 IFN receptor, administered intravenously) are FDA-approved approaches shown to improve outcomes in these clinical domains, with better ability to taper glucocorticoids. Thalidomide and lenalidomide may be used for some cutaneous subtypes but with caution due to risk for polyneuropathy and teratogenicity. Methotrexate (a folinic acid antagonist) is another option for arthritis associated with SLE.

RENAL DISEASE: SUCCESSES WITH COMBINATION THERAPIES

Treatment approaches have been recommended by EULAR for class III and IV lupus nephritis, as well as combination of class V lupus nephritis with either class III or IV (Fig. 368-5). The recommended targets are a 3-month target of 25% or more reduction in UPCr level, a 6-month target of 50% or more reduction in UPCr to <3 g/d, and a 12- to 24-month target in UPCr to <0.5–0.7 g/d, with each of these targets associated with an eGFR within 10% from baseline. All patients with class III, IV, or V nephritis should receive an angiotensin-converting enzyme (ACE) inhibiting agent or an angiotensin receptor blocking (ARB) agent to decrease proteinuria and reduce systemic and glomerular hypertension. With a baseline decreased eGFR, use of a sodium-glucose cotransporter 2 (SGLT2) inhibitor should be considered to reduce glomerular hyperfiltration, lower blood pressure, and decrease proteinuria. Patients with active lupus nephritis with concomitant antiphospholipid nephropathy as noted by biopsy should receive a vitamin K antagonist or heparin.

Current initial regimens include either cyclophosphamide (an alkylating agent administered intravenously) or MMF (administered orally) plus glucocorticoids to reduce progression to ESRD and death. Rates of infection and death are similar in both treatment approaches. Induction in Black or Hispanic patients with MMF 2–3 g daily is appropriate given better response to MMF in these ethnic groups, while use of either MMF or cyclophosphamide may be an option for White and Asian patients. Low doses of cyclophosphamide (500 mg every 2 weeks for six doses) followed by either azathioprine or MMF maintenance are as effective as standard high doses of cyclophosphamide with improved tolerability. High-dose cyclophosphamide monthly for 6 months followed by azathioprine or MMF may be more appropriate for patients with crescentic proliferative glomerulonephritis or rapidly progressive glomerulonephritis. Ovarian failure is a potential consequence of high-dose cyclophosphamide, and therefore, a gonadotropin-releasing hormone agonist may be given prior to each cyclophosphamide dose. Dosing of glucocorticoids as part of the initial therapeutic approach is controversial, with more recent data suggesting the use of intermediate doses for shorter periods of time with more rapid dose reduction. For subsequent therapy, MMF is superior to azathioprine in maintaining renal function and survival, but both may be used, and azathioprine is the preferred agent for lupus nephritis in pregnancy. Patients using azathioprine should be screened for homozygous deficiency of the thiopurine *S*-methyltransferase (*TMPT*) gene, which increases the risk for severe bone marrow suppression, as this gene is responsible for the TPMT enzyme necessary for metabolizing thiopurine drugs. Calcineurin inhibitors such as tacrolimus and cyclosporine may be beneficial for decreasing proteinuria, and tacrolimus can be used in pregnancy.

Two immunomodulating therapies have recently been approved by the FDA for use in lupus nephritis for patients on standard therapy for active nephritis: belimumab, approved in 2020, and the cyclosporine analogue voclosporin, approved in 2021. The randomized controlled trials for these medications demonstrate favorable outcomes without the use of high-dose steroids. Voclosporin-based triple immunotherapy (with MMF and glucocorticoids) or belimumab-based triple therapy (with either low-dose cyclophosphamide or MMF and glucocorticoids) can be considered as an initial approach in active lupus nephritis. Patients with active lupus nephritis and inadequate response to conventional therapies may also consider alternative biologics targeting B cells such as rituximab. While rituximab studies in lupus nephritis did not demonstrate efficacy compared to standard therapies, issues with the trial designs may have hampered the demonstration of an effect.

Treatment approaches for patients with pure class V membranous glomerulonephritis have not been well studied, and commonly patients with this presentation are treated only in the setting of significant proteinuria. All patients with class III, IV, or V nephritis, including combinations of class III or IV with class V, should be considered for treatment with ARB or ACE inhibition to reduce proteinuria.

■ THERAPIES FOR OTHER ORGAN-BASED SYSTEMS

Few studies have focused on central and peripheral nervous system, cardiac, pulmonary, serosal, gastrointestinal, or hematologic manifestations of SLE. Anti-CD20 agents such as rituximab are used for lupus-related hemolytic anemia and severe thrombocytopenia. Colchicine can be used to treat pleural, pericardial, or peritoneal inflammatory symptoms. MMF, cyclophosphamide, and rituximab plus corticosteroids are the mainstay of treatment for serious central and peripheral inflammatory nervous system presentations including psychosis, vasculitis, and transverse myelitis. Plasmapheresis and high levels of glucocorticoids can be lifesaving in atypical hemolytic-uremic syndrome. In refractory patients, rituximab and a C5 inhibitor, eculizumab, are used. Concomitant APS should be treated with warfarin for long-term anticoagulation rather than direct oral anticoagulants **(Chap. 369)**.

■ EMERGING THERAPIES AND TREATMENT APPROACHES

SGLT2 inhibitors are being studied in lupus nephritis for their effects on preserving function through tubular protection with effects on inflammation and fibrosis. Obinutuzumab, a CD20-targeted monoclonal antibody that is commercially available for oncologic indications, has recently demonstrated efficacy in a phase 3 study for lupus nephritis, and is also being investigated for use in SLE without nephritis. Anecdotal experience with obinutuzumab by rheumatologists in treating SLE patients off-label suggests that this drug may be more clinically effective than rituximab due to improved CD20 tissue depletion. Deucravacitinib, an oral Tyk2 inhibitor, is in phase 3 study for SLE. Telitacicept, a TACI-Fc fusion protein that targets BLyS and APRIL growth factors for B cells, has been approved in China for use in patients with SLE, and global phase 3 studies are underway. Litifilimab, a drug in late-stage development with initial promising results, is a monoclonal antibody targeting BDCA2, a protein expressed on plasmacytoid dendritic cells, that suppresses activation of type I interferons and other proinflammatory cytokines. Initial results show achievement of remission, according to the Definition of Remission in Systemic Lupus Erythematosus (DORIS), from nonrandomized CD19-directed chimeric antigen receptor T-cell treatment, which has prompted interest in further study of this approach. Other cell-based therapeutic programs under investigation include both allogenic and autologous natural killer cell–based programs and mesenchymal embryonic stem cell transplantation for severe SLE.

■ GUIDELINES

Besides the recent EULAR (2013) and Pan-American League of Associations for Rheumatology (2018) guidelines for SLE and lupus nephritis, the ACR (2012) and the Kidney Disease Improving Global Outcomes (KDIGO) organization (2021) have published guidelines specifically for lupus nephritis. These guidelines address medication-based, lifestyle, and nonpharmacologic approaches, as well as infection screening and vaccination recommendations. The KDIGO guidelines include recommendations prior to the approval of belimumab and voclosporin. An important concept that the KDIGO group emphasizes is the unsettled question of duration of immunosuppression. The group notes that the risk of lupus relapse should be balanced with risk of adverse events secondary to immunosuppressive drugs. They suggest that for patients who have achieved a complete renal response and have no ongoing extrarenal SLE manifestations, the total duration of immunosuppression (initial plus maintenance) should not be <36 months. The EULAR guidelines provide consensus guidance for SLE management that combines evidence and expert opinion concerning the use of HCQ, glucocorticoids, immunosuppressive drugs, calcineurin inhibitors, and biologics, with suggestions for treatment strategies, response assessment, combination approaches, and therapy tapering (see Figures 368-4 and 368-5). Most of the concepts and strategies presented in these guidelines are described in the sections above.

DRUG-INDUCED LUPUS

Drug-induced lupus (DIL) is an autoimmune phenomenon in which a positive ANA and clinical features of SLE arise from a drug exposure. DIL accounts for ~6–12% of all lupus cases. Symptoms may include fever, rash, arthralgia, myalgia, and serositis. Antihistone antibodies are commonly associated with DIL, with dsDNA antibodies occurring less frequently. DIL tends to be less severe than SLE, rarely involving kidneys or brain, and symptoms typically resolve within several weeks after discontinuation of the offending agent.

The first agent identified to cause lupus symptoms was hydralazine in 1954. Over the years, >100 drugs have been identified as potential triggers for DIL. The highest incidence of DIL occurs with the antiarrhythmic procainamide, with risk reported to be as high as 30%. Other substances associated with DIL include the antihypertensives hydralazine and methyldopa; other antiarrhythmics; antibiotics including minocycline, isoniazid, rifampin, and nitrofurantoin; the antirheumatic drug sulfasalazine; anticonvulsants such as phenytoin and carbamazepine; antipsychotics including chlorpromazine and lithium; several ACE inhibitors and beta blockers; the diuretic hydrochlorothiazide; the antithyroid propylthiouracil; proton pump inhibitors; NSAIDs; and oral contraceptives. Biologic agents include antitumor necrosis factor α medications, particularly infliximab and etanercept, and IFN-α.

Diagnosis can be difficult, as DIL symptoms may start weeks to several months after drug initiation. Furthermore, skin biopsy may be indistinguishable from SLE, and laboratory characteristics including histone antibody may be seen in the majority of SLE cases. In DIL, the ANA pattern is most commonly homogeneous, and histone is positive in 75% of DIL cases. Cytopenias may be present but are typically mild. The presence of other autoantibodies may help to distinguish SLE from DIL, as these are rarely seen in DIL. The ANA test typically appears before symptoms, although many of the same medications will induce a positive ANA in patients who never develop symptoms. If DIL is suspected, these laboratory tests should help to inform the clinical picture to decide if medication discontinuation is appropriate. If symptoms resolve after medication discontinuation, this helps to confirm a diagnosis of DIL.

PREGNANCY AND REPRODUCTIVE HEALTH

FERTILITY

Lupus activity can affect the menstrual and ovulation cycle. Lower measures of ovarian reserve are seen in SLE patients compared to healthy controls. Women with SLE have a higher risk for infertility if they have active disease, comorbidities such as APS or nephritis, and prior exposure to treatments such as cyclophosphamide. Cyclophosphamide use may lead to amenorrhea, ovarian insufficiency, and early menopause.

CONTRACEPTION/HORMONE REPLACEMENT

Contraception and family planning should be discussed with all women of reproductive age with SLE. SLE patients with antiphospholipid antibodies and those with moderate-to-severe or active lupus should avoid estrogen compounds due to increased risk for thromboembolism. Progestin-only contraceptive options such as the levonorgestrel intrauterine device or etonogestrel implant, as well as emergency contraception, may be used safely in women with SLE. Estrogen compounds as a part of postmenopausal hormone replacement therapy should similarly be avoided in the presence of antiphospholipid antibodies.

PREGNANCY

SLE pregnancies are high risk. Disease activity in pregnancy is unpredictable, and the disease can flare during any trimester. Active disease at time of conception is associated with more flares during pregnancy. HCQ reduces the risk of flares during pregnancy and postpartum. Women with SLE are at a higher risk for adverse pregnancy outcomes including pregnancy loss, preterm birth, preeclampsia and pregnancy-induced hypertension, intrauterine growth restriction, cesarean section, and maternal death. Flaring or active disease increases the risk for these adverse outcomes. The presence of maternal hypertension, use of glucocorticoids during pregnancy, maternal renal disease, discontinuation of HCQ, non-White ethnicity, and lupus anticoagulant positivity are all associated with worse pregnancy outcomes. SLE patients at the highest risk for maternal death are those with pulmonary arterial hypertension, prior arterial thrombosis, and severe end-organ damage, and these patients should strongly consider avoiding pregnancy.

Outcomes are generally favorable for women with SLE with mild-to-moderate disease activity on pregnancy-compatible medications. Low-dose aspirin is used during pregnancy for preeclampsia prevention. HCQ is safe and the standard of care for all SLE pregnancies. Azathioprine and tacrolimus may be used safely in pregnancy for more severe disease and/or renal disease. MMF and methotrexate are teratogenic and should be avoided. Belimumab, anifrolumab, and voclosporin should currently be avoided due to lack of safety data. Patients who get pregnant on these medications may be switched to azathioprine or tacrolimus. Rituximab should be discontinued at conception but may be used during pregnancy in organ-threatening disease. Prednisone may be used for disease flares and can suppress disease activity, but the dose should be tapered to the lowest effective dose by adding pregnancy-compatible immunosuppressant medications. Warfarin is teratogenic; low-molecular-weight heparin is used in SLE pregnancies with APS or prior pregnancy morbidity.

NEONATAL LUPUS

SS-A (Ro) antibodies may cross the placenta into the fetal circulation and put the infant at risk for neonatal lupus. Neonatal lupus can consist of a transient cutaneous rash or lab abnormalities that present after birth or, less commonly, congenital heart block that presents in utero. The presence of Ro antibodies, particularly in high titer, should prompt referral to high-risk obstetricians who can perform fetal echocardiography to screen for congenital heart block. The use of fluorinated steroids may be considered if first- or second-degree heart block is detected. The use of HCQ in pregnancy significantly reduces the risk of congenital heart block in women who have had a prior fetus with congenital heart block.

OUTCOMES

Measuring SLE outcomes both in clinical trials and in clinical practice is complex given the multisystem nature of the disease. Various aspects of disease require unique treatment approaches and often at differing rates, which makes measuring response challenging. Modern SLE treatment outcomes include improvement in inflammatory type 1 and chronic type 2 symptoms, quality of life, and the prevention of frailty and damage.

New drugs and interventions are constantly being explored as potential treatments for SLE in clinical trials. Goals of clinical trials for novel therapies include improved disease activity and acceptable safety. Long-term extension studies focus on long-term safety and maintenance of improved disease activity. The Systemic Lupus Erythematosus Responder Index (SRI) and the British Isles Lupus Assessment Group (BILAG)-based Composite Lupus Assessment (BICLA) are validated instruments for use in clinical trials. The Cutaneous Lupus Disease Area and Severity Index (CLASI) is also utilized specifically for trials involving cutaneous lupus. Several instruments are utilized in clinical trials to assess quality of life and fatigue, although few are specific to SLE.

TREAT-TO-TARGET APPROACHES

Treat-to-target goals for clinical trials and clinical practice have been developed and validated, including the definition of a lupus low disease activity state (LLDAS) and remission (DORIS). Both LLDAS and DORIS remission scoring utilize the SLEDAI-2K instrument, a validated instrument that measures SLE disease activity. Achievement of remission by these measures is defined by low disease activity (LLDAS) or no disease activity (DORIS), with minimal use of steroids and stable use of antimalarials, immunosuppressants, and/or biologics and by low levels (LLDAS) or no evidence of disease activity (DORIS) as assessed by the evaluating physician. LLDAS has been shown to be attainable in up to 80% of patients, and the achievement of sustained LLDAS for 2 more years is associated with significantly less overall damage.

DAMAGE

Organ or structural damage occurs in SLE due to the use of medications to treat the disease or because of persistent disease activity. Use of glucocorticoids is associated with the development of irreversible organ damage that is independent of disease activity. Cataracts, osteoporosis-associated fractures, avascular necrosis, and diabetes mellitus are the complications from glucocorticoid treatment that are most commonly observed. Development of damage accrues in a dose-dependent manner, but even low doses of 5 mg of prednisone or equivalent appear to be associated with damage accrual. Accelerated atherosclerosis is a major cause of death after 5 years of SLE disease duration. Increased myocardial infarction and cerebrovascular event rates are due to both traditional cardiovascular risk factors and non-traditional factors including presence of antiphospholipid antibodies, increased disease activity, low C3 levels, high glucocorticoid dose, and high homocysteine and leptin levels. Statin use has been demonstrated to reduce all-cause mortality in SLE patients with renal transplantation. Studies of malignancy rates in SLE have been inconsistent; however, recent meta-analyses suggest an increased risk of overall cancer and cancer-related death in SLE patients. Seventeen site-specific cancers noted include digestive cancers (esophagus, colon, anus, hepatobiliary, liver, and pancreas), hematologic cancers (lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma, leukemia, and multiple myeloma), and cancer involving lung, larynx, cervix, vagina/vulva, kidney, bladder, skin, and thyroid.

FURTHER READING

Aringer M et al: 2019 European League Against Rheumatism/American College of Rheumatology classification for systemic lupus erythematosus. Ann Rheum Dis 78:1151, 2019.

Aringer M et al: A glimpse into the future of systemic lupus erythematosus. Ther Adv Musculoskelet Dis 14:1759720X221086719, 2022.

Barber MRW et al: Global epidemiology of systemic lupus erythematosus. Nat Rev Rheumatol 17: 515, 2021.

Fanouriakis A et al: EULAR recommendations for the management of systemic lupus erythematosus: 2023 update. Ann Rheum Dis 83:15, 2024.

Izmirly PM et al: Prevalence of systemic lupus erythematosus in the United States: Estimates from a meta-analysis of the centers for Disease Control and Prevention National Lupus Registries. Arthritis Rheumatol 73:991, 2021.

Morand EF et al: Advances in the management of systemic lupus erythematosus. BMJ 383:e073980, 2023.

Petri M et al: Derivation and validation of Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 64:2677, 2012.

Pisetsky DS et al: A novel system to categorize the symptoms of systemic lupus erythematosus. Arthritis Care Res (Hoboken) 71:735, 2019.

Shumilova A, Vital EM: Musculoskeletal manifestations of systemic lupus erythematosus. Best Pract Res Clin Rheumatol 22:101859, 2023.

Vale ECSD, Garcia LC: Cutaneous lupus erythematosus: A review of etiopathogenic, clinical, diagnostic and therapeutic aspects. An Bras Dermatol 98:355, 2023.

369

Antiphospholipid Syndrome

Haralampos M. Moutsopoulos,
Maria G. Tektonidou

DEFINITIONS

Antiphospholipid syndrome (APS) is an autoantibody-mediated acquired thrombophilia characterized by recurrent arterial or venous thrombosis, microvascular manifestations, and/or pregnancy morbidity. It affects primarily young females. APS may occur alone (primary) or in association with other autoimmune diseases, mainly systemic lupus erythematosus (SLE). Catastrophic APS (CAPS) is a rare, life-threatening, rapidly progressive thromboembolic disease involving simultaneously three or more organs.

The major autoantibodies detected in APS patients are directed against phospholipid-binding plasma proteins such as β_2-glycoprotein I (β_2GPI) and prothrombin. The plasma protein β_2GPI is a 43-kDa plasma apolipoprotein, which consists of 326 amino acids arranged in five domains (I through V). The presence of anti–domain I IgG antibodies has been associated with increased thrombotic risk. Recent evidence from animal studies has shown that the antithrombotic function of β_2GPI is dependent on its fifth domain (domain V). Another group of antibodies against phospholipid-binding proteins, termed *lupus anticoagulant* (LA), prolongs clotting time in vitro, which is not corrected by adding normal plasma.

Anticardiolipin and anti-β_2GPI IgG/IgM antibodies are measured by standardized enzyme-linked immunosorbent assays (ELISA); β_2GPI in combination with cardiolipin or β_2GPI in the absence of cardiolipin serve as target antigens, respectively. LA testing includes a three-step procedure: (1) screening including testing with activated partial thromboplastin time (aPTT) and dilute Russel viper venom test (DRVVT); (2) mixing study; and (3) confirmation. In patients receiving anticoagulation treatment, LA test positivity should be interpreted with caution. In patients with features highly reminiscent of APS in the absence of classical antiphospholipid antibodies (aPL), testing for antiphosphatidylserine/prothrombin antibodies may have a valuable diagnostic role.

EPIDEMIOLOGY

The incidence of APS is estimated to be 1–2 cases per 100,000 persons per year. The prevalence of APS in the general population is estimated to be 40–50 per 100,000. APL antibodies occur in 1–5% of the general population. Their prevalence increases with age; however, it is questionable whether they are able to induce thrombotic events in elderly individuals. Moreover, one-third of patients with SLE **(Chap. 368)** possess these antibodies and 10–15% of them develop APS.

PATHOGENESIS

The initiating events for the induction of antibodies to phospholipid-binding proteins seem to be endothelial injury induced by infections, oxidative stress, and major physical stresses such as surgery or trauma in an appropriate genetic background, given the previously demonstrated associations with alleles within the human leukocyte antigen (HLA) locus. These contributors seem to induce increased apoptosis of the vessel endothelial cells, autoantigen presentation, and subsequent initiation of autoimmune response. The binding of aPL to the disrupted endothelial cells leads to a proinflammatory and prothrombotic state involving monocytes and platelets activation, adhesion molecules and proinflammatory cytokines production, complement and neutrophil activation, neutrophil extracellular trap (NET) formation, and thrombus formation. Recently, type I interferon pathway activation and an imbalance between type I and III interferons have been proposed as potential mechanisms of thrombotic events and APS-related obstetric complications.

CLINICAL MANIFESTATIONS AND LABORATORY FINDINGS

Clinical manifestations include venous or arterial thrombosis, microvascular events, and/or pregnancy morbidity **(Table 369-1)**. Deep venous thrombosis occurs primarily in the lower extremities and often causes pulmonary emboli. Thrombosis of the pulmonary arteries leads to pulmonary hypertension and thrombosis of hepatic veins to Budd-Chiari syndrome. Cerebral venous thrombosis presents with signs and symptoms of intracranial hypertension and focal neurologic deficits. Other venous thrombosis manifestations include retinal vein thrombosis and renal, mesenteric, or splanchnic vein thrombosis. Arterial thrombosis affects more commonly the arteries of the brain and is manifested as transient ischemic attack, stroke, migraines, or cognitive dysfunction. Peripheral artery thrombosis presents with acute limb ischemia, ischemic leg ulcers, or digital gangrene. Thrombosis of other arteries leads to myocardial infarction, retinal artery occlusion, renal artery stenosis, and infarcts of spleen, liver, and adrenals.

Microvascular manifestations include livedo reticularis, which consists of a mottled reticular vascular pattern that appears as a lace-like, purplish discoloration of the skin; livedo racemosa, characterized by a broken, irregular, and asymmetric pattern; and livedoid vasculopathy, which presents with painful papules or lower limb ulcers. Another microvascular feature of APS is the so-called aPL-nephropathy, manifested with hypertension, proteinuria (mild to nephrotic range), hematuria, and mild renal insufficiency. Histologically, the acute phase is characterized by thrombotic microangiopathy lesions in the glomerular capillaries and/or arterioles. In a chronic phase, fibrous intima hyperplasia, fibrous and/or fibrocellular arteriolar occlusions, and focal cortical atrophy are present (Table 369-1). Additional rare microvascular manifestations are pulmonary hemorrhage, myocardial disease confirmed by magnetic resonance imaging or histologically, and adrenal hemorrhage.

Pregnancy morbidity manifests with prefetal (<10 weeks of gestation) or fetal death, preeclampsia, eclampsia, and placental insufficiency with intrauterine fetal growth restriction, abnormal umbilical arterial Doppler, and/or infarctions of the placenta

Cardiac valve manifestations include valve thickening or vegetations. Libman-Sacks endocarditis consists of small (usually <1 cm) valve vegetations, histologically characterized by organized platelet-fibrin microthrombi surrounded by growing fibroblasts and macrophages. Premature atherosclerosis has been also recognized as a feature of APS. Thrombocytopenia and autoimmune hemolytic anemia are the main hematologic manifestations of APS. Musculoskeletal

TABLE 369-1 Clinical Features of Antiphospholipid Syndrome

MANIFESTATION	%
Venous Thrombosis and Skin Manifestations	
Deep-vein thrombosis	39
Livedo reticularis	24
Pulmonary embolism	14
Superficial thrombophlebitis	12
Thrombosis in various other sites	11
Arterial Thrombosis and Cardiac Manifestations	
Stroke	20
Cardiac valve thickening/dysfunction and/or Libman-Sacks vegetations	14
Transient ischemic attack	11
Myocardial ischemia (infarction or angina) and coronary bypass graft thrombosis	10
Leg ulcers and/or digital gangrene	9
Arterial thrombosis in the extremities	7
Retinal artery thrombosis/amaurosis fugax	7
Ischemia of visceral organs or avascular necrosis of bone	6
Multi-infarct dementia	3
Neurologic Manifestations of Uncertain Etiology	
Migraine	20
Epilepsy	7
Chorea	1
Cerebellar ataxia	1
Transverse myelopathy	0.5
Renal Manifestations Due to Various Reasons (Renal Artery/Renal Vein/Glomerular Thrombosis, Fibrous Intima Hyperplasia)	**3**
Musculoskeletal Manifestations	
Arthralgias	39
Arthritis	27
Obstetric Manifestations (Referred to the Number of Pregnancies)	
Preeclampsia	10
Eclampsia	4
Fetal Manifestations (Referred to the Number of Pregnancies)	
Early fetal loss (<10 weeks)	35
Late fetal loss (≥10 weeks)	17
Premature birth among the live births	11
Hematologic Manifestations	
Thrombocytopenia	30
Autoimmune hemolytic anemia	10

Source: Adapted from R Cervera et al: Arthritis Rheum 46:1019, 2002.

manifestations may also occur in APS including arthralgias/arthritis, avascular bone necrosis, bone marrow necrosis, nontraumatic fractures, and osteoporosis.

■ DIAGNOSIS AND DIFFERENTIAL DIAGNOSIS

The diagnosis of APS should be seriously considered in cases of thrombosis, cerebral vascular accidents in individuals <55 years of age, or pregnancy morbidity, in the presence of livedo reticularis or thrombocytopenia. In these cases, aPL antibodies should be measured. According to the 2023 American College of Rheumatology/European Alliance of Associations for Rheumatology (ACR/EULAR) classification criteria for APS, the presence of at least 3 points from clinical domains and 3 points from laboratory domains is required to classify APS. Clinical manifestations with other equally or more likely explanations than APS will not be counted. Clinical domains include (1) venous thromboembolism; (2) arterial thrombosis; (3) microvascular events; (4) obstetric complications; (5) cardiac valve disease; and (6) hematologic manifestations, as shown in **Table 369-2**. Laboratory domains include (1) LA, (2) anticardiolipin (aCL), and/or (3) anti-β_2GPI antibodies, at moderate (40-79 units) or high titers (≥ 80 units) on two occasions 12 weeks apart.

Differential diagnosis is based on the exclusion of other inherited or acquired causes of thrombophilia **(Chap. 121)**, Coombs-positive hemolytic anemia **(Chap. 105)**, and thrombocytopenia **(Chap. 120)**. Livedo reticularis with or without a painful ulceration on the lower extremities may be also a manifestation of disorders affecting (1) the vascular wall, such as atherosclerosis, polyarteritis nodosa, SLE, cryoglobulinemia, and lymphomas; or (2) the vascular lumen, such as

TABLE 369-2 2023 American College of Rheumatology/European Alliance of Associations for Rheumatology (ACR/EULAR) Classification Criteria for Antiphospholipid Syndrome

Clinical Domains	
DOMAINS/CRITERIA	**POINTS**
Venous thromboembolism (VTE)	
• With high-risk VTE profile	1
• Without a high-risk VTE profile	3
Microvascular events	
Suspected (any of below)	2
• Livedo racemosa (exam)	
• Livedoid vasculopathy lesions (exam)	
• Acute/chronic aPL-nephropathy (exam or lab)	
• Pulmonary hemorrhage (symptoms and imaging)	
Established (any of below)	5
• Livedoid vasculopathy (pathology)	
• Acute/chronic aPL-nephropathy (pathology)	
• Pulmonary hemorrhage (BAL or pathology)	
• Myocardial disease (imaging or pathology)	
• Adrenal hemorrhage (imaging or pathology)	
Cardiac valve disease	
• Thickening	2
• Vegetation	4
Arterial thrombosis	
• With high-risk CVD profile	2
• Without high-risk CVD profile	4
Obstetric complications	
• >3 consecutive prefetal (<10 w) and/or early fetal deaths (10w0d–15w6d)	1
• Fetal death (16w0d–33w6d) in the absence of preeclampsia (PEC) with severe features or placental insufficiency (PI) with severe features	1
• PEC with severe features (<34w0d) or PI with severe features (<34w0d) with/without fetal death	3
• PEC with severe features (<34w0d) and PI with severe features (<34w0d) with/without fetal death	4
Hematologic manifestations	
• Thrombocytopenia (otherwise unexplained lowest platelet count ever between 20 and 130 × 10^9/L)	2
Laboratory Domains (aPL)	
DOMAINS/CRITERIA	**POINTS**
Lupus anticoagulant (LA) test	
• One time positive	1
• Persistent	5
Anticardiolipin (aCL) and/or anti-β_2GPI antibodies (persistent)	
• Moderate (40–79 units) or high (≥80 units) IgM aCL and/or anti-β_2GPI	1
• Moderate (40–79 units) IgG aCL and/or anti-β_2GPI	4
• High (≥80 units) positive IgG aCL or anti-β_2GPI	5
• High (≥80 units) IgG aCL and anti-β_2GPI	7

Abbreviations: aPL, antiphospholipid antibody; anti-β_2GPI, anti-β_2-glycoprotein I; BAL, bronchoalveolar lavage; CVD, cardiovascular disease; Exam, physical examination; Lab, laboratory tests.

Source: Modified with permission from R Zuo et al: The 2023 ACR/EULAR Antiphospholipid Syndrome Classification Criteria. Arthritis Rheumatol 75:1687, 2023.

myeloproliferative disorders, hypercholesterolemia, or other causes of thrombophilia.

Diagnosis of CAPS should be considered in patients with mainly microvascular thrombotic disease involving multiple organs in a short time period. A patient is classified as definite CAPS if all four of the following criteria are fulfilled and as probable CAPS when a combination of these criteria is present: (1) thrombosis in three or more organs/systems; (2) development in less than a week; (3) histologic evidence of small-vessel thrombosis in at least one organ; and (4) aPL presence. Because thrombotic microangiopathy (microangiopathic hemolytic anemia and severe thrombocytopenia) is a typical finding in CAPS, the differential diagnosis includes thrombotic thrombocytopenic purpura, hemolytic-uremic syndrome, disseminated intravascular coagulation, and heparin-induced thrombocytopenia.

TREATMENT

Antiphospholipid Syndrome

It has been increasingly appreciated that the risk of thrombotic and obstetric events is closely related to the underlying aPL profile. The latter depends on the type of autoantibodies (IgG high risk vs IgM low risk), the number of aPL antibodies (simultaneous presence of two or three classical autoantibodies vs a single antibody denotes a higher risk profile), their titer (moderate-high titer vs low), and the persistence of aPL positivity in repeated measurements.

Following the first thrombotic event, APS patients should be placed on vitamin K antagonists (VKAs) for life, aiming to achieve an international normalized ratio (INR) ranging from 2.0 to 3.0 in case of an unprovoked venous thrombosis. For patients with arterial thrombosis, the corresponding INR target should be 3.0–4.0 or 2.0–3.0 with or without low-dose aspirin (LDA, 75–100 mg daily), depending on the thrombotic/hemorrhagic patient profile. Administration of direct oral anticoagulants (DOACs) recently has been shown to increase the risk of arterial events, especially in patients with triple positivity or previous arterial thrombosis. A recent meta-analysis of four open-label randomized controlled trials showed that prior thrombosis type (arterial vs venous) did not affect the increased odds of arterial thrombosis associated with DOACs compared to VKAs in APS. In pregnant women with a history of obstetric APS, combination treatment with LDA and prophylactic dose of low-molecular-weight heparin (LMWH) is recommended, whereas in cases of thrombotic APS, LDA plus therapeutic LMWH dose should be administered. When recurrent obstetric complications occur despite standard treatment, increasing the LMWH dose (from prophylactic to therapeutic) or administering oral hydroxychloroquine 400 mg/d or low-dose prednisolone in the first trimester are alternative options.

For asymptomatic individuals or SLE patients with a high-risk aPL profile and no evidence of a previous thrombotic event or pregnancy morbidity, prophylactic treatment with LDA is recommended. In nonpregnant women with a history of APS-related obstetric complications, independently of the presence of underlying SLE diagnosis, treatment with LDA seems to reduce the risk of a subsequent thrombotic event.

Patients with CAPS should be treated with combination therapy including glucocorticoids, heparin, and plasma exchange or intravenous immunoglobulin (IVIG) together with appropriate management of triggering events such as infections. For refractory CAPS, B-cell depletion (e.g., with rituximab) or complement inhibition (e.g., with eculizumab) therapies are alternative options.

■ FURTHER READING

Barbhaiya M et al: 2023 ACR/EULAR antiphospholipid syndrome classification criteria. Arthritis Rheumatol 75:1687, 2023.

Knight JS et al: Antiphospholipid syndrome: Advances in diagnosis, pathogenesis, and management. BMJ 380:e069717, 2023.

Tektonidou MG et al: EULAR recommendations for the management of antiphospholipid syndrome in adults. Ann Rheum Dis 78:1296, 2019.

370 Rheumatoid Arthritis

Ankoor Shah, E. William St. Clair

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic inflammatory disease characterized by a symmetric, erosive polyarthritis. It is the most common form of chronic inflammatory arthritis. Since persistently active RA often results in articular cartilage and bone destruction and functional disability, it is vital to diagnose and treat this disease early and aggressively before damage ensues. Although predominately affecting joints, RA is a systemic disease that may lead to a variety of extraarticular manifestations, including fatigue, subcutaneous nodules, lung involvement, pericarditis, peripheral neuropathy, vasculitis, and hematologic abnormalities.

Insights gained over the past two decades have revolutionized the contemporary paradigms for the diagnosis and management of RA. Testing for serum antibodies to anti-citrullinated protein antibodies (ACPA) and rheumatoid factor continues to be valuable in the diagnostic evaluation of patients with suspected RA, and these autoantibodies serve as biomarkers of prognostic significance. Advances in imaging modalities assist clinical decision-making by improving the detection of joint inflammation and monitoring the progression of damage. The science of RA has taken major leaps forward by illuminating new disease-related genes, environmental interactions, and the molecular components and pathways of disease pathogenesis. The relative contribution of these cellular and inflammatory mediators in disease pathogenesis has been further brought to light by the efficacy of the approved biologic and targeted synthetic disease-modifying therapies. Despite these major strides, incomplete understanding of the initiating events of RA and the factors perpetuating the chronic inflammatory response remains a barrier to finding a cure.

The past 20 years have witnessed a remarkable improvement in the outcomes of RA. The crippling arthritis of years past is encountered much less frequently today. Much of this progress can be traced to the expanded therapeutic armamentarium and the renewed sense of urgency to identify patients earlier in the disease process, enabling prompt treatment intervention. This shift in treatment strategy dictates a new mindset for primary care practitioners—namely, one that demands early referral of patients with inflammatory arthritis to a rheumatologist for confirmation of the diagnosis and initiation of disease-modifying therapy, since delays in starting effective treatment are associated with worse outcomes. Reaching a clinical state of low disease activity and remission is now an achievable goal for most patients.

CLINICAL FEATURES

The incidence of RA increases between 25 and 55 years of age, after which it plateaus until the age of 75 and then decreases. The presenting symptoms are typically related to inflammation of the joints, tendons, and bursae. Patients often complain of early morning joint stiffness lasting >1 h that eases with physical activity. The earliest involved joints are mostly the small joints of the hands and feet. The initial pattern of joint involvement may be monoarticular, oligoarticular (≤4 joints), or polyarticular (>5 joints), usually in a symmetric distribution. Some patients with inflammatory arthritis will present with too few affected joints to be classified as having RA—so-called undifferentiated inflammatory arthritis. Those with an undifferentiated arthritis who are most likely to be diagnosed later with RA have a higher number of tender and swollen joints, test positive for serum rheumatoid factor (RF) or ACPA, and have higher scores for physical disability.

In established RA, the most frequently involved joints are the wrists and metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joints **(Fig. 370-1)**. Distal interphalangeal (DIP) joint involvement may occur in patients with RA, but it is usually due to coexistent osteoarthritis. Flexor tendon tenosynovitis is a frequent hallmark of RA and leads to decreased range of motion, reduced grip strength,

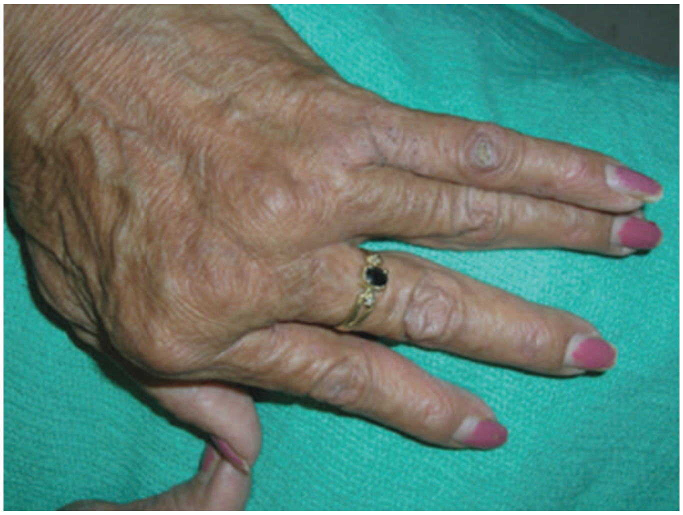

FIGURE 370-1 Metacarpophalangeal joint swelling and subluxation. *(Reproduced with permission from RP Usatine, MA Smith, EJ Mayeaux: The Color Atlas and Synopsis of Family Medicine, 3rd ed. New York, McGraw Hill, 2019; Fig. 97.5.)*

and "trigger" fingers. Flexor tendon involvement may also lead to tendon rupture, most commonly the flexor pollicis longus. Progressive destruction of the joints and soft tissues may result in chronic, irreversible deformities. Ulnar deviation is a common deformity of long-standing RA that is caused by subluxation of the MCP joints, with subluxation, or partial dislocation, of the proximal phalanx to the volar side of the hand. Hyperextension of the PIP joint with flexion of the DIP joint ("swan-neck deformity"), flexion of the PIP joint with hyperextension of the DIP joint ("boutonnière deformity"), and subluxation of the first MCP joint with hyperextension of the first interphalangeal (IP) joint ("Z-line deformity") also may result from damage to the tendons, joint capsule, and other soft tissues. Inflammation about the ulnar styloid and tenosynovitis of the extensor carpi ulnaris may cause subluxation of the distal ulna, resulting in a "piano-key movement" of the ulnar styloid. Although metatarsophalangeal (MTP) joint involvement is an early feature of disease, chronic inflammation of the ankle and midtarsal regions usually comes later and may lead to pes planovalgus ("flat feet"). Large joints, including the knees and shoulders, are often affected in established disease and may remain asymptomatic for many years after onset.

Atlantoaxial involvement of the cervical spine is clinically noteworthy because of its potential to cause compressive myelopathy and neurologic dysfunction. Neurologic manifestations are rarely a presenting sign or symptom of atlantoaxial disease, but they may slowly evolve over time with progressive instability of C1 on C2. The prevalence of atlantoaxial subluxation has been declining in recent years and occurs now in <10% of patients. Unlike the spondyloarthritides **(Chap. 374)**, RA rarely affects the thoracic and lumbar spine.

Extraarticular manifestations may develop during the clinical course of RA in up to 40% of patients, even prior to the onset of arthritis **(Fig. 370-2)**. Patients most likely to develop extraarticular disease have a history of cigarette smoking, have early onset of significant physical disability, and test positive for serum RF or ACPA. Subcutaneous nodules, Sjögren's disease, interstitial lung disease (ILD), pulmonary nodules, and anemia are among the most frequently observed extraarticular manifestations. Recent studies have shown a decrease in the incidence and severity of at least some extraarticular manifestations, particularly Felty's syndrome and vasculitis.

The most common systemic and extraarticular features of RA are described in more detail in the sections below.

CONSTITUTIONAL

These signs and symptoms include weight loss, fever, fatigue, malaise, depression, and in the most severe cases, cachexia; they generally reflect a high degree of inflammation and may even precede the onset of joint symptoms. In general, the presence of a fever of >38.3°C (101°F) at any time during the clinical course should raise suspicion of systemic vasculitis (see below) or infection.

NODULES

Subcutaneous nodules have been reported to occur in 30–40% of patients and more commonly in those with the highest levels of disease activity, who carry the disease-related shared epitope (SE) (see below), who have a positive test for serum RF, and who show radiographic evidence of joint erosions. However, more recent cohort studies suggest a declining prevalence of subcutaneous nodules, perhaps related to early and more aggressive disease-modifying therapy. When palpated, the nodules are generally firm; nontender; and adherent to periosteum, tendons, or bursae; they develop in areas of the skeleton subject to repeated trauma or irritation such as the forearm, sacral prominences, and Achilles tendon. They may also occur in the lungs, pleura, pericardium, and peritoneum. Nodules are typically benign, although they can be associated with infection, ulceration, and gangrene. Accelerated growth of smaller nodules may occur in up to 10% of patients taking long-term methotrexate, although the mechanisms behind this phenomenon are unclear.

SJÖGREN'S SYNDROME

Secondary Sjögren's syndrome, or Sjögren's disease **(Chap. 373)**, is defined by the presence of either keratoconjunctivitis sicca (dry eyes) or xerostomia (dry mouth) in association with another connective tissue disease, such as RA. Approximately 10% of patients with RA have secondary Sjögren's syndrome.

PULMONARY

Pleuritis, the most common pulmonary manifestation of RA, may result in pleuritic chest pain and dyspnea, as well as a finding of a pleural friction rub and radiographic evidence of a pleural effusion. Pleural effusions tend to be exudative with increased numbers of monocytes and neutrophils. ILD may also occur in patients with RA and is heralded by symptoms of dry cough and progressive shortness of breath. ILD can be associated with cigarette smoking and is generally found in patients with higher disease activity, although it may be diagnosed in up to 3.5% of patients prior to the onset of joint symptoms. Recent studies have shown the overall prevalence of ILD in RA to be as high as 12%. Diagnosis is readily made by high-resolution chest computed tomography (CT) scan, which shows infiltrative opacification, or ground-glass opacities, in the periphery of both lungs. Usual interstitial pneumonia (UIP) and nonspecific interstitial pneumonia (NSIP) are the main histologic and radiologic patterns of ILD. UIP causes progressive scarring of the lungs that, on chest CT scan, produces honeycomb changes in the periphery and lower portions of the lungs. In contrast, the most common radiographic changes in NSIP are relatively symmetric and bilateral ground-glass opacities with associated fine reticulations, volume loss, and traction bronchiectasis. In both cases, pulmonary function testing shows a restrictive pattern (e.g., reduced total lung capacity) and a reduced diffusing capacity for carbon monoxide (DL_{CO}). The presence of ILD confers a poor prognosis. The prognosis of ILD in RA, however, is not quite as poor as that of idiopathic pulmonary fibrosis (e.g., usual interstitial pneumonitis) and responds better to immunosuppressive therapy **(Chap. 304)**. Pulmonary nodules are also common in patients with RA and may be solitary or multiple. Caplan's syndrome is a rare subset of pulmonary nodulosis characterized by the development of nodules and pneumoconiosis following silica exposure. Less commonly, RA may be associated with respiratory bronchiolitis and bronchiectasis.

CARDIAC

The most frequent site of cardiac involvement in RA is the pericardium. However, clinical manifestations of pericarditis occur in <10% of patients with RA despite pericardial involvement being detectable in nearly one-half of cases by echocardiogram or at autopsy. Up to 20% of patients with RA may have asymptomatic pericardial effusions on echocardiography. Cardiomyopathy, another clinically important manifestation of RA, may result from necrotizing or granulomatous myocarditis, coronary artery disease, or diastolic dysfunction. This

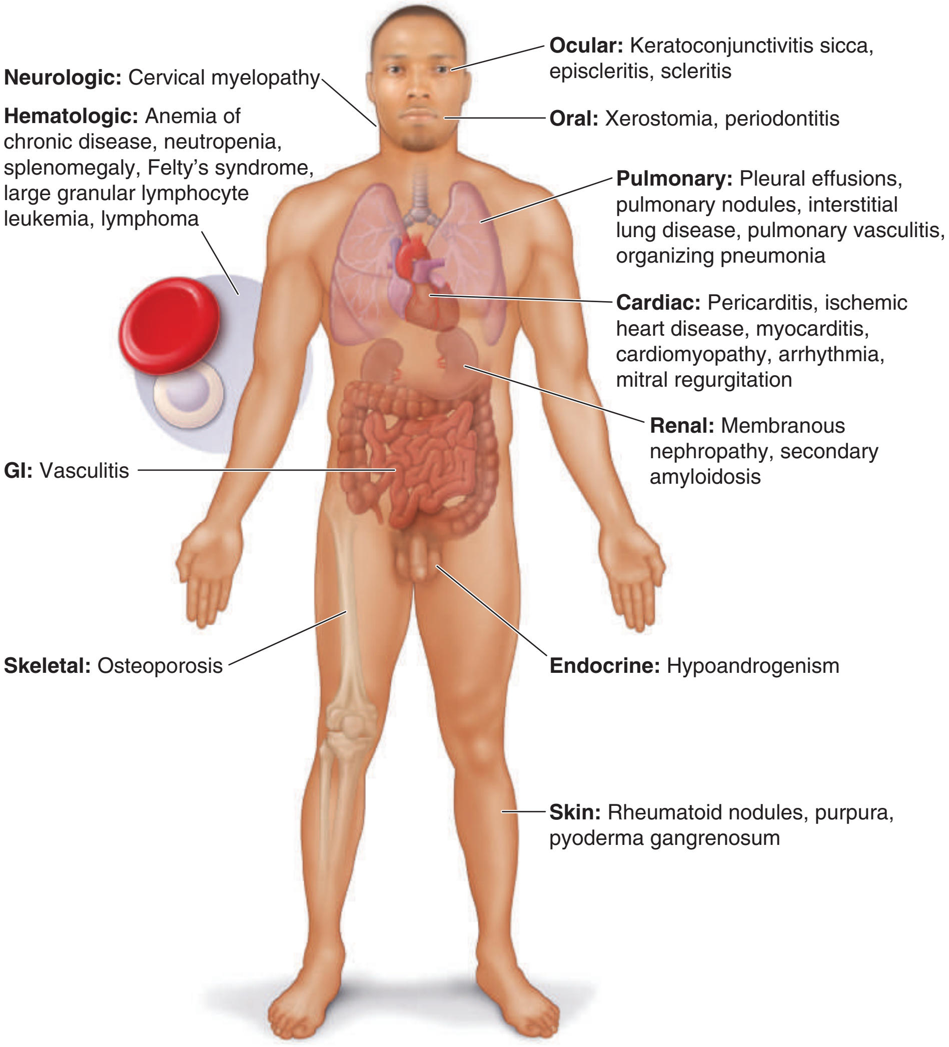
Neurologic: Cervical myelopathy
Hematologic: Anemia of chronic disease, neutropenia, splenomegaly, Felty's syndrome, large granular lymphocyte leukemia, lymphoma
GI: Vasculitis
Skeletal: Osteoporosis
Ocular: Keratoconjunctivitis sicca, episcleritis, scleritis
Oral: Xerostomia, periodontitis
Pulmonary: Pleural effusions, pulmonary nodules, interstitial lung disease, pulmonary vasculitis, organizing pneumonia
Cardiac: Pericarditis, ischemic heart disease, myocarditis, cardiomyopathy, arrhythmia, mitral regurgitation
Renal: Membranous nephropathy, secondary amyloidosis
Endocrine: Hypoandrogenism
Skin: Rheumatoid nodules, purpura, pyoderma gangrenosum

FIGURE 370-2 Extraarticular manifestations of rheumatoid arthritis.

involvement, too, may be subclinical and only identified by echocardiography or cardiac magnetic resonance imaging (MRI). Rarely, the heart muscle may contain rheumatoid nodules or be infiltrated with amyloid.

■ VASCULITIS

Rheumatoid vasculitis **(Chap. 375)** typically occurs in patients with long-standing disease, a positive test for serum RF or ACPA, and hypocomplementemia. The overall incidence has decreased significantly in the past decade to <1% of patients. The cutaneous signs vary and include petechiae, purpura, digital infarcts, gangrene, livedo reticularis, and in severe cases large, painful lower extremity ulcerations. Vasculitic ulcers, which may be difficult to distinguish from those caused by venous insufficiency, may be treated successfully with immunosuppressive agents including cytotoxic treatment and skin grafting in severe cases. Sensorimotor polyneuropathies, such as mononeuritis multiplex, may occur in association with systemic rheumatoid vasculitis; they usually present with new onset of numbness, tingling, or focal muscle weakness.

■ HEMATOLOGIC

A normochromic, normocytic anemia often develops in patients with RA and is the most common hematologic abnormality. The degree of anemia parallels the degree of inflammation, correlating with the levels of serum C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR). Platelet counts may also be elevated in RA as an acute-phase reactant; immune-mediated thrombocytopenia is rare.

Felty's syndrome is defined by the clinical triad of neutropenia, splenomegaly, and nodular RA and is seen in <1% of patients, although its incidence appears to be declining in the face of more aggressive treatment of the joint disease. It typically occurs in the late stages of severe RA and is more common in whites than other racial groups. T-cell large granular lymphocyte leukemia (T-LGL) may have a similar clinical presentation and often occurs in association with RA. T-LGL is characterized by a chronic, indolent clonal growth of LGL cells, leading to neutropenia and splenomegaly. As opposed to Felty's syndrome, T-LGL may develop early in the course of RA. Leukopenia apart from these disorders is uncommon and most often a side effect of drug therapy.

■ LYMPHOMA

Large cohort studies have shown a two- to fourfold increased risk of lymphoma in RA patients compared with the general population. The most common histopathologic type of lymphoma is a diffuse large B-cell lymphoma. The risk of developing lymphoma increases if the patient has high levels of disease activity or Felty's syndrome.

■ ASSOCIATED CONDITIONS

In addition to extraarticular manifestations, several conditions associated with RA contribute to disease morbidity and mortality rates. They are worthy of mention because they affect chronic disease management.

Cardiovascular Disease The most common cause of death in patients with RA is cardiovascular disease. The incidence of coronary artery disease and carotid atherosclerosis is higher in RA patients than in the general population even when controlling for traditional cardiac risk factors, such as hypertension, obesity, hypercholesterolemia, diabetes, and cigarette smoking. Furthermore, congestive heart failure (including both systolic and diastolic dysfunction) occurs at an approximately twofold higher rate in RA than in the general population. The presence of elevated serum inflammatory markers appears to confer an increased risk of cardiovascular disease.

Osteoporosis Osteoporosis is more common in patients with RA than an age- and sex-matched population, with an incidence rate of nearly double that of the healthy population and a prevalence of approximately one-third in postmenopausal women with RA. There is also an increased risk of fragility fracture, with a greater risk among women. The inflammatory milieu of the joint promotes generalized bone loss by activating osteoclasts. Both trabecular and cortical bone are affected by the inflammatory response, with cortical sites more susceptible to bone loss. Chronic use of glucocorticoids and disability-related immobility also contribute to osteoporosis. Hip fractures are more likely to occur in patients with RA and are significant predictors of increased disability and mortality rate in this disease.

EPIDEMIOLOGY

RA affects ~0.5–1% of the adult population worldwide, although the majority of epidemiologic studies have been done in Western countries. There is evidence that the overall incidence of RA has been decreasing in recent decades, whereas the prevalence has remained the same because individuals with RA are living longer. The incidence and prevalence of RA vary based on geographic location, both globally and among certain ethnic groups within a country **(Fig. 370-3)**. For example, the indigenous Yakima, Pima, and Chippewa tribes of North America have reported prevalence rates in some studies of nearly 7%. In contrast, many population studies from Africa and Asia show lower prevalence rates for RA in the range of 0.2–0.4%.

Like many other autoimmune diseases, RA occurs more commonly in females than in males, with a 2–3:1 ratio. Interestingly, studies of RA from some of the Latin American and African countries show an even greater predominance of disease in females compared to males, with ratios of 6–8:1. Given this preponderance of females, various theories have been proposed to explain the possible role of estrogen in disease pathogenesis. Broadly speaking, most of the theories center on the role of estrogens and androgens in enhancing and suppressing the immune response, respectively. However, estrogens have both stimulatory and inhibitory effects on the immune system, and their role, if any, on the development of RA is unknown.

GENETIC CONSIDERATIONS

It has been recognized for >30 years that genetic factors contribute to the occurrence of RA as well as to its severity. The likelihood that a first-degree relative of a patient will share the diagnosis of RA is 2–10 times greater than in the general population. There remains, however, some uncertainty in the extent to which genetics plays a role in the causative mechanisms of RA. Heritability estimates range from 40 to 60%, although this may be higher in ACPA-positive patients compared to those without ACPA. The estimate of genetic influence has varied across studies, probably due to gene–environment interactions.

The alleles known to confer the greatest risk of RA are located within the major histocompatibility complex (MHC) and, in particular, those encoding the MHC class II molecules. MHC class II molecules are typically expressed on antigen-presenting cells and are comprised of α and β chains. Most, but probably not all, of this risk is associated with allelic variation in the HLA-DRB1 gene, which encodes the MHC II β-chain molecule. The disease-associated HLA-DRB1 alleles share an amino acid sequence at positions 70–74 in the third hypervariable regions of the HLA-DR β-chain, termed the *shared epitope* (SE). These amino acids are located in the antigen-binding grove within the hypervariable regions of the HLA-DRβ1 molecule. These hypervariable regions are important not only for determining antigen recognition but also for binding of the MHC-peptide complex to the T-cell receptor (TCR). Peptides derived from posttranslationally modified proteins (via citrullination, acetylation, or carbamylation, for example) may bind with greater avidity to the SE, providing a potential explanation for increased disease risk at a molecular level. The risk of RA is four times higher in persons carrying a single SE allele and eight times higher in those carrying two alleles compared to SE-negative individuals.

Carriership of the SE alleles is associated with production of ACPAs and worse disease outcomes. Some HLA-DRB1 alleles bestow a high risk of disease (*0401), whereas others confer a more moderate risk (*0101, 0404, 1001, and 0901). Over 90% of patients with RA express at least one of these variants. Interestingly, HLA-DRB1*1301 and to a lesser extent HLA-DRB1*1302 confer protection from ACPA-positive RA.

Additionally, there is geographic variation in disease susceptibility and the identity of the HLA-DRB1 risk alleles. In Greece, for example,

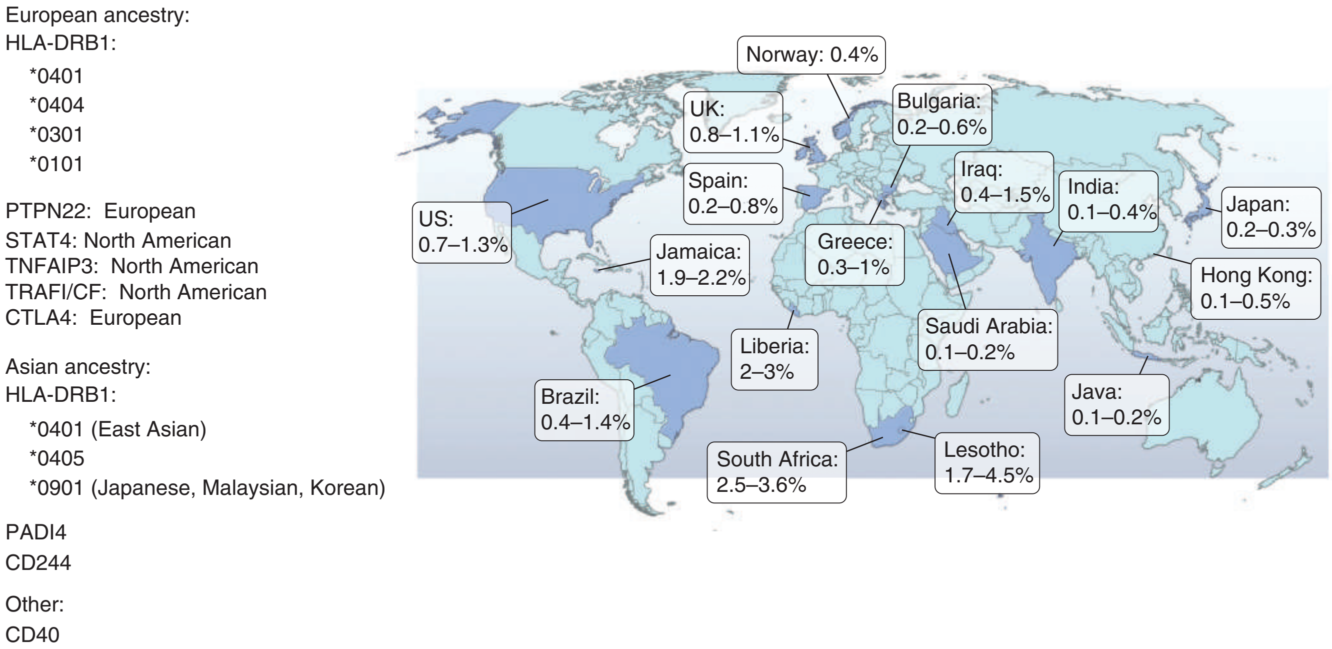

FIGURE 370-3 Global prevalence rates of rheumatoid arthritis (RA) with genetic associations. Listed are the major genetic alleles associated with RA. Although human leukocyte antigen (HLA)-DRB1 mutations are found globally, some alleles have been associated with RA in only certain ethnic groups.

where RA tends to be milder than in western European countries, RA susceptibility has been associated with the *0101 SE allele. By comparison, the *0401 or *0404 alleles are found in ~50–70% of northern Europeans and are the predominant risk alleles in this group. The most common disease susceptibility SE alleles in Asians, namely the Japanese, Koreans, and Chinese, are *0405 and *0901. Lastly, disease susceptibility of Native American populations such as the Pima and Tlingit Indians, where the prevalence of RA can be as high as 7%, is associated with the SE allele *1042. The risk of RA conferred by these SE alleles is less in African and Hispanic Americans than in individuals of European ancestry.

Genome-wide association studies (GWAS) have made possible the identification of several non-MHC-related genes that contribute to RA susceptibility. GWASs are based on the detection of single nucleotide polymorphisms (SNPs), which allow for examination of the genetic architecture of complex diseases such as RA. Overall, several themes have emerged from GWAS in RA. First, among the >100 non-MHC loci identified as risk alleles for RA, they individually have only a modest effect on risk; they also contribute to the risk for developing other autoimmune diseases, such as type 1 diabetes mellitus, systemic lupus erythematosus, and multiple sclerosis. Second, although most of the non-HLA associations are described in patients with ACPA-positive disease, there are several risk loci that are unique to ACPA-negative disease. Third, risk alleles vary among ethnic groups. And fourth, the risk loci mostly reside in genes encoding proteins involved in the regulation of the immune response. However, the risk alleles identified by GWAS only account at present for ~5% of the genetic risk, suggesting that rare variants or other classes of DNA variants, such as variants in copy number, may be yet found that significantly contribute to the overall risk model.

Among the best examples of the non-MHC genes contributing to the risk of RA is the gene encoding protein tyrosine phosphatase non-receptor 22 (*PTPN22*). This gene varies in frequency among patients from different parts of Europe (e.g., 3–10%) but is absent in patients of East Asian ancestry. *PTPN22* encodes lymphoid tyrosine phosphatase, a protein that regulates T- and B-cell function. Inheritance of the risk allele for *PTPN22* produces a gain-of-function in the protein that is hypothesized to result in the abnormal thymic selection of autoreactive T and B cells and appears to be associated exclusively with ACPA-positive disease. The peptidyl arginine deiminase type IV (*PADI4*) gene is another risk allele that encodes an enzyme involved in the conversion of arginine to citrulline and is postulated to play a role in the development of antibodies to citrullinated antigens. A polymorphism in *PADI4* has been associated with a twofold increase in the risk of RA, primarily in those of East Asian descent. Recently, polymorphisms in apolipoprotein M (*APOM*) in an East Asian population were found confer an increased risk for RA as well as risk for dyslipidemia, independent of RA disease activity.

Epigenetics is the study of heritable traits that affect gene expression but do not modify DNA sequence. Epigenetic mechanisms are theoretically involved in three important aspects of RA: contribution to disease etiology, perpetuation of chronic inflammatory responses, and disease severity. The best-studied epigenetic mechanisms are those regulating posttranslational histone modifications and DNA methylation. DNA methylation patterns have been shown to differ between RA patients and healthy controls, including monozygotic twins, as well as from patients with osteoarthritis. Epigenetics may also offer a mechanistic explanation for how cigarette smoking confers an increased risk for ACPA-positive RA (see below) as those patients with the SE have higher levels of DNA methylation compared to nonsmokers. MicroRNAs, which are noncoding RNAs that function as posttranscriptional regulators of gene expression, represent an additional epigenetic mechanism that may potentially influence cellular responses.

ENVIRONMENTAL FACTORS

In addition to genetic predisposition, a host of environmental factors have been implicated in the pathogenesis of RA. The most reproducible of these environmental links is cigarette smoking. Numerous cohort and case-control studies have demonstrated that smoking confers a relative risk for developing RA of 1.5–3.5 times. Smoking-related risk interacts in a synergistic manner with MHC risk alleles. The classic SE alleles alone increase the likelihood of developing RA by four- to sixfold; however, this risk increases to 20- to 40-fold when combined with smoking. In particular, women who smoke cigarettes have a nearly 2.5 times greater risk of RA, a risk that persists even 15 years after smoking cessation. A twin who smokes will have a significantly higher risk for RA than their monozygotic co-twin, theoretically with the same genetic risk, who does not smoke. Interestingly, the risk from smoking is almost exclusively related to RF and ACPA-positive disease. However, it has not been shown that smoking cessation, while having many health benefits, reduces the severity or extent of joint inflammation. Inhalant-related occupations and silica inhalants also may increase RA risk. These observations have led to the theory that subclinical lung disease may play a critical early role in the initial development of autoreactive immune cells and account for the occurrence of autoantibodies more than a decade prior to the clinical development of joint disease.

Researchers began to aggressively seek an infectious etiology for RA after the discovery in 1931 that sera from patients with this disease could agglutinate strains of streptococci. Certain viruses such as Epstein-Barr virus (EBV) have garnered the most interest over the past 30 years given their ubiquity, ability to persist for many years in the host, and frequent association with arthritic complaints. For example, titers of IgG antibodies against EBV antigens in the peripheral blood and saliva are significantly higher in patients with RA than in the general population. EBV DNA has also been found in synovial fluid and synovial cells of RA patients. Because the evidence for these links is largely circumstantial, it has not been possible to directly implicate infection as a causative factor in RA.

An attractive hypothesis raised by newer studies is that microbial dysbiosis of the oral or gut microbiome may predispose to the development of RA. Recent work suggests that periodontitis in the oral cavity may play a role in disease mechanisms. Multiple studies provide evidence for a link between ACPA-positive RA and cigarette smoking, periodontal disease, and the oral microbiome, specifically *Porphyromonas gingivalis*. It has been hypothesized that the immune response to *P. gingivalis* may trigger the development of RA and the bacterial enzyme peptidyl arginine deiminase (PAD) induces ACPA by catalyzing the citrullination of arginine residues on the stimulating autoantigens. Interestingly, *P. gingivalis* is the only oral bacterial species known to harbor this enzyme. Some studies have shown a relationship between the presence of circulating antibodies to *P. gingivalis* and RA, as well as with first-degree relatives at risk for this disease. However, it remains unproven whether the observed dysbiosis in the oral cavity precedes the development of disease.

There are also limited data suggesting a role for the gut microbiome in the etiology of RA. Some studies have found that the gut microbiome is different in patients with early RA compared with controls. In particular, *Prevotella copri* was reported to be enriched in early untreated RA as well as in an "at-risk" population. On the other hand, a common dysbiotic signature does not seem to predominate in patients with RA, and evidence is lacking for a direct immune-modulating effect.

PATHOLOGY

RA affects the synovial lining tissue of the diarthrodial joints and the underlying cartilage and bone. The synovial membrane, which covers most articular surfaces, tendon sheaths, and bursae, normally is a thin layer of connective tissue. In joints, it faces the bone and cartilage, bridging the opposing bony surfaces and inserting at periosteal regions close to the articular cartilage. It consists primarily of two cell types—type A synoviocytes (macrophage-derived) and type B synoviocytes (fibroblast-derived). The synovial fibroblasts are the most abundant and produce the structural components of joints, including collagen, fibronectin, and laminin, as well as other extracellular constituents of the synovial matrix. The sublining layer consists of blood vessels and a sparse population of mononuclear cells within a loose network of connective tissue. Synovial fluid, an ultrafiltrate of blood, diffuses through the subsynovial lining tissue across the synovial membrane and into the joint cavity. Its main constituents are hyaluronan and lubricin. Hyaluronan is a glycosaminoglycan that contributes to the viscous nature of synovial fluid, which, along with lubricin, lubricates the surface of the articular cartilage.

The pathologic hallmarks of RA are synovial inflammation and proliferation, focal bone erosions, and thinning of articular cartilage. Chronic inflammation leads to synovial lining hyperplasia and the formation of pannus, a thickened cellular membrane consisting of multiple layers of fibroblast-like synoviocytes and granulation-reactive fibrovascular tissue that has a propensity for invading underlying cartilage and bone. The inflammatory infiltrate is made up of no less than six cell types: T cells, B cells, plasma cells, dendritic cells, mast cells, and, to a lesser extent, granulocytes. The T cells compose 30–50% of the infiltrate, with the other cells accounting for the remainder. The topographical organization of these cells is complex and may vary among individuals with RA. Most often, the lymphocytes are diffusely organized among the tissue resident cells; however, in some cases, the B cells, T cells, and dendritic cells may form higher levels of organization, such as lymphoid follicles and germinal center–like structures. Growth factors secreted by synovial fibroblasts and macrophages promote the formation of new blood vessels in the synovial sublining that supply the increasing demands for oxygenation and nutrition required by the infiltrating leukocytes and expanding synovial tissue.

The structural damage to the mineralized cartilage and subchondral bone is mediated by the osteoclast. Osteoclasts are multinucleated giant cells that appear at the pannus-bone interface where they eventually form resorption lacunae. These lesions typically localize where the synovial membrane inserts into the periosteal surface at the edges of bone close to the rim of articular cartilage and at the attachment sites of ligaments and tendon sheaths. This process most likely explains why bone erosions usually develop at the radial sites of the MCP joints juxtaposed to the insertion sites of the tendons, collateral ligaments, and synovial membrane. Periarticular osteopenia, another form of bone loss, occurs in joints with active inflammation. It is associated with substantial thinning of the bony trabeculae along the metaphyses of bones from inflammation of the bone marrow cavity. These lesions can be visualized on MRI scans, where they appear as signal alterations in the bone marrow adjacent to inflamed joints. Their signal characteristics show they are water-rich with a low-fat content and are consistent with highly vascularized inflammatory tissue. These bone marrow lesions are often the forerunner of bone erosions. The cortical bone layer that separates the bone marrow from the invading pannus is relatively thin and susceptible to penetration by the inflamed synovium. The bone marrow lesions visualized on MRI scans are associated with an endosteal bone response characterized by the accumulation of osteoblasts and deposition of osteoid. Finally, generalized osteoporosis, which results in the thinning of trabecular bone throughout the body, is a third form of bone loss found in patients with RA and can lead to fragility fractures.

Articular cartilage is an avascular tissue comprised of a specialized matrix of collagens, proteoglycans, and other proteins. It is organized in four distinct regions (superficial, middle, deep, and calcified cartilage zones)—chondrocytes constitute the unique cellular component in these layers. Originally, cartilage was considered to be an inert tissue, but it is now known to be a highly responsive tissue that reacts to inflammatory mediators and mechanical factors, which, in turn, alter the balance between cartilage anabolism and catabolism. In RA, the initial areas of cartilage degradation are juxtaposed to the synovial pannus. The cartilage matrix is characterized by a generalized loss of proteoglycan, most evident in the superficial zones adjacent to the synovial fluid. Degradation of cartilage may also take place in the perichondrocytic zone and in regions adjacent to the subchondral bone.

PATHOGENESIS

The pathogenic mechanisms of synovial inflammation are likely the result of a complex interplay of genetic, environmental, and immunologic factors that cause immune system dysregulation and a breakdown in self-tolerance **(Fig. 370-4)**. Precisely what triggers these initiating events and what genetic and environmental factors disrupt the immune system remains a mystery. However, a detailed molecular picture is emerging of the mechanisms underlying the chronic inflammatory response and the resulting destruction of the articular cartilage and bone.

In RA, the preclinical stage appears to be characterized by a breakdown in self-tolerance. This idea is supported by the finding that autoantibodies, such as RF and ACPA, may be found in sera from patients many years before onset of clinical disease. However, the antigenic targets of ACPA and RF are not restricted to the joint. ACPAs are directed against deaminated peptides, which result from posttranslational modification by the enzyme PADI4. They recognize citrulline-containing regions of several different matrix proteins, including filaggrin, keratin, fibrinogen, and vimentin, and are present at higher levels in the joint fluid compared to the serum. Antibodies binding to carbamylated peptides and mutant citrullinated vimentin, as well as the 14-3-3 family of proteins, have also been detected in a minority of RA patients.

In theory, environmental triggers may synergize with other factors to bring about inflammation in RA. People who smoke display higher citrullination of proteins in bronchoalveolar fluid than those who do not smoke. Thus, it has been speculated that long-term exposure to tobacco smoke might induce citrullination of cellular proteins via increased PADI expression in the lung and generate a neoepitope capable of inducing self-reactivity, which in turns, leads to formation of immune complexes that trigger joint inflammation and joint damage.

Microbiota may also be involved in the initiating events of RA. The immune system is alerted to the presence of microbial infections when pathogen-associated molecular patterns (PAMPs) on their surface bind to Toll-like receptors (TLRs) on host cells. There are 10 TLRs in humans that recognize a variety of microbial products, including bacterial cell-surface lipopolysaccharides and heat-shock proteins (TLR4), lipoproteins (TLR2), double-strand RNA viruses (TLR3), and unmethylated CpG DNA from bacteria (TLR9). TLR2, 3, and 4 are abundantly expressed by synovial fibroblasts in early RA and, when bound by their ligands, upregulate the cell's production of proinflammatory cytokines. Although TLR ligands may theoretically amplify inflammatory pathways in RA, their specific role in disease pathogenesis remains uncertain.

The pathogenesis of RA is built upon the concept that self-reactive T cells drive the chronic inflammatory response. In theory, self-reactive T cells might arise in RA from abnormal central (thymic) selection or intrinsic defects lowering the threshold in the periphery for T-cell activation. Either mechanism might result in abnormal expansion of a self-reactive T-cell repertoire and a breakdown in T-cell tolerance. The support for these ideas comes mainly from studies of arthritis in mouse models. It has not been shown that patients with RA have abnormal thymic selection of T cells or defective apoptotic pathways regulating cell death. At least some antigen stimulation inside the joint seems likely, owing to the fact that T cells in the synovium express a cell-surface phenotype indicating prior antigen exposure and show evidence of clonal expansion.

There is substantial evidence of a role for CD4+ T cells in the pathogenesis of RA. First, the co-receptor CD4 on the surface of T cells binds to invariant sites on MHC class II molecules, stabilizing the MHC-peptide–T-cell receptor complex during T-cell activation. Because the SE on MHC class II molecules is a risk factor for RA, it follows that CD4+ T-cell activation plays a role in the pathogenesis of this disease. Second, CD4+ memory T cells are enriched in the synovial tissue from patients with RA and therefore are implicated through "guilt by association." Third, CD4+ T cells have been shown to be important in the initiation of arthritis in animal models. Fourth, some, but not all, T cell–directed therapies have shown clinical efficacy in this disease. Taken together, these lines of evidence suggest that CD4+ T cells play an important role in orchestrating the chronic inflammatory response in RA. However, other cell types, such as CD8+ T cells, natural killer (NK) cells, and B cells are present in synovial tissue and may also influence pathogenic responses.

In the rheumatoid joint, by mechanisms of cell-cell contact and release of soluble mediators, activated T cells stimulate macrophages and fibroblast-like synoviocytes to generate proinflammatory

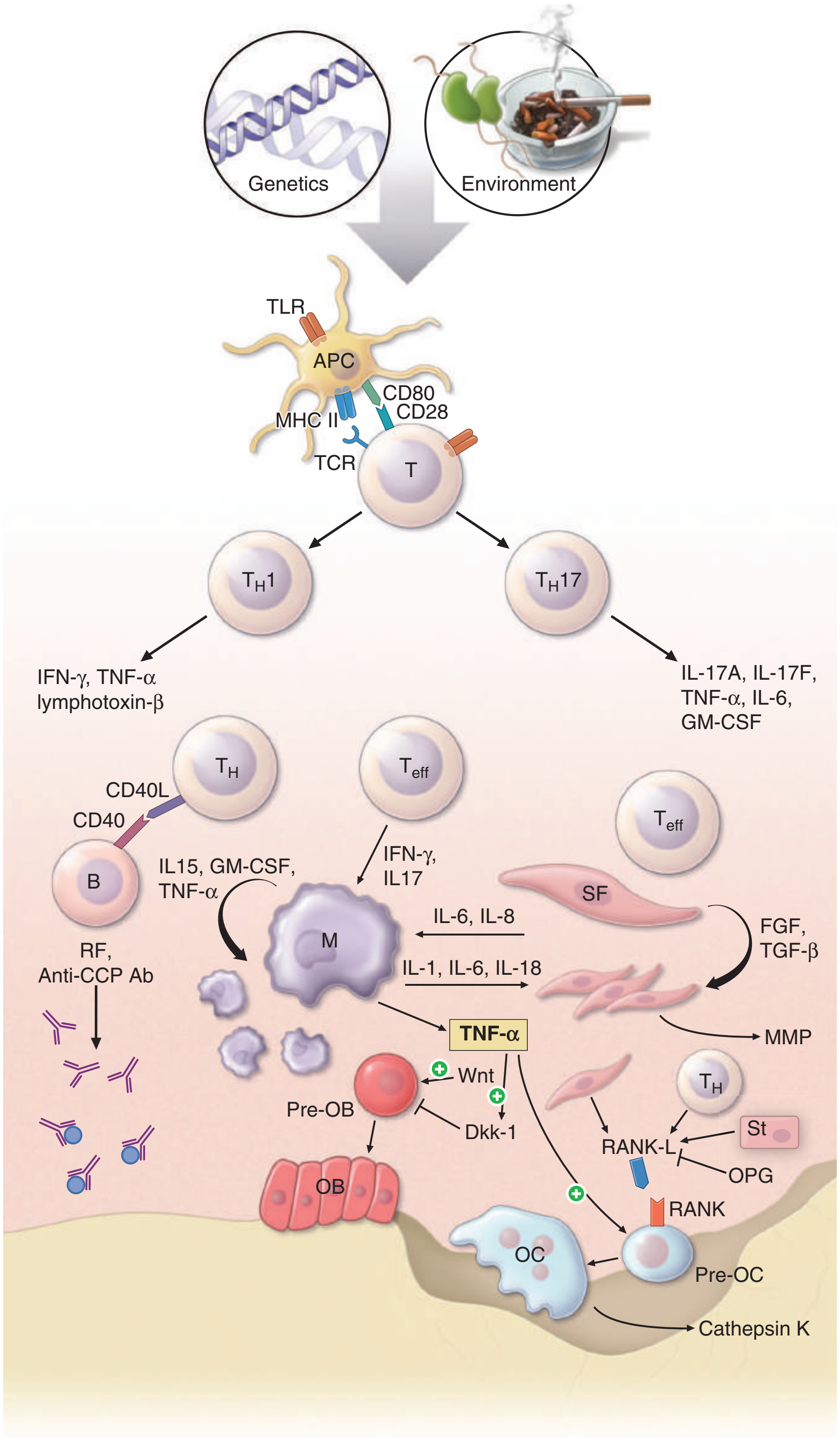

FIGURE 370-4 Pathophysiologic mechanisms of inflammation and joint destruction. Genetic predisposition along with environmental factors may trigger the development of rheumatoid arthritis (RA), with subsequent synovial T-cell activation. CD4+ T cells become activated by antigen-presenting cells (APCs) through interactions between the T-cell receptor and class II MHC-peptide antigen (signal 1) with co-stimulation through the CD28-CD80/86 pathway, as well as other pathways (signal 2). In theory, ligands binding Toll-like receptors (TLRs) may further stimulate activation of APCs inside the joint. Synovial CD4+ T cells differentiate into T_H1 and T_H17 cells, each with their distinctive cytokine profile. CD4+ T_H cells in turn activate B cells, some of which are destined to differentiate into autoantibody-producing plasma cells. Immune complexes, possibly comprised of rheumatoid factors (RFs) and anti–cyclic citrullinated peptides (CCP) antibodies, may form inside the joint, activating the complement pathway and amplifying inflammation. T effector cells stimulate synovial macrophages (M) and fibroblasts (SF) to secrete proinflammatory mediators, among which is tumor necrosis factor α (TNF-α). TNF-α upregulates adhesion molecules on endothelial cells, promoting leukocyte influx into the joint. It also stimulates the production of other inflammatory mediators, such as interleukin 1 (IL-1), IL-6, and granulocyte-macrophage colony-stimulating factor (GM-CSF). TNF-α has a critically important function in regulating the balance between bone destruction and formation. It upregulates the expression of dickkopf-1 (DKK-1), which can then internalize Wnt receptors on osteoblast precursors. Wnt is a soluble mediator that promotes osteoblastogenesis and bone formation. In RA, bone formation is inhibited through the Wnt pathway, presumably due to the action of elevated levels of DKK-1. In addition to inhibiting bone formation, TNF-α stimulates osteoclastogenesis. However, it is not sufficient by itself to induce the differentiation of osteoclast precursors (Pre-OC) into activated osteoclasts capable of eroding bone. Osteoclast differentiation requires the presence of macrophage colony-stimulating factor (M-CSF) and receptor activator of nuclear factor-κB (RANK) ligand (RANKL), which binds to RANK on the surface of Pre-OC. Inside the joint, RANKL is mainly derived from stromal cells, synovial fibroblasts, and T cells. Osteoprotegerin (OPG) acts as a decoy receptor for RANKL, thereby inhibiting osteoclastogenesis and bone loss. FGF, fibroblast growth factor; IFN, interferon; MMP, matrix metalloproteinase; TGF, transforming growth factor.

mediators and proteases that drive the synovial inflammatory response and destroy the cartilage and bone. CD4+ T-cell activation is dependent on two signals: (1) T-cell receptor binding to peptide-MHC on antigen-presenting cells; and (2) CD28 binding to CD80/86 on antigen-presenting cells. This interaction then leads to downstream signals that differentiate CD4+ T cells into effector and memory cell populations, as well as activate CD8+ T cells. Certain subsets of CD4+ T cells, called T helper cells, enable B cells to differentiate into antibody-secreting cells. An earlier T cell–centric model for the pathogenesis of RA was based on a T_H1-driven paradigm, which came from studies indicating that CD4+ T helper (T_H) cells differentiated into T_H1 and T_H2 subsets, each with their distinctive cytokine profiles. T_H1 cells were found to mainly produce interferon γ (IFN-γ), lymphotoxin β, and tumor necrosis factor (TNF)-α, whereas T_H2 cells predominately secreted interleukin (IL)-4, IL-5, IL-6, IL-10, and IL-13. In humans, naïve T cells may also be induced to differentiate into T_H17 cells by exposure to transforming growth factor β (TGF-β), IL-1, IL-6, and IL-23. Upon activation, T_H17 cells secrete a variety of proinflammatory mediators such as IL-17, IL-21, IL-22, TNF-α, IL-26, IL-6, and granulocyte-macrophage colony-stimulating factor (GM-CSF). Substantial evidence now exists from studies in both animal models and humans that IL-17 plays an important role not only in promoting joint inflammation but also in destroying cartilage and subchondral bone. However, therapeutic antibodies that interfere with the IL-17/IL-17 receptor pathway were not particularly effective in clinical trials for reducing joint inflammation in RA.

The immune system has evolved mechanisms to counterbalance the potentially harmful immune-mediated inflammatory responses provoked by infectious agents and other triggers. Among these negative regulators are regulatory T (Treg) cells, which are produced in the thymus and induced in the periphery to suppress immune-mediated inflammation. They are characterized by the surface expression of CD25 and the expression of the transcription factor forkhead box P3 (FOXP3). Tregs orchestrate dominant tolerance through contact with other immune cells and secretion of inhibitory cytokines, such as TGF-β, IL-10, and IL-35. They are heterogeneous and capable of suppressing distinct classes (T_H1, T_H2, T_H17) of the immune response, although data supporting their role in suppressing inflammation in RA are inconclusive.

Cytokines, chemokines, antibodies, and endogenous danger signals bind to receptors on the surface of immune cells and stimulate a cascade of intracellular signaling events that can amplify the inflammatory response. Examples of signaling molecules in these critical inflammatory pathways include Janus kinase (JAK)/signal transducers and activators of transcription (STAT), spleen tyrosine kinase (Syk), mitogen-activated protein kinases (MAPKs), and nuclear factor-κB (NF-κB). These pathways exhibit significant crosstalk and are found in many cell types. Some signal transducers, such as the JAKs, are expressed in hematopoietic cells and play an important role in the inflammatory response in RA.

Activated, autoreactive B cells are also important players in the chronic inflammatory process. B cells give rise to plasma cells, which, in turn, produce antibodies, including RF and ACPA. RFs may form large immune complexes inside the joint that contribute to the pathogenic process by fixing complement and promoting the release of proinflammatory cytokines and chemokines. In mouse models of arthritis, RF-containing immune complexes and ACPA-containing immune complexes synergize with other mechanisms to exacerbate the synovial inflammatory response.

RA is often considered to be a macrophage-driven disease because this cell type is the predominant source of proinflammatory cytokines inside the joint. Key proinflammatory cytokines released by macrophage-like synoviocytes include TNF-α, IL-1, IL-6, IL-12, IL-15, IL-18, and IL-23. Synovial fibroblasts, the other major cell type in this microenvironment, produce the cytokines IL-1 and IL-6 as well as TNF-α. TNF-α is a pivotal cytokine in the pathobiology of synovial inflammation and has many proinflammatory effects. It upregulates adhesion molecules on endothelial cells, promoting the influx of leukocytes into the synovial microenvironment; activates synovial fibroblasts; stimulates angiogenesis; promotes pain receptor sensitizing pathways; and drives osteoclastogenesis. Fibroblasts secrete matrix metalloproteinases (MMPs) as well as other proteases that are chiefly responsible for the breakdown of articular cartilage; they also promote inflammation and synovial proliferation by secreting cytokines such as IL-6, IL-1, IL-18, and GM-CSF, chemokines, and vascular endothelial growth factor.

Recent studies of synovial tissue samples from patients with RA have combined histology and single-cell RNA expression data to identify distinct cell populations that drive joint inflammation. These studies reveal that synovial tissue samples can be distinguished among patients with RA by the relative abundance of myeloid cells, T cells, B cells, synovial fibroblasts, and endothelial cells and hence can be divided into different subsets. Thus far, these data highlight the importance of IFN-γ-secreting granzyme K+ CD8+ T cells, peripheral helper CD4+ T cells, activated sublining fibroblasts, and proinflammatory monocytes as key mediators of joint inflammation. In these studies, most of the synovial tissue samples were found to be populated by T cells, fibroblasts, and myeloid cells and had cellular phenotypes classified as predominately myeloid, T and B cell, and T cell and myeloid, while others had a paucity of immune cells and were dominated by endothelial cells and fibroblasts. Further research along these lines will enrich our understanding of the active immune pathways in RA and may help to determine in the future if knowledge of these cellular phenotypes can be used to guide a personalized treatment approach.

Osteoclast activation at the site of the pannus is closely tied to the presence of focal bone erosion. Receptor activator of nuclear factor-κB ligand (RANKL) is expressed by stromal cells, synovial fibroblasts, and T cells. Upon binding to its receptor RANK on osteoclast progenitors, RANKL stimulates osteoclast differentiation and bone resorption. RANKL activity is regulated by osteoprotegerin (OPG), a decoy receptor of RANKL that blocks osteoclast formation. Monocytic cells in the synovium serve as the precursors of osteoclasts and, when exposed to macrophage colony-stimulating factor (M-CSF) and RANKL, fuse to form polykaryons termed *preosteoclasts*. These precursor cells undergo further differentiation into osteoclasts with a characteristic ruffled membrane. Cytokines such as TNF-α, IL-1, IL-6, and IL-17 can promote osteoclastogenesis by increasing the expression of RANKL inside the joint. Osteoclasts also secrete cathepsin K, a cysteine protease that degrades the bone matrix by cleaving collagen and thus contributes to generalized bone loss and osteoporosis.

Increased bone loss is only part of the story in RA, as decreased bone formation plays a crucial role in bone remodeling at sites of inflammation. Recent evidence shows that inflammation suppresses bone formation. TNF-α plays a key role in actively suppressing bone formation by enhancing the expression of dickkopf-1 (DKK-1). DKK-1 is an important inhibitor of the Wnt pathway, which acts to promote osteoblast differentiation and bone formation. The Wnt system is a family of soluble glycoproteins that binds to cell-surface receptors known as frizzled (fz) and low-density lipoprotein (LDL) receptor–related proteins (LRPs) and promotes cell growth. In animal models, increased levels of DKK-1 are associated with decreased bone formation, whereas inhibition of DKK-1 protects against structural damage in the joint. Wnt proteins also induce the formation of OPG and thereby shut down bone resorption, emphasizing their key role in tightly regulating the balance between bone resorption and formation.

DIAGNOSIS

The clinical diagnosis of RA is largely made by recognizing the signs and symptoms of a chronic inflammatory arthritis and the typical pattern of joint involvement, with laboratory and radiographic results providing important corroborating information. In 2010, a collaborative effort between the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR) revised the 1987 ACR classification criteria for RA in an effort to improve early diagnosis with the goal of identifying patients who would benefit from early introduction of disease-modifying therapy **(Table 370-1)**. The items in the newly revised criteria yield a score of 0–10, with a score of ≥6 fulfilling the requirements for definite RA. The new classification

TABLE 370-1 Classification Criteria for Rheumatoid Arthritis

		SCORE
Joint involvement	1 large joint (shoulder, elbow, hip, knee, ankle)	0
	2–10 large joints	1
	1–3 small joints (MCP, PIP, thumb IP, MTP, wrists)	2
	4–10 small joints	3
	>10 joints (at least 1 small joint)	5
Serology	Negative RF and negative ACPA	0
	Low-positive RF or low-positive anti-CCP antibodies (≤3 times ULN)	2
	High-positive RF or high-positive anti-CCP antibodies (>3 times ULN)	3
Acute-phase reactants	Normal CRP and normal ESR	0
	Abnormal CRP or abnormal ESR	1
Duration of symptoms	<6 weeks	0
	≥6 weeks	1

Note: These criteria are aimed at classification of newly presenting patients who have at least one joint with definite clinical synovitis that is not better explained by another disease. A score of ≥6 fulfills requirements for definite RA.

Abbreviations: ACPA, anti-citrullinated peptide antibodies; CCP, cyclic citrullinated peptides; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; IP, interphalangeal joint; MCP, metacarpophalangeal joint; MTP, metatarsophalangeal joint; PIP, proximal interphalangeal joint; RF, rheumatoid factor; ULN, upper limit of normal.

Source: Reproduced with permission from D Aletaha et al: 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 62:2569, 2010.

criteria differ in several ways from the older criteria sets. Previous classification schemes required symptoms to be present for >6 weeks. There are several conditions, including virus-related syndromes, that may cause a polyarthritis mimicking RA and be associated with transient production of RF. Such conditions usually last only 2–3 weeks. The 2010 criteria, however, do not mandate symptoms be present for >6 weeks. They also include as an item a positive test for serum ACPA, which carries greater specificity for the diagnosis of RA than a positive test for RF. About three-fourths of patients with the clinical and radiographic features of RA test positive for RF and/or ACPA (seropositive), while the remaining one-fourth of patients with RA test negative for RF and/or ACPA (seronegative). The presence of rheumatoid nodules or radiographic joint damage as items were not included as they occur only rarely in early RA. It is important to emphasize that the 2010 ACR-EULAR criteria are "classification criteria" as opposed to "diagnostic criteria" and serve to distinguish patients at the onset of disease who have a high likelihood of evolution to chronic disease with persistent synovitis and joint damage. The presence of radiographic joint erosions or subcutaneous nodules may inform the diagnosis in the later stages of the disease.

The differential diagnosis for RA includes all types of acute and chronic inflammatory arthritides, many of which may be differentiated from RA based on the clinical course, pattern of joint involvement, the presence of disease in other organ systems, and ancillary laboratory studies. Patients with primary Sjögren's syndrome whose predominate clinical manifestations are dry eyes and dry mouth often also have symptoms of polyarthralgia and may show a mild inflammatory synovitis similar to RA. Moreover, 50% of patients with primary Sjögren's syndrome test positive for RF and, therefore, may be confused with early RA. However, most patients with primary Sjögren's syndrome test positive for ANA and approximately two-thirds have detectable anti-Ro/SSA antibodies. Spondyloarthropathies, such as psoriatic arthritis or enteropathy-associated arthritis, may present similarly to RA and may be distinguished by the presence of sacroiliitis, other enthesopathic features, and the presence of psoriasis or inflammatory bowel disease, respectively. In elderly patients, seronegative RA may be difficult at times to distinguish from polymyalgia rheumatica (PMR), a chronic inflammatory condition of the elderly characterized by neck/shoulder girdle and lower back/hip girdle pain and stiffness **(Chap. 375)**. While PMR may occasionally feature distal limb involvement, RA predominately affects the wrists/hands and ankles/feet, especially at disease onset. Similarly, the relatively rare condition called remitting seronegative symmetrical synovitis with pitting edema (RS3PE) and paraneoplastic syndromes may also be confused with early RA. RS3PE is typically characterized by prominent distal limb pitting edema, which is unusual in RA, and responds promptly to treatment with low doses of prednisone. Chronic tophaceous gout may mimic severe RA in some cases, and tophi may be confused with rheumatoid nodules. Hepatitis C–related arthropathy often involves the small joints of the hands and is associated in half the cases with a positive RF, but generally not ACPA.

LABORATORY FEATURES

Patients with systemic inflammatory diseases such as RA will often present with elevated nonspecific inflammatory markers such as an ESR or CRP. Testing for serum RF and anti-cyclic citrullinated peptide (CCP) antibodies is important in differentiating RA from other polyarticular diseases, although RF lacks diagnostic specificity and may be found in association with other chronic inflammatory diseases in which arthritis figures in the clinical manifestations.

IgM, IgG, and IgA isotypes of RF occur in sera from patients with RA, although the IgM isotype is the one most frequently measured by commercial laboratories. Serum IgM RF has been found in 75% of patients with RA; therefore, a negative result does not exclude the presence of this disease. It is also found in other connective tissue diseases, such as primary Sjögren's syndrome, systemic lupus erythematosus, and type II mixed essential cryoglobulinemia, as well as chronic infections such as subacute bacterial endocarditis and hepatitis B and C. Serum RF, mostly at low levels, may also be detected in 1–5% of the healthy population.

The presence of serum anti-CCP antibodies has about the same sensitivity as serum RF for the diagnosis of RA. However, its diagnostic specificity approaches 95%, so a positive test for anti-CCP antibodies in the setting of an early inflammatory arthritis is useful for distinguishing RA from other forms of arthritis. There is some incremental value in testing for the presence of both RF and anti-CCP, as some patients with RA are positive for RF but negative for anti-CCP and vice versa. The presence of RF or anti-CCP antibodies also has prognostic significance, with anti-CCP antibodies showing the most value for predicting worse outcomes.

Patients with RA may also have other antibodies associated with autoimmune disease. Approximately 30% of patients with RA test positive for antinuclear antibodies (ANAs), and some sera from some patients contain antineutrophil cytoplasmic antibodies (ANCAs; particularly p-ANCAs). However, patients with RA would not be expected to test positive for anti-MPO or anti-PR3 antibodies as they would if they had an ANCA-associated vasculitis.

■ SYNOVIAL FLUID ANALYSIS

Typically, the cellular composition of synovial fluid from patients with RA reflects an acute inflammatory state. Synovial fluid white blood cell (WBC) counts can vary widely but generally range between 5000 and 50,000 WBC/μL, compared with <2000 WBC/μL for a noninflammatory condition such as osteoarthritis. In contrast to the synovial tissue, the overwhelming cell type in the synovial fluid is the neutrophil. Clinically, the analysis of synovial fluid is most useful for confirming an inflammatory arthritis (as opposed to osteoarthritis), while at the same time excluding infection or a crystal-induced arthritis such as gout or pseudogout **(Chap. 384)**.

■ JOINT IMAGING

Joint imaging is a valuable tool not only for diagnosing RA but also for tracking progression of any joint damage. Plain x-ray is the most common imaging modality, but it is limited to visualization of the bony structures and inferences about the state of the articular cartilage based on the amount of joint space narrowing. MRI and ultrasound techniques offer the added value of detecting changes in the soft tissues such as synovitis, tenosynovitis, and effusions, as well as providing greater sensitivity for identifying bony abnormalities. Plain

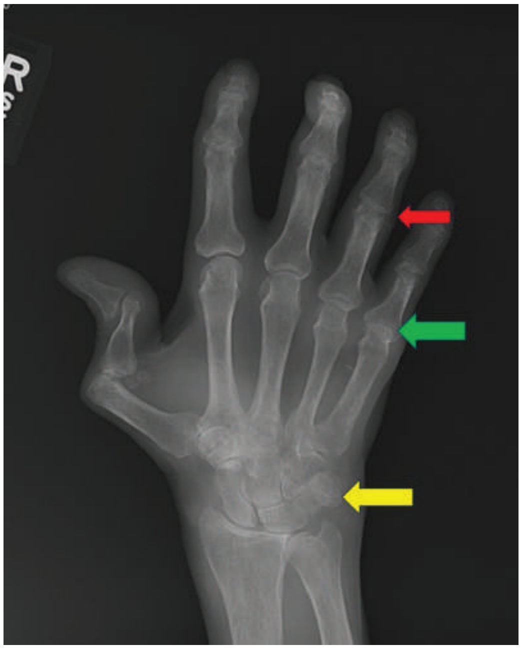

FIGURE 370-5 X-ray of the right hand demonstrating features of advanced rheumatoid arthritis. This x-ray is notable for carpal crowding of the wrist (*yellow arrow*), loss of joint space and subluxation of the fifth metacarpophalangeal joint (*green arrow*), and an erosion of the fourth proximal interphalangeal joint (*red arrow*).

radiographs are usually relied upon in clinical practice for the purpose of diagnosis and monitoring affected joints. However, in selected cases, MRI and ultrasound can provide additional diagnostic information that may guide clinical decision-making. Musculoskeletal ultrasound with power Doppler is increasingly used in rheumatology clinical practice for detecting synovitis and bone erosion.

Plain Radiography Classically in RA, the initial radiographic finding is periarticular osteopenia. Practically speaking, however, this finding is difficult to appreciate on plain films and particularly their digitalized images. Other findings on plain radiographs include soft tissue swelling, symmetric joint space loss, and juxtaarticular and subchondral erosions, most frequently in the wrists and hands (MCPs and PIPs) and the feet (MTPs). In the feet, the lateral aspect of the fifth MTP is often targeted first, but other MTP joints may be involved at the same time. X-ray imaging of advanced RA may reveal signs of severe destruction, including joint subluxation and collapse (Fig. 370-5).

MRI MRI offers the greatest sensitivity for detecting synovitis and joint effusions, as well as early bone and bone marrow changes. These soft tissue abnormalities often occur before osseous changes are noted on x-ray. Presence of bone marrow edema has been recognized to be an early sign of inflammatory joint disease and can predict the subsequent development of erosions on plain radiographs as well as MRI scans. Cost and availability of MRI are the main factors limiting its routine clinical use.

Ultrasound Ultrasound, including power color Doppler, can detect more erosions than plain radiography, especially in easily accessible joints. It can also reliably detect synovitis, including increased joint vascularity indicative of inflammation. The usefulness of ultrasound is dependent on the experience of the sonographer; however, it does offer the advantages of portability, lack of radiation, and low expense relative to MRI, factors that make it attractive as a clinical tool (Fig. 370-6). See Videos 370-1, 370-2, and 370-3.

CLINICAL COURSE

The natural history of RA is complex and affected by a number of factors including age of onset, gender, genotype, phenotype (i.e., extraarticular manifestations or variants of RA), and comorbid conditions, which make for a truly heterogeneous disease. It is important to realize that as many as 10% of patients with inflammatory arthritis fulfilling ACR classification criteria for RA will undergo a spontaneous remission within 6 months (particularly seronegative patients). However, the vast majority of patients will exhibit a pattern of persistent and progressive disease activity that waxes and wanes in intensity over time. A minority of patients will show intermittent and recurrent explosive attacks of inflammatory arthritis interspersed with periods of disease quiescence. Finally, an aggressive form of RA may occur in an unfortunate few with inexorable progression of severe erosive joint disease, although this highly destructive course is less common in the modern treatment era.

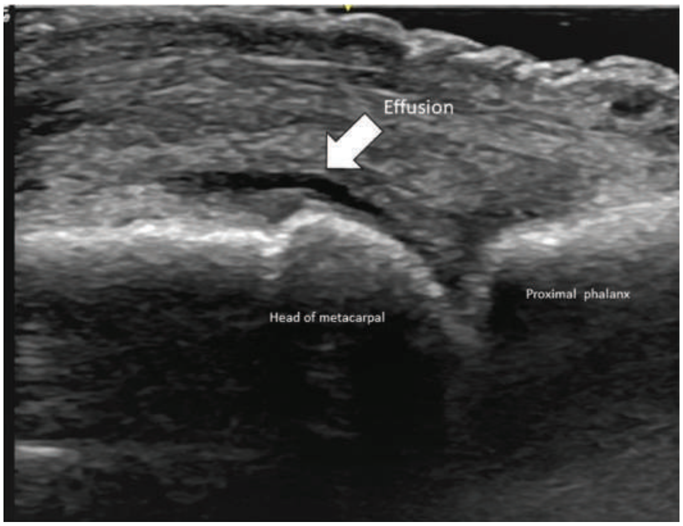

FIGURE 370-6 Ultrasound demonstrating an effusion (*arrow*) within the metacarpophalangeal joint. *(Courtesy of Dr. Ryan Jessee.)*

Disability, as measured by the Health Assessment Questionnaire (HAQ), shows gradual worsening over time in the face of poorly controlled disease activity and disease progression. Disability may result from both a disease activity–related component that is potentially reversible with therapy and a joint damage–related component owing to the cumulative and largely irreversible effects of soft tissue, cartilage, and bone breakdown. Early in the course of disease, the extent of joint inflammation is the primary determinant of disability, while in the later stages of disease, the amount of joint damage is the dominant contributing factor. Previous studies have shown that more than one-half of patients with RA are unable to work 10 years after the onset of their disease; however, increased employability and less work absenteeism have been reported recently with the use of newer therapies and earlier treatment intervention.

The overall mortality rate in RA is two times greater than in the general population. In particular, patients are at significantly increased risk for death due to respiratory illness, cardiovascular disease, and infection. Median life expectancy is shortened by an average of 7 years for men and 3 years for women compared with control populations. Patients at higher risk for shortened survival are those with systemic extraarticular involvement, low functional capacity, low socioeconomic status, low education, and chronic prednisone use.

TREATMENT

Rheumatoid Arthritis

The amount of clinical disease activity in patients with RA reflects the overall burden of inflammation and is the variable that most influences treatment decisions. Joint inflammation is the main driver of joint damage and is the most important cause of functional disability in the early stages of disease. Several composite indices have been developed to assess clinical disease activity. The ACR 20, 50, and 70 improvement criteria (which correspond to a 20, 50, and 70% improvement, respectively, in joint counts, physician/patient assessment of disease severity, pain scale, serum levels of acute-phase reactants [ESR or CRP], and a functional assessment of disability using a self-administered patient questionnaire) are a composite index with a dichotomous response variable. The ACR improvement criteria are commonly used in clinical trials as an endpoint for comparing the proportion of responders between treatment groups. In contrast, the Disease Activity Score (DAS), Simplified Disease Activity Index (SDAI), the Clinical Disease Activity Index (CDAI), and the Routine Assessment of Patient Index Data 3 (RAPID3) are continuous measures of disease activity that are used in clinical practice for tracking disease status and documenting treatment response.

The identification of genetic and environmental risk factors as well as the presence of autoantibodies creates a phenotype of a person who may be at higher risk for development of RA, or "preclinical RA." Several trials have enrolled patients meeting a definition of preclinical RA in an attempt to prevent development of disease with disease-modifying therapies approved for the treatment of RA. While some of these disease-modifying therapies may delay the development of RA compared to a placebo, no strategy has yet been successful in preventing disease.

Several developments during the past two decades have changed the therapeutic landscape in RA. They include (1) the emergence of methotrexate as the disease-modifying antirheumatic drug (DMARD) of first choice for the treatment of early RA; (2) the development of novel highly efficacious biologicals that can be used alone or in combination with methotrexate; and (3) the proven superiority of combination DMARD regimens over methotrexate alone. The medications used for the treatment of RA may be divided into broad categories: nonsteroidal anti-inflammatory drugs (NSAIDs); glucocorticoids, such as prednisone and methylprednisolone; conventional DMARDs; and biologic DMARDs **(Table 370-2)**. Although disease for some patients with RA is managed adequately with a single DMARD, such as methotrexate, it calls in most cases for the use of a combination DMARD regimen that may vary in its components over the treatment course depending on fluctuations in disease activity, loss of treatment efficacy, and emergence of drug-related toxicities and comorbidities.

NSAIDS

NSAIDs were formerly viewed as the core of RA therapy, but they are now considered to be adjunctive agents for management of symptoms uncontrolled by other measures. NSAIDs exhibit both analgesic and anti-inflammatory properties. The anti-inflammatory effects of NSAIDs derive from their ability to nonselectively inhibit cyclooxygenase (COX)-1 and COX-2. Although the results of clinical trials suggest that NSAIDs are roughly equivalent in their efficacy, experience suggests that some individuals may preferentially respond to a particular NSAID. Chronic use should be minimized due to the possibility of side effects, particularly gastritis and peptic ulcer disease and kidney injury.

GLUCOCORTICOIDS

Glucocorticoids may serve in several ways to control disease activity in RA. First, they may be administered in low to moderate doses to achieve rapid disease control before the onset of fully effective DMARD therapy, which often takes several weeks or even months. Second, a 1- to 2-week burst of glucocorticoids may be prescribed for the management of acute disease flares, with dose and duration guided by the severity of the exacerbation. Chronic administration of low doses (5–10 mg/d) of prednisone (or its equivalent) may also be warranted to control disease activity in patients with an inadequate response to DMARD therapy. As much as possible, chronic glucocorticoid therapy should be avoided in favor of finding an effective DMARD regimen that adequately controls the disease. Best practices minimize chronic use of low-dose prednisone therapy owing to the risk of osteoporosis and other long-term complications; however, the use of chronic prednisone therapy is unavoidable in some cases. High-dose glucocorticoids may be necessary for the treatment of severe extraarticular manifestations of RA, such as ILD. Finally, if a patient exhibits one or a few actively inflamed joints, the clinician may consider intraarticular injection of an intermediate-acting glucocorticoid such as triamcinolone acetonide. This approach may allow for rapid control of inflammation in a limited number of affected joints. Caution must be exercised to appropriately exclude joint infection as it often mimics a monoarticular RA flare.

Osteoporosis ranks as an important long-term complication of chronic prednisone use. Based on a patient's risk factors, including total prednisone dosage, length of treatment, gender, race, and bone density, treatment with an oral (e.g., alendronate, risedronate) or parenteral (e.g., zoledronic acid) bisphosphonate may be appropriate for primary prevention of glucocorticoid-induced osteoporosis. Other agents, including denosumab and teriparatide, have also been approved for the treatment of steroid-induced osteoporosis and may be indicated in certain cases. Although prednisone use is known to increase the risk of peptic ulcer disease, especially with concomitant NSAIDs, no evidence-based guidelines recommend routine co-administration of proton pump inhibitors for gastrointestinal ulcer prophylaxis.

DMARDS

DMARDs are so named because of their ability to slow or prevent structural progression of RA. The conventional DMARDs include hydroxychloroquine, sulfasalazine, methotrexate, and leflunomide; they exhibit a delayed onset of action of ~6–12 weeks or longer. Methotrexate is the DMARD of choice for the treatment of RA and is the anchor drug for most combination therapies. It was approved for the treatment of RA in 1988 and remains a benchmark for comparison with the newer disease-modifying therapies. At the dosages used for the treatment of RA, methotrexate has been shown to stimulate adenosine release from cells, producing an anti-inflammatory effect. Methotrexate is administered weekly by the oral or subcutaneous route. Folic acid is taken as co-therapy to mitigate some of methotrexate's side effects. The clinical efficacy of leflunomide, an inhibitor of pyrimidine synthesis, appears similar to that of methotrexate; it has been shown in well-designed trials to be effective for the treatment of RA as monotherapy or in combination with methotrexate and other DMARDs.

Although similar to the other DMARDs in its slow onset of action, hydroxychloroquine has not been shown to delay radiographic progression of disease and thus is not considered to be a true DMARD. In clinical practice, hydroxychloroquine is generally used for treatment of early, mild disease or as adjunctive therapy in combination with other DMARDs. It is prescribed at a dosage of 5 mg/kg of body weight or less to decrease the risk of retinal toxicity. Sulfasalazine has been shown in randomized, controlled trials to reduce joint inflammation and radiographic progression of disease. A regimen termed oral triple therapy, which is comprised of hydroxychloroquine, sulfasalazine, and methotrexate, is an effective combination regimen often used in patients with an inadequate response to treatment with methotrexate alone.

BIOLOGICALS

Biologic DMARDs have revolutionized the treatment of RA over the past two decades (Table 370-2). They are protein therapeutics designed to target cytokines and cell-surface molecules. The TNF inhibitors were the first biologicals approved for the treatment of RA, but several others have come to market since their introduction in 1999. Additionally, biosimilars are agents that have been designed with the same active properties as the licensed biopharmaceutical product. Several of these biosimilars have been studied for the treatment of RA and demonstrate a similar efficacy and side-effect profile to the reference biological.

Anti-TNF Agents The development of TNF inhibitors was originally spurred by the experimental finding that TNF-α is a critical upstream mediator of joint inflammation. Five original "innovator" TNF inhibitors with various molecular designs have been approved for the treatment of RA. Three of these products are anti-TNF

TABLE 370-2 DMARDs Used for the Treatment of Rheumatoid Arthritis

DRUG	DOSAGE	SERIOUS TOXICITIES	OTHER COMMON SIDE EFFECTS	INITIAL EVALUATION	MONITORING
Hydroxychloroquine	200–400 mg/d orally (≤5 mg/kg)	Irreversible retinal damage Cardiotoxicity Blood dyscrasia	Nausea Diarrhea Headache Rash	Eye examination if >40 years old or prior ocular disease	Optical coherence tomography and visual field testing every 12 months
Sulfasalazine	Initial: 500 mg orally twice daily Maintenance: 1000–1500 mg twice daily	Granulocytopenia Hemolytic anemia (with G6PD deficiency)	Nausea Diarrhea Headache	CBC, LFTs G6PD level	CBC every 2–4 weeks for first 3 months, then every 3 months
Methotrexate	10–25 mg/week orally or SQ Folic acid 1 mg/d to reduce toxicities	Hepatotoxicity Myelosuppression Infection Interstitial pneumonitis Pregnancy category X	Nausea Diarrhea Stomatitis/mouth ulcers Alopecia Fatigue	CBC, LFTs Viral hepatitis panel[a] Chest x-ray	CBC, creatinine, LFTs every 2–3 months
Leflunomide	10–20 mg/d	Hepatotoxicity Myelosuppression Infection Pregnancy category X	Alopecia Diarrhea	CBC, LFTs Viral hepatitis panel[a]	CBC, creatinine, LFTs every 2–3 months
TNF-α inhibitors	Infliximab: 3 mg/kg IV at weeks 0, 2, 6, then every 8 weeks. May increase dose up to 10 mg/kg every 4 weeks	↑ Risk bacterial, fungal infections Reactivation of latent tuberculosis ↑ Lymphoma risk (controversial) Drug-induced lupus Neurologic deficits	Infusion reaction ↑ LFTs	Tuberculosis screening[b]	LFTs periodically
	Etanercept: 50 mg SQ weekly, or 25 mg SQ biweekly	As above	Injection site reaction	Tuberculosis screening	Monitor for injection site reactions
	Adalimumab: 40 mg SQ every other week	As above	Injection site reaction	Tuberculosis screening	Monitor for injection site reactions
	Golimumab: 50 mg SQ monthly	As above	Injection site reaction	Tuberculosis screening	Monitor for injection site reactions
	Certolizumab: 400 mg SQ weeks 0, 2, 4, then 200 mg every other week	As above	Injection site reaction	Tuberculosis screening	Monitor for injection site reactions
Abatacept	Weight based: <60 kg: 500 mg 60–100 kg: 750 mg >100 kg: 1000 mg IV dose at weeks 0, 2, and 4, and then every 4 weeks OR 125 mg SQ weekly	↑ Risk bacterial, viral infections	Headache Nausea	Tuberculosis screening	Monitor for infusion reactions
Anakinra	100 mg SQ daily	↑ Risk bacterial, viral infections Reactivation of latent tuberculosis Neutropenia	Injection site reaction Headache	Tuberculosis screening CBC with differential	CBC every month for 3 months, then every 4 months for 1 year Monitor for injection site reactions
Rituximab	1000 mg IV × 2, days 0 and 14 May repeat course every 24 weeks or more Premedicate with methylprednisolone 100 mg to decrease infusion reaction	↑ Risk bacterial, viral infections Infusion reaction Cytopenia Hepatitis B reactivation	Rash Fever	CBC Viral hepatitis panel[a]	CBC at regular intervals
Interleukin-6 inhibitors	Tocilizumab: 4–8 mg/kg IV monthly OR 162 mg SQ every other week (<100 kg weight) 162 mg SQ every week (≥100 kg weight) Sarilumab: 200 mg SQ every other week	Risk of infection Infusion reaction LFT elevation Dyslipidemia Cytopenias		Tuberculosis screening	CBC and LFTs at regular intervals

(*Continued*)

TABLE 370-2 DMARDs Used for the Treatment of Rheumatoid Arthritis (Continued)

DRUG	DOSAGE	SERIOUS TOXICITIES	OTHER COMMON SIDE EFFECTS	INITIAL EVALUATION	MONITORING
JAK inhibitors	Tofacitinib: 5 mg orally twice daily OR 11 mg orally daily Upadacitinib: 15 mg orally daily Baricitinib: 2 mg orally daily	Risk of infection LFT elevation Dyslipidemia Neutropenia Thrombosis	Upper respiratory tract infections Diarrhea Headache Nasopharyngitis	Tuberculosis screening	CBC, LFTs, and lipids at regular intervals

[a]Viral hepatitis panel: hepatitis B surface antigen, hepatitis C viral antibody. [b]Tuberculosis screening can be performed using a Mantoux tuberculin skin test or blood interferon-gamma release assay.

Abbreviations: CBC, complete blood count; DMARDs, disease-modifying antirheumatic drugs; G6PD, glucose-6-phosphate dehydrogenase; IV, intravenous; LFTs, liver function tests; JAK, Janus kinase; SQ, subcutaneous.

monoclonal antibodies: infliximab, a chimeric (part mouse and human) monoclonal antibody; and adalimumab and golimumab, humanized monoclonal antibodies. Certolizumab pegol, a fourth product, is a pegylated Fab′ fragment of a humanized monoclonal antibody, while the other, etanercept, is a soluble fusion protein consisting of the TNF receptor 2 in covalent linkage with the Fc portion of IgG1. All of these TNF inhibitors have been shown in randomized controlled clinical trials to reduce the signs and symptoms of RA, slow radiographic progression of joint damage, and improve physical function and quality of life. Anti-TNF drugs may be used alone or in combination with background methotrexate therapy. Patients with RA who have an inadequate response to methotrexate therapy are often treated with a TNF inhibitor, usually in combination with methotrexate.

Anti-TNF agents should be avoided in patients with active infection or a history of hypersensitivity to these agents and are contraindicated in patients with chronic hepatitis B infection or class III/IV congestive heart failure. A major concern is the increased risk for infection, including serious bacterial infections, opportunistic fungal infection, and reactivation of latent tuberculosis infection (LTBI). For this reason, all patients are screened for LTBI according to national guidelines with either an intradermal injection of purified protein derivative (PPD) or an IFN-γ release assay prior to starting anti-TNF therapy **(Chap. 183)**. If positive, treatment for LTBI is given for at least 1 month prior to starting anti-TNF therapy.

Anakinra Anakinra is the recombinant form of the naturally occurring IL-1 receptor antagonist. Anakinra has seen limited use for the treatment of RA because other biologic DMARDs have seen greater efficacy in practice.

Abatacept Abatacept is a soluble fusion protein consisting of the extracellular domain of human CTLA-4 linked to the modified portion of human IgG. It inhibits the co-stimulation of T cells by blocking CD28-CD80/86 interactions and may also inhibit the function of antigen-presenting cells by reverse signaling through CD80 and CD86. Abatacept has been shown in clinical trials to reduce disease activity, slow radiographic progression of damage, and improve functional disability. Most patients receive abatacept in combination with a conventional DMARD. Abatacept therapy has been associated with an increased risk of infection and can inhibit a vaccine response.

Rituximab Rituximab is a chimeric monoclonal antibody directed against CD20, a cell-surface molecule expressed by most mature B lymphocytes. It works by depleting B cells, which, in turn, leads to a reduction in the inflammatory response by unknown mechanisms. These mechanisms may include a reduction in autoantibodies, inhibition of T-cell activation, and alteration of cytokine production. Rituximab has been approved for the treatment of refractory RA (failure of treatment with a TNF-α inhibitor) in combination with methotrexate and has been shown to be more effective for patients with seropositive than seronegative disease. Its administration has been associated with mild to moderate infusion reactions, as well as an increased risk of infection. Notably, there have been rare isolated reports of a potentially lethal brain disorder, progressive multifocal leukoencephalopathy (PML), in association with rituximab therapy, although the absolute risk of this complication appears to be very low in patients with RA (~1:25,000). Most of these cases have occurred on a background of previous or current exposure to other potent immunosuppressive drugs. Studies in recent years have also demonstrated that rituximab significantly inhibits a vaccine response (e.g., response to SARS-CoV-2 vaccination).

Anti-IL-6 Receptor Agents IL-6 is a proinflammatory cytokine implicated in the pathogenesis of RA, with effects on both joint inflammation and damage. IL-6 binding to its receptor activates intracellular signaling pathways that affect the acute-phase response, cytokine production, and osteoclast activation. Tocilizumab and sarilumab are both monoclonal antibodies directed against the membrane and soluble forms of the IL-6 receptor. Clinical trials attest to the clinical efficacy of these therapies for RA, both as monotherapy and in combination with methotrexate and other conventional DMARDs. Anti-IL-6 receptor agents have been associated with an increased risk of infection, neutropenia, and thrombocytopenia; the hematologic abnormalities appear to be reversible upon stopping the drug. In addition, this agent has been shown to increase LDL cholesterol. However, it is not known if this effect on lipid levels increases the risk for atherosclerotic disease. Anti-IL-6 receptor agents have also been associated with an increased rate of gastrointestinal perforation and should be avoided if possible in patients with a history of diverticulitis and a recent gastric or duodenal ulcer.

TARGETED SYNTHETIC DMARDs

Because treatment with conventional DMARDs and/or biologic therapies may not result in optimal disease control, other therapeutic strategies have been investigated to fill this gap. The so-called targeted synthetic DMARDs have been designed to target intracellular signaling pathways, which transduce the positive signals from cell-surface receptors activated by cytokines and other inflammatory mediators. They have an advantage over biologics in their oral formulation.

JAK Inhibitors Although several different kinases have been evaluated as potential treatment targets, only inhibitors of the JAKs have demonstrated safety and efficacy for the treatment of RA. The JAK family comprises four members (JAK1, JAK2, JAK3, and tyrosine kinase 2 [Tyk2]) and link a host of extracellular cytokine receptors with their intracellular signaling domains. They form homodimeric and heterodimeric duplexes that mediate signaling of the receptors for the common γ-chain-related cytokines IL-2, -4, -7, -9, -15, and -21, as well as IFN-γ, IL-6, IL-12, IL-23, IL-10, the type I IFNs, and the hemopoietic cytokines and growth factors. Many of these cytokines play critical roles in promoting T- and B-cell activation as well as chronic inflammation in RA.

Tofacitinib is a potent JAK1 and JAK3 inhibitor with minor inhibitory effects on JAK2 and Tyk2; baricitinib is a JAK1 and JAK2 inhibitor with moderate inhibition of Tyk2 and minimal inhibition of JAK3; and upadacitinib is a predominately selective inhibitor of JAK1. Each of the JAK inhibitors can be used as monotherapy or in combination with methotrexate. However, they are associated with an increased risk of infections, including bacterial infections and herpes zoster. Other possible side effects of these agents include elevated liver transaminases, neutropenia, increased cholesterol levels, hypertension, and elevations in serum creatinine. Recent studies have also found an increased risk of thrombosis, major adverse cardiovascular events, and malignancies in patients taking tofacitinib compared with TNF inhibitors, with the highest of these risks in the elderly.

TREATMENT OF EXTRAARTICULAR MANIFESTATIONS

In general, treatment of the underlying RA favorably modifies its extraarticular manifestations, and observational studies suggest aggressive management of early disease can potentially prevent their occurrence in established disease. The treatment of patients with RA-ILD, however, can be particularly challenging because some of the DMARDs used for the treatment of RA are also associated with pulmonary toxicity, such as methotrexate and leflunomide. RA-ILD is often treated with high doses of corticosteroids and adjunctive immunosuppressive agents, such as azathioprine, mycophenolate mofetil, or rituximab.

APPROACH TO THE PATIENT

Rheumatoid Arthritis

The treatment of RA adheres to the following principles and goals: (1) early, aggressive therapy to prevent joint damage and disability; (2) frequent modification of DMARD therapy to achieve treatment goals with utilization of combination therapy where appropriate; (3) individualization of DMARD therapy in an attempt to maximize response and minimize side effects; (4) minimal use of long-term glucocorticoid therapy; and (5) achieving, whenever possible, low disease activity or clinical remission. A considerable amount of evidence supports this intensive treatment approach and has been termed “treat to target,” with a target of achieving low disease activity or remission.

As mentioned earlier, methotrexate is the DMARD of first choice for initial treatment of moderate to severe RA. Failure to achieve adequate improvement with methotrexate therapy calls for a change in DMARD therapy, usually a transition to an effective combination regimen. Effective combinations include methotrexate, sulfasalazine, and hydroxychloroquine (oral triple therapy); methotrexate and leflunomide; and methotrexate plus a biological. The combination of methotrexate and an anti-TNF agent, for example, has been shown in randomized controlled trials to be superior to methotrexate alone, not only for reducing signs and symptoms of disease but also for retarding the progression of structural joint damage. Predicting which patients are at higher risk for developing radiologic joint damage is imprecise at best, although some factors such as an elevated serum level of acute-phase reactants, high burden of joint inflammation, and the presence of erosive disease are associated with increased likelihood of developing structural injury.

In 2021, the ACR updated their guidelines for the treatment of RA. These recommendations do not make a distinction between the treatment of patients with early (<6 months of disease duration) or established disease and highlight the use of a treat-to-target approach and the need to switch or add therapies for worsening or persistent moderate/high disease activity. For example, in patients with early RA who have persistent moderate/high disease activity despite DMARD monotherapy, providers should consider escalation to combination DMARD therapy or switching to an anti-TNF +/− methotrexate or a non-TNF biologic +/− methotrexate. Since a more intensive initial approach (e.g., combination DMARD therapy) has been shown to produce superior long-term outcomes compared with starting methotrexate alone, the usual approach is to begin with methotrexate and, in the absence of an adequate therapeutic response, rapidly step up (e.g., after 3–6 months) to a combination of conventional DMARDs or add an anti-TNF or non-TNF biological agent.

Some patients may not respond to an anti-TNF drug or may be intolerant of its side effects. Initial responders to an anti-TNF agent who later experience worsening of their condition may benefit from switching to another anti-TNF agent or an alternative biological with a different mechanism of action. Indeed, some studies suggest that switching to an alternative biological such as abatacept is more effective than switching to another anti-TNF drug. Unacceptable toxicity from an anti-TNF agent may also call for switching to another biological or targeted synthetic DMARD with a different mechanism of action or a conventional DMARD regimen.

Studies have also shown that oral triple therapy (hydroxychloroquine, methotrexate, and sulfasalazine) may be used effectively for the treatment of early RA. Treatment may be initiated with methotrexate alone and, lacking an adequate treatment response, followed within 6 months by a step-up to oral triple therapy.

A clinical state defined as low disease activity or remission is the optimal goal of therapy, although most patients never achieve complete remission despite every effort to achieve it. Composite indices, such as the Disease Activity Score-28 (DAS-28), are useful for classifying states of low disease activity and remission; however, they are imperfect tools due to the limitations of the clinical joint examination in which low-grade synovitis may escape detection. Complete remission has been stringently defined as the total absence of all articular and extraarticular inflammation and immunologic activity related to RA. However, evidence for this state can be difficult to demonstrate in clinical practice. In an effort to standardize and simplify the definition of remission for clinical trials, the ACR and EULAR developed two provisional operational definitions of remission in RA **(Table 370-3)**. A patient may be considered in remission if the patient (1) meets all the clinical and laboratory criteria listed in Table 370-3 or (2) has a composite SDAI score of <3.3. The SDAI is calculated by taking the sum of a tender joint and swollen joint count (using 28 joints), patient global assessment (0–10 scale), physician global assessment (0–10 scale), and CRP (in mg/dL). This definition of remission does not consider the possibility of subclinical synovitis or that damage alone may produce a tender or swollen joint. Ignoring the semantics of these definitions, the aforementioned remission criteria are nonetheless useful for setting a level of disease control that will likely result in minimal or no progression of structural damage and disability. With an early, aggressive approach to treatment, a significant number of patients who reach remission may have a reduction in treatment without an increased risk of flare or loss of remission. However, complete withdrawal of DMARD therapy during established disease almost invariably leads to a relapse.

TABLE 370-3 ACR/EULAR Provisional Definition of Remission in Rheumatoid Arthritis

At any time point, patient must satisfy all of the following:
- Tender joint count ≤1
- Swollen joint count ≤1
- C-reactive protein ≤1 mg/dL
- Patient global assessment ≤1 (on a 0–10 scale)

OR

At any time point, patient must have a Simplified Disease Activity Index score of ≤3.3

Abbreviations: ACR, American College of Rheumatology; EULAR, European League Against Rheumatism.

Source: Reproduced with permission from DT Felson et al; American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Arthritis Rheum 63:573, 2011.

PHYSICAL ACTIVITY AND ASSISTIVE DEVICES

In principle, all patients with RA should receive a prescription for exercise and physical activity. The ACR has developed guidelines for integrative interventions to accompany disease-modifying therapy and includes evidence-based recommendations for diet and exercise. Dynamic strength training, community-based comprehensive physical therapy, and physical-activity coaching (emphasizing achieving 150 min of moderate-to-vigorous physical activity per week) have all been shown to improve muscle strength and perceived health status, as well as improve DAS-28 scores and inflammatory markers. Foot orthotics for painful valgus deformity decrease foot pain and may reduce disability and functional limitations. Judicious use of wrist splints can also decrease pain; however, their benefits may be offset by decreased dexterity and variably curb grip strength.

SURGERY

Surgical procedures may improve pain and disability in RA with varying degrees of reported long-term success—most notably for the hands, wrists, and feet. For large joints, such as the knee, hip, shoulder, or elbow, the preferred option for advanced joint disease may be total joint arthroplasty. A few surgical options exist for dealing with the smaller hand joints. Silicone implants are the most common prosthetic for MCP arthroplasty and are generally implanted in patients with severe decreased arc of motion, marked flexion contractures, MCP joint pain with radiographic abnormalities, and severe ulnar drift. Arthrodesis and total wrist arthroplasty are reserved for patients with severe disease who have substantial pain and functional impairment. Numerous surgical options exist for correction of hallux valgus in the forefoot, including arthrodesis and arthroplasty, as well as primarily arthrodesis for refractory hindfoot pain.

OTHER MANAGEMENT CONSIDERATIONS

Pregnancy Up to 75% of female RA patients will note overall improvement in symptoms during pregnancy but they will often flare shortly after delivery. Flares during pregnancy are generally treated with low doses of prednisone. Hydroxychloroquine and sulfasalazine are probably the safest DMARDs to use during pregnancy, although anti-TNF agents are generally accepted to be safe in this setting. However, current guidelines call for anti-TNF drugs to be discontinued during the third trimester of pregnancy. Methotrexate and leflunomide therapy are contraindicated during pregnancy due to their teratogenicity in animals and humans. The experience with other biologic agents, such as abatacept, anti-IL-6 receptor antibodies, rituximab, and JAK inhibitors has been insufficient to make specific recommendations for their use during pregnancy and are discontinued at conception. Active inflammatory disease is associated with worse pregnancy outcomes, and thus, controlling disease activity remains a priority.

Elderly Patients RA presents in up to one-third of patients after the age of 60; however, older individuals may receive less aggressive treatment due to concerns about increased risks of drug toxicity. Studies suggest that conventional DMARDs and biological agents are equally effective and safe in younger and older patients (with the possible exception of JAK inhibitors). Due to comorbidities, many elderly patients have an increased risk of infection. Aging also leads to a gradual decline in renal function that may raise the risk for side effects from NSAIDs and some DMARDS, such as methotrexate. Renal function must be taken into consideration before prescribing methotrexate, which is mostly cleared by the kidneys. To reduce the risks of side effects, methotrexate doses may need to be adjusted downward for the drop in renal function that usually comes with the seventh and eighth decades of life.

GLOBAL CHALLENGES

Developing countries are finding an increase in the incidence of noncommunicable, chronic diseases such as diabetes, cardiovascular disease, and RA in the face of ongoing poverty, rampant infectious disease, and poor access to modern health care facilities. In these areas, patients tend to have a greater delay in diagnosis and limited access to specialists and, thus, greater disease activity and disability at presentation. In addition, infection risk remains a significant issue for the treatment of RA in developing countries because of the immunosuppression associated with the use of glucocorticoids and most DMARDs. For example, in some developing countries, patients undergoing treatment for RA have a substantial increase in the incidence of tuberculosis, which demands the implementation of far more comprehensive screening practices and liberal use of isoniazid prophylaxis than in developed countries. The increased prevalence of hepatitis B and C, as well as human immunodeficiency virus (HIV), in these developing countries also poses challenges. Reactivation of viral hepatitis has been observed in association with some of the DMARDs, such as rituximab. Also, reduced access to antiretroviral therapy may limit the control of HIV infection and therefore the choice of DMARD therapies.

Despite these challenges, one should attempt to initiate early treatment of RA in the developing countries with the resources at hand. Hydroxychloroquine, sulfasalazine, and methotrexate are all reasonably accessible throughout the world where they can be used as both monotherapy and in combination with other drugs. The availability of biological agents is increasing in the developed countries as well as in other areas around the world, although their use is limited by high cost; national protocols restrict their use, and concerns remain about the risk for opportunistic infections.

SUMMARY

Improved understanding of the pathogenesis of RA and its treatment has dramatically revolutionized the management of this disease. The outcomes of patients with RA are vastly superior to those of the prebiologic modifier era; more patients than in years past are able to avoid significant disability and continue working, albeit with some job modifications. The need for early and aggressive treatment of RA as well as frequent follow-up visits for monitoring of drug therapy has implications for our health care system. Primary care physicians and rheumatologists must be prepared to work together as a team to implement a treat-to-target approach with the goal of reaching remission or low disease activity. In many settings, rheumatologists have reengineered their practice in a way that places high priority on consultations for any new patient with early inflammatory arthritis.

The therapeutic regimens for RA are becoming increasingly complex with the rapidly expanding armamentarium. Patients receiving these therapies must be carefully monitored by both the primary care physician and the rheumatologist to minimize the risk of side effects and identify quickly any complications of chronic immunosuppression. Also, prevention and treatment of RA-associated conditions such as ischemic heart disease and osteoporosis will likely benefit from a team approach owing to the value of multidisciplinary care.

Research will continue to search for new therapies with superior efficacy and safety profiles and investigate treatment strategies that can bring the disease under control more rapidly and nearer to remission. However, prevention and cure of RA will likely require new breakthroughs in our understanding of disease pathogenesis. Several prevention trials in RA are currently underway and focus on a variety of prevention strategies in individuals who have serologic and/or clinical features at higher risk than the general population for developing RA. Equally important is the identification of predictive biomarkers in RA that will enable a personalized approach to DMARD therapy, a foreseeable goal in the future that will take advantage of the advances in technology and scientific understanding of the disease.

■ FURTHER READING

Aletaha D, Smolen JS: Diagnosis and management of rheumatoid arthritis: A review. JAMA 320:1360, 2018.

Catrina AI et al: Mechanisms leading from systemic autoimmunity to joint-specific disease in rheumatoid arthritis. Nat Rev Rheumatol 13:79, 2017.

Erickson AR et al: Clinical features of rheumatoid arthritis, in *Kelley and Firestein's Textbook of Rheumatology*, 10th ed. Firestein GS et al (eds). Philadelphia, Elsevier, 2017, pp 1167–1186.

Fraenkel L et al: 2021 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Rheumatol 73:1108, 2021.

Gravallese EM, Firestein GS: Rheumatoid arthritis: Common origins, divergent mechanisms. N Engl J Med 388:529, 2023.

Karimi J et al: Genetic implications in the pathogenesis of rheumatoid arthritis; an updated review. Gene 702:8, 2019.

Moreland LW et al: A randomized comparative effectiveness study of oral triple therapy versus methotrexate plus etanercept in early aggressive rheumatoid arthritis: The Treatment of Early Aggressive Rheumatoid Arthritis Trial. Arthritis Rheum 64:2824, 2012.

VIDEO 370-1 Transverse view of 2nd MCP demonstrating synovial hypertrophy with abnormal power doppler signal (yellow) consistent with synovitis. Synovial proliferation is displacing the overlying extensor tendon of the digit above the joint space. This view demonstrates small erosions of the dorsal metacarpal head. *(Courtesy of Dr. Philip Chu.)*

VIDEO 370-2 Longitudinal dorsal view of the 2nd MCP demonstrating synovial hypertrophy with abnormal power doppler signal (yellow) consistent with synovitis. *(Courtesy of Dr. Philip Chu)*

VIDEO 370-3 Transverse view of 2nd MCP demonstrating synovial hypertrophy and intrasynovial effusion with abnormal power doppler signal (yellow) consistent with synovitis. Fluid and synovial proliferation is displacing the overlying extensor tendon of the digit above the joint space. *(Courtesy of Dr. Philip Chu.)*

371

Acute Rheumatic Fever

Joseph Kado, Jonathan Carapetis

Acute rheumatic fever (ARF) is a multisystem disease resulting from an autoimmune reaction to infection with group A *Streptococcus*. Although many parts of the body may be affected, almost all of the manifestations resolve completely. The major exception is cardiac valvular damage (rheumatic heart disease [RHD]), which may persist after the other features have disappeared.

GLOBAL CONSIDERATIONS

ARF and RHD are diseases of poverty. They were common in all countries until the early twentieth century, when their incidence began to decline in industrialized nations. This decline was largely attributable to improved living conditions—particularly less crowded housing and better hygiene—which resulted in reduced transmission of group A streptococci. The introduction of antibiotics and improved systems of medical care had a supplemental effect.

The virtual disappearance of ARF and reduction in the incidence of RHD in high-income countries during the first half of the twentieth century unfortunately was not replicated in low- and middle-income countries (LMICs), where these diseases continue unabated. RHD is the most common cause of acquired heart disease in children in LMICs and is a major cause of mortality and morbidity in adults as well. It is estimated that >40 million people worldwide are affected by RHD, with >300,000 deaths occurring each year. Some 95% of ARF cases and RHD deaths now occur in developing countries, with particularly high burdens in sub-Saharan Africa, Pacific nations, Australasia, China, and South and Central Asia. The pathogenetic pathway from exposure to group A *Streptococcus* followed by pharyngeal or superficial skin infection and subsequent development of ARF, ARF recurrences, and development of RHD and its complications is associated with a range of risk factors and, therefore, potential interventions at each point **(Fig. 371-1)**. In affluent countries, many of these risk factors are well controlled, and where needed, interventions are in place. Unfortunately, the greatest burden of disease is found in LMICs, most of which do not have the resources, capacity, and/or interest to tackle this multifaceted disease. In particular, few of these countries have coordinated, register-based RHD control programs, which have been proven to be cost-effective in reducing the burden of RHD. Enhancing awareness of RHD and mobilizing resources for its control in LMICs are issues requiring international attention. In 2018, member states of the World Health Organization (WHO) unanimously adopted a Global Resolution on Rheumatic Fever and Rheumatic Heart Disease, calling on all states as well as international stakeholders and the WHO itself to take practical actions to control these diseases.

EPIDEMIOLOGY

ARF is mainly a disease of children aged 5–14 years. Initial episodes become less common in older adolescents and young adults and are rare in persons aged >30 years. By contrast, recurrent episodes of ARF remain relatively common in adolescents and young adults. This pattern contrasts with the prevalence of RHD, which peaks between 25 and 40 years. There is no clear gender association for ARF, but RHD more commonly affects females, sometimes up to twice as frequently as males.

PATHOGENESIS

ORGANISM FACTORS

Conventional teaching has it that ARF is exclusively caused by infection of the upper respiratory tract with group A streptococci **(Chap. 153)**. Although classically, certain M-serotypes (particularly types 1, 3, 5, 6, 14, 18, 19, 24, 27, and 29) were associated with ARF, recent evidence demonstrates that many more M-serotypes are rheumatogenic and that so-called “rheumatogenic motifs” are found in only a minority of serotypes associated with rheumatic fever. This epidemiologic evidence also points to a clear role of skin infection in the pathogenesis of ARF. The potential role of groups C and G streptococci is unclear at this time.

HOST FACTORS

Based on epidemiologic evidence, ~3–6% of any population may be susceptible to ARF, and this proportion does not vary dramatically between populations. Findings of familial clustering of cases and concordance in monozygotic twins—particularly for chorea—confirm that susceptibility to ARF is an inherited characteristic, with 44% concordance in monozygotic twins compared to 12% in dizygotic twins and heritability more recently estimated at 60%. Most evidence for host factors focuses on immunologic determinants. Initial studies found associations with human leukocyte antigen (HLA) class II alleles, some protective and some associated with increased susceptibility, as well as polymorphisms in tumor necrosis factor and mannose binding lectin. Recent genome-wide association studies and genomic sequencing analyses have identified associations with a range of genes including the immunoglobulin heavy chain (IGH) locus (specifically the IGHV4-61*02 allele), complement factor H, and some HLA class II (a range of HLA-DQ A and B alleles) and III loci. The associations are often population dependent, although increasing consistency is being found across populations in meta-analyses of genomic studies. Over coming years, further genomic analyses are expected to provide additional insights into ARF and RHD pathogenesis, as well as host protective and susceptibility factors.

THE IMMUNE RESPONSE

The most widely accepted theory of rheumatic fever pathogenesis is based on the concept of molecular mimicry, whereby an immune response targeted at streptococcal antigens (mainly thought to be on the M protein) also recognizes human tissues. In this model, antigen processing cells of the innate immune system present streptococcal antigens after throat (and possibly skin) group A streptococcal infection to T cells, which then lead to activation of both humoral and cellular immunity. Cross-reactive antibodies bind to endothelial cells on the heart valve, leading to activation of the adhesion molecule VCAM-1, with resulting recruitment of activated lymphocytes and lysis of endothelial cells in the presence of complement. The latter leads to release of peptides including laminin, keratin, and tropomyosin, which, in turn, activates cross-reactive T cells that invade the heart, amplifying the damage and causing epitope spreading. An alternative hypothesis proposes that the initial damage is due to streptococcal invasion of epithelial surfaces, with binding of M protein to type IV collagen

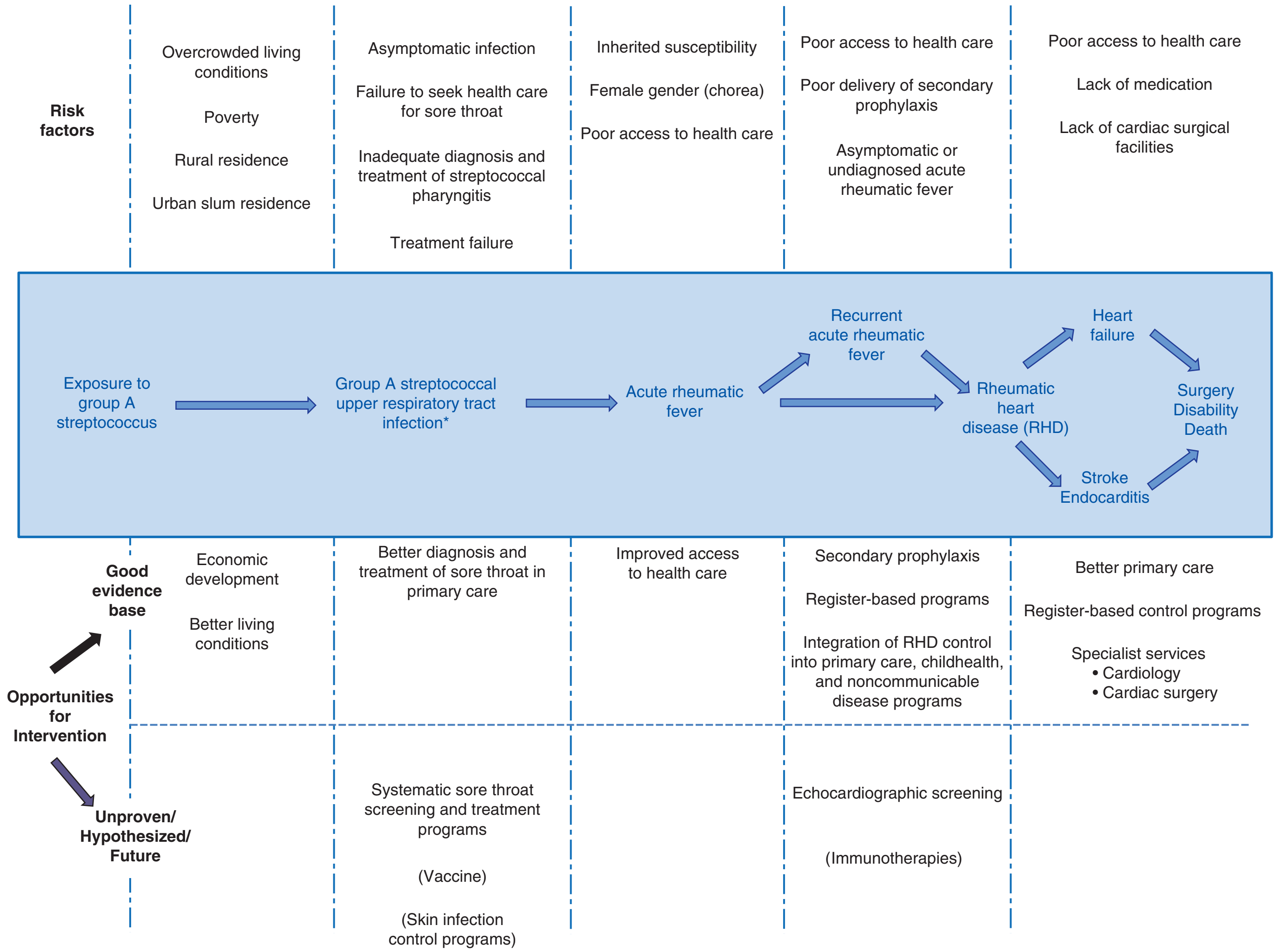
Risk factors
Overcrowded living conditions
Poverty
Rural residence
Urban slum residence
Asymptomatic infection
Failure to seek health care for sore throat
Inadequate diagnosis and treatment of streptococcal pharyngitis
Treatment failure
Inherited susceptibility
Female gender (chorea)
Poor access to health care
Poor access to health care
Poor delivery of secondary prophylaxis
Asymptomatic or undiagnosed acute rheumatic fever
Poor access to health care
Lack of medication
Lack of cardiac surgical facilities
Exposure to group A streptococcus
Group A streptococcal upper respiratory tract infection*
Acute rheumatic fever
Recurrent acute rheumatic fever
Rheumatic heart disease (RHD)
Heart failure
Stroke Endocarditis
Surgery Disability Death
Good evidence base
Opportunities for Intervention
Unproven/ Hypothesized/ Future
Economic development
Better living conditions
Better diagnosis and treatment of sore throat in primary care
Improved access to health care
Secondary prophylaxis
Register-based programs
Integration of RHD control into primary care, childhealth, and noncommunicable disease programs
Better primary care
Register-based control programs
Specialist services
• Cardiology
• Cardiac surgery
Systematic sore throat screening and treatment programs
(Vaccine)
(Skin infection control programs)
Echocardiographic screening
(Immunotherapies)

*Increasing evidence of the role of streptococcal skin infection

FIGURE 371-1 Pathogenetic pathway for acute rheumatic fever and rheumatic heart disease (RHD), with associated risk factors and opportunities for intervention at each step. Interventions in *parentheses* are either unproven or currently unavailable.

allowing it to become immunogenic, but not through the mechanism of molecular mimicry.

CLINICAL FEATURES

There is a latent period of ~3 weeks (1–5 weeks) between the precipitating group A streptococcal infection and the appearance of the clinical features of ARF. The exceptions are chorea and indolent carditis, which may follow prolonged latent periods lasting up to 6 months. Although many patients report a prior sore throat, the preceding group A streptococcal infection is commonly subclinical; in these cases, it can only be confirmed using streptococcal antibody testing. The most common clinical features are polyarthritis (present in 60–75% of cases) and carditis (50–75%). The prevalence of chorea in ARF varies substantially between populations, ranging from <2 to 30%. Erythema marginatum and subcutaneous nodules are now rare, being found in <5% of cases.

■ HEART INVOLVEMENT

Up to 75% of patients with ARF progress to RHD. The endocardium, pericardium, or myocardium may be affected. Valvular damage is the hallmark of rheumatic carditis. The mitral valve is almost always affected, sometimes together with the aortic valve; isolated aortic valve involvement is rare. Damage to the pulmonary or tricuspid valves is usually secondary to increased pulmonary pressures resulting from left-sided valvular disease. Early valvular damage leads to regurgitation. Over ensuing years, usually as a result of recurrent episodes, leaflet thickening, scarring, calcification, and valvular stenosis may develop **(Fig. 371-2)**. **See Videos 371-1 and 371-2.** Therefore, the characteristic manifestation of carditis in previously unaffected individuals is mitral regurgitation, sometimes accompanied by aortic regurgitation. Myocardial inflammation may affect electrical conduction pathways, leading to P-R interval prolongation (first-degree atrioventricular block or rarely higher-level block) and softening of the first heart sound.

People with RHD are often asymptomatic for many years before their valvular disease progresses to cause cardiac failure. Moreover, particularly in resource-poor settings, the diagnosis of ARF is often not made, so children, adolescents, and young adults may have RHD but not know it. These cases can be diagnosed using echocardiography; auscultation is poorly sensitive and specific for RHD diagnosis in asymptomatic patients. Echocardiographic screening of school-aged children in populations with high rates of RHD is becoming more widespread and has been facilitated by improving technologies in portable echocardiography and the availability of consensus guidelines for the diagnosis of RHD on echocardiography **(Table 371-1)**. These guidelines replace the previous “definite,” “borderline,” and “latent” diagnostic category terms with a classification based on the risk of progression to more advanced valvular heart disease and provide recommendations on secondary prophylaxis for each group.

■ JOINT INVOLVEMENT

The most common form of joint involvement in ARF is arthritis, i.e., objective evidence of inflammation, with hot, swollen, red, and/or tender joints, and involvement of more than one joint (i.e., polyarthritis). Polyarthritis is typically migratory, moving from one joint to another over a period of hours. ARF almost always affects the large joints—most commonly the knees, ankles, hips, and elbows—and is asymmetric. The pain is severe and usually disabling until anti-inflammatory medication is commenced.

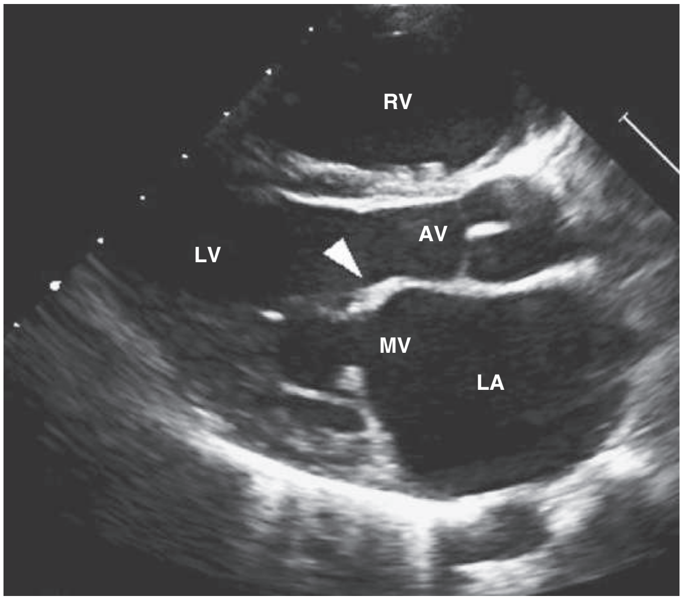

FIGURE 371-2 Transthoracic echocardiographic image from a 5-year-old boy with chronic rheumatic heart disease. This diastolic image demonstrates leaflet thickening, restriction of the anterior mitral valve leaflet tip, and doming of the body of the leaflet toward the interventricular septum. This appearance (marked by the *arrowhead*) is commonly described as a "hockey stick" or an "elbow" deformity. AV, aortic valve; LA, left atrium; LV, left ventricle; MV, mitral valve; RV, right ventricle. *(Courtesy of Dr. Bo Remenyi, Department of Paediatric and Congenital Cardiac Services, Starship Children's Hospital, Auckland, New Zealand.)*

Less severe joint involvement is also relatively common and has been recognized as a potential major manifestation in high-risk populations in the most recent revision of the Jones criteria. Arthralgia without objective joint inflammation usually affects large joints in the same migratory pattern as polyarthritis. In some populations, aseptic monoarthritis may be a presenting feature of ARF, which may, in turn, result from early commencement of anti-inflammatory medication before the typical migratory pattern is established.

The joint manifestations of ARF are highly responsive to salicylates and other nonsteroidal anti-inflammatory drugs (NSAIDs).

CHOREA

Sydenham's chorea commonly occurs in the absence of other manifestations, follows a prolonged latent period after group A streptococcal infection, and is found mainly in females. The choreiform movements affect particularly the head (causing characteristic darting movements of the tongue) and the upper limbs **(Chap. 447)**. They may be generalized or restricted to one side of the body (hemi-chorea). In mild cases, chorea may be evident only on careful examination, whereas in the most severe cases, the affected individuals are unable to perform activities of daily living. There is often associated emotional lability or obsessive-compulsive traits, which may last longer than the choreiform movements (which usually resolve within 6 weeks but sometimes may take up to 6 months). More than 50% of patients presenting with chorea will have carditis, for which reason echocardiography should be part of the workup.

SKIN MANIFESTATIONS

The classic rash of ARF is *erythema marginatum* **(Chap. 21)**, which begins as pink macules that clear centrally, leaving a serpiginous, spreading edge. The rash is evanescent, appearing and disappearing before the examiner's eyes. It occurs usually on the trunk, sometimes on the limbs, but almost never on the face.

Subcutaneous nodules occur as painless, small (0.5–2 cm), mobile lumps beneath the skin overlying bony prominences, particularly of the hands, feet, elbows, occiput, and occasionally the vertebrae. They are a delayed manifestation, appearing 2–3 weeks after the onset of disease, last for just a few days up to 3 weeks, and are commonly associated with carditis.

TABLE 371-1 Staging of Rheumatic Heart Disease Detected by Echocardiography[a,b]

Stage A: Minimal Echocardiographic Criteria for RHD
Applies only to individuals aged ≤20 years old • Clinical risk: might be at risk of valvular heart disease progression • Echocardiographic features: the presence of mild[c] MR or AR without morphologic features
Stage B: Mild RHD
Can apply to any age • Clinical risk: at moderate or high risk of progression and at risk of developing symptoms of valvular heart disease • Echocardiographic features: evidence of mild[c] valvular regurgitation plus at least one morphologic feature in individuals aged ≤20 years and at least two morphologic features in individuals aged >20 years[d]; or mild regurgitation in both mitral and aortic valves
Stage C: Advanced RHD at Risk of Clinical Complications
Can apply to any age • Clinical risk: at high risk of developing clinical complications that require medical or surgical intervention • Echocardiographic features: moderate or severe MR, moderate or severe AR, any MS or AS[e], pulmonary hypertension, and decreased LV systolic function
Stage D: Advanced RHD with Clinical Complications
Can apply to any age • Clinical risk: established clinical complications include cardiac surgery, heart failure, arrhythmia, stroke, and infective endocarditis • Echocardiographic features: moderate or severe MR, moderate or severe AR, any MS or AS[e], pulmonary hypertension, and decreased LV systolic function

[a]To be applied in high-risk settings and requires other causes of valvular heart disease to have been excluded. [b]After the application of the confirmatory echocardiographic criteria, diagnostic categories might include "normal" and "other," which encompasses other diseases such as congenital heart disease, cardiomyopathies, and pericardial effusion. [c]Fulfilling the confirmatory criteria for pathologic regurgitation (see Source). [d]This cutoff value is derived from expert consensus. [e]Aortic stenosis is defined in accordance with international guidelines on valvular heart disease. A diagnosis of rheumatic aortic stenosis requires the exclusion of other causes, including bicuspid aortic valve and degenerative calcific AS.

Abbreviations: AR, aortic regurgitation; AS, aortic stenosis; LV, left ventricular; MR, mitral regurgitation; MS, mitral stenosis, RHD, rheumatic heart disease.

Source: Reproduced with permission from J Rwebembera et al: 2023 World Heart Federation guidelines for the echocardiographic diagnosis of rheumatic heart disease. Nat Rev Cardiol 21:250; 2023.

OTHER FEATURES

Fever occurs in most cases of ARF, although rarely in cases of pure chorea. Although high-grade fever (≥39°C) is the rule, lower grade temperature elevations are not uncommon. Elevated acute-phase reactants are also present in most cases.

EVIDENCE OF A PRECEDING GROUP A STREPTOCOCCAL INFECTION

With the exception of chorea and low-grade carditis, both of which may become manifest many months later, evidence of a preceding group A streptococcal infection is essential in making the diagnosis of ARF. Because most cases do not have a positive throat swab culture or rapid antigen test, serologic evidence is usually needed. The most common serologic tests are the anti-streptolysin O (ASO) and anti-DNase B (ADB) titers. Where possible, age-specific reference ranges should be determined in a local population of healthy people without a recent group A streptococcal infection.

CONFIRMING THE DIAGNOSIS

Because there is no definitive test, the diagnosis of ARF relies on the presence of a combination of typical clinical features together with evidence of the precipitating group A streptococcal infection, and the exclusion of other diagnoses. This uncertainty led Dr. T. Duckett Jones in 1944 to develop a set of criteria (subsequently known as the *Jones criteria*) to aid in the diagnosis. The most recent revision of the Jones criteria **(Table 371-2)** requires the clinician to determine if the patient is from a setting or population known to experience low rates of ARF.

TABLE 371-2 Jones Criteria

A. For All Patient Populations with Evidence of Preceding Group A Streptococcal Infection	
Diagnosis: initial ARF	2 major manifestations or 1 major plus 2 minor manifestations
Diagnosis: recurrent ARF	2 major or 1 major and 2 minor or 3 minor
B. Major Criteria	
Low-risk populations[a]	Moderate- and high-risk populations
Carditis[b]	Carditis
• Clinical and/or subclinical	• Clinical and/or subclinical
Arthritis	Arthritis
• Polyarthritis only	• Monoarthritis or polyarthritis
	• Polyarthralgia[c]
Chorea	Chorea
Erythema marginatum	Erythema marginatum
SC nodules	SC nodules
C. Minor Criteria	
Low-risk populations[a]	Moderate- and high-risk populations
Polyarthralgia	Monoarthralgia
Fever (≥38.5°C)	Fever (≥38°C)
ESR ≥60 mm in the first hour and/or CRP ≥3.0 mg/dL[d]	ESR ≥30 mm/h and/or CRP ≥3.0 mg/dL[d]
Prolonged PR interval[e], after accounting for age variability (unless carditis is a major criterion)	Prolonged PR interval[e], after accounting for age variability (unless carditis is a major criterion)

[a] Low-risk populations are those with ARF incidence ≤2 per 100,000 school-age children or all-age rheumatic heart disease prevalence of ≤1 per 1000 population per year. [b] Subclinical carditis indicates echocardiographic valvulitis. (See source document.) [c] Polyarthralgia should only be considered as a major manifestation in moderate- to high-risk populations after exclusion of other causes. As in past versions of the criteria, erythema marginatum and SC nodules are rarely "stand-alone" major criteria. Additionally, joint manifestations can only be considered in either the major or minor categories but not both in the same patient. (See source document for more information.) [d] CRP value must be greater than upper limit of normal for laboratory. Also, because ESR may evolve during the course of ARF, peak ESR values should be used. [e] Prolonged PR interval can only be considered in the absence of carditis as a major criterion.

Abbreviations: ARF, acute rheumatic fever; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; SC, subcutaneous.

Source: Reproduced with permission from MH Gewitz et al: Revision of the Jones criteria for the diagnosis of acute rheumatic fever in the era of Doppler echocardiography: A scientific statement from the American Heart Association. Circulation 131(20):1806, 2015. *https://www.ahajournals.org/doi/full/10.1161/CIR.0000000000000205.*

For this group, there is a set of "low-risk" criteria; for all others, there is a set of more sensitive criteria.

TREATMENT

Acute Rheumatic Fever

Patients with possible ARF should be followed closely to ensure that the diagnosis is confirmed, treatment of heart failure and other symptoms is undertaken, and preventive measures including commencement of secondary prophylaxis, inclusion on an ARF registry, and health education are commenced. Echocardiography should be performed on all possible cases to aid in making the diagnosis and to determine the severity at baseline of any carditis. Other tests that should be performed are listed in **Table 371-3**.

There is no treatment for ARF that has been proven to alter the likelihood of developing, or the severity of, RHD. With the exception of treatment of heart failure, which may be lifesaving in cases of severe carditis, the treatment of ARF is symptomatic.

ANTIBIOTICS

All patients with ARF should receive antibiotics sufficient to treat the precipitating group A streptococcal infection **(Chap. 153)**. Penicillin is the drug of choice and can be given orally (as phenoxymethyl penicillin, 500 mg [250 mg for children ≤27 kg] PO twice daily, or amoxicillin, 50 mg/kg [maximum, 1 g] daily, for 10 days) or as a single dose of 1.2 million units (600,000 units for children ≤27 kg) IM benzathine penicillin G.

TABLE 371-3 Testing and Monitoring of ARF in the Acute Setting

Investigations

Always request:

- Electrocardiogram (ECG)
- Echocardiogram
- Complete blood count (CBC)
- C-reactive protein (CRP)
- Streptococcal serology (antistreptolysin and anti-DNase B)

In relevant situations:

- Throat swab
- Skin sore swab
- Blood cultures
- Synovial fluid aspirate
  - Ensure sample does not clot by using correct tubes that have been well mixed and transported promptly to the laboratory
  - Include request for cell count, microscopy, culture, and gonococcal polymerase chain reaction (PCR)
- Pregnancy test
- Creatinine test (UEC [urea, electrolytes, creatinine]) since nonsteroidal anti-inflammatory drugs can affect renal function

Tests to exclude alternative diagnoses, depending on clinical presentation and locally endemic infections:

- Autoantibodies, double-stranded DNA, anti–cyclic citrullinated peptide (anti-CCP) antibodies
- Urine for *Neisseria gonorrhoeae* molecular test
- Urine for *Chlamydia trachomatis* molecular test
- Serologic or other testing for viral hepatitis, *Yersinia* spp., cytomegalovirus (CMV), parvovirus B19, respiratory viruses, Ross River virus, Barmah Forest virus

Source: Reproduced with permission from RDHAustralia, Menzies School of Health Research. RHDAustralia (ARF/RHD writing group). The 2020 Australian guideline for prevention, diagnosis and management of acute rheumatic fever and rheumatic heart disease (3rd edition); 2020. Available at *https://www.rhdaustralia.org.au/arf-rhd-guideline.*

SALICYLATES AND NSAIDS

These may be used for the treatment of arthritis, arthralgia, and fever once the diagnosis is confirmed. They are of no proven value in the treatment of carditis or chorea. Aspirin has traditionally been the first-line choice, delivered at a dose of 50–60 mg/kg per day, up to a maximum of 80–100 mg/kg per day (4–8 g/d in adults) in 4–5 divided doses. At higher doses, the patient should be monitored for symptoms of salicylate toxicity such as nausea, vomiting, or tinnitus; if symptoms appear, lower doses should be used. Owing to the frequency of gastrointestinal side effects and the potential of more severe adverse effects of aspirin, many clinicians now prefer to use naproxen at a dose of 10–20 mg/kg per day because it may be safer than aspirin and has the advantage of twice-daily dosing. When the acute symptoms are substantially resolved, usually within the first 2 weeks, patients on higher doses of anti-inflammatory medications can have the dose reduced for a further 2–4 weeks. Fever, joint manifestations, and elevated acute-phase reactants sometimes recur up to 3 weeks after the medication is discontinued. This does not indicate a recurrence and can be managed by recommencing anti-inflammatory agents for a brief period.

CONGESTIVE HEART FAILURE

Glucocorticoids The use of glucocorticoids in ARF remains controversial. Two meta-analyses have failed to demonstrate a benefit of glucocorticoids compared to placebo or salicylates in improving the short- or longer-term outcome of carditis. However, the studies included in these meta-analyses all took place >40 years ago and did not use medications in common usage today. There are some recent data that suggest corticosteroids improve laboratory, radiologic, and echocardiographic parameters in carditis. Many clinicians treat cases of severe carditis (causing heart failure) with glucocorticoids in the belief that they may reduce the acute inflammation and result in more rapid resolution of failure. However, the potential benefits of this treatment should be balanced against the possible adverse effects. If used, prednisone or prednisolone is recommended at a dose of 1–2 mg/kg per day (maximum, 80 mg), usually for a few days or up to a maximum of 3 weeks.

MANAGEMENT OF HEART FAILURE

See Chap. 265.

BED REST

Traditional recommendations for long-term bed rest, once the cornerstone of management, are no longer widely practiced. Instead, bed rest should be prescribed as needed while arthritis and arthralgia are present and for patients with heart failure. Once symptoms are well controlled, gradual mobilization can commence as tolerated.

CHOREA

Medications to control the abnormal movements do not alter the duration or outcome of chorea. Milder cases can usually be managed by providing a calm environment. In patients with severe chorea, carbamazepine or sodium valproate is preferred to haloperidol. A response may not be seen for 1–2 weeks, and medication should be continued for 1–2 weeks after symptoms subside. There is recent evidence that corticosteroids are effective and lead to more rapid symptom reduction in chorea. They should be considered in severe or refractory cases. Prednisone or prednisolone may be commenced at 0.5 mg/kg daily, with weaning as early as possible, preferably after 1 week if symptoms are reduced, although slower weaning or temporary dose escalation may be required if symptoms worsen.

INTRAVENOUS IMMUNOGLOBULIN (IVIG)

Small studies have suggested that IVIg may lead to more rapid resolution of chorea but have shown no benefit on the short- or long-term outcome of carditis in ARF without chorea. In the absence of better data, IVIg is *not* recommended except in cases of severe chorea refractory to other treatments.

PROGNOSIS

Untreated, ARF lasts on average 12 weeks. With treatment, patients are usually discharged from hospital within 1–2 weeks. Inflammatory markers should be monitored every 1–2 weeks until they have normalized (usually within 4–6 weeks), and an echocardiogram should be performed after 1 month to determine if there has been progression of carditis. Cases with more severe carditis need close clinical and echocardiographic monitoring in the longer term.

Once the acute episode has resolved, the priority in management is to ensure long-term clinical follow-up and adherence to a regimen of secondary prophylaxis. Patients should be entered onto the local ARF registry (if present) and contact made with primary care practitioners to ensure a plan for follow-up and administration of secondary prophylaxis before the patient is discharged. Patients and their families should also be educated about their disease, emphasizing the importance of adherence to secondary prophylaxis.

PREVENTION

PRIMARY PREVENTION

Ideally, primary prevention would entail elimination of the major risk factors for streptococcal infection, particularly overcrowded housing. This is difficult to achieve in most places where ARF is common but must remain a priority in ongoing efforts to achieve global control of ARF and RHD.

Concerted international efforts are underway to develop a vaccine against group A *Streptococcus* that would prevent infection of the throat or skin and consequently prevent ARF in the absence of a suitable vaccine; however, the mainstay of primary prevention for ARF remains primary prophylaxis (i.e., the timely and complete treatment of group A streptococcal sore throat with antibiotics). If commenced within 9 days of sore throat onset, a course of penicillin (as outlined above for treatment of ARF) will prevent almost all cases of ARF that would otherwise have developed. In settings where ARF and RHD are common but microbiologic diagnosis of group A streptococcal pharyngitis is not available, such as in resource-poor countries, primary care guidelines sometimes recommend that all patients with sore throat be treated with penicillin (an approach that has the potential drawbacks that come from antibiotic overuse, including side effects and increasing pressure on antimicrobial resistance in group A *Streptococcus* or bystander pathogens) or, alternatively, that a clinical algorithm be used to identify patients with a higher likelihood of group A streptococcal pharyngitis. Although imperfect, such approaches recognize the importance of ARF prevention at the expense of overtreating many cases of sore throat that are not caused by group A *Streptococcus*. Although there is no proof that antibiotic treatment of group A streptococcal skin infections can prevent ARF, the increasing evidence that impetigo is strongly associated with ARF in some populations argues for a focus on treatment and prevention of group A streptococcal skin infections as part of a comprehensive ARF control strategy in regions with endemic impetigo.

SECONDARY PREVENTION

The mainstay of controlling ARF and RHD is secondary prevention. Because patients with ARF are at dramatically higher risk than the general population of developing a further episode of ARF after a group A streptococcal infection, they should receive long-term penicillin prophylaxis to prevent recurrences. The best antibiotic for secondary prophylaxis is benzathine penicillin G (1.2 million units, or 600,000 units if ≤27 kg) delivered intramuscularly every 4 weeks. It can be given every 3 weeks, or even every 2 weeks, to persons considered to be at particularly high risk, although in settings where good compliance with an every-4-week dosing schedule can be achieved, more frequent dosing is rarely needed. Evidence has emerged recently that subcutaneous delivery of benzathine penicillin G provides more optimal pharmacokinetics than intramuscular delivery and may have added advantages of reducing pain and allowing for larger doses to be delivered less frequently, although this approach is yet to be recommended in clinical guidelines. Oral penicillin V (250 mg) can be given twice daily instead but is less effective than benzathine penicillin G. Penicillin-allergic patients can receive erythromycin (250 mg) twice daily.

The duration of secondary prophylaxis is determined by many factors, in particular the duration since the last episode of ARF (recurrences become less likely with increasing time), age (recurrences are less likely with increasing age), and the severity of RHD (if severe, it may be prudent to avoid even a very small risk of recurrence because of the potentially serious consequences) **(Table 371-4)**. Secondary prophylaxis is best delivered as part of a coordinated RHD control program, based around a registry of patients. Registries improve the ability to follow patients and identify those who default from prophylaxis and to institute strategies to improve adherence.

TABLE 371-4 American Heart Association Recommendations for Duration of Secondary Prophylaxis[a]

CATEGORY OF PATIENT	DURATION OF PROPHYLAXIS
Rheumatic fever without carditis	For 5 years after the last attack or 21 years of age (whichever is longer)
Rheumatic fever with carditis but no residual valvular disease	For 10 years after the last attack, or 21 years of age (whichever is longer)
Rheumatic fever with persistent valvular disease, evident clinically or on echocardiography	For 10 years after the last attack, or 40 years of age (whichever is longer); sometimes lifelong prophylaxis

[a] These are only recommendations and must be modified by individual circumstances as warranted. Note that some organizations recommend a minimum of 10 years of prophylaxis after the most recent episode, or until 21 years of age (whichever is longer), regardless of the presence of carditis with the initial episode.

Source: Reproduced with permission from MA Gerber et al: Prevention of rheumatic fever and diagnosis and treatment of acute streptococcal pharyngitis. Circulation 119:1541, 2009. *https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.109.191959?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed.*

FURTHER READING

Carapetis JR et al: Acute rheumatic fever and rheumatic heart disease. Nat Rev Dis Primers 14:15084, 2016.

Gewitz MH et al: Revision of the Jones Criteria for the diagnosis of acute rheumatic fever in the era of Doppler echocardiography: A scientific statement from the American Heart Association. Circulation 131:1806, 2015.

Institute for Health Metrics and Evaluation. Rheumatic heart disease interactive visualization. Global Burden of Disease Results Tool. Accessed May 2024. *https://vizhub.healthdata.org/gbd-results/*.

RHD Australia (ARF/RHD Writing Group): The 2020 Australian guideline for prevention, diagnosis and management of acute rheumatic fever and rheumatic heart disease (3rd edition); 2020. Available at *https://www.rhdaustralia.org.au/arf-rhd-guideline*.

Rwebembera J et al: 2023 World Heart Federation guidelines for the echocardiographic diagnosis of rheumatic heart disease. Nat Rev Cardiol 21:250, 2024.

Walkinshaw DR et al: The *Streptococcus pyogenes* vaccine landscape. NPJ Vaccines 8:16, 2023.

VIDEO 371-1A Transthoracic echocardiographic images of a 9-year-old girl with first episode of acute rheumatic fever. Images demonstrate the typical echocardiographic findings of acute rheumatic carditis. The valve leaflets are relatively thin and highly mobile. The failure of coaptation of the mitral valve leaflets is the result of chordal elongation and annular dilatation. The mitral valve regurgitation is moderate with a typical posterolaterally directed regurgitant jet of rheumatic carditis. ***A.*** Acute rheumatic carditis (apical four-chamber view echocardiogram).

VIDEO 371-1B Transthoracic echocardiographic images of a 9-year-old girl with first episode of acute rheumatic fever. Images demonstrate the typical echocardiographic findings of acute rheumatic carditis. The valve leaflets are relatively thin and highly mobile. The failure of coaptation of the mitral valve leaflets is the result of chordal elongation and annular dilatation. The mitral valve regurgitation is moderate with a typical posterolaterally directed regurgitant jet of rheumatic carditis. ***B.*** Acute rheumatic carditis (apical four-chamber view color Doppler echocardiogram).

VIDEO 371-1C Transthoracic echocardiographic images of a 9-year-old girl with first episode of acute rheumatic fever. Images demonstrate the typical echocardiographic findings of acute rheumatic carditis. The valve leaflets are relatively thin and highly mobile. The failure of coaptation of the mitral valve leaflets is the result of chordal elongation and annular dilatation. The mitral valve regurgitation is moderate with a typical posterolaterally directed regurgitant jet of rheumatic carditis. ***C.*** Acute rheumatic carditis (parasternal long-axis view echocardiogram).

VIDEO 371-1D Transthoracic echocardiographic images of a 9-year-old girl with first episode of acute rheumatic fever. Images demonstrate the typical echocardiographic findings of acute rheumatic carditis. The valve leaflets are relatively thin and highly mobile. The failure of coaptation of the mitral valve leaflets is the result of chordal elongation and annular dilatation. The mitral valve regurgitation is moderate with a typical posterolaterally directed regurgitant jet of rheumatic carditis. ***D.*** Acute rheumatic carditis (parasternal long-axis view color Doppler echocardiogram).

VIDEO 371-2A Transthoracic echocardiographic images are from a 5-year-old boy with chronic rheumatic heart disease with severe mitral valve regurgitation and moderate mitral valve stenosis. Images demonstrate the typical echocardiographic findings in advanced chronic rheumatic heart disease. Both the anterior and posterior mitral valve leaflets are markedly thickened. During diastole, the motion of the anterior mitral valve leaflet tip is restricted with doming of the body of the leaflet toward the interventricular septum. This appearance is commonly described as a “hockey stick” or an “elbow” deformity. ***A.*** Chronic rheumatic heart disease (parasternal long-axis view).

VIDEO 371-2B Transthoracic echocardiographic images are from a 5-year-old boy with chronic rheumatic heart disease with severe mitral valve regurgitation and moderate mitral valve stenosis. Images demonstrate the typical echocardiographic findings in advanced chronic rheumatic heart disease. Both the anterior and posterior mitral valve leaflets are markedly thickened. During diastole, the motion of the anterior mitral valve leaflet tip is restricted with doming of the body of the leaflet toward the interventricular septum. This appearance is commonly described as a “hockey stick” or an “elbow” deformity. ***B.*** Chronic rheumatic heart disease (apical two-chamber view echocardiogram).

372

Systemic Sclerosis (Scleroderma) and Related Disorders

John Varga

DEFINITION AND CLASSIFICATION

Systemic sclerosis (SSc) is an orphan disease of unknown etiology, complex pathogenesis, and variable clinical presentations. SSc follows a chronic and often progressive course and is associated with significant disability and mortality. Virtually every organ can be affected **(Fig. 372-1)**.

There is marked variability among SSc patients in patterns of skin involvement and organ complications, rates of disease progression, response to treatment, disease severity, and survival. The early stages of SSc may be associated with prominent inflammatory features such as edema, but over time, structural alterations in multiple vascular beds and visceral organ dysfunction due to fibrosis and atrophy progressively dominate the clinical picture. Classification criteria for the diagnosis of SSc (shown in **Table 372-1**) are >90% specific and sensitive, although they are intended to be used in clinical research rather than clinical practice.

Although thickened and indurated skin (*scleroderma*) is the distinguishing hallmark of SSc, similar skin changes can also be seen in localized forms of scleroderma, along with a variety of metabolic, inherited, autoimmune, and iatrogenic conditions **(Table 372-2)**. Patients with SSc can be broadly segregated into two major subsets defined by the pattern of skin involvement and associated with characteristic clinical and serologic features and natural history **(Table 372-3)**. Patients with diffuse cutaneous SSc (dcSSc) have extensive skin induration, starting in the fingers (sclerodactyly) and ascending from distal to proximal limbs and the trunk. In these patients, progressive skin disease, interstitial lung disease (ILD), and less commonly acute renal involvement may develop relatively early. In contrast, in patients with limited cutaneous SSc (lcSSc), Raynaud’s phenomenon generally precedes sclerodactyly and other disease manifestations, sometimes by years. In these patients, skin involvement remains confined to the fingers, distal limbs, and face, while the trunk is spared. A subset of patients with lcSSc display a distinct constellation of findings (calcinosis cutis, Raynaud’s phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia), historically termed the *CREST syndrome*. In lcSSc, visceral organ involvement tends to follow a more insidious and occasionally benign course, while digital ischemic ulcers, pulmonary arterial hypertension (PAH), hypothyroidism, Sjogren’s symptoms, and primary biliary cirrhosis may occur as late complications. In some patients, Raynaud’s phenomenon and characteristic clinical and laboratory features of SSc occur in the absence of detectable skin thickening. This relatively benign disease subset has been termed *SSc sine scleroderma*.

INCIDENCE AND PREVALENCE

SSc is an acquired sporadic disease with a worldwide distribution affecting all races and has a rising incidence rate. In the United States, the incidence is 9–46 cases per million per year. There are an estimated 100,000 U.S. cases, although this number may be significantly higher if patients who do not fulfill formal SSc classification criteria are also included. There are large regional variations in incidence rates of SSc, potentially reflecting differences in case definition, environmental exposures, or genetic susceptibility genes in populations with different ancestries. Prevalence rates in England, northern Europe, and Japan appear to be lower than in North America and Australia. Age, sex, and ancestry influence disease susceptibility, and black people have higher age-specific incidence rates and mortality. In common with most autoimmune diseases, SSc shows a strong female sex bias (4.6:1),

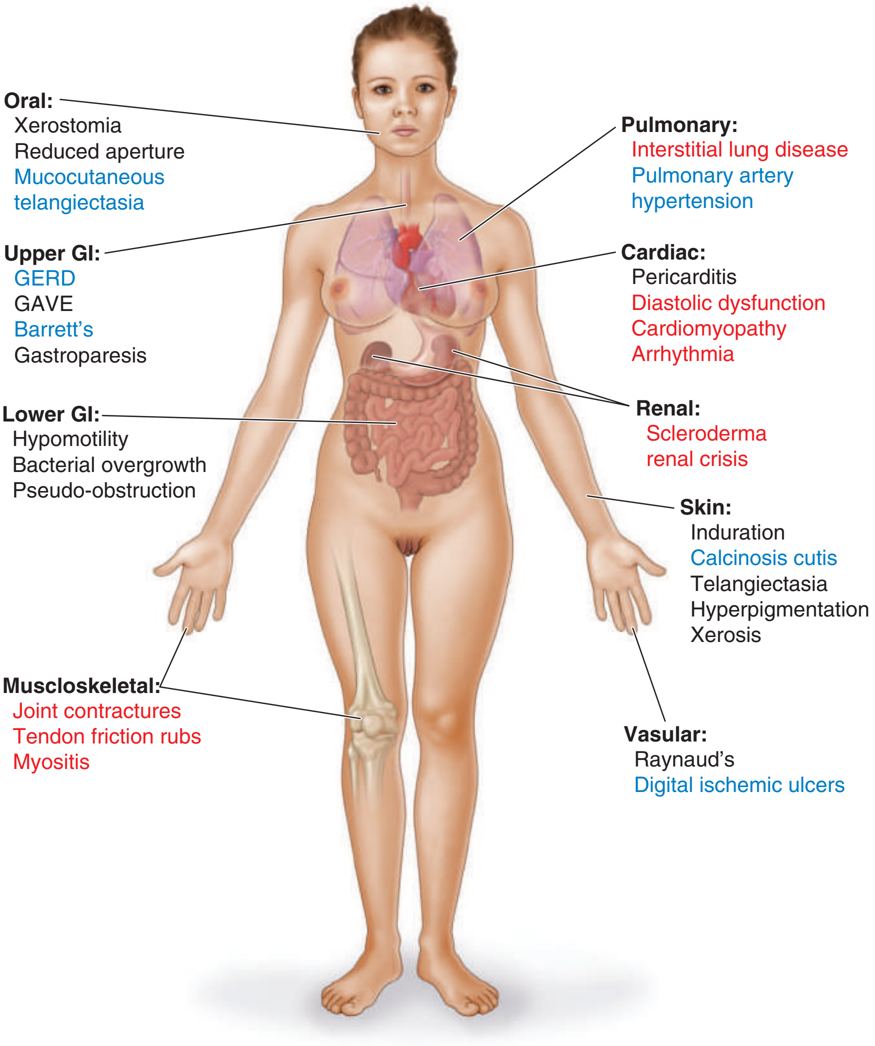
Oral:
Xerostomia
Reduced aperture
Mucocutaneous telangiectasia
Upper GI:
GERD
GAVE
Barrett's
Gastroparesis
Lower GI:
Hypomotility
Bacterial overgrowth
Pseudo-obstruction
Muscloskeletal:
Joint contractures
Tendon friction rubs
Myositis
Pulmonary:
Interstitial lung disease
Pulmonary artery hypertension
Cardiac:
Pericarditis
Diastolic dysfunction
Cardiomyopathy
Arrhythmia
Renal:
Scleroderma renal crisis
Skin:
Induration
Calcinosis cutis
Telangiectasia
Hyperpigmentation
Xerosis
Vasular:
Raynaud's
Digital ischemic ulcers

FIGURE 372-1 Multiorgan involvement in systemic sclerosis (SSc). Complications of SSc; *Red*, complications more prominent in patients with diffuse cutaneous SSc; *Blue*, those more prominent in limited cutaneous SSc; *Black*, complications common in both limited and diffuse cutaneous SSc subsets. GAVE, gastric antral vascular ectasia; GERD, gastroesophageal reflux disease; GI, gastrointestinal.

TABLE 372-1 ACR/EULAR Classification Criteria for Diagnosis of Systemic Sclerosis (SSc)

ITEM	SUBITEM	WEIGHT/SCORE
Skin thickening (bilateral)—fingers extending proximal to MCP joints		9
Skin thickening of fingers only	Puffy fingers Sclerodactyly (skin thickened distal to MCP joints)	2 4
Fingertip lesions	Digital tip ulcer or pitting scar	2 3
Mucocutaneous telangiectasia		2
Abnormal nailfold capillary pattern		2
Lung involvement	PAH Interstitial lung disease	2 2
Raynaud's phenomenon		3
SSc-specific autoantibodies	ACA Scl-70 RNA polymerase III	3

Abbreviations: ACA, anticentromere antibodies; ACR, American College of Rheumatology; EULAR, European Alliance of Associations for Rheumatology; MCP, metacarpophalangeal joint; PAH, pulmonary arterial hypertension.

TABLE 372-2 Conditions Associated with Scleroderma-Like Skin Induration

- Systemic sclerosis (SSc)
  - Limited cutaneous SSc
  - Diffuse cutaneous SSc
- Localized scleroderma
  - Guttate (plaque) morphea, diffuse (pansclerotic) morphea, bullous morphea
  - Linear scleroderma, coup de sabre, hemifacial atrophy
- Pansclerotic morphea
- Overlap syndromes
  - Mixed connective tissue disease
  - SSc/polymyositis
- Diabetic scleredema and scleredema of Buschke
- Scleromyxedema (papular mucinosis)
- Chronic graft-versus-host disease
- Diffuse fasciitis with eosinophilia (Shulman's disease, eosinophilic fasciitis)
- Stiff skin syndrome
- Pachydermatoperiostosis (primary hypertrophic osteoarthropathy)
- Chemically induced and drug-associated scleroderma-like conditions
  - Vinyl chloride–induced disease
  - Eosinophilia-myalgia syndrome (associated with L-tryptophan contaminant exposure)
  - Nephrogenic systemic fibrosis (associated with gadolinium exposure)
- Paraneoplastic syndrome

TABLE 372-3 Subsets of Systemic Sclerosis (SSc): Features of Limited Cutaneous Versus Diffuse Cutaneous Disease

CHARACTERISTIC FEATURE	LIMITED CUTANEOUS SSc	DIFFUSE CUTANEOUS SSc
Skin involvement	Indolent onset. Limited to fingers, distal to elbows, face; slow progression	Rapid onset. Diffuse: fingers, extremities, face, trunk; rapid progression
Raynaud phenomenon	Antedates skin involvement, sometimes by years; may be associated with critical ischemia in the digits	Onset coincident with skin involvement; critical ischemia less common
Musculoskeletal	Mild arthralgia	Severe arthralgia, carpal tunnel syndrome, tendon friction rubs
Interstitial lung disease	Slowly progressive, generally mild	Frequent, early onset and progression, can be severe
Pulmonary arterial hypertension	Frequent, late, may occur as an isolated complication	Often occurs in association with interstitial lung disease
Scleroderma renal crisis	Very rare	Occurs in 15%; generally early (<4 years from disease onset)
Calcinosis cutis	Frequent, prominent	Less common, mild
Characteristic autoantibodies	Anti-centromere	Anti–topoisomerase I (Scl-70), anti-RNA polymerase III

which is most pronounced in the childbearing years and declines after menopause. Males and females show numerous differences in immune cell composition and function, with females exhibiting greater innate, humoral, and cellular immune responses. The female bias in SSc may be in part due to sex hormones, as estrogens have immunostimulatory effects, including stimulation of interleukin (IL)-6, RF5, and TLR7 expression. An alternate, but not mutually exclusive, hypothesis for sex-biased disease susceptibility implicated skewed inactivation of the X chromosome in females, resulting in aberrant expression of immune-related genes residing on the X chromosome, including TLR7 and TLR8 and CD40L. Moreover, while autosomal genes are the same in males and females, their cell type–specific expression can vary between the sexes. Finally, differing environmental exposures between males and females might also contribute to sexual dimorphism in SSc susceptibility, disease course, and response to therapy. Having an affected first-degree family member with SSc increases disease risk 13-fold. Although SSc can present at any age, the peak age of onset in women with both limited and diffuse cutaneous forms is 65–74 years, although in black patients, disease onset occurs at an earlier age. Furthermore, black patients with SSc are more likely to have diffuse cutaneous disease, ILD, a worse prognosis, and greater premature death.

GENETIC CONTRIBUTION TO DISEASE SUSCEPTIBILITY

In general, the genetic associations of SSc identified to date make only a relatively modest contribution to disease susceptibility. Studies of SSc showed low rates of disease concordance (4.7%) in monozygotic twins, a rate much lower than in other autoimmune diseases such as rheumatoid arthritis (12.3%). On the other hand, supporting the genetic contribution to disease susceptibility is the observation that 1.6% of SSc patients have a first-degree relative with SSc, a prevalence rate that is markedly increased compared to the general population (0.026). Moreover, the risk of Raynaud's phenomenon, ILD, and other autoimmune diseases, including systemic lupus erythematosus (SLE) **(Chap. 368)**, rheumatoid arthritis **(Chap. 370)**, and autoimmune thyroiditis **(Chap. 395)**, is also increased in first-degree relatives of patients with SSc. Current approaches to uncover the genetic factor contribution to SSc include DNA sequencing and single nucleotide polymorphism (SNP) analysis of candidate genes and SNP analysis of the entire genome in a hypothesis-free manner. Genome-wide association studies (GWAS) involve large multicenter, multinational, and multiethnic cohorts. A majority of the robustly validated genetic susceptibility loci for SSc are genes at the highly polymorphic human leukocyte antigen (HLA) region and other genes involved in innate and adaptive immunity and interferon responses, highlighting the importance of autoimmunity as the initial trigger for the disease. Genetic studies have shown associations with common (small effect size) variants related to B and T lymphocyte activation (*BANK1*, *BLK*, *CD247*, *IL2RA*, *CCR6*, *IDO1, TNFSF4/OX40L*, *PTPN22,* and *TNIP1*), as well as *STAT4* and *IRF5*, both implicated in the secretion of type I interferon from plasmacytoid dendritic cells. In addition, candidate gene studies and GWAS identified a strong and consistent association with HLA class II haplotypes on chromosome 6, including _HLA-DRB1*11:04, DQA1*05:01,_ and _DQB1*03:01_, and the non-HLA major histocompatibility complex (MHC) genes *NOTCH4* and *PSORSC1*. Other genetic variants associated with SSc are involved in innate immunity (*IRF7*, *TNFAIP3*, and *TLR2*) and secretion of type interferons, particularly from plasmacytoid dendritic cells (*STAT4, IRF5*). Additional associations with *IL12RB2, IL-21,* the autophagy and apoptosis-related genes *DNASE1L3* and *SOX5*, and the extracellular matrix–related genes *CSK, CAV1, PPARG,* and *GRB10* have been reported. In addition to SSc susceptibility, some of these genetic loci are associated with particular disease manifestations or serologic subsets, including ILD (*CTGF*, *CD226*), PAH (*TNIP1*), scleroderma renal crisis (_HLA-DRB1*_), and anticentromere antibodies (_HLA-DPB1*05:01_). While the functional consequences of these gene variants and their potential roles in pathogenesis are not well understood, it seems likely that in combination they cause a state of altered immune regulation, leading to increased susceptibility to autoimmunity and persistent inflammation. Of note, several of the SSc genetic variants are also linked with other autoimmune disorders, including SLE, Sjögren's syndrome, rheumatoid arthritis, multiple sclerosis, and psoriasis, suggesting common pathogenic pathways shared among these phenotypically dissimilar conditions. A population electronic health record–based GWAS identified *ADRA2A*, implicated in α_{2a}-adrenoreceptor signaling, as a candidate causal gene for Raynaud's phenomenon. Notably, the genetic associations identified to date only explain a fraction of the heritability of SSc and focus on common variants. By contrast, next-generation sequencing, such as whole exome sequencing, may help identify additional genetic susceptibility factors in SSc, particularly rare (and potentially causal) coding variants and their association with specific phenotypes.

ENVIRONMENTAL FACTORS AND OCCUPATIONAL EXPOSURES

The etiology of SSc is unknown. Given the relatively modest genetic contribution to susceptibility in SSc, environmental exposures, such as infectious agents, microbiome, and occupational, dietary, lifestyle, and drug exposures, are likely to play a major role. Studies suggest potential links with viruses including parvovirus B19, Epstein-Barr virus (EBV), adeno-associated virus 5, and cytomegalovirus (CMV). However, evidence to directly implicate these infectious agents in the pathogenesis of SSc is lacking. Occupational exposures tentatively linked with SSc include particulate silica (quartz), polyvinyl chloride, epoxy resins, welding fumes and organic solvents, and aromatic hydrocarbons including paint thinners, toluene, xylene, and trichloroethylene. These exposures might elicit cell type–specific stable and heritable epigenetic modifications such as DNA methylation and histone modifications to drive altered cellular gene expression **(Fig. 372-2)**. Because epigenetic modifications are often reversible, they represent potential targets for therapy. Drugs implicated in SSc-like illnesses include bleomycin, pentazocine, cocaine, and appetite suppressants linked with PAH. Most recently, immune checkpoint inhibitors, such as PD-1 blockers used in cancer therapy, have been implicated as triggers for SSc-like illnesses. Radiation therapy for cancer has been linked with de novo onset of SSc as well as with exacerbation of preexisting SSc. In contrast to rheumatoid arthritis, cigarette smoking does not increase the risk of SSc. Although case reports and series describing the occurrence of SSc in women with silicone breast implants had raised concern regarding a putative pathogenic role of silicone, large-scale epidemiologic

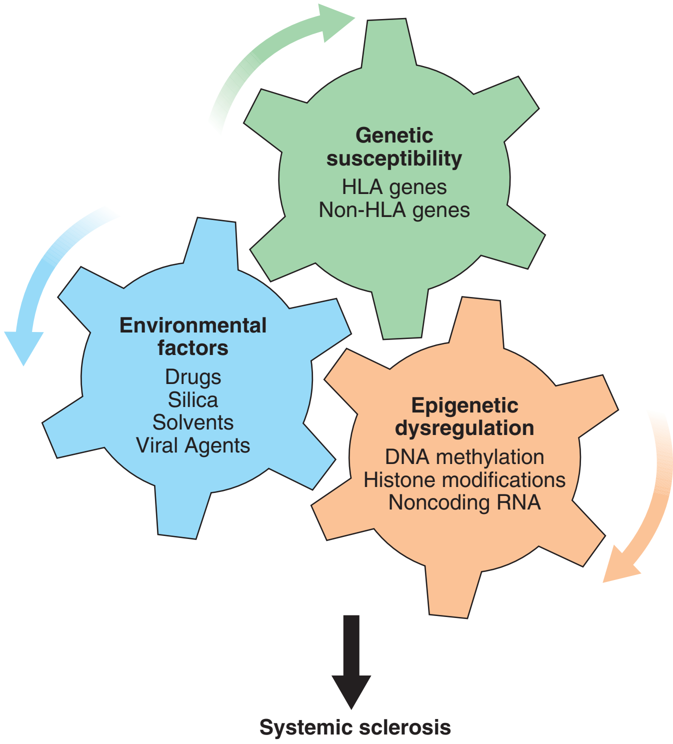

FIGURE 372-2 The interplay of genetic risk factors and environmental exposure-induced epigenetic modifications underlying the complex pathogenesis of systemic sclerosis. HLA, human leukocyte antigen. *(Reproduced with permission from Drs. Amr Sawalha and Pei-Suen Tsou, University of Michigan.)*

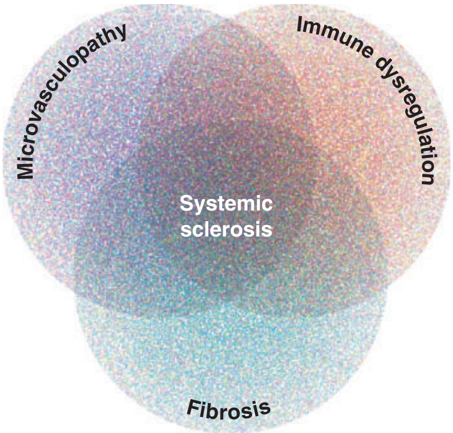

FIGURE 372-3 The pathogenic systemic sclerosis (SSc) triad. The characteristic constellation of synchronously occurring vasculopathy, autoimmunity/inflammation, and fibrosis distinguishes SSc and underlies its protean clinical manifestations.

investigations did not find increased prevalence of SSc in such populations.

PATHOGENESIS

The pathogenesis of SSc involves putative environmental insults that trigger epigenetic modifications in a genetically predisposed host, causing alterations in gene expression underlying durable changes in the behavior of multiple cell types (Fig. 372-2).

Three cardinal pathomechanistic processes underlie the protean clinical manifestations and pathologic changes of SSc: (1) diffuse microangiopathy, (2) inflammation and autoimmunity, and (3) visceral and vascular fibrosis affecting multiple organs **(Fig. 372-3)**. While each of these distinct processes may be synchronously active in a given SSc patient, their relative severity progression and contribution to the overall clinical picture varies among individual patients and over time. In general, autoimmunity and reversible vascular reactivity are early features of SSc, while fibrosis and atrophy occur later in the disease.

■ MODELS OF DISEASE

There is no single animal model of SSc that fully reproduces the three cardinal processes that underlie pathogenesis or phenocopies their temporal evolution. In mice, skin and lung fibrosis accompanied by variable vasculopathy and autoimmunity can be elicited by repeated injection of bleomycin or angiotensin II or by transplantation of HLA-mismatched bone marrow or spleen cells, which elicit a chronic graft-versus-host disease–like condition. Targeted genetic modifications in mice give rise to new disease models for investigating the pathogenetic roles of individual molecules, pathways, and cell types. These animal models have utility as experimental tools to discover and validate novel therapeutic targets, cell types, and biomarkers. Recent efforts focus on the development of three-dimensional ex vivo disease models, including skin equivalents and organoids.

■ MICROANGIOPATHY

In a progressive model of disease pathogenesis **(Fig. 372-4)**, vascular injury is an early and possibly primary pathogenic event that underlies protean manifestations of small vessel vasculopathy, including mucocutaneous telangiectasia, Raynaud's phenomenon, ischemic digital ulcers, acro-osteolysis, scleroderma renal crisis, myocardial involvement, and PAH.

Raynaud's phenomenon, commonly the earliest clinical manifestation of SSc, is characterized by altered blood-flow response to cold challenge in small digital arteries. This reversible functional abnormality is associated with autonomic and peripheral nervous system alterations, including impaired production of the neuropeptide calcitonin gene-related peptide from sensory afferent nerves and heightened sensitivity of α_2-adrenergic receptors on vascular smooth-muscle cells. Isolated (or primary) Raynaud's disease is common, generally benign, and nonprogressive. In contrast, SSc-associated (or secondary) Raynaud's phenomenon often progresses to irreversible structural changes in the small blood vessels, culminating in ischemic digital tip ulcers, necrosis, and finger amputation. Genetic studies have identified *ADRA2A* as a candidate causative gene in Raynaud's phenomenon.

Viruses, cytotoxic factors, chemokines, thrombogenic circulating microvesicles, alternate complement pathway activation, neutrophil extracellular trap (NET) formation (NETosis), and autoantibodies targeting endothelial cells, phospholipids, and β_2-glycoprotein I (β_2GPI) have all been considered as possible triggers of endothelial cell injury in patients with SSc. Endothelial damage disrupts the production of vasodilatory (nitric oxide and prostacyclin) and vasoconstricting (endothelin-1) substances, while causing upregulation of intercellular adhesion molecule 1 (ICAM-1) and other surface adhesion molecules. Microvessels show enhanced permeability and transendothelial leukocyte diapedesis, activation of coagulation cascades, formation of NETs, elevated thrombin production, and impaired fibrinolysis. Spontaneous platelet aggregation causes release of serotonin, platelet-derived growth factor (PDGF), platelet factor 4 (PF4, also known as CXCL4), and platelet alpha granules including thromboxane, a potent vasoconstrictor. Smooth-muscle cell–like myointimal cells arising from endothelial cells through a process called endothelial-mesenchymal transition (EndoMT) accumulate within the vascular media. The basement membrane is thickened and reduplicated, and perivascular adventitial fibrosis develops. The vasculopathic process primarily affects capillaries, arterioles, and, less commonly, even large vessels in many organs, resulting in impaired blood flow and tissue ischemia. Progressive luminal occlusion due to intimal and medial hypertrophy, combined with persistent endothelial cell damage and adventitial fibrosis, establishes a vicious cycle that underlies fibroproliferative vasculopathy and culminates in the striking absence of small blood vessels (rarefaction) in late-stage disease. Recurrent ischemia-reperfusion generates reactive oxygen species (ROS) that further damage the endothelium through peroxidation of membrane lipids. Paradoxically, the process of revascularization

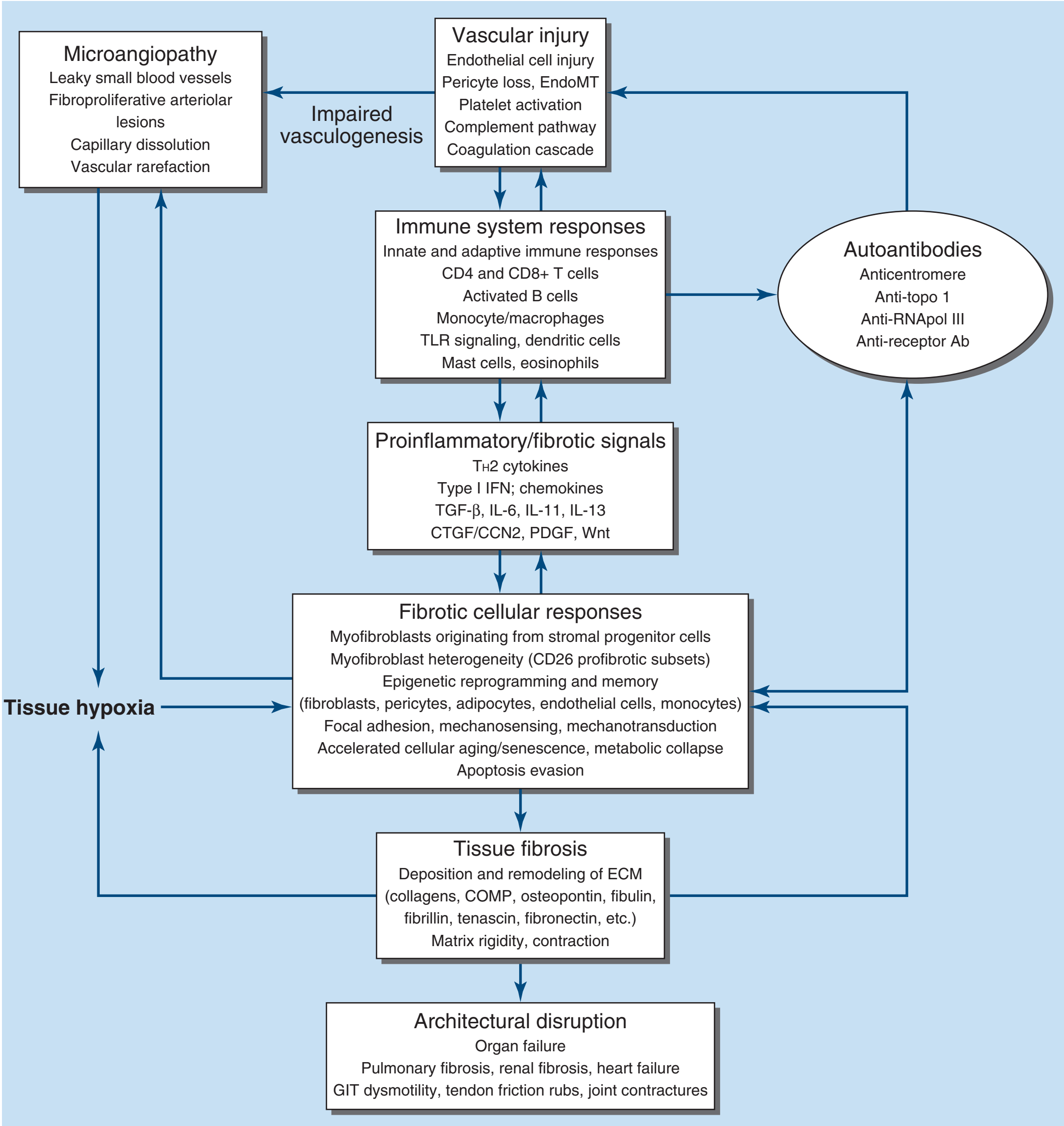
Vascular injury
Endothelial cell injury
Pericyte loss, EndoMT
Platelet activation
Complement pathway
Coagulation cascade
Microangiopathy
Leaky small blood vessels
Fibroproliferative arteriolar lesions
Capillary dissolution
Vascular rarefaction
Impaired vasculogenesis
Immune system responses
Innate and adaptive immune responses
CD4 and CD8+ T cells
Activated B cells
Monocyte/macrophages
TLR signaling, dendritic cells
Mast cells, eosinophils
Autoantibodies
Anticentromere
Anti-topo 1
Anti-RNApol III
Anti-receptor Ab
Proinflammatory/fibrotic signals
TH2 cytokines
Type I IFN; chemokines
TGF-β, IL-6, IL-11, IL-13
CTGF/CCN2, PDGF, Wnt
Fibrotic cellular responses
Myofibroblasts originating from stromal progenitor cells
Myofibroblast heterogeneity (CD26 profibrotic subsets)
Epigenetic reprogramming and memory
(fibroblasts, pericytes, adipocytes, endothelial cells, monocytes)
Focal adhesion, mechanosensing, mechanotransduction
Accelerated cellular aging/senescence, metabolic collapse
Apoptosis evasion
Tissue hypoxia
Tissue fibrosis
Deposition and remodeling of ECM
(collagens, COMP, osteopontin, fibulin, fibrillin, tenascin, fibronectin, etc.)
Matrix rigidity, contraction
Architectural disruption
Organ failure
Pulmonary fibrosis, renal fibrosis, heart failure
GIT dysmotility, tendon friction rubs, joint contractures

FIGURE 372-4 Integrated progressive model of systemic sclerosis (SSc) pathogenesis. Initial vascular insult in a genetically predisposed individual triggers a cascade of functional and structural vascular alterations associated with inflammation and autoimmunity. Early immune responses elicit fibroblast activation and differentiation, resulting in sustained and progressive fibrogenesis, irreversible tissue damage, and functional failure of affected organs. Vascular damage also leads to platelet activation, hypercoagulability, and tissue ischemia that further contributes to progressive fibrosis and atrophy. Ab, antibody; CTGF, connective tissue growth factor; ECM, extracellular matrix; EndoMT, endothelial-mesenchymal transition; GIT, gastrointestinal tract; IFN, interferon; IL, interleukin; PDGF, platelet-derived growth factor; TGF-β, transforming growth factor β; TLR, toll-like receptor.

that normally reestablishes blood flow to ischemic tissue is defective in SSc despite elevated levels of other angiogenic factors. Moreover, bone marrow–derived circulating endothelial progenitor cells are reduced in number and impaired in function.

There is increasing evidence implicating EndoMT in the pathogenesis of SSc vasculopathy. The process is characterized by transition of endothelial cells into myofibroblasts accompanied by loss of endothelial cell markers and nuclear localization of the transcription factor Snail1. In arterioles and small arteries, EndoMT leads to accumulation of endothelial cell–derived myofibroblasts in the intima and media, resulting in fibroproliferative vasculopathy and luminal occlusion. In contrast, EndoMT affecting the capillary vessels leads to a destructive vasculopathy characterized by loss of endothelial cells and accumulation of profibrotic interstitial myofibroblasts derived from endothelial cells, resulting in fibrosis and microvessel rarefaction, as can be observed by nailfold capillaroscopy. Widespread capillary loss combined with fibroproliferative vasculopathy affecting arterioles and arteries and impaired ability to repair and replace damaged vessels are hallmarks of SSc.

INFLAMMATION AND AUTOIMMUNITY

Cellular Immunity While the initial events triggering both innate and adaptive autoimmune responses in SSc are unknown, a number of observations support the inflammatory/autoimmune nature of SSc. These include the near-universal presence of circulating autoantibodies with defined specificities; clustering of SSc with other autoimmune diseases; activated immune cells, including autoreactive T cells with oligoclonal antigen receptors, within target organs; prominent type I interferon (IFN) signatures, characterized by elevated expression of IFN-regulated genes, in a variety of immune cell types in circulation and in skin biopsies; elevated circulating levels and spontaneous mononuclear cell secretion of cytokines such as IL-6, tumor necrosis factor α, IL-4, IL-10, IL-17, IL-33, and chemokines CCL2, CCL9, and CXCL4; genetic association of SSc (shared with other autoimmune diseases) with variants of MHC and other immune response genes; and the rapid clinical response, fibrosis resolution, and vascular regeneration observed in some SSc patients treated with immunomodulatory or immunoablative and chimeric antigen receptor (CAR) T-cell (CAR-T) therapies.

In early (edematous) stage SSc, mononuclear cell infiltrates composed of activated T cells, monocytes/macrophages, and dendritic cells can be detected in skin, lungs, and other affected organs even prior to fibrosis or vascular damage. Dendritic cells are professional antigen-presenting cells that bridge the adaptive and innate immune systems. They respond to injury by sensing tissue damage via their pattern recognition receptors such as Toll-like receptors (TLRs), NOD-like receptors (NLRs), and RIG-I like receptors. In SSc skin biopsies, dendritic cells located in close proximity to activated myofibroblasts express TLR8 and other endosomal TLRs that are activated by self-nucleic acids. TLR stimulation induces mediators including IFN, IL-10, and thymic stromal lymphopoietin (TSLP). A principal product secreted from SSc plasmacytoid dendritic cells is the atypical pleiotropic chemokine CXCL4, also known as platelet factor 4 (PAF). Produced in platelets and other cell types, CXCL4 is a short cationic peptide implicated in multiple autoimmune, inflammatory, and fibrotic diseases. In contrast to other chemokines that signal through G protein–coupled receptors on leukocytes, endothelial cells, and other target cells, CXCL4 can, independent of its receptor CXCR3, bind to extracellular self-nucleic acids and act as a chaperone to facilitate their cellular uptake and activation of endosomal TLRs. By inducing potent proinflammatory and antiangiogenic activities mediated via type 2 cytokines and IL-6, CXCL4 elicits a profibrotic monocyte phenotype and directly promotes myofibroblast differentiation from stromal progenitor cells. Levels of CXCL4 are elevated in the serum of patients with SSc and can predict future disease progression. In light of their crucial pathogenic roles in SSc, plasmacytoid dendritic cells and CXCL4 are potential therapeutic targets. Tissue-infiltrating T cells express CD45 and HLA-DR activation markers and display restricted T-cell receptor signatures indicative of oligoclonal expansion in response to recognition of as-yet-unknown antigen. Of note, in patients who are diagnosed with RNA polymerase III antibody–positive SSc in close temporal association with cancer, the tumor commonly harbors genomic alterations in RNApol3, which leads to the generation of autoantigen-specific T-cell immunity and cross-reactive antibodies.

Circulating T cells in SSc express chemokine receptors and α_1 integrin, accounting for their enhanced binding to endothelium and to fibroblasts, while endothelial cells express ICAM-1 and other adhesion molecules that facilitate leukocyte diapedesis. Activated T cells show a T_H2-polarized immune response driven by dendritic cells. The T_H2 cytokines IL-4, IL-13, IL-33, and TSLP induce fibroblast activation, whereas the T_H1 cytokine IFN-γ blocks cytokine-mediated fibroblast activation and exhibits antifibrotic properties. Alternately activated macrophages expressing CD163, which produce transforming growth factor (TGF)-β and promote angiogenesis and tissue remodeling, are increased in the skin and lung in SSc. Regulatory T cells (Tregs) enforce immune tolerance; although the frequency is elevated in the circulation and tissues in SSc patients, their immunosuppressive function appears to be defective. A recent report demonstrated that Tregs play a critical role in preventing spontaneous fibrosis in the skin and fascia, possibly by sequestering activated TGF-β. Studies implicate altered B-cell homeostasis and function in SSc. Circulating B cells show elevated CD19 and co-stimulatory molecules CD80 and CD86, suggesting B-cell chronic activation. Serum levels of proliferation-inducing ligand (APRIL) and B-cell activating factor (BAFF) are elevated in SSc and associated with extent of skin and lung involvement. In addition to autoantibodies, B cells also secrete IL-6, TGF-β, and other profibrotic cytokines implicated in pathogenesis. Thus, B-cell hyperactivity might directly contribute to the inflammatory and fibrotic processes in SSc, as well as generation of autoantibodies.

While polymorphonuclear leukocytes have not been extensively investigated in SSc, neutrophils can secrete fibrogenic TGF-β and IL-6, and release ROS; moreover, SSc neutrophils are prone to release extracellular traps that can trigger thrombotic and immune responses. Mast cells attracted into the fibrotic skin via plasminogen activator inhibitor (PAI-1) signaling from keratinocytes can release fibrogenic mediators including TGF-β. Moreover, mast cells can directly activate fibroblasts via serpine 1.

Humoral Autoimmunity Circulating antinuclear antibodies (ANAs) can be detected in virtually all patients with SSc, even in early and possibly preclinical stages of disease. In addition, several SSc-specific autoantibodies with distinct patterns of immunofluorescence show strong associations with unique disease phenotypes as well as HLA haplotype **(Table 372-4)**. Owing to their exquisite specificity, mutual exclusivity, and association with distinct disease manifestations, SSc-associated autoantibodies have clinical utility for diagnosis and risk stratification. Recently, circulating antibodies directed against fibrillin-1, matrix metalloproteinases, cell surface markers, and stimulatory antibodies to angiotensin II receptor, endothelin-1 receptor, or the PDGF receptor have been identified in patients with SSc, although their clinical relevance is not yet established. These functional antibodies show receptor agonist activity in a variety of cell types and might have direct pathogenic roles.

A variety of mechanisms have been proposed to account for the generation of SSc-associated autoantibodies. Proteolytic cleavage, increased expression or altered subcellular localization of normal proteins, or their alterations due to mutation in the case of certain tumors could lead to their immune recognition as neoepitopes, resulting in break of immune tolerance.

■ FIBROSIS

Fibrosis synchronously affecting multiple organs is a distinguishing feature of SSc. The process is characterized by replacement of normal tissue architecture with rigid, avascular, and relatively acellular connective tissue. Fibrosis in SSc follows, and is a consequence of, inflammation and microvascular damage (Fig. 372-4). Fibroblasts are mesenchymal cells primarily responsible for the functional and structural integrity of connective tissue. Upon their activation by extracellular or mechanical

TABLE 372-4 Major Systemic Sclerosis (SSc)-Specific Autoantibodies and Principal Associated Features

TARGET ANTIGEN; CHARACTERISTIC IMMUNOFLUORESCENCE PATTERN	—	SSc SUBSET	PROMINENT CHARACTERISTIC CLINICAL ASSOCIATION
Topoisomerase I (Scl-70) Speckled pattern	—	dcSSc	Tendon friction rubs, digital ischemic ulcers, extensive skin involvement, early-onset ILD, cardiac involvement, scleroderma renal crisis
Centromere proteins Discreet speckled (centromere) pattern	—	lcSSc	Digital ischemic ulcers, calcinosis cutis, isolated PAH; renal crisis rare
RNA polymerase III Speckled pattern	—	dcSSc	Rapidly progressive skin involvement, tendon friction rubs, joint contractures, GAVE, renal crisis, contemporaneous cancers; digital ulcers rare
U3-RNP (fibrillarin) Nucleolar pattern	—	dc/lcSSc	PAH, ILD, scleroderma renal crisis, gastrointestinal tract involvement, myositis
Th/T_0 Nucleolar pattern	—	lcSSc	ILD, PAH
PM/Scl Nucleolar pattern	—	lcSSc	Calcinosis cutis, ILD, myositis overlap
Ro52/TRIM21	—	Overlap	Older-age onset, ILD
Ku Speckled pattern	—	Overlap	SLE, myositis overlap
U1-RNP Speckled pattern	—	MCTD	PAH, inflammatory arthritis, myositis overlap
U11/U12 RNP Speckled pattern	—	dc/lcSSc	ILD

Abbreviations: dcSSc, diffuse cutaneous SSc; GAVE, gastric antral vascular ectasia; ILD, interstitial lung disease; lcSSc, limited cutaneous SSc; MCTD, mixed connective tissue disease; PAH, pulmonary arterial hypertension; SLE, systemic lupus erythematosus.

cues, fibroblasts proliferate; migrate; secrete collagens and other extracellular matrix molecules, growth factors, chemokines, and cytokines; and transdifferentiate into contractile myofibroblasts with prominent cytoskeletal stress fibers. Under normal conditions, these are responses self-limited to accomplish proper (adaptive) tissue repair and regeneration. In contrast, when these responses become sustained and amplified (maladaptive repair), pathologic fibrosis results. A panoply of stimulatory signals are potentially implicated in SSc pathogenesis. In addition to TGF-β, which plays a central role in fibrogenesis, mediating fibrosis in SSc are soluble mediators including IL-6, IL-11, IL-13, and IL-23; the chemokine CXCL4; morphogens and Wnt ligands; connective tissue growth factor (CTGF); PDGF; lysophosphatidic acid; and endothelin. Moreover, hypoxia, ROS, thrombin, and mechanical forces might also contribute to sustained fibroblast activation underlying maladaptive repair in SSc. Damage-associated endogenous ligands for the innate immune receptor TLR4, such as the extracellular matrix–associated macromolecules EDA-fibronectin and tenascin-C, HMGB1, and endogenous ligands for TLR8 and TLR9 (mitochondrial DNA and other self-nucleic acids), further contribute to nonresolving fibrosis via unchecked TLR-mediated fibroblast activation and innate immune signaling.

In addition to tissue-resident fibroblasts and transformed myofibroblasts, bone marrow–derived circulating mesenchymal progenitor cells also contribute to fibrosis. The factors that regulate the differentiation of circulating mesenchymal progenitor cells and their trafficking into lesional tissue are unknown. Endothelial cells in injured arterioles and small arteries undergo EndoMT, giving rise to myofibroblasts that drive perivascular fibrosis. Studies also demonstrate that intradermal and subcutaneous mesenchymal stem cells and mature adipocytes can undergo transdifferentiation into myofibroblasts in the skin. This process of adipocyte-mesenchymal transition could explain the attenuation of subcutaneous white adipose tissue characteristic of fibrotic skin in SSc patients and in mouse models of the disease. Epithelial cells and tissue fibroblasts are all putative sources of pathogenic myofibroblasts. Although myofibroblasts are transiently found in normal wound healing, their persistence in fibrotic tissue, possibly due to their ability to evade apoptosis via upregulation of the antiapoptotic protein BCL-X_L, contributes to scar formation.

Explanted SSc fibroblasts display an abnormally activated phenotype ex vivo, characterized by increased collagen production, spontaneous ROS generation, shortening of their primary (nonmotile) cilia, prominent stress fibers, and constitutive expression of alpha smooth-muscle actin. Persistence of the "scleroderma phenotype" during serial ex vivo passage of SSc fibroblasts may reflect autocrine TGF-β stimulatory loops, deregulated microRNA expressions, cell-autonomous metabolic changes, or stable acquired epigenetic modifications. Single-cell RNA sequencing of SSc skin biopsies has revealed marked functional heterogeneity and diverse cell states across fibroblasts.

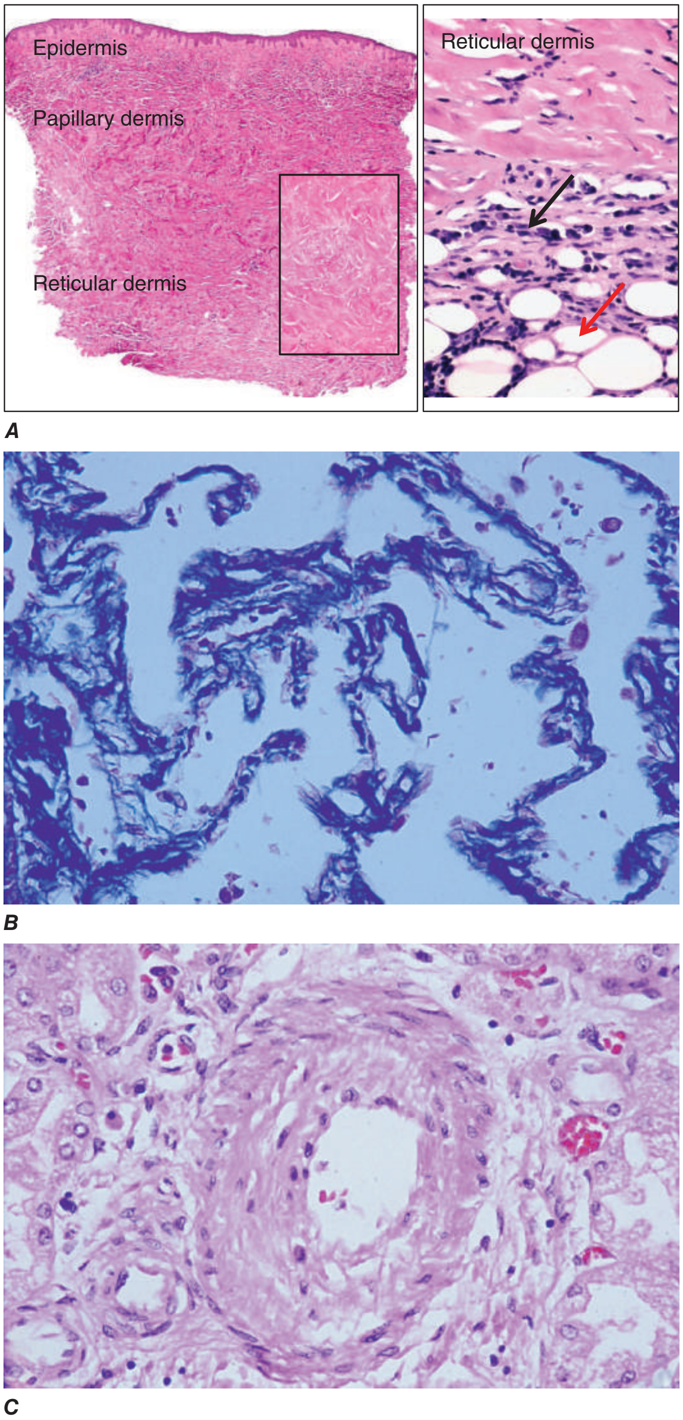

FIGURE 372-5 Pathologic findings in multiple organs in systemic sclerosis (SSc). ***A. Left panel:*** Skin is thickened due to fibrotic expansion of the dermis. Inset, higher magnification showing thick hyalinized collagen bundles replacing skin appendages. ***Right panel:*** Mononuclear inflammatory cells within the dermis and intradermal adipose tissue (indicated by *arrow*). ***B.*** Early SSc interstitial lung disease. Diffuse fibrosis of the alveolar septae and chronic inflammatory cell infiltrates. Trichrome stain. ***C.*** Pulmonary artery obliterative vasculopathy. Striking intimal and medial hyperplasia and luminal narrowing of small artery, with sparse inflammation and minimal interstitial lung fibrosis, in a patient with SSc pulmonary arterial hypertension.

PATHOLOGY

While pathologic findings vary across anatomic sites, the distinguishing hallmark of SSc irrespective of the organ system is widespread microangiopathy (fibroproliferative vasculopathy) capillary loss and obliteration, combined with fibrosis. In early-stage disease, perivascular inflammatory cell infiltrates composed of T and B lymphocytes, activated monocytes and macrophages, and mast cells may be detected as distinct cellular niches in multiple organs. A noninflammatory obliterative microangiopathy in the heart, lungs, kidneys, and gastrointestinal tract is a prominent late finding. Fibrosis most prominently affects the skin, lungs, cardiovascular system, gastrointestinal tract, tendon sheaths, perifascicular tissue surrounding skeletal muscle, and some endocrine organs such as the thyroid gland.

■ SKIN

In the skin, the dermis is thickened due to accumulation of broad bundles of homogenized collagen oriented parallel to the epithelium (Fig. 372-5*A*). Adnexal glands are atrophic, and loss of periadnexal and intradermal white adipose tissue and its replacement with collagen can be striking. While perivascular mononuclear cell infiltrates may be seen early, established skin fibrosis shows absence of inflammation.

■ LUNGS

Autopsy studies universally show evidence of lung involvement in both limited and diffuse cutaneous subsets of SSc. Most common is nonspecific interstitial pneumonia (NSIP) pattern characterized by variable interstitial fibrosis and mild chronic inflammation with

T lymphocytes, macrophages, and eosinophils. With progression, interstitial fibrosis and vascular damage dominate, often coexisting within the same biopsy. The usual interstitial pneumonia (UIP) pattern of spatial and temporal heterogeneity of inflammation and fibrosis and fibrotic foci, a hallmark of idiopathic pulmonary fibrosis, is less common in SSc **(Fig. 372-5*B*)**. Fibrosis of the alveolar septae results in obliteration of the airspaces and loss of pulmonary blood vessels. This process impairs gas exchange and contributes to pulmonary hypertension. Intimal thickening of the pulmonary arteries, best seen with elastin stain, underlies SSc-associated PAH **(Fig. 372-5C)** and, at autopsy, is often associated with pulmonary emboli and myocardial fibrosis. Patients with SSc-associated PAH also show fibrosis and intimal proliferation in preseptal venules and veins in the lung, accounting for venoocclusive disease. Lymphocytic bronchiolitis involving the submucosa of the terminal bronchioles is occasionally seen.

■ GASTROINTESTINAL TRACT

Pathologic changes can be found at any level from the mouth to the rectum. Atrophy and fibrosis of the muscularis propria and characteristic vascular lesions are prominent in the lower esophagus, while striated muscle in the upper third of the esophagus is generally spared. Collagenous replacement of the intestinal tract architecture results in impaired smooth-muscle contractility and diminished peristaltic activity, underlying gut dysmotility, bacterial overgrowth, small-bowel pseudo-obstruction, and perforation.

■ KIDNEYS

In the kidneys, vascular lesions affecting the interlobular and arcuate arteries predominate. Chronic renal ischemia is associated with shrunken glomeruli. Patients with scleroderma renal crisis, a life-threatening acute complication of SSc, show acute fibrinoid necrosis of afferent arterioles, followed by intimal proliferation (onion-skin pattern) and ischemic collapse of glomeruli. These changes are reminiscent of thrombotic microangiopathies such as atypical hemolytic-uremic syndrome **(Chaps. 105 and 120)** and are accompanied by thrombosis, thrombocytopenia due to platelet consumption, and intravascular hemolysis. Evidence of complement activation may be seen in kidney biopsies. Extensive vascular thrombosis, glomerular collapse, and peritubular capillary deposits predict irreversible renal failure.

■ HEART

Subclinical cardiac pathology is common in SSc and may affect the myocardium and pericardium. The characteristic arteriolar lesions in the heart are concentric intimal hypertrophy and luminal narrowing, patchy contraction band necrosis, loss of cardiac myocytes, and myocardial fibrosis due to microvascular involvement and ischemia-reperfusion injury. Fibrosis of the conduction system is also common, especially at the sinoatrial node. The frequency of epicardial atherosclerotic coronary artery disease may be increased in SSc compared to the general population, similar to other systemic inflammatory diseases. Pericardial involvement with chronic inflammatory infiltrates and fibrinous exudates is common and is sometimes associated with pericardial effusions.

■ PATHOLOGY IN OTHER ORGANS

Synovitis of the hands may occur in SSc; with disease progression, the synovium becomes fibrotic, and in contrast to rheumatoid disease, pannus formation or bone resorption (distinct from acro-osteolysis) is uncommon. Fibrosis of tendon sheaths and fascia, sometimes accompanied by calcifications, produces palpable and sometimes audible tendon friction rubs. Inflammation and, in later stages, atrophy and fibrosis of skeletal muscles are common findings and are similar to those in polymyositis. Fibrosis of the thyroid and of the minor salivary glands may be seen and underlie hypothyroidism and the sicca syndrome. Placentas from SSc pregnancies show decidual vasculopathy, which is associated with poor perinatal outcomes and fetal death.

TABLE 372-5 Frequency of Clinical Organ Involvement in Limited Cutaneous and Diffuse Cutaneous Systemic Sclerosis (SSc)

CLINICAL FEATURES	LIMITED CUTANEOUS SSc (%)	DIFFUSE CUTANEOUS SSc (%)
Skin involvement	90[a]	100
Raynaud's phenomenon	99	98
Ischemic digital ulcers	50	25
Esophageal involvement	90	80
Interstitial lung disease	35	65
Pulmonary arterial hypertension	15	15
Myopathy	11	23
Clinical cardiac involvement	9	12
Scleroderma renal crisis	2	10-15
Calcinosis cutis	30	50

[a]Approximately 10% of SSc patients have SSc sine scleroderma.

CLINICAL FEATURES

■ OVERVIEW

SSc is truly a systemic disease that can affect virtually any organ (Fig 372-1 and **Table 372-5**). Although a dichotomous stratification of patients into diffuse and limited cutaneous subsets (Table 372-2) is useful and used widely in clinical practice, SSc is in fact far more complex, and multiple distinct clusters or endophenotypes of SSc with characteristic manifestations and trajectories and outcomes can be recognized within each subset. Unique endophenotypes are associated with autoantibodies with distinct and mutually exclusive specificities (Table 372-4). Moreover, patients with "overlap" have typical features of SSc coexisting with clinical and laboratory evidence of another autoimmune disease, most commonly polymyositis, Sjögren's syndrome, polyarthritis, autoimmune liver disease, or SLE.

■ INITIAL CLINICAL PRESENTATION

Characteristic initial presentation is quite different in patients with the diffuse (dcSSc) versus limited (lcSSc) cutaneous forms of the disease. In dcSSc, the interval between Raynaud's phenomenon and subsequent onset of other disease manifestations is typically brief (weeks to months). Soft tissue swelling, puffy fingers, and pruritus are signs of the early inflammatory "edematous" phase. The fingers, distal limbs, and face are usually affected first. Diffuse hyperpigmentation of the skin, carpal tunnel syndrome, arthralgias, muscle weakness, fatigue, and decreased joint mobility are common. During the ensuing weeks to months, the inflammatory edematous phase evolves into the "fibrotic" phase, with skin induration associated with hair loss, reduced production of skin oils, and decline in sweating capacity. Progressive flexion contractures of the fingers ensue. The wrists, elbows, knees, and ankles become stiff due to fibrosis of the supporting joint structures. While proximally advancing skin involvement is the most visible manifestation of early dcSSc, important and clinically silent internal organ involvement can occur during this stage. The initial 4 years from disease onset is the period of most rapidly evolving and potentially irreversible lung and renal damage. If organ failure does not occur during this phase of dcSSc, the systemic process may plateau and stabilize. Increasingly recognized is an early stage of the disease, termed *very early diagnosis of SSc* (VEDOSS). These patients have puffy fingers, Raynaud's phenomenon, and SSc-specific autoantibodies and generally show progression to frank disease.

Compared to dcSSc, the course of lcSSc tends to be more indolent. The interval between onset of Raynaud's phenomenon and disease manifestations such as gastroesophageal reflux disease (GERD), cutaneous telangiectasia, ischemic digital ulcers, or soft tissue calcifications can be as long as years. Scleroderma renal crisis, significant ILD, and tendon friction rubs occur rarely in lcSSc, whereas PAH and overlap with keratoconjunctivitis sicca, polyarthritis, cutaneous vasculitis, and biliary cirrhosis can develop, even many years after disease onset.

ORGAN INVOLVEMENT

RAYNAUD'S PHENOMENON

Raynaud's phenomenon (also called secondary Raynaud's), the most frequent extracutaneous complication of SSc, is characterized by episodic vasoconstriction in the fingers and toes, sometimes also affecting the tip of the nose and earlobes. These attacks are typically self-limited and fully reversible, can be triggered by a decrease in temperature as well as emotional stress and vibration, and commonly start with pallor of the fingers, followed by cyanosis of variable duration. Hyperemia ensues spontaneously or with rewarming of the digit. The characteristic temporal progression of the three-color phases reflects the underlying vasoconstriction in arteriovenous anastomoses involved in thermoregulation, followed by ischemia and reperfusion. It is important to note that up to 5% of the general population has Raynaud's phenomenon. Recent GWAS revealed a robust link of Raynaud's with the *ADRA2A* gene, implicating deregulated α-adrenergic signaling in pathogenesis. Secondary Raynaud's occurs in SSc and other connective tissue diseases, as well as hematologic and endocrine conditions and occupational disorders, and can complicate treatment with beta blockers and anticancer drugs such as cisplatin and bleomycin. Distinguishing primary Raynaud's disease from secondary Raynaud's phenomenon can present a diagnostic challenge. Raynaud's disease is supported by the following: absence of an underlying cause; a family history of Raynaud's phenomenon; absence of digital tissue necrosis or ulceration; and a negative ANA test. Secondary Raynaud's phenomenon tends to occur at an older age, is more severe (episodes more frequent, prolonged, and painful), and is frequently complicated by ischemic digital ulcers and loss of digits **(Fig. 372-6)**.

Nailfold capillaroscopy using a low-power stereoscopic microscope or ophthalmoscope permits visualization of nailbed cutaneous capillaries under immersion oil **(Fig. 372-7)**. Raynaud's disease is associated with evenly spaced parallel vascular loops, whereas in secondary Raynaud's phenomenon, nailfold capillaries are distorted with widened and irregular loops, dilated lumen, microhemorrhages, and areas of vascular "dropout." Thus, nailfold capillaroscopy can be helpful for both differentiating primary from secondary Raynaud's phenomenon and for establishing the early diagnosis of SSc.

SKIN FEATURES

Bilateral symmetrical skin thickening is the hallmark that distinguishes SSc from other connective tissue diseases. Skin involvement starts in the fingers and characteristically advances from distal to proximal extremities in an ascending fashion. Some patients note diffuse tanning in the absence of sun exposure as a very early manifestation. In dark-skinned individuals, vitiligo-like hypopigmentation may occur. Pigment loss sparing the perifollicular areas gives rise to a "salt-and-pepper" appearance of the skin, most prominently on the scalp, upper back, and chest. Dermal sclerosis obliterating hair follicles, sweat glands, and eccrine and sebaceous glands causes hair loss, decreased sweating, and xerosis and itching in affected areas of the skin. Transverse creases on the dorsum of the fingers disappear **(Fig. 372-8)**. Fixed flexion contractures of the fingers cause reduced hand mobility and lead to muscle atrophy. Skin and subjacent tendon fibrosis accounts for fixed contractures of the wrists, elbows, and knees. Thick ridges at the neck due to firm adherence of skin to the underlying platysma muscle interfere with neck extension.

Patients with established SSc may show a characteristic "mauskopf" facial appearance with taut and shiny skin and occasionally an expressionless facies due to reduced mobility of the eyelids, cheeks, and mouth. Thinning of the lips with accentuation of the central incisor teeth and prominent perioral radial furrowing (rhytides) complete the picture. Reduced oral aperture (microstomia) interferes with eating and oral hygiene. The nose assumes a pinched, beak-like appearance. In late-stage disease, the skin becomes thin and atrophic and is firmly bound to the subcutaneous fat (tethering). Dilated skin capillaries 2–20 mm in diameter (telangiectasia), reminiscent of hereditary hemorrhagic telangiectasia, are frequently on the face, hands, lips, and oral

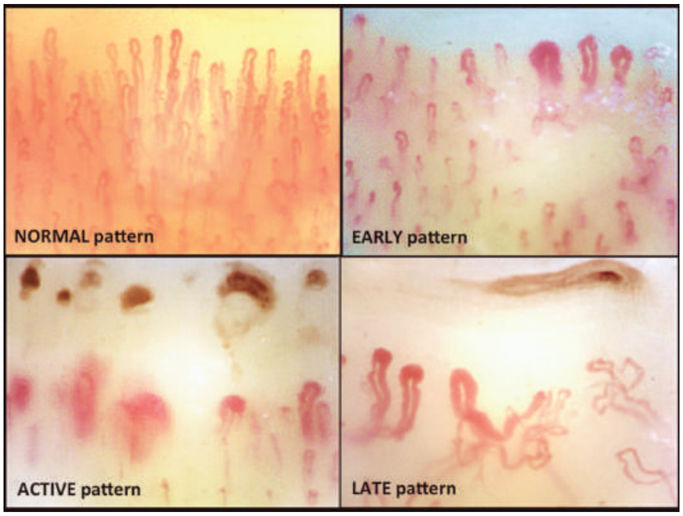

FIGURE 372-7 Systemic sclerosis–associated nailfold capillary alterations. In healthy subjects (normal pattern), note regularly arrayed and uniform-size "hairpin" microvessels. In early pattern, note dilations and symmetrically increased microvessels (giant capillaries). In active pattern, note giant capillaries, collapse with microhemorrhages, and loss of capillaries. In late pattern, note massive loss of capillaries, fibrosis, and neoangiogenesis with secondary dilations (nailfold videocapillaroscopy; magnification 220×). *(Courtesy Professor Maurizio Cutolo, University of Genoa, Italy.)*

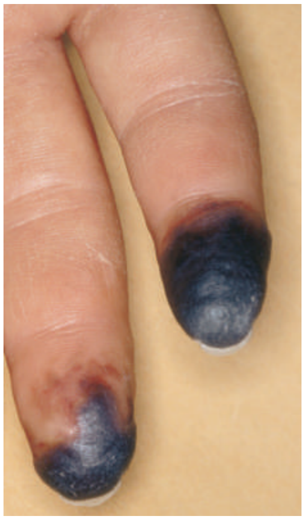

FIGURE 372-6 Digital necrosis. Sharply demarcated ischemic necrosis of the fingertip in a patient with limited cutaneous systemic sclerosis (SSc) associated with severe Raynaud's phenomenon.

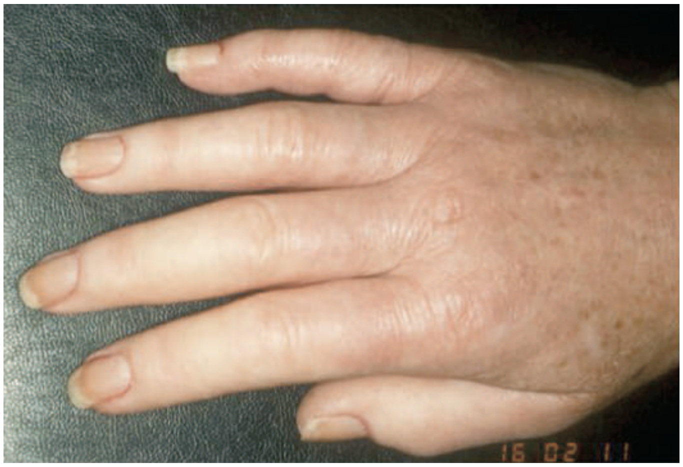

FIGURE 372-8 Sclerodactyly. Note puffy and indurated fingers and fixed flexion contractures of proximal interphalangeal joints in a patient with limited cutaneous systemic sclerosis (lcSSc).

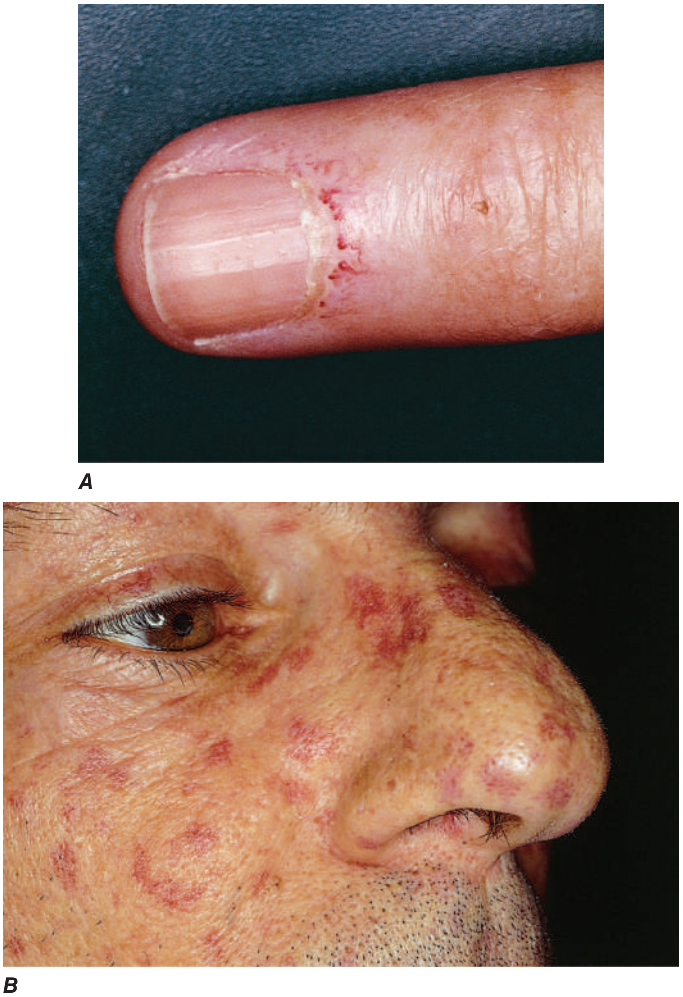

FIGURE 372-9 Cutaneous vascular changes in systemic sclerosis. ***A.*** Vascular changes at the finger nailfold in a patient with limited cutaneous systemic sclerosis (lcSSc). ***B.*** Telangiectasia on the face.

mucosa **(Fig. 372-9)**. The number of telangiectasias correlates with the severity of microvascular disease, including PAH. Breakdown of atrophic skin leads to chronic ulcerations at the extensor surfaces of the proximal interphalangeal joints, the volar pads of the fingertips, and bony prominences such as elbows and malleoli. Ulcers are often painful, heal slowly, and become secondarily infected, resulting in osteomyelitis. Healing of ischemic fingertip ulcerations leaves characteristic fixed digital "pits." Loss of soft tissue at the fingertips due to ischemia may be associated with striking resorption of the terminal phalanges (acro-osteolysis) **(Fig. 372-10)**.

Dystrophic calcifications in the skin and subcutaneous and soft tissues (calcinosis cutis) in the presence of normal serum calcium and phosphate levels occur in up to 40% of patients, most commonly in those with long-standing anti-centromere antibody–positive lcSSc. Calcific deposits, composed of calcium hydroxyapatite crystals, vary in size from tiny punctate lesions to large conglomerate masses and can be readily visualized on plain radiographs or dual-energy computed tomography (CT) imaging. These deposits occur when calcium precipitates in tissue damaged by inflammation, hypoxia, or local trauma. Common locations include the finger pads, palms, extensor surfaces of the forearms, and the olecranon and prepatellar bursae **(Fig. 372-11)**. Large calcific deposits can cause pain and nerve compression and may ulcerate through the overlying skin with drainage of chalky white material and secondary infections. Paraspinal sheet calcifications may cause neurologic complications.

PULMONARY FEATURES

The two principal forms of lung involvement in SSc, ILD and pulmonary vascular disease, are frequent and together account for a majority of SSc-related deaths. Survival is particularly poor in SSc patients with concurrent presence of these two processes. Less common pulmonary complications of SSc include aspiration pneumonitis associated with chronic gastroesophageal reflux, pulmonary hemorrhage due to endobronchial telangiectasia, obliterative bronchiolitis, pleural reactions, restrictive physiology due to chest wall fibrosis, spontaneous pneumothorax, and drug-induced lung toxicity. The incidence of lung cancer is increased in SSc.

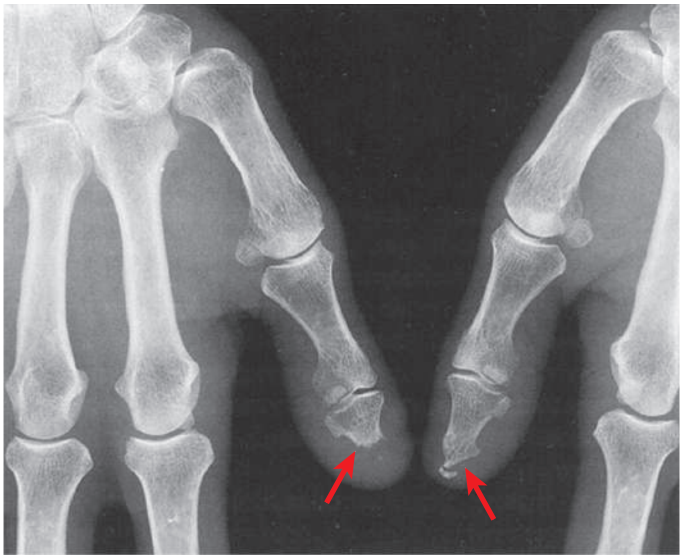

FIGURE 372-10 Acro-osteolysis. Note dissolution of distal terminal phalanges *(arrows).* Acro-osteolysis is associated with digital ischemia and is seen in patients with long-standing limited cutaneous systemic sclerosis (lcSSc) and Raynaud's phenomenon.

Interstitial Lung Disease While evidence of ILD can be found in up to 65% of SSc patients by high-resolution CT (HRCT), clinically significant ILD develops in 16–43%; the frequency varies depending on the detection method used. Risk factors for significant ILD include male sex, African-American race, diffuse skin involvement, severe gastroesophageal reflux, and presence of topoisomerase I autoantibodies, whereas SSc patients who are anti-centromere antibody–positive have a lower risk of ILD. Additional risk factors for significant ILD include low forced vital capacity (FVC) or single-breath diffusing capacity of the lung for carbon monoxide (DL_{CO}) at initial presentation. Esophageal dilatation with chronic acid reflux, common in SSc, causes recurrent microaspiration, which is a risk factor for development and progression of ILD. The most rapid progression in ILD generally occurs early in the disease course (within the first 3–5 years), when the FVC can decline by 30% per year. In contrast, new offset of ILD is rare in SSc patients with long-standing disease.

Pulmonary involvement may remain asymptomatic until it is advanced. The most common presenting respiratory symptoms—exertional dyspnea, fatigue, and reduced exercise tolerance—are subtle and slowly progressive. A chronic dry cough may be present. Physical examination may reveal fine inspiratory "Velcro" crackles at the lung bases. Pulmonary function testing (PFT) is relatively insensitive for detecting early pulmonary involvement, although it is central for monitoring its course. In patients with established SSc-ILD, PFT typically shows a restrictive ventilatory defect (FVC $<70\%$ predicted and/or forced expiratory volume in 1 s [FEV_1]/FVC ratio >0.8 and reduced total lung capacity [TLC] and DL_{CO}). A reduction in DL_{CO} that is significantly out of proportion to the reduction in lung volumes should raise suspicion for pulmonary vascular disease but may also be due to anemia. Oxygen desaturation with exercise is common.

Chest radiography can be used to rule out infection and other causes of respiratory symptoms in SSc, but compared to HRCT, it is relatively insensitive for detection of early ILD. Characteristic imaging findings include bilateral lower lobe subpleural reticular linear opacities and ground-glass opacifications with an apicobasal gradient,

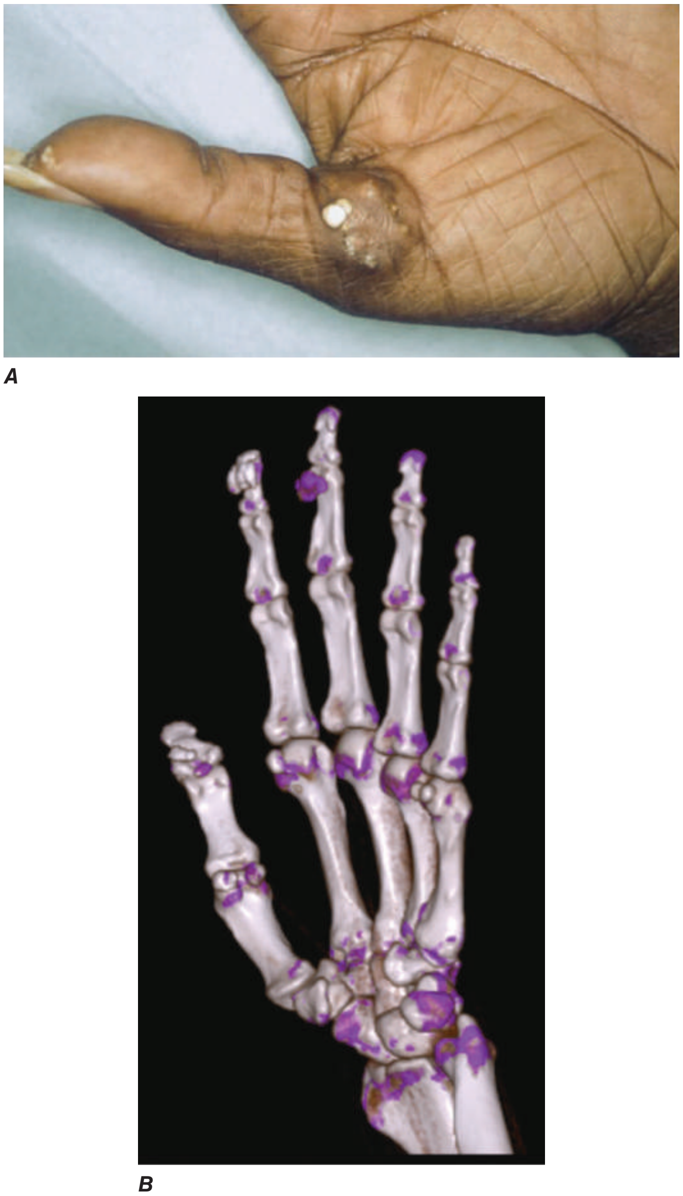

FIGURE 372-11 Calcinosis cutis in systemic sclerosis. ***A.*** Note calcific deposit breaking through the skin in a patient with limited cutaneous systemic sclerosis (lcSSc). ***B.*** Dual-energy computed tomography showing calcific deposits at the proximal interphalangeal joints.

even in asymptomatic patients with normal PFTs **(Fig. 372-12)**. Additional HRCT findings include mediastinal lymphadenopathy, pulmonary nodules, traction bronchiectasis, and uncommonly, honeycomb changes. The extent of interstitial changes on chest HRCT at baseline is a predictor of ILD progression and mortality. While bronchoalveolar lavage (BAL) can demonstrate inflammatory cells in the lower respiratory tract and may be useful for ruling out tuberculosis and other infections, it has limited utility for SSc diagnosis or for identifying reversible alveolitis. Surgical lung biopsy in SSc is indicated only in patients with atypical findings on chest imaging. The histologic pattern on lung biopsy may predict the risk of progression of ILD, with the NSIP pattern having a better prognosis than UIP.

Pulmonary Arterial Hypertension (PAH) PAH results from vascular remodeling of small (<500 μm) pulmonary arteries. PAH develops in 8–12% of patients with SSc as a late complication and can be an isolated abnormality or associated with ILD. PAH is defined as a mean pulmonary artery pressure (mPAP) >20 mmHg with a pulmonary capillary wedge pressure ≤15 mmHg and pulmonary vascular resistance >2.0 Wood units. The natural history of SSc-associated PAH is variable but often follows a progressive course leading to right heart failure. The 3-year survival of SSc patients with untreated PAH is <50%. Risk factors include limited cutaneous disease, older age at disease onset, high numbers of cutaneous telangiectasia, and antibodies to centromere, U3-RNP (fibrillarin), and B23. Mutations in the *BMPR2* gene implicated in idiopathic PAH are not associated with SSc-PAH.

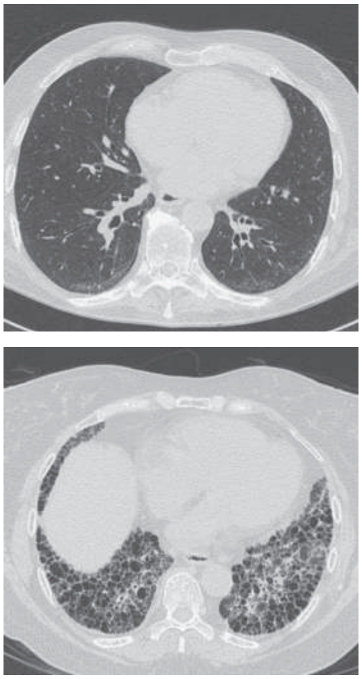

FIGURE 372-12 High-resolution chest computed tomography scans in systemic sclerosis. ***Top panel:*** Early-stage interstitial lung disease with subtle subpleural reticulations and ground-glass opacities in the lower lobes. Patient in supine position. ***Bottom panel:*** Extensive lung fibrosis with coarse reticular honeycombing and traction bronchiectasis. *(Courtesy of Dr. Rishi Agrawal, Northwestern University.)*

In SSc, PAH is often asymptomatic in early stages. Patients present with exertional dyspnea and reduced exercise capacity. With progression, angina, near-syncope, and symptoms and signs of right-sided heart failure appear. Physical examination may show tachypnea, a loud pulmonic component of the S_2 heart sound, pulmonic/tricuspid regurgitation murmur, palpable right ventricular heave, elevated jugular venous pressure, and dependent edema. Doppler echocardiography is a noninvasive screening method for estimating the pulmonary arterial pressure. In light of the poor prognosis of untreated PAH and better therapeutic response in patients with early diagnosis, SSc patients should be screened for PAH at initial evaluation and annually thereafter. Estimated pulmonary artery systolic pressure >40 mmHg at rest or tricuspid regurgitation jet velocities >3 m/s suggest PAH. PFT may show a reduced DL_{CO} in isolation or out of proportion with the severity of restriction. Because echocardiography can over- or underestimate pulmonary artery pressures, cardiac catheterization is required to confirm the diagnosis of suspected PAH; assess its severity, including the degree of right heart dysfunction; rule out venoocclusive disease and other cardiac (postcapillary) causes of pulmonary hypertension; and provide prognostic parameters. Distinguishing PAH in SSc patients from pulmonary hypertension secondary to pulmonary fibrosis and hypoxia can be difficult. Serum levels of N-terminal pro-brain natriuretic peptide (NT-proBNP) correlate with the presence and severity of PAH, as well as survival. While NT-proBNP can be useful in PAH screening and in monitoring the response to treatment, elevated levels are not specific for PAH and also occur in other forms of right and left heart disease. Despite more favorable hemodynamics of SSc-associated PAH, its prognosis is worse, and treatment response poorer, than idiopathic PAH. This is most likely due to frequent concurrence of ILD and cardiac involvement in SSc patients.

■ GASTROINTESTINAL INVOLVEMENT

Variable involvement of the gastrointestinal (GI) tract, which can affect any level, is the most common internal organ manifestation of

TABLE 372-6 Prominent Gastrointestinal Manifestations of Systemic Sclerosis (SSc) and Their Management

SITE IN GASTROINTESTINAL TRACT	PRINCIPAL MANIFESTATION	MANAGEMENT
Oropharynx	Diminished oral aperture Dry mouth Periodontitis, gingivitis Swallowing	Periodontal care Artificial saliva Swallowing therapy
Esophagus	Reflux Dysphagia Strictures Barrett's metaplasia	Lifestyle modifications, Prokinetic drugs Proton pump inhibitors Endoscopic procedures
Stomach	Gastroparesis Gastric antral vascular ectasia (GAVE; watermelon stomach)	Prokinetic agents Endoscopic laser cryotherapy
Small and large intestines	Bacterial overgrowth (small intestinal bacterial overgrowth) Diarrhea/constipation Pseudo-obstruction Pneumatosis intestinalis Malabsorption Colonic pseudo-diverticula	Laxatives Prokinetic agents Rotating antibiotics Octreotide Parenteral nutritional support
Anorectum	Sphincter incompetence	Biofeedback, sacral nerve stimulation, surgery

SSc, seen in up to 90% of SSc patients with both limited and diffuse cutaneous disease **(Table 372-6)**. Severe GI tract involvement is associated with male sex and specific autoantibodies. Pathologic findings of GI involvement in SSc include fibrosis, smooth-muscle atrophy, and obliterative small-vessel vasculopathy throughout the length of the GI tract. Together, they have a major impact on quality of life, malnutrition, and mortality.

Upper GI Tract Involvement Decreased oral aperture interferes with regular dental hygiene. Loss of periodontal ligament attaching teeth to the alveolar bone leads to teeth loosening. Additional oropharyngeal manifestations due to a combination of xerostomia, shortened frenulum, and resorption of the mandibular condyles cause much distress. Most SSc patients show symptoms of GERD: heartburn, regurgitation, and dysphagia, due to a combination of reduced lower esophageal sphincter pressure resulting in reflux, impaired esophageal clearance of refluxed gastric contents due to diminished motility, and delayed gastric emptying. Calcium channel antagonists and phosphodiesterase inhibitors commonly used to treat Raynaud's phenomenon in SSc can further aggravate reflux. Esophageal manometry shows abnormal motility in most patients, even in the absence of symptoms. Common extraesophageal GERD manifestations in SSc include hoarseness, cough, and chronic or recurrent microaspiration, which can aggravate underlying ILD. Characteristic chest CT findings include dilated patulous esophagus with intraluminal air. Esophageal dilation is associated with the severity of interstitial lung disease. Endoscopy may be indicated in patients with dysphagia in order to rule out opportunistic infections with *Candida*, herpes virus, and cytomegalovirus. Severe erosive esophagitis may be found in patients with minimal symptoms. Esophageal strictures and Barrett's esophagus may complicate chronic GERD in SSc patients. Because Barrett's metaplasia is associated with increased risk of adenocarcinoma, these patients require regular surveillance endoscopy with biopsy.

Gastroparesis with early satiety, abdominal distention, and aggravated reflux symptoms is common. Barium contrast studies are neither sensitive nor specific for evaluation of gastric involvement in SSc. Gastric antral vascular ectasia (GAVE) in the antrum may occur. These subepithelial lesions, reflecting the widespread small-vessel vasculopathy of SSc, are called "watermelon stomach" due to their endoscopic appearance. GAVE may present with acute life-threatening or recurrent episodes of GI bleeding, resulting in chronic unexplained iron-deficiency anemia. Patients with SSc show increased prevalence of *Helicobacter pylori* infection in the stomach.

Lower GI Tract and Anorectal Involvement Weight loss and malnutrition due to a combination of impaired intestinal motility, malabsorption, and chronic diarrhea secondary to bacterial overgrowth are common. Fat and protein malabsorption as well as vitamin B_{12} and vitamin D deficiency ensue and may be further exacerbated by pancreatic insufficiency. Disturbed intestinal motor function associated with delayed colonic transit may present with intestinal pseudo-obstruction. Patients present with acute abdominal pain, nausea, and vomiting, and radiographic studies show acute intestinal obstruction. A major diagnostic challenge is differentiating pseudo-obstruction, which often responds to supportive care and intravenous nutritional supplementation, from mechanical obstruction. Colonic involvement in SSc may cause constipation, occasionally complicated by sigmoid volvulus. Anorectal dysfunction commonly presents with fecal incontinence. GI bleeding can result from telangiectasia and rectal prolapse. In late-stage SSc, wide-mouth sacculations or diverticula occur in the colon and occasionally cause perforation and bleeding. An unusual radiologic finding is pneumatosis cystoides intestinalis, which is due to air trapping in the bowel wall that may rarely rupture and cause benign pneumoperitoneum. Studies of the fecal microbiota in SSc show a reduction in protective butyrate-producing bacteria, potentially promoting a proinflammatory intestinal microenvironment. Although the liver is rarely affected in patients with SSc, primary biliary cirrhosis may occur, particularly in patients with limited cutaneous disease.

RENAL INVOLVEMENT: SCLERODERMA RENAL CRISIS

Scleroderma renal crisis presents with accelerated hypertension accompanied by acute kidney injury and progressive failure. This acute life-threatening complication occurs in <15% of SSc patients, almost always within 4 years of disease onset. Scleroderma renal crisis can occasionally even be the presenting manifestation of SSc. While short-term survival in scleroderma renal crisis was <10% prior to the advent of angiotensin-converting enzyme (ACE) inhibitors, outcomes for this serous complication have shown great improvement. The pathogenesis involves obliterative vasculopathy of the renal arcuate and interlobular arteries, with consequent intravascular hemolysis **(Fig. 372-13)**. Progressive reduction in renal blood flow, aggravated by vasospasm, leads to increased juxtaglomerular renin secretion and angiotensin II generation, with further renal vasoconstriction resulting in a vicious cycle that culminates in accelerated hypertension. Risk factors for scleroderma renal crisis include African-American race, male sex, and diffuse or progressive skin involvement. Up to 50% of patients with scleroderma renal crisis have antibodies to anti-RNA polymerase III, whereas patients with anti-centromere antibodies appear protected. Palpable tendon friction rubs, pericardial effusion, new unexplained anemia, and thrombocytopenia may be harbingers of impending scleroderma renal crisis. High-risk patients with early-stage SSc should monitor their blood pressure daily. Because glucocorticoid use is associated with scleroderma renal crisis, prednisone in high-risk SSc patients should be taken only when absolutely required and at low doses (<10 mg/d).

Patients with scleroderma renal crisis characteristically present with accelerated hypertension (generally >150/90 mmHg) and progressive oliguric renal insufficiency. However, ~10% of patients present with blood pressure in the normal range. Normotensive renal crisis is generally associated with a poor outcome. Headache, blurred vision, congestive heart failure, and pulmonary edema may accompany elevation of blood pressure. Moderate thrombocytopenia and microangiopathic hemolysis with fragmented red blood cells can be seen, and urinalysis typically shows mild proteinuria, granular casts, and microscopic hematuria. Progressive oliguric renal failure over several days generally follows. Scleroderma renal crisis is sometimes initially misdiagnosed

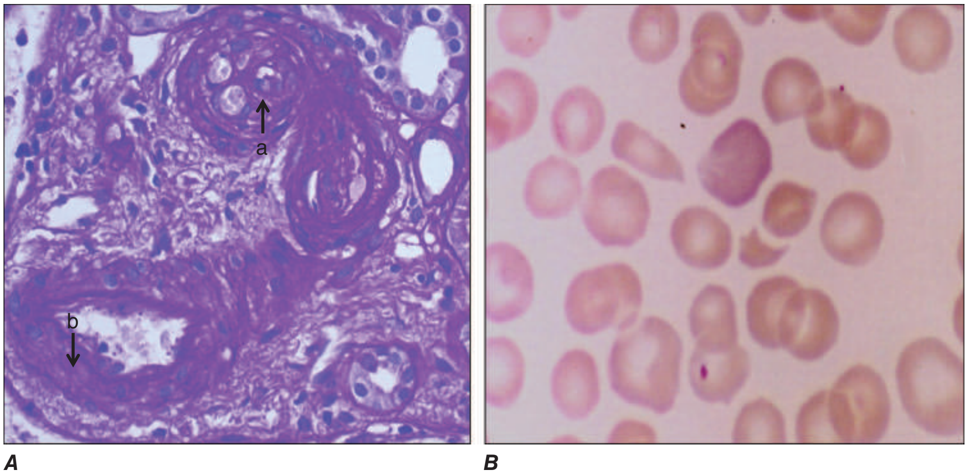

FIGURE 372-13 Pathologic findings in scleroderma renal crisis. ***A.*** Kidney biopsy demonstrating intimal proliferation *(arrows)* and myxoid changes in medium-sized renal arteries. ***B.*** Fragmentation of circulating red blood cells due to intravascular hemolysis. *(Courtesy of Drs. Edward Stern and Christopher Denton, Royal Free Hospital, London, UK.)*

as thrombotic thrombocytopenic purpura (TTP) or other forms of thrombotic microangiopathy. In such cases, renal biopsy and determination of serum von Willebrand factor–cleaving protease activity may be of benefit. The presence of oliguria or a creatinine >3 mg/dL at initial presentation predicts poor outcome (permanent hemodialysis and mortality), as do vascular thrombosis and glomerular ischemic collapse on renal biopsy.

■ CARDIAC INVOLVEMENT

Although it is often silent, cardiac involvement is detected in 10–50% of SSc patients screened with sensitive diagnostic tools. Clinical cardiac involvement is more frequent in dcSSc than in lcSSc and may be primary or secondary to PAH, ILD, or renal involvement. Cardiac involvement in SSc is associated with poor outcomes. The endocardium, myocardium, and pericardium may each be affected separately or together. Pericardial involvement is manifested as pericarditis, pericardial effusions, constrictive pericarditis, and rarely, cardiac tamponade. Conduction system fibrosis is common and may be silent or manifested by heart block. Other arrhythmias include premature ventricular contractions, atrial fibrillation, and supraventricular tachycardia. Microvascular involvement, recurrent vasospasm, and ischemia-reperfusion injury contribute to patchy myocardial fibrosis, and the resulting systolic or diastolic left ventricular dysfunction may progress to overt heart failure. Acute or subacute myocarditis leading to left ventricular dysfunction may occur, best diagnosed by cardiac magnetic resonance imaging or endomyocardial biopsy. While conventional echocardiography has low sensitivity for detecting preclinical heart involvement in SSc, newer modalities such as tissue Doppler echocardiography (TDE), cardiac magnetic resonance imaging (cMRI), and nuclear imaging (single photon emission CT [SPECT]) reveal a high prevalence of abnormal myocardial function or perfusion. The serum levels of NT-proBNP, a ventricular hormone elevated in SSc-PAH, may also have utility as markers of primary cardiac involvement.

■ MUSCULOSKELETAL COMPLICATIONS

Musculoskeletal complications are commonly seen in SSc. Carpal tunnel syndrome may be a presenting disease manifestation, and generalized arthralgia and stiffness are prominent in early disease. Mobility of both small and large joints is progressively impaired, and fixed contractures at the proximal interphalangeal joints and wrists ensue. Large joint contractures occur in patients with dcSSc and are frequently accompanied by tendon friction rubs. These are characterized by coarse leathery crepitation heard or palpated upon passive joint movement and are caused by fibrosis and adhesion of the tendon sheaths and fascial planes at the affected joint. The presence of tendon friction rubs is associated with increased risk for renal and cardiac complications and reduced survival. Synovitis seen on ultrasound or MRI is common; occasional SSc patients develop erosive polyarthritis in the hands, and some may have a seropositive rheumatoid arthritis overlap. Muscle weakness is common and multifactorial; deconditioning, disuse atrophy, malnutrition, inflammation, and fibrosis may all contribute. A chronic noninflammatory form of myopathy characterized by atrophy and fibrosis with mildly elevated muscle enzymes occurs in late-stage SSc. Bone resorption in the terminal phalanges, possibly reflecting recurrent ischemia, underlies the characteristic loss of the distal tufts (acro-osteolysis) seen on plain radiographs (Fig. 372-5). Resorption of the mandibular condyles can lead to bite difficulties. Osteolysis can also affect the ribs and distal clavicles.

■ LESS RECOGNIZED DISEASE MANIFESTATIONS

Dry eyes and dry mouth (sicca complex) are common in SSc. Biopsy of the minor salivary glands in these cases shows fibrosis rather than focal lymphocytic infiltration characteristic of primary Sjögren's syndrome **(Chap. 373)**. Hypothyroidism resulting from Graves or Hashimoto disease is common, particularly in lcSSc, and may be underrecognized. Whereas the central nervous system is generally spared in SSc, unilateral or bilateral sensory trigeminal neuropathy can occur. Erectile dysfunction is a frequent and occasionally initial disease manifestation. Inability to attain or maintain penile erection is due to vascular insufficiency and fibrosis of corporeal smooth muscle and responds poorly to medical therapy. Sexual performance is also adversely affected in women. While fertility is not impaired in SSc, pregnancy is associated with higher risk of adverse fetal outcomes. Furthermore, cardiopulmonary involvement may worsen during pregnancy, and new onset of scleroderma renal crisis has been described.

Cancer Epidemiologic studies indicate an increased cancer risk in patients with SSc. Lung cancer and esophageal adenocarcinoma typically occur in the setting of long-standing ILD or GERD and may be linked to chronic inflammation and maladaptive repair. In contrast, breast, lung, and ovarian carcinomas and lymphomas tend to occur in close temporal association with the onset of SSc, particularly in patients who are positive for autoantibodies to RNA polymerase III. In this scenario, SSc is thought to represent a paraneoplastic syndrome that is triggered by the antitumor immune response.

■ LABORATORY EVALUATION AND BIOMARKERS

Mild microcytic anemia is frequent and may indicate recurrent GI bleeding caused by GAVE or chronic esophagitis. Macrocytic anemia may be caused by folate and vitamin B_{12} deficiency due to small-bowel bacterial overgrowth and malabsorption or by drugs such as methotrexate. Microangiopathic hemolytic anemia caused by mechanical fragmentation of red blood cells during their passage through microvessels coated with fibrin or platelet thrombi is a hallmark of scleroderma renal crisis. Elevation of C-reactive protein (CRP) in SSc is indicative of higher risk of progressive fibrosis and mortality. Similarly, an elevated serum level of IL-6 is associated with worse outcomes. The erythrocyte sedimentation rate (ESR) is generally normal in SSc; an elevation may signal coexisting myositis or malignancy.

Circulating antinuclear autoantibodies are detected in almost all patients with SSc. Anti-topoisomerase I (Scl-70) and anti-centromere antibodies are mutually exclusive and highly specific for SSc. Topoisomerase I antibodies are associated with increased risk of ILD and poor outcomes. Anti-centromere antibodies are associated with PAH, but only infrequently with significant cardiac, pulmonary, or renal involvement. Nucleolar immunofluorescence pattern may indicate antibodies to U3-RNP (fibrillarin), Th/To, or PM/Scl, whereas speckled immunofluorescence indicates antibodies to RNA polymerase III,

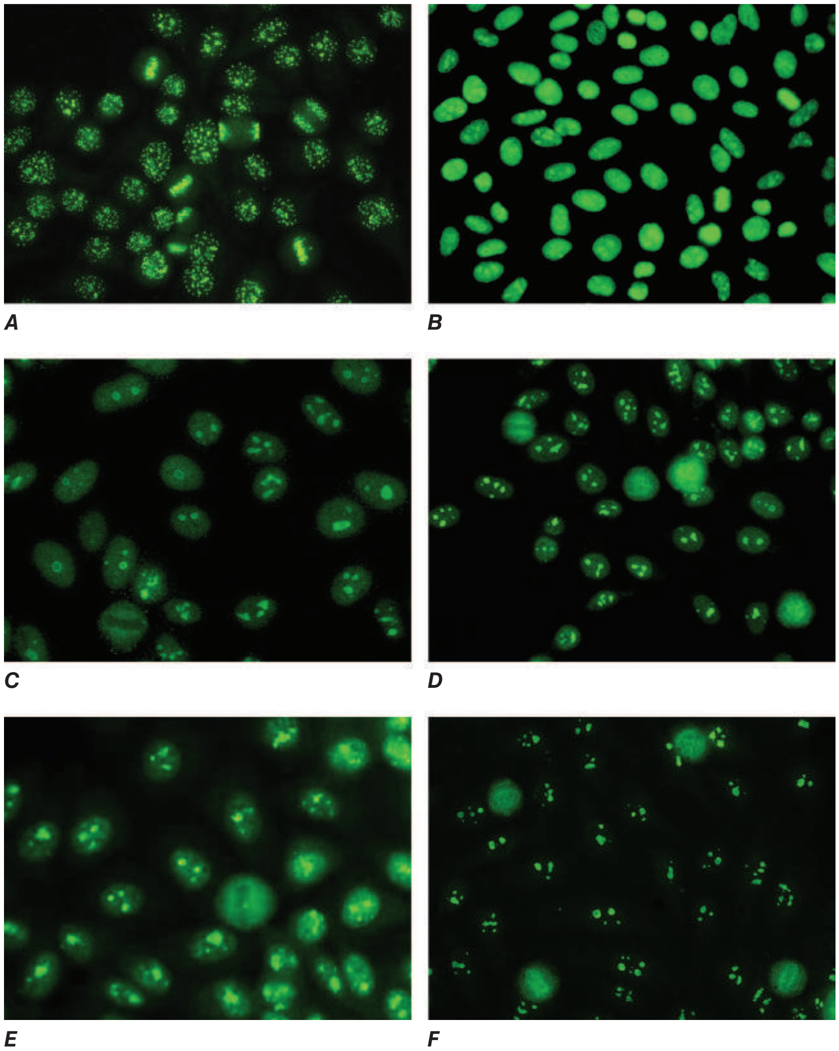

FIGURE 372-14 Systemic sclerosis (SSc)-associated autoantibodies: immunofluorescence. Indirect immunofluorescence of SSc serum samples using HEp-2 substrate. Representative images showing ***(A)*** anti-centromere; ***(B)*** anti-Scl-70/topoisomerase I; ***(C)*** anti-PM/Scl; ***(D)*** anti-Th/To; ***(E)*** anti-RNA polymerase III; and ***(F)*** anti-fibrillarin/U3RNP antibodies. Variations in immunostaining are clues to autoantibody specificity, but immunoassays with purified autoantigens are needed to confirm antigen specificity. *(Courtesy of Marvin Fritzler and Susan Copple, Inova Diagnostics Inc., San Diego, CA.)*

associated with increased risk of scleroderma renal crisis and malignancy **(Fig. 372-14)**.

■ DIAGNOSIS, STAGING, AND MONITORING

The diagnosis of SSc is made primarily on clinical grounds and is generally straightforward in patients with established disease. Diagnostic criteria developed for classification are >90% specific and sensitive for SSc. The presence of skin induration with a characteristic symmetric distribution pattern associated with typical visceral organ manifestations establishes the diagnosis of SSc with a high degree of certainty. In lcSSc, a history of Raynaud's phenomenon and GERD symptoms, coupled with sclerodactyly and nailfold capillary changes, often in combinations with cutaneous telangiectasia and calcinosis cutis, helps to establish the diagnosis. Primary Raynaud's disease is a benign condition that must be differentiated from early or limited SSc. Nailfold microscopy can be particularly helpful in this situation, because in contrast to SSc, nailfold capillaries are normal. Diagnosing SSc at an early stage may be a particular challenge. In dcSSc, initial symptoms are often nonspecific, Raynaud's phenomenon may be absent, and physical examination may show only upper extremity edema and puffy fingers. Early-stage SSc might therefore occasionally be misdiagnosed as arthritis, SLE, myositis, or, most commonly, undifferentiated connective tissue disease leading to delays in diagnosis. Within weeks to months, Raynaud's phenomenon and advancing skin induration appear. Detection of SSc-specific autoantibodies provides a high degree of diagnostic certainty. Raynaud's phenomenon with fingertip ulcerations or other evidence of digital ischemia, coupled with telangiectasia, distal esophageal dysmotility, unexplained ILD or PAH, or accelerated hypertension with renal failure in the absence of clinically evident skin induration, suggests the diagnosis of SSc sine scleroderma. Patients with a new diagnosis of SSc should be screened for ILD, followed by regular pulmonary monitoring for several years **(Fig. 372-15)**.

MANAGEMENT OF SYSTEMIC SCLEROSIS

■ OVERVIEW: GENERAL PRINCIPLES

To date, few therapies have been shown to significantly alter the natural history of SSc. In contrast, a variety of interventions are effective in alleviating symptoms, slowing progression of the cumulative organ damage, and reducing disability. Moreover, a significant decrease in disease-related mortality has been noted during the past 25 years. In light of the marked heterogeneity in disease manifestations, course, and outcomes, optimized care for patients with SSc requires a thoughtful "precision medicine" approach that is tailored specifically to each individual patient's unique needs.

The following general principles should guide management of patients with SSc **(Table 372-7)**: prompt and accurate diagnosis; disease subclassification and risk stratification based on clinical, radiologic, and laboratory evaluation, including autoantibody profiles and prognostic and predictive biomarkers; early recognition of organ-based complications and careful assessment of their extent, severity, and likelihood of deterioration; regular monitoring for disease progression, new complications,

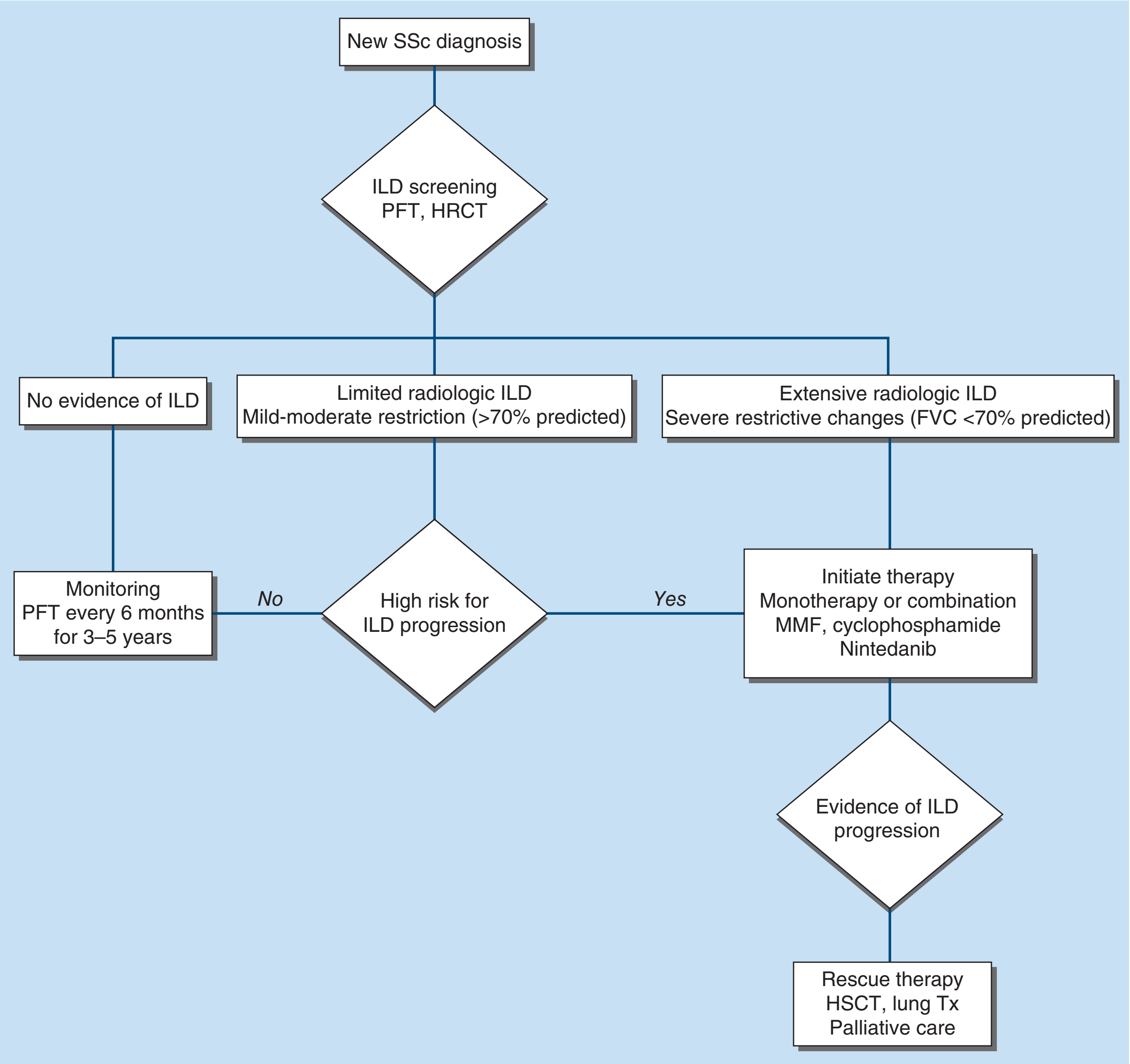
New SSc diagnosis
ILD screening
PFT, HRCT
No evidence of ILD
Limited radiologic ILD
Mild-moderate restriction (>70% predicted)
Extensive radiologic ILD
Severe restrictive changes (FVC <70% predicted)
Monitoring
PFT every 6 months
for 3–5 years
No
High risk for
ILD progression
Yes
Initiate therapy
Monotherapy or combination
MMF, cyclophosphamide
Nintedanib
Evidence of ILD
progression
Rescue therapy
HSCT, lung Tx
Palliative care

FIGURE 372-15 Algorithm for screening, monitoring, and treatment of systemic sclerosis (SSc)-associated interstitial lung disease (ILD). FVC, forced vital capacity; HRCT, high-resolution computed tomography; HSCT, hematopoietic stem cell transplantation; MMF, mycophenolate mofetil; PFT, pulmonary function testing; Tx, transplantation. *(Adapted from A Perelas et al: Lancet Respir Med 8:304, 2020.)*

and response to and side effects of therapy; adjusting therapy; and patient education. To minimize the risk of irreversible organ damage, management should be proactive, with regular screening and initiation of appropriate intervention at the earliest possible opportunity. In light of the complex and multisystemic nature of SSc, a team-oriented management approach integrating appropriate specialists should be pursued. Patients should be encouraged to become familiar with potential complications and empowered to partner with their treating physicians. This requires a long-term relationship between patient and physician, with ongoing counseling, encouragement, and two-way dialogue.

DISEASE-MODIFYING THERAPIES: IMMUNOMODULATORY AGENTS

Immunomodulatory agents commonly used in other autoimmune diseases have generally shown modest or no benefit in SSc. Glucocorticoids alleviate stiffness and aching in early inflammatory-stage dcSSc but do not influence the progression of skin or internal organ involvement. Since their use is associated with an increased risk of scleroderma renal crisis, glucocorticoids should be given only when absolutely necessary, at the lowest dose possible, and for brief periods only.

TABLE 372-7 Key Principles in Management

- Establish early and accurate diagnosis.
- Detect and evaluate internal organ involvement.
- Define clinical disease stage and activity.
- Tailor individualized therapy to each patient's unique needs.
- Assess treatment response, and adjust therapy as needed; monitor for disease activity, progression, and new complications.

Cyclophosphamide has been extensively studied in light of its efficacy in the treatment of vasculitis **(Chap. 375)**, SLE **(Chap. 368)**, and other autoimmune diseases **(Chap. 367)**. Both oral and intravenous cyclophosphamide have been shown to reduce the progression of SSc-associated ILD, with stabilization and, less commonly modest improvement, of pulmonary function, HRCT findings, respiratory symptoms, and skin induration. These benefits need to be balanced against cyclophosphamide's potential toxicity, including bone marrow suppression, opportunistic infections, hemorrhagic cystitis and bladder cancer, premature ovarian failure, and late secondary malignancies.

Methotrexate had only modest effect on SSc skin involvement in small studies. Mycophenolate mofetil, a prodrug of mycophenolic acid, is an inhibitor of inosine monophosphate dehydrogenase, which impairs B- and T-cell proliferation. Mycophenolate was evaluated in both open-label and randomized controlled trials. Both skin induration and ILD improved in patients treated with mycophenolate mofetil, and the drug was well tolerated. Tocilizumab, a monoclonal antibody that blocks IL-6 receptor signaling, also showed some benefit on both skin and lung involvement in randomized SSc trials. Tocilizumab was well tolerated and has a role in the treatment of SSc-associated ILD in relatively early disease. Open-label studies and small clinical trials support the use of rituximab, a monoclonal antibody directed against CD20 that depletes peripheral B cells but spares plasma cells and hematopoietic stem cells. Other treatments occasionally used include extracorporeal photopheresis, IV immunoglobulin, and abatacept, a fusion protein that inhibits T-cell co-stimulation and function. The use of cyclosporine, azathioprine, hydroxychloroquine, thalidomide, and rapamycin for SSc therapy is currently not well supported by the literature. Intensive immune ablation using high-dose chemotherapy, followed by autologous hematopoietic stem cell reconstitution therapy (HSCT), was associated with durable remission and improved long-term survival in multiple randomized clinical trials. Currently, HSCT is indicated for selected patients with severe SSc but carries the risk of treatment-related morbidity and mortality, as well as significant cost. Cellular therapies and CD19-targeting CAR-T immunotherapies for SSc are still in early stages of development but show promise for inducing sustained disease remission.

Antifibrotic Therapy Because tissue fibrosis underlies organ damage in SSc, drugs that interfere with the fibrotic process represent a rational therapeutic approach. In older retrospective studies, D-penicillamine was shown to stabilize skin induration, prevent new internal organ involvement, and improve survival. However, a randomized controlled clinical trial in early active SSc found no difference in the extent of skin involvement between patients treated with standard-dose (750 mg/d) or very-low-dose (125 mg every other day) D-penicillamine. Clinical trials show benefit of the tyrosine kinase inhibitor nintedanib, used alone or in combination with mycophenolate, in patients with SSc-ILD, with significant slowing of lung function decline, although no significant benefit on skin involvement.

Vascular Therapy The goal of Raynaud's therapy is to control episodes, prevent and enhance the healing of ischemic complications, and slow the progression of obliterative vasculopathy. Patients should dress warmly, minimize cold exposure, and avoid drugs that precipitate or exacerbate vasospastic episodes. Extended-release dihydropyridine calcium channel blockers such as amlodipine and diltiazem ameliorate Raynaud's phenomenon, but their use is often limited by side effects (palpitations, dependent edema, worsening gastroesophageal reflux). While ACE inhibitors do not reduce the frequency or severity of episodes, angiotensin II receptor blockers such as losartan are effective and well tolerated. Patients with Raynaud's phenomenon unresponsive to these therapies may require the addition of α_1-adrenergic receptor blockers (e.g., prazosin), 5-phosphodiesterase inhibitors (e.g., sildenafil), topical nitroglycerine, and intermittent IV infusions of prostaglandins. Low-dose aspirin and dipyridamole prevent platelet aggregation and may have a role as adjunctive agents. In patients with ischemic digital tip ulcerations, the endothelin-1 receptor antagonist bosentan reduces the risk of new ulcers. Digital sympathectomy and intradigital injections of botulinum type A (Botox) may be considered in patients with severe ongoing ischemia. Empirical long-term therapy with statins and antioxidants may retard the progression of vascular damage and obliteration. There is limited evidence-based information for the treatment of cardiac complications of SSc, which should be guided by specialists experienced in their diagnosis and management. While selective beta blockers such as metoprolol can precipitate vasospasm, nondihydropyridine calcium channel blockers can be used for rate control in atrial arrhythmias, and nonselective alpha/beta blockers such as carvedilol can be used for improving myocardial perfusion and left ventricular systolic function.

TREATMENT

TREATMENT OF SSc-ASSOCIATED INTERSTITIAL LUNG DISEASE

ILD remains a leading cause of death in patients with SSc. The radiographic extent of ILD on HRCT and the FVC at initial evaluation, along with decline in FVC during the preceding 12-month period, are helpful in identifying patients with likely progression. Additionally, male sex, older age at initial presentation, anti-topoisomerase I positivity, progressive skin involvement, and myocardial disease may be risk factors for ILD progression. Patients at high risk for ILD should, in addition to baseline screening HRCT, be monitored by performing PFTs every 3–6 months (Fig. 372-15); serial HRCT imaging is not recommended. Cyclophosphamide, given as monthly IV infusions or orally for 6–12 months, and mycophenolate mofetil both slow the decline in lung function and improve respiratory symptoms. As cyclophosphamide is associated with more frequent side effects, mycophenolate mofetil is generally preferred. In patients who show ILD progression despite immunotherapy, the antifibrotic drug nintedanib might be added. The optimal duration of nintedanib for SSc-associated ILD is currently under investigation. In patients who show continued progression of ILD despite medical therapy, lung transplantation might be considered as a life-prolonging procedure, although significant GERD contributing to organ rejection is a concern in SSc. Recurrence of SSc-ILD in transplanted lung allografts has not been reported.

TREATMENT OF GASTROINTESTINAL COMPLICATIONS

Because oral problems including decreased oral aperture, decreased saliva production, gum recession, periodontal disease, and teeth loss are common, regular dental care is recommended. Gastroesophageal reflux is very common in SSc and may be present even in the absence of GERD symptoms. Patients should be instructed to elevate the head of the bed, eat frequent small meals, and avoid alcohol, caffeine, and known reflux exacerbants or meals before bedtime. Proton pump inhibitors to reduce acid reflux may need to be given in relatively high doses. Prokinetic agents such as erythromycin (a motilin agonist), prucalopride, and domperidone may help in SSc, but are frequently associated with side effects. Botulinum toxin injection sometimes ameliorates impaired gastric emptying. Surgical antireflux procedures such as Nissen fundoplication can result in secondary achalasia and generally should be avoided. Episodic bleeding from gastric antral vascular ectasia (watermelon stomach) may be treated with endoscopic laser or argon plasma photocoagulation, although it commonly recurs. In some patients, enteral feeding and/or decompression via percutaneous gastrostomy or jejunostomy may become necessary. Small intestinal bacterial overgrowth secondary to gut dysmotility causes abdominal bloating and diarrhea and may lead to malabsorption and severe malnutrition. Short courses of rotating broad-spectrum antibiotics such as metronidazole, erythromycin, and rifaximin can eradicate bacterial overgrowth. Small-bowel hypomotility may respond to octreotide, but pseudo-obstruction is difficult to treat. Fecal incontinence, a frequent but underreported complication, may respond to antidiarrheal medication, biofeedback, sphincter augmentation, and sacral neuromodulation. Potential malnutrition should be routinely assessed.

TREATMENT OF PULMONARY ARTERIAL HYPERTENSION

In SSc, PAH carries an extremely poor prognosis and accounts for 30% of deaths. Because PAH is asymptomatic until advanced, patients with SSc should be screened at initial evaluation and yearly thereafter. Treatment is generally started with an oral endothelin-1 receptor antagonist such as bosentan or a phosphodiesterase-5 inhibitor such as sildenafil. Recently, the soluble guanylate cyclase stimulator riociguat, which acts by increasing the production of nitric oxide, and the selective IP prostacyclin receptor agonist selexipag were shown to improve PAH symptoms and survival. Patients may also require diuretics and digoxin. If hypoxemia is documented, supplemental oxygen should be prescribed in order to avoid secondary pulmonary vasoconstriction. Prostacyclin analogues such as epoprostenol or treprostinil can be given by continuous IV or SC infusion or via intermittent nebulized inhalations. Combination therapy with different classes of agents acting additively or synergistically is often necessary. Lung transplantation remains an option for selected SSc patients with PAH who fail medical therapy, and 2-year survival rates (64%) are comparable to those of idiopathic ILD or PAH.

TREATMENT OF RENAL CRISIS

Scleroderma renal crisis represents a medical emergency. Since its outcome is largely determined by the extent of renal damage at the time that aggressive therapy is initiated, prompt recognition of impending or early scleroderma renal crisis is essential, and efforts should be made to avoid its occurrence. High-risk SSc patients (early disease, extensive and progressive skin involvement, tendon friction rubs, and anti-RNA polymerase III antibodies) should be instructed to monitor their blood pressure daily and report significant alterations immediately. Potentially nephrotoxic drugs should be avoided, and glucocorticoids should be used only when absolutely necessary and at low doses. Patients presenting with scleroderma renal crisis should be immediately hospitalized. Once other causes of acute renal disease are excluded, treatment should be started promptly with titration of short-acting ACE inhibitors, with the goal of rapid normalization of the blood pressure. In patients with persistent hypertension, addition of angiotensin II receptor blockers, calcium channel blockers, endothelin-1 receptor blockers, prostacyclins, and direct renin inhibitors should be considered. In light of evidence for intrarenal complement pathway activation in some patients with scleroderma renal crisis, addition of eculizumab to ACE inhibitors may be considered. Up to two-thirds of patients with scleroderma renal crisis necessitate dialysis. Substantial renal recovery can occur following an episode of scleroderma renal crisis, and in 30–50% of patients, renal replacement therapy can be ultimately discontinued. Kidney transplantation is appropriate for patients unable to discontinue dialysis after 2 years. Survival of transplanted SSc patients is comparable to that of other diseases, and recurrence of renal crisis is rare.

SKIN CARE

Because involvement of the skin in SSc is never life-threatening and stabilizes and may even regress spontaneously, disease management should not be dictated by its cutaneous manifestations. The inflammatory symptoms of early skin involvement can be controlled with antihistamines and short-term use of low-dose glucocorticoids (<5 mg/d of prednisone). Cyclophosphamide and methotrexate have modest effects on skin induration. Because the skin is dry, the use of hydrophilic ointments and bath oils is encouraged, and regular skin massage is helpful. Telangiectasia, which present a cosmetic problem, especially on the face, can be treated with pulsed dye laser. Ischemic digital ulcerations should be protected by occlusive dressing to promote healing and prevent infection. Infected skin ulcers are treated with topical antibiotics and surgical debridement. While no therapy has been shown to be effective in preventing soft tissue calcific depositions or promoting their dissolution, reports support the use of minocycline, as well as bisphosphonates and topical, intralesional, or IV sodium thiosulfate (STS). Additional approaches include carbon dioxide laser treatment, extracorporeal shock-wave lithotripsy, and surgical high-speed microdrilling.

TREATMENT OF MUSCULOSKELETAL COMPLICATIONS

Arthralgia and joint stiffness are very common and distressing manifestations in early-stage disease. Short courses of nonsteroidal anti-inflammatory agents, methotrexate, and hydroxychloroquine and cautious use of low-dose glucocorticoids may alleviate symptoms. Physical and occupational therapy can be effective for preventing loss of musculoskeletal function and joint contractures and should be initiated early.

COURSE

The natural history of SSc is highly variable and difficult to predict, especially in early stages of the disease. Patients with dcSSc tend to have a more rapidly progressive course and worse prognosis than those with lcSSc. Inflammatory symptoms of early dcSSc, such as fatigue, edema, joint pain, and pruritus, subside, and skin thickening reaches a plateau at 2–4 years after disease onset. It is during the early edematous/inflammatory stage that life-threatening visceral organ involvement may develop. While existing visceral organ involvement, such as ILD, may progress even after skin involvement peaks, new organ involvement is rare. Scleroderma renal crisis generally occurs within the first 4 years of disease. In late-stage disease (>6 years), the skin is usually soft and atrophic. Skin regression characteristically occurs in an order that is the reverse of initial involvement, with softening on the trunks followed by proximal and finally distal extremities; however, sclerodactyly and fixed finger contractures generally persist. Relapse or recurrence of skin thickening after peak skin involvement has been reached is uncommon. Patients with lcSSc follow a clinical course that is markedly different than that of dcSSc. Raynaud's phenomenon typically precedes other disease manifestations by years or even decades, and visceral organ complications such as PAH generally develop late and progress slowly.

PROGNOSIS

SSc confers a substantial increase in the risk of premature death. Age- and gender-adjusted mortality rates are fivefold to eightfold higher compared to the general population, and more than half of all patients with SSc die from their disease. In one population-based study of SSc, the median survival was 11 years. In patients with dcSSc, 5- and 10-year survival rates are 70 and 55%, respectively, whereas in patients with lcSSc, 5- and 10-year survival rates are 90 and 75%, respectively. The prognosis correlates with the extent of skin involvement, which itself is a surrogate for visceral organ involvement. Major causes of death are PAH, pulmonary fibrosis, GI involvement, and cardiac disease. Scleroderma renal crisis is associated with a 30% 3-year mortality. Lung cancer and excess cardiovascular deaths also contribute to increased mortality. Markers of poor prognosis include male gender, African ancestry, older age at disease onset, extensive skin thickening with truncal involvement, palpable tendon friction rubs, and evidence of significant or progressive visceral organ involvement. Laboratory predictors of increased mortality at initial evaluation include an elevated ESR, anemia, proteinuria, and anti–topoisomerase I antibodies. In one study, SSc patients with extensive skin involvement, lung vital capacity <55% predicted, significant GI involvement (pseudo-obstruction or malabsorption), clinical cardiac involvement, or scleroderma renal crisis had a 9-year survival <40%. The severity of PAH predicts mortality, and patients with mean pulmonary arterial pressure ≥45 mmHg on cardiac catheterization had a 33% 3-year survival. The advent of ACE inhibitors for scleroderma renal crisis had a dramatic impact on survival, increasing from <10% survival at 1 year in the pre–ACE inhibitor era to >70% 3-year survival at the present time. Moreover, 10-year survival in SSc has improved from <60% in the 1970s to >66–78% in the 1990s, a trend that reflects both earlier detection and better management of complications.

LOCALIZED SCLERODERMA

The term *scleroderma* describes a group of localized skin disorders (Table 372-1). These occur more commonly in children than in adults and, in marked contrast to SSc, are generally not complicated by Raynaud's phenomenon or significant internal organ involvement. Morphea presents as solitary or multiple circular patches of thick skin or, rarely, as widespread induration (generalized or pansclerotic morphea); the fingers are generally spared. Linear scleroderma may affect subcutaneous tissues, leading to fibrosis and atrophy of supporting structures, tendons, muscle, and even bone. In children, the growth of affected long bones can be slowed. When linear scleroderma crosses large joints, significant contractures can develop.

MIXED CONNECTIVE TISSUE DISEASE

Patients who have lcSSc coexisting with features of SLE, polymyositis, and rheumatoid arthritis may have mixed connective tissue disease (MCTD), generally associated with high titers of autoantibodies to U1-RNP. The characteristic initial presentation is Raynaud's phenomenon associated with puffy fingers and myalgia. Over time, sclerodactyly, soft tissue calcinosis, and cutaneous telangiectasia may appear. Skin rash suggestive of SLE (malar erythema, photosensitivity) or dermatomyositis (heliotrope rash on the eyelids, erythematous rash on knuckles) may occur. Arthralgia is common, and some patients develop erosive polyarthritis. Pulmonary fibrosis and isolated or secondary PAH may develop. Other manifestations include esophageal dysmotility, pericarditis, Sjögren's syndrome, and renal disease, especially membranous glomerulonephritis. Laboratory evaluation shows elevated ESR and hypergammaglobulinemia. In contrast to SSc, MCTD often responds to glucocorticoids, and the long-term prognosis is better than that of SSc. Whether MCTD is truly a distinct entity or is a subset of SLE or SSc remains controversial.

EOSINOPHILIC FASCIITIS (DIFFUSE FASCIITIS WITH EOSINOPHILIA)

Eosinophilic fasciitis is a rare idiopathic disorder of adults associated with abrupt skin induration. The skin characteristically shows a coarse cobblestone "peau d'orange" appearance. In contrast to SSc, Raynaud's phenomenon and SSc-associated internal organ involvement and autoantibodies are absent. Furthermore, skin involvement spares the fingers. Full-thickness biopsy of the lesional skin reveals fibrosis of the subcutaneous fascia, with variable inflammation and eosinophil infiltration. In the acute phase of the illness, peripheral blood eosinophilia may be prominent. MRI appears to be a sensitive tool for the diagnosis of eosinophilic fasciitis. Eosinophilic fasciitis can occur in association with, or preceding, various myelodysplastic syndromes or multiple myeloma. Although glucocorticoids cause prompt resolution of eosinophilia, the skin shows slow and variable improvement. The prognosis of patients with eosinophilic fasciitis is generally good.

FURTHER READING

Allanore Y et al: Systemic sclerosis. Nat Rev Dis Primers 1:15002, 2015.

Herzog EL et al: Interstitial lung disease associated with systemic sclerosis and idiopathic pulmonary fibrosis: How similar or distinct? Arthritis Rheum 66:1967, 2014.

Joseph CG et al: Association of the autoimmune disease scleroderma with an immunologic response to cancer. Science 343:152, 2014.

Martyanov V, Whitfield ML: Molecular stratification and precision medicine in systemic sclerosis from genomic and proteomic data. Curr Opin Rheumatol 28:83, 2016.

Tashkin DP et al: Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): A randomised controlled, double-blind, parallel group trial. Lancet Respir Med 4:708, 2016.

373

Sjögren's Disease

Haralampos M. Moutsopoulos,
Clio P. Mavragani

DEFINITION, INCIDENCE, AND PREVALENCE

Sjögren's disease is a prototype autoimmune disease characterized by lymphocytic infiltration of the exocrine glands resulting in xerostomia, dry eyes (keratoconjunctivitis sicca), and profound B-cell hyperactivity. The disease has unique features since it presents with a wide clinical spectrum from organ-specific to systemic disease; can occur alone or in association with other systemic rheumatic diseases, more commonly rheumatoid arthritis, limited scleroderma, and systemic lupus erythematosus; and displays high probability of the development of lymphoma. Because of all these characteristics, it is an ideal model to study not only autoimmunity but also lymphoid malignancy.

Middle-aged women (female-to-male ratio 10–20:1) are primarily affected, although Sjögren's disease may occur at any age, including childhood. Patients with earlier disease onset express a more aggressive disease phenotype manifested by a high occurrence of systemic manifestations and serum autoantibodies. The prevalence of Sjögren's disease is ~0.5–1%, while 5–20% of patients with other autoimmune diseases can express sicca manifestations.

PATHOGENESIS

The autoimmune phenomena observed in Sjögren's disease include lymphocytic infiltration of the exocrine glands (primarily salivary and lachrymal glands) and B lymphocyte hyperreactivity. The latter is mainly manifested by hypergammaglobulinemia and the presence of serum autoantibodies toward non-organ-specific antigens such as immunoglobulins (rheumatoid factors) and extractable cellular antigens (Ro52, Ro60, and La). The major infiltrating cells in the affected exocrine glands are activated T lymphocytes. In labial minor salivary gland tissues with extensive lymphocytic infiltrations, B-cell populations prevail. Other cellular subsets detected in the labial minor salivary gland histopathologic lesion of Sjogren's disease include follicular, myeloid, and plasmacytoid dendritic cells, as well as macrophages. The latter along with inflammasome activation have been shown to be associated with increased risk for lymphoma development.

The interplay of endogenous (e.g., intracellular stress, inappropriate overexpression of endogenous nucleic acids) and exogenous triggers (e.g., viruses, hormonal triggers, stressful life events) in a background of a genetically determined hyperactive immune response seems to be crucial for the initiation and perpetuation of the disease. Ductal and acinar epithelial cells appear to play a significant role in the initiation and perpetuation of autoimmune injury. These cells (1) express inappropriately costimulatory molecules and the intracellular autoantigens Ro and La on their cell surfaces, acquiring the capacity to provide signals essential for lymphocytic activation; (2) produce proinflammatory cytokines and lymphocyte-attracting chemokines necessary for sustaining the autoimmune lesion and allowing the formation of ectopic germinal centers; (3) express functional receptors of innate immunity, particularly Toll-like receptors (TLRs) 3, 7, and 9 molecules, which may account for the initiation of the autoimmune reactivity; and (4) display immunoregulatory molecules such as ICAM and CD40. Glandular epithelial cells also seem to have an active role in the production of B cell–activating factor (BAFF), which is induced after stimulation with type I and II interferons. Circulating BAFF has been found to be elevated also in the serum of Sjögren's disease patients, especially those with hypergammaglobulinemia and serum autoantibodies, and probably accounts for the antiapoptotic effect on B lymphocytes.

In contrast to B and T lymphocytes, glandular epithelial cells display increased rates of apoptotic death. Established risk genetic loci implicated in Sjögren's disease include the human leukocyte antigen DQA1*0501 allele, variations involved in the interferon/BAFF axis (*IRF5, STAT4, BAFF*) and B-cell function (*EBF1, BLK*), as well as combined genetic deficiencies in the classical complement pathway.

CLINICAL MANIFESTATIONS

Most patients with Sjögren's disease have symptoms related to impaired exocrine gland, particularly lacrimal and salivary gland, function. The disease evolution is slow and, in the majority of patients, runs a benign course. Studies have shown that prior to disease onset, patients with Sjögren's disease experience major stressful life events with which they cannot cope adequately.

The principal oral symptom of Sjögren's disease is dryness (xerostomia). Patients report difficulty in swallowing dry food, a burning mouth sensation, an increase in dental caries, and problems in wearing complete dentures. Physical examination shows a dry, erythematous sticky oral mucosa. In patients with severe oral dryness, the filiform papillae on the tongue dorsum are atrophic, and the tongue is deeply fissured **(Fig. 373-1)**. Furthermore, saliva from the major glands is either not expressible or cloudy. Intermittent or persistent enlargement of the parotid or other major salivary glands occurs in two-thirds of patients with Sjögren's disease. Diagnostic tests include sialometry and several imaging techniques, including ultrasound, magnetic resonance

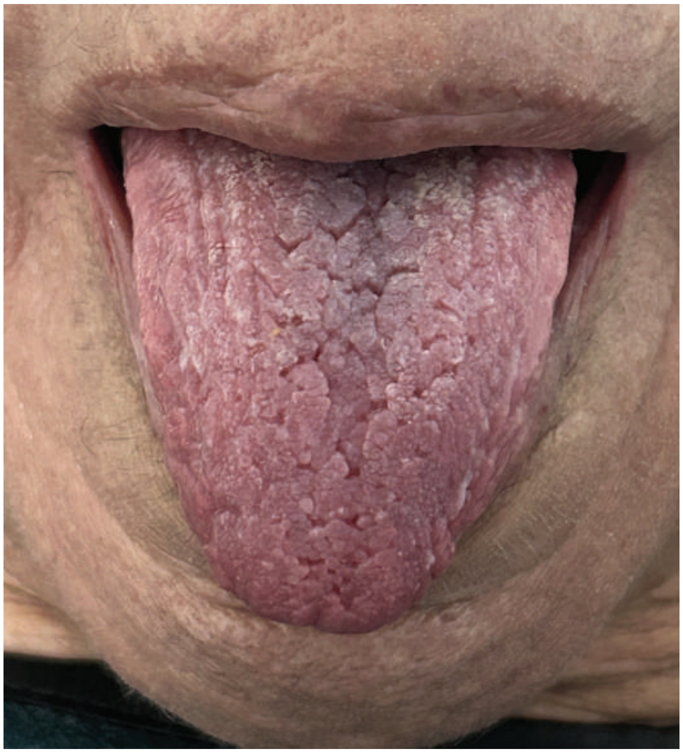

FIGURE 373-1 A dry deeply fissured tongue from a patient with Sjögren's disease.

TABLE 373-1 Prevalence of Extraglandular Manifestations in Primary Sjögren's disease

CLINICAL MANIFESTATION	PERCENT	REMARKS
Nonspecific		
Fatigability/myalgias	25	Fibromyalgia
Arthralgias/arthritis	60	Usually nonerosive, leading to Jaccoud's arthropathy
Raynaud's phenomenon	37	In one-third of patients, precedes sicca manifestations
Periepithelial		
Lung involvement	14	Small airway disease/lymphocyte interstitial pneumonitis
Kidney involvement	9	Interstitial kidney disease is usually asymptomatic
Liver involvement	6	Primary biliary cirrhosis stage I
Immune complex–mediated		
Small vessel vasculitis	9	Purpura, urticarial lesions
Peripheral neuropathy	2	Polyneuropathy, either sensory or sensorimotor
Glomerulonephritis	2	Membranoproliferative
Lymphoma		
Lymphoma	6	Glandular MALT[a] lymphoma is most common

[a] Mucosa-associated lymphoid tissue.

imaging (MRI), and magnetic resonance sialography of the major salivary glands. In particular, salivary gland ultrasound is an emerging tool of both diagnostic and prognostic utility. Biopsy of the labial minor salivary gland allows histopathologic confirmation of focal lymphocytic infiltrates.

Ocular involvement is the other major manifestation of Sjögren's disease. Patients usually report a sandy or gritty feeling under the eyelids. Other ocular symptoms include burning, accumulation of secretions in thick strands at the inner canthi, decreased tearing, redness, itching, eye fatigue, and increased photosensitivity. These symptoms are attributed to the destruction of corneal and bulbar conjunctival epithelium, a pathology termed *keratoconjunctivitis sicca*. Diagnostic evaluation of keratoconjunctivitis sicca includes measurement of tear flow by Schirmer's I test, determination of tear composition by tear breakup time or tear lysozyme content, and slit-lamp examination of the cornea and conjunctiva after lissamine green or Rose Bengal staining that reveals punctuate corneal and bulbar conjunctival ulcerations and attached filaments.

Involvement of other exocrine glands, which occurs less frequently, includes a decrease in mucous gland secretions of the upper and lower respiratory tree, resulting in dry nose, throat, and trachea (xerotrachea). In addition, diminished secretion of the exocrine glands of the gastrointestinal tract leads to esophageal mucosal dysmotility and atrophic gastritis. Dyspareunia, in premenopausal women, due to dryness of the external genitalia and dry skin also may occur.

Extraglandular (systemic) manifestations are seen in one-third of patients with Sjögren's disease **(Table 373-1)** and can be classified as follows: *nonspecific, periepithelial* (surrounding of epithelial tissues by lymphocytes), *immune complex–mediated*, and *lymphoma*. *Nonspecific* manifestations include fatigability, low-grade fever, Raynaud's phenomenon, myalgias, arthralgias, and arthritis. Arthritis in patients with primary Sjögren's disease is nonerosive. *Periepithelial* pathology due to periepithelial accumulation of lymphocytes results from the involvement of parenchymal organs such as the lungs, kidneys, and liver. On the basis of this observation, one of the authors (H.M.M.) has coined the term *autoimmune epithelitis*. Lung involvement is usually manifested with dry cough and rarely with dyspnea. The underlying lung pathology includes peribronchial infiltrates (bronchitis sicca) and interstitial pneumonitis. Renal involvement includes interstitial nephritis, clinically manifested by hyposthenuria and renal tubular dysfunction with or without acidosis. Untreated acidosis may lead to nephrocalcinosis. *Immune complex–mediated* disease is expressed with vasculitis affecting primarily small-sized vessels, mainly manifested with purpura and rarely with urticarial rash, skin ulcerations, mononeuritis multiplex, and membranoproliferative glomerulonephritis associated with mixed type II or III cryoglobulinemia. Central nervous system involvement is rarely recognized. A few cases of myelitis associated with antibodies to aquaporin 4 have been described.

Sjögren's disease is characterized by the highest risk for lymphoma development among all autoimmune diseases with clinical, laboratory, and histopathologic features being designated as risk factors **(Fig. 373-2)**. Despite that the pathogenesis of lymphoma in the setting of Sjögren's disease remains to be elucidated, genetic alterations involved in chronic inflammatory, B-cell activation, and the type I and II interferon pathways, as well as epigenetic abnormalities, have been shown to be significant contributors. Of interest, B cells producing cryoglobulins were shown to harbor mutations in genes recurrently mutated in B-cell malignancies. Most lymphomas are extranodal, low-grade, marginal zone B-cell lymphomas and are usually detected incidentally during evaluation of the labial minor salivary gland biopsy. The affected lymph nodes are usually peripheral. Survival rates are decreased in patients with B symptoms, lymph node mass >7 cm in diameter, and high or intermediate histologic grade. Fluorodeoxyglucose positron emission tomography (FDG-PET)/computed tomography (CT) can be a useful tool to exclude lymphoma when suspected. In line with observations in rheumatoid arthritis and systemic lupus erythematosus, patients with Sjögren's disease also display an increased risk of cardiovascular disease.

Routine laboratory tests in Sjögren's disease can reveal leukopenia and infrequently lymphopenia. In two-thirds of patients, elevated erythrocyte sedimentation rate, hypergammaglobulinemia, antinuclear antibodies, rheumatoid factors, and antibodies against Ro52/Ro60 and La autoantigens are detected. Anticentromere autoantibodies are present in Sjögren's patients with a clinical picture similar to that of limited scleroderma **(Chap. 372)**, while the presence of antimitochondrial antibodies may connote liver involvement in the form of autoimmune cholangitis **(Chap. 357)**. Autoantibodies to 21-hydroxylase are found in patients with a blunted adrenal response, while autoantibodies to citrullinated peptides are seen in Sjögren's patients with

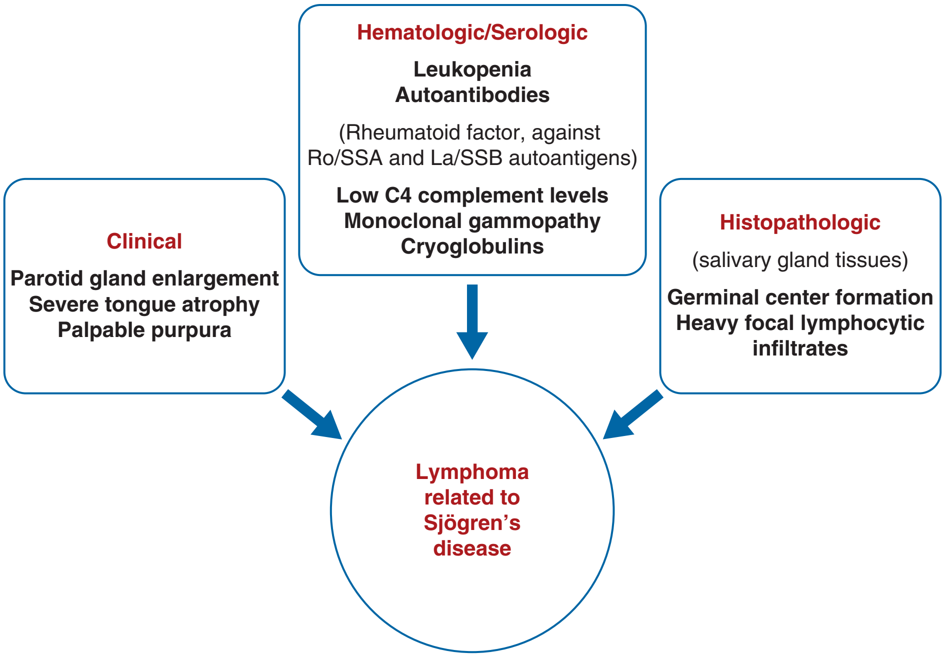

FIGURE 373-2 Sjögren's disease risk factors for lymphoma development.

arthritis. Anticalponin-3 antibodies have been recently associated with the occurrence of peripheral neuropathies.

DIAGNOSIS AND DIFFERENTIAL DIAGNOSIS

Sjögren's disease should be suspected if a patient presents with eye and/or mouth dryness, major salivary gland enlargement, or systemic manifestations such as Raynaud's phenomenon, palpable purpura, or symptomatology of renal tubular acidosis. A careful history of medications causing dryness should be obtained. Recently, cases of Sjögren's disease were triggered by PD-1/PD-L1 checkpoint inhibitors.

The workup should include eye tests that might reveal keratoconjunctivitis sicca, salivary flow tests or ultrasonography, and serum evaluation for specific autoantibodies. Testing for chronic viral infections (hepatitis C virus, HIV), chest x-ray to rule out sarcoidosis, protein electrophoresis, IgG4 serum levels, and autoantibodies to thyroid antigens can be also offered. Labial biopsy is valuable to rule out conditions that may cause dry mouth, dry eyes, or parotid gland enlargement **(Tables 373-2** and **373-3)**. A diagnostic algorithm based on recent classification criteria (2016 American College of Rheumatology–European League Against Rheumatism [EULAR] Classification Criteria) is presented in **Fig. 373-3**.

TABLE 373-2 Differential Diagnosis of Sicca Symptoms

XEROSTOMIA	DRY EYE	BILATERAL PAROTID GLAND ENLARGEMENT
Viral infections (HCV, HIV) Drugs Psychotherapeutic Parasympatholytic Antihypertensive Psychogenic origin Irradiation Diabetes mellitus Trauma Sjögren's disease Amyloidosis Autoimmune thyroid disease	Inflammation Stevens-Johnson syndrome Pemphigoid Chronic conjunctivitis Chronic blepharitis Sjögren's Disease Toxicity Burns Drugs Neurologic conditions Impaired lacrimal gland function Impaired eyelid function Miscellaneous Trauma Hypovitaminosis A Blink abnormality Anesthetic cornea Lid scarring Epithelial irregularity Autoimmune thyroid disease	Viral infections Mumps Influenza Epstein-Barr virus Coxsackievirus A Cytomegalovirus HIV, HCV Sarcoidosis, tuberculosis IgG4-related disease Sjögren's disease Metabolic disorders Diabetes mellitus Hyperlipoproteinemias (types IV and V) Chronic pancreatitis Hepatic cirrhosis Endocrine Acromegaly Gonadal hypofunction Eosinophilic sialodochitis Lymphoma

Abbreviation: HCV, hepatitis C virus.

TREATMENT

Sjögren's Disease

Treatment of Sjögren's disease aims to relieve symptoms and limit the damage from chronic xerostomia and keratoconjunctivitis sicca through substitution or stimulation of impaired secretions.

To replace deficient tears, several ophthalmic preparations are readily available (hydroxypropyl methylcellulose; polyvinyl alcohol; 0.5% methylcellulose; Hypo Tears). If corneal ulcerations are present, eye patching and boric acid ointments are recommended, as well as cyclosporine eye drops. Certain drugs that may decrease lacrimal and salivary secretions, such as diuretics, antihypertensive drugs, anticholinergics, and antidepressants, should be avoided.

For xerostomia, the best replacement is water. Propionic acid gels may be used to treat vaginal dryness. To stimulate secretions, orally administered pilocarpine (5 mg thrice daily) or cevimeline (30 mg

TABLE 373-3 Differential Diagnosis of Sjögren's disease

HIV INFECTION AND SICCA SYNDROME	SJÖGREN'S DISEASE	SARCOIDOSIS
Predominant in young males	Predominant in middle-aged women	No age or sex preference
Lack of autoantibodies to Ro and/or La	Presence of autoantibodies	Lack of autoantibodies to Ro and/or La
Lymphoid infiltrates of salivary glands by CD8+ T lymphocytes	Lymphoid infiltrates of salivary glands by CD4+ T lymphocytes	Granulomas in salivary glands
Association with HLA-DR5	Association with HLA-DR3 and DRw52	Unknown
Positive serologic tests for HIV	Negative serologic tests for HIV	Negative serologic tests for HIV

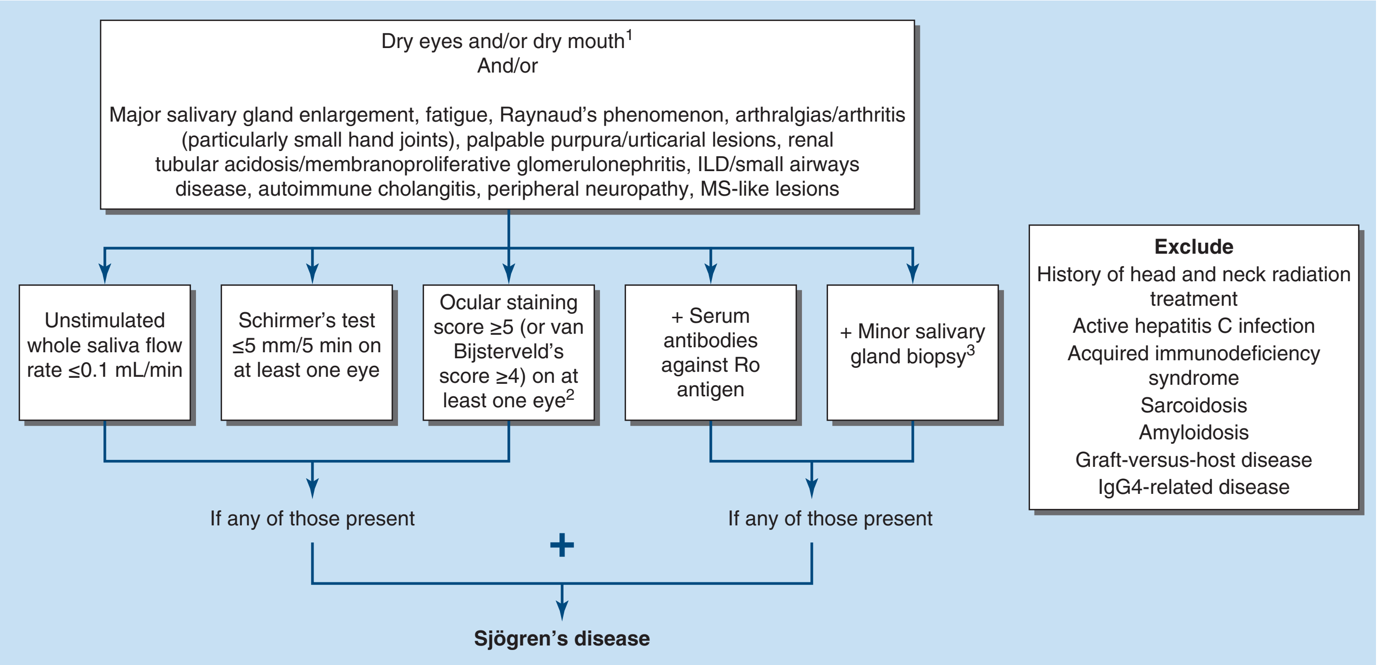
Dry eyes and/or dry mouth[1]
And/or
Major salivary gland enlargement, fatigue, Raynaud's phenomenon, arthralgias/arthritis (particularly small hand joints), palpable purpura/urticarial lesions, renal tubular acidosis/membranoproliferative glomerulonephritis, ILD/small airways disease, autoimmune cholangitis, peripheral neuropathy, MS-like lesions
Unstimulated whole saliva flow rate ≤0.1 mL/min
Schirmer's test ≤5 mm/5 min on at least one eye
Ocular staining score ≥5 (or van Bijsterveld's score ≥4) on at least one eye[2]
+ Serum antibodies against Ro antigen
+ Minor salivary gland biopsy[3]
Exclude
History of head and neck radiation treatment
Active hepatitis C infection
Acquired immunodeficiency syndrome
Sarcoidosis
Amyloidosis
Graft-versus-host disease
IgG4-related disease
If any of those present
+
If any of those present
Sjögren's disease

FIGURE 373-3 Diagnostic algorithm for Sjögren's disease. [1] Defined as a positive response to at least one of the following questions: (a) Have you had daily, persistent, troublesome dry eyes for more than 3 months? (b) Do you have a recurrent sensation of sand or gravel in the eyes? (c) Do you use tear substitutes more than three times a day? (d) Have you had a daily feeling of dry mouth for more than 3 months? (e) Do you frequently drink liquids to aid in swallowing dry food? [2] Ocular staining score described in Whitcher et al. van Bijsterveld score described in van Bijsterveld et al. [3] Focus score count ≥1 (based on the number of foci per 4 mm^2 of salivary gland tissue) following a protocol described in Daniels et al. ILD, interstitial lung disease; MS, multiple sclerosis.

thrice daily) appears to improve sicca manifestations, and both are well tolerated. Hydroxychloroquine (200 to 400 mg daily) and/or methotrexate (0.2–0.3 mg/kg body weight/weekly) are helpful for arthralgias and mild arthritis.

Patients with renal tubular acidosis should receive sodium bicarbonate by mouth (0.5–2 mmol/kg in four divided doses). Glucocorticoids and monoclonal antibody to CD20 (rituximab) appear to be effective in patients with systemic disease, particularly in those with purpura and persistent arthritis. Novel monoclonal antibodies targeting the CD40L/CD40 costimulatory pathway or the BAFF receptor or sequential treatment of belimumab and rituximab seem to be promising therapeutic strategies. Data for a beneficial role of other conventional immunosuppressants are limited. Treatment of lymphoma in the setting of Sjögren's disease follows the general guidelines for lymphoma management in the general population. Detailed recommendations on Sjögren's disease management have been issued by EULAR and British Society of Rheumatology.

■ FURTHER READING

Daniels TE et al: Associations between salivary gland histopathologic diagnoses and phenotypic features of Sjögren's syndrome among 1,726 registry participants. Arthritis Rheum 63:2021, 2011.

Fragkioudaki S et al: Predicting the risk for lymphoma development in Sjogren syndrome: An easy tool for clinical use. Medicine (Baltimore) 95:e3766, 2016.

Mavragani CP, Moutsopoulos HM: Sjögren's syndrome. CMAJ 186:579, 2014.

Mavragani CP, Moutsopoulos HM: Sjogren's syndrome: Old and new therapeutic targets. J Autoimmun 110:102364, 2020.

Moutsopoulos HM: Sjögren's syndrome: A forty-year scientific journey. J Autoimmun 51:1, 2014.

Nocturne G, Mariette X: Expert perspective: Challenges in Sjögren's disease. Arthritis Rheumatol 75:2078, 2023.

Price EJ et al: British Society for Rheumatology guideline on management of adult and juvenile onset Sjögren disease. Rheumatology (Oxford)16:keae152, 2024.

Shiboski CH et al: 2016 American College of Rheumatology/European League Against Rheumatism Classification Criteria for Primary Sjögren's Syndrome: A consensus and data-driven methodology involving three international patient cohorts. Arthritis Rheumatol 69:35, 2017.

van Bijsterveld OP: Diagnostic tests in the Sicca syndrome. Arch Ophthalmol 82:10, 1969.

Whitcher JP et al: A simplified quantitative method for assessing keratoconjunctivitis sicca from the Sjögren's Syndrome International Registry. Am J Ophthalmol 149:405, 2010.

374 Spondyloarthritis

Atul Deodhar, Dirk Elewaut

Spondyloarthritis (SpA) refers to a family of immune-mediated inflammatory arthritis disorders that share many clinical, genetic, and pathologic characteristics. Moll and Wright are credited in recognizing these diseases in 1974 as distinct from rheumatoid arthritis (RA), a well-recognized inflammatory arthritis at that time. While axial skeletal involvement is hinted at in the word "spondyloarthritis," it is not mandatory. The musculoskeletal manifestations of SpA include sacroiliitis, inflammatory spinal lesions, peripheral inflammatory arthritis, enthesitis (inflammation at the attachment of tendons and ligaments to the bones), tendonitis, tenosynovitis, and dactylitis ("sausage digits"). Other clinical manifestations include uveitis, skin psoriasis, inflammatory bowel disease (IBD, Crohn's, or ulcerative colitis), absence of rheumatoid factor (RF) and nodules (to differentiate from RA), familial correlation, and association with the gene human leukocyte antigen *(HLA)-B27*. The 2009 Assessment of Spondyloarthritis International Society (ASAS) classification criteria divided SpA under "axial" (axSpA) and "peripheral" (pSpA). "Peripheral" SpA may seem a misnomer, but it simply indicates that peripheral musculoskeletal

TABLE 374-1 Spectrum of Spondyloarthritis

SPONDYLOARTHRITIS (SPA)			
AXIAL SPONDYLOARTHRITIS (axSpA)		PERIPHERAL SPONDYLOARTHRITIS (pSpA)	
CONDITION	CLINICAL FEATURES	CONDITION	CLINICAL FEATURES
		Conditions commonly included under pSpA	
Radiographic axSpA (r-axSpA, also called ankylosing spondylitis [AS])	Sacroiliitis grade 2 (bilateral), grade 3 or 4 (unilateral or bilateral)	Psoriatic arthritis	Skin psoriasis and nail involvement usually preceding arthritis, enthesitis, dactylitis
Nonradiographic axSpA (nr-axSpA)	Sacroiliitis grade 1 (unilateral or bilateral) or grade 2 (unilateral)	Inflammatory bowel disease–associated arthritis	Crohn's disease or ulcerative colitis with inflammatory arthritis
		Reactive arthritis	Urethritis, history of preceding infection with *Salmonella, Shigella, Yersinia, Campylobacter,* and *Chlamydia,* mucosal ulcers, conjunctivitis, keratoderma blennorrhagica
		Undifferentiated peripheral SpA	Enthesitis, dactylitis, family history of SpA, *HLA-B27,* acute anterior uveitis, and not fitting in any of the other conditions mentioned above
		Conditions sometimes included under pSpA	
		SAPHO syndrome	Synovitis, acne, pustulosis, hyperostosis, and osteitis; typical involvement of anterior chest wall joints
		Acne-associated arthritis	Acne conglobata, peripheral arthritis
		Hidradenitis suppurativa–associated arthritis	Hidradenitis suppurativa, peripheral arthritis, rarely asymmetric sacroiliitis

manifestations are predominant and more troublesome to the patient than the spinal involvement.

AxSpA is further divided into radiographic (r-axSpA, also known as ankylosing spondylitis [AS]), and nonradiographic (nr-axSpA). The conditions included under pSpA are psoriatic arthritis (PsA), reactive arthritis (ReA), arthritis associated with IBD, and undifferentiated SpA, where patients cannot be classified under any known categories. Patients with axSpA can have peripheral musculoskeletal manifestations, and patients with pSpA may have additional axial skeletal involvement at presentation or may develop it later in the course of their disease. Other conditions sometimes included under pSpA are SAPHO (synovitis, acne, pustulosis, hyperostosis, and osteitis) syndrome, acne-associated arthritis, and arthritis associated with hidradenitis suppurativa. **Table 374-1** shows the full spectrum of conditions included under "spondyloarthritis."

Pattern recognition is important for the diagnosis of SpA. Typical patterns include asymmetric, mono- or oligoarticular (four or less joints) inflammatory arthritis of large joints in the lower limb (usually knee), with enthesitis, tendonitis, and dactylitis. This contrasts with RA, where polyarticular symmetric involvement of metacarpophalangeal (MCP)/proximal interphalangeal (PIP) joints predominates. The cumulative prevalence of all conditions under SpA is 2–3% of the population, much more common than RA.

PATHOLOGY OF SPONDYLOARTHRITIS

SpA is a prototypic chronic immune-mediated inflammatory disease that couples inflammation to structural damage. A unique feature of SpA is that inflammation-induced bone loss coexists with pathologic new bone formation, which occurs at specific sites in the skeleton. These sites include not only the sacroiliac joint but also the anterior spinal ligaments or at entheses in the peripheral skeleton. If uncontrolled, axSpA may ultimately lead to a complete ankylosis of the sacroiliac joints and the spine, termed "bamboo spine." With increasing awareness of SpA resulting from early diagnosis and treatment, there has been a steady decrease in the amount of structural damage observed over the past decades with a gradual shift of higher prevalence from r-AxSpA to nr-AxSpA.

While histopathologic studies have led to an increased understanding of local inflammatory and structural changes, the accessibility to target tissue particularly in the axial skeleton and tendons/entheses has been hampered by technical issues. Most of the information to date comes from surgically derived specimens from patients with longstanding disease. Accordingly, many of the described features represent disease under treatment, which inevitably impacts histopathology. Obtaining synovial tissue from an array of joints such as knees, wrists, and small joints in fingers has become much more readily accessible through ultrasound-guided synovial tissue biopsy, which permits study of synovial tissue alterations in an early treatment-naïve stage.

PATHOGENESIS OF SPONDYLOARTHRITIS

Although the exact mechanisms that initiate the disease are not fully delineated, genetic and environmental risk factors and molecular and cellular alterations at key disease sites have been implicated in the pathophysiology of SpA **(Fig. 374-1)**. At these sites, which include the axial and peripheral joints, entheses, skin, and gut, variation is seen in the disease mechanisms and perturbations of the key cytokine pathways. These tissue-based variations are most likely associated with the dissimilar clinical responses observed between the spine, peripheral joints, and gut during treatment with cytokine inhibitors.

■ GENETIC ASSOCIATIONS IN AXIAL SPONDYLOARTHRITIS AND OTHER SPONDYLOARTHRITIS: *HLA-B27* AND BEYOND

HLA-B27 is a major histocompatibility complex (MHC) class I allele that forms the strongest genetic risk factor for SpA and specifically axSpA. MHC class I molecules are expressed on the surface of nucleate cells and can present peptides to CD8+ T cells. *HLA-B27* is present in 85–90% of the patients with r-axSpA and 50–90% in nr-axSpA, in contrast to about 5% of the healthy Caucasian population.

Besides *HLA-B27*, other MHC class I genes and non-MHC coding genes have been identified in large genome-wide association studies (GWAS) in axSpA. Within the group of non-MHC coding genes, genes linked to innate immune processes, interleukin (IL)-23/IL-17 immunity, joint and bone remodeling, epithelial function, and antigen presentation are seen, which are processes relevant in the immunopathogenesis of axSpA. Whether all of these represent independent risk factors is still not clear, although some appear to be interconnected. For example, genetic epistatic interaction has been suggested between ERAP1, a peptide-trimming enzyme, of which single nucleotide polymorphisms (SNPs) are associated with AS and *HLA-B27*. How *HLA-B27* fosters development of SpA remains unclear despite the discovery of this strong genetic association 50 years ago. At present, several theories exist with limitations to each of them.

The "arthritogenic peptide" theory proposes that *HLA-B27* presents pathological microbial peptides to CD8+ cytotoxic T cells that

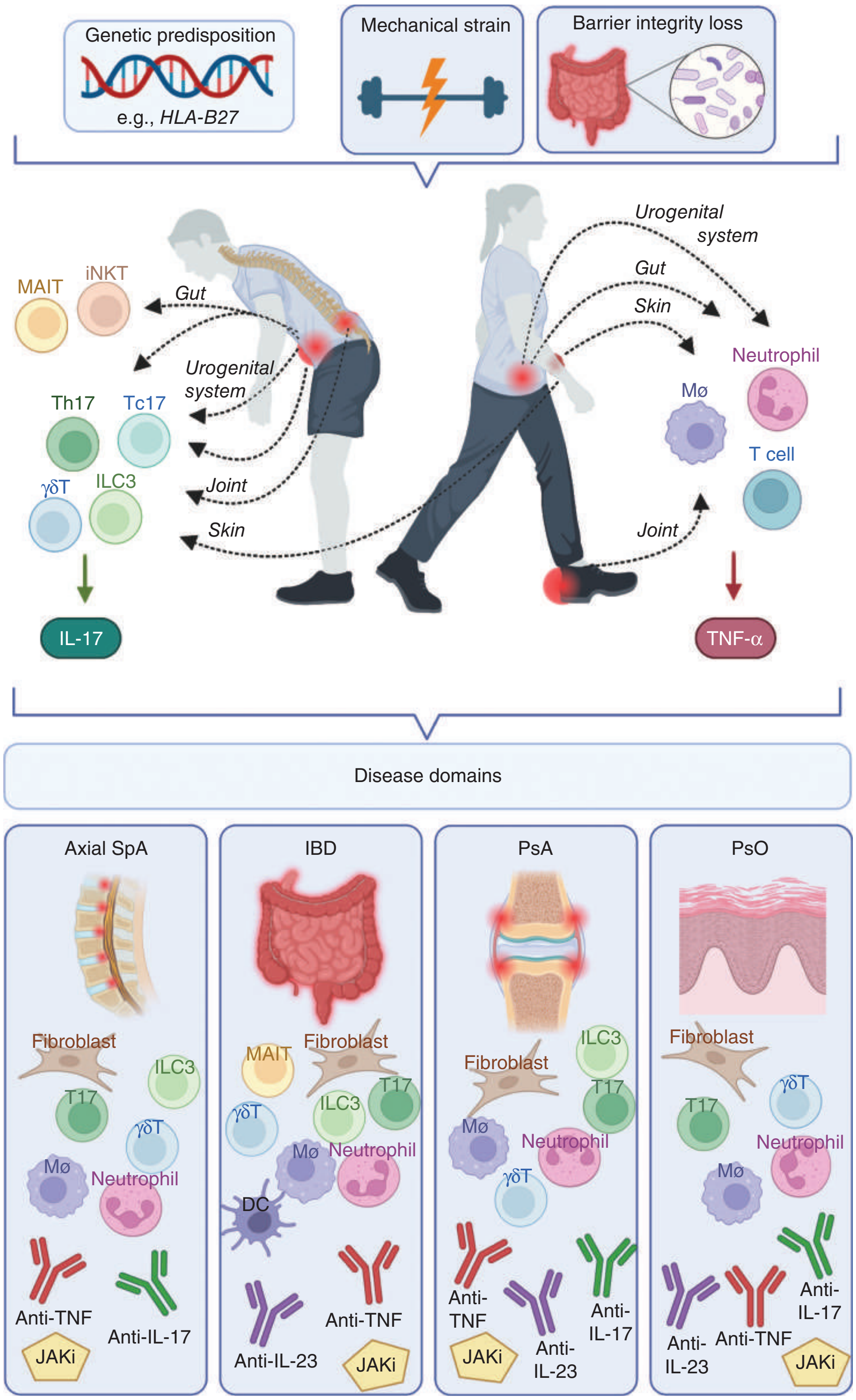
Genetic predisposition
e.g., *HLA-B27*
Mechanical strain
Barrier integrity loss
MAIT
iNKT
Gut
Urogenital system
Th17
Tc17
γδT
ILC3
Joint
Skin
IL-17
Urogenital system
Gut
Skin
Neutrophil
Mø
T cell
Joint
TNF-α
Disease domains
Axial SpA
Fibroblast
ILC3
T17
γδT
Mø
Neutrophil
Anti-TNF
Anti-IL-17
JAKi
IBD
MAIT
Fibroblast
γδT
ILC3
T17
Mø
Neutrophil
DC
Anti-TNF
Anti-IL-23
JAKi
PsA
Fibroblast
ILC3
T17
Mø
Neutrophil
γδT
Anti-TNF
Anti-IL-23
Anti-IL-17
JAKi
PsO
Fibroblast
γδT
T17
Neutrophil
Mø
Anti-IL-17
Anti-IL-23
Anti-TNF
JAKi

FIGURE 374-1 Pathogenesis and signature cytokines across different disease domains in spondyloarthritis (SpA). In genetically predisposed individuals, SpA may arise from mechanical stress, a well-known disease precipitator in psoriatic arthritis and axial SpA, or by altered host-microbial interplay at inner (gut, urogenital tract) or outer (skin) surfaces culminating in a cytokine crosstalk characterized by both tumor necrosis factor (TNF)- and interleukin (IL)-17-driven immune responses. Drivers of disease include cells of both innate and adaptive immune system, along with tissue-resident stromal cells. Different disease domains (gut, spin, peripheral joints, and skin) harbor overlapping (e.g., TNF-α) but also distinct (e.g., IL-17, IL-23) signature cytokines, which has major therapeutic implications. Janus kinase (JAK) inhibitors, which interfere with signaling of several cytokines, are effective across all domains.

then target cross-reactive human peptides leading to inflammation and resulting in molecular mimicry. Presumably, this first antigenic exposure could take place in the gut or at the urogenital tract. Several allotypes of *HLA-B27* exist, but not all of them represent risk factors for axSpA. For example, HLA-B*2704 and HLA-B*2705 subtypes confer a genetic risk, whereas the HLA-B*2706 and HLA-B*2709 allotypes are not associated with axSpA even though they only differ in one peptide residue or amino acid. This would suggest that there could be specific peptides that can only be presented by the pathogenic *HLA-B27* subtypes to induce the "arthritogenic" response. In the arthritogenic peptide hypothesis, inflammation is driven by antigen presentation to CD8+ T cells. Existence of *HLA-B27*-restricted CD8+ T cells in the synovial fluid of AS patients was shown three decades ago, whereas more recent next-generation sequencing T-cell receptor (TCR) studies showed expansion of both CD4+ and CD8+ TCR clonotypes in blood and synovium of *HLA-B27*-positive SpA. Expansion of CD8+ TCR clonotypes with the CAS **STDTQYF CDR3 motif were found in AS patients versus *HLA-B27*-positive controls in blood and synovium, which was recently suggested as a target for therapeutic intervention.

The second theory relates to the ability of *HLA-B27* to homodimerize. These *HLA-B27* homodimers can be expressed at the cell surface and interact with innate immune receptors, such as the KIR3DL2 receptor, on T and natural killer (NK) cells resulting in IL-17 production, one of the key pathogenic cytokines in axSpA. However, there is still some controversy regarding whether various allotypes of *HLA-B27* differ in their ability to form homodimers.

A third theory links *HLA-B27* to unfolded protein responses (UPR) and autophagy, which may promote type 17 immune responses through release of IL-23. Support comes from the *HLA-B27* transgenic rat model, while no evidence for an increase in UPR in synovial tissue or blood of *HLA-B27*-positive AS patients was found in steady state, which questions its role in humans.

Although *HLA-B27* has the strongest genetic link, other MHC class I associations have also been found, for example (HLA-Cw6) in PsA. IL-23R gene polymorphisms are shared across most forms of SpA. This suggests overlapping yet distinct features across different disease domains.

FACTORS INSTIGATING THE ONSET OF SPONDYLOARTHRITIS IN GENETICALLY PREDISPOSED INDIVIDUALS

Despite the strong genetic association between *HLA-B27* and AS, only a minority of *HLA-B27*-positive subjects develop disease. In addition, the risk of developing AS in *HLA-B27*-positive first-degree relatives of AS patients is only 20%. This indicates that additional factors are needed to induce disease. Because SpA affects seemingly distinct tissues (spine, peripheral joints, enthesis, anterior eye chamber, intestine, skin), it has been a challenge to seek a common denominator between these different organs. Two main mechanisms are currently under consideration: the role of intestinal microbiota following a loss of barrier integrity and an aberrant response to mechanical stress.

Barrier Integrity Loss and the Intestinal Microbiota in SpA In view of the link between gut inflammation and SpA, much attention has been focused on barrier integrity loss in the intestine and the role of the intestinal microbiota. Evidence supporting this mechanism has been the development of ReA following an infectious gastroenteritis with *Salmonella typhimurium, Shigella enteritidis, Yersinia enterocolitica*, or *Campylobacter enteritidis*. Studies have highlighted the loss of barrier integrity, which may facilitate translocation of bacteria and microbial peptides that can elicit an immune response. The human microbiome is shaped by genetic and environmental factors. Changes in diversity and composition of the gut microbiome have been identified in SpA, pinpointing an intestinal dysbiosis that is similar to patients with IBD. While most information on host-microbial interplay in SpA relates to the intestine, other mucosal surfaces and particularly the urogenital tract (e.g., *Chlamydia*) may serve as an entry route for pathogens to instigate the onset of SpA.

Germ-free *HLA-B27* transgenic rats fail to develop SpA, suggesting a crucial pathogenic role for intestinal microbiota. In humans, microbiome profiling in stool and biopsy samples of *HLA-B27*–positive individuals found that *HLA-B27* affects the gut microbiome even in the absence of disease. This supports the concept that in *HLA-B27*–positive individuals, intestinal dysbiosis occurs prior to the onset of clinical manifestations and may contribute to the risk of disease. Earlier research suggested that *HLA-B27* may promote intracellular survival and persistence of pathogens such as *Chlamydia trachomatis* in monocytes, which could travel to joints and promote synovial inflammation. More recent studies indicate that not all preclinical models of combined gut and joint disease mimicking SpA impact these sites similarly: in a tumor necrosis factor (TNF) cytokine-dependent model with Crohn's-like ileitis and peripheral arthritis, gut inflammation was found to be microbiota dependent but not joint inflammation. This suggests that intestinal microbiota may play a role in some, but not all, SpA patients or disease features.

The skin, as the outer barrier of the body, is also vulnerable to the development of immune-mediated inflammatory disease. Like the inner barriers of the intestines and joints, the skin plays a key role in preserving tissue homeostasis at a site exposed to microbial, chemical, and mechanical challenges. Alterations in the cutaneous microbiome of patients with psoriasis and PsA have been described with striking similarities between nonlesional and lesional skin. However, the functional impact of skin microbiota on the disease trajectory is still ill defined.

Mechanical Stress as Gateway to Joint Disease in SpA The entheses in SpA and PsA are key inflammatory target tissues, and patients can present with both axial and peripheral entheseal inflammation. In both SpA patients and in individuals without inflammatory conditions, mechanical stress in entheses may lead to microdamage that leads to immune activation and tissue repair. The importance of mechanical strain in the pathophysiology of enthesitis in SpA is supported by observations in patients and animal models. Enthesitis occurs more often in the lower limbs, which are subject to higher mechanical strain. In patients with AS, exposure to physical labor was related to more structural damage. In mouse models, mechanical strain was shown to be causally related to development of enthesitis and SpA-related new bone formation through mechanisms that support chronicity of inflammation.

Responses to mechanical stress are physiologic and self-limited, but they appear to be enhanced and prolonged in SpA. The reason for this is unclear, but this may be influenced by genetic factors and/or impaired intestinal barrier function.

Chronic entheseal inflammation in axSpA also leads to excessive local bone responses. This is associated with osteoblast differentiation initiated by prostaglandin E_2, IL-17-A and -F, and IL-22 followed by activation of downstream bone morphogenic proteins and Wnt proteins, which are thought to play a role in pathological new bone formation. This culminates in the generation of bony spurs at vertebral bodies and sacroiliac joints, eventually leading to ankylosis.

Common Versus Private Cytokine Hubs in Disease Domains in SpA Regardless of the route of disease initiation, several common effector pathways have emerged, with TNF and IL-17 representing signature cytokines in the immunologic basis of axSpA.

The impact of TNF-α as a common effector pathway that acts downstream in the inflammatory process has been long recognized given its broad therapeutic efficacy across all disease domains (gut, enthesis, skin, eye, spine, and joints). TNF-α is an important activator and product of macrophages that stimulates cytokine production in immune cells and activates fibroblasts. TNF-α can also be made by neutrophils and activated T cells in the inflamed synovium, enthesis, and intestine. By activating osteoclasts, it promotes inflammation-induced bone loss and bone erosions in SpA.

Even though axSpA and IBD share a similar genetic predisposition in IL-23R polymorphisms, the two disorders differ substantially in their cytokine dependency. IL-17A blockade is effective in axSpA, whereas IL-23 inhibition does not provide benefit. Conversely, IL-23 blockers are approved in IBD, but agents targeting IL-17A are contraindicated in active IBD, due to the barrier protective roles of IL-17. In PsA, IL-23 and IL-17 inhibitors both appear to be effective.

These observations instigated a search into cellular sources of IL-17 in SpA and their dependency on IL-23. Several IL-17A–producing cell types, including T17 cells (consisting of both CD4+ T_H17 and CD8+ cytotoxic T17 [Tc17] cells), γ/δ T cells, and ILC3, are present at entheseal sites in steady-state conditions. In synovial fluid of SpA patients, innate-like T cells such as γδ T cells were found to be the major T-cell source of IL-17 in the presence or absence of IL-23. This is of particular interest as these γδ T cells also survey epithelial sites such as skin or intestine. Thus, entheses, synovium, and mucosal sites appear to harbor cells that may boost IL-17A responses in the absence of IL-23.

Janus kinase (JAK) inhibitors, which block signaling of several cytokines, are effective in a wide range of SpA including all clinical domains (Fig. 374-1).

AXIAL SPONDYLOARTHRITIS

AxSpA is a chronic, immune-mediated inflammatory disorder predominantly involving the axial skeleton (sacroiliac joints and the spine) and also the peripheral skeleton. The axial skeleton is always involved, and involvement of the peripheral joints occurs in ~30–40% of patients. The hips, historically considered "root joints," are the most common nonspinal joints affected. Enthesitis in the axial and peripheral skeleton is a common feature, but dactylitis is relatively rare. Extraarticular manifestations include psoriasis, acute anterior uveitis, and IBD. Some patients with axSpA experience gradual spinal bony fusion over several years, which leads to reduced spine and neck flexibility. Despite these osteoproliferative changes in the skeleton, osteoporosis is a common morbidity. AxSpA significantly affects patients' well-being, function, productivity, and health-related quality of life.

■ DEFINITIONS AND NOMENCLATURE

AxSpA is divided into r-axSpA (also called AS) and nr-axSpA, the difference based on the presence or absence of definitive sacroiliitis on plain radiographs. "Definitive sacroiliitis" is defined by the modified New York criteria **(Table 374-2)** as grade 2 bilateral or grade 3 or 4 unilateral or bilateral **(Table 374-3)**. Patients with sacroiliitis of lower grade are classified as nr-axSpA. Since the differentiation between r-axSpA and nr-axSpA is based on the degree of sacroiliitis, it is open to disagreements even between trained musculoskeletal radiologists. The interreader reliability as measured by the kappa statistic is between 0.35 and 0.45. nr-axSpA and r-axSpA have the same clinical features and share similar disease burden, though in late-stage r-axSpA, fusion of spine and hip involvement can add to functional loss.

■ EPIDEMIOLOGY

The male-to-female ratio is 2:1 for r-axSpA and 1:1 for nr-axSpA. The population prevalence of axSpA depends on the prevalence of *HLA-B27* in that geographic region. In Europe, *HLA-B27* frequency in the population increases from south to north, and so does the prevalence of axSpA. For example, the prevalence of *HLA-B27* in Spain and Norway is 7 and 9%, respectively, and the prevalence of AS and axSpA in these two countries is 0.1 and 0.56%. The 2009–2010 National Health and Nutritional Examination Survey (NHANES) estimated that in the United States, the population prevalence of *HLA-B27* is 6% and that of axSpA is between 0.9 and 1.4%, of which 0.5% were reported to have AS. The population prevalence of axSpA in the rest of the world is not well studied and is believed to be between 0.02 and 1.5%. Diagnostic prevalence of axSpA in the United States, which includes only the diagnosed cases in a population, is one-tenth that of population prevalence, suggesting a large undiagnosed population. Based on insurance claims databases in the United States, the diagnostic prevalence of axSpA increased between 2006 to 2019, which could be attributed to increased disease awareness. The incidence rates of AS are 0.005–0.01 per 100 patient-years globally and are not available for nr-axSpA.

The risk of axSpA is 5% in *HLA-B27*-positive people but increases to 20% in *HLA-B27*-positive first-degree relatives of affected individuals with axSpA.

TABLE 374-2 Modified New York Classification Criteria for Ankylosing Spondylitis (AS)

Clinical Criteria
1. Low back pain >3 months
Improved with exercise
Not relieved by rest
2. Limited lumbar motion in fontal and lateral planes
3. Reduced chest expansion
Radiographic Criteria
1. Bilateral grade >2 sacroiliitis on x-ray
2. Unilateral or bilateral grade 3 or 4 on x-ray
Definite AS requires ≥1 clinical criterion plus 1 radiographic criterion

Source: Adapted from S van der Linden et al: Arthritis Rheum 27:361, 1984.

TABLE 374-3 Grading of Sacroiliitis

GRADE	DESCRIPTION
Grade 0	Normal
Grade 1	Suspicious change
Grade 2	Minimum abnormality (small, localized areas with erosions or sclerosis, without alterations in the joint width)
Grade 3	Unequivocal abnormality (moderate or advanced sacroiliitis with erosions, evidence of sclerosis, widening, narrowing, or partial ankylosis)
Grade 4	Severe abnormality (total ankylosis)

Source: Reproduced with permission from S van der Linden et al: Evaluation of diagnostic criteria for ankylosing spondylitis. Arthritis Rheum 27:361, 1984.

■ CLINICAL MANIFESTATIONS

The most common first manifestation of axSpA is low back, hip, or buttock pain, which starts before the age of 40–45 years, usually in the 20s to 30s. The typical characteristics of this back pain include any combination of the following: insidious onset, chronicity (>3 months' duration), age of onset <45 years, improvement with exercise or activity and no improvement with rest, worse back pain in the second half of the night with improvement after getting out of bed and walking around, and prolonged (>30 min) morning stiffness. Back pain with these characteristics is termed "inflammatory back pain" (IBP) to differentiate it from the common "mechanical back pain" **(Table 374-4)**, but the term is a misnomer because only 15% of patients with IBP have demonstrable inflammation in their axial skeleton. Rarely, thoracic or neck pain can be a presenting feature, and when present, it is more commonly seen in women. Fatigue and stiffness are the two most bothersome symptoms after back pain in axSpA. Fatigue is multifactorial, secondary to sleep disturbance, active inflammation, and anemia, and may also indicate underlying depression and anxiety of chronic disease.

Peripheral arthritis is seen in 30–40% of patients with axSpA. Typically, this is an asymmetric, oligoarticular inflammatory arthritis involving the large joints of the lower extremities and rarely the small joints of hands and feet. Hip involvement, presenting as pain in the groin with radiation to medial thigh or the knee, is common and is associated with significant functional impairment. Enthesitis is a common manifestation of axSpA. Enthesitis may present as heel pain (plantar fasciitis, Achilles tendon insertion pain), chest wall pain, intercostal muscle insertions, medial or lateral epicondylitis, quadriceps or patellar tendon insertion pain around knees, or pelvic rim pain. Dactylitis is an uncommon manifestation of axSpA. Persistent axial inflammation may lead to bony fusion of sacroiliac joints, apophyseal joints, and development of syndesmophytes by osteoproliferation in the outer fibers of the annulus fibrosus of the intervertebral disks. These bony changes lead to limitation of spinal, including neck, mobility in all directions and are a major cause of significant functional impairment in late stages of the disease.

Extramusculoskeletal manifestations of axSpA, such as psoriasis, IBD, and acute anterior uveitis, can also be the presenting symptom of axSpA in some patients, and such patients may first present to a dermatologist, gastroenterologist, or an ophthalmologist. Uveitis, the most

TABLE 374-4 Clinical Features of Inflammatory Versus Mechanical Back Pain

FEATURE	INFLAMMATORY BACK PAIN	MECHANICAL BACK PAIN
Age at onset	Before 40–45 years	20–65 years
Onset	Insidious	Acute or insidious
Morning stiffness	Prolonged (more than 30 min)	Less than 30 min
Pain at night	Yes, usually after midnight	No, usually late in the day
Exercise/activity	Improves pain and stiffness	Worsens pain
Rest/inactivity	Worsens pain and stiffness	Improves pain
Duration	Chronic	Acute or chronic
Response to full-dose NSAIDs	More than 50% relief in 48 hours	Limited relief

Note: Not all features are required for diagnosis of inflammatory back pain, nor are they present in every patient.

Abbreviation: NSAIDs: nonsteroidal anti-inflammatory drugs.

common extramusculoskeletal manifestation, is seen in 40% of patients with axSpA and presents as eye discomfort followed by redness, pain, photophobia, and miosis. The typical phenotype of uveitis associated with axSpA is acute, anterior, unilateral, and episodic. While frank IBD is seen in 10%, histologic evidence of subclinical gut inflammation is seen in up to 50% of patients. Psoriasis is seen in 10%, and osteoporosis can be present in 40%.

AxSpA can affect multiple organ systems. The two renal manifestations of axSpA are IgA nephropathy, which can present at any time in the disease course as microscopic hematuria, and nephrotic syndrome secondary to renal (AA) amyloidosis, which is a late complication only seen after prolonged uncontrolled inflammation. Conduction abnormalities (heart blocks) can be seen at any stage of the disease, whereas aortic valve insufficiency and cauda equina syndrome presenting as urinary hesitancy and/or saddle anesthesia are late complications. Pulmonary manifestations of axSpA include apical fibrosis, cavitary lung lesion, or fibrotic parenchymal lesions. Sleep apnea syndrome and restrictive lung disease seen in late-stage axSpA are mostly related to osteoproliferative structural changes of the cervical spine and rib cage.

The course of axSpA can be variable, with close to 50% of patients progressing from nonradiographic to radiographic stage over 20 years and <10% progressing to significant spinal involvement with bamboo spine. Risk factors for osteoproliferation include male sex, persistent inflammation (seen on magnetic resonance imaging [MRI] of sacroiliac joints and spine, high C-reactive protein [CRP]), syndesmophytes, presence of *HLA-B27*, and smoking.

DIAGNOSIS

A diagnostic delay of 5–14 years from symptom onset of axSpA reported in various parts of the world is mostly related to the commonality of mechanical back pain in the general population. Traditionally, AS is thought to be a disease of males, and back pain and enthesitis in women are commonly mistaken for fibromyalgia. This delays the diagnosis in females even further. Lower prevalence of *HLA-B27* in nonwhite populations also adds to the delay in diagnosis.

The diagnosis of axSpA is based on pattern recognition, ruling out common causes for the symptoms, and clinical reasoning. **Figure 374-2** shows a schema for the diagnosis of axSpA in a patient complaining of chronic (>3 months) back pain. While elevated inflammatory markers, erythrocyte sedimentation rate (ESR), and CRP help in making the diagnosis of axSpA, these tests are neither sensitive (seen in only 30–40% of patients with active axSpA) nor specific for axSpA.

Imaging plays a very important role in the diagnosis of axSpA. A single anteroposterior (AP) view or a Ferguson view x-ray of the pelvis is sufficient to image sacroiliac joints. Multiple views (e.g., oblique) of the sacroiliac joints add little in making the diagnosis; in addition, they increase radiation risk to gonads. Radiographic features of sacroiliitis include marginal sclerosis, erosions, narrowing and widening of the joints, and in late stages, fusion (**Fig. 374-3*A*** and Table 374-3). As a general rule, plain radiographs of the spine should be avoided in patients with chronic back pain. However, there are some characteristic changes of axSpA seen in the lateral radiograph of the spine, and they include Romanus lesions or the shiny corner sign and squaring of the vertebral body seen as a result of erosions at the attachments of the spinal ligaments at the vertebral corners. Andersson lesion is an uncommon finding and is characterized by vertebral body erosion and sclerosis at the intervertebral disk level. In the late stages, ossification of the outer layer of annulus fibrosus leads to syndesmophyte formation **(Fig. 374-3*B*)**. Typical syndesmophyte orientation is vertical, differentiating it from the horizontal orientation of osteophytes, which is commonly seen with osteoarthritis (OA) of the spine. Bamboo spine, or ankylosis of the entire spine, is seen in a very small percentage of late-stage AS patients.

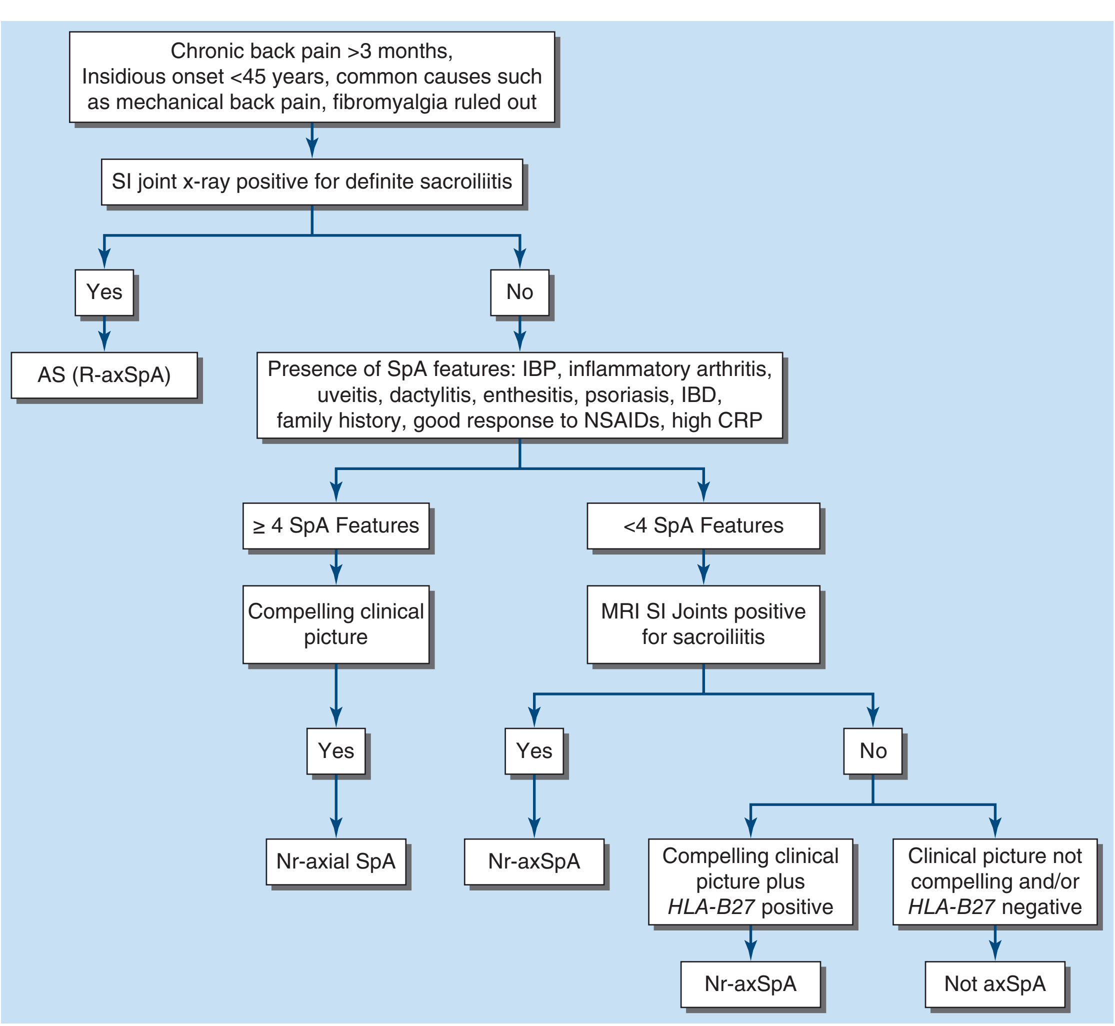
Chronic back pain >3 months, Insidious onset <45 years, common causes such as mechanical back pain, fibromyalgia ruled out
SI joint x-ray positive for definite sacroiliitis
Yes
No
AS (R-axSpA)
Presence of SpA features: IBP, inflammatory arthritis, uveitis, dactylitis, enthesitis, psoriasis, IBD, family history, good response to NSAIDs, high CRP
≥ 4 SpA Features
<4 SpA Features
Compelling clinical picture
MRI SI Joints positive for sacroiliitis
Yes
Yes
No
Nr-axial SpA
Nr-axSpA
Compelling clinical picture plus *HLA-B27* positive
Clinical picture not compelling and/or *HLA-B27* negative
Nr-axSpA
Not axSpA

FIGURE 374-2 Schema for the diagnosis of axial spondyloarthritis in a patient complaining of chronic back pain (back pain lasting >3 months).

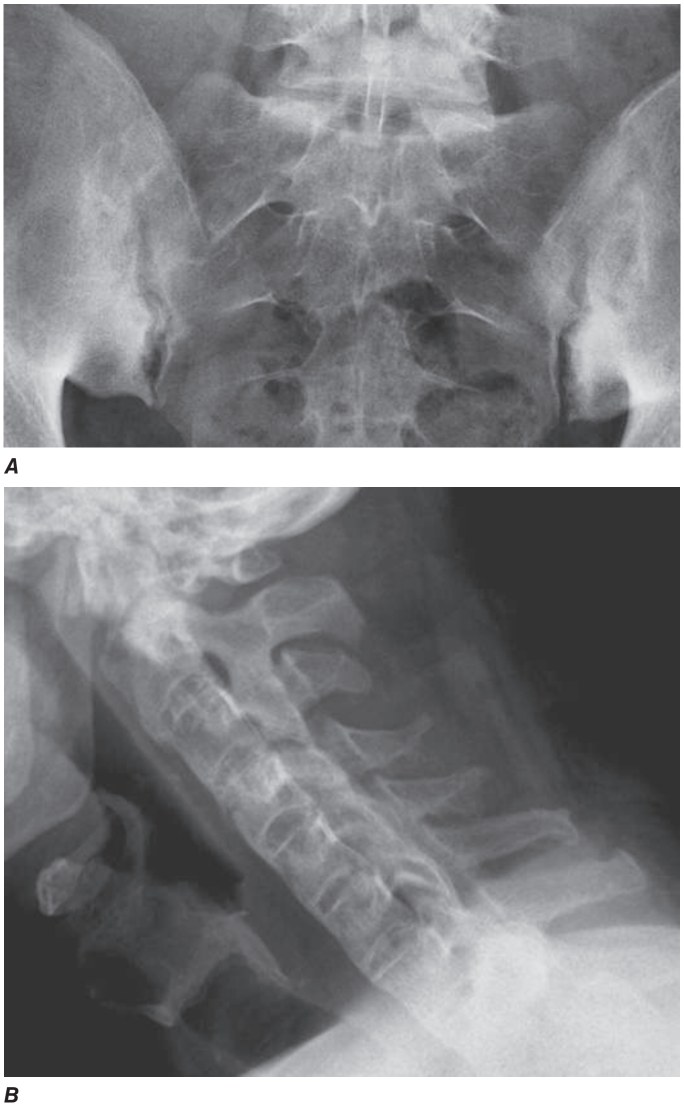

FIGURE 374-3 Radiographic axial spondyloarthritis. ***A.*** Bilateral sacroiliitis (modified New York grade 3) with sclerosis and erosions. ***B.*** Lateral view cervical spine in advanced radiographic axial spondyloarthritis showing anterior syndesmophytes and fused facet joints.

Low-dose computed tomography (CT) of the sacroiliac joints and spine in the diagnosis of axSpA is emerging as an alternative to plain radiography for diagnosis and assessing progression **(Fig. 374-4)**.

Magnetic resonance imaging (MRI) of the sacroiliac joints and spine has evolved as an important tool in making the diagnosis of axSpA. During the early stage of disease, MRI of the sacroiliac joints may show evidence of active inflammation, or "osteitis," and that may be the only abnormality. As time passes, the active inflammatory changes transition to fatty metaplasia, and this may further transform to new bone formation. The structural changes of sclerosis, fat metaplasia, erosions, fat metaplasia in an erosion cavity (backfill), and new bone formation, in addition to the inflammatory changes seen on MRI, aid in making the diagnosis of axSpA **(Fig. 374-5)**. Sole presence of inflammatory lesions in early stages or sole presence of ankylosis in very late stage may be enough for the diagnosis of axSpA, but as a general rule, multiple types of inflammatory and structural lesions increase the suspicion of axSpA. In the spine, multiple corner inflammatory lesions and/or multiple corner fatty lesions increase the confidence of axSpA diagnosis.

MRI is a very sensitive imaging technique, and mechanical stress on the sacroiliac joints in professional athletes, postpartum women, and even in normal individuals, especially above the age of 40, may show changes of osteitis. Fat metaplasia-type changes in the sacroiliac joints are also seen in degenerative arthritis as well as in normal individuals. Inappropriate utilization of MRI can lead to overdiagnosis.

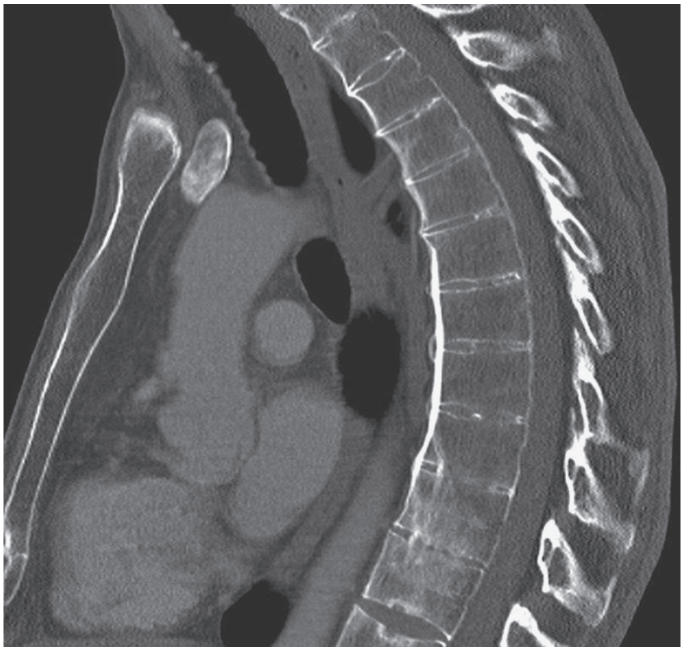

FIGURE 374-4 Computed tomography of the thoracic spine sagittal view in radiographic axial spondyloarthritis showing anterior and posterior syndesmophytes.

DIFFERENTIAL DIAGNOSIS

Chronic nonspecific "mechanical" back pain is common in the general population, and axSpA is the etiology in only 4–5% of such patients. Mechanical causes of back pain therefore should be considered first in a patient presenting with chronic back pain. The so-called "red flag"

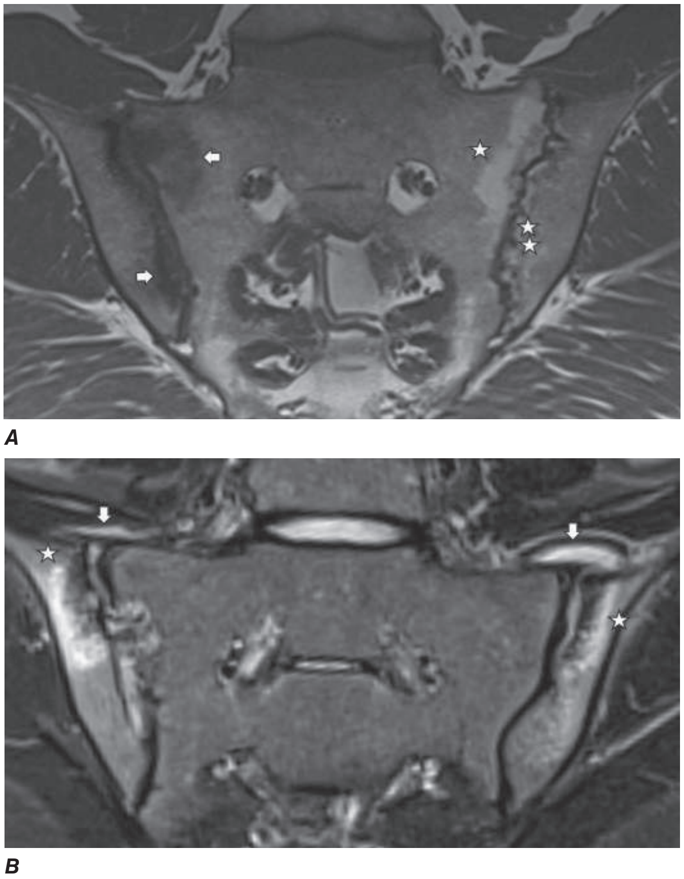

FIGURE 374-5 Magnetic resonance images of spondyloarthritis. ***A.*** T1-weighted image of sacroiliac joints with bilateral erosions (left-sided erosions marked by **), left fatty metaplasia (single *), and right bone marrow edema (marked with *white arrows*). ***B.*** Short tau inversion recovery (STIR) image with bilateral bone marrow edema suggestive of joint space inflammation (shown by *) and anterior capsulitis (shown by a *white arrow*).

signs of fever, weight loss, advanced age, and past history of malignancy should alert the examiner to look for osteomyelitis, osteoporotic fracture, or metastatic disease. In young patients with generalized body pain, fibromyalgia, central sensitization, hypermobility, hypothyroidism, and hypovitaminosis D may be considered.

Abnormalities on incidental imaging in a patient with or even without chronic back pain lead to other differential diagnoses such as diffuse idiopathic skeletal hyperostosis (DISH) or osteitis condensans ilii (OCI). DISH is a noninflammatory, degenerative condition affecting the spine, with exuberant new bone formation in the form of anterior and posterior longitudinal ligament ossification of at least four contiguous vertebral bodies and bulky "flowing" osteophytes typically on the right side of the thoracic spine, but normal sacroiliac joints. It is generally seen in obese, diabetic males, often older than 50 years of age. DISH is sometimes mistaken for bamboo spine of AS. OCI is usually an asymptomatic condition of multiparous women characterized by radiographic findings of a triangular area of dense sclerosis on the lower and inferior part of the iliac side of the sacroiliac joints. This can be mistaken for sacroiliitis, and in early stages, MRI of the sacroiliac joints may be indistinguishable from that of axSpA. Lack of erosions should help distinguish OCI from axSpA. OCI can be seen in nulliparous women and even men. While DISH and OCI are generally asymptomatic, some patients with either condition may present with chronic back pain.

Inflammatory back pain with sclerotic changes on sacroiliac joint radiographs can also be seen in conditions such as postpartum insufficiency fracture of the sacrum; septic sacroiliitis from tuberculosis, brucellosis, fungi, and other infectious agents; and rarely, malignancies such as acute lymphoblastic leukemia.

■ MONITORING

AxSpA patients should be monitored at every visit for disease activity, function, and medication safety. The frequency of monitoring is individualized. Assessment of disease activity can be performed by using the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) or the Ankylosing Spondylitis Disease Activity Score (ASDAS). Functional impairment can be measured by the Bath Ankylosing Spondylitis Functional Index (BASFI). This is a patient-administered questionnaire that includes 10 activities of daily living. The Bath Ankylosing Spondylitis Metrology Index is an index of spinal and hip mobility in patients with AS.

The ESR and CRP may be useful in monitoring disease activity. MRI of the sacroiliac joints or spine is expensive and is not indicated to monitor disease activity of AxSpA. Spine radiographs and MRI are not used for monitoring purposes in daily practice, although they may be used if back pain persists or relapses to rule out active inflammation despite therapy or complications such as spinal fractures.

TREATMENT

Axial Spondyloarthritis

Principles of treatment of nr-axSpA and r-axSpA are identical and are described as one disease. Some agents are approved to treat only AS (r-axSpA) and are mentioned below.

The targets for treatment include low disease activity or remission. Traditionally, the treatment of axSpA was limited to exercise, physical therapy, and nonsteroidal anti-inflammatory drugs (NSAIDs), but with the advent of biologics and targeted synthetic disease-modifying antirheumatic drugs (tsDMARDs), great strides have been made in in the last two decades. Shared decision-making regarding all modes of therapy is the key to success.

Treatment of axSpA can be divided into nonpharmacologic, pharmacologic, and surgical.

NONPHARMACOLOGIC INTERVENTIONS

Physical exercise in the form of active physical therapy (on land or aquatic) and self-directed physical exercise are strongly recommended by all international guidelines and should be promoted. Physical therapy improves fatigue, mobility, and posture. Smoking cessation is encouraged in all patients.

PHARMACOLOGIC INTERVENTIONS

The American College of Rheumatology (ACR), Spondylitis Association of America (SAA), Spondyloarthritis Research and Treatment Network (SPARTAN) treatment guidelines, and ASAS-EULAR (European Alliance of Associations for Rheumatology) treatment recommendations have many commonalities regarding the order of treatments, choices of agents, and precautions. **Table 374-5** outlines pharmacologic interventions for the treatment of axSpA.

The use of immunosuppressive biologics and tsDMARDs has been associated with an increased risk of serious infections. While bacterial infections are seen in both classes, herpes zoster is mostly seen with JAK inhibitors (JAKis), and reactivation of latent tuberculosis is mostly seen with TNF inhibitors (TNFis). Mucosal and skin *Candida* infections are associated with IL-17 inhibitor (IL-17i) use.

Other side effects of biologic therapy include injection-site reactions with subcutaneous injections or infusion reactions with intravenous (IV) medication. TNFi adverse effects include development of demyelinating disease, worsening of congestive heart failure, and paradoxical development of psoriasis-like skin lesions. Adverse events associated with IL-17i therapy include leukopenia and development or worsening of IBD. JAKis have a higher risk of major cardiovascular adverse events and cancers (lymphoma and lung cancers), especially in patients aged 65 years or older and in those with current or past history of smoking, with history of cardiovascular disease, or with malignancy. As a result, the U.S. Food and Drug Administration (FDA) suggests reserving these medicines for patients who have had an inadequate response or intolerance to one or more TNFi.

SURGICAL INTERVENTIONS

Hip involvement in axSpA leads to significant functional impairment, and total hip arthroplasty should be considered in patients with refractory hip pain or functional decline in the presence of radiographic evidence of structural damage. Spinal osteotomies should be reserved for severe fixed kyphotic deformities in advanced AS that affect horizontal vision and should be performed in centers of expertise.

■ COMPLICATIONS

The prevalence of osteoporosis is ~25% after 10 years of axSpA. Vertebral fractures are seen in ~10% of patients with AS. The lower cervical spine is the most commonly involved area, where there is often also neurologic compromise. One should consider vertebral fractures in patients with neck and back pain that has changed in intensity or character from baseline. Another rare complication is the cauda equina syndrome from long-standing AS resulting from inflammation of the lumbosacral nerve roots caused by arachnoiditis. Cardiac manifestations include aortic valve insufficiency and conduction abnormalities. Pulmonary manifestations are rare and include chest expansion restriction from ossification of costovertebral joints and upper lobe–predominant interstitial lung disease. Patients with AS can have IgA nephropathy or, in late stages, renal amyloidosis if the disease is left untreated. Cardiovascular disease is being recognized as a common comorbidity in patients with AS and is linked to chronic inflammation. All patients with AS should be counseled about the traditional risk factors for cardiovascular disease, and appropriate treatment should be instituted if necessary.

PSORIATIC ARTHRITIS

PsA is a relatively common immune-mediated inflammatory condition that may lead to progressive joint damage. If left untreated, it is associated with poor quality of life, loss of function, and increased morbidity and mortality. PsA, associated with cutaneous and nail psoriasis, involves peripheral joints, axial skeleton (sacroiliac joints and spine), entheses, and tendon sheaths. In addition, uveitis and inflammatory bowel disease may coexist with PsA. Common comorbidities of PsA

TABLE 374-5 Pharmacologic Management of Axial Spondyloarthritis (axSpA)

- Nonsteroidal anti-inflammatory drugs (NSAIDs) should be used in highest tolerable doses continuously in active disease and on an as-needed basis if the disease is stable. No particular NSAID is preferred over any other. Side effects to monitor are gastric ulcer disease, hypertension, renal insufficiency, and cardiovascular disease.
- Conventional synthetic DMARDs like sulfasalazine, up to 3 g/d, and methotrexate, up to 25 mg/wk, are useful for the treatment of peripheral joint and entheses involvement, but not for axial disease in axSpA.
- If the disease remains active despite a 4-week trial of full-dose NSAIDs, treatment should be escalated to start a biologic from either the TNFi or IL-17i class.
- Before initiating therapy with biologics or JAKi, screening for latent tuberculosis, hepatitis B, and hepatitis C should be performed.
- All five TNFi agents are approved for the treatment of AS, while selective TNFis are approved for nr-axSpA in various countries.
  - Soluble receptor of TNF–etanercept 50 mg subcutaneous (SC) injection once weekly.
  - TNFi monoclonal antibodies:
    - Adalimumab 40 mg SC every 2 weeks.
    - Certolizumab pegol 200 mg SC twice a month or 400 mg SC once a month.
    - Golimumab 50 mg SC once a month or golimumab 2 mg/kg intravenous (IV) at weeks 0, 4, and then every 8 weeks thereafter.
    - Infliximab 5 mg/kg IV infusion at weeks 0, 2, 6, and then every 6 weeks.
  - No TNFi is preferred, except in patients with concomitant uveitis or IBD in whom TNF monoclonal antibodies are preferred over soluble receptor TNFi therapy.
  - In the case of primary failure of TNFi, treatment can be switched to IL-17i or JAKi. In the case of secondary failure of TNFi therapy (failing after initial benefit), an alternative TNFi can be used.
- Two IL-17Ais are approved for the treatment of nr-axSpA and AS.
  - Secukinumab 150 mg SC at weeks 0, 1, 2, 3, and 4, followed by 150 or 300 mg every 4 weeks.
  - Ixekizumab 160 mg SC once, followed by 80 mg every 4 weeks.
  - In case of primary failure of IL-17Ai, treatment can be switched to TNFi or JAKi. In the case of secondary failure of IL-17Ai therapy, an alternative IL-17i can be used.
- Two JAKis are approved for the treatment of AS, while only one is approved for nr-axSpA. JAKi should be used in patients who have failed TNFi.
  - Tofacitinib 5 mg orally two times a day is approved for the treatment of AS only.
  - Upadacitinib 15 mg once a day is approved for the treatment of AS and nr-axSpA.
- One IL-17A and IL-17F inhibitor is approved in certain parts of the world to treat both nr-axSpA and radiographic axSpA.
  - Bimekizumab 160 mg SC every 4 weeks.
- Local glucocorticoid injections for upper extremity enthesitis or intraarticular injections for peripheral arthritis are indicated. Injections around lower limb weight-bearing entheses are not recommended because of the risk of rupture.
- Systemic glucocorticoids do not have any efficacy in axSpA and should be avoided.

Abbreviations: AS, ankylosing spondylitis; DMARDs, disease-modifying antirheumatic drugs; IBD, inflammatory bowel disease; IL-17i, interleukin 17 inhibitor; JAKi, Janus kinase inhibitor; nr-axSpA, nonradiographic axial spondyloarthritis; TNFi, tumor necrosis factor inhibitor.

include depression, metabolic syndrome, diabetes mellitus, and cardiovascular disease. Sometimes the term "psoriatic disease" is used to emphasize these varied manifestations.

In 1973 Moll and Wright defined PsA as an inflammatory arthritis (peripheral arthritis and/or sacroiliitis or spondylitis), with presence of cutaneous psoriasis and absence of RF. They demonstrated the association between PsA and the gene $HLA\text{-}B^{*}27$ and described five phenotypes of PsA, namely asymmetric oligoarticular inflammatory arthritis, symmetric polyarticular arthritis, distal interphalangeal joint arthritis, spondylitis, and arthritis mutilans. Over the years, enthesitis and dactylitis have been added to these musculoskeletal manifestations. It is important to remember that these clinical phenotypes are not exclusive and frequently coexist.

EPIDEMIOLOGY

With a prevalence of 2–3%, skin psoriasis is one of the most common immune-mediated diseases. The prevalence of PsA in patients with psoriasis is 30% (range, 6–42%). The incidence of PsA in patients with psoriasis is 3 per 100 patient-years (range 2–10) and depends on the severity of skin psoriasis. The likelihood of developing PsA in psoriasis patients does not change over time.

CLINICAL FEATURES

Clinical features of PsA include peripheral arthritis, axial arthritis (spondylitis), enthesitis, dactylitis, and tenosynovitis. Inflammatory arthritis presents with joint pain, swelling, prolonged stiffness, and reduced mobility. Typically, early PsA is oligoarticular (<4 joints) and often asymmetric, and as the number of affected joints increases, the disease becomes polyarticular and the distribution becomes more symmetric. Distal interphalangeal joint involvement is typical of PsA and is associated with psoriatic nail changes since the nail bed is closely linked to the distal interphalangeal (DIP) joints.

Inflammatory arthritis of the axial skeleton occurs in 5% of PsA patients in early stage and can be seen in 50% of patients after 20 years of disease. Axial inflammatory arthritis presents with inflammatory back and/or neck pain and stiffness, which is typically worse after periods of prolonged inactivity and improves with activity. However, axial arthritis can be asymptomatic or overshadowed by patients' peripheral arthritis symptoms. Most patients with axial involvement have sacroiliitis, although axial involvement without sacroiliitis may distinguish axial PsA from axial SpA. While many of the radiographic features may resemble AS, presence of asymmetric sacroiliitis, nonmarginal and asymmetric bulky syndesmophytes, and frequent involvement of cervical spine have been more often assigned to axial PsA then AS. These phenotypic differences have led to the suggestion that this may reflect potential pathogenic distinctions. Whether axial PsA pathogenesis differs substantially from axSpA and especially whether this has therapeutic consequences are the subjects of investigations by several international consortia.

Other important musculoskeletal manifestations of PsA include dactylitis, enthesitis, and tenosynovitis. Dactylitis is defined as inflammatory swelling of an entire finger or toe. Dactylitis is due to inflammation of the joints, tendons, bones, and soft tissues within the digit. Dactylitis is a marker of severity of PsA, and persistent dactylitis leads to structural damage to the joints within that digit. Arthritis mutilans is an end manifestation of a severe arthritis process.

Enthesitis is another important manifestation of PsA. The most common sites affected by enthesitis are the plantar fascia and Achilles tendon insertion on the calcaneum (plantar fasciitis and Achilles enthesitis). Other common sites for enthesitis include quadriceps tendon insertion on the patella, patellar tendon insertion on tibia, medial and lateral epicondyle, pelvic rim, spinous processes, and intercostal muscle insertion on ribs. Tenosynovitis, or inflammation of the tendon sheath, may affect tendons in the hands, wrists, and around the ankles.

Most patients with PsA have psoriasis vulgaris. In ~80% of people with PsA, psoriasis develops first, and the musculoskeletal manifestations develop after a variable duration. In ~10%, both arthritis and psoriasis develop simultaneously. While patients with PsA have psoriasis by definition, in 10% of patients, the arthritis develops first and cutaneous psoriasis manifests a few years later. Diagnosis of PsA is challenging in such patients, and family history of psoriasis, with typical musculoskeletal involvement, can help make the diagnosis. Although 40% of patients with psoriasis without PsA have nail lesions, nail involvement is much more frequent in patients with PsA, affecting close to 90% of patients. Psoriatic nail manifestations may be classified into nail matrix disease and nail bed disease. Nail pitting, leukonychia, crumbling, and red spots in the lunula are manifestations of nail matrix disease, whereas onycholysis, subungual hyperkeratosis, oil drop change, and splinter hemorrhages are manifestations of nail bed disease.

Patients with PsA may also have involvement of the eye and gastrointestinal tract. Uveitis may occur in 5% of patients, and conjunctivitis is rarely seen. While unilateral acute anterior uveitis is the commonest form, posterior uveitis and bilateral eye involvement are also reported. IBD (Crohn's disease or rarely ulcerative colitis) may coexist in patients with PsA. Mucous membrane inflammation presents with painful mouth ulcers and rarely urethritis.

The disease course of PsA is variable. Whereas some patients do well with few joints involved and no significant damage, others progress very quickly to develop marked joint damage within a few months. Patients with PsA have a worse quality of life and function compared with the general population as well as to patients with psoriasis without PsA.

Patients with psoriasis and PsA have an increased risk of cardiovascular disease, as high as the risk in people with diabetes mellitus. The increased risk is independent of cardiovascular risk factors and correlates with more severe skin disease, disease duration, and increased inflammatory markers. Patients with PsA have more severe subclinical atherosclerosis compared with psoriasis patients without PsA. The prevalence of hypertension, type 2 diabetes, obesity, metabolic syndrome, fatty liver disease, angina, and myocardial infarction is higher among patients with PsA compared with the general population. Obesity has an additional adverse impact on PsA disease activity and treatment response. Patients with psoriasis and PsA have increased rates of depression and suicidality compared to the general population. The prevalence of malignancy in PsA is not increased, and mortality studies have shown conflicting results.

DIAGNOSIS

The diagnosis of PsA is based on clinical features of cutaneous psoriasis with inflammatory musculoskeletal disease such as arthritis, dactylitis, enthesitis, or spondylitis. Presence of dactylitis is an important sign, and careful examination of feet is important since dactylitis of a single toe or enthesitis around the heel may be the only musculoskeletal manifestation of PsA in a patient with psoriasis. Psoriasis may be hidden in areas such as the scalp, umbilicus, below breasts, or in the natal cleft, or it may be present in the nail only and should be looked for carefully. Typical involvement of the joints of the fingers or toes is in a "ray" distribution (involvement of all joints in a finger or toe), with one finger showing dactylitis, another with bony ankylosis, and the neighboring finger showing arthritis mutilans. The pattern of involvement in PsA is often asymmetric compared with RA, in which the pattern tends to be symmetric. The presence of spondylitis with asymmetric oligoarticular peripheral arthritis would virtually rule out RA and would make the diagnosis of PsA more likely. The diagnosis of PsA may sometimes be made even in the absence of psoriasis if the above characteristic features are present and if there is a family history of psoriasis or PsA.

Laboratory tests only play a minor role in the diagnosis of PsA. Tests for acute-phase reactants such as ESR or CRP are abnormal in ~50% of patients despite clinically active disease. High levels of ESR or CRP are markers of severe disease and are predictors of radiographic damage and mortality. RF is typically negative, but low titer RF or low titer anticyclic citrullinated peptide antibody may be positive in <10% of patients. *HLA-B27* is positive in 20% of all patients with PsA and 50% of patients with axial PsA. A patient with PsA may present with monoarticular arthritis of knee or ankle, and in this situation, synovial fluid analysis will differentiate between infection ("septic arthritis"), crystal-induced arthritis, OA, or immune-mediated inflammatory arthritis such as PsA. White blood cell counts of >2000/mL of synovial fluid with no crystals and negative Gram stain and culture are diagnostic of inflammatory arthritis and, in this clinical scenario, should raise the suspicion of SpA such as PsA, reactive arthritis, or IBD-associated arthritis. Routine laboratory tests such as complete blood counts or kidney or liver function tests do not help in making the diagnosis of PsA but can be useful for monitoring treatment or assessing comorbidities.

IMAGING FEATURES

Imaging is an important modality in the assessment of PsA and may help in confirming the diagnosis and determining disease severity. Radiographs of the hands, feet, pelvis, spine, and other affected joints should be performed to look for changes suggestive of PsA. In early disease, x-rays of the hands and feet show soft tissue swelling around the involved joints. Periarticular osteopenia is generally absent. Marginal erosions are markers of disease severity and predict further radiographic progression, deformity, and disability. In PsA, erosions are often accompanied by "fluffy" new bone formation. The combination of erosions at joint margins with new bone formation is characteristic of PsA **(Fig. 374-6)**. New bone formation may also cause ankylosis of the joints. There also may be joint space narrowing, "pencil-in-cup" deformity, and in late stages, such as arthritis mutilans, total joint destruction or joint lysis and acro-osteolysis (resorption of the terminal phalanx) are seen.

Certain characteristic radiographic features such as "pencil-in-cup" deformity, bony ankylosis in a ray distribution, or "fluffy" new bone formation close to sites of erosions can be pathognomonic of PsA.

X-ray of the pelvis (anteroposterior view) may show changes of sacroiliitis. Sacroiliitis in PsA is more likely to be unilateral compared to axSpA, where it is usually bilateral. Axial involvement in PsA may show changes of shiny vertebral corners, erosions, squaring of vertebrae, and syndesmophyte formation on the lateral view of the spine. While this may mimic radiographic axSpA, often in PsA, the syndesmophytes are chunky and "nonmarginal" and develop from sites away from the vertebral corners **(Fig. 374-7)**. The presence of such syndesmophytes is characteristic of PsA. Cervical and lumbar vertebrae are frequently involved. Rarely, atlantoaxial subluxation develops with potentially serious consequences.

Ultrasound combined with Doppler evaluation can help in differentiating enthesitis from fibromyalgia tender points or in differentiating mechanical versus inflammatory etiology. Ultrasound can also detect erosions before they appear on x-rays, especially in the hand joints. Some of the findings in inflammatory enthesitis in PsA include thickening of the tendon with loss of the regular fibrillar architecture, hypoechoic lesion, and neovascularization around insertions of tendons, ligaments, and joint capsules at the bone. Bony irregularities and erosions would distinguish inflammatory from mechanical enthesitis. Ultrasound, however, is highly operator dependent and is useful in expert hands only, which remains a limitation. MRI with IV contrast of the peripheral joints can detect synovitis, and MRI without contrast in axial joints may show bone marrow edema (on the short tau inversion recovery [STIR] image) suggestive of active inflammation before any abnormality is visualized on x-rays. MRI can also show erosions in the joints even before they are seen on plain x-rays. MRI is also very useful in visualizing enthesitis. CT scans are useful when an MRI is unavailable or contraindicated, although the risk of considerable radiation exposure needs to be balanced with the benefits. Nuclear medicine bone scans are neither specific nor sensitive to diagnose PsA. Use of fluorodeoxyglucose positron emission tomography scans to assess inflammatory burden in PsA is experimental.

CLASSIFICATION CRITERIA

The Classification of Psoriatic Arthritis (CASPAR) criteria **(Table 374-6)** have a specificity of 98.7% and sensitivity of 91.4%. To meet the CASPAR criteria for PsA, individuals must first have inflammatory articular disease (joint, spine, or entheseal) and must have at least three points from the five categories described in Table 374-6.

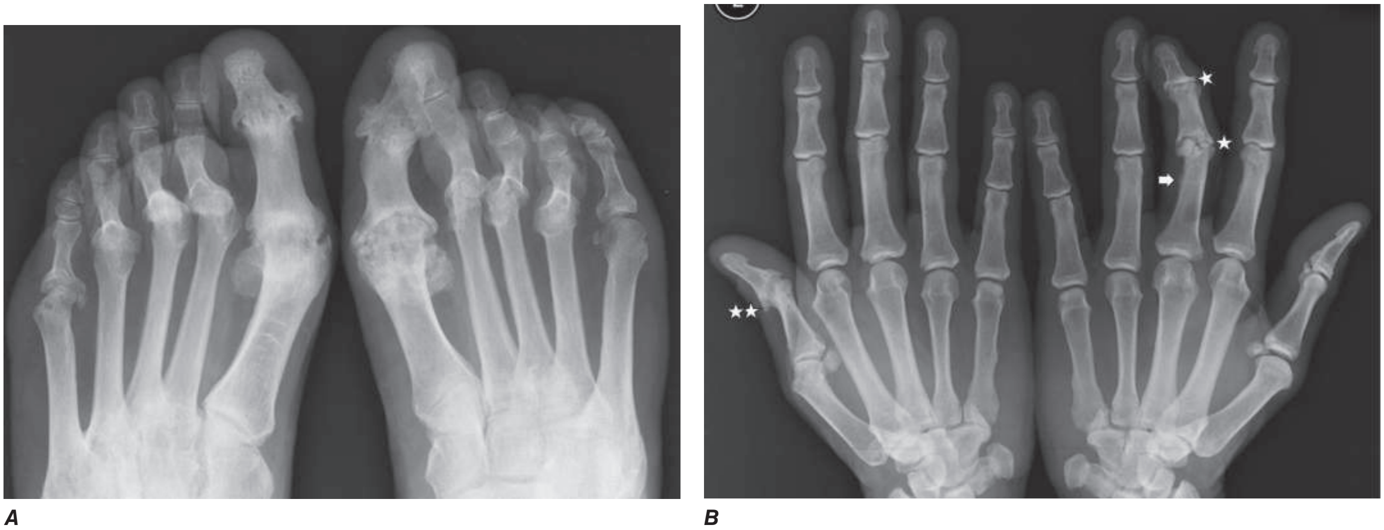

FIGURE 374-6 ***A.*** X-ray of the feet in psoriatic arthritis. These images show erosions, loss of cartilage in bilateral first metatarsophalangeal (MTP) joints, bilateral interphalangeal joint of great toes, and right fifth MTP joint. There are changes of secondary osteoarthritis with sclerosis and osteophyte formation in these joints and early pencil-in-cup deformity in the left fifth MTP joint. ***B.*** X-ray of the hands in psoriatic arthritis. Left thumb interphalangeal joint showing early pencil-in-cup deformity (marked with **). Right third finger has dactylitis and shows periosteal new bone formation in the third proximal phalanx (*white arrow*), along with erosions in the proximal (marked with *) and distal interphalangeal joints (marked with *). This combination of periosteal new bone formation and erosions in "ray" distribution is typical of psoriatic arthritis.

■ DIFFERENTIAL DIAGNOSIS

OA, the most common form of arthritis, occurs in ~5% of the population, and it may coexist with psoriasis especially in the DIP joints. Thus, the presence of Heberden nodules in the context of patients with hand OA with concomitant psoriasis is an important differential diagnosis to be made. Psoriatic nail involvement would differentiate PsA from inflammatory OA of the DIP joints, along with imaging abnormalities such as on ultrasonography. Patients with "seronegative RA" should be carefully examined for presence of psoriasis in hidden areas (scalp, umbilicus, gluteal cleft, genitals, below breasts, and nails) and questioned about family history of psoriasis. Many patients with seronegative RA may have pSpA such as PsA, ReA, or IBD-associated arthritis. The high skin turnover in psoriasis can lead to hyperuricemia, and gout can coexist with PsA. Flare of monoarticular PsA should be assessed with synovial fluid analysis to rule out gout or septic arthritis. PsA shares several clinical, laboratory, and imaging features with other SpA conditions such as ReA, IBD-associated arthritis, and axSpA, and patients may fulfill classification criteria of multiple SpAs simultaneously. Calcium pyrophosphate deposition disease (CPPD) arthritis affecting the MCP joints may coexist in the elderly with skin psoriasis. Chondrocalcinosis of the triangular fibrocartilage in the wrist, "hook-like" osteophytes in the MCP joints, and absence of erosions or fluffy new bone formation should rule out PsA. Synovitis, acne, pustulosis, SAPHO, and rarely hidradenitis suppurativa may mimic palmer plantar psoriasis and PsA.

■ ASSESSING DISEASE ACTIVITY IN PSORIATIC ARTHRITIS

Disease Activity in Psoriatic Arthritis (DAPSA) and Minimal Disease Activity (MDA) are the most common instruments used to assess disease activity in clinical practice. A simple scale used to assess plaque psoriasis is to ask the patient, "How many palms will be covered by the current psoriasis on your body?" With one palm of the patient counted as 1% body surface area (BSA), the total BSA involved by active psoriasis can be assessed. Nail psoriasis and dactylitis can be

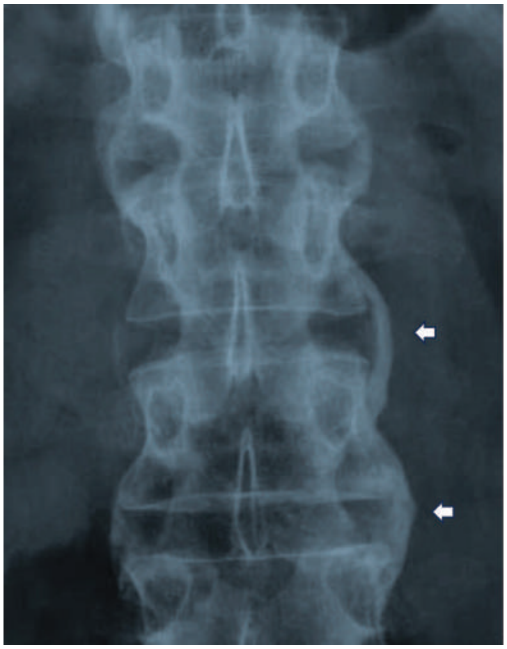

FIGURE 374-7 X-ray of the lumbar spine anteroposterior view in psoriatic arthritis. The images show nonmarginal "chunky" syndesmophytes (marked with *white arrows*).

TABLE 374-6 Classification Criteria for Psoriatic Arthritis (CASPAR) Patients with Inflammatory Articular Disease (Joint, Spine, or Entheseal) Plus ≥3 Points Meet CASPAR Criteria for Psoriatic Arthritis

CATEGORY	DESCRIPTION	POINT VALUE
Psoriasis[a]	Current psoriasis as determined by a rheumatologist or dermatologist	2
	Personal history of psoriasis obtained from patient or qualified health care provider	1
	A family history of psoriasis as reported by patient in a first- or second-degree relative	1
Nail changes	Typical psoriatic nail dystrophy (onycholysis, pitting, and hyperkeratosis) present at assessment	1
Rheumatoid factor	Rheumatoid factor negative	1
Dactylitis (swelling of entire digit)	Current dactylitis or history of dactylitis recorded by a rheumatologist	1
Radiographic evidence	Juxtaarticular new bone formation (excluding osteophyte formation) on plain radiograph of hand or foot; appears as ill-defined ossification near joint margins	1

[a] Psoriasis should account for only one of the three descriptors.

Source: Adapted from WJ Taylor et al: Arthritis Rheum 54: 2665, 2006.

assessed by simply counting the digits with affected fingernails or with dactylitis. A number of enthesitis indices have been developed, such as the Spondyloarthritis Research Consortium of Canada Index and the Leeds Enthesitis Index. Axial arthritis is assessed using tools borrowed from axSpA assessment.

TREATMENT

Psoriatic Arthritis

The treatment target is remission or low disease activity. This is achieved by nonpharmacologic and pharmacologic interventions including novel advanced therapies. A multidisciplinary team consisting of rheumatologists, physical and occupational therapists, dermatologists, and depending upon the domains involved, gastroenterologists, ophthalmologists, psychologists/psychiatrists, dieticians, endocrinologists, and cardiologists may be needed to manage common comorbidities. All pharmacologic interventions need to be based on shared decision-making between the patient and their provider.

The four main classes of pharmacotherapy used in the management of PsA include NSAIDs, conventional synthetic disease-modifying antirheumatic drugs (csDMARDs), biologic DMARDs, and tsDMARDs. Biologic and tsDMARDs have revolutionized the management of PsA by effectively controlling signs and symptoms, decreasing joint damage progression, and improving health-related quality of life. Biologic DMARDs used in the management of PsA include TNFi (etanercept, adalimumab, infliximab, certolizumab, or golimumab), antibodies against IL-12/23 (ustekinumab), IL-17A (secukinumab, ixekizumab), IL-17A and IL-17F (bimekizumab), IL-23 (guselkumab, risankizumab), and costimulatory blockade agent (abatacept). tsDMARDs include the phosphodiesterase-4 (PDE4) inhibitor apremilast; JAKis such as tofacitinib, baricitinib, and upadacitinib; and the tyrosine kinase 2 (TYK2) inhibitor deucravacitinib. The IL-17A and IL-17A+F inhibitor nanobodies izokibep and sonelokimab are currently undergoing clinical trials.

TREATMENT RECOMMENDATIONS

The ACR along with National Psoriasis Foundation (ACR-NPF), EULAR, and Group for the Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) have published separate recommendations for the pharmacologic management of PsA, though there are many similarities. They all recommend that the treatment should be tailored according to the disease activity assessed by level of symptoms and clinical findings in peripheral joints, skin, nails, axial skeleton, entheses, and presence of dactylitis. Individual clinical features (age, gender, concomitant medications, and psychosocial factors), comorbidities (especially IBD, uveitis, metabolic syndrome, and heart disease), prognostic indicators, and patient preferences, values, and cultural background need to be considered in this shared decision-making process. Patients should be monitored at regular intervals and treatment adjusted as appropriate. Treatment is generally continued indefinitely for this lifelong disease since there is high risk of flare once treatment is discontinued.

Principles and recommendations for PsA management from these recommendations are summarized in **Table 374-7**.

For musculoskeletal manifestations of PsA, TNFi and IL-17i have demonstrated comparable efficacy, whereas IL-17i, IL-12/23i, and IL-23i have shown superior efficacy in clearing skin psoriasis compared to TNFi in head-to-head studies. JAKis have comparable efficacy to TNFis for musculoskeletal manifestations of PsA. All biologics except abatacept, and JAKis have shown efficacy in reducing radiographic progression in PsA, but none of the csDMARDs, PDE-4 inhibitors, or TYK2 inhibitors have shown this. While there are no comparative studies on safety, IL-17i, IL-12/23i, IL-23i, PDE-4 inhibitors, and TYK2 inhibitors are perceived to be safer than TNFis in clinical practice.

REACTIVE ARTHRITIS

The interrelation between infection and arthritic disease has been established for many decades. ReA refers to the development of arthritis following urogenital or gastrointestinal infection involving well-defined triggering microorganisms that are normally not cultivable from the joints affected. Prototypical inducers of reactive arthritis include *Yersinia, Campylobacter, Salmonella, Shigella,* and *Chlamydia*.

CLINICAL FEATURES

A crucial factor in the diagnosis relates to recognition of the clinical picture and the identification of causative organisms. A hallmark feature is the time delay between initial infection and onset of arthritis: the interval ranges between 1 day and 4 weeks after infection. Inflammatory musculoskeletal disease (joint, spine, or entheseal) is essential for the diagnosis of ReA. The clinical spectrum may include a typical asymmetrical oligoarticular arthritis preferentially involving lower limbs, enthesitis, tendonitis, bursitis, dactylitis, inflammatory back pain, and sacroiliitis. Mucocutaneous lesions may also occur in addition to the musculoskeletal manifestations and include uveitis/conjunctivitis, erythema nodosum, oral ulcers, circinate balanitis, and keratoderma blenorrhagica.

ReA is strongly associated with the presence of *HLA-B27*, with up to 50–80% of all reactive arthritis patients carrying *HLA-B27*. Accordingly, prevalence of ReA matches the prevalence of *HLA-B27*, which has a clear North-South gradient; e.g., the highest prevalence of *HLA-B27* is observed in the Northern Hemisphere, which was reflected by historical large outbreaks of ReA in, for example, Scandinavian regions. However, over the past decades, there has been a steady decline in the number of such outbreaks with improved hygiene and changes in microbiota. Recent updates on incidence of reactive arthritis point to an overall incidence rate estimate of 3.4 cases per 100,000 person-years. Overall, the estimated percentage of cases to develop ReA following an enteric infection with *Campylobacter, Salmonella, Shigella,* or *Yersinia* is estimated around 2.6%.

Expert consensus refers to ReA only if clinical picture and causative microbes are *HLA-B27* and SpA related. Hence many other infections that trigger onset of arthritis fall outside of the SpA construct. These "infection-related arthritides" constitute all forms of arthritis associated with infections except septic arthritis. They include, among others, acute rheumatic fever, meningococcus, *Mycoplasma genitalium*, *Ureaplasma urealyticum*, *Chlamydia pneumoniae*, beta-hemolytic streptococci, and some live vaccines and COVID. These should not be described as ReA since the clinical pattern does not fit the SpA spectrum and is unrelated to *HLA-B27*. Thus, the term ReA should be restricted to an *acute SpA* that is linked to an acute genitourinary or gastrointestinal infection. The other forms should be referred to as postinfectious arthritides and will not be discussed further in this chapter.

DIAGNOSIS

ReA is diagnosed based on clinical features and identification of an antecedent infection from medical history and laboratory tests and exclusion of other different diagnosis. While diagnostic procedures are focused around *HLA-B27* and molecular testing for *Chlamydia trachomatis* infection, additional individualized diagnostic tests must also be considered to rule out other causes (e.g., postinfectious arthritides). Laboratory tests such as stool cultures to test for *Salmonella, Shigella*, *Campylobacter,* and *Yersinia* can sometimes confirm a preceding or concomitant infection with a causative pathogen, but since gastrointestinal symptoms often have resolved when rheumatic features develop, pathogens may no longer be retrievable. Serology for enteric pathogens is used primarily in epidemiologic studies to test for preceding infections, but there are some questions on its utility in daily clinical practice given the overall low yield of positive results. All patients should be offered screening for *Chlamydia*.

TREATMENT

Reactive Arthritis

Treatment of ReA is oriented to management of SpA features, but treatment guidelines for ReA do not exist currently. While antibiotics are not used in the routine care of ReA, one study demonstrated

TABLE 374-7 Pharmacologic Management of Psoriatic Arthritis (PsA)

PRINCIPLES AND RECOMMENDATIONS

1. The primary goal of treating patients with PsA is to maximize long-term health-related quality of life through control of symptoms, abrogation of inflammation, prevention of structural damage, normalization of function, and social participation.
2. PsA requires multidisciplinary treatment (rheumatologist, dermatologist, psychiatrist, gastroenterologist, ophthalmologist, dietician, endocrinologist, cardiologist, and physical/occupational therapists), and it should be based on a shared decision-making between the patient and the rheumatologist.
3. "Treat to target" is recommended, with the target being minimal disease activity or remission. Patients should be regularly monitored, and treatment should be adjusted appropriately to reach the target.
4. NSAIDs may be used as an initial short-term treatment to relieve symptoms related to peripheral arthritis, enthesitis, dactylitis, and axial disease. NSAIDs alone are very rarely sufficient to treat PsA but can be combined with cs-, b-, or tsDMARDs.
5. Local injections of glucocorticoids should be considered for active arthritis, enthesitis, and/or dactylitis. Glucocorticoid injections around lower extremity weight-bearing entheses should be avoided due to the risk of tendon rupture.
6. Systemic glucocorticoids should be avoided. In rare circumstances such as acute flares of arthritis, they may be used with caution at the lowest effective dose for the shortest period.
7. In patients with active disease with peripheral arthritis, enthesitis, and/or dactylitis, treatment with csDMARDs, such as methotrexate, sulfasalazine, or leflunomide, should be considered at an early stage. Methotrexate is preferred if clinically relevant psoriasis is present.
8. PDE-4 inhibitor or TYK-2 inhibitor may be considered as the initial treatment instead of csDMARD for active musculoskeletal disease (except axial disease) or active skin disease.
9. In patients with predominant axial disease, csDMARDs should be avoided, and either bDMARDs or JAKi should be considered after inadequate response to NSAIDs.
10. In rare circumstances, in patients with very active musculoskeletal disease (multiple swollen joints, structural damage in the presence of active inflammation, and/or clinically relevant extraarticular manifestations such as IBD or uveitis) or extensive skin involvement, treatment with bDMARDs can be started before csDMARDs at the discretion of the treating provider.
11. The choice of initial biologic for active musculoskeletal disease (TNF, IL-12/23, IL-17, or IL-23 inhibitors) should be based on pros and cons of each therapy class, patients' comorbidities, and their comfort with risks/benefit.
12. For patients with predominant skin disease along with musculoskeletal disease, IL-12/23, IL-17, or IL-23 inhibitors are preferred over TNF inhibitors.
13. For PsA patients with comorbid IBD, monoclonal antibodies against TNF, IL-12/23 inhibitors, IL-23 inhibitors, or JAKi therapy is preferred. IL-17 inhibitors should be avoided in patients with active IBD but can be used in patients with history of IBD.
14. For PsA patients with comorbid uveitis, therapy with monoclonal antibodies against TNF is preferred.
15. JAKi should be used after failure of TNF inhibitors according to the black-box warning issued by regulatory authorities based on their CV risk. Comorbidities of metabolic syndrome and CV disease should be considered before JAKi treatment.
16. Combination of methotrexate with bDMARDs such as TNF inhibitors has not been shown to improve efficacy compared to bDMARDs alone.
17. Abatacept has no efficacy against the disease domains of skin, nail, and the axial skeleton. Its efficacy against peripheral arthritis is modest.
18. Agents that effectively treat skin and nail involvement (but with minimal efficacy on the musculoskeletal domains) include topical glucocorticoids such as clobetasol, retinoic acid derivatives, PUVA, and cyclosporin.
19. Biosimilars may be used in place of the original biologics.
20. Tapering of DMARDs in a PsA patient with long-term minimal disease activity or remission should be a shared decision. Stopping DMARD therapy is not recommended.

Abbreviations: bDMARDs, biologic disease-modifying antirheumatic drugs; csDMARDs, conventional synthetic disease-modifying antirheumatic drugs; CV, cardiovascular; IBD, inflammatory bowel disease; IL, interleukin; JAKi, Janus kinase inhibitor; NSAIDs, nonsteroidal anti-inflammatory drugs; PUVA, psoralene plus ultraviolet light A; TNF, tumor necrosis factor; tsDMARDs, targeted synthetic disease-modifying antirheumatic drugs.

that a 6-month course of combination antibiotics (doxycycline or azithromycin, combined with rifampicin) yielded higher response rates then placebo in chronic *Chlamydia*-induced ReA. However, these results have not been replicated, and the routine use of long-term antibiotics to treat chronic ReA is not recommended.

In acute ReA, management of joint symptoms include NSAIDs, intraarticular glucocorticoids for monoarthritis or oligoarthritis, and short-term low-dose systemic glucocorticoids for polyarticular joint involvement. While ReA is thought to be self-limiting and resolving in a large proportion of patients, ReA may also evolve into a chronic phase. The presence of *HLA-B27* has been linked to the evolution into chronicity in part because *HLA-B27* has been suggested to promote intracellular survival of causative microorganisms in infected macrophages, providing a roadmap to chronicity. In chronic forms of ReA (e.g., lasting longer than 6 months and with inadequate response to management outline in acute ReA), DMARDs such as sulfasalazine and methotrexate and even TNFis can be considered. Overall, the management of chronic forms of ReA follows treatment principles of peripheral SpA.

IBD-ASSOCIATED ARTHRITIS

An intriguing relationship between bowel and joint inflammation has emanated for >50 years that points to a prominent role of barrier integrity loss and dysbiosis in SpA, initially recognized by uncovering the occurrence of arthritis in IBD patients. The link between gut and joint inflammation, however, is reciprocal. Hence, patients with a diagnosis of SpA but without associated gastrointestinal symptoms were also found to display histologic signs of inflammation in ileocolonoscopic studies, primarily in the terminal ileum. Here, two types of lesions were described: an acute type of inflammation resembling an infectious enterocolitis and a chronic type of inflammation with many features of Crohn's disease. Overall, the frequency of microscopic inflammatory lesions mounts up to 50% of new-onset treatment-naïve SpA patients and is associated with higher degrees of bone marrow edema in MRI of sacroiliac joints. Furthermore, the presence of chronic lesions represents a risk factor for evolution into full-blown Crohn's disease and progression to AS. These microscopic lesions are accompanied by elevated fecal calprotectin and CRP levels.

Both ulcerative colitis and Crohn's disease are associated with SpA. AxSpA as well as pSpA commonly occur in both of them, with pSpA being the most prevalent form. R-axSpA develops in up to 10% of IBD patients, whereas prevalence of pSpA varies largely according to various studies, ranging from 10 to 30%.

Conversely, patients with axSpA have a substantially higher risk of developing IBD, with a prevalence varying between 5 and 10% and with an observed increase with disease duration. This reciprocal relationship between IBD and AS is reflected by a considerable overlap in the genetic predisposition between these diseases, including polymorphisms in genes linked to type 3 immunity (e.g., IL-23R, IL-12p40, RORc), innate immunity (TNFAIP3, TLR4), and epithelial integrity (GPR35, PTGER4). However, *HLA-B27* is not a risk factor for IBD.

■ CLINICAL FEATURES

Axial skeletal involvement associated with IBD is clinically similar to idiopathic axSpA. There may be, however, a disconnect in time in the occurrence of both. Thus, axSpA may develop after diagnosis of IBD but, in other cases, precedes onset of IBD. Peripheral involvement in IBD resembles the clinical picture of peripheral SpA, which may include more acute forms of oligoarthritis that co-occur with relapses of IBD. More chronic forms of SpA may also occur that may cause substantial structural damage if not appropriately treated. A rare symmetric polyarticular arthritis that runs an independent course has also been described.

In addition to IBD patients with a firm diagnosis of SpA, a comparable fraction of IBD patients have arthralgias or fibromyalgia symptoms.

■ DIAGNOSIS

IBD-related arthritis is diagnosed based on clinical and imaging features in patients with a history of IBD and follows the same diagnostic principles as other forms of SpA. It is important to note, however, that most patients with IBD-related arthritis are seen while being treated for IBD. Thus, special caution should be taken in interpretation of imaging in patients treated with biologics, immune suppressive drugs, and glucocorticoids.

TREATMENT

IBD-Associated Arthritis

Management of IBD-related arthritis follows the general principles of axSpA and pSpA treatment with some special considerations, given the observed disconnect that may occur between gut and joint symptoms with some targeted therapies. Within the targeted therapies, the monoclonal antibodies, but not the soluble receptor, blocking TNF-α are effective in IBD as well as in both axial and peripheral SpA. Other biologics with efficacy in IBD that may have proven efficacy in peripheral SpA include ustekinumab (anti-IL12/23p40), guselkumab, and risankizumab (anti-IL23p19), but anti-IL17 therapy is contraindicated. The JAKi tofacitinib and upadacitinib are approved in ulcerative colitis and upadacitinib in Crohn's disease and are also effective in axSpA and pSpA. Other treatments for IBD include sulfasalazine and related drugs as well as systemic and local glucocorticoids. NSAIDs, especially COX2-selective formulations, are helpful for arthritis and generally well tolerated but may induce IBD flares, so they should be considered only for temporary use. Vedolizumab is a gut-selective integrin inhibitor approved for both Crohn's disease and ulcerative colitis. However, up to 15% of IBD patients treated with vedolizumab may develop new-onset disease or flare of axSpA or pSpA.

UNDIFFERENTIATED PERIPHERAL SPONDYLOARTHRITIS

The term "undifferentiated pSpA" is used to describe patients who do not fit into any of the above well-defined conditions such as axSpA, PsA, ReA, or IBD-associated arthritis but have the typical SpA phenotype (oligoarticular inflammatory arthritis, usually involving lower limbs, asymmetric, or peripheral enthesitis or dactylitis). The ASAS group developed classification criteria for pSpA in 2011 (Table 374-8). These criteria allow the classification of a patient with peripheral enthesitis or dactylitis but without peripheral arthritis. The prevalence and incidence of undifferentiated pSpA are not known since there are no epidemiologic studies available. Some of these patients are likely to have ReA, where the inciting infection is silent, and in some patients, features of psoriasis and/or IBD may develop later. Treatment of pSpA should be individualized based on the predominant musculoskeletal manifestation.

TABLE 374-8 Assessment of Spondyloarthritis International Society (ASAS) Criteria for Peripheral Spondyloarthritis (SpA)

Arthritis or **enthesitis** or **dactylitis**	
Plus ≥1 of:	**Plus ≥2 of the remaining:**
- Psoriasis	- Arthritis
- Inflammatory bowel disease	- Enthesitis
- Preceding infection	- Dactylitis
- *HLA-B27*	- IBP in the past
- Uveitis	- Positive family history for SpA
- Sacroiliitis on imaging (radiographs or MRI)	

Note: Peripheral arthritis: usually lower limb, asymmetric arthritis; enthesitis: clinically assessed; dactylitis: clinically assessed.

Abbreviations: IBP, Inflammatory back pain; MRI, magnetic resonance imaging.

SPONDYLOARTHRITIS ASSOCIATED WITH SAPHO SYNDROME AND HIDRADENITIS SUPPURATIVA

SAPHO is a clinical entity that is sometimes associated with *Propionibacterium acne* infection on bone biopsies. It is included under SpA since some patients have asymmetric inflammatory arthritis (typically sternoclavicular joints and chest wall costochondral junctions), sacroiliitis, and spinal hyperostosis. The skin manifestations include acne conglobate and palmoplantar pustular lesions that are pathologically identical to psoriasis. Some consider SAPHO as the adult phenotype of juvenile chronic recurrent multifocal osteomyelitis (CRMO), where the osteomyelitis is sterile. Apart from TNFi, IL-17i, and JAKi, there are case reports of pamidronate, a bisphosphonate, being effective in treatment. Hidradenitis suppurativa patients may develop a SpA phenotype of musculoskeletal involvement, including asymmetric sacroiliitis.

■ FURTHER READING

Coates L et al: Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA): Updated treatment recommendations for psoriatic arthritis 2021. Nat Rev Rheumatol 18:465, 2022.

Danve A et al: Treatment of axial spondyloarthritis: An update. Nat Rev Rheumatol 18:205, 2022.

Gracey E et al. Revisiting the gut-joint axis: Links between gut inflammation and spondyloarthritis. Nat Rev Rheumatol 16:415, 2020.

Gracey E et al: Tendon and ligament mechanical loading in the pathogenesis of inflammatory arthritis. Nat Rev Rheumatol 16:193, 2020.

Maksymowych WP et al: Data-driven definitions for active and structural MRI lesions in the sacroiliac joint in spondyloarthritis and their predictive utility. Rheumatology (Oxford) 60:4778, 2021.

Moll JM et al: Psoriatic arthritis. Semin Arthritis Rheum 3:55, 1973.

Rudwaleit M et al: The development of assessment of SpondyloArthritis International Society classification criteria for axial spondyloarthritis (part II): Validation and final selection. Ann Rheum Dis 68:777, 2009.

Schett G et al: Reframing immune-mediated inflammatory diseases through signature cytokine hubs. N Engl J Med 385:628, 2021.

van der Linden S et al: Evaluation of diagnostic criteria for ankylosing spondylitis: A proposal for modification of the New York criteria. Arthritis Rheum 27:361, 1984.

Ward MM et al: 2019 update of the American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network recommendations for the treatment of ankylosing spondylitis and non-radiographic axial spondyloarthritis. Arthritis Rheumatol 71:1519, 2019.

375 The Vasculitis Syndromes

Carol A. Langford, Anthony S. Fauci

DEFINITION

Vasculitis is a clinicopathologic process characterized by inflammation of and damage to blood vessels. The vessel lumen is usually compromised, and this is associated with ischemia of the tissues supplied by the involved vessel. A broad and heterogeneous group of syndromes may result from this process, since any type, size, and location of blood vessel may be involved. Vasculitis and its consequences may be the primary or sole manifestation of a disease; alternatively, vasculitis may be a secondary component of another disease. Vasculitis may be confined to a single organ, such as the skin, or it may simultaneously involve several organ systems.

CLASSIFICATION

The vasculitic syndromes as a group have a great deal of heterogeneity while also possessing a considerable degree of overlap. Table 375-1 lists the major vasculitis syndromes. The distinguishing and overlapping features of these syndromes are discussed below.

PATHOPHYSIOLOGY AND PATHOGENESIS

Generally, most of the vasculitic syndromes are assumed to be mediated at least in part by immunopathogenic mechanisms that occur in response to certain antigenic stimuli. However, evidence supporting this hypothesis is for the most part indirect and may reflect epiphenomena as opposed to true causality. Furthermore, it is unknown why some individuals might develop vasculitis in response to certain antigenic stimuli, whereas others do not. It is likely that a number of factors are involved in the ultimate expression of a vasculitic syndrome. These include the genetic predisposition, environmental exposures, and the regulatory mechanisms associated with immune response to certain antigens. Although immune complex formation, antineutrophil cytoplasmic antibodies (ANCA), and pathogenic T lymphocyte responses (Table 375-2) have been among the prominent hypothesized mechanisms, the pathogenesis of individual forms of vasculitis remains complex and varied.

TABLE 375-1 Vasculitis Syndromes

PRIMARY VASCULITIS SYNDROMES	SECONDARY VASCULITIS SYNDROMES
Granulomatosis with polyangiitis	Vasculitis associated with probable etiology
Microscopic polyangiitis	Drug-induced vasculitis
Eosinophilic granulomatosis with polyangiitis (Churg-Strauss)	Hepatitis C virus–associated cryoglobulinemic vasculitis
IgA vasculitis (Henoch-Schönlein)	Hepatitis B virus–associated vasculitis
Cryoglobulinemic vasculitis	Cancer-associated vasculitis
Polyarteritis nodosa	Vasculitis associated with systemic disease
Kawasaki disease	Lupus vasculitis
Giant cell arteritis	Rheumatoid vasculitis
Takayasu arteritis	Sarcoid vasculitis
Behçet's disease	Others
Cogan's syndrome	
Single-organ vasculitis	
Cutaneous leukocytoclastic angiitis	
Cutaneous arteritis	
Primary central nervous system vasculitis	
Isolated aortitis	

Source: Adapted from JC Jennette et al: Arthritis Rheum 65:1, 2013.

TABLE 375-2 Potential Mechanisms of Vessel Damage in Vasculitis Syndromes

- Pathogenic immune-complex formation and/or deposition
  - IgA vasculitis (Henoch-Schönlein)
  - Lupus vasculitis
  - Serum sickness and cutaneous vasculitis syndromes
  - Hepatitis C virus–associated cryoglobulinemic vasculitis
  - Hepatitis B virus–associated vasculitis
- Production of antineutrophilic cytoplasmic antibodies
  - Granulomatosis with polyangiitis
  - Microscopic polyangiitis
  - Eosinophilic granulomatosis with polyangiitis (Churg-Strauss)
- Pathogenic T lymphocyte responses and granuloma formation
  - Giant cell arteritis
  - Takayasu arteritis
  - Granulomatosis with polyangiitis
  - Eosinophilic granulomatosis with polyangiitis (Churg-Strauss)

Source: Reproduced with permission from MC Sneller, AS Fauci: Pathogenesis of vasculitis syndromes. Med Clin North Am 81:221, 1997.

PATHOGENIC IMMUNE-COMPLEX FORMATION

Deposition of immune complexes was the first and most widely accepted pathogenic mechanism of vasculitis. However, the causal role of immune complexes has not been clearly established in most of the vasculitic syndromes. Circulating immune complexes need not result in deposition of the complexes in blood vessels with ensuing vasculitis, and many patients with active vasculitis do not have demonstrable circulating or deposited immune complexes. The actual antigen contained in the immune complex has only rarely been identified in vasculitic syndromes. In this regard, hepatitis B antigen has been identified in both the circulating and deposited immune complexes in a subset of patients who have features of a systemic vasculitis clinically similar to polyarteritis nodosa (see "Polyarteritis Nodosa"). Cryoglobulinemic vasculitis is strongly associated with hepatitis C virus infection; hepatitis C virions and hepatitis C virus antigen-antibody complexes have been identified in the cryoprecipitates of these patients (see "Cryoglobulinemic Vasculitis").

The mechanisms of tissue damage in immune complex–mediated vasculitis resemble those described for serum sickness. In this model, antigen-antibody complexes are formed in antigen excess and are deposited in vessel walls whose permeability has been increased by vasoactive amines such as histamine, bradykinin, and leukotrienes released from platelets or from mast cells as a result of IgE-triggered mechanisms. The deposition of complexes results in activation of complement components, particularly C5a, which is strongly chemotactic for neutrophils. These cells then infiltrate the vessel wall, phagocytose the immune complexes, and release their intracytoplasmic enzymes, which damage the vessel wall. As the process becomes subacute or chronic, mononuclear cells infiltrate the vessel wall resulting in compromise of the vessel lumen with ischemic changes in the tissues supplied by the involved vessel. Several variables may explain why only certain types of immune complexes cause vasculitis and why only certain vessels are affected in individual patients. These include the ability of the reticuloendothelial system to clear circulating complexes from the blood, the size and physicochemical properties of immune complexes, the relative degree of turbulence of blood flow, the intravascular hydrostatic pressure in different vessels, and the preexisting integrity of the vessel endothelium.

ANTINEUTROPHIL CYTOPLASMIC ANTIBODIES

ANCA are antibodies directed against certain proteins in the cytoplasmic granules of neutrophils and monocytes. These autoantibodies are present in a high percentage of patients with active granulomatosis with polyangiitis and microscopic polyangiitis and in a lower percentage of patients with eosinophilic granulomatosis with polyangiitis. Because these diseases share the presence of ANCA and small-vessel vasculitis, they have been grouped collectively as "ANCA-associated vasculitis." However, since these diseases possess unique clinical phenotypes, they should continue to be viewed as separate entities.

There are two major categories of ANCA based on different targets for the antibodies. The terminology of *cytoplasmic ANCA* (cANCA) refers to the diffuse, granular cytoplasmic staining pattern observed by immunofluorescence microscopy when serum antibodies bind to indicator neutrophils. Proteinase-3, a 29-kDa neutral serine proteinase present in neutrophil azurophilic granules, is the major cANCA antigen. More than 90% of patients with active granulomatosis with polyangiitis have detectable antibodies to proteinase-3 (see below). The terminology of *perinuclear ANCA* (pANCA) refers to the more localized perinuclear or nuclear staining pattern of the indicator neutrophils. The major target for pANCA is the enzyme myeloperoxidase; other targets that can produce a pANCA pattern include elastase, cathepsin G, lactoferrin, lysozyme, and bactericidal/permeability-increasing protein. However, only antibodies to myeloperoxidase have been convincingly associated with vasculitis. Antimyeloperoxidase antibodies have been reported in a variable percentage of patients with microscopic polyangiitis, eosinophilic granulomatosis with polyangiitis, isolated necrotizing crescentic glomerulonephritis, and granulomatosis with polyangiitis (see below). A pANCA pattern that is not due to antimyeloperoxidase antibodies has been associated with nonvasculitic entities such as rheumatic and nonrheumatic autoimmune diseases, inflammatory bowel disease, certain drugs, and infections such as endocarditis and bacterial airway infections in patients with cystic fibrosis.

It is unclear why patients with these vasculitis syndromes develop antibodies to myeloperoxidase or proteinase-3 or what role these antibodies play in disease pathogenesis. There are a number of in vitro observations that suggest possible mechanisms whereby these antibodies can contribute to the pathogenesis of the vasculitis syndromes. Proteinase-3 and myeloperoxidase reside in the azurophilic granules and lysosomes of resting neutrophils and monocytes, where they are apparently inaccessible to serum antibodies. However, when neutrophils or monocytes are primed by tumor necrosis factor α (TNF-α) or interleukin 1 (IL-1), or activation of the alternative complement cascade, proteinase-3 and myeloperoxidase translocate to the cell membrane, where they can interact with extracellular ANCA. The neutrophils then degranulate and produce reactive oxygen species that can cause tissue damage. Furthermore, ANCA-activated neutrophils can adhere to and kill endothelial cells in vitro. Activation of neutrophils and monocytes by ANCA also induces the release of proinflammatory cytokines such as IL-1 and IL-8. Adoptive transfer experiments in genetically engineered mice provide further evidence for a direct pathogenic role of ANCA in vivo. In contradiction, however, a number of clinical and laboratory observations argue against a primary pathogenic role for ANCA. Patients may have active granulomatosis with polyangiitis in the absence of ANCA; the absolute height of the antibody titers does not correlate well with disease activity; and patients with granulomatosis with polyangiitis in remission may continue to have high ANCA levels for years (see below).

■ PATHOGENIC T LYMPHOCYTE RESPONSES AND GRANULOMA FORMATION

The histopathologic feature of granulomatous vasculitis has provided evidence to support a role of pathogenic T lymphocyte responses and cell-mediated immune injury. Vascular endothelial cells can express human leukocyte antigen (HLA) class II molecules following activation by cytokines such as interferon (IFN) γ. This allows these cells to participate in immunologic reactions such as interaction with CD4+ T lymphocytes in a manner similar to antigen-presenting macrophages. Endothelial cells can secrete IL-1, which may activate T lymphocytes and initiate or propagate in situ immunologic processes within the blood vessel. In addition, IL-1 and TNF-α are potent inducers of endothelial-leukocyte adhesion molecule 1 (ELAM-1) and vascular cell adhesion molecule 1 (VCAM-1), which may enhance the adhesion of leukocytes to endothelial cells in the blood vessel wall.

APPROACH TO THE PATIENT

General Principles of Diagnosis

The diagnosis of vasculitis should be considered in any patient with an unexplained systemic illness. However, there are certain clinical features that should suggest a diagnosis of vasculitis. These include palpable purpura, pulmonary infiltrates and microscopic hematuria, chronic sinusitis, mononeuritis multiplex, unexplained ischemic events, and glomerulonephritis with evidence of multisystem disease. As a number of nonvasculitic diseases may also produce these abnormalities, the first step in the workup of a patient with suspected vasculitis is to exclude other diseases that can mimic vasculitis **(Table 375-3)**. It is particularly important to exclude infectious diseases with features that overlap those of vasculitis, especially if the patient's clinical condition is deteriorating rapidly and empirical immunosuppressive treatment is being contemplated.

Once diseases that mimic vasculitis have been excluded, the workup should follow a series of progressive steps that establish the diagnosis of vasculitis and determine, where possible, the category of the vasculitis syndrome **(Fig. 375-1)**. This approach is of considerable importance since several of the vasculitis syndromes require aggressive therapy with glucocorticoids and other immunosuppressive agents, whereas other syndromes usually resolve spontaneously and require symptomatic treatment only. The definitive diagnosis of vasculitis is usually made based on biopsy of involved tissue. The yield of "blind" biopsies of organs with no subjective or objective evidence of involvement is very low and should be avoided.

TABLE 375-3 Conditions That Can Mimic Vasculitis

Infectious Diseases
Bacterial endocarditis
Disseminated gonococcal infection
Pulmonary histoplasmosis
Coccidioidomycosis
Syphilis
Lyme disease
Rocky Mountain spotted fever
Whipple's disease
Coagulopathies/Thrombotic Microangiopathies
Antiphospholipid syndrome
Thrombotic thrombocytopenic purpura
Neoplasms
Atrial myxoma
Lymphoma
Carcinomatosis
Drug Toxicity
Cocaine
Levamisole
Amphetamines
Ergot alkaloids
Methysergide
Arsenic
Other
Sarcoidosis
Atheroembolic disease
Antiglomerular basement membrane disease (Goodpasture's syndrome)
Amyloidosis
Migraine
Fibromuscular dysplasia
Heritable disorders of connective tissue
Segmental arterial mediolysis (SAM)
Reversible cerebral vasoconstrictive syndrome

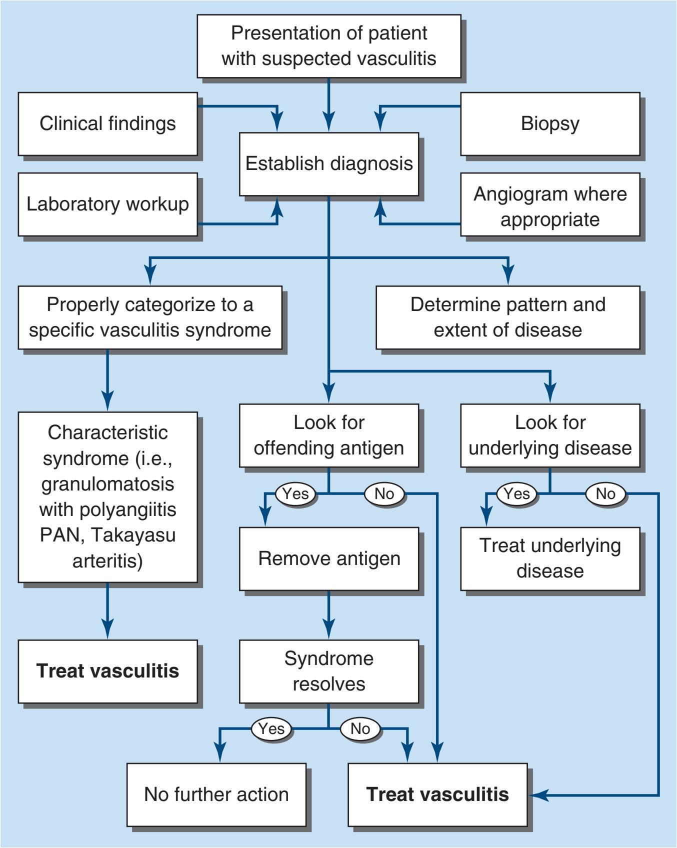
Presentation of patient with suspected vasculitis
Clinical findings
Biopsy
Establish diagnosis
Laboratory workup
Angiogram where appropriate
Properly categorize to a specific vasculitis syndrome
Determine pattern and extent of disease
Characteristic syndrome (i.e., granulomatosis with polyangiitis PAN, Takayasu arteritis)
Look for offending antigen
Look for underlying disease
Yes
No
Yes
No
Remove antigen
Treat underlying disease
Treat vasculitis
Syndrome resolves
Yes
No
No further action
Treat vasculitis

FIGURE 375-1 Algorithm for the approach to a patient with suspected diagnosis of vasculitis. PAN, polyarteritis nodosa.

When syndromes such as polyarteritis nodosa, Takayasu arteritis, or primary central nervous system (CNS) vasculitis are suspected, arteriogram of organs with suspected involvement should be performed.

GENERAL PRINCIPLES OF TREATMENT

Once a diagnosis of vasculitis has been established, a decision regarding therapeutic strategy must be made (Fig. 375-1). If an offending antigen that precipitates the vasculitis is recognized, the antigen should be removed where possible. If the vasculitis is associated with an underlying disease such as an infection, neoplasm, or connective tissue disease, the underlying disease should be treated. If the syndrome represents a primary vasculitic disease, treatment should be initiated according to the category of the vasculitis syndrome. Specific therapeutic regimens are discussed below for the individual vasculitis syndromes; however, certain general principles regarding therapy should be considered. Decisions regarding treatment should be based on the use of regimens for which there has been published literature supporting efficacy for that particular vasculitic disease. Since the potential toxic side effects of certain therapeutic regimens may be substantial, the risk-versus-benefit ratio of any therapeutic approach should be weighed carefully. Glucocorticoids and/or other immunosuppressive agents should be instituted immediately in diseases where irreversible organ system dysfunction and high morbidity and mortality rates have been clearly established. Granulomatosis with polyangiitis is the prototype of a severe systemic vasculitis requiring such a therapeutic approach (see below). Conversely, aggressive therapy should be avoided for vasculitic manifestations that rarely result in irreversible organ system dysfunction such as isolated idiopathic cutaneous vasculitis. Glucocorticoids should be initiated in those systemic vasculitides that cannot be specifically categorized or for which there is no established standard therapy, with other immunosuppressive agents being added if an adequate response does not result or if remission can only be achieved and maintained with an unacceptably toxic regimen of glucocorticoids. Following improvement, one should continually attempt to taper glucocorticoids and discontinue when possible. When using other immunosuppressive regimens, one should base the choice of agent upon the available therapeutic data supporting efficacy in that disease, the site and severity of organ involvement, and the toxicity profile of the drug.

Most forms of vasculitis carry the potential for relapse such that treatment planning needs to include assessment of relapse risk for that vasculitis and the medication regimen being used.

Physicians should be thoroughly aware of the acute and long-term side effects associated with the agents that are commonly used to treat different forms of vasculitis **(Table 375-4)**.

Morbidity and mortality can occur as a result of treatment, and strategies to monitor for and prevent toxicity represent an essential part of patient care.

Addressing the risk of bone loss is important in all patients receiving glucocorticoids. Daily cyclophosphamide should be taken all at once in the morning with a large amount of fluid throughout the day to reduce the risk of bladder injury, and monitoring for bladder cancer should continue indefinitely.

Maintaining the white blood cell (WBC) count at >3000/μL and the neutrophil count at >1500/μL is essential to reduce the risk of life-threatening infections. Monitoring of the complete blood count every 1–2 weeks for as long as the patient receives cyclophosphamide can effectively prevent cytopenias. Methotrexate, azathioprine, and mycophenolate mofetil are also associated with bone marrow suppression, and complete blood counts should be obtained every 1–2 weeks for the first 1–2 months after their initiation and once a month thereafter. To lessen toxicity, methotrexate is often given together with folic acid, 1 mg daily, or folinic acid, 5–10 mg once a week 24 h following methotrexate. Methotrexate is eliminated by the kidney and contraindicated in renal insufficiency as this increases the risk for toxicity. Prior to initiation of azathioprine, thiopurine methyltransferase (TPMT), an enzyme involved in the metabolism of azathioprine, should be assayed because inadequate levels may result in severe cytopenia.

Rituximab can be associated with infusion reactions. In addition to administering this within a skilled infusion center, these reactions can be lessened by the use of premedications. There is a risk of hepatitis B reactivation with rituximab such that all patients should be screened for this infection prior to its use. Immunoglobulins should be monitored with rituximab use.

Tocilizumab and sarilumab are associated with cytopenias, hepatotoxicity, and hyperlipidemia. Laboratory monitoring for drug toxicity should be performed 4–8 weeks after start of therapy and every 3 months thereafter.

Infection represents a significant toxicity for all vasculitis patients treated with immunosuppressive therapy. Infections with *Pneumocystis jirovecii* and certain fungi can be seen even when the WBC count is within normal limits, particularly in patients receiving glucocorticoids. Use of prophylactic therapy to prevent *P. jirovecii* infection should be based on the treatment regimen being used. Patients with granulomatosis with polyangiitis or microscopic polyangiitis who are receiving induction treatment with glucocorticoids in combination with another immunosuppressive agent should receive trimethoprim-sulfamethoxazole (TMP-SMX) or a different preventive medication. Continuation of *Pneumocystis* prophylaxis during maintenance therapy should be strongly considered in most patients, particularly for those receiving rituximab or who have persistent lymphopenia.

In recent years, national and regional organizations have published treatment guidelines that can provide additional direction to clinicians. It should be emphasized that each patient is unique and requires individual decision-making. Information from guidelines as well as this chapter should serve as a framework for the application of evidence-based approaches with incorporation of patient-specific goals to provide maximal therapeutic efficacy with minimal toxic side effects.

TABLE 375-4 Major Toxic Side Effects of Drugs Used in the Treatment of Vasculitis[a]

CONVENTIONAL IMMUNOSUPPRESSIVE AGENTS	
Glucocorticoids	
Osteoporosis	Growth suppression in children
Cataracts	Hypertension
Glaucoma	Avascular necrosis of bone
Diabetes mellitus	Myopathy
Electrolyte abnormalities	Alterations in mood
Metabolic abnormalities	Psychosis
Severe and opportunistic infections	Pseudotumor cerebri
Cushingoid features	Peptic ulcer diathesis
	Pancreatitis
Cyclophosphamide	
Bone marrow suppression	Hypogammaglobulinemia
Cystitis	Pulmonary fibrosis
Bladder carcinoma	Myelodysplasia
Gonadal suppression	Oncogenesis
Gastrointestinal intolerance	Teratogenicity
	Severe and opportunistic infections
Methotrexate	
Gastrointestinal intolerance	Pneumonitis
Stomatitis	Teratogenicity
Bone marrow suppression	Severe and opportunistic infections
Hepatotoxicity (may lead to fibrosis or cirrhosis)	
Azathioprine	
Gastrointestinal intolerance	Severe and opportunistic infections
Bone marrow suppression	Hypersensitivity
Hepatotoxicity	
Mycophenolate mofetil	
Bone marrow suppression	Severe and opportunistic infections
Gastrointestinal intolerance	Teratogenicity
BIOLOGIC AGENTS AND SMALL-MOLECULE INHIBITORS	
Rituximab (granulomatosis with polyangiitis and microscopic polyangiitis)	
Infusion reactions	Severe and opportunistic infections
Progressive multifocal leukoencephalopathy	Hepatitis B reactivation
Mucocutaneous reactions	Tumor lysis syndrome
Hypogammaglobulinemia	Late-onset neutropenia
Impaired vaccine response	
Tocilizumab (giant cell arteritis)	
Sarilumab (polymyalgia rheumatica)	
Bone marrow suppression	Severe and opportunistic infections
Hepatotoxicity	Gastrointestinal perforation
Hyperlipidemia	Hypersensitivity reactions
Mepolizumab (eosinophilic granulomatosis with polyangiitis [Churg-Strauss])	
Hypersensitivity reactions	Opportunistic infections: herpes zoster
Apremilast (Behçet's syndrome; see Chap. 376)	
Diarrhea, nausea, and vomiting	Weight decrease
Depression	
Avacopan (severe granulomatosis with polyangiitis and microscopic polyangiitis)	
Hepatotoxicity	Hepatitis B reactivation
Hypersensitivity reactions	Serious infections

[a] Consult the drug package insert for a full listing of side effects.

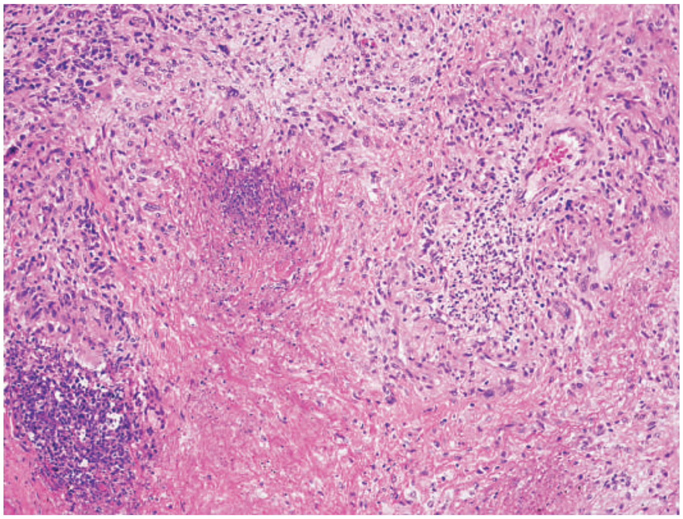

FIGURE 375-2 Lung histology in granulomatosis with polyangiitis. This area of geographic necrosis has a serpiginous border of histiocytes and giant cells surrounding a central necrotic zone. Vasculitis is also present with neutrophils and lymphocytes infiltrating the wall of a small arteriole (*upper right*). *(Courtesy of William D. Travis, MD; with permission.)*

GRANULOMATOSIS WITH POLYANGIITIS

DEFINITION

Granulomatosis with polyangiitis is a distinct clinicopathologic entity characterized by granulomatous vasculitis of the upper and lower respiratory tracts together with glomerulonephritis. In addition, variable degrees of disseminated vasculitis involving both small arteries and veins may occur.

INCIDENCE AND PREVALENCE

Granulomatosis with polyangiitis has an estimated prevalence of 3 per 100,000. It is extremely rare in blacks compared with whites; the male-to-female ratio is 1:1. The disease can be seen at any age; ~15% of patients are <19 years of age, but only rarely does the disease occur before adolescence; the mean age of onset is 40–60 years.

PATHOLOGY AND PATHOGENESIS

The histopathologic hallmarks of granulomatosis with polyangiitis are necrotizing vasculitis of small arteries and veins together with granuloma formation, which may be either intravascular or extravascular **(Fig. 375-2)**. Lung involvement typically appears as multiple, bilateral, nodular or cavitary infiltrates **(Fig. 375-3)**, which on biopsy can reveal necrotizing granulomatous vasculitis. Upper airway lesions, particularly those in the sinuses and nasopharynx, typically reveal inflammation, necrosis, and granuloma formation, with or without vasculitis.

In its earliest form, renal involvement is characterized by a focal and segmental glomerulitis that may evolve into a rapidly progressive crescentic glomerulonephritis. Granuloma formation is only rarely seen on

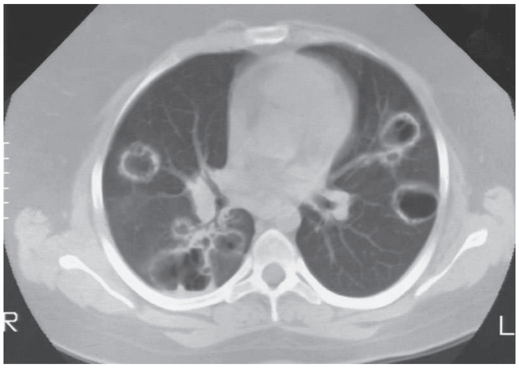

FIGURE 375-3 Computed tomography scan of a patient with granulomatosis with polyangiitis. The patient developed multiple, bilateral, and cavitary infiltrates.

renal biopsy. In contrast to other forms of glomerulonephritis, immune complex deposition is not found in the renal lesion of granulomatosis with polyangiitis (pauci-immune glomerulonephritis). In addition to the classic triad of disease of the upper and lower respiratory tracts and kidney, virtually any organ can be involved with vasculitis, granuloma, or both.

The immunopathogenesis of this disease is unclear, although the involvement of upper airways and lungs with granulomatous vasculitis suggests an aberrant cell-mediated immune response to an exogenous or even endogenous antigen that enters through or resides in the upper airway. Chronic nasal carriage of *Staphylococcus aureus* has been reported to be associated with a higher relapse rate of granulomatosis with polyangiitis; however, there is no evidence for a role of this organism in the pathogenesis of the disease.

Peripheral blood mononuclear cells obtained from patients with granulomatosis with polyangiitis manifest increased secretion of IFN-γ but not of IL-4, IL-5, or IL-10 compared to healthy controls. In addition, TNF-α production from peripheral blood mononuclear cells and CD4+ T cells is elevated and monocytes from patients with granulomatosis with polyangiitis produce increased amounts of IL-12. These findings indicate an unbalanced T_H1-type T-cell cytokine pattern in this disease that may have pathogenic and perhaps therapeutic implications.

A high percentage of patients with granulomatosis with polyangiitis develop ANCA, and these autoantibodies may play a role in the pathogenesis of this disease (see above).

CLINICAL AND LABORATORY MANIFESTATIONS

Involvement of the upper airways occurs in 95% of patients with granulomatosis with polyangiitis. Patients often present with severe upper respiratory tract findings such as paranasal sinus pain and drainage and purulent or bloody nasal discharge, with or without nasal mucosal ulceration **(Table 375-5)**. Nasal septal perforation may follow, leading to saddle nose deformity. Serous otitis media may occur as a result of eustachian tube blockage. Subglottic stenosis resulting from active disease or scarring occurs in ~16% of patients and may result in severe airway obstruction.

Pulmonary involvement (85–90% of patients) may be clinically expressed as cough, hemoptysis, dyspnea, and chest discomfort, or active disease may be asymptomatic in up to 30% of cases. Endobronchial disease, either in its active form or as a result of fibrous scarring, may lead to obstruction with atelectasis.

Eye involvement (52% of patients) may range from a mild conjunctivitis to dacryocystitis, episcleritis, scleritis, granulomatous sclerouveitis, ciliary vessel vasculitis, and retroorbital mass lesions leading to proptosis.

Skin lesions (46% of patients) appear as papules, vesicles, palpable purpura, ulcers, or subcutaneous nodules; biopsy reveals vasculitis, granuloma, or both. Cardiac involvement (8% of patients) manifests as pericarditis, coronary vasculitis, or, rarely, cardiomyopathy. Nervous system manifestations (23% of patients) include cranial neuritis, mononeuritis multiplex, or, rarely, cerebral vasculitis and/or granuloma.

Renal disease (77% of patients) generally dominates the clinical picture and, if left untreated, accounts directly or indirectly for most of the mortality in this disease. Although it may smolder in some cases as a mild glomerulitis with proteinuria, hematuria, and red blood cell casts, once clinically detectable renal functional impairment occurs, rapidly progressive renal failure usually ensues unless appropriate treatment is instituted.

While the disease is active, most patients have nonspecific symptoms and signs such as malaise, weakness, arthralgias, anorexia, and weight loss. Fever may indicate activity of the underlying disease but more often reflects secondary infection, usually of the upper airway.

Characteristic laboratory findings include an elevated erythrocyte sedimentation rate (ESR) and/or C-reactive protein (CRP), anemia and leukocytosis, and mild hypergammaglobulinemia. Thrombocytosis may be seen as an acute-phase reactant. Approximately 90% of patients with active granulomatosis with polyangiitis have a positive antiproteinase-3 ANCA. However, in the absence of active disease, the sensitivity drops to ~60–70%. A small percentage of patients with granulomatosis with polyangiitis may have antimyeloperoxidase rather than antiproteinase-3 antibodies, and up to 20% may lack ANCA.

TABLE 375-5 Granulomatosis with Polyangiitis: Frequency of Clinical Manifestations in 158 Patients Studied at the National Institutes of Health

MANIFESTATION	PERCENTAGE AT DISEASE ONSET	PERCENTAGE THROUGHOUT COURSE OF DISEASE
Kidney		
Glomerulonephritis	18	77
Ear/Nose/Throat	**73**	**92**
Sinusitis	51	85
Nasal disease	36	68
Otitis media	25	44
Hearing loss	14	42
Subglottic stenosis	1	16
Ear pain	9	14
Oral lesions	3	10
Lung	**45**	**85**
Pulmonary infiltrates	25	66
Pulmonary nodules	24	58
Hemoptysis	12	30
Pleuritis	10	28
Eyes		
Conjunctivitis	5	18
Dacryocystitis	1	18
Scleritis	6	16
Proptosis	2	15
Eye pain	3	11
Visual loss	0	8
Retinal lesions	0	4
Corneal lesions	0	1
Iritis	0	2
Other[a]		
Arthralgias/arthritis	32	67
Fever	23	50
Cough	19	46
Skin abnormalities	13	46
Weight loss (>10% body weight)	15	35
Peripheral neuropathy	1	15
Central nervous system disease	1	8
Pericarditis	2	6
Hyperthyroidism	1	3

[a] Fewer than 1% had parotid, pulmonary artery, breast, or lower genitourinary (urethra, cervix, vagina, testicular) involvement.

Source: GS Hoffman et al: Ann Intern Med 116:488, 1992.

Patients with granulomatosis with polyangiitis have been found to have an increased incidence of venous thrombotic events. Although routine anticoagulation for all patients is not recommended, a heightened awareness for any clinical features suggestive of deep-vein thrombosis or pulmonary emboli is warranted.

DIAGNOSIS

The diagnosis of granulomatosis with polyangiitis can be established by the demonstration of necrotizing granulomatous vasculitis on tissue biopsy in a patient with compatible clinical features. Pulmonary tissue offers the highest diagnostic yield, almost invariably revealing granulomatous vasculitis. Biopsy of upper airway tissue usually reveals acute and chronic inflammation but can show granulomas and/or vasculitis. Renal biopsy can confirm a pauci-immune glomerulonephritis

(Chaps. A4 and A14), which differentiates this from other causes of glomerular disease **(Chap. 326)**.

The utility of ANCA in diagnosis varies based on the clinical presentation. The predictive value of a positive ANCA is 90% in patients with active glomerulonephritis. However, in patients who present with sinus and/or lung disease where renal disease is absent, this decreases to 30–60%. Tissue biopsy is typically necessary in these settings to rule out infection or malignancy, which can have a similar presentation and be associated with a false-positive ANCA. Biopsy should also be pursued when clinically inconsistent features are present or when ANCA is absent.

In its typical presentation, the clinicopathologic complex of granulomatosis with polyangiitis usually provides ready differentiation from other disorders. However, if all the typical features are not present at once, it needs to be differentiated from the other vasculitides, anti-glomerular basement membrane disease (Goodpasture's syndrome) **(Chap. 326)**, relapsing polychondritis **(Chap. 378)**, tumors of the upper airway or lung, and infectious diseases such as histoplasmosis **(Chap. 218)**, endocarditis **(Chap. 133)**, mucocutaneous leishmaniasis **(Chap. 233)**, and rhinoscleroma **(Chap. 166)** as well as noninfectious granulomatous diseases.

Of particular note is the differentiation from other *midline destructive diseases*. These diseases lead to extreme tissue destruction and mutilation localized to the midline upper airway structures including the sinuses; erosion through the skin of the face commonly occurs, a feature that is extremely rare in granulomatosis with polyangiitis. Although blood vessels may be involved in the intense inflammatory reaction and necrosis, primary vasculitis is not seen. *Upper airway neoplasms* and specifically *extranodal natural killer (NK)/T-cell lymphoma (nasal type)* are important causes of midline destructive disease. These lesions are diagnosed based on histology, which reveals polymorphous atypical lymphoid cells with an NK cell immunophenotype, typically Epstein-Barr virus **(Chap. 199)**. Such cases are treated based on their degree of dissemination, and localized lesions have responded to irradiation. Upper airway lesions should never be irradiated in granulomatosis with polyangiitis. Cocaine-induced tissue injury can be another important mimic of granulomatosis with polyangiitis in patients who present with isolated midline destructive disease. ANCA that target human neutrophil elastase can be found in patients with cocaine-induced midline destructive lesions and can complicate the differentiation from granulomatosis with polyangiitis. This has been further confounded by the high frequency of levamisole adulteration of cocaine, which can result in cutaneous infarction and serologic changes that may mimic vasculitis. Granulocytopenia is a common finding in levamisole-induced disease that is not associated with granulomatosis with polyangiitis.

Granulomatosis with polyangiitis must also be differentiated from *lymphomatoid granulomatosis*, which is an Epstein-Barr virus–positive B-cell proliferation that is associated with an exuberant T-cell reaction. Lymphomatoid granulomatosis is characterized by lung, skin, CNS, and kidney involvement in which atypical lymphocytoid and plasmacytoid cells infiltrate nonlymphoid tissue in an angioinvasive manner. In this regard, it clearly differs from granulomatosis with polyangiitis in that it is not an inflammatory vasculitis in the classic sense but an angiocentric perivascular infiltration of atypical mononuclear cells. Up to 50% of patients may develop a true malignant lymphoma.

TREATMENT
Granulomatosis with Polyangiitis

Prior to the introduction of effective therapy, granulomatosis with polyangiitis was universally fatal within a few months of diagnosis. Glucocorticoids alone led to some symptomatic improvement, with little effect on the ultimate course of the disease. The development of treatment with cyclophosphamide dramatically changed patient outcome such that marked improvement was seen in >90% of patients, complete remission was seen in 75% of patients, and 5-year patient survival was seen in >80%.

Despite the ability to successfully induce remission, 50–70% of remissions are later associated with one or more relapses. The determination of relapse should be based on objective evidence of disease activity, taking care to rule out other features that may have a similar appearance such as infection, medication toxicity, or chronic disease sequelae. Many patients who achieve remission continue to have a positive ANCA for years, and changes in ANCA should not be used as a measure of disease activity. Results from a large prospective study found that only 43% of patients relapsed within 1 year of an increase in ANCA levels. Thus, a rise in ANCA by itself is not a harbinger of immediate disease relapse and should not lead to reinstitution or increase in immunosuppressive therapy. Reinduction of remission after relapse is almost always achieved; however, a high percentage of patients ultimately have some degree of damage from irreversible features of their disease, such as varying degrees of renal insufficiency, neurologic impairment, hearing loss, subglottic stenosis, saddle nose deformity, and chronic sinus dysfunction. Patients who developed irreversible renal failure but who achieved subsequent remission have undergone successful renal transplantation.

Treatment of granulomatosis with polyangiitis is viewed as having two phases: *induction*, where active disease is put into remission, followed by *maintenance*. Decisions regarding induction and maintenance agents are guided by published data, individual patient factors that include contraindications, relapse history, and comorbidities, and stratification based on disease severity. Severe disease is associated with life- or organ-threatening manifestations and is based not only on the degree of involvement but also the organ affected. Pulmonary involvement can be nonsevere when presenting as nodular disease without respiratory compromise or severe in the setting of alveolar hemorrhage. In contrast, active glomerulonephritis is always considered to be severe disease even when renal function remains normal.

REMISSION INDUCTION OF SEVERE DISEASE

Regimens to induce remission of severe active disease consist of either cyclophosphamide or rituximab given in combination with glucocorticoids and consideration of adjunctive avacopan. Glucocorticoids were historically given as prednisone 1 mg/kg per day for the first month, followed by tapering. In a randomized trial, use of a reduced-dose glucocorticoid regimen was noninferior to a standard-dose regimen and was associated with a lower rate of serious infection.

Avacopan (a C5a receptor inhibitor) was investigated in a randomized trial as an alternative to glucocorticoids in patients receiving induction with either cyclophosphamide or rituximab. At 52 weeks, sustained remission was higher in those who received avacopan as compared to prednisone with a similar rate of serious adverse events. Although glucocorticoids were given within the first few weeks to most patients receiving avacopan, glucocorticoid exposure in the avacopan group remained markedly less than those randomized to the prednisone treatment arm. This supported that glucocorticoids can be rapidly withdrawn in patients with severe active disease receiving avacopan together with either rituximab or cyclophosphamide.

In patients presenting with disease that is life-threatening, methylprednisolone 1000 mg daily for 3 days has been used. Adjunctive plasma exchange was found to provide no added benefit in reducing the composite outcome of end-stage renal disease or death in a large, randomized trial. Whether it may still play a role in selected patients with the most fulminant disease remains controversial. Although a low number of patients in this trial were enrolled with alveolar hemorrhage requiring mechanical ventilation, a retrospective study also showed no benefit. From a meta-analysis, patients with active glomerulonephritis and serum creatinine >5.7 mg/dL who received plasma exchange had improvement in renal function but an increased risk of serious infection. Plasma exchange should be used in patients with granulomatosis with polyangiitis or microscopic polyangiitis who are also positive for antiglomerular basement membrane autoantibody **(Chap. 326)**. When plasma exchange is used concurrently with rituximab, these treatments must be separated by at least 48 h to avoid removal of rituximab.

CYCLOPHOSPHAMIDE INDUCTION FOR SEVERE DISEASE

Daily cyclophosphamide combined with glucocorticoids was the first regimen proven to effectively induce remission and prolong survival. Cyclophosphamide is given in doses of 2 mg/kg per day orally, but because it is renally eliminated, dosage reduction should be considered in patients with renal insufficiency. Although we continue to favor the use of daily cyclophosphamide, a randomized trial found IV cyclophosphamide 15 mg/kg, three infusions given every 2 weeks, then every 3 weeks thereafter, to have a similar rate of remission but higher frequency of relapse compared to cyclophosphamide 2 mg/kg daily given for 3 months followed by 1.5 mg/kg daily. Although leukopenia occurred more often with daily cyclophosphamide, this was likely related to a reduced frequency of blood count monitoring after the first month.

RITUXIMAB INDUCTION FOR SEVERE DISEASE

Rituximab (anti-CD20) has been of proven effectiveness for the treatment of granulomatosis with polyangiitis and microscopic polyangiitis. In two randomized trials that enrolled patients with severe active granulomatosis with polyangiitis or microscopic polyangiitis, rituximab 375 mg/m^2 once a week for 4 weeks in combination with glucocorticoids was found to be as effective as cyclophosphamide with glucocorticoids for inducing disease remission. In the trial that also enrolled patients with relapsing disease, rituximab was found to be statistically superior to cyclophosphamide. Although rituximab does not have the bladder toxicity or infertility risks of cyclophosphamide, in both randomized trials, the rate of adverse events was similar in the rituximab and cyclophosphamide arms.

The decision about whether to utilize cyclophosphamide or rituximab for remission induction must be individually based. Factors to consider include disease severity, whether the patient has newly diagnosed or relapsing disease, medication contraindications, and individual patient factors particularly fertility concerns and risk factors for toxicity. In patients with rapidly progressive glomerulonephritis with a creatinine >4.0 mg/dL or pulmonary hemorrhage requiring mechanical ventilation, daily cyclophosphamide and glucocorticoids may be favored.

There has been no evidence to date that using cyclophosphamide and rituximab together for induction is more efficacious, and combined use raises concern for additive toxicity, particularly infection. However, in instances of toxicity or rare occurrences of inefficacy, there may be a need to switch from cyclophosphamide to rituximab or vice versa.

REMISSION MAINTENANCE OF SEVERE DISEASE

When cyclophosphamide is given for induction, it should be stopped after 3–6 months and switched to another agent for remission maintenance. Medications used in this setting with which there has been published experience from randomized trials are rituximab, azathioprine, methotrexate, and mycophenolate mofetil. A lower rate of relapse was seen with rituximab given at 500 mg for two doses followed by 500 mg every 6 months when compared to azathioprine 2 mg/kg per day. In a randomized trial comparing methotrexate to azathioprine for remission maintenance, similar rates of toxicity and relapse were seen. Methotrexate is administered orally or subcutaneously at a starting dose of 15 mg/week, which is increased by 2.5 mg every 2 weeks up to a dosage of 20–25 mg/week. In patients who are unable to receive methotrexate or azathioprine or who have experienced relapse on such treatment, mycophenolate mofetil 1000 mg twice a day may also sustain remission, but it is associated with a higher rate of relapse compared to azathioprine.

For patients who receive rituximab for remission induction, a randomized trial found that rituximab had a lower rate of relapse compared to azathioprine. Although rituximab maintenance doses have not been directly compared, use of 1000 mg every 4 months in this trial was not found to be associated with a longer length of sustained remission after discontinuation.

The optimal duration of maintenance therapy is uncertain. With regard to glucocorticoids, it has been unclear whether maintaining patients on prednisone 5 mg/d has greater risks or benefits. Maintenance therapy with azathioprine, methotrexate, or mycophenolate mofetil is usually given for a minimum of 2 years. Because there is evidence that the risk of relapse is higher once maintenance medication has been stopped, the decision is individualized regarding whether to continue treatment or taper these agents over a 6- to 12-month period until discontinuation. Patients with significant organ damage or a history of relapse may benefit from longer-term maintenance therapy. Although rituximab has been found to have a lower relapse rate, studies have similarly shown that relapses occur once treatment is discontinued. Longer term use of rituximab can be associated with hypogammaglobulinemia such that the decision for how long to continue this agent beyond 2 years must be weighed in each patient.

REMISSION INDUCTION AND MAINTENANCE OF NONSEVERE DISEASE

For patients whose disease is not immediately organ- or life-threatening, methotrexate or mycophenolate mofetil together with glucocorticoids, often initiated at a lower dose, may be given to induce and then maintain remission. Rituximab can also be used in this setting weighing the risks and benefits. Treatment with cyclophosphamide is rarely, if ever, justified for the treatment of nonsevere granulomatosis with polyangiitis. There have been no published studies to date with the use of avacopan in nonsevere disease.

OTHER BIOLOGIC AGENTS

Etanercept (TNF inhibitor) was not found to sustain remission when used adjunctively to standard therapy and should not be used in granulomatosis with polyangiitis. Belimumab (anti-B lymphocyte stimulator) was examined as an adjunctive therapy to azathioprine for remission maintenance but showed no added benefit in reducing the risk of relapse. Abatacept (CTLA4-Ig) was examined in a randomized trial in patients with nonsevere relapsing disease where it did not reduce the risk of relapse, disease worsening, or failure to achieve remission.

ORGAN-SPECIFIC TREATMENT

Not all manifestations of granulomatosis with polyangiitis require or respond to immunosuppressive therapy, and differentiation of active disease from damage is necessary. As sinus disease can disrupt the mucociliary barrier, patients should be instructed on the use of local care with moisturization and humidification. Although certain reports have indicated that TMP-SMX may be of benefit in the treatment of granulomatosis with polyangiitis isolated to the sinonasal tissues, it should never be used alone to treat active granulomatosis with polyangiitis involving other organs. Subglottic stenosis can often scar and responds optimally to nonmedical intervention with dilation and glucocorticoid injection.

MICROSCOPIC POLYANGIITIS

DEFINITION

The term *microscopic polyarteritis* was introduced into the literature by Davson in 1948 in recognition of the presence of glomerulonephritis in patients with polyarteritis nodosa. In 1992, the Chapel Hill Consensus Conference on the Nomenclature of Systemic Vasculitis adopted the term *microscopic polyangiitis* to connote a necrotizing vasculitis with few or no immune complexes affecting small vessels (capillaries, venules, or arterioles). Glomerulonephritis is very common in microscopic polyangiitis, and pulmonary capillaritis often occurs. The absence of granulomatous inflammation in microscopic polyangiitis is said to differentiate it from granulomatosis with polyangiitis.

INCIDENCE AND PREVALENCE

The incidence of microscopic polyangiitis is estimated to be 3–5/100,000. The mean age of onset is ~57 years, and males are slightly more frequently affected than females.

PATHOLOGY AND PATHOGENESIS

Microscopic polyangiitis has a predilection to involve capillaries and venules in addition to small- and medium-sized arteries. Immunohistochemical staining reveals a paucity of immunoglobulin deposition in the vascular lesion of microscopic polyangiitis, suggesting that immune-complex formation does not play a role in the pathogenesis of this syndrome. The renal lesion seen in microscopic polyangiitis is histologically identical to that of granulomatosis with polyangiitis. Like granulomatosis with polyangiitis, microscopic polyangiitis is highly associated with ANCA, which may play a role in pathogenesis of this syndrome (see above).

CLINICAL AND LABORATORY MANIFESTATIONS

Because of its predilection to involve the small vessels, microscopic polyangiitis and granulomatosis with polyangiitis share similar clinical features. Disease onset may be gradual, with initial symptoms of fever, weight loss, and musculoskeletal pain; however, it is often acute. Glomerulonephritis occurs in at least 79% of patients and can be rapidly progressive, leading to renal failure. Hemoptysis may be the first symptom of alveolar hemorrhage, which occurs in 12% of patients. Other manifestations include mononeuritis multiplex and gastrointestinal tract and cutaneous vasculitis. Upper airway disease and pulmonary nodules are not typically found in microscopic polyangiitis and, if present, suggest granulomatosis with polyangiitis.

Features of inflammation may be seen, including an elevated ESR and/or CRP, anemia, leukocytosis, and thrombocytosis. ANCA are present in 75% of patients with microscopic polyangiitis, with antimyeloperoxidase antibodies being the predominant antigen association.

DIAGNOSIS

The diagnosis is based on histologic evidence of vasculitis or pauci-immune glomerulonephritis in a patient with compatible clinical features of multisystem disease. Although microscopic polyangiitis is strongly ANCA-associated, tissue biopsy should continue to be pursued in patients who do not have a clinically compatible picture.

TREATMENT

Microscopic Polyangiitis

The 5-year survival rate for patients with treated microscopic polyangiitis is 74%, with disease-related mortality occurring from alveolar hemorrhage or gastrointestinal, cardiac, or renal disease. Studies on treatment have come from trials that have included patients with granulomatosis with polyangiitis or microscopic polyangiitis. Currently, the treatment approach for microscopic polyangiitis is the same as is used for granulomatosis with polyangiitis (see "Granulomatosis with Polyangiitis" for a detailed description of this therapeutic regimen). Disease relapse has been observed in at least 34% of patients. Treatment for such relapses would be based on site and severity of disease.

EOSINOPHILIC GRANULOMATOSIS WITH POLYANGIITIS (CHURG-STRAUSS)

DEFINITION

Eosinophilic granulomatosis with polyangiitis was described in 1951 by Churg and Strauss and is characterized by asthma, peripheral and tissue eosinophilia, extravascular granuloma formation, and vasculitis of multiple organ systems.

INCIDENCE AND PREVALENCE

Eosinophilic granulomatosis with polyangiitis is an uncommon disease with an estimated annual incidence of 1–3 per million. The disease can occur at any age with the possible exception of infants. The mean age of onset is 48 years, with a female-to-male ratio of 1.2:1.

PATHOLOGY AND PATHOGENESIS

The necrotizing vasculitis of eosinophilic granulomatosis with polyangiitis involves small- and medium-sized muscular arteries, capillaries, veins, and venules. A characteristic histopathologic feature of eosinophilic granulomatosis with polyangiitis is granuloma that may be present in the tissues or even within the walls of the vessels themselves. These are usually associated with infiltration of the tissues with eosinophils. This process can occur in any organ in the body; lung involvement is predominant, with skin, cardiovascular system, kidney, peripheral nervous system, and gastrointestinal tract also commonly involved. Although the precise pathogenesis of this disease is uncertain, its strong association with asthma and its clinicopathologic manifestations, including eosinophilia, granuloma, and vasculitis, point to aberrant immunologic phenomena.

CLINICAL AND LABORATORY MANIFESTATIONS

Patients with eosinophilic granulomatosis with polyangiitis often exhibit nonspecific manifestations such as fever, malaise, anorexia, and weight loss, which are characteristic of a multisystem disease. The pulmonary findings in eosinophilic granulomatosis with polyangiitis dominate the clinical picture with asthmatic attacks and the presence of pulmonary infiltrates. Mononeuritis multiplex is the second most common manifestation and occurs in up to 72% of patients. Allergic rhinitis and sinusitis develop in up to 61% of patients and are often observed early in the course of disease. Clinically recognizable heart disease with myocarditis, pericarditis, endocarditis, or coronary vasculitis occurs in ~14% of patients and is an important cause of mortality. Skin lesions occur in ~51% of patients and include purpura in addition to cutaneous and subcutaneous nodules. Renal disease in eosinophilic granulomatosis with polyangiitis is less common than that of granulomatosis with polyangiitis and microscopic polyangiitis.

The characteristic laboratory finding in virtually all patients with eosinophilic granulomatosis with polyangiitis is a striking eosinophilia, which reaches levels >1000 cells/μL in >80% of patients. Evidence of inflammation as evidenced by elevated ESR and/or CRP, fibrinogen, or α_2-globulins can be found in 81% of patients. The other laboratory findings reflect the organ systems involved. Approximately 48% of patients with eosinophilic granulomatosis with polyangiitis have circulating ANCA that is usually antimyeloperoxidase.

DIAGNOSIS

Although the diagnosis of eosinophilic granulomatosis with polyangiitis is optimally made by biopsy in a patient with the characteristic clinical manifestations (see above), histologic confirmation can be challenging because the pathognomonic features often do not occur simultaneously. In order to be diagnosed with eosinophilic granulomatosis with polyangiitis, a patient should have evidence of asthma, peripheral blood eosinophilia, and clinical features consistent with vasculitis.

TREATMENT

Eosinophilic Granulomatosis with Polyangiitis

The prognosis of untreated eosinophilic granulomatosis with polyangiitis is poor, with a reported 5-year survival of 25%. With treatment, prognosis is favorable, with one study finding a 78-month actuarial survival rate of 72%. Myocardial involvement is the most frequent cause of death and is responsible for 39% of patient mortality. Echocardiography should be performed in all newly diagnosed patients because this may influence therapeutic decisions.

Glucocorticoids alone appear to be effective in many patients. Dosage tapering is often limited by asthma, with asthma often persisting after clinical recovery from vasculitis. In patients who present with fulminant multisystem disease, particularly cardiac involvement, the treatment of choice is daily cyclophosphamide and prednisone followed by azathioprine or methotrexate (see "Granulomatosis with Polyangiitis" for a detailed description of this therapeutic regimen). Use of rituximab and prednisone can be considered, particularly in ANCA-positive patients who have glomerulonephritis or other features of severe small-vessel vasculitis.

Mepolizumab (anti-IL-5 antibody) 300 mg given subcutaneously once a month was studied in a randomized trial and found to be more effective than placebo, with a higher number of weeks in remission (28% vs 3%) and a higher proportion of patients in remission (32% vs 3%). Patients with life-threatening eosinophilic granulomatosis with polyangiitis were excluded from this trial, and mepolizumab should not be used for the treatment of severe disease. Mepolizumab may have a particularly beneficial role in the setting of relapsing or resistant asthma requiring glucocorticoids, sinus disease, mild to moderate eosinophilic disease features, or mild vasculitis. In a recent study, benralizumab, which also inhibits IL-5, demonstrated noninferior rates of remission compared to mepolizumab in patients with relapsing or refractory, non-life-threatening eosinophilic granulomatosis with polyangiitis.

POLYARTERITIS NODOSA

DEFINITION

Polyarteritis nodosa was described in 1866 by Kussmaul and Maier. It is a multisystem, necrotizing vasculitis of small- and medium-sized muscular arteries characteristically involving the renal and visceral arteries. Polyarteritis nodosa does not involve pulmonary arteries, although bronchial vessels may be involved; glomerulonephritis, granuloma, and significant eosinophilia are not seen.

INCIDENCE AND PREVALENCE

It is difficult to establish an accurate incidence of polyarteritis nodosa because previous reports have included microscopic polyangiitis as well as other related vasculitides. Polyarteritis nodosa, as currently defined, is felt to be a very uncommon disease.

PATHOLOGY AND PATHOGENESIS

The vascular lesion in polyarteritis nodosa is a necrotizing inflammation of small- and medium-sized muscular arteries. The lesions are segmental and tend to involve bifurcations and branchings of arteries. They may spread circumferentially to involve adjacent veins. However, involvement of venules is not seen in polyarteritis nodosa and, if present, suggests microscopic polyangiitis (see below). In the acute stages of disease, polymorphonuclear neutrophils infiltrate all layers of the vessel wall and perivascular areas, which results in intimal proliferation and degeneration of the vessel wall. Mononuclear cells infiltrate the area as the lesions progress to the subacute and chronic stages. Fibrinoid necrosis of the vessels ensues with compromise of the lumen, thrombosis, infarction of the tissues supplied by the involved vessel, and, in some cases, hemorrhage. As the lesions heal, there is collagen deposition, which may lead to further occlusion of the vessel lumen. Aneurysmal dilations up to 1 cm in size along the involved arteries are characteristic.

Multiple organ systems are involved, and the clinicopathologic findings reflect the degree and location of vessel involvement and the resulting ischemic changes. As mentioned above, pulmonary arteries are not involved, and bronchial artery involvement is uncommon. The pathology in the kidney is that of arteritis without glomerulonephritis. In patients with significant hypertension, typical pathologic features of glomerulosclerosis may be seen. In addition, pathologic sequelae of hypertension may be found elsewhere in the body.

The presence of a polyarteritis nodosa–like vasculitis in patients with hepatitis B together with the isolation of circulating immune complexes composed of hepatitis B antigen and immunoglobulin and the demonstration by immunofluorescence of hepatitis B antigen, IgM, and complement in the blood vessel walls strongly suggest the role of immunologic phenomena in the pathogenesis of this disease. A polyarteritis nodosa–like vasculitis has also been reported in patients with hepatitis C. Hairy cell leukemia can be associated with polyarteritis nodosa; the pathogenic mechanisms of this association are unclear.

A polyarteritis nodosa–like vasculitis has been described in conjunction with deficiency of adenosine deaminase type 2 (DADA2). Patients with DADA2 usually present in childhood with a variable pattern of clinical features and vascular pathology that is responsive to TNF inhibitors. As this differs from the usual treatment for polyarteritis nodosa, DADA2 should be considered in patients with suggestive clinical features, particularly those with early-onset disease.

TABLE 375-6 Clinical Manifestations Related to Organ System Involvement in Polyarteritis Nodosa

ORGAN SYSTEM	PERCENT INCIDENCE	CLINICAL MANIFESTATIONS
Renal	60	Renal failure, hypertension
Musculoskeletal	64	Arthritis, arthralgia, myalgia
Peripheral nervous system	51	Peripheral neuropathy, mononeuritis multiplex
Gastrointestinal tract	44	Abdominal pain, nausea and vomiting, bleeding, bowel infarction and perforation, cholecystitis, hepatic infarction, pancreatic infarction
Skin	43	Rash, purpura, nodules, cutaneous infarcts, livedo reticularis, Raynaud's phenomenon
Cardiac	36	Congestive heart failure, myocardial infarction, pericarditis
Genitourinary	25	Testicular, ovarian, or epididymal pain
Central nervous system	23	Cerebral vascular accident, altered mental status, seizure

Source: Reproduced with permission from TR Cupps, AS Fauci: The vasculitides. Major Probl Intern Med 21:1, 1981.

CLINICAL AND LABORATORY MANIFESTATIONS

Nonspecific signs and symptoms are the hallmarks of polyarteritis nodosa. Fever, weight loss, and malaise are present in over one-half of cases. Patients usually present with vague symptoms such as weakness, malaise, headache, abdominal pain, and myalgias that can rapidly progress to a fulminant illness. Specific complaints related to the vascular involvement within a particular organ system may also dominate the presenting clinical picture as well as the entire course of the illness **(Table 375-6)**. Renal involvement most commonly manifests as hypertension, renal insufficiency, or hemorrhage due to microaneurysms.

There are no diagnostic serologic tests for polyarteritis nodosa. In >75% of patients, the leukocyte count is elevated with a predominance of neutrophils. Eosinophilia is rare and, when present at high levels, suggests the diagnosis of eosinophilic granulomatosis with polyangiitis. The anemia of chronic disease may be seen, and an elevated ESR and/or CRP is almost always present. Other laboratory findings reflect the particular organ involved. Hypergammaglobulinemia may be present, and all patients should be screened for hepatitis B and C. ANCA are rarely found.

DIAGNOSIS

The diagnosis of polyarteritis nodosa is based on the demonstration of characteristic findings of vasculitis on biopsy material of involved organs. Biopsy of symptomatic organs such as nodular skin lesions, painful testes, and nerve/muscle provides the highest diagnostic yields. In the absence of easily accessible tissue for biopsy, the arteriographic demonstration of involved vessels, particularly in the form of aneurysms of small- and medium-sized arteries in the renal, hepatic, and visceral vasculature, is sufficient to make the diagnosis. This should consist of a catheter-directed dye arteriogram because magnetic resonance and computed tomography arteriograms do not have sufficient resolution at the current time to visualize the affected vessels. Aneurysms of vessels are not pathognomonic of polyarteritis nodosa; furthermore, aneurysms may not always be present, and arteriographic findings may be limited to stenotic segments and obliteration of vessels.

TREATMENT

Polyarteritis Nodosa

The prognosis of untreated polyarteritis nodosa is extremely poor, with a reported 5-year survival rate between 10 and 20%. Death usually results from gastrointestinal complications, particularly bowel infarcts and perforation, and cardiovascular causes. Intractable hypertension often compounds dysfunction in other organ systems, such as the kidneys, heart, and CNS, leading to additional late morbidity and mortality. The combination of prednisone and cyclophosphamide has been found to significantly improve the survival rate (see "Granulomatosis with Polyangiitis" for a description of this therapeutic regimen). In less severe cases, glucocorticoids alone have resulted in disease remission. Treatment of hypertension can lessen vascular complications. Following remission, relapse has been estimated to occur in 10–20% of patients. In patients with hepatitis B or C who have a polyarteritis nodosa–like vasculitis, antiviral therapy is an essential part of management, with immunosuppression being given concurrently for severe vasculitis and then withdrawn once the vasculitis is controlled. Plasma exchange has been used in some series of hepatitis virus-associated polyarteritis nodosa.

GIANT CELL ARTERITIS AND POLYMYALGIA RHEUMATICA

DEFINITION

Giant cell arteritis, historically referred to as *temporal arteritis*, is an inflammation of medium- and large-sized arteries. It characteristically involves one or more branches of the carotid artery, particularly the temporal artery. However, it is a systemic disease that can involve arteries in multiple locations, particularly the aorta and its main branches.

Giant cell arteritis is closely associated with *polymyalgia rheumatica*, which is characterized by stiffness, aching, and pain in the muscles of the neck, shoulders, lower back, hips, and thighs. Most commonly, polymyalgia rheumatica occurs in isolation, but it may be seen in 40–50% of patients with giant cell arteritis. In addition, ~10–20% of patients who initially present with features of isolated polymyalgia rheumatica later go on to develop giant cell arteritis. A careful history and physical examination at initial and subsequent visits should be performed in patients with polymyalgia rheumatica to look for features suggestive of giant cell arteritis. This strong clinical association together with data from pathophysiologic studies has supported that giant cell arteritis and polymyalgia rheumatica represent differing clinical spectrums of a single disease process.

INCIDENCE AND PREVALENCE

Giant cell arteritis occurs almost exclusively in individuals aged >50 years. It is more common in women than in men and is rare in blacks. The incidence of giant cell arteritis varies widely in different studies and in different geographic regions. A high incidence has been found in Scandinavia and in regions of the United States with large Scandinavian populations, compared to a lower incidence in southern Europe. The annual incidence rates in individuals aged ≥50 years range from 6.9 to 32.8 per 100,000 population. Familial aggregation has been reported, as has an association with HLA-DR4. In addition, genetic linkage studies have demonstrated an association of giant cell arteritis with alleles at the HLA-DRB1 locus, particularly HLA-DRB1*04 variants. In Olmsted County, Minnesota, the annual incidence of polymyalgia rheumatica in individuals aged ≥50 years is 58.7 per 100,000 population.

PATHOLOGY AND PATHOGENESIS

Although the temporal artery is most frequently involved in giant cell arteritis, patients often have a systemic vasculitis of multiple medium- and large-sized arteries, which may go undetected. Histopathologically, the disease is a panarteritis with inflammatory mononuclear cell infiltrates within the vessel wall with frequent giant cell formation. There is proliferation of the intima and fragmentation of the internal elastic lamina. Pathophysiologic findings in organs result from the ischemia related to the involved vessels.

Experimental data support that giant cell arteritis is an antigen-driven disease in which activated T lymphocytes, macrophages, and dendritic cells play a critical role in pathogenesis. Sequence analysis of the T-cell receptor of tissue-infiltrating T cells in lesions of giant cell arteritis indicates restricted clonal expansion, suggesting the presence of an antigen residing in the arterial wall. Giant cell arteritis is believed to be initiated in the adventitia where CD4+ T cells enter through the vasa vasorum, become activated, and orchestrate macrophage differentiation. T cells recruited to vasculitic lesions in patients with giant cell arteritis produce predominantly IL-2 and IFN-γ, and the latter has been suggested to be involved in the progression to arteritis. Laboratory-based data demonstrate that at least two separate lineages of CD4 T cells—IFN-γ-producing T_H1 cells and IL-17-producing T_H17 cells—participate in vascular inflammation and may have differing levels of responsiveness to glucocorticoids.

CLINICAL AND LABORATORY MANIFESTATIONS

Giant cell arteritis is most commonly characterized clinically by the complex of fever, anemia, high ESR and/or CRP, and headaches in a patient aged >50 years. Four phenotypic manifestations can be seen including cranial arteritis, polymyalgia rheumatica, large-vessel disease, and features of systemic inflammation. These can occur together, in isolation, or at different times.

In patients with involvement of the cranial arteries, headache is the predominant symptom and may be associated with a tender, thickened, or nodular artery, which may pulsate early in the disease but may become occluded later. Scalp pain and claudication of the jaw and tongue may occur. A well-recognized and dreaded complication of giant cell arteritis, particularly in untreated patients, is ischemic optic neuropathy, which may lead to serious visual symptoms, including sudden blindness in some patients. However, most patients have complaints relating to the head or eyes before visual loss. Attention to such symptoms with institution of appropriate therapy (see below) lessens the risk of this complication. Other cranial ischemic complications include strokes and scalp or tongue infarction.

Up to one-third of patients can have large-vessel disease that can be the primary presentation of giant cell arteritis or can emerge at a later point in patients who have had previous cranial arteritis features or polymyalgia rheumatica. Manifestations of large-vessel disease can include subclavian artery stenosis that can present as arm claudication or aortic aneurysms involving the thoracic and to a lesser degree the abdominal aorta, which carry risks of rupture or dissection.

Clinical features of systemic inflammation can include malaise, fatigue, anorexia, weight loss, and sweats. Laboratory findings can also support an inflammatory process. An elevated ESR and/or CRP will be seen in most patients. Therapeutic blockade of IL-6 signaling prevents the hepatic production of CRP and other acute-phase proteins like ESR. These values will therefore normalize in patients who are treated with an IL-6 inhibitor, eliminating the ability to utilize CRP and ESR in monitoring disease activity or infection assessment in this setting. Other laboratory features include a normochromic or slightly hypochromic anemia. Liver function abnormalities are common, particularly increased alkaline phosphatase levels. Increased levels of IgG and complement have been reported.

DIAGNOSIS

The diagnosis of giant cell arteritis can often be suggested clinically by the demonstration of the complex of fever, anemia, and high ESR and/or CRP with or without symptoms of polymyalgia rheumatica in a patient >50 years old. The diagnosis can be confirmed by biopsy of the temporal artery but may not be positive in all patients due to patchy histologic findings. Since involvement of the vessel may be segmental, positive yield is increased by obtaining a biopsy segment of 3–5 cm together with serial sectioning of biopsy specimens. Ultrasonography of the temporal arteries alone or in combination with the axillary arteries demonstrating a noncompressible "halo" sign has been used in supporting the diagnosis. Therapy should not be delayed pending the performance of diagnostic studies. In this regard, it has been reported that temporal artery biopsies may show vasculitis even after ~14 days of glucocorticoid therapy. A dramatic clinical response to a trial of glucocorticoid therapy can further support the diagnosis.

Large-vessel disease may be suggested by symptoms and findings on physical examination such as diminished pulses or bruits. It is confirmed by vascular imaging, most commonly through magnetic resonance or computed tomography. Positron emission tomography has become increasingly investigated, although its role in diagnosis and monitoring remains unclear.

Isolated polymyalgia rheumatica is a clinical diagnosis made by the presence of typical symptoms of stiffness, aching, and pain in the muscles of the hip and shoulder girdle, an increased ESR and/or CRP, the absence of clinical features suggestive of giant cell arteritis, and a prompt therapeutic response to low-dose prednisone. Polymyalgia rheumatica can be associated with a peripheral arthritis that can mimic rheumatoid arthritis **(Chap. 370)**. Rheumatoid factor and anti–cyclic citrullinated peptide (CCP) should be negative. In patients who develop a worsening pattern of peripheral arthritis, the potential for a seronegative rheumatoid arthritis or other inflammatory arthropathy should be considered. Levels of enzymes indicative of muscle damage such as serum creatine kinase are not elevated.

TREATMENT

Giant Cell Arteritis

Acute disease–related mortality directly from giant cell arteritis is uncommon, with fatalities occurring from cerebrovascular events or myocardial infarction. However, patients are at risk of late mortality from aortic aneurysm rupture or dissection as patients with giant cell arteritis are 18 times more likely to develop thoracic aortic aneurysms than the general population.

The goals of treatment in giant cell arteritis are to reduce symptoms and, most importantly, to prevent visual loss. The treatment approach for cranial arteritis and large-vessel disease in giant cell arteritis is currently the same.

Giant cell arteritis and its associated symptoms are responsive to glucocorticoid therapy. Treatment should begin with prednisone 40–60 mg/d for ~1 month, followed by a gradual tapering. When ocular signs and symptoms occur, consideration should be given for the use of methylprednisolone 1000 mg daily for 3 days to protect remaining vision. Although the optimal duration of glucocorticoid therapy has not been established, most series have found that patients require treatment for ≥2 years. Symptom recurrence during prednisone tapering develops in 60–85% of patients with giant cell arteritis, requiring a dosage increase. For patients treated with prednisone alone, the ESR and/or CRP can serve as an indicator of inflammatory disease activity in monitoring and tapering therapy and can be used to guide the tapering schedule. However, minor increases in the ESR and/or CRP can occur as glucocorticoids are being tapered and do not necessarily reflect an exacerbation of arteritis, particularly if the patient remains symptom-free. Under these circumstances, the tapering should continue with caution. An increase in ESR and/or CRP can also occur for other reasons, including infection. Glucocorticoid toxicity occurs in 35–65% of patients and represents an important cause of patient morbidity.

Tocilizumab (anti-IL-6 receptor) 162 mg given subcutaneously once every week or once every other week combined with a 26-week prednisone taper was found to be superior to prednisone alone in sustaining glucocorticoid-free remission in a randomized trial. The decision about when to use tocilizumab in giant cell arteritis is individually based, taking into account comorbidities, glucocorticoid toxicity, and the side effects of tocilizumab. Pharmacokinetic studies supported that tocilizumab 6 mg/kg given intravenously every 4 weeks could also be used as a treatment option in giant cell arteritis.

The use of methotrexate as a glucocorticoid-sparing agent was examined in two randomized placebo-controlled trials that reached conflicting conclusions. It may be considered in select patients with glucocorticoid toxicity who are unable to take or are intolerant of tocilizumab.

Abatacept (CTLA4-Ig) was examined in a small randomized trial in giant cell arteritis and demonstrated greater efficacy than glucocorticoids alone. Infliximab, a monoclonal antibody to TNF, was studied in a randomized trial and was not found to provide benefit.

Aspirin 81 mg daily has been found in retrospective studies to reduce the occurrence of cranial ischemic complications in giant cell arteritis, and adjunctive use should be considered in patients who do not have contraindications.

Polymyalgia Rheumatica

Patients with isolated polymyalgia rheumatica respond promptly to prednisone given at a dose of 10–20 mg/d. Similar to giant cell arteritis, the ESR and/or CRP can serve as a useful indicator in monitoring prednisone reduction in those treated with glucocorticoids alone. Recurrent polymyalgia symptoms develop in the majority of patients during prednisone tapering.

Sarilumab (anti-IL-6 receptor) was found to be effective in a randomized trial of patients with glucocorticoid-resistant polymyalgia rheumatica who relapsed on prednisone ≥7.5 mg/d. In this trial, sarilumab 200 mg given subcutaneously every 2 weeks plus a 14-week glucocorticoid taper was associated with a higher rate of sustained remission compared to glucocorticoids alone. As most patients with polymyalgia rheumatica can be successfully treated with glucocorticoids alone, use of sarilumab should be restricted to those who have had an inadequate response to glucocorticoids or who relapse with a glucocorticoid taper.

One study of methotrexate found that this reduced the prednisone dose on average by only 1 mg and did not decrease prednisone-related side effects. A randomized trial in polymyalgia rheumatica did not find infliximab to lessen relapse or glucocorticoid requirements.

TAKAYASU ARTERITIS

DEFINITION

Takayasu arteritis is an inflammatory disease of large-sized arteries characterized by a strong predilection for the aortic arch and its branches.

INCIDENCE AND PREVALENCE

Takayasu arteritis is an uncommon disease with an estimated annual incidence rate of 1.2–2.6 cases per million. It is most prevalent in adolescent girls and young women. Although it is more common in Asia, it is neither racially nor geographically restricted.

PATHOLOGY AND PATHOGENESIS

The disease involves large-sized arteries, with a strong predilection for the aortic arch and its branches; the pulmonary artery may also be involved. The most commonly affected arteries seen by arteriography are listed in **Table 375-7**. The involvement of the major branches of

TABLE 375-7 Frequency of Arteriographic Abnormalities and Potential Clinical Manifestations of Arterial Involvement in Takayasu Arteritis

ARTERY	PERCENTAGE OF ARTERIOGRAPHIC ABNORMALITIES	POTENTIAL CLINICAL MANIFESTATIONS
Subclavian	93	Arm claudication, Raynaud's phenomenon
Common carotid	58	Visual changes, syncope, transient ischemic attacks, stroke
Abdominal aorta[a]	47	Abdominal pain, nausea, vomiting
Renal	38	Hypertension, renal failure
Aortic arch or root	35	Aortic insufficiency, congestive heart failure
Vertebral	35	Visual changes, dizziness
Coeliac axis[a]	18	Abdominal pain, nausea, vomiting
Superior mesenteric[a]	18	Abdominal pain, nausea, vomiting
Iliac	17	Leg claudication
Pulmonary	10–40	Atypical chest pain, dyspnea
Coronary	<10	Chest pain, myocardial infarction

[a] Arteriographic lesions at these locations are usually asymptomatic but may potentially cause these symptoms.

Source: G Kerr et al: Ann Intern Med 120:919, 1994.

the aorta is much more marked at their origin than distally. The disease is a panarteritis with inflammatory mononuclear cell infiltrates and occasionally giant cells. There are marked intimal proliferation and fibrosis, scarring and vascularization of the media, and disruption and degeneration of the elastic lamina. The vasa vasorum are frequently involved. Narrowing of the lumen occurs with or without thrombosis. Pathologic changes in various organs typically reflect reduced blood flow through these vascular stenoses. Aneurysms can also occur due to weakening of the vessel wall.

Immunopathogenic mechanisms, the precise nature of which is uncertain, are suspected in this disease. As with several of the vasculitis syndromes, circulating immune complexes have been demonstrated, but their pathogenic significance is unclear.

■ CLINICAL AND LABORATORY MANIFESTATIONS

Takayasu arteritis is a systemic disease with generalized as well as vascular symptoms. The generalized symptoms include malaise, fever, night sweats, arthralgias, anorexia, and weight loss, which may occur months before vessel involvement is apparent. These symptoms may merge into those related to vascular compromise and organ ischemia. Pulses are commonly absent in the involved vessels, particularly the subclavian artery. The frequency of arteriographic abnormalities and the potentially associated clinical manifestations are listed in Table 375-7. Hypertension occurs in 32–93% of patients and contributes to renal, cardiac, and cerebral injury.

Characteristic laboratory findings include an elevated ESR and/or CRP, mild anemia, and elevated immunoglobulin levels.

■ DIAGNOSIS

The diagnosis of Takayasu arteritis should be suspected strongly in a young woman who develops a decrease or absence of peripheral pulses, a discrepancy in blood pressure between arms, or arterial bruits. The diagnosis is confirmed by the characteristic pattern on arteriography, which includes irregular vessel walls, stenosis, poststenotic dilation, aneurysm formation, occlusion, and evidence of increased collateral circulation. Complete imaging of the aorta and its major branches by magnetic resonance or computed tomography arteriography should be obtained to fully delineate the distribution and degree of arterial disease. Because of the involvement of the large vessels, tissue is rarely available or necessary as a means of diagnosis and obtained only if vascular surgery is necessary. IgG4-related disease **(Chap. 380)** is a potential cause of aortitis and periaortitis that is histologically differentiated from Takayasu arteritis by a dense lymphoplasmacytic infiltrate rich in IgG4-positive plasma cells, a storiform pattern of fibrosis, and obliterative phlebitis.

TREATMENT

Takayasu Arteritis

The long-term outcome of patients with Takayasu arteritis has varied widely between studies. Although two North American reports found overall survival to be ≥94%, the 5-year mortality rate from other studies has ranged from 0 to 35%. Disease-related mortality most often occurs from congestive heart failure, cerebrovascular events, myocardial infarction, aneurysm rupture, or renal failure. Even in the absence of life-threatening disease, Takayasu arteritis can be associated with significant morbidity. The course of the disease is variable, and although spontaneous remissions may occur, Takayasu arteritis is most often chronic and relapsing. Although glucocorticoids, given as prednisone 40–60 mg/d followed by a taper, alleviate symptoms, there are no convincing studies that indicate that it increases survival. The combination of glucocorticoid therapy for acute signs and symptoms and an aggressive surgical and/or arterioplastic approach to stenosed vessels has markedly improved outcome and decreased morbidity by lessening the risk of stroke, correcting hypertension due to renal artery stenosis, and improving blood flow to ischemic viscera and limbs. Unless it is urgently required, surgical correction of stenosed arteries should be undertaken only when the vascular inflammatory process is well controlled with medical therapy.

In individuals who are refractory to or unable to taper glucocorticoids, methotrexate in doses up to 25 mg per week has been used with positive results. Retrospective series with anti-TNF therapies have shown benefit, but these agents have not been studied through randomized trials to determine efficacy.

Tocilizumab has been investigated in a randomized trial where it did not reach its primary efficacy endpoint, although it was found to be associated with glucocorticoid-sparing effects and improvements in well-being. Encouraging results have also been seen in retrospective studies such that the utility of this agent remains an active question. Abatacept was examined in a randomized trial in Takayasu arteritis but did not demonstrate efficacy beyond glucocorticoids alone.

IGA VASCULITIS (HENOCH-SCHÖNLEIN)

■ DEFINITION

IgA vasculitis (Henoch-Schönlein) is a small-vessel vasculitis characterized by palpable purpura (most commonly distributed over the buttocks and lower extremities), arthralgias, gastrointestinal signs and symptoms, and glomerulonephritis.

■ INCIDENCE AND PREVALENCE

IgA vasculitis is usually seen in children ages 4–7 years; however, the disease may also be seen in infants and adults. It is not a rare disease; in one series, it accounted for between 5 and 24 admissions per year at a pediatric hospital. The male-to-female ratio is 1.5:1. A seasonal variation with a peak incidence in spring has been noted.

■ PATHOLOGY AND PATHOGENESIS

The presumptive pathogenic mechanism for IgA vasculitis is immune-complex deposition. A number of inciting antigens have been suggested including upper respiratory tract infections, various drugs, foods, insect bites, and immunizations. IgA is the antibody class most often seen in the immune complexes and has been demonstrated in the renal biopsies of these patients.

■ CLINICAL AND LABORATORY MANIFESTATIONS

In pediatric patients, palpable purpura is seen in virtually all patients; most patients develop polyarthralgias in the absence of frank arthritis. Gastrointestinal involvement, which is seen in almost 70% of pediatric patients, is characterized by colicky abdominal pain usually associated with nausea, vomiting, diarrhea, or constipation, and is frequently accompanied by the passage of blood and mucus per rectum; bowel intussusception may occur. Renal involvement occurs in 10–50% of patients and is usually characterized by mild glomerulonephritis leading to proteinuria and microscopic hematuria, with red blood cell casts in the majority of patients; it usually resolves spontaneously without therapy. Progressive glomerulonephritis rarely develops. In adults, presenting symptoms are most frequently related to the skin and joints, while initial complaints related to the gut are less common. Although certain studies have found that renal disease is more frequent and more severe in adults, this has not been a consistent finding. However, the course of renal disease in adults may be more insidious and thus requires close follow-up. Myocardial involvement can occur in adults but is rare in children.

Laboratory studies generally show a mild leukocytosis, a normal platelet count, and occasionally eosinophilia. Serum complement components are normal, and IgA levels are elevated in about one-half of patients.

■ DIAGNOSIS

The diagnosis of IgA vasculitis is based on clinical signs and symptoms. Skin biopsy specimen can be useful in confirming leukocytoclastic vasculitis with IgA and C3 deposition by immunofluorescence. Renal biopsy is rarely needed for diagnosis but may provide prognostic information in some patients.

TREATMENT
IgA Vasculitis

The prognosis of IgA vasculitis is excellent. Mortality is exceedingly rare, and 1–5% of children progress to end-stage renal disease, although renal failure may be more common in adults. Most patients recover completely, and some do not require therapy. When glucocorticoids are required, prednisone, 1 mg/kg per day and tapered according to clinical response, has been shown to be useful in decreasing tissue edema, arthralgias, and abdominal discomfort; however, it has not proved beneficial in the treatment of skin or renal disease and does not appear to shorten the duration of active disease or lessen the chance of recurrence. Patients with rapidly progressive glomerulonephritis have been anecdotally reported to benefit from glucocorticoids used in combination with another immunosuppressive agent. Disease recurrences have been reported in 10–40% of patients.

CRYOGLOBULINEMIC VASCULITIS

DEFINITION
Cryoglobulins are cold-precipitable monoclonal or polyclonal immunoglobulins. Cryoglobulinemia may be associated with a systemic vasculitis characterized by palpable purpura, arthralgias, weakness, neuropathy, and glomerulonephritis. The most common association has been with hepatitis C, although cryoglobulinemia can be observed in association with a variety of underlying disorders including multiple myeloma, lymphoproliferative disorders, connective tissue diseases, infection, and liver disease and can be idiopathic.

INCIDENCE AND PREVALENCE
The incidence of cryoglobulinemic vasculitis has not been established. It has been estimated that 5% of patients with chronic hepatitis C will develop cryoglobulinemic vasculitis.

PATHOLOGY AND PATHOGENESIS
Skin biopsies in cryoglobulinemic vasculitis reveal an inflammatory infiltrate surrounding and involving blood vessel walls, with fibrinoid necrosis, endothelial cell hyperplasia, and hemorrhage. Deposition of immunoglobulin and complement is common. Abnormalities of uninvolved skin including basement membrane alterations and deposits in vessel walls may be found. Membranoproliferative glomerulonephritis is responsible for 80% of all renal lesions in cryoglobulinemic vasculitis.

The association between hepatitis C and cryoglobulinemic vasculitis has been supported by the high frequency of documented hepatitis C infection, the presence of hepatitis C RNA and anti–hepatitis C antibodies in serum cryoprecipitates, evidence of hepatitis C antigens in vasculitic skin lesions, and the effectiveness of antiviral therapy. Current evidence suggests that in the majority of cases, cryoglobulinemic vasculitis occurs when an aberrant immune response to hepatitis C infection leads to the formation of immune complexes consisting of hepatitis C antigens, polyclonal hepatitis C–specific IgG, and monoclonal IgM rheumatoid factor. The deposition of these immune complexes in blood vessel walls triggers an inflammatory cascade that results in cryoglobulinemic vasculitis.

CLINICAL AND LABORATORY MANIFESTATIONS
The most common clinical manifestations of cryoglobulinemic vasculitis are cutaneous vasculitis, arthritis, peripheral neuropathy, and glomerulonephritis. Renal disease develops in 10–30% of patients. Life-threatening rapidly progressive glomerulonephritis or vasculitis of the CNS, gastrointestinal tract, or heart occurs infrequently.

The presence of circulating cryoprecipitates is the fundamental finding in cryoglobulinemic vasculitis. Rheumatoid factor is almost always found and may be a useful clue to the disease when cryoglobulins are not detected. Hypocomplementemia occurs in 90% of patients. An elevated ESR and/or CRP and anemia occur frequently. Evidence for hepatitis C infection must be sought in all patients by testing for hepatitis C antibodies and hepatitis C RNA.

TREATMENT
Cryoglobulinemic Vasculitis

Acute mortality directly from cryoglobulinemic vasculitis is uncommon, but the presence of glomerulonephritis is a poor prognostic sign for overall outcome. In such patients, 15% progress to end-stage renal disease, with 40% later experiencing fatal cardiovascular disease, infection, or liver failure. As indicated above, the majority of cases are associated with hepatitis C infection. In such patients, treatment with antiviral therapy **(Chap. 352)** is first-line therapy for hepatitis C–associated cryoglobulinemic vasculitis, particularly given the efficacy of current hepatitis C therapies. Clinical improvement with antiviral therapy is dependent on the virologic response. Patients who clear hepatitis C from the blood have objective improvement in their vasculitis along with significant reductions in levels of circulating cryoglobulins, IgM, and rheumatoid factor. While transient improvement can be observed with glucocorticoids, a complete response is seen in only 7% of patients. Plasma exchange and cytotoxic agents have been used in anecdotal reports. These observations have not been confirmed, and such therapies carry significant risks. Randomized trials with rituximab in hepatitis C–associated cryoglobulinemic vasculitis showed evidence of benefit such that this should be considered in patients with active vasculitis either in combination with antiviral therapy or alone in patients who have relapsed through, are intolerant to, or have contraindications to antiviral agents.

SINGLE-ORGAN VASCULITIS
Single-organ vasculitis has been defined as vasculitis in arteries or veins of any size in a single organ that has no features that indicate that it is a limited expression of a systemic vasculitis. Examples include isolated aortitis, testicular vasculitis, vasculitis of the breast, isolated cutaneous vasculitis, and primary CNS vasculitis. In some instances, this may be discovered at the time of surgery such as orchiectomy for a testicular mass where there is concern for neoplasm that is found instead to be vasculitis. Some patients originally diagnosed with single-organ vasculitis may later develop additional manifestations of a more systemic disease. In instances where there is no evidence of systemic vasculitis and the affected organ has been removed in its entirety, the patient may be followed closely without immunosuppressive therapy. In other instances, such as primary CNS vasculitis, medical intervention is necessary.

IDIOPATHIC CUTANEOUS VASCULITIS

DEFINITION
The term *cutaneous vasculitis* is defined broadly as inflammation of the blood vessels of the dermis. Because of its heterogeneity, cutaneous vasculitis has been described by a variety of terms including *hypersensitivity vasculitis* and *cutaneous leukocytoclastic angiitis*. However, cutaneous vasculitis is not one specific disease but a manifestation that can be seen in a variety of settings. In >70% of cases, cutaneous vasculitis occurs either as part of a primary systemic vasculitis or as a secondary vasculitis related to an inciting agent or an underlying disease (see "Secondary Vasculitis," below). In the remaining 30% of cases, cutaneous vasculitis occurs idiopathically.

INCIDENCE AND PREVALENCE
Cutaneous vasculitis represents the most commonly encountered vasculitis in clinical practice. The exact incidence of idiopathic cutaneous vasculitis has not been determined due to the predilection for cutaneous vasculitis to be associated with an underlying process and the variability of its clinical course.

PATHOLOGY AND PATHOGENESIS
The typical histopathologic feature of cutaneous vasculitis is the presence of vasculitis of small vessels. Postcapillary venules are the most commonly involved vessels; capillaries and arterioles may be involved less frequently. This vasculitis is characterized by a *leukocytoclasis*, a

term that refers to the nuclear debris remaining from the neutrophils that have infiltrated in and around the vessels during the acute stages. In the subacute or chronic stages, mononuclear cells predominate; in certain subgroups, eosinophilic infiltration is seen. Erythrocytes often extravasate from the involved vessels, leading to palpable purpura. *Cutaneous arteritis* can also occur, which involves slightly larger-sized vessels within the dermis.

■ CLINICAL AND LABORATORY MANIFESTATIONS

The hallmark of idiopathic cutaneous vasculitis is the predominance of skin involvement. Skin lesions may appear typically as palpable purpura; however, other cutaneous manifestations of the vasculitis may occur, including macules, papules, vesicles, bullae, subcutaneous nodules, ulcers, and recurrent or chronic urticaria. The skin lesions may be pruritic or painful, with a burning or stinging sensation. Lesions most commonly occur in the lower extremities in ambulatory patients or in the sacral area in bedridden patients due to the effects of hydrostatic forces on the postcapillary venules. Edema may accompany certain lesions, and hyperpigmentation often occurs in areas of recurrent or chronic lesions.

There are no specific laboratory tests diagnostic of idiopathic cutaneous vasculitis. A mild leukocytosis with or without eosinophilia is characteristic, as is an elevated ESR and/or CRP. Laboratory studies should be aimed toward ruling out features to suggest an underlying disease or a systemic vasculitis.

■ DIAGNOSIS

The diagnosis of cutaneous vasculitis is made by the demonstration of vasculitis on biopsy. An important diagnostic principle in patients with cutaneous vasculitis is to search for an etiology of the vasculitis—be it an exogenous agent, such as a drug or an infection, or an endogenous underlying disease (Fig. 375-1). In addition, a careful history, physical, and laboratory examination should be performed to rule out the possibility of coexisting systemic vasculitis.

TREATMENT

Idiopathic Cutaneous Vasculitis

When an antigenic stimulus is recognized as the precipitating factor in the cutaneous vasculitis, it should be removed; if this is a microbe, appropriate antimicrobial therapy should be instituted. If the vasculitis is associated with another underlying disease, treatment of the latter often results in resolution of the former. In situations where disease is apparently self-limited, no therapy, except possibly symptomatic therapy, is indicated. When cutaneous vasculitis persists and when there is no evidence of an inciting agent, an associated disease, or an underlying systemic vasculitis, the decision to treat should be based on weighing the balance between the degree of symptoms and the risk of treatment. Some cases of idiopathic cutaneous vasculitis resolve spontaneously, whereas others remit and relapse. In patients with persistent vasculitis, a variety of therapeutic regimens have been tried with variable results. In general, the treatment of idiopathic cutaneous vasculitis has not been satisfactory. Fortunately, since the disease is generally limited to the skin, this lack of consistent response to therapy usually does not lead to a life-threatening situation. Agents with which there have been anecdotal reports of success include dapsone, colchicine, hydroxychloroquine, and nonsteroidal anti-inflammatory agents. Glucocorticoids are often used in the treatment of idiopathic cutaneous vasculitis. Therapy is usually instituted with the lowest effective dose of prednisone with rapid tapering to discontinuation. In cases that prove refractory to glucocorticoids, a trial of another immunosuppressive agent may be indicated. Patients with chronic vasculitis isolated to cutaneous venules rarely respond dramatically to any therapeutic regimen, and other immunosuppressive agents should be used only as a last resort in these patients. Methotrexate and azathioprine have been used in such situations in anecdotal reports. Cyclophosphamide should almost never be used for idiopathic cutaneous vasculitis because of the potential toxicity.

PRIMARY CENTRAL NERVOUS SYSTEM VASCULITIS

Primary CNS vasculitis is an uncommon clinicopathologic entity characterized by vasculitis restricted to the vessels of the CNS without other apparent systemic vasculitis. The inflammatory process is usually composed of mononuclear cell infiltrates with or without granuloma formation.

Patients may present with headaches, altered mental function, and focal neurologic defects. Systemic symptoms are generally absent. Devastating neurologic abnormalities may occur depending on the extent of vessel involvement. The diagnosis can be suggested by abnormal magnetic resonance imaging of the brain, an abnormal lumbar puncture, and/or demonstration of characteristic vessel abnormalities on arteriography **(Fig. 375-4)**, but it is confirmed by biopsy of the brain parenchyma and leptomeninges. In the absence of a brain biopsy, care should be taken not to misinterpret as true primary vasculitis arteriographic abnormalities that might actually be related to another cause. An important entity in the differential diagnosis is reversible cerebral vasoconstrictive syndrome, which typically presents with "thunderclap" headache and is associated with reversible arteriographic abnormalities that mimic primary CNS vasculitis. Other diagnostic considerations include infection, atherosclerosis, emboli, connective tissue disease, sarcoidosis, malignancy, and drug-associated causes. The prognosis of granulomatous primary CNS vasculitis is poor; however, some reports indicate that glucocorticoid therapy, alone or together with cyclophosphamide administered as described above, has induced clinical remissions. Following disease remission, cyclophosphamide should be switched to azathioprine or mycophenolate mofetil as these have good penetration into the CNS.

BEHÇET'S DISEASE

Behçet's disease is a clinicopathologic entity characterized by recurrent episodes of oral and genital ulcers, iritis, and cutaneous lesions. The underlying pathologic process is a leukocytoclastic venulitis, although vessels of any size and in any organ can be involved. **This disorder is described in detail in Chap. 376.**

COGAN'S SYNDROME

Cogan's syndrome is characterized by interstitial keratitis together with vestibuloauditory symptoms. It may be associated with a systemic vasculitis, particularly aortitis with involvement of the aortic valve.

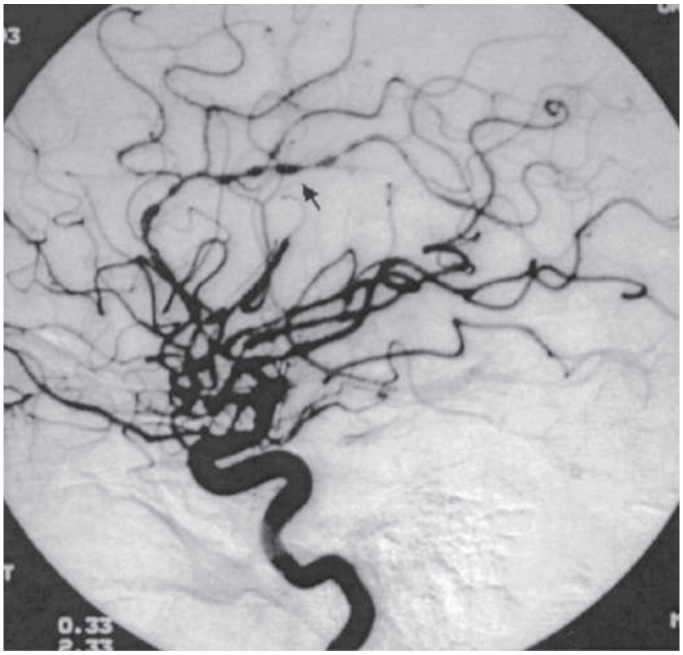

FIGURE 375-4 Cerebral arteriogram from a 32-year-old man with primary central nervous system vasculitis. Dramatic beading (*arrow*) typical of vasculitis is seen.

Glucocorticoids are the mainstay of treatment. Initiation of treatment as early as possible after the onset of hearing loss improves the likelihood of a favorable outcome.

KAWASAKI'S DISEASE

Kawasaki's disease is an acute, febrile, multisystem disease of children. Some 80% of cases occur prior to the age of 5, with the peak incidence occurring at ≤2 years. It is characterized by nonsuppurative cervical adenitis and changes in the skin and mucous membranes such as edema; congested conjunctivae; erythema of the oral cavity, lips, and palms; and desquamation of the skin of the fingertips. Although the disease is generally benign and self-limited, it is associated with coronary artery aneurysms in ~25% of cases, with an overall case-fatality rate of 0.5–2.8%. These complications usually occur between the third and fourth weeks of illness during the convalescent stage. Vasculitis of the coronary arteries is seen in almost all the fatal cases that have been autopsied and can cause complications into adulthood. There is typical intimal proliferation and infiltration of the vessel wall with mononuclear cells. Beadlike aneurysms and thromboses may be seen along the artery. Other manifestations include pericarditis, myocarditis, myocardial ischemia and infarction, and cardiomegaly.

Apart from the up to 2.8% of patients who develop fatal complications, the prognosis of this disease for uneventful recovery is excellent. High-dose IV γ-globulin (2 g/kg as a single infusion over 10 h) together with aspirin (100 mg/kg per day for 14 days followed by 3–5 mg/kg per day for several weeks) has been shown to be effective in reducing the prevalence of coronary artery abnormalities when administered early in the course of the disease. Surgery may be necessary for Kawasaki disease patients who have giant coronary artery aneurysms or other coronary complications. Surgical treatment most commonly includes thromboendarterectomy, thrombus clearing, aneurysmal reconstruction, and coronary artery bypass grafting.

Multisystem inflammatory syndrome (MIS-C), a serious condition that may resemble Kawasaki's disease, has been observed with infections due to SARS-CoV-2 **(Chap. 205)**. Although clinical features consistent with Kawasaki's disease have been observed, these patients can also have manifestations atypical for Kawasaki's disease, including gastrointestinal symptoms, myocarditis, neurocognitive symptoms, and shock. Any patient who presents with a clinical picture suggestive of Kawasaki's disease should be tested for SARS-CoV-2 to guide management.

POLYANGIITIS OVERLAP SYNDROMES

Some patients with systemic vasculitis manifest clinicopathologic characteristics that do not fit into a specific disease but have overlapping features of different vasculitides. The diagnostic and therapeutic considerations as well as the prognosis for these patients depend on the sites and severity of active vasculitis. Patients with vasculitis that could cause irreversible damage to a major organ system should be treated as described under "Granulomatosis with Polyangiitis."

SECONDARY VASCULITIS

■ DRUG-INDUCED VASCULITIS

Vasculitis associated with drug reactions usually presents as palpable purpura that may be generalized or limited to the lower extremities or other dependent areas; however, urticarial lesions, ulcers, and hemorrhagic blisters may also occur **(Chap. 63)**. Signs and symptoms may be limited to the skin, although systemic manifestations such as fever, malaise, and polyarthralgias may occur. Although the skin is the predominant organ involved, systemic vasculitis may result from drug reactions. Drugs that have been implicated in vasculitis include allopurinol, thiazides, gold, sulfonamides, phenytoin, and penicillin **(Chap. 63)**.

An increasing number of drugs have been reported to cause vasculitis associated with ANCA. Of these, the best evidence of causality exists for hydralazine and propylthiouracil. The clinical manifestations in ANCA-positive drug-induced vasculitis can range from cutaneous lesions to glomerulonephritis and pulmonary hemorrhage. Outside of drug discontinuation, treatment should be based on the severity of the vasculitis. Patients with immediately life-threatening small-vessel vasculitis should initially be treated with glucocorticoids and cyclophosphamide as described for granulomatosis with polyangiitis. Following clinical improvement, consideration may be given for tapering such agents along a more rapid schedule.

■ SERUM SICKNESS AND SERUM SICKNESS–LIKE REACTIONS

These reactions are characterized by the occurrence of fever, urticaria, polyarthralgias, and lymphadenopathy 7–10 days after primary exposure and 2–4 days after secondary exposure to a heterologous protein (classic serum sickness) or a nonprotein drug such as penicillin or sulfa (serum sickness–like reaction). Most manifestations are not due to vasculitis; however, occasional patients will have typical cutaneous venulitis that may progress rarely to a systemic vasculitis.

■ VASCULITIS ASSOCIATED WITH OTHER UNDERLYING DISEASES

Certain *infections* may directly trigger an inflammatory vasculitic process. For example, rickettsias can invade and proliferate in the endothelial cells of small blood vessels causing a vasculitis **(Chap. 192)**. In addition, the inflammatory response around blood vessels associated with certain systemic fungal diseases such as histoplasmosis **(Chap. 218)** may mimic a primary vasculitic process. A cutaneous leukocytoclastic vasculitis with occasional involvement of other organs may be a component of some infections. These include *subacute bacterial endocarditis*, *Epstein-Barr virus infection*, *HIV infection*, and a number of other infections.

Vasculitis can be associated with certain *malignancies*, particularly leukemic, lymphoid, or reticuloendothelial neoplasms. Leukocytoclastic venulitis confined to the skin is the most common finding; however, systemic vasculitis may occur.

A number of *connective tissue diseases* have vasculitis as a secondary manifestation of the underlying primary process. Foremost among these are *systemic lupus erythematosus* **(Chap. 368)**, *rheumatoid arthritis* **(Chap. 370)**, *inflammatory myositis* **(Chap. 377)**, *relapsing polychondritis* **(Chap. 378)**, and *Sjögren's syndrome* **(Chap. 373)**. The most common form of vasculitis in these conditions is the small-vessel venulitis isolated to the skin. However, certain patients may develop a fulminant systemic necrotizing vasculitis.

Secondary vasculitis has also been observed in association with *ulcerative colitis*, *congenital deficiencies of various complement components*, *sarcoidosis*, *primary biliary cirrhosis*, α_1*-antitrypsin deficiency*, and *intestinal bypass surgery*.

■ FURTHER READING

Finkielman JD et al: Antiproteinase 3 antineutrophil cytoplasmic antibodies and disease activity in Wegener granulomatosis. Ann Intern Med 147:611, 2007.

Jayne D et al: A randomized trial of maintenance therapy for vasculitis associated with antineutrophil cytoplasmic autoantibodies. N Engl J Med 349:36, 2003.

Jayne DRW et al: Avacopan for the treatment of ANCA-associated vasculitis. N Engl J Med 18:599, 2021.

Jennette JC et al: 2012 revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides. Arthritis Rheum 65:1, 2013.

Smith RM et al: Rituximab versus azathioprine for maintenance of remission for patients with ANCA-associated vasculitis and relapsing disease: An international randomised controlled trial. Ann Rheum Dis 82:937, 2023.

Spiera RF et al: Sarilumab for relapse of polymyalgia rheumatica during glucocorticoid taper. N Engl J Med 389:1263, 2023.

Stone JH et al: Rituximab versus cyclophosphamide for ANCA-associated vasculitis. N Engl J Med 363:221, 2010.

Stone JH et al: Trial of tocilizumab in giant-cell arteritis. N Engl J Med 377:317, 2017.

Walsh M et al: Plasma exchange and glucocorticoids in severe ANCA-associated vasculitis. N Engl J Med 382:622, 2020.

Wechsler ME et al: Mepolizumab or placebo for eosinophilic granulomatosis with polyangiitis. N Engl J Med 376:1921, 2017.

376 Behçet Syndrome

Yusuf Yazici

Behçet syndrome is a systemic vasculitis, first described by Hulusi Behçet, a Turkish dermatologist. It can present with oral and genital ulcers, skin lesions, uveitis, arthritis, major arterial and venous vessel disease, and gastrointestinal and neurologic manifestations. These manifestations can be present in various combinations and sequences over time. Patients are most commonly from the Middle East, the Mediterranean region, and the Far East; it is most prevalent in Turkey, with a prevalence of 1 in 250 adults. It is relatively rare before the late teens and after age 50. Males and females are equally affected; however, males frequently have more severe disease and poorer outcomes. Some manifestations may show regional differences; for example, gastrointestinal involvement, rare in Turkey, is more common in Japan and is seen in ~30% of patients in the United States.

DIAGNOSIS

Behçet syndrome is diagnosed clinically. There are no specific laboratory, imaging, or histologic features that can help in the diagnosis of a patient with suggestive symptoms. However, these can be used in ruling out conditions that may mimic Behçet syndrome and the diagnosis is based on a combination of clinical features in the setting of ruling out other potential causes. In this regard, some patients may require months to years to develop the array of symptoms that would lead to a definitive diagnosis, although a tentative diagnosis may be made well before. The most commonly used and best performing diagnostic criteria are the International Study Group (ISG) criteria (sensitivity ~95%, specificity ~96%); patients need to have recurrent oral ulcers plus two of the following four clinical manifestations: recurrent genital ulcers, skin lesions, eye lesions, or a positive pathergy test (Table 376-1). Additional clinical manifestations may involve various organ systems, including the gastrointestinal, vascular, pulmonary, and central nervous systems. Up to 50–60% of patients, depending on where they are from, can be positive for HLA B*51; however, it is not used as a diagnostic test because it is also found in around 20% of the normal population.

PATHOGENESIS

The pathogenesis and etiology of Behçet syndrome are unknown. Family studies show a possible genetic predisposition, and increased inflammation and immunologic mechanisms play a role. Both innate and adaptive immune systems may be involved. Unlike other autoimmune diseases, however, Behçet syndrome is not typically associated with autoantibodies, Raynaud's phenomenon, Sjögren's syndrome, thrombocytopenia, hemolytic anemia, sun hypersensitivity, serosal involvement, or an increased risk for other autoimmune diseases. On the other hand, features that separate it from autoinflammatory conditions include tendency to abate with time, absence of mutations associated with autoinflammatory diseases, and higher prevalence than typical autoinflammatory diseases such as familial Mediterranean fever (Chap. 381). There is neutrophil hyperreactivity; however, it is not clear whether this is primary or secondary to cytokine-directed activation. There is also evidence from retrospective patient cohort analyses that there may be different clusters of disease presentation; for example, acne lesions are more commonly seen with arthritis and associated with enthesitis, and each of these clusters may have a different pathogenesis.

TABLE 376-1 International Study Group Criteria for the Diagnosis of Behçet Syndrome

CRITERIA	FREQUENCY	COMMENTS
Oral ulcers	~98%	At least 3 times in a 12-month period
Plus 2 out of 4 from below:		
Recurrent genital ulcers	~80%	Usually scarring
Skin lesions	~80%	Erythema nodosum, pseudofolliculitis, papulopastular or acneiform nodules (postadolescent, not receiving corticosteroids)
Eye lesions	~50%	Anterior or posterior uveitis, cells in vitreous or retinal vasculitis
Pathergy	~50%	Evaluated in 24–48 h, after dermal insertion of a 20-gauge needle

CLINICAL PRESENTATION

The most common symptoms are associated with mucocutaneous tissues. Oral ulcers are seen in virtually all patients and are commonly the first manifestation (Fig. 376-1). Commonly, like ordinary canker sores, they are usually multiple. They last around 10 days but recur unless treated. Only the uncommon, major ulcers tend to scar. Beneficial effects of dental and periodontal therapies suggest that decreased oral health is associated with disease severity.

Genital ulcers are the most specific lesions, most commonly occurring on the scrotum or labia (Fig. 376-1). They are larger and deeper and take longer to heal than oral ulcers and tend to form scars.

Acne-like or papulopustular lesions are indistinguishable from acne vulgaris in appearance and pathology. They are seen both at the usual acne sites as well as at uncommon sites such as lower extremities. Other skin findings are the nodular lesions, which are of two types: erythema nodosum lesions due to panniculitis and superficial vein thromboses. Superficial thrombophlebitis often occurs in men and is associated with deep-vein thrombosis; it should trigger workup for other vascular involvement, including pulmonary artery aneurysms.

Pathergy reaction is a nonspecific hyperreactivity of the skin to trauma. Typically, a papule or pustule forms in 24–48 h after a

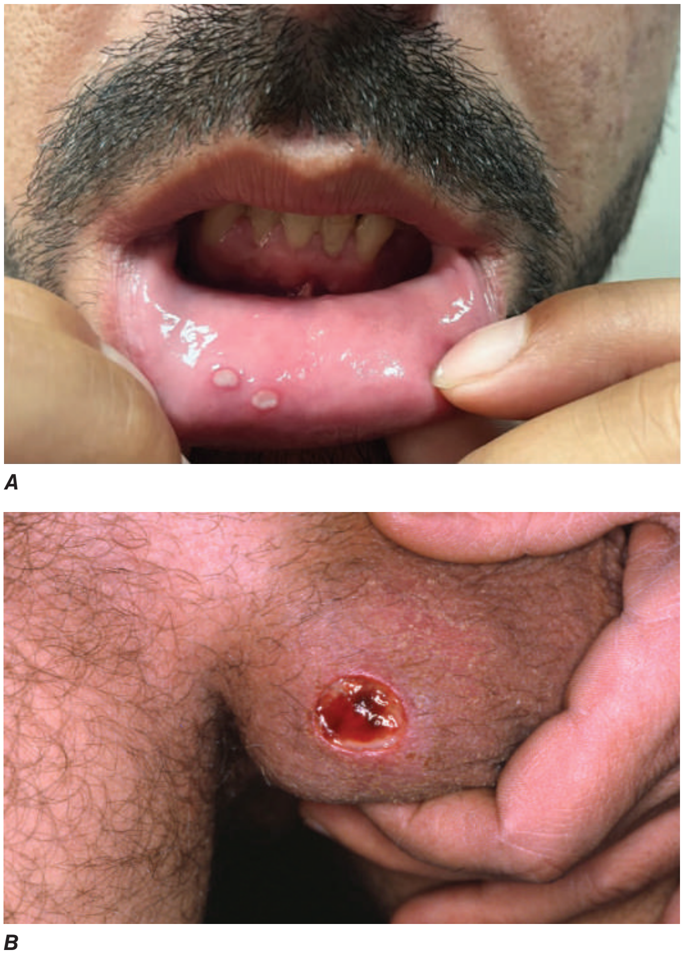

FIGURE 376-1 Clinical findings in Behçet syndrome. ***A.*** Behçet oral ulcer. ***B.*** Behçet scrotal ulcer.

20–21-gauge needle prick. It is rather unique for Behçet syndrome and is part of the ISG diagnostic criteria.

Arthralgia or arthritis is seen in about half of patients; it is usually a mono- or oligoarthritis in the lower extremities and does not usually cause erosions or deformity.

Eye involvement is seen in half of all patients and may be seen in up to ~70% of males. It is most commonly a bilateral panuveitis. A hypopyon, seen in ~10% of patients with eye disease, is an intense inflammation in the anterior chamber and is quite specific for Behçet syndrome. Ocular involvement develops usually in the first 2 years after fulfillment of diagnostic criteria and is most severe during the first few years and then tends to abate. Male gender, posterior involvement, frequent attacks (>3 per year), strong vitreous opacity, and macular edema are poor prognostic factors.

Vascular disease is seen in up to 40% of patients. It is associated with intensive thrombosis and runs a relapsing course. Several well-defined venous vascular associations are seen; superficial and deep-vein thrombosis, Budd-Chiari syndrome, inferior vena cava syndrome, pulmonary artery involvement, intracardiac thrombosis, and cerebral venous sinus thrombosis frequently cluster in various combinations. Pulmonary artery aneurysms carry a 5-year mortality rate of 20–25%.

Prevalence of neurologic involvement is ~5%, with about three-quarters of patients presenting with parenchymal involvement, while the remaining cases present with cerebral venous sinus thrombosis. These two forms rarely occur together. Parenchymal involvement usually affects the telencephalic-diencephalic junction, brainstem, and spinal cord. Patients may present with a subacute onset of severe headache, cranial nerve palsy, dysarthria, ataxia, and hemiparesis.

Prevalence of gastrointestinal involvement changes significantly across different populations (up to 50% in the Far East but rare in the Middle East). Clinical and endoscopic appearance of intestinal involvement can be similar to, and thus cannot easily be differentiated from, Crohn's disease. Ulcers tend to be single or less than five, are usually confined to the ileocecal area, are more likely to be deep and round, and are prone to perforate; perianal and rectal area involvement are rare. In practice, it is difficult to distinguish Behçet syndrome from Crohn's disease unless extraintestinal manifestations are present.

TREATMENT

Behçet Syndrome

Treatment is guided by type and severity of involvement, with the goal of preventing long-term damage. Most new manifestations tend to present within the first 5 years, and for most patients, the natural course is one of diminishing symptoms culminating in potential remission, frequently not requiring ongoing treatment with medications. Patient characteristics, such as being young and male, need to be kept in mind when making treatment decisions, as these patients tend to have a worse prognosis. In most patients, tapering and/or stopping their medications in 2–3 years after the symptoms have improved should be attempted.

Oral ulcers can be managed with topical glucocorticoids and on an as-needed basis if mild. Lesions resistant to local measures may require systemic treatment with colchicine, oral glucocorticoids, immunosuppressants such as apremilast, azathioprine, or a tumor necrosis factor (TNF)-α inhibitor such as infliximab. A similar treatment approach can be used for genital ulcers and other mucocutaneous manifestations. Patients may need a combination of medications, at least initially, to control disease activity.

Eye involvement, given its frequency and potential morbidity, requires early and aggressive treatment with brief courses of glucocorticoids and longer-term treatment with an immunosuppressant. Azathioprine is usually the preferred agent in clinical practice. TNF inhibitors infliximab or adalimumab can also be used, either as first-line monotherapy or more commonly in combination with systemic glucocorticoids and azathioprine, for control of disease activity. Cyclosporine can also be considered in combination regimens; monotherapy with interferon is another option. Glucocorticoids can be tapered in many patients after active disease has been controlled, whereas immunosuppressants are generally continued for at least 2 years with plans to potentially taper them also based on treatment response and ongoing disease activity.

Gastrointestinal involvement is treated with glucocorticoids plus an immunosuppressant such as azathioprine alone or in combination with infliximab.

Venous thrombotic events are treated by controlling systemic inflammation with immunosuppressive medications (usually azathioprine or, for more severe cases, cyclophosphamide), rather than using anticoagulants. However, if venous thrombotic events occur, standard anticoagulation treatment can be given, provided there is a low risk of bleeding and there are no coexistent pulmonary artery aneurysms. For central nervous system involvement, the combination of azathioprine and a TNF inhibitor is usually the first choice.

FURTHER READING

Hatemi G et al: 2018 update of the EULAR recommendations for the management of Behçet's syndrome. Ann Rheum Dis 77:808, 2018.

Kural-Seyahi E et al: The long-term mortality and morbidity of Behçet syndrome: A 2-decade outcome survey of 387 patients followed at a dedicated center. Medicine (Baltimore) 82:60, 2003.

Yazici H et al: Behçet syndrome: A contemporary view. Nat Rev Rheumatol 14:107, 2018.

Yazici Y et al: Behçet syndrome. Nat Rev Dis Primers 7:67, 2021.

377

Inflammatory Myopathies

Steven A. Greenberg, Anthony A. Amato

This chapter focuses on the major types of inflammatory myopathies (IMs), including dermatomyositis (DM), polymyositis (PM), immune-mediated necrotizing myopathy (IMNM), antisynthetase syndrome (ASyS), and inclusion body myositis (IBM) **(Table 377-1)**. Other IMs include those caused by infection, eosinophilic myositis, granulomatous myositis, and myositis triggered by checkpoint inhibitors. Of note, inflammatory cell infiltrates can also be occasionally seen in muscle biopsies of hereditary myopathies (e.g., muscular dystrophies, metabolic myopathies), sporadic late-onset nemaline myopathy (SLONM), and toxic myopathies.

Epidemiologic studies suggest that the incidence of IM grouped together is up to 16 cases per 100,000 with prevalence in the range of 14–32 per 100,000. Defining the actual incidence and prevalence of the individual myositides is limited, however, by different diagnostic criteria employed in various epidemiologic studies, an increasing recognition of ASyS and IMNM, as well as the frequent misdiagnosis of IBM. Idiopathic PM without signs of an overlap syndrome is quite rare compared to DM, ASyS, IBM, and IMNM that occur in roughly similar frequencies. DM can occur in children (juvenile DM), while IBM always occurs in adults and is the most common cause of myopathy in those aged >50. DM, PM, and ASyS are more common in women, while IBM is more common in men.

DIAGNOSTIC APPROACH AND DIFFERENTIAL DIAGNOSIS

The approach to patients with suspected myopathy is detailed in **Chap. 460**. In any patient presenting with weakness, the first step is to localize the site of the lesion by history and clinical findings **(Chap. 26)**. Weakness could be caused by a process in the cerebral hemispheres, spinal cord **(Chap. 453)**, anterior horn cell **(Chap. 448)**, peripheral nerve

TABLE 377-1 Inflammatory Myopathies: Clinical and Laboratory Features

DISORDER	SEX	AGE OF ONSET	RASH	PATTERN OF WEAKNESS	LABORATORY FEATURES	MUSCLE BIOPSY	CELLULAR INFILTRATE	RESPONSE TO IS THERAPY	COMMON ASSOCIATED CONDITIONS
DM	F > M	Childhood and adult	Yes	Proximal > distal	Normal or increased CK (up to 50× normal or higher); various MSAs (anti-MDA5, anti-TIF1, anti-Mi-2, anti-NXP2)	Perimysial and perivascular inflammation; IFN-1 regulated proteins (MHC-1, MxA), MAC deposition on capillaries	CD4+ dendritic cells; B cells; macrophages	Yes	Myocarditis, ILD, malignancy, vasculitis, other CTDs
PM	F > M	Adult	No	Proximal > distal	Increased CK (up to 50× normal or higher)	Endomysial and perivascular inflammation; ubiquitous expression of MHC-1	CD8+ T cells; macrophages; plasma cells	Yes	Myocarditis, ILD, other CTDs
IMNM	M = F	Children and adults	No	Proximal > distal	Elevated CK (>10× normal or higher); anti-HMGCR or anti-SRP antibodies	Necrotic muscle fibers; minimal inflammatory infiltrate; MHC-I and MAC deposition on sarcolemma of scattered nonnecrotic muscle fibers	Macrophages in necrotic fibers undergoing phagocytosis	Yes	Malignancy, CTD, HMGCR antibody cases can be triggered by statin use
ASyS	F > M	Children and adults	Sometimes	Proximal > distal	Elevated CK (>10× normal or higher); antisynthetase antibodies	Perimysial and perivascular inflammation; perimysial fragmentation with alkaline phosphatase staining; perimysial muscle damage with necrosis; MHC-I, HLA-DR, and MAC deposition on sarcolemma of perifascicular muscle fibers	CD4+ dendritic cells; B cells; macrophages	Yes	Nonerosive arthritis, ILD, Raynaud's phenomenon, mechanic hands, and fever
IBM	M > F	Older adults (>50 years)	No	Proximal and distal; predilection for: finger/wrist flexors, knee extensors	Normal or mildly increased CK (usually <10× normal); anti-cN-1A antibodies; large granular lymphocytes on flow cytometry and reduced CD4/CD8 ratio with increased CD8 count	Endomysial and perivascular inflammation; ubiquitous expression of MHC-1 and HLA-DR; rimmed vacuoles; p62, LC3, TDP-43 aggregates; EM: 15–18 nm tubulofilaments; ragged red and COX-negative fibers	CD8+ T cells; macrophages; plasma cells; myeloid dendritic cells; large granular lymphocytes	None or minimal	Granular lymphocytic leukemia/lymphocytosis, sarcoidosis, sicca or Sjögren's syndrome

Abbreviations: CK, creatine kinase; cN-1A, cytosolic 5′-nucleotidase 1A; CTDs, connective tissue diseases; COX, cytochrome oxidase; DM, dermatomyositis; F, female; g, immunoglobulin; IBM, inclusion body myositis; IFN-1, type 1 interferon; ILD, interstitial lung disease; IS, immunosuppressive; M, male; MAC, membrane attack complex; MDA5, melanoma differentiation antigen; MHC-1, major histocompatibility antigen 1; MSA, myositis-specific autoantibodies; NCP2, nuclear matrix protein 2 (NXP2); NM, necrotizing myopathy; PM, polymyositis; TIF1, transcriptional intermediary factor 1.

Source: Reproduced with permission from AA Amato, JA Russell (eds): *Neuromuscular Disorders*. 2nd ed. New York: McGraw-Hill Education; 2016.

(Chaps. 457–458), neuromuscular junction **(Chap. 459)**, or muscle **(Chap. 460)**. Past medical history, medication use, and family history, combined with a detailed clinical examination and an appreciation for the pattern of muscle involvement (e.g., what muscles are weak and atrophic or hypertrophic as well as the presence of scapular winging, early contractures, sensory abnormalities, fasciculations, or rash), help differentiate myopathies from other neuromuscular disorders and the different types of myopathies from each other **(see Chap. 460)**. For example, atrophy with fasciculations suggests a neurogenic process such as amyotrophic lateral sclerosis, fatigable weakness on examination points to a neuromuscular junction defect such as myasthenia gravis, and concomitant sensory symptoms suggest a central process such as a spinal cord disorder or a polyneuropathy. Scapular winging, calf hypertrophy or atrophy, and early contractures before significant weakness develops would strongly suggest a muscular dystrophy, particularly if there is a positive family history. A heliotrope rash combined with Gottron papules (below) and dilated nailfold capillaries is diagnostic for DM. The presence of atrophy and weakness of the flexor forearm muscles and quadriceps in a person aged >50 years is most likely IBM.

When the site of the lesion cannot be localized based on history and clinical examination alone, laboratory testing is required. Serum creatine kinase (CK) is the most sensitive laboratory marker of muscle destruction. Not all myopathies are associated with elevated CK levels, but a markedly elevated CK (e.g., >2000 U/L) is almost always due to a myopathy. A slightly elevated CK can also be seen in neurogenic disorders, however. Myositis-associated and myositis-specific antibodies (MSAs) help to distinguish subtypes of IM, as discussed below. Electromyography (EMG) and nerve conductions studies (NCS) are useful in localizing the site of the lesion but are less specific in helping to determine the actual cause of a myopathy. EMG can be useful at times in guiding what muscle to biopsy, especially if muscles typically biopsied are normal on clinical examination. Imaging skeletal muscle can be helpful in assessing muscle involvement and revealing fatty replacement, atrophy, or edema within muscle or surrounding fascia.

A muscle biopsy is often required to definitively distinguish one myopathy from another, if there is no characteristic dermatomyositis rash or myositis specific autoantibody. However, a muscle biopsy should be performed in every case of suspected PM to exclude IBM (if not clinically apparent) and other causes of myopathy. Diagnosis of IMNM is by definition based upon histologic findings but again is not needed if a patient has clinical features and anti-3-hydroxy-3-methylglutaryl-coenzyme reductase (HMGCR) or anti–signal recognition particle (SRP) antibodies. It is important to biopsy a muscle that is clinically affected but not too weak (e.g., Medical Research Council grade 4 out of 5 in strength); otherwise, one may just see end-stage muscle. A biopsy should always be coordinated with an experienced muscle histopathology laboratory.

Patients with severe muscle pain, subjective weakness, and fatigue with normal strength and function on examination are not likely to have an IM. Polymyalgia rheumatica should be considered in older individuals with an elevated erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP) but normal CK and EMG. Fibromyalgia is likely in patients with a normal laboratory workup. In general, a muscle biopsy is not indicated unless there is objective weakness, an abnormal EMG, or elevated CK.

SPECIFIC DISORDERS

DERMATOMYOSITIS

Clinical Features DM manifests with symmetric, proximal greater than distal weakness along with a characteristic rash that includes the heliotrope rash (erythematous discoloration of eyelids with periorbital edema), Gottron sign (erythematous rash over the extensor surfaces of joints such as the knuckles, elbows, knees, and ankles), Gottron papules (raised erythematous rash over knuckles) **(Fig. 377-1)**, V-sign (rash on the sun-exposed anterior neck and chest), shawl sign over the back of the neck and shoulders, nail bed telangiectasias, and subcutaneous calcium deposits. The weakness and rash usually accompany one another but can be separated by several months. Furthermore, beyond "classic DM" with prominent muscle and skin manifestations, there is a spectrum of involvement such that some patients have skin-predominant disease (only with a rash called amyopathic DM, or minimal muscle disease called hypomyopathic DM), while others may present mainly with weakness and little or no visible skin changes. Patients may also have myalgias, arthralgias, dysphagia, and dysarthria. Cutaneous disease activity is highly relevant in DM; in comparison to other debilitating skin diseases including cutaneous lupus erythematosus, psoriasis, and atopic dermatitis, skin symptoms in DM patients are associated with an overall reduction in life quality. Pruritus can be especially debilitating. Dyspnea can occur from ventilatory muscle weakness or intrinsic lung involvement including interstitial lung disease (ILD), bronchopneumonia, and alveolitis. Pulmonary manifestations are often associated with anti-MDA-5 antibodies or with antisynthetase antibodies; myositis associated with the ASyS is now considered a distinct disorder (discussed below). DM can present in children (juvenile DM) or in adults. There is a higher risk for malignancy in adult-onset cases, ~15% within the first 2–3 years.

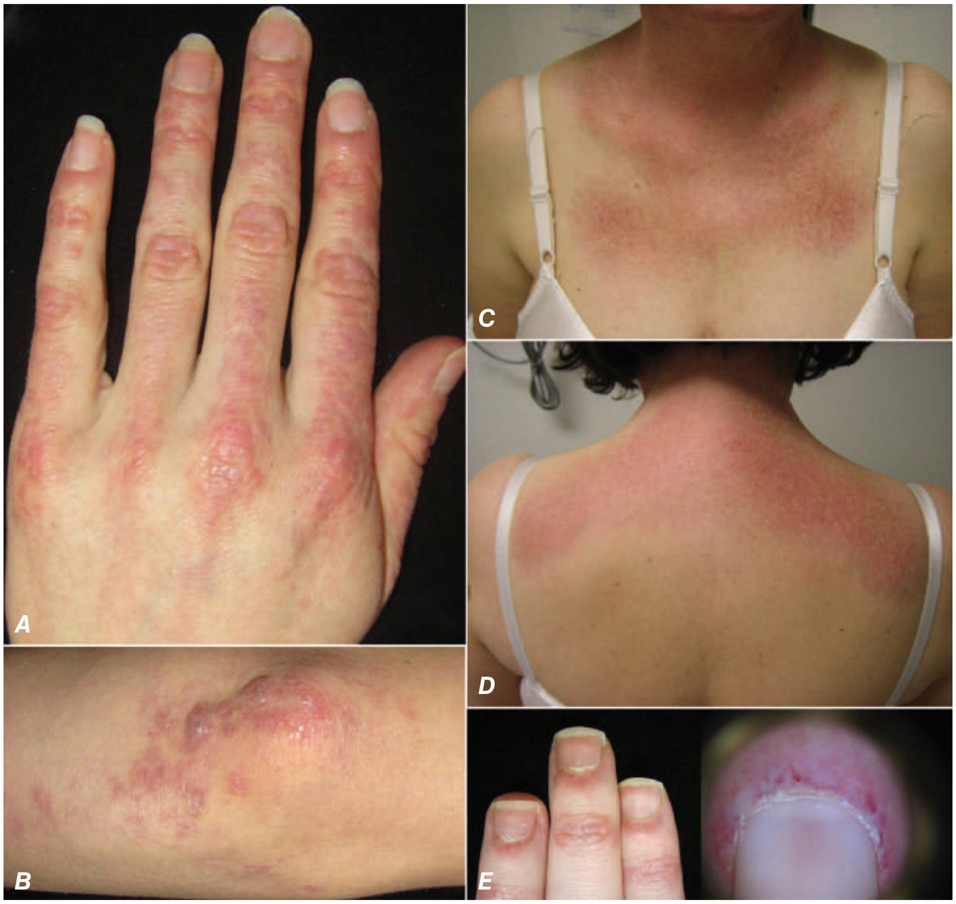

FIGURE 377-1 Cutaneous manifestations of dermatomyositis. ***A.*** Macular erythema plaques (Gottron sign) and erythematous papules (Gottron papules) on extensor surface of fingers and ***B.*** elbow. ***C.*** Macular erythema plaques over anterior neck and chest (V-sign) and ***D.*** the posterior neck, shoulder, and upper back (Shawl sign). ***E.*** Nail bed changes with dilated capillaries.

Laboratory Features Serum CK levels are elevated in 70–80% of patients; in 10% of those with normal CK, serum aldolase may be increased. Antinuclear antibodies can be positive but are a nonspecific finding. The myositis-specific antibodies (MSAs) that are specific for DM include anti–complex nucleosome remodeling histone deacetylase (anti-Mi-2), anti–transcription intermediary factor 1-γ (anti-TIF1-γ), anti–melanoma differentiation-associated gene 5 (anti-MDA5), anti–nuclear matrix protein 2 (anti-NXP-2), and anti–small ubiquitin-like modifier activating enzyme (anti-SAE). These antibodies are usually associated with characteristic clinical features, and recent studies suggest that they are also directly involved in the pathogenesis of DM. Mi-2 antibodies are found in 15–20% of patients with DM and are typically associated with an acute onset, a florid rash, and prominent weakness but a good response to therapy and a favorable prognosis. Anti-MDA5 antibodies are found in 10–20% of DM patients and up to 65% of patients with clinically amyopathic DM. This antibody is associated with palmar rash, severe skin ulcerations from ischemia, and rapidly progressive ILD. Anti-TIF1-γ, also known as p155, antibodies are found in adult cancer-associated DM with an 89% specificity and 70% sensitivity. Thus, enhanced vigilance for underlying cancer is especially important in these patients. Anti-NXP-2 antibodies are found in as many as 17% of patients with DM and are also associated with calcinosis, subcutaneous edema, distal weakness, and dysphagia, as well as with cancer. Anti-SAE antibodies are present in 1.5–8% of DM and are associated with an underlying cancer in 14–57% of patients. Most manifest with a skin rash alone, and CK is often normal, but approximately one-third of patients have elevated aldolase levels. ILD can also be seen in anti-SAE DM, but unlike anti-MDA-5 amyopathic DM, the ILD is usually mild.

EMG of weak muscles shows increased insertional and spontaneous activity in the form of positive sharp waves and fibrillation potentials, or complex repetitive discharges along with early recruitment of small-amplitude, short-duration, polyphasic motor units. These findings are nonspecific and can also be seen in other myopathies. Skeletal muscle magnetic resonance imaging (MRI muscle) reveals edema in affected muscles and sometimes more specific findings of abnormalities of fascia suggesting fasciitis.

Histopathology and Pathogenesis The characteristic histopathologic abnormality on muscle biopsy is perifascicular atrophy **(Fig. 377-2*A*)**; however, this finding is present in perhaps only 50% of patients. Immunohistochemical staining for myxovirus resistance protein A (MxA) is diagnostically more sensitive and highly specific **(Fig. 377-2*B*)**. The inflammatory cell infiltrate is predominantly perivascular and located in the perimysium and is composed primarily of macrophages, B cells, and plasmacytoid dendritic cells. Recent studies have highlighted some variability in histologic abnormalities associated with different MSAs. Skin biopsies reveal cell-poor interface dermatitis, which is analogous to the perifascicular atrophy in that the basal

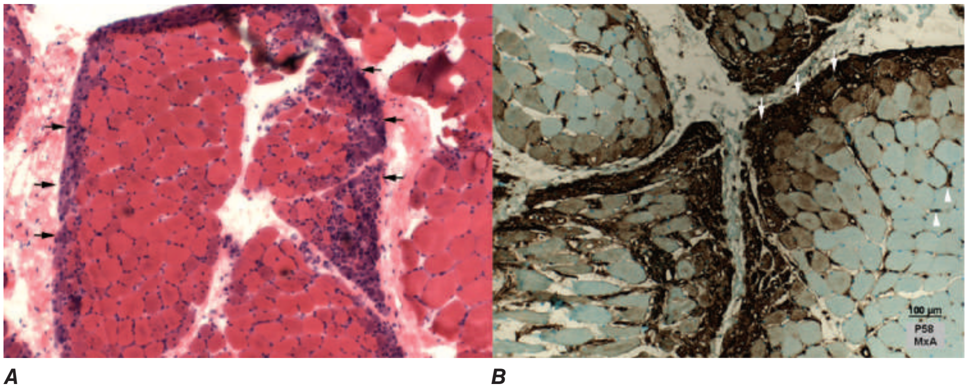

FIGURE 377-2 Perifascicular atrophy and myxovirus resistance protein A (MxA) expression in dermatomyositis. ***A.*** Perifascicular myofibers (*black arrows*) bordering on disrupted perimysial connective tissue are atrophic and basophilic on hematoxylin and eosin (H&E) stains. ***B.*** Perifascicular myofibers (*white arrows*) show intense staining for MxA protein along a gradient from superficial to deep; all capillaries show intense MxA expression (*white arrowheads*).

layer of keratinocytes is most damaged; the inflammatory infiltrate is typically absent or minimal and, when present, is located mainly at the border zone of the dermis and epidermis.

The pathogenesis of DM was traditionally attributed to an antibody-mediated attack on endothelial cells, followed by complement-mediated destruction of capillaries and watershed ischemia of muscle fibers. However, subsequent studies suggest that this is not likely the case. Immunoglobulin deposition is largely absent on endothelial cells, and complement deposition may be a secondary phenomenon. There is increasing evidence that the microvasculopathy and skin and muscle damage associated with DM are primarily due to toxicity from type I interferon (IFN)–mediated pathways, most likely IFN-β. As mentioned, there is increasing evidence that the MSAs are directly pathogenic. For example, anti-Mi-2 antibodies appear to be capable of entering myonuclei and inhibiting the CHD4/NuRD complex in the nucleosome.

Prognosis In the absence of malignancy, prognosis is generally favorable in patients with DM, with 5-year survival rates ranging from 70 to 93%. Poor prognostic features are increased age, associated ILD, cardiac disease, and late or previous inadequate treatment.

■ POLYMYOSITIS

Clinical Features PM is a heterogenous group of disorders that usually presents with symmetric and proximal weakness that worsens over several weeks to months. As with DM, there can be associated heart, lung, and joint involvement as well as an increased risk of cancer. Some epidemiologic studies suggest that the risk of cancer in PM is less than that in DM, but these older series likely included patients with IBM and dystrophies with inflammation who were misdiagnosed as having PM.

Laboratory Features CK levels are always elevated in uncontrolled PM. A normal CK should alert clinicians to the possibility of IBM. EMG and skeletal muscle imaging can be abnormal, but the findings are not specific **(Fig. 377-3)**.

Histopathology and Pathogenesis Because PM is a heterogeneous category, muscle pathology varies substantially. Most often, patients with nonspecific inflammatory cells present in perimysial more often than endomysial locations have been categorized as PM. A small minority of patients have a mononuclear inflammatory infiltrate that surrounds fibers with sarcolemmal expression of major histocompatibility complex (MHC-I) molecules **(Fig. 377-4)**.

There is debate as to whether a true invasion of myofibers occurs in PM or rather always indicates IBM. The inflammatory infiltrate predominantly consists of CD8+ T cells and macrophages located in the endomysial, perimysial, and perivascular regions. PM is heterogeneous, and its varied forms of pathogenesis are poorly understood.

Prognosis Most patients with PM improve with immunotherapies but usually require lifelong treatment. Some retrospective studies suggest that PM does not respond as well as DM to these therapies. However, many of these older series of "PM" likely included patients who actually had IMNM, IBM, or other myopathies (including muscular dystrophies) that do not respond to immunotherapies. As in DM, poor prognostic features are cancer, increased age, lung or cardiac involvement, and late or previously inadequate treatment.

■ OVERLAP SYNDROMES

The term *overlap syndrome* is applied when an inflammatory myopathy is associated with other well-defined connective tissue diseases (CTDs) such as scleroderma, mixed connective tissue disease (MCTD), Sjögren's syndrome, systemic lupus erythematosus (SLE), or rheumatoid arthritis. Overlap syndromes are usually responsive to immunotherapies. The exception is Sjogren's syndrome with coexisting IBM.

■ IMMUNE-MEDIATED NECROTIZING MYOPATHY

Clinical Features IMNM, or autoimmune necrotizing myopathy, is characterized by the acute or insidious onset of symmetric, proximal more than distal weakness. Dysphagia, dysarthria, or myalgia may occur. Patients may have an underlying CTD (usually scleroderma or MCTD) or cancer (paraneoplastic necrotizing myopathy), or the condition may be idiopathic. There are at least two distinct forms of IMNM associated with specific autoantibodies (anti-HMGCR and anti-SRP). Anti-HMGCR myopathy can be seen in patients receiving statins, inhibitors of HMGCR, particularly in those aged >50 years. However, anti-HMGCR myopathy can develop in children and young adults without a history of statin use and can mimic a limb girdle muscular dystrophy. Unlike the more common "toxic" myopathy associated with statin use, anti-HMGCR myopathy does not improve when statins are discontinued. Anti-SRP myopathies are notable for the presence of anti-SRP antibodies and a typically subacute, aggressive, and relatively refractory course.

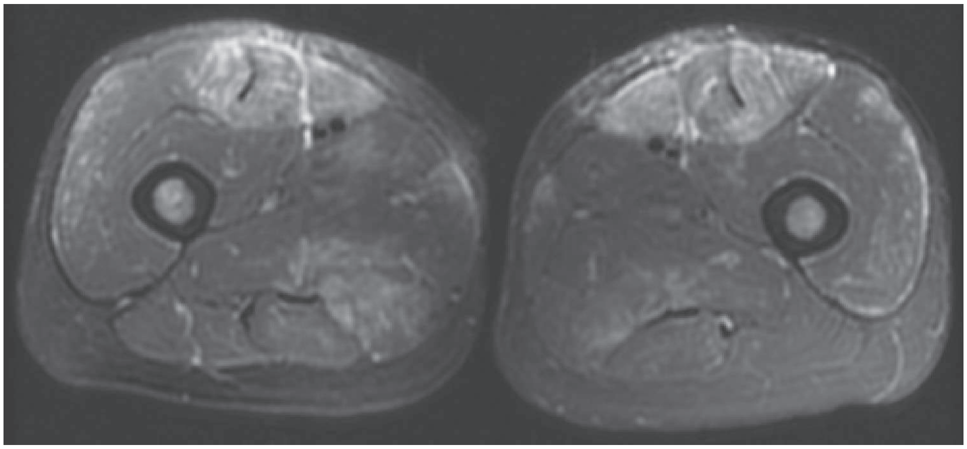

FIGURE 377-3 Skeletal muscle magnetic resonance imaging (MRI) with short T1 inversion recovery (STIR) imaging in polymyositis. MRI of the thigh demonstrates bright signal indicative of edema/inflammation, particularly in the rectus femoris muscle. This contrasts with MRI in inclusion body myositis in which there is more selective involvement of the vastus lateralis and medialis with relative sparing of the rectus femoris (see Fig. 377-7*F* and *G*).

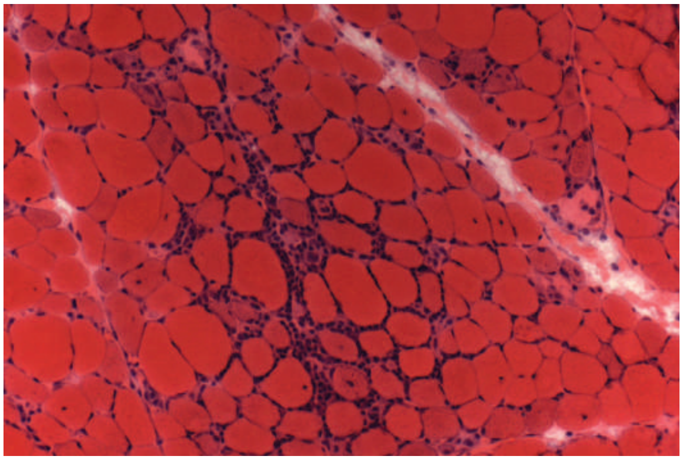

FIGURE 377-4 Pathology of polymyositis. Muscle biopsy demonstrates endomysial infiltrates surrounding nonnecrotic muscle fibers.

Laboratory Features CK levels are markedly elevated (usually >10 × normal) in IMNM and are associated with titers of anti-HMGCR or anti-SRP antibodies. EMG often shows increased insertional and spontaneous activity, including myotonic discharges. Skeletal muscle imaging findings are nonspecifically abnormal.

Histopathology and Pathogenesis Muscle biopsies reveal multifocal necrotic and regenerating muscle fibers with a paucity of inflammatory cells **(Fig. 377-5)**. However, some patients with anti-HMGCR myopathy have endomysial, macrophage-predominant infiltrates similar to what is seen in PM. Overexpression of MHC-I and membrane attack complex (MAC) molecules may be evident on sarcolemma of nonnecrotic fibers, and MAC deposition present on capillaries. The pathogenesis of IMNM is not completely understood and likely varies depending on subtype. A trial of a complement inhibitor in both anti-HMGCR and anti-SRP myopathies failed to demonstrate any efficacy, so IMNM in these subtypes does not appear to be primarily complement driven. Interestingly, pathogenic biallelic variants in the *HMGCR* gene result in proximal weakness, myalgias, high CK, and dystrophic changes on skeletal muscle MRI and biopsies. Recent studies suggest that HMGCR antibodies may bind to the receptor on the sarcolemma and lead to accumulation of acetyl-CoA and subsequently an increase in lipids within muscle fibers. Recent studies also suggest that anti-SRP antibodies are directly causal to the associated IMNM.

Prognosis Anti-HMGCR myopathy is often responsive to intravenous immunoglobulin (IVIG) monotherapy. However, anti-SRP myopathy and seronegative IMNM are generally much more difficult to treat, and aggressive immunotherapy is usually required. The progressive course despite immunotherapy and the marked weakness with atrophy can lead to a misdiagnosis of a limb girdle muscular dystrophy. There may be an increased incidence of cancer in patients with anti-HMGCR myopathy; thus, patients should undergo a malignancy workup.

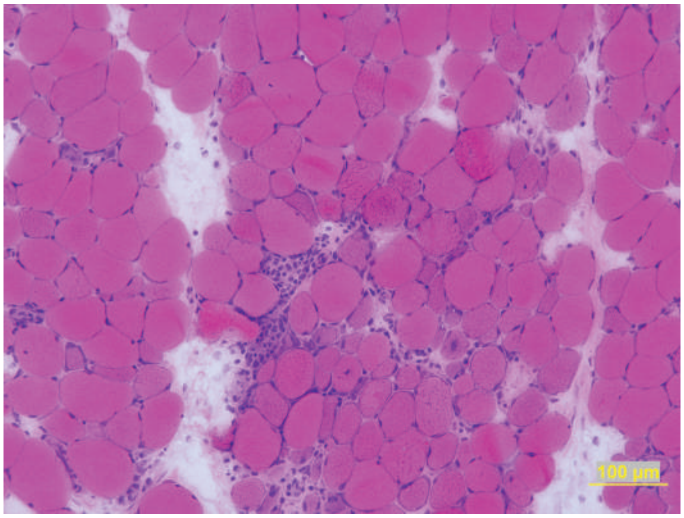

FIGURE 377-5 Pathology of immune-mediated necrotizing myopathy. Muscle biopsy demonstrates scattered necrotic fibers with inflammatory infiltrate confined to those fibers undergoing myophagocytosis along with a few regenerating fibers.

■ ANTISYNTHETASE SYNDROME

Clinical Features The presence of myositis, nonerosive arthritis, ILD, Raynaud's phenomenon, mechanic hands, and fever associated with antibodies against aminoacyl-tRNA synthetase constitute the ASyS. Some patients have an erythematous rash, and muscle biopsies share histopathologic features of DM, which likely accounts for many of these patients being classified as having DM.

Laboratory Features Antibodies against aminoacyl-tRNA synthetases are the most common MSA, present in 25–35% of patients with myositis. These include anti-Jo-1 (histidyl), anti-PL-7 (threonyl), anti-PL-12 (alanyl), anti-EJ (glycyl), anti-OJ (isoleucyl), anti-KS (asparaginyl), anti-Ha (tyrosyl), anti-Zo (phenylalanyl), and other less common antibodies. The most common aminoacyl-tRNA synthetase antibody is anti-Jo-1. CK is usually elevated in patients with ASyS and myositis. Those with ILD demonstrate reduced forced vital capacity and diffusion capacity on pulmonary function tests. Spiral chest computed tomography (CT) scans are best at demonstrating the honeycomb pattern of ILD. Skeletal muscle MRI and EMG show abnormalities similar to DM, PM, and IMNM.

Histopathology and Pathogenesis Muscle biopsies demonstrate a predilection for perimysial damage including perimysial fragmentation and staining with alkaline phosphatase **(Fig. 377-6)**, plasmacytoid dendritic cells and macrophages in the perimysium and around blood vessels, and MAC deposition on capillaries. Also similar to DM, there is perifascicular muscle fiber damage, but with ASyS, there is more perifascicular muscle fiber necrosis compared to DM, in which perifascicular atrophy is more prominent. MHC-I, MHC-II (HLA-DR), and MAC deposits on muscle fibers may be seen on sarcolemma of perifascicular muscle fibers. The HLA-DR expression on muscle fibers suggests that ASyS is more driven by gamma-IFN than type 1 IFN. Recent studies suggest that the various antibodies may be directly pathogenic by binding to specific aminoacyl-tRNA synthetases thereby impairing protein synthesis. More work needs to be done to confirm these observations.

Prognosis Most patients respond to treatment, although ASyS can be difficult to treat, in particular for patients with interstitial lung disease. Aggressive treatment, often with early rituximab, is warranted in such cases. There does not appear to be an increased risk of malignancy.

■ INCLUSION BODY MYOSITIS

Clinical Features IBM usually manifests in patients over the age of 50 years and is slightly more common in men than women. It is associated with slowly progressive weakness and muscle atrophy that has a predilection for early involvement of the wrist and finger flexors in the arms and quadriceps in the legs **(Fig. 377-7)**. Weakness is often asymmetric. Dysphagia is common and rarely can be the presenting feature. These clinical features can help distinguish IBM from PM and other forms of myopathy. The mean duration from onset of symptoms to use of wheelchair or scooter is ~15 years. There is no known increase in risk of malignancy.

Laboratory Features CK levels can be normal or only slightly elevated (usually <10 times normal). Antibodies targeting cytosolic 5′-nucleotidase 1A (cN-1A) are detected in the blood in a third to more than two-thirds of IBM patients and are a highly specific diagnostic biomarker for IBM among patients with myopathy. Other blood biomarkers for IBM include the presence of an abnormal population of large granular lymphocytes on flow cytometry and a reduced CD4/CD8

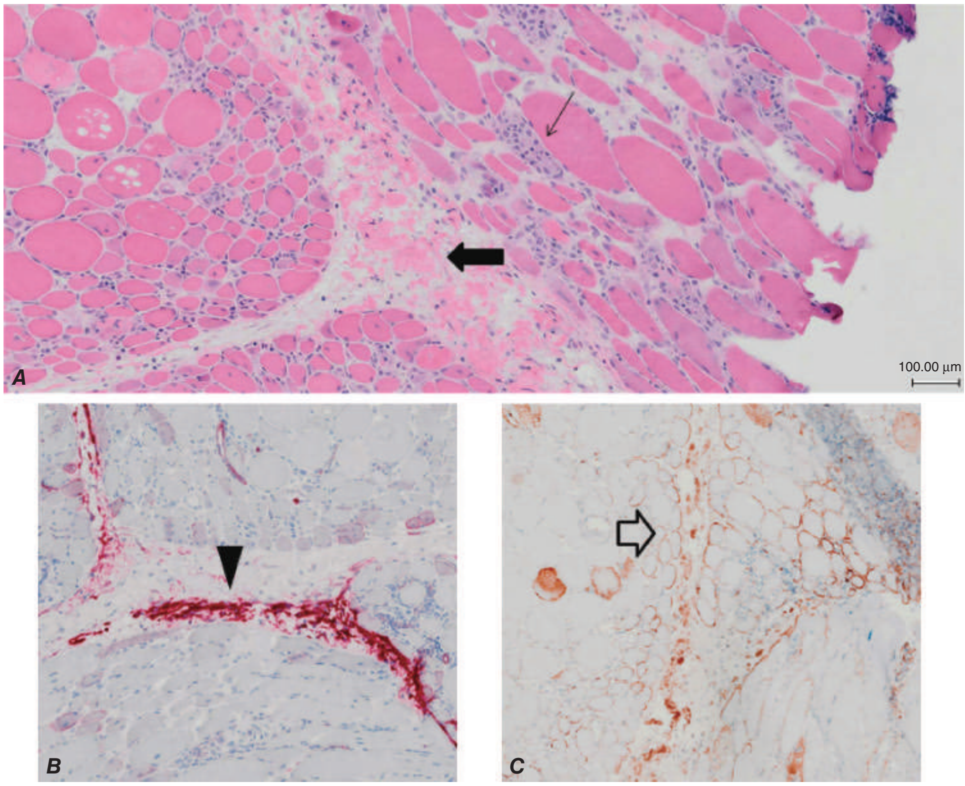

FIGURE 377-6 Pathology of myositis with anti-Jo-1 antibodies (antisynthetase syndrome). ***A.*** Perifascicular/perimysial muscle fiber atrophy and necrosis (*thin arrow*) associated with perimysial connective tissue is edematous and fragmented in appearance (*thick arrow*), hematoxylin and eosin stain. ***B.*** The perimysial connective tissue intensely stains red with alkaline phosphatase stain (*arrowhead*). ***C.*** Immunostaining demonstrates deposition of membrane attack complex (MAC) deposits on the sarcolemma of nonnecrotic perifascicular muscle fibers (*open arrow*).

ratio with an increased CD8 count. Needle EMG may demonstrate large-amplitude, long-duration motor unit potentials that can be misinterpreted as neurogenic but reflect the chronicity of the myopathy. Muscle MRI may show a predilection for involvement of the flexor digitorum profundus in the arms and the vastus medialis and lateralis muscles with sparing of the rectus femoris muscle.

Histopathology and Pathogenesis Muscle biopsies demonstrate endomysial inflammatory infiltrates predominantly composed of highly differentiated CD8+ CD57+ T cells that express killer cell lectin-like receptor G1 (KLRG1) molecules, plus macrophages, that surround and invade nonnecrotic muscle fibers expressing MHC-I and MHC-II on the sarcolemma, along with fibers with rimmed vacuoles, cytochrome oxidase (COX)–negative fibers, and inclusions on light or electron microscopy **(Fig. 377-8)**. The inclusions contain beta-sheet misfolded proteins (amyloid) but are difficult to appreciate with routine Congo red stain (they are seen on frozen but not paraffin sections). Immunostaining for p62 appears to be the most sensitive stain for detection of these inclusions. Importantly, rimmed vacuoles may not be seen in as many as 20–30% of muscle biopsies. In such cases, the presence of mitochondrial abnormalities (ragged red and COX-negative fibers) and immunostaining demonstrating p62 inclusions are

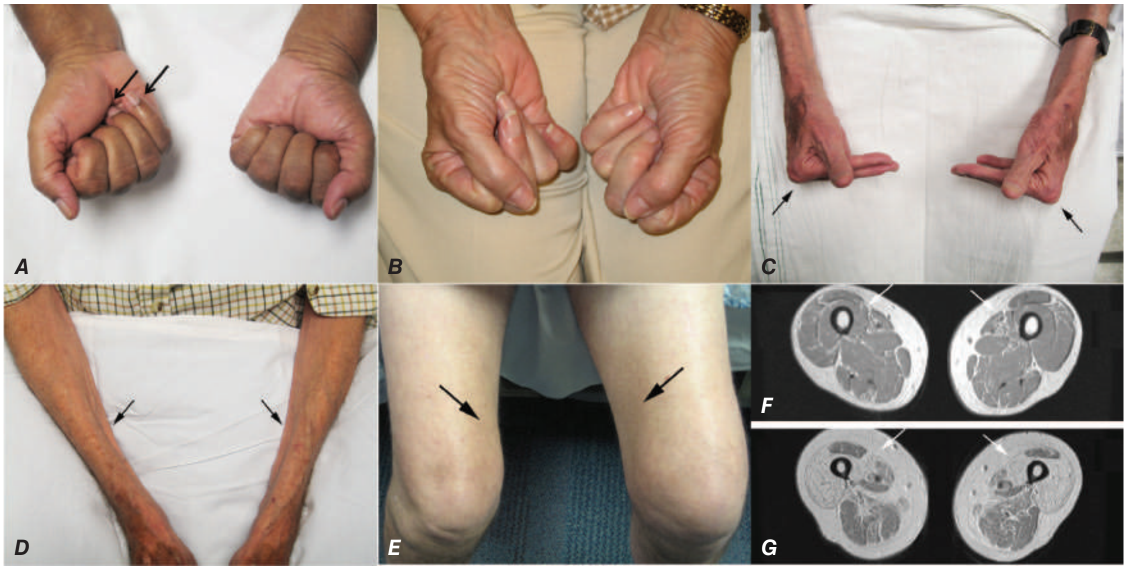

FIGURE 377-7 Muscle manifestations of inclusion body myositis (IBM; *A–C*). Finger flexor weakness can be (***A***) subtle and multifocal (*black arrows*), (***B***) moderate, or (***C***) severe. Note that even with complete paralysis of deep and superficial finger flexors, metacarpophalangeal joint flexion (*arrows*) is often maintained due to preservation of lumbricals. ***D.*** Ventral forearm atrophy (*arrows*). ***E.*** Atrophy of medial thighs due to loss of vastus medialis (*arrows*). ***F.*** Early IBM, with relatively preserved vastus medialis (*arrows*), in contrast to (***G***) advanced IBM with marked fibrous replacement of vastus medialis (*arrows*).

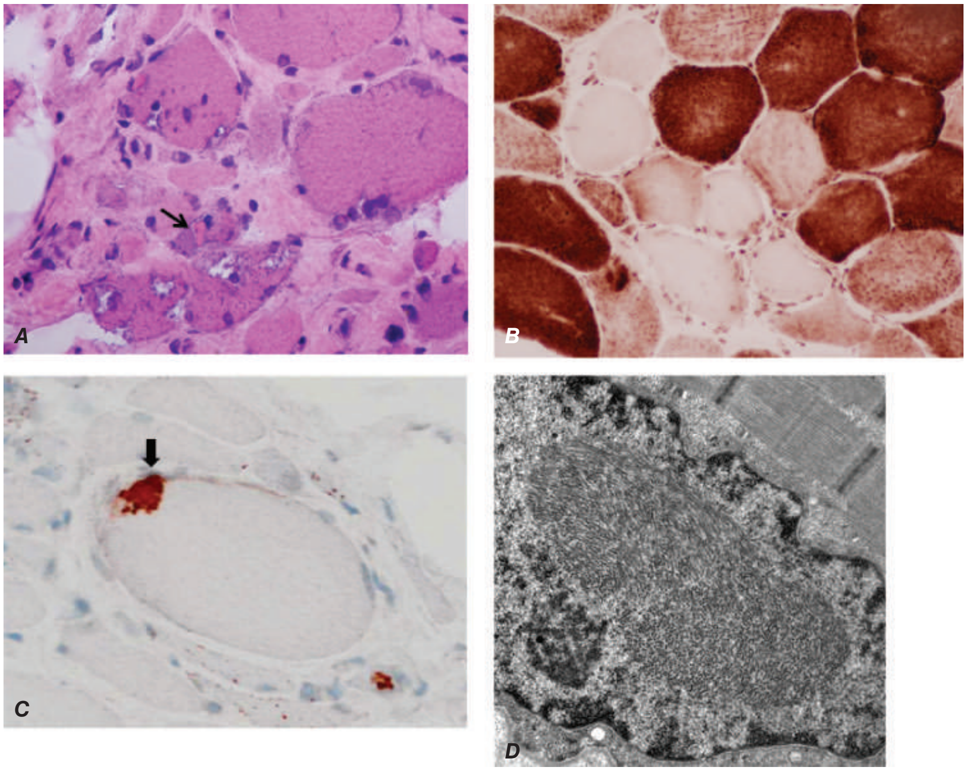

FIGURE 377-8 Pathology of inclusion body myositis. ***A.*** Scattered muscle fibers with rimmed vacuoles and rare fibers with eosinophilic inclusions (*arrow*), hematoxylin and eosin stain. ***B.*** Cytochrome oxidase stain demonstrates an increased number of pale-staining or COX-negative muscle fibers. ***C.*** Cytoplasmic inclusions stain positive with p62 within a muscle fiber (*thick arrow*). ***D.*** Electromicroscopy reveals 15- to 21-nm tubulofilamentous inclusions within a myonucleus.

helpful in distinguishing IBM from PM (aside from the clinical pattern of muscle weakness). Immunostaining also demonstrates that TAR DNA-binding protein 43 (TDP-43), an intranuclear RNA/DNA-binding protein involved in the regulation of RNA processing, is extruded from the myonuclei in IBM. This is similar to what is found in neurons of patients with neurodegenerative disorders such as frontotemporal dementia **(Chap. 443)** and amyotrophic lateral sclerosis **(Chap. 448)**.

The pathogenesis of IBM is poorly understood. The prominent adaptive immune system abnormalities related to T-cell inflammation and the presence of a relatively specific autoantibody against a muscle protein indicate an autoimmune attack on muscle. The chronic and highly inflammatory environment within muscles in IBM may alter protein synthesis and degradation pathways in part via aberrant immunoproteasome expression. Additional histologic features, typically referred to as "degenerative," include aggregation of various proteins including markers of endoplasmic reticulum (ER) stress and autophagy (e.g., p62 and LC3). Involvement of ER stress and autophagy has also been observed in other autoimmune diseases, such as primary biliary cholangitis (PBC), inflammatory bowel disease, and ankylosing spondylitis, some of which can be highly refractory to immunotherapy. As noted above, TDP-43, which is important for normal splicing of messenger RNA, is extruded from myonuclei in IBM; loss of TDP-43–mediated splicing repression likely leads to abnormal inclusion of cryptic exons in skeletal muscle and aberrant translation of muscle proteins. Whether this tissue damage results directly from a pathogenic immune response or a secondary neurodegenerative process is unclear at this time.

Prognosis The myopathy is slowly progressive and is not typically responsive to immunotherapies. Most patients require a scooter or wheelchair within 10–15 years of onset of symptoms.

TREATMENT OF INFLAMMATORY MYOPATHIES (TABLE 377-2)

DM, PM, ASyS, and IMNM are typically responsive to immunotherapy. High-dose glucocorticoids are considered the first-line treatment. There is uncertainty regarding when to start second-line agents (e.g., methotrexate, azathioprine, mycophenolate, immunoglobulin, or rituximab). The clinician must weigh with the patient the increased risks of immunosuppression versus possible benefits (e.g., faster improvement, steroid-sparing effects, and/or avoidance of morbidities associated with long-term glucocorticoid use). It is our general practice to start a second-line agent (typically methotrexate) with glucocorticoids in patients with severe weakness or other organ system involvement (e.g., myocarditis, ILD), those with increased risk of steroid complications (e.g., diabetics, osteoporosis, or postmenopausal women), and patients with IMNM who are known to have difficult-to-treat myositis. When treatment is initiated with prednisone alone, a second-line agent is added in patients who fail to significantly improve after 2–4 months of treatment or in those who cannot be tapered to a low dose of prednisone. Many patients with IMNM do not respond to prednisone alone or even prednisone plus a second-line agent in combination. Many require triple therapy with prednisone, methotrexate, and IVIG and, if this fails, rituximab. In our experience and that of others, anti-HMGCR myopathy often responds to monotherapy with IVIG.

Unfortunately, IBM does not typically respond to any known immunotherapy. The mainstay of treatment is physical and occupational therapy to improve function, and swallowing therapy (and sometimes esophageal dilation or cricopharyngeal myotomy) in those with dysphagia.

■ GENERAL GUIDELINES FOR USE OF SPECIFIC IMMUNOTHERAPIES

Glucocorticoids Treatment is initiated with prednisone (0.75–1.5 mg/kg up to 100 mg) administered as a daily morning single dose (the most common dose used in adults is 60 mg daily). In patients with severe weakness or comorbidities (e.g., ILD, myocarditis), treatment with a short course of intravenous methylprednisolone (1 g daily for 3 days) is recommended prior to starting oral glucocorticoids. Patients are generally maintained on high-dose prednisone until strength normalizes or until improvement in strength has reached a plateau (usually 3–6 months). Subsequently, prednisone can be tapered by 5 mg every 2–4 weeks. Once the dose is reduced to 20 mg every day or every other day,

TABLE 377-2 Immunotherapies for Inflammatory Myopathies

THERAPY	ROUTE	DOSE	SIDE EFFECTS	MONITOR
Prednisone	Oral	0.75–1.5 mg/kg per day to start	Hypertension, fluid and weight gain, hyperglycemia, hypokalemia, cataracts, gastric irritation, osteoporosis, infection, aseptic femoral necrosis	Weight, blood pressure, serum glucose/potassium, cataract formation
Methylprednisolone	Intravenous	1 g in 100 mL/normal saline over 1–2 h, daily or every other day for 3–6 doses	Arrhythmia, flushing, dysgeusia, anxiety, insomnia, fluid and weight gain, hyperglycemia, hypokalemia, infection	Heart rate, blood pressure, serum glucose/potassium
Azathioprine	Oral	2–3 mg/kg per day; single a.m. dose	Flu-like illness, hepatotoxicity, pancreatitis, leukopenia, macrocytosis, neoplasia, infection, teratogenicity	Blood count, liver enzymes
Methotrexate	Oral	7.5–20 mg weekly, single or divided doses; 1 day a week dosing	Hepatotoxicity, pulmonary fibrosis, infection, neoplasia, infertility, leukopenia, alopecia, gastric irritation, stomatitis, teratogenicity	Liver enzymes, blood count
	Subcutaneously	20–50 mg weekly; 1 day a week dosing	Same as oral	Same as oral
Cyclophosphamide	Oral Intravenous	1.5–2 mg/kg per day; single a.m. dose 0.5–1.0 g/m^2 per month × 6–12 months	Bone marrow suppression, infertility, hemorrhagic cystitis, alopecia, infections, neoplasia, teratogenicity	Blood count, urinalysis
Cyclosporine	Oral	4–6 mg/kg per day, split into two daily doses	Nephrotoxicity, hypertension, infection, hepatotoxicity, hirsutism, tremor, gum hyperplasia, teratogenicity	Blood pressure, creatinine/BUN, liver enzymes, cyclosporine levels
Tacrolimus	Oral	0.1–0.2 mg/kg per day in two divided doses	Nephrotoxicity, hypertension, infection, hepatotoxicity, hirsutism, tremor, gum hyperplasia, teratogenicity	Blood pressure, creatinine/BUN, liver enzymes, tacrolimus levels
Mycophenolate mofetil	Oral	Adults (1–1.5 g BID) Children (600 mg/m^2 per dose BID) (no >1 g/d in patients with renal failure)	Bone marrow suppression, hypertension, tremor, diarrhea, nausea, vomiting, headache, sinusitis, confusion, amblyopia, cough, teratogenicity, infection, neoplasia	Blood count
Intravenous immunoglobulin	Intravenous	2 g/kg over 2–5 days; then 1 g/kg every 4–8 weeks as needed	Hypotension, arrhythmia, diaphoresis, flushing, nephrotoxicity, headache, aseptic meningitis, anaphylaxis, stroke	Heart rate, blood pressure, creatinine/BUN
Rituximab	Intravenous	A course is typically 750 mg/m^2 (up to 1 g) and repeated in 2 weeks Courses are then repeated usually every 6–18 months	Infusion reactions (as per IVIG), infection, progressive multifocal leukoencephalopathy	Some check B-cell count prior to subsequent courses (but this may not be warranted)

Abbreviations: BUN, blood urea nitrogen; IVIG, intravenous immunoglobulin.

Source: Reproduced with permission from AA Amato, JA Russell (eds): *Neuromuscular Disorders*. 2nd ed. New York: McGraw-Hill Education; 2016.

the taper is slowed to 2.5 mg every 2–4 weeks. The goal is to taper prednisone to ≤10 mg daily. Although most patients improve, the response may not be complete and many will require at least a small dose of prednisone or a second-line agent to have a sustained remission. Serum CK levels are monitored; however, dose adjustments of prednisone and other immunotherapies are primarily based on the objective clinical examination and not the CK levels or the patients' subjective response. When no response is noted after an adequate trial of high-dose prednisone, alternative diagnoses (e.g., IBM or an inflammatory muscular dystrophy) and a repeat muscle biopsy should be considered.

Relapse of the myositis needs to be distinguished from steroid myopathy. Features suggesting a steroid myopathy include weakness developing while on high dosage, a normal serum CK, clinical features of steroid excess such as ecchymoses and "moon facies," and absence of muscle membrane irritability on EMG. By contrast, patients experiencing relapses of myositis may become weaker during the prednisone taper, have increasing serum CK levels, and display abnormal spontaneous activity on EMG.

Intravenous Immunoglobulin IVIG is most often used in patients refractory to prednisone and at least one second-line immunosuppressive agent. However, the ProDERM Trial Group found IVIG to be effective in a randomized, controlled trial in patients with DM, leading to U.S. Food and Drug Administration (FDA) approval. Thus, IVIG can be administered as first-line therapy in DM. In addition, IVIG is effective as a monotherapy in anti-HMGCR myopathy and may be the treatment of choice. A dose of 2 g/kg is divided over 2–5 days, and repeat infusions are given at monthly intervals for at least 3 months. Subsequently, intervals can be lengthened or dosage decreased: 2 g/kg every 2 months or 1 g/kg per month.

SECOND-LINE THERAPIES

Methotrexate Methotrexate is usually the second-line treatment of choice because most authorities believe it works faster than other agents. An oral dose of 5 or 7.5 mg/week is initiated and then gradually increased as needed up to 25 mg/week. If there is no improvement after 1 month of 25 mg/week of oral methotrexate, a switch to weekly parenteral (usually subcutaneous) methotrexate is the next step, with dose escalation by 5 mg weekly; only rarely is a dose >35 mg/week used. The major side effects of methotrexate are alopecia, stomatitis, ILD, teratogenicity, oncogenicity, risk of infection, and pulmonary fibrosis, along with bone marrow, renal, and liver toxicity. Patients are concomitantly treated with folate or folinic acid.

Azathioprine A recommended initial dose is 50 mg/d in adults, which can be increased by 50 mg every 2 weeks up to 2–3 mg/kg per day. Approximately 12% of patients develop a systemic reaction characterized by fever, abdominal pain, nausea, vomiting, and anorexia that requires discontinuation of the drug. The major practical limitation of azathioprine is that 6–18 months of treatment are usually required before benefit can be seen. Patients can be prescreened for thiopurine methyltransferase (TPMT) deficiency that is associated with severe bone marrow toxicity from this drug.

Mycophenolate Mofetil This drug inhibits the proliferation of T and B lymphocytes by blocking purine synthesis. It appears to be effective in different forms of myositis and is the second-line treatment of choice for myositis patients with ILD. The starting dose is 1.0 g twice daily and can be increased to 3 g daily in divided doses, if necessary. Mycophenolate is excreted through the kidneys; therefore, the dose should be decreased in patients with renal insufficiency. An advantage of mycophenolate compared to other immunosuppressive agents is the lack of renal or hepatic toxicity.

Rituximab Rituximab is a monoclonal antibody directed against CD20+ B cells. A large randomized controlled trial found no benefit, but there were flaws in the study design. Most authorities feel that rituximab can be beneficial in some patients who are refractory to prednisone and at least one of the other second-line agents. The typical dosage is 750 mg/m^2 (up to 1 g) IV with a second infusion 2 weeks later and with repeat courses (375 mg/m^2 as a single infusion or with a second infusion 2 weeks apart) every 6–18 months as needed.

Drugs in Clinical Trials Trials of a monoclonal antibody to block IFN-β and drugs to target downstream pathways (e.g., JAK-1) are ongoing in DM and other forms of myositis. Studies targeting the neonatal Fc receptor (FcRn) to decrease myositis-specific antibodies are underway in DM, ASyS, and IMNM, as are trials with chimeric antigen receptor (CAR) T-cell therapy. Ongoing trials in IBM include rapamycin as well as a trial of a monoclonal antibody binding to KLRG1 to deplete highly differentiated muscle-invading T cells.

MYOSITIS ASSOCIATED WITH CHECKPOINT INHIBITORS

Autoimmune neurologic complications, including inflammatory neuropathy **(Chap. 458)**, myasthenia gravis **(Chap. 459)**, and myositis, can occur with use of immune checkpoint inhibitors (anti-CTLA-4, anti-PD-1, and anti-PD-L1) to treat various cancers. Patients with myositis often develop muscle pain and weakness (axial musculature and proximal limbs) after one or two cycles. Myocarditis can also develop. Additionally, diplopia with extraocular weakness along with dysphagia and dysarthria suggesting the co-occurrence of myasthenia gravis (MG) may be present. In such cases, an elevated CK level helps support the diagnosis of myositis, while acetylcholine receptor antibodies or decremental response on slow repetitive nerve stimulation can establish the diagnosis of MG. Endomysial inflammatory cell infiltrates composed of macrophages expressing PD-L1 and CD8+ lymphocytes expressing PD-1, overexpression of MHC-I on sarcolemma of muscle fibers, and scattered necrotic and regenerating fibers can be found on muscle biopsies.

The immune checkpoint inhibitor should be discontinued, but most patients require concurrent treatment with glucocorticoids or IVIG. Patients generally improve over several months, during which time immunotherapy can be tapered. There are rare reports of patients with mild myositis who were able to be successfully re-treated with an immune checkpoint inhibitor.

MYOSITIS ASSOCIATED WITH COVID-19 INFECTION

Early series of patients hospitalized with COVID-19 report that as many as 44% of patients experience muscle pain or fatigue and 33% have elevated CK levels. Rare cases are complicated by myoglobinuria. Histopathology can demonstrate inflammatory cell infiltration and necrotic muscle fibers. In autopsy series, SARS-CoV-2 has not been demonstrated in muscle fibers, and the myositis is felt in most, if not all, to be due to cytokine storm. Whether COVID-19 infection can induce a chronic autoimmune myositis is controversial; there have been a number of reports of various types of IM, some with MSA, in patients following COVID-19 infection, but temporal associations do not equal causation.

GLOBAL ISSUES

There is a lack of epidemiologic data with regard to the incidence and prevalence of various subtypes of IM throughout the world. Complicating the issue is disease awareness and the inability to obtain and process muscle biopsies and MSAs, particularly in less developed countries. Nevertheless, each of these disorders occurs throughout the world. The specific environmental triggers and genetic risk factors are likely variable. Interestingly, a report from Japan found that 28% of IBM patients had evidence of exposure to hepatitis C, which was much higher than seen in the Western Hemisphere and also more common than seen in PM and healthy population controls in Japan. HIV-associated PM and IBM are more commonly encountered in areas endemic for HIV, and recent studies suggest most of these "PM" patients turn out to have IBM and can develop symptoms at an earlier age (e.g., in the 30s). Interestingly, while most cases of anti-HMGCR myopathy in North America and Europe are associated with prior statin use, the majority of such cases in Asia are not. Pyomyositis and parasitic myositis are clearly more common in the tropics. The prevalence of different types of cancers varies in different parts of the world, an important consideration with respect to paraneoplastic myositis. For example, nasopharyngeal cancer is particularly common in Asia; thus, assessment for this type of cancer should be considered in the workup of patients from high-risk regions.

■ FURTHER READING

Aggarwal R et al: ProDERM Trial Group. Trial of intravenous immune globulin in dermatomyositis. N Engl J Med 387:1264, 2022.

Allenbach Y et al: Immune-mediated necrotizing myopathy: Clinical features and pathogenesis. Nat Rev Rheumatol 16:689, 2020.

Amato AA, Russell JA (eds): *Neuromuscular Disorders*, 2nd ed. New York, McGraw-Hill Education, 2016, pp. 827–871.

Britson KA et al: Loss of TDP-43 function and rimmed vacuoles persist after T cell depletion in a xenograft model of sporadic inclusion body myositis. Sci Transl Med 14:eabi9196, 2022.

Greenberg SA: Inclusion body myositis: clinical features and pathogenesis. Nat Rev Rheumatol 15:257, 2019.

Julien S et al: Immune-mediated necrotizing myopathy (IMNM): A story of antibodies. Antibodies (Basel) 13:12, 2024.

Mariampillai K et al: Development of a new classification system for idiopathic inflammatory myopathies based on clinical manifestations and myositis-specific autoantibodies. JAMA Neurol 75:1528, 2018.

Müller F et al: CD19 CAR T-cell therapy in autoimmune disease: A case series with follow-up. N Engl J Med 390:687, 2024.

Pinal-Fernandez I et al: Pathogenic autoantibody internalization in myositis. medRxiv [Preprint]. 17:2024.01.15.24301339, 2024.

Puwanant A et al: Clinical spectrum of neuromuscular complications after immune checkpoint inhibition. Neuromuscul Disord 29:127, 2019.

Suh J et al: Skeletal muscle and peripheral nerve histopathology in COVID-19. Neurology 97:e849, 2021.

378

Relapsing Polychondritis

Marcela A. Ferrada, Peter C. Grayson

Relapsing polychondritis (RP) is a rare systemic disease characterized by recurrent inflammation in cartilaginous structures. Involvement of the ears, nose, respiratory tract, and joints are hallmark features of the disease; however, other organs can be affected including the eyes, inner ear, nervous system, skin, and cardiovascular system. Some patients with RP may be concomitantly diagnosed with other rheumatologic diseases, such as Sjögren's syndrome or systemic lupus erythematosus. Recurrent chondritis of the ears and nose in older male patients with severe systemic inflammation and progressive bone marrow failure is associated with acquired hematologic mutations in *UBA1*, a condition now known as the VEXAS (vacuoles, E1 enzyme, X-linked, autoinflammatory, somatic) syndrome. Prompt recognition and accurate diagnosis of RP or VEXAS syndrome are essential to prevent life-threatening complications of these diseases.

The epidemiology of RP is poorly characterized. The disease is commonly reported to primarily affect middle-aged adults without a strong sex predilection; some cohort studies report female predilection, but large, population-based studies typically show equal sex distribution. Children can be affected; however, these data are limited to case reports. Rare instances of familial aggregation have also been reported. The estimated incidence rate of disease is 0.7–3.5 cases per million per year, and prevalence estimates range from 4.5–25 cases per million. Epidemiologic data in RP are limited to older studies, typically evaluating population distributions more than three decades ago in cohorts from Europe and North America. Given the diagnostic challenges encountered in this condition, these data may underestimate the true prevalence of disease. Although RP has been reported in many regions of the world, whether racial or ethnic differences are associated with variable clinical features or outcomes is not known.

PATHOLOGY AND PATHOPHYSIOLOGY

Histologic findings from biopsy of affected cartilage are dependent on the timing of tissue sampling. In acute disease, a mixed inflammatory infiltrate of myeloid and lymphoid populations is observed at the cartilage interface. Over time, cartilage destruction is marked by lacunar breakdown and loss of chondrocytes, replaced by fibrosis. Biopsy of affected cartilage may cause morbidity, and there are no features on histology that are specifically diagnostic for RP. Therefore, biopsy is often reserved when there is a high suspicion of conditions that mimic RP, particularly infectious diseases and malignancies.

Although genetic, environmental, and immunologic aspects of the disease have been studied in humans and in animal models, the exact mechanisms that drive the disease are unclear. Limited studies in animal models support the concept that autoimmunity to cartilage components may play a role in RP. Immunization of specific animals to type II collagen, type IX collagen, or matrillin-1 can recapitulate various aspects of the clinical phenotypes observed in patients with RP; however, the performance characteristics of these antibodies in clinical practice are poor to differentiate RP from other diseases. Autoantibodies to type II collagen have been reported in small cohorts of patients with RP but are neither sensitive nor specific to be considered as an acceptable diagnostic marker of disease. Observational cohort data suggest that B-cell–depleting therapies are not particularly effective to treat the condition, arguing against a strong primary role of antibody-mediated disease.

Acute phase reactants (e.g., erythrocyte sedimentation rate or C-reactive protein) are not reliably elevated in patients with RP. Proinflammatory cytokines and chemokines related to both innate and adaptive immunity have been identified in association with RP; however, these findings are not consistent and, therefore, are not routinely measured to guide management in clinical practice.

CLINICAL MANIFESTATIONS AND ORGAN INVOLVEMENT

RP is associated with a wide range of clinical manifestations. In the absence of a diagnostic blood test, a detailed medical history and physical examination are the most essential tools for diagnosis and, most importantly, further management. Because disease activity is intermittent and patients may not exhibit clear signs of inflammation during direct evaluation, it is crucial to consider the patient's symptoms over time and to review any provided photographic documentation of disease features.

Ears, Nose, and Throat Ear involvement is one of the most common clinical manifestations of relapsing polychondritis, affecting 80–90% of the patients. However, there are no clear definitions of ear chondritis, and patient descriptions of pain and associated symptoms are variable. The typical description includes involvement of only the cartilaginous aspect of the ear (pinnae) **(Fig. 378-1*A* and *B*)**. However, a subset of patients may report pain in the ears with minimal associated findings visualized on physical examination, such as mild swelling or erythema **(Fig. 378-1*C*)**. Although damage to the cartilaginous portion of the external ear (i.e., "cauliflower ear") is observed in RP **(Fig. 378-1*D*)**, it is present only in a small number of patients, usually in the setting of recurrent episodes of ear inflammation. Questions that can be helpful to elicit the nature of ear pain include asking about onset, duration, and triggers. Common triggers include minimal trauma to the ear (such as from lying on the effected side or wearing glasses) or temperature changes.

Patients usually describe a sensation of pressure at the bridge of the nose that may not be accompanied by other associated symptoms. Some patients may have overt redness and swelling of the bridge or tip of nose, but this is less common **(Fig. 378-2*A*)**. While a "saddle nose" has been described as a typical clinical finding in RP, many patients will not develop an obvious nasal deformity **(Fig. 378-2*B*)**. The nasal septum and nasal passages should always be evaluated in patients with suspected RP, as many patients can have septal ulcers and mucosal inflammation. In contrast to granulomatosis with polyangiitis (GPA) where a saddle nose deformity is usually associated with a septal perforation **(Chap. 375)**, patients with RP who have a saddle nose deformity usually do not have

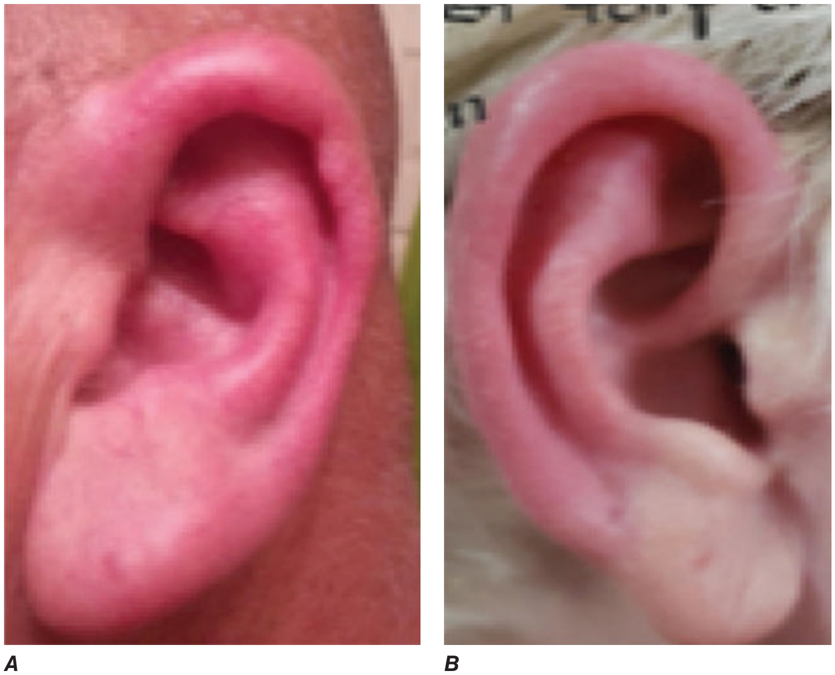

FIGURE 378-1 Representation of ear involvement in relapsing polychondritis. ***A and B.*** Examples of typical ear chondritis, with significant inflammation of the pinnae. ***C.*** Mild ear cartilage inflammation. ***D.*** Cartilage damage.

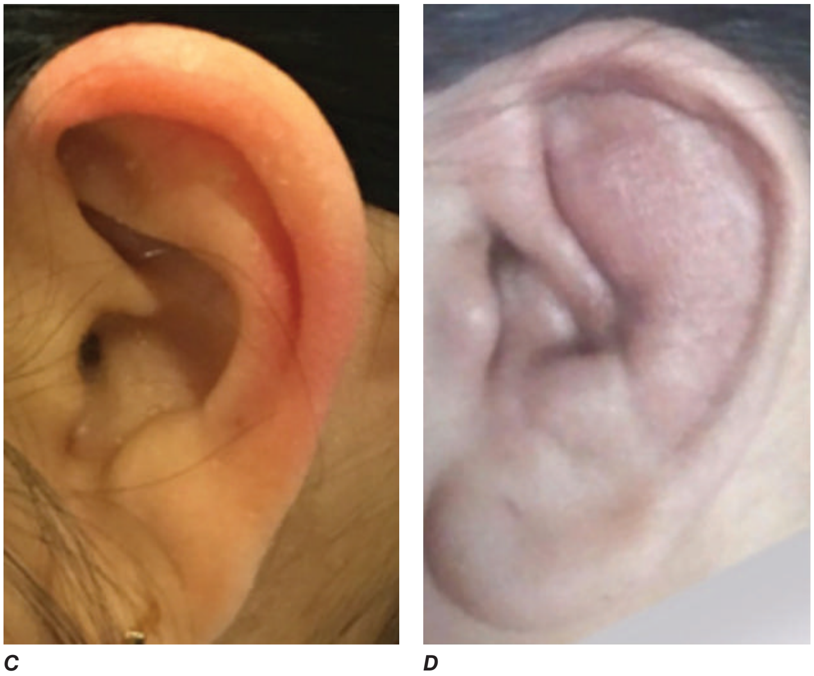

FIGURE 378-1 (Continued)

a septal perforation, as most of the nasal perforations in patients with RP are located anteriorly.

Patients can have complaints of throat pain or a sensation of globus, usually described as a "choking sensation." Patients can also report pain on the anterior aspect of the neck, in some cases associated with erythema, usually located at the level of the thyroid cartilage.

Upper and Lower Airway Conditions like subglottic stenosis may manifest acutely or subacutely, leading to cough, voice changes, or breathlessness. Severe narrowing can result in stridor, requiring urgent medical attention to avoid mortality **(Fig. 378-3*A*)**. Often misdiagnosed as adult-onset asthma, timely recognition of airway involvement is important to prevent chronic damage to the large airways, which can result in tracheomalacia, bronchomalacia, or tracheal calcification **(Fig. 378-3*B* and *C*)**. Patients with intermittent episodes of wheezing and with a review of symptoms positive for possible RP should be followed closely and further evaluated for lower airway involvement.

Musculoskeletal • **JOINTS** RP is associated with a nonerosive inflammatory arthritis that can affect small and large peripheral joints and as well as the axial skeleton. Erosions have been described in the sacroiliac joints but not in other joints. The most common large joint affected is the knees, often presenting with effusions. The pattern of small joint involvement emulates rheumatoid arthritis, without erosions or deformities.

TENOSYNOVITIS Different tendons can be affected including Achilles and peroneal tendons.

COSTOCHONDRAL JOINTS Located at the rib-sternum junction, these joints are frequently inflamed in RP, producing constant pain that

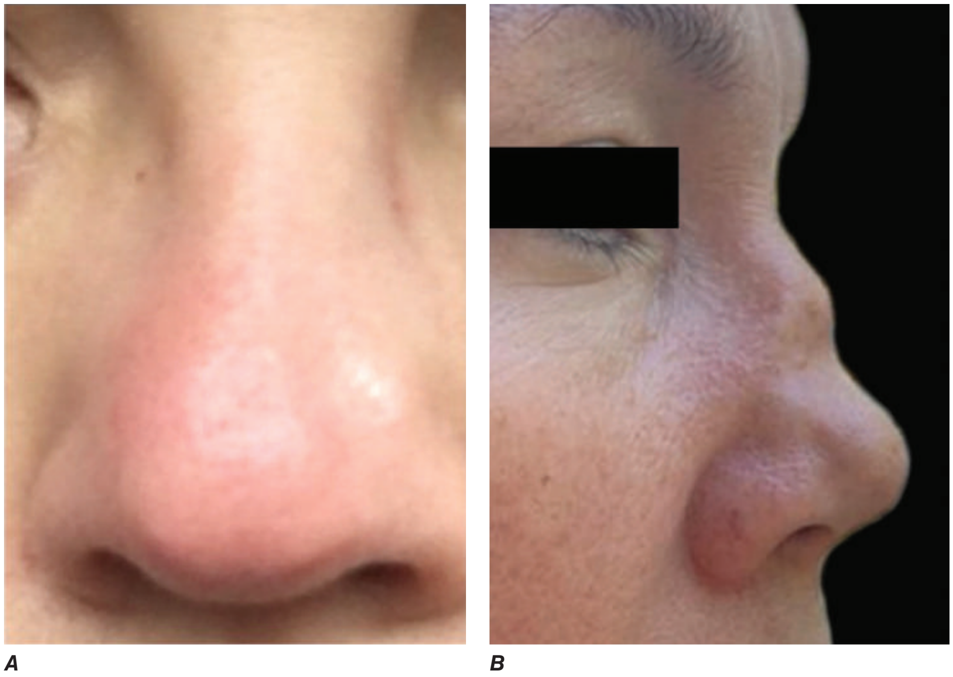

FIGURE 378-2 Nose involvement in relapsing polychondritis. *A.* Redness and swelling of the base and tip of the nose. ***B.*** Saddle nose deformity.

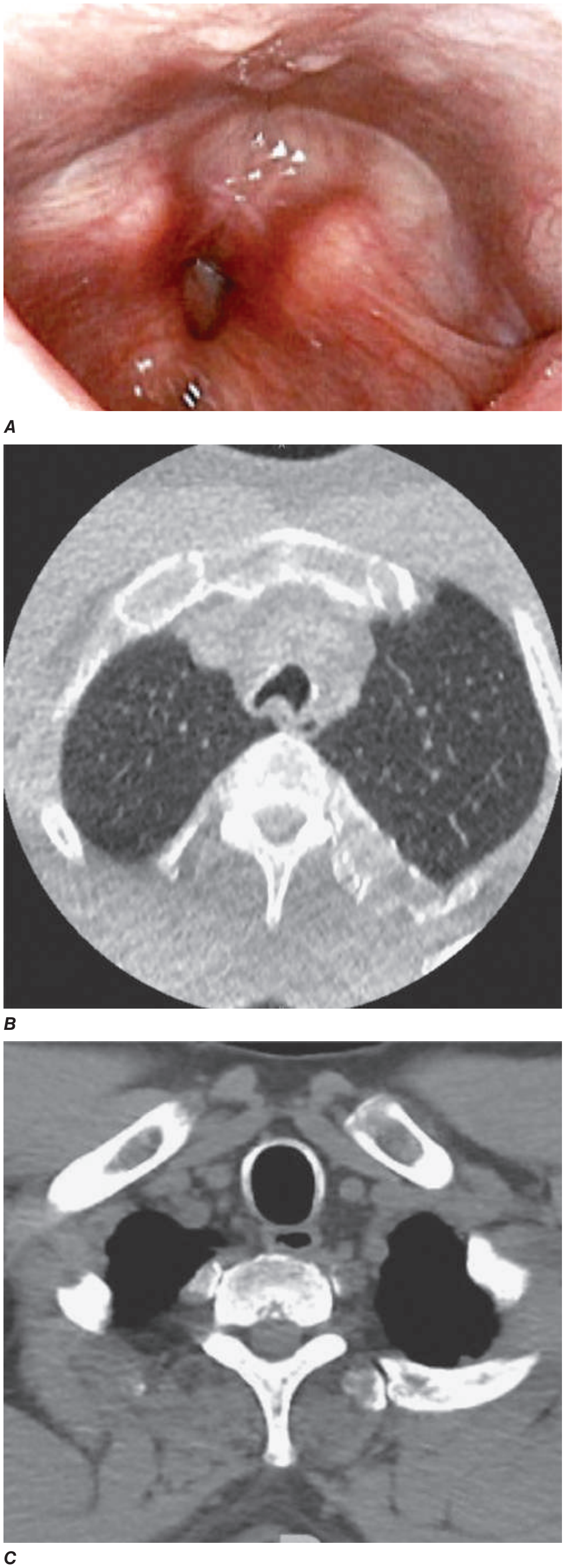

FIGURE 378-3 Airway involvement in relapsing polychondritis. ***A.*** Subglottic stenosis. ***B and C.*** Computed tomography demonstrating tracheal calcification and tracheomalacia.

stands out due to its unique severe quality. Pain due to costochondritis can be so severe that patients will seek evaluation in the emergency room due to concerns for an acute coronary event. The pain is bilateral, reproducible with palpation, and not associated with movement of the affected rib.

Inner Ear All patients with RP should undergo hearing tests to evaluate for conductive and sensorineural hearing loss. During the physical exam, the ear canal should be evaluated as some patients can have severe inflammation or eustachian tube dysfunction leading to conductive hearing loss. The presence of dizziness necessitates further evaluation through maneuvers like finger-to-nose, Romberg, and nystagmus to rule out vestibular involvement.

Ocular Involvement Episcleritis is the most frequent type of ocular inflammation in RP. It might coincide with other flare symptoms or occur independently. Although rare, features suggesting scleritis mandate urgent ophthalmologic evaluation due to its potential catastrophic outcomes, including vision loss, scleromalacia, or global rupture necessitating enucleation **(Fig. 378-5*A*)**.

Cardiovascular Patients with mouth and genital ulcers with inflamed cartilage (MAGIC) syndrome may exhibit symptoms seen in Behçet's syndrome **(Chap. 376)**, including pathergy and gastrointestinal involvement. These patients can also have large-vessel vasculitis involving the aorta.

Neurologic The diagnosis of neurologic involvement due to RP is one of exclusion. All possible etiologies including infection and malignancy should be ruled out before attributing the clinical findings to RP; however, encephalitis and meningitis can be seen in RP.

DIAGNOSIS

There are no circulating or tissue-based biomarkers with adequate performance characteristics to function as a diagnostic test for RP. As such, diagnosis is based primarily on clinical pattern recognition informed by a comprehensive review of systems and physical examination. Several diagnostic criteria have been proposed; however, these criteria were developed in small cohorts, are largely based on expert opinion, and have not been formally validated. *McAdam's criteria* require presence of three of six symptoms, including bilateral auricular chondritis, nonerosive seronegative inflammatory polyarthritis, nasal chondritis, ocular inflammation, respiratory tract chondritis, or vestibular/cochlear dysfunction. *Damiani and Levine* modified McAdam's criteria to include either three clinical features or one clinical feature with histologic evidence of chondritis or two clinical features with clinical response to glucocorticoids, dapsone, or both. *Michet's criteria* propose three major criteria (inflammation of ear, nose, or respiratory tract) and four minor criteria (ocular inflammation, hearing loss, vestibular dysfunction, and seronegative inflammatory arthritis) and require fulfillment of two major criteria or one major plus two minor criteria to establish the diagnosis. Another important limitation of these criteria is the lack of precise definitions of organ involvement.

The differential diagnosis for RP includes conditions that mimic chondritis in specific cartilaginous organs and systemic diseases for which chondritis may be a manifestation. *Ear chondritis* can be mimicked by infectious and cutaneous diseases of the external ear, traumatic otohematoma, or red ear syndrome. Various infectious diseases and angiocentric centrofacial lymphoma can mimic *nasal chondritis*. Asthma, traumatic airway stenosis, infectious diseases, and congenital disorders involving the airway can mimic *airway chondritis*. Systemic diseases that mimic RP include GPA **(Chap. 375)** and autoinflammatory diseases **(Chap. 381)**. GPA can be differentiated from RP in part by the presence of antineutrophil cytoplasmic autoantibodies (ANCA) and glomerulonephritis, which are not features of RP. Presence of cytopenia, most notably macrocytosis and lymphopenia, in older male patients should trigger consideration of *genetic testing* and *bone marrow assessment* for VEXAS.

Audiometry to assess hearing loss and *dynamic expiratory phase computed tomography of the chest* to assess tracheobronchomalacia are critical studies that should be performed in all patients with suspected

RP to define the extent of disease involvement. *Direct laryngoscopy* can be useful to visualize cartilage damage and inflammation in the upper airway; however, *bronchoscopy* to visualize the lower airway can exacerbate disease and should be performed with caution. *Echocardiography* can assess valvular heart disease. *Fluorodeoxyglucose–positron emission tomography* may have utility to detect inflammation in the large arteries and the larger airways. *Biopsy of affected cartilage* often yields nonspecific findings and should be reserved to exclude mimicking conditions.

TREATMENT

Relapsing Polychondritis

Patients with RP can present to diverse specialists prior to diagnosis that can include multiple emergency room visits. Internal medicine doctors are typically among the first to encounter patients who present with the initial manifestations of RP and are instrumental in starting the initial treatment regimen.

Because RP is a systemic and heterogeneous disease, clinical management should typically involve multidisciplinary teams, including otolaryngologists (ear, nose, and throat specialists) and pulmonologists. Given the disease's propensity to progressively involve various organs, it is crucial to perform initial testing to define organ damage. If possible, patients should be referred to a rheumatologist for comprehensive evaluation, diagnosis confirmation, and the establishment of a tailored treatment plan.

Glucocorticoids are the cornerstone of initial treatment for inflammation in RP, with dosages tailored to the severity of the disease and the specific organs involved. In cases where life-threatening organ involvement, such as the airway, is present, high doses of glucocorticoids (e.g., 1 mg/kg of prednisone or an equivalent glucocorticoid) should be rapidly initiated. For patients with milder disease, lower doses of glucocorticoids can be trialed. Different disease-modifying agents can be used to treat the disease including methotrexate, leflunomide, mycophenolate, anti-interleukin-6, or tumor necrosis factor inhibitors. Due to the heterogeneity of the disease, treatment strategies are often determined by the predominant clinical phenotype.

■ FURTHER READING

Beck DB et al: Somatic mutations in UBA1 and severe adult-onset autoinflammatory disease. N Engl J Med 383:2628, 2020.

Ferrada M et al: Defining clinical subgroups in relapsing polychondritis: A prospective observational cohort study. Arthritis Rheum 72:1396, 2020.

379

Sarcoidosis

Alicia K. Gerke

Sarcoidosis is a systemic inflammatory disease of unknown cause characterized by the formation of nonnecrotizing granulomatous inflammation. Sarcoidosis affects the lung in >90% of cases but can afflict almost any other organ system as well. The disease was first described >150 years ago by Dr. Jonathan Hutchinson of London, England, as a skin malady of symmetrical purple plaques (later termed lupus pernio) and raised red lesions (Mortimer's malady). In 1899, the skin lesions were noted to be “sarkoid” by Caeser Boeck, a Norwegian dermatologist, as they resembled sarcoma, albeit the lesions were largely benign and granulomatous in their histology. The advent of imaging and other clinical technology later unveiled the multiorgan involvement associated with sarcoidosis. However, despite even modern advances in immunology, genetics, and genomics, there is still considerable mystery regarding the exact etiology of the inflammation and the factors that contribute to variability of disease presentation and clinical course. Because of this considerable heterogeneity and lack of large, randomized treatment trials, decisions regarding diagnosis and management continue to be challenging.

EPIDEMIOLOGY

Sarcoidosis occurs throughout the world, affecting all genders, races, and ages. Epidemiologic studies regarding the incidence, prevalence, and mortality rates of sarcoidosis are difficult to conduct and compare; this is due to the potential for nonstandardized case definition, differing methods of case acquisition, and lack of known etiology. Regional variabilities due to race and gender, as well as phenotypic variability, further confound epidemiologic measurements. Taken as a whole, epidemiologic studies highlight the variation in disease globally.

■ INCIDENCE AND PREVALENCE

The incidence and prevalence of sarcoidosis vary by race, gender, geographic region, ethnicity, and season. An increased familial risk has been noted. The highest incidence in the world occurs in African Americans (35.5 per 100,000) and Northern European populations (24 per 100,000) as described by studies from the United States and Sweden. A higher incidence occurring in women is consistent across most studies, although the ratio does not usually exceed 2:1. The U.S. Black Women's Health Study cohort found a particularly high annual incidence of 71 per 100,000 and a prevalence of 2%. Some studies suggest an increase in incidence over time in the populations of some countries such as the United States and Japan. Peak incidence of the disease occurs between 35 and 45 years of age, but a significant proportion of cases do happen after age 55, with men manifesting disease earlier than women. The overall perceived incidence rate may be underestimated, as identification of subclinical sarcoidosis may only be possible in screening population studies or at autopsy. Certain occupations and exposures and an elevated body mass index have been identified as risk factors for development of sarcoidosis, whereas smoking is associated with reduced odds of incident sarcoidosis. Seasonal variation is also reported, with a predominance of diagnoses occurring in the springtime.

Differences in presentation and severity are also seen among populations. Cardiac and ocular involvement are more commonly observed in Japanese populations, whereas Europeans tend to develop erythema nodosum, which is uncommon among Japanese and black populations. Several studies suggest that overall severity is higher in black populations, with black patients presenting with more frequent involvement of extrapulmonary disease and three times the likelihood of familial disease. Women are at increased risk for developing erythema nodosum and ocular, musculoskeletal, and neurologic disease, while men are at higher risk for hypercalcemia. While most sarcoidosis patients have lung involvement, black and female patients tend to have more severe lung disease. Finally, lower income has also been associated with severity of disease and potentially greater disability and impairment (particularly increased dyspnea and onset of new organ involvement), suggesting that socioeconomic status may impact disease course.

■ MORTALITY

Mortality rates vary considerably based on the study setting, with population screenings noting lower rates of mortality than those reported from referral centers. These occurrences may reflect differences in disease severity. Analysis of death certificates from the United States shows an average age- and gender-adjusted mortality rate of 4.32 per 1,000,000. Mortality rates are higher in women and in African Americans. Pulmonary fibrosis is the leading cause of death in this population. Lower mortality rates are seen in Japan, with estimates at 0.1 per million people. In Japan, mortality from sarcoidosis itself is usually attributed to cardiac involvement, whereas in Europeans and Americans, respiratory mortality predominates.

ETIOLOGY

The cause of sarcoidosis is unknown; however, current evidence strongly indicates a heightened host immune response to an undetermined environmental exposure in a genetically susceptible individual. It is unknown if there are multiple different antigens that induce the same disease process. The immunologic response in sarcoidosis appears to be physiologic, although it is either exaggerated or is deficient in its regulatory mechanisms. The transmissibility of the immune response also supports the presence of an antigen. For example, the Kveim-Siltzbach reagent (historically used to diagnose sarcoidosis) is a homogenate of human splenic tissue from patients with sarcoidosis that, when injected into the skin, induces a systemic granulomatous response in patients with early sarcoidosis. In a similar fashion, transplant recipients can develop a sarcoid-like inflammatory response after receiving organ tissue from another person with sarcoidosis.

Histologic inflammatory responses like sarcoidosis occur in the lungs of patients with known inorganic and organic exposures such as beryllium, mycobacteria, or fungi, resulting in these antigens being studied as potential causes. Mycobacterial proteins (e.g., mKatG, ESAT-6, mycobacterial superoxide dismutase A, antigen 85A, heat shock proteins) have been found disproportionately in sarcoidosis granulomas. Heightened macrophage immune responses against these peptides within blood and bronchoalveolar lavage (BAL) have also been noted in patients with sarcoidosis. In a similar manner in other cohorts, *Cutibacterium acnes* (formerly *Propionibacterium acnes*) has been implicated. A genetic predisposition is also suspected, as antigen recognition and presentation molecule polymorphisms have been associated with sarcoidosis development and phenotype. Similarly, gene–environment interactions have been identified, suggesting a complex interaction of a susceptible host in a predisposing environment.

Epidemiologically, the role of antigens in pulmonary sarcoidosis is further supported by case-cluster events and occupational association. Whereas numerous familial case clusters support a genetic predisposition to sarcoidosis in at least a proportion of cases, spatial clustering of nonrelated sarcoidosis cases suggests exposure to an inciting environmental antigen. For example, a baseline higher incidence is noted in U.S. firefighters, with an annual incidence of sarcoidosis occurring in 13–15 per 100,000 individuals. Following the World Trade Center event and building collapses, the rates of "sarcoid-like" disease were 86 per 100,000 in the first year and 22 per 100,000 over the next 4 years, with the highest risk associated with working in the debris piles. Interestingly, the cases included not only those with lung disease but also those with multisystem involvement, which would not be expected from a localized inhalational exposure. Additionally, employment in agriculture, teaching, metalworking, radiation, cotton ginning, and automobiles has also been associated with sarcoidosis.

Increased rates of sarcoidosis have been found in association with environmental exposures. Musty or humid environments, mold and mildew, insecticides, air conditioning, raising birds, and exposure to wood stoves/burning and fireplaces have been associated with increased risk of sarcoidosis. On the other hand, smoking is consistently a negative predictor of disease; the mechanism of this is unknown, although it is perhaps related to immunomodulatory effects of nicotine and smoke. These epidemiologic studies would indicate that the inciting exposure may be a bioaerosol or an inorganic agent, which may explain some of the epidemiologic variability of clinical presentation by race, age, gender, and ethnicity. Additionally, the inciting antigen may not be a single agent; rather, differing antigens may exist for each individual or population.

PATHOGENESIS

Like other granulomatous processes, granulomas of sarcoidosis are presumably forming around a poorly degradable or insoluble material that is identified and presented to lymphocytes by dendritic cells or macrophages via the major histocompatibility complex **(Fig. 379-1)**. The initial innate immune response triggers the production of cytokines and chemokines that begin the cascade of granulomatous inflammation, attracting cells of the adaptive immune response. Activated lymphocytes and mononuclear cells migrate to the site of granulomatous inflammation in a process driven by amplification of oligoclonal T cells of undefined antigenic specificity. Type 1 (T_H1) immune polarization is the immunopathogenic hallmark of sarcoidosis, characterized by highly polarized CD4+ T_H1 lymphocytes and less so by CD8+ lymphocytes,

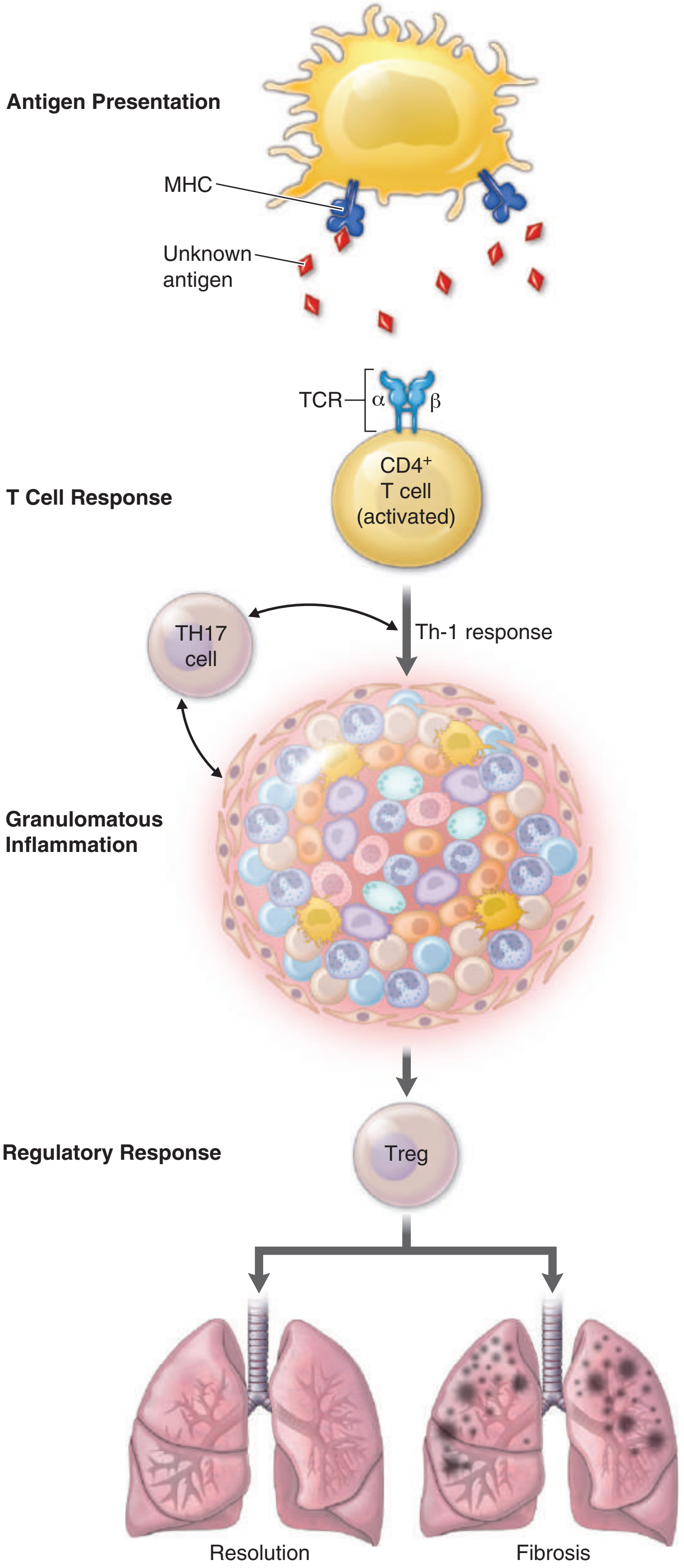

FIGURE 379-1 Immunopathogenesis of sarcoidosis. Granulomatous inflammation of sarcoidosis is presumed to be due to a heightened immune response to an undetermined antigen in a genetically predisposed individual. The antigen presentation via the major histocompatibility complex (MHC) activates CD4+ T cells, leading to a polarized T_H1 response. Activated lymphocytes, macrophages, and monocytes migrate to sites of inflammation, leading to a cascade of tightly formed noncaseating granulomas. Altered T-regulatory (Treg) and T_H17 responses are thought to be involved in lack of resolution of granulomatous inflammation or progression to fibrosis in some cases.

producing interferon-gamma (IFN-γ). T_H1-promoting immunoregulatory cytokines and chemokines (interleukin [IL]-1, IL-2, IL-15, IL-18, CXCR3) are consistently upregulated at sites of inflammation, propagating the immune response, activating macrophages, and forming granulomas. The accumulation of inflammatory cells at granuloma formation sites appears to be the result of both in situ proliferation and redistribution of the peripheral blood to the lung. Alterations in regulatory T cells and T_H17 cells may also play a role in the body's inability to control the inflammation. Although most patients with sarcoidosis eventually recover, the critical step as to why some patients with sarcoidosis transition to fibrosis is not understood.

CLINICAL PRESENTATION

The clinical presentation of patients varies considerably from asymptomatic to progressive organ dysfunction. Acute-onset presentation is seen in Löfgren syndrome in up to 10% of sarcoidosis cases (fever, erythema nodosum, ankle arthralgias, bilateral hilar lymphadenopathy) and, less commonly, Heerfordt syndrome (facial nerve palsy, fever, anterior uveitis, and bilateral enlargement of the parotid glands). These acute presentations typically have an excellent prognosis with full spontaneous resolution. In subacute or chronic cases, symptoms such as chronic cough, dyspnea, atypical chest pain, or intermittent joint or muscle pain tend to have a more insidious onset. In rare cases, presentation of sarcoidosis can be fatal, usually related to heart failure, sudden cardiac death, or neurologic involvement.

Symptoms vary depending on organ involvement and intensity of inflammatory response; therefore, patients can present to different clinical specialists. Nonspecific symptoms of fevers, malaise, night sweats, and weight loss are often present. More fulminant and organ-threatening symptoms can include acute hypercalcemia, blindness, arrhythmias, heart block, or neurologic demise, necessitating hospitalization. Cardiopulmonary symptoms such as cough, chest pain, or shortness of breath can mimic more common diseases, contributing to delays in diagnosis. In many cases, findings of sarcoidosis are found incidentally on imaging, prompting further workup.

ESTABLISHING A DIAGNOSIS

The diagnosis of sarcoidosis is made using three main criteria: (1) a compatible clinical presentation, (2) the presence of nonnecrotizing granulomatous inflammation in one or more tissues, and (3) exclusion of alternative causes of granulomatous inflammation. The diagnostic workup for sarcoidosis should also include assessment of extent and severity of organ involvement, a review of prognostic factors, and a determination of whether therapy will benefit the patient.

If a compatible clinical picture or radiographic finding suggests sarcoidosis, then biopsy should be considered as the next step. There are no well-established, clinically useful diagnostic serum biomarkers that can secure a diagnosis. Noninvasive sites for biopsy can include skin, conjunctiva, or superficial lymph nodes, but in cases where lung involvement and mediastinal/hilar lymphadenopathy are the prominent feature, bronchoscopy is the preferred method for diagnosis. Endobronchial ultrasound-guided lymph node sampling has high diagnostic yield and low procedural risk. BAL studies of lymphocyte subpopulations can be performed and characteristically reveal an elevated percentage of lymphocytes. A CD4:CD8 lymphocyte ratio >3.5 supports a diagnosis of sarcoidosis, particularly when combined with consistent clinical and radiographic findings (albeit a normal ratio does not exclude sarcoidosis). BAL is also important to exclude infection. Granulomas can line all segments of airway walls, providing the "cobblestone" appearance seen during bronchoscopy. If needed, transbronchial lung tissue biopsies can also be done to confirm sarcoidosis with lung infiltrates. Endoscopic ultrasound through the esophagus can be considered as an alternative biopsy approach based on clinical presentation and location of inflammation. Rarely, surgical lung biopsy or mediastinoscopy may be indicated when less invasive procedures have proven nondiagnostic and no other accessible sites for biopsy have been identified. ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET)–computed tomography (CT) can also be helpful to identify active tissue inflammation and a site to biopsy in difficult cases.

In some cases, the clinical presentation can be so compelling that lymph node sampling may not be necessary. For example, this can occur in patients with classic Löfgren syndrome, Heerfordt syndrome, or lupus pernio. Similarly, deferring biopsy in asymptomatic patients with symmetrical bilateral hilar lymphadenopathy may be considered based on patient-oriented discussions. In all cases, however, close clinical follow-up is still recommended, as the diagnosis of sarcoidosis is never fully secure.

PATHOLOGY

The sarcoidosis granuloma is an organized collection of macrophages, epithelioid cells, and multinucleated giant cells (which form as the epithelioid cells fuse). These cells are surrounded by a well-circumscribed rim of lymphocytes and fibroblasts. CD4+ T lymphocytes are also found interspersed within the granuloma center, while CD8+ T lymphocytes and B lymphocytes reside at the periphery. Sarcoidosis granulomas are nonnecrotizing; however, punctuate necrosis may be present on rare occasions. More abundant necrosis strongly indicates alternative causes of granulomatous inflammation such as infection, and loosely formed nonnecrotizing granulomas would suggest an alternative diagnosis such as hypersensitivity pneumonitis **(Fig. 379-2)**. Additionally, inclusions occasionally appear in sarcoidosis granulomas; these can include asteroid bodies, birefringent calcium carbonate or oxalate crystals, Schaumann bodies, and Hamazaki-Wesenberg bodies (especially in lymph nodes). Although most inflammatory granulomas tend to resolve, hyalinization and fibrosis of the granulomas can occur.

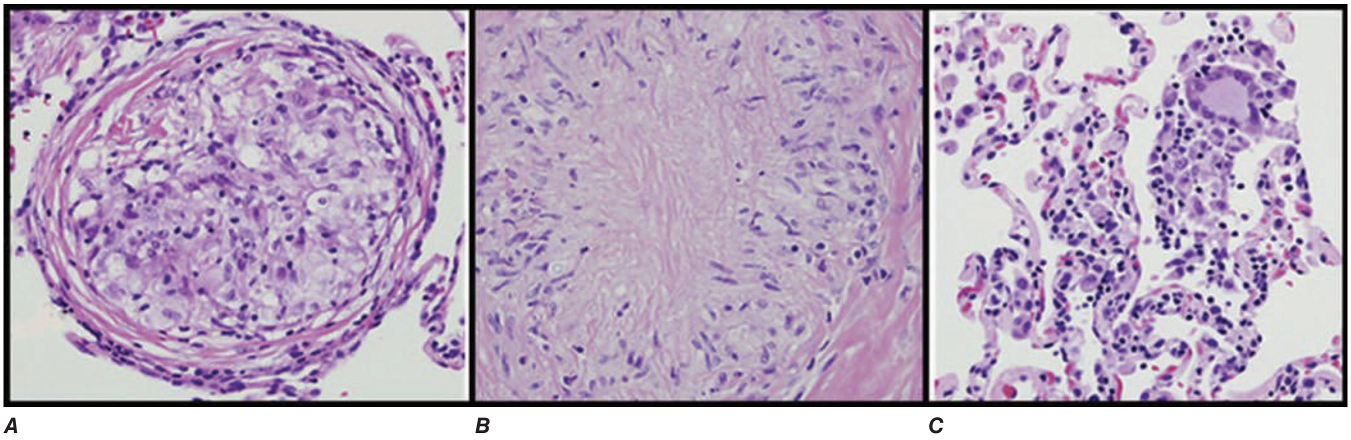

FIGURE 379-2 Comparison of granulomas. (***A***) The sarcoidosis granuloma is tightly formed and nonnecrotizing. (***B***) The infectious granuloma as seen in tuberculosis or fungal infections is tightly formed but with a necrotizing center. (***C***) The granuloma in hypersensitivity pneumonitis is loosely formed and without necrosis.

TABLE 379-1 Common Exclusionary Causes of Granulomatous Disease in the Diagnosis of Sarcoidosis

INFECTIONS	OTHER INFLAMMATORY DISEASES	MALIGNANCY RELATED
Mycobacterial infection • Tuberculosis • Atypical mycobacteria Fungal infection • Histoplasmosis • Coccidiomycosis • Blastomycosis • Aspergillosis • Cryptococcus Parasitic infection • Toxoplasmosis • Schistosomiasis Other bacterial and viral infection (less common) • Herpes zoster • Tularemia (*Francisella tularensis*) • Q fever (*Coxiella burnetii*) • *Bartonella henselae*	Environmental exposure • Hypersensitivity pneumonitis • Aspiration pneumonitis • Chronic beryllium disease • Silicosis • Talc inhalation or injection • Local or systemic reaction to tattoo ink Medication sarcoid-like reaction • Immune checkpoint inhibitors • Antiretroviral therapy • Interferon • TNF-α antagonist Autoimmune/Inflammatory • ANCA-associated vasculitis • Granulomatous-lymphocytic interstitial lung disease (associated with common variable immunodeficiency) • IgG4-related disease • Rheumatoid nodules • Inflammatory bowel disease • Bronchocentric granulomatosis	Malignancy • Local granulomatous reaction surrounding tumor • Systemic sarcoid-like reaction to malignancy Chemotherapy or immunotherapy • Sarcoid-like reaction

Abbreviations: ANCA, antineutrophil cytoplasmic antibodies; TNF, tumor necrosis factor.

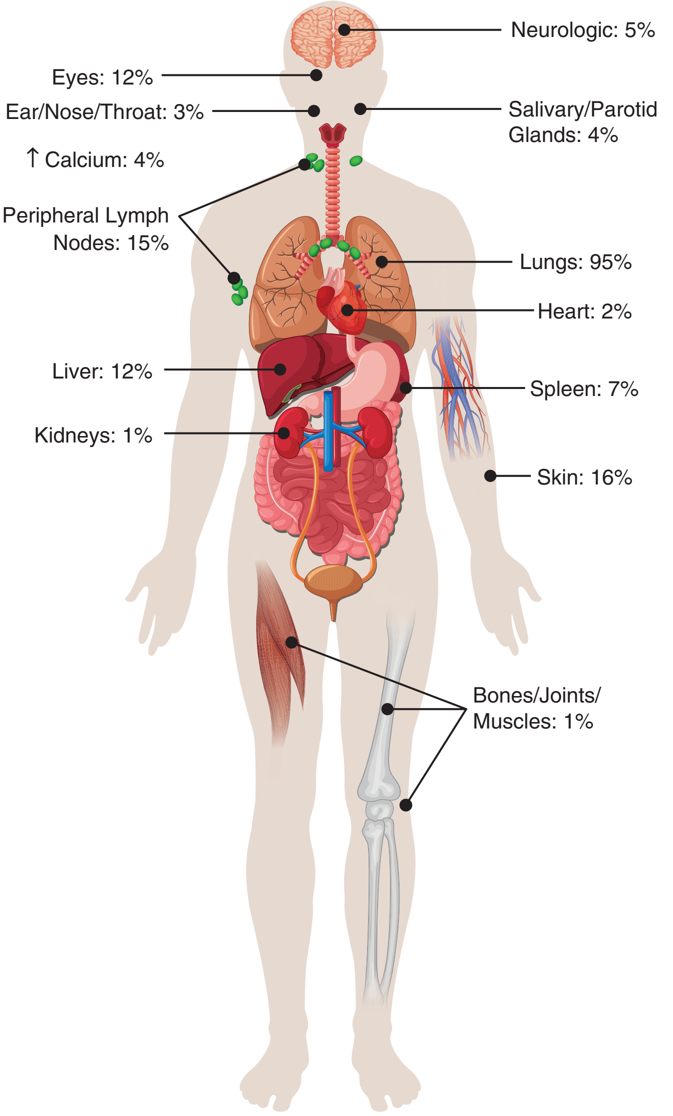

FIGURE 379-3 Frequency of organ involvement in sarcoidosis. Organ involvement upon presentation. *(RP Baughman et al: Am J Respir Crit Care Med 164:1885, 2001.)*

A histopathologic examination of the involved organ tissue reveals granulomas in sarcoidosis; however, diagnosis heavily relies on exclusion of other causes of granulomas. The differential diagnosis can be broad and must be considered in the clinical context **(Table 379-1)**. A careful history assessment for environmental exposures or other inflammatory diseases is imperative to secure confidence in the diagnosis of sarcoidosis. Infections can be difficult to differentiate from sarcoidosis and require special histologic stains and cultures. Necrotizing sarcoidosis granulomatosis, characterized by necrosis, is a rare and less well-established entity that often involves vasculitis. Given the numerous causes of granulomatous inflammation and disease mimics, the diagnosis of sarcoidosis is never made with absolute certainty, and new symptoms or signs that occur during the disease process should be fully evaluated.

DIAGNOSTIC EVALUATION

The initial evaluation of sarcoidosis patients includes a complete history and physical exam, as well as assessment of organ involvement and severity of disease. Once diagnosis is suspected, a full review of symptoms is warranted to elucidate any potential pulmonary and extrapulmonary involvement, as about half of patients will have extrapulmonary involvement. Almost every organ system in the body can be affected by sarcoidosis **(Fig. 379-3)**. A baseline set of testing (at minimum) should be done in all patients on initial evaluation, including pulmonary function tests, chest imaging, a baseline ophthalmologic investigation to evaluate for clinically silent uveitis, and an electrocardiogram (ECG) **(Table 379-2)**. Serum tests for abnormal calcium metabolism and any significant hepatic, renal, or hematologic involvement should also be performed. Hematologic abnormalities are frequent and can be attributed to redistribution of T cells to sites of disease, splenic sequestration, granulomatous bone marrow involvement, or immune mediation. Routine asymptomatic follow-up testing is more controversial regarding its utility, but repeat testing should be done to evaluate the onset of any new signs or symptoms as the disease evolves. Abnormal initial testing may require further workup.

ORGAN INVOLVEMENT

LUNGS

The lungs are involved in >90% of patients with sarcoidosis. Pulmonary function tests are important to measure initial lung impairment and to assess improvement or deterioration over time. Many patients have normal lung function testing or a normal lung exam despite the presence of granulomatous inflammation and abnormal imaging findings. A restrictive pattern is often seen in those with more

TABLE 379-2 Baseline Testing upon Initial Diagnosis of Sarcoidosis

Full history, review of systems, physical exam
Pulmonary function testing
Chest imaging
Eye exam
Serum creatinine, alkaline phosphatase, calcium level, complete blood counts
25- and 1,25-hydroxyvitamin D levels if assessing vitamin D metabolism
Electrocardiogram

Note: Further testing is dependent upon signs or symptoms indicating potential organ involvement.

advanced disease, and airflow obstruction can occur in up to one-third of patients. Sarcoidosis tends to be upper-lobe predominant on lung imaging. Chest x-rays are classically used to "stage," or describe the pattern of presentation in the lungs, known as Scadding stages (Fig. 379-4). Although chest CT is not indicated in every patient, it is commonly performed to determine extent of lung disease and to guide biopsies. Sarcoidosis granulomas have a lymphangitic distribution in the lung along the pleural and subpleural areas and interlobular septa and around the bronchovascular bundles. Classic findings of sarcoidosis on chest CT scans are widespread small nodules with a bronchovascular and subpleural distribution (Fig. 379-5*A, B*), thickened interlobular septa, mediastinal and hilar lymphadenopathy, and conglomerate masses with architectural distortion (Fig. 379-5*C*). In advanced cases of pulmonary involvement, CT findings include honeycombing, cyst or cavity formation, and bronchiectasis. Nodular sarcoidosis or necrotizing sarcoid granulomatosis can present with nodular masses. Pleural effusions tend to be indirectly related (e.g., infections, heart failure), but rarely, sarcoidosis can cause lymphocytic, exudative pleural effusions. Upper respiratory tract involvement can also occur rarely within the larynx, pharynx, and sinuses and is associated with more chronic severe disease.

Progressive pulmonary fibrosis with irreversible scarring is a significant cause of complications and death in patients with sarcoidosis. Fibrosis can be complicated by cavitary lung disease, chronic pulmonary aspergillosis or aspergillomas, and pulmonary hypertension. Airflow obstruction is common in these patients due to central airway scarring and distortion of the lung structure. Fibrosis can occur even in the setting of anti-inflammatory therapy.

■ HEART

Granulomas can infiltrate any part of the heart, causing rhythm abnormalities, heart failure, pericardial effusion, and even sudden death. In the broader general population, cardiac sarcoidosis is the cause of a significant proportion of cases of new-onset atrioventricular block, cardiomyopathy, and ventricular dysrhythmias, particularly in younger age groups. Clinical evidence of cardiac involvement in patients with

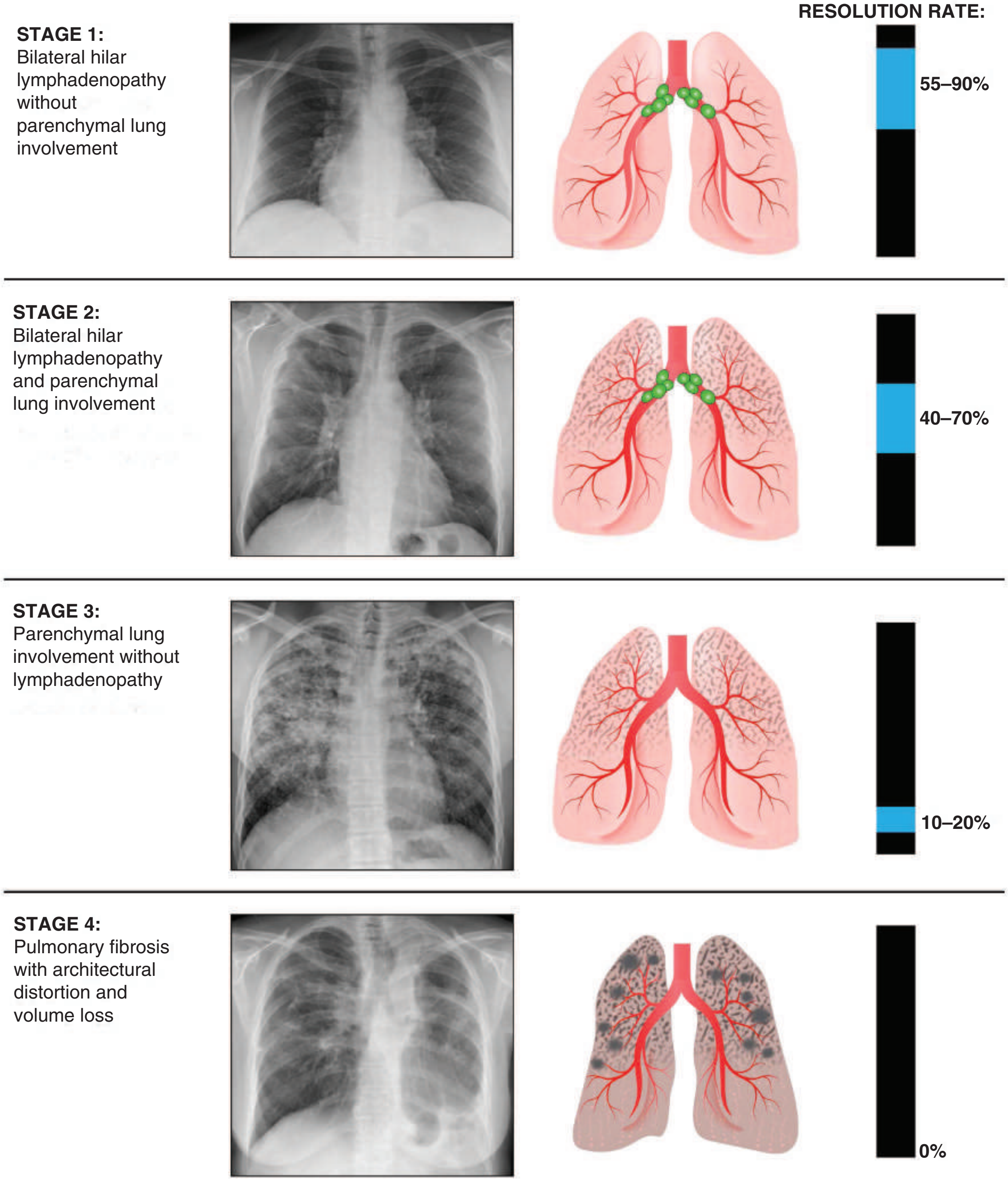
RESOLUTION RATE:
STAGE 1:
Bilateral hilar lymphadenopathy without parenchymal lung involvement
55–90%
STAGE 2:
Bilateral hilar lymphadenopathy and parenchymal lung involvement
40–70%
STAGE 3:
Parenchymal lung involvement without lymphadenopathy
10–20%
STAGE 4:
Pulmonary fibrosis with architectural distortion and volume loss
0%

FIGURE 379-4 Scadding stages in sarcoidosis.

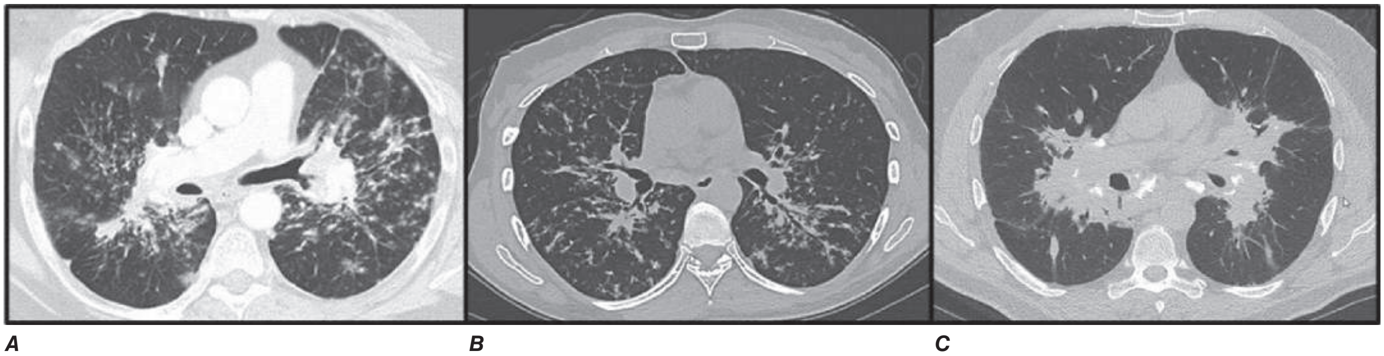

FIGURE 379-5 Imaging findings in sarcoidosis. (***A and B***) Chest computed tomography images of sarcoidosis show widespread small nodules in a perilymphatic and subpleural distribution and prominent bronchovascular bundle thickening. (***C***) Large peribronchovascular mass-like consolidations extending along the hila along with partially calcified mediastinal and hilar lymphadenopathy in a patient with sarcoidosis.

known sarcoidosis is found in <5% of U.S. sarcoidosis patients and 23% of Japanese patients, although autopsy studies suggest a higher rate of subclinical involvement. Severity of pulmonary involvement does not appear to predict the presence of cardiac sarcoidosis, and therefore every patient should be queried for potential cardiac symptoms, as cardiac sarcoidosis can have life-threatening complications. Isolated cardiac sarcoidosis can also rarely occur.

Sudden death is the most common cause of death among cardiac sarcoidosis patients, and it can occur as the initial presentation of the disease. Ventricular tachycardia is the most frequent tachyarrhythmia, but atrial tachycardias are also found in relation to ventricular dysfunction, dilated atria, or severe pulmonary disease. Heart failure is most often due to involvement of the myocardium but can also be caused by infiltration of valves or papillary muscles. Predictors of sudden death include severity of heart failure, left ventricular end-diastolic diameter, and sustained ventricular tachycardia.

Diagnosis of cardiac sarcoidosis can be difficult, since gold-standard biopsies lack sensitivity due to the patchy presence of granulomas in cardiac tissue, predominant location at base of the heart, and the lack of right ventricular involvement. ECG should be obtained on initial evaluation of every patient with sarcoidosis to reveal the presence of any conduction block, ST-T wave abnormalities, Q waves, frequent premature ventricular complexes, or resting tachycardia. Symptoms or unexpected abnormalities on ECG should prompt further cardiac evaluation, which can include a Holter monitor, signal-averaged ECG, or further cardiac imaging.

Several imaging tests are used for evaluating and diagnosing cardiac sarcoidosis. Historically, thallium-201 scintigraphy or technetium-sestamibi single-photon emission CT nuclear scans have been used to identify segmental defects that correspond either to a granulomatous disease or a fibrous scar. In contrast to coronary artery disease, patients with cardiac sarcoidosis have perfusion defects that reverse with exercise or dipyridamole, termed *reverse redistribution*. More current imaging studies now support the use of enhanced cardiac magnetic resonance imaging (MRI) as a preferred diagnostic tool due to its superior resolution and ability to detect early myocardial involvement with high sensitivity and specificity. Cardiac PET/CT can also be used in some cases to detect the presence of active inflammation in cases where cardiac MRI is unavailable, unable to be done (noncompatible devices or claustrophobia), or inconclusive. The use of these imaging techniques has largely obviated the need for myocardial biopsy in the diagnosis of cardiac involvement.

In patients with rhythm disturbances, prompt referral to an electrophysiologist is warranted for consideration of potential mapping or implantation of cardiac devices. Echocardiography can help identify ventricular dysfunction, valve dysfunction, abnormal septal wall thickness, or wall motion abnormalities that may also be contributing to the clinical picture.

Traditional treatments for cardiac arrhythmias and heart failure are indicated depending on the presentation, as well as immunosuppressive medications. The optimal amount of time dedicated to the course of immunosuppressive therapy is unclear, although the presence of cardiac sarcoidosis generally portends a more chronic and protracted course. Serial imaging and response to therapy can help guide potential therapy duration. Heart transplantation can be considered for refractory severe cardiac sarcoidosis, resulting in better short- and intermediate-term survival rates as compared to other transplant recipients.

PULMONARY ARTERY HYPERTENSION

Pulmonary hypertension (PH) in sarcoidosis, classified as a World Health Organization group 5 cause of PH, can complicate the clinical course and contribute significantly to morbidity and mortality. Pulmonary hypertension in sarcoidosis may be due to granulomatous vasculitis, pulmonary arterial occlusion by lymphadenopathy, thrombosis, pulmonary venous occlusion, destruction of the vascular bed, left heart dysfunction due to myocardial involvement, portopulmonary hypertension from associated liver cirrhosis, or hypoxic vasoconstriction related to parenchymal abnormalities. In some cases, precapillary PH can occur in the absence of other cardiopulmonary disease, resembling idiopathic PH. Evaluation for PH should be sought in patients who have symptoms that are disproportionate to the amount of parenchymal disease, have extensive parenchymal abnormalities, desaturate on 6-min walk test, or have a low diffusing capacity.

Echocardiography is a useful noninvasive technique screening test to evaluate pulmonary pressures, systolic or diastolic function, and valvular disease; however, right heart catheterization is the gold standard for diagnosis. Treating granulomatous disease with corticosteroids may result in the improvement of pulmonary pressures in some, but not all, cases, and pulmonary vasodilator therapy can be considered on a case-by-case basis.

NERVOUS SYSTEM INVOLVEMENT

Sarcoidosis affecting the nervous system is highly associated with systemic disease but also can occur in the absence of significant pulmonary or systemic disease. Cranial nerve involvement, particularly with the facial and optic nerves, is the most common manifestation and can be a result of granulomatous basal meningitis or direct involvement of the nerve. Facial nerve palsy is associated with good prognosis compared to other neurologic findings. Neurosarcoidosis has a predilection for the base of the brain. Inflammation of the hypothalamus and pituitary gland can cause neuroendocrine abnormalities that can persist even after immunosuppressive treatment. Other findings can include space-occupying masses, acute or chronic meningitis, peripheral neuropathy, and neuromuscular or spinal cord involvement. Rare presentations include seizures or neuropsychiatric symptoms.

Gadolinium-enhanced MRI is the preferred test for evaluating brain parenchyma, meninges, dura, and the spinal cord. MRI findings of leptomeningeal or parenchymal enhancement or multiple white matter lesions in a periventricular distribution are the more common abnormalities and are sensitive diagnostic findings. However, the appearance of sarcoidosis lesions is often nonspecific, mimicking malignancies or infections. Chest imaging can be helpful in diagnosis since many patients will have a pulmonary abnormality that can be biopsied. Cerebrospinal fluid (CSF) analysis most often reveals mononuclear cell pleocytosis and/or elevated proteins. Elevated CSF angiotensin-converting enzyme, high soluble IL-2 receptor, increased lysozyme, decreased glucose, and oligoclonal bands can also be seen but are not diagnostic. CSF analysis is important to help exclude infection. Histologic confirmation of disease in the brain or spinal cord is occasionally necessary when significant diagnostic uncertainty exists or a patient is not responding to therapy. Otherwise, a diagnosis can be established with an acceptable degree of certainty by biopsy of a nonneurologic site, such as the lung, together with consistent clinical features and imaging and exclusion of other causes.

Peripheral nerves can also be affected in sarcoidosis. Pain is a common symptom among sarcoidosis patients and may be related to neuropathic or fibromyalgia syndromes. Small-fiber neuropathy has increasingly been noted to be a cause of pain and autonomic dysfunction. Large-fiber neuropathies can be evaluated by performing nerve biopsy or nerve conduction studies. Unfortunately, pain syndromes, regardless of their underlying cause, often do not respond well to immunosuppressive medications.

CUTANEOUS

Sarcoidosis causes many forms of skin lesions, with maculopapular, nodular eruptions and plaques being the most common types involving granulomatous inflammation of the tissue itself. Other skin lesions include infiltration of old scars or tattoos, hypo- and hyperpigmented areas, scales, annular lesions, ulcerations, and subcutaneous nodules or masses. Most of the time, lesions are not painful or pruritic. Nonspecific, nongranulomatous lesions of erythema nodosum are a well-established manifestation of sarcoidosis, commonly found in Europeans and women. Lesions are raised, tender, erythematous bumps or nodules found on the anterior legs and often a prominent feature of Löfgren syndrome. Lupus pernio is a distinctive chronic lesion and consists of indurated plaques and violaceous discoloration of the nose, cheeks, lips, and ears, and frequently involves the nasal mucosa **(Fig. 379-6)**. This lesion appears most commonly among members of the African-American female population. Lupus pernio is often associated with more severe disease including bone cysts, pulmonary fibrosis, and a prolonged clinical course. Overall, skin lesions can be treated with topical or injected corticosteroids, although disfiguring or symptomatic lesions may need systemic therapy.

EYE INVOLVEMENT

Ocular sarcoidosis can affect any part of the eye or orbit, with clinical manifestations ranging from asymptomatic to permanent vision loss. Given the potential for permanent vision loss, ophthalmologic examination is recommended on initial exam in sarcoidosis patients and with any change in visual symptoms. Uveitis (anterior, posterior, or panuveitis) is the most common manifestation and is often bilateral. Chronic uveitis may lead to synechiae formation, glaucoma, cataracts, cystoid macular edema, and blindness. Optic neuropathy can lead to sudden permanent vision loss and should be treated as an emergency with immediate initiation of immunosuppressive therapy. Other eye lesions include conjunctival or eyelid granulomas, lacrimal gland enlargement, keratoconjunctivitis sicca, dacryocystitis, retinal vasculitis, and periorbital swelling. If conjunctival nodules are present, biopsy can be a relatively noninvasive way to obtain diagnostic tissue.

LIVER

Hepatic granulomas are found very frequently in patients with pulmonary sarcoidosis (up to 80% in some autopsy studies) but are much less often clinically significant (up to 15%). Conversely, hepatic sarcoidosis can occur without involving the lungs. Liver function tests are abnormal in up to one-third of sarcoidosis cases, with an elevated alkaline phosphatase as the most common lab abnormality. Elevated transaminases are also frequent, while elevated bilirubin levels are more indicative of significant progressive sarcoidosis liver disease. Hepatomegaly is commonly seen. Patients who do present with symptoms experience abdominal pain, pruritus, fevers, weight loss, and jaundice. Hepatic sarcoidosis can infrequently progress to chronic cholestasis, hepatocellular disease, portal hypertension, Budd-Chiari syndrome, or cirrhosis. The pattern of presentation depends on the extent and location of granulomatous inflammation and fibrosis in the liver. Imaging of the liver often shows hepatomegaly and/or small, innumerable nodules of low attenuation. Extensive evaluation of alternative causes should be performed when cases of isolated hepatic granulomas are involved, because liver granulomas are associated with a large differential diagnosis. Immunosuppressives may be effective when a patient experiences symptomatic liver dysfunction, but not in every case. Liver transplantation remains a viable option when hepatic failure occurs.

MUSCULOSKELETAL

Acute arthritis (such as that seen in Löfgren syndrome) is common in sarcoidosis, especially early in the course of disease. Pain is a potential indication to treat. Chronic arthritis or synovitis is rarer than the acute presentations. The most common joints affected by sarcoidosis are the ankles, knees, elbows, wrists, and small joints of the hands and feet. Bone lesions, characteristically identified as lacy, reticular osteolytic lesions on radiographs, can be frequently asymptomatic but occasionally cause pain. Bone marrow involvement can cause hematologic abnormalities such as anemia and lymphopenia.

Cases of symptomatic muscle involvement present with pain, tenderness, or weakness and have physical findings that are consistent with myositis, chronic myopathy, or muscle nodules. Myopathy is more commonly found in women and can be the sole presentation of disease. MRI or nuclear bone scans are useful to identify active inflammation in patients with musculoskeletal symptoms or to differentiate nonsarcoidosis causes. Electromyogram studies and biopsy may also be helpful in the diagnostic workup.

HYPERCALCEMIA

Hypercalcemia occurs in only about 10–20% of patients with sarcoidosis, but hypercalciuria is much more frequent and is seen in up to 50%

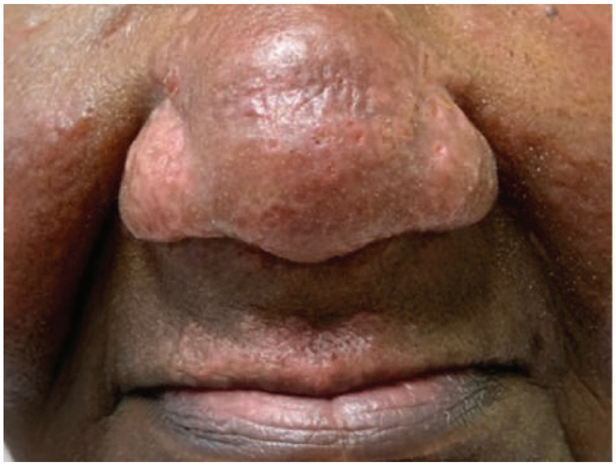

FIGURE 379-6 Cutaneous sarcoidosis. Extensive involvement of the nose and philtrum extending onto the cheeks is seen. The lesions include confluent papules and plaques, some of which are vaguely annular on the philtrum. *(Courtesy of Dr. Misha Rosenbach, MD.)*

of patients. High calcium levels are due to dysregulated production of the active form of vitamin D (1,25-hydroxyvitamin D_3 or calcitriol) by activated macrophages and granulomas in the setting of abundant cytokines such as IFN-γ. Normally, conversion of vitamin D to its active form is a highly regulated process that occurs in the kidney via the action of 1-alpha-hydroxylase (CYP27B1) controlled by parathyroid gland. The conversion of vitamin D to calcitriol by macrophages occurs unregulated, causing increased intestinal absorption and bone resorption and thereby raising serum calcium despite low parathyroid hormone levels. Hypercalcemia, left untreated, leads to nephrocalcinosis and subsequent renal dysfunction. Patients with evidence of hypercalcemia should monitor vitamin D intake, avoid excessive sunlight, and maintain hydration. In severe acute symptomatic cases, hypercalcemia is a cause of hospital admission requiring more urgent corticosteroid therapy.

■ KIDNEY

Renal dysfunction is mostly related to hypercalcemia and nephrocalcinosis associated with sarcoidosis, rather than direct granulomatous infiltration. However, sarcoidosis can cause interstitial nephritis (with and without granulomas), which is generally responsive to treatment. Urinalysis usually reveals sterile pyuria, bland sediment, or mild proteinuria.

■ LYMPHOID SYSTEM

Peripheral lymphadenopathy can be the presenting sign of sarcoidosis in a minority of cases. The extrathoracic glands most frequently involved are the cervical, axillary, epitrochlear, and inguinal glands. Enlarged glands are discrete, movable, and nontender without ulceration or drainage. Splenomegaly in sarcoidosis is common but is usually clinically silent. Hypodense nodules can be seen on imaging. Infrequently, splenic involvement may cause pressure, pain, constitutional symptoms, or hematologic abnormalities that require splenectomy.

■ PAROTID GLANDS

Parotid or salivary involvement occurs in 4% of sarcoidosis patients, often associated with Heerfordt syndrome (parotitis, uveitis, fever, facial palsy). Patients with parotid gland involvement can be asymptomatic or present with painful, swollen glands or xerostomia. Parotid enlargement is self-limiting in many cases and can present like Sjögren's syndrome.

■ GASTROINTESTINAL TRACT

Gastrointestinal tract involvement is a rare manifestation of sarcoidosis and most commonly involves the stomach than any other parts. Patients may present with dysphagia, bleeding, nausea or vomiting, pain, weight loss, or obstructive symptoms.

■ REPRODUCTIVE ORGANS

Granulomas may occur in the reproductive system, most commonly appearing within the female uterus or the male epididymis. Given the rarity of involvement, any new nodules or masses in these organs require careful attention to exclude malignancy.

CLINICAL COURSE AND PROGNOSIS

In just over half of patients, the disease resolves spontaneously; however, a significant proportion (30–40%) of patients will have chronic disease that threatens vital organs or causes debilitating symptoms, necessitating treatment. Chronic symptoms can be progressive or waxing and waning. Predicting clinical course on presentation in any one individual is difficult, given lack of known prognostic biomarkers. Research studies are often confounded by the variable clinical course and the lack of correlation of serum and radiographic measures to the intensity of granulomatous response. Most cases of sarcoidosis will show improvement in chest radiography or spirometry after 2 years of follow-up, including treated and untreated. Serious extrapulmonary involvement (e.g., cardiac, central nervous system, hepatic) occurs in 4–7% of sarcoidosis patients at presentation and can develop as the disease evolves. Relapses are common and are associated with shorter treatment duration. Acute exacerbations of pulmonary sarcoidosis are thought to occur in over one-third of patients, although they are not well-defined.

Death related to sarcoidosis itself does occur in 1–5% patients because of progressive respiratory insufficiency, complications of pulmonary scarring (e.g., hemoptysis, infection, aspergilloma), central nervous system involvement, liver failure, or myocardial involvement (heart failure, cardiac block, sudden cardiac death, or arrhythmias). Differing mortality rates reflect variances in disease severity, referral bias, and diverse genetic and epidemiologic factors of the populations studied. For instance, most deaths occurring in the United States and Northern Europe are due to pulmonary complications, whereas a vast majority of deaths due to sarcoidosis in Japanese patients are from cardiac involvement. Complications of immunosuppressive therapies can also contribute indirectly to mortality. Transplantation of the lungs, heart, liver, or kidneys can be lifesaving in severe progressive cases.

■ CLINICAL FACTORS OF PROGNOSTIC SIGNIFICANCE

Overall, patients with a more robust initial immune response have a better prognosis than those with more subacute or chronic presentation. Patients that present with Löfgren or Heerfordt syndromes tend to have an excellent prognosis with spontaneous remission in up to 80%. Löfgren syndrome occurs in 20–30% of Caucasians with sarcoidosis, particularly Northern Europeans, and <5% of Asians and African Americans. Erythema nodosum and fever usually remit spontaneously within 6 weeks, while resolution of lymphadenopathy may be delayed for over a year. Immunosuppressive therapy is rarely necessary. On the other hand, adverse clinical prognostic factors include lupus pernio, chronic uveitis, older age, chronic hypercalcemia, nephrocalcinosis, African origins, pulmonary hypertension, fibrotic pulmonary sarcoidosis, sinus involvement, cystic bone lesions, neurosarcoidosis, and cardiac involvement. Parenchymal infiltrates associated with Scadding radiographic stages do not predict activity, progression of lung disease, or extrapulmonary involvement but do reveal overall resolution rates (Fig. 379-4).

■ PREDICTIVE VALUE OF SERUM, BRONCHOALVEOLAR LAVAGE, AND GENETICS

There are no current clinically used biomarkers that can predict clinical course in any one case. Serum angiotensin-converting enzyme (ACE) levels, derived from macrophages, are not largely useful as a diagnostic or prognostic tool due to low sensitivity. However, ACE levels have been employed to monitor systemic disease activity, as they correlate roughly with the granulomatous burden. Similarly, the T-cell–derived soluble IL-2 receptor lacks sensitivity for diagnosis but correlates with disease activity and may predict relapse rate. In BAL fluid, the high CD4:CD8 ratio, although supporting a diagnosis, does not correlate with disease progression or need for treatment, but higher values seem to implicate a better overall prognosis. Increased percentages of BAL neutrophils (or eosinophils) appear to be present in a higher proportion of individuals with progression. In genetics, certain human leukocyte antigen (HLA) haplotypes have been associated with acute and chronic phenotypes. Similarly, variants of butyrophilin-like 2 (*BTNL2*), tumor necrosis factor alpha (*TNF*-α), and annexin A11 (*ANXA11*) genes have been associated with worse prognosis. Unfortunately, these and multiple other serum and BAL markers lack generalizability and utility in the clinical realm for any one individual.

FOLLOW-UP AND MONITORING

To assess prognosis and determine the need for therapy, longitudinal surveillance of sarcoidosis should be most intensive during the 2-year period following presentation with decreasing frequency as time goes on. When therapy has started, follow-up of response and side effects should be done in 1–3 months and, once stable on therapy, every 3–6 months. Once in full remission, patients should be followed for a few years to assess for relapses; however, lengthier observation may be necessary based on severity of disease or residual organ dysfunction. Follow-up testing should include symptoms, a complete physical exam, imaging of affected organs, periodic pulmonary function testing, and markers of specific organ dysfunction. Choice of imaging depends on clinical scenario and can include chest x-rays, CT scans, and in some rarer cases, PET/CT or MRI. Any development of new symptoms over time should prompt full evaluation of that organ system.

TREATMENT

Sarcoidosis

Treatment decisions in sarcoidosis are complex, and medication options tend to be guided by expert opinion, uncontrolled trials, and only a few larger randomized controlled trials. There is no cure for sarcoidosis; the goal of treatment is to suppress granulomatous inflammation, prevent irreversible organ fibrosis, and alleviate debilitating symptoms as the disease evolves. With the high side effect profile of corticosteroids and other immunosuppressive therapy, it is first important to identify those cases in which therapy is indicated. For example, treatment is not indicated for asymptomatic stage I disease, which has an inherent high rate of spontaneous resolution. Mild symptoms may be closely monitored, and topical immunosuppressive therapies are preferred for skin disease if possible. Oral immunosuppressive therapy should be reserved for those who have significant symptoms, disfigurement, or dangerous impending organ damage. In pulmonary disease, this may include worsening cough and shortness of breath, moderate to severe parenchymal lung involvement, or a progressive decline in pulmonary function. In each case of sarcoidosis, the risks and toxicity of immunosuppression must be considered in comparison to the potential benefit of treatment.

Corticosteroids remain the mainstay of first-line treatment, although controversy exists regarding whether they alter the natural history. Oral steroids have been shown to improve chest x-ray imaging scores, symptoms, and spirometry in the short term. There are limited data beyond 2 years to indicate whether oral steroids have any modifying effect on long-term disease progression or mortality. In cases of intolerance or toxicity to corticosteroids, inability to taper corticosteroids, or refractory disease, alternative steroid-sparing immunosuppressives can be considered **(Table 379-3)**. In choosing a therapy, clinicians should take into consideration the presence of comorbid conditions such as osteoporosis, diabetes, cataracts, or glaucoma that may be aggravated by corticosteroids. Organ transplantation is occasionally necessary in cases of progressive organ damage including lungs, liver, heart, and kidney. Sarcoidosis can recur in transplanted organs but is rarely a cause of organ failure after transplant.

CORTICOSTEROIDS

When it is determined that treatment is indicated, corticosteroids are generally used as first-line therapy given efficacy and ease of use. Generally, prednisone or an equivalent corticosteroid is started at 0.5 mg/kg per day or ~20–40 mg/d for 4 weeks, and then reduced to a maintenance dose that will control symptoms and disease progression, ideally below 10 mg/d. Higher starting doses are not largely effective in most cases of sarcoidosis and are associated with increased morbidity. Lower induction doses can be considered based on severity of presentation, side effect tolerance, and rate of disease progression. More aggressive therapeutic approaches with higher doses or in association with other immunosuppression can be considered in life- or organ-threatening situations such as seen in central nervous system involvement, ocular involvement leading to vision loss, or extensive myocardial inflammation. In a subgroup of cases with very mild pulmonary symptoms such as cough, inhaled corticosteroids may be considered for symptom control but have not been shown to be of benefit in the larger sarcoidosis population.

STEROID-SPARING THERAPIES

As the landscape of drug development broadens, treatment options have expanded beyond corticosteroids to include several anti-inflammatories drawn from treatment of other rheumatologic conditions (Table 379-3). The choice of steroid-sparing agent often depends on several clinical factors (e.g., alcohol use, desire for pregnancy, known pharmacogenetic profiles, comorbidities such as liver or renal dysfunction), but methotrexate is the most recommended second-line therapy due to its efficacy and favorable side effect profile. Azathioprine appears to be equally effective, although

TABLE 379-3 Treatments for Sarcoidosis

	DRUG NAME	DOSE RANGE	CONSIDERATIONS
Initial therapy	Corticosteroids (prednisone, prednisolone)	20–40 mg/d initial, tapered to 7.5–15 mg/d	Consider taper early based on clinical improvement Monitor bone health Consider implications of weight gain Assess risk of diabetes Monitor eye exams (glaucoma and cataracts)
Common second-line therapies	Methotrexate	7.5–20 mg/wk orally or subcutaneously	Concurrent administration with folic acid Monitor liver and renal function and blood counts Contraindicated with significant alcohol use
	Leflunomide	10–20 mg/d	Monitor liver and renal function and blood counts In cases of toxicity, consider cholestyramine to accelerate clearance
	Azathioprine	50–200 mg/d or 1–2 mg/kg/d	Monitor liver and renal function and blood counts Consider thiopurine methyltransferase activity level
	Mycophenolate	1000–3000 mg/d	Monitor blood counts Enteric-coated option available (note: different dose range)
	Hydroxychloroquine	200–400 mg/d	Periodic eye exams for retinopathy Assess QT (monitor drug interactions) Effective for skin or hypercalcemia, but not for pulmonary
Refractory disease or inability to taper corticosteroids	Infliximab	3–5 mg/kg intravenously at weeks 0 and 2, and then every 4–8 weeks	Tuberculosis screening prior to use Avoid use in heart failure Allergic reactions possible with infusion Longer term association with demyelination syndrome and malignancy Can induce sarcoid-like reactions
	Adalimumab	40 mg subcutaneous every 1–2 weeks (exact dose unknown)	Similar precautions and adverse events as infliximab

potentially has higher rates of infectious complications. Other medications that can be considered second-line treatment include leflunomide or mycophenolate mofetil. Hydroxychloroquine can also be an effective immunomodulator, particularly for skin disease or hypercalcemia, but is not recommended for lung disease.

If all second-line therapies have been exhausted due to intolerance, inefficacy, or inability to taper corticosteroids, tumor necrosis factor (TNF) inhibitors can be considered. TNF-α is produced by the activated macrophages of the granuloma and contributes to propagation of inflammation in sarcoidosis. Infliximab and adalimumab are monoclonal antibodies that target TNF-α. Infliximab has garnered the most supporting data, with randomized controlled trials and larger case series showing positive effects on pulmonary function, imaging, and inflammatory cytokine levels, particularly in those with more severe disease. Adalimumab has shown efficacy in smaller series, particularly in ocular sarcoidosis. Etanercept, on the other hand, a TNF receptor antagonist, has been shown to be ineffective for sarcoidosis and should not be used for treatment. Other therapies targeting different aspects of the immune response are also considered at times in refractory cases including anti-B-cell therapy and antifibrotic medications.

IMMUNOSUPPRESSIVE MONITORING

While on immunosuppressive therapy, close attention to prevention of toxicity is important in the care plan, and monitoring should follow published rheumatologic guideline recommendations. Bisphosphonates should be considered to minimize steroid-induced osteoporosis, and regular eye exams should be obtained with corticosteroid use and hydroxychloroquine. When assessing a medication choice, it is also important to consider pregnancy risk, as many second-line agents have adverse reproductive effects.

DURATION OF TREATMENT

The exact duration of treatment may differ between individuals based on severity of disease but is ~1 year for most patients. Shorter courses of therapy have been associated with higher relapse rates, although they can be considered for particularly responsive cases. For pulmonary involvement, the greatest amount of improvement is seen within the first weeks to months of therapy; therefore, tapering of corticosteroids to a lower maintenance dose should be considered with early follow-up to minimize long-term toxicity. Longer courses of treatment may be necessary for life-threatening or refractory disease. Relapses are treated with the last known effective dosing regimen.

APPROACH TO THE PATIENT

Sarcoidosis

Comprehensive treatment of the patient with sarcoidosis includes not only suppression of granulomatous inflammation but also consideration of medication side effects and evaluation of fatigue, psychological health, and pain. Up to 80% of patients with sarcoidosis experience multifactorial fatigue, which can persist even after treatment and remission. Fatigue can be due to medications, sleep disorders, pain, and the granulomatous inflammation itself. It is associated with extrapulmonary involvement. Neuropathic pain or autonomic dysfunction may require symptomatic care. Formalized physical training has been shown to improve exercise capacity, strength, and fatigue in patients with sarcoidosis. Depression is also highly prevalent in chronic sarcoidosis, associated with female sex, dyspnea, and poor access to care. Cognitive impairment contributing to decreased quality of life is well-noted but poorly understood. Patient-centered care of the entire constellation of effects of sarcoidosis can include neuropsychiatric evaluation, sleep therapy, exercise therapy, pain relief, and lifestyle modifications. An individualized plan is necessary for every patient to address the benefits of anti-inflammatory treatments, associated comorbidities, and maintenance of quality of life while living with sarcoidosis.

FURTHER READING

Baughman RP et al: Clinical characteristics of patients in a case control study of sarcoidosis. Am J Respir Crit Care Med 164:1885, 2001.

Baughman RP et al: European Respiratory Society clinical practice guidelines on treatment of sarcoidosis. Eur Respir J 58: 2004079, 2021.

Crouser ED et al: Diagnosis and Detection of Sarcoidosis. An Official American Thoracic Society Clinical Practice Guideline. Am J Respir Crit Care Med 201:e26, 2020.

Drent M et al: Challenges of sarcoidosis and its management. N Engl J Med 385:1018, 2021.

Judson MA: Environmental risk factors for sarcoidosis. Front Immunol 11:1340, 2020.

380

IgG4-Related Disease

John H. Stone

IgG4-related disease (IgG4-RD) is a fibroinflammatory condition characterized by a tendency to form tumefactive lesions. The clinical manifestations of this disease, however, are protean, as IgG4-RD can affect virtually any organ system. The disease has a particular predilection for targeting blood vessels of any size on either the venous or arterial side of the circulation and is therefore regarded as a variable-vessel vasculitis. Commonly affected organs are the pancreas, biliary tree, major salivary glands (submandibular, parotid), periorbital tissues, kidneys, lungs, lymph nodes, and retroperitoneum. In addition, IgG4-RD involvement of the meninges, aorta, prostate, thyroid, pericardium, skin, and other organs is well described. The disease affects the brain parenchyma, the joints, the bone marrow, and the bowel mucosa only rarely.

The pathologic findings are consistent across all affected organs. These findings include a lymphoplasmacytic infiltrate with a high percentage of IgG4-positive plasma cells; a characteristic pattern of fibrosis termed "storiform" (from the Latin *storea*, for "woven mat"); a tendency to target blood vessels, particularly veins, through an obliterative process ("obliterative phlebitis"); and a mild to moderate tissue eosinophilia. Although the pathology is consistent from organ to organ, it is essentially never diagnostic in and of itself. Classification criteria emphasize the importance of careful correlation among clinical, serologic, radiologic, and pathologic findings in deciding whether a patient should be classified as having IgG4-RD. Biopsy is not required in order to establish the diagnosis in classic cases, but most patients undergo a biopsy at some point in the evaluation in order to exclude malignancy.

IgG4-RD encompasses a number of conditions previously regarded as separate, organ-specific entities. A condition once known as "lymphoplasmacytic sclerosing pancreatitis" became the paradigm of IgG4-RD in 2000, when Japanese investigators recognized that these patients had elevated serum concentrations of IgG4. This form of sclerosing pancreatitis is now termed type 1 (IgG4-related) autoimmune pancreatitis (AIP). By 2003, extrapancreatic disease manifestations had been identified in patients with type 1 AIP, and descriptions of IgG4-RD in other organs followed. *Mikulicz's disease*, once considered to be a subset of Sjögren's syndrome that affected the lacrimal, parotid, and submandibular glands, is one of the most common presentations of IgG4-RD.

CLINICAL FEATURES

The major organ lesions are summarized in **Table 380-1**. IgG4-RD usually presents subacutely, and even in the setting of multiorgan disease, most patients do not have fevers or high elevations of C-reactive protein levels. Some patients, however, experience substantial weight loss over periods of months. This is generally because of exocrine

TABLE 380-1 Organ Manifestations of IgG4-Related Disease

ORGAN	MAJOR CLINICAL FEATURES
Orbits and periorbital tissues	Painless eyelid or periocular tissue swelling; orbital pseudotumor; dacryoadenitis; dacryocystitis; orbital myositis; and mass lesions extending into the pterygopalatine fossa and infiltrating along the trigeminal nerve
Ears, nose, and sinuses	Allergic phenomena (nasal polyps, asthma, allergic rhinitis, peripheral eosinophilia); nasal obstruction, rhinorrhea, anosmia, chronic sinusitis; occasional bone-destructive lesions
Salivary glands	Submandibular and/or parotid gland enlargement (isolated bilateral submandibular gland involvement more common); minor salivary glands sometimes involved
Meninges	Headache, radiculopathy, cranial nerve palsies, or other symptoms resulting from spinal cord compression; tendency to form mass lesions; magnetic resonance imaging (MRI) shows marked thickening and enhancement of dura
Hypothalamus and pituitary	Clinical syndromes resulting from involvement of the hypothalamus and pituitary, e.g., anterior pituitary hormone deficiency, central diabetes insipidus, or both; imaging reveals thickened pituitary stalk or mass formation on the stalk, swelling of the pituitary gland, or mass formation within the pituitary
Lymph nodes	Generalized lymphadenopathy or localized disease adjacent to a specific affected organ; the lymph nodes involved are generally 1–2 cm in diameter and nontender
Thyroid gland	Riedel's thyroiditis; fibrosing variant of Hashimoto's thyroiditis
Lungs	Asymptomatic finding on lung imaging; cough, hemoptysis, dyspnea, pleural effusion, or chest discomfort; associated with parenchymal lung involvement, pleural disease, or both; four main clinical lung syndromes: inflammatory pseudotumor, paravertebral mass often extending over several vertebrae, central airway disease, localized or diffuse interstitial pneumonia; pleural lesions have severe, nodular thickening of the visceral or parietal pleura with diffuse sclerosing inflammation, sometimes associated with pleural effusion
Aorta	Asymptomatic finding on radiologic studies; surprise finding at elective aortic surgery; aortic dissection; clinicopathologic syndromes described include lymphoplasmacytic aortitis of thoracic or abdominal aorta, aortic dissection, periaortitis and periarteritis, and inflammatory abdominal aneurysm
Retroperitoneum	Backache, lower abdominal pain, lower extremity edema, hydronephrosis from ureteral involvement, asymptomatic finding on radiologic studies. Classic radiologic appearance is periaortic inflammation extending caudally to involve the iliac vessels.
Kidneys	Tubulointerstitial nephritis; membranous glomerulonephritis in a small minority; asymptomatic tumoral lesions, typically multiple and bilateral, are sometimes detected on radiologic studies; renal involvement strongly associated with hypocomplementemia
Pancreas	Type 1 autoimmune pancreatitis, presenting as mild abdominal pain; weight loss; acute, obstructive jaundice, mimicking adenocarcinoma of the pancreas (including a pancreatic mass); between 20 and 50% of patients present with acute glucose intolerance; imaging shows diffuse (termed "sausage-shaped pancreas") or segmental pancreatic enlargement, with loss of normal lobularity; a mass often raises the suspicion of malignancy
Biliary tree and liver	Obstructive jaundice associated with autoimmunity in most cases; weight loss; steatorrhea; abdominal pain; and new-onset diabetes mellitus; mimicker of primary sclerosing cholangitis and cholangiocarcinoma
Other organs involved	Gallbladder, liver (mass), breast (pseudotumor), prostate (prostatism), pericardium (constrictive pericarditis), mesentery (sclerosing mesenteritis), mediastinum (fibrosing mediastinitis), skin (erythematous or flesh-colored papules), peripheral nerve (perineural inflammation)

pancreatic failure and the resulting inability of the pancreas to produce sufficient quantities of digestive enzymes. Failure of the endocrine pancreas, resulting in diabetes mellitus, is also common. Clinically apparent disease can evolve over months, years, or even decades before the manifestations within a given organ become sufficiently severe to bring the patient to medical attention. Some patients have disease that is marked by the appearance and then resolution or temporary improvement in symptoms within a particular organ. Other patients accumulate new organ involvement as their disease persists in previously affected organs. Initial misdiagnosis, particularly of cancer, is extremely common because of the disease's tendency to cause mass lesions in the organs it affects. IgG4-RD is also often identified incidentally through radiologic findings.

Multiorgan disease may be evident at diagnosis but can also evolve over months to years. Some patients, however, have disease that appears to be confined to a single organ at the time of diagnosis. Others have subclinical involvement of other organs when the major clinical feature presents. For example, patients with type 1 AIP may have their major disease focus in the pancreas, but thorough investigation through the history and physical examination, blood tests, and cross-sectional imaging may demonstrate disease in multiple other organs.

Two common characteristics of IgG4-RD are allergic disease and the tendency to form tumefactive lesions that mimic malignancies **(Fig. 380-1)**. Many IgG4-RD patients have allergic features such as atopy, eczema, asthma, nasal polyps, sinusitis, and modest peripheral eosinophilia. IgG4-RD also appears to account for a significant proportion of tumorous swellings—pseudotumors—in many organ systems **(Fig. 380-2)**. Some patients undergo major surgeries (e.g., modified Whipple procedures or thyroidectomy) for the purpose of resecting malignancies before the correct diagnosis is identified.

IgG4-RD often causes major morbidity and can lead to organ failure; however, its general pattern is to cause damage in a subacute manner. Destructive bone lesions in the sinuses, head, and middle ear spaces that mimic granulomatosis with polyangiitis occur occasionally in IgG4-RD, but less aggressive lesions are the rule in most organs. In regions such as the retroperitoneum, substantial fibrosis often occurs before the diagnosis is established, leading to ureteral entrapment, hydronephrosis, postobstructive uropathy, and renal atrophy. Undiagnosed or undertreated IgG4-related sclerosing cholangitis can lead to hepatic failure within months. Similarly, IgG4-related aortitis can cause aneurysms and dissections. Substantial renal dysfunction and even renal failure can ensue from IgG4-related tubulointerstitial nephritis, and renal atrophy is a frequent sequel to this disease complication even following apparently effective immunosuppressive therapy. IgG4-related membranous glomerulonephropathy, a less common renal manifestation than tubulointerstitial nephritis, must be distinguished from idiopathic membranous glomerulonephropathy. As a variable-vessel vasculitis, IgG4-RD can target small-, medium-, and large-sized arteries and veins. Patients with longstanding IgG4-RD marked by substantial elevations of serum IgG4 concentrations and other biomarkers are at risk for developing coronary arteritis, marked by wall thickening, periarterial soft tissue encasement, stenosis, calcification, and aneurysms or ectasia.

SEROLOGIC FINDINGS

The majority of patients with IgG4-RD have elevated serum IgG4 concentrations; however, the range of elevation varies widely. Serum concentrations of IgG4 as high as 30 or 40 times the upper limit of normal sometimes occur, usually in patients with disease that affects multiple organ systems simultaneously. Approximately 30% of patients have normal serum IgG4 concentrations despite classic histopathologic and immunohistochemical findings. Such patients tend to have disease that affects fewer organs. Patients with IgG4-related retroperitoneal fibrosis often have normal serum IgG4 concentrations, perhaps because the

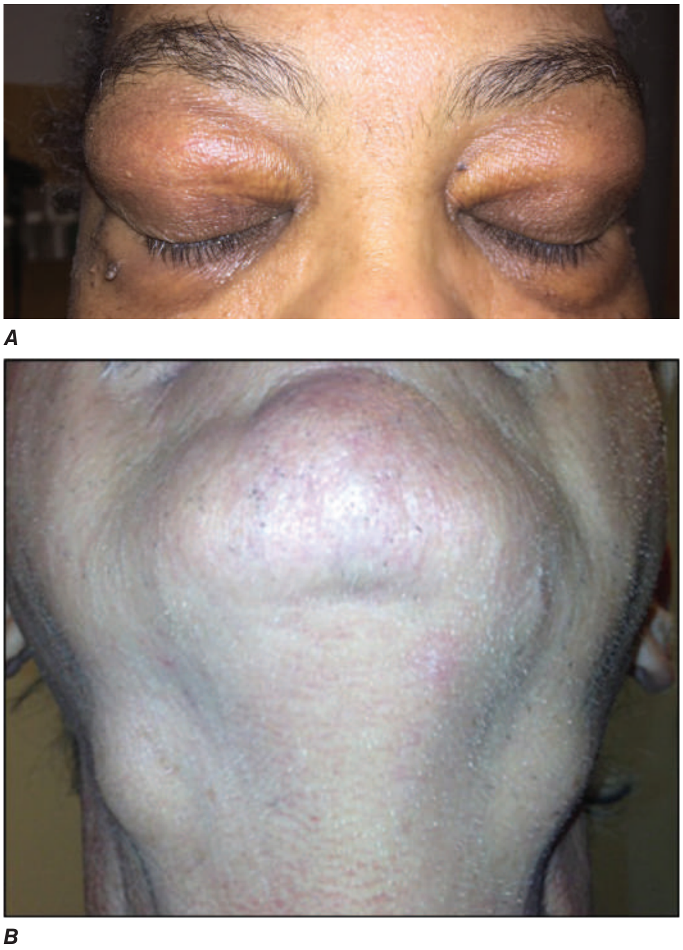

FIGURE 380-1 A major clinical feature of IgG4-related disease is its tendency to form tumefactive lesions. Shown here are mass lesions of the lacrimal glands, causing supraorbital swelling (***A***) and the submandibular glands (***B***).

process has advanced to a fibrotic stage by the time the diagnosis is considered.

Correlations between serum IgG4 concentrations, disease activity, and the need for treatment are imperfect. Serum IgG4 concentrations typically decline swiftly with the institution of therapy but often do not normalize completely. Patients can achieve clinical remissions yet have persistently elevated serum IgG4 concentrations. Following treatment and a disease response, however, steadily rising serum IgG4 concentrations are useful in identifying patients at risk for clinical flares who should be considered for re-treatment. Clinical relapses occur in some patients despite persistently normal IgG4 concentrations.

IgG4 concentrations in serum are usually measured by nephelometry assays. In the setting of extremely high serum IgG4 concentrations, these assays can generate spuriously low IgG4 values because of the prozone effect. Failure to identify dramatic serum IgG4 elevations can have a substantial impact on patients, because that subset of patients is at greatest risk for multiorgan disease and substantial end-organ injury. The prozone effect should be considered when the results of serologic testing for IgG4 concentrations are normal despite the presence of clinical features that strongly suggest IgG4-RD. This effect can be corrected by dilution of the serum sample in the laboratory.

EPIDEMIOLOGY

The typical patient with IgG4-RD is a middle-aged to elderly man. This epidemiology stands in stark contrast to that of many classic autoimmune conditions, which tend to affect young women. Male patients are approximately 5 years older at the time of diagnosis compared to female patients and have higher degrees of serologic activity, e.g., IgG4 hypergammaglobulinemia or hypocomplementemia. The male-to-female ratio appears to be on the order of 2:1, with even more striking male predominance reported in certain types of internal organ involvement (e.g., the pancreas, kidneys, or retroperitoneum). Among IgG4-RD manifestations that involve organs of the head and neck—the orbits, lacrimal glands, and major salivary glands—the sex ratio may be closer to 1:1.

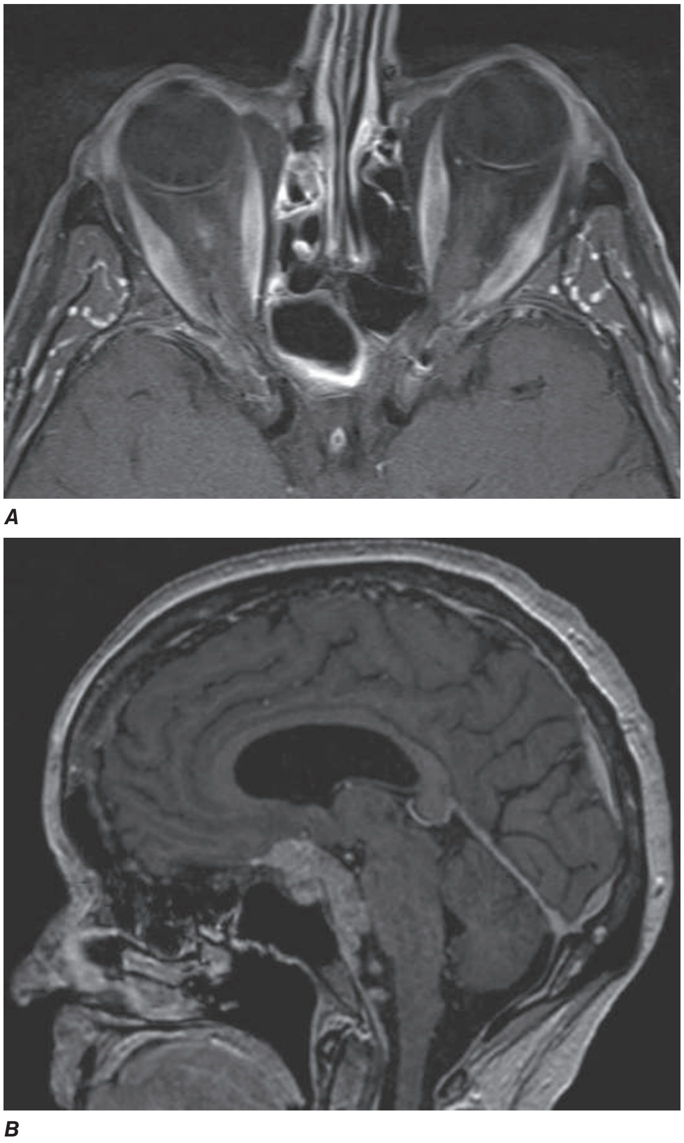

FIGURE 380-2 Thickening of extraocular muscles and meninges. (***A***) Magnetic resonance imaging study of the orbits, showing enhancement and enlargement of extraocular muscles in a patient with IgG4-related orbital disease (orbit axial T1 postcontrast fat-saturated image). (***B***) Magnetic resonance imaging study of the brain, showing thickening of the pachymeninges. There is nodular pachymeningeal enhancement at the base of the brain (blue dot), anterior to the pons. There is also abnormal enhancement and thickening of the tentorium cerebelli, interposed between the superior cerebellum and the occipital lobes (sagittal MPRAGE volumetric T1 postcontrast image).

PATHOLOGY

The key histopathology characteristics of IgG4-RD are a dense lymphoplasmacytic infiltrate (Fig. 380-3) that is organized in a storiform pattern, obliterative phlebitis, and a mild to moderate eosinophilic infiltrate. Lymphoid follicles and germinal centers are frequently observed. The infiltrate tends to aggregate around ductal structures when it affects glands. The inflammatory lesion often aggregates into tumefactive masses that destroy the involved tissue.

Obliterative arteritis is observed in some organs, particularly the lung; however, venous involvement is more common (and is indeed a hallmark of IgG4-RD). Several histopathology features are uncommon in IgG4-RD and, when detected, mitigate against the diagnosis of IgG4-RD. These include intense neutrophilic infiltration, leukocytoclasis, granulomatous inflammation, multinucleated giant cells, and fibrinoid necrosis.

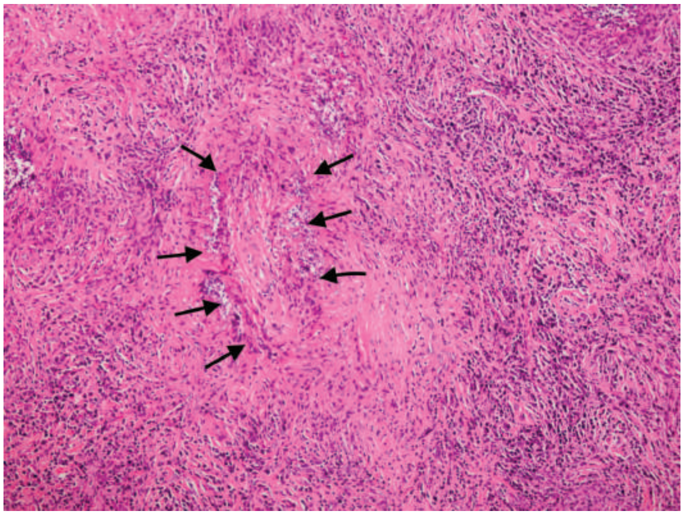

FIGURE 380-3 Hallmark histopathology characteristics of IgG4-related disease (IgG4-RD) are a dense lymphoplasmacytic infiltrate and a mild to moderate eosinophilic infiltrate. The cellular inflammation is often encased in a distinctive type of fibrosis termed "storiform," which often has a basket weave pattern. Abundant fibroblasts and strands of fibrosis accompany the lymphoplasmacytic infiltrate in this figure. Outlined by the *arrowheads* is a vein demonstrating obliterative phlebitis, underscoring the vascular tropism of this disease, which is classified as a variable-vessel vasculitis. This biopsy is from a patient with IgG4-related hypertrophic pachymeningitis. However, the findings are identical to the pathology found in the pancreas, kidneys, lungs, salivary glands, and other organs affected by IgG4-RD.

The inflammatory infiltrate is composed of an admixture of B and T lymphocytes. B cells are typically organized in germinal centers. Plasma cells staining for CD19, CD138, and IgG4 appear to radiate from the germinal centers. In contrast, the T cells, usually CD4+, are distributed more diffusely throughout the lesion and generally represent the most abundant cell type. Fibroblasts, histiocytes, and eosinophils can all be observed in moderate numbers. Some biopsy samples are particularly enriched with eosinophils. In other samples, particularly from long-standing cases, fibrosis predominates.

The histologic appearance of IgG4-RD, although characteristic, should be supplemented by immunohistochemical stains for IgG4 and IgG. IgG4-positive plasma cells predominate within the lesion, but plasma cells containing immunoglobulins from each subclass can be found. The number of IgG4-positive plasma cells can be quantified by either counting the number of cells per high-power field (HPF) or by calculating the ratio of IgG4- to IgG-bearing plasma cells. Tissue fibrosis predominates in the latter phases of organ involvement, and in this relatively acellular phase of inflammation, both the IgG4:total IgG ratio and the pattern of tissue fibrosis are more important than the number of IgG4-positive cells per HPF in establishing the diagnosis. No matter how strongly the histopathology and immunohistochemistry studies favor a diagnosis of IgG4-RD, however, the pathology findings must always be interpreted in the context of clinical, serologic, and radiologic findings. Pathology findings alone are never sufficient for the diagnosis.

PATHOPHYSIOLOGY

Despite the emphasis of IgG4 in the name of this disease, the IgG4 molecule is not believed to play a direct role in the pathophysiology of disease within most organs. The IgG4 molecule can undergo Fab exchange, a phenomenon in which the two halves of the molecule dissociate from each other and reassociate with hemi-molecules of different antigen specificity that have originated from other dissociated IgG4 molecules. Partly as a result of this Fab exchange, IgG4 antibodies do not bind antigen tightly. Moreover, the molecules have low affinities for Fc receptors and C1q and are regarded generally as noninflammatory immunoglobulins. The low affinities for Fc receptors and C1q impair the ability of IgG4 antibodies to induce phagocyte activation, antibody-dependent cellular cytotoxicity, and complement-mediated damage. It is possible, therefore, that the role of IgG4 in this disease is actually as a counterregulatory mechanism rather than part of the primary inflammatory process.

Next-generation sequencing studies of CD4+ effector T cells have demonstrated a unique CD4+ cytotoxic T cell. This cell, also found in abundance at tissue sites of disease, makes interferon gamma, T-cell growth factor-beta, and interleukin-1, all of which may contribute to the storiform fibrosis found in this condition. The cells also elaborate perforin, granzyme A and B, and granulysin, products capable of inducing cytotoxicity. The pronounced oligoclonal expansion of this CD4+ cytotoxic T cell at tissue sites suggests that this cell is a major disease driver.

Oligoclonal expansions of plasmablasts are also present within the blood of patients with IgG4-RD. Continuous antigen presentation by B cells and plasmablasts may support CD4+ cytotoxic T cells, which in turn produce profibrotic cytokines and other molecules, thereby directly mediating tissue injury.

TREATMENT

Vital organ involvement must be treated aggressively, because IgG4-RD can lead to serious organ dysfunction and failure. Aggressive disease can lead quickly to end-stage liver disease, permanent impairment of pancreatic function, renal atrophy, aortic dissection or aneurysms, and destructive lesions in the sinuses and nasopharynx. Not every disease manifestation of IgG4-RD requires immediate treatment, however, because the disease may take an indolent form in some patients. IgG4-related lymphadenopathy, for example, can be asymptomatic for years, without evolution to other disease manifestations. Watchful waiting is prudent in some cases, but monitoring is essential because serious organ involvement may evolve over time, particularly in the setting of persistently rising serum IgG4 concentrations.

Glucocorticoids are the first line of therapy. Treatment regimens, extrapolated from experience with the management of type 1 AIP, generally begin with 40 mg/d of prednisone, with tapering to discontinuation or maintenance doses of 5 mg/d within 2 or 3 months. Although the clinical response to glucocorticoids is usually swift and striking, prolonged steroid-free remissions are uncommon and the risk of steroid-induced morbidity in this middle-aged to elderly patient population is high, particularly in those with baseline comorbidities and pancreatic involvement by IgG4-RD. Few data exist to support the utility of conventional glucocorticoid-sparing agents in this disease.

For patients with relapsing or glucocorticoid-resistant disease, B-cell depletion with rituximab is an excellent second-line therapy. Rituximab treatment (two doses of 1 g IV, separated by ~15 days) leads to a swift decline in serum IgG4 concentrations, suggesting that rituximab achieves its effects in part by preventing the repletion of short-lived plasma cells that produce IgG4. More important than its effects on IgG4 concentrations, however, may be the effect of B-cell depletion on T-cell function. Specific effects of rituximab on the CD4+ cytotoxic T cell described above have been documented in IgG4-RD. The rapidly evolving understanding of the pathophysiology of IgG4-RD suggests several novel targeted approaches to treating the disease, some of which are in clinical trials. Phase 3 clinical trials of treatments targeting CD19+ B lymphocytes are now at advanced stages of recruitment. B-cell–targeted treatment strategies may be appropriate first-line therapy options for some patients following confirmation of their efficacy in clinical trials, particularly for patients at high risk for glucocorticoid toxicity and for those with immediately organ-threatening disease.

FURTHER READING

Jha I et al: Sex as a predictor of clinical phenotype and determinant of immune response in IgG4-Related disease: A retrospective study of 328 patients fulfilling the American College of Rheumatology/European League Against Rheumatism classification criteria. Lancet Rheumatol 6:E460, 2024.

Katz G et al: IgG4-related disease as a variable-vessel vasculitis: A case series of 13 patients with medium-sized coronary artery involvement. Semin Arthritis Rheum 60:152184, 2023.

Katz G et al: Proliferative features of IgG4-related disease. Lancet Rheumatol 6:e481, 2024.

Lanzillotta M et al: B-cell depletion therapy in IgG4-related disease: State of the art and future perspectives. Mod Rheumatol 33:258, 2023.

Lanzillotta M et al: Fibrotic phenotype of IgG4-related disease. Lancet Rheumatol 6:e469, 2024.

Perugino CA et al: Efficacy and safety of inebilizumab for IgG4-related disease: Protocol for a randomized clinical trial. Rheumatol Ther 10:1795, 2023.

Perugino CA et al: Evaluation of the safety, efficacy, and mechanism of action of obexelimab for the treatment of patients with IgG4-related disease: An open-label, single-arm, single centre, phase 2 pilot trial. Lancet Rheumatol 5:e442, 2023.

Peyronel F et al: IgG4-related disease: Advances in pathophysiology and treatment. Expert Rev Clin Immunol 19:537, 2023.

381

Familial Mediterranean Fever and Other Hereditary Autoinflammatory Diseases

Daniel L. Kastner

Familial Mediterranean fever (FMF) is the prototype of a group of inherited diseases **(Table 381-1)** that are characterized by recurrent episodes of fever with serosal, synovial, or cutaneous inflammation and, in some individuals, the eventual development of systemic AA amyloidosis **(Chap. 117)**. Because of the relative infrequency of high-titer autoantibodies or antigen-specific T cells, the term *autoinflammatory* has been proposed to describe these disorders, rather than autoimmune. The innate immune system, with its myeloid effector cells and germline receptors for pathogen-associated molecular patterns and endogenous danger signals, plays a predominant role in the pathogenesis of the autoinflammatory diseases. Although the hereditary recurrent fevers comprise an important category of the autoinflammatory diseases, other inherited disorders of inflammation in which recurrent fever plays a less prominent role are now also considered to be autoinflammatory.

BACKGROUND AND PATHOPHYSIOLOGY

FMF was first recognized among Armenians, Arabs, Turks, and non-Ashkenazi (primarily North African and Iraqi) Jews. With the advent of genetic testing, FMF has been documented with increasing frequency among Ashkenazi Jews, Italians, and other Mediterranean and Middle Eastern populations, and occasional cases have been confirmed even in the absence of known Mediterranean ancestry. FMF is generally regarded as recessively inherited, but there is an increasing awareness of clear-cut clinical cases with only a single demonstrable genetic mutation, and for certain relatively rare FMF mutations, there is strong evidence for dominant inheritance. Particularly in countries where families are small, a positive family history can only be elicited in ~50% of cases. DNA testing demonstrates carrier frequencies as high as 1:10 among affected populations, most likely due to some selective advantage.

The FMF gene encodes a 781-amino-acid, ~95-kDa protein denoted *pyrin* that is expressed in granulocytes, eosinophils, monocytes, dendritic cells, and synovial and peritoneal fibroblasts. The N-terminal 92 amino acids of pyrin define a motif, the PYRIN domain, that mediates homotypic protein-protein interactions and has been found in several other proteins, including cryopyrin (NLRP3), which is mutated in three other recurrent fever syndromes. Through the interaction of its PYRIN domain with an intermediary adaptor protein, pyrin nucleates the formation of a macromolecular *pyrin inflammasome* to activate caspase-1 (interleukin [IL] 1β–converting enzyme) and thereby IL-1β and IL-18 secretion and gasdermin D-mediated cell death (pyroptosis). Certain bacterial toxins that block leukocyte cytoskeletal assembly by inactivating RhoA GTPase trigger pyrin inflammasome activation as a part of the normal host defense. At a molecular level, FMF-associated mutations are considered gain-of-function, and in FMF patients, the threshold for pyrin inflammasome activation is thus reduced. Population genetic and immune functional studies support a role for bubonic plague pandemics in selecting for FMF founder mutations that had arisen in Biblical times in the Middle East.

ACUTE ATTACKS

Febrile episodes in FMF may begin even in early infancy; 90% of patients have had their first attack by age 20. Typical FMF episodes generally last 24–72 h, with arthritic attacks tending to last somewhat longer. In some patients, the episodes occur with great regularity, but more often, the frequency of attacks varies over time, ranging from as often as once every few days to remissions lasting several years. Attacks are often unpredictable, although some patients relate them to physical exertion, emotional stress, or menses; pregnancy may be associated with remission.

If measured, fever is nearly always present throughout FMF attacks. Severe hyperpyrexia and even febrile seizures may be seen in infants, and fever is sometimes the only manifestation of FMF in young children.

Over 90% of FMF patients experience abdominal attacks at some time. Episodes range in severity from dull, aching pain and distention with mild tenderness on direct palpation to severe generalized pain with absent bowel sounds, rigidity, rebound tenderness, and air-fluid levels on upright radiographs. Computed tomography (CT) scanning may demonstrate a small amount of fluid in the abdominal cavity. If such patients undergo exploratory laparotomy, a sterile, neutrophil-rich peritoneal exudate is present, sometimes with adhesions from previous episodes. Ascites is rare.

Pleural attacks are usually manifested by unilateral, sharp, stabbing chest pain. Radiographs may show atelectasis and sometimes an effusion. If performed, thoracentesis demonstrates an exudative fluid rich in neutrophils. After repeated attacks, pleural thickening may develop.

FMF arthritis is most frequent among individuals homozygous for the M694V mutation, which is especially common in the non-Ashkenazi Jewish population. Acute arthritis in FMF is usually monoarticular, affecting the knee, ankle, or hip, although other patterns can be seen. Large sterile effusions rich in neutrophils are frequent, without commensurate erythema or warmth. Even after repeated arthritic attacks, radiographic changes are rare. Before the advent of colchicine prophylaxis, chronic arthritis of the knee or hip was seen in ~5% of FMF patients with arthritis. Chronic sacroiliitis can occur in FMF irrespective of the HLA-B27 antigen, even in the face of colchicine therapy. In the United States, FMF patients are much more likely to have arthralgia than arthritis.

The most characteristic cutaneous manifestation of FMF is erysipelas-like erythema, a raised erythematous rash that most commonly occurs on the dorsum of the foot, ankle, or lower leg alone or in combination with abdominal pain, pleurisy, or arthritis. Biopsy demonstrates perivascular infiltrates of granulocytes and monocytes. This rash is seen most often in M694V homozygotes.

Exercise-induced (nonfebrile) myalgia is common in FMF, and a small percentage of patients develop a protracted febrile myalgia that can last several weeks. Symptomatic pericardial disease is rare, although small pericardial effusions may be noted on echocardiography. Unilateral acute scrotal inflammation may occur in prepubertal boys. Aseptic meningitis has been reported in FMF, but the causal connection is controversial. Vasculitis, including IgA vasculitis (Henoch-Schönlein purpura) and polyarteritis nodosa **(Chap. 375)**, may be seen at increased frequency in FMF. The M694V FMF mutation has been shown to be a risk factor for Behçet's syndrome **(Chap. 376)** and ankylosing spondylitis **(Chap. 374)**.

Laboratory features of FMF attacks are consistent with acute inflammation and include an elevated erythrocyte sedimentation rate,

TABLE 381-1 The Hereditary Recurrent Fever Syndromes

	FMF	TRAPS	HIDS/MKD	MWS	FCAS	NOMID
Ethnicity	Jewish, Arab, Turkish, Armenian, Italian	Any ethnic group	Predominantly Dutch, northern European	Any ethnic group	Any ethnic group	Any ethnic group
Inheritance	Recessive or dominant[a]	Dominant	Recessive	Dominant	Dominant	Most commonly de novo mutations; somatic mosaicism in a significant minority
Gene/chromosome	*MEFV*/16p13.3	*TNFRSF1A*/12p13	*MVK*/12q24	*NLRP3*/1q44	*NLRP3*/1q44	*NLRP3*/1q44
Protein	Pyrin	p55 TNF receptor	Mevalonate kinase	NLRP3 (cryopyrin)	NLRP3 (cryopyrin)	NLRP3 (cryopyrin)
Attack length	1–3 days	Often >7 days	3–7 days	1–2 days	Minutes–3 days	Continuous, with flares
Serosa	Pleurisy, peritonitis; asymptomatic pericardial effusions	Pleurisy, peritonitis, pericarditis	Abdominal pain, but seldom peritonitis; pleurisy, pericarditis uncommon	Abdominal pain; pleurisy, pericarditis rare	Rare	Rare
Skin	Erysipeloid erythema	Centrifugally migrating erythema	Diffuse maculopapular rash; oral ulcers	Diffuse urticaria-like rash	Cold-induced urticaria-like rash	Diffuse urticaria-like rash
Joints	Acute monoarthritis; chronic hip arthritis (rare)	Acute monoarthritis, arthralgia	Arthralgia, oligoarthritis	Arthralgia, large joint oligoarthritis	Polyarthralgia	Epiphyseal, patellar overgrowth, clubbing
Muscle	Exercise-induced myalgia common; protracted febrile myalgia rare	Migratory myalgia	Uncommon	Myalgia common	Sometimes myalgia	Sometimes myalgia
Eyes, ears	Uncommon	Periorbital edema, conjunctivitis	Uncommon	Conjunctivitis, episcleritis, optic disc edema; sensorineural hearing loss	Conjunctivitis	Conjunctivitis, uveitis, optic disc edema, blindness, sensorineural hearing loss
CNS	Aseptic meningitis rare	Headache	Headache	Headache	Headache	Aseptic meningitis, seizures
Amyloidosis	Most common in M694V homozygotes	~15% of cases, most often cysteine mutations, T50M	Sometimes associated with V377I/I268T *MVK* genotype	~25% of cases	Uncommon	Late complication
Treatment	Oral colchicine prophylaxis, IL-1 inhibitors for refractory cases	Glucocorticoids, IL-1 inhibitors, etanercept	NSAIDs for fever; IL-1 inhibitors	Canakinumab, rilonacept, anakinra	Canakinumab, rilonacept, anakinra	Anakinra

[a]A substantial percentage of patients with clinical FMF have only a single demonstrable *MEFV* mutation on DNA sequencing.

Abbreviations: CNS, central nervous system; FCAS, familial cold autoinflammatory syndrome; FMF, familial Mediterranean fever; HIDS/MKD, hyperimmunoglobulinemia D with periodic fever syndrome, also known as mevalonate kinase deficiency; IL, interleukin; MWS, Muckle-Wells syndrome; NOMID, neonatal-onset multisystem inflammatory disease; NSAIDs, nonsteroidal anti-inflammatory drugs; TNF, tumor necrosis factor; TRAPS, TNF receptor-associated periodic syndrome.

leukocytosis, thrombocytosis (in children), and elevations in C-reactive protein, fibrinogen, haptoglobin, and serum immunoglobulins. Transient albuminuria and hematuria may also be seen.

AMYLOIDOSIS

Before the advent of colchicine prophylaxis, systemic amyloidosis was a common complication of FMF. It is caused by deposition of a fragment of serum amyloid A, an acute-phase reactant, in the kidneys, adrenals, intestine, spleen, lung, and testes **(Chap. 117)**. Amyloidosis should be suspected in patients who have proteinuria between attacks; renal or rectal biopsy is used most often to establish the diagnosis. Risk factors include the M694V homozygous genotype, positive family history (independent of FMF mutational status), the *SAA1* genotype, male gender, noncompliance with colchicine or IL-1 inhibitory therapy, and national origin (with amyloid risk paralleling infant mortality rate).

DIAGNOSIS

For typical cases, physicians experienced with FMF can often make the diagnosis on clinical grounds alone. Clinical criteria sets for FMF have been shown to have high sensitivity and specificity in parts of the world where the pretest probability of FMF is high. Genetic testing can provide a useful adjunct in ambiguous cases or for physicians not experienced in FMF. Most of the more severe disease-associated FMF mutations are in exon 10 of the gene. An updated, expert-curated list of mutations (as well as benign nonpathogenic polymorphisms) for FMF and other hereditary autoinflammatory diseases can be found online at *http://fmf.igh.cnrs.fr/infevers/*.

Genetic testing has permitted a broadening of the clinical spectrum and geographic distribution of FMF and may be of prognostic value. Most studies indicate that M694V homozygotes have an earlier age of onset and a higher frequency of arthritis, rash, and amyloidosis. In contrast, the E148Q variant in exon 2 is quite common in certain Asian populations and is more likely to affect overall levels of inflammation than to cause clinical FMF. E148Q is sometimes found in *cis* with exon 10 mutations, which may increase the penetrance of the exon 10 mutation. Only ~70% of patients with clinically typical FMF have two identifiable mutations in *trans*, consistent with the concept that FMF mutations are gain-of-function with regard to inflammasome activation, with a dosage effect. In those cases in which only a single mutation is identified, clinical judgment is very important, and sometimes a therapeutic trial of colchicine or an IL-1 inhibitor may help to confirm the diagnosis.

If a patient is seen during their first attack, the differential diagnosis may be broad, although delimited by the specific organ involvement. After several attacks, the differential diagnosis may include the other hereditary recurrent fever syndromes (Table 381-1); the syndrome of periodic fever with aphthous ulcers, pharyngitis, and cervical adenopathy (PFAPA); systemic-onset juvenile rheumatoid arthritis or adult Still's disease; porphyria; hereditary angioedema; inflammatory bowel disease; and, in women, gynecologic disorders.

TREATMENT

Familial Mediterranean Fever

The initial treatment of choice for FMF is daily oral colchicine, which decreases the frequency and intensity of attacks and prevents the development of amyloidosis in compliant patients. Intermittent dosing at the onset of attacks is not as effective as daily prophylaxis and is of unproven value in preventing amyloidosis. The usual adult dose of colchicine is 1.2–1.8 mg/d, which causes substantial reduction in symptoms in two-thirds of patients and some improvement in >90%. Children may require lower doses, although not proportionately to body weight.

Common side effects of colchicine include bloating, abdominal cramps, lactose intolerance, and diarrhea. They can be minimized by starting at a low dose and gradually advancing as tolerated, splitting the dose, use of simethicone for flatulence, and avoidance of dairy products. If taken by either parent at the time of conception, colchicine may cause a small increase in the risk of trisomy 21 (Down's syndrome). Colchicine is usually continued during pregnancy, because the risk of miscarriage due to FMF attacks is thought to outweigh any effect of colchicine on fetal development. In elderly patients with renal insufficiency, colchicine can cause a myoneuropathy characterized by proximal muscle weakness and elevation of the creatine kinase. Cyclosporine inhibits hepatic excretion of colchicine by its effects on the multidrug resistance 1 (MDR1) transport system, sometimes leading to colchicine toxicity in patients who have undergone renal transplantation for amyloidosis. Intravenous colchicine should not be administered to patients already taking oral colchicine, because severe, sometimes fatal, toxicity has been observed in this setting.

For FMF patients who do not respond to colchicine or cannot tolerate therapeutic doses, injectable IL-1 inhibitors may be used. Based on a randomized placebo-controlled phase 3 trial, monthly subcutaneous administration of the monoclonal anti-IL-1β antibody canakinumab received U.S. Food and Drug Administration (FDA) approval for this indication. In a small randomized placebo-controlled trial, weekly subcutaneous rilonacept, a recombinant IL-1 receptor fusion protein, significantly reduced the frequency of attacks. There is also substantial anecdotal experience with daily subcutaneous anakinra, a recombinant IL-1 receptor antagonist, in preventing the acute attacks of FMF and, in some cases, reducing established amyloid deposits. Bone marrow transplantation has been suggested for refractory FMF, but the risk-benefit ratio is usually regarded as unacceptable.

OTHER HEREDITARY RECURRENT FEVERS

TNF RECEPTOR–ASSOCIATED PERIODIC SYNDROME

Tumor necrosis factor (TNF) receptor–associated periodic syndrome (TRAPS) is caused by dominantly inherited mutations in the extracellular domains of the 55-kDa TNF receptor (TNFR1, p55). Although originally described in a large Irish family (and hence the name *familial Hibernian fever*), TRAPS has a broad ethnic distribution. TRAPS episodes often begin in childhood. The duration of attacks ranges from 1 to 2 days to as long as several weeks, and in severe cases, symptoms may be nearly continuous. In addition to peritoneal, pleural, and synovial attacks similar to FMF, TRAPS patients frequently have ocular inflammation (most often conjunctivitis and/or periorbital edema), and a distinctive migratory myalgia with overlying painful erythema may be present. TRAPS patients generally respond better to glucocorticoids than to prophylactic colchicine. Untreated, ~15% develop amyloidosis. The diagnosis of TRAPS is based on the demonstration of a *TNFRSF1A* mutation in the presence of characteristic symptoms. Two particular variants, R92Q and P46L, are common in certain populations and may act more as functional polymorphisms than as disease-causing mutations. In contrast, pathogenic *TNFRSF1A* mutations, including a number of substitutions at highly conserved cysteine residues, are associated with intracellular TNFR1 misfolding, aggregation, and retention, with consequent ligand-independent kinase activation, mitochondrial reactive oxygen species production, and proinflammatory cytokine release. Etanercept, a TNF inhibitor, ameliorates TRAPS attacks, but the long-term experience with this agent has been less favorable. IL-1 inhibition has been beneficial in a large percentage of the patients in whom it has been used, and monthly, subcutaneous canakinumab has received FDA approval for the treatment of TRAPS. Monoclonal anti-TNF antibodies should be avoided, because they may exacerbate TRAPS attacks.

HYPERIMMUNOGLOBULINEMIA D WITH PERIODIC FEVER SYNDROME (ALSO KNOWN AS MEVALONATE KINASE DEFICIENCY)

Hyperimmunoglobulinemia D with periodic fever syndrome (HIDS) is a recessively inherited recurrent fever syndrome found primarily in individuals of northern European ancestry. It is caused by mutations in mevalonate kinase (*MVK*), encoding an enzyme involved in the synthesis of cholesterol and nonsterol isoprenoids, including geranylgeranyl pyrophosphate. The latter compound is essential for proper localization of RhoA GTPase to the cell membrane, and the mislocalization of RhoA leads to its inactivation and the consequent activation of the pyrin inflammasome. HIDS attacks usually begin in infancy and last 3–5 days. Clinically distinctive features include painful cervical adenopathy, a diffuse maculopapular rash sometimes affecting the palms and soles, and aphthous ulcers; pleurisy is rare. Amyloidosis has been observed associated with the V377I/I268T *MVK* genotype. Although originally defined by the persistent elevation of serum IgD, disease activity is not related to IgD levels, and some patients with FMF or TRAPS may have modestly increased serum IgD. Moreover, occasional patients with *MVK* mutations and recurrent fever have normal IgD levels, while all patients with mutations have markedly elevated urinary mevalonate levels during their attacks. For these reasons, some have proposed renaming this disorder *mevalonate kinase deficiency* (MKD). Monthly, subcutaneous canakinumab has been approved by the FDA for the treatment of HIDS/MKD.

NLRP3-ASSOCIATED AUTOINFLAMMATORY DISEASE (ALSO KNOWN AS THE CRYOPYRINOPATHIES OR CRYOPYRIN-ASSOCIATED PERIODIC SYNDROMES)

Three hereditary febrile syndromes, familial cold autoinflammatory syndrome (FCAS), Muckle-Wells syndrome (MWS), and neonatal-onset multisystem inflammatory disease (NOMID), are all caused by mutations in *NLRP3* (formerly known as *CIAS1*), the gene encoding cryopyrin (or NLRP3), and represent a clinical spectrum of disease. FCAS patients develop chills, fever, headache, arthralgia, conjunctivitis, and an urticaria-like rash in response to generalized cold exposure. In MWS, an urticarial rash is noted, but it is not usually induced by cold; MWS patients also develop fevers, abdominal pain, limb pain, arthritis, conjunctivitis, and, over time, sensorineural hearing loss. NOMID is the most severe of the three disorders, with chronic aseptic meningitis, a characteristic arthropathy, and rash. Like the FMF protein pyrin, cryopyrin has an N-terminal PYRIN domain, allowing the formation of an *NLRP3 inflammasome* that mediates caspase-1 activation, IL-1β and IL-18 release, and pyroptosis. Peripheral blood leukocytes from patients with FCAS, MWS, and NOMID release increased amounts of IL-1β upon in vitro stimulation, relative to healthy controls. Macrophages from cryopyrin-deficient mice exhibit decreased IL-1β production in response to certain gram-positive bacteria, bacterial RNA, and monosodium urate crystals. Patients with all three cryopyrinopathies or cryopyrin-associated periodic syndromes (CAPS) show a dramatic response to injections of IL-1 inhibitors. Bimonthly canakinumab and weekly rilonacept are approved by the FDA for the treatment of FCAS and MWS, whereas daily anakinra is approved for the treatment of NOMID.

Approximately one-third of patients with clinical manifestations of NOMID do not have germline mutations in *NLRP3*, but they have been found to be mosaic for somatic *NLRP3* mutations. Such patients also respond dramatically to IL-1 inhibition. Somatic mosaicism in *NLRP3* has been reported rarely in Schnitzler's syndrome, which presents in middle age with recurrent fever, urticarial rash, elevated acute-phase reactants, monoclonal IgM gammopathy, and abnormal bone remodeling. IL-1 inhibition is the treatment of choice for Schnitzler's syndrome.

PERIODIC FEVER WITH APHTHOUS STOMATITIS, PHARYNGITIS, AND CERVICAL ADENITIS

Periodic fever with aphthous stomatitis, pharyngitis, and cervical adenitis (PFAPA) is the most common periodic fever syndrome in children, notable for the almost clock-like regularity of episodes and the tendency for resolution of attacks by early adulthood. PFAPA tends to run in families, but not in a Mendelian fashion. Common shared variants in the *IL12A*, *IL10*, *STAT4,* and *CCR1-CCR3* loci confer susceptibility for a spectrum of phenotypes ranging from common aphthous ulcers to PFAPA to Behçet's syndrome. Therapeutic options for PFAPA include intermittent glucocorticoids; daily oral colchicine, cimetidine, or apremilast; or tonsillectomy/adenoidectomy.

OTHER INHERITED AUTOINFLAMMATORY DISEASES

In the quarter century since the concept of autoinflammation was first proposed, over 60 monogenic autoinflammatory diseases have been described. These discoveries have enabled the understanding of the innate immune system through the lens of human disease and highlight the fact that many cellular homeostatic pathways have evolved to double as direct or indirect innate sensing mechanisms for microbial infection.

The inflammasomopathies are caused by germline or somatic mutations in genes encoding structural components of one of the inflammasomes or of proteins that regulate inflammasome activity, and include FMF, HIDS, and CAPS. Dominantly inherited mutations in the *PSTPIP1* gene lead to noncanonical activation of the pyrin inflammasome and the syndrome of pyogenic arthritis with pyoderma gangrenosum and acne (PAPA). Patients present with episodes of sterile pyogenic arthritis often induced by trauma, severe pyoderma gangrenosum, and severe cystic acne usually beginning in puberty. Arthritic manifestations often respond to IL-1 inhibition. Dominantly inherited mutations in *NLRC4* activate the NLRC4 inflammasome, causing increased IL-1 and IL-18 production. Patients may present with infantile enterocolitis or life-threatening recurrent macrophage activation syndrome later in life. Milder mutations are associated with a cold urticaria syndrome. Inflammasomopathies caused by mutations in *NLRP1* and in the IL-1 receptor antagonist gene (*IL1RN*) have also been described.

Another subset of autoinflammatory diseases is characterized by excessive production of type I interferons. These interferonopathies may be caused by mutations in sensors of intracellular nucleic acids that, under physiologic conditions, would trigger a response to viral infection. For example, de novo germline gain-of-function mutations in *TMEM173* cause constitutive activation of STING, a sensor for intracellular double-stranded DNA, causing severe vasculopathy and pulmonary fibrosis. Retrograde transport of STING from the Golgi apparatus to the endoplasmic reticulum is important for STING inactivation. Loss-of-function mutations in *COPA,* which encodes a chaperone protein involved in retrograde STING transport, cause an autoinflammatory disease characterized by diffuse alveolar hemorrhage or interstitial lung disease, arthritis, and glomerulopathy (COPA syndrome). Mutations in several other genes involved in nucleic acid sensing and/or metabolism cause type I interferonopathies collectively denoted Aicardi-Gutières syndrome. Yet another class of type I interferonopathy is caused by Mendelian or digenic mutations in various components of the proteasome and is sometimes denoted chronic atypical neutrophilic dermatosis with lipodystrophy and elevated temperature (CANDLE) syndrome.

A third subset of autoinflammatory diseases, the relopathies, involves mutations in genes encoding proteins that mediate or regulate NF-κB activation or TNF-induced cell death. These include the genes encoding several ubiquitin ligases and deubiquitinases that, when mutated, may cause clinical syndromes of panniculitis (*OTULIN*), Behçet's-like mucosal ulcerations (*TNFAIP3, RELA*), arthritis (*SHARPIN*), or lymphadenopathy/organomegaly (*RIPK1*). *Blau's syndrome* is caused by mutations in *NOD2* (also known as *CARD15*), which senses certain bacterial products and induces NF-κB activation. Blau's syndrome is caused by gain-of-function mutations in *NOD2/CARD15* and is characterized by granulomatous dermatitis, uveitis, and arthritis; distinct *NOD2/CARD15* variants predispose to Crohn's disease in a non-Mendelian fashion. Autoinflammatory disease associated with certain common variants of *NOD2/CARD15* has been denoted Yao's syndrome, but the causal relationship, if any, between these variants and autoinflammatory phenotypes remains uncertain. ROSAH (retinal dystrophy, optic nerve edema, splenomegaly, anhidrosis, and headaches) syndrome is a relatively newly recognized dominantly inherited autoinflammatory disease caused by gain-of-function mutations in *ALPK1,* encoding a novel kinase that serves as a sensor for bacterial seven-carbon sugars and activates NF-κB. Many patients lose their vision by adulthood; fevers, arthritis, and elevated acute-phase reactants are also frequently observed.

Although the exact disease mechanism remains to be established, recessive loss-of-function germline mutations in the gene encoding adenosine deaminase 2 (ADA2) cause a vasculopathy that can manifest as livedoid rash, early-onset lacunar strokes, and/or polyarteritis nodosa **(Chap. 375)**. TNF inhibitors are highly effective in preventing strokes in these patients. Deficiency of ADA2 (DADA2) may also present with immunodeficiency or bone marrow failure.

Myeloid-restricted somatic mutations in an essential ubiquitylation enzyme have recently been implicated in a severe adult-onset autoinflammatory disease called VEXAS (vacuoles, E1 enzyme, X-linked, autoinflammatory, somatic) syndrome, which can present as relapsing polychondritis **(Chap. 378)**, Sweet's syndrome, vasculitis, or myelodysplastic syndrome. VEXAS syndrome is estimated to affect ~1 in 4000 men over the age of 50.

Finally, it should be noted that several common, genetically complex disorders are now sometimes considered autoinflammatory, because of evidence that elements of the innate immune system, such as the inflammasome, play a role in the pathogenesis and may thus lead to novel therapeutic strategies. Two prominent examples are gout and atherosclerosis. Recent data have implicated endothelial inflammasome activation in the disruption of the blood-brain barrier induced by sepsis.

GLOBAL CONSIDERATIONS

All the disorders discussed in this chapter have been observed in multiple populations. However, as noted herein, FMF is most frequently observed in Mediterranean and Middle Eastern populations and HIDS in northern European populations, particularly the Dutch. A recessive founder mutation in *ADA2* is particularly common in the Middle East and is associated with polyarteritis nodosa.

FURTHER READING

Beck DB et al: Estimated prevalence and clinical manifestations of UBA1 variants associated with VEXAS syndrome in a clinical population. JAMA 329:318, 2023.

Davidson S et al: Dominant negative OTULIN-related autoinflammatory syndrome. J Exp Med 221:e20222171, 2024.

Harapas CR et al: Organellar homeostasis and innate immune sensing. Nat Rev Immunol 22:535, 2022.

Lee PY et al: Evaluation and management of deficiency of adenosine deaminase 2: An international consensus statement. JAMA Netw Open 6:e2315894, 2023.

Lee W et al: Interrupting an IFN-γ-dependent feedback loop in the syndrome of pyogenic arthritis with pyoderma gangrenosum and acne. Ann Rheum Dis 83:787, 2024.

Oda H et al: Biallelic human SHARPIN loss of function induces autoinflammation and immunodeficiency. Nat Immunol 25:764, 2024.

Reygaerts T et al: Pyrin variant E148Q potentiates inflammasome activation and the effect of pathogenic mutations in cis. Rheumatology (Oxford) 63:882, 2024.

Wei C et al: Brain endothelial GSDMD activation mediates inflammatory BBB breakdown. Nature 629:893, 2024.

Section 3 Disorders of the Joints and Adjacent Tissues

382 Approach to Articular and Musculoskeletal Disorders

John J. Cush

Musculoskeletal disorders affect one in three (approximately 127 million) in the U.S., and account for 15–25% of all outpatient visits in the United States. The Centers for Disease Control and Prevention estimates that 54.4 million U.S. adults (22.7%) have physician-diagnosed arthritis. While many patients will have self-limited conditions requiring minimal evaluation and symptomatic therapy, specific musculoskeletal presentations may herald a more serious condition that requires further evaluation or laboratory testing to establish a diagnosis. The goal of the musculoskeletal evaluation is to formulate a differential diagnosis that leads to an accurate diagnosis and timely therapy, while avoiding excessive diagnostic testing and unnecessary treatment (Table 382-1). There are several urgent conditions that must be diagnosed promptly to avoid damage and morbidity. These "red flag" diagnoses include septic arthritis, acute crystal-induced arthritis (e.g., gout), and fracture. Each is suspected by an acute onset of monoarticular or focal musculoskeletal pain.

Most individuals with musculoskeletal complaints can be diagnosed with a thorough history and a comprehensive physical and musculoskeletal examination. The initial encounter should determine whether the musculoskeletal complaint signals a red flag condition (septic arthritis, gout, or fracture) or not. The evaluation should ascertain if the complaint is (1) *articular* or *nonarticular*, (2) *inflammatory* or *noninflammatory*, (3) *acute* or *chronic*, and (4) *localized (monoarticular)* or *widespread (polyarticular)*. The term arthralgia (joint pain) is not synonymous with arthritis, with the latter implying pathology of articular structures. Arthralgia is common and may be a symptom of arthritis.

By answering these four questions, the musculoskeletal presentation can be characterized (e.g., acute inflammatory monoarthritis or a chronic noninflammatory, nonarticular widespread pain) to narrow the diagnostic possibilities. Many musculoskeletal disorders resemble each other at the outset, and some may take weeks or months (but not years) to evolve into a recognizable diagnostic entity. This consideration should temper the desire to establish a definitive diagnosis at the first encounter.

TABLE 382-1 Evaluation of Patients with Musculoskeletal Complaints

- Goals
  - Accurate diagnosis
  - Timely provision of therapy
  - Avoidance of unnecessary diagnostic testing
  - Identification of "red flag" conditions
- Approach
  - Determine the chronology (acute vs chronic)
  - Determine the nature of the pathologic process (inflammatory vs noninflammatory)
  - Anatomic localization of complaint (articular vs nonarticular)
  - Determine the extent of involvement (monoarticular, polyarticular, focal, widespread)
  - Consider the most common disorders first
  - Consider the need for diagnostic testing
  - Formulate a differential diagnosis

ARTICULAR OR NONARTICULAR

The musculoskeletal evaluation must discriminate the anatomic origin(s) of the patient's complaint. For example, ankle pain may result from pathologic conditions involving disparate anatomic structures, including gouty arthritis, calcaneal fracture, Achilles tendinitis, plantar fasciitis, or entrapment neuropathy. Distinguishing between articular and nonarticular conditions requires a thorough examination. Articular structures include the synovium, synovial fluid, articular cartilage, intraarticular ligaments, joint capsule, and juxtaarticular bone. Nonarticular (or periarticular) structures, such as supportive extraarticular ligaments, tendons, bursae, muscle, fascia, bone, nerve, and overlying skin, may be involved in the pathologic process. Musculoskeletal complaints ascribed to the joints are more frequently of nonarticular origin and may be misconstrued as arthritis. Distinguishing between articular and nonarticular/periarticular pain may be challenging to the inexperienced examiner. Articular disorders are distinguished by deep or diffuse pain, limited range of motion on active and passive movement, and swelling (from synovial proliferation, effusion, or bony enlargement), crepitation, instability, "locking," or deformity. By contrast, nonarticular disorders tend to be painful on active, but not passive (or assisted), range of motion. Periarticular conditions often demonstrate point or focal tenderness in regions adjacent to articular structures, may radiate or be elicited with a specific movement or position, and have physical findings remote from the joint capsule. Nonarticular disorders seldom demonstrate swelling, crepitus, instability, or deformity.

INFLAMMATORY OR NONINFLAMMATORY

Early on, the examiner should determine whether the underlying pathologic process is inflammatory or noninflammatory. Inflammatory disorders may be infectious (*Neisseria gonorrhoeae* or *Mycobacterium tuberculosis*), crystal-induced (gout, pseudogout), immune-related (rheumatoid arthritis [RA], systemic lupus erythematosus [SLE]), reactive (rheumatic fever, reactive arthritis), or idiopathic. Inflammatory disorders may be suggested by any of the four cardinal signs of inflammation (erythema, warmth, pain, or swelling), systemic symptoms (fever, rash, weight loss), or laboratory evidence of inflammation (elevated erythrocyte sedimentation rate [ESR] or C-reactive protein [CRP], thrombocytosis, anemia of chronic disease, or hypoalbuminemia). Articular stiffness often accompanies chronic musculoskeletal disorders. The duration of stiffness may be prolonged (hours) with inflammatory disorders (such as RA or polymyalgia rheumatica [PMR]) and improves with activity. By contrast, short-lived or intermittent stiffness (also known as gel phenomenon) is typical of noninflammatory conditions (such as osteoarthritis [OA]), is brief in duration (≤45 min), and is exacerbated by activity. Fatigue may be profound with inflammation (as seen in RA and PMR) but may also be seen with fibromyalgia (a noninflammatory disorder), chronic pain, poor sleep, depression, anemia, cardiac failure, endocrinopathy, malnutrition, deconditioning, or chronic illness.

Noninflammatory disorders may be related to trauma (rotator cuff tear), repetitive use (bursitis, tendinitis), degeneration or ineffective repair (OA), neoplasm (pigmented villonodular synovitis), or pain amplification (fibromyalgia). Noninflammatory disorders are often characterized by pain without synovial swelling or warmth, absence of inflammatory or systemic features, intermittent gel phenomena and normal (for age) or negative laboratory investigations.

The musculoskeletal complaint can be refined and categorized by noting the chronology, pathology, locale, and distribution of the complaint (e.g., acute inflammatory monoarthritis, chronic noninflammatory, nonarticular widespread pain). Narrowing the diagnostic considerations can clarify the need for immediate diagnostic or therapeutic intervention or for continued observation. Figure 382-1 presents an algorithmic approach to the evaluation of patients with musculoskeletal complaints. This approach relies on clinical and historic features, rather than laboratory testing, to diagnose the most common rheumatic disorders.

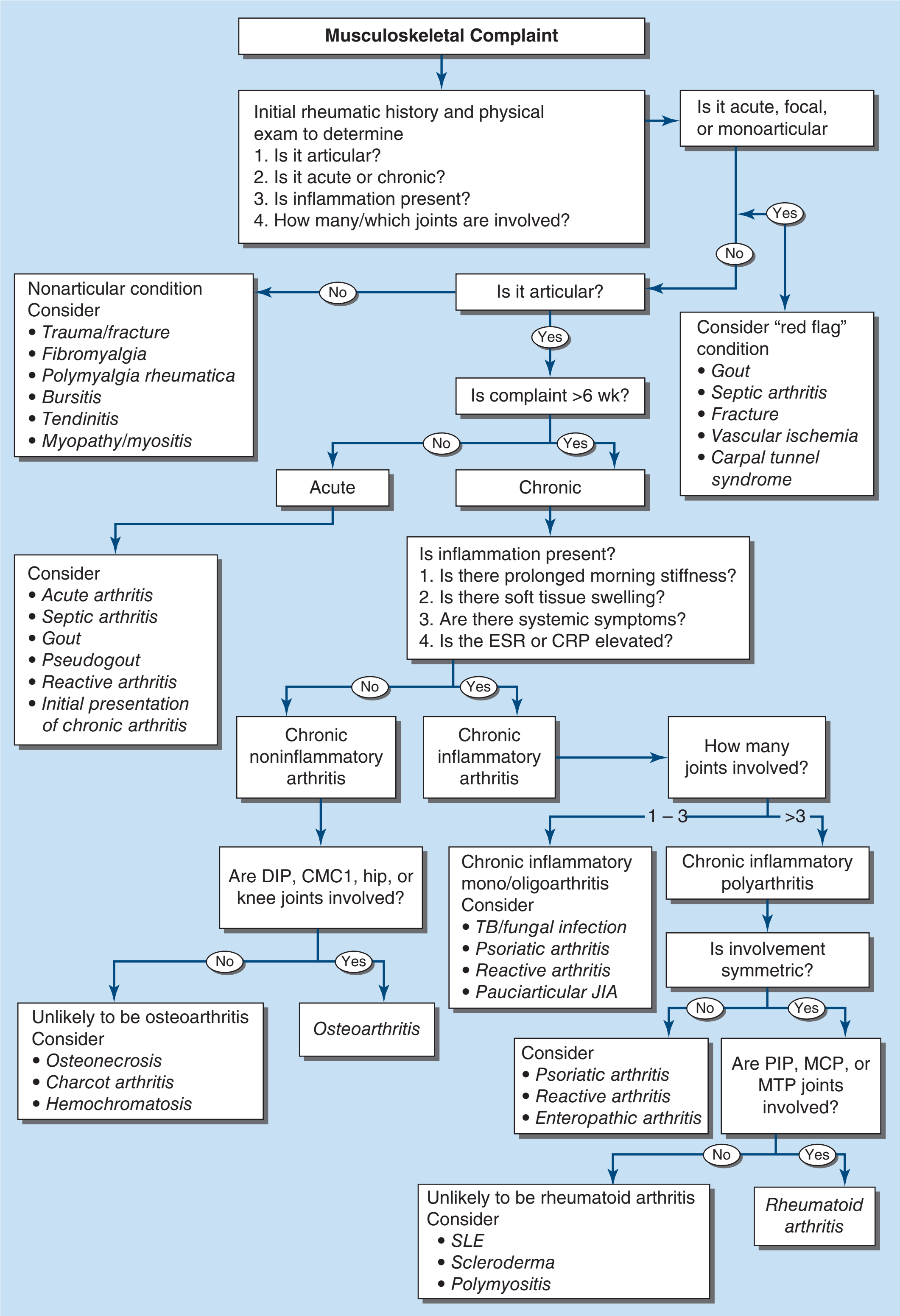
Musculoskeletal Complaint
Initial rheumatic history and physical exam to determine
1. Is it articular?
2. Is it acute or chronic?
3. Is inflammation present?
4. How many/which joints are involved?
Is it acute, focal, or monoarticular
Yes
No
Consider "red flag" condition
• Gout
• Septic arthritis
• Fracture
• Vascular ischemia
• Carpal tunnel syndrome
Is it articular?
No
Nonarticular condition
Consider
• Trauma/fracture
• Fibromyalgia
• Polymyalgia rheumatica
• Bursitis
• Tendinitis
• Myopathy/myositis
Yes
Is complaint >6 wk?
No
Yes
Acute
Chronic
Consider
• Acute arthritis
• Septic arthritis
• Gout
• Pseudogout
• Reactive arthritis
• Initial presentation of chronic arthritis
Is inflammation present?
1. Is there prolonged morning stiffness?
2. Is there soft tissue swelling?
3. Are there systemic symptoms?
4. Is the ESR or CRP elevated?
No
Yes
Chronic noninflammatory arthritis
Chronic inflammatory arthritis
How many joints involved?
1 – 3
>3
Are DIP, CMC1, hip, or knee joints involved?
No
Yes
Unlikely to be osteoarthritis
Consider
• Osteonecrosis
• Charcot arthritis
• Hemochromatosis
Osteoarthritis
Chronic inflammatory mono/oligoarthritis
Consider
• TB/fungal infection
• Psoriatic arthritis
• Reactive arthritis
• Pauciarticular JIA
Chronic inflammatory polyarthritis
Is involvement symmetric?
No
Yes
Consider
• Psoriatic arthritis
• Reactive arthritis
• Enteropathic arthritis
Are PIP, MCP, or MTP joints involved?
No
Yes
Unlikely to be rheumatoid arthritis
Consider
• SLE
• Scleroderma
• Polymyositis
Rheumatoid arthritis

FIGURE 382-1 Algorithm for the diagnosis of musculoskeletal complaints. An approach to formulating a differential diagnosis (shown in italics). CMC, carpometacarpal; CRP, C-reactive protein; DIP, distal interphalangeal; ESR, erythrocyte sedimentation rate; JIA, juvenile idiopathic arthritis; MCP, metacarpophalangeal; MTP, metatarsophalangeal; PIP, proximal interphalangeal; PMR, polymyalgia rheumatica; SLE, systemic lupus erythematosus; TB, tuberculosis.

A simpler, alternative approach would consider the most commonly encountered conditions first, based on the age of the subject. The most prevalent causes of musculoskeletal complaints are shown in **Fig. 382-2**. Because trauma, fracture, low back pain, overuse syndromes, and fibromyalgia are among the most common causes of musculoskeletal pain, these should be considered with each new encounter. If excluded, other frequently occurring disorders should be considered according to the patient's age. Hence, those under age 60 years are commonly affected by repetitive use/strain disorders, gout (men only), RA, spondyloarthritis, and uncommonly, infectious arthritis. Patients over 60 years old are frequently affected by OA, crystal (gout and pseudogout) arthritis, PMR, osteoporotic fracture, and uncommonly, septic arthritis. Notably, these conditions are between 10 and 100 times more prevalent than other serious autoimmune conditions, such as SLE, systemic sclerosis (scleroderma), polymyositis, and vasculitis.

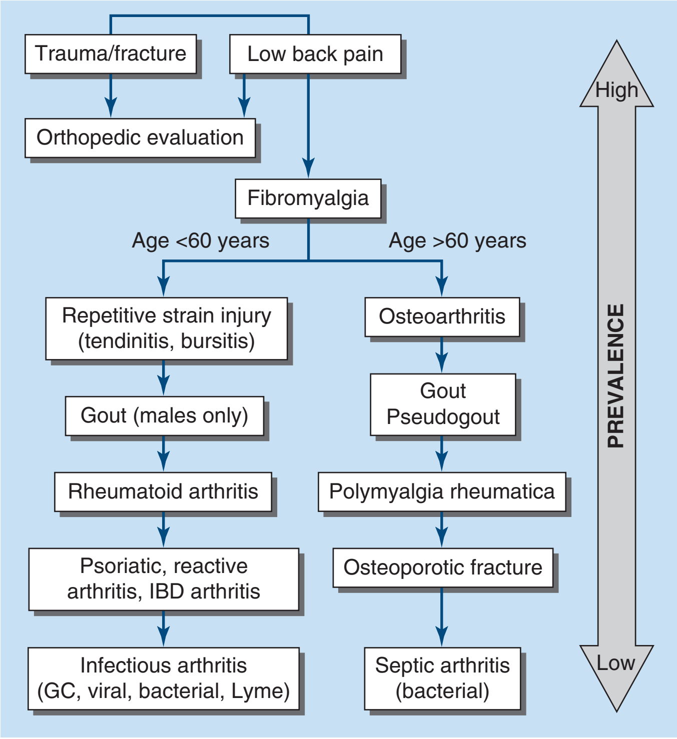

FIGURE 382-2 Algorithm for consideration of the most common musculoskeletal conditions. GC, gonococcal; IBD, inflammatory bowel disease.

CLINICAL HISTORY

Historic features may reveal important insights into the diagnosis. Aspects of the patient profile, complaint chronology, extent of joint involvement, and precipitating factors can provide important information. Certain diagnoses are more frequent in different *age* groups. SLE and gonococcal arthritis occur more frequently in young adults, whereas fibromyalgia and RA are frequent in middle age, and OA and PMR are more prevalent in the elderly. *Sex* and *race* should also be considered. Gout, spondyloarthritis, and ankylosing spondylitis are more common in men, whereas RA, lupus, fibromyalgia, osteoarthritis, and osteoporosis are more frequent in women. *Racial predilections* are common. Thus, PMR, giant cell arteritis, and granulomatosis with polyangiitis (GPA) commonly affect whites, whereas sarcoidosis and SLE more commonly affect African Americans. *Familial aggregation* is unlikely for most arthropathies but may be seen with ankylosing spondylitis, gout, and Heberden's nodes of OA.

The *onset, duration, and evolution* of the complaint are equally important. An abrupt onset is typical of septic arthritis or gout, whereas OA, RA, and fibromyalgia may have more indolent presentations. Musculoskeletal disorders are typically classified by symptom duration as either acute (≤6 weeks) or chronic (>6 weeks). Acute arthropathies tend to be infectious, crystal-induced, or reactive. Chronic conditions may include noninflammatory or immune-mediated (e.g., OA, RA, polymyositis) and nonarticular disorders (e.g., fibromyalgia).

The patients' complaints may evolve differently and be classified as chronic (OA), additive (RA, psoriatic arthritis), intermittent (crystal or Lyme arthritis), or migratory (rheumatic fever, gonococcal or viral arthritis).

The *distribution* of articular involvement is often informative. The number of joints involved defines the complaint as *monoarticular* (one joint), *oligoarticular* or *pauciarticular* (two or three joints), or *polyarticular* (four or more joints). Monoarticular and oligoarticular conditions share similar etiologies. While crystal and infectious arthritis are often mono- or oligoarticular, OA and RA are usually polyarticular. Nonarticular disorders may be classified as either focal or widespread. Tendinitis and carpal tunnel syndrome are typically focal, whereas weakness and myalgia (from polymyositis or fibromyalgia) have more widespread presentations. Joint involvement in RA tends to be symmetric and polyarticular. By contrast, spondyloarthritis, reactive arthritis, and gout are often asymmetric and oligoarticular. OA and psoriatic arthritis may be either symmetric or asymmetric and oligo- or polyarticular. The upper extremities are frequently involved in RA and OA, whereas lower extremity arthritis is characteristic of reactive arthritis and gout at their onset. Involvement of the spine is common in OA and axial spondyloarthritis but is infrequent in RA, with the notable exception of the cervical spine.

The clinical history may also identify *precipitating events*, such as trauma (osteonecrosis, meniscal tear), drug administration (Table 382-2),

TABLE 382-2 Drug-Induced Musculoskeletal Conditions

Arthralgias	Quinidine, cimetidine, beta blockers, quinolones, chronic acyclovir, interferons, IL-2, nicardipine, vaccines, rifabutin, aromatase inhibitors, HIV protease inhibitors, DPP-4 inhibitors (sitagliptin, linagliptin, alogliptin), paclitaxel, checkpoint inhibitors (ipilimumab, pembrolizumab, nivolumab, atezolizumab, durvalumab, cemiplimab), denosumab, teriparatide, abaloparatide, voriconazole
Myalgias/myopathy	Glucocorticoids, penicillamine, hydroxychloroquine, AZT, lovastatin, simvastatin, atorvastatin, pravastatin, clofibrate, amiodarone, interferon, IL-2, alcohol, cocaine, paclitaxel, docetaxel, imatinib mesylate, colchicine, quinolones, cyclosporine, tacrolimus, protease inhibitors, checkpoint inhibitors
Drug-induced dermatomyositis	Hydroxyurea, statins, lipid-lowering agents, penicillamine, immune checkpoint inhibitors, trastuzumab, and tumor necrosis factor inhibitors
Tendon rupture/tendinitis	Quinolones, glucocorticoids, isotretinoin, statins, aromatase inhibitors, collagenase injections
Drug-induced enthesitis	Dupilumab, voriconazole, immune checkpoint inhibitors, fluoroquinolones, retinoids and fluorinated agents
Gout	Diuretics, aspirin, cytotoxics, cyclosporine, tacrolimus, alcohol, moonshine, pyrazinamide, ethambutol, fructose-containing soft drinks, nicotinic acid.
Drug-induced lupus	Hydralazine, procainamide, quinidine, phenytoin, carbamazepine, methyldopa, isoniazid, chlorpromazine, lithium, penicillamine, tetracyclines, TNF inhibitors, ACE inhibitors, ticlopidine, terbinafine, aromatase inhibitors, ethosuximide, vedolizumab, immune checkpoint inhibitors, trastuzumab
Drug-induced subacute lupus	Proton pump inhibitors, calcium channel blockers (diltiazem), ACE inhibitors, TNF inhibitors, terbinafine, interferons (α and β-1a), paclitaxel, docetaxel, gemcitabine, capecitabine, aromatase inhibitors, HCTZ
Osteonecrosis/atypical fractures	Glucocorticoids, alcohol, radiation, bisphosphonates
Osteopenia	Glucocorticoids, chronic heparin, phenytoin, aromatase inhibitors, antiandrogen therapy, thiazolidinediones
Psoriasis	TNF inhibitors, beta blockers, lithium, hydroxychloroquine, chloroquine, minocycline, ACE inhibitors, terbinafine, checkpoint inhibitors
Systemic sclerosis (scleroderma)	Vinyl chloride, bleomycin, baricitinib, pentazocine, organic solvents, carbidopa, tryptophan, rapeseed oil
Raynaud's phenomenon	Cisplatin, bleomycin, beta blockers, clonidine, bromocriptine, ergot alkaloids, cocaine, methylphenidate, dextroamphetamine, phentermine, interferon therapy, tyrosine kinase inhibitors
Vasculitis	Allopurinol, amphetamines, cocaine (often levamisole adulterated), cannabis, thiazides, penicillamine, propylthiouracil, montelukast, TNF inhibitors, hepatitis B vaccine, trimethoprim/sulfamethoxazole, minocycline, hydralazine, apixaban

Abbreviations: ACE, angiotensin-converting enzyme; AZT, zidovudine; HCTZ, hydrochlorothiazide; IL-2, interleukin 2; TNF, tumor necrosis factor.

recent infection (reactive arthritis, hepatitis, chikungunya), or illnesses that may underlie the patient's complaint. Certain comorbidities may have musculoskeletal consequences. This is especially so for diabetes mellitus (carpal tunnel syndrome), renal insufficiency (gout), depression or insomnia (fibromyalgia), myeloma (spinal pain), cancer (myositis), and osteoporosis (fracture). Similarly, there may be musculoskeletal effects with drug use, such as glucocorticoids (osteonecrosis, septic arthritis), diuretics, or chemotherapy (gout) (Table 382-2).

Lastly, an adapted and thorough *rheumatic review of systems* may identify musculoskeletal diagnoses by their associated systemic features, such as fever (Still's disease, infection), rash (SLE, psoriatic arthritis), nail abnormalities (psoriatic or reactive arthritis), myalgias (fibromyalgia, statin- or drug-induced myopathy), or weakness (polymyositis, neuropathy). Musculoskeletal and autoimmune conditions can be suggested by involvement of other organ systems including the eyes (Behçet's disease, sarcoidosis, spondyloarthritis), gastrointestinal tract (systemic sclerosis, inflammatory bowel disease), genitourinary tract (reactive arthritis, gonococci), or nervous system (Lyme disease, vasculitis).

■ FIBROMYALGIA

Fibromyalgia is a common masquerader of musculoskeletal complaints and is often misdiagnosed or not considered. Fibromyalgia **(see Chap. 385)** is a pain amplification disorder heralded by sleep disturbance, exaggerated pain and sensitivity (owing to lowered pain thresholds), and a multiplicity of symptoms, yet a paucity of abnormalities on clinical examination or laboratory testing. Tender "trigger points" may be elicited over the epicondyles, trochanteric bursae, anserine bursae, and specific muscles (gluteal, trapezius, supraspinatus). Fibromyalgia is characterized by widespread aches and pains, even though presenting symptoms may be fewer or focal. Fibromyalgia often coexists with numerous comorbidities including irritable bowel syndrome, dysmenorrhea, migraine, depression, anxiety, memory loss, nonanatomic paresthesia or dysesthesia, fatigue, myalgias, temporomandibular joint pain, hypermobility, and multiple chemical sensitivities. Fibromyalgia affects over 5 million Americans. Early consideration of this very common disorder can avert needless investigation, therapy, and concern for those afflicted (Fig. 382-2).

THE ELDERLY PATIENT

The incidence of rheumatic diseases rises with age, such that 58% of those >65 years will have joint complaints. Musculoskeletal complaints in the elderly are usually ascribed to aging and arthritis, but may also be related to sarcopenia, deconditioning, poor sleep, and malnutrition. Musculoskeletal disorders may also be overlooked because the signs and symptoms may be insidious or overshadowed by comorbidities or care access issues. These difficulties are compounded by the diminished reliability of laboratory testing in the elderly, who often manifest nonpathologic abnormal results. For example, the ESR (that rises with age) may be misleadingly elevated, and low-titer positive tests for rheumatoid factor (RF) and antinuclear antibodies (ANAs) may be seen in up to 15% of elderly patients. Although nearly all rheumatic disorders afflict the elderly, geriatric patients are particularly prone to OA, osteoporosis, osteoporotic fractures, gout, pseudogout, rotator-cuff disorders, PMR, vasculitis, septic arthritis, depression, and drug-induced disorders (Table 382-2). The elderly should be approached in the same manner as other patients with musculoskeletal complaints, but with an emphasis on identifying the potential rheumatic consequences of comorbidities and therapies used in the elderly. The physical examination should identify the nature of the musculoskeletal complaint as well as coexisting diseases that may influence diagnosis and choice of treatment.

THE HOSPITALIZED PATIENT

The hospitalized patient with rheumatic complaints may pose a diagnostic challenge owing to symptom severity, acute plus chronic disease, and greater interplay of comorbidities. Patients with rheumatic disorders tend to be admitted for one of several reasons: (1) acute onset of inflammatory arthritis (possibly gout or septic arthritis); (2) undiagnosed systemic or febrile illness; (3) musculoskeletal trauma; (4) exacerbation or deterioration of an existing musculoskeletal disorder (e.g., SLE); or (5) new medical diagnoses (e.g., thrombotic event, lymphoma, infection) arising in patients with an established rheumatic disorder. Notably, patients with chronic inflammatory disorders (e.g., RA, SLE, psoriatic arthritis) are at increased risk of infection, cardiovascular events, pulmonary disorders, and neoplasia. Rheumatic patients need not be admitted to hospital for widespread pain, serologic abnormalities, or the initiation of new therapies.

Acute monoarticular inflammatory arthritis is a "red flag" presentation (e.g., septic arthritis, gout, pseudogout) that may require arthrocentesis and hospitalization if infection is suspected. New-onset inflammatory polyarthritis has a wide differential diagnosis (e.g., RA, hepatitis-related arthritis, chikungunya arthritis, serum sickness, drug-induced lupus, SLE, polyarticular septic arthritis) and may require targeted laboratory investigations more so than synovial fluid analyses. Patients with febrile, multisystem disorders require exclusion of crystal (gout, pseudogout), infectious, or neoplastic etiologies and an evaluation driven by the dominant symptom/finding with the greatest specificity. Febrile conditions worthy of consideration may include vasculitis (PMR, giant cell arteritis, GPA), adult-onset Still's disease, SLE, antiphospholipid syndrome, IgG4-related disease, and sarcoidosis. A preexisting rheumatic diagnosis (e.g., SLE, RA, ankylosing spondylitis) should be confirmed by careful history and examination and review of medical records, as this will influence the ensuing in-patient evaluation. Notably, when patients with established rheumatic disease are admitted to the hospital, it is usually not for their autoimmune disorder, but rather a comorbid or new medical problem or complication of drug therapy.

Acute gout, pseudogout, or infection, may arise during hospitalization and immobilization. Unfortunately, overly aggressive and unfocused laboratory testing may yield abnormal findings that are better explained by the patient's preexisting condition (chronic lung, renal, or liver disease) rather than a new inflammatory or autoimmune disorder (lupus, vasculitis).

MUSCULOSKELETAL EXAMINATION

The goal of the physical examination is to ascertain the structures involved, the nature of the underlying pathology, the functional consequences of the process, and the presence of systemic or extraarticular manifestations. A knowledge of topographic and functional anatomy is necessary to identify the primary site(s) of involvement and differentiate articular from nonarticular disorders. The musculoskeletal examination depends on careful inspection, palpation, contralateral comparison, and a variety of specific physical maneuvers to elicit diagnostic signs **(Table 382-3)**. Although most articulations of the appendicular skeleton can be examined in this manner, inspection and palpation are not possible for many axial (e.g., zygapophyseal) and inaccessible (e.g., sacroiliac or hip) joints. For such joints, there is a greater reliance on specific maneuvers and imaging for assessment.

Examination of involved and uninvolved joints will determine whether *pain, warmth, erythema*, or *swelling* is present. The locale and level of pain elicited by palpation or movement should be quantified. Examination of 28 easily palpated joints (proximal interphalangeals [PIPs], metacarpophalangeals [MCPs], wrists, elbows, shoulders, and knees) can quantify the number of tender or swollen joints (0–28) involved. Careful examination should distinguish between true articular swelling (caused by bony hypertrophy, synovial effusion or proliferation) and nonarticular (or periarticular) involvement, which usually extends beyond the normal joint margins. Cautious palpation can distinguish synovial effusion (fluctuant swelling) from synovial hypertrophy (grape-like compressibility) and bony hypertrophy (firm as a nut). Small to moderate knee effusions may be identified by the "bulge sign" or "ballottement of the patellae." Bursal effusions (e.g., effusions of the olecranon or prepatellar bursa) are often focal and periarticular, overlie

TABLE 382-3 Glossary of Musculoskeletal Terms

Crepitus	A palpable (less commonly audible) vibratory or crackling sensation elicited with joint motion; fine joint crepitus is common and often insignificant in large joints; coarse joint crepitus indicates advanced cartilaginous and degenerative changes (as in osteoarthritis)
Subluxation	Alteration of joint alignment such that articulating surfaces incompletely approximate each other
Dislocation	Abnormal displacement of articulating surfaces such that the surfaces are not in contact
Range of motion	For diarthrodial joints, the arc of measurable movement through which the joint moves in a single plane
Contracture	Loss of full movement resulting from a fixed resistance caused either by tonic spasm of muscle (reversible) or by fibrosis of periarticular structures (permanent)
Deformity	Abnormal shape, size, or alignment resulting from bony hypertrophy, malalignment of articulating structures, or damage to periarticular supportive structures
Enthesitis	Inflammation of the entheses (tendinous or ligamentous insertions on bone)
Epicondylitis	Infection or inflammation involving an epicondyle

bony prominences, and are fluctuant with defined borders. Joint *stability* (indicative of periarticular support and integrity) can be assessed by stabilizing the proximal joint and applying manual stress to the distal appendage in different planes. *Subluxation* or *dislocation*, which may be secondary to traumatic, mechanical, or inflammatory causes, can be assessed by inspection and palpation. Joint *swelling* or *volume* can be assessed by palpation. Distention of the articular capsule usually causes pain and evident enlargement or fluctuance. The patient will attempt to minimize the pain by maintaining the joint in the position of least intraarticular pressure and greatest volume (usually partial flexion). With chronicity, this may lead to flexion contracture. Active and passive *range of motion* should be assessed in all planes, with contralateral comparison. A goniometer may be used to accurately quantify the arc of movement. Each joint should be passively moved through its full range of motion (including, as appropriate, flexion, extension, rotation, abduction, adduction, lateral bending, inversion, eversion, supination, pronation, medial/lateral deviation, and plantar- or dorsiflexion). Extremes in range of motion and connective tissue laxity may be seen with hypermobility syndrome, Ehlers-Danlos syndrome, or Marfan's syndrome. Limitation of motion or contractures are frequently caused by inflammation, effusion, pain, deformity, or neuromyopathic causes. If passive motion exceeds active motion, a periarticular process (e.g., tendinitis, tendon rupture, or myopathy) should be considered. *Contractures* may reflect antecedent synovial inflammation, trauma, or chronic immobilization. Minor joint *crepitus* is common during joint palpation and maneuvers, but if it becomes coarser or louder, it may indicate significant cartilage degeneration and degenerative changes (e.g., OA). Joint *deformity* usually indicates a long-standing or aggressive pathologic process. Deformities may result from ligamentous destruction, soft tissue contracture, bony enlargement, ankylosis, erosive disease, subluxation, trauma, or loss of proprioception. Examination of the musculature will document strength, atrophy, pain, or spasm. Appendicular muscle weakness should be characterized as proximal or distal. Muscle strength should be assessed by observing the patient's performance (e.g., walking, rising from a chair, grasping, writing). Motor strength may also be graded on a 5-point scale: 0 for no movement; 1 for trace movement or twitch; 2 for movement with gravity eliminated; 3 for movement against gravity only; 4 for movement against gravity and resistance; and 5 for normal strength. The examiner should assess for (often overlooked) nonarticular or periarticular involvement, especially when articular complaints are not supported by objective findings referable to the joint capsule. The identification of soft tissue or nonarticular pain will prevent unwarranted and often expensive laboratory or imaging evaluations. Specific maneuvers may reveal common nonarticular abnormalities, such as a carpal tunnel syndrome (best identified by Durkan's test or Tinel's sign). Other examples of soft tissue abnormalities include olecranon bursitis, epicondylitis (e.g., tennis elbow), enthesitis (e.g., Achilles tendinitis), and tender trigger points associated with fibromyalgia.

REGIONAL RHEUMATIC COMPLAINTS

Focal musculoskeletal complaints may exhibit a predictable pattern of onset, evolution, and localization and are best diagnosed by their localization, along with select maneuvers or tests. Four commonly involved anatomic regions—the hand, shoulder, hip, and knee—are reviewed here.

HAND PAIN

Focal or unilateral hand pain may result from trauma, overuse, infection, or reactive or crystal-induced arthritis. By contrast, bilateral hand complaints commonly suggest a degenerative (e.g., OA), systemic, or inflammatory/immune (e.g., RA) etiology. The distribution or pattern of joint involvement is suggestive of certain disorders **(Fig. 382-3)**. Thus, OA (or degenerative arthritis) may manifest as distal interphalangeal (DIP) and PIP joint pain with bony hypertrophy leading to Heberden's and Bouchard's nodes, respectively. Pain, with or without bony swelling, involving the base of the thumb (first carpometacarpal joint) is also highly suggestive of OA. By contrast, RA manifests as an additive, symmetric, polyarticular arthritis involving the PIP, MCP, intercarpal, and carpometacarpal joints (wrist) with pain and palpable synovial tissue hypertrophy. Psoriatic arthritis may mimic the pattern of joint involvement seen in OA (DIP and PIP joints) but can be

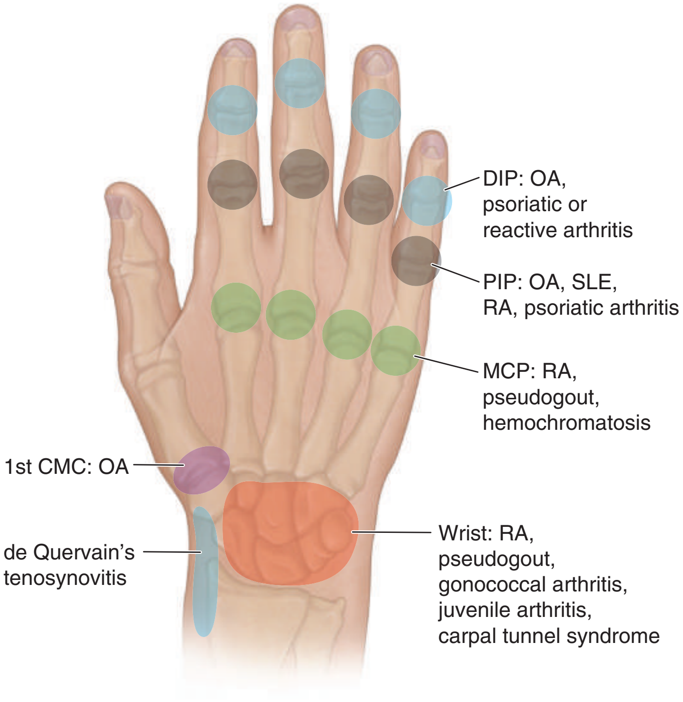
DIP: OA, psoriatic or reactive arthritis
PIP: OA, SLE, RA, psoriatic arthritis
MCP: RA, pseudogout, hemochromatosis
1st CMC: OA
de Quervain's tenosynovitis
Wrist: RA, pseudogout, gonococcal arthritis, juvenile arthritis, carpal tunnel syndrome

FIGURE 382-3 Sites of hand or wrist involvement and their potential disease associations. CMC, carpometacarpal; DIP, distal interphalangeal; MCP, metacarpophalangeal; OA, osteoarthritis; PIP, proximal interphalangeal; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus. *(Reproduced with permission from JJ Cush et al [eds], Evaluation of musculoskeletal complaints, in Rheumatology: Diagnosis and Therapeutics, 2nd ed. Philadelphia, Lippincott Williams & Wilkins, 2005.)*

distinguished by the presence of inflammation (erythema, warmth, synovial swelling), with or without carpal involvement, nail pitting, or onycholysis. Whereas lateral or medial subluxations at the PIP or DIP joints are most likely due to inflammatory OA or psoriatic arthritis, dorsal or ventral deformities (swan neck or boutonnière deformities) are typical of RA. Hemochromatosis should be considered when degenerative changes (bony hypertrophy) are seen at the second and third MCP joints with associated radiographic chondrocalcinosis or episodic, inflammatory wrist arthritis.

Dactylitis manifests as soft tissue swelling of the whole digit and may have a sausage-like appearance. Common causes of dactylitis include psoriatic arthritis, spondyloarthritis, juvenile spondylitis, enthesitis-related arthropathy, mixed connective tissue disease, systemic sclerosis, sarcoidosis, and sickle cell disease. Soft tissue swelling over the dorsum of the hand and wrist may suggest an inflammatory extensor tendon tenosynovitis possibly caused by gonococcal infection, gout, tuberculosis, RA, or giant cell tumor. Tenosynovitis is suggested by swelling or pitting edema (with or without warmth) and may be confirmed when the soft tissue swelling tracks with tendon movement during flexion and extension of fingers, or when pain is induced while stretching the extensor tendon sheaths (flexing the digits distal to the MCP joints and maintaining the wrist in a fixed, neutral position). Volar (palmar) tendon swellings may be from tenosynovitis, Dupuytren's contractures, tendon nodules, or synovial cysts.

Focal wrist pain localized to the radial aspect may be caused by de Quervain's tenosynovitis resulting from inflammation of the tendon sheath(s) involving the abductor pollicis longus or extensor pollicis brevis (Fig. 382-3). This commonly results from occupational or repetitive overuse or follows pregnancy and can be diagnosed with Finkelstein's test. A positive test is present when radial wrist pain is induced after the thumb is flexed and placed inside a clenched fist and the patient actively deviates the hand downward with ulnar deviation at the wrist. Carpal tunnel syndrome is another common disorder of the upper extremity and results from compression of the median nerve within the carpal tunnel. Manifestations include pain in the wrist that may radiate with paresthesia to the thumb, second and third fingers, and radial half of the fourth finger and, at times, atrophy of thenar musculature. Carpal tunnel syndrome is commonly associated with pregnancy, edema, trauma, OA, inflammatory arthritis, and infiltrative disorders (e.g., amyloidosis). The diagnosis may be suggested by a positive Durkan's test or Tinel's sign. With each test, paresthesia in a median nerve distribution is induced or increased by 30 s of compression over the carpal tunnel (Durkan's test) or "thumping" the volar aspect of the wrist and median nerve (Tinel's sign). The low sensitivity and moderate specificity of these tests may require nerve conduction velocity testing to confirm a suspected diagnosis.

SHOULDER PAIN

Shoulder involvement is common in those with a history of trauma, fibromyalgia, infection, inflammatory arthritis, occupational risks, or previous cervical surgery or disease. The patient should be questioned as to the activities or movement(s) that elicit shoulder pain. While arthritis is suggested by pain on movement in all planes, pain with specific or active (not passive) motion suggests a periarticular (nonarticular) process. Shoulder pain may originate in the glenohumeral or acromioclavicular joints, subacromial (subdeltoid) bursa, or periarticular soft tissues (e.g., fibromyalgia, rotator cuff tear/tendinitis) (Fig. 382-4). Shoulder pain may be referred from the cervical spine but may also be referred from intrathoracic lesions (e.g., a Pancoast tumor), gallbladder, liver, or diaphragmatic disease. These same visceral causes may also manifest as focal scapular pain. Fibromyalgia should be suspected when glenohumeral pain is accompanied by diffuse periarticular (i.e., subacromial, bicipital) pain, tender points (i.e., trapezius or supraspinatus), and a sleep disturbance. The shoulder should be put through its full range of motion, both active and passive (with examiner assistance) movements: forward flexion, extension, abduction, adduction, and internal and external rotation. Glenohumeral pain and pathology are best detected by placing the thumb anteriorly over the glenohumeral joint, just medial and inferior to the coracoid process, and applying pressure anteriorly while internally and externally rotating the humeral head. Synovial effusion or tissue is seldom palpable in the shoulder, but, if present, may suggest infection, RA, amyloidosis, or an acute tear of the rotator cuff. The examiner should apply direct manual pressure over the subacromial bursa that lies lateral to and immediately beneath the acromion (Fig. 382-4). Subacromial bursitis is a frequent cause of shoulder pain. Anterior to the subacromial bursa, the bicipital tendon traverses the bicipital groove. This tendon is best identified by palpating it in its groove as the patient rotates the humerus internally and externally. Direct pressure over the tendon may reveal pain indicative of bicipital tendinitis. Palpation of the acromioclavicular joint may disclose local pain, bony hypertrophy, or, uncommonly, synovial swelling. Whereas OA and RA commonly involve the acromioclavicular joint, OA only involves the glenohumeral joint with trauma, occupational causes, or osteonecrosis.

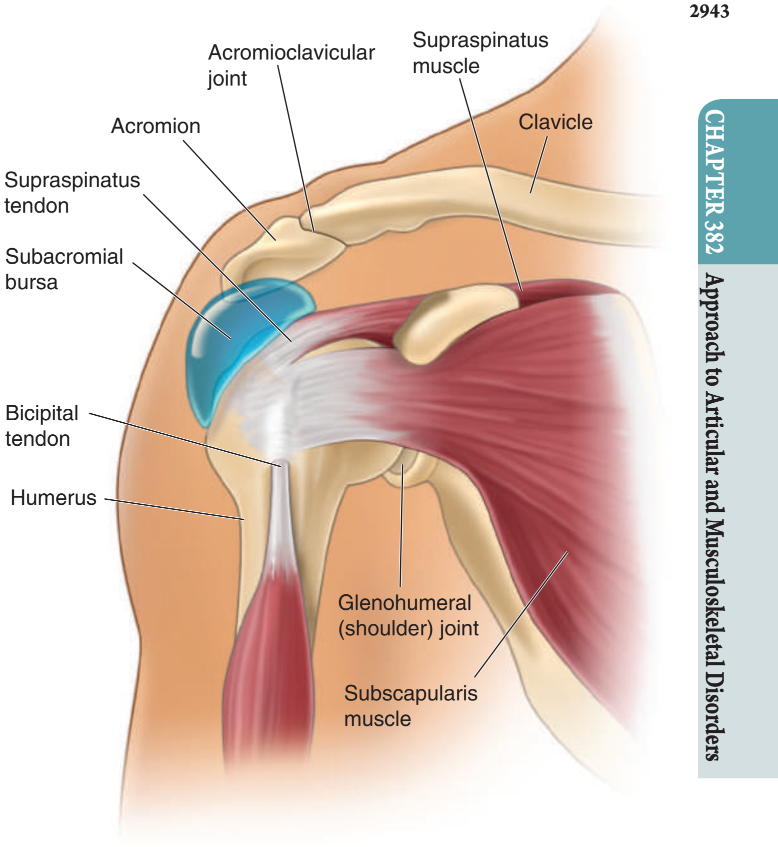

FIGURE 382-4 Origins of shoulder pain. The schematic diagram of the shoulder indicates, with *arrows*, the anatomic origins of shoulder pain.

Rotator cuff tendinitis or tears are a very common cause of shoulder pain, especially in those who are active, manual laborers or have inflammatory arthritis. Nearly 30% of the elderly with shoulder pain will have rotator cuff tendinitis or tear as a primary cause. The rotator cuff is formed by four tendons that attach the scapula to the proximal humerus (supraspinatus, infraspinatus, teres minor, and subscapularis tendons). Of these, the supraspinatus muscle is the most commonly damaged. Rotator cuff tendinitis is suggested by pain on active abduction (but not passive abduction), pain over the lateral deltoid muscle, night pain, and evidence of the impingement signs (pain with overhead arm activities). The Neer test for impingement is performed by the examiner raising the patient's arm into forced flexion while stabilizing and preventing rotation of the scapula. A positive sign is present if pain develops before 180° of forward flexion. Tear of the rotator cuff is common in the elderly and often results from trauma; it may manifest in the same manner as tendinitis. The drop arm test is abnormal with supraspinatus pathology and is demonstrated by passive abduction of the arm to 90° by the examiner. If the patient is unable to hold the arm up actively or unable to lower the arm slowly without dropping, the test is positive. Tendinitis or tear of the rotator cuff is best confirmed by magnetic resonance imaging (MRI) or ultrasound.

KNEE PAIN

Knee pain may result from intraarticular (OA, RA) or periarticular (anserine bursitis, collateral ligament strain) processes or be referred from hip pathology (fracture). A careful history should delineate the chronology of the knee complaint and whether there are predisposing conditions, trauma, medications, or activities that may underlie the complaint. For example, patellofemoral disease (e.g., OA) may cause anterior knee pain that worsens with climbing stairs. Observing the patient's gait may identify a knee, hip, or foot disorder. The knee should be carefully inspected in the upright (weight-bearing) and supine positions for swelling, erythema, malalignment, visible trauma, muscle wasting, and leg length discrepancy (measuring from the anterior superior iliac crest to medial malleolus, with contralateral comparison). The most common malalignment in the knee is *genu varum* (bowlegs) or *genu valgum* (knock-knees) resulting from asymmetric cartilage loss medially or laterally, respectively. Bony swelling of the knee joint commonly results from hypertrophic osseous changes seen with disorders such as OA and neuropathic arthropathy. Swelling caused by hypertrophy of the synovium or synovial effusion may manifest as a fluctuant, ballotable, or soft tissue enlargement in the suprapatellar pouch (suprapatellar reflection of the synovial cavity) or regions lateral and medial to the patella. Synovial effusions may also be detected by balloting the patella downward toward the femoral groove or by eliciting a "bulge sign." With the knee extended, the examiner should manually compress, or "milk," synovial fluid down from the suprapatellar pouch and lateral to the patellae. The application of manual pressure lateral to the patella may cause an observable shift in synovial fluid (bulge) to the medial aspect. The examiner should note that this maneuver is only effective in detecting small to moderate effusions (<100 mL). Inflammatory disorders such as RA, gout, pseudogout, and psoriatic arthritis may involve the knee joint and produce significant pain, swelling, or warmth. A popliteal or *Baker's cyst* may be palpated with the knee partially flexed and is best viewed posteriorly with the patient standing and knees fully extended to visualize isolated or unilateral popliteal swelling or fullness.

Anserine bursitis is an often-missed periarticular cause of knee pain in adults. The pes anserine bursa underlies the insertion of the conjoined tendons (sartorius, gracilis, semitendinosus) on the anteromedial proximal tibia and may be painful following trauma, overuse, or inflammation (bursitis). It is often tender (rarely inflamed) in patients with fibromyalgia, obesity, and knee OA. Other forms of bursitis may also present as knee pain. The prepatellar bursa is superficial and is located over the inferior portion of the patella. The infrapatellar bursa is deeper and lies beneath the patellar ligament before its insertion into the tibial tubercle.

Internal derangement of the knee, from trauma or degenerative processes, implies mechanical disruption of articulating surfaces by damage to cartilage or ligaments or fracture. Damage to the meniscal cartilage (medial or lateral) frequently presents as chronic or intermittent knee pain. Such injury should be suspected when there is a history of trauma, athletic activity, or chronic knee arthritis, and when the patient relates symptoms of "locking," "buckling," or "giving way" of the knee. With the knee flexed 90° and the patient's foot on the table, pain elicited during palpation over the joint line or when the knee is stressed laterally or medially may suggest a meniscal tear. A positive McMurray test may also indicate a meniscal tear. To perform this test, the knee is first flexed at 90°, and the leg is then extended while the lower extremity is simultaneously torqued medially or laterally. A painful click during inward rotation may indicate a lateral meniscus tear, and pain during outward rotation may indicate a tear in the medial meniscus. Lastly, damage to the cruciate ligaments should be suspected with acute onset of pain, possibly with swelling, a history of trauma, or a synovial fluid aspirate that is grossly bloody. Examination of the cruciate ligaments is best accomplished by eliciting a drawer sign. With the patient recumbent, the knee should be partially flexed and the foot stabilized on the examining surface. The examiner should manually attempt to displace the tibia anteriorly or posteriorly with respect to the femur. If anterior movement is detected, then anterior cruciate ligament damage is likely. Conversely, significant posterior movement may indicate posterior cruciate damage. Contralateral comparison will assist the examiner in detecting significant anterior or posterior movement. Collateral ligament damage is suggested by pain on palpation or manual stretch of the ligament (medial or lateral).

HIP PAIN

The hip is best evaluated by observing the patient's gait and assessing range of motion. The vast majority of patients reporting "hip pain" localize their pain unilaterally to the posterior gluteal musculature (Fig. 382-5) with radiation down the posterolateral aspect of the thigh,

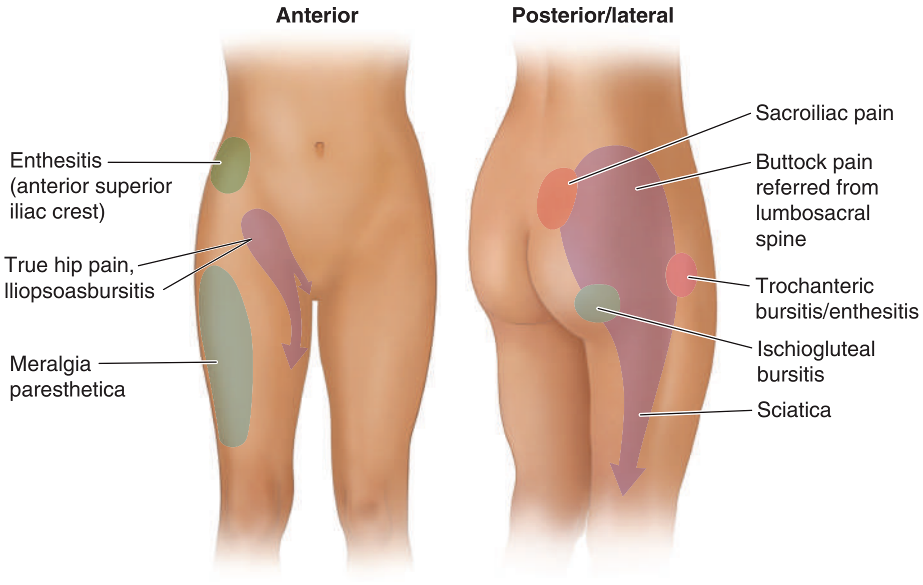
Anterior
Posterior/lateral
Enthesitis (anterior superior iliac crest)
True hip pain, Iliopsoasbursitis
Meralgia paresthetica
Sacroiliac pain
Buttock pain referred from lumbosacral spine
Trochanteric bursitis/enthesitis
Ischiogluteal bursitis
Sciatica

FIGURE 382-5 Origins of hip pain and dysesthesias. *(Reproduced with permission from JJ Cush et al [eds], Evaluation of musculoskeletal complaints, in Rheumatology: Diagnosis and Therapeutics, 2nd ed. Philadelphia, Lippincott Williams & Wilkins, 2005.)*

with or without associated low back pain. This presentation frequently results from degenerative arthritis of the lumbosacral spine or disks and commonly follows a dermatomal distribution with involvement of nerve roots between L4 and S1. Sciatica is caused by impingement of the L4, L5, or S1 nerve (i.e., from a herniated disk) and manifests as unilateral neuropathic pain extending from the gluteal region down the posterolateral leg to the foot. Some individuals instead localize their "hip pain" laterally to the area overlying the trochanteric bursa. Because of the depth of this bursa, swelling and warmth are usually absent. Diagnosis of trochanteric bursitis or enthesitis can be confirmed by inducing point tenderness over the trochanteric bursa. Gluteal and trochanteric pain are common findings in fibromyalgia. Range of movement may be limited by pain. Pain from the "true" hip joint is less common and tends to be located anteriorly, over the inguinal ligament; may radiate medially to the groin; and is worsened by adduction or rotation of the hip. Uncommonly, iliopsoas bursitis may mimic true hip joint pain. Pain associated with iliopsoas bursitis is localized to the groin or anterior thigh and tends to worsen with hyperextension of the hip; many patients prefer to flex and externally rotate the hip to reduce the pain from a distended iliopsoas bursa.

TELEHEALTH EVALUATION

Telemedicine has grown significantly in recent years as a means of remote patient assessment and clinical care. Engagement of patients via electronic telecommunication allows for virtual patient evaluations with a time, cost, and convenience advantage, but can be a disadvantage for the elderly or technologically disadvantaged populations. Telemedicine requires quality audio and video connections and is appropriate for patient education, monitoring, and routine disease assessments. Patients evaluated by telemedicine should undergo the same medical history and inquiry as a routine clinic evaluation but will differ in the scope of examination and maneuvers. The virtual (televideo) musculoskeletal examination may be appropriate and effective in many, with a reliance on functional observation, range of motion, patient self-examination, and contralateral comparison. Specifically, the televideo exam should assess the following: (1) patient gait (to assess lower extremity abnormalities); (2) rising from a seated position (to assess for weakness); (3) cervical range of motion (flexion, extension, lateral bending); (4) upper extremity assessment by contralateral comparison (e.g., "praying hands," "make a fist," "flex wrists," and "hands on ears, elbows out" to assess shoulder range of motion); and (5) side-by-side inspection of knees, ankles, and feet. Telehealth evaluations are contingent on clinical experience, knowledge of and familiarity with the patient, and training the patient on the use of technology and goals and limits of a televideo visit. Telephone-only assessments do not qualify as telehealth visits.

LABORATORY INVESTIGATIONS

Most musculoskeletal disorders can be logically diagnosed by a complete history and physical examination. An additional objective of the initial encounter is to determine whether additional investigations or immediate therapy is required. Laboratory evaluation is indicated with (1) monoarticular presentations; (2) traumatic or inflammatory conditions; (3) the presence of neurologic findings; (4) systemic manifestations; or (5) chronic symptoms (>6 weeks) and a lack of response to symptomatic measures. The extent and nature of the additional investigation should be dictated by the clinical features and suspected pathologic process. Laboratory tests should be used to confirm a specific clinical diagnosis and not be used to screen or evaluate patients with vague rheumatic complaints. Indiscriminate use of broad batteries of diagnostic tests and radiographic procedures is rarely a useful or cost-effective means to establish a diagnosis.

Besides a complete blood count, including a white blood cell (WBC) and differential count, the routine evaluation should include a determination of an acute-phase reactant such as the ESR or CRP, which can be useful in discriminating inflammatory from noninflammatory disorders. Both are inexpensive, easily obtained, and may be elevated with infection, inflammation, autoimmune disorders, neoplasia, pregnancy, renal insufficiency, advanced age, or hyperlipidemia. Extreme elevation of the acute-phase reactants (CRP, ESR) is seldom seen without evidence of serious illness (e.g., sepsis, pleuropericarditis, PMR, giant cell arteritis, Still's disease). Conversely, not all active, inflammatory or autoimmune conditions will have an elevated ESR or CRP; in active SLE and RA patients, nearly 50% will have normal ESR and CRP results.

Serum uric acid determinations are useful in the diagnosis of gout and in monitoring the response to urate-lowering therapy. Uric acid, the product of purine metabolism, is primarily excreted in the urine. Serum values range from 238 to 516 μmol/L (4.0–8.6 mg/dL) in men; lower values (178–351 μmol/L [3.0–5.9 mg/dL]) seen in women are caused by the uricosuric effects of estrogen. Urinary uric acid levels are normally <750 mg per 24 h. Although hyperuricemia is associated with an increased incidence of gout and nephrolithiasis, levels may not correlate with the severity of disease. Uric acid levels (and the risk of gout) may be increased by inborn errors of metabolism (Lesch-Nyhan syndrome), disease states (renal insufficiency, myeloproliferative disease, psoriasis), or drugs (alcohol, cytotoxic therapy, thiazides). Although nearly all patients with gout will demonstrate hyperuricemia at some time during their illness, up to 50% of patients will have normal serum uric acid levels during an acute gouty attack. Monitoring serum uric acid is useful in assessing the response to urate-lowering therapy or chemotherapy, with the target goal being a serum urate <6 mg/dL (<5 mg/dL if tophi are present).

Serologic tests for RF, cyclic anti-citrullinated peptide (CCP or ACPA) antibodies, ANAs, complement levels, Lyme and antineutrophil cytoplasmic antibodies (ANCAs), or antistreptolysin O (ASO) titer should be carried out only when there is sufficient clinical evidence to suggest a specific diagnosis. These have poor predictive value when used for screening, particularly when the pretest probability is low. There is little value to repeated or serial serologic testing. Although 4–5% of a healthy population will have positive tests for RF and ANAs, only 1% and <0.4% of the population will have RA or SLE, respectively. IgM RF (autoantibodies against the Fc portion of IgG) is found in 80% of patients with RA and is moderately specific as it may also be seen in low titers in patients with chronic infections (tuberculosis, leprosy, hepatitis); other autoimmune diseases (SLE, Sjögren's syndrome); and chronic pulmonary, hepatic, or renal disease. When considering RA, both serum RF and anti-CCP antibodies should be obtained, as these are complementary. Both are comparably sensitive, but CCP antibodies are more specific than RF. In RA, the presence of high titers of anti-CCP or RF antibodies, or double positivity, may indicate a greater risk for more severe, erosive polyarthritis. ANAs are found in nearly all patients with SLE and may also be seen in patients with other autoimmune diseases (polymyositis, systemic sclerosis, antiphospholipid syndrome, Sjögren's syndrome), drug-induced lupus (Table 382-2), chronic thyroid, liver, or renal disorders, and advanced age. Positive ANAs are found in 5% of adults, up to 15% of elderly or chronically ill individuals, and up to 80% of patients with thyroid disease. The ANA test is very sensitive but poorly specific for lupus, as only 1–2% of all positive results will be caused by lupus alone. The interpretation of a positive ANA test may depend on the magnitude of the titer and the pattern observed by immunofluorescence microscopy **(Table 382-4)**. Diffuse and speckled patterns are least specific, whereas a peripheral, or rim, pattern (related to autoantibodies against double-strand [native] DNA) is highly specific and suggestive of lupus. Centromeric patterns are seen in patients with limited cutaneous systemic sclerosis (*c*alcinosis, *R*aynaud's phenomenon, *e*sophageal involvement, *s*clerodactyly, *t*elangiectasia [CREST] syndrome), primary biliary sclerosis, Sjögren's syndrome, or thyroiditis, and nucleolar patterns may be seen in patients with diffuse cutaneous systemic sclerosis or inflammatory myositis. More detailed serologic autoantibody profiling may be appropriate when specific autoimmune disorders (Sjögren's, dermatomyositis, systemic sclerosis) are diagnosed, but should not be employed as screening tests.

Aspiration and analysis of synovial fluid are always indicated in acute monoarthritis or when an infectious or crystal-induced arthropathy is suspected. Synovial fluid may distinguish between noninflammatory

TABLE 382-4 Antinuclear Antibody (ANA) Patterns and Clinical Associations

ANA PATTERN	ANTIGEN IDENTIFIED	CLINICAL CORRELATE
Diffuse	Deoxyribonucleoprotein	Nonspecific
	Histones	Drug-induced lupus, lupus
Peripheral (rim)	ds-DNA	50% of SLE (specific)
Speckled	U1-RNP	>90% of MCTD
	Sm	30% of SLE (specific)
	Ro (SS-A)	Sjögren's 60%, SCLE, neonatal lupus, ANA (–) lupus
	La (SS-B)	50% of Sjögren's, 15% lupus
	Scl-70 (topoisomerase I)	40% of diffuse cutaneous systemic sclerosis
	Jo-1 (histidyl t-RNA synthetase)	PM with pneumonitis + arthritis
Nucleolar	RNA polymerase I, others	40% of PSS
Centromere	Kinetochore	75% CREST (limited cutaneous systemic sclerosis), PBC, Sjögren's, thyroiditis

Abbreviations: CREST, *c*alcinosis, *R*aynaud's phenomenon, *e*sophageal involvement, *s*clerodactyly, and *t*elangiectasia; MCTD, mixed connective tissue disease; PBC, primary biliary cirrhosis; PSS, progressive systemic sclerosis; SCLE, subacute cutaneous lupus erythematosus; SLE, systemic lupus erythematosus.

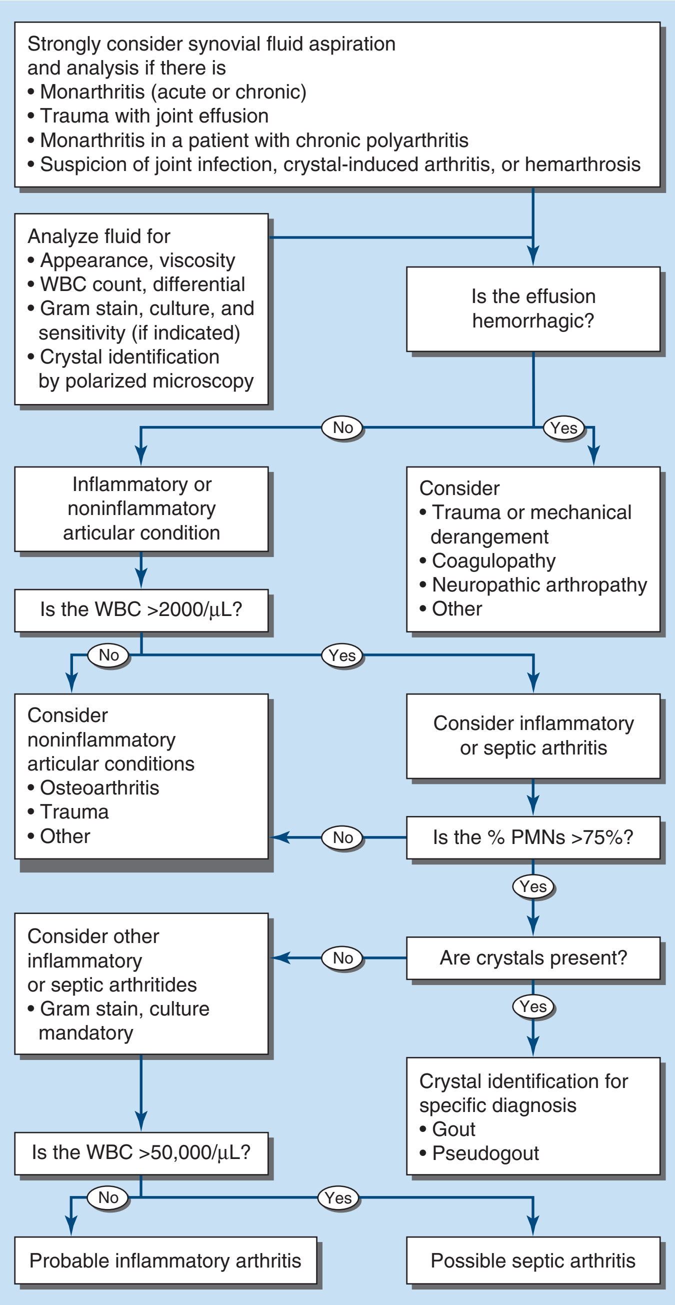
Strongly consider synovial fluid aspiration and analysis if there is
• Monarthritis (acute or chronic)
• Trauma with joint effusion
• Monarthritis in a patient with chronic polyarthritis
• Suspicion of joint infection, crystal-induced arthritis, or hemarthrosis
Analyze fluid for
• Appearance, viscosity
• WBC count, differential
• Gram stain, culture, and sensitivity (if indicated)
• Crystal identification by polarized microscopy
Is the effusion hemorrhagic?
No
Yes
Inflammatory or noninflammatory articular condition
Consider
• Trauma or mechanical derangement
• Coagulopathy
• Neuropathic arthropathy
• Other
Is the WBC >2000/µL?
No
Yes
Consider noninflammatory articular conditions
• Osteoarthritis
• Trauma
• Other
Consider inflammatory or septic arthritis
No
Is the % PMNs >75%?
Yes
Consider other inflammatory or septic arthritides
• Gram stain, culture mandatory
No
Are crystals present?
Yes
Crystal identification for specific diagnosis
• Gout
• Pseudogout
Is the WBC >50,000/µL?
No
Yes
Probable inflammatory arthritis
Possible septic arthritis

FIGURE 382-6 Algorithmic approach to the use and interpretation of synovial fluid aspiration and analysis. PMNs, polymorphonuclear (leukocytes); WBC, white blood cell count.

and inflammatory processes by analysis of the appearance, viscosity, and cell count. Tests for synovial fluid glucose, protein, lactate dehydrogenase, lactic acid, or autoantibodies are not recommended as they have no diagnostic value. Normal synovial fluid is clear or a pale straw color and is viscous, primarily because of the high levels of hyaluronate. Noninflammatory synovial fluid is clear, viscous, and amber-colored, with a WBC count of <2000/μL and a predominance of mononuclear cells. The viscosity of synovial fluid is assessed by expressing fluid from the syringe one drop at a time. Normally, there is a stringing effect, with a long tail behind each synovial drop. Effusions caused by OA or trauma will have normal viscosity or "string sign." Inflammatory fluid is turbid and yellow, with an increased WBC count (2000–50,000/μL) and a polymorphonuclear leukocyte predominance. Inflammatory fluid has reduced viscosity (no stringing), diminished hyaluronate, and little or no tail following each drop of synovial fluid. Such effusions are found in RA, gout, and other inflammatory arthritides. Septic fluid is opaque and purulent, with a WBC count usually >50,000/μL, a predominance of polymorphonuclear leukocytes (>75%), and low viscosity. Such effusions are typical of septic arthritis but may also occur with RA or gout. Lastly, hemorrhagic synovial fluid (hemarthrosis) may be seen with trauma (ligament or cartilage tears), osteochondral fracture, neuropathic arthritis, or coagulopathy. An algorithm for synovial fluid aspiration and analysis is shown in Fig. 382-6. Synovial fluid should be analyzed immediately for appearance, viscosity, and cell count. Monosodium urate crystals (observed in gout) are seen by polarized microscopy and are long, needle-shaped, negatively birefringent, and usually intracellular. In chondrocalcinosis and pseudogout, calcium pyrophosphate dihydrate crystals are usually short, rhomboid-shaped, and positively birefringent. Whenever infection is suspected, synovial fluid should be Gram stained and cultured and tested by metagenomic sequencing, when appropriate. If gonococcal arthritis is suspected, nucleic acid amplification tests should be used to detect either *Chlamydia trachomatis* or *N. gonorrhoeae* infection. Synovial fluid from patients with chronic monoarthritis should also be cultured for *M. tuberculosis* and fungi. Last, it should be noted that crystal-induced arthritis and septic arthritis occasionally occur together in the same joint.

DIAGNOSTIC IMAGING

Conventional radiography has been a valuable tool in the diagnosis and staging of articular disorders. Plain x-rays are most appropriate and cost effective when there is a history of trauma, suspected chronic infection, progressive disability, or monoarticular involvement; when therapeutic alterations are considered; or when a baseline assessment is desired for what appears to be a chronic process. However, in acute or early arthritis, early radiography is rarely helpful in establishing a diagnosis and may only reveal soft tissue swelling or juxtaarticular demineralization. As the disease progresses, conventional radiographs may disclose calcification (of soft tissues, cartilage, or periarticular bone), joint space narrowing, erosions, bony ankylosis, new bone formation (sclerosis, osteophytes, or periostitis), or subchondral cysts may develop and suggest specific clinical entities. Advanced imaging with computed tomography (CT) or magnetic resonance imaging (MRI) is more expensive but may reveal the extent of articular or periarticular involvement. Consultation with a radiologist will help define the optimal imaging modality, technique, or positioning to optimize interpretation and prevent the need for further studies.

Additional imaging techniques may be indicated in selected circumstances or when conventional radiography is inadequate or nondiagnostic **(Table 382-5)**. *Ultrasonography* is useful in the detection of soft tissue abnormalities (tendinitis, tenosynovitis, enthesitis, bursitis), crystal deposition, and entrapment neuropathies. Lower cost, better technology, and enhanced site-specific transducers now allow for wider use and diagnostic specificity, especially when considering synovial (Baker's) cysts, rotator cuff tears, tendinitis (PMR) and tendon injury, and crystal deposition on cartilage. Use of power Doppler allows for early detection of synovitis and bony erosions. *Radionuclide scintigraphy* is a very sensitive, but poorly specific, means of detecting inflammatory or metabolic alterations in bone or periarticular soft tissue structures (Table 382-5). Scintigraphy is best suited for total-body assessment (extent and distribution) of skeletal involvement (neoplasia, Paget's disease) or the assessment of patients with undiagnosed polyarthralgias, looking for occult arthritis or inflammation. The use of scintigraphy has declined with greater use, specificity, and declining cost of ultrasound and MRI. MRI has largely replaced scintigraphy in diagnosing osseous infection, neoplasia, inflammation, increased blood flow, bone remodeling, heterotopic bone formation, or avascular necrosis. Gallium scanning uses ^{67}Ga, which binds serum and cellular transferrin and lactoferrin and is preferentially taken up by neutrophils, macrophages, bacteria, and tumor tissue (e.g., lymphoma). As such, it is primarily used in the identification of occult infection or malignancy. Scanning with ^{111}In-labeled WBCs has been used to detect osteomyelitis and infectious or inflammatory arthritis. Despite their utility, ^{111}In-labeled WBC and ^{67}Ga scanning have also been largely replaced by MRI, except when there is a suspicion of septic joint or prosthetic joint infections.

CT provides detailed visualization of the axial skeleton. Articulations previously considered difficult to visualize by radiography (e.g., zygapophyseal, sacroiliac, sternoclavicular, hip joints) can be evaluated using CT. CT has been demonstrated to be useful in the diagnosis of low back pain (e.g., spinal stenosis vs herniated disk), sacroiliitis, osteoid osteoma, and stress fractures. Helical or spiral CT (with or without contrast angiography) is a novel technique that is rapid, cost effective, and sensitive in diagnosing pulmonary embolism or obscure fractures, often in the setting of initially equivocal findings. High-resolution CT can be advocated in the evaluation of suspected or established infiltrative lung disease (e.g., systemic sclerosis or rheumatoid lung). The recent use of hybrid (positron emission tomography [PET] or single-photon emission CT [SPECT]) CT scans in metastatic evaluations has incorporated CT to provide better anatomic localization of scintigraphic abnormalities.

^{18}F-Fluorodeoxyglucose (FDG) is the most used radiopharmaceutical in PET scanning. FDG-PET/CT scans are rarely indicated in septic or inflammatory arthritis but have been useful in the evaluation of patients with undiagnosed fever, systemic illness, or suspected large vessel vasculitis. FDG-PET may be better at assessing inflammation, whereas MRI and CT can better define the extent of tissue damage. Dual-energy CT (DECT) scanning, developed in urology to identify urinary calculi, has been a highly sensitive and specific method used to identify and quantify uric acid deposition in tissues **(Fig. 382-7)**.

MRI has significantly revolutionized the imaging of musculoskeletal structures, especially in patients with traumatic, systemic, infectious, or neoplastic disorders. MRI has the advantages of providing multiplanar images with fine anatomic detail and contrast resolution **(Fig. 382-8)** to visualize fascia, vessels, nerve, muscle, cartilage, ligaments, tendons, pannus, synovial effusions, and bone marrow.

Because of its sensitivity to changes in marrow fat, MRI is a sensitive but nonspecific means of detecting osteonecrosis, osteomyelitis, and marrow inflammation indicating overlying synovitis or osteitis (Fig. 382-8). Because of its enhanced soft tissue resolution, MRI is more sensitive than arthrography or CT in the diagnosis of soft tissue injuries (e.g., meniscal and rotator cuff tears); intraarticular derangements; marrow abnormalities (osteonecrosis, myeloma); and spinal cord or nerve root damage, synovitis, or cartilage damage or loss. Although more costly with a longer procedural time than CT, MRI has become the preferred technique when evaluating trauma and complex musculoskeletal disorders.

TABLE 382-5 Diagnostic Imaging Techniques for Musculoskeletal Disorders

METHOD	COST[a]	CURRENT INDICATIONS	CAVEATS
Ultrasound	++	Synovial (Baker's) cysts Rotator cuff tears Bursitis, tendinitis, tendon injury Enthesitis Carpal tunnel syndrome Urate or calcium pyrophosphate deposition on cartilage Early detection of synovial inflammation or erosions Ultrasound-guided injection/arthrocentesis	Operator dependent
Radionuclide scintigraphy	++	Metastatic bone survey Evaluation of Paget's disease Identifying occult arthritis in patients with undiagnosed polyarthralgia Acute infection Prosthetic infection Acute osteomyelitis Acute and chronic infection Acute osteomyelitis	Low sensitivity Resolution issues
^{99m}Tc	++	Arthritis Inflammation Degeneration Metastases Infection	Ionizing radiation exposure
^{111}In-WBC	+++	Septic arthritis	
^{67}Ga	++++	Prosthetic joint infection	
Computed tomography (CT)	++++	Herniated intervertebral disk Sacroiliitis Spinal stenosis Spinal trauma Osteoid osteoma Stress fracture	Ionizing radiation exposure Contrast use Availability of CT angiography
Dual-energy CT	+++	Uric acid deposition Tophus localization	Not widely available
Magnetic resonance imaging	++++	Avascular necrosis Osteomyelitis Septic arthritis, infected prosthetic joints Sacroiliitis Intraarticular derangement and soft tissue injury Derangements of axial skeleton and spinal cord Herniated intervertebral disk Pigmented villonodular synovitis Inflammatory and metabolic muscle pathology	Cost Contrast use No ionizing radiation Soft tissue detail
FDG-PET CT	+++++++		Very expensive Whole-body scanning Limited availability Ionizing radiation exposure

[a] Imaging cost, relative to conventional radiography.

Abbreviations: FDG-PET, fluorodeoxyglucose–positron emission tomography; WBC, white blood cell.

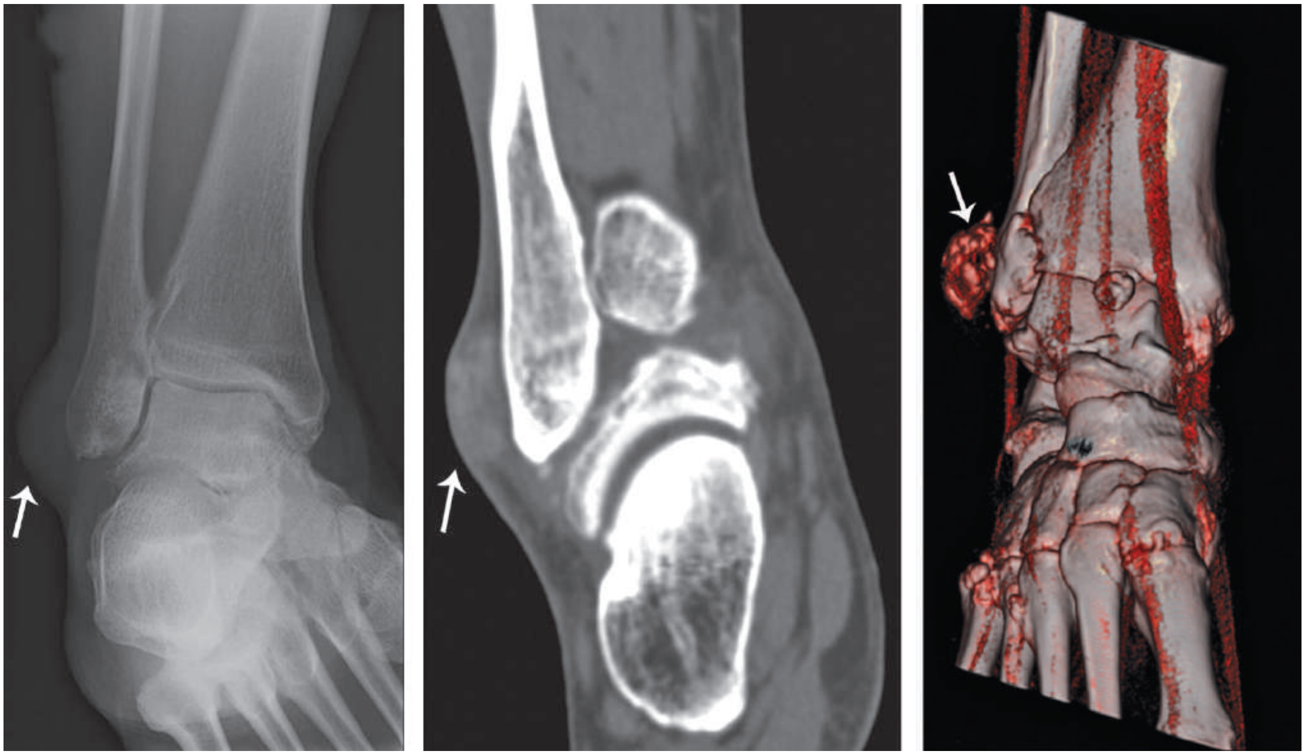

FIGURE 382-7 Dual-energy computed tomography (DECT) scan from a 45-year-old woman with right ankle swelling around the lateral malleolus. Three-dimensional volume-rendered coronal reformatted DECT image shows that the mass is composed of monosodium urate (*red*) in keeping with tophus (*arrow*). *(Reprinted from S Nicolaou et al: Dual-energy CT as a potential new diagnostic tool in the management of gout in the acute setting. AJR Am J Roentgenol 194:1072, 2010.)*

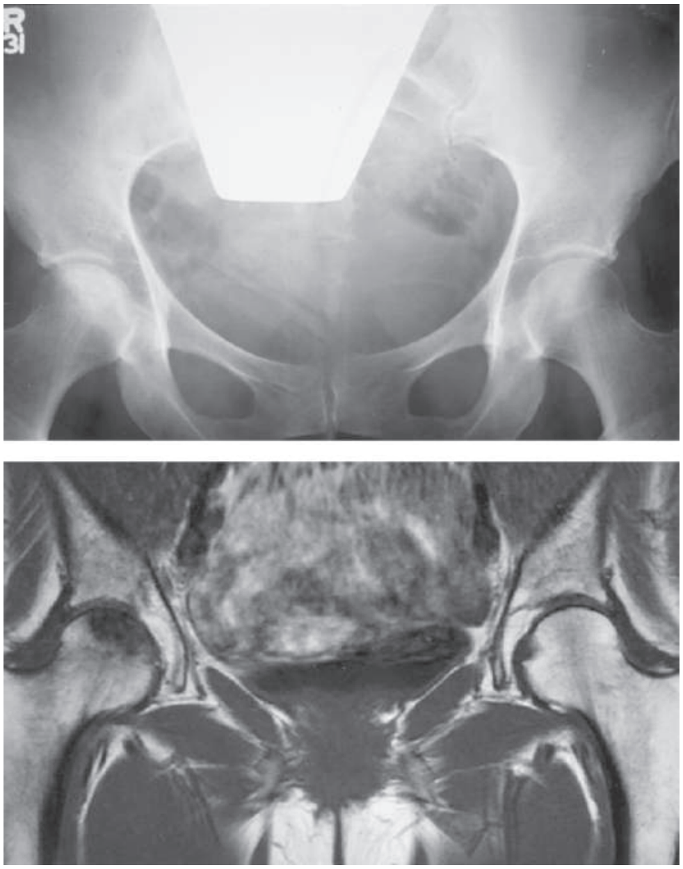

FIGURE 382-8 Superior sensitivity of magnetic resonance imaging (MRI) in the diagnosis of osteonecrosis of the femoral head. A 45-year-old woman receiving high-dose glucocorticoids developed right hip pain. Conventional x-rays (*top*) demonstrated only mild sclerosis of the right femoral head. T1-weighted MRI (*bottom*) demonstrated low-density signal in the right femoral head, diagnostic of osteonecrosis.

ACKNOWLEDGMENT

The author acknowledges the sage contributions of Dr. Peter E. Lipsky to this chapter in previous editions.

FURTHER READING

Cush JJ et al: Evaluation of musculoskeletal complaints. Available from *http://www.rheumaknowledgy.com/evaluation-of-musculoskeletal-complaints*. Accessed April 22, 2024.

Hubbard MJ et al: Common soft tissue musculoskeletal pain disorders. Prim Care 45:289, 2018.

Laskowski ER et al: The telemedicine musculoskeletal examination. Mayo Clinic Proc 95:1715, 2020.

Nguyen AT et al: Musculoskeletal health: An ecological study assessing disease burden and research funding. Lancet Reg Health Am 29:100661, 2024.

Olsen NJ, Karp DR: Finding lupus in the ANA haystack. Lupus Sci Med 7:e000384, 2020.

Rönnelid J et al: Autoantibodies in rheumatoid arthritis: Laboratory and clinical perspectives. Front Immunol 12:685312, 2021.

Rudwaleit M et al: How to diagnose axial spondyloarthritis early. Ann Rheum Dis 63:535, 2004.

Sciascia S et al: Autoantibodies testing in autoimmunity: Diagnostic, prognostic and classification value. Autoimmun Rev 22:103356, 2023.

Simpfendorfer CS: Radiologic approach to musculoskeletal infections. Infect Dis Clin North Am 31:299, 2017.

383 Osteoarthritis

David T. Felson, Tuhina Neogi

Osteoarthritis (OA) is the most common type of arthritis. Its high prevalence, especially in the elderly, and its negative impact on physical function make it a leading cause of disability in the elderly. Because of the aging of Western populations and because obesity, a major risk factor, is increasing in prevalence, the occurrence of OA is on the rise.

OA affects certain joints yet spares others **(Fig. 383-1)**. Commonly affected joints include the hip, knee, and first metatarsal phalangeal joint (MTP) and cervical and lumbosacral spine. In the hands, the distal and proximal interphalangeal joints and the base of the thumb are often affected. Usually spared are the wrist, elbow, and ankle. Our joints were designed, in an evolutionary sense, for brachiating apes, animals that still walked on four limbs. We thus develop OA in joints that were ill designed for human tasks such as pincer grip (OA in the thumb base) and walking upright (OA in knees and hips). Some joints, like the ankles, may be spared because their articular cartilage may be uniquely resistant to loading stresses.

OA can be diagnosed based on structural abnormalities or on the symptoms these abnormalities evoke. According to cadaveric studies, by elderly years, structural changes of OA are nearly universal. These include cartilage loss (seen as joint space loss on x-rays) and osteophytes. Many persons with x-ray evidence of OA have no joint symptoms, and although the prevalence of structural abnormalities is of interest in understanding disease pathogenesis, what matters more from a clinical perspective is the prevalence of symptomatic OA. Symptoms, usually joint pain, determine disability, visits to clinicians, and disease costs.

Symptomatic OA of the knee (pain on most days of a recent month plus x-ray evidence of OA in that knee) occurs in ~12% of persons age ≥60 in the United States and 6% of all adults age ≥30. Symptomatic hip OA is roughly one-third as common as disease in the knee. Although radiographic hand OA and the appearance of bony enlargement in affected hand joints **(Fig. 383-2)** are extremely common in older persons, most affected persons have no pain. Even so, painful hand OA occurs in ~10% of elderly individuals and often produces measurable limitation in function.

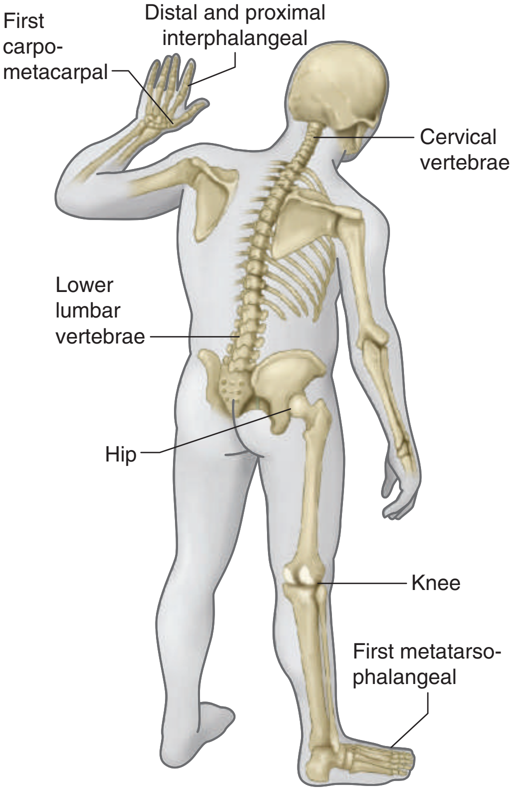

FIGURE 383-1 Joints commonly affected by osteoarthritis.

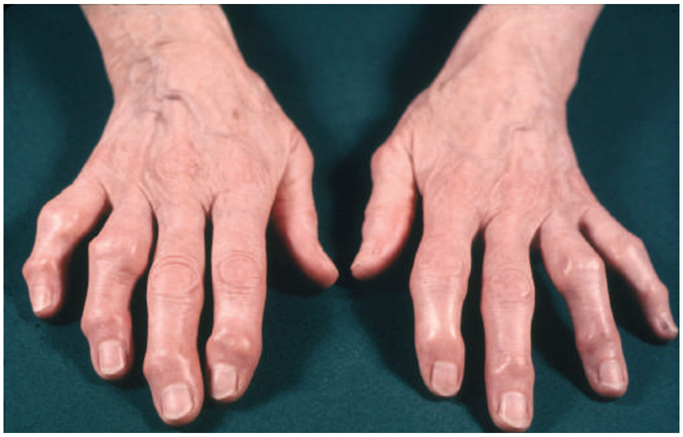

FIGURE 383-2 Severe osteoarthritis of the hands affecting the distal interphalangeal joints (Heberden's nodes) and the proximal interphalangeal joints (Bouchard's nodes). There is no clear bony enlargement of the other common site in the hands, the thumb base.

The prevalence of OA rises strikingly with age, being uncommon in adults aged <40 and highly prevalent in those aged >60. It is also a disease that, at least in middle-aged and elderly persons, is much more common in women than in men.

X-ray evidence of OA is common in the lower back and neck, but back pain and neck pain have not been tied to findings of OA on x-ray. Thus, back pain and neck pain are treated separately **(Chaps. 18 and 19)**.

■ GLOBAL CONSIDERATIONS

With the aging of the populations, both the prevalence of OA and the amount of disability worldwide related to OA have been increasing, especially in developed countries where many are living into old age. Hip OA is rare in China and in immigrants from China to the United States. Anatomic differences between Chinese and white hips may account for much of the difference in hip OA prevalence, with white hips having a higher prevalence of anatomic predispositions to the development of OA. However, OA in the knees is at least as common, if not more so, in Chinese as in whites from the United States, and knee OA represents a major cause of disability in China, especially in rural areas.

DEFINITION

OA is joint failure, a disease in which all structures of the joint have undergone pathologic change, often in concert. The pathologic sine qua non of disease is hyaline articular cartilage loss, present in a focal and, initially, nonuniform manner. This is accompanied by increasing thickness and sclerosis of the subchondral bony plate, by outgrowth of osteophytes at the joint margin, by stretching of the articular capsule, by variable degrees of synovitis, and by weakness of muscles bridging the joint. In knees, meniscal degeneration is part of the disease. There are numerous pathways that lead to joint failure, but the initial step is often joint injury in the setting of a failure of protective mechanisms.

JOINT PROTECTIVE MECHANISMS AND THEIR FAILURE

Joint protectors include joint capsule and ligaments, muscle, sensory afferents, and underlying bone. Joint capsule and ligaments serve as joint protectors by providing a limit to excursion, thereby fixing the range of joint motion.

Synovial fluid reduces friction between articulating cartilage surfaces, thereby serving as a protector against friction-induced cartilage wear. This lubrication function depends on *hyaluronic acid* and on *lubricin*, a mucinous glycoprotein secreted by synovial fibroblasts whose concentration diminishes after joint injury and in the face of synovial inflammation.

The ligaments, along with overlying skin and tendons, contain mechanoreceptor sensory nerves. These mechanoreceptors fire at different frequencies throughout a joint's range of motion, providing feedback by way of the spinal cord to muscles and tendons. As a consequence, these muscles and tendons assume the right tension at appropriate points in joint excursion to act as optimal joint protectors, anticipating joint loading.

Muscles and tendons that bridge the joint are key joint protectors. Focal stress across the joint is minimized by muscle contraction that decelerates the joint before impact and assures that when joint impact arrives, it is distributed broadly across the joint surface.

Failure of these joint protectors increases the risk of joint injury and OA. For example, in animals, OA develops rapidly when a sensory nerve to the joint is sectioned and joint injury induced. Similarly, in humans, Charcot's arthropathy, a severe and rapidly progressive OA, develops when minor joint injury occurs in the presence of posterior column peripheral neuropathy. Another example of joint protector failure is rupture of ligaments, a well-known cause of the early development of OA.

■ CARTILAGE AND ITS ROLE IN JOINT FAILURE

A thin rim of tissue at the ends of two opposing bones, cartilage is lubricated by synovial fluid to provide an almost frictionless surface across which these two bones move. The compressible stiffness of cartilage compared to bone provides the joint with impact-absorbing capacity.

The earliest changes of OA may occur in cartilage, and abnormalities there can accelerate disease development. The two major macromolecules in cartilage are type 2 collagen, which provides cartilage its tensile strength, and aggrecan, a proteoglycan macromolecule linked with hyaluronic acid, which consists of highly negatively charged glycosaminoglycans. In normal cartilage, type 2 collagen is woven tightly, constraining the aggrecan molecules in the interstices between collagen strands, forcing these highly negatively charged molecules into close proximity. The aggrecan molecule, through electrostatic repulsion of its negative charges, gives cartilage its compressive stiffness. Chondrocytes, the cells within this avascular tissue, synthesize all elements of the matrix and produce enzymes that break it down **(Fig. 383-3)**. Cartilage matrix synthesis and catabolism are in a dynamic equilibrium influenced by the cytokine and growth factor environment. Mechanical and osmotic stress on chondrocytes induces these cells to alter gene expression and increase production of inflammatory cytokines and matrix-degrading enzymes. While chondrocytes synthesize numerous enzymes, matrix metalloproteinases (MMPs; especially collagenases and ADAMTS-5) are critical enzymes in the breakdown of cartilage matrix.

Local inflammation accelerates the development and progression of osteoarthritis and increases the likelihood that an osteoarthritic joint will be painful. Some of this inflammation may be induced by mechanical stimuli, so called mechanoinflammation. The synovium, cartilage, and bone all influence disease development through cytokines, chemokines, and even complement activation (Fig. 383-3). Matrix fragments released from cartilage stimulate synovium, which releases inflammatory cytokines, and they, in turn, induce chondrocytes to synthesize other proinflammatory molecules. Ultimately, the combination of effects on chondrocytes triggers matrix degradation. Growth factors are also part of this complex network, with bone morphogenetic protein 2 (BMP-2) and transforming growth factor β (TGF-β) playing prominent roles in stimulating the development of osteophytes. Triggered by local vascular endothelial growth factor (VEGF) synthesis, blood vessels invade cartilage and, with them, come nerves that may bring nociceptive innervation.

With aging, articular chondrocytes develop a senescence-associated secretory phenotype, exhibiting a decline in synthetic capacity and producing proinflammatory mediators and matrix-degrading enzymes. These chondrocytes are unable to maintain tissue homeostasis (such as after insults of a mechanical or inflammatory nature). Thus, with age, cartilage is easily damaged by minor sometimes unnoticed injuries, including those that are part of daily activities.

OA cartilage is characterized by gradual depletion of aggrecan, an unfurling of the tightly woven collagen matrix, and loss of type 2 collagen. With these changes comes increasing vulnerability of cartilage, which loses its compressive stiffness.

RISK FACTORS

Risk factors for OA can be understood in terms of their effect either on joint vulnerability or joint loading. On the one hand, a vulnerable joint whose protectors are dysfunctional can develop OA with minimal levels of loading, perhaps even levels encountered during everyday activities. On the other hand, in a young joint with competent protectors, a

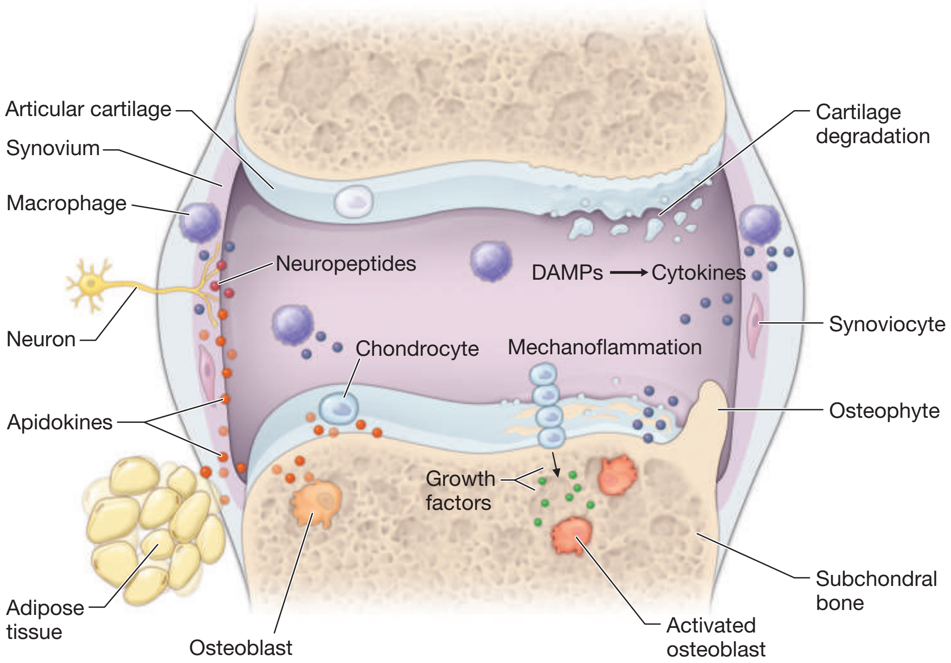

FIGURE 383-3 Selected factors involved in the osteoarthritic process including chondrocytes, bone, and synovium. Synovitis causes release of cytokines, alarmins, damage-associated molecular pattern (DAMP) molecules, and complement, which activate chondrocytes through cell-surface receptors. Chondrocytes produce matrix molecules (collagen type 2, aggrecan) and the enzymes responsible for the degradation of the matrix (e.g., ADAMTS-5 and matrix metalloproteinases [MMPs]). Bone invasion occurs through the calcified cartilage, triggered by vascular endothelial growth factor (VEGF) and other growth factors. *(Reproduced with permission from De Roover A et al: Fundamentals of osteoarthritis: Inflammatory mediators in osteoarthritis. Osteoarthritis Cart 31:1303, 2023.)*

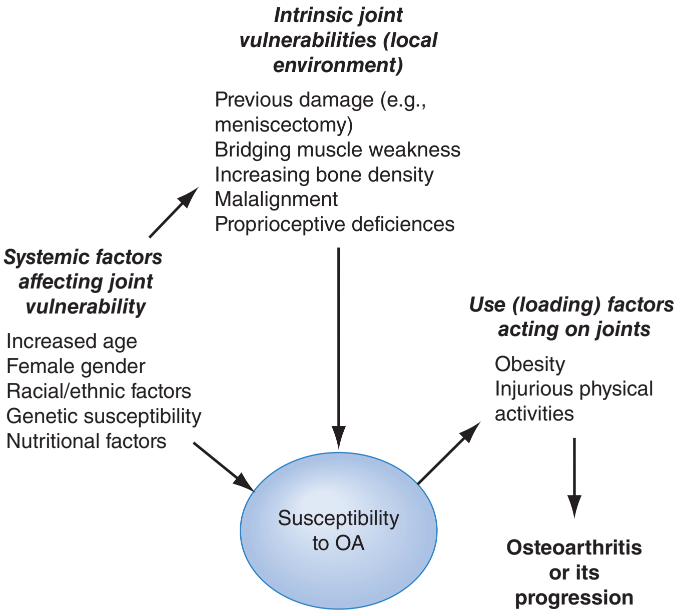
Intrinsic joint vulnerabilities (local environment)
Previous damage (e.g., meniscectomy)
Bridging muscle weakness
Increasing bone density
Malalignment
Proprioceptive deficiences
Systemic factors affecting joint vulnerability
Increased age
Female gender
Racial/ethnic factors
Genetic susceptibility
Nutritional factors
Use (loading) factors acting on joints
Obesity
Injurious physical activities
Susceptibility to OA
Osteoarthritis or its progression

FIGURE 383-4 Risk factors for osteoarthritis (OA) either contribute to the susceptibility of the joint (systemic factors or factors in the local joint environment) or increase risk by the load they put on the joint. Usually, a combination of loading and susceptibility factors is required to cause disease or its progression.

major acute injury or long-term overloading is necessary to precipitate disease **(Fig. 383-4)**.

SYSTEMIC RISK FACTORS THAT AFFECT JOINT VULNERABILITY

Age is the most potent risk factor for OA. Radiographic evidence of OA is rare in individuals aged <40; however, in some joints, such as the hands, OA occurs in >50% of persons aged >70. Aging increases joint vulnerability through several mechanisms. Whereas dynamic loading of joints stimulates matrix synthesis by chondrocytes in young cartilage, aged cartilage is less responsive to these stimuli. As a consequence of this failure to synthesize matrix with loading, cartilage thins with age, and thinner cartilage experiences higher shear stress and is at greater risk of damage. Also, joint protectors fail more often with age. Muscles that bridge the joint become weaker with age and respond less quickly to oncoming impulses. Sensory nerve input slows with age, retarding the feedback loop of mechanoreceptors to muscles and tendons related to their tension and position. Ligaments stretch with age, making them less able to absorb impulses. These factors work in concert to increase the vulnerability of older joints to OA.

Older women are at high risk of OA in all joints, a risk that increases as women reach their sixth decade. Although hormone loss with menopause may contribute to this risk, there is little understanding of the unique vulnerability of older women versus men.

HERITABILITY AND GENETICS AND THEIR RELATION TO JOINT VULNERABILITY

OA is a heritable disease, but its heritability is mostly joint specific. Nearly 60–65% of OA in hips or hands may be attributed to shared genetics within a family. However, heritability of knee OA is, at most, 30%, with some studies suggesting no heritability at all. Knees are susceptible to nongenetic factors like activities that affect joint loading and risk of injury. On the other hand, many people with OA develop "generalized OA" or multisite OA, but the involvement of multiple joints with OA is usually a consequence of aging rather than genetics.

The best replicated genetic variant known to increase OA risk lies in the locus for growth differentiation factor 5 (GDF5). It affects epigenetic regulation of GDF5 activity, resulting in reduction in GDF5 expression. GDF5 probably affects joint shape, a mechanism by which genes increase disease risk. Minor abnormalities in joint shape can make a joint vulnerable to damage if focal stresses across the joint increase.

RISK FACTORS IN THE JOINT ENVIRONMENT

Some risk factors increase vulnerability of the joint through local effects on the joint environment. With changes in joint anatomy, for example, load across the joint is no longer distributed evenly across the joint surface, but rather shows an increase in focal stress. In the hip, three uncommon developmental abnormalities occurring in utero or in childhood—congenital dysplasia, Legg-Perthes disease, and slipped capital femoral epiphysis—leave a child with distortions of hip joint anatomy that often lead to OA later in life. Girls are predominantly affected by acetabular dysplasia, a mild form of congenital dysplasia, whereas the other abnormalities more often affect boys. Depending on the severity of the anatomic abnormalities, hip OA occurs either in young adulthood (severe abnormalities) or middle age (mild abnormalities). Femoroacetabular impingement can develop during adolescence. It is a clinical syndrome in which an outgrowth of bone at the femur's head/neck junction thought to develop during closure of the growth plate results in abnormal contact between the femur and acetabulum, especially during hip flexion and rotation. This leads to cartilage and labral damage, to hip pain, and ultimately in later life, to an increased risk of hip OA.

Major injuries to a joint also can produce anatomic abnormalities that leave the joint susceptible to OA. For example, a fracture through the joint surface often causes OA in joints in which the disease is otherwise rare such as the ankle and the wrist. Avascular necrosis can lead to collapse of dead bone at the articular surface, producing anatomic irregularities and subsequent OA.

Tears of ligamentous and fibrocartilaginous structures that protect the joints, such as the anterior cruciate ligament or meniscus in the knee and the labrum in the hip, can lead to premature OA. Meniscal tears increase with age and, when chronic, are often asymptomatic but lead to adjacent cartilage damage and accelerated OA. Even recalled injuries in which the affected person never received a diagnosis may increase risk of OA. For example, in the Framingham Study subjects, men with a history of major knee injury, but no surgery, had a 3.5-fold increased risk for subsequent knee OA.

Another source of anatomic abnormality is malalignment across the joint **(Fig. 383-5)**, a factor best studied in the knee. Varus (bowlegged) knees with OA are at exceedingly high risk of cartilage loss in the medial or inner compartment of the knee, whereas valgus (knock-kneed) malalignment predisposes to rapid cartilage loss in the lateral compartment. Malalignment causes this effect by increasing stress on a focal area of cartilage, which then breaks down; it also causes damage to bone underlying the cartilage, producing bone marrow lesions seen on magnetic resonance imaging (MRI). Malalignment in the knee often produces such a substantial increase in focal stress within the knee (as evidenced by its destructive effects on subchondral bone) that severely malaligned knees may be destined to progress regardless of the status of other risk factors.

Weakness in the quadriceps muscles bridging the knee increases the risk of the development of painful OA in the knee, especially in women.

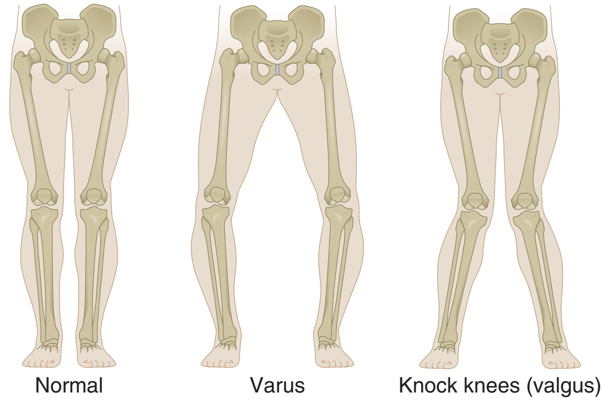

FIGURE 383-5 The two types of limb malalignment in the frontal plane: varus, in which the stress is placed across the medial compartment of the knee joint, and valgus, which places excess stress across the lateral compartment of the knee.

High bone density also increases OA risk, especially risk of a subtype of OA characterized by large osteophytes.

LOADING FACTORS

Obesity Three to six times body weight is transmitted across the knee during single-leg stance. Any increase in weight may be multiplied by this factor to reveal the excess force across the knee in overweight persons during walking. Obesity is a potent risk factor for the development of knee OA and, less so, for hip OA. It is a stronger risk factor for disease in women than in men, and for women, the relationship of weight to the risk of disease is linear, so that with each pound increase in weight, there is a commensurate increase in risk. Not only is obesity a risk factor for OA in weight-bearing joints, but obese persons have more pain from the disease. Weight loss greater than 10% of body weight reduces cartilage loss and, in most affected persons, alleviates pain.

Obesity's effect on the development and progression of disease, especially in knees, is mediated mostly through the increased loading in weight-bearing joints that occurs in overweight persons. However, high levels of leptin, an adipokine, are associated with increasing pain severity in OA regardless of the joint affected.

Repeated Use of Joint and Exercise There are two categories of repetitive joint use: occupational use and leisure time physical activities. Workers who, over many years, perform repetitive tasks as part of their occupations are at high risk of developing OA in joints they use repeatedly. Workers whose jobs require regular knee bending or lifting or carrying heavy loads have a high rate of knee OA. One reason why workers may get disease is that during long days at work, their muscles may gradually become exhausted, no longer serving as effective joint protectors.

It is widely recommended for people to adopt an exercise-filled lifestyle, and long-term studies of exercise suggest no consistent association of exercise with OA risk in most persons. However, persons who already have injured joints may put themselves at greater risk by engaging in certain types of exercise. For example, persons who have already sustained major knee injuries are at increased risk of progressive knee OA as a consequence of running. In addition, compared to nonrunners, elite runners (professional runners and those on Olympic teams) have high risks of both knee and hip OA. Lastly, although recreational runners are not at increased risk of knee OA, studies suggest that they have a modest increased risk of disease in the hip.

PATHOLOGY

The pathology of OA provides evidence of the involvement of many joint structures in disease. Cartilage initially shows surface fibrillation and irregularity. As disease progresses, focal erosions develop there, and these eventually extend to the subjacent bone. With further progression, cartilage erosion down to bone expands to involve a larger proportion of the joint surface, even though OA remains a focal disease with nonuniform loss of cartilage.

After an injury to cartilage, chondrocytes undergo mitosis and clustering. Although the metabolic activity of these chondrocyte clusters is high, the net effect of this activity is to promote proteoglycan depletion in the matrix surrounding the chondrocytes. This is because the catabolic activity is greater than the synthetic activity. As disease develops, collagen matrix becomes damaged, the negative charges of proteoglycans get exposed, and cartilage swells from ionic attraction to water molecules. Because in damaged cartilage proteoglycans are no longer forced into close proximity, cartilage does not bounce back after loading as it did when healthy, and cartilage becomes vulnerable to further injury.

With loss of cartilage comes alteration in subchondral bone. Stimulated by growth factors and cytokines, osteoclasts and osteoblasts in the bony plate just underneath cartilage become activated. Bone formation produces a thickening of the subchondral plate that occurs even before cartilage ulcerates. Trauma to bone during joint loading may be the primary factor driving this bone response, with healing from injury (including microcracks) inducing remodeling. Small areas of osteonecrosis usually exist in joints with advanced disease. Bone death may also be caused by bone trauma with shearing of microvasculature, leading to a cutoff of vascular supply to some bone areas.

At the margin of the joint, near areas of cartilage loss, osteophytes form. These start as outgrowths of new cartilage, and with neurovascular invasion from the bone, this cartilage ossifies. Osteophytes are an important radiographic hallmark of OA.

The synovium produces lubricating fluids that minimize shear stress during motion. In healthy joints, the synovium consists of a single discontinuous layer filled with fat and containing two types of cells, macrophages and fibroblasts, but in OA, it can sometimes become edematous and inflamed. There is a migration of macrophages from the periphery into the tissue, and cells lining the synovium proliferate. Inflammatory cytokines and alarmins secreted by the synovium activate chondrocytes to produce enzymes that accelerate destruction of matrix.

Additional pathologic changes occur in the capsule, which stretches, becomes edematous, and can become fibrotic.

The pathology of OA is not identical across joints. In hand joints with severe OA, for example, there are often cartilage erosions in the center of the joint probably produced by bony pressure from the opposite side of the joint.

Basic calcium phosphate and calcium pyrophosphate dihydrate crystals are present microscopically in most joints with end-stage OA. Their role in osteoarthritic cartilage is unclear, but their release from cartilage into the joint space and joint fluid likely triggers synovial inflammation, which can, in turn, produce release of cytokines and trigger nociceptive stimulation.

SOURCES OF PAIN

Because healthy cartilage is aneural, cartilage loss alone is not accompanied by much pain. Thus, pain in OA likely arises from structures outside the cartilage. Innervated structures in the joint include the synovium, ligaments, joint capsule, muscles, and subchondral bone. Most of these are not visualized by x-ray, and the severity of x-ray changes in OA correlates poorly with pain severity. However, in later stages of OA, loss of cartilage integrity accompanied by neurovascular invasion may contribute more to pain.

Based on MRI studies in osteoarthritic knees comparing those with and without pain and on studies mapping tenderness in unanesthetized joints, likely sources of pain include synovial inflammation, joint effusions, and bone marrow edema. Modest synovitis develops in many but not all osteoarthritic joints. The presence of synovitis on MRI is correlated with the presence and severity of knee pain, and potentially with pain sensitization. Capsular stretching from fluid in the joint stimulates nociceptive fibers there, inducing pain. Increased focal loading as part of the disease not only damages cartilage but probably also injures the underlying bone. As a consequence, bone marrow edema appears on the MRI; histologically, this edema signals the presence of microcracks and scar, which are the consequences of trauma. These lesions may stimulate bone nociceptive fibers. Pain may arise from outside the joint also, including bursae near the joints. Common sources of pain near the knee are anserine bursitis and iliotibial band syndrome.

Much of the pain experienced in OA occurs when nociceptors in the joint are stimulated during weight-bearing activities. However, the pain may eventually become more constant and present at rest, and this suggests other mechanisms contribute to the pain experience. The pathologic changes of OA may lead to alterations in nervous system signaling **(Chap. 18)**. Specifically, peripheral nociceptors can become more responsive to sensory input, known as peripheral sensitization, and there can also be an increase in facilitated central ascending nociceptive signaling, known as central sensitization. Individuals with OA may also have insufficient descending inhibitory modulation. Some individuals may be genetically predisposed to becoming sensitized; however, regardless of the etiology, these nervous system alterations are associated with more severe pain, contribute to the presence of allodynia and hyperalgesia in patients with OA, and may predispose to development of chronic pain.

CLINICAL FEATURES

Joint pain from OA is primarily activity-related in the early stages of the disease. Pain comes on either during or just after joint use and then gradually resolves. Examples include knee or hip pain with going up or

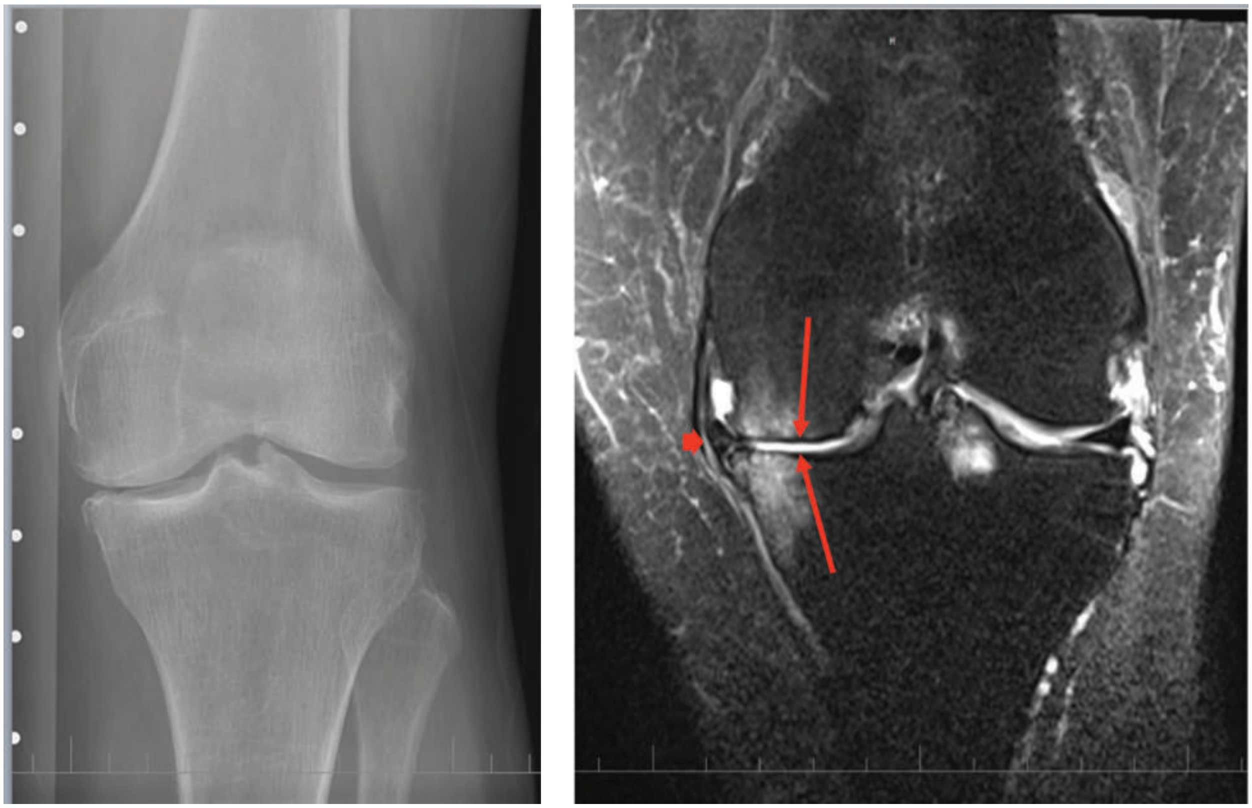

FIGURE 383-6 X-ray and magnetic resonance imaging (MRI) of knee with medial osteoarthritis. X-ray shows osteophytes at the medial and lateral tibia and femur and joint space narrowing of the medial tibiofemoral joint. Coronal intermediate-weighted fat-suppressed MRI confirms the presence of medial and lateral osteophytes and the medial tibiofemoral joint space narrowing. There is diffuse denuded area with no cartilage remaining at the weight-bearing medial tibiofemoral joint (*arrows*). There is also a severe medial meniscus extrusion (*arrowhead*). Bone marrow lesions, which provide evidence of bone injury, are present at medial tibia, medial femur, and intraspinous tibial region. Cartilage focal defects are also seen at the lateral weight-bearing femur and tibia.

down stairs, pain in weight-bearing joints when walking, and, for hand OA, pain when cooking. Early in disease, pain is episodic, triggered often by overactive use of a diseased joint, such as a person with knee OA taking a long run and noticing a few days of pain thereafter. As disease progresses, the pain becomes continuous and even begins to be bothersome at night. Stiffness of the affected joint may be prominent, but morning stiffness is usually brief (<30 min).

In knees, buckling may occur, in part, from weakness of muscles crossing the joint. Mechanical symptoms, such as buckling, catching, or locking, could also signify internal derangement, like an anterior cruciate ligament or meniscal tear; however, these symptoms, which are common in persons with knee OA, need to be further evaluated only if they develop after an acute knee injury. In the knee, pain with activities requiring knee flexion, such as stair climbing and arising from a chair, often emanates from the patellofemoral compartment of the knee, which does not actively articulate until the knee is bent ~35°.

OA is the most common cause of chronic knee pain in persons aged >45, but the differential diagnosis is long. Inflammatory arthritis is likely if there is prolonged morning stiffness, and many other joints are affected. Bursitis occurs commonly around knees and hips. A physical examination should focus on whether tenderness is over the joint line (at the junction of the two bones around which the joint is articulating) or outside of it. Anserine bursitis, medial and distal to the knee, is an extremely common cause of chronic knee pain that may respond to a glucocorticoid injection. Prominent nocturnal pain in the absence of end-stage OA merits a distinct workup. For hip pain, OA can be detected by loss of internal rotation on passive movement, and pain isolated to an area lateral to the hip joint usually reflects the presence of trochanteric bursitis.

No blood tests are routinely indicated for workup of patients with OA unless symptoms and signs suggest inflammatory arthritis. Examination of the synovial fluid is often more helpful diagnostically than an x-ray. If the synovial fluid white count is >1000/μL, inflammatory arthritis or gout or pseudogout is likely, the latter two being also identified by the presence of crystals.

Radiographs are not indicated in the workup of OA. They should be ordered only when joint pain and physical findings are not typical of OA or if pain persists after inauguration of treatment effective for OA. In OA, imaging findings **(Fig. 383-6)** correlate poorly with the presence and severity of pain. Further, in both knees and hips, radiographs may be normal in early disease as they are insensitive to cartilage loss and other early findings.

Although MRI may reveal the extent of pathology in an osteoarthritic joint, it is not indicated as part of the diagnostic workup. Findings such as meniscal tears and cartilage and bone lesions occur not only in most patients with OA in the knee but also in most older persons without joint pain. MRI findings rarely warrant a change in therapy.

TREATMENT

Osteoarthritis

The goals of the treatment of OA are to alleviate pain and minimize loss of physical function. To the extent that pain and loss of function are consequences of inflammation, of weakness across the joint, and of laxity and instability, the treatment of OA involves addressing each of these impairments. Comprehensive therapy consists of a multimodality approach including physical modalities and pharmacologic elements.

Foundational to OA management is patient education and self-management strategies. Patients with mild and intermittent symptoms may need only symptomatic management and/or treatments aimed at weight loss, physical activity, exercise, and self-management strategies. Patients with ongoing, disabling pain are likely to need both physical modalities and pharmacotherapy.

Treatments for knee OA have been more completely evaluated than those for hip and hand OA or for disease in other joints. Thus, although the principles of treatment are identical for OA in all joints, we shall focus below on the treatment of knee OA, noting specific recommendations for disease in other joints, especially when they differ from those for the knee.

PHYSICAL MANAGEMENT MODALITIES

Because OA is a mechanically driven disease, the mainstay of treatment involves altering loading across the painful joint and improving the function of joint protectors, so they can better distribute load across the joint. Ways of lessening focal load across the joint include:

1. Avoiding painful activities as these are usually activities that overload the joint
2. Improving the strength and conditioning of muscles that bridge the joint to optimize their function
3. Unloading the joint, either by redistributing load within the joint with a brace or a splint or by unloading the joint during weight bearing with a cane or a crutch

The simplest treatment for many patients is to avoid activities that precipitate pain. For example, for the middle-aged patient whose long-distance running brings on symptoms of knee OA, a less demanding form of weight-bearing activity may alleviate all symptoms. For an older person whose daily walks up and down hills bring on knee pain, routing these away from hills might eliminate symptoms.

Weight loss is a central strategy for those who are overweight or obese, particularly for knee OA. Each pound of weight loss has a multiplier effect, unloading both knees and hips and probably relieving pain in those joints. Depending on how much weight loss occurs, bariatric surgery and drugs that cause weight loss such as glucagon-like peptide 1 receptor agonists (GLP1R agonists) often reduce pain and may slow structural disease progression.

In hand joints affected by OA, splinting, by limiting motion, often minimizes pain for patients with involvement especially in the base of the thumb. Weight-bearing joints such as knees and hips can be unloaded by using a cane in the hand opposite the affected joint for partial weight bearing. A physical therapist can help teach the patient how to use the cane optimally, including ensuring that its height is optimal for unloading. Crutches or walkers can serve a similar beneficial function.

Exercise Osteoarthritic pain in knees or hips during weight bearing results in lack of activity and poor mobility, and because OA is so common, the inactivity that results increases the risk of cardiovascular disease and obesity. Aerobic capacity is poor in most elders with symptomatic knee OA, worse than others of the same age.

Weakness in muscles that bridge osteoarthritic joints is multifactorial in etiology. First, there is a decline in strength with age. Second, with limited mobility comes disuse muscle atrophy. Third, patients with painful knee or hip OA alter their gait to lessen loading across the affected joint, and this further diminishes muscle use. Fourth, “arthrogenous inhibition” may occur, whereby contraction of muscles bridging the joint is inhibited by a nerve afferent feedback loop emanating in a swollen and stretched joint capsule; this prevents attainment of voluntary maximal strength. Because adequate muscle strength and conditioning are critical to joint protection, weakness in a muscle that bridges a diseased joint makes the joint more susceptible to further damage and pain. The degree of weakness correlates strongly with the severity of joint pain and the degree of physical limitation. One of the cardinal elements of the treatment of OA is to improve the functioning of muscles surrounding the joint.

Trials in knee and hip OA have shown that exercise lessens pain and improves physical function. Most effective exercise regimens consist of aerobic and/or resistance training, the latter of which focuses on strengthening muscles across the joint. Treatment guidelines strongly recommend exercise for knee and hip OA with no hierarchy regarding type of exercise due to lack of sufficient head-to-head data. Exercises are likely to be effective especially if they train muscles for the activities a person performs daily. Activities that increase pain in the joint should be avoided. Range-of-motion exercises, which do not strengthen muscles, and isometric exercises that strengthen muscles, but not through range of motion, are unlikely to be effective by themselves. Low-impact exercises, including water aerobics and water resistance training, are often better tolerated by patients than exercises involving impact loading, such as running or treadmill exercises. Evidence suggests that high-intensity strengthening is no more effective in reducing knee pain than less aggressive strengthening regimens. The exercise regimen needs to be individualized to optimize effectiveness, and patients should be referred to an exercise class or a therapist who can create an individualized regimen. In addition to conventional exercise regimens, tai chi may be effective for knee OA. However, there is no strong evidence that patients with hand OA benefit from therapeutic exercise.

Adherence over the long term is the major challenge to an exercise prescription. In trials involving patients with knee OA who were engaged in exercise treatment, from a third to over half of patients stopped exercising by 6 months. Less than 50% continued regular exercise at 1 year. The strongest predictor of a patient’s continued exercise is a previous personal history of successful exercise. Physicians should reinforce the exercise prescription at each clinic visit, help the patient recognize barriers to ongoing exercise, and identify convenient times for exercise to be done routinely. Mobile health and wearable technologies are increasingly being used to encourage adherence to exercise. The combination of exercise with calorie restriction and weight loss is especially effective in lessening pain.

Correction of Malalignment Malalignment in the frontal plane (varus-valgus) markedly increases the stress across the joint, which can lead to progression of disease and to pain and disability (Fig. 383-5). Correcting varus-valgus malalignment, either surgically or with bracing, may relieve pain in persons whose knees are malaligned. However, correcting malalignment is often challenging. Fitted braces that straighten varus knees by putting valgus stress across the knee can be effective. Unfortunately, many patients are unwilling to wear a realigning knee brace; in addition, in patients with obese legs, braces may slip with usage and lose their realigning effect. Braces are indicated for willing patients who can learn to put them on correctly and on whom they do not slip. Shoes modified with rubber hemispheres on the sole that alter alignment of the proximal knee have shown efficacy in trials especially if worn for brief periods daily over several months.

Pain from the patellofemoral compartment of the knee can be caused by tilting of the patella or patellar malalignment with the patella riding laterally in the femoral trochlear groove. Using a patellar brace to realign the patella, or tape to pull the patella back into the trochlear sulcus or reduce its tilt, can lessen patellofemoral pain. However, patients may find it difficult to apply tape, with skin irritation common, and, like realigning braces, patellar braces may slip.

Although their effect on malalignment is questionable, neoprene sleeves pulled up to cover the knee lessen pain and are easy to use and popular among patients. The explanation for their therapeutic effect on pain is unclear.

In patients with knee OA, acupuncture produces modest pain relief compared to placebo needles and may be an adjunctive treatment, though placebo effect is likely high. In patients with refractory joint pain from OA, radiofrequency ablation of the nerves innervating the joint has been shown to provide prolonged pain relief, although long-term safety is unknown.

PHARMACOTHERAPY

Although approaches involving physical modalities constitute its mainstay, pharmacotherapy serves an important adjunctive role in OA treatment for symptom management. Available drugs are administered using oral, topical, and intraarticular routes. To date, there are no available drugs that alter the disease process itself.

Acetaminophen, Nonsteroidal Anti-Inflammatory Drugs (NSAIDs), and Cyclooxygenase-2 (COX-2) Inhibitors The treatment effect of acetaminophen (paracetamol) in OA is small and not considered clinically meaningful **(Table 383-1)**. However, for a minority of patients, it is adequate to control symptoms, in which case more toxic drugs such as oral NSAIDs can be avoided.

NSAIDs are the most popular drugs to treat osteoarthritic pain. They can be administered either topically or orally. In clinical trials, oral NSAIDs produce ~30% greater improvement in pain than high-dose acetaminophen. Response to NSAIDs varies greatly across patients. Occasional patients treated with NSAIDs experience dramatic pain relief, whereas others experience little improvement. Initially, NSAIDs should be administered topically or taken orally on an "as-needed" basis because side effects are less frequent with low intermittent doses. For those with mild symptoms, topical NSAIDs may be sufficient to reduce pain. If topicals or occasional oral medication use is insufficiently effective, then daily treatment may be indicated, with an anti-inflammatory dose selected (Table 383-1).

Topical NSAIDs absorbed through the skin have plasma concentrations an order of magnitude lower than the same amount of drug administered orally. However, when these drugs are administered topically over a superficial joint (knees, hands, but not hips), the drug can be found in joint tissues such as the synovium and cartilage. Generally, topical NSAIDs are slightly less efficacious than oral agents, but have far fewer gastrointestinal (GI) and systemic side effects. Topical NSAIDs often cause local skin irritation where the medication is applied, inducing redness, burning, or itching (see Table 383-1).

TABLE 383-1 Pharmacologic Treatment for Osteoarthritis

TREATMENT	MAXIMAL DOSAGE	COMMENTS
Oral NSAIDs and COX-2 inhibitors		Take with food. Increased risk of myocardial infarction and stroke for some NSAIDs. High rates of gastrointestinal side effects, including ulcers and bleeding. Patients at high risk for gastrointestinal side effects should also take either a proton pump inhibitor or misoprostol.[a] There is an increase in gastrointestinal side effects or bleeding when taken with acetylsalicylic acid. Can also cause edema and renal insufficiency.
Naproxen	375–500 mg bid	
Salsalate	1500 mg bid	
Ibuprofen	600–800 mg qid/tid	
Celecoxib	100–200 mg qd	
Acetaminophen	Up to 2 g/d	Of limited efficacy; conditionally recommended.
Opiates		Common side effects include dizziness, sedation, nausea, vomiting, dry mouth, constipation, urinary retention, and pruritus. Addiction risk. Less efficacious than oral NSAIDs.
Topical NSAIDs		Rub onto hands/knees. Few systemic side effects.
Diclofenac Na 1% gel	4 g qid	Skin irritation common.
Capsaicin	0.025–0.075% cream tid/qid	Can irritate mucous membranes.
Intraarticular injections		
Steroids	Various	
Hyaluronans	Varies from 3 to 5 weekly injections	Mild to moderate pain at injection site. Controversy exists regarding efficacy.

[a] Patients at high risk include those with previous gastrointestinal events, persons ≥60 years, and persons taking glucocorticoids. Trials have shown the efficacy of proton pump inhibitors and misoprostol in the prevention of ulcers and bleeding. Misoprostol is associated with a high rate of diarrhea and cramping; therefore, proton pump inhibitors are more widely used to reduce NSAID-related gastrointestinal symptoms.

Abbreviations: COX-2, cyclooxygenase-2; NSAIDs, nonsteroidal anti-inflammatory drugs.

Source: From DT Felson: Osteoarthritis of the Knee. N Engl J Med 354:841, 2006. Copyright © 2006 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

Oral NSAIDs are often effective in reducing moderate or severe joint pain, but they have substantial and frequent side effects, the most common of which is upper GI toxicity, including dyspepsia, nausea, bloating, GI bleeding, and ulcer disease. Certain oral agents including celecoxib and nabumetone are safer to the stomach than others. Major NSAID-related GI side effects can occur in patients who do not complain of upper GI symptoms. Patients should be reminded to take low-dose aspirin and ibuprofen or naproxen at different times during the day to avoid a drug interaction.

Because of the increased rates of cardiovascular events associated with conventional NSAIDs such as diclofenac, many of these drugs are not appropriate long-term treatment choices for older persons with OA, especially those at high risk of heart disease or stroke. The American Heart Association has identified COX-2 inhibitors as putting patients at high risk, although low doses of celecoxib (≤200 mg/d) are not associated with an elevation of risk. Naproxen is a safe NSAID from a cardiovascular perspective, but it does have GI toxicity.

There are other common side effects of NSAIDs, including the tendency to develop edema because of prostaglandin inhibition of afferent blood supply to glomeruli in the kidneys and, for similar reasons, a predilection toward reversible renal insufficiency. Blood pressure may increase modestly in some NSAID-treated patients. Oral NSAIDs should not be used in patients with stage 3, 4, or 5 renal disease.

Intraarticular Injections: Glucocorticoids, Hyaluronic Acid, and Other Products Because synovial inflammation is likely to be a major cause of pain in patients with OA, local anti-inflammatory treatments administered intraarticularly may be effective in ameliorating pain for up to 3 months. Glucocorticoid injections provide such efficacy, but response is variable, with some patients having little relief of pain, whereas most experience pain relief lasting up to several months. Synovitis, a major cause of joint pain in OA, may abate after an injection, and this correlates with the reduction in knee pain severity. Glucocorticoid injections are useful to get patients over acute flares of pain. Repeated injections may cause minor amounts of cartilage loss, but these do not appear to increase risk of disease progression, worsening pain, functional limitations, or the need for surgery.

Hyaluronic acid injections can be given for treatment of symptoms in knee and hip OA, but most evidence suggests they have little efficacy versus placebo (Table 383-1).

Few rigorous studies of intraarticular stem cell therapy and platelet-rich plasma (PRP) have been conducted to date. Further, the composition of the injected biologic materials reflecting stem cells or PRP has not been standardized, making it challenging to compare across studies due to variability in formulation.

Other Classes of Drugs and Nutraceuticals Opioids have only modest short-term efficacy in treating pain in hip or knee OA without evidence of long-term benefit and, given concerns about opioid dependency, should be avoided. If NSAIDs are ineffective, one option is the use of duloxetine, which is U.S. Food and Drug Administration approved for OA. Duloxetine may be particularly efficacious when knee pain is part of a syndrome of widespread pain.

Biologic agents and disease-modifying agents used for rheumatoid arthritis have been tested for OA, but trials have mostly been negative. Notable exceptions are two randomized trials for hand OA, one of 6 weeks of prednisolone and the other of 6 months of methotrexate. Both trials demonstrated reductions in hand pain compared with placebo. In a large randomized trial, GLP1R agonists reduced knee pain in obese patients.

Recent guidelines recommend against the use of glucosamine or chondroitin for OA. Large publicly supported trials have failed to show that, compared with placebo, these compounds relieve pain in persons with disease. Curcumin has anti-inflammatory effects and may alleviate pain in OA.

Optimal pharmacologic therapy for OA is often achieved by trial and error, with each patient having idiosyncratic responses to specific treatments. Placebo (or contextual) effects may account for 50% of more of treatment effects in OA, and certain modes of treatment delivery including intraarticular injections have greater contextual effects than others such as pills. When medical therapies have failed and the patient has an unacceptable reduction in their quality of life and ongoing pain and disability, then at least for knee and hip OA, total joint arthroplasty is indicated.

SURGERY

Based on data from randomized trials, the efficacy of arthroscopic debridement and lavage in persons with OA is no greater than that of sham surgery for relief of pain or disability. Further, if a meniscal tear is present, as is often the case in persons with knee OA, trials have shown that arthroscopic meniscectomies do not relieve knee pain or improve function long-term, nor do they reduce catching or locking symptoms.

On the other hand, for patients with knee OA isolated to the medial compartment, operations to realign the knee to lessen medial loading can relieve pain. These include a high tibial osteotomy, in which the tibia is broken just below the tibial plateau and realigned to load the lateral, nondiseased compartment, or a unicompartmental replacement with realignment. Each surgery may provide the patient with years of pain relief before a total knee replacement is required.

Ultimately, when the patient with knee or hip OA has failed nonsurgical treatments with limitations of pain or function that compromise their quality of life, patients with reasonable expectations and readiness for surgery should be referred for total knee or hip arthroplasty. These are highly efficacious operations that relieve pain and improve function in the vast majority of patients, although pain elimination occurs in almost all patients getting a hip replacement but only ~80% of those undergoing knee replacement. Currently, failure rates from loosening or infection for both are ~1% per year, with higher rates in obese patients. The chance of surgical success is greater in centers where at least 25 such operations are performed yearly or with surgeons who perform multiple operations annually. The timing of knee or hip replacement is critical. If the patient suffers for many years until their functional status has declined substantially, with considerable muscle weakness, postoperative functional status may not improve to a level achieved by others who underwent operation earlier in their disease course.

Cartilage Regeneration Chondrocyte transplantation has not been found to be efficacious in OA, perhaps because OA includes pathology of joint mechanics, which is not corrected by chondrocyte transplants. Similarly, abrasion arthroplasty (chondroplasty) has not been well studied for efficacy in OA, but it produces fibrocartilage in place of damaged hyaline cartilage. Both surgical attempts to regenerate and reconstitute articular cartilage are more likely to be efficacious early in disease when joint malalignment and many of the other noncartilage abnormalities that characterize OA have not yet developed.

FURTHER READING

De Roover A et al: Fundamentals of osteoarthritis: Inflammatory mediators in osteoarthritis. Osteoarthritis Cart 31:1303, 2023.

Katz JN et al: Diagnosis and treatment of hip and knee osteoarthritis: A review. JAMA 325:568, 2021.

Kolasinski SL et al: 2019 American College of Rheumatology/Arthritis Foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Rheumatol 72:220, 2020.

384 Gout and Other Crystal-Associated Arthropathies

Hyon K. Choi

The use of polarizing light microscopy during synovial fluid analysis in 1961 by McCarty and Hollander and the subsequent application of other crystallographic techniques, such as electron microscopy, energy-dispersive elemental analysis, and x-ray diffraction, have allowed investigators to identify the pathogenic roles of different microcrystals, including monosodium urate (MSU), calcium pyrophosphate (CPP), calcium apatite (apatite), and calcium oxalate (CaOx), in inducing acute or chronic arthritis or periarthritis **(Table 384-1)**. The clinical manifestations that result from these crystals have substantial similarities but also have notable differences. Given their therapeutic implications, the need to perform synovial fluid analysis to distinguish the type of crystal involved should be emphasized. Polarized light microscopy alone can identify most typical crystals, except for apatite. Aspiration and analysis of effusions are also important to assess the possibility of infection. Crystal shedding from inert deposits triggered by certain factors is considered a key process behind episodic manifestation of acute inflammation (gout or pseudogout) involving activation of inflammasome and potent proinflammatory cytokines such as interleukin (IL) 1β. Furthermore, physical, inflammatory, and catalytic effects (involving metalloproteinase, collagenase, or prostaglandin E_2) of crystal deposits on the cartilage and other joint structures can lead to chronic erosive or destructive changes in the articular structures.

GOUT

PATHOGENESIS

Gout is a hyperuricemic metabolic condition, typically manifested by episodic inflammatory arthritis with disabling pain, typically among middle-aged to elderly men and postmenopausal women. It stems from an increased body pool of urate due to chronic hyperuricemia, leading to supersaturation and crystal formation of MSU within the joints and connective tissue and inflammatory reactions to the crystals **(Fig. 384-1)**. Although chronic hyperuricemia is a prerequisite for the development of gout, other factors influence subsequent MSU crystallization and inflammatory reactions (Fig. 384-1); a minority of hyperuricemic individuals develop gout during their lifetime. Considered often as a part of the insulin resistance syndrome, hyperuricemia and gout are associated with multiple cardiometabolic-kidney comorbidities, including obesity, hypertension, type 2 diabetes, myocardial infarction, stroke, chronic kidney disease (CKD), and urate nephrolithiasis; modifiable

TABLE 384-1 Musculoskeletal Manifestations of Crystal-Induced Arthritis

- Acute arthritis (episodic)
  - Mono-, oligo-, or polyarthritis
- Periarticular inflammation
  - Bursitis
  - Tendinitis
  - Enthesitis
- Tophaceous deposits
- Chronic arthropathy
  - Destructive arthropathies
  - Chronic inflammatory arthritis
  - Peculiar type of osteoarthritis
- Spinal arthritis
- Carpal tunnel syndrome

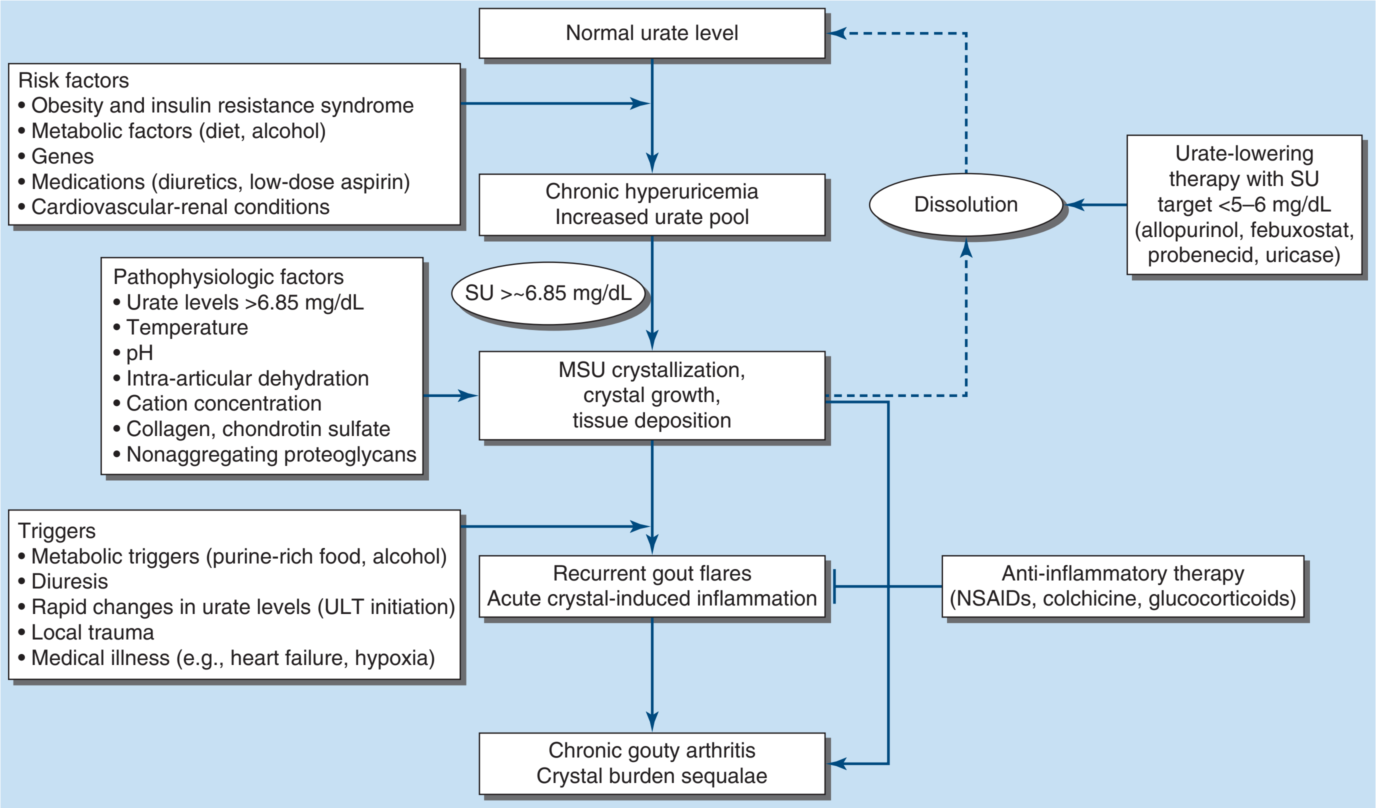
Normal urate level
Risk factors
• Obesity and insulin resistance syndrome
• Metabolic factors (diet, alcohol)
• Genes
• Medications (diuretics, low-dose aspirin)
• Cardiovascular-renal conditions
Chronic hyperuricemia
Increased urate pool
Dissolution
Urate-lowering therapy with SU target <5–6 mg/dL (allopurinol, febuxostat, probenecid, uricase)
Pathophysiologic factors
• Urate levels >6.85 mg/dL
• Temperature
• pH
• Intra-articular dehydration
• Cation concentration
• Collagen, chondrotin sulfate
• Nonaggregating proteoglycans
SU >~6.85 mg/dL
MSU crystallization, crystal growth, tissue deposition
Triggers
• Metabolic triggers (purine-rich food, alcohol)
• Diuresis
• Rapid changes in urate levels (ULT initiation)
• Local trauma
• Medical illness (e.g., heart failure, hypoxia)
Recurrent gout flares
Acute crystal-induced inflammation
Anti-inflammatory therapy
(NSAIDs, colchicine, glucocorticoids)
Chronic gouty arthritis
Crystal burden sequalae

FIGURE 384-1 Pathogenesis of gout and therapeutic targets. Metabolic and some genetic factors contribute to the development of chronic hyperuricemia. Gout stems from an increased body pool of urate due to chronic hyperuricemia, leading to supersaturation and crystal formation and deposition of monosodium urate (MSU; tophi) within the joints and connective tissue. Synovial tophi are usually walled off, but certain triggers may loosen them from the organic matrix, leading to crystal shedding and interaction with synovial cell lining and residential inflammatory cells, initiating an acute gout flare. In some individuals, growing MSU crystal deposition leads to chronic gouty arthritis with subcutaneous tophi. Anti-inflammatory therapy targets the downstream process of crystal-induced inflammation, whereas ultimate control of gout requires correction of the underlying central cause, chronic hyperuricemia with increased urate pool. NSAIDs, nonsteroidal anti-inflammatory drugs; SU, sodium urate; ULT, urate-lowering therapy.

risk factors include obesity, Western diet, alcohol, sedentary lifestyle, and diuretics (Fig. 384-1).

Humans are the only mammals known to develop gout spontaneously as they often develop hyperuricemia with their evolutionary species-wide loss of uricase, which converts urate into the highly water-soluble compound allantoin. At physiologic pH, uric acid exists largely as urate, the ionized form, given its weak acidic property (pKa, 5.8).

The systemic urate concentration reflects a balance between dietary intake, endogenous synthesis, and excretion. Hyperuricemia arises predominantly from impaired urate excretion (90%) or, less commonly, from overproduction (10%), with many cases involving both mechanisms. The pathways leading to urate overproduction, implicated in hyperuricemia and gout, are delineated in **Fig. 384-2**. Most instances of endogenous urate overproduction are linked to salvaged purines from increased cell turnover in proliferative and inflammatory disorders, such as hematologic malignancies and psoriasis, pharmacologic agents that elevate urate synthesis, like chemotherapy, or tissue hypoxia. A minority of hyperuricemia cases (10%) are due to inherited metabolic disorders, including heightened activity of 5′-phosphoribosyl 1-pyrophosphate (PRPP) synthetase or deficits in hypoxanthine-guanine phosphoribosyl transferase (HPRT) (Fig. 384-2). Genetic mutations causing PRPP synthetase overactivity result in an accelerated production of PRPP, purine nucleotides, and urate, culminating in conditions like gout and urate nephrolithiasis.

Approximately two-thirds of daily uric acid elimination occurs through renal excretion, with the remaining one-third via the gastrointestinal tract. Within the renal proximal tubule, distinct transporter systems are responsible for urate reabsorption and secretion. The reabsorption of urate at the apical membrane of the proximal tubule is sodium-dependent. This process involves the sodium-driven uptake of certain anions, which in turn promotes apical urate-anion exchange primarily through URAT1 and OAT10 transporters. Subsequently, urate is excreted across the basolateral membrane via the GLUT9 transporter. This mechanism underlies the notable anti-uricosuric actions of agents like pyrazinoate (PZA) and nicotinate. Meanwhile, urate entry into the intestinal tract is facilitated in part by ABCG2, a high-capacity urate efflux transporter encoded by the *ABCG2* gene and also present in the renal proximal tubular epithelium. In the intestine, urate undergoes degradation by gut bacteria (intestinal uricolysis).

In gout, urate crystals trigger a robust inflammatory response by stimulating the synthesis and release of inflammatory mediators. Central to this process is the phagocytosis of MSU crystals by neutrophils, which amplifies inflammation. Preceding neutrophil recruitment are events in the synovium and joint vasculature, involving activation of resident synovial phagocytes, including monocytes/macrophages and mast cells, and the NOD-like receptor protein 3 (NLRP3) inflammasome. This leads to the generation of pro-interleukin (IL)-1 and IL-1β, as well as other proinflammatory molecules like IL-6, IL-8, and tumor necrosis factor (TNF)-α.

Neutrophils respond to MSU crystals with a respiratory burst, releasing inflammatory agents such as lysosomal enzymes, superoxide anion, leukotriene B4, IL-1, and other cytokines and chemokines. The cellular mechanisms for neutrophil chemotaxis and activation involve kinases, phospholipases, chemoattractants, adhesion molecules, and other factors, highlighting the intense nature of the inflammatory response in gout.

If left untreated, gout can progress to chronic gouty arthritis, frequently with low-grade persistent synovitis and erosive deformities due to growing deposition of MSU crystals surrounded by granulomatous inflammation (tophi). Tophi, often found in proteoglycan-rich tissues like joints and tendons, encapsulate this chronic inflammatory process. These tophi engage both innate and adaptive immunity, representing a dynamic response to MSU crystals. Although occasionally exhibiting acute inflammatory signs, tophi predominantly signify a chronic inflammatory state in gout.

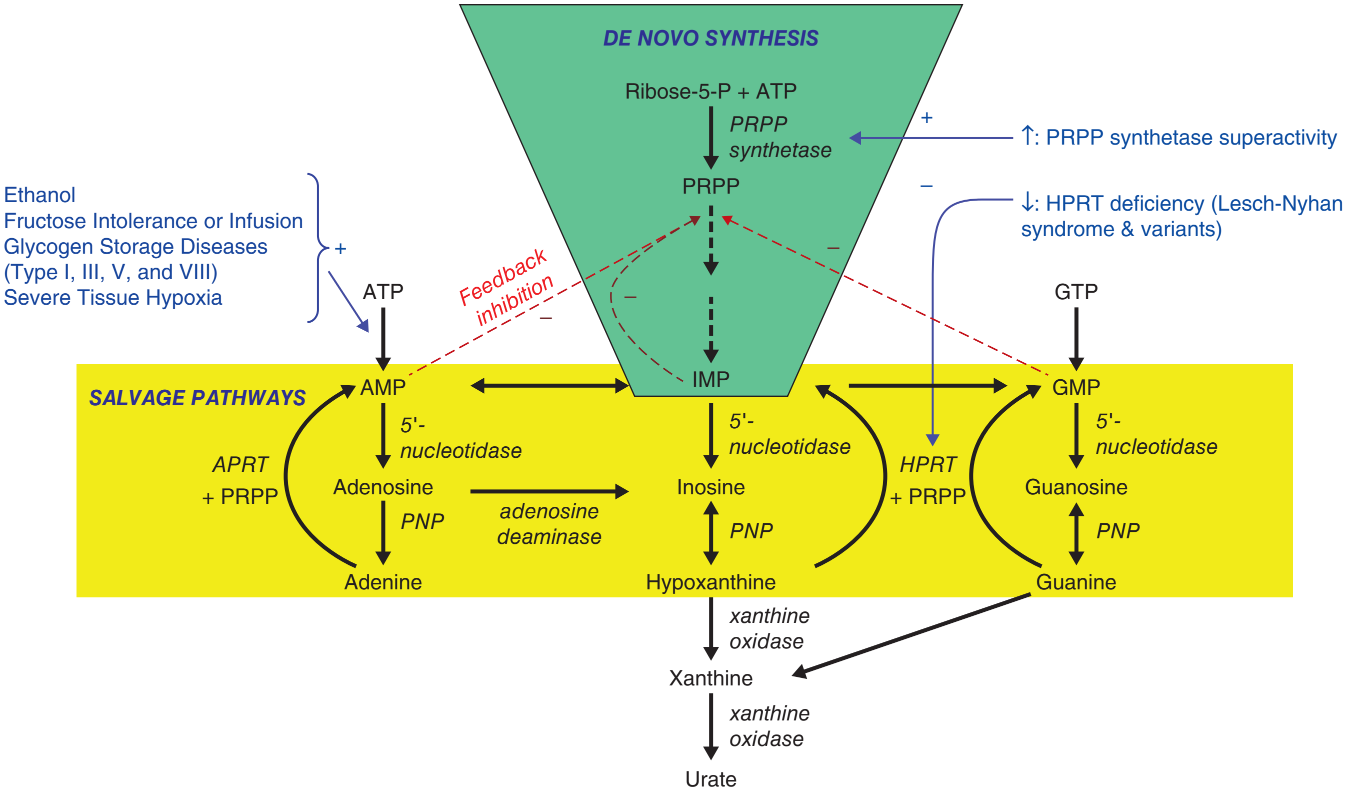

FIGURE 384-2 Urate production pathways implicated in the pathogenesis of gout. The de novo synthesis starts with 5′-phosphoribosyl 1-pyrophosphate (PRPP). PRPP is produced by addition of a further phosphate group from adenosine triphosphate (ATP) to the modified sugar ribose-5-phosphate. This step is performed by the family of PRPP synthetase (PRS) enzymes. Mutations in the PRS genes can result in overactivity of the pathway. Superactivity of PRS is an X chromosome-linked inborn error of metabolism, which leads to increased rate of PRPP, purine nucleotide, and urate production, in association with gout and urate urolithiasis. In addition, purine bases derived from tissue nucleic acids are reutilized through the salvage pathway. The enzyme hypoxanthine-guanine phosphoribosyl transferase (HPRT) salvages hypoxanthine to inosine monophosphate (IMP) and guanine to guanosine monophosphate (GMP). Mutations in the gene encoding HPRT are associated with a spectrum of disease that ranges from hyperuricemia alone to hyperuricemia with profound neurological and behavioral dysfunction (Lesch-Nyhan syndrome). Hypoxanthine cannot be reutilized without HPRT, and can only be degraded to urate. Both the underutilization of PRPP and decrease in IMP and GMP levels in the salvage pathway further contributes to hyperuricemia by feedback inhibition on de novo purine synthesis. The HPRT is located on the X chromosome and 250 different mutations of the gene have been described. Further, conditions associated with net ATP degradation lead to the accumulation of adenosine diphosphate (ADP) and adenosine monophosphate (AMP), which can be rapidly degraded to uric acid. These conditions are displayed in left upper corner. PNP, Purine nucleotide phosphorylase; APRT, adenine phosphoribosyl transferase.

CLINICAL MANIFESTATIONS

Early clinical manifestation of gout is characterized by acute recurrent gout flares. Usually, only one joint is affected initially, although oligo- and polyarticular gout flares can develop over time. The metatarsophalangeal joint of the first toe is involved in 70–90% of cases (podagra), followed by tarsal joints, ankles, and knees. Finger, wrist, and elbow joints can also be involved, although more often in elderly patients or in advanced disease. The gout flares often begin at night to early morning, constituting one of the most painful conditions experienced by humans. The affected joints and periarticular areas rapidly become warm, red, tender, and substantially swollen with a clinical appearance that often mimics that of cellulitis (pseudocellulitis). Typical flares tend to subside spontaneously within 1–2 weeks, and most patients have intervals of varying length with no residual symptoms until the next episode (intercritical gout). Triggers of gout flares include purine-rich food, alcohol, diuretic use, initial introduction of urate-lowering therapy, local trauma, and medical illnesses such as congestive heart failure and respiratory hypoxic conditions (Fig. 384-1).

Usually after years of gout flares without treatment, chronic gouty arthritis can develop, associated with ongoing synovitis, subcutaneous tophi, deformity, and bony destruction. Less commonly, chronic gouty arthritis will be the only manifestation, and more rarely, gout can manifest only as tophi. Women represent only 5–20% of all patients with gout. Most women with gout are postmenopausal and elderly; tend to have osteoarthritis, hypertension, or mild renal insufficiency; and usually are receiving diuretics. Premenopausal gout is rare, although kindreds of precocious gout in young women caused by decreased renal urate clearance and renal insufficiency have been described.

Genetic mutations in the *HPRT* gene are linked to a diverse range of clinical manifestations, extending from isolated hyperuricemia to severe neurologic and behavioral abnormalities (Lesch-Nyhan syndrome). In the absence of functional HPRT, hypoxanthine cannot be recycled and is exclusively converted to urate. This enzymatic deficiency leads to underuse of PRPP and a reduction in inosine monophosphate (IMP) and guanosine monophosphate (GMP) levels in the purine salvage pathway. These alterations contribute to hyperuricemia through feedback inhibition of de novo purine synthesis (Fig. 384-2). Owing to the X-linked inheritance of these enzyme defects, males who are homozygous for the mutation exhibit the disease phenotype. Carrier females, particularly postmenopausal, may develop gout and urinary tract stones. Notably, the presence of hyperuricemia in prepubertal boys should prompt consideration of these specific enzymatic deficits.

DIAGNOSIS

Laboratory Diagnosis Even if characteristic clinical features strongly suggest gout, the presumptive diagnosis ideally should be confirmed by needle aspiration of involved joints or tophaceous deposits. Acute septic arthritis, other crystal-associated arthropathies, palindromic rheumatism, and psoriatic arthritis can mimic clinical presentations of gout. During acute gout flares, needle-shaped MSU crystals typically are present both intracellularly and extracellularly **(Fig. 384-3)**. Under compensated polarized light, these crystals show bright, negative birefringence. Synovial fluid appears cloudy due to the increased numbers of leukocytes (e.g., from 5000–75,000/μL). Large amounts of crystals occasionally produce a thick, pasty, chalky joint fluid or drainage from distended tophus. Because bacterial infection can coexist with MSU crystals in synovial fluid, joint fluid is often stained and cultured for potential septic arthritis. MSU crystals also can often be demonstrated in the first metatarsophalangeal joint and

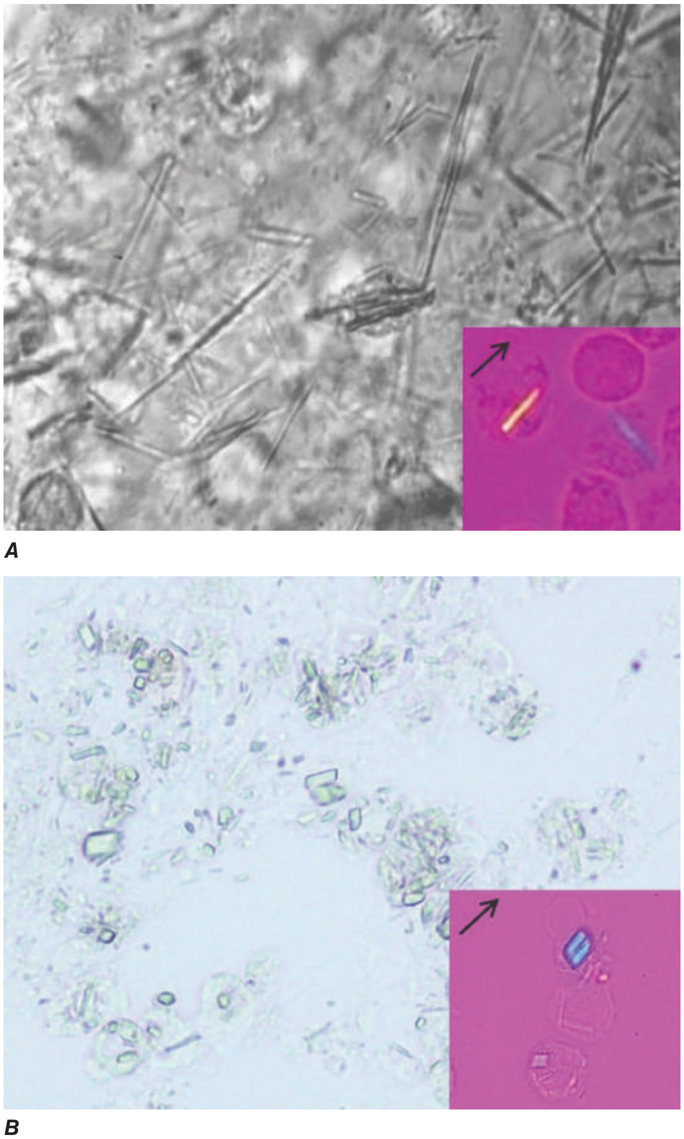

FIGURE 384-3 ***A.*** Extracellular and intracellular monosodium urate crystals, as seen in a fresh preparation of synovial fluid, illustrate needle- and rod-shaped crystals (400×). These crystals are strongly negative birefringent crystals under compensated polarized light microscopy. *(Inset 400×, provided by Eliseo Pascual.)* ***B.*** Intracellular and extracellular calcium pyrophosphate (CPP) crystals, as seen in a fresh preparation of synovial fluid, illustrate rectangular, rod-shaped, and rhomboid crystals (400×). These crystals are weakly positively or nonbirefringent crystals under compensated polarized light microscopy (inset 600×). *(Images provided by Eliseo Pascual.)*

in knees not acutely involved with gout, making arthrocentesis of these joints useful for the diagnosis of gout between flares.

While chronic hyperuricemia is a prerequisite in the pathogenesis of gout, serum urate levels can be normal or low at the time of an acute flare, as inflammatory cytokines' uricosuric property (e.g., IL-6) can lower the level by ~2 mg/dL. This tends to limit the value of serum urate testing in the setting of an acute flare. Nevertheless, serum urate levels are almost always elevated at some time in a gout patient's lifetime, and thus, it is important to seek historical or postflare serum urate values as a diagnostic clue or therapeutic target to urate-lowering therapy. Serum creatinine, liver function tests, hemoglobin, white blood cell (WBC) count, hemoglobin A_{1c}, and serum lipids are usually obtained at baseline to assess possible risk factor and comorbidities requiring treatment or monitored for potential adverse effects of gout treatments.

Radiographic Features Musculoskeletal ultrasound can timely aid early diagnosis by revealing a double-contour sign overlying the articular cartilage (signifying MSU deposition). Similarly, dual-energy computed tomography (CT) that utilizes two different energy beams and identifies MSU based on its chemical composition can indicate specific presence of MSU crystals. In plain radiography, cystic changes, well-defined erosions with sclerotic margins (often with overhanging bony edges), and soft tissue masses are characteristic features of advanced gout with tophaceous deposits, although these findings are typically absent in earlier stages of gout.

TREATMENT

Acute Gout Care

Although nonpharmacologic measures, such as ice pack application and rest of the involved joints, can be helpful, the mainstay of acute gout care is the administration of anti-inflammatory drugs such as colchicine, nonsteroidal anti-inflammatory drugs (NSAIDs), and glucocorticoids (Fig. 384-1). The choice of these options largely depends on the patients' comorbidities, concurrent medications, and previous response if recurrent flares. Early initiation of anti-inflammatories helps abort or reduce the severity of flares. Thus, for recurrent flares in established gout patients, patients can be provided a supply of their medications ready to start at the first sign of a flare. Oral colchicine is effective, particularly if used early in a gout flare. A low-dose regimen (1.2 mg with the first sign of a flare, followed by 0.6 mg in 1 h and subsequent-day dosing depending on response) is as effective as, and better tolerated than, the formerly used higher-dose regimens. Colchicine is eliminated from the body through P-glycoprotein (also known as MDR1) in the liver, small intestine, and kidneys as well as via enteric and hepatic cytochrome P450 3E4 (CYP3A4). As such, the dose, particularly for prolonged use, is reduced among patients with renal disease or when used with P-glycoprotein or CYP3A4 inhibitors such as clarithromycin or tacrolimus; additional caution is warranted among patients with hepatorenal impairment. NSAIDs are used most often in individuals without complicating comorbidities; NSAIDs given in full anti-inflammatory doses are effective in the vast majority of patients (e.g., indomethacin, 25–50 mg tid; naproxen, 500 mg bid; ibuprofen, 800 mg tid; and celecoxib, 800 mg followed by 400 mg 12 h later, then 400 mg bid).

Glucocorticoids given orally (e.g., prednisone, 30–50 mg/d as the initial dose and gradually tapered with the resolution of the attack) or as an intramuscular injection can be effective even in polyarticular gout. For a single joint or a few involved joints, intraarticular glucocorticoid injection is effective and well tolerated. With the central role of the inflammasome and IL-1β in gout flares, anakinra or canakinumab is a useful option when other treatments are contraindicated or have failed.

URATE-LOWERING THERAPY

Ultimate control of gout requires correction of the underlying chronic hyperuricemia with increased urate pool, the central cause for gout. Attempts to normalize serum urate to a subsaturation point (typically, <300–360 μmol/L [5.0–6.0 mg/dL]) to prevent recurrent gout flares and eliminate tophi are critical and entail a commitment to urate-lowering regimens that are required usually for life (Fig. 384-1). Urate-lowering drug therapy should be considered when, as in most patients, the hyperuricemia cannot be corrected by risk factor interventions (control of body weight, healthy diet, limitation of alcohol use, decreased consumption of fructose-rich foods and beverages, and avoidance of thiazide and loop diuretics). The decision to initiate urate-lowering drug therapy usually is made considering the number of gout flares (e.g., two or more attacks yearly), severity and duration of flares, or the patient's willingness to commit to long-term therapy. Urate-lowering therapy should be initiated in any patient who already has subcutaneous tophi or chronic gouty arthritis or known uric acid nephrolithiasis.

Allopurinol, a xanthine oxidase inhibitor, is the first-line urate-lowering drug among gout patients. Allopurinol can be given in a single morning dose, starting at 100 mg daily or less and titrating up (to 800 mg daily) to achieve a target serum urate level <5–6 mg/dL (i.e., a subsaturation point of MSU crystals) (Fig. 384-1). In patients with CKD, the starting allopurinol dose should be lowered depending on the CKD levels; for example, with an estimated glomerular filtration rate of <30 mL/min, one would start at 50 mg daily and titrate up slowly. Starting at a low dose and titrating up reduces the risk of severe cutaneous adverse reactions (SCARs), including Stevens-Johnson syndrome and toxic epidermal necrolysis, as well as the risk of gout flares associated with rapid serum urate reduction due to introduction of urate-lowering therapy (Fig. 384-1). Allopurinol is generally well tolerated, but mild cutaneous reactions can develop in ~2% of users. SCARs to allopurinol are rare but can be life-threatening, and thus, appropriate precaution is advised. Presence of CKD, higher allopurinol initial dosing (e.g., >100 mg daily in CKD patients), and *HLA-B***5801* carriage are important risk factors. As the *HLA-B***5801* carriage rate is high among Southeast Asian (except for Japanese descendants), Pacific Islander/Native Hawaiian, and black patients (leading to racial disparity in the risk of SCARs), screening for *HLA-B***5801* should be performed before starting allopurinol in these populations. If patients carry the *HLA-B***5801* allele, an alternative urate-lowering agent should be administered. Febuxostat is another xanthine oxidase inhibitor that is predominantly metabolized by glucuronide formation and oxidation in the liver and considered to not require dose adjustment in moderate to severe chronic kidney disease. It has also been used in patients who carry the *HLA-B***5801* allele.

Uricosuric agents such as probenecid are considered second-line urate-lowering therapies for gout and can be used in patients with good renal function either alone or in combination with xanthine oxidase inhibitors such as allopurinol. Probenecid can be started at a dose of 250 mg twice daily and increased gradually as needed up to 3 g/d to achieve and maintain a target serum urate level. Probenecid is generally not effective in patients with serum creatinine levels >177 μmol/L (2 mg/dL); benzbromarone (not available in the United States) is another uricosuric drug that is more effective in patients with CKD. Pegloticase is a pegylated uricase that is available for patients who do not tolerate or fail full doses of other treatments. It is often used in combination with methotrexate to reduce immunogenicity, which can lead to a loss of efficacy and infusion reactions. Therapies used to treat common cardiometabolic comorbidities of gout can also help lower serum urate levels, including losartan, amlodipine, fenofibrate, and sodium-glucose cotransporter-2 inhibitors (SGLT2i). Given their approved multiple cardiometabolic-kidney indications, SGLT2i can be beneficial for gout patients, offering both urate reduction and comprehensive comorbidity care.

Urate-lowering drugs are generally not initiated during active ongoing gout flares, given the potential worsening of the flare by acutely lowering serum urate levels. However, urate-lowering therapy can be started during the resolution phase of or after a gout flare, together with anti-inflammatory prophylaxis (e.g., colchicine 0.6 mg one to two times daily or naproxen 250 mg twice daily) to reduce the risk of the flares that often occur with the initiation of urate-lowering therapy. These paradoxical flares are presumed to be inflammatory reactions to MSU crystals shed from dissolution of organized MSU deposits (synovial tophi) induced by rapid serum urate reduction (Fig. 384-1). As such, faster urate reduction has been associated with higher risk of flares in clinical trials of urate-lowering drugs. Continuing concomitant anti-inflammatory prophylaxis is usually recommended until the patient is normouricemic and without gouty attacks for 3–6 months or tophi disappear.

CALCIUM PYROPHOSPHATE DEPOSITION DISEASE

CPP deposition (CPPD) predominately affects the elderly, with a doubling prevalence of CPPD in articular tissues with each decade >60 years of age (e.g., prevalence of nearly 50% for those >85 years old). In most cases, this process is asymptomatic with uncertain underlying etiology, although it is likely that biochemical changes in aging or diseased cartilage favor CPP crystal nucleation. Increased levels of inorganic pyrophosphate in cartilage matrix are thought to be a central pathogenic process in patients with CPPD arthritis, analogous to hyperuricemia in gout. This pyrophosphate can combine with calcium to form CPP crystals in cartilage matrix vesicles or on collagen fibers. Most inorganic pyrophosphate in cartilage matrix originates from breakdown of extracellular ATP. The ATP efflux (and thus the levels of inorganic pyrophosphate) is tightly regulated by a multipass membrane protein, ANKH. As such, mutations in the *ANKH* gene, as described in both familial and sporadic cases, have been found to increase elaboration and extracellular transport of pyrophosphate. The increase in pyrophosphate production is also related to enhanced activity of ATP pyrophosphohydrolase and 5′-nucleotidase, which catalyze the reaction of ATP to adenosine and pyrophosphate. Decreased levels of cartilage glycosaminoglycans and increased activities of transglutaminase enzymes may contribute to the deposition of CPP crystals.

TABLE 384-2 Factors and Conditions Associated with Calcium Pyrophosphate Crystal Deposition Disease

- Aging
- Osteoarthritis
- Postmeniscectomy or joint trauma
- Familial-genetic
- Endocrine-metabolic conditions
  - Primary hyperparathyroidism
  - Hemochromatosis
  - Hypophosphatasia
  - Hypomagnesemia
    - Certain drugs: thiazide and loop diuretics, potentially proton pump inhibitors
    - Malabsorption
    - Gitelman's syndrome
  - Gout
  - X-linked hypophosphatemic rickets
  - Familial hypocalciuric hypercalcemia

Similar to MSU crystals, release of CPP crystals into the joint space is followed by the phagocytosis of those crystals by monocyte-macrophages and neutrophils, which respond by releasing chemotactic and inflammatory substances and activating the inflammasome.

A minority of patients with CPPD arthropathy have metabolic abnormalities or hereditary CPPD **(Table 384-2)**. As such, the presence of CPPD arthritis in individuals aged <50 years should lead to consideration of these metabolic disorders and inherited forms of disease, including those identified in a variety of ethnic groups. Included among endocrine-metabolic conditions are hyperparathyroidism, hemochromatosis, hypophosphatasia, and hypomagnesemia. These associations suggest that a variety of different metabolic products may enhance CPP crystal deposition either by directly altering cartilage or by inhibiting inorganic pyrophosphatases. Investigation of younger patients with CPPD should include inquiry for evidence of familial aggregation and evaluation of serum calcium, phosphorus, alkaline phosphatase, magnesium, iron, and transferrin.

■ CLINICAL MANIFESTATIONS

Acute CPP crystal arthritis, originally termed *pseudogout* by McCarty and co-workers, often mimics acute gout flares with similar articular findings of substantial inflammation. There are several clinical clues that suggest CPP crystal arthritis. The knee is the joint most commonly affected, followed by the wrist, while the first metatarsophalangeal joint (podagra) is rarely affected. Other affected joints include the shoulder, ankle, elbow, and hands. Also, initial episodes of acute attacks tend to last longer than typical gout flares, even up to weeks to months. Acute attacks present sometimes with systemic signs such as fever, chills, and elevated acute-phase reactants. Acute CPP crystal arthritis may be precipitated by trauma, severe medical illness, or surgery, especially parathyroidectomy, as it leads to rapid diminution of serum calcium and magnesium levels, causing dissolution and shedding of crystal.

Chronic CPP crystal arthritis has several patterns; the most common form is a polyarticular arthritis resembling osteoarthritis (pseudoosteoarthritis). The clinical picture mimics that of progressive osteoarthritis, characterized by unusually severe joint damage in atypical joints for osteoarthritis, such as metacarpophalangeal, wrist, elbow, shoulder, or ankle joints. Other less common forms include chronic symmetric synovitis that is clinically similar to rheumatoid arthritis, severe destructive disease that may radiographically mimic neuropathic arthritis, intervertebral disk and ligament calcification with restriction of spine mobility, the crowned dens syndrome, spinal stenosis (most commonly seen in the elderly), and rarely, tumoral deposits of CPP crystals in soft tissues.

If radiographs or ultrasound reveal punctate and/or linear radiodense deposits within fibrocartilaginous joint menisci or articular hyaline cartilage (*chondrocalcinosis*), the diagnostic likelihood of CPPD disease is further increased. *Definitive diagnosis* requires demonstration of typical rhomboid or rodlike crystals (generally weakly positively birefringent or nonbirefringent with polarized light) in synovial fluid or articular tissue (Fig. 384-3). In the absence of joint effusion or indications to obtain a synovial biopsy, chondrocalcinosis is presumptive of CPPD. One exception is chondrocalcinosis due to CaOx in some patients with chronic renal failure.

In as many as 50% of cases, episodes of CPP crystal-induced inflammation are associated with low-grade fever and, on occasion, temperatures as high as 40°C (104°F). In such cases, synovial fluid analysis with microbial cultures is essential to rule out the possibility of infection. In fact, infection in a joint with any microcrystalline deposition process can lead to crystal shedding and subsequent synovitis from both crystals and microorganisms. The leukocyte count in synovial fluid in acute CPPD can range from several thousand cells to 100,000 cells/μL, with the mean being ~24,000 cells/μL and the predominant cell being the neutrophil. CPP crystals may be seen inside tissue fragments and fibrin clots and in neutrophils (Fig. 384-3). CPP crystals may coexist with MSU and apatite in some cases.

TREATMENT

CPPD Disease

Anti-inflammatory treatment for acute CPP crystal arthritis is largely adapted from that for gout flares, including ice pack applications with rest, joint fluid aspiration and glucocorticoid injection, oral colchicine, NSAIDs, and systemic glucocorticoids. Severe polyarticular attacks can also be treated with the IL-1β antagonist anakinra. For patients with frequent recurrent attacks, prophylactic treatment with daily colchicine can be helpful. For chronic CPP crystal arthritis, there is no effective way to remove CPP crystal deposits from cartilage and synovium (unlike urate-lowering agents in gout). Limited studies suggest that NSAIDs (with a gastric protective agent if required), colchicine, low-dose glucocorticoids, hydroxychloroquine, or methotrexate may be helpful for chronic synovitis. Patients with progressive destructive large-joint arthropathy may require joint replacement.

CALCIUM APATITE DEPOSITION DISEASE

PATHOGENESIS

Apatite is the primary mineral of normal bone and teeth. Abnormal accumulation of basic calcium phosphates, largely carbonate substituted apatite, can occur in areas of tissue damage (dystrophic calcification), hypercalcemic or hyperparathyroid states (metastatic calcification), connective tissue diseases (calcinosis), and other conditions **(Table 384-3)**. In chronic renal failure, hyperphosphatemia can contribute to extensive apatite deposition both in and around joints (tumoral calcinosis, calciphylaxis). Familial aggregation is rarely seen. Synovial lining cell or fibroblast cultures exposed to apatite (or CPP) crystals can undergo mitosis and markedly increase the release of prostaglandin E_2, various cytokines, and collagenases and neutral proteases, underscoring the destructive potential of abnormally stimulated synovial lining cells.

Although the true incidence of apatite arthritis is not known, 30–50% of patients with osteoarthritis have apatite microcrystals in their synovial fluid. Such crystals frequently can be identified in clinically stable osteoarthritic joints, but they are more likely to come to attention in persons experiencing acute or subacute worsening of joint pain and swelling.

TABLE 384-3 Clinical Manifestations of Apatite Crystal Deposition

- Periarticular
  - Calcific periarthritis (e.g., supraspinatus tendon apatite deposit rupture)
  - Bursitis and tendinitis
  - Hydroxyapatite pseudopodagra
  - Polyarticular involvement
  - Asymptomatic deposition
- Articular
  - Hemorrhagic shoulder effusions in the elderly (Milwaukee shoulder)
  - Chronic destructive arthropathy
  - Chronic erosive monoarthritis (resembling erosive osteoarthritis)
  - Acute synovitis
  - Association with osteoarthritis
- Secondary apatite crystal deposition
  - Tumoral calcinosis
  - Hyperparathyroidism
  - Calciphylaxis (renal failure/long-term dialysis)
  - Connective tissue diseases (e.g., systemic sclerosis, dermatomyositis)
  - Heterotopic calcification after neurologic catastrophes (e.g., stroke, spinal cord injury)
  - Fibrodysplasia ossificans progressiva

CLINICAL MANIFESTATIONS

Periarticular or articular deposits can present with acute reversible inflammation and/or chronic damage to the joint capsule, tendons, bursa, or articular surfaces (Table 384-3). The most common sites of apatite deposition include bursae and tendons in and/or around the knees, shoulders, hips, and fingers. These deposits can also be asymptomatic radiographic abnormalities.

Apatite aggregates are often present in synovial fluid in an extremely destructive chronic arthropathy of the elderly that occurs most often in the shoulders (Milwaukee shoulder) and in a similar process in hips, knees, and erosive osteoarthritis of fingers (Table 384-3). Joint destruction is associated with damage to cartilage and supporting structures, leading to instability and deformity. Progression tends to be indolent. Symptoms range from minimal to severe pain and disability that may lead to joint replacement surgery.

DIAGNOSIS

Intra- and/or periarticular calcifications with or without erosive, destructive, or hypertrophic changes may be seen on radiographs **(Fig. 384-4)**. They should be distinguished from the linear calcifications typical of CPPD.

The synovial fluid leukocyte count in apatite arthritis is usually low (<2000/μL) despite dramatic symptoms, with predominance of mononuclear cells. Definitive diagnosis of apatite arthropathy, also called basic calcium phosphate disease, depends on identification of crystals from synovial fluid or tissue (Fig. 384-4). Individual crystals are very small and can be seen only by electron microscopy. Clumps of crystals may appear as 1- to 20-μm shiny intra- or extracellular nonbirefringent globules or aggregates that stain purplish with Wright's stain and bright red with alizarin red S. Absolute identification depends on electron microscopy with energy-dispersive elemental analysis, x-ray diffraction, infrared spectroscopy, or Raman microspectroscopy, but these techniques usually are not required in clinical diagnosis.

TREATMENT

Calcium Apatite Deposition Disease

Treatment of apatite arthritis or periarthritis is nonspecific. Acute attacks of bursitis or synovitis may be self-limiting, resolving in

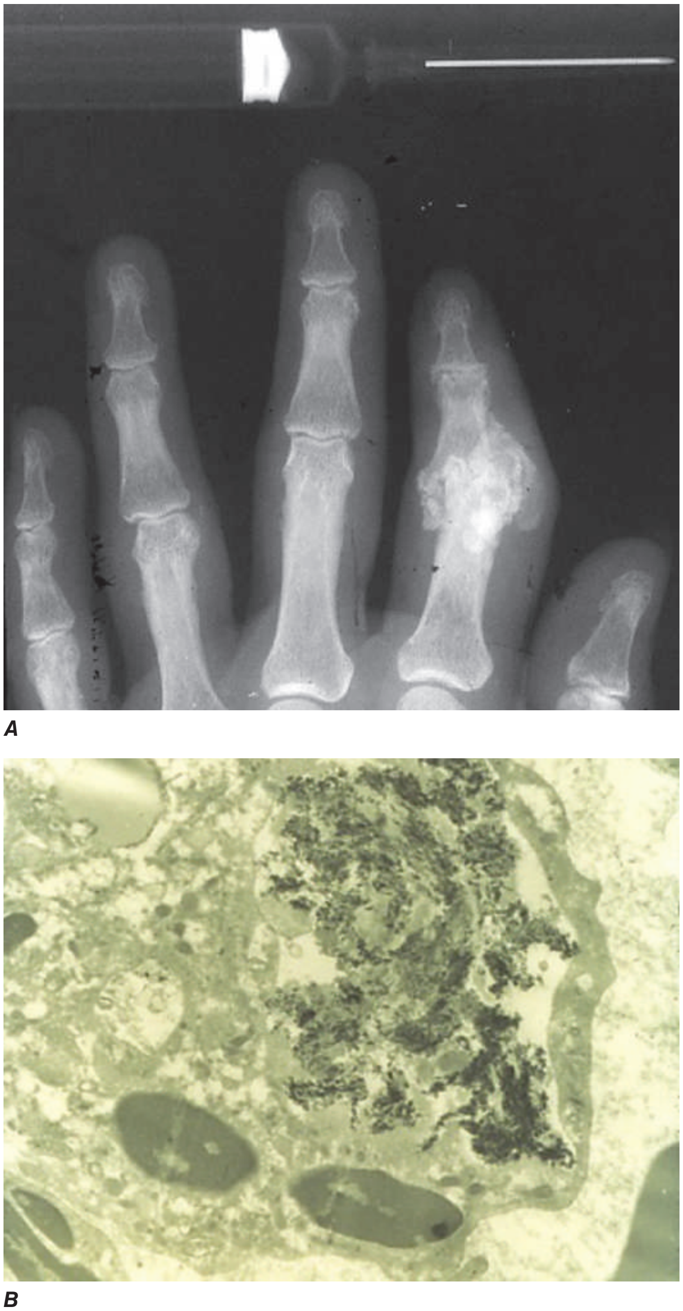

FIGURE 384-4 ***A.*** Radiograph showing calcification due to apatite crystals surrounding an eroded joint. ***B.*** An electron micrograph demonstrates dark needle-shaped apatite crystals within a vacuole of a synovial fluid mononuclear cell (30,000×).

days to several weeks. Aspiration of effusions and the use of either NSAIDs or oral colchicine for 2 weeks or intra- or periarticular injection of a depot glucocorticoid appear to shorten the duration and intensity of symptoms. Periarticular apatite deposits may be resorbed with resolution of attacks. Agents to lower serum phosphate levels may lead to resorption of deposits in renal failure patients receiving hemodialysis. In patients with underlying severe destructive articular changes, response to medical therapy is usually less rewarding.

CaOx DEPOSITION DISEASE

PATHOGENESIS

In chronic renal disease, CaOx deposits have long been recognized in visceral organs, blood vessels, bones, and cartilage and are now known to be one of the causes of arthritis in chronic renal failure. Thus far, reported patients have been dependent on long-term hemodialysis or peritoneal dialysis **(Chap. 323)**, and many had received ascorbic acid supplements. Ascorbic acid is metabolized to oxalate, which is inadequately cleared in uremia and by dialysis. Such supplements and foods high in oxalate content usually are avoided in dialysis programs because of the risk of enhancing hyperoxalosis and its sequelae.

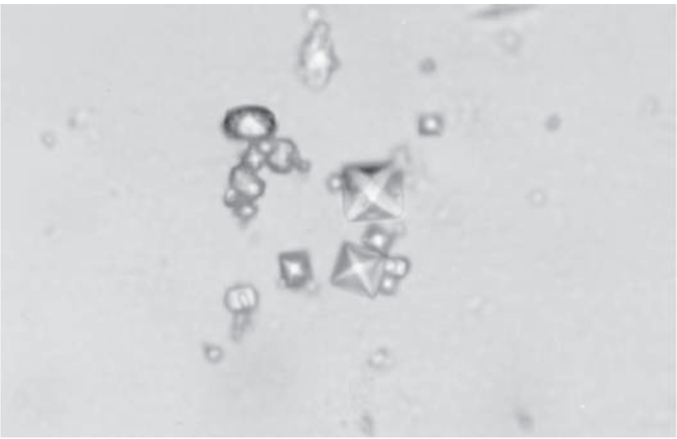

FIGURE 384-5 Bipyramidal and small polymorphic calcium oxalate crystals from synovial fluid are a classic finding in calcium oxalate arthropathy (ordinary light microscopy; 400×).

CaOx aggregates can be found in bone, articular cartilage, synovium, and periarticular tissues. From these sites, crystals may be shed, causing acute synovitis. Persistent aggregates of CaOx can, like apatite and CPP crystals, stimulate synovial cell proliferation and enzyme release, resulting in progressive articular destruction. *Primary oxalosis* is a rare hereditary metabolic disorder **(Chap. 431)** that can lead to acute or chronic CaOx arthritis, periarthritis, and bone disease during later years of illness.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Clinical features of acute CaOx arthritis may not be distinguishable from those due to MSU, CPP, or apatite. Deposits have been documented in fingers, wrists, elbows, knees, ankles, and feet. Radiographs may reveal chondrocalcinosis or soft tissue calcifications. CaOx-induced synovial effusions are usually noninflammatory, with <2000 leukocytes/μL, or mildly inflammatory. Neutrophils or mononuclear cells can predominate. CaOx crystals have a variable shape and variable birefringence to polarized light. The most easily recognized forms are bipyramidal, have strong birefringence **(Fig. 384-5)**, and stain with alizarin red S.

TREATMENT

Calcium Oxalate Deposition Disease

Treatment of CaOx arthropathy with NSAIDs, colchicine, intraarticular glucocorticoids, and/or an increased frequency of dialysis has produced only slight improvement. In primary oxalosis, liver transplantation has induced a significant reduction in crystal deposits **(Chaps. 356 and 431)**.

ACKNOWLEDGMENT

This chapter includes content on purine and pyrimidine metabolism written by John N. Mecchella and Christopher M. Burns in previous editions of this textbook.

FURTHER READING

Choi HK et al: Pathogenesis of gout. Ann Intern Med 143:499, 2005.
Dalbeth N et al: Gout. Nat Rev Dis Primers 5:69, 2019.
Rosenthal AK, Ryan LM: Calcium pyrophosphate deposition disease. N Engl J Med 374:2575, 2016.

385 Fibromyalgia

Leslie J. Crofford

DEFINITION

Fibromyalgia (FM) is characterized by chronic widespread musculoskeletal pain and tenderness. Although FM is defined primarily as a pain syndrome, patients also commonly report associated neuropsychological symptoms of fatigue, unrefreshing sleep, cognitive dysfunction, anxiety, and depression. Patients with FM have an increased prevalence of other syndromes associated with pain and fatigue, including myalgic encephalitis/chronic fatigue syndrome **(Chap. 461)**, temporomandibular disorder, chronic headaches, irritable bowel syndrome, interstitial cystitis/painful bladder syndrome, and other pelvic pain syndromes. These are collectively referred to as chronic primary or chronic overlapping pain conditions. Available evidence implicates the central nervous system as key to maintaining pain and other core symptoms of FM and related conditions. The presence of FM is associated with substantial negative consequences for physical and social functioning.

EPIDEMIOLOGY

Worldwide prevalence is ~2%, with a prevalence of ~4% in women and <1% in men. There is some variability depending on the method of ascertainment; however, the prevalence data are similar across world regions and socioeconomic classes. Cultural factors may play a role in determining whether patients with FM symptoms seek medical attention; however, even in cultures in which secondary gain is not expected to play a significant role, the prevalence of FM remains in this range. In clinical settings, a diagnosis of FM is far more common in women than in men, with a ratio of ~8:1. In population studies, the ratio of women to men is closer to 3:1. The prevalence of FM is much higher in patients with rheumatic diseases such as rheumatoid arthritis or systemic lupus erythematosus, where up to 30% have comorbid FM. Additional risk factors include sleep disturbances, physical inactivity, and overweight or obesity.

CLINICAL MANIFESTATIONS

Pain and Tenderness At presentation, patients with FM most commonly report "pain all over." Widespread pain is operationalized as being present both above and below the waist on both sides of the body and involving the axial skeleton (neck, back, or chest). The pain attributable to FM is poorly localized, difficult to ignore, severe in its intensity, and associated with a reduced functional capacity. For a diagnosis of FM, pain should have been present most of the day on most days for at least 3 months.

The pain of FM is associated with tenderness and increased evoked pain sensitivity. In clinical practice, this elevated sensitivity may be identified by pain induced by the pressure of a blood pressure cuff or skin roll tenderness. More formally, an examiner may complete a tender-point examination in which the examiner uses the thumbnail to exert pressure of ~4 kg/m^2 (or the amount of pressure leading to blanching of the tip of the thumbnail) on well-defined musculotendinous sites. Previously, the classification criteria of the American College of Rheumatology required that 11 of 18 sites be perceived as painful on exam for a diagnosis of FM. In practice, tenderness is a continuous variable, and strict application of a categorical threshold for diagnosis is not necessary. Newer criteria eliminate the need for identification of tender points and focus instead on patient-reported clinical symptoms of widespread or multisite pain and neuropsychological symptoms **(Fig. 385-1)**. The newer criteria perform well in clinical settings in comparison to the older, tender-point criteria. When subjective criteria are applied to populations, the result is an increase in prevalence of FM and a change in the sex ratio (see "Epidemiology," earlier).

Patients with FM often have peripheral pain generators that are thought to serve as triggers for the more widespread pain attributed to central nervous system factors. Potential pain generators such as arthritis, bursitis, tendinitis, neuropathies, and other inflammatory or degenerative conditions should be identified by history and physical examination. More subtle pain generators may include joint hypermobility and scoliosis. In addition, patients may have chronic myalgias triggered by infectious, metabolic, or psychiatric conditions that can serve as triggers for the development of FM. These conditions are often identified in the differential diagnosis of patients with FM, and a major challenge is to distinguish the ongoing pain of a triggering condition from FM pain that is occurring as a consequence of a comorbid condition and that should itself be treated.

Neuropsychological Symptoms In addition to widespread pain, FM patients typically report fatigue, stiffness, sleep disturbance, cognitive dysfunction, anxiety, and depression. These symptoms are present to varying degrees in most FM patients but are not present in every patient or at all times in a given patient. Relative to pain, such symptoms may, however, have an equal or even greater impact on function and quality of life. Fatigue is highly prevalent in patients in primary care who ultimately are diagnosed with FM. Pain, stiffness, and fatigue often are worsened by exercise or unaccustomed activity. The sleep complaints include difficulty falling asleep, difficulty staying asleep, and early-morning awakening. Regardless of the specific complaint, patients awake feeling unrefreshed. Patients with FM may meet criteria for restless legs syndrome and sleep-disordered breathing; frank sleep apnea can also be documented. Cognitive issues are characterized as difficulties with attention or concentration, problems with word retrieval, and short-term memory loss. Studies have demonstrated altered cognitive function in these domains in patients with FM, although speed of processing is age appropriate. Symptoms of anxiety and depression are common, and the lifetime prevalence of mood disorders in patients with FM approaches 80%. Although depression is neither necessary nor sufficient for the diagnosis of FM, it is important to screen for major depressive disorders by querying for depressed mood and anhedonia. Analysis of genetic factors that are likely to predispose to FM reveals shared neurobiologic pathways with mood disorders, providing the basis for comorbidity (see later in this chapter).

Overlapping Syndromes FM is considered as part of a group of conditions called chronic overlapping pain syndromes because of the propensity to coexist with other syndromes that may share underlying mechanisms. Review of systems often reveals headaches, facial/jaw pain, regional myofascial pain particularly involving the neck or back, and arthritis. Visceral pain involving the gastrointestinal tract, bladder, and pelvic or perineal region is often present as well. It is important for patients to understand that shared pathways may mediate symptoms and treatment strategies effective for one condition may help with global symptom management.

Comorbid Conditions FM is often comorbid with chronic musculoskeletal, infectious, metabolic, or psychiatric conditions. Whereas FM affects only ~2% of the general population, it occurs in ~10–30% of patients with degenerative or inflammatory rheumatic disorders, likely because these conditions serve as peripheral pain generators to alter central pain-processing pathways. The proposition that there may be an inflammatory or autoimmune etiology for FM in some patients has not been rigorously tested to date. It is particularly important for clinicians to be sensitive to pain management of these comorbid conditions so that when FM emerges—characterized by pain outside the boundaries of what could reasonably be explained by the triggering condition, development of neuropsychological symptoms, or tenderness on physical examination—treatment of central pain processes will be undertaken as opposed to a continued focus on treatment of peripheral or inflammatory causes of pain.

Psychosocial Considerations Symptoms of FM often have their onset and are exacerbated during periods of perceived stress. This pattern may reflect an interaction among central stress physiology, vigilance or anxiety, and central pain-processing pathways. An understanding of current psychosocial stressors will aid in patient

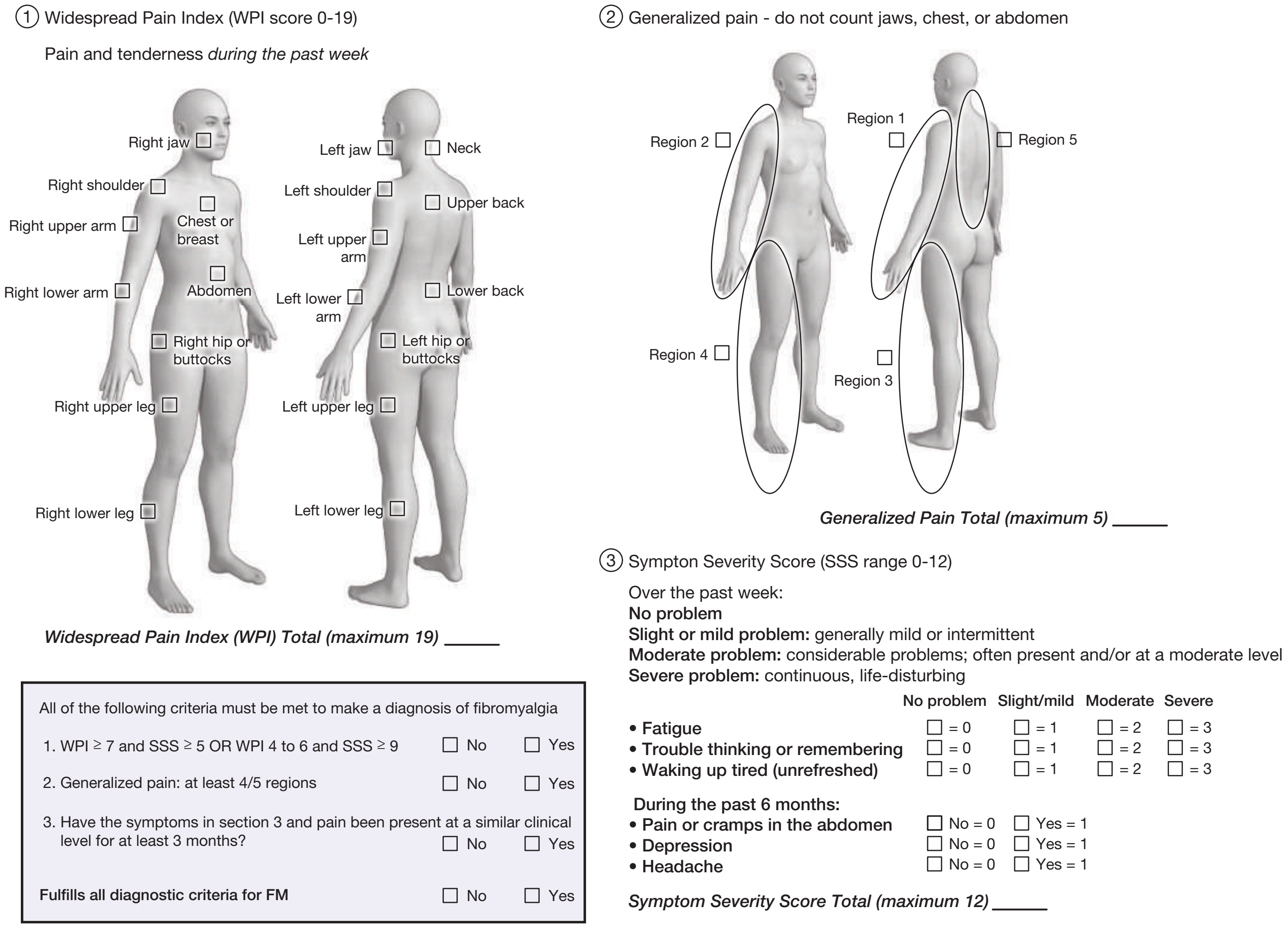
① Widespread Pain Index (WPI score 0-19)
Pain and tenderness *during the past week*
Right jaw ☐
Right shoulder ☐
Right upper arm ☐
☐ Chest or breast
Right lower arm ☐
☐ Abdomen
☐ Right hip or buttocks
Right upper leg ☐
Right lower leg ☐
Left jaw ☐
☐ Neck
Left shoulder ☐
☐ Upper back
Left upper arm ☐
Left lower arm ☐
☐ Lower back
☐ Left hip or buttocks
Left upper leg ☐
Left lower leg ☐
Widespread Pain Index (WPI) Total (maximum 19) ______
② Generalized pain - do not count jaws, chest, or abdomen
Region 2 ☐
Region 1 ☐
☐ Region 5
Region 4 ☐
☐ Region 3
Generalized Pain Total (maximum 5) ______

③ Sympton Severity Score (SSS range 0-12)

Over the past week:
No problem
Slight or mild problem: generally mild or intermittent
Moderate problem: considerable problems; often present and/or at a moderate level
Severe problem: continuous, life-disturbing

	No problem	Slight/mild	Moderate	Severe
• Fatigue	☐ = 0	☐ = 1	☐ = 2	☐ = 3
• Trouble thinking or remembering	☐ = 0	☐ = 1	☐ = 2	☐ = 3
• Waking up tired (unrefreshed)	☐ = 0	☐ = 1	☐ = 2	☐ = 3
During the past 6 months:				
• Pain or cramps in the abdomen	☐ No = 0	☐ Yes = 1		
• Depression	☐ No = 0	☐ Yes = 1		
• Headache	☐ No = 0	☐ Yes = 1		

Symptom Severity Score Total (maximum 12) ______

All of the following criteria must be met to make a diagnosis of fibromyalgia

1. WPI ≥ 7 and SSS ≥ 5 OR WPI 4 to 6 and SSS ≥ 9	☐ No	☐ Yes
2. Generalized pain: at least 4/5 regions	☐ No	☐ Yes
3. Have the symptoms in section 3 and pain been present at a similar clinical level for at least 3 months?	☐ No	☐ Yes
Fulfills all diagnostic criteria for FM	☐ No	☐ Yes

FIGURE 385-1 Fibromyalgia (FM) 2016 diagnostic criteria. *(Figure created using data from F Wolfe et al: Semin Arthritis Rheum 46:319, 2016.)*

management, as many factors that exacerbate symptoms cannot be addressed by pharmacologic approaches. Furthermore, there is a high prevalence of exposure to previous interpersonal and other forms of violence in patients with FM and related conditions. If posttraumatic stress disorder is an issue, the clinician should be aware of it and consider treatment options.

Functional Impairment It is crucial to evaluate the impact of FM symptoms on function and role fulfillment. In defining the success of a management strategy, improved function is a key measure. Functional assessment should include physical, mental, and social domains. Recognition of the ways in which role functioning falls short will be helpful in establishing treatment goals.

DIFFERENTIAL DIAGNOSIS

Because musculoskeletal pain is such a common complaint, the differential diagnosis of FM is broad. Table 385-1 lists some of the more common conditions that should be considered. Patients with inflammatory causes for widespread pain should be identifiable on the basis of specific history, physical findings, and laboratory or radiographic tests.

LABORATORY OR RADIOGRAPHIC TESTING

Routine laboratory and radiographic tests yield normal results in FM without comorbidities. Thus, diagnostic testing is focused on identification of other diagnoses and evaluation for pain generators or comorbid conditions (Table 385-2). Most patients with new chronic widespread pain should be assessed for the most common entities in the differential diagnosis. Radiographic testing should be used very sparingly and only for diagnosis of inflammatory arthritis. After the patient has been evaluated thoroughly, repeat testing is discouraged unless the symptom complex changes. Particularly to be discouraged is magnetic resonance imaging (MRI) of the spine unless there are features suggesting inflammatory spine disease or neurologic symptoms.

GENETICS AND PHYSIOLOGY

As in most complex diseases, it is likely that a number of genes contribute to vulnerability to the development of FM. To date, these genes appear to be in pathways controlling pain and stress responses. Some of the genetic underpinnings of FM are shared across other chronic pain conditions. Genes associated with metabolism, transport, and receptors of serotonin and other monoamines have been implicated in FM and overlapping conditions. Genes associated with other pathways involved in pain transmission have also been described as vulnerability factors for FM. Taken together, the pathways in which polymorphisms have been identified in FM patients further implicate central factors in mediation of the physiology that leads to the clinical manifestations of FM.

Psychophysical testing of patients with FM has demonstrated altered sensory afferent pain processing and impaired descending noxious inhibitory control leading to hyperalgesia and allodynia. Functional MRI and other research imaging procedures clearly demonstrate activation of the brain regions involved in the experience of pain in response to stimuli that are innocuous in study participants without FM. Pain perception in FM patients is influenced by the emotional and cognitive dimensions, such as catastrophizing and perceptions of control, providing a solid basis for recommendations for cognitive and behavioral treatment strategies.

Studies have indicated that some patients meeting criteria for FM may have a small fiber neuropathy. There have also been early reports of a possible autoimmune etiology for changes in the

TABLE 385-1 Common Conditions in the Differential Diagnosis of Fibromyalgia

Inflammatory
Polymyalgia rheumatica
Inflammatory arthritis: rheumatoid arthritis, spondyloarthritides
Connective tissue diseases: systemic lupus erythematosus, Sjögren's syndrome
Infectious
Hepatitis C
HIV infection
Lyme disease
Parvovirus B19 infection
Epstein-Barr virus infection
Noninflammatory
Degenerative joint/spine/disk disease
Myofascial pain syndromes
Bursitis, tendinitis, repetitive strain injuries
Endocrine
Hypo- or hyperthyroidism
Hyperparathyroidism
Neurologic Diseases
Multiple sclerosis
Neuropathic pain syndromes
Psychiatric Disease
Major depressive disorder
Drugs
Statins
Aromatase inhibitors

peripheral nervous system in patients with FM. Other studies have identified alterations in expressed gene or metabolic signatures in peripheral blood. These studies raise the possibility that confirmatory diagnostic testing could be developed in the future to assist in the diagnosis of FM.

APPROACH TO THE PATIENT

Fibromyalgia

FM is common and has an extraordinary impact on the patient's function and health-related quality of life. Optimal management requires prompt diagnosis and assessment of pain, function, and psychosocial context. Physicians and other health professionals can be helpful in managing some of the symptoms and impact of FM. Developing a partnership with patients is essential for improving the outcome of FM, with a goal of understanding the factors involved, implementing a treatment strategy, and choosing appropriate nonpharmacologic and pharmacologic treatments.

TABLE 385-2 Laboratory and Radiographic Testing in Patients with Fibromyalgia Symptoms

Routine
Erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP)
Complete blood count (CBC)
Thyroid-stimulating hormone (TSH)
Guided by History and Physical Examination
Complete metabolic panel
Antinuclear antibody (ANA)
Anti-SSA (anti–Sjögren's syndrome A) and anti-SSB
Rheumatoid factor and anti–cyclic citrullinated peptide (anti-CCP)
Creatine phosphokinase (CPK)
Viral (e.g., hepatitis C, HIV) and bacterial (e.g., Lyme) serologies
Spine and joint radiographs

Source: LM Arnold et al: J Women's Health 21:231, 2012; MA Fitzcharles et al: J Rheumatol 40:1388, 2013.

TREATMENT

Fibromyalgia

NONPHARMACOLOGIC TREATMENT

Patients with chronic pain, fatigue, and other neuropsychological symptoms require a framework for understanding the symptoms that have such an important impact on their function and quality of life. Explaining the genetics, triggers, and physiology of FM can be an important adjunct in relieving associated anxiety and in reducing the overall cost of health care resources. In addition, patients must be educated regarding expectations for treatment. The physician should focus on improved function and quality of life rather than elimination of pain. Illness behaviors, such as frequent physician visits, should be discouraged and behaviors that focus on improved function strongly encouraged.

Treatment strategies should include physical conditioning, with encouragement to begin at low levels of aerobic exercise and to proceed with slow but consistent advancement. Physical activity and exercise are consistently found to be the most helpful strategies. Exercise programs are helpful in reducing tenderness and enhancing self-efficacy. Patients who have been physically inactive may do best in supervised or water-based programs at the start. Strength training may be recommended after patients reach their aerobic goals. The U.S. Food and Drug Administration has approved devices including a laser therapy device and a transcutaneous electric nerve stimulation (TENS) device. A large randomized, placebo-controlled trial showed that TENS reduces movement-evoked pain and fatigue. Meditative movement therapies, such as qigong, yoga, or Tai Chi, may be helpful to manage symptoms. Other defined physical therapies such as acupuncture or hydrotherapy may also be considered. Cognitive-behavioral strategies to improve sleep hygiene and reduce illness behaviors can also be helpful in management.

PHARMACOLOGIC APPROACHES

It is essential for the clinician to treat any comorbid triggering condition and to clearly delineate for the patient the treatment goals for each medication. For example, glucocorticoids or nonsteroidal anti-inflammatory drugs may be useful for management of inflammatory triggers but are not effective against FM-related symptoms. At present, the treatment approaches that have proved most successful in FM patients target afferent or descending pain pathways. **Table 385-3** lists the drugs with demonstrated effectiveness. It should be emphasized that strong opioid analgesics are to be avoided in patients with FM. These agents have no demonstrated efficacy in FM and are associated with adverse effects that can worsen both symptoms and function. Tramadol, an opioid with mild serotonin-noradrenaline reuptake inhibitor activity, has been studied in this population with indication of efficacy, however its use is generally discouraged due to opioid-related adverse effects. Use of single agents to treat multiple symptom domains is strongly encouraged. For example, if a patient's symptom complex is dominated by pain and sleep disturbance, use of an agent that exerts both analgesic and sleep-promoting effects is desirable. These agents include cyclobenzaprine, sedating antidepressants such as amitriptyline, and alpha-2-delta ligands such as gabapentin and pregabalin. For patients whose pain is associated with fatigue, anxiety, or depression, drugs that have both analgesic and antidepressant/anxiolytic effects, such as duloxetine or milnacipran, may be the best first choice.

TABLE 385-3 Pharmacologic Agents Effective for Treatment of Fibromyalgia

DRUG	DOSE
Muscle relaxant	
Cyclobenzaprine	10–40 mg daily, at bedtime or in divided doses up to 4×/d
Antidepressants: balanced serotonin-norepinephrine reuptake inhibitors	
Amitryptiline[a]	10–50 mg/d; recommended to take at bedtime
Duloxetine[b,c]	60 mg/d; begin at 30 mg/d for 1 week
Milnacipran[b,c]	50 mg bid; titrate over 7 days[d]
Anticonvulsants: ligand of the alpha-2-delta subunit of voltage-gated calcium channels	
Pregabalin[b]	150–450 mg/d; begin at 75 mg bid for at least 1 week
Analgesic	
Tramadol	37.5 mg tramadol + 325 mg acetaminophen up to 4×/d

[a] RA Moore et al: Cochrane Database Syst Rev 12:CD008242, 2012. [b] Approved by the U.S. Food and Drug Administration. [c] W Hauser et al: Cochrane Database Syst Rev 1:CD010292, 2013. [d] Day 1: 12.5 mg once; days 2–3: 12.5 mg bid; days 4–7: 25 mg bid; after day 7: 50 mg bid.

Source: GJ Macfarlane et al: EULAR revised recommendations for the management of fibromyalgia. Ann Rheum Dis 76:318, 2017.

FURTHER READING

Bair MJ, Krebs EE: In the clinic: Fibromyalgia. Ann Intern Med 172:ITC33, 2020.

Macfarlane GJ et al: EULAR revised recommendations for the management of fibromyalgia. Ann Rheum Dis 76:318, 2017.

Wolfe F et al: 2016 Revisions to the 2010/2011 fibromyalgia diagnostic criteria. Semin Arthritis Rheum 46:319, 2016.

386

Arthritis Associated with Systemic Disease, and Other Arthritides

Carol A. Langford, Brian F. Mandell

ARTHRITIS ASSOCIATED WITH SYSTEMIC DISEASE

ARTHROPATHY OF ACROMEGALY

Acromegaly is the result of excessive production of growth hormone by a pituitary gland adenoma **(Chap. 392)**. The excessive secretion of growth hormone along with insulin-like growth factor I stimulates cartilage, periarticular connective tissue, and bone, resulting osteoarthritis, back pain, muscle weakness, and carpal tunnel syndrome. Diagnosis is usually between 40 and 50 years, but 5% are younger than 20.

Osteoarthritis is a common feature, affecting one or multiple knees, shoulders, hips, wrists, and fingers. Importantly, wrist involvement is not prominent in primary osteoarthritis. Initial radiographic widening of the joint space results from cartilage hypertrophy. The newly synthesized cartilage is abnormally susceptible to fissuring and damage. Ligamental laxity of joints further contributes to joint discomfort. As cartilage degrades, the joint space narrows and radiographic osteoarthritis develops. Joint fluid, if obtained, is usually noninflammatory. Chondrocalcinosis may occur, and calcium pyrophosphate dihydrate crystals present in the cartilage, when shed into the joint, can elicit attacks of pseudogout (calcium pyrophosphate arthropathy). Back pain is extremely common, perhaps as a result of spine hypermobility. Spine radiographs show normal or widened intervertebral disk spaces, anterior osteophytes, and ligamental calcification. The latter changes may mimic diffuse idiopathic skeletal hyperostosis. Dorsal kyphosis in conjunction with rib enlargement contributes to the development of the barrel chest in acromegalic patients. The hands and feet become enlarged as a result of soft tissue proliferation. The fingers are thickened and have "spadelike" distal tufts. Joint pains may persist after treatment of the acromegaly. One-third of patients have a thickened heel pad. Approximately one-third of patients exhibit Raynaud's phenomenon. Carpal tunnel syndrome occurs in about half of patients due to median nerve compression by excess connective tissue. Patients with acromegaly may develop proximal muscle weakness, which is thought to be caused by a noninflammatory effect of growth hormone on muscle. Serum muscle enzymes and electromyograms are normal.

ARTHROPATHY OF HEMOCHROMATOSIS

Hemochromatosis is a disorder of iron storage. Excessive intestinal absorption of iron leads to iron deposition in parenchymal cells, which results in impairment of organ function **(Chap. 426)**. Symptoms of hemochromatosis usually begin between the ages of 40 and 60 but can appear earlier. Arthropathy (20–40% of patients) usually begins after the age of 50 and may be the first clinical feature of hemochromatosis. The arthropathy is an osteoarthritis-like disorder affecting the small joints of the hands and later the larger joints, such as knees, ankles, shoulders, and hips. The second and third metacarpophalangeal joints of both hands are often the first and most prominent joints affected, providing a clue to the possibility of hemochromatosis because these joints are not predominantly affected by "routine" osteoarthritis. Patients experience some morning stiffness and pain with use of involved joints. The affected joints are enlarged and mildly tender. Hand pain in patients with hemochromatosis starts at an earlier age than in patients with primary osteoarthritis and may cause less disability due to the general sparing of the proximal thumb joint. The joint space narrows with subchondral sclerosis, subchondral cysts, and juxtaarticular proliferation of bone. Hooklike osteophytes of the metacarpophalangeal (MCP) joints are seen in up to 20% of patients; although they are regarded as a characteristic feature of hemochromatosis, they can also occur in primary osteoarthritis. However, dominant radiographic and clinical findings in the second and third MCP joints warrant evaluation of ferritin and iron/total iron-binding capacity levels. The typical radiographic changes of primary osteoarthritis in the proximal interphalangeal, distal interphalangeal, and first carpometacarpal joints are often not striking. The synovial fluid is noninflammatory. The synovium shows mild to moderate proliferation of iron-containing lining cells, fibrosis, and minimal inflammation. In approximately half of patients, there is calcium pyrophosphate deposition, and some patients may develop calcium pyrophosphate arthritis **(Chap. 384)**. An early diagnosis is suggested by high serum transferrin saturation, which is more sensitive than ferritin elevation.

Iron may damage the articular cartilage in several ways. Iron catalyzes superoxide-dependent lipid peroxidation, which may play a role in joint damage. In animal models, ferric iron has been shown to interfere with collagen formation and increase the release of lysosomal enzymes from cells in the synovial membrane. Iron inhibits synovial tissue pyrophosphatase (in vitro) and thus may result in chondrocalcinosis.

TREATMENT

Arthropathy of Hemochromatosis

The treatment of hemochromatosis is repeated phlebotomy. Phlebotomy often has little effect on established arthritis, which, along with chondrocalcinosis, may progress. Symptom-based treatment of the arthritis consists of administration of acetaminophen and nonsteroidal anti-inflammatory drugs (NSAIDs), as tolerated. Acute calcium pyrophosphate arthropathy flares are treated with high doses of an NSAID or a short course of glucocorticoids. Colchicine may have efficacy in treating and limiting the number of flares. Advanced osteoarthritis can be treated with arthroplasty.

HEMOPHILIC ARTHROPATHY

Hemophilia is a sex-linked recessive genetic disorder characterized by the absence or deficiency of factor VIII (hemophilia A, or classic hemophilia) or factor IX (hemophilia B, or Christmas disease) **(Chap. 121)**. Hemophilia A constitutes 85% of cases. Spontaneous hemarthrosis is a common problem with both types of untreated hemophilia and can lead to a deforming arthritis. The severity of hemarthrosis is related to the degree of clotting factor deficiency. Hemarthrosis is not common in other disorders of coagulation such as von Willebrand disease, factor V deficiency, oral anticoagulation, or thrombocytopenia.

Hemarthrosis occurs after 1 year of age, when a child begins to walk and run. The joints most commonly affected are the knees, ankles, elbows, shoulders, and hips. Small joints of the hands and feet are occasionally involved.

In the initial stage of arthropathy, hemarthrosis produces a warm, tensely swollen, and painful joint. Blood in the joint remains liquid because of the absence of intrinsic clotting factors and the absence of tissue thromboplastin in the synovium. The synovial blood is resorbed over a period of ≥1 week, with the precise interval depending on the size of the hemarthrosis. Joint function usually returns to normal or baseline in ~2 weeks. Low-grade temperature elevation may accompany hemarthrosis, but a fever >38.3°C (101°F) warrants concern about infection, and joint aspiration should be considered.

Recurrent hemarthrosis may result in chronic noninflammatory, fibrotic arthropathy. The joints remain swollen, and flexion deformities may develop. Joint motion may be restricted and function severely limited.

Bleeding into muscle and soft tissue also causes musculoskeletal dysfunction. When bleeding into the iliopsoas muscle occurs, the hip is held in flexion because of the pain, resulting in a hip flexion contracture. Rotation of the hip is generally preserved, distinguishing this from hemarthrosis or hip synovitis. Expansion of the hematoma may place pressure on the femoral nerve, resulting in femoral neuropathy. Hemorrhage into a closed compartment space, such as the calf or the volar compartment in the forearm, can result in muscle necrosis, neuropathy, and flexion deformities of the ankles, wrists, and fingers. When bleeding involves periosteum or bone, a painful pseudotumor forms. These pseudotumors occur distal to the elbows or knees in children and improve with treatment of hemophilia. Surgical removal is indicated if the pseudotumor continues to enlarge. In adults, pseudotumors develop in the femur and pelvis and are usually refractory to treatment. When bleeding occurs in muscle, cysts may develop within the muscle. Needle aspiration of a cyst is relatively contraindicated because it can induce further bleeding; however, if the cyst becomes secondarily infected, drainage may be necessary (after factor repletion).

Septic arthritis is rare in hemophilia but is difficult to distinguish from acute hemarthrosis on physical examination. If there is serious suspicion of an infected joint, the joint should be aspirated immediately, the fluid cultured, and empiric treatment with broad-spectrum antibiotics administered, with coverage including *Staphylococcus*, until culture results become available. Clotting factor deficiency should be corrected before arthrocentesis to minimize the risk of traumatic bleeding. Notably, low-grade fevers can occur in the setting of acute hemarthrosis.

Radiographs of joints reflect the stage of disease. Initially there is only capsule distention; later, juxtaarticular osteopenia, marginal erosions, and subchondral cysts may develop. Late in the disease, the joint space is narrowed and there is bony overgrowth similar to that in osteoarthritis. Blood collections may calcify.

TREATMENT

Musculoskeletal Bleeding

The treatment of musculoskeletal bleeding is initiated with the immediate infusion of factor VIII or IX at the first sign of joint or muscle hemorrhage. Patients who have developed factor inhibitors are at elevated risk for joint damage and may benefit from receiving recombinant activated factor VII or activated prothrombin complex concentrate. The joint should be splinted in gentle forced extension, as tolerated, to avoid contracture. Analgesia should be provided; nonselective NSAIDs, which can diminish platelet function, should be avoided if possible. Selective cyclooxygenase-2 inhibitors do not interfere with platelet function, although cardiovascular, renal, and gastrointestinal concerns should be considered. Synovectomy—open or arthroscopic—may be attempted to limit future bleeding and fibrosis in patients with chronic symptomatic synovial proliferation and recurrent hemarthrosis, although hypertrophied synovium is highly vascular and subject to bleeding. Open surgical synovectomy has been associated with decreased range of motion. Both require aggressive prophylaxis against bleeding. Radiosynovectomy with either yttrium-90 silicate or phosphorus-31 colloid has been effective and may be attempted if available when surgery is not advised. Total joint replacement is indicated for severe joint destruction and incapacitating pain.

ARTHROPATHIES ASSOCIATED WITH HEMOGLOBINOPATHIES

Sickle Cell Disease Sickle cell disease **(Chap. 103)** is associated with several musculoskeletal abnormalities **(Table 386-1)**. Children aged <5 years may develop diffuse swelling, tenderness, and warmth of the hands and feet lasting 1–3 weeks. This condition, referred to as *sickle cell dactylitis* or *hand-foot syndrome*, has also been observed in sickle cell thalassemia. Dactylitis is believed to result from infarction of the bone marrow and cortical bone leading to periostitis and soft tissue swelling. Radiographs show periosteal elevation, subperiosteal new-bone formation, and areas of radiolucency and increased density involving the metacarpals, metatarsals, and proximal phalanges. These bone changes disappear after several months. The syndrome leaves little or no residual damage. This syndrome is rarely seen after age 5.

Sickle cell crisis is associated with periarticular and long bone pain, occasionally with joint effusions. The joint and periarticular areas are warm and tender. Knees and elbows are most often affected, but other joints can be involved. Joint effusions are usually noninflammatory, but acute synovial infarction can cause a sterile joint effusion with a high neutrophil count. Scintigraphic studies have shown decreased marrow uptake adjacent to the involved joint. The treatment for sickle cell crisis is detailed in **Chap. 103**.

Patients with sickle cell disease seem predisposed to osteomyelitis, perhaps related to splenic dysfunction, which commonly involves the long tubular bones **(Chap. 136)**; *Salmonella* is a particularly associated cause **(Chap. 171)**. Radiographs of the involved site initially show periosteal elevation, with subsequent disruption of the cortex. Treatment of the infection results in healing of the bone lesion. Sterile bone infarction resulting from vasoocclusion secondary to the sickling of red cells can occur and contribute to the painful crises. Bone infarction also occurs in patients with sickle cell trait and sickle cell thalassemia **(Chap. 103)**. In children, infarction of the epiphyseal growth plate may interfere with normal growth of the affected extremity. Radiographically, infarction of the bone cortex results in periosteal elevation and irregular cortical thickening. Infarction in the bone marrow leads to lysis, fibrosis, and new bone formation. Clinical distinction between osteomyelitis and bone infarctions can be difficult; nuclear imaging may be helpful.

TABLE 386-1 Musculoskeletal Abnormalities in Sickle Cell Disease

Sickle cell dactylitis	Avascular necrosis
Joint effusions in sickle cell crises	Bone changes secondary to marrow hyperplasia
Osteomyelitis	Septic arthritis
Infarction of bone	Gouty arthritis
Infarction of bone marrow	Synovial infarction

Avascular necrosis of the femoral head occurs in ~5% of patients. It also occurs in the humeral head and less commonly in the distal femur, tibial condyles, distal radius, vertebral bodies, and other juxtaarticular sites. Periarticular damage may contribute to the development of osteoarthritis. Radiography of the affected joint may show patchy radiolucency and density followed by collapse or change in bone shape, but magnetic resonance imaging (MRI) is a far more sensitive technique for detecting early avascular necrosis. Joint replacements may markedly reduce pain and improve function.

Septic arthritis is occasionally encountered in sickle cell disease **(Chap. 135)**. Multiple joints may be infected. Joint infection may result from bacteremia due to splenic dysfunction or from contiguous osteomyelitis. The more common microorganisms include *Staphylococcus aureus* and *Streptococcus*. *Salmonella* does not cause septic arthritis as frequently as it causes osteomyelitis. Acute gouty arthritis may occur in young patients with sickle cell disease, as up to 40% of patients are hyperuricemic due to increased urate generation and interstitial renal disease. Attacks may be polyarticular, and diagnostic arthrocentesis should be performed to distinguish infection from gout or synovial infarction.

The bone marrow hyperplasia in sickle cell disease results in widening of the medullary cavities, thinning of the cortices, and coarse trabeculations and central cupping of the vertebral bodies. These changes are also seen to a lesser degree in patients with sickle cell thalassemia. In normal individuals, red marrow is located mostly in the axial skeleton, but in sickle cell disease, red marrow is found in the bones of the extremities and even in the tarsal and carpal bones. Vertebral compression may lead to dorsal kyphosis, and softening of the bone in the acetabulum may result in protrusio acetabuli.

Thalassemia A congenital disorder of hemoglobin synthesis, β thalassemia is characterized by impaired production of β chains **(Chap. 103)**. Bone and joint abnormalities occur in β thalassemia, being most common in the major and intermedia types. In one study, ~50% of patients with β thalassemia had evidence of symmetric ankle arthropathy characterized by a dull aching pain that was aggravated by weight bearing. The onset came most often in the second or third decade of life. The degree of ankle pain in these patients varied. Symptoms eventually abated in a few patients. Synovial fluid from two patients was noninflammatory. Radiographs of the ankle showed osteopenia, widened medullary spaces, thin cortices, and coarse trabeculations—findings that are consistent with bone marrow expansion. The joint space was preserved. Iron staining was found in the bone trabeculae, in osteoid, in the cement line, and in the synovium. This arthropathy was considered to be secondary to the underlying bone pathology and was treated with analgesics and splints as well as transfusions in some patients to decrease hematopoiesis and bone marrow expansion.

In patients with β-thalassemia major and β-thalassemia intermedia, other joints are also involved, including the knees, hips, and shoulders. Gouty arthritis and septic arthritis have been reported. Avascular necrosis is not a feature of thalassemia because there is vasoocclusion leading to thrombosis and infarction.

β-Thalassemia minor (also known as *β-thalassemia trait*) is likewise associated with joint manifestations. Chronic seronegative oligoarthritis affecting predominantly ankles, wrists, and elbows has been described; the affected patients had mild persistent synovitis without large effusions or joint erosions. Recurrent episodes of acute asymmetric arthritis have also been reported; episodes last <1 week and may affect the knees, ankles, shoulders, elbows, wrists, and metacarpal phalangeal joints. The mechanism underlying this arthropathy is unknown. Treatment with NSAIDs is not particularly effective. Patients receiving frequent transfusions may be more likely to have circulating antinuclear antibodies, but not systemic lupus erythematosus.

■ MUSCULOSKELETAL DISORDERS ASSOCIATED WITH HYPERLIPIDEMIA

(See also Chap. 419) Musculoskeletal or cutaneous manifestations may be the first clinical indication of a hereditary disorder of lipoprotein metabolism. Patients with rare familial hypercholesterolemia (previously referred to as *type II hyperlipoproteinemia*) may have recurrent migratory polyarthritis involving the knees and other large peripheral joints and, to a lesser degree, peripheral small joints. Pain ranges from moderate to incapacitating. The involved joints can be warm, erythematous, swollen, and tender. Arthritis usually has a sudden onset, lasts from a few days to 2 weeks, and does not cause joint damage and suggest an acute gout flare. These may be recurrent. Synovial fluid from involved joints is not inflammatory and contains no crystals. Flares may be inflammatory periarthritis or peritendinitis and not true arthritis. Attacks of tendinitis, including the large Achilles and patellar tendons, may come on gradually and last only a few days or may be acute as described above. Patients may be asymptomatic between attacks. Achilles tendinitis and other joint manifestations often precede the appearance of xanthomas and may be the first clinical indication of hyperlipoproteinemia. Attacks of tendinitis have been reported following initiation of a lipid-lowering drug. Over time, patients may develop tendinous xanthomas in the Achilles, patellar, and extensor tendons of the hands and feet. Xanthomas have also been reported in the peroneal tendon, the plantar aponeurosis, and the periosteum overlying the distal tibia. These xanthomas are located within tendon fibers. Tuberous xanthomas are soft subcutaneous masses located over the extensor surfaces of the elbows, knees, and hands as well as on the buttocks. They appear during childhood in homozygous patients and after the age of 30 in heterozygous patients. Patients with elevated plasma levels of very-low-density lipoprotein (VLDL) and triglycerides (previously referred to as *type IV hyperlipoproteinemia*) may also have a mild inflammatory arthritis affecting large and small peripheral joints, usually in an asymmetric pattern, with only a few joints involved at a time. The onset of arthritis usually comes in middle age. Radiographs may show juxtaarticular osteopenia and cystic lesions. Large bone cysts have been noted in a few patients. Xanthoma and bone cysts are also observed in other lipoprotein disorders. The pathogenesis of arthritis in patients with familial hypercholesterolemia or with elevated levels of VLDL and triglycerides is not well understood. NSAIDs or analgesics usually provide adequate relief of symptoms when used on an as-needed basis.

Patients may improve clinically as they are treated with lipid-lowering agents; however, patients treated with an HMG-CoA reductase inhibitor may experience myalgias, and rare patients develop myopathy or myositis. Testing for anti-HMGCR autoantibodies in patients with elevated muscle enzymes on treatment may identify the patients with statin-induced necrotizing autoimmune myopathy who require aggressive immunosuppressive therapy. Myositis (and gout) has been reported with the use of niacin **(Chap. 377)** but is less common than myalgias.

Musculoskeletal syndromes have not clearly been associated with the more common mixed hyperlipidemias seen in general practice. The incidence of gout is increased in patients taking bempedoic acid.

OTHER ARTHRITIDES

■ NEUROPATHIC JOINT DISEASE

Neuropathic joint disease (Charcot joint) is a progressive destructive arthritis associated with loss of pain sensation, proprioception, or both. Normal muscular reflexes that modulate joint movement are impaired. Without these protective mechanisms, joints are subjected to repeated trauma, resulting in progressive cartilage and bone damage. Diabetes mellitus is the most frequent cause of neuropathic joint disease **(Fig. 386-1)**. A variety of other disorders are associated with neuropathic arthritis, including tabes dorsalis, leprosy, yaws, syringomyelia, meningomyelocele, congenital indifference to pain, peroneal muscular atrophy (Charcot-Marie-Tooth disease), and amyloidosis. An arthritis resembling neuropathic joint disease has been reported in patients who have received frequent intraarticular glucocorticoid injections, but this was not observed in a series of patients with knee osteoarthritis who received intraarticular glucocorticoid injections every 3 months for 2 years. The distribution of joint involvement depends on the underlying neurologic disorder **(Table 386-2)**. In tabes dorsalis, the knees,

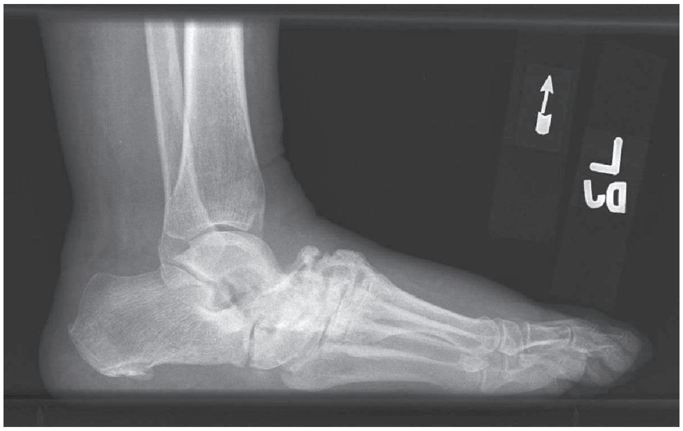

FIGURE 386-1 Charcot arthropathy associated with diabetes mellitus. Lateral foot radiograph demonstrating complete loss of the arch due to bony fragmentation and dislocation in the midfoot. *(Courtesy of Andrew Neckers, MD, and Jean Schils, MD; with permission.)*

hips, and ankles are most commonly affected; in syringomyelia, the glenohumeral joint, elbow, and wrist; and in diabetes mellitus, the tarsal and tarsometatarsal joints.

PATHOLOGY AND PATHOPHYSIOLOGY

The pathologic changes in the neuropathic joint are similar to those found in the severe osteoarthritic joint. There is fragmentation and eventual loss of articular cartilage with eburnation of the underlying bone. Osteophytes are found at the joint margins. With more advanced disease, erosions are present on the joint surface. Fractures, devitalized bone, intraarticular loose bodies, and microscopic fragments of cartilage and bone may be present.

At least two underlying mechanisms are believed to be involved in the pathogenesis of neuropathic arthritis. An abnormal autonomic nervous system is thought to be responsible for the dysregulated blood flow to the joint with subsequent resorption of bone. Loss of bone, particularly in the diabetic foot, may be the initial finding. With the loss of deep pain, proprioception, and protective neuromuscular reflexes, the joint is subjected to repeated microtrauma, resulting in ligamental tears and bone fractures. The injury that may follow frequent intraarticular glucocorticoid injections is thought to be due to the analgesic effect of glucocorticoids, leading to overuse of an already damaged joint with accelerated cartilage damage, although steroid-induced cartilage damage is more common in animal species than in humans. There have been reports of accelerated joint damage occurring in some animals and occasional patients with the therapeutic use of nerve growth factor antagonists. It is not understood why only a few patients with neuropathy develop clinically evident neuropathic arthritis.

CLINICAL MANIFESTATIONS

Neuropathic joint disease usually begins in a single joint and then becomes apparent in other joints, depending on the underlying neurologic disorder. The involved joint becomes progressively enlarged as a result of bony overgrowth and synovial effusion. Free-floating cartilage and bone fragments may cause mechanical joint symptoms. Neuropathic joints may develop rapidly, and a totally disorganized joint with multiple bony fragments may evolve within weeks or months. Less pain is reported by patients than would be anticipated from the degree of joint damage, but they may experience sudden pain from intraarticular fractures of osteophytes.

TABLE 386-2 Disorders Associated with Neuropathic Joint Disease

Diabetes mellitus	Amyloidosis
Tabes dorsalis	Leprosy
Meningomyelocele	Congenital indifference to pain
Syringomyelia	Peroneal muscular atrophy

Neuropathic arthritis is estimated to occur in patients with diabetes mellitus with an incidence of ~0.5%, usually at ≥50 years in a patient who has had diabetes for several years, but exceptions occur. The tarsal and tarsometatarsal joints are most often affected, with the metatarsophalangeal and talotibial joints next most commonly involved. Patients often attribute the onset of foot pain to antecedent trauma such as twisting of the foot. Neuropathic changes may develop rapidly after a foot fracture or dislocation. The foot and ankle are often swollen. Downward collapse of the tarsal bones leads to convexity of the sole, referred to as a "rocker foot." Large osteophytes may protrude from the top of the foot. Calluses frequently form over the metatarsal heads and may lead to infected ulcers and osteomyelitis. The value of protective inserts and orthotics, as well as regular foot examination, cannot be overstated. Radiographs may show resorption and tapering of the distal metatarsal bones. The term *Lisfranc fracture-dislocation* is sometimes used to describe the destructive changes at the tarsometatarsal joints.

DIAGNOSIS

The diagnosis of neuropathic arthritis is based on the clinical features and characteristic radiographic findings in a patient with underlying sensory neuropathy. The differential diagnosis of neuropathic arthritis depends on the severity of the process and includes osteomyelitis, avascular necrosis, advanced osteoarthritis, stress fractures, and calcium pyrophosphate deposition disease. Patients may have more than one of these conditions. Imaging in neuropathic arthritis initially shows changes of osteoarthritis with joint space narrowing, subchondral bone sclerosis, osteophytes, and joint effusions; marked destructive and hypertrophic changes occur later. The radiographic findings of neuropathic arthritis may be difficult to differentiate from those of osteomyelitis, especially in the diabetic foot. The joint margins in a neuropathic joint tend to be distinct, while in osteomyelitis, they are blurred. Imaging studies may be helpful, but cultures of tissue from the joint are often required to exclude osteomyelitis. MRI and nuclear scans using indium-111–labeled white blood cells, which may show increased uptake in osteomyelitis but not in a neuropathic joint, may occasionally be useful. A technetium bone scan will not distinguish osteomyelitis from neuropathic arthritis, as increased uptake is observed in both. The joint fluid in neuropathic arthritis is noninflammatory; it may be xanthochromic or even bloody and contain fragments of synovium, cartilage, and bone. The finding of calcium pyrophosphate dihydrate crystals supports the diagnosis of crystal-associated arthropathy. In the absence of such crystals, an increased number of synovial neutrophils suggests possible osteomyelitis.

TREATMENT

Neuropathic Joint Disease

The primary focus of treatment is to stabilize the joint. Treatment of the underlying disorder, even if successful, does not usually affect established joint disease. Braces and splints are helpful. Their use requires close surveillance, because patients may be unable to appreciate pressure from a poorly adjusted brace. In the diabetic patient, early recognition of Charcot foot and its treatment—prohibition of weight bearing by the foot for at least 8 weeks—may prevent severe disease from developing. Fusion of an unstable joint may improve function and reduce pain, but nonunion is frequent, especially when immobilization of the joint is inadequate.

HYPERTROPHIC OSTEOARTHROPATHY AND CLUBBING

Hypertrophic osteoarthropathy (HOA) is characterized by clubbing of digits and, in more advanced stages, by periosteal new-bone formation. Primary or familial HOA may begin in childhood. Secondary HOA is associated with intrathoracic malignancies, suppurative and some hypoxemic lung diseases, congenital heart disease, hepatopulmonary disease, and a variety of other disorders. Clubbing is almost always a feature of HOA but can also occur as an isolated finding **(Fig. 386-2)**.

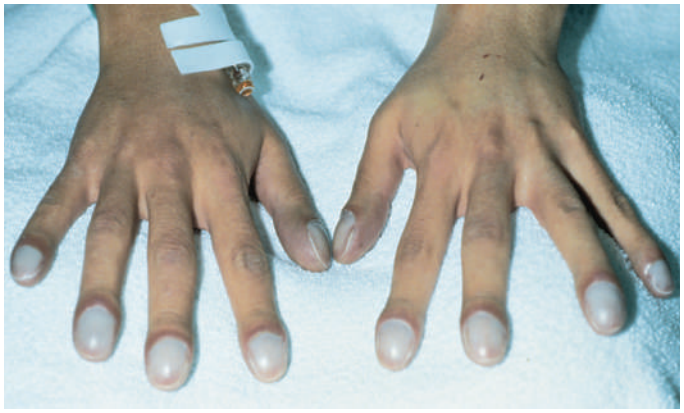

FIGURE 386-2 Clubbing of the fingers. *(Photo contributor Alan B. Storrow, MD.)*

The presence of clubbing in isolation may be congenital or represent either an early stage or one element in the spectrum of HOA. Isolated acquired clubbing has the same clinical significance as clubbing associated with periostitis.

Pathology and Pathophysiology of Acquired HOA In HOA, bone changes in the distal extremities begin as periostitis followed by new bone formation. At this stage, a radiolucent area may be observed between the new periosteal bone and the subjacent cortex. As the process progresses, multiple layers of new bone are deposited and become contiguous with the cortex, with consequent cortical thickening. The outer portion of the bone is laminated in appearance, with an irregular surface. Initially, the process of periosteal new-bone formation involves the proximal and distal diaphyses of the tibia, fibula, radius, and ulna and, less frequently, the femur, humerus, metacarpals, metatarsals, and phalanges. Occasionally, scapulae, clavicles, ribs, and pelvic bones are also affected. Radionuclide studies show pericortical linear uptake along the cortical margins of long bones that may precede any radiographic changes. The distribution of bone manifestations is usually bilateral and symmetric. The soft tissue overlying the distal third of the arms and legs may be thickened. Proliferation of connective tissue occurs in the nail bed and volar pad of digits, giving the distal phalanges a clubbed appearance. Small blood vessels in the clubbed digits are dilated and have thickened walls. In addition, the number of arteriovenous anastomoses is increased.

Many theories have been suggested for the pathogenesis of HOA, but several have been disproved or have not explained the condition's development in all clinical disorders with which it is associated. Current hypotheses have in common the concept that circulating cells or growth factors bypass clearance by the lungs and circulate to the periphery where they stimulate connective tissue proliferation. A role for platelets in the development of HOA has been proposed. Megakaryocytes and large platelet particles present in the venous circulation are fragmented in their passage through normal lung; in patients with cyanotic congenital heart disease or other disorders associated with right-to-left shunts, these large platelet particles bypass the lung and reach the distal extremities, where they interact with endothelial cells. Increased levels of platelet-derived growth factor (PDGF), vascular endothelial growth factor (VEGF), and other factors stimulate the proliferation of connective tissue and periosteum. Elevated plasma levels of von Willebrand factor antigen have been found in patients with both primary and secondary forms of HOA, supporting the concept of endothelial activation or damage.

Clinical Manifestations A form of primary familial HOA, referred to as *pachydermoperiostitis* or *Touraine-Solente-Golé syndrome*, usually begins insidiously at puberty. The disorder is inherited as an autosomal dominant trait with variable expression and is nine times more common among boys than among girls. Approximately one-third of patients have a family history of primary HOA.

Primary HOA is characterized by clubbing, periostitis, and unusual skin features. A small number of patients with this syndrome do not express clubbing. The skin changes and periostitis are prominent features of this syndrome. The skin becomes thickened and coarse. Deep nasolabial folds develop, and the forehead may become furrowed. The skin is often greasy, and there may be excessive sweating of the hands and feet. Patients may also experience acne vulgaris, seborrhea, and folliculitis. In a few patients, the skin over the scalp becomes very thick and corrugated, a feature that has been descriptively termed *cutis verticis gyrata*. The distal extremities, particularly the legs, become markedly thickened as a consequence of the proliferation of new bone and soft tissue. The periostitis usually is not painful, distinguishing it from secondary HOA. Clubbing of the fingers may be extensive, producing large, bulbous deformities and clumsiness. Clubbing also affects the toes. Patients may experience articular and periarticular pain, especially in the ankles and knees, and joint motion may be mildly restricted by periarticular bone overgrowth. Noninflammatory effusions without synovial proliferation occur in the wrists, knees, and ankles. Associated abnormalities observed in patients with primary HOA include hypertrophic gastropathy, bone marrow failure, gynecomastia, and cranial suture defects. In patients with primary HOA, the many symptoms disappear when adulthood is reached.

HOA secondary to an underlying disease occurs more frequently than primary HOA. It accompanies a variety of disorders and may precede clinical features of the associated disorder by months. Clubbing is more frequent than the full syndrome of HOA in patients with associated illnesses. Because clubbing evolves over months and is usually asymptomatic, it is often recognized first by the physician and not the patient. Patients may experience a burning sensation in their fingertips. Clubbing is characterized by widening of the fingertips, enlargement of the distal volar pad, convexity of the nail contour, and the loss of the normal 15° angle between the proximal nail and cuticle. The thickness of the digit at the base of the nail is greater than the thickness at the distal interphalangeal joint. An objective measurement of finger clubbing can be made by determining the diameter at the base of the nail and at the distal interphalangeal joint of all 10 digits. Clubbing is present when the sum of the individual digit ratios is >10. At the bedside, clubbing can be appreciated by having the patient place the dorsal surface of the distal phalanges of the fourth fingers together with the nails opposing each other. Normally, an open area is visible between the bases of the opposing fingernails; when clubbing is present, this open space is no longer visible. The base of the nail feels spongy when compressed, and the nail can be easily rocked on its bed. When clubbing is advanced, the finger may have a drumstick appearance, and the distal interphalangeal joint can be hyperextended. Periosteal involvement in the distal extremities may produce a burning or deep-seated aching pain. The pain, which can be quite incapacitating, is aggravated by dependency and relieved by elevation of the affected limbs. Pressure applied over the distal forearms and legs or gentle percussion of distal long bones like the tibia may be quite painful.

Patients may experience joint pain, most often in the ankles, wrists, and knees. The small joints of the hands are rarely affected. Severe joint or long bone pain may be the presenting symptom of an underlying lung malignancy and may precede the appearance of clubbing. In addition, the progression of HOA tends to be more rapid when associated with malignancies, most notably bronchogenic carcinoma. Noninflammatory but variably painful knee effusions may occur prior to the appearance of clubbing and symptoms of distal periostitis. Unlike primary HOA, secondary HOA does not commonly include excessive sweating and oiliness of the skin or thickening of the facial skin.

HOA occurs in 5–10% of patients with intrathoracic malignancies, the most common being bronchogenic carcinoma and pleural tumors **(Table 386-3)**. Lung metastases infrequently cause HOA. HOA is also seen in patients with intrathoracic infections, including lung abscesses, empyema, and bronchiectasis, but is uncommon in pulmonary tuberculosis. HOA may accompany chronic interstitial pneumonitis, sarcoidosis, and cystic fibrosis. In cystic fibrosis, clubbing is more common than the full syndrome of HOA. Other causes of clubbing include congenital heart disease with right-to-left shunts, bacterial endocarditis, Crohn's disease, ulcerative colitis, sprue, and neoplasms of the esophagus, liver, and small and large bowel. In patients who

TABLE 386-3 Disorders Associated with Hypertrophic Osteoarthropathy

Pulmonary
Bronchogenic carcinoma and other neoplasms
Lung abscesses, empyema, bronchiectasis
Chronic interstitial pneumonitis
Cystic fibrosis
Sarcoidosis
Gastrointestinal
Inflammatory bowel disease
Sprue
Neoplasms: esophagus, liver, bowel
Cardiovascular
Cyanotic congenital heart disease
Subacute bacterial endocarditis
Infected arterial grafts[a]
Aortic aneurysm[b]
Aneurysm of major extremity artery[a]
Patent ductus arteriosus[b]
Arteriovenous fistula of major extremity vessel[a]
Thyroid (thyroid acropachy)
Hyperthyroidism (Graves' disease)

[a]Unilateral involvement. [b]Bilateral lower-extremity involvement.

have congenital heart disease with right-to-left shunts, clubbing alone occurs more often than the full syndrome of HOA.

Unilateral clubbing has been found in association with aneurysms of major extremity arteries, with infected arterial grafts, and with arteriovenous fistulas of brachial vessels. Clubbing of the toes but not the fingers has been associated with an infected abdominal aortic aneurysm and patent ductus arteriosus. Clubbing of a single digit may follow trauma and has been reported in tophaceous gout and sarcoidosis. While clubbing occurs more commonly than the full syndrome in most diseases, periostitis in the absence of clubbing has been observed in the affected limb of patients with infected arterial grafts.

Hyperthyroidism (Graves' disease), treated or untreated, is occasionally associated with clubbing and periostitis of the bones of the hands and feet. This condition is referred to as *thyroid acropachy* and may result from similar factors that stimulate the growth of retrobulbar connective tissue. Periostitis may be asymptomatic and occurs in the midshaft and diaphyseal portion of the metacarpal and phalangeal bones. Significant hand-joint pain may occur; this pain may respond to successful therapy for thyroid dysfunction. The long bones of the extremities are seldom affected. Elevated levels of long-acting thyroid stimulator are found in the sera of these patients.

TREATMENT

Hypertrophic Osteoarthropathy

The treatment of HOA aims to identify the associated disorder and treat it appropriately. The symptoms and signs of HOA may disappear completely with removal of or effective chemotherapy for a tumor or with antibiotic therapy for a chronic pulmonary infection and drainage of the infected site. Vagotomy or percutaneous block of the vagus nerve leads to symptomatic relief in some patients. NSAIDs or analgesics may help control symptoms of HOA.

COMPLEX REGIONAL PAIN SYNDROME

The reflex sympathetic dystrophy syndrome is now referred to as *complex regional pain syndrome, type 1*. This syndrome is characterized by pain, sometimes with swelling, usually of a distal extremity, accompanied by variable vasomotor instability, trophic skin changes, and the rapid development of bony demineralization. Complex regional pain syndrome, including its treatment, is covered in greater detail in **Chap. 451**.

TIETZE SYNDROME AND COSTOCHONDRITIS

Tietze syndrome is manifested by painful swelling of one or more costochondral articulations. The age of onset is usually before 40, and both sexes are affected equally. In most patients, only one joint is involved, usually the second or third costochondral joint. The onset of anterior chest pain may be sudden or gradual. The pain may radiate to the arms or shoulders and is aggravated by sneezing, coughing, deep inspirations, or twisting motions of the chest. The term *costochondritis* is often used interchangeably with *Tietze syndrome*, but some authors restrict the latter term to pain of the costochondral articulations with swelling. Costochondritis is observed in patients aged >40 years; it tends to affect the third, fourth, and fifth costochondral joints, and occurs more often in women. Both syndromes may superficially mimic cardiac or pulmonary-vascular causes of pain. Rheumatoid and psoriatic arthritis, ankylosing spondylitis, and reactive arthritis may involve costochondral joints but are readily distinguished by the presence of other clinical features. Other anterior chest wall pain syndromes are xiphoidalgia and the slipping rib syndrome, which usually involves the tenth or lower ribs and causes pain below the rib cage that can be reproduced or exacerbated by gently hooking the examiner's fingers around the rib and exerting mild caudal traction. Malignancies can invade the ribs, thoracic spine, or chest wall and produce symptoms suggesting Tietze syndrome, as can thoracic herpes zoster. Patients with osteomalacia may have significant rib pain, with or without documented microfractures; this can be suggested by documenting very low vitamin D levels and nuclear bone imaging. Analgesics, anti-inflammatory drugs, and local glucocorticoid injections usually relieve symptoms of costochondritis/Tietze syndrome. Care should be taken to avoid overdiagnosing these syndromes in patients with acute chest pain syndromes; many patients are tender to overly vigorous palpation of the costochondral joints.

MYOFASCIAL PAIN SYNDROME

Myofascial pain syndrome is characterized by multiple areas of localized musculoskeletal pain and tenderness often in association with tender points that can fluctuate in intensity. The pain is deep and aching and may be accompanied by a burning sensation. Myofascial pain may be regional following trauma, overuse, or prolonged static contraction of a muscle or muscle group, which may occur when an individual is reading or writing at a desk or working at a computer. This type of pain may be associated with underlying osteoarthritis of the neck or low back. Pain may be referred from tender points to defined areas distant from the area of original tenderness. Palpation of the tender point reproduces or accentuates the pain. The tender points are usually located in the center of a muscle belly, but they can occur at other sites such as costosternal junctions, the xiphoid process, ligamentous and tendinous insertions, fascia, and fatty areas. Notably tender point sites in muscles are also present in pain-free patients. Myofascial pain most often involves the posterior neck, low back, shoulders, and chest. Chronic pain in the muscles of the posterior neck may involve referral of pain from a tender point in the erector neck muscle or upper trapezius to the head, leading to persistent headaches that may last for days. Tender points in the paraspinal muscles of the low back may refer pain to the buttock. Pain may be referred down the leg from a tender point in the gluteus medius and can mimic sciatica. Lateral hip trochanteric area tenderness may radiate down the lateral leg. A tender point in the infraspinatus muscle may produce local and referred pain over the lateral deltoid and down the outside of the arm into the hand, and a tender point medial to the scapula can radiate to the anterior chest or arm in a distribution that can mimic coronary pain. Injection of a local anesthetic such as 1% lidocaine into the tender point site often results in at least transient pain relief. Another useful technique is first to spray an agent such as ethyl chloride from the tender point toward the area of referred pain and then to stretch the muscle. This maneuver may need to be repeated several times. Massage and application of ultrasound to the affected area also may be beneficial. Attention to posture and resting positions can be helpful in preventing abnormal muscle tension. The prognosis in most patients is good. In some patients, regionally localized myofascial pain syndrome may seem to evolve into or be part of previously unrecognized generalized fibromyalgia **(Chap. 385)**. Abnormal or nonrestorative sleep is a common accompaniment in these patients and should be specifically addressed, as fatigue and "brain fog" are often present.

NEOPLASIAS AND ARTHRITIS

Primary tumors and tumor-like disorders of synovium are uncommon but should be considered in the differential diagnosis of chronic monoarticular joint disease. In addition, metastases to bone and primary bone tumors adjacent to a joint may produce joint symptoms, occasionally with joint effusions.

Pigmented villonodular synovitis (*PVNS*), likely the same process causing tenosynovial giant cell tumors, is characterized by slowly progressive, benign proliferation of tenosynovial tissue. This usually involves a single large joint or tendon. The most common age of onset is in the third decade, and women are affected slightly more often than men. The proliferating tissue usually displays clonal chromosomal translocations, often involving the colony stimulating factor-1 (CSF-1) pathway, which influences the proliferation and maturation of mononuclear cells and macrophages. This provides a target for potential therapy.

The synovium has a brownish color and numerous large, fingerlike villi and nodules. There is marked hyperplasia of synovial cells in the stroma of the villi. Hemosiderin and lipids are found in the cytoplasm of macrophages and in the interstitial tissue. Multinucleated giant cells may be present. The proliferative synovium may behave as a simple mass or as a more diffuse invasive tissue growing into the adjacent cartilage and bone.

The clinical picture of PVNS is characterized by the insidious onset of progressive swelling and pain in affected joints or tendons, most commonly the knee or flexor tendons of the hand. Other commonly affected joints include the hips, ankles, calcaneocuboid joints, elbows, and small joints of the fingers or toes. A multifocal form of PVNS occurs less commonly. Tendon sheaths in the wrist, ankle, or foot may be involved. Symptoms of pain, a catching sensation, and stiffness are often mild and intermittent for years before the patient seeks medical attention. Radiographs may show joint space narrowing, erosions, and subchondral cysts. The diagnosis of PVNS is strongly suggested by gradient echo MRI, which reveals a synovial mass lesion of low signal intensity typical of tissue containing hemosiderin **(Fig. 386-3)**. The joint fluid may contain blood and is often dark red or almost black in color. Lipid-containing macrophages may be present in the fluid.

The treatment for PVNS, if needed, is open or arthroscopic synovectomy. Irradiation of the involved joint has been successful in some patients but may cause a delayed malignant transformation. Treatment directed at inhibiting the CSF-1 pathway activated kinase has demonstrated efficacy.

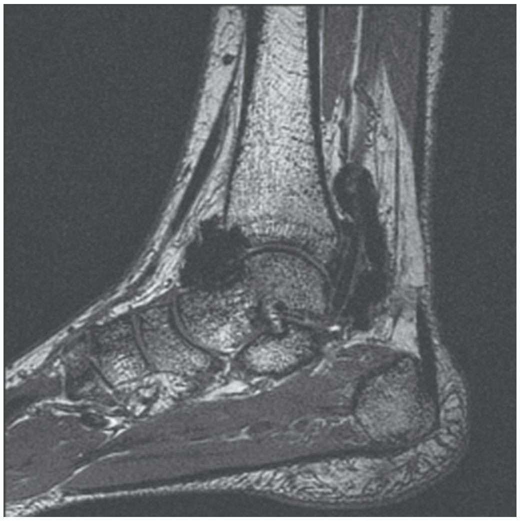

FIGURE 386-3 Pigmented villonodular synovitis. Magnetic resonance imaging gradient echo sagittal image showing a mass that abuts the neck of the talus with marked low signal typical of tissue containing hemosiderin. *(Courtesy of Donald Flemming, MD; with permission.)*

Synovial chondromatosis is a disorder characterized by multiple focal metaplastic growths of normal-appearing cartilage in the synovium or tendon sheath. Segments of cartilage break loose and continue to grow as loose bodies. When calcification and ossification of loose bodies occur, the disorder is referred to as *synovial osteochondromatosis*. The disorder is usually monoarticular and affects young to middle-aged individuals. The knee is most often involved, followed by hip, elbow, and shoulder. Symptoms are pain, swelling, and decreased motion of the joint. Radiographs may show several rounded calcifications within the joint cavity. Treatment is synovectomy; as in PVNS, the growths may recur.

Synovial sarcoma is a malignant neoplasm often found near a large joint of upper or lower extremities, commonly in the legs. It seldom arises within the joint itself. Synovial sarcomas constitute 10% of soft tissue sarcomas. The tumor is believed to arise from primitive mesenchymal tissue that differentiates into epithelial and/or spindle cells. Small foci of calcification may be present in the tumor mass. Synovial sarcoma occurs most often in young male adults. The tumor presents as a slowly growing deep-seated mass near a joint, without much pain. The area of the knee is the most common site, followed by the foot, ankle, elbow, and shoulder. Other primary sites include the buttocks, abdominal wall, retroperitoneum, and mediastinum. The tumor spreads along tissue planes. The most common site of visceral metastasis is the lung. The diagnosis is made by biopsy and must be distinguished from PVNS. Treatment consists of wide resection of the tumor, including adjacent muscle and regional lymph nodes, followed by chemotherapy and radiation therapy. Amputation of the involved distal extremity may be required. Chemotherapy may be beneficial in some patients with metastatic disease. Isolated pulmonary metastases can be surgically removed. The 5-year survival rate with treatment depends on the staging of the tumor, ranging from ~25% to ≥60%. Synovial sarcomas tend to recur locally and metastasize to regional lymph nodes, lungs, and skeleton.

In addition to the rare direct metastases of solid cell tumors to the highly vascular synovium, neoplasia arising from nonarticular organ sites can affect joints in other ways. Acute leukemias in children can mimic juvenile systemic inflammatory arthritis with severe joint pain and fever. In adults, chronic and acute myeloid leukemia rarely infiltrates the synovium. Hairy cell leukemia has a peculiar tendency to cause episodic inflammatory oligoarthritis and tenosynovitis; these episodes are dramatic and mimic acute gout flares. They respond to potent anti-inflammatory therapy with glucocorticoids; with remission of the leukemia, they may abate. Lymphomas, usually of T-cell origin, may also involve the synovium.

Carcinomas can be associated with several paraneoplastic articular syndromes, including HOA (discussed above). Acute palmar fasciitis with polyarthritis is a well-described but rare condition associated with certain cancers, mainly adenocarcinomas. Clinically, this syndrome is abrupt in onset, with pain in the metacarpophalangeal and proximal interphalangeal joints of the hands and rapidly evolving contractures of the fingers due to thickening of the palmar (flexor) tendons. A similar syndrome, although less dramatic in onset, can occur in diabetics. Paraneoplastic arthritis has been described and may occur in several patterns: asymmetric disease predominantly affecting the lower extremity joints and symmetric polyarthritis with hand joint involvement. Tumors are often found after the onset of the arthritis, and many patients have a preceding period of malaise or weight loss. The onset is often acute, and patients tend to be older men. These features should raise the specter of an underlying malignancy (or a viral infection such as hepatitis C) as the cause of the arthritis. In one series, the symptoms resolved with successful therapy for the malignancy and did not recur with relapse of the malignancy. Dermatomyositis and necrotizing myopathy have well-described associations with neoplasms and may include joint pain and arthritis as part of the clinical presentation. Malignancy-associated arthritis may be initially responsive to NSAIDs and to treatment of the primary neoplasm.

Immune checkpoint inhibitors, increasingly used to treat various malignancies, are now well recognized to elicit severe, autoimmune, inflammatory, organ-targeted reactions including myositis, polymyalgia rheumatica, and polyarthritis.

FURTHER READING

Aguilar C et al: Bone and joint disease in sickle cell disease. Hematol Oncol Clin North Am 19:929, 2005.
Botek G et al: Charcot neuroarthropathy: An often overlooked complication of diabetes. Cleve Clin J Med 77:593, 2010.
Dallos T et al: Idiopathic hand osteoarthritis vs haemochromatosis arthropathy: A clinical, functional and radiographic study. Rheumatology 52:910, 2013.
Guggenbuhl P et al: Miscellaneous non-inflammatory musculoskeletal conditions. Haemochromatosis: The bone and the joint. Best Pract Res Clin Rheumatol 25:649, 2011.
Kedar E, Gardner GC: Lipid-associated rheumatologic syndromes. Rheum Dis Clin North Am 39:481, 2013.
Killinger Z et al: Arthropathy in acromegaly. Rheum Dis Clin North Am 36:713, 2010.
Pineda C, Martínez-Lavín M: Hypertrophic osteoarthropathy: What a rheumatologist should know about this uncommon condition. Rheum Dis Clin North Am 39:383, 2013.
Stephan SR et al: Pigmented villonodular synovitis. A comprehensive review and proposed treatment algorithm. JBJS Rev 4:1, 2016.
Vanderhave KL et al: Musculoskeletal care of the hemophiliac patient. J Am Acad Orthop Surg 20:553, 2012.

387

Periarticular Disorders of the Extremities

Carol A. Langford

Periarticular disorders are common musculoskeletal abnormalities that can affect people of all ages. This chapter reviews some of the more frequently encountered periarticular disorders.

BURSITIS

Bursitis is inflammation of a bursa, which is a thin-walled sac lined with synovial tissue. Adults typically have over 150 bursae that function to facilitate movement of tendons and muscles over bony prominences. Excessive frictional forces from overuse, trauma, systemic disease (e.g., rheumatoid arthritis, gout), or infection may cause bursitis. Although any bursa has the ability to become inflamed, certain bursal structures are more prone to developing bursitis because of their location and the mechanical stresses they are subject to. *Subacromial bursitis* (subdeltoid bursitis) is the most common form of bursitis and often accompanies rotator cuff tendinitis. The subacromial bursa, which is contiguous with the subdeltoid bursa, is located between the undersurface of the acromion and the humeral head and is covered by the deltoid muscle. Patients most often experience pain in the anterolateral aspect of the shoulder. As it can develop in the setting of other shoulder disorders, directed examination and imaging are particularly important to rule these out. *Trochanteric bursitis* (also called *greater trochanteric pain syndrome*) is a tendinopathy of the gluteus medius and/or minimus muscles that occurs with or without inflammation of the bursa located around the insertion of the gluteus medius onto the greater trochanter of the femur. Patients experience pain over the lateral aspect of the hip and upper thigh and have tenderness over the posterior aspect of the greater trochanter, often most noticeable when lying on their side. External rotation and resisted abduction of the hip elicit pain, as will direct pressure applied to the bursa. *Olecranon bursitis* involves the posterior elbow, often presenting with visible swelling over the olecranon. In settings where this area becomes acutely inflamed with redness and/or warmth, infection or gout should be excluded by aspirating the bursa and performing a Gram stain and culture on the fluid as well as examining the fluid for urate crystals. *Achilles bursitis* (also called subcutaneous calcaneal bursitis) impacts the bursa located between the skin and the Achilles tendon and is most often caused by overuse and ill-fitting shoes. *Retrocalcaneal bursitis* impacts a deeper bursa that is located between the calcaneus and posterior surface of the Achilles tendon. This bursa facilitates sliding of the Achilles tendon over the calcaneus and can become inflamed as a result of repetitive movement, trauma, or in association with spondyloarthritis, rheumatoid arthritis, or gout. Patients experience pain at the back of the heel and swelling along the medial and/or lateral side of the tendon. *Ischial bursitis* affects the bursa separating the gluteus medius from the ischial tuberosity and typically occurs from prolonged sitting on hard surfaces. *Iliopsoas bursitis* affects the iliopsoas bursa, the largest bursa in the body, which lies between the anterior aspect of the hip joint and the iliopsoas muscle. Bursitis in this location is associated with overuse or an underlying systemic disease process. Patients experience pain over the area, made worse by hip extension and flexion, which must be differentiated from pain originating within the hip joint. *Anserine bursitis* is an inflammation of the bursa located over the medial side of the tibia just below the knee joint under the conjoint tendon of the sartorius, gracilis, and semitendinosus muscles. Pain can be elicited on climbing stairs and by direct pressure over the bursa about 5 cm below the medial knee joint line. *Prepatellar bursitis* occurs in the bursa situated between the patella and overlying skin and is caused by kneeling on hard surfaces. Gout or infection may also occur at this site. Bursitis is typically diagnosed by history and physical examination, but ultrasound may play a useful role in selected instances for diagnosis and directed guidance of glucocorticoid injection. Treatment of bursitis consists of prevention of aggravating factors, rest of the involved part, administration of a topical or systemic nonsteroidal anti-inflammatory drug (NSAID) where appropriate for an individual patient, or local glucocorticoid injection.

ROTATOR CUFF TENDINITIS AND IMPINGEMENT SYNDROME

Tendinitis of the rotator cuff (also called *impingement syndrome*) is the major cause of a painful shoulder. The rotator cuff consists of the tendons of the supraspinatus, infraspinatus, subscapularis, and teres minor muscles, and inserts on the humeral tuberosities. Of the tendons forming the rotator cuff, the supraspinatus tendon is the most often affected, because of its potential for impingement between the humeral head and the undersurface of the anterior third of the acromion and coracoacromial ligament above **(Fig. 387-1)**. The tendon of the infraspinatus and that of the long head of the biceps are less commonly involved. Subacromial bursitis often accompanies this syndrome. Symptoms can appear without a triggering cause or after injury or

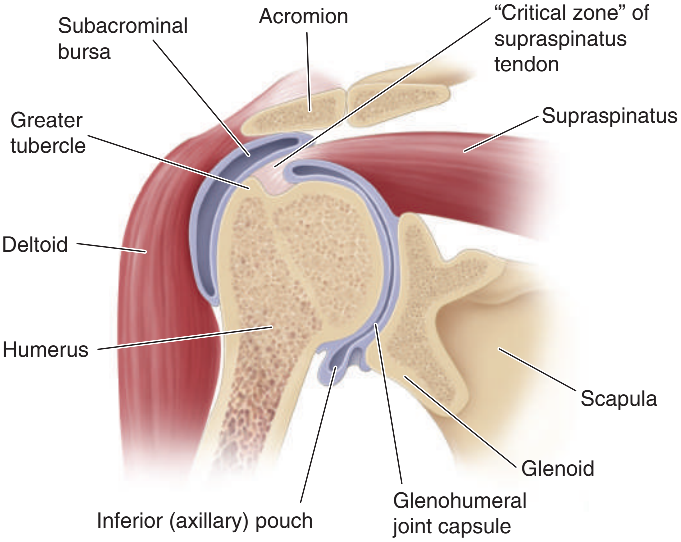
Subacrominal bursa
Acromion
"Critical zone" of supraspinatus tendon
Greater tubercle
Supraspinatus
Deltoid
Humerus
Scapula
Glenoid
Glenohumeral joint capsule
Inferior (axillary) pouch

FIGURE 387-1 Coronal section of the shoulder illustrating the relationships of the glenohumeral joint, the joint capsule, the subacromial bursa, and the rotator cuff (supraspinatus tendon). *(Reproduced with permission from F Kozin, in WJ Koopman [ed]: Arthritis and Allied Conditions, 13th ed, Baltimore, Williams & Wilkins, 1997.)*

overuse, especially with occupations or sports that involve elevation of the arm with forward flexion such as baseball, tennis, or swimming. Work and home ergonomic issues have also increasingly been raised as contributing factors. Patients typically experience a dull aching in the shoulder, which may interfere with sleep. Severe pain occurs when the arm is actively abducted into an overhead position, with the arc between 60° and 120° being especially painful. Tenderness is present over the lateral aspect of the humeral head just below the acromion. NSAIDs, local glucocorticoid injection, and physical therapy may relieve symptoms. Surgical decompression of the subacromial space may be necessary in patients refractory to conservative treatment.

ROTATOR CUFF TEARS

Rotator cuff tears may be acute, subacute, or chronic. Acute tears can occur related to sports, falling on an outstretched arm, or lifting a heavy object. Subacute and chronic tears are more common and develop in middle-aged and older patients, where the cause of the tear is likely multifactorial. Osteoarthritis is a risk factor for rotator cuff tears. Symptoms are pain along with weakness of abduction and external rotation of the shoulder. There is no single examination finding that can diagnose a torn rotator cuff, although this can be suggested by a range of maneuvers that include a painful abduction arc above 90°, failure to control shoulder adduction (drop arm test), or weakness in external rotation. The diagnosis is established by ultrasound or magnetic resonance imaging (MRI). Multiple factors influence management including symptom duration, degree of the tear (partial or full thickness), level of pain and disability, functional status, severity of osteoarthritis if present, and individual patient characteristics. Surgical referral should be pursued in patients with acute or acute-on-chronic tears that are moderate to severe with functional loss. Surgery may also be indicated in patients who fail to respond to conservative measures.

CALCIFIC TENDINITIS

Calcific tendinitis is characterized by deposition of calcium salts, primarily hydroxyapatite, within a tendon. The shoulder is the main site of involvement with calcific deposits in or around the rotator cuff tendons, with the supraspinatus tendon being involved in 63% of patients. The condition usually develops after age 40. Although the mechanism through which calcification is initiated is unknown, the process is felt to go through different phases. It is not felt to be due to overuse or systemic disease, although links have been seen with diabetes and thyroid disorders and a genetic predisposition has been questioned. Calcification within the tendon may evoke acute inflammation, producing sudden and severe shoulder pain, but may also be asymptomatic. Diagnosis of calcific tendinitis can be made by radiograph or ultrasound. Most cases are self-limited and respond to conservative therapy with physical therapy and/or NSAIDs. For patients with refractory pain, ultrasound-guided percutaneous needle aspiration and lavage, extracorporeal shock wave therapy, or surgery has been used.

BICIPITAL TENDINITIS AND RUPTURE

Bicipital tendinitis is produced by friction on the tendon of the long head of the biceps as it passes through the bicipital groove. When the inflammation is acute, patients experience anterior shoulder pain that radiates down the biceps into the forearm. Abduction and external rotation of the arm are painful and limited with tenderness over the bicipital groove. Pain may be elicited along the course of the tendon by resisting supination of the forearm with the elbow at 90° (Yergason's test). Acute rupture of the tendon may occur with vigorous exercise of the arm and is often painful. In a healthy and active patient, it should be repaired surgically as soon as possible after the rupture occurs. Rupture of the tendon in an older person may be associated with little or no pain and is recognized by the presence of persistent swelling of the biceps produced by the retraction of the long head of the biceps. Surgery is usually not necessary in this setting.

DE QUERVAIN'S TENOSYNOVITIS

De Quervain's tenosynovitis is the result of inflammation involving the abductor pollicis longus and the extensor pollicis brevis as these tendons pass through a fibrous sheath at the radial styloid process. The usual cause is repetitive twisting of the wrist. It may also occur in pregnancy and in parents who hold their babies with the thumb outstretched. Patients experience pain on grasping or pinching with their thumb. Swelling and tenderness are often present over the radial styloid process. A positive Finkelstein test can be elicited by having the patient place the thumb in the palm, closing the fingers over it, and then ulnar deviating the wrist, which results in pain over the involved tendon sheath in the area of the radial styloid. Initial treatment consists of wrist splinting and NSAIDs. When severe or refractory, local glucocorticoid injections can be effective.

PATELLAR TENDINITIS

Patellar tendinitis involves the patellar tendon at its attachment to the lower pole of the patella. Patients may experience pain when jumping during sports, going up stairs, or doing deep knee squats. Tenderness is noted on examination over the lower pole of the patella. Treatment consists of rest, icing, and NSAIDs, followed by physical therapy.

DRUG-INDUCED TENDINOPATHIES

With the broadening range of pharmacologic agents, the potential for drug-induced tendinopathies has become increasingly recognized. The drug classes most associated with tendinopathies include quinolones, glucocorticoids, aromatase inhibitors, and statins. Although any tendon can be affected, the tendons of the lower extremities are most often impacted, particularly the Achilles tendon. The pathophysiologic mechanisms responsible for drug-induced tendinopathies remain unknown. Presenting features include pain and potentially swelling over the tendon, although some patients may first present with tendon rupture. Ultrasound and MRI can provide information on tendon structure and integrity in support of the diagnosis. When suspected, the potential agent should be withdrawn and not reintroduced where possible. Tendon ruptures may require surgery.

ILIOTIBIAL BAND SYNDROME

The iliotibial band is a thick connective tissue that runs from the ilium to the tibia. Iliotibial band syndrome most commonly presents with aching or burning pain where the band courses over the lateral femoral condyle of the knee; pain may also radiate up the thigh, toward the hip. Predisposing factors for iliotibial band syndrome include a varus alignment of the knee, distance running, poorly fitted shoes, or running on uneven terrain. Treatment consists of rest, NSAIDs, physical therapy, and addressing modifiable risk factors. Glucocorticoid injection into the area of tenderness can provide relief, but running must be avoided for at least 2 weeks after the injection. Surgical release of the iliotibial band has been helpful in rare patients for whom conservative treatment has failed.

ADHESIVE CAPSULITIS

Adhesive capsulitis ("frozen shoulder") is characterized by pain and restricted movement of the shoulder, usually in the absence of intrinsic shoulder disease. It most often develops in the setting of reduced arm mobility following bursitis or tendinitis of the shoulder, fractures, or recovery from surgery but can occur without an antecedent event. Adhesive capsulitis has been associated with systemic disorders such as diabetes mellitus, chronic pulmonary disease, myocardial infarction, and thyroid disease. Pathologically, the capsule of the shoulder is thickened, and a mild chronic inflammatory infiltrate and fibrosis may be present.

Adhesive capsulitis occurs more commonly in women over age 50. Pain and stiffness usually develop gradually but can rapidly progress. Night pain is often present in the affected shoulder and interferes with sleep. The shoulder is tender to palpation, and both active and passive movements are restricted. The diagnosis is based on physical examination showing a reduced range of motion in the shoulder that is not due to other causes. There are no diagnostic findings by plain radiograph, ultrasound, or MRI, but these can be used to rule out other conditions.

In most patients, the condition improves spontaneously 1–3 years after onset. While pain usually improves, many patients are left with some limitation of shoulder motion. Early mobilization of the arm following a shoulder injury may prevent the development of this disorder. Physical therapy provides the foundation of treatment for adhesive capsulitis. Local glucocorticoid injection and NSAIDs may also provide relief of symptoms. Slow but forceful injection of contrast material into the joint may lyse adhesions and stretch the capsule, resulting in improvement of shoulder motion. Manipulation under anesthesia may be helpful in some patients.

LATERAL EPICONDYLITIS

Lateral epicondylitis, also known as tennis elbow, is a painful condition involving the soft tissue over the lateral aspect of the elbow. The pain originates at or near the site of attachment of the common extensors to the lateral epicondyle and may radiate into the forearm and dorsum of the wrist. The pain is usually linked to work or recreational activities that involve repeated motions of wrist extension and supination against resistance. Most patients with this disorder injure themselves in activities other than tennis, such as pulling weeds, carrying suitcases or briefcases, or using a screwdriver. The injury in tennis usually occurs when hitting a backhand with the elbow flexed. The pain can be reproduced by shaking hands or opening doors.

The treatment is usually rest with avoidance of activities that would aggravate the elbow along with an NSAID. Ultrasound, icing, and friction massage may also help relieve pain. When pain is severe, the elbow is placed in a sling or splinted at 90° of flexion. A one-time local glucocorticoid injection can be considered for severe, acute pain, but repeat injections should be avoided. After 4–6 weeks of rest, the patient should begin physical therapy to strengthen and increase flexibility of the extensor muscles. A forearm band placed 2.5–5.0 cm below the elbow may help to reduce tension on the extensor muscles at their attachment to the lateral epicondyle. The patient should be advised to restrict activities requiring forcible extension and supination of the wrist. The patient may continue to experience mild pain but, with care, can usually avoid the return of debilitating pain. In recalcitrant cases, ultrasound-guided needle tenotomy and injection of platelet-rich plasma have been used. Occasionally, surgical release of the extensor aponeurosis may be necessary.

MEDIAL EPICONDYLITIS

Medial epicondylitis is an overuse syndrome resulting in pain over the medial side of the elbow with radiation into the forearm. The cause of this syndrome is considered to be repetitive resisted motions of wrist flexion and pronation, which lead to microtears at the origin of the pronator teres and forearm flexors, particularly the flexor carpi radialis. This overuse syndrome is usually seen in patients aged >35 years and is much less common than lateral epicondylitis. It occurs most often in work-related repetitive activities and with recreational activities such as swinging a golf club or throwing a baseball. On physical examination, there is tenderness just distal to the medial epicondyle over the origin of the forearm flexors. Pain can be reproduced by resisting wrist flexion and pronation with the elbow extended. Radiographs are usually normal.

The initial treatment of medial epicondylitis is conservative, involving rest, NSAIDs, friction massage, ultrasound, and icing. Some patients may require splinting. For severe pain, a one-time local glucocorticoid injection can be considered. Patients should be instructed to rest for at least 4–6 weeks followed by physical therapy once the pain has subsided. In patients with chronic debilitating medial epicondylitis that remains unresponsive after at least a year of treatment, surgical release of the flexor muscle at its origin may be necessary and is often successful.

PLANTAR FASCIITIS

Plantar fasciitis is a common cause of foot pain in adults. The pain originates at or near the site of the plantar fascia attachment to the medial tuberosity of the calcaneus. Factors that increase the risk of developing plantar fasciitis include obesity, pes planus (flat foot or absence of the foot arch when standing), pes cavus (high-arched foot), limited dorsiflexion of the ankle, prolonged standing, walking on hard surfaces, and faulty shoes. In runners, excessive running and a change to a harder running surface may precipitate plantar fasciitis.

The diagnosis of plantar fasciitis can usually be made on the basis of history and physical examination alone. Patients experience severe pain with the first steps on arising in the morning or following inactivity during the day. The pain usually lessens with weight-bearing activity but then increases with continued activity. Pain is made worse on walking barefoot or going up stairs. On examination, tenderness is elicited on palpation over the inferior heel at the attachment site of the plantar fascia to the calcaneus. Imaging may be indicated when the diagnosis is not clear. Plain radiographs may show heel spurs, which are of little diagnostic significance. Ultrasonography in plantar fasciitis can demonstrate thickening of the fascia and diffuse hypoechogenicity, indicating edema at the attachment of the plantar fascia to the calcaneus. MRI is a sensitive method for detecting plantar fasciitis, but it is usually not required to establish the diagnosis.

Resolution of symptoms occurs within 12 months in >80% of patients with plantar fasciitis. Initial treatment consists of ice, heat, massage, stretching, and eliminating inciting activities. Orthotics provide medial arch support and can be effective. Some patients may benefit from foot strapping or taping or by wearing a night splint designed to keep the ankle in a neutral position. A short course of NSAIDs can be given to patients who do not have contraindications. Local glucocorticoid injections have also been shown to be efficacious but may carry an increased risk for plantar fascia rupture. Plantar fasciotomy is reserved for patients who have failed to improve after at least 6–12 months of conservative treatment.

FURTHER READING

Buchbinder R: Plantar fasciitis. N Engl J Med 350:2159, 2004.

Greis AC et al: Evaluation and nonsurgical management of rotator cuff calcific tendinopathy. Orthop Clin North Am 46:293, 2015.

Harrison AK, Flatow EL: Subacromial impingement syndrome. J Am Acad Orthop Surg 19:701, 2011.

Kirchgesner T et al: Drug-induced tendinopathy: From physiology to clinical applications. Joint Bone Spine 81:485, 2014.

Neviaser AS, Neviaser RJ: Adhesive capsulitis of the shoulder. J Am Acad Orthop Surg 19:536, 2011.

This page intentionally left blank

Endocrinology and Metabolism PART 12

Section 1 Endocrinology

388 Approach to the Patient with Endocrine Disorders

J. Larry Jameson

The management of endocrine disorders has applied principles of precision medicine before the term was commonly used **(Chap. 5)**. A general goal is to maintain or restore homeostasis using precise homone measurements to titrate treatment regimens. Effective patient management requires a broad understanding of intermediary metabolism, reproductive physiology, bone metabolism, and growth. Accordingly, the practice of endocrinology is intimately linked to a conceptual framework for understanding hormone secretion, hormone action, and principles of feedback control **(Chap. 389)**. The endocrine system is evaluated primarily by measuring hormone concentrations, arming the clinician with valuable diagnostic information. Most disorders of the endocrine system are amenable to effective treatment once the correct diagnosis is established. Endocrine deficiency disorders are treated with physiologic hormone replacement; hormone excess conditions, which usually are caused by benign glandular adenomas, are managed by removing tumors surgically or reducing hormone levels medically.

SCOPE OF ENDOCRINOLOGY

Classically, the specialty of endocrinology encompasses the study of glands and the hormones they produce. Over time, the field has expanded because of the discovery of hormones and growth factors produced by the brain, gastrointestinal (GI) tract, musculoskeletal system, and other nonglandular organs. The term *endocrine* was coined by Starling to contrast the actions of hormones secreted internally (*endocrine*) with those secreted externally (*exocrine*) or into a lumen, such as the GI tract. The term *hormone*, derived from a Greek phrase meaning "to set in motion," aptly describes the dynamic actions of hormones as they elicit cellular responses and regulate physiologic processes through feedback mechanisms.

Unlike many other specialties in medicine, it is not possible to define endocrinology strictly along anatomic lines. The classic endocrine glands—pituitary, thyroid, parathyroid, pancreatic islets, adrenals, and gonads—communicate broadly with other organs through the nervous system, hormones, cytokines, and growth factors. In addition to its traditional synaptic functions, the brain produces a vast array of peptide hormones, and this has led to the discipline of neuroendocrinology. Through the production of hypothalamic releasing factors, the central nervous system (CNS) exerts a major regulatory influence over pituitary hormone secretion **(Chap. 390)**. The peripheral nervous system stimulates the adrenal medulla. The immune and endocrine systems are also intimately intertwined. The adrenal hormone cortisol is a powerful immunosuppressant. Cytokines and interleukins (ILs) have profound effects on the functions of the pituitary, adrenal, thyroid, and gonads. Common endocrine diseases such as autoimmune thyroid disease and type 1 diabetes mellitus are caused by dysregulation of immune surveillance and tolerance. Less common diseases such as polyglandular failure, Addison's disease, and lymphocytic hypophysitis also have an immunologic basis. Immune therapies for cancer and various autoimmune diseases can initiate autoimmune endocrine disease as a side effect of treatment.

The interdigitation of endocrinology with physiologic processes in other specialties sometimes blurs the role of hormones. For example, hormones play an important role in maintenance of blood pressure, intravascular volume, and peripheral resistance in the cardiovascular system. Vasoactive substances such as catecholamines, angiotensin II, endothelin, and nitric oxide are involved in dynamic changes of vascular tone in addition to their multiple roles in other tissues. The heart is the principal source of atrial natriuretic peptide, which acts in classic endocrine fashion to induce natriuresis at a distant target organ (the kidney). Erythropoietin, a traditional circulating hormone, is made in the kidney and stimulates erythropoiesis in bone marrow **(Chap. 66)**. The kidney is also integrally involved in the renin-angiotensin axis **(Chap. 398)** and is a primary target of several hormones, including parathyroid hormone (PTH), mineralocorticoids, fibroblast growth factor 23 (FGF23), and vasopressin. The GI tract produces a vast array of peptide hormones, such as glucagon-like peptide 1 (GLP1), cholecystokinin, ghrelin, gastrin, secretin, and vasoactive intestinal peptide, among many others. Carcinoid and islet tumors can secrete excessive amounts of these hormones, leading to specific clinical syndromes **(Chap. 89)**. Many of these GI hormones are also produced in the CNS, where their functions are poorly understood. Adipose tissue produces leptin, which acts centrally to control appetite, along with adiponectin, resistin, and other hormones that regulate metabolism. As hormones such as inhibin, ghrelin, and leptin are discovered, they become integrated into the science and practice of medicine on the basis of their functional roles rather than their tissues of origin.

Characterization of hormone receptors frequently reveals unexpected relationships to factors in nonendocrine disciplines. The growth hormone (GH) and leptin receptors, for example, are members of the cytokine receptor family. The G protein–coupled receptors (GPCRs), which mediate the actions of many peptide hormones, are used in numerous physiologic processes, including vision, smell, and neurotransmission.

PATHOLOGIC MECHANISMS OF ENDOCRINE DISEASE

Endocrine diseases can be divided into three major types of conditions: (1) hormone excess, (2) hormone deficiency, and (3) hormone resistance **(Table 388-1)**.

■ CAUSES OF HORMONE EXCESS

Syndromes of hormone excess can be caused by neoplastic growth of endocrine cells, autoimmune disorders, and excess hormone administration. Benign endocrine tumors, including parathyroid, pituitary, and adrenal adenomas, often retain the capacity to produce hormones, reflecting the fact that these tumors are relatively well differentiated. Many endocrine tumors exhibit subtle defects in their "set points" for feedback regulation. For example, in Cushing's disease, impaired feedback inhibition of adrenocorticotropic hormone (ACTH) secretion is associated with autonomous function. However, the tumor cells are less sensitive to feedback inhibition, as evidenced by ACTH suppression at higher doses of dexamethasone (e.g., high-dose dexamethasone test) **(Chap. 398)**. Similar set point defects are also typical of parathyroid adenomas and autonomously functioning thyroid nodules.

The molecular basis of some endocrine tumors, such as the multiple endocrine neoplasia (MEN) syndromes (MEN1, 2A, 2B), has provided important insights into tumorigenesis **(Chap. 402)**. MEN1 is characterized primarily by the triad of parathyroid, pancreatic islet, and pituitary tumors. MEN2 predisposes to medullary thyroid carcinoma, pheochromocytoma, and hyperparathyroidism. The *MEN1* gene, located on chromosome 11q13, encodes a tumor-suppressor gene, menin. Analogous to the paradigm first described for retinoblastoma, the affected individual inherits a mutant copy of the *MEN1* gene, and tumorigenesis ensues after a somatic "second hit" leads to loss of function of the normal *MEN1* gene (through deletion or point mutations).

In contrast to inactivation of a tumor-suppressor gene, as occurs in MEN1 and most other inherited cancer syndromes, MEN2 is caused by activating mutations in a single allele. In this case, activating mutations of the *RET* protooncogene, which encodes a receptor tyrosine kinase,

TABLE 388-1 Causes of Endocrine Dysfunction

TYPE OF ENDOCRINE DISORDER	EXAMPLES
Hyperfunction	
Neoplastic	
Benign	Pituitary adenomas, hyperparathyroidism, autonomous thyroid or adrenal nodules
Malignant	Adrenal cancer, medullary thyroid cancer, carcinoid
Ectopic	Ectopic ACTH, SIADH secretion
Genetic predisposition	MEN1, MEN2
Autoimmune	Graves' disease
Iatrogenic	Cushing's syndrome, hypoglycemia
Infectious/inflammatory	Subacute thyroiditis
Activating receptor mutations	LH, TSH, Ca^{2+}, PTH receptors, $G_s\alpha$
Hypofunction	
Autoimmune	Hashimoto's thyroiditis, type 1 diabetes mellitus, Addison's disease, polyglandular failure
Iatrogenic	Radiation-induced hypopituitarism, hypothyroidism, surgical
Infectious/inflammatory	Adrenal insufficiency, hypothalamic sarcoidosis
Hormone mutations	GH, LHβ, FSHβ, vasopressin
Enzyme defects	21-Hydroxylase deficiency
Developmental defects	Kallmann's syndrome, Turner's syndrome, transcription factors
Nutritional/vitamin deficiency	Vitamin D deficiency, iodine deficiency
Hemorrhage/infarction	Sheehan's syndrome, adrenal insufficiency
Hormone Resistance	
Receptor mutations	
Membrane	GH, vasopressin, LH, FSH, ACTH, GnRH, GHRH, PTH, leptin, Ca^{2+}
Nuclear	AR, TR, VDR, ER, GR, PPARγ
Signaling pathway mutations	Albright's hereditary osteodystrophy
Postreceptor	Type 2 diabetes mellitus, leptin resistance

Abbreviations: ACTH, adrenocorticotropic hormone; AR, androgen receptor; ER, estrogen receptor; FSH, follicle-stimulating hormone; GH, growth hormone; GHRH, growth hormone–releasing hormone; GnRH, gonadotropin-releasing hormone; GR, glucocorticoid receptor; LH, luteinizing hormone; MEN, multiple endocrine neoplasia; PPAR, peroxisome proliferator activated receptor; PTH, parathyroid hormone; SIADH, syndrome of inappropriate antidiuretic hormone; TR, thyroid hormone receptor; TSH, thyroid-stimulating hormone; VDR, vitamin D receptor.

lead to thyroid C-cell hyperplasia in childhood before the development of medullary thyroid carcinoma. Elucidation of this pathogenic mechanism has allowed early genetic screening for *RET* mutations in individuals at risk for MEN2, permitting identification of those who may benefit from prophylactic thyroidectomy and biochemical screening for pheochromocytoma and hyperparathyroidism.

Mutations that activate hormone receptor signaling have been identified in several GPCRs. For example, activating mutations of the luteinizing hormone (LH) receptor cause a dominantly transmitted form of male-limited precocious puberty, reflecting premature stimulation of testosterone synthesis in Leydig cells **(Chap. 403)**. Activating mutations in these GPCRs are located predominantly in the transmembrane domains and induce receptor coupling to $G_s\alpha$ even in the absence of hormone. Consequently, adenylate cyclase is activated, and cyclic adenosine monophosphate (AMP) levels increase in a manner that mimics hormone action. A similar phenomenon results from activating mutations in $G_s\alpha$. When these mutations occur early in development, they cause McCune-Albright syndrome. When they occur only in somatotropes, the activating $G_s\alpha$ mutations cause GH-secreting tumors and acromegaly **(Chap. 392)**.

In autoimmune Graves' disease, antibody interactions with the thyroid-stimulating hormone (TSH) receptor mimic TSH action, leading to hormone overproduction **(Chap. 394)**. Analogous to the effects of activating mutations of the TSH receptor, these stimulating autoantibodies induce conformational changes in the TSH receptor that release it from a constrained state, thereby triggering receptor coupling to G proteins, and signaling induces excess thyroid hormone synthesis and secretion.

CAUSES OF HORMONE DEFICIENCY

Most examples of hormone deficiency states can be attributed to glandular destruction caused by autoimmunity, surgery, infection, inflammation, infarction, hemorrhage, or tumor infiltration (Table 388-1). Autoimmune damage to the thyroid gland (Hashimoto's thyroiditis) and pancreatic islet β cells (type 1 diabetes mellitus) are examples of relatively common endocrine diseases. Mutations in a number of hormones, hormone receptors, transcription factors, enzymes, and channels can also lead to hormone deficiencies.

HORMONE RESISTANCE

Most severe hormone resistance syndromes are due to inherited defects in membrane receptors, nuclear receptors, or the pathways that transduce receptor signals. These disorders are characterized by defective hormone action despite the presence of increased hormone levels. In complete androgen resistance, for example, mutations in the androgen receptor result in a female phenotypic appearance in genetic (XY) males, even though LH and testosterone levels are increased **(Chap. 402)**. In addition to these relatively rare genetic disorders, more common acquired forms of functional hormone resistance include insulin resistance in type 2 diabetes mellitus, leptin resistance in obesity, and GH resistance in catabolic states. The pathogenesis of functional resistance involves receptor downregulation and postreceptor desensitization of signaling pathways; functional forms of resistance are generally reversible.

CLINICAL EVALUATION OF ENDOCRINE DISORDERS

Because most glands are relatively inaccessible, the physical examination usually focuses on the manifestations of hormone excess or deficiency as well as direct examination of palpable glands, such as the thyroid and gonads. For these reasons, it is important to evaluate patients in the context of their presenting symptoms, review of systems, family and social history, and exposure to medications that may affect the endocrine system. Astute clinical skills are required to detect subtle symptoms and signs suggestive of underlying endocrine disease. For example, a patient with Cushing's syndrome may manifest specific findings, such as central fat redistribution, skin striae, and proximal muscle weakness, in addition to features seen commonly in the general population, such as obesity, plethora, hypertension, and glucose intolerance. Similarly, the insidious onset of hypothyroidism—with fatigue, mental slowing, dry skin, and other features—can be difficult to distinguish from similar, nonspecific findings in the general population. Clinical judgment that is based on knowledge of disease prevalence and pathophysiology is required to decide when to embark on more extensive evaluation of these disorders. Laboratory testing plays an essential role in endocrinology by allowing quantitative assessment of hormone levels and dynamics. Radiologic imaging tests such as computed tomography (CT) scan, magnetic resonance imaging (MRI),

TABLE 388-2 Examples of Prevalent Endocrine and Metabolic Disorders in the Adult

DISORDER	APPROXIMATE PREVALENCE IN ADULTS[a]	SCREENING/TESTING RECOMMENDATIONS[b]	CHAPTER(S)
Obesity	40% Obese, BMI ≥30 70% Overweight, BMI ≥25	Calculate BMI Measure waist circumference Exclude secondary causes Consider comorbid complications	414
Type 2 diabetes mellitus	>10%	Beginning at age 45, screen every 3 years, or earlier in high-risk groups: FPG >126 mg/dL Random plasma glucose >200 mg/dL An elevated HbA_{1c} Consider comorbid complications	415
Hyperlipidemia	20–25%	Cholesterol screening at least every 5 years; more often in high-risk groups Lipoprotein analysis (LDL, HDL) for increased cholesterol, CAD, diabetes Consider secondary causes	419
Metabolic syndrome	35%	Measure waist circumference, FPG, BP, lipids	420
Hypothyroidism	5–10%, women 0.5–2%, men	TSH; confirm with free T_4	396
Graves' disease	1–3%, women 0.1%, men	TSH, free T_4	395
Thyroid nodules and neoplasia	2–5% palpable >25% by ultrasound	Physical examination or ultrasound of thyroid Fine-needle aspiration biopsy	397
Osteoporosis	5–10%, women 2–5%, men	Bone mineral density measurements in women >65 years or in postmenopausal women or men at risk Exclude secondary causes	423
Hyperparathyroidism	0.1–0.5%, women > men	Serum calcium PTH, if calcium is elevated Assess comorbid conditions	422
Infertility	10%, couples	Investigate both members of couple Semen analysis in male Assess ovulatory cycles in female Specific tests as indicated	403,404
Polycystic ovarian syndrome	5–10%, women	Free testosterone, DHEAS Consider comorbid conditions	404
Hirsutism	5–10%	Free testosterone, DHEAS Exclude secondary causes Additional tests as indicated	406
Menopause	Median age, 51	FSH	407
Hyperprolactinemia	15% in women with amenorrhea or galactorrhea	PRL level MRI, if not medication-related	392
Erectile dysfunction	10–25%	Careful history, PRL, testosterone Consider secondary causes (e.g., diabetes)	409
Hypogonadism, male	1–2%	Testosterone, LH	403
Gynecomastia	15%	Often, no tests are indicated Consider Klinefelter's syndrome Consider medications, hypogonadism, liver disease	403
Klinefelter's syndrome	0.2%, men	Karyotype Testosterone	402
Vitamin D deficiency	10%	Measure serum 25-OH vitamin D Consider secondary causes	421
Turner's syndrome	0.03%, women	Karyotype Consider comorbid conditions	402

[a] The prevalence of most disorders varies among ethnic groups and with aging. Data based primarily on U.S. population. [b] See individual chapters for additional information on evaluation and treatment. Early testing is indicated in patients with signs and symptoms of disease and in those at increased risk.

Abbreviations: BMI, body mass index; BP, blood pressure; CAD, coronary artery disease; DHEAS, dehydroepiandrosterone; FPG, fasting plasma glucose; FSH, follicle-stimulating hormone; HbA_{1C}, hemoglobin A_{1C}; HDL, high-density lipoprotein; LDL, low-density lipoprotein; LH, luteinizing hormone; MRI, magnetic resonance imaging; PRL, prolactin; PTH, parathyroid hormone; TSH, thyroid-stimulating hormone.

ultrasound, and thyroid scan are also used for the diagnosis of endocrine disorders. However, these tests generally are employed only after a hormonal abnormality has been established by biochemical testing.

HORMONE MEASUREMENTS AND ENDOCRINE TESTING

Immunoassays are the most important diagnostic tool in endocrinology, as they allow sensitive, specific, and quantitative determination of steady-state and dynamic changes in hormone concentrations. Immunoassays use antibodies to detect specific hormones. For many peptide hormones, these measurements are now configured to use two different antibodies to increase binding affinity and specificity. There are many variations of these assays; a common format involves using one antibody to capture the antigen (hormone) onto an immobilized surface and a second antibody, coupled to a chemiluminescent (immunochemiluminescent assay [ICMA]) or radioactive (immunoradiometric assay [IRMA]) signal, to detect the antigen. These assays are sensitive enough to detect plasma hormone concentrations in the picomolar to nanomolar range, and they can readily distinguish structurally related proteins, such as PTH from PTH-related peptide (PTHrP). A variety of other techniques are used to measure specific hormones, including mass spectroscopy, various forms of chromatography, and enzymatic methods; bioassays are now used rarely. Mass spectroscopy is increasingly being used given its ability to quantitatively measure large numbers of peptides or steroids simultaneously.

Most hormone measurements are based on plasma or serum samples. However, urinary hormone determinations remain useful for the evaluation of some conditions. Urinary collections over 24 h provide an integrated assessment of the production of a hormone or metabolite, many of which vary during the day. It is important to ensure complete collections of 24-h urine samples; simultaneous measurement of creatinine provides an internal control for the adequacy of collection and can be used to normalize some hormone measurements. A 24-h urine-free cortisol measurement largely reflects the amount of unbound cortisol, thus providing a reasonable index of biologically available hormone. Other commonly used urine determinations include 17-hydroxycorticosteroids, 17-ketosteroids, vanillylmandelic acid, metanephrine, catecholamines, 5-hydroxyindoleacetic acid, and calcium.

The value of quantitative hormone measurements lies in their correct interpretation in a clinical context. The normal range for most hormones is relatively broad, often varying by a factor of two- to tenfold. The wide normal range reflects the effects of binding proteins as well as circadian rhythms and other physiologic variables. The normal ranges for many hormones are sex- and age-specific. Thus, using the correct normative database is an essential part of interpreting hormone tests. The pulsatile nature of hormones and factors that can affect their secretion, such as sleep, meals, and medications, must also be considered. Cortisol values increase fivefold between midnight and dawn; reproductive hormone levels vary dramatically during the female menstrual cycle.

For many endocrine systems, much information can be gained from basal hormone testing, particularly when different components of an endocrine axis are assessed simultaneously. For example, low testosterone and elevated LH levels suggest a primary gonadal disease, whereas a hypothalamic-pituitary disorder is likely if both LH and testosterone are low. Because TSH is a sensitive indicator of thyroid function, it is generally recommended as a first-line test for thyroid disorders. An elevated TSH level is almost always the result of primary hypothyroidism, whereas a low TSH is most often caused by thyrotoxicosis. These predictions can be confirmed by determining the free thyroxine level. In the less common circumstance when free thyroxine and TSH are both low, it is important to consider secondary hypopituitarism caused by hypothalamic-pituitary disease. Elevated calcium and PTH levels suggest hyperparathyroidism, whereas PTH is suppressed in hypercalcemia caused by malignancy or granulomatous diseases. A suppressed ACTH in the setting of hypercortisolemia, or increased urine free cortisol, is seen with hyperfunctioning adrenal adenomas.

It is not uncommon, however, for baseline hormone levels associated with pathologic endocrine conditions to overlap with the normal range. In this circumstance, dynamic testing is useful to separate the two groups further. There are a multitude of dynamic endocrine tests, but all are based on principles of feedback regulation, and most responses can be rationalized based on principles that govern the regulation of endocrine axes. *Suppression tests* are used in the setting of suspected endocrine hyperfunction. An example is the dexamethasone suppression test used to evaluate Cushing's syndrome **(Chaps. 392 and 398)**. *Stimulation tests* generally are used to assess endocrine hypofunction. The ACTH stimulation test, for example, is used to assess the adrenal gland response in patients with suspected adrenal insufficiency. Other stimulation tests use hypothalamic-releasing factors such as corticotropin-releasing hormone (CRH) and growth hormone–releasing hormone (GHRH) to evaluate pituitary hormone reserve **(Chap. 392)**. Insulin-induced hypoglycemia evokes pituitary ACTH and GH responses. Stimulation tests based on reduction or inhibition of endogenous hormones are now used infrequently. Examples include metyrapone inhibition of cortisol synthesis and clomiphene inhibition of estrogen feedback.

SCREENING AND ASSESSMENT OF COMMON ENDOCRINE DISORDERS

Many endocrine disorders are prevalent in the adult population **(Table 388-2)** and can be diagnosed and managed by general internists, family practitioners, or other primary health care providers. The high prevalence and clinical impact of certain endocrine diseases justify vigilance for features of these disorders during routine physical examinations; laboratory screening is indicated in selected high-risk populations.

FURTHER READING

Endocrine Society: The Endocrine Society Clinical Practice Guidelines. Available from *https://www.endocrine.org/clinical-practice-guidelines*.

Loriaux DL: *A Biographical History of Endocrinology*. Hoboken, Wiley Blackwell, 2016.

Robertson RP (ed): *DeGroot's Endocrinology: Adult and Pediatric*, 8th ed. Philadelphia, Elsevier, 2023.

389

Mechanisms of Hormone Action

J. Larry Jameson

The endocrine system, composed of various glands and the hormones they produce, regulates growth, metabolism, homeostasis, and reproduction. Because hormones circulate and act via receptors in target tissues, they serve to coordinate physiologic responses to external or internal cues. For example, the light-dark cycle, sensed through the visual system, modulates hypothalamic corticotropin-releasing hormone (CRH), which increases pituitary adrenocorticotropin hormone (ACTH) production, leading to increased adrenal cortisol production before the time of waking in the morning. Increased cortisol, in turn, circulates throughout the body, acting via the nuclear glucocorticoid receptor, to activate numerous genetic programs that influence metabolism, the cardiovascular system, behavior, and the immune system. This chapter provides an overview of the different types of hormones and how they function at the cellular level to control myriad physiologic processes.

CLASSES OF HORMONES

Hormones can be divided into five major types: (1) *amino acid derivatives* such as dopamine, catecholamine, and thyroid hormone; (2) *small neuropeptides* such as gonadotropin-releasing hormone (GnRH), thyrotropin-releasing hormone (TRH), somatostatin, and vasopressin; (3) *large proteins* such as insulin, luteinizing hormone (LH), and parathyroid hormone (PTH); (4) *steroid hormones* such as cortisol and estrogen that are synthesized from cholesterol-based precursors; and (5) *vitamin derivatives* such as retinoids (vitamin A) and vitamin D. A variety of *peptide growth factors*, such as insulin-like growth factor 1 (IGF1), share actions with hormones but often act more locally. As a rule, amino acid derivatives and peptide hormones interact with cell-surface membrane receptors. Steroids, thyroid hormones, vitamin D, and retinoids are lipid-soluble and bind to intracellular nuclear receptors, although many also interact with membrane receptors or intracellular signaling proteins as well.

HORMONE AND RECEPTOR FAMILIES

Hormones and receptors can be grouped into families, reflecting structural similarities and evolutionary origins **(Table 389-1)**. The evolution of these families generates diverse but highly selective pathways of hormone action. Understanding these relationships is useful to extrapolate structural and mechanistic insights gleaned from one hormone or receptor to other family members.

The glycoprotein hormone family, consisting of thyroid-stimulating hormone (TSH), follicle-stimulating hormone (FSH), LH, and human chorionic gonadotropin (hCG), illustrates many features of evolutionarily related hormones. The glycoprotein hormones are heterodimers that share the α subunit in common; the β subunits are distinct and confer specific biologic actions. The overall three-dimensional architecture of the β subunits is similar, reflecting the locations of conserved disulfide bonds that constrain protein conformation. Evolutionary analysis suggests that the β-subunit genes arose from a common ancestral gene through gene duplication and divergence to evolve new biologic functions.

As hormone families expand and diverge, their receptors have co-evolved to create new biologic functions. Related G protein–coupled receptors (GPCRs), for example, have evolved for each of the glycoprotein hormones. These receptors are also structurally similar, and each is coupled predominantly to the $G_s\alpha$ signaling pathway. Because of co-evolution with respective hormones to achieve specificity, there is minimal overlap of hormone binding. For example, TSH binds with high specificity to the TSH receptor but interacts minimally with the LH or FSH receptors. Nonetheless, there can be subtle physiologic consequences of hormone cross-reactivity with other receptors. Very high levels of hCG during pregnancy weakly stimulate the TSH receptor and increase thyroid hormone levels, resulting in feedback inhibition and a compensatory decrease in TSH.

TABLE 389-1 Examples of Membrane Receptor Families and Signaling Pathways

RECEPTORS	EFFECTORS	SIGNALING PATHWAYS
G Protein–Coupled Seven-Transmembrane Receptor (GPCR)		
LH, FSH, TSH, β-adrenergic	$G_s\alpha$, adenylate cyclase	Stimulation of cyclic AMP production, protein kinase A
Glucagon, PTH, PTHrP, ACTH, MSH, GHRH, CRH	Ca^{2+} channels	Calmodulin, Ca^{2+}-dependent kinases
Somatostatin, α-adrenergic	$G_i\alpha$	Inhibition of cyclic AMP production Activation of K^+, Ca^{2+} channels
TRH, GnRH	G_q, G_{11}	Phospholipase C, diacylglycerol, IP_3, protein kinase C, voltage-dependent Ca^{2+} channels
Receptor Tyrosine Kinase		
Insulin, IGF-I	Tyrosine kinases, IRS	MAP kinases, PI 3-kinase; AKT
Cytokine Receptor–Linked Kinase		
GH, PRL	JAK, tyrosine kinases	STAT, MAP kinase, PI 3-kinase, IRS-1
Serine Kinase		
Activin, TGF-β, MIS	Serine kinase	Smads

Abbreviations: IP_3, inositol triphosphate; IRS, insulin receptor substrates; MAP, mitogen-activated protein; MSH, melanocyte-stimulating hormone; PI, phosphatidylinositol; RSK, ribosomal S6 kinase; TGF-β, transforming growth factor β. For all other abbreviations, see text. Note that most receptors interact with multiple effectors and activate networks of signaling pathways.

IGF1 and IGF2 have structural similarities that are most apparent when precursor forms of the proteins are compared. In contrast to the high degree of specificity seen with the glycoprotein hormones, there is moderate cross-talk among the members of the insulin/IGF family. High concentrations of an IGF2 precursor produced by certain tumors (e.g., sarcomas) can cause hypoglycemia, partly because of binding to insulin and IGF1 receptors. High concentrations of insulin also bind to the IGF1 receptor, accounting for some of the clinical manifestations seen in conditions with chronic hyperinsulinemia.

Another important example of receptor cross-talk is seen with PTH and parathyroid hormone–related peptide (PTHrP) **(Chap. 422)**. PTH is produced by the parathyroid glands, whereas PTHrP is expressed at high levels during development and by a variety of tumors **(Chap. 98)**. These hormones have amino acid sequence similarity, particularly in their amino-terminal regions. Both hormones bind to the PTH1R receptor that is expressed in bone and kidney. Excessive production of either hormone results in hypercalcemia and hyperphosphatemia, making it difficult to distinguish hyperparathyroidism from hypercalcemia of malignancy solely on the basis of serum chemistries. However, sensitive and specific assays for PTH and PTHrP now allow these disorders to be distinguished.

Based on their specificities for DNA-binding sites, the nuclear receptor family can be subdivided into type 1 receptors (glucocorticoid receptor, mineralocorticoid receptor, androgen receptor, estrogen receptor, progesterone receptor) that bind steroids and type 2 receptors (thyroid hormone receptor, vitamin D receptor, retinoic acid receptor, peroxisome proliferator activated receptor) that bind thyroid hormone, vitamin D, retinoic acid, or lipid derivatives, respectively. Certain functional domains in nuclear receptors, such as the zinc finger DNA-binding domains, are highly conserved. However, selective amino acid differences within this domain confer DNA sequence specificity. The hormone-binding domains are more variable, providing great diversity in the array of small molecules that bind to different nuclear receptors. With few exceptions, hormone binding is highly specific for a single type of nuclear receptor. One exception involves the glucocorticoid and mineralocorticoid receptors. Because the mineralocorticoid receptor also binds glucocorticoids with high affinity, an enzyme (11β-hydroxysteroid dehydrogenase) in renal tubular cells inactivates glucocorticoids, allowing selective renal responses to mineralocorticoids such as aldosterone. However, when very high glucocorticoid concentrations occur, as in Cushing's syndrome, the glucocorticoid degradation pathway becomes saturated, allowing excessive cortisol levels to bind mineralocorticoid receptors leading to sodium retention and potassium wasting. This phenomenon is particularly pronounced in ectopic ACTH syndromes **(Chap. 398)**. Another example of relaxed nuclear receptor specificity involves the estrogen receptor, which can bind an array of compounds, some of which have little apparent structural similarity to the high-affinity ligand estradiol. This feature of the estrogen receptor makes it susceptible to activation by "environmental estrogens" such as resveratrol, octylphenol, and many other aromatic hydrocarbons. However, this lack of specificity provides an opportunity to synthesize clinically useful antagonists (e.g., tamoxifen) and selective estrogen response modulators (SERMs) such as raloxifene. These compounds generate distinct estrogen receptor conformations that alter receptor interactions with components of the transcription machinery (see below), thereby conferring their unique actions.

HORMONE SYNTHESIS AND PROCESSING

The synthesis of peptide hormones and their receptors occurs through a classic pathway of gene expression: transcription → mRNA → protein → posttranslational protein processing → intracellular sorting, followed by membrane integration or secretion.

Many hormones are embedded within larger precursor polypeptides that are proteolytically processed to yield the biologically active hormone. Examples include proopiomelanocortin (POMC) → ACTH; proglucagon → glucagon; proinsulin → insulin; and pro-PTH → PTH, among others. In many cases, such as POMC and proglucagon, these precursors generate multiple biologically active peptides. For example, proglucagon generates glucagon, as well as glucagon-like peptide 1 (GLP1), among other peptides. It is provocative that hormone precursors are typically inactive, presumably adding an additional level of control through peptide processing. Prohormone conversion occurs not only for peptide hormones but also for certain steroids (testosterone → dihydrotestosterone) and thyroid hormone ($T_4 \rightarrow T_3$).

Peptide precursor processing is intimately linked to intracellular sorting pathways that transport proteins to appropriate vesicles and enzymes, resulting in specific cleavage steps, followed by protein folding and translocation to secretory vesicles. Hormones destined for secretion are translocated across the endoplasmic reticulum guided by an amino-terminal signal sequence that subsequently is cleaved. Cell-surface receptors are inserted into the membrane via short segments of hydrophobic amino acids that remain embedded within the lipid bilayer. During translocation through the Golgi and endoplasmic reticulum, hormones and receptors are subject to a variety of post-translational modifications, such as glycosylation and phosphorylation, which can alter protein conformation, modifying circulating half-life and biological activity.

Synthesis of most steroid hormones is based on modifications of the precursor, cholesterol. Multiple regulated enzymatic steps are required for the synthesis of testosterone **(Chap. 403)**, estradiol **(Chap. 404)**, cortisol **(Chap. 398)**, and vitamin D **(Chap. 421)**. This large number of synthetic steps predisposes to multiple genetic and acquired disorders of steroidogenesis.

Endocrine genes contain regulatory DNA elements similar to those found in many other genes, but their exquisite control by hormones reflects the presence of specific hormone response elements. For example, the TSH genes are repressed directly by thyroid hormones acting through the thyroid hormone receptor (TR), a member of the nuclear receptor family. Steroidogenic enzyme gene expression requires specific transcription factors, such as steroidogenic factor 1 (SF1), acting in conjunction with signals transmitted by trophic hormones (e.g., ACTH or LH). Once activated, SF1 functions as a master regulator, inducing a large array of genes required for steroidogenic and metabolic pathways required for steroid synthesis. For some hormones, substantial regulation occurs at the level of translational efficiency. Insulin biosynthesis, although it requires ongoing gene transcription, is regulated primarily at the translational and secretory levels in response to the levels of glucose or amino acids.

HORMONE SECRETION, TRANSPORT, AND DEGRADATION

The circulating level of a hormone is determined by its rate of secretion and its half-life. After protein processing, peptide hormones (e.g., GnRH, insulin, growth hormone [GH]) are stored in secretory granules. As these granules mature, they are poised beneath the plasma membrane for imminent release into the circulation. In most instances, the stimulus for hormone secretion is a releasing factor or neural signal that induces rapid changes in voltage-gated channel activity or intracellular calcium concentrations, leading to secretory granule fusion with the plasma membrane and release of its contents into the extracellular environment and bloodstream. Steroid hormones, in contrast, diffuse into the circulation as they are synthesized. Thus, their secretory rates are closely aligned with rates of synthesis. For example, ACTH and LH induce steroidogenesis by stimulating the activity of the *st*eroidogenic *a*cute *r*egulatory (StAR) protein, which transports cholesterol into the mitochondrion. These hormones also induce other rate-limiting enzymatic steps (e.g., cholesterol side-chain cleavage enzyme, CYP11A1) in specific steroidogenic pathways.

Hormone transport and degradation dictate the rapidity with which a hormonal signal decays. Some hormone signals are evanescent (e.g., somatostatin), whereas others are longer-lived (e.g., TSH). Because somatostatin exerts effects in virtually every tissue, a short half-life allows its concentrations and actions to be controlled locally. Structural modifications that impair somatostatin degradation have been useful for generating long-acting therapeutic analogues such as octreotide **(Chap. 392)**. In contrast, the actions of TSH are highly specific for the thyroid gland. Its prolonged half-life generates relatively constant serum levels even though TSH is secreted in discrete pulses.

An understanding of circulating hormone half-life is important for achieving physiologic hormone replacement, as the frequency of dosing and the time required to reach steady state are intimately linked to rates of hormone decay. T_4, for example, has a circulating half-life of 7 days. Consequently, >1 month is required to reach a new steady state, and single daily doses are sufficient to achieve constant hormone levels. T_3, in contrast, has a half-life of 1 day. Its administration is associated with more dynamic serum levels, and it must be administered two to three times per day. Similarly, synthetic glucocorticoids vary widely in their half-lives; those with longer half-lives (e.g., dexamethasone) are associated with greater suppression of the hypothalamic-pituitary-adrenal (HPA) axis. Most protein hormones (e.g., ACTH, GH, prolactin [PRL], PTH, LH) have relatively short half-lives (<20 min), leading to sharp peaks of secretion and decay. The only accurate way to profile the pulse frequency and amplitude of these hormones is to measure levels in frequently sampled blood (every 10 min or less) over long durations (8–24 h). Because this is not practical in a clinical setting, an alternative strategy is to pool three to four blood samples drawn at about 30-min intervals or interpret the results in the context of a relatively wide normal range. Rapid hormone decay is useful in certain clinical settings. For example, the short half-life of PTH allows the use of intraoperative PTH levels to confirm successful removal of a parathyroid adenoma. This is particularly valuable diagnostically when there is a possibility of multicentric disease or parathyroid hyperplasia, as occurs with multiple endocrine neoplasia (MEN) or renal insufficiency.

Many hormones circulate in association with serum-binding proteins. Examples include (1) T_4 and T_3 binding to thyroxine-binding globulin (TBG), albumin, and thyroxine-binding prealbumin (TBPA); (2) cortisol binding to cortisol-binding globulin (CBG); (3) androgen and estrogen binding to sex hormone–binding globulin (SHBG); (4) IGF1 and IGF2 binding to multiple IGF-binding proteins (IGFBPs); (5) GH interactions with GH-binding protein (GHBP), a circulating fragment of the GH receptor extracellular domain; and (6) activin binding to follistatin. These interactions provide a hormone reservoir, prevent otherwise rapid degradation of unbound hormones, restrict hormone access to certain sites (e.g., IGFBPs), and modulate the levels of unbound, or "free," hormone concentrations. Although a variety of binding protein abnormalities have been identified, most have little clinical consequence aside from creating diagnostic problems. For example, TBG deficiency can reduce total thyroid hormone levels greatly, but the free concentrations of T_4 and T_3 remain normal. Liver disease and certain medications can also influence binding protein levels (e.g., estrogen increases TBG) or cause displacement of hormones from binding proteins (e.g., salsalate displaces T_4 from TBG). In general, only unbound hormone is available to interact with receptors and thus elicit a biologic response. Short-term perturbations in binding proteins change the free hormone concentration, which in turn induces compensatory adaptations through feedback loops. SHBG changes in women are an exception to this self-correcting mechanism. When SHBG decreases because of insulin resistance or androgen excess, the unbound testosterone concentration is increased, potentially contributing to hirsutism in women with polycystic ovary syndrome (PCOS) **(Chap. 406)**. The increased unbound testosterone level does not result in an adequate compensatory feedback correction because estrogen, not testosterone, is the primary regulator of the reproductive axis.

An additional exception to the unbound hormone hypothesis involves megalin, a member of the low-density lipoprotein (LDL) receptor family that serves as an endocytotic receptor for thyroglobulin, carrier-bound vitamins A and D, and SHBG-bound androgens and estrogens. After internalization, the carrier proteins are degraded in lysosomes and release their bound ligands within the cells. Other membrane transporters have also been identified for thyroid hormones.

Hormone degradation can be an important mechanism for regulating concentrations locally. As noted above, 11β-hydroxysteroid dehydrogenase inactivates glucocorticoids in renal tubular cells, preventing actions through the mineralocorticoid receptor. Thyroid hormone deiodinases convert T_4 to T_3 and can inactivate T_3. During development, degradation of retinoic acid by Cyp26b1 prevents primordial germ cells in the male from entering meiosis, as occurs in the female ovary.

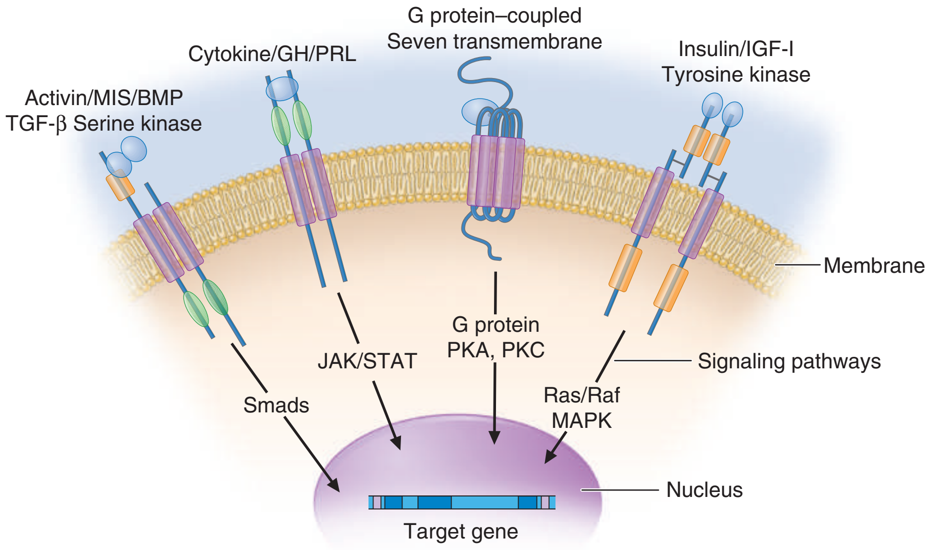

FIGURE 389-1 Membrane receptor signaling. MAPK, mitogen-activated protein kinase; PKA, C, protein kinase A, C; TGF, transforming growth factor. For other abbreviations, see text.

HORMONE ACTION THROUGH RECEPTORS

Receptors for hormones are divided into two major classes: membrane and nuclear. *Membrane receptors* primarily bind peptide hormones and catecholamines. *Nuclear receptors* bind small molecules that can diffuse across the cell membrane, such as steroids and vitamin D. Certain general principles apply to hormone-receptor interactions regardless of the class of receptor. Hormones bind to receptors with specificity and an affinity that generally coincides with the dynamic range of circulating hormone concentrations. Low concentrations of free hormone (usually 10^{-12} to 10^{-9} *M*) rapidly associate and dissociate from receptors in a bimolecular reaction such that the occupancy of the receptor at any given moment is a function of hormone concentration and the receptor's affinity for the hormone. Receptor numbers vary greatly in different target tissues, providing one of the major determinants of tissue-specific responses to circulating hormones. For example, ACTH receptors are located almost exclusively in the adrenal cortex, and LH receptors are found predominantly in the gonads. In contrast, insulin and TRs are widely distributed, reflecting the need for metabolic responses in all tissues.

MEMBRANE RECEPTORS

Membrane receptors for hormones can be divided into several major groups: (1) seven-transmembrane GPCRs, (2) tyrosine kinase receptors, (3) cytokine receptors, and (4) serine kinase receptors **(Fig. 389-1)**. The *seven-transmembrane GPCR family* binds a huge array of hormones, including large proteins (e.g., LH, PTH), small peptides (e.g., TRH, somatostatin), catecholamines (epinephrine, dopamine), and even minerals (e.g., calcium). The extracellular domains of GPCRs vary widely in size and are the major binding site for large hormones. The transmembrane-spanning regions are composed of hydrophobic α-helical domains that traverse the lipid bilayer. Like some channels, these domains are thought to circularize and form a hydrophobic pocket into which certain small ligands fit. Hormone binding induces conformational changes in these domains, transducing structural changes to the intracellular domain, which is a docking site for G proteins.

The large family of *G proteins*, so named because they bind guanine nucleotides (guanosine triphosphate [GTP], guanosine diphosphate [GDP]), provides great diversity for coupling receptors to different signaling pathways. G proteins form a heterotrimeric complex that is composed of various α and βγ subunits **(Fig. 389-2)**. The α subunit contains the guanine nucleotide–binding site and an intrinsic GTPase that hydrolyzes GTP → GDP. The βγ subunits are tightly associated and modulate the activity of the α subunit as well as mediating their own effector signaling pathways. G protein activity is regulated by a cycle that involves GTP hydrolysis and dynamic interactions between the α and βγ subunits. Hormone binding to the receptor induces GDP dissociation, allowing Gα to bind GTP and dissociate from the βγ complex. Under these conditions, the Gα subunit is activated and mediates signal transduction through various enzymes, such as adenylate cyclase and phospholipase C. GTP hydrolysis to GDP allows reassociation with the βγ subunits and restores the inactive state. G proteins interact with other cellular proteins, including kinases, channels, G protein–coupled receptor kinases (GRKs), and arrestins, that mediate signaling as well as receptor desensitization and recycling.

A variety of endocrinopathies result from mutations in GPCRs that alter their interactions with G proteins **(Table 389-2)**. Loss-of-function mutations are generally recessive and inactivate the relevant hormone signaling pathway. Because many of these receptors are important for development as well as signaling, patient presentations resemble glandular failure syndromes (e.g., mutations in LH-R, FSH-R, TSH-R). Gain-of-function (GOF) mutations are more complex. Selected GOF mutations induce conformational changes in the GPCR that mimic the activated state normally induced by hormone binding. These GOF mutations result in a constitutively active state in which G protein coupling stimulates cell signaling pathways, most commonly via cyclic adenosine 5′-monophosphate (cAMP) and protein kinase A. When mutations occur in the germline, the conditions are heritable and present in early life (e.g., LH-R, TSH-R). Sporadic, somatic mutations can also occur and result in clonal expansion of hyperfunctioning cells.

Mutations in the TSH-R illustrate the range of possible clinical consequences of GPCR mutations. Recessive inactivating mutations in the TSH-R cause congenital hypothyroidism with thyroid gland hypoplasia and resistance to TSH. Clinically, the hormone profile resembles primary hypothyroidism with low T_4 and high TSH. On the other hand, germline activating mutations cause congenital hyperthyroidism. The disorder is autosomal dominant because an activating mutation of one TSH-R allele is sufficient to induce cellular hyperfunction and disease. Because the TSH-R is activated in every cell of the thyroid, there is hyperplastic growth and hyperfunction that resembles the pathology seen in Graves' disease. This unusual disorder presents in infancy and must be distinguished from the more common clinical circumstance in which maternal antibodies in women with active or previously treated Graves' disease cross the placenta and stimulate the thyroid gland of the fetus. If an activating TSH-R mutation occurs later in life, in the somatic tissue, there is clonal expansion of the thyrocyte harboring the mutation, ultimately leading to an autonomous hyperfunctioning thyroid nodule. Of note, a similar condition can be caused by somatic mutations in $G_s\alpha$. In this case, the $G_s\alpha$ GTPase is inactivated and GTP cannot be converted to GDP. Consequently, the $G_s\alpha$ signaling pathway in this particular cell is constitutively active, mimicking chronic TSH stimulation and again leading to clonal expansion and an autonomous hyperfunctioning thyroid nodule. About one-third of hyperfunctioning "hot" thyroid nodules harbor sporadic mutations in either the TSH-R or $G_s\alpha$ (TSH-R mutations are more common).

$G_s\alpha$ mutations in tissues other than the thyroid can also cause endocrine disease. For example, $G_s\alpha$ mutations in pituitary somatotropes

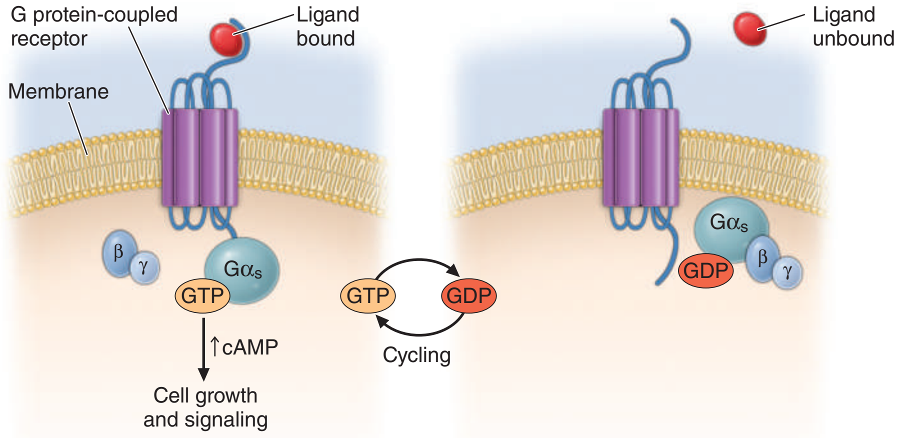

FIGURE 389-2 G protein signaling. G protein–coupled receptors (GPCRs) signal via the family of G proteins, so named because they bind guanylyl nucleotides. In the example shown, a GPCR bound to a ligand induces GDP dissociation, allowing $G_s\alpha$ to bind GTP and dissociate from the βγ complex. GTP-bound $G_s\alpha$ increases cAMP production by adenylyl cyclase and activates the protein kinase A pathway. Not shown are separate signaling pathways activated by the βγ complex. When GTP is converted to GDP by an intrinsic GTPase, the βγ subunits reassociate with GDP-bound $G_s\alpha$ and the complex returns to an inactive state. As noted in the text, mutations in $G_s\alpha$ that eliminate GTPase activity result in constitutive activation of receptor signaling pathways because GTP-bound $G_s\alpha$ cannot be converted to its GDP-bound inactive state. cAMP, cyclic adenosine 5′-monophosphate; GDP, guanosine diphosphate; $G_s\alpha$, G protein α; GTP, guanosine triphosphate.

mimic activation of the growth hormone–releasing hormone (GHRH) pathway and lead to GH-producing adenomas and acromegaly. Rarely, mutations in other components of the protein kinase A pathway in somatotropes can also cause GH-producing adenomas. $G_s\alpha$ mutations that occur early in development (typically mosaic) cause McCune-Albright syndrome **(Chap. 424)**, and the clinical features are manifest because the activated G protein pathway mimics the actions of various hormones (PTH, melanocyte-stimulating hormone [MSH], TSH, GHRH) in different tissues. Germline inactivating $G_s\alpha$ mutations cause a range of disorders that are transmitted and expressed in a complex manner because the locus is imprinted **(Chap. 422)**. These conditions include Albright's hereditary osteodystrophy (AHO), pseudopseudohypoparathyroidism (PPHP), and pseudohypoparathyroidism types 1b, 1c, and 2.

The *tyrosine kinase receptors* transduce signals for insulin and a variety of growth factors, such as IGF1, epidermal growth factor (EGF), nerve growth factor, platelet-derived growth factor, and fibroblast growth factors. The cysteine-rich extracellular domains contain binding sites for the growth factors. After ligand binding, this class of receptors undergoes autophosphorylation, inducing interactions with intracellular adaptor proteins such as Shc and insulin receptor substrates (IRS). In the case of the insulin receptor, multiple kinases are activated, including the Raf-Ras-MAPK and the Akt/protein kinase B pathways. The tyrosine kinase receptors play a prominent role in cell growth and differentiation as well as in intermediary metabolism.

The GH and PRL receptors belong to the *cytokine receptor* family. Analogous to the tyrosine kinase receptors, ligand binding induces receptor interaction with intracellular kinases—the Janus kinases (JAKs), which phosphorylate members of the signal transduction and activators of transcription (STAT) family—as well as with other signaling pathways (Ras, PI3-K, MAPK). The activated STAT proteins translocate to the nucleus and stimulate expression of target genes.

The *serine kinase receptors* mediate the actions of activins, transforming growth factor β, müllerian-inhibiting substance (MIS; also known as anti-müllerian hormone [AMH]), and bone morphogenic proteins (BMPs). This family of receptors (consisting of type I and II subunits) signals through proteins termed *smads* (fusion of terms for *Caenorhabditis elegans* sma + mammalian mad). Like the STAT proteins, the smads serve a dual role of transducing the receptor signal and acting as transcription factors. The pleomorphic actions of these growth factors dictate that they act primarily in a local (paracrine or autocrine) manner. Binding proteins such as follistatin (which binds activin and other members of this family) function to inactivate the growth factors and restrict their distribution.

Disease-causing mutations also occur in each of these classes of receptors. For example, insulin receptor mutations cause an extreme form of insulin resistance. GH receptor mutations cause Laron-type dwarfism, characterized by low IGF1 and high GH. AMH receptor mutations cause persistent müllerian duct syndrome. These hormone resistance syndromes are autosomal recessive and relatively uncommon. Unlike the GPCRs, activating mutations are unusual, although they do occur for the RET tyrosine kinase receptor, which causes the autosomal dominant disorder MEN type 2 (MEN2) **(Chap. 400)**.

NUCLEAR RECEPTORS

The family of nuclear receptors has nearly 100 members, many of which are still classified as orphan receptors because their ligands, if they exist, have not been identified **(Fig. 389-3)**. Otherwise, most nuclear receptors are classified on the basis of their ligands. Although all nuclear receptors ultimately act to increase or decrease gene transcription, some (e.g., glucocorticoid receptor) reside primarily in the cytoplasm, whereas others (e.g., TR) are located in the nucleus. After ligand binding, the cytoplasmically localized receptors translocate to the nucleus. There is growing evidence that certain ligands and their nuclear receptors (e.g., glucocorticoid, estrogen) can also act at the membrane or in the cytoplasm to modulate signal transduction pathways, providing a mechanism for cross-talk between membrane and nuclear receptors.

The structures of nuclear receptors have been studied extensively, including by x-ray crystallography. The DNA-binding domain, consisting of two zinc fingers, contacts specific DNA recognition sequences in target genes. Most nuclear receptors bind to DNA as dimers. Consequently, each monomer recognizes an individual DNA motif, referred to as a "half-site." The steroid receptors, including the glucocorticoid, estrogen, progesterone, and androgen receptors, bind to DNA as homodimers. Consistent with this twofold symmetry, their DNA recognition half-sites are palindromic. The thyroid, retinoid, peroxisome proliferator activated, and vitamin D receptors bind to DNA preferentially as heterodimers in combination with retinoid X receptors (RXRs). Their DNA half-sites are typically arranged as direct repeats.

The carboxy-terminal hormone-binding domains mediate transcriptional control. For type II receptors such as TR and retinoic acid receptor (RAR), co-repressor proteins bind to the receptor in the absence of ligand and silence gene transcription. Hormone binding induces conformational changes in the receptor, triggering the release of co-repressors and the recruitment of coactivators that stimulate transcription. Thus, these receptors are capable of mediating dynamic changes in the level of gene activity. Disease states can be associated with defective regulation of these events. For example, in promyelocytic leukemia, fusion of RARα to other nuclear proteins causes aberrant gene silencing that prevents normal cellular differentiation. Treatment with retinoic acid reverses this repression and allows cellular differentiation and apoptosis to occur. Most type 1 steroid receptors interact weakly with co-repressors, but ligand binding still induces interactions with an array of coactivators. X-ray crystallography shows that various SERMs induce distinct estrogen receptor conformations. The tissue-specific responses caused by these agents in breast, bone, and uterus appear to reflect distinct interactions with various coactivators. The receptor-coactivator complex stimulates gene transcription by several pathways, including (1) recruitment of enzymes

TABLE 389-2 Genetic Causes of G protein Receptor Disorders

RECEPTOR	DISORDER	GENETICS
LH	Leydig cell hypoplasia (male) Primary amenorrhea, resistance to LH (female) Familial male precocious puberty (male) Leydig cell adenoma, precocious puberty (male)	AR, inactivating AR, inactivating AD, activating Sporadic, activating
FSH	Hypergonadotropic ovarian failure (female) Hypospermia (male) Ovarian hyperstimulation (female)	AR, inactivating AR, inactivating Sporadic, activating
TSH	Congenital hypothyroidism, TSH resistance Nonautoimmune familial hyperthyroidism Hyperfunctioning thyroid adenoma	AR, AD, inactivating AD, activating Sporadic, activating
GnRH	Hypogonadotropic hypogonadism	AR, inactivating
Kisspeptin	Hypogonadotropic hypogonadism Precocious puberty	AR, inactivating AD, activating
Prokineticin	Precocious puberty	Sporadic, activating
TRH	Central hypothyroidism	AR, inactivating
GHRH	GH deficiency	AR, inactivating
PTH	Blomstrand chondrodysplasia Jansen metaphyseal chondrodysplasia	AR, inactivating AD, activating
Calcium sensing receptor	Familial hypocalciuric hypercalcemia Neonatal severe hyperparathyroidism Familial hypocalcemic hypercalciura	AD, inactivating AR, inactivating AD, activating
Arginine vasopressin receptor 2	Nephrogenic diabetes insipidus Nephrogenic SIADH	XL, inactivating XL, activating
ACTH	Familial ACTH resistance ACTH-independent Cushing syndrome	AR, inactivating Sporadic, activating
Melanocortin 4	Severe obesity	Codominant, inactivating

Abbreviations: ACTH, adrenocorticotropin hormone; AD, autosomal dominant; AR, autosomal recessive; FSH, follicle-stimulating hormone; GH, growth hormone; GHRH, growth hormone–releasing hormone; GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone; PTH, parathyroid hormone; SIADH, syndrome of inappropriate antidiuretic hormone secretion; TRH, thyrotropin-releasing hormone; TSH, thyroid-stimulating hormone; XL, X-linked.

(histone acetyl transferases) that modify chromatin structure, (2) interactions with additional transcription factors on the target gene, and (3) direct interactions with components of the general transcription apparatus to enhance the rate of RNA polymerase II–mediated transcription. Studies of nuclear receptor–mediated transcription reveal relatively rapid (e.g., 30–60 min) cycling of transcription complexes on any specific target gene.

Nuclear receptor mutations are an important cause of endocrine disease. Androgen receptor mutations cause androgen insensitivity syndrome (AIS) **(Chap. 402)**. Because the androgen receptor is located on the X chromosome, phenotypic expression is more commonly manifest than with other nuclear receptor disorders. Affected individuals with AIS are XY phenotypic females with retained testes and male-range testosterone levels. Tissue insensitivity to androgens varies based on the severity of the mutation. Müllerian structures are absent because Sertoli cells of the testis produce AMH during development. Female carriers of androgen receptor mutations are phenotypically normal. Recessive mutations of the estrogen, glucocorticoid, and vitamin D receptors occur but are rare.

Thyroid hormone receptor β (TRβ) mutations have an unusual pathophysiology. They are autosomal dominant and function via a “dominant negative” mechanism to cause resistance to thyroid hormone (RTH) **(Chap. 394)**. The mutations occur in selected regions of the TRβ hormone-binding domain and preserve the ability of the mutant receptor to heterodimerize with RXR, interact with corepressors, and bind to DNA regulatory sites. The mutant receptors function as antagonists of receptors from the normal copy of the TRβ gene. Affected patients have high T_4 and T_3 and inappropriately elevated (unsuppressed) TSH, reflecting impaired feedback regulation of the hypothalamic-pituitary-thyroid axis. Organ systems are variably resistant to thyroid hormones based on the relative expression of TRβ and TRα. Mutations in the genes encoding TRα and PPARγ can also cause disease by functioning in an analogous dominant negative manner.

FUNCTIONS OF HORMONES

The functions of individual hormones are described in detail in subsequent chapters. Nevertheless, it is useful to illustrate how most biologic responses require the integration of several different hormone pathways. The physiologic functions of hormones can be divided into three general types: (1) growth and differentiation, (2) maintenance of homeostasis, and (3) reproduction.

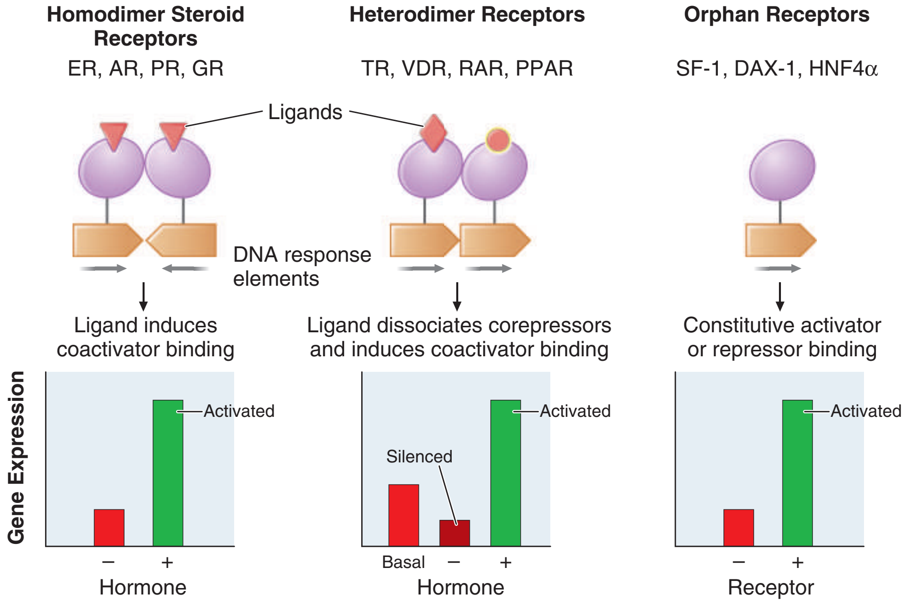

FIGURE 389-3 Nuclear receptor signaling. AR, androgen receptor; DAX, *d*osage-sensitive sex-reversal, *a*drenal hypoplasia congenita, *X* chromosome; ER, estrogen receptor; GR, glucocorticoid receptor; HNF4α, hepatic nuclear factor 4α; PPAR, peroxisome proliferator activated receptor; PR, progesterone receptor; RAR, retinoic acid receptor; SF-1, steroidogenic factor-1; TR, thyroid hormone receptor; VDR, vitamin D receptor.

GROWTH

Multiple hormones and nutritional factors mediate the complex phenomenon of growth **(Chap. 390)**. Short stature may be caused by GH deficiency, hypothyroidism, Cushing's syndrome, precocious puberty, malnutrition, chronic illness, or genetic abnormalities that affect the epiphyseal growth plates (e.g., *FGFR3* and *SHOX* mutations). Many factors (GH, IGF1, thyroid hormones) stimulate growth, whereas others (sex steroids) lead to epiphyseal closure. Understanding these hormonal interactions is important in the diagnosis and management of growth disorders. For example, delaying exposure to high levels of sex steroids may enhance the efficacy of GH treatment.

MAINTENANCE OF HOMEOSTASIS

Although virtually all hormones affect homeostasis, the most important among them are the following:

1. Thyroid hormone—controls ~25% of basal metabolism in most tissues.
2. Cortisol—exerts a permissive action for many hormones in addition to its own direct effects.
3. PTH—regulates calcium and phosphorus levels.
4. Vasopressin—regulates serum osmolality by controlling renal free-water clearance.
5. Mineralocorticoids—control vascular volume and serum electrolyte (Na^+, K^+) concentrations.
6. Insulin—maintains euglycemia in the fed and fasted states.

The defense against hypoglycemia is an impressive example of integrated hormone action **(Chap. 418)**. In response to the fasting state and falling blood glucose, insulin secretion is suppressed, resulting in decreased glucose uptake and enhanced glycogenolysis, lipolysis, proteolysis, and gluconeogenesis to mobilize fuel sources. If hypoglycemia develops (usually from insulin administration or sulfonylureas), an orchestrated counterregulatory response occurs—glucagon and epinephrine rapidly stimulate glycogenolysis and gluconeogenesis, whereas GH and cortisol act over several hours to raise glucose levels and antagonize insulin action.

Although free-water clearance is controlled primarily by vasopressin, cortisol and thyroid hormone are also important for facilitating renal tubular responses to vasopressin **(Chap. 393)**. PTH and vitamin D function in an interdependent manner to control calcium metabolism **(Chap. 421)**. PTH stimulates renal synthesis of 1,25-dihydroxyvitamin D, which increases calcium absorption in the gastrointestinal tract and enhances PTH action in bone. Increased calcium, along with vitamin D, feeds back to suppress PTH, thus maintaining calcium balance.

Depending on the severity of a specific stress and whether it is acute or chronic, multiple endocrine and cytokine pathways are activated to mount an appropriate physiologic response. In severe acute stress such as trauma or shock, the sympathetic nervous system is activated, and catecholamines are released, leading to increased cardiac output and a primed musculoskeletal system. Catecholamines also increase mean blood pressure and stimulate glucose production. Multiple stress-induced pathways converge on the hypothalamus, stimulating several hormones, including vasopressin and CRH. These hormones, in addition to cytokines (tumor necrosis factor α, interleukin [IL] 2, IL-6), increase ACTH and GH production. ACTH stimulates the adrenal gland, increasing cortisol, which in turn helps sustain blood pressure and dampen the inflammatory response. Increased vasopressin acts to conserve free water.

REPRODUCTION

The stages of reproduction include (1) sex determination during fetal development **(Chap. 402)**; (2) sexual maturation during puberty **(Chaps. 403 and 404)**; (3) conception, pregnancy, lactation, and child rearing **(Chap. 404)**; and (4) cessation of reproductive capability at menopause **(Chap. 407)**. Each of these stages involves an orchestrated interplay of multiple hormones, a phenomenon well illustrated by the dynamic hormonal changes that occur during each 28-day menstrual cycle. In the early follicular phase, pulsatile secretion of LH and FSH stimulates the progressive maturation of the ovarian follicle. This results in gradually increasing estrogen and progesterone levels, leading to enhanced pituitary sensitivity to GnRH, which, when combined with accelerated GnRH secretion, triggers the LH surge and rupture of the mature follicle. Inhibin, a protein produced by the granulosa cells, enhances follicular growth and feeds back to the pituitary to selectively suppress FSH without affecting LH. Growth factors such as EGF and IGF1 modulate follicular responsiveness to gonadotropins. Vascular endothelial growth factor and prostaglandins play a role in follicle vascularization and rupture.

During pregnancy, the increased production of PRL, in combination with placentally derived steroids (e.g., estrogen and progesterone), prepares the breast for lactation. Estrogens induce the production of progesterone receptors, allowing for increased responsiveness to progesterone. In addition to these and other hormones involved in lactation, the nervous system and oxytocin mediate the suckling response and milk release.

HORMONAL FEEDBACK REGULATORY SYSTEMS

Feedback control, both negative and positive, is a fundamental feature of endocrine systems. Each of the major hypothalamic-pituitary-hormone axes is governed by negative feedback, a process that maintains hormone levels within a relatively narrow range **(Chap. 390)**. Examples of hypothalamic-pituitary negative feedback include (1) thyroid hormones on the TRH-TSH axis, (2) cortisol on the CRH-ACTH axis, (3) gonadal steroids on the GnRH-LH/FSH axis, and (4) IGF1 on the GHRH-GH axis **(Fig. 389-4)**. These regulatory loops include both positive (e.g., TRH, TSH) and negative (e.g., T_4, T_3) components, allowing for exquisite control of hormone levels. As an example, a small reduction of thyroid hormone triggers a rapid increase of TRH and TSH secretion, resulting in thyroid gland stimulation and increased thyroid hormone production. When thyroid hormone reaches a normal level, it feeds back to suppress TRH and TSH, and a new steady state is attained. Feedback regulation also occurs for endocrine systems that do not involve the pituitary gland, such as calcium feedback on PTH, glucose inhibition of insulin secretion, and leptin feedback on the hypothalamus. An understanding of feedback regulation provides important insights into endocrine testing paradigms (see below).

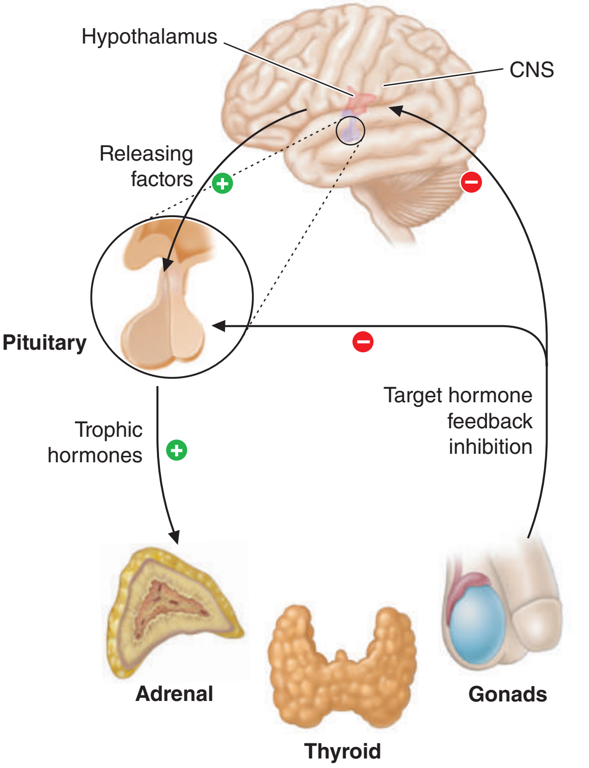

FIGURE 389-4 Feedback regulation of endocrine axes. CNS, central nervous system.

Positive feedback control also occurs but is not well understood. The primary example is estrogen-mediated stimulation of the midcycle LH surge. Although chronic low levels of estrogen are inhibitory, gradually rising estrogen levels stimulate LH secretion. This effect, which is illustrative of an endocrine rhythm (see below), involves activation of the hypothalamic GnRH pulse generator. In addition, estrogen-primed gonadotropes are extraordinarily sensitive to GnRH, leading to amplification of LH release.

PARACRINE AND AUTOCRINE CONTROL

The previously mentioned examples of feedback control involve classic endocrine pathways in which hormones are released by one gland and act on a distant target gland. However, local regulatory systems, often involving growth factors, are increasingly recognized. *Paracrine regulation* refers to factors released by one cell that act on an adjacent cell in the same tissue. For example, somatostatin secretion by pancreatic islet δ cells inhibits insulin secretion from nearby β cells. *Autocrine regulation* describes the action of a factor on the same cell from which it is produced. IGF1 acts on many cells that produce it, including chondrocytes, breast epithelium, and gonadal cells. Unlike endocrine actions, paracrine and autocrine control are difficult to document because local growth factor concentrations cannot be measured readily.

Anatomic relationships of glandular systems also greatly influence hormonal exposure: the physical organization of islet cells enhances their intercellular communication; the portal vasculature of the hypothalamic-pituitary system exposes the pituitary to high concentrations of hypothalamic releasing factors; testicular seminiferous tubules gain exposure to high testosterone levels produced by the interdigitated Leydig cells; the pancreas receives nutrient information and local exposure to peptide hormones (incretins) from the gastrointestinal tract; and the liver is the proximal target of insulin action because of portal drainage from the pancreas.

HORMONAL RHYTHMS

The feedback regulatory systems described above are superimposed on hormonal rhythms that are used for adaptation to the environment. Seasonal changes, the daily occurrence of the light-dark cycle, sleep, meals, and stress are examples of the many environmental events that affect hormonal rhythms. The *menstrual cycle* is repeated on average every 28 days, reflecting the time required to follicular maturation, ovulation, and potential implantation **(Chap. 390)**. Essentially all pituitary hormone rhythms are entrained to sleep and to the *circadian cycle*, generating reproducible patterns that are repeated approximately every 24 h. The HPA axis, for example, exhibits characteristic peaks of ACTH and cortisol production in the early morning, with a nadir during the night. Recognition of these rhythms is important for endocrine testing and treatment. Patients with Cushing's syndrome characteristically exhibit increased midnight cortisol levels compared with normal individuals **(Chap. 398)**. In contrast, morning cortisol levels are similar in these groups, as cortisol is normally high at this time of day in normal individuals. The HPA axis is more susceptible to suppression by glucocorticoids administered at night as they blunt the early-morning rise of ACTH. Understanding these rhythms allows glucocorticoid replacement that mimics diurnal production by administering larger doses in the morning than in the afternoon. Disrupted sleep rhythms can alter hormonal regulation. For example, sleep deprivation causes mild insulin resistance, food craving, and hypertension, which are reversible, at least in the short term. Emerging evidence indicates that circadian clock pathways not only regulate sleep-wake cycles but also play important roles in virtually every cell type. For example, tissue-specific deletion of clock genes alters rhythms and levels of gene expression, as well as metabolic responses in liver, adipose, and other tissues.

Other endocrine rhythms occur on a more rapid time scale. Many peptide hormones are secreted in discrete bursts every few hours. LH and FSH secretion are exquisitely sensitive to GnRH pulse frequency. Intermittent pulses of GnRH are required to maintain pituitary sensitivity, whereas continuous exposure to GnRH causes pituitary gonadotrope desensitization. This feature of the hypothalamic-pituitary-gonadotrope axis forms the basis for using long-acting GnRH agonists to treat central precocious puberty or to decrease testosterone levels in the management of prostate cancer. It is important to be aware of the pulsatile nature of hormone secretion and the rhythmic patterns of hormone production in relating serum hormone measurements to normal values. For some hormones, integrated markers have been developed to circumvent hormonal fluctuations. Examples include 24-h urine collections for cortisol, the measurement of IGF1 as a biologic marker of GH action, and HbA_{1c} as an index of long-term (weeks to months) blood glucose control.

Often, one must interpret endocrine data only in the context of other hormones. For example, PTH levels typically are assessed in combination with serum calcium concentrations. A high serum calcium level in association with elevated PTH is suggestive of hyperparathyroidism, whereas a suppressed PTH in the setting of hypercalcemia is more likely to be caused by hypercalcemia of malignancy, or other causes of hypercalcemia. Similarly, when T_4 and T_3 concentrations are low, TSH should be elevated, reflecting reduced feedback inhibition. When this is not the case, it is important to consider secondary hypothyroidism, which is caused by a defect at the level of the pituitary.

FURTHER READING

Fukami M et al: Gain-of-function mutations in G-protein-coupled receptor genes associated with human endocrine disorders. Clin Endocrinol 88:351, 2018.

Herbison AE: A simple model of estrous cycle negative and positive feedback regulation of GnRH secretion. Frontiers Neuroendocrinol 57:100837, 2020.

Kim YH, Lazar MA: Transcriptional control of circadian rhythms and metabolism: A matter of time and space. Endocr Rev 41:707, 2020.

Robertson RP (ed): *DeGroot's Endocrinology: Adult and Pediatric*, 8th ed. Philadelphia, Elsevier, 2023.

Scholtes C, Giguère V: Transcriptional control of energy metabolism by nuclear receptors. Nature Rev Mol Cell Biol 23:750, 2022.

390

Physiology of Anterior Pituitary Hormones

Shlomo Melmed, J. Larry Jameson

The anterior pituitary often is referred to as the "master gland" because, together with the hypothalamus, it orchestrates the complex regulatory functions of many other endocrine glands. The anterior pituitary gland produces six major hormones: (1) prolactin (PRL), (2) growth hormone (GH), (3) adrenocorticotropic hormone (ACTH), (4) luteinizing hormone (LH), (5) follicle-stimulating hormone (FSH), and (6) thyroid-stimulating hormone (TSH) **(Table 390-1)**. Pituitary hormones are secreted in a pulsatile manner, reflecting regulation by an array of specific hypothalamic releasing factors. Each of these pituitary hormones elicits specific trophic responses in peripheral target tissues including the adrenal, thyroid, and gonads, as well as tissues involved in metabolism (e.g., liver, breast, bone). Elicited hormonal products of peripheral glands, in turn, exert feedback control at the level of the hypothalamus and pituitary to modulate pituitary function **(Fig. 390-1)**. Pituitary tumors cause characteristic hormone excess syndromes. Hormone deficiency may be inherited or acquired. Fortunately, there are efficacious treatments for many pituitary hormone excess and deficiency syndromes. Nonetheless, these diagnoses are often elusive; this emphasizes the importance of recognizing subtle clinical manifestations and performing the correct laboratory diagnostic tests. **For discussion of disorders of the posterior pituitary or neurohypophysis, see Chap. 393.**

TABLE 390-1 Anterior Pituitary Hormone Expression and Regulation

CELL	CORTICOTROPE	SOMATOTROPE	LACTOTROPE	THYROTROPE	GONADOTROPE
Tissue-specific transcription factor	T-Pit	Prop-1, Pit-1	Prop-1, Pit-1	Prop-1, Pit-1, TEF	SF-1, DAX-1
Developmental timing	6 weeks	8 weeks	12 weeks	12 weeks	12 weeks
Hormone	POMC	GH	PRL	TSH	FSH, LH
Protein	Polypeptide	Polypeptide	Polypeptide	Glycoprotein α, β subunits	Glycoprotein α, β subunits
Amino acids	266 (ACTH 1–39)	191	198	211	210, 204
Stimulators	CRH, AVP, cytokines	GHRH, ghrelin	Estrogen, TRH, VIP	TRH	GnRH, activins, estrogen
Inhibitors	Glucocorticoids	Somatostatin, IGF-1	Dopamine	T_3, T_4, dopamine, somatostatin, glucocorticoids	Sex steroids, inhibin
Target gland	Adrenal	Liver, bone, other tissues	Breast, other tissues	Thyroid	Ovary, testis
Trophic effect	Steroid production	IGF-1 production, growth induction, insulin antagonism	Milk production	T_4 synthesis and secretion	Sex steroid production, follicle growth, germ cell maturation
Normal range	ACTH, 4–22 pg/L	<0.5 μg/L[a]	M <15 μg/L; F <20 μg/L	0.1–5 mU/L	M, 5–20 IU/L; F (basal), 5–20 IU/L

[a] Hormone secretion integrated over 24 h.

Abbreviations: F, female; M, male. For other abbreviations, see text.

Source: Courtesy of Elsevier.

ANATOMY AND DEVELOPMENT

ANATOMY

The pituitary gland weighs ~600 mg and is located within the sella turcica ventral to the diaphragma sella; it consists of anatomically and functionally distinct anterior and posterior lobes. The bony sella is contiguous to vascular and neurologic structures, including the cavernous sinuses, cranial nerves, and optic chiasm. Thus, expanding intrasellar pathologic processes may have significant central mass effects in addition to their endocrinologic impact.

Hypothalamic neural cells synthesize specific releasing and inhibiting hormones that are secreted directly into the portal vessels of the pituitary stalk. Blood supply of the pituitary gland comes from the superior and inferior hypophyseal arteries **(Fig. 390-2)**. The hypothalamic-pituitary portal plexus provides the major blood source for the anterior pituitary, allowing reliable transmission of hypothalamic peptide pulses without significant systemic dilution; consequently, anterior pituitary cells are exposed to specific releasing or inhibiting factors and in turn release their respective hormones as discrete pulses into the systemic circulation **(Fig. 390-3)**.

The posterior pituitary is supplied by the inferior hypophyseal arteries. In contrast to the anterior pituitary, the posterior lobe is directly innervated by hypothalamic neurons (supraopticohypophyseal and tuberohypophyseal nerve tracts) via the pituitary stalk **(Chap. 393)**. Thus, posterior pituitary production of arginine vasopressin (AVP) and oxytocin is particularly sensitive to neuronal damage by lesions that affect the pituitary stalk or hypothalamus.

PITUITARY DEVELOPMENT

The embryonic differentiation and maturation of anterior pituitary cells have been elucidated in considerable detail. Pituitary development from Rathke's pouch involves a complex interplay of lineage-specific transcription factors expressed in pluripotent Sox2-expressing precursor cells and gradients of locally produced growth factors (Table 390-1). The transcription factor Prop-1 induces pituitary development of Pit-1-specific lineages as well as gonadotropes. The transcription factor Pit-1 determines cell-specific expression of GH, PRL, and TSH in somatotropes, lactotropes, and thyrotropes. Expression of high levels of estrogen receptors in cells that contain Pit-1 favors PRL expression, whereas thyrotrope embryonic factor (TEF) induces TSH expression. Pit-1 binds to GH, PRL, and TSH gene regulatory elements, providing a mechanism for determining specific pituitary hormone phenotypic stability. Gonadotrope cell development is further defined by the cell-specific expression of the nuclear receptors steroidogenic factor (SF-1) and *d*osage-sensitive sex reversal, *a*drenal hypoplasia critical region, on chromosome *X*, gene *1* (DAX-1). Development of corticotrope cells, which express the proopiomelanocortin (*POMC*) gene, requires the T-Pit transcription factor. Abnormalities of pituitary development can be caused by inherited mutations of developmental transcription factors including Pit-1, Prop-1, SF-1, DAX-1, and T-Pit, resulting in selective or combined pituitary hormone deficit syndromes.

ANTERIOR PITUITARY HORMONES

Each anterior pituitary hormone is under unique control, and each exhibits highly specific normal and dysregulated secretory characteristics.

PROLACTIN

Synthesis PRL consists of 198 amino acids and has a molecular mass of 21,500 kDa; it is weakly homologous to GH and human placental lactogen (hPL), reflecting the duplication and divergence of a common GH-PRL-hPL precursor gene. PRL is synthesized in lactotropes, which constitute ~20% of anterior pituitary cells. Lactotropes and somatotropes are derived from a common precursor cell that may give rise to a tumor that secretes both PRL and GH. Lactotrope cell hyperplasia develops during pregnancy and the first few months of lactation. These transient functional changes in the lactotrope population are induced by estrogen to increase PRL production.

Secretion Normal adult serum PRL levels are about 10–25 μg/L in women and 10–20 μg/L in men. PRL secretion is pulsatile, with the highest secretory peaks occurring during non–rapid eye movement (non-REM) sleep. Peak serum PRL levels (up to 30 μg/L) occur between 4:00 and 6:00 A.M. The circulating half-life of PRL is ~50 min.

PRL is unique among the pituitary hormones in that the predominant hypothalamic control mechanism is inhibitory, reflecting tonic dopamine-mediated suppression of PRL release. This regulatory pathway accounts for the spontaneous PRL hypersecretion that occurs with pituitary stalk section, often a consequence of head trauma or compressive mass lesions at the skull base. Pituitary dopamine type 2 (D_2) receptors mediate inhibition of PRL synthesis and secretion. Targeted disruption (gene knockout) of the murine D_2 receptor in mice results in hyperprolactinemia and lactotrope proliferation. As discussed below, dopamine agonists play a central role in the management of hyperprolactinemic disorders.

Thyrotropin-releasing hormone (TRH) (pyro Glu-His-Pro-NH_2) is a hypothalamic tripeptide that elicits PRL release within 15–30 min after intravenous injection. TRH primarily regulates TSH, and the physiologic relevance of TRH for PRL regulation is unclear **(Chap. 394)**.

Serum PRL levels rise transiently after exercise, meals, sexual intercourse, minor surgical procedures, general anesthesia, chest wall injury, acute myocardial infarction, and other forms of acute stress. PRL levels increase markedly (about tenfold) during pregnancy and

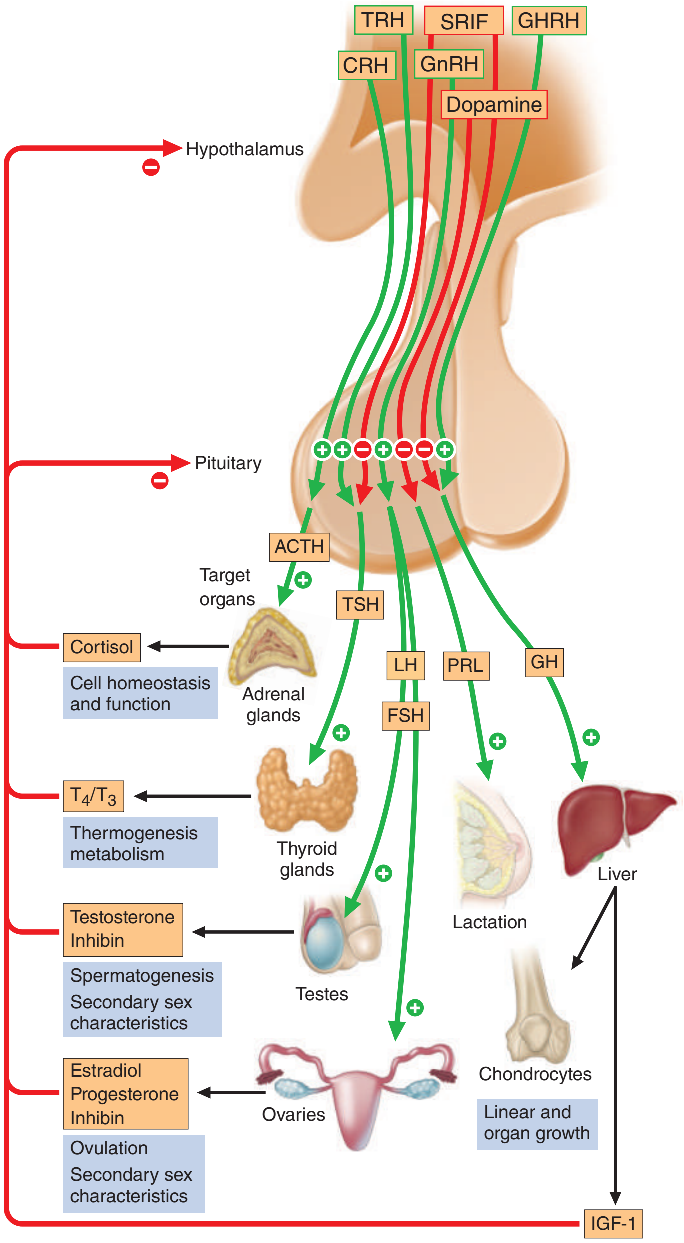
TRH
SRIF
GHRH
CRH
GnRH
Dopamine
Hypothalamus
Pituitary
ACTH
Target organs
TSH
Cortisol
Cell homeostasis and function
Adrenal glands
LH
PRL
GH
FSH
T4/T3
Thermogenesis metabolism
Thyroid glands
Liver
Lactation
Testosterone
Inhibin
Spermatogenesis
Secondary sex characteristics
Testes
Chondrocytes
Estradiol
Progesterone
Inhibin
Ovaries
Linear and organ growth
Ovulation
Secondary sex characteristics
IGF-1

FIGURE 390-1 Diagram of pituitary axes. Hypothalamic hormones regulate anterior pituitary trophic hormones that in turn determine target gland secretion. Peripheral hormones feed back to negatively regulate hypothalamic and pituitary hormones. For abbreviations, see text.

decline rapidly within 2 weeks of parturition. If breast-feeding is initiated, basal PRL levels remain elevated; suckling stimulates transient reflex increases in PRL levels that last for ~30–45 min. Breast suckling activates afferent neural pathways in the hypothalamus that induce PRL release. With time, suckling-induced responses diminish and interfeeding PRL levels return to normal.

Action The PRL receptor is a member of the type I cytokine receptor family that also includes GH and interleukin (IL) 6 receptors. Ligand binding induces receptor dimerization and intracellular signaling by Janus kinase (JAK), which stimulates translocation of the signal transduction and activators of transcription (STAT) family to activate target genes. Mutations of the PRL receptor result in PRL insensitivity, hyperprolactinemia, and oligomenorrhea. When homozygous, PRL receptor mutations cause agalactia, demonstrating that PRL action is necessary for lactation. In the breast, the lobuloalveolar epithelium proliferates in response to PRL, placental lactogens, estrogen, progesterone, and local paracrine growth factors, including insulin-like growth factor 1 (IGF-1).

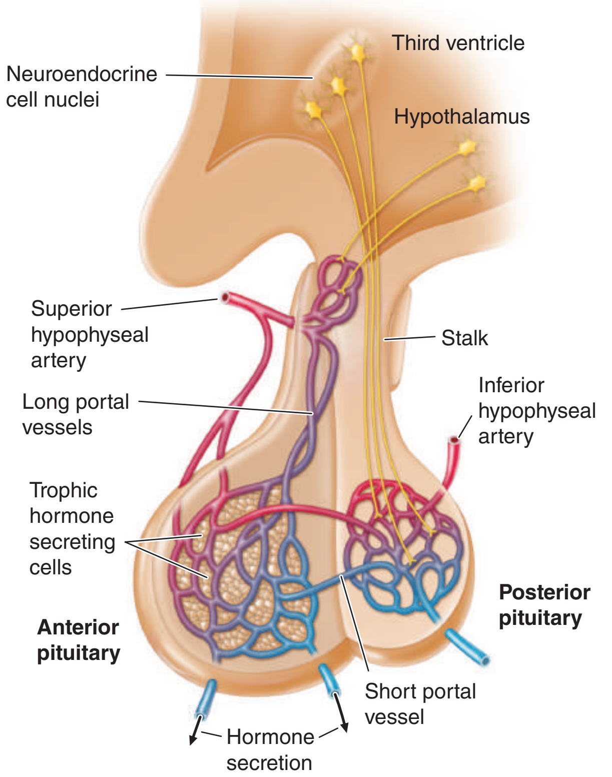
Third ventricle
Neuroendocrine cell nuclei
Hypothalamus
Superior hypophyseal artery
Stalk
Long portal vessels
Inferior hypophyseal artery
Trophic hormone secreting cells
Posterior pituitary
Anterior pituitary
Short portal vessel
Hormone secretion

FIGURE 390-2 Diagram of hypothalamic-pituitary vasculature. The hypothalamic nuclei produce hormones that traverse the portal system and impinge on anterior pituitary cells to regulate pituitary hormone production. Posterior pituitary hormones are derived from direct neural extensions.

PRL acts to induce and maintain lactation and to suppress both reproductive function and sexual drive. These functions are geared toward ensuring that maternal lactation is sustained and not interrupted by pregnancy. PRL inhibits reproductive function by suppressing hypothalamic gonadotropin-releasing hormone (GnRH) and pituitary gonadotropin secretion and by impairing gonadal steroidogenesis in both women and men. In the ovary, PRL blocks folliculogenesis and inhibits granulosa cell aromatase activity, leading to hypoestrogenism and anovulation. PRL also has a luteolytic effect, generating a shortened, or inadequate, luteal phase of the menstrual cycle. In men, attenuated LH secretion leads to low testosterone levels and decreased spermatogenesis. These hormonal changes decrease libido and reduce fertility in patients with hyperprolactinemia.

GROWTH HORMONE

Synthesis GH is the most abundant anterior pituitary hormone, and GH-secreting somatotrope cells constitute up to 50% of the total anterior pituitary cell population. Mammosomatotrope cells, which coexpress PRL with GH, can be identified by using double immunostaining techniques. Somatotrope development and GH transcription

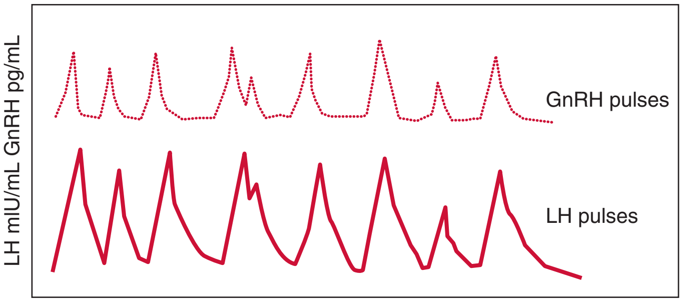

FIGURE 390-3 Hypothalamic gonadotropin-releasing hormone (GnRH) pulses induce secretory pulses of luteinizing hormone (LH).

are determined by expression of the cell-specific Pit-1 nuclear transcription factor. Five distinct genes encode GH and related proteins. The pituitary GH gene (*hGH-N*) produces two alternatively spliced products that give rise to 22-kDa GH (191 amino acids) and a less abundant 20-kDa GH molecule with similar biologic activity. Placental syncytiotrophoblast cells express a GH variant (*hGH-V*) gene; the related hormone human chorionic somatotropin (HCS) is expressed by distinct members of the gene cluster.

Secretion GH secretion is controlled by complex hypothalamic and peripheral factors. *GH-releasing hormone* (GHRH) is a 44-amino-acid hypothalamic peptide that stimulates GH synthesis and release. Ghrelin, an octanoylated gastric-derived peptide, and synthetic agonists of the *GHS-R* induce GHRH and also directly stimulate GH release. *Somatostatin* (somatotropin-release inhibiting factor [SRIF]) is synthesized in the medial preoptic area of the hypothalamus and inhibits GH secretion. GHRH is secreted in discrete spikes that elicit GH pulses, whereas SRIF sets basal GH secretory tone. SRIF also is expressed in many extrahypothalamic tissues, including the central nervous system (CNS), gastrointestinal tract, and pancreas, where it also acts to inhibit islet hormone secretion. *IGF-1*, the peripheral target hormone for GH, feeds back to inhibit GH; estrogen induces GH, whereas chronic glucocorticoid excess suppresses GH release, leading to growth delay in children.

Surface receptors on the somatotrope regulate GH synthesis and secretion. The GHRH receptor is a G protein–coupled receptor (GPCR) that signals through the intracellular cyclic AMP pathway to stimulate somatotrope cell proliferation as well as GH production. Inactivating mutations of the GHRH receptor cause profound growth deficiency (dwarfism). A distinct surface receptor for ghrelin, the gastric-derived GH secretagogue, is expressed in both the hypothalamus and pituitary. Somatostatin binds to five distinct receptor subtypes (SST1 to SST5); SST2 and SST5 subtypes preferentially suppress GH (and TSH) secretion, while SST5 predominantly suppresses ACTH secretion.

GH secretion is pulsatile, with highest peak levels occurring at night, generally correlating with sleep onset. GH secretory rates decline markedly with age so that hormone levels in middle age are ~15% of pubertal levels. These changes are paralleled by an age-related decline in lean muscle mass. GH secretion is also reduced in obese individuals, although IGF-1 levels may not be suppressed, suggesting a change in the setpoint for feedback control. Elevated GH levels occur within an hour of deep sleep onset as well as after exercise, physical stress, and trauma and during sepsis. Integrated 24-h GH secretion is higher in women and is also enhanced by estrogen replacement, likely reflective of increased peripheral GH resistance. Using standard assays, random GH measurements are undetectable in ~50% of daytime samples obtained from healthy subjects and are also undetectable (<1 μg/L) in most obese and elderly subjects. Thus, single random GH measurements do not distinguish patients with adult GH deficiency from those with GH levels in the normal range.

GH secretion is profoundly influenced by nutritional factors. Using ultrasensitive GH assays with a sensitivity of 0.002 μg/L, a glucose load suppresses GH to <0.7 μg/L in women and to <0.07 μg/L in men. Increased GH pulse frequency and peak amplitudes occur with chronic malnutrition or prolonged fasting. GH is stimulated by oral ghrelin receptor agonists, intravenous L-arginine, dopamine, and apomorphine (a dopamine receptor agonist), as well as by α-adrenergic pathways. β-Adrenergic blockade induces basal GH and enhances GHRH- and insulin-evoked GH release.

Action The pattern of GH secretion may affect tissue responses. The higher GH pulsatility observed in men compared with the relatively continuous basal GH secretion in women may be an important biologic determinant of linear growth patterns and liver enzyme induction.

The 70-kDa peripheral GH receptor protein has structural homology with the cytokine/hematopoietic superfamily. A fragment of the receptor extracellular domain generates a soluble GH binding protein (GHBP) that binds to circulating GH. The liver and cartilage express the greatest number of GH receptors. GH binding to preformed receptor dimers is followed by internal rotation and subsequent signaling through the JAK/STAT pathway. Activated STAT proteins translocate to the nucleus, where they modulate expression of GH-regulated target genes. GH analogues that bind to the receptor but are incapable of mediating receptor signaling are potent antagonists of GH action. A GH receptor antagonist (pegvisomant) is approved for treatment of acromegaly.

GH induces protein synthesis and nitrogen retention and also impairs glucose tolerance by antagonizing insulin action. GH also stimulates lipolysis, leading to increased circulating fatty acid levels, reduced omental fat mass, and enhanced lean body mass. GH promotes sodium, potassium, and water retention and elevates serum levels of inorganic phosphate. Linear bone growth occurs as a result of complex hormonal and growth factor actions, including those of IGF-1. GH stimulates epiphyseal prechondrocyte differentiation. These precursor cells produce IGF-1 locally, and their proliferation is also responsive to the growth factor.

Insulin-Like Growth Factors Although GH exerts direct effects in target tissues, many of its physiologic effects are mediated indirectly through IGF-1, a potent growth and differentiation factor. The liver is the major source of circulating IGF-1. In peripheral tissues, IGF-1 also exerts local paracrine actions that appear to be both dependent on and independent of GH. Thus, GH administration induces circulating IGF-1 as well as stimulating local IGF-1 production in multiple tissues.

Both IGF-1 and IGF-2 are bound to high-affinity circulating IGF-binding proteins (IGFBPs) that regulate IGF availability and bioactivity. Levels of IGFBP3 are GH dependent, and it serves as the major carrier protein for circulating IGF-1. GH deficiency and malnutrition usually are associated with low IGFBP3 levels. IGFBP1 and IGFBP2 regulate local tissue IGF action but do not bind appreciable amounts of circulating IGF-1.

Serum IGF-1 concentrations are profoundly affected by physiologic factors. Levels increase during puberty, peak at 16 years, and subsequently decline by >80% during the aging process. IGF-1 concentrations are higher in women than in men. Because GH is the major determinant of hepatic IGF-1 synthesis, abnormalities of GH synthesis or action (including pituitary failure, GHRH receptor defect, GH receptor defect, or pharmacologic GH receptor blockade) lead to reduced IGF-1 levels. Hypocaloric states are associated with GH resistance; IGF-1 levels are therefore low with cachexia, malnutrition, and sepsis. In acromegaly, IGF-1 levels are high and reflect a log-linear relationship with circulating GH concentrations.

IGF-1 PHYSIOLOGY Injected IGF-1 (100 μg/kg) induces hypoglycemia, and lower doses improve insulin sensitivity in patients with severe insulin resistance and diabetes. In cachectic subjects, IGF-1 infusion (12 μg/kg per h) enhances nitrogen retention and lowers cholesterol levels. Longer-term subcutaneous IGF-1 injections enhance protein synthesis and are anabolic. Although bone formation markers are induced, bone turnover also may be stimulated by IGF-1. IGF-1 is approved for use in patients with GH-resistance syndromes.

IGF-1 side effects are dose dependent, and overdose may result in hypoglycemia, hypotension, fluid retention, temporomandibular jaw pain, and increased intracranial pressure, all of which are reversible. Retinal damage and avascular femoral head necrosis have been reported. Chronic excess IGF-1 administration presumably would result in features of acromegaly.

ADRENOCORTICOTROPIC HORMONE

(See also Chap. 398).

Synthesis ACTH-secreting corticotrope cells constitute ~20% of the pituitary cell population. ACTH (39 amino acids) is derived from the POMC precursor protein (266 amino acids) that also generates several other peptides, including β-lipotropin, β-endorphin, met-enkephalin, α-melanocyte-stimulating hormone (α-MSH), and corticotropin-like intermediate lobe protein (CLIP). The *POMC* gene is potently suppressed by glucocorticoids and induced by corticotropin-releasing hormone (CRH), AVP, and proinflammatory cytokines, including IL-6, as well as leukemia inhibitory factor.

CRH, a 41-amino-acid hypothalamic peptide synthesized in the paraventricular nucleus as well as in higher brain centers, is the predominant stimulator of ACTH synthesis and release. The CRH receptor is a GPCR that is expressed on the corticotrope and signals to induce *POMC* transcription.

Secretion ACTH secretion is pulsatile and exhibits a characteristic circadian rhythm, peaking at about 6:00 A.M. and reaching a nadir about midnight. Adrenal glucocorticoid secretion, which is driven by ACTH, follows a parallel diurnal pattern. ACTH circadian rhythmicity is determined by variations in secretory pulse amplitude rather than changes in pulse frequency. Superimposed on this endogenous rhythm, ACTH levels are increased by physical and psychological stress, exercise, acute illness, and insulin-induced hypoglycemia.

Glucocorticoid-mediated negative regulation of the hypothalamic-pituitary-adrenal (HPA) axis occurs as a consequence of both hypothalamic CRH suppression and direct attenuation of pituitary *POMC* gene expression and ACTH release. In contrast, loss of cortisol feedback inhibition, as occurs in primary adrenal failure, results in extremely high ACTH levels.

Acute inflammatory or septic insults activate the HPA axis through the integrated actions of proinflammatory cytokines, bacterial toxins, and neural signals. The overlapping cascade of ACTH-inducing cytokines (tumor necrosis factor [TNF]; IL-1, -2, and -6; and leukemia inhibitory factor) activates hypothalamic CRH and AVP secretion, pituitary *POMC* gene expression, and local pituitary paracrine cytokine networks. The resulting cortisol elevation restrains the inflammatory response and enables host protection. Concomitantly, cytokine-mediated central glucocorticoid receptor resistance impairs glucocorticoid suppression of the HPA. Thus, the neuroendocrine stress response reflects the net result of highly integrated hypothalamic, intrapituitary, and peripheral hormone and cytokine signals acting to regulate cortisol secretion.

Action The major function of the HPA axis is to maintain metabolic homeostasis and mediate the neuroendocrine stress response, largely by inducing adrenal cortisol production. ACTH induces adrenocortical steroidogenesis by sustaining adrenal cell proliferation and function. The receptor for ACTH, designated *melanocortin-2 receptor*, is a GPCR that induces steroidogenesis by stimulating a cascade of steroidogenic enzymes **(Chap. 398)**.

GONADOTROPINS: FSH AND LH

Synthesis and Secretion Gonadotrope cells constitute ~10% of anterior pituitary cells and produce two gonadotropin hormones—LH and FSH. Like TSH and human chorionic gonadotropin, LH and FSH are glycoprotein hormones that contain α and β subunits. The α subunit is common to these glycoprotein hormones; specificity of hormone function is conferred by the β subunits, which are expressed by separate genes.

Gonadotropin synthesis and release are dynamically regulated. This is particularly true in women, in whom rapidly fluctuating gonadal steroid levels vary throughout the menstrual cycle. Hypothalamic GnRH, a 10-amino-acid peptide, regulates the synthesis and secretion of both LH and FSH. Brain kisspeptin, a product of the *KISS1* gene, regulates hypothalamic GnRH release. GnRH is secreted in discrete pulses every 60–120 min, and the pulses in turn elicit LH and FSH pulses (Fig. 390-3). The pulsatile mode of GnRH input is essential to its action; pulses prime gonadotrope responsiveness, whereas continuous GnRH exposure induces desensitization. Based on this phenomenon, long-acting GnRH agonists are used to suppress gonadotropin levels in children with precocious puberty and in men with prostate cancer **(Chap. 92)** and are used in some ovulation-induction protocols to reduce levels of endogenous gonadotropins **(Chap. 404)**. Estrogens act at both the hypothalamus and the pituitary to modulate gonadotropin secretion. Chronic estrogen exposure is inhibitory, whereas rising estrogen levels, as occur during the preovulatory surge, exert positive feedback to enhance pituitary responsiveness and to increase gonadotropin pulse frequency and amplitude. Progesterone slows GnRH pulse frequency but enhances gonadotropin responses to GnRH. Testosterone feedback in men also occurs at the hypothalamic and pituitary levels and is mediated in part by its conversion to estrogens.

Although GnRH is the main regulator of LH and FSH secretion, FSH synthesis is also under distinct control by the gonadal peptides inhibin and activin, members of the transforming growth factor β (TGF-β) family. Inhibin selectively suppresses FSH, whereas activin stimulates FSH synthesis **(Chap. 404)**.

Action The gonadotropin hormones interact with their respective GPCRs expressed in the ovary and testis, evoking germ cell development and maturation and steroid hormone biosynthesis. In women, FSH regulates ovarian follicle development and stimulates ovarian estrogen production. LH mediates ovulation and maintenance of the corpus luteum. In men, LH induces Leydig cell testosterone synthesis and secretion, and FSH stimulates seminiferous tubule development and regulates spermatogenesis.

THYROID-STIMULATING HORMONE

Synthesis and Secretion TSH-secreting thyrotrope cells constitute 5% of the anterior pituitary cell population. TSH shares a common α subunit with LH and FSH but contains a specific TSH β subunit. TRH is a hypothalamic tripeptide (pyroglutamyl histidylprolinamide) that acts through a pituitary GPCR to stimulate TSH synthesis and secretion; it also stimulates the lactotrope cell to secrete PRL. TSH secretion is stimulated by TRH, whereas thyroid hormones, dopamine, somatostatin, and glucocorticoids suppress TSH by overriding TRH induction. Thyroid hormones are the predominant negative regulator of TSH production.

Thyrotrope cell proliferation and TSH secretion are both induced when negative feedback inhibition by thyroid hormones is removed. Thus, thyroid damage (including surgical thyroidectomy), radiation-induced hypothyroidism, chronic thyroiditis, and prolonged goitrogen exposure are associated with increased TSH levels. Long-standing untreated hypothyroidism can lead to elevated TSH levels, which may be associated with thyrotrope hyperplasia and pituitary enlargement and may sometimes be evident on magnetic resonance imaging.

Action TSH is secreted in pulses, although the excursions are modest in comparison to other pituitary hormones because of the low amplitude of the pulses and the relatively long half-life of TSH. Consequently, single determinations of TSH suffice to precisely assess its circulating levels. TSH binds to a GPCR on thyroid follicular cells to stimulate thyroid hormone synthesis and release **(Chap. 394)**.

FURTHER READING

Bernard V et al: Prolactin: A pleiotropic factor in health and disease. Nat Rev Endocrinol 15:356, 2019.

Das N, Kumar TR: Molecular regulation of follicle-stimulating hormone synthesis, secretion and action. J Mol Endocrinol 60:R131, 2018.

Langlais D et al: Adult pituitary cell maintenance: Lineage-specific contribution of self-duplication. Mol Endocrinol 27:1103, 2013.

Le Tissier P et al: The process of anterior pituitary hormone pulse generation. Endocrinology 159:3524, 2018.

Ho KY et al: The physiology of growth hormone (GH) in adults: Translational journey to GH replacement therapy. J Endocrinol 257:e220197, 2023.

Ranke MB, Wit JM: Growth hormone: Past, present and future. Nat Rev Endocrinol 14:285, 2018.

Zhang S et al: Single-cell transcriptomics identifies divergent developmental lineage trajectories during human pituitary development. Nat Commun 11:5275, 2020.

391 Hypopituitarism

Shlomo Melmed, J. Larry Jameson

Deficient production of anterior pituitary hormones leads to features of hypopituitarism. Impaired production of one or more of the anterior pituitary trophic hormones can result from inherited disorders; more commonly, adult hypopituitarism is acquired and reflects the compressive mass effects of tumors or the consequences of local pituitary or hypothalamic traumatic, autoimmune, inflammatory, or vascular damage. These processes also may impair synthesis or secretion of hypothalamic hormones, with resultant pituitary failure **(Table 391-1)**.

TABLE 391-1 Etiology of Hypopituitarism[a]

Development/structural	
	Midline cerebral defect syndromes
	Pituitary dysplasia/aplasia
	Primary empty sella
	Congenital hypothalamic disorders (septo-optic dysplasia, Prader-Willi syndrome, Bardet-Biedl syndrome, Kallmann syndrome)
	Congenital central nervous system mass, encephalocele
Genetic	
	Combined pituitary hormone deficiencies
	Isolated primary hormone deficiencies
Traumatic	
	Surgical resection
	Radiotherapy damage
	Head injuries
Neoplastic	
	Pituitary adenoma
	Parasellar mass (germinoma, ependymoma, glioma)
	Rathke's cyst
	Craniopharyngioma
	Hypothalamic hamartoma, gangliocytoma
	Pituitary metastases (breast, lung, colon carcinoma)
	Lymphoma and leukemia
	Meningioma
Infiltrative/inflammatory	
	Lymphocytic hypophysitis
	Hemochromatosis
	Sarcoidosis
	Histiocytosis X
	Granulomatous hypophysitis
	Transcription factor antibodies
	Immunotherapy
Vascular	
	Pituitary apoplexy
	Pregnancy-related (infarction with diabetes; postpartum necrosis)
	Subarachnoid hemorrhage
	Sickle cell disease
	Arteritis
	Snake bite venom
Infections	
	Fungal (histoplasmosis)
	Parasitic (toxoplasmosis)
	Tuberculosis
	Pneumocystis jirovecii
Drug-induced	
	CTLA-4 inhibitors
	PD-1/PD-L1 inhibitors

[a] Trophic hormone failure associated with pituitary compression or destruction usually occurs sequentially: growth hormone > follicle-stimulating hormone > luteinizing hormone > thyroid-stimulating hormone > adrenocorticotropic hormone. During childhood, growth retardation is often the presenting feature, and in adults, hypogonadism is the earliest symptom.

Abbreviations: CTLA-4, cytotoxic T lymphocyte antigen 4; PD-1, programmed cell death protein 1; PD-L1, programmed cell death protein ligand 1.

DEVELOPMENTAL CAUSES OF HYPOPITUITARISM

Pituitary dysplasia may result in aplastic, hypoplastic, or ectopic pituitary gland development. Because pituitary development follows midline cell migration from the nasopharyngeal Rathke's pouch, midline craniofacial disorders may be associated with pituitary dysplasia. Acquired pituitary failure in the newborn also can be caused by birth trauma, including cranial hemorrhage, asphyxia, and breech delivery.

A large number (>50) of transcription factors and growth factors are critical for the development of the hypothalamus and pituitary gland and the function of differentiated anterior pituitary cell lineages. Mutations have been described in the *HESX1, SOX2, SOX3, LHX3, LHX4, OTX, GLI2, PAX6, BMP4, ARNT2, FGF8, FGFR1, SHH, PROKR2, GPR161, IGSF1, PITX2,* and *CHD7* genes, among others. Heterozygous loss-of-function or autosomal recessive mutations disrupt hypothalamic and pituitary development at different developmental stages, causing a wide array of phenotypes ranging from severe syndromic midline and other defects to combined pituitary hormone defects or isolated hormone deficiencies. Depending on the gene involved, the pituitary may be hypoplastic, hyperplastic, or ectopic. Midline defects include variable combinations of abnormal development of the eyes, corpus collosum, vertebrae, and genital systems. Pituitary dysfunction ranges from isolated hormone deficiency to combined pituitary hormone deficiency (CPHD) and arginine vasopressin deficiency (AVP-D).

In addition to these syndromic developmental disorders, some mutations affect specific pituitary cell lineages. For example, Pit-1 mutations cause combined growth hormone (GH), prolactin (PRL), and thyroid-stimulating hormone (TSH) deficiencies. These patients usually present with growth failure and varying degrees of hypothyroidism. The pituitary may appear hypoplastic on magnetic resonance imaging (MRI). Prop-1 is expressed early in pituitary development and appears to be required for Pit-1 function. Familial and sporadic *PROP1* mutations result in combined GH, PRL, TSH, and gonadotropin deficiency. Over 80% of these patients have growth retardation; by adulthood, all are deficient in TSH and gonadotropins, and a small minority later develop adrenocorticotropic hormone (ACTH) deficiency. Because of gonadotropin deficiency, these individuals do not enter puberty spontaneously. In some cases, the pituitary gland appears enlarged on MRI. *TPIT* mutations result in ACTH deficiency associated with hypocortisolism. Mutations in *NR5A1* (also known as steroidogenic factor 1 [*SF1*]) impair development of gonadotropes, as well as adrenal/gonadal development.

HYPOTHALAMIC ENDOCRINE DYSFUNCTION

Hypothalamic disorders can affect temperature regulation, appetite, sleep-wake cycles, autonomic systems, behavior, and memory, as well as multiple endocrine systems. Selected examples of hypothalamic disorders that affect the endocrine system are described below.

Kallmann Syndrome Kallmann syndrome results from defective hypothalamic gonadotropin-releasing hormone (GnRH) synthesis and is associated with anosmia or hyposmia due to olfactory bulb agenesis or hypoplasia **(Chap. 403)**. Classically, the syndrome may also be associated with color blindness, optic atrophy, nerve deafness, cleft palate, renal abnormalities, cryptorchidism, and neurologic abnormalities such as mirror movements. The initial genetic cause was the X-linked *KAL* gene, mutations of which impair embryonic migration of GnRH neurons from the hypothalamic olfactory placode to the hypothalamus. Since then, more than a dozen additional genetic abnormalities, in addition to *KAL* mutations, have been found to cause isolated GnRH deficiency. Autosomal recessive (i.e., *GPR54, KISS1*) and dominant (i.e., *FGFR1*) modes of transmission have been described, and there is a growing list of genes associated with GnRH deficiency (including *GNRH1, PROK2, PROKR2, CHD7, PCSK1, FGF8, NELF, WDR11, TAC3, TACR3,* and *SEMA3E*). Some patients have oligogenic mutations in which mutations in a combination of different genes lead to the phenotype. Associated clinical features, in addition to GnRH deficiency, vary depending on the genetic cause. GnRH deficiency prevents progression through puberty. Males present with delayed puberty and pronounced hypogonadal features, including micropenis, probably the result of low testosterone levels during infancy. Females present with primary amenorrhea and failure of secondary sexual development.

Kallmann syndrome and other causes of congenital GnRH deficiency are characterized by low luteinizing hormone (LH) and follicle-stimulating hormone (FSH) levels and low concentrations of sex steroids (testosterone or estradiol). In sporadic cases of isolated gonadotropin deficiency, the diagnosis is often one of exclusion after other known causes of hypothalamic-pituitary dysfunction have been eliminated. Repetitive GnRH administration restores normal pituitary gonadotropin responses, pointing to a hypothalamic defect in these patients.

Long-term treatment of males with human chorionic gonadotropin (hCG) or testosterone restores pubertal development and secondary sex characteristics; women can be treated with cyclic estrogen and progestin. Fertility may be restored by the administration of gonadotropins or by using a portable infusion pump to deliver subcutaneous, pulsatile GnRH.

Bardet-Biedl Syndrome This very rare genetically heterogeneous disorder is characterized by intellectual disability, renal abnormalities, obesity, and hexadactyly, brachydactyly, or syndactyly. Central AVP-D may or may not be associated. GnRH deficiency occurs in 75% of males and half of affected females. Retinal degeneration begins in early childhood, and most patients are blind by age 30. Numerous subtypes of Bardet-Biedl syndrome have been identified, with genetic linkage to at least nine different loci. Several of the loci encode genes involved in basal body cilia function, and this may account for the diverse clinical manifestations.

Leptin and Leptin Receptor Mutations Deficiencies of leptin or its receptor cause a broad spectrum of hypothalamic abnormalities, including hyperphagia, obesity, and central hypogonadism **(Chap. 413)**. Decreased GnRH production in these patients results in attenuated pituitary FSH and LH synthesis and release.

Prader-Willi Syndrome This is a contiguous gene syndrome that results from deletion of the paternal copies of the imprinted *SNRPN* gene, the *NECDIN* gene, and possibly other genes on chromosome 15q. Prader-Willi syndrome is associated with hypogonadotropic hypogonadism, hyperphagia-obesity, chronic muscle hypotonia, intellectual disability, and adult-onset diabetes mellitus. Multiple somatic defects also involve the skull, eyes, ears, hands, and feet. Diminished hypothalamic oxytocin- and AVP-producing nuclei have been reported. Deficient GnRH synthesis is suggested by the observation that chronic GnRH treatment restores pituitary LH and FSH release.

ACQUIRED HYPOPITUITARISM

Hypopituitarism may be caused by accidental or neurosurgical trauma; vascular events such as apoplexy; pituitary or hypothalamic neoplasms, craniopharyngioma, lymphoma, or metastatic tumors; inflammatory disease such as lymphocytic hypophysitis; autoimmune hypophysitis associated with checkpoint inhibitor cancer immunotherapy; infiltrative disorders such as sarcoidosis, hemochromatosis **(Chap. 426)**, and tuberculosis; or irradiation.

Patients with brain injury, including from contact sports trauma, motor vehicle accidents, explosive causes, subarachnoid hemorrhage, and irradiation, can experience transient or long-term hypopituitarism. These traumatic and vascular conditions likely account for about 5% of hypopituitarism, reflecting the growing prevalence and recognition of these disorders. Long-term periodic endocrine follow-up is indicated because hypothalamic or pituitary dysfunction will develop in 25–40% of these patients.

Hypothalamic Infiltration Disorders Sarcoidosis, histiocytosis X, amyloidosis, and hemochromatosis frequently involve both hypothalamic and pituitary neuronal and neurochemical tracts. Consequently, AVP-D is a common presentation, reported in half of patients with these disorders. Growth retardation is seen if attenuated GH secretion occurs before puberty. Hypogonadotropic hypogonadism and hyperprolactinemia are also common.

Inflammatory Lesions Pituitary damage and subsequent secretory dysfunction can be seen with chronic site infections such as tuberculosis, with opportunistic fungal infections associated with AIDS, and in tertiary syphilis. Other inflammatory processes, such as granulomas and sarcoidosis, should be considered in the differential diagnosis of imaging studies suggestive of a pituitary adenoma. These lesions may cause extensive hypothalamic and pituitary damage, leading to hormone deficiencies.

Cranial Irradiation Cranial irradiation may result in long-term hypothalamic and pituitary dysfunction, especially in children and adolescents, as they are more susceptible to damage after whole-brain or head and neck therapeutic irradiation. The development of subsequent hormonal abnormalities correlates strongly with irradiation dosage and the time interval after completion of radiotherapy. Up to two-thirds of patients ultimately develop hormone insufficiency after a median dose of 50 Gy (5000 rad) directed at the skull base. The development of hypopituitarism occurs over 5–15 years and usually reflects hypothalamic damage rather than primary destruction of pituitary cells. Although the pattern of hormone loss is variable, GH deficiency is most common, followed by gonadotropin, TSH, and ACTH deficiency. When deficiency of one or more hormones is documented, the possibility of diminished reserve of other hormones is likely. Accordingly, anterior pituitary function should be continually evaluated over the long term in previously irradiated patients, and replacement therapy instituted when appropriate (see below).

Lymphocytic Hypophysitis This occurs most often in postpartum women; it usually presents with hyperprolactinemia and MRI evidence of a prominent pituitary mass that often resembles an adenoma, with mildly elevated PRL levels. Pituitary failure caused by diffuse lymphocytic infiltration may be transient or permanent but requires immediate evaluation and treatment. Rarely, isolated pituitary hormone deficiencies have been described, suggesting a selective autoimmune process targeted to specific cell types. Most patients manifest symptoms of progressive mass effects with headache and visual disturbance. The erythrocyte sedimentation rate often is elevated. Because it may be indistinguishable from a pituitary adenoma on MRI, hypophysitis should be considered in a postpartum woman with a newly diagnosed pituitary mass before an unnecessary surgical intervention is undertaken. The inflammatory process often resolves after several months of glucocorticoid treatment, and pituitary function may be restored, depending on the extent of damage.

Immunotherapy and Hypophysitis Pituitary cells express cytotoxic T lymphocyte antigen-4 (CTLA-4), and up to 20% of patients receiving cancer immunotherapy with CTLA-4 inhibitors (e.g., ipilimumab) may develop hypophysitis with heterogeneously associated thyroid, adrenal, islet, and gonadal failure. Hypophysitis is also reported with PD-1/PD-L1 inhibitors (e.g., pembrolizumab and nivolumab) and may show delayed presentation. HLA type DQ0602 is associated with checkpoint inhibitor–associated hypophysitis in 39% of such patients. Pituitary hormone replacement, with or without high-dose glucocorticoids, may be safely tolerated with continued immunotherapy.

Pituitary Apoplexy Acute intrapituitary hemorrhagic vascular events can cause substantial damage to the pituitary and surrounding sellar structures. Pituitary apoplexy may occur spontaneously in a preexisting pituitary adenoma; postpartum (Sheehan's syndrome); or in association with diabetes, hypertension, sickle cell anemia, or acute shock. The hyperplastic enlargement of the pituitary, which occurs normally during pregnancy, increases the risk for hemorrhage and infarction. Apoplexy is an endocrine emergency that may result in severe hypoglycemia, hypotension and shock, central nervous system

(CNS) hemorrhage, and death. Acute symptoms may include severe headache with signs of meningeal irritation, bilateral visual changes, ophthalmoplegia, and, in severe cases, cardiovascular collapse and loss of consciousness. Pituitary computed tomography (CT) or MRI may reveal signs of intratumoral or sellar hemorrhage, with pituitary stalk deviation and compression of pituitary tissue.

Patients with no evident visual loss or impaired consciousness can be observed and managed conservatively with high-dose glucocorticoids. Those with significant or progressive visual loss, cranial nerve palsy, or loss of consciousness require urgent surgical decompression. Visual recovery after sellar surgery is inversely correlated with the length of time after the acute event. Therefore, severe ophthalmoplegia or visual deficits are indications for early surgery. Hypopituitarism is common after apoplexy.

Empty Sella A partial or apparently totally empty sella is often an incidental MRI finding and may sometimes be associated with intracranial hypertension. These patients usually have normal pituitary function, implying that the surrounding rim of pituitary tissue is fully functional. Hypopituitarism, however, may develop insidiously. Pituitary adenomas also may undergo clinically silent infarction and involution with development of a partial or totally empty sella by cerebrospinal fluid (CSF) filling the dural herniation. Rarely, small but functional pituitary adenomas may arise within the rim of normal pituitary tissue, and they are not always visible on MRI.

PRESENTATION AND DIAGNOSIS

The clinical manifestations of hypopituitarism depend on which hormones are lost and the extent of the hormone deficiency (see below). GH deficiency causes growth disorders in children and leads to abnormal body composition in adults. Gonadotropin deficiency causes menstrual disorders and infertility in women and decreased sexual function, infertility, and loss of secondary sexual characteristics in men. TSH and ACTH deficiencies usually develop later in the course of pituitary failure. TSH deficiency causes growth retardation in children and features of hypothyroidism in children and adults. Secondary adrenal insufficiency caused by ACTH deficiency leads to hypocortisolism with relative preservation of mineralocorticoid production. PRL deficiency causes failure of lactation. When lesions involve the posterior pituitary, polyuria and polydipsia reflect loss of AVP secretion. In patients with long-standing pituitary damage, epidemiologic studies document an increased mortality rate, primarily from increased cardiovascular and cerebrovascular disease. Previous head or neck irradiation is also a determinant of increased mortality rates in patients with hypopituitarism, especially from cerebrovascular disease.

LABORATORY INVESTIGATION

Biochemical diagnosis of pituitary insufficiency is made by demonstrating low levels of respective pituitary trophic hormones in the setting of low levels of target organ hormones. For example, low free thyroxine in the setting of a low or inappropriately normal TSH level suggests secondary hypothyroidism. Similarly, a low testosterone level without elevation of gonadotropins suggests hypogonadotropic hypogonadism. Provocative tests may be required to assess pituitary reserve **(Table 391-2)**. GH responses to insulin-induced hypoglycemia, arginine, glucagon, L-dopa, growth hormone–releasing hormone (GHRH), or growth hormone–releasing orally active ghrelin receptor agonist macimorelin can be used to assess GH reserve. Corticotropin-releasing hormone (CRH) administration induces ACTH release, and administration of synthetic ACTH (cosyntropin) evokes adrenal cortisol release as an indirect indicator of pituitary ACTH reserve **(Chap. 398)**. ACTH reserve is most reliably assessed by measuring ACTH and cortisol levels during insulin-induced hypoglycemia. However, this test should be performed cautiously in patients with suspected adrenal insufficiency because of enhanced susceptibility to hypoglycemia and hypotension. Administering insulin to induce hypoglycemia is contraindicated in patients with active coronary artery disease or known seizure disorders.

TREATMENT

Hypopituitarism

Hormone replacement therapy, including glucocorticoids, thyroid hormone, sex steroids, GH, and AVP, is usually safe and free of complications. Treatment regimens that mimic physiologic hormone production allow for maintenance of satisfactory clinical homeostasis. Effective dosage schedules are outlined in **Table 391-3**. Patients in need of glucocorticoid replacement require especially careful dose adjustments during stressful events such as acute illness, dental procedures, trauma, and hospitalization.

DISORDERS OF GROWTH AND DEVELOPMENT

Skeletal Maturation and Somatic Growth The growth plate is dependent on a variety of hormonal stimuli, including GH, insulin-like growth factor (IGF)-1, sex steroids, thyroid hormones, paracrine and circulating growth factors (e.g., fibroblast growth factor family), and cytokines. The growth-promoting process also requires caloric energy, amino acids, vitamins, and trace metals and consumes ~10% of normal energy production. Malnutrition impairs chondrocyte activity, increases GH resistance, and leads to reduced circulating IGF-1 and IGF binding protein (IGFBP)-3 levels.

Linear bone growth rates are very high in infancy and are pituitary dependent. Mean growth velocity is ~6 cm/year in later childhood and usually is maintained within a given range on a standardized percentile chart. Peak growth rates occur during midpuberty when bone age is 12 (girls) or 13 (boys). Secondary sexual development is associated with elevated sex steroids that cause progressive epiphyseal growth plate closure. *Bone age* is delayed in patients with all forms of true GH deficiency or GH receptor defects that result in attenuated GH action.

Short stature may occur as a result of constitutive intrinsic growth defects or because of acquired extrinsic factors that impair growth. In general, delayed bone age in a child with short stature is suggestive of a hormonal or systemic disorder, whereas normal bone age in a short child is more likely to be caused by a genetic cartilage dysplasia or growth plate disorder **(Chap. 425)**.

GH Deficiency in Children Isolated GH deficiency is characterized by short stature, micropenis, increased fat, high-pitched voice, and a propensity to hypoglycemia due to relatively unopposed insulin action. Familial modes of inheritance are seen in at least one-third of these individuals and may be autosomal dominant, recessive, or X-linked. About 10% of children with GH deficiency have mutations in the *GH-N* gene, including gene deletions and a wide range of point mutations. Mutations in transcription factors Pit-1 and Prop-1, which control somatotrope development (see above), result in GH deficiency in combination with other pituitary hormone deficiencies, which may become manifest only in adulthood. The diagnosis of *idiopathic GH deficiency* should be made only after known molecular defects have been rigorously excluded.

GHRH RECEPTOR MUTATIONS Recessive mutations of the GHRH receptor gene in subjects with severe proportionate dwarfism are associated with low basal GH levels that cannot be stimulated by exogenous GHRH, GH-releasing peptide, or insulin-induced hypoglycemia, as well as anterior pituitary hypoplasia. The syndrome exemplifies the importance of the GHRH receptor for determining somatotrope proliferation and hormonal responsiveness.

GH INSENSITIVITY This is caused by defects of GH receptor structure or signaling. Homozygous or heterozygous mutations of the GH receptor are associated with partial or complete GH insensitivity and growth failure (*Laron syndrome*). The diagnosis is based on normal or high GH levels, with decreased circulating GH-binding protein (GHBP), and low IGF-1 levels. Very rarely, defective IGF-1, IGF-1 receptor, or IGF-1 signaling defects are also encountered. *STAT5B* mutations result in both immunodeficiency as well as abrogated GH signaling, leading to short stature with normal or elevated GH levels and low IGF-1 levels. Circulating GH receptor antibodies may rarely cause peripheral GH insensitivity.

TABLE 391-2 Tests of Pituitary Sufficiency

HORMONE	TEST	BLOOD SAMPLES	INTERPRETATION
GH	Insulin tolerance test: Regular insulin (0.05–0.15 U/kg IV)	–30, 0, 30, 60, 120 min for glucose and GH	Glucose <40 mg/dL; GH should be >3 μg/L
	GHRH/L-arginine test: GHRH 1 μg/kg IV and arginine 30 g IV over 30 min	0, 15, 30, 45, 60, 120 min for GH	Normal GH response is BMI dependent: 11 μg/L if BMI <25 kg/m^2, 8 μg/L if BMI 25-20, and 4 μg/L if BMI ≥30 Not available in the United States
	Ghrelin receptor agonist test: 0.5 mg/kg PO	0, 30, 45, 60, 90 min for GH	Normal response is GH >2.8 μg/L
	Glucagon test: 1 mg IM (1.5 mg if body weight >90 kg)	0, 30, 60, 90, 120, 150, 180, 210, 240 for GH	Normal response is GH >3.0 μg/L if BMI <25 kg/m^2 or if BMI 25-30 and low pretest probability, and GH >1.0 μg/L if BMI 25-30 and high pretest probability or if BMI >30
	L-Dopa test: 500 mg PO	0, 30, 60, 120 min for GH	Normal response is GH >3 μg/L
PRL	TRH test: 200–500 μg IV	0, 20, and 60 min for TSH and PRL	Normal PRL is >2 μg/L and increase >200% of baseline
ACTH	Insulin tolerance test: Regular insulin (0.05–0.15 U/kg IV)	–30, 0, 30, 60, 90 min for glucose and cortisol	Glucose <40 mg/dL Cortisol should increase by >7 μg/dL or to >20 μg/dL
	CRH test: 1 μg/kg CRH IV at 8 A.M.	0, 15, 30, 60, 90, 120 min for ACTH and cortisol	Basal ACTH increases 2- to 4-fold and peaks at 20–100 pg/mL Cortisol levels >20–25 μg/dL
	Metyrapone test: Metyrapone (30 mg/kg) at midnight	Plasma 11-deoxycortisol and cortisol at 8 A.M.; ACTH can also be measured	Plasma cortisol should be <4 g/dL to assure an adequate response Normal response is 11-deoxycortisol >7.5 μg/dL or ACTH >75 pg/mL
	Standard ACTH stimulation test: ACTH 1-24 (cosyntropin), 0.25 mg IM or IV	0, 30, 60 min for cortisol and aldosterone	Normal response is cortisol >21 μg/dL and aldosterone response >4 ng/dL above baseline
	Low-dose ACTH test: ACTH 1-24 (cosyntropin), 1 μg IV	0, 30, 60 min for cortisol	Cortisol should be >21 μg/dL
	3-day ACTH stimulation test consists of 0.25 mg ACTH 1-24 given IV over 8 h each day		Cortisol >21 μg/dL
TSH	Basal thyroid function tests: T_4, T_3, TSH	Basal measurements	Low free thyroid hormone levels in the setting of TSH levels that are not appropriately increased indicate pituitary insufficiency
	TRH test: 200–500 μg IV	0, 20, 60 min for TSH and PRL[a]	TSH should increase by >5 mU/L unless thyroid hormone levels are increased
LH, FSH	LH, FSH, testosterone, estrogen	Basal measurements	Basal LH and FSH should be increased in postmenopausal women Low testosterone levels in the setting of low LH and FSH indicate pituitary insufficiency
	GnRH test: GnRH (100 μg) IV	0, 30, 60 min for LH and FSH	In most adults, LH should increase by 10 IU/L and FSH by 2 IU/L Normal responses are variable
Multiple hormones	Combined anterior pituitary test: GHRH (1 μg/kg), CRH (1 μg/kg), GnRH (100 μg), TRH (200 μg) are given IV	–30, 0, 15, 30, 60, 90, 120 min for GH, ACTH, cortisol, LH, FSH, and TSH	Combined or individual releasing hormone responses must be elevated in the context of basal target gland hormone values and may not be uniformly diagnostic (see text)

[a] Evoked PRL response indicates lactotrope integrity.

Abbreviations: T_3, triiodothyronine; T_4, thyroxine; TRH, thyrotropin-releasing hormone. For other abbreviations, see text.

NUTRITIONAL SHORT STATURE Caloric deprivation and malnutrition, uncontrolled diabetes, and chronic renal failure represent secondary causes of abrogated GH receptor function. These conditions also stimulate production of proinflammatory cytokines, which act to exacerbate the block of GH-mediated signal transduction. Children with these conditions typically exhibit features of acquired short stature with normal or elevated GH and low IGF-1 levels.

PSYCHOSOCIAL SHORT STATURE Emotional and social deprivation lead to growth retardation accompanied by delayed speech, discordant hyperphagia, and an attenuated response to administered GH. A nurturing environment restores growth rates.

PRESENTATION AND DIAGNOSIS

Short stature is commonly encountered in clinical practice, and the decision to evaluate these children requires clinical judgment in association with auxologic data and family history. Short stature should be evaluated comprehensively if a patient's height is >3 standard deviations below the mean for age or if the growth rate has decelerated. Skeletal maturation is best evaluated by measuring a radiologic bone age, which is based mainly on the degree of wrist bone growth plate fusion. Final height can be predicted using standardized scales (Bayley-Pinneau or Tanner-Whitehouse) or estimated by adding 6.5 cm (boys) or subtracting 6.5 cm (girls) from the midparental height.

LABORATORY INVESTIGATION

Because GH secretion is pulsatile, GH deficiency is best assessed by examining the response to provocative stimuli, including exercise, insulin-induced hypoglycemia, and other pharmacologic tests that normally increase GH to >7 μg/L in children. Random GH measurements do not distinguish children with normal GH levels from those with true GH deficiency. Adequate adrenal and thyroid hormone replacement should be assured before testing. Age-matched IGF-1 levels are not sufficiently sensitive or specific to make the diagnosis but can be useful to confirm GH deficiency. Pituitary MRI may reveal pituitary mass lesions or structural defects. Molecular analyses for known mutations should be undertaken when the cause of short stature remains cryptic or when additional clinical features suggest a genetic cause.

TABLE 391-3 Hormone Replacement Therapy for Adult Hypopituitarism[a]

HORMONE DEFICIT	HORMONE REPLACEMENT
ACTH	Hydrocortisone (10–20 mg/d in divided doses)
	Cortisone acetate (15–25 mg/d in divided doses)
	Prednisone (5 mg A.M.)
TSH	L-Thyroxine (0.075–0.15 mg daily)
FSH/LH	Males
	Testosterone gel (5–10 g/d)
	Testosterone skin patch (5 mg/d)
	Testosterone enanthate (200 mg IM every 2 weeks)
	Females
	Conjugated estrogen (0.65–1.25 mg qd for 25 days)
	Progesterone (5–10 mg qd) on days 16–25
	Estradiol skin patch (0.025–0.1 mg every week), adding progesterone on days 16–25 if uterus intact
	For fertility: menopausal gonadotropins, human chorionic gonadotropins
GH	Adults: Somatotropin (0.1–1.25 mg SC qd)
	Children: Somatotropin (0.02–0.05 mg/kg per day)
AVP	Intranasal desmopressin (5–20 g twice daily) Oral 300–600 µg qd

[a]All doses shown should be individualized for specific patients and should be reassessed during stress, surgery, or pregnancy. Male and female fertility requirements should be managed as discussed in Chaps. 403 and 404.

Note: For abbreviations, see text.

TABLE 391-4 Features of Adult Growth Hormone Deficiency

Clinical
Impaired quality of life
Decreased energy and drive
Poor concentration
Low self-esteem
Social isolation
Body composition changes
Increased body fat mass
Central fat deposition
Increased waist-to-hip ratio
Decreased lean body mass
Reduced exercise capacity
Reduced maximum O_2 uptake
Impaired cardiac function
Reduced muscle mass
Cardiovascular risk factors
Impaired cardiac structure and function
Abnormal lipid profile
Decreased fibrinolytic activity
Atherosclerosis
Omental obesity
Imaging
Pituitary: mass or structural damage
Bone: reduced bone mineral density
Abdomen: excess omental adiposity
Laboratory
Evoked GH <3 ng/mL
IGF-1 and IGFBP-3 low or normal
Increased LDL cholesterol
Concomitant gonadotropin, TSH, and/or ACTH reserve deficits may be present

Abbreviation: LDL, low-density lipoprotein. For other abbreviations, see text.

TREATMENT

Disorders of Growth and Development

Replacement therapy with recombinant GH (0.02–0.05 mg/kg per day SC) restores growth velocity in GH-deficient children to ~10 cm/year. If pituitary insufficiency is documented, other associated hormone deficits should be corrected, especially adrenal steroids. In selected situations, GH treatment may be combined with strategies to delay puberty (e.g., GnRH agonist) or reduce sex steroids (e.g., aromatase inhibitors) as a means to mitigate sex steroid effect on epiphyseal closure. GH treatment is also moderately effective for accelerating growth rates in children with Turner syndrome and chronic renal failure. Treating psychosocial or constitutional (idiopathic) short stature with GH is not uniformly recommended as these children may only experience modest additive growth, which should be weighed against GH cost and side effect profiles.

In patients with GH insensitivity and growth retardation due to mutations of the GH receptor, treatment with IGF-1 bypasses the dysfunctional GH receptor.

ADULT GH DEFICIENCY

Adult GH deficiency (AGHD) usually is caused by acquired hypothalamic or pituitary somatotrope damage. Acquired pituitary hormone deficiency follows a typical pattern in which loss of adequate GH reserve foreshadows subsequent hormone deficits. The sequential order of hormone loss is usually GH → FSH/LH → TSH → ACTH. Patients previously diagnosed with childhood-onset GH deficiency should be retested as adults to affirm the diagnosis.

PRESENTATION AND DIAGNOSIS

The clinical features of AGHD include changes in body composition, lipid metabolism, and quality of life as well as cardiovascular dysfunction (Table 391-4). Body composition changes are common and include reduced lean body mass, increased fat mass with selective deposition of intraabdominal visceral fat, and increased waist-to-hip ratio. Hyperlipidemia, left ventricular dysfunction, hypertension, and increased plasma fibrinogen levels also may be present. Bone mineral content is reduced, with resultant increased fracture rates. Patients may experience social isolation, depression, and difficulty maintaining gainful employment. Adult hypopituitarism is associated with a threefold increase in cardiovascular mortality rates in comparison to age- and sex-matched controls, and this may be due to GH deficiency, as patients in these studies were replaced with other deficient pituitary hormones.

LABORATORY INVESTIGATION

AGHD is rare, and in light of the nonspecific nature of associated clinical symptoms, patients appropriate for testing should be selected carefully on the basis of well-defined criteria. With few exceptions, testing should be restricted to patients with the following predisposing factors: (1) pituitary surgery, (2) pituitary or hypothalamic tumor or granulomas, (3) history of cranial irradiation, (4) radiologic evidence of a pituitary lesion, and (5) childhood requirement for GH replacement therapy. The transition of a GH-deficient adolescent to adulthood requires retesting to document subsequent AGHD. Up to 20% of patients previously treated for childhood-onset GH deficiency are found to be GH sufficient on repeat testing as adults.

A significant proportion (~25%) of truly GH-deficient adults have low-normal IGF-1 levels. Thus, as in the evaluation of GH deficiency in children, valid age-matched IGF-1 measurements provide a useful index of therapeutic responses but are not sufficiently precise for diagnostic purposes. The most validated test to distinguish pituitary-sufficient patients from those with AGHD is insulin-induced (0.05–0.1 U/kg) hypoglycemia. After glucose reduction to ~40 mg/dL, most individuals experience neuroglycopenic symptoms (Chap. 418), and peak GH release occurs at 60 min and remains elevated for up to 2 h. About 90% of healthy adults exhibit GH responses >5 µg/L; AGHD is defined by a peak GH response to hypoglycemia of <3 µg/L. Although insulin-induced hypoglycemia is safe when performed under appropriate

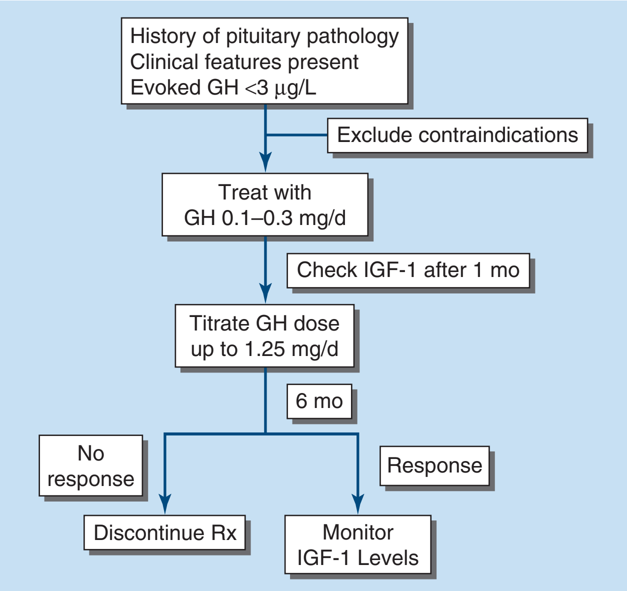

FIGURE 391-1 Management of adult growth hormone (GH) deficiency. IGF, insulin-like growth factor; Rx, treatment.

supervision, it is contraindicated in patients with diabetes, ischemic heart disease, cerebrovascular disease, or epilepsy and in elderly patients. Alternative stimulatory tests include intravenous arginine (30 g), GHRH (1 µg/kg), oral ghrelin receptor agonist (0.5 mg/kg), and glucagon (1 mg). Combinations of these tests may evoke GH secretion in subjects who are not responsive to a single test.

TREATMENT

Adult GH Deficiency

Once the diagnosis of AGHD is unequivocally established, replacement of GH may be indicated. Contraindications to therapy include the presence of an active neoplasm, intracranial hypertension, and uncontrolled diabetes and retinopathy. The starting adult dose of 0.1–0.2 mg/d should be titrated (up to a maximum of 1.25 mg/d) to maintain IGF-1 levels in the mid-normal range for age- and sex-matched controls **(Fig. 391-1)**. Women require higher doses than men, and elderly patients require less GH. Long-term GH maintenance sustains normal IGF-1 levels and is associated with persistent body composition changes (e.g., enhanced lean body mass and lower body fat). High-density lipoprotein cholesterol increases, but total cholesterol and insulin levels may not change significantly. Lumbar spine bone mineral density increases, but this response is gradual (>1 year). Many patients note significant improvement in quality of life when evaluated by standardized questionnaires. The effect of GH replacement on mortality rates in GH-deficient patients is currently the subject of long-term prospective investigation. Recently approved long-acting GH preparations for patients with AGHD require weekly injections. Ideally, dosing should be titrated to achieve normal but not supra-normal IGF-1 levels. Early reports indicate that side effects appear similar to subcutaneous formulations.

About 30% of patients exhibit reversible dose-related fluid retention, joint pain, and carpal tunnel syndrome, and up to 40% exhibit myalgias and paresthesia. Patients receiving insulin require careful monitoring for dosing adjustments, as GH is a potent counterregulatory hormone for insulin action. Patients with type 2 diabetes mellitus may initially develop further insulin resistance. However, glycemic control usually improves with the sustained loss of abdominal fat associated with long-term GH replacement. Headache, increased intracranial pressure, hypertension, and tinnitus occur rarely. Pituitary tumor regrowth and progression of skin lesions or other tumors have not been encountered in long-term surveillance programs with appropriate replacement doses.

ACTH DEFICIENCY

PRESENTATION AND DIAGNOSIS

Secondary adrenal insufficiency occurs as a result of pituitary ACTH deficiency. It is characterized by fatigue, weakness, anorexia, nausea, vomiting, and, occasionally, hypoglycemia. In contrast to primary adrenal failure, hypocortisolism associated with pituitary failure usually is not accompanied by hyperpigmentation or mineralocorticoid deficiency.

ACTH deficiency is commonly due to glucocorticoid withdrawal after treatment-associated suppression of the hypothalamic-pituitary-adrenal (HPA) axis. Isolated ACTH deficiency may occur after surgical resection of an ACTH-secreting pituitary adenoma that has suppressed the HPA axis; this phenomenon is in fact suggestive of a surgical cure. The mass effects of other pituitary adenomas or sellar lesions may lead to ACTH deficiency, usually in combination with other pituitary hormone deficiencies. Partial ACTH deficiency may be unmasked in the presence of an acute medical or surgical illness, when clinically significant hypocortisolism reflects diminished ACTH reserve. Rarely, *TPIT* or *POMC* mutations result in primary ACTH deficiency.

LABORATORY DIAGNOSIS

Inappropriately low ACTH levels in the setting of low cortisol levels are characteristic of diminished ACTH reserve. Low basal serum cortisol levels are associated with blunted cortisol responses to ACTH stimulation and impaired cortisol response to insulin-induced hypoglycemia or testing with metyrapone or CRH. **For a description of provocative ACTH tests, see Chap. 398.**

TREATMENT

ACTH Deficiency

Glucocorticoid replacement therapy improves most features of ACTH deficiency. The total daily dose of hydrocortisone replacement preferably should not exceed 20 mg daily, divided into two or three doses. Prednisone (5 mg each morning) is longer acting and has fewer mineralocorticoid effects than hydrocortisone. Some authorities advocate lower maintenance doses in an effort to avoid cushingoid side effects. Doses should be increased severalfold during periods of acute illness or stress. Patients should wear medical alert bracelets and/or carry identification cards with information about their glucocorticoid requirements.

GONADOTROPIN DEFICIENCY

Hypogonadism is the most common presenting feature of adult hypopituitarism even when other pituitary hormones are also deficient. It is often a harbinger of hypothalamic or pituitary lesions that impair GnRH production or delivery through the pituitary stalk. As noted below, hypogonadotropic hypogonadism is a common presenting feature of hyperprolactinemia.

A variety of inherited and acquired disorders are associated with *isolated hypogonadotropic hypogonadism* **(Chap. 403)**. Hypothalamic defects associated with GnRH deficiency include Kallmann syndrome and mutations in more than a dozen genes that regulate GnRH neuron migration, development, and function (see above). Mutations in *GPR54, DAX1, NR5A1*, kisspeptin, the GnRH receptor, and the LHβ or FSHβ subunit genes also cause pituitary gonadotropin deficiency. Acquired forms of GnRH deficiency leading to hypogonadotropism are seen in association with anorexia nervosa, stress, starvation, and extreme exercise but also may be idiopathic. Hypogonadotropic hypogonadism in these disorders is reversed by removal of the stressful stimulus or by caloric replenishment.

PRESENTATION AND DIAGNOSIS

In premenopausal women, hypogonadotropic hypogonadism presents as diminished ovarian function leading to oligomenorrhea or amenorrhea, infertility, decreased vaginal secretions, decreased libido, and breast atrophy. In hypogonadal adult men, secondary testicular failure is associated with decreased libido and potency, infertility, decreased muscle mass with weakness, reduced beard and body hair growth, soft testes, and characteristic fine facial wrinkles. Osteoporosis occurs in both untreated hypogonadal women and men.

LABORATORY INVESTIGATION

Central hypogonadism is associated with low or inappropriately normal serum gonadotropin levels in the setting of low sex hormone concentrations (testosterone in men, estradiol in women). Because gonadotropin secretion is pulsatile, valid assessments may require repeated measurements or the use of pooled serum samples. Men have reduced sperm counts.

Intravenous GnRH (100 μg) stimulates gonadotropes to secrete LH (which peaks within 30 min) and FSH (which plateaus during the ensuing 60 min). Normal responses vary according to menstrual cycle stage, age, and sex of the patient. Generally, LH levels increase about threefold, whereas FSH responses are less pronounced. In the setting of gonadotropin deficiency, a normal gonadotropin response to GnRH indicates intact pituitary gonadotrope function and suggests a hypothalamic abnormality. An absent response, however, does not reliably distinguish pituitary from hypothalamic causes of hypogonadism. For this reason, GnRH testing usually adds little to the information gained from baseline evaluation of the hypothalamic-pituitary-gonadotrope axis except in cases of isolated GnRH deficiency (e.g., Kallmann syndrome).

MRI examination of the sellar region and assessment of other pituitary functions usually are indicated in patients with documented central hypogonadism.

TREATMENT

Gonadotropin Deficiency

In males, testosterone replacement is necessary to achieve and maintain normal growth and development of the external genitalia, secondary sex characteristics, male sexual behavior, and androgenic anabolic effects, including maintenance of muscle function and bone mass. Testosterone may be administered by intramuscular injections every 1–4 weeks or by using skin patches or testosterone gels **(Chap. 403)**. Gonadotropin injections (hCG or human menopausal gonadotropin [hMG]) over 12–18 months are used to restore fertility. Pulsatile GnRH therapy (25–150 ng/kg every 2 h), administered by a subcutaneous infusion pump, is also effective for treatment of hypothalamic hypogonadism when fertility is desired.

In premenopausal women, cyclical replacement of estrogen and progesterone maintains secondary sexual characteristics and integrity of genitourinary tract mucosa and prevents premature osteoporosis **(Chap. 404)**. Gonadotropin therapy is used for ovulation induction. Follicular growth and maturation are initiated using hMG or recombinant FSH; hCG or human luteinizing hormone is subsequently injected to induce ovulation. As in men, pulsatile GnRH therapy can be used to treat hypothalamic causes of gonadotropin deficiency.

ARGININE VASOPRESSIN DEFICIENCY

See Chap. 393 for diagnosis and treatment of AVP-D.

FURTHER READING

Fleseriu M et al: Hormonal replacement in hypopituitarism in adults: An Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 101:3888, 2016.

Gregory LC, Dattani MT: The molecular basis of congenital hypopituitarism and related disorders. J Clin Endocrinol Metab 105:e2103, 2020.

Melmed S: Pathogenesis and diagnosis of growth hormone deficiency in adults. N Engl J Med 380:2551, 2019.

Miller BS et al: Long-acting growth hormone preparations-current status and future considerations. J Clin Endocrinol Metab 105:e2121, 2020.

Prodam F et al: Insights into non-classic and emerging causes of hypopituitarism. Nat Rev Endocrinol 17:114, 2021.

Quandt Z et al: Spectrum of clinical presentations, imaging findings, and HLA types in immune checkpoint inhibitor-induced hypophysitis. J Endocr Soc 7:bvad012, 2023.

Tanriverdi F et al: Pituitary dysfunction after traumatic brain injury: A clinical and pathophysiological approach. Endocr Rev 36:305, 2015.

Yamamoto M et al: Autoimmune pituitary disease: New concepts with clinical implications. Endocr Rev 41:261, 2020.

392

Pituitary Tumor Syndromes

Shlomo Melmed, J. Larry Jameson

HYPOTHALAMIC, PITUITARY, AND OTHER SELLAR MASSES

EVALUATION OF SELLAR MASSES

Local Mass Effects Clinical manifestations of sellar lesions vary, depending on the anatomic location of the mass and the direction of its extension **(Table 392-1)**. The dorsal sellar diaphragm presents the least resistance to soft tissue expansion from the sella; consequently, pituitary adenomas frequently extend in a suprasellar direction. Bony invasion may occur as well, especially through the sellar floor to the sphenoid sinus **(Fig. 392-1)**.

Headaches are common features of small intrasellar tumors, even with no demonstrable suprasellar extension. Because of the confined nature of the pituitary, small changes in intrasellar pressure stretch the dural plate; however, headache severity correlates poorly with adenoma size or extension.

Suprasellar extension can lead to visual loss by several mechanisms, the most common being compression of the optic chiasm. Rarely, direct invasion of the optic nerves or obstruction of cerebrospinal fluid (CSF) flow leading to secondary visual disturbances can occur. Pituitary stalk compression by a hormonally active or inactive intrasellar mass may compress the portal vessels, disrupting pituitary access to hypothalamic hormones and dopamine; this results in early hyperprolactinemia and later concurrent loss of other pituitary hormones. This "stalk section" phenomenon may also be caused by trauma, whiplash injury with posterior clinoid stalk compression, or skull base fractures. Lateral mass invasion may impinge on the cavernous sinus and compress its neural contents, leading to cranial nerve III, IV, and VI palsies as well as effects on the ophthalmic and maxillary branches of the fifth cranial nerve **(Chap. 452)**. Patients may present with diplopia, ptosis, ophthalmoplegia, and decreased facial sensation, depending on the extent of neural damage. Extension into the sphenoid sinus indicates that the pituitary mass has eroded through the sellar floor (Fig. 392-1). Aggressive tumors rarely invade the palate roof and cause nasopharyngeal obstruction, infection, and CSF leakage. Temporal and frontal lobe involvement may rarely lead to uncinate seizures, personality disorders, and anosmia. Direct hypothalamic encroachment by an invasive pituitary mass may cause important metabolic sequelae, including precocious puberty or hypogonadism, arginie vasopressin deficiency (AVP-D), sleep disturbances, dysthermia, and appetite disorders.

Magnetic Resonance Imaging Sagittal and coronal T1-weighted magnetic resonance imaging (MRI) before and after administration of gadolinium allows precise visualization of the pituitary gland with

TABLE 392-1 Features of Sellar Mass Lesions[a]

IMPACTED STRUCTURE	CLINICAL IMPACT
Pituitary	Hypogonadism Hypothyroidism Growth failure, adult growth hormone deficiency Hypoadrenalism Hyperprolactinema (stalk compression)
Optic chiasm	Loss of red perception Bitemporal hemianopia Superior or bitemporal field defect Scotoma Blindness
Hypothalamus	Temperature dysregulation Appetite and thirst disorders Obesity Arginine vasopression deficiency Sleep disorders Behavioral dysfunction Autonomic dysfunction
Cavernous sinus	Ophthalmoplegia with or without ptosis or diplopia Facial numbness
Frontal lobe	Personality disorder Anosmia
Brain	Headache Hydrocephalus Psychosis Dementia Laughing seizures

[a]As the intrasellar mass expands, it first compresses intrasellar pituitary tissue, then usually invades dorsally through the dura to lift the optic chiasm or laterally to the cavernous sinuses. Bony erosion is rare, as is direct brain compression. Microadenomas may present with headache.

clear delineation of the hypothalamus, pituitary stalk, pituitary tissue and surrounding suprasellar cisterns, cavernous sinuses, sphenoid sinus, and optic chiasm. Pituitary gland height ranges from 6 mm in children to 8 mm in adults; during pregnancy and puberty, the height may reach 10–12 mm. The upper aspect of the adult pituitary is flat or slightly concave, but in adolescent and pregnant individuals, this surface may be convex, reflecting physiologic pituitary enlargement. The stalk should be midline and vertical.

Anterior pituitary gland soft tissue consistency is slightly heterogeneous on MRI, and signal intensity resembles that of brain matter on T1-weighted imaging **(Fig. 392-2)**. Adenoma density is usually lower than that of surrounding normal tissue on T1-weighted imaging, and the signal intensity increases with T2-weighted images. Computed tomography (CT) scan is reserved to define the extent of bony erosion or the presence of calcification.

Sellar masses are encountered commonly as incidental findings on MRI, and most are pituitary adenomas (incidentalomas). In the absence of hormone hypersecretion, these small intrasellar lesions can be monitored safely with MRI, which is performed annually and then less often if there is no evidence of further growth. Resection should be considered for incidentally discovered larger macroadenomas, because about one-third become invasive or cause local pressure effects. If hormone hypersecretion is identified, specific therapies are indicated as described below. When larger masses (>1 cm) are encountered, they should also be distinguished from nonadenomatous lesions. Meningiomas often are associated with bony hyperostosis; craniopharyngiomas may have calcifications and are usually hypodense, whereas gliomas are hyperdense on T2-weighted images.

Ophthalmologic Evaluation Because optic tracts may be contiguous to an expanding pituitary mass, reproducible visual field assessment using perimetry techniques should be performed on all patients with sellar mass lesions that impinge the optic chiasm **(Chap. 34)**. Bitemporal hemianopia, often more pronounced superiorly, is observed classically. It occurs because nasal ganglion cell fibers, which cross in the optic chiasm, are especially vulnerable to compression of the ventral optic chiasm. Occasionally, homonymous hemianopia occurs from postchiasmal compression or monocular temporal field loss from prechiasmal compression. Invasion of the cavernous sinus can produce diplopia from ocular motor nerve palsy. Early diagnosis reduces the risk of optic atrophy, vision loss, or eye misalignment.

Laboratory Investigation The presenting clinical features of functional pituitary adenomas (e.g., acromegaly, prolactinoma, or Cushing's disease) should guide the laboratory studies **(Table 392-2)**. However, for a sellar mass with no obvious clinical features of hormone excess, laboratory studies are geared toward determining the nature of the tumor and assessing the possible presence of hypopituitarism. When a pituitary adenoma is suspected based on MRI, initial hormonal evaluation usually includes (1) basal prolactin (PRL); (2) insulin-like growth factor (IGF)-1; (3) 24-h urinary free cortisol (UFC) and/or overnight oral dexamethasone (1 mg) suppression test; (4) α subunit, follicle-stimulating hormone (FSH), and luteinizing hormone (LH); and (5) thyroid function tests. Additional hormonal evaluation may be indicated based on the results of these tests. Pending more detailed assessment of hypopituitarism, a menstrual history, measurement of testosterone and 8 A.M. cortisol levels, and thyroid function tests usually identify patients with pituitary hormone deficiencies that require hormone replacement before further testing or surgery **(Chap. 391)**.

Histologic Evaluation Immunohistochemical staining of pituitary tumor specimens obtained at transsphenoidal surgery for hormones as well as cell-type specific transcription factors confirms clinical and laboratory studies and provides a histologic diagnosis when hormone studies are equivocal and in cases of clinically nonfunctioning tumors.

TREATMENT

Hypothalamic, Pituitary, and Other Sellar Masses

OVERVIEW

Successful management of sellar masses requires accurate diagnosis as well as selection of optimal therapeutic modalities. Most pituitary tumors are benign and slow growing. Clinical features result from local mass effects and hormonal hyper- or hyposecretion syndromes caused directly by the adenoma or occurring as a consequence of treatment. Thus, lifelong management and follow-up are necessary for these patients.

MRI with gadolinium enhancement for pituitary visualization, new advances in transsphenoidal surgery and in stereotactic radiotherapy, and novel therapeutic agents have improved pituitary tumor management. The goals of pituitary tumor treatment include normalization of excess pituitary secretion, amelioration of symptoms and signs of hormonal hypersecretion syndromes, and shrinkage or ablation of large tumor masses with relief of adjacent structure compression. Residual anterior pituitary function should be preserved during treatment and sometimes can be restored by removing the tumor mass. Ideally, adenoma recurrence should be prevented.

TRANSSPHENOIDAL SURGERY

Transsphenoidal resection is the desired surgical approach for pituitary tumors, except for the rare invasive suprasellar mass surrounding the frontal or middle fossa or the optic nerves or invading posteriorly behind the clivus, which may require transcranial approaches. Intraoperative microscopy facilitates visual distinction between adenomatous and normal pituitary tissue as well as microdissection of small tumors that may not be visible by MRI **(Fig. 392-3)**. Endoscopic techniques with three-dimensional intraoperative localization enable better visualization and access to tumor tissue. Transsphenoidal surgery also avoids cranial invasion and manipulation of brain tissue required by subfrontal surgical

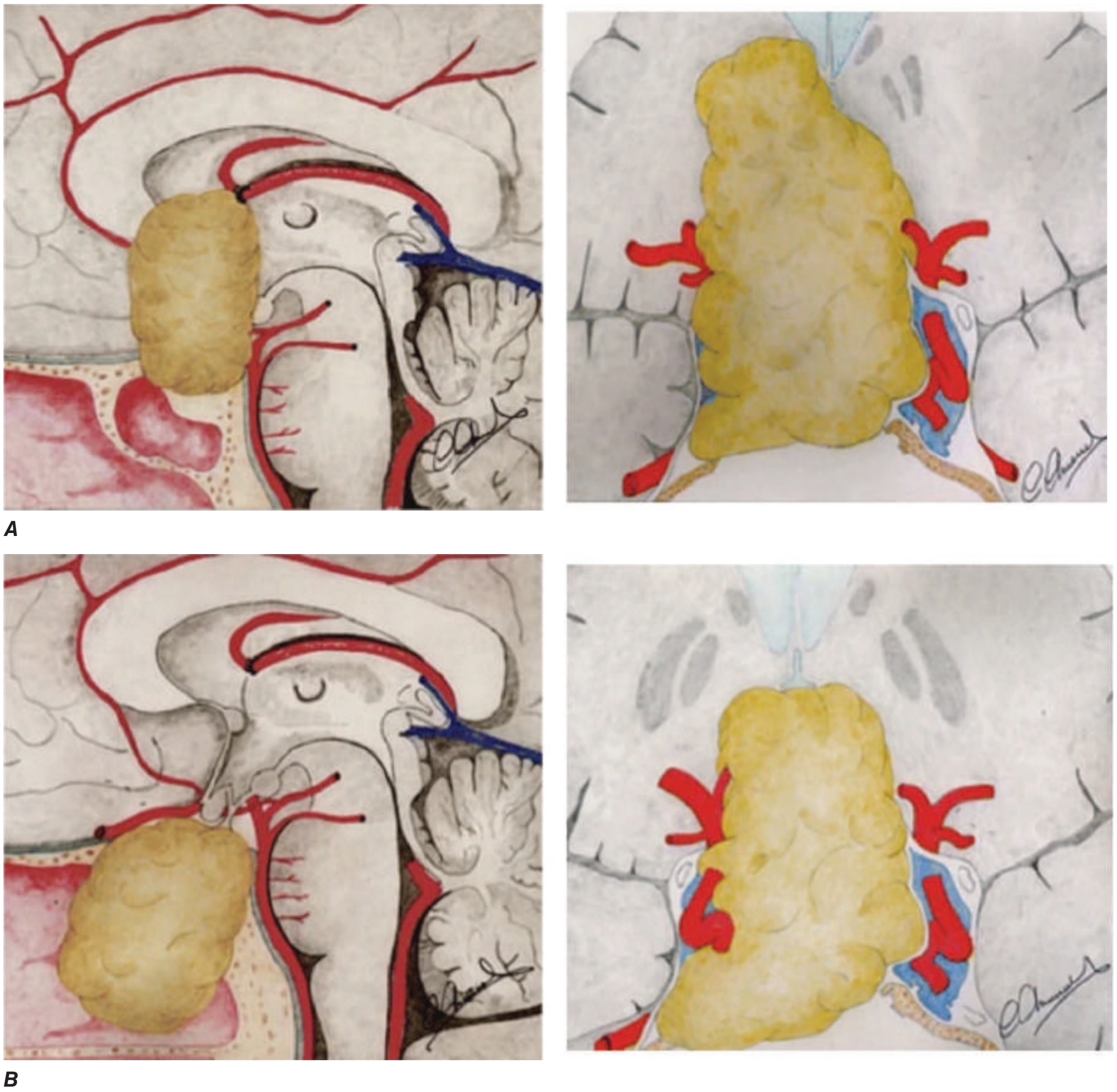

FIGURE 392-1 Expanding pituitary mass. Pituitary mass expansion may ***(A)*** impinge vital soft tissue structures and ***(B)*** invade the sphenoid sinus. *(Reproduced with permission from P Cappabianca et al: Size does not matter. The intrigue of giant adenomas: a true surgical challenge. Acta Neurochir (Wien) 156:2217, 2014.)*

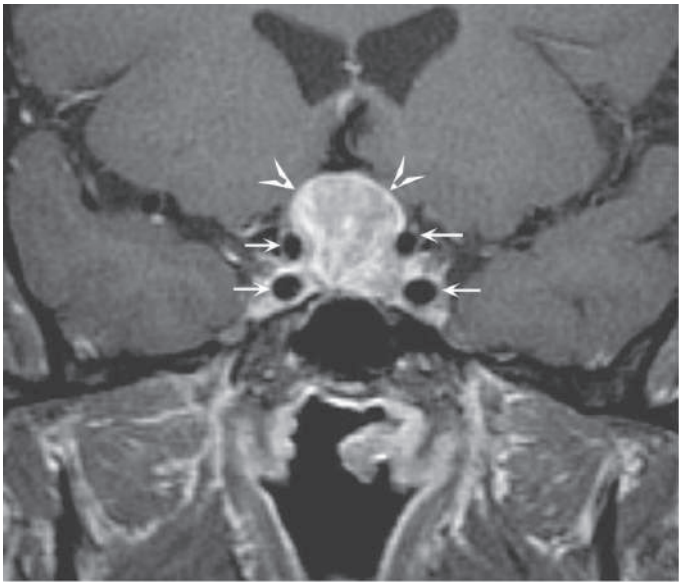

FIGURE 392-2 Pituitary adenoma. Coronal T1-weighted postcontrast magnetic resonance image shows a homogeneously enhancing mass (*arrowheads*) in the sella turcica and suprasellar region compatible with a pituitary adenoma; the *small arrows* outline the carotid arteries.

approaches. Individual surgical experience is a major determinant of outcome efficacy with these techniques.

In addition to correction of hormonal hypersecretion, pituitary surgery is indicated for mass lesions that impinge on surrounding structures. Surgical decompression and resection are required for an expanding pituitary mass, which may be asymptomatic or accompanied by persistent headache, progressive visual field defects, cranial nerve palsies, hydrocephalus, and, occasionally, intrapituitary hemorrhage and apoplexy. Transsphenoidal surgery rarely is used for pituitary tissue biopsy to establish a histologic diagnosis. Whenever possible, the pituitary mass lesion should be selectively excised; normal pituitary tissue should be manipulated or resected only when critical for effective mass dissection. Nonselective hemihypophysectomy or total hypophysectomy may be indicated if no hypersecreting mass lesion is clearly discernible, multifocal lesions are present, or the remaining nontumorous pituitary tissue is obviously necrotic. This strategy, however, increases the likelihood of postoperative hypopituitarism and the need for lifelong hormone replacement.

Preoperative mass effects, including visual field defects and compromised pituitary function, may be reversed by surgery, particularly when the deficits are not long-standing. For large and invasive tumors, it is necessary to determine the optimal balance between maximal tumor resection and preservation of anterior pituitary hormonal function, especially for preserving growth and

TABLE 392-2 Screening Tests for Functional Pituitary Adenomas

	TEST	COMMENTS
Acromegaly	Serum IGF-1 Oral glucose tolerance test with GH obtained at 0, 30, and 60 min	Interpret IGF-1 relative to age- and sex-matched controls Normal subjects should suppress growth hormone to <1 μg/L
Prolactinoma	Serum PRL	Exclude medications MRI of the sella should be ordered if PRL is elevated
Cushing's disease	24-h urinary free cortisol Dexamethasone (1 mg) at 11 P.M. and fasting plasma cortisol measured at 8 A.M. Late night salivary cortisol ACTH assay CRH stimulation test with measurements of cortisol and ACTH from peripheral and/or petrosal sinus blood	Ensure urine collection is total and accurate Normal subjects suppress to <5 μg/dL Distinguishes adrenal adenoma (ACTH suppressed) from ectopic ACTH or Cushing's disease (ACTH normal or elevated) The CRH test is used primarily to distinguish pituitary adenomas from ectopic ACTH sources
Gonadotropinoma	Baseline FSH, LH, free α subunit, ovarian hyperstimulation, estrogen (females), testosterone (males) TRH stimulation test with assays for LH, FSH, free α subunit, free LHβ, free FSHβ subunits	Rare; more commonly nonfunctioning adenomas Consider screening for hypopituitarism Some gonadotropinomas exhibit an inappropriate gonadotropin response to TRH
TSH-producing adenoma	Free T_4, free T_3, TSH, free α subunit	Key feature is an inappropriately normal or high TSH in the setting of elevated free T_4 and T_3

Abbreviations: ACTH, adrenocorticotropin hormone; CRH, corticotropin-releasing hormone; FSH, follicle-stimulating hormone; GH, growth hormone; IGF-1, insulin-like growth factor 1; LH, luteinizing hormone; MRI, magnetic resonance imaging; PRL, prolactin; TSH, thyroid-stimulating hormone.

reproductive function in younger patients. Tumor invasion outside the sella is rarely amenable to surgical cure, and the surgeon must judge the risk-versus-benefit ratio of extensive tumor resection.

Side Effects Tumor size, the degree of invasiveness, and experience of the surgeon largely determine the incidence of surgical complications. Operative mortality rate is ~1%. Transient AVP-D and hypopituitarism occur in up to 20% of patients. Permanent AVP-D, cranial nerve damage, nasal septal perforation, or visual disturbances may be encountered in up to 10% of patients. CSF leaks occur in 4% of patients. Less common complications include carotid artery injury, loss of vision, hypothalamic damage, and meningitis. Permanent side effects are rare after surgery for microadenomas.

RADIATION

Radiation is used either as a primary therapy for pituitary or parasellar masses or, more commonly, as an adjunct to surgery or medical therapy. Focused megavoltage irradiation is achieved by precise MRI localization, using a high-voltage linear accelerator and accurate isocentric rotational arcing. A major determinant of accurate irradiation is reproduction of the patient's head position during multiple visits and maintenance of absolute head immobility. A total of <50 Gy (5000 rad) is given as 180 cGy (180 rad) fractions divided over ~6 weeks. Stereotactic radiosurgery delivers a large single high-energy dose from a cobalt-60 source (Gamma Knife), linear accelerator, or cyclotron. Long-term effects of Gamma Knife surgery appear to be similar to those encountered with conventional radiation. Proton beam therapy is available in some centers and provides concentrated radiation doses within a localized region.

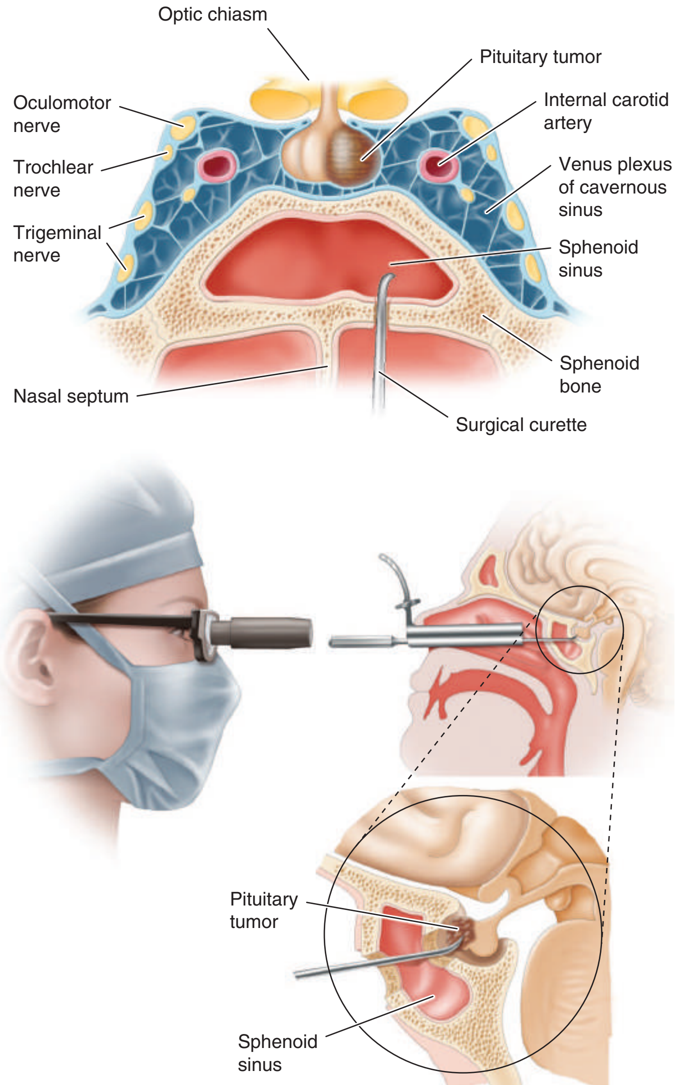

FIGURE 392-3 Transsphenoidal resection of pituitary mass via the endonasal approach.

The role of radiation therapy in pituitary tumor management depends on the nature and anatomic location of the tumor, the age of the patient, and the availability of surgical and radiation expertise. Because of its relatively slow onset of action, radiation therapy is usually reserved for postsurgical management. As an adjuvant to surgery, radiation is used to treat residual tumor in an attempt to prevent persistent growth or recurrence. Irradiation offers the only means for potentially ablating significant postoperative residual nonfunctioning tumor tissue. By contrast, PRL-, growth hormone (GH)–, adrenocorticotropin hormone (ACTH)–, and thyrotropin (thyroid-stimulating hormone [TSH])–secreting residual tumor tissues are amenable to medical therapy.

Side Effects In the short term, radiation may cause transient nausea and weakness. Alopecia and loss of taste and smell may be more long-lasting. Failure of pituitary hormone synthesis is common in patients who have undergone head and neck or pituitary-directed irradiation. More than 50% of patients develop loss of GH, ACTH, TSH, and/or gonadotropin secretion within 10 years, usually due to hypothalamic damage. Lifelong follow-up with testing of anterior pituitary hormone reserve is therefore required after radiation treatment. Optic nerve damage with impaired vision due to optic neuritis is reported in ~2% of patients who undergo pituitary irradiation. Cranial nerve damage is uncommon now that radiation doses are <2 Gy (200 rad) at any one treatment session and the maximum dose is <50 Gy (5000 rad). The use of stereotactic radiotherapy reduces the risk of damage to adjacent structures. Conventional radiotherapy for pituitary tumors has been associated with adverse mortality rates, mainly from cerebrovascular disease. The cumulative risk of developing a secondary tumor after conventional radiation is 1.3% after 10 years and 1.9% after 20 years.

MEDICAL

Medical therapy for pituitary tumors is highly specific and depends on tumor type. For prolactinomas, dopamine agonists are the treatment of choice. For acromegaly, somatostatin receptor ligands (SRLs) and a GH receptor antagonist are indicated. For TSH-secreting tumors, SRLs and occasionally dopamine agonists are indicated. ACTH-secreting tumors may respond to SRLs, and adrenal-directed therapy may also be of benefit. Nonfunctioning tumors are generally not responsive to medications and require surgery and/or irradiation.

SELLAR MASSES

Sellar masses may arise from brain, hypothalamic, or pituitary tissues. Each exhibit features related to the lesion location but also unique to the specific etiology. Unique MRI characteristics inform the differential diagnosis of pituitary masses **(Fig. 392-4)**.

Lesions involving the anterior and preoptic hypothalamic regions cause paradoxical vasoconstriction, tachycardia, and hyperthermia. Acute hyperthermia usually is due to a hemorrhagic insult, but poikilothermia may also occur. Central disorders of thermoregulation result from posterior hypothalamic damage. The *periodic hypothermia syndrome* is characterized by episodic attacks of rectal temperatures <30°C (86°F), sweating, vasodilation, vomiting, and bradycardia **(Chap. 477)**. Damage to the ventromedial hypothalamic nuclei by craniopharyngiomas, hypothalamic trauma, or inflammatory disorders may be associated with *hyperphagia* and *obesity*. This region appears to contain an energy-satiety center where melanocortin receptors are influenced by leptin, insulin, pro-opiomelanocortin (POMC) products, and gastrointestinal peptides **(Chap. 413)**. Polydipsia and hypodipsia are associated with damage to central osmoreceptors located in preoptic nuclei **(Chap. 393)**. Slow-growing hypothalamic lesions can cause increased somnolence and disturbed sleep cycles as well as obesity, hypothermia, and emotional outbursts. Lesions of the central hypothalamus may stimulate sympathetic neurons, leading to elevated serum catecholamine and cortisol levels. These patients are predisposed to cardiac arrhythmias, hypertension, and gastric erosions.

Craniopharyngiomas are benign, suprasellar cystic masses that present with headaches, visual field deficits, and variable degrees of hypopituitarism. They are derived from Rathke's pouch and arise near the pituitary stalk, commonly extending into the suprasellar cistern. Craniopharyngiomas are often large, cystic, and locally invasive. Many are partially calcified, exhibiting a characteristic appearance on skull x-ray and CT images. More than half of all patients present before age 20, usually with signs of increased intracranial pressure, including headache, vomiting, papilledema, and hydrocephalus. Associated symptoms include visual field abnormalities, personality changes and cognitive deterioration, cranial nerve damage, sleep difficulties, and weight gain accompanied by features of the metabolic syndrome. Hypopituitarism is documented in ~90%, and AVP-D occurs in ~10% of patients. About half of affected children present with growth retardation. MRI is generally superior to CT for evaluating cystic structure and tissue components of craniopharyngiomas. CT is useful to define calcifications and evaluate invasion into surrounding bony structures and sinuses.

Treatment usually involves transcranial or transsphenoidal surgical resection followed by postoperative radiation of residual tumor. Surgery alone is curative in less than half of patients because of recurrences due to adherence to vital structures or because of small tumor deposits in the hypothalamus or brain parenchyma. The goal of surgery is to remove as much tumor as possible without risking complications associated with efforts to remove firmly adherent or inaccessible tissue. In the absence of radiotherapy, ~75% of craniopharyngiomas recur, and 10-year survival is <50%. In patients with incomplete resection, radiotherapy improves 10-year survival to 70–90% but is associated with increased risk of secondary malignancies. Most patients require lifelong pituitary hormone replacement. As some craniopharyngiomas (particularly papillary) are associated with activated *BRAF* V600E mutations, use of BRAF inhibitors (dabrafenib or vemurafenib) either alone or in combination with MEK inhibitors (trametinib or cobimetinib) has resulted in long-term growth responses in some patients.

Developmental failure of Rathke's pouch obliteration may lead to *Rathke's cysts*, which are small (<5 mm) cysts entrapped by squamous epithelium and are found in ~20% of individuals at autopsy. Although

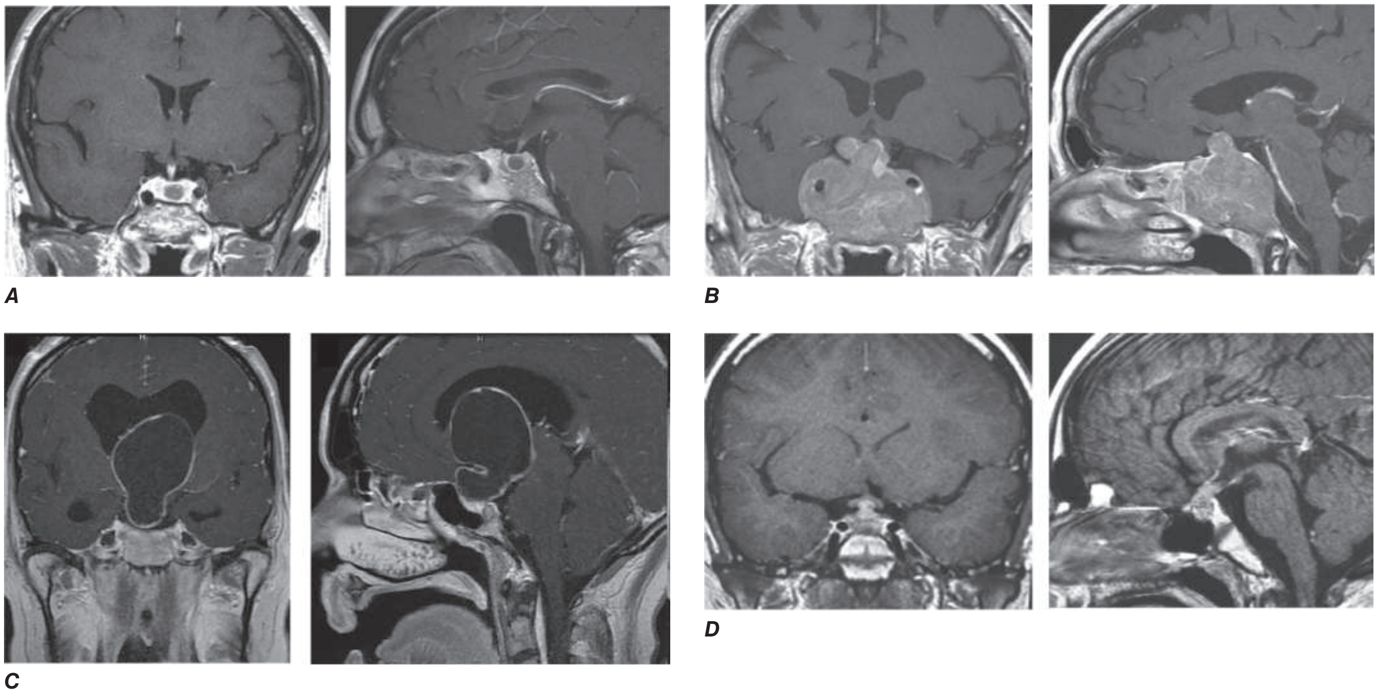

FIGURE 392-4 Imaging differential diagnosis of sellar masses. ***A.*** Microadenoma. ***B.*** Macroadenoma. ***C.*** Craniopharyngioma. ***D.*** Hypophysitis with stalk thickening. *(A, B, D: Used with permission from Vivien Bonert, MD. C: Reproduced with permission from Muller HL: Childhood craniopharyngioma. Recent advances in diagnosis, treatment and follow-up. Horm Res 69:193, 2008.)*

Rathke's cysts do not usually grow and are often diagnosed incidentally, about a third present in adulthood with compressive symptoms, AVP-D, and hyperprolactinemia due to stalk compression. Rarely, hydrocephalus develops. The diagnosis is suggested preoperatively by visualizing the cyst wall on MRI, which distinguishes these lesions from craniopharyngiomas. Cyst contents range from CSF-like fluid to mucoid material. *Arachnoid cysts* are rare and generate an MRI image that is isointense with CSF.

Sellar chordomas usually present with bony clival erosion, local invasiveness, and, on occasion, calcification. Normal pituitary tissue may be visible on MRI, distinguishing chordomas from aggressive pituitary adenomas. Mucinous material may be obtained by fine-needle aspiration.

Meningiomas arising in the sellar region may be difficult to distinguish from nonfunctioning pituitary adenomas. Meningiomas typically enhance on MRI and may show evidence of calcification or bony erosion. Meningiomas may cause compressive symptoms.

Histiocytosis X includes a variety of syndromes associated with foci of eosinophilic granulomas. AVP-D, exophthalmos, and punched-out lytic bone lesions (*Hand-Schüller-Christian disease*) are associated with granulomatous lesions visible on MRI, as well as a characteristic axillary skin rash. Rarely, the pituitary stalk may be involved.

Pituitary metastases occur in ~3% of cancer patients. Bloodborne metastatic deposits are found almost exclusively in the posterior pituitary. Accordingly, AVP-D can be a presenting feature of lung, gastrointestinal, breast, and other pituitary metastases. About half of pituitary metastases originate from breast cancer; ~25% of patients with metastatic breast cancer have such deposits. Rarely, pituitary stalk involvement results in anterior pituitary insufficiency. The MRI diagnosis of a metastatic lesion may be difficult to distinguish from an aggressive pituitary adenoma; the diagnosis may require histologic examination of excised tumor tissue. Primary or metastatic lymphoma, leukemias, and plasmacytomas also occur within the sella.

Hypothalamic hamartomas and *gangliocytomas* may arise from astrocytes, oligodendrocytes, and neurons with varying degrees of differentiation. These tumors may overexpress hypothalamic neuropeptides, including gonadotropin-releasing hormone (GnRH), growth hormone–releasing hormone (GHRH), and corticotropin-releasing hormone (CRH). With GnRH-producing tumors, children present with precocious puberty, psychomotor delay, and laughing-associated seizures. Medical treatment of GnRH-producing hamartomas with long-acting GnRH analogues effectively suppresses gonadotropin secretion and controls premature pubertal development. Rarely, hamartomas also are associated with craniofacial abnormalities; imperforate anus; cardiac, renal, and lung disorders; and pituitary failure as features of *Pallister-Hall syndrome*, which is caused by mutations in the carboxy terminus of the *GLI3* gene. Hypothalamic hamartomas are often contiguous with the pituitary, and preoperative MRI diagnosis may not be possible. Histologic evidence of hypothalamic neurons in tissue resected at transsphenoidal surgery may be the first indication of a primary hypothalamic lesion.

Hypothalamic gliomas and *optic gliomas* occur mainly in childhood and usually present with visual loss. Adults have more aggressive tumors; about a third are associated with neurofibromatosis.

Brain germ cell tumors may arise within the sellar region. They include *dysgerminomas*, which frequently are associated with AVP-D and visual loss. They rarely metastasize. *Germinomas, embryonal carcinomas*, *teratomas*, and *choriocarcinomas* may arise in the parasellar region and produce human chorionic gonadotropin (hCG). These germ cell tumors present with precocious puberty, AVP-D, visual field defects, and thirst disorders. Many patients are GH deficient with short stature.

PITUITARY ADENOMAS AND HYPERSECRETION SYNDROMES

Pituitary adenomas are the most common cause of pituitary hormone hypersecretion and hyposecretion syndromes in adults. They account for ~15% of all intracranial neoplasms and have been identified with a population prevalence of ~80/100,000. At autopsy, up to one-quarter of all pituitary glands harbor an unsuspected microadenoma (<10 mm diameter). Similarly, pituitary imaging detects small clinically inapparent pituitary lesions in at least 10% of individuals.

Pathogenesis Pituitary adenomas are benign neoplasms that arise from one of the five anterior pituitary cell types. The clinical and biochemical phenotypes of pituitary adenomas depend on the cell type from which they are derived. Thus, tumors arising from lactotrope (PRL), somatotrope (GH), corticotrope (ACTH), thyrotrope (TSH), or gonadotrope (LH, FSH) cells hypersecrete their respective hormones **(Table 392-3)**. Plurihormonal tumors express various combinations of GH, PRL, TSH, ACTH, or the glycoprotein hormone α or β subunits. They may be diagnosed by careful immunocytochemistry of specific hormone and transcription factor expression or may manifest as clinical syndromes that combine features of these hormonal hypersecretory syndromes. Morphologically, these tumors may arise from a single polysecreting cell type or include cells with mixed function within the same tumor.

Hormonally active tumors are characterized by autonomous hormone secretion with diminished feedback responsiveness to physiologic inhibitory pathways. Hormone production does not always correlate with tumor size. Small hormone-secreting adenomas may cause significant clinical perturbations, whereas larger adenomas that produce less hormone may be clinically silent and remain undiagnosed (if no central compressive effects occur). About one-third of all adenomas are clinically nonfunctioning and produce no distinct clinical hypersecretory syndrome. Most of them arise from gonadotrope cells and may secrete small amounts of α- and β-glycoprotein hormone subunits or, very rarely, intact circulating gonadotropins. True pituitary carcinomas with documented extracranial metastases are exceedingly rare.

Almost all pituitary adenomas are monoclonal in origin, implying the acquisition of one or more somatic mutations that confer a selective growth advantage. Consistent with their clonal origin, complete surgical resection of small pituitary adenomas usually cures hormone hypersecretion. Nevertheless, hypothalamic hormones such as GHRH and CRH also enhance mitotic activity of their respective pituitary target cells in addition to their role in pituitary hormone regulation. Thus, patients who harbor rare abdominal or chest tumors that produce ectopic GHRH or CRH may present with somatotrope or corticotrope hyperplasia with GH or ACTH hypersecretion.

TABLE 392-3 Classification of Pituitary Adenomas[a]

ADENOMA CELL ORIGIN	HORMONE PRODUCT	CLINICAL SYNDROME
Lactotrope	PRL	Hypogonadism, galactorrhea
Gonadotrope	FSH, LH, subunits	Silent/nonfunctioning, ovarian hyperstimulation, hypogonadism
Somatotrope	GH	Acromegaly/gigantism
Corticotrope	ACTH/none	Cushing's disease or silent
Mixed lactotrope and somatotrope	GH, PRL	Acromegaly, hypogonadism, galactorrhea
Other plurihormonal cell	Any	Mixed
Acidophil stem cell	PRL, GH	Hypogonadism, galactorrhea, acromegaly
Mammosomatotrope	PRL, GH	Hypogonadism, galactorrhea, acromegaly
Thyrotrope	TSH	Thyrotoxicosis
Null cell	None	Hypopituitarism/none
Oncocytoma	None	Hypopituitarism/none

[a]Hormone-secreting tumors are listed in decreasing order of frequency. All tumors may cause local pressure effects, including visual disturbances, cranial nerve palsy, and headache.

Note: For abbreviations, see text.

Source: Reproduced with permission from S Melmed: Pathogenesis of pituitary tumors. Nat Rev Endocrinol 7:257, 2011.

Several etiologic genetic events have been implicated in the development of pituitary tumors. The pathogenesis of sporadic forms of acromegaly has been particularly informative as a model of tumorigenesis. GHRH, after binding to its G protein–coupled somatotrope receptor, uses cyclic adenosine monophosphate (AMP) as a second messenger to stimulate GH secretion and somatotrope proliferation. A subset (~35%) of GH-secreting pituitary tumors contains sporadic mutations in $G_s\alpha$. These mutations attenuate intrinsic GTPase activity, resulting in constitutive elevation of cyclic AMP, Pit-1 induction, and activation of cyclic AMP response element binding protein (CREB), thereby promoting somatotrope cell proliferation and GH secretion.

Growth factors may also promote pituitary tumor proliferation. Basic fibroblast growth factor (bFGF) is abundant in the pituitary and stimulates pituitary cell mitogenesis, whereas epidermal growth factor receptor (EGFR) signaling induces both hormone synthesis and cell proliferation. Mutations of *USP8* may result in overexpressed EGFR in a subset of ACTH-secreting tumors. Other factors involved in initiation and promotion of pituitary tumors include loss of negative-feedback inhibition (as seen with primary hypothyroidism or hypogonadism) and estrogen-mediated or paracrine angiogenesis. Growth characteristics and neoplastic behavior also may be influenced by activated oncogenes, including *RAS* and pituitary tumor transforming gene (*PTTG*), or inactivation of growth suppressor genes, including *MEG3*. Pituitary adenomas exhibit lineage-specific features of cell-cycle disruption, including cellular senescence, with chromosomal instability and copy number alterations as well as elevated levels of CDK inhibitors. These features underlie the invariably benign nature of these adenomas.

Genetic Syndromes Associated with Pituitary Tumors Several familial syndromes are associated with pituitary tumors, and the genetic mechanisms for some of them have been unraveled **(Table 392-4)**.

Multiple endocrine neoplasia (MEN) 1 is an autosomal dominant syndrome characterized primarily by a genetic predisposition to parathyroid, pancreatic islet, and pituitary adenomas **(Chap. 400)**. MEN 1 is caused by inactivating germline mutations in *MENIN*, a constitutively expressed tumor-suppressor gene located on chromosome 11q13. Loss of heterozygosity or a somatic mutation of the remaining normal *MENIN* allele leads to tumorigenesis. About half of affected patients develop prolactinomas; acromegaly and Cushing's disease are less commonly encountered.

Carney complex is characterized by spotty skin pigmentation, myxomas, and endocrine tumors, including testicular, adrenal, and pituitary adenomas. Acromegaly occurs in ~20% of these patients. A subset of patients has mutations in the R1α regulatory subunit of protein kinase A (*PRKAR1A*).

TABLE 392-4 Familial Pituitary Tumor Syndromes (See Chap. 400)

	GENE MUTATED	CLINICAL FEATURES
Multiple endocrine neoplasia 1 (MEN 1)	*MEN1* (11q13)	Hyperparathyroidism Pancreatic neuroendocrine tumors Foregut carcinoids Adrenal adenomas Skin lesions Pituitary adenomas (40%)
Multiple endocrine neoplasia 4 (MEN 4)	*CDKNIB* (12p13)	Hyperparathyroidism Pituitary adenomas Other tumors
Carney complex	*PRKAR1A* (17q23-24)	Pituitary hyperplasia and adenomas (10%) Atrial myxomas Schwannomas Adrenal hyperplasia Lentigines
Familial pituitary adenomas	*AIP* (11q13.2)	Acromegaly/gigantism (~15% of afflicted families)

McCune-Albright syndrome consists of polyostotic fibrous dysplasia, pigmented skin patches, and a variety of endocrine disorders, including acromegaly, adrenal adenomas, and autonomous ovarian function **(Chap. 424)**. Hormonal hypersecretion results from constitutive cyclic AMP production caused by inactivation of the GTPase activity of $G_s\alpha$. The $G_s\alpha$ mutations occur postzygotically, leading to a mosaic pattern of mutant expression.

Familial acromegaly is a rare disorder in which family members may manifest either acromegaly or gigantism. A subset of families with a predisposition for familial pituitary tumors, especially acromegaly, has been found to harbor germline mutations in the *AIP* gene, which encodes the aryl hydrocarbon receptor interacting protein.

HYPERPROLACTINEMIA

Etiology Hyperprolactinemia is the most common pituitary hormone hypersecretion syndrome in both men and women. PRL-secreting pituitary adenomas (prolactinomas) are the most common cause of PRL levels >200 μg/L (see below). Less pronounced PRL elevation can also be seen with microprolactinomas but is more commonly caused by drugs, pituitary stalk compression, hypothyroidism, or renal failure **(Table 392-5)**.

Pregnancy and lactation are the important physiologic causes of hyperprolactinemia. Sleep-associated hyperprolactinemia reverts to normal within an hour of awakening. Nipple stimulation and sexual orgasm also may increase PRL. Chest wall stimulation or trauma (including chest surgery and herpes zoster) invokes the reflex suckling arc with resultant hyperprolactinemia. Chronic renal failure elevates PRL by decreasing peripheral clearance. Primary hypothyroidism is associated with mild hyperprolactinemia, probably because of compensatory TRH secretion. Mutation of the PRL receptor is a rare cause of hyperprolactinemia.

Lesions of the hypothalamic-pituitary region that disrupt hypothalamic dopamine synthesis, portal vessel delivery, or lactotrope responses are associated with hyperprolactinemia. Thus, hypothalamic tumors, cysts, infiltrative disorders, and radiation-induced damage cause elevated PRL levels, usually in the range of 30–100 μg/L. Plurihormonal adenomas (including GH and ACTH tumors) may hypersecrete PRL directly. Pituitary masses, including clinically nonfunctioning pituitary tumors, may compress the pituitary stalk to cause hyperprolactinemia.

Drug-induced inhibition or disruption of dopaminergic receptor function is a common cause of hyperprolactinemia (Table 392-5). Thus, antipsychotics and antidepressants are a relatively common cause of mild hyperprolactinemia. Most patients receiving risperidone have elevated PRL levels, sometimes exceeding 200 μg/L. Methyldopa inhibits dopamine synthesis, and verapamil blocks dopamine release, also leading to hyperprolactinemia. Hormonal agents that induce PRL include estrogens and thyrotropin-releasing hormone (TRH).

Presentation and Diagnosis Amenorrhea, galactorrhea, and infertility are the hallmarks of hyperprolactinemia in women. If hyperprolactinemia develops before menarche, primary amenorrhea results. More commonly, hyperprolactinemia develops later in life and leads to oligomenorrhea and ultimately to amenorrhea. If hyperprolactinemia is sustained, vertebral bone mineral density can be reduced compared with age-matched controls, particularly when it is associated with pronounced hypoestrogenemia. Galactorrhea is present in up to 80% of hyperprolactinemic women. Although usually bilateral and spontaneous, it may be unilateral or expressed only manually. Patients also may complain of decreased libido, weight gain, and mild hirsutism.

In men with hyperprolactinemia, diminished libido, infertility, and visual loss (from optic nerve compression) are the usual presenting symptoms. Gonadotropin suppression leads to reduced testosterone, impotence, and oligospermia. True galactorrhea is uncommon in men with hyperprolactinemia. If the disorder is long-standing, secondary effects of hypogonadism are evident, including osteopenia, reduced muscle mass, and decreased beard growth.

The diagnosis of idiopathic hyperprolactinemia is made by exclusion of known causes of hyperprolactinemia in the setting of a normal

TABLE 392-5 Etiology of Hyperprolactinemia

- **I. Physiologic hypersecretion**
  - Pregnancy
  - Lactation
  - Chest wall stimulation
  - Sleep
  - Stress
- **II. Hypothalamic-pituitary stalk damage**
  - Pituitary adenoma with stalk compression
  - Suprasellar mass
    - Craniopharyngioma
    - Meningioma
    - Dysgerminoma
    - Metastases
  - Empty sella
  - Lymphocytic hypophysitis
  - Granulomas
  - Rathke's cyst
  - Irradiation
  - Trauma
    - Pituitary stalk section
    - Suprasellar surgery
- **III. Pituitary adenoma hypersecretion**
  - Prolactinoma
  - Acromegaly
- **IV. Systemic disorders**
  - Chronic renal failure
  - Hypothyroidism
  - Cirrhosis
  - Pseudocyesis
  - Epileptic seizures
- **V. Drug-induced hypersecretion**
  - Dopamine receptor blockers
    - Atypical antipsychotics: risperidone
    - Phenothiazines: chlorpromazine, perphenazine
    - Butyrophenones: haloperidol
    - Thioxanthenes
    - Metoclopramide
  - Dopamine synthesis inhibitors
    - α-Methyldopa
  - Catecholamine depletors
    - Reserpine
  - Opiates
  - H_2 antagonists
    - Cimetidine, ranitidine
  - Imipramines
    - Amitriptyline, amoxapine
  - Serotonin reuptake inhibitors
    - Fluoxetine
  - Calcium channel blockers
    - Verapamil
    - Estrogens
    - Thyrotropin-releasing hormone

Note: Hyperprolactinemia >200 μg/L almost invariably is indicative of a prolactin-secreting pituitary adenoma. Physiologic causes, hypothyroidism, and drug-induced hyperprolactinemia should be excluded before extensive evaluation.

pituitary MRI. Some of these patients may harbor small microadenomas below visible MRI sensitivity (~2 mm).

■ GALACTORRHEA

Galactorrhea, the inappropriate discharge of milk-containing fluid from the breast, is considered abnormal if it persists longer than 6 months after childbirth or discontinuation of breast-feeding. Postpartum galactorrhea associated with amenorrhea is a self-limiting disorder usually associated with moderately elevated PRL levels. Galactorrhea may occur spontaneously, or it may be elicited by nipple pressure. In both men and women, galactorrhea may vary in color and consistency (transparent, milky, or bloody) and arise either unilaterally or bilaterally. Mammography or ultrasound is indicated for bloody discharges (particularly from a single nipple), which may be caused by breast cancer. Galactorrhea is commonly associated with hyperprolactinemia caused by any of the conditions listed in Table 392-5. Acromegaly is associated with galactorrhea in about one-third of patients. Treatment of galactorrhea usually involves managing the underlying disorder (e.g., replacing T_4 for hypothyroidism, discontinuing a medication, treating prolactinoma).

Laboratory Investigation Basal, fasting morning PRL levels (normally <20 μg/L) should be measured to assess hypersecretion. Both false-positive and false-negative results may be encountered. In patients with markedly elevated PRL levels (>1000 μg/L), reported results may be falsely lowered because of assay artifacts; sample dilution is required to measure these high values accurately. Falsely elevated values may be caused by aggregated forms of circulating PRL, which are usually biologically inactive (macroprolactinemia). Hypothyroidism should be excluded by measuring TSH and T_4 levels.

TREATMENT

Hyperprolactinemia

Treatment of hyperprolactinemia depends on the cause of elevated PRL levels. Regardless of the etiology, however, treatment should be aimed at normalizing PRL levels to alleviate suppressive effects on gonadal function, halt galactorrhea, and preserve bone mineral density. Dopamine agonists are effective for most causes of hyperprolactinemia (see the treatment section for prolactinoma, below) regardless of the underlying cause.

If the patient is taking a medication known to cause hyperprolactinemia, the drug should be withdrawn, if possible. For psychiatric patients who require neuroleptic agents, supervised dose titration or the addition of a dopamine agonist can help restore normoprolactinemia and alleviate reproductive symptoms. However, dopamine agonists may worsen the underlying psychiatric condition, especially at high doses. Hyperprolactinemia usually resolves after adequate thyroid hormone replacement in hypothyroid patients or after renal transplantation in patients undergoing dialysis. Resection of hypothalamic or sellar mass lesions can reverse hyperprolactinemia caused by stalk compression and reduced dopamine tone. Granulomatous infiltrates occasionally respond to glucocorticoid administration. In patients with irreversible hypothalamic damage, no treatment may be warranted. In up to 30% of patients with hyperprolactinemia—usually without a visible pituitary microadenoma—the condition may resolve spontaneously.

■ PROLACTINOMA

Etiology and Prevalence Tumors arising from lactotropes account for about half of all functioning pituitary tumors, with a population prevalence of ~10/100,000 in men and ~30/100,000 in women. Mixed tumors that secrete combinations of GH and PRL, ACTH and PRL, and rarely TSH and PRL are also seen. These plurihormonal tumors are usually recognized by immunohistochemistry, sometimes without apparent clinical manifestations from the production of additional hormones. Microadenomas are classified as <1 cm in diameter and usually do not invade the parasellar region. Macroadenomas are ≥1 cm in diameter and may be locally invasive and impinge on adjacent structures. The female-to-male ratio for microprolactinomas is 20:1, whereas the sex ratio is near 1:1 for macroadenomas. Tumor size generally correlates directly with PRL concentrations; values >250 μg/L usually are associated with macroadenomas. Men tend to present with larger tumors than women, possibly because the features of male hypogonadism are less readily evident. PRL levels remain stable in most patients, reflecting the slow growth of these tumors. About 5% of microadenomas progress in the long term to macroadenomas.

Presentation and Diagnosis Women usually present with amenorrhea, infertility, and galactorrhea. If the tumor extends outside the sella, visual field defects or other mass effects may be seen. Men often present with impotence, loss of libido, infertility, or signs of central nervous system (CNS) compression, including headaches and visual defects. Assuming that physiologic and medication-induced causes of hyperprolactinemia are excluded (Table 392-5), the diagnosis of prolactinoma is likely with a PRL level >200 μg/L. PRL levels <100 μg/L may be caused by microadenomas, other sellar lesions that decrease dopamine inhibition, or nonneoplastic causes of hyperprolactinemia. For this reason, an MRI should be performed in all patients with hyperprolactinemia. It is important to remember that hyperprolactinemia caused secondarily by the mass effects of nonlactotrope lesions is also corrected by treatment with dopamine agonists despite failure to shrink the underlying mass. Consequently, PRL suppression by dopamine agonists does not necessarily indicate that the underlying lesion is a prolactinoma.

TREATMENT

Prolactinoma

Because microadenomas rarely progress to become macroadenomas, no treatment may be needed if patients are asymptomatic and fertility is not desired; these patients should be monitored by regular serial PRL measurements and MRI scans. For symptomatic microadenomas, therapeutic goals include control of hyperprolactinemia, reduction of tumor size, restoration of menses and fertility, and resolution of galactorrhea. Dopamine agonist doses should be titrated to achieve maximal PRL suppression and restoration of reproductive function (Fig. 392-5). A normalized PRL level does not ensure reduced tumor size. However, tumor shrinkage usually is not seen in those who do not respond with lowered PRL levels. For macroadenomas, formal visual field testing should be performed before initiating dopamine agonists. MRI and visual fields should be assessed at 6- to 12-month intervals until the mass shrinks and annually thereafter until maximum size reduction has occurred.

Oral dopamine agonists (cabergoline and bromocriptine) are the mainstay of therapy for patients with micro- or macroprolactinomas. Dopamine agonists suppress PRL secretion and synthesis as well as lactotrope proliferation. In patients with microadenomas who have achieved normoprolactinemia and significant reduction of tumor mass, the dopamine agonist may be withdrawn after 2 years. These patients should be monitored carefully for evidence of prolactinoma recurrence. About 20% of patients (especially males) are resistant to dopaminergic treatment; these adenomas may exhibit decreased D_2 dopamine receptor numbers or a postreceptor defect. D_2 receptor gene mutations in the pituitary have not been reported.

Cabergoline An ergoline derivative, cabergoline is a long-acting dopamine agonist with high D_2 receptor affinity. The drug effectively suppresses PRL for >14 days after a single oral dose and induces prolactinoma shrinkage in most patients. Cabergoline (0.5–1.0 mg twice weekly) achieves normoprolactinemia and resumption of normal gonadal function in ~80% of patients with microadenomas; galactorrhea improves or resolves in 90% of patients. Cabergoline normalizes PRL and shrinks ~70% of macroprolactinomas. Mass effect symptoms, including headaches and visual disorders, usually improve dramatically within days after cabergoline initiation; improvement of sexual function requires several weeks of treatment but may occur before complete normalization of PRL levels. MRI should be repeated within 16 weeks after initial therapy of macroadenomas as shrinkage of invasive adenomas may be striking (Fig. 392-6). After initial control of PRL levels has been achieved, cabergoline should be reduced to the lowest effective maintenance dose. In ~5% of treated patients harboring a microadenoma, hyperprolactinemia may resolve and not recur when dopamine agonists are discontinued after long-term treatment. Cabergoline also may be effective in patients resistant to bromocriptine. Adverse effects and drug intolerance are encountered less commonly than with bromocriptine.

Bromocriptine The ergot alkaloid bromocriptine mesylate is a dopamine receptor agonist that suppresses PRL secretion. Because it is short-acting, the drug is preferred when pregnancy is desired. Therapy is initiated by administering a low bromocriptine dose (0.625–1.25 mg) at bedtime with a snack, followed by gradually

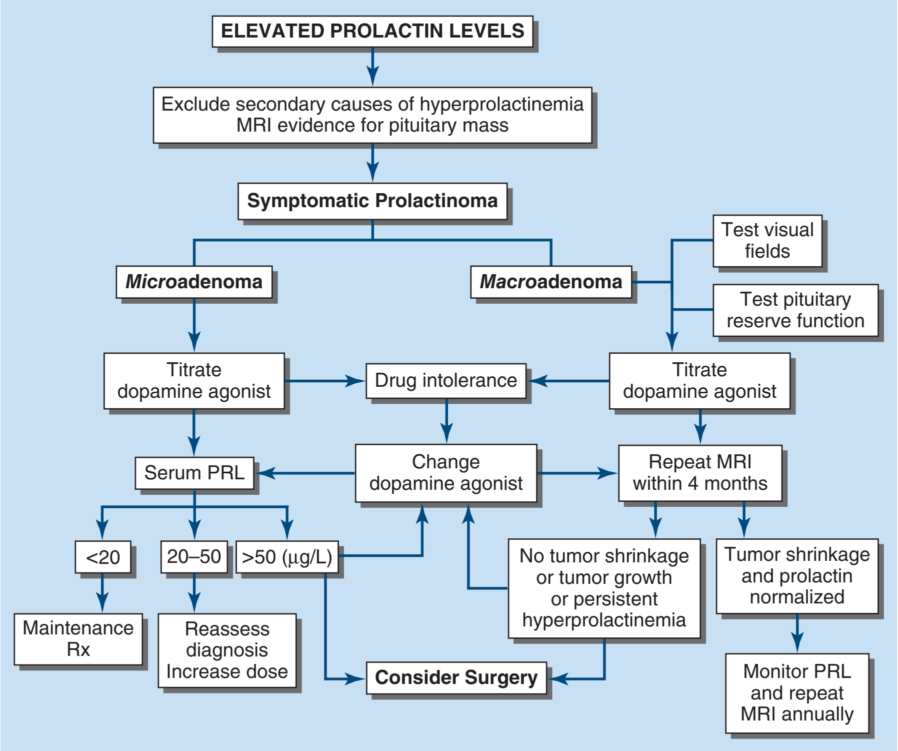
ELEVATED PROLACTIN LEVELS
Exclude secondary causes of hyperprolactinemia
MRI evidence for pituitary mass
Symptomatic Prolactinoma
Microadenoma
Macroadenoma
Test visual fields
Test pituitary reserve function
Titrate dopamine agonist
Drug intolerance
Titrate dopamine agonist
Serum PRL
Change dopamine agonist
Repeat MRI within 4 months
<20
20–50
>50 (μg/L)
No tumor shrinkage or tumor growth or persistent hyperprolactinemia
Tumor shrinkage and prolactin normalized
Maintenance Rx
Reassess diagnosis Increase dose
Consider Surgery
Monitor PRL and repeat MRI annually

FIGURE 392-5 Management of prolactinoma. MRI, magnetic resonance imaging; PRL, prolactin.

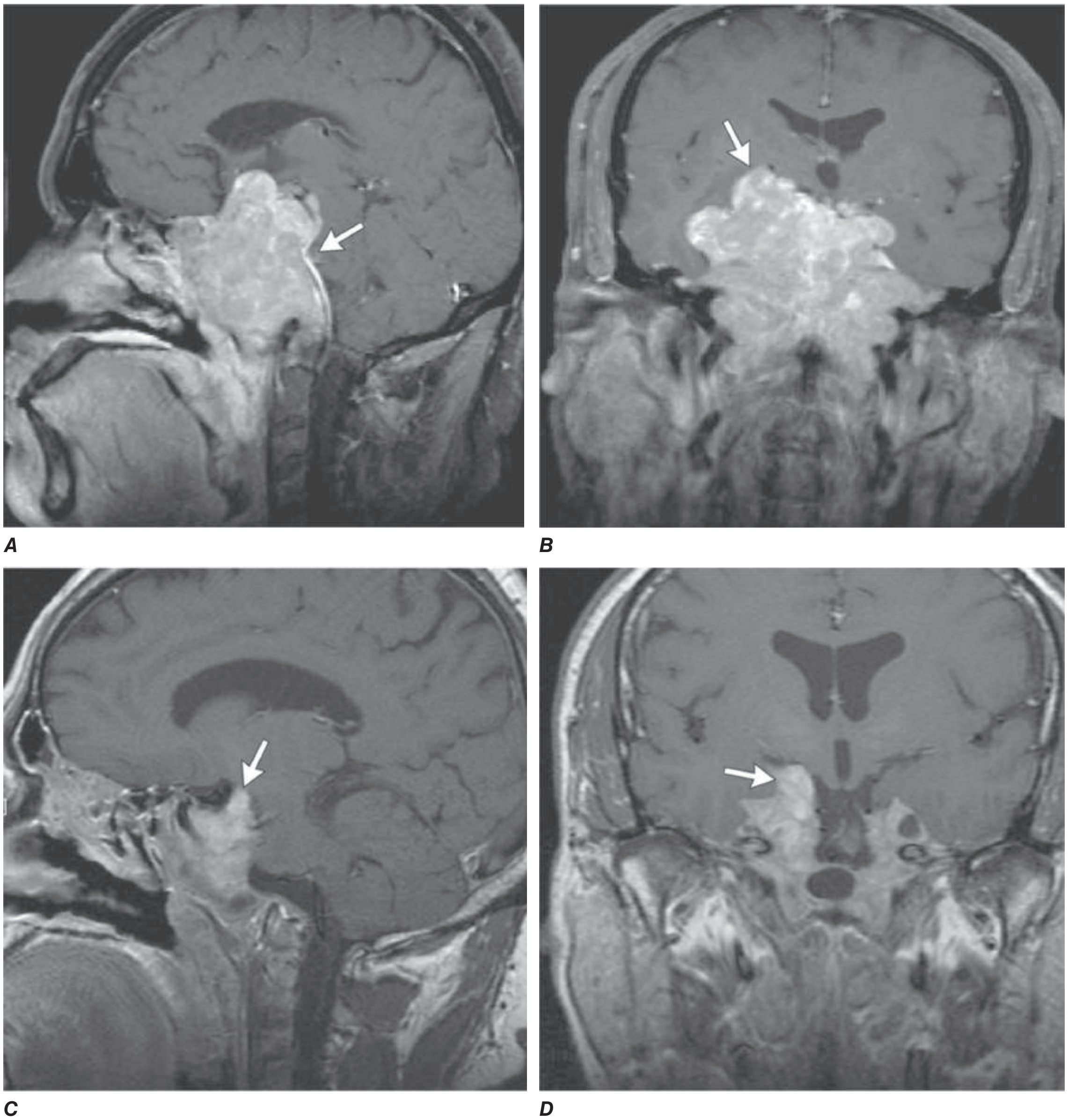

FIGURE 392-6 Large invasive prolactinoma successfully treated with cabergoline. ***A–B.*** Prolactin-secreting macroadenoma in a 32-year-old male measuring 5.6 × 6.9 cm invading the skull base. PRL level was 122,260 μg/L. Four days after cabergoline was started, PRL was 10,823 μg/L and dropped to 772 μg/L after 3 weeks. ***C–D.*** Substantial tumor regression after 40 months of treatment, with PRL levels stable at 25 μg/L. *(Reproduced with permission from M Ahmed, O Al-Nozha: Images in clinical medicine. Large prolactinoma. N Engl J Med 363:177, 2010.)*

increasing the dose. Most patients are controlled with a daily dose of <7.5 mg (2.5 mg tid).

SIDE EFFECTS

Side effects of dopamine agonists include constipation, nasal stuffiness, dry mouth, nightmares, insomnia, and vertigo; decreasing the dose usually alleviates these problems. Nausea, vomiting, and postural hypotension with faintness may occur in ~25% of patients after the initial dose. These symptoms may persist in some patients. In general, fewer side effects are reported with cabergoline. For the ~15% of patients who are intolerant of oral bromocriptine, cabergoline may be better tolerated. Intravaginal administration of bromocriptine is often efficacious in patients with intractable gastrointestinal side effects. Auditory hallucinations, delusions, mood swings, and impulse control disorders have been reported in up to 5% of patients and may be due to the dopamine agonist properties or to the lysergic acid derivative of the compounds. Rare reports of leukopenia, thrombocytopenia, pleural fibrosis, cardiac arrhythmias, and hepatitis have been described. Patients with Parkinson disease who receive at least 3 mg of cabergoline daily have been reported to be at risk for development of cardiac valve regurgitation. Studies analyzing >500 prolactinoma patients receiving recommended doses of cabergoline (up to 2 mg weekly) have shown no evidence for an increased incidence of valvular disorders. Nevertheless, because no controlled prospective studies in pituitary tumor patients are available, it is prudent to perform echocardiograms before initiating standard-dose cabergoline therapy.

Surgery Surgical adenoma debulking may be indicated for dopamine resistance or intolerance as well as the presence of an invasive macroadenoma with compromised vision that fails to improve after drug treatment. Initial PRL normalization is achieved in ~70% of microprolactinomas after surgical resection, but only 40% of macroadenomas can be resected successfully. Follow-up studies have shown that hyperprolactinemia recurs in up to 20% of patients within the first year after surgery; long-term recurrence rates may exceed 50% for macroadenomas. Radiotherapy for prolactinomas is reserved for patients with aggressive tumors that do not respond to maximally tolerated dopamine agonists and/or surgery.

PREGNANCY

The pituitary increases in size during pregnancy, reflecting the stimulatory effects of estrogen and perhaps other growth factors on pituitary vascularity and lactotrope hyperplasia. About 5% of microadenomas significantly increase in size, but 15–30% of macroadenomas grow during pregnancy. Bromocriptine has been used for >30 years to restore fertility in women with hyperprolactinemia, without evidence of teratogenic effects. Nonetheless, most authorities recommend strategies to minimize fetal exposure to the drug. For women taking bromocriptine who desire pregnancy, mechanical contraception should be used through three regular menstrual cycles to allow for conception timing. When pregnancy is confirmed, bromocriptine should be discontinued and PRL levels followed serially, especially if headaches or visual symptoms occur. For women harboring macroadenomas, regular visual field testing is recommended, and the drug should be reinstituted if tumor growth is apparent. Although pituitary MRI may be safe during pregnancy, this procedure should be reserved for symptomatic patients with severe headache and/or visual field defects. Surgical decompression may be indicated if vision is threatened. Although comprehensive data support the efficacy and relative safety of bromocriptine-facilitated fertility, patients should be advised of potential unknown deleterious effects and the risk of tumor growth during pregnancy. Because cabergoline is long-acting with a high D_2-receptor affinity, it is not recommended for use in women when fertility is desired.

ACROMEGALY

Etiology GH hypersecretion is usually the result of a somatotrope adenoma but may rarely be caused by extrapituitary lesions (Table 392-6). In addition to the more common GH-secreting somatotrope adenomas, mixed mammosomatotrope tumors and acidophilic stem cell adenomas secrete both GH and PRL. In patients with acidophilic stem cell adenomas, features of hyperprolactinemia (hypogonadism and galactorrhea) predominate over the less clinically evident signs of acromegaly. Occasionally, mixed plurihormonal tumors are encountered that also secrete ACTH, the glycoprotein hormone α subunit, or TSH in addition to GH. Patients with partially empty sellae may present with GH hypersecretion due to a small GH-secreting adenoma within the compressed rim of pituitary tissue; some of these may reflect the spontaneous necrosis of tumors that were previously larger. GH-secreting tumors rarely arise from ectopic pituitary tissue remnants in the nasopharynx or midline sinuses.

TABLE 392-6 Causes of Acromegaly

	PREVALENCE, %
Excess Growth Hormone Secretion	
Pituitary	98
Densely or sparsely granulated GH cell adenoma	60
Mixed GH cell and PRL cell adenoma	25
Mammosomatotrope cell adenoma	10
Plurihormonal adenoma	
GH cell carcinoma or metastases	
Multiple endocrine neoplasia 1 (GH cell adenoma)	
McCune-Albright syndrome	
Ectopic sphenoid or parapharyngeal sinus pituitary adenoma	
Extrapituitary tumor	<1
Pancreatic islet cell tumor	
Lymphoma	
Excess Growth Hormone–Releasing Hormone Secretion	
Central	<1
Hypothalamic hamartoma, choristoma, ganglioneuroma	
Peripheral	<1
Bronchial carcinoid, pancreatic islet cell tumor, small-cell lung cancer, adrenal adenoma, medullary thyroid carcinoma, pheochromocytoma	

Abbreviations: GH, growth hormone; PRL, prolactin.

Source: Data from S Melmed: Medical progress: Acromegaly. N Engl J Med 355:2558, 2006.

There are case reports of ectopic GH secretion by tumors of pancreatic, ovarian, lung, or hematopoietic origin. Rarely, excess GHRH production may cause acromegaly because of chronic stimulation of somatotropes. These patients present with classic features of acromegaly, elevated GH levels, pituitary enlargement on MRI, and pathologic characteristics of pituitary hyperplasia. The most common cause of GHRH-mediated acromegaly is a chest or abdominal carcinoid tumor. Although these tumors usually express positive GHRH immunoreactivity, clinical features of acromegaly are evident in only a minority of patients with carcinoid disease. Excessive GHRH also may be elaborated by hypothalamic tumors, usually choristomas or neuromas.

Presentation and Diagnosis Protean manifestations of GH and IGF-1 hypersecretion are indolent and often are not clinically diagnosed for 10 years or more. Acral bony overgrowth results in frontal bossing, increased hand and foot size, mandibular enlargement with prognathism, and widened space between the lower incisor teeth. In children and adolescents, initiation of GH hypersecretion before epiphyseal long bone closure is associated with development of pituitary gigantism (Fig. 392-7). Soft tissue swelling results in increased heel pad thickness, increased shoe or glove size, ring tightening, characteristic coarse facial features, and a large fleshy nose. Other commonly encountered clinical features include hyperhidrosis, a deep and hollow-sounding voice, oily skin, arthropathy, kyphosis, carpal tunnel syndrome, proximal muscle weakness and fatigue, acanthosis nigricans, and skin tags. Generalized visceromegaly occurs, including cardiomegaly, macroglossia, and thyroid gland enlargement.

The most significant clinical impact of GH excess occurs with respect to the cardiovascular system. Cardiomyopathy with arrhythmias, left ventricular hypertrophy, decreased diastolic function, and hypertension ultimately occur in most patients if untreated. Upper airway obstruction with sleep apnea occurs in >60% of patients and is associated with both soft tissue laryngeal airway obstruction and central sleep dysfunction. Diabetes mellitus develops in 25% of patients with acromegaly, and most patients are intolerant of a glucose load (as GH counteracts the action of insulin). Acromegaly is associated with an increased risk of colon polyps and mortality from colonic malignancy; polyps are diagnosed in up to one-third of patients. Overall mortality is increased about threefold and is due primarily to cardiovascular and cerebrovascular disorders and respiratory disease. Unless GH levels are controlled, survival is reduced by an average of 10 years compared with an age-matched control population.

Laboratory Investigation Age-matched serum IGF-1 levels are elevated in acromegaly. Consequently, an IGF-1 level provides a useful laboratory screening measure when clinical features raise the possibility of acromegaly. Owing to the pulsatility of GH secretion, measurement of a single random GH level is not useful for the diagnosis or exclusion of acromegaly and does not correlate with disease severity. The diagnosis of acromegaly is confirmed by demonstrating the failure of GH suppression to <0.4 μg/L within 1–2 h of an oral glucose load (75 g). When ultrasensitive GH assays are used, normal nadir GH levels are even lower (<0.05 μg/L). About 20% of patients exhibit a paradoxical GH rise after glucose. PRL should be measured, as it is elevated in ~25% of patients with acromegaly. Thyroid function, gonadotropins, and sex steroids may be attenuated because of tumor mass effects. Because most patients will undergo surgery with glucocorticoid coverage, tests of ACTH reserve in asymptomatic patients are more efficiently deferred until after surgery.

TREATMENT

Acromegaly

The goal of treatment is to control GH and IGF-1 hypersecretion, ablate or arrest tumor growth, ameliorate comorbidities, restore mortality rates to normal, and preserve pituitary function.

Surgical resection of GH-secreting adenomas is the initial treatment for most patients (Fig. 392-8). SRLs are used as adjuvant

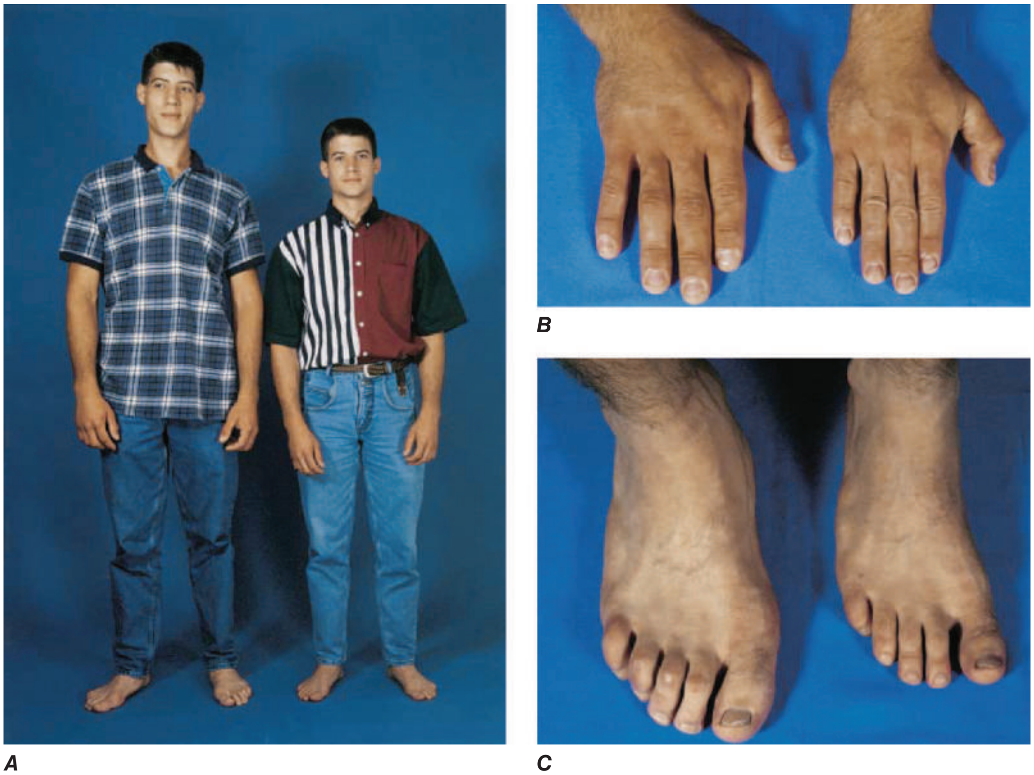

FIGURE 392-7 Features of acromegaly/gigantism. A 22-year-old man with gigantism due to excess growth hormone is shown to the left of his identical twin. The increased height and prognathism (***A***) and enlarged hand (***B***) and foot (***C***) of the affected twin are apparent. Their clinical features began to diverge at the age of ~13 years. *(Reproduced with permission from RF Gagel, IE McCutcheon. Images in clinical medicine. Pituitary gigantism. N Engl J Med 340:524, 1999.)*

treatment for preoperative shrinkage of large invasive macroadenomas, immediate relief of debilitating symptoms, and reduction of GH hypersecretion; in frail patients experiencing morbidity; and in patients who decline surgery or when surgery fails to achieve biochemical control. Irradiation or repeat surgery may be required for patients who cannot tolerate or do not respond to adjunctive medical therapy. The high rate of late hypopituitarism and the slow rate (5–15 years) of biochemical response are the main disadvantages of radiotherapy. Irradiation is also relatively ineffective in normalizing IGF-1 levels. Stereotactic ablation of GH-secreting adenomas by Gamma Knife radiotherapy is promising, but long-term results and side effects appear similar to those observed with conventional radiation. SRLs may be required while awaiting the full benefits of radiotherapy. Systemic comorbid sequelae of acromegaly, including cardiovascular disease, diabetes, and arthritis, should be managed aggressively. Mandibular surgical repair may be indicated.

SURGERY

Transsphenoidal surgical resection by an experienced surgeon is the preferred primary treatment for both microadenomas (remission rate ~70%) and macroadenomas (<50% in remission). Soft tissue swelling improves immediately after tumor resection. GH levels return to normal within an hour, and IGF-1 levels are normalized within 3–4 days. In ~10% of patients, acromegaly may recur several years after apparently successful surgery; hypopituitarism develops in up to 15% of patients after surgery.

SOMATOSTATIN RECEPTOR LIGANDS

SRLs exert their therapeutic effects through SST2 and SST5 receptor subtypes, both expressed by GH-secreting tumors.

The preferred medical treatments for patients with acromegaly include long-acting injectable SRL depot formulations of octreotide and lanreotide as well as oral octreotide capsules. Although responses vary widely in individual patients, meta-analyses indicate that GH and IGF-1 levels are normalized in ~50% of patients. Octreotide acetate is an eight-amino-acid synthetic somatostatin analogue. In contrast to native somatostatin, the analogue is relatively resistant to plasma degradation. It has a 2-h serum half-life and possesses 40-fold greater potency than native somatostatin to suppress GH. *Octreotide LAR* is a sustained-release, long-acting formulation of octreotide incorporated into microspheres that sustain drug levels for several weeks after intramuscular injection. GH suppression occurs for as long as 6 weeks after a 30-mg intramuscular injection; long-term monthly treatment sustains GH and IGF-1 suppression and also reduces pituitary tumor size in ~50% of patients. *Lanreotide*, in a slow-release depot SRL preparation, is a cyclic somatostatin octapeptide analogue that suppresses GH and IGF-1 hypersecretion after a 60-mg subcutaneous injection. Long-term (every 4–6 weeks) administration controls GH hypersecretion in about two-thirds of treated patients and improves patient compliance because of the long interval required between drug injections. *Oral octreotide capsules* (40–80 mg daily) maintain biochemical control in patients previously maintained on injectable formulations. Rapid relief of headache and soft tissue swelling occurs in ~75% of patients within days to weeks of SRL initiation. Most patients report symptomatic improvement, including amelioration of headache, perspiration, obstructive apnea, and cardiac failure. *Pasireotide LAR*, a multireceptor ligand with preferential SST5 binding, has been shown to exhibit efficacy in achieving biochemical control in patients resistant to octreotide or lanreotide preparations.

Side Effects

SRLs are well tolerated in most patients. Adverse effects are similar for injectable octreotide and lanreotide as well as for oral octreotide formulation. They are short-lived and mostly relate to drug-induced suppression of gastrointestinal motility and secretion. Transient nausea, abdominal discomfort, fat malabsorption, diarrhea, and flatulence occur in one-third of patients, and these symptoms usually remit within 2 weeks. Gallbladder contractility and emptying are attenuated; up to 30% of patients develop long-term echogenic sludge or asymptomatic cholesterol gallstones. Other side effects include mild glucose intolerance due to transient insulin suppression, asymptomatic bradycardia, hypothyroxinemia, and local injection site discomfort. Pasireotide is associated with similar gastrointestinal side effects but with a higher prevalence of glucose intolerance and new-onset diabetes mellitus.

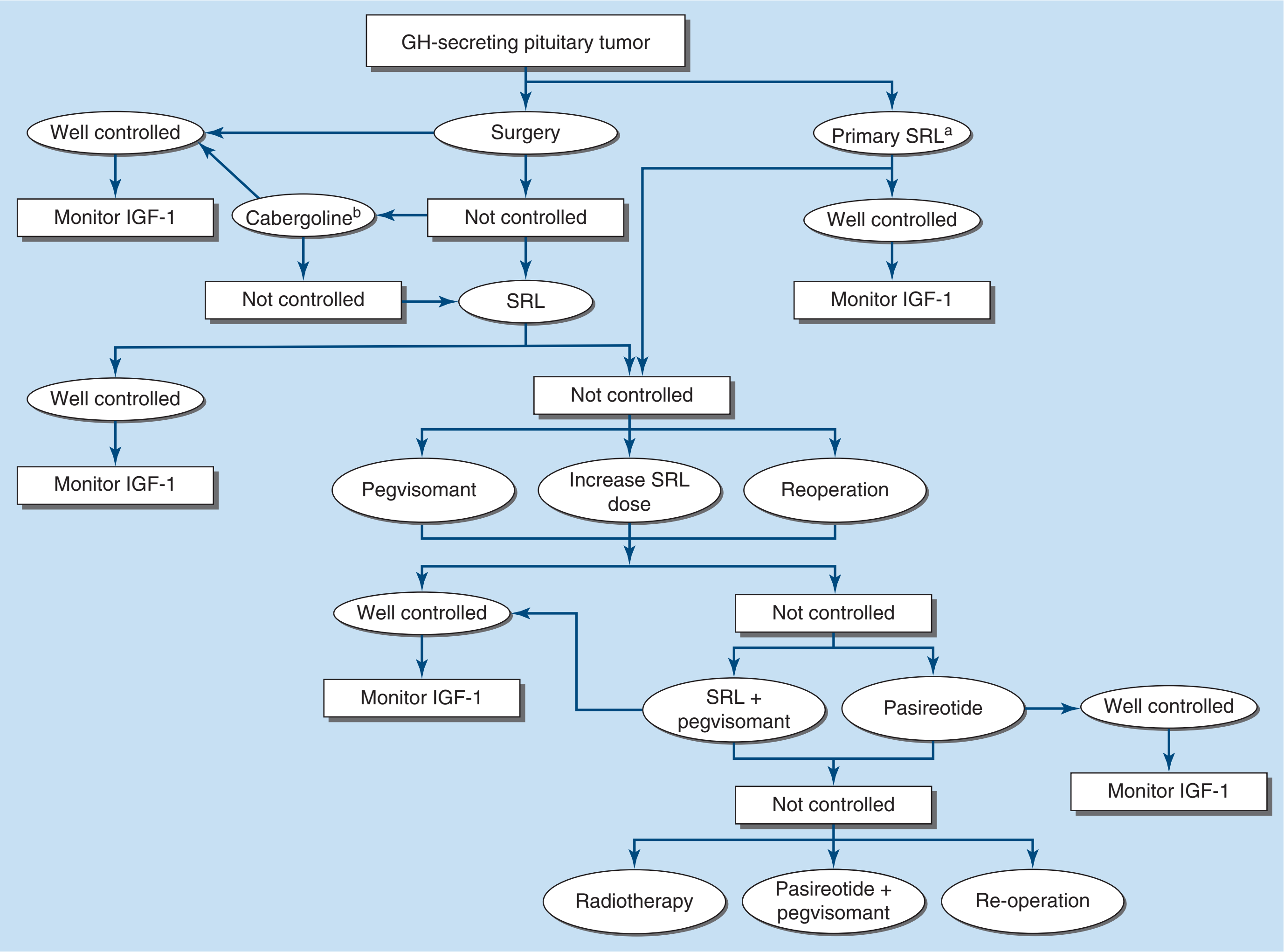
GH-secreting pituitary tumor
Surgery
Primary SRL^a
Well controlled
Monitor IGF-1
Cabergoline^b
Not controlled
Well controlled
Monitor IGF-1
Not controlled
SRL
Well controlled
Monitor IGF-1
Not controlled
Pegvisomant
Increase SRL dose
Reoperation
Well controlled
Not controlled
Monitor IGF-1
SRL + pegvisomant
Pasireotide
Well controlled
Monitor IGF-1
Not controlled
Radiotherapy
Pasireotide + pegvisomant
Re-operation

FIGURE 392-8 Management of acromegaly. ^{a}If curative surgery is not feasible. ^{b}Consider in cases of mild postoperative GH/IGF-1 elevations. GH, growth hormone; IGF, insulin-like growth factor; SRL, somatostatin receptor ligand (injectable or oral octreotide, or lanreotide).

GH RECEPTOR ANTAGONIST

Pegvisomant antagonizes endogenous GH action by blocking peripheral GH binding to its receptor. Consequently, serum IGF-1 levels are suppressed, reducing the deleterious effects of excess endogenous GH. Pegvisomant is administered by daily subcutaneous injection (10–30 mg) and normalizes IGF-1 in ~70% of patients. GH levels, however, remain elevated as the drug does not target the pituitary adenoma. Side effects include reversible liver enzyme elevation, lipodystrophy, and injection site pain. Tumor size should be monitored by MRI.

Combined treatment with monthly SRLs and weekly or biweekly pegvisomant injections has been used effectively in treatment-resistant patients.

DOPAMINE AGONISTS

Very high doses of cabergoline (0.5 mg/d) may achieve short-lived and modest GH therapeutic efficacy. Combined treatment with octreotide and cabergoline may induce additive biochemical control compared with either drug alone.

RADIATION THERAPY

External radiation therapy or high-energy stereotactic techniques are used as adjuvant therapy for acromegaly. An advantage of radiation is that patient compliance with long-term treatment is not required. Tumor mass is reduced, and GH levels are attenuated over time. However, 50% of patients require at least 8 years for GH levels to be suppressed to <5 μg/L; this level of GH reduction is achieved in ~90% of patients after 18 years but represents suboptimal GH suppression. Patients may require interim medical therapy for several years before attaining maximal radiation benefits. Most patients also experience hypothalamic-pituitary damage, leading to gonadotropin, ACTH, and/or TSH deficiency within 10 years of therapy.

SUMMARY

Surgery is the preferred primary treatment for GH-secreting microadenomas (Fig. 392-8). The high frequency of residual GH hypersecretion after macroadenoma resection usually necessitates adjuvant or primary medical therapy for these larger tumors. Patients unable to receive or respond to unimodal medical treatment may benefit from combined treatments, or they can be offered radiation. Very rarely, repeat surgery may be required.

■ CUSHING'S DISEASE (ACTH-PRODUCING ADENOMA)

(See also Chap. 398)

Etiology and Prevalence Pituitary corticotrope adenomas (Cushing's disease) account for 70% of patients with endogenous causes of Cushing's syndrome. However, it should be emphasized that iatrogenic hypercortisolism is the most common cause of cushingoid features. Ectopic tumor ACTH production, cortisol-producing adrenal adenomas, adrenal carcinoma, and adrenal hyperplasia account for the other causes; rarely, ectopic tumor CRH production is encountered.

ACTH-producing adenomas account for ~10–15% of all pituitary tumors. Because the clinical features of Cushing's syndrome often lead to early diagnosis, most ACTH-producing pituitary tumors are

TABLE 392-7 Clinical Features of Cushing's Syndrome (All Ages)

SYMPTOMS/SIGNS	FREQUENCY, %
Obesity or weight gain (>115% ideal body weight)	80
Thin skin	80
Moon facies	75
Hypertension	75
Purple skin striae	65
Hirsutism	65
Menstrual disorders (usually amenorrhea)	60
Plethora	60
Abnormal glucose tolerance	55
Impotence	55
Proximal muscle weakness	50
Truncal obesity	50
Acne	45
Bruising	45
Mental changes	45
Osteoporosis	40
Edema of lower extremities	30
Hyperpigmentation	20
Hypokalemic alkalosis	15
Diabetes mellitus	15

Source: Adapted with permission from MA Magiokou et al, in Wierman ME: *Diseases of the Pituitary.* Totowa, NJ: Humana; 1997.

relatively small microadenomas. However, macroadenomas also are seen and some ACTH-expressing adenomas are clinically silent. Cushing's disease is 5–10 times more common in women than in men. These pituitary adenomas exhibit unrestrained ACTH secretion, with resultant hypercortisolemia. However, they retain partial suppressibility in the presence of high doses of administered glucocorticoids, providing the basis for dynamic testing to distinguish pituitary from nonpituitary causes of Cushing's syndrome.

Presentation and Diagnosis The diagnosis of Cushing's syndrome presents two great challenges: (1) to distinguish patients with pathologic cortisol excess from those with physiologic or other disturbances of cortisol production and (2) to determine the etiology of pathologic cortisol excess.

Typical features of chronic cortisol excess include thin skin, central obesity, hypertension, plethoric moon facies, purple striae and easy bruisability, glucose intolerance or diabetes mellitus, gonadal dysfunction, osteoporosis, proximal muscle weakness, signs of hyperandrogenism (acne, hirsutism), and psychological disturbances (depression, mania, and psychoses) **(Table 392-7)**. Hematopoietic features of hypercortisolism include leukocytosis, lymphopenia, and eosinopenia. Immune suppression includes delayed hypersensitivity and infection propensity. These protean yet commonly encountered manifestations of hypercortisolism make it challenging to decide which patients mandate formal laboratory evaluation. Certain features make pathologic causes of hypercortisolism more likely; they include characteristic central redistribution of fat, thin skin with striae and bruising, and proximal muscle weakness. In children and young females, early osteoporosis may be particularly prominent. The primary cause of death is cardiovascular disease, but life-threatening infections and risk of suicide are also increased.

Rapid development of features of hypercortisolism associated with skin hyperpigmentation and severe myopathy suggests an ectopic tumor source of ACTH. Hypertension, hypokalemic alkalosis, glucose intolerance, and edema are also more pronounced in these patients. Serum potassium levels <3.3 mmol/L are evident in ~70% of patients with ectopic ACTH secretion but are seen in <10% of patients with pituitary-dependent Cushing's syndrome.

Laboratory Investigation The diagnosis of Cushing's disease is based on laboratory documentation of endogenous hypercortisolism. Measurement of 24-h UFC is a precise and cost-effective screening test. Alternatively, the failure to suppress plasma cortisol after an overnight 1-mg dexamethasone suppression test can be used to identify patients with hypercortisolism. As nadir levels of cortisol occur at night, elevated midnight serum or salivary samples of cortisol are suggestive of Cushing's disease. Basal plasma ACTH levels often distinguish patients with ACTH-independent (adrenal or exogenous glucocorticoid) from those with ACTH-dependent (pituitary, ectopic ACTH) Cushing's syndrome. Mean basal ACTH levels are about eightfold higher in patients with ectopic ACTH secretion than in those with pituitary ACTH-secreting adenomas. However, extensive overlap of ACTH levels in these two disorders precludes using ACTH measurements to make the distinction. Preferably, dynamic testing based on differential sensitivity to glucocorticoid feedback or ACTH stimulation in response to CRH or cortisol reduction is used to distinguish ectopic from pituitary sources of excess ACTH **(Table 392-8)**. Very rarely, circulating CRH levels are elevated, reflecting ectopic tumor-derived secretion of CRH and often ACTH. **For further discussion of dynamic testing for Cushing's syndrome, see Chap. 398.**

Most ACTH-secreting pituitary tumors are <5 mm in diameter, and about half are undetectable by sensitive MRI. The high prevalence of incidental pituitary microadenomas diminishes the ability to distinguish ACTH-secreting pituitary tumors accurately from nonsecreting incidentalomas.

Inferior Petrosal Venous Sampling Because pituitary MRI with gadolinium enhancement is insufficiently sensitive to detect small (<2 mm) pituitary ACTH-secreting adenomas, bilateral inferior petrosal sinus ACTH sampling before and after CRH administration may be required to distinguish these lesions from ectopic ACTH-secreting tumors that may have similar clinical and biochemical characteristics. Simultaneous assessment of ACTH in each inferior petrosal vein and

TABLE 392-8 Differential Diagnosis of ACTH-Dependent Cushing's Syndrome[a]

	ACTH-SECRETING PITUITARY TUMOR	ECTOPIC ACTH SECRETION
Etiology	Pituitary corticotrope adenoma Plurihormonal adenoma	Bronchial, abdominal carcinoid Small-cell lung cancer Thymoma, other sources
Sex	F > M	M > F
Clinical features	Slow onset	Rapid onset Pigmentation Severe myopathy
Serum potassium <3.3 μg/L	<10%	75%
24-h UFC	High	High
Basal ACTH level	Inappropriately high	Very high
Dexamethasone suppression		
1 mg overnight		
Low-dose (0.5 mg q6h)	Cortisol >5 μg/dL	Cortisol >5 μg/dL
High-dose (2 mg q6h)	Cortisol <5 μg/dL	Cortisol >5 μg/dL
UFC >80% suppressed	Microadenomas: 90% Macroadenomas: 50%	10%
Inferior petrosal sinus sampling		
Basal		
central: peripheral	>2	<2
CRH-induced		
central: peripheral	>3	<3

[a] ACTH-independent causes of Cushing's syndrome are diagnosed by suppressed ACTH levels and an adrenal mass in the setting of hypercortisolism. Iatrogenic Cushing's syndrome is excluded by history.

Abbreviations: ACTH, adrenocorticotropic hormone; CRH, corticotropin-releasing hormone; F, female; M, male; UFC, urinary free cortisol.

in the peripheral circulation provides a strategy for confirming and localizing pituitary ACTH production. Sampling is performed at baseline and 2, 5, and 10 min after intravenous CRH (1 μg/kg) injection. An increased central:peripheral ACTH ratio (>2) before and a peak central:peripheral ACTH ratio >3 after CRH injection confirm the presence of a pituitary ACTH-secreting adenoma. The sensitivity of this test is >95%, with very rare false-positive results. False-negative results may be encountered in patients with aberrant venous drainage. Petrosal sinus catheterizations are technically difficult, and ~0.05% of patients develop neurovascular complications. The procedure should not be performed in patients with hypertension, in patients with known cerebrovascular disease, or in the presence of a well-visualized pituitary adenoma on MRI.

TREATMENT

Cushing's Disease

Selective transsphenoidal resection is the treatment of choice for Cushing's disease **(Fig. 392-9)**. The remission rate for this procedure is ~80% for microadenomas but <50% for macroadenomas. However, surgery is rarely successful when the adenoma is not visible on MRI. After successful tumor resection, most patients experience a postoperative period of symptomatic ACTH deficiency that may last up to 12 months. This usually requires low-dose cortisol replacement, as patients experience steroid withdrawal symptoms and have a suppressed hypothalamic-pituitary-adrenal axis. Biochemical recurrence occurs in ~5% of patients in whom surgery was initially successful. As persistent hypercortisolemia may cause blood clotting defects, prophylactic postoperative thromboembolic management has been advocated for vulnerable patients.

When initial surgery is unsuccessful, repeat surgery is sometimes indicated, particularly when a pituitary source for ACTH is well documented. In older patients, in whom issues of growth and fertility are less important, hemi- or total hypophysectomy may be necessary if a discrete pituitary adenoma is not recognized. Pituitary irradiation may be used after unsuccessful surgery, but it cures only ~15% of patients. Because the effects of radiation are slow and only partially effective in adults, adrenal-targeted steroidogenic inhibitors are used in combination with pituitary irradiation to block adrenal responses to persistently high ACTH levels.

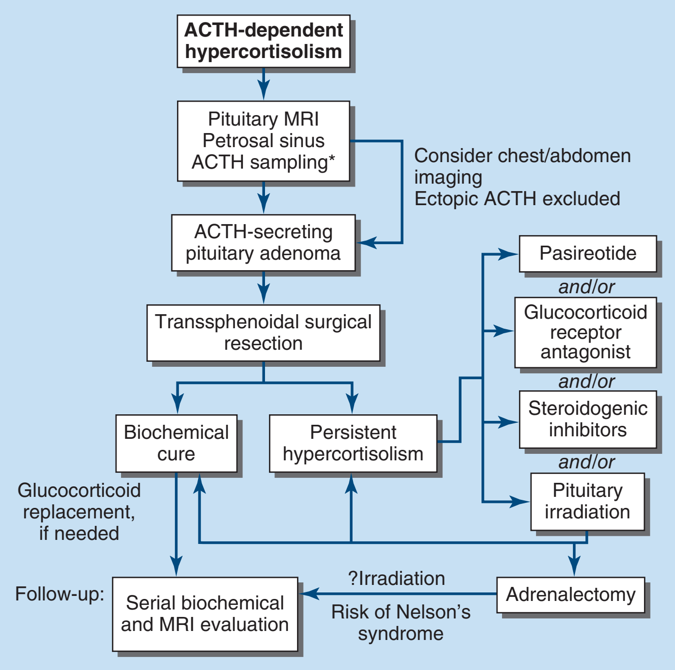
ACTH-dependent hypercortisolism
Pituitary MRI
Petrosal sinus
ACTH sampling*
Consider chest/abdomen imaging
Ectopic ACTH excluded
ACTH-secreting pituitary adenoma
Transsphenoidal surgical resection
Pasireotide
and/or
Glucocorticoid receptor antagonist
and/or
Steroidogenic inhibitors
and/or
Pituitary irradiation
Biochemical cure
Persistent hypercortisolism
Glucocorticoid replacement, if needed
Follow-up:
Serial biochemical and MRI evaluation
?Irradiation
Risk of Nelson's syndrome
Adrenalectomy

FIGURE 392-9 Management of Cushing's disease. ACTH, adrenocorticotropin hormone; MRI, magnetic resonance imaging; *, Not usually required.

Pasireotide LAR 10–40 mg intramuscularly, an SRL with high affinity for SST5 > SST2 receptor subtypes, may control hypercortisolemia in a subset of patients with ACTH-secreting pituitary tumors when surgery is not an option or has not been successful. The drug lowers plasma ACTH levels and normalizes 24-h UFC levels in ~20% of patients, and up to 40% of patients may experience pituitary tumor shrinkage. Side effects are similar to those encountered for other SRLs and include transient abdominal discomfort, diarrhea, nausea, and gallstones (20% of patients). Notably, hyperglycemia and new-onset diabetes develop in up to 70% of patients, likely due to suppressed pancreatic secretion of insulin and incretins. Because patients with hypercortisolism are insulin-resistant, hyperglycemia should be rigorously managed. The drug requires consistent long-term administration.

Osilodrostat (2 mg twice daily titrated up to 30 mg twice daily), an oral 11β-hydroxylase inhibitor that blocks adrenal gland cortisol biosynthesis, normalized 24-h UFC in 86% of patients. Mild, mostly transient gastrointestinal symptoms are common. Patients should be closely monitored for development of hypocortisolism and adrenal insufficiency. Elevated adrenal hormone precursors may lead to hypokalemia and hypertension. QTc prolongation and possibly increased tumor volume are also reported.

Ketoconazole, an imidazole derivative antimycotic agent, inhibits several P450 enzymes and effectively lowers cortisol in most patients with Cushing's disease when administered twice daily (600–1200 mg/d). Elevated hepatic transaminases, gynecomastia, impotence, gastrointestinal upset, and edema are common side effects.

Levoketoconazole, a 2S,4R enantiomer of ketoconazole, is administered at the same dose/schedule as ketoconazole and has a similar side effect profile.

Mifepristone (300–1200 mg/d), a glucocorticoid receptor antagonist, blocks peripheral cortisol action and is approved to treat hyperglycemia in Cushing's disease. Because the drug does not target the pituitary tumor, both ACTH and cortisol levels remain elevated, thus obviating a reliable circulating biomarker. Side effects are largely due to general antagonism of other steroid hormones and include hypokalemia, endometrial hyperplasia, hypoadrenalism, and hypertension.

Metyrapone (2–4 g/d) inhibits 11β-hydroxylase activity and normalizes plasma cortisol in up to 75% of patients. Side effects include nausea and vomiting, rash, and exacerbation of acne or hirsutism. *Mitotane* (3–6 g/d orally in four divided doses) suppresses cortisol hypersecretion by inhibiting 11β-hydroxylase and cholesterol side-chain cleavage enzymes and by destroying adrenocortical cells. Side effects of mitotane include gastrointestinal symptoms, dizziness, gynecomastia, hyperlipidemia, skin rash, and hepatic enzyme elevation. It also may lead to hypoaldosteronism. Other agents include *aminoglutethimide* (250 mg tid), *trilostane* (200–1000 mg/d), *cyproheptadine* (24 mg/d), and IV *etomidate* (0.3 mg/kg per h). Glucocorticoid insufficiency is a potential side effect of agents used to block steroidogenesis.

The use of steroidogenic inhibitors has decreased the need for bilateral adrenalectomy. Surgical removal of both adrenal glands corrects hypercortisolism but may be associated with significant morbidity rates and necessitates permanent glucocorticoid and mineralocorticoid replacement. Adrenalectomy in the setting of residual corticotrope adenoma tissue predisposes to the development of *Nelson's syndrome*, a disorder characterized by rapid pituitary tumor enlargement and increased pigmentation secondary to high ACTH levels. Prophylactic radiation therapy may be indicated to prevent the development of Nelson's syndrome after adrenalectomy.

NONFUNCTIONING AND GONADOTROPIN-PRODUCING PITUITARY ADENOMAS

Etiology and Prevalence Nonfunctioning pituitary adenomas include those that secrete little or no pituitary hormones into the systemic circulation, as well as tumors that produce too little hormone to result in recognizable clinical features. They are the most common type of pituitary adenoma and are usually macroadenomas at the time of diagnosis because clinical features are not apparent until tumor mass effects occur. Based on immunohistochemistry, most clinically nonfunctioning adenomas can be shown to originate from gonadotrope cells or from pituitary null cells. These tumors typically produce small amounts of intact gonadotropins (usually FSH) as well as uncombined α, LH β, and FSH β subunits. Tumor secretion may lead to elevated α and FSH β subunits and, very rarely, to increased LH β subunit levels. Some adenomas express α subunits without FSH or LH. A TRH stimulation test often induces an atypical increase of tumor-derived gonadotropins or subunits.

Presentation and Diagnosis Clinically nonfunctioning tumors often present with optic chiasm pressure and other symptoms of local expansion or may be incidentally discovered on an MRI performed for another indication (incidentaloma). Rarely, menstrual disturbances or ovarian hyperstimulation occur in women with large tumors that produce FSH and LH. In these cases, ovaries may have features that resemble polycystic ovarian syndrome and may produce very high levels of estrogen. More commonly, adenoma compression of the pituitary stalk or surrounding pituitary tissue leads to attenuated LH and features of hypogonadism. PRL levels are usually slightly increased, also because of stalk compression. It is important to distinguish this circumstance from true prolactinomas, as nonfunctioning tumors do not shrink in response to treatment with dopamine agonists.

Laboratory Investigation The goal of laboratory testing in clinically nonfunctioning tumors is to classify the type of tumor, identify hormonal markers of tumor activity, and detect possible hypopituitarism. Free α subunit levels may be elevated in 10–15% of patients with nonfunctioning tumors. In female patients, peri- or postmenopausal basal FSH concentrations are difficult to distinguish from tumor-derived FSH elevation. Premenopausal women have cycling FSH levels, also preventing clear-cut diagnostic distinction from tumor-derived FSH. In men, gonadotropin-secreting tumors may be diagnosed because of slightly increased gonadotropins (FSH > LH) in the setting of a pituitary mass. Testosterone levels are usually low despite the normal or increased LH level, perhaps reflecting reduced LH bioactivity or the loss of normal LH pulsatility. Because this pattern of hormone test results is also seen in primary gonadal failure and, to some extent, with aging (Chap. 403), the finding of increased gonadotropins alone is insufficient for the diagnosis of a gonadotropin-secreting tumor. In the majority of patients with gonadotrope adenomas, TRH administration stimulates LH β subunit secretion; this response is not seen in normal individuals. GnRH testing, however, is not helpful for making the diagnosis. For nonfunctioning and gonadotropin-secreting tumors, the diagnosis usually rests on immunohistochemical analyses of surgically resected tumor tissue, as the mass effects of these tumors usually necessitate resection.

Although acromegaly or Cushing's disease usually presents with unique clinical features, clinically inapparent (silent) somatotrope or corticotrope adenomas may only be diagnosed by immunostaining of resected tumor tissue. These silent tumors usually grow more aggressively and account for up to 20% of all nonfunctioning adenomas. If PRL levels are <100 μg/L in a patient harboring a pituitary mass, a nonfunctioning adenoma causing pituitary stalk compression should be considered.

TREATMENT

Nonfunctioning and Gonadotropin-Producing Pituitary Adenomas

As the probability of nonfunctioning microadenoma growth is very low, asymptomatic small nonfunctioning microadenomas with no threat to vision may be followed with an MRI after 3 years, with serial MRI and visual field testing thereafter as needed. However, for macroadenomas, transsphenoidal surgery is indicated to reduce tumor size and relieve compressive mass effects (Fig. 392-10). Although it is not usually possible to remove all adenoma tissue surgically, vision improves in 70% of patients with preoperative visual field defects. Preexisting hypopituitarism that results from tumor mass effects may improve or resolve completely. Beginning ~6 months postoperatively, MRI scans should be performed yearly to detect whether tumor regrowth is occurring. Within 5–6 years after successful surgical resection, ~15% of nonfunctioning tumors recur. When substantial tumor remains after transsphenoidal surgery, adjuvant radiotherapy may be indicated to prevent persistent tumor regrowth. Radiotherapy may be deferred if no postoperative residual mass is evident. Nonfunctioning pituitary tumors respond poorly to dopamine agonist treatment, and SRLs are largely ineffective for shrinking these tumors.

■ TSH-SECRETING ADENOMAS

TSH-producing macroadenomas are very rare but are often large and locally invasive when they occur. Patients usually present with thyroid goiter and hyperthyroidism, reflecting chronic overproduction of TSH.

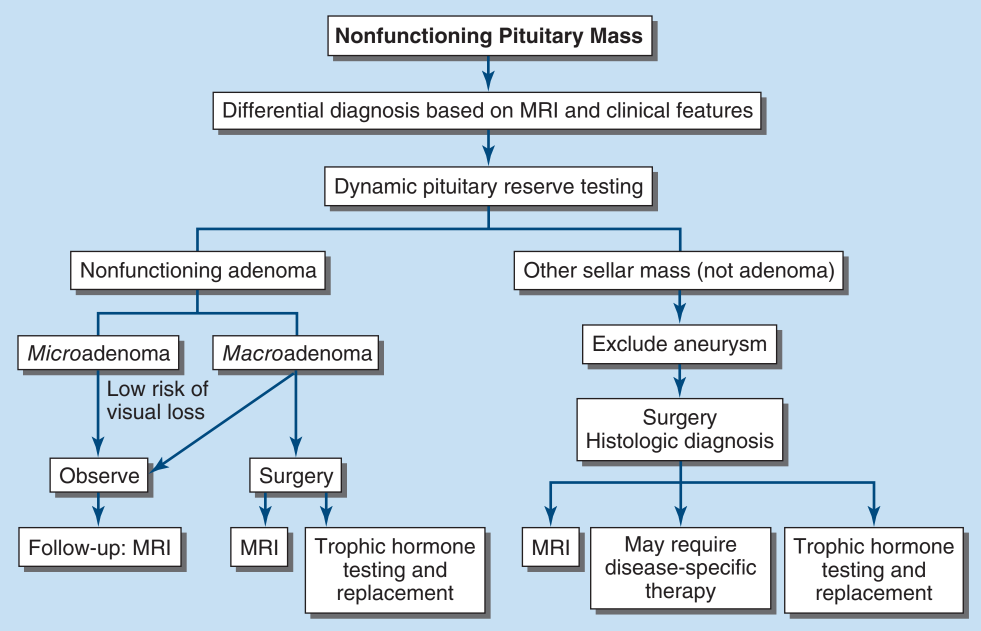
Nonfunctioning Pituitary Mass
Differential diagnosis based on MRI and clinical features
Dynamic pituitary reserve testing
Nonfunctioning adenoma
Other sellar mass (not adenoma)
*Micro*adenoma
*Macro*adenoma
Low risk of visual loss
Exclude aneurysm
Observe
Surgery
Surgery Histologic diagnosis
Follow-up: MRI
MRI
Trophic hormone testing and replacement
MRI
May require disease-specific therapy
Trophic hormone testing and replacement

FIGURE 392-10 Management of a nonfunctioning pituitary mass. MRI, magnetic resonance imaging.

Diagnosis is based on demonstrating elevated serum free T_4 levels, inappropriately normal or high TSH secretion, and MRI evidence of a pituitary adenoma. Elevated free glycoprotein hormone α subunits are seen in many patients.

It is important to exclude other causes of inappropriate TSH secretion, such as resistance to thyroid hormone, an autosomal dominant disorder caused by mutations in the thyroid hormone β receptor **(Chap. 394)**. The presence of a pituitary mass and elevated β subunit levels are suggestive of a TSH-secreting tumor. Dysalbuminemic hyperthyroxinemia syndromes, caused by mutations in serum thyroid hormone binding proteins, are also characterized by elevated thyroid hormone levels, but with normal rather than suppressed TSH levels. Moreover, free thyroid hormone levels are normal in these disorders, most of which are familial.

TREATMENT

TSH-Secreting Adenomas

The initial therapeutic approach is to remove or debulk the tumor mass surgically, usually using a transsphenoidal approach. Total resection is not often achieved as most of these adenomas are large and locally invasive. Normal circulating thyroid hormone levels are achieved in about two-thirds of patients after surgery. Thyroid ablation or antithyroid drugs (methimazole and propylthiouracil) can be used to reduce thyroid hormone levels. SRL treatment effectively normalizes TSH and α subunit hypersecretion, shrinks the tumor mass in 50% of patients, and improves visual fields in 75% of patients; euthyroidism is restored in most patients. Because SRLs markedly suppress TSH, biochemical hypothyroidism often requires concomitant thyroid hormone replacement, which may also further control tumor growth.

AGGRESSIVE ADENOMAS

Despite the rarity of malignant transformation and metastatic lesions, a subset of pituitary adenomas undergoes aggressive local growth and central nervous system invasion with high Ki67 levels (>4%). Silent corticotrope and somatotrope tumors, as well as prolactinomas occurring in middle-aged men, are particularly prone to aggressive growth and recurrence. Patients with these tumors usually require an integrated management approach including repeat surgeries and irradiation. Temozolomide has also been used with variable responses.

FURTHER READING

Coopmans EC et al: Multivariable prediction model for biochemical response to first-generation somatostatin receptor ligands in acromegaly. J Clin Endocrinol Metab 105:2964, 2020.

Elbelt U et al: Efficacy of temozolomide therapy in patients with aggressive pituitary adenomas and carcinomas: A German survey. J Clin Endocrinol Metab 105:e660, 2020.

Fleseriu M et al: Acromegaly: Pathogenesis, diagnosis, and management. Lancet Diabetes Endocrinol 10:804, 2022.

Fleseriu M et al: An individualized approach to the management of Cushing disease. Nat Rev Endocrinol 19:581, 2023.

Hamblin R et al: Natural history of non-functioning pituitary microadenomas: Results from the UK non-functioning pituitary adenoma consortium. Eur J Endocrinol 189:87, 2023.

Melmed S: Pituitary-tumor endocrinopathies. N Engl J Med 382:937, 2020.

Neou M et al: Pangenomic classification of pituitary neuroendocrine tumors. Cancer Cell 37:123, 2020.

Petersenn S et al: Diagnosis and management of prolactin-secreting pituitary adenomas: A Pituitary Society international Consensus Statement. Nat Rev Endocrinol 19:722, 2023.

Samson SL et al: Maintenance of acromegaly control in patients switching from injectable somatostatin receptor ligands to oral octreotide. J Clin Endocrinol Metab 105:e3785, 2020.

393

Disorders of the Neurohypophysis

Mirjam Christ-Crain, Mark Sherlock

The posterior pituitary consists of the distal axons of the hypothalamic magnocellular neurons that make up the neurohypophysis. The perikarya (cell bodies) of these axons are located in paired paraventricular and supraoptic nuclei of the hypothalamus. Some of these neurons produce arginine vasopressin (AVP), also known as antidiuretic hormone (ADH); others produce oxytocin. AVP acts on the renal tubules to reduce water loss by concentrating the urine. Oxytocin stimulates postpartum milk letdown in response to suckling, and also elicits socioemotional responses. A deficiency of AVP secretion or action causes a syndrome characterized by the production of large amounts of dilute urine. Excessive or inappropriate AVP production impairs urinary water excretion and predisposes to hyponatremia.

VASOPRESSIN

SYNTHESIS AND SECRETION

AVP is a nonapeptide composed of a six-member disulfide ring and a tripeptide tail **(Fig. 393-1)**. It is synthesized via a polypeptide precursor that includes AVP, neurophysin, and copeptin, all encoded by a single gene on chromosome 20. After preliminary processing and folding, the precursor is packaged in neurosecretory vesicles, where it is transported down the axon. It is further processed to AVP, neurophysin, and copeptin, and stored in neurosecretory vesicles in the posterior pituitary until it is released by exocytosis into peripheral blood.

In healthy individuals, AVP secretion is regulated primarily by the "effective" osmotic pressure, which is determined largely by the plasma concentration of sodium and its anions. This regulation is mediated by specialized cells in the anteromedial hypothalamus, known as *osmoreceptors*. The osmoreceptors receive blood from small perforating branches of the anterior communicating artery. They are extremely sensitive to small changes in the plasma concentration of sodium and its anions but normally are insensitive to other naturally occurring plasma solutes such as urea and glucose. This osmoregulatory system includes inhibitory as well as stimulatory components that function in concert to create an osmotic threshold, or set point, control system. Below this osmotic threshold, plasma AVP is suppressed to levels that permit the development of a maximum water diuresis. Above the threshold, plasma AVP rises steeply in direct proportion to plasma osmolarity, quickly reaching levels sufficient to produce maximum antidiuresis. The absolute levels of plasma osmolarity/sodium at which minimally and maximally effective levels of plasma AVP occur differ from person to person, apparently due to genetic influences on the set and sensitivity of the system. However, the average threshold, or set point, for AVP release corresponds to a plasma osmolarity and sodium of ~275 mosmol/L and 135 meq/L, respectively; levels only 2–4% higher normally result in maximum antidiuresis.

AVP is also secreted in response to a decrease in blood pressure or by volume loss of >10–20%. These hemodynamic (baroregulated) influences are mediated by neuronal afferents that originate in transmural pressure receptors of the heart and large arteries and project via the vagus and glossopharyngeal nerves to the brainstem, which sends postsynaptic projections to the hypothalamus. AVP secretion also can be stimulated by nausea, acute hypoglycemia, glucocorticoid deficiency, smoking, and possibly angiotensin. Emetic stimuli are extremely potent in comparison to osmotic stimuli; they typically elicit immediate, 50- to 100-fold increases in plasma AVP even when the nausea is transient and not associated with vomiting or other symptoms. They act via the emetic center in the medulla and can be blocked completely by treatment with antiemetics such as fluphenazine. There is no evidence that pain or other noxious stresses have any effect on AVP unless they elicit a vasovagal reaction and its associated nausea and hypotension.

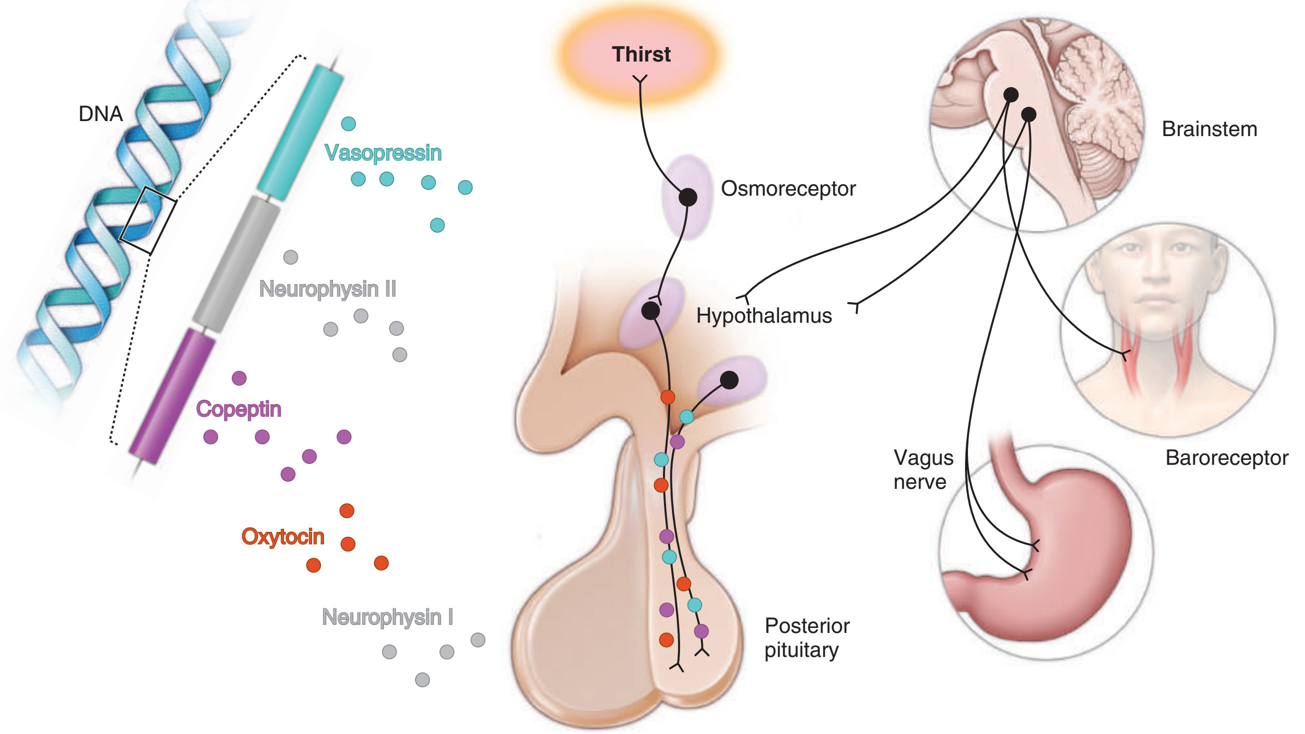

FIGURE 393-1 Primary structure, production, and release of arginine vasopressin (AVP). AVP is a nonapeptide composed of a six-member disulfide ring and a tripeptide tail. It is synthesized via a polypeptide precursor that includes AVP, neurophysin, and copeptin, all encoded by a single gene on chromosome 20. The precursor hormone, pre-pro-vasopressin consists of three peptides: AVP, neurophysin 2, and copeptin.

■ ACTION

The most important physiologic action of AVP is to reduce water excretion by promoting the concentration of urine. Other physiologic actions of AVP include stimulating ACTH and vasoconstriction. This antidiuretic effect is achieved primarily by increasing the hydro-osmotic permeability of principal cells that line the distal tubule and medullary collecting ducts of the kidney **(Fig. 393-2)**. In the absence of AVP, these cells are impermeable to water and reabsorb little, if any, of the relatively large volume of dilute filtrate that enters from the proximal nephron. In this condition, the rate of urine output can be as high as 0.2 mL/kg per min and the specific gravity and osmolarity as low as ~1.000 and 50 mOsmol/L, respectively. When AVP is secreted, it binds to V_2 receptors on the basal surface of principal cells causing water channels composed of aquaporin-2 (AQ-2) to be inserted into the apical surface of the cell. These channels allow water to flow passively from the lumen through the cell down the osmotic gradient created by the hypertonicity of the renal medulla. The magnitude of this antidiuretic effect varies in direct proportion to plasma AVP, the rate of solute excretion, and the level of hypertonicity in the renal medulla. The maximum antidiuresis achievable in healthy humans occurs at plasma AVP concentrations in the range of 1 to 3 pg/mL and results in a urine osmolarity as high as 1200 mOsmol/L. However, maximum concentrating capacity varies considerably depending on the level of hypertonicity in the renal medulla and that, in turn, is a function of the level and duration of AVP receptor 2 (AVPR2)–stimulated reabsorption of urea in the distal nephron. Hence, if basal AVP stimulation of AVPR2 is low (e.g., a high basal fluid intake in primary polydipsia), the rise in urine osmolarity that occurs immediately after an increase in AVP concentrations may be so blunted as to suggest a defect in antidiuretic function. This reduced concentrating capacity accounts for the shortcomings of the traditional indirect methods for the differential diagnosis of polyuric states (see below).

At high concentrations, AVP also causes contraction of smooth muscle in blood vessels in the skin and gastrointestinal tract, induces glycogenolysis in the liver, and potentiates ACTH release by

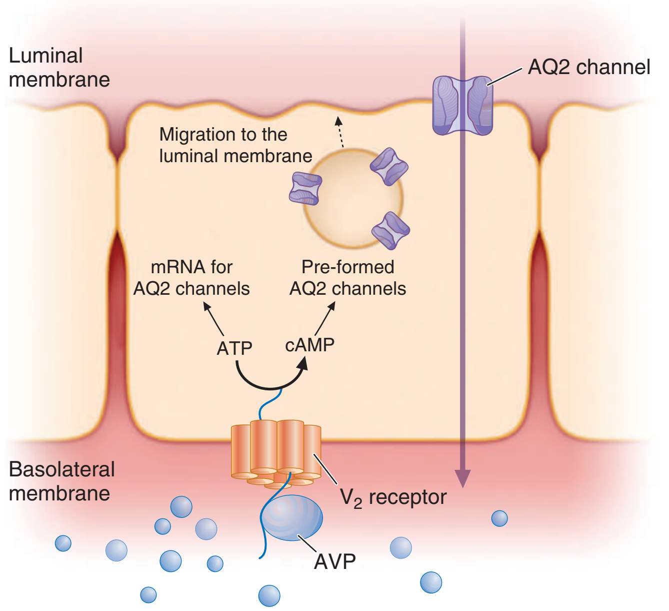

FIGURE 393-2 Antidiuretic effect of arginine vasopressin (AVP) in the regulation of urine volume. In a typical 70-kg adult, the kidney filters ~180 L/d of plasma. Of this, ~144 L (80%) is reabsorbed isosmotically in the proximal tubule and another 8 L (4–5%) is reabsorbed without solute in the descending limb of Henle's loop. In the presence of AVP, solute-free water is reabsorbed osmotically through the principal cells of the collecting ducts, resulting in the excretion of a much smaller volume of concentrated urine. This antidiuretic effect is mediated via a G protein–coupled V_2 receptor that increases intracellular cyclic AMP, thereby inducing translocation of aquaporin 2 (AQP 2) water channels into the apical membrane. The resultant increase in permeability permits an influx of water that diffuses out of the cell through AQP 3 and AQP 4 water channels on the basal-lateral surface. The net rate of flux across the cell is determined by the number of AQP 2 water channels in the apical membrane and the strength of the osmotic gradient between tubular fluid and the renal medulla.

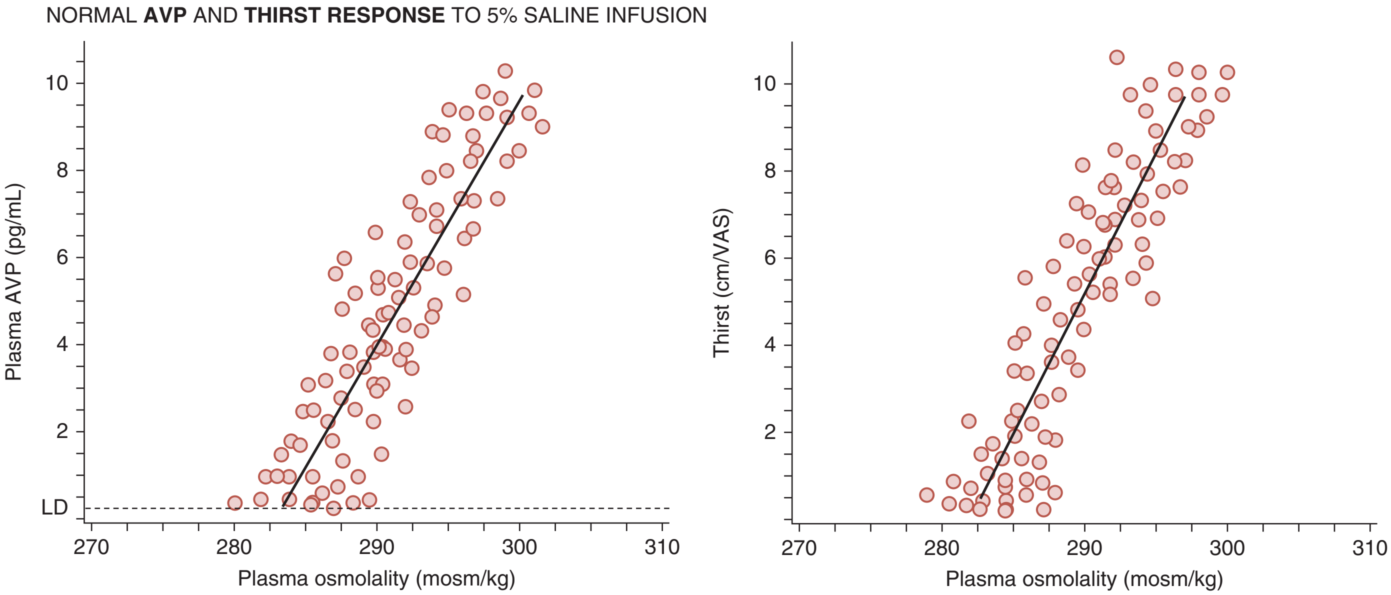

FIGURE 393-3 The relationship of plasma osmolality to arginine vasopressin (AVP) secretion and thirst. VAS, visual analogue scale.

corticotropin-releasing factor. These effects are mediated by V_{1a} or V_{1b} receptors that are coupled to phospholipase C. They may also affect the sensitivity of the baroreceptor and influence sympathetic and parasympathetic outflows to a variety of target organs, including the heart, the peripheral vasculature, and the kidneys.

■ METABOLISM

AVP distributes rapidly into a space approximately equal to the extracellular fluid volume. It is cleared irreversibly with a half-life ($t_{1/2}$) of 10–30 min. Most AVP clearance is due to degradation in the liver and kidneys. During pregnancy, the metabolic clearance of AVP is increased three- to fourfold due to placental production of an N-terminal peptidase (vasopressinase). Importantly, the synthetic vasopressin analogue desmopressin (DDAVP) is more resistant to N-terminal peptidases, resulting in a longer half-life.

THIRST

Because AVP cannot reduce water loss below a certain minimum level obligated by urinary solute load and evaporation from skin and lungs, a mechanism for ensuring adequate water intake is essential for preventing dehydration. This vital function is performed by the thirst mechanism. Like AVP, thirst and fluid intake are regulated primarily by an osmostat that is localized in the anteromedial hypothalamus and detects very small changes in the plasma concentration of sodium and its anions. The thirst osmostat appears to be "set" about 3% higher than the AVP osmostat **(Fig. 393-3)**. This relationship ensures that thirst, polydipsia, and dilution of body fluids do not occur until plasma osmolarity/sodium exceeds the defensive capacity of the antidiuretic mechanism. Defects in thirst result in hypodipsia/adipsia. The gastrointestinal tract also has a mechanism that detects fluid intake and inhibits thirst and AVP secretion before water is absorbed sufficiently to lower plasma osmolarity/sodium. However, the resultant inhibition of thirst and AVP is transient unless plasma osmolarity/sodium is reduced, and the role of this system in clinical disorders of water balance has not been determined.

OXYTOCIN

Oxytocin (OXT) is also a nonapeptide that differs from AVP at positions 3 and 8 (Fig. 393-1). However, it has relatively little antidiuretic effect and seems to act mainly on mammary ducts to facilitate milk letdown during nursing. It also may help initiate or facilitate labor by stimulating contraction of uterine smooth muscle, but it is not clear if this action is physiologic or necessary for normal delivery. In addition to its release from axonal terminals, OXT is dendritically released into the central extracellular space and directly projected to other brain regions, where it acts as a neurotransmitter. The central oxytocinergic system is key in regulating socioemotional functioning, including attachment and pair bonding, fear extinction, emotion recognition, and empathy.

DEFICIENCIES OF AVP SECRETION AND ACTION

■ DIABETES INSIPIDUS (AVP DEFICIENCY AND AVP RESISTANCE)

Clinical Characteristics Deficiencies in AVP secretion or action result in the excretion of abnormally large volumes of dilute urine. The 24-h urine volume exceeds 40–50 mL/kg body weight and consequently leads to polydipsia. Signs and symptoms of dehydration (and biochemical hypernatremia) are uncommon unless thirst and/or water intake are also impaired.

Etiology AVP deficiency and AVP resistance should be differentiated from increased AVP metabolism in pregnancy and from primary polydipsia. Primary deficiency of AVP secretion was formerly called *neurogenic, pituitary*, *cranial*, or *central diabetes insipidus* but is now referred to as AVP deficiency. It can be caused by a variety of acquired, congenital, or genetic disorders but is often idiopathic **(Table 393-1)**. The most common genetic form is transmitted in an autosomal dominant mode and is caused by diverse mutations in the coding region of one allele of the AVP–neurophysin II (or *AVP-NPII*) gene. Renal insensitivity to the antidiuretic action of AVP leads to AVP resistance, which was formerly known as nephrogenic diabetes insipidus. It can be caused by a drug such as lithium, a disorder such as hypokalemia and hypercalcemia, or by a genetic mutation.

In pregnancy, increased metabolism of AVP may occur due to AVP degradation by an N-terminal aminopeptidase (vasopressinase) produced in the placenta. This is referred to as *gestational AVP deficiency* because the signs and symptoms manifest during pregnancy and usually remit several weeks after delivery.

These forms of AVP deficiency and AVP resistance should be differentiated from excessive intake of fluids, which is commonly referred to as *primary polydipsia.* This disorder is common in patients with neurodevelopmental or psychotic disorders, particularly chronic schizophrenia. Outside the psychiatric setting, it is increasingly seen in the general population owing to the popularity of lifestyle programs and the belief that drinking large amounts of water is healthy and improves cognition (Table 393-1).

TABLE 393-1 Etiology of Polyuria–Polydipsia Syndromes

BASIC DEFECT	ACQUIRED CAUSES	HEREDITARY CAUSES
AVP Deficiency		
Deficiency in AVP synthesis or secretion	• Trauma (surgery, deceleration injury) • Neoplasia (craniopharyngioma, meningioma, germinoma, metastases) • Vascular (cerebral or hypothalamic hemorrhage, infarction or ligation of anterior communicating artery aneurysm) • Granulomatous (histiocytosis, sarcoidosis) • Infectious (meningitis, encephalitis, tuberculosis) • Inflammatory or autoimmune (lymphocytic infundibuloneurohypophysitis, IgG4 neurohypophysitis) • Drug or toxin exposure • Osmoreceptor dysfunction (adipsic DI) • Others (hydrocephalus, ventricular or suprasellar cyst, trauma, and degenerative diseases) • Idiopathic	• Autosomal dominant: *AVP* mutations • Autosomal recessive, type a and b: *AVP* mutations • Autosomal recessive, type c: *WFS1* mutations • Autosomal recessive, type d: *PCSK1* mutations • X-linked recessive: gene unknown
AVP Resistance		
Reduced renal sensitivity to antidiuretic effect of physiologic AVP levels	• Drug exposure (lithium, demeclocycline, cisplatin, etc.) • Hypercalcemia or hypokalemia • Infiltrating lesions (sarcoidosis, amyloidosis, multiple myeloma, etc.) • Vascular disorders (sickle cell anemia) • Mechanical (polycystic kidney disease and urethral obstruction)	• X-linked: *AVPR2* mutations • Autosomal recessive or dominant: *AQP2* mutations
Primary Polydipsia		
Excessive fluid intake at a diminished set point	• Dipsogenic[a] (idiopathic or similar lesions as with central DI) • Psychosis intermittent hyponatremia–polydipsia (PIP) syndrome • Compulsive water drinking • Health enthusiasts	NA
Gestational AVP Deficiency		
Increased enzymatic metabolism of circulating AVP hormone	Pregnancy	NA

[a] Downward resetting of the thirst threshold.

Abbreviations: AVP, arginine vasopressin; AVPR, AVP receptor; DI, diabetes insipidus; NA, not applicable; *PCSK1*, proprotein convertase subtilisin/kexin type 1; *WFS1*, Wolfram syndrome 1.

Source: Reproduced with permission from M Christ-Crain et al: Diabetes insipidus. Nat Rev Dis Primers 5:54, 2019.

Pathophysiology In AVP deficiency and resistance, the defect in urine concentration increases the rate of water excretion and causes a small (1–2%) decrease in body water and a commensurate increase in plasma osmolarity/sodium, which stimulates thirst and a compensatory increase in water intake. The severity of the defect in antidiuretic function varies significantly from patient to patient. In some patients, AVP deficiency is nearly complete and cannot be overcome by even an intense stimulus such as nausea or severe dehydration. In others, AVP deficiency is incomplete, and a modest stimulus such as a few hours of fluid deprivation, smoking, or a vasovagal reaction is sufficient to concentrate the urine. However, even in patients with a partial defect, the maximum level of urine osmolarity produced by these stimuli is usually less than normal partly because the prior deficiency in basal AVP stimulation temporarily diminishes renal concentrating capacity. Nevertheless, the underlying cause of the AVP deficiency/resistance can be determined by analyzing the relationship of urine osmolarity to plasma AVP/copeptin and of plasma AVP/copeptin to plasma osmolarity/sodium.

The pathophysiology of primary polydipsia is the reverse of that in AVP deficiency or resistance. The increase in fluid intake reduces plasma osmolarity/sodium and leads to a physiologic decrease in AVP secretion. The resultant urinary dilution produces a compensatory increase in urinary free-water excretion that usually offsets the increase in intake and stabilizes plasma osmolarity/sodium at a level below basal. Thus, hyponatremia is uncommon unless the polydipsia is very severe or the compensatory water diuresis is impaired (by another contributory factor that causes AVP release). In diagnostic tests, fluid deprivation or hypertonic saline infusion produces a normal rise in plasma AVP, but the resultant increase in urine concentration is usually subnormal because the capacity of the kidney to concentrate the urine is temporarily diminished by the prior lack of AVP stimulation. Thus, the maximum level of urine osmolarity achieved is often indistinguishable from that produced by fluid deprivation and/or administration of ADH in partial pituitary or partial nephrogenic diabetes insipidus. However, unlike AVP deficiency or resistance, the relationships of the rise in plasma AVP to the rise in plasma and urine osmolarity are both normal in primary polydipsia.

Differential Diagnosis If symptoms of polyuria, nocturia, and/or persistent thirst are present in the absence of glucosuria, the possibility of AVP deficiency or AVP resistance should be evaluated by collecting a 24-h urine on unrestricted fluid intake. If the volume is >40–50 mL/kg per day and/or >3 L/d, further investigations are indicated. If sodium levels are below the normal reference range (<135 mmol/L), this suggests primary polydipsia since these patients can drink themselves into hyponatremia. If sodium levels are above the normal reference range, the diagnosis of AVP deficiency or resistance is likely, and a test with desmopressin (2 μg) followed by a repeat measurement of urine osmolarity will determine if hypotonic polyuria is due to a AVP deficiency or AVP resistance. This is subcutaneous and should be done in hospital to allow reassessment of urinary osmolality. However, in most patients, sodium levels will be in the normal range, making further tests for differential diagnosis necessary **(Fig. 393-4)**.

The indirect water deprivation test was the gold standard for differential diagnosis for many years. This test is based on indirect assessment of AVP activity by measurement of the urine concentration capacity during a prolonged period of dehydration and again after a subsequent injection of an exogenous synthetic AVP analogue, desmopressin. However, the published criteria for interpretation were based on post hoc data from a small number of patients with an overall

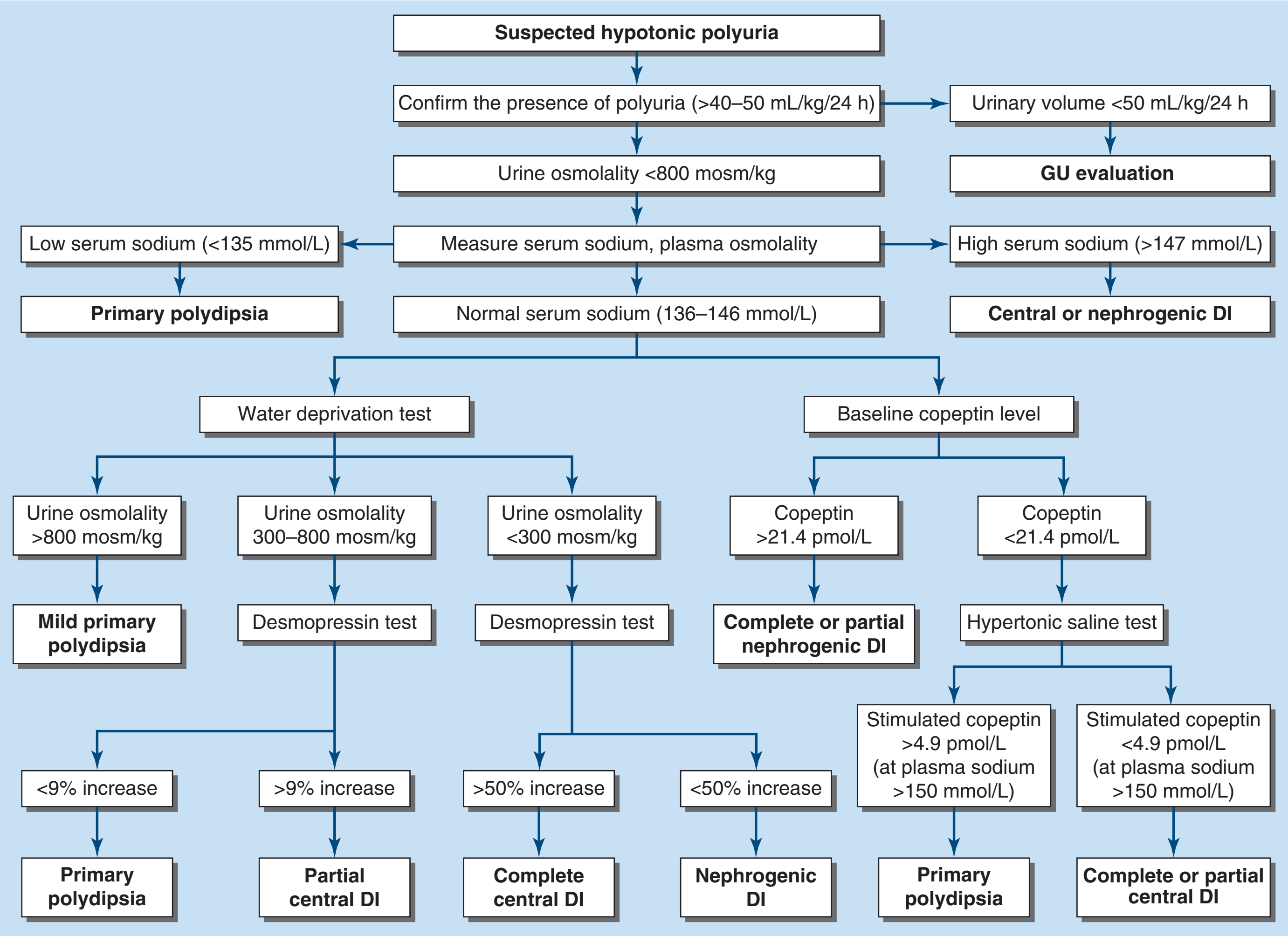
Suspected hypotonic polyuria
Confirm the presence of polyuria (>40–50 mL/kg/24 h)
Urinary volume <50 mL/kg/24 h
Urine osmolality <800 mosm/kg
GU evaluation
Low serum sodium (<135 mmol/L)
Measure serum sodium, plasma osmolality
High serum sodium (>147 mmol/L)
Primary polydipsia
Normal serum sodium (136–146 mmol/L)
Central or nephrogenic DI
Water deprivation test
Baseline copeptin level
Urine osmolality >800 mosm/kg
Urine osmolality 300–800 mosm/kg
Urine osmolality <300 mosm/kg
Copeptin >21.4 pmol/L
Copeptin <21.4 pmol/L
Mild primary polydipsia
Desmopressin test
Desmopressin test
Complete or partial nephrogenic DI
Hypertonic saline test
Stimulated copeptin >4.9 pmol/L (at plasma sodium >150 mmol/L)
Stimulated copeptin <4.9 pmol/L (at plasma sodium >150 mmol/L)
<9% increase
>9% increase
>50% increase
<50% increase
Primary polydipsia
Partial central DI
Complete central DI
Nephrogenic DI
Primary polydipsia
Complete or partial central DI

FIGURE 393-4 Algorithm for differential diagnosis of polyuria polydipsia syndrome. If symptoms of polyuria, nocturia, and/or persistent thirst are present in the absence of glucosuria, the possibility of arginine vasopressin (AVP) deficiency or AVP resistance should be evaluated by collecting a 24-h urine on unrestricted fluid intake. If the volume is >50 mL/kg per day with a concomitant urinary osmolality <800 mOsm/kg, serum sodium and plasma osmolality should be measured. If sodium levels are below the normal reference range, it suggests primary polydipsia since these patients can drink themselves into hyponatremia. If sodium levels are above the normal reference range, the diagnosis of AVP deficiency or resistance can be made, and a test with desmopressin (2 μg) followed by a repeat measurement of urine osmolarity will determine if hypotonic polyuria is due to a AVP deficiency or AVP resistance. However, in most patients, sodium levels will be in the normal range, making further tests for differential diagnosis necessary. If copeptin measurement is available, a copeptin-based diagnostic algorithm is used. High baseline copeptin level of >21.4 pmol/L without prior water deprivation identifies AVP resistance. For the more difficult differential diagnosis of AVP deficiency and primary polydipsia, a copeptin level of >4.9 pmol/L at a high sodium level (≥150 mmol/L) after hypertonic saline infusion has an overall diagnostic accuracy of 97% to diagnose AVP deficiency. DI, diabetes insipidus; GU, genitourinary.

diagnostic accuracy of 70% and only 41% for patients with primary polydipsia. To overcome these limitations, direct measurement of AVP was proposed, but despite initial promising results, this method is not in routine clinical use, mainly because of technical limitations of the AVP assay. Copeptin is the C-terminal segment of the AVP prohormone and is an AVP surrogate that is very stable ex vivo (Fig. 393-1). Studies have shown that a high baseline copeptin level of >21.4 pmol/L, without prior water deprivation, unequivocally identifies AVP resistance. For the more difficult differential diagnosis of AVP deficiency and primary polydipsia, a copeptin level of >4.9 pmol/L at a high sodium level (≥150 mmol/L) after hypertonic saline infusion has an overall diagnostic accuracy of 96.5%. Importantly, this test requires close monitoring of sodium levels. Copeptin levels after arginine infusion have also shown promising results in differentiating AVP deficiency from primary polydipsia, but with a lower diagnostic accuracy. Currently, copeptin assays are commercially available in Europe, Australia, India, and Mexico, and tests are pending in several other countries.

Once AVP deficiency has been diagnosed, the underlying pathology must be identified by magnetic resonance imaging (MRI) of the sella and suprasellar regions. Also, assessment of the posterior pituitary and the pituitary stalk can be helpful in the differential diagnosis of AVP deficiency. The pituitary bright spot on MRI (a radiologic marker of neurosecretory vesicles containing AVP) is an area of hyperintensity seen in most healthy individuals, but may be lacking in patients with AVP deficiency. However, the absence of the pituitary bright spot is not sufficient to establish a diagnosis of AVP deficiency since it can be present in early stages of AVP deficiency or can be absent in elderly patients.

Treatment The signs and symptoms of uncomplicated AVP deficiency can be eliminated by treatment with DDAVP, a synthetic analogue of AVP. DDAVP acts selectively at V_2 receptors to increase urine concentration and decrease urine flow in a dose-dependent manner. It is also more resistant to degradation than is AVP and has a three- to fourfold longer duration of action. The dose of DDAVP impacts on the duration of action (i.e., the higher the dose, the greater the duration of action). DDAVP can be given by IV or SC injection, nasal inhalation, or orally by means of a tablet or melt. The doses required to treat AVP deficiency vary depending on the patient and the route of administration. Among adults, doses usually range from 1–2 μg qd or bid by injection, 10–20 μg bid or tid by nasal spray, or 100–400 μg bid or tid orally. The onset of antidiuresis is rapid, ranging from as little as 15 min after injection to 60 min after oral administration. Hyponatremia is the most common complication of desmopressin therapy, with mild depression of plasma sodium concentration (131–134 mmol/L) reported in about a quarter of patients with intact thirst. In 15% of patients, hyponatremia is more severe, with plasma sodium concentration of <130 mmol/L. Desmopressin escape, which involves intermittently delaying DDAVP for a number of hours to allow a transient aquaresis, reduces the risk of hyponatremia.

Treatment of primary polydipsia focuses on the reduction of excessive fluid intake, optimally in a graded fashion to allow patients to slowly reduce fluids. Treatments to reduce mouth dryness (e.g., ice chips, hard candy to stimulate salivary flow) are also useful to reduce thirst. Pharmacologic therapies have been tried without consistent success. A recent study suggests that glucagon-like peptide 1 (GLP-1) analogues reduce fluid intake, urine output, and thirst perception. AVP resistance is difficult to treat. Patients typically do not respond to desmopressin treatment; however, some patients may respond to high doses if their resistance is only partial. Treatment with conventional doses of a thiazide diuretic and/or amiloride in conjunction with a low-sodium diet and coadministration of nonsteroidal anti-inflammatory drugs (NSAIDs) usually reduces the polyuria and polydipsia, but this combination is nephrotoxic, and careful monitoring of renal function is important. Drug-induced AVP resistance should be treated by discontinuation of the causative agent—most commonly lithium—where possible. Persistent lithium-induced AVP resistance can be treated by hydrochlorothiazide and amiloride. It important to be aware that plasma volume contraction produced by thiazide diuretics can decrease lithium excretion and predispose to lithium toxicity.

HYPODIPSIC/ADIPSIC HYPERNATREMIA

An increase in plasma osmolarity/sodium above the normal range (hypertonic hypernatremia) can be due to a decrease in total body water or an increase in total body sodium. The former results from a failure to drink enough water to replace normal or increased urinary and insensible loss due either to water deprivation or a lack of thirst (hypodipsia/adipsia).

Clinical Characteristics Hypodipsic/adipsic hypernatremia is a rare syndrome characterized by chronic or recurrent hypertonic dehydration that most frequently coexists with AVP deficiency (adipsic diabetes insipidus or adipsic AVP deficiency). The hypernatremia varies widely in severity and is often associated with signs of hypovolemia such as tachycardia, postural hypotension, azotemia, hyperuricemia, and hypokalemia due to secondary hyperaldosteronism. Muscle weakness, pain, rhabdomyolysis, hyperglycemia, hyperlipidemia, thromboembolic disease, acute renal failure, and obtundation can also occur.

Etiology Hypodipsia/adipsia is usually due to abnormalities of osmoreceptors in the anterior hypothalamus that regulate thirst. The defect can result from various congenital malformations of midline brain structures or may be acquired due to diseases such tumors (primary or secondary, and their associated surgery) or aneurysms of the anterior communicating artery, head trauma, granulomatous diseases such as sarcoidosis and histiocytosis, AIDS, and cytomegalovirus encephalitis. Adipsic hypernatremia without demonstrable hypothalamic lesions has also been associated with autoantibodies directed against the subfornical organ.

Pathophysiology A deficiency in osmotically induced thirst results in a failure to drink enough water to replenish obligatory renal and extrarenal losses with resultant significant hypernatremia. Rarely, the regulation of AVP secretion is completely normal, suggesting that the lack of thirst is due to a defect in postosmoreceptor neural pathways to higher cognitive centers.

Differential Diagnosis Hypodipsic/adipsic hypernatremia with or without coexisting AVP deficiency usually can be distinguished from other causes of inadequate fluid intake (e.g., coma, paralysis, restraints, absence of fresh water) by the clinical history and setting as well as measurements of serum and urine osmolality. Previous episodes and/or denial of thirst and failure to drink spontaneously when the patient is conscious, unrestrained, and hypernatremic are virtually diagnostic.

Treatment Hypodipsic hypernatremia can be corrected by administering water orally if the patient is alert and cooperative or by infusing hypotonic fluids (0.45% saline or 5% dextrose) if the patient is not. The amount of free water in liters required to correct the free water deficit should be estimated from body weight in kg and the serum sodium concentration in mmol/L. This amount plus an allowance for continuing insensible and urinary losses should be given over a 24- to 48-h period with close monitoring of serum sodium to ensure that it does not correct too rapidly. Plasma urea/creatinine should be monitored closely for signs of acute renal failure caused by rhabdomyolysis, hypovolemia, and hypotension. Once the patient has been rehydrated, an MRI of the brain and tests of anterior pituitary function should be performed to look for the cause and collateral defects in other hypothalamic functions. A long-term management plan to prevent or minimize recurrence of the fluid and electrolyte imbalance also should be developed. This should include a practical method to regulate fluid intake in accordance with variations in water balance as indicated by changes in body weight or serum sodium determined by home monitoring analyzers, if available. Another potential treatment approach is summarized in Table 393-2.

TABLE 393-2 Approach to the Management of Water Balance for Patients with Adipsic Arginine Vasopressin (AVP) Deficiency

1. Replace AVP with sufficient vasopressin (DDAVP).
2. Monitor fluid input/output initially as an inpatient to achieve eunatremia.
3. Weigh and record patient's eunatremic weight.
4. Recommend 1.5–2 L of fluid intake per day assuming urinary losses are less.
5. Weigh daily.
6. If below eunatremic weight, then replace with equivalent volume of fluid to restore eunatremic weight.
7. Recommend increased fluid intake in times of increased perspiration or ambient temperatures.
8. Regular plasma sodium measurements.

INAPPROPRIATE ANTIDIURESIS (SEE IN MORE DETAIL CHAP. 56)

Clinical Characteristics Syndrome of inappropriate antidiuresis (SIAD) is produced when plasma levels of AVP are elevated at times when the physiologic secretion of AVP from the posterior pituitary would normally be osmotically suppressed. The clinical abnormality is a decrease in the osmotic pressure of body fluids, such that the hallmark of SIAD is hypoosmolality. If hyponatremia is severe or develops acutely, it can cause a variety of neurologic symptoms and signs, such as headache, confusion, anorexia, nausea, vomiting, coma, and convulsions. If the hyponatremia develops gradually or exists for more than a few days, it may be apparently asymptomatic, but even mild hyponatremia is associated with an increased rate of falls and fractures, neurocognitive and neuromuscular symptoms, and increased morbidity and mortality.

Etiology SIAD has many different etiologies, which are summarized in Chap. 56.

Pathophysiology In SIAD, the failure to mount a water diuresis when intake exceeds urinary and insensible loss results in a slight expansion of total body water followed by a modest increase in urinary sodium excretion. As a result, expansion of extracellular volume is minimal, and clinically detectable edema does not develop. However, intracellular volume increases in proportion to the severity and rapidity of the change in plasma sodium. In the brain, this cellular swelling causes an increase in pressure that triggers a variety of symptoms. After several days, the swelling and symptoms may subside due to inactivation of some intracellular solutes and resultant decrease in cellular volume.

Differential Diagnosis Evaluation of urine and serum osmolality and sodium is the most useful investigation in establishing whether the diagnostic criteria for SIAD are met, alongside clinical assessment of volume status. Measurement of serum osmolality is important to exclude non-hypotonic causes of hyponatremia, such as pseudohyponatremia (due to raised lipids or protein) or translocational hyponatremia (due to raised glucose or exogenous effective osmolytes [e.g., mannitol]). SIAD must then be differentiated from other types of hypotonic hyponatremia. Measurement of urinary osmolality helps to distinguish SIAD from primary polydipsia where urinary osmolality is usually <100–200 mosm/L. Measurement of urinary sodium levels are useful to differentiate hyponatremia with low effective arterial blood volume from hyponatremia with normal effective arterial blood volume. In hypovolemic or hypervolemic hyponatremia with low effective blood volume, the mechanisms to conserve sodium remain intact, resulting in low urine sodium excretion, whereas in SIAD, there is ongoing renal sodium loss. According to guidelines, levels of <30 mmol/L argue for low effective arterial blood volume. Assessment of volume status has been used to classify hyponatremia into hypo-, hyper-, or euvolemic hyponatremia, though the accuracy of the clinical examination is poor. Hypervolemic hyponatremia typically occurs in patients with generalized edema due to severe congestive heart failure or cirrhosis. Hypovolemic hyponatremia occurs in patients with loss of sodium and water due to severe vomiting, diarrhea, diuretic use or primary adrenal insufficiency. In case of euvolemic hyponatremia, secondary adrenal insufficiency should be excluded before SIAD is diagnosed. Measurement of plasma AVP or copeptin has no diagnostic value. It is widely recommended to evaluate thyroid function in the workup of hyponatremia. There is, however, limited evidence for hypothyroidism as a significant cause of hyponatremia.

TREATMENT

Syndrome of Inappropriate Antidiuresis

The management of SIAD differs depending on the underlying etiology as well as the severity and duration of symptoms. In the presence of severe neurologic symptoms, urgent intervention and treatment with hypertonic saline are indicated, aiming for an initial sodium rise of 4–6 mmol with a 100-mL bolus of hypertonic (3%) sodium chloride administered over ~15 min. An initial dose should be followed by clinical reassessment and a repeat dose if there is no clinical response until this target increment has been achieved.

First-line treatment of chronic hyponatremia is fluid restriction, usually in the range of 500 to 1000 mL/d including all liquids, not solely water. However, fluid restriction is often ineffective, especially if urinary osmolality is >500 mosm/L. Second-line treatments when fluid restriction is inadequate are urea or tolvaptan. Urea is a product of hepatic nitrogen metabolism that is renally excreted and exerts an osmotic effect to promote free water excretion. Usually, doses of 30–60 g are sufficient to raise sodium levels. Tolvaptan is an oral vasopressin V_2-receptor antagonist that blocks AVP action in the kidney, inducing a water diuresis and raising sodium levels. Due to the risk of overly rapid correction, low-dose tolvaptan (7.5 mg) has become more widely prescribed in recent years. Other concerns with tolvaptan include cost and risk of liver function derangement seen with sustained higher-dose therapy as used for autosomal dominant polycystic kidney disease.

Demeclocycline is recommended by some as an alternative therapy, but European guidelines do not recommend it due to side effects. Demeclocycline can induce renal AVP resistance, thereby promoting dilute urine excretion. This effect is unpredictable and potential adverse effects include gastrointestinal intolerance, nephrotoxicity, and a photosensitive skin rash. According to a recent study, oral sodium chloride tablets and loop diuretics have a questionable efficacy. SGLT-2 inhibitors act at the sodium-glucose cotransporter in the proximal tubule to reduce resorption of glucose and sodium. SGLT-2 inhibitors induce glycosuria, which is accompanied by increased water excretion due to an osmotic diuresis. This has been shown to increase sodium levels in hospitalized as well as in outpatients with SIAD.

ACKNOWLEDGMENTS

The authors are grateful to Gary L. Robertson and Daniel G. Bichet for their contributions to this chapter in previous editions of Harrison's.

FURTHER READING

Adrogué HJ, Madias NE: The syndrome of inappropriate antidiuresis. N Engl J Med 389:1499, 2023.

Christ-Crain M et al: Diabetes insipidus. Nat Rev Dis Primers 5:54, 2019.

Fenske W et al: A copeptin based approach in the diagnosis of diabetes insipidus. N Engl J Med 379:428, 2018.

Refardt J et al: Arginine or hypertonic saline-stimulated copeptin to diagnose AVP deficiency. N Engl J Med 389:1877, 2023.

Tomkins M et al: Diagnosis and management of central diabetes insipidus in adults. J Clin Endocrinol Metab 107:2701, 2022.

Warren A et al: Syndrome of inappropriate antidiuresis: From pathophysiology to management. Endocr Rev 44:819, 2023.

394

Thyroid Gland Physiology and Testing

J. Larry Jameson, Anthony P. Weetman, Susan J. Mandel

The thyroid gland produces two related hormones, thyroxine (T_4) and triiodothyronine (T_3) **(Fig. 394-1)**. Acting through thyroid hormone receptors (TR) α and β, these hormones play a critical role in cell differentiation and organogenesis during development and help maintain thermogenic and metabolic homeostasis in the adult. Autoimmune disorders of the thyroid gland can stimulate overproduction of thyroid hormones (*thyrotoxicosis*) **(Chap. 396)** or cause glandular destruction and hormone deficiency (*hypothyroidism*) **(Chap. 395)**. Benign nodules and various forms of thyroid cancer are relatively common and amenable to detection by physical examination, ultrasound, and other imaging techniques **(Chap. 397)**.

ANATOMY AND DEVELOPMENT

The thyroid (Greek *thyreos*, shield, plus *eidos*, form) consists of two lobes connected by an isthmus. It is located anterior to the trachea between the cricoid cartilage and the suprasternal notch. The normal thyroid is 12–20 g in size, highly vascular, and soft in consistency. Four parathyroid glands, which produce parathyroid hormone **(Chap. 422)**, are located posterior to each pole of the thyroid. The recurrent laryngeal nerves traverse the lateral borders of the thyroid gland and must be identified during thyroid surgery to avoid injury and vocal cord paralysis.

The thyroid gland develops from the floor of the primitive pharynx during the third week of gestation. The developing gland migrates along the thyroglossal duct to reach its final location in the neck. This feature accounts for the rare ectopic location of thyroid tissue at the base of the tongue (lingual thyroid) as well as the occurrence of thyroglossal duct cysts along this developmental tract. Thyroid hormone synthesis begins at about 11 weeks' gestation.

Neural crest derivatives from the ultimobranchial body give rise to thyroid medullary C cells that produce calcitonin, a calcium-lowering hormone. The C cells are interspersed throughout the thyroid gland, although their density is greatest in the juncture of the upper one-third and lower two-thirds of the gland. Calcitonin plays a minimal role in calcium homeostasis in humans, but the C cells are clinically important because they are the cellular origin of medullary thyroid cancer **(Chap. 400)**.

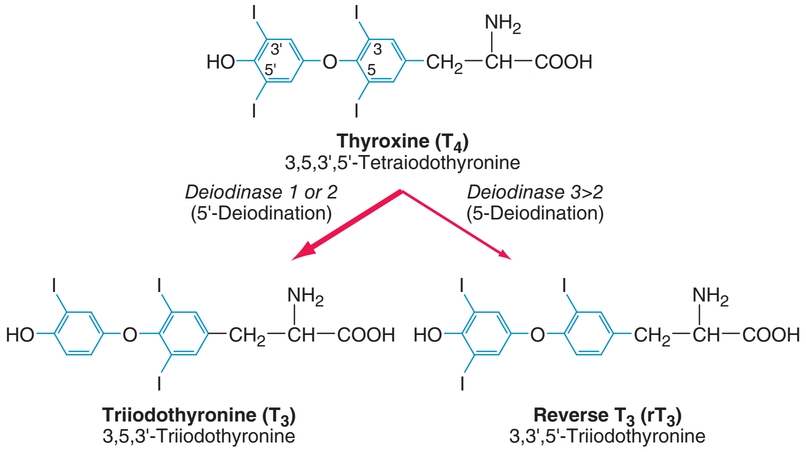

FIGURE 394-1 Structures of thyroid hormones. Thyroxine (T_4) contains four iodine atoms. Deiodination leads to production of the potent hormone triiodothyronine (T_3) or the inactive hormone reverse T_3.

Thyroid gland development is orchestrated by the coordinated expression of several developmental transcription factors. Thyroid transcription factor genes *TTF1*, *TTF2*, *NKX2-1*, *FOXE1*, and paired homeobox-8 (*PAX8*) are expressed selectively, but not exclusively, in the thyroid gland. In combination, they dictate thyroid cell development and the induction of thyroid-specific proteins such as thyroglobulin (Tg), thyroid peroxidase (TPO), the sodium iodide symporter (Na^+/I^-, NIS), and the thyroid-stimulating hormone (TSH) receptor (TSH-R). Mutations in these developmental transcription factors or their downstream target genes are rare causes of thyroid agenesis or dyshormonogenesis, although the causes of most forms of congenital hypothyroidism remain unknown **(see Chap. 395, Table 395-1)**. Because congenital hypothyroidism occurs in ~1 in 4000 newborns, neonatal screening is now performed in most industrialized countries. Transplacental passage of maternal thyroid hormone occurs before the fetal thyroid gland begins to function and provides significant hormone support to a fetus with congenital hypothyroidism. Early thyroid hormone replacement in newborns with congenital hypothyroidism prevents potentially severe developmental abnormalities.

The thyroid gland consists of numerous spherical follicles composed of thyroid follicular cells that surround secreted colloid, a proteinaceous fluid containing large amounts of thyroglobulin, the protein precursor of thyroid hormones **(Fig. 394-2)**. The thyroid follicular cells are polarized—the basolateral surface is apposed to the bloodstream and an apical surface faces the follicular lumen. Increased demand for thyroid hormone is regulated by TSH, which binds to its receptor on the basolateral surface of the follicular cells. This binding leads to Tg reabsorption from the follicular lumen and proteolysis within the cytoplasm, yielding thyroid hormones for secretion into the bloodstream.

REGULATION OF THE THYROID AXIS

TSH, secreted by the thyrotrope cells of the anterior pituitary, plays a pivotal role in control of the thyroid axis and serves as the most useful physiologic marker of thyroid hormone action. TSH is a 31-kDa hormone composed of α and β subunits; the α subunit is common to the other glycoprotein hormones (luteinizing hormone, follicle-stimulating hormone, human chorionic gonadotropin [hCG]), whereas the TSH β subunit is unique to TSH. The extent and nature of carbohydrate modification are modulated by thyrotropin-releasing hormone (TRH) and influence the biologic activity of the hormone.

The thyroid axis is a classic example of an endocrine feedback loop **(Chap. 389)**. Hypothalamic TRH stimulates pituitary production of TSH, which, in turn, stimulates thyroid hormone synthesis and secretion. Thyroid hormones act via negative feedback predominantly through thyroid hormone receptor β2 (TRβ2) to inhibit TRH and TSH production (Fig. 394-2). The "set point" in this axis is established by TSH. TRH is the major positive regulator of TSH synthesis and secretion. Peak TSH secretion occurs ~15 min after administration of exogenous TRH. Dopamine, glucocorticoids, and somatostatin suppress TSH but are not of major physiologic importance except when these agents are administered in pharmacologic doses. Reduced levels of thyroid hormone increase basal TSH production and enhance TRH-mediated stimulation of TSH. High thyroid hormone levels rapidly and directly suppress TSH gene expression and inhibit TRH stimulation of TSH secretion, indicating that thyroid hormones are the dominant regulator of TSH production. Like other pituitary hormones, TSH is released in a pulsatile manner and exhibits a diurnal rhythm; its highest levels occur at night. However, these TSH excursions are modest in comparison to those of other pituitary hormones, in part, because TSH has a relatively long plasma half-life (50 min). Consequently, single measurements of TSH are adequate for assessing its circulating level. TSH is measured using immunoradiometric assays that are highly sensitive and specific. These assays readily distinguish between normal and suppressed TSH values; thus, TSH can be used for the diagnosis of primary hyperthyroidism (low TSH) or primary hypothyroidism (high TSH).

THYROID HORMONE SYNTHESIS, METABOLISM, AND ACTION

THYROID HORMONE SYNTHESIS

Thyroid hormones are derived from Tg, a large iodinated glycoprotein. After secretion into the thyroid follicle, Tg is iodinated on tyrosine

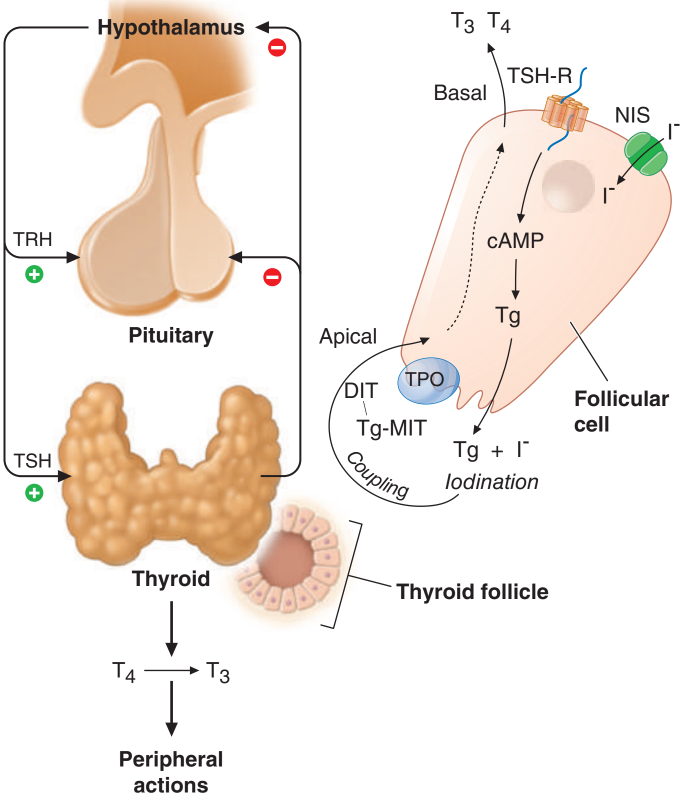

FIGURE 394-2 Regulation of thyroid hormone synthesis. ***Left.*** Thyroid hormones T_4 and T_3 feed back to inhibit hypothalamic production of thyrotropin-releasing hormone (TRH) and pituitary production of thyroid-stimulating hormone (TSH). TSH stimulates thyroid gland production of T_4 and T_3. ***Right.*** Thyroid follicles are formed by thyroid epithelial cells surrounding proteinaceous colloid, which contains thyroglobulin. Follicular cells, which are polarized, synthesize thyroglobulin and carry out thyroid hormone biosynthesis (see text for details). DIT, diiodotyrosine; MIT, monoiodotyrosine; NIS, sodium iodide symporter; Tg, thyroglobulin; TPO, thyroid peroxidase; TSH-R, thyroid-stimulating hormone receptor.

residues that are subsequently coupled via an ether linkage. Reuptake of Tg into the thyroid follicular cell allows proteolysis and the release of newly synthesized T_4 and T_3.

Iodine Metabolism and Transport Iodide uptake is a critical first step in thyroid hormone synthesis. Ingested iodine is bound to serum proteins, particularly albumin. Unbound iodine is excreted in the urine. The thyroid gland extracts iodine from the circulation in a highly efficient manner. For example, 10–25% of radioactive tracer (e.g., ^{123}I) is taken up by the normal thyroid gland over 24 h in an iodine-replete state; this value can rise to 70–90% in Graves' disease. Iodide uptake is mediated by NIS, which is expressed at the basolateral membrane of thyroid follicular cells. NIS is most highly expressed in the thyroid gland, but low levels are present in the salivary glands, lactating breast, and placenta. The iodide transport mechanism is highly regulated, allowing adaptation to variations in dietary supply. Low iodine levels increase the amount of NIS and stimulate uptake, whereas high iodine levels suppress NIS expression and uptake. The selective expression of NIS in the thyroid allows isotopic scanning, treatment of hyperthyroidism, and ablation of thyroid cancer with radioisotopes of iodine, without significant effects on other organs. Mutation of the *NIS* gene is a rare cause of congenital hypothyroidism, underscoring its importance in thyroid hormone synthesis. Another iodine transporter, pendrin, is located on the apical surface of thyroid cells and mediates iodine efflux into the lumen. Mutation of the *pendrin* gene causes *Pendred syndrome*, a disorder characterized by defective organification of iodine, goiter, and sensorineural deafness.

Iodine deficiency is prevalent in many mountainous regions and in central Africa, central South America, and northern Asia (Fig. 394-3). Europe remains mildly iodine-deficient, and health surveys indicate that iodine intake has been falling in the United States and Australia. The World Health Organization (WHO) estimates that about 2 billion people are iodine-deficient, based on urinary excretion data. In areas of relative iodine deficiency, there is an increased prevalence of goiter and, when deficiency is severe, hypothyroidism and cretinism. *Cretinism* is characterized by intellectual disability and growth retardation and occurs when children who live in iodine-deficient regions are not treated with iodine or thyroid hormone to restore normal thyroid hormone levels during early life. These children are often born to mothers with iodine deficiency, and it is likely that maternal thyroid hormone deficiency worsens the condition. The physiologic iodine requirement is higher in lactating women, and iodine turnover is high in infants. Concomitant selenium deficiency may also contribute to the neurologic manifestations of cretinism. Iodine supplementation of salt, bread, and other food substances has markedly reduced the prevalence of cretinism. Unfortunately, however, iodine deficiency remains a common cause of preventable intellectual disability, often because of societal resistance to food additives or the cost of supplementation. In

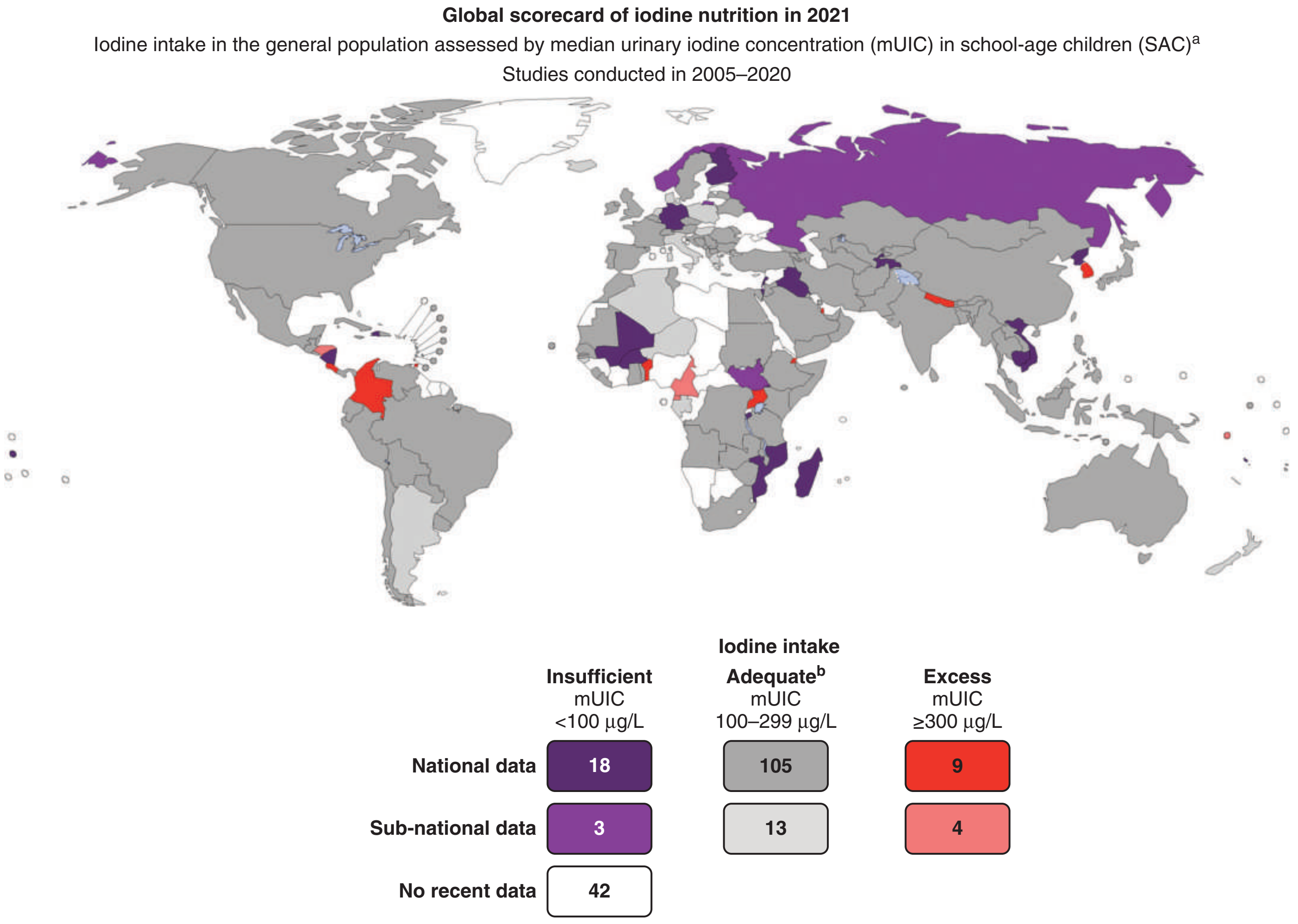
Global scorecard of iodine nutrition in 2021
Iodine intake in the general population assessed by median urinary iodine concentration (mUIC) in school-age children (SAC)[a]
Studies conducted in 2005–2020
Iodine intake

	Insufficient mUIC <100 µg/L	Adequate[b] mUIC 100–299 µg/L	Excess mUIC ≥300 µg/L
National data	18	105	9
Sub-national data	3	13	4
No recent data	42		

FIGURE 394-3 Worldwide iodine nutrition. [a]In population monitoring of iodine status using urinary iodine concentration (UIC), school-age children (SAC) serve as a proxy for the general population; therefore, preference has been given to studies carried out in SAC. The UIC data have been selected for each country in the following order of priority: data from the most recent known nationally representative survey carried out between 2005 and 2020 in (1) SAC, (2) SAC and adolescents, (3) adolescents, (4) women of reproductive age, (5) other adults (excluding pregnant or lactating women), and (6) other eligible populations. In the absence of recent national surveys, subnational data were used in the same order of priority. Subnational UIC surveys are commonly carried out to provide a rapid assessment of population iodine status, but due to a lack of sampling rigor, they may over- or underestimate the iodine status at the national level and should be interpreted with caution. [b]Adequate iodine intake in SAC corresponds to median UIC values in the range of 100–299 µg/L and includes categories previously referred to as "adequate" (100–199 µg/L) and "more than adequate" (200–299 µg/L). *(Reproduced with permission from The Iodine Global Network. Global scorecard of iodine nutrition in 2021 in the general population based on data in school-age children (SAC). IGN: Ottawa, Canada. 2021.)*

addition to overt cretinism, mild iodine deficiency can lead to subtle reduction of IQ. Oversupply of iodine, through supplements or foods enriched in iodine (e.g., shellfish, kelp), is associated with an increased incidence of autoimmune thyroid disease. The Recommended Dietary Allowance (RDA) is 220 μg iodine per day for pregnant women and 290 μg iodine per day for breastfeeding women. Because the effects of iodine deficiency are most severe in pregnant women and their babies, the American Thyroid Association has recommended that all pregnant and breastfeeding women in the United States and Canada take a prenatal multivitamin containing 150 μg iodine per day. Urinary iodine is >100 μg/L in iodine-sufficient populations.

Organification, Coupling, Storage, and Release After iodide enters the thyroid, it is trapped and transported to the apical membrane of thyroid follicular cells, where it is oxidized in an organification reaction that involves TPO and hydrogen peroxide produced by dual oxidase (DUOX) and DUOX maturation factor (DUOXA). The reactive iodine atom is added to specific tyrosyl residues within Tg, a large (660 kDa) dimeric protein that consists of 2769 amino acids. The iodotyrosines in Tg are then coupled via an ether linkage in a reaction that is also catalyzed by TPO. Either T_4 or T_3 can be produced by this reaction, depending on the number of iodine atoms present in the iodotyrosines. After coupling, Tg is taken back into the thyroid cell, where it is processed in lysosomes to release T_4 and T_3. Uncoupled mono- and diiodotyrosines (MIT, DIT) can be deiodinated by the enzyme iodotyrosine deiodinase, thereby recycling any iodide that is not converted into thyroid hormones.

Disorders of thyroid hormone synthesis are rare causes of congenital hypothyroidism **(Chap. 395)**. The vast majority of these disorders are due to recessive mutations in TPO or Tg, but defects have also been identified in the TSH-R, NIS, pendrin, hydrogen peroxide generation, and iodotyrosine deiodinase, as well as genes involved in thyroid gland development. In the case of biosynthetic defects, the gland is incapable of synthesizing adequate amounts of hormone, leading to increased TSH and a large goiter.

TSH Action TSH regulates thyroid gland function through the TSH-R, a seven-transmembrane G protein–coupled receptor (GPCR). The TSH-R is coupled to the α subunit of stimulatory G protein ($G_s\alpha$), which activates adenylyl cyclase, leading to increased production of cyclic adenosine monophosphate (cAMP). TSH also stimulates phosphatidylinositol turnover by activating phospholipase C. Recessive loss-of-function TSH-R mutations cause thyroid hypoplasia and congenital hypothyroidism. Dominant gain-of-function mutations are rare causes of sporadic or familial hyperthyroidism that is characterized by goiter, thyroid cell hyperplasia, and autonomous function **(Chap. 396)**. Most of these activating mutations occur in the transmembrane domain of the receptor. They mimic the conformational changes induced by TSH binding or the interactions of thyroid-stimulating immunoglobulins (TSIs) in Graves' disease. Activating TSH-R mutations also occur as somatic events, leading to clonal selection and expansion of the affected thyroid follicular cell and autonomously functioning thyroid nodules.

Other Factors That Influence Hormone Synthesis and Release Although TSH is the dominant hormonal regulator of thyroid gland growth and function, a variety of growth factors, most produced locally in the thyroid gland, also influence thyroid hormone synthesis. These include insulin-like growth factor 1 (IGF-1), epidermal growth factor, transforming growth factor β (TGF-β), endothelins, and various cytokines. The quantitative roles of these factors are not well understood, but they are important in selected disease states. In acromegaly, for example, increased levels of growth hormone and IGF-1 are associated with goiter and predisposition to multinodular goiter (MNG). Certain cytokines and interleukins (ILs) produced in association with autoimmune thyroid disease induce thyroid growth, whereas others lead to apoptosis. Iodine deficiency increases thyroid blood flow and upregulates the NIS, stimulating more efficient iodine uptake. Excess iodide transiently inhibits thyroid iodide organification, a phenomenon known as the *Wolff-Chaikoff effect*. In individuals with a normal thyroid, the gland escapes from this inhibitory effect and iodide organification resumes; the suppressive action of high iodide may persist, however, in patients with underlying autoimmune thyroid disease.

THYROID FUNCTION IN PREGNANCY

Five factors alter thyroid function in pregnancy: (1) the transient increase in hCG during the first trimester, which weakly stimulates the TSH-R; (2) the estrogen-induced rise in thyroxine-binding globulin (TBG) during the first trimester, which is sustained during pregnancy; (3) alterations in the immune system, leading to the onset, exacerbation, or amelioration of an underlying autoimmune thyroid disease; (4) increased thyroid hormone metabolism by the placental type III deiodinase; and (5) increased urinary iodide excretion, which can cause impaired thyroid hormone production in areas of marginal iodine sufficiency. Women with a precarious iodine intake (<50 μg/d) are most at risk of developing a goiter during pregnancy or giving birth to an infant with a goiter and hypothyroidism. The WHO recommends a daily iodine intake of 250 μg during pregnancy and lactation, and prenatal vitamins should contain 150 μg per tablet.

The rise in circulating hCG levels during the first trimester is accompanied by a reciprocal fall in TSH that persists into the middle of pregnancy. This reflects the weak binding of hCG, which is present at very high levels, to the TSH-R. Rare individuals have variant TSH-R sequences that enhance hCG binding and TSH-R activation. hCG-induced changes in thyroid function can result in transient gestational hyperthyroidism that may be associated with *hyperemesis gravidarum*, a condition characterized by severe nausea and vomiting and risk of volume depletion. However, antithyroid drugs are not indicated unless concomitant Graves' disease is suspected. Parenteral fluid replacement usually suffices until the condition resolves.

Normative values for most thyroid function tests differ during pregnancy, and if available, trimester-specific reference ranges should be used when diagnosing thyroid dysfunction during pregnancy. TSH levels decrease at the end of the first trimester and then rise as gestation progresses so that the nonpregnant reference ranges can be used from mid-gestation to delivery. Total T_4 and T_3 levels are ~1.5× higher throughout pregnancy, but the free T_4, which is the same or slightly higher at the end of the first trimester, progressively decreases so that by the third trimester, values are often below the nonpregnant lower reference cutoff.

During pregnancy, subclinical hypothyroidism occurs in 2% of women, but overt hypothyroidism is present in only 1 in 500. Prospective randomized controlled trials have not shown a benefit for universal thyroid disease screening in pregnancy. Targeted TSH testing for hypothyroidism is recommended for women planning a pregnancy if they have a strong family history of autoimmune thyroid disease, other autoimmune disorders (e.g., type 1 diabetes), infertility, prior preterm delivery or recurrent miscarriage, or signs or symptoms of thyroid disease, or are older than 30 years. Thyroid hormone requirements are increased by up to 45% during pregnancy in levothyroxine-treated hypothyroid women.

THYROID HORMONE TRANSPORT AND METABOLISM

Serum-Binding Proteins T_4 is secreted from the thyroid gland in about 15-fold excess over T_3 **(Table 394-1)**. Both hormones are bound to plasma proteins, including TBG, transthyretin (TTR, formerly known as thyroxine-binding prealbumin [TBPA]), and albumin. The plasma-binding proteins increase the pool of circulating hormone, delay hormone clearance, and may modulate hormone delivery to selected tissue sites. The concentration of TBG is relatively low (1–2 mg/dL), but because of its high affinity for thyroid hormones ($T_4 > T_3$), it carries ~80% of the bound hormones. Albumin has relatively low affinity for thyroid hormones but has a high plasma concentration (~3.5 g/dL), and it binds up to 10% of T_4 and 30% of T_3. TTR carries ~10% of T_4 but little T_3.

When the effects of the various binding proteins are combined, ~99.98% of T_4 and 99.7% of T_3 are protein-bound. Because T_3 is less

TABLE 394-1 Characteristics of Circulating T_4 and T_3

HORMONE PROPERTY	T_4	T_3
Serum concentrations		
Total hormone	8 μg/dL	0.14 μg/dL
Fraction of total hormone in the unbound form	0.02%	0.3%
Unbound (free) hormone	$21 \times 10^{-12} M$	$6 \times 10^{-12} M$
Serum half-life	7 d	2 d
Fraction directly from the thyroid	100%	20%
Production rate, including peripheral conversion	90 μg/d	32 μg/d
Intracellular hormone fraction	~20%	~70%
Relative metabolic potency	0.3	1
Receptor binding	$10^{-10} M$	$10^{-11} M$

tightly bound than T_4, the fraction of unbound T_3 is greater than unbound T_4, but there is less unbound T_3 in the circulation because it is produced in smaller amounts and cleared more rapidly than T_4. The unbound or "free" concentrations of the hormones are $\sim 2 \times 10^{-11}$ M for T_4 and $\sim 6 \times 10^{-12}$ M for T_3, which roughly correspond to the thyroid hormone receptor–binding constants for these hormones (see below). The unbound hormone is thought to be biologically available to tissues. The homeostatic mechanisms that regulate the thyroid axis are directed toward maintenance of normal concentrations of unbound hormones.

Abnormalities of Thyroid Hormone–Binding Proteins A number of inherited and acquired abnormalities affect thyroid hormone–binding proteins. X-linked TBG deficiency is associated with very low levels of total T_4 and T_3. However, because unbound hormone levels are normal, patients are euthyroid and TSH levels are normal. It is important to recognize this disorder to avoid efforts to normalize total T_4 levels, because this leads to thyrotoxicosis and is futile because of rapid hormone clearance in the absence of TBG. TBG levels are elevated by estrogen, which increases sialylation and delays TBG clearance. Consequently, in women who are pregnant or taking estrogen-containing contraceptives, elevated TBG increases total T_4 and T_3 levels; however, unbound T_4 and T_3 levels are normal. These features are part of the explanation for why women with hypothyroidism require increased amounts of L-thyroxine replacement as TBG levels are increased by pregnancy or estrogen treatment. Mutations in TBG, TTR, and albumin may increase the binding affinity for T_4 and/or T_3 and cause disorders known as *euthyroid hyperthyroxinemia* or *familial dysalbuminemic hyperthyroxinemia* (FDH) **(Table 394-2)**. These disorders result in increased total T_4 and/or T_3, but unbound hormone levels are normal. The familial nature of the disorders, and the fact that TSH levels are normal rather than suppressed, should suggest this diagnosis. Unbound hormone levels (ideally measured by dialysis) are normal in FDH. The diagnosis can be confirmed by using tests that measure the affinities of radiolabeled hormone binding to specific transport proteins or by performing DNA sequence analyses of the abnormal transport protein genes.

Certain medications, such as salicylates and salsalate, can displace thyroid hormones from circulating binding proteins. Although these drugs transiently perturb the thyroid axis by increasing free thyroid hormone levels, TSH is suppressed until a new steady state is reached, thereby restoring euthyroidism. Circulating factors associated with acute illness may also displace thyroid hormone from binding proteins **(Chap. 396)**.

Deiodinases T_4 may be thought of as a precursor for the more potent T_3. T_4 is converted to T_3 by the deiodinase enzymes (Fig. 394-1). Type I deiodinase, which is located primarily in thyroid, liver, and kidneys, has a relatively low affinity for T_4. Type II deiodinase has a higher affinity for T_4 and is found primarily in the pituitary gland, brain, brown fat, and thyroid gland. Expression of type II deiodinase allows it to regulate T_3 concentrations locally, a property that may be important in the context of levothyroxine (T_4) replacement. Type II deiodinase is also regulated by thyroid hormone; hypothyroidism induces the enzyme, resulting in enhanced $T_4 \rightarrow T_3$ conversion in tissues such as brain and pituitary. $T_4 \rightarrow T_3$ conversion is impaired by fasting, systemic illness or acute trauma, oral contrast agents, and a variety of medications (e.g., propylthiouracil, propranolol, amiodarone, glucocorticoids). Type III deiodinase inactivates T_4 and T_3 and is the most important source of reverse T_3 (rT_3), including in the sick euthyroid syndrome. This enzyme is expressed in the human placenta but is not active in healthy individuals. In the sick euthyroid syndrome, especially with hypoperfusion, the type III deiodinase is activated in muscle and liver. Massive hemangiomas and other tumors that express type III deiodinase are a rare cause of consumptive hypothyroidism.

THYROID HORMONE ACTION

Thyroid Hormone Transport Circulating thyroid hormones enter cells by passive diffusion and via specific transporters such as the monocarboxylate 8 transporter (MCT8), MCT10, and organic anion-transporting polypeptide 1C1. Mutations in the *MCT8* gene have been identified in patients with X-linked psychomotor retardation and thyroid function abnormalities (low T_4, high T_3, and high TSH). After entering cells, thyroid hormones act primarily through nuclear receptors, although they also have nongenomic actions through stimulating mitochondrial enzymatic responses and may act directly on blood vessels and the heart through integrin receptors.

Nuclear Thyroid Hormone Receptors Thyroid hormones bind with high affinity to nuclear TRs α and β. Both TRα and TRβ are expressed in most tissues, but their relative expression levels vary among organs; TRα is particularly abundant in brain, kidneys, gonads, muscle, and heart, whereas TRβ expression is relatively high in the pituitary and liver. Both receptors are variably spliced to form unique isoforms. The TRβ2 isoform, which has a unique amino terminus, is selectively expressed in the hypothalamus and pituitary, where it plays a role in feedback control of the thyroid axis (see above). The TRα2

TABLE 394-2 Conditions Associated with Euthyroid Hyperthyroxinemia

DISORDER	CAUSE	TRANSMISSION	CHARACTERISTICS
Familial dysalbuminemic hyperthyroxinemia (FDH)	Albumin mutations, usually R218H	AD	Increased T_4 Normal unbound T_4 Rarely increased T_3
TBG			
Familial excess	Increased TBG production	XL	Increased total T_4, T_3 Normal unbound T_4, T_3
Acquired excess	Medications (estrogen), pregnancy, cirrhosis, hepatitis	Acquired	Increased total T_4, T_3 Normal unbound T_4, T_3
Transthyretin[a]			
Excess	Islet tumors	Acquired	Usually normal T_4, T_3
Mutations	Increased affinity for T_4 or T_3	AD	Increased total T_4, T_3 Normal unbound T_4, T_3
Medications: propranolol, ipodate, iopanoic acid, amiodarone	Decreased $T_4 \rightarrow T_3$ conversion	Acquired	Increased T_4 Decreased T_3 Normal or increased TSH
Resistance to thyroid hormone (RTH)	Thyroid hormone receptor β mutations	AD	Increased unbound T_4, T_3 Normal or increased TSH Some patients clinically thyrotoxic

[a] Also known as thyroxine-binding prealbumin (TBPA).

Abbreviations: AD, autosomal dominant; TBG, thyroxine-binding globulin; TSH, thyroid-stimulating hormone; XL, X-linked.

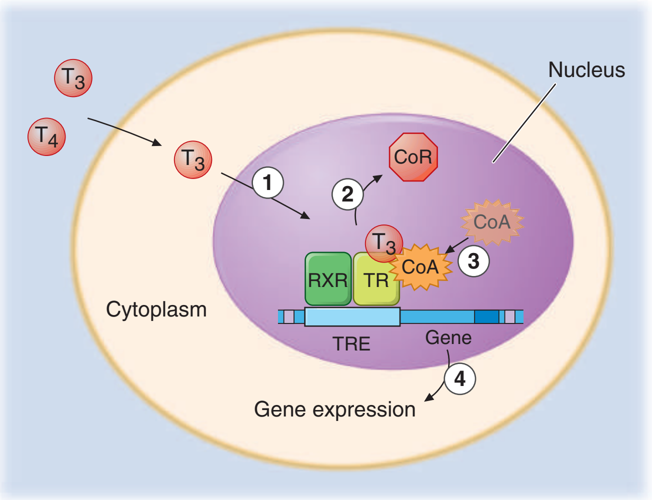

FIGURE 394-4 Mechanism of thyroid hormone receptor action. The thyroid hormone receptor (TR) and retinoid X receptor (RXR) form heterodimers that bind specifically to thyroid hormone response elements (TRE) in the promoter regions of target genes. In the absence of hormone, TR binds co-repressor (CoR) proteins that silence gene expression. The numbers refer to a series of ordered reactions that occur in response to thyroid hormone: (1) T_4 or T_3 enters the nucleus; (2) T_3 binding dissociates CoR from TR; (3) co-activators (CoA) are recruited to the T_3-bound receptor; and (4) gene expression is altered.

isoform contains a unique carboxy terminus that precludes thyroid hormone binding.

The TRs contain a central DNA-binding domain and a C-terminal ligand-binding domain. They bind to specific DNA sequences, termed *thyroid response elements* (TREs), in target genes **(Fig. 394-4)**. The receptors bind as homodimers or, more commonly, as heterodimers with retinoic acid X receptors (RXRs) **(Chap. 389)**. The activated receptor can either stimulate gene transcription (e.g., myosin heavy chain α) or inhibit transcription (e.g., TSH β-subunit gene), depending on the nature of the regulatory elements in the target gene.

Thyroid hormones (T_3 and T_4) bind with similar affinities to TRα and TRβ. However, structural differences in the ligand-binding domains provide the potential for developing receptor-selective agonists or antagonists, and these are under investigation. T_3 is bound with 10–15 times greater affinity than T_4, which explains its increased potency. Although T_4 is produced in excess of T_3, receptors are occupied mainly by T_3, reflecting $T_4 \rightarrow T_3$ conversion by peripheral tissues, T_3 bioavailability in the plasma, and the greater affinity of receptors for T_3. After binding to TRs, thyroid hormone induces conformational changes in the receptors that modify its interactions with accessory transcription factors. Importantly, in the absence of thyroid hormone binding, the aporeceptors bind to co-repressor proteins that inhibit gene transcription. Hormone binding dissociates the co-repressors and allows the recruitment of co-activators that enhance transcription. The discovery of TR interactions with co-repressors explains the fact that TR silences gene expression in the absence of hormone binding. Consequently, hormone deficiency has a profound effect on gene expression because it causes gene repression as well as loss of hormone-induced stimulation. This concept has been corroborated by the finding that targeted deletion of the TR genes in mice has a less pronounced phenotypic effect than hormone deficiency.

Thyroid Hormone Resistance Resistance to thyroid hormone (RTH) is an autosomal dominant disorder characterized by elevated thyroid hormone levels and inappropriately normal or elevated TSH. Individuals with RTH do not, in general, exhibit signs and symptoms that are typical of hypothyroidism because hormone resistance is partial and is compensated by increased levels of thyroid hormone. The clinical features of RTH can include goiter, attention deficit disorder, mild reduction in IQ, delayed skeletal maturation, tachycardia, and impaired metabolic responses to thyroid hormone.

Classical forms of RTH are caused by mutations in the TRβ gene. These mutations, located in restricted regions of the ligand-binding domain, cause loss of receptor function. However, because the mutant receptors retain the capacity to dimerize with RXRs, bind to DNA, and recruit co-repressor proteins, they function as antagonists of the remaining normal TRβ and TRα receptors. This property, referred to as "dominant negative" activity, explains the autosomal dominant mode of transmission. The diagnosis is suspected when unbound thyroid hormone levels are increased without suppression of TSH. Similar hormonal abnormalities are found in other affected family members, although the TRβ mutation arises de novo in ~20% of patients. DNA sequence analysis of the TRβ gene provides a definitive diagnosis. RTH must be distinguished from other causes of euthyroid hyperthyroxinemia (e.g., FDH) and inappropriate secretion of TSH by TSH-secreting pituitary adenomas **(Chap. 392)**. In most patients, no treatment is indicated; the importance of making the diagnosis is to avoid inappropriate treatment of mistaken hyperthyroidism and to provide genetic counseling.

A distinct form of RTH is caused by mutations in the TRα gene. Affected patients have many clinical features of congenital hypothyroidism including growth retardation, skeletal dysplasia, severe constipation, and delayed neurocognitive development. In contrast to RTH caused by mutations in TRβ, thyroid function tests include normal TSH, low or normal T_4, and normal or elevated T_3 levels. These distinct clinical and laboratory features underscore the different tissue distribution and functional roles of TRβ and TRα. Thyroxine treatment appears to alleviate some of the clinical manifestations of patients with RTH caused by TRα mutations.

PHYSICAL EXAMINATION

In addition to the examination of the thyroid itself, the physical examination should include a search for signs of abnormal thyroid function and the extrathyroidal features of ophthalmopathy and dermopathy **(Chap. 396)**. Examination of the neck begins by inspecting the seated patient from the front and side and noting any surgical scars, obvious masses, or distended veins. The thyroid can be palpated with both hands from behind or while facing the patient, using the thumbs to palpate each lobe. It is best to use a combination of these methods, especially when nodules are small. The patient's neck should be slightly flexed to relax the neck muscles. After locating the cricoid cartilage, the isthmus, which is attached to the lower one-third of the thyroid lobes, can be identified and then followed laterally to locate either lobe (normally, the right lobe is slightly larger than the left). By asking the patient to swallow sips of water, thyroid consistency can be better appreciated as the gland moves beneath the examiner's fingers.

Features to be noted include thyroid size, consistency, nodularity, and any tenderness or fixation. An estimate of thyroid size (normally 12–20 g) should be made, and a drawing is often the best way to record findings. Ultrasound imaging provides the most accurate measurement of thyroid volume and nodularity and is useful for assessment of goiter prevalence in iodine-deficient regions. However, ultrasound is not indicated if the thyroid physical examination is normal. The size, location, and consistency of any nodules should also be defined. A bruit or thrill over the gland, located over the insertion of the superior and inferior thyroid arteries (supero- or inferolaterally), indicates increased vascularity, associated with turbulent rather than laminar blood flow, as occurs in hyperthyroidism. If the lower borders of the thyroid lobes are not clearly felt, a goiter may be retrosternal. Large retrosternal goiters can cause venous distention over the neck and difficulty breathing, especially when the arms are raised (Pemberton's sign). With any central mass above the thyroid, the tongue should be extended, as thyroglossal cysts then move upward. The thyroid examination is not complete without assessment for lymphadenopathy in the supraclavicular and cervical regions of the neck.

LABORATORY EVALUATION

Measurement of Thyroid Hormones The enhanced sensitivity and specificity of *TSH assays* have greatly improved laboratory assessment of thyroid function. Because TSH levels change dynamically in response to alterations of T_4 and T_3, a logical approach to thyroid testing is to first determine whether TSH is suppressed, normal, or elevated. With rare exceptions (see below), a normal TSH level excludes a primary abnormality of thyroid function. This strategy depends on the use of assays for TSH that are sensitive enough to discriminate between the lower limit of the reference interval and the suppressed values that occur with thyrotoxicosis. Extremely sensitive assays can detect TSH levels ≤0.004 mIU/L, but, for practical purposes, assays sensitive to ≤0.01 mIU/L are sufficient. The widespread availability of sensitive TSH assays has rendered the TRH stimulation test obsolete, because the failure of TSH to rise after an intravenous bolus of 200–400 μg TRH has the same implications as a suppressed basal TSH. Because the antibodies used in many TSH assays are biotinylated, biotin supplements containing 1000 μg or more, including biotin in some multivitamins, can interfere with TSH measurements, resulting in falsely low TSH values and falsely high T_4 or T_3 levels. Therefore, patients should be advised to stop taking biotin for at least 2 days prior to thyroid function testing.

The finding of an abnormal TSH level must be followed by measurements of circulating thyroid hormone levels to confirm the diagnosis of hyperthyroidism (suppressed TSH) or hypothyroidism (elevated TSH). Automated immunoassays are widely available for serum *total* T_4 and *total* T_3. T_4 and T_3 are highly protein-bound, and numerous factors (illness, medications, genetic factors) can influence protein binding. It is useful, therefore, to measure the free, or unbound, hormone levels, which correspond to the biologically available hormone pool. Two direct methods are used to measure *unbound thyroid hormones*: (1) unbound thyroid hormone competition with radiolabeled T_4 (or an analogue) for binding to a solid-phase antibody, and (2) physical separation of the unbound hormone fraction by ultracentrifugation or equilibrium dialysis. Although early unbound hormone immunoassays suffered from artifacts, newer assays correlate well with the results of the more technically demanding and expensive physical separation methods. An indirect method that is now less commonly used to estimate unbound thyroid hormone levels is to calculate the free T_3 or free T_4 index from the total T_4 or T_3 concentration and the *thyroid hormone binding ratio* (THBR). The latter is derived from the *T_3-resin uptake test*, which determines the distribution of radiolabeled T_3 between an absorbent resin and the unoccupied thyroid hormone–binding proteins in the sample. The binding of the labeled T_3 to the resin is increased when there is reduced unoccupied protein binding sites (e.g., TBG deficiency) or increased total thyroid hormone in the sample; it is decreased under the opposite circumstances. The product of THBR and total T_3 or T_4 provides the *free T_3 or T_4 index*. In effect, the index corrects for anomalous total hormone values caused by variations in hormone-protein binding.

Total thyroid hormone levels are *elevated* when TBG is increased due to estrogens (pregnancy, oral contraceptives, hormone therapy, tamoxifen, selective estrogen receptor modulators, inflammatory liver disease) and *decreased* when TBG binding is reduced (androgens, nephrotic syndrome). Genetic disorders and acute illness can also cause abnormalities in thyroid hormone–binding proteins, and various drugs (phenytoin, carbamazepine, salicylates, and nonsteroidal anti-inflammatory drugs [NSAIDs]) can interfere with thyroid hormone binding. Because unbound thyroid hormone levels are normal and the patient is euthyroid in all of these circumstances, assays that measure unbound hormone are preferable to those for total thyroid hormones.

For most purposes, the unbound T_4 level is sufficient to confirm thyrotoxicosis, but 2–5% of patients have only an elevated T_3 level (T_3 toxicosis). Thus, unbound T_3 levels should be measured in patients with a suppressed TSH but normal unbound T_4 levels.

There are several clinical conditions in which the use of TSH as a screening test may be misleading, particularly without simultaneous unbound T_4 determinations. Any severe nonthyroidal illness can cause abnormal TSH levels. Although hypothyroidism is the most common cause of an elevated TSH level, rare causes include a TSH-secreting pituitary tumor **(Chap. 392)**, thyroid hormone resistance, and assay artifact. Conversely, a suppressed TSH level, particularly <0.01 mIU/L, usually indicates thyrotoxicosis. However, subnormal TSH levels between 0.01 and 0.1 mIU/L may be seen during the first trimester of pregnancy (due to hCG secretion), after treatment of hyperthyroidism (because TSH can remain suppressed for several months), and in response to certain medications (e.g., high doses of glucocorticoids or dopamine). As above, biotin supplements, generally those containing more than 1000 ug, taken <18 h prior to a blood draw can interfere with biotinylated TSH antibodies and the subsequent streptavidin capture. Importantly, secondary hypothyroidism, caused by hypothalamic-pituitary disease, is associated with a variable (low to high-normal) TSH level, which is inappropriate for the low T_4 level. Thus, *TSH should not be used as an isolated laboratory test to assess thyroid function in patients with suspected or known hypothalamic or pituitary disease*.

Tests for the end-organ effects of thyroid hormone excess or depletion, such as estimation of basal metabolic rate, tendon reflex relaxation rates, or serum cholesterol, are relatively insensitive and are not useful as clinical determinants of thyroid function.

Tests to Determine the Etiology of Thyroid Dysfunction

Autoimmune thyroid disease is detected most easily by measuring circulating antibodies against TPO and Tg. Because antibodies to Tg alone are less common, it is reasonable to measure only TPO antibodies. About 5–15% of euthyroid women and up to 2% of euthyroid men have thyroid antibodies; such individuals are at increased risk of developing thyroid dysfunction. Almost all patients with autoimmune hypothyroidism, and up to 80% of those with Graves' disease, have TPO antibodies, usually at high levels.

TSIs are antibodies that stimulate the TSH-R in Graves' disease. They are most commonly measured by commercially available tracer displacement assays called TRAb (TSH receptor antibody) with the assumption that elevated levels in the setting of clinical hyperthyroidism reflect stimulatory effects on the TSH receptor. A bioassay is less commonly used. Remission rates in patients with Graves' disease after antithyroid drug cessation are higher in patients with disappearance TRAb. Furthermore, the TRAb assay is used to predict both fetal and neonatal thyrotoxicosis caused by transplacental passage of high maternal levels of TRAb or TSI (>3× upper limit of normal) in the last trimester of pregnancy.

Serum Tg levels are increased in all types of thyrotoxicosis except *thyrotoxicosis factitia* caused by self-administration of thyroid hormone. Tg levels are particularly increased in thyroiditis, reflecting thyroid tissue destruction and release of Tg. The main role for Tg measurement, however, is in the follow-up of thyroid cancer patients. After total thyroidectomy and radioablation for patients with thyroid cancer, Tg levels should be <0.2 ng/mL in the absence of anti-Tg antibodies; measurable levels indicate incomplete ablation or recurrent cancer.

Radioiodine Uptake and Thyroid Scanning The thyroid gland selectively transports and organifies radioisotopes of iodine (^{123}I, ^{125}I, ^{131}I) and transports ^{99m}Tc pertechnetate, allowing thyroid imaging for all these isotopes, but quantitation of radioactive tracer fractional uptake for iodine isotopes only.

Nuclear imaging of Graves' disease is characterized by an enlarged gland and increased tracer uptake that is distributed homogeneously. Toxic adenomas appear as focal areas of increased uptake, with suppressed tracer uptake in the remainder of the gland (reflecting suppressed TSH). In toxic MNG, the gland is enlarged—often with distorted architecture—and there are multiple areas of relatively increased (functioning nodules) or decreased tracer uptake (suppressed thyroid parenchyma or nonfunctioning nodules). Subacute, viral, and postpartum thyroiditis are associated with very low uptake because of follicular cell damage and TSH suppression. Thyrotoxicosis factitia is also associated with low uptake because exogenous hormone suppresses TSH. In addition, excessive circulating exogenous iodine (e.g., from dietary sources of iodinated contrast dye) reduces radionuclide uptake even in the presence of increased thyroid hormone production.

Thyroid scintigraphy is not used in the routine evaluation of patients with thyroid nodules but should be performed if the serum TSH level is subnormal to determine if functioning thyroid nodules are present. Functioning or "hot" nodules are almost never malignant, and fine-needle aspiration (FNA) biopsy is not indicated **(see Chap. 397, Fig. 397-3*B*)**. The vast majority of thyroid nodules do not produce thyroid hormone ("cold" nodules, see Chap. 397, Fig 397-3*A*), and these are more likely to be malignant (~5–10%). Whole-body and thyroid scanning is also used in the treatment and, now less frequently, in the surveillance of thyroid cancer. After thyroidectomy for thyroid cancer, the TSH level is raised by either using a thyroid hormone withdrawal protocol or recombinant human TSH injection (Chap. 397). Administration of either ^{131}I or ^{123}I (in higher activities than used to image the thyroid gland alone) allows whole-body scanning (WBS) to detect the thyroid remnant. WBS imaging is also performed after therapeutic administration of ^{131}I, which confirms remnant ablation and may reveal iodine-avid metastases.

Thyroid Ultrasound Ultrasonography is the most valuable tool for the diagnosis and evaluation of patients with nodular thyroid disease (Chap. 397). Evidence-based guidelines recommend thyroid ultrasonography for all patients suspected of having thyroid nodules by either physical examination or another imaging study. Using 10- to 12-MHz linear transducers, resolution and image quality are excellent, allowing the characterization of nodules and cysts >3 mm. Sonographic patterns that combine suspicious sonographic features are highly suggestive of malignancy (e.g., hypoechoic solid nodules with irregular borders and punctate echogenic foci >90% cancer risk), whereas other patterns correlate with a lower likelihood of cancer (isoechoic solid nodules, 5–10% cancer risk). Some patterns suggest benignity (e.g., spongiform nodules, defined as those with multiple small internal cystic areas, or simple cysts, <3% cancer risk) (see Chap. 397, Fig. 397-2). These patterns have been incorporated into validated risk stratification systems (RSSs) for sonographic imaging of thyroid nodules (American College of Radiology [ACR] Thyroid Imaging Reporting and Data System [TI-RADS], American Thyroid Association, European Thyroid Association [EU-TIRADS] and others) (see Chap. 397, Fig. 397-1). These systems are relatively concordant in the classification of thyroid nodules; they differ in size cutoff recommendations for FNA. Not surprisingly, the RSSs with lower size cutoffs have higher sensitivity and lower specificity for thyroid cancer diagnosis than those with higher cutoffs. Nevertheless, all have been shown to reduce unnecessary FNAs by at least 45%, in part due to the recommendation not to perform FNA of spongiform nodules.

In addition to evaluating thyroid nodules, ultrasound is useful for monitoring nodule size and for the aspiration of nodules or cystic lesions. Ultrasound-guided FNA biopsy of thyroid lesions lowers the rate of inadequate sampling and decreases sample error, thereby reducing both the nondiagnostic and false-negative rates of FNA cytology. Ultrasonography of the central and lateral cervical lymph node compartments is indispensable in the evaluation of thyroid cancer patients, preoperatively and during follow-up. In addition, the all of the RSSs recommend a survey of the cervical lymph nodes as part of every diagnostic thyroid sonographic examination.

FURTHER READING

Alexander EK et al: 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. Thyroid 27:315, 2017.

Andersson M, Braegger CP: The role of iodine for thyroid function in lactating women and infants. Endocr Rev 43:469, 2022.

Carvalho DP, Dupuy C: Thyroid hormone biosynthesis and release. Mol Cell Endocrinol 458:S0303, 2017.

Groeneweg S et al: Thyroid hormone transporters. Endocr Rev 41:146, 2020.

Lee SY, Pearce EN: Testing, monitoring, and treatment of thyroid dysfunction in pregnancy. J Clin Endocrinol Metab 106:883, 2021.

Mio C et al: Molecular defects in thyroid dysgenesis. Clin Genet 97:222, 2020.

Moran C et al: Genetic disorders of thyroid development, hormone biosynthesis and signalling. Clin Endocrinol 97:502, 2022.

Pappa T, Refetoff S: Resistance to thyroid hormone beta: a focused review. Front Endocrinol 12:1, 2021.

Rugge JB et al: Screening and treatment of thyroid dysfunction: An evidence review for the U.S. Preventive Services Task Force. Ann Intern Med 162:35, 2015.

395 Hypothyroidism

Anthony P. Weetman, Susan J. Mandel,
J. Larry Jameson

HYPOTHYROIDISM

Iodine deficiency remains a common cause of hypothyroidism worldwide. In areas of iodine sufficiency, autoimmune disease (Hashimoto's thyroiditis) and iatrogenic causes (treatment of hyperthyroidism) are most common (Table 395-1).

CONGENITAL HYPOTHYROIDISM

Prevalence Hypothyroidism occurs in about 1 in 2000–4000 newborns, and neonatal screening is performed in most industrialized countries. It may be transient, especially if the mother has thyroid-stimulating hormone (TSH) receptor (TSH-R)–blocking antibodies or has received antithyroid drugs, but permanent hypothyroidism occurs in the majority. The causes of neonatal hypothyroidism include thyroid gland dysgenesis in 65%, inborn errors of thyroid hormone synthesis in 30%, and TSH-R antibody mediated in 5% of affected newborns. The developmental abnormalities are twice as common in girls. Mutations that cause congenital hypothyroidism are being increasingly identified, but most remain idiopathic (Chap. 394). These can be broadly categorized as mutations causing (1) central hypothyroidism because of abnormal hypothalamic-pituitary development or the loss of specific components of the thyrotropin-releasing hormone (TRH)/TSH hormonal pathways; (2) abnormal thyroid gland development or dysgenesis; or (3) abnormal thyroid hormone synthesis and processing, or dyshormonogenesis (Table 395-2). Transplacental passage of maternal thyroid hormone occurs before the fetal thyroid gland begins to function and provides partial hormone support to a fetus with congenital hypothyroidism.

TABLE 395-1 Causes of Hypothyroidism

Primary
Autoimmune hypothyroidism: Hashimoto's thyroiditis, atrophic thyroiditis
Iatrogenic: ^{131}I treatment, subtotal or total thyroidectomy, external irradiation of neck for lymphoma or cancer
Drugs: iodine excess (including iodine-containing contrast media), amiodarone, lithium, antithyroid drugs, *p*-aminosalicylic acid, interferon α and other cytokines, aminoglutethimide, tyrosine kinase inhibitors (e.g., sunitinib), immune checkpoint inhibitors (e.g., ipilimumab, nivolumab, pembrolizumab)
Congenital hypothyroidism: absent or ectopic thyroid gland, dyshormonogenesis, TSH-R mutation
Iodine deficiency
Infiltrative disorders: amyloidosis, sarcoidosis, hemochromatosis, scleroderma, cystinosis, Riedel's thyroiditis
Overexpression of type 3 deiodinase in infantile hemangioma and other tumors
Transient
Silent thyroiditis, including postpartum thyroiditis
Subacute thyroiditis
Withdrawal of supraphysiologic thyroxine treatment in individuals with an intact thyroid
After ^{131}I treatment or subtotal thyroidectomy for Graves' disease
Secondary
Hypopituitarism: tumors, pituitary surgery or irradiation, infiltrative disorders, Sheehan's syndrome, trauma, genetic forms of combined pituitary hormone deficiencies
Isolated TSH deficiency or inactivity
Drugs: bexarotene, mitotane
Hypothalamic disease: tumors, trauma, infiltrative disorders, Prader-Willi syndrome

Abbreviations: TSH, thyroid-stimulating hormone; TSH-R, TSH receptor.

TABLE 395-2 Examples of Genetic Causes of Congenital Hypothyroidism

DEFECTIVE GENE PROTEIN	TYPE OF HYPOTHYROIDISM	INHERITANCE	CONSEQUENCES
PROP-1	Central, hypothyroidism	Homozygous recessive	Combined pituitary hormone deficiencies, including thyroid-stimulating hormone (TSH), with preservation of adrenocorticotropic hormone
PIT-1	Central, hypothyroidism	Homozygous or heterozygous loss of function	Combined deficiencies of growth hormone, prolactin, TSH
IGSF1	Central, hypothyroidism	X-linked loss of function	Loss of TSH receptor (TSH-R) expression, testicular enlargement
TSHβ	Central, hypothyroidism	Heterozygous loss of function	TSH deficiency
TTF-1 (TITF-1)	Primary, thyroid dysgenesis	Heterozygous loss of function	Variable thyroid hypoplasia, choreoathetosis, pulmonary problems
TTF-2 (FOXE-1)	Primary, thyroid dysgenesis	Homozygous recessive	Thyroid agenesis, choanal atresia, spiky hair
PAX-8	Primary, thyroid dysgenesis	Heterozygous loss of function	Thyroid dysgenesis, kidney abnormalities
NKX2-1	Primary, thyroid dysgenesis	Heterozygous loss of function	Thyroid dysgenesis, brain, lung abnormalities
NKX2-5	Primary, thyroid dysgenesis	Heterozygous loss of function	Thyroid dysgenesis, heart abnormalities
GLIS3	Primary, thyroid dysgenesis	Homozygous recessive	Thyroid dysgenesis, neonatal diabetes, facial abnormalities
JAG-1	Primary, thyroid dysgenesis	Heterozygous loss of function	Thyroid dysgenesis, Alagille syndrome type 1, heart abnormalities
TSH receptor	Primary, thyroid dysgenesis and dyshormonogenesis	Homozygous recessive	Resistance to TSH
$G_s\alpha$ (Albright hereditary osteodystrophy)	Primary, thyroid dyshormonogenesis	Heterozygous loss of function, imprinting	Resistance to TSH
Na^+/I^- symporter (SLC5A5)	Primary, thyroid dyshormonogenesis	Homozygous recessive	Inability to transport iodide
DUOX2 (THOX2)	Primary, thyroid dyshormonogenesis	Heterozygous loss of function	Organification defect
DUOXA2	Primary, thyroid dyshormonogenesis	Homozygous recessive	Organification defect
Thyroid peroxidase	Primary, thyroid dyshormonogenesis	Homozygous recessive	Defective organification of iodide
Thyroglobulin	Primary, thyroid dyshormonogenesis	Homozygous recessive	Defective synthesis of thyroid hormone
Pendrin (SLC26A4)	Primary, thyroid dyshormonogenesis	Homozygous recessive	Pendred syndrome: sensorineural deafness and partial organification defect in thyroid
Dehalogenase 1 (IYD)	Primary, thyroid dyshormonogenesis	Homozygous recessive	Loss of iodide reutilization

Clinical Manifestations The majority of infants appear normal at birth, and with the use of biochemical screening, few cases are now diagnosed based on clinical features, which include prolonged jaundice, feeding problems, hypotonia, enlarged tongue, delayed bone maturation, and umbilical hernia. Importantly, permanent neurologic damage results if treatment is delayed. Typical features of adult hypothyroidism may also be present (Table 395-3). Other congenital malformations, especially cardiac, are four times more common in congenital hypothyroidism.

Diagnosis and Treatment Because of the severe neurologic consequences of untreated congenital hypothyroidism, neonatal screening programs have been established. These are generally based on measurement of TSH or T_4 levels in heel-prick blood specimens. When the diagnosis is confirmed, T_4 is instituted at a dose of 10–15 μg/kg per d, and the dose is adjusted by close monitoring of TSH levels. T_4 requirements are relatively great during the first year of life, and a high circulating T_4 level is usually needed to normalize TSH. Early treatment with T_4 results in normal IQ levels, but subtle neurodevelopmental abnormalities may occur in those with the most severe hypothyroidism at diagnosis or when treatment is delayed or suboptimal. If transient hypothyroidism is suspected, or the diagnosis is unclear, treatment can be stopped safely after the age of 3 years followed by further evaluation.

TABLE 395-3 Signs and Symptoms of Hypothyroidism (Descending Order of Frequency)

SYMPTOMS	SIGNS
Tiredness, weakness	Dry coarse skin; cool peripheral extremities
Dry skin	Puffy face, hands, and feet (myxedema)
Feeling cold	Diffuse alopecia
Hair loss	Bradycardia
Difficulty concentrating and poor memory	Peripheral edema
Constipation	Delayed tendon reflex relaxation
Weight gain with poor appetite	Carpal tunnel syndrome
Dyspnea	Serous cavity effusions
Hoarse voice	
Menorrhagia (later oligomenorrhea or amenorrhea)	
Paresthesia	
Impaired hearing	

AUTOIMMUNE HYPOTHYROIDISM

Classification Autoimmune hypothyroidism (Hashimoto's thyroiditis) may be associated with a goiter (*goitrous thyroiditis*) or minimal residual thyroid tissue (*atrophic thyroiditis*). Because the autoimmune process gradually reduces thyroid function, there is a phase of compensation when normal thyroid hormone levels are maintained by a rise in TSH. Although some patients may have minor symptoms, this state is called *subclinical hypothyroidism*. Later, unbound T_4 levels fall and TSH levels rise further; symptoms become more readily apparent at this stage (usually TSH >10 mIU/L), which is referred to as *clinical hypothyroidism* or *overt hypothyroidism*.

Prevalence The mean annual incidence rate of autoimmune hypothyroidism is up to 4 per 1000 women and 1 per 1000 men. It is more common in certain populations, such as the Japanese, probably because of genetic factors and chronic exposure to a high-iodine diet. Hypothyroidism typically occurs between 30 and 50 years of age, and the prevalence of increases with age. The mean age at diagnosis is 60 years, and the prevalence of overt hypothyroidism increases with age. Subclinical hypothyroidism is found in 6–8% of women (10% over the age of 60) and 3% of men. The annual risk of developing clinical hypothyroidism is ~4% when subclinical hypothyroidism is associated with positive thyroid peroxidase (TPO) antibodies.

Pathogenesis In Hashimoto's thyroiditis, there is a marked lymphocytic infiltration of the thyroid with germinal center formation, atrophy of the thyroid follicles accompanied by oxyphil metaplasia, absence of colloid, and mild to moderate fibrosis. In atrophic thyroiditis, the fibrosis is much more extensive, lymphocyte infiltration is less pronounced, and thyroid follicles are almost completely absent. Atrophic thyroiditis usually represents the end stage of Hashimoto's thyroiditis rather than a separate disorder, although a distinct form of marked fibrosis occurs in which the gland is infiltrated with IgG4-positive plasma cells.

As with most autoimmune disorders, susceptibility to autoimmune hypothyroidism is determined by a combination of genetic and environmental factors, and the risk of either autoimmune hypothyroidism or Graves' disease is increased among siblings. HLA-DR polymorphisms are the best documented genetic risk factors for autoimmune hypothyroidism, especially HLA-DR3, DR4, and DR5 in Caucasians. A weak association also exists between polymorphisms in *PTPN22* and *CTLA-4*, which have immunoregulatory functions, and autoimmune hypothyroidism. All these genetic associations are shared by other autoimmune diseases, which may explain the relationship between autoimmune hypothyroidism and other autoimmune diseases, especially type 1 diabetes mellitus, Addison's disease, pernicious anemia, and vitiligo. The role of other contributory loci remains to be clarified. A gene on chromosome 21 may be responsible for the association between autoimmune hypothyroidism and Down's syndrome. The female preponderance of thyroid autoimmunity is most likely due to sex steroid effects on the immune response, but an X chromosome–related genetic factor is also possible and may account for the high frequency of autoimmune hypothyroidism in Turner's syndrome. Environmental susceptibility factors are poorly defined at present. A high iodine or low selenium intake and decreased exposure to microorganisms in childhood increase the risk of autoimmune hypothyroidism. Smoking cessation transiently increases incidence, whereas alcohol intake seems protective. These factors may account for the increase in prevalence over the past two to three decades.

The thyroid lymphocytic infiltrate in autoimmune hypothyroidism is composed of activated T cells as well as B cells. Thyroid cell destruction is primarily mediated by the CD8+ cytotoxic T cells, but local production of cytokines, such as tumor necrosis factor (TNF), interleukin-1 (IL-1), and interferon γ (IFN-γ), derived from the inflammatory infiltrate may render thyroid cells more susceptible to apoptosis mediated by death receptors, such as Fas, and by oxidative stress. These cytokines also impair thyroid cell function directly and induce the expression of other proinflammatory molecules by the thyroid cells themselves, such as cytokines, human leukocyte antigen (HLA) class I and class II molecules, adhesion molecules, CD40, and nitric oxide. Administration of high concentrations of cytokines for therapeutic purposes (especially IFN-α) is associated with increased autoimmune thyroid disease, possibly through mechanisms similar to those in sporadic disease. Novel anticancer and immunomodulatory treatments, such as tyrosine kinase inhibitors, immune checkpoint inhibitors, and alemtuzumab, can also induce thyroiditis via their effects on T-cell regulation.

Antibodies to TPO and thyroglobulin (Tg) are clinically useful markers of thyroid autoimmunity, but any pathogenic effect is restricted to a secondary role in amplifying an ongoing autoimmune response. TPO antibodies fix complement, and complement membrane-attack complexes are present in the thyroid in autoimmune hypothyroidism. However, transplacental passage of Tg or TPO antibodies has no effect on the fetal thyroid, which suggests that T cell–mediated injury is required to initiate autoimmune damage to the thyroid.

Up to 20% of patients with autoimmune hypothyroidism have antibodies against the TSH-R, which, in contrast to thyroid-stimulating immunoglobulin (TSI), do not stimulate the receptor but prevent the binding of TSH. These TSH-R-blocking antibodies, therefore, cause hypothyroidism and, especially in Asian patients, thyroid atrophy. Their transplacental passage may induce transient neonatal hypothyroidism. Rarely, patients have a mixture of TSI and TSH-R-blocking antibodies, and thyroid function can oscillate between hyperthyroidism and hypothyroidism as one or the other antibody becomes dominant. Predicting the course of disease in such individuals is difficult, and they require close monitoring of thyroid function. Bioassays can be used to document that TSH-R-blocking antibodies reduce the cyclic AMP–inducing effect of TSH on cultured TSH-R-expressing cells, but these assays are difficult to perform. Immunoassays for TSH receptor antibodies, which measure whether the patient's serum contains an antibody that can displace either labeled TSH or a monoclonal TSH receptor antibody from the TSH receptor, do not distinguish between these types of functional antibodies, but a positive result in a patient with spontaneous hypothyroidism is strong evidence for the presence of blocking antibodies. The use of these assays does not generally alter clinical management, although it may be useful to confirm the cause of transient neonatal hypothyroidism.

Clinical Manifestations The main clinical features of hypothyroidism are summarized in Table 395-3. The onset is usually insidious, and the patient may become aware of symptoms only when euthyroidism is restored. Patients with Hashimoto's thyroiditis may present because of goiter rather than symptoms of hypothyroidism. The goiter may not be large, but it is usually irregular and firm in consistency. Rarely, uncomplicated Hashimoto's thyroiditis is associated with pain.

Patients with atrophic thyroiditis or the later stage of Hashimoto's thyroiditis present with symptoms and signs of hypothyroidism. The skin is dry, and there is decreased sweating, thinning of the epidermis, and hyperkeratosis of the stratum corneum. Increased dermal glycosaminoglycan content traps water, giving rise to skin thickening without pitting (*myxedema*). Typical features include a puffy face with edematous eyelids and nonpitting pretibial edema **(Fig. 395-1)**. There is pallor, often with a yellow tinge to the skin due to carotene accumulation. Nail growth is retarded, and hair is dry, brittle, difficult to manage, and falls out easily. In addition to diffuse alopecia, there is thinning of the outer third of the eyebrows, although this is not a specific sign of hypothyroidism.

Other common features include constipation and weight gain (despite a poor appetite). In contrast to popular perception, the weight gain is usually modest and due mainly to fluid retention in the myxedematous tissues. Libido is decreased in both sexes, and there may be oligomenorrhea or amenorrhea in long-standing disease, but

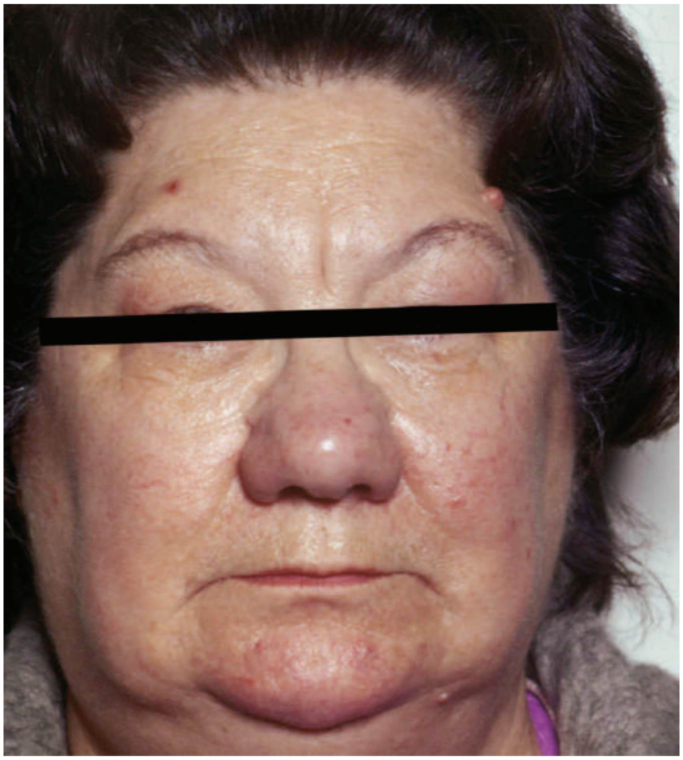
[redacted]

FIGURE 395-1 Facial appearance in hypothyroidism. Note puffy eyes and thickened skin.

menorrhagia may occur at an early stage. Fertility is reduced, and the incidence of miscarriage is increased. Prolactin levels are often modestly increased **(Chap. 392)** and may contribute to alterations in libido and fertility and cause galactorrhea.

Myocardial contractility and pulse rate are reduced, leading to a reduced stroke volume and bradycardia. Increased peripheral resistance may be accompanied by hypertension, particularly diastolic. Blood flow is diverted from the skin, producing cool extremities. Pericardial effusions occur in up to 30% of patients but rarely compromise cardiac function. Although alterations in myosin heavy chain isoform expression have been documented, cardiomyopathy is rare. Fluid may also accumulate in other serous cavities and in the middle ear, giving rise to conductive deafness; sensorineural deafness may also occur. Pulmonary function is generally normal, but dyspnea may be caused by pleural effusion, impaired respiratory muscle function, diminished ventilatory drive, or sleep apnea.

Carpal tunnel and other entrapment syndromes are common, as is impairment of muscle function with stiffness, cramps, and pain. On examination, there may be slow relaxation of tendon reflexes and pseudomyotonia. Memory and concentration are impaired. Experimentally, positron emission tomography (PET) scans examining glucose metabolism in hypothyroid subjects show lower regional activity in the amygdala, hippocampus, and perigenual anterior cingulate cortex, among other regions, and this activity corrects after thyroxine replacement. Rare neurologic problems include reversible cerebellar ataxia, dementia, psychosis, and myxedema coma. *Hashimoto's encephalopathy* has been defined as a steroid-responsive syndrome associated with TPO antibodies, myoclonus, and slow-wave activity on electroencephalography, but the relationship with thyroid autoimmunity or hypothyroidism is not established, and if a patient is euthyroid, levothyroxine (LT4) therapy has not been shown to be efficacious in treatment. The hoarse voice and occasionally clumsy speech of hypothyroidism reflect fluid accumulation in the vocal cords and tongue.

The features described above are the consequence of thyroid hormone deficiency. However, autoimmune hypothyroidism may be associated with signs or symptoms of other autoimmune diseases, particularly vitiligo, pernicious anemia, Addison's disease (Schmidt's syndrome), alopecia areata, and type 1 diabetes mellitus (T1DM). In the polygenic disorder autoimmune polyendocrine syndrome type 2, autoimmune thyroid disease is present in 70–75%, T1DM in 40–60%, and Addison's disease in 40–50%. Less common associations include celiac disease, dermatitis herpetiformis, chronic active hepatitis, rheumatoid arthritis, systemic lupus erythematosus (SLE), myasthenia gravis, autoimmune hypoparathyroidism, primary hypogonadism, and Sjögren's syndrome. Thyroid-associated ophthalmopathy usually occurs in Graves' disease **(Chap. 396)**, but in ~5% of patients, it is associated with autoimmune hypothyroidism.

Autoimmune hypothyroidism is uncommon in children and usually presents with slow growth and delayed facial and dental maturation. The pituitary may be enlarged due to thyrotroph hyperplasia. Myopathy, with muscle swelling, is more common in children than in adults. In most cases, puberty is delayed, but precocious puberty sometimes occurs. There may be intellectual impairment if the onset is before 3 years and the hormone deficiency is severe.

Laboratory Evaluation A summary of the investigations used to determine the existence and cause of hypothyroidism is provided in **Fig. 395-2**. A normal TSH level excludes primary (but not secondary or central) hypothyroidism. If the TSH is elevated, a free or unbound T_4 level (FT_4) is needed to confirm the presence of clinical hypothyroidism, but T_4 is inferior to TSH when used as a screening test because it will not detect subclinical hypothyroidism. Circulating unbound T_3 levels are normal in ~25% of patients, reflecting adaptive deiodinase responses to hypothyroidism. T_3 measurements are, therefore, not indicated.

Once clinical or subclinical hypothyroidism is confirmed, the etiology is usually easily established by demonstrating the presence of TPO and Tg antibodies, which are present in >95% of patients with autoimmune hypothyroidism. Other abnormal laboratory findings in hypothyroidism may include increased creatine phosphokinase, elevated cholesterol and triglycerides, and anemia (usually normocytic or macrocytic) depending upon the degree and duration of hypothyroidism. Except when accompanied by iron deficiency or B_{12} deficiency from concomitant pernicious anemia, the anemia and other abnormalities gradually resolve with thyroxine replacement.

Differential Diagnosis An asymmetric goiter in Hashimoto's thyroiditis may be confused with a multinodular goiter (MNG) or thyroid carcinoma, in which thyroid antibodies may also be present. Primary thyroid lymphoma **(Chap. 397)** is rare but strongly associated with preexisting autoimmune thyroiditis. Ultrasound can be used to show the presence of solitary or multiple nodules rather than the thyroid enlargement with heterogeneous echogenicity typical of Hashimoto's thyroiditis. However, ultrasound imaging may also detect pseudonodules, hypoechoic areas likely representing areas of lymphocytic infiltrates, which need to be distinguished from true superimposed nodules as fine-needle aspirate is not warranted for these. Fine-needle aspiration

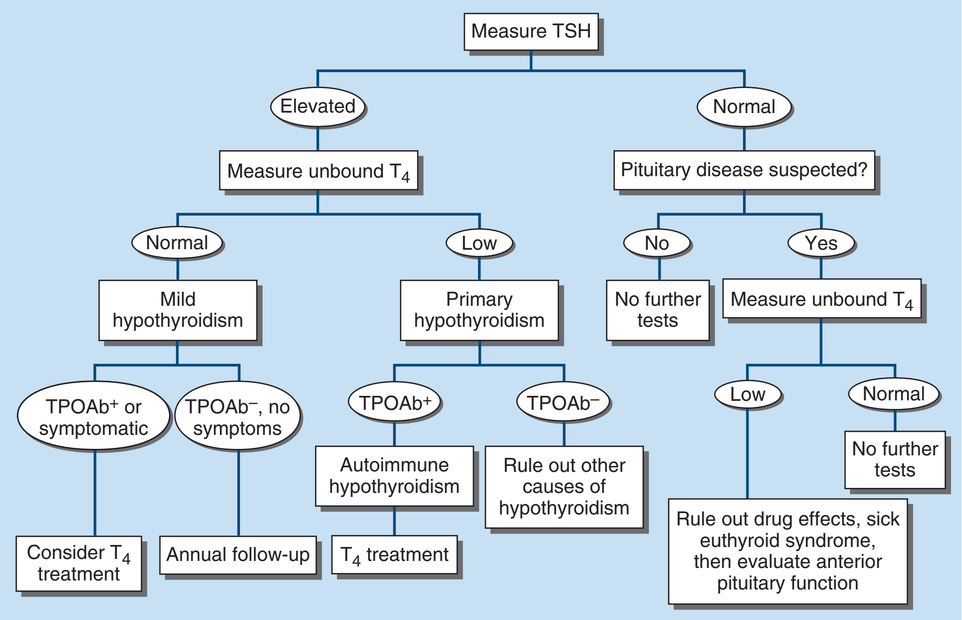

FIGURE 395-2 Evaluation of hypothyroidism. TPOAb⁺, thyroid peroxidase antibodies present; TPOAb⁻, thyroid peroxidase antibodies not present; TSH, thyroid-stimulating hormone.

(FNA) biopsy should be performed for all true nodules meeting FNA criteria **(Chap. 397)**. Other causes of hypothyroidism are discussed below and in Table 395-1 but rarely cause diagnostic confusion.

OTHER CAUSES OF HYPOTHYROIDISM

Iatrogenic hypothyroidism is a common cause of hypothyroidism and can often be detected by screening before symptoms develop. In the first 3–4 months after radioiodine treatment for Graves' disease, transient hypothyroidism may occur due to reversible radiation damage. Low-dose thyroxine treatment can be withdrawn if recovery occurs. Because TSH levels are suppressed by hyperthyroidism, unbound T_4 levels are a better measure of thyroid function than TSH in the months following radioiodine treatment. Mild hypothyroidism after subtotal thyroidectomy or lobectomy may also resolve after several months, as the gland remnant is stimulated by increased TSH levels.

Iodine deficiency is responsible for endemic goiter and cretinism but is an uncommon cause of adult hypothyroidism unless the iodine intake is very low or there are complicating factors, such as the consumption of thiocyanates in cassava or selenium deficiency. Although hypothyroidism due to iodine deficiency can be treated with thyroxine, public health measures to improve iodine intake should be advocated to eliminate this problem. Iodized salt or bread or a single bolus of oral or intramuscular iodized oil have all been used successfully.

Paradoxically, chronic iodine excess can also induce goiter and hypothyroidism. The intracellular events that account for this effect are unclear, but individuals with autoimmune thyroiditis are especially susceptible. Iodine excess is responsible for the hypothyroidism that occurs in patients treated with amiodarone **(Chap. 396)**. Other drugs, particularly lithium, may also cause hypothyroidism. Transient hypothyroidism may also be caused by thyroiditis **(Chap. 396)**.

Secondary or central hypothyroidism is usually diagnosed in the context of other anterior pituitary hormone deficiencies; isolated TSH deficiency is very rare **(Chap. 391)**. TSH levels may be low, normal, or even slightly increased in secondary hypothyroidism; the latter is due to secretion of immunoactive but bioinactive forms of TSH. The diagnosis is confirmed by detecting a low unbound T_4 level. The goal of treatment is to maintain free T_4 levels in the upper half of the reference interval because TSH levels cannot be used to monitor therapy.

TREATMENT

Hypothyroidism

CLINICAL HYPOTHYROIDISM

If there is no residual thyroid function, the daily replacement dose of LT4 is usually 1.6 μg/kg body weight (typically 100–150 μg), ideally taken at least 30 min before breakfast. In many patients, however, lower doses suffice until residual thyroid tissue is destroyed. In patients who develop hypothyroidism after the treatment of Graves' disease, there is often underlying autonomous function, necessitating lower replacement doses (typically 75–125 μg/d).

Adult patients under 60 years old without evidence of heart disease may be started on 50–100 μg of LT4 daily. The dose is adjusted on the basis of TSH levels, with the goal of treatment being a normal TSH, ideally in the lower half of the reference range. TSH responses are gradual and should be measured about 6–8 weeks after instituting treatment or after any subsequent change in LT4 dosage. The clinical effects of LT4 replacement are slow to appear. Patients may not experience full relief from symptoms until several months after normal TSH levels are restored. Adjustment of LT4 dosage is made in 12.5- or 25-μg increments if the TSH is high; decrements of the same magnitude should be made if the TSH is suppressed. Patients with a suppressed TSH of any cause, including LT4 overtreatment, have an increased risk of atrial fibrillation and reduced bone density.

About 10–15% of patients may have persistent symptoms despite restoration of euthyroidism with LT4 for reasons that remain unclear. Although desiccated animal thyroid preparations (thyroid extract USP) are available, they are not recommended because the ratio of T_3 to T_4 is nonphysiologic. The use of LT4 combined with liothyronine (triiodothyronine, T_3) has been investigated, but benefit has not been confirmed in prospective studies. There is no place for liothyronine alone as long-term replacement, because the short half-life necessitates three or four daily doses and is associated with fluctuating T_3 levels.

Once full replacement is achieved and TSH levels are stable, follow-up measurement of TSH is recommended at annual intervals. It is important to ensure ongoing adherence as patients do not feel any symptomatic difference after missing a few doses of LT4, and this sometimes leads to self-discontinuation.

In patients of normal body weight who are taking ≥200 μg of LT4 daily, an elevated TSH level is often a sign of poor adherence to treatment. This is also the likely explanation for fluctuating TSH levels, despite a constant LT4 dosage. Such patients often have normal or high free T_4 levels, despite an elevated TSH, because they remember to take medication for a few days before testing; this is sufficient to normalize T_4, but not TSH levels. It is important to consider variable adherence because this pattern of thyroid function tests is otherwise suggestive of disorders associated with inappropriate TSH secretion **(Chap. 394)**. Because T_4 has a long half-life (7 days), patients who miss a dose can be advised to take two doses of the skipped tablets at once. Other causes of increased LT4 requirements must be excluded, particularly malabsorption (e.g., celiac disease, small-bowel surgery, atrophic or *Helicobacter pylori*–related gastritis), oral estrogen-containing medications or selective estrogen receptor modulator therapy, ingestion with a meal, and drugs that interfere with T_4 absorption or metabolism such as bile acid sequestrants, ferrous sulfate, calcium supplements, sevelamer, sucralfate, proton pump inhibitors, lovastatin, aluminum hydroxide, rifampicin, amiodarone, carbamazepine, phenytoin, and tyrosine kinase inhibitors.

SUBCLINICAL HYPOTHYROIDISM

Subclinical hypothyroidism refers to biochemical evidence of thyroid hormone deficiency in patients who have few or no apparent clinical features of hypothyroidism. There are no universally accepted recommendations for the management of subclinical hypothyroidism, but LT4 is recommended if the patient is a woman who wishes to conceive or is pregnant or when TSH levels are >10 mIU/L. Most other patients can simply be monitored annually. A trial of treatment may be considered when young or middle-aged patients have symptoms of hypothyroidism or risk of heart disease. It is important to confirm that any elevation of TSH is sustained over a 3-month period before treatment is given. Treatment is administered by starting with a low dose of LT4 (25–50 μg/d) with the goal of normalizing TSH.

SPECIAL TREATMENT CONSIDERATIONS

Rarely, LT4 replacement is associated with pseudotumor cerebri in children. Presentation appears to be idiosyncratic and occurs months after treatment has begun.

Because maternal hypothyroidism may both adversely affect fetal neural development and be associated with adverse gestational outcomes (miscarriage, preterm delivery), thyroid function should be monitored to preserve euthyroidism in women with a history or a high risk of hypothyroidism. Although epidemiologic studies have demonstrated the association of miscarriage and preterm delivery with the presence of thyroid autoantibodies detected either during or prior to gestation in euthyroid women, randomized controlled multicenter trials evaluating LT4 therapy prior to conception in this population have not demonstrated benefit. Because of the known increase in thyroid hormone requirements during pregnancy in hypothyroid women, LT4 therapy should be targeted to maintain a serum TSH in the normal range but <2.5 mIU/L prior to conception.

In women without evidence of thyroid dysfunction, serum TSH decreases in the late first trimester, and if trimester-specific ranges are not available, an appropriate range for 7–12 weeks' gestation can be approximated by decreasing the upper limit of the nonpregnant reference range by 0.5 mIU/L (~4.0 mIU/L) and the lower limit by 0.4 mIU/L (~0.1 mIU/L). However, it is important to recognize that the normal TSH range in pregnancy for the second and third trimesters is not significantly different from the nonpregnancy reference range. Therefore, when caring for LT4-replaced women with hypothyroidism, thyroid function should be evaluated immediately after pregnancy is confirmed and every 4 weeks during the first half of the pregnancy, with less frequent testing after 20 weeks' gestation (every 6–8 weeks depending on whether LT4 dose adjustment is ongoing). The increment of LT4 dosage increase depends upon the etiology of hypothyroidism, with athyreotic women requiring more (~45%) than those with Hashimoto's who may have some residual thyroid function. Women should increase LT4 from once-daily dosing to nine doses per week as soon as pregnancy is confirmed to anticipate this change. Thereafter, dosage should be closely monitored with a goal TSH in the lower half of the trimester-specific normative range, if available, or <2.5 mIU/L, which allows for reserve if additional LT4 dosage increases are required as pregnancy progresses. After delivery, LT4 doses typically return to prepregnancy levels. Pregnant women should be counseled to separate ingestion of prenatal vitamins and iron supplements from LT4.

Elderly patients may require 20% less thyroxine than younger patients. In the elderly, especially patients with known coronary artery disease, the starting dose of LT4 is 12.5–25 μg/d with similar increments every 2–3 months until TSH is normalized. In some patients, it may be impossible to achieve full replacement despite optimal antianginal treatment. *Emergency surgery* is generally safe in patients with untreated hypothyroidism, although routine surgery in a hypothyroid patient should be deferred until euthyroidism is achieved.

Myxedema coma still has a 20–40% mortality rate, despite intensive treatment, and outcomes are independent of the T_4 and TSH levels. Clinical manifestations include reduced level of consciousness, sometimes associated with seizures, as well as the other features of hypothyroidism (Table 395-3). Hypothermia can reach 23°C (74°F). There may be a history of treated hypothyroidism with poor compliance, or the patient may be previously undiagnosed. Myxedema coma almost always occurs in the elderly and is usually precipitated by factors that impair respiration, such as drugs (especially sedatives, anesthetics, and antidepressants), pneumonia, congestive heart failure, myocardial infarction, gastrointestinal bleeding, or cerebrovascular accidents. Sepsis should also be suspected. Exposure to cold may also be a risk factor. Hypoventilation, leading to hypoxia and hypercapnia, plays a major role in pathogenesis; hypoglycemia and dilutional hyponatremia also contribute to the development of myxedema coma.

LT4 can initially be administered as a single IV bolus of 200–400 μg, which serves as a loading dose, followed by a daily oral dose of 1.6 μg/kg per d, reduced by 25% if administered IV. If a suitable IV preparation is not available, the same initial dose of LT4 can be given by nasogastric tube (although absorption may be impaired in myxedema). Because $T_4 \rightarrow T_3$ conversion is impaired in myxedema coma, there is a rationale for adding liothyronine (T_3) intravenously or via nasogastric tube to LT4 treatment, although excess liothyronine has the potential to provoke arrhythmias. An initial loading dose of 5–20 μg liothyronine should be followed by 2.5–10 μg every 8 h, with lower doses chosen for smaller or older patients and those at cardiovascular risk.

Supportive therapy should be provided to correct any associated metabolic disturbances. External warming is indicated only if the temperature is <30°C, as it can result in cardiovascular collapse **(Chap. 477)**. Space blankets should be used to prevent further heat loss. Parenteral hydrocortisone (50 mg every 6 h) should be administered because there is impaired adrenal reserve in profound hypothyroidism. Any precipitating factors should be treated, including the early use of broad-spectrum antibiotics, pending the exclusion of infection. Ventilatory support with regular blood gas analysis is usually needed during the first 48 h. Hypertonic saline or IV glucose may be needed if there is severe hyponatremia or hypoglycemia; hypotonic IV fluids should be avoided because they may exacerbate water retention secondary to reduced renal perfusion and inappropriate vasopressin secretion. The metabolism of most medications is impaired, and sedatives should be avoided if possible or used in reduced doses. Medication blood levels should be monitored, when available, to guide dosage.

■ FURTHER READING

Biondi B et al: Subclinical hypothyroidism in older individuals. Lancet Diabetes Endocrinol 10:129, 2022.

Chaker L et al: Hypothyroidism. Nature Rev Dis Primers 8:30, 2022.

Hegedüs L et al: Primary hypothyroidism and quality of life. Nature Rev Endocrinol 18:230, 2022.

Jonklaas J et al: Guidelines for the treatment of hypothyroidism: Prepared by the American Thyroid Association Task Force on thyroid hormone replacement. Thyroid 24:1670, 2014.

Lee SY et al: Assessment and treatment of thyroid disorders in pregnancy and the postpartum period. Nat Rev Endocrinol 18:158, 2022.

van Trotsenburg P et al: Congenital hypothyroidism: A 2020-2021 consensus guidelines update-an ENDO-European reference network initiative endorsed by the European Society for Pediatric Endocrinology and the European Society for Endocrinology. Thyroid 31:387, 2021.

396

Hyperthyroidism and Other Causes of Thyrotoxicosis

Anthony P. Weetman, Susan J. Mandel, J. Larry Jameson

THYROTOXICOSIS

Thyrotoxicosis is defined as the state of thyroid hormone excess and is not synonymous with *hyperthyroidism*, which is the result of excessive thyroid function. However, the major etiologies of thyrotoxicosis are hyperthyroidism caused by Graves' disease, toxic multinodular goiter (MNG), and toxic adenomas. Other causes are listed in **Table 396-1**.

■ GRAVES' DISEASE

Epidemiology Graves' disease accounts for 60–80% of thyrotoxicosis. The prevalence varies among populations, reflecting genetic factors and iodine intake (high iodine intake is associated with an increased prevalence of Graves' disease). Graves' disease occurs in up to 2% of women but is one-tenth as frequent in men. The disorder rarely begins before adolescence and typically occurs between 20 and 50 years of age; it also occurs in the elderly.

Pathogenesis As in autoimmune hypothyroidism, a combination of environmental and genetic factors, including polymorphisms in HLA-DR, the immunoregulatory genes *CTLA-4*, *CD25*, *CD40*, *PTPN22*, *FCRL3*, and *CD226*, as well as the gene encoding the thyroid-stimulating hormone (TSH) receptor (TSH-R), contributes to Graves' disease susceptibility. The concordance for Graves' disease in monozygotic twins is 20–30%, compared to <5% in dizygotic twins. Indirect evidence suggests that stress is an important environmental factor, presumably operating through neuroendocrine effects on the immune system. Smoking is a moderate risk factor for Graves' disease and a major risk factor for the development of ophthalmopathy. Sudden increases in iodine intake may precipitate Graves' disease, and there is

TABLE 396-1 Causes of Thyrotoxicosis

Primary Hyperthyroidism
Graves' disease
Toxic multinodular goiter
Toxic adenoma
Functioning thyroid carcinoma metastases
Activating mutation of the TSH receptor
Activating mutation of $G_s\alpha$ (McCune-Albright syndrome)
Struma ovarii
Drugs: iodine excess (Jod-Basedow phenomenon)
Thyrotoxicosis without Hyperthyroidism
Subacute thyroiditis
Silent thyroiditis
Other causes of thyroid destruction: drugs (amiodarone, cytokines, tyrosine kinase inhibitors, immune checkpoint inhibitors), radiation, infarction of adenoma
Ingestion of excess thyroid hormone (thyrotoxicosis factitia) or thyroid tissue
Secondary Hyperthyroidism
TSH-secreting pituitary adenoma
Thyroid hormone resistance syndrome: occasional patients may have features of thyrotoxicosis
Chorionic gonadotropin-secreting tumors[a]
Gestational thyrotoxicosis[a]

[a] Circulating TSH levels are low in these forms of secondary hyperthyroidism.

Abbreviation: TSH, thyroid-stimulating hormone.

a threefold increase in the occurrence of Graves' disease in the postpartum period. Graves' disease may occur during the immune reconstitution phase after highly active antiretroviral therapy (HAART) or alemtuzumab treatment and following treatment with immune checkpoint inhibitors (e.g., nivolumab, pembrolizumab).

The hyperthyroidism of Graves' disease is caused by thyroid-stimulating immunoglobulins (TSIs) that are synthesized by lymphocytes in the thyroid gland as well as in bone marrow and lymph nodes. Such antibodies can be detected by bioassays or by using the more widely available immunoassays (TSH receptor antibodies [TRAb]) that measure whether the patient's serum contains an antibody that can displace either labeled TSH or a monoclonal TSH receptor antibody from the TSH receptor. The presence of TRAb in a patient with thyrotoxicosis implies the existence of TSI, and these assays are useful in monitoring pregnant Graves' patients in whom high levels of TSI can cross the placenta and cause neonatal thyrotoxicosis. Other thyroid autoimmune responses, similar to those in autoimmune hypothyroidism (see above), occur concurrently in patients with Graves' disease. In particular, thyroid peroxidase (TPO) and thyroglobulin (Tg) antibodies occur in up to 80% of cases. Because the coexisting lymphocytic thyroiditis can also affect thyroid function, there is no direct correlation between the level of TSI and thyroid hormone levels in Graves' disease.

Cytokines appear to play a major role in thyroid-associated ophthalmopathy. There is infiltration of the extraocular muscles by activated T cells; the release of cytokines such as interferon γ (IFN-γ), tumor necrosis factor (TNF), and interleukin 1 (IL-1) results in fibroblast activation and increased synthesis of glycosaminoglycans that trap water, thereby leading to characteristic muscle swelling. Late in the disease, there is irreversible fibrosis of the muscles. Increased fat is an additional cause of retrobulbar tissue expansion. The increase in intraorbital pressure can lead to proptosis, diplopia, and optic neuropathy. Although the pathogenesis of thyroid-associated ophthalmopathy is incompletely understood, the TSH-R is a thyroid autoantigen and is expressed in orbital tissues. In addition, aberrant signaling via insulin-like growth factor 1 receptors (IGF-1R) on orbital fibroblasts has also been implicated. These mechanisms are the basis for new monoclonal antibody treatments (e.g., teprotumumab) that reduce the levels of TSH-R/IGF-1R complexes and attenuate signaling.

Clinical Manifestations Signs and symptoms include features that are common to any cause of thyrotoxicosis **(Table 396-2)** as well as those specific for Graves' disease. The clinical presentation depends on the severity of thyrotoxicosis, the duration of disease, individual susceptibility to excess thyroid hormone, and the patient's age. In the elderly, features of thyrotoxicosis may be subtle or masked, and patients may present mainly with fatigue and weight loss, a condition known as *apathetic thyrotoxicosis*.

TABLE 396-2 Signs and Symptoms of Thyrotoxicosis (Descending Order of Frequency)

SYMPTOMS	SIGNS[a]
Hyperactivity, irritability, dysphoria	Tachycardia; atrial fibrillation in the elderly
Heat intolerance and sweating	Tremor
Palpitations	Goiter
Fatigue and weakness	Warm, moist skin
Weight loss with increased appetite	Muscle weakness, proximal myopathy
Diarrhea	Lid retraction or lag
Polyuria	Gynecomastia
Oligomenorrhea, loss of libido	

[a] Excludes the signs of ophthalmopathy and dermopathy specific for Graves' disease.

Thyrotoxicosis may cause unexplained weight loss, despite an enhanced appetite, due to the increased metabolic rate. Weight gain occurs in 5% of patients, however, because of increased food intake. Other prominent features include hyperactivity, nervousness, and irritability, ultimately leading to a sense of easy fatigability in some patients. Insomnia and impaired concentration are common; apathetic thyrotoxicosis may be mistaken for depression in the elderly. Fine tremor is a frequent finding, best elicited by having patients stretch out their fingers while feeling the fingertips with the palm. Common neurologic manifestations include hyperreflexia, muscle wasting, and proximal myopathy without fasciculation. Chorea is rare. Thyrotoxicosis is sometimes associated with a form of hypokalemic periodic paralysis; this disorder is particularly common in Asian males with thyrotoxicosis, but it occurs in other ethnic groups as well.

The most common cardiovascular manifestation is sinus tachycardia, often associated with palpitations, occasionally caused by supraventricular tachycardia. The high cardiac output produces a bounding pulse, widened pulse pressure, and an aortic systolic murmur and can lead to worsening of angina or heart failure in the elderly or those with preexisting heart disease. Atrial fibrillation is more common in patients >50 years of age. Treatment of the thyrotoxic state alone converts atrial fibrillation to normal sinus rhythm in up to 75% of patients without an underlying cardiac problem.

The skin is usually warm and moist, and the patient may complain of sweating and heat intolerance, particularly during warm weather. Palmar erythema, onycholysis, and, less commonly, pruritus, urticaria, and diffuse hyperpigmentation may be evident. Hair texture may become fine, and diffuse alopecia occurs in up to 40% of patients, persisting for months after restoration of euthyroidism. Gastrointestinal transit time is decreased, leading to increased stool frequency, often with diarrhea and occasionally mild steatorrhea. Women frequently experience oligomenorrhea or amenorrhea; in men, there may be impaired sexual function and, rarely, gynecomastia. The direct effect of thyroid hormones on bone resorption leads to osteopenia in long-standing thyrotoxicosis; mild hypercalcemia occurs in up to 20% of patients, but hypercalciuria is more common. There is a small increase in fracture rate in patients with a previous history of thyrotoxicosis.

In Graves' disease, the thyroid is usually diffusely enlarged to two to three times its normal size. The consistency is firm, but not nodular. There may be a thrill or bruit, best detected at the inferolateral margins of the thyroid lobes, due to the increased vascularity of the gland and the hyperdynamic circulation.

Lid retraction, causing a staring appearance, can occur in any form of thyrotoxicosis and is the result of sympathetic overactivity. However, Graves' disease is associated with specific eye signs that comprise *Graves' ophthalmopathy* **(Fig. 396-1*A*)**. This condition is also called *thyroid eye disease* (TED) because it occurs in the absence

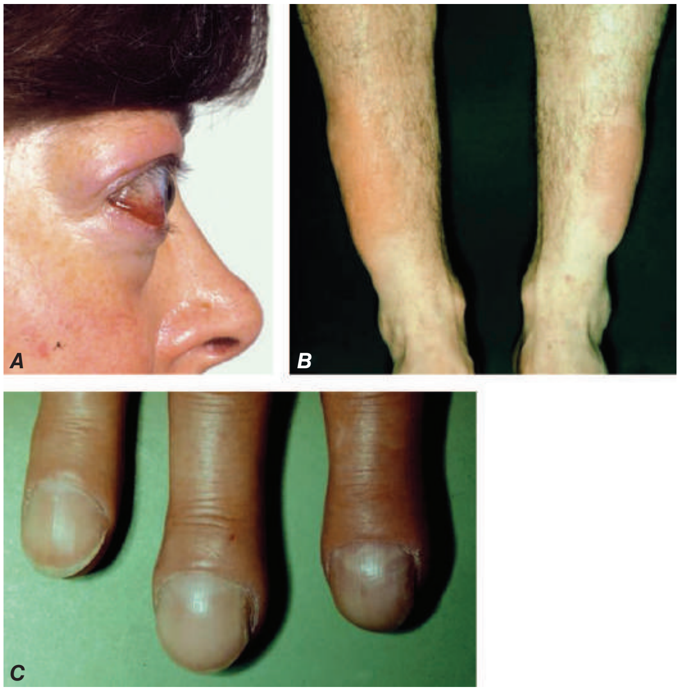

FIGURE 396-1 Features of Graves' disease. ***A.*** Ophthalmopathy in Graves' disease; lid retraction, periorbital edema, conjunctival injection, and proptosis are marked. ***B.*** Thyroid dermopathy over the lateral aspects of the shins. ***C.*** Thyroid acropachy.

of hyperthyroidism in 10% of patients. Most of these individuals have autoimmune hypothyroidism or thyroid antibodies. The onset of Graves' ophthalmopathy occurs within the year before or after the diagnosis of thyrotoxicosis in 75% of patients but can sometimes precede or follow thyrotoxicosis by several years, accounting for some cases of euthyroid ophthalmopathy.

About one-third of patients with Graves' disease have clinical evidence of ophthalmopathy. However, the enlarged extraocular muscles typical of the disease, and other subtle features, can be detected in most patients when investigated by ultrasound or computed tomography (CT) imaging of the orbits. Unilateral signs are found in up to 10% of ophthalmopathy patients. The earliest manifestations of ophthalmopathy are usually a sensation of grittiness, eye discomfort, and excess tearing. About one-third of patients have proptosis, best detected by visualization of the sclera between the lower border of the iris and the lower eyelid, with the eyes in the primary position. Proptosis can be measured using an exophthalmometer. In severe cases, proptosis may cause corneal exposure and damage, especially if the lids fail to close during sleep. Periorbital edema, scleral injection, and chemosis are also frequent. In 5–10% of patients, the muscle swelling is so severe that diplopia results, typically, but not exclusively, when the patient looks up and laterally. The most serious manifestation is compression of the optic nerve at the apex of the orbit, leading to papilledema; peripheral field defects; and, if left untreated, permanent loss of vision.

The "NO SPECS" scoring system to evaluate ophthalmopathy is an acronym derived from the following changes:

0 = **N**o signs or symptoms
1 = **O**nly signs (lid retraction or lag), no symptoms
2 = **S**oft tissue involvement (periorbital edema)
3 = **P**roptosis (>22 mm)
4 = **E**xtraocular muscle involvement (diplopia)
5 = **C**orneal involvement
6 = **S**ight loss

Although useful as a mnemonic, the NO SPECS scheme is inadequate to describe the eye disease fully, and patients do not necessarily progress from one class to another; alternative scoring systems (e.g., the EUGOGO system developed by the European Group on Graves' Orbitopathy) that assess disease activity are preferable for monitoring and treatment purposes. When Graves' eye disease is active and moderate to severe, evaluation and management with an ophthalmologist is indicated, and objective measurements are needed, such as lid-fissure width; corneal staining with fluorescein; evaluation of extraocular muscle function (e.g., Hess chart), intraocular pressure and visual fields, acuity, and color vision; and orbital imaging with CT or magnetic resonance imaging (MRI).

Thyroid dermopathy occurs in <5% of patients with Graves' disease **(Fig. 396-1*B*)**, almost always in the presence of moderate or severe ophthalmopathy. Although most frequent over the anterior and lateral aspects of the lower leg (hence the term *pretibial myxedema*), skin changes can occur at other sites, particularly after trauma. The typical lesion is a noninflamed, indurated plaque with a deep pink or purple color and an "orange skin" appearance. Nodular involvement can occur, and the condition can rarely extend over the whole lower leg and foot, mimicking elephantiasis. *Thyroid acropachy* refers to a form of clubbing found in <1% of patients with Graves' disease **(Fig. 396-1*C*)**. It is so strongly associated with thyroid dermopathy that an alternative cause of clubbing should be sought in a Graves' patient without coincident skin and orbital involvement. Ophthalmopathy, dermopathy, and acropachy have declined in incidence, probably due to better recognition and prompt treatment of the underlying thyroid disease.

Laboratory Evaluation Investigations used to determine the existence and cause of thyrotoxicosis are summarized in **Fig. 396-2**. In Graves' disease, the TSH level is suppressed, and total and unbound thyroid hormone levels are increased. In 2–5% of patients (and more in areas of borderline iodine intake), only T_3 is increased (T_3 toxicosis). The converse state of T_4 toxicosis, with elevated total and unbound T_4 and normal T_3 levels, is occasionally seen when hyperthyroidism is induced by excess iodine, providing surplus substrate for thyroid hormone synthesis. Measurement of TRAb is also useful. Associated abnormalities that may cause diagnostic confusion in thyrotoxicosis include elevation of bilirubin, liver enzymes, and ferritin. Microcytic anemia and thrombocytopenia may occur.

Differential Diagnosis Diagnosis of Graves' disease is straightforward in a patient with biochemically confirmed thyrotoxicosis, diffuse goiter on palpation, ophthalmopathy, and often a personal or family history of autoimmune disorders. TRAb measurement is usually used to confirm the diagnosis of Graves' disease in patients with thyrotoxicosis who lack these features, but the diagnosis can also be established by a radionuclide (^{99m}Tc, ^{123}I, or ^{131}I) scan and uptake of the thyroid, which will distinguish the diffuse, high uptake of Graves' disease from destructive thyroiditis, ectopic thyroid tissue, and factitious thyrotoxicosis, as well as diagnose a toxic adenoma or toxic MNG. Alternatively, color-flow Doppler ultrasonography distinguishes between hyperthyroidism (with increased blood flow) and destructive thyroiditis and avoids using radioactivity. In secondary hyperthyroidism due to a TSH-secreting pituitary tumor, there is also a diffuse goiter. The presence of a nonsuppressed TSH level and the finding of a pituitary tumor on CT or MRI scan suggest this diagnosis.

Clinical features of thyrotoxicosis can mimic certain aspects of other disorders, including panic attacks, mania, pheochromocytoma, and weight loss associated with malignancy. The diagnosis of thyrotoxicosis can be easily excluded if the TSH and unbound T_4 and T_3 levels are normal. A normal TSH also excludes Graves' disease as a cause of diffuse goiter.

Clinical Course Clinical features generally worsen without treatment; mortality was 10–30% before the introduction of satisfactory therapy. Some patients with mild Graves' disease experience spontaneous relapses and remissions. Rarely, there may be fluctuation between hypo- and hyperthyroidism due to changes in the functional activity of TSH-R antibodies. About 15% of patients who enter remission after treatment develop hypothyroidism 10–15 years later because of the destructive autoimmune process.

The clinical course of ophthalmopathy does not follow that of the thyroid disease, although thyroid dysfunction can worsen eye signs. Ophthalmopathy typically worsens over the initial 3–6 months, followed by a plateau phase over the next 12–18 months, and then some

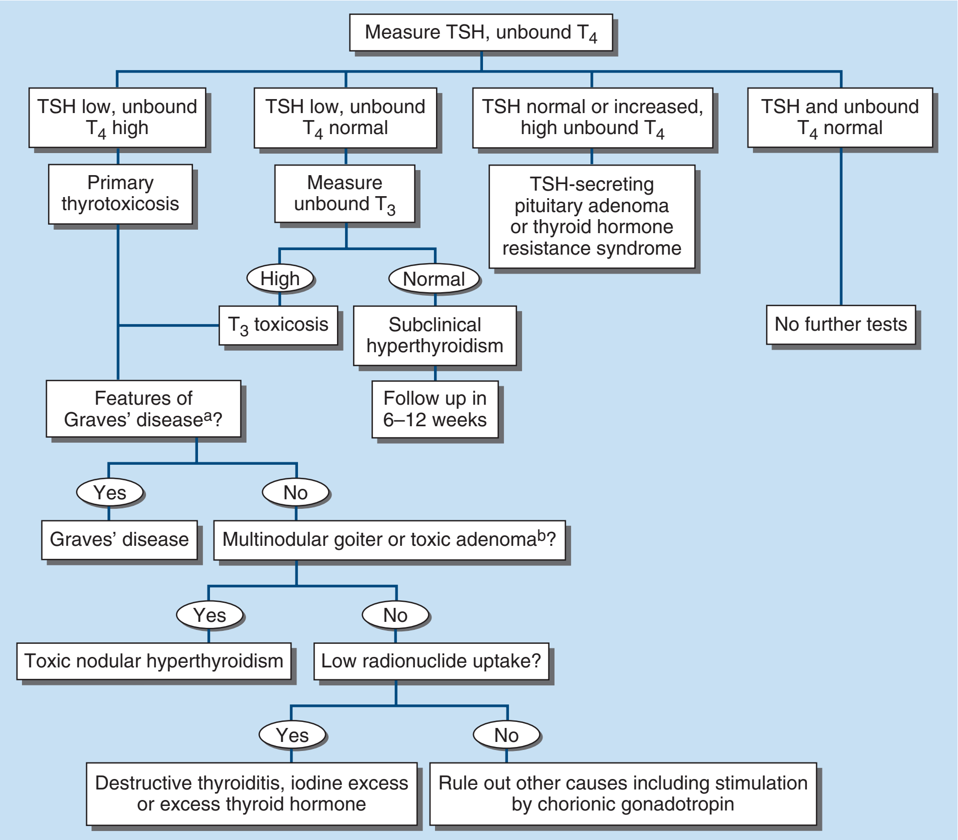
Measure TSH, unbound T_4
TSH low, unbound T_4 high
TSH low, unbound T_4 normal
TSH normal or increased, high unbound T_4
TSH and unbound T_4 normal
Primary thyrotoxicosis
Measure unbound T_3
TSH-secreting pituitary adenoma or thyroid hormone resistance syndrome
High
Normal
T_3 toxicosis
Subclinical hyperthyroidism
No further tests
Features of Graves' disease[a]?
Follow up in 6–12 weeks
Yes
No
Graves' disease
Multinodular goiter or toxic adenoma[b]?
Yes
No
Toxic nodular hyperthyroidism
Low radionuclide uptake?
Yes
No
Destructive thyroiditis, iodine excess or excess thyroid hormone
Rule out other causes including stimulation by chorionic gonadotropin

FIGURE 396-2 Evaluation of thyrotoxicosis. [a]Diffuse goiter, positive thyroid peroxidase (TPO) antibodies or thyroid-stimulating hormone (TSH) receptor antibody (TRAb), ophthalmopathy, dermopathy. [b]Can be confirmed by radionuclide scan.

spontaneous improvement, particularly in the soft tissue changes. However, the course is more fulminant in up to 5% of patients, requiring intervention in the acute phase if there is optic nerve compression or corneal ulceration. Diplopia may appear late in the disease due to fibrosis of the extraocular muscles. Radioiodine treatment for hyperthyroidism worsens the eye disease in a small proportion of patients (especially smokers). Antithyroid drugs and surgery have no adverse effects on the clinical course of ophthalmopathy. Thyroid dermopathy, when it occurs, usually appears 1–2 years after the development of Graves' hyperthyroidism; it may improve spontaneously.

TREATMENT

Graves' Disease

The *hyperthyroidism* of Graves' disease is treated by reducing thyroid hormone synthesis, using an antithyroid drug, or reducing the amount of thyroid tissue with radioiodine (^{131}I) treatment or by thyroidectomy. Antithyroid drugs are the predominant initial therapy in many centers in Europe, Latin America, and Japan, whereas radioiodine is more often the first line of treatment in North America. These differences reflect the fact that no single approach is optimal and that patients may require multiple treatments to achieve remission.

The main *antithyroid drugs* are thionamides: propylthiouracil, carbimazole (not available in the United States), and the active metabolite of the latter, methimazole. All inhibit the function of TPO, reducing oxidation and organification of iodide. These drugs also reduce thyroid antibody levels by mechanisms that remain unclear, and they appear to enhance spontaneous rates of remission. Propylthiouracil inhibits deiodination of $T_4 \rightarrow T_3$. However, this effect is of minor benefit, except in the most severe thyrotoxicosis, and is offset by the much shorter half-life of this drug (90 min) compared to methimazole (6 h). Due to the hepatotoxicity of propylthiouracil, the U.S. Food and Drug Administration (FDA) has limited indications for its use to the first trimester of pregnancy, the treatment of thyroid storm, and patients with minor adverse reactions to methimazole. If propylthiouracil is used, monitoring of liver function tests is recommended.

There are many variations of antithyroid drug regimens. The initial dose of carbimazole or methimazole is usually 10–20 mg every 12 h, but once-daily dosing is possible after euthyroidism is restored. Propylthiouracil is given at a dose of 100–200 mg every 6–8 h, and divided doses are usually given throughout the course. Lower doses of each drug may suffice in areas of low iodine intake. The starting dose of an antithyroid drug can be gradually reduced (titration regimen) as thyrotoxicosis improves. Less commonly, high doses may be given combined with levothyroxine (LT4) supplementation (block-replace regimen) to avoid drug-induced hypothyroidism. The titration regimen is often preferred to minimize the dose of antithyroid drug and provide an index of treatment response.

Thyroid function tests and clinical manifestations are reviewed 4–6 weeks after starting treatment, and the dose is titrated based on unbound T_4 levels. Most patients do not achieve euthyroidism until 6–8 weeks after treatment is initiated. TSH levels often remain suppressed for several months and therefore do not provide a sensitive index of treatment response. The usual daily maintenance doses of antithyroid drugs in the titration regimen are 2.5–10 mg of carbimazole or methimazole and 50–100 mg of propylthiouracil. In the block-replace regimen, the initial dose of antithyroid drug is held constant, and the dose of LT4 is adjusted to maintain normal unbound T_4 levels. When TSH suppression is alleviated, TSH levels can also be used to monitor therapy.

Maximum remission rates (30–60%) are achieved by 12–18 months for the titration regimen and are higher in patients in whom TRAb levels are no longer detected than in those with TRAb persistence. For unclear reasons, remission rates appear to vary in different geographic regions. Younger patients, males, smokers, and patients with a history of allergy, severe hyperthyroidism, or large goiters are most likely to relapse when treatment stops, as are those with enduring TRAb, but outcomes are difficult to predict. All patients should be followed closely for relapse during the first year after treatment and at least annually thereafter. Prolonged treatment for up to 10 years with small doses of an antithyroid drug has been used as an alternative to ablative therapies following relapse.

The common minor side effects of antithyroid drugs are rash, urticaria, fever, and arthralgia (1–5% of patients). These may resolve spontaneously or after substituting an alternative antithyroid drug; rashes may respond to an antihistamine. Rare but major side effects include hepatitis (especially with propylthiouracil; avoid use in children) and cholestasis (methimazole and carbimazole); vasculitis; and, most important, agranulocytosis (<1%). It is essential that antithyroid drugs are stopped and not restarted if a patient develops major side effects. Written instructions should be provided regarding the symptoms of possible agranulocytosis (e.g., sore throat, fever, mouth ulcers) and the need to stop treatment pending an urgent complete blood count to confirm that agranulocytosis is not present. Management of agranulocytosis is described in **Chap. 107**. It is not useful to monitor blood counts prospectively, because the onset of agranulocytosis is idiosyncratic and abrupt.

Propranolol (20–40 mg every 6 h) or longer-acting selective β_1 receptor blockers such as atenolol may be helpful to control adrenergic symptoms, especially in the early stages before antithyroid drugs take effect. Beta blockers are also useful in patients with thyrotoxic periodic paralysis, pending correction of thyrotoxicosis. In consultation with a cardiologist or using a risk score like CHA_2DS_2-VASc, anticoagulation should be considered in all patients with atrial fibrillation; the majority revert spontaneously to sinus rhythm with control of hyperthyroidism, and long-term anticoagulation is not usually needed. Decreased warfarin doses are required when patients are thyrotoxic. If digoxin is used, increased doses are often needed in the thyrotoxic state.

Radioiodine causes progressive destruction of thyroid cells and can be used as initial treatment or for relapses after a trial of antithyroid drugs. There is a small risk of thyrotoxic crisis (see below) after radioiodine, which can be minimized by pretreatment with antithyroid drugs for at least a month before treatment. Antecedent treatment with an antithyroid drug and a beta blocker should be considered for all elderly patients or for those with cardiac problems. Carbimazole or methimazole must be stopped 2–3 days before radioiodine administration to achieve optimum iodine uptake and can be restarted 3–7 days after radioiodine in those at risk of complications from worsening thyrotoxicosis. Propylthiouracil appears to have a prolonged radioprotective effect and should be stopped for a longer period before radioiodine is given, or a larger dose of radioiodine will be necessary.

Efforts to calculate an optimal dose of radioiodine that achieves euthyroidism without a high incidence of relapse or progression to hypothyroidism have not been successful. Some patients inevitably relapse after a single dose because the biologic effects of radiation vary between individuals, and hypothyroidism cannot be uniformly avoided even using accurate dosimetry. A practical strategy is to give a fixed dose based on clinical features, such as the severity of thyrotoxicosis, the size of the goiter (larger goiters need a higher dosage) and the radioiodine uptake (higher uptake decreases the dosage needed). ^{131}I dosage generally ranges between 370 MBq (10 mCi) and 555 MBq (15 mCi). Most authorities favor an approach aimed at thyroid ablation (as opposed to euthyroidism), given that LT4 replacement is straightforward and most patients ultimately progress to hypothyroidism over 5–10 years.

Certain radiation safety precautions are necessary in the first few days after radioiodine treatment, but the exact guidelines vary depending on local protocols. In general, patients need to avoid close, prolonged contact with children and pregnant women for 5–7 days because of possible transmission of residual isotope and exposure to radiation emanating from the gland. Rarely, there may be mild pain due to radiation thyroiditis 1–2 weeks after treatment. Hyperthyroidism can persist for 2–3 months before radioiodine takes full effect. For this reason, β-adrenergic blockers or antithyroid drugs, which can be restarted 5–7 days after radioiodine administration, can be used to control symptoms during this interval. Persistent hyperthyroidism can be treated with a second dose of radioiodine, usually 6 months after the first dose. The risk of hypothyroidism after radioiodine depends on the dosage but is at least 10–20% in the first year and 5% per year thereafter and 5% per year thereafter, with higher rates after ablative treatment. Patients should be informed of this possibility before treatment and require close follow-up during the first year followed by annual thyroid function testing.

Pregnancy and breast-feeding are absolute contraindications to radioiodine treatment, but patients can conceive safely 6 months after treatment. The presence of ophthalmopathy, especially in smokers, requires caution. Prednisone, 0.2–0.5 mg/kg per d (depending on ophthalmopathy severity), at the time of radioiodine treatment, tapered over 6–12 weeks, may prevent exacerbation of ophthalmopathy, but radioiodine should generally be avoided in patients with active moderate-to-severe eye disease. Although many physicians avoid radioiodine in children and adolescents because of the potential risks of malignancy, others have advocated radioiodine use in older children. There is no overall increase in cancer risk after radioiodine. Patients may be advised that the risks are small enough to be indistinguishable from antithyroid drugs or surgery.

Total or near-total thyroidectomy is an option for patients who relapse after antithyroid drugs and prefer this treatment to radioiodine. Some experts recommend surgery in young individuals, particularly when the goiter is very large. Careful control of thyrotoxicosis with antithyroid drugs, followed by potassium iodide (SSKI; 1–2 drops orally tid for 10 days), is needed prior to surgery to avoid thyrotoxic crisis and to reduce the vascularity of the gland. The major complications of surgery—bleeding, laryngeal edema, hypoparathyroidism, and damage to the recurrent laryngeal nerves—are unusual when the procedure is performed by highly experienced surgeons. Recurrence rates in the best series are <2%, but the rate of hypothyroidism is similar to that following radioiodine treatment, especially with the current trend away from subtotal thyroidectomy.

Antithyroid drugs should be used to manage active Graves' disease in *pregnancy.* Because transplacental passage of these drugs may produce fetal hypothyroidism and goiter if the maternal dose is excessive, maternal antithyroid dose titration should target serum free or total T_4 levels at or just above the pregnancy reference range. If available, propylthiouracil should be used until 14–16 weeks' gestation because of the risk of rare cases of methimazole/carbimazole embryopathy, including *aplasia cutis* and other defects, such as choanal atresia and tracheoesophageal fistulae. Because of the potential for teratogenic effects, antithyroid medication should be discontinued in any newly pregnant woman with Graves' disease who is euthyroid on a low dose of methimazole (<5–10 mg/d) or propylthiouracil (<100–200 mg/d), after evaluating recent thyroid function tests, disease history, goiter size, duration of therapy, and TRAb measurement. Following cessation, careful monitoring of maternal thyroid function tests is essential. On the other hand, for women at high risk of developing thyrotoxicosis if antithyroid drugs are discontinued (large goiter, requirement for higher antithyroid drug dosage, elevated TRAb), continued therapy is necessary, with propylthiouracil (if available) administration in the first trimester. Because of its rare association with hepatotoxicity, propylthiouracil should be limited to the first trimester and then maternal therapy should be converted to methimazole (or carbimazole) at a ratio of 15–20 mg of propylthiouracil to 1 mg of methimazole. It is often possible to stop treatment in the last trimester because TSI tend to decline in pregnancy. Nonetheless, the transplacental transfer of these antibodies, if present, at levels three times higher than the normative range may rarely cause *fetal* or *neonatal thyrotoxicosis*. Poor intrauterine growth, a fetal heart rate of >160 beats/min, advanced bone age, fetal goiter, and high levels of maternal TSI after 26 weeks' gestation may herald this complication. Antithyroid drugs given to the mother can be used to treat the fetus and may be needed for 1–3 months after delivery, until the maternal antibodies disappear from the baby's circulation. The postpartum period is a time of major risk for relapse of Graves' disease. Breast-feeding is safe with low doses of antithyroid drugs. Graves' disease in *children* is usually managed initially with methimazole or carbimazole (avoid propylthiouracil), often given as a course of the titration regimen for at least 3 years. Surgery or radioiodine may be indicated for severe or relapsing disease.

Thyrotoxic crisis, or *thyroid storm*, is rare and presents as a life-threatening exacerbation of hyperthyroidism, accompanied by fever, delirium, seizures, coma, vomiting, diarrhea, and jaundice. The mortality rate due to cardiac failure, arrhythmia, or hyperthermia is high (4–17%) even with treatment. Thyrotoxic crisis is usually precipitated by acute illness (e.g., stroke, infection, trauma, diabetic ketoacidosis), surgery (especially on the thyroid), or radioiodine treatment of a patient with partially treated or untreated hyperthyroidism. Management requires intensive monitoring and supportive care, identification and treatment of the precipitating cause, and measures that reduce thyroid hormone synthesis. Large doses of propylthiouracil (500–1000 mg loading dose and 250 mg every 4 h) should be given orally or by nasogastric tube or per rectum; the drug's inhibitory action on $T_4 \rightarrow T_3$ conversion makes it the antithyroid drug of choice. If not available, methimazole can be used in doses of 20 mg every 6 h. One hour after the first dose of propylthiouracil or methimazole, stable iodide (5 drops SSKI every 6 h) is given to block thyroid hormone synthesis via the Wolff-Chaikoff effect (the delay allows the antithyroid drug to prevent the excess iodine from being incorporated into new hormone). Propranolol should also be given to reduce tachycardia and other adrenergic manifestations (60–80 mg PO every 4 h, or 2 mg IV every 4 h). Although other β-adrenergic blockers can be used, high doses of propranolol decrease $T_4 \rightarrow T_3$ conversion, and the doses can be easily adjusted. Caution is needed to avoid acute negative inotropic effects, but controlling the heart rate is important, as some patients develop a form of high-output heart failure. Short-acting IV esmolol can be used to decrease heart rate while monitoring for signs of heart failure. Additional therapeutic measures include glucocorticoids (e.g., hydrocortisone 300 mg IV bolus, then 100 mg every 8 h), antibiotics if infection is present, cholestyramine to sequester thyroid hormones, cooling, oxygen, and IV fluids.

Mild ophthalmopathy requires no active treatment, because there is usually spontaneous improvement. General measures include meticulous control of thyroid hormone levels, cessation of smoking, and an explanation of the natural history of ophthalmopathy. Discomfort can be relieved with artificial tears (e.g., hypromellose 0.3% or carbomer 0.2% ophthalmic gel), paraffin-based eye ointment, and the use of dark glasses with side frames. Periorbital edema may respond to a more upright sleeping position or a diuretic. Corneal exposure during sleep can be avoided by using patches or taping the eyelids shut. Minor degrees of diplopia improve with prisms fitted to spectacles. Some authorities also advocate selenium 100 μg bid. *Moderate-to-severe ophthalmopathy,* in which the eye disease has sufficient impact on daily life to justify the risks of treatment, is usually treated with IV methylprednisolone (e.g., 500 mg of methylprednisolone once weekly for 6 weeks, then 250 mg once weekly for 6 weeks). This is preferable to oral glucocorticoids. A poor response at 6 weeks generally indicates the need for alternative treatment. Once the eye disease has stabilized, surgery may be indicated for relief of diplopia and correction of appearance. External beam radiotherapy of the orbits has been used for many years, but the efficacy of this therapy remains unclear, and it is best reserved for those who are not responsive to glucocorticoid therapy. Teprotumumab, a monoclonal antibody inhibitor of the IGF-1 receptor, improves proptosis, diplopia, clinical activity score, and quality of life and is recommended as first-line treatment in patients with active moderate-to-severe ophthalmopathy with significant proptosis or diplopia, but use may be limited by availability and cost. Relapse rates are similar to glucocorticoids (30%). Rituximab and tocilizumab are other monoclonal antibodies that have been used as second-line treatment after glucocorticoids. *Sight-threatening ophthalmopathy* due to optic nerve compression or corneal damage is an emergency that requires immediate high-dose IV glucocorticoids (e.g., methylprednisolone 500–1000 mg on alternate days). If there is a poor response after 2 weeks, surgical orbital decompression should be considered by removing bone from the medial and inferior orbital wall, thereby allowing displacement of fat and swollen extraocular muscles. In addition to nerve decompression, proptosis recedes an average of 5 mm, but there may be residual or even worsened diplopia.

Thyroid dermopathy does not usually require treatment, but it can cause cosmetic problems or interfere with the fit of shoes. Surgical removal is not indicated. If necessary, treatment consists of topical, high-potency glucocorticoid ointment under an occlusive dressing or compression dressing. Teprotumumab has been beneficial in some cases.

OTHER CAUSES OF THYROTOXICOSIS

Destructive thyroiditis (subacute or silent thyroiditis) typically presents with a short thyrotoxic phase due to the release of preformed thyroid hormones and catabolism of Tg (see "Subacute Thyroiditis," below). True hyperthyroidism is absent, as demonstrated by a low radionuclide uptake. Circulating Tg levels are typically increased. Other causes of thyrotoxicosis with low or absent thyroid radionuclide uptake include *thyrotoxicosis factitia*, iodine excess, and, rarely, ectopic thyroid tissue, particularly teratomas of the ovary (*struma ovarii*) and functional metastatic follicular carcinoma. Whole-body radionuclide studies can demonstrate ectopic thyroid tissue, and thyrotoxicosis factitia can be distinguished from destructive thyroiditis by the clinical features and low levels of Tg. Amiodarone treatment is associated with thyrotoxicosis in up to 10% of patients, particularly in areas of low iodine intake (see below).

TSH-secreting pituitary adenoma is a rare cause of thyrotoxicosis. It is characterized by the presence of an inappropriately normal or increased TSH level in a patient with hyperthyroidism, diffuse goiter, and elevated T_4 and T_3 levels **(Chap. 392)**. Elevated levels of the α subunit of TSH, released by the TSH-secreting adenoma, support this diagnosis, which can be confirmed by demonstrating the pituitary tumor on MRI or CT scan. A combination of transsphenoidal surgery, sella irradiation, and octreotide may be required to normalize TSH, because many of these tumors are large and locally invasive at the time of diagnosis. Radioiodine or antithyroid drugs can be used to control thyrotoxicosis.

Thyrotoxicosis caused by *toxic MNG* and *hyperfunctioning solitary nodules* is discussed below **(Chap. 396)**.

THYROIDITIS

There are several classification systems to describe the clinical syndromes of thyroiditis. One is based on the onset and duration of disease **(Table 396-3)**; others are based on the absence or presence of pain.

ACUTE THYROIDITIS

Acute thyroiditis is rare and due to suppurative infection of the thyroid. In children and young adults, the most common cause is the presence of a piriform sinus, a remnant of the fourth branchial pouch that connects the oropharynx with the thyroid. Such sinuses are predominantly left-sided. A long-standing goiter and degeneration in a thyroid malignancy are risk factors in the elderly. The patient presents with thyroid pain, often referred to the throat or ears, and a small, tender goiter that may be asymmetric. Fever, dysphagia, and erythema over the thyroid are common, as are systemic symptoms of a febrile illness and lymphadenopathy.

The differential diagnosis of *thyroid pain* includes subacute or, rarely, chronic thyroiditis; hemorrhage into a cyst; malignancy including lymphoma; and, rarely, amiodarone-induced thyroiditis or amyloidosis. However, the abrupt presentation and clinical features of acute thyroiditis rarely cause confusion. The erythrocyte sedimentation rate (ESR) and white cell count are usually increased, but thyroid function is generally normal. Fine-needle aspiration (FNA) biopsy shows infiltration by polymorphonuclear leukocytes; culture of the sample

TABLE 396-3 Causes of Thyroiditis

Acute
Bacterial infection: especially *Staphylococcus, Streptococcus,* and *Enterobacter*
Fungal infection: *Aspergillus, Candida, Coccidioides, Histoplasma,* and *Pneumocystis*
Radiation thyroiditis after ^{131}I treatment
Amiodarone (may also be subacute or chronic)
Subacute
Viral (or granulomatous) thyroiditis
Silent thyroiditis (including postpartum thyroiditis)
Mycobacterial infection
Drug-induced (interferon, amiodarone, tyrosine kinase inhibitors, immune checkpoint inhibitors)
Chronic
Autoimmunity: focal thyroiditis, Hashimoto's thyroiditis, atrophic thyroiditis
Riedel's thyroiditis
Parasitic thyroiditis: echinococcosis, strongyloidiasis, cysticercosis
Traumatic: after palpation

can identify the organism. Caution is needed in immunocompromised patients as fungal, mycobacterial, or *Pneumocystis* thyroiditis can occur in this setting. Antibiotic treatment is guided initially by Gram stain and, subsequently, by cultures of the FNA biopsy. Surgery may be needed to drain an abscess, which can be localized by CT scan or ultrasound. Tracheal obstruction, septicemia, retropharyngeal abscess, mediastinitis, and jugular venous thrombosis may complicate acute thyroiditis but are uncommon with prompt use of antibiotics.

SUBACUTE THYROIDITIS

This is also termed *de Quervain's thyroiditis*, *granulomatous thyroiditis*, or *viral thyroiditis*. Many viruses have been implicated, including mumps, coxsackie, influenza, adenoviruses, and echoviruses. Subacute thyroiditis may also occur with the SARS-CoV-2 illness or after receiving the COVID vaccine. The diagnosis may be overlooked because the symptoms can mimic pharyngitis. Attempts to identify the causative virus in an individual patient are often unsuccessful and do not influence management. The peak incidence occurs at 30–50 years, and women are affected three times more frequently than men.

Pathophysiology The thyroid shows a characteristic patchy inflammatory infiltrate with disruption of the thyroid follicles and multinucleated giant cells within some follicles. The follicular changes progress to granulomas accompanied by fibrosis. Finally, the thyroid returns to normal, usually several months after onset. During the initial phase of follicular destruction, there is release of Tg and thyroid hormones, leading to increased circulating T_4 and T_3 and suppression of TSH **(Fig. 396-3)**. During this destructive phase, radioactive iodine uptake is low or undetectable. After several weeks, the thyroid is depleted of stored thyroid hormone and a phase of hypothyroidism typically occurs, with low unbound T_4 (and sometimes T_3) and moderately increased TSH levels. Radioactive iodine uptake returns to normal or is even increased as a result of the rise in TSH. Finally, thyroid hormone and TSH levels return to normal as the disease subsides.

Clinical Manifestations The patient usually presents with a painful and enlarged thyroid, sometimes accompanied by fever. There may be features of thyrotoxicosis or hypothyroidism, depending on the phase of the illness. Malaise and symptoms of an upper respiratory tract infection may precede the thyroid-related features by several weeks. In other patients, the onset is acute, severe, and without obvious antecedent. The patient typically complains of a sore throat, and examination reveals a small goiter that is exquisitely tender. Pain is often referred to the jaw or ear. Complete resolution is the usual outcome, but late-onset permanent hypothyroidism occurs in 15% of cases, particularly in those with coincidental thyroid autoimmunity. A prolonged course over many months, with one or more relapses, occurs in a small percentage of patients.

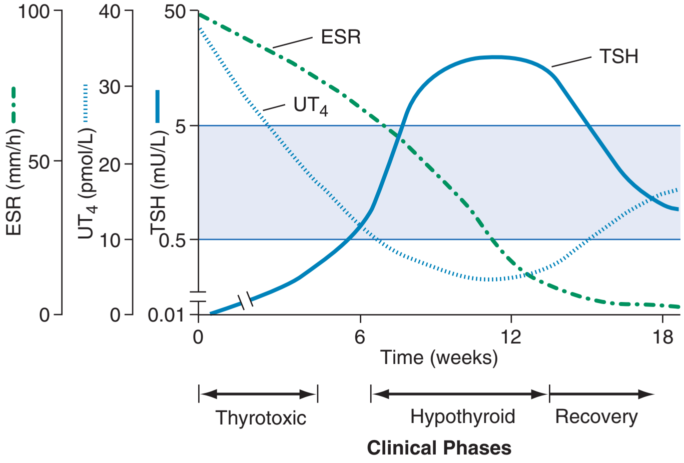

FIGURE 396-3 Clinical course of subacute thyroiditis. The release of thyroid hormones is initially associated with a thyrotoxic phase and suppressed thyroid-stimulating hormone (TSH). A hypothyroid phase then ensues, with low T_4 and TSH levels that are initially low but gradually increase. During the recovery phase, increased TSH levels combined with resolution of thyroid follicular injury lead to normalization of thyroid function, often several months after the beginning of the illness. ESR, erythrocyte sedimentation rate; UT_4, free or unbound T_4.

Laboratory Evaluation As depicted in Fig. 396-3, thyroid function tests characteristically evolve through three distinct phases over about 6 months: (1) thyrotoxic phase, (2) hypothyroid phase, and (3) recovery phase. In the thyrotoxic phase, T_4 and T_3 levels are increased, reflecting their discharge from the damaged thyroid cells, and TSH is suppressed. The T_4/T_3 ratio is lower than in Graves' or thyroid autonomy, in which T_3 is often disproportionately increased. The diagnosis is confirmed by a high ESR and low uptake of radioiodine (<5%) or ^{99m}Tc pertechnetate (as compared to salivary gland pertechnetate concentration). The white blood cell count may be increased, and thyroid antibodies are negative. If the diagnosis is in doubt, FNA biopsy may be useful, particularly to distinguish unilateral involvement from bleeding into a cyst or neoplasm.

TREATMENT

Subacute Thyroiditis

Relatively large doses of aspirin (e.g., 600 mg every 4–6 h) or nonsteroidal anti-inflammatory drugs (NSAIDs) are sufficient to control symptoms in many cases. Generally, gastroprotective medications, such as proton pump inhibitors, are also prescribed. If this treatment is inadequate, or if the patient has marked local or systemic symptoms, glucocorticoids should be given. The usual starting dose is 15–40 mg of prednisone, depending on severity. The dose is gradually tapered over 6–8 weeks, in response to improvement in symptoms and the ESR. If a relapse occurs during glucocorticoid withdrawal, the dosage should be increased and then withdrawn more gradually. Thyroid function should be monitored every 2–4 weeks using TSH and free T_4 levels. Symptoms of thyrotoxicosis improve spontaneously but may be ameliorated by β-adrenergic blockers; antithyroid drugs play no role in treatment of the thyrotoxic phase. LT4 replacement may be needed if the hypothyroid phase is prolonged, but doses should be low enough (50–100 μg daily) to allow TSH-mediated recovery.

SILENT THYROIDITIS

Painless thyroiditis, or "*silent*" *thyroiditis*, occurs in patients with underlying autoimmune thyroid disease and has a clinical course similar to that of subacute thyroiditis. The condition occurs in up to 5% of women 3–6 months after pregnancy and is then termed *postpartum thyroiditis*. Typically, patients have a brief phase of thyrotoxicosis lasting 2–4 weeks, followed by hypothyroidism for 4–12 weeks, and then resolution; often, however, only one phase is apparent. The condition is associated with the presence of TPO antibodies antepartum, and it is three times more common in women with type 1 diabetes mellitus. As in subacute thyroiditis, the uptake of ^{99m}Tc pertechnetate or radioactive iodine is initially suppressed. In addition to the painless goiter, silent thyroiditis can be distinguished from subacute thyroiditis by a normal ESR and the presence of TPO antibodies. Glucocorticoid treatment is not indicated for silent thyroiditis. Severe thyrotoxic symptoms can be managed with a brief course of propranolol, 20–40 mg three or four times daily. Thyroxine replacement may be needed for the hypothyroid phase but should be withdrawn after 6–9 months, as recovery is the rule. Annual follow-up thereafter is recommended because a proportion of these individuals develop permanent hypothyroidism. The condition may recur in subsequent pregnancies.

■ DRUG-INDUCED THYROIDITIS

Patients receiving cytokines, such as IFN-α, tyrosine kinase inhibitors, such as sorafenib, and immune checkpoint inhibitors may develop painless thyroiditis. IFN-α, which is used to treat chronic hepatitis B or C and hematologic and skin malignancies, causes thyroid dysfunction in up to 5% of treated patients. It has been associated with painless thyroiditis, hypothyroidism, and Graves' disease and is most common in women with TPO antibodies prior to treatment. Thyroiditis occurs in 5–20% of cancer patients treated with the immune checkpoint inhibitors pembrolizumab or nivolumab. In all cases, treatment is the same as silent thyroiditis. Routine monitoring of thyroid function tests is recommended by the American Society of Clinical Oncology. For discussion of amiodarone, see "Amiodarone Effects on Thyroid Function," below.

■ CHRONIC THYROIDITIS

Focal thyroiditis is present in 20–40% of euthyroid autopsy cases and is associated with serologic evidence of autoimmunity, particularly the presence of TPO antibodies. The most common clinically apparent cause of chronic thyroiditis is *Hashimoto's thyroiditis*, an autoimmune disorder that often presents as a firm or hard goiter of variable size **(Chap. 395)**. *Riedel's thyroiditis* is a rare disorder that typically occurs in middle-aged women. It presents with an insidious, painless goiter with local symptoms due to compression of the esophagus, trachea, neck veins, or recurrent laryngeal nerves. Dense fibrosis disrupts normal gland architecture and can extend outside the thyroid capsule. Despite these extensive histologic changes, thyroid dysfunction is uncommon. The goiter is hard, nontender, often asymmetric, and fixed, leading to suspicion of a malignancy. Diagnosis requires open biopsy as FNA biopsy is usually inadequate. Treatment is with glucocorticoids, other immunomodulatory treatments, tamoxifen, or surgical relief of compressive symptoms. There is an association between Riedel's thyroiditis and IgG4-related disease causing idiopathic fibrosis at other sites (retroperitoneum, mediastinum, biliary tree, lung, and orbit) **(Chap. 380)**.

SICK EUTHYROID SYNDROME (NONTHYROIDAL ILLNESS)

Any acute, severe illness can cause abnormalities of circulating TSH or thyroid hormone levels in the absence of underlying thyroid disease, making these measurements potentially misleading. The major cause of these hormonal changes is the release of cytokines such as IL-6. Unless a thyroid disorder is strongly suspected, the routine testing of thyroid function should be avoided in acutely ill patients.

The most common hormone pattern in sick euthyroid syndrome (SES), also called nonthyroidal illness (NTI), is a decrease in total and unbound T_3 levels (low T_3 syndrome) with normal levels of T_4 and TSH. The magnitude of the fall in T_3 correlates with the severity of the illness. T_4 conversion to T_3 via peripheral 5′ (outer ring) deiodination is impaired, leading to increased reverse T_3 (rT_3). Since rT_3 is metabolized by 5′ deiodination, its clearance is also reduced. Thus, decreased clearance rather than increased production is the major basis for increased rT_3. Also, T_4 is alternately metabolized to the hormonally inactive T_3 sulfate. It is generally assumed that this low T_3 state is adaptive, because it can be induced in normal individuals by fasting. Teleologically, the fall in T_3 may limit catabolism in starved or ill patients.

Very sick patients may exhibit a dramatic fall in total T_4 and T_3 levels (low T_4 syndrome). With decreased tissue perfusion, muscle and liver expression of the type 3 deiodinase leads to accelerated T_4 and T_3 metabolism. This state has a poor prognosis. Another key factor in the fall in T_4 levels is altered binding to thyroxine-binding globulin (TBG). The commonly used free T_4 assays are subject to artifact when serum binding proteins are low and underestimate the true free T_4 level. Fluctuation in TSH levels also creates challenges in the interpretation of thyroid function in sick patients. TSH levels may range from 0.01 to 0.1 mIU/L in very ill patients, especially with dopamine or glucocorticoid therapy, to >20 mIU/L during the recovery phase of SES. The exact mechanisms underlying the subnormal TSH seen in 10% of sick patients and the increased TSH seen in 5% remain unclear but may be mediated by cytokines including IL-12 and IL-18. However, if the serum TSH is undetectable (<0.01 mIU/L), primary thyroid disease is more likely and endocrine evaluation should be done.

Any severe illness can induce changes in thyroid hormone levels, but certain disorders exhibit a distinctive pattern of abnormalities. Acute liver disease is associated with an initial rise in total (but not unbound) T_3 and T_4 levels due to TBG release; these levels become subnormal with progression to liver failure. A transient increase in total and unbound T_4 levels, usually with a normal T_3 level, is seen in 5–30% of acutely ill psychiatric patients. TSH values may be transiently low, normal, or high in these patients. In the early stage of HIV infection, T_3 and T_4 levels rise, even if there is weight loss. T_3 levels fall with progression to AIDS, but TSH usually remains normal. Renal disease is often accompanied by low T_3 concentrations, but with normal rather than increased rT_3 levels, due to an unknown factor that increases uptake of rT_3 into the liver.

The diagnosis of NTI is challenging. Historic information may be limited, and patients often have multiple metabolic derangements. Useful features to consider include previous history of thyroid disease and thyroid function tests, evaluation of the severity and time course of the patient's acute illness, documentation of medications that may affect thyroid function or thyroid hormone levels, and measurements of rT_3 together with unbound thyroid hormones and TSH. The diagnosis of NTI is frequently presumptive, given the clinical context and pattern of laboratory values; only resolution of the test results with clinical recovery can clearly establish this disorder. Treatment of NTI with thyroid hormone (T_4 and/or T_3) is controversial, but most authorities recommend monitoring the patient's thyroid function tests during recovery, without administering thyroid hormone, unless there is historic or clinical evidence suggestive of hypothyroidism. Sufficiently large randomized controlled trials using thyroid hormone are unlikely to resolve this therapeutic controversy in the near future, because clinical presentations and outcomes are highly variable.

AMIODARONE EFFECTS ON THYROID FUNCTION

Amiodarone is a commonly used type III antiarrhythmic agent **(Chap. 259)**. It is structurally related to thyroid hormone and contains 39% iodine by weight. Thus, typical doses of amiodarone (200 mg/d) are associated with very high iodine intake, leading to greater than 40-fold increases in plasma and urinary iodine levels. Moreover, because amiodarone is stored in adipose tissue, high iodine levels persist for >6 months after discontinuation of the drug. Amiodarone inhibits deiodinase activity, and its metabolites function as weak antagonists of thyroid hormone action. Amiodarone has the following effects on thyroid function: (1) acute, transient suppression of thyroid function; (2) inhibition of T_4 to T_3 conversion causing either euthyroid hyperthyroxinemia or increased dosage requirement in LT4-treated hypothyroid patients; (3) hypothyroidism in patients susceptible to the inhibitory effects of a high iodine load; and (4) thyrotoxicosis that may be caused by either a Jod-Basedow effect from the iodine load, in the setting of MNG or incipient Graves' disease, or a thyroiditis-like condition due to a toxic effect on thyroid follicular cells.

The initiation of amiodarone treatment is associated with a transient decrease of T_4 levels, reflecting the inhibitory effect of iodine on T_4 release. Soon thereafter, most individuals escape from iodide-dependent suppression of the thyroid (Wolff-Chaikoff effect), and the inhibitory effects on deiodinase activity and thyroid hormone receptor action become predominant. These events lead to the following pattern of thyroid function tests: increased T_4, decreased T_3, increased rT_3, and a transient TSH increase (up to 20 mIU/L). TSH levels normalize or are slightly suppressed within 1–3 months.

The incidence of hypothyroidism from amiodarone varies geographically, apparently correlating with iodine intake. Hypothyroidism occurs in up to 13% of amiodarone-treated patients in iodine-replete countries, such as the United States, but is less common (<6% incidence) in areas of lower iodine intake, such as Italy or Spain. The pathogenesis appears to involve an inability of the thyroid gland to escape from the Wolff-Chaikoff effect in autoimmune thyroiditis. Consequently, amiodarone-associated hypothyroidism is more common in women and individuals with positive TPO antibodies. It is usually unnecessary to discontinue amiodarone for this side effect, because LT4 can be used to normalize thyroid function. TSH levels should be monitored, because T_4 levels are often increased for the reasons described above. In addition, TSH levels need to be monitored in LT4-replaced hypothyroid patients because a dosage increase is often required.

The management of amiodarone-induced thyrotoxicosis (AIT) is complicated by the fact that there are different causes of thyrotoxicosis and because the increased thyroid hormone levels exacerbate underlying arrhythmias and coronary artery disease. Amiodarone treatment causes thyrotoxicosis in 10% of patients living in areas of low iodine intake and in 2% of patients in regions of high iodine intake. There are two major forms of AIT, although some patients have features of both. Type 1 AIT is associated with an underlying thyroid abnormality (preclinical Graves' disease or nodular goiter). Thyroid hormone synthesis becomes excessive as a result of increased iodine exposure (Jod-Basedow phenomenon). Type 2 AIT occurs in individuals with no intrinsic thyroid abnormalities and is the result of drug-induced lysosomal activation leading to destructive thyroiditis with histiocyte accumulation in the thyroid; the incidence rises as cumulative amiodarone dosage increases. Mild forms of type 2 AIT can resolve spontaneously or can occasionally lead to hypothyroidism. Color-flow Doppler ultrasonography shows increased vascularity in type 1 AIT but decreased vascularity in type 2 AIT. Thyroid scintiscans are difficult to interpret in this setting because the high endogenous iodine levels diminish tracer uptake. However, the presence of normal or rarely increased uptake favors type 1 AIT.

In AIT, because of amiodarone's prolonged half-life and storage in adipose tissue, there is no immediate benefit from discontinuing the drug. Treatment studies have reported a similar time course for AIT type 2 resolution whether amiodarone was continued or stopped. Therefore, the decision to discontinue the drug should be based upon the severity of the arrhythmia. High doses of antithyroid drugs can be used in type 1 AIT but are often ineffective. Potassium perchlorate, 500 mg bid, has been used to reduce thyroidal iodide content. Perchlorate treatment has been associated with agranulocytosis, although the risk appears relatively low with short-term use. This drug is also no longer available in many countries. Glucocorticoids (initial dose usually 40–60 mg/d of prednisone) are effective in treating in type 2 AIT. Near-total thyroidectomy rapidly decreases thyroid hormone levels and may be the most effective long-term solution if the patient can undergo the procedure safely.

■ FURTHER READING

Alexander EK et al: 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. Thyroid 27:315, 2017.

Biondi B, Cooper DS: Subclinical hyperthyroidism. N Engl J Med 378:2411, 2018.

Burch HB et al: Management of thyroid eye disease: A Consensus Statement by the American Thyroid Association and European Thyroid Association. Thyroid 32:1439, 2022.

Kim BW: Does radioactive iodine therapy for hyperthyroidism cause cancer? J Clin Endocrinol Metab 107:e448, 2022.

Ross DS et al: 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. Thyroid 26:1343, 2016.

Wiersinga WM et al: Hyperthyroidism: Aetiology, pathogenesis, diagnosis, management, complications, and prognosis. Lancet Diabetes Endocrinol 11:282, 2023.

397

Thyroid Nodular Disease and Thyroid Cancer

Susan J. Mandel, Anthony P. Weetman, J. Larry Jameson

■ GOITER AND THYROID NODULAR DISEASE

Goiter refers to an enlarged thyroid gland. Biosynthetic defects, iodine deficiency, autoimmune disease, and nodular diseases can each lead to goiter, although by different mechanisms. Biosynthetic defects and iodine deficiency are associated with reduced efficiency of thyroid hormone synthesis, leading to increased thyroid-stimulating hormone (TSH), which stimulates thyroid growth as a compensatory mechanism to overcome the block in hormone synthesis. Graves' disease and Hashimoto's thyroiditis are also associated with goiter. In Graves' disease, the goiter results mainly from the TSH-R–mediated effects of thyroid-stimulating immunoglobulins. The goitrous form of Hashimoto's thyroiditis occurs because of acquired defects in hormone synthesis, leading to elevated levels of TSH and its consequent growth effects. Lymphocytic infiltration and immune system–induced growth factors also contribute to thyroid enlargement in Hashimoto's thyroiditis.

Thyroid nodular disease is characterized by the disordered growth of thyroid cells, which can be either hyperplastic or neoplastic. A patient may have a multinodular goiter (MNG) in which thyroid nodules (generally hyperplastic) replace the majority of the normal thyroid parenchyma; this presentation is more common in areas of borderline iodine deficiency. Or, the thyroid gland may be normal in size and contain discrete thyroid nodules. Because the management of goiter depends on the etiology, the detection of thyroid enlargement on physical examination should prompt further evaluation to identify its cause.

Nodular thyroid disease is common, occurring in about 3–7% of adults when assessed by physical examination. Using ultrasound, nodules are present in up to 50% of adults, with the majority being <1 cm in diameter. Thyroid nodules may be solitary or multiple, and they may be functional or nonfunctional.

■ DIFFUSE NONTOXIC (SIMPLE) GOITER

Etiology and Pathogenesis When diffuse enlargement of the thyroid occurs in the absence of nodules and hyperthyroidism, it is referred to as a *diffuse nontoxic goiter*. This is sometimes called *simple goiter* and is characterized by the presence of uniform follicles that are filled with colloid. Worldwide, diffuse goiter is most commonly caused by iodine deficiency and is termed *endemic goiter* when it affects >5% of the population. In nonendemic regions, *sporadic goiter* occurs, and the cause is usually unknown. Thyroid enlargement in teenagers is sometimes referred to as *juvenile goiter*. In general, goiter is more common in women than men, probably because of the greater prevalence of underlying autoimmune disease and the increased iodine demands associated with pregnancy.

In *iodine-deficient areas*, thyroid enlargement reflects a compensatory effort to trap iodide and produce sufficient hormone under conditions in which hormone synthesis is relatively inefficient. Somewhat surprisingly, TSH levels are usually normal or only slightly increased, suggesting increased sensitivity to TSH or activation of other pathways that lead to thyroid growth. Iodide appears to have direct actions on thyroid vasculature and may indirectly affect growth through vasoactive substances such as endothelins and nitric oxide. Endemic goiter may also be caused by exposure to environmental *goitrogens* such as cassava root, which contains a thiocyanate; vegetables of the Cruciferae family (known as cruciferous vegetables) (e.g., Brussels sprouts, cabbage, and cauliflower); and milk from regions where goitrogens are present in grass. Although relatively rare, inherited defects in thyroid hormone synthesis lead to a diffuse nontoxic goiter. Abnormalities at each step of hormone synthesis, including iodide transport (sodium/iodide symporter [NIS]), thyroglobulin (Tg) synthesis, organification and coupling (thyroid peroxidase [TPO]), and the regeneration of iodide (dehalogenase), have been described.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

If thyroid function is preserved, most goiters are asymptomatic. Examination of a diffuse goiter reveals a symmetrically enlarged, nontender, generally soft gland without palpable nodules. Goiter is defined, somewhat arbitrarily, as a lateral lobe with a volume greater than the thumb of the individual being examined. On ultrasound, total thyroid volume exceeding 30 mL is considered abnormal. If the thyroid is markedly enlarged, it can cause tracheal or esophageal compression. These features are unusual, however, in the absence of nodular disease and fibrosis. *Substernal goiter* may obstruct the thoracic inlet. *Pemberton's sign* refers to facial and neck congestion due to jugular venous obstruction when the arms are raised above the head, a maneuver that draws the thyroid into the thoracic inlet. Respiratory flow measurements and computed tomography (CT) or magnetic resonance imaging (MRI) should be used to evaluate substernal goiter in patients with obstructive signs or symptoms.

Thyroid function tests should be performed in all patients with goiter to exclude thyrotoxicosis or hypothyroidism. It is not unusual, particularly in iodine deficiency, to find a low total T_4, with normal T_3 and TSH, reflecting enhanced $T_4 \rightarrow T_3$ conversion, as well as preferential T_3 production. A low TSH with a normal free T_3 and free T_4, particularly in older patients, suggests the possibility of thyroid autonomy or undiagnosed Graves' disease, and is termed *subclinical thyrotoxicosis*. The benefit of treatment (typically with radioiodine) in subclinical thyrotoxicosis, versus follow-up and implementing treatment if free T_3 or free T_4 levels become abnormal, is unclear, but treatment is increasingly recommended in the elderly to reduce the risk of atrial fibrillation and bone loss. Low urinary iodine levels (<50 μg/L) support a diagnosis of iodine deficiency. Thyroid scanning is not generally necessary for euthyroid patients. If preformed, scintigraphy demonstrates increased uptake in iodine deficiency and most cases of dyshormonogenesis.

TREATMENT

Diffuse Nontoxic (Simple) Goiter

Iodine replacement induces variable regression of goiter in iodine deficiency, depending on duration and the degree of hyperplasia, with accompanying fibrosis, and autonomous function that may have developed. Surgery is rarely indicated for diffuse goiter. Exceptions include documented evidence of tracheal compression or obstruction of the thoracic inlet, which are more likely to be associated with substernal MNGs (see below). Subtotal or near-total thyroidectomy for these or cosmetic reasons should be performed by an experienced surgeon to minimize complication rates. Surgery should be followed by replacement with levothyroxine (LT4).

NONTOXIC MULTINODULAR GOITER

Etiology and Pathogenesis Depending on the population studied, MNG or the presence of nodules in a thyroid of normal size occurs in up to 12% of adults. MNG should be distinguished from the presence of nodules in a normal-size thyroid gland (see "Approach to the Patient with Thyroid Nodules"). MNG is more common in women than men and increases in prevalence with age. It is more common in iodine-deficient regions but also occurs in regions of iodine sufficiency, reflecting multiple genetic, autoimmune, and environmental influences on the pathogenesis.

There is typically wide variation in nodule size. Histology reveals a spectrum of morphologies ranging from hypercellular, hyperplastic regions to cystic areas filled with colloid. Fibrosis is often extensive, and areas of hemorrhage or lymphocytic infiltration may be seen. Using molecular techniques, most nodules within an MNG are polyclonal in origin, suggesting a hyperplastic response to locally produced growth factors and cytokines. TSH, which is usually not elevated, may play a permissive or contributory role. Monoclonal neoplastic lesions may also occur, reflecting mutations in genes that confer a selective growth advantage to the progenitor cell.

Clinical Manifestations Most patients with nontoxic MNG are asymptomatic and euthyroid. MNG typically develops over many years and is detected on routine physical examination, when an individual notices an enlargement in the neck, or as an incidental finding on imaging. If the goiter is large enough, it can ultimately lead to compressive symptoms including difficulty swallowing, respiratory distress (tracheal compression), or plethora (venous congestion), but these symptoms are uncommon. Symptomatic MNGs are usually large and/or develop fibrotic areas that cause compression. Sudden pain in an MNG is usually caused by hemorrhage into a nodule. Hoarseness, reflecting laryngeal nerve involvement, may suggest malignancy but more commonly is due to others causes such as gastroesophageal reflux.

Diagnosis On examination, thyroid architecture is distorted, and multiple nodules of varying size can be appreciated. Because many nodules are deeply embedded in thyroid tissue or reside in posterior or substernal locations, it is not possible to palpate all nodules. Pemberton's sign, characterized by facial suffusion when the patient's arms are elevated above the head, suggests that the goiter has increased pressure in the thoracic inlet. A TSH level should be measured to exclude subclinical hyper- or hypothyroidism, but thyroid function is usually normal. Tracheal deviation is common, but compression must usually exceed 70% of the tracheal diameter before there is significant airway compromise. Pulmonary function testing can be used to assess the functional effects of compression, which characteristically causes inspiratory stridor. CT or MRI can be used to evaluate the anatomy of the goiter and the extent of substernal extension or tracheal narrowing. A barium swallow may reveal the extent of esophageal compression. The risk of malignancy in MNG is similar to that in solitary nodules. Ultrasonography should be used to identify which nodules should be biopsied based on a combination of size and sonographic pattern **(Fig. 397-1) (Chap. 394)**. For nodules with more suspicious sonographic patterns (e.g., hypoechoic solid nodules with irregular borders), biopsy is recommended at a lower size cutoff than those with less suspicious imaging features (**Figs.** 397-1 and **397-2**).

TREATMENT

Nontoxic Multinodular Goiter

Most nontoxic MNGs can be managed conservatively. T_4 suppression is rarely effective for reducing goiter size and introduces the risk of subclinical or overt thyrotoxicosis, particularly if there is underlying autonomy or if it develops during treatment. Contrast agents and other iodine-containing substances should be avoided because of the risk of inducing the *Jod-Basedow effect*, characterized by enhanced thyroid hormone production by autonomous nodules. Radioiodine has been used when surgery

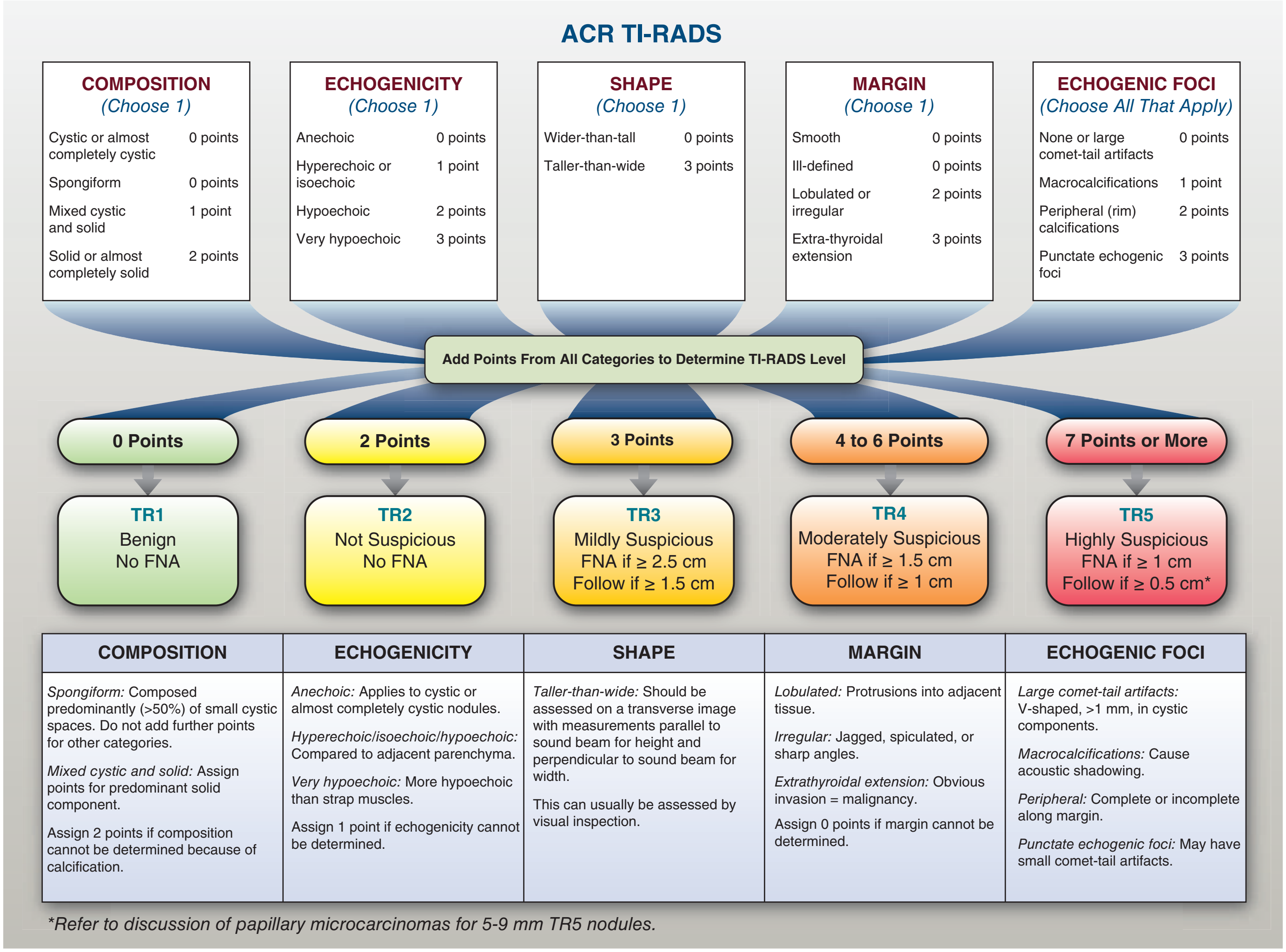

ACR TI-RADS

COMPOSITION (Choose 1)

Cystic or almost completely cystic	0 points
Spongiform	0 points
Mixed cystic and solid	1 point
Solid or almost completely solid	2 points

ECHOGENICITY (Choose 1)

Anechoic	0 points
Hyperechoic or isoechoic	1 point
Hypoechoic	2 points
Very hypoechoic	3 points

SHAPE (Choose 1)

Wider-than-tall	0 points
Taller-than-wide	3 points

MARGIN (Choose 1)

Smooth	0 points
Ill-defined	0 points
Lobulated or irregular	2 points
Extra-thyroidal extension	3 points

ECHOGENIC FOCI (Choose All That Apply)

None or large comet-tail artifacts	0 points
Macrocalcifications	1 point
Peripheral (rim) calcifications	2 points
Punctate echogenic foci	3 points

Add Points From All Categories to Determine TI-RADS Level

0 Points	2 Points	3 Points	4 to 6 Points	7 Points or More
TR1 Benign No FNA	TR2 Not Suspicious No FNA	TR3 Mildly Suspicious FNA if ≥ 2.5 cm Follow if ≥ 1.5 cm	TR4 Moderately Suspicious FNA if ≥ 1.5 cm Follow if ≥ 1 cm	TR5 Highly Suspicious FNA if ≥ 1 cm Follow if ≥ 0.5 cm*

COMPOSITION	ECHOGENICITY	SHAPE	MARGIN	ECHOGENIC FOCI
Spongiform: Composed predominantly (>50%) of small cystic spaces. Do not add further points for other categories. *Mixed cystic and solid:* Assign points for predominant solid component. Assign 2 points if composition cannot be determined because of calcification.	*Anechoic:* Applies to cystic or almost completely cystic nodules. *Hyperechoic/isoechoic/hypoechoic:* Compared to adjacent parenchyma. *Very hypoechoic:* More hypoechoic than strap muscles. Assign 1 point if echogenicity cannot be determined.	*Taller-than-wide:* Should be assessed on a transverse image with measurements parallel to sound beam for height and perpendicular to sound beam for width. This can usually be assessed by visual inspection.	*Lobulated:* Protrusions into adjacent tissue. *Irregular:* Jagged, spiculated, or sharp angles. *Extrathyroidal extension:* Obvious invasion = malignancy. Assign 0 points if margin cannot be determined.	*Large comet-tail artifacts:* V-shaped, >1 mm, in cystic components. *Macrocalcifications:* Cause acoustic shadowing. *Peripheral:* Complete or incomplete along margin. *Punctate echogenic foci:* May have small comet-tail artifacts.

**Refer to discussion of papillary microcarcinomas for 5-9 mm TR5 nodules.*

FIGURE 397-1 American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS). TI-RADS is a five-tiered system categorizing the sonographic appearance of thyroid nodules based on increased risk for malignancy. For each level (TR1–5), there are recommendations for both fine-needle aspiration (FNA) minimum size cutoffs and follow-up. *(Reproduced with permission from FN Tessler et al: ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. J Am Coll Radiol 14:587, 2017.)*

is contraindicated in areas where large nodular goiters are more prevalent (e.g., some areas of Europe and Brazil) because it can decrease MNG volume and may selectively ablate regions of autonomy. Dosage of ^{131}I depends on the size of the goiter and radioiodine uptake but is usually about 3.7 MBq (0.1 mCi) per gram of tissue, corrected for uptake (typical dose 370–1070 MBq [10–29 mCi]). Repeat treatment may be needed, and effectiveness may be increased by concurrent administration of low-dose recombinant TSH (0.1 mg IM). It is possible to achieve a 40–50% reduction in goiter size in most patients. Earlier concerns about radiation-induced thyroid swelling and tracheal compression have diminished, as studies have shown this complication to be rare. When acute compression occurs, glucocorticoid treatment or surgery may be needed. Radiation-induced hypothyroidism is less common than after treatment for Graves' disease. However, posttreatment autoimmune thyrotoxicosis may occur in up to 5% of patients treated for nontoxic MNG. Surgery remains highly effective but is not without risk, particularly in older patients with underlying cardiopulmonary disease.

■ TOXIC MULTINODULAR GOITER

The pathogenesis of toxic MNG appears to be similar to that of nontoxic MNG; the major difference is the presence of functional autonomy in toxic MNG. The molecular basis for autonomy in toxic MNG remains unknown. As in nontoxic goiters, many nodules are polyclonal, whereas others are monoclonal and vary in their clonal origins. Genetic abnormalities known to confer functional autonomy, such as activating TSH-R or $G_s\alpha$ mutations (see below), are not usually found in the autonomous regions of toxic MNG goiter.

In addition to features of goiter, the clinical presentation of toxic MNG includes subclinical or mild overt hyperthyroidism. The patient is usually elderly and may present with atrial fibrillation or palpitations, tachycardia, nervousness, tremor, or weight loss. Recent exposure to iodine, from contrast dyes or other sources, may precipitate or exacerbate thyrotoxicosis. The TSH level is low. The free T_4 level may be normal or minimally increased; T_3 is often elevated to a greater degree than T_4. Thyroid scan shows heterogeneous uptake with multiple regions of increased and decreased uptake; 24-h uptake of radioiodine may not be increased but is usually in the upper normal range in the presence of the low TSH level.

Prior to definitive treatment of the hyperthyroidism, ultrasound imaging should be performed to assess the presence of discrete nodules corresponding to areas of decreased uptake ("cold" nodules) **(Fig. 397-3A)**. If present, fine-needle aspiration (FNA) may be indicated based on sonographic patterns and size cutoffs (see "Approach to the Patient with Thyroid Nodules"). The cytology results, if indeterminate or suspicious, may direct the therapy to surgery.

TREATMENT

Toxic Multinodular Goiter

Antithyroid drugs normalize thyroid function and are particularly useful in the elderly or ill patients with limited life span.

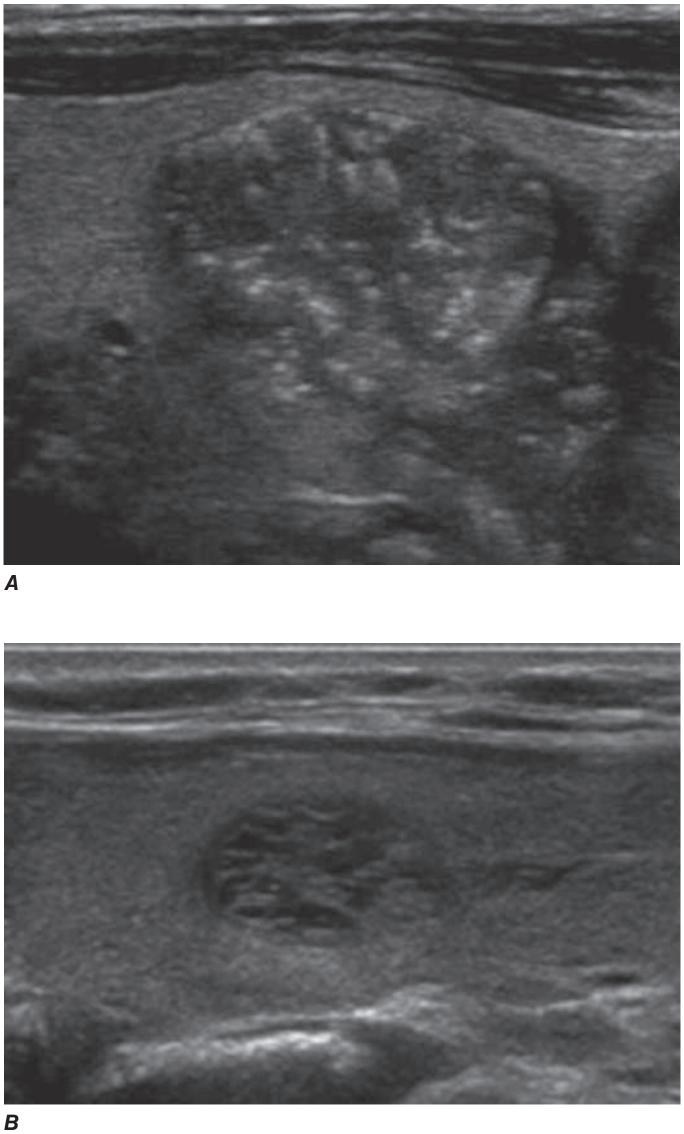

FIGURE 397-2 Sonographic patterns of thyroid nodules. ***A.*** High suspicion ultrasound pattern for thyroid malignancy ACR TI-RADS TR5 (hypoechoic solid nodule with irregular borders and punctate echogenic foci). ***B.*** Very low suspicion ultrasound pattern for thyroid malignancy ACR TI-RADS TR1 (spongiform nodule with microcystic areas comprising >50% of nodule volume). ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System.

In contrast to Graves' disease, spontaneous remission does not occur, and low-dose antithyroid drug therapy is well tolerated for years. Radioiodine is generally the treatment of choice; it treats areas of autonomy as well as decreasing the mass of the goiter by ablating the functioning nodules. Sometimes, however, a degree of autonomy may persist, presumably because multiple autonomous regions may emerge after others are treated, and further radioiodine treatment may be necessary. Surgery provides definitive treatment of underlying thyrotoxicosis as well as goiter. Patients should be rendered euthyroid using an antithyroid drug before operation.

■ HYPERFUNCTIONING SOLITARY NODULE

A solitary, autonomously functioning thyroid nodule is referred to as *toxic adenoma*. The pathogenesis of this disorder has been unraveled by demonstrating the functional effects of mutations that stimulate the TSH-R signaling pathway. Most patients with solitary hyperfunctioning nodules have acquired somatic, activating mutations in the TSH-R **(Fig. 397-4)**. These mutations, located primarily in the receptor transmembrane domain, induce constitutive receptor coupling to $G_s\alpha$, increasing cyclic adenosine monophosphate (AMP) levels and leading to enhanced thyroid follicular cell proliferation and function. Less commonly, somatic mutations are identified in $G_s\alpha$. These mutations, which are similar to those seen in McCune-Albright syndrome **(Chap. 424)** or in a subset of somatotrope adenomas **(Chap. 392)**, impair guanosine triphosphate (GTP) hydrolysis, causing constitutive activation of the cyclic AMP signaling pathway. In most series, activating mutations in either the TSH-R or the $G_s\alpha$ subunit genes are identified in >90% of patients with solitary hyperfunctioning nodules.

Thyrotoxicosis is usually mild and is generally only detected when a nodule is >3 cm. The disorder is suggested by a subnormal TSH level; the presence of the thyroid nodule, often large enough to be palpable; and the absence of clinical features suggestive of Graves' disease or other causes of thyrotoxicosis. A thyroid scan provides a definitive diagnostic test, demonstrating focal uptake in the hyperfunctioning nodule and diminished uptake in the remainder of the gland, as activity of the normal thyroid is suppressed.

TREATMENT

Hyperfunctioning Solitary Nodule

Radioiodine ablation is usually the treatment of choice. Because normal thyroid function is suppressed, ^{131}I is concentrated in the

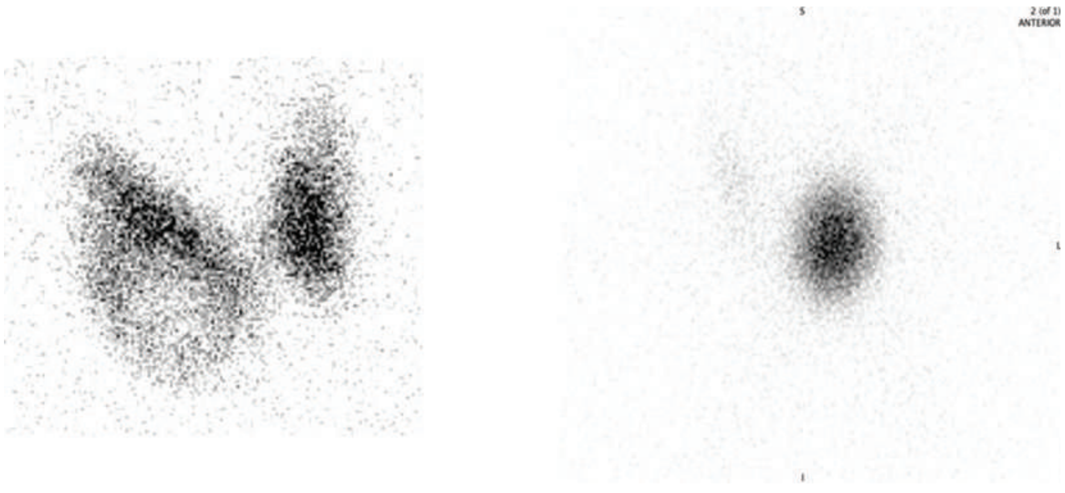

Panel ***A***

Panel ***B***

FIGURE 397-3 Scintigraphic scans of thyroid nodules. I^{123} scan (anterior view) of right lower pole nonfunctioning "cold" nodule (***A***) in a euthyroid patient and left hyperfunctioning "hot" nodule (***B***) causing thyrotoxicosis with suppression of I^{123} uptake in the extranodular thyroid. *(Courtesy of Dan Pryma.)*

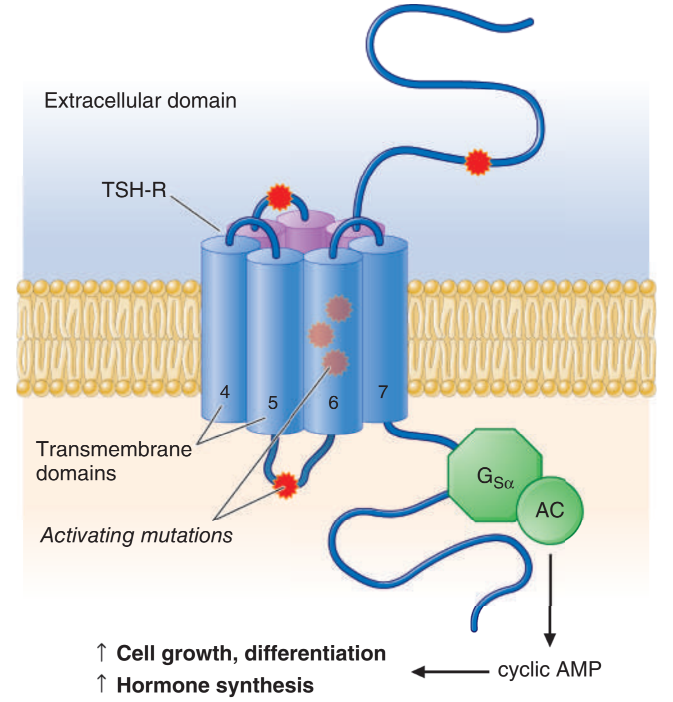

FIGURE 397-4 Activating mutations of the thyroid-stimulating hormone receptor (TSH-R). Mutations (*) that activate TSH-R reside mainly in transmembrane 5 and intracellular loop 3, although mutations have occurred in a variety of different locations. The effect of these mutations is to induce conformational changes that mimic TSH binding, thereby leading to coupling to stimulatory G protein ($G_{S\alpha}$) and activation of adenylate cyclase (AC), an enzyme that generates cyclic AMP.

hyperfunctioning nodule with minimal uptake and damage to normal thyroid tissue. Relatively large radioiodine doses (e.g., 370–1110 MBq [10–29.9 mCi] ^{131}I) have been shown to correct thyrotoxicosis in ~75% of patients within 3 months. Hypothyroidism occurs in <10% of those patients over the next 5 years. Surgical resection is also effective and is usually limited to lobectomy, thereby preserving thyroid function and minimizing risk of hypoparathyroidism or damage to the recurrent laryngeal nerves. Medical therapy using antithyroid drugs and beta blockers can normalize thyroid function but is not an optimal long-term treatment. Using ultrasound guidance, percutaneous radiofrequency ablation has been used successfully in some centers to ablate hyperfunctioning nodules, and this technique has also been used to reduce the size of nonfunctioning thyroid nodules.

TABLE 397-1 WHO Classification of Thyroid Neoplasms

Developmental abnormalities

1. Thyroglossal duct cyst
2. Other congenital thyroid abnormalities

Follicular cell–derived neoplasms

1. Benign tumors
    a. Thyroid follicular nodular disease
    b. Follicular adenoma
    c. Follicular adenoma with papillary architecture
    d. Oncocytic adenoma of the thyroid
2. Low-risk neoplasms
    a. Noninvasive follicular thyroid neoplasm with papillary-like nuclear features
    b. Thyroid tumors of uncertain malignant potential
    c. Hyalinizing trabecular tumor
3. Malignant neoplasms
    a. Follicular thyroid carcinoma
    b. Invasive encapsulated follicular variant papillary carcinoma
    c. Papillary thyroid carcinoma
    d. Oncocytic carcinoma of the thyroid
    e. Follicular-derived carcinomas, high-grade
        i. Differentiated high-grade thyroid carcinoma
        ii. Poorly differentiated thyroid carcinoma
    f. Anaplastic follicular cell–derived thyroid carcinoma

Thyroid C-cell–derived carcinoma

1. Medullary thyroid carcinoma

Mixed medullary and follicular cell–derived carcinomas

Salivary gland–type carcinomas of the thyroid

1. Mucoepidermoid carcinoma of the thyroid
2. Secretory carcinoma of salivary gland type

Thyroid tumors of uncertain histogenesis

1. Sclerosing mucoepidermoid carcinoma with eosinophilia
2. Cribriform morular thyroid carcinoma

Thymic tumors within the thyroid

1. Thymoma family
2. Spindle epithelial tumor with thymus-like elements
3. Thymic carcinoma family

Embryonal thyroid neoplasms

1. Thyroblastoma

Abbreviation: WHO, World Health Organization.

Source: Reprinted from International Agency for Research on Cancer, WHO Classification of Tumours, *https://whobluebooks.iarc.fr/structures/endocrine-and-neuroendocrine-tumours/*

BENIGN LESIONS

The various types of benign thyroid nodules are listed in **Table 397-1**. Benign nodules may be hyperplastic and reflect a combination of both macro- and microfollicular architecture, or they may be neoplastic, encapsulated adenomas that generally have a more monotonous microfollicular pattern. If the adenoma is composed of oncocytic follicular cells arranged in a follicular pattern, this is termed an oncocytic (formerly termed Hürthle cell) adenoma. Hyperplastic nodules generally appear as mixed cystic/solid or spongiform lesions on ultrasound. The definition of spongiform requires the presence of microcystic areas comprising >50% of the nodule volume, with the concept that this microcystic sonographic pattern recapitulates the histology of macrofollicles containing colloid **(Fig. 397-2*B*)**. However, the majority of solid nodules (whether hypo-, iso-, or hyperechoic) are also benign. FNA, usually performed with ultrasound guidance, is the diagnostic procedure of choice to evaluate thyroid nodules (see the "Approach to the Patient with Thyroid Nodules" section). Pure thyroid cysts, <1% of all thyroid growths, consist of colloid and are benign as well. Cysts frequently recur, even after repeated aspiration, and may require surgical excision if they are large. Ethanol ablation to sclerose the cyst has been used successfully for patients who are symptomatic.

TSH suppression with LT4 therapy does not decrease thyroid nodule size in iodine-sufficient populations. However, if there is relative iodine deficiency, both iodine and LT4 therapy have been demonstrated to decrease nodule volume. If LT4 is administered in this situation, the TSH should be maintained at or just below the lower limit of normal, but not frankly suppressed. If the nodule has not decreased in size after 6–12 months of therapy, treatment should be discontinued because little benefit is likely to accrue from long-term treatment; the risk of iatrogenic subclinical thyrotoxicosis should also be considered.

THYROID CANCER

Thyroid carcinoma is the most common malignancy of the endocrine system. Malignant tumors derived from the follicular epithelium are classified according to histologic features. Differentiated tumors, such as papillary thyroid cancer (PTC) or follicular thyroid cancer (FTC), are often curable, and the prognosis is good for patients identified with early-stage disease. In contrast, anaplastic thyroid cancer (ATC) is aggressive, responds poorly to treatment, and is associated with a bleak prognosis.

Over the past 30 years, the incidence of thyroid cancer has increased from 4.9 to >15 cases per 100,000 individuals in the United States. However, disease-specific mortality has only minimally increased. The increased incidence is predominantly attributable to small T1 papillary cancer tumors (<2 cm) and has led experts to consider that thyroid cancer is being overdiagnosed, suggesting that cancers are being detected that would otherwise be unlikely to harm a patient. The concept of cancer overdiagnosis is predicated upon the presence of a disease reservoir (the autopsy prevalence of PTC is ~25%), activities leading to disease detection (increased diagnostic imaging with incidental detection of nodules), and a mismatch in the directional rate between diagnosis and mortality (thyroid cancer disease-specific mortality not changed in 40 years). Similar trends have been observed worldwide, especially in countries with a higher proportion of privately financed health care, leading to increased resource utilization including imaging. The 20-year disease-specific mortality for low-risk thyroid cancer is 1%. Fortunately, epidemiologic data in the United States document a decrease in new thyroid cancer diagnoses (62,450 cases in 2015 and 43,720 cases in 2023), and this trend correlates with the implementation of evidence-based guidelines that recommend higher size thresholds for nodule FNA.

Current trends in thyroid cancer care focus on (1) avoiding overdiagnosis by limiting FNA by sonographic risk stratification with size cutoffs; (2) limiting surgery, radioiodine, and subsequent surveillance for low-risk tumors; and (3) identifying patients at higher recurrence risk for more aggressive treatment and monitoring. Prognosis is generally worse in older persons (>65 years). Thyroid cancer is twice as common in women as men, but male gender is associated with a worse prognosis. Additional important risk factors include a history of childhood (before age 18) head or neck irradiation, evidence for local tumor fixation or gross metastatic involvement of lymph nodes, and the presence of distant metastases (Table 397-2).

Several unique features of thyroid cancer facilitate its management: (1) thyroid nodules are amenable to biopsy by FNA; (2) iodine radioisotopes can be used to diagnose (^{123}I and ^{131}I) and potentially treat (^{131}I) differentiated thyroid cancer, reflecting the unique uptake of this anion by the thyroid gland; and (3) serum markers allow the detection of residual or recurrent disease, including the use of Tg levels for PTC FTC and oncocytic carcinoma and calcitonin for medullary thyroid cancer (MTC).

CLASSIFICATION

Thyroid neoplasms can arise in each of the cell types that populate the gland, including thyroid follicular cells, calcitonin-producing C cells, lymphocytes, and stromal and vascular elements, as well as metastases from other sites (Table 397-1). The American Joint Committee on Cancer (AJCC) staging system using the tumor, node, metastasis (TNM) classification is most commonly used. This system classifies different types of thyroid cancers (papillary, follicular, poorly differentiated, oncocytic cell, or anaplastic) based upon: 1) tumor size and extrathyroidal invasion; 2) regional lymph node metastases; and 3) distant metastases.[1]

TABLE 397-2 Risk Factors for Thyroid Carcinoma in Patients with Thyroid Nodule from History and Physical Examination

History of head and neck irradiation before the age of 18, including mantle radiation for Hodgkin's disease and brain radiation for childhood leukemia or other cranial malignancies Exposure to ionizing radiation from fallout in childhood or adolescence Age <20 or >65 years Rapidly enlarging neck mass Male gender ^{18}FDG or Ga-68 Dotatate PET avidity	Family history of papillary thyroid cancer in two or more first-degree relatives, MEN 2, or other genetic syndromes associated with thyroid malignancy (e.g., Cowden's syndrome, familial adenomatous polyposis, Carney complex, PTEN [phosphatase and tensin homolog] hamartoma tumor) Vocal cord paralysis, hoarse voice Nodule fixed to adjacent structures Lateral cervical lymphadenopathy (ipsilateral to the nodule)

Abbreviations: FDG, fluorodeoxyglucose; MEN, multiple endocrine neoplasia; PET, positron emission tomography.

PATHOGENESIS AND GENETIC BASIS

Radiation Early studies of the pathogenesis of thyroid cancer focused on the role of external radiation, which predisposes to chromosomal breaks, leading to genetic rearrangements and loss of tumor-suppressor genes. External radiation of the mediastinum, face, head, and neck region was administered in the past to treat an array of conditions, including acne and enlargement of the thymus, tonsils, and adenoids. Radiation exposure increases the risk of benign and malignant thyroid nodules, is associated with multicentric cancers, and shifts the incidence of thyroid cancer to an earlier age group. Radiation from nuclear fallout also increases the risk of thyroid cancer through absorption of radioactive iodine isotopes. Children seem more predisposed to the effects of radiation than adults.

TSH and Growth Factors Many differentiated thyroid cancers express TSH receptors and, therefore, remain responsive to TSH. Higher serum TSH levels, even within normal range, are associated with increased thyroid cancer risk in patients with thyroid nodules. These observations provide the rationale for T_4 suppression of TSH in patients with thyroid cancer. Residual expression of TSH receptors also allows TSH-stimulated uptake of ^{131}I therapy (see below).

Oncogenes and Tumor-Suppressor Genes Thyroid cancers are monoclonal in origin, consistent with the idea that they originate as a consequence of mutations that confer a growth advantage, or resistance to cell death, to a single progenitor cell. In addition to increased rates of proliferation, some thyroid cancers exhibit impaired apoptosis and features that enhance invasion, angiogenesis, and metastasis. Thyroid neoplasms have been analyzed for a variety of genetic alterations, but without clear evidence of an ordered acquisition of somatic mutations as they progress from the benign to the malignant state. On the other hand, certain mutations, such as *RET/PTC* and *PAX8-PPARγ1* rearrangements, are relatively specific for thyroid neoplasia.

As described above, activating mutations of the TSH-R and the $G_s\alpha$ subunit are associated with autonomously functioning nodules. Mutations in the enhancer of zeste homolog 1 (*EZH1*), which plays a role in chromatin remodeling, are also seen in about a quarter of autonomous nodules. Although these mutations induce thyroid cell growth, this type of nodule is almost always benign, likely because they drive differentiation pathways.

Activation of the RET-RAS-BRAF signaling pathway is seen in up to 70% of PTCs, although the types of mutations are heterogeneous. A variety of rearrangements involving the *RET* gene on chromosome 10 bring this receptor tyrosine kinase under the control of other promoters, leading to receptor overexpression. *RET* rearrangements occur in 15% of PTCs in different series and were observed with increased frequency in tumors developing after the Chernobyl radiation accident. Rearrangements in PTC have also occurred for another tyrosine kinase gene, *TRK1*, which is located on chromosome 1. *BRAF* V600E mutations are a common genetic alteration, occurring in up to 60% of PTC. These mutations activate the kinase, which stimulates the mitogen-activated protein kinase (MAPK) cascade. *RAS* mutations, which also stimulate the MAPK cascade, are found in ~15% of thyroid neoplasms (*NRAS* > *HRAS* > *KRAS*), including both PTC follicular variant and FTC. Of note, simultaneous *RET*, *BRAF*, and *RAS* mutations rarely occur in the same tumor, suggesting that activation of the MAPK cascade is critical for tumor development, independent of the step that initiates the cascade. Activation of the phosphatidylinositol 3-kinase (PI3K) pathway, often by deletion of the *PTEN* phosphatase gene, also occurs in both differentiated thyroid cancers.

---

[1] RM Tuttle et al: Updated American Joint Committee on Cancer/Tumor-Node-Metastasis Staging System for Differentiated and Anaplastic Thyroid Cancer (Eighth Edition): What changed and why? Thyroid 27:751, 2017.

BRAF and *RAS* mutations also occur in FTCs. In addition, a rearrangement of the thyroid developmental transcription factor PAX8 with the nuclear receptor PPARγ is identified in a significant fraction of FTCs, usually independent of *RAS* pathway mutations. Overall, ~70% of follicular cancers have mutations or genetic rearrangements. Loss of heterozygosity of 3p or 11q, consistent with deletions of tumor-suppressor genes, is also common in FTCs. Upregulation and downregulation of various microRNAs also influence gene expression patterns in both differentiated and dedifferentiated thyroid cancers.

Oncocytic tumors, which contain large numbers of mitochondria, are characterized by mutations in mitochondrial genes, primarily encoding enzymes in the electron transport chain. They also exhibit chromosomal copy number alterations and near haploidization.

Most of the mutations seen in differentiated thyroid cancers have also been detected in ATCs. *TERT* promoter mutations occur in <10% of differentiated PTCs but are more common in ATC. *BRAF* mutations are seen in up to 50% of ATCs. Mutations in *CTNNB1*, which encodes β-catenin, occur in about two-thirds of ATCs, but not in PTC or FTC. Mutations of the tumor-suppressor *P53* also play an important role in the development of ATC. Because *P53* plays a role in cell-cycle surveillance, DNA repair, and apoptosis, its loss may contribute to the rapid acquisition of genetic instability as well as poor treatment responses **(Chap. 77)**.

MTC, when associated with multiple endocrine neoplasia (MEN) type 2, harbors an inherited mutation of the *RET* gene. Unlike the rearrangements of *RET* seen in PTC, the mutations in MEN 2 are point mutations that induce constitutive activity of the tyrosine kinase **(Chap. 400)**. MTC is preceded by hyperplasia of the C cells, raising the likelihood that as-yet-unidentified “second hits” lead to cellular transformation. A subset of ~40% of sporadic MTC contains somatic mutations that activate *RET*.

Molecular diagnostics are being used more commonly in the clinical management of thyroid nodules, particularly to distinguish benign from malignant lesions after FNA and to reduce the number of diagnostic surgeries for indeterminate nodules. These tools are now offered as diagnostic panels by specialized referral laboratories. The panels also distinguish parathyroid and medullary nodules from thyroid follicular cell–derived lesions.

■ WELL-DIFFERENTIATED THYROID CANCER

Papillary PTC is the most common type of thyroid cancer, accounting for 80–85% of well-differentiated thyroid malignancies. Microscopic PTC is present in up to 25% of thyroid glands at autopsy, but most of these lesions are very small (several millimeters) and are not clinically significant. Characteristic cytologic features of PTC help make the diagnosis by FNA or after surgical resection; these include large, clear nuclei with powdery chromatin (described as an “Orphan Annie eye” appearance) with nuclear grooves and prominent nucleoli. The histologic finding of these cells arranged in either papillary structures or follicles distinguishes the classic and follicular variants of PTC, respectively. There are several subtypes of papillary thyroid cancer. The more differentiated classic and follicular variants are likely to have an indolent course in the absence of angioinvasion or metastatic adenopathy. The aggressive variants (tall cell, columnar cell, hobnail, poorly differentiated) require more intensive therapy and closer follow-up. Recently, a subtype previously known as the encapsulated PTC follicular variant, without capsular or angioinvasion, is no longer considered malignant and has been renamed noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP).

PTC may be multifocal and invade locally within the thyroid gland as well as through the thyroid capsule and into adjacent structures in the neck. It has a propensity to spread via the lymphatic system but can metastasize hematogenously as well, particularly to bone and lung. Because of the relatively slow growth of the tumor, a significant burden of pulmonary metastases may accumulate, sometimes with remarkably few symptoms. The prognostic implication of lymph node spread depends on the volume of metastatic disease. Micrometastases, defined as <2 mm of cancer in a lymph node, do not affect prognosis. However, gross metastatic involvement of multiple 2- to 3-cm lymph nodes indicates a 25–30% chance of recurrence and may increase mortality in older patients. Most papillary cancers are identified in the early stages (>95% stages I or II) and have an excellent prognosis, with survival curves similar to expected survival **(Fig. 397-5)**. Mortality is markedly increased in stage IV disease, especially in the presence of distant metastases (stage IVB), but this group comprises only about 1% of patients. The treatment of PTC is described below.

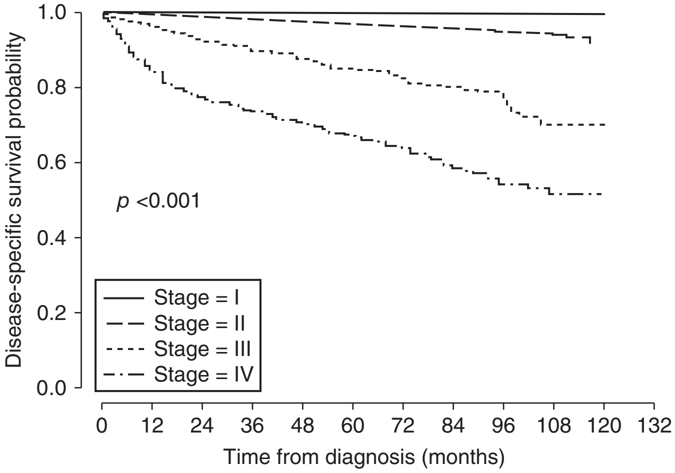

FIGURE 397-5 Unadjusted disease-specific survival curves for patients with papillary thyroid cancer in the American Joint Commission on Cancer/Union for International Cancer Control eighth edition TNM staging system. *(Reproduced with permission from LN Pontius et al: Projecting survival in papillary thyroid cancer: A comparison of the seventh and eighth editions of the American Joint Commission on Cancer/Union for International Cancer Control Staging Systems in two contemporary national patient cohorts. Thyroid 27:1408, 2017. The publisher for this copyrighted material is MaryAnn Liebert, Inc. publishers.)*

Follicular The incidence of FTC varies widely in different parts of the world; it is more common in iodine-deficient regions. Currently, FTC accounts for only about 5% of all thyroid cancers diagnosed in the United States. FTC is difficult to diagnose by FNA because the distinction between benign and malignant follicular neoplasms requires histology because the nuclear features of follicular adenomas and carcinomas do not differ. Rather, follicular carcinoma is diagnosed by the presence of capsular and/or vascular invasion. Follicular carcinomas with only capsular invasion have a very low risk of metastasis, and lobectomy alone suffices. Angioinvasive FTC is more aggressive and may metastasize to bone, lung, and the central nervous system. Mortality rates associated with angioinvasive FTC are less favorable than for PTC, in part because a larger proportion of patients present with stage IV disease. Poor prognostic features include distant metastases, age >55 years, primary tumor size >4 cm, and the presence of marked vascular invasion.

TREATMENT

Surgery for Well-Differentiated Thyroid Cancer

All well-differentiated thyroid cancers >1–1.5 cm (T1b or larger) should be surgically excised. Active surveillance is an option for small (up to 1.5 cm) intrathyroidal micropapillary thyroid cancers without cervical lymph node metastases, extrathyroidal extension, or location near the posterior thyroid capsule (near the recurrent laryngeal nerve). In addition to removing the primary lesion, surgery allows accurate histologic diagnosis and staging. Because there is no compelling evidence that bilateral thyroid surgery improves survival, the initial surgical procedure may be either a unilateral (lobectomy) or bilateral (near-total thyroidectomy) procedure for patients with intrathyroidal cancers >1 cm and <4 cm

(T1b and T2 tumors) in the absence of metastatic disease and after a careful sonographic evaluation for metastatic cervical adenopathy. For patients at high risk for recurrence, bilateral surgery allows administration of radioiodine for remnant ablation and potential treatment of iodine-avid metastases, if indicated, as well as for monitoring of serum Tg levels. Therefore, near-total thyroidectomy is appropriate in the presence of metastases or clinical evidence of extrathyroidal invasion and for most tumors >4 cm. In addition, for patients found to have a high-risk tumor after lobectomy based upon aggressive pathology features (e.g., vascular invasion or a less differentiated subtype), completion surgery should be performed to allow for effective radioiodine administration. Surgical complication rates are acceptably low if the surgeon is highly experienced in the procedure. Preoperative sonography should be performed in all patients to assess the central and lateral cervical lymph node compartments for suspicious adenopathy, which if present, should undergo FNA and be removed, as indicated, at surgery.

TSH SUPPRESSION THERAPY

Because most tumors are still TSH-responsive, LT4 suppression of TSH has been mainstay of thyroid cancer treatment. The degree of TSH suppression should be individualized based on a patient's risk of recurrence. It should be adjusted over time as surveillance blood tests and imaging confirm absence of disease or, alternatively, indicate possible residual/recurrent cancer. For patients at low risk of recurrence, TSH should be maintained in the lower normal limit (0.5–2.0 mIU/L). For patients either at intermediate or high risk of recurrence, TSH levels should be kept to 0.1–0.5 mIU/L and <0.1 mIU/L, respectively, if there are no strong contraindications to mild thyrotoxicosis. TSH should be <0.1 mIU/L for those with known metastatic disease.

RADIOIODINE TREATMENT

After near-total thyroidectomy, <1 g of thyroid tissue remains in the thyroid bed. Postsurgical radioablation of the remnant thyroid eliminates residual normal thyroid, facilitating the use of Tg determinations. In addition, well-differentiated thyroid cancer often incorporates radioiodine, although less efficiently than normal thyroid follicular cells. Radioiodine uptake is determined primarily by expression of the NIS and is stimulated by TSH, requiring expression of the TSH-R. The retention time for radioactivity is influenced by the extent to which the tumor retains differentiated functions such as iodide trapping and organification. Consequently, for patients at higher risk of recurrence and for those with known distant metastatic disease, ^{131}I therapy may provide an adjuvant role and potentially treat residual tumor cells.

Indications Not all patients benefit from radioiodine therapy. Neither recurrence nor survival rates are improved in stage I patients with T1 tumors (≤2 cm) confined to the thyroid. No benefit has been demonstrated for larger (>2 cm but <4 cm) low-risk tumors. However, in higher risk patients (larger tumors, more aggressive variants of papillary cancer, tumor vascular invasion, extrathyroidal invasion, presence of large-volume lymph node metastases), radioiodine may reduce recurrence and may increase survival for older patients.

^{131}I Thyroid Ablation and Treatment As noted above, the decision to use ^{131}I for thyroid ablation should be coordinated with the surgical approach, because radioablation is much more effective when there is minimal remaining normal thyroid tissue. Radioiodine is administered after iodine depletion (patient follows a low-iodine diet for 1–2 weeks) and in the presence of elevated serum TSH levels to stimulate uptake of the isotope into both the remnant and potentially any residual tumor. To achieve high serum TSH levels, there are two approaches. A patient may be withdrawn from thyroid hormone so that endogenous TSH is secreted and, ideally, the serum TSH level is >25 mIU/L at the time of ^{131}I therapy. A typical strategy is to treat the patient for several weeks postoperatively with liothyronine (25 μg qd or bid), followed by thyroid hormone withdrawal for 2 weeks. Alternatively, recombinant human TSH (rhTSH) is administered as two daily consecutive injections (0.9 mg) with administration of ^{131}I 24 h after the second injection. The patient can continue to take LT4 and remains euthyroid. Both approaches have equal success in achieving remnant ablation.

A pretreatment scanning dose of ^{131}I (usually 111 MBq [3 mCi]) or ^{123}I (74 MBq [2 mCi]) can reveal the amount of residual tissue and provides guidance about the dose needed to accomplish ablation. However, because of concerns about radioactive "stunning" that impairs subsequent treatment, there is a trend to avoid pretreatment scanning with ^{131}I and use either ^{123}I or proceed directly to ablation, unless there is suspicion that the amount of residual tissue will alter therapy or that there is distant metastatic disease. In the United States, outpatient doses of up to 6475 MBq (175 mCi) can be given at most centers. The administered dose depends on the indication for therapy, with lower doses of 1100 MBq (30 mCi) given for remnant ablation but higher doses of up to 5500 MBq (150 mCi) reserved for use as adjuvant therapy when residual disease is suspected or present. Whole-body scanning (WBS) following radioiodine treatment is used to confirm the ^{131}I uptake in the remnant and to identify possible metastatic disease.

Surveillance Testing Serum thyroglobulin (Tg) is a sensitive marker of residual/recurrent thyroid cancer after ablation of the residual postsurgical thyroid tissue. Current Tg assays have functional sensitivities as low as 0.1 ng/mL, as opposed to older assays with functional sensitivities of 1–2 ng/mL, reducing the number of patients with truly undetectable serum Tg levels. Because the vast majority of PTC recurrences are in cervical lymph nodes, a neck ultrasound should be performed about 6 months after thyroid ablation; ultrasound has been shown to be more sensitive than WBS in this scenario.

In low-risk patients who have no clinical evidence of residual disease after ablation, negative cervical sonography, and a basal Tg <0.2 ng/mL on LT4, the risk of structural recurrence is <3% at 5 years, and the frequency of follow-up testing can be decreased to annual TSH and Tg testing, with only periodic ultrasound examination. The serum TSH should be maintained in the lower half of the normal range.

The use of WBS for thyroid cancer surveillance is reserved for patients with known iodine-avid metastases or those with elevated serum thyroglobulin levels and negative imaging with ultrasound, chest CT, neck cross-sectional imaging, and positron emission tomography (PET) CT who may require additional ^{131}I therapy.

In addition to radioiodine, external beam radiotherapy is also used to treat gross residual neck disease or specific metastatic lesions, particularly when they cause bone pain or threaten neurologic injury (e.g., vertebral metastases).

New Potential Therapies Kinase inhibitors target pathways known to be active in thyroid cancer, including the RAS, BRAF, RET, EGFR, VEGFR, and angiogenesis pathways. Treatment has been shown to stabilize progressive metastatic disease that is refractory to radioiodine therapy, although only one study has demonstrated improved survival. Given the significant associated toxicities and the need for ongoing therapy, patient selection is critical to limit systemic therapy to those with significant morbidity risk. The American Thyroid Association guidelines recommend active surveillance for asymptomatic patients with metastatic tumors between 1 and 2 cm and then intervention as the rate of tumor growth increases. In addition, based on genetic analyses of metastases, mutation-selective kinase inhibitors are now being used. In addition to multikinase inhibitors, new therapies target tumor-specific mutational changes including *BRAF* V600E point mutations and *NTRK* and *RET* gene fusions. Immune checkpoint inhibitors have also shown efficacy for some thyroid cancers with high tumor mutation burden (TMB-H). Ongoing trials are also exploring whether differentiation protocols, targeting the MAPK pathway, might enhance radioiodine uptake and efficacy.

ANAPLASTIC AND OTHER FORMS OF THYROID CANCER

Anaplastic Thyroid Cancer As noted above, ATC is a poorly differentiated and aggressive cancer. The prognosis is poor, and most patients die within 6 months of diagnosis. Because of the undifferentiated state of these tumors, the uptake of radioiodine is usually negligible, but it can be used therapeutically if there is residual uptake. Chemotherapy has been attempted with multiple agents, including anthracyclines and paclitaxel, but it is usually ineffective. External beam radiation therapy can be attempted and continued if tumors are responsive. Both multitargeted and mutation-directed kinase inhibitors are in clinical trials and may prolong survival by a few months.

Thyroid Lymphoma Lymphoma in the thyroid gland often arises in the background of Hashimoto's thyroiditis. A rapidly expanding thyroid mass suggests the possibility of this diagnosis. Diffuse large-cell lymphoma is the most common type in the thyroid. Biopsies reveal sheets of lymphoid cells that can be difficult to distinguish from small-cell lung cancer or ATC. These tumors are often highly sensitive to external radiation. Surgical resection should be avoided as initial therapy because it may spread disease that is otherwise localized to the thyroid. If staging indicates disease outside of the thyroid, treatment should follow guidelines used for other forms of lymphoma **(Chap. 113)**.

MEDULLARY THYROID CARCINOMA

MTC can be sporadic or familial and accounts for ~5% of thyroid cancers. There are three familial forms of MTC: MEN 2A, MEN 2B, and familial MTC without other features of MEN **(Chap. 400)**. In general, MTC is more aggressive in MEN 2B than in MEN 2A, and familial MTC is more aggressive than sporadic MTC. Elevated serum calcitonin provides a marker of residual or recurrent disease. All patients with MTC should be tested for *RET* mutations, because genetic counseling and testing of family members can be offered to those individuals who test positive for mutations.

The management of MTC is primarily surgical. Prior to surgery, pheochromocytoma should be excluded in all patients with a *RET* mutation. Unlike tumors derived from thyroid follicular cells, these tumors do not take up radioiodine. External radiation treatment and targeted kinase inhibitors may provide palliation in patients with advanced disease **(Chap. 400)**.

APPROACH TO THE PATIENT

Thyroid Nodules

Palpable thyroid nodules are found in ~5% of adults, but the prevalence varies considerably worldwide. Given this high prevalence rate, practitioners may identify thyroid nodules on physical examination. However, the increased usage of diagnostic medical imaging (e.g., carotid ultrasound, cervical spine MRI) has led to an increased frequency of incidental nodule detection, accounting for the majority of patients currently presenting for nodule evaluation. The main goal of this evaluation is to identify, in a cost-effective manner, the small subgroup of individuals with malignant lesions that have the potential to be clinically significant.

Nodules are more common in iodine-deficient areas, in women, and with aging. Most palpable nodules are >1 cm in diameter, but the ability to feel a nodule is influenced by its location within the gland (superficial vs deeply embedded), the anatomy of the patient's neck, and the experience of the examiner. More sensitive methods of detection, such as CT, thyroid ultrasound, and pathologic studies, reveal thyroid nodules in up to 50% of glands in individuals aged >50 years. The presence of these thyroid incidentalomas has led to much debate about how to detect nodules and which nodules to investigate further.

An approach to the evaluation of thyroid nodules detected by either palpation or imaging is outlined in **Fig. 397-6**. Most patients with thyroid nodules have normal thyroid function tests. Nonetheless, thyroid function should be assessed by measuring a TSH level, which may be suppressed by one or more autonomously functioning nodules. If the TSH is suppressed, a radionuclide scan is indicated to determine if the identified nodule is "hot," as lesions with increased uptake are almost never malignant and FNA is unnecessary **(Fig. 397-3*B*)**. Otherwise, the next step in evaluation is performance of a thyroid ultrasound for three reasons: (1) For nodules detected on physical examination, ultrasound will confirm if the palpable nodule is indeed a nodule. About 15% of "palpable" nodules are not confirmed on imaging, and therefore no further evaluation is required. (2) Ultrasound will assess if there are additional nonpalpable nodules for which FNA may be recommended based on imaging features and size. (3) Ultrasound will characterize the imaging pattern of the nodule, which, combined with the nodule's size, facilitates decision-making about FNA. There are several validated risk stratification systems (RSS) for sonographic imaging of thyroid nodules (American College of Radiology [ACR] Thyroid Imaging Reporting and Data System [TI-RADS], American Thyroid Association, European Thyroid Association [EU-TIRADS], among others). These demonstrate consistent risk estimates for thyroid cancer based on certain sonographic patterns. All provide size cutoff recommendations for nodule FNA based on sonographic patterns, with lower size cutoffs for nodules with more suspicious ultrasound patterns, but the specific size cutoff criteria differ among the RSS. Not surprisingly, the RSSs with lower size cutoffs have higher sensitivity and lower specificity for thyroid cancer diagnosis than those with higher cutoffs. Nevertheless, all have been shown to reduce unnecessary FNAs by at least 45%, in part due to the recommendation not to perform FNA for spongiform nodules. ACR TI-RADS is currently the most widely used RSS in the United States, and nodules are classified from TR1 to TR5 (Fig. 397-1).

For example, a spongiform nodule (TR1, Fig. 397-2B) has a <3% chance of cancer, and observation rather than FNA is generally recommended by all RSSs, whereas 10–20% of solid hypoechoic nodules with smooth borders (TR4) are malignant and FNA is recommended at size cutoffs ranging from 1 to 1.5 cm. All the RSSs recommend FNA at 1 cm for the highest suspicion pattern nodule, TR5 (Figs. 397-1 and 397-2A). Given what is known about the prevalence and generally indolent behavior of small thyroid cancers <1 cm, none of the RSSs recommend routine FNA for any nodule <1 cm unless metastatic cervical lymph nodes are present.

FNA biopsy, ideally performed with ultrasound guidance, is the best diagnostic test when performed by physicians familiar with the procedure and when the results are interpreted by experienced cytopathologists. The technique is particularly useful for detecting PTC. However, the distinction between benign and malignant follicular patterned lesions is often not possible using cytology alone because of the absence of characteristic nuclear features in follicular carcinoma. Using the current ultrasound RSS for FNA decision-making, FNA biopsies yield the following spectrum of cytology diagnoses: 50–60% benign, 5% malignant or suspicious for malignancy, 5–7% nondiagnostic or yielding insufficient material for diagnosis, and 25–40% indeterminate. The Bethesda System is now widely used to provide more uniform terminology for reporting thyroid nodule FNA cytology results. This six-tiered classification system (newest version 3) with the respective estimated malignancy rates is shown in **Table 397-3**. Importantly, because NIFTP can only be diagnosed by surgical pathology, NIFTP is included in the malignancy estimates. Specifically, the Bethesda System subcategorized cytology specimens previously labeled as indeterminate into three categories: atypia undetermined significance (AUS), follicular neoplasm (FN), and suspicious for malignancy (SFM).

Cytology results indicative of malignancy generally mandate surgery, after performing preoperative sonography to evaluate the cervical lymph nodes. Nondiagnostic cytology specimens most often result from cystic lesions but may also occur in fibrous longstanding nodules or very vascular nodules where a longer needle

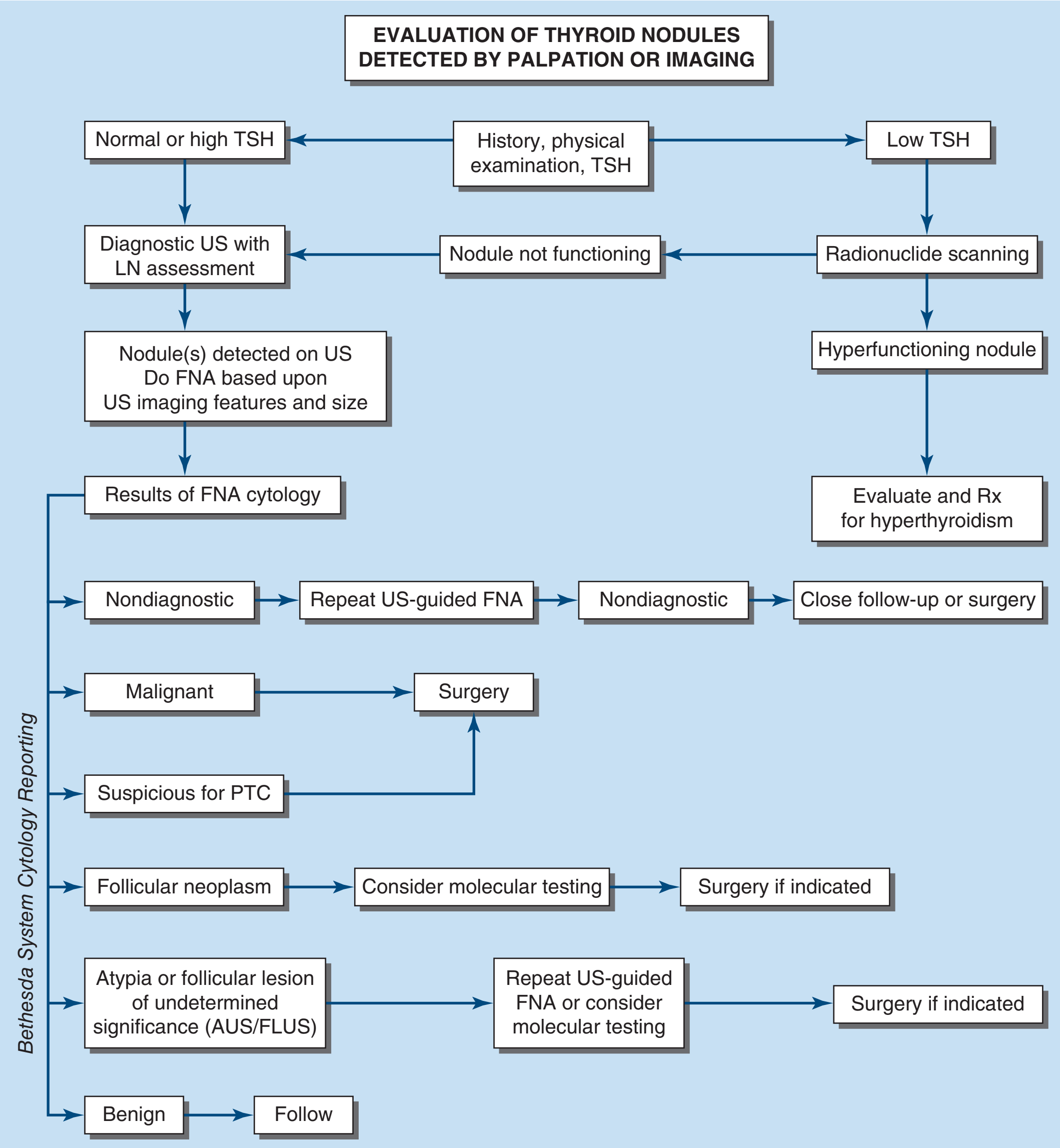
EVALUATION OF THYROID NODULES DETECTED BY PALPATION OR IMAGING
Normal or high TSH
History, physical examination, TSH
Low TSH
Diagnostic US with LN assessment
Nodule not functioning
Radionuclide scanning
Nodule(s) detected on US Do FNA based upon US imaging features and size
Hyperfunctioning nodule
Results of FNA cytology
Evaluate and Rx for hyperthyroidism
Bethesda System Cytology Reporting
Nondiagnostic
Repeat US-guided FNA
Nondiagnostic
Close follow-up or surgery
Malignant
Surgery
Suspicious for PTC
Follicular neoplasm
Consider molecular testing
Surgery if indicated
Atypia or follicular lesion of undetermined significance (AUS/FLUS)
Repeat US-guided FNA or consider molecular testing
Surgery if indicated
Benign
Follow

FIGURE 397-6 Approach to the patient with a thyroid nodule. See text and references for details. FNA, fine-needle aspiration; LN, lymph node; PTC, papillary thyroid cancer; Rx, therapy; TSH, thyroid-stimulating hormone; US, ultrasound.

dwell time may result in a hemorrhagic specimen. Ultrasound-guided FNA is indicated when a repeat FNA is necessary. Repeat FNA will yield a diagnostic cytology in ~50% of cases. Given the low false-negative rate of a benign cytology (<3%), benign nodules with a lower suspicion sonographic pattern (TR2, TR3, TR4) can be followed. Those with more worrisome ultrasound features, especially TR5 nodules, should undergo repeat FNA because of a higher likelihood of a missed malignancy. The use of LT4 to suppress serum TSH is not effective in shrinking nodules in iodine-replete populations, and therefore, LT4 suppression should not be used. The three indeterminate cytology classifications introduced by the Bethesda System are associated with different risks of malignancy (Table 397-3). For nodules with suspicious for malignancy cytology, surgery is recommended after ultrasound assessment of cervical lymph nodes. Options to be discussed with the patient include lobectomy versus total thyroidectomy.

TABLE 397-3 Bethesda Classification for Thyroid Cytology Version 3

DIAGNOSTIC CATEGORY	RISK OF MALIGNANCY (INCLUDING NIFTP) MEAN % (RANGE)
I. Nondiagnostic or unsatisfactory	13 (5–20)
II. Benign	4 (2–7)
III. Atypia of unknown significance (AUS)	22 (13–30)
IV. Follicular neoplasm (FN)	30 (23–34)
V. Suspicious for malignancy (SFM)	74 (67–83)
VI. Malignant	97 (97–100)

Abbreviation: NIFTP, noninvasive follicular thyroid neoplasm with papillary-like nuclear features.

On the other hand, the majority of nodules with AUS and FN cytology results are benign; the range of malignancy (ROM) varies from 13 to 34%. The traditional approach for these patients is diagnostic lobectomy for histopathologic diagnosis. Therefore, many patients undergo surgery for benign nodules. Over the past decade, the uncertainty about the ROM for indeterminate cytology nodules has been the driver for the development of molecular testing, which can better differentiate benign from malignant nodules. Based on results from next-generation sequencing, which includes point mutations, small insertions/deletions, and gene fusions, as well as results from microRNA analyses and gene expression, the current validated and commercially available molecular tests combine these techniques with the following two goals: (1) risk stratification of thyroid nodules based on a positive result; and (2) reduction in cancer risk to an acceptable level for nonsurgical surveillance based on a negative result. Assuming a 25–30% ROM for nodules with indeterminate cytology, the negative predictive values for the currently validated molecular tests are >95%.

The evaluation of a thyroid nodule is stressful for most patients. They are concerned about the possibility of thyroid cancer, whether verbalized or not. It is constructive, therefore, to review the diagnostic approach and to reassure patients when no malignancy is found. When a suspicious lesion or thyroid cancer is identified, the generally favorable prognosis and available treatment options can be reassuring.

FURTHER READING

Ali SZ et al: The 2023 Bethesda system for reporting thyroid cytopathology. Thyroid 33:1039, 2023.

Baloch ZW et al: Overview of the 2022 WHO classification of thyroid neoplasms. Endocr Pathol 33:27, 2022.

Davies L, Hoang J: Thyroid cancer in the USA: Current trends and outstanding questions. Lancet Diab Endocrinol 9:11, 2021.

Dunn LA et al: Vemurafenib redifferentiation of BRAF mutant, RAI-refractory thyroid cancers. J Clin Endocrinol Metab 104:1417, 2019.

Durante C et al: The diagnosis and management of thyroid nodules: A review. JAMA 319:914, 2018.

Fagin JA, Nikirov YE: Progress in thyroid cancer genomics: A 40-year journey. Thyroid 33:1271, 2023.

Haugen BR et al: 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. Thyroid 26:1, 2016.

Tessler FN et al: ACR Thyroid Imaging Reporting and Data System (TI-RADS): White paper of the ACR TI-RADS committee. J Am Coll Radiol 14:587, 2017.

Tuttle RM et al: Updated American Joint Committee and Cancer/Tumor-Node-Metastasis Staging System for differentiated and anaplastic thyroid cancer (eighth edition): What changed and why? Thyroid 27:751, 2017.

398

Disorders of the Adrenal Cortex

Wiebke Arlt, Alessandro Prete

The adrenal cortex produces three classes of corticosteroid hormones: glucocorticoids (e.g., cortisol), mineralocorticoids (e.g., aldosterone), and adrenal androgen precursors (e.g., dehydroepiandrosterone [DHEA]) **(Fig. 398-1)**. Glucocorticoids and mineralocorticoids act through specific nuclear receptors, regulating aspects of the physiologic stress response as well as blood pressure and electrolyte homeostasis. Adrenal androgen precursors are converted in the gonads and peripheral target cells to sex steroids that act via nuclear androgen and estrogen receptors.

Disorders of the adrenal cortex are characterized by deficiency or excess of one or several of the three major corticosteroid classes. Hormone deficiency can be caused by inherited glandular or enzymatic disorders or by destruction of the pituitary or adrenal gland by autoimmune disorders, infection, infarction, or iatrogenic events such as surgery or hormonal suppression. Hormone excess is usually the result of neoplasia, leading to increased production of adrenocorticotropic hormone (ACTH) by the pituitary or neuroendocrine ectopic ACTH-producing cells or increased production of glucocorticoids, mineralocorticoids, or adrenal androgen precursors by adrenal nodules or hyperplasia. Adrenal nodules are increasingly identified incidentally during cross-sectional imaging of the chest or abdomen performed for other reasons.

ADRENAL ANATOMY AND DEVELOPMENT

The normal adrenal glands weigh 6–11 g each. They are located above the kidneys and have their own blood supply. Arterial blood flows initially to the subcapsular region and then meanders from the outer cortical zona glomerulosa through the intermediate zona fasciculata to the inner zona reticularis and eventually to the adrenal medulla. The right adrenal (or suprarenal) vein drains directly into the vena cava, while the left adrenal vein drains into the left renal vein.

During early embryonic development, the adrenals originate from the urogenital ridge and then separate from gonads and kidneys at about the sixth week of gestation. Concordant with the time of sexual differentiation (seventh to ninth week of gestation, **Chap. 402**), the adrenal cortex starts to produce cortisol and the adrenal sex steroid precursor DHEA. The orphan nuclear receptors SF1 (steroidogenic factor 1; encoded by the gene *NR5A1*) and DAX1 (dosage-sensitive sex reversal gene 1; encoded by the gene *NR0B1*), among others, play a crucial role during this period of development, as they regulate a multitude of adrenal genes involved in steroidogenesis.

REGULATORY CONTROL OF STEROIDOGENESIS

Production of glucocorticoids and adrenal androgens is under the control of the hypothalamic-pituitary-adrenal (HPA) axis, whereas mineralocorticoids are regulated by the renin-angiotensin-aldosterone (RAA) system.

Glucocorticoid synthesis is under inhibitory feedback control by the hypothalamus and the pituitary **(Fig. 398-2)**. Hypothalamic release of corticotropin-releasing hormone (CRH) occurs in response to endogenous or exogenous stress. CRH stimulates the cleavage of the 241–amino acid polypeptide proopiomelanocortin (POMC) by pituitary-specific prohormone convertase 1 (PC1), yielding the 39–amino acid peptide ACTH. ACTH is released by the corticotrope cells of the anterior pituitary and acts as the pivotal regulator of adrenal cortisol synthesis, with additional short-term effects on mineralocorticoid and adrenal androgen synthesis. The release of CRH, and subsequently ACTH, occurs in a pulsatile fashion that follows a circadian rhythm under the control of the hypothalamus, specifically its suprachiasmatic nucleus (SCN), with additional regulation by a complex network of cell-specific clock genes. The circadian release of ACTH is mostly regulated by the CRH, but arginine-vasopressin (AVP) also exerts a secretagogue role. Reflecting the pattern of ACTH secretion, adrenal cortisol secretion exhibits a distinct circadian rhythm, starting to rise in the early morning hours prior to awakening, with peak levels in the morning and low levels in the evening **(Fig. 398-3)**.

Diagnostic tests assessing the HPA axis make use of the fact that it is regulated by negative feedback. Glucocorticoid excess is diagnosed by employing a dexamethasone suppression test. Dexamethasone, a potent synthetic glucocorticoid, suppresses CRH/ACTH by binding hypothalamic-pituitary glucocorticoid receptors (GRs) and, therefore, results in downregulation of endogenous cortisol synthesis. Various versions of the dexamethasone suppression test are described in detail in **Chap. 392**. If cortisol production is autonomous (e.g., adrenal nodule), ACTH is already suppressed, and dexamethasone has little additional effect. If cortisol production is driven by an ACTH-producing pituitary adenoma, dexamethasone suppression is ineffective at lower doses but usually induces suppression at higher doses. If cortisol production is driven by an ectopic source of ACTH, the tumors are usually resistant to dexamethasone suppression. Thus, the dexamethasone suppression test is useful to establish the diagnosis of Cushing's syndrome and assist with the differential diagnosis of cortisol excess.

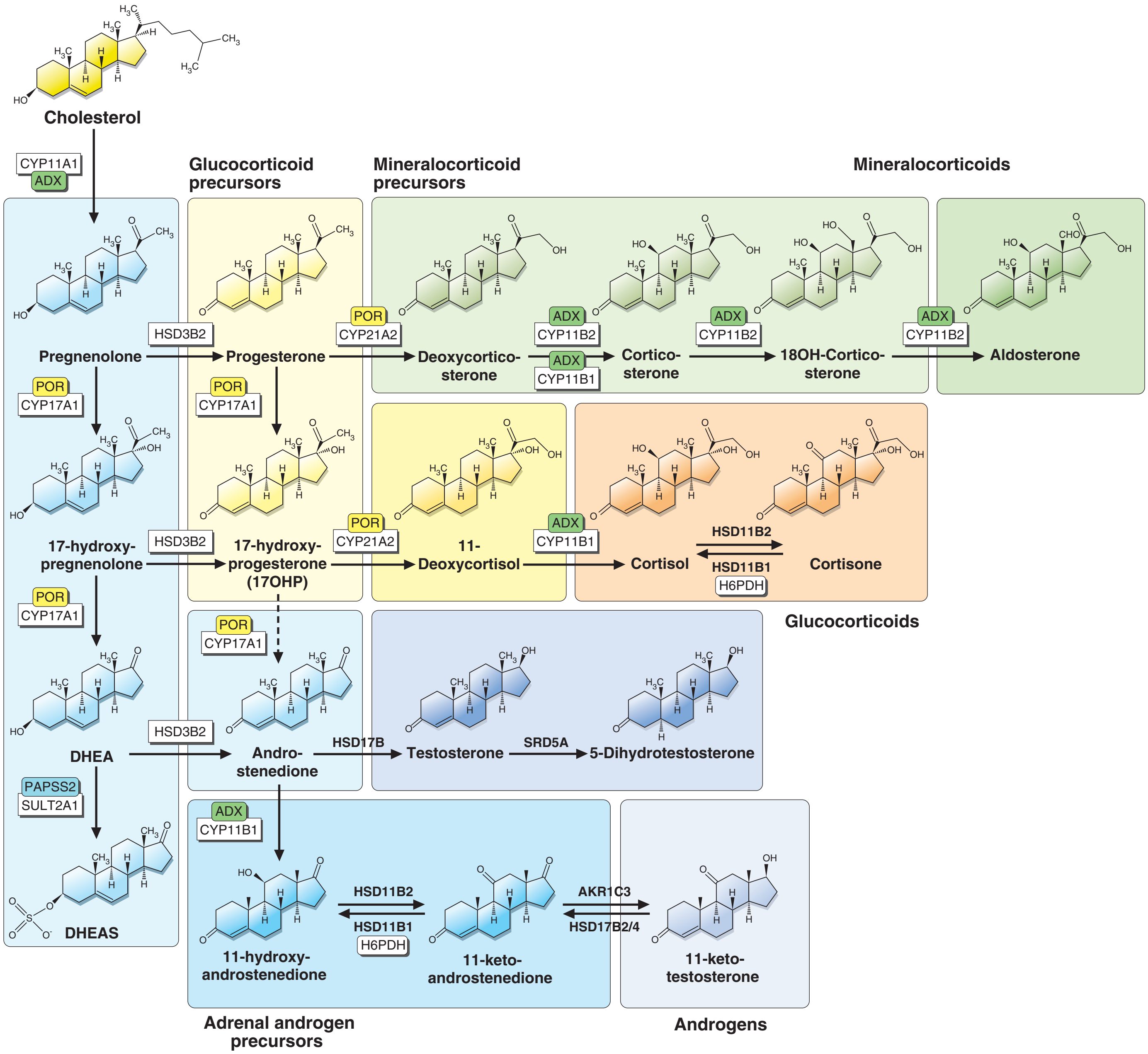
Cholesterol
CYP11A1
ADX
Glucocorticoid precursors
Mineralocorticoid precursors
Mineralocorticoids
Pregnenolone
HSD3B2
Progesterone
POR
CYP21A2
Deoxycorticosterone
ADX
CYP11B1
Corticosterone
ADX
CYP11B2
18OH-Corticosterone
ADX
CYP11B2
Aldosterone
POR
CYP17A1
POR
CYP17A1
17-hydroxy-pregnenolone
HSD3B2
17-hydroxy-progesterone (17OHP)
POR
CYP21A2
11-Deoxycortisol
ADX
CYP11B1
Cortisol
HSD11B2
HSD11B1
H6PDH
Cortisone
Glucocorticoids
POR
CYP17A1
POR
CYP17A1
DHEA
HSD3B2
Androstenedione
HSD17B
Testosterone
SRD5A
5-Dihydrotestosterone
PAPSS2
SULT2A1
DHEAS
ADX
CYP11B1
11-hydroxy-androstenedione
HSD11B2
HSD11B1
H6PDH
11-keto-androstenedione
AKR1C3
HSD17B2/4
11-keto-testosterone
Adrenal androgen precursors
Androgens

FIGURE 398-1 Adrenal steroidogenesis. ADX, adrenodoxin; AKR1C3, aldo-keto reductase 1C3; CYP11A1, side chain cleavage enzyme; CYP11B1, 11β-hydroxylase; CYP11B2, aldosterone synthase; CYP17A1, 17α-hydroxylase/17,20 lyase; CYP21A2, 21-hydroxylase; DHEA, dehydroepiandrosterone; DHEAS, dehydroepiandrosterone sulfate; H6PDH, hexose-6-phosphate dehydrogenase; HSD11B1, 11β-hydroxysteroid dehydrogenase type 1; HSD11B2, 11β-hydroxysteroid dehydrogenase type 2; HSD17B, 17β-hydroxysteroid dehydrogenase; HSD3B2, 3β-hydroxysteroid dehydrogenase type 2; PAPSS2, PAPS synthase type 2; POR, P450 oxidoreductase; SRD5A, 5α-reductase; SULT2A1, DHEA sulfotransferase.

Conversely, to assess glucocorticoid deficiency, ACTH stimulation of cortisol production is used. The ACTH peptide contains 39 amino acids, but the first 24 are sufficient to elicit a physiologic response. The standard ACTH stimulation test involves the administration of cosyntropin (ACTH 1-24), 250 μg IM or IV, and collection of blood samples at 0, 30, and optionally 60 min for cortisol. A normal response is defined as a cortisol level >15–20 μg/dL (>400–550 nmol/L) 30–60 min after cosyntropin stimulation, with the precise cutoff dependent on the assay used. A low-dose (1 μg cosyntropin IV) version of this test has been advocated; however, it has no superior diagnostic value and is more cumbersome to carry out due to the lack of commercially available preparations of 1 μg cosyntropin. Alternatively, an insulin tolerance test (ITT) can be used to assess adrenal function. It involves injection of insulin to induce hypoglycemia, which represents a strong stress signal that triggers hypothalamic CRH release and activation of the entire HPA axis. The ITT involves administration of regular insulin 0.1 U/kg IV (dose should be lower if hypopituitarism is likely) and collection of blood samples at 0, 30, 60, and 120 min for glucose, cortisol, and growth hormone (GH), if also assessing the GH axis. Oral or IV glucose is administered after the patient has achieved symptomatic hypoglycemia (usually plasma glucose <40 mg/dL). A normal response is defined as a cortisol >20 μg/dL and GH >5.1 μg/L, again with assay-specific cutoff variability. The ITT requires careful clinical monitoring and sequential measurements of glucose. It is contraindicated in patients with coronary disease, cerebrovascular disease, or seizure disorders, which has made the cosyntropin test the commonly accepted first-line test. The overnight metyrapone test is alternatively used in some centers: metyrapone—a drug blocking the conversion of 11-deoxycortisol to cortisol by 11β-hydroxylase (see treatment of Cushing's syndrome)—is administered orally at midnight (2500 mg or 30 mg/kg). In healthy individuals, the decrease in cortisol stimulates CRH and ACTH production, resulting in an increase in 11-deoxycortisol. A normal response consists of an early morning serum 11-deoxycortisol concentration >287 nmol/L (10 μg/dL). The use of home-waking salivary cortisone as a measure of adrenal cortisol reserve looks highly promising and represents a potential alternative noninvasive test that can be undertaken at home.

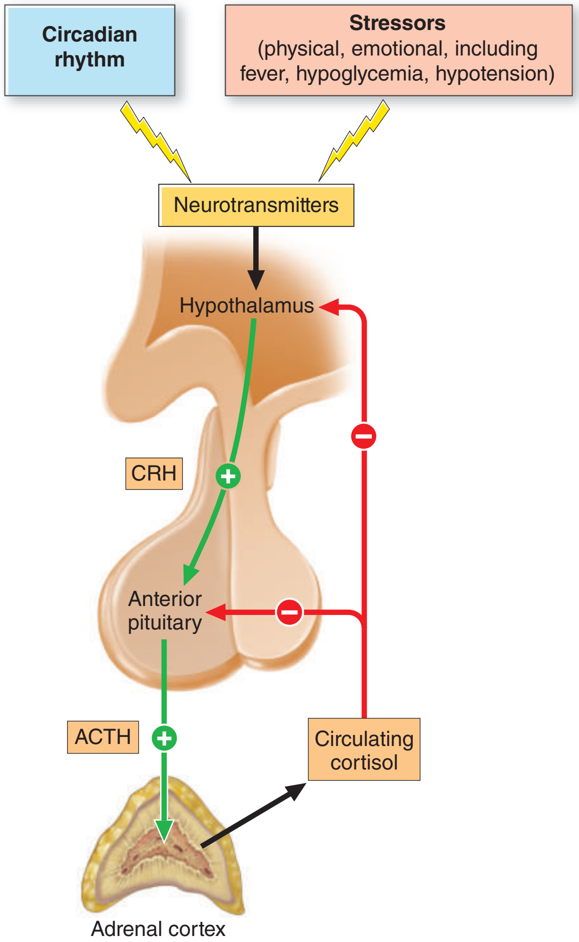

FIGURE 398-2 Regulation of the hypothalamic-pituitary-adrenal (HPA) axis. ACTH, adrenocorticotropic hormone; CRH, corticotropin-releasing hormone.

Mineralocorticoid production is controlled by the RAA regulatory cycle, which is initiated by the release of the enzyme renin from the juxtaglomerular cells in the kidney, resulting in cleavage of hepatic angiotensinogen to angiotensin I **(Fig. 398-4)**. The angiotensin-converting enzyme (ACE) cleaves angiotensin I to angiotensin II, which binds and activates the angiotensin II receptor type 1 (AT1 receptor [AT1R]), resulting in increased adrenal aldosterone production and vasoconstriction. Aldosterone enhances renal sodium retention and potassium excretion resulting in volume expansion and increased renal perfusion, which in turn regulates renin release. Because mineralocorticoid synthesis is primarily under the control of the RAA system, hypothalamic-pituitary damage does not significantly impact the capacity of the adrenal to synthesize aldosterone.

Similar to the HPA axis, the assessment of the RAA system can be used for diagnostic purposes. If mineralocorticoid excess is present, there is a counterregulatory downregulation of plasma renin (see below for testing). Conversely, in mineralocorticoid deficiency, plasma renin is markedly increased. Physiologically, oral or IV sodium loading results in suppression of aldosterone, a response that is attenuated or absent in patients with autonomous mineralocorticoid excess.

■ STEROID HORMONE BIOSYNTHESIS, METABOLISM, AND ACTION

ACTH stimulation is required for the initiation of steroidogenesis. The ACTH receptor MC2R (melanocortin 2 receptor) interacts with the MC2R-accessory protein MRAP, and the complex is transported to the adrenocortical cell membrane, where it binds to ACTH **(Fig. 398-5)**. ACTH stimulation generates cyclic AMP (cAMP), which upregulates the protein kinase A (PKA) signaling pathway. Inactive PKA is a tetramer of two regulatory and two catalytic subunits that is dissociated by cAMP into a dimer of two regulatory subunits bound to cAMP and two free and active catalytic subunits. PKA activation impacts steroidogenesis in three distinct ways: (1) increases the import of cholesterol esters; (2) increases the activity of hormone-sensitive lipase, which cleaves cholesterol esters to cholesterol for import into the mitochondrion; and (3) increases the availability and phosphorylation of CREB (cAMP response element binding), a transcription factor that enhances transcription of CYP11A1 and other enzymes required for glucocorticoid synthesis.

Adrenal steroidogenesis occurs in a zone-specific fashion, with mineralocorticoid synthesis occurring in the outer zona glomerulosa, glucocorticoid synthesis in the zona fasciculata, and adrenal androgen biosynthesis in the inner zona reticularis serving as precursors for both classic and 11-oxygenated androgens (Fig. 398-1). All steroidogenic pathways require cholesterol import into the mitochondrion, a process initiated by the action of the steroidogenic acute regulatory (StAR) protein, which shuttles cholesterol from the outer to the inner mitochondrial membrane. The majority of steroidogenic enzymes are cytochrome P450 (CYP) enzymes, which are either located in the mitochondrion (side chain cleavage enzyme, CYP11A1; 11β-hydroxylase, CYP11B1; aldosterone synthase, CYP11B2) or in the endoplasmic reticulum membrane (17α-hydroxylase, CYP17A1; 21-hydroxylase, CYP21A2; aromatase, CYP19A1). These enzymes require electron donation via specific redox cofactor enzymes, P450 oxidoreductase (POR), and adrenodoxin/adrenodoxin reductase (ADX/ADR) for the microsomal and mitochondrial CYP enzymes, respectively. In addition, the short-chain dehydrogenase 3β-hydroxysteroid dehydrogenase type 2 (3β-HSD2), also termed Δ4, Δ5 isomerase, plays a major role in adrenal steroidogenesis.

The cholesterol side chain cleavage enzyme CYP11A1 generates pregnenolone. Glucocorticoid synthesis requires conversion of pregnenolone to progesterone by 3β-HSD2, followed by conversion to 17-hydroxyprogesterone (17OHP) by CYP17A1, further hydroxylation at carbon 21 by CYP21A2, and eventually, 11β-hydroxylation by CYP11B1 to generate active cortisol (Fig. 398-1). Mineralocorticoid synthesis also requires progesterone, which is first converted to deoxycorticosterone (DOC) by CYP21A2 and then converted via corticosterone and 18-hydroxycorticosterone to aldosterone in three steps catalyzed by CYP11B2. For adrenal androgen synthesis, pregnenolone undergoes conversion by CYP17A1, which uniquely catalyzes two enzymatic reactions. Via its 17α-hydroxylase activity, CYP17A1

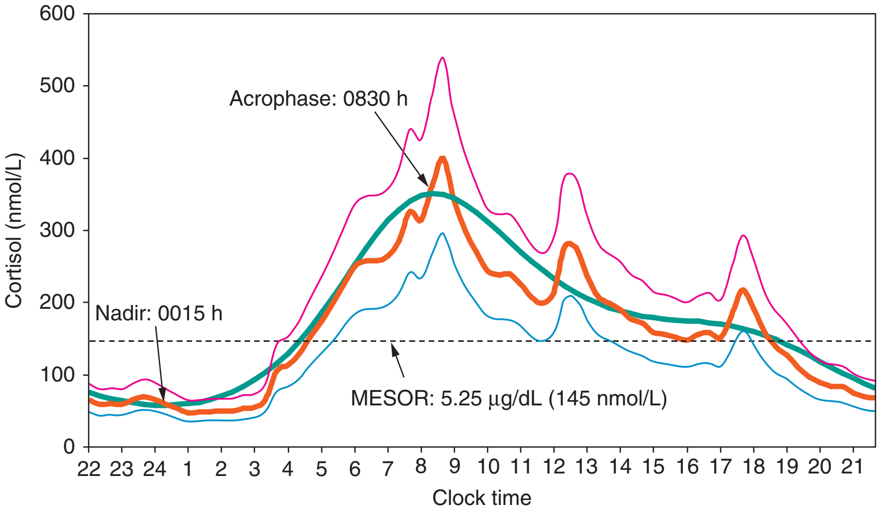

FIGURE 398-3 Physiologic cortisol circadian rhythm. Circulating cortisol concentrations (geometrical mean ± standard deviation values and fitted cosinor) drop under the rhythm-adjusted mean (MESOR) in the early evening hours, with nadir levels around midnight and a rise in the early morning hours; peak levels are observed ~8:30 A.M. (acrophase). *(Reproduced with permission from M Debono et al: Modified-release hydrocortisone to provide circadian cortisol profiles. J Clin Endocrinol Metab 94:1548, 2009.)*

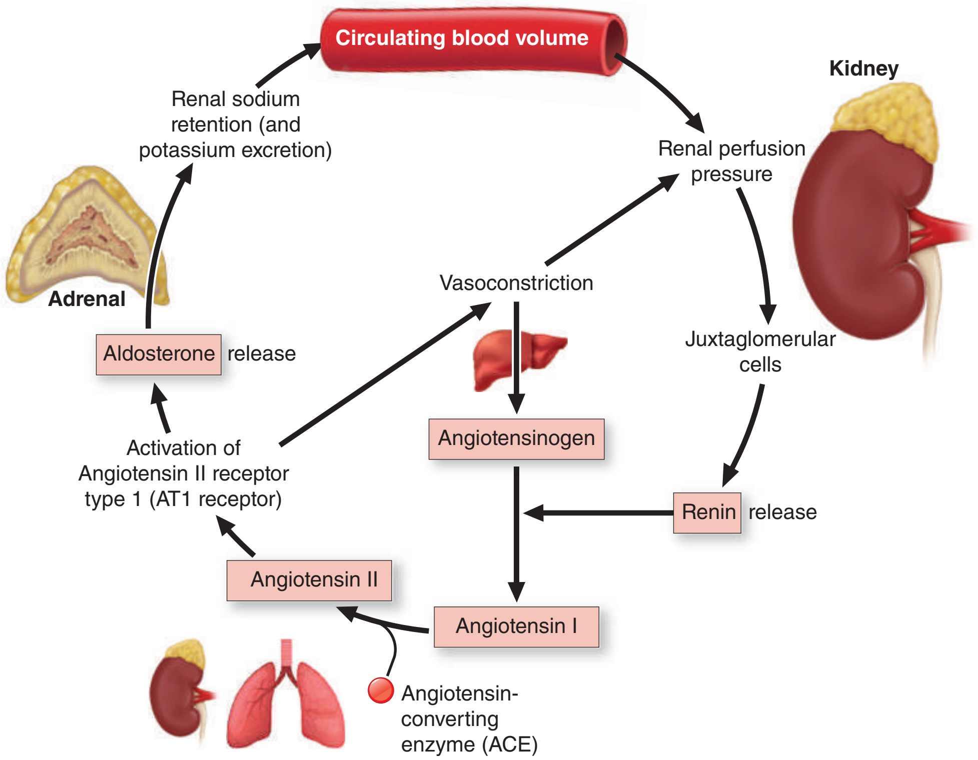
Circulating blood volume
Kidney
Renal perfusion pressure
Juxtaglomerular cells
Renin release
Renal sodium retention (and potassium excretion)
Adrenal
Aldosterone release
Vasoconstriction
Angiotensinogen
Activation of Angiotensin II receptor type 1 (AT1 receptor)
Angiotensin II
Angiotensin I
Angiotensin-converting enzyme (ACE)

FIGURE 398-4 Regulation of the renin-angiotensin-aldosterone (RAA) system.

converts pregnenolone to 17-hydroxypregnenolone, followed by generation of the universal sex steroid precursor DHEA via CYP17A1 17,20 lyase activity. The majority of DHEA is secreted by the adrenal in the form of its sulfate ester, DHEAS, generated by DHEA sulfotransferase (SULT2A1). DHEA is converted to androstenedione, which can be activated to testosterone or channeled into the 11-oxygenated androgen pathway by 11β-hydroxylation (CYP11B1).

Following its release from the adrenal, cortisol circulates in the bloodstream mainly bound to cortisol-binding globulin (CBG) and, to a lesser extent, to albumin, with only a minor fraction circulating as free, unbound hormone. Free cortisol is thought to enter cells directly, not requiring active transport. In addition, in a multitude of peripheral target tissues of glucocorticoid action, including adipose, liver, muscle, and brain, cortisol is generated from inactive cortisone within the cell by the enzyme 11β-hydroxysteroid dehydrogenase type 1 (11β-HSD1) **(Fig. 398-6)**. Thereby, 11β-HSD1 functions as a tissue-specific prereceptor regulator of glucocorticoid action. For the conversion of inactive cortisone to active cortisol, 11β-HSD1 requires nicotinamide adenine

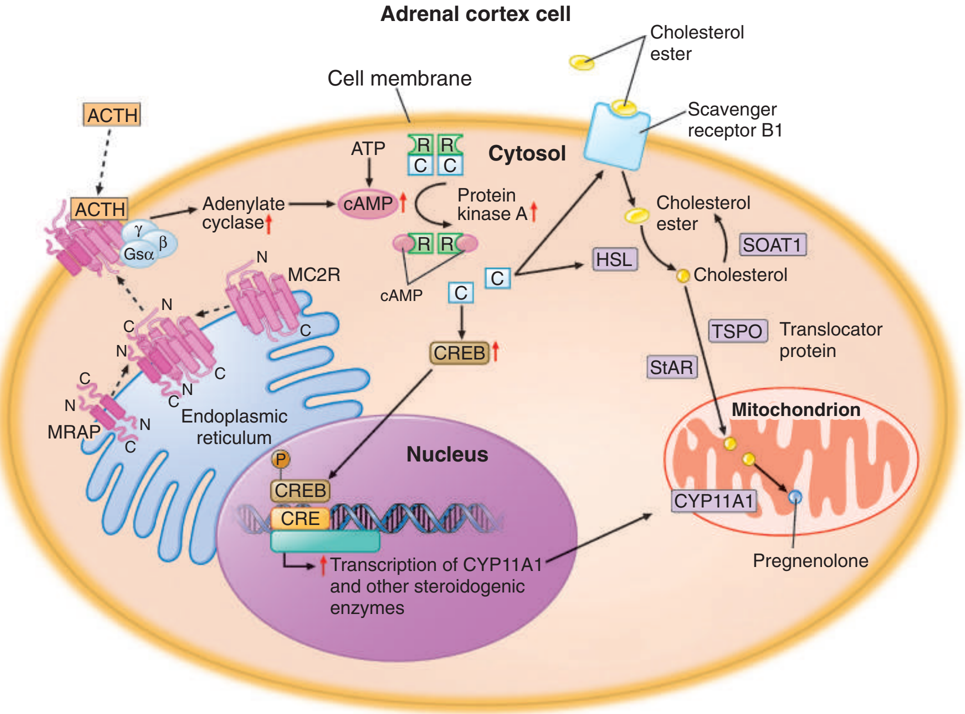
Adrenal cortex cell
Cell membrane
Cholesterol ester
Scavenger receptor B1
ACTH
ATP
R R
C C
Cytosol
ACTH
Adenylate cyclase↑
cAMP↑
Protein kinase A↑
γ
β
Gsα
R R
cAMP
C
C
HSL
Cholesterol ester
SOAT1
Cholesterol
N
MC2R
C
N
C
N
C
N
C
N
C
N
MRAP
N
C
Endoplasmic reticulum
CREB↑
TSPO
Translocator protein
StAR
Mitochondrion
Nucleus
P
CREB
CRE
CYP11A1
Pregnenolone
↑Transcription of CYP11A1 and other steroidogenic enzymes

FIGURE 398-5 ACTH effects on adrenal steroidogenesis. ACTH, adrenocorticotropic hormone; CREB, cAMP response element–binding protein; HSL, hormone-sensitive lipase; MRAP, MC2R-accessory protein; protein kinase A catalytic subunit (C; *PRKACA*), PKA regulatory subunit (R; *PRKAR1A*); SOAT1, sterol O-acyltransferase 1; StAR, steroidogenic acute regulatory (protein); TSPO, translocator protein.

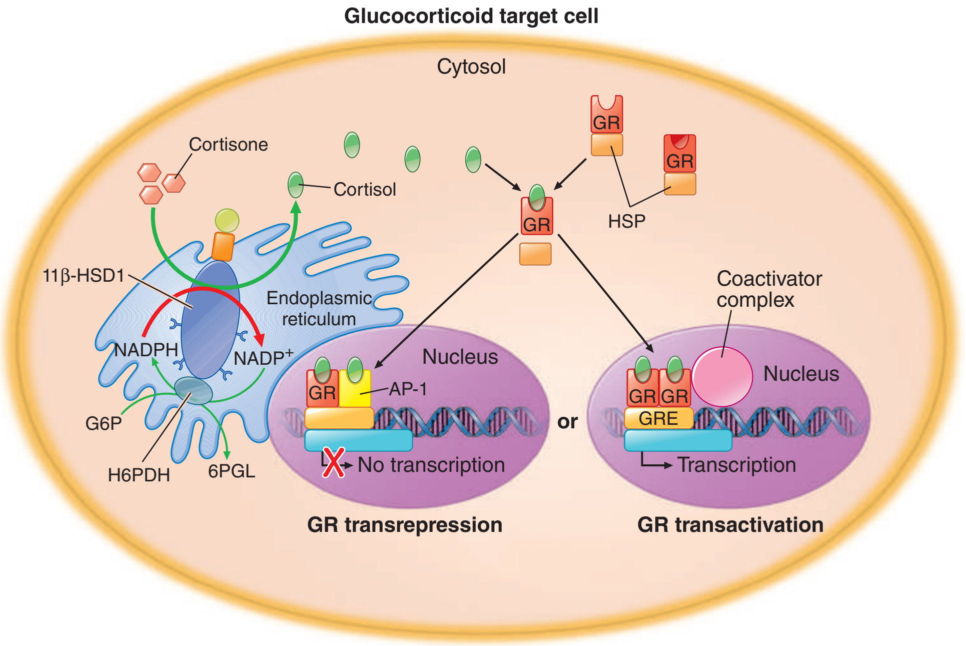

FIGURE 398-6 Prereceptor activation of cortisol and glucocorticoid receptor (GR) action. AP-1, activator protein-1; G6P, glucose-6-phosphate; GREs, glucocorticoid response elements; HSPs, heat shock proteins; NADPH, nicotinamide adenine dinucleotide phosphate (reduced form); 6PGL, 6-phosphogluconate.

dinucleotide phosphate (NADPH [reduced form]), which is provided by the enzyme hexose-6-phosphate dehydrogenase (H6PDH). Like the catalytic domain of 11β-HSD1, H6PDH is located in the lumen of the endoplasmic reticulum and converts glucose-6-phosphate (G6P) to 6-phosphogluconate (6PGL), thereby regenerating $NADP^+$ to NADPH, which drives the activation of cortisol from cortisone by 11β-HSD1.

In the cytosol of target cells, cortisol binds and activates the GR, which results in dissociation of heat shock proteins (HSPs) from the receptor and subsequent dimerization (Fig. 398-6). Cortisol-bound GR dimers translocate to the nucleus and activate glucocorticoid response elements (GREs) in the DNA sequence, thereby enhancing transcription of glucocorticoid-regulated genes (GR transactivation). However, cortisol-bound GR can also form heterodimers with transcription factors such as AP-1 or nuclear factor-κB (NF-κB), resulting in transrepression of proinflammatory genes, a mechanism of major importance for the anti-inflammatory action of glucocorticoids. It is important to note that corticosterone also exerts glucocorticoid activity, albeit much weaker than cortisol itself. However, in rodents, corticosterone is the major glucocorticoid, and in patients with 17-hydroxylase deficiency, lack of cortisol can be compensated for by higher concentrations of corticosterone that accumulates as a consequence of the enzymatic block.

Cortisol is inactivated to cortisone by the microsomal enzyme 11β-hydroxysteroid dehydrogenase type 2 (11β-HSD2) **(Fig. 398-7)**, mainly in the kidney, but also in the colon, salivary glands, and other mineralocorticoid target tissues. Cortisol and aldosterone bind the mineralocorticoid receptor (MR) with equal affinity; however, cortisol circulates in the bloodstream at about a 1000-fold higher concentration. Thus, only rapid inactivation of cortisol to cortisone by 11β-HSD2 prevents MR activation by excess cortisol, thereby acting as a tissue-specific modulator of the MR pathway. In addition to cortisol and aldosterone, DOC (Fig. 398-1) also exerts mineralocorticoid activity. DOC accumulation due to 11β-hydroxylase deficiency or due to tumor-related excess production can result in mineralocorticoid excess.

Aldosterone synthesis in the adrenal zona glomerulosa cells is driven by the enzyme aldosterone synthase (CYP11B2). The binding of angiotensin II to the AT1 receptor causes glomerulosa cell membrane depolarization by increasing intracellular sodium through inhibition of sodium-potassium (Na^+/K^+) ATPase enzymes as well as potassium channels. This drives an increase in intracellular calcium by opening voltage-dependent calcium channels or inhibition of calcium (Ca^{2+}) ATPase enzymes. Consequently, the calcium signaling pathway is triggered, resulting in upregulation of CYP11B2 transcription **(Fig. 398-8)**.

Analogous to cortisol action via the GR, aldosterone (or cortisol) binding to the MR in the kidney tubule cell dissociates the HSP-receptor complex, allowing homodimerization of the MR and translocation of the hormone-bound MR dimer to the nucleus (Fig. 398-7). The activated MR enhances transcription of the epithelial sodium channel (ENaC) and serum glucocorticoid-inducible kinase 1 (SGK-1). In the cytosol, interaction of ENaC with Nedd4 prevents cell surface expression of ENaC. However, SGK-1 phosphorylates serine residues within the Nedd4 protein, reduces the interaction between Nedd4 and ENaC, and consequently, enhances the trafficking of ENaC to the cell surface, where it mediates sodium retention at the expense of potassium extrusion through the renal outer medullary potassium channel (ROMK) channel.

CUSHING'S SYNDROME

(See also Chap. 392) Cushing's syndrome reflects a constellation of clinical features that result from chronic exposure to excess glucocorticoids of any etiology. The disorder can be ACTH-dependent (e.g., pituitary corticotrope adenoma, ectopic secretion of ACTH by nonpituitary tumor) or ACTH-independent (e.g., adrenocortical adenoma, adrenocortical carcinoma [ACC], nodular adrenal hyperplasia), as well as iatrogenic (e.g., administration of exogenous glucocorticoids to treat various inflammatory conditions). The term *Cushing's disease* refers specifically to Cushing's syndrome caused by a pituitary corticotrope adenoma.

Epidemiology Iatrogenic (or exogenous) Cushing's syndrome is by far the most common etiology, considering that around 1% of the population uses chronic glucocorticoid treatments for their anti-inflammatory and immunosuppressive properties. On the contrary, endogenous Cushing's syndrome is generally considered a rare disease; it occurs with an incidence of 1.8–3.2 per million population per year. However, it is debated whether mild cortisol excess may be more prevalent among patients with features of Cushing's such as centripetal obesity, type 2 diabetes, and osteoporotic vertebral fractures, recognizing that these are relatively nonspecific and common in the population.

Etiology In the overwhelming majority of patients with endogenous Cushing's syndrome, the underlying cause is an ACTH-producing pituitary neuroendocrine tumor, i.e., a corticotrope adenoma **(Table 398-1)**, as initially described by Harvey Cushing in 1912. Cushing's disease more frequently affects women, with the exception of prepubertal cases, where it is more common in boys. By contrast, ectopic

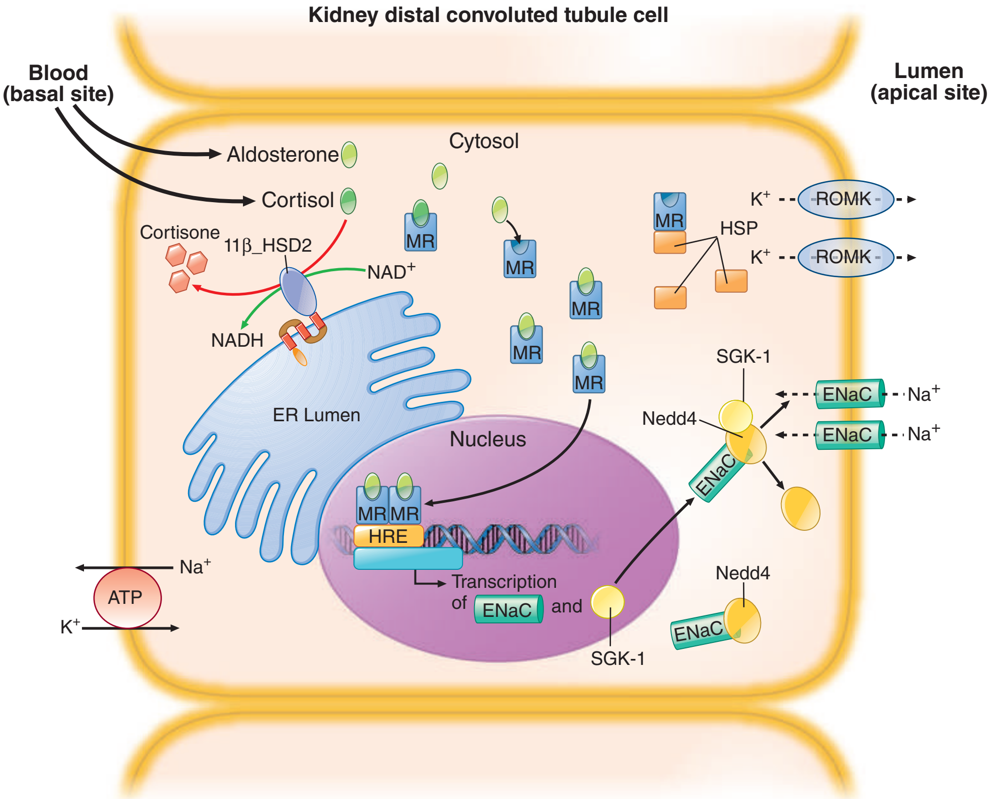
Kidney distal convoluted tubule cell
Blood
(basal site)
Lumen
(apical site)
Aldosterone
Cortisol
Cytosol
Cortisone
11β_HSD2
NAD+
NADH
MR
HSP
K+
ROMK
K+
ROMK
SGK-1
ENaC
Na+
ENaC
Na+
Nedd4
ENaC
ER Lumen
Nucleus
MR MR
HRE
Transcription
of ENaC and
SGK-1
Nedd4
ENaC
Na+
ATP
K+

FIGURE 398-7 Prereceptor inactivation of cortisol and mineralocorticoid receptor action. ENaC, epithelial sodium channel; HRE, hormone response element; Na^+/K^+-ATPase, sodium-potassium adenosine triphosphatase; NADH, nicotinamide adenine dinucleotide; ROMK, renal outer medullary potassium channel; SGK-1, serum glucocorticoid-inducible kinase-1.

ACTH syndrome is more frequently identified in men. Only 10% of patients with Cushing's syndrome have a primary, adrenal cause of their disease (i.e., autonomous cortisol excess independent of ACTH), and most of these patients are women.

In at least 90% of patients with Cushing's disease, ACTH excess is caused by a corticotrope pituitary microadenoma, often only a few millimeters in diameter. Pituitary macroadenomas (i.e., tumors >1 cm in size) are found in only 5–10% of patients and may have invasive features and affect the optic chiasm and the cavernous sinuses. Pituitary corticotrope adenomas usually occur sporadically but very rarely can be found in the context of multiple endocrine neoplasia type 1 (MEN 1) **(Chap. 400)**, MEN 4, familial isolated pituitary adenomas, Carney complex, DICER 1 syndrome, and Lynch syndrome. Pituitary adenomas causative of Cushing's disease frequently harbor activating somatic variants in the *USP8* gene, encoding the deubiquitinating enzyme ubiquitin-specific protease 8, which lead to constitutive activation of epidermal growth factor (EGF) signaling and consequent upregulated expression of the ACTH precursor POMC. *USP8* mutations are found in 11–62% of corticotrope adenomas and more frequently in adults and women with Cushing's disease. Less frequently, somatic variants of the glucocorticoid receptor gene (*NR3C1*), *BRAF*, *USP48*, and *TP53* are observed in the tumor tissue of *USP8* wild-type Cushing's disease cases.

Ectopic ACTH production is predominantly caused by occult carcinoid tumors, most frequently in the lung, but also in the thymus or pancreas. Because of their small size, these tumors are often difficult to locate. Advanced small-cell lung cancer can cause ectopic ACTH production. In rare cases, ectopic CRH and/or ACTH production has been found to originate from medullary thyroid carcinoma or pheochromocytoma, the latter co-secreting catecholamines and ACTH.

The majority of patients with endogenous ACTH-independent cortisol excess harbor a cortisol-producing adrenal adenoma, and

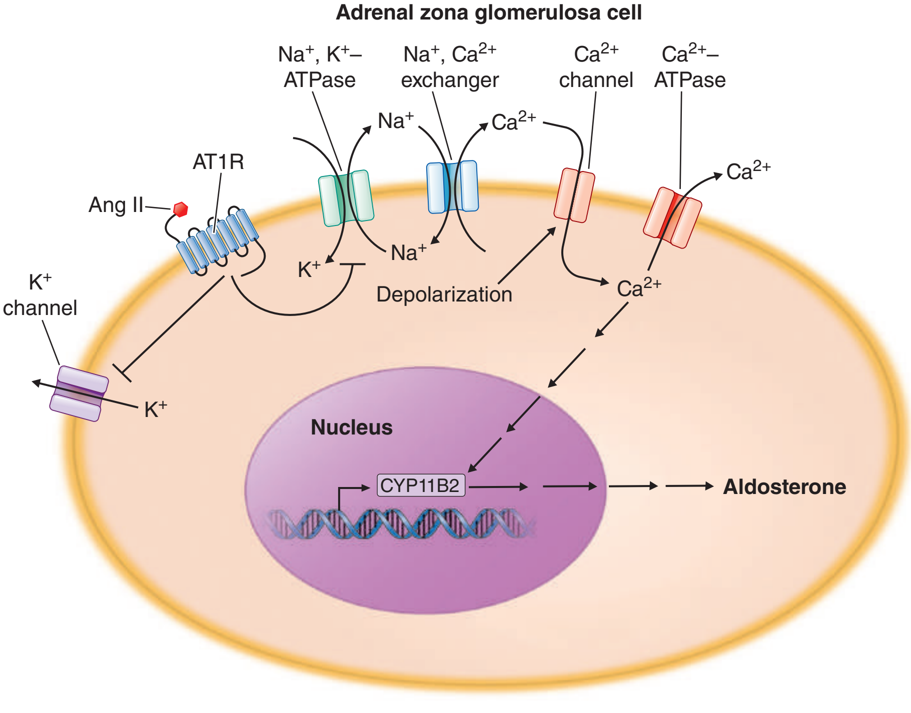
Adrenal zona glomerulosa cell
Na+, K+–
ATPase
Na+, Ca2+
exchanger
Ca2+
channel
Ca2+–
ATPase
Na+
Ca2+
Ca2+
AT1R
Ang II
K+
Na+
K+
channel
Depolarization
Ca2+
K+
Nucleus
CYP11B2
Aldosterone

FIGURE 398-8 Regulation of adrenal aldosterone synthesis. Ang II, angiotensin II; AT1R, angiotensin II receptor type 1; CYP11B2, aldosterone synthase.

TABLE 398-1 Causes of Endogenous Cushing's Syndrome

	FEMALE:MALE RATIO	%
ACTH-Dependent Cushing's Syndrome		**90**
Cushing's disease (= ACTH-producing pituitary adenoma)	4:1	75
Ectopic ACTH syndrome (due to ACTH secretion by bronchial or pancreatic carcinoid tumors, small-cell lung cancer, medullary thyroid carcinoma, pheochromocytoma, and others)	1:1	15
ACTH-Independent Cushing's Syndrome	**4:1**	**10**
Adrenocortical adenoma		5–10
Adrenocortical carcinoma		1
Rare causes: macronodular adrenal hyperplasia; primary pigmented nodular adrenal disease (micro- and/or macronodular); McCune-Albright syndrome		<1

Abbreviation: ACTH, adrenocorticotropic hormone.

activating somatic mutations in the PKA catalytic subunit *PRKACA* have been identified as the cause of disease in 40% of these tumors. Somatic inactivating defects of *PRKAR1A*, encoding one of the regulatory subunits of PKA, and *GNAS*, encoding the stimulatory G protein alpha subunit 1 GNAS-1 (guanine nucleotide-binding protein alpha stimulating activity polypeptide 1), have been observed less frequently in unilateral cortisol-producing adenomas. ACCs may also cause ACTH-independent cortisol excess and are often large, with excess production of several corticosteroid classes.

A rare but notable cause of adrenal cortisol excess is primary bilateral macronodular adrenal hyperplasia (PBMAH) with low circulating ACTH but with evidence for autocrine stimulation of cortisol production via intra-adrenal ACTH production. These hyperplastic nodules are often also characterized by aberrant expression of G protein–coupled receptors not usually found in the adrenal, including receptors for luteinizing hormone, vasopressin, serotonin, interleukin 1, catecholamines, or gastric inhibitory peptide (GIP), the cause of food-dependent Cushing's. Activation of these receptors results in upregulation of PKA signaling, as physiologically occurs with ACTH, with a subsequent increase in cortisol production. A combination of germline and somatic mutations in the tumor-suppressor gene *ARMC5* have been identified as a prevalent cause of Cushing's due to bilateral macronodular adrenal hyperplasia; these patients often present with biochemical evidence of Cushing's but lack specific clinical signs, which develop slowly over decades and accelerate cardiovascular risk. Constitutively activating *PRKACA* mutations and inactivating *KDM1A* variants (lysine-specific histone demethylase 1A) can also be a rare cause of bilateral macronodular adrenal hyperplasia, the latter associated with GIP-dependent Cushing's syndrome. Bilateral macronodular adrenal hyperplasia with cortisol excess has also been described in other familial syndromes: MEN 1, hereditary leiomyomatosis and renal cell cancer (*FH* gene), and familial adenomatous polyposis (*APC* gene).

Germline mutations in *PRKAR1A* are found in patients with primary pigmented nodular adrenal disease (PPNAD) as part of *Carney's complex*, an autosomal dominant multiple neoplasia condition associated with cardiac and cutaneous myxomas, hyperlentiginosis, Sertoli cell tumors, ovarian tumors, breast tumors, thyroid nodules, Schwannomas, and PPNAD. PPNAD can present as micronodular or macronodular hyperplasia, or both. Phosphodiesterases can influence intracellular cAMP and can thereby impact PKA activation. Mutations in *PDE11A* and *PDE8B* have been identified in patients with bilateral adrenal hyperplasia and Cushing's, with and without evidence of PPNAD.

Another rare cause of ACTH-independent Cushing's is *McCune-Albright syndrome*, also associated with polyostotic fibrous dysplasia, unilateral café-au-lait spots, and precocious puberty. McCune-Albright syndrome is caused by *GNAS* activating mutations (Table 398-1; **Chap. 424**).

TABLE 398-2 Signs and Symptoms of Cushing's Syndrome

BODY COMPARTMENT/ SYSTEM	SIGNS AND SYMPTOMS
Body fat	Weight gain, central obesity, rounded face, fat pad on back of neck ("buffalo hump")
Skin	Facial plethora, thin and brittle skin, easy bruising, broad and purple stretch marks, acne, hirsutism
Bone	Osteopenia, osteoporosis (vertebral fractures), decreased linear growth in children
Muscle	Weakness, proximal myopathy (prominent atrophy of gluteal and upper leg muscles with difficulty climbing stairs or getting up from a chair)
Cardiovascular system	Hypertension, hypokalemia, edema, atherosclerosis
Metabolism	Glucose intolerance/diabetes, dyslipidemia
Reproductive system	Decreased libido; in women, amenorrhea (due to cortisol-mediated inhibition of gonadotropin release)
Central nervous system	Irritability, emotional lability, depression, insomnia and disrupted sleep, sometimes cognitive defects; in severe cases, paranoid psychosis
Blood and immune system	Increased susceptibility to infections, increased white blood cell count, eosinopenia, hypercoagulation with increased risk of deep vein thrombosis and pulmonary embolism

Clinical Manifestations Glucocorticoids affect almost all cells of the body; thus, signs of cortisol excess impact multiple physiologic systems **(Table 398-2)**, with upregulation of gluconeogenesis, lipolysis, and protein catabolism causing the most prominent features. In addition, excess glucocorticoid secretion overcomes the ability of 11β-HSD2 to rapidly inactivate cortisol to cortisone in the kidney, thereby exerting mineralocorticoid actions, manifest as hypertension, hypokalemia, and edema. Excess glucocorticoids also interfere with central regulatory systems, leading to suppression of gonadotropins with subsequent hypogonadism and amenorrhea and suppression of the hypothalamic-pituitary-thyroid axis, resulting in decreased thyroid-stimulating hormone (TSH) secretion.

The majority of clinical signs and symptoms observed in Cushing's syndrome are relatively nonspecific and include features such as obesity, diabetes, diastolic hypertension, hirsutism, and depression that are commonly found in patients who do not have Cushing's. Therefore, careful clinical assessment is an important aspect of evaluating suspected cases. A diagnosis of Cushing's should be considered when several clinical features are found in the same patient, particularly when more specific features are found or manifest at an unusual age, e.g., osteoporosis in a young patient. Distinct features include thinning and fragility of the skin, with easy bruising and broad (>1 cm), purplish striae **(Fig. 398-9)**, and signs of proximal myopathy, which becomes most obvious when trying to stand up from a chair without the use of hands or when climbing stairs. Clinical manifestations of Cushing's do not differ substantially among the different causes of Cushing's. In ectopic ACTH syndrome, hyperpigmentation of the knuckles, scars, or skin areas exposed to increased friction can be observed (Fig. 398-9) and is caused by stimulatory effects of excess ACTH and other POMC cleavage products on melanocyte pigment production. Furthermore, patients with ectopic ACTH syndrome, and some with ACC as the cause of Cushing's, may have a more rapid onset and progression of clinical signs and symptoms, namely edema, hypokalemia, and hypertension.

Patients with Cushing's syndrome can be acutely endangered by deep vein thrombosis, with subsequent pulmonary embolism, due to a hypercoagulable state associated with cortisol excess. The majority of patients also experience psychiatric symptoms, mostly in the form of anxiety or depression, but acute paranoid or depressive psychosis may occur. Even after cure, long-term health may be affected by persistently impaired health-related quality of life and increased risk of cardiovascular disease and osteoporosis with vertebral fractures, depending on the duration and degree of exposure to significant cortisol excess.

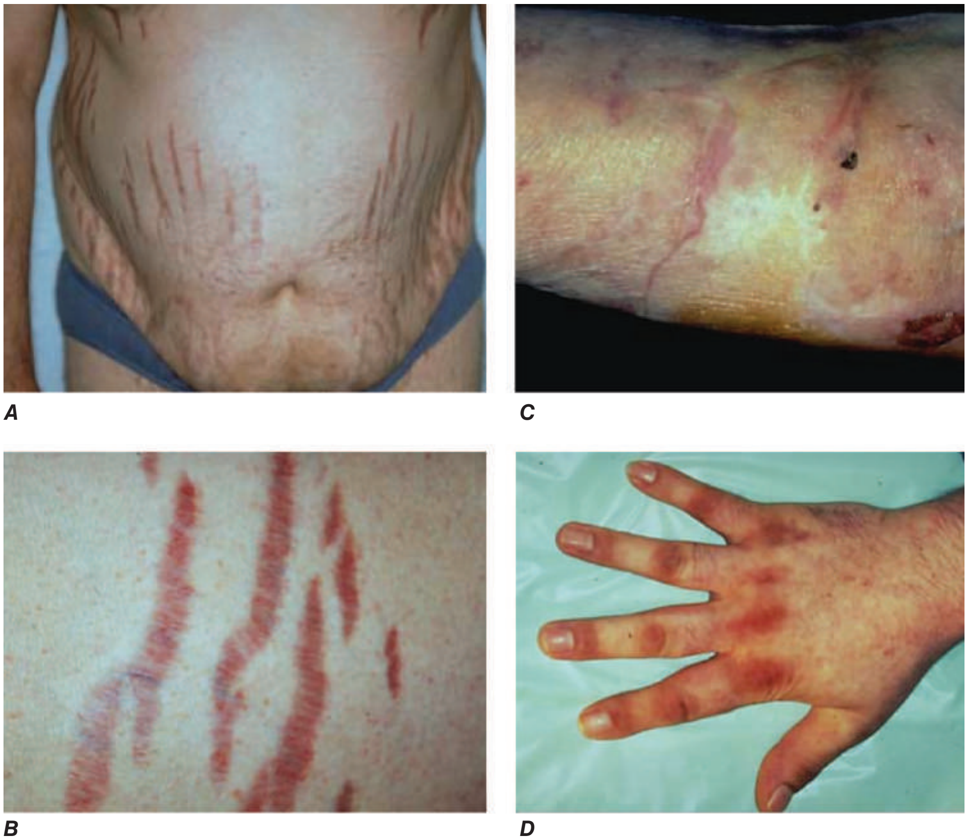

FIGURE 398-9 Clinical features of Cushing's syndrome. ***A.*** Note central obesity and broad, purple stretch marks (***B.*** close-up). ***C.*** Note thin and brittle skin in an elderly patient with Cushing's syndrome. ***D.*** Hyperpigmentation of the knuckles in a patient with ectopic adrenocorticotropic hormone (ACTH) excess.

Diagnosis The most important first step in the management of patients with suspected Cushing's syndrome is to establish the correct diagnosis. Most mistakes in clinical management, leading to unnecessary imaging or surgery, are made because the diagnostic protocol is not followed **(Fig. 398-10)**. This protocol requires establishing the diagnosis of Cushing's beyond doubt prior to employing any tests used for the differential diagnosis of the condition. In principle, after excluding exogenous glucocorticoid use as the cause of clinical signs and symptoms, suspected cases should be tested if there are multiple and progressive features of Cushing's, particularly features with a potentially higher discriminatory value. Exclusion of cortisol excess is also indicated in patients with incidentally discovered adrenal masses.

A diagnosis of Cushing's can be considered as established if the results of several tests are consistently suggestive of Cushing's. These tests may include increased 24-h urinary free cortisol excretion in three separate collections, failure to appropriately suppress morning serum cortisol after overnight exposure to dexamethasone, and evidence of loss of diurnal cortisol secretion with high levels of serum or salivary cortisol at midnight, the time of the physiologically lowest secretion (Fig. 398-10). Factors potentially affecting the outcome of these diagnostic tests have to be excluded such as incomplete 24-h urine collection or rapid inactivation of dexamethasone due to concurrent intake of CYP3A4-inducing drugs (e.g., antiepileptics, rifampicin). Concurrent intake of oral contraceptives that raise CBG and thus total cortisol can cause failure to suppress after dexamethasone. If in doubt, testing should be repeated after 4–6 weeks off oral estrogens. Patients with pseudo-Cushing states, e.g., alcohol-related nonneoplastic hypercortisolism or associated with major depression or morbid obesity, as well as those with cyclic Cushing's syndrome, may require further testing to safely confirm or exclude the diagnosis of Cushing's. In addition, the biochemical assays employed can affect the test results, with specificity representing a common problem with antibody-based assays for the measurement of urinary free cortisol. These assays have been greatly improved by the introduction of highly specific tandem mass spectrometry.

Differential Diagnosis The evaluation of patients with confirmed Cushing's should be carried out by an endocrinologist and begins with the differential diagnosis of ACTH-dependent and ACTH-independent cortisol excess (Fig. 398-10). Generally, plasma ACTH levels are suppressed in cases of autonomous adrenal cortisol excess, consequent to enhanced negative feedback to the hypothalamus and pituitary. By contrast, patients with ACTH-dependent Cushing's have normal or increased plasma ACTH, with very high levels being found in some patients with ectopic ACTH syndrome. Importantly, imaging should only be used after it is established whether the cortisol excess is ACTH-dependent or ACTH-independent because nodules in the pituitary or the adrenal are a common finding in the general population. In patients with confirmed ACTH-independent excess, adrenal imaging is indicated **(Fig. 398-11)**, preferably using an unenhanced computed tomography (CT) scan. This allows assessment of adrenal morphology and also the determination of tumor density, i.e., the attenuation value measured in Hounsfield units (HUs), which helps to distinguish between benign and malignant adrenal lesions (note, this only applies to unenhanced CT; postcontrast CT does not convey this information).

For ACTH-dependent cortisol excess **(Chap. 392)**, a magnetic resonance image (MRI) of the pituitary is the investigation of choice, but it may not show an abnormality in up to 40% of cases because of small tumors below the sensitivity of detection. Characteristically, pituitary corticotrope adenomas fail to enhance following gadolinium administration on T1-weighted MRI images. In all cases of confirmed ACTH-dependent Cushing's, further tests are required for the differential diagnosis of pituitary Cushing's disease and ectopic ACTH syndrome. These tests exploit the fact that most pituitary corticotrope adenomas still display regulatory features, including residual ACTH suppression by high-dose glucocorticoids and responsiveness to CRH or desmopressin, a synthetic analog of vasopressin. In contrast, ectopic sources of ACTH are typically resistant to dexamethasone suppression and unresponsive to CRH or desmopressin (Fig. 398-10). However, it should be noted that a small minority of ectopic ACTH-producing

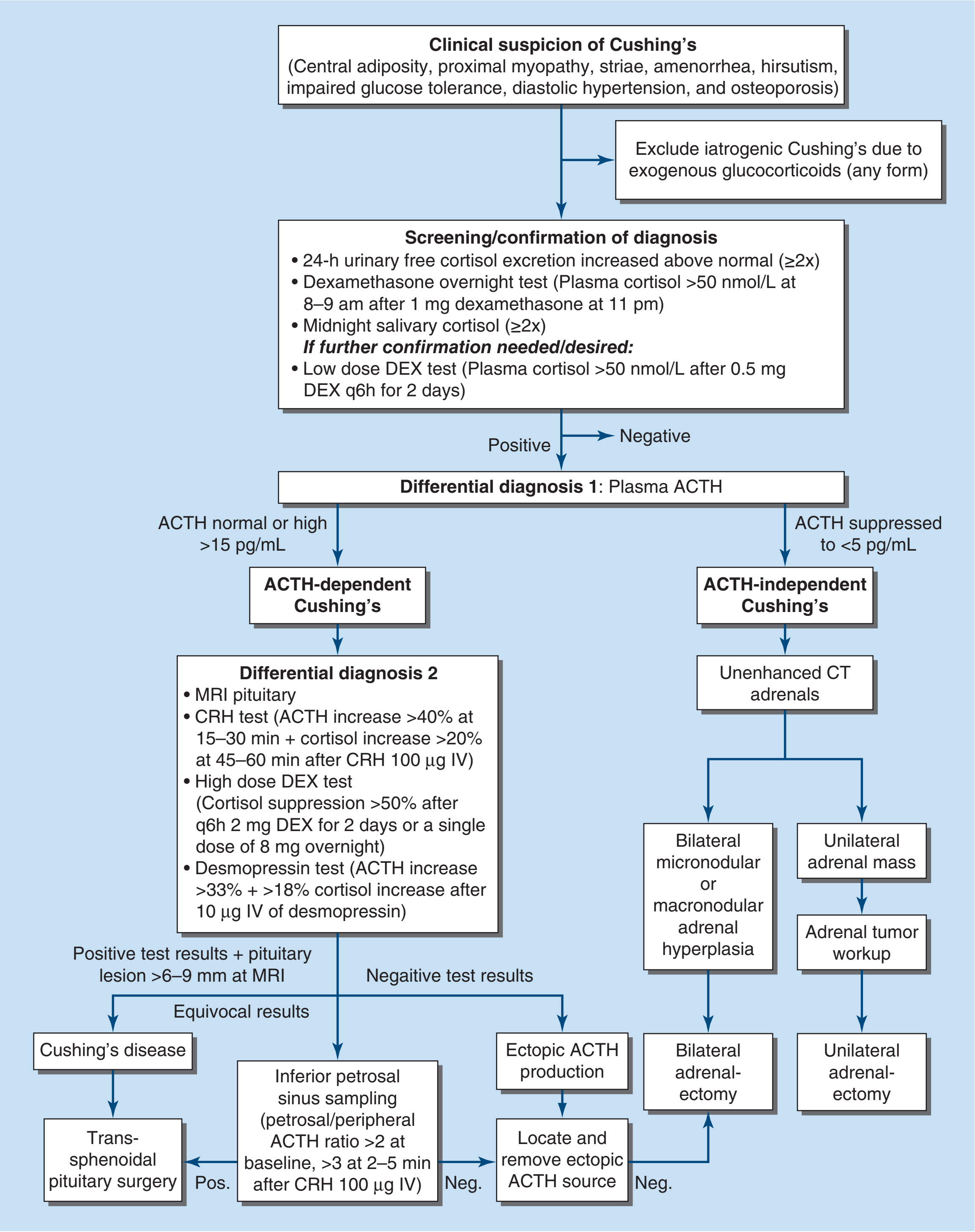
Clinical suspicion of Cushing's
(Central adiposity, proximal myopathy, striae, amenorrhea, hirsutism, impaired glucose tolerance, diastolic hypertension, and osteoporosis)
Exclude iatrogenic Cushing's due to exogenous glucocorticoids (any form)
Screening/confirmation of diagnosis
• 24-h urinary free cortisol excretion increased above normal (≥2x)
• Dexamethasone overnight test (Plasma cortisol >50 nmol/L at 8–9 am after 1 mg dexamethasone at 11 pm)
• Midnight salivary cortisol (≥2x)
If further confirmation needed/desired:
• Low dose DEX test (Plasma cortisol >50 nmol/L after 0.5 mg DEX q6h for 2 days)
Positive
Negative
Differential diagnosis 1: Plasma ACTH
ACTH normal or high >15 pg/mL
ACTH suppressed to <5 pg/mL
ACTH-dependent Cushing's
ACTH-independent Cushing's
Differential diagnosis 2
• MRI pituitary
• CRH test (ACTH increase >40% at 15–30 min + cortisol increase >20% at 45–60 min after CRH 100 μg IV)
• High dose DEX test (Cortisol suppression >50% after q6h 2 mg DEX for 2 days or a single dose of 8 mg overnight)
• Desmopressin test (ACTH increase >33% + >18% cortisol increase after 10 μg IV of desmopressin)
Unenhanced CT adrenals
Bilateral micronodular or macronodular adrenal hyperplasia
Unilateral adrenal mass
Adrenal tumor workup
Positive test results + pituitary lesion >6–9 mm at MRI
Negaitive test results
Equivocal results
Cushing's disease
Ectopic ACTH production
Bilateral adrenal-ectomy
Unilateral adrenal-ectomy
Inferior petrosal sinus sampling (petrosal/peripheral ACTH ratio >2 at baseline, >3 at 2–5 min after CRH 100 μg IV)
Trans-sphenoidal pituitary surgery
Pos.
Neg.
Locate and remove ectopic ACTH source
Neg.

FIGURE 398-10 Management of the patient with suspected Cushing's syndrome. ACTH, adrenocorticotropic hormone; CRH, corticotropin-releasing hormone; CT, computed tomography; DEX, dexamethasone; MRI, magnetic resonance imaging.

tumors exhibit dynamic responses similar to pituitary corticotrope tumors. If the tests show discordant results or if there is any other reason for doubt, the differential diagnosis can be further clarified by performing bilateral inferior petrosal sinus sampling (IPSS) with concurrent blood sampling for ACTH in the right and left inferior petrosal sinus and a peripheral vein. An increased central/peripheral plasma ACTH ratio >2 at baseline and >3 at 2–5 min after CRH injection is indicative of Cushing's disease (Fig. 398-10), with very high sensitivity and specificity. Of note, the results of the IPSS cannot be reliably used for lateralization (i.e., prediction of the location of the tumor within the pituitary) because there is broad interindividual variability in the venous drainage of the pituitary region. Importantly, no cortisol-lowering agents should be used prior to IPSS.

If the differential diagnostic testing indicates ectopic ACTH syndrome, then further imaging should include high-resolution, fine-cut CT scanning of the chest and abdomen for scrutiny of the lung, thymus, and pancreas. If no lesions are identified, an MRI of the chest can be considered because carcinoid tumors usually show high signal intensity on T2-weighted images. Furthermore, octreotide scintigraphy and ^{68}Ga DOTATATE positron emission tomography (PET)/CT can be helpful in some cases because ectopic ACTH-producing tumors often express somatostatin receptors. Depending on the suspected cause, patients with ectopic ACTH syndrome should also undergo blood sampling for fasting gut hormones, chromogranin A, calcitonin, and biochemical exclusion of pheochromocytoma.

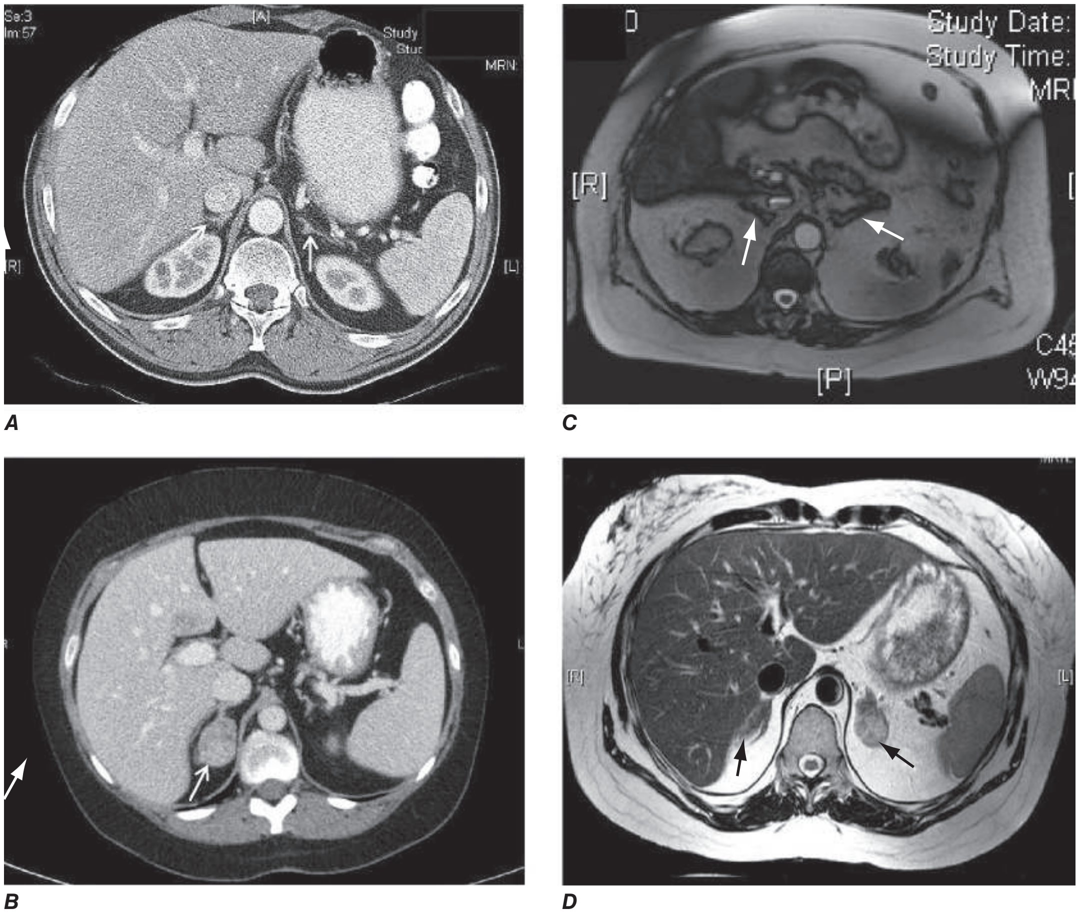

FIGURE 398-11 Adrenal imaging in Cushing's syndrome. ***A.*** Adrenal computed tomography (CT) showing normal bilateral adrenal morphology (*arrows*). ***B.*** CT scan depicting a right adrenocortical adenoma (*arrow*) causing Cushing's syndrome. ***C.*** Magnetic resonance imaging (MRI) showing bilateral adrenal hyperplasia due to excess adrenocorticotropic hormone stimulation in Cushing's disease. ***D.*** MRI showing bilateral macronodular hyperplasia causing Cushing's syndrome.

TREATMENT

Cushing's Syndrome

Overt Cushing's is associated with a poor prognosis if left untreated. In ACTH-independent disease, treatment consists of surgical removal of the adrenal tumor or nodular hyperplasia. For benign tumors, a minimally invasive laparoscopic approach can be used, whereas for those suspected of malignancy, an open approach is preferred.

In Cushing's disease, the treatment of choice is selective removal of the pituitary corticotrope tumor, usually via an endoscopic transsphenoidal approach. This results in an initial cure rate of 60–80% when performed by a highly experienced surgeon, although remission rates are much lower (12–70%) in patients presenting with larger or invasive tumors. Even after initial remission following surgery, long-term follow-up is important because late relapse occurs in a significant number of patients. If pituitary disease persists after surgery or recurs after initial remission, there are several options including second surgery, fractionated radiotherapy, stereotactic radiosurgery, pharmacologic therapy, and bilateral adrenalectomy. These options need to be applied in a highly individualized fashion.

In some patients with very severe, overt Cushing's (e.g., difficult-to-control hypokalemic hypertension, acute psychosis, or life-threatening infections), it may be necessary to introduce medical therapy to rapidly control the cortisol excess during the period leading up to surgery, which also can help to alleviate hypercoagulability and, thus, operative risk. Similarly, patients with metastasized, glucocorticoid-producing carcinomas may require long-term antiglucocorticoid drug treatment. In the case of ectopic ACTH syndrome, in which the tumor cannot be located, one must carefully weigh whether drug treatment or bilateral adrenalectomy is the most appropriate choice, with the latter facilitating immediate cure but requiring life-long corticosteroid replacement. In this instance, it is paramount to ensure regular imaging follow-up for identification of the ectopic ACTH source.

Oral agents with established efficacy in Cushing's syndrome are osilodrostat, metyrapone, ketoconazole, levoketoconazole, and mifepristone. Osilodrostat and metyrapone inhibit cortisol synthesis at the level of 11β-hydroxylase (Fig. 398-1). The antimycotic drug ketoconazole, and its stereoisomer levoketoconazole, inhibit the early steps of steroidogenesis. Mifepristone is a GR blocker, hence mitigating the effects of cortisol excess on target tissues. Typical starting doses are 1–2 mg bid for osilodrostat (maximum daily dose, 60 mg), 500 mg tid for metyrapone (maximum daily dose, 6 g), 200 mg tid for ketoconazole (maximum daily dose, 1200 mg), 150 mg bid for levoketoconazole (maximum daily dose, 1200 mg), and 300 mg for mifepristone (maximum daily dose, 1200 mg). Mitotane, a derivative of the insecticide o,p'DDD, is an adrenolytic agent that is also effective for reducing cortisol. Because of its side effect profile, it is most commonly used in the context of ACC, but low-dose treatment (500–1000 mg/d) has also been used in benign Cushing's. In severe cases of cortisol excess, etomidate, an agent that potently blocks 11β-hydroxylase and aldosterone synthase, can be used to rapidly lower cortisol. It is administered by continuous IV infusion in low, nonanesthetic doses. For Cushing's disease, the subcutaneous or intramuscular administration of pasireotide, a somatostatin receptor agonist, represents another therapeutic option, if surgical cure cannot be achieved. The dopamine agonist cabergoline is also used off-label to treat certain patients with Cushing's disease, especially those with mild cortisol excess and/or residual tumor due to the potential for tumor shrinkage.

After the successful removal of an ACTH- or cortisol-producing tumor, the HPA axis will remain suppressed. Thus, hydrocortisone replacement needs to be initiated at the time of surgery and slowly tapered following recovery, to allow physiologic adaptation to normal cortisol levels. Depending on the degree and duration of cortisol excess, the HPA axis may require many months or even years to resume normal function and sometimes does not recover. Generally, ectopic ACTH syndrome shows the best recovery rate (80%; median time to recovery 7 months) and adrenal Cushing's has the lowest (40%; median time to recovery 30 months), with Cushing's disease intermediate (60%; median time to recovery 17 months).

PRIMARY MINERALOCORTICOID EXCESS

Epidemiology Following the first description of a patient with an aldosterone-producing adrenal adenoma (*Conn's syndrome*), mineralocorticoid excess was thought to represent a rare cause of hypertension. However, in studies systematically screening all patients with hypertension, a much higher prevalence is now recognized, ranging from 5 to 12%. The prevalence is higher when patients are preselected for hypokalemic hypertension or drug-resistant hypertension.

Etiology The most common cause of primary (i.e., adrenal) mineralocorticoid excess is primary aldosteronism, reflecting excess production of aldosterone by the adrenal zona glomerulosa. Traditionally, the main subtypes of the disease are bilateral primary aldosteronism due to bilateral micronodular adrenal hyperplasia (the most common form) and unilateral aldosterone-producing adenoma (Conn's syndrome) **(Table 398-3)**. Although recent evidence has shown that primary aldosteronism exists on a spectrum blurring the line between bilateral and unilateral forms, their binary distinction underpins clinical management because unilateral forms are amenable to potentially curative surgery, whereas mineralocorticoid receptor antagonist therapy is the treatment of choice for bilateral forms. Somatic mutations in channels and enzymes responsible for increasing sodium and calcium influx in adrenal zona glomerulosa cells have been identified as prevalent causes of aldosterone-producing adrenal adenomas (Table 398-3) and, in the case of germline mutations, also of familial forms of primary aldosteronism. Sporadic bilateral adrenal hyperplasia as a cause of mineralocorticoid excess is usually micronodular but can also contain larger nodules that might be mistaken for a unilateral adenoma. Genetic forms of primary aldosteronism, which account for up to 6% of cases, should be suspected in patients with severe hypertension diagnosed

TABLE 398-3 Causes of Primary Mineralocorticoid Excess

CAUSES OF PRIMARY MINERALOCORTICOID EXCESS	MECHANISM	%
Sporadic Primary Aldosteronism		**94–99**
Adrenal (Conn's) adenoma	Autonomous aldosterone excess can be caused by somatic (intratumor) mutations in the potassium channel GIRK4 (encoded by *KCNJ5*; identified as cause of disease in 40–70% of aldosterone-producing adenomas and associated with overproduction of hybrid steroids). Further causes include somatic mutations affecting the α-subunit of the Na$^+$/K$^+$-ATPase (encoded by *ATP1A1*), the plasma membrane calcium-transporting ATPase 3 (encoded by *ATP2B3*), the voltage-gated calcium channel CaV1.3 and CaV3.2 (encoded by *CACNA1D* and *CACNA1H*, respectively), and the ClC-2 chloride channel (encoded by *CLCN2*). All mutations result in the upregulation of CYP11B2 and hence aldosterone synthesis. Rarely, somatic mutations affecting β-catenin (encoded by *CTNNB1*) have been observed alone or in combination with *CACNA1D* mutations and are presumed to promote an increase in the number of aldosterone-producing cells.	30–40
Bilateral adrenal hyperplasia	Bilateral autonomous aldosterone excess due to aldosterone-producing diffuse hyperplasia and/or multiple aldosterone-producing micronodules. *CACNA1D* somatic mutations have been described in 58% of aldosterone-producing micronodules of patients with bilateral aldosterone excess who underwent unilateral adrenalectomy as a nonstandard treatment.	60–70
Familial Primary Aldosteronism		**1–6**
Type 1 (also known as glucocorticoid-remediable hyperaldosteronism or dexamethasone-suppressible hyperaldosteronism)	Autosomal dominant. Crossover between the *CYP11B1* and *CYP11B2* genes results in ACTH-driven aldosterone production. Characterized by severe hypertension in childhood or young adults (often <20 years), with a high risk of cardiovascular events including hemorrhagic stroke at a young age due to ruptured intracranial aneurysm. Analysis of adrenal steroidogenesis shows overproduction of hybrid steroids.	
Type 2	Autosomal dominant. The most common form (1–6% of adults with primary aldosteronism). Due to germline *CLCN2* mutations. Variable phenotypic presentation, which typically includes early-onset hypertension.	
Type 3	Autosomal dominant. Germline *KCNJ5* mutations. Characterized by severe early-onset hypertension (<20 years) with massive bilateral macronodular adrenal hyperplasia. Analysis of adrenal steroidogenesis shows overproduction of hybrid steroids.	
Type 4	Autosomal dominant. Germline *CACNA1H* mutations. Characterized by severe early-onset hypertension (<20 years), which can be associated with developmental disorders in some cases.	
Primary aldosteronism, seizures, and neurologic abnormalities (PASNA)	De novo *CACNA1D* mutations. Characterized by childhood-onset hypertension, seizures, neurologic abnormalities, congenital hyperinsulinism, and cardiac abnormalities.	
Other Causes (Rare)		**<1**
Syndrome of apparent mineralocorticoid excess (SAME)	Mutations in *HSD11B2* result in lack of renal inactivation of cortisol to cortisone, leading to excess activation of the MR by cortisol (inhibition of 11β-hydroxysteroid dehydrogenase type 2 by excess licorice ingestion can have similar effects).	
Cushing's syndrome	Cortisol excess overcomes the capacity of HSD11B2 to inactivate cortisol to cortisone, consequently flooding the MR.	
Glucocorticoid resistance	Upregulation of cortisol production due to GR mutations results in flooding of the MR by cortisol.	
Adrenocortical carcinoma	Autonomous aldosterone and/or DOC excess.	
Congenital adrenal hyperplasia	Accumulation of DOC due to mutations in *CYP11B1* or *CYP17A1*.	
Progesterone-induced hypertension	Progesterone acts as an abnormal ligand due to mutations in the MR gene.	
Liddle's syndrome	Mutant ENaC β or γ subunits resulting in reduced degradation of ENaC keeping the membrane channel in open conformation for longer, enhancing mineralocorticoid action.	

Abbreviations: ACTH, adrenocorticotropic hormone; DOC, deoxycorticosterone; ENaC, epithelial sodium channel; GR, glucocorticoid receptor; HSD11B2, 11β-hydroxysteroid dehydrogenase type 2; MR, mineralocorticoid receptor.

during childhood or young adulthood (Table 398-3). In rare instances, primary aldosteronism is caused by an ACC. Carcinomas should be considered in younger patients and in those with larger tumors because benign aldosterone-producing adenomas usually measure <2 cm in diameter.

A rare cause of aldosterone excess is glucocorticoid-remediable aldosteronism (GRA), or type 1 familial primary aldosteronism, which is caused by a chimeric gene resulting from the crossover of promoter sequences between the *CYP11B1* and *CYP11B2* genes that are involved in glucocorticoid and mineralocorticoid synthesis, respectively (Fig. 398-1). This rearrangement brings *CYP11B2* transcription under the control of ACTH receptor signaling; consequently, aldosterone production is regulated by ACTH rather than by renin. The family history can be helpful because there may be evidence for dominant transmission of hypertension. Recognition of the disorder is important because it can be associated with early-onset hypertension and strokes. In addition, glucocorticoid suppression can reduce aldosterone production. The adrenal glands of patients with GRA produce high levels of the hybrid steroids 18-oxocortisol and 18-hydroxycortisol due to the coexistence of CYP11B1 and CYP11B2 enzymatic activities in the same steroidogenic cells, which are normally segregated to the zona fasciculata and glomerulosa, respectively. Similarly, high hybrid steroid levels are observed in patients with aldosterone-producing adenomas harboring somatic *KCNJ5* mutations (encoding the potassium channel GIRK4) and type 3 familial primary aldosteronism due to germline *KCNJ5* mutations (Table 398-3). Hybrid steroids are produced by normal adrenal glands in very low concentrations and can therefore be measured in the blood and urine to identify these conditions.

Other rare causes of primary mineralocorticoid excess are listed in Table 398-3. An important cause is excess binding and activation of the MR by a steroid other than aldosterone. Cortisol acts as a potent mineralocorticoid if it escapes efficient inactivation to cortisone by 11β-HSD2 in the kidney (Fig. 398-7). This can be caused by inactivating mutations in the *HSD11B2* gene resulting in the syndrome of apparent mineralocorticoid excess (SAME) that characteristically manifests with severe hypokalemic hypertension in childhood. However, milder mutations may cause normokalemic hypertension manifesting in adulthood (type II SAME). Inhibition of 11β-HSD2 by excess licorice ingestion also results in hypokalemic hypertension, as does overwhelming of 11β-HSD2 conversion capacity by cortisol excess in Cushing's syndrome. DOC also binds and activates the MR and can cause hypertension if its circulating concentrations are increased. This can arise through autonomous DOC secretion by an ACC, but also when DOC accumulates as a consequence of an adrenal enzymatic block, as seen in congenital adrenal hyperplasia (CAH) due to CYP11B1 (11β-hydroxylase) or CYP17A1 (17α-hydroxylase) deficiency (Fig. 398-1). Progesterone can cause hypokalemic hypertension in rare individuals who harbor a MR mutation that enhances binding and activation by progesterone; physiologically, progesterone normally exerts antimineralocorticoid activity. Finally, excess mineralocorticoid activity can be caused by mutations in the β or γ subunits of the ENaC, disrupting its interaction with Nedd4 (Fig. 398-7), and thereby decreasing receptor internalization and degradation. The constitutively active ENaC drives hypokalemic hypertension, resulting in an autosomal dominant disorder termed *Liddle's syndrome*.

Clinical Manifestations Excess activation of the MR leads to potassium depletion and increased sodium retention, with the latter causing an expansion of extracellular and plasma volume. Increased ENaC activity also results in hydrogen depletion, which can cause metabolic alkalosis. Aldosterone also has direct effects on the vascular system, where it increases cardiac remodeling and decreases compliance. Aldosterone excess may cause direct damage to the myocardium and the kidney glomeruli, in addition to secondary damage due to systemic hypertension.

The clinical hallmark of mineralocorticoid excess is hypokalemic hypertension; however, only 10–40% of patients with primary aldosteronism exhibit hypokalemia. Serum sodium tends to be normal due to the concurrent fluid retention, which in some cases can lead to peripheral edema. Hypomagnesemia is also a common finding. Hypokalemia can be exacerbated by thiazide drug treatment, which leads to increased delivery of sodium to the distal renal tubule, thereby driving potassium excretion. Severe hypokalemia can be associated with polyuria, glucose intolerance, muscle weakness, overt proximal myopathy, or even arrhythmias, rhabdomyolysis, and hypokalemic paralysis. Severe alkalosis contributes to muscle cramps and, in severe cases, can cause tetany.

Of note, patients with primary aldosteronism show increased rates of osteoporosis, type 2 diabetes, and cognitive dysfunction. A significant proportion of patients with primary aldosteronism suffer from concurrent mild autonomous cortisol secretion (MACS), a constellation also termed *Connshing syndrome*.

Diagnosis Diagnostic screening for mineralocorticoid excess is not currently recommended for all patients with hypertension but should be restricted to those who exhibit hypertension associated with drug resistance, hypokalemia, an adrenal mass, onset of disease before the age of 40 years, or family history of primary aldosteronism **(Fig. 398-12)**. The accepted screening test is concurrent measurement of plasma renin and aldosterone with subsequent calculation of the aldosterone-renin ratio (ARR) (Fig. 398-12); serum potassium needs to be normalized prior to testing. Stopping antihypertensive medication can be cumbersome, particularly in patients with severe hypertension. Thus, for practical purposes, in the first instance, the patient can remain on the usual antihypertensive medications, with the exception that MR antagonists need to be ceased at least 4–6 weeks prior to ARR measurement. The remaining antihypertensive drugs may also affect the outcome of ARR testing (e.g., beta-blocker treatment can cause false-positive results and ACE/AT1R inhibitors can cause false-negative results in milder cases and may have to be discontinued for 2–4 weeks in case of repeat testing) **(Table 398-4)**. The decision to arrange washout from potentially interfering medications should be individualized.

Current international guidelines suggest a positive ARR screening result is >750 pmol/L per ng/mL per hour, with a concurrently high normal or increased aldosterone (Fig. 398-12). However, lower ARR cutoffs are routinely used in clinical practice since it is increasingly recognized that milder forms of primary aldosteronism may otherwise be missed. Another consideration is that if one relies on the ARR only, the likelihood of a false-positive ARR becomes greater when renin levels are very low. The characteristics of the biochemical assays are also important. Some labs measure plasma renin activity, whereas others measure plasma renin concentrations. Antibody-based assays for the measurement of serum aldosterone lack the reliability of tandem mass spectrometry assays, but these are not yet ubiquitously available.

Diagnostic confirmation of mineralocorticoid excess in a patient with a positive ARR screening result should be undertaken by an endocrinologist as the tests lack optimized validation. The most straightforward is the saline infusion test, which involves the IV administration of 2 L of physiologic saline over a 4-h period. Failure of aldosterone to suppress <170 pmol/L (6 ng/dL) in the seated position or <140 pmol/L (5 ng/dL) in the recumbent position is indicative of autonomous mineralocorticoid excess. Alternative tests are the oral sodium loading test (300 mmol NaCl/d for 3 days) or the fludrocortisone suppression test (0.1 mg q6h with 30 mmol NaCl q8h for 4 days); the latter can be difficult because of the risk of profound hypokalemia and increased hypertension. In patients with overt hypokalemic hypertension, strongly positive ARR, and concurrently increased aldosterone levels, confirmatory testing is usually not necessary.

Differential Diagnosis and Treatment After the diagnosis of hyperaldosteronism is established, the next step is to use adrenal imaging to further assess the cause. Fine-cut CT scanning of the adrenal region is the method of choice because it provides excellent visualization of adrenal morphology, and most aldosterone-producing adenomas are <2 cm. CT will readily identify larger tumors suspicious of malignancy but may miss lesions <5 mm. The differentiation between bilateral micronodular hyperplasia and a unilateral adenoma is only required if a surgical approach is feasible and desired. Consequently, selective adrenal vein sampling (AVS) should only be carried out in

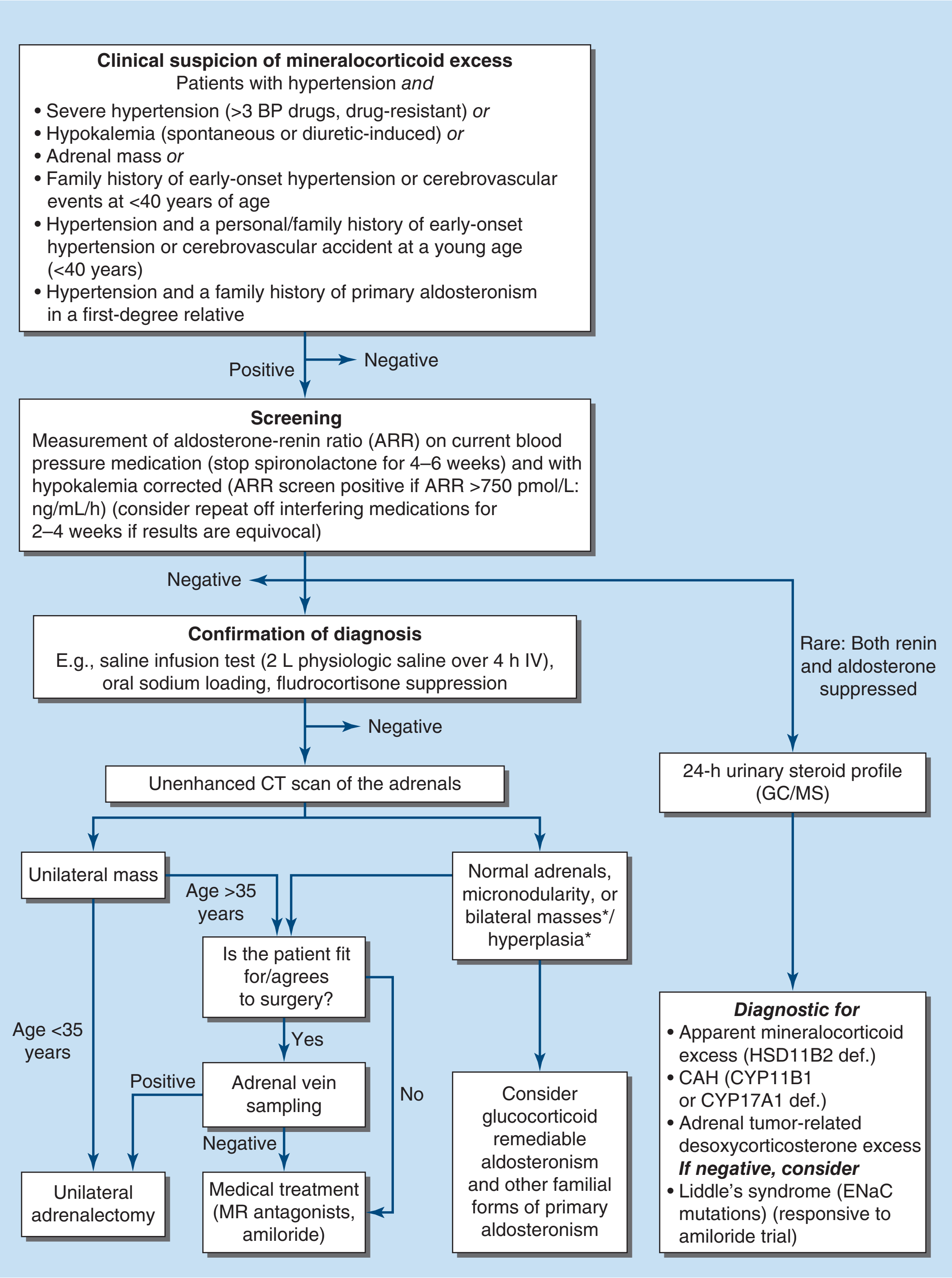
Clinical suspicion of mineralocorticoid excess
Patients with hypertension and
• Severe hypertension (>3 BP drugs, drug-resistant) or
• Hypokalemia (spontaneous or diuretic-induced) or
• Adrenal mass or
• Family history of early-onset hypertension or cerebrovascular events at <40 years of age
• Hypertension and a personal/family history of early-onset hypertension or cerebrovascular accident at a young age (<40 years)
• Hypertension and a family history of primary aldosteronism in a first-degree relative
Positive
Negative
Screening
Measurement of aldosterone-renin ratio (ARR) on current blood pressure medication (stop spironolactone for 4–6 weeks) and with hypokalemia corrected (ARR screen positive if ARR >750 pmol/L: ng/mL/h) (consider repeat off interfering medications for 2–4 weeks if results are equivocal)
Negative
Rare: Both renin and aldosterone suppressed
Confirmation of diagnosis
E.g., saline infusion test (2 L physiologic saline over 4 h IV), oral sodium loading, fludrocortisone suppression
Negative
Unenhanced CT scan of the adrenals
24-h urinary steroid profile (GC/MS)
Unilateral mass
Age >35 years
Normal adrenals, micronodularity, or bilateral masses*/ hyperplasia*
Is the patient fit for/agrees to surgery?
Age <35 years
Yes
No
Positive
Adrenal vein sampling
Negative
Unilateral adrenalectomy
Medical treatment (MR antagonists, amiloride)
Consider glucocorticoid remediable aldosteronism and other familial forms of primary aldosteronism
Diagnostic for
• Apparent mineralocorticoid excess (HSD11B2 def.)
• CAH (CYP11B1 or CYP17A1 def.)
• Adrenal tumor-related desoxycorticosterone excess
If negative, consider
• Liddle's syndrome (ENaC mutations) (responsive to amiloride trial)

FIGURE 398-12 Management of patients with suspected mineralocorticoid excess. Perform adrenal tumor workup (see Fig. 398-13). BP, blood pressure; CAH, congenital adrenal hyperplasia; CT, computed tomography; ENaC, epithelial sodium channel; GC/MS, gas chromatography/mass spectrometry; MR, mineralocorticoid receptor; PRA, plasma renin activity.

TABLE 398-4 Effects of Medications and Other Conditions on the Aldosterone-Renin Ratio (ARR)

EFFECT ON THE ARR	ANTIHYPERTENSIVE DRUGS	OTHER MEDICATIONS	OTHER CONDITIONS
Possible *false-negative* results (due to increase of renin and/or aldosterone)	MR antagonists; diuretics; DHP-calcium antagonists; ACE inhibitors; AT1R blockers; aliskiren[a]	SGLT2 inhibitors; SSRIs	Hypokalemia; dietary salt restriction; malignant hypertension; renovascular hypertension; pregnancy
Possible *false-positive* results (due to reduction of renin and/or aldosterone)	β-Blockers; clonidine; α-methyldopa; aliskiren[a]	NSAIDs; estrogen-containing oral contraceptives	Impaired renal function with hyperkalemia; luteal phase of the menstrual cycle in premenopausal women
Negligible effect	Non-DHP-calcium antagonists; α_1-adrenergic receptor antagonists; hydralazine	/	/

[a] The direct renin inhibitor aliskiren can cause false-negative results if the direct renin concentration is measured, as it causes a reduction of plasma renin activity.

Abbreviations: ACE, angiotensin-converting enzyme; AT1R, angiotensin II receptor type 1; DHP, dihydropyridine; MR, mineralocorticoid receptor; NSAID, nonsteroidal anti-inflammatory drug; SGLT2, sodium-glucose cotransporter 2; SSRI, selective serotonin reuptake inhibitor.

surgical candidates with either no obvious lesion on CT or evidence of a unilateral lesion but with age >35 years because the latter patients have a higher likelihood of harboring a coincidental, endocrine-inactive adrenal adenoma (Fig. 398-12). AVS is used to compare aldosterone levels in the inferior vena cava and between the right and left adrenal veins. AVS requires concurrent measurement of cortisol to document the correct placement of the catheter in the adrenal veins and should demonstrate a cortisol gradient >2–3 in baseline conditions between the vena cava and each adrenal vein. Lateralization is confirmed by an aldosterone/cortisol ratio that is at least twofold higher on one side than the other in baseline conditions. AVS is a complex procedure that requires a highly skilled interventional radiologist. Even then, the right adrenal vein can be difficult to cannulate correctly, which, if not achieved, invalidates the procedure. There is also no agreement as to whether the two adrenal veins should be cannulated simultaneously or successively and whether ACTH stimulation enhances the diagnostic value of AVS. Recently, PET-CT with a labeled form of metomidate, a methyl analogue of etomidate that binds to both CYP11B1 and CYP11B2, has been validated as a noninvasive alternative to AVS. [^{11}C] metomidate PET-CT with dexamethasone pretreatment to suppress CYP11B1 activity can differentiate unilateral from bilateral forms of primary aldosteronism and has similar performance to AVS in predicting biochemical and clinical success following adrenalectomy.

Patients <35 years with confirmed primary aldosteronism, a unilateral lesion on CT, and no clinical suspicion of familial forms can go straight to surgery, which is also indicated in patients with confirmed lateralization documented by a valid AVS procedure or [^{11}C]metomidate PET-CT. Laparoscopic adrenalectomy is the preferred approach. Patients who are not surgical candidates, or with evidence of bilateral hyperplasia based on CT or AVS, should be treated medically (Fig. 398-12). Medical treatment, which can also be considered prior to surgery to avoid postsurgical hypoaldosteronism, consists primarily of the MR antagonist spironolactone. It can be started at 12.5–50 mg daily and titrated up to a maximum of 400 mg/d to control blood pressure and normalize potassium. Side effects include menstrual irregularity, decreased libido, and gynecomastia. The more selective MR antagonist eplerenone can also be used. Doses start at 25 mg bid, and it can be titrated up to 200 mg/d. Another useful drug is the potassium-sparing diuretic amiloride (5–10 mg bid).

In patients with normal adrenal morphology and family history of early-onset, severe hypertension, a diagnosis of GRA should be considered and can be evaluated using genetic testing. Treatment of GRA consists of administering dexamethasone, using the lowest dose possible to control blood pressure. Some patients also require additional MR antagonist treatment.

The diagnosis of non-aldosterone-related mineralocorticoid excess is based on documentation of suppressed renin and suppressed aldosterone in the presence of hypokalemic hypertension. This testing is best carried out by employing urinary steroid metabolite profiling by gas chromatography/mass spectrometry (GC/MS). An increased free cortisol over free cortisone ratio is suggestive of SAME and can be treated with dexamethasone. Steroid profiling by GC/MS also detects the steroids associated with CYP11B1 and CYP17A1 deficiency or the irregular steroid secretion pattern in a DOC-producing ACC (Fig. 398-12). If the GC/MS profile is normal, Liddle's syndrome should be considered. It is very sensitive to amiloride treatment but will not respond to MR antagonist treatment because the defect is due to a constitutively active ENaC.

APPROACH TO THE PATIENT: INCIDENTALLY DISCOVERED ADRENAL MASS

Epidemiology Incidentally discovered adrenal masses, commonly termed adrenal "incidentalomas," are common, with a prevalence of 2–5% in the general population as documented in CT and autopsy series. The prevalence increases with age, with 1% of 40-year-olds and 7% of 70-year-olds harboring an adrenal mass. The widespread use of cross-sectional imaging has also increased the recognized prevalence.

Etiology Most solitary adrenal tumors are monoclonal neoplasms. Several genetic syndromes, including MEN 1 (*MEN1*), MEN 2 (*RET*), Carney's complex (*PRKAR1A*), McCune-Albright (*GNAS1*), Li Fraumeni (*TP53*), Lynch (*MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM*), and familial adenomatous polyposis (*APC*) can have adrenocortical tumors as one of their features. On the other hand, Von-Hippel-Lindau disease (*VHL*), MEN 2, paraganglioma syndrome type 1/4/5 (*SDHD*/*SDHB*/*SDHA*), neurofibromatosis type 1 (*NF1*), and hereditary leiomyomatosis and renal cell cancer (*FH*) are among the genetic syndromes associated with pheochromocytomas.

TABLE 398-5 Classification of Unilateral Adrenal Masses

MASS	APPROXIMATE PREVALENCE (%)
Benign	
Adrenocortical adenoma	
Endocrine-inactive	40–70
Cortisol-producing (mild autonomous cortisol secretion)	20–50
Aldosterone-producing	2–5
Cortisol-producing (overt Cushing's syndrome)	1–4
Pheochromocytoma	1–5
Adrenal myelolipoma	3-6
Adrenal ganglioneuroma	1
Adrenal cyst and pseudocyst	1
Adrenal hematoma/hemorrhagic infarction	<1
Adrenal hemangioma	<0.1
Indeterminate	
Adrenocortical oncocytoma	<1
Malignant	
Metastases (most frequent: breast, lung)	3–7
Adrenocortical carcinoma	0.4–4
Malignant pheochromocytoma	<1
Adrenal neuroblastoma	<0.1
Lymphomas (including primary adrenal lymphoma)	<0.1

Note: Bilateral adrenal enlargement/masses may be caused by congenital adrenal hyperplasia, bilateral macronodular hyperplasia, bilateral hemorrhage (due to antiphospholipid syndrome or sepsis-associated Waterhouse-Friderichsen syndrome), granuloma, amyloidosis, or infiltrative disease including tuberculosis.

Up to 50% of adrenal nodules are hormonally active, due to a cortisol- or aldosterone-producing adrenocortical adenoma or a pheochromocytoma associated with catecholamine excess **(Table 398-5)**. ACC is rare but is the cause of an adrenal mass in up to 4% of patients. However, metastases originating from another solid tissue tumor are an additional cause of adrenal incidentaloma and account for up to 40% of adrenal masses in patients undergoing imaging for tumor staging or follow-up monitoring (Table 398-5).

Differential Diagnosis and Treatment Patients with an adrenal mass >1 cm require a diagnostic evaluation. Two key questions need to be addressed: (1) Does the tumor autonomously secrete hormones that could have a detrimental effect on health? (2) Is the adrenal mass benign or malignant?

Hormone secretion by an adrenal mass occurs along a continuum, with a gradual increase in clinical manifestations in parallel with hormone levels. Exclusion of catecholamine excess from a pheochromocytoma arising from the adrenal medulla is a mandatory part of the diagnostic workup of lipid-poor masses **(Fig. 398-13)**. In case of hypertension, primary aldosteronism should be ruled out. Overproduction of adrenal androgen precursors, DHEA and its sulfate, is rare and most frequently seen in the context of ACC, as are increased levels of steroid precursors such as 17OHP. Cortisol excess is the most observed hormonal abnormality in adrenal masses, ranging from rare clinically overt Cushing's syndrome to much more prevalent MACS (early-morning cortisol >50 nmol/L [1.8 μg/dL] after overnight administration of dexamethasone 1 mg in the absence of clinical features of Cushing's). Patients with MACS may exhibit one or more components of the metabolic syndrome (e.g., obesity, type 2 diabetes, hypertension,

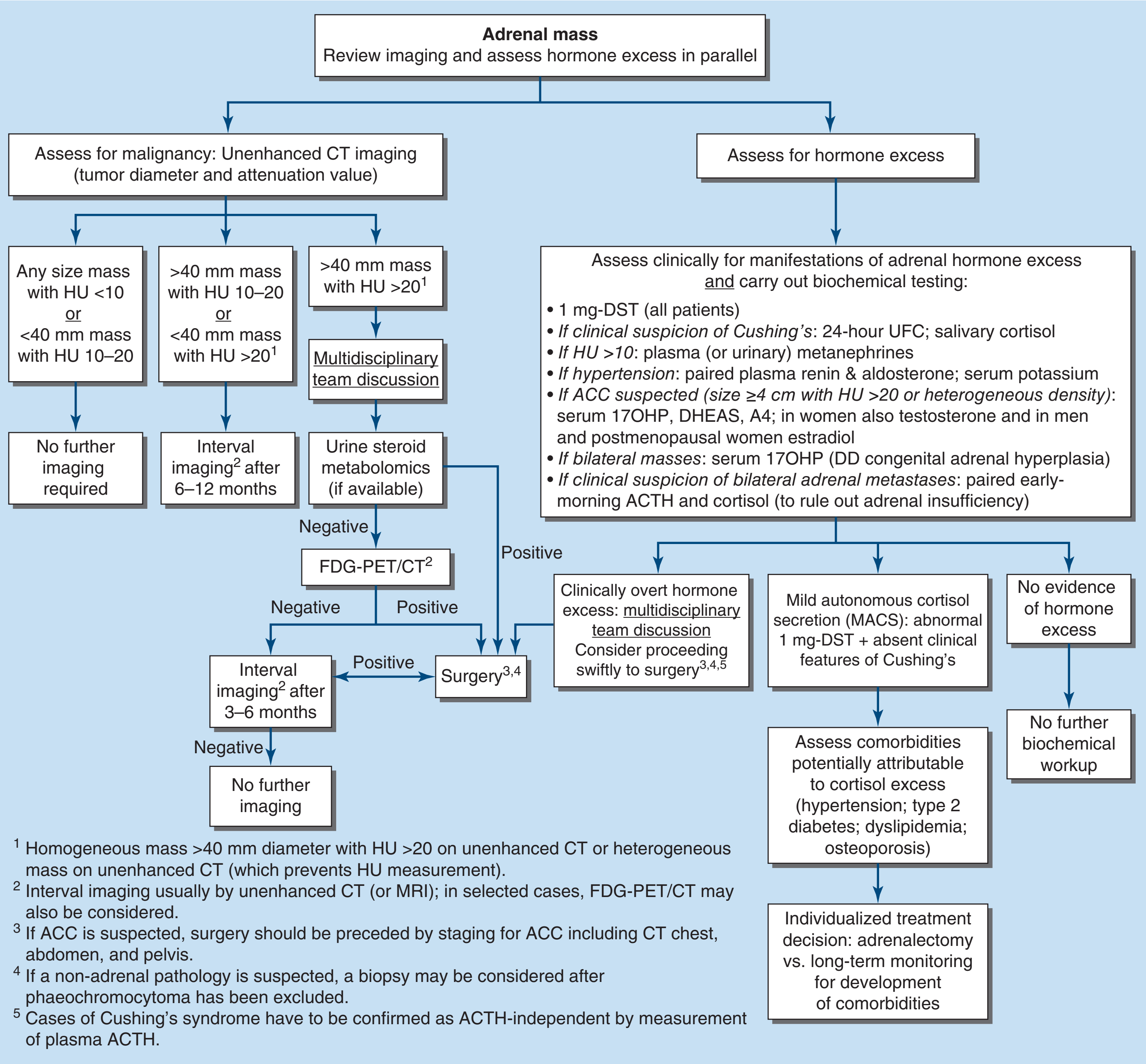
Adrenal mass
Review imaging and assess hormone excess in parallel
Assess for malignancy: Unenhanced CT imaging (tumor diameter and attenuation value)
Assess for hormone excess
Any size mass with HU <10 or <40 mm mass with HU 10–20
>40 mm mass with HU 10–20 or <40 mm mass with HU >20¹
>40 mm mass with HU >20¹
Multidisciplinary team discussion
No further imaging required
Interval imaging² after 6–12 months
Urine steroid metabolomics (if available)
Negative
Positive
FDG-PET/CT²
Negative
Positive
Interval imaging² after 3–6 months
Positive
Surgery³,⁴
Negative
No further imaging
Assess clinically for manifestations of adrenal hormone excess and carry out biochemical testing:
• 1 mg-DST (all patients)
• If clinical suspicion of Cushing's: 24-hour UFC; salivary cortisol
• If HU >10: plasma (or urinary) metanephrines
• If hypertension: paired plasma renin & aldosterone; serum potassium
• If ACC suspected (size ≥4 cm with HU >20 or heterogeneous density): serum 17OHP, DHEAS, A4; in women also testosterone and in men and postmenopausal women estradiol
• If bilateral masses: serum 17OHP (DD congenital adrenal hyperplasia)
• If clinical suspicion of bilateral adrenal metastases: paired early-morning ACTH and cortisol (to rule out adrenal insufficiency)
Clinically overt hormone excess: multidisciplinary team discussion Consider proceeding swiftly to surgery³,⁴,⁵
Mild autonomous cortisol secretion (MACS): abnormal 1 mg-DST + absent clinical features of Cushing's
No evidence of hormone excess
Assess comorbidities potentially attributable to cortisol excess (hypertension; type 2 diabetes; dyslipidemia; osteoporosis)
No further biochemical workup
Individualized treatment decision: adrenalectomy vs. long-term monitoring for development of comorbidities

[1] Homogeneous mass >40 mm diameter with HU >20 on unenhanced CT or heterogeneous mass on unenhanced CT (which prevents HU measurement).

[2] Interval imaging usually by unenhanced CT (or MRI); in selected cases, FDG-PET/CT may also be considered.

[3] If ACC is suspected, surgery should be preceded by staging for ACC including CT chest, abdomen, and pelvis.

[4] If a non-adrenal pathology is suspected, a biopsy may be considered after phaeochromocytoma has been excluded.

[5] Cases of Cushing's syndrome have to be confirmed as ACTH-independent by measurement of plasma ACTH.

FIGURE 398-13 Management of the patient with an incidentally discovered adrenal mass. This pathway is primarily designed for adrenal incidentalomas; in adrenal nodules identified by screening or staging for an extra-adrenal primary malignancy, the likelihood of a metastasis needs to be considered when arranging further imaging or a biopsy. ACC, adrenocortical carcinoma; ACTH, adrenocorticotropic hormone; CT, computed tomography; DD, differential diagnosis; DHEAS, dehydroepiandrosterone sulfate; 1 mg-DST, 1 mg dexamethasone overnight suppression test; FDG-PET, fluorodeoxyglucose positron emission tomography; HU, Hounsfield Units; 17-OHP, 17-hydroxyprogesterone; MACS, mild autonomous cortisol excess secretion; MRI, magnetic resonance imaging; UFC, urinary free cortisol.

dyslipidemia), osteoporosis, increased risk of cardiovascular events, frailty, and impaired quality of life, and have increased mortality risk. There is an ongoing debate about the optimal treatment for these patients, which includes adrenalectomy or conservative management with long-term monitoring and treatment of comorbidities potentially attributable to the cortisol excess.

For the differentiation of benign from malignant adrenal masses, imaging is relatively sensitive, although specificity is suboptimal. Unenhanced CT is the procedure of choice for imaging the adrenal glands (Fig. 398-11). A diagnosis of ACC, pheochromocytoma, and benign adrenal myelolipoma becomes more likely with increasing diameter of the adrenal mass. However, size alone is of poor predictive value, with only 80% specificity for the differentiation of benign from malignant masses when using a 4-cm cutoff. Metastases are rare but are found with similar frequency in adrenal masses of all sizes. The tumor attenuation value on unenhanced CT is of high diagnostic value, as many adrenocortical adenomas are lipid rich and thus present with low attenuation values (i.e., densities of <20 HUs). However, similar numbers of adrenocortical adenomas are lipid poor and present with higher HUs, making it difficult to differentiate them from ACCs, as well as pheochromocytomas, both of which invariably have high attenuation values (i.e., densities >20 HU on precontrast scans). Generally, benign lesions are rounded and homogenous, whereas most malignant lesions appear lobulated and inhomogeneous. Pheochromocytoma and adrenomyelolipoma may also exhibit lobulated and inhomogeneous features. MRI also allows for the visualization of the adrenal glands with somewhat lower resolution than CT. However, because it does not involve exposure to ionizing radiation, it is preferred in children, young adults, and during pregnancy. MRI has a valuable role in the characterization of indeterminate adrenal lesions using chemical shift analysis, with malignant tumors rarely showing loss of signal on opposed-phase MRI; however, this may also be observed in a proportion of benign adrenocortical adenomas, and chemical shift analysis results usually are similar to those obtained by measuring tumor attenuation on unenhanced CT. The evidence base for the utility of fluorodeoxyglucose (FDG)-PET was initially scarce but is now expanding, with recent data indicating that a lack of FDG uptake reliably indicates a benign adrenocortical mass; however, care has to be taken in patients with a history of extra-adrenal cancer as adrenal metastases arising from those sometimes do not take up FDG, i.e., as typically seen in renal cell carcinoma. Adrenocortical carcinoma is invariably FDG avid; however, specificity is poor, with many benign adrenal adenomas also taking up FDG.

Fine-needle aspiration (FNA) or CT-guided biopsy of an adrenal mass is very rarely indicated. FNA of a pheochromocytoma can cause a life-threatening hypertensive crisis. FNA of an ACC violates the tumor capsule and can cause needle track metastasis. FNA should only be considered in a patient with a history of nonadrenal malignancy and a newly detected adrenal mass, after careful exclusion of pheochromocytoma, and if the outcome will influence therapeutic management. It is important to recognize that in 25% of patients with a previous history of nonadrenal malignancy, a newly detected mass on CT is not a metastasis. While FNA can diagnose extra-adrenal malignancies, it has very limited ability to differentiate between benign and malignant adrenocortical lesions and hence should not be used for the diagnosis of ACC.

Adrenal masses associated with confirmed hormone excess or suspected malignancy are usually treated surgically (Fig. 398-13) or, if adrenalectomy is not feasible or desired, with medication. Preoperative exclusion of glucocorticoid excess is particularly important for the prediction of postoperative suppression of the contralateral adrenal gland, which requires glucocorticoid replacement peri- and postoperatively. Adrenal masses of any size with normal endocrine biochemistry at diagnosis and an attenuation value of <10 HU on unenhanced CT can be considered benign (malignancy rate <0.5%) and do not require further follow-up (Fig. 398-13). This similarly applies to adrenal masses <4 cm in diameter and an attenuation value of 10–20 HU. In adrenal masses <4 cm and >20 HU attenuation as well as in masses >4 cm and 10–20 HU, the likelihood of malignancy is still very low (<5%) and can be mitigated by arranging for interval imaging (Fig. 398-13). The overwhelming majority of risk of malignancy applies to adrenal masses >4 cm with an attenuation >20 HU, and these masses require swift attention (Fig. 398-13), with a 50% risk of underlying malignancy. The risk of adrenocortical carcinoma is higher in young patients (<40 years), whereas metastases of extra-adrenal primaries occur at all ages. In adrenal masses with suspicious imaging findings (>4 cm and >20 HU), surgery and/or biopsy are feasible options; however, the latter will still result in unnecessary surgery for many benign tumors. A recently introduced diagnostic test, urine steroid metabolomics, has a twofold higher positive predictive value than imaging in detecting adrenocortical carcinoma, based on a distinct "malignant steroid fingerprint" with accumulating precursor steroid metabolites in 24-h urine.

■ ADRENOCORTICAL CARCINOMA

ACC is a rare malignancy with an annual incidence of 1–2 per million population. ACC is generally considered a highly malignant tumor; however, it presents with broad interindividual variability with regard to biologic characteristics and clinical behavior. Somatic mutations in the tumor-suppressor gene *TP53* are found in 25% of apparently sporadic ACC. Germline *TP53* mutations are the cause of the Li-Fraumeni syndrome associated with multiple solid organ cancers including ACC and are found in 25% of pediatric ACC cases; the *TP53* mutation R337H is found in almost all pediatric ACC in Brazil. Other genetic changes identified in ACC include germline mutations of genes affecting DNA mismatch repair (Lynch syndrome), *PRKAR1A* (Carney complex), *MEN1*, *APC* (familial adenomatous polyposis), and somatic alterations in the Wnt/β-catenin pathway and in the insulin-like growth factor 2 (IGF2) cluster. IGF2 overexpression is found in 90% of ACCs.

Patients with large adrenal tumors suspicious of malignancy should be managed by a multidisciplinary specialist team, including an endocrinologist, an oncologist, a surgeon, a radiologist, and a histopathologist. FNA is not indicated in suspected ACC: first, cytology and also histopathology of a core biopsy cannot differentiate between benign and malignant primary adrenal masses; second, FNA violates the tumor capsule and may even cause needle canal metastasis. Even when the entire tumor specimen is available, the histopathologic differentiation between benign and malignant adrenocortical lesions is a diagnostic challenge. The most common histopathologic classification is the Weiss score, taking into account high nuclear grade; mitotic rate (>5/high-power field [HPF]); atypical mitosis; <25% clear cells; diffuse architecture; and presence of necrosis, venous invasion, and invasion of sinusoidal structures and tumor capsule. The presence of three or more elements suggests ACC. However, FNA is a feasible option if looking for metastases of an extra-adrenal primary or other adrenal tumor entities, such as ganglioneuroma.

TABLE 398-6 Classification System for Staging of Adrenocortical Carcinoma

ENSAT STAGE	TNM STAGE	TNM DEFINITIONS
I	T1,N0,M0	T1, tumor ≤5 cm N0, no positive lymph node M0, no distant metastases
II	T2,N0,M0	T2, tumor >5 cm N0, no positive lymph node M0, no distant metastases
III	T1–T2,N1,M0 T3–T4,N0–N1,M0	N1, positive lymph node(s) M0, no distant metastases T3, tumor infiltration into surrounding tissue T4, tumor invasion into adjacent organs *or* venous tumor thrombus in vena cava or renal vein
IV	T1–T4,N0–N1,M1	M1, presence of distant metastases

Abbreviations: ENSAT, European Network for the Study of Adrenal Tumors; TNM, tumor, node, metastasis.

Although 60–70% of ACCs show biochemical evidence of steroid overproduction, in many patients, this is not clinically apparent due to the relatively inefficient steroid production by the adrenocortical cancer cells. Excess production of glucocorticoids and adrenal androgen precursors are most common and indicative of malignancy.

Tumor staging at ACC diagnosis **(Table 398-6)** has important prognostic implications and requires scanning of the chest and abdomen for local organ invasion, lymphadenopathy, and metastases. Intravenous contrast medium is necessary for maximum sensitivity for hepatic metastases. An adrenal origin may be difficult to determine on standard axial CT imaging if the tumors are large and invasive, but CT reconstructions and MRI are more informative **(Fig. 398-14)** using multiple planes and different sequences. Vascular and adjacent organ invasion is diagnostic of malignancy. 18-Fluoro-2-deoxy-D-glucose PET (18-FDG-PET) is highly sensitive for the detection of malignancy and can be used to detect small metastases or local recurrence that may not be obvious on CT (Fig. 398-14). However, FDG-PET has limited specificity and therefore cannot be used for differentiating benign from malignant adrenal lesions. Metastasis in ACC most frequently occurs to liver and lung.

There is no established grading system for ACC, and the Weiss score carries no prognostic value; the most important prognostic histopathologic parameter is the Ki67 proliferation index, with Ki67 <10% indicative of slow to moderate growth velocity, whereas a Ki67 ≥10% is associated with poor prognosis including high risk of recurrence and rapid progression.

Cure of ACC can only be achieved by early detection and complete surgical removal. Capsule violation during primary surgery, metastasis at diagnosis, and primary treatment in a nonspecialist center and by a nonspecialist surgeon are major determinants of poor survival. If the primary tumor invades adjacent organs, en bloc removal of kidney and spleen should be considered to reduce the risk of recurrence, and regional lymph node dissection may further reduce this risk. Surgery

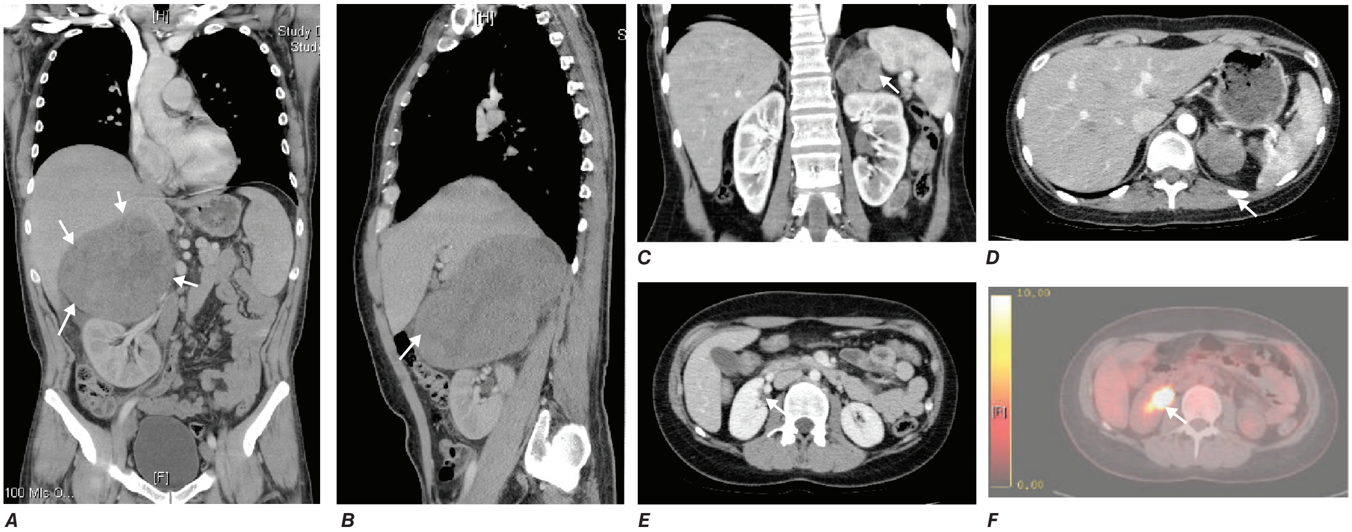

FIGURE 398-14 Imaging in adrenocortical carcinoma (ACC). Magnetic resonance imaging scan with (***A***) frontal and (***B***) lateral views of a right ACC that was detected incidentally. Computed tomography (CT) scan with (***C***) coronal and (***D***) transverse views depicting a right-sided ACC. Note the irregular border and inhomogeneous structure. CT scan (***E***) and positron emission tomography/CT (***F***) visualizing a peritoneal metastasis of an ACC in close proximity to the right kidney (*arrow*).

can also be considered in a patient with metastases if there is severe tumor-related hormone excess. This indication needs to be carefully weighed against surgical risk, including thromboembolic complications, and the resulting delay in the introduction of other therapeutic options. Patients with confirmed ACC and successful removal of the primary tumor should receive adjuvant treatment with mitotane (o,p'DDD), particularly in patients with a high risk of recurrence as determined by European Network for the Study of Adrenal Tumors (ENSAT) stage III and IV or Ki67 proliferation index ≥10%. Adjuvant mitotane should be continued for at least 2 years, if side effects are tolerated. Regular monitoring of plasma mitotane levels is mandatory (therapeutic range 14–20 mg/L; neurotoxic complications more frequent at >20 mg/L). Mitotane is usually started at 500 mg tid, with stepwise increases to a maximum dose of 2000 mg tid in days (high-dose saturation) or weeks (low-dose saturation) as tolerated. Once therapeutic range plasma mitotane levels are achieved, the dose can be tapered to maintenance doses mostly ranging from 1000 to 1500 mg tid. Mitotane treatment results in disruption of cortisol synthesis and thus requires glucocorticoid replacement; glucocorticoid replacement dose should be at least double of that usually used in adrenal insufficiency (i.e., 40–60 mg daily in two to three divided doses) because mitotane induces hepatic CYP3A4 activity, resulting in rapid inactivation of glucocorticoids. Mitotane also increases circulating CBG, thereby decreasing the available free cortisol fraction. Single metastases can be addressed surgically or with radiofrequency ablation as appropriate. If the tumor recurs or progresses during mitotane treatment, cytotoxic chemotherapy should be considered; the established first-line chemotherapy regimen is the combination of cisplatin, etoposide, and doxorubicin plus continuing mitotane. Painful bone metastasis responds to irradiation. Overall survival in ACC is still poor, with 5-year survival rates of 30–40% and a median survival of 15 months in metastatic ACC.

■ ADRENAL INSUFFICIENCY

Epidemiology The prevalence of well-documented, permanent adrenal insufficiency is 5 in 10,000 in the general population. Hypothalamic-pituitary origin of the disease is most frequent, with a prevalence of 3 in 10,000, whereas primary adrenal insufficiency has a prevalence of 2 in 10,000. Approximately one-half of the latter cases are acquired, mostly caused by autoimmune destruction of the adrenal glands; the other one-half are genetic, most commonly caused by distinct enzymatic blocks in adrenal steroidogenesis affecting glucocorticoid synthesis (i.e., CAH).

Adrenal insufficiency arising from suppression of the HPA axis consequent to exogenous glucocorticoid treatment is much more common, occurring in 0.5–2% of the population in developed countries.

Etiology *Primary adrenal insufficiency* is most commonly caused by autoimmune adrenalitis. Isolated autoimmune adrenalitis accounts for 30–40%, whereas 60–70% develop adrenal insufficiency as part of autoimmune polyglandular syndromes (APSs) **(Chap. 400) (Table 398-7)**. APS1, also termed APECED (autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy), is the underlying cause in 10% of patients affected by APS. APS1 is transmitted in an autosomal recessive manner and is caused by mutations in the autoimmune regulator gene *AIRE*. Associated autoimmune conditions overlap with those seen in APS2 but may also include total alopecia, primary hypoparathyroidism, and, in rare cases, lymphoma. APS1 patients invariably develop chronic mucocutaneous candidiasis, usually manifested in childhood and preceding adrenal insufficiency by years or decades. The much more prevalent APS2 is of polygenic inheritance, with confirmed associations with the *HLA-DR3* gene region in the major histocompatibility complex and distinct gene regions involved in immune regulation (*CTLA-4, PTPN22, CLEC16A*). Coincident autoimmune disease most frequently includes thyroid autoimmune disease, vitiligo, and premature ovarian failure. Less commonly, additional features may include type 1 diabetes and pernicious anemia caused by vitamin B_{12} deficiency.

X-linked adrenoleukodystrophy has an incidence of 1:20,000 males and is caused by mutations in the *X-ALD* gene encoding the peroxisomal membrane transporter protein ABCD1; its disruption results in the accumulation of very-long-chain (>24 carbon atoms) fatty acids. Approximately 50% of cases manifest in early childhood with rapidly progressive white matter disease (cerebral adrenoleukodystrophy); 35% present during adolescence or in early adulthood with neurologic features indicative of myelin and peripheral nervous system involvement (adrenomyeloneuropathy [AMN]). In the remaining 15%, adrenal insufficiency is the sole manifestation of disease. Of note, distinct mutations manifest with variable penetrance and phenotypes within affected families.

Rarer causes of adrenal insufficiency involve the destruction of the adrenal glands as a consequence of infection, hemorrhage, or infiltration (Table 398-7); tuberculous adrenalitis is still a frequent cause of disease in developing countries. Adrenal metastases rarely cause adrenal insufficiency, and this occurs only with bilateral, bulky metastases.

Inborn causes of primary adrenal insufficiency other than CAH are rare, causing <1% of cases. However, their elucidation provides important insights into adrenal gland development and physiology. Mutations causing primary adrenal insufficiency (Table 398-7) include factors regulating adrenal development and steroidogenesis (DAX-1, SF-1), cholesterol synthesis, import and cleavage (DHCR7, StAR, CYP11A1), elements of the adrenal ACTH response pathway (MC2R, MRAP) (Fig. 398-5), and factors involved in redox regulation (NNT, TXNRD2) and DNA repair (MCM4, CDKN1C).

TABLE 398-7 Causes of Primary Adrenal Insufficiency

DIAGNOSIS	GENE	ASSOCIATED FEATURES
Autoimmune polyglandular syndrome 1 (APS1)	*AIRE*	Hypoparathyroidism, chronic mucocutaneous candidiasis, other autoimmune disorders, rarely lymphomas
Autoimmune polyglandular syndrome 2 (APS2)	Associations with HLA-DR3, CTLA-4	Hypothyroidism, hyperthyroidism, premature ovarian failure, vitiligo, type 1 diabetes mellitus, pernicious anemia
Isolated autoimmune adrenalitis	Associations with HLA-DR3, CTLA-4	
Congenital adrenal hyperplasia (CAH)	*CYP21A2, CYP11B1, CYP17A1, HSD3B2, POR*	See Table 398-10 (see also **Chap. 402**)
Congenital lipoid adrenal hyperplasia (CLAH)	*STAR, CYP11A1*	46,XY DSD, gonadal failure (see also **Chap. 402**)
Adrenal hypoplasia congenita (AHC)	*NR0B1 (DAX-1), NR5A1 (SF-1)*	46,XY DSD, gonadal failure (see also **Chap. 402**)
Adrenoleukodystrophy (ALD), adrenomyeloneuropathy (AMN)	*ABCD1*	Demyelination of central nervous system (ALD) or spinal cord and peripheral nerves (AMN)
Familial glucocorticoid deficiency	*MC2R*	Tall stature
	MRAP	None
	STAR	None
	NNT	None
	TXNRD2	None
	MCM4	Growth retardation, natural killer cell deficiency
Triple A syndrome	*AAAS*	Alacrima, achalasia, neurologic impairment
Smith-Lemli-Opitz syndrome	*SLOS*	Cholesterol synthesis disorder associated with mental retardation, craniofacial malformations, growth failure
Kearns-Sayre syndrome	Mitochondrial DNA deletions	Progressive external ophthalmoplegia, pigmentary retinal degeneration, cardiac conduction defects, gonadal failure, hypoparathyroidism, type 1 diabetes,
IMAGe syndrome	*CDKN1C*	Intrauterine growth retardation, metaphyseal dysplasia, genital anomalies
MIRAGE syndrome	*SAMD9*	Myelodysplasia, infection, restriction of growth, genital phenotypes, and enteropathy
Sphingosine-1-phosphate lyase deficiency	*SGPL1*	Steroid-resistant nephrotic syndrome, immunodeficiency, neurologic defects, ichthyosis, primary hypothyroidism, cryptorchidism
Adrenal infections		Tuberculosis, HIV, CMV, cryptococcosis, histoplasmosis, coccidioidomycosis
Adrenal infiltration		Metastases, lymphomas, sarcoidosis, amyloidosis, hemochromatosis
Adrenal hemorrhage		Meningococcal sepsis (Waterhouse-Friderichsen syndrome), primary antiphospholipid syndrome
Drug-induced		Mitotane, aminoglutethimide, abiraterone, trilostane, etomidate, ketoconazole, levoketoconazole, osilodrostat, suramin, mifepristone (RU486), interferon-alpha, immune checkpoint inhibitors (rare)
Bilateral adrenalectomy		E.g., in the management of Cushing's syndrome or after bilateral nephrectomy

Abbreviations: AIRE, autoimmune regulator; CMV, cytomegalovirus; DSD, disordered sex development; MC2R, ACTH receptor; MCM4, mini chromosome maintenance-deficient 4 homologue; MRAP, MC2R-accessory protein; NNT, nicotinamide nucleotide transhydrogenase.

Secondary (or central) adrenal insufficiency is the consequence of dysfunction of the hypothalamic-pituitary component of the HPA axis **(Table 398-8)**. Excluding iatrogenic suppression, the overwhelming majority of cases are caused by pituitary or hypothalamic tumors or their treatment by surgery or irradiation **(Chap. 392)**. Rarer causes include pituitary apoplexy, either as a consequence of an infarcted pituitary adenoma or transient reduction in the blood supply of the pituitary during surgery or after rapid blood loss associated with parturition, also termed Sheehan's syndrome. Isolated ACTH deficiency is rarely caused by autoimmune disease or pituitary infiltration (Table 398-8). Mutations in the ACTH precursor POMC or in factors regulating pituitary development are genetic causes of ACTH deficiency (Table 398-8).

Clinical Manifestations In principle, the clinical features of primary adrenal insufficiency (Addison's disease) are characterized by the loss of both glucocorticoid and mineralocorticoid secretion **(Table 398-9)**. In secondary adrenal insufficiency, only glucocorticoid deficiency is present, as the adrenal itself is intact and thus still amenable to regulation by the RAA system. Adrenal androgen secretion is disrupted in both primary and secondary adrenal insufficiency (Table 398-9). Hypothalamic-pituitary disease can lead to additional clinical manifestations due to involvement of other endocrine axes (thyroid, gonads, GH, prolactin) or visual impairment with bitemporal hemianopia caused by chiasmal compression. It is important to recognize that iatrogenic adrenal insufficiency caused by exogenous glucocorticoid suppression of the HPA axis may result in all symptoms associated with glucocorticoid deficiency (Table 398-9), if exogenous glucocorticoids are stopped abruptly. However, patients will appear clinically cushingoid as a result of the preceding overexposure to glucocorticoids.

Chronic adrenal insufficiency manifests with relatively nonspecific signs and symptoms, such as fatigue and loss of energy, often resulting in delayed or missed diagnoses (e.g., as depression or anorexia). A distinguishing feature of primary adrenal insufficiency is hyperpigmentation, which is caused by excess ACTH stimulation of melanocytes. Hyperpigmentation is most pronounced in skin areas exposed to increased friction or shear stress and is increased by sunlight **(Fig. 398-15)**. Conversely, in secondary adrenal insufficiency, the skin has an alabaster-like paleness due to lack of ACTH secretion.

Hyponatremia is a characteristic biochemical feature in primary adrenal insufficiency and is found in 80% of patients at presentation. Hyperkalemia is present in 40% of patients at initial diagnosis. Hyponatremia is primarily caused by mineralocorticoid deficiency but can also occur in secondary adrenal insufficiency due to diminished inhibition of antidiuretic hormone (ADH) release by cortisol, resulting in mild syndrome of inappropriate secretion of antidiuretic hormone (SIADH). Glucocorticoid deficiency also results in slightly increased TSH concentrations that normalize within days to weeks after initiation of glucocorticoid replacement.

TABLE 398-8 Causes of Secondary Adrenal Insufficiency

DIAGNOSIS	GENE	ASSOCIATED FEATURES
Pituitary tumors (endocrine active and inactive adenomas, very rare: carcinoma)		Depending on tumor size and location: visual field impairment (bilateral hemianopia), hyperprolactinemia, secondary hypothyroidism, hypogonadism, growth hormone deficiency
Other mass lesions affecting the hypothalamic-pituitary region		Craniopharyngioma, meningioma, ependymoma, metastases
Pituitary irradiation		Radiotherapy administered for pituitary tumors, brain tumors, or craniospinal irradiation in leukemia
Autoimmune hypophysitis		Often associated with pregnancy; may present with panhypopituitarism or isolated ACTH deficiency; can be associated with autoimmune thyroid disease, more rarely with vitiligo, premature ovarian failure, type 1 diabetes, pernicious anemia
Pituitary apoplexy/hemorrhage		Hemorrhagic infarction of large pituitary adenomas or pituitary infarction consequent to traumatic major blood loss (e.g., surgery or pregnancy: Sheehan's syndrome)
Pituitary infiltration		Tuberculosis, actinomycosis, sarcoidosis, histiocytosis X, granulomatosis with polyangiitis (Wegener's), metastases
Drug-induced		Chronic glucocorticoid excess (endogenous or exogenous), immune check point inhibitors, opioids, interferon alpha, ribavirin, megestrol acetate
Congenital isolated ACTH deficiency	*TBX19* (Tpit)	
Combined pituitary hormone deficiency (CPHD)	*PROP-1*	Progressive development of CPHD in the order GH, PRL, TSH, LH/FSH, ACTH
	HESX1	CPHD and septo-optic dysplasia
	LHX3	CPHD and limited neck rotation, sensorineural deafness
	LHX4	CPHD and cerebellar abnormalities
	SOX3	CPHD and variable mental retardation
Proopiomelanocortin (POMC) deficiency	*POMC*	Early-onset obesity, red hair pigmentation

Abbreviations: ACTH, adrenocorticotropic hormone; GH, growth hormone; LH/FSH, luteinizing hormone/follicle-stimulating hormone; PRL, prolactin; TSH, thyroid-stimulating hormone.

TABLE 398-9 Signs and Symptoms of Adrenal Insufficiency

Signs and Symptoms Caused by Glucocorticoid Deficiency
Fatigue, lack of energy
Weight loss, anorexia
Myalgia, joint pain
Fever
Normochromic anemia, lymphocytosis, eosinophilia
Slightly increased TSH (due to loss of feedback inhibition of TSH release)
Hypoglycemia (more frequent in children)
Low blood pressure, postural hypotension
Hyponatremia (due to loss of feedback inhibition of AVP release)
Signs and Symptoms Caused by Mineralocorticoid Deficiency (Primary Adrenal Insufficiency Only)
Abdominal pain, nausea, vomiting
Dizziness, postural hypotension
Salt craving
Low blood pressure, postural hypotension
Increased serum creatinine (due to volume depletion)
Hyponatremia
Hyperkalemia
Signs and Symptoms Caused by Adrenal Androgen Deficiency
Lack of energy
Dry and itchy skin (in women)
Loss of libido (in women)
Loss of axillary and pubic hair (in women)
Other Signs and Symptoms
Hyperpigmentation (primary adrenal insufficiency only) (due to excess of proopiomelanocortin [POMC]-derived peptides)
Alabaster-colored pale skin (secondary adrenal insufficiency only) (due to deficiency of POMC-derived peptides)

Abbreviations: AVP, arginine vasopressin; TSH, thyroid-stimulating hormone.

Acute adrenal insufficiency, also termed adrenal crisis, usually occurs after a prolonged period of nonspecific complaints and is more frequently observed in patients with primary adrenal insufficiency, due to the loss of both glucocorticoid and mineralocorticoid secretion. Postural hypotension may progress to hypovolemic shock. Adrenal insufficiency may mimic features of acute abdomen with abdominal tenderness, nausea, vomiting, and fever. In some cases, the primary presentation may resemble neurologic disease, with decreased responsiveness progressing to stupor and coma. An adrenal crisis can be triggered by an intercurrent illness, surgical or other stress, or increased glucocorticoid inactivation (e.g., hyperthyroidism). Prospective data indicate 8.3 adrenal crises and 0.5 adrenal crisis–related deaths per 100 patient-years.

Diagnosis The diagnosis of adrenal insufficiency is established by the short cosyntropin test, a safe and reliable tool with excellent predictive diagnostic value **(Fig. 398-16)**. The cutoff for failure is usually defined at cortisol levels of <450–500 nmol/L (16–18 μg/dL) sampled 30–60 min after ACTH stimulation; the exact cutoff is dependent on the locally available assay, with generally lower cutoffs for mass spectrometry–based assays. During the early phase of HPA disruption (e.g., within 4 weeks of pituitary insufficiency), patients may still respond to exogenous ACTH stimulation. In this circumstance, the ITT is an alternative choice but is more invasive and should be carried out only under a specialist's supervision (see above). Induction of hypoglycemia is contraindicated in individuals with diabetes mellitus, cardiovascular disease, or history of seizures. Random serum cortisol measurements are of limited diagnostic value because baseline cortisol levels may be coincidentally low due to the physiologic diurnal rhythm of cortisol secretion (Fig. 398-3). Similarly, many patients with secondary adrenal insufficiency have relatively normal baseline cortisol levels but fail to mount an appropriate cortisol response to ACTH, which can only be revealed by stimulation testing. Importantly, tests to establish the diagnosis of adrenal insufficiency should never delay treatment. Thus, in a patient with suspected adrenal crisis, it is reasonable to draw baseline cortisol levels, provide replacement therapy, and defer formal stimulation testing until a later time.

Once adrenal insufficiency is confirmed, measurement of plasma ACTH is the next step, with increased or inappropriately low levels defining primary and secondary origin of disease, respectively (Fig. 398-16). In primary adrenal insufficiency, increased plasma renin will confirm the presence of mineralocorticoid deficiency. At initial presentation, patients with primary adrenal insufficiency should undergo screening for steroid autoantibodies as a marker of autoimmune adrenalitis. If these tests are negative, adrenal imaging by CT is indicated to investigate possible hemorrhage, infiltration, or masses. In male patients with negative autoantibodies in the plasma, very-long-chain fatty acids should be measured to exclude X-ALD. Patients with inappropriately low ACTH, in the presence of confirmed cortisol deficiency, should undergo hypothalamic-pituitary imaging by MRI. Features suggestive of preceding pituitary apoplexy, such as sudden-onset severe headache or history of previous head trauma, should be carefully explored, particularly in patients with no obvious MRI lesion.

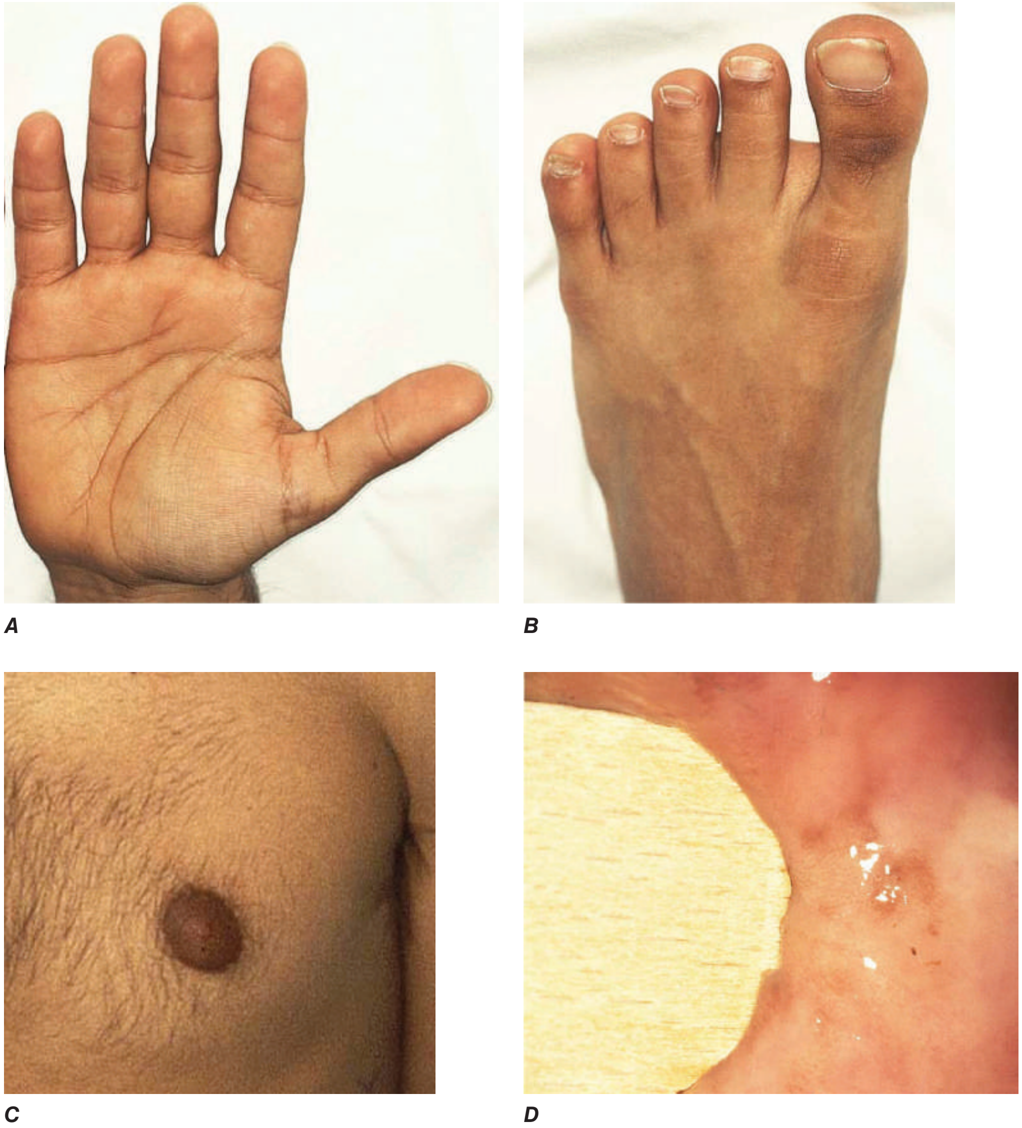

FIGURE 398-15 Clinical features of Addison's disease. Note the hyperpigmentation in areas of increased friction including (***A***) palmar creases, (***B***) dorsal foot, (***C***) nipples and axillary region, and (***D***) patchy hyperpigmentation of the oral mucosa.

TREATMENT
Acute Adrenal Insufficiency

Acute adrenal insufficiency requires immediate initiation of rehydration, usually carried out by saline infusion at initial rates of 1 L/h with continuous cardiac monitoring. Glucocorticoid replacement should be initiated by bolus injection of 100 mg hydrocortisone, followed by the administration of 200 mg hydrocortisone over 24 h, preferably by continuous infusion or alternatively by bolus IV or IM injections. Mineralocorticoid replacement can be initiated once the daily hydrocortisone dose has been reduced to <50 mg because at higher doses hydrocortisone provides sufficient stimulation of MRs.

Glucocorticoid replacement for the treatment of chronic adrenal insufficiency should be administered at a dose that replaces the physiologic daily cortisol production, which is usually achieved by the oral administration of 15–25 mg hydrocortisone in two to three divided doses. Pregnancy may require an increase in hydrocortisone dose by 50% during the last trimester. In all patients, at least one-half of the daily dose should be administered in the morning upon awakening. Long-acting glucocorticoids such as prednisolone or dexamethasone are not preferred because they result in increased glucocorticoid exposure due to extended GR activation at times of physiologically low cortisol secretion. There are no well-established dose equivalencies, but as a guide, equipotency can be assumed for 1 mg hydrocortisone, 1.6 mg cortisone acetate, 0.2 mg prednisolone, 0.25 mg prednisone, and 0.025 mg dexamethasone. Currently available standard glucocorticoid preparations fail to mimic the physiologic cortisol secretion rhythm (Fig. 398-3). However, this is overcome by recent modified release hydrocortisone preparations, with promising results emerging from the treatment of patients with primary adrenal insufficiency due to congenital adrenal hyperplasia.

Monitoring of glucocorticoid replacement is mainly based on the history and examination for signs and symptoms suggestive of glucocorticoid over- or underreplacement, including assessment of body weight and blood pressure. Plasma ACTH, 24-h urinary free cortisol, or serum cortisol day curves reflect whether hydrocortisone has been taken or not but do not convey reliable information about replacement quality. In patients with isolated primary adrenal insufficiency, monitoring should include screening for autoimmune thyroid disease, and female patients should be made aware of the possibility of premature ovarian failure. Supraphysiologic glucocorticoid treatment with doses equivalent to 30 mg hydrocortisone or more will affect bone metabolism, and these patients should undergo regular bone mineral density evaluation. All patients with adrenal insufficiency need to be instructed about the requirement for stress-related glucocorticoid dose adjustments. These generally consist of doubling the routine oral glucocorticoid dose in the case of intercurrent illness with fever and bed rest and the need for immediate IV or IM injection of 100 mg hydrocortisone followed

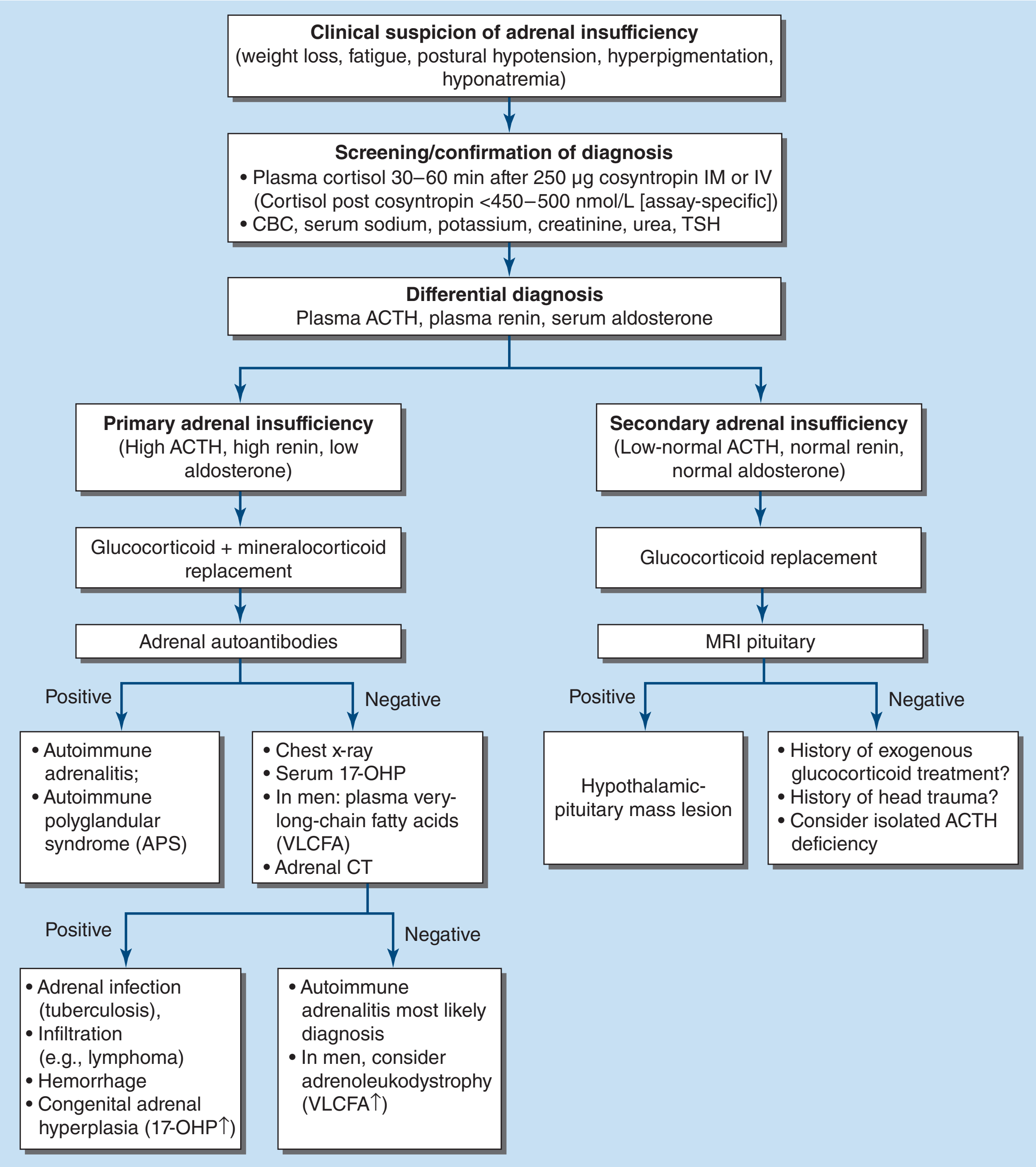
Clinical suspicion of adrenal insufficiency
(weight loss, fatigue, postural hypotension, hyperpigmentation, hyponatremia)
Screening/confirmation of diagnosis
• Plasma cortisol 30–60 min after 250 μg cosyntropin IM or IV (Cortisol post cosyntropin <450–500 nmol/L [assay-specific])
• CBC, serum sodium, potassium, creatinine, urea, TSH
Differential diagnosis
Plasma ACTH, plasma renin, serum aldosterone
Primary adrenal insufficiency
(High ACTH, high renin, low aldosterone)
Secondary adrenal insufficiency
(Low-normal ACTH, normal renin, normal aldosterone)
Glucocorticoid + mineralocorticoid replacement
Glucocorticoid replacement
Adrenal autoantibodies
MRI pituitary
Positive
Negative
Positive
Negative
• Autoimmune adrenalitis;
• Autoimmune polyglandular syndrome (APS)
• Chest x-ray
• Serum 17-OHP
• In men: plasma very-long-chain fatty acids (VLCFA)
• Adrenal CT
Hypothalamic-pituitary mass lesion
• History of exogenous glucocorticoid treatment?
• History of head trauma?
• Consider isolated ACTH deficiency
Positive
Negative
• Adrenal infection (tuberculosis),
• Infiltration (e.g., lymphoma)
• Hemorrhage
• Congenital adrenal hyperplasia (17-OHP↑)
• Autoimmune adrenalitis most likely diagnosis
• In men, consider adrenoleukodystrophy (VLCFA↑)

FIGURE 398-16 Management of the patient with suspected adrenal insufficiency. ACTH, adrenocorticotropic hormone; CBC, complete blood count; CT, computed tomography; MRI, magnetic resonance imaging; 17-OHP, 17-hydroxyprogesterone; PRA, plasma renin activity; TSH, thyroid-stimulating hormone.

by intravenous infusion of 200 mg hydrocortisone/24 h in cases of prolonged vomiting, surgery, or trauma. All patients, but in particular those living or traveling in regions with delayed access to acute health care, should carry a hydrocortisone self-injection emergency kit, in addition to their usual steroid emergency cards and bracelets, and should receive training in its use.

Mineralocorticoid replacement in primary adrenal insufficiency should be initiated at a dose of 100–150 μg fludrocortisone. The adequacy of treatment can be evaluated by measuring blood pressure, sitting and standing, to detect a postural drop indicative of hypovolemia. In addition, serum sodium, potassium, and plasma renin should be measured regularly. Renin levels should be kept in the upper normal reference range. Changes in glucocorticoid dose may also impact on mineralocorticoid replacement as cortisol also binds the MR; 40 mg of hydrocortisone is equivalent to 100 μg of fludrocortisone. It is important to note that prednisone and prednisolone have reduced mineralocorticoid activity and dexamethasone has none. In patients living or traveling in areas with hot or tropical weather conditions, the fludrocortisone dose should be increased by 50–100 μg during the summer. Mineralocorticoid dose may also need to be adjusted during pregnancy due to the anti-mineralocorticoid activity of progesterone, but this is less often required than hydrocortisone dose adjustment. Plasma renin cannot serve as a monitoring tool during pregnancy because renin rises physiologically during gestation.

Adrenal androgen replacement is an option in patients with lack of energy, despite optimized glucocorticoid and mineralocorticoid replacement. It may also be indicated in women with features of androgen deficiency, including loss of libido. Adrenal androgen replacement can be achieved by once-daily administration of 25–50 mg DHEA. Treatment is monitored by measurement of DHEAS, androstenedione, testosterone, and sex hormone–binding globulin (SHBG) 24 h after the last DHEA dose.

■ GLUCOCORTICOID-INDUCED ADRENAL INSUFFICIENCY

A possible unwanted effect of glucocorticoid treatment is suppression of the hypothalamic-pituitary-adrenal axis. Glucocorticoid-induced adrenal insufficiency (GI-AI) can be observed with any route of exogenous administration, including oral, topical, inhaled, intranasal, and intraarticular. Factors affecting the risk of GI-AI include the duration of glucocorticoid therapy, mode of administration, glucocorticoid dose and potency, concomitant drugs that interfere with glucocorticoid metabolism, and individual susceptibility. Up to half of long-term oral glucocorticoid users (0.5–1.8% of the general population in Western countries) develop GI-AI: patients treated with supraphysiologic doses for longer than 3–4 weeks carry the highest risk, whereas short-term therapy is generally safe. GI-AI has been described in around 8% of inhaled glucocorticoid users, but the risk increases substantially (21–27%) when high doses are used and for treatment duration >1 year. GI-AI has also been reported in approximately half of patients receiving intra-articular steroid injections, but evidence in this population is very limited. The use of multiple glucocorticoid formulations at the same time and the concomitant administration of strong cytochrome P450 3A4 inhibitors, which include several antibiotics, antifungals, and the protease inhibitor ritonavir, are among the factors dramatically increasing the risk of developing GI-AI. Ritonavir is the most reported offending medication, used as part of antiviral combinations to treat HIV infection, hepatitis C infection, and COVID-19. Other drugs that can increase the risk are opioids, which can also blunt the HPA axis response.

Clinicians who prescribe glucocorticoids should educate patients about the different aspects of glucocorticoid therapy, including the risk of GI-AI. If glucocorticoids are no longer required for the control of the underlying disease and the treatment duration is <3–4 weeks, glucocorticoids can be stopped abruptly. Conversely, patients on long-term treatment should be tapered down gradually until approaching a physiologic dose (e.g., 3–5 mg of prednisone). Patients who have developed a dependence on supraphysiologic glucocorticoid doses can develop withdrawal symptoms during tapering including malaise, fatigue, nausea, muscle and joint pain, and sleep and mood disturbances. Symptoms can be severe, and patients can sometimes benefit from a temporary increase in their glucocorticoid dose and a subsequent slower taper. Once a physiologic glucocorticoid dose has been achieved, patients should be tested for HPA axis recovery through an early-morning serum cortisol measurement, which provides information on the likelihood of GI-AI. There are no established cortisol cutoffs, which are influenced by the assay used and other factors including concurrent major stress and altered CBG levels leading to falsely elevated (e.g., oral estrogens, pregnancy) or falsely reduced (e.g., advanced liver cirrhosis, nephrotic syndrome) cortisol levels. Current international guidelines suggest that an early-morning cortisol <150 nmol/L (or 5 μg/dL) is highly suggestive of GI-AI, whereas values >300 nmol/L (or 10 μg/dL) are in keeping with HPA axis recovery. For indeterminate values, cosyntropin stimulation may be considered. Several clinicians do not routinely measure cortisol levels in patients tapering glucocorticoids; in such cases, patients must be educated and closely monitored for clinical manifestations of adrenal insufficiency, and biochemical testing should be swiftly arranged if there are any clinical concerns.

Patients with established GI-AI must be treated as any other patient with secondary adrenal insufficiency. Other than daily glucocorticoid replacement, patients should receive stress dose coverage when they are exposed to stress to avoid acute adrenal insufficiency. Mineralocorticoid therapy is not required due to preserved aldosterone production.

CONGENITAL ADRENAL HYPERPLASIA

(See also Chap. 402) CAH is caused by mutations in genes encoding steroidogenic enzymes involved in glucocorticoid synthesis (*CYP21A2, CYP17A1, HSD3B2, CYP11B1*) or in the cofactor enzyme P450 oxidoreductase that serves as an electron donor to CYP21A2 and CYP17A1 (Fig. 398-1). Invariably, patients affected by CAH exhibit glucocorticoid deficiency. Depending on the exact step of enzymatic block, they may also have excess production of mineralocorticoids or deficient production of sex steroids **(Table 398-10)**. The diagnosis of CAH is readily established by measurement of the steroids accumulating before the distinct enzymatic block, either in serum or in urine, preferably by the use of mass spectrometry–based assays (Table 398-10).

Mutations in *CYP21A2* are the most prevalent cause of CAH, responsible for 90–95% of cases. 21-Hydroxylase deficiency disrupts glucocorticoid and mineralocorticoid synthesis (Fig. 398-1), resulting in diminished negative feedback via the HPA axis. This leads to increased pituitary ACTH release, which drives increased synthesis of adrenal androgen precursors and subsequent androgen excess. The degree of impairment of glucocorticoid and mineralocorticoid secretion depends on the severity of mutations. Major loss-of-function mutations result in combined glucocorticoid and mineralocorticoid deficiency (classic CAH, neonatal presentation), whereas less severe mutations affect glucocorticoid synthesis only (simple virilizing CAH, neonatal or early childhood presentation). The mildest mutations result in the least severe clinical phenotype, nonclassic CAH, usually presenting during adolescence and early adulthood and with preserved glucocorticoid production.

Androgen excess is present in all patients and manifests with broad phenotypic variability, ranging from severe virilization of the external genitalia in neonatal girls (e.g., 46,XX disordered sex development [DSD]) to hirsutism and oligomenorrhea resembling a polycystic ovary syndrome phenotype in young women with nonclassic CAH. In countries without neonatal screening for CAH, boys with classic CAH usually present with life-threatening adrenal crisis in the first few weeks of life (salt-wasting crisis); a simple-virilizing genotype manifests with precocious pseudo-puberty and advanced bone age in early childhood, whereas men with nonclassic CAH are usually detected only through family screening.

Glucocorticoid treatment is more complex than for other causes of primary adrenal insufficiency as it not only needed to replace missing glucocorticoids but also to control the increased ACTH drive and subsequent androgen excess. Current treatment is hampered by the lack of glucocorticoid preparations that mimic the diurnal cortisol secretion profile, resulting in a prolonged period of ACTH stimulation and subsequent androgen production during the early morning hours. In childhood, optimization of growth and pubertal development are important goals of glucocorticoid treatment, in addition to prevention of adrenal crisis and treatment of 46,XX DSD. In adults, the focus shifts

TABLE 398-10 Variants of Congenital Adrenal Hyperplasia

VARIANT	GENE	IMPACT ON STEROID SYNTHESIS	DIAGNOSTIC MARKER STEROIDS IN SERUM (AND URINE)
21-Hydroxylase deficiency (21OHD)	*CYP21A2*	Glucocorticoid deficiency, mineralocorticoid deficiency, adrenal androgen excess	17-Hydroxyprogesterone, 21-deoxycortisol (pregnanetriol, 17-hydroxypregnanolone, pregnanetriolone)
11β-Hydroxylase deficiency (11OHD)	*CYP11B1*	Glucocorticoid deficiency, mineralocorticoid excess, adrenal androgen excess	11-deoxycortisol, 11-deoxycorticosterone (tetrahydro-11-deoxycortisol, tetrahydro-11-deoxycorticosterone)
17α-Hydroxylase deficiency (17OHD)	*CYP17A1*	(Glucocorticoid deficiency), mineralocorticoid excess, androgen deficiency	11-Deoxycorticosterone, corticosterone, pregnenolone, progesterone (tetrahydro-11-deoxycorticosterone, tetrahydrocorticosterone, pregnenediol, pregnanediol)
3β-Hydroxysteroid dehydrogenase deficiency (3bHSDD)	*HSD3B2*	Glucocorticoid deficiency, (mineralocorticoid deficiency), adrenal androgen excess (females and males), gonadal androgen deficiency (males)	17-Hydroxypregnanolone (pregnanetriol)
P450 oxidoreductase deficiency (PORD)	*POR*	Glucocorticoid deficiency, (mineralocorticoid excess), prenatal androgen excess and postnatal androgen deficiency, skeletal malformations	Pregnenolone, progesterone, 17-hydroxyprogesterone (pregnanediol, pregnanetriol)

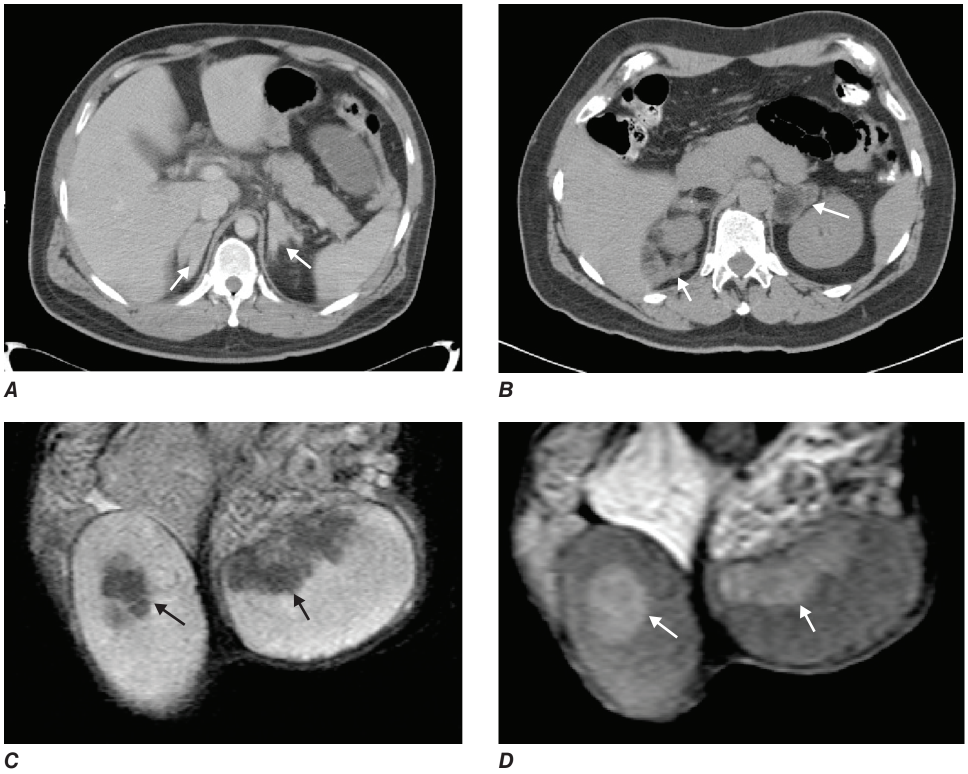

FIGURE 398-17 Imaging in congenital adrenal hyperplasia (CAH). Adrenal computed tomography scans showing homogenous bilateral hyperplasia in a young patient with classic CAH (***A***) and macronodular bilateral hyperplasia (***B***) in a middle-aged patient with classic CAH with longstanding poor disease control. Magnetic resonance imaging scan with T1-weighted (***C***) and T2-weighted (***D***) images showing bilateral testicular adrenal rest tumors (*arrows*) in a young patient with salt-wasting CAH. *(Used with permission from N. Reisch.)*

to preserving fertility and preventing side effects of glucocorticoid overtreatment, namely, the metabolic syndrome and osteoporosis. Fertility can be compromised in women due to oligomenorrhea/amenorrhea with chronic anovulation as a consequence of androgen excess. Men may develop testicular adrenal rest tissue (TART) **(Fig. 398-17)** consisting of hyperplastic cells with shared adrenal and gonadal characteristics located in the rete testis, which should not be confused with testicular tumors. TART can compromise sperm production and induce testicular fibrosis that may be irreversible.

TREATMENT

Congenital Adrenal Hyperplasia

Hydrocortisone is a good treatment option for the prevention of adrenal crisis, but longer acting prednisolone may be needed to control androgen excess. In children, hydrocortisone is given in divided doses at 1–1.5 times the normal cortisol production rate (~10–13 mg/m^2 per day). In adults, if hydrocortisone does not suffice, intermediate-acting glucocorticoids (e.g., prednisone) may be given, using the lowest dose necessary to suppress excess androgen production. For achieving fertility, dexamethasone treatment may be required but should only be given for the shortest possible time period to limit adverse metabolic side effects. The recent introduction of modified and delayed-release hydrocortisone, which mimics the endogenous physiologic cortisol release pattern, is promising, providing effective control of steroid precursor excess while the daily hydrocortisone dose is lower than required for immediate-release hydrocortisone. An oral corticotropin-releasing factor antagonist, Crinecerfont, has shown efficacy to reduce glucocorticoid replacement doses needed to suppress adrenal androgens.

Biochemical monitoring should include androstenedione and testosterone, aiming for the normal sex-specific reference range. 17OHP is a useful marker of overtreatment, indicated by 17OHP levels within the normal range of healthy controls. Glucocorticoid overtreatment may suppress the hypothalamic-pituitary-gonadal axis. Thus, treatment needs to be carefully titrated against clinical features of disease control. Stress-dose glucocorticoids should be given at double or triple the daily dose for surgery, acute illness, or severe trauma. Poorly controlled CAH can result in adrenocortical hyperplasia, which gave the disease its name, and may present as macronodular hyperplasia subsequent to long-standing ACTH excess (Fig. 398-17). The nodular areas can develop autonomous adrenal androgen production and may be unresponsive to glucocorticoid treatment. The prevalence of adrenomyelolipomas is increased in CAH; these are benign but can require surgical intervention due to lack of self-limiting growth.

Mineralocorticoid requirements change during life and are higher in children, explained by relative mineralocorticoid resistance that diminishes with ongoing maturation of the kidney. Children with CAH usually receive mineralocorticoid and salt replacement. However, young adults with CAH should undergo reassessment of their mineralocorticoid reserve. Plasma renin should be regularly monitored and kept within the upper half of the normal reference range.

FURTHER READING

Bancos I et al: Urine steroid metabolomics for the differential diagnosis of adrenal incidentalomas in the EURINE-ACT study: A prospective test validation study. Lancet Diabetes Endocrinol 8:773, 2020.

Beuschlein F et al: European Society of Endocrinology and Endocrine Society Joint Clinical Guideline: Diagnosis and therapy of glucocorticoid-induced adrenal insufficiency. Eur J Endocrinol 109:1657, 2024.

Claahsen-Van Der Grinten HL et al: Congenital adrenal hyperplasia: Current insights in pathophysiology, diagnostics and management. Endocr Rev 43:91, 2022.

Debono M et al: Home waking salivary cortisone to screen for adrenal insufficiency. NEJM Evid 2:EVIDoa2200182, 2023.

Fassnacht M et al: European Society of Endocrinology clinical practice guidelines on the management of adrenal incidentalomas, in collaboration with the European Network for the Study of Adrenal Tumors. Eur J Endocrinol 189:G1, 2023.

Fleseriu M et al: Consensus on diagnosis and management of Cushing's disease: A guideline update. Lancet Diabetes Endocrinol 9:847, 2021.

Hahner S et al: Adrenal insufficiency. Nat Rev Dis Primers 7:19, 2021.

Merke DP et al: Modified-release hydrocortisone in congenital adrenal hyperplasia. J Clin Endocrinol Metab 106:e2063, 2021.

Prete A et al: Prevention of adrenal crisis: Cortisol responses to major stress compared to stress dose hydrocortisone delivery. J Clin Endocrinol Metab 105:2262, 2020.

Prete A et al: Cardiometabolic disease burden and steroid excretion in benign adrenal tumors: A cross-sectional multicenter study. Ann Intern Med 175:325, 2022.

Wu X et al: [11C]metomidate PET-CT versus adrenal vein sampling for diagnosing surgically curable primary aldosteronism: a prospective, within-patient trial. Nat Med 29:190, 2023.

399

Pheochromocytoma

Frederic Castinetti,
Hartmut P. H. Neumann

Pheochromocytomas and paragangliomas (PPGLs) are catecholamine-producing tumors derived from the sympathetic or parasympathetic nervous system. These tumors may arise sporadically or be inherited as features of multiple endocrine neoplasia type 2 (MEN 2), von Hippel–Lindau (VHL) disease, or several other pheochromocytoma-associated syndromes. The diagnosis of pheochromocytomas identifies a potentially correctable cause of hypertension, and their removal can prevent hypertensive crises that can be lethal. The clinical presentation is variable, ranging from an adrenal incidentaloma to a hypertensive crisis with associated cerebrovascular or cardiac complications. Given the wide range of clinical signs, diagnosis may be delayed for years.

EPIDEMIOLOGY

The incidence of PPGL ranges from 0.04 to 0.95 cases per 100,000 per year, with a gradual increase over time probably due to genetic familial screening, changes in imaging modalities, and more frequent diagnosis as an incidentaloma. About ~0.1% of hypertensive patients harbor a pheochromocytoma. The mean age at diagnosis is ~40 years, although the tumors can be detected from early childhood, particularly when genetically determined, until late in life. The classic "rule of tens" for pheochromocytomas states that ~10% are bilateral, 10% are extra-adrenal, and 10% are metastatic.

ETIOLOGY AND PATHOGENESIS

PPGLs are well-vascularized tumors that arise from cells derived from the sympathetic (e.g., adrenal medulla or sympathetic trunk) or parasympathetic (e.g., carotid body, glomus tympanicum, glomus jugulare, glomus vagale) paraganglia **(Fig. 399-1)**. The name *pheochromocytoma* reflects the formerly used black-colored staining caused by chromaffin oxidation of catecholamines. The World Health Organization (WHO) applies the term *pheochromocytoma* to adrenal tumors (usually secreting) and the term *paraganglioma* to tumors at all other sites including head and neck, thoracic, extra-adrenal retroperitoneal, and pelvic sites.

The etiology of sporadic PPGLs is unknown. However, 25–33% of patients have an inherited condition, including germline mutations in the classically recognized *RET* (*re*arranged during *t*ransfection), *VHL*, *NF1* (neurofibromatosis type 1), *SDHB, SDHC,* and *SDHD* (subunits of SDH) genes or in the more recently recognized *SDHA, SDHAF2, TMEM127* (transmembrane protein 127), *MAX* (myc-associated factor X), *FH* (fumarate hydratase), *PDH1, PDH2* (pyruvate dehydrogenase), *HIF1α* and *HIF2α* (hypoxia-inducible factor), *MDH2* (malate dehydrogenase), *KIF1Bβ* (kinesin family member), *IDH1*, (isocitrate dehydrogenase 1), *SLC25A11* (oxoglutarate/malate), *H-RAS* (transforming protein p21), and *DNMTA3* (DNA methyltransferase 3 alpha) genes. Biallelic gene inactivation, a characteristic of tumor-suppressor genes, has been demonstrated for the *VHL, NF1, SDHx, TMEM127, MAX, FH, PDH1, PDH2, MDH2,* and *KIF1Bβ* genes. In contrast, *RET* is a protooncogene, and mutations activate receptor tyrosine kinase activity. Succinate dehydrogenase (SDH) is an enzyme of the Krebs cycle and the mitochondrial respiratory chain. The VHL protein is a component of a ubiquitin E3 ligase. *VHL* mutations reduce protein degradation, resulting in upregulation of components involved in cell-cycle progression, glucose metabolism, and oxygen sensing.

In addition to germline mutations, somatic mutations have been observed in >20 genes, broadly grouped into three different clusters of pathogenetically relevant genes: cluster 1, the pseudohypoxia group comprising mainly the genes *SDHx* (subunits of SDH), *FH*, *VHL*, and *HIF2A*; cluster 2, the kinase signaling group (*RET*, *NF1*, *TMEM127*, *MAX*, *HRAS*, *KIF1Bβ*, *PDH*); and cluster 3, the Wnt signaling group (*CSDE1*, *MAML3*).

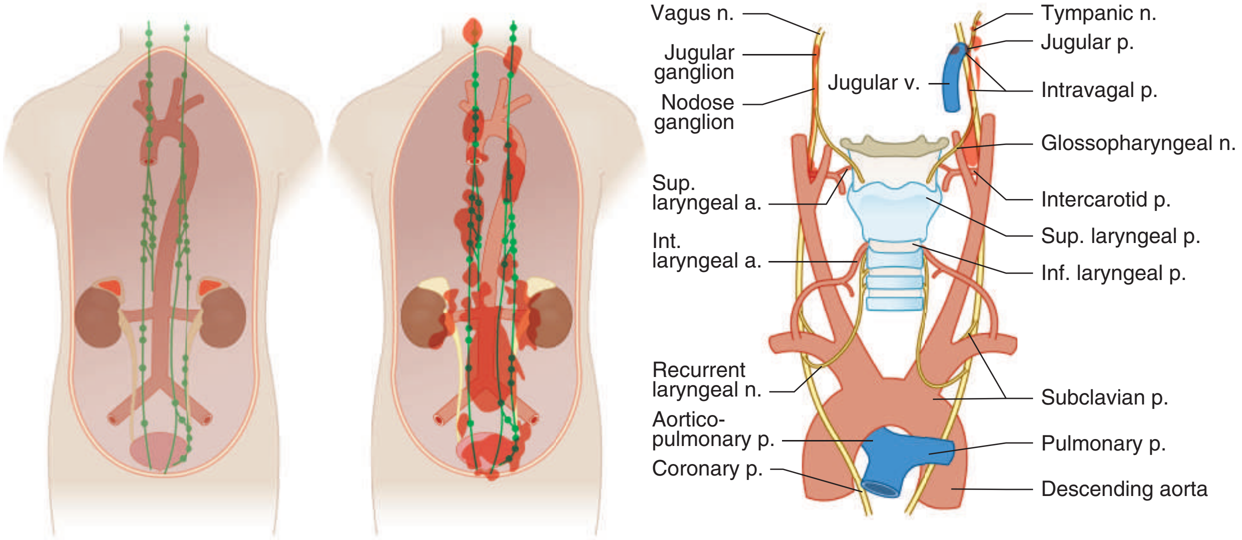

A Adrenal pheochromocytoma

B Extra-adrenal pheochromocytoma

C Head and neck paraganglioma

FIGURE 399-1 The paraganglial system and topographic sites (in red) of pheochromocytomas and paragangliomas. *(Parts A and B reproduced with permission from WM Manger, RW Gifford: Clinical and experimental pheochromocytoma. Cambridge: Blackwell Science; 1996.)*

TABLE 399-1 Clinical Features Associated with Pheochromocytoma, Listed by Frequency of Occurrence

1. Headaches
2. Profuse sweating
3. Palpitations and tachycardia
4. Hypertension, sustained or paroxysmal
5. Anxiety and panic attacks
6. Pallor
7. Nausea
8. Abdominal pain
9. Weakness
10. Weight loss
11. Paradoxical response to antihypertensive drugs
12. Polyuria and polydipsia
13. Constipation
14. Orthostatic hypotension
15. Dilated cardiomyopathy
16. Erythrocytosis
17. Elevated blood sugar
18. Hypercalcemia

TABLE 399-2 Biochemical and Imaging Methods Used for Diagnosis of Pheochromocytoma and Paraganglioma

DIAGNOSTIC METHOD	SENSITIVITY	SPECIFICITY
24-h urinary tests		
Catecholamines	+++	+++
Fractionated metanephrines	++++	++
Total metanephrines	+++	++++
Plasma tests		
Catecholamines	+++	++
Free metanephrines	++++	+++
Imaging		
CT	++++	+++
MRI	++++	+++
MIBG scintigraphy	++	++++
Somatostatin receptor scintigraphy[a]	++	++
^{18}Fluoro-DOPA PET/CT	++++	++++
^{68}Gallium-DOTATOC or DOTATATE PET/CT	++++	++++

[a]Values are particularly high in head and neck paragangliomas.

Abbreviations: CT, computed tomography; MIBG, metaiodobenzylguanidine; MRI, magnetic resonance imaging; PET/CT, positron emission tomography plus CT. For the biochemical tests, the ratings correspond globally to sensitivity and specificity rates as follows: ++, <85%; +++, 85–95%; and ++++, >95%.

CLINICAL FEATURES

Its clinical presentation is so variable that pheochromocytoma has been termed "the great masquerader" **(Table 399-1)**. Among the presenting manifestations, episodes of palpitation, headache, and profuse sweating are typical, and these manifestations constitute a classic triad that is seen in roughly a third of patients with pheochromocytoma. The presence of all three manifestations in association with hypertension makes pheochromocytoma a likely diagnosis. However, a pheochromocytoma can be asymptomatic for years or can be identified through imaging screening in a patient presenting with a hereditary syndrome. Some tumors grow to a considerable size before patients note symptoms.

The dominant sign is hypertension. Classically, patients have episodic hypertension, but sustained hypertension is also common. Catecholamine crises can lead to heart failure, pulmonary edema, arrhythmias, and intracranial hemorrhage. During episodes of hormone release, which can occur at widely divergent intervals, patients are anxious and pale, and they experience tachycardia and palpitations. These paroxysms generally last <1 h and may be precipitated by surgery, positional changes, exercise, pregnancy, urination (particularly with bladder pheochromocytomas), and various medications (e.g., tricyclic antidepressants, opiates, metoclopramide). Patients may present with true panic attacks that may be mistakenly attributed to psychiatric illness.

DIAGNOSIS

The diagnosis is based on documentation of catecholamine excess by biochemical testing and localization of the tumor by imaging. These two criteria are of equal importance, although measurement of catecholamines or metanephrines (their methylated metabolites) is traditionally the first step in diagnosis.

Biochemical Testing PPGLs synthesize and store catecholamines, which include norepinephrine (noradrenaline), epinephrine (adrenaline), and dopamine. Elevated plasma and urinary levels of catecholamines and metanephrines form the cornerstone of diagnosis. The characteristic fluctuations in the hormonal activity of tumors result in considerable variation in serial catecholamine measurements. However, most tumors continuously leak O-methylated metabolites, which are detected by measurement of metanephrines.

Catecholamines and metanephrines can be measured by different methods, including high-performance liquid chromatography, enzyme-linked immunosorbent assay, and liquid chromatography/mass spectrometry. When pheochromocytoma is suspected on clinical grounds (i.e., when values are three times the upper limit of normal), this diagnosis is highly likely regardless of the assay used. However, as summarized in **Table 399-2**, the sensitivity and specificity of available biochemical tests vary greatly, and these differences are important in assessing patients with borderline elevations of different compounds. Urinary tests for metanephrines (total or fractionated) and catecholamines are widely available and are used commonly for initial evaluation. Among these tests, those for the fractionated metanephrines and catecholamines are the most sensitive. Plasma tests are more convenient and include measurements of catecholamines and metanephrines. Measurements of plasma metanephrines are the most sensitive and are less susceptible to false-positive elevations from stress, including venipuncture. Although the incidence of false-positive test results has been reduced by the introduction of newer assays, physiologic stress responses and medications that increase catecholamine levels still can confound testing. Because the tumors are relatively rare, borderline elevations are likely to represent false-positive results. In this circumstance, it is important to exclude dietary or drug-related factors (withdrawal of levodopa or use of sympathomimetics, diuretics, tricyclic antidepressants, alpha and beta blockers) that might cause false-positive results and then to repeat testing.

Diagnostic Imaging A variety of methods have been used to localize PPGLs (Table 399-2, **Figs. 399-2, 399-3,** and **399-4**). Computed tomography (CT) and magnetic resonance imaging (MRI) are similar in sensitivity and should be performed with contrast. T2-weighted MRI with gadolinium contrast is optimal for detecting pheochromocytomas and is somewhat better than CT for imaging extra-adrenal PPGLs. About 5% of adrenal incidentalomas, which usually are detected by CT or MRI, prove to be pheochromocytomas upon endocrinologic evaluation, but the presence of pheochromocytomas is unlikely if unenhanced CT reveals an attenuation of <10 Hounsfield units (HU).

Tumors also can be localized by procedures using radioactive tracers, including ^{131}I- or ^{123}I-metaiodobenzylguanidine (MIBG) scintigraphy, ^{18}F-DOPA positron emission tomography (PET), ^{68}Ga-DOTATATE PET, or ^{18}F-fluorodeoxyglucose (FDG) PET (Fig. 399-2*B* and 399-4*A* and *B*). For PET-CT with both ^{68}Ga-DOTATATE and ^{18}F-DOPA, the sensitivity and specificity are very high (>95%). These agents are particularly useful in the documentation of hereditary syndromes but also in metastatic pheochromocytoma, because uptake is exhibited also in paragangliomas and metastases.

Pathology PPGLs are found at the classical sites of the adrenal medulla (Fig. 399-2) and paraganglia (Fig. 399-3). Histologically, the tumors often show a characteristic "Zellballen" pattern, consisting of nests of neuroendocrine chief cells with peripheral glial-like sustentacular cells. However, a broad spectrum of architectural and cytologic features can be seen. Immunohistochemistry is positive for chromogranin and synaptophysin in the chief cells and S-100 in the sustentacular cells **(Fig. 399-5*A*–*D*)**. Increasingly, staining with antibodies against the proteins encoded by susceptibility genes for hereditary pheochromocytomas, such as SDHB, is used to histologically demonstrate defects of these proteins, thereby making germline mutations more likely **(Fig. 399-5*E* and *F*)**.

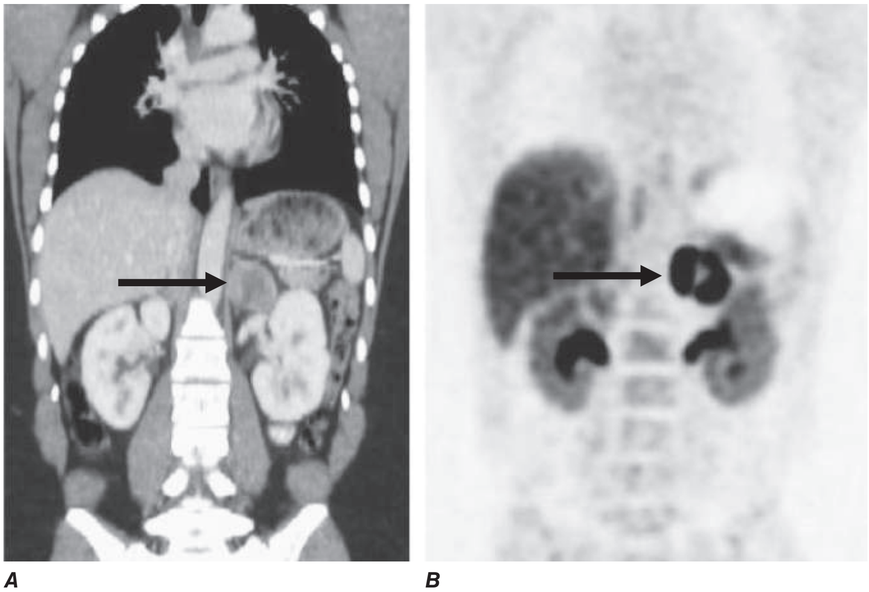

FIGURE 399-2 Typical pheochromocytoma (adrenal unilateral). ***A.*** Magnetic resonance imaging. ***B.*** ^{18}F-DOPA positron emission tomography (PET). Tumor marked by *arrows*. *(Part A was provided courtesy of Dr. Tobias Krauss, Freiburg. Part B was provided courtesy of Dr. Juri Ruf, Freiburg.)*

Differential Diagnosis When the possibility of a pheochromocytoma is being entertained, other disorders to consider include essential hypertension, anxiety attacks, use of cocaine or amphetamines, mastocytosis or carcinoid syndrome (usually without hypertension), intracranial lesions, clonidine withdrawal, autonomic epilepsy, and factitious crises (usually from use of sympathomimetic amines). When an asymptomatic adrenal mass is identified, likely diagnoses other than pheochromocytoma include a nonfunctioning adrenal adenoma, an aldosteronoma, and a cortisol-producing adenoma (Cushing's syndrome).

■ TREATMENT

Complete tumor removal, the ultimate therapeutic goal, can be achieved by partial or total adrenalectomy. It is important to preserve the normal adrenal cortex in order to prevent Addison's disease, particularly in hereditary disorders in which bilateral pheochromocytomas are most likely. Preoperative preparation of the patient has to be considered, and blood pressure should be consistently <160/90 mmHg. Classically, blood pressure has been controlled by α-adrenergic blockers (oral phenoxybenzamine, 0.5–4 mg/kg of body weight). Because patients are volume-constricted, liberal salt intake and hydration are necessary to avoid severe orthostasis. Oral prazosin or intravenous phentolamine can be used to manage paroxysms while adequate alpha blockade is awaited. Beta blockers (e.g., 10 mg of propranolol three or four times per day) should not be used as first-line treatment because of the risk of increased hypertension. Other antihypertensives, such as calcium channel blockers

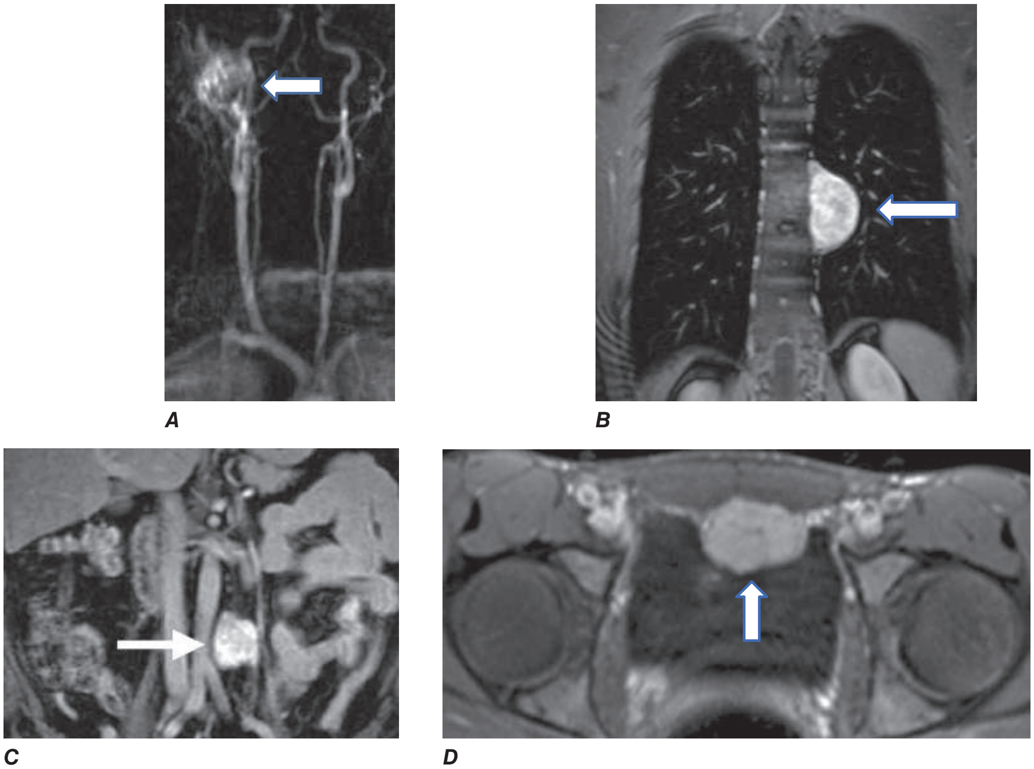

FIGURE 399-3 Paragangliomas (extra-adrenal pheochromocytomas). ***A.*** Carotid body tumor. ***B.*** Thoracic tumor. ***C.*** Paraaortal tumor. ***D.*** Pelvic tumor at the anterior wall of the urinary bladder. Tumors marked by *arrows*. *(Part A was provided courtesy of Dr. Carsten Boedeker, Stralsund. Parts B and D were provided courtesy of Dr. Tobias Krauss, Freiburg. Part C was provided courtesy of Dr. Martin Walz, Essen.)*

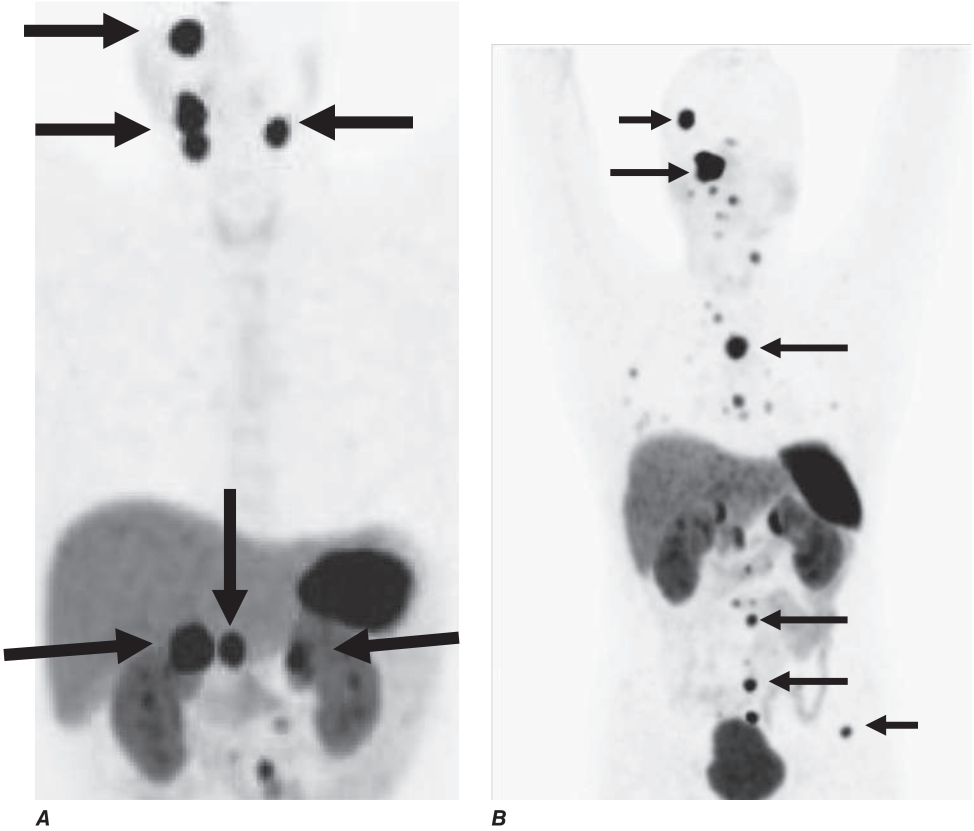

FIGURE 399-4 Multiple and metastatic pheochromocytoma. ***A.*** Paraganglioma syndrome. A patient with the *SDHD* W5X mutation and PGL1 ^{68}Ga-DOTATATE positron emission tomography (PET) demonstrating tumor uptake in the right jugular glomus, the right and left carotid body, both adrenal glands, and an interaortocaval paraganglion (*arrows*). Note the physiologic accumulation of the radiopharmaceutical agent in the kidneys and the liver. ***B.*** ^{18}F-DOPA PET of a patient with metastatic pheochromocytoma. Several metastases marked by *arrows*. *(Parts A and B were provided courtesy of Dr. Juri Ruf, Freiburg.)*

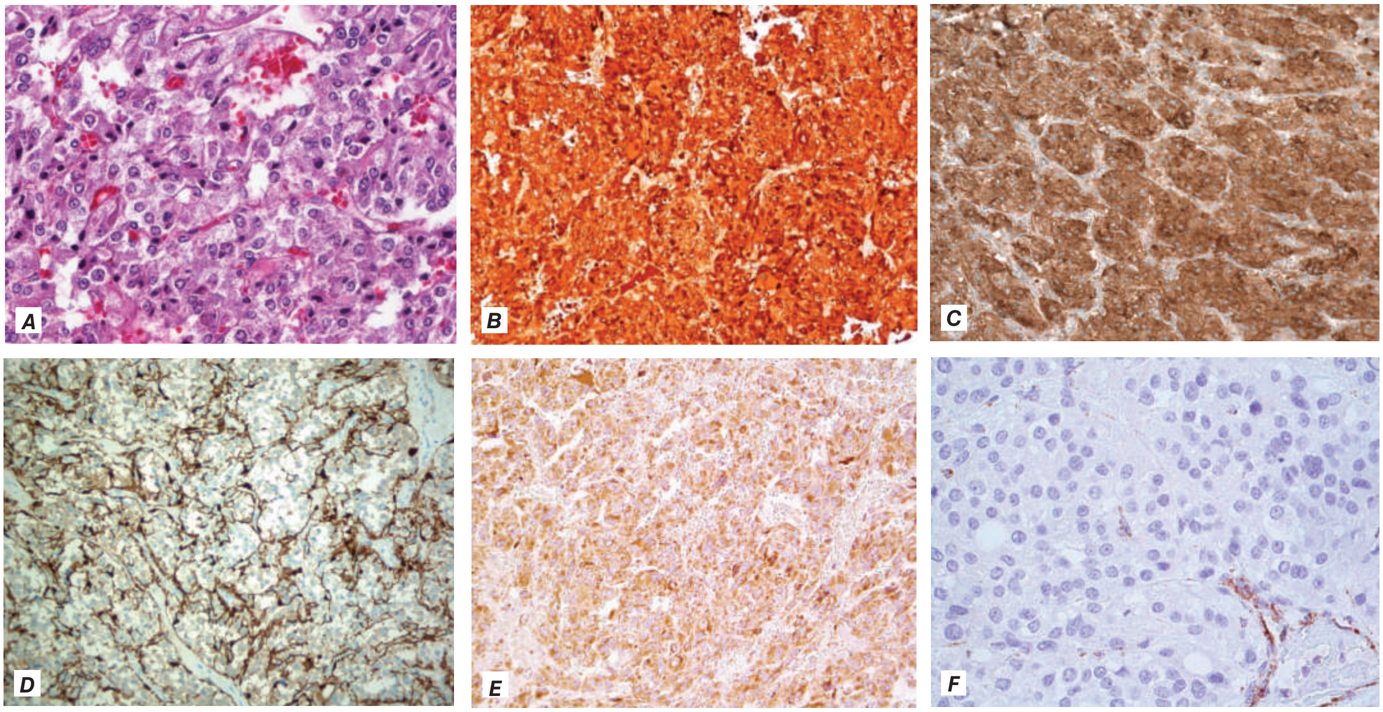

FIGURE 399-5 Histology and immunohistochemistry of pheochromocytoma. ***A.*** Hematoxylin and eosin, ***B.*** chromogranin, ***C.*** synaptophysin, ***C*** and ***B*** stain chief cells; ***D.*** S-100 stains sustentacular cells. ***E, F.*** Immunohistochemistry with SDHB antibody: positive staining (granular cytoplasmic staining) indicates intact SDHB (***E***), whereas negative staining (endothelial cells positive as internal control) (***F***) indicates structurally changed or absent SDHB due to a germline mutation in the *SDHB* gene, which was confirmed by molecular genetic analysis of a blood sample. *(Parts A–D and F were used with permission from Dr. Helena Leijon, Helsinki. Part E was provided courtesy of Dr. Kurt Werner Schmid, Essen.)*

or angiotensin-converting enzyme inhibitors, have also been used effectively.

Surgery should be performed by teams of surgeons and anesthesiologists with experience in the management of pheochromocytomas. Blood pressure can be labile during surgery, particularly at the outset of intubation or when the tumor is manipulated. Nitroprusside infusion is useful for intraoperative hypertensive crises, and hypotension usually responds to volume infusion. The latter side effect can, however, be avoided in normotensive pheochromocytoma patients by having only standby intraoperative nitroprusside, which has been shown to be safe and avoids postoperative hypotension often caused by alpha blockers. The long-lasting guideline for obligatory preoperative treatment with alpha blockers is under discussion and seemingly not needed.

Minimally invasive techniques (laparoscopy or retroperitoneoscopy) have become the standard approaches in pheochromocytoma surgery. They are associated with fewer complications, a faster recovery, and optimal cosmetic results. Extra-adrenal abdominal and most thoracic pheochromocytomas can also be removed endoscopically. In this setting, adrenal sparing surgery should be considered. Postoperatively, catecholamine normalization should be documented. An adrenocorticotropic hormone (ACTH) test should be used to exclude cortisol deficiency when bilateral adrenal cortex–sparing surgery has been performed.

Head and neck paragangliomas are a challenge for surgeons, since damage of adjacent tissue, mainly vessels or cranial nerves II, VII, IX, X, XI, and XII, is a frequent permanent side effect. Careful consideration of best management is important, and radiotherapy may be an alternative, especially for large head and neck paragangliomas. Tympanic paragangliomas are symptomatic early, and most of these tumors can easily be resected, with subsequent improvement of hearing and alleviation of tinnitus. Asymptomatic paraganglial tumors, now often detected in patients with hereditary tumors and their relatives, are challenging to manage. Watchful waiting strategies have been introduced, but they should consider any genetic syndrome, such as *SDHB*, that might be associated with a higher degree of malignancy (see below).

METASTATIC PHEOCHROMOCYTOMA

About 5–10% of PPGLs are metastatic. The diagnosis of malignant pheochromocytoma is problematic. The typical histologic criteria of cellular atypia, presence of mitoses, and invasion of vessels or adjacent tissues are insufficient for the diagnosis of malignancy in pheochromocytoma. Size >5 cm, high pheochromocytoma of the adrenal gland scaled score (PASS) and grading system for adrenal pheochromocytoma and paraganglioma (GAPP) score, and SDHB-positive status have been considered as markers of risk of recurrence, but they remain controversial. Thus, the term *malignant pheochromocytoma* has been replaced by *metastatic pheochromocytoma* as suggested by the WHO and is restricted to tumors with lymph node or distant metastases, the latter most commonly found by nuclear medicine imaging in lungs, bone, or liver locations, suggesting a vascular pathway of spread (Fig. 399-4*B*). Because hereditary syndromes are associated with multifocal tumor sites, these features should be anticipated in patients with germline mutations, especially of *SDHB, SDHD, VHL,* and *RET.* However, distant metastases also occur in these syndromes, especially in carriers of *SDHB* mutations.

Treatment of metastatic pheochromocytoma or paraganglioma is challenging. Options include tumor mass reduction, alpha blockers for symptoms, chemotherapy including tyrosine kinase inhibitors, nuclear medicine radiotherapy, and stereotactic radiation. Nuclear medicine therapy is the treatment of choice for scintigraphically documented metastases, preferably with ^{131}I-MIBG in 100–300 mCi doses over 3–6 cycles, or somatostatin receptor ligands, e.g., DOTATOC labeled with yttrium-90 or lutetium-177. Averbuch's chemotherapy protocol includes dacarbazine (600 mg/m^2 on days 1 and 2), cyclophosphamide (750 mg/m^2 on day 1), and vincristine (1.4 mg/m^2 on day 1), all repeated every 21 days for 3–6 cycles. Palliation (stable disease to shrinkage) is achieved in about one-half of patients. Due to increasing insights in the genetics of pheochromocytoma and their molecular pathways, new targeted chemotherapeutic options such as sunitinib and temozolomide are under investigation. The prognosis of metastatic pheochromocytoma or paraganglioma is variable, with 5-year survival rates of 30–60%.

PHEOCHROMOCYTOMA IN PREGNANCY

Pheochromocytomas occasionally are diagnosed in pregnancy and can be very challenging to manage. The pathogenesis of adrenergic crises during pregnancy in previously asymptomatic women might be linked to human chorionic gonadotropin (hCG)-induced stimulation of epinephrine production by pheochromocytoma, as some of these express luteinizing hormone/chorionic gonadotropin (LHCG) receptors. Endoscopic removal, preferably in the fourth to sixth month of gestation, is possible and can be followed by uneventful childbirth. Regular screening in families with inherited pheochromocytomas provides an opportunity to identify and remove such tumors in women of reproductive age.

PHEOCHROMOCYTOMA-ASSOCIATED SYNDROMES

About 25–33% of patients with a PPGL have an inherited syndrome. At diagnosis, patients with inherited syndromes are a mean of ~15 years younger than patients with sporadic tumors.

The best-known pheochromocytoma-associated syndrome is the autosomal dominant disorder MEN 2 **(Chap. 400)**. Both types of MEN 2 (2A and 2B) are caused by mutations in *RET*, which encodes a tyrosine kinase. The locations of *RET* mutations correlate with the age of disease onset, the aggressiveness, and the type of MEN 2 **(Chap. 400)**. MEN 2A is characterized by medullary thyroid carcinoma (MTC), pheochromocytoma, and hyperparathyroidism. MEN 2B also includes MTC (more aggressive than in MEN 2A), pheochromocytoma, and multiple mucosal neuromas, marfanoid habitus, and other developmental disorders, although it typically lacks hyperparathyroidism. MTC is found in virtually all patients with MEN 2, but pheochromocytoma occurs in only ~50% of these patients. Nearly all pheochromocytomas in MEN 2 are benign and located in the adrenals, often bilaterally. Pheochromocytoma may be symptomatic before the diagnosis of MTC is made. Prophylactic thyroidectomy is being performed in many carriers of *RET* mutations, and the recommended age to perform thyroidectomy usually depends on the mutation and/or the level of calcitonin; pheochromocytomas should be excluded before any surgery in these patients.

VHL is an autosomal dominant disorder that predisposes to retinal and cerebellar hemangioblastomas, which also occur in the brainstem and spinal cord **(Fig. 399-6)**. Other important features of VHL are clear cell renal carcinomas, pancreatic neuroendocrine tumors, endolymphatic sac tumors of the inner ear, cystadenomas of the epididymis and broad ligament, and multiple pancreatic or renal cysts. Although the *VHL* gene can be inactivated by all types of mutations, patients with pheochromocytoma predominantly have missense mutations. About 20–30% of patients with VHL have pheochromocytomas, but in some families, the incidence can reach 90%. The recognition of pheochromocytoma as a VHL-associated feature provides an opportunity to diagnose retinal, central nervous system, renal, and pancreatic tumors at a stage when effective treatment may still be possible.

NF1 was the first described pheochromocytoma-associated syndrome. The *NF1* gene functions as a tumor suppressor by regulating the Ras signaling cascade. Classic features of neurofibromatosis include multiple neurofibromas, café au lait spots, axillary freckling of the skin, and Lisch nodules of the iris. Pheochromocytomas occur in only ~1% of these patients and are located predominantly in the adrenals. Metastatic pheochromocytoma is not uncommon in NF1.

The *paraganglioma syndromes* (PGLs) have been classified by genetic analyses of families with head and neck paragangliomas. The susceptibility genes encode subunits of the enzyme SDH, a component in the Krebs cycle and the mitochondrial electron transport chain. SDH is formed by four subunits (A–D). Mutations of *SDHA* (PGL5), *SDHB* (PGL4), *SDHC* (PGL3), *SDHD* (PGL1), and *SDHAF2* (PGL2) predispose to the PGLs. The transmission of the disease in carriers of

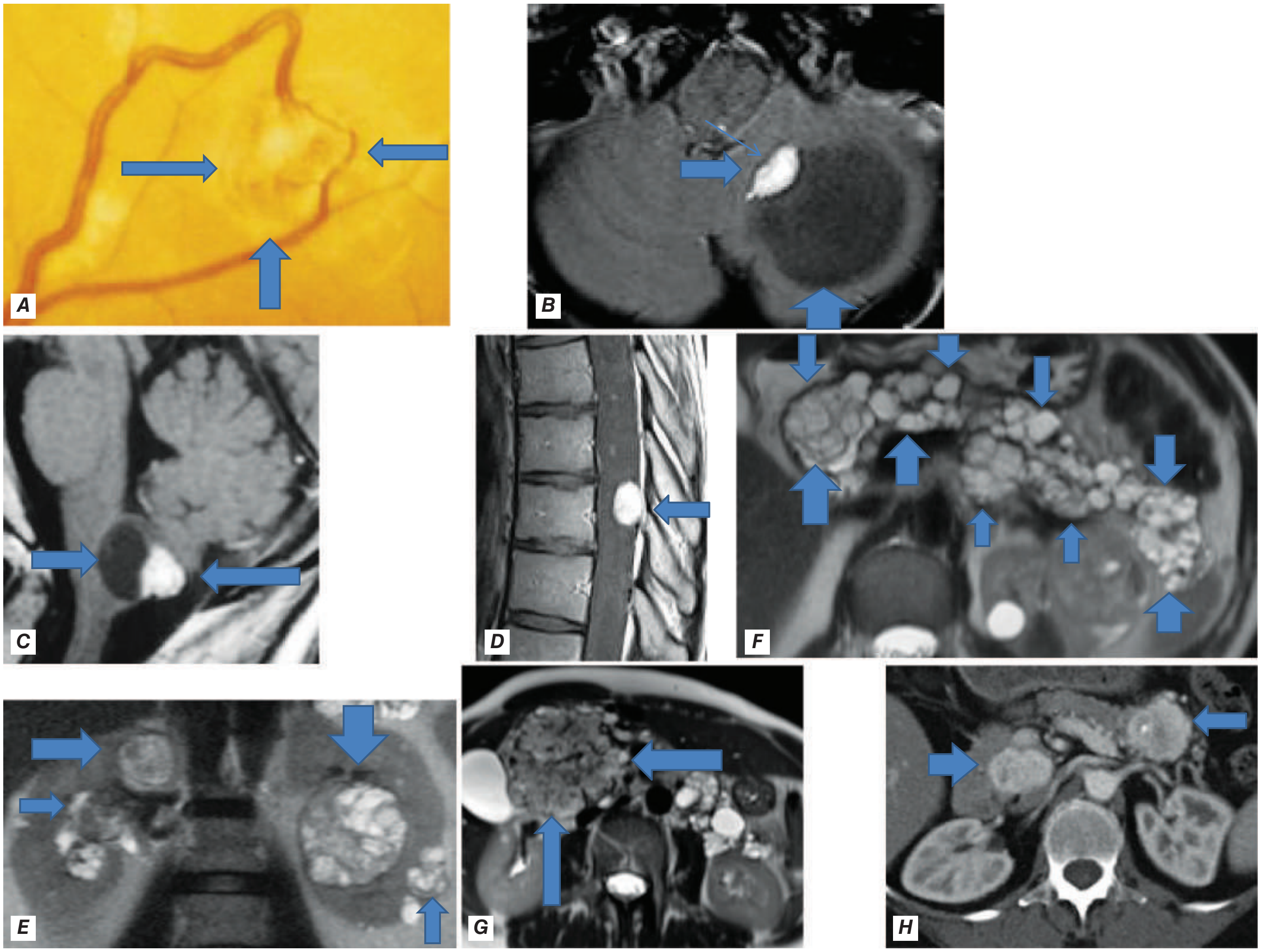

FIGURE 399-6 von Hippel–Lindau disease. Tumors and cysts marked by *arrows*. ***A.*** Retinal angioma (*arrows* with a pair of feeding vessels). All subsequent panels show findings on magnetic resonance imaging. ***B–D.*** Hemangioblastomas of the cerebellum (large cyst and a solid mural tumor) (***B***) in brainstem (in part cystic) (***C***) and spinal cord (thoracic) (***D***). ***E.*** Bilateral renal clear cell carcinomas with two tumors on each side ***F.*** Multiple pancreatic cysts. ***G.*** Microcystic serous pancreatic cystadenoma (with multiple tiny spaces). ***H.*** Two pancreatic islet cell tumors. *(Part A was provided courtesy of Dr. Dieter Schmidt. Part B was provided courtesy of Dr. Christian Taschner, Freiburg. Part C was provided courtesy of Dr. Sven Glaesker, Brussels. Part D was used with permission from Dr. Jan-Helge Klingler, Freiburg. Part E was provided courtesy of Dr. Cordula Jilg, Freiburg. Parts F–H were provided courtesy of Dr. Tobias Krauss, Freiburg.)*

SDHA, SDHB, and *SDHC* germline mutations is autosomal dominant. In contrast, in virtually all *SDHD* and *SDHAF2* families, only the progeny of affected fathers develops tumors if they inherit the mutation. PGL1 is most common, followed by PGL4; PGL2, PGL3, and PGL5 are rare. Adrenal, extra-adrenal abdominal, and thoracic pheochromocytomas, which are components of PGL1, PGL4, and PGL5, are rare in PGL3 and absent in PGL2 (Fig. 399-4*A*). About one-third of patients with PGL4 develop metastases, which is the highest rate in pheochromocytoma-associated syndromes. However, the penetrance of the disease is usually low, raising questions about the optimal way to monitor asymptomatic carriers on a long-term basis. Other syndromes with metastatic pheochromocytomas are mainly VHL, NF1, and PGL1.

Other familial pheochromocytoma has been attributed to hereditary, mainly adrenal tumors in patients with germline mutations in the genes *TMEM127* and *MAX*. Transmission is also autosomal dominant, and mutations of *MAX*, like those of *SDHD,* cause tumors only if inherited from the father. Pituitary neuroendocrine tumors have been described as an association with *SDHx* (pituitary adenoma and pheochromocytoma/paraganglioma 3PA syndrome) as well as *MAX* mutations (MEN 5), but they seem to occur very rarely.

■ GENETIC SCREENING OF PATIENTS WITH PHEOCHROMOCYTOMA OR PARAGANGLIOMA

Universal germline panel testing is now the gold standard to characterize the genetic factors involved in PPGL. It usually identifies a genetic etiology in up to 30% of the cases. This rate can be even higher in patients with early age of onset, extra-adrenal location, multiple tumors, metastatic tumors, or family history of PPGL. Effective preventive medicine for pheochromocytoma and pheochromocytoma-associated diseases requires management according to identified germline mutations in susceptibility genes **(Table 399-3)**. Because of the relatively high prevalence of familial syndromes among patients who present with pheochromocytoma or paraganglioma, it is useful to identify germline mutations even in patients without a known family history. Despite the use of a universal germline panel testing performed for all patients with a PPGL, a first step remains to search for clinical features of inherited syndromes and to obtain an in-depth, multigenerational family history. Each of these syndromes exhibits autosomal dominant transmission with variable penetrance, but a proband with a mother affected by paraganglial tumors is not predisposed to PLG1 and PGL2 (*SDHD* and *SDHAF2* mutation carrier). Cutaneous neurofibromas, café au lait spots, and axillary freckling suggest neurofibromatosis. Germline mutations in *NF1* have nearly never been reported in patients with sporadic pheochromocytomas. Thus, *NF1* testing is not needed in the absence of other clinical features of neurofibromatosis. A personal or family history of MTC or an elevation of serum calcitonin strongly suggests MEN 2 and should prompt testing for *RET* mutations. A history of visual impairment or tumors of the cerebellum, brainstem, spinal cord, or the kidney suggests the possibility of VHL. A personal and/or family history of head and neck paraganglioma suggests PGL1 or PGL4. Of note, sequencing protocols may not detect large deletions of one or more exons.

TABLE 399-3 Patterns of Occurrence in Inherited Pheochromocytoma and Paraganglioma–Associated Syndromes

MUTATED GENE	ADRENAL TUMORS	HEAD AND NECK TUMORS	EXTRA-ADRENAL RETROPERITONEAL OR PELVIC TUMORS	THORACIC TUMORS	MULTIPLE TUMORS	BILATERAL ADRENAL TUMORS	METASTATIC TUMORS	FAMILY HISTORY IN PROBANDS FOR COMPONENTS OF THE GIVEN SYNDROME
MAX	+++++	<x	+	<x	+++++	++++	++	+++
NF1	+++++	<+	+	<+	+	++	+	++
RET	+++++	<+	<+	<+	++++	++++	<+	+
SDHA	++	++++	++	+	+	<+	+	+
SDHB	++++	+++	+++	+	++	<+	+++	++
SDHC	<+	+++++	<+	+	+	<+	<+	++
SDHD	++	+++++	+	+	++++	<+	+	+++
VHL	+++++	<+	+	+	++++	+++	+	++++
TMEM127	+++++	+	+	+	++	++	<+	+

Note: Frequencies in percentage (<+: 0–4%; +: 5–19%; ++: 20–39%; +++: 40–59%; ++++: 60–79%; +++++: 80–100%) of clinical characteristics of pheochromocytomas/paragangliomas of patients with germline mutations of the genes *MAX, NF1, RET, SDHA, SDHB, SDHC, SDHD, VHL*, and *TMEM127*; for other genes, the data are too limited to include in this summary.

A single adrenal pheochromocytoma in a patient with an otherwise unremarkable history may still be associated with mutations of *VHL, RET, SDHB*, or *SDHD* (in decreasing order of frequency). Two-thirds of extra-adrenal tumors are associated with one of these syndromes, and multifocal tumors occur with decreasing frequency in carriers of *RET, SDHD, VHL, SDHB*, and *MAX* mutations. About 30% of head and neck paragangliomas are associated with germline mutations of one of the SDH subunit genes (most often *SDHD*) and are rare in carriers of *VHL, RET, MAX*, and *TMEM127* mutations. Immunohistochemistry is helpful in the preselection of hereditary pheochromocytoma. Negative immunostaining with antibodies to SDHB (Fig. 399-5*F*), TMEM127, and MAX may predict mutations of the *SDHx* (PGL1-5), *TMEM127*, and *MAX* genes, respectively.

Once the underlying syndrome is diagnosed, the benefit of genetic testing can be extended to relatives. For this purpose, it is necessary to identify the germline mutation in the proband and, after genetic counseling, to perform DNA sequence analyses of the responsible gene in relatives to determine whether they are affected. Other family members may benefit when individuals who carry a germline mutation are biochemically screened for paraganglial tumors.

FUTURE PERSPECTIVES

About 15% of sporadic PPGLs present with recurrence, including some ocurring >10 years after initial management; this rate is higher in genetically determined PPGL. Defining the clinical outcome of a PPGL remains difficult, and machine learning models including several factors might be helpful in the future to personalize the follow-up of these patients. The recent American Joint Committee on Cancer tumor-node-metastasis classification could help select patients at risk of recurrence. The pathogenesis of PPGL remains imperfectly understood, and as for other endocrine tumors, microenvironment and immune cells might influence PPGL aggressiveness or metastatic behavior; they could also constitute some new therapeutic targets in the future.

FURTHER READING

Al Subhi AR et al: Systematic review: Incidence of pheochromocytoma and paraganglioma over 70 years. J Endocr Soc 6:bvac105, 2022.

Amar L et al: International consensus on initial screening and follow-up of asymptomatic SDHx mutation carriers. Nat Rev Endocrinol 17:435, 2021.

Bancos I et al: Maternal and fetal outcomes in pheochromocytoma and pregnancy: A multi-center retrospective cohort study and systematic review of literature. Lancet Diabetes Endocrinol 2021; 9:13.

Calsina B et al: Genomic and immune landscape of metastatic pheochromocytoma and paraganglioma. Nat Commun 14:1122, 2023.

Castinetti F et al: Controversies about the systematic preoperative pharmacological treatment before pheochromocytoma or paraganglioma surgery. Eur J Endocrinol 186:D17, 2022.

Horton C et al: Universal germline panel testing for individuals with pheochromocytoma and paraganglioma produces high diagnostic yield. J Clin Endocrinol Metab 107:e1917, 2022.

Lenders JW et al: Genetics, diagnosis, management and future directions of research of phaeochromocytoma and paraganglioma: A position statement and consensus of the Working Group on Endocrine Hypertension of the European Society of Hypertension. J Hypertens 38:1443, 2020.

Lopez AG et al: Expression of LHCGR in pheochromocytomas unveils an endocrine mechanism connecting pregnancy and epinephrine overproduction. Hypertension 79:1006, 2022.

Neumann HPH et al: Comparison of pheochromocytoma-specific morbidity and mortality among adults with bilateral pheochromocytomas undergoing total adrenalectomy vs cortical-sparing adrenalectomy. JAMA Netw Open 2:e198898, 2019.

Pamporaki C et al: Prediction of metastatic pheochromocytoma and paraganglioma: A machine learning modelling study using data from a cross-sectional cohort. Lancet Digit Health 5:e551, 2023.

Taïeb D et al: European Association of Nuclear Medicine Practice Guideline/Society of Nuclear Medicine and Molecular Imaging Procedure Standard 2019 for radionuclide imaging of phaeochromocytoma and paraganglioma. Eur J Nucl Med Mol Imaging 46:2112, 2019.

Taïeb D et al: Clinical consensus guideline on the management of phaeochromocytoma and paraganglioma in patients harbouring germline SDHD pathogenic variants. Lancet Diabetes Endocrinol 11:345, 2023.

400

Multiple Endocrine Neoplasia Syndromes

Rajesh V. Thakker

Multiple endocrine neoplasia (MEN) is characterized by a predilection for tumors involving two or more endocrine glands. Five major forms of MEN are recognized and referred to as MEN types 1–5 (MEN 1–5) **(Table 400-1)**. Each type of MEN is inherited as an autosomal dominant syndrome or may occur sporadically, that is, without a family history. However, this distinction between familial and sporadic forms is often difficult because family members with the disease may have died before symptoms developed. In addition to MEN 1–5, at least six other syndromes are associated with multiple endocrine and other organ neoplasias (MEONs) **(Table 400-2)**. These MEONs include the hyperparathyroidism-jaw tumor (HPT-JT) syndrome, Carney complex, von Hippel–Lindau disease **(Chap. 399)**, neurofibromatosis

TABLE 400-1 Multiple Endocrine Neoplasia (MEN) Syndromes

TYPE (CHROMOSOMAL LOCATION)	TUMORS (ESTIMATED PENETRANCE)	GENE AND MOST FREQUENTLY MUTATED CODONS
MEN 1 (11q13)	Parathyroid adenoma (90%) Enteropancreatic tumor (30–70%) • Gastrinoma (>50%) • Insulinoma (10–30%) • Nonfunctioning and PPoma (20–55%) • Glucagonoma (<3%) • VIPoma (<1%) Pituitary adenoma (15–50%) • Prolactinoma (60%) • Somatotrophinoma (25%) • Corticotrophinoma (<5%) • Nonfunctioning (<5%) Associated tumors • Adrenal cortical tumor (20–70%) • Pheochromocytoma (<1%) • Bronchopulmonary NET (2%) • Thymic NET (2%) • Gastric NET (10%) • Lipomas (>33%) • Angiofibromas (85%) • Collagenomas (70%) • Meningiomas (8%)	*MEN1* 83/84, 4-bp del (≈4%) 119, 3-bp del (≈3%) 209-211, 4-bp del (≈8%) 418, 3-bp del (≈4%) 514-516, del or ins (≈7%) Intron 4 ss (≈10%)
MEN 2 (10 cen-10q11.2)		
MEN 2A	MTC (90%) Pheochromocytoma (>50%) Parathyroid adenoma (10–25%)	*RET* 634, e.g., Cys → Arg (~85%)
MTC only	MTC (100%)	*RET* 618, missense (>50%)
MEN 2B (also known as MEN 3)	MTC (>90%) Pheochromocytoma (>50%) Associated abnormalities (40–50%) • Mucosal neuromas • Marfanoid habitus • Medullated corneal nerve fibers • Megacolon	*RET* 918, Met → Thr (>95%)
MEN 4 (12p13)	Parathyroid adenoma[a] Pituitary adenoma[a] Reproductive organ tumors[a] (e.g., testicular cancer, neuroendocrine cervical carcinoma) ?Adrenal + renal tumors[a]	*CDKN1B*; no common mutations identified to date
MEN5 (14q23.3)	Pheochromocytoma[a] Pituitary adenoma[a] Parathyroid adenomas?[a] Neural crest tumors (e.g., ganglioneuroma, neuroblastoma) (other tumors? – renal cell carcinoma, renal oncocytoma, pancreatic NETs, chondrosarcoma)	*MAX*; no common mutations identified to date

[a] Insufficient numbers reported to provide prevalence information.

Note: Autosomal dominant inheritance of the MEN syndromes has been established.

Abbreviations: del, deletion; ins, insertion; MTC, medullary thyroid cancer; NET, neuroendocrine tumor; PPoma, pancreatic polypeptide–secreting tumor; VIPoma, vasoactive intestinal polypeptide–secreting tumor.

Source: Adapted with permission from Thakker RV. Multiple endocrine neoplasia–syndromes of the twentieth century. J Clin Endocrinol Metab 83:2617, 1998.

TABLE 400-2 Multiple Endocrine and Other Organ Neoplasia (MEON) Syndromes

DISEASE[a]	GENE PRODUCT	CHROMOSOMAL LOCATION
Hyperparathyroidism-jaw tumor (HPT-JT)	Parafibromin	1q31.2
Carney complex		
CNC1	PRAKAR1A	17q24.2
CNC2	?[b]	2p16
von Hippel–Lindau disease (VHL)	pVHL (elongin)	3p25
Neurofibromatosis type 1 (NF1)	Neurofibromin	17q11.2
Cowden's syndrome (CWS)		
CWS1	PTEN	10q23.31
CWS2	SDHB	1p36.13
CWS3	SDHD	11q23.1
CWS4	KLLN	10q23.31
CWS5	PIK3CA	3q26.32
CWS6	AKT1	14q32.33
CWS7	SEC23B	20p11.23
McCune-Albright syndrome (MAS)	$G_s\alpha$	20q13.32

[a] The inheritance for these disorders is autosomal dominant, except MAS, which is due to mosaicism that results from the postzygotic somatic cell mutation of the *GNAS1* gene, encoding $G_s\alpha$. [b] ?, unknown.

type 1 **(Chap. 95)**, Cowden's syndrome (CWS), and McCune-Albright syndrome (MAS) **(Chap. 424)**; all of these are inherited as autosomal dominant disorders, except for MAS, which is caused by mosaic expression of a postzygotic somatic cell mutation (Table 400-2).

A diagnosis of a MEN or MEON syndrome may be established in an individual by one of three criteria: (1) clinical features (two or more of the associated tumors [or lesions] in an individual); (2) familial pattern (one of the associated tumors [or lesions] in a first-degree relative of a patient with a clinical diagnosis of the syndrome); and (3) genetic analysis (a germline mutation in the associated gene in an individual, who may be clinically affected or asymptomatic). Mutational analysis in MEN and MEON syndromes is helpful in clinical practice to (1) confirm the clinical diagnosis; (2) identify family members who harbor the mutation and require screening for relevant tumor detection and early/appropriate treatment; and (3) identify the ~50% of family members who do not harbor the germline mutation and can, therefore, be alleviated of the anxiety of developing associated tumors. This latter aspect also helps to reduce health care costs by reducing the need for unnecessary biochemical and radiologic investigations.

MULTIPLE ENDOCRINE NEOPLASIA TYPE 1

Clinical Manifestations MEN type 1 (MEN 1), which is also referred to as Wermer's syndrome, is characterized by the triad of tumors involving the parathyroids, pancreatic islets, and anterior pituitary. In addition, adrenal cortical tumors, carcinoid tumors usually of the foregut, meningiomas, facial angiofibromas, collagenomas, and lipomas may also occur in some patients with MEN 1. Combinations of the affected glands and their pathologic features (e.g., hyperplastic adenomas of the parathyroid glands) may differ in members of the same family and even between identical twins. In addition, a nonfamilial (e.g., sporadic) form occurs in 8–14% of patients with MEN 1, and molecular genetic studies have confirmed the occurrence of de novo mutations of the *MEN1* gene in ~10% of patients with MEN 1. The prevalence of MEN 1 is ~0.25% based on randomly chosen postmortem studies but is 1–18% among patients with primary hyperparathyroidism, 16–38% among patients with pancreatic islet tumors, and <3% among patients with pituitary tumors. The disorder affects all age groups, with a reported age range of 5–81 years, with clinical and biochemical manifestations developing in the vast majority by the fifth decade. The clinical manifestations of MEN 1 are related to the sites of tumors and their hormonal products. In the absence of treatment, endocrine tumors are associated with an earlier

TABLE 400-3 Biochemical and Radiologic Screening in Multiple Endocrine Neoplasia Type 1

TUMOR	AGE TO BEGIN (YEARS)	BIOCHEMICAL TEST (PLASMA OR SERUM) ANNUALLY	IMAGING TEST (TIME INTERVAL)
Parathyroid	8	Calcium, PTH	None
Pancreatic NETs			
Gastrinoma	20	Gastrin (± gastric pH)	None
Insulinoma	5	Fasting glucose, insulin	None
Other pancreatic NET	<10	Chromogranin A; pancreatic polypeptide, glucagon, vasoactive intestinal peptide	MRI, CT, or EUS (annually)
Anterior pituitary	5	Prolactin, IGF-I	MRI (every 3 years)
Adrenal	<10	None unless symptoms or signs of functioning tumor and/or tumor >1 cm identified on imaging	MRI or CT (annually with pancreatic imaging)
Thymic and bronchial carcinoid	15	None	CT or MRI (every 1–2 years)

Abbreviations: CT, computed tomography; EUS, endoscopic ultrasound; IGF-I, insulin-like growth factor I; MRI, magnetic resonance imaging; NET, neuroendocrine tumor; PTH, parathyroid hormone.

Source: Data from PJ Newey, RV Thakker: Role of multiple endocrine neoplasia type 1 mutational analysis in clinical practice. Endocr Pract 17, 2011 and RV Thakker: Multiple endocrine neoplasia type 1 (MEN1). Translational Endocrinology and Metabolism, Vol 2. Chevy Chase, MD: The Endocrine Society; 2011.

mortality in patients with MEN 1, with a 50% probability of death by the age of 50 years. The cause of death is usually a malignant tumor, often from a pancreatic neuroendocrine tumor (NET) or foregut carcinoid. In addition, the treatment outcomes of patients with MEN 1–associated tumors are not as successful as those in patients with non–MEN 1 tumors. This is because MEN 1–associated tumors, with the exception of pituitary NETs, are usually multiple, making it difficult to achieve a successful surgical cure. Occult metastatic disease is also more prevalent in MEN 1, and the tumors may be larger, more aggressive, and resistant to treatment.

Parathyroid Tumors (See also Chap. 422) Primary hyperparathyroidism occurs in ~90% of patients and is the most common feature of MEN 1. Patients may have asymptomatic hypercalcemia or vague symptoms associated with hypercalcemia (e.g., polyuria, polydipsia, constipation, malaise, or dyspepsia). Nephrolithiasis and osteitis fibrosa cystica (less commonly) may also occur. Biochemical investigations reveal hypercalcemia, usually in association with elevated circulating parathyroid hormone (PTH) (Table 400-3). The hypercalcemia is usually mild, and severe hypercalcemia or parathyroid cancer is a rare occurrence. Additional differences in the primary hyperparathyroidism of patients with MEN 1, as opposed to those without MEN 1, include an earlier age at onset (20–25 vs 55 years) and an equal male-to-female ratio (1:1 vs 1:3). Preoperative imaging (e.g., neck ultrasound with ^{99m}Tc-sestamibi parathyroid scintigraphy) is of limited benefit because all parathyroid glands may be affected, and neck exploration may be required irrespective of preoperative localization studies.

TREATMENT

Parathyroid Tumors

Surgical removal of the abnormally overactive parathyroids in patients with MEN 1 is the definitive treatment, and most specialist centers recommend performing a subtotal (e.g., removal of 3.5 glands) parathyroidectomy. Minimally invasive parathyroidectomy is not recommended because all four parathyroid glands are usually affected with multiple adenomas or hyperplasia. Surgical experience should be taken into account given the variability in pathology in MEN 1. Calcimimetics (e.g., cinacalcet), which act via the calcium-sensing receptor, have been used to treat primary hyperparathyroidism in some patients when surgery is unsuccessful or contraindicated.

Pancreatic Tumors (See also Chap. 89) The incidence of pancreatic islet cell tumors, which are NETs, in patients with MEN 1 ranges from 30 to 80% in different series. Most of these tumors (Table 400-1) produce excessive amounts of hormone (e.g., gastrin, insulin, glucagon, vasoactive intestinal polypeptide [VIP]) and are associated with distinct clinical syndromes, although some are nonfunctioning or nonsecretory. These pancreatic islet cell tumors have an earlier age at onset in patients with MEN 1 than in patients without MEN 1.

Gastrinoma Gastrin-secreting tumors (gastrinomas) are associated with marked gastric acid production and recurrent peptic ulcerations, a combination referred to as Zollinger-Ellison syndrome. Gastrinomas occur more often in patients with MEN 1 who are aged >30 years. Recurrent severe multiple peptic ulcers, which may perforate, and cachexia are major contributors to the high mortality. Patients with Zollinger-Ellison syndrome may also suffer from diarrhea and steatorrhea. The diagnosis is established by demonstration of an elevated fasting serum gastrin concentration in association with increased basal gastric acid secretion (Table 400-3). However, the diagnosis of Zollinger-Ellison syndrome may be difficult in hypercalcemic MEN 1 patients because hypercalcemia can also cause hypergastrinemia. Ultrasonography, endoscopic ultrasonography, computed tomography (CT), nuclear magnetic resonance imaging (MRI), selective abdominal angiography, venous sampling, and somatostatin receptor scintigraphy (SRS) are helpful in localizing the tumor prior to surgery. Gastrinomas represent >50% of all pancreatic NETs in patients with MEN 1, and ~20% of patients with gastrinomas will be found to have MEN 1. Gastrinomas, which may also occur in the duodenal mucosa, are the major cause of morbidity and mortality in patients with MEN 1.

TREATMENT

Gastrinoma

Medical treatment of patients with MEN 1 and Zollinger-Ellison syndrome is directed toward reducing basal acid output to <10 mmol/L. Parietal cell H^+-K^+-adenosine triphosphatase (ATPase) inhibitors (e.g., omeprazole or lansoprazole) reduce acid output and are the drugs of choice for gastrinomas. Some patients may also require additional treatment with the histamine H_2 receptor antagonists cimetidine or ranitidine. The role of surgery in the treatment of gastrinomas in patients with MEN 1 is controversial. The goal of surgery is to reduce the risk of distant metastatic disease and improve survival. For a nonmetastatic gastrinoma situated in the pancreas, surgical excision is often effective. However, the risk of hepatic metastases increases with tumor size, such that 25–40% of patients with pancreatic NETs >4 cm develop hepatic metastases, and 50–70% of patients with tumors 2–3 cm in size have lymph node metastases. Survival in MEN 1 patients with gastrinomas <2.5 cm in size is 100% at 15 years, but 52% at 15 years, if metastatic disease is present. The presence of lymph node metastases does not appear to adversely affect survival. Surgery for gastrinomas that are >2–2.5 cm has been recommended, because the disease-related survival in these patients is improved following surgery. In addition, duodenal gastrinomas, which occur more frequently in patients with MEN 1, have been treated successfully with surgery. However, in most patients with MEN 1, gastrinomas are multiple or extrapancreatic, and with the exception of duodenal gastrinomas, surgery is rarely successful. For example, the results of one study revealed that only ~15% of patients with MEN 1 were free of disease immediately after surgery, and at 5 years, this number had decreased to ~5%; the respective outcomes in patients without MEN 1 were better, at 45 and 40%. Given these findings, most specialists recommend a nonsurgical management for gastrinomas in MEN 1, except as noted earlier for smaller, isolated lesions. Treatment of disseminated gastrinomas is difficult. Chemotherapy with streptozotocin and 5-fluorouracil; hormonal therapy with octreotide or lanreotide, which are human somatostatin analogues (SSAs); selected internal radiation therapy (SIRT); radiofrequency ablation; peptide radio receptor therapy (PRRT); hepatic artery embolization; administration of human leukocyte interferon; and removal of all resectable tumor have been successful in some patients.

Insulinoma These β islet cell insulin-secreting tumors represent 10–30% of all pancreatic tumors in patients with MEN 1. Patients with an insulinoma present with hypoglycemic symptoms (e.g., weakness, headaches, sweating, faintness, seizures, altered behavior, weight gain) that typically develop after fasting or exertion and improve after glucose intake. The most reliable test is a supervised 72-h fast. Biochemical investigations reveal increased plasma insulin concentrations in association with hypoglycemia (Table 400-3). Circulating concentrations of C peptide and proinsulin, which are also increased, are useful in establishing the diagnosis. It also is important to demonstrate the absence of sulfonylureas in plasma and urine samples obtained during the investigation of hypoglycemia (Table 400-3). Surgical success is greatly enhanced by preoperative localization by endoscopic ultrasonography, CT scanning, or celiac axis angiography. Additional localization methods may include preoperative and perioperative percutaneous transhepatic portal venous sampling, selective intraarterial stimulation with hepatic venous sampling, and intraoperative direct pancreatic ultrasonography. Insulinomas occur in association with gastrinomas in 10% of patients with MEN 1, and the two tumors may arise at different times. Insulinomas occur more often in patients with MEN 1 who are aged <40 years, and some arise in individuals aged <20 years. In contrast, in patients without MEN 1, insulinomas generally occur in those aged >40 years. Insulinomas may be the first manifestation of MEN 1 in 10% of patients, and ~4% of patients with insulinomas will have MEN 1.

TREATMENT

Insulinoma

Medical treatment, which consists of frequent carbohydrate meals and diazoxide or octreotide, is not always successful, and surgery is the optimal treatment. Surgical treatment, which ranges from enucleation of a single tumor to a distal pancreatectomy or partial pancreatectomy, has been curative in many patients. Chemotherapy (streptozotocin, 5-fluorouracil, and doxorubicin), PRRT (e.g., with ^{177}Lu-DOTATATE), or hepatic artery embolization has been used for metastatic disease.

Glucagonoma These glucagon-secreting pancreatic NETs occur in <3% of patients with MEN 1. The characteristic clinical manifestations of a skin rash (necrolytic migratory erythema), weight loss, anemia, and stomatitis may be absent. The tumor may have been detected in an asymptomatic patient with MEN 1 undergoing pancreatic imaging or by the finding of glucose intolerance and hyperglucagonemia.

TREATMENT

Glucagonoma

Surgical removal of the glucagonoma is the treatment of choice. However, treatment may be difficult because ~50–80% of patients have metastases at the time of diagnosis. Medical treatment with SSAs (e.g., octreotide or lanreotide) or chemotherapy with streptozotocin and 5-fluorouracil has been successful in some patients, and hepatic artery embolization has been used to treat metastatic disease.

Vasoactive Intestinal Peptide (VIP) Tumors (VIPomas) VIPomas have been reported in only a few patients with MEN 1. This clinical syndrome is characterized by watery diarrhea, hypokalemia, and achlorhydria (WDHA syndrome), which is also referred to as the Verner-Morrison syndrome, or the VIPoma syndrome. The diagnosis is established by excluding laxative and diuretic abuse, confirming a stool volume in excess of 0.5–1.0 L/d during a fast, and documenting a markedly increased plasma VIP concentration.

TREATMENT

VIPomas

Surgical management of VIPomas, which are mostly located in the tail of the pancreas, can be curative. However, in patients with unresectable tumor, SSAs, such as octreotide and lanreotide, may be effective. Streptozotocin with 5-fluorouracil may be beneficial, along with hepatic artery embolization for the treatment of metastases.

Pancreatic Polypeptide-Secreting Tumors (PPomas) and Nonfunctioning Pancreatic NETs PPomas are found in a large number of patients with MEN 1. No pathologic sequelae of excessive polypeptide (PP) secretion are apparent, and the clinical significance of PP is unknown. Many PPomas may have been unrecognized or classified as nonfunctioning pancreatic NETs, which likely represent the most common enteropancreatic NET associated with MEN 1 **(Fig. 400-1)**. The absence of both a clinical syndrome and specific biochemical abnormalities may result in a delayed diagnosis of nonfunctioning pancreatic NETs, which are associated with a worse prognosis than other functioning tumors, including insulinoma and gastrinoma. The optimum screening method and its timing interval for nonfunctioning pancreatic NETs remain to be established. At present, endoscopic ultrasound likely represents the most sensitive method of detecting small pancreatic tumors, but SRS is the most reliable method for detecting metastatic disease (Table 400-3).

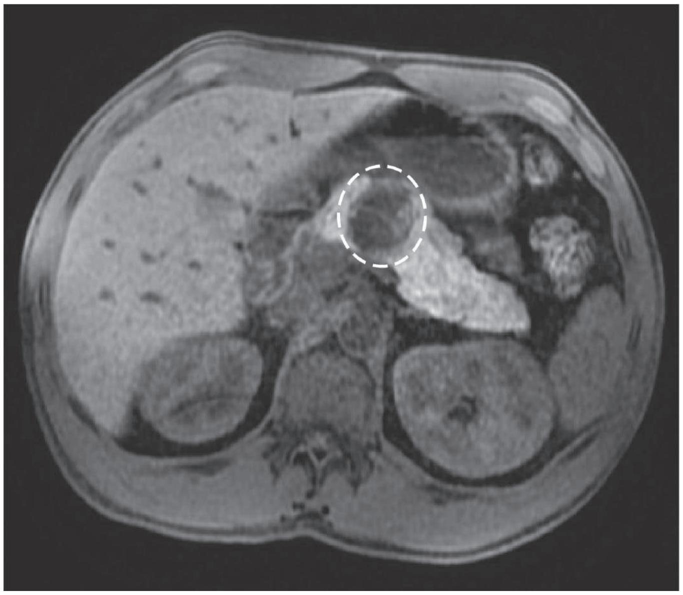

FIGURE 400-1 Pancreatic nonfunctioning neuroendocrine tumor (NET) in a 32-year-old patient with multiple endocrine neoplasia type 1 (MEN 1). An abdominal magnetic resonance imaging (MRI) scan revealed a low-intensity >3.0 cm (anteroposterior maximal diameter) tumor within the body of pancreas. There was no evidence of invasion of adjacent structures or metastases. The tumor is indicated by *white dashed circle*.

TREATMENT

PPomas and Nonfunctioning Pancreatic NETs

The management of nonfunctioning pancreatic NETs in the asymptomatic patient is controversial. One recommendation is to undertake surgery irrespective of tumor size after biochemical assessment is complete. Alternatively, other experts recommend surgery based on tumor size, using either >1 cm or >2 cm at different centers. Pancreatoduodenal surgery is successful in removing the tumors in 80% of patients, but >40% of patients develop complications, including diabetes mellitus, frequent steatorrhea, early and late dumping syndromes, and other gastrointestinal symptoms. However, ~50–60% of patients treated surgically survive >5 years. When considering these recommendations, it is important to consider that occult metastatic disease (e.g., tumors not detected by imaging investigations) is likely to be present in a substantial proportion of these patients at the time of presentation. Inhibitors of tyrosine kinase receptors (TKRs) and of the mammalian target of rapamycin (mTOR) signaling pathway have been reported to be effective in treating pancreatic NET metastases and in doubling the progression-free survival time. Additional treatments for metastatic disease include PRRT using ^{177}Lu-DOTATATE, chemotherapy, radiofrequency ablation, transarterial chemoembolization, and SIRT.

Other Pancreatic NETs NETs secreting growth hormone–releasing hormone (GHRH), GHRHomas, have been reported rarely in patients with MEN 1. It is estimated that ~33% of patients with GHRHomas have other MEN 1–related tumors. GHRHomas may be diagnosed by demonstrating elevated serum concentrations of growth hormone and GHRH. More than 50% of GHRHomas occur in the lung, 30% occur in the pancreas, and 10% are found in the small intestine. Somatostatinomas secrete somatostatin, a peptide that inhibits the secretion of a variety of hormones, resulting in hyperglycemia, cholelithiasis, low acid output, steatorrhea, diarrhea, abdominal pain, anemia, and weight loss. Although 7% of pancreatic NETs secrete somatostatin, the clinical features of somatostatinoma syndrome are unusual in patients with MEN 1.

Pituitary Tumors (See also Chap. 392) Pituitary tumors occur in 15–50% of patients with MEN 1 (Table 400-1), and ~75% of these are microadenomas (<1 cm diameter). The tumors occur as early as 5 years of age or as late as the ninth decade. MEN 1 pituitary adenomas are more frequent in women than men, in whom they are often macroadenomas (>1 cm diameter). There are no specific histologic parameters that differentiate between MEN 1 and non–MEN 1 pituitary tumors. Approximately 60% of MEN 1–associated pituitary tumors secrete prolactin, <25% secrete growth hormone, 5% secrete adrenocorticotropic hormone (ACTH), and the remainder appear to be nonfunctioning, with some secreting glycoprotein subunits (Table 400-1). However, pituitary tumors derived from MEN 1 patients may exhibit immunoreactivity to several hormones. In particular, there is a greater frequency of somatolactotrope tumors. Prolactinomas are the first manifestation of MEN 1 in ~15% of patients, whereas somatotrope tumors occur more often in patients aged >40 years. Fewer than 3% of patients with anterior pituitary tumors will have MEN 1. Clinical manifestations are similar to those in patients with sporadic pituitary tumors without MEN 1 and depend on the hormone secreted and the size of the pituitary tumor. Thus, patients may have symptoms of hyperprolactinemia (e.g., amenorrhea, infertility, and galactorrhea in women, or impotence and infertility in men) or have features of acromegaly or Cushing's disease. In addition, enlarging pituitary tumors may compress adjacent structures such as the optic chiasm or normal pituitary tissue, causing visual disturbances and/or hypopituitarism. In asymptomatic patients with MEN 1, periodic biochemical monitoring of serum prolactin and insulin-like growth factor 1 (IGF-1) levels, as well as MRI of the pituitary, can lead to early identification of pituitary tumors (Table 400-3). In patients with abnormal results, hypothalamic-pituitary testing should characterize the nature of the pituitary lesion and its effects on the secretion of other pituitary hormones.

TREATMENT

Pituitary Tumors

Treatment of pituitary tumors in patients with MEN 1 consists of therapies similar to those used in patients without MEN 1 and includes appropriate medical therapy (e.g., bromocriptine or cabergoline for prolactinoma; or octreotide or lanreotide for somatotrope tumors) or selective transsphenoidal adenomectomy, if feasible, with radiotherapy reserved for residual unresectable tumor tissue.

Associated Tumors Patients with MEN 1 may also develop carcinoid tumors, adrenal cortical tumors, facial angiofibromas, collagenomas, thyroid tumors, and lipomatous tumors.

Carcinoid Tumors (See also Chap. 89) Carcinoid tumors occur in >3% of patients with MEN 1 (Table 400-1). The carcinoid tumor may be located in the bronchi, gastrointestinal tract, pancreas, or thymus. At the time of diagnosis, most patients are asymptomatic and do not have clinical features of the carcinoid syndrome. Importantly, no hormonal or biochemical abnormality (e.g., plasma chromogranin A) is consistently observed in individuals with thymic or bronchial carcinoid tumors. Thus, screening for these tumors is dependent on radiologic imaging. The optimum method for screening has not been established. Low-dose CT and MRI are sensitive for detecting thymic and bronchial tumors (Table 400-3), although repeated CT scanning raises concern about exposure to repeated doses of ionizing radiation. Octreotide scintigraphy may also reveal some thymic and bronchial carcinoids, although there is insufficient evidence to recommend its routine use. Gastric carcinoids, of which the type II gastric enterochromaffin-like (ECL) cell carcinoids (ECLomas) are associated with MEN 1 and Zollinger-Ellison syndrome, may be detected incidentally at the time of gastric endoscopy for dyspeptic symptoms in MEN 1 patients. These tumors, which may be found in >10% of MEN 1 patients, are usually multiple and sized <1.5 cm. Bronchial carcinoids in patients with MEN 1 occur predominantly in women (male-to-female ratio, 1:4). In contrast, thymic carcinoids in European patients with MEN 1 occur predominantly in men (male-to-female ratio, 20:1), with cigarette smokers having a higher risk for these tumors; thymic carcinoids in Japanese patients with MEN 1 have a less marked sex difference (male-to-female ratio 2:1). The course of thymic carcinoids in MEN 1 appears to be particularly aggressive. The presence of thymic tumors in patients with MEN 1 is associated with a median survival after diagnosis of ~9.5 years, with 70% of patients dying as a direct result of the tumor.

TREATMENT

Carcinoid Tumors

If resectable, surgical removal of carcinoid tumors is the treatment of choice. For patients with unresectable tumors and those with metastatic disease, treatment with SSAs, radiotherapy, chemotherapeutic agents (e.g., fluorouracil, temozolomide, cisplatin, etoposide), mTOR inhibitors (e.g., everolimus), or PRRT therapy has resulted in symptom improvement and regression of some tumors. Little is known about the malignant potential of gastric type II ECLomas, but treatment with SSAs has resulted in regression of these ECLomas.

Adrenocortical Tumors (See also Chap. 398) Asymptomatic adrenocortical tumors occur in 20–70% of patients with MEN 1 depending on the radiologic screening methods used (Table 400-1). Most of these tumors, which include cortical adenomas, hyperplasia, multiple adenomas, nodular hyperplasia, cysts, and carcinomas, are nonfunctioning. Indeed, <10% of patients with enlarged adrenal glands have hormonal hypersecretion, with primary hyperaldosteronism and ACTH-independent Cushing's syndrome being encountered most commonly. Occasionally, hyperandrogenemia may occur in association with adrenocortical carcinoma. Pheochromocytoma in association with MEN 1 is rare. Biochemical investigation (e.g., plasma renin and aldosterone concentrations, low-dose dexamethasone suppression test, urinary catecholamines, and/or metanephrines) should be undertaken in those with symptoms or signs suggestive of functioning adrenal tumors or in those with tumors >1 cm. Adrenocortical carcinoma occurs in ~1% of MEN 1 patients but increases to >10% for adrenal tumors >1 cm.

TREATMENT

Adrenocortical Tumors

Consensus has not been reached about the management of MEN 1–associated nonfunctioning adrenal tumors, because the majority are benign. However, the risk of malignancy increases with size, particularly for tumors with a diameter >4 cm. Indications for surgery for adrenal tumors include size >4 cm in diameter, atypical or suspicious radiologic features (e.g., increased Hounsfield unit on unenhanced CT scan) and size of 1–4 cm in diameter, or significant measurable growth over a 6-month period. The treatment of functioning (e.g., hormone-secreting) adrenal tumors is similar to that for tumors occurring in non–MEN 1 patients.

Meningioma Central nervous system (CNS) tumors, including ependymomas, schwannomas, and meningiomas, have been reported in MEN 1 patients (Table 400-1). Meningiomas are found in <10% of patients with other clinical manifestations of MEN 1 (e.g., primary hyperparathyroidism) for >15 years. The majority of meningiomas are not associated with symptoms, and 60% do not enlarge. The treatment of MEN 1–associated meningiomas is similar to that in non–MEN 1 patients.

Lipomas Subcutaneous lipomas occur in >33% of patients with MEN 1 (Table 400-1) and are frequently multiple. In addition, visceral, pleural, or retroperitoneal lipomas may occur in patients with MEN 1. Management is conservative. However, when surgically removed for cosmetic reasons, they typically do not recur.

Facial Angiofibromas and Collagenomas The occurrence of multiple facial angiofibromas in patients with MEN 1 may range from >20 to >90%, and occurrence of collagenomas may range from 0 to >70% (Table 400-1). These cutaneous findings may allow presymptomatic diagnosis of MEN 1 in the relatives of a patient with MEN 1. Treatment for these cutaneous lesions is usually not required.

Thyroid Tumors Thyroid tumors, including adenomas, colloid goiters, and carcinomas, have been reported to occur in >25% of patients with MEN 1. However, the prevalence of thyroid disorders in the general population is high, and it has been suggested that the association of thyroid abnormalities in patients with MEN 1 may be incidental. The treatment of thyroid tumors in MEN 1 patients is similar to that for non–MEN 1 patients.

Genetics and Screening The *MEN1* gene is located on chromosome 11q13 and consists of 10 exons, which encode a 610–amino acid protein, menin, that regulates transcription, genome stability, cell division, and proliferation. The pathophysiology of MEN 1 follows the Knudson two-hit hypothesis with a tumor-suppressor role for menin. Inheritance of a germline *MEN1* mutation predisposes an individual to developing a tumor that arises following a somatic mutation, which may be a point mutation or more commonly a deletion, leading to loss of heterozygosity (LOH) in the tumor DNA. The germline mutations of the *MEN1* gene are scattered throughout the entire 1830-bp coding region and splice sites, and there is no apparent correlation between the location of *MEN1* mutations and clinical manifestations of the disorder, in contrast with the situation in patients with MEN 2 (Table 400-1). More than 10% of *MEN1* germline mutations arise de novo and may be transmitted to subsequent generations. Some families with MEN 1 mutations develop parathyroid tumors as the sole endocrinopathy, and this condition is referred to as familial isolated hyperparathyroidism (FIHP). However, between 5 and 25% of patients with MEN 1 do not harbor germline mutations or deletions of the *MEN1* gene. Such patients with MEN 1–associated tumors but without *MEN1* mutations may represent phenocopies or have mutations involving other genes. Other genes associated with MEN 1–like features include *CDKN1B*, which encodes p27kip1; mutations result in MEN4 (see below) (Table 400-1). Mutations in *CDC73*, which encodes parafibromin, result in the HPT-JT syndrome; mutations in the calcium-sensing receptor gene (*CaSR*) result in familial benign hypocalciuric hypercalcemia (FBHH); mutations in *MAX*, which encodes Myc-associated factor X, result in MEN5, and mutations in the aryl hydrocarbon receptor interacting protein gene (*AIP*), a tumor suppressor located on chromosome 11q13, are associated with familial isolated pituitary adenomas (FIPA). Genetic testing to determine the *MEN1* mutation status in symptomatic family members within a MEN 1 kindred, as well as in all index cases (e.g., patients) with two or more endocrine tumors, is advisable. If a *MEN1* mutation is not identified in the index case with two or more endocrine tumors, clinical and genetic tests for other disorders such as HPT-JT syndrome, FBHH, FIPA, MEN 2, or MEN 4 should be considered because these patients may represent phenocopies for MEN 1.

The current guidelines recommend that *MEN1* mutational analysis should be undertaken in (1) an index case with two or more MEN 1–associated endocrine tumors (e.g., parathyroid, pancreatic, or pituitary tumors); (2) asymptomatic first-degree relatives of a known *MEN1* mutation carrier; and (3) first-degree relatives of a *MEN1* mutation carrier with symptoms, signs, or biochemical or radiologic evidence for one or more MEN 1–associated tumors. In addition, *MEN1* mutational analysis should be considered in patients with suspicious or atypical MEN 1. This would include individuals with parathyroid adenomas before the age of 30 years or multigland parathyroid disease; individuals with gastrinoma or multiple pancreatic NETs at any age; or individuals who have two or more MEN 1–associated tumors that are not part of the classical triad of parathyroid, pancreatic islet, and anterior pituitary tumors (e.g., parathyroid tumor plus adrenal tumor). Family members, including asymptomatic individuals who have been identified to harbor a *MEN1* mutation, will require biochemical and radiologic screening (Table 400-3). In contrast, relatives who do not harbor the *MEN1* mutation have a risk of developing MEN 1–associated endocrine tumors that is similar to that of the general population; thus, relatives without the *MEN1* mutation do not require repeated screening.

Mutational analysis in asymptomatic individuals should be undertaken at the earliest opportunity and, if possible, in the first decade of life because tumors have developed in some children by the age of 5 years. Appropriate biochemical and radiologic investigations (Table 400-3) aimed at detecting the development of tumors should then be undertaken in affected individuals. Mutant gene carriers should undergo biochemical screening at least once per annum and also have baseline pituitary and abdominal imaging (e.g., MRI or CT), which should then be repeated at 1- to 3-year intervals (Table 400-3). Screening should commence after 5 years of age and should continue for life because the disease may develop as late as the eighth decade. The screening history and physical examination elicit the symptoms and signs of hypercalcemia; nephrolithiasis; peptic ulcer disease; neuroglycopenia; hypopituitarism; galactorrhea and amenorrhea in women; acromegaly; Cushing's disease; and visual field loss and the presence of subcutaneous lipomas, angiofibromas, and collagenomas. Biochemical screening should include measurements of serum calcium, PTH, gastrointestinal hormones (e.g., gastrin, insulin with a fasting glucose, glucagon, VIP, PP), chromogranin A, prolactin, and IGF-1 in all individuals. More specific endocrine function tests should be undertaken in individuals who have symptoms or signs suggestive of a specific clinical syndrome. Biochemical screening for the development of MEN 1 tumors in asymptomatic members of families with MEN 1 is of great importance to reduce morbidity and mortality from the associated tumors.

MULTIPLE ENDOCRINE NEOPLASIA TYPE 2 AND TYPE 3

Clinical Manifestations MEN type 2 (MEN 2), which is also called Sipple's syndrome, is characterized by the association of medullary

TABLE 400-4 Recommendations for Tests and Surgery in MEN 2 and MEN 3[a]

RET MUTATION, EXON (EX) LOCATION, AND CODON INVOLVED	RISK[b]	RECOMMENDED AGE (YEARS) FOR TEST/INTERVENTION				
		RET MUTATIONAL ANALYSIS	FIRST SERUM CALCITONIN AND NECK ULTRASOUND	PROPHYLACTIC THYROIDECTOMY	SCREENING FOR PHEOCHROMOCYTOMA	SCREENING FOR PHPT
Ex8 (533)[c]; Ex10 (609, 611, 618, 620)[c]; Ex11 (630, 631, 666)[c]; Ex13 (768, 790)[c]; Ex14 (804)[c]; Ex15 (891)[c]; EX16 (912)[c]	+	<3–5	5	<5[d]	16[e]	16
Ex11 (634)[c]; Ex15 (883)[c]	++	<3	<3	<5[f]	11[e]	11
Ex15 (883)[g]; Ex16 (918)[g]	+++	ASAP and by <1	ASAP and by <0.5–1	ASAP and by <1	11[e]	—[h]

[a]Data from American Thyroid Association Guidelines Task Force, RT Kloos et al: Medullary thyroid cancer: management guidelines of the American Thyroid Association. Thyroid 19:565, 2009 and revised from SA Wells Jr et al: Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. Thyroid 25:567, 2015. [b]Risk for early development of metastasis and aggressive growth of medullary thyroid cancer: +++, highest; ++, high; and + moderate. [c]Mutations associated with MEN 2A (or medullary thyroid carcinoma only). [d]Timing of surgery to be based on elevation of serum calcitonin and/or joint discussion with pediatrician, surgeon, and parent/family. Later surgery may be appropriate if serum calcitonin and neck ultrasound are normal. [e]Presence of pheochromocytoma must be excluded prior to any surgical intervention and also in women with *RET* mutation who are planning pregnancy or are pregnant. [f]Surgery earlier than 5 years based on elevation of serum calcitonin. Optimal timing of surgery should be decided by the surgeon and pediatrician, in consultation with the child's parent. [g]Mutations associated with MEN 2B (MEN 3). [h]Not required because PHPT is not a feature of MEN 2B (MEN 3).

Abbreviations: ASAP, as soon as possible; MEN, multiple endocrine neoplasia; PHPT, primary hyperparathyroidism.

thyroid carcinoma (MTC), pheochromocytomas, and parathyroid tumors (Table 400-1). Three clinical variants of MEN 2 are recognized: MEN 2A, MEN 2B, and MTC only. MEN 2A, which is often referred to as MEN 2, is the most common variant. In MEN 2A, MTC is associated with pheochromocytomas in 50% of patients (may be bilateral) and with parathyroid tumors in 20% of patients. MEN 2A may rarely occur in association with Hirschsprung's disease, caused by the absence of autonomic ganglion cells in the terminal hindgut, resulting in colonic dilatation, severe constipation, and obstruction. MEN 2A may also be associated with cutaneous lichen amyloidosis, which is a pruritic lichenoid lesion that is usually located on the upper back. MEN 2B, which is also referred to as MEN 3, represents 5% of all cases of MEN 2 and is characterized by the occurrence of MTC and pheochromocytoma in association with a Marfanoid habitus; mucosal neuromas of the lips, tongue, and eyelids; medullated corneal fibers; and intestinal autonomic ganglion dysfunction leading to multiple diverticulae and megacolon. Parathyroid tumors do not usually occur in MEN 2B. MTC only (FMTC) is a variant in which MTC is the sole manifestation of the syndrome. However, the distinction between FMTC and MEN 2A is difficult and should only be considered if there are at least four family members aged >50 years who are affected by MTC but not pheochromocytomas or primary hyperparathyroidism. All of the MEN 2 variants are due to mutations of the rearranged during transfection (*RET*) protooncogene, which encodes a TKR. Moreover, there is a correlation between the locations of *RET* mutations and MEN 2 variants. Thus, ~95% of MEN 2A patients have mutations involving the cysteine-rich extracellular domain, with mutations of codon 634 accounting for ~85% of MEN 2A mutations; FMTC patients also have mutations of the cysteine-rich extracellular domain, with most mutations occurring in codon 618. In contrast, ~95% of MEN 2B/MEN 3 patients have mutations of codon 918 of the intracellular tyrosine kinase domain (Table 400-1 and Table 400-4).

Medullary Thyroid Carcinoma MTC is the most common feature of MEN 2A and MEN 2B and occurs in almost all affected individuals. MTC represents 5–10% of all thyroid gland carcinomas, and 20% of MTC patients have a family history of the disorder. The use of *RET* mutational analysis to identify family members at risk for hereditary forms of MTC has altered the presentation of MTC from that of symptomatic tumors to a preclinical disease for which prophylactic thyroidectomy (Table 400-4) is undertaken to improve the prognosis and ideally result in cure. However, in patients who do not have a known family history of MEN 2A, FMTC, or MEN 2B, and therefore have not had *RET* mutational analysis, MTC may present as a palpable mass in the neck, which may be asymptomatic or associated with symptoms of pressure or dysphagia in >15% of patients. Diarrhea occurs in 30% of patients and is associated either with elevated circulating concentrations of calcitonin or tumor-related secretion of serotonin and prostaglandins. Some patients may also experience flushing. In addition, ectopic ACTH production by MTC may cause Cushing's syndrome. The diagnosis of MTC relies on the demonstration of hypercalcitoninemia (>90 pg/mL in the basal state); stimulation tests using IV pentagastrin (0.5 mg/kg) and or calcium infusion (2 mg/kg) are rarely used now, reflecting improvements in the assay for calcitonin. Neck ultrasonography with fine-needle aspiration of the nodules can confirm the diagnosis. Radionucleotide thyroid scans may reveal MTC tumors as "cold" nodules. Radiography may reveal dense irregular calcification within the involved portions of the thyroid gland and in lymph nodes involved with metastases. Positron emission tomography (PET) may help to identify the MTC and metastases **(Fig. 400-2)**. Metastases of MTC usually occur to the cervical lymph nodes in the early stages and to the mediastinal nodes, lung, liver, trachea, adrenal, esophagus, and bone in later stages. Elevations in serum calcitonin concentrations are often the first sign of recurrence or persistent disease, and the serum calcitonin doubling time is useful for determining prognosis. MTC can have an aggressive clinical course, with early metastases and death in ~10% of patients. A family history of aggressive MTC or MEN 2B may be elicited.

TREATMENT

Medullary Thyroid Carcinoma

Individuals with *RET* mutations who do not have clinical manifestations of MTC should be offered prophylactic surgery between the ages of <1 and 5 years. The timing of surgery will depend on the type of *RET* mutation and its associated risk for early development, metastasis, and aggressive growth of MTC (Table 400-4). Such patients should have a total thyroidectomy with a systematic central neck dissection to remove occult nodal metastasis, although the value of undertaking a central neck dissection has been subject to debate. Prophylactic thyroidectomy, with lifelong thyroxine replacement, has dramatically improved outcomes in patients with MEN 2 and MEN 3, such that ~90% of young patients with *RET* mutations who had a prophylactic thyroidectomy have no evidence of persistent or recurrent MTC at 7 years after surgery. In patients with clinically evident MTC, a total thyroidectomy with bilateral central resection is recommended, and an ipsilateral lateral neck dissection should be undertaken if the primary tumor is >1 cm in size or there is evidence of nodal metastasis in the central neck. Surgery is the only curative therapy for MTC. The 10-year survival in patients with metastatic MTC is ~20%. For inoperable MTC or metastatic disease, TKR inhibitors (e.g., vandetanib, cabozantinib, selpercatinib) have improved the progression-free survival times. Among these, selpercatinib is more selective for the RET kinase and is most effective. PRRT with ^{177}Lu-DOTATATE has been reported to be beneficial for metastatic MTCs that were found by SRS to express somatostatin receptors. Other types of chemotherapy are of limited efficacy, but radiotherapy may help to palliate local disease.

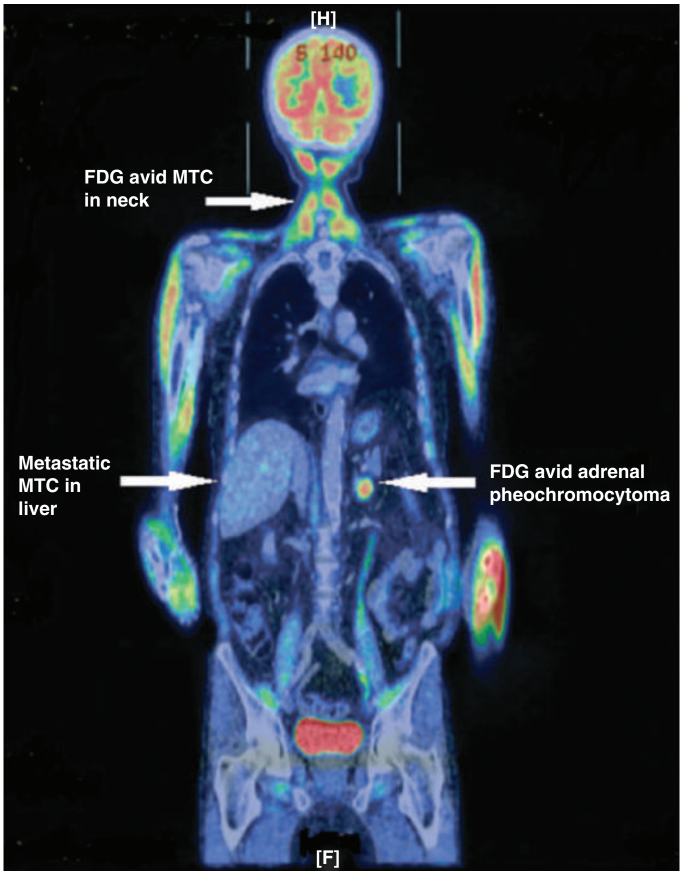

FIGURE 400-2 Fluorodeoxyglucose (FDG) positron emission tomography scan in a patient with multiple endocrine neoplasia type 2A, showing medullary thyroid cancer (MTC) with hepatic and skeletal (left arm) metastasis and a left adrenal pheochromocytoma. Note the presence of excreted FDG compound in the bladder. *(Reproduced with permission from A Naziat et al: Confusing genes: A patient with MEN2A and Cushing's disease. Clin Endocrinol (Oxf) 78:966, 2013.)*

Pheochromocytoma (See also Chap. 399) These noradrenaline- and adrenaline-secreting tumors occur in >50% of patients with MEN 2A and MEN 2B and are a major cause of morbidity and mortality. Patients may have symptoms and signs of catecholamine secretion (e.g., headaches, palpitations, sweating, poorly controlled hypertension), or they may be asymptomatic with detection through biochemical screening based on a history of familial MEN 2A, MEN 2B, or MTC. Pheochromocytomas in patients with MEN 2A and MEN 2B differ significantly in distribution when compared with patients without MEN 2A and MEN 2B. Extra-adrenal pheochromocytomas, which occur in 10% of patients without MEN 2A and MEN 2B, are observed rarely in patients with MEN 2A and MEN 2B. Malignant pheochromocytomas are much less common in patients with MEN 2A and MEN 2B. The biochemical and radiologic investigation of pheochromocytoma in patients with MEN 2A and MEN 2B is similar to that in non–MEN 2 patients and includes the measurement of plasma (obtained from supine patients) and urinary free fractionated metanephrines (e.g., normetanephrine and metanephrines measured separately), CT or MRI scanning, radionuclide scanning with meta-iodo-(^{123}I or ^{131}I)-benzyl guanidine (MIBG), and PET using (^{18}F)-fluorodopamine or (^{18}F)-fluoro-2-dexoxy-D-glucose (Fig. 400-2).

TREATMENT

Pheochromocytoma

Surgical removal of pheochromocytoma, using α and β adrenoreceptor blockade before and during the operation, is the recommended treatment. Other antihypertensive agents, including calcium channel blockers, are sometimes required for adequate blood pressure control. Endoscopic adrenal-sparing surgery, which decreases postoperative morbidity, hospital stay, and expense, as opposed to open surgery, has become the method of choice.

Parathyroid Tumors (See also Chap. 422) Parathyroid tumors occur in 10–25% of patients with MEN 2A. However, >50% of these patients do not have hypercalcemia. The presence of abnormally enlarged parathyroids, which are unusually hyperplastic, is often seen in the normocalcemic patient undergoing thyroidectomy for MTC. The biochemical investigation and treatment of hypercalcemic patients with MEN 2A is similar to that of patients with MEN 1.

Genetics and Screening To date, ~50 different *RET* mutations have been reported, and these are located in exons 5, 8, 10, 11, 13, 14, 15, and 16. *RET* germline mutations are detected in >95% of MEN 2A, FMTC, and MEN 2B families, with Cys634Arg being most common in MEN 2A, Cys618Arg being most common in FMTC, and Met918Thr being most common in MEN 2B (Tables 400-1 and 400-4). Between 5 and 10% of patients with MTC or MEN 2A–associated tumors have de novo *RET* germline mutations, and ~50% of patients with MEN 2B have de novo *RET* germline mutations. These de novo *RET* germline mutations always occur on the paternal allele. Approximately 5% of patients with sporadic pheochromocytoma have a germline *RET* mutation, but such germline *RET* mutations do not appear to be associated with sporadic primary hyperparathyroidism. Thus, *RET* mutational analysis should be performed in (1) all patients with MTC who have a family history of tumors associated with MEN 2, FMTC, or MEN 3, such that the diagnosis can be confirmed and genetic testing offered to asymptomatic relatives; (2) all patients with MTC and pheochromocytoma without a known family history of MEN 2 or MEN 3; (3) all patients with MTC, but without a family history of MEN 2, FMTC, or MEN 3, because these patients may have a de novo germline *RET* mutations; (4) all patients with bilateral pheochromocytoma; and (5) patients with unilateral pheochromocytoma, particularly if this occurs with increased calcitonin levels.

Screening for MEN 2/MEN 3–associated tumors in patients with *RET* germline mutations should be undertaken annually and include serum calcitonin measurements, a neck ultrasound for MTC, plasma (or 24-h urinary) fractionated metanephrines for pheochromocytoma, and albumin-corrected serum calcium or ionized calcium with PTH for primary hyperparathyroidism. In patients with MEN 2–associated *RET* mutations, screening for MTC should begin by 1–5 years, for pheochromocytoma by 11–16 years, and for primary hyperparathyroidism by 11–16 years of age (Table 400-4).

MULTIPLE ENDOCRINE NEOPLASIA TYPE 4

Clinical Manifestations Patients with MEN 1–associated tumors, such as parathyroid adenomas, pituitary adenomas, and pancreatic NETs, occurring in association with gonadal, adrenal, renal, and thyroid tumors have been reported to have mutations of the gene encoding the 196–amino acid cyclin-dependent kinase inhibitor (CK1) p27 kip1 (*CDKN1B*). Such families with MEN 1–associated tumors and *CDKN1B* mutations are designated to have MEN 4 (Table 400-1). The investigations and treatments for the MEN 4–associated tumors are similar to those for MEN 1 and non–MEN 1 tumors.

Genetics and Screening To date, 50 MEN patients (from <20 kindreds) with mutations of *CDKN1B*, which is located on chromosome 12p13, have been reported, and all of these are predicted to result in a loss of function. These MEN 4 patients may represent ~3% of the 5–10% of patients with MEN 1 who do not have mutations of the *MEN1* gene. Germline *CDKN1B* mutations may rarely be found in patients with sporadic (i.e., nonfamilial) forms of primary hyperparathyroidism.

MULTIPLE ENDOCRINE NEOPLASIA TYPE 5

Clinical Manifestations Two kindreds in whom multiple endocrine tumors were associated with heterozygous germline *MAX* mutations have been reported. Although *MAX* mutations have been described in rare cases of hereditary pheochromocytoma and paraganglioma, the occurrence of these tumors in these kindreds was also associated with pituitary adenomas, parathyroid adenomas, and nonendocrine tumors including chondrosarcoma, lung adenocarcinoma, ganglioneuroma, and neuroblastoma. Additional tumors reported in patients with germline *MAX* mutations include pancreatic NETs, renal oncocytomas, and renal carcinoma.

Genetics and Screening The *MAX* gene is located on chromosome 14q23.3 and encodes the ubiquitously expressed MYC-associated factor X transcription factor, which is presumed to act as a tumor suppressor. The spectrum of clinical manifestations associated with germline *MAX* mutations and MEN 5 remains to be fully defined, although most patients appear to have development of early-onset pheochromocytoma.

HYPERPARATHYROIDISM-JAW TUMOR SYNDROME (SEE ALSO CHAP. 422)

Clinical Manifestations Hyperparathyroidism-jaw tumor (HPT-JT) syndrome is an autosomal dominant disorder characterized by the development of parathyroid tumors (15% are carcinomas) and fibro-osseous jaw tumors. In addition, some patients may also develop Wilms' tumors, renal cysts, renal hamartomas, renal cortical adenomas, renal cell carcinoma (RCC), pancreatic adenocarcinomas, uterine tumors, testicular mixed germ cell tumors with a major seminoma component, and Hürthle cell thyroid adenomas. The parathyroid tumors may occur in isolation and without any evidence of jaw tumors, and this may cause confusion with other hereditary hypercalcemic disorders, such as MEN 1. However, genetic testing to identify the causative mutation will help to establish the correct diagnosis. The investigation and treatment for HPT-JT–associated tumors are similar to those in non-HPT-JT patients, except that early parathyroidectomy is advisable because of the increased frequency of parathyroid carcinoma.

Genetics and Screening The gene that causes HPT-JT is located on chromosome 1q31.2 and encodes a 531–amino acid protein, parafibromin (Table 400-2). Parafibromin is also referred to as cell division cycle protein 73 (CDC73) and has a role in transcription. Genetic testing in families helps to identify mutation carriers who should be periodically screened for the development of tumors (Table 400-5).

VON HIPPEL–LINDAU DISEASE (SEE ALSO CHAP. 399)

Clinical Manifestations von Hippel–Lindau (VHL) disease is an autosomal dominant disorder characterized by hemangioblastomas of the retina and CNS; cysts involving the kidneys, pancreas, and epididymis; RCC; pheochromocytomas; and pancreatic islet cell tumors. The retinal and CNS hemangioblastomas are benign vascular tumors that may be multiple; those in the CNS may cause symptoms by compressing adjacent structures and/or increasing intracranial pressure. In the CNS, the cerebellum and spinal cord are the most frequently involved sites. The renal abnormalities consist of cysts and carcinomas, and the lifetime risk of RCC in VHL is 70%. The endocrine tumors in VHL consist of pheochromocytomas and pancreatic islet cell tumors. The clinical presentation of pheochromocytoma in VHL disease is similar to that in sporadic cases, except that there is a higher frequency of bilateral or multiple tumors, which may involve extra-adrenal sites in VHL disease. The most frequent pancreatic lesions in VHL are multiple cyst-adenomas, which rarely cause clinical disease. However, nonsecreting pancreatic islet cell tumors occur in <10% of VHL patients, who are usually asymptomatic. The pancreatic tumors in these patients are often detected by regular screening using abdominal imaging. Pheochromocytomas should be investigated and treated as described earlier for MEN 2. The pancreatic islet cell tumors frequently become malignant, and early surgery is recommended.

TABLE 400-5 HPT-JT Screening Guidelines

TUMOR[a]	TEST	FREQUENCY[b]
Parathyroid	Serum Ca, PTH	6–12 months
Ossifying jaw fibroma	Panoramic jaw x-ray with neck shielding[c]	5 years
Renal	Abdominal MRI[c,d]	5 years
Uterine	Ultrasound (transvaginal or transabdominal) and additional imaging ± D&C if indicated[e]	Annual

[a]Screening for most common HPT-JT–associated tumors is considered. Assessment for other reported tumor types may be indicated (e.g., pancreatic, thyroid, testicular tumors). [b]Frequency of repeating test after baseline tests performed. [c]X-rays and imaging involving ionizing radiation should ideally be avoided to minimize risk of generating subsequent mutations. [d]Ultrasound scan recommended if MRI unavailable. [e]Such selective pelvic imaging should be considered after obtaining a detailed menstrual history.

Abbreviations: Ca, calcium; D&C, dilation and curettage; HPT-JT, hyperparathyroidism-jaw tumor syndrome; MRI, magnetic resonance imaging; PTH, parathyroid hormone.

Source: Reproduced with permission from PJ Newey, MR Bowl, T Cranston et al: Cell division cycle protein 73 homolog (CDC73) mutations in the hyperparathyroidism-jaw tumor syndrome (HPT-JT) and parathyroid tumors. Hum Mutat 31:295, 2010.

Genetics and Screening The *VHL* gene, which is located on chromosome 3p26-p25, is widely expressed in human tissues and encodes a 213–amino acid protein (pVHL) (Table 400-2). A wide variety of germline *VHL* mutations have been identified. *VHL* acts as a tumor-suppressor gene. A correlation between the type of mutation and the clinical phenotype has been reported; large deletions and protein-truncating mutations are associated with a low incidence of pheochromocytomas, whereas some missense mutations in VHL patients are associated with pheochromocytoma (referred to as VHL type 2C). Other missense mutations may be associated with hemangioblastomas and RCC but not pheochromocytoma (referred to as VHL type 1), whereas distinct missense mutations are associated with hemangioblastomas, RCC, and pheochromocytoma (VHL type 2B). VHL type 2A, which refers to the occurrence of hemangioblastomas and pheochromocytoma without RCC, is associated with rare missense mutations. The basis for these complex genotype-phenotype relationships remains to be elucidated. One major function of pVHL, which is also referred to as elongin, is to downregulate the expression of vascular endothelial growth factor (VEGF) and other hypoxia-inducible mRNAs. Thus, pVHL, in complex with other proteins, regulates the expression of hypoxia-inducible factors (HIF-1 and HIF-2) such that loss of functional pVHL leads to a stabilization of the HIF protein complexes, resulting in VEGF overexpression and tumor angiogenesis. Screening for the development of pheochromocytomas and pancreatic islet cell tumors is as described earlier for MEN 2 and MEN 1, respectively (Tables 400-3 and 400-4).

NEUROFIBROMATOSIS

Clinical Manifestations Neurofibromatosis type 1 (NF1), which is also referred to as von Recklinghausen's disease, is an autosomal dominant disorder characterized by the following manifestations: neurologic (e.g., peripheral and spinal neurofibromas); ophthalmologic (e.g., optic gliomas and iris hamartomas such as Lisch nodules); dermatologic (e.g., café au lait macules); skeletal (e.g., scoliosis, macrocephaly, short stature, pseudoarthrosis); vascular (e.g., stenoses of renal and intracranial arteries); and endocrine (e.g., pheochromocytoma, carcinoid tumors, precocious puberty). Neurofibromatosis type 2 (NF2) is also an autosomal dominant disorder but is characterized by the development of bilateral vestibular schwannomas (acoustic neuromas) that lead to deafness, tinnitus, or vertigo. Some patients with NF2 also develop meningiomas, spinal schwannomas, peripheral nerve neurofibromas, and café au lait macules. Endocrine abnormalities are not found in NF2 and are associated solely with NF1. Pheochromocytomas, carcinoid tumors, and precocious puberty occur in ~1% of patients with NF1, and growth hormone deficiency has also been reported. The features of pheochromocytomas in NF1 are similar to those in non-NF1 patients, with 90% of tumors being located within the adrenal medulla and the remaining 10% at an extra-adrenal location, which often involves the para-aortic region. Primary carcinoid tumors are often periampullary and may also occur in the ileum but rarely in the pancreas, thyroid, or lungs. Hepatic metastases are associated with symptoms of the carcinoid syndrome, which include flushing, diarrhea, bronchoconstriction, and tricuspid valve disease.

Precocious puberty is usually associated with the extension of an optic glioma into the hypothalamus with resultant early activation of gonadotropin-releasing hormone secretion. Growth hormone deficiency has also been observed in some NF1 patients, who may or may not have optic chiasmal gliomas, but it is important to note that short stature is frequent in the absence of growth hormone deficiency in patients with NF1. The investigation and treatment for tumors are similar to those undertaken for each respective tumor type in non-NF1 patients.

Genetics and Screening The *NF1* gene, which is located on chromosome 17q11.2 and acts as a tumor suppressor, consists of 60 exons that span >350 kb of genomic DNA (Table 400-2). Mutations in *NF1* are of diverse types and are scattered throughout the exons. The NF1 gene product is the protein neurofibromin, which has homologies to the p120GAP (GTPase activating protein) and acts on p21ras by converting the active GTP bound form to its inactive GDP form. Mutations of *NF1* impair this downregulation of the p21ras signaling pathways, which in turn results in abnormal cell proliferation. Screening for the development of pheochromocytomas and carcinoid tumors is as described earlier for MEN 2 and MEN 1, respectively (Tables 400-3 and 400-4).

CARNEY COMPLEX

Clinical Manifestations Carney complex (CNC) is an autosomal dominant disorder characterized by spotty skin pigmentation (usually of the face, labia, and conjunctiva), myxomas (usually of the eyelids and heart, but also the tongue, palate, breast, and skin), psammomatous melanotic schwannomas (usually of the sympathetic nerve chain and upper gastrointestinal tract), and endocrine tumors that involve the adrenals, Sertoli cells, somatotropes, thyroid, and ovary. Cushing's syndrome, the result of primary pigmented nodular adrenal disease (PPNAD), is the most common endocrine manifestation of CNC and may occur in one-third of patients. Patients with CNC and Cushing's syndrome often have an atypical appearance by being thin (as opposed to having truncal obesity). In addition, they may have short stature, muscle and skin wasting, and osteoporosis. These patients often have levels of urinary free cortisol that are normal or increased only marginally. Cortisol production may fluctuate periodically with days or weeks of hypercortisolism; this pattern is referred to as "periodic Cushing's syndrome." Patients with Cushing's syndrome usually have loss of the circadian rhythm of cortisol production. Acromegaly, the result of a somatotrope tumor, affects ~10% of patients with CNC. Testicular tumors may also occur in one-third of patients with CNC. These may either be large-cell calcifying Sertoli cell tumors, adrenocortical rests, or Leydig cell tumors. The Sertoli cell tumors occasionally may be estrogen-secreting and lead to precocious puberty or gynecomastia. Some patients with CNC have been reported to develop thyroid follicular tumors, ovarian cysts, or breast duct adenomas.

Genetics and Screening CNC type 1 (CNC1) is due to mutations of the protein kinase A (PKA) regulatory subunit 1 α (R1α) (*PRAKAR1A*), a tumor suppressor, whose gene is located on chromosome 17q.24.2 (Table 400-2). The gene causing CNC type 2 (CNC2) is located on chromosome 2p16 and has not yet been identified. It is interesting to note, however, that some tumors do not show LOH of 2p16 but instead show genomic instability, suggesting that this CNC gene may not be a tumor suppressor. Screening and treatment of these endocrine tumors are similar to those described earlier for patients with MEN 1 and MEN 2 (Tables 400-3 and 400-4).

COWDEN'S SYNDROME

Clinical Manifestations Multiple hamartomatous lesions, especially of the skin, mucous membranes (e.g., buccal, intestinal, colonic), breast, and thyroid, are characteristic of Cowden's syndrome (CWS), which is an autosomal dominant disorder. Thyroid abnormalities occur in two-thirds of patients with CWS, and these usually consist of multinodular goiters or benign adenomas, although <10% of patients may have a follicular thyroid carcinoma. Breast abnormalities occur in >75% of patients and consist of either fibrocystic disease or adenocarcinomas. The investigation and treatment for CWS tumors are similar to those undertaken for non-CWS patients.

Genetics and Screening CWS is genetically heterogenous, and seven types (CWS1–7) are recognized (Table 400-2). CWS1 is due to mutations of the phosphate and tensin homologue deleted on chromosome 10 (*PTEN*) gene, located on chromosome 10q23.31. CWS2 is caused by mutations of the succinate dehydrogenase subunit B (*SDHB*) gene, located on chromosome 1p36.13; and CWS3 is caused by mutations of the *SDHD* gene, located on chromosome 11q13.1. *SDHB* and *SDHD* mutations are also associated with pheochromocytoma. CWS4 is caused by hypermethylation of the Killin (*KLLN*) gene, the promoter of which shares the same transcription site as *PTEN* on chromosome 10q23.31. CWS5 is caused by mutations of the phosphatidylinositol 3-kinase catalytic alpha (*PIK3CA*) gene on chromosome 3q26.32. CWS6 is caused by mutations of the V-Akt murine thymoma viral oncogene homolog 1 (*AKT1*) gene on chromosome 14q32.33, and CWS7 is caused by mutations of the saccharomyces cerevisiae homology of B (*SEC23B*) gene on chromosome 20p11.23. Screening for thyroid abnormalities entails neck ultrasonography and fine-needle aspiration with analysis of cell cytology.

MCCUNE-ALBRIGHT SYNDROME (SEE ALSO CHAP. 424)

Clinical Manifestations McCune-Albright syndrome (MAS) is characterized by the triad of polyostotic fibrous dysplasia, which may be associated with hypophosphatemic rickets; café au lait skin pigmentation; and peripheral precocious puberty. Other endocrine abnormalities include thyrotoxicosis, which may be associated with a multinodular goiter, somatotrope tumors, and Cushing's syndrome (due to adrenal tumors). Investigation and treatment for each endocrinopathy are similar to those used in patients without MAS.

Genetics and Screening MAS is a disorder of mosaicism that results from postzygotic somatic cell mutations of the G protein α-stimulating subunit ($G_s\alpha$), encoded by the *GNAS1* gene, located on chromosome 20q13.32 (Table 400-2). The $G_s\alpha$ mutations, which include Arg201Cys, Arg201His, Glu227Arg, or Glu227His, are activating and are found only in cells of the abnormal tissues. Screening for hyperfunction of relevant endocrine glands and development of hypophosphatemia, which may be associated with elevated serum fibroblast growth factor 23 (FGF23) concentrations, is undertaken in MAS patients.

ACKNOWLEDGMENT

The author is grateful to the National Institute of Health Research (NIHR) Oxford Biomedical Research Centre Programme for support and to Mrs. Tracey Walker for typing the manuscript.

FURTHER READING

Binderup MLM et al: von Hippel-Lindau disease: Updated guideline for diagnosis and surveillance. Eur J Med Genet 65:104538, 2022.

Bouys L, Bertherat J: Management of endocrine disease: Carney complex: Clinical and genetic update 20 years after the identification of the CNC1 (PRKAR1A) gene. Eur J Endocrinol 184:R99, 2021.

Brandi ML et al: Multiple endocrine neoplasia type 1: Latest insights. Endocr Rev 42:133, 2021.

Legius E et al: Revised diagnostic criteria for neurofibromatosis type 1 and Legius syndrome: An international consensus recommendation. Genet Med 23:1506, 2021.

Hadoux J et al: Phase 3 trial of selpercatinib in advanced RET-mutant medullary thyroid cancer. N Engl J Med 389:1851, 2023.

Minisola S et al: Epidemiology, pathophysiology, and genetics of primary hyperparathyroidism. J Bone Miner Res 37:2315, 2022.

Ruggeri RM et al: Multiple endocrine neoplasia type 4 (MEN4): A thorough update on the latest and least known MEN syndrome. Endocrine 82:480, 2023.

Seabrook AJ et al: Multiple endocrine tumors associated with germline MAX mutations: Multiple endocrine neoplasia type 5? J Clin Endocrinol Metab 106:1163, 2021.

Thakker RV et al: Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). J Clin Endocrinol Metab 97:2990, 2012.

401

Autoimmune Polyendocrine Syndromes

Peter A. Gottlieb, Aaron W. Michels

Polyglandular deficiency syndromes have been given many different names, reflecting the wide spectrum of disorders that have been associated with these syndromes and the heterogeneity of their clinical presentations. The name used in this chapter for this group of disorders is *autoimmune polyendocrine syndrome* (APS). In general, these disorders are divided into two major categories, APS type 1 (APS-1) and APS type 2 (APS-2). Some groups have further subdivided APS-2 into APS type 3 (APS-3) and APS type 4 (APS-4) depending on the type of autoimmunity involved. For the most part, this additional classification does not clarify our understanding of disease pathogenesis or prevention of complications in individual patients. Importantly, there are many nonendocrine disease associations included in these syndromes, suggesting that although the underlying autoimmune disorder predominantly involves endocrine targets, it does not exclude other tissues. The disease associations found in APS-1 and APS-2 are summarized in Table 401-1. Understanding these syndromes and their disease manifestations can lead to early diagnosis and treatment of additional disorders in patients and their family members.

TABLE 401-1 Disease Associations with Autoimmune Polyendocrine Syndromes

AUTOIMMUNE POLYENDOCRINE SYNDROME TYPE 1	AUTOIMMUNE POLYENDOCRINE SYNDROME TYPE 2	OTHER AUTOIMMUNE POLYENDOCRINE DISORDERS
Endocrine	**Endocrine**	IPEX (immune dysfunction polyendocrinopathy X-linked)
Addison's disease	Addison's disease	Thymic tumors
Hypoparathyroidism	Type 1 diabetes	Anti-insulin receptor antibodies
Hypogonadism	*Graves' disease or autoimmune thyroiditis*	POEMS syndrome
Graves' disease or autoimmune thyroiditis	*Hypogonadism*	Insulin autoimmune syndrome (Hirata's syndrome)
Type 1 diabetes		Adult combined pituitary hormone deficiency (CPHD) with anti-Pit1 autoantibodies
		Kearns-Sayre syndrome
		DIDMOAD syndrome
Nonendocrine	**Nonendocrine**	Congenital rubella associated with thyroiditis and/or diabetes
Mucocutaneous candidiasis	Celiac disease, dermatitis herpetiformis	
Chronic active hepatitis	Pernicious anemia	
Pernicious anemia	Vitiligo	
Vitiligo	*Alopecia*	
Asplenism	*Myasthenia gravis*	
Ectodermal dysplasia	*IgA deficiency*	
Alopecia	*Parkinson's disease*	
Malabsorption syndromes	*Idiopathic thrombocytopenia*	
IgA deficiency		

Abbreviations: DIDMOAD, *d*iabetes *i*nsipidus, *d*iabetes *m*ellitus, progressive bilateral *o*ptic *a*trophy, and sensorineural *d*eafness; POEMS, *p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathy, *M*-protein, and *s*kin changes.

Note: Italics denote less common disorders.

APS-1

APS-1 (Online Mendelian Inheritance in Man [OMIM] 240300) has also been called autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy (APECED). Mucocutaneous candidiasis, hypoparathyroidism, and Addison's disease form the three major components of this disorder. However, as summarized in Table 401-1, many other organ systems can be involved over time. APS-1 is rare, with <500 cases reported in the literature.

The classical form of APS-1 is an autosomal recessive disorder caused by mutations in the *AIRE* gene (autoimmune regulator gene) found on chromosome 21. This gene is most highly expressed in thymic medullary epithelial cells (mTECs) where it controls the expression of tissue-specific self-antigens (e.g., insulin). Deletion of this regulator leads to decreased expression of tissue-specific self-antigens and is hypothesized to allow autoreactive T cells to avoid central deletion, which normally occurs during T-cell maturation in the thymus. The *AIRE* gene is also expressed in epithelial cells found in peripheral lymphoid organs, but its role in these extrathymic cells remains controversial. To date, >100 mutations have been described in this gene, and there is a higher frequency within certain ethnic groups including Iranian Jews, Sardinians, Finns, Norwegians, and Irish. Recently, several autosomal dominant mutations have been identified and are localized primarily in the PHD1 domain of the *AIRE* gene, rather than the CARD region, where the autosomal recessive mutations have been found. Individuals with this nonclassical form of APS-1 may have a later onset of symptoms and less aggressive disease, without the full spectrum of autoimmune components being expressed.

Clinical Manifestations Classical APS-1 develops very early in life, often in infancy (Table 401-2). Chronic mucocutaneous candidiasis without signs of systemic disease is often the first manifestation. It affects the mouth and nails more frequently than the skin and esophagus. Chronic oral candidiasis can result in atrophic disease with areas suggestive of leukoplakia, which can pose a risk for future carcinoma. The etiology is associated with anticytokine autoantibodies (anti-interleukin [IL] 17A, IL-17F, and IL-22) related to T helper (T_H) 17 T cells and depressed production of these cytokines by peripheral blood mononuclear cells. Hypoparathyroidism usually develops next, followed by adrenal insufficiency. The time from development of one component of the disorder to the next can be many years, and the order of disease appearance is variable.

Chronic candidiasis is nearly always present and is not very responsive to treatment. Hypoparathyroidism is found in >85% of cases, and Addison's disease is found in nearly 80%. Gonadal failure appears to affect women more than men (70 vs 25%, respectively), and hypoplasia of the dental enamel also occurs frequently (77% of patients).

TABLE 401-2 Comparison of APS-1 and APS-2

APS-1	APS-2
Early onset: infancy	Later onset
Siblings often affected and at risk	Multigenerational
Equivalent sex distribution	Females > males affected
Monogenic: *AIRE* gene, chromosome 21, autosomal recessive	Polygenic: *HLA, MICA, PTNP22, CTLA4*
Not HLA associated for entire syndrome, some specific component risk	DR3/DR4 associated; other HLA class III gene associations noted
Autoantibodies to type 1 interferons and IL-17 and IL-22	No autoantibodies to cytokines
Autoantibodies to specific target organs	Autoantibodies to specific target organs
Asplenism	No defined immunodeficiency
Mucocutaneous candidiasis	Association with other nonendocrine immunologic disorders like myasthenia gravis and idiopathic thrombocytopenic purpura

Abbreviations: APS, autoimmune polyendocrine syndrome; HLA, human leukocyte antigen; IL, interleukin.

Other endocrine disorders that occur less frequently include type 1 diabetes (23%) and autoimmune thyroid disease (18%). Nonendocrine manifestations that present less frequently include alopecia (40%), vitiligo (26%), intestinal malabsorption (18%), pernicious anemia (31%), chronic active hepatitis (17%), and nail dystrophy. An unusual and debilitating manifestation of the disorder is the development of refractory diarrhea/obstipation that may be related to autoantibody-mediated destruction of enterochromaffin or enterochromaffin-like cells. The incidence rates for many of these disorders peak in the first or second decade of life, but the individual disease components continue to emerge over time. Therefore, prevalence rates may be higher than originally reported.

Diagnosis The diagnosis of APS-1 is usually made clinically when two of the three major component disorders are found in an individual patient. Siblings of individuals with APS-1 should be considered affected even if only one component disorder has been detected due to the known inheritance of the syndrome. Genetic analysis of the *AIRE* gene should be undertaken to identify mutations. Detection of anti–interferon α and anti–interferon ω antibodies can identify nearly 100% of cases with APS-1. The autoantibody arises independent of the type of *AIRE* gene mutation and is not found in other autoimmune disorders.

Diagnosis of each underlying disorder should be done based on their typical clinical presentations (Table 401-3). Mucocutaneous candidiasis may present throughout the gastrointestinal tract, and it may be detected in the oral mucosa or from stool samples. Evaluation by a gastroenterologist to examine the esophagus for candidiasis or secondary stricture may be merited based on symptoms. Other gastrointestinal manifestations of APS-1, including malabsorption and obstipation, may also bring these young patients to the attention of gastroenterologists for first evaluation. Specific physical examination findings of hyperpigmentation, vitiligo, alopecia, tetany, and signs of hyper- or hypothyroidism should be considered as signs of development of component disorders.

The development of disease-specific autoantibody assays can help confirm disease and also detect risk for future disease. For example, where possible, detection of anticytokine antibodies to IL-17 and IL-22 would confirm the diagnosis of mucocutaneous candidiasis due to APS-1. The presence of anti-21-hydroxylase antibody or anti-17-hydroxylase antibody (which may be found more commonly in adrenal insufficiency associated with APS-1) would confirm the presence or risk for Addison's disease. Other autoantibodies found in type 1 diabetes (e.g., anti-GAD65), pernicious anemia, and other component conditions should be screened for on a regular basis (6- to 12-month intervals depending on the age of the subject).

Laboratory tests, including a complete metabolic panel, phosphorous and magnesium, thyroid-stimulating hormone (TSH), adrenocorticotropic hormone (ACTH; morning), hemoglobin A_{1c}, plasma vitamin B_{12} level, and complete blood count with peripheral smear looking for Howell-Jolly bodies (asplenism), should also be performed at these time points. Detection of abnormal physical findings or test results should prompt subsequent examinations of the relevant organ system (e.g., presence of Howell-Jolly bodies indicates need for ultrasound of spleen).

TABLE 401-3 Clinical Features and Recommended Follow-Up for APS-1 and APS-2

COMPONENT DISEASE	RECOMMENDED EVALUATION
APS-1	
Addison's disease	Sodium, potassium, ACTH, cortisol, 21- and 17-hydroxylase autoantibodies
Diarrhea	History
Ectodermal dysplasia	Physical examination
Hypoparathyroidism	Serum calcium, phosphate, PTH
Hepatitis	Liver function tests
Hypothyroidism/Graves' disease	TSH; thyroid peroxidase and/or thyroglobulin autoantibodies and anti-TSH receptor Ab
Male hypogonadism	FSH/LH, testosterone
Malabsorption	Physical examination, anti-IL-17 and anti-IL-22 autoantibodies
Mucocutaneous candidiasis	Physical examination, mucosal swab, stool samples
Obstipation	History
Ovarian failure	FSH/LH, estradiol
Pernicious anemia	CBC, vitamin B_{12} levels
Splenic atrophy	Blood smear for Howell-Jolly bodies; platelet count; ultrasound if positive
Type 1 diabetes	Glucose, hemoglobin A_{1c}, diabetes-associated autoantibodies (insulin, GAD65, IA-2, ZnT8)
APS-2	
Addison's disease	21-Hydroxylase autoantibodies, ACTH stimulation testing if positive
Alopecia	Physical examination
Autoimmune hyper- or hypothyroidism	TSH; thyroid peroxidase and/or thyroglobulin autoantibodies, anti-TSH receptor Ab
Celiac disease	Transglutaminase autoantibodies; small intestine biopsy if positive
Cerebellar ataxia	Dictated by signs and symptoms of disease
Chronic inflammatory demyelinating polyneuropathy	Dictated by signs and symptoms of disease
Hypophysitis	Dictated by signs and symptoms of disease, anti-Pit1 autoantibody
Idiopathic heart block	Dictated by signs and symptoms of disease
IgA deficiency	IgA level
Myasthenia gravis	Dictated by signs and symptoms of disease, antiacetylcholinesterase Ab
Myocarditis	Dictated by signs and symptoms of disease
Pernicious anemia	Anti–parietal cell autoantibodies
	CBC, vitamin B_{12} levels if positive
Serositis	Dictated by signs and symptoms of disease
Stiff man syndrome	Dictated by signs and symptoms of disease
Vitiligo	Physical examination, NALP-1 polymorphism

Abbreviations: Ab, antibody; ACTH, adrenocorticotropic hormone; APS, autoimmune polyendocrine syndrome; CBC, complete blood count; FSH, follicle-stimulating hormone; IL, interleukin; LH, luteinizing hormone; PTH, parathyroid hormone; TSH, thyroid-stimulating hormone.

TREATMENT

APS-1

Therapy of individual disease components is carried out as outlined in other relevant chapters. Replacement of deficient hormones (e.g., adrenal, pancreas, ovaries/testes) will treat most of the endocrinopathies noted. Several unique issues merit special emphasis. Adrenal insufficiency can be masked by primary hypothyroidism by prolonging the half-life of cortisol. The caveat therefore is that replacement therapy with thyroid hormone can precipitate an adrenal crisis in an undiagnosed individual. Hence, all patients with hypothyroidism and the possibility of APS should be screened for adrenal insufficiency to allow treatment with glucocorticoids prior to the initiation of thyroid hormone replacement. Treatment of mucocutaneous candidiasis with ketoconazole in an individual with subclinical adrenal insufficiency may also precipitate adrenal crisis. Furthermore, mucocutaneous candidiasis may be difficult to eradicate entirely. Severe cases of disease involvement may require systemic immunomodulatory therapy, but this is not commonly needed.

APS-2

APS-2 (OMIM 269200) is more common than APS-1, with a prevalence of 1–2 in 100,000. It has a gender bias and occurs more often in female patients, with a ratio of at least 3:1 compared to male patients. In contrast to APS-1, APS-2 often has its onset in adulthood, with a peak incidence between 20 and 60 years of age. It shows a familial, multigenerational heritage (Table 401-2). The presence of two or more of the following endocrine deficiencies in the same patient defines the presence of APS-2: primary adrenal insufficiency (Addison's disease; 50–70%), Graves' disease or autoimmune thyroiditis (15–69%), type 1 diabetes mellitus (T1D; 40–50%), and primary hypogonadism. Frequently associated autoimmune conditions include celiac disease (3–15%), myasthenia gravis, vitiligo, alopecia, serositis, and pernicious anemia. These conditions occur with increased frequency in affected patients but are also found in their family members (Table 401-3).

Genetic Considerations The overwhelming risk factor for APS-2 has been localized to the genes in the human lymphocyte antigen (HLA) complex on chromosome 6. Primary adrenal insufficiency in APS-2, but not APS-1, is strongly associated with both HLA-DR3 and HLA-DR4. Other class I and class II genes and alleles, such as HLA-B8, HLA-DQ2 and HLA-DQ8, and HLA-DR subtypes such as *DRB1*04:04*, appear to contribute to organ-specific disease susceptibility **(Table 401-4)**. HLA-B8- and HLA-DR3-associated illnesses include selective IgA deficiency, juvenile dermatomyositis, dermatitis herpetiformis, alopecia, scleroderma, autoimmune thrombocytopenia purpura, hypophysitis, metaphyseal osteopenia, and serositis.

Several other immune genes have been proposed to be associated with Addison's disease and therefore with APS-2 (Table 401-3). The "5.1" allele of a major histocompatibility complex (MHC) gene is an atypical class I HLA molecule MIC-A. The MIC-A5.1 allele has a very strong association with Addison's disease that is not accounted for by linkage disequilibrium with DR3 or DR4. Its role is complicated because certain HLA class I genes can offset this effect. *PTPN22* codes for a polymorphism in a protein tyrosine phosphatase, which acts on intracellular signaling pathways in both T and B lymphocytes. It has been implicated in T1D, Addison's disease, and other autoimmune conditions. *CTLA4* is a receptor on the T-cell surface that modulates the activation state of the cell as part of the signal 2 pathway (i.e., binding to CD80/86 on antigen presenting cells). Polymorphisms of this gene appear to cause downregulation of the cell surface expression of the receptor, leading to decreased T-cell activation and proliferation. This appears to contribute to Addison's disease and potentially other components of APS-2. Allelic variants of the IL-2Rα are linked to development of T1D and autoimmune thyroid disease and could contribute to the phenotype of APS-2 in certain individuals.

Diagnosis When one of the component disorders is present, a second associated disorder occurs more commonly than in the general population (Table 401-3). There is controversy as to which tests to use and how often to screen individuals for disease. A strong family history of autoimmunity should raise suspicion in an individual with an initial component diagnosis. The development of a rarer form of autoimmunity, such as Addison's disease, should prompt more extensive screening for other linked disorders, as ~50% of Addison's disease patients develop another autoimmune disease during their lifetime.

Circulating autoantibodies, as previously discussed, can precede the development of clinical disease by many years but would allow the clinician to follow the patient and identify the disease onset at its earliest time point (Tables 401-3 and 401-4). For each of the endocrine components of the disorder, appropriate autoantibody assays are listed and, if positive, should prompt physiologic testing to diagnose clinical or subclinical disease. For Addison's disease, antibodies to 21-hydroxylase antibodies are highly diagnostic for risk of adrenal insufficiency. However, individuals may take many years to develop overt symptoms of hypoadrenalism. Screening of 21-hydroxylase antibody–positive patients can be performed measuring morning ACTH and cortisol on a yearly basis. Rising ACTH values over time or low morning cortisol in association with signs or symptoms of adrenal insufficiency should prompt testing via the cosyntropin stimulation test **(Chap. 398)**. T1D can be screened for by measuring autoantibodies directed against insulin, GAD65, IA-2, and ZnT8. Risk for progression to disease is based on the number of antibodies (≥2 islet autoantibodies with normal glucose tolerance is now defined as stage 1 of T1D as the lifetime risk for developing clinical symptoms is nearly 100%) and metabolic factors (impaired oral glucose tolerance test). Many efforts are ongoing and underway to screen relatives of T1D patients and those in the general population for islet autoantibodies to identify individuals with

TABLE 401-4 APS-2 and Other Polyendocrine Disorder Associations

DISEASE	HLA ASSOCIATION	INITIATING FACTOR	MECHANISM	AUTOANTIGEN
Graves' disease	DR3	Iodine Anti-CD52	Antibody	TSH receptor
Myasthenia gravis	DR3, DR7	Thymoma Penicillamine	Antibody	Acetylcholine receptor
Anti-insulin receptor	?	SLE or other autoimmune disease	Antibody	Insulin receptor
Hypoparathyroidism	?	?	Antibody	Cell surface inhibitor
Insulin autoimmune syndrome	DR4, DRB1*0406	Methimazole Sulfhydryl-containing drugs	Antibody	Insulin
Celiac disease	DQ2/DQ8	Gluten diet	T cell	Transglutaminase
Type 1 diabetes	DR3/DR4 DQ2/DQ8	? Congenital rubella	T cell	Insulin, GAD65, IA-2, ZnT8, IGRP
Addison's disease	DR3/DR4 DRB1*0404	Unknown	T cell	21-Hydroxylase P450-5cc
Thyroiditis	DR3/DQB1*0201 DQA1*0301	Iodine Interferon α	T cell	Thyroglobulin Thyroid peroxidase
Pernicious anemia	?	?	T cell	Intrinsic factor H+/K+ ATPase
Vitiligo	?	Melanoma Antigen Immunization	?	Melanocyte
Chromosome dysgenesis–trisomy 21 and Turner's syndrome	DQA1*0301	?	?	Thyroid, islet, transglutaminase
Hypophysitis	?	Pit-1, TDRD6	?	Pituitary, Pit-1

Abbreviations: APS, autoimmune polyendocrine syndrome; SLE, systemic lupus erythematosus; TSH, thyroid-stimulating hormone.

preclinical disease who may elect to have treatment with teplizumab, anti-CD3 monoclonal antibody, to delay the clinical onset of diabetes.

Screening tests for thyroid disease can include anti–thyroid peroxidase (TPO) or anti-thyroglobulin autoantibodies or anti-TSH receptor antibodies for Graves' disease. Yearly measurements of TSH can then be used to follow these individuals. Celiac disease can be screened for using the anti–tissue transglutaminase (tTg) antibody test. For those <20 years of age, testing every 1–2 years should be performed, whereas less frequent testing is indicated after the age of 20 because the majority of individuals who develop celiac disease have the antibody earlier in life. Positive tTg antibody test results should be confirmed on repeat testing, followed by small-bowel biopsy to document pathologic changes of celiac disease. Many patients have asymptomatic celiac disease that is nevertheless associated with osteopenia and impaired growth. If left untreated, symptomatic celiac disease has been reported to be associated with an increased risk of gastrointestinal malignancy, especially lymphoma, and osteoporosis later in life.

The knowledge of the particular disease associations should guide other autoantibody or laboratory testing. A complete history and physical examination should be performed every 1–3 years including complete blood count, metabolic panel, TSH, and vitamin B_{12} levels to screen for most of the possible abnormalities. More specific tests should be based on specific findings from the history and physical examination.

TREATMENT

APS-2

With the exception of Graves' disease, the management of each endocrine component of APS-2 involves hormone replacement and is covered in detail in the chapters on adrenal **(Chap. 398)**, thyroid **(Chap. 394)**, gonadal **(Chaps. 403 and 404)**, and parathyroid diseases **(Chap. 422)**. As noted for APS-1, adrenal insufficiency can be masked by primary hypothyroidism and should be considered and treated as discussed above. In patients with T1D, decreasing insulin requirements or hypoglycemia, without obvious secondary causes, may indicate the emergence of adrenal insufficiency. Hypocalcemia in APS-2 patients is more likely due to malabsorption, potentially from undiagnosed celiac disease, than hypoparathyroidism.

Immunotherapy for autoimmune endocrine disease has been reserved for T1D, for the most part, reflecting the lifetime burden of the disease for the individual patient and society. Although several immunotherapies (e.g., modified anti-CD3, rituximab, abatacept, alefacept, low-dose antithymocyte globulin, TNF-α inhibitors, and JAK inhibitors) can prolong the honeymoon phase of T1D, none has achieved long-term success. Notably, the anti-CD3 monoclonal antibody (teplizumab) does delay the onset of clinical diabetes by an average of 3 years when administered to individuals with stage 2 T1D (e.g., those with autoantibodies and impaired glucose tolerance) and is now approved for clinical use in the United States. Active basic and clinical research using novel therapies and combinations may change the treatment landscape of this disease and other autoimmune conditions that share similar pathways.

■ IMMUNE CHECKPOINT INHIBITOR–INDUCED ENDOCRINE AUTOIMMUNITY

Therapies that block immune checkpoints, such as programmed cell death protein 1 (PD-1), its ligand (PD-L1), or CTLA-4, are beneficial immunotherapies for many advanced-stage cancers. These immune checkpoint inhibitors (ICIs) block negative immune regulation, thereby allowing for an immune response directed against tumor cells. However, immune-related adverse events also occur, especially autoimmunity directed toward self-tissues. ICI-induced T1D, thyroid disease, hypophysitis, and adrenal insufficiency have all been reported with these therapies and in combination. Hypothyroidism occurs in ~10% and T1D in 1–2% of those receiving monoclonal antibodies directed against PD-1 or PD-L1, and hypophysitis and adrenal insufficiency occur in <1% of treated patients. These autoimmune side effects can develop during or after therapy, mostly within a few weeks to months following the start of therapy. ICI-induced T1D has a very rapid onset, presents with diabetic ketoacidosis, is permanent, and requires lifelong exogenous insulin therapy for treatment. There is a strong genetic association, with HLA-DR4 being present in ~70% of patients, and islet autoantibodies may be present at diagnosis. The pathogenesis is immune mediated as T lymphocyte infiltration has been documented in the pancreatic islets of an ICI-T1D patient. Determining the mechanisms of autoimmune disease development following ICI therapies and developing biomarkers to stratify risk for autoimmune side effects prior to therapy are active areas of research.

■ IPEX

*I*mmune dysregulation, *p*olyendocrinopathy, *e*nteropathy, and *X*-linked disease (IPEX; OMIM 304790) is a rare X-linked recessive disorder. The disease onset is in infancy and is characterized by severe enteropathy, T1D, and skin disease, as well as variable association with several other autoimmune disorders. Many infants die within the first days of life, but the course is variable, with some children surviving for 12–15 years. Early onset of T1D, often at birth, is highly suggestive of the diagnosis because nearly 80% of IPEX patients develop T1D. Although treatment of the individual disorders can temporarily improve the situation, treatment of the underlying immune deficiency is required and includes immunosuppressive therapy generally followed by hematopoietic stem cell transplantation. Transplantation is the only life-saving form of therapy and can be fully curative by normalizing the imbalanced immune system found in this disorder.

IPEX is caused by mutations in the *FOXP3* gene, which is also mutated in the Scurfy mouse, an animal model that shares much of the same phenotype of IPEX patients. The FOXP3 transcription factor is expressed in regulatory T cells designated CD4+CD25+FOXP3+ (Treg). Lack of this factor causes a profound deficiency of this Treg population and results in rampant autoimmunity due to the lack of peripheral tolerance normally provided by these cells. Certain mutations may lead to varying forms of expression of the full syndrome, and there are rare cases where the *FOXP3* gene is intact but other genes involved in this pathway (e.g., CD25, IL-2Rα) may be causative. Future therapy with autologous CD4+ T cells transfected with a functioning *FOXP3* gene may offer a better long-term outcome than has been seen in those treated with stem cell transplantation.

■ THYMIC TUMORS

Thymomas and thymic hyperplasia are associated with several autoimmune diseases, with the most common being myasthenia gravis (44%) and red cell aplasia (20%). Graves' disease, T1D, and Addison's disease may also be associated with thymic tumors. Patients with myasthenia gravis and thymoma may have unique anti–acetylcholine receptor autoantibodies. Most thymomas lack AIRE expression within the thymoma, and this could be a potential factor in the development of autoimmunity. In support of this concept, thymoma is the one other disease with "frequent" development of anticytokine antibodies and mucocutaneous candidiasis in adults. The majority of tumors are malignant, and temporary remissions of the autoimmune condition can occur with resection of the tumor.

■ ANTI-INSULIN RECEPTOR ANTIBODIES

This is a very rare disorder where severe insulin resistance (type B) is caused by the presence of anti-insulin receptor antibodies. It is associated with acanthosis nigricans, which can also be associated with other forms of less severe insulin resistance. About one-third of patients have an associated autoimmune illness such as systemic lupus erythematosus or Sjögren's syndrome. Therefore, the presence of antinuclear antibodies, elevated erythrocyte sedimentation rate, hyperglobulinemia, leukopenia, and hypocomplementemia may accompany the presentation. The presence of anti-insulin receptor autoantibodies leads to marked insulin resistance, requiring >100,000 units of insulin to be given daily with only partial control of hyperglycemia. Patients can also have severe hypoglycemia due to partial activation of the insulin receptor by the antibody. The course of the disease is variable, and several patients have had spontaneous remissions. A therapeutic approach that targets B lymphocytes, including rituximab, cyclophosphamide, and pulse steroids, has been validated in follow-on case reports to induce remission of the disease.

INSULIN AUTOIMMUNE SYNDROME (HIRATA'S SYNDROME)

The insulin autoimmune syndrome, associated with Graves' disease and methimazole therapy (or other sulfhydryl-containing medications), is of particular interest due to a remarkably strong association with a specific HLA haplotype. Such patients with elevated titers of anti-insulin antibodies frequently present with hypoglycemia. In Japan, the disease is restricted to HLA-DR4-positive individuals with DRB1*04:06, while Caucasian patients predominantly have *DRB1*04:03* (which is related to *DRB1*04:06*). In Hirata's syndrome, the anti-insulin antibodies are often polyclonal. Discontinuation of the medication generally leads to resolution of the syndrome over time. There are very rare cases of insulin autoimmune syndrome not associated with sulfhydryl-containing medications that result in profound, life-threatening hypoglycemia. Treatment involves treating the underlying condition that causes anti-insulin antibodies, such as a B lymphocyte lymphoma (tend to have monoclonal insulin antibodies) or systemic lupus erythematosus. As hypoglycemia is profound when elevated titers of high-affinity insulin antibodies bind secreted insulin and then release it into circulation, treatment that begins with high-dose glucocorticoids and rituximab to target B lymphocytes has been shown to be effective.

POEMS SYNDROME

POEMS (*p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathy, *M*-protein, and *s*kin changes; also known as Crow-Fukase syndrome; OMIM 192240) patients usually present with a progressive sensorimotor polyneuropathy, diabetes mellitus (50%), primary gonadal failure (70%), and a plasma cell dyscrasia with sclerotic bony lesions. Associated findings can be hepatosplenomegaly, lymphadenopathy, and hyperpigmentation. Patients often present in the fifth to sixth decade of life and have a median survival after diagnosis of <3 years. The syndrome is assumed to be secondary to circulating immunoglobulins, but patients have excess vascular endothelial growth factor as well as elevated levels of other inflammatory cytokines such as IL-1β, IL-6, and tumor necrosis factor α. Patients have been treated with thalidomide, and more recently lenalidomide, leading to a decrease in vascular endothelial growth factor. Hyperglycemia responds to small, subcutaneous doses of insulin. The hypogonadism is due to primary gonadal disease with elevated plasma levels of follicle-stimulating hormone and luteinizing hormone. Temporary resolution of the features of POEMS, including normalization of blood glucose, may occur after radiotherapy for localized plasma cell lesions of bone or after chemotherapy, lenalidomide and dexamethasone, or autologous stem cell transplantation.

OTHER DISORDERS

Other diseases can exhibit polyendocrine deficiencies, including Kearns-Sayre syndrome, DIDMOAD syndrome (*d*iabetes *i*nsipidus, *d*iabetes *m*ellitus, progressive bilateral *o*ptic *a*trophy, and sensorineural *d*eafness; also termed Wolfram's syndrome), Down's syndrome or trisomy 21 (OMIM 190685), Turner's syndrome (monosomy X, 45,X0), and congenital rubella.

Kearns-Sayre syndrome (OMIM 530000) is a rare mitochondrial DNA disorder characterized by myopathic abnormalities leading to ophthalmoplegia and progressive weakness in association with several endocrine abnormalities, including hypoparathyroidism, primary gonadal failure, diabetes mellitus, and hypopituitarism. Crystalline mitochondrial inclusions are found in muscle biopsy specimens, and such inclusions have also been observed in the cerebellum. Antiparathyroid antibodies have not been described; however, antibodies to the anterior pituitary gland and striated muscle have been identified, and the disease may have autoimmune components. These mitochondrial DNA mutations occur sporadically and do not appear to be associated with a familial syndrome.

Wolfram's syndrome (OMIM 222300, chromosome 4; OMIM 598500, mitochondrial) is a rare autosomal recessive disease that is also called DIDMOAD. Neurologic and psychiatric disturbances are prominent in most patients and can cause severe disability. The disease is caused by defects in the *Wolfram syndrome 1* (*WFS1*) gene, which encodes a 100-kDa transmembrane protein that has been localized to the endoplasmic reticulum and is found in neuronal and neuroendocrine tissue. Its expression induces ion channel activity with a resultant increase in intracellular calcium and may play an important role in intracellular calcium homeostasis. Wolfram's syndrome appears to be a slowly progressive neurodegenerative process, and there is nonautoimmune selective destruction of the pancreatic beta cells. Diabetes mellitus with an onset in childhood is usually the first manifestation. Diabetes mellitus and optic atrophy are present in all reported cases, but expression of the other features is variable. Treatments targeting endoplasmic reticulum dysfunction are being tested and may be a bridge until gene therapy can be developed to treat the most severely affected cases.

Down's syndrome, or trisomy 21 (OMIM 190685), is associated with the development of T1D, thyroiditis, and celiac disease. Patients with Turner's syndrome also appear to be at increased risk for the development of thyroid disease and celiac disease. It is recommended to screen patients with trisomy 21 and Turner's syndrome for associated autoimmune diseases on a regular basis.

GLOBAL CONSIDERATIONS

Identification of these syndromes requires access to central laboratories with the ability to detect unique autoantibodies and to sequence the specific genes that may underlie these disorders. Early recognition of the clinical features of these disorders and timely referral and/or consultation with tertiary care centers to confirm the diagnosis and initiate therapy are important to improving outcomes. The *AIRE* recessive gene mutations found in APS-1 were originally described in high frequency in several populations including Finns, Iranian Jews, Sardinians, Norwegians, and Irish. Although individuals from many other countries have now been found to have these mutations and the newly identified dominant *AIRE* gene mutations, understanding the frequency in the background population may raise the clinician's level of suspicion for these rare disorders. Hirata's syndrome was originally reported in Japanese populations but also may be found in other populations, as noted.

FURTHER READING

Anderson MS, Su MA: AIRE expands: New roles in immune tolerance and beyond. Nat Rev Immunol 16:247, 2016.

Dispenzieri A: POEMS syndrome: 2021 update on diagnosis, risk-stratification, and management. Am J Hematol 96:872, 2021.

Husebye ES et al: Autoimmune polyendocrine syndromes. N Engl J Med 378:1132, 2018.

Oftedal BE et al: A partial form of AIRE deficiency underlies a mild form of autoimmune polyendocrine syndrome type 1. J Clin Invest 133:e169704, 2023.

Postow MA et al: Immune-related adverse events associated with immune checkpoint blockade. N Engl J Med 378:158, 2018.

Ramos EL et al: Teplizumab and β-cell function in newly diagnosed type 1 diabetes. N Engl J Med 389:2151, 2023.

Zhao Z et al: Autoimmune polyendocrine syndrome induced by immune checkpoint inhibitors: A systematic review. Cancer Immunol Immunother 70:1527, 2021.

Section 2 Sex- and Gender-Based Medicine

402 Sex Development

Courtney Finlayson, J. Larry Jameson, John C. Achermann

Sex development begins in utero but continues into young adulthood with the achievement of sexual maturity and reproductive capability. The major determinants of sex development can be divided into three components: chromosomal sex, gonadal sex (sex determination), and phenotypic sex (sex differentiation) **(Fig. 402-1)**. Variations at each of these stages can result in differences (or disorders) of sex development (DSDs) **(Table 402-1)**. In the newborn period, ~1 in 5000 babies undergo investigation because of atypical or ambiguous genitalia. Urgent assessment is indicated, because some causes such as congenital adrenal hyperplasia (CAH) can be associated with life-threatening adrenal crises. An experienced multidisciplinary team is important for counseling, planning appropriate investigations, discussing long-term well-being, supporting parents, and providing clear communication about the diagnosis and management options. DSDs can also present at other ages and to a range of health professionals, including with either absent puberty, primary amenorrhea, or androgenization in the teen years **(Table 402-2)**. Other forms of gonadal dysfunction (e.g., Klinefelter syndrome [KS], Turner syndrome [TS]) often are diagnosed later in life by internists. Because DSDs are associated with a variety of psychological, reproductive, and potential medical consequences, an open dialogue must be established between the patient and health care providers to ensure continuity and attention to these issues across the life span. Gender variance and gender dysphoria are more common among some individuals with DSD than in the general population, though not high. Thus, attention to and comfort discussing gender identity is important. Support groups also have a valuable role to play for many patients and families.

Care of individuals with DSDs should be holistic, often involving medical, psychosocial, and urogynecologic expertise, while acknowledging that the best way to care for individuals with DSD is not always clear and should be individualized. There are many controversies, particularly concerning whether genitoplasty or prophylactic gonadectomy in selected conditions should be performed for infants and young children prior to the age of consent. Accepted nomenclature is also controversial. Previous terms such as *intersex* and *hermaphrodite* were changed by the 2006 Consensus Statement to *disorder of sex development* and *ovotesticular DSD*. The term "disorder" is often considered negative and stigmatizing and thus has shifted toward *difference of sex development*, but no term is universally accepted.

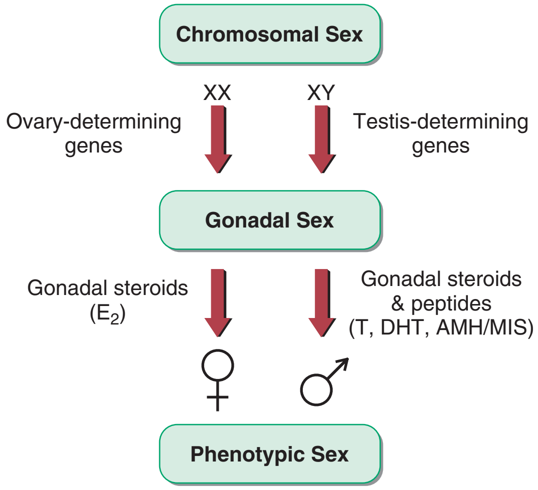

FIGURE 402-1 Sex development can be divided into three major components: chromosomal sex, gonadal sex, and phenotypic sex. AMH, anti-müllerian hormone also known as Müllerian-inhibiting substance, MIS; DHT, dihydrotestosterone; T, testosterone.

SEX DEVELOPMENT

Chromosomal sex, defined by a karyotype, describes the X and/or Y chromosome complement (46,XY; 46,XX) established at the time of fertilization. The presence of a normal Y chromosome determines that testis development will occur even in the presence of multiple X chromosomes (e.g., 47,XXY). Loss of an X chromosome impairs gonad development (45,X or 45,X/46,XY mosaicism). Fetuses with no X chromosome (45,Y) are not viable.

Gonadal sex refers to the histologic and functional characteristics of gonadal tissue as testis or ovary. The embryonic gonad is initially "bipotential" and can develop into either a testis or an ovary **(Fig. 402-2)**. Testis development is initiated by expression of the gene *SRY* (sex-determining region on the Y chromosome) (from ~42 days after conception). Disruption of *SRY* prevents testis development in 46,XY individuals, whereas translocation of *SRY* in 46,XX individuals induces testis development and a male phenotype. SRY regulates *SOX9* (*SRY*-related HMG-box gene 9), leading to expression of a cascade of genes involved in testis development, including in Sertoli cell maturation and Leydig cell differentiation/steroidogenesis. Disruption of some of these genes can influence both the development of the testis and other organs, such as kidney (*WT1*), adrenal/spleen (*SF1*, *NR5A1*), brain (*PPP1R12A*), or heart (*GATA4*). Chromosomal segment duplications (e.g., Xp21 containing *DAX1*/*NR0B1*) can also impair testis development, revealing the sensitivity of testis-determining pathway to gene dosage effects.

Ovarian development is not a "passive" process. Many specific genes are expressed during early ovary development, some of which may repress testis development (e.g., WNT4, R-spondin-1) (Fig. 402-2). Once the ovary has formed, additional factors are required for normal follicular development (e.g., follicle-stimulating hormone [FSH] receptor). Steroidogenesis in the ovary requires the development of follicles that contain granulosa cells and theca cells surrounding the oocytes **(Chap. 404)**. Thus, there is relatively limited ovarian steroidogenesis until puberty.

Germ cells also develop in a sex dimorphic manner. In the developing ovary, primordial germ cells (PGCs) show marked proliferation and enter meiosis, whereas they undergo mitotic arrest in the developing testis. Approximately 7 million germ cells are present in the fetal ovary toward the end of the second trimester, and 1 million remain at birth. Only 400 are ovulated during a woman's reproductive life span **(Chap. 404)**.

Phenotypic sex refers to the structures of the external and internal genitalia and secondary sex characteristics. In early gestation, internal and external genitalia are initially similar in both sexes ("indifferent"). Sex-specific development occurs as a result of hormone action **(Fig. 402-3)**. The developing testis releases anti-müllerian hormone (AMH; also known as müllerian-inhibiting substance [MIS]) from Sertoli cells and testosterone from Leydig cells. AMH acts through specific receptors to cause regression of the müllerian structures from 60–80 days after conception. At ~60–140 days after conception, testosterone supports the maintenance of wolffian structures, including the epididymides, vasa deferentia, and seminal vesicles. Testosterone is the precursor for dihydrotestosterone (DHT), a potent androgen that promotes development of the external genitalia, including the penis and scrotum (60–100 days, and thereafter) (Fig. 402-3). The urogenital sinus develops into the prostate and prostatic urethra in the male and into the urethra and lower portion of the vagina in the female. The genital tubercle becomes the glans penis in the male and the clitoris in the female. The urogenital swellings form the scrotum or the labia majora, and the urethral folds fuse to form the shaft of the penis and the male urethra or the labia minora. In the female, wolffian ducts regress and the müllerian ducts form the fallopian tubes, uterus, and upper segment of the vagina. A female phenotype will develop in the

TABLE 402-1 Classification of Differences (Disorders) of Sex Development (DSDs)

SEX CHROMOSOME DSD	46,XY DSD (SEE TABLE 402-3)	46,XX DSD (SEE TABLE 402-4)
47,XXY (Klinefelter syndrome and variants) 45,X (Turner syndrome and variants) 45,X/46,XY mosaicism (mixed gonadal dysgenesis) 46,XX/46,XY (chimerism/mosaicism)	**Gonadal (testis) development** Complete or partial gonadal dysgenesis Impaired fetal Leydig cell function Ovotesticular DSD Testis regression **Disruption in androgen synthesis or action** Disruption of androgen biosynthesis LH receptor (*LHCGR*) Smith-Lemli-Opitz syndrome (*DHCR7*) Steroidogenic acute regulatory (*StAR*) protein Cholesterol side-chain cleavage (*CYP11A1*) 3β-Hydroxysteroid dehydrogenase II (*HSD3B2*) 17α-Hydroxylase/17,20-lyase (*CYP17A1*) P450 oxidoreductase (*POR*) Cytochrome b5 (*CYB5A*) 17β-Hydroxysteroid dehydrogenase III (*HSD17B3*) 5α-Reductase II (*SRD5A2*) Aldo-keto reductase 1C2 (*AKR1C2*) Disruption of androgen action Androgen insensitivity syndrome **Others include:** Syndromic associations of male genital development Associated with fetal growth restriction Persistent müllerian duct syndrome Vanishing testis syndrome Isolated hypospadias Congenital hypogonadotropic hypogonadism Cryptorchidism Environmental influences	**Gonadal (ovary) development** Gonadal dysgenesis Ovotesticular DSD Testicular DSD **Androgen excess** Fetal 3β-Hydroxysteroid dehydrogenase II (*HSD3B2*) 21-Hydroxylase (*CYP21A2*) P450 oxidoreductase (*POR*) 11β-Hydroxylase (*CYP11B1*) Fetoplacental Aromatase deficiency (*CYP19A1*) Oxidoreductase deficiency (*POR*) Maternal Maternal virilizing tumors (e.g., luteomas) Androgenic drugs **Others include:** Syndromic associations (e.g., cloacal anomalies) Müllerian agenesis/hypoplasia (e.g., MRKH) Uterine abnormalities (e.g., MODY5) Vaginal atresia (e.g., McKusick-Kaufman) Labial adhesions

Abbreviations: LH, luteinizing hormone; MODY, maturity-onset diabetes of the young; MRKH, Mayer-Rokitansky-Küster-Hauser syndrome.

Source: Reproduced with permission from IA Hughes et al: Consensus statement on management of intersex disorders. J Pediatr Urol 2:148, 2006.

absence of the gonad, but estrogen is needed for maturation of the uterus and breast at puberty.

The prenatal hormone environment is likely one of many factors influencing aspects of gender identity and behavior. This is an area of ongoing research and is beyond the scope of this chapter.

DIFFERENCES OF CHROMOSOMAL SEX

Variations in sex chromosome number and structure can present as DSDs (e.g., 45,X/46,XY). KS (47,XXY) and TS (45,X) do not usually present with genital ambiguity but are associated with gonadal dysfunction **(Table 402-3)**.

KLINEFELTER SYNDROME (47,XXY) (SEE ALSO CHAP. 403)

Pathophysiology The classic form of KS (47,XXY) occurs after meiotic nondisjunction of the sex chromosomes during gametogenesis (40% during spermatogenesis, 60% during oogenesis). Other forms of aneuploidy have similar features to KS (including mosaic 46,XY/47,XXY, 48,XXYY, and 48,XXXY) but are less common. KS has an incidence of at least 1 in 1000 men, but ~75% of cases are not diagnosed. Of those diagnosed, historically only 10% were identified prepubertally. However, noninvasive prenatal testing is leading to increased early detection.

TABLE 402-2 Presentation of Differences of Sex Development (DSD) at Different Stages of Life

PRESENTATION	FEATURES	PROFESSIONAL	EXAMPLES
Prenatal	Karyotype-phenotype discordance	Obstetrician; fetal medicine	Many
Neonatal	Atypical genitalia	Obstetrician; neonatal medicine	Many
	Salt-losing crisis	Pediatrician	CAH *(CYP21A2)*
Childhood	Hernia	Surgeon	CAIS
	Androgenization	Endocrinologist	CAH *(CYP21A2, CYP11B1)*
	Poor growth	Pediatrician	Turner, 45,X/46,XY
	Associated features	Oncologist/nephrologist	Wilms' tumor
Puberty	Androgenization	Endocrinologist	17β-HSD, 5α-reductase, SF1
	Estrogenization	Endocrinologist	Ovotestis
	Absent puberty	Endocrinologist	Gonadal dysgenesis, CAH *(CYP17A1)*, Turner
Post-puberty	Amenorrhea	Gynecologist	CAIS
Adult	Infertility	Andrologist	Klinefelter, 45,X/46,XY, SF1

Abbreviations: CAH, congenital adrenal hyperplasia; CAIS, complete androgen insensitivity syndrome; 17β-HSD, 17β-hydroxysteroid dehydrogenase deficiency; SF1, steroidogenic factor 1 (*NR5A1*).

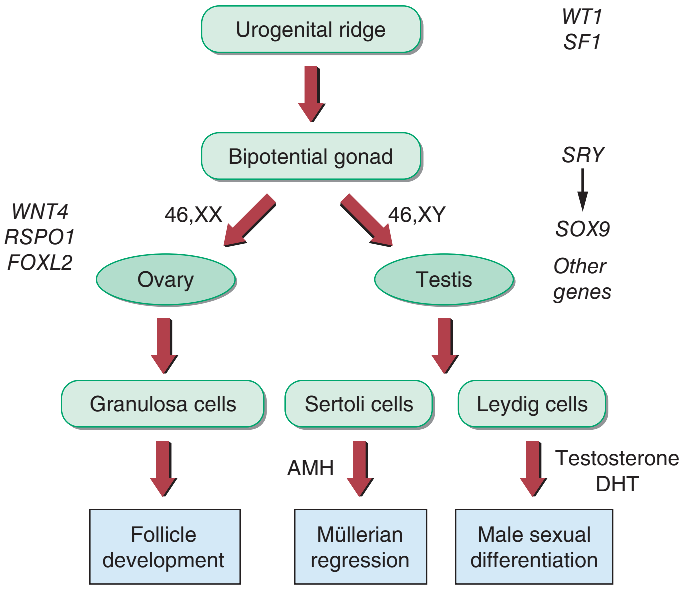
Urogenital ridge
WT1
SF1
Bipotential gonad
SRY
SOX9
Other genes
WNT4
RSPO1
FOXL2
46,XX
46,XY
Ovary
Testis
Granulosa cells
Sertoli cells
Leydig cells
AMH
Testosterone
DHT
Follicle development
Müllerian regression
Male sexual differentiation

FIGURE 402-2 The genetic regulation of gonadal development. See text for additional genes involved. AMH, anti-müllerian hormone (müllerian-inhibiting substance); DHT, dihydrotestosterone; *FOXL2*, forkhead transcription factor L2; *RSPO1*, R-spondin 1; *SF1*, steroidogenic factor 1 (also known as NR5A1); *SOX9*, *SRY*-related HMG-box gene 9; *SRY*, sex-determining region on the Y chromosome; *WNT4*, wingless-type MMTV integration site 4; *WT1*, Wilms' tumor–related gene 1.

Clinical Features KS is most commonly characterized by small testes, infertility, gynecomastia, tall stature/increased leg length, and hypogonadism in phenotypic males. At birth, most infants with KS do not have clinical features, although there are higher rates of cryptorchidism and hypospadias. Most patients present in puberty with arrested pubertal development caused by testicular insufficiency. Others are diagnosed after puberty, based on low androgens, gynecomastia, or infertility. Testes are small and firm (median length 2.5 cm [4 mL volume]; almost always <3.5 cm [12 mL]) and typically seem inappropriately small for the degree of androgenization. Biopsies are not usually necessary but typically reveal seminiferous tubule hyalinization and azoospermia. Other clinical features of KS are listed in Table 402-3. Plasma concentrations of FSH and luteinizing hormone (LH) are increased in most adults with 47,XXY, and plasma testosterone is decreased (50–75%), reflecting primary gonadal insufficiency. Estradiol is often increased, resulting in gynecomastia **(Chap. 403)**. Those with mosaic forms of KS have less severe clinical features, have larger testes, and sometimes achieve spontaneous fertility.

TREATMENT

Klinefelter Syndrome

Growth, endocrine function, and bone mineralization should be monitored, especially from adolescence. Educational and psychological support is important for many individuals with KS.

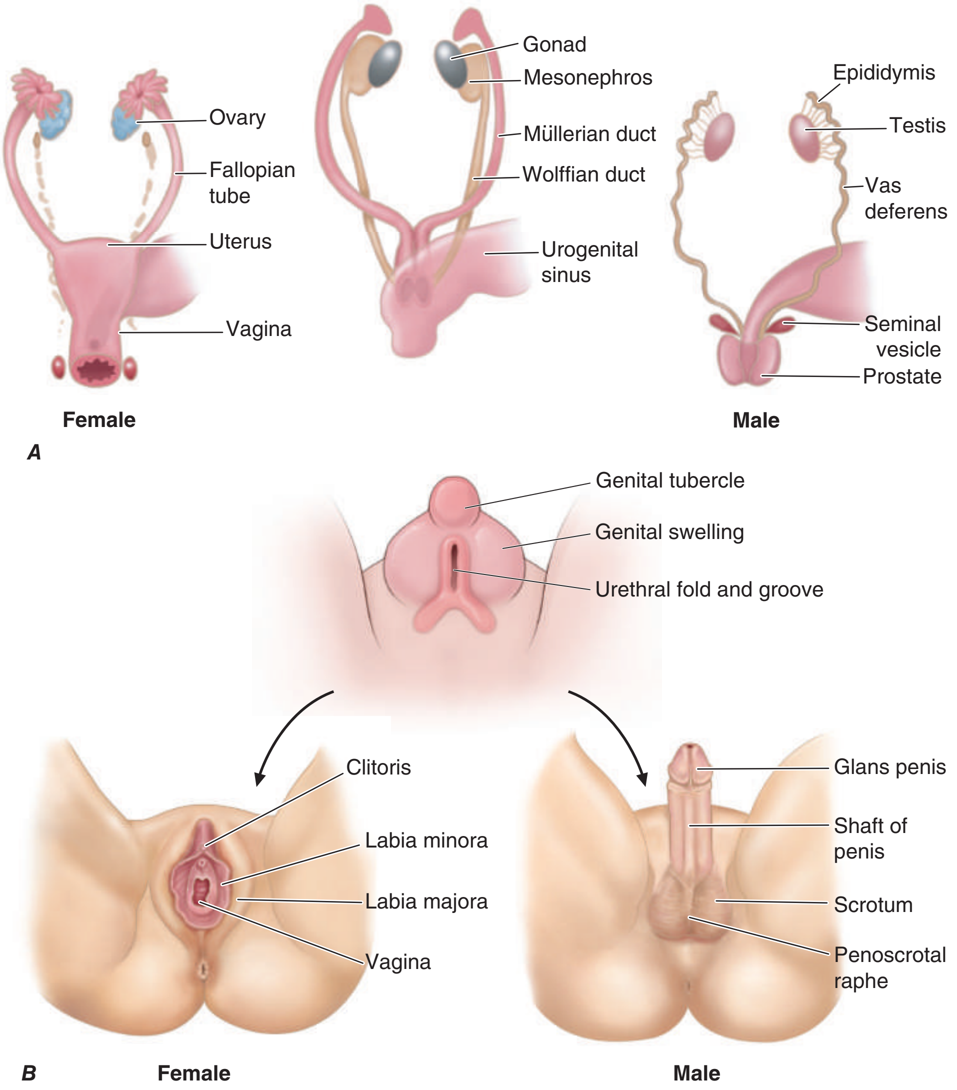
Gonad
Mesonephros
Müllerian duct
Wolffian duct
Urogenital sinus
Ovary
Fallopian tube
Uterus
Vagina
Epididymis
Testis
Vas deferens
Seminal vesicle
Prostate
Female
Male
A
Genital tubercle
Genital swelling
Urethral fold and groove
Clitoris
Labia minora
Labia majora
Vagina
Glans penis
Shaft of penis
Scrotum
Penoscrotal raphe
B
Female
Male

FIGURE 402-3 Sex development. ***A.*** Internal urogenital tract. ***B.*** External genitalia.

TABLE 402-3 Possible Associated Clinical Features of Sex Chromosome Variations

			GENITALIA		
DISORDER	**COMMON CHROMOSOMAL COMPLEMENT**	**GONAD**	**EXTERNAL**	**INTERNAL**	**BREAST DEVELOPMENT**
Klinefelter syndrome	47,XXY or 46,XY/47,XXY	Hyalinized testes	Male	Male	Increased incidence of gynecomastia
	Clinical Features				
	Small testes, azoospermia, decreased facial and axillary hair, decreased libido, tall stature and increased leg length, decreased penile length, increased risk of breast tumors, thromboembolic disease, learning difficulties, anxiety, speech delay and decreased verbal IQ, obesity, diabetes mellitus, metabolic syndrome, varicose veins, hypothyroidism, systemic lupus erythematosus, epilepsy				
Turner syndrome	45,X or 45,X/46,XX	Streak gonad or immature ovary	Female	Hypoplastic female	Immature female
	Clinical Features				
	Infancy: lymphedema, web neck, shield chest, low-set hairline, cardiac defects and coarctation of the aorta, urinary tract malformations, and horseshoe kidney				
	Childhood: short stature, cubitus valgus, short neck, short fourth metacarpals, hypoplastic nails, micrognathia, scoliosis, otitis media and sensorineural hearing loss, ptosis and amblyopia, multiple nevi and keloid formation, autoimmune thyroid disease, visuospatial learning difficulties				
	Adulthood: absent puberty and primary amenorrhea, hypertension, obesity, dyslipidemia, impaired glucose tolerance and insulin resistance, autoimmune thyroid disease, cardiovascular disease, aortic root dilation, osteoporosis, inflammatory bowel disease, chronic hepatic dysfunction, increased risk of colon cancer, hearing loss				
45,X/46,XY mosaicism	45,X/46,XY	Testis or streak gonad	Variable	Variable	Usually male
	Clinical Features				
	Short stature, increased risk of gonadal tumors, some Turner syndrome features				
Ovotesticular DSD	46,XX/46,XY	Testis and ovary or ovotestis	Variable	Variable	Gynecomastia
	Clinical Features				
	Possible increased risk of gonadal tumors				

Abbreviation: DSD, difference of sex development.

Androgen supplementation improves virilization, libido, energy, hypofibrinolysis, and bone mineralization in men with low testosterone levels but may occasionally worsen gynecomastia **(Chap. 403)**. Gynecomastia can be treated by surgical reduction if it causes concern **(Chap. 403)**. Fertility has been achieved by using in vitro fertilization in men with oligospermia or with intracytoplasmic sperm injection (ICSI) after retrieval of spermatozoa by testicular sperm extraction techniques **(Chap. 403)**. Long-term monitoring of men with KS is important given the increased risk of breast cancer, cardiovascular disease, metabolic syndrome, osteoporosis, and autoimmune disorders. Because most men with KS are never diagnosed, it is important that all internists consider this diagnosis in men with such features.

TURNER SYNDROME (GONADAL DYSGENESIS; 45,X)

Pathophysiology TS is caused by complete or partial loss of one X chromosome and affects ~1 in 2500 women. Approximately one-half of women with TS have a 45,X karyotype, ~20% have 45,X/46,XX mosaicism, and the remainder have structural abnormalities of the X chromosome, or Y chromosome material. The clinical features of TS likely result at least in part from haploinsufficiency of X chromosomal genes (especially in the pseudoautosomal region), but the exact mechanisms remain poorly understood.

Clinical Features TS is characterized by female-typical external genitalia, short stature, hypergonadotropic hypogonadism, infertility, and other phenotypic features (Table 402-3). Infants may present with lymphedema, nuchal folds, low hairline, or left-sided cardiac defects or later in childhood with unexplained growth failure or delayed puberty. Although limited spontaneous pubertal development occurs in up to 30% of girls with TS (10%, 45,X; 60%, 45,X/46,XX) and up to 20% have menarche, the vast majority of women with TS develop complete ovarian insufficiency. Therefore, this diagnosis should be considered in all women who present with primary or secondary amenorrhea and elevated gonadotropin levels.

TREATMENT

Turner Syndrome

The management of girls and women with TS requires a multidisciplinary approach to address many potentially affected organ systems according to TS practice guidelines. Individuals require long-term monitoring by an experienced cardiologist to follow congenital heart defects (CHDs) (30%) (bicuspid aortic valve, 30–50%; coarctation of the aorta, 30%; aortic root dilation, 5%), antibiotic prophylaxis for dental or surgical procedures, and serial magnetic resonance imaging (MRI) of aortic root dimensions, as progressive aortic root dilation is associated with increased risk of aortic dissection. Individuals found to have congenital renal and urinary tract malformations (30%) are at risk for urinary tract infections, hypertension, and nephrocalcinosis. Hypertension can occur independently of cardiac and renal malformations and should be monitored and treated as in other patients with essential hypertension. Regular assessment of thyroid function, weight, dentition, hearing, speech, vision, and educational issues should be performed during childhood. Counseling about long-term growth and fertility issues should be provided. Patient support groups are active throughout the world and can play an invaluable role.

Short stature is common, and untreated final height rarely exceeds 150 cm in nonmosaic 45,X TS. Recombinant growth hormone is used to increase adult height. Girls with evidence of ovarian insufficiency require estrogen replacement to induce breast and uterine development, support growth, and maintain bone mineralization. Most physicians now initiate low-dose estrogen therapy to induce puberty at an age-appropriate time (~11 years). Doses of estrogen are increased gradually to allow development over a 2- to 4-year period. Progestins are added later to regulate withdrawal bleeds. A very small percentage of women with TS have had spontaneous pregnancy, whereas others have achieved successful pregnancy after ovum donation and in vitro fertilization, but the risks of cardiac complications are high, and expert counseling and management are needed. The existence of Y chromosome material in individuals with TS increases the risk for germ cell tumors, and gonadectomy has traditionally been advised. Increasingly, it is recommended that this decision also weigh potential for gonadal function and the importance of autonomous decision-making. Long-term follow-up of women with TS includes careful surveillance of sex hormone replacement and reproductive function, bone mineralization, cardiac function and aortic root dimensions, blood pressure, weight and glucose tolerance, hepatic and lipid profiles, thyroid function, celiac disease screening, skin examination, and hearing. This service is provided by a dedicated TS clinic in some centers.

45,X/46,XY MOSAICISM

The phenotype of individuals with 45,X/46,XY mosaicism (sometimes called *mixed gonadal dysgenesis*) can vary considerably. Some have a predominantly female phenotype (see TS above). Most 45,X/46,XY individuals have a male phenotype and testes, and the diagnosis is made incidentally after amniocentesis or during investigation of infertility. In practice, most newborns referred for assessment have atypical genitalia and variable somatic features. There is often marked asymmetry, with a streak gonad and hemiuterus on one side and a partially descended dysgenetic testis and hemiscrotum on the other side. Many children are raised as boys, but in some children, sex designation (whether to raise the baby as male or female) must be decided by parents and the multidisciplinary team. There is an increased risk of germ cell cancer (GCC), up to 35% in intraabdominal gonads, so prophylactic removal of intraabdominal gonads is usually considered. Individuals raised as males often have hypospadias surgery and removal of dysgenetic or streak gonads if the gonads cannot be brought down into the scrotum. Scrotal testes can be preserved but require regular examination for tumor development and sonography (and possibly biopsy) at the time of puberty. Testosterone supplementation may be required in puberty or adulthood if low testosterone is detected. Potential associated features (e.g., cardiac, renal) should be monitored according to TS guidelines. Infertility is typical, but non-azoospermia or focal spermatogenesis has been reported, highlighting the importance of individualized approaches to management.

OVOTESTICULAR DSD

Ovotesticular DSD (OTDSD) is a condition in which an individual has both ovarian and testicular tissue, either by having both an ovary and a testis or by having an ovotestis. Most individuals with this diagnosis have a 46,XX karyotype (especially in individuals of African ancestry), although 46,XX/46,XY chimerism and very rarely a 46,XY karyotype is also possible. OTDSD usually presents with atypical genitalia at birth and sometimes breast development, cyclical hematuria, and/or phallic development at puberty. Progressive regression of the ovarian and/or testicular component can occur over time. Gender identity varies in OTDSD but often aligns with assigned sex. Risk of GCC is also elevated in OTDSD (~3%). Infertility is typical (especially in 46,XX testes with no Y chromosome), but births have occurred via ovum or sperm from individuals with other forms of OTDSD.

DIFFERENCES OF GONADAL AND PHENOTYPIC SEX

Differences of gonadal and phenotypic sex can result in reduced androgen production or action in individuals with a 46,XY karyotype (46,XY DSD) or excess androgen production in individuals with a 46,XX karyotype (46,XX DSD) (Table 402-1). These conditions cover a spectrum of phenotypes ranging from phenotypic females with a Y chromosome to phenotypic males with a 46,XX karyotype to individuals with atypical genitalia. Karyotype is a useful starting investigation for diagnosis with basic biochemical profiling. High-throughput genetic testing can help in reaching a definitive diagnosis.

46,XY DSD

Underandrogenization of the 46,XY fetus reflects defects in androgen production or action. It can result from disruption of testis development, defects of androgen synthesis, or resistance to testosterone and DHT (Table 402-1).

Disruption of Testis Development • **TESTICULAR DYSGENESIS** *Complete gonadal dysgenesis* (CGD, *Swyer's syndrome*) is associated with streak gonads, müllerian structures (due to insufficient AMH/MIS secretion), and a complete absence of androgenization. Phenotypic females with this condition usually present because of absent pubertal development and are found to have a 46,XY karyotype. Serum sex steroids, AMH/MIS, and inhibin B are low, and LH and FSH are elevated. Individuals with CGD typically identify as female. The risk of GCC is high, and intraabdominal gonads should be removed. In contrast, patients with *partial gonadal dysgenesis* (PGD, *dysgenetic testes*) may produce enough MIS to regress the uterus and sufficient testosterone for partial androgenization and, therefore, usually present in the newborn period with atypical genitalia, highlighting the spectrum of features that are typically seen with many DSDs.

Testicular dysgenesis can result from disruption of testis-promoting genes (e.g., *WT1*, *SF1*, *SRY*, *SOX9*, *MAP3K1*, *DHH*, *DHX37*, and others) or, rarely, duplication of chromosomal loci containing "antitestis" genes (e.g., *DAX1*) **(Table 402-4)**. Among these, deletions or mutations of *SRY* and heterozygous mutations of *SF1* (*NR5A1*) or *DHX37* appear to be most common but still account collectively for <30% of cases. Associated clinical features may be present (Table 402-4). A family history of DSD, hypospadias, infertility, or early menopause is important because variations in some genes (e.g., *SF1/NR5A1*, *SOX8*) can be associated with a range of reproductive phenotypes. *SF1* variants are sometimes inherited from a mother in a sex-limited dominant manner (which can mimic X-linked inheritance), and a woman may later develop primary ovarian insufficiency because of the effect of *SF1* on the ovary. Gender identity can be variable in PGD. Dysgenetic testes have an increased risk of GCC. For descended testes, monitoring via physical examination is appropriate. If testes are intraabdominal and not able to be brought down, they may be removed to prevent GCC (risk up to 35% if intraabdominal). Dysgenetic testes may or may not produce sufficient testosterone for puberty. In those who identify as male, testosterone replacement may be needed. In those who identify as female, estrogen replacement will be needed for female-typical pubertal development and ongoing sex steroid requirements. *Absent (vanishing) testis syndrome* (*bilateral anorchia*) reflects regression of the testis during development. The absence of müllerian structures indicates adequate secretion of AMH early in utero. Usually, androgenization of the external genitalia is normal. The etiology is often unknown but sometimes associated with pathogenic variants in *DHX37*. These individuals can be offered testicular prostheses and androgen replacement in adolescence and typically identify as male.

Disruption of Androgen Synthesis Defects in the pathway that regulates androgen synthesis **(Fig. 402-4)** cause underandrogenization of the 46,XY fetus (Table 402-1). Müllerian regression is unaffected because Sertoli cell function is preserved, and no uterus is found. These conditions can present with a spectrum of genital appearances, ranging from female-typical external genitalia or clitoromegaly in some individuals to penoscrotal hypospadias or a small phallus in others.

LH RECEPTOR Pathogenic variants in the LH receptor (LHCGR) cause Leydig cell hypoplasia and testosterone insufficiency due to LH resistance.

STEROIDOGENIC ENZYME PATHWAYS Mutations in *steroidogenic acute regulatory protein* (*StAR*) and *CYP11A1* affect both adrenal and gonadal steroidogenesis (Fig. 402-4) **(Chap. 398)**. Affected individuals (46,XY) usually have severe early-onset salt-losing adrenal failure and a female phenotype, although later-onset milder variants are increasingly reported. Defects in *3β-hydroxysteroid dehydrogenase type 2* (*HSD3B2*) also cause adrenal insufficiency in severe cases, but the accumulation of dehydroepiandrosterone (DHEA) has a mild androgenizing effect, resulting in atypical genitalia or hypospadias. Salt loss occurs in many but not all children. Patients with CAH due to *17α-hydroxylase (CYP17A1) deficiency* have variable underandrogenization and develop hypertension and hypokalemia due to the potent salt-retaining effects

TABLE 402-4 Selected Genetic Causes of 46,XY Differences of Sex Development (DSDs)

GENE	INHERITANCE	GONAD	UTERUS	EXTERNAL GENITALIA	ASSOCIATED FEATURES
Disruption of Testis Development					
WT1	AD	Dysgenetic testis	+/–	Female or ambiguous	Wilms' tumor, renal abnormalities, gonadal tumors (WAGR, Denys-Drash and Frasier syndromes)
SF1/NR5A1	AR/AD (SL)	Dysgenetic testis/ Leydig dysfunction	+/–	Female, ambiguous or male	Primary adrenal failure (rare); hyposplenia (common); primary ovarian insufficiency in female (46,XX) relatives
SRY	Y	Dysgenetic testis or ovotestis	+/–	Female or ambiguous	
SOX9	AD	Dysgenetic testis or ovotestis	+/–	Female or ambiguous	Campomelic dysplasia
DHX37	AD	Dysgenetic testis	+/–	Female, ambiguous or male	Testicular regression syndrome
Other causes of testicular dysgenesis include: *DMRT1*, *CBX2*, *MAP3K1*, *SOX8*, *ZNRF3*, *GATA4*/*ZFPM2* (congenital heart disease), *DHH* (neuropathy), *ARX* (X-linked lissencephaly), *TSPYL1* (sudden infant death), *MYRF* (diaphragmatic hernia), *ESR2/NR3A2*, *SAMD9* (MIRAGE syndrome), *ATRX* (blood), *PPP1R12A* (brain, gastrointestinal), *MAMLD1*, dupXp21, dup1p35, del9p24, del10q23, and in several other congenital syndromes					
Disruption of Androgen Synthesis					
LHCGR	AR	Testis	–	Female, ambiguous or micropenis	Leydig cell hypoplasia
DHCR7	AR	Testis	–	Variable	Smith-Lemli-Opitz syndrome: coarse facies, second-third toe syndactyly, failure to thrive, developmental delay, cardiac and visceral abnormalities
STAR	AR	Testis	–	Female or ambiguous	Congenital lipoid adrenal hyperplasia (primary adrenal insufficiency)
CYP11A1	AR	Testis	–	Female or ambiguous	Primary adrenal insufficiency
HSD3B2	AR	Testis	–	Ambiguous	CAH, primary adrenal insufficiency ± salt loss, partial androgenization due to ↑ DHEA
CYP17A1	AR	Testis	–	Female or ambiguous	CAH, hypertension due to ↑ corticosterone and 11-deoxycorticosterone, except in isolated 17,20-lyase deficiency
CYB5A	AR	Testis	–	Ambiguous	Apparent isolated 17,20-lyase deficiency; methemoglobinemia
POR	AR	Testis	–	Ambiguous or male	Mixed features of 21-hydroxylase deficiency and 17α-hydroxylase/17,20-lyase deficiency, sometimes associated with Antley-Bixler craniosynostosis
HSD17B3	AR	Testis	–	Female or ambiguous	Partial androgenization at puberty, ↑ androstenedione-to-testosterone ratio
SRD5A2	AR	Testis	–	Ambiguous or micropenis	Partial androgenization at puberty, ↑ testosterone-to-dihydrotestosterone ratio
AKR1C2 (*AKR1C4*)	AR	Testis	–	Female or ambiguous	Decreased fetal DHT production
Disruption of Androgen Action					
Androgen receptor	X	Testis	–	Female, ambiguous, micropenis or normal male	Phenotypic spectrum from complete androgen insensitivity syndrome (female external genitalia) and partial androgen insensitivity (ambiguous) to normal male genitalia and infertility

Abbreviations: AD, autosomal dominant; *AKR1C2*, aldo-keto reductase family 1 member 2; AR, autosomal recessive; *ARX*, aristaless related homeobox, X-linked; CAH, congenital adrenal hyperplasia; *CBX2*, chromobox homologue 2; *CYB5A*, cytochrome b5; *CYP11A1*, P450 cholesterol side-chain cleavage; *CYP17A1*, cytochrome P450 family 17 subfamily A member 1; *DAX1*, dosage sensitive sex-reversal, adrenal hypoplasia congenita on the X chromosome, gene 1; DHEA, dehydroepiandrosterone; *DHCR7*, sterol 7 δ reductase; *DHH*, desert hedgehog; *DMRT1*,doublesex and mab3-related transcription factor 1; *GATA4*, GATA binding protein 4; *HSD17B3*, 17β-hydroxysteroid dehydrogenase type 3; *HSD3B2*, 3β-hydroxysteroid dehydrogenase type 2; *LHR*, LH receptor; *MAP3K1*, mitogen-activated protein kinase kinase kinase 1; MIRAGE, myelodysplasia, infection, restriction of growth, adrenal hypoplasia, genital phenotypes, and enteropathy; *MYRF*, myelin regulatory factor; POR, P450 oxidoreductase; *SF1*, steroidogenic factor 1; SL, sex-limited; *SOX8*, *SRY*-related HMG-box gene 8; *SOX9*, *SRY*-related HMG-box gene 9; *SRD5A2*, 5α-reductase type 2; *SRY*, sex-related gene on the Y chromosome; *StAR*, steroidogenic acute regulatory protein; *TSPYL1*, testis-specific Y-encoded-like protein 1; WAGR, Wilms' tumor, aniridia, genitourinary anomalies, and mental retardation; *WNT4*, wingless-type mouse mammary tumor virus integration site, 4; *WT1*, Wilms' tumor–related gene 1; *ZFPM2*, zinc finger protein, multitype 2; *ZNRF3*, zinc and ring finger 3.

of corticosterone and 11-deoxycorticosterone. Patients with complete loss of 17α-hydroxylase function often present as phenotypic females who do not enter puberty and are found to have inguinal testes and hypertension in adolescence. Some mutations in *CYP17A1* selectively impair 17,20-lyase activity without altering 17α-hydroxylase activity, leading to underandrogenization without mineralocorticoid excess and hypertension. Disruption of the coenzyme, *cytochrome b5* (*CYB5A*), can present similarly, and methemoglobinemia is usually present. Mutations in *P450 oxidoreductase* (*POR*) affect multiple steroidogenic enzymes, leading to reduced androgen production and a biochemical pattern of apparent combined 21-hydroxylase and 17α-hydroxylase deficiency, sometimes with skeletal abnormalities (Antley-Bixler craniosynostosis). Defects in *17β-hydroxysteroid dehydrogenase type 3* (*HSD17B3*) and *5α-reductase type 2* (*SRD5A2*) interfere with the synthesis of testosterone and DHT, respectively. These conditions are characterized by minimal or absent androgenization in utero, but some phallic development can occur during adolescence due to the action of other enzyme isoforms. Individuals with 5α-reductase type 2 deficiency have normal wolffian structures and usually do not develop significant breast tissue. At puberty, the increase in testosterone may lead to virilizing features despite DHT deficiency. DHT gel can improve prepubertal phallic growth in patients raised as male. Prevention of testosterone exposure (by gonadectomy or pubertal suppression) in adolescence and estrogen replacement at puberty can be considered in individuals who identify as female. Disruption of alternative pathways to fetal DHT production might also present with 46,XY DSD (*AKR1C2/AKR1C4*).

Disruption of Androgen Action • **ANDROGEN INSENSITIVITY SYNDROME** Pathogenic variants in the androgen receptor cause resistance to androgen (testosterone, DHT) action or the *androgen insensitivity syndrome* (*AIS*). AIS is a spectrum of disorders that affects

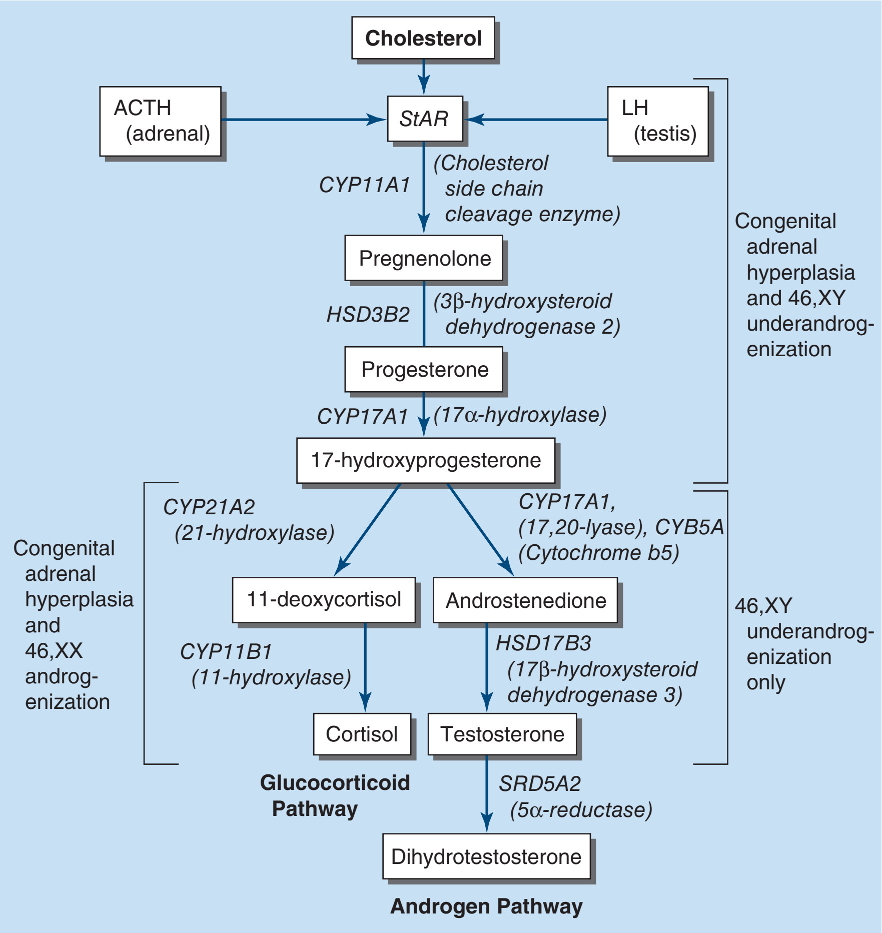
Cholesterol
ACTH (adrenal)
StAR
LH (testis)
CYP11A1
(Cholesterol side chain cleavage enzyme)
Pregnenolone
HSD3B2
(3β-hydroxysteroid dehydrogenase 2)
Progesterone
CYP17A1
(17α-hydroxylase)
17-hydroxyprogesterone
Congenital adrenal hyperplasia and 46,XY underandrogenization
CYP21A2
(21-hydroxylase)
CYP17A1, (17,20-lyase), CYB5A (Cytochrome b5)
Congenital adrenal hyperplasia and 46,XX androgenization
11-deoxycortisol
Androstenedione
46,XY underandrogenization only
CYP11B1
(11-hydroxylase)
HSD17B3
(17β-hydroxysteroid dehydrogenase 3)
Cortisol
Testosterone
Glucocorticoid Pathway
SRD5A2
(5α-reductase)
Dihydrotestosterone
Androgen Pathway

FIGURE 402-4 Simplified overview of glucocorticoid and androgen synthesis pathways. Defects in *CYP21A2* and *CYP11B1* shunt steroid precursors into the androgen pathway and cause androgenization of the 46,XX fetus. Testosterone is synthesized in the testicular Leydig cells and converted to dihydrotestosterone peripherally. Defects in enzymes involved in androgen synthesis result in underandrogenization of the 46,XY fetus. ACTH, adrenocorticotropic hormone; LH, luteinizing hormone; *StAR*, steroidogenic acute regulatory protein.

at least 1 in 100,000 46,XY individuals. Because the androgen receptor is X-linked, only 46,XY offspring are affected. The condition is usually inherited from a mother who carries the sequence variant but can also arise de novo. XY individuals with *complete AIS* (formerly called *testicular feminization syndrome*) have a female phenotype, normal breast development (due to aromatization of testosterone), a short vagina but no uterus (because AMH/MIS production is normal), sometimes scanty pubic and axillary hair, and typically a female gender identity and sex role behavior. Gonadotropins can be variable, but testosterone is usually elevated and AMH/MIS levels in childhood are normal or high. CAIS sometimes presents as inguinal hernias (containing testes) in childhood or more often with primary amenorrhea in late adolescence. In the past, gonadectomy was recommended early in childhood. Increasingly, risk of tumor development is weighed against potential for endogenous hormone function. Thus gonadectomy is often delayed and gonads are left in situ until breast development is complete. Subsequently, the adolescent or young adult should be counseled about the risk of malignancy and the option for gonadectomy (with estrogen replacement), especially because early detection of premalignant changes by imaging or biomarkers is currently not possible. The use of graded dilators in adolescence or young adulthood is often sufficient to dilate the vagina for sexual activity.

Partial AIS (*Reifenstein's syndrome*) results from androgen receptor mutations that maintain residual function. Patients often present in infancy with penoscrotal hypospadias and undescended testes and with gynecomastia at the time of puberty. Gender identity can be variable. At puberty, testes produce testosterone with variable phenotypic development. For those who identify as male, high-dose testosterone has been given to support development if puberty does not progress, but long-term data are limited. For those raised as female, testosterone effects at puberty can be prevented (by pubertal suppression) and female-typical puberty induced with estrogen. They also have an increased risk of GCC, again raising the question of if and when to perform gonadectomy. *Azoospermia* and male-factor infertility also have been described in association with mild loss-of-function mutations in the androgen receptor. Androgen resistance without AR variants is called type 2 AIS (e.g., *DAAM2*).

OTHER DISORDERS AFFECTING 46,XY MALES

Persistent müllerian duct syndrome is the presence of a uterus in an otherwise phenotypic male. This condition can result from pathogenic variants in AMH or its receptor (AMHR2). *Isolated hypospadias* occurs in ~1 in 250 males. Many less severe cases are idiopathic. However, evidence of penoscrotal hypospadias, poor phallic development, bilateral cryptorchidism, and/or other clinical features requires investigation for an underlying DSD (e.g., partial gonadal dysgenesis, partial defect in testosterone action, or even severe forms of 46,XX CAH) or one of many DSD/hypospadias-associated syndromes. Unilateral undescended testes (cryptorchidism) affect >3% of boys at birth. Orchidopexy should be considered if the testis has not descended by 6–9 months of age. Bilateral cryptorchidism occurs less frequently and should raise suspicion of gonadotropin deficiency or DSD. *Syndromic associations* and *intrauterine growth retardation* also occur relatively frequently in association with impaired testicular function or target tissue responsiveness, but the underlying etiology of many of these conditions is unknown except in specific syndromes (e.g., MIRAGE [myelodysplasia, infection, restriction of growth, adrenal hypoplasia, genital phenotypes, and enteropathy], IMAGe [intrauterine growth restriction, metaphyseal dysplasia, adrenal hypoplasia congenita, and genitourinary abnormalities]).

46,XX DSD

Androgenization (virilization) of the 46,XX fetus occurs when the gonad (ovary) contains androgen-secreting testicular tissue or after increased androgen exposure, which is usually adrenal in origin (Table 402-1).

46,XX Testicular DSD Testicular tissue can develop in 46,XX testicular DSD (46,XX males) most often following translocation of *SRY*. This may be diagnosed with karyotype/phenotype discordance or later in life during evaluation for hypogonadism or infertility. Individuals with this condition develop testes with normal testosterone production, leading to external male phenotype in utero, and produce AMH/MIS to regress müllerian structures. They have azoospermia due to lack of the AZF region of the Y chromosome. Progressive testicular regression and hypogonadism are common. Gender identity is typically male.

46,XX OTDSD Ovotestes (or testes) can also develop in individuals with a 46,XX karyotype following upregulation of *SOX9* or *SOX3* or defects in *RSPO1*, *NR2F2*, *WT1*, or *SF1/NR5A1* **(Table 402-5)**. OTDSD is discussed above under "Differences of Chromosomal Sex."

Increased Androgen Exposure • **21-HYDROXYLASE DEFICIENCY (CONGENITAL ADRENAL HYPERPLASIA)** The *classic form* of 21-hydroxylase deficiency (21-OHD) is the most common cause of CAH **(Chap. 398)**, and it is the most common cause of androgenization in chromosomal 46,XX children (incidence around 1 in 15,000

TABLE 402-5 Selected Genetic Causes of 46,XX Differences of Sex Development (DSDs)

GENE	INHERITANCE	GONAD	UTERUS	EXTERNAL GENITALIA	ASSOCIATED FEATURES
Testicular/Ovotesticular DSD					
SRY	Translocation	Testis or ovotestis	–	Male or ambiguous	
SF1/NR5A1 (codon 92)	AD	Testis or ovotestis	+/–	Male or ambiguous	
Other causes of testicular/ovotesticular DSD include: *COUP-TF2/NR2F2* (congenital heart disease), *RSPO1* (palmar plantar hyperkeratosis, squamous cell skin carcinoma), *WNT4* (SERKAL syndrome), *WT1* (in zinc finger 4), dysregulation/duplication of *SOX3* (Xq27) and *SOX9* (dup17q24)					
Increased Androgen Synthesis					
HSD3B2	AR	Ovary	+	Clitoromegaly	CAH, primary adrenal insufficiency, mild androgenization due to ↑ DHEA
CYP21A2	AR	Ovary	+	Ambiguous	CAH, phenotypic spectrum from severe salt-losing forms associated with adrenal insufficiency to simple virilizing forms with compensated adrenal function, ↑ 17-hydroxyprogesterone
POR	AR	Ovary	+	Ambiguous or female	Mixed features of 21-hydroxylase deficiency and 17α-hydroxylase/17,20-lyase deficiency, sometimes associated with Antley-Bixler craniosynostosis
CYP11B1	AR	Ovary	+	Ambiguous	CAH, hypertension due to ↑ 11-deoxycorticosterone
CYP19	AR	Ovary	+	Ambiguous	Maternal virilization during pregnancy, absent breast development at puberty

Abbreviations: ACTH, adrenocorticotropin; AD, autosomal dominant; AR, autosomal recessive; CAH, congenital adrenal hyperplasia; *COUP-TF2*, chicken ovalbumin upstream promoter transcription factor 2; *CYP11B1*, 11β-hydroxylase; *CYP19*, aromatase; *CYP21A2*, 21-hydroxylase; DHEA, dehydroepiandrosterone; *HSD3B2*, 3β-hydroxysteroid dehydrogenase type 2; *POR*, P450 oxidoreductase; *RSPO1*, R-spondin 1; SERKAL, sex reversion, kidneys, adrenal and lung dysgenesis; *SF1*, steroidogenic factor 1; *SOX3*, *SRY*-related HMG-box gene 3; *SOX9*, *SRY*-related HMG-box gene 9; *SRY*, sex-related gene on the Y chromosome; *WT1*, Wilms' tumor–related gene 1.

live births) (Table 402-5). Affected individuals are homozygous or compound heterozygous for severely disruptive sequence variants in the gene (*CYP21A2*) encoding the enzyme 21-hydroxylase. Impaired 21-hydroxylase activity prevents adrenal glucocorticoid and mineralocorticoid synthesis, thus shunting steroid precursors into the androgen synthesis pathway (Fig. 402-4). Increased androgen synthesis in utero causes androgenization of the 46,XX fetus in the first trimester. Atypical genitalia are usually seen at birth, with varying degrees of clitoral enlargement and labial fusion.

A salt-wasting crisis usually manifests between 5 and 21 days of life and is a potentially life-threatening event that requires urgent fluid resuscitation and steroid treatment. Thus, a diagnosis of 21-OHD should be considered in any baby with atypical genitalia with bilateral nonpalpable gonads. Males (46,XY) with 21-OHD have no genital abnormalities at birth but are equally susceptible to adrenal insufficiency and salt-losing crises. Excess androgen production can cause gonadotropin-independent precocious puberty in males with 21-OHD.

Nonclassic 21-OHD is discussed in **Chap. 398**.

TREATMENT

Congenital Adrenal Hyperplasia

Acute salt-wasting crises require fluid resuscitation, IV hydrocortisone, and correction of hypoglycemia. Once the patient is stabilized, glucocorticoids must be given to correct the cortisol insufficiency and suppress adrenocorticotropic hormone (ACTH) stimulation, thereby preventing further virilization, rapid skeletal maturation, and the development of polycystic ovaries. Mineralocorticoid replacement may be needed. Salt supplements may be required in early life. In childhood, treatment is also titrated carefully to prevent impairment of linear growth. In the future, different forms of glucocorticoid replacement and multimodal therapies may improve treatment options. **See Chap. 398 for detailed discussion of hormone replacement.**

Individuals with 46,XX CAH due to classic 21-OHD historically often underwent genitoplasty in infancy, but if and when these procedures should be performed is debated. Concerns have arisen about the importance of assent/consent by the individual for genital surgery, potential long-term side effects related to sexual function and ability to achieve orgasm, and the increased incidence of nonfemale gender identity. Surgical options include vaginoplasty and clitoroplasty. When vaginoplasty is performed in infancy, surgical revision or vaginal dilation may still be needed in adolescence or adulthood and, if deferred, may be necessary for menstrual flow or intercourse. Current clinical practice guidelines can vary in different countries, including supporting parents to defer surgery or for individuals to have no surgery. Women with 21-OHD frequently develop polycystic ovaries and have subfertility. The latter occurs due to multiple factors including anatomic barriers, hormone imbalances, and psychological effects of the condition. Preconception genetic counseling is recommended. Due to concerns about fetal neurologic development, prenatal treatment with dexamethasone to prevent androgenization of a fetus is currently not recommended unless in a study protocol that allows long-term follow-up of all children treated.

The treatment of other forms of CAH (including in 46,XY individuals) includes mineralocorticoid and glucocorticoid replacement for salt-losing conditions (e.g., *StAR*, *CYP11A1*, *HSD3B2*), suppression of ACTH drive with glucocorticoids in disorders associated with hypertension (e.g., *CYP11B1*), and appropriate sex hormone replacement in adolescence and adulthood, when necessary.

OTHER CAUSES Increased androgen synthesis can also occur in CAH due to defects in *POR*, *11β-hydroxylase* (*CYP11B1*), and *3β-hydroxysteroid dehydrogenase type 2* (*HSD3B2*) and with mutations in the genes encoding *aromatase* (*CYP19A1*). Increased androgen exposure in utero can occur with maternal virilizing tumors, luteomas, and ingestion of androgenic compounds.

OTHER DISORDERS AFFECTING 46,XX FEMALES

Congenital absence of the vagina occurs in association with *müllerian agenesis* or *hypoplasia* as part of the Mayer-Rokitansky-Kuster-Hauser (MRKH) syndrome. This diagnosis should be considered in otherwise phenotypic females with primary amenorrhea. Associated features include renal (agenesis) and cervical spinal abnormalities.

LONG-TERM FOLLOW-UP

Whether a DSD is diagnosed during childhood or adulthood, many individuals will require ongoing care. Medical care may include hormone replacement, optimization of bone health, screening related to other associated clinical features, and care for infertility or sterility. There is increasing attention to fertility preservation and investigational gonadal tissue cryopreservation in the pediatric and young adult population, which may allow for increased fertility potential in adulthood. More frequently, individuals retain gonadal tissue at increased risk for germ cell cancer, requiring monitoring and potential gonadectomy. Long-term urogynecologic follow-up may be needed. Psychosocial and psychosexual evaluation and management are also important, particularly to address the reported increased rates of mental health symptoms and poorer quality of life sometimes related to social relationships.

■ GLOBAL CONSIDERATIONS

The approach to an individual with atypical genitalia or another DSD requires cultural sensitivity, as the concepts of sex and gender vary widely around the world. Rare genetic DSDs can occur more frequently in specific populations (e.g., *5α-reductase type 2* in the Dominican Republic). Different forms of CAH also show ancestral and geographic variability. In many countries, appropriate biochemical or genetic tests may not be readily available, and access to appropriate forms of treatment and support may be limited.

■ FURTHER READING

Ahmed SF et al: Turner HE. Society for Endocrinology UK Guidance on the initial evaluation of a suspected difference or disorder of sex development (Revised 2021). Clin Endocrinol (Oxf) 95:818, 2021.

Claahsen-van der Grinten HL et al: Congenital adrenal hyperplasia-current insights in pathophysiology, diagnostics and management. Endocr Rev 43:91, 2022.

Cools M et al: Caring for individuals with a difference of sex development (DSD): A consensus statement. Nat Rev Endocrinol 14:415, 2018.

Gravholt CH et al: The changing face of Turner syndrome. Endocr Rev 12:44, 2023.

Gravholt CH et al: Clinical practice guidelines for the care of girls and women with Turner syndrome. Eur J Endocrinol 190:G53, 2024.

Mongan NP et al: Androgen insensitivity syndrome. Best Pract Res Clin Endocrinol Metab 29:569, 2015.

Morin J et al: Gonadal malignancy in patients with differences of sex development. Transl Androl Urol 9:2408, 2020.

Zitzmann M et al: European Academy of Andrology guidelines on Klinefelter syndrome: Endorsing organization: European Society of Endocrinology. Andrology 9:145, 2021.

403

Disorders of the Testes and Male Reproductive System

Shalender Bhasin, J. Larry Jameson

The male reproductive system regulates sex differentiation, androgenization, and the hormonal changes that accompany puberty, ultimately leading to spermatogenesis and fertility. Under the control of the pituitary hormones—luteinizing hormone (LH) and follicle-stimulating hormone (FSH)—the Leydig cells of the testes produce testosterone and germ cells are nurtured by Sertoli cells to divide, differentiate, and mature into sperm. During embryonic development, testosterone and dihydrotestosterone (DHT) induce the wolffian duct and virilization of the external genitalia. During puberty, testosterone promotes somatic growth and the development of secondary sex characteristics. In the adult, testosterone is necessary for spermatogenesis, libido and normal sexual function, and maintenance of muscle and bone mass. This chapter focuses on the physiology of the testes and disorders associated with decreased androgen production, which may be caused by gonadotropin deficiency or by primary testis dysfunction. Infertility occurs in ~5% of men and is increasingly amenable to treatment by hormone replacement or by using sperm transfer techniques. **For further discussion of sexual dysfunction, disorders of the prostate, and testicular cancer, see Chaps. 409, 92, and 93, respectively.**

DEVELOPMENT AND STRUCTURE OF THE TESTIS

The fetal testis develops from a single bipotential progenitor cell population in the undifferentiated gonad after expression of a genetic cascade that is initiated by the gene encoding SRY (sex-related gene on the Y chromosome) **(Chap. 402)**. SRY, whose expression is regulated by histone modification and DNA methylation, induces differentiation of Sertoli cells, which surround germ cells and, together with peritubular myoid cells, form testis cords that will later develop into seminiferous tubules. Sertoli cells direct the differentiation of fetal germ cells into spermatogenic precursors. Fetal Leydig cells and endothelial cells migrate into the gonad from the adjacent mesonephros but may also arise from interstitial cells that reside between testis cords. Fetal Leydig cells atrophy after birth and do not contribute to the origin of adult Leydig cells, which originate from undifferentiated progenitor cells that appear in the testis after birth and acquire full steroidogenic function during puberty. Testosterone produced by the fetal Leydig cells supports the growth and differentiation of Wolffian duct structures that develop into the epididymis, vas deferens, and seminal vesicles. Testosterone is also converted to DHT (see below), which induces formation of the prostate and the external male genitalia, including the penis, urethra, and scrotum. The development of the germ cells is dependent on normally functioning Sertoli and Leydig cells and cannot occur without them; in contrast, the somatic cells can develop and function without the germ cells. Testicular descent through the inguinal canal is controlled in part by Leydig cell production of insulin-like factor 3 (INSL3), which along with testosterone, acts through their cognate receptors—relaxin/insulin-like family peptide receptor 2 (*RXFP2*) and androgen receptor (AR)—to guide the inguinoscrotal descent of the testes. Sertoli cells produce Müllerian inhibiting substance (MIS), which causes regression of the Müllerian structures, including the fallopian tube, uterus, and upper segment of the vagina.

NORMAL MALE PUBERTAL DEVELOPMENT

Puberty commonly refers to the maturation of the reproductive axis, initiation of gametogenesis, rhythmic sex hormone secretion, and the development of secondary sex characteristics. In addition to reproductive hormones, it requires a coordinated response of multiple hormonal systems including metabolic signals (e.g., leptin), as well as the adrenal and growth hormone (GH, IGF-1) axes **(Fig. 403-1)**. The development of secondary sex characteristics is initiated by *adrenarche*, which usually occurs between 6 and 8 years of age when the adrenal gland begins to produce greater amounts of androgens from the zona reticularis, the principal site of dehydroepiandrosterone (DHEA) production. The sex maturation process is accelerated by the activation of the hypothalamic-pituitary axis and the production of gonadotropin-releasing hormone (GnRH). The GnRH pulse generator in the infundibular

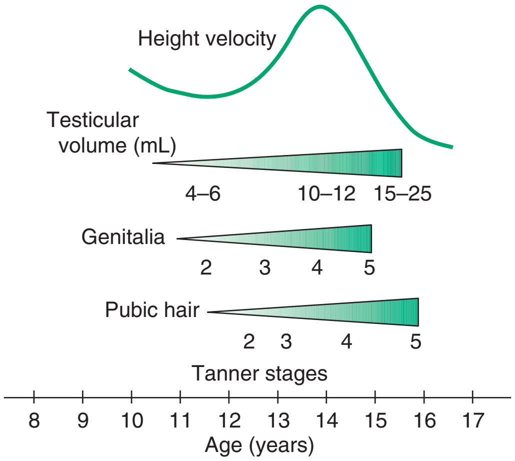

FIGURE 403-1 Pubertal events in males. Sexual maturity ratings for genitalia and pubic hair and divided into five stages.

nucleus is active during fetal life and early infancy but is restrained until the early stages of puberty by a neuroendocrine brake imposed by the inhibitory actions of glutamate and γ-aminobutyric acid (GABA) in the mediobasal hypothalamus and neuropeptide Y. Kisspeptin 1 (kiss1), neurokinin B, dynorophin (KNDy) neurons in the arcuate nucleus and possibly other kiss1 neurons in the anteroventral periventricular nucleus (AVPV), and neurons producing neuropeptide Y (NPY), agouti-related protein (AgRP), pro-opiomelanocortin (POMC), and cocaine and amphetamine-regulated transcript (CART) play an important role in the control of puberty. The Kisspeptin pathway plays a critical role in puberty. Individuals with mutations of G protein-coupled kisspeptin receptor fail to enter puberty. Infusion of the Kiss1 is sufficient to induce premature puberty in nonhuman primates. Kisspeptin signaling plays an important role in mediating the feedback action of sex steroids on gonadotropin secretion and in regulating the tempo of sexual maturation at puberty. The metabolic control of puberty is mediated via the action of leptin secreted by adipocytes, and by neurons expressing NPY and AgRP, and by POMC and CART expressing neurons that transmit the energy sensing signals communicated by leptin to the Kiss1 neurons. *MKRN3*, an imprinted gene encoding makorin ring finger protein 3, serves as an upstream inhibitor of GnRH secretion and plays an important role in directing the onset of puberty. Leptin plays a permissive role in the resurgence of GnRH secretion at the onset of puberty, as leptin-deficient individuals also fail to enter puberty **(Chap. 413)**. The adipocyte hormone leptin, the gut hormone ghrelin, neuropeptide Y, AgRP, CART, and kiss 1 integrate the signals originating in energy stores and metabolic tissues and control the onset of puberty through regulation of GnRH secretion. Energy deficit and excess and metabolic stress are associated with disturbed reproductive maturation and timing of puberty.

The early stages of puberty are characterized by nocturnal surges of LH and FSH. Growth of the testes is usually the first clinical sign of puberty, reflecting an increase in seminiferous tubule volume. Increasing levels of testosterone deepen the voice and stimulate muscle growth. Conversion of testosterone to DHT leads to growth of the external genitalia and pubic hair. DHT also stimulates prostate and facial hair growth and initiates recession of the temporal hairline. The growth spurt occurs at a testicular volume of ~10–12 mL. GH increases early in puberty and is stimulated in part by the rise in gonadal steroids. GH increases the level of insulin-like growth factor 1 (IGF-1), which enhances linear bone growth. The prolonged pubertal exposure to gonadal steroids (mainly estradiol) ultimately induces epiphyseal closure and limits further bone growth.

REGULATION OF TESTICULAR FUNCTION

■ REGULATION OF THE HYPOTHALAMIC-PITUITARY-TESTIS AXIS IN ADULT MAN

Pulsatile secretion of GnRH in the hypothalamus is regulated by kisspeptin, neurokinin B (NKB), and dynorphin (DYN) **(Fig. 403-2)** that are co-expressed within the KNDy neurons in the arcuate nucleus and are important regulators of mammalian reproduction. Kisspeptin binds to the kisspeptin (GPR54) receptors in the cell bodies of the GnRH neurons as well as in the GnRH nerve terminals in the median eminence to induce pulsatile GnRH secretion into the portal blood. As a component of this autocrine/paracrine loop, NKB released by the KNDy neurons activates NK3R to stimulate kisspeptin release. KNDy neurons also produce dynorphin A, which inhibits basal as well as NKB-stimulated kisspeptin release through the mediation of K-type opioid receptor. Exogenous opioids suppress LH and FSH and inhibit LH pulsatility likely by mimicking the effects of DYN, an endogenous opioid peptide that acts primarily through the κ-opioid receptor (KOR). The negative feedback effects of testosterone, estradiol, and progesterone are mediated through KNDy neurons in the preoptic area by inhibition of kisspeptin release. Mutations in the genes encoding NKB, tachykinin 3 (TAC3), or its receptor (TACR3) are associated with hypogonadotrophic hypogonadism.

Hypothalamic GnRH regulates the production of the pituitary gonadotropins LH and FSH (Fig. 403-2). GnRH is released in discrete pulses approximately every 2 h, resulting in corresponding pulses of LH and FSH. These dynamic hormone pulses account in part for the wide variations in LH and testosterone, even within the same individual. LH acts primarily on the Leydig cell to stimulate testosterone synthesis. The regulatory control of androgen synthesis is modulated by dynamic integration of the feedforward elements exerted on the testis by LH

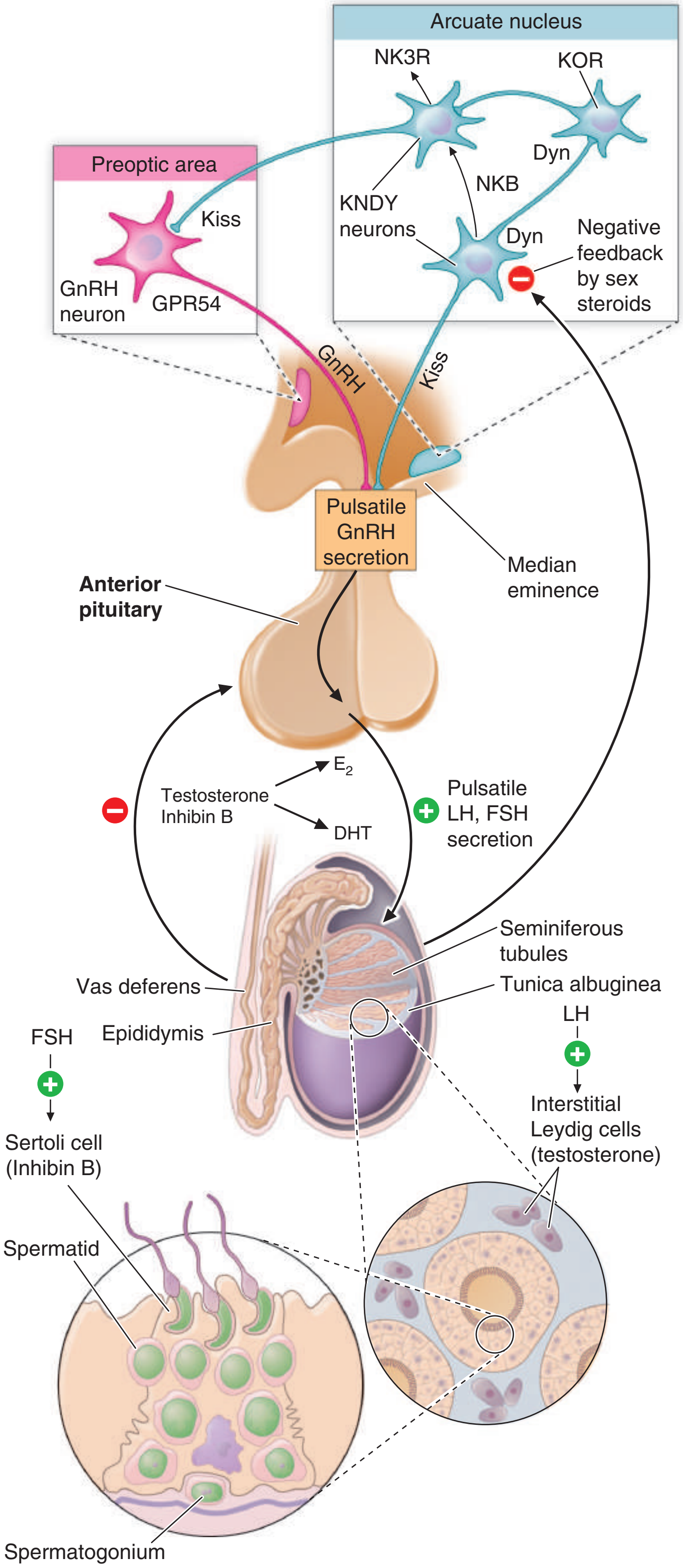

FIGURE 403-2 Hypothalamic-pituitary-gonadotropin axis, structure of testis, seminiferous tubule. DHT, dihydrotestosterone; Dyn, dynorphin A; E_2, 17β-estradiol; FSH, follicle-stimulation hormone; GnRH, gonadotropin-releasing hormone; KISSR, G protein–coupled receptor 54 for kisspeptin; Kiss, kisspeptin; KNDY, kisspeptin, neurokinin B, dynorphin neurons; KOR, K-type opioid receptor; LH, luteinizing hormone; NKB, neurokinin B; NK3R, neurokinin 3 receptor.

and FSH and the feedback exerted by testosterone and estrogen on both the hypothalamus and the pituitary. FSH acts on the Sertoli cell to regulate spermatogenesis and the production of inhibin B, which acts to selectively suppress pituitary FSH. Despite these somewhat distinct Leydig and Sertoli cell–regulated pathways, testis function is integrated at several levels: GnRH regulates both gonadotropins; spermatogenesis requires high levels of testosterone; and numerous paracrine interactions between Leydig, Sertoli, and germ cells are necessary for normal testis function.

THE LEYDIG CELL: ANDROGEN SYNTHESIS

LH binds to its transmembrane G protein–coupled receptor to activate the adenylyl cyclase/cyclic AMP/protein kinase A pathway. Stimulation of the LH receptor induces *st*eroid *a*cute *r*egulatory (StAR) protein, along with several steroidogenic enzymes involved in androgen synthesis. LH receptor mutations cause Leydig cell hypoplasia or agenesis, underscoring the importance of this pathway for Leydig cell development and function. The rate-limiting process in testosterone synthesis is the transport of intracellular cholesterol to the inner mitochondrial membrane by a protein complex that contains the steroidogenic acute regulatory (StAR) protein and a translocator protein (TSPO), previously called peripheral benzodiazepine receptor protein. Mutations of the StAR protein are associated with congenital lipoid adrenal hyperplasia, a rare form of congenital adrenal hyperplasia (CAH) characterized by very low adrenal and gonadal steroids. The major enzymatic steps involved in testosterone synthesis are summarized in **Fig. 403-3**. After cholesterol transport into the mitochondrion, the formation of pregnenolone, catalyzed by CYP11A1 (side chain cleavage enzyme) and auxiliary electron transferring proteins, is a limiting enzymatic step. The 17α-hydroxylase and the 17,20-lyase

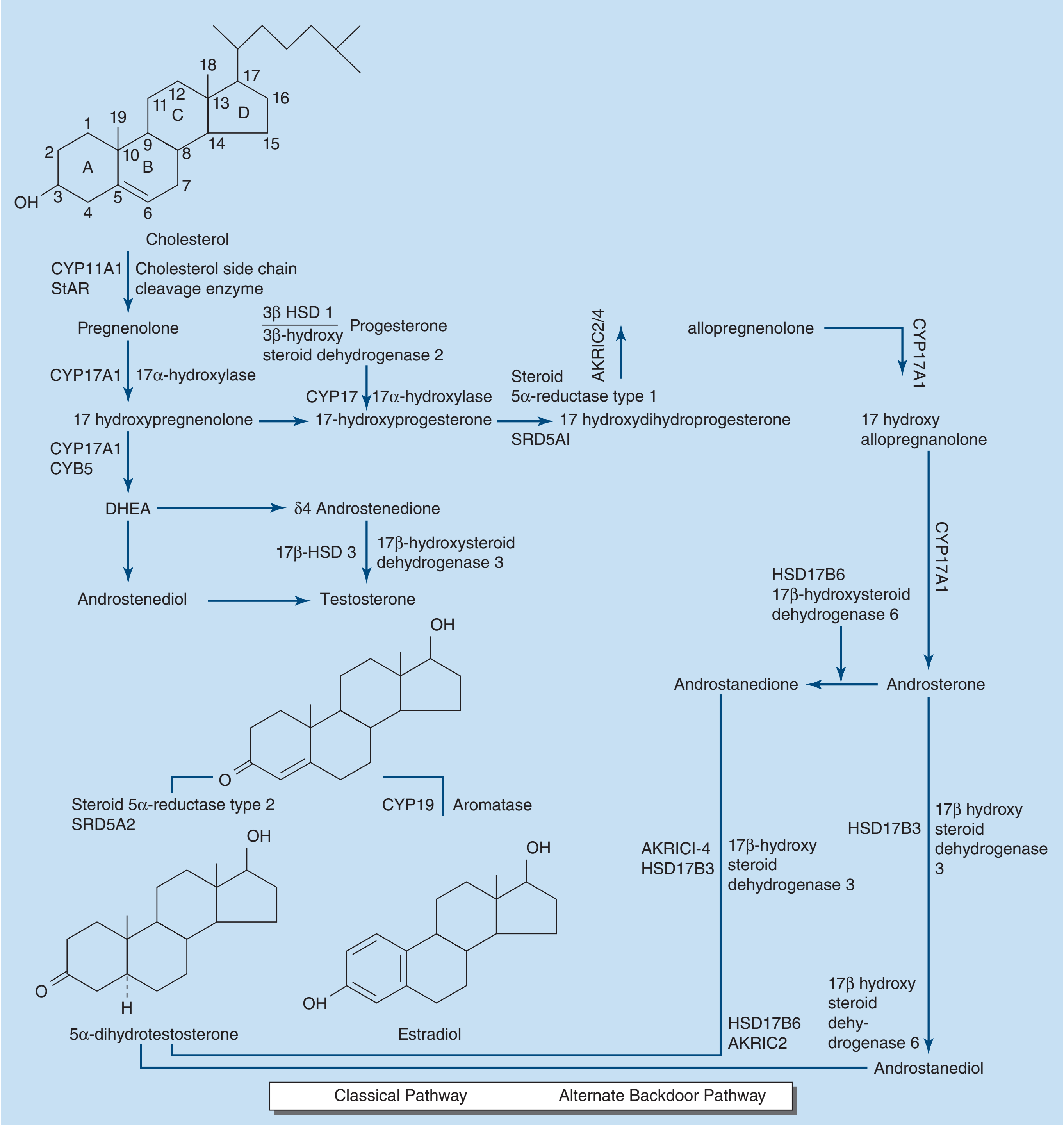
18
17
16
11
12
C
13
D
1
19
9
14
15
2
10
8
A
B
7
OH
3
5
4
6
Cholesterol
CYP11A1
StAR
Cholesterol side chain
cleavage enzyme
Pregnenolone
3β HSD 1
3β-hydroxy
steroid dehydrogenase 2
Progesterone
AKRIC2/4
allopregnenolone
CYP17A1
CYP17A1
17α-hydroxylase
Steroid
5α-reductase type 1
CYP17
17α-hydroxylase
17 hydroxypregnenolone
17-hydroxyprogesterone
17 hydroxydihydroprogesterone
SRD5AI
17 hydroxy
allopregnanolone
CYP17A1
CYB5
DHEA
δ4 Androstenedione
17β-HSD 3
17β-hydroxysteroid
dehydrogenase 3
CYP17A1
HSD17B6
17β-hydroxysteroid
dehydrogenase 6
Androstenediol
Testosterone
OH
Androstanedione
Androsterone
O
Steroid 5α-reductase type 2
SRD5A2
CYP19
Aromatase
OH
OH
AKRICI-4
HSD17B3
17β-hydroxy
steroid
dehydrogenase 3
HSD17B3
17β hydroxy
steroid
dehydrogenase
3
O
H
OH
17β hydroxy
steroid
dehy-
drogenase 6
5α-dihydrotestosterone
Estradiol
HSD17B6
AKRIC2
Androstanediol
Classical Pathway
Alternate Backdoor Pathway

FIGURE 403-3 The biochemical pathway in the conversion of 27-carbon sterol cholesterol to androgens and estrogens.

reactions are catalyzed by a single enzyme, CYP17A1; posttranslational modification (phosphorylation) of this enzyme and the presence of specific enzyme cofactors, such as cytochrome B, confer 17,20-lyase activity selectively in the testis and zona reticularis of the adrenal gland. Although CYP17A1 is able to catalyze the conversion of progesterone to 17α-hydroxyprogesterone, most of δ4-androstenedione in humans is derived from the conversion of 17α-hydroxypregnenolone to DHEA in the δ5 pathway and further conversion of DHEA to δ4-androstenedione. Abiraterone is a dual inhibitor of 17α-hydroxylase and 17,20-lyase activities, which play an important role in androgen synthesis in castration-resistant prostate cancers. Testosterone can be converted to the more potent DHT by a family of steroid 5α-reductase enzymes, and it is aromatized to estradiol by CYP19 (aromatase). At least two isoforms of steroid 5α-reductase, SRD5A1 and SRD5A2, have been described; most known patients with 5α-reductase deficiency have had mutations in SRD5A2, the predominant form in the prostate and the skin. Finasteride predominantly inhibits SRD5A2, whereas dutasteride is a dual inhibitor of both SRD5A1 and SRD5A2. DHT can also be derived through the backdoor pathway in which 17α-hydroxyprogesterone is converted to androsterone and eventually to DHT. Recent reports of mutations in the *AKR1C2/4* genes in undervirilized 46,XY individuals suggest that the backdoor pathway for DHT formation, which was originally described in the tammar wallaby, is active in the human fetal testis. The placental progesterone serves as a substrate for the synthesis of androsterone via the backdoor pathway, which is then converted to DHT in the genital tubercle.

Testosterone Transport and Metabolism In males, 95% of circulating testosterone is derived from testicular production (3–10 mg/d). Direct secretion of testosterone by the adrenal and the peripheral conversion of androstenedione to testosterone collectively account for another 0.5 mg/d of testosterone. Only a small amount of DHT (70 μg/d) is secreted directly by the testis; most circulating DHT is derived from peripheral conversion of testosterone. Most of the daily production of estradiol (~45 μg/d) in men is derived from aromatase-mediated peripheral conversion of testosterone and androstenedione.

Circulating testosterone is bound predominantly to sex hormone–binding globulin (SHBG) and albumin **(Fig. 403-4)** and, to a lesser extent, to cortisol-binding globulin (CBG) and orosomucoid (also known as α_1 acid glycoprotein). SHBG binds testosterone with much greater affinity than albumin, CBG, and orosomucoid. The binding proteins regulate the transport and bioavailability of testosterone. SHBG circulates as a dimer, and testosterone's binding to SHBG involves intermonomeric allostery such that neither the conformation nor the binding affinity of the two monomers is equivalent. Similarly, estradiol binding to SHBG involves bidirectional, intermonomeric allostery that changes the distribution of both monomers among various energy and conformational states. Intermonomeric allostery offers a mechanism to extend the binding range of SHBG and regulate hormone bioavailability as sex hormone concentrations vary widely during life. Albumin contains multiple, allosterically coupled binding sites for testosterone. Testosterone shares these binding sites on albumin with free fatty acids. Commonly used drugs such as ibuprofen and some antibiotics can displace testosterone from albumin under various physiologic states or disease conditions, affecting its bioavailability. SHBG concentrations are decreased by androgens, obesity, diabetes mellitus, hypothyroidism, nephrotic syndrome, and genetic factors. Conversely, polymorphisms in the SHBG gene, estrogen administration, hyperthyroidism, chronic inflammatory illnesses, infections such as HIV or hepatitis B and C, aging, and the use of some anticonvulsants are associated with high SHBG concentrations.

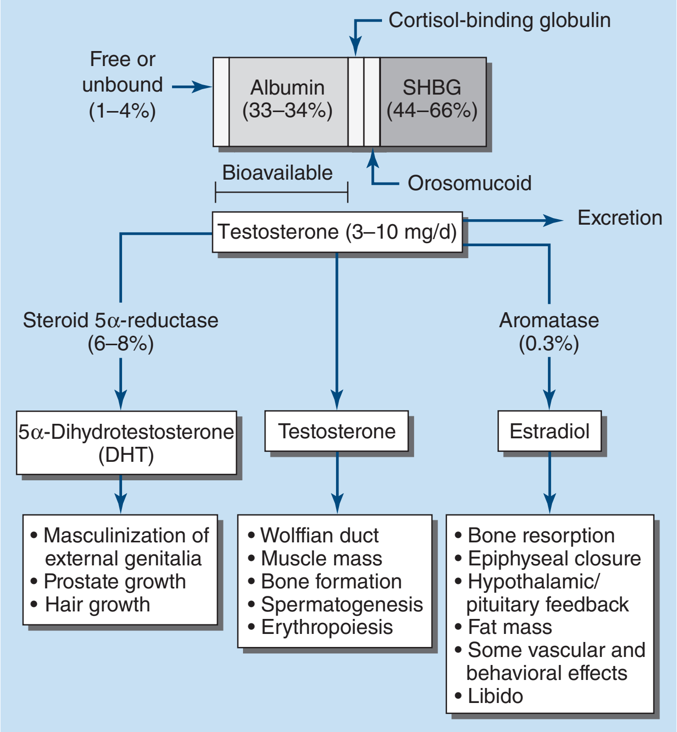

FIGURE 403-4 Androgen metabolism and actions. SHBG, sex hormone–binding globulin.

Testosterone is metabolized predominantly in the liver, although some degradation occurs in peripheral tissues, particularly the prostate and the skin. In the liver, testosterone is converted by a series of enzymatic steps that involve 5α and 5β-reductases, 3α- and 3β-hydroxysteroid dehydrogenases, and 17β-hydroxysteroid dehydrogenase into androsterone, etiocholanolone, DHT, and 3α-androstanediol. These compounds undergo glucuronidation or sulfation before being excreted by the kidneys. Polymorphisms of these enzymes are associated with altered testosterone metabolism and circulating testosterone levels.

Mechanism of Androgen Action Testosterone exerts its biologic effects by either binding to androgen receptor (AR) directly or after its conversion to its metabolites, DHT, and 11-ketotestosterone. 11-Ketotestosterone and 11-ketoandrostenedione are 11-oxygenated steroids and potent androgens that are produced in the testes, adrenals, and some peripheral tissues by the action of CYP11B1 and CYP11B2. DHT is converted in some tissues by 3β-hydroxysteroid dehydrogenase to 5α-androstane-3β,17β-diol, which is a high-affinity ligand and agonist of estrogen receptor β. 5α-DHT can also be converted by 3α-hydroxysteroid dehydrogenase in some cell types to 5α-androstane-3α,17β-diol, a modulator of $GABA_A$ receptors. Testosterone also is aromatized by CYP19 to 17β estradiol, a ligand for estrogen receptor, that can be further converted in tissues to catechol estrogens by addition of hydroxyl groups at C2 and C4 that are inhibitors of the catecholamine methyltransferase and affect angiogenesis and cell proliferation **(Fig. 403-5)**. 16-Hydroestradiol levels have been associated with breast cancer risk in women. The actions of testosterone on the Wolffian structures, skeletal muscle, erythropoiesis, and bone in men do not require its obligatory conversion to DHT. However, the conversion of testosterone to DHT is necessary for the masculinization of the urogenital sinus and genital tubercle but is not obligatory for mediating its effects on the muscle, bone, or hematopoiesis. Aromatization of testosterone to estradiol mediates additional effects of testosterone on bone resorption, epiphyseal closure, sexual desire, vascular endothelium, and fat.

The AR is structurally related to the nuclear receptors for estrogen, glucocorticoids, and progesterone **(Chap. 389)**. The AR, a 919–amino acid protein with a molecular mass of ~110 kDa, is encoded by a gene on the long arm of the X chromosome. A polymorphic region in the amino terminus of the receptor, which contains a variable number of glutamine and glycine repeats, modifies the transcriptional activity of the receptor. The ligand binding to the AR induces conformational changes that dissociates it from heat shock protein and allows the recruitment and assembly of tissue-specific cofactors and causes it to translocate into the nucleus, where it binds to specific androgen response elements in the DNA or other transcription factors already bound to DNA. Thus, the AR is a ligand-activated transcription factor that regulates the expression of androgen-dependent genes in a tissue-specific manner. Some

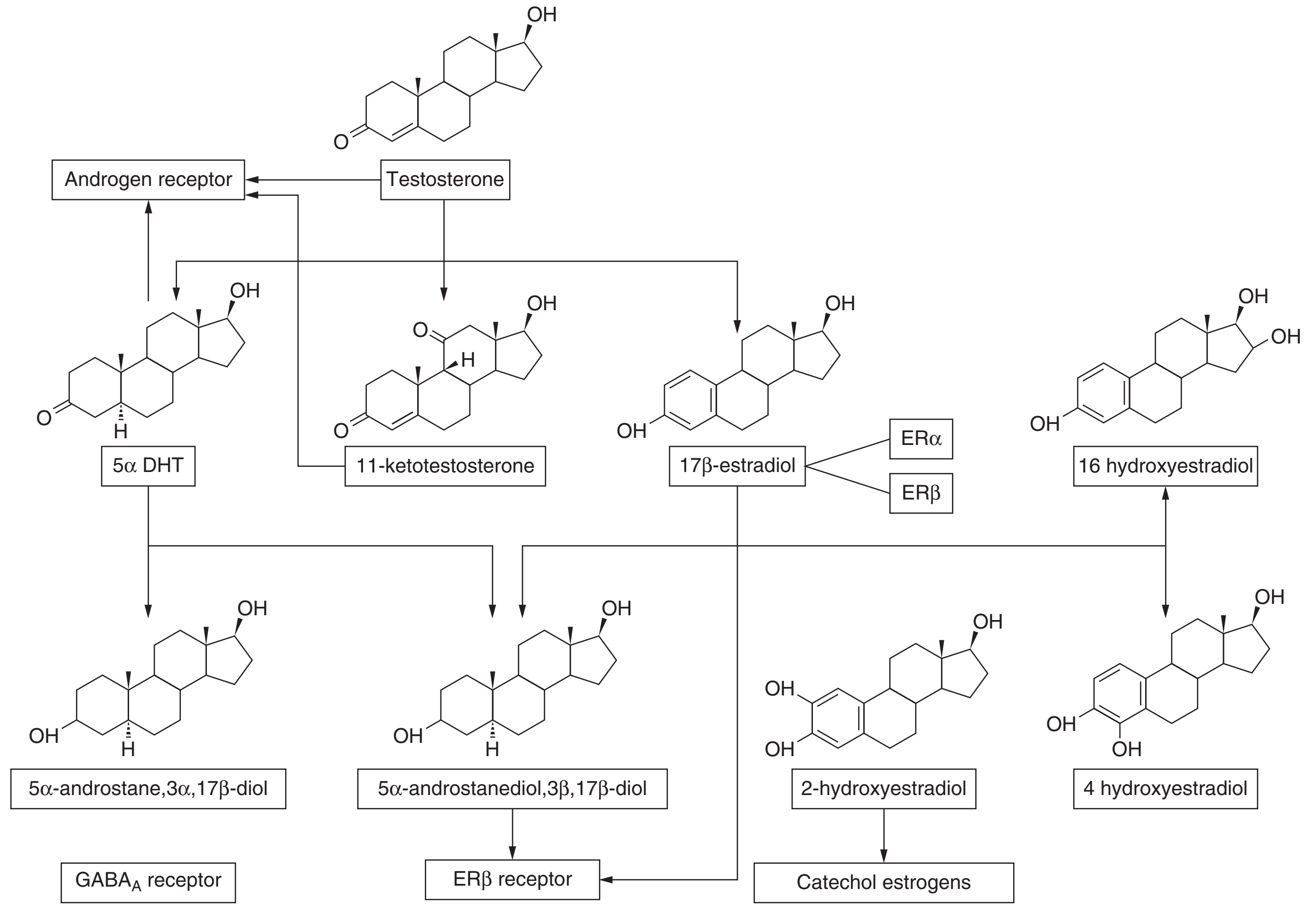

FIGURE 403-5 Testosterone as an androgen and precursor of other steroids with diverse actions. DHT, dihydrotestosterone; ERα, estrogen receptor α; Eαβ, estrogen receptor β.

androgen effects, such as those on the smooth muscle, may be mediated by nongenomic AR signal transduction pathways. The nongenomic actions of testosterone involve direct activation of kinase signaling cascades such as the Src tyrosine kinase that activates the epidermal growth factor receptor and induces the downstream mitogen-activated protein kinase cascade (RAF, MEK, and ERK), resulting in increased transcription of the cyclic AMP response element binding protein transcription factor-regulated genes. Some effects of testosterone on cell proliferation and autophagy require the mediation of GPRC6A. Testosterone and DHT improve penile blood flow and erections by rapid nongenomic endothelial-dependent (increased nitric oxide production) and endothelium-independent inhibition of l-type calcium channels and/or activation of potassium channels in the vascular smooth muscle.

THE SEMINIFEROUS TUBULES: SPERMATOGENESIS

The seminiferous tubules are convoluted, closed loops that empty into the rete testis, a network of progressively larger efferent ducts that ultimately form the epididymis (Fig. 403-2). The seminiferous tubules total ~600 m in length and compose about two-thirds of testis volume. The walls of the tubules are formed by polarized Sertoli cells that are apposed to peritubular myoid cells. Tight junctions between Sertoli cells create the blood-testis barrier. Germ cells compose the majority of the seminiferous epithelium (~60%) and are intimately embedded within the cytoplasmic extensions of the Sertoli cells, which function as "nurse cells." Germ cells progress through characteristic stages of mitotic and meiotic divisions. A pool of type A spermatogonia serve as stem cells capable of self-renewal. Primary spermatocytes are derived from type B spermatogonia and undergo meiosis before progressing to spermatids that undergo spermiogenesis (a differentiation process involving chromatin condensation, acquisition of an acrosome, elongation of cytoplasm, and formation of a tail) and are released from Sertoli cells as mature spermatozoa. The complete differentiation process into mature sperm requires 74 days. Peristaltic-type action by peritubular myoid cells transports sperm into the efferent ducts. The spermatozoa spend an additional 21 days in the epididymis, where they undergo further maturation and capacitation. The normal adult testes produce >100 million sperm per day.

Naturally occurring mutations in FSHβ or in the FSH receptor confirm an important, but not essential, role for this pathway in spermatogenesis. Females with mutations in FSHβ or the FSH receptor are hypogonadal and infertile because ovarian follicles do not mature; males with these mutations exhibit variable degrees of reduced spermatogenesis, presumably because of impaired Sertoli cell function. Because Sertoli cells produce inhibin B, an inhibitor of FSH, seminiferous tubule damage (e.g., by radiation) causes a selective increase of FSH. Testosterone reaches very high concentrations locally in the testis and is essential for spermatogenesis. The cooperative actions of FSH and testosterone are important in the progression of meiosis and spermiation. In the prepubertal testis, testosterone alone is insufficient for completion of spermatogenesis; however, in men with postpubertal onset of gonadotropin deficiency, human chorionic gonadotropin (hCG) or recombinant LH can reinitiate spermatogenesis without FSH. FSH and testosterone regulate germ cell survival via the intrinsic and extrinsic apoptotic mechanisms. FSH may also play an important role in supporting spermatogonia. Gonadotropin-regulated testicular RNA helicase (GRTH/DDX25), a testis-specific gonadotropin/androgen-regulated RNA helicase, is present in germ cells and Leydig cells and may be an important factor in the paracrine regulation of germ cell development. Several cytokines and growth factors are also involved in the regulation of spermatogenesis by paracrine and autocrine mechanisms.

The human Y chromosome contains two pseudoautosomal regions that are located at the two tips of Y chromosome and can recombine with homologous regions of the X chromosome **(Fig. 403-6)**. The genes

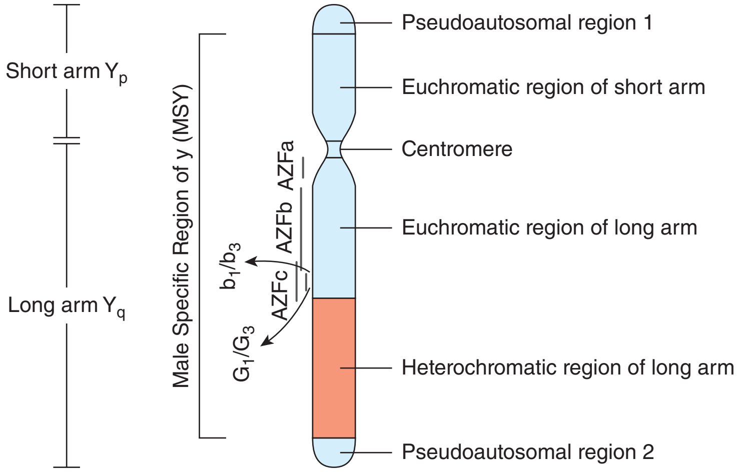
Short arm Y_p
Long arm Y_q
Male Specific Region of y (MSY)
b_1/b_3
G_1/G_3
AZFc
AZFb
AZFa
Pseudoautosomal region 1
Euchromatic region of short arm
Centromere
Euchromatic region of long arm
Heterochromatic region of long arm
Pseudoautosomal region 2

FIGURE 403-6 Structure of the Y chromosome relevant for spermatogenesis.

in the pseudoautosomal regions are involved in cell signaling, transcriptional regulation, and mitochondrial function. Mutations of genes in pseudoautosomal region 1 are associated with mental disorders and short stature. The euchromatic part of the Y chromosome that does not recombine with the X chromosome is referred to as the male-specific region of the Y chromosome (MSY) and contains X-transposed sequences that have 99% homology with the X chromosome, single-copy genes or pseudogene homologues of X-linked genes, and ampliconic sequences that are organized in large palindromes. The MSY contains nine families of Y-specific multicopy genes; many of these Y-specific genes are testis-specific and necessary for spermatogenesis.

Nonallelic homologous recombination between highly homologous repeated sequences in the MSY region renders the Y chromosome susceptible to microdeletions. Microdeletions in four major nonoverlapping subregions of the Y chromosome—AZFa, AZFb, AZFbc, and AZFc—which contain many spermatogenic genes, are associated with oligospermia or azoospermia. Microdeletions involving the AZFc region are the most common Y microdeletions in infertile men. Several partial deletions of the AZFc region have been described including the gr/gr deletion, which is associated with infertility among Caucasian men in Europe and the Western Pacific region, whereas the b2/b3 deletion is associated with male infertility in African and Dravidian men. Approximately 6 to 15% of infertile men with nonobstructive azoospermia or severe oligozoospermia harbor a Y microdeletion. Complete deletions of the AZFa and AZFb subregions are typically associated with Sertoli cells only and azoospermia and a poor prognosis for sperm retrieval. In contrast, AZFc subregion microdeletions are typically associated with oligozoospermia and higher success rates for sperm retrieval.

TREATMENT

Male Factor Infertility

Treatment options for male factor infertility have expanded greatly in recent years. Secondary hypogonadism is highly amenable to treatment with pulsatile GnRH or gonadotropins (see below). Assisted reproductive technologies, such as in vitro fertilization (IVF) and intracytoplasmic sperm injection (ICSI), have provided new opportunities for patients with primary testicular failure and disorders of sperm transport. Choice of initial treatment options depends on sperm concentration and motility. Expectant management should be attempted initially in men with mild male factor infertility (sperm count of 15–20 × 10^6/mL and normal motility). Treatment of moderate male factor infertility (10–15 × 10^6/mL and 20–40% motility) should begin with intrauterine insemination alone or in combination with treatment of the female partner with clomiphene or gonadotropins, but it may require IVF with or without ICSI. For men with a severe defect (sperm count of <10 × 10^6/mL, 10% motility), IVF with ICSI or donor sperm has become the treatment of choice. Yq microdeletions will be transmitted through ICSI from the affected father to his male offspring if sperm carrying the Yq microdeletion is used.

CLINICAL AND LABORATORY EVALUATION OF MALE REPRODUCTIVE FUNCTION

HISTORY AND PHYSICAL EXAMINATION

The history should focus on developmental stages such as puberty and linear growth, as well as androgen-dependent events such as early morning erections, frequency and intensity of sexual thoughts, and frequency of masturbation or intercourse. Although libido and the overall frequency of sexual acts are decreased in androgen-deficient men, young hypogonadal men can achieve erections in response to visual erotic stimuli. Men with acquired androgen deficiency often report decreased energy, low mood, and less commonly hot flushes.

The physical examination should focus on secondary sex characteristics such as hair growth, gynecomastia, testicular volume, prostate, and height and body proportions. *Eunuchoid proportions* are defined as an arm span >2 cm greater than height and suggest that androgen deficiency occurred before epiphyseal fusion. Hair growth in the face, axilla, chest, and pubic regions is androgen-dependent; however, changes may not be noticeable unless androgen deficiency is severe and prolonged. Ethnicity also influences the intensity of hair growth **(Chap. 406)**. Testicular volume is measured by using a Prader orchidometer. Testes range from 3.5 to 5.5 cm in length, which corresponds to a volume of 12–25 mL. Advanced age does not influence testicular size, although the consistency becomes less firm. Because of its possible role in infertility, the presence of varicocele should be sought by palpation while the patient is standing; it is more common on the left side. Patients with Klinefelter syndrome have markedly reduced testicular volumes (1–2 mL). In congenital hypogonadotropic hypogonadism, testicular volumes provide a good index of the degree of gonadotropin deficiency and the likelihood of response to gonadotropin therapy.

GONADOTROPIN AND INHIBIN MEASUREMENTS

LH and FSH are measured using two-site immunoradiometric, immunofluorometric, or chemiluminescent assays, which have very low cross-reactivity with other pituitary glycoprotein hormones and hCG and have sufficient sensitivity to measure the low levels present in patients with hypogonadotropic hypogonadism. In men with a low testosterone level, an LH level can distinguish primary (high LH) versus secondary (low or inappropriately normal LH) hypogonadism. An elevated LH level indicates a primary defect at the testicular level, whereas a low or inappropriately normal LH level suggests a defect at the hypothalamic-pituitary level. LH pulses occur about every 1–3 h in normal men. Thus, gonadotropin levels fluctuate, and samples should be pooled or repeated when results are equivocal. FSH is less pulsatile than LH because it has a longer half-life. Selective increase in FSH suggests damage to the seminiferous tubules. Inhibin B, a Sertoli cell product that suppresses FSH, is reduced with seminiferous tubule damage. Inhibin B is a dimer with α-β_B subunits and is measured by two-site immunoassays.

GnRH Stimulation Testing The GnRH test is performed by measuring LH and FSH concentrations at baseline and at 30 and 60 min after intravenous administration of 100 μg of GnRH. A minimally acceptable response is a twofold LH increase and a 50% FSH increase. In the prepubertal period or with severe GnRH deficiency, the gonadotrope may not respond to a single bolus of GnRH; in these patients, GnRH responsiveness may be restored by chronic, pulsatile GnRH administration. With the availability of sensitive and specific LH assays, GnRH stimulation testing is used rarely.

TESTOSTERONE ASSAYS

Total Testosterone Total testosterone includes both unbound and protein-bound testosterone and is measured by radioimmunoassays, immunometric assays, or liquid chromatography tandem mass spectrometry (LC-MS/MS). LC-MS/MS involves extraction of serum by organic solvents, separation of testosterone from other steroids by high-performance liquid chromatography and mass spectrometry, and quantitation of unique testosterone fragments by mass spectrometry. LC-MS/MS provides accurate and precise measurements of testosterone levels even in the low range and has emerged as the method of choice for testosterone measurement. The use of LC-MS/MS for the measurement of testosterone in laboratories that have been certified by the Centers for Disease Control and Prevention's (CDC) Hormone Standardization Program for Testosterone (HoST) can ensure that testosterone measurements are accurate and calibrated to an international standard. Serum testosterone levels fluctuate because of its pulsatile, diurnal, and circannual secretory rhythms. Testosterone is generally lower in the late afternoon, or after eating a meal, and is reduced by acute illness. The harmonized normal range for total testosterone, measured using LC-MS/MS in nonobese populations of European and American men aged 19–39 years, is 264–916 ng/dL. This harmonized reference range can be applied to values from laboratories that are certified by the CDC's HoST program.

Alterations in SHBG levels due to aging, obesity, diabetes mellitus, hyperthyroidism, some types of medications, chronic illness, or genetic factors can affect total testosterone levels. Heritable factors contribute substantially to the population-level variation in testosterone levels. Genome-wide association studies have revealed polymorphisms in the SHBG gene as well as other loci that are uniquely associated with total and bioavailable testosterone levels, but not with SHBG levels, as important contributors to variation in testosterone levels.

Measurement of Unbound Testosterone Levels Most circulating testosterone is bound to SHBG and to albumin; only 2.0–4.0% of circulating testosterone is unbound, or "free." Free testosterone should ideally be measured by equilibrium dialysis under standardized conditions using an accurate and reliable assay for total testosterone. The unbound testosterone concentration also can be calculated from total testosterone, SHBG, and albumin concentrations. Recent research has shown that testosterone binding to SHBG is a multistep process that involves complex allosteric interactions between the two binding sites within the SHBG dimer; a novel ensemble allosteric model of testosterone's binding to SHBG dimers provides good estimates of free testosterone concentrations. The previous law-of-mass-action equations based on linear models of testosterone binding to SHBG used assumptions that have been shown to be erroneous. Tracer analogue methods are relatively inexpensive and convenient, but they are inaccurate. The term *bioavailable testosterone* refers to unbound testosterone plus testosterone bound loosely to albumin and reflects the concept that albumin-bound testosterone can dissociate at the capillary level, especially in tissues with long transit time, such as the liver and the brain. *Bioavailable testosterone* can be determined by the ammonium sulfate precipitation method. However, the measurements of bioavailable testosterone using the ammonium sulfate precipitation are technically challenging, susceptible to imprecision, and not recommended.

hCG Stimulation Test The hCG stimulation test is performed by administering a single injection of 1500–4000 IU of hCG intramuscularly and measuring testosterone levels at baseline and 24, 48, 72, and 120 h after hCG injection. An alternative regimen involves three injections of 1500 units of hCG on 3 successive days and measuring testosterone levels 24 h after the last dose. An acceptable response to hCG is a doubling of the testosterone concentration in adult men. In prepubertal boys, an increase in testosterone to >150 ng/dL indicates the presence of testicular tissue. No response may indicate an absence of testicular tissue or marked impairment of Leydig cell function. Measurement of Anti-Mullerian hormone, a Sertoli cell product, is also used to detect the presence of testes in prepubertal boys with cryptorchidism.

■ SEMEN ANALYSIS

Semen analysis is the most important step in the evaluation of male infertility. Samples are collected by masturbation following abstinence for 2–3 days. Semen volumes and sperm concentrations vary considerably among fertile men, and several samples may be needed before concluding that the results are abnormal. Analysis should be performed within an hour of collection. Using semen samples from >4500 men in 14 countries, whose partners had a time to pregnancy of <12 months, the World Health Organization (WHO) has generated the following one-sided reference limits for semen parameters: semen volume, 1.5 mL; total sperm number, 39 million per ejaculate; sperm concentration, 15 million per mL; vitality, 58% live; progressive motility, 32%; total (progressive + nonprogressive) motility, 40%; and morphologically normal forms, 4.0%. Some men with low sperm counts are fertile. Some studies suggest that sperm counts have declined in recent decades. A variety of tests for sperm function can be performed in specialized laboratories, but these add relatively little to the treatment options.

■ TESTICULAR BIOPSY

Testicular biopsy is useful in some patients with oligospermia or azoospermia as an aid in diagnosis and in assessing the feasibility of treatment. Fine-needle aspiration biopsy is performed under local anesthesia to aspirate tissue for histology. Alternatively, open biopsies can be performed under local or general anesthesia when more tissue is required. A normal biopsy in an azoospermic man with a normal FSH level suggests obstruction of the vas deferens, which may be correctable surgically. Biopsies are also used to harvest sperm for ICSI and to classify disorders such as hypospermatogenesis (all stages present but in reduced numbers), germ cell arrest (usually at primary spermatocyte stage), and Sertoli cell–only syndrome (absent germ cells) or hyalinization (sclerosis with absent cellular elements).

Testing for Y Chromosome Microdeletions Testing for Y chromosome microdeletions is indicated in infertile men with nonobstructive azoospermia or severe oligozoospermia (sperm density < 5 million/mL). Y chromosome microdeletions are detected by extracting DNA from peripheral blood leukocytes and using polymerase chain reaction (PCR) analysis of a small number of sequence-tagged sites on the Y chromosome (typically sY84 and sY86 in AZFa region, sY127 and SY134 in AZFb region, and sY254 and sY255 in AZFc region). Because these Y chromosome markers account for only a small fraction of the 23 million base pairs on the Y chromosome, a negative test does not exclude microdeletions in other subregions of the Y chromosome.

DISORDERS OF SEXUAL DIFFERENTIATION

See Chap. 402.

DISORDERS OF PUBERTY

The onset and tempo of puberty vary greatly in the general population and are affected by genetic, nutritional, and environmental factors. Although the mean age for reaching Tanner stage 2 in boys is 11.5 years, the range is much wider (9–14 years). A substantial fraction of the variance in the timing of puberty is explained by heritable factors. The timing of pubertal onset is highly correlated within families and between twins. Genome-wide association studies for age of menarche in girls and age of voice deepening in boys have identified nearly 400 independent loci in girls and boys. Over the past century, the age of onset of puberty has fallen by 3 years and this secular trend towards earlier onset of puberty has continued during the past decade, in part due to improved nutrition as well as increasing obesity rates in children.

■ PRECOCIOUS PUBERTY

Precocious puberty in boys is defined by progressive testicular enlargement (>4 mL) before 9 years of age associated with acceleration of linear growth and bone age. Earlier onset of puberty is associated with increased risk of breast and endometrial cancer, cardiovascular disease, hypertension, type 2 diabetes, hair pigmentation, and shorter life span.

Isosexual precocity refers to premature sexual development consistent with phenotypic sex and includes features such as the development of facial hair and phallic growth. Isosexual precocity is divided into gonadotropin-dependent and gonadotropin-independent causes of androgen excess **(Table 403-1)**. *Heterosexual precocity* refers to the premature development of estrogenic features in boys, such as breast development.

TABLE 403-1 Causes of Precocious or Delayed Puberty in Boys

I. Precocious puberty
   A. Gonadotropin-dependent
      1. Idiopathic
      2. CNS tumors such as hypothalamic hamartoma, optic glioma, arachnoid cysts, astrocytoma, ependymoma, or tunerous sclerosis
      3. Inflammatory and infectious lesions
      4. Mutations in genes that regulate GnRH secretion, such as kisspeptin (*KISS1*), kisspeptin receptor (*KISS1R*), and makorin ring finger protein 3 (*MKRN3*)
   B. Gonadotropin-independent
      1. Congenital adrenal hyperplasia
      2. hCG-secreting tumor
      3. McCune-Albright syndrome
      4. Activating LH receptor mutations (familial male-limited precocious puberty)
      5. Exogenous androgens
      6. Androgen producing tumors of the adrenal or the testis
II. Delayed puberty
   A. Constitutional delay of growth and puberty
   B. Systemic disorders
      1. Chronic disease
      2. Malnutrition
      3. Eating disorders
   C. CNS tumors and their treatment (radiotherapy and surgery)
   D. Hypothalamic-pituitary causes of pubertal failure (low gonadotropins)
      1. Congenital disorders associated with GnRH or gonadotropin deficiency (Table 403-2)
      2. Acquired disorders
         a. Pituitary tumors
         b. Hyperprolactinemia
         c. Infiltrative disorders, such as hemochromatosis
   E. Gonadal causes of pubertal failure (elevated gonadotropins)
      1. Klinefelter syndrome
      2. Bilateral undescended testes
      3. Orchitis
      4. Chemotherapy or radiotherapy
      5. Anorchia
   F. Androgen insensitivity

Abbreviations: CNS, central nervous system; GnRH, gonadotropin-releasing hormone; hCG, human chronic gonadotropin; LH, luteinizing hormone.

Gonadotropin-Dependent Precocious Puberty This disorder, also called *central precocious puberty* (CPP), is less common in boys than in girls. It is caused by premature activation of the GnRH pulse generator, sometimes because of central nervous system (CNS) lesions such as hypothalamic hamartomas, but it is often idiopathic. CPP is characterized by gonadotropin levels that are inappropriately elevated for age and LH and FSH responses to GnRH typical of those seen in postpubertal boys or adults. Magnetic resonance imaging (MRI) should be performed to exclude a mass, structural defect, infection, or inflammatory process. Mutations in *MKRN3*, an imprinted gene encoding makorin ring finger protein 3, that serves as an upstream inhibitor of GnRH secretion and is expressed only from the paternally inherited allele, have been associated with CPP. Loss-of-function mutations in *MKRN3* remove the brake that restrains pulsatile GnRH secretion, resulting in precocious puberty. Additional activating mutations in kisspeptin and kisspeptin receptor and loss-of-function mutations in delta-like homolog (*DLK1*) have been identified in patients with CPP.

Gonadotropin-Independent Precocious Puberty In boys with gonadotropin-independent precocious puberty, circulating androgens are increased but gonadotropins are low. This group of disorders includes hCG-secreting tumors; CAH; sex steroid–producing tumors of the testis, adrenal, and ovary; accidental or deliberate exogenous sex steroid administration; hypothyroidism; and activating mutations of the LH receptor or $G_s\alpha$ subunit.

Familial Male-Limited Precocious Puberty Also called *testotoxicosis*, familial male-limited precocious puberty is an autosomal dominant disorder caused by activating mutations in the LH receptor, leading to constitutive stimulation of the cyclic AMP pathway and testosterone production. Clinical features include premature androgenization in boys, linear growth acceleration in early childhood, and advanced bone age followed by premature epiphyseal fusion. Testosterone is elevated, and LH is suppressed. Treatment options include inhibitors of testosterone synthesis (e.g., ketoconazole, abiraterone, medroxyprogesterone acetate), AR antagonists (e.g., flutamide and bicalutamide), and aromatase inhibitors (e.g., anastrozole).

MCCUNE-ALBRIGHT SYNDROME This is a sporadic disorder caused by somatic (postzygotic) activating mutations in the $G_s\alpha$ subunit that links G protein–coupled receptors to intracellular signaling pathways **(Chap. 424)**. The mutations impair the guanosine triphosphatase activity of the $G_s\alpha$ protein, leading to ligand-independent signaling of the G_s-coupled receptor, constitutive activation of adenylyl cyclase, and stimulation of testosterone production. In addition to sexual precocity, affected individuals may have autonomy in the adrenals, pituitary, and thyroid glands. *Café au lait* spots are characteristic skin lesions that reflect the onset of the somatic mutations in melanocytes during embryonic development. Constitutive $G_s\alpha$ activation in the postnatal multipotent skeletal stem cells leads to the formation of immature woven bone and replacement of the bone marrow with fibrotic stroma (polyostotic fibrous dysplasia). Treatment is similar to that in patients with activating LH receptor mutations. Bisphosphonates have been used to treat bone lesions.

CONGENITAL ADRENAL HYPERPLASIA Boys with CAH who are not well controlled with glucocorticoid suppression of adrenocorticotropic hormone (ACTH) can develop premature virilization because of excessive androgen production by the adrenal gland **(Chaps. 398 and 402)**. LH is low, and the testes are small. Some children with CAH, who are poorly controlled with glucocorticoid treatment, may develop gonadotropin-dependent precocious puberty with early maturation of the hypothalamic-pituitary-gonadal axis, elevated gonadotropins, and testicular growth. Similar development of gonadotropin-dependent precocious puberty has been reported in children with androgen-secreting tumors or other conditions associated with androgen excess, likely due to androgen-induced skeletal maturation. Adrenal rests may develop within the testis of poorly controlled patients with CAH because of chronic ACTH stimulation; adrenal rests do not require surgical removal and regress with effective glucocorticoid therapy.

Heterosexual Sexual Precocity Breast enlargement in prepubertal boys can result from familial aromatase excess, estrogen-producing tumors in the adrenal gland, Sertoli cell tumors in the testis, marijuana smoking, or exogenous estrogens or androgens. Occasionally, germ cell tumors that secrete hCG can be associated with breast enlargement due to excessive stimulation of estrogen production (see "Gynecomastia," below).

APPROACH TO THE PATIENT

Precocious Puberty

After verification of precocious development, serum testosterone, LH, and FSH levels should be measured to determine whether gonadotropins are increased in relation to chronologic age (gonadotropin-dependent) or whether sex steroid secretion is occurring independent of LH and FSH (gonadotropin-independent). In children with gonadotropin-dependent precocious puberty, CNS lesions should be excluded by history, neurologic examination, and MRI scan of the head. If an organic cause is not found, one is left with the diagnosis of idiopathic central precocious puberty. Patients with high testosterone but suppressed LH concentrations have gonadotropin-independent precocious puberty; in these patients, DHEA sulfate (DHEAS) and 17α-hydroxyprogesterone should be measured. High levels of testosterone and 17α-hydroxyprogesterone suggest the possibility of CAH due to 21α-hydroxylase or 11β-hydroxylase deficiency. If testosterone and DHEAS are elevated, adrenal tumors should be excluded by obtaining a computed tomography (CT) scan of the adrenal glands. Patients with elevated testosterone but without increased 17α-hydroxyprogesterone or DHEAS should undergo evaluation of the testis by palpation and ultrasound to exclude a Leydig cell neoplasm. Activating mutations of the LH receptor or $G_s\alpha$ subunit should be considered in children with gonadotropin-independent precocious puberty in whom CAH, androgen abuse, and adrenal and testicular neoplasms have been excluded.

TREATMENT

Precocious Puberty

In patients with a known cause (e.g., a CNS lesion or a testicular tumor), therapy should be directed toward the underlying disorder. In patients with idiopathic CPP, treatment with long-acting GnRH analogues is indicated in boys showing rapid pubertal progression, who are more advanced in pubertal development (e.g., Tanner stage 3 or greater genital development) and experiencing rapid linear growth or psychological distress. GnRH analogues suppress gonadotropins and testosterone, halt early pubertal development, delay accelerated bone maturation, prevent early epiphyseal closure, promote final height gain, and mitigate the psychosocial consequences of early pubertal development without causing osteoporosis. The treatment is most effective for increasing final adult height if it is initiated before age 6. Puberty resumes after discontinuation of the GnRH analogue. Counseling is an important aspect of the overall treatment strategy.

In children with gonadotropin-independent precocious puberty, inhibitors of steroidogenesis, such as ketoconazole or abiraterone, AR antagonists, and aromatase inhibitors have been used empirically. Long-term treatment with spironolactone (a weak androgen antagonist) and ketoconazole also has been reported to normalize growth rate and bone maturation and to improve predicted height in small, nonrandomized trials in boys with familial male-limited precocious puberty. Aromatase inhibitors, such as testolactone and letrozole, have been used as adjuncts to antiandrogen therapy for children with familial male-limited precocious puberty, CAH, and McCune-Albright syndrome. More potent novel inhibitors of testosterone synthesis, such as abiraterone, have not been systematically evaluated in boys with gonadotropin-independent precocious puberty.

DELAYED PUBERTY

Puberty is considered delayed in boys if it has not ensued by age 14, an age that is 2–2.5 standard deviations above the mean for healthy children. Pubertal delay is not necessarily pathologic and may be a variant of normal pubertal development in some children. Delayed puberty has been associated with lower peak bone mass, higher risk for metabolic and cardiovascular disorders, and lower risk for breast and endometrial cancer in women.

Delayed puberty is more common in boys than in girls. There are four main categories of delayed puberty: (1) self-limited delayed puberty (previously called constitutional delay of puberty) (~60% of cases); (2) functional hypogonadotropic hypogonadism caused by systemic illness, malnutrition, or eating disorders (~20% of cases); (3) hypogonadotropic hypogonadism caused by congenital or acquired defects in the hypothalamic-pituitary region (~10% of cases); and (4) hypogonadism secondary to primary gonadal failure (~15% of cases) (Table 403-1). The self-limited delayed puberty is the most common cause that accounts for nearly two-thirds of boys and one-third of girls with delayed puberty. The self-limited puberty clusters in families and displays a complex inheritance pattern, having an autosomal dominant pattern of inheritance in some families, but autosomal recessive, X-linked, or bilineal pattern in other families. Self-limited delayed puberty and congenital hypogonadotropic hypogonadism share similarities in their clinical presentation but differ in their clinical course and treatment. In boys with self-limited delayed puberty, pubertal progression will occur eventually, while in those with congenital hypogonadotropin hypogonadism, hormone treatment will be needed to induce secondary sex characteristics. Genetic studies of patients with congenital hypogonadotropic hypogonadism have identified >65 genes that are involved in regulating one or more of the following: (1) development and migration of GnRH neurons (e.g., *ANOS1, PROK2, PROK2R, FGFR1, FGF8, FGF17, WDR11,* and *CHD7*); (2) regulation of GnRH secretion (*KISS1, KISS1R, TAC3,* and *TACR3*); and (3) GnRH function (e.g., *GNRH1, RNRHR*). The mutations in some of these genes are associated with dysmorphic features or occur as a part of a syndromic constellation that can facilitate their identification. Some of the genes associated with self-limited delayed puberty are found uniquely in kindreds of children with this disorder, while others are similar to those found in families with congenital hypogonadotropic hypogonadism. Furthermore, 10 to 20% of children with congenital hypogonadotropic hypogonadism may experience spontaneous restoration of GnRH and gonadotropin secretion in adulthood.

Functional hypogonadotropic hypogonadism is more common in girls than in boys. Permanent causes of congenital hypogonadotropic hypogonadism or primary testicular failure are identified in <25% of boys with delayed puberty.

APPROACH TO THE PATIENT

Delayed Puberty

History of systemic illness, eating disorders, excessive exercise, social and psychological problems, and abnormal patterns of linear growth during childhood should be verified. Boys with pubertal delay may have accompanying emotional and physical immaturity relative to their peers, which can be a source of anxiety. Physical examination should focus on height; arm span; weight; visual fields; and secondary sex characteristics, including hair growth, testicular volume, phallic size, and scrotal reddening and thinning. Testicular size >2.5 cm generally indicates that the child has entered puberty.

The main diagnostic challenge is to distinguish those with self-limited delayed puberty, who will progress through puberty at a later age, from those with congenital hypogonadotropic hypogonadism as both conditions present with low testosterone and low gonadotropin levels, delayed bone age and short stature. Pituitary priming by pulsatile GnRH is required before LH and FSH are synthesized and secreted normally. Thus, blunted responses to exogenous GnRH can be seen in patients with self-limited delayed puberty, congenital hypogonadotropic hypogonadism, or pituitary disorders. On the other hand, low-normal basal gonadotropin levels or a normal response to exogenous GnRH is consistent with an early stage of puberty, which is often heralded by nocturnal GnRH secretion. The presence of micropenis or history of cryptorchidism are more suggestive of congenital hypogononadotropic hypogonadism. However, self-limited delayed puberty remains a diagnosis of exclusion that requires ongoing evaluation until the onset of puberty and the growth spurt.

TREATMENT

Delayed Puberty

A major challenge in the treatment of children with delayed puberty is to determine whether and when to initiate hormonal treatment. Reassurance without hormonal treatment is appropriate for many individuals with presumed self-limited delayed puberty. However, the impact of delayed growth and pubertal progression on a child's social relationships and school performance should be weighed.

Boys with self-limited delayed puberty are less likely to achieve their full genetic height potential and have reduced total body bone mass as adults, mainly due to narrow limb bones and vertebrae as a result of impaired periosteal expansion during puberty. Furthermore, the time of onset of puberty is negatively associated with bone mineral content and density in boys at skeletal maturity. Judicious use of testosterone therapy in carefully selected boys with self-limited delayed puberty can induce pubertal induction and progression and promote short-term growth without compromising final height, and when administered with an aromatase inhibitor, it may improve final height.

If therapy is considered appropriate, it can begin with 25–50 mg of testosterone enanthate or testosterone cypionate every 2–4 weeks or by using a 2.5-mg testosterone patch or 25-mg testosterone gel. Because aromatization of testosterone to estrogen is obligatory for mediating androgen effects on epiphyseal fusion, concomitant treatment with aromatase inhibitors may allow attainment of greater final adult height. Testosterone treatment should be interrupted after 6 months to determine if endogenous LH and FSH secretion have ensued. Other causes of delayed puberty should be considered when there are associated clinical features or when boys do not enter puberty spontaneously after a year of observation or treatment.

Boys with confirmed congenital hypogonadotropic hypogonadism require an earlier institution of hormonal treatment than those with self-limited delayed puberty. Optimization of future fertility potential may require initial treatment with hCG to induce development of secondary sex characteristics as well as spermatogenesis to enable harvesting and cryopreservation of sperm.

DISORDERS OF THE MALE REPRODUCTIVE AXIS DURING ADULTHOOD

HYPOGONADOTROPIC HYPOGONADISM

Because LH and FSH are trophic hormones for the testes, impaired secretion of these pituitary gonadotropins results in secondary hypogonadism, which is characterized by low testosterone in the setting of low or inappropriately normal LH and FSH. Hypogonadotropic hypogonadism can be classified into congenital and acquired disorders. Congenital disorders most commonly involve GnRH deficiency, which leads to gonadotropin deficiency. Acquired disorders are more common than congenital disorders and may result from sellar mass lesions, infiltrative diseases of the hypothalamus or pituitary, drugs, nutritional or psychiatric disorders, or systemic diseases. Those with the most severe congenital gonadotropin deficiency have complete absence of pubertal development, and, in some cases, hypospadias, undescended testes, and micropenis. Patients with congenital or prepubertal onset of partial gonadotropin deficiency have delayed or arrested sex development. The 24-h LH secretory profiles are heterogeneous in patients with hypogonadotropic hypogonadism, reflecting variable abnormalities of LH pulse frequency or amplitude. In severe cases, basal LH is low, and there are no LH pulses. A smaller subset of patients has low-amplitude LH pulses or markedly reduced pulse frequency. Occasionally, only sleep-entrained LH pulses occur, reminiscent of the pattern seen in the early stages of puberty.

Congenital Disorders Associated with Gonadotropin Deficiency (See Chap. 391) Congenital hypogonadotropic hypogonadism is a heterogeneous group of disorders characterized by decreased gonadotropin secretion and testicular dysfunction either due to impaired function of the GnRH pulse generator or the gonadotrope. The disorders characterized by GnRH deficiency represent a family of monogenic or oligogenic disorders whose phenotype spans a wide spectrum. Some individuals with GnRH deficiency may suffer from complete absence of pubertal development, while others may manifest varying degrees of gonadotropin deficiency and pubertal delay, and a subset that carries the same mutations as their affected family members may even have normal reproductive function. In ~10% of men with congenital isolated hypogonadotropic hypogonadism (IHH), reversal of gonadotropin deficiency may occur in adult life after sex steroid therapy. Also, a small fraction of men with IHH may present with testosterone deficiency and infertility in adult life after having gone through apparently normal pubertal development. Nutritional, emotional, or metabolic stress may unmask gonadotropin deficiency and reproductive dysfunction (e.g., hypothalamic amenorrhea) in some patients who harbor mutations in the candidate genes but who previously had normal reproductive function. Oligogenicity and gene-gene and gene-environment interactions may contribute to variations in clinical phenotype.

Mutations in >60 genes involved in the development and migration of GnRH neurons (e.g., *ANOS1, PROK2, PROK2R, FGFR1, FGF8, FGF17, WDR11, CHD7*), GnRH secretion (*KISS1, KISS1R, TAC3, TACR3*), or GnRH function (e.g., *GNRH1, RNRHR*) have been linked to congenital IHH. The genetic defect remains elusive in nearly two-thirds of cases. Familial hypogonadotropic hypogonadism can be transmitted as an X-linked (20%), autosomal recessive (30%), or autosomal dominant (50%) trait. Some individuals with IHH have sporadic mutations in the same genes that cause inherited forms of the disorder. The genetic defects associated with GnRH deficiency can be classified as anosmic (Kallmann syndrome) or normosmic (Table 403-2), although the occurrence of both anosmic and normosmic forms of GnRH deficiency in the same families suggests commonality of pathophysiologic mechanisms. *Kallmann syndrome*, the anosmic form

TABLE 403-2 Causes of Congenital Hypogonadotropic Hypogonadism

GENE	LOCUS	INHERITANCE	ASSOCIATED FEATURES
A. Hypogonadotropic Hypogonadism due to GnRH Deficiency			
A1. GnRH Deficiency Commonly Associated with Hyposmia or Anosmia			
ANOS1 (KAL1)	Xp22	X-linked	Anosmia, renal agenesis, synkinesia, cleft lip/palate, oculomotor/visuospatial defects, gut malformations
NELF (NCMF)	9q34.3	AR	Anosmia, hypogonadotropic hypogonadism (some may be normosmic)
FGF8	10q24	AR	Anosmia (some patients may be normosmic), skeletal abnormalities
FGF17	8p21.3	AR	Anosmia (some patients may be normosmic)
FGFR1	8p11-p12	AD	Anosmia, cleft lip/palate, synkinesia, syndactyly
PROK2	3p21	AR	Anosmia/sleep dysregulation
PROKR2	20p12.3	AR	Variable
CHD7	8q12.1		Anosmia, other features of CHARGE syndrome
FEZ1	8p22	AR	Anosmia, olfactory bulb aplasia
WDR11	10q26	AD	Anosmia
SOX10	22q13		Deafness (Wardenburg Syndrome)
HS6ST1	2q14	Complex	Anosmia

(*Continued*)

TABLE 403-2 Causes of Congenital Hypogonadotropic Hypogonadism (Continued)

GENE	LOCUS	INHERITANCE	ASSOCIATED FEATURES
TUBB3	Tubulin β 3	AR	Anosmia
DUSP6	12q21.33	AR	Anosmia
GLCE	15q23	AR	Anosmia (some patients may be normosmic)
FLRT3	20p12.1	AR	Anosmia (some patients may be normosmic)
SPRY4	5q31.3	AR	Anosmia (some patients may be normosmic)
IL17RD	3p14.3	AR	Anosmia
SEMA3A	7q21.11		Anosmia
SEMA7A	15q24.1		Anosmia as well as normosmia
SEMA3E	7q21.11		Anosmia
PLXNA1	3q21.3		Anosmia as well as normosmia
DCC	18q21.2	AR	Anosmia
NTN1	17p13.1	AR	Anosmia
KLB	4p14	AR	Metabolic disorders
A2. GnRH Deficiency with Normal Sense of Smell			
GNRHR	4q21	AR	None
GnRH1	8p21	AR	None
KISS1	1q32.1	AR	None
KISS1R	19p13	AR	None
TAC3	12q13	AR	Microphallus, cryptorchidism, reversal of GnRH deficiency
TAC3R	4q25	AR	Microphallus, cryptorchidism, reversal of GnRH deficiency
LEPR	1p31	AR	Obesity
LEP	7q31	AR	Obesity
DMXL2	15q21.2	AR	Polyendocrine polyneuropathy syndrome
OTUD4	4q31.21	AR	Ataxia (Gordon Holmes syndrome)
RNF216	7p22.1	AR	Ataxia (Gordon Holmes syndrome)
STUB1	16p13.3	AR	Ataxia
POLR3B	12q23.3	AR	Ataxia (hypomyelinating leukodystrophy-8 with hypogonadotropic hypogonadism [4H] syndrome)
PNPLA6	19p13.2	AR	Ataxia (Laurence-Moon syndrome)
NR0B1 (Dax1)	Xp21.2	X-linked	Primary adrenal failure
CCDC141	2q31.2	AR	Normosmia
IGSF10	3q25.1	AR	Reversible and normosmic form of hypogonadotropic hypogonadism
RNF216	7p22.1	AR	Atxia, dementia, and hypogonadotropic hypogonadism (Gordon Holmes syndrome)
B. Hypogonadotropic Hypogonadism Not Due to GnRH Deficiency			
PCSK1	5q15-21	AR	Obesity, diabetes mellitus, ACTH deficiency
HESX1	3p21	AR	Septo-optic dysplasia, CPHD
LHX3	9q34	AR	CPHD (ACTH spared), cervical spine rigidity
PROP1	5q35	AR	CPHD (ACTH usually spared)
FSHβ	11p13	AR	↑ LH
LHβ	19q13	AR	↑ FSH
SF1 (*NR5A1*)	9p33	AD/AR	Primary adrenal failure, XY sex reversal

Abbreviations: ACTH, adrenocorticotropic hormone; AD, autosomal dominant; AR, autosomal recessive; *CCDC141*, coiled-coil domain containing 141; CHARGE syndrome, eye coloboma, heart defects, choanal atresia, growth and developmental retardation, genitourinary anomalies, ear anomalies; CPHD, combined pituitary hormone deficiency; *DAX1*, dosage-sensitive sex-reversal, adrenal hypoplasia congenita, X chromosome; *DCC*, deleted in colon cancer; *DMXL2*, DMX like 2; *DUSP6*, dual specificity phosphatase 6; *FGFR1*, fibroblast growth factor receptor 1; *FGF17*, fibroblast growth factor 17; *FSHβ*, follicle-stimulating hormone β-subunit; *FLRT3*, fibronectin like domain containing leucine rich transmembrane protein 3; GH, growth hormone; *GLCE,* glucuronic acid epimerase; *GNRHR*, gonadotropin-releasing hormone receptor; *GPR54*, G protein–coupled receptor 54; *HESX1*, homeobox gene expressed in embryonic stem cells 1; *IGSF10,* immunoglobulin superfamily member 10; *KAL1*, Kallmann syndrome interval gene 1, also known as anosmin 1; *KISS1*, kisspeptin 1; *KLB,* klotho-1; *LEP*, leptin; *LEPR*, leptin receptor; *LHX3,* LIM homeobox gene 3; *LHβ*, luteinizing hormone β-subunit; *NELF*, nasal embryonic luteinizing hormone–releasing hormone factor; *NSMF*, NMDA receptor synaptonuclear signaling and neuronal migration factor; *NR0B1*, nuclear receptor subfamily 0, group B, member 1; *NTN1*, netrin-1; *OTUD4*, OUT domain containing protein 4; *PCSK1*, proprotein convertase subtilisin/kexin type 1; *PLXNA1, plexin A1; PNPLA6*, patatin-like phospholipase domain-containing protein 6; *PC1*, prohormone convertase 1; *PROK2*, prokineticin 2; *PROKR2*, prokineticin 2 receptor; *PROP1*, prophet of pit 1; *RNF216*, ring finger protein 216; *POLR3B*, polymerase III RNA subunit B; *SEMA3A, semaphorin 3A; SEMA7A, semaphorin 7A; SEMA3E, semoaphorin 3E; SF1*, steroidogenic factor 1; *SPRY4*, sprouty RTK signaling antagonist 4; *STUB1*, srip 1 homologous and U box containing protein 1; *TUBB3*, tubulin beta 3; *IL17RD*, interleukin 17 receptor D.

of GnRH deficiency, can result from mutations in one or more neurodevelopmental genes associated with olfactory bulb morphogenesis or the migration of GnRH neurons from their origin in the region of the olfactory placode, along the scaffold established by the olfactory nerves, through the cribriform plate into their final location into the preoptic region of the hypothalamus. Thus, mutations in *KAL1*, NMDA receptor synaptonuclear signaling and neuronal migration factor (*NSMF*), genes involved in fibroblast growth factor (FGF) signaling (*FGF8*, *FGFR1*, *FGF17*, *IL17RD*, *DUSP6*, *SPRY4*, and *FLRT3*), *NELF*, genes involved in *PROK* signaling (*PROK2* and *PROK2R*), *WDR11*, *SOX10*, *TUBB3 SEMA3*, *HS6ST1*, *CHD7*, and *FEZF1* have been described in patients with Kallmann syndrome; among these mutations in *ANOS1*, *CHD7*, *FGF8*, *FGFR1*, *PRROK2*, and *PROLR2* are the most common. An X-linked form of IHH is caused by mutations in the *ANOS1* gene, which encodes anosmin, a protein that mediates the migration of neural progenitors of the olfactory bulb and GnRH-producing neurons. These individuals have GnRH deficiency and variable combinations of anosmia or hyposmia. Proteins such as those involved in FGF and prokineticin signaling and *KAL1*, which account for the great majority of Kallmann syndrome cases, interact with heparin sulfate glycosaminoglycan compounds within the extracellular matrix in supporting GnRH neuronal migration. Mutations in the *FGFR1* gene cause an autosomal dominant form of hypogonadotropic hypogonadism that clinically resembles Kallmann syndrome; mutations in its putative ligand, the *FGF8* gene product, have also been associated with IHH. Craniofacial tissues and olfactory ensheathing cells also play important roles in neurogenesis and migration of the GnRH neurons, and additional proteins that regulate these cell types may also be involved in the pathogenesis of Kallmann syndrome. The co-occurrence of tooth anomalies, cleft palate, craniofacial anomalies, pigmentation, and neurologic defects in patients with Kallmann syndrome suggests that the syndrome may be a part of the spectrum of neurocristopathies. Other dysmorphic features associated with some forms of IHH include renal agenesis, hearing loss, synkinesia, short metacarpals, eye movement abnormalities, cerebellar ataxia, and dental agenesis. The presence of these dysmorphic features can offer clues to the underlying genetic abnormality and guide genetic testing.

Normosmic GnRH deficiency results from defects in pulsatile GnRH secretion, its regulation, or its action on the gonadotrope and has been associated with mutations in *GnRHR, GNRH1, KISS1R, TAC3, TACR3, NROB1* (*DAX1*), leptin, or leptin receptor. Some mutations, such as those in *PROK2, PROKR2, NSMF, FGFR1, FGF8, SEMA3A, WDR11*, and *CHD7*, have been associated with both anosmic and normosmic forms of IHH; it is possible that these genes are involved in GnRH neuronal migration as well in regulation of GnRH secretion. *GnRHR* mutations, the most frequent identifiable cause of normosmic IHH, account for ~40% of autosomal recessive and 10% of sporadic cases of hypogonadotropic hypogonadism. These patients have decreased LH response to exogenous GnRH. Some receptor mutations alter GnRH binding affinity, allowing apparently normal responses to pharmacologic doses of exogenous GnRH, whereas other mutations may alter signal transduction downstream of hormone binding. Mutations of the *GnRH1* gene have also been reported in patients with hypogonadotropic hypogonadism, although they are rare. The G protein–coupled receptor *KISS1R* (*GPR54*) and its cognate ligand, kisspeptin (*KISS1*), are important regulators of sexual maturation in primates. Recessive mutations in *GPR54* cause gonadotropin deficiency without anosmia. The genes encoding NKB (*TAC3*), which is involved in preferential activation of GnRH release in early development, and its receptor (*TAC3R*) have been implicated in some families with normosmic IHH. The tachykinin pathway plays an important role in GnRH activation during "mini-puberty" as well as in puberty. Prokineticin 2 (PROK2) and its receptor (PROK2R) are highly expressed in the olfactory ventricle and subventricular zone of the lateral ventricle and are associated with neurogenesis of the olfactory bulbs and the migration of the olfactory neuronal cells. Mutations in the *CHD7* gene that encodes for the chromodomain helicase DNA binding protein 7 causes CHARGE syndrome characterized by eye coloboma, heart anomalies, choanal atresia, growth and developmental retardation, genitourinary anomalies, hypogonadism, and ear abnormalities. X-linked hypogonadotropic hypogonadism also occurs in *adrenal hypoplasia congenita*, a disorder caused by mutations in the *DAX1* gene, which encodes a nuclear receptor in the adrenal gland and reproductive axis. Adrenal hypoplasia congenita is characterized by absent development of the adult zone of the adrenal cortex, leading to neonatal adrenal insufficiency. Puberty usually does not occur or is arrested, reflecting variable degrees of gonadotropin deficiency. Although sexual differentiation is normal, some patients have testicular dysgenesis and impaired spermatogenesis despite gonadotropin replacement. Less commonly, adrenal hypoplasia congenita, sex reversal, and hypogonadotropic hypogonadism can be caused by mutations of steroidogenic factor 1 (SF1). Rarely, recessive mutations in the *LHβ* or *FSHβ* genes have been described in patients with selective deficiencies of these gonadotropins.

A number of homeodomain transcription factors are involved in the development and differentiation of the specialized hormone-producing cells within the pituitary gland (Table 403-2). Patients with mutations of *PROP1* have combined pituitary hormone deficiency that includes GH, prolactin (PRL), thyroid-stimulating hormone (TSH), LH, and FSH, but not ACTH. *LHX3* mutations cause combined pituitary hormone deficiency in association with cervical spine rigidity. *HESX1* mutations cause septo-optic dysplasia and combined pituitary hormone deficiency. Mutations of *ARNT1*, inherited as an autosomal recessive disorder, are associated with diabetes insipidus; ACTH, GH, LH, and FSH deficiency; anterior pituitary hypoplasia; hypoplastic frontal and temporal lobes; thin corpus callosum; prominent forehead; and retrognathia. Patients with *SOX2* mutations can have gonadotropin deficiency, variable deficiencies of TSH and ACTH, pituitary hypoplasia, microphthalmia, and intellectual disability.

Prader-Willi syndrome (PWS) is a genomic imprinting disorder caused by deletions of the proximal portion of the paternally derived chromosome 15q11-15q13 region, which contains a bipartite imprinting center (65–75%); uniparental disomy of the maternal alleles (20–30%); or mutations of the genes/loci involved in imprinting (1–3%) **(Chap. 479)**. The imprinting gene implicated in PWS is a ubiquitin pathway gene known as *UBE3A*, for which the maternal allele is normally expressed while the paternal allele is silenced. PWS is characterized by obesity, hypotonic musculature, intellectual disability, hypogonadism, short stature, and small hands and feet. Some of the major manifestations of PWS may be due to reduced expression of prohormone convertase 1.

Laurence-Moon syndrome is an autosomal recessive disorder characterized by obesity, hypogonadism, mental retardation, polydactyly, and retinitis pigmentosa. Recessive mutations of leptin, or its receptor, cause severe obesity and pubertal arrest, apparently because of hypothalamic GnRH deficiency **(Chap. 413)**.

Acquired Hypogonadotropic Disorders • **SEVERE ILLNESS, STRESS, MALNUTRITION, AND EXERCISE** These may cause reversible gonadotropin deficiency. Although gonadotropin deficiency and reproductive dysfunction are well documented in these conditions in women, men exhibit similar but less pronounced responses. Unlike women, most male runners and other endurance athletes have normal gonadotropin and sex steroid levels, despite low body fat and frequent intensive exercise. Testosterone levels fall at the onset of illness and recover during recuperation. The magnitude of gonadotropin suppression generally correlates with the severity of illness. Although hypogonadotropic hypogonadism is the most common cause of testosterone deficiency in patients with acute illness, some have elevated levels of LH and FSH, which suggests primary gonadal dysfunction. The pathophysiology of reproductive dysfunction during acute illness is unknown but likely involves the combined effects of reductive stress, inflammation, cytokines, and/or glucocorticoid effects. There is a high frequency of low testosterone levels in patients with chronic illnesses such as HIV infection, end-stage renal disease, chronic obstructive lung disease, and many types of cancer and in patients receiving glucocorticoids. About 20% of HIV-infected men with low testosterone levels have elevated LH and FSH levels; these patients presumably have primary testicular dysfunction. The remaining 80% have either normal or low LH and FSH levels; these men have a central hypothalamic-pituitary defect or a dual defect involving both the testis and the hypothalamic-pituitary centers. Muscle wasting is common in chronic diseases associated with hypogonadism, which also leads to debility, poor quality of life, and adverse outcome of disease.

Men using opioids for relief of cancer or noncancerous pain or because of addiction often have suppressed testosterone and LH levels and high prevalence of sexual dysfunction and osteoporosis; the degree of suppression is dose-related and particularly severe with long-acting opioids such as methadone. Exogenous opioids bind to the κ-opioid receptors on the KNDy neurons, suppress GnRH secretion, and alter the sensitivity to feedback inhibition by gonadal steroids. Heavy marijuana use has been associated with reduced sperm density, motility, morphology, and capacitation, and increased risk of subfertility.

The effects of marijuana use on circulating testosterone levels have been inconsistent in humans. Gynecomastia observed in marijuana users can be caused by plant estrogens in crude preparations. Androgen deprivation therapy in men with prostate cancer has been associated with increased risk of bone fractures, diabetes mellitus, cardiovascular events, fatigue, sexual dysfunction, tender gynecomastia, and poor quality of life.

OBESITY In men with mild to moderate obesity, SHBG levels decrease in proportion to the degree of obesity, resulting in lower total testosterone levels. However, free testosterone levels usually remain within the normal range. SHBG production in the liver is inhibited by hepatic lipids and by tumor necrosis factor α and interleukin 1, but it is not affected by insulin. Thus, the low SHBG levels seen in obesity and diabetes are likely the result of low-grade inflammation and the increased amount of hepatic lipids rather than high insulin levels. Inflammation and gliosis in the pituitary, hypothalamic leptin resistance, and increased estradiol levels because of aromatization of testosterone to estradiol in adipose tissue may impair hypothalamic-pituitary function; obesity also may directly affect testicular function. Weight gain in adult men can accelerate the rate of age-related decline in testosterone levels. Weight loss is associated with reversal of these abnormalities including an increase in total and free testosterone levels and a decrease in estradiol levels.

HYPERPROLACTINEMIA **(See also Chap. 392)** Elevated PRL levels are associated with hypogonadotropic hypogonadism. PRL inhibits hypothalamic GnRH secretion either directly or through modulation of tuberoinfundibular dopaminergic pathways and also attenuates LH and FSH response to GnRH. A PRL-secreting tumor may also destroy the surrounding gonadotropes by invasion or compression of the pituitary stalk. Treatment with dopamine agonists reverses gonadotropin deficiency, although there may be a delay relative to PRL suppression.

SELLAR MASS LESIONS Neoplastic and nonneoplastic lesions in the hypothalamus or pituitary can directly or indirectly affect gonadotrope function. In adults, pituitary adenomas constitute the largest category of space-occupying lesions affecting gonadotropin and other pituitary hormone production. Pituitary adenomas that extend into the suprasellar region can impair GnRH secretion and mildly increase PRL secretion (usually <50 μg/L) because of impaired tonic inhibition by dopaminergic pathways. These tumors that cause hyperprolactinemia by stalk compression should be distinguished from prolactinomas, which typically are associated with higher PRL levels. The presence of diabetes insipidus suggests the possibility of a craniopharyngioma, infiltrative disorder, or other hypothalamic lesions **(Chap. 393)**.

HEMOCHROMATOSIS **(See also Chap. 426)** Both the pituitary and testis can be affected by excessive iron deposition. However, the pituitary defect is the predominant lesion in most patients with hemochromatosis and hypogonadism. The diagnosis of hemochromatosis is suggested by the association of characteristic skin discoloration, hepatic enlargement or dysfunction, diabetes mellitus, arthritis, cardiac conduction defects, hypogonadism, and increased serum ferritin (>300 ng/mL for men) and transferrin saturation (>45%, threshold varies in different guidelines).

■ PRIMARY TESTICULAR CAUSES OF HYPOGONADISM

Common causes of primary testicular dysfunction include Klinefelter syndrome, uncorrected cryptorchidism, cancer chemotherapy, radiation to the testes, trauma, torsion, infectious orchitis, HIV infection, anorchia syndrome, and myotonic dystrophy. Primary testicular disorders may be associated with impaired spermatogenesis, decreased androgen production, or both. **See Chap. 402 for disorders of testis development, androgen synthesis, and androgen action.**

Klinefelter Syndrome **(See also Chap. 402)** Klinefelter syndrome is the most common chromosomal disorder associated with 47,XXY karyotype, testicular dysfunction, and male infertility. It occurs in about 1 in 600 live-born males. Azoospermia is the rule in men with Klinefelter syndrome who have the 47,XXY karyotype due to the progressive loss of 47,XXY spermatogonial stem cells; however, men with mosaicism (46,XY/47,XXY) and even some with 47,XXY karyotype may have germ cells, especially at a younger age. The clinical phenotype of Klinefelter syndrome can be variable, possibly because of mosaicism, polymorphisms in AR gene, the parental origin of the X chromosome, X-linked copy number variations, gene-dosage effects in conjunction with X chromosome inactivation, variable testosterone levels, and variable changes in the transcriptome and methylome in various tissues. Testicular histology shows hyalinization of seminiferous tubules and germ cell aplasia. However, spermatogenesis can be observed in a small number of tubules from which sperm can be harvested during testicular sperm extraction for IVF. Although their function is impaired, the number of Leydig cells appears to increase. Testosterone is decreased and estradiol is increased, leading to clinical features of undervirilization and gynecomastia. Men with Klinefelter syndrome are at increased risk of systemic lupus erythematosus, Sjögren's syndrome, breast cancer, diabetes mellitus, osteoporosis, non-Hodgkin's lymphoma, and some types of lung cancers and at reduced risk of prostate cancer. Periodic mammography for breast cancer surveillance is recommended for men with Klinefelter syndrome. Fertility can be achieved by intracytoplasmic injection of sperm retrieved surgically from the testes of men with Klinefelter syndrome, including some men with nonmosaic form of Klinefelter syndrome. Although sperm retrieval in adolescence for fertility preservation offers no benefit over harvesting in adulthood, fertility counseling, including the potential for sperm retrieval, should be offered prior to starting testosterone replacement therapy. The karyotypes 48,XXXY and 49,XXXXY are associated with a more severe phenotype, increased risk of congenital malformations, and lower intelligence than 47,XXY individuals.

Cryptorchidism Cryptorchidism occurs when there is incomplete descent of the testis from the abdominal cavity into the scrotum. About 1–4% of full-term and 30% of premature male infants have at least one undescended testis at birth, but descent is usually complete by the first few weeks of life. Fifty percent of undescended testes at birth will descend spontaneously within the first 6–18 months of life. Consequently, the incidence of cryptorchidism is <1% by 9 months of age. Cryptorchidism should be distinguished from retractile testes that can be pulled down into the scrotum during physical examination and require no treatment.

Testosterone and insulin-like 3 (*INSL3*) and their cognate receptors, androgen receptor and relaxin/insulin-like family peptide receptor 2 (*RXFP2*), regulate the development and growth of the fetal gubernaculum-cremaster muscle complex and the inguinoscrotal descent of the testes. Mutations in *INSL3* and *RXFP2* have been found in some patients with nonsyndromic cryptorchidism. However, genome-wide association studies have not found a consistent relation of nonsyndromic cryptorchidism with any single locus, suggesting complex, multilocus genetic susceptibility.

Cryptorchidism is associated with increased risk of malignancy, infertility, inguinal hernia, and torsion. Unilateral cryptorchidism, even when corrected before puberty, is associated with decreased sperm count, possibly reflecting unrecognized damage to the fully descended testis or other genetic factors. Surgical correction is usually performed between 6 and 18 months of age depending on the location of the testes, the child's body size, and parental preference. Epidemiologic, clinical, and molecular evidence supports the idea that cryptorchidism, hypospadias, impaired spermatogenesis, and testicular cancer may be causally related to common genetic and environment perturbations and are components of the testicular dysgenesis syndrome.

Acquired Testicular Defects *Viral orchitis* may be caused by the mumps virus, echovirus, lymphocytic choriomeningitis virus, and group B arboviruses. Orchitis occurs in as many as one-fourth of adult men with mumps; the orchitis is unilateral in about two-thirds and bilateral in the remainder. Orchitis usually develops a few days after the onset of parotitis but may precede it. The testis may return to normal size and function or undergo atrophy. Semen analysis returns to normal for three-fourths of men with unilateral involvement but for only one-third of men with bilateral orchitis. *Trauma*, including testicular torsion, can also cause secondary atrophy of the testes. The exposed position of the testes in the scrotum renders them susceptible to both thermal and physical trauma, particularly in men with hazardous occupations.

The late-term adverse effects of cancer treatment on reproductive health have emerged as an important concern among cancer survivors. Many professional cancer societies have published guidelines on the fertility preservation in patients with cancer and strongly endorsed consideration of fertility preservation measures prior to initiation of cancer treatment. The testes are sensitive to *radiation damage* due to the direct effects of ionized radioactive particles as well as indirect effects of free radicals generated from water. Radiation doses <1.0 Gy are associated with only a transient decline in sperm density; doses between 1.0 and 2.0 Gy are associated with temporary azoospermia, and doses >2.0 Gy are generally associated with permanent azoospermia. Leydig cells are relatively resistant to the effects of radiation and only doses >20 Gy are associated with Leydig cell dysfunction and increased FSH and LH levels. Permanent testosterone deficiency in adult men is uncommon after therapeutic radiation; however, most boys given direct testicular radiation therapy for acute lymphoblastic leukemia have permanently low testosterone levels. Direct testicular radiation and whole-body radiation before bone marrow transplantation pose the greatest risk of permanent testicular damage.

Combination chemotherapy for acute leukemia, Hodgkin's disease, and testicular and other cancers may impair Leydig cell function and cause infertility. The degree of gonadal dysfunction depends on the type of chemotherapeutic agent and the dose and duration of therapy. Because of the high response rates and the young age of these men, infertility and testosterone deficiency have emerged as important long-term complications of cancer chemotherapy. Cyclophosphamide, procarbazine, ifosfamide, busulfan, chlorambucil, chlormethine cisplatin, carboplatin, and doxorubicin are associated with moderate to high risk of infertility. Azoospermia is uncommon with cyclophosphamide equivalent doses of <4000 mg/m^2, and higher doses of alkylating agents are generally associated with varying degree of damage to germ cells. The outcomes of assisted reproductive technologies in cancer survivors using cryopreserved sperm collected prior to cancer treatment are similar to those in infertile men who do not have cancer. A smaller subset of cancer survivors treated with large doses of alkylating agents may also suffer from testosterone deficiency and sexual dysfunction.

Drugs interfere with testicular function by several mechanisms, including inhibition of testosterone synthesis (e.g., ketoconazole, abiraterone, GnRH agonists and antagonists), blockade of androgen action (e.g., spironolactone), increased estrogen (e.g., marijuana), and toxic effects on spermatogenesis (e.g., chemotherapy).

Alcohol, when consumed in small to moderate amounts, modestly increases testosterone levels, but ingestion of large quantities decreases testosterone, independent of liver disease or malnutrition. Chronic ethanol use in large amounts lowers testosterone levels by inducing inflammation and reductive stress and by its effects on the hypothalamus, pituitary, and liver.

The occupational and recreational history should be evaluated in all men with infertility because of the toxic effects of many *chemical agents* on spermatogenesis. Known environmental hazards include pesticides (e.g., vinclozolin, dicofol, atrazine), sewage contaminants (e.g., ethinyl estradiol in birth control pills, surfactants such as octylphenol, nonylphenol), plasticizers (e.g., phthalates), flame retardants (e.g., polychlorinated biphenyls, polybrominated diphenol ethers), industrial pollutants (e.g., heavy metals such as cadmium and lead, dioxins, polycyclic aromatic hydrocarbons), microwaves, and ultrasound. In some populations, sperm density is said to have declined by as much as 40% in the past 50 years. Environmental estrogens or antiandrogens may be partly responsible.

Sperm antibodies can cause isolated male infertility. In some instances, these antibodies are secondary phenomena resulting from duct obstruction or vasectomy. Granulomatous diseases can affect the testes, and testicular atrophy occurs in 10–20% of men with lepromatous leprosy because of direct tissue invasion by the mycobacteria. The tubules are involved initially, followed by endarteritis and destruction of Leydig cells.

Systemic disease can cause primary testis dysfunction in addition to suppressing gonadotropin production. In cirrhosis, a combined testicular and pituitary abnormality leads to decreased testosterone production independent of the direct toxic effects of ethanol. Impaired hepatic extraction of adrenal androstenedione leads to extraglandular conversion to estrone and estradiol, which partially suppresses LH. Testicular atrophy and gynecomastia are present in approximately one-half of men with cirrhosis. In chronic renal failure, testosterone synthesis and sperm production decrease despite elevated gonadotropins. The elevated LH level is due to reduced clearance. About one-fourth of men with renal failure have hyperprolactinemia. Improvement in testosterone production with hemodialysis is incomplete, but successful renal transplantation may return testicular function to normal. Testicular atrophy is present in one-third of men with sickle cell anemia. The defect may be at either the testicular or the pituitary level, although hypogonadotropic hypogonadism is more common. Sperm density can decrease temporarily after acute febrile illness in the absence of a change in testosterone production. Infertility in men with celiac disease is associated with a hormonal pattern typical of androgen resistance, namely, elevated testosterone and LH levels.

Neurologic diseases such as myotonic dystrophy, spinobulbar muscular atrophy (Kennedy's disease), and spinal cord injury are often associated with associated with impaired testicular function. In myotonic dystrophy, small testes may be associated with impairment of both spermatogenesis and Leydig cell function. Spinobulbar muscular atrophy (Kennedy's disease) is caused by an expansion of the glutamine repeat tract in the amino-terminal region of the AR; this expansion impairs function of the AR. Men with spinobulbar muscular atrophy often have undervirilization and infertility as a late manifestation. Spinal cord injury that causes paraplegia is often associated with low testosterone levels and may cause persistent defects in spermatogenesis; some patients retain the capacity for penile erection and ejaculation.

ANDROGEN INSENSITIVITY SYNDROMES

Mutations in the AR cause resistance to the action of testosterone and DHT. These X-linked mutations are associated with variable degrees of defective male phenotypic development and undervirilization **(Chap. 402)**. Although not technically hormone-insensitivity syndromes, two genetic disorders associated with mutations in the steroid 5α-reductase type 2 or the *CYP19* (aromatase) gene impair testosterone conversion to active sex steroids. Mutations in the *SRD5A2* gene, which encodes the steroid 5α-reductase type 2, prevent the conversion of testosterone to DHT, which is necessary for the normal development of the male external genitalia. Mutations in the *CYP19* gene, which encodes aromatase, prevent testosterone conversion to estradiol. Males with *CYP19* mutations have delayed epiphyseal fusion, tall stature, eunuchoid proportions, visceral adiposity, and osteoporosis, consistent with evidence from an estrogen receptor–deficient individual that these testosterone actions are mediated via estrogen.

GYNECOMASTIA

Gynecomastia refers to enlargement of the male breast. It is caused by excess estrogen action and is usually the result of an increased estrogen/androgen ratio. True gynecomastia is associated with glandular breast tissue that is >4 cm in diameter and often tender. Glandular tissue enlargement should be distinguished from excess adipose tissue: glandular tissue is firmer and contains fibrous-like cords. Gynecomastia occurs as a normal physiologic phenomenon in the newborn (due to transplacental transfer of maternal and placental estrogens), during puberty (high estrogen-to-androgen ratio in early stages of puberty), and with aging (increased fat tissue and increased aromatase activity along with the age-related decline in testosterone levels), but it can also result from pathologic conditions associated with testosterone deficiency or estrogen excess. The prevalence of gynecomastia increases with age and body mass index (BMI), likely because of increased aromatase activity in adipose tissue. Medications that alter androgen metabolism or action may also cause gynecomastia. The relative risk of breast cancer is increased in men with gynecomastia, although the absolute risk is relatively small.

PATHOLOGIC GYNECOMASTIA

Any cause of *androgen deficiency* can lead to gynecomastia, reflecting an increased estrogen/androgen ratio, as estrogen synthesis still occurs by aromatization of residual adrenal and gonadal androgens. Gynecomastia is a characteristic feature of Klinefelter syndrome **(Chap. 402)**. *Androgen insensitivity* disorders also cause gynecomastia. Androgen deprivation therapy using GnRH analogues with and without androgen receptor blockers in men with prostate cancer is often associated with painful breast enlargement. *Excess estrogen production* may be caused by tumors, including Leydig cell tumors; Sertoli cell tumors in isolation or in association with Peutz-Jeghers syndrome or Carney complex; granulosa cell tumors; and adrenal tumors producing estrogen precursors. Tumors that produce hCG, including some testicular germ cell tumors, stimulate Leydig cell estrogen synthesis. *Increased conversion of androgens to estrogens* can be a result of increased availability of substrate (androstenedione) for extraglandular estrogen formation (CAH, hyperthyroidism, and most feminizing adrenal tumors) or of diminished catabolism of androstenedione (liver disease) so that estrogen precursors are shunted to aromatase in peripheral sites. Obesity is associated with increased aromatization of androgen precursors to estrogens. Extraglandular aromatase activity can also be increased in tumors of the liver or adrenal gland or rarely as an inherited disorder. Several families with *increased peripheral aromatase activity* inherited as an autosomal dominant or as an X-linked disorder have been described. In some families with this disorder, a chimeric CYP19 due to an inversion in chromosome 15q21.2-3 causes the *CYP19* gene to be activated by the regulatory elements of contiguous genes (e.g., transient receptor potential cation channel subfamily M member 7 [*TRPM7*], *TMOD3*, or *FLJ14957*), resulting in excessive estrogen production in the fat and other extragonadal tissues. The familial aromatase excess syndrome due to *CYP19* mutation or chromosomal rearrangement is characterized by pre- or peripubertal onset of gynecomastia, advanced bone age, short adult height due to premature epiphyseal closure, and hypogonadotropic hypogonadism. Peutz-Jeghers syndrome is characterized by intestinal hamartomas, mucocutaneous pigmentation, calcifying Sertoli cell tumors, premature epiphyseal closure, and prepubertal gynecomastia due to increased aromatization. Hyperthyroidism is associated with elevated SHBG levels, which increase the free estradiol–to–free testosterone ratio even though total testosterone and estradiol levels are often normal. *Drugs* can cause gynecomastia by acting directly as estrogenic substances (e.g., oral contraceptives, phytoestrogens, digitalis) or inhibiting androgen synthesis (e.g., GnRH agonists, ketoconazole) or action (e.g., spironolactone, AR blockers such as enzalutamide); for many drugs, such as cimetidine, imatinib, or some antiretroviral drugs for HIV, the precise mechanism is unknown. Unintentional exposure to estrogenic agents in skin care products has been reported as a cause of gynecomastia in prepubertal children.

Because up to two-thirds of pubertal boys and about half of hospitalized men have palpable glandular tissue that is benign, detailed investigation or intervention is not indicated in all men presenting with gynecomastia **(Fig. 403-7)**. In addition to the extent of gynecomastia, recent onset, rapid growth, tender tissue, and occurrence in a lean subject should prompt more extensive evaluation. This should include a careful drug history, examination of the testes, assessment

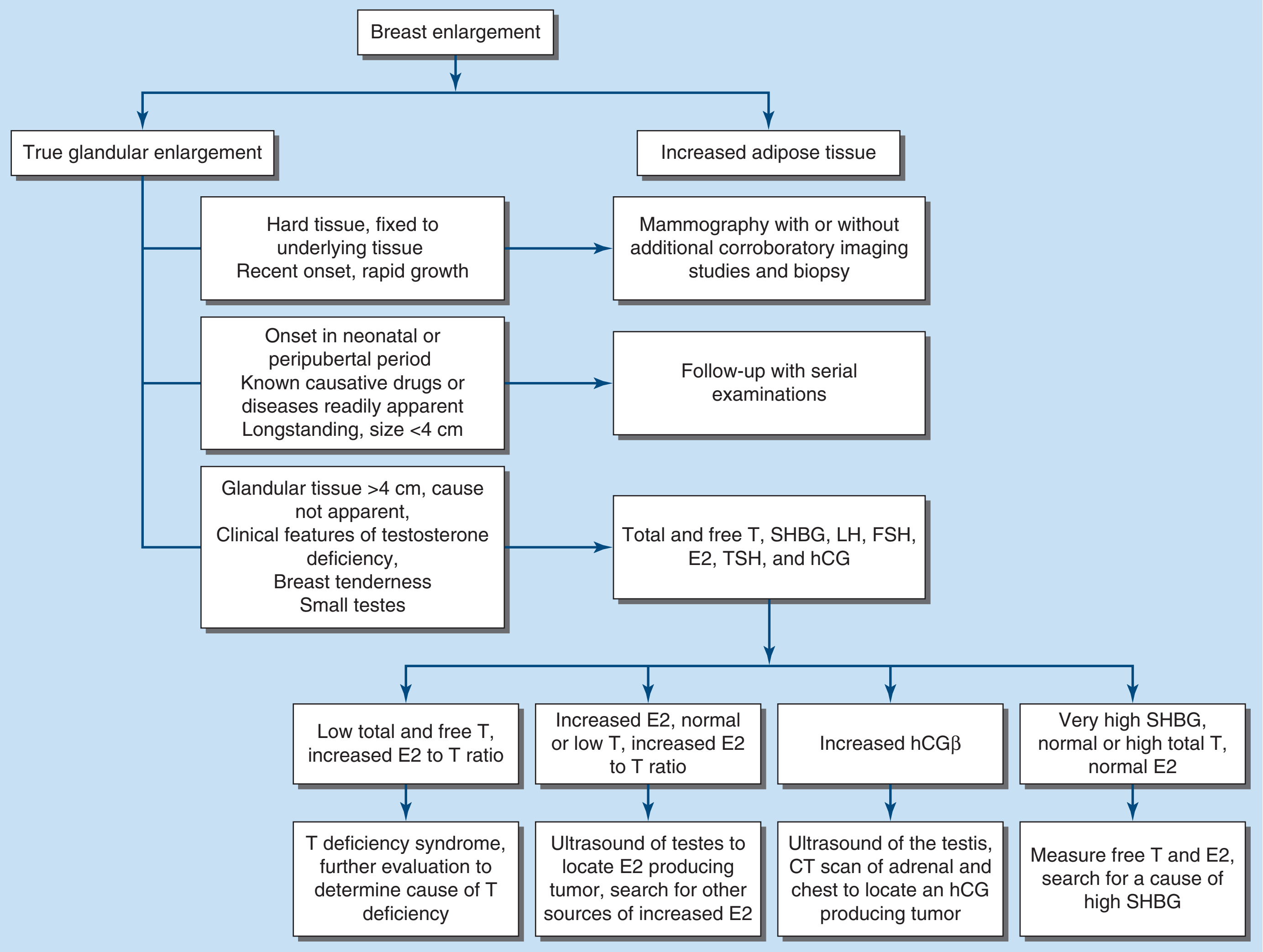
Breast enlargement
True glandular enlargement
Increased adipose tissue
Hard tissue, fixed to underlying tissue
Recent onset, rapid growth
Mammography with or without additional corroboratory imaging studies and biopsy
Onset in neonatal or peripubertal period
Known causative drugs or diseases readily apparent
Longstanding, size <4 cm
Follow-up with serial examinations
Glandular tissue >4 cm, cause not apparent,
Clinical features of testosterone deficiency,
Breast tenderness
Small testes
Total and free T, SHBG, LH, FSH, E2, TSH, and hCG
Low total and free T, increased E2 to T ratio
Increased E2, normal or low T, increased E2 to T ratio
Increased hCGβ
Very high SHBG, normal or high total T, normal E2
T deficiency syndrome, further evaluation to determine cause of T deficiency
Ultrasound of testes to locate E2 producing tumor, search for other sources of increased E2
Ultrasound of the testis, CT scan of adrenal and chest to locate an hCG producing tumor
Measure free T and E2, search for a cause of high SHBG

FIGURE 403-7 Evaluation of gynecomastia. E_2, 17β-estradiol; FSH, follicle-stimulating hormones; hCGβ, human chorionic gonadotropin β; LH, luteinizing hormone; T, testosterone.

of virilization, evaluation of liver function, and hormonal measurements including testosterone, estradiol, estrone, androstenedione, LH, and hCG. Markedly elevated estradiol concentrations along with suppressed LH should prompt a search for a testicular or adrenal estrogen-secreting tumor. A karyotype should be obtained in men with very small testes to exclude Klinefelter syndrome. Despite extensive evaluation, the etiology is established in only a small proportion of patients.

TREATMENT

Gynecomastia

When the primary cause can be identified and corrected shortly after the onset of gynecomastia, breast enlargement usually subsides over several months. However, if gynecomastia is of long duration, surgery is the most effective therapy. Indications for surgery include severe psychological distress, continued growth or tenderness, failure to respond to medical therapy, or suspected malignancy. In patients who have painful gynecomastia and in whom surgery cannot be performed, treatment with antiestrogens such as tamoxifen (20 mg/d) can reduce pain and breast tissue size in over half the patients. The estrogen receptor antagonists tamoxifen and raloxifene have been reported in small trials to reduce breast size in men with pubertal gynecomastia, although complete regression of breast enlargement is unusual with the use of estrogen receptor antagonists. Aromatase inhibitors can be effective in the early proliferative phase of the disorder. However, in a randomized trial in men with established gynecomastia, anastrozole proved no more effective than placebo in reducing breast size. Tamoxifen is effective in prevention and treatment of breast enlargement and breast pain in men with prostate cancer who are receiving androgen deprivation therapy. Long-standing gynecomastia is usually not responsive to drug therapy and requires mammoplasty if it is associated with pain and distress.

AGING-RELATED CHANGES IN MALE REPRODUCTIVE FUNCTION

A number of cross-sectional and longitudinal studies (e.g., the Baltimore Longitudinal Study of Aging, the Framingham Heart Study, the Massachusetts Male Aging Study, and the European Male Aging Study [EMAS]) have established that testosterone concentrations decrease with advancing age. This age-related decline starts in the third decade of life and progresses slowly; the rate of decline in testosterone concentrations is greater in men with obesity, chronic illness, and in those taking medications. Because SHBG concentrations are higher in older men than in younger men, free or bioavailable testosterone concentrations decline with aging to a greater extent than total testosterone concentrations. The age-related decline in testosterone is due to defects at all levels of the hypothalamic-pituitary-testicular axis: pulsatile GnRH secretion is attenuated, LH response to GnRH is reduced, and testicular response to LH is impaired. However, the gradual rise of LH with aging suggests that testis dysfunction is the main cause of declining androgen levels. The term *andropause* has been used to denote age-related decline in testosterone concentrations; this term is a misnomer because there is no discrete time when testosterone concentrations decline abruptly.

Several epidemiologic studies, such as the Framingham Heart Study, the EMAS, and the Study of Osteoporotic Fractures in Men (MrOS), that used mass spectrometry for measuring testosterone levels have reported ~10% prevalence of low testosterone levels in middle-aged and older men; the prevalence of unequivocally low testosterone and sexual symptoms in men aged 40–70 years in the EMAS was 2.1% and increased with age from 0.1% for men aged 40–49 years of age to 5.1% for those aged 70–79 years. Low total and bioavailable testosterone concentrations have been associated with decreased sexual desire, decreased appendicular skeletal muscle mass and strength, decreased self-reported physical function, higher whole body and visceral fat mass and insulin resistance, and increased risk of type 2 diabetes, coronary artery disease, and all-cause mortality. An analysis of signs and symptoms in older men in the EMAS revealed a syndromic association of sexual symptoms with total testosterone levels <320 ng/dL and free testosterone levels <64 pg/mL in community-dwelling older men.

Two placebo-controlled randomized trials—the Testosterone Trials (TTrials) and the Testosterone Replacement Therapy for Assessment of Long-Term Vascular Events and efficacy Response in Hypogonadal Men (TRAVERSE) study—have provided important information about the efficacy and safety of testosterone replacement therapy (TRT) in middle-aged and older men with hypogonadism. In these and other randomized trials, TRT of men with hypogonadism consistently improved sexual activity and sexual desire and alleviated hypogonadal symptoms. TRT also improves mood and energy level. TRT increases hemoglobin levels and corrects anemia in a greater proportion of middle-aged and older men with hypogonadism who have unexplained anemia, compared to placebo. Testosterone treatment significantly increases vertebral as well as femoral volumetric and areal bone mineral density and estimated bone strength. Testosterone treatment does not improve memory or other measures of cognition in men who do not have cognitive deficits. Testosterone treatment increases lean body mass, maximal voluntary muscle strength and muscle power, and self-reported mobility, and reduces whole body and visceral fat mass. In men with hypogonadism, testosterone treatment has not been shown to prevent progression from prediabetes to diabetes or induce glycemic remission in those with diabetes. In systematic reviews of randomized controlled trials, testosterone therapy of healthy older men with low or low-normal testosterone levels was associated with greater increments in lean body mass, grip strength, and self-reported physical function than that associated with placebo.

Neither the testosterone trials nor a randomized trial of the effects of testosterone on atherosclerosis progression in aging men (TEAAM trial) with low or low normal testosterone levels found significant differences between testosterone and placebo arms in the rates of change in either the coronary artery calcium scores or the common carotid artery intima-media thickness. However, in the TTrials, testosterone treatment was associated with a significantly greater increase in coronary artery noncalcified plaque volume, as measured by coronary artery computed tomography angiography. Neither of the trials was long enough or large enough to determine the effects of TRT on prostate or major adverse cardiovascular events.

The TRAVERSE trial compared the effects of TRT and placebo on major adverse cardiovascular events (MACE) in middle-aged and older men with hypogonadism and either preexisting cardiovascular disease (CVD) or increased risk of CVD. During a mean follow-up period of 33 months, the incidence of MACE did not differ between the testosterone and placebo groups. Testosterone treatment was associated with higher incidence of venous thromboembolism and atrial fibrillation than placebo. The incidences of high-grade prostate cancer (Gleason grade 4+3 or higher) or any prostate cancer, acute urinary retention, invasive surgical procedures for benign prostatic hyperplasia, and initiation of new pharmacologic therapy for benign prostatic hyperplasia were low and did not differ between the testosterone and placebo groups. Testosterone treatment was associated with significantly greater increase in serum prostate-specific antigen (PSA) level in the first 12 months of testosterone but the rate of change in PSA after 12 months of treatment was similar in the testosterone and placebo-treated men. Testosterone treatment did not worsen lower urinary tract symptoms. Testosterone treatment, by increasing PSA levels, could increase the risk of referral for prostate biopsy and detection of low-grade prostate cancer.

Population screening of all older men for low testosterone levels is not recommended, and testing should be restricted to men who have symptoms or signs attributable to androgen deficiency. Testosterone therapy is not recommended for all older men with low testosterone levels. In older men with significant symptoms of testosterone deficiency who have unequivocally low testosterone levels, testosterone treatment should be individualized based on consideration of the burden of symptoms, the presence of comorbid conditions that may increase the risk of harm from testosterone treatment, the risks versus benefits of treatment and monitoring, and clinician and patient values (see "Testosterone Replacement," below).

Testicular morphology, semen production, and fertility are maintained up to a very old age in men. Although concern has been expressed about age-related increases in germ cell mutations and impairment of DNA repair mechanisms, there is no clear evidence that the frequency of chromosomal aneuploidy is increased in the sperm of older men. However, the incidence of autosomal dominant diseases, such as achondroplasia, polyposis coli, Marfan syndrome, and Apert syndrome, increases in the offspring of men who are advanced in age, consistent with transmission of sporadic missense mutations. Advanced paternal age may be associated with increased rates of de novo mutations, which may contribute to an increased risk of neurodevelopmental diseases such as schizophrenia and autism. The somatic mutations in male germ cells that enhance the proliferation of germ cells could lead to within-testis expansion of mutant clonal lines, thus favoring the propagation of germ cells carrying these pathogenic mutations and increasing the risk of mutations in the offspring of older fathers (the "selfish spermatogonial selection" hypothesis).

APPROACH TO THE PATIENT

Testosterone Deficiency

Hypogonadism is often characterized by decreased sex drive, reduced frequency of sexual activity, inability to maintain erections, fatigue, low mood, hot flushes, reduced beard growth, loss of muscle mass, decreased testicular size, and gynecomastia. Erectile dysfunction and testosterone deficiency are two distinct clinical disorders that can coexist in middle-aged and older men. Some, but not all, patients with erectile dysfunction have testosterone deficiency. Thus, it is useful to evaluate men presenting with erectile dysfunction for testosterone deficiency. Except when extreme, these clinical features of testosterone deficiency may be difficult to distinguish from changes that occur with normal aging. Moreover, testosterone deficiency may develop gradually. When symptoms or clinical features suggest possible testosterone deficiency, the laboratory evaluation is initiated by the measurement of total testosterone in a fasting specimen, preferably in the morning using a reliable assay, such as LC-MS/MS, that has been calibrated to an international testosterone standard **(Fig. 403-8)**. A consistently low total testosterone level below the lower limit of the normal male range, measured by an LC-MS/MS assay in a CDC-certified laboratory, in association with symptoms, is evidence of testosterone deficiency. An early-morning testosterone level >400 ng/dL makes the diagnosis of androgen deficiency unlikely. In men with testosterone levels between 200 and 400 ng/dL, the total testosterone level should be repeated, and a free testosterone level should be measured. In older men and in patients with other clinical states that are associated with alterations in SHBG levels, a direct measurement of free testosterone level by equilibrium dialysis is needed to diagnose testosterone deficiency.

When testosterone deficiency has been confirmed by the consistently low testosterone concentrations, LH should be measured to determine whether the patient has primary (high LH) or secondary (low or inappropriately normal LH) hypogonadism. Elevated LH and FSH levels indicate a defect at the testicular level. Common causes of primary testicular failure include Klinefelter syndrome, HIV infection, uncorrected cryptorchidism, cancer chemotherapeutic agents, radiation, surgical orchiectomy, or prior infectious orchitis. A karyotype should be performed in men with low testosterone and elevated LH to diagnose Klinefelter syndrome. Men who have a low testosterone but "inappropriately normal" or low LH levels have secondary hypogonadism; their defect resides at the hypothalamic-pituitary level. Common causes of acquired secondary hypogonadism include space-occupying lesions of the sella,

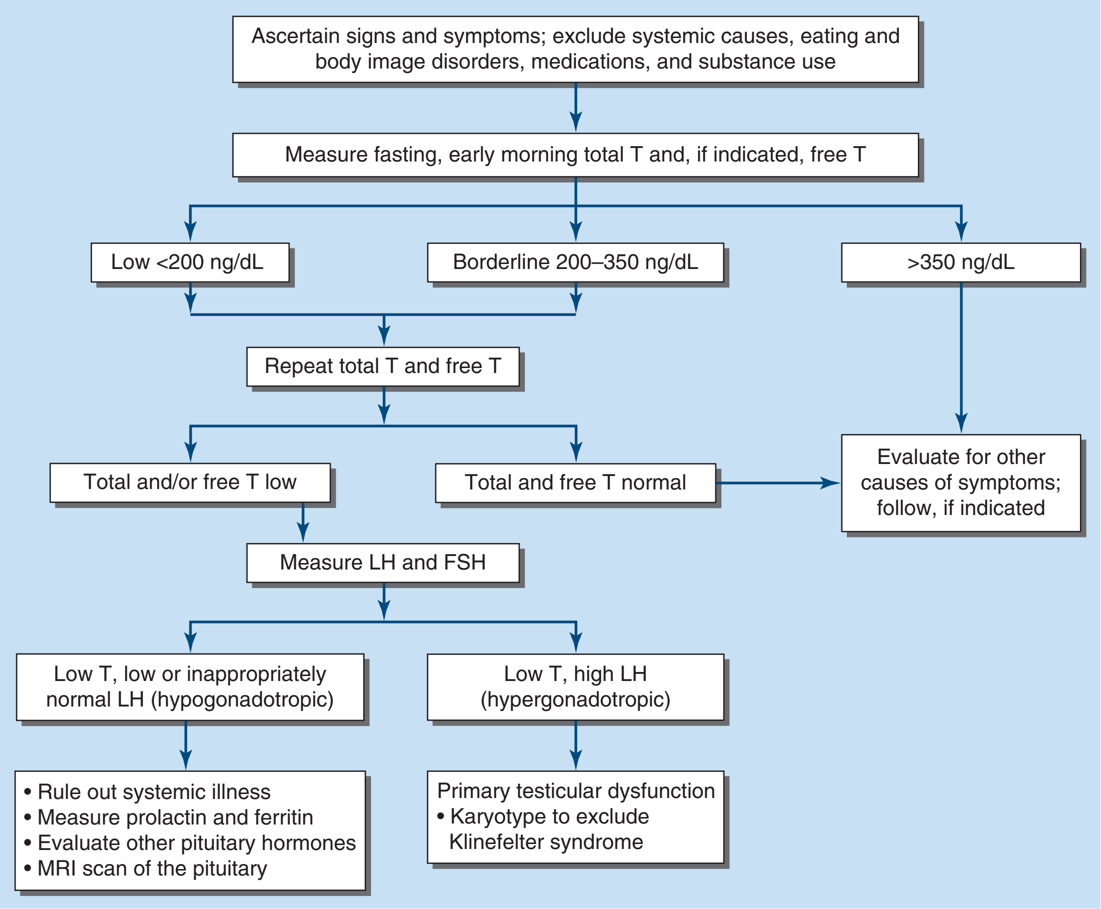
Ascertain signs and symptoms; exclude systemic causes, eating and body image disorders, medications, and substance use
Measure fasting, early morning total T and, if indicated, free T
Low <200 ng/dL
Borderline 200–350 ng/dL
>350 ng/dL
Repeat total T and free T
Total and/or free T low
Total and free T normal
Evaluate for other causes of symptoms; follow, if indicated
Measure LH and FSH
Low T, low or inappropriately normal LH (hypogonadotropic)
Low T, high LH (hypergonadotropic)
• Rule out systemic illness
• Measure prolactin and ferritin
• Evaluate other pituitary hormones
• MRI scan of the pituitary
Primary testicular dysfunction
• Karyotype to exclude Klinefelter syndrome

FIGURE 403-8 Evaluation of hypogonadism. FSH, follicle-stimulating hormone; GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone; T, testosterone.

hyperprolactinemia, chronic illness, hemochromatosis, excessive exercise, and the use of anabolic-androgenic steroids, opioids, marijuana, glucocorticoids, and alcohol. Measurement of PRL and MRI scan of the hypothalamic-pituitary region can help exclude the presence of a space-occupying lesion. Patients in whom known causes of hypogonadotropic hypogonadism have been excluded are classified as having IHH. It is not unusual for congenital causes of hypogonadotropic hypogonadism, such as Kallmann syndrome, to be diagnosed in young adults.

TREATMENT

Testosterone Deficiency

GONADOTROPINS

Gonadotropin therapy is used to establish or restore fertility in patients with gonadotropin deficiency of any cause. Several gonadotropin preparations are available. Human menopausal gonadotropin (hMG; purified from the urine of postmenopausal women) contains 75 IU of FSH and 75 IU of LH per vial. hCG (purified from the urine of pregnant women) has minimal FSH activity and resembles LH in its ability to stimulate testosterone production by Leydig cells. Recombinant LH is also available, but its half-life is relatively short (10 h). Treatment is usually begun with hCG alone. If hCG alone does not induce spermatogenesis, hMG is added to promote the FSH-dependent stages of spermatid development. Recombinant human FSH (hFSH) is available and is indistinguishable from purified urinary hFSH in its biologic activity and pharmacokinetics in vitro and in vivo, although the mature β subunit of recombinant hFSH has seven fewer amino acids. Recombinant hFSH is available in ampoules containing 75 IU (~7.5 μg FSH), which accounts for >99% of protein content. Once spermatogenesis is restored using combined FSH and LH therapy, hCG alone is often sufficient to maintain spermatogenesis.

Although a variety of treatment regimens are used, 1500 IU of hCG administered intramuscularly three times weekly is a reasonable starting dose. Testosterone levels should be measured 6–8 weeks later and 48–72 h after the hCG injection; the hCG dose should be adjusted to achieve testosterone levels in the mid-normal range. Sperm counts should be monitored on a monthly basis. It may take several months for spermatogenesis to be restored; therefore, it is important to forewarn patients about the potential length and expense of the treatment and to provide conservative estimates of success rates. If testosterone levels are in the mid-normal range but the sperm concentrations are low after 6 months of therapy with hCG alone, FSH should be added. This can be done by using hMG, highly purified urinary hFSH, or recombinant hFSH. A common practice is to start with the addition of 75 IU FSH three times each week in conjunction with the hCG injections. If sperm densities are still low after 3 months of combined treatment, the FSH dose should be increased to 150 IU. Occasionally, it may take ≥18–24 months for spermatogenesis to be restored.

The two best predictors of the success of gonadotropin therapy in hypogonadotropic men are testicular volume at presentation and time of onset of gonadotropin deficiency. In general, men with testicular volumes >8 mL have better response rates than those who have testicular volumes <4 mL. Patients who develop gonadotropin deficiency after puberty experience higher success rates than those who have never undergone pubertal changes. Spermatogenesis can usually be reinitiated by hCG alone, with high rates of success for men with postpubertal onset of hypogonadotropism. The presence of a primary testicular abnormality, such as cryptorchidism, will attenuate testicular response to gonadotropin therapy. Prior testosterone therapy does not preclude subsequent response to gonadotropin therapy, although some studies suggest that it may attenuate response to subsequent gonadotropin therapy.

TESTOSTERONE REPLACEMENT

TRT is indicated to restore testosterone levels to normal and correct features of testosterone deficiency in men with hypogonadism. TRT induces secondary sex characteristics; improves libido and overall sexual activity; relieves symptoms of hypogonadism; improves mood and energy level; increases skeletal muscle mass, muscle strength and power, and some measures of physical function; hemoglobin and hematocrit; and volumetric and areal bone mineral density; corrects the mild anemia associated with testosterone deficiency; and decreases fat mass. The benefits of TRT have only been proven in men with confirmed hypogonadism, as demonstrated by testosterone levels that are below the lower limit of normal and one or more symptoms of hypogonadism.

Testosterone is available in a variety of formulations with distinct pharmacokinetics **(Table 403-3)**. Testosterone serves as a prohormone and is converted to 17β-estradiol by aromatase and to 5α-DHT by steroid 5α-reductase. Therefore, when evaluating testosterone formulations, it is important to consider whether the formulation being used can achieve physiologic estradiol and DHT concentrations, in addition to normal testosterone concentrations. The current recommendation is to restore testosterone levels to the mid-normal range.

Oral Derivatives of Testosterone Testosterone is well absorbed after oral administration but is quickly degraded during the first pass through the liver. Therefore, it is difficult to achieve sustained blood levels of testosterone after oral administration of crystalline testosterone. 17α-Alkylated derivatives of testosterone (e.g., 17α-methyl testosterone, oxandrolone, fluoxymesterone) are relatively resistant to hepatic degradation and can be administered orally; however, because of the potential for hepatotoxicity, including cholestatic jaundice, peliosis, and hepatocellular neoplasms, these formulations should not be used for testosterone replacement. Hereditary angioedema due to C1 esterase deficiency is the only exception to this general recommendation; in this condition, oral 17α-alkylated androgens are useful because they stimulate hepatic synthesis of the C1 esterase inhibitor.

Injectable Forms of Testosterone The esterification of testosterone at the 17β-hydroxy position makes the molecule hydrophobic and extends its duration of action. The slow release of testosterone ester from an oily depot in the muscle accounts for its extended duration of action. The longer the side chain, the greater is the hydrophobicity of the ester and longer the duration of action. Thus, testosterone enanthate, cypionate, and undecanoate with longer side chains have longer durations of action than testosterone propionate. Within 24 h after intramuscular administration of 200 mg testosterone enanthate or cypionate, testosterone levels rise into the high-normal or supraphysiologic range and then gradually decline into the hypogonadal range over the next 2 weeks. A bimonthly regimen of testosterone enanthate or cypionate therefore results in peaks and troughs in testosterone levels that may be accompanied by changes in a patient's mood, sexual desire, and energy level; weekly administration of testosterone enanthate or cypionate can reduce these variations in testosterone levels during the dosing interval. The kinetics of testosterone enanthate and cypionate are similar. Estradiol and DHT levels are normal if testosterone replacement is physiologic.

A long-acting testosterone undecanoate in oil, administered at an initial priming dose of 750 mg intramuscularly followed by a second dose of 750 mg 4 weeks later, and then at a maintenance dose of 750 mg every 10 weeks, maintains serum testosterone, estradiol, and DHT in the normal male range and corrects symptoms of testosterone deficiency in a majority of treated men. Its relative drawbacks are the large injection volume and the risk of pulmonary oil microembolism (POME) reaction in a small proportion of patients.

Testosterone Gel Several transdermal testosterone gels, such as Androgel, Testim, Fortesta, and Axiron, and some generic versions, when applied topically to the skin in appropriate doses (Table 403-3), can maintain total and free testosterone concentrations in the normal range in men with hypogonadism. The current

TABLE 403-3 Clinical Pharmacology of Some Testosterone Formulations

FORMULATION	REGIMEN	PHARMACOKINETIC PROFILE	DHT AND E2	ADVANTAGES	DISADVANTAGES
T enanthate or cypionate	140–200 mg IM q2wk or 70–100 mg/wk	After a single IM injection, serum T levels rise into the supraphysiologic range, then decline gradually into the low-normal or the hypogonadal range by the end of the dosing interval	DHT and E2 levels rise in proportion to the increase in T levels; T:DHT and T:E2 ratios do not change	Corrects symptoms of testosterone deficiency; relatively inexpensive if self-administered; flexibility of dosing	Requires IM injection; peaks and valleys in serum T levels that are associated with fluctuations in patient's mood, energy level, and sex drive
Topical T gels	Available in sachets, tubes, and pumps	When used in appropriate doses, these topical formulations restore serum T and E2 levels to the physiologic male range	Serum DHT levels and DHT:T ratio are higher in hypogonadal men treated with the transdermal gels than in healthy eugonadal men	Corrects symptoms of testosterone deficiency, ease of application, good skin tolerability	Potential of transfer to a female partner or child by direct skin-to-skin contact; skin irritation in a small proportion of treated men; moderately high DHT levels; considerable interindividual and intraindividual variation in on-treatment testosterone levels
T pellets	Several pellets implanted SC; dose and regimen vary with formulation	Serum T peaks at 1 month and then is sustained in normal range for 3–4 months, depending on formulation	T:DHT and T:E2 ratios do not change	Corrects symptoms of testosterone deficiency	Requires surgical incision for insertions; pellets may extrude spontaneously
Oral T undecanoate (TU)	Two different formulations of oral testosterone are available each taken orally bid with food	TU formulated in a self-emulsifying drug delivery system that includes hydrophilic and lipophilic excipients to enable the solubilization of TU and its absorption through the lymphatics after oral ingestion with a typical meal. After each administration, serum T levels rise and return to baseline by 12 h. When administered at the recommended dose, average serum T levels are maintained in the normal range in a majority of treated men	High DHT:T ratio	Convenience of oral administration	High DHT:T ratio
Injectable long-acting TU in oil	U.S. regimen 750 mg IM, followed by 750 mg at 4 weeks, and 750 mg every 10 weeks	When administered at the recommended dose, serum T levels are maintained in the normal range in a majority of treated men	DHT and E2 levels rise in proportion to the increase in T levels; T:DHT and T:E2 ratios do not change	Corrects symptoms of testosterone deficiency; requires infrequent administration	Requires IM injection of a large volume; serious pulmonary oil microembolism (POME) reactions, characterized by cough, dyspnea, throat tightening, chest pain, dizziness, and syncope, and episodes of anaphylaxis have been reported to occur during or immediately after the injection in a very small number of patients; patients should be watched for POME reaction for 30 min after each injection
Intranasal T	2 actuations of the metered-dose pump (11 mg) applied into the nostrils 3 times daily	Restores T into the normal male range	T:DHT and T:E2 ratio in the physiologic range		Requires 3 times daily application; nasal irritation, epistaxis, nasopharyngitis

Abbreviations: DHT, dihydrotestosterone; E2, estradiol; T, testosterone.

recommendations are to begin with an initial U.S. Food and Drug Administration–recommended dose and adjust the dose based on testosterone and hematocrit levels. The advantages of the testosterone gel include the ease of application. A major concern is the potential for transfer of the gel to a sexual partner or to children who may come in close contact with the patient. The ratio of DHT to testosterone concentrations is higher in men treated with the testosterone gel than in healthy men. Also, there is considerable intra- and interindividual variation in serum testosterone levels in men treated with the transdermal gel due to variations in transdermal absorption and plasma clearance of testosterone. Therefore, monitoring of serum testosterone levels and multiple dose adjustments are required to achieve and maintain testosterone levels in the target range.

Pellets of crystalline testosterone can be inserted in the subcutaneous tissue through a small skin incision. Testosterone is released by surface erosion of the implanted pellets and absorbed into the systemic circulation, and testosterone levels can be maintained in the normal range for 3–4 months. Potential drawbacks include the need for skin incision for insertion and removal and spontaneous extrusions and fibrosis at the insertion.

Testosterone undecanoate, formulated in a self-emulsifying drug delivery system that includes hydrophilic and lipophilic excipients to enable its solubilization in the gut, is absorbed through the lymphatics after oral ingestion with a fatty meal and is spared the first-pass degradation in the liver. After each administration, serum testosterone levels rise and return to baseline by 12 h. When administered at the recommended dose, average serum testosterone levels are maintained in the normal range in a majority of treated men, but DHT-to-testosterone ratios are higher in men with hypogonadism treated with oral testosterone undecanoate, as compared to eugonadal men.

An intranasal testosterone gel is available as a metered-dose pump and is administered typically at a starting dose of 11 mg testosterone in the form of two pump actuations, one in each nostril three times daily. Formulation-specific adverse effects include rhinorrhea, nasal discomfort, epistaxis, nasopharyngitis, and nasal scab.

Novel Androgen Formulations Selective AR modulators (SARMs) are a class of AR ligands that bind the AR and display tissue-selective actions. A number of nonsteroidal SARMs that act as agonists on the muscle and bone and that spare the prostate to varying degrees have advanced to phase 3 human trials. Nonsteroidal SARMs do not serve as substrates for either the steroid 5α-reductase or the CYP19 (aromatase). SARM binding to AR induces specific conformational changes in the AR protein, which then modulates protein–protein interactions between AR and its coregulators, resulting in tissue-specific regulation of gene expression. SARMs that are strong agonists for the muscle, bone, and sexual function and antagonists for the prostate may be valuable in treating men with prostate cancer who are receiving androgen deprivation therapy. 7α-Methyl-19-nortestosterone is an androgen that cannot be 5α-reduced; therefore, compared to testosterone, it has relatively greater agonist activity in muscle and gonadotropin suppression but lesser activity on the prostate.

Pharmacologic Uses of Androgens Androgens and SARMs are being evaluated as anabolic therapies for functional limitations associated with aging and chronic illness. Testosterone supplementation increases skeletal muscle mass, maximal voluntary strength, and muscle power in healthy men, hypogonadal men, older men with low testosterone levels, HIV-infected men with weight loss, and men receiving glucocorticoids. These anabolic effects of testosterone are related to testosterone dose and circulating concentrations. Systematic reviews have confirmed that testosterone therapy of HIV-infected men with weight loss promotes improvements in body weight, lean body mass, muscle strength, and depression indices, leading to the recommendation that testosterone may be considered as an adjunctive therapy in HIV-infected men who are experiencing unexplained weight loss and who have low testosterone levels. It is unknown whether testosterone therapy of older men with functional limitations is safe and effective in improving physical function and reducing disability.

Testosterone administration induces hypertrophy of both type 1 and 2 fibers and increases satellite cell (muscle progenitor cells) and myonuclear number. Androgens promote the differentiation of mesenchymal, multipotent progenitor cells into the myogenic lineage and inhibit their differentiation into the adipogenic lineage. Testosterone binding to AR promotes the association of liganded AR with β-catenin and its translocation into the nucleus where it binds TCF-4 and activates Wnt-target genes, including follistatin, which blocks signaling through the transforming growth factor β pathway, thereby promoting myogenic differentiation of muscle progenitor cells. Testosterone may have additional effects on satellite cell replication and polyamine pathway, which may contribute to an increase in skeletal muscle mass.

Other indications for androgen therapy are in selected patients with anemia due to bone marrow failure (an indication largely supplanted by erythropoietin) and hereditary angioedema.

Recommended Regimens for Testosterone Replacement Therapy Testosterone esters are administered typically at doses of 70–100 mg intramuscularly every week or 140–200 mg every 2 weeks. Testosterone undecanoate is administered at an initial dose of 750 mg followed 4 weeks later by a second injection of 750 mg and then 750 mg every 10 weeks. Testosterone gels are typically applied over a covered area of skin at initial doses that vary with the formulation. Patients should wash their hands after gel application and keep the area of gel application covered with clothing to minimize the risk of gel transfer to another person. Oral testosterone undecanoate is taken with meals at a starting dose of 237 mg twice daily. Intranasal testosterone is administered as a spray in each nostril three times a day (33 mg/d).

Evaluating Efficacy of Testosterone Replacement Therapy The goals of TRT are to restore serum testosterone levels into the mid-normal range for healthy young men, correct symptoms of hypogonadism, and induce and maintain secondary sex characteristics. Testosterone should be measured 3 months after initiating therapy to assess adequacy of therapy. There is substantial interindividual variability in serum testosterone levels, especially with transdermal gels, presumably due to genetic differences in testosterone clearance and substantial variation in transdermal absorption. In patients who are treated with testosterone enanthate or cypionate, testosterone levels should be 350–550 ng/dL 1 week after the injection. If testosterone levels are outside this range, adjustments should be made either in the dose or in the interval between injections. In men on transdermal testosterone patch or gel, testosterone levels should be in the mid-normal range (400–750 ng/dL) 4–12 h after application. If testosterone levels are outside this range, the dose should be adjusted. Multiple dose adjustments are often necessary to achieve testosterone levels in the desired therapeutic range.

Restoration of sexual function, induction and maintenance of secondary sex characteristics, well-being, and maintenance of muscle and bone health are important objectives of TRT. The patient should be asked about sexual desire and activity, the presence of early morning erections, and the ability to achieve and maintain erections adequate for sexual intercourse. The hair growth in response to androgen replacement is variable and depends on ethnicity. Men with prepubertal onset of testosterone deficiency who begin testosterone therapy in their late twenties or thirties may find it difficult to adjust to their newly found sexuality and may benefit from counseling. If the patient has a sexual partner, the partner should be included in counseling because of the dramatic physical and sexual changes that occur with TRT.

Contraindications for Testosterone Administration Testosterone administration is contraindicated in men with prostate or breast cancer **(Table 403-4)**. Testosterone therapy should not be administered without further urologic evaluation to men with a palpable prostate nodule or induration, a PSA >3 ng/mL, or severe lower urinary tract symptoms (American Urological Association lower urinary tract symptom score >19). Testosterone should not be administered to men with a hypercoagulable condition, baseline hematocrit ≥50%, severe untreated obstructive sleep apnea, or uncontrolled or poorly controlled congestive heart failure, or with myocardial infarction, stroke, or acute coronary syndrome in the preceding 4 months.

Monitoring Potential Adverse Experiences The clinical effectiveness and safety of TRT should be assessed 3–6 months after initiating testosterone therapy and annually thereafter **(Table 403-5)**. Potential adverse effects include acne, oiliness of skin, erythrocytosis, venous thromboembolism, breast tenderness, leg edema, and increased risk of detection of prostate events. In addition, there may be formulation-specific adverse effects such as skin irritation with transdermal patch; risk of gel transfer to a sexual partner with testosterone gels; pain and mood fluctuation with injectable testosterone esters; cough and injection site pain with long-acting

TABLE 403-4 Conditions in Which Testosterone Administration Should not be Used or Used with Increased Caution

Conditions in which testosterone administration is associated with very high risk of serious adverse outcomes:
Metastatic prostate cancer
Some types of breast cancers
Thrombophilia
Conditions in which testosterone administration is associated with moderate to high risk of adverse outcomes and should be used with increased caution and only after further evaluation
Undiagnosed prostate nodule or induration
PSA >3
Erythrocytosis (hematocrit >50%)
Severe lower urinary tract symptoms associated with benign prostatic hypertrophy as indicated by American Urological Association/International prostate symptom score >19
Uncontrolled or poorly controlled congestive heart failure
Myocardial infarction, stroke, or acute coronary syndrome in the preceding 4 months

Abbreviation: PSA, prostate-specific antigen.

TABLE 403-5 Monitoring Men Receiving Testosterone Therapy

1. Evaluate the patient 3–6 months after treatment initiation and then annually to assess whether symptoms have responded to treatment and whether the patient is suffering from any adverse effects.
2. Monitor testosterone level 3–6 months after initiation of testosterone therapy:
   - Therapy should aim to raise average serum testosterone level into the mid-normal range (~400–750 ng/dL).
     - Injectable testosterone enanthate or cypionate: Measure serum testosterone level midway between injections. If testosterone is >750 ng/dL (26.2 nmol/L) or <400 ng/dL (14.0 nmol/L), adjust dose or frequency.
     - Transdermal gels and solution: Assess testosterone level 2–12 h after patient has been on treatment for at least 2 weeks; adjust dose to achieve serum testosterone level in the mid-normal range (400–750 ng/dL).
     - Testosterone pellets: Measure testosterone levels at the end of the dosing interval. Adjust the number of pellets and/or the dosing interval to achieve serum testosterone levels in the normal range.
     - Oral testosterone undecanoate: Measure testosterone levels 6–8 h after an oral dose.
     - Injectable testosterone undecanoate: Measure serum testosterone level just prior to each subsequent injection and adjust the dosing interval to maintain serum testosterone in mid-normal range.
3. Check hematocrit at baseline, at 3–6 months, and then annually. If hematocrit is >54%, stop therapy until hematocrit decreases to a safe level; evaluate the patient for hypoxia and sleep apnea; reinitiate therapy with a reduced dose.
4. Measure bone mineral density of lumbar spine and/or femoral neck after 1–2 years of testosterone therapy in hypogonadal men with osteoporosis or low trauma fracture, consistent with regional standard of care.
5. In men aged ≥55 years, perform digital rectal examination and check PSA level before initiating treatment, at 3–6 months, and then in accordance with guidelines for prostate cancer screening depending on the age and race of the patient.
6. Obtain urologic consultation if there is:
   - An increase in serum PSA concentration >1.4 ng/mL within the first 12 months after starting testosterone treatment, confirmed by repeating the test.
   - A PSA level >4 ng/mL any time during treatment, confirmed by repeating the test.
   - Detection of a prostatic abnormality on digital rectal examination.
   - Severe lower urinary tract symptoms
7. Evaluate formulation-specific adverse effects at each visit:
   - Injectable testosterone esters (enanthate, cypionate, and undecanoate): Ask about fluctuations in mood or libido, and rarely cough after injections.
   - Testosterone gels: Advise patients to cover the application sites with a shirt and to wash the skin with soap and water before having skin-to-skin contact because testosterone gels leave a testosterone residue on the skin that can be transferred to a woman or child who might come in close contact. Serum testosterone levels are maintained when the application site is washed 4–6 h after application of the testosterone gel.
   - Testosterone undecanoate injection: Observe patients for POME reaction for 30 min after each injection.
   - Testosterone pellets: Look for signs of infection, fibrosis, or pellet extrusion.
   - Intranasal testosterone: Look for signs of nasal irritation or scab.

Abbreviations: AUA/IPSS, American Urological Association International Prostate Symptom Score; POME, pulmonary oil microembolism; PSA, prostate-specific antigen.

testosterone undecanoate; and nasal irritation, epistaxis, and nasal scab with intranasal formulation. In the TRAVERSE trial, a randomized controlled trial of TRT in middle-aged and older men with hypogonadism, TRT was associated with increased risk of atrial fibrillation, acute kidney injury, and bone fractures.

Hemoglobin Levels Administration of testosterone to men with hypogonadism is typically associated with only a small (~3%) increase in hemoglobin levels, due to direct effects of testosterone on hematopoietic progenitors in the bone marrow, stimulation of erythropoietin, suppression of hepcidin, and increased iron availability for erythropoiesis. The magnitude of hemoglobin increase during TRT is greater in older men than younger men and in men who have sleep apnea, a significant smoking history, or chronic obstructive lung disease or who live at high altitude. The frequency of erythrocytosis is higher in men with hypogonadism treated with injectable testosterone esters than in those treated with transdermal formulations, presumably due to the higher testosterone dose delivered by the typical regimens of testosterone esters. Erythrocytosis is a frequent adverse event reported in testosterone trials in middle-aged and older men and is also the most frequent cause of treatment discontinuation in these trials. If hematocrit rises above 54%, testosterone therapy should be stopped until hematocrit has fallen to <50%. After evaluation of the patient for hypoxia and sleep apnea, testosterone therapy may be reinitiated at a lower dose.

Prostate and Serum PSA Levels TRT increases prostate volume to the size seen in age-matched controls. In the TRAVERSE trial, the incidences of high-grade or any prostate cancer were low and did not differ between the testosterone- and placebo-treated men. The incidences of acute urinary retention, invasive surgical procedure for benign prostatic hyperplasia, prostate biopsy, or initiation of new pharmacologic therapy for benign prostatic hyperplasia also were similar in the testosterone and placebo groups. Testosterone treatment did not worsen lower urinary tract symptoms. There is no evidence that TRT causes prostate cancer. However, testosterone treatment can exacerbate preexisting metastatic prostate cancer. Many older men harbor microscopic foci of cancer in their prostates. It is not known whether long-term TRT will induce these microscopic foci to grow into clinically significant cancers.

PSA levels are lower in testosterone-deficient men and increase after initiation of TRT. An increase in PSA levels after starting TRT can lead to urologic referral, increased risk of prostate biopsy, and detection of a low-grade prostate cancer that is common among middle-aged and older men. Increments in PSA levels after testosterone supplementation in androgen-deficient men are generally <0.5 ng/mL, and increments >1.0 ng/mL over a 3- to 6-month period are unusual. The 90% confidence interval for the change in PSA values in men with benign prostatic hyperplasia, measured 3–6 months apart, is 1.4 ng/mL. Therefore, the Endocrine Society expert panel suggested that an increase in PSA >1.4 ng/mL in the first year after starting TRT, if confirmed, should lead to urologic evaluation. There is considerable test-retest variability in PSA measurements. Therefore, PSA elevations after starting TRT should be confirmed by repeating the PSA test no sooner than 4 weeks after the first test. PSA level >4 ng/mL at any time during treatment, if confirmed by repeat testing, requires further urologic evaluation. Because prostate biopsy has the potential for harm (infection, bleeding, and diagnosis of a low-grade prostate cancer), the institution of PSA monitoring and urologic referral for a prostate biopsy should be a shared decision of the patient and the clinician.

Cardiovascular Risk In 2015, the U.S. Food and Drug Administration (FDA), because of concerns about the cardiovascular risk of TRT, required the testosterone manufacturers to conduct a randomized controlled trial to compare the effects of TRT versus placebo on cardiovascular events. The TRAVERSE trial, a large prospective randomized controlled trial, conducted in response to the FDA's requirement, has provided strong evidence that testosterone treatment does not increase the risk of major cardiovascular events (death due to cardiovascular cause, nonfatal myocardial infarction, or nonfatal stroke) relative to placebo treatment in middle-aged and older men with hypogonadism and preexisting cardiovascular disease or increased risk of cardiovascular disease during a mean follow-up duration of 33 months. The incidence of a secondary composite cardiovascular endpoint that included death due to cardiovascular cause, nonfatal myocardial infarction, nonfatal stroke, or coronary revascularization procedure also did not differ between the testosterone and placebo groups. Testosterone-treated men experienced higher rates of venous thromboembolic events, nonfatal arrythmias clinical bone fractures, and acute kidney injury than placebo-treated men. The TRAVERSE trial also found a small but statistically significantly greater increase in blood pressure in the testosterone than in the placebo group, similar to the findings of some other studies.

Androgen Use by Athletes and Recreational Bodybuilders The illicit use of androgenic-anabolic steroids (AAS) to enhance athletic performance first surfaced in the 1950s among powerlifters and spread rapidly to other sports, professional as well as high school athletes, and recreational bodybuilders. In the early 1980s, the use of AAS spread beyond the athletic community into the general population, and now, as many as 3–4 million Americans—most of them men—have likely used these compounds. Most AAS users are not athletes, but rather recreational weightlifters, almost all men, who use these drugs to look lean and more muscular. A subset of AAS users suffers from muscle dysmorphia, a form of body image disorder characterized by excessive preoccupation with leanness and muscularity and poor functioning in social and occupational life. A secular transformation of idealized body image toward greater muscularity and leanness has contributed to increasing prevalence of body image disorders and the use of muscle-building anabolic drugs in young men.

The most commonly used AAS include testosterone esters, nandrolone, trenbolone, stanozolol, methandienone, boldenone, and methenolone. AAS users generally use large doses of multiple steroids in cycles, a practice known as stacking. AAS users may also typically use other drugs that are perceived to be muscle building or performance enhancing, such as human GH; erythropoiesis-stimulating agents; insulin; stimulants such as amphetamine, clenbuterol, cocaine, ephedrine, and thyroxine; and drugs perceived to reduce adverse effects such as hCG, aromatase inhibitors, or estrogen antagonists. Recent years have witnessed increasing use of unapproved nonsteroidal SARMs and GH secretagogues purchased from internet sites.

Most of the information about the adverse effects of AAS has emerged from case reports, uncontrolled studies, or clinical trials that used replacement doses of testosterone. The adverse event data from clinical trials using physiologic replacement doses of testosterone have been extrapolated unjustifiably to AAS users, who may administer 10–100 times the replacement doses of testosterone over many years, to support the claim that AAS use is safe and manageable. The adverse events associated with AAS use may be due to AAS themselves, concomitant use of other drugs, high-risk behaviors, and host characteristics that may render these individuals more susceptible to AAS use or to other high-risk behaviors.

Four categories of adverse events associated with AAS abuse are of particular concern: cardiovascular events, psychiatric disorders, prolonged suppression of the hypothalamic-pituitary-testicular axis, and potential neurotoxicity. The high rates of premature mortality observed in AAS users are alarming. One Finnish study reported 4.6 times the risk of death among elite powerlifters compared with age-matched men from the general population. The causes of death among powerlifters included suicides, myocardial infarction, and liver failure. A retrospective review of patient records in Sweden also reported higher standardized mortality ratios for AAS users than for nonusers and increased death rates due to suicide, homicide, and accidents. High doses of AAS may induce proatherogenic dyslipidemia, accelerate atherogenesis, increase thrombosis risk via effects on clotting factors and platelets, and induce vasospasm through their effects on vascular nitric oxide. Long-term AAS use may be associated with myocardial hypertrophy and fibrosis. Myocardial tissue of powerlifters using AAS has been shown to be infiltrated with fibrous tissue and fat droplets. Current AAS users display significantly reduced left ventricular systolic and diastolic function compared to previous users and nonusers. Additionally, studies using CT angiography have reported higher coronary artery plaque volume in AAS users than in nonusers. Lifetime AAS dose is strongly associated with coronary atherosclerotic burden. Power athletes using AAS often have short QT intervals but increased QT dispersion, which may predispose them to ventricular arrhythmias.

Unlike replacement doses of testosterone, which are associated with only a small decrease in high-density lipoprotein (HDL) cholesterol and little or no effect on total cholesterol, low-density lipoprotein (LDL) cholesterol, and triglyceride levels, supraphysiologic doses of testosterone and orally administered 17α-alkylated, nonaromatizable AAS are associated with marked reductions in HDL cholesterol and increases in LDL cholesterol.

Some AAS users develop hypomanic and manic symptoms (irritability, aggressiveness, reckless behavior, and occasional psychotic symptoms, sometimes associated with violence) during AAS exposure, and depression, sometimes associated with suicidality, during AAS withdrawal. Users may also be susceptible to other forms of illicit drug use.

Long-term AAS use suppresses LH, FSH, and testosterone levels and spermatogenesis. Men who have used AAS for more than a few months experience marked suppression of the hypothalamic-pituitary-testicular (HPT) axis after stopping AAS that may be associated with sexual dysfunction, fatigue, infertility, depressed mood, and even suicidality. In some long-term AAS users, recovery of the HPT axis may take a long time, may be incomplete, or may never occur. The symptoms of testosterone deficiency caused by AAS withdrawal may cause some men to revert back to using AAS, leading to continued use and AAS dependence. As many as 30% of AAS users develop a syndrome of AAS dependence, characterized by long-term AAS use despite adverse medical and psychiatric effects. AAS withdrawal hypogonadism has emerged as an important cause of androgen deficiency, accounting for a substantial fraction of testosterone prescriptions in many men's health clinics; therefore, AAS use should be considered in the differential diagnosis of hypogonadism in young men.

Supraphysiologic doses of testosterone may also impair insulin sensitivity. Orally administered androgens also have been associated with insulin resistance and diabetes.

AAS users are more likely to engage in high-risk behaviors such as unsafe injection practices and have increased rates of incarceration that may render them at increased risk of HIV and hepatitis B and C. AAS users are more likely to report high-risk unprotected anal sex than nonusers.

Elevated liver enzymes, cholestatic jaundice, hepatic neoplasms, and peliosis hepatis have been reported with oral, 17α-alkylated AAS. AAS use may cause muscle hypertrophy without compensatory adaptations in tendons, ligaments, and joints, thus increasing the risk of tendon, ligament, and joint injuries. Upper extremity tendon ruptures are observed predominantly among weightlifters who use AAS. AAS use is associated with acne, baldness, and increased body hair.

APPROACH TO THE PATIENT

Androgenic-Anabolic Steroids Use

The suspicion of AAS use should be raised by increased hemoglobin and hematocrit levels, suppressed LH and FSH and testosterone levels, low HDL cholesterol and SHBG levels, and low testicular volume and sperm density in a person who looks highly muscular. In AAS users seeking medical attention, evaluation using the Appearance and Performance Enhancing Drug Use Schedule (APEDUS), a validated semi-structured interview, is sufficient to assess the associated body image or eating disorder, psychiatric symptoms, and the use of AAS and other substances; formal testing for AAS usually is rarely needed in clinical practice. History of AAS use should be obtained in all young men being evaluated for hypogonadism. As AAS use is often associated with the use of other substances, a urine dug screen for other substances is helpful in guiding treatment. If needed, accredited laboratories use gas chromatography–mass spectrometry or liquid chromatography–tandem mass spectrometry to detect anabolic steroid abuse. High-resolution mass spectrometry and tandem mass spectrometry have improved the sensitivity of detecting AAS use. Illicit testosterone use is detected generally by the measurement of urinary testosterone-to-epitestosterone ratio and further confirmed by the ^{13}C:^{12}C ratio in testosterone using isotope ratio combustion mass spectrometry. Exogenous testosterone administration increases urinary testosterone glucuronide excretion and consequently the testosterone-to-epitestosterone ratio. Ratios >6 are highly suggestive of exogenous testosterone use but can also reflect genetic variation. Genetic variations in uridine diphospho-glucuronyltransferase 2B17 (*UGT2B17*), the major enzyme for testosterone glucuronidation, affect the testosterone-to-epitestosterone ratio. Synthetic testosterone has a lower ^{13}C:^{12}C ratio than endogenously produced testosterone, and these differences in ^{13}C:^{12}C ratio can be detected by isotope ratio combustion mass spectrometry, which is used to confirm exogenous testosterone use in individuals with a high testosterone-to-epitestosterone ratio.

The treatment of AAS use disorder requires a multidisciplinary team that includes an endocrinologist or an internist to treat the AAS withdrawal hypogonadism and other medical problems; a mental health expert to treat the substance use disorder and depressive symptoms and to address suicide risk and body image disorder; and sometimes a social worker for care coordination. In patients who are willing to stop or who have already stopped AAS use, the initial step is to restore the hypothalamic-pituitary-gonadal axis by administering either clomiphene (or its enantiomer trans enclomiphene), a partial estrogen agonist, at an initial dose of 50 mg daily or hCG at a dose of 1000–2000 IU three times weekly. Some men may not respond to clomiphene and may require hCG. AAS users also need evaluation and treatment of the underlying body image disorder. Mirror exposure therapy in which the patient stands in front of a mirror and describes his body appearance to the mental health provider has been moderately efficacious in small, randomized trials. Body dysmorphia may require cognitive-behavioral therapy or pharmacotherapy using selective serotonin uptake inhibitors or tricyclic antidepressants.

■ FURTHER READING

Argenta J et al: Molecular basis of normal and pathological puberty: From basic mechanisms to clinical implications. Lancet Diabetes Endocrinol 11:203, 2023.

Bhasin S et al: Testosterone therapy in men with hypogonadism: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab 103:1715, 2018.

Bhasin S et al: Prostate safety events during testosterone replacement therapy in men with hypogonadism: A randomized controlled trial. JAMA Netw Open 6:e2348692, 2023.

Finkelstein JS et al: Gonadal steroids and body composition, strength, and sexual function in men. N Engl J Med 369:1011, 2013.

Hildebrandt T et al: Body image disturbance in 1000 male appearance and performance enhancing drug users. J Psychiatr Res 44:841, 2010.

Hughes JF, Page DC: The biology and evolution of mammalian Y chromosomes. Annu Rev Genet 49:507, 2015.

Jasuja R et al: Estradiol binding induces bidirectional allosteric coupling and repartitioning of sex hormone binding globulin monomers among various conformational states. iScience 24:102414, 2021.

Krausz C et al: EAA/EMQN best practice guidelines for molecular diagnosis of Y-chromosomal microdeletions: State-of-the-art 2013. Andrology 1:5, 2014.

Lincoff AM et al: Cardiovascular safety of testosterone-replacement therapy. N Engl J Med 389:107, 2023.

Mäkelä JA et al: Testis development. Endocr Rev 40:857, 2019.

O'Shaughnessy PJ et al: Alternative (backdoor) androgen production and masculinization in the human fetus. PLoS Biol 17:e3000002, 2019.

Patel B et al: The emerging therapeutic potential of kisspeptin and neurokinin B. Endocr Rev 45:30, 2024.

Pope HG Jr et al: Adverse health consequences of performance-enhancing drugs: An Endocrine Society scientific statement. Endocr Rev 35:341, 2014.

Snyder PJ et al: Effects of testosterone treatment in older men. N Engl J Med 74:611, 2016.

Stamou MI et al: Kallmann syndrome: Phenotype and genotype of hypogonadotropic hypogonadism. Metabolism 86:124, 2018.

Travison TG et al: Harmonized reference ranges for circulating testosterone levels in men of four cohort studies in the USA and Europe. J Clin Endocrinol Metab 102:1161, 2017.

Zakharov MN et al: A multi-step, dynamic allosteric model of testosterone's binding to sex hormone binding globulin. Mol Cell Endocrinol 399:190, 2015.

404

Disorders of the Female Reproductive System

Janet E. Hall, Anuja Dokras

The female reproductive system regulates the hormonal changes responsible for puberty and adult reproductive function. Normal reproductive function in women requires the dynamic integration of hormonal signals from the hypothalamus, pituitary, and ovary, resulting in repetitive cycles of follicle development, ovulation, and preparation of the endometrial lining of the uterus for implantation should conception occur.

For further discussion of related topics, see the following chapters: amenorrhea and pelvic pain **(Chap. 405)**, infertility and contraception **(Chap. 408)**, menopause **(Chap. 407)**, disorders of sex development **(Chap. 402)**, and disorders of the male reproductive system **(Chap. 403)**.

DEVELOPMENT OF THE OVARY AND EARLY FOLLICULAR GROWTH

The ovary orchestrates the development and release of a mature oocyte and secretes hormones (e.g., estrogen, progesterone, inhibins A and B, relaxin) that play critical roles in a variety of target tissues, including breast, bone, and uterus, in addition to the hypothalamus and pituitary. To achieve these functions in repeated monthly cycles, the ovary undergoes some of the most dynamic changes of any organ in the body. Primordial germ cells can be identified by the third week of gestation, and their migration to the genital ridge is complete by 6 weeks of gestation. Germ cells persist within the genital ridge, are then referred to as *oogonia*, and are essential for induction of ovarian development. In patients with 45,X Turner syndrome, primordial germ cells proliferate and migrate to the genital ridge but do not persist because their survival requires pregranulosa cells that are dependent on the presence of both X chromosomes **(Chap. 402)**.

The germ cell population expands, and starting at ~8 weeks of gestation, oogonia begin to enter prophase of the first meiotic division and become primary oocytes. Entry into meiosis provides some degree of protection from programmed cell death. It also allows the oocyte to be surrounded by a single layer of flattened granulosa cells to form a primordial follicle. Granulosa cells are derived from mesonephric cells that migrate into the ovary early in its development, pushing the germ cells to the periphery. Although there is evidence that both oocyte-like cells and follicle-like structures can form from embryonic stem cells in culture, there is no clear evidence that this occurs in vivo, and thus, the ovary appears to contain a nonrenewable pool of germ cells. Through

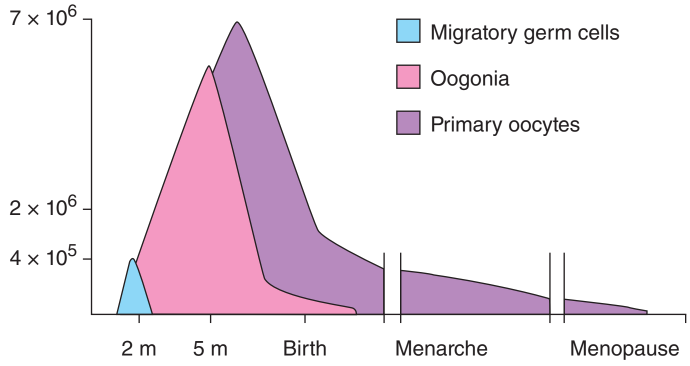

FIGURE 404-1 Ovarian germ cell number is maximal at mid-gestation and decreases precipitously thereafter.

the combined processes of mitosis, meiosis, and atresia, the population of oogonia reaches its maximum of 6–7 million by 20 weeks in the fetus, after which there is a progressive loss of both oogonia and primordial follicles through the process of atresia. At birth, oogonia are no longer present in the ovary, and only 1–2 million germ cells remain in the form of primordial follicles **(Fig. 404-1)**. The oocyte persists in prophase of the first meiotic division until just before ovulation, when meiosis resumes.

The quiescent primordial follicles are recruited to further growth and differentiation through a highly regulated process that limits the size of the developing cohort to ensure that folliculogenesis can continue throughout the reproductive life span. Transformation of primordial follicles to form primary follicles **(Fig. 404-2)** is characterized by growth of the oocyte and the transition from squamous to cuboidal granulosa cells. The theca interna cells that surround the developing follicle begin to form as the primary follicle grows. Acquisition of a zona pellucida by the oocyte, and the presence of several layers of surrounding cuboidal granulosa cells, further surrounded by theca cells, mark the development of secondary follicles. It is at this stage that granulosa cells develop follicle-stimulating hormone (FSH), estradiol, and androgen receptors and communicate with one another through the development of gap junctions.

Bidirectional signaling between the germ cells and the somatic cells in the ovary is a necessary component underlying the maturation of the oocyte and the capacity for hormone secretion. Members of the transforming growth factor beta (TGF-β) family of proteins are involved in this bidirectional signaling; oocyte-derived growth differentiation factor 9 (GDF-9) and bone morphogenic protein-15 (BMP-15), also known as GDF-9b, are required for migration of pregranulosa and pretheca cells to the outer surface of the developing follicle initial follicle formation. GDF-9 is also required for formation of secondary follicles, as are granulosa cell–derived KIT ligand (KITL) and the forkhead transcription factor (FOXL2). A significant number of genes have been identified that are required for development of the normal complement of oogonia in the ovary, initial follicle development, and resistance to follicle loss; all are candidates for premature ovarian insufficiency (POI), and mutations in >50 genes have been identified in patients with POI, with even more that have been associated with an earlier age at natural menopause.

DEVELOPMENT OF A MATURE FOLLICLE

The early stages of follicle growth are primarily driven by intraovarian factors. Further maturation to the preovulatory stage, including the resumption of meiosis in the oocyte, requires the combined stimulus of FSH and luteinizing hormone (LH) **(Figs.** 404-2 and **404-3)**. Recruitment from the resting pool of secondary follicles, termed *cyclic recruitment*, requires the direct action of FSH, whereas anti-müllerian hormone (AMH) produced from small growing preantral follicles restrains this effect of FSH, controlling the number of follicles entering the actively growing pool. Accumulation of follicular fluid between the layers of granulosa cells creates an antrum that divides the granulosa cells into two functionally distinct groups: mural cells that line the follicle wall and cumulus cells that surround the oocyte (Fig. 404-3). As the follicle develops into a preovulatory, or Graffian, follicle, mural granulosa cells in proximity to theca cells express the greatest steroidogenic activity. Cumulus cells surround the oocyte and express mammalian target of rapamycin (mTOR), which regulates cellular metabolism and increases the transfer of nutrients to the oocyte. In addition to its role in normal development of the müllerian system, the WNT signaling pathway is required for normal antral follicle development and may also play a role in ovarian steroidogenesis. Recruitment to the small antral stage generally occurs over several cycles with further growth to follicle sizes of >4–7 mm in waves during a single cycle. However, recruitment over a single cycle can occur as evidenced by a normal follicular phase length in gonadotropin-releasing hormone (GnRH)-deficient women in response to a first cycle of treatment with a physiologic regimen of pulsatile GnRH administration and normalization of FSH and LH. A single dominant follicle emerges from the growing follicle pool within the first 5–7 days after the onset of menses, while the majority of follicles

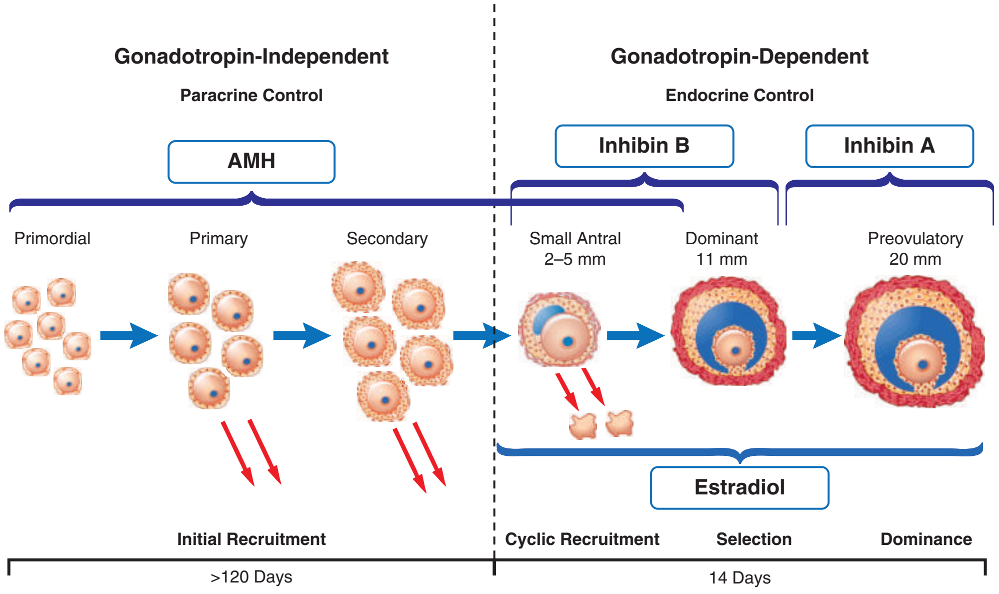

FIGURE 404-2 Gonadotropin-independent and gonadotropin-dependent ovarian follicle development that ultimately results in ovulation of a mature oocyte. AMH, anti-müllerian hormone. *(Reproduced with permission from Donna Jeanne Corcoran.)*

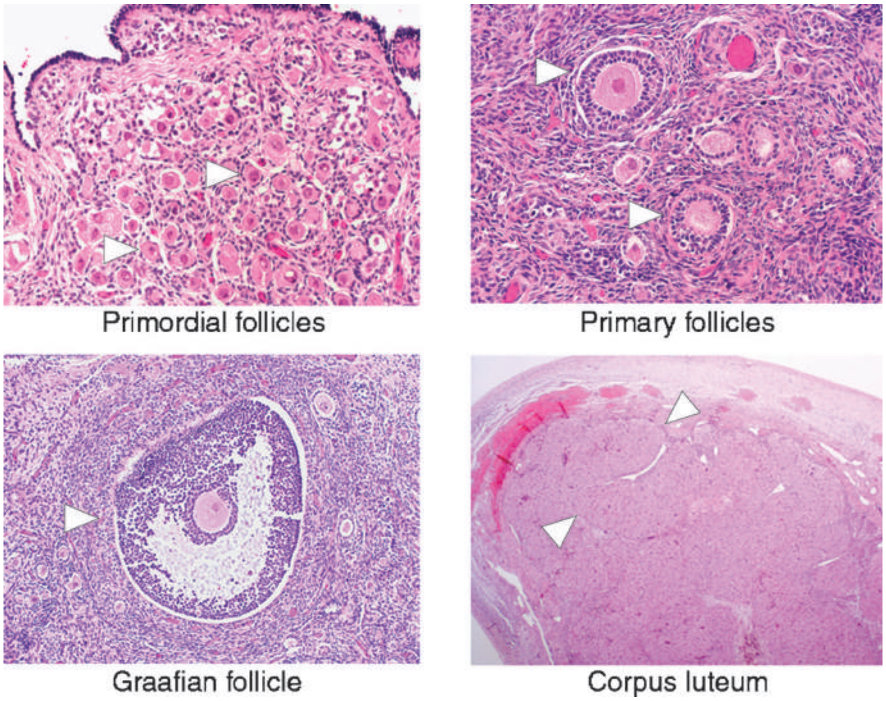

FIGURE 404-3 Development of ovarian follicles. The Graafian follicle is also known as a tertiary or preovulatory follicle. *(Courtesy of JH Eichhorn and D Roberts, Massachusetts General Hospital; with permission.)*

fall off their growth trajectory and become atretic. Autocrine actions of activin and BMP-6, derived from the granulosa cells, and paracrine actions of GDF-9, BMP-15, BMP-6, and Gpr149, derived from the oocyte, are involved in granulosa cell proliferation and modulation of FSH responsiveness. Differential exposure to these factors, and to vascular endothelial growth factor (VEGF), can alter vascular density and permeability, likely explaining the mechanism whereby a given follicle is selected for continued growth to the preovulatory stage. The dominant follicle can be distinguished by its size, evidence of granulosa cell proliferation, large number of FSH receptors, high aromatase activity, and elevated concentrations of estradiol and inhibin A in follicular fluid. In addition, secretion of estradiol and inhibin from the dominant follicle inhibits FSH and the growth of other follicles.

The dominant follicle undergoes rapid expansion during the 5–6 days prior to ovulation, reflecting granulosa cell proliferation and accumulation of follicular fluid. FSH induces LH receptors on the granulosa cells, and the preovulatory, or Graafian, follicle moves to the outer ovarian surface in preparation for ovulation. The LH surge triggers the resumption of meiosis, the suppression of granulosa cell proliferation, and the induction of cyclooxygenase 2 (COX-2), prostaglandins, the progesterone receptor (PR), and the epidermal growth factor–like growth factors amphiregulin, epiregulin, betacellulin, and neuroregulin 1, all of which are required for ovulation. Ovulation requires production of extracellular matrix, leading to expansion of the cumulus cell population that surrounds the oocyte and the controlled expulsion of the egg and follicular fluid. Both progesterone and prostaglandins (induced by the ovulatory stimulus) are essential for this process, as are members of the matrix metalloproteinase family. After ovulation, luteinization of theca and granulosa cells is induced by LH in conjunction with the acquisition of a rich vascular network in response to VEGF and basic fibroblast growth factor (FGF). Traditional regulators of central reproductive control, GnRH and its receptor (GnRHR), as well as kisspeptin, are also produced in the ovary and may be involved in corpus luteum function.

REGULATION OF OVARIAN FUNCTION

■ HYPOTHALAMIC AND PITUITARY SECRETION

GnRH neurons derive from cells in the olfactory placode and, to a lesser extent, the neural crest. They migrate into the brain across the cribriform plate along with the olfactory neurons which then form the olfactory bulb, while the GnRH neurons continue their journey to the hypothalamus. The importance of the initial migratory pathway into the brain is evidenced in patients born without a nose (bosma arhinia microphthalmia syndrome) who lack olfactory foramina in the cribriform plate and are both anosmic and have hypogonadotropic hypogonadism. Studies in these and other GnRH-deficient patients who fail to undergo puberty have provided insights into genes that control the ontogeny and function of GnRH neurons **(Fig. 404-4)**. *ANOS1* (also known as *KAL1*), *FGF8/FGFR1*,

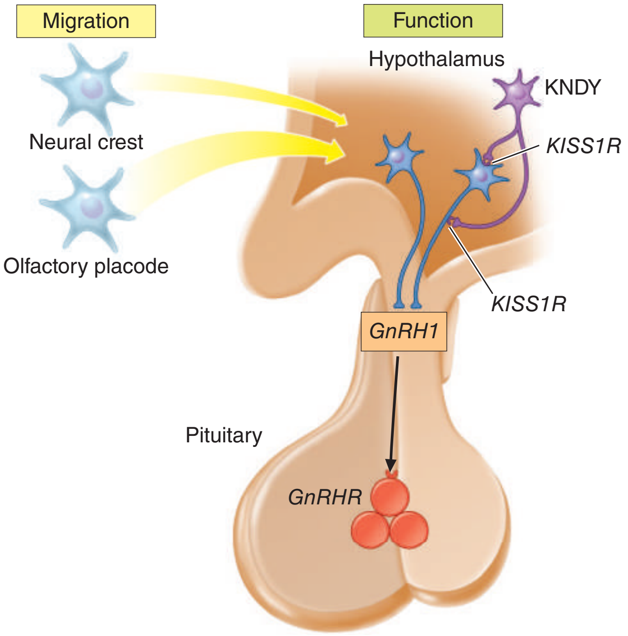

FIGURE 404-4 Genetic studies in patients with congenital forms of hypogonadotropic hypogonadism have expanded our understanding of the development and migration of gonadotropin-releasing hormone (GnRH) neurons from the olfactory placode and possibly the neural crest to the hypothalamus as well as the upstream regulation of GnRH secretion by kisspeptin (*KISS1*), neurokinin B (*TAC3*), and dynorphin (*Dyn*), which are co-expressed in the KNDY neurons. GnRHR, GnRH receptor.

PROK2/PROKR2, NSMF, HS6SD1, and *CDH7*, among a host of others **(Chap. 403)**, have been implicated in the migration of GnRH neurons to the hypothalamus, whereas *KISS*, *TAC3*, *Dyn*, and their receptors are involved in the upstream regulation of GnRH secretion. Approximately 7000 GnRH neurons, scattered throughout the medial basal hypothalamus, establish contacts with capillaries of the pituitary portal system in the median eminence. GnRH is secreted into the pituitary portal system in discrete pulses to stimulate synthesis and secretion of LH and FSH from pituitary gonadotropes, which comprise ~10% of cells in the pituitary **(Chap. 390)**. Functional connections of GnRH neurons with the portal system are established by the end of the first trimester, coinciding with the production of pituitary gonadotropins. Thus, like the ovary, the hypothalamic and pituitary components of the reproductive system are present before birth. However, the high levels of estradiol and progesterone produced by the placenta suppress hypothalamic-pituitary stimulation of ovarian hormonal secretion in the fetus.

After birth and the loss of placenta-derived steroids, gonadotropin levels rise. FSH levels are much higher in girls than in boys. The rise in FSH in girls results in circulating estradiol and inhibin B in what has been termed the mini-puberty of infancy, but without terminal follicle maturation or ovulation. Studies that have identified mutations in *TAC3*, which encodes neurokinin B, and its receptor, *TAC3R*, in patients with GnRH deficiency indicate that both are involved in control of GnRH secretion and may be particularly important at this early stage of development. By 12–20 months of age, the reproductive axis is again suppressed, and a period of relative quiescence persists until puberty **(Fig. 404-5)**. At the onset of puberty, pulsatile GnRH secretion induces pituitary gonadotropin production. In the early stages of puberty, LH and FSH secretion are apparent only during sleep, but as puberty develops, pulsatile LH secretion, a faithful marker of GnRH secretion, occurs throughout the day and night.

The mechanisms responsible for the childhood quiescence and pubertal reactivation of the reproductive axis remain incompletely understood. GnRH neurons in the hypothalamus respond to both excitatory and inhibitory factors. Increased sensitivity to the inhibitory influence of gonadal steroids has long been implicated in the inhibition of GnRH secretion during childhood but has not been definitively established in the human. Metabolic signals, including adipocyte-derived leptin, play a permissive role in reproductive function but are not sufficient to induce puberty **(Chap. 413)**.

Studies of patients with isolated GnRH deficiency reveal that mutations in the G protein–coupled receptor 54 (*GPR54*) gene (now known as *KISS1R*) preclude the onset of puberty. Kisspeptin, the ligand for this receptor, is derived from the parent peptide, kisspeptin-1 (*KISS1*), and is a potent stimulant for GnRH release. The tachykinin 3 gene (*TAC3*), which encodes neurokinin-B (NKB), stimulates GnRH secretion through kisspeptin signaling, while dynorphin (*Dyn*), which acts mainly through kappa opioid receptors, plays an inhibitory role in GnRH control. *TAC3* and *Dyn* are frequently co-expressed with *KISS1* in KNDy neurons of the median eminence that project to GnRH neurons. This system is intimately involved in both estrogen and progesterone negative feedback regulation of GnRH secretion as well as metabolic and stress signaling to the reproductive axis. Kisspeptin release is pulsatile, and secretion is generated by alternate stimulation and inhibition of kisspeptin by NKB and Dyn, respectively. While pulsatility is an intrinsic property of GnRH neurons, the current model suggests that pulsatile secretion of GnRH is coordinated by pulsatile kisspeptin secretion, with potential input from upstream glutaminergic neurons.

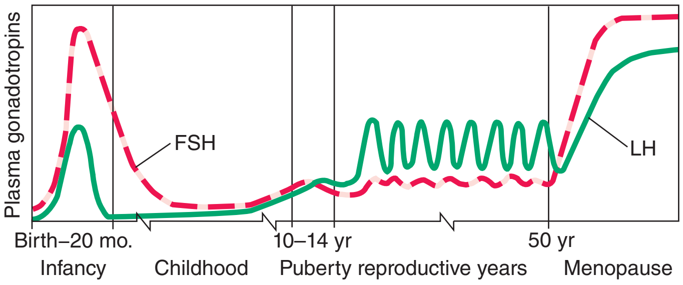

FIGURE 404-5 Follicle-stimulating hormone (FSH) and luteinizing hormone (LH) are increased during the neonatal years but go through a period of childhood quiescence before increasing again during puberty. Gonadotropin levels are cyclic during the reproductive years and increase dramatically with the loss of negative feedback that accompanies menopause.

A role for kisspeptin in the onset of puberty is suggested by upregulation of *KISS1* and *KISS1R* transcripts in the hypothalamus at the time of puberty. The onset of puberty may occur via a switch from epigenetic repression to epigenetic activation in kisspeptin neurons. Two imprinted genes that were initially identified in families with central precocious puberty, *MKRN3* and *DLK1*, may be involved in this switch given their known functions and their association with age at onset of menarche in a large cohort of women.

While short-term infusion of kisspeptin restored LH pulsatility in patients with hypothalamic amenorrhea and hyperprolactinemia, there is also evidence that tachyphylaxis occurs, potentially limiting its use as a treatment for these conditions.

RFamide-related peptides (RFRPs) are the mammalian orthologues of gonadotropin inhibitory hormone (GnIH), which was initially discovered in the quail. In lower animal species, these peptides decrease gonadal function and sexual motivation in addition to increasing feeding behavior and mediating the inhibitory actions of stress on reproduction. While RFRP-1 and RFRP-3 neurons send axonal projections to GnRH neurons in humans and RFRPs are secreted into the pituitary portal system, further studies are required to determine their potential physiologic role in the human.

OVARIAN STEROIDS

Ovarian steroid-producing cells do not store hormones but produce them in response to FSH and LH during the normal menstrual cycle. The sequence of steps and the enzymes involved in the synthesis of steroid hormones are similar in the ovary, adrenal, and testis. However, the enzymes required to catalyze specific steps are compartmentalized and may not be abundant or even present in all cell types. Within the developing ovarian follicle, estrogen synthesis from cholesterol requires close integration between theca and granulosa cells—the two-cell model for steroidogenesis **(Fig. 404-6)**. FSH receptors are confined to the granulosa cells, whereas LH receptors are restricted to the theca

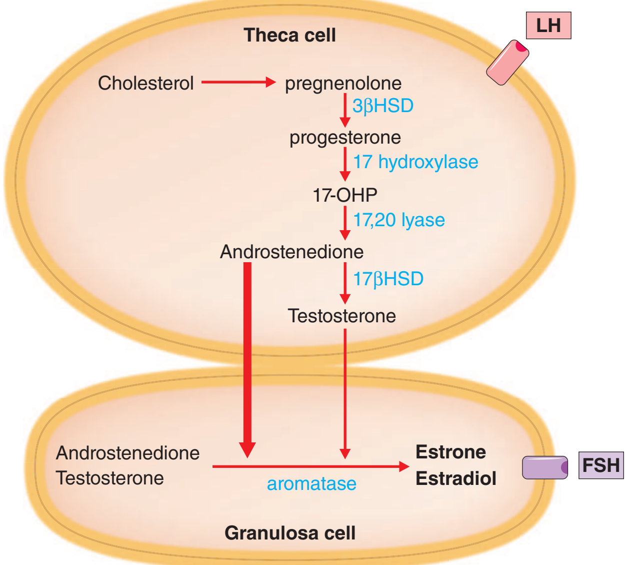

FIGURE 404-6 Estrogen production in the ovary requires the cooperative function of the theca and granulosa cells under the control of luteinizing hormone (LH) and follicle-stimulating hormone (FSH). HSD, hydroxysteroid dehydrogenase; OHP, hydroxyprogesterone.

cells until the late stages of follicular development, when they are also found on granulosa cells. The theca cells surrounding the follicle are highly vascularized and use cholesterol, derived primarily from circulating lipoproteins, as the starting point for the synthesis of androstenedione and testosterone under the control of LH. There is some evidence that 11-oxo-androgens, which are generally thought to be produced only in the adrenal, may also be produced to some degree in the ovary. These steroid precursors cross the basal lamina to the granulosa cells, which receive no direct blood supply. The mural granulosa cells are particularly rich in aromatase and, under the control of FSH, produce estradiol, the primary steroid secreted from the follicular phase ovary and the most potent estrogen. Theca cell–produced androstenedione and, to a lesser extent, testosterone are also secreted into peripheral blood, where they can be converted to dihydrotestosterone in skin and to estrogens in adipose tissue. The hilar interstitial cells of the ovary are functionally similar to Leydig cells and are also capable of secreting androgens. Stromal cells proliferate in response to androgens (as in polycystic ovary syndrome [PCOS]) but do not secrete androgens. However, high levels of androgens may be produced by luteinized theca cells within the stroma in women with hyperthecosis.

Development of the rich capillary network following rupture of the follicle at the time of ovulation makes it possible for large molecules such as low-density lipoprotein (LDL) to reach the luteinized granulosa and theca lutein cells. As in the follicle, both cell types are required for steroidogenesis in the corpus luteum. The luteinized granulosa cells are the main source of progesterone production, whereas the smaller theca lutein cells produce 17-hydroxyprogesterone and androgenic substrates for aromatization to estradiol by the luteinized granulosa cells. Production of estrogen metabolites by the corpus luteum plays a significant role in maintenance of the vascularization required for its function. LH is critical for formation and maintenance of corpus luteum structure and function. LH and human chorionic gonadotropin (hCG) bind to a common receptor; thus, in conception cycles, hCG produced upon fertilization rescues the declining function of the corpus luteum, maintaining steroid and peptide secretion for the first 10 weeks of pregnancy. hCG is commonly used for luteal phase support in the treatment of infertility.

Steroid Hormone Actions Both estrogen and progesterone play critical roles in the expression of secondary sexual characteristics in women **(Chap. 389)**. Estrogen promotes development of the ductule system in the breast, whereas progesterone is responsible for glandular development. In the reproductive tract, estrogens create a receptive environment for fertilization and support pregnancy and parturition through carefully coordinated changes in the endometrium, thickening of the vaginal mucosa, thinning of the cervical mucus, and uterine growth and contractions. Progesterone induces secretory activity in the estrogen-primed endometrium, increases the viscosity of cervical mucus, and inhibits uterine contractions. Both gonadal steroids play critical roles in negative and positive feedback of gonadotropin secretion. Progesterone also increases basal body temperature, which is used clinically as a marker of ovulation.

The vast majority of circulating estrogens and androgens are carried in the blood bound to carrier proteins, which restrain their free diffusion into cells and prolong their clearance, serving as a reservoir. High-affinity binding proteins include sex hormone–binding globulin (SHBG), which binds androgens with somewhat greater affinity than estrogens, and corticosteroid-binding globulin (CBG), which also binds progesterone. Modulations in binding protein levels by insulin, androgens, and estrogens contribute to high bioavailable testosterone levels in PCOS and to high circulating total estrogen and progesterone levels during pregnancy.

Estrogens act primarily through binding to the nuclear receptors, estrogen receptors (ER) α and β. Transcriptional coactivators and co-repressors modulate ER action **(Chap. 389)**. Both ER subtypes are present in the hypothalamus, pituitary, ovary, and reproductive tract. Although ERα and ERβ exhibit some functional redundancy, there is also a high degree of specificity, particularly in expression within cell types. For example, ERα functions in ovarian theca cells, whereas ERβ is critical for granulosa cell function. There is also evidence for membrane-initiated signaling by estrogen. Similar signaling mechanisms pertain for progesterone with evidence of transcriptional regulation through PR-A and PR-B protein isoforms, as well as rapid membrane signaling.

OVARIAN PEPTIDES

Inhibin was initially isolated from gonadal fluids based on its ability to selectively inhibit FSH secretion from pituitary cells. Inhibin is a heterodimer composed of an α subunit and a βA or βB subunit to form inhibin A or inhibin B, both of which are secreted from the ovary. Activin is a homodimer of inhibin β subunits with the capacity to stimulate the synthesis and secretion of FSH. Inhibins and activins are members of the TGF-β superfamily of growth and differentiation factors. During the purification of inhibin, follistatin, an unrelated monomeric protein that inhibits FSH secretion, was discovered. Within the pituitary, follistatin inhibits FSH secretion indirectly by binding and neutralizing activin.

Inhibin B is constitutively secreted from the granulosa cells of small antral follicles, and its serum levels increase in conjunction with granulosa cell proliferation during recruitment of secondary follicles under the control of FSH (Fig. 404-2). Inhibin B is an important inhibitor of FSH, independent of estradiol, during the menstrual cycle. Inhibin A is present in both granulosa and theca cells and is secreted by the dominant follicle. Inhibin A is also present in luteinized granulosa cells and is a major secretory product of the corpus luteum. Synthesis and secretion of inhibin A are directly controlled by FSH and LH. Although activin is also secreted from the ovary, the excess of follistatin in serum, combined with its nearly irreversible binding of activin, make it unlikely that ovarian activin plays an endocrine role in FSH regulation. However, there is evidence that activin plays an autocrine/paracrine role in the ovary in germ cell survival, follicle assembly, and inhibition of androgen production, in addition to its intrapituitary role in modulation of FSH production.

AMH (also known as müllerian-inhibiting substance) is important in ovarian biology in addition to the function from which it derived its name (i.e., promotion of the degeneration of the müllerian system during embryogenesis in the male). AMH is produced by granulosa cells from small preantral and early antral follicles (Fig. 404-2) and is a marker of ovarian reserve with advantages over inhibin B because of its relative stability across the menstrual cycle. AMH inhibits FSH in the recruitment of primordial follicles into the follicle pool and counters FSH stimulation of aromatase expression. AMH levels increase during puberty, are highest in the early twenties, and decrease markedly by menopause. AMH is increased in PCOS in conjunction with the abundance of small follicles in this disorder.

Gonadotropin surge attenuating factor (GnSAF) is an ovarian factor that attenuates GnRH-induced gonadotropin secretion. Its role is not yet fully understood, but there is an inverse relationship between GnSAF and follicle size, suggesting that its primary role involves the early stages of follicle development rather than curtailing the gonadotropin surge as its name implies.

Relaxin is produced primarily by the theca lutein cells of the corpus luteum. Both relaxin and its receptor, encoded by the gene *RXFP1*, are highly expressed in the uterus during the peri-implantation period and its primary role appears to be in promoting decidualization and vascularization of the endometrium prior to implantation. Relaxin was named for its ability to suppress myometrial contractility in pigs and rodents, but it does not appear to exert this activity in women.

HORMONAL INTEGRATION OF THE NORMAL MENSTRUAL CYCLE

The sequence of changes responsible for mature reproductive function is coordinated through a series of negative and positive feedback loops that alter pulsatile GnRH secretion, the pituitary response to GnRH, and the relative secretion of LH and FSH from gonadotropes **(Fig. 404-7)**. The frequency and amplitude of pulsatile GnRH secretion differentially modulate the synthesis and secretion of LH and FSH. Slow GnRH pulse frequencies favor FSH synthesis, whereas increased

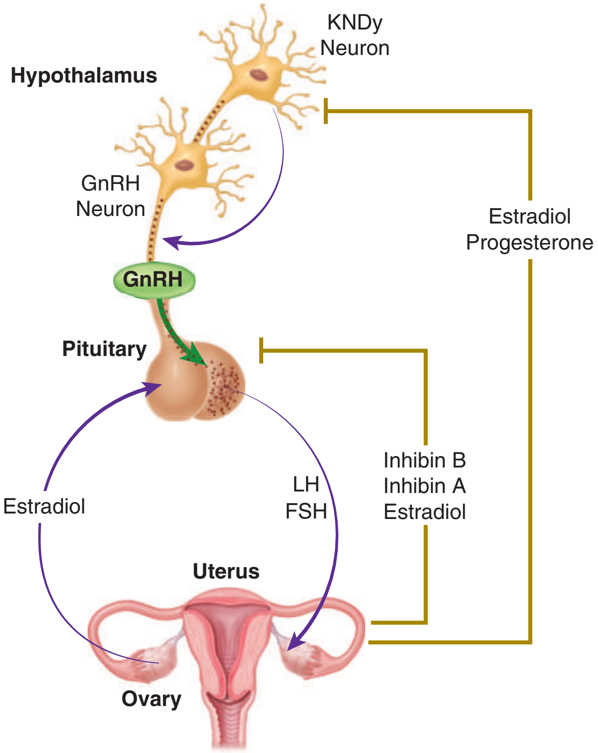

FIGURE 404-7 The reproductive system in women is critically dependent on both negative feedback of gonadal steroids and inhibin to modulate follicle-stimulating hormone (FSH) secretion and on estrogen positive feedback to generate the preovulatory luteinizing hormone (LH) surge. GnRH, gonadotropin-releasing hormone.

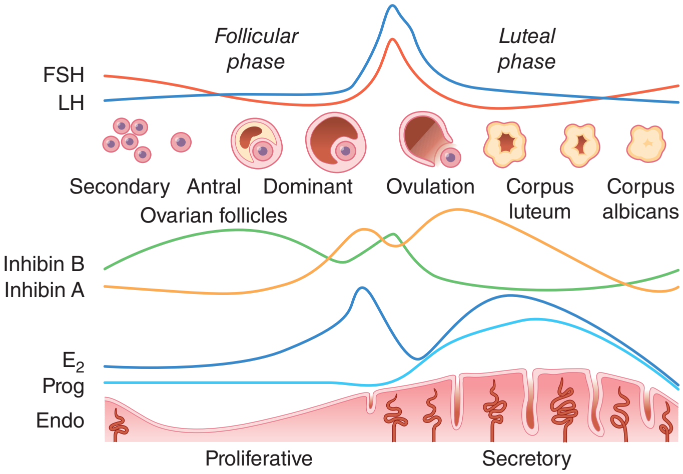

FIGURE 404-8 Relationship between gonadotropins, follicle development, gonadal secretion, and endometrial changes during the normal menstrual cycle. E_2, estradiol; Endo, endometrium; FSH, follicle-stimulating hormone; LH, luteinizing hormone; Prog, progesterone.

GnRH pulse frequency and amplitude favor LH synthesis. FSH synthesis is also controlled by the activin-inhibin-follistatin system within the pituitary. Activin is produced in both pituitary gonadotropes and folliculostellate cells and stimulates the synthesis and secretion of FSH through autocrine-paracrine mechanisms that are modulated by follistatin, which is also produced in folliculostellate cells. Inhibins function as potent antagonists of activins through sequestration of the activin receptors. Although inhibin is expressed in the pituitary, gonadal inhibin is the principal source of feedback inhibition of FSH.

For the majority of the cycle, the reproductive system functions in a classic endocrine negative feedback mode. Estradiol and progesterone inhibit GnRH secretion, acting through kisspeptin and dynorphin in the KNDy neurons, while the inhibins act at the pituitary to selectively inhibit FSH synthesis and secretion (Fig. 404-7). Estradiol also contributes to negative feedback at the pituitary with an effect that is greater for FSH than LH. This tightly regulated negative feedback control of FSH is critical for development of the single mature follicle that characterizes normal reproductive cycles in women. In addition to these negative feedback controls, the menstrual cycle is uniquely dependent on estrogen-induced positive feedback to produce an LH surge that is essential for ovulation of a mature follicle. Estrogen negative feedback in women occurs primarily at the hypothalamus with a small pituitary contribution, whereas estrogen positive feedback occurs at the pituitary in women with upregulation of GnRH signaling and responsiveness. In women, hypothalamic GnRH secretion plays a permissive role in generating the preovulatory gonadotropin surge, a mechanism that differs significantly from that in rodents and other species that rely on seasonal and circadian cues, in which a surge of GnRH also occurs.

THE FOLLICULAR PHASE

The follicular phase is characterized by *cyclic recruitment* of a cohort of secondary follicles and the ultimate selection of a preovulatory follicle **(Figs.** 404-2 **and 404-8)**. The follicular phase begins, by convention, on the first day of menses. However, follicle recruitment is initiated by the rise in FSH that begins in the late luteal phase of the previous cycle in conjunction with the loss of negative feedback of gonadal steroids and likely inhibin A. The fact that an ~20% increase in FSH is adequate for follicular recruitment speaks to the marked sensitivity of the resting follicle pool to FSH. The resultant granulosa cell proliferation is responsible for increasing early follicular phase levels of inhibin B. Inhibin B, in conjunction with rising levels of estradiol and inhibin A, restrains FSH secretion during this critical period such that only a single follicle matures in the vast majority of cycles. The increased risk of multiple gestation associated with the higher levels of FSH characteristic of advanced maternal age or with exogenous gonadotropin administration in the treatment of infertility attests to the importance of the precise negative feedback regulation of FSH that is necessary for monofollicular development. With further growth of the dominant follicle, estradiol and inhibin A increase. Increasing levels of estradiol across the follicular phase are responsible for proliferative changes in the endometrium. Acquisition of FSH-induced LH receptors on granulosa cells allows LH to complete the final stages of maturation of the preovulatory follicle, leading to secretion of low levels of progesterone and 17α-hydroxyprogesterone.

The exponential rise in estradiol in the mid-late luteal phase results in positive feedback on the pituitary gonadotropes, leading to generation of an LH surge (and a smaller FSH surge). The low levels of progesterone secreted from the peri-ovulatory follicle are not required for the gonadotropin surge but appear to play a role in its timing. The LH surge may precede follicle rupture by 24–36 h, during which time the oocyte uncouples from the granulosa cells, and genes involved in inflammation and tissue remodeling are induced. Further alterations in steroidogenesis accompany luteinization of theca and granulosa cells.

THE LUTEAL PHASE

The luteal phase begins with the formation of the corpus luteum from the ruptured follicle (Fig. 404-8). Progesterone, 17α-hydroxyprogesterone, and inhibin A are produced from the luteinized granulosa cells. The granulosa-lutein cells continue to aromatize androgen precursors derived from theca-lutein cells, producing estradiol. The combined actions of estrogen and progesterone, as well as relaxin, are responsible for the secretory changes in the endometrium that are necessary for implantation. The corpus luteum is supported by LH but has a finite life span. Both the progesterone-induced decrease in LH pulse frequency and diminished postreceptor signaling are likely to contribute to the demise of the corpus luteum. The progressive decline in hormonal support of the endometrium results in inflammation, local hypoxia and ischemia, and subsequent vascular changes with release of cytokines, cell death, and shedding of the endometrium.

If conception occurs, hCG produced by the trophoblast binds to LH receptors on the corpus luteum, maintaining steroid hormone production and preventing involution of the corpus luteum until the luteal-placental shift in hormone production that occurs 6–10 weeks after conception.

CLINICAL ASSESSMENT OF OVARIAN FUNCTION

Menstrual bleeding should become regular within 2–4 years of menarche, although anovulatory and irregular cycles are common before that. For the remainder of adult reproductive life, the cycle length counted from the first day of menses to the day preceding subsequent menses is ~28 days, with a range of 25–35 days. However, cycle-to-cycle variability for an individual woman is ±2 days. Luteal phase length is relatively constant between 12 and 14 days in normal cycles; thus, the major variability in cycle length is due to variations in follicular phase length. The duration of menstrual bleeding in ovulatory cycles varies between 4 and 6 days. There is a gradual shortening of cycle length with age such that women aged >35 years have cycles that are shorter than during their younger reproductive years. Anovulatory cycles increase as women approach menopause, and bleeding patterns may be erratic.

Women who report regular monthly bleeding generally have ovulatory cycles, but several other clinical signs can be used to assess the likelihood of ovulation. Some women experience *mittelschmerz*, described as midcycle pelvic discomfort that is thought to be caused by the rapid expansion of the dominant follicle at the time of ovulation. A constellation of premenstrual moliminal symptoms such as bloating, breast tenderness, mood changes, and food cravings often occur several days before menses in ovulatory cycles, but their absence cannot be used as evidence of anovulation. Methods that can be used to determine whether ovulation occurred include a serum progesterone level >3 ng/mL ~7 days after ovulation, an increase in basal body temperature of 0.24°C (>0.5°F) in the second half of the cycle due to the thermoregulatory effect of progesterone, or detection of the urinary LH surge using ovulation predictor kits, although the mere presence of an apparent LH surge does not guarantee that ovulation will occur. However, because ovulation occurs ~36 h after the LH surge, urinary LH can be helpful in timing intercourse to coincide with ovulation.

Ultrasound can be used to detect the growth of the fluid-filled antrum of the developing follicle and to assess endometrial thickness in response to increasing estradiol levels in the follicular phase. It can also be used to provide evidence of ovulation by documenting collapse of the dominant follicle and/or the presence of a corpus luteum as well as the characteristic echogenicity of the secretory endometrium of the luteal phase.

PUBERTY

NORMAL PUBERTAL DEVELOPMENT IN GIRLS

The first menstrual period (*menarche*) occurs relatively late in the series of developmental milestones that characterize normal pubertal development. Menarche is preceded by the appearance of pubic and then axillary hair (*adrenarche*) as a result of maturation of the zona reticularis in the adrenal gland and increased adrenal androgen secretion, particularly dehydroepiandrosterone (DHEA). The triggers for adrenarche remain unknown but may involve increases in body mass index, as well as in utero and neonatal factors. Menarche is also preceded by breast development (*thelarche*). The breast is exquisitely sensitive to the very low levels of estrogen that result from peripheral conversion of adrenal androgens and the low levels of estrogen secreted from the ovary early in pubertal maturation. This estrogen sensitivity also explains why infants occasionally develop breast tissue in response to endogenous or environmental estrogens. Breast development precedes the appearance of pubic and axillary hair in ~60% of girls. The interval between the onset of breast development and menarche is ~2 years. There has been a gradual decline in the age of puberty attributed to improved nutrition, however more recent changes indicate a decline in the age of thelarche but not menarche which is associated with obesity. In the United States, menarche occurs at an average age of 12.5 years.

Much of the variation in the timing of puberty is due to genetic factors. Heritability estimates from twin studies range between 50 and 80%. Adrenarche and thelarche occur ~1 year earlier in black girls compared with white girls, although the difference in the timing of menarche is less pronounced. Genome-wide association studies have identified over a hundred genes associated with pubertal timing in boys and girls, attesting to the high degree of coordination of this reproductive and growth milestone. These findings include genes involved in GnRH secretion (e.g., *TACR3*, and the maternally imprinted gene, *MKRN3*, that has been associated with familial precocious puberty), pituitary development and function (e.g., *POU1F1*), hormone synthesis and bioactivity (e.g., *STARD4, ESR1, RXRG*), gonadal feedback (e.g., *INHBA, ESR1*), and energy homeostasis and growth, including *LIN28B*, a sentinel puberty gene, which is a potent regulator of microRNA processing.

Other important hormonal changes also occur in conjunction with puberty. Growth hormone (GH) levels increase early in puberty, stimulated in part by the pubertal increase in estrogen secretion. GH increases insulin-like growth factor-1 (IGF-1), which enhances linear growth. The growth spurt is generally less pronounced in girls than in boys, with a peak growth velocity of ~7 cm/year. Linear growth is ultimately limited by closure of epiphyses in the long bones as a result of prolonged exposure to estrogen. Puberty is also associated with mild insulin resistance.

DISORDERS OF PUBERTY

The differential diagnosis of precocious and delayed puberty is similar in boys **(Chap. 403)** and girls. However, there are differences in the timing of normal puberty and differences in the relative frequency of specific disorders in girls compared with boys.

Precocious Puberty Traditionally, precocious puberty has been defined as the development of secondary sexual characteristics before the age of 8 in girls based on data from Marshall and Tanner in British girls studied in the 1960s. More recent studies led to recommendations that girls be evaluated for precocious puberty if breast development or pubic hair is present at <7 years of age for white girls or <6 years for black girls; however, these guidelines have not been widely accepted in favor of careful follow-up in all girls presenting at <8 years.

Precocious puberty in girls is most often centrally mediated **(Table 404-1)**, resulting from early activation of the hypothalamic-pituitary-ovarian axis. It is characterized by pulsatile LH secretion (which is initially associated with deep sleep) and an enhanced LH and FSH response to exogenous GnRH or a GnRH agonist (two- to threefold stimulation) **(Table 404-2)**. True precocity is marked by advancement in bone age of >2 standard deviations, a recent history of growth acceleration, and progression of secondary sexual characteristics.

TABLE 404-1 Differential Diagnosis of Precocious Puberty

CENTRAL (GnRH DEPENDENT)	PERIPHERAL (GnRH INDEPENDENT)
Idiopathic	Congenital adrenal hyperplasia
CNS tumors	Estrogen-producing tumors
Hamartomas	Adrenal tumors
Astrocytomas	Ovarian tumors
Adenomyomas	Gonadotropin/hCG-producing tumors
Gliomas	Exogenous exposure to estrogen or androgen or lavender or tea-tree oil
Germinomas	McCune-Albright syndrome
CNS infection	Aromatase excess syndrome
Genetic, i.e., *KISS1, KISS1R, MKRN3, DLK1*	
Head trauma	
Iatrogenic	
Radiation	
Chemotherapy	
Surgical	
CNS malformation	
Arachnoid or suprasellar cysts	
Septo-optic dysplasia	
Hydrocephalus	

Abbreviations: CNS, central nervous system; *DLK1*, delta-like 1 homolog gene; GnRH, gonadotropin-releasing hormone; hCG, human chorionic gonadotropin; *KISS1*, kisspeptin gene; *KISS1R*, kisspeptin receptor gene; *MKRN3*, makorin ring finger protein 3 gene.

TABLE 404-2 Evaluation of Precocious and Delayed Puberty

	PRECOCIOUS	DELAYED
Initial Screening Tests		
History and physical	×	×
Assessment of growth velocity	×	×
Bone age	×	×
LH, FSH	×	×
Estradiol, testosterone	×	×
DHEAS	×	×
17-Hydroxyprogesterone	×	
TSH, T_4	×	×
Complete blood count		×
Sedimentation rate, C-reactive protein		×
Electrolytes, renal function		×
Liver enzymes		×
IGF-1, IGFBP-3		×
Urinalysis		×
Secondary Tests		
Pelvic ultrasound	×	×
Cranial MRI	×	×
β-hCG	×	
GnRH/agonist stimulation test	×	×
ACTH stimulation test	×	
Inflammatory bowel disease panel	×	×
Celiac disease panel		×
Prolactin		×
Karyotype		×

Abbreviations: ACTH, adrenocorticotropic hormone; DHEAS, dehydroepiandrosterone sulfate; FSH, follicle-stimulating hormone; GnRH, gonadotropin-releasing hormone; hCG, human chorionic gonadotropin; IGF-1, insulin-like growth factor 1; IGFBP-3, IGF-binding protein 3; LH, luteinizing hormone; MRI, magnetic resonance imaging; TSH, thyroid-stimulating hormone; T_4, thyroxine.

In girls, centrally mediated precocious puberty (CPP) is categorized as idiopathic in ~85% of cases; however, neurogenic causes must be considered. Loss-of-function mutations in *MKRN3* and *DLK1*, both of which are imprinted genes, have been reported in familial CPP. Activating mutations in *KISS1, KISS1R,* and *PROKR2* have also been found in a small number of patients with CPP. However, the frequency of these mutations is insufficient to justify their use in routine clinical testing. GnRH agonists that induce pituitary desensitization are the mainstay of treatment to prevent premature epiphyseal closure and preserve adult height, as well as to manage psychosocial repercussions of precocious puberty.

Peripherally mediated precocious puberty does not involve activation of the hypothalamic-pituitary-ovarian axis and is characterized by suppressed gonadotropins in the presence of elevated estradiol. Management of peripheral precocious puberty involves treating the underlying disorder (Table 404-1) and limiting the effects of gonadal steroids using aromatase inhibitors, inhibitors of steroidogenesis, and ER blockers. It is important to be aware that central precocious puberty can also develop in girls whose precocity was initially peripherally mediated, as in McCune-Albright syndrome and congenital adrenal hyperplasia.

Incomplete and intermittent forms of precocious puberty may also occur. For example, premature breast development are common in girls before the age of 2 years, with no further progression and without significant advancement in bone age, estrogen production, or compromised height. Premature adrenarche can also occur in the absence of progressive pubertal development, but it must be distinguished from late-onset congenital adrenal hyperplasia and androgen-secreting tumors, in which case it may be termed *heterosexual precocity*. Premature adrenarche may be associated with obesity, hyperinsulinemia, and the subsequent predisposition to PCOS.

Delayed Puberty Delayed puberty (Table 404-3) is defined as the absence of secondary sexual characteristics by age 13 in girls. The diagnostic considerations are very similar to those for primary amenorrhea (Chap. 405). Between 25 and 40% of delayed puberty in girls is of ovarian origin, with Turner syndrome accounting for the majority of such patients. Delayed puberty may occur in the setting of systemic illnesses, including celiac disease and chronic renal disease, and endocrinopathies such as diabetes and hypothyroidism. In addition, girls appear to be particularly susceptible to the adverse effects of decreased energy balance resulting from exercise, dieting, and/or eating disorders, and thus, functional hypothalamic amenorrhea (HA) can present with primary amenorrhea. Together, these reversible conditions account for ~25% of delayed puberty in girls. Congenital hypogonadotropic hypogonadism in girls or boys can be caused by mutations in several different genes or combinations of genes (Fig. 404-4, Chap. 391, Table 404-3).

TABLE 404-3 Differential Diagnosis of Delayed Puberty

Hypergonadotropic

- Ovarian
  - Turner's syndrome
  - Gonadal dysgenesis
  - Chemotherapy/radiation therapy
  - Galactosemia
  - Autoimmune oophoritis
  - Congenital lipoid hyperplasia
- Steroidogenic enzyme abnormalities
  - 17α-Hydroxylase deficiency
  - Aromatase deficiency
- Gonadotropin/receptor mutations
  - *FSHβ, LHR, FSHR*
- Androgen resistance syndrome

Hypogonadotropic

- Genetic
  - Hypothalamic syndromes
    - Leptin/leptin receptor
    - *HESX1* (septo-optic dysplasia)
    - *PC1* (prohormone convertase)
  - IHH and Kallmann's syndrome
    - *KAL1, FGF8, FGFR1, NSMF, PROK2, PROKR2, SEM3A, HS6ST1, WDR11, CHD7*
    - *KISS1, KISS1R, TAC3, TAC3R, GnRH1, GnRHR,* and others
  - Abnormalities of pituitary development/function
    - *PROP1*
- CNS tumors/infiltrative disorders
  - Craniopharyngioma
  - Astrocytoma, germinoma, glioma
  - Prolactinomas, other pituitary tumors
  - Histiocytosis X
- Chemotherapy/radiation
- Functional
  - Chronic diseases
  - Malnutrition
  - Excessive exercise
  - Eating disorders

Abbreviations: CHD7, chromodomain-helicase-DNA-binding protein 7; CNS, central nervous system; *FGF8*, fibroblast growth factor 8; *FGFR1*, fibroblast growth factor 1 receptor; *FSH*β, follicle-stimulating hormone β chain; *FSHR*, FSH receptor; *GNRHR*, gonadotropin-releasing hormone receptor; *HESX1*, homeobox, embryonic stem cell expressed 1; *HS6ST1*, heparin sulfate 6-O sulfotransferase 1; IHH, idiopathic hypogonadotropic hypogonadism; KAL, Kallmann; *KISS1*, kisspeptin 1; *KISSR1*, KISS1 receptor; *LHR*, luteinizing hormone receptor; *NSMF*, NMDA receptor synaptonuclear signaling and neuronal migration factor; *PROK2*, prokineticin 2; *PROKR2*, prokineticin receptor 2; *PROP1*, prophet of Pit1, paired-like homeodomain transcription factor; *SEMA3A*, semaphorin-3A; *WDR11*, WD repeat-containing protein 11.

Approximately 50% of girls with congenital hypogonadotropic hypogonadism, with or without anosmia, have a history of some degree of breast development, and 10% report one to two episodes of vaginal bleeding. Family studies suggest that genes identified in association with absent puberty may also cause delayed puberty, and recent reports have further suggested that a genetic susceptibility to environmental stresses such as diet and exercise may account for at least some cases of functional HA, including in girls who present with primary amenorrhea. Although neuroanatomic causes of delayed puberty are considerably less common in girls than in boys, it is always important to rule these out in the setting of primary hypogonadotropic hypogonadism.

■ FURTHER READING

Balasubramanian R, Crowley WF Jr: Isolated gonadotropin-releasing hormone (GnRH) deficiency. 2007 May 23 [Updated 2022 May 12], in GeneReviews® [Internet]. Adam MP et al (eds). Seattle, WA, University of Washington, Seattle, 1993-2025. Available from: *https://www.ncbi.nlm.nih.gov/books/NBK1334/*.

Brito VN et al: The congenital and acquired mechanisms implicated in the etiology of central precocious puberty. Endocr Rev 44:193, 2023.

Cedars MI: Evaluation of female fertility-AMH and ovarian reserve testing. J Clin Endocrinol Metab 107:1510, 2022.

Lippincott MF et al: HU206 constitutional delay of puberty and idiopathic hypogonadotropic hypogonadism: differential contributions of common genetic variants. J Clin Endocrinol Metab 7:bvad114.1457, 2023.

Louden ED et al: Genetics of hypogonadotropic hypogonadism: Human and mouse genes, inheritance, oligogenicity, and genetic counseling. Mol Cell Endocrinol 534:111334, 2021.

Moore AM et al: KNDy neurons of the hypothalamus and their role in GnRH pulse generation: An update. Endocrinology 165:bqad194, 2023.

405

Menstrual Disorders and Pelvic Pain

Janet E. Hall, Anuja Dokras

Menstrual dysfunction can signal an underlying abnormality that may have long-term health consequences. Although frequent or prolonged bleeding usually prompts a woman to seek medical attention, infrequent or absent bleeding may seem less troubling, and the patient may not bring it to the attention of the physician. Thus, a focused menstrual history is a critical part of every encounter with a female patient. Pelvic pain is a common complaint that may relate to an abnormality of the reproductive organs but also may be of gastrointestinal, urinary tract, or musculoskeletal origin. Depending on its cause, pelvic pain may require urgent surgical attention. Recent guidelines no longer recommend routine pelvic examination in asymptomatic, average-risk women other than periodic cervical cancer screening. However, pelvic examination is an important part of the evaluation of amenorrhea, abnormal uterine bleeding, and pelvic pain.

MENSTRUAL DISORDERS

■ DEFINITION AND PREVALENCE

Amenorrhea refers to the absence of menstrual periods and is classified as *primary* if menstrual bleeding has never occurred in the absence of hormonal treatment or *secondary* if menstrual periods cease for 3–6 months. Primary amenorrhea is a rare disorder that occurs in <1% of the female population. However, between 3 and 5% of women experience at least 3 months of secondary amenorrhea in any specific year. There is no evidence that race or ethnicity influences the prevalence of amenorrhea. However, because of the importance of adequate nutrition for normal reproductive function, both the age at menarche and the prevalence of secondary amenorrhea vary significantly in different parts of the world.

Abnormal uterine bleeding (AUB) has replaced the term dysfunctional uterine bleeding and describes irregularities in the menstrual cycle involving frequency, cyclicity, duration, and volume of flow outside of pregnancy. A menstrual cycle typically occurs every 21–35 days, lasting between 4 and 7 days, with up to 80 mL of blood loss. Variations in any of these parameters constitutes a diagnosis of AUB, with up to one-third of women between menarche and menopause experiencing these symptoms. The acronym PALM-COEIN, which was developed to describe the etiologies for AUB, includes the structural causes (polyp, adenomyosis, leiomyoma [submucosal or other myoma], and malignancy and hyperplasia) and nonstructural causes (coagulopathy, ovulatory dysfunction, endometrial, iatrogenic, and not yet classified). Oligo- and anovulation are most frequently associated with polycystic ovary syndrome (PCOS).

Primary Amenorrhea The absence of menarche (the first menstrual period) by age 16 has been used traditionally to define primary amenorrhea. However, other factors, such as growth, secondary sexual characteristics, and the presence of cyclic pelvic pain, also influence the age at which primary amenorrhea should be investigated. Recent studies suggest that puberty is occurring at an earlier age, particularly in obese girls. However, it is important to note that these data reflect earlier breast development alone with minimal change in the age of menarche. Thus, an evaluation for amenorrhea should be initiated by age 15 or 16 in the presence of normal growth and secondary sexual characteristics; age 13 in the absence of secondary sexual characteristics or if height is less than the third percentile; age 12 or 13 in the presence of breast development and cyclic pelvic pain; or within 2 years of breast development if menarche has not occurred.

Secondary Amenorrhea or Oligomenorrhea Irregular cycles are relatively common for up to 3 years after menarche and for 1–2 years before the final menstrual period. In the intervening years, menstrual cycle length is ~28 days. Cycle-to-cycle variability in an individual woman who is ovulating consistently is generally +/– 2 days. Pregnancy should be excluded early in any evaluation of menstrual irregularity. However, many women occasionally miss a single period. Three months of secondary amenorrhea, or 6 months in women with previously irregular cycles, should prompt an evaluation, as should a history of intermenstrual intervals >35 or <21 days or bleeding that persists for >7 days.

■ DIAGNOSIS

Pregnancy is the most common cause of amenorrhea and must be excluded in all cases, regardless of patient history. Evaluation of menstrual dysfunction depends on understanding the interrelationships between the four critical components of the reproductive tract: (1) the hypothalamus, (2) the pituitary, (3) the ovaries, and (4) the uterus and outflow tract **(Fig. 405-1; Chap. 404)**. This system is maintained by complex negative and positive feedback loops involving the ovarian steroids (estradiol and progesterone) and peptides (inhibin B and inhibin A) and the hypothalamic (gonadotropin-releasing hormone [GnRH]) and pituitary (follicle-stimulating hormone [FSH] and luteinizing hormone [LH]) components of this system (Fig. 405-1).

Disorders of menstrual function fall into two main categories: disorders of the uterus and outflow tract and disorders of ovulation. Many of the conditions that cause primary amenorrhea are congenital but go unrecognized until the time of normal puberty (e.g., genetic, chromosomal, and anatomic abnormalities). All causes of secondary amenorrhea also can cause primary amenorrhea.

Disorders of the Uterus or Outflow Tract Abnormalities of the uterus and outflow tract typically present as primary amenorrhea. In patients with normal pubertal development and a blind vagina,

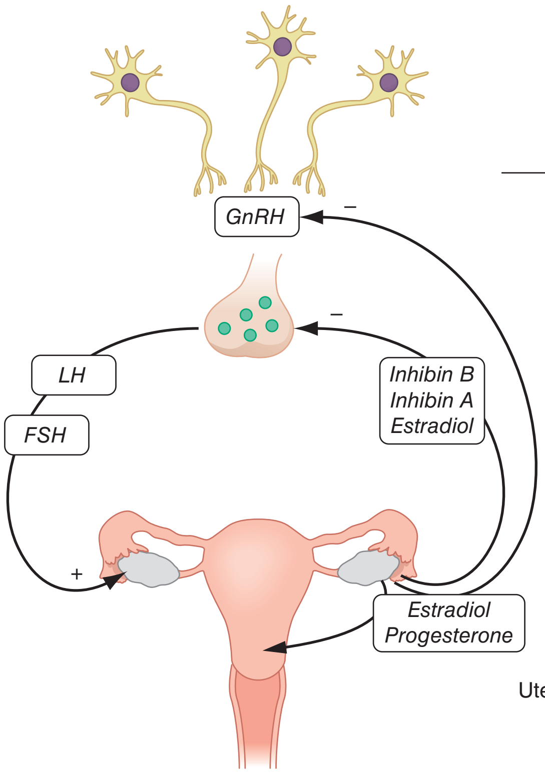

	Primary	Secondary
Hypothalamus	28%	36%
Pituitary	2%	15%
PCOS	8%	30%
Ovary	43%	12%
Uterus/outflow tract	19%	7%

FIGURE 405-1 Role of the hypothalamic-pituitary-gonadal axis in the etiology of amenorrhea. Gonadotropin-releasing hormone (GnRH) secretion from the hypothalamus stimulates follicle-stimulating hormone (FSH) and luteinizing hormone (LH) secretion from the pituitary to induce ovarian folliculogenesis and steroidogenesis. Ovarian secretion of estradiol and progesterone controls the shedding of the endometrium, resulting in menses, and, in combination with the inhibins, provides feedback regulation of the hypothalamus and pituitary to control secretion of FSH and LH. The prevalence of amenorrhea resulting from abnormalities at each level of the reproductive system (hypothalamus, pituitary, ovary, uterus, and outflow tract) varies depending on whether amenorrhea is primary or secondary. PCOS, polycystic ovarian syndrome.

the differential diagnosis includes *obstruction* by a transverse vaginal septum or imperforate hymen; *müllerian agenesis* (Mayer-Rokitansky-Kuster-Hauser syndrome), which can be caused by mutations in the *WNT4* gene, and *androgen insensitivity syndrome* (AIS), which is an X-linked recessive disorder that accounts for ~10% of all cases of primary amenorrhea (Chap. 403). Patients with AIS have a 46,XY karyotype, but because of the lack of androgen receptor responsiveness, those with complete AIS lack features of androgenization and have female external genitalia. The absence of pubic and axillary hair distinguishes them clinically from patients with müllerian agenesis, as does a testosterone level in the male range. The rare patient with 5α reductase type 2 enzyme deficiency has a similar presentation but undergoes virilization at the time of puberty. *Asherman's syndrome* presents as secondary amenorrhea or hypomenorrhea and results from partial or complete obliteration of the uterine cavity by adhesions that prevent normal growth and shedding of the endometrium. Curettage performed for pregnancy complications accounts for >90% of cases; genital tuberculosis is an important cause in regions where it is endemic.

TREATMENT

Disorders of the Uterus or Outflow Tract

Obstruction of the outflow tract usually presents as dysmenorrhea or lower abdominal cyclic pain with no menses. Evaluation of the patient includes a medical history, physical examination including a perineal examination, and ultrasound imaging. In many cases, a magnetic resonance imaging (MRI) scan can more accurately identify the reproductive tract anomaly prior to surgery. It is important that surgery be performed as soon as the diagnosis is made as the risk of endometriosis is increased with retrograde menstrual flow. *Müllerian agenesis* may require surgical intervention to allow sexual intercourse, although vaginal dilatation is adequate in some patients. In these patients, because ovarian function is normal, assisted reproductive techniques can be used with a surrogate carrier. More recently, there have been a few cases of successful uterine transplantation in women with müllerian agenesis. AIS (Chap. 402) requires gonadectomy because there is risk of gonadoblastoma in the dysgenetic gonads, although surgery is generally delayed until after breast development and the pubertal growth spurt. Estrogen replacement is indicated after gonadectomy, and vaginal dilatation may be required to allow sexual intercourse.

Disorders of Ovulation Once uterus and outflow tract abnormalities have been excluded, other causes of amenorrhea involve disorders of ovulation. The differential diagnosis is based on the results of initial tests, including a pregnancy test, an FSH level (to determine whether the cause is likely to be ovarian or central), and assessment of hyperandrogenism (Fig. 405-2).

HYPOGONADOTROPIC HYPOGONADISM Low estrogen levels in combination with normal or low levels of LH and FSH are seen with anatomic, genetic, or functional abnormalities that interfere with hypothalamic GnRH secretion or normal pituitary responsiveness to GnRH. Although relatively uncommon, tumors and infiltrative diseases should be considered in the differential diagnosis of hypogonadotropic hypogonadism (Chap. 392). These disorders may present with primary or secondary amenorrhea. They may occur in association with other features suggestive of hypothalamic or pituitary dysfunction, such as short stature, diabetes insipidus, galactorrhea, and headache. Hypogonadotropic hypogonadism also may be seen after cranial irradiation. In the postpartum period, amenorrhea occurs normally in association with breast feeding but may also be caused by pituitary necrosis (Sheehan's syndrome) or lymphocytic hypophysitis. Because reproductive dysfunction is commonly associated with hyperprolactinemia from neuroanatomic lesions or medications, prolactin should be measured in all patients with hypogonadotropic hypogonadism (Chap. 392).

Isolated hypogonadotropic hypogonadism (IHH) occurs in women, although it is three times more common in men. IHH generally presents with primary amenorrhea, although 50% have some degree of

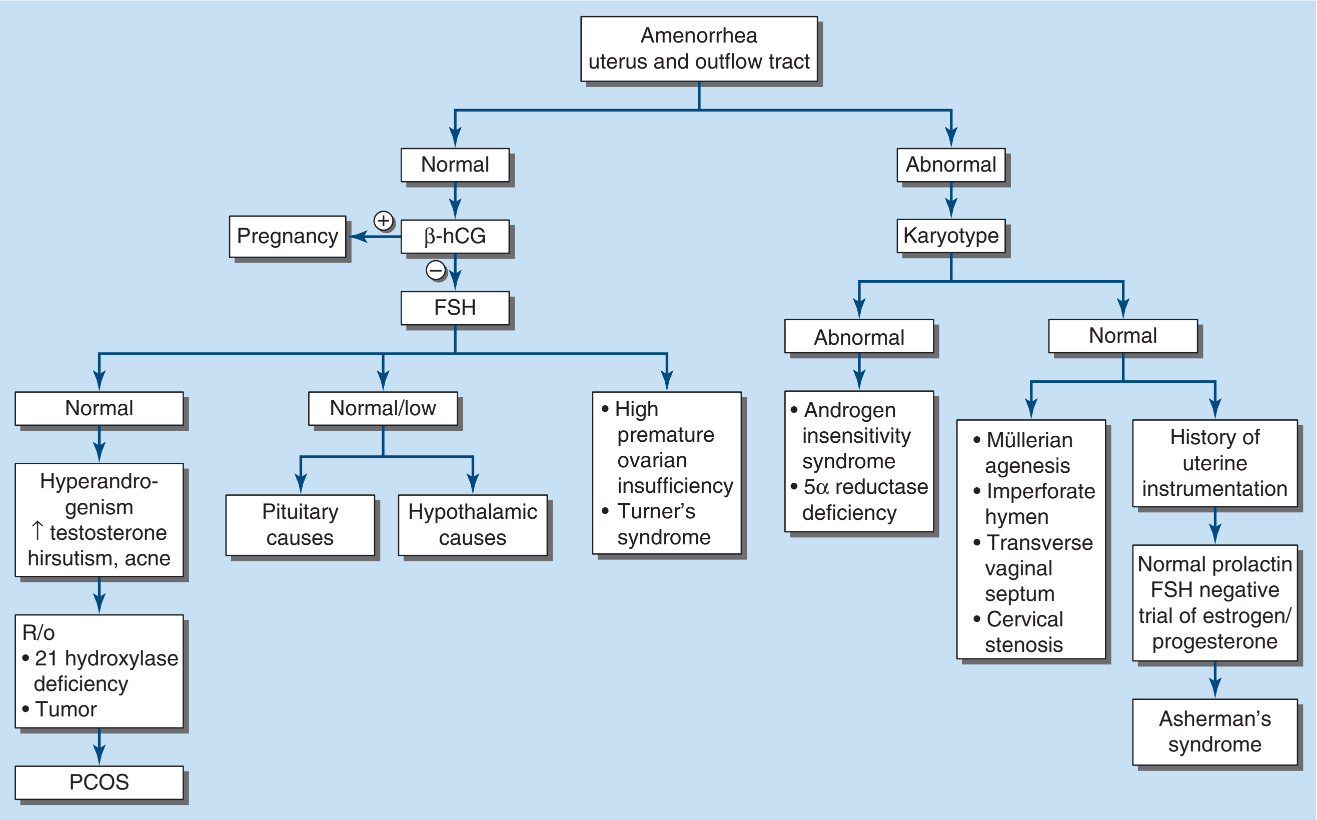
Amenorrhea
uterus and outflow tract
Normal
Abnormal
Pregnancy
+
β-hCG
−
FSH
Karyotype
Abnormal
Normal
Normal
Normal/low
• High premature ovarian insufficiency
• Turner's syndrome
• Androgen insensitivity syndrome
• 5α reductase deficiency
• Müllerian agenesis
• Imperforate hymen
• Transverse vaginal septum
• Cervical stenosis
History of uterine instrumentation
Hyperandrogenism
↑ testosterone
hirsutism, acne
Pituitary causes
Hypothalamic causes
Normal prolactin
FSH negative
trial of estrogen/
progesterone
R/o
• 21 hydroxylase deficiency
• Tumor
Asherman's syndrome
PCOS

FIGURE 405-2 Algorithm for evaluation of amenorrhea. β-hCG, β-human chorionic gonadotropin; FSH, follicle-stimulating hormone; GYN, gynecologist; MRI, magnetic resonance imaging; PRL, prolactin; R/O, rule out; TSH, thyroid-stimulating hormone.

breast development, and ~10% report one to two menses. IHH is associated with anosmia in half of women (termed Kallmann's syndrome). Genetic causes of IHH have been identified in ~50% of patients **(Chaps. 403 and 404)**.

Functional hypothalamic amenorrhea (HA) is a diagnosis of exclusion of other causes of hypogonadotropic hypogonadism including chronic diseases (type 1 diabetes, celiac disease, hyperthyroidism, Cushing's syndrome) and use of opioids, glucocorticoids, or psychotropic medications that increase prolactin levels. Functional HA is most commonly associated with conditions causing a mismatch between energy expenditure and energy intake and/or significant stress leading to increased corticotropin-releasing hormone (CRH), suppression of GnRH, and decreased thyrotropin-releasing hormone (TRH) input. Variants in genes associated with IHH may increase susceptibility to these environmental inputs, accounting in part for the clinical variability in this disorder. Metabolic and stress signaling is transduced to the reproductive axis, at least in part, through leptin signaling from the periphery and via hypothalamic kisspeptin, neurokinin B, and dynorphin control of GnRH. The diagnosis of HA generally can be made on the basis of a careful history, a physical examination, and the demonstration of low levels of gonadotropins and normal prolactin levels. Eating disorders, excessive exercise, and chronic disease must be specifically excluded. An atypical history, headache, signs of other hypothalamic dysfunction, or hyperprolactinemia, even if mild, necessitates cranial MRI to exclude a neuroanatomic cause. Up to 10% of women with HA may have some features of PCOS (irregular menses, increased ovarian volume with polycystic appearing ovaries, higher anti-müllerian hormone [AMH] levels, and slightly elevated androgen levels).

HYPERGONADOTROPIC HYPOGONADISM Ovarian failure is considered premature when it occurs in women <40 years old and accounts for ~10% of secondary amenorrhea. *Primary ovarian insufficiency* (POI) has replaced the terms *premature menopause* and *premature ovarian failure* in recognition of the continuum of impaired ovarian function encompassed by this disorder. Ovarian insufficiency is associated with the loss of negative feedback restraint on the hypothalamus and pituitary, resulting in increased FSH and LH levels. FSH is a better marker of ovarian failure because of loss of negative feedback effects of both estradiol and the inhibins and because its levels are less variable than those of LH. AMH levels will also be low in patients with POI. As with natural menopause, POI may wax and wane, and serial measurements may be necessary to establish the diagnosis. The presentation may include irregular menses or complete cessation of menses, hot flashes, and vaginal dryness.

Once the diagnosis of POI has been established, further evaluation is indicated because of other health problems that may be associated with POI. Although POI is most commonly of unknown cause, it also occurs in association with a variety of chromosomal abnormalities (most often Turner's syndrome), autoimmune polyglandular failure syndromes, and other rare disorders. Radiotherapy and chemotherapy may reduce ovarian reserve, with effects on both the oocytes and the supporting granulosa cells. New approaches, including ovarian, oocyte, and embryo cryopreservation, should be offered to women of reproductive age prior to gonadotoxic chemotherapy or pelvic radiation treatment. The recognition that early ovarian insufficiency occurs in premutation carriers of the fragile X syndrome is important because of the increased risk of severe intellectual disability in male children with *FMR1* mutations. Thus, follow-up testing should include a karyotype in all POI patients, serum anti-cortical and 21-hydroxylase antibodies (specific but not sensitive for subsequent adrenal insufficiency), thyroid function and thyroid peroxidase antibodies, *FMR1* premutation screening, and assessment of bone mineral density. Ovarian biopsy is not indicated. Although the number of genetic causes of POI is increasing, routine testing for mutations other than *FMR1* is currently not recommended.

Hypergonadotropic hypogonadism occurs rarely in other disorders, such as mutations in the FSH or LH receptors. Aromatase deficiency and 17α-hydroxylase deficiency are associated with decreased estrogen and elevated gonadotropins and with hyperandrogenism and hypertension, respectively. Gonadotropin-secreting tumors in women of reproductive age generally present with high, rather than low, estrogen levels and cause ovarian hyperstimulation or dysfunctional bleeding.

TREATMENT

Hypo- and Hypergonadotropic Causes of Amenorrhea

Amenorrhea almost always is associated with chronically low levels of estrogen, whether it is caused by hypogonadotropic hypogonadism or ovarian insufficiency. Development of secondary sexual characteristics requires gradual titration of estradiol replacement with eventual addition of progestin. Hormone replacement with either low-dose estrogen/progesterone regimens or oral contraceptive pills is recommended until the usual age of menopause for bone and cardiovascular protection. In women with functional HA or anorexia nervosa, hormone replacement alone may not be sufficient to restore or maintain bone density. A more long-term multidisciplinary approach including behavioral health professionals is essential. Patients with hypogonadotropic hypogonadism who are interested in fertility require treatment with both exogenous FSH and LH. Patients with POI can consider oocyte donation, which has a high rate of success in this population, although its use in women with Turner's syndrome is limited by increased cardiovascular risk in pregnancy.

POLYCYSTIC OVARY SYNDROME The diagnosis of PCOS is made in adult women using the updated Rotterdam criteria (published in the 2023 international guidelines). These include irregular menses (<8 menses per year), clinical or biochemical hyperandrogenism (elevated total or free testosterone, modified Ferriman-Gallwey score >4–6 depending on ethnicity, see **Chap. 406**), and polycystic-appearing ovaries on ultrasound (≥20 antral follicles or ovarian volume ≥10 cm^3 in at least one ovary) or elevated AMH. The presence of two of the three criteria will confirm the diagnosis, resulting in different phenotypes, namely, hyperandrogenic or non-hyperandrogenic. PCOS is a diagnosis of exclusion, and other etiologies for irregular menses and hyperandrogenism should be excluded (hypothyroidism, hyperprolactinemia, adrenal sources for hyperandrogenism). Diagnosis in adolescents may be difficult to establish, and it is recommended to wait at least 3 years after menarche before confirming the diagnosis. In adolescents, the diagnosis is based on irregular menses and hyperandrogenism criteria only, as the ultrasound and AMH criteria are not established for this age group. Lean oligo-ovulatory patients with PCOS generally have high LH levels in the presence of normal to low levels of FSH and estradiol, although given the pulsatility of LH secretion, a random serum LH/FSH ratio is not included in the diagnostic criteria. The prevalence of obesity is high in PCOS and significantly increases the risk of comorbidities including metabolic syndrome, type 2 diabetes, dyslipidemia, hypertension, and cardiovascular disease. Frequent anovulation results in irregular menses and increased risk of endometrial hyperplasia and endometrial cancer (two- to sixfold increased risk). Abnormalities in GnRH pulsatility result in elevated LH levels resulting in increased production of ovarian androgens. Insulin resistance, especially in skeletal muscle and adipose tissue, also contributes to increased insulin-stimulated ovarian androgen production. An alternate source of androgens, namely 11-oxygenated androgens, may also be elevated in this population. Genome-wide association studies in diverse populations and PCOS phenotypes have identified several loci (~19) associated with PCOS, and cluster analyses suggest the presence of reproductive and metabolic abnormalities. Symptoms generally begin in adolescence and are modified by obesity and age, such that by the fourth decade of life, most women with PCOS will have regular menses and normal serum androgens. There is a high prevalence of depression and anxiety disorders, as well as disordered eating and body image distress. PCOS is also associated with an increased risk of obstructive sleep apnea and metabolic dysfunction-associated steatotic liver disease (MDSLD), independent of body mass index (BMI).

TREATMENT

Polycystic Ovary Syndrome

The first-line treatment of women with PCOS not attempting pregnancy is combined hormonal contraceptives to regulate menstrual cycles and decrease serum androgens by increasing sex hormone–binding globulin levels. Although serum androgens decrease by 2–3 months after initiating hormonal therapy, it may take longer to observe the beneficial effects on hirsutism and acne. Patients should be prescribed hormonal contraceptives containing the lowest effective dose of estrogen, either in a cyclic or continuous manner. If there is an inadequate response to hormonal contraceptives after 6 months for management of hyperandrogenic symptoms, antiandrogens, such as spironolactone and flutamide, can be considered **(Chap. 406)**. Endometrial protection can also be achieved with the use of progestins (medroxyprogesterone acetate,10 mg, or Prometrium [progesterone], 200 mg daily for 10–14 days at least every 3 months, or a levonorgestrel intrauterine device [IUD]). All women with PCOS should be screened for obesity, hypertension, glycemic control, depression, and anxiety at the time of diagnosis and then at regular intervals. Overweight and obese women should also have a fasting lipid profile at the time of diagnosis. Lifestyle management should be recommended in all women with PCOS, and metformin should be considered for prevention of cardiometabolic risk factors in those with overweight and obesity **(Chap. 420)**. Women with PCOS are at an increased risk of early miscarriage, gestational diabetes, gestational hypertension, preeclampsia, and preterm birth. Lifestyle management and prepregnancy counseling should be offered prior to attempting pregnancy **(Chap. 408)**. Letrozole, an aromatase inhibitor, is the first-line treatment for ovulation induction followed by clomiphene citrate, a selective estrogen response modulator, with or without metformin. Injectable gonadotropins can be used judiciously by experienced practitioners to induce monofollicular growth as PCOS increases the risk of hyperstimulation. Metformin can be used as an adjunct with diet and exercise for obese women with PCOS or for treatment of diabetes or impaired glucose tolerance, as in non-PCOS patients. However, metformin alone is not recommended for endometrial protection or treatment of hyperandrogenic symptoms, infertility, pregnancy loss, or prevention of gestational diabetes.

PELVIC PAIN

The mechanisms that cause pelvic pain are similar to those that cause abdominal pain **(Chap. 16)** and include inflammation of the parietal peritoneum, obstruction of hollow viscera, vascular disturbances, and pain originating in the abdominal wall. Pelvic pain may reflect pelvic disease per se but also may reflect extrapelvic disorders that refer pain to the pelvis. In up to 60% of cases, pelvic pain can be attributed to gastrointestinal problems, including appendicitis, cholecystitis, infections, intestinal obstruction, diverticulitis, and inflammatory bowel disease. Urinary tract and musculoskeletal disorders are also common causes of pelvic pain.

APPROACH TO THE PATIENT

Pelvic Pain

As with all types of abdominal pain, the first priority is to identify life-threatening conditions (shock, peritoneal signs) that may require emergent surgical management. The possibility of pregnancy should be identified as soon as possible by menstrual history and β-human chorionic gonadotropin (β-hCG) testing. A thorough history that includes the type, location, radiation, and recurrence can help identify the cause of acute pelvic pain. Specific associations with vaginal bleeding, sexual activity, defecation, urination, movement, or eating should be specifically sought. Determination of whether the pain is acute versus chronic and cyclic versus noncyclic will direct further investigation **(Table 405-1)**. However, disorders that cause cyclic pain occasionally may cause noncyclic pain, and the converse is also true.

ACUTE PELVIC PAIN

Pelvic inflammatory disease (PID) refers to infection of the upper genital tract and may present with a spectrum of symptoms. In the acute setting, the most common presentation is bilateral lower abdominal pain of recent onset that may be exacerbated with sexual activity. Risk factors for PID include age <25 years and history of multiple sexual partners, sexually transmitted infections (STIs), or recent uterine procedures. However, any sexually active woman can be at risk for PID. PID associated with tubo-ovarian abscess or peritonitis may present with severe pain, fever, and peritoneal signs. Abnormal uterine bleeding may occur in about one-third of patients. Cervical motion tenderness, uterine and adnexal pain, and vaginal discharge are common findings on pelvic examination. The presence of right upper quadrant pain is suggestive of perihepatitis (Fitz-Hugh–Curtis syndrome).

The diagnosis of PID is established based on symptoms and clinical examination and can be aided by a wet mount preparation of vaginal discharge and nucleic acid amplification tests for *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. Of note, a presumptive clinical diagnosis is sufficient to prescribe treatment even in the absence of positive test results, as PID can occur due to other vaginal and enteric pathogens. Pelvic imaging can be obtained based on symptoms, findings of the pelvic examination, or if there is lack of response to therapy. With public health efforts to control STIs, the incidence and severity of PID have declined in the United States and Europe; however, this is not the case in the developing world. Subclinical PID with its attendant risks of infertility and ectopic pregnancy remains a significant problem worldwide. Public health and professional organizations recommend annual testing for *C. trachomatis* in all sexually active women <25 years old and both *C. trachomatis* and *N. gonorrhoeae* in all women at increased risk. *Adnexal pathology* can present acutely and may be due to rupture, bleeding, or torsion of ovarian cysts or, much less commonly, the fallopian tubes. Rupture of an ovarian cyst may be diagnosed based on the acute presentation in a reproductive-age woman and pelvic ultrasound findings of a simple, collapsed or hemorrhagic cyst, with or without free fluid in the pelvis. Ovarian torsion typically presents as acute onset of unilateral, intermittent pain and is a diagnosis of exclusion unless absent blood flow to the ovary is demonstrated via Doppler ultrasound imaging. Neoplasms of the ovary or fallopian tube are much less common causes of acute pain.

Ectopic pregnancy represents 1–2% of all pregnancies and most commonly occurs in the fallopian tubes. It may present with acute lower abdominal pain, hemodynamic instability, and peritoneal signs. The index of suspicion should be high in any reproductive-age woman presenting with abdominal pain or vaginal bleeding irrespective of current use of contraception. Risk factors for an ectopic pregnancy include history of tubal disease, pelvic infection, tubal surgery, previous ectopic pregnancy, infertility, smoking, and current use of IUD, although a large proportion may have no risk factors. Rupture of the fallopian tube remains a life-threatening emergency; the incidence depends on access to care but is ~18% in developed countries. Diagnosis of an ectopic pregnancy can be established by assessing the patient's menstrual history and symptoms, measuring a single or serial β-hCG levels, and performing pelvic ultrasound imaging. β-hCG levels typically double every 48 h in early first trimester, and 99% of viable intrauterine pregnancies are associated with an increase in hCG levels of at least 53% in 2 days. The discriminatory zone refers to β-hCG values above which the landmarks of a normal intrauterine pregnancy should be seen on ultrasound (1500–3000 IU/mL). Absence of an intrauterine pregnancy and presence of an adnexal mass or free fluid increase the likelihood of an ectopic pregnancy. *Threatened abortion* may also present with amenorrhea, abdominal pain, and vaginal bleeding with no cervical dilation in the setting of an intrauterine pregnancy with cardiac activity in the first trimester of pregnancy. Although more common than ectopic pregnancy, it is rarely associated with systemic signs. *Uterine pathology* includes endometritis, and less frequently, degenerating leiomyomas (fibroids) present with acute pain. Endometritis often is associated with vaginal bleeding and systemic signs of infection. It occurs in the setting of STIs, uterine instrumentation, or postpartum infection.

TABLE 405-1 Gynecologic Causes of Pelvic Pain

	ACUTE	CHRONIC
Cyclic pelvic pain		Mittelschmerz Dysmenorrhea
Noncyclic pelvic pain	Pelvic inflammatory disease Ruptured or hemorrhagic ovarian cyst, endometrioma, or ovarian torsion Ectopic pregnancy Endometritis Acute growth or degeneration of uterine myoma Threatened abortion	Endometriosis Uterine fibroids Adenomyosis Pelvic adhesions Pelvic malignancy Vulvodynia Chronic pelvic inflammatory disease Tuberculous salpingitis History of sexual abuse Pelvic congestion syndrome

TREATMENT

Acute Pelvic Pain

Treatment of acute pelvic pain depends on the suspected etiology but may require surgical or medical intervention. Immediate treatment of PID is indicated upon diagnosis, even if the diagnosis is presumed or the symptoms are mild, due to long-term complications resulting in increased risk of ectopic pregnancy and infertility. Treatment in patients eligible for outpatient management includes 250 mg IM ceftriaxone and a 14-day course of oral doxycycline 100 mg twice daily. If the presentation is acute with high fever, nausea, vomiting, severe abdominal pain, or presence of tubo-ovarian abscess, inpatient therapy is recommended **(Chap. 141)**. Conservative management is an important consideration for *ovarian cysts*, if torsion is not suspected, to avoid unnecessary surgery and associated risks of reduced fertility due to cystectomy or adhesions. If surgery is performed, it is preferable to perform a cystectomy, removing the cyst wall and leaving the remaining ovary, in a reproductive-age woman. Combined hormonal contraceptives are recommended in women with a history of recurrent ovarian cyst formation. Surgical treatment may be required for *ectopic pregnancies* when the patient presents with acute pain, is hemodynamically unstable, or has signs of intraperitoneal bleeding. The choice of salpingectomy versus salpingostomy is based on patient's presentation, desire for future child-bearing, and prior pelvic infections. Clinically stable women presenting with unruptured ectopic pregnancies may be appropriate for treatment with methotrexate, which is effective in ~90% of cases when multiple doses are used. Threatened abortion is managed conservatively even in the presence of a subchorionic hemorrhage. The treatment of endometritis is similar to PID. Pain from a degenerating fibroid, if visualized on pelvic sonography, can be managed with nonsteroidal anti-inflammatory drugs (NSAIDs).

CHRONIC PELVIC PAIN

Chronic pelvic pain is a complex condition resulting from gynecologic, urologic, or gastrointestinal organs and contributes to significant frustration and burden of disease. Common gynecologic conditions contributing to chronic pain are endometriosis, fibroids, adenomyosis, and adnexal pathology. In addition to a detailed history and physical exam, the evaluation of chronic pelvic pain typically includes a pelvic ultrasound. As causes other than those related to the female reproductive system are common, referral should be made to other specialists, as appropriate. Neuromuscular and psychosomatic etiologies should also be considered.

Some women experience discomfort at the time of ovulation (*mittelschmerz* or *ovulation pain*). The pain can be quite intense but is generally of short duration. The mechanism is thought to involve rapid expansion of the dominant follicle, although it also may be caused by peritoneal irritation by follicular fluid released at the time of ovulation.

Dysmenorrhea typically refers to the crampy lower abdominal midline discomfort that begins with the onset of menstrual bleeding and gradually decreases over 12–72 h. It may be associated with nausea, diarrhea, fatigue, and headache and occurs in 60–93% of adolescents, beginning with the establishment of regular ovulatory cycles. Its prevalence decreases after pregnancy and can be treated effectively with hormonal contraceptives. *Primary dysmenorrhea* results, in a majority of cases, from hormone-dependent prostaglandin (PG) pathway mechanisms that cause intense uterine contractions, decreased blood flow, and increased peripheral nerve hypersensitivity, resulting in pain. However, variability in response to cyclooxygenase inhibitors suggests that PG-independent pathways, such as platelet activating factor, may also mediate inflammation. *Secondary dysmenorrhea* refers to pain caused by underlying pelvic pathology.

Endometriosis results from the presence of endometrial glands and stroma outside the uterus. These deposits of ectopic endometrium respond to hormonal stimulation and may be associated with dysmenorrhea, painful intercourse, and painful bowel movements. On pelvic exam, adnexal tenderness may be present or tender nodules may be palpated along the uterosacral ligaments. Pain associated with endometriosis can be cyclic or continuous, and the stage/severity of endometriosis, as determined by laparoscopy, does not always correlate with the extent of pain. Transvaginal pelvic ultrasound is part of the initial workup and may detect an endometrioma within the ovary. Additional sonographic techniques can be used to identify deep endometriosis including nonmobile ovaries and rectovaginal or bladder nodules. The CA-125 level may be increased, but it has low negative predictive value. Pelvic MRI has higher sensitivity and specificity for diagnosis of endometriosis. Diagnostic laparoscopy is performed when patients do not respond adequately to empiric treatment and is considered the gold standard for diagnosis. If endometriosis is detected, the severity can be staged and the endometriotic lesions ablated or excised.

Large *fibroids* can cause chronic pelvic pain or pressure, and submucosal fibroids may be associated with dysmenorrhea. *Other secondary causes* of dysmenorrhea include adenomyosis, a condition caused by the presence of ectopic endometrial glands and stroma within the myometrium. Chronic PID may be associated with ongoing pelvic pain and is associated with tuberculosis or actinomycosis. *Pelvic congestion syndrome* is associated with pelvic varicosities with low blood flow, resulting in pelvic venous congestion. However, this is no clear evidence to indicate that this finding is associated with chronic pelvic pain.

TREATMENT

Chronic Pelvic Pain

DYSMENORRHEA

Local application of heat, exercise, sexual activity, a vegetarian diet, use of vitamins D, B$_1$, B$_6$, and E and fish oil, acupuncture, and yoga have all been suggested to be of benefit, but studies are not adequate to provide recommendations. However, NSAIDs are very effective and provide >80% sustained response rates. Ibuprofen, naproxen, ketoprofen, mefenamic acid, and nimesulide are all superior to placebo. For best response, treatment should be initiated prior to the onset of menses and continued for at least 2–3 days. Combined or progestin-only hormonal contraceptives taken cyclically or continuously effectively reduce symptoms of dysmenorrhea.

ENDOMETRIOSIS

Combined hormonal contraceptives or continuous progestin (either orally, implants, or a levonorgestrel IUD) is used for the treatment of endometriosis. Evidence of an endometrioma on ultrasound imaging can be medically managed and does not require surgical removal unless it increases in size or there is persistent pain. Patients who do not respond to medical management and laparoscopic resection of endometriotic lesions can be offered GnRH agonist suppression with add-back therapy or aromatase inhibitors.

FIBROIDS

Chronic pain and dysmenorrhea associated with *fibroids* can be managed medically or surgically depending on the number and location of fibroids and associated symptoms. The U.S. Food and Drug Administration (FDA) approved the first two oral treatments for uterine fibroids (elagolix with estradiol/norethindrone acetate and relugolix with estradiol/norethindrone acetate), and the selective progesterone receptor modulator ulipristal acetate was withdrawn in Europe and Canada. Medical management includes oral hormonal contraceptives, tranexamic acid, NSAIDs, progestins, IUDs, and GnRH agonists and antagonists. Surgical management includes myomectomy, endometrial ablation/myolysis, radiofrequency volumetric thermal ablation, laparoscopic ablation, or hysterectomy. Chronic pain and dysmenorrhea associated with adenomyosis can be managed with combined hormonal treatment, levonorgestrel IUD, or hysterectomy after child-bearing is complete.

FURTHER READING

Bartels CB et al: An evidence-based approach to the medical management of fibroids: A systematic review. Clin Obstet Gynecol 59:30, 2016.

Bloomfield H et al: Screening pelvic examinations in asymptomatic average risk adult women. WA-ESP Project #09-009; 2013.

Bouilly J et al: Identification of multiple gene mutations accounts for the new genetic architecture of ovarian insufficiency. J Clin Endocrinol Metab 101:4541, 2016.

Brunham RC et al: Pelvic inflammatory disease. N Engl J Med 372:2039, 2015.

Fourman LR, Fazeli PK: Neuroendocrine causes of amenorrhea—An update. J Clin Endocrinol Metab 100:812, 2015.

Ju H et al: The prevalence and risk factors of dysmenorrhea. Epidem Rev 36:104, 2014.

Lee IT, Barnhart KT: What is an ectopic pregnancy? JAMA 329:434, 2023.

Oladosu FA et al: Nonsteroidal anti-inflammatory drug resistance in dysmenorrhea: Epidemiology, causes, and treatment. Am J Obstet Gynecol 218:390, 2018.

Taylor HS et al: Endometriosis is a chronic systemic disease: Clinical challenges and novel innovations. Lancet 397:839, 2021.

Teede HJ et al: International PCOS Network. Recommendations from the 2023 International Evidence-based Guideline for the Assessment and Management of Polycystic Ovary Syndrome. Fertil Steril 120:767, 2023.

406

Hirsutism

David A. Ehrmann

DEFINING HIRSUTISM

Body hair can be categorized as either *vellus* (fine, soft, and not pigmented) or *terminal* (long, coarse, and pigmented). Approximately 10% of reproductive-age women have hirsutism, defined by the presence of excessive terminal hair growth. Hirsutism is most often idiopathic or the consequence of androgen excess associated with polycystic ovary

TABLE 406-1 Causes of Hirsutism

- Gonadal hyperandrogenism
  - Ovarian hyperandrogenism
  - Polycystic ovary syndrome/functional ovarian hyperandrogenism
  - Ovarian steroidogenic blocks
  - Syndromes of extreme insulin resistance
  - Ovarian neoplasms
  - Hyperthecosis
- Adrenal hyperandrogenism
  - Premature adrenarche
  - Functional adrenal hyperandrogenism
  - Congenital adrenal hyperplasia (nonclassic and classic)
  - Abnormal cortisol action/metabolism
  - Adrenal neoplasms
- Other endocrine disorders
  - Cushing's syndrome
  - Hyperprolactinemia
  - Acromegaly
- Peripheral androgen overproduction
  - Obesity
  - Idiopathic
- Pregnancy-related hyperandrogenism
  - Hyperreactio luteinalis
  - Thecoma of pregnancy
- Drugs
  - Androgens
  - Oral contraceptives containing androgenic progestins
  - Minoxidil
  - Phenytoin
  - Diazoxide
  - Cyclosporine
  - Valproic acid
- Ovotesticular disorders of sex development

syndrome (PCOS). Less frequently, it results from adrenal androgen overproduction as occurs in nonclassic congenital adrenal hyperplasia (CAH) **(Table 406-1)**.

Virilization refers to a condition that may the result from benign hyperplasia of ovarian theca and stroma cells (e.g., *hyperthecosis*); it may also be a harbinger of a serious underlying condition, such as an ovarian or adrenal neoplasm. In women with virilization, androgen levels are sufficiently high to cause deepening of the voice, breast atrophy, increased muscle bulk, clitoromegaly, and increased libido. Cutaneous manifestations commonly associated with hirsutism include acne and hair thinning or pattern hair loss (androgenic alopecia).

■ HAIR FOLLICLE GROWTH AND DIFFERENTIATION

The number of hair follicles remains unchanged over the life span, but follicle size and the type of hair can change in response to numerous factors, particularly androgens. Androgens are necessary for terminal hair and sebaceous gland development and mediate differentiation of pilosebaceous units (PSUs) into a terminal hair follicle and/or a sebaceous gland. In the former case, androgens transform the vellus hair into a terminal hair; in the latter case, the sebaceous component proliferates and the hair remains vellus.

There are three phases in the cycle of hair growth: (1) *anagen* (growth phase), (2) *catagen* (involution phase), and (3) *telogen* (rest phase). Depending on the body site, hormonal regulation may play an important role in the hair growth cycle. Hair growth on the face, chest, upper abdomen, and back typically requires elevated androgen concentrations. However, there is only a modest correlation between androgen levels and the quantity of hair growth. This is due to the fact that hair growth from the follicle also depends on local growth factors, and the variability in end-organ (PSU) sensitivity to androgens. Genetic factors and ethnic background also influence hair growth. Androgen excess in women may result in hair thinning or loss because androgens cause scalp hairs to spend less time in the anagen phase.

In general, dark-haired individuals tend to be more hirsute than blond or fair individuals. Asians and Native Americans have relatively sparse hair in regions sensitive to high androgen levels, whereas people of Mediterranean descent are more hirsute.

■ CLINICAL ASSESSMENT

Historic elements relevant to the assessment of hirsutism include the age at onset and rate of progression of hair growth and associated symptoms or signs (e.g., menstrual irregularity and acne). Depending on the cause, excess hair growth typically is first noted during the second and third decades of life. The growth is usually slow but progressive. Sudden development and rapid progression of hirsutism suggest the possibility of an androgen-secreting neoplasm, in which case virilization may also be present.

The age at onset of menstrual cycles (menarche) and the pattern of the menstrual cycle should be ascertained. Menses may be irregular in the first 2 years after menarche; oligomenorrhea (<8 cycles per calendar year) thereafter is more likely to result from ovarian than adrenal androgen excess. Associated symptoms such as galactorrhea should prompt evaluation for hyperprolactinemia **(Chap. 392)** or possibly hypothyroidism **(Chap. 394)**. Hypertension, striae, easy bruising, and centripetal weight gain suggest hypercortisolism (Cushing's syndrome; **Chap. 398**). Rarely, patients with acromegaly present with hirsutism. Medications such as phenytoin, minoxidil, and cyclosporine may be associated with androgen-*independent* excess hair growth (i.e., hypertrichosis). A family history of infertility and/or hirsutism may indicate inherited disorders such as nonclassic CAH **(Chap. 398)**.

Physical examination should include measurement of height and weight and calculation of body mass index (BMI). A BMI >25 kg/m^2 is indicative of excess weight for height, and values >30 kg/m^2 are often seen in association with hirsutism, probably the result of increased conversion of androgen precursors to testosterone. Notation should be made of blood pressure, as adrenal causes may be associated with hypertension. Cutaneous signs sometimes associated with androgen excess and insulin resistance include acanthosis nigricans and skin tags.

An objective clinical assessment of hair distribution and quantity is central to the evaluation in any woman presenting with concerns about excessive hair growth. This assessment permits the distinction between hirsutism and hypertrichosis and provides a baseline reference point to gauge the response to treatment. A simple and commonly used method to grade hair growth is the modified scale of Ferriman and Gallwey **(Fig. 406-1)**, in which each of nine androgen-sensitive sites is graded from 0 (no hair growth) to 4 (hair growth typically seen in adult men). Although it is normal for most women to have some hair growth in androgen-sensitive sites, ~95% of non-Hispanic white and African American women have a score <8 on this scale. Scores >8 suggest excess androgen-mediated hair growth, a finding that should be assessed further by means of hormonal evaluation (see below). Asian and Native American women are less likely to manifest hirsutism, and the only cutaneous evidence of androgen excess may be pustular acne and thinning scalp hair.

■ HORMONAL EVALUATION

Androgens are secreted by the ovaries and adrenal glands in response to their respective tropic hormones: luteinizing hormone (LH) and adrenocorticotropic hormone (ACTH). Testosterone is the principal circulating steroid involved in the etiology of hirsutism; other steroids that may contribute to the development of hirsutism include androstenedione and dehydroepiandrosterone (DHEA) and its sulfated form (DHEAS). The ovaries and adrenal glands normally contribute about equally to testosterone production. Approximately half of the total testosterone originates from direct glandular secretion, and the remainder is derived from the peripheral conversion of androstenedione and DHEA **(Chap. 393)**.

Testosterone is the most important circulating androgen, but it is a precursor hormone in mediating hirsutism. Testosterone is converted to dihydrotestosterone (DHT) by the enzyme 5α-reductase, which is

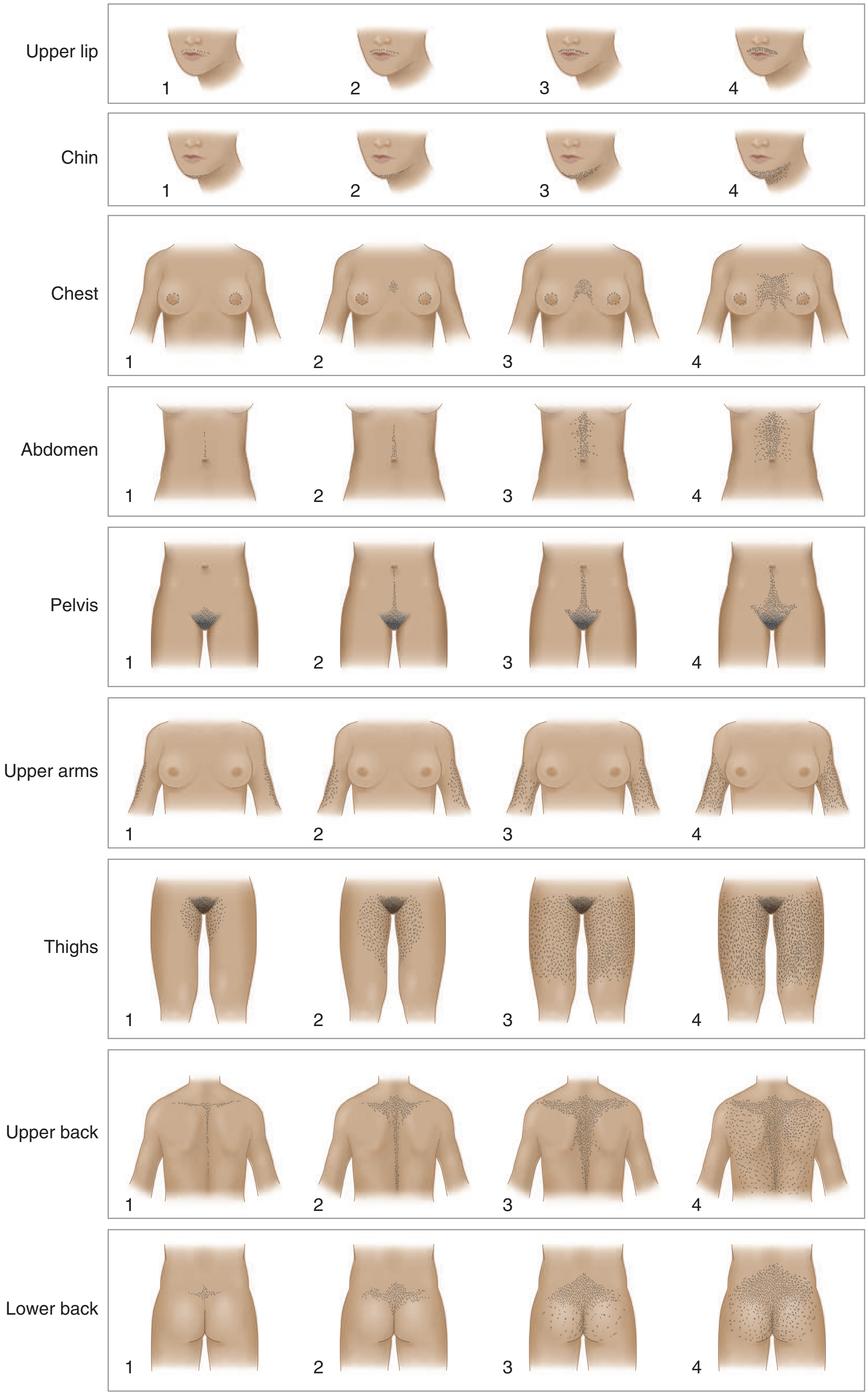

FIGURE 406-1 Hirsutism scoring scale of Ferriman and Gallwey. The nine body areas that have androgen-sensitive areas are graded from 0 (no terminal hair) to 4 (frankly virile) to obtain a total score. A normal hirsutism score is <8. *(Modified with permission from LJ DeGroot, JL Jameson: Endocrinology, 5th ed. Philadelphia, PA: Saunders; 2006.)*

located in the PSU. DHT is more potent than testosterone as it has a higher affinity for, and slower dissociation from, the androgen receptor. The local production of DHT allows it to serve as the primary mediator of androgen action at the level of the PSU. There are two isoenzymes of 5α-reductase: type 2 is found in the prostate gland and in hair follicles, and type 1 is found primarily in sebaceous glands.

One approach to the evaluation and treatment of hirsutism is depicted in Fig. 406-2. In addition to measuring blood levels of testosterone and DHEAS, it is often important to measure the level of free (or unbound) testosterone, i.e., the fraction of testosterone that is not bound to its carrier protein, sex hormone–binding globulin (SHBG). Unbound testosterone is biologically available for conversion to DHT

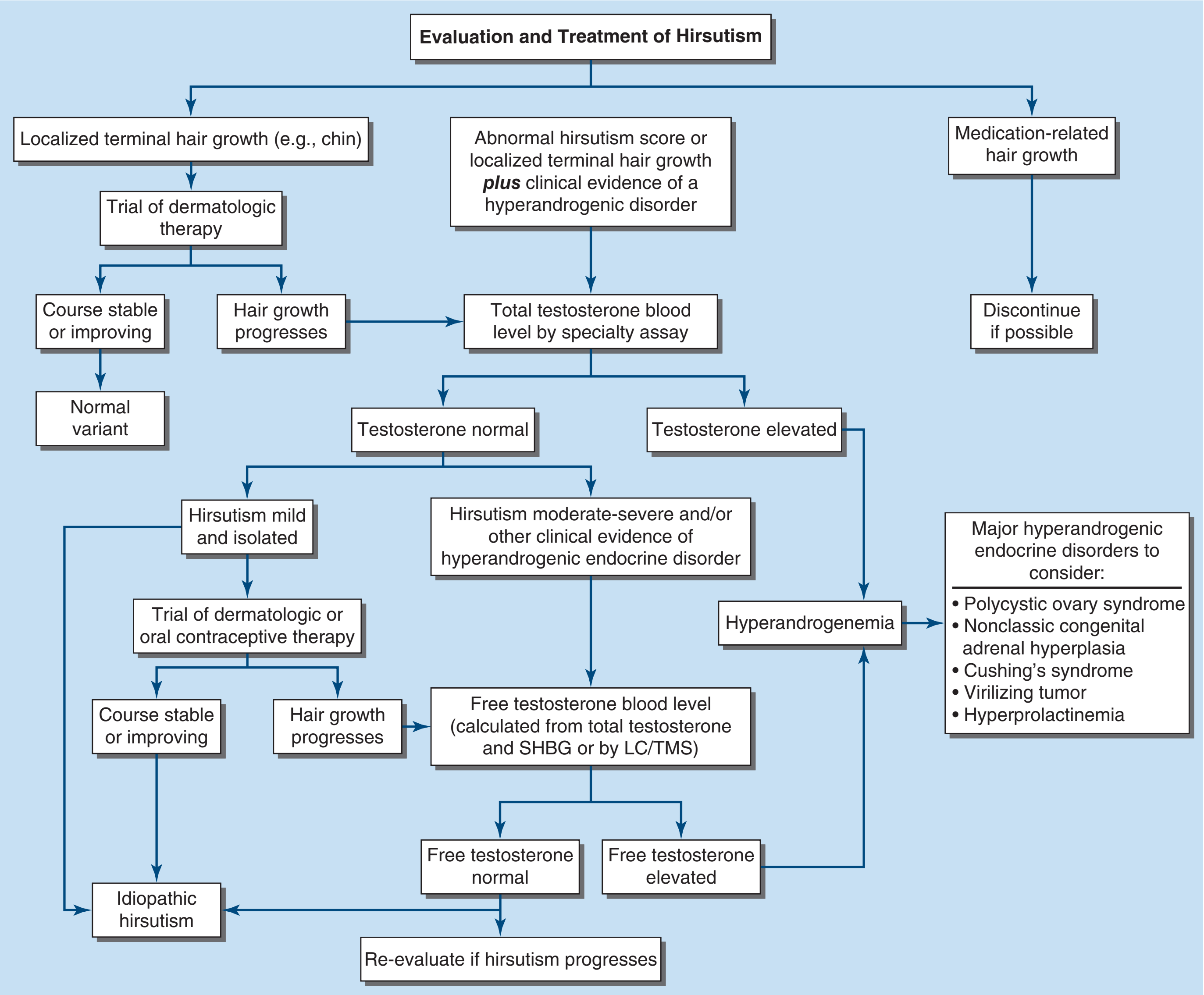
Evaluation and Treatment of Hirsutism
Localized terminal hair growth (e.g., chin)
Abnormal hirsutism score or localized terminal hair growth *plus* clinical evidence of a hyperandrogenic disorder
Medication-related hair growth
Trial of dermatologic therapy
Course stable or improving
Hair growth progresses
Total testosterone blood level by specialty assay
Discontinue if possible
Normal variant
Testosterone normal
Testosterone elevated
Hirsutism mild and isolated
Hirsutism moderate-severe and/or other clinical evidence of hyperandrogenic endocrine disorder
Major hyperandrogenic endocrine disorders to consider:
• Polycystic ovary syndrome
• Nonclassic congenital adrenal hyperplasia
• Cushing's syndrome
• Virilizing tumor
• Hyperprolactinemia
Trial of dermatologic or oral contraceptive therapy
Hyperandrogenemia
Course stable or improving
Hair growth progresses
Free testosterone blood level (calculated from total testosterone and SHBG or by LC/TMS)
Free testosterone normal
Free testosterone elevated
Idiopathic hirsutism
Re-evaluate if hirsutism progresses

FIGURE 406-2 Algorithm for the evaluation and treatment of hirsutism. LC/TMS, liquid chromatography/tandem mass spectrometry; SHBG, sex hormone–binding globulin. *(Reproduced with permission from KA Martin et al: Evaluation and treatment of hirsutism in premenopausal women: An endocrine society clinical practice guideline. J Clin Endocrinol Metab 103:1233, 2018.)*

and binding to androgen receptors. Both hyperinsulinemia and androgen excess decrease hepatic production of SHBG, resulting in levels of total testosterone within the high-normal range, whereas the unbound hormone is elevated more substantially. Although there is a decline in ovarian testosterone production after menopause, ovarian estrogen production decreases to an even greater extent, and the concentration of SHBG is reduced. Consequently, there is an increase in the relative proportion of unbound testosterone, and it may exacerbate hirsutism after menopause.

A baseline plasma total testosterone level >12 nmol/L (>3.5 ng/mL) usually indicates an androgen-producing tumor, whereas a level >7 nmol/L (>2 ng/mL) is suggestive of tumor but may also be observed in women with hyperthecosis. A basal DHEAS level >18.5 μmol/L (>7000 μg/L) suggests an adrenal tumor. Although DHEAS has been proposed as a "marker" of predominant adrenal androgen excess, it is not unusual to find modest elevations in DHEAS among women with PCOS. Computed tomography (CT) or magnetic resonance imaging (MRI) should be used to localize an adrenal mass, and transvaginal ultrasound usually suffices to identify an ovarian mass if clinical evaluation and hormonal levels suggest these possibilities.

PCOS is the most common cause of ovarian androgen excess **(Chap. 404)**. An increased ratio of LH to follicle-stimulating hormone (FSH) is characteristic in carefully studied patients with PCOS. However, because of the pulsatile nature of gonadotropin secretion, a random measurement of LH and FSH may be misleading and is not recommended. Transvaginal ultrasound classically shows enlarged ovaries, increased stroma, and multiple "cysts" in women with PCOS. These so-called cysts are, in fact, preantral and early antral follicles that result from abnormal follicular maturation. "Cystic" ovaries also may be found in women with hypothalamic amenorrhea **(Chap. 404)** and even among women without clinical or laboratory features of PCOS. Thus, ultrasonography is often not needed to diagnose PCOS given its relatively low specificity and its high degree of operator dependence.

Because adrenal androgens are readily suppressed by low doses of glucocorticoids, the dexamethasone androgen-suppression test may broadly distinguish ovarian from adrenal androgen overproduction. A blood sample is obtained before and after the administration of dexamethasone (0.5 mg orally every 6 h for 4 days). An adrenal source is suggested by suppression of unbound testosterone into the normal range; incomplete suppression suggests ovarian androgen excess. An overnight 1-mg dexamethasone suppression test, with measurement of 8:00 A.M. serum cortisol, is useful when there is clinical suspicion of Cushing's syndrome **(Chap. 398)**.

Nonclassic CAH is most commonly due to 21-hydroxylase deficiency but also can be caused by autosomal recessive defects in other steroidogenic enzymes necessary for adrenal corticosteroid synthesis **(Chap. 398)**. Because of the enzyme defect, the adrenal gland cannot secrete glucocorticoids (especially cortisol) efficiently. This results in diminished negative feedback inhibition of ACTH, leading to compensatory adrenal hyperplasia and the accumulation of steroid precursors that subsequently are converted to androgen. Deficiency of 21-hydroxylase can be reliably excluded by determining a morning 17-hydroxyprogesterone level <6 nmol/L (<2 μg/L) (drawn in the follicular phase). Alternatively, 21-hydroxylase deficiency can be diagnosed by measurement of 17-hydroxyprogesterone 1 h after the administration of 250 μg of synthetic ACTH (cosyntropin) intravenously.

TREATMENT

Hirsutism

Treatment of hirsutism may be accomplished pharmacologically or by mechanical means of hair removal. Nonpharmacologic treatments should be considered in all patients either as the only treatment or as an adjunct to drug therapy.

Nonpharmacologic treatments include (1) bleaching, (2) removal of the hair from the skin surface by shaving or with chemical treatments, and (3) depilatory (removal of the hair including the root) such as plucking, waxing, electrolysis, laser, and intense pulsed light (IPL). Despite perceptions to the contrary, shaving does not increase the rate or density of hair growth. Chemical depilatory treatments may be useful for mild hirsutism that affects only limited skin areas, although they can cause skin irritation. Wax treatment removes hair temporarily but is uncomfortable. Electrolysis is effective for more permanent hair removal, particularly in the hands of a skilled electrologist. Laser and IPL are used to treat large areas of pigmented, terminal hair. Light of specific wavelength, duration, and energy is absorbed by melanin in the hair shaft and follicle leading to photothermolysis. Properly delivered, this treatment delays hair regrowth and causes permanent hair removal in many patients.

Pharmacologic therapy is directed at interrupting one or more of the steps in the pathway of androgen synthesis and action: (1) suppression of adrenal and/or ovarian androgen production, (2) enhancement of androgen-binding to plasma-binding proteins, particularly SHBG, (3) impairment of the peripheral conversion of androgen precursors to active androgen, and (4) inhibition of androgen action at the target tissue level. Attenuation of hair growth is typically not evident until 4–6 months after initiation of medical treatment and, in most cases, leads to only a modest reduction in hair growth.

Combination estrogen-progestin therapy in the form of an oral contraceptive is usually the first-line endocrine treatment for hirsutism and acne, after dermatologic management. The estrogenic component of most oral contraceptives currently in use is either ethinyl estradiol or mestranol. The suppression of LH leads to reduced production of ovarian androgens. The reduced androgen levels also result in a dose-related increase in SHBG, thus lowering the fraction of unbound plasma testosterone. Estrogens also have a direct, dose-dependent suppressive effect on sebaceous cell function.

The choice of a specific oral contraceptive should be predicated on the progestational component, as progestins vary in their suppressive effect on SHBG levels and in their androgenic potential. Ethynodiol diacetate has relatively low androgenic potential, whereas progestins such as norgestrel and levonorgestrel are particularly androgenic, as judged from their attenuation of the estrogen-induced increase in SHBG. Norgestimate exemplifies the newer generation of progestins that are virtually “nonandrogenic.” Drospirenone, an analogue of spironolactone that has both antimineralocorticoid and antiandrogenic activities, is often used as a progestational agent in combination with ethinyl estradiol, although concern remains about is prothrombotic effects.

Oral contraceptives are contraindicated in women with a history of thromboembolic disease and women with increased risk of breast or other estrogen-dependent cancers **(Chap. 407)**. There is a relative contraindication to the use of oral contraceptives in smokers and those with hypertension or a history of migraine headaches. Improvements in hirsutism are typically in the range of 20%, but there may be an arrest of further progression of hair growth. In most trials, estrogen-progestin therapy alone improves the extent of acne by an average of 50%. The effect on hair growth may not be evident for 6 months, and the maximum effect may require 9–12 months owing to the length of the hair growth cycle

Because oral contraceptives are efficacious and have fewer side effects, they are recommended over glucocorticoids as first-line treatment of hirsutism in CAH. If the response to oral contraceptives is inadequate, glucocorticoids may be used. The lowest effective dose of glucocorticoid should be used (e.g., dexamethasone [0.2–0.5 mg] or prednisone [5–10 mg]) taken at bedtime to achieve maximal suppression by inhibiting the nocturnal surge of ACTH.

Cyproterone acetate is the prototypic antiandrogen. It acts mainly by competitive inhibition of the binding of testosterone and DHT to the androgen receptor. In addition, it may enhance the metabolic clearance of testosterone by inducing hepatic enzymes. Although not available for use in the United States, cyproterone acetate is widely used in Canada, Mexico, and Europe. Cyproterone (50–100 mg) is given on days 1–15, and ethinyl estradiol (50 μg) is given on days 5–26 of the menstrual cycle. Side effects include irregular uterine bleeding, nausea, headache, fatigue, weight gain, and decreased libido.

Spironolactone, which usually is used as a mineralocorticoid antagonist, is also a weak antiandrogen. It is almost as effective as cyproterone acetate when used at high enough doses (100–200 mg daily). Patients should be monitored intermittently for hyperkalemia or hypotension, though these side effects are uncommon. Pregnancy should be avoided because of the risk of feminization of a male fetus. Spironolactone can also cause menstrual irregularity. It often is used in combination with an oral contraceptive, which suppresses ovarian androgen production and helps prevent pregnancy.

Flutamide is a potent nonsteroidal antiandrogen that is effective in treating hirsutism, but concerns about the induction of hepatocellular dysfunction preclude its use. Finasteride is a competitive inhibitor of 5α-reductase type 2. Beneficial effects on hirsutism have been reported, but the predominance of 5α-reductase type 1 in the PSU appears to account for its limited efficacy. Finasteride would also be expected to impair sexual differentiation in a male fetus, and it should not be used in women who may become pregnant. Although studies of dutasteride are limited in number, it appears that this agent may have efficacy in treating scalp hair thinning and loss as well as hirsutism. Dutasteride differs from finasteride as it targets both 5α-reductase types 1 and 2.

Ultimately, the choice of any specific agent(s) must be tailored to the unique needs of the patient being treated. As noted previously, pharmacologic treatments for hirsutism should be used in conjunction with nonpharmacologic approaches. It is also helpful to review the pattern of female hair distribution in the normal population to dispel unrealistic expectations.

FURTHER READING

Azarchi S et al: Androgens in women: Hormone-modulating therapies for skin disease. J Am Acad Derm 80:1509, 2019.

Brown DL et al: Ovarian stromal hyperthecosis: Sonographic features and histologic associations. J Ultrasound Med 28:587, 2009.

Forslund M et al: Different kinds of oral contraceptive pills in polycystic ovary syndrome: A systematic review and meta-analysis Eur J Endocrinol 189:S1, 2023.

Haak CS et al: Hair removal in hirsute women with normal testosterone levels: A randomized controlled trial of long-pulsed diode laser vs. intense pulsed light. Br J Dermatol 163:1007, 2010.

Martin KA et al: Evaluation and treatment of hirsutism in premenopausal women: An Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 103:1233, 2018.

McCartney CR, Marshall JC: Polycystic ovary syndrome. N Engl J Med 375:1398, 2016.

Rosenfield RL, Ehrmann DA: The pathogenesis of polycystic ovary syndrome (PCOS): The hypothesis of PCOS as functional ovarian hyperandrogenism revisited. Endocr Rev 37:467, 2016.

407

Menopause and Postmenopausal Hormone Therapy

JoAnn E. Manson, Shari S. Bassuk

Menopause is the permanent cessation of menstruation due to loss of ovarian follicular function. It is diagnosed retrospectively after 12 months of amenorrhea. The average age at menopause is 51 years among U.S. women. *Perimenopause* refers to the time period preceding menopause, when fertility wanes and menstrual cycle irregularity increases, until the first year after cessation of menses. The onset of perimenopause precedes the final menses by 2–8 years, with a mean duration of 4 years. Smoking accelerates the menopausal transition by 2 years.

Although the peri- and postmenopausal transitions share many symptoms, the physiology and clinical management of the two differ. Low-dose oral contraceptives have become a therapeutic mainstay in perimenopause, whereas postmenopausal hormone therapy (HT) has been a common method of symptom alleviation after menstruation ceases.

PERIMENOPAUSE

PHYSIOLOGY

Ovarian mass and fertility decline sharply after age 35 and even more precipitously during perimenopause; depletion of primary follicles, a process that begins before birth, occurs steadily until menopause **(Chap. 404)**. In perimenopause, intermenstrual intervals shorten significantly (typically by 3 days) as a result of an accelerated follicular phase. Follicle-stimulating hormone (FSH) levels rise because of altered folliculogenesis and reduced inhibin secretion. In contrast to the consistently high FSH and low estradiol levels seen in menopause, perimenopause is characterized by “irregularly irregular” hormone levels. The propensity for anovulatory cycles can produce a hyperestrogenic, hypoprogestagenic environment that may account for the increased incidence of endometrial hyperplasia or carcinoma, uterine polyps, and leiomyoma observed among women of perimenopausal age. Mean serum levels of selected ovarian and pituitary hormones during the menopausal transition are shown in **Fig. 407-1**. With transition into menopause, estradiol levels fall markedly, whereas estrone levels are relatively preserved, a pattern reflecting peripheral aromatization of adrenal and ovarian androgens. Levels of FSH increase more than those of luteinizing hormone, presumably because of the loss of inhibin as well as estrogen feedback.

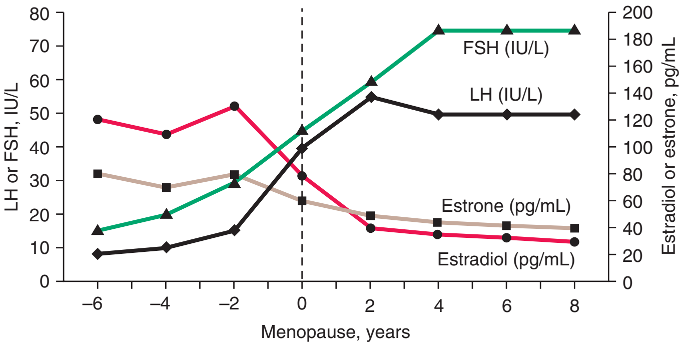

FIGURE 407-1 Mean serum levels of ovarian and pituitary hormones during the menopausal transition. FSH, follicle-stimulating hormone; LH, luteinizing hormone. *(Data from G Rannevik et al: A longitudinal study of the perimenopausal transition: Altered profiles of steroid and pituitary hormones, SHBG and bone mineral density. Maturitas 21:103, 1995.)*

DIAGNOSTIC TESTS

The Stages of Reproductive Aging Workshop +10 (STRAW+10) classification provides a comprehensive framework for the clinical assessment of ovarian aging. As shown in **Fig. 407-2**, menstrual cycle characteristics are the principal criteria for characterizing the menopausal transition, with biomarker measures as supportive criteria. Because of their extreme intraindividual variability, FSH and estradiol levels are imperfect diagnostic indicators of perimenopause in menstruating women. However, a consistently low FSH level in the early follicular phase (days 2–5) of the menstrual cycle does not support a diagnosis of perimenopause, while levels >25 IU/L in a random blood sample are characteristic of the late menopause transition. FSH measurement can also aid in assessing fertility; levels of <20 IU/L, 20 to <30 IU/L, and ≥30 IU/L measured on day 3 of the cycle indicate a good, fair, and poor likelihood of achieving pregnancy, respectively. Anti-müllerian hormone and inhibin B may also be useful for assessing reproductive aging.

SYMPTOMS

Determining whether symptoms that develop in midlife are due to ovarian senescence or to other age-related changes is difficult. There is strong evidence that the menopausal transition can cause hot flashes, night sweats, irregular bleeding, and vaginal dryness, and there is moderate evidence that it can cause sleep disturbances in some women. There is inconclusive or insufficient evidence that ovarian aging is a major cause of mood swings, depression, impaired memory or concentration, somatic symptoms, urinary incontinence, or sexual dysfunction. In one U.S. study, nearly 60% of women reported hot flashes in the 2 years before their final menses. Symptom intensity, duration, frequency, and effects on quality of life are highly variable.

TREATMENT

Perimenopause

PERIMENOPAUSAL THERAPY

For women with irregular or heavy menses or hormone-related symptoms that impair quality of life, low-dose combined oral contraceptives are a staple of therapy. Static doses of estrogen and progestin (e.g., 20 μg of ethinyl estradiol and 1 mg of norethindrone acetate daily for 21 days each month) can eliminate vasomotor symptoms and restore regular cyclicity. Oral contraceptives provide other benefits, including protection against ovarian and endometrial cancers and increased bone density, although it is not clear whether use during perimenopause decreases fracture risk later in life. Moreover, the contraceptive benefit is important, given that the unintentional pregnancy rate among women in their forties rivals that of adolescents. Contraindications to oral contraceptive use include cigarette smoking, liver disease, a history of thromboembolism or cardiovascular disease, breast cancer, or unexplained vaginal bleeding. Progestin-only formulations (e.g., 0.35 mg of norethindrone daily) or medroxyprogesterone (Depo-Provera) injections (e.g., 150 mg IM every 3 months) may provide an alternative for the treatment of perimenopausal menorrhagia in women who smoke or have cardiovascular risk factors. Although progestins neither regularize cycles nor reduce the number of bleeding days, they reduce the volume of menstrual flow.

Nonhormonal strategies to reduce menstrual flow include the use of nonsteroidal anti-inflammatory agents such as mefenamic acid (an initial dose of 500 mg at the start of menses, then 250 mg qid for 2–3 days) or, when medical approaches fail, endometrial ablation. It should be noted that menorrhagia requires an evaluation to rule out uterine disorders. Transvaginal ultrasound with saline enhancement is useful for detecting leiomyomata or polyps, and endometrial aspiration can identify hyperplastic changes.

TRANSITION TO MENOPAUSE

For sexually active women using contraceptive hormones to alleviate perimenopausal symptoms, the question of when and if to

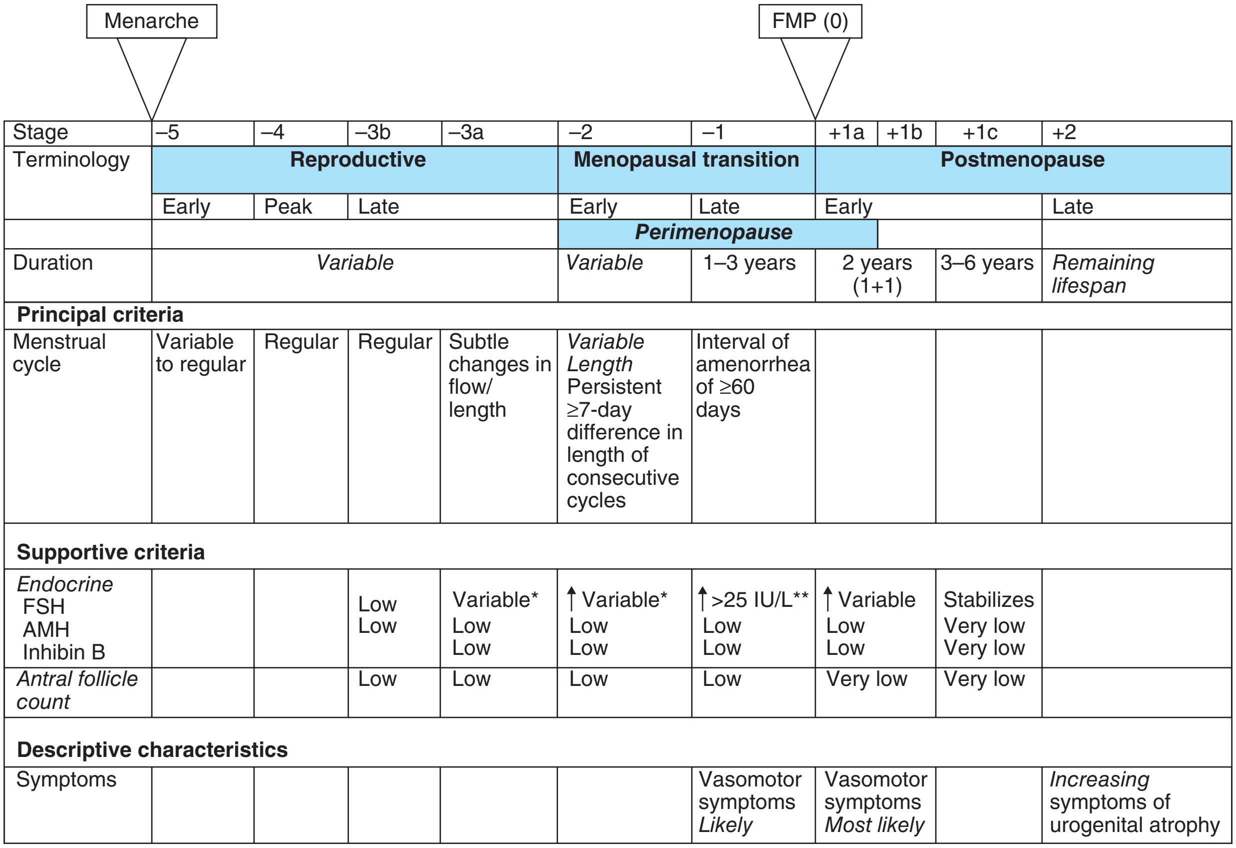
Menarche
FMP (0)

Stage	–5	–4	–3b	–3a	–2	–1	+1a	+1b	+1c	+2
Terminology	**Reproductive**				**Menopausal transition**		**Postmenopause**			
	Early	Peak	Late		Early	Late	Early			Late
					Perimenopause					
Duration	*Variable*				*Variable*	1–3 years	2 years (1+1)		3–6 years	*Remaining lifespan*
Principal criteria										
Menstrual cycle	Variable to regular	Regular	Regular	Subtle changes in flow/ length	*Variable Length* Persistent ≥7-day difference in length of consecutive cycles	Interval of amenorrhea of ≥60 days				
Supportive criteria										
Endocrine FSH AMH Inhibin B			 Low Low	Variable* Low Low	↑ Variable* Low Low	↑ >25 IU/L** Low Low	↑ Variable Low Low		Stabilizes Very low Very low	
Antral follicle count			Low	Low	Low	Low	Very low		Very low	
Descriptive characteristics										
Symptoms						Vasomotor symptoms *Likely*	Vasomotor symptoms *Most likely*			*Increasing* symptoms of urogenital atrophy

*Blood draw on cycle days 2–5 ↑ = elevated.
**Approximate expected level based on assays using current international pituitary standard.

FIGURE 407-2 The Stages of Reproductive Aging Workshop +10 (STRAW +10) staging system for reproductive aging in women. AMH, anti-müllerian hormone; FSH, follicle-stimulating hormone. *(Reproduced with permission from SD Harlow et al: Executive summary of the Stages of Reproductive Aging Workshop + 10: addressing the unfinished agenda of staging reproductive aging. Menopause 19:387, 2012.)*

switch to HT must be individualized. Doses of estrogen and progestogen (either synthetic progestins or natural forms of progesterone) in HT are lower than those in oral contraceptives and have not been documented to prevent pregnancy. Although a 1-year absence of spontaneous menses reliably indicates ovulation cessation, it is not possible to assess the natural menstrual pattern while a woman is taking an oral contraceptive. Women willing to switch to a barrier method of contraception should do so; if menses occur spontaneously, oral contraceptive use can be resumed. The average age of final menses among relatives can serve as a guide for when to initiate this process, which can be repeated yearly until menopause has occurred.

MENOPAUSE AND POSTMENOPAUSAL HT

One of the most complex health care decisions facing women is whether to use postmenopausal HT. Once prescribed primarily to relieve vasomotor symptoms, HT has been promoted as a strategy to forestall various disorders that accelerate after menopause, including osteoporosis and cardiovascular disease. In 2000, nearly 40% of postmenopausal women aged 50–74 in the United States had used HT. This widespread use occurred despite the paucity of conclusive data, until recently, on the health consequences of such therapy. Although many women rely on their health care providers for a definitive answer to the question of whether to use postmenopausal hormones, balancing the benefits and risks for an individual patient is challenging.

Although observational studies suggest that HT prevents cardiovascular and other chronic diseases, the apparent benefits may result at least in part from differences between women who opt to take postmenopausal hormones and women who do not. Those choosing HT tend to be healthier, have greater access to medical care, are more compliant with prescribed treatments, and maintain a more health-promoting lifestyle. Randomized trials, which eliminate these confounding factors, have not consistently confirmed the benefits found in observational studies. Indeed, the largest HT trial to date, the Women's Health Initiative (WHI), which examined >27,000 postmenopausal women aged 50–79 (mean age, 63) for an average of 5–7 years, was stopped early because of an overall unfavorable benefit-risk ratio in the estrogen-progestin arm and an excess risk of stroke that was not offset by a reduced risk of coronary heart disease (CHD) in the estrogen-only arm.

The following summary offers a decision-making guide based on a synthesis of currently available evidence. Prevention of cardiovascular disease is eliminated from the equation due to lack of evidence for such benefits in randomized clinical trials.

■ BENEFITS AND RISKS OF POSTMENOPAUSAL HT

See Table 407-1.

Definite Benefits • SYMPTOMS OF MENOPAUSE Compelling evidence, including data from randomized clinical trials, indicates that estrogen therapy is highly effective for controlling vasomotor and genitourinary symptoms. Alternative approaches, including the use of antidepressants (such as paroxetine mesylate, 7.5 mg/d; venlafaxine, 37.5–75 mg/d; desvenlafaxine, 100 mg/d; fluoxetine, 20–30 mg/d; sertraline, 50–100 mg/d; citalopram, 10–30 mg/d; and escitalopram, 10–20 mg/d); the neurokinin-3 receptor antagonist fezolinetant (45 mg/d); the γ-aminobutyric acid analogue gabapentin (300 mg nightly, up to 900 mg in divided doses); and the antispasmodic anticholinergic oxybutynin (2.5–5 mg twice per day) may also alleviate vasomotor symptoms, although they are less effective than HT. Evidence is inconsistent on the efficacy of pregabalin (75–150 mg/d twice per day) and clonidine (oral, 0.1–1 mg/d, or transdermal, 0.1–0.3 mg weekly). Paroxetine mesylate and fezolinetant are the only nonhormonal drugs approved

TABLE 407-1 Benefits and Risks of Postmenopausal Hormone Therapy in the Overall Study Population of Women aged 50–79 Years in the Intervention Phase of the Women's Health Initiative (WHI) Estrogen-Progestin and Estrogen-Alone Trials[a]

		ESTROGEN-PROGESTIN		ESTROGEN ALONE	
OUTCOME	EFFECT	RELATIVE BENEFIT OR RISK	ABSOLUTE BENEFIT OR RISK[b]	RELATIVE BENEFIT OR RISK	ABSOLUTE BENEFIT OR RISK[b]
Definite Benefits					
Symptoms of menopause	Definite improvement	↓65–90% decreased risk[c]		↓65–90% decreased risk[c]	
Osteoporosis	Definite increase in bone mineral density and decrease in fracture risk	↓33% decreased risk for hip fracture	6 fewer cases (11 vs 17) of hip fracture	↓33% decreased risk for hip fracture	6 fewer cases (13 vs 19) of hip fracture
Definite Risks[h]					
Endometrial cancer	Definite increase in risk with estrogen alone (see below for estrogen-progestin)	See below	See below		4.6 excess cases (observational studies)
Pulmonary embolism	Definite increase in risk	↑98% increased risk	9 excess cases (18 vs 9)	↑35% increased risk (n.s.)	4 excess cases (14 vs 10)
Deep-vein thrombosis	Definite increase in risk	↑87% increased risk	11.5 excess cases (25 vs 14)	↑48% increased risk	7.5 excess cases (23 vs 15)
Breast cancer	Definite increase in risk with long-term use (≥5 years) of estrogen-progestin	↑24% increased risk	8.5 excess cases (43 vs 35)	↓21% decreased risk (n.s.)	7 fewer cases (28 vs 35)
Gallbladder disease	Definite increase in risk	↑57% increased risk	47 excess cases (131 vs 84)	↑55% increased risk	58 excess cases (164 vs 106)
Probable or Uncertain Risks and Benefits[h]					
Coronary heart disease[d]	Probable increase in risk among older women and women many years past menopause; possible decrease in risk or no effect in younger or recently menopausal women[e]	↑18% increased risk (n.s.)	6 excess cases (41 vs 35)	No increase in risk	No difference in risk
Myocardial infarction	Significant interaction by age group for estrogen alone, with reduced risk in younger—but not older—women (*p* for trend by age = .02)	↑24% increased risk (n.s.)	6 excess cases (35 vs 29)	No increase in risk[e]	No difference in risk[e]
Stroke	Probable increase in risk	↑37% increased risk	9 excess cases (33 vs 24)	↑35% increased risk	11 excess cases (45 vs 34)
Ovarian cancer	Probable increase in risk with long-term use (≥5 years)	↑41% increased risk (n.s.)	1 excess case (5 vs 4)	Not available	Not available
Endometrial cancer	Probable decrease in risk with estrogen-progestin during long-term follow-up (see above for estrogen alone)	↓33% decreased risk[f]	3 fewer cases (7 vs 10)	See above	See above
Urinary incontinence	Probable increase in risk	↑49% increased risk	549 excess cases (1661 vs 1112)	↑61% increased risk	852 excess cases (2255 vs 1403)
Colorectal cancer	Probable decrease in risk with estrogen-progestin; possible increase in risk in older women with estrogen alone (*p* for trend by age = .02 for estrogen alone)	↓38% decreased risk	6.5 fewer cases (10 vs 17)	No increase or decrease in risk[e]	No difference in risk[e]
Type 2 diabetes	Probable decrease in risk	↓19% decreased risk	16 fewer cases (72 vs 88)	↓14% decreased risk	21 fewer cases (134 vs 155)
Dementia (age ≥65)	Increase in risk in older women (but inconsistent data from observational studies and randomized trials)	↑101% increased risk	23 excess cases (46 vs 23)	↑47% increased risk (n.s.)	15 excess cases (44 vs 29)
Total mortality	Possible increase in risk among older women and women many years past menopause; possible decrease in risk or no effect in younger or recently menopausal women (*p* for trend by age <.05 for both trials combined)	No increase in risk	No difference in risk	No increase in risk[e]	No difference in risk[e]
Global index[g]	Probable increase in risk or no effect among older women and women many years past menopause; possible decrease in risk or no effect in younger or recently menopausal women (*p* for trend by age = .02 for estrogen alone)	↑12% increased risk	20.5 excess cases (189 vs 168)	No increase in risk[e]	No difference in risk[e]

[a]The estrogen-progestin arm of the WHI assessed 5.6 years of conjugated equine estrogens (0.625 mg/d) plus medroxyprogesterone acetate (2.5 mg/d) versus placebo. The estrogen-alone arm of the WHI assessed 7.1 years of conjugated equine estrogens (0.625 mg/d) versus placebo. [b]Number of cases per 10,000 women per year. [c]The WHI was not designed to assess the effect of hormone therapy (HT) on menopausal symptoms. Data from other randomized trials suggest that HT reduces risk for menopausal symptoms by 65–90%. [d]*Coronary heart disease* is defined as nonfatal myocardial infarction or coronary death. [e]There was a significant interaction by age; that is, the association between HT and the specified outcome was different in younger women and older women. [f]This is the risk reduction that was observed during a cumulative 13-year follow-up period (5.6 years of treatment plus 8.2 years of postintervention observation). [g]The *global index* is a composite outcome representing the first event for each participant from among the following: coronary heart disease, stroke, pulmonary embolism, breast cancer, colorectal cancer, endometrial cancer (estrogen-progestin arm only), hip fracture, and death. Because participants can experience more than one type of event, the global index cannot be derived by a simple summing of the component events. [h]Includes some outcomes where results were divergent between the estrogen-progestin arm and the estrogen-alone arm.

Abbreviation: n.s., not statistically significant.

Source: Data from JE Manson et al: Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women's Health Initiative randomized trials. JAMA 310:1353, 2013.

by the U.S. Food and Drug Administration for treatment of vasomotor symptoms. Bazedoxifene, an estrogen agonist/antagonist, in combination with conjugated estrogens has also received approval for this use. Cognitive behavioral therapy and clinical hypnosis have been shown in randomized trials to help with vasomotor symptom management. The consumption of phytoestrogens, including soy isoflavones, may be effective in women who are able to metabolize the isoflavone daidzein to the biologically active metabolite equol. Weight loss, mindfulness-based stress reduction, and stellate ganglion block are also promising strategies, although more trials are needed. For genitourinary syndrome of menopause, the efficacy of low-dose vaginal estrogen is similar to that of oral or transdermal estrogen; oral ospemifene and vaginal prasterone are additional options.

OSTEOPOROSIS (See also Chap. 423)

Bone density By reducing bone turnover and resorption rates, estrogen slows the aging-related bone loss experienced by most postmenopausal women. More than 50 randomized trials have demonstrated that postmenopausal estrogen therapy, with or without a progestogen, rapidly increases bone mineral density at the spine by 4–6% and at the hip by 2–3% and that those increases are maintained during treatment.

Fractures Data from observational studies indicate a 50–80% lower risk of vertebral fracture and a 25–30% lower risk of hip, wrist, and other peripheral fractures among current estrogen users; addition of a progestogen does not appear to modify this benefit. In the WHI, 5–7 years of either combined estrogen-progestin or estrogen-only therapy was associated with a 33% reduction in hip fractures and 25–30% fewer total fractures among a population unselected for osteoporosis. Bisphosphonates (such as alendronate, 10 mg/d or 70 mg once per week; risedronate, 5 mg/d or 35 mg once per week; ibandronate, 2.5 mg/d or 150 mg once per month or 3 mg every 3 months IV; or zoledronic acid, 5 mg once per year IV) and denosumab (60 mg twice per year SC) increase bone mass density by reducing bone resorption and have been shown in randomized trials to decrease fracture rates. Other treatment options include bazedoxifene in combination with conjugated estrogens; the selective estrogen receptor modulator (SERM) raloxifene (60 mg/d); and parathyroid hormone (teriparatide, 20 μg/d SC). Unlike estrogen, these alternative therapies do not appear to have adverse effects on the endometrium or breast. Increased weight-bearing and resistance exercise; adequate calcium intake (1000–1200 mg/d through diet or supplements in two or three divided doses); and adequate vitamin D intake (600–1000 IU/d) may also reduce the risk of osteoporosis-related fractures. According to a 2011 report by the Institute of Medicine (now the National Academy of Medicine), 25-hydroxyvitamin D blood levels of ≥50 nmol/L are sufficient for bone-density maintenance and fracture prevention. The Fracture Risk Assessment (FRAX®) score, an algorithm that combines an individual's bone-density score with age and other risk factors to predict her 10-year risk of hip and major osteoporotic fracture, may be of use in guiding decisions about pharmacologic treatment (see *www.sheffield.ac.uk/FRAX/*).

Definite Risks • **ENDOMETRIAL CANCER (WITH ESTROGEN ALONE)** A combined analysis of 30 observational studies found a tripling of endometrial cancer risk among short-term users (1–5 years) of unopposed estrogen and a nearly 10-fold increased risk among long-term users (≥10 years). These findings are supported by results from the randomized Postmenopausal Estrogen/Progestin Interventions (PEPI) trial, in which 24% of women assigned to unopposed estrogen for 3 years developed atypical endometrial hyperplasia—a premalignant lesion—as opposed to only 1% of women assigned to placebo. Use of a progestogen, which opposes the effects of estrogen on the endometrium, eliminates these risks and may even reduce risk (see later).

VENOUS THROMBOEMBOLISM A meta-analysis of observational studies found that current oral estrogen use was associated with a 2.5-fold increase in risk of venous thromboembolism in postmenopausal women. A meta-analysis of randomized trials, including the WHI, found a 2.1-fold increase in risk. Results from the WHI indicate a nearly twofold increase in risk of pulmonary embolism and deep-vein thrombosis with estrogen-progestin and a 35–50% increase in these risks with estrogen-only therapy. Transdermal estrogen, taken alone or with certain progestogens (micronized progesterone or pregnane derivatives), appears to be a safer alternative with respect to thrombotic risk.

BREAST CANCER (WITH ESTROGEN-PROGESTIN) An increased risk of breast cancer has been found among current or recent estrogen users in observational studies; this risk is directly related to duration of use. In contrast to findings for endometrial cancer, combined estrogen-progestin regimens appear to increase breast cancer risk more than estrogen alone. In a 2019 meta-analysis of 58 observational studies (108,647 breast cancer cases) published between 1992 and 2017, the relative risks (RRs) for incident breast cancer with current use of estrogen-progestin for 1–4, 5–9, 10–14, and ≥15 years were 1.60 (95% confidence interval [CI], 1.52–1.69), 1.97 (95% CI, 1.90–2.04), 2.26 (95% CI, 2.16–2.36), and 2.51 (95% CI, 2.35–2.68), respectively, compared with never use. The corresponding statistics for incident breast cancer with current use of estrogen alone were 1.17 (1.10–1.26), 1.22 (1.17–1.28), 1.43 (1.37–1.50), and 1.58 (1.51–1.66), respectively. Data from randomized trials also indicate that estrogen-progestin raises breast cancer risk. In the WHI, women assigned to receive combination hormones for an average of 5.6 years were 24% more likely to develop breast cancer than women assigned to placebo, but 7.1 years of estrogen-only therapy did not increase risk. Indeed, the WHI showed a trend toward a reduction in breast cancer risk with estrogen alone, although it is unclear whether this finding would pertain to formulations of estrogen other than conjugated equine estrogens or to treatment durations of >7 years. In the Heart and Estrogen/Progestin Replacement Study (HERS), combination therapy for 4 years was associated with a 27% increase in breast cancer risk. Although the latter finding was not statistically significant, the totality of evidence strongly implicates estrogen-progestin therapy in breast carcinogenesis.

Some observational data suggest that the length of the interval between menopause onset and HT initiation may influence the association between such therapy and breast cancer risk, with a "gap time" of <3–5 years conferring a higher HT-associated breast cancer risk. (This pattern of findings contrasts with that for CHD, as discussed later in this chapter.) However, this association remains inconclusive and may be a spurious finding attributable to higher rates of screening mammography and thus earlier cancer detection in HT users than in nonusers, especially in early menopause. Indeed, in the WHI trial, relative risk (RRs) for HT and breast cancer risk did not differ among women 50–59, those 60–69, and those 70–79 years of age at trial entry. (There was insufficient power to examine finer age categories.) Additional research is needed to clarify the issue.

GALLBLADDER DISEASE Large observational studies report a two- to threefold increased risk of gallstones or cholecystectomy among postmenopausal women taking oral estrogen. In the WHI, women randomized to estrogen-progestin or estrogen alone were ~55% more likely to develop gallbladder disease than those assigned to placebo. Risks were also increased in HERS. Transdermal HT might be a safer alternative, but further research is needed.

Probable or Uncertain Risks and Benefits • **CORONARY HEART DISEASE/STROKE** Until recently, HT had been enthusiastically recommended as a possible cardioprotective agent. In the past three decades, multiple observational studies suggested, in the aggregate, that estrogen use leads to a 35–50% reduction in CHD incidence among postmenopausal women. The biologic plausibility of such an association is supported by data from randomized trials demonstrating that exogenous estrogen lowers plasma low-density lipoprotein (LDL) cholesterol levels and raises high-density lipoprotein (HDL) cholesterol levels by 10–15%. Administration of estrogen also favorably affects lipoprotein(a) levels, LDL oxidation, endothelial vascular function, fibrinogen, and plasminogen activator inhibitor 1. However, estrogen therapy has unfavorable effects on other biomarkers of cardiovascular risk: it boosts triglyceride levels; promotes coagulation via factor VII, prothrombin fragments 1 and 2, and fibrinopeptide A elevations; and raises levels of the inflammatory marker C-reactive protein.

Randomized trials of estrogen or combined estrogen-progestin in women with preexisting cardiovascular disease have not confirmed the benefits reported in observational studies. In HERS (a secondary-prevention trial designed to test the efficacy and safety of estrogen-progestin therapy with regard to clinical cardiovascular outcomes), the 4-year incidence of coronary death and nonfatal myocardial infarction was similar in the active-treatment and placebo groups, and a 50% increase in risk of coronary events was noted during the first year among participants assigned to the active-treatment group. Although it is possible that progestin may mitigate estrogen's benefits, the Estrogen Replacement and Atherosclerosis (ERA) trial indicated that angiographically determined progression of coronary atherosclerosis was unaffected by either opposed or unopposed estrogen treatment. Moreover, no cardiovascular benefit was found in the Papworth Hormone Replacement Therapy Atherosclerosis Study, a trial of transdermal estradiol with and without norethindrone; the Women's Estrogen for Stroke Trial (WEST), a trial of oral 17β-estradiol; or the Estrogen in the Prevention of Reinfarction Trial (ESPRIT), a trial of oral estradiol valerate. Thus, in clinical trials, HT has not proved effective for the secondary prevention of cardiovascular disease in postmenopausal women.

Primary-prevention trials also suggest an early increase in cardiovascular risk and an absence of cardioprotection with postmenopausal HT. In the WHI, women assigned to 5.6 years of estrogen-progestin therapy were 18% more likely to develop CHD (defined in primary analyses as nonfatal myocardial infarction or coronary death) than those assigned to placebo, although this risk elevation was not statistically significant. However, during the trial's first year, there was a significant 80% increase in risk, which diminished in subsequent years (*p* for trend by time = .03). In the estrogen-only arm of the WHI, no overall effect on CHD was observed during the 7.1 years of the trial or in any specific year of follow-up. This pattern of results was similar to that for the outcome of total myocardial infarction.

However, a closer look at available data suggests that timing of initiation of HT may critically influence the association between such therapy and CHD. Estrogen may slow early stages of atherosclerosis but have adverse effects on advanced atherosclerotic lesions. It has been hypothesized that the prothrombotic and proinflammatory effects of estrogen manifest themselves predominantly among women with subclinical lesions who initiate HT well after the menopausal transition, whereas women with less arterial damage who start HT early in menopause may derive cardiovascular benefit because they have not yet developed advanced lesions. Data from experiments in nonhuman primates and from some randomized trials in humans support this concept. Conjugated estrogens had no effect on the extent of coronary artery plaque in cynomolgus monkeys assigned to receive estrogen alone or combined with progestin starting 2 years (~6 years in human terms) after oophorectomy and well after the establishment of atherosclerosis. However, administration of exogenous hormones immediately after oophorectomy, during the early stages of atherosclerosis, reduced the extent of plaque by 70%. In the Early versus Late Intervention Trial with Estradiol (ELITE), a 6-year trial among 643 healthy postmenopausal women that was designed to test whether effects of estrogen on the development and progression of atherosclerosis depend on age at initiation of therapy, oral 17β-estradiol administered with or without vaginal micronized progesterone significantly slowed carotid atherosclerotic progression in women within 6 years of menopause onset (mean age, 55.4 years) but not in women >10 years past menopause onset (mean age, 65.4 years) (*p* for interaction = .007). On the other hand, in the Kronos Early Estrogen Prevention Study (KEEPS), a 4-year trial among 729 healthy postmenopausal women within 3 years of menopause onset at trial entry (mean age, 53 years), neither oral conjugated estrogens nor transdermal estradiol, administered with oral micronized progesterone, affected carotid atherosclerotic progression. However, the low prevalence of this endpoint in the overall study population may have curtailed power to detect a treatment difference.

Lending further credence to the timing hypothesis are results of subgroup analyses of data from observational studies and large clinical trials. For example, among women who entered the WHI trial with a relatively favorable cholesterol profile, estrogen with or without progestin led to a 40% lower risk of incident CHD. Among women who entered with a worse cholesterol profile, therapy resulted in a 73% higher risk (*p* for interaction = .02). The presence or absence of the metabolic syndrome **(Chap. 420)** also strongly influenced the relation between HT and incident CHD. Among women with the metabolic syndrome, HT more than doubled CHD risk, whereas no association was observed among women without the syndrome. Moreover, although there was no association between estrogen-only therapy and CHD in the WHI trial cohort as a whole, such therapy was associated with a CHD risk reduction of 40% among participants aged 50–59; in contrast, a risk reduction of only 5% was observed among those aged 60–69, and a risk increase of 9% was found among those aged 70–79 (*p* for trend by age = .08). For the outcome of total myocardial infarction, estrogen alone was associated with a borderline-significant 45% reduction and a nonsignificant 24% increase in risk among the youngest and oldest women, respectively (*p* for trend by age = .02). Estrogen was also associated with lower levels of coronary artery calcified plaque in the younger age group. Although age did not have a similar effect in the estrogen-progestin arm of the WHI, CHD risks increased with years since menopause (*p* for trend = .08), with a significantly elevated risk among women who were ≥20 years past menopause. For the outcome of total myocardial infarction, estrogen-progestin was associated with a 9% risk reduction among women <10 years past menopause as opposed to a 16% increase in risk among women 10–19 years past menopause and a twofold increase in risk among women >20 years past menopause (*p* for trend = .01). In the large observational Nurses' Health Study, women who chose to start HT within 4 years of menopause experienced a lower risk of CHD than did nonusers, whereas those who began therapy ≥10 years after menopause appeared to receive little coronary benefit. Observational studies include a high proportion of women who begin HT within 3–4 years of menopause, whereas clinical trials include a high proportion of women ≥12 years past menopause; this difference helps to reconcile some of the apparent discrepancies between the two types of studies.

For the outcome of stroke, WHI participants assigned to estrogen-progestin or estrogen alone were ~35% more likely to suffer a stroke than those assigned to placebo. Whether or not age at initiation of HT influences stroke risk is not well understood. In the WHI and the Nurses' Health Study, HT was associated with an excess risk of stroke in all age groups. Further research is needed on age, time since menopause, and other individual characteristics (including biomarkers) that predict increases or decreases in cardiovascular risk associated with exogenous HT. Furthermore, it remains uncertain whether different doses, formulations, or routes of administration of HT will produce different cardiovascular effects.

COLORECTAL CANCER Observational studies have suggested that HT reduces risks of colon and rectal cancer, although the estimated magnitudes of the relative benefits have ranged from 8 to 34% in various meta-analyses. In the WHI (the sole trial to examine the issue), estrogen-progestin was associated with a significant 38% reduction in colorectal cancer over a 5.6-year period, although no benefit was seen with 7 years of estrogen-only therapy. However, a modifying effect of age was observed, with a doubling of risk with HT in women aged 70–79 but no risk elevation in younger women (*p* for trend by age = .02).

COGNITIVE DECLINE AND DEMENTIA A meta-analysis of 10 case-control and two cohort studies suggested that postmenopausal HT is associated with a 34% decreased risk of dementia. Subsequent randomized trials (including the WHI), however, have failed to demonstrate any benefit of estrogen or estrogen-progestin therapy on the progression of mild to moderate Alzheimer's disease and/or have indicated a potential adverse effect of HT on the incidence of dementia, at least in women ≥65 years of age. Among women randomized to HT (as opposed to placebo) at age 50–55 in the WHI, no effect on cognition was observed during the postintervention phase. Determining whether timing of initiation of HT influences cognitive outcomes will require further study.

OVARIAN CANCER AND OTHER DISORDERS On the basis of limited observational and randomized data, it has been hypothesized that HT increases the risk of ovarian cancer and reduces the risk of type 2 diabetes mellitus. Results from the WHI support these hypotheses. The WHI also found that HT use was associated with an increased risk of urinary incontinence and that estrogen-progestin was associated with increased rates of lung cancer mortality.

ENDOMETRIAL CANCER (WITH ESTROGEN-PROGESTIN) In the WHI, use of estrogen-progestin was associated with a nonsignificant 17% reduction in risk of endometrial cancer. A significant reduction in risk emerged during the postintervention period (see later).

ALL-CAUSE MORTALITY In the overall WHI cohort, estrogen with or without progestin was not associated with all-cause mortality. However, there was a trend toward reduced mortality in younger women, particularly with estrogen alone. For women aged 50–59, 60–69, and 70–79 years, RRs associated with estrogen-only therapy were 0.70, 1.01, and 1.21, respectively (*p* for trend = .04).

OVERALL BENEFIT-RISK PROFILE Estrogen-progestin was associated with an unfavorable benefit-risk profile (excluding relief from menopausal symptoms) as measured by a "global index"—a composite outcome including CHD, stroke, pulmonary embolism, breast cancer, colorectal cancer, endometrial cancer, hip fracture, and death (Table 407-1)—in the WHI cohort as a whole, and this association did not vary by 10-year age group. Estrogen-only therapy was associated with a neutral benefit-risk profile in the WHI cohort as a whole. However, there was a significant trend toward a more favorable benefit-risk profile among younger women and a less favorable profile among older women, with RRs of 0.84, 0.99, and 1.17 for women aged 50–59, 60–69, and 70–79 years, respectively (*p* for trend by age = .02). The balance of benefits and risks of estrogen with and without progestin among women aged 50–59 is shown in Fig. 407-3.

CHANGES IN HEALTH STATUS AFTER DISCONTINUATION OF HT In the WHI, many but not all risks and benefits associated with active use of HT dissipated within 5–7 years after discontinuation of therapy. For estrogen-progestin, an elevated risk of breast cancer persisted (RR = 1.28; 95% CI, 1.11–1.48) during a median cumulative 13-year follow-up period (5.6 years of treatment plus 8.2 years of postintervention observation), but most cardiovascular disease risks became neutral. A reduction in hip fracture risk persisted (RR = 0.81; 95% CI, 0.68–0.97), and a significant reduction in endometrial cancer risk emerged (RR = 0.67; 95% CI, 0.49–0.91). For estrogen alone, the reduction in breast cancer risk became statistically significant (RR = 0.79; 95% CI, 0.65–0.97) during a median cumulative 13-year follow-up period (6.8 years of treatment plus 6.6 years of postintervention observation), and significant differences by age group persisted for total myocardial infarction and the global index, with more favorable results for younger women. During a median cumulative 18-year follow-up, estrogen alone was

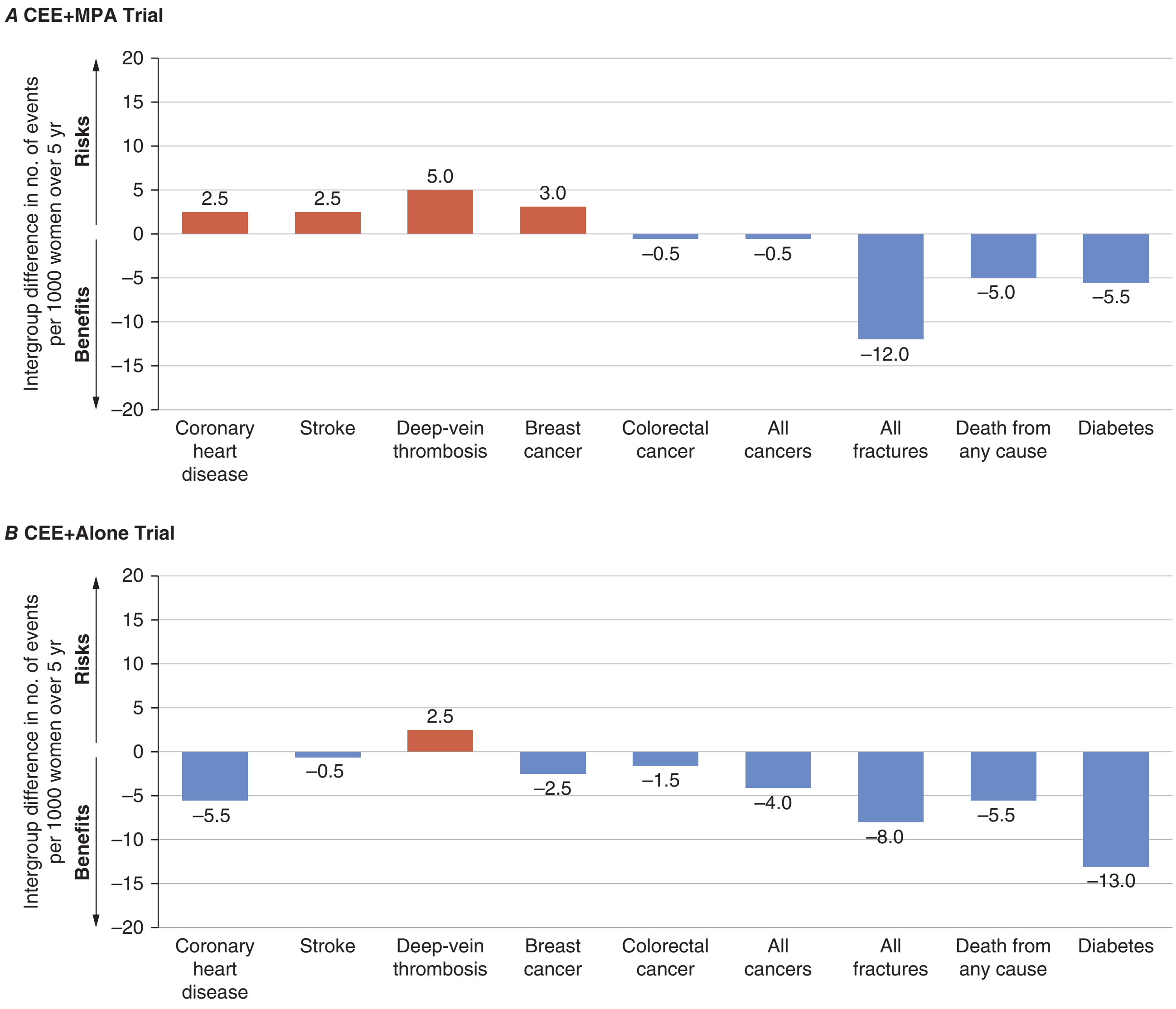

FIGURE 407-3 Benefits and risks of the two hormone therapy (HT) formulations evaluated in the Women's Health Initiative, in women aged 50–59 years. Results are shown for the two formulations, conjugated equine estrogens (CEE) alone or in combination with medroxyprogesterone acetate (MPA). Risks and benefits are expressed as the difference in number of events (number in the HT group minus the number in the placebo group) per 1000 women over 5 years. *(Reproduced with permission from JE Manson, AM Kaunitz: Menopause Management–Getting Clinical Care Back on Track. N Engl J Med 374:803, 2016.)*

associated with a significant reduction in all-cause mortality in women aged 50–59 years (RR = 0.79; 95% CI, 0.64–0.96); the protective effect was seen primarily in those with bilateral oophorectomy (RR = 0.68; 95% CI, 0.48–0.96). During a median cumulative 20-year follow-up, estrogen-progestin was associated with a significant elevation in breast cancer risk (RR = 1.28; 95% CI, 1.13–1.45) and a suggestive elevation in breast cancer mortality (RR =1.35; 95% CI, 0.94–1.95; p = .11), whereas estrogen alone was associated with a significant reduction in breast cancer risk (RR = 0.78; 95% CI, 0.65–0.93) and breast cancer mortality (RR = 0.60; 95% CI, 0.37–0.97).

APPROACH TO THE PATIENT

Postmenopausal HT

The rational use of postmenopausal HT requires balancing the potential benefits and risks. Table 407-2 provides one approach to decision-making. This approach applies to women with menopausal symptoms who are age 45 years and older or to women who have had removal of both ovaries, regardless of age. Women below age 45 years or those with uncertain menopausal status may need additional clinical evaluation before determining a management plan. The clinician should first assess whether the patient has moderate to severe hot flashes and/or night sweats—the primary indication for initiation of systemic HT—that do not subside in response to behavioral/lifestyle modifications, such as lowering the thermostat, dressing in layers, avoiding warm beverages, and not smoking. Systemic HT may also be used to prevent osteoporosis in women at high risk of fracture who cannot tolerate alternative osteoporosis therapies. (Vaginal estrogen or other medications may be used to treat genitourinary syndrome of menopause in the absence of vasomotor symptoms [see later].) The benefits and risks of such therapy should be reviewed with the patient, giving more emphasis to absolute than to relative measures of effect and pointing out uncertainties in clinical knowledge where relevant. Because chronic disease rates generally increase with age, absolute risks tend to be greater in older women, even when RRs remain similar. Potential side effects—especially vaginal bleeding that may result from use of the combined estrogen-progestogen formulations recommended for women with an intact uterus—should be noted. The patient's own preference regarding therapy should be elicited and factored into the decision. Contraindications should be assessed routinely and include unexplained vaginal bleeding; liver dysfunction or disease; venous thromboembolism; known blood clotting disorder or thrombophilia (transdermal estrogen may be an option); untreated hypertension; history of endometrial cancer (except stage 1 without deep invasion), breast cancer, or other estrogen-dependent cancer; and history of CHD, stroke, or transient ischemic attack. Relative contraindications to systemic HT include an elevated risk of breast cancer (e.g., women who have one or more first-degree relatives with breast cancer, susceptibility genes such as *BRCA1* or *BRCA2*, or a personal history of cellular atypia detected by breast biopsy); hypertriglyceridemia (>400 mg/dL); an elevated risk of cardiovascular disease; and active gallbladder disease (transdermal estrogen may be an option in the latter three cases because it has a less adverse effect on triglyceride levels, clotting factors, and inflammation factors than oral HT). Primary prevention of heart disease should not be viewed as an expected benefit of HT, and an increase in the risk of stroke as well as a small early increase in the risk of coronary artery disease should be considered. Nevertheless, such therapy may be appropriate if the noncoronary benefits of treatment clearly outweigh the risks. Reassess benefits and risks at least once every 6–12 months, assuming the patient's continued preference for HT, or if the patient's health status changes. A woman who suffers an acute coronary event or stroke while taking HT should discontinue therapy immediately.

TABLE 407-2 Approach to Initiating Menopausal Hormone Therapy for Vasomotor Symptom Management

1. Vasomotor symptom assessment		
Confirm that hot flashes and/or night sweats are adversely affecting sleep, daytime functioning, or quality of life.		
2. Risk factor assessment		
Confirm that there are no absolute contraindications to menopausal hormone therapy		
Breast, endometrial, or other estrogen-dependent cancer		
Cardiovascular disease (heart disease, stroke, transient ischemic attack)		
Active liver disease		
Undiagnosed vaginal bleeding		
3. Menopausal hormone therapy initiation		
RECOMMEND	**CONSIDER WITH CAUTION**	**AVOID**
Age <60 years and Menopause onset within 10 years and Low risk of breast cancer[a] and cardiovascular disease[b]	Age ≥60 years ••••••• OR ••••••• Menopause onset >10 years prior ••••••• OR ••••••• Moderate risk of breast cancer[a] or cardiovascular disease[a]	High risk of breast cancer[a] or cardiovascular disease[b] ••••••••• OR ••••••••• Age ≥60 years or menopause onset >10 years prior and Moderate risk of breast cancer[a] or cardiovascular disease[b]

[a]For online tools to assess breast cancer risk, see AH McClintock et al: Breast cancer risk assessment: A step-wise approach for primary care providers on the front lines of shared decision making. Mayo Clin Proc 95:126, 2020. [b]For online tools to assess cardiovascular disease risk, see D Lloyd-Jones et al: Use of risk assessment tools to guide decision-making in the primary prevention of atherosclerotic cardiovascular disease: A special report from the American Heart Association and American College of Cardiology. Circulation 139:e1162, 2019.

Source: Adapted from JL Shifren et al: JAMA 2019; 321:2458-2459.

Many options for systemic HT are available. Estrogen alone is recommended for women with hysterectomy, whereas estrogen plus progestogen is recommended for women with a uterus. In the United States, the most commonly prescribed oral estrogens for systemic treatment of vasomotor symptoms are 17β-estradiol (1.0 or 0.5 mg/d or other doses) and conjugated equine estrogens (CEE; 0.625, 0.45, or 0.3 mg/d or other doses). The most commonly prescribed transdermal estrogen products are 17β-estradiol skin patches (0.035 or 0.05 mg/d or other doses). The most commonly prescribed progestogens are medroxyprogesterone acetate (MPA; 2.5, 5, or 10 mg/d) and micronized progesterone (100 or 200 mg/d). Also available are oral estrogen-progestin combinations, such as oral CEE and MPA, oral 17β-estradiol or ethinyl estradiol with norethindrone acetate, oral estradiol with progesterone, and other options. CEE/bazedoxifene may be an option for women with a uterus, especially those with concerns about breast tenderness, breast density, or uterine bleeding. Contraindications to CEE/bazedoxifene are similar to those for systemic HT.

Short-term use (<5 years for estrogen-progestogen and <7 years for estrogen alone) is appropriate for relief of menopausal symptoms among women without contraindications to such use. However, such therapy should be avoided by women with an elevated baseline risk of future cardiovascular events. Women who have contraindications for or are opposed to HT may derive benefit from the use of certain antidepressants (including paroxetine, mesylate, venlafaxine, fluoxetine, and others), fezolinetant, gabapentin, or oxybutynin.

Long-term use (≥5 years for estrogen-progestogen and ≥7 years for estrogen alone) is more problematic because a heightened risk of breast cancer must be factored into the decision, especially for estrogen-progestogen. Reasonable candidates for such use include postmenopausal women who have persistent severe vasomotor symptoms along with an increased risk of osteoporosis (e.g., those with osteopenia, a personal or family history of nontraumatic fracture, or a weight <125 lb), who also have no personal or family history of breast cancer in a first-degree relative or other contraindications, and who have a strong personal preference for therapy. Poor candidates are women with elevated cardiovascular risk, those at increased risk of breast cancer, and those at low risk of osteoporosis. Even for reasonable candidates, strategies to minimize dose and duration of use should be employed. For example, women using HT to relieve intense vasomotor symptoms in early postmenopause should consider discontinuing therapy within 5 years, resuming it only if such symptoms persist. Because of the role of progestogens in increasing breast cancer risk, regimens that employ cyclic rather than continuous progestogen exposure as well as formulations other than MPA should be considered if treatment is extended. For prevention of osteoporosis, alternative therapies such as bisphosphonates or SERMs should be considered. Research on alternative progestogens and androgen-containing preparations has been limited, particularly with respect to long-term safety. Additional research on the effects of these agents on cardiovascular disease, glucose tolerance, and breast cancer will be of particular interest.

For genitourinary symptoms such as vaginal dryness or pain with intercourse/sexual activity, intravaginal estrogen creams, tablets, or rings; prasterone (vaginal dehydroepiandrosterone); and ospemifene are options. Contraindications to low-dose vaginal estrogen include unexplained vaginal bleeding or breast cancer, endometrial cancer, or other estrogen-dependent cancer. Contraindications to ospemifene and prasterone are the same as those for low-dose vaginal estrogen, and contraindications for ospemifene additionally include venous or arterial thromboembolic disease, severe liver disease, and use of estrogens or estrogen agonists-antagonists.

In addition to HT, lifestyle choices such as smoking abstention, adequate physical activity, and a healthy diet can play a role in controlling symptoms and preventing chronic disease. An expanding array of pharmacologic options (e.g., bisphosphonates, SERMs, and other agents for osteoporosis; cholesterol-lowering or antihypertensive agents for cardiovascular disease) should also reduce the widespread reliance on hormone use. However, short-term HT may still benefit some women.

FURTHER READING

Bassuk SS, Manson JE: Menopausal hormone therapy and cardiovascular disease risk: Utility of biomarkers and clinical factors for risk stratification. Clin Chem 60:68, 2014.

Chlebowski RT et al: Association of menopausal hormone therapy with breast cancer incidence and mortality during long-term follow-up of the Women's Health Initiative randomized clinical trials. JAMA 324:369, 2020.

Crandall CM et al: Management of menopausal symptoms: A review. JAMA 329:5, 2023.

Duralde ER et al: Management of perimenopausal and menopausal symptoms. BMJ 382:e072612, 2023.

Manson JE, Bassuk SS: Hot Flashes, *Hormones and Your Health*. New York, McGraw-Hill, 2007.

Manson JE et al: Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women's Health Initiative randomized trials. JAMA 310:1353, 2013.

Manson JE et al: The Women's Health Initiative trials of menopausal hormone therapy: Lessons learned. Menopause 27:918, 2020.

North American Menopause Society: The 2020 genitourinary syndrome of menopause position statement of the North American Menopause Society. Menopause 27:976, 2020.

North American Menopause Society: The 2022 hormone therapy position statement of the North American Menopause Society. Menopause 29:767, 2022.

Shifren JL et al: Menopausal hormone therapy. JAMA 321:2458, 2019.

408 Infertility and Contraception

Anuja Dokras, Janet E. Hall

INFERTILITY

The World Health Organization (WHO) categorizes infertility as a disease of the reproductive system. Infertility is the third most common disease worldwide, affecting ~48–72 million couples. It is defined as the inability to achieve a pregnancy over 12 months of unprotected intercourse. The prevalence of infertility, ~17.5% globally, has remained relatively stable over the past few decades. Primary infertility occurs in couples who have never achieved a pregnancy, whereas secondary infertility refers to infertility after achieving at least one pregnancy. During the first year of attempting pregnancy, the fecundability rate, defined as the ability to achieve a pregnancy within one menstrual cycle, is highest in the first 3 months and declines over the next 9 months. Approximately 85% of couples will achieve pregnancy after 12 months, and 95% will achieve pregnancy after 24 months. Increasing trends toward later childbearing can have significant implications due to age-related decrease in the fecundability rate. Compared to women aged 30–31 years of age, fecundability is reduced by 14% in women aged 34–35 years, 19% in women aged 36–37 years, 53% in women aged 40–41 years, and 59% in women aged 42–44 years.

ETIOLOGY

The causes for infertility are generally classified as female factors, male factors, and unexplained infertility **(Fig. 408-1)**. The female causes include tubal factors (pelvic inflammatory disease, endometriosis, prior surgery, salpingitis isthmica nodosum), uterine etiology (fibroids, congenital malformations, uterine scarring), ovulatory dysfunction (polycystic ovary syndrome [PCOS], diminished ovarian reserve, premature ovarian insufficiency), and endocrine dysfunction (hypothyroidism, hyperprolactinemia). Although the probability of achieving a pregnancy decreases after the age of 35 in women, primarily due to chromosomal abnormalities in the oocyte during meiosis, a similar decline has not been observed in men <50 years of age. The male causes of infertility include anatomic factors in the reproductive system (vasectomy, infection, absence of the vas), endocrine factors (hypogonadotropic hypogonadism, hypothyroidism, hyperprolactinemia, morbid obesity, use of certain medications), sexual dysfunction (erectile or ejaculatory dysfunction, decreased libido), and genetic factors contributing to primary testicular dysfunction, including defects in spermatogenesis (Klinefelter's syndrome, Y chromosome microdeletions). The distribution of these causes varies significantly across the world. Overall, female factors are present in 30–40% of couples with infertility, male factors are present in 40–50%, and both male and female factors are identified in 20–30%. Unexplained infertility refers to the absence of any identified abnormality after completing the fertility workup and occurs in up to 30% of couples. As a result, a complete workup of both partners is recommended in all couples presenting with infertility.

FERTILITY EVALUATION

Diagnostic evaluation for infertility is typically initiated after 1 year of unprotected intercourse because 80–85% of couples will achieve a pregnancy over this time period. Evaluation can be initiated even prior to meeting the definition of infertility, especially if one of the partners has risk factors for infertility. If the female partner's age is >35 years, it is recommended to initiate evaluation after 6 months of attempting pregnancy. If the age of the female partner is >40 years, it is recommended to start evaluating the couple immediately. The initial evaluation should include detailed medical history, laboratory testing, radiologic evaluation, and preconception counseling for both partners. As multiple causes for infertility may be identified, it is best to perform the complete diagnostic evaluation prior to initiating treatment.

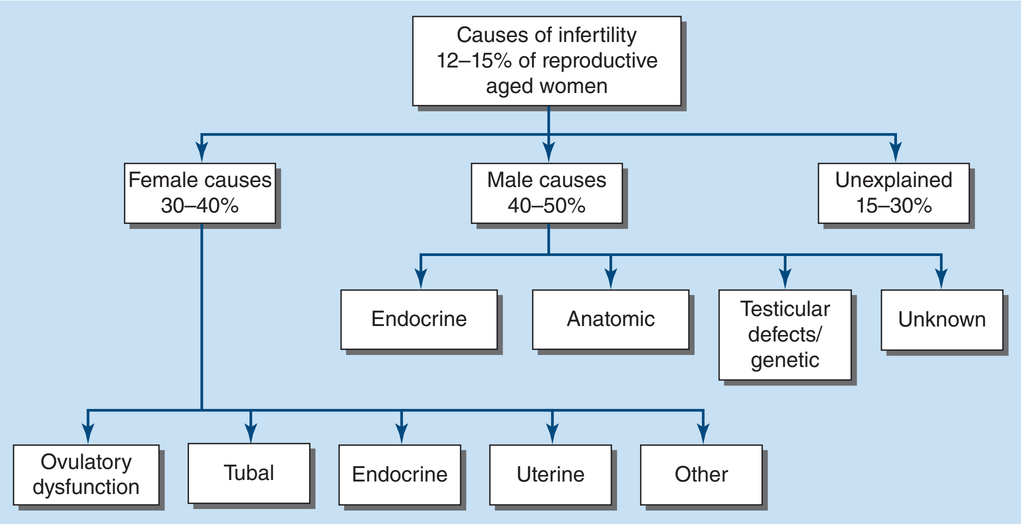

FIGURE 408-1 Causes of infertility.

History and Physical Exam A detailed history obtained from both partners is essential to identify risk factors for infertility. In the female partner, gynecologic history (menstrual frequency, menorrhagia, dysmenorrhea, history of sexually transmitted infections, endometriosis), medical and endocrine history, exposure to pelvic radiation, abdominal or pelvic surgeries, tobacco and alcohol use, medication use including cytotoxic drugs, family history of early menopause, and prior history of pregnancy should be assessed. In addition, frequency of intercourse, timing of intercourse, use of methods to detect ovulation, and concerns regarding sexual dysfunction over the past several months should be ascertained. Physical exam in the female partner should include assessment of weight and blood pressure (BP), thyroid and breast exam, assessment for signs of hyperandrogenism, and pelvic exam to assess uterine size, adnexal masses, and factors that might impact intercourse. Similarly, a detailed history should be obtained in the male partner with specific questions regarding injuries and surgery in the reproductive tract; mumps orchitis; exposure to pelvic radiation; use of androgens, cytotoxic drugs, and other medications; and fertility with any prior partner. The exam in the male partner should include body mass index (BMI), BP, and complete physical exam including testicular exam.

Ultrasound An abdominal and transvaginal pelvic ultrasound can assess uterine (myomas, adenomyosis, müllerian anomalies) and adnexal abnormalities (endometriosis, polycystic-appearing ovaries) and evaluate ovarian reserve (number of antral follicles in both ovaries).

Ovulation Assessment Women who have regular menstrual cycles between 25 and 35 days will typically have ovulatory cycles. Ovulation can be assessed by using ovulation detection strips at home to detect urinary luteinizing hormone (LH) or by measuring a serum progesterone level 7 days after ovulation. Basal body temperatures can also be used to confirm ovulation when a rise in temperature is noted in the luteal phase. However, basal body temperature measurements are less reliable than the above methods.

Hysterosalpingogram An hysterosalpingogram (HSG) is performed during the follicular phase to assess the patency of fallopian tubes by injecting radiopaque contrast through the cervix into the uterus and imaging the flow of contrast through one or both tubes. In addition to identifying tubal pathology, an HSG may identify intrauterine abnormalities such as polyps, submucosal myomas, and adhesions. Although the negative predictive value of HSG for assessing tubal patency is high, the positive predictive value is relatively low. Interestingly, pregnancy rates have been shown to be higher after an HSG test compared to no testing and higher when oil-based contrast was used compared to water-based contrast, likely related to tubal flushing of mucus plugs. Alternate options that are increasingly used include injection of agitated saline contrast through the cervix into the uterus. Tubal patency is assessed by demonstrating passage of agitated saline contrast through the tubes or accumulation in the cul de sac as visualized by ultrasonography. A saline infusion sonogram is more accurate in assessing intrauterine pathology such as polyps and intrauterine scarring compared to HSG and can be combined with ultrasound assessment of the pelvis.

Ovarian Reserve Evaluation Assessment of ovarian reserve includes measurement of serum FSH and estradiol on day 2 or 3 of the menstrual cycle and serum anti-müllerian hormone (AMH). These screening tests combined with age of the female partner and antral follicle counts measured by ultrasound can identify diminished ovarian reserve and provide information on the urgency to initiate treatment. AMH and antral follicle counts are also used to determine starting doses of gonadotropins for fertility treatments. These markers of ovarian reserve, however, do not predict the likelihood of pregnancy and live birth.

Endocrine Tests In women with irregular menses, serum TSH, prolactin, and androgens (total and free testosterone) should be measured to identify other causes for anovulation.

Semen Analysis (see Chap. 403) The semen sample is collected after 2–7 days of abstinence and provides an assessment of sperm count, motility, morphology, volume, and pH. None of the individual sperm parameters are predictive of fertility, but the likelihood of infertility increases with multiple abnormalities. Those with abnormal sperm parameters based on the WHO criteria (*oligoasthenozoospermia* is defined as sperm counts <15 million/mL, motility <40%, and normal morphology <4%) should have a physical exam and endocrine evaluation (serum follicle-stimulating hormone [FSH], LH, prolactin, and thyroid-stimulating hormone [TSH]); those with azoospermia or severe oligospermia (<5 million/mL) should have genetic evaluation (karyotype and Y chromosome microdeletion). Although a DNA sperm fragmentation assay is not part of the initial evaluation, it may be indicated in patients with recurrent pregnancy loss. Sperm antibody testing and scrotal ultrasound should not be routinely performed in infertile men.

Genetic Screening All couples can be offered preconception genetic screening based on ethnicity, family history, or common autosomal recessive conditions.

Of note, diagnostic laparoscopy, postcoital test, endometrial biopsy, thrombophilia, and immunologic testing and karyotype are not indicated as part of the initial workup of infertility.

COUNSELING AND TREATMENT

Preconception Counseling All patients seeking fertility care should be provided with preconception counseling to identify modifiable risks and optimize pregnancy outcomes. This includes counseling about eating disorders or lifestyle modifications for weight management as obesity in women is associated with an increase in anovulatory cycles, miscarriage rates, and maternal and fetal complications in pregnancy. Obesity in men is associated with abnormal sperm parameters. Preconception counseling regarding smoking cessation is important as evidence suggests that smoking cessation can reverse the detrimental impact of smoking on fecundity. Smoking decreases fertility rates by a direct impact on oocyte DNA and also increases the risk of miscarriage and ectopic pregnancy. In addition, smoking during pregnancy is associated with an increased risk of placental abruption and intrauterine growth restriction (IUGR). Moreover, the impact of smoking on ovarian reserve has been shown to accelerate the time to menopause by 1–4 years. As high levels of caffeine consumption increase the risk of infertility and miscarriage, women should be counseled to restrict caffeine consumption to ≤2 cups while attempting pregnancy and during pregnancy. Use of testosterone products, which are widely used for the treatment of hypoandrogenism and sexual dysfunction in men, should be stopped. Inquiries should be made about possible misuse of androgens for physical appearance or performance enhancement **(Chap. 411)**. As part of the preconception counseling, patients should be informed that the fertile window is typically 5–6 days prior to ovulation, and therefore, intercourse every 1–2 days during this time period will increase the chance of pregnancy. Various methods are used by women to detect ovulation, including basal body temperature measurements, assessment of changes in cervical mucus, and urinary LH kits. A rise in basal body temperatures indicates that ovulation has occurred and therefore cannot be used to time intercourse. LH kits can be used to detect the start of ovulation and subsequently time intercourse on the day of the LH surge and the following day. Physicians should counsel patients that advanced maternal age (>35 years) is associated with a higher risk of aneuploidy and advanced paternal age (>40 years) is associated with adverse health outcomes in the offspring.

Treatment Treatment recommendations depend on the results of the fertility evaluation described above **(Table 408-1)**. The success of different treatments depends on several factors including age of the female partner, assessment of ovarian reserve, history of smoking, BMI, and race.

Tubal Factor Infertility Tubal factor infertility constitutes 30–35% of cases of female infertility, and a large majority are secondary to tubal obstruction resulting from sexually transmitted infections. In vitro fertilization (IVF) was first developed as a treatment for tubal factor infertility as it bypasses the fallopian tubes and allows fertilization of oocytes in the laboratory prior to transcervical transfer into the uterus. IVF offers the highest success rates for couples with tubal factor infertility. Tubal repair or reconstruction is typically not recommended in cases associated with tubal infections or hydrosalpinx, due to both the low success rate in achieving tubal patency and increased risk of ectopic pregnancy. In fact, removal of hydrosalpinges by salpingectomy will improve pregnancy rates in subsequent IVF treatments as it prevents efflux of tubal fluid into the uterine cavity. If a proximal tubal blockage is observed on HSG, radiographically guided cannulation of fallopian tubes can be attempted. In women with bilateral tubal ligation, the decision between microsurgical reanastomosis versus IVF will depend on a number of factors including patient's age, ovarian reserve, number of children desired, partner's semen parameters, experience of the surgeon, and cost of procedure.

TABLE 408-1 Assisted Reproductive Technologies

Ovulation induction Oral agents Injectable hormones	Clomiphene citrate (selective estrogen response modulator) Letrozole (aromatase inhibitor) FSH, LH (gonadotropins)
Intrauterine insemination (IUI)	Office-based procedure by which washed and concentrated ejaculated sperm is deposited in the uterine cavity via a soft catheter passed through the cervix
In vitro fertilization (IVF)	Oocytes are harvested transvaginally under local anesthesia or intravenous sedation and incubated with sperm to facilitate fertilization. The fertilized embryos are cultured for 3 days (cleavage stage) or 5 days (blastocyst stage) prior to transcervical placement of one or more embryos, depending on the age of the female patient, into the uterine cavity under ultrasound guidance.
Intracytoplasmic sperm injection (ICSI)	In cases of severe male factor infertility, a single motile morphologically normal appearing sperm is injected into the oocyte for potential fertilization.

Abbreviations: FSH, follicle-stimulating hormone; LH, luteinizing hormone.

Ovulatory Dysfunction Endocrine conditions such as hypothyroidism and hyperprolactinemia should be treated prior to use of ovulation induction medications. Lifestyle modifications should be recommended in patients with low BMI or obesity. Weight loss in obese women has been shown to increase the likelihood of spontaneous or drug-induced ovulation. First-line treatment for anovulatory infertility (most common etiology is PCOS) includes use of letrozole followed by clomiphene citrate to induce ovulation **(Chap. 406)**. A large majority of women with PCOS (60–80%) respond to these oral medications, and the addition of metformin, combined with the above medications as a second-line agent, may further increase the chance of ovulation, particularly in obese women. In women with hypothalamic amenorrhea, behavioral modifications such as weight gain and decreased exercise may resume ovulation. If there is no response, judicious use of low-dose injectable gonadotropins can induce monofollicular growth. In women with diminished ovarian reserve, treatment can be escalated from ovulation induction with oral medications and intrauterine insemination (IUI) to IVF, as the overall live birth rates are lower. In both women with diminished ovarian reserve and women with premature ovarian insufficiency, the option of using donor oocytes can be offered. In that case, the egg donor will undergo the IVF procedure, the harvested eggs are fertilized with the male partner's sperm, and the fertilized embryos will be transferred to the patient's uterus.

Male Infertility Given the high prevalence of male factor infertility (40–50%), timely evaluation and treatment are recommended. Men with abnormal semen parameters have associated health risks and should have a detailed evaluation by specialists in male reproduction. In men with no sperm (azoospermia) in the ejaculate, further evaluation including a repeat semen analysis followed by physical examination, endocrine tests, and genetics studies should be performed to identify obstructive (40% prevalence) versus nonobstructive etiology. First-line treatment for mild to moderate male factor infertility includes IUI alone or IUI combined with ovulation induction, depending on the female partner's age and other causes of infertility. In men with severe male factor infertility (sperm count <5 million/mL or motility <20%), IVF with intracytoplasmic sperm injection (ICSI) is recommended. In men with obstructive azoospermia, sperm can be procured by direct aspiration from the epididymis or testis. In men with congenital bilateral absence of the vas deferens (CBAVD), testing for *CFTR* mutations and genetic counseling are indicated before to offering IVF with ICSI. In men with nonobstructive azoospermia, microsurgical sperm retrieval from the testes may result in successful pregnancies after IVF-ICSI; however, the use of donor sperm for IUI is an alternate option. Men with hypogonadotropic hypogonadism (e.g., Kallmann's syndrome) can be treated with gonadotropins to initiate spermatogenesis followed by IUI or IVF. Treatment of male sexual dysfunction and avoidance of exogenous androgens are effective strategies for addressing male factor infertility. Repair of a moderate to large varicocele is recommended when associated with abnormal semen parameters or if the patient is symptomatic from the varicocele; however, it may take several months to detect an improvement in semen parameters.

Unexplained Infertility In 15–30% of couples, no clear causes of infertility are identified. In such cases, it is appropriate to initiate ovarian stimulation with oral medications to increase the number of developing oocytes combined with IUI timed to ovulation in order to increase the number of motile sperm in the reproductive tract. Depending on the age of the female partner, this approach offers modest success rates limiting its use to 3–6 months before recommending IVF. Overall, IVF is associated with a low risk of complications; the risk of ovarian hyperstimulation syndrome is significantly decreased by judiciously monitoring stimulation and using gonadotropin-releasing hormone (GnRH) to trigger ovulation instead of human chorionic gonadotropin (hCG). Multiple pregnancy remains the highest risk associated with IVF despite improvements in cryopreservation of embryos and age-based guidelines for limiting the number of embryos to transfer. In some couples, the IVF treatment may reveal an underlying cause of infertility such as lower fertilization, embryo cleavage, or blastocyst formation rates. Of note, guidelines from different medical societies around the world vary in the rapidity of offering IVF for unexplained infertility.

Uterine Factors Fibroids are the most common benign tumors of the reproductive tract and occur in 50–70% of reproductive-age women. It is not clear whether fibroids decrease the likelihood of pregnancy; submucosal fibroids and intramural fibroids that distort the endometrial cavity may lower pregnancy rates and increase the risk of pregnancy loss. Removal of submucosal fibroids, uterine polyps, and intrauterine adhesions hysteroscopically may improve subsequent pregnancy rates.

Endometriosis Endometriosis is a common gynecologic condition associated with pelvic pain and dysmenorrhea, and in severe cases, it is associated with tubo-ovarian infertility. Approximately 25–50% of infertile women have endometriosis, and 30–50% of women with endometriosis have infertility. Prolonged medical management to suppress endometriotic lesions and surgical treatment of stage 1 and 2 endometriosis have not been shown to improve subsequent fertility rates. Surgical removal of endometriotic lesions or endometriomas in women with stage 3 or 4 endometriosis may improve subsequent pregnancy rates. First-line treatment of infertility associated with endometriosis alone includes use of oral ovulation induction medications and IUI.

PSYCHOLOGICAL ASPECTS OF INFERTILITY

It is well recognized that infertility is associated with psychological stress related not only to the diagnostic and therapeutic procedures themselves but also to repeated cycles of hope and loss associated with each new procedure or cycle of treatment that does not result in the birth of a child. These feelings are often combined with a sense of isolation from friends and family. Counseling and stress-management techniques should be offered early in the evaluation of infertility as many patients do not pursue treatments after the initial consultation. Importantly, infertility and its treatment do not appear to be associated with long-term psychological sequelae.

CONTRACEPTION

The desired ideal number of children per family varies around the globe and is approximately 2.6 in the United States. Couples not using any form of contraception have an 85% chance of achieving a pregnancy over 1 year. Based on these data, couples spend most of their reproductive life preventing a pregnancy and a much smaller proportion attempting to become or being pregnant. It is therefore not surprising that a majority of women who have been sexually active will have used some form of contraception to prevent a pregnancy. Unintended pregnancies primarily occur due to lack of use or inconsistent use of contraceptives rather than failure of the contraceptive method used. Of the different forms of contraception used worldwide in 2022, tubal sterilization was the most common (~219 million) followed by use of male condom (208 million), intrauterine device (IUD) (161 million), and the (birth control) pill (150 million). Contraceptive methods used by married women differ from those used by single women, and the most widely used contraceptive methods differ by world regions. The rates of female sterilization increased steadily in the last century and now show a slight decrease, likely due to the increasing use of long-acting reversible contraceptive (LARC) agents, such as IUDs and implants, which are as effective as sterilization. The convenience of use of contraceptives determines their compliance and efficacy; contraceptives requiring daily and coitus-related use have higher failure rates compared to long-acting reversible and permanent methods. The U.S. Medical Eligibility Criteria (USMEC) for contraceptive use are evidence-based guidelines to help health care providers recommend appropriate contraceptives to women with chronic medical conditions (Table 408-2). This excellent resource is adapted from the WHO guidance and is kept up to date through continual review of published literature.

TABLE 408-2 U.S. Medical Eligibility Criteria (USMEC) for Contraceptive Use

USMEC Category 4 (a condition that represents an unacceptable health risk if the contraceptive method is used)
Smoking: women age ≥35 years who smoke ≥15 cigarettes per day
Known ischemic heart disease or multiple risk factors for cardiovascular disease (older age, smoking, diabetes, low HDL, high LDL, high triglycerides, and hypertension)
Acute DVT
Previous thromboembolic event; high risk of recurrent DVT
Stroke or known thrombogenic mutations
Complicated valvular heart disease
Peripartum cardiomyopathy (<6 months, moderately to severely impaired cardiac function)
Complicated solid organ transplantation
Hypertension (systolic ≥160 mmHg or diastolic ≥100 mmHg, vascular disease)
Systemic lupus erythematous (positive or unknown antiphospholipid antibodies)
Cirrhosis, hepatocellular adenoma or hepatoma (malignant)
Viral hepatitis, acute flare
Pregnancy and early postpartum (<21 days)
Breast-feeding <21days postpartum
Breast cancer
Diabetes: neuropathy/retinopathy/nephropathy
Migraines with aura
USMEC Category 3 (a condition for which the theoretical or proven risks outweigh the advantages for using the method)
Smoking: women ≥35 years who smoke <15 cigarettes/day
Previous thromboembolic event; lower risk of recurrent DVT
Superficial thrombosis (acute or history of)
Past history of breast cancer and no evidence for 5 years
Hypertension (adequately controlled or systolic 140–159 mmHg or diastolic 90–99 mmHg)
Anticonvulsant drug therapy (certain anticonvulsants (phenytoin, carbamazepine, barbiturates, primidone, topiramate, oxcarbazepine)
Antimicrobial therapy: rifampin or rifabutin
Antiretroviral therapy for prevention (preexposure prophylaxis) or treatment of HIV
Bariatric surgery (Roux-en-Y gastric bypass or biliopancreatic diversion)
Breast-feeding 21–42 days postpartum with or without risk factors for VTE

Abbreviation: DVT, deep-vein thrombosis; HDL, high-density lipoprotein; LDL, low-density lipoprotein; VTE, venous thromboembolism.

TYPES OF CONTRACEPTION

These can be classified in a number of ways, such as permanent versus reversible, hormonal versus nonhormonal, or barrier versus nonbarrier (Table 408-3).

Permanent Contraception The permanent forms of contraception include tubal sterilization and vasectomy. Male sterilization has declined globally with a rate of <2% in 2022. Vasectomy is a low-risk procedure typically performed in an outpatient setting with a very low failure rate of 0.1 pregnancies per 100 women per year. It is not immediately effective, and patients should be told to use other forms of contraception for a minimum of 3 months after the procedure. Globally, tubal sterilization rates have also declined steadily and represent 23% of all methods. Tubal sterilization can be performed in the postpartum period or as an interval procedure and has a failure rate of 0.5 pregnancies per 100 women per year. Postpartum sterilization can be

TABLE 408-3 Effectiveness of Different Forms of Contraception

METHOD OF CONTRACEPTION	THEORETICAL EFFECTIVENESS (%)	ACTUAL EFFECTIVENESS (%)	CONTINUED USE AT 1 YEAR (%)	USE OF CONTRACEPTIVE METHOD BY U.S. WOMEN AGE 15–49 (%)
No method	15	15		34.7
Fertility awareness	96	76	47	1.2
Withdrawal	96	78	46	4.4
Barrier methods				
Condoms	98	82	43	8.4
Diaphragm	94	82	57	2
Spermicides	82	72	43	1
Sterilization				
Female	99.5	99.5	100	18.1
Male	99.5	99.9	100	5.6
Intrauterine device				10.4
Copper T	99.4	99.8	85	
Progestin-containing	99.8	99.8	88	
Hormonal contraceptives				
Combined and progestin only	99.7	91	67	14
Transdermal patch	99.7	91	67	0.5
Vaginal ring	99.7	91	67	1.8
Implant				3.1
Depo-Provera	99.8	94	56	
Subdermal implant	99.5	99.5	84	
Emergency contraception	95	-	-	11

Sources: Data from J Trussell et al: Contraceptive Efficacy, in *Contraceptive Technology*, 20th revised ed. RA Hatcher et al (eds). New York, Ardent Media, 2011; CDC. NCHS National Survey of Family Growth, 2011-2013; J Jones et al: Current contraceptive use in the United States, 2006-2010, and changes in patterns of use since 1995. Natl Health Stat Report 60:1, 2012; and NE Birgisson et al: Preventing unintended pregnancy: The contraceptive CHOICE project in review. J Womens Health (Larchmt) 24:349, 2015. Current Contraceptive Status Among Women Aged 15–49: United States, 2017–2019 NCHS Data Brief No. 388, October 2020. Available at: *https://www.cdc.gov/nchs/products/databriefs/db388.htm*.

performed during a cesarean section or after a vaginal delivery via mini-laparotomy. Interval procedures can be performed laparoscopically or via mini-laparotomy and include partial or complete salpingectomy or occlusion of the fallopian tubes using electrocoagulation or mechanical devices such as clips. These permanent methods of contraception are highly effective as they avoid the need for user-dependent contraception. All patients should undergo preprocedure counseling regarding risk of failure, permanence of the procedure, regret, and alternatives.

Hormonal Contraceptives • **COMBINED ESTROGEN- AND PROGESTIN-CONTAINING CONTRACEPTIVES** The mechanism of action of the hormonal contraceptives involves negative feedback from continuous estrogen administration, thereby decreasing FSH secretion, follicular development, and formation of a dominant follicle. The continuous progestin suppresses LH secretion and inhibits ovulation, alters endometrial receptivity, thickens the cervical mucus, and impairs tubal motility. These hormones can be delivered via oral pills to be taken daily, as a transdermal patch that is changed weekly, or a vaginal ring that is replaced monthly or annually. There are numerous *pills* available containing different doses of estrogen (<50 μg) and types of estrogen and progestins and varying doses within a pack (monophasic vs multiphasic); the pills can be taken in a cyclic or extended cycle schedule. The contraceptive efficacy is similar with varying doses of estrogen and progestin. Decreasing the duration of hormone-free days may decrease some side effects associated with menses, such as menstrual migraines and dysmenorrhea. The overall failure rate for combined hormonal contraceptives is 8 pregnancies per 100 women per year, although compliance with daily use of pills may be lower, affecting efficacy. The contraceptive patch and vaginal ring have higher compliance compared to daily pills. Use of the *contraceptive patch* is associated with a low risk of skin reactions and a lower efficacy in women weighing >90 kg. The transdermal mode of delivery is associated with a higher steady state comparable to that of a 40-μg ethinyl estradiol oral contraceptive. Hormonal contraceptives offer additional benefits such as regulation of menstrual cycles; suppression of ovarian cysts; and decrease in menorrhagia, dysmenorrhea, and hyperandrogenism symptoms; in addition, they reduce the risk of both endometrial (50% reduction) and ovarian cancer (40% reduction). Common side effects include nausea, breast tenderness, bloating, and intermenstrual bleeding. There may be a mild increase in BP in some patients, and it is recommended to check BP at follow-up visits. In large studies and meta-analyses, hormonal contraceptives are not associated with significant weight gain, mood changes, or effect on libido. Prior to administering hormonal contraceptives, a detailed patient history should be obtained to determine any absolute or relative contraindications to their use. Due to the low but slightly increased risk of deep-vein thrombosis (DVT) associated with estrogen-containing hormonal contraceptives (3–15 per 10,000 women-years), they are contraindicated in the immediate postpartum period, in smokers over the age of 35 years, and in women with a history of hereditary thrombophilias or DVT. The association between risk of DVT and different doses of estrogen (ethinyl estradiol <35 μg) or different routes of administration (transdermal patch) is weak. There is, however, some association between third- and fourth-generation progestins and increased risk of DVT. Routine screening for familial thrombotic disorders is not recommended prior to prescribing hormonal contraceptives. Although obesity is associated with decreased fertility, the vast majority of women with obesity do not experience infertility. The USMEC classifies obesity alone as risk category 2, where the benefits of taking hormonal contraceptives outweigh any theoretical risk.

PROGESTIN-ONLY HORMONAL CONTRACEPTION Different types of progestins are used for contraception in oral pills, injectable forms, subdermal implants, and IUDs and may be an option for women who have contraindications to the use of estrogen-containing contraceptives (e.g., migraine with aura, DVT, stroke, breast-feeding). The failure rate with *progestin-only pills* is 9 pregnancies per 100 women per year, whereas the failure rate of progestin IUDs is 0.1 pregnancies per 100 women per year. In addition to acting as a spermicidal, the *levonorgestrel IUD* also thickens the cervical mucus and thins the endometrium, thereby decreasing its receptivity. The common side effects include irregular bleeding, acne, breast tenderness, and pain, with higher rates of expulsion when IUDs are inserted in the immediate postpartum period. Breakthrough bleeding or unscheduled bleeding is commonly reported, as estrogen usually serves to stabilize the endometrial lining and prolonged exposure to progestin alone results in a thinner decidualized lining. Depending on the device used, the progestin IUD is effective for 3–7 years. The *injectable form of progesterone* (medroxyprogesterone acetate) is administered intramuscularly or subcutaneously every 3 months with a failure rate of 3 pregnancies per 100 women per year. Its side effects include weight gain, irregular menses, amenorrhea, and mood changes, and there is a slow return to ovulation and fertility after discontinuation (6–9 months). The *subdermal implant* contains etonogestrel and is placed easily over the triceps muscle in the inner arm using local anesthesia. It lasts up to 5 years and has a failure rate of 0.05 pregnancies per 100 women per year. Findings from the Contraceptive Choice research project showed that continuation rates were higher for LARC (IUDs and implants) compared to short-acting methods. LARCs are the most effective reversible form of contraception with high continuation and satisfaction rates; hence, they are a good choice in adolescents and nulliparous women.

Nonhormonal IUD IUDs are a commonly used form of contraception worldwide and are available as hormonal and nonhormonal devices. The nonhormonal copper IUD works as a spermicidal and is effective for up to 12 years with a failure rate of <1 pregnancy per 100 women per year. Patients should be counseled regarding the increased risk of heavy vaginal bleeding and dysmenorrhea resulting in higher discontinuation rates compared to the levonorgestrel-containing IUDs. IUDs can be used in adolescents and adult women and are typically inserted and removed as an office procedure with use of mild analgesics. They can be inserted anytime during a menstrual cycle, referred to as interval insertion, and in the immediate postpartum and postabortion period.

Barrier Contraception The barrier forms of contraception include condoms (male, female) and diaphragm and cervical cap and have lower effectiveness secondary to inconsistent and incorrect use. They offer several advantages including minimal side effects, lower cost, no requirement for a prescription, and protection from sexually transmitted infections. The failure rate for male and female condoms is 17–21 pregnancies per 100 women per year. Spermicidals can be used in conjunction with barrier methods to improve effectiveness.

Lactational Contraception Lactation may serve as an effective form of contraception during the first 6 postpartum months if there is exclusive breast-feeding and menstrual cycles have not resumed. The contraceptive effect occurs due to suppression of GnRH pulsatility associated with suckling. The failure rate under these circumstances can be as low as 0.5–1.5 pregnancies per 100 women per year.

Fertility Awareness The standard days method is typically used by women with regular menstrual cycles whereby they track their cycles to avoid intercourse from cycle days 8–19. The rhythm and withdrawal method are also referred to as traditional methods of contraception.

Emergency Contraception Also known as postcoital contraception, this method is used after an unprotected or inadequately protected act of intercourse. The probability of pregnancy independent of the time of the month is 8%, but the probability varies significantly in relation to proximity to ovulation and may be as high has 30%. Many women are not aware of the availability of emergency contraception and its appropriate use. As the probability of pregnancy is highest if there has been unprotected intercourse during the 3 days prior to ovulation, the timing of administration and type of emergency contraceptive used determine the efficacy. Emergency contraception options include the copper IUD and oral medications such as ulipristal acetate, levonorgestrel, and combined hormonal pills. The copper IUD prevents fertilization and implantation and is the most effective choice if inserted within 5 days of unprotected intercourse. It can be offered to obese women in whom other hormonal forms of emergency contraceptive may be less effective. Ulipristal acetate, a progesterone receptor antagonist, blocks the ability of endogenous progesterone to act on its receptors and inhibits the LH surge, delaying or inhibiting ovulation, and may directly inhibit follicular rupture. It is administered as a 30-mg single dose up to 5 days after unprotected intercourse. Levonorgestrel administered as a single dose will prevent or delay ovulation and is associated with fewer side effects compared to combined hormonal pills. Overall, the failure rate for all hormonal emergency contraception is 1–3%, with ulipristal acetate being the most effective. Side effects are mild and may include nausea, irregular vaginal bleeding, and fatigue. Emergency contraception should be offered to all women who ask for it up to 5 days after unprotected intercourse and not delayed in order to obtain a pregnancy test or perform a clinical examination. Although body weight can affect the efficacy of emergency hormonal contraception, treatment should not be withheld from overweight and obese women.

CONTRACEPTION COUNSELING

Patients should be provided information regarding the different methods of contraception, side effects, noncontraceptive benefits, efficacy, need for strict compliance, and impact on future fertility. In order to facilitate patient-centric care, the provider should discuss plans for future pregnancy and whether childbearing is complete. A detailed patient history should be reviewed to identify potential contraindications such as migraines with aura, smoking, and hypertension. Providers should refer to the most updated USMEC or WHO Medical Eligibility Criteria for Contraceptive Use guidelines when counseling patients with associated comorbidities. As part of the shared decision-making approach, the patient's choice should be the guiding factor, and the discussion should be nonjudgmental. Adolescents should be offered access to the full range of contraceptive options. In a low-risk patient, hormonal contraceptives can be prescribed from menarche to menopause; regular evaluation of side effects and assessment of changes in the patient's medical history, however, are required.

FURTHER READING

American College of Obstetricians and Gynecologists: Effectiveness of birth control. Available at *https://www.acog.org/womens-health/infographics/effectiveness-of-birth-control-methods*. Accessed June 30, 2024.

Centers for Disease Control and Prevention: Reproductive health. Available at *https://www.cdc.gov/reproductivehealth/Infertility/#e*. Accessed November 22, 2023.

Curtis KM, Peipert JF: Long-acting reversible contraception. N Engl J Med 376:461, 2017.

Curtis KM et al: U.S. medical eligibility criteria for contraceptive use, 2016. MMWR Recomm Rep 65:1, 2016.

Infertility Workup for the Women's Health Specialist: ACOG Committee Opinion, Number 781. Obstet Gynecol 133:e377, 2019. Reaffirmed in February 2023.

Kulkarni AD et al: Fertility treatments and multiple births in the United States. N Engl J Med 369:2218, 2013.

Mascarenhas MN et al. National, regional, and global trends in infertility prevalence since 1990: A systematic analysis of 277 health surveys. PLoS Med 9:e1001356, 2012.

Slama R et al: Estimation of the frequency of involuntary infertility on a nation-wide basis. Hum Reprod 27:1489, 2012.

Steiner AZ et al: Association between biomarkers of ovarian reserve and infertility among older women of reproductive age. JAMA 318:1367, 2017.

World Family Planning 2022. Available at *https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/documents/2023/Feb/undesa_pd_2022_world-family-planning.pdf*. Accessed November 22, 2023.

World Health Organization: Infertility. Available at *https://www.who.int/news-room/fact-sheets/detail/infertility*. Accessed December 23, 2020.

World Health Organization: WHO Laboratory Manual for the Examination and Processing of Human Semen. 6th ed. WHO Press; Geneva, Switzerland: 2021. Available at *https://www.who.int/publications/i/item/9789240030787*. Accessed December 3, 2021.

409 Sexual Dysfunction

Kevin T. McVary

Male sexual dysfunction affects up to 31% of middle-aged and elderly men, whereas female sexual dysfunction, although studied less intensely, has a higher prevalence (43%) than male sexual dysfunction. Demographic changes, the popularity of newer treatments, and greater awareness of sexual dysfunction by patients and society have led to increased diagnosis and associated health care expenditures for the management of this common disorder. Sexual health and satisfaction with sex life are important aspects of quality of life for many, including those in poor health. Because many patients are reluctant to initiate discussion of their sex lives, physicians should address this topic directly to elicit a history of sexual dysfunction. Specifically addressing sexual health should be a routine part of the clinical encounter, particularly in those with cardiovascular risk factors.

MALE SEXUAL DYSFUNCTION

PHYSIOLOGY OF MALE SEXUAL RESPONSE

An erection is a neurovascular event, and the cardiovascular system needs to be intact for sexual stimulation to successfully result in an erection. Normal male sexual function includes (1) sufficient libido, (2) the ability to achieve and maintain penile erection, (3) ejaculation, and (4) detumescence. *Libido* refers to sexual desire and is influenced by a variety of visual, olfactory, tactile, auditory, imaginative, and hormonal stimuli. Sex steroids, particularly testosterone, act to increase libido. Libido can be diminished by emotional context, systemic illness, hormonal disturbances, psychiatric disorders, and medications.

Penile tumescence leading to erection depends on an increased flow of blood into the lacunar network accompanied by complete relaxation of the arteries and corporal smooth muscle. The microarchitecture of the corpora is composed of a mass of smooth muscle (trabecula) that contains a network of endothelial-lined vessels (lacunar spaces). Subsequent compression of the trabecular smooth muscle against the fibroelastic tunica albuginea causes a passive closure of the emissary veins and accumulation of blood in the corpora. In the presence of a full erection and a competent valve mechanism, the corpora become noncompressible cylinders from which blood does not escape. This cascade of relaxation and venous occlusion culminates in a rigid erection.

The central nervous system (CNS) exerts an important influence by either stimulating or antagonizing spinal pathways that mediate erectile function and ejaculation. The erectile response is mediated by a combination of central (psychogenic) innervation and peripheral (reflexogenic) innervation. Sensory nerves that originate from receptors in the penile skin and glans converge to form the dorsal nerve of the penis, which travels to the S2-S4 dorsal root ganglia via the pudendal nerve. Parasympathetic nerve fibers to the penis arise from neurons in the intermediolateral columns of the S2-S4 sacral spinal segments. Sympathetic innervation originates from the T-11 to the L-2 spinal segments and descends through the hypogastric plexus.

Neural input to smooth-muscle tone is crucial to the initiation and maintenance of an erection. There is also an intricate interaction between the corporal smooth-muscle cell and its overlying endothelial cell lining **(Fig. 409-1)**. Nitric oxide, which induces vascular relaxation, promotes erection and is opposed by endothelin 1 (ET-1) and Rho kinase, which mediate vascular contraction. Nitric oxide is synthesized from L-arginine by nitric oxide synthase (NOS) and is released from the nonadrenergic, noncholinergic (NANC) autonomic nerve supply to act postjunctionally on smooth-muscle cells. Nitric oxide increases the production of cyclic 3′,5′-guanosine monophosphate (cyclic GMP), which induces relaxation of smooth muscle **(Fig. 409-2)**. Cyclic GMP is metabolized by phosphodiesterase type 5 (PDE-5). Inhibitors of PDE-5 such as the oral medications sildenafil, tadalafil, vardenafil, and avanafil maintain erections by reducing the breakdown of cyclic GMP. However, if nitric oxide is not produced at some level, PDE-5 inhibitors are ineffective, as these drugs facilitate, but do not initiate, the initial enzyme cascade. In addition to nitric oxide, vasoactive prostaglandins (PGE_1, $PGF_{2\alpha}$) are synthesized within the cavernosal tissue and increase cyclic AMP levels, also leading to relaxation of cavernosal smooth-muscle cells.

Ejaculation is stimulated by the sympathetic nervous system; this results in contraction of the epididymis, vas deferens, seminal vesicles, and prostate, causing seminal fluid to enter the urethra. Seminal fluid emission is followed by rhythmic contractions of the bulbocavernosus and ischiocavernosus muscles, leading to ejaculation. This is followed by expulsion, characterized by stereotypic rhythmic contractions of the striated perineal muscles, leading to forceful expulsion of semen with the bladder neck closed. This emission and expulsion are controlled by the autonomic (parasympathetic and sympathetic) and somatic spinal centers, respectively. The synchronization between autonomic and somatic spinal centers is orchestrated by interneurons that form a spinal ejaculation generator that is present in mammals including man.

Premature ejaculation usually is related to anxiety or a learned behavior and is amenable to behavioral therapy or treatment with medications such as selective serotonin reuptake inhibitors (SSRIs). *Retrograde ejaculation* (RE) results when the internal urethral sphincter does not close; it may occur in men with diabetes or after surgery involving the bladder neck. *Anejaculation*, the failure of a portion or the whole of the emission process often confused with RE, is commonly the result of selective alpha blockers used in male voiding dysfunction (e.g., tamsulosin, silodosin).

Detumescence is mediated by norepinephrine from the sympathetic nerves, endothelin from the vascular surface, and smooth-muscle contraction induced by postsynaptic α-adrenergic receptors and activation of Rho kinase. These events increase venous outflow and restore the flaccid state. Venous leak can cause premature detumescence and is caused by insufficient relaxation of the corporal smooth muscle rather than a specific anatomic defect. *Priapism* refers to a persistent and painful erection and may be associated with sickle cell anemia, hypercoagulable states, spinal cord injury, or injection of vasodilator agents into the penis.

ERECTILE DYSFUNCTION

Epidemiology Erectile dysfunction (ED) is not considered a normal part of the aging process. Nonetheless, it is associated with certain physiologic and psychological changes related to age. In the Massachusetts Male Aging Study (MMAS), a community-based survey of men aged 40–70, 52% of responders reported some degree of ED. Complete ED occurred in 10% of respondents, moderate ED in 25%, and minimal ED in 17%. The incidence of moderate or severe ED more than doubled between the ages of 40 and 70. In the National Health and Social Life Survey (NHSLS), which included a sample of men and women aged 18–59, 10% of men also reported being unable to maintain an erection. Incidence was highest among men in the age group 50–59 (21%) and men who were poor (14%), divorced (14%), and less educated (13%).

The incidence of ED is also higher among men with certain medical disorders, such as diabetes mellitus, obesity, lower urinary tract symptoms secondary to benign prostatic hyperplasia (LUTS/BPH), heart disease, hypertension, decreased high-density lipoprotein (HDL) levels, and diseases associated with general systemic inflammation (e.g., rheumatoid arthritis). Cardiovascular disease and ED share etiologies as well as pathophysiology (e.g., endothelial dysfunction), and the degree of ED appears to correlate with the severity of cardiovascular disease. Consequently, ED represents a "sentinel symptom" in patients with occult cardiovascular and peripheral vascular disease.

Smoking is also a significant risk factor in the development of ED. Medications used in treating diabetes or cardiovascular disease are additional risk factors (see below). There is a higher incidence of ED among men who have undergone radiation or surgery for prostate cancer and in those with a lower spinal cord injury. Psychological causes of ED include depression, anger, stress from employment or relationships, anxiety, and other stress-related causes.

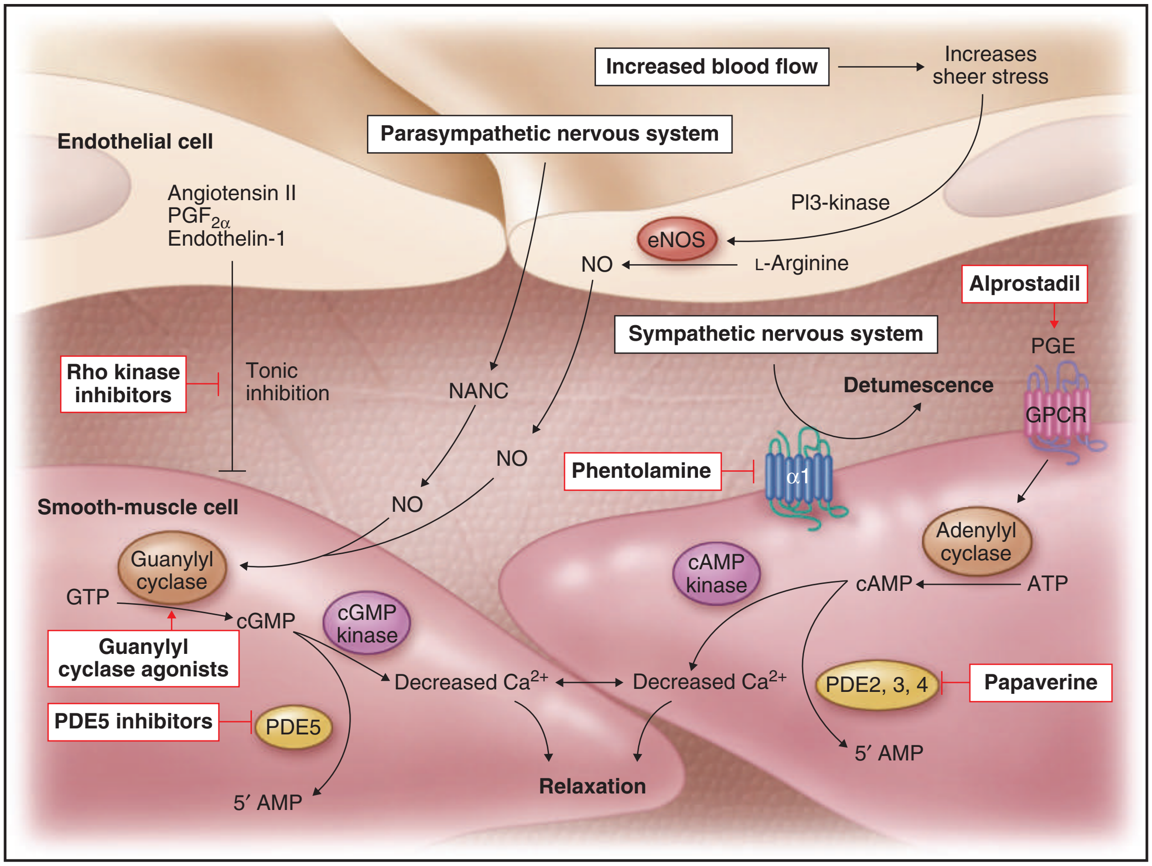
Increased blood flow
Increases sheer stress
Endothelial cell
Parasympathetic nervous system
Angiotensin II
PGF$_{2\alpha}$
Endothelin-1
PI3-kinase
eNOS
NO
L-Arginine
Alprostadil
Sympathetic nervous system
PGE
Rho kinase inhibitors
Tonic inhibition
NANC
Detumescence
GPCR
NO
Phentolamine
α1
Smooth-muscle cell
NO
Guanylyl cyclase
GTP
cGMP
Guanylyl cyclase agonists
cGMP kinase
cAMP kinase
Adenylyl cyclase
cAMP
ATP
Decreased Ca^{2+}
Decreased Ca^{2+}
PDE2, 3, 4
Papaverine
PDE5 inhibitors
PDE5
5′ AMP
Relaxation
5′ AMP

FIGURE 409-1 Pathways that control erection and detumescence. Outflow from the parasympathetic nervous system leads to relaxation of the cavernous sinusoids in two ways, both of which increase the concentration of nitric oxide (NO) in smooth-muscle cells. First, NO is the neurotransmitter in nonadrenergic, noncholinergic (NANC) fibers; second, stimulation of endothelial nitric oxide synthase (eNOS) through cholinergic output causes increased production of NO. The NO produced in the endothelium then diffuses into the smooth-muscle cells and decreases its intracellular calcium concentration through a pathway mediated by cyclic guanosine monophosphate (cGMP), leading to relaxation. A separate mechanism that decreases the intracellular calcium level is mediated by cyclic adenosine monophosphate (cAMP). With increased cavernosal blood flow, as well as increased levels of vascular endothelial growth factor (VEGF), the endothelial release of NO is further sustained through the phosphatidylinositol 3 (PI3) kinase pathway. Active treatments (red boxes) include drugs that affect the cGMP pathway (phosphodiesterase type 5 [PDE-5] inhibitors and guanylyl cyclase agonists), the cAMP pathway (alprostadil), or both pathways (papaverine), along with neural-tone mediators (phentolamine and Rho kinase inhibitors). Agents that are being developed include guanylyl cyclase agonists (to bypass the need for endogenous NO) and Rho kinase inhibitors (to inhibit tonic contraction of smooth-muscle cells mediated through endothelin). α1, α-adrenergic receptor; GPCR, G protein–coupled receptor; GTP, guanosine triphosphate; iCa^{2+}, intracellular calcium; NOS, nitric oxide synthase; PGE, prostaglandin E; PGF, prostaglandin F. *(Reproduced with permission from KT McVary: Clinical practice. Erectile dysfunction. N Engl J Med 357:2472, 2007.)*

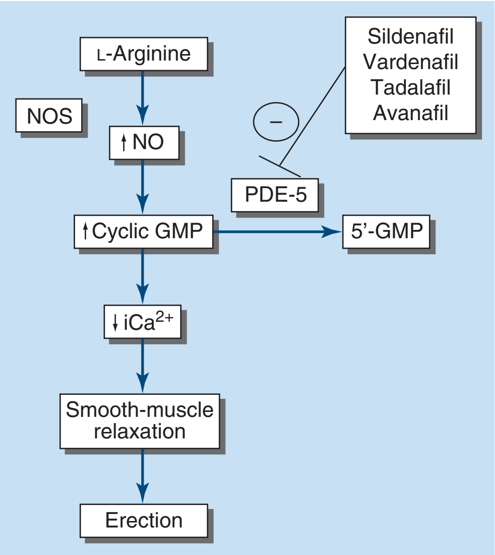
L-Arginine
NOS
↑NO
Sildenafil
Vardenafil
Tadalafil
Avanafil
–
PDE-5
↑Cyclic GMP
5'-GMP
↓iCa^{2+}
Smooth-muscle relaxation
Erection

FIGURE 409-2 Biochemical pathways modified by phosphodiesterase type 5 (PDE-5) inhibitors. Sildenafil, vardenafil, tadalafil, and avanafil enhance erectile function by inhibiting PDE-5, thereby maintaining high levels of cyclic 3′,5′-guanosine monophosphate (cyclic GMP). iCa^{2+}, intracellular calcium; NO, nitric oxide; NOS, nitric oxide synthase.

Pathophysiology ED may result from three basic mechanisms: (1) failure to initiate (psychogenic, endocrinologic, or neurogenic), (2) failure to fill (arteriogenic), and (3) failure to store adequate blood volume within the lacunar network (veno-occlusive dysfunction). These categories are not mutually exclusive, and multiple factors contribute to ED. Psychogenic factors frequently coexist with other etiologic factors and should be considered in all cases. Diabetic, atherosclerotic, and drug-related causes account for >80% of cases of ED in older men.

Vasculogenic The most common organic cause of ED is a disturbance of blood flow to and from the penis. Atherosclerotic or traumatic arterial disease can decrease flow to the lacunar spaces, resulting in decreased rigidity and an increased time to full erection. Excessive outflow through the veins despite adequate inflow also may contribute to ED. Structural alterations to the fibroelastic components of the corpora may cause a loss of compliance and inability to compress the tunical veins. This condition may result from aging, increased cross-linking of collagen fibers induced by nonenzymatic glycosylation, hypoxemia, or altered synthesis of collagen associated with hypercholesterolemia.

Neurogenic Disorders that affect the sacral spinal cord or the autonomic fibers to the penis preclude nervous system relaxation of penile smooth muscle, thus leading to ED. In patients with spinal cord injury, the degree of ED depends on the completeness and level of the lesion. Patients with incomplete lesions or injuries to the upper part of the spinal cord are more likely to retain erectile capabilities than are those with complete lesions or injuries to the lower part. Although 75% of patients with spinal cord injuries have some erectile capability, only 25% have erections sufficient for penetration. Other neurologic disorders commonly associated with ED include multiple sclerosis and peripheral neuropathy. The latter is often due to either diabetes or alcoholism. Pelvic surgery may cause ED through disruption of the autonomic nerve supply.

Endocrinologic Androgens increase libido, but their exact role in erectile function is unclear. Individuals with castrate levels of testosterone can achieve erections from visual or sexual stimuli. Nonetheless, normal levels of testosterone appear to be important for erectile function, in which the upregulation of nitric oxide synthase and the nitric oxide cascade is optimized (Fig. 409-1*A*). Androgen replacement therapy can improve depressed erectile function when it is secondary to hypogonadism; however, *it is not useful for ED when endogenous testosterone levels are normal and should be avoided.* Increased prolactin may decrease libido by suppressing gonadotropin-releasing hormone (GnRH) resulting in decreased testosterone levels. Treatment of hyperprolactinemia with dopamine agonists can restore libido and eugonadism.

Diabetic ED occurs in 35–75% of men with diabetes mellitus. Pathologic mechanisms are related primarily to diabetes-associated vascular and neurologic complications. Diabetic macrovascular complications are related mainly to age, whereas microvascular complications correlate with the duration of diabetes and the degree of glycemic control **(Chap. 415)**. Individuals with diabetes also have reduced amounts of nitric oxide synthase in both endothelial and neural tissues.

Psychogenic Two mechanisms contribute to the inhibition of erections in psychogenic ED. First, psychogenic stimuli to the sacral cord may inhibit reflexogenic responses, thereby blocking activation of vasodilator outflow to the penis. Second, excess sympathetic stimulation in an anxious man may increase penile smooth-muscle tone. The most common causes of psychogenic ED are performance anxiety, depression, relationship conflict, loss of attraction, sexual inhibition, conflicts over sexual preference, sexual abuse in childhood, and fear of pregnancy or sexually transmitted disease. Almost all patients with ED, even when it has a defined organic basis, develop a psychogenic component as a reaction to ED.

Medication-Related Medication-induced ED **(Table 409-1)** is estimated to occur in 25% of men seen in general medical clinics. The adverse effects related to drug therapy are additive, especially in older men. In addition to the drug itself, the underlying disease being treated is likely to contribute to sexual dysfunction (e.g., hypertension). Among the antihypertensive agents, the thiazide diuretics and beta blockers have been implicated most frequently. Calcium channel blockers and angiotensin-converting enzyme inhibitors are cited less frequently. These drugs may act directly at the corporal level (e.g., calcium channel blockers) or indirectly by reducing pelvic blood pressure, which is important in the development of penile rigidity. α-Adrenergic blockers are less likely to cause ED. Estrogens, GnRH agonists, H_2 antagonists, and spironolactone cause ED by suppressing gonadotropin production or by blocking androgen action. Antidepressant and antipsychotic agents—particularly neuroleptics, tricyclics, and SSRIs—are associated with erectile, ejaculatory, orgasmic, and sexual desire difficulties. Among the SSRIs, paroxetine and escitalopram have been associated with the highest risk of sexual dysfunction. Bupropion, nefazodone, and mirtazapine appear less likely to cause sexual dysfunction. A number of molecular pathways have been implicated in antidepressant-induced sexual adverse events. Serotonin has been hypothesized to inhibit normal sexual response by decreasing dopamine-enhanced libido, arousal, and erection and by increasing prolactin release. SSRIs have also been shown to be potent inhibitors of nitric oxide synthase.

If there is a strong association between the institution of a drug and the onset of ED, alternative medications should be considered. Otherwise, it is often practical to treat the ED without attempting multiple changes in medications as it may be difficult to establish a causal role for a drug.

TABLE 409-1 Drugs Associated with Erectile Dysfunction

CLASSIFICATION	DRUGS	POSSIBLE SUBSTITUTES
Diuretics	Thiazides	
	Spironolactone	
Antihypertensives	Calcium channel blockers	α-Adrenergic blockers Prazosin Terazosin Doxazosin ACE inhibitors
	Methyldopa	
	Clonidine	
	Reserpine	
	Beta blockers	
	Guanethidine	
Cardiac/antihyperlipidemics	Digoxin	
	Gemfibrozil	
	Clofibrate	
Antidepressants	Selective serotonin reuptake inhibitors	Bupropion Nefazodone Mirtazapine
	Tricyclic antidepressants	
	Lithium	
	Monoamine oxidase inhibitors	
Tranquilizers	Butyrophenones	
	Phenothiazines	
H_2 antagonists	Ranitidine	Proton pump inhibitors (PPI) Omeprazole Esomeprazole Pantoprazole Rabeprazole
	Cimetidine	
Hormones	Progesterone	
	Estrogens	
	Corticosteroids	
	GnRH agonists	
	5α-Reductase inhibitors	
	Cyproterone acetate	
Cytotoxic agents	Cyclophosphamide	
	Methotrexate	
	Roferon-A	
Anticholinergics	Disopyramide	
	Anticonvulsants	
Recreational	Ethanol	
	Cocaine	
	Marijuana	

Abbreviations: ACE, angiotensin-converting enzyme; GnRH, gonadotropin-releasing hormone.

APPROACH TO THE PATIENT

Erectile Dysfunction

A good physician–patient relationship helps unravel the possible causes of ED, many of which require discussion of personal and sensitive topics. For this reason, a primary care provider is often ideally suited to initiate the evaluation. However, a significant percentage of men experience ED and remain undiagnosed unless specifically questioned about this issue. By far the two most common reasons for underreporting of ED are patient embarrassment and perceptions of physicians' inattention to the disorder. Once

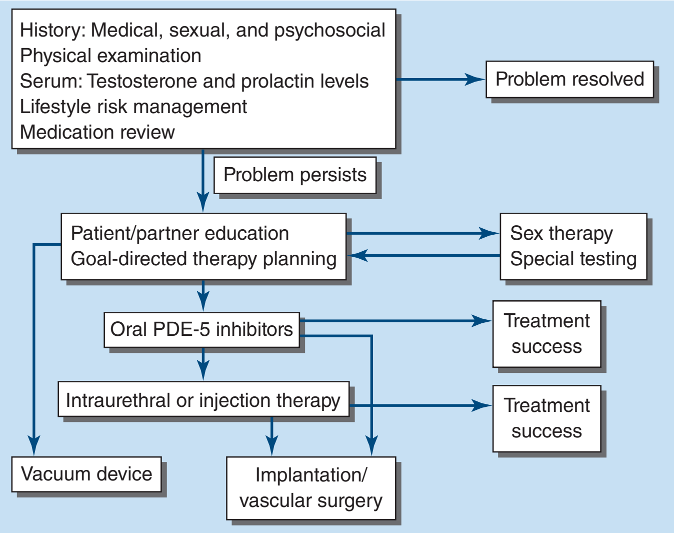
History: Medical, sexual, and psychosocial
Physical examination
Serum: Testosterone and prolactin levels
Lifestyle risk management
Medication review
Problem resolved
Problem persists
Patient/partner education
Goal-directed therapy planning
Sex therapy
Special testing
Oral PDE-5 inhibitors
Treatment success
Intraurethral or injection therapy
Treatment success
Vacuum device
Implantation/ vascular surgery

FIGURE 409-3 Algorithm for the evaluation and management of patients with erectile dysfunction. PDE, phosphodiesterase.

the topic is initiated by the physician, patients are more willing to discuss their potency issues. A complete medical and sexual history should be taken in an effort to assess whether the cause of ED is organic, psychogenic, or multifactorial **(Fig. 409-3)**.

Both the patient and his sexual partner should be interviewed regarding sexual history. ED should be distinguished from other sexual problems, such as premature ejaculation. Lifestyle factors such as sexual orientation, the patient's distress from ED, performance anxiety, and details of sexual techniques should be addressed. Validated questionnaires are available to assess ED, including the International Index of Erectile Function (IIEF) and the more easily administered Sexual Health Inventory for Men (SHIM), a validated abridged version of the IIEF. These can assess the severity of ED, measure treatment effectiveness, and guide future management. The initial evaluation of ED begins with a review of the patient's medical, surgical, sexual, and psychosocial histories. The history should note whether the patient has experienced pelvic trauma, surgery, or radiation. In light of the increasing recognition of the relationship between lower urinary tract symptoms (LUTS/BPH) and ED, it is advisable to evaluate for the presence of associated urinary symptoms. Questions should focus on the onset of symptoms, the presence and duration of partial erections, and the progression of ED. A history of nocturnal or early morning erections may be useful for distinguishing physiologic ED from psychogenic ED. Nocturnal erections occur during rapid eye movement (REM) sleep and require intact neurologic and circulatory systems. Organic causes of ED generally are characterized by a gradual and persistent change in rigidity or the inability to sustain nocturnal, coital, or self-stimulated erections. It is also important to address libido, as decreased sexual drive and ED are sometimes the earliest signs of decreased testosterone levels. It is useful to ask whether the problem is confined to coitus with one partner or also involves other partners; ED not uncommonly arises with new or extramarital sexual relationships. Situational ED, as opposed to consistent ED, suggests psychogenic causes. For men with recalcitrant ED, referral to a mental health professional may promote treatment adherence, reduce performance anxiety, and integrate treatments into a sexual relationship. Ejaculation is much less commonly affected than erection, but questions should be asked about whether ejaculation is normal, premature, delayed, or absent. Relevant risk factors should be identified, such as diabetes mellitus, cardiovascular disease, and neurologic disorders. The patient's surgical history should be explored with an emphasis on bowel, bladder, prostate, and vascular procedures. A complete drug history, including tobacco, alcohol, marijuana, and illicit drug inquiries, is also important. Social changes that may precipitate ED are also crucial to the evaluation, including health worries, spousal death, divorce, relationship difficulties, and financial concerns.

Because ED commonly involves a host of endothelial cell risk factors, men with ED report higher rates of overt and silent myocardial infarction. Therefore, ED in an otherwise asymptomatic male warrants consideration of other vascular disorders, including coronary disease. It is now widely recognized that ED often precedes the development of symptomatic coronary disease by several years. In addition, several studies now suggest that ED itself is an independent risk factor for the development of CVD even after correcting for other cardiovascular risk factors.

Men who suffer from ED are at high risk for concomitant LUTS from BPH and vice versa. Given that some treatments of one disorder will impact the other, the clinician should consider an assessment of LUTS in any man with ED.

The physical examination is an essential element in the assessment of ED. Signs of hypertension as well as evidence of thyroid, hepatic, hematologic, cardiovascular, or renal diseases should be sought. An assessment should be made of the endocrine and vascular systems, the external genitalia, and the prostate gland. The penis should be palpated carefully along the corpora to detect fibrotic plaques. Reduced testicular size and loss of secondary sexual characteristics are suggestive of hypogonadism. Neurologic examination should include assessment of anal sphincter tone, investigation of the bulbocavernosus reflex, and testing for peripheral neuropathy.

Although hyperprolactinemia is uncommon, a serum prolactin level should be measured in hypogonadal men, as decreased libido and/or ED may be the presenting symptoms of a prolactinoma or another mass lesion of the sella **(Chap. 392)**. The serum testosterone level should be measured, and if it is low, gonadotropins should be measured to determine whether hypogonadism is primary (testicular) or secondary (hypothalamic-pituitary) in origin **(Chap. 403)**. If not performed recently, serum chemistries, hemoglobin A_{1c}, and lipid profiles may be of value, as they can yield evidence of diabetes, hyperlipidemia, or other systemic diseases associated with ED.

Additional diagnostic testing is rarely necessary in the evaluation of ED. However, in selected patients, specialized testing may provide insight into pathologic mechanisms of ED and aid in the selection of treatment options. Optional specialized testing includes (1) studies of nocturnal penile tumescence and rigidity, (2) vascular testing (in-office injection of vasoactive substances, penile Doppler ultrasound), (3) neurologic testing (biothesiometry-graded vibratory perception, somatosensory-evoked potentials), and (4) psychological diagnostic tests. The information potentially gained from these procedures must be balanced against their invasiveness, cost, and impact on ultimate treatment outcome.

Clinicians should counsel men with ED who have comorbidities known to negatively affect erectile function that lifestyle modifications, including changes in diet and increased physical activity, improve overall health and can improve ED.

TREATMENT

Male Sexual Dysfunction

PATIENT EDUCATION

Patient and partner education is essential in the treatment of ED. In goal-directed therapy, education facilitates understanding of the disease, the results of the tests, and the selection of treatment. Discussion of treatment options helps clarify how treatment is best offered and stratify first- and second-line therapies. Patients with high-risk lifestyle issues such as obesity, smoking, alcohol abuse, and recreational drug use should be counseled on the role those factors play in the development of ED.

Therapies currently employed for the treatment of ED include oral PDE-5 inhibitor (PDE-5i) therapy (most commonly used), injection therapies, testosterone therapy, penile devices, and

TABLE 409-2 PDE-5 Inhibitors[a]

DRUG	ONSET OF ACTION	$T_{1/2}$	DOSE	ADVERSE EFFECTS	CONTRAINDICATIONS
Sildenafil	T_{max} 30–120 min Duration 4 h High-fat meal decreases absorption Alcohol use may affect efficacy	2–5 h	25–100 mg Starting dose 50 mg	Headache, flushing, dyspepsia, nasal congestion, altered vision	Nitrates Hypotension Cardiovascular risk factors Retinitis pigmentosa Change dose with some antiretrovirals Should be on stable dose of alpha blockers
Vardenafil	T_{max} 30–120 min Duration 4–5 h High-fat meal decreases absorption ETOH may affect efficacy	4.5 h	5–10 mg	Headache, flushing, rhinitis, dyspepsia	Same as sildenafil May have minor prolongation of QT interval Concomitant use of class I antiarrhythmic
Tadalafil	T_{max} 30–60 min Duration 12–36 h Plasma concentration not affected by food or ETOH	17.5 h	10 or 20 mg; 2.5 or 5 mg for daily dose	Headache, dyspepsia, backpain, nasal congestion, myalgia	Same as sildenafil
Avanafil	T_{max} 30 min Duration 2 h Plasma concentration not affected by food	3–5 h	50, 100, and 200 mg dose	Headache, flushing, nasal congestion nasopharyngitis back pain	Same as sildenafil

[a] Sildenafil, vardenafil, tadalafil, and the newest option, avanafil, appear to be equally effective, but tadalafil has a longer duration of action and avanafil has a more rapid onset.

Abbreviations: ETOH, ethanol; PDE-5, phosphodiesterase type 5.

psychological therapy. In addition, limited data suggest that treatments for underlying risk factors and comorbidities—for example, weight loss, exercise, stress reduction, and smoking cessation—may improve erectile function.

ORAL AGENTS

Sildenafil, tadalafil, vardenafil, and avanafil are the only approved and effective oral agents for the treatment of ED. These four medications have markedly improved the management of ED because they are effective for the treatment of a broad range of etiologies. They belong to a class of medications that are selective and potent inhibitors of PDE-5, the predominant phosphodiesterase isoform found in the penis. They are administered in graduated doses and enhance erections after sexual stimulation (Fig. 409-2). The onset of action is ~30–120 min, depending on the medication used and other factors, such as recent food intake. Reduced initial doses should be considered for patients who are elderly, are taking concomitant alpha blockers, have renal insufficiency, or are taking medications that inhibit the CYP3A4 metabolic pathway in the liver (e.g., erythromycin, cimetidine, ketoconazole, clarithromycin, diltiazem, itraconazole, ritonavir, verapamil, grapefruit, and possibly itraconazole and mibefradil), as they may increase the serum concentration of the PDE-5i or promote hypotension. Initially, there were concerns about the cardiovascular safety of these drugs. It is known that these agents can act as mild vasodilators. Earlier concerns that the use of PDE-5is would increase cardiovascular events have been reversed as more studies support that a general population of men with ED who take PDE-5is have lower rates of major adverse cardiovascular events and lower overall mortality rates than men not exposed to PDE-5is, even after correcting for baseline cardiovascular risks and concomitant medicines. These studies suggest that PDE-5is may have a significant cardiovascular protective effect and may have potential as preventive cardiology agents.

Several randomized trials have demonstrated the efficacy of this class of medications. There are no compelling data to support the superiority of one PDE-5i over another. Subtle differences between agents have variable clinical relevance **(Table 409-2)**.

Patients may fail to respond to a PDE-5i for several reasons **(Table 409-3)**. Some patients may not tolerate PDE-5i secondary to adverse events from vasodilation in nonpenile tissues expressing PDE-5 or from the inhibition of homologous nonpenile isozymes (i.e., PDE-6 found in the retina). Abnormal vision attributed to the effects of PDE-5i on retinal PDE-6 is of short duration, reported only with sildenafil, and not clinically significant. A more serious concern is the possibility that PDE-5is may cause nonarteritic anterior ischemic optic neuropathy (NAION); although data to support that association are limited, it is prudent to avoid the use of these agents in men with a prior history of NAION.

Testosterone supplementation combined with a PDE-5i may be beneficial in improving erectile function in hypogonadal men with ED who are unresponsive to PDE-5i alone. Side effects associated with PDE-5is include headaches (19%), facial flushing (9%), dyspepsia (6%), and nasal congestion (4%). Approximately 7% of men using sildenafil may experience transient altered color vision (blue halo effect), and 6% of men taking tadalafil may experience loin pain. PDE-5i is contraindicated in men receiving nitrate therapy for cardiovascular disease, including agents delivered by the oral, sublingual, transnasal, and topical routes as they can potentiate its hypotensive effect. Likewise, amyl/butyl nitrate “poppers” may have a fatal synergistic effect on blood pressure. PDE-5is also should be avoided in patients with congestive heart failure and cardiomyopathy because of the risk of vascular collapse. Because sexual activity leads to an increase in physiologic expenditure (5–6 metabolic equivalent tasks [METs]), physicians have been advised to exercise caution in prescribing any drug for sexual activity to those with active coronary disease, heart failure, borderline hypotension, or hypovolemia and to those on complex antihypertensive regimens.

Although the various forms of PDE-5is have a common mechanism of action, there are a few differences among the four agents (Table 409-2). Tadalafil is unique in its longer half-life, and avanafil appears to have the fastest onset of action. Although there are pharmacokinetic and pharmacodynamic differences among these agents, clinically relevant differences are not clear.

TABLE 409-3 Issues to Consider if Patients Report Failure of Phosphodiesterase Type 5 Inhibitor (PDE-5i) to Improve Erectile Dysfunction

1. A trial of medication on at least 6 different days at the maximal dose should be performed before declaring patient nonresponsive to PDE-5i use.
2. Confirm that the patient did not partake in a high-fat meal prior to taking medication; pertains to sildenafil.
3. Failure to include physical and psychic stimulation at the time of foreplay to induce endogenous NO.
4. Took medications at an appropriate time frame *prior* to step 3: half-hour prior for avanafil, 1 h for sildenafil/vardenafil, or 2.5 h for tadalafil.
5. Unrecognized hypogonadism.

Abbreviation: NO, nitric oxide.

ANDROGEN THERAPY

Testosterone replacement is used to treat both primary and secondary causes of hypogonadism **(Chap. 403)**. Men with ED and testosterone deficiency (TD) who are considering ED treatment with a PDE-5i should be informed that PDE-5is may be more effective if combined with testosterone therapy. Androgen supplementation in the setting of normal testosterone is not efficacious in the treatment of ED and is discouraged secondary to additional risk for toxicity without benefit.

The increased scrutiny of testosterone caused the U.S. Food and Drug Administration (FDA) to issue a warning that there is a "weak signal" that testosterone replacement therapy increases the risk of thromboembolic events and may have addictive properties. Although testosterone therapy has known risks, such as water retention in heart failure patients and worsening sleep apnea, increasing evidence suggests that, when monitored appropriately, this therapy decreases the risk for metabolic syndrome, changes body composition by increasing lean muscle mass, and improves insulin sensitivity and average hemoglobin A_{1c}. This evidence, combined with the fact that hypogonadism is a known risk factor for metabolic syndrome and cardiovascular disease, has led to the conclusion that testosterone therapy for age-related hypogonadism in fact improves overall health and decreases the risk of cardiovascular events. It is important to note that men with secondary hypogonadism who desire fertility should not be treated directly with testosterone, but with an alternative such as the selective estrogen receptor modulator (SERM) clomiphene citrate, which increases gonadotropin levels, stimulating testicular testosterone production.

Testosterone circulates in the body in two forms: free and unbound or that bound to proteins such as albumin or sex hormone–binding globulin (SHBG). SHBG has a very high affinity for testosterone, and thus, testosterone bound to SHBG does not bind to the androgen receptor and is not bioavailable. Bioavailable testosterone is any testosterone that is not bound to SHBG. Unfortunately, reliable assays to directly measure bioavailable testosterone or free testosterone are expensive, difficult to perform, and thus not offered by most laboratories. However, direct measurement of SHBG is inexpensive and reliable, allowing free and bioavailable testosterone to be calculated.

Men who receive testosterone should be reevaluated after 3–6 months and at least annually thereafter for testosterone levels, erectile function, and adverse effects, which may include gynecomastia, sleep apnea, development or exacerbation of LUTS or BPH, prostate cancer, lowering of HDL, erythrocytosis, and elevations of liver function tests. Periodic reevaluation should include measurement of hemoglobin, liver function tests, prostate-specific antigen, and digital rectal examination. Therapy should be discontinued in patients who do not respond within 6 months without an alternate explanation (e.g., elevated estradiol).

VACUUM CONSTRICTION DEVICES

Vacuum constriction devices (VCDs) are a well-established noninvasive therapy. They are a reasonable treatment alternative for select patients who cannot take PDE-5is or do not desire other interventions. VCDs draw venous blood into the penis and use a constriction ring to restrict venous return and maintain tumescence. Adverse events with VCD include pain, numbness, bruising, and altered ejaculation. Additionally, many patients complain that the devices are cumbersome and that the induced erections have a nonphysiologic appearance and feel.

INTRAURETHRAL ALPROSTADIL

If a patient fails to respond to oral agents, a reasonable next choice is intraurethral or self-injection of vasoactive substances. Intraurethral prostaglandin E_1 (alprostadil), in the form of a semisolid pellet (doses of 125–1000 μg), is delivered with an applicator. Approximately 65% of men receiving intraurethral alprostadil respond with an erection when tested in the office, but <50% achieve successful coitus at home.

INTRACAVERNOSAL SELF-INJECTION

Injection of synthetic formulations of alprostadil is effective in 70–80% of patients with ED, but discontinuation rates are high because of the invasive nature of administration. Doses range between 1 and 40 μg. Injection therapy is contraindicated in men with a history of hypersensitivity to the drug and men at risk for priapism (hypercoagulable states, sickle cell disease). Side effects include local adverse events, prolonged erections, pain, and fibrosis with chronic use. Various combinations of alprostadil, phentolamine, and/or papaverine sometimes are used.

SURGERY

An important but less frequently used form of therapy for ED involves the surgical implantation of a semirigid or inflatable penile prosthesis. Because of the permanence of prosthetic devices, patients should first consider less invasive options for treatment. These surgical treatments are associated with a low rate of complications and are used for those who do not want the less spontaneous medical treatments, in PDE-5i–refractory ED, or in men who cannot tolerate such medications. Despite the requirement for surgery, penile prostheses are associated with very high rates of patient and partner satisfaction.

SEX THERAPY

A course of sex therapy may be useful for addressing specific interpersonal factors that may affect sexual functioning. These approaches may be useful in patients who have psychogenic or social components to their ED, although data from randomized trials are inconsistent. It is preferable to include both partners in therapy if the patient is involved in an ongoing relationship.

FEMALE SEXUAL DYSFUNCTION

Female sexual dysfunction (FSD) includes chronic sexual conditions in the domains of desire, arousal, pain, and muted orgasm. The associated risk factors for FSD are similar to those in males: cardiovascular disease, endocrine disorders, hypertension, neurologic disorders, and smoking **(Table 409-4)**. Women with hypertension report significantly lower sexual satisfaction (especially younger women).

EPIDEMIOLOGY

Epidemiologic data are limited, but the available estimates suggest that as many as 43% of women complain of at least one sexual problem. Despite their frequency and impact, FSDs are substantially undetected by clinicians and undertreated even when recognized. Despite the recent interest in organic causes of FSD, desire and arousal phase disorders (including lubrication complaints) remain the most common presenting problems when surveyed in a community-based population.

TABLE 409-4 Risk Factors for Female Sexual Dysfunction

Neurologic disease: stroke, spinal cord injury, parkinsonism
Trauma, genital surgery, radiation
Endocrinopathies: diabetes, hyperprolactinemia
Liver and/or renal failure
Cardiovascular disease, especially hypertension
Psychological factors and interpersonal relationship disorders: sexual abuse, life stressors
Medications
Antiandrogens: cimetidine, spironolactone
Antidepressants, alcohol, hypnotics, sedatives
Antiestrogens or GnRH antagonists
Antihistamines, sympathomimetic amines
Antihypertensives: diuretics, calcium channel blockers
Alkylating agents
Anticholinergics

Abbreviation: GnRH, gonadotropin-releasing hormone.

PHYSIOLOGY OF THE FEMALE SEXUAL RESPONSE

The normal female sexual response requires the presence of estrogens. A role for androgens is also likely but less well established. In the CNS, estrogens and androgens work synergistically to enhance sexual arousal and response. A number of studies report enhanced libido in women during preovulatory phases of the menstrual cycle, suggesting that hormones involved in the ovulatory surge (e.g., estrogens) increase desire.

Sexual motivation is heavily influenced by context, including the environment and partner factors. Once sufficient sexual desire is reached, sexual arousal is mediated by the central and autonomic nervous systems. Cerebral sympathetic outflow is thought to increase desire, and peripheral parasympathetic activity results in clitoral vasocongestion and vaginal secretion (lubrication).

The neurotransmitters for clitoral corporal engorgement are similar to those in the male penile tissues, with a prominent role for neural, smooth-muscle, and endothelial released nitric oxide (NO). A fine network of vaginal nerves and arterioles promotes a vaginal transudate. The major transmitters of this complex vaginal response are not certain, but roles for NO and vasoactive intestinal polypeptide (VIP) are supported. There are doubts concerning the construct of a linear relationship between initial desire, arousal, vasocongestion, lubrication, and orgasm. Caregivers should consider a paradigm of a positive emotional and physical outcome with one, many, or no orgasmic peak and release.

Although there are anatomic differences as well as variation in the density of vascular and neural beds in males and females, the primary effectors of sexual response are strikingly similar. Intact sensation is important for arousal. Thus, reduced levels of sexual functioning are more common in women with peripheral neuropathies (e.g., diabetes). Vaginal lubrication is a transudate of serum that results from the increased pelvic blood flow associated with arousal. Vascular insufficiency from a variety of causes may compromise adequate lubrication and result in dyspareunia. Cavernosal and arteriole smooth-muscle relaxation occurs via increased NO synthase (NOS) activity and produces engorgement in the clitoris and the surrounding vestibule. Orgasm requires an intact sympathetic outflow tract; hence, orgasmic disorders are common in female patients with spinal cord injuries.

APPROACH TO THE PATIENT

Female Sexual Dysfunction

Many women do not volunteer information about their sexual response. Open-ended questions in a supportive atmosphere are helpful in initiating a discussion of sexual integrity. Once a complaint has been voiced, a comprehensive evaluation should be performed, including a medical history, a psychosocial history, a physical examination, and limited laboratory testing.

The history should include the usual medical, surgical, obstetric, psychological, gynecologic, sexual, and social information. Past experiences, intimacy, knowledge, and partner availability should also be ascertained. Medical disorders that may affect sexual health should be delineated. They include diabetes, cardiovascular disease, gynecologic conditions, obstetric history, depression, anxiety disorders, and neurologic disease. Medications should be reviewed as they may affect arousal, libido, and orgasm. The need for counseling and recognizing life stresses should be identified. The physical examination should assess the genitalia, including the clitoris. Pelvic floor examination may identify prolapse or other disorders. Laboratory studies are needed, especially if menopausal status is uncertain. Estradiol, follicle-stimulating hormone (FSH), and luteinizing hormone (LH) are usually obtained, and dehydroepiandrosterone (DHEA) should be considered as it reflects adrenal androgen secretion. A complete blood count, liver function assessment, and lipid studies may be useful, if not otherwise obtained. Complicated diagnostic evaluation such as clitoral Doppler ultrasonography and biothesiometry require expensive equipment and are of uncertain utility. It is important for the patient to identify which symptoms are most distressing.

The evaluation of FSD previously occurred exclusively in a psychosocial context. However, inconsistencies between diagnostic categories based only on psychosocial considerations and the emerging recognition of organic etiologies have led to a new classification of FSD. This diagnostic scheme is based on four components that are not mutually exclusive: (1) *hypoactive sexual desire*—the persistent or recurrent lack of sexual thoughts and/or receptivity to sexual activity; hypoactive sexual desire may result from endocrine failure or may be associated with psychological or emotional disorders; (2) *sexual interest arousal disorder*—the persistent or recurrent inability to attain or maintain sexual excitement; (3) *orgasmic disorder*—the persistent or recurrent loss of orgasmic potential after sufficient sexual stimulation and arousal; and (4) *sexual pain disorder*—persistent or recurrent genital pain associated with noncoital sexual stimulation. This newer classification emphasizes "personal distress" as a requirement for dysfunction and provides clinicians with an organized framework for evaluation before or in conjunction with more traditional counseling methods.

TREATMENT

Female Sexual Dysfunction

GENERAL

An open discussion with the patient is important as couples may need to be educated about normal anatomy and physiologic responses, including the role of orgasm, in sexual encounters. Physiologic changes associated with aging and/or disease should be explained. Couples may need to be reminded that clitoral stimulation rather than coital intromission may be more beneficial.

Behavioral modification and nonpharmacologic therapies should be a first step. Patient and partner counseling may improve communication and relationship strains. Lifestyle changes involving known risk factors can be an important part of the treatment process. Emphasis on maximizing physical health and avoiding lifestyles (e.g., smoking, alcohol abuse) and medications likely to produce FSD is important (Table 409-3). The use of topical lubricants may address complaints of dyspareunia and dryness. Contributing medications such as antidepressants may need to be altered, including the use of medications with less impact on sexual function, dose reduction, medication switching, or drug holidays.

HORMONAL THERAPY

In postmenopausal women, estrogen replacement therapy may be helpful in treating vaginal atrophy, decreasing coital pain, and improving clitoral sensitivity **(Chap. 407)**. Menopause and its transition represent significant risk factors for the development of vulvovaginal atrophy–related sexual dysfunction. Available vaginal estrogen preparations include conjugated equine estrogens, estradiol vaginal cream, a sustained-release intravaginal estradiol ring, or a low-dose estradiol tablet. Vaginal estrogen preparations with the lowest systemic absorption rate may be preferred in women with history of breast cancer and severe vaginal atrophy. Vaginal lubricants and moisturizers applied on a regular basis have an efficacy comparable to that of local estrogen therapy and should be offered to women wishing to avoid the use of vaginal estrogens. If a hormonal supplement is chosen, then estrogen replacement in the form of local cream is the preferred method as it avoids systemic side effects. Androgen levels in women decline substantially before menopause. However, low levels of testosterone or DHEA are not effective predictors of a positive therapeutic outcome with androgen therapy. The widespread use of exogenous androgens is not supported by the literature except in select circumstances (premature ovarian failure or menopausal states) and in secondary arousal disorders.

Atrophic vaginitis is very common in postmenopausal women and is most commonly treated with estrogen-based treatments. However, many women are hesitant to use estrogen-based treatments due to health concerns or are unable to use them due to a history of breast cancer or endometrial cancer. Hyaluronic acid vaginal gel has been found to be efficacious in treating atrophic vaginitis.

ORAL AGENTS

Flibanserin, originally developed as an antidepressant, is approved by the FDA as a treatment for low sexual desire in premenopausal women. Flibanserin, a postsynaptic agonist of serotonin receptor 1A and antagonist of serotonin receptor 2A, increases sexual desire and reduces resultant stress in women with hyposexual desire disorder (HSDD) with few adverse effects. Flibanserin has two principal pharmacologic actions in neural microcircuits: it acts as a full agonist at postsynaptic 5-HT_{1A} receptors and an antagonist at postsynaptic 5-HT_{2A} receptors. Exclusive binding at these receptors differentiates flibanserin from buspirone and bupropion. This action in the prefrontal cortex causes the downstream release of dopamine and norepinephrine and reduction of serotonin. Flibanserin acts selectively on pyramidal neurons that excite brainstem 5-HT neurons yet also selectively on pyramidal neurons that inhibit brainstem norepinephrine and dopamine neurons.

Flibanserin may boost sex drive in women who experience low sexual desire and who find the experience distressing. The drug should be discontinued if there is no improvement in sex drive after 8 weeks. Potentially serious side effects include low blood pressure and dizziness, particularly if it is mixed with alcohol. Other common adverse events include nausea, fatigue, sleepiness, and insomnia. Health care professionals and pharmacies dealing with flibanserin have to undergo a certification (risk evaluation and mitigation strategy [REMS]) process, and patients need to submit a written agreement to abstain from alcohol. The goal of the flibanserin REMS is to inform patients about the increased risk of hypotension and syncope due to an interaction with alcohol.

Bremelanotide, a melanocortin 4 receptor agonist, is also approved for HSDD. It demonstrates significant improvement in desire and a significant decrease in distress related to lack of desire. The most common adverse effects include nausea (40%), facial flushing (20%), and headache (10%). Bremelanotide's place in therapy is unclear, as the trials met statistical significance for change in sexual desire elements and distress related to sexual desire, yet the clinical benefit may only be modest. It is a subcutaneous injection given 45 min prior to sexual activity. Bremelanotide has no clinically significant interactions with ethanol. Prescribing guidelines recommend no more than one dose in 24 h and no more than eight doses per month. Individuals should discontinue use after 8 weeks without benefit.

There is no role for PDE-5is in FDS and should be discouraged.

CLITORAL VACUUM DEVICE

In patients with arousal and orgasmic difficulties, the option of using a clitoral vacuum device may be explored. This handheld battery-operated device has a small soft plastic cup that applies a vacuum over the stimulated clitoris. This causes increased cavernosal blood flow, engorgement, and vaginal lubrication.

FURTHER READING

Burnett AL et al: Erectile dysfunction: AUA Guideline. J Urol 200:633, 2018.

Geerkens MJM et al: Sexual dysfunction and bother due to erectile dysfunction in the healthy elderly male population: Prevalence from a systematic review. Eur Urol Focus 6:776, 2020.

Kloner RA et al: Effect of phosphodiesterase type 5 inhibitors on major adverse cardiovascular events and overall mortality in a large nationwide cohort of men with erectile dysfunction and cardiovascular risk factors: A retrospective, observational study based on healthcare claims and national death index data. J Sex Med 20:38, 2023.

McVary KT: Clinical practice. Erectile dysfunction. N Engl J Med 357:2472, 2007.

Mulhall JP et al: Evaluation and management of testosterone deficiency: AUA guideline. J Urol 200:423, 2018.

Nappi RE et al: Medical treatment of female sexual dysfunction. Urol Clin North Am 49:299, 2022.

Zhao B et al: Erectile dysfunction predicts cardiovascular events as an independent risk factor: A systematic review and meta-analysis. J Sex Med 16:1005, 2019.

410

Women's Health

Emily Nosova, Andrea Dunaif

The clinical discipline of women's health is well established. Indeed, its emphasis on greater attention to patient education and medical decision-making is a paradigm for what has become known as patient-centered health care. Moreover, the recognition of sex differences in gene expression, disease processes, and health outcomes is an important example of precision medicine. Sex difference refers to the biologic differences conferred by sex chromosomes and hormones. In contrast, gender differences are related to psychosocial roles and cultural expectations. The study of sex differences continues to grow as a scientific discipline. In 2016, the National Institutes of Health recognized its importance by implementing the expectation that sex should be considered as a biologic variable in study designs, analyses, and reporting in not only human but also vertebrate animal research. Strong scientific justification must be provided to limit research to only one sex.

DISEASE RISK: REALITY AND PERCEPTION

The leading causes of death are the same in women and men: (1) heart disease and (2) cancer **(Fig. 410-1)**. In 2020, COVID-19 emerged as the third leading cause of death, representing >10% of all deaths that year. Mortality rates due to COVID-19 were slightly lower in women (9.8%) than in men (10.9%). The leading cause of cancer death, lung cancer, is the same in both sexes. Breast cancer is the second leading cause of cancer death in women. Men are more likely than women to die from suicide and accidents. Maternal mortality continues to be higher in the United States than in other industrialized nations and is associated with substantial health disparities in maternal deaths. U.S. maternal mortality rates declined for the majority of the twentieth century given improvements in maternity care and safer surgical techniques; however, the rates began to rise again in 2000. Over the past decade, the mortality rate has remained relatively stable with the exception of a slight rise in 2021–2022, followed by a decline to typical rates again in 2023.

In June 2022, an historic Supreme Court decision overturned *Roe v. Wade* after nearly 50 years, declaring that women no longer had the constitutional right to abortion in the United States. In the year that followed this landmark ruling, more than a dozen states banned abortion, while several enacted abortion protection laws, and at least 70 clinics ceased offering abortion-related services. These changes were projected to affect nearly 30% of women of reproductive age in the United States. A 2022 study found that maternal mortality in states that restricted abortion was 62% higher compared to that of states with more widespread access (28.8 compared to 17.8 per 100,000 live births). Further longitudinal observation is needed to assess the impact of changes in abortion access on maternal and fetal mortality and outcomes.

Women's risk for many diseases increases at menopause. The median age of menopause in Caucasian women from industrialized countries is between 50 and 52 years, where women spend one-third of their lives in the postmenopausal period. Menopause occurs at earlier ages in Hispanic and African-American women as well as in women of lower socioeconomic status. Estrogen levels fall abruptly at menopause,

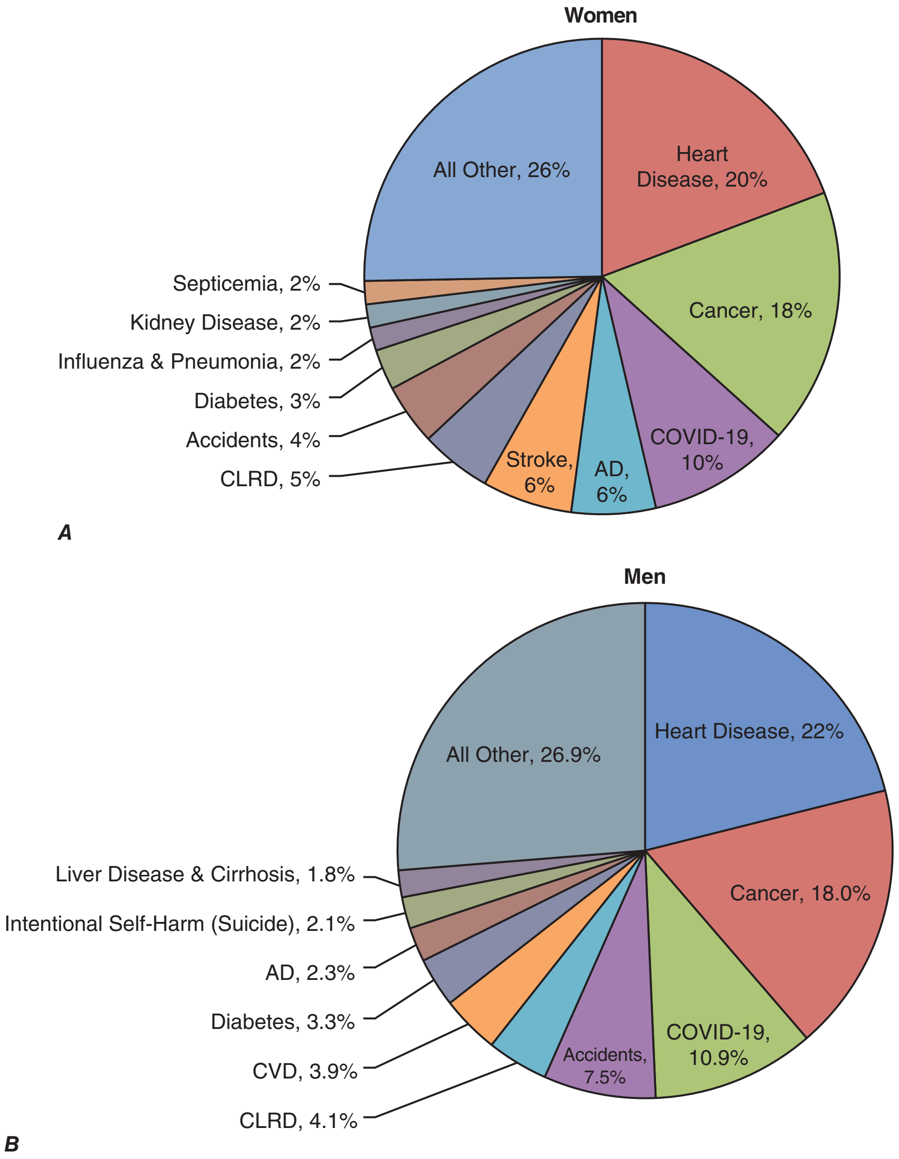

FIGURE 410-1 Percent distribution of 10 leading causes of death in (*A*) women compared to (*B*) men in the United States in 2020. In both women and men, the first, second, and third leading causes of death are the same: heart disease, cancer, and COVID-19, respectively. Causes of death then diverge by sex. Chronic lower respiratory disease (CLRD), stroke, and Alzheimer's disease (AD) cause a larger percentage of deaths in women than in men. Suicide is among the 10 leading causes of death in men but not in women. CVD, cardiovascular disease. *(Data from https://www.cdc.gov/nchs/data/nvsr/nvsr72/nvsr72-14.pdf.)*

inducing a variety of physiologic and metabolic responses. Rates of cardiovascular disease (CVD) increase, and bone density decreases rapidly after menopause.

In the United States, women live on average 5.9 years longer than men, with a life expectancy at birth in 2021 of 79.1 years in women compared with 73.2 years in men of all races. Notably, these estimates represent a decrease in life expectancy from prior years related to the COVID-19 pandemic. Life expectancy is lower in African Americans of both sexes and higher in Hispanics of both sexes than their Caucasian counterparts. Accordingly, elderly women outnumber elderly men, so that age-related conditions, such as hypertension, have a female preponderance.

SEX DIFFERENCES IN HEALTH AND DISEASE

■ ALZHEIMER'S DISEASE

(See also Chap. 442) Alzheimer's disease (AD) affects approximately twice as many women as men. Because the risk for AD increases with age, part of this sex difference is accounted for by the fact that women live longer than men. However, even in relatively younger groups (60–70 years of age), there is still a higher incidence of AD among women. Additional factors may contribute to the increased risk for AD in women, including sex differences in brain size, structure, and functional organization. Multimodal neuroimaging has demonstrated that certain biomarkers of the preclinical phase of AD, including a decline in neuronal mitochondrial function and impaired cerebral glucose metabolism, are evident earlier in women and are even distinguishable during the perimenopausal endocrine transition. There is emerging evidence for sex-specific differences in gene expression, not only for genes on the X and Y chromosomes but also for some autosomal genes. These genetic differences may translate into variable severity of AD, with women experiencing greater deficits in cognition. The ε4 allele of the apolipoprotein E gene (APOε4), a cholesterol carrier integral for lipid transport in the brain, is a major risk factor for AD. Recent studies show that the APOε4 genotype is strongly linked to development of sporadic AD in women. Women who carry either the APOε4 homo- or heterozygous isoform have an increased risk of progressing from healthy aging patterns to cognitive impairment or AD, whereas men who carry either isoform experience marginal impact on their memory or cognition.

Estrogens have pleiotropic genomic and nongenomic effects on the central nervous system, including neurotrophic actions in key areas involved in cognition and memory. Women with AD have lower endogenous estrogen levels than do women without AD. These observations have led to the hypothesis that estrogen is neuroprotective. The Women's Health Initiative Memory Study (WHIMS), an ancillary study in the Women's Health Initiative (WHI) in women aged ≥65 years, found significantly increased risk for both dementia and mild cognitive impairment in women receiving estrogen alone or estrogen with progestin compared to placebo. However, the Kronos Early Estrogen Prevention Study (KEEPS), a randomized clinical trial of early hormone therapy (HT) initiation after menopause that compared conjugated equine estrogen (CEE), transdermal estradiol (both estrogen arms included cyclic oral micronized progesterone), and placebo, found no adverse effect of HT on cognitive function. In summary, there is no evidence from placebo-controlled trials that HT improves cognitive function.

While studies have shown a link between female sex and AD, other neurodegenerative disorders, including Parkinson's disease (PD) and amyotrophic lateral sclerosis (ALS), exhibit a stronger association with male sex. Men are 1.5 times more likely to develop PD than women across all age groups. A possible explanation for the male predilection may be the effect of Y-chromosome exclusive gene sex-determining region Y (SRY) on nigrostriatal dopaminergic (NSDA) neurons: the SRY upregulates neuronal numbers, synthesis of dopamine, and metabolism of neurons.

■ CVD AND STROKE

(See also Chap. 284) There are major sex differences in CVD, the leading cause of death in developed countries. However, there are also major gender differences because of perceptions by both women and their health care providers that women are at lower risk for CVD. As a result of these misconceptions, women are less likely to seek medical help when they experience symptoms of CVD. Health care providers are less likely to suspect CVD, so women receive fewer interventions for modifiable risk factors as well as fewer acute interventions than do men. Women and their health care providers are also less aware that prodromal symptoms of cardiac disease differ in women compared to men. Women are less likely than men to present with chest pain and more likely to present with fatigue, shortness of breath, indigestion/nausea, and anxiety.

Sex steroids have major effects on the cardiovascular system and lipid metabolism. Estrogen increases high-density lipoprotein (HDL) and lowers low-density lipoprotein (LDL), whereas androgens have the opposite effect. Estrogen has direct vasodilatory effects on the vascular endothelium, enhances insulin sensitivity, and has antioxidant and anti-inflammatory properties. There is a striking increase in CVD after both natural and surgical menopause, suggesting that endogenous estrogens are cardioprotective. Women also have longer QT intervals on electrocardiograms, and this increases their susceptibility to certain arrhythmias.

CVD presents differently in women, who are usually 10–15 years older than their male counterparts and are more likely to have comorbidities such as hypertension, congestive heart failure, and diabetes mellitus (DM). In the Framingham study, angina was the most common initial symptom of CVD in women, whereas myocardial infarction (MI) was the most common initial presentation in men. Women more often have atypical symptoms such as fatigue, anxiety, nausea, indigestion, and upper back pain. Although awareness that heart disease is the leading cause of death in women has nearly doubled over the past 15 years, women remain less aware that its symptoms are often atypical and are less likely to contact 9-1-1 when they experience such symptoms.

A type of acute coronary syndrome (ACS) termed takotsubo syndrome, initially described in 1990 and characterized as a transient and reversible stress cardiomyopathy, disproportionately affects women. National cohort studies demonstrate that ~2% of all patients presenting to acute care centers with symptoms of ACS are diagnosed with takotsubo, and importantly, 80–90% are postmenopausal women. If stratified by sex, ~10% of patients with suspected ACS are ultimately diagnosed with takotsubo syndrome. Recurrence of this condition is also more common among women. Interestingly, morbidity and mortality rates are higher in men, with cardiogenic shock, cardiac arrest, and mortality occurring more frequently than in women. The pathophysiology that leads to takotsubo is complex: potential inciting mechanisms involve microvascular dysfunction and impaired vascular reactivity, followed by reversible abnormalities in the coronary flow reserve and microvascular resistance. One potential explanation for the striking sex difference is that postmenopausal women have age- and estrogen deficiency–related coronary vasomotor dysfunction. Estrogen improves coronary blood flow through endothelium-dependent and -independent mechanisms; however, its deficiency results in increased sympathetic drive and endothelial dysfunction. An animal study suggested that estrogen supplementation may partially attenuate an excessive cardiovascular response to stress; however, this has not been studied clinically. Furthermore, hypotheses related to estrogen do not explain why takotsubo occurs in men and also appears to be associated with higher mortality rates in men. The specific mechanisms underlying these sex differences in disease prevalence and outcomes remain unknown and require further study.

Deaths from CVD have decreased markedly in men since 1980, whereas CVD deaths only started to decrease substantially in women beginning in 2000. After 2010, death rates from CVD among both sexes stabilized and even began to increase slightly in men. Women with MI are more likely to present with cardiac arrest or cardiogenic shock, whereas men are more likely to present with ventricular tachycardia. Further, younger women with MI are more likely to die than are men of similar age. However, this mortality gap has decreased in recent years because younger women have experienced greater improvements in survival after MI than men. The improvement in survival is due largely to a reduction in comorbidities, suggesting a greater attention to modifiable risk factors in women.

Sex differences account for more variable short-term outcomes observed among women with CVD who receive therapeutic intervention, as compared to men. Women undergoing coronary artery bypass graft surgery have more advanced disease, a higher perioperative mortality rate, less relief of angina, and less graft patency; however, 5- and 10-year survival rates are similar. Women undergoing percutaneous transluminal coronary angioplasty have lower rates of initial angiographic and clinical success than men, but they also have a lower rate of re-stenosis and a better long-term outcome. Women may benefit less and have more frequent serious bleeding complications from thrombolytic therapy compared with men. Factors such as older age, more comorbid conditions, smaller body size, and more severe CVD in women at the time of events or procedures account in part for the observed sex differences.

Important risk factors for CVD in both men and women include elevated cholesterol levels, hypertension, smoking, obesity, low HDL cholesterol levels, DM, and lack of physical activity **(Fig. 410-2)**. Total triglyceride levels are an independent risk factor for CVD in women but not in men. Low HDL cholesterol and DM are more important risk factors for CVD in women than in men. Several disorders conferring increased CVD risk affect women exclusively, such as pregnancy-associated hypertension, preeclampsia, gestational DM, and polycystic ovary syndrome, or predominantly, such as rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE). Cholesterol-lowering drugs are equally effective in men and women for primary and secondary prevention of CVD. In contrast to men, randomized trials showed that aspirin was not effective in the primary prevention of CVD in women; it did significantly reduce the risk of ischemic stroke.

Psychosocial stressors and loneliness may also be important risk factors for the development of CVD in women. A recent cohort study of nearly 60,000 women older than 65 years showed that loneliness and social isolation were associated with a 5% and 8% higher risk of CVD, respectively, even after adjusting for health behaviors and outcomes. Women with greater loneliness and social isolation had a 13–27% higher risk of incident CVD compared with women with relatively less social isolation and less loneliness.

Sex-specific factors related to reproductive and pregnancy history are now recognized as important risk-enhancing factors for CVD in women. Recent studies demonstrate that the greatest risk for CVD

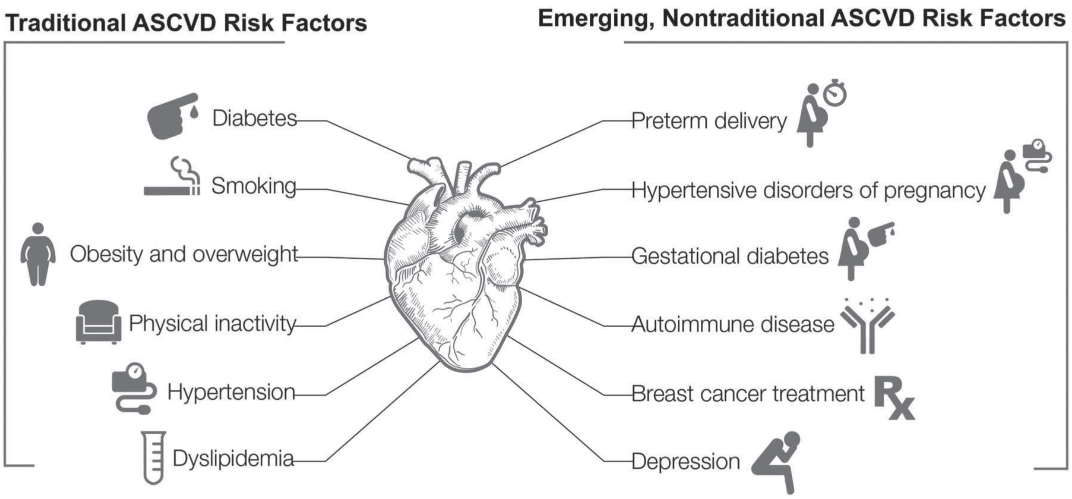

FIGURE 410-2 Traditional and nontraditional risk factors for atherosclerotic cardiovascular disease (ASCVD) in women. *(Reproduced with permission from M Garcia et al: Cardiovascular disease in women: Clinical perspectives. Circ Res 118:1273, 2016.)*

(at least twofold) was conferred by adverse pregnancy outcomes (APO), including a history of stillbirth, preterm birth, or preeclampsia, followed by a 1.5- to 1.9-fold risk with gestational diabetes and hypertension, premature ovarian insufficiency, and placental abruption; the lowest risk (<1.5 fold) was associated with early menarche, early menopause, parity, and polycystic ovary syndrome. Given these strong associations, targeted counseling and increased surveillance of CVD risk factors is warranted for women in high-risk groups. Accordingly, the 2019 American Heart Association (AHA) guidelines included a recommendation to assess for pregnancy complications as a part of routine cardiovascular risk assessment for primary prevention.

The sex differences in CVD prevalence, beneficial biologic effects of estrogen on the cardiovascular system, and reduced risk for CVD in observational studies led to the hypothesis that HT was cardioprotective. However, the WHI, which studied >16,000 women on CEE plus medroxyprogesterone acetate (MPA) or placebo and >10,000 women with hysterectomy on CEE alone or placebo, did not demonstrate a benefit of HT for the primary or secondary prevention of CVD. In addition, CEE plus MPA was associated with an increased risk for CVD, particularly in the first year of therapy, whereas CEE alone neither increased nor decreased CVD risk. Both HT groups were associated with an increased risk for ischemic stroke. In a subgroup analysis of the WHI estrogen-alone trial, a relatively younger age (50–59 years) combined with a history of bilateral salpingo-oophorectomy (BSO) was associated with a >30% CEE treatment–associated reduction in all-cause mortality, whereas CEE-treated older women with prior BSO did not see a significant reduction in any other outcomes, including incidence of coronary artery disease, invasive breast cancer, all-cause mortality, and a composite index of the aforementioned outcomes plus stroke, hip fracture, pulmonary embolism, and colorectal cancer.

These results suggest that postmenopausal women younger than 60 with prior BSO may have mortality benefit from HT, while women older than 60 with BSO may suffer consequences associated with HT. More recent data from KEEPS indicate that even if estrogen therapy is initiated shortly after the menopausal transition, it does not reduce atherosclerotic progression or impact CVD outcomes. Additionally, HT and placebo groups have similar outcomes with respect to venous thromboembolism and breast cancer. Although HT does not slow CVD development as previously thought, findings from KEEPS suggest that treated women experience significant improvements in vasomotor symptoms, mood, sexual function, and bone density, especially when therapy is started sooner after menopause onset. HT is discussed further in **Chap. 407**.

DIABETES MELLITUS

(See also Chap. 415) Women are more sensitive to insulin than men. Despite this, the prevalence of type 2 DM is similar in men and women. There is a sex difference in the relationship between endogenous androgen levels and DM risk. Higher bioavailable testosterone levels are associated with increased risk in women, whereas lower bioavailable testosterone levels are associated with increased risk in men. This observation has been confirmed in a recent Mendelian randomization that found that genetically determined higher testosterone increases risk for DM in women but reduces risk in men. Polycystic ovary syndrome, preeclampsia, pregnancy-associated hypertension, and gestational DM—common conditions in premenopausal women—are associated with a significantly increased risk for type 2 DM. Among individuals with DM, women have a greater risk for MI than do men. Women with DM have a sixfold greater risk of dying of CVD compared to women without DM. Premenopausal women with DM lose the cardioprotective effect of female sex and have rates of CVD identical to those in males. These women have impaired endothelial function and reduced coronary vasodilatory responses, which may predispose to cardiovascular complications. Women with DM are more likely to have left ventricular hypertrophy. Women with DM receive less aggressive treatment for modifiable CVD risk factors than men with DM.

HYPERTENSION

(See also Chap. 288) After age 60, hypertension is more common in U.S. women than in men, largely because of the high prevalence of hypertension in older age groups and the longer survival of women. Isolated systolic hypertension is present in 30% of women >60 years old. Sex hormones affect blood pressure. Both normotensive and hypertensive women have higher blood pressure levels during the follicular phase than during the luteal phase. In the Nurses' Health Study, the relative risk of hypertension was 1.8 in current users of oral contraceptives, but this risk is lower with the newer low-dose contraceptive preparations. Long-term data from this cohort also demonstrate that having gestational hypertension or preeclampsia during a first pregnancy doubled the rate of developing chronic hypertension. HT is not associated with hypertension. Among secondary causes of hypertension, there is a female preponderance of renal artery fibromuscular dysplasia.

The benefits of treatment for hypertension have been dramatic in both women and men. A meta-analysis of the effects of hypertension treatment, the Individual Data Analysis of Antihypertensive Intervention Trial, found a reduction of risk for stroke and major cardiovascular events in women. The effectiveness of various antihypertensive drugs appears to be comparable in women and men; however, women may experience more side effects, such as cough with angiotensin-converting enzyme inhibitors.

AUTOIMMUNE DISORDERS

(See also Chap. 367) Most autoimmune disorders occur more commonly in women than in men; they include autoimmune thyroid and liver diseases, Hashimoto's hypothyroidism, Graves' disease (GD), SLE, RA, scleroderma, multiple sclerosis (MS), and idiopathic thrombocytopenic purpura. However, there is no sex difference in the incidence of type 1 DM, and ankylosing spondylitis occurs more commonly in men. Sex differences in both immune responses and adverse reactions to vaccines have been reported. For example, there is a female preponderance of postvaccination arthritis.

Adaptive immune responses are more robust in women than in men; this may be explained by the stimulatory actions of estrogens and the inhibitory actions of androgens on the cellular mediators of immunity. Consistent with an important role for sex hormones, there is variation in immune responses during the menstrual cycle, and the activity of certain autoimmune disorders is altered by castration or pregnancy (e.g., RA and MS may remit during pregnancy). Nevertheless, the majority of studies show that exogenous estrogens and progestins in the form of HT or oral contraceptives do not alter autoimmune disease incidence or activity. Exposure to fetal antigens, including circulating fetal cells that persist in certain tissues, has been speculated to increase the risk of autoimmune responses. There is clearly an important genetic component to autoimmunity, as indicated by the familial clustering and HLA association of many such disorders. X chromosome genes also contribute to sex differences in immunity. Indeed, nonrandom X chromosome inactivation may be a risk factor for autoimmune diseases.

■ HIV INFECTION

(See also Chap. 208) Women (sex assigned at birth) accounted for 18% (6600) of the ~36,100 new HIV diagnoses in the United States in 2021. This represents a similar incidence observed in recent years. Black/African-American women accounted for 54% of new diagnoses among people assigned as female at birth, as compared to white women, who accounted for 23%, and Hispanic/Latino women, who represented about 18% of the newly diagnosed. AIDS remains an important cause of death in younger women, particularly African-American women aged 25–44 years. Heterosexual contact with an at-risk partner is the fastest-growing transmission category, and women are more susceptible to HIV infection during vaginal sex than men. This increased susceptibility is accounted for in part by an increased prevalence of sexually transmitted diseases, i.e., gonorrhea and syphilis, in women.

Some studies have suggested that hormonal contraceptives may increase the risk of HIV transmission. Progesterone has been shown to increase susceptibility to infection in nonhuman primate models of HIV. Women are also more likely to be infected by multiple variants of the virus than men. Women with HIV have more rapid decreases in their CD4 cell counts than do men. Compared with men, HIV-infected women more frequently develop candidiasis, but Kaposi's sarcoma is less common than it is in men. Women have more adverse reactions, such as lipodystrophy, dyslipidemia, and rash, with antiretroviral therapy than do men. This observation is explained in part by sex differences in the pharmacokinetics of certain antiretroviral drugs, resulting in higher plasma concentrations in women.

■ OBESITY

(See also Chap. 414) The prevalence of both obesity (body mass index ≥30 kg/m^2) and abdominal obesity (waist circumference ≥88 cm in women) are similar in U.S. women and men. According to the most recent National Health and Nutrition Examination Survey data spanning 2017 until prepandemic March 2020, the age-adjusted prevalence of obesity among U.S. adults was 41.8%, and there were no significant differences observed between women and men, even across different age groups. However, some sex-specific differences were observed: the prevalence of obesity was highest among non-Hispanic black women (57.9%) as compared with non-Hispanic white (39.6%), Hispanic (45.7%), and non-Hispanic Asian women (14.5%). Non-Hispanic black women had a higher prevalence of obesity compared to non-Hispanic black men. There were no significant differences in prevalence between men and women among non-Hispanic white, non-Hispanic Asian, or Hispanic adults. More than 80% of patients who undergo bariatric surgery are women. Pregnancy and menopause are risk factors for obesity.

There are major sex differences in body fat distribution. Women characteristically have a gluteal and femoral or gynoid pattern of fat distribution, whereas men typically have a central or android pattern. Women have more subcutaneous fat than men. In women, endogenous androgen levels are positively associated with abdominal obesity, and androgen administration increases visceral fat. In contrast, there is an inverse relationship between endogenous androgen levels and abdominal obesity in men. Further, androgen administration decreases visceral fat in these obese men. The reasons for these sex differences in the relationship between visceral fat and androgens are unknown; however, emerging evidence suggests that there is a contribution of genetic variation. Studies in humans also suggest that sex steroids play a role in modulating food intake and energy expenditure.

In men and women, abdominal obesity characterized by increased visceral fat is associated with an increased risk for CVD and DM. Obesity increases a woman's risk for certain cancers, in particular postmenopausal breast and endometrial cancer, in part because adipose tissue provides an extragonadal source of estrogen through aromatization of circulating adrenal and ovarian androgens, especially the conversion of androstenedione to estrone. Obesity increases the risk of infertility, miscarriage, and complications of pregnancy.

In addition to sex-specific epidemiologic differences in obesity, notable patterns have emerged with the advent and rising popularity of the novel obesity pharmacotherapies, namely the glucagon-like 1 peptide (GLP1) receptor agonists and the combined GLP1/gastric inhibitory polypeptide (GIP) receptor agonists **(Chap. 414)**. In the recent Semaglutide Treatment Effect in People with Obesity (STEP 1–4) studies from 2021, in which weekly semaglutide was compared to placebo for weight management, women comprised 55–80% of the study populations. Subgroup analyses evaluating the efficacy of semaglutide by sex have demonstrated a greater average weight reduction in women participants compared to men. Possible explanations for these findings include a lower average baseline body weight in women, potential differences in eating behavior that may be regulated by differences in sex hormones, and differential rates of gastric emptying. Sex-related differences in gastrointestinal side effects may also play a role. A possible connection may exist at the neurologic level: a recent study in mice demonstrated sex-specific dynamics in GLP1 receptor and GIP receptor expression patterns in metabolically responsive areas of the hypothalamus and the amygdala induced by exposure to a high-fat diet. It remains unclear whether mechanistic differences may explain the differential outcomes in women compared to men who use these medications.

■ OSTEOPOROSIS

(See also Chap. 423) Osteoporosis is about five times more common in postmenopausal women than in age-matched men, and osteoporotic hip fractures are a major cause of morbidity in elderly women. Men accumulate more bone mass and lose bone more slowly than do women. Sex differences in bone mass are found as early as infancy. Calcium intake, vitamin D, and estrogen all play important roles in bone formation and bone loss. Particularly during adolescence, calcium intake is an important determinant of peak bone mass. Vitamin D deficiency is surprisingly common in elderly women, occurring in >40% of women living in northern latitudes. Receptors for estrogens and androgens have been identified in bone. Estrogen deficiency is associated with increased osteoclast activity and a decreased number of bone-forming units, leading to net bone loss. The aromatase enzyme, which converts androgens to estrogens, is also present in bone. Estrogen is an important determinant of bone mass in men (derived from the aromatization of androgens) as well as in women.

■ PHARMACOLOGY

On average, women have lower body weights, smaller organs, a higher percentage of body fat, and lower total-body water than men. There are also important sex differences in drug action and metabolism that are not accounted for by these differences in body size and composition. Sex steroids alter the binding and metabolism of a number of drugs. Further, menstrual cycle phase and pregnancy can alter drug action.

Women also take more medications than men, including over-the-counter formulations and supplements. The greater use of medications combined with these biologic differences may account for the reported higher frequency of adverse drug reactions in women than in men.

Two-thirds of cases of drug-induced torsades des pointes, a rare, life-threatening ventricular arrhythmia, occur in women because they have a longer, more vulnerable QT interval. These drugs, which include certain antihistamines, antibiotics, antiarrhythmics, and antipsychotics, can prolong cardiac repolarization by blocking cardiac voltage-gated potassium channels.

PSYCHOLOGICAL DISORDERS

(See also Chap. 463) Depression, anxiety, and affective and eating disorders (bulimia and anorexia nervosa) are more common in women than in men. Epidemiologic studies from both developed and developing nations consistently find major depression to be twice as common in women as in men, with the sex difference becoming evident in early adolescence. Depression occurs in 10% of women during pregnancy and in 10–15% of women during the postpartum period. There is a high likelihood of recurrence of postpartum depression with subsequent pregnancies. The incidence of major depression diminishes after the age of 45 years and does not increase with the onset of menopause. Depression in women appears to have a worse prognosis than does depression in men; episodes last longer, and there is a lower rate of spontaneous remission. Schizophrenia and bipolar disorders occur at equal rates in men and women, although there may be sex differences in symptoms.

Both biologic and social factors account for the greater prevalence of depressive disorders in women. Men have higher levels of the neurotransmitter serotonin. Sex steroids also affect mood, and fluctuations during the menstrual cycle have been linked to symptoms of premenstrual syndrome. Sex hormones differentially affect the hypothalamic-pituitary-adrenal responses to stress. Testosterone appears to blunt cortisol responses to corticotropin-releasing hormone. Both low and high levels of estrogen can activate the hypothalamic-pituitary-adrenal axis.

COVID-19 INFECTION

(See also Chap. 205) Soon after the discovery of COVID-19, which was identified in November 2019 in Wuhan, China, as being caused by the novel coronavirus SARS-CoV-2, it was evident that there were appreciable sex differences in severity and outcomes. Indeed, observational data from the early pandemic, spanning late 2019 to early 2020, demonstrated a higher overall incidence of infectious cases, hospitalizations, intensive care unit admissions, and case-fatality rates among men as compared to women. More pronounced sex differences were observed with advanced age, with a higher overall incidence in older male age groups. These sex differences have persisted among different racial, ethnic, and socioeconomic groups and across all continents, as SARS-CoV-2 became a global pandemic.

There are several potential mechanisms for these sex-specific effects of SARS-CoV-2 infection **(Fig. 410-3)**. The virus's entry point into cells is the membrane-bound angiotensin-converting enzyme 2 (ACE2) receptor, and it also harnesses the primer TMPRSS2, a cellular serine protease. Circulating levels of ACE2, which is expressed in a variety of tissues including the lung, heart, and kidneys, have been reported to be relatively higher in men who have diabetes and/or kidney disease, as well as in healthy men; however, not all studies have reported similar sex differences. One hypothesis is that upregulation of the ACE2 receptor in men may provide greater opportunity for cellular entry, viral replication, and development of symptoms and deleterious sequelae.

ACE2 plays a critical role in bronchial transient secretory cells/type II alveolar cells as well as the renin-angiotensin-aldosterone system (RAAS). In the RAAS, ACE2 opposes angiotensin II's vasoconstrictive actions by converting angiotensin II to the vasodilatory angiotensin 1-7 in critical tissues, including cardiac myocytes, cardiac fibroblasts, and coronary endothelial cells. Importantly, recent evidence has shown the impact of sex and sex hormones on the RAAS and ACE2: estrogen downregulates angiotensin II receptor type 1 and regulates renin activity, as well as modulates local RAAS in the atrial myocardium. Furthermore, it has been shown that ovariectomized females have increased ACE2 activity and expression in their kidney and adipose tissue and that estradiol replacement reduces ACE2 expression. On the contrary, orchiectomized males have decreased ACE2 activity. Estrogen appears to reduce ACE2 expression in the heart and kidney in both rodent and human studies.

TMPRSS2, a vital contributor to SARS-CoV-2 cellular invasion, is a protein that is most abundantly expressed in prostate epithelial tissue, including high-grade prostate cancers and metastases. Accordingly, the protein's involvement in viral priming is thought to be an important reason for the higher case-fatality rate observed in men; however, the association has not yet been proven. TMPRSS2 is also expressed in airway epithelia, where its physiologic function is not entirely clear. Transcription of the cellular protein is regulated by androgenic ligands and androgen receptor binding element; it is unknown whether estrogen plays a role in its regulation. Emerging in vitro studies have demonstrated that a TMPRSS2 inhibitor blocks viral entry into cells. These data may serve as an important foundation for sex-specific and personalized therapeutic approaches in the future.

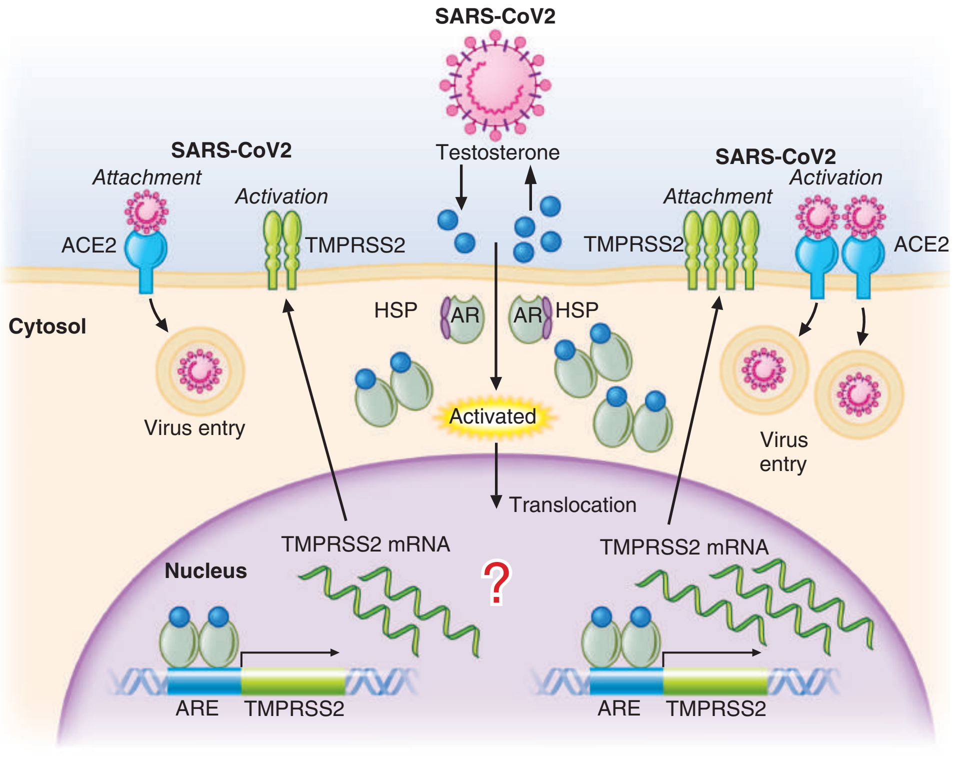

FIGURE 410-3 Proposed sex hormone differences in TMPRSS2-mediated SARS-CoV-2 host cell entry. The virus entry point into cells is the membrane-bound angiotensin-converting enzyme 2 (ACE2) receptor. The cell-membrane protease, TMPRSS2, is also vital for host cell entry. Circulating levels of ACE2, expressed abundantly in the lung, heart, and kidney tissues, have been reported to be relatively higher in men. Upregulation of the ACE2 receptor in men may provide greater opportunity for cellular entry, viral replication, symptom development, and multiorgan involvement. *(Adapted from C Gebhard et al: Impact of sex and gender on COVID-19 outcomes in Europe. Biol Sex Differ 11:29, 2020.)*

SUBSTANCE ABUSE AND TOBACCO

(See also Chaps. 464 and 465) Substance abuse is more common in men than in women. However, nearly one-third of Americans who suffer from alcoholism are women. Women are less likely to be diagnosed with alcoholism than men. A greater proportion of men than women seek help for alcohol and drug abuse. Men are more likely to go to an alcohol or drug treatment facility, whereas women tend to approach a primary care physician or mental health professional for help under the guise of a psychosocial problem. Blood alcohol levels are higher in women than in men after drinking equivalent amounts of alcohol, adjusted for body weight. This greater bioavailability of alcohol in women is due to both the smaller volume of distribution and the slower gastric metabolism of alcohol secondary to lower activity of gastric alcohol dehydrogenase than is the case in men. Women with alcoholism have a higher mortality rate than do women and men without alcoholism. Women also appear to develop alcoholic liver disease and other alcohol-related diseases with shorter drinking histories and lower levels of alcohol consumption. Alcohol abuse also poses special risks to a woman, adversely affecting fertility and the health of the baby (fetal alcohol syndrome). Even moderate alcohol use increases the risk of breast cancer, hypertension, and stroke in women.

More men than women smoke tobacco, but this sex difference continues to decrease. Women have a much larger burden of smoking-related disease. Smoking markedly increases the risk of CVD in premenopausal women and is also associated with a decrease in the age of menopause. Women who smoke are more likely to develop chronic obstructive pulmonary disease and lung cancer than men and at lower levels of tobacco exposure. Postmenopausal women who smoke have lower bone density than women who never smoked. Smoking during pregnancy increases the risk of preterm deliveries and low birth weight infants.

VIOLENCE AGAINST WOMEN

Approximately 15% of women in the United States have experienced rape, physical violence, and/or stalking by an intimate partner, as compared to 4% of men with similar experiences. Adult women are much more likely to be raped by a spouse, ex-spouse, or acquaintance than by a stranger. Intimate partner violence (IPV) is a leading cause of death among young women. Rates of reported IPV in the United States increased dramatically amid stay-at-home orders during the COVID-19 pandemic but have since declined again to prepandemic rates. IPV is an important risk factor for depression, substance abuse, and suicide in women. Screening instruments can identify women experiencing IPV and should be administered in settings that ensure adequate privacy and safety.

SUMMARY

Women's health is now a mature discipline, and the importance of sex differences in biologic processes is well recognized. Nevertheless, ongoing misperceptions about disease risk, not only among women but also among their health care providers, result in insufficient attention to modifiable risk factors. Research into the fundamental mechanisms of sex differences will provide important biologic insights. Further, those insights will have an impact on both women's and men's health.

FURTHER READING

Howell E: Reducing disparities in severe maternal morbidity and mortality. Clin Obstet Gynecol 61:2, 2018.

Lott N et al: Sex hormones in SARS-CoV-2 susceptibility: Key players or confounders? Nat Rev Endocrinol 19:217, 2023.

Rich-Edwards J et al: Sex and gender differences research design for basic, clinical, and population studies: Essentials for investigators. Endocr Rev 39:4, 2018.

Rubino D et al: Effect of continued weekly subcutaneous semaglutide vs placebo on weight loss maintenance in adults with overweight or obesity: The STEP 4 randomized clinical trial. JAMA 325:14, 2021.

Singh T et al: Takotsubo syndrome: Pathophysiology, emerging concepts, and clinical implications. Circulation 145:13, 2022.

411

Men's Health

Shalender Bhasin

The emergence of men's health as a distinct discipline within internal medicine is founded on the wide consensus that men and women differ across their life span in their susceptibility to disease, in the clinical manifestations of the disease, and in their response to treatment. Furthermore, men and women weigh the health consequences of illness differently and have different motivation for seeking care. Men and women experience different types of disparities in access to health care services and in the manner in which health care is delivered to them because of a complex array of socioeconomic and cultural factors. Attitudinal and institutional barriers to accessing care, fear and embarrassment due to the perception that it is not manly to seek medical help, and reticence on the part of patients and physicians in discussing issues related to sexuality, substance use, and aging have heightened the need for programs tailored to address the specific health needs of men.

The sex differences in disease prevalence, susceptibility, and clinical manifestations of the disease were discussed in **Chap. 410** (Women's Health) and will not be discussed here. It is notable that the two leading causes of death in both men and women—heart disease and cancer—are the same. However, men have higher prevalence of neurodevelopmental and degenerative disorders; substance use disorders, including the use of performance-enhancing drugs and alcohol use disorder; diabetes; cardiovascular disease; liver cirrhosis; and some types of cancer such as prostate, melanoma, and pancreatic cancer; and women have a higher prevalence of autoimmune disorders, depression, rheumatologic disorders, and osteoporosis. Men are substantially more likely to die from accidents, suicides, and homicides than women. Among men 15–34 years of age, unintentional injuries, homicides, and suicides account for over three-fourths of all deaths. Among men 35–64 years of age, heart disease, cancer, and unintentional injuries are the leading causes of death. Among men ≥65 years of age, heart disease, cancer, lower respiratory tract infections, and stroke are the major causes of death.

From 1999 to 2010, the mortality rates in the United States decreased for men and women of all age groups, largely due to reduced death rates from heart attacks, cancer, motor vehicle injuries, and HIV infection. However, since 2010, troubling disparities in sex-specific mortality rates have emerged among middle-aged men in the United States. From 2010 to 2017, the death rates have risen and life expectancy has decreased for young and middle-aged men. The increase in mortality rates among people aged 25–64 was the highest in the Ohio Valley and Appalachia, and in the New England states of New Hampshire, Maine, and Vermont. The rising mortality rates in young and middle-aged men have been attributed to an increase in deaths due to drug overdose, alcohol-related liver disease, and suicide. The rising rates of "deaths of despair" among young and middle-aged men, especially white non-Hispanics, have been associated with deterioration in economic and social well-being, reduced rates of marriage and labor force participation, and poor physical and mental health.

The biologic bases of sex differences in disease susceptibility, progression, and manifestation remain incompletely understood and are likely multifactorial. Undoubtedly, sex-specific differences in the genetic architecture and circulating sex hormones influence disease phenotype; additionally, epigenetic effects of sex hormones during fetal life, infancy, and pubertal development may epigenetically imprint sexual and nonsexual behaviors, body composition, and disease susceptibility. The circulating and tissue concentrations of sex hormones differ substantially in men and women, and these hormonal differences may affect gene expression in cells of males and females in all parts of the body. The presence of only one X chromosome in men renders them more susceptible to X-linked disorders than women. Due to the X inactivation of one randomly chosen X chromosome, women's bodies contain two epigenetically different cell populations. The genes that do not undergo X inactivation exhibit dosage differences between male and female cells. Expression of the Y chromosome genes in men may affect the function of somatic cells containing the Y chromosome. The loss of Y chromosome with aging in men is associated with increased risk of heart disease, especially heart failure, certain types of cancers, and shortened life span. The differences in the imprinting of maternally and paternally derived genes may also contribute to sex differences in the expression of disease. Reproductive load and physiologic changes during pregnancy, including profound hormonal and metabolic shifts and microchimerism (transfer of cells from the mother to the fetus and from the fetus to the mother), may affect disease susceptibility and disease severity in women. Sociocultural norms of child-rearing practices, societal expectations of gender roles, and the long-term economic impact of these practices and gender roles influence health behaviors and disease risk. Furthermore, the trajectories of age-related changes in sex hormones during the reproductive and postreproductive years vary substantially between men and women and influence the sex-specific patterns of the temporal evolution of age-related conditions such as osteoporosis, breast cancer, and autoimmune disease.

In a reflection of the growing attention on issues related to men's health, men's health clinics have mushroomed all over the country. Although the major threats to men's health have not changed—heart disease, cancer, and unintentional injury continue to dominate the list of major medical causes of morbidity and mortality in men—the men who attend men's health clinics do so largely for sexual, reproductive, and urologic health concerns involving common conditions, such as testosterone deficiency syndromes, age-related decline in testosterone levels, sexual dysfunction, muscle dysmorphia and anabolic-androgenic steroid (AAS) use, lower urinary tract symptoms (LUTS), and medical complications of prostate cancer therapy, which are the subjects of this chapter. Additionally, we are witnessing the emergence of new categories of body image disorders in men that had not been recognized until the 1980s, such as the body dysmorphia syndrome and the use of performance-enhancing drugs to increase muscularity and lean appearance. Although menopause and women's health have been the subject of intense investigation for more than five decades, these issues that are specific to men's health are just beginning to gain the attention that they deserve because of their high prevalence and impact on overall health, well-being, and quality of life.

AGING-RELATED CHANGES IN MALE REPRODUCTIVE FUNCTION

A number of cross-sectional and longitudinal studies (e.g., the Baltimore Longitudinal Study of Aging, the Framingham Heart Study [FHS], the Massachusetts Male Aging Study, and the European Male Aging Study [EMAS]) have established that testosterone concentrations decrease with advancing age. This age-related decline starts in the third decade of life and progresses slowly **(Fig. 411-1)**; the magnitude and trajectory of age-related decline in testosterone levels are affected by adiposity and weight change, comorbid conditions, and genetic factors. Because sex hormone–binding globulin (SHBG) concentrations are higher in older men than in younger men, free or bioavailable testosterone concentrations decline with aging to a greater extent than total testosterone concentrations. In the EMAS, 2.1% of community-dwelling men aged 40–70 years had total testosterone levels <317 ng/dL and a free testosterone level of <64 pg/mL, as well as sexual symptoms. The age-related decline in testosterone is due to defects at all levels of the hypothalamic-pituitary-testicular (HPT) axis. Pulsatile gonadotropin-releasing hormone (GnRH) secretion is attenuated, luteinizing hormone (LH) response to GnRH is reduced, and the frequency, amplitude, and area of LH pulses are reduced in older men. Leydig cell number is reduced and testicular response to LH is impaired. The gradual rise of LH with aging suggests that testis dysfunction is the main cause of declining androgen levels.

In epidemiologic surveys, low total and bioavailable testosterone concentrations in middle-aged and older men are associated with decreased sexual desire, poor erections, and diminished early morning erections; lower appendicular skeletal muscle mass, muscle strength, and self-reported physical function; increased risk of mobility limitation and falls; higher visceral fat mass, insulin resistance, and type 2 diabetes; reduced telomere length and increased all-cause and cardiovascular mortality; lower areal and volumetric bone mineral density and bone quality; and higher rates of bone fractures **(Table 411-1)**. Middle-aged and older men with hypogonadism report a high frequency of depressive symptoms. However, testosterone levels have not been consistently associated with major depressive disorder. An analysis of signs and symptoms in older men in the EMAS revealed a syndromic association of sexual symptoms with total testosterone levels <320 ng/dL and free testosterone levels <64 pg/mL in community-dwelling older men. Men with hypogonadism or Klinefelter's syndrome have reduced prostate cancer mortality.

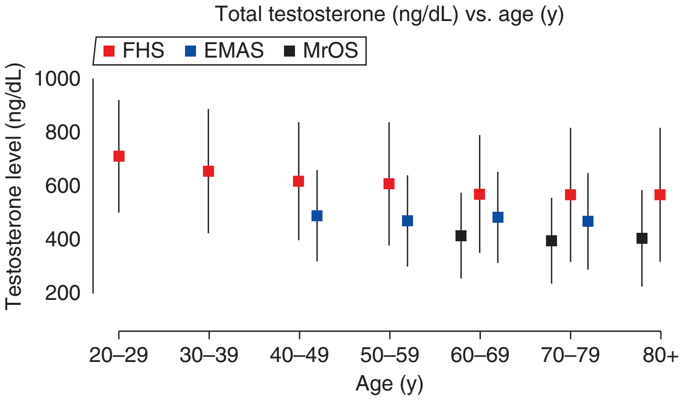

FIGURE 411-1 Age-related decline in total testosterone levels. Total testosterone levels measured using liquid chromatography–tandem mass spectrometry in men of the Framingham Heart Study (FHS), the European Male Aging Study (EMAS), and the Osteoporotic Fractures in Men Study (MrOS). *(Reproduced with permission from S Bhasin et al: Reference ranges for testosterone in men generated using liquid chromatography tandem mass spectrometry in a community-based sample of healthy nonobese young men in the Framingham Heart Study and applied to three geographically distinct cohorts. J Clin Endocrinol Metab 96:2430, 2011.)*

Mendelian randomization studies using data from the United Kingdom Biobank Study found a sexual dimorphic relation between genetically determined testosterone levels and the risk of type 2 diabetes; in men, lower genetically determined testosterone levels were associated with higher risk of type 2 diabetes, but in women, higher genetically determined testosterone levels were associated with higher risk of type 2 diabetes. Higher genetically determined testosterone levels were also associated with increased risk of prostate cancer in men in this study.

TABLE 411-1 Association of Testosterone Levels with Outcomes in Older Men

1. Positively associated with:
    - Muscle mass and muscle strength
    - Self-reported and performance-based measures of physical function
    - Sexual desire
    - Bone mineral density, bone geometry and quality, and volumetric bone mineral density
2. Negatively associated with risk of:
    - Coronary artery disease
    - Type 2 diabetes mellitus
    - Metabolic syndrome
    - All-cause mortality
    - Falls and fracture risk
    - Dementia and Alzheimer's disease
    - Frailty
    - Late-onset low-grade persistent depressive disorder (dysthymia)
3. Not associated with:
    - Lower urinary tract symptoms
    - Erectile dysfunction
    - Major depressive disorder

EFFICACY OF TESTOSTERONE REPLACEMENT THERAPY IN MIDDLE-AGED AND OLDER MEN WITH HYPOGONADISM

Two large, randomized trials—the Testosterone Trials (TTrials) and the Testosterone Replacement Therapy for Assessment of Long-Term Vascular Events and Efficacy Response in Hypogonadal Men (TRAVERSE) trial—have provided the most comprehensive data on the efficacy and safety of testosterone treatment in middle-aged and older men with hypogonadism. The TTrials, a set of seven coordinated placebo-controlled efficacy trials of testosterone replacement therapy, enrolled 788 men, aged 65 years or older, with an average of two morning, fasting total testosterone levels <275 ng/dL and one or more of the following: low sexual desire, mobility limitation, and/or fatigue. The primary aim of the TRAVERSE trial was to compare the effects of testosterone replacement therapy and placebo treatment on major adverse cardiovascular events in middle-aged and older men, 45–80 years, with two fasting, morning testosterone concentrations <300 ng/dL, one or more symptoms of hypogonadism, and preexisting coronary artery disease (CAD) or increased risk of CAD. The eligible participants were randomized with stratification for preexisting CAD to receive either 1.62% transdermal testosterone gel or a placebo gel daily for up to 5 years. Because of its large sample size and long duration, the TRAVERSE trial has provided important data on the efficacy and long-term safety of testosterone replacement therapy. In these and other trials, testosterone treatment was associated with greater improvement in overall sexual activity, sexual desire, and hypogonadal symptoms than placebo (Table 411-2). The effects of testosterone alone on erectile function were small and inconsistent across studies. Nearly 15% of men enrolled in the TRAVERSE trial had anemia; among men with anemia, a greater proportion of testosterone-treated men than placebo-treated men experienced correction of anemia. Testosterone treatment also reduced the incidence of anemia in men who were not anemic at baseline. Fifty percent of men with hypogonadism reported significant depressive symptoms at baseline; however, only 1.5% of men in the TRAVERSE trial met the rigorous definition of late-life-onset, low-grade persistent depressive disorder (previously called dysthymia). Testosterone treatment was more efficacious in improving depressive symptoms and energy level than placebo. Testosterone treatment improved volumetric as well as areal bone mineral density and estimated bone strength; the improvement in volumetric bone density in the spine was greater than in the hip and greater in the trabecular than the peripheral bone. Testosterone treatment also increased lean body mass, muscle strength, and some measures of physical function (Fig. 411-2). In the TTrials, testosterone treatment of older men with mobility limitation improved self-reported walking ability and modestly improved 6-min walking distance but did not affect falls.

In TRAVERSE trial participants who had prediabetes, the incidence of progression from prediabetes to diabetes was similar between testosterone- and placebo-treated groups. Testosterone treatment did not improve glycemic control in men with hypogonadism and prediabetes or diabetes at baseline. Thus, testosterone treatment should not be used alone to improve glycemic control in men with diabetes or prediabetes. These findings of the TRAVERSE trial are in contrast to those of the Testosterone for Diabetes Mellitus (T4DM) trial in men, aged 50–74 years without hypogonadism who had a newly diagnosed type 2 diabetes or were at increased risk for type 2 diabetes. Testosterone treatment in conjunction with a lifestyle program for 2 years reduced the proportion of participants with type 2 diabetes more than

TABLE 411-2 The Potential Benefits of Testosterone Replacement Therapy in Middle-Aged and Older Men with Hypogonadism

Strong Evidence of Efficacy

1. Improves overall sexual activity and sexual desire in men with low libido
2. Relieves hypogonadal symptoms
3. Improves depressive symptoms in men with significant depressive symptoms
4. Improves energy level
5. Corrects unexplained anemia and prevents the development of anemia
6. Increases skeletal muscle mass, reduces whole-body and abdominal fat mass
7. Increases areal and volumetric bone mineral density and estimated bone strength

Suggestive Evidence of Efficacy

1. Modestly improves mobility in older men with mobility limitation

Evidence of Lack of Efficacy

1. Improves cognitive function in older men without cognitive deficit
2. Improves depressive symptoms with major depressive disorder

Note: The Testosterone Trials (TTrials) and the TRAVERSE trial have provided robust evidence of efficacy of testosterone replacement therapy in middle-aged and older men with hypogonadism. The TTrials were a set of seven coordinated placebo-controlled trials whose primary aim was to determine whether testosterone treatment for 1 year of men aged 65 years or older with an average of two fasting morning testosterone levels <275 ng/dL plus one or more of low sexual desire, mobility limitation, and/or low vitality was more efficacious than placebo in improving sexual function, mobility, and/or vitality. The primary aim of the TRAVERSE trial was to compare the effect of testosterone replacement therapy with placebo on major adverse cardiovascular events in middle-aged and older men, 45–80 years of age, with at least two fasting morning testosterone level <300 ng/dL and preexisting coronary artery disease or increased risk of coronary artery disease. Because of its large sample size and longer treatment duration, the TRAVERSE trial has provided some of the most comprehensive evidence of the efficacy and safety of testosterone replacement therapy.

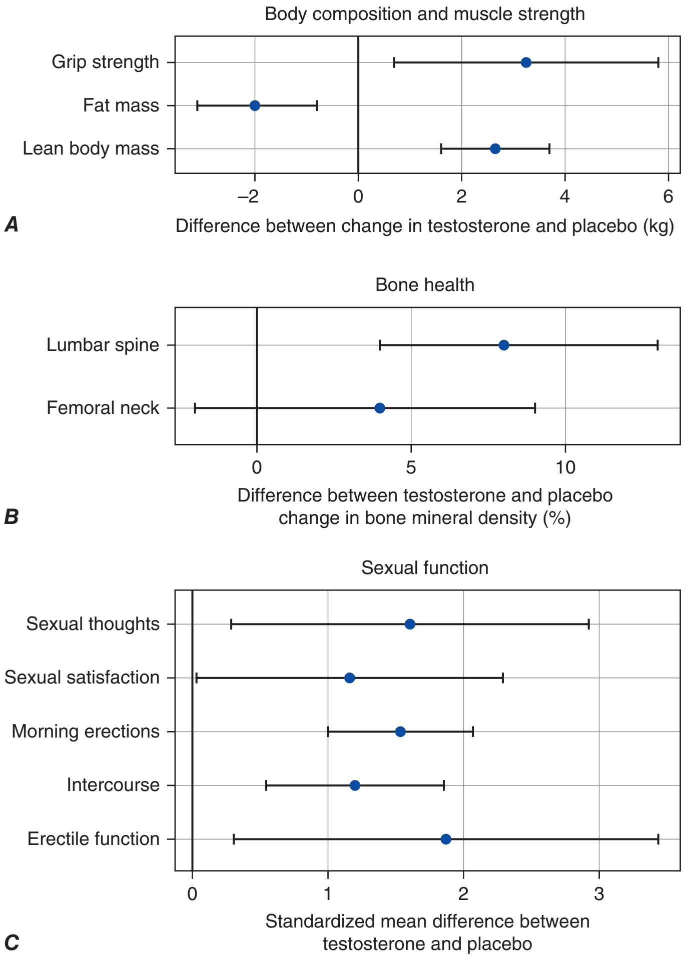

FIGURE 411-2 The effects of testosterone therapy on body composition, muscle strength, bone mineral density (BMD), and sexual function in intervention trials. The point estimates and the associated 95% confidence intervals are shown. ***A.*** The effects of testosterone therapy on lean body mass, grip strength, and fat mass in a meta-analysis of randomized trials. ***B.*** The effects of testosterone therapy on lumbar and femoral BMD in a meta-analysis of randomized trials. ***C.*** The effects of testosterone therapy on measures of sexual function in men with baseline testosterone <10 nmol/L (290 ng/dL). (***A.*** *Data from S Bhasin et al: Drug insight: Testosterone and selective androgen receptor modulators as anabolic therapies for chronic illness and aging. Nat Clin Pract Endocrinol Metab 2:146, 2006.* ***B.*** *Data from MJ Tracz et al: Testosterone use in men and its effects on bone health. A systematic review and meta-analysis of randomized placebo-controlled trials. J Clin Endocrinol Metab 91:2011, 2006.* ***C.*** *Data from AM Isidori et al: Effects of testosterone on sexual function in men: results of a meta-analysis. Clin Endocrinol (Oxf) 63:381, 2005.*)

TABLE 411-3 Adverse Events Associated with Testosterone Replacement Therapy in Middle-Aged and Older Men

Adverse Events for Which There Is Strong Evidence
Erythrocytosis
Venous thromboembolic events
Growth of metastatic prostate cancer
Reduced sperm production and fertility
Acne
Increased risk of detection of low-grade prostate cancer
Adverse Events for Which There Is Strong Evidence That Testosterone Treatment Does NOT Increase Risk
Major adverse cardiovascular events
Lower urinary tract symptoms
Adverse Events for Which There Is Weak Evidence
Atrial fibrillation
Bone fractures
Acute kidney injury
Male pattern balding
Adverse Events for Which There Is Insufficient Evidence to Assess Risk
Gynecomastia
Growth of breast cancer

placebo plus lifestyle program. Testosterone therapy has not been shown to improve cognitive function in men who do not have cognitive dysfunction.

SAFETY OF TESTOSTERONE REPLACEMENT THERAPY IN MIDDLE-AGED AND OLDER MEN WITH HYPOGONADISM

Erythrocytosis is the most frequent adverse event associated with testosterone treatment **(Table 411-3)**. In two placebo-controlled trials, the rates of atherosclerosis progression did not differ significantly between the testosterone and placebo groups. In the Cardiovascular Trial of the TTrials, testosterone treatment was associated with a greater increase in the volume of the noncalcified plaque than placebo. In the TRAVERSE trial, which was designed and powered specifically to assess the cardiovascular safety of testosterone, the incidence of major adverse cardiovascular events, which included nonfatal myocardial infarction, cardiovascular death, and nonfatal stroke, was similar between the testosterone- and placebo-treated men. The incidences of high-grade prostate cancer or any prostate cancer, acute urinary retention, invasive surgical procedures for benign prostatic hyperplasia, or initiation of new pharmacologic treatment for benign prostatic hyperplasia were low and did not differ significantly between the testosterone and placebo groups. Testosterone treatment did not worsen lower urinary tract symptoms but was associated with a higher incidence of venous thromboembolic events, atrial fibrillation, and acute kidney injury. The number of clinical bone fractures was greater in the testosterone group than in the placebo group; this finding was surprising because testosterone treatment increases areal and volumetric bone mineral density and estimated bone strength.

APPROACH TO THE PATIENT

Older Men with Age-Related Decline in Testosterone

Population screening of all older men for low testosterone levels is not recommended; testing should be restricted to men who have symptoms or physical features suggestive of hypogonadism. Testosterone treatment of older men with hypogonadism offers some clinical benefits (e.g., improvement of sexual symptoms in men with low libido, improvements in mood and energy, correction of anemia), but also has some potential for adverse effects (e.g., increased risk of venous thromboembolism and atrial fibrillation). An expert panel of the Endocrine Society recommended against testosterone treatment of *all* older men with low testosterone levels. Instead, the expert panel recommended that "testosterone therapy should be offered on an individualized basis … in men >65 years who have symptoms or conditions suggestive of testosterone deficiency (e.g., low libido, significant depressive symptoms, fatigue, or unexplained anemia) and consistently low testosterone." The decision to offer testosterone treatment to older men with hypogonadism should be a shared decision, guided by an individualized assessment of potential benefits and risks and careful weighing of the burden of symptoms/conditions against the potential benefits and risks **(Fig. 411-3)**. Evaluate whether the patient has clear evidence of testosterone deficiency confirmed by two or more early morning, fasting testosterone levels below the lower limit of normal for healthy young men plus the presence of symptoms. Weigh the burden of symptoms/conditions against the known benefits and the uncertainty of long-term harm. Ascertain whether the patient has any conditions that might increase the risk of harm, such as prostate cancer, severe LUTS, erythrocytosis, or a hypercoagulable condition. Older men considering testosterone treatment should undergo baseline evaluation of prostate cancer risk. The initiation of prostate screening and monitoring should be a shared decision because prostate cancer screening has some risks. A shared decision to treat should be accompanied by a standardized monitoring plan to optimize the benefit-to-risk ratio.

AGE-RELATED CHANGES IN FECUNDITY

Although testicular morphology, semen production, and fertility are maintained up to a very old age in men, advanced paternal age is a risk factor for reduced fertility. Compared to men aged 21–25 years, men >50 years have lower sperm motility and sperm morphology, a higher frequency of sperm tail defects, and lower fecundity. The fecundity is reduced when both parents are >40 years old. Increased workforce participation and changing career expectations of women, a higher age at reproductive union, and the availability of contraceptives that enable couples to separate their sexual and procreative lives have underpinned powerful secular trends toward postponement of childbearing to an older age. The median age at first childbirth has been increasing steadily across the world; postponement of childbirth to an older age increases the risk of involuntary childlessness because of the adverse effects of advanced maternal and paternal age on fecundity, increased risk of comorbidities that may indirectly affect fecundity, and the age-related changes in reproductive behaviors. Increased paternal age is associated with increased risk of germline mutations in the *FGFR2*, *FGFR3*, and *RET* genes and the associated autosomal dominant diseases, such as achondroplasia, Pfeiffer's syndrome, thanatophoric dysplasia, Crouzon's syndrome, Apert's syndrome, multiple endocrine neoplasia (MEN) 2A, and MEN 2B. Advanced paternal age also increases the offspring's risk of Klinefelter's syndrome, trisomy 13 and 18, neurodevelopmental disorders such as schizophrenia, autism, bipolar disorders, and cardiac malformations such as ventricular septal defects, atrial septal defects, and patent ductus arteriosus.

Sexual Dysfunction Various forms of sexual dysfunction are a major motivating factor for men seeking care at men's health clinics. The landmark descriptions of the human sexual response cycle by Masters and Johnson demonstrating that men and women display predictable physiologic responses after sexual stimulation provided the basis for rational classification of human sexual disorders. Subsequently, this model was expanded to include sexual desire, resulting in a three-stage model comprising of desire, arousal, and orgasm. Whipple and Brash-McGreer later proposed a circular model of sexual response whereby a positive sexual experience reinforces desire.

The classification of sexual disorders in men has been revised in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) and the eleventh revision of the International Classification of Diseases and Related Health Problems (ICD-11). DSM-5 includes three major categories of male sexual disorders: sexual desire disorders, sexual arousal disorders, and orgasmic disorders. The DSM-5 also

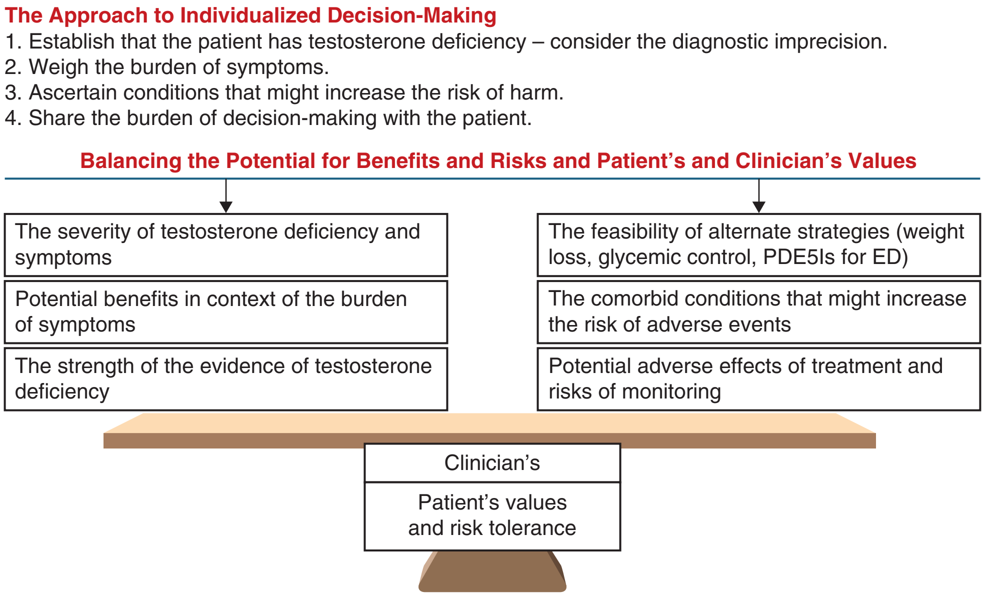
The Approach to Individualized Decision-Making
1. Establish that the patient has testosterone deficiency – consider the diagnostic imprecision.
2. Weigh the burden of symptoms.
3. Ascertain conditions that might increase the risk of harm.
4. Share the burden of decision-making with the patient.
Balancing the Potential for Benefits and Risks and Patient's and Clinician's Values
The severity of testosterone deficiency and symptoms
Potential benefits in context of the burden of symptoms
The strength of the evidence of testosterone deficiency
The feasibility of alternate strategies (weight loss, glycemic control, PDE5Is for ED)
The comorbid conditions that might increase the risk of adverse events
Potential adverse effects of treatment and risks of monitoring
Clinician's
Patient's values and risk tolerance

FIGURE 411-3. An approach to individualized, shared treatment decision-making in older men with testosterone deficiency. Testosterone treatment is not indicated in all older men with testosterone deficiency. Consider the diagnostic imprecision in establishing the diagnosis, the burden of symptoms, and the presence of comorbid conditions that might increase the risk of harm. Weigh the potential benefits against the burden of symptoms, the potential of harm from treatment and monitoring in the context of other comorbid conditions, and the feasibility of alternate strategies such as weight loss, improved glycemic control, and the use of selective phosphodiesterase-5 inhibitors (PDE5Is). A shared treatment decision takes into consideration the patient's and clinician's values. ED, erectile dysfunction.

includes an additional category of "other specified sexual dysfunction" due to causes other than those included in the above three categories. Male erectile dysfunction is now referred to as erectile disorder.

Classification of the patient's disorder into these categories is important as the etiologic factors, diagnostic tests, and therapeutic strategies vary for each class of sexual disorder. Historically, the classification and nomenclature for sexual disorders were based on the DSM, based on the erroneous belief that sexual disorders in men are largely psychogenic in origin. However, the recognition of erectile disorder as a manifestation of systemic disease and the availability of easy-to-use oral selective phosphodiesterase-5 (PDE5) inhibitors have placed sexual disorders in men within the purview of the primary care provider. These disorders have been discussed in Chap. 409 (Sexual Dysfunction).

MUSCLE DYSMORPHIA SYNDROME IN MEN—A FORM OF BODY IMAGE DISORDER

Muscle dysmorphia is a form of body image disorder characterized by a pathologic preoccupation with muscularity and leanness. The men with muscle dysmorphia express a strong desire to be more muscular and lean and report dissatisfaction and embarrassment about their body size and shape, preoccupation with bodybuilding and muscularity, and impairment of social and occupational functioning. Patients with muscle dysmorphia also report higher rates of mood and anxiety disorders and obsessive and compulsive behaviors than individuals without muscle dysmorphia.

Patients with muscle dysmorphia syndrome—nearly all men—are almost always engaged in weightlifting and body building and are more likely to use performance-enhancing drugs, especially AASs, than men in the general population or even weightlifters without body dysmorphia. Muscle dysmorphia disorder exposes men to an increased risk of disease due to the combined interactive effects of the intensity of physical exercise, the use of performance-enhancing drugs, and other lifestyle factors associated with weightlifting and the use of performance-enhancing drugs. No randomized trials of any treatment modalities have been conducted; anecdotally, behavioral and cognitive therapies have been tried with varying degrees of success.

AAS Abuse by Athletes and Recreational Bodybuilders The illicit use of AASs to enhance athletic performance first surfaced in the 1950s among powerlifters and spread rapidly to other sports, professional as well as high school athletes, and recreational bodybuilders. In the early 1980s, the use of AAS spread beyond the athletic community into the general population. Most AAS users are not athletes, but rather recreational weightlifters who use these drugs to look lean and more muscular.

The most commonly used AASs include testosterone esters, nandrolone, stanozolol, methandienone, trenbolone, boldenone, and oxandrolone. AAS users generally use increasing doses of multiple steroids in a practice known as stacking.

The adverse effects of long-term AAS abuse remain poorly understood. Most of the information about the adverse effects of AAS has emerged from case reports, uncontrolled studies, or clinical trials that used replacement doses of testosterone (Table 411-4). The adverse event data from clinical trials using physiologic replacement doses of testosterone have been extrapolated unjustifiably to AAS users who may administer 10–100 times the replacement doses of testosterone over many years and to support the claim that AAS use is safe. A substantial fraction of AAS users also use other drugs that are perceived to be muscle-building or performance-enhancing, such as growth hormone; erythropoiesis-stimulating agents; insulin; stimulants such as amphetamine, clenbuterol, cocaine, ephedrine, and thyroxine; and drugs perceived to reduce adverse effects such as human chorionic gonadotropin (hCG), aromatase inhibitors, or estrogen antagonists. Men who abuse AAS are more likely to engage in other high-risk behaviors than nonusers. The adverse events associated with AAS use may be due to AAS themselves, concomitant use of other drugs, or high-risk behaviors.

The high rates of mortality and morbidities observed in AAS users are alarming. One Finnish study reported 4.6 times the risk of death among elite power lifters than in age-matched men from the general population. The causes of death among power lifters included suicides, myocardial infarction, hepatocellular failure, and non-Hodgkin's lymphoma. A retrospective review of patient records in Sweden also reported higher standardized mortality ratios for AAS users than for nonusers. Studies indicate that 32% of deaths among AAS users were suicidal, 26% homicidal, and 35% accidental. The median age of death among AAS users in this study—24 years—was even lower than that for heroin or amphetamine users.

Numerous reports of cardiac death among young AAS users raise concerns about the adverse cardiovascular effects of AAS. High doses of AAS may induce proatherogenic dyslipidemia, increase thrombosis risk via effects on clotting factors and platelets, induce vasospasm

TABLE 411-4 Potential Adverse Effects Associated with the Use of Anabolic-Androgenic Steroids

A. Four Major Categories of Serious Adverse Effects Associated with the Use of Anabolic-Androgenic Steroids

1. *Cardiovascular*
   a. Hypertension
   b. Sudden death
   c. Cardiomyopathy
   d. Diastolic dysfunction
   e. Accelerated atherosclerosis
   f. Dyslipidemia
2. *Endocrine*
   a. Suppression of the hypothalamic-pituitary-testicular axis
      i. Anabolic steroid withdrawal hypogonadism
      ii. Infertility
      iii. Testicular atrophy
   b. Gynecomastia
3. *Neuropsychiatric*
   a. Major mood disorders
      i. Hypomania and mania during periods of high-dose anabolic-androgenic steroids use
      ii. Depression and even suicidality during anabolic steroid withdrawal
   b. Other behavioral disorders
      i. Rage reactions
      ii. Aggressive responding and violence
   c. Dependence and addiction behaviors
   d. Cognitive deficits
4. *Musculoskeletal Injuries*
   a. Tendon injuries
   b. Other musculoskeletal injuries

B. Other Adverse Events

1. *Hematologic*
   a. Erythrocytosis
   b. Venous thromboembolic events
2. *Hepatotoxicity*
   a. Inflammatory or cholestatic hepatic damage with the use of 17α-alkyl-substituted steroids
   b. Peliosis hepatis
   c. Hepatic adenoma
   d. Hepatocellular carcinoma
3. *Hair and skin*
   a. Acne
   b. Scalp hair loss in genetically predisposed men
   c. Increased body hair
   d. Striae
4. *Renal*
   a. Acute kidney injury due to rhabdomyolysis
   b. Focal glomerulosclerosis

through their effects on vascular nitric oxide, and induce myocardial hypertrophy and fibrosis. The supraphysiologic doses of testosterone and orally administered, 17-α-alkylated, nonaromatizable AAS are associated with marked reductions in high-density lipoprotein (HDL) cholesterol and increases in low-density lipoprotein (LDL) cholesterol.

Studies using tissue Doppler and strain imaging and magnetic resonance imaging (MRI) have reported diastolic and systolic dysfunction, including lower early and late diastolic tissue velocities, reduced E/A ratio, and reduced peak systolic strain in AAS users than in nonusers. Power athletes using AAS often have short QT intervals but increased QT dispersion, which may predispose them to ventricular arrhythmias. Long-term AAS use may be associated with myocardial hypertrophy and fibrosis. Myocardial tissue of power lifters using AAS has been shown to be infiltrated with fibrous tissue and fat droplets. AAS users demonstrate higher coronary artery plaque volume than nonusers, and lifetime AAS dose is associated with coronary atherosclerotic burden.

Long-term AAS use suppresses LH and follicle-stimulating hormone (FSH) secretion and inhibits endogenous testosterone production and spermatogenesis. Men who have used AAS for more than a few months experience marked suppression of the HPT axis after stopping AAS that may be associated with sexual dysfunction, fatigue, infertility, and depressive symptoms. In some long-term AAS users, HPT suppression may last more than a year, and in a few individuals, complete recovery may not occur. The symptoms of androgen deficiency during AAS withdrawal may cause some men to revert back to using AAS, leading to continued use and AAS dependence. As many as 30% of AAS users develop a syndrome of AAS dependence, characterized by long-term AAS use despite adverse medical and psychiatric effects. In some men's health clinics, as many as 25% of young men receiving testosterone replacement therapy have anabolic steroid withdrawal hypogonadism.

Supraphysiologic doses of testosterone may also impair insulin sensitivity. Orally administered androgens have been associated with insulin resistance and diabetes.

Unsafe injection practices, high-risk behaviors, and increased rates of incarceration render AAS users at increased risk of HIV and hepatitis B and C. In one survey, nearly 1 in 10 gay men had injected AAS or other substances, and AAS users were more likely to report high-risk unprotected anal sex than other men. Surveys of male prisoners find high rates of AAS use.

Some AAS users develop hypomanic and manic symptoms during AAS exposure (irritability, aggressiveness, reckless behavior, and psychotic symptoms, sometimes associated with violence) and major depression (sometimes associated with suicidality) during AAS withdrawal. Users may also develop other forms of illicit drug use.

AAS use has been associated with difficulties with spatial as well as working memory, problem solving, and attention, and structural and functional changes in many brain regions involved in inhibitory control and emotional regulation. A structural MRI study of users of high doses of AAS reported smaller cortical, gray matter, putamen, and corpus callosum volumes. Both low and high androgen levels have been associated with increased Aβ and tau-P levels and Aβ toxicity. These data have raised concern that long-term AAS use may increase the risk of Alzheimer's disease and related dementias.

Elevated liver enzymes, cholestatic jaundice, hepatic neoplasms, and peliosis hepatis have been reported with oral, 17-α-alkylated AAS. AAS use may cause muscle hypertrophy without compensatory adaptations in tendons, ligaments, and joints, thus increasing the risk of tendon and joint injuries. AAS use is associated with acne, baldness, and increased body hair.

APPROACH TO THE PATIENT

Detection of AAS Use

AAS users generally mistrust physicians and seek medical help infrequently; when they do seek medical help, it is often for the treatment of AAS withdrawal syndrome, infertility, gynecomastia, or other medical or psychiatric complications of AAS use. The suspicion of AAS use should be raised by the increased hemoglobin and hematocrit levels; suppressed LH, FSH, SHBG and testosterone levels; low HDL cholesterol; and low testicular volume and sperm density in a person who looks highly muscular (Table 411-5). A combination of these findings along with self-report of their use by the patient—which usually can be elicited by a tactful interview—are often sufficient to establish a diagnosis in clinical practice.

Accredited laboratories use gas chromatography and mass spectrometry or liquid chromatography and mass spectrometry to detect AAS abuse. Illicit testosterone use is detected generally by the measurement of the urinary testosterone-to-epitestosterone ratio and confirmed by the use of the ^{13}C:^{12}C ratio in testosterone by the use of isotope ratio combustion mass spectrometry. Exogenous testosterone administration increases urinary testosterone glucuronide excretion and consequently the testosterone-to-epitestosterone

TABLE 411-5 Detection of the Use of Anabolic-Androgenic Steroids

Clinical indicators that should raise suspicion of anabolic-androgenic steroid use
1. "V" muscular phenotype
2. Reduced testicular volume (<15 mL)
3. Excessive concern about leanness and muscularity.
Laboratory indicators
1. Suppressed LH, FSH, and SHBG levels
2. Increased hematocrit
Detection of anabolic-androgenic steroids
1. LC-MS/MS analysis of urine
Detection of exogenous testosterone use
1. Urinary testosterone-to-epitestosterone ratio
2. Isotope ratio mass spectrometry analysis to detect differences in ^{13}C:^{12}C ratio in exogenous and endogenous testosterone

Note: In clinical settings, the use of anabolic-androgenic steroids can often be ascertained simply by direct questioning. Reduced testicular volume, suppressed LH and FSH, and increased hematocrit in an unusually muscular man should raise suspicion of anabolic-androgenic steroid use. Although rarely needed in clinical practice, recent use of anabolic-androgenic steroids can be confirmed by LC-MS/MS analysis of urine. Exogenous testosterone use can be detected using the urinary testosterone-to-epitestosterone ratio and isotope ratio mass spectrometry analysis to detect differences in ^{13}C:^{12}C ratio in exogenous and endogenous testosterone.

Abbreviations: FSH, follicle-stimulating hormone; LC-MS/MS, liquid chromatography–tandem mass spectrometry; LH, luteinizing hormone.

ratio. Ratios >4 suggest exogenous testosterone use but can also reflect genetic variation. Genetic variations in the uridine diphospho-glucuronyl transferase 2B17 (*UGT2B17*), the major enzyme for testosterone glucuronidation, CYP17 (the aromatase gene), SHBG, and 3', 5' cyclic nucleotide phosphodiesterase (PDE7B) affect the testosterone-to-epitestosterone ratio and increase the risk of a false-negative test. Synthetic testosterone has a lower ^{13}C:^{12}C ratio than endogenously produced testosterone, and these differences in the ^{13}C:^{12}C ratio can be detected by isotope ratio combustion mass spectrometry and used to confirm exogenous testosterone use in individuals with a high testosterone-to-epitestosterone ratio.

TREATMENT

Integrated Management of Patients with AAS Use

The nonathlete weightlifters who abuse AAS frequently do not seek medical treatment and often mistrust physicians. They also do not view these drugs and the associated lifestyle as deleterious to their health. Many clinicians erroneously view AAS abuse as largely a problem of cheating in competitive sports, while, in fact, most AAS users are not athletes at all.

In addition to treating the underlying body dysmorphia disorder that motivates the use of these drugs and the addiction behaviors, the treatment should be directed at the symptoms or the condition for which the patient seeks therapy, such as infertility, sexual dysfunction, gynecomastia, or depressive symptoms. Accordingly, therapy may include some combination of cognitive and behavioral therapy for muscle dysmorphia syndrome, antidepressant therapy for depression, selective PDE5 inhibitors for erectile dysfunction, and/or use of selective estrogen receptor modulators, aromatase inhibitors, or hCG to restore testosterone levels.

As discussed above, AASs suppress the male hypothalamic-pituitary-gonadal axis, and men with long-term AAS use may experience symptoms of profound androgen deficiency such as sexual dysfunction, fatigue, and depressive symptoms during AAS withdrawal. Some of these patients may resume AAS use or start using other drugs to combat the distressing withdrawal symptoms. There are no randomized trials of any therapies for AAS withdrawal. Case reports and clinical experience suggest that administration of selective estrogen receptor modulators, CYP19 aromatase inhibitors, or hCG may restore circulating testosterone levels. Clomiphene citrate, a partial estrogen receptor agonist, administered in a dose of 25–50 mg on alternate days, can increase LH and FSH levels and restore testosterone levels in a vast majority of men with AAS withdrawal syndrome. However, the recovery of sexual function during clomiphene administration is variable despite improvements in testosterone levels. Anecdotally, other aromatase inhibitors such as anastrozole have also been used. hCG, administered by intramuscular injections of 1500 to 2000 IU three times each week, can raise testosterone levels into the normal range. Some patients may not respond to either clomiphene or hCG therapy, raising the possibility of irreversible long-term toxic effects of AAS on Leydig cell function.

Cognitive and behavioral therapy to address the body image disorder and addiction behaviors and antidepressants to treat depression may be needed. The opioid antagonist naltrexone blocks AAS dependence in animals. Therefore, treatments for human opioid dependence might also benefit AAS dependence. Many patients who abuse AAS suffer from body image disorder and require psychiatric treatment for this underlying disorder.

LUTS IN MEN

LUTS in men include storage symptoms (urgency, daytime as well as nighttime frequency, and urgency incontinence), voiding disturbances (slow or intermittent stream, difficulty in initiating micturition, straining to void, pain or discomfort during the passage of urine, and terminal dribbling), or postmicturition symptoms (a sense of incomplete voiding after passing urine and postmicturition dribble). The overactive bladder syndrome refers to urgency with or without urgency incontinence, usually with urinary frequency and nocturia, and is often due to detrusor muscle overactivity. A presumptive diagnosis of benign prostatic hyperplasia should be made only in men with LUTS, who have demonstrable evidence of prostate enlargement and obstruction based on the size of the prostate. LUTS have historically been attributed to benign prostatic hyperplasia, although it has become apparent that the pathophysiologic mechanisms of LUTS are complex and multifactorial and may include structural or functional abnormalities of the bladder, bladder neck, prostate, distal sphincter mechanism, and urethra, as well as abnormalities in the neural control of the lower urinary tract and autonomic dysfunction. Diuretics, antihistamines, antidepressants, and other medications that have anticholinergic properties can cause or exacerbate LUTS in older men. The intensity of LUTS tends to fluctuate over time.

LUTS is highly prevalent in older men, affecting nearly 50% of men >65 and 70% of men >80 years old. LUTS adversely affect quality of life because of their impact on sleep, ability to perform activities of daily living, and depressive symptoms. LUTS are often associated with erectile dysfunction.

APPROACH TO THE PATIENT

Lower Urinary Tract Symptoms

Medical evaluation should include assessment of potential causes of symptoms; medications including herbal and over-the-counter products that might contribute to symptoms; the symptom severity and bother using an International Prostate Symptom Score; and in some patients, a frequency-volume chart. The impact of LUTS on sleep, activities of daily living, and quality of life should be evaluated. Evaluation should also include digital prostate examination, neurologic examination focused on perineum and lower extremities, urinalysis, fasting blood glucose, electrolytes, creatinine, and prostate-specific antigen (PSA). Urodynamic studies are not required in most patients but are recommended when invasive surgical therapies are being considered. A urologic referral may be appropriate if the patient has hydronephrosis, renal insufficiency, recurrent urinary tract infections, hematuria, or history of acute urinary retention.

TREATMENT

Patients with LUTS

Considerations of the severity of symptoms; the impact of symptoms on sleep, activities of daily living, and quality of life; the natural history of the disease; and potential adverse effects of the intervention should guide the decision to intervene. In men with mild to moderately severe LUTS, the symptoms typically progress slowly over many years and may remain stable or even improve in some men. The men who have mild symptoms can usually be reassured and followed. Several simple steps such as reducing caffeine and alcohol intake, especially late in the day, taking the diuretic medication early in the day, avoiding excessive water intake close to bedtime, bladder training, pelvic floor exercises including biofeedback to promote pelvic floor relaxation, and timed voiding regimens or double voiding to ensure complete emptying of the bladder may be helpful in reducing the severity of symptoms. Men with mild to moderate bothersome LUTS can be treated effectively using α-adrenergic antagonists, steroid 5α-reductase inhibitors, PDE5 inhibitors, or anticholinergic agents alone or in combination. Selective α-adrenergic antagonists are typically the first line of therapy; their side effects include hypotension, dizziness, nasal congestion, retrograde or delayed ejaculation, and rarely floppy iris syndrome. In men with probable benign prostate obstruction with gland enlargement and LUTS, therapy with steroid 5α-reductase inhibitors, finasteride or dutasteride, improves urinary symptoms and flow rate and reduces prostatic volume. Long-term treatment with 5α-reductase inhibitors can reduce the risk of acute urinary retention and need for prostate surgery. Combined administration of a steroid 5α-reductase inhibitor and an α_1-adrenergic blocker can rapidly improve urinary symptoms and reduce the relative risk of acute urinary retention and surgery. PDE5 inhibitors, when administered chronically alone or in combination with α-adrenergic blockers, are effective in improving LUTS and erectile dysfunction through their effects on nitric oxide–cyclic guanosine monophosphate (cGMP) in the bladder, urethra, and prostate. PDE5 inhibitors do not improve uroflow parameters, and their hypotensive effect may be potentiated by α_1-adrenergic blockers. Anticholinergic drugs are used for the treatment of overactive bladder in men with prominent irritative symptoms, such as frequency, urgency, and incontinence, and no evidence of elevated postvoid residual urine. Containment products, such as pads, can help improve social life in men who have severe storage symptoms, including incontinence. Surgery is indicated when medical therapy fails, symptoms progress despite medical therapy, or the patient develops acute urinary retention, hydronephrosis, renal insufficiency, or recurrent urinary tract infections, or if the patient has postvoid residual urine volume >25% of the urinary bladder volume.

MEDICAL COMPLICATIONS OF PROSTATE CANCER THERAPY

Prostate cancer is the most common malignancy in American men. The majority of these men have low-grade, organ-confined prostate cancer; are treated with radical prostatectomy, radiation, or active surveillance; and have excellent prospects of long-term survival. Substantial improvement in survival in men with prostate cancer has focused attention on the high prevalence of sexual dysfunction, physical dysfunction, and low vitality in the men, which are important contributors to poor quality of life among patients treated for prostate cancer. The pathophysiology of these symptoms after radical prostatectomy or radiation therapy is multifactorial, but denervation and testosterone deficiency are important contributors to these symptoms.

Testosterone deficiency is common in men with prostate cancer. Testosterone levels decline with age, and men with prostate cancer are at risk of having low testosterone levels simply by virtue of their age. However, total and free testosterone levels are even lower in men with prostate cancer who have undergone prostatectomy, when compared to noncancer age-matched controls. Testosterone deficiency in men with prostate cancer is associated with fatigue, sexual dysfunction, mobility limitation, and decreased physical function. A majority of men treated with surgery or radiation therapy will develop sexual dysfunction and incontinence. Although there is some recovery of sexual function with passage of time, 40–50% of men undergoing radical prostatectomy find their sexual performance to be a moderate to large problem 18 months after surgery. Sexual problems are a source of psychosocial distress in men with localized prostate cancer. The men with locally advanced or metastatic prostate cancer who undergo androgen deprivation therapy (ADT) encounter even more distressing symptoms. In addition to fatigue, sexual dysfunction, and hot flushes, these men are at increased risk for diabetes, metabolic syndrome, coronary heart disease, and frailty.

Testosterone Therapy in Men with a History of Prostate Cancer A history of prostate cancer has historically been considered a contraindication for testosterone therapy. This guidance is based on observations that testosterone promotes the growth of metastatic prostate cancer and metastatic prostate cancer generally regresses after orchiectomy and ADT. Androgen receptor signaling plays a central role in maintaining growth of normal prostate and prostate cancer.

The role of testosterone in prostate cancer is complex. Epidemiologic studies and their meta-analyses have not revealed a consistent relationship between serum testosterone or dihydrotestosterone levels and prostate cancer. However, men with hypogonadism or Klinefelter's syndrome have lower prostate cancer mortality than the general population. In a Mendelian randomization analysis of men in the UK Biobank, higher genetically determined bioavailable testosterone levels were associated with increased risk of prostate cancer. Testosterone treatment of older men with low testosterone does not affect intraprostatic androgen levels or the expression of androgen-dependent prostatic genes. The suppression of circulating testosterone levels by a GnRH antagonist also does not affect intraprostatic androgen concentrations. Open-label trials and retrospective analyses of testosterone therapy in men with prostate cancer who have undergone radical prostatectomy and have undetectable PSA levels after radical prostatectomy have found very low rates of PSA recurrence.

A majority of men diagnosed with prostate cancer today have localized disease that can be potentially cured by radical prostatectomy. The men with organ-confined prostate cancer (pT2, N0, M0) and Gleason score 6 or 7 (3+4) are at a very low risk of disease recurrence after radical prostatectomy, with 0.5% biochemical recurrence rate and 0.2% local recurrence rate at >10–15 years. Therefore, men with organ-confined prostate cancer (pT2), Gleason score 6 or 7 (3+4), and a preoperative PSA of <10 ng/mL, who have had undetectable PSA levels (<0.1 ng/mL) for >2 years after radical prostatectomy, have very low risk of disease recurrence (<0.5% at 10 years) and may be considered for testosterone therapy on an individualized basis. If testosterone therapy is instituted, it should be associated with careful monitoring of PSA levels and close consultation with a urologist.

MEDICAL COMPLICATIONS OF ADT

In patients with prostate cancer and distant metastases, ADT improves survival. In patients with locally advanced disease, ADT in combination with external beam radiation or as an adjuvant therapy (postprostatectomy and pelvic lymphadenectomy) also has been shown to improve survival. However, ADT along with radiation is being increasingly used as primary therapy in men with localized disease and in men encountering biochemical recurrence. The overall use of ADT in men with prostate cancer has increased in the past two decades, and its use in men with localized disease and biochemical recurrence accounts for a substantial fraction of this increase. Since most men with prostate cancer die of conditions other than their primary malignancy, recognition and management of these adverse effects are paramount.

Profound hypogonadism resulting from ADT is associated with sexual dysfunction, vasomotor symptoms, gynecomastia, decreased muscle mass and strength, frailty, increased fat mass, anemia, fatigue, bone loss, loss of body hair, depressive symptoms, and reduced quality of life. Increased risk of diabetes and cardiovascular disease has

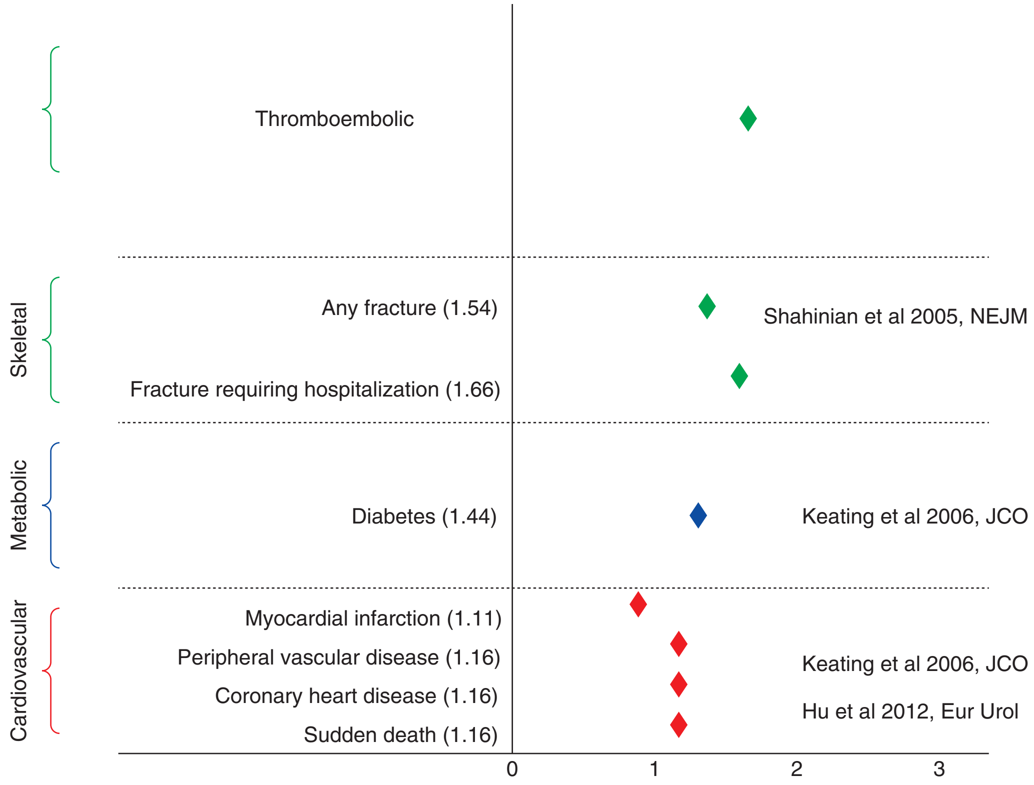

FIGURE 411-4 Adverse cardiometabolic and skeletal effects of androgen deprivation therapy (ADT) in men receiving ADT for prostate cancer. Administration of ADT has been associated with increased risk of thromboembolic events, fractures, and diabetes. Some, but not all, studies have reported increased risk of cardiovascular events in men receiving ADT. *(Data from VB Shahinian et al: N Engl J Med 352:154, 2005; NL Keating et al: J Clin Oncol 24:4448, 2006; JC Hu et al: Eur Urol 61:1119, 2012.)*

recently been added to the list of these complications **(Fig. 411-4)**. Treatment with GnRH agonists in men with prostate cancer is associated with rapid induction of insulin resistance, hyperinsulinemia, and a significant increase in the risk of incident diabetes. Metabolic syndrome is prevalent in >50% of men undergoing long-term ADT when compared to age-matched men with prostate cancer not undergoing ADT and their age-matched eugonadal counterparts. Some but not all studies have reported an increased risk of cardiovascular events, death due to cardiovascular events, and peripheral vascular disease in men undergoing ADT. Some reports suggest that men receiving ADT are at an increased risk of thromboembolic events and cognitive dysfunction. The rates of acute kidney injury are higher in men currently receiving ADT than in men not receiving ADT; the increased risk appears to be particularly associated with the use of combined regimens of a GnRH agonist plus an antiandrogen. ADT also is associated with substantially increased risk of osteoporosis and bone fractures.

APPROACH TO THE PATIENT

Men Receiving ADT

The benefits of ADT in treating nonmetastatic prostate cancer should be carefully weighed against the risks of ADT-induced adverse events **(Table 411-6)**. If ADT is medically indicated, consider whether intermittent ADT is a feasible option. Men being considered for ADT should undergo assessment of cardiovascular, diabetes, and fracture risk; this assessment may include measurement of blood glucose, plasma lipids, and bone mineral density by dual energy x-ray absorptiometry. Institute measures to prevent bone loss, including physical activity, adequate calcium and vitamin D intake, and pharmacologic therapy in men with a previous minimal trauma fracture and those with 10-year risk of a major osteoporotic fracture >20%, unless contraindicated. Bisphosphonates and denosumab have been shown to reduce fracture risk in men undergoing ADT, and zoledronic acid and denosumab have been approved by the U.S. Food and Drug Administration for the prevention of metastasis-related skeletal-related events in this population. Men with prostate cancer who are receiving ADT should be monitored for weight gain and diabetes. Encourage lifestyle interventions, including physical activity and exercise, and attention to weight, blood pressure, lipid profile, blood glucose, and smoking cessation to reduce the risk of cardiometabolic complications. Metformin plus lifestyle intervention can prevent and improve metabolic dysregulation. In randomized trials, gabapentin, medroxyprogesterone, and the serotonin reuptake inhibitor venlafaxine have been shown to be more efficacious than placebo in alleviating hot flushes. The side effects of these medications—increased appetite and weight gain with medroxyprogesterone, gynecomastia with estrogenic compounds, and dry mouth with venlafaxine—should be weighed against their relative efficacy. Acupuncture, soy products, vitamin E, herbal medicines, and transdermal estradiol have been used empirically for the treatment of vasomotor symptoms without clear evidence of efficacy. Gynecomastia can be

TABLE 411-6 Checklist for Men Undergoing Androgen Deprivation Therapy (ADT)

1. Weigh the risks and benefits of ADT and whether intermittent ADT is a feasible and safe option.
2. Perform a baseline assessment including fasting glucose, plasma lipids, blood pressure, bone mineral density, and FRAX score.
3. Optimize calcium and vitamin D intake, encourage structured physical activity and exercise, and consider pharmacologic therapy in men with a previous minimal trauma fracture and those with a 10-year risk of a major osteoporotic fracture >20%, unless contraindicated.
4. Monitor body weight, fasting glucose, plasma lipids, blood pressure, and bone mineral density, and encourage smoking cessation and physical exercise. Consider metformin in those with metabolic disorder.
5. In men who are receiving ADT and who experience bothersome hot flushes, as indicated by sleep disturbance or interference with work or activities of daily living, consider initial therapy with venlafaxine. If ineffective, add medroxyprogesterone acetate or gabapentin.
6. In men who experience painful breast enlargement, consider therapy with an estrogen receptor antagonist, such as tamoxifen.

prevented by the use of an antiestrogen, an aromatase inhibitor, or local radiation therapy; these therapies are effective in alleviating pain and tenderness but are less effective in reducing established gynecomastia. For long-standing gynecomastia that persists after cessation of ADT and is bothersome, mammoplasty is an effective treatment option. Intermittent ADT can reduce the frequency of adverse effects associated with ADT, but its long-term efficacy and safety need further investigation.

PREVENTION OF SEXUALLY TRANSMITTED DISEASES

Adolescent boys and young men aged 15–24 years; men who have sex with men, have multiple sex partners, have unprotected sex without condom, or have sex with sex workers; men who use illicit drugs; men who have a history of previous sexually transmitted infection (STI); and transgender men are at increased risk for STIs. STIs increase the risk of oropharyngeal and anogenital cancers, liver disease, pelvic pain, infertility, inadvertent transmission of infection to others, and emergency department visits and are a preventable cause of excess morbidity and mortality. HIV, hepatitis B and C infections, and syphilis can have additional disease-specific complications. The prevention and treatment of STIs are discussed in **Chap. 141**. Additionally, the Centers for Disease Control and Prevention (CDC) and U.S. Preventive Services Task Force (USPSTF) have published guidelines on the prevention, treatment, and pre- and postexposure prophylaxis of STIs. The approach to the prevention of STIs includes a structured risk assessment; counseling about safe sex practices including condom use; immunization of individuals at risk; diagnosis and treatment of infected individuals whether or not they are symptomatic; detection and treatment of sexual partners; and targeted sex education of adolescents and young men who are at high risk for STIs. The USPSTF recommends screening for HIV in all men aged 15–65 years, regardless of risk, and for hepatitis B virus and syphilis in men at increased risk. Because more than half of STIs occur in persons aged 15–24 years, the USPSTF also recommends behavioral counseling for all sexually active adolescents and adult men at increased risk of STIs to encourage condom use and other protective behaviors, including consideration of abstinence, reducing the number of sex partners, and avoidance of unsafe sex practices. Consistent and correct condom use is the most important method of preventing STIs. Effective immunizations are available against hepatitis B, human papillomavirus (HPV), and *Neisseria meningitides*. The CDC's Advisory Committee on Immunization Practices (ACIP) recommends universal hepatitis B immunization for all unvaccinated adults presenting to an STI clinic, all HIV-infected adults, and health workers. Although ACIP recommends HPV vaccination in males aged 9–21 years and in men aged 9–26 if they have sex with men or have an immunocompromising condition, recent data suggest that the prevalence of HPV and its complications continue to increase until middle age, and some experts recommend extending the age limit for HPV vaccination. Meningococcal vaccination is indicated for men who have sex with men from an area of outbreak and for all HIV-infected men.

Because men seeking care in men's health clinics often do so for sexual and urogenital problems, these visits offer opportunities for counseling, screening, and treatment of STIs and institution of immunization and other preventive measures for STIs.

SEX DIFFERENCES IN COVID-19 DISEASE OUTCOMES

The COVID-19 pandemic has highlighted sex differences in the susceptibility to respiratory viral infections. Men infected with SARS-CoV-2 virus are more likely to have a more severe disease, require mechanical ventilation, have disease complications, and die of the disease than women. Somewhat similar sex differences in morbidity and mortality have been reported for influenza infection. In the United States, the incidence and rates of hospitalization for influenza are higher in men than in women across all age groups. However, the sex-specific mortality rates associated with influenza vary substantially across countries and age groups. The sex differences in susceptibility to SARS-CoV-2 infection and morbidity have been attributed to behavioral factors, such as higher rates of smoking and alcohol use in men; biologic factors, such as higher rates of comorbid conditions in men than in women; sex differences in immune responses, including a poor T lymphocyte response to SARS-CoV-2 infection; and lower expression levels in men of X-linked genes that are involved in the innate detection of RNA viruses and that escape X inactivation in women, resulting in higher expression levels in women. Additionally, the expression of angiotensin-converting enzyme 2 (ACE2) and the cell surface transmembrane protease serine 2 (TMPRSS2), the two host proteins that facilitate the entry of SARS-CoV-2 into the alveolar cells, is regulated by androgens in subsets of lung epithelial cells, and it is possible that higher testosterone levels in men may contribute to increased susceptibility to the infection.

FURTHER READING

Abrams P et al: Evaluation and treatment of lower urinary tract symptoms in older men. J Urol 189:S93, 2013.

Ahmadi H et al: Androgen deprivation therapy: Evidence-based management of side effects. BJU Int 111:543, 2013.

Baggish A et al: Cardiovascular toxicity of illicit anabolic-androgenic steroid use. Circulation 135:1991, 2017.

Bhasin S et al: Testosterone therapy in men with hypogonadism: An Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 103:1715, 2018.

Bhasin S et al: The implications of reproductive aging for the health, vitality and economic welfare of human societies. J Clin Endocrinol Metab 104:3821, 2019.

Bhasin S: Testosterone replacement in aging men: An evidence-based patient-centric perspective. J Clin Invest 131:e146607, 2021.

Bhasin S et al: Prostate safety events during testosterone replacement therapy in men with hypogonadism: A randomized clinical trial. JAMA Netw Open 6:e2348692, 2023.

Centers for Disease Control and Prevention: National Vital Statistics System: Mortality tables. *https://www.cdc.gov/nchs/nvss/mortality_tables.htm*. Accessed March 17, 2024.

Lincoff AM et al: Cardiovascular safety of testosterone replacement therapy. N Engl J Med 389:107, 2023.

Nelson BS et al: Anabolic-androgenic steroid use is associated with psychopathy, risk-taking anger, and physical problems. Sci Rep 12:9133, 2022.

Pope HG Jr et al: Adverse health consequences of performance-enhancing drugs: An endocrine society scientific statement. Endocr Rev 35:341, 2014.

Rudd RA et al: Increases in drug and opioid-involved overdose deaths—United States, 2010–2015. MMWR Morb Mortal Wkly Rep 65:1445, 2016.

Ruth KS et al: Using human genetics to understand the disease impacts of testosterone in men and women. Nat Med 26:252, 2020.

Snyder PJ et al: Effects of testosterone treatment in older men. N Engl J Med 374:611, 2016.

U.S. Preventive Health Services Task Force. Final recommendation statement sexually transmitted infections: Behavioral counseling. *https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/sexually-transmitted-infections-behavioral-counseling1*. Accessed June 21, 2017.

Wittert G et al: Testosterone treatment to prevent or revert type 2 diabetes in men enrolled in a lifestyle programme (T4DM): A randomised, double-blind, placebo-controlled, 2-year, phase 3b trial. Lancet Diabetes Endocrinol 9:32, 2021.

Woolf SH, Schoomaker H: Life expectancy and mortality rates in the United States, 1959–2017. JAMA 322:1996, 2019.

Workowski KA et al: Sexually transmitted infections treatment guidelines, 2021. MMWR Recomm Rep 70:1, 2021.

412 Lesbian, Gay, Bisexual, Transgender, Queer or Questioning, Intersex, Asexual, and More (LGBTQIA+) Health

Baligh R. Yehia, Todd P. Levin

UNDERSTANDING LGBTQIA+ HEALTH DISPARITIES

The acceptance of the lesbian, gay, bisexual, transgender, queer or questioning, intersex, asexual, and more (LGBTQIA+) community has greatly increased over the past decade in certain communities and parts of the world. However, numerous studies highlight health disparities involving the care of LGBTQIA+ people. Lesbian and bisexual women are less likely to receive recommended preventive screenings such as breast, cervical, and colorectal cancer screenings. Among men who have sex with men, rates of human papillomavirus–associated anal cancers are 17 times higher than those of heterosexual men. In addition, gay and bisexual men accounted for 68% of all new HIV diagnoses in the United States in 2020 according to the Centers for Disease Control and Prevention, and they disproportionately contract sexually transmitted infections. In 2021, men who have sex with men accounted for 47% of primary and secondary syphilis infections in the United States where the sex of the sexual partner was known. Transgender individuals have a higher prevalence of HIV infection and suicide compared with other groups.

Research has found that LGBTQIA+ individuals are more likely to experience depression, anxiety, and alcohol and drug use than their counterparts. Most concerning are the rates of suicide attempts and ideation among the LGBTQIA+ community, particularly youth. In a 2021 survey of 35,000 LGBTQIA+ youth ages 13–24, 42% seriously considered attempting suicide in the last year, including more than half of transgender and nonbinary youth. Lesbian, gay, and bisexual (LGB) youth are four times more likely to attempt suicide than their heterosexual peers, and 61% of gender variant youth reported suicidal ideation at some point in their life. Additionally, the recent U.S. Transgender Survey found that 40% of transgender young adults and adults reported attempting suicide at some point in their lives.

U.S. studies indicate that substance abuse is twice as common in LGBTQIA+ youth compared with their counterparts. These findings are mirrored among LGBTQIA+ adults: the prevalence of substance abuse disorders is 20–30% compared with ~9% in the general population.

These health issues are compounded by structural barriers to health care, including decreased access to medical care, lack of awareness to the unique health needs of LGBTQIA+ individuals, and stigma and discrimination toward the LGBTQIA+ community. Many LGBTQIA+ individuals perceive the health care setting and providers as threatening, which may lead to avoiding needed medical care or withholding important medical information. A large U.S. survey identified that 8% of LGB and 27% of transgender individuals were refused needed health care, and almost 11% of LGB and 21% of transgender people reported being subjected to harsh or abusive language by health care professionals. Apart from health care settings, more than two-thirds of LGB people report discrimination in their personal lives, and 90% of transgender individuals report harassment, mistreatment, or discrimination at work. Chronic exposure to high levels of stress from real or anticipated discrimination, referred to as "minority stress," may be an important factor contributing to the poor health outcomes experienced by LGBT populations.

While some research on LGBTQIA+ health has been conducted, there remains a great opportunity to better understand the needs and experiences of LGBTQIA+ individuals. Moreover, many LGBTQIA+ individuals experience health disparities across their life cycle (e.g., LGBTQIA+ youth are at greater risk of suicide and homelessness, whereas elderly LGBTQIA+ individuals face barriers to health because of isolation and fewer family supports), necessitating a longitudinal approach to examining LGBTQIA+ health issues. There are more limited data on the health of LGBTQIA+ individuals outside the United States and Europe. However, studies demonstrate that problems are greatest when people cannot be open about their sexual orientation and gender identity. Encouraging greater LGBTQIA+ acceptance and access to health care will be critical to improving outcomes and experiences for LGBTQIA+ communities.

CREATING POSITIVE HEALTH EXPERIENCES FOR LGBTQIA+ PATIENTS

Understanding Gender Identity and Sexual Orientation

Addressing health disparities and creating positive health care experiences require an understanding of the diversity of cultural expression and lives of LGBTQIA+ persons. First, providers must be able to distinguish *gender identity* from *sexual orientation*. Gender identity is a person's internal sense of their gender. It should not be confused with sex assigned at birth, which is based on anatomy and biology. Gender identity expands beyond the binary male and female and includes persons who think of their gender as containing elements of both, neither, or another gender. Many individuals who do not identify with the gender that correlates with their sex assigned at birth often use the terms *transgender* or *trans-male* or *trans-female* to identify themselves. Sexual orientation refers to how one thinks of their physical or emotional attraction to others. Sexual orientation has three dimensions: attraction, behavior, and identity. *Attraction* refers to one's desire to be with someone, regardless of one's behavior or stated identity. For example, a woman may be attracted to another woman, but this attraction may never be acted upon and may not form part of her sexual identity. *Behavior* refers to a person's interactions with sexual and romantic partners. Although sexual identity often aligns with behavior, some individuals who identify as heterosexual may have same-gender partners and some individuals who identify as lesbian or gay may have different-gender partners. Lastly, *identity* refers to how a person defines their own sexuality. Common terms for sexual identity include *gay*, *lesbian*, *bisexual*, *straight*, *heterosexual*, *homosexual*, and *asexual* **(Tables 412-1 and 412-2)**. As individuals go through the process of understanding their sexuality and self-identity over time, they may change how they define their sexual identity.

The creation of a welcoming environment requires not making any assumptions about an individual's gender identity or sexual orientation. Both front-line staff and clinicians should be cognizant of patient communication. For example:

- *Instead of saying* "How may I help you, sir?", *say* "How may I help you?"
- *Instead of saying* "She is here for her appointment," *say* "The patient is here in the waiting room."
- *Instead of saying* "Do you have a wife?", *say* "Are you in a relationship?"
- *Instead of saying* "What are your mother's and father's names?", *say* "What are your parents' names?"

Developing Comfort and Competency in Sexual Health Developing comfort discussing sexual health and intimacy is critical to providing appropriate care. After inquiring about healthy relationships and relationship status, a good starting place is to ask if a patient is sexually active and, if so, with whom, how often, and what types of physical interactions and types of sex they have with their partner(s). These discussions can allow clinicians to focus subsequent conversations on issues most relevant to a patient's health. For example, a gay man with multiple sexual partners who engages in receptive anal sex without condoms is at high risk for HIV and sexually transmitted infections (STIs). It will be important to recommend more frequent screenings for STIs and discuss the use of HIV preexposure prophylaxis (PrEP), along with postexposure prophylaxis (PEP), and condoms

TABLE 412-1 Common LGBTQIA+ Terminology and Definitions

TERM	DEFINITION
Agender	Identifying as having no gender, or does not experience gender as a primary identity component.
Asexual	Experiencing little or no sexual attraction to others. Asexual people may still engage in sexual activity.
Assigned female at birth/ assigned male at birth	Refers to the sex that is assigned to an infant, most often based on the infant's anatomic and other biological characteristics. Commonly abbreviated as AFAB (assigned female at birth) or AMAB (assigned male at birth).
Bisexual	A sexual orientation that describes a person who is emotionally and physically attracted to more than one gender.
Cisgender	A person whose gender identity is consistent in a traditional sense with their sex assigned at birth; for example, a person assigned female sex at birth whose gender identity is woman/female. The term cisgender comes from the Latin prefix cis, meaning "on the same side of."
Cisnormativity	Cisnormativity is the assumption that everyone is cisgender, or that being cisgender is the only "normal" gender identity. Cisnormativity puts pressure on people to conform to binary notions of gender expression and gender identity. Cisnormativity can manifest as cissexism, the biased belief that cisgender identity is the superior gender identity. Individuals, communities, policies, and institutions can perpetuate cisnormativity and cissexism, and thus marginalize transgender and gender diverse people and communities.
Gay	A sexual orientation describing people who are primarily emotionally and physically attracted to people of the same gender identity as themselves. Commonly used as an umbrella term for "not straight."
Gender-affirming hormone therapy (GAHT)	Feminizing and masculinizing hormone treatment to align secondary sex characteristics with gender identity.
Gender-affirming surgery (GAS)	Surgeries to modify a person's body to be more aligned with that person's gender identity. Types of GAS include chest and genital surgeries, facial feminization, body sculpting, and hair removal.
Gender-affirming chest surgery	Surgeries to remove and/or construct a person's chest to be more aligned with that person's gender identity. Also referred to as top surgery. Types of chest surgeries include: • Feminizing breast surgery: breast augmentation, chest construction, or breast mammoplasty • Masculinizing chest surgery: mastectomy (removal of breast tissue) and chest contouring
Gender-affirming genital surgeries	Surgeries that help align a person's genitals and/or internal reproductive organs with that person's gender identity, including: • Clitoroplasty (creation of a clitoris) • Hysterectomy (removal of the uterus; may also include removal of the cervix, ovaries, and fallopian tubes) • Labiaplasty (creation of inner and outer labia) • Metoidioplasty (creation of a masculine phallus using testosterone-enlarged clitoral tissue) • Oophorectomy (removal of ovaries) • Orchiectomy (removal of testicles) • Penectomy (removal of the penis) • Phalloplasty (creation of a phallus) • Scrotoplasty (creation of a scrotum and often paired with testicular implants) • Urethral lengthening (to allow voiding while standing) • Vaginectomy (removal of the vagina) • Vaginoplasty (creation of a neo-vagina) • Vulvoplasty (creation of a vulva)
Gender dysphoria	Distress experienced by some people whose gender identity does not align with their sex assigned at birth based on societal expectations; or distress experienced when a person's gender identity and/or gender expression is not affirmed. The degree and severity of gender dysphoria is highly variable among transgender and gender diverse people.
Gender expression	The ways in which people present their gender identity to others, such as through gait, mannerisms, voice, dress, and hairstyle. Cultural and historical contexts influence how people interpret and express gender identity.
Gender identity	A person's inner sense of being a girl/woman/female, boy/man/male, both, neither, or beyond the gender binary.
Gender queer	Describes people whose gender identity goes beyond the gender binary.
Heterosexual	A sexual orientation that describes women who are emotionally and sexually attracted to men, and men who are emotionally and sexually attracted to women. Also referred to as straight.
Heteronormativity	Heteronormativity is the assumption that everyone is heterosexual, or that being heterosexual is the only "normal" sexual orientation. Heteronormative social systems incentivize conforming to the behavior and appearance of heterosexuality. Heteronormativity can manifest as heterosexism, the biased belief that heterosexual identity is the only valid sexual orientation. Individuals, communities, policies, and institutions can perpetuate heteronormativity and heterosexism, and thus marginalize queer people and communities.
Intersex	Describes a group of congenital conditions in which the reproductive organs, genitals, and/or other sexual anatomy do not develop according to traditional expectations for females or males. Intersex can also be used as an identity term for someone with one of these conditions. The medical community sometimes uses the term differences of sex development (DSD) to describe intersex conditions; however, the term intersex is recommended by several intersex community members and groups **(Chap. 402)**.
Lesbian	A sexual orientation describing a woman or non-binary person who is primarily emotionally and physically attracted to women.
Men who have sex with men (MSM)/women who have sex with women (WSW)	Categories used in research and public health to describe those who engage in same-sex sexual behavior, regardless of their sexual orientation. People rarely use the terms MSM or WSW to describe themselves.
Pangender	Describes a person who has multiple gender identities.
Pansexual	A sexual orientation describing a person who can be emotionally and/or physically attracted to people of any gender identity or gender expression.
Pronouns	Pronouns are words used in place of a person's name. Examples of pronouns are she/her/hers, they/them/theirs, he/him/his. Neopronouns are new pronouns that people may use to affirm their gender identity; examples include: xe/xem/xyr and ze/hir/hirs. Some people use their name instead of pronouns.

(*Continued*)

TABLE 412-1 Common LGBTQIA+ Terminology and Definitions (Continued)

TERM	DEFINITION
Queer	An umbrella term used by some to describe people who think of their sexual orientation or gender identity as outside of societal norms. Some people view the term as more fluid and inclusive than traditional categories for sexual orientation and gender identity. Although queer was historically used as a slur, it has been reclaimed by many as a term of empowerment. Nonetheless, some still find the term offensive.
Questioning	Describes an individual who is unsure about or is exploring their own sexual orientation and/or gender identity.
Same-sex attraction (SSA)	Describes the experience of a person who is emotionally and/or physically attracted to people of the same gender. Used most commonly by people who live in religious communities that are not accepting of LGBTQIA+ identities.
Sexual orientation	How a person experiences their physical, emotional, and romantic attachments to other people.
Trans man/transgender man	A transgender person whose gender identity is boy/man/male may use these terms to describe themselves. Some will use the term man.
Trans woman/transgender woman	A transgender person whose gender identity is girl/woman/female may use these terms to describe themselves. Some will use the term woman.
Transgender	An umbrella term for people whose gender identity does not align with the sex they were assigned at birth, based on societal expectations.
Two-spirit	Describes a person who embodies both a masculine and a feminine spirit. This is a culture-specific term used among some Native American, American Indian, and First Nations people. Abbreviated as 2S.

Note: It is important to note that definitions vary across communities, that they change over time, and that not all LGBT people agree with all these definitions.

Source: From *https://www.lgbtqiahealtheducation.org/glossary/en/.*

to prevent HIV and STIs. Additionally, if you are seeing a transgender man, it will be important to know if he still has natal female genitalia to ensure appropriate cancer screening, desire for having biologic children, and contraceptive needs. Notably, many if not most transgender people have not had gender affirmation surgery and retain their natal sex organs.

Creating a Welcoming and Safe Health Care Environment Hospitals and clinics can take a number of steps to create a welcoming and safe space for LGBTQIA+ patients. This starts by establishing and communicating a nondiscrimination policy that clearly includes gender identity, gender expression, and sexual orientation protections. Additionally, hospitals and clinics can develop and implement an equal visitation policy to ensure equal visitation for LGBTQIA+ patients from same-sex partners, parents, and other family and friends. Staff training in LGBTQIA+ patient-centered care also is a key component of creating inclusive health environments. This includes covering LGBTQIA+ cultural competency, caring for LGBTQIA+ patients, creating an inclusive environment for LGBTQIA+ patients and staff, and other topics important for LGBTQIA+ health.

TABLE 412-2 Outdated and Insensitive Terms to Replace

OUTDATED TERM	RECOMMENDED TERM
Berdache	Two-spirit
Biological female/male	Assigned female/male at birth
Cross-sex hormone therapy; hormone replacement therapy	Gender-affirming hormone therapy
Disorders/differences of sex development	Intersex
Female-to-male (FTM) and male-to-female (MTF)	Transgender man and transgender woman
Gender nonconforming	Gender nonbinary
Hermaphrodite/ambiguous genitalia	Intersex
Homosexual	Gay or lesbian
Legal name	Administrative name or name on legal documents
Preferred name	Chosen name or name used
Preferred pronouns	Pronouns
Sex change/sex reassignment surgery/gender reconstruction surgery	Gender-affirming surgery
Sexual preference/lifestyle	Sexual orientation
Transgendered	Transgender

Note: Most people consider the listed terms outdated and insensitive. We recommend replacing these words with the suggested terms provided.

Source: Reproduced with permission from LGBTQIA+ Glossary of Terms for Health Care Teams 2020. *https://www.lgbtqiahealtheducation.org/publication/lgbtqia-glossary-of-terms-for-health-care-teams/.*

As hospitals and clinics continue to adopt electronic health records, collecting sexual orientation and gender identity information becomes increasingly important to delivering personalized care to LGBTQIA+ individuals. It allows providers to monitor quality of care, track population-based outcomes, and collect more accurate information for research purposes. Most of this information can be captured by three questions:

- *How do you think of yourself?* As straight or heterosexual; lesbian, gay, or homosexual; bisexual; something else; don't know; choose not to disclose.
- *What is your current gender identity?* Male; female; transgender male/trans man; transgender female/trans woman; genderqueer, neither exclusively male nor female; additional category, please specify; choose not to disclose.
- *What sex were you assigned at birth on your original birth certificate?* Male; female; choose not to disclose.

The physical environment of a hospital or clinic is important, but the majority of clinical spaces do not signal that they are safe spaces for LGBTQIA+ patients. Most health care posters, pamphlets, and materials feature heterosexual individuals or couples; adding LGBTQIA+-friendly images and text can help signal that the hospital or clinic is a safe space for sexual and gender minorities. In addition, easily identifying LGBTQIA+-competent providers by using websites, buttons, and pins can help patients select providers and feel at ease when attending appointments. Lastly, designating all-gender bathrooms is important to creating welcoming spaces, particularly for transgender and gender-nonconforming individuals.

FUTURE DIRECTION IN LGBT HEALTH

While social and cultural acceptance of the LGBTQIA+ community has improved in certain parts of the world, many LGBTQIA+ individuals continue to experience discrimination, stigmatization, and violence. Inequitable health care policies and practices, lack of awareness of LGBTQIA+ health issues, and limited understanding of the unique health needs of LGBTQIA+ individuals contribute to decreased access to care and disparities in health outcomes for LGBTQIA+ individuals. Addressing these barriers will require improved data collection on the LGBTQIA+ population; understanding of the intersectionality of gender identity, sexual orientation, race/ethnicity, and other sociocultural determinants of health; and outcomes-focused research across the life course. In striving to deliver high-quality care experiences for all patients, hospitals, clinics, and providers need to focus on meeting the needs of the LGBTQIA+ community.

FURTHER READING

Centers for Disease Control and Prevention: Lesbian, Gay, Bisexual, and Transgender Health. 2018. Available at *www.cdc.gov/lgbthealth/*. Accessed December 10, 2019.

Fenway Health: Glossary of LGBT terms for health care teams. February 2024. Available at *https://www.lgbtqiahealtheducation.org/publication/lgbtqia-glossary-of-terms-for-health-care-teams/*.

Institute of Medicine: The health of lesbian, gay, bisexual, and transgender (LGBT) people: Building a foundation for better understanding. 2011. Available at *www.nap.edu/catalog.php?record_id=13128*.

Institute of Medicine: Collecting sexual orientation and gender identity data in electronic health records: Workshop summary. 2013. Available at *https://www.nap.edu/catalog/18260/collecting-sexual-orientation-and-gender-identity-data-in-electronic-health-records*.

Joint Commission: Advancing effective communication, cultural competence, patient- and family-centered care for the lesbian, gay, bisexual, and transgender community: A field guide. 2011. Available at *www.jointcommission.org/lgbt*.

National Center for Transgender Equality: The report of the 2015 U.S. Transgender Survey. 2015. Available at *https://transequality.org/sites/default/files/docs/usts/USTS-Full-Report-Dec17.pdf*.

Safer JD, Tangpricha V: Care of transgender persons. N Engl J Med 381:2451, 2019.

Substance Abuse and Mental Health Services Administration: Top health issues for LGBT populations information & resource kit. 2012. Available at *https://store.samhsa.gov/product/top-health-issues-lgbt-populations/sma12-4684*.

Section 3 Obesity, Diabetes Mellitus, and Metabolic Syndrome

413

Pathobiology of Obesity

Stephen O'Rahilly, I. Sadaf Farooqi

Adipose tissue evolved as a solution to the challenge of the intermittent availability of food. At times when food is plentiful, excess calories are converted to triglycerides and efficiently stored in the unilocular lipid droplets that occupy most of the volume of fat cells. When needed, the triglyceride is rapidly broken down to free fatty acids and glycerol, which provide an energy source to other sites throughout the body. However, in environments where food is abundant and when individuals tend to be sedentary, the chronic excess of energy intake over expenditure leads to obesity. The risks of developing obesity under those circumstances and of developing the illnesses associated with obesity vary greatly between individuals, with that variation having a strong genetic basis.

DEFINITION OF OBESITY AND OVERWEIGHT

Obesity is defined as a state of excess adipose tissue mass that adversely affects health. The direct measurement of fat mass is not something that is readily undertaken in routine clinical practice, so a proxy measure, the body mass index (BMI), is generally used. This is calculated as weight/height^2 (in kg/m^2) **(Fig. 413-1)**. BMI-based definitions of obesity and overweight have been established based on associations with certain morbidities and excess mortality. These definitions have been based largely on studies of predominantly white, Western populations, and there is growing evidence that the relationship between BMI and adverse outcomes is different in people from other ethnic groups, usually in the direction of worse health outcomes being seen at lower levels of BMI. The World Health Organization (WHO) defines a BMI of 30 kg/m^2 as the cutoff point for obesity, while individuals with values between 25 and 30 kg/m^2 are classified as overweight. For individuals with a very muscular body habitus, the BMI may overestimate the amount of body fat. For any given BMI, women will generally have a higher percentage of body fat than men.

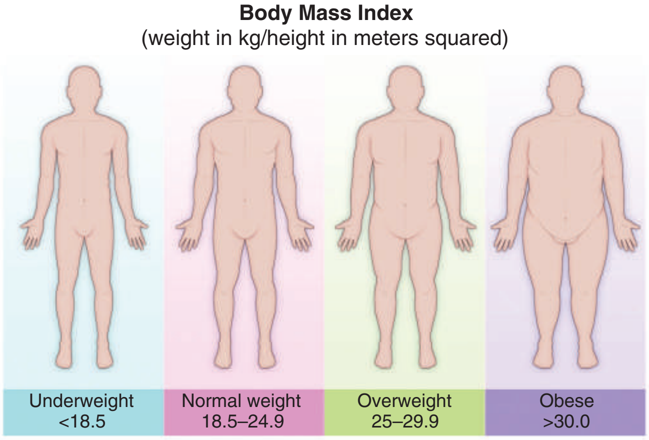

FIGURE 413-1 Definitions of overweight and obesity. The World Health Organization defines obesity based on body mass index (BMI), which is calculated as weight in kilograms divided by the height in meters squared.

The extent to which different adipose depots expand in response to chronic overnutrition varies markedly between people. In general, females store more fat in subcutaneous tissues, especially on buttocks, thighs, and upper arms, whereas men are more prone to store fat in intraabdominal and truncal subcutaneous sites. A simple measure of fat distribution is provided by a measurement of the waist-to-hip ratio. Independent of the degree of obesity, a waist-to-hip ratio >0.9 in women and >1.0 in men is associated with adverse health outcomes such as type 2 diabetes and dyslipidemia.

EPIDEMIOLOGY

The annual National Health and Nutrition Examination Survey (NHANES) provides an ongoing record of the prevalence of obesity in the United States. In 2017–2018, 42.4% of U.S. adults aged ≥20 years old had obesity with no significant differences in prevalence by age group. Non-Hispanic black people had the highest prevalence of obesity at 49.6%, followed by Hispanic (44.8%), non-Hispanic white (42.2%), and non-Hispanic Asian (17.4%) people. In the United States, Asians represent a highly heterogeneous group encompassing both East and South Asia as well as a substantial Filipino community. The risks of obesity and its complications may differ greatly between people from different parts of Asia; in general, the prevalence of obesity is somewhat higher in women than in men, with black women having the highest prevalence at 56.9%. There has been a marked increase in the prevalence of obesity over time. For example, between 1976 and 1980, the NHANES survey reported a prevalence of 14.5%, indicating a near threefold increase over the past 40 years.

This trend is seen globally. According to the WHO, obesity has nearly tripled worldwide since 1975. In 2016, >1.9 billion adults aged ≥18 years old were overweight. Of these, >650 million were obese; 39% of adults aged ≥18 years old were overweight in 2016, and 13% were obese. Most of the world's population lives in countries where overweight and obesity kills more people than underweight.

During this time, one of the most striking changes has been in the prevalence of obesity in children. In children, the relationship between BMI and body fat varies considerably with age and with pubertal maturation; however, when adjusted for age and sex, BMI is a reasonable proxy for fat mass. Using age- and sex-specific BMI cutoffs (overweight ≥91st percentile; obesity ≥99th percentile), in 2019, the WHO estimated that 38 million children under the age of 5 were overweight or obese, and in 2016, they reported that 340 million children and adolescents aged 5–19 were overweight or obese.

PHYSIOLOGIC REGULATION OF ENERGY BALANCE

Discussions about obesity so frequently focus on the issues of personal choice or the obesogenic environment that it can be easy to forget that the amount of stored energy in our bodies is subject to homeostatic control by fundamental physiologic processes essential to our survival. In the 1940s, it was demonstrated that rodents defend their level of body fat; once returned to ad libitum diets after a short period of enforced caloric restriction or excess, animals either overconsumed or underconsumed calories until they returned to their previous status. Since that time, research has progressively dissected the signals that sense nutrient stores and the contents of our diets and how this information is integrated to control hunger, satiety, and the expenditure of energy. The key locus for the integration of these signals is the hypothalamus, an area of the brain at least partially outside the blood-brain barrier that facilitates its ability to receive hormonal signals and combine these with sensory, cognitive, and other neural inputs.

The hypothalamus receives multiple hormonal signals relevant to energy balance **(Fig. 413-2)**. The circulating concentration of leptin, a peptide hormone produced by fat cells, increases as fat stores increase and declines as fat stores are depleted. Importantly, under conditions of caloric restriction, circulating leptin levels fall faster than the disappearance of fat. Humans born without functional leptin or leptin receptors, although normal weight at birth, develop severe obesity from an early age, largely as a result of an intense drive to eat (hyperphagia). Clearly, a reduction of leptin below normal level is a powerful stimulus to food intake and largely explains the rebound overeating and weight regain that occurs after a period of starvation or dieting. The hypothalamus also receives hormonal signals that are more immediately related to the amount and type of food that has been ingested. Peripheral hormones such as cholecystokinin (CCK) from the stomach, glucagon-like peptide 1 (GLP-1) and gastric inhibitory polypeptide (GIP) from enteroendocrine cells of the small intestine, and peptide YY (PYY) and oxyntomodulin from the large intestine are secreted in response to eating a meal and/or the presence of nutrients in the intestinal lumen. Their release together with neural signals from the vagus nerve and the enteric nervous system contributes to satiety, often indirectly acting on the hypothalamus via projections from the brainstem. Insulin and amylin, produced by the pancreas in response to carbohydrate and protein-rich meals, also have effects on neurons controlling energy balance.

The propeptide pro-opiomelanocortin (POMC) is expressed in a highly restricted population of hypothalamic neurons that project widely throughout the brain **(Fig. 413-3)**. These neurons are responsive to the endocrine signals described above and are critical to the regulation of energy balance. The POMC-derived peptides α- and β-melanocyte-stimulating hormone (MSH) act on the melanocortin 4 receptor (MC4R) to regulate both food intake and aspects of energy expenditure that are influenced by the sympathetic nervous system. γ-MSH, acting mostly through the MC3 receptor, appears to play more of a role in controlling linear growth and the disposition of nutrients into lean versus fat tissues. Signaling through both these melanocortin receptors is also subject to negative control by a different population of neurons, which make and release agouti-related peptide (AGRP), neuropeptide Y (NPY), and the inhibitory neurotransmitter γ-aminobutyric acid (GABA). AGRP actively switches off melanocortin receptors. Leptin, which suppresses food intake, simultaneously stimulates POMC neurons and inhibits NPY/AGRP neurons. Human energy balance is highly sensitive to signaling through this system as people who have a genetic defect in just one of the two copies of the *MC4R* gene are very prone to overeat (hyperphagia) and to gain weight.

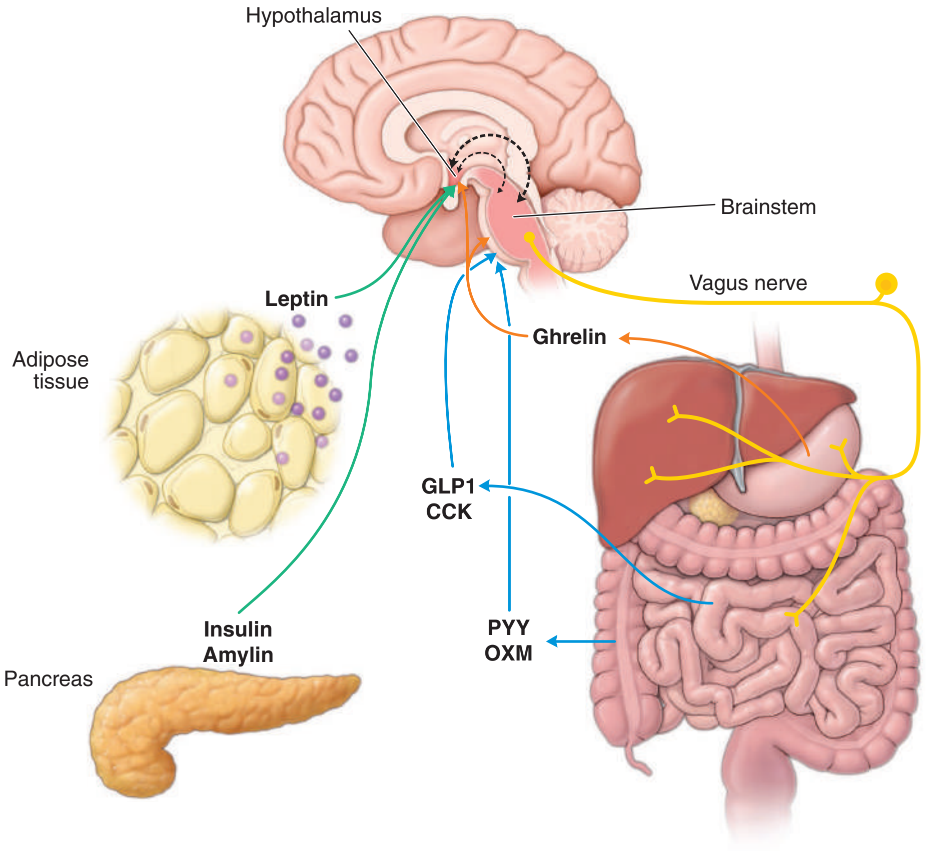

FIGURE 413-2 The homeostatic regulation of body weight. In most people, body weight remains stable over long periods of time despite fluctuations in the amount of food we eat and the amount of activity we undertake. This homeostatic regulation of body weight is controlled by the neurons in the hypothalamus, which receive hormonal signals from adipose tissue such as leptin and neural and hormonal signals from the gut in response to meals. Glucagon-like peptide 1 (GLP1) and cholecystokinin (CCK) from enteroendocrine cells of the small intestine and peptide YY (PYY) and oxyntomodulin (OXM) from the large intestine are secreted in response to eating a meal and/or the presence of nutrients in the intestinal lumen. Their release, together with neural signals from the vagus nerve and the enteric nervous system, contributes to satiety, acting on the hypothalamus via projections from the brainstem. Insulin, produced by the pancreas in response to carbohydrate- and protein-rich meals and potentiated by the action of some of the gut hormones, also has effects on the hypothalamic neurons controlling energy balance, whereas amylin acts predominantly via the brainstem. The release of the hormone ghrelin from the stomach increases in the unfed state and induces appetite by acting on hypothalamic neurons as well as on receptors in the brainstem.

■ THE PHYSIOLOGY OF NUTRIENT STORAGE IN ADIPOSE TISSUE

When energy intake exceeds energy expenditure, a small amount of that excess energy is stored as glycogen in liver and skeletal muscle. But if the imbalance is greater, then our bodies are designed to store that excess energy in a more efficient way as triacylglycerol (triglyceride). This fat is more efficient because, unlike glycogen, it does not need accompanying water, and when metabolized, it generates more than twice as much energy per gram than does carbohydrate. Adipocytes (fat cells) have evolved to contain a highly specialized organelle, the unilocular fat droplet, which holds the triglyceride within a single-layer of phospholipid that contains all the components needed for enzymes that make and breakdown triglycerides in a manner that is rapidly responsive to metabolic requirements. No other type of cell is specifically designed to store fat safely in this manner, and many of the adverse consequences of obesity are likely caused not by having too much fat in adipocytes but by "nonprofessional" cells being forced to take up and store fat. During weight gain, the amount of lipid in each fat cell increases. Some new fat cells can also be made in adulthood when ~10% of our fat cell population turns over every year.

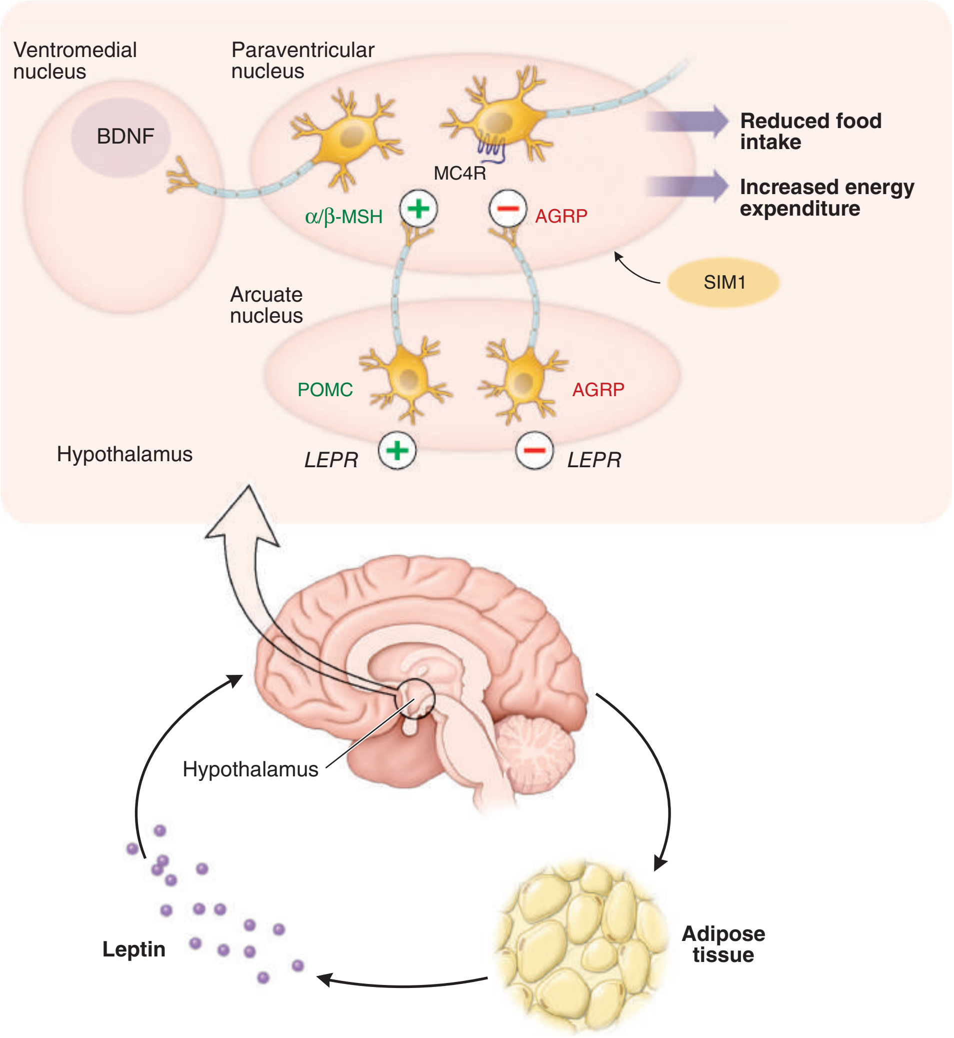

FIGURE 413-3 Hypothalamic pathways regulating body weight. Neurons in the hypothalamus regulate energy intake and expenditure in response to leptin and other hormones. In the fed state, leptin stimulates primary neurons in the arcuate nucleus of the hypothalamus that express pro-opiomelanocortin (POMC). The POMC-derived peptides α- and β-melanocyte-stimulating hormone (MSH) act on the melanocortin 4 receptor (MC4R) expressed on neurons in the paraventricular nucleus to reduce energy intake and increase energy expenditure. At the same time, leptin inhibits neurons expressing agouti-related peptide (AGRP), which switches off melanocortin receptors. When these and other key molecules, such as brain-derived neurotrophic factor (BDNF) and single minded-1 (SIM1), are disrupted by inherited mutations, affected individuals have hyperphagia and severe obesity.

THE CAUSES OF OBESITY: AN INTERACTION OF GENES AND ENVIRONMENT

For a person to develop obesity, energy intake must exceed energy expenditure in a manner that is sufficiently sustained to result in the accumulation of a large excess of triglyceride in adipose tissue. As obesity is a cumulative pathology, if energy intake exceeds energy expenditure by even a small amount (as little as 7 kcal/d), this is sufficient to develop obesity over a matter of years or decades. Even where obesity is common, there are many people who are not overweight. Economic and social factors are likely to play a role as there are more normal-weight people in wealthier and more socially advantaged groups, at least in Western societies. It is also true, however, that because of discrimination, people with obesity may become socially and economically disadvantaged, which complicates interpretation of that data. We can, however, state with considerable certainty that genetic factors play a major role in predisposing people to a range of adiposity. We know this from a large number of studies comparing identical and nonidentical twins. It is particularly telling that the degree of adiposity in adult life of identical twins brought up in different families is very similar between the twins but is not at all correlated with that of the adoptive siblings with whom they were raised.

THE RELATIVE ROLES OF EXCESS INTAKE AND LOWER ENERGY EXPENDITURE IN CONFERRING BIOLOGIC PREDISPOSITION

Do these heritable factors influence energy intake, energy expenditure, or both? It is clear that by the time a person develops obesity the amount of energy they expend in the resting state is more, not less, than a normal weight person. However, if a person with obesity loses weight by dieting, there is some evidence that they tend to be more “energy efficient” than a person who has never been obese, particularly in terms of how many calories they burn during a defined bout of activity. However, the effects are subtle. It seems very likely that there are some individuals who are predominantly predisposed to develop obesity by virtue of a lower metabolic rate, but thus far, apart from severe hypothyroidism, concrete examples are scarce. In contrast, a much more consistent and compelling body of evidence supports the idea that the genetic predisposition to obesity is largely mediated through the brain’s control of food intake. When studied in controlled settings, individuals who carry genetic variants that predispose to obesity tend to eat more and be less readily satiated. This is very readily demonstrable when the mutation has a major effect on obesity predisposition, but similar data are now emerging in the case of common genetic variants with smaller effects.

ENVIRONMENTAL FACTORS PREDISPOSING TO OBESITY

Obesity cannot exist in the absence of sufficient food to lay down and maintain excess fat stores. That fact not infrequently leads to the belief that the principal cause of obesity must be either a person’s ignorance of the role of excess caloric intake or their conscious choice to prioritize the immediate pleasures of eating over the long-term health harms associated with obesity. Taken to extremes, these views can engender serious social, economic, and medical discrimination against people with obesity. It is clear that genetic factors, however important they are in an individual’s predisposition to obesity, cannot explain the marked increase in obesity prevalence that has occurred in the past few decades. We have to look to an environment that has become increasingly obesogenic to explain that phenomenon. In most developed and developing countries, energy-dense and highly palatable food and beverages have been aggressively marketed, made cheaper than ever before, provided in larger portions, and made available ubiquitously and continuously. This has been combined with the reduction in physical activity in work and domestic life due to mechanization and the change in the nature of employment. Even the control of our external temperature by artificial heating and cooling has meant less energy expended on thermoregulation. Taken together, these are likely to be the major factors driving the recent increase in obesity. It is important to remember, however, that a substantial proportion of the population remains normal weight under these circumstances and a large part of that is attributable to their genetic good fortune.

There is much current investigation into other environmental factors that might influence the development of obesity. Heated debates continue about the optimal balance of macronutrients in the diet to maintain normal weight and good health. Much of this revolves around the potential benefits of reducing the relative proportion of carbohydrates in the diet **(Chap. 414)**. There seems to be reasonable consensus that, in the short term, diets that are rich in protein and fat and lower in carbohydrates more readily result in quick weight loss. This may be because the appetite-suppressing gut hormones discussed above increase more in response to protein than to carbohydrate, thus inducing earlier satiation. However, longer-term studies to date are less compelling, and the long-term increases in protein and fat intake are not without at least theoretical risks. A growing body of evidence suggests that exposures early in life, either in utero or in early postnatal life, might "program" individuals to develop obesity and/or cardiometabolic disease through effects that are often attributed to "epigenetics" **(Chap. 497)**. This is an attractive idea, and if true, it would mean that time-limited and affordable interventions early in life might have lifelong benefits. Inevitably, it will take time to see if the promise of such interventions will be fulfilled. Much excitement has been generated by the increasing recognition of the diversity of our intestinal microbiome, which clearly has relevance to gastrointestinal health **(Chap. 484)**. At present, it is premature to ascribe any significant role to the human microbiome in obesity or its adverse consequences.

WHY DOESN'T LEPTIN PREVENT OBESITY?

Leptin is known to suppress food intake, and its levels rise as fat stores expand. So why does this not prevent us from developing obesity? The most plausible explanation lies in the evolutionary history of leptin and the fact that it appears to defend strongly against the loss of body fat stores, with a fall in circulating leptin below a person's habitual level being a powerful stimulus to food intake, whereas the response to rises in leptin above the normal level is less pronounced. At higher levels of leptin, administering extra amounts of the hormone may have no discernible effect at all—a phenomenon that has come to be called leptin resistance. It is important to remember that even though a person appears to be leptin resistant, some leptin action is occurring; otherwise, the person would become as insatiably hungry and progressively obese as someone with congenital leptin deficiency (see below). It also seems likely that a subgroup of people may have relatively low leptin levels, which plays a role in the etiology of their obesity. There are likely other hormonal signals produced in severe obesity that, unlike leptin, continue to exert a suppressive effect on food intake and help to ensure that the expansion of adipose tissue does not become indefinitely cumulative.

SINGLE-GENE DISORDERS LEADING TO OBESITY

The assessment of severely obese children and, indeed, adults should be directed at screening for potentially treatable endocrine and neurologic conditions and identifying genetic conditions so that appropriate genetic counseling and, in some cases, treatment can be started. Clinically, it remains useful to categorize the genetic obesity syndromes as those with dysmorphism and/or developmental delay and those without these features **(Tables 413-1 and 413-2)**. Although individually these monogenic disorders are rare, cumulatively, up to 20% of children with severe obesity have rare chromosomal abnormalities and/or highly penetrant genetic mutations that drive their obesity. This figure is likely to increase with wider accessibility to genetic testing and as new genes are identified. A genetic diagnosis can inform management (many such patients find it very difficult to lose weight through diet and exercise) and can inform clinical decision-making regarding the use of bariatric surgery (feasible in some; high risk in others) **(Chap. 414)**. There are a number of drugs in clinical trials targeted specifically at patients with genetic obesity syndromes. Specifically, setmelanotide, a MC4R agonist, has been used effectively in phase 2/3 clinical trials in children who are genetically deficient in POMC, PCSK1, and the leptin receptor. It is also being explored for the treatment of other genetic obesity syndromes affecting the melanocortin pathway and in acquired hypothalamic obesity caused by tumors such as craniopharyngiomas.

TABLE 413-1 Classical Genetic Obesity Syndromes

SYNDROME	INHERITANCE	ADDITIONAL CLINICAL FEATURES
Prader-Willi	Autosomal dominant	Hypotonia, failure to thrive in infancy, developmental delay, short stature, hypogonadotropic hypogonadism, sleep disturbance, obsessive behavior
Albright's hereditary osteodystrophy or pseudohypoparathyroidism	Autosomal dominant	Short stature in some, skeletal defects, developmental delay, shortened metacarpals; hormone resistance when mutation on maternally inherited allele
Bardet-Biedl	Autosomal recessive	Syndactyly/brachydactyly/polydactyly, developmental delay, retinal dystrophy or pigmentary retinopathy, hypogonadism, renal abnormalities
Cohen's	Autosomal recessive	Facial dysmorphism, microcephaly, hypotonia, developmental delay, retinopathy
Carpenter's	Autosomal recessive	Acrocephaly, brachydactyly, developmental delay, congenital heart defects; growth retardation, hypogonadism
Alström's	Autosomal recessive	Progressive cone-rod dystrophy, sensorineural hearing loss, hyperinsulinemia, early type 2 diabetes mellitus, dilated cardiomyopathy, pulmonary, hepatic and renal fibrosis
Tubby	Autosomal recessive	Progressive cone-rod dystrophy, hearing loss

CLASSICAL SYNDROMIC DISORDERS

A number of syndromes were identified by clinicians long before their exact genetic cause was known. In these syndromes, obesity is associated with a stereotyped set of other anomalies, often neurodevelopmental in type. The precise genetic basis for the majority of these syndromes is now known. Prader-Willi syndrome (PWS) is the most common syndromic cause of obesity, with an estimated prevalence of ~1 in 25,000. It is an autosomal dominant disorder caused by deletion of an imprinted region on the paternal chromosome 15 **(Chap. 479)**. The characteristic clinical features are hypotonia, feeding difficulties in infancy, developmental delay, hypogonadotropic hypogonadism, hyperphagia (increased food intake), and obesity. Children with PWS are short with reduced lean body mass and increased fat mass, features resembling those seen in growth hormone (GH) deficiency; GH treatment decreases body fat and increases linear growth and muscle mass and is now standard of care in this condition. Low levels of brain expression of the neuropeptide oxytocin and the nerve growth factor brain-derived neurotrophic factor (BDNF) in PWS patients have suggested new therapeutic opportunities for these patients.

Inherited or de novo (not found in either parent) mutations in another imprinted gene, *GNAS1*, which encodes Gsα protein, cause a syndrome known as Albright's hereditary osteodystrophy (AHO) or pseudohypoparathyroidism (PHP) **(Chap. 424)**. Maternal transmission of *GNAS1* mutations leads to short stature, obesity, and skeletal defects plus resistance to several hormones (e.g., parathyroid hormone), whereas paternal transmission leads only to the AHO phenotype. The clinical spectrum is very broad, and some patients may present with obesity alone.

Bardet-Biedl syndrome (BBS) is a rare autosomal recessive disease characterized by obesity, developmental delay, polydactyly, retinal dystrophy or pigmentary retinopathy, hypogonadism, and renal

TABLE 413-2 Obesity Syndromes due to Mutations in Genes Controlling Energy Homeostasis Pathways

GENE AFFECTED	INHERITANCE	ADDITIONAL CLINICAL FEATURES
Leptin	Autosomal recessive	Severe hyperphagia, frequent infections, hypogonadotropic hypogonadism, mild hypothyroidism
Leptin receptor	Autosomal recessive	Severe hyperphagia, frequent infections, hypogonadotropic hypogonadism, mild hypothyroidism
Proopiomelanocortin	Autosomal recessive	Hyperphagia, cholestatic jaundice or adrenal crisis due to ACTH deficiency, pale skin and red hair
Prohormone convertase 1	Autosomal recessive	Small-bowel enteropathy, postprandial hypoglycemia, hypothyroidism, ACTH deficiency, hypogonadism, central diabetes insipidus
Carboxypeptidase E	Autosomal recessive	Severe insulin resistance
Melanocortin 4 receptor	Autosomal dominant	Hyperphagia, accelerated linear growth
Single-minded 1	Autosomal dominant	Hyperphagia, accelerated linear growth, speech and language delay, autistic traits
BDNF	Autosomal dominant	Hyperphagia, developmental delay, hyperactivity, behavioral problems including aggression
TrkB	Autosomal dominant	Hyperphagia, speech and language delay, variable developmental delay, hyperactivity, behavioral problems including aggression
SH2B1	Autosomal dominant	Hyperphagia, disproportionate hyperinsulinemia, early type 2 diabetes mellitus, behavioral problems including aggression

Abbreviations: ACTH, adrenocorticotropic hormone; *BDNF*, brain-derived neurotrophic factor; *SH2B1*, Src-homology-2 (SH2) B-adaptor protein-1 (SH2B1); *TrkB*, tropomyosin receptor kinase B.

abnormalities. The same clinical features can arise from mutations in >26 genes, which disrupt signaling in primary cilia. Melanocortin receptor agonists may be useful in treating hyperphagia and obesity in patients with BBS. Overlapping clinical features are seen in a number of other genetic obesity syndromes (Table 413-1).

■ DISORDERS OF LEPTIN-MELANOCORTIN SIGNALING

Homozygous mutations that disrupt the production or action of leptin are rare but result in a disorder that is treatable. Children with homozygous loss-of-function leptin mutations have rapid weight gain in the first few months of life, resulting in severe obesity due to an intense drive to eat (hyperphagia) and impaired satiety with food-seeking behavior soon after the end of a meal. Congenital leptin deficiency can be treated with subcutaneous injections of recombinant leptin, which reduce hunger, increase satiety, and lead to weight loss. Similar clinical features are seen in patients with homozygous mutations in the leptin receptor gene, but they are not responsive to leptin treatment (Table 413-2). Normal pubertal development rarely occurs in adults with leptin or leptin receptor deficiency, with biochemical evidence of hypogonadotropic hypogonadism. However, there is some evidence for the delayed but spontaneous onset of menses in a small number of leptin- and leptin receptor–deficient adults. Leptin treatment permits progression of pubertal development, suggesting that leptin is a permissive factor for the development of puberty. An MC4R agonist (setmelanotide) is licensed for chronic weight management in leptin receptor deficiency.

Homozygous or compound heterozygous mutations in POMC lead to hyperphagia and early-onset obesity. As adrenocorticotropic hormone (ACTH) is produced in the pituitary gland by cleavage from POMC, patients also present with isolated ACTH deficiency (neonatal hypoglycemia and cholestatic jaundice). In the skin, POMC-derived melanocortin peptides act on melanocortin 1 receptors to induce pigmentation. For this reason, the lack of POMC-derived peptides in obese patients with POMC deficiency results in hypopigmentation of skin and hair, which is more noticeable in people of Caucasian ancestry who often have red hair. Prohormone convertase 1 (PCSK1) is an enzyme involved in the cleavage of POMC into ACTH, which is then further cleaved to make α-MSH by carboxypeptidase E. Impaired processing of POMC contributes to the hyperphagic severe early-onset obesity and ACTH deficiency in people lacking PCSK1 who also have hypogonadotropic hypogonadism, postprandial hypoglycemia (due to impaired processing of proinsulin to insulin), and a neonatal enteropathy in early childhood. Heterozygous mutations that impair the function of MC4R are found in 5–6% of patients with severe early-onset obesity and at a frequency of ~1 in 300 in the general population, making this the most common gene in which variants contribute to obesity. MC4R mutations are inherited in a co-dominant manner, with variable penetrance and expression in heterozygous carriers; homozygous carriers are severely obese. Patients are often hyperphagic from early childhood and hyperinsulinemic and have increased lean mass and increased linear growth.

■ GENETIC SUBTYPES OF OBESITY ASSOCIATED WITH NEUROBEHAVIORAL ABNORMALITIES

Both PWS patients and patients with mutations in *SIM1* (a gene that acts downstream of MC4R) exhibit a spectrum of behavioral abnormalities that overlap with autism-like features that could be related to reduced oxytocin signaling (Table 413-2). Mutations affecting BDNF and its receptor tropomyosin receptor kinase B (TrkB) cause speech and language delay, hyperphagia, and severe obesity, as well as hyperactivity, autistic traits, and impaired short-term memory. Interestingly, a common variant in BDNF (V66M), found in heterozygous form in ~20% of the population, is associated with a number of traits and neuropsychiatric disorders including anxiety and depression. Chromosomal deletion and mutations affecting Src-homology-2 (SH2) B-adaptor protein-1 (*SH2B1*) are associated with dominantly inherited, severe, early-onset obesity, disproportionate insulin resistance, early-onset type 2 diabetes, and behavioral problems including aggressive behavior.

■ OBESITY SECONDARY TO OTHER DISORDERS

Endocrine Disorders Patients with hypothyroidism may gain weight and develop obesity, although it is rarely the sole cause of severe obesity. It is nonetheless prudent always to measure thyroid function in a patient presenting with obesity. Measurement of thyroid-stimulating hormone (TSH) will detect significant primary disease of the thyroid, but for rare secondary hypothyroidism, additional measurement of free thyroxine levels is needed **(Chap. 395)**. Weight gain can also be a presenting feature of Cushing's syndrome. Clinically, the presence of spontaneous bruising, livid striae, myopathy, and marked centripetal distribution of body fat helps to distinguish true endogenous hypercortisolism from common obesity. This condition is usually reasonably straightforward to diagnose based on tests that approximate cortisol production rates (24-h urine free cortisol) or the suppression of serum cortisol by dexamethasone **(Chap. 398)**. Occasionally, in patients with severe obesity, effects of adiposity on glucocorticoid metabolism can make it difficult to interpret results, and more sophisticated tests, including those measuring diurnal rhythm of cortisol, may be necessary to establish or exclude the diagnosis with confidence. Weight gain can also be a presenting feature of patients with insulinoma, driven largely by the need to eat more frequently than normal to avoid hypoglycemia.

Hypothalamic Damage The hypothalamic regions that control energy balance can be disrupted by tumors (such as craniopharyngiomas), inflammatory masses, or after a severe head injury **(Chap. 391)**. In such cases, there is often some accompanying evidence of disruption of the hormonal functions of the anterior or posterior pituitary, although it may be subtle and the history of hyperphagia and weight gain is often short. It is worth noting that in common obesity, GH levels in response

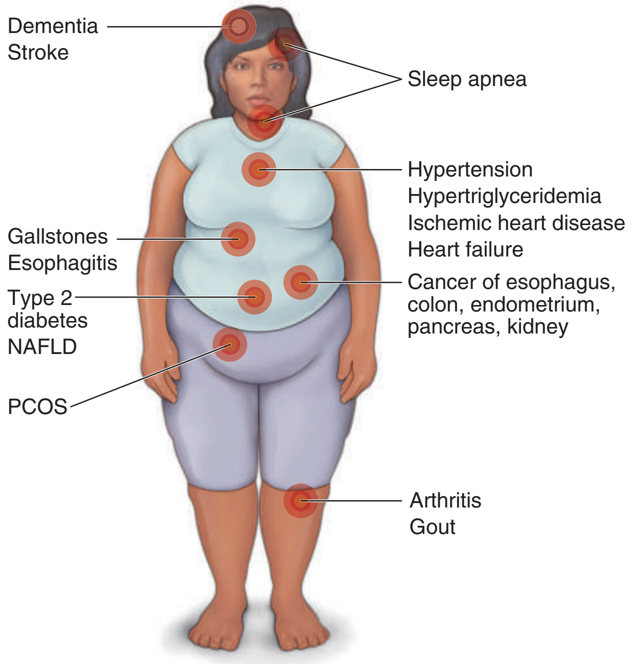

FIGURE 413-4 Obesity-related complications. The expanded fat mass that characterizes obesity predisposes to certain obesity-related complications (e.g., osteoarthritis of knees, reflux esophagitis, and obstructive sleep apnea) directly through its mass and/or volume. However, in the case of the metabolic, endocrine, and cardiovascular complications, the link is less clear. Further research is needed to establish whether some features of the expanded fat mass influence the development of these complications or whether other aspects of the chronically overnourished state, such as excess fat outside the fat depot, are more relevant. NAFLD, nonalcoholic fatty liver disease; PCOS, polycystic ovarian syndrome.

to provocative testing may be somewhat lower than normal, but this does not necessarily suggest the presence of a structural lesion.

ADVERSE CONSEQUENCES OF OBESITY

Mechanistic Considerations Obesity is associated with a wide range of pathologies that can adversely impact morbidity and mortality (Chap. 420). Some of these consequences are related, at least in part, to the direct mechanical or gravitational effects of the expanded fat mass itself (Fig. 413-4). However, the principal mechanisms behind many of the complications of obesity are less likely to be due to the expanded fat mass itself but more closely related to the chronic state of overnutrition itself and its effects on tissues throughout the body.

As people develop obesity, one of the first and most prominent biochemical abnormalities that develops is the need for increased circulating concentrations of insulin to maintain glucose homeostasis. This state of insulin resistance generally worsens with a greater degree of obesity, but there is a high degree of interindividual variability. It is more prominent when fat is distributed more centrally. Insulin resistance/hyperinsulinemia is likely to play a major role in the predisposition to metabolic endocrine and cardiovascular diseases seen more frequently in obesity and may even play a role in the predisposition of people with obesity to develop certain cancers.

The main sites of insulin action in the body are the liver and skeletal muscle. Thus, for insulin resistance to be discernible at the level of the whole body, the action of insulin must be disturbed in one or both of these tissues. It seems unlikely that an expanded fat cell mass would do that directly. How then does obesity lead to a state of insulin resistance? One hypothesis suggests a leading role for the inflammation that occurs in the adipose tissue in obesity (Fig. 413-5). This undoubtedly happens, as there are more macrophages in obese than nonobese adipose tissues, and this is associated with higher levels of inflammatory markers in the circulation of people with obesity. The majority of macrophages in obese adipose tissue are found in clusters around dead or dying adipocytes, so it appears that these cells are clearing debris after cell death. Studies in animal models provide strong support for the notion that this inflammatory state is mechanistically linked to insulin resistance, but evidence from humans for this is not as strong.

An alternative hypothesis is that as individuals develop obesity they become less able to safely store nutrients in their adipose tissue and begin to redirect macronutrients to other tissues that are not designed for fat storage and may be damaged by the nutrient excess. This certainly happens to people who are born with a lack of adipose tissue (lipodystrophy) who, early in life, develop severe versions of all the metabolic complications that are seen in obesity as they have no safe depot in which to store excess nutrients. There are stronger human data from both genetic and pharmacologic studies for the existence of the latter mechanism. How ectopic fat leads to insulin resistance and other damaging effects is still a puzzle, but it is very likely a major driver of pathology associated with obesity.

Metabolic Complications • DYSLIPIDEMIA The insulin resistance of obesity is frequently associated with dyslipidemia characterized by high circulating triglycerides and low high-density lipoprotein cholesterol (Chap. 419). Occasionally, the hypertriglyceridemia may be severe enough to put the patient at risk of pancreatitis. Although there is a relationship between obesity and raised circulating levels of low-density lipoprotein cholesterol (which is the major risk factor for coronary artery disease), genetic factors independent of obesity and the type of dietary fat consumed probably have an even greater impact.

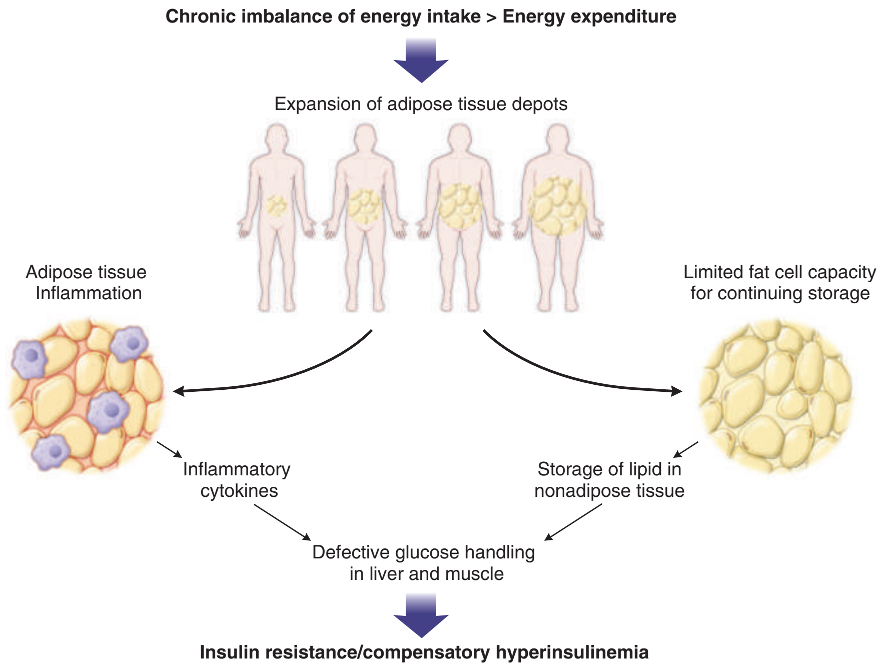
Chronic imbalance of energy intake > Energy expenditure
Expansion of adipose tissue depots
Adipose tissue Inflammation
Limited fat cell capacity for continuing storage
Inflammatory cytokines
Storage of lipid in nonadipose tissue
Defective glucose handling in liver and muscle
Insulin resistance/compensatory hyperinsulinemia

FIGURE 413-5 How does obesity cause metabolic disease? Insulin resistance is one of the earliest complications of obesity and underlies and precedes many of its adverse health consequences. The disposal and production of glucose by the most important tissues, muscle and liver, respectively, become less sensitive to insulin, and this results in a compensatory increase in insulin secretion from the pancreas. There are two main theories for the association of obesity with insulin resistance. In the first, products of macrophages and other inflammatory cells that are more abundant in obese adipose tissue can, through paracrine or endocrine routes, disturb insulin's action in muscle and liver cells. In the second, as adipose storage deposits fill up, they become less able to take on excessive calories, which end up being stored as ectopic lipid in tissues such as muscle and liver, which are not primarily designed to store nutrients of this type. The evidence in humans is stronger for the latter hypothesis.

FATTY LIVER DISEASE Obesity is strongly associated with the presence of ectopic fat in hepatocytes. In a subset of patients, this can progress to nonalcoholic steatohepatitis (NASH), which can progress to fibrosis, which is a precursor to cirrhosis **(Chap. 354)**. The reported incidence of NASH-related cirrhosis and of hepatocellular carcinoma has increased markedly in step with the increase in the prevalence of obesity in adolescents and adults.

TYPE 2 DIABETES The insulin resistance characteristic of the overnourished state strongly predisposes to the development of type 2 diabetes in people who, largely for genetic reasons, are less able to maintain the high levels of insulin secretion over many decades. Impaired glucose tolerance and type 2 diabetes are among the most common complications of obesity **(Chap. 415)**.

Endocrine Complications In females, the insulin resistance/hyperinsulinemia frequently found in obesity strongly predisposes to the development of polycystic ovaries, characterized by irregular menstruation, anovulatory infertility, and hirsutism due to hyperandrogenism. In males, obesity is more often associated with a degree of central hypogonadism, where low circulating testosterone is associated with levels of luteinizing hormone and follicle-stimulating hormone that do not rise appropriately to compensate for the testosterone-deficient state.

Dermatologic Complications Obesity can result in problems with excessive skin folds that can cause discomfort through mechanical irritation and can also become infected with fungi. Insulin resistance/hyperinsulinemia is associated with acanthosis nigricans, where areas such as axilla, groin, and the back of neck develop velvety hyperpigmentation. Hidradenitis suppurativa is a potentially disabling skin condition strongly associated with obesity. It is characterized by recurrent boils often with chronically draining sinus tracts affecting skin areas containing apocrine sweat glands.

Cardiovascular Complications People with obesity, even if they do not have diabetes, have increased morbidity and mortality from atherothrombotic vascular disease, including coronary artery disease and stroke. The factors that result in this are complex and involve increased prevalence of hypertension, dyslipidemia, and insulin resistance/hyperinsulinemia. The rare condition of thrombotic thrombocytopenic purpura, which causes microvascular platelet thrombosis, thrombocytopenia, and hemolytic anemia due to the presence of abnormally large von Willebrand factor multimers, is strongly associated with obesity.

Independent of occlusive arterial disease, people with obesity are also at increased risk of heart failure, particularly characterized primarily by diastolic dysfunction, and of atrial fibrillation, the most common arrhythmia.

Respiratory Complications Exertional dyspnea is common in obesity, contributed to by the increased work required to move a greater mass as well as impacts of pressure on the diaphragm and thoracic cage on chest wall compliance. Enlargement of soft tissue of the mouth and throat and adipose depots around the airways contribute to the high prevalence of sleep apnea, although other factors such as central nocturnal hypoventilation, also contribute in some people.

Gastrointestinal Disorders Reflux esophagitis is the most common gastrointestinal complication of obesity, particularly occurring in those with high intraabdominal pressure. Gallstones are also more common in people with obesity, bringing increased risks of biliary colic, cholecystitis, pancreatitis, and gallbladder cancer.

Rheumatologic Disorders Osteoarthritis of the knee and gout are the two most common rheumatologic conditions clearly associated with obesity. Interestingly, despite obesity being described as a proinflammatory state, there is no evidence for an increase in rheumatoid arthritis or the seronegative arthritides among people with obesity.

Cancers Obesity is a risk factor for a number of common cancers. Indeed, it has recently been calculated that, at least in some countries, obesity has overtaken smoking as the greatest risk factor for developing cancer. Recent research has found that as the BMI increases by 5 kg/m^2, cancer mortality increases by 10%. The largest effects are on colorectal, kidney, and pancreatic cancer, adenocarcinoma of the esophagus, and, in women, endometrial carcinoma. The recent rapid increase in the prevalence of esophageal adenocarcinoma is likely related to the marked recent increase in reflux esophagitis due to the raised intraabdominal pressure (with or without hiatus hernia) characteristic of central obesity.

Response to Infection The fact that obesity can influence the outcome of some infections has become very apparent with the COVID-19 pandemic. Obese patients have a substantially worse outcome if infected by SARS-CoV-2 through mechanisms that are as yet unclear. Obese patients also appear to be more susceptible to bacterial wound infections and postoperative sepsis.

Disorders of the Central Nervous System There is increasing evidence that obesity is a risk factor for dementia in later life, although how that risk is mediated is not clear. Idiopathic intracranial hypertension is a rare disorder that is strongly associated with obesity.

CONCLUSION

Obesity is a medical disorder that has been greatly increasing in prevalence due to environmental factors that are ubiquitous in developed and developing countries. However, it is important to bear in mind that it is a highly heterogeneous condition, which in some people is attributable entirely to genetic causes, and that underlying genetic variation strongly influences the risk of obesity in all people. It is a serious condition leading to multiple adverse health outcomes and considerable human suffering. As our understanding of its pathogenesis increases, our duty to treat obese patients with understanding and compassion and to develop new and better options for its treatment and prevention is worthy of emphasis.

FURTHER READING

Farooqi IS, O'Rahilly S: The genetics of obesity in humans, in *Endotext*. KR Feingold et al (eds). South Dartmouth, MA, 2000.

Friedman JM: Leptin and the endocrine control of energy balance. Nat Metab 1:754, 2019.

Hall KD et al: The energy balance model of obesity: Beyond calories in, calories out. Am J Clin Nutr 115:1243, 2022.

Heymsfield SB, Wadden TA: Mechanisms, pathophysiology, and management of obesity. N Engl J Med 376:1492, 2017.

Leibel RL et al: Changes in energy expenditure resulting from altered body weight. N Engl J Med 332:621, 1995.

NCD Risk Factor Collaboration (NCD-RISC): Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: A pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. Lancet 390:2627, 2017.

O'Rahilly S: Harveian Oration 2016: Some observations on the causes and consequences of obesity. Clin Med (Lond) 16:551, 2016.

414

Evaluation and Management of Obesity

Robert F. Kushner

More than 70% of U.S. adults are considered to be overweight or have obesity, and the prevalence of obesity is increasing rapidly in most of the industrialized world. Children and adolescents also are developing greater rates of obesity, indicating that the current trends will accelerate over time. Obesity is associated with an increased risk of multiple health problems, including hypertension, type 2 diabetes, dyslipidemia, obstructive sleep apnea, metabolic dysfunction-associated steatotic liver disease, degenerative joint disease, and some malignancies. Thus, it is important for health care professionals to identify, evaluate, and treat patients for obesity and associated complications and comorbid conditions.

■ EVALUATION

Health care professionals should screen all adult patients for obesity and offer intensive lifestyle counseling including behavioral interventions to promote sustained weight loss. The four main steps in the evaluation of obesity, as described below, are (1) a focused obesity-related history that includes lifestyle questions about diet, physical activity, sleep, and stress; (2) a physical examination to determine the degree and type of obesity; (3) assessment of complications and comorbid conditions; and (4) assessment of the patient's readiness to engage in weight management.

The Obesity-Focused History The first step in taking an obesity-focused history is to approach the topic in a sensitive manner. The reason for this concern is that the word *obesity* is a highly charged, emotive term. It has a significant pejorative meaning for many patients, leaving them feeling judged and blamed when labeled as such. This is not the case when patients are told that they have other chronic diseases such as diabetes or hypertension. Patients prefer that clinicians use more neutral words or terms such as *weight*, *excess weight*, *body mass index (BMI)*, or *unhealthy weight*, versus more perceived stigmatizing terms such as *obesity*, *morbid obesity*, or *fatness*.

Information from the history should address the following seven questions:

- What factors contribute to the patient's weight gain and obesity?
- How is obesity affecting the patient's health?
- What is the patient's level of risk from obesity?
- What does the patient find difficult about managing weight?
- What are the patient's goals and expectations?
- What is the patient's motivation to begin a weight management program?
- What kind of help does the patient need?

Although the majority of cases of obesity are promoted by biopsychosocial and behavioral factors that affect diet and physical activity patterns, the history may suggest secondary causes that merit further evaluation. Disorders to consider include polycystic ovarian syndrome, hypothyroidism, Cushing's syndrome, and hypothalamic disease. Drug-induced weight gain also should be considered. Common causes include medications for diabetes (insulin, sulfonylureas, thiazolidinediones), steroid hormones, antipsychotic agents (clozapine, olanzapine, risperidone), mood stabilizers (lithium), antidepressants (tricyclics, monoamine oxidase inhibitors, paroxetine, mirtazapine), and antiepileptic drugs (valproate, gabapentin, carbamazepine). Other medications, such as nonsteroidal anti-inflammatory drugs and calcium channel blockers, may cause peripheral edema but do not increase body fat.

The patient's current diet and physical activity patterns may reveal factors that contribute to the development of obesity and may identify behaviors to target for treatment. Physical fitness and sedentary lifestyle, in particular, are important predictors of all-cause mortality rate independent of BMI and body composition, which highlights the importance of taking a physical activity and exercise history during examination as well as emphasizing physical activity as a treatment approach.

Inquiring about sleep health that addresses regularity, duration, efficiency, and satisfaction is also important. Although the mechanisms are uncertain, sleep deprivation is associated with metabolic alterations in appetite regulation, sympathetic nervous system overactivity, insulin sensitivity, and changes in circadian rhythm. Stress may also contribute to obesity, in part due to activation of the adrenal cortical axis and elevated cortisol levels and its impact on emotional health and behaviors. This historic information is best obtained by the combination of a questionnaire and an interview.

TABLE 414-1 Classification of Weight Status and Disease Risk

CLASSIFICATION	BODY MASS INDEX (kg/m^2)	OBESITY CLASS	DISEASE RISK
Underweight	<18.5	—	—
Healthy weight	18.5–24.9	—	—
Overweight	25.0–29.9	—	Increased
Obesity	30.0–34.9	I	High
Obesity	35.0–39.9	II	Very high
Extreme obesity	≥40	III	Extremely high

Source: Reproduced with permission from WHO Consultation on Obesity (1997): Geneva, Switzerland), World Health Organization; 1997.

BMI and Waist Circumference Three key anthropometric measurements are important in evaluating the degree of obesity: weight, height, and waist circumference. The BMI, calculated as weight (kg)/height (m)^2 or as weight (lb)/height (in)^2 × 703, is used to classify weight status and risk of disease **(Table 414-1)**. BMI is correlated with body fat and is related to disease risk. Lower BMI thresholds for overweight and obesity have been proposed for the Asia-Pacific region since this population appears to be at risk for glucose and lipid abnormalities at lower body weights. The problem with BMI is that it only measures the size of an individual. It does *not* measure body composition, distribution of body fat, health, quality of life, or any individual characteristics. BMI has many limitations but is still useful for screening and as a population estimate of increased morbidity and mortality. Currently, direct measurement of excess body fat is not universally practical in the clinical setting.

Excess abdominal fat, assessed by measurement of waist circumference or waist-to-hip ratio, is independently associated with a higher risk for metabolic syndrome, diabetes mellitus, and cardiovascular disease. Measurement of the waist circumference is a surrogate for visceral adipose tissue and should be performed in the horizontal plane above the iliac crest in individuals with a BMI ≤35 kg/m^2 **(Table 414-2)**.

Obesity-Associated Complications and Comorbid Conditions The evaluation of complications and comorbid conditions should be based on presentation of symptoms, risk factors, and index of suspicion. For all patients, a fasting lipid profile (total, low-density lipoprotein, and high-density lipoprotein cholesterol and triglyceride levels), chemistry panel, and glycated hemoglobin should be performed, and blood pressure determined. Symptoms and diseases that are directly or indirectly related to obesity are listed in **Table 414-3**.

TABLE 414-2 Ethnic-Specific Cutpoint Values for Waist Circumference

ETHNIC GROUP	WAIST CIRCUMFERENCE
Europeans	
Men	>94 cm (>37 in)
Women	>80 cm (>31.5 in)
South Asians and Chinese	
Men	>90 cm (>35 in)
Women	>80 cm (>31.5 in)
Japanese	
Men	>85 cm (>33.5 in)
Women	>90 cm (>35 in)
Ethnic South and Central Americans	Use South Asian recommendations until more specific data are available.
Sub-Saharan Africans	Use European data until more specific data are available.
Eastern Mediterranean and Middle Eastern (Arab) populations	Use European data until more specific data are available.

Source: Reproduced with permission from KG Alberti, P Zimmet, J Shaw; IDF Epidemiology Task Force Consensus Group. The metabolic syndrome-a new worldwide definition. Lancet 366:1059, 2005.

TABLE 414-3 Obesity-Related Organ Systems Review

- Cardiovascular
  - Hypertension
  - Congestive heart failure
  - Cor pulmonale
  - Varicose veins
  - Pulmonary embolism
  - Coronary artery disease
- Endocrine
  - Metabolic syndrome
  - Type 2 diabetes
  - Dyslipidemia
  - Polycystic ovarian syndrome
- Musculoskeletal
  - Hyperuricemia and gout
  - Immobility
  - Osteoarthritis (knees and hips)
  - Low back pain
  - Carpal tunnel syndrome
- Psychological
  - Depression/low self-esteem
  - Body image disturbance
  - Social stigmatization
- Integument
  - Striae distensae
  - Stasis pigmentation of legs
  - Lymphedema
  - Cellulitis
  - Intertrigo, carbuncles
  - Acanthosis nigricans
  - Acrochordons (skin tags)
  - Hidradenitis suppurativa
- Respiratory
  - Dyspnea
  - Obstructive sleep apnea
  - Hypoventilation syndrome
  - Pickwickian syndrome
  - Asthma
- Gastrointestinal
  - Gastroesophageal reflux disease
  - Nonalcoholic fatty liver disease
  - Cholelithiasis
  - Hernias
  - Colon cancer
- Genitourinary
  - Urinary stress incontinence
  - Obesity-related glomerulopathy
  - Hypogonadism (male)
  - Breast and uterine cancer
  - Pregnancy complications
- Neurologic
  - Stroke
  - Idiopathic intracranial hypertension
  - Meralgia paresthetica
  - Dementia

Although individuals vary, the number and severity of organ-specific complications and comorbid conditions usually rise with increasing levels of obesity.

Identifying the High-Risk Patient Efforts are under way to develop more practical and useful assessments to identify patients who are at high risk in addition to using BMI alone. Analogous to other staging systems commonly used for congestive heart failure or chronic kidney disease, the American Association of Clinical Endocrinology (AACE) and the American College of Endocrinology (ACE) guidelines advocate a simple and clinically useful obesity disease staging system that is based on ethnic-specific BMI cutoffs in conjunction with assessment for adiposity-related complications (Fig. 414-1). Stage 0 is assigned to individuals who are overweight or have obesity by BMI classification but have no complications, whereas stages 1 and 2 are defined as individuals who are overweight or have obesity by BMI classification and have one or more mild-moderate complications (stage 1) or at least one severe complication (stage 2). A different functional staging system for obesity, called the Edmonton Obesity Staging System (EOSS), classifies individuals with obesity into five graded categories (0–4), based on their morbidity and health-risk profile along three domains—medical, functional, and psychological. In this system, staging occurs independent of BMI.

Assessing the Patent's Readiness to Change An attempt to initiate lifestyle changes when the patient is not ready usually leads to frustration and may hamper future weight-loss efforts. Assessment includes patient motivation and support, stressful life events, psychiatric status, time availability and constraints, and appropriateness of goals and expectations. Readiness can be viewed as the balance of two opposing forces: (1) motivation, or the patient's desire to change; and (2) resistance, or the patient's barriers to change.

A helpful method to begin a readiness assessment is to use the motivational interviewing technique of "anchoring" the patient's interest and confidence to change on a numerical scale. With this technique, the patient is asked to rate—on a scale from 0 to 10, with 0 being not so important (or confident) and 10 being very important (or confident)—their level of interest in and confidence about engaging in

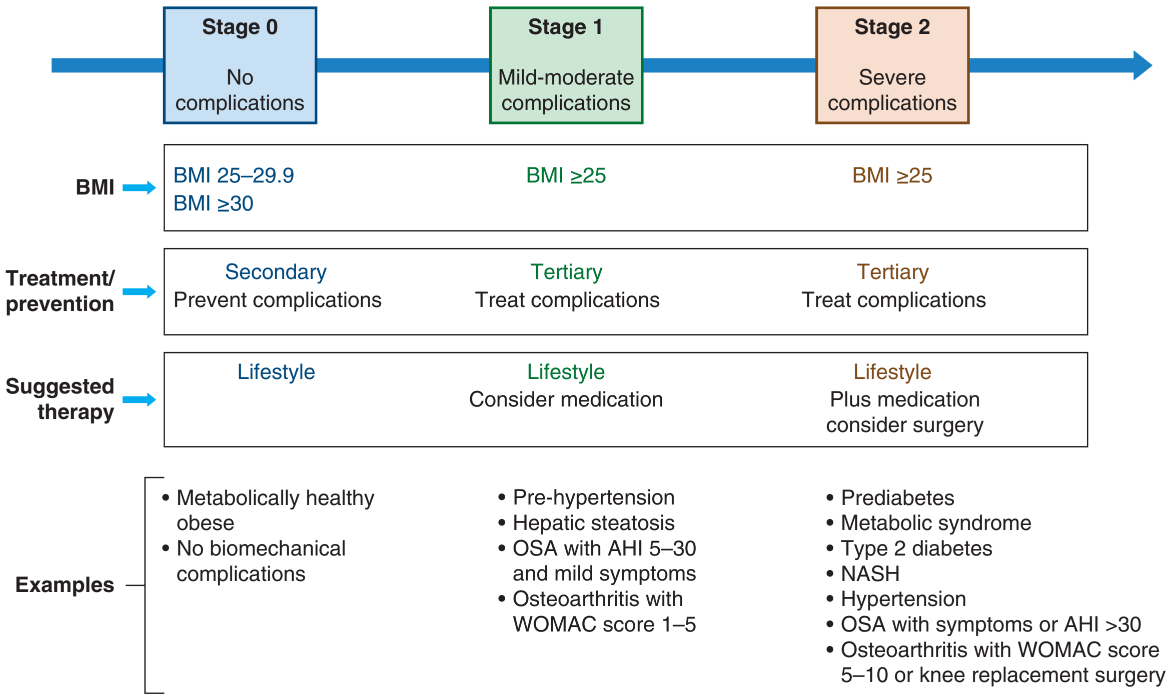

	Stage 0 No complications	Stage 1 Mild-moderate complications	Stage 2 Severe complications
BMI	BMI 25–29.9 BMI ≥30	BMI ≥25	BMI ≥25
Treatment/ prevention	Secondary Prevent complications	Tertiary Treat complications	Tertiary Treat complications
Suggested therapy	Lifestyle	Lifestyle Consider medication	Lifestyle Plus medication consider surgery
Examples	• Metabolically healthy obese • No biomechanical complications	• Pre-hypertension • Hepatic steatosis • OSA with AHI 5–30 and mild symptoms • Osteoarthritis with WOMAC score 1–5	• Prediabetes • Metabolic syndrome • Type 2 diabetes • NASH • Hypertension • OSA with symptoms or AHI >30 • Osteoarthritis with WOMAC score 5–10 or knee replacement surgery

FIGURE 414-1 Staging the severity of obesity using the American Association of Clinical Endocrinology clinical practice guidelines. AHI, apnea-hypopnea index; BMI, body mass index; NASH, nonalcoholic steatohepatitis; OSA, obstructive sleep apnea; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index (a patient-reported outcome measure for osteoarthritis registering pain, stiffness, and function). *(Data from WT Garvey et al: American Association of Clinical Endocrinologists and American College of Endocrinology comprehensive clinical practice guidelines for medical care of patients with obesity. Endocr Pract 22(Suppl 3):1, 2016.)*

weight management at this time. This exercise helps establish readiness to change and also serves as a basis for further dialogue.

TREATMENT

Obesity

THE GOAL OF THERAPY

The primary goals of treatment are to improve obesity-related complications and comorbid conditions and quality of life and reduce the risk of developing future obesity-related complications. Information obtained from the history, physical examination, and diagnostic tests is used to determine risk and develop a treatment plan. The decision of how aggressively to treat the patient and which modalities to employ is determined by using shared decision-making that includes the patient's risk status, expectations and personal values, and available resources. Not all patients who are deemed to have obesity by BMI screening need to be treated, since BMI alone is an imperfect measurement of the disease of obesity. However, patients who present with obesity-related complications and comorbidities and who would benefit from weight-loss intervention should be managed proactively. Therapy for obesity always begins with lifestyle management and may include pharmacotherapy or bariatric surgery, depending on BMI risk category **(Table 414-4)**. Setting an initial weight-loss goal of 8–10% over 6 months is a realistic and practical target.

LIFESTYLE MANAGEMENT

Obesity care involves attention to three essential elements of lifestyle: dietary habits, physical activity, and behavior modification. Because obesity is fundamentally a disease of energy imbalance, all patients must learn how and when energy is consumed (diet), how and when energy is expended (physical activity), and how to incorporate this information into their daily lives (behavioral therapy). Lifestyle management has been shown to result in a modest (typically 3–5 kg) weight loss when compared with no treatment or usual care.

Diet Therapy The primary focus of diet therapy is to reduce overall calorie consumption. Guidelines from the American Heart Association/American College of Cardiology/The Obesity Society (AHA/ACC/TOS) recommend initiating treatment with a calorie deficit of 500–750 kcal/d compared with the patient's habitual diet. Alternatively, a diet of 1200–1500 kcal/d for women and 1500–1800 kcal/d for men (adjusted for the individual's body weight) can be prescribed. This reduction is consistent with a goal of losing ~1–2 lb/week. The calorie deficit can be instituted through dietary substitutions or alternatives. Examples include choosing smaller portion sizes, eating more fruits and vegetables, consuming more whole-grain cereals, selecting leaner cuts of meat and skimmed dairy products, reducing consumption of fried foods and other foods with added fats and oils, and drinking water instead of sugar-sweetened beverages. It is important that dietary counseling remains patient centered and that the selected goals are SMART (specific, measurable, agreed upon, realistic, timely).

The macronutrient composition of the diet will vary with the patient's preference and medical condition. The 2020 U.S. Department of Agriculture Dietary Guidelines for Americans **(Chap. 343)**, which focus on health promotion and risk reduction, can be applied to treatment of patients who are overweight or have obesity. The recommendations include maintaining a diet rich in whole grains, fruits, vegetables, and dietary fiber; decreasing sodium intake to <2300 mg/d; consuming fat-free or low-fat dairy products; and keeping added sugars and saturated fat intake to <10% of daily calories. Application of these guidelines to specific calorie goals can be found on the website *www.choosemyplate.gov*. Since portion control is one of the most difficult strategies for patients to manage, the use of preprepared products such as meal replacements is a simple and convenient suggestion. Examples include frozen entrees, protein shakes, and bars. Use of meal replacements in the diet has been shown to result in a 7–8% weight loss.

Numerous randomized trials comparing diets of different macronutrient composition (e.g., low-carbohydrate, low-fat, Mediterranean) have shown that weight loss depends primarily on reduction of total caloric intake and adherence to the prescribed diet, not the specific proportions of carbohydrate, fat, and protein in the diet. The macronutrient composition will ultimately be determined by the patient's taste preferences, cooking style, and culture. However, the patient's underlying medical problems are also important in guiding the recommended dietary composition. The dietary prescription will vary according to the patient's metabolic profile and risk factors. A consultation with a registered nutritionist for medical nutrition therapy is particularly useful in considering patient preference and treatment of comorbid diseases.

Another dietary approach to consider is based on the concept of *energy density*, which refers to the number of calories (i.e., amount of energy) a food contains per unit of weight. People tend to ingest a constant volume of food regardless of caloric or macronutrient content. Adding water or fiber to a food decreases its energy density by increasing weight without affecting caloric content. Examples of foods with low energy density include soups, fruits, vegetables, oatmeal, and lean meats. Dry foods and high-fat foods such as pretzels, cheese, egg yolks, potato chips, and red meat have a high energy density. Diets containing low-energy-dense foods have been shown to control hunger and thus to result in decreased caloric intake and weight loss.

Occasionally, very-low-calorie diets (VLCDs) are prescribed as a form of aggressive dietary therapy. The primary purpose of a VLCD is to promote a rapid and significant (13- to 23-kg) short-term weight loss over a 3- to 6-month period. The proprietary formulas designed for this purpose typically supply ≤800 kcal, 50–80 g of protein, and 100% of the recommended daily intake for vitamins and minerals. Indications for initiating a VLCD include the involvement of well-motivated individuals who have moderate to severe obesity, have failed at more conservative approaches to weight loss, and have a medical condition that would be immediately improved with rapid weight loss. These conditions include poorly controlled type 2 diabetes, hypertriglyceridemia, obstructive sleep apnea, and symptomatic peripheral edema. In the DiRECT trial of patients with type 2 diabetes and obesity, a low-energy formula diet (825–853 kcal/d) was administered for 3 months following by a structured monthly program. At 12 months, almost half of the participants achieved remission to a nondiabetic state off of all antidiabetic drugs. Use of formula diets should be prescribed by trained practitioners in a medical care setting where medical monitoring and high-intensity lifestyle intervention can be provided.

Physical Activity Therapy Although exercise alone is only moderately effective for weight loss, the combination of dietary

TABLE 414-4 A Guide to Opting for Treatment for Obesity

	BMI CATEGORY (kg/m^2)				
TREATMENT	**25–26.9**	**27–29.9**	**30–34.9**	**35–39.9**	**≥40**
Diet, exercise, behavioral therapy	With comorbidities	With comorbidities	+	+	+
Pharmacotherapy	—	With comorbidities	+	+	+
Surgery	—	—	—	With comorbidities	+

Source: Reproduced from U.S. Department of Health and Human Services Public Health Service. National Institute of Health National Heart, Lung and Blood Institute. The Practical Guide Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. NIH Publication Number 00-4084. October 2000.

modification and exercise is the most effective behavioral approach for the treatment of obesity. The most important role of exercise appears to be in the maintenance of weight loss. The 2018 Physical Activity Guidelines for Americans (*www.health.gov/paguidelines*) recommend that adults should engage in 150 min of moderate-intensity or 75 min a week of vigorous-intensity aerobic physical activity per week, preferably spread throughout the week. Focusing on simple ways to add physical activity into the normal daily routine through leisure activities, travel, and domestic work should be suggested. Examples include brisk walking, using the stairs, doing housework and yard work, and engaging in sports. Additionally, it is important to reduce sedentary behavior, which is associated with all-cause and cardiovascular disease mortality in adults. Asking the patient to use a wearable activity tracker to monitor total accumulation of steps or kcal expended as part of the activities of daily living is a useful strategy. Step counts are highly correlated with activity level. Studies have demonstrated that lifestyle activities are as effective as structured exercise programs for improving cardiorespiratory fitness and weight loss. A high level of physical activity (>300 min of moderate-intensity activity per week) is often needed to lose weight and sustain weight loss. These exercise recommendations are daunting to most patients and need to be implemented gradually. Consultation with an exercise physiologist or personal trainer may be helpful.

Behavioral Therapy Cognitive behavioral therapy is used to help change and reinforce new dietary and physical activity behaviors. Strategies include self-monitoring techniques (e.g., journaling, weighing, and measuring food and activity); stress management; stimulus control (e.g., using smaller plates, not eating in front of the television or in the car); social support; problem solving; and cognitive restructuring to help patients develop more positive and realistic thoughts about themselves. When recommending any behavioral lifestyle change, the patient should be asked to identify what, when, where, and how the behavioral change will be performed. The patient should keep a record of the anticipated behavioral change so that progress can be reviewed at the next office visit. Because these techniques are time consuming to implement, their supervision is often undertaken by ancillary office staff, such as an advanced practice provider or registered nutritionist.

PHARMACOTHERAPY

Adjuvant pharmacologic treatments should be considered for patients with a BMI ≥ 30 kg/m^2 or a BMI ≥ 27 kg/m^2 who have concomitant obesity-related diseases and for whom dietary and physical activity therapy has not been successful. When an antiobesity medication is prescribed, patients should be actively engaged in a lifestyle program that provides the strategies and skills needed to use the drug effectively since such support increases total weight loss.

Medications for obesity have traditionally fallen into two major categories: those that affect appetite and those that inhibit gastrointestinal fat absorption. However, since the introduction of more effective nutrient-stimulated hormone-based therapeutics in 2021, an additional designation of first- and second-generation medications has been adopted. Antiobesity medications are approved by the U.S. Food and Drug Administration (FDA) with an indication for chronic weight management, with the exception of phentermine and other sympathomimetics, which are only approved for short-term use. The centrally active medications work biologically to suppress appetite, affecting hunger, satiety, and response to highly rewarding foods, and thus making it easier for patients to follow their dietary intentions to reduce caloric intake. In addition, one capsule that is considered a medical device was marketed in 2020. Characteristics of the currently approved antiobesity medications are shown in **Table 414-5**.

First-Generation Centrally Acting Medications This class of medications directly targets neurotransmitters in the hypothalamus and reward centers in the central nervous system **(Chap. 413)** to affect satiety (feeling of fullness after a meal), hunger (the biologic sensation that prompts eating), and craving (intense desire for a specific food). By controlling appetite, these agents help patients reduce caloric intake without a sense of deprivation. The classic sympathomimetic adrenergic agents (benzphetamine, phendimetrazine, diethylpropion, and phentermine) function by stimulating norepinephrine release or by blocking its reuptake. Among these agents, phentermine is the most commonly prescribed; however, there is limited long-term data on its effectiveness. A 2002 review of six randomized, placebo-controlled trials of phentermine for weight control found that patients lost 0.6–6.0 additional kg of weight over 2–24 weeks of treatment. The most common side effects of the amphetamine-derived agents are restlessness, insomnia, dry mouth, constipation, and increased blood pressure and heart rate.

Phentermine/topiramate (PHEN/TPM) is a combination drug that contains a catecholamine releaser (phentermine) and an anticonvulsant (topiramate). Topiramate is approved by the FDA as an anticonvulsant for the treatment of epilepsy and for the prophylaxis of migraine headaches. Weight loss was identified as an unintended side effect of topiramate during clinical trials for epilepsy. The mechanism responsible for weight loss is uncertain but is thought to be mediated through the drug's modulation of γ-aminobutyric acid receptors, inhibition of carbonic anhydrase, and antagonism of glutamate. PHEN/TPM has undergone two 1-year pivotal randomized, placebo-controlled, double-blind trials of efficacy and safety: EQUIP and CONQUER. In a third study, SEQUEL, 78% of CONQUER participants continued to receive their blinded treatment for an additional year. All participants received diet and exercise counseling. Mean percent weight loss for participants randomized to medication and placebo are displayed in **Fig. 414-2**. Intention-to-treat 1-year placebo-subtracted weight loss for PHEN/TPM was 9.3% (15-mg/92-mg dose) and 6.6% (7.5-mg/46-mg dose), respectively, in the EQUIP and CONQUER trials. Clinical and statistical dose-dependent improvements were seen in selected cardiovascular and metabolic outcome measurements that were related to the weight loss. The most common adverse events experienced by the drug-randomized group were paresthesias, dry mouth, constipation, dysgeusia, and insomnia. Because of an increased risk of congenital fetal oral-cleft formation from topiramate, women of childbearing age should have a negative pregnancy test before treatment and monthly thereafter and use effective contraception consistently during medication therapy.

Naltrexone SR/bupropion SR (NB) is a combination of an opioid antagonist and a dopamine and norepinephrine reuptake inhibitor, respectively. Individually, naltrexone is approved by the FDA for the treatment of alcohol dependence and for the blockade of the effects of exogenously administered opioids, whereas bupropion is approved as an antidepressant and smoking cessation aid. As a combination drug, each component works in consort: bupropion stimulates secretion of α-melanocyte-stimulating hormone (MSH) from proopiomelanocortin (POMC), whereas naltrexone blocks the feedback inhibitory effects of opioid receptors activated by the β-endorphin released in the hypothalamus, thus allowing the inhibitory effects of MSH to reduce food intake.

The medication has undergone three randomized, placebo-controlled, double-blind trials for efficacy and safety. Participants were randomized to receive NB (8 mg/90 mg two tablets bid) or placebo in the three COR studies. Whereas participants received standardized nutritional and exercise counseling in COR-I and COR-II, a more intensive behavior modification program was provided in COR-BMOD (Table 414-5). Intention-to-treat 1-year placebo-subtracted weight loss was 4.8%, 5.1%, and 4.2%, respectively, in the COR-I, COR-II, and COR-BMOD trials. Clinical and statistical dose-dependent improvements were seen in selected cardiovascular and metabolic outcome measurements that were related to the weight loss. However, the medication led to slight increases or smaller decreases in blood pressure and pulse than placebo. The most common adverse events experienced by the drug-randomized groups were nausea, constipation, headache, vomiting, dizziness, diarrhea, insomnia, and dry mouth.

TABLE 414-5 Antiobesity Medications

DRUG CHARACTERISTIC	PHENTERMINE	ORLISTAT	PHEN/TPM	NAL/BUP	LIRAGLUTIDE	SEMAGLUTIDE	TIRZEPATIDE
Mechanism of action	Sympathomimetic, increases norepinephrine release in CNS	Gastrointestinal lipase enzyme inhibitor	Phen: sympathomimetic in CNS; TPM: modulates GABA receptors in the CNS	Nal: blocks opioid-mediated POMC autoinhibition; bup: activates POMC in the hypothalamus	GLP-1 receptor agonist	GLP-1 receptor agonist	GLP-1/GIP dual receptor agonist
Route of administration, frequency, and dose	Oral, once to 3 times daily, doses of 8, 15, 30, and 37.5 mg	Oral, 3 times daily, within 1 h of fat-containing meals	Oral, once daily. Start 3.75/23 mg/d × 2 weeks, then 6.5/46 mg/d; can escalate to max dose of 15/2 mg/d	Oral, 1 tablet (8 mg/90 mg) qAM × 1 week; 1 tablet in morning and evening × 1 week; 2 tablets in morning and 1 tablet in evening × 1 week; then 2 tablets in morning and 2 tablets in evening	Subcutaneous, once daily; initiate at 0.6 mg/d × 1 week; increase by 0.6 mg weekly to 3 mg/d	Subcutaneous, once weekly; 0.25 mg qW × 4 weeks, then 0.5 mg qW × 4 weeks, then 1 mg qW × 4 weeks, then 1.7 mg qW × 4 weeks, then 2.4 mg qW	Subcutaneous, once weekly; 2.5 mg qW × 4 weeks, then 5 mg qW × 4 weeks, then 7.5 mg qW × 4 weeks, then 10 mg qW × 4 weeks, then 12.5 mg qW, then 15 mg qW
Percent weight loss (placebo subtracted)[a]	4.4%	4.1%	8.0%	5.1%	5.4%	12.5%	17.8%
Most common adverse effects	Dry mouth Insomnia Constipation Headache Dizziness	Steatorrhea Increased defecation Oily spotting Liquid stool Fecal urgency	Paresthesia Dry mouth Constipation Headache Insomnia Dizziness	Nausea Constipation Headache Vomiting Dizziness Insomnia	Nausea Diarrhea Constipation Dyspepsia Vomiting	Nausea Diarrhea Constipation Dyspepsia Vomiting	Nausea Diarrhea Constipation Abdominal pain
Contraindications	Uncontrolled hypertension, untreated hyperthyroidism, within 14 d of MAOI use	Chronic malabsorption	Uncontrolled hypertension, untreated hyperthyroidism, history of glaucoma, calcium oxalate nephrolithiasis, within 14 d of MAOI use	Uncontrolled hypertension, history of seizures, bulimia or anorexia nervosa, within 14 d of MAOI use, long-term opioid use	Personal or family history of medullary thyroid cancer, MEN type 2; pancreatitis is a caution	Personal or family history of medullary thyroid cancer; MEN type 2; pancreatitis is a caution	Personal or family history of medullary thyroid cancer, MEN type 2; pancreatitis is a caution

[a] Based on maximal dose.

Abbreviations: CNS, central nervous system; GABA, γ-aminobutyric acid; GIP, gastric inhibitory polypeptide; GLP-1, glucagon-like peptide 1; MAOI, monoamine oxidase inhibitor; MEN, multiple endocrine neoplasia; Nal/Bup, naltrexone/bupropion; Phen/TPM, phentermine/topiramate; POMC, proopiomelanocortin; qW, every week.

Liraglutide was the first glucagon-like peptide 1 receptor agonist (GLP-1 RA) with 97% homology to human GLP-1 that introduces nutrient-stimulated, hormone-based therapy for the treatment of obesity. In addition to its effect as an incretin hormone (glucose-induced insulin secretion), liraglutide inhibits both gastric emptying and glucagon secretion and stimulates GLP-1 receptors in the arcuate nucleus of the hypothalamus, the nucleus tractus solitarius of the brainstem, and projections to other appetite-modulating relay nuclei to reduce appetite. As a result of molecular modifications of the structure, liraglutide has reduced susceptibility to DPP-4 and can be dosed once daily by subcutaneous (SC) administration with a half-life of ~11–15 h. Liraglutide was initially approved for the treatment of type 2 diabetes in the United States in 2010 at doses up to 1.8 mg once daily. It was subsequently approved for obesity treatment at doses up to 3.0 mg once daily in 2014 for adults and, in 2020, for adolescents (aged ≥12 years).

Specifically targeting obesity, liraglutide has undergone five randomized, double-blind, placebo-controlled trials in adults called SCALE (Satiety and Clinical Adiposity–Liraglutide Evidence) involving >5000 adult participants to evaluate its efficacy and safety for weight management. All participants received diet and physical activity counseling and were randomized to receive liraglutide (3.0 mg SC daily) or placebo with the primary outcome of change in body weight. Intention-to-treat 1-year placebo-subtracted weight loss for these trials ranged from 3.4 to 6.1%. Clinical and statistical dose-dependent improvements were seen in selected cardiovascular and metabolic outcome measurements. In SCALE TEENS, which involved adolescents with obesity (average age 14.6 years), the liraglutide-treated group demonstrated a placebo-subtracted weight loss of 5%. The most common adverse events from the SCALE trials were nausea, diarrhea, constipation, and vomiting, which were reported as mild and transient. GLP-1 agonists should not be prescribed in patients with a family or personal history of medullary thyroid cancer or multiple endocrine neoplasia.

Setmelanotide is a melancortin-4 (MC4) receptor agonist that was FDA approved in 2020 for daily SC administration for chronic weight management in adults and pediatric patients 6 years of age and older with monogenic or syndromic obesity due to Bardet-Biedl syndrome (BBS) or POMC, proprotein convertase subtilisin/kexin type 1 (PCSK1), or leptin receptor (LEPR) deficiency. The medication addresses the underlying hyperphagia and specific molecular mechanism of these rare genetic diseases. The most common side effects include injection site reactions, skin hyperpigmentation, nausea, and spontaneous penile erections.

First-Generation Peripherally Acting Medication Orlistat is currently the only medication in this class. It is a synthetic hydrogenated derivative of a naturally occurring lipase inhibitor, lipostatin, that is produced by the mold *Streptomyces toxytricini*. This drug is a potent, slowly reversible inhibitor of pancreatic, gastric, and carboxylester lipases and phospholipase A_2, which are required for the hydrolysis of dietary fat into fatty acids and monoacylglycerols.

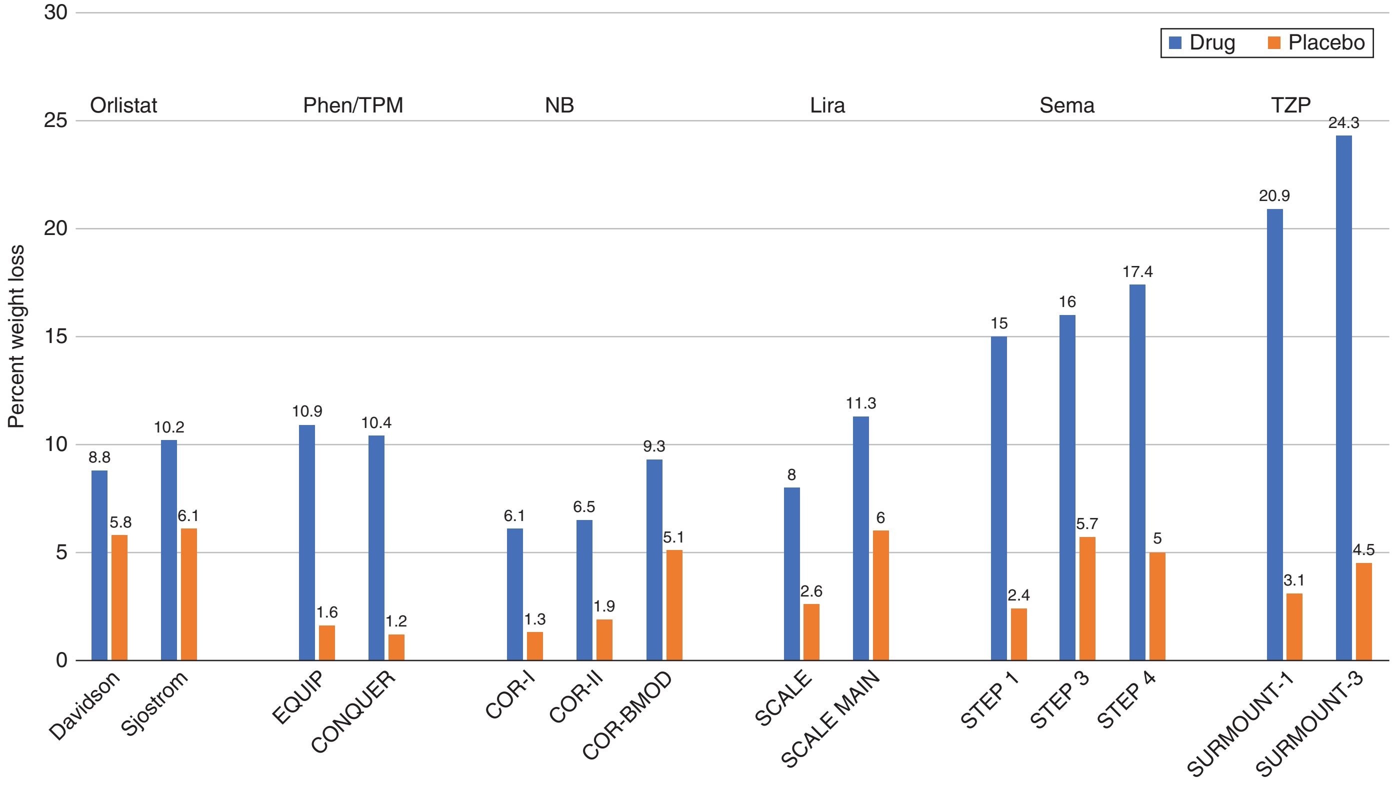

FIGURE 414-2 One-year mean weight loss for antiobesity medications compared to placebo. Lira, liraglutide; NB, naltrexone/bupropion; Phen/TPM, phentermine/topiramate; Sema, semaglutide; TZP, tirzepatide.

Orlistat acts in the lumen of the stomach and small intestine by forming a covalent bond with the active site of these lipases. Taken at a therapeutic dose of 120 mg tid, orlistat blocks the digestion and absorption of ~30% of dietary fat. After discontinuation of the drug, fecal fat content usually returns to normal within 48–72 h.

Multiple randomized, double-blind, placebo-controlled studies have shown an intention-to-treat 1-year placebo-subtracted weight loss of 2.7–4.1%. Because orlistat is minimally (<1%) absorbed from the gastrointestinal tract, it has no systemic side effects. Tolerability is related to the malabsorption of dietary fat, and this is generally diminished as patients control their dietary fat intake. Because serum concentrations of the fat-soluble vitamins D and E and β-carotene may be reduced by orlistat treatment, vitamin supplements are recommended to prevent potential deficiencies. Orlistat was approved for over-the-counter use in 2007.

Second-Generation Medications Semaglutide is generally recognized as the first drug in this category based on its greater weight loss efficacy and further chemical modification that allows for once-weekly SC administration with a longer half-life. Semaglutide was initially approved for the treatment of type 2 diabetes in the United States at doses up to 1.0 mg once weekly in 2017 and at 2.0 mg once weekly in 2022. It was subsequently approved at doses up to 2.4 mg once weekly for chronic weight management for adults in 2021 and for adolescents in 2022.

Semaglutide has undergone multiple prospective, randomized, placebo-controlled trials in the STEP (Semaglutide Treatment Effect in People with Obesity) program that was designed to investigate the effect of semaglutide 2.4 mg SC weekly versus placebo on weight loss, safety, and tolerability in adults with obesity or overweight. Intention-to-treat placebo-subtracted weight loss for the STEP 1 to 4 trials ranged from 6.2 to 14.8% at 68 weeks depending on the population and study design, and loss of 12.6% occurred after 104 weeks for STEP 5. Clinical and statistical dose-dependent improvements were seen in selected cardiovascular and metabolic outcome measurements. The most common adverse effects include nausea, diarrhea, constipation, and vomiting. GLP-1 RAs should not be prescribed in patients with a family or personal history of medullary thyroid cancer or multiple endocrine neoplasia.

In the STEP TEENS trial, adolescents with overweight or obesity randomized to semaglutide 2.4 mg achieved a 16.1% reduction in BMI versus a 0.6% increase for placebo. The recent SELECT (Semaglutide Effects on Cardiovascular Outcomes in People with Overweight or Obesity) trial demonstrated a 20% reduction in major adverse cardiovascular events in patients with preexisting cardiovascular disease and overweight or obesity but without diabetes who were randomized to semaglutide 2.4 mg versus placebo, and the STEP-HFpEF trial observed improved heart failure–related symptoms, physical limitations, and exercise function and reduced body weight and inflammation in patients randomized to semaglutide versus placebo.

Tirzepatide is the first long-acting, weekly, injectable, dual GLP-1/gastric inhibitory polypeptide (GIP) peptide analogue, engineered from the native GIP sequence with agonist activity at both the GLP-1 and GIP receptors with a half-life of ~117 h. Whereas GIP receptor agonism is equal to native GIP, the molecule's GLP-1 receptor affinity is approximately five times weaker than native GLP-1. GIP in the brain appears to act synergistically with GLP-1 receptor activation to allow greater weight loss. Tirzepatide was FDA approved for type 2 diabetes in 2022 and subsequently approved for chronic weight management in adults in 2023, at 5-, 10-, and 15-mg doses.

Tirzepatide has undergone multiple prospective, randomized, placebo-controlled trials in the SURMOUNT clinical development program that was designed to investigate the effect of tirzepatide 5-, 10-, and 15-mg SC weekly doses versus placebo on weight loss, safety, and tolerability in adults with obesity or overweight. Intention-to-treat placebo-subtracted weight loss for SURMOUNT 1 to 4 trials at the 15-mg dose ranged from 11.6 to 21.4% depending on the study population and design. Clinical and statistical dose-dependent improvements were seen in selected cardiovascular and metabolic outcome measurements. Similar to GLP-1 RAs, the most common adverse events are gastrointestinal (nausea, diarrhea, constipation, and vomiting) and reported as mild and transient; tirzepatide should not be prescribed in patients with a family or personal history of medullary thyroid cancer or multiple endocrine neoplasia.

Oral Device Gelesis100 is a nonsystemic, water-soluble gel that was approved by the FDA in 2019. In the stomach, the capsule releases the cellulose microgel, which absorbs water and forms a matrix with the consistency of food, occupying ~25% of the stomach. In the large intestine, it is broken down by enzymes and the cellulose is excreted. Gelesis100 and placebo were evaluated over 24 weeks in patients with BMI of 27 to ≤40 kg/m^2 and fasting plasma glucose of 90–145 mg/dL. Intention-to-treat, 24-week, placebo-subtracted weight loss was 2.1% (6.4 vs 4.4%). Gelesis100 treatment had no apparent increased safety risks. The capsules are approved for patients with a BMI of ≥25 kg/m^2, with or without comorbidities.

SURGERY

Bariatric surgery **(Fig. 414-3)** can be considered for patients with severe obesity (BMI ≥40 kg/m^2) or for those with moderate obesity (BMI ≥35 kg/m^2) associated with a number of comorbid conditions. The clinical benefits of bariatric surgery in achieving weight loss and alleviating metabolic complications and comorbidities have been attributed largely to changes in the physiologic responses of gut hormones, bile acid metabolism, the microbiota, and adipose tissue metabolism. Metabolic effects resulting from bypassing the foregut include altered responses of ghrelin, GLP-1, peptide YY3-36, and oxyntomodulin. Additional effects on food intake and body weight control may be attributed to changes in vagal signaling. The loss of fat mass, particularly visceral fat, is associated with multiple metabolic, adipokine, and inflammatory changes that include improved insulin sensitivity and glucose disposal; reduced free fatty acid flux; increased adiponectin levels; and decreased interleukin 6, tumor necrosis factor α, and high-sensitivity C-reactive protein levels.

Restrictive surgeries limit the amount of food the stomach can hold and slow the rate of gastric emptying. *Laparoscopic adjustable gastric banding* is the prototype of this category. The first banding device was approved for use in the United States in 2001. In contrast to previous devices, this band has a diameter that is adjustable by way of its connection to a reservoir that is implanted under the skin. Injection of saline into the reservoir and removal of saline from the reservoir tighten and loosen the band's internal diameter, respectively, thus changing the size of the gastric opening. Although the mean percentage of total body weight lost at 5 years is estimated at 20–25%, longer-term follow-up has been more disappointing, leading to near abandonment of the procedure. In the *laparoscopic sleeve gastrectomy*, the stomach is restricted by stapling and dividing it vertically, removing ~80% of the greater curvature and leaving a slim banana-shaped remnant stomach along the lesser curvature. Weight loss after this procedure is superior to that after laparoscopic adjustable gastric banding.

The three restrictive-malabsorptive bypass procedures combine the elements of gastric restriction and selective malabsorption: Roux-en-Y gastric bypass, biliopancreatic diversion, and biliopancreatic diversion with duodenal switch (Fig. 414-3). These procedures are routinely performed by laparoscopy.

These procedures generally produce a 28–33% average total body weight loss at 12–18 months followed by variable weight recurrence. Significant improvement in multiple obesity-related comorbid conditions, including type 2 diabetes, hypertension, dyslipidemia, obstructive sleep apnea, quality of life, and long-term cardiovascular events, has been reported. A meta-analysis of controlled clinical trials comparing bariatric surgery versus no surgery showed that surgery was associated with a reduced odds ratio (OR) risk of global mortality (OR = 0.55), cardiovascular death (OR = 0.58), and all-cause mortality (OR = 0.70).

Among the observed improvements in comorbidities, the prevention and treatment of type 2 diabetes resulting from bariatric

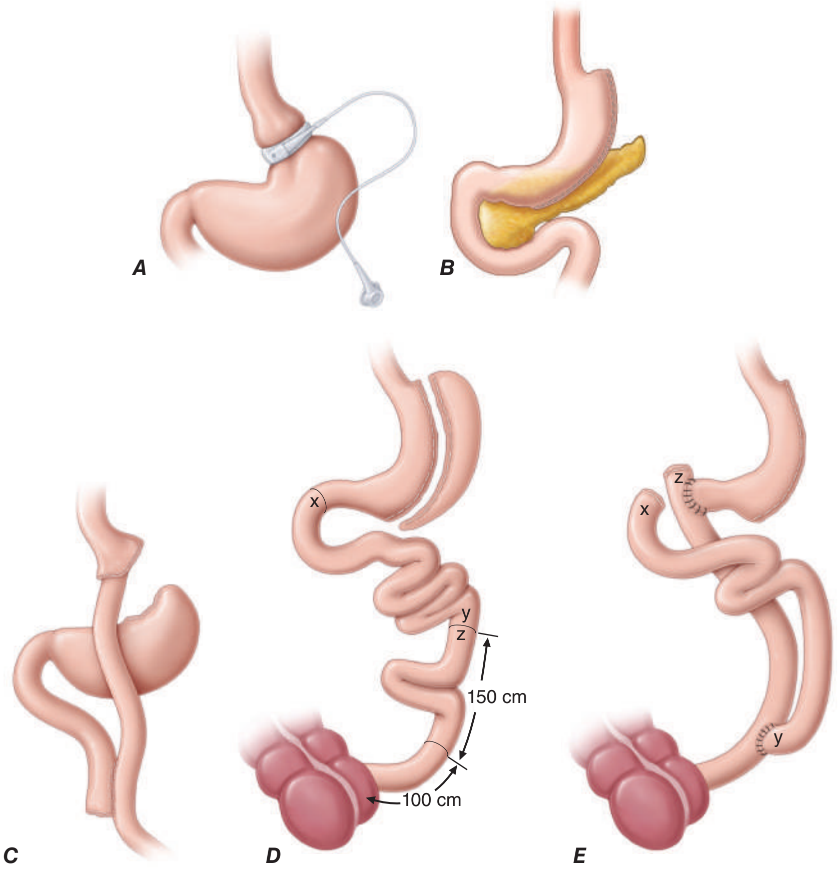

FIGURE 414-3 Bariatric surgical procedures. Examples of operative interventions used for surgical manipulation of the gastrointestinal tract. ***A.*** Laparoscopic adjustable gastric banding. ***B.*** Laparoscopic sleeve gastrectomy. ***C.*** The Roux-en-Y gastric bypass. ***D.*** Biliopancreatic diversion with duodenal switch. ***E.*** Biliopancreatic diversion.

surgery have garnered the most attention. Fifteen-year data from the Swedish Obese Subjects study demonstrated a marked reduction (i.e., by 78%) in the incidence of type 2 diabetes development among patients with obesity who underwent bariatric surgery. Multiple randomized controlled studies have shown greater weight loss and more improved glycemic control from 1 and 12 years among surgical patients than among patients receiving conventional medical therapy. A retrospective cohort study of >4000 adults with diabetes found that, overall, 68.2% of patients experienced an initial complete remission of type 2 diabetes within 5 years after surgery. However, among these patients, one-third redeveloped type 2 diabetes within 5 years. Patients with earlier-stage type 2 diabetes (i.e., those who do not need insulin, with shorter-duration disease, and with lower hemoglobin A_{1c}) appear to have better improvement after bariatric surgery. The rapid improvement seen in diabetes after bariatric surgery is thought to be due to caloric restriction, reduced insulin resistance, and surgery-specific effects on glucose homeostasis brought about by alteration of gut hormones.

The mortality rate from bariatric surgery is generally <1% but varies with the procedure, the patient's age and comorbid conditions, and the experience of the surgical team. The most common surgical complications resulting from the RYGB procedure include stomal stenosis or marginal ulcers (occurring in 5–15% of patients) that present as prolonged nausea and vomiting after eating or inability to advance the diet to solid foods. These complications typically are treated by endoscopic balloon dilation and acid suppression therapy, respectively. For patients who undergo laparoscopic adjustable gastric banding, there are no intestinal absorptive abnormalities other than mechanical reduction in gastric size and outflow. Therefore, selective deficiencies are uncommon unless eating habits become unbalanced. In contrast, the restrictive-malabsorptive procedures carry an increased risk for micronutrient deficiencies of vitamin B_{12}, iron, folate, calcium, and vitamin D. Patients with restrictive-malabsorptive procedures require lifelong supplementation with these micronutrients.

NONSURGICAL ENDOSCOPIC TECHNIQUES

Endoscopic techniques offer potentially reversible and minimally invasive methods for treatment of obesity. These techniques include intraluminal gastric balloons, which are space-occupying devices that limit gastric volume, decrease hunger, and reduce food intake. The balloons generally need to be removed after 6 months, and produce a weight loss of 6–12% at the time of device removal. Adverse effects include nausea, vomiting, and abdominal pain. Gastric remodeling procedures such as the endoscopic sleeve gastroplasty (ESG), which induces early satiety and reduces gastric emptying, has demonstrated a weight loss of 18% at 2 years, comparable to the bariatric sleeve gastrectomy.

FURTHER READING

Apovian CM et al: Pharmacological management of obesity: An Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 100:342, 2015.

Chakhtoura M et al: Pharmacotherapy of obesity: An update on the available medications and drugs under investigation. Lancet 58:101882, 2023.

Garvey WT et al: American Association of Clinical Endocrinologists and American College of Endocrinology Comprehensive clinical practice guidelines for medical care of patients with obesity. Endocr Pract 22(Suppl 3):1, 2016.

Grunvald E et al: AGA clinical practice guideline on pharmacological interventions for adults with obesity. Gastroenterology 163:1198, 2022.

Jensen MD et al: 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Obesity Society. Circulation 129:S102, 2014.

Obesity Canada: Canadian Adult Obesity Clinical Guidelines (CPGs). Available at *https://obesitycanada.ca/guidelines/*. Accessed December 25, 2020.

415

Diabetes Mellitus: Diagnosis, Classification, and Pathophysiology

Alvin C. Powers, Kevin D. Niswender, Carmella Evans-Molina

Diabetes mellitus (DM) refers to a group of common metabolic disorders that share the phenotype of hyperglycemia. Several distinct types of DM are caused by a complex interaction of genetics and environmental factors. Depending on the etiology of the DM, factors contributing to hyperglycemia include reduced insulin secretion, decreased glucose utilization, and increased glucose production. The metabolic dysregulation associated with DM causes secondary pathophysiologic changes in multiple organ systems that impose a tremendous burden on the individual with diabetes and on the health care system. In the United States, DM is the leading cause of end-stage renal disease (ESRD), nontraumatic lower-extremity amputations, and adult blindness. Persons with diabetes are at increased risk for cardiovascular disease, which is the main cause of morbidity and mortality in this population.

CLASSIFICATION

DM is classified on the basis of the pathogenic process leading to hyperglycemia **(Table 415-1)**. There are two broad categories of DM, designated as either type 1 or type 2 DM. However, there is increasing recognition of other forms of diabetes in which the molecular pathogenesis is better understood and may be associated with a single-gene defect or a combination of genetic loci. These alternative forms, as well as other "atypical" forms, may share features of type 1 and/or type 2 DM. Type 1 DM develops as a result of autoimmunity against the insulin-producing beta cells, resulting in insulin deficiency. Type 2 DM is a heterogeneous group of disorders characterized by variable degrees of insulin resistance, impaired insulin secretion, and increased hepatic glucose production. Defects in insulin action and/or secretion give rise to the common phenotype of hyperglycemia in type 2 DM and have important therapeutic implications now that pharmacologic agents are available to target specific metabolic derangements. Both type 1 and type 2 DM are preceded by a period of progressive worsening of glucose homeostasis, followed by the development of hyperglycemia that exceeds the threshold for clinical diagnosis. In terms of type 2 DM, this phase is referred to as prediabetes and is more specifically classified as impaired fasting glucose (IFG) or impaired glucose tolerance (IGT) **(Fig. 415-1)**. Recently, three distinct stages of type 1 DM have been defined based on the development of autoantibodies against pancreatic beta cell antigens or the development of worsening dysglycemia (discussed below).

OTHER TYPES OF DM

Other etiologies of DM include specific genetic defects in insulin secretion or action, metabolic abnormalities that impair insulin secretion, mitochondrial abnormalities, and a host of conditions that impair glucose tolerance (Table 415-1). *Monogenic diabetes, sometimes referred to as maturity-onset diabetes of the young (MODY),* are types of DM characterized by autosomal dominant inheritance, early onset of hyperglycemia (usually <25 years; sometimes in neonatal period), and impaired insulin secretion (discussed below). Mutations in the insulin receptor cause a group of rare disorders characterized by severe insulin resistance.

DM may also develop as a result of cystic fibrosis or chronic pancreatitis, in which the islets become damaged from a primary pathologic process originating in the pancreatic exocrine tissue. Hormones that antagonize insulin action can lead to DM. Hyperglycemia is often a feature of endocrinopathies such as acromegaly and Cushing's disease. Viral infections have been implicated in pancreatic islet destruction but are an extremely rare cause of DM. A form of acute onset of type 1

TABLE 415-1 Etiologic Classification of Diabetes Mellitus

I. Type 1 DM (immune-mediated beta cell destruction, usually leading to absolute insulin deficiency)

II. Type 2 DM (may range from predominantly insulin resistance with relative insulin deficiency to a predominantly insulin secretory defect with insulin resistance)

III. Other types of diabetes caused by:

- A. Genetic defects of beta cell development or function characterized by mutations such as (e.g., monogenic, MODY):
  1. Hepatocyte nuclear transcription factor (HNF) 4α
  2. Glucokinase
  3. HNF-1α
  4. Insulin promoter factor-1, HNF-1β, NeuroD1, and other pancreatic islet regulators/proteins such as *KLF11, PAX4, BLK, GATA4, GATA6, SLC2A2* (GLUT2), *GLIS3*
  5. Mitochondrial DNA
- B. Neonatal diabetes (onset <6 months of age) due to mutations in subunits of ATP-sensitive potassium channel, *RFX6*, insulin
- C. Transient neonatal diabetes
- D. Diseases of the exocrine pancreas—pancreatitis, pancreatectomy, neoplasia, cystic fibrosis, hemochromatosis, fibrocalculous pancreatopathy, mutations in carboxyl ester lipase
- E. Genetic defects in insulin action, including type A insulin resistance, leprechaunism, Rabson-Mendenhall syndrome, lipodystrophy syndromes
- F. Endocrinopathies—acromegaly, Cushing's syndrome, glucagonoma, pheochromocytoma, hyperthyroidism, somatostatinoma, aldosteronoma
- G. Drug- or chemical-induced—glucocorticoids, calcineurin and mTOR inhibitors (after organ transplantation), pentamidine, nicotinic acid, statins, HIV therapies (antiretroviral, protease inhibitors, nucleoside/nucleotide reverse transcriptase inhibitors), diazoxide, β-adrenergic agonists, thiazides, proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors, hydantoins, asparaginase, α-interferon, protease inhibitors, antipsychotics (second-generation atypicals and others), epinephrine, vacor (a rodenticide)
- H. Infections—congenital rubella, cytomegalovirus, coxsackievirus
- I. Uncommon forms of immune-mediated diabetes—"stiff-person" syndrome, anti-insulin receptor antibodies, treatment of malignancy with immune checkpoint inhibitors
- J. Other genetic syndromes sometimes associated with diabetes—Wolfram syndrome, Down's syndrome, Klinefelter's syndrome, Turner's syndrome, Friedreich's ataxia, Huntington's chorea, Laurence-Moon-Biedl syndrome, myotonic dystrophy, porphyria, Prader-Willi syndrome

IV. Gestational diabetes mellitus (GDM)

Abbreviation: MODY, maturity-onset diabetes of the young or monogenic diabetes; see text.

Source: Data from American Diabetes Association: Standards of medical care in diabetes–2014. Diabetes Care 37:S81, 2014 and other sources.

diabetes, termed *fulminant diabetes*, has been noted in Japan and may be related to viral infection of the islets.

GESTATIONAL DM

Glucose intolerance developing during the second or third trimester of pregnancy is classified as gestational diabetes mellitus (GDM). Insulin resistance is related to the metabolic and hormonal changes of pregnancy, during which the increased insulin demands and insulin resistance may lead to IGT or diabetes. The American Diabetes Association (ADA) recommends that diabetes diagnosed within the first trimester be classified as preexisting pregestational diabetes rather than GDM. In 2021, the International Diabetes Federation (IDF) estimated that 16% of pregnancies worldwide were affected by either GDM or preexisting DM. Most women with GDM revert to normal glucose tolerance postpartum but have a substantial risk (35–60%) of developing DM in the next 10–20 years. In addition, children born to a mother with GDM also have an increased risk of developing metabolic syndrome and type 2 DM later in life. Currently, the ADA recommends that women with a history of GDM undergo lifelong screening for the development of diabetes or prediabetes at least every 3 years.

		Hyperglycemia			
		Prediabetes*	Diabetes Mellitus		
	Normal glucose tolerance	Impaired fasting glucose or impaired glucose tolerance	Not insulin requiring	Insulin required for control	Insulin required for survival
			Symptoms of diabetes + random blood glucose concentration ≥11.1 mmol/L (200 mg/dL)[a]		
HbA1C	<5.6% (<41 mmol/mol)	5.7–6.4% (42–47 mmol/mol)	≥6.5%[b] (≥48 mmol/mol)		
FPG	<5.6 mmol/L (100 mg/dL)	5.6–6.9 mmol/L (100–125 mg/dL)	≥7.0 mmol/L (126 mg/dL)[c]		
2-h PG	<7.8 mmol/L (140 mg/dL)	7.8–11.0 mmol/L (140–199 mg/dL)	≥11.1 mmol/L (200 mg/dL)[d]		

FIGURE 415-1 Spectrum of glucose homeostasis and diagnosis of diabetes mellitus (DM). Glucose homeostasis is a spectrum from normal glucose tolerance (*left portion of figure*) to diabetes (*right portion of figure*) including type 1 DM, type 2 DM, specific types of diabetes, and gestational DM. The diagnostic criteria for diabetes are shown in the right portion of the figure and include the hemoglobin A_{1c} (HbA_{1c}), the fasting plasma glucose (FPG), and the 2-h plasma glucose (PG) after a glucose challenge. In most types of DM, the individual traverses from normal glucose tolerance to impaired glucose tolerance to overt diabetes (these should be viewed not as abrupt categories but as a spectrum). Changes in glucose tolerance may be bidirectional in some individuals or types of diabetes. For example, individuals with type 2 DM may return to the impaired glucose tolerance category with weight loss; in gestational DM, diabetes may revert to impaired glucose tolerance or normal glucose tolerance after delivery. [a]Random is defined as without regard to time since the last meal. [b]HbA_{1c} test should be performed in a laboratory using a method approved by the National Glycohemoglobin Standardization Program that can be related to the reference assay of the Diabetes Control and Complications Trial. Some point-of-care HbA_{1c} assays meet these criteria and can be used. [c]Fasting is defined as no caloric intake for at least 8 h. [d]The test should be performed using a glucose load containing the equivalent of 75 g of anhydrous glucose dissolved in water, not recommended for routine clinical use. Assessment of 1-h glucose may be helpful in diabetes risk prediction in individuals with cystic fibrosis or other forms of pancreatic disease. In the absence of unequivocal hyperglycemia and acute metabolic decompensation, the blood glucose criteria should be confirmed by repeat testing on a different day. These values do not apply to the diagnosis of gestational DM. *(Modified from American Diabetes Association Professional Practice Committee. 2. Diagnosis and classification of diabetes: Standards of Care in Diabetes—2024. Diabetes Care 47[Suppl. 1]:S20, 2024.)*

ATYPICAL DIABETES

It is increasingly recognized that some forms of diabetes have features of both type 1 and type 2 DM. These are distinct from monogenic forms (MODY) as they have not been linked to single-gene defects. The development of a type 2 DM phenotype before puberty, a type 2 DM phenotype in very lean individuals, or a type 1 diabetes phenotype in individuals with no autoantibodies are examples of atypical diabetes. An additional example is ketosis-prone diabetes, where individuals present with ketoacidosis but do not require long-term exogenous insulin therapy. Many of these individuals are African American or Asian in heritage. Mechanisms underlying atypical forms of diabetes are being actively studied.

EPIDEMIOLOGY AND GLOBAL CONSIDERATIONS

The worldwide prevalence of DM has risen dramatically over the past four decades, from an estimated 30 million cases in 1985 to 537 million in 2021 (prevalence of 10.5%) **(Fig. 415-2)**. Based on current trends, the IDF projects that 643 million adults aged 20–79 years will have diabetes by the year 2030 (prevalence of 11.3%) and 783 million by 2045 (prevalence of 12.2%) (see *http://www.idf.org/*). Although the prevalence of both type 1 and type 2 DM is increasing worldwide, the prevalence of type 2 DM is rising much more rapidly, presumably because of dietary changes and increasing obesity, reduced activity levels as countries become more industrialized, and aging of the population. The countries with the greatest number of individuals with diabetes in 2021 were China (140.9 million), India (74.2 million),

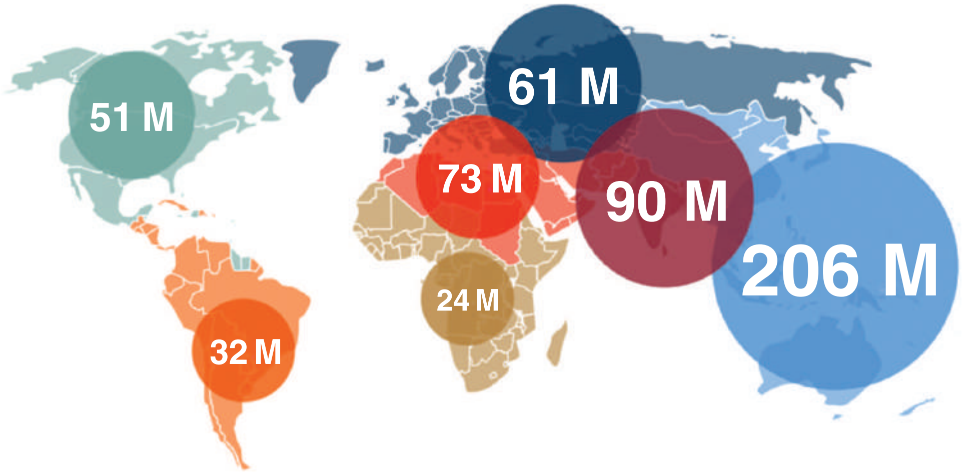

FIGURE 415-2 Worldwide prevalence of diabetes mellitus. Regional estimates of the number of individuals with diabetes (20–79 years of age) are shown (2021). *(Data from the IDF Diabetes Atlas, 10th ed. The International Diabetes Federation; 2021.)*

Pakistan (33.0 million), the United States (32.2 million), Indonesia (19.5 million), Brazil (15.7 million), and Mexico (14.1 million). In the most recent estimate for the United States (2021), the Centers for Disease Control and Prevention (CDC) estimated that 11.6% of the population had diabetes and as many as 38% of U.S. adults had prediabetes. Approximately 22.8% of adults with diabetes in the United States were undiagnosed. Globally, it is estimated that ~50% of those with diabetes who live in middle-income and low-income countries are undiagnosed. The prevalence of DM increases with age. For example, the DM prevalence in the United States was estimated to be 3.0% in individuals aged 18–44 years, 14.5% in persons aged 45–44 years, and 24.4% in those >65 years of age.

There is also considerable geographic variation in the incidence of both type 1 and type 2 DM. Currently, Scandinavia and the Middle East have the highest incidence of type 1 DM, with the lowest incidence in the Pacific Rim. Northern Europe and the United States have an intermediate rate. Much of the risk of type 1 DM is believed to reflect the frequency of high-risk human leukocyte antigen (HLA) alleles among ethnic groups in different geographic locations, but the type 1 DM incidence is increasing in non-European populations and in populations where the presence of the classic high-risk HLA alleles is less. The countries with the greatest number of individuals with type 1 DM, defined as diabetes onset at <19 years of age, are India, the United States, Brazil, and China. Adult-onset type 1 DM is increasingly recognized, but its incidence and prevalence are incompletely defined.

The prevalence of type 2 DM and its harbinger, IGT, is highest in certain Pacific islands and the Middle East and intermediate in countries such as India and the United States. This variability is likely due to genetic, behavioral, and environmental factors. DM prevalence also varies among different ethnic populations within a given country, with indigenous populations usually having a greater prevalence of diabetes than the general population of the country. For example, the CDC estimated that the age-adjusted prevalence of DM in the United States (age >20 years; 2019–2021) was 10.3% in non-Hispanic whites, 9.2% in Asians, 10.3% in Hispanics, 12.5% in non-Hispanic blacks, and 16% in Native American and Alaskan native populations. The onset of type 2 DM occurs, on average, at an earlier age in most non-Hispanic groups. In Asian populations (Southeast Asian and Western Pacific), the prevalence of type 2 DM is increasing rapidly, with onset at a lower body mass index (BMI) and younger age, greater visceral adiposity, and reduced insulin secretory capacity. Approximately 75% of individuals with diabetes live in low- or middle-income countries.

Diabetes is a major cause of mortality. In 2021, diabetes was listed as the eighth leading cause of death in the United States, but several studies indicate that diabetes-related deaths are likely underreported. In 2021, the IDF estimated that diabetes was responsible for nearly 6.7 million deaths worldwide, accounting for 12.2% of global all-cause mortality in adults aged 20–79 years. About one-third of diabetes-related deaths occur in individuals in their working years (<60 years of age). Diabetes also has important economic implications. In 2021, it was estimated that almost $1 trillion of health care expenditures worldwide were spent on individuals with diabetes.

DIAGNOSIS

Glucose tolerance, as assessed using hemoglobin A$_{1c}$ (HbA$_{1c}$), the fasting plasma glucose (FPG), or the response to oral glucose challenge, is classified into three broad categories: normal glucose homeostasis, impaired glucose homeostasis, or DM (Fig. 415-1). An HbA$_{1c}$ <5.7%, an FPG <5.6 mmol/L (100 mg/dL), and a plasma glucose <7.9 mmol/L (140 mg/dL) following an oral glucose challenge are considered to define normal glucose tolerance. An HbA1c ≥6.5% (> 48 mmol/mol), an FPG ≥7.0 mmol/L (126 mg/dL), or a glucose ≥11.1 mmol/L (200 mg/dL) 2 h after an oral glucose challenge meets the criteria for the diagnosis of DM (Fig. 415-1). A random plasma glucose concentration ≥11.1 mmol/L (200 mg/dL) accompanied by classic symptoms of DM (e.g., polyuria, polydipsia, weight loss) is also sufficient for the diagnosis of DM. DM is defined as the level of glycemia at which a diabetes-specific complication (e.g., retinopathy) occurs rather than a deviation from a population-based mean.

Impaired glucose homeostasis includes an HbA$_{1c}$ of 5.7–6.4%, IFG as defined as an FPG of 5.6–6.9 mmol/L (100–125 mg/dL), or IGT as defined as a plasma glucose level of 7.8–11 mmol/L (140–199 mg/dL) following an oral glucose challenge. The term *prediabetes* refers to those with an HbA$_{1c}$ of 5.7–6.4%, IFG, or IGT. These categories reflect the position of the American Diabetes Association (ADA). The World Health Organization defines IFG as an FPG >6.1 mmol/L (110 mg/dL). Although an HbA$_{1c}$ of 5.7–6.4%, IFG, and IGT do not identify the same individuals (i.e., different biologic mechanisms involved), individuals in all three groups are at greater risk of progressing to type 2 DM, have an increased risk of cardiovascular disease, and should be counseled about ways to decrease these risks. The risk for subsequent diabetes in individuals with prediabetes is related to the HbA$_{1c}$, with those with a value of 6.0–6.5% (42–48 mmol/mol) having a 25.5% 5-year risk of progressing to type 2 DM. It is important to recognize that HbA$_{1c}$ ≥6.5%, FPG, and the glucose level following an oral glucose challenge are continuous rather than discrete variables, with the risk for comorbidities increasing continuously rather than discretely by diagnostic category.

The current criteria for the diagnosis of DM emphasize the HbA$_{1c}$ and the FPG as the most reliable and convenient tests for identifying DM in asymptomatic individuals (Fig. 415-1). Also, it is important to note that some clinical conditions (e.g., hemoglobinopathies) and ancestry may influence the HbA$_{1c}$ **(Chap. 416)**. An oral glucose tolerance test (OGTT) is not often used in routine clinical care with the exception of pregnancy care and screening for gestational diabetes or cystic fibrosis–related diabetes **(Chap. 302)**. Continuous glucose monitoring (discussed in **Chap. 416**) is being examined as a method to diagnose diabetes, especially type 1 DM in its asymptomatic stages.

The diagnosis of DM has profound implications for an individual from both a medical and a financial standpoint. Thus, abnormalities on screening tests for diabetes should be repeated before making a definitive diagnosis of DM, unless acute metabolic derangements or a markedly elevated plasma glucose are present. These criteria also allow for the diagnosis of DM to be withdrawn in situations when the glucose intolerance reverts to normal.

■ SCREENING

Widespread use of HbA$_{1c}$ or FPG as a screening test for type 2 DM is recommended because (1) a large number of individuals who meet the

TABLE 415-2 Criteria for Screening for Type 2 DM in Adults

1. Consider testing in adults of any age who are overweight or obese (BMI ≥25 kg/m^2 or ≥23 kg/m^2 in Asian Americans) who have one of these risk factors:
   - Family history of diabetes (i.e., parent or sibling with type 2 DM)
   - Race/ethnicity (e.g., African American, Latino, Native American, Asian American, Pacific Islander)
   - Hypertension (blood pressure ≥130/80 mmHg)
   - HDL cholesterol level <35 mg/dL (0.90 mmol/L) and/or a triglyceride level >250 mg/dL (2.82 mmol/L)
   - Polycystic ovary syndrome or acanthosis nigricans
   - History of cardiovascular disease
   - Physical inactivity
   - Other condition associated with insulin resistance (severe obesity, acanthosis nigricans)
2. Individuals with previously identified IFG, IGT, or a hemoglobin A_{1c} of 5.7–6.4% should be screened annually.
3. Women who had GDM should be screened at least every 3 years.
4. For other individuals, initiate testing at 35 years of age and repeat every 3 years.
5. Individuals taking HIV medications or medications associated with diabetes (second-generation antipsychotic, glucocorticoids).
6. History of pancreatitis (screen within 3–6 months of episode of pancreatitis).
7. Cystic fibrosis (beginning at age 10); oral GTT recommended.
8. After organ transplantation; oral GTT recommended.

Abbreviations: BMI, body mass index; GDM, gestational diabetes mellitus; GTT, glucose tolerance test; HDL, high-density lipoprotein; HIV, human immunodeficiency virus; IFG, impaired fasting glucose; IGT, impaired glucose tolerance.

Source: Data from American Diabetes Association: 2. Diagnosis and classification of diabetes: Standards of care in diabetes—2024. Diabetes Care 47:S20, 2024, and other sources.

current criteria for DM are asymptomatic and unaware that they have the disorder, (2) epidemiologic studies suggest that type 2 DM may be present for up to a decade before diagnosis, (3) some individuals with type 2 DM have one or more diabetes-specific complications at the time of their diagnosis, (4) treatment of type 2 DM may favorably alter the natural history of DM or complications related to diabetes, and (5) diagnosis of prediabetes should spur efforts for diabetes prevention **(Table 415-2)**. The ADA recommends screening all individuals aged >35 years every 3 years and screening individuals at an earlier age if they are overweight and have one additional risk factor for diabetes. Screening for cystic fibrosis–related diabetes (beginning at age 10 years) and posttransplantation diabetes should use OGTT rather than HbA_{1c} **(Chap. 302)**. Screening for GDM is not discussed in this chapter. While not discussed here, the rapid increase in type 2 DM incidence and prevalence in children and adolescents, especially in racial and ethnic minority populations, has led to recommendations for screening and therapies.

REGULATION OF GLUCOSE HOMEOSTASIS

OVERALL REGULATION OF GLUCOSE HOMEOSTASIS

Glucose homeostasis reflects a balance between energy intake from ingested food, hepatic glucose production (gluconeogenesis), and peripheral tissue glucose uptake and utilization. Insulin is the most important regulator of this metabolic equilibrium, but neural input, metabolic signals, and other hormones (e.g., glucagon) result in integrated control of glucose supply and utilization **(Fig. 415-3)**. The organs that regulate glucose and lipids communicate by neural and humoral mechanisms with fat and muscle, producing adipokines, myokines, and metabolites that influence liver function. In the fasting state, low insulin levels, together with modest increases in glucagon, increase glucose production by promoting hepatic gluconeogenesis and glycogen breakdown (glycogenolysis). In parallel, glucose uptake in insulin-sensitive tissues (skeletal muscle and fat) is reduced, and there is increased mobilization of gluconeogenic precursors such as amino acids and free fatty acids (lipolysis). Under normal conditions, alpha cells increase glucagon secretion only when blood glucose or insulin levels are low or during exercise, but it is increased in fasting and postprandially in DM and stimulates excess glycogenolysis and gluconeogenesis by the liver and to a small degree by the renal medulla **(Chap. 418)**. Conversely, in healthy people, the postprandial glucose load elicits a rise in insulin and fall in glucagon, leading to optimized glucose disposal. Insulin, an anabolic hormone, promotes the storage of carbohydrate and fat and protein synthesis. The major portion of postprandial glucose is utilized by skeletal muscle, an effect of insulin-stimulated glucose uptake. Other tissues, most notably the brain, utilize glucose in an insulin-independent fashion. Factors secreted by skeletal myocytes, adipocytes (e.g., leptin, resistin, adiponectin), and bone also influence glucose homeostasis.

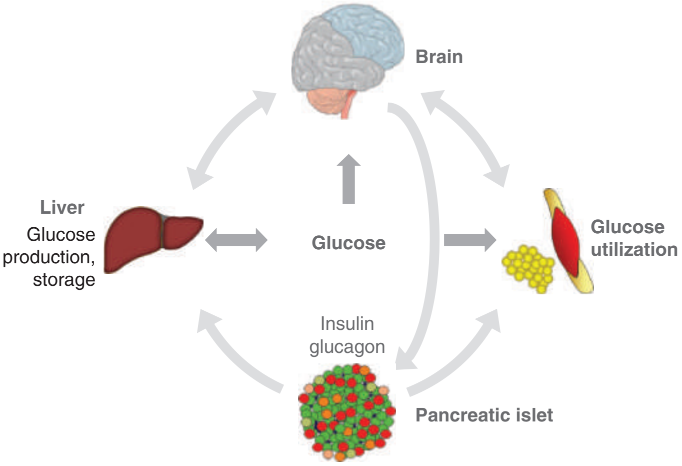

FIGURE 415-3 Regulation of glucose homeostasis. The organs shown contribute to glucose utilization, production, or storage. See text for a description of the communications (*arrows*), which can be neural or humoral. Although not shown, the gastrointestinal tract and bone produce factors that influence glucose homeostasis.

INSULIN BIOSYNTHESIS

Insulin, produced by the beta cells of the pancreatic islets, is initially synthesized as a single-chain 86-amino-acid precursor polypeptide, preproinsulin. Subsequent proteolytic processing removes the amino-terminal signal peptide, giving rise to proinsulin. Proinsulin is structurally related to insulin-like growth factors I and II, which bind weakly to the insulin receptor. Cleavage of an internal 31-residue fragment from proinsulin generates C-peptide with the A (21 amino acids) and B (30 amino acids) chains of insulin being connected by disulfide bonds. The mature insulin molecule and C-peptide are stored together and co-secreted from secretory granules in the beta cells. Because C-peptide is cleared more slowly than insulin, it is a useful marker of insulin secretion and allows discrimination of endogenous and exogenous sources of insulin in the evaluation of hypoglycemia **(Chaps. 89 and 418)**. Elevated levels of serum proinsulin have been observed in both type 1 and 2 DM and are thought to be indicative of beta cell dysfunction. Pancreatic beta cells co-secrete islet amyloid polypeptide (IAPP) or amylin, a 37-amino-acid peptide, along with insulin. The role of IAPP in normal physiology is incompletely defined, but it is the major component of the amyloid fibrils found in the islets of patients with type 2 DM, and an analogue is sometimes used in treating type 1 and type 2 DM **(Chap. 416)**.

INSULIN SECRETION

Glucose is the key regulator of insulin secretion by the pancreatic beta cell, although amino acids, ketones, various nutrients, gastrointestinal peptides, and neurotransmitters also influence insulin secretion. Glucose levels >3.9 mmol/L (70 mg/dL) stimulate insulin synthesis, primarily by enhancing protein translation and processing. Insulin secretion occurs in two phases, a rapid first-phase response, and a more prolonged second phase. Impaired first-phase insulin responses are among the earliest detectable abnormalities during the progression in both type 1 and type diabetes. Glucose stimulation of insulin secretion begins with its transport into the beta cell by a facilitative glucose transporter **(Fig. 415-4)**. Glucose phosphorylation by glucokinase is

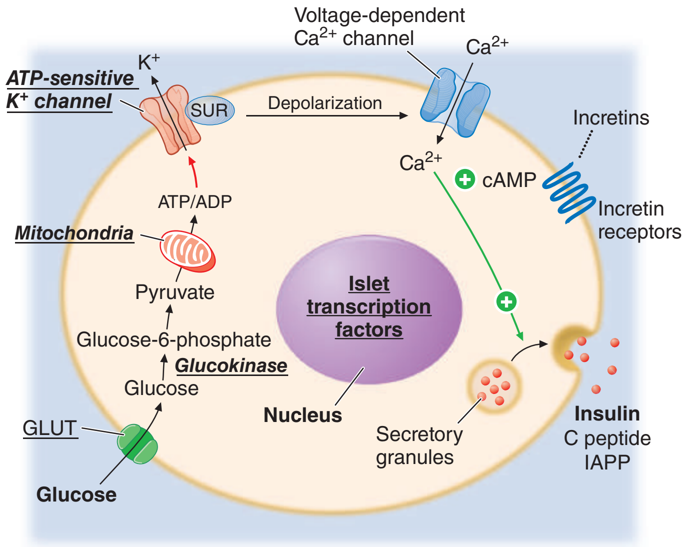

FIGURE 415-4 Mechanisms of glucose-stimulated insulin secretion and abnormalities in diabetes. Glucose and other nutrients regulate insulin secretion by the pancreatic beta cell. Glucose is transported by a glucose transporter (GLUT1 and/or GLUT2 in humans, GLUT2 in rodents); subsequent glucose metabolism by the beta cell alters ion channel activity, leading to insulin secretion. The SUR receptor is the binding site for some drugs that act as insulin secretagogues. Mutations in the events or proteins underlined are a cause of monogenic forms of diabetes. ADP, adenosine diphosphate; ATP, adenosine triphosphate; cAMP, cyclic adenosine monophosphate; IAPP, islet amyloid polypeptide or amylin; SUR, sulfonylurea receptor.

the rate-limiting step that controls glucose-regulated insulin secretion. Further metabolism of glucose-6-phosphate via glycolysis generates ATP, which inhibits the activity of an ATP-sensitive K^+ channel. This channel consists of two separate proteins: one is the binding site for certain oral hypoglycemics (e.g., sulfonylureas, meglitinides); the other is an inwardly rectifying K^+ channel protein (Kir6.2). Inhibition of this K^+ channel induces beta cell membrane depolarization, which opens voltage-dependent calcium channels (leading to an influx of calcium) and stimulates insulin secretion. An alternative model of stimulated insulin secretion does not utilize oxidative phosphorylation, but instead proposes that glycolysis, pyruvate kinase, and spatially confined glucose metabolism impact the K^+ channel protein. A number of metabolic pathways internal to the beta cell as well as external cues amplify glucose-stimulated insulin secretion. For example, calcium release from the endoplasmic reticulum and other internal stores contributes to the rise in cytosolic calcium in response to glucose and contributes to the insulin secretory response. Glucagon-like peptide 1 (GLP-1) and glucose-dependent insulinotropic peptide (GIP) are incretin hormones that bind specific receptors on the beta cell to stimulate insulin secretion through cyclic AMP production, but they have this effect only when the blood glucose is above the fasting level. Incretin hormones also suppress glucagon production and secretion. Incretin analogues or pharmacologic agents that prolong the activity of endogenous GLP-1 are used to treat type 2 DM. Classically, GLP-1 release was thought to occur solely from neuroendocrine L-cells of the gastrointestinal tract following food ingestion. However, recent preclinical studies suggest that intraislet production of GLP-1 from alpha cells may play a role in the regulation of insulin secretion.

INSULIN ACTION

Insulin is secreted into the portal venous system and acts to suppress endogenous hepatic glucose production and increase hepatic glucose uptake. A large portion (50%) of secreted insulin is cleared by the liver in this first pass, yielding a portal-to-peripheral insulin concentration gradient of ~2:1, with important implications for the clinical use of exogenous insulin **(Chap. 416)**. Uncleared insulin enters the systemic circulation where it binds to receptors in peripheral target tissues such as skeletal muscle and adipose. Insulin binding to its receptor stimulates intrinsic tyrosine kinase activity, leading to receptor autophosphorylation and the recruitment of intracellular signaling molecules, including the important insulin receptor substrates (IRS). IRS and other adaptor proteins initiate a complex cascade of phosphorylation and dephosphorylation reactions, resulting in the widespread metabolic and mitogenic effects of insulin. As an example, activation of the phosphatidylinositol-3′-kinase (PI3-kinase) pathway stimulates translocation of a facilitative glucose transporter (e.g., GLUT4) to the cell surface, an event that is crucial for glucose uptake by skeletal muscle and fat. Activation of other insulin receptor signaling pathways induces glycogen synthesis, protein synthesis, lipogenesis, and regulation of various genes in insulin-responsive cells.

PATHOGENESIS OF DM

TYPE 1 DM

Type 1 DM is the result of interactions of genetic, environmental, and immunologic factors that ultimately lead to immune-mediated destruction of the pancreatic beta cells and insulin deficiency. Individuals with type 1 DM, especially children, often have dramatic and acute symptoms of diabetes with marked elevations in the blood glucose, with 25–50% presenting with diabetic ketoacidosis (DKA). Type 1 DM can develop at any age. While classically thought as having its onset in prepuberty and adolescence, recent data estimate that as many as 40% of individuals who develop type 1 DM have onset at over the age of 30. Many individuals with adult-onset type 1 DM are initially misdiagnosed with type 2 DM, highlighting the need for vigilance when classifying the individual's diabetes. Most, but not all, individuals with type 1 DM have evidence of islet-directed autoimmunity, which is detected by the presence of autoantibodies against islet cell antigens in the blood. The presence of two or more autoantibodies is now designated as stage 1 type 1 DM **(Fig. 415-5)**. The temporal decline of beta cell function and mass preceding the development of type 1 DM is shown schematically in Fig. 415-5 as a smooth, gradual reduction in beta cell mass and function. However, natural history studies suggest an initial loss of first-phase insulin response that occurs years in advance of the onset of Stage 3 type 1 DM. This is followed by a gradual reduction in insulin secretion, until ~1-2 years before the onset of Stage 3 type 1 DM, when there is a more accelerated decline in beta cell function,

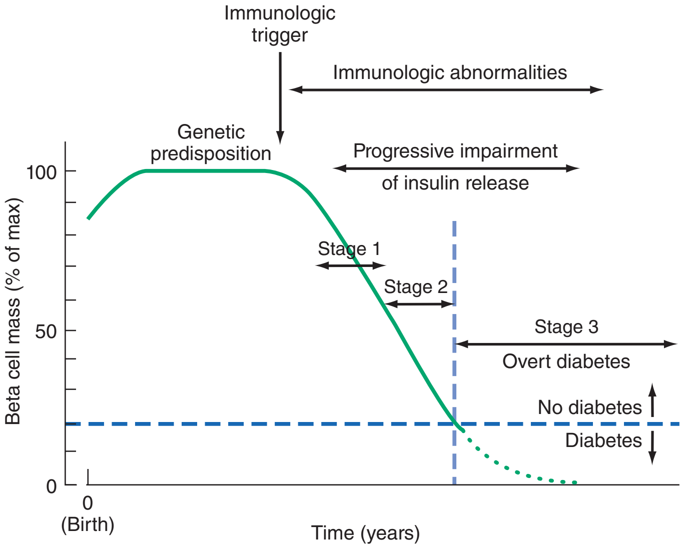

FIGURE 415-5 Temporal model for development of type 1 DM. Individuals with a genetic predisposition are exposed to a trigger that initiates an autoimmune process, resulting in the development of islet autoantibodies and a gradual decline in beta cell function and mass. Stage 1 disease is characterized by the development of two or more islet cell autoantibodies but the maintenance of normoglycemia. Stage 2 disease is defined by continued autoimmunity and the development of dysglycemia. Stage 3 is defined by the development of hyperglycemia that exceeds the diagnostic criteria for the diagnosis of diabetes. The downward slope of the beta cell function varies among individuals and may not be continuous. A "honeymoon" phase may be seen in the first 1 or 2 years after the onset of diabetes and is associated with reduced insulin requirements.

culminating in significant metabolic decompensation at the time of clinical diagnosis. In susceptible individuals, the autoimmune process is thought to be triggered by an infectious or environmental stimulus, but definitive proof is lacking. In the majority of patients, autoantibodies against islet cell antigens appear after this triggering event, followed by a progressive loss of insulin secretion. The time course between this triggering event and the development of Stage 3 type 1 DM varies widely among individuals, with some progressing rapidly to clinical diabetes and others evolving to diabetes more slowly, over a period of several years. Autopsy studies suggest the degree of loss of beta cell mass is variable at the time of disease presentation. At this point, residual functional beta cells exist but are insufficient in number and quality to maintain glucose tolerance. In some individuals, the events that trigger the transition from dysglycemia to frank hyperglycemia are often associated with increased insulin requirements, as might occur during infections or puberty. After the initial clinical presentation of type 1 DM, a "honeymoon" phase may ensue during which time glycemic control is achieved with modest doses of insulin or, rarely, insulin is not needed. However, this fleeting phase of endogenous insulin production from residual beta cells disappears and the individual becomes insulin deficient. Many individuals with long-standing type 1 DM produce a small amount of insulin (as reflected by C-peptide production), and autopsy studies show that beta cells can persist in the pancreas decades after diagnosis.

■ GENETIC CONSIDERATIONS

Susceptibility to type 1 DM involves multiple genes. The concordance of type 1 DM in identical twins ranges from 30–70%, indicating that additional modifying factors are likely involved in determining whether diabetes develops. The major susceptibility gene for type 1 DM is located in the HLA region on chromosome 6. Polymorphisms in the HLA complex account for ~50% of the genetic risk of developing type 1 DM. This region contains genes that encode the class II major histocompatibility complex (MHC) molecules, which present antigen to helper T cells and thus are involved in initiating the immune response **(Chap. 360)**. The ability of class II MHC molecules to present antigen is dependent on the amino acid composition of their antigen-binding sites. Refinements in genotyping of HLA loci have shown several genetic loci are predisposing (e.g., DRB1*0301-DQB1*0201 [DR3-DQ2] and DRB1*0401-DQB1*0302 [DR4-DQ8]). These haplotypes are present at greater frequency in those with type 1 DM than in the general population, although most individuals with predisposing haplotypes do not develop diabetes. An improved understanding of the genetics of type 1 diabetes has also identified protective HLA haplotypes (e.g., DRB1*1501 and DQA1* 0102-DQB1*).

In addition to MHC class II associations, genome-wide association studies have identified >60 additional genetic loci that contribute susceptibility to type 1 DM (e.g., polymorphisms in the promoter region of the insulin gene, the CTLA-4 gene, interleukin 2 receptor, and PTPN22). Combined assessment of HLA and non-HLA loci using genetic risk scores has been used to improve prediction of type 1 DM risk, with polygenic risk scores being applied in population-based screening studies. Notably, among recent cohorts of individuals with new-onset type 1 DM, there is a decreased representation of the highest-risk HLA alleles and increasing penetrance of disease in genotypes classically associated with lower risk, suggesting environmental factors may have an increasing role in disease pathogenesis.

Although the risk of developing type 1 DM is increased in relatives of individuals with the disease, the risk is relatively low: 1–9% if the parent has type 1 DM and 6–7% in a sibling (depending on which HLA haplotypes are shared). Hence, the majority of individuals with type 1 DM (>80%) do not have a relative with this disorder.

Pathophysiology Pathologically, the pancreatic islets demonstrate a modest infiltration of lymphocytes (a process termed *insulitis*); however, the frequency of insulitis is heterogeneous both within and between individuals. Studies of the autoimmune process have identified the following abnormalities in the innate and adaptive arms of the immune system: (1) islet cell autoantibodies (ICAs); (2) activated lymphocytes in the islets and peripancreatic lymph nodes; (3) T lymphocytes that proliferate when stimulated with islet proteins; and (4) release of cytokines within the insulitis. ICAs are distinct autoantibodies directed at pancreatic islet molecules such as insulin, glutamic acid decarboxylase (GAD), islet antigen 2 (IA-2), or zinc transporter 8 (ZnT8) and serve as a marker of the autoimmune process of type 1 DM. Testing for ICAs is useful in classifying the type of DM as type 1 as they are present in the majority of individuals (>85%) diagnosed with new-onset type 1 DM. ICAs can also identify nondiabetic individuals at risk for developing type 1 DM, although their use for this purpose has been restricted until recently mostly to research studies. In children with high genetic risk followed from birth as part of several cohort studies, the presence of two or more ICAs, which often appeared in the first 3 years of life, was associated with a nearly 70% risk of developing type 1 DM after 10 years of follow-up and a > 80% risk of developing diabetes after 15 years of follow-up. These observations led to a revision in the staging system for type 1 DM (Fig. 415-5), in which the development of multiple autoantibodies is now defined as the onset of stage 1 type 1 DM. While ICAs can be detected in the serum, and their presence is an important biomarker of type 1 DM risk, the antibodies do not have a direct role in beta cell death. Beta cell destruction is mostly likely mediated by direct CD8+ T cell–mediated cytotoxicity. Beta cells may exacerbate this process through the development of modified proteins or "neoantigens" and through increased presentation of these antigens on their cell surface via upregulation of MHC class I molecules. In addition, beta cells may be damaged by the toxic effects of cytokines (i.e., tumor necrosis factor α [TNF-α], interferon γ, and interleukin 1 [IL-1]) as well as reactive oxygen species generated by infiltrating immune cells. Efforts to suppress the autoimmune process at the time of diagnosis of diabetes have partially slowed beta cell destruction as reflected by a slower decline in C-peptide levels. There is now an increased emphasis on identifying high-risk individuals who have multiple ICAs in first-degree relatives of individuals with type 1 DM or in the general population to allow interventions earlier in the disease course (i.e., during stage 1 and 2 disease; Fig. 415-5). An example is a single 14-day course of teplizumab, an Fc receptor–nonbinding anti-CD3 monoclonal antibody, which delayed the onset of stage 3 type 1 DM in high-risk individuals with multiple autoantibodies and dysglycemia (i.e., stage 2 type 1 DM) by a median of 32.5 months. This treatment, recently approved by the U.S. Food and Drug Administration (FDA) for stage 2 type 1 DM, is being used by selected centers in some individuals with stage 2 type 1 DM. Screening for presymptomatic type 1 DM using ICAs is rapidly evolving and is becoming more widely available. If ICAs are detected, referral to a specialized center for testing for dysglycemia and possible treatment to delay clinical diabetes should be considered.

Measurement of ICAs is recommended for adults who have phenotypic features similar to those in classic type 1 DM (e.g., younger age at diagnosis, ketoacidosis, or rapid need for insulin treatment) as these individuals may have adult-onset type 1 DM. See additional discussion below in the section "Classification of DM in an Individual Patient."

Although other islet cell types (alpha cells [glucagon-producing], delta cells [somatostatin-producing], or PP cells [pancreatic polypeptide–producing]) are functionally and embryologically similar to beta cells, they are spared from the autoimmune destruction. However, altered patterns of hormone secretion from these other cell types in type 1 DM likely contribute to metabolic instability. Alpha cell dysfunction is reflected by fasting and postprandial hyperglucagonemia but an impaired glucagon response to hypoglycemia.

Environmental Factors Numerous environmental events have been proposed to trigger the autoimmune process in genetically susceptible individuals; however, none has been conclusively linked to diabetes. Identification of an environmental trigger has been difficult because the event may precede the onset of DM by several years (Fig. 415-5). Putative environmental triggers include viruses (coxsackie, rubella, and enteroviruses most prominently), bovine milk proteins, nitrosourea compounds, vitamin D deficiency, and environmental toxins. There is increasing interest in the microbiome and type 1 DM **(Chap. 484)**.

■ TYPE 2 DM

Type 2 DM likely encompasses a range of disorders with the common phenotype of hyperglycemia. Insulin resistance and abnormal insulin secretion are central to the development of type 2 DM. Although the primary defect is controversial, most studies support the view that insulin resistance precedes an insulin secretory defect but that diabetes develops only when insulin secretion becomes inadequate for the degree of insulin resistance. Historically, our understanding of the pathophysiology and genetics is mostly based on studies of individuals of European descent. Studies in more diverse populations are beginning to yield new insights into pathophysiologic differences among ethnic groups. In general, Latinos have greater insulin resistance and East Asians and South Asians have more beta cell dysfunction, but both defects are present in both populations. East and South Asians appear to develop type 2 DM at a younger age and a lower BMI. In some groups, DM that is ketosis prone (often in obese individuals) or ketosis resistant (often lean) is sometimes seen. For example, African Americans can be more prone to nonketotic hyperosmolar presentation of diabetes exacerbations. In many forms of type 2 DM around the world, the social determinants of health play a major role in the rates of type 2 DM.

■ GENETIC CONSIDERATIONS

Type 2 DM has a strong genetic component. The concordance of type 2 DM in identical twins is between 70 and 90%. Individuals with a parent with type 2 DM have an increased risk of diabetes; if both parents have type 2 DM, the risk approaches 70%. Insulin resistance, as demonstrated by reduced glucose utilization in skeletal muscle, is present in many nondiabetic, first-degree relatives of individuals with type 2 DM. The disease is polygenic and multifactorial, because in addition to genetic susceptibility, environmental factors (e.g., obesity, poor nutrition, reduced physical inactivity, increasing age, and socioeconomic status) modulate the frequency and the phenotype. Shared environmental and lifestyle factors also contribute to the high concordance in families. In addition, the in utero environment contributes, with either increased or reduced birth weight increasing the risk of type 2 DM in adult life. Children of pregnancies complicated by gestational hyperglycemia also exhibit an increased risk of type 2 DM.

The genes that predispose to type 2 DM are incompletely identified, but genome-wide association studies have identified a large number of genetic loci (>600) that convey a relatively small risk for type 2 DM (relative risk of 1.06–1.5). Most prominent is a variant of the transcription factor 7-like 2 gene that has been associated with both type 2 DM and IGT in several populations. The mechanisms by which these genetic loci, which are found mostly in noncoding regions of the genome, increase the susceptibility to type 2 DM are not clear, but most are predicted to alter islet function or development or insulin secretion. Although the genetic susceptibility to type 2 DM is under active investigation (it is estimated that <10% of genetic risk is determined by loci identified thus far), it is currently not possible to use a combination of known genetic loci to reliably predict type 2 DM.

Pathophysiology Type 2 DM is characterized by impaired insulin secretion, insulin resistance, excessive hepatic glucose production, abnormal fat metabolism, and systemic low-grade inflammation. Obesity, particularly visceral or central (as evidenced by the hip-waist ratio), is very common in type 2 DM (≥80% of patients are obese). In the early stages of the disorder, glucose tolerance remains near-normal, despite insulin resistance, because the pancreatic beta cells compensate by increasing insulin output **(Fig. 415-6)**. A number of pathophysiologic mechanisms contribute to type 2 DM, and their relative importance varies from individual to individual. As insulin resistance and compensatory hyperinsulinemia progress, the pancreatic islets in certain individuals are unable to sustain the hyperinsulinemic state, manifesting as IGT, defined as elevations in postprandial glucose. A decline in insulin secretion and/or increased glucagon secretion cause an increase in hepatic glucose production leading to fasting hyperglycemia. Ultimately, frank beta cell failure ensues as a combination of these mechanisms leading to the manifestation of type 2 DM.

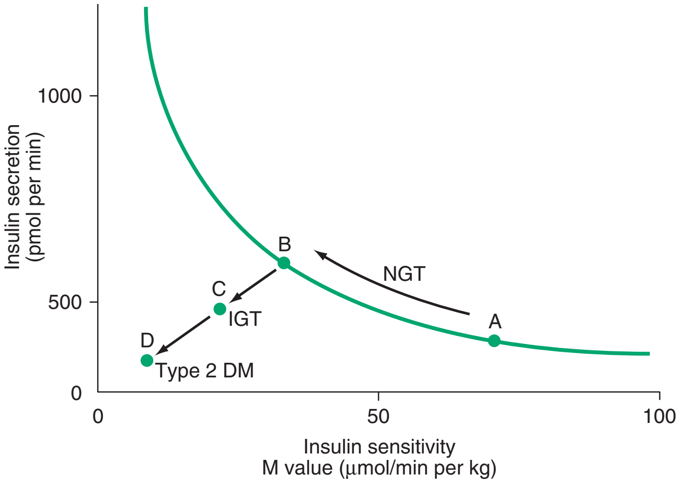

FIGURE 415-6 Metabolic changes during the development of type 2 DM. Insulin secretion and insulin sensitivity are related, and as an individual becomes more insulin resistant (by moving from point A to point B), insulin secretion increases. A failure to compensate by increasing the insulin secretion results initially in impaired glucose tolerance (IGT; point C) and ultimately in type 2 DM (point D). NGT, normal glucose tolerance. *(Data from SE Kahn: Clinical review 135: The importance of beta-cell failure in the development and progression of type 2 diabetes. J Clin Endocrinol Metab 86:4047, 2001 and RN Bergman, M Ader: Free fatty acids and pathogenesis of type 2 diabetes mellitus. Trends Endocrinol Metab 11:351, 2000.)*

Metabolic Abnormalities Insulin resistance, the decreased ability of insulin to act effectively on target tissues (especially muscle, liver, and fat), is a prominent feature of type 2 DM and results from a combination of genetic susceptibility, obesity, and metabolic inflammation. Insulin resistance is relative, however, because supranormal levels of circulating insulin will normalize the plasma glucose. In type 2 DM, glucose utilization is 30–60% lower than in nondiabetic individuals. Insulin resistance impairs glucose utilization by insulin-sensitive tissues (skeletal muscle) and, in liver, coupled with elevated glucagon, leads to increased hepatic glucose output. Increased hepatic glucose output predominantly accounts for increased FPG levels, whereas decreased peripheral glucose utilization results in postprandial hyperglycemia.

The precise molecular mechanism leading to insulin resistance in type 2 DM has not been elucidated. Insulin receptor levels and tyrosine kinase activity in skeletal muscle are reduced, but these alterations are most likely secondary to hyperinsulinemia and are not a primary defect. Therefore, “postreceptor” defects in insulin-regulated phosphorylation/dephosphorylation appear to play the predominant role in insulin resistance. Abnormalities include the accumulation of lipid intermediates within skeletal myocytes, which may impair mitochondrial oxidative phosphorylation and reduce insulin-stimulated mitochondrial ATP production. Impaired fatty acid oxidation and lipid accumulation within skeletal myocytes also may generate reactive oxygen species such as lipid peroxides. These and other mechanisms also generate low-grade metabolic inflammation that feeds back and directly worsens insulin resistance. Of note, not all insulin signal transduction pathways are resistant to the effects of insulin (e.g., those controlling cell growth and differentiation using the mitogenic-activated protein kinase pathway). Consequently, hyperinsulinemia may increase the insulin action through these pathways, potentially accelerating diabetes-related conditions such as atherosclerosis.

The obesity accompanying type 2 DM, particularly in a central or visceral location, is thought to be part of the pathogenic process **(Chap. 413)**. In addition to these white fat depots, humans have brown fat, which has much greater thermogenic capacity. Efforts are underway to increase the activity or quantity of brown fat. The increased adipocyte mass leads to increased levels of circulating free fatty acids and other fat cell products. For example, adipocytes secrete a number of biologic products (nonesterified free fatty acids, retinol-binding protein 4, leptin, TNF-α, resistin, IL-6, and adiponectin). Further, adipose resident macrophages are an important source of metabolic inflammation in diabetes. In addition to regulating body weight, appetite, and energy expenditure, adipokines also modulate insulin sensitivity. The increased production of free fatty acids and some adipokines may cause insulin resistance in skeletal muscle and liver. The venous drainage of the visceral adipose beds is the portal circulation, and this likely contributes to hepatic dysfunction. Free fatty acids also impair glucose utilization in skeletal muscle, promote glucose production by the liver, and impair beta cell function. In contrast, the production by adipocytes of adiponectin, an insulin-sensitizing peptide, is reduced in obesity, and this may contribute to hepatic insulin resistance. Adipocyte products and adipokines also produce an inflammatory state and may explain why markers of inflammation such as IL-6 and C-reactive protein are often elevated in type 2 DM.

IMPAIRED INSULIN SECRETION Insulin secretion and sensitivity are interrelated (Fig. 415-6). In type 2 DM, insulin secretion initially increases in response to insulin resistance to maintain normal glucose tolerance. Initially, the insulin secretory defect is mild and selectively involves glucose-stimulated insulin secretion, including a greatly reduced first secretory phase. The response to other nonglucose secretagogues, such as arginine, is preserved, but overall beta cell function is reduced by as much as 50% at the onset of type 2 DM. Abnormalities in proinsulin processing are reflected by increased secretion of proinsulin in type 2 DM. Eventually, the insulin secretory defect is progressive.

The reason(s) for the decline in insulin secretory capacity in type 2 DM is unclear. The assumption is that a second genetic defect—superimposed upon insulin resistance—leads to defects in beta cell function, mass, and potentially cellular identity and differentiation status. Islet amyloid polypeptide or amylin, co-secreted by the beta cell, forms amyloid fibrillar deposits found in the islets of individuals with long-standing type 2 DM. Whether such islet amyloid deposits are a primary or secondary event is not known. The metabolic environment of diabetes also negatively impacts islet function. For example, chronic hyperglycemia paradoxically impairs islet function (“glucose toxicity”) and leads to a worsening of hyperglycemia. Improvement in glycemic control is often associated with improved islet function, an important clinical consideration. In addition, elevated levels of free fatty acids (“lipotoxicity”) and systemic and local elevations in proinflammatory cytokines from increased numbers of islet-associated macrophages may also worsen islet function.

INCREASED HEPATIC GLUCOSE AND LIPID PRODUCTION In type 2 DM, insulin resistance in the liver reflects the failure of hyperinsulinemia to suppress gluconeogenesis, which results in fasting hyperglycemia and decreased glycogen storage by the liver in the postprandial state. Increased hepatic glucose production occurs early in the course of diabetes, although likely after the onset of insulin and glucagon secretory abnormalities and insulin resistance in skeletal muscle. As a result of insulin resistance in adipose tissue, lipolysis and free fatty acid flux from adipocytes are increased and efficiently cleared by liver, leading to increased very-low-density lipoprotein (VLDL)-triglyceride synthesis in hepatocytes and secretion from liver. This is also responsible for the dyslipidemia found in type 2 DM (elevated triglycerides, reduced high-density lipoprotein [HDL], and increased small dense low-density lipoprotein [LDL] particles). If this lipid is retained, steatosis in the liver may lead to metabolic dysfunction-associated steatotic liver disease (MASLD) and abnormal liver function tests.

Insulin Resistance Syndromes The insulin resistance condition comprises a spectrum of disorders, with hyperglycemia representing one of the most readily diagnosed features. The *metabolic syndrome* and the *insulin resistance syndrome* are terms used to describe a constellation of metabolic derangements that includes insulin resistance, hypertension, dyslipidemia (decreased HDL and elevated triglycerides), central or visceral obesity, type 2 DM or IGT/IFG, and accelerated cardiovascular disease. This syndrome is discussed in Chap. 420.

A number of relatively rare forms of severe insulin resistance include features of type 2 DM or IGT (Table 415-1). Mutations in the insulin receptor that interfere with binding or signal transduction are a rare cause of insulin resistance. Acanthosis nigricans and signs of hyperandrogenism (hirsutism, acne, and oligomenorrhea in women) are also common physical features. Two distinct syndromes of severe insulin resistance have been described in adults: (1) type A, which affects young women more severely and is characterized by severe hyperinsulinemia, obesity, and features of hyperandrogenism; and (2) type B, which affects middle-aged women and is characterized by severe hyperinsulinemia, features of hyperandrogenism, and autoimmune disorders. Individuals with type A insulin resistance syndrome have an undefined defect in the insulin-signaling pathway; individuals with type B insulin resistance syndrome have autoantibodies directed at the insulin receptor. These receptor autoantibodies may block insulin binding or may stimulate the insulin receptor, leading to intermittent hypoglycemia.

Polycystic ovary syndrome (PCOS) is a common disorder that affects premenopausal women and is characterized by chronic anovulation and hyperandrogenism **(Chap. 404)**. Insulin resistance is seen in a significant subset of women with PCOS, and the disorder substantially increases the risk for type 2 DM, independent of the effects of obesity.

Lipodystrophies are a group of heterogeneous disorders characterized by selective loss of adipose tissue, leading to severe insulin resistance and hypertriglyceridemia. Lipodystrophies can be inherited or acquired and associated with variable degrees of adipose tissue loss.

Prevention Type 2 DM is preceded by a period of IGT or IFG, and a number of lifestyle modifications and pharmacologic agents prevent or delay the onset of DM. Individuals with prediabetes or increased risk of diabetes should be referred to a structured program to reduce body weight and increase physical activity and for consideration for treatment with a weight-lowering medication such as a GLP-1 agonist (discussed in Chap. 416). In addition, they should be screened for cardiovascular disease. The Diabetes Prevention Program (DPP) demonstrated that intensive changes in lifestyle (diet and exercise for 30 min/d five times/week) in individuals with IGT prevented or delayed the development of type 2 DM by 58% compared to placebo. This effect was seen in individuals regardless of age, sex, or ethnic group. In the same study, metformin prevented or delayed diabetes by 31% compared to placebo. The lifestyle intervention group lost 5–7% of their body weight during the 3 years of the study; the effects of the intervention persisted for at least 15 years. Studies in Finnish and Chinese populations noted similar efficacy of diet and exercise in preventing or delaying type 2 DM. A number of agents, including α-glucosidase inhibitors, metformin, thiazolidinediones, GLP-1 receptor pathway modifiers, SGLT-2 inhibitors, and orlistat, prevent or delay type 2 DM. Individuals with a strong family history of type 2 DM and individuals with IFG or IGT should be strongly encouraged to achieve a normal BMI and engage in regular physical activity. Pharmacologic therapy for individuals with prediabetes is currently undergoing a transformation because of the greater degree of weight loss with GLP-1 receptor agonists and the combined GLP-1 and glucose-dependent insulinotropic peptide (GIP) receptor agonist, tirzepatide. In addition, intensive diet interventions (very-low-calorie intake) have been effective.

OTHER FORMS OF DM

Genetically Defined, Monogenic Forms of DM Several monogenic forms of DM have been identified. Monogenic diabetes, sometimes referred to as MODY, are caused by mutations in genes encoding islet-enriched transcription factors, glucokinase, insulin, and other islet factors (Fig. 415-5; Table 415-1) and present with an autosomal dominant mode of transmission. For example, monogenic diabetes may be caused by mutations in hepatocyte nuclear transcription factor (HNF) 4α, HNF-1α, or HNF-1β. As their names imply, these transcription factors are expressed in the liver but also in other tissues, including the pancreatic islets and kidney. These factors most likely affect islet development, the expression of genes important in glucose-stimulated insulin secretion, or the maintenance of beta cell mass. For example, individuals with an HNF-1α mutation (MODY 3) have a progressive decline in glycemic control but may respond to sulfonylureas. In fact, some of these patients were initially thought to have type 1 DM but were later shown to respond to a sulfonylurea, and insulin was discontinued. Individuals with an HNF-1β mutation have progressive impairment of insulin secretion and hepatic insulin resistance and require insulin treatment with minimal response to sulfonylureas. These individuals often have other abnormalities such as renal cysts, mild pancreatic exocrine insufficiency, and abnormal liver function tests. Individuals with monogenic diabetes as a result of mutations in the glucokinase gene have mild-to-moderate but stable hyperglycemia that does not respond to oral hypoglycemic agents and otherwise does not require treatment, except possibly during pregnancy. Glucokinase catalyzes the formation of glucose-6-phosphate from glucose, a reaction that is important for glucose sensing by the beta cells (Fig. 415-5) and for glucose utilization by the liver. As a result of glucokinase mutations, higher glucose levels are required to elicit insulin secretory responses, thus altering the set point for insulin secretion. A form of monogenic diabetes is caused by mutations in pancreatic and duodenal homeobox 1, a transcription factor that regulates pancreatic development and insulin gene transcription. Homozygous inactivating mutations cause pancreatic agenesis, whereas heterozygous mutations may result in DM. Studies of populations with type 2 DM suggest that mutations in MODY-associated genes are a relatively uncommon cause of type 2 DM (<5%). Anyone with atypical features of type 1 or type 2 DM should be considered for genetic screening for monogenic diabetes.

Permanent neonatal diabetes (onset <6 months of age) is a heterogeneous group of disorders caused by genetic mutations that impact beta cell function and/or pancreatic development (Fig. 415-5). Affected individuals typically require treatment with insulin and exhibit phenotypic overlap with type 1 DM. Activating mutations in the ATP-sensitive potassium channel subunits (Kir6.2 and ABCC8) impair glucose-stimulated insulin secretion. Importantly, these individuals may respond to sulfonylureas and can be treated with these agents, allowing discontinuation of insulin therapy. Mutations in the transcription factor *GATA6* are the most common cause of pancreatic agenesis. Homozygous glucokinase mutations cause a severe form of neonatal diabetes, whereas mutations in mitochondrial DNA are associated with diabetes and deafness. A number of mutations identified in the coding sequence of the insulin gene have been found to interfere with proinsulin folding, processing, and bioactivity and are designated as mutant *Ins*-gene-induced diabetes of youth (MIDYs). Some of the neonatal diabetes syndromes are associated with a spectrum of neurologic dysfunction and a variety of extrapancreatic manifestations. Any individual who developed diabetes at <6 months of age should have genetic screening.

COVID-19 and DM Early in the COVID-19 pandemic, a marked connection between DM and SARS-CoV-2 infection was noted, with the presence of hyperglycemia being more common in those infected with the virus and the presence of DM worsening clinical outcomes in those with COVID (**Chap. 205**). Initial studies early in the COVID-19 pandemic suggested an increase in the incidence of type 1 DM. Although data are still accumulating, it is most likely that SARS-CoV-2 infection does not lead directly to beta cell destruction and type 1 DM or cause type 2 DM but rather increases the DM ascertainment and creates clinical conditions that lead to or worsen hyperglycemia as in other serious illnesses in individuals with undiagnosed DM.

Pancreatic Exocrine Disease and DM Several types of pancreatic disease are now recognized to increase the risk of diabetes, sometimes termed type 3 or 3c DM, pancreatic DM, or pancreatogenic DM. The features of this form of DM are distinct from type 1 and type 2 DM, with its pathogenesis likely involving pancreatic islet damage or destruction, leading to insulin and glucagon deficiency. There is an increasing recognition that the pancreatic endocrine and pancreatic exocrine compartments interact, but the exact nature of this communication is still unclear. An episode of pancreatitis should prompt screening for diabetes. Pancreatic cancer and new-onset type 2 DM have an association. As the survival with cystic fibrosis has improved, the prevalence of cystic fibrosis–related diabetes has risen to almost 50%. In most forms of pancreatic diabetes, insulin is the preferred therapy.

Other Forms of DM One rare form of diabetes listed in Table 415-1 is immune checkpoint–related diabetes. This form of insulin-deficient diabetes has a rapid onset of hyperglycemia, often with DKA, and occurs at a variable time after initiation of immune checkpoint therapy, usually anti-programmed cell death 1 (PD-1) or anti-programmed death ligand 1 (PD-L1) alone or in combination with anti-CTLA-4 (<1% of individuals treated). While there are some similarities to type 1 DM in terms of genetics and occasional presence of ICA, it likely has a different, as yet undetermined, pathogenesis. Insulin therapy is required.

APPROACH TO THE PATIENT

Diabetes Mellitus

Once the diagnosis of DM is made, attention should be directed to symptoms related to diabetes (acute and chronic) and classifying the type of diabetes. DM and its complications produce a wide range of symptoms and signs; those secondary to acute hyperglycemia may occur at any stage of the disease, whereas those related to chronic hyperglycemia typically begin to appear during the second decade of hyperglycemia (**Chap. 417**). Because of long delays in clinical recognition, individuals with previously undetected type 2 DM may present with chronic complications of DM at the time of diagnosis. The history and physical examination should assess for symptoms or signs of acute hyperglycemia and screen for chronic microvascular and macrovascular complications and conditions associated with DM (**Chap. 417**).

HISTORY

A complete medical history should be obtained with special emphasis on DM-relevant aspects such as current weight as well as any recent changes in weight, family history of DM and its complications, sleep history, risk factors for cardiovascular disease, exercise, smoking status, history of pancreatic disease, and ethanol use. Symptoms of hyperglycemia include polyuria, polydipsia, weight loss, fatigue, weakness, blurry vision, frequent superficial infections (vaginitis, fungal skin infections), and slow healing of skin lesions after minor trauma. Metabolic derangements relate mostly to hyperglycemia (osmotic diuresis) and to the catabolic state of the patient (urinary loss of glucose and calories, muscle breakdown due to protein degradation and decreased protein synthesis). Blurred vision results from changes in the water content of the lens and resolves as hyperglycemia is controlled.

In a patient with established DM, the initial assessment should include a review of symptoms at the time of the initial diabetes diagnosis. This is an essential part of the history that can help define whether the correct type of DM has been diagnosed. Special emphasis should be placed on prior diabetes care, including types of therapies tried, the nature of any intolerance to previous therapies, prior HbA_{1c} levels, self-monitoring glucose results, frequency of hypoglycemia (<3.0 mmol/L, <54 mg/dL), presence of DM-specific complications, and assessment of the patient's knowledge about diabetes, exercise, nutrition, and sleep history. Diabetes-related complications may afflict several organ systems, and an individual patient may exhibit some, all, or none of the symptoms related to the complications of DM (**Chap. 417**). In addition, the presence of DM-related comorbidities should be established (cardiovascular disease, hypertension, dyslipidemia). Pregnancy plans should be ascertained in women of childbearing age. The ADA recommends that all women of childbearing age be counseled about the importance of tight glycemic control (HbA_{1c} <6.5%) prior to conception.

PHYSICAL EXAMINATION

In addition to a complete physical examination, special attention should be given to DM-relevant aspects such as weight and BMI, retinal examination, orthostatic blood pressure, foot examination, peripheral pulses, and insulin injection sites. Because periodontal disease is more frequent in DM, the teeth and gums should also be examined.

An annual foot examination should (1) assess blood flow (pedal pulses), sensation (vibratory sensation [128-Hz tuning fork at the base of the great toe], the ability to sense touch with a monofilament [5.07, 10-g monofilament]), pinprick sensation, ankle reflexes, and nail care; (2) look for the presence of foot deformities such as hammer or claw toes and Charcot foot; and (3) identify sites of potential ulceration. The ADA recommends annual screening for distal symmetric polyneuropathy beginning with the initial diagnosis of diabetes and annual screening for autonomic neuropathy 5 years after diagnosis of type 1 DM and at the time of diagnosis of type 2 DM. This testing is aimed at detecting loss of protective sensation (LOPS) caused by diabetic neuropathy **(Chap. 417)**.

CLASSIFICATION OF DM IN AN INDIVIDUAL PATIENT

The etiology and classification of diabetes in an individual with new-onset disease is usually assigned based on clinical criteria. DM classification is critical in efforts to provide personalized therapy, but some individuals cannot be neatly classified as having type 1 or type 2 DM at the time of diagnosis. Even if type 1 DM is classified correctly, there is heterogeneity in terms of immune markers, insulin requirements, and the rapidity of beta cell loss. Long-used features such as age of onset, obesity, and the need for insulin treatment are often not helpful in DM classification, especially in adults with new-onset DM. Features such as younger age at diagnosis (<35 years), nonobese BMI, ketoacidosis, markedly elevated plasma glucose at diagnosis, and a family or personal history of other autoimmune disorders such as autoimmune thyroid disease, adrenal insufficiency, pernicious anemia, celiac disease, and vitiligo should prompt consideration of type 1 DM and measurement of ICA. However, these are only broad guidelines and may not be present in every individual with type 1 DM. For example, now many individuals with new-onset type 1 DM are overweight or obese. In contrast, individuals with type 2 DM often exhibit the following features: (1) obesity; 80% are obese, but elderly individuals and those from South and East Asia may be lean; (2) may not require insulin therapy initially; and (3) may have associated conditions such as insulin resistance, hypertension, cardiovascular disease, dyslipidemia, or PCOS. In type 2 DM, insulin resistance is often associated with abdominal obesity (as opposed to hip and thigh obesity) and hypertriglyceridemia. Although most individuals diagnosed with type 2 DM are older, the age of diagnosis is declining, and there is a marked increase among overweight children and adolescents. Some individuals with phenotypic type 2 DM present with DKA but lack autoimmune markers and may be later treated with oral glucose-lowering agents rather than insulin (this clinical picture is sometimes referred to as *ketosis-prone type 2 DM*). On the other hand, some individuals (5–10%) with the phenotypic appearance of type 2 DM do not initially have absolute insulin deficiency but have autoimmune markers (GAD and other ICAs) suggestive of type 1 DM (sometimes termed *autoimmune diabetes of adults*). Such individuals are more likely to require insulin treatment within 5 years of diabetes diagnosis and likely have a form of autoimmune diabetes. Monogenic forms of diabetes should be considered in those with diabetes onset in childhood or early adulthood and especially those diagnosed within the first 6 months of life those with an autosomal pattern of diabetes inheritance, diabetes without typical features of type 1 or 2 DM, and stable mild fasting hyperglycemia. Despite recent advances in the understanding of the pathogenesis of diabetes, it often remains difficult to categorize some patients unequivocally. There is increasing recognition that many forms of diabetes are atypical and do not fit in the type 1 or type 2 DM pattern. Individuals who deviate from the clinical profile of type 1 and type 2 DM, or who have other associated defects such as deafness, pancreatic exocrine disease (type 3c DM), and other endocrine disorders, should be classified accordingly (Table 415-1). While a major goal is personalized or precision medicine in the diagnosis and treatment of diabetes, this is currently an aspirational goal in many individuals with diabetes.

LABORATORY ASSESSMENT

The laboratory assessment should first determine whether the patient meets the diagnostic criteria for DM (Fig. 415-1) and then assess the degree of glycemic control **(Chap. 416)**. In addition to the standard laboratory evaluation, the patient should be screened for DM-associated conditions (e.g., albuminuria, dyslipidemia, thyroid dysfunction).

The classification of the type of DM may be facilitated by laboratory assessments. Serum C-peptide measurements may be useful but should always be interpreted with a concurrent blood glucose level. A low C-peptide in the setting of an elevated blood glucose level may confirm a patient's need for insulin. However, C-peptide levels are unable to completely distinguish type 1 from type 2 DM as many individuals with type 1 DM retain some C-peptide production. Measurement of islet cell antibodies at the time of diabetes onset may be useful if the type of DM is not clear based on the characteristics described above.

■ FURTHER READING

American Diabetes Association Professional Practice Committee: 2. Diagnosis and classification of diabetes: Standards of Care in Diabetes—2024. Diabetes Care 47:S20, 2024.

Buzzetti R et al: Adult-onset autoimmune diabetes. Nat Rev Dis Primers 22:63, 2022

Faruqi J, Balasubramanyam A: COVID-19 and diabetes mellitus: A review of the incidence, pathophysiology and management of diabetes during the pandemic. Expert Rev Endocrinol Metab 18:167, 2023.

Hill-Briggs F et al: Social determinants of health and diabetes: A scientific review. Diabetes Care 44:258, 2021.

International Diabetes Federation: *IDF Diabetes Atlas: 2021*, 10th ed. Available at *www.diabetesatlas.org*.

Ke C et al: Pathophysiology, phenotypes and management of type 2 diabetes mellitus in Indian and Chinese populations. Nat Rev Endocrinol 18:413, 2022.

Kreienkamp RJ et al: Genetics of type 2 diabetes, in *Diabetes in America*, Lawrence JM, et al (eds). Bethesda, Maryland, National Institute of Diabetes and Digestive and Kidney Diseases, 2023. Available at *https://www.ncbi.nlm.nih.gov/books/NBK597726/*.

Merrins MJ, Kibbey RG: Glucose regulation of β-cell KATP channels: It is time for a new model! Diabetes 73:856, 2024.

Powers AC: Type 1 diabetes mellitus: Much progress, many opportunities. J Clin Invest 131:142242, 2021.

Redondo MJ et al: Genetics of type 1 diabetes, in *Diabetes in America*, Lawrence JM et al (eds). Bethesda, Maryland, National Institute of Diabetes and Digestive and Kidney Diseases, 2023. Available at *https://www.ncbi.nlm.nih.gov/books/NBK597726/*.

Salguero MV et al: Monogenic forms of diabetes, in *Diabetes in America*, Lawrence JM et al (eds). Bethesda, Maryland, National Institute of Diabetes and Digestive and Kidney Diseases, 2023. Available at *https://www.ncbi.nlm.nih.gov/books/NBK597726/*.

Stone SI et al: Atypical diabetes: What have we learned and what does the future hold? Diabetes Care 47:770, 2024.

Tobias DK et al: Second international consensus report on gaps and opportunities for the clinical translation of precision diabetes medicine. Nat Med 29:2438, 2023.

416 Diabetes Mellitus: Management and Therapies

Alvin C. Powers, Kevin D. Niswender, Michael R. Rickels

OVERALL GOALS

The goals of therapy for all forms of diabetes mellitus (DM) are to (1) eliminate symptoms related to hyperglycemia, (2) reduce or eliminate the long-term microvascular and macrovascular complications of DM (Chap. 417), and (3) allow the patient to achieve as normal a lifestyle as possible. To reach these goals, the physician and health care team should identify a target level of glycemic control for each patient, provide the patient with the educational and pharmacologic resources necessary to reach this target, avoid hypoglycemia, and monitor/prevent/treat DM-related complications. Symptoms of diabetes usually resolve when the plasma glucose is <11.1 mmol/L (200 mg/dL), and thus most DM treatment focuses on achieving the second and third goals.

The care of an individual with either type 1 or type 2 DM requires a multidisciplinary team. Central to the success of this team are the patient's participation, input, and enthusiasm, all of which are essential for optimal diabetes management. Members of the health care team usually include the primary care provider and/or the endocrinologist or diabetologist, an advanced practice provider (APP), a pharmacist, a certified diabetes educator, a nutritionist, a behavioral health professional, and possibly a social worker. In addition, when the complications of DM arise, subspecialists (including ophthalmologists, neurologists, podiatrists, nephrologists, cardiologists, and cardiovascular and transplant surgeons) with experience in DM-related complications are essential. The American Diabetes Association (ADA) suggests applying the Chronic Care Model to diabetes with an emphasis on these elements: a proactive, team-based delivery and health system design that involves self-management, decision support with evidence-based guidelines for person-specific and population-based approaches, and community resources and policies that support healthy lifestyles. Space limitations do not allow a discussion of all these elements, so this chapter first reviews the ongoing treatment of diabetes in the outpatient setting and then discusses the treatment of severe hyperglycemia, as well as the treatment of diabetes in hospitalized patients.

ONGOING ASPECTS OF COMPREHENSIVE DIABETES CARE

A number of names are sometimes applied to different approaches to diabetes care, such as intensive insulin therapy or intensive glycemic control. The current chapter, and other sources, uses the term *Comprehensive diabetes care* to emphasize the fact that optimal diabetes therapy involves much more than glucose management and medications and that individualized, patient-centered care is essential. Although glycemic control is central to optimal diabetes therapy, comprehensive diabetes care of both type 1 and type 2 DM should also detect and modify risk factors for DM-associated disorders and manage DM-specific complications (Chap. 417). The key elements of comprehensive diabetes care are summarized in Table 416-1. The morbidity and mortality of DM can be greatly reduced by timely and consistent surveillance, including the detection, prevention, and management of DM-related complications (Table 416-1 and Chap. 417). Such approaches are indicated for all individuals with DM, but many individuals with diabetes do not receive these or comprehensive diabetes care. The social determinants of health and family, financial, cultural, and employment-related issues may negatively impact diabetes care. This chapter, while recognizing that resources available for diabetes care vary widely throughout the world, provides guidance for comprehensive diabetes care in health care settings with considerable societal resources.

TABLE 416-1 Guidelines for Ongoing, Comprehensive Medical Care for Individuals with Diabetes

- Individualized glycemic goal and therapeutic plan with an emphasis on shared decision-making
- Blood glucose measurement using continuous glucose monitoring (CGM) or capillary fingerstick device
- HbA_{1c} testing (2–4 times/year)
- Lifestyle management in the care of diabetes, including:
  - Diabetes self-management education and support
  - Nutrition therapy
  - Physical activity
  - Psychosocial care, including evaluation for depression, anxiety, diabetes distress
- Detection, prevention, or management of diabetes-related complications, including:
  - Diabetes-related eye examination (annual or biannual; Chap. 417)
  - Diabetes-related foot examination (1–2 times/year by provider; daily by patient; Chap. 415)
  - Diabetes-related neuropathy examination (annual; Chap. 415)
  - Diabetes-related kidney disease testing (annual; Chap. 417)
  - Screen for other diabetes-related complications (annual; see Table 417-1)
  - Assessment of fracture risk in older adults with diabetes (consider measurement of bone mineral density)
- Manage or treat diabetes-relevant conditions, including:
  - Blood pressure (assess 2–4 times/year; Chap. 417)
  - Lipids (1–2 times/year; Chap. 417)
  - Consider screening individuals with type 2 diabetes or prediabetes for metabolic dysfunction–associated steatotic liver disease if other risk factors are present
  - Consider antiplatelet therapy with low-dose aspirin (Chap. 417)
  - Immunizations, including influenza, pneumococcal, hepatitis B, coronavirus, and respiratory syncytial virus (>60 years of age) (Chap. 6)

Abbreviation: HbA_{1c}, glycated hemoglobin A_{1c}.

Patient-oriented websites also offer important resources for patients and their caregivers. Examples of these resources include: *https://www.tidepool.org/about*; *https://diatribe.org/understanding-diabetes/diabetes-technology*; and *https://pro.diabeteswise.org/en/*.

Lifestyle Management in Diabetes Care The patient with type 1 or type 2 DM should receive education about nutrition, physical activity, psychosocial support, care of diabetes during illness, medications used to control the glucose, methods for glucose monitoring, and strategies to prevent diabetes-related complications. Patient education allows and encourages individuals with DM to assume greater responsibility for their care, leading to improved compliance.

Diabetes Self-Management Education and Support (DSMES) DSMES refers to ways to improve the patient's knowledge, skills, and abilities necessary for diabetes self-care and should also emphasize psychosocial issues and emotional well-being. Patient education is a continuing process with regular visits for reinforcement; it is not a process completed after one or two visits. It should receive special emphasis at the diagnosis of diabetes, annually, or at times when diabetes treatment goals are not attained, and during transitions in life or medical care. DSMES is delivered by a diabetes educator who is a health care professional (nurse, dietician, or pharmacist) with specialized patient-education skills and who is certified in diabetes education (e.g., Association of Diabetes Care and Education Specialists or Certification Board for Diabetes Care and Education). Education topics important for optimal diabetes self-care include continuous glucose monitoring (CGM) or blood glucose monitoring (BGM); urine or blood ketone monitoring (type 1 DM); insulin administration; guidelines for diabetes management during illnesses; prevention and management of hypoglycemia (Chap. 418); foot and skin care; diabetes management before, during, and after exercise; and risk factor–modifying activities. The focus is providing patient-centered, individualized education. More frequent contact between the patient and the diabetes management team (e.g., electronic, telephone, video) improves glycemic control.

Nutrition Therapy *Medical nutrition therapy* (MNT) is a term used by the ADA to describe the optimal coordination of caloric intake with other aspects of diabetes therapy (e.g., insulin, exercise, and weight loss). Some aspects of MNT are directed at preventing or delaying the onset of type 2 DM in high-risk individuals (obese or with prediabetes) by promoting weight reduction. Other measures of MNT are directed at improving glycemic control through monitoring carbohydrate intake, avoiding simple sugars and fructose, and managing diabetes-related complications (atherosclerotic cardiovascular disease [ASCVD], nephropathy). Medical treatment of obesity, including pharmacologic approaches that facilitate weight loss and metabolic surgery, should be considered in some patients (Chaps. 413 and 414).

In general, the components of optimal MNT are similar for individuals with type 1 or type 2 DM—high-quality, nutrient-dense foods with limits on carbohydrate intake and weight management (Table 416-2). The data are currently inconclusive about various eating patterns (e.g., intermittent fasting). Sleep deprivation and shift work are risk factors for weight gain and insulin resistance. Dietary advice should be individualized, acknowledging personal preferences, cultural, and religious traditions. Use of the *glycemic index*, an estimate of the postprandial rise in the blood glucose when a certain amount of that food is consumed, may reduce postprandial glucose excursions and improve glycemic control.

The goal of MNT in type 1 DM is to coordinate and match the carbohydrate intake, both temporally and quantitatively, with the appropriate amount of insulin. MNT in type 1 DM is informed by CGM and/or BGM that should be integrated to define the optimal insulin regimen. Based on the patient's estimate of the carbohydrate content of a meal, an insulin-to-carbohydrate ratio determines the bolus insulin dose for a meal or snack. MNT must be flexible enough to allow for exercise, and the insulin regimen must allow for variations in caloric intake. An important component of MNT in type 1 DM is to minimize the weight gain often associated with intensive insulin therapy and is best achieved by placing limits on carbohydrate intake.

TABLE 416-2 Nutritional Recommendations for Adults with Diabetes or Prediabetes[a]

General dietary guidelines

- Vegetables, fruits, whole grains, legumes, low-fat dairy products and food higher in fiber and lower in glycemic content; optimal diet composition and eating pattens are not known.

Fat in diet (optimal percentage of diet is not known; should be individualized)

- Encourage Mediterranean-style diet rich in monounsaturated and polyunsaturated fatty acids.
- Minimal or no trans fat consumption.

Carbohydrate in diet (optimal percentage of diet is not known; should be individualized)

- Monitor carbohydrate intake in regard to calories and set limits for meals to reduce postprandial glycemia.
- Consider limiting overall carbohydrate intake in adults with diabetes as this may improve glycemia.
- Avoid fructose- and sucrose-containing beverages and minimize consumption of foods with added sugar that may displace healthier, more nutrient-dense food choices and elevate postprandial glycemia.
- Estimate grams of carbohydrate in diet for flexible insulin dosing (type 1 diabetes and insulin-dependent type 2 diabetes).
- Consider using glycemic index to predict how consumption of a particular food may affect blood glucose.

Protein in diet (optimal percentage of diet is not known; should be individualized)

Other components

- Reduced-calorie and nonnutritive sweeteners may be useful.
- Routine supplements of vitamins, antioxidants, or trace elements not supported by evidence.
- Vitamin D and calcium supplemental as recommended to promote bone health.
- Sodium intake as advised for general population (<2300 mg/d).
- Minimize disruption to sleep and eating patterns (chrononutrition), and note risk of hypoglycemia associated with religious fasting.

[a]See text for differences for patients with type 1 or type 2 diabetes.

Source: Data from American Diabetes Association: Facilitating positive health behaviors and well-being to improve health outcomes: Standards of care in diabetes—2024. Diabetes Care 47:S77, 2024.

The goals of MNT in type 2 DM should focus on weight loss and address the greatly increased prevalence of cardiovascular risk factors (hypertension, dyslipidemia, obesity) and disease in this population. The majority of these individuals are obese, and weight loss is strongly encouraged. Very-low-carbohydrate diets that induce weight loss may result in rapid and dramatic glucose lowering in individuals with new-onset type 2 DM. MNT for type 2 DM should emphasize modest caloric reduction, increased physical activity, and weight loss (goal of at least 5–10% loss). Weight loss and exercise each independently improve insulin sensitivity.

Fasting for religious reasons, such as during Ramadan, presents a challenge for individuals with diabetes, especially those taking medications to lower the plasma glucose. Under most guidelines for fasting during Ramadan, individuals are risk-stratified based on a pre-Ramadan risk assessment for people with diabetes as those who can safely fast with medical evaluation and supervision and those in whom fasting is not advised. Thus, patient education and regular glucose monitoring are critical.

Physical Activity Exercise has multiple positive benefits, including cardiovascular risk reduction, reduced blood pressure, maintenance of muscle mass, reduction in body fat, and weight loss. For individuals with type 1 or type 2 DM, exercise is also useful for lowering plasma glucose (during and following exercise) and increasing insulin sensitivity. In patients with diabetes, the ADA recommends 150 min/week (distributed over at least 3 days) of moderate aerobic physical activity with no gaps longer than 2 days. Resistance exercise, flexibility and balance training, and reduced sedentary behavior throughout the day are also advised.

Despite its benefits, exercise may present challenges for some individuals with DM because they lack the normal glucoregulatory mechanisms (normally, insulin falls and glucagon rises during exercise). Skeletal muscle is a major site for metabolic fuel consumption in the resting state, and the increased muscle activity during vigorous aerobic exercise greatly increases fuel requirements. Individuals with type 1 DM are prone to either hyperglycemia or hypoglycemia during exercise, depending on the pre-exercise plasma glucose, the circulating insulin level, lactate, and the level of exercise-induced catecholamines. If the insulin level is too low, the delivery of lactate to the liver and rise in catecholamines may increase the plasma glucose excessively, promote ketone body formation, and possibly lead to ketoacidosis. Conversely, if the circulating insulin level is excessive, this relative hyperinsulinemia may reduce hepatic glucose production (decreased glycogenolysis, decreased gluconeogenesis) and increase glucose entry into muscle, leading to hypoglycemia.

To avoid exercise-related hyper- or hypoglycemia, individuals with type 1 DM should (1) monitor blood glucose before, during, and after exercise (see below related to CGM and possible blood glucose discordance); (2) delay exercise if blood glucose is >14 mmol/L (250 mg/dL) and ketones are present; (3) if the blood glucose is <5.0 mmol/L (90 mg/dL), ingest carbohydrate before exercising; (4) monitor glucose during exercise and ingest carbohydrate as needed to prevent hypoglycemia; (5) decrease insulin doses (based on previous experience) before and after exercise and inject insulin into a nonexercising area; and (6) learn individual glucose responses to different types of exercise. In individuals with type 2 DM, exercise-related hypoglycemia is less common but can occur in individuals taking either insulin or insulin secretagogues. Untreated proliferative retinopathy is a relative contraindication to vigorous exercise because this may lead to vitreous hemorrhage or retinal detachment (Chap. 417).

Psychosocial Care Because the individual with DM faces challenges that affect many aspects of daily life, psychosocial assessment and support are a critical part of comprehensive diabetes care. The patient should view himself/herself/themself as an essential member of the diabetes care team and not as someone who is cared for by the diabetes management team. Even with considerable effort, normoglycemia can be an elusive goal, and solutions to worsening glycemic control may not be easily identifiable. Depression, anxiety, or "diabetes distress," defined by the ADA as "negative psychological reactions related to emotional burdens … in having to manage a chronic disease like diabetes," should be recognized and may require the care of a mental health specialist. Emotional stress may provoke a change in behavior so that individuals no longer adhere to a dietary, exercise, or therapeutic regimen. Eating disorders, including binge eating disorders, bulimia, and anorexia nervosa, appear to occur more frequently in individuals with type 1 or type 2 DM.

MONITORING THE LEVEL OF GLYCEMIC CONTROL

Optimal monitoring of glycemic control involves CGM or BGM (blood collected by fingerstick) by the patient and an assessment of long-term control by providers using measurement of hemoglobin A$_{1c}$ (HbA$_{1c}$). These measurements are complementary: the patient's measurements provide a picture of short-term glycemic control, whereas the HbA$_{1c}$ reflects average glycemic control over the previous 2–3 months. By integrating glycemic measurements with diet and exercise history into comprehensive diabetes care, the diabetes management team and patient can improve glycemic control and reduce diabetes-related complications.

Assessment of Short-Term Glycemic Control All individuals with diabetes should be offered a device to assess their short-term patterns of glycemia. CGM technology utilizes a sensor or electrode to detect interstitial glucose, which is in equilibrium with the blood glucose but may lag when the blood glucose changes rapidly or during exercise. Glucose sensors are placed subcutaneously by the patient and replaced every 10–14 days; a different device can be placed subcutaneously by a minor surgical procedure and replaced every 6–12 months. Some CGMs require calibration by fingerstick blood glucose measurement. CGM provides glucose data every 5 minutes, and the device's output can be provided in an ambulatory glucose profile (AGP) that is a standardized, single page summary that includes the percentage of time in the desired glycemic range (TIR, or time in range), the percentage of time above the target range, the percentage of time below the target range, the glucose management indicator (GMI), which correlates with HbA$_{1c}$ **(Table 416-3)**, and glucose variability. CGM in real time also allows the patient to monitor the trend of glucose change (upward or downward), with this trend being used to avoid predicted hyper- or hypoglycemia.

CGM device technology is rapidly evolving with the number of available CGM devices increasing and features expanding. This chapter uses the term CGM to encompass both real-time CGM and intermittently scanned CGM, with real-time CGM being more effective. Most CGMs require a provider's prescription, but CGM devices are approved for purchase without a provider's prescription. The selection of CGM type should consider the implications for the individual with diabetes (e.g., cost, convenience, insurance coverage), the provider (e.g., access to patient's CGM data), and the health system (e.g., integration of patient CGM data into the electronic medical record). Selection is best optimized by the involvement of a certified diabetes educator knowledgeable about these technologies. Unfortunately, certified diabetes educators are not available in all care settings, so the individual with diabetes and the provider may need to investigate CGM options using online resources and see suggestions in the reference list. Many are vendor specific; some resources provide information on multiple vendors, technologies, and devices. The selection of a CGM device should be individualized based on patient preference and skill level, capability to collect and use the data, and ability to upload data to adjust therapy. CGM may be used in individuals whose diabetes is partially or completely managed by someone else, such as a caregiver (e.g., in a child or an individual with cognitive impairment). After the selection of a CGM device, the individual with diabetes (and/or the caregiver) should receive education and training on a regular basis.

TABLE 416-3 Relationship of HbA$_{1c}$ and Estimated Average Glucose (eAG)

HEMOGLOBIN A$_{1C}$ (%)	ESTIMATED AVERAGE GLUCOSE (AVERAGE, RANGE)	
	mmol/L	mg/dL
5	5.4 (4.2–6.7)	97 (76–120)
6	7.0 (5.5–8.5)	126 (100–152)
7	8.6 (6.8–10.3)	154 (123–185)
8	10.2 (8.1–12.1)	183 (147–217)
9	11.8 (9.4–13.9)	212 (170–249)
10	13.4 (10.7–15.7)	240 (193–282)
11	14.9 (12.0–17.5)	269 (217–314)
12	16.5 (13.3–19.3)	298 (240–347)

Source: Data adapted from Diabetes Care 31:1473, 2008.

BGM devices use a small drop of blood (<2 μL) and an enzymatic reaction to rapidly measure the capillary blood glucose. It is critical that individuals using CGM also have a capillary (fingerstick) device for times when the CGM technology is malfunctioning, CGM results are questionable, or for CGM calibration (if required or desired). Substances can interfere with the accuracy of measurements by CGM devices (e.g., hydroxyurea, acetaminophen, ascorbic acid, mannitol, and sorbitol) and BGM devices (e.g., uric acid, galactose, xylose, acetaminophen, L-DOPA, or ascorbic acid).

Individuals with type 1 DM or individuals with type 2 DM taking insulin injections each day should monitor their blood glucose by CGM. CGM in type 1 DM, especially in those with hypoglycemia unawareness, can decrease the frequency of serious hypoglycemia (especially nocturnal hypoglycemia). The combination of an insulin-infusion device (discussed below) and a CGM can automate insulin delivery with either predictive suspension of insulin delivery to avoid hypoglycemia or closed-loop control that automatically adjusts insulin delivery by a predictive algorithm **(Fig. 416-1)**. Some CGM/insulin pump manufacturers offer the patient a way to upload glucose data into the manufacturer's server, which can then be securely accessed by the provider's staff. It is critical for the patient's glycemic data to be securely accessible to the provider and uploaded into the electronic health record of the provider and health system, but approaches and systems to accomplish this need further improvement.

Individuals with type 2 DM treated on oral therapy and/or only with diet/lifestyle require less intense glycemic monitoring and can use CGM or BGM to measure the glucose at a lower frequency (e.g., 3–5 times/week). Many individuals with type 1 or type 2 DM report that real-time access to glycemic information via CGM assists in lifestyle choices, diet, and activity in addition to insulin or medication management, thereby improving control.

Assessment of Long-Term Glycemic Control Measurement of glycated hemoglobin (HbA$_{1c}$) is the standard method for assessing long-term glycemic control. When plasma glucose is consistently elevated, there is an increase in nonenzymatic glycation of hemoglobin; this alteration reflects the glycemic history over the previous 2–3 months, because erythrocytes have an average life span of ~120 days (glycemic level in the preceding month contributes about 50% to the HbA$_{1c}$ value). Laboratory standards for the HbA$_{1c}$ test should be correlated to the reference assay of the Diabetes Control and Complications Trial (DCCT). Measurement of HbA$_{1c}$ at the "point of care" allows for more rapid feedback and may therefore assist in adjustment of therapy.

As the primary predictor of long-term complications of DM, the HbA$_{1c}$ should mirror the short-term measurement by CGM or BGM. HbA$_{1c}$ should be measured in all individuals with DM during their initial evaluation and as part of their comprehensive diabetes care. In patients achieving their glycemic goal, the ADA recommends measurement of the HbA$_{1c}$ at least twice per year. More frequent testing (every 3 months) is warranted when glycemic control is inadequate or when therapy has changed. The HbA$_{1c}$ correlates with the average plasma glucose value or estimated average glucose (eAG) (Table 416-3) but does not detect glycemic variability or recent intercurrent illnesses as CGM or BGM can. There is interindividual variability in the HbA$_{1c}$ to mean glucose relationship, likely genetically determined; there is controversy and some uncertainty about the influence of race or ancestry

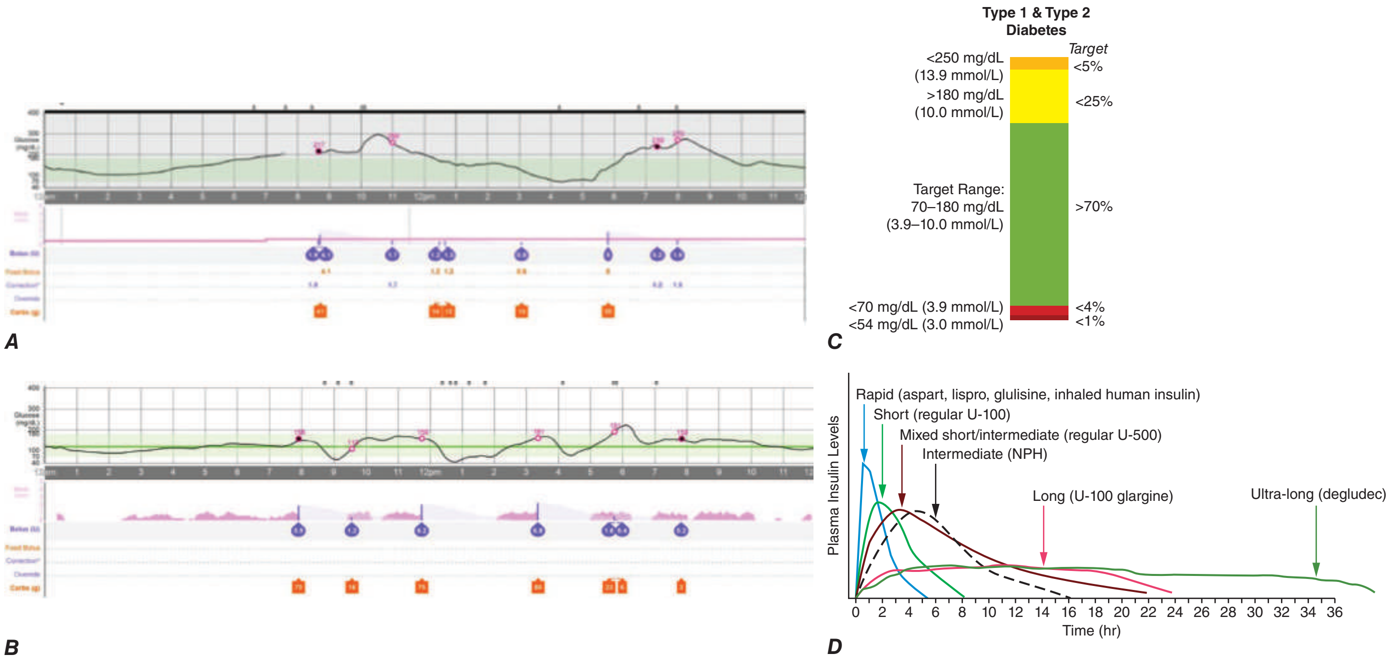

FIGURE 416-1 Glycemic monitoring and insulin administration options for treatment of diabetes. ***A.*** Continuous glucose monitoring (CGM) profile and delivery of rapid-acting insulin analogue by continuous subcutaneous insulin infusion pump involves a basal rate (*light purple line*) and prandial and correction boluses (*purple circles*) based on estimated carbohydrate intake (*orange squares*) and an insulin sensitivity factor. ***B.*** CGM profile with sensor-communicating insulin pump that automates insulin delivery by suspending delivery for predicted hypoglycemia and increasing basal delivery for predicted hyperglycemia (*light purple curves*) while still requiring user input for estimated carbohydrate intake (*orange squares*) to provide prandial insulin boluses (*purple circles*). ***C.*** CGM profile is used to generate an estimate of time-in-range with glycemic goal shown on the left side of the bar and target percent time in that glycemic range shown on the right side of the bar. ***D.*** Pharmacokinetic profile of selected insulin formulations. The duration of action of an insulin may vary among individuals. *(Part C: Reproduced with permission from T Battelino et al: Clinical targets for continuous glucose monitoring data interpretation: Recommendations from the International Consensus on Time in Range. Diabetes Care 42:1593, 2019; Part D: Reproduced with permission from JJ Neumiller: Insulin Update: New and Emerging Insulins, American Diabetes Association, 2018.)*

on HbA_{1c}. For example, the HbA_{1c} in African Americans is slightly higher (~0.3%) than in non-Hispanic white or Hispanic individuals for the same mean glucose. Clinical conditions leading to abnormal red blood cell (RBC) parameters such as hemoglobinopathies, anemias, reticulocytosis, transfusions, uremia, variants in glucose-6-phosphate dehydrogenase, hemodialysis, erythropoietin therapy, and HIV treatment may alter the HbA_{1c} result. Glycemic control can also be assessed by the degree of glycation of other proteins, such as fructosamine or glycated albumin, that reflect glycemia over the prior 2–4 weeks.

PHARMACOLOGIC TREATMENT OF DIABETES

Comprehensive care of type 1 and type 2 DM requires an emphasis on nutrition, exercise, and monitoring of glycemic control in addition to glucose-lowering medication(s). Medications to prevent and manage diabetes-related complications are discussed in **Chap. 417**. This chapter discusses classes of such medications but does not describe all glucose-lowering agents available worldwide. The initial step is to select an individualized glycemic goal for the patient.

■ ESTABLISHMENT OF TARGET LEVEL OF GLYCEMIC CONTROL

Because the complications of DM are related to glycemic control, normoglycemia or near-normoglycemia is the desired, but often elusive, goal for most patients. Normalization or near-normalization of the plasma glucose for long periods of time had been extremely difficult, as demonstrated by the DCCT and United Kingdom Prospective Diabetes Study (UKPDS), but new technologies and medications are making this goal more feasible. Regardless of the level of hyperglycemia, improvement in glycemic control will lower the risk of diabetes-related complications, most notably the microvascular complications **(Chap. 417)**.

The target for glycemic control (as reflected by the HbA_{1c}) should be individualized, and the goals of therapy should be developed in consultation with the individuals with diabetes after considering a number of medical, social, and lifestyle issues (ADA terms this patient-centered care) such as age, ability to understand and implement a treatment regimen, presence and severity of complications of diabetes such as ASCVD, ability to recognize hypoglycemic symptoms, presence of other medical conditions or treatments that might affect survival or the response to therapy, lifestyle and occupation (e.g., possible consequences of experiencing hypoglycemia on the job), and level of support available from family and friends.

In general, the ADA suggests that the goal is to achieve an HbA_{1c} as close to normal as possible without significant hypoglycemia. In most individuals, the target HbA_{1c} should be <7% **(Table 416-4)** with a more stringent (≤6.5%) target for some patients. With current treatment and devices, the level of HbA_{1c} is no longer inversely related to the frequency and severity of hypoglycemia as seen in the DCCT. A higher HbA_{1c} target of <7.5 or 8% is appropriate for individuals with cognitive impairment, those with reduced ability to sense hypoglycemia, or those with limited life span, realizing that these factors represent a spectrum across individuals (Table 416-4). Approximately one in four individuals over the age of 65 years has diabetes. Thus, the glycemic goal in elderly individuals (>65 years) should be individualized and consider the overall clinical state of the individual. For example, in an elderly individual with robust cognition and few major health issues, the glycemic goal may be the same as in younger individuals (HbA_{1c} target <7.0%), while in an individual with impaired cognition or a resident of a long-term facility, the major goal is avoidance of hypoglycemia and severe hyperglycemia (Table 416-4).

Large clinical trials (UKPDS, Action to Control Cardiovascular Risk in Diabetes [ACCORD], Action in Diabetes and Vascular Disease: Preterax and Diamicron MR Controlled Evaluation [ADVANCE], Veterans Affairs Diabetes Trial [VADT]; DCCT/EDIC study in type 1 DM; see **Chap. 417**) examined glycemic control in type 2 DM in individuals with low risk of ASCVD, a high risk of ASCVD, or established ASCVD. Overall, these studies indicate that (1) improved glycemic control reduces microvascular complications of diabetes; (2) improved glycemic control in individuals early in the course of type 1 DM led to reduction in nonfatal myocardial infarction, stroke, and cardiovascular death almost two decades after the period of improved glycemic control had ended; (3) intense glycemic control is beneficial for ASCVD in

TABLE 416-4 Glycemic Goals for Adults with Diabetes[a]

INDEX OF GLYCEMIC CONTROL	ADULTS (NONPREGNANT)	ELDERLY ADULTS WITH INTACT COGNITION AND FUNCTIONAL STATUS	ELDERLY ADULTS WITH OTHER SERIOUS COMORBIDITIES[b]	ELDERLY ADULTS WITH COMPLEX COMORBIDITIES, POOR HEALTH, OR IMPAIRED COGNITION[c]
HbA_{1c}	<7.0% (53 mmol/mol)	<7.0–7.5% (53–57 mmol/mol)	<8.0% (64 mmol/mol)	<8.5% (64 mmol/mol) with avoidance of hypoglycemia
CGM metrics				
% Time within indicated range[c]	>70% within 3.9–10.0 mmol/L (70–180 mg/dL)[c]	>70% within 4.4–10.0 mmol/L (80–180 mg/dL)[c]	>50% within 5.5–10.0 mmol/L (100–200 mg/dL)[c]	>40% within 6.7–12.2 mmol/L (120–220 mg/dL)[c]
Time below 3.9 mmol/L (70 mg/dL) but >54 mg/dL (>3 mmol/L) indicating level 1 or mild hypoglycemia)[c]	<4%	<1%	0%	0%
Time below <54 mg/dL (<3 mmol/L) indicating level 2 or moderate/severe hypoglycemia[c]	<1%	<1%	0%	0%
Glucose variability, % coefficient of variation[d]	≤36%	<33%	N/A	N/A
Preprandial capillary blood glucose	4.4–7.2 mmol/L (80–130 mg/dL)	4.4–7.2 mmol/L (80–130 mg/dL)	5.0–8.3 mmol/L (90–150 mg/dL)	5.6–10.0 mmol/L (100–180 mg/dL)
Postprandial capillary blood glucose[e]	<10.0 mmol/L (<180 mg/dL)	<11.1 mmol/L (200 mg/dL)	<13.9 mmol/L (250 mg/dL)	<13.9 mmol/L (250 mg/dL)

[a] Glycemic goal should be individualized for each patient; elderly >65 years. Some suggest different glycemic targets such as the American Association of Clinical Endocrinology (AACE), which suggests an HbA_{1c} goal <6.5%, and American College of Physicians (ACP), which suggests a goal of 7–8%. [b] Multiple chronic illnesses, impaired activities of daily living, or cognitive impairment. [c] Overall poor health with complex comorbidities, cognitive impairment, or limited life span or resident in skilled nursing facility or a long-term care facility. [d] As determined by CGM. See **Chap. 418** for hypoglycemia definitions. [e] 1–2 h after beginning of a meal.

Abbreviations: CGM, continuous glucose monitoring; HbA_{1c}, hemoglobin A_{1c}; N/A, not applicable to glycemia management.

Source: Adapted from several sources, including Diabetes Care 47:S111, 2024, and Diabetes Care 47:S244, 2024.

some populations with type 2 DM; and (4) hypoglycemia in high-risk populations with ASCVD should be avoided as it is associated with cardiovascular events and mortality. Thus, near-normal glycemia is not the goal in this population (Table 416-4). As will be discussed later, these studies were conducted before the advent of glucagon-like peptide 1 (GLP-1) receptor agonists (GLP-1RAs) and sodium-glucose cotransporter 2 (SGLT-2) inhibitors, which have greater cardiovascular benefit than agents utilized in these earlier clinical trials.

TYPE 1 DIABETES MELLITUS

General Aspects The goal is to design and implement an insulin regimen that mimics physiologic insulin secretion. Because individuals with type 1 DM partially or completely lack endogenous insulin production, administration of basal insulin is essential for regulating glycogen breakdown, gluconeogenesis, lipolysis, and ketogenesis (e.g., fine-tuning hepatic and adipose metabolism). Likewise, insulin replacement for meals should be appropriate for the carbohydrate intake and insulin sensitivity, promoting normal glucose utilization and storage. The continued increase in insulin costs over the past decade has been a major challenge for individuals with diabetes. Recent federal and state legislative action has begun to address this; however, the cost of insulin continues to be a major issue in diabetes care.

Intensive Management of Glycemia Intensive insulin therapy in type 1 DM seeks to achieve normal or near-normal glycemia. This goal requires the integration of multiple resources and efforts, including thorough and continuing patient education, comprehensive recording of glucose measurements and nutrition intake by the patient, and a variable insulin regimen that matches carbohydrate intake and exercise and insulin dose. Insulin is delivered subcutaneously via multiple daily injections (MDIs), continuous subcutaneous insulin infusion (CSII), a sensor-augmented system, or an automated insulin delivery system (AID). Insulin delivery by CSII requires a manual entry into the pump to alter the basal infusion rate or direct an insulin bolus. A sensor-augmented system has a pump and a CGM device, assisted by an algorithm that suspends the insulin infusion when the glucose is low or predicted to be low in 30 min based on the glucose trajectory. An AID system, using a pump, CGM, and algorithm, increases or decreases the basal insulin infusion rate in real time based on CGM data. Some AID systems deliver a correction insulin bolus, but these

TABLE 416-5 Properties of Insulin Preparations[a]

PREPARATION	TIME OF ACTION		
	ONSET, h	PEAK, h	EFFECTIVE DURATION, h
Rapid-acting, injected			
Aspart[b]	<0.25	0.5–1.5	3–5
Glulisine	<0.25	0.5–1.5	3–5
Lispro[c]	<0.25	0.5–1.5	3–5
Short-acting, injected			
Regular[d]	0.5–1.0	2–3	4–8
Rapid-acting, inhaled			
Inhaled human insulin	<0.25	1–2	3
Intermediate-acting, injected			
NPH	2–4	4–10	10–16
Long-acting or Ultralong-acting, injected			
Degludec	1–9	—[e]	42[f]
Glargine[g]	2–4	—[e]	20–24
Examples of insulin combinations[h]			
75/25–75% protamine lispro, 25% lispro	<0.25	Dual[i]	10–16
70/30–70% protamine aspart, 30% aspart	<0.25	Dual[i]	15–18
50/50–50% protamine lispro, 50% lispro	<0.25	Dual[i]	10–16
70/30–70% NPH, 30% regular	0.5–1	Dual[i]	10–16
Combination of long-acting insulin and GLP-1RA	See text		

[a] Injectable insulin preparations (with exception of inhaled formulation) available in the United States; others are available in the United Kingdom and Europe. Standard formulations are U-100 (100 units of insulin per mL solution). Insulin detemir, a long-acting insulin, will soon not be available and is not included in this table. [b] Formulation with niacinamide (vitamin B_3) has a slightly more rapid onset and offset. [c] Lispro-aabc formulation has a slightly more rapid onset and offset. Several forms of insulin (e.g., degludec, insulin lispro, Lispro-aabc) are also available in U-200 concentration. [d] Formulation also available in U-500 concentration with delayed onset and offset. [e] Degludec and glargine have minimal peak activity. [d] Duration is dose-dependent. [g] Formulation also available in U-300 concentration, which has longer duration. [h] Other insulin combinations are available. [i] Dual: two peaks—one at 2–3 h and the second one several hours later.

Abbreviations: GLP-1RA, glucagon-like peptide 1 receptor agonist; NPH, neutral protamine Hagedorn.

are not a completely closed-loop system as the patient must input carbohydrate intake data and projected activity or exercise. This is a rapidly evolving area of type 1 DM–related technology, algorithms, and artificial intelligence, with some individuals having considerable success using do-it-yourself (DIY) approaches that are based in the type 1 DM user community and not U.S. Food and Drug Administration (FDA) approved.

The benefits of intensive insulin therapy and improved glycemic control include a reduction in the acute metabolic and chronic microvascular complications of DM. From a psychological standpoint, the patient experiences greater control over their diabetes and often notes an improved sense of well-being, greater flexibility in the timing and content of meals, and the capability to alter insulin dosing with exercise. Intensive insulin therapy prior to and during pregnancy reduces the risk of fetal malformations and morbidity. Intensive insulin therapy is encouraged in newly diagnosed patients with type 1 DM, including the use of CGM. Although intensive management confers impressive benefits, it may not be appropriate at all times for all individuals with T1D (Table 416-4). Some individuals with diabetes prefer subcutaneous, intermittent insulin injections combined with a CGM to being connected continuously to an insulin infusion device, highlighting the need for individualized diabetes care.

Insulin Preparations Insulin preparations are generated by recombinant DNA technology and consist of the amino acid sequence of human insulin or variations thereof. In the United States, most insulin is formulated as U-100 (100 units/mL); short-acting insulin formulated as U-200 (200 units/mL; lispro) and long-acting as U-300 (300 units/mL; glargine) are available in order to limit injection volumes for patients with high insulin requirements. Regular insulin formulated as U-500 (500 units/mL) is sometimes used in patients with severe insulin resistance. Human insulin has been formulated with distinctive pharmacokinetics (regular and neutral protamine Hagedorn [NPH] insulin have the native insulin amino acid sequence) or genetically modified to alter insulin absorption and hence the onset and duration of insulin action. Insulins can be classified as rapid-acting, short-acting, intermediate-acting, long-acting, or ultralong-acting (**Table 416-5;** Fig. 416-1*D*). For example, one rapid-acting insulin formulation, insulin lispro, is an insulin analogue in which the 28th and 29th amino acids (lysine and proline) on the insulin B chain have been reversed. Insulin aspart and insulin glulisine are modified insulin analogues with properties similar to lispro. A biosimilar version of lispro is available. These insulin analogues have full biologic activity but less tendency for self-aggregation, resulting in more rapid absorption and onset of action and a shorter duration of action. These characteristics are particularly advantageous for allowing entrainment of insulin injection and action to the rising plasma glucose levels following meals. The shorter duration of action also appears to be associated with a decreased number of hypoglycemic episodes, primarily because the decay of insulin action corresponds to the decline in plasma glucose after a meal. Thus, insulin aspart, lispro, or glulisine is preferred over regular insulin for prandial coverage. Insulin glargine is a long-acting biosynthetic human insulin that differs from normal insulin in that asparagine is replaced by glycine at amino acid 21, and two arginine residues are added to the C terminus of the B chain, leading to the formation of microprecipitates at physiologic pH in subcutaneous tissue. Compared to NPH insulin, the onset of insulin glargine action is later, the duration of action is longer (~24 h), and there is a less pronounced peak. A lower incidence of hypoglycemia, especially at night, has been reported with insulin glargine when compared to NPH insulin. A biosimilar version is available. Twice-daily injections of glargine are sometimes required to provide optimal 24-h basal insulin coverage. Because of modification and extension of the carboxy-terminus of the B chain, insulin degludec forms multihexamers in subcutaneous tissue and binds albumin, prolonging its duration of action (>42 h); it provides similar glycemic control as glargine but with less frequent nocturnal and severe hypoglycemia. Other modified insulins, such as one with a duration of action of 1 week, are in clinical trials and will likely soon be available.

Basal insulin requirements, largely fine-tuning hepatic glucose metabolism, are provided by long-acting insulin formulations (NPH insulin, insulin glargine, or insulin degludec) (Fig. 416-1*D*; Table 416-5). These are usually prescribed with rapid-acting insulin in an attempt to mimic physiologic insulin release with meals (prandial insulin requirement). In the past, NPH and short-acting insulin formulations were mixed in the same syringe, but this is not common now. The miscibility of some insulins allows for the production of combination insulins that contain 70% NPH and 30% regular (70/30), or equal mixtures of NPH and regular (50/50). By including the insulin analogue mixed with protamine, several additional combinations have a rapid-acting and long-acting profile (Table 416-5; Fig. 416-1*D*). Although more convenient for the patient (only two injections a day), combination insulin formulations do not allow independent adjustment of short-acting and long-acting activity and are not appropriate in type 1 DM management. Most insulin formulations are available as insulin "pens," which are more convenient and accurate than syringes; "smart pens" can assist with insulin dose tracking. Insulin delivery by inhalation to provide mealtime insulin has a more rapid onset of action than insulin injected subcutaneously. Prior to its use, the forced expiratory volume in 1 s (FEV_1) should be measured, and then monitored periodically during treatment. Inhaled insulin can cause bronchospasm and cough and should not be used by individuals with lung disease or those who smoke. Long-acting insulin/GLP-1RA combinations in fixed doses (degludec plus liraglutide or glargine plus lixisenatide) are effective and are associated with less weight gain.

Insulin Regimens There is considerable patient-to-patient variation in the peak and duration. In all regimens, long-acting insulins (NPH, glargine, or degludec) supply basal insulin, whereas regular, insulin aspart, glulisine, or lispro provide prandial insulin (Fig. 416-1*D*; Table 416-5). Rapid-acting insulin analogues should be injected just before (<10 min) and regular insulin 30–45 min prior to a meal. Sometimes rapid-acting insulin analogues are injected just after a meal (gastroparesis, unpredictable food intake). A consensus statement from ADA and the European Association for the Study of Diabetes provides guidance about different insulin regimens used in type 1 DM.

A shortcoming of current insulin regimens is that injected insulin immediately enters the systemic circulation, whereas endogenous insulin is secreted into the portal venous system. Thus, exogenous insulin administration exposes the liver to subphysiologic insulin levels, and requires achieving higher peripheral levels of insulin to restrain hepatic glucose production. No current insulin regimen reproduces the precise insulin secretory pattern of the pancreatic islet. However, the most physiologic regimens entail more frequent insulin injections, greater reliance on rapid-acting insulin, and CGM or more frequent BGM. In general, individuals with type 1 DM require 0.4–1.0 units/kg per day of insulin divided into multiple doses, with 30–50% of daily insulin given as basal insulin with the remainder as prandial insulin. All individuals with type 1 DM should have a filled glucagon prescription **(Chap. 416)**.

MDI regimens refer to the combination of basal insulin and bolus insulin (preprandial rapid-acting insulin). The timing and dose of rapid-acting, preprandial insulin are altered to accommodate the CGM or BGM results, anticipated food intake, and physical activity. Such regimens offer the patient with type 1 DM more flexibility in terms of lifestyle and the best chance for achieving near normoglycemia. Most often, basal insulin with glargine or degludec is used in conjunction with preprandial lispro, glulisine, or insulin aspart. The dose of long-acting insulin is adjusted based on the fasting glucose. The insulin aspart, glulisine, or lispro dose is based on individualized algorithms that integrate the preprandial glucose and the anticipated carbohydrate intake. To determine the meal component of the preprandial insulin dose, the patient uses an insulin-to-carbohydrate ratio (a common ratio for type 1 DM is 1 unit/10–15 g of carbohydrate, but this must be determined for each individual). To this insulin dose is added the supplemental or correcting insulin based on the preprandial blood glucose (one formula uses 1 unit of insulin for every 1.6–3.3 mmol/L [30–60 mg/dL] over the preprandial glucose target; this correction factor can be estimated from 1500/[total daily insulin dose]). Such

TABLE 416-6 Agents Used for Treatment of Type 1 or Type 2 Diabetes

	MECHANISM OF ACTION	EXAMPLES[a]	HBA$_{1C}$ REDUCTION (%)[b]	AGENT-SPECIFIC ADVANTAGES	AGENT-SPECIFIC DISADVANTAGES	CONSIDERATIONS
Oral						
Biguanides[c*]	↓ Hepatic glucose production, ↑ insulin sensitivity, influence gut function	Metformin	1–2	Weight neutral, do not cause hypoglycemia, inexpensive, extensive experience, modest ↓ CV events	Diarrhea, nausea, lactic acidosis, vitamin B_{12} deficiency	Renal insufficiency (see text for GFR <30 mL/min), CHF, radiographic contrast studies, hospitalized patients, acidosis
Sodium-glucose cotransporter 2 (SGLT-2) inhibitors[c***]	↑ Renal glucose excretion	Canagliflozin, dapagliflozin, empagliflozin, ertugliflozin, bexagliflozin, sotagliflozin (SGLT-1/2 inhibitor)	0.5–1.0	Renal protective, ↓ CV events, ↓ heart failure, do not cause hypoglycemia, modest ↓ weight and blood pressure	Increased risk genital mycotic infections and necrotizing fasciitis of perineum; polyuria, dehydration; increased risk of euglycemic DKA[f] (see text); exacerbate tendency to hyperkalemia	Moderate renal insufficiency; discontinue 3–4 days before surgery, during serious illness
Dipeptidyl peptidase-4 inhibitors[c***]	Prolong endogenous GLP-1 action; ↑ insulin, ↓ glucagon	Alogliptin, linagliptin, saxagliptin, sitagliptin, vildagliptin	0.5–0.8	Well tolerated, do not cause hypoglycemia	Angioedema/urticarial and immune-mediated dermatologic effects; rarely associated with pancreatitis	Reduced dose with renal insufficiency
Insulin secretagogues: Sulfonylureas[c*]	↑ Insulin secretion	Glimepiride, glipizide, gliquidone, glyburide	1–2	Short onset of action, lower postprandial glucose, inexpensive	Hypoglycemia, weight gain	Renal/liver insufficiency
Insulin secretagogues: Nonsulfonylureas[c***]	↑ Insulin secretion	Nateglinide, repaglinide	0.5–1.0	Short onset of action, lower postprandial glucose	Hypoglycemia	Renal/liver insufficiency (except repaglinide)
Thiazolidinediones[c****]	↓ Insulin resistance, ↑ glucose utilization	Pioglitazone, rosiglitazone	0.5–1.4	Lower insulin requirements	Peripheral edema, CHF, weight gain, fractures, macular edema	CHF, renal/liver insufficiency
α-Glucosidase inhibitors[c**]	↓ GI glucose absorption	Acarbose, miglitol	0.5–0.8	Reduce postprandial glycemia	GI flatulence, elevated liver function tests	Renal/liver insufficiency
Parenteral/Oral (GLP-1RA-related agents)						
GLP-1RAs[c***]	↑ Insulin, ↓ glucagon, slow gastric emptying, satiety	Dulaglutide, exenatide, liraglutide, lixisenatide, semaglutide (oral formulation available)	0.5–1.0	Weight loss, do not cause hypoglycemia (unless combined with another insulin secretagogue or insulin); ↓ CV events, modest renoprotection	Nausea, GI intolerance; possibly associated with pancreatitis[e], possibly worsen retinopathy	Renal disease, agents that also slow GI motility; medullary carcinoma of thyroid, previous ileus
GLP-1/GIP receptor agonist[c***]	↑ Insulin, ↓ glucagon, slow gastric emptying, satiety	Tirzepatide	1.8-2.4	Weight loss, do not cause hypoglycemia (unless combined with insulin secretagogue or insulin); ↓ CV events	Nausea, GI intolerance, possibly associated pancreatitis[e], possibly worsen retinopathy	Renal disease, agents that also slow GI motility; medullary carcinoma of thyroid, pancreatic disease, history of gastroparesis
Parenteral						
Amylin agonists[c,d***]	Slow gastric emptying, ↓ glucagon	Pramlintide	0.25–0.5	Reduce postprandial glycemia, weight loss	Injection, nausea, ↑ risk of hypoglycemia with insulin	Agents that also slow GI motility
Insulin[c,d****]	↑ Glucose utilization, ↓ hepatic glucose production, and other anabolic actions	See text and Table 416-4	Not limited	Known safety profile	Injection, weight gain, hypoglycemia	Can be combined with most other agents
Medical nutrition therapy and physical activity[c*]	↓ Insulin resistance, ↑ insulin secretion	Low-calorie, carbohydrate-controlled diet, exercise	1–3	Other health benefits	Compliance difficult, long-term success of sustained weight loss low	Other health benefits

[a] Examples are approved for use in the United States; others are available in other countries. Examples may not include all agents in the class. [b] HbA_{1c} reduction (absolute) depends partly on starting HbA_{1c}. [c] Used for treatment of type 2 diabetes. [d] Used in conjunction with insulin for treatment of type 1 diabetes. Cost of agent in the United States: *low, **moderate, ***high, ****variable. [e] Degree of risk uncertain, avoid in individuals with risk factors for pancreatitis. [f] Risk of euglycemic DKA in patients with insulin deficiency.

Note: Some agents used to treat type 2 diabetes are not included in table (see text).

Abbreviations: CHF, congestive heart failure; CV, cardiovascular; DKA, diabetic ketoacidosis; GFR, glomerular filtration rate; GI, gastrointestinal; GIP, gastric inhibitory polypeptide; GLP-1, glucagon-like peptide 1; GLP-1RA, glucagon-like peptide 1 receptor agonist; HbA_{1c}, glycated hemoglobin A_{1c}.

calculations must be adjusted based on each individual's sensitivity to insulin. CGM or BGM is essential for these types of insulin regimens.

AID is the preferred insulin delivery mechanism for most individuals with type 1 DM (Fig. 416-1), but cost and insurance coverage are critical considerations. To the basal insulin infusion, a preprandial insulin ("bolus") is delivered by the insulin infusion device based on instructions from the patient or an algorithm that incorporates the preprandial plasma glucose and anticipated carbohydrate intake. These sophisticated devices can accurately deliver small doses of insulin (microliters per hour) and have several advantages: (1) multiple basal infusion rates can be programmed to accommodate nocturnal versus daytime basal insulin requirement; (2) basal infusion rates can be altered during periods of exercise; (3) different waveforms of insulin infusion with meal-related bolus allow better matching of insulin depending on meal composition; and (4) programmed algorithms consider ongoing action of prior insulin administration and blood glucose values in calculating the insulin dose. As mentioned, the technology, algorithms, and integration of the different components are changing rapidly, indicating the need to match these with patients' desires, for instruction by a health professional with considerable experience with insulin infusion devices, and for frequent patient interactions with the diabetes management team. Insulin infusion devices may present unique challenges, such as infection at the infusion site, unexplained hyperglycemia because the infusion set becomes obstructed, or diabetic ketoacidosis (DKA) if the insulin infusion device becomes disconnected. Because most physicians use lispro, glulisine, or insulin aspart in CSII or AID, the short half-life of these insulins quickly leads to insulin deficiency if the delivery system is interrupted. Essential to the safe use of infusion devices is thorough patient education, CGM or frequent BGM, and a backup plan for injecting long- and/or rapid-acting insulins in the event of insulin infusion device failure. CGM sensor-augmented insulin infusion devices integrate the information from the CGM to inform insulin delivery (Fig. 416-1). Currently, sensor communicating functions can interrupt basal insulin delivery during hypoglycemia (threshold suspension) or when hypoglycemia is anticipated (predictive suspension), which may be particularly useful for preventing nocturnal hypoglycemia. Hybrid closed-loop systems can combine patient-directed preprandial boluses with automated adjustment of between-meal and basal insulin delivery based on CGM. Clinical experience with closed-loop systems is rapidly increasing and expanding. Bihormonal infusion devices that deliver both insulin and glucagon are being tested.

Other Agents That Improve Glucose Control The role of amylin, a 37-amino-acid peptide co-secreted with insulin from pancreatic beta cells in normal glucose homeostasis is uncertain. However, based on the rationale that patients who are insulin deficient are also amylin deficient, an analogue of amylin (pramlintide) was created and found to reduce postprandial glycemic excursions in individuals with type 1 or type 2 DM taking insulin. Pramlintide injected just before a meal slows gastric emptying and suppresses glucagon but does not alter insulin levels. Pramlintide is approved for insulin-treated patients with type 1 or type 2 DM. The addition of pramlintide produces a modest reduction in the HbA$_{1c}$ and seems to dampen meal-related glucose excursions. In type 1 DM, pramlintide is started as a 15-μg SC injection before each meal and titrated up to a maximum of 30–60 μg as tolerated. In type 2 DM, pramlintide is started as a 60-μg SC injection before each meal and may be titrated up to a maximum of 120 μg. The major side effects are nausea and vomiting, and dose escalations should be slow to limit these side effects. Because pramlintide slows gastric emptying, it may influence absorption of other medications and should not be used in combination with other drugs that slow gastrointestinal (GI) motility. The rapid-acting insulin given before the meal should initially be reduced to avoid hypoglycemia and then titrated as the effects of the pramlintide become evident. Because pramlintide suppresses glucagon, it may worsen hypoglycemia recovery and should not be used in patients with hypoglycemia unawareness. GLP-1RAs and SGLT-2 inhibitors modestly improve the HbA$_{1c}$ in type 1 DM, but the SGLT-2 inhibitors increase the risk of DKA and in general should not be used.

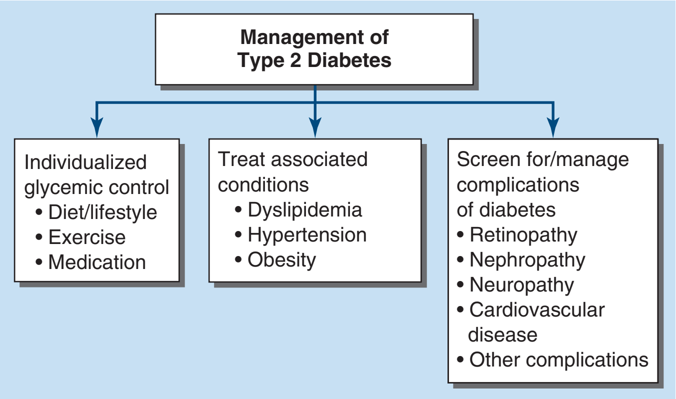

FIGURE 416-2 Essential elements in comprehensive care of type 2 diabetes.

TYPE 2 DIABETES MELLITUS

General Aspects The goals of glucose-directed therapy for type 2 DM are similar to those in type 1 DM and, likewise, should be individualized for each patient. Whereas glycemic control tends to dominate the management of type 1 DM, the care of individuals with type 2 DM must also include even greater attention to the treatment of conditions associated with type 2 DM (e.g., obesity, hypertension, dyslipidemia, ASCVD) and prevention/detection/management of DM-related complications **(Fig. 416-2; Chap. 417)**. Reduction in cardiovascular risk is of paramount importance because this is the leading cause of mortality in these individuals. One approach to pharmacology of glucose-directed therapies in type 2 DM is shown in Fig. 416-3.

Type 2 DM management should begin with MNT (discussed above). An exercise regimen to increase insulin sensitivity and promote weight loss should also be instituted. Pharmacologic approaches to the management of type 2 DM include oral glucose-lowering agents, insulin, and other agents that improve glucose control. Any therapy that improves glycemic control reduces "glucose toxicity" to beta cells and may improve endogenous insulin secretion. However, type 2 DM is a progressive disorder and ultimately requires multiple therapeutic agents and sometimes insulin.

Glucose-Lowering Agents Advances in the therapy of type 2 DM have led to glucose-lowering agents that target different pathophysiologic processes in type 2 DM. Based on their mechanisms of action, glucose-lowering agents are subdivided into agents that increase insulin secretion, reduce glucose production, increase insulin sensitivity, act as a GLP-1 receptor agonist, or promote urinary excretion of glucose (Table 416-6). Insulin is sometimes the initial glucose-lowering agent in type 2 DM if there is severe hyperglycemia or the patient is catabolic.

BIGUANIDES Metformin, representative of this class of agents, reduces hepatic glucose production and improves peripheral glucose utilization slightly (Table 416-6) and is relatively low cost. Metformin acts in multiple tissues, but its mechanism of action remains incompletely defined. Metformin reduces fasting plasma glucose (FPG) and insulin levels, improves the lipid profile, and promotes modest weight loss. An extended-release form is available and may have fewer GI side effects (diarrhea, anorexia, nausea, metallic taste). Because of metformin's relatively slow onset of action and GI symptoms with higher doses, the initial dose should be low and then escalated every 1–2 weeks to a maximally tolerated dose of 2000 mg daily. Metformin is effective as monotherapy and can be used in combination with other glucose lowering agents. The major toxicity of metformin, lactic acidosis, is very rare and can be prevented by careful patient selection. Vitamin B$_{12}$ levels are lower during metformin treatment and should be monitored. Metformin should not be used in patients with moderate renal insufficiency (glomerular filtration rate [GFR] <30 mL/min), any form of acidosis, unstable congestive heart failure (CHF), liver disease, or severe hypoxemia. Metformin should be discontinued in hospitalized patients, in patients who can take nothing orally, and in those receiving radiographic contrast material. Insulin should be used as needed until metformin can be restarted.

INSULIN SECRETAGOGUES—AGENTS THAT AFFECT THE ATP-SENSITIVE K$^+$ CHANNEL Insulin secretagogues stimulate insulin secretion by interacting with the ATP-sensitive potassium channel on the beta cell **(Chap. 415)**. These drugs are most effective in individuals with type 2 DM of relatively recent onset (<5 years) who have residual endogenous insulin production. Sulfonylureas reduce both fasting and postprandial glucose and should be initiated at low doses and increased at 1- to 2-week intervals based on CGM or BGM. Glimepiride and glipizide can be given in a single daily dose and are preferred over glyburide, especially in the elderly. Repaglinide and nateglinide are not sulfonylureas but also interact with the ATP-sensitive potassium channel. Because of their short half-life, these glinide agents are given immediately before each meal to reduce meal-related glucose excursions.

Insulin secretagogues, especially the longer-acting ones, have the potential to cause hypoglycemia, especially in elderly individuals. Hypoglycemia is usually related to delayed meals, increased physical activity, alcohol intake, or renal insufficiency. Individuals who ingest an overdose of some agents develop prolonged and serious hypoglycemia and should be monitored closely in the hospital **(Chap. 418)**. Most sulfonylureas are metabolized in the liver to compounds (some of which are active, such as those of glyburide and the glinide nateglinide) that are cleared by the kidney. Thus, their use in individuals with significant hepatic or renal dysfunction is not advisable. For patients with chronic kidney disease the glinide repaglinide may be used with caution. Weight gain, a common side effect of sulfonylurea therapy, results from the increased insulin levels and improvement in glycemic control. Some sulfonylureas have significant drug interactions with alcohol and some medications including warfarin, aspirin, ketoconazole, α-glucosidase inhibitors, and fluconazole. Sulfonylureas interact with some antibiotics such as fluoroquinolones, clarithromycin, sulfamethoxazole-trimethoprim, metronidazole, and fluconazole, so the sulfonylureas should be discontinued when these antimicrobials are added to the patient's medications.

GLP-1RAS ALONE OR IN COMBINATION WITH GIP RECEPTOR AGONIST "Incretins" amplify glucose-stimulated insulin secretion and suppress inappropriate glucagon secretion **(Chap. 415)**. Agents that either act as a GLP-1RA or enhance endogenous GLP-1 activity are approved for the treatment of type 2 DM and obesity (Table 416-6). Agents in this class do not cause hypoglycemia because of the glucose-dependent nature of incretin-stimulated insulin secretion (unless there is concomitant use of an agent that can lead to hypoglycemia—sulfonylureas, etc.). GLP-1RAs increase glucose-stimulated insulin secretion, suppress glucagon, and slow gastric emptying, but the GLP-1 receptor is expressed in several tissues, including the brain. These agents promote weight loss **(see Chaps. 413 and 414)** and reduce cardiovascular events in those with type 2 DM and ASCVD (see **Chap. 417** for additional discussion about the effect on diabetes-related complications). Thus, these agents are particularly advantageous in type 2 DM. Long-acting GLP-1RAs include sustained-release exenatide, dulaglutide, lixisenatide, and semaglutide, all administered weekly, and are the ones most commonly used. Daily oral semaglutide is available that allows gastric absorption to avoid proteolytic degradation in the small intestine. All are modified to avoid enzymatic inactivation by dipeptidyl peptidase IV (DPP-4) in the circulation. Higher doses of liraglutide and semaglutide than used for glucose-lowering effects are effective for weight-loss therapy for obesity. Liraglutide treatment has also been associated with a decrease in cardiovascular disease (CVD) events in patients with type 2 DM and established CVD and with lower rates of diabetic kidney disease. In similar patient populations, semaglutide treatment has been associated with fewer CVD events and reduced diabetic kidney disease, but with an increased rate of retinopathy-related complications. Dulaglutide treatment has been associated with both a reduction in CVD events and a reduction in composite microvascular retinopathy and nephropathy-related complications primarily driven by prevention of renal events. Treatment with GLP-1RAs should start at a low dose to minimize initial side effects (nausea being the limiting one). GLP-1RAs can be used as combination therapy with metformin, sulfonylureas, and thiazolidinediones. Some patients taking insulin or an insulin secretagogue may require a reduction in those agents to prevent hypoglycemia. The major side effects are nausea and vomiting. Some formulations carry a black box warning from the FDA because of an increased risk of thyroid C-cell tumors in rodents and are contraindicated in individuals with medullary carcinoma of the thyroid or multiple endocrine neoplasia. Because GLP-1RAs slow gastric emptying, they may influence the absorption of other drugs. Whether GLP-1RAs enhance beta cell survival or promote beta cell proliferation is not known. It is not clear if these agents alter the natural history of type 2 DM.

Tirzepatide, a once-weekly subcutaneous injectable peptide engineered to have dual agoniism at both the glucose-dependent insulinotropic polypeptide receptor (GIPR) and the GLP-1R, promotes greater weight loss than a GLP-1RA alone. Additional dual-acting and triple-acting molecules are in development and clinical trials.

DPP-4 inhibitors inhibit degradation of native GLP-1 and GIP and thus enhance the incretin effect. DPP-4, which is widely expressed on the cell surface of endothelial cells and some lymphocytes, degrades a wide range of peptides (not incretin specific). DPP-4 inhibitors promote insulin secretion in the absence of hypoglycemia or weight gain and appear to have a preferential effect on postprandial blood glucose. The levels of GLP-1 action in the patient are greater with the GLP-1RAs than with DPP-4 inhibitors. DPP-4 inhibitors are used either alone or in combination with other oral agents in type 2 DM. Reduced doses should be given to patients with renal insufficiency. Allergy, including rash, hypersensitivity reactions (including anaphylaxis, angioedema, and Stevens-Johnson syndrome), and severe joint pain have been reported in association with DPP-4 inhibitors. There is evidence concerning a potentially increased risk for acute pancreatitis with GLP-1RAs and less so with DPP-4 inhibitors. It is prudent to avoid these agents in patients with pancreatic disease or with other significant risk factors for acute pancreatitis (e.g., heavy alcohol use, severely elevated serum triglycerides, hypercalcemia).

SGLT-2 INHIBITORS These agents (Table 416-6) lower the blood glucose by selectively inhibiting this co-transporter, which is expressed almost exclusively in the proximal convoluted tubule in the kidney. This inhibits glucose reabsorption, lowers the renal threshold for glucose excretion, and leads to increased urinary glucose loss. Thus, the glucose-lowering effect is insulin independent and not related to changes in insulin sensitivity or secretion. The loss of urinary glucose may promote modest weight reduction. Since these agents also impair proximal reabsorption of sodium, their use is associated with a diuretic effect and a 3- to 6-mmHg reduction in systolic blood pressure. Due to the increased urinary glucose, urinary and genital mycotic infections are more common in both men and women, and the diuretic effect can lead to reduced intravascular volume and acutely impaired kidney function. Inhibition of SGLT-2 may lead to increased glucagon and, consequently, liver production of glucose and ketones. Euglycemic DKA may occur during illness or when ongoing glucosuria masks stress-induced requirements for insulin. Patients should be educated about this possibility, and providers should be vigilant about detection. These agents should not be prescribed for patients with type 1 DM or pancreatogenic forms of DM associated with insulin deficiency. Empagliflozin or canagliflozin reduces ASCVD events and all-cause cardiovascular mortality in patients with type 2 DM and established ASCVD. SGLT-2 inhibitors may reduce hospitalization for CHF. Empagliflozin, canagliflozin, and dapagliflozin have all been shown to reduce progression of diabetic kidney disease but should not be initiated in patients with stage 3b chronic kidney disease (CKD; estimated GFR [eGFR] <45 mL/min per 1.73 m^2) and should not be used in stage 4 CKD (eGFR <30 mL/min per 1.73 m^2). A possible increased risk of bladder cancer has been seen with dapagliflozin. **The impact of SGLT-2 inhibitors on diabetes-related complications is discussed in Chap. 417 and below.**

THIAZOLIDINEDIONES Thiazolidinediones (Table 416-6) reduce insulin resistance by binding to the peroxisome proliferator-activated receptor γ (PPAR-γ) nuclear receptor (which forms a heterodimer with the retinoid X receptor). The PPAR-γ receptor is found at highest levels in adipocytes but is expressed at lower levels in many other tissues. Agonists of this receptor regulate a large number of genes, promote adipocyte differentiation, reduce hepatic fat accumulation, and promote fatty acid storage. Thiazolidinediones promote a redistribution of fat from central to peripheral locations. Circulating insulin levels decrease with use of the thiazolidinediones, indicating a reduction in insulin resistance.

Rosiglitazone raises low-density lipoprotein (LDL), high-density lipoprotein (HDL), and triglycerides slightly. Pioglitazone raises HDL to a greater degree and LDL to a lesser degree but lowers triglycerides. The clinical significance of the lipid changes with these agents is not known.

Thiazolidinediones are associated with weight gain (2–3 kg), a small reduction in the hematocrit, and a mild increase in plasma volume. Peripheral edema and CHF are more common in individuals treated with these agents. These agents are contraindicated in patients with hepatic insufficiency or CHF (class III or IV). The FDA has issued an alert that rare patients taking these agents may experience a worsening of diabetic macular edema. An increased risk of fractures has been noted in postmenopausal women taking these agents. Thiazolidinediones have been shown to induce ovulation in premenopausal women with polycystic ovary syndrome. Women should be warned about the risk of pregnancy because the safety of thiazolidinediones in pregnancy is not established. According to an FDA review, pioglitazone may be associated with an increased risk of bladder cancer. In one study, pioglitazone lowered the risk for recurrent stroke or myocardial infarction in insulin-resistant individuals without diabetes who had a prior stroke or transient ischemic attack.

α-GLUCOSIDASE INHIBITORS α-Glucosidase inhibitors reduce postprandial hyperglycemia by delaying glucose absorption (Table 416-6). Postprandial hyperglycemia, secondary to impaired hepatic and peripheral glucose disposal, contributes significantly to the hyperglycemic state in type 2 DM. These drugs, taken just before each meal, reduce glucose absorption by inhibiting the enzyme that cleaves oligosaccharides into simple sugars in the intestinal lumen. Therapy should be initiated at a low dose with the evening meal and increased to a maximal dose over weeks to months. The major side effects (diarrhea, flatulence, abdominal distention) are related to increased delivery of oligosaccharides to the large bowel and can be reduced somewhat by gradual upward dose titration. α-Glucosidase inhibitors may increase levels of sulfonylureas and increase the incidence of hypoglycemia. Simultaneous treatment with bile acid resins and antacids should be avoided. These agents should not be used in individuals with inflammatory bowel disease, gastroparesis, or a serum creatinine >177 μmol/L (2 mg/dL). This class of agents is not as potent as other oral agents in lowering the HbA_{1c} but is unique because it reduces the postprandial glucose rise. If hypoglycemia from other diabetes treatments occurs while taking these agents, the patient should consume glucose because the degradation and absorption of complex carbohydrates will be slowed.

OTHER THERAPIES FOR TYPE 2 DM • Bile Acid–Binding Resins Evidence indicates that bile acids, by signaling through nuclear receptors, may have a role in metabolism. Bile acid metabolism is abnormal in type 2 DM. The bile acid–binding resin colesevelam has been approved for the treatment of type 2 DM (already approved for treatment of hypercholesterolemia). The role of this class of drugs in the treatment of type 2 DM is not yet defined.

Bromocriptine A formulation of the dopamine receptor agonist bromocriptine (Cycloset) has been approved by the FDA for the treatment of type 2 DM. However, its role in the treatment of type 2 DM is uncertain.

INSULIN THERAPY IN TYPE 2 DM Insulin should be considered for initial therapy in type 2 DM, particularly in lean individuals or those with severe weight loss, in individuals with underlying renal or hepatic disease that precludes oral glucose-lowering agents, or in individuals who are hospitalized or acutely ill. Insulin therapy is ultimately required by a substantial number of individuals with type 2 DM because of the progressive nature of the disorder and the relative insulin deficiency that develops with long-standing diabetes. Both physician and patient reluctance often delay the initiation of insulin therapy, but glucose control and individual well-being are improved by insulin therapy in patients who have not reached their glycemic target.

Because endogenous insulin secretion is capable of providing some coverage of mealtime caloric intake, insulin is usually initiated in a single dose of long-acting insulin (0.1–0.4 U/kg per day), given in the evening or just before bedtime (NPH, glargine, or degludec). Because fasting hyperglycemia and increased hepatic glucose production are prominent features of type 2 DM, bedtime insulin is more effective in clinical trials than a single dose of morning insulin. Glargine given at bedtime has less nocturnal hypoglycemia than NPH insulin. Some physicians prefer a relatively low, fixed starting dose of long-acting insulin (10–15 units) or a weight-based dose (0.1 units/kg). The insulin dose may then be adjusted in 10–20% increments as dictated by CGM or BGM results. Both morning and bedtime long-acting insulin may be used in combination with oral glucose-lowering agents. Initially, basal insulin may be sufficient, but often prandial insulin coverage with multiple insulin injections is needed as diabetes progresses (see insulin regimens used for type 1 DM). Other insulin formulations that have a combination of rapid-acting and long-acting insulin (Table 416-5) are sometimes used in patients with type 2 DM because of convenience but do not allow independent adjustment of rapid-acting and long-acting insulin dose and often do not achieve the same degree of glycemic control as basal/bolus regimens. AID in selected individuals with type 2 DM should be considered, especially in those who are insulin-deficient. CGM should be used in all individuals taking insulin.

CHOICE OF INITIAL GLUCOSE-LOWERING AGENT The level of hyperglycemia and the patient's individualized goal (see "Establishment of Target Level of Glycemic Control") should influence the initial choice of therapy. Patients with mild hyperglycemia (FPG <7.0–11.0 mmol/L [126–199 mg/dL]) often respond well to a single, oral glucose-lowering agent, while those with moderate hyperglycemia (FPG 11.1–13.9 mmol/L [200–250 mg/dL]) will usually require more than one oral agent or insulin. Patients with more severe hyperglycemia (FPG >13.9 mmol/L [250 mg/dL]) may respond partially but are unlikely to achieve normoglycemia with oral therapy. Insulin can be used as initial therapy in individuals with severe hyperglycemia (FPG <13.9–16.7 mmol/L [250–300 mg/dL]) or in those who are symptomatic from the hyperglycemia. This approach is based on the rationale that more rapid glycemic control will reduce "glucose toxicity" to the islet cells, improve endogenous insulin secretion, and possibly allow oral glucose-lowering agents to be more effective. If this occurs, the insulin may be discontinued. Treatment algorithms by several professional societies (ADA/ European Association for the Study of Diabetes [EASD], International Diabetes Federation, American Association of Clinical Endocrinology) suggest metformin as initial therapy because of its efficacy, known side effect profile, and low cost **(Fig. 416-3)**. Initiation of pharmacologic therapy should be accompanied by an emphasis on lifestyle modification (e.g., MNT, increased physical activity, and weight loss). Metformin's advantages are that it promotes mild weight loss, lowers insulin levels, and improves the lipid profile slightly. Based on CGM or BGM results and the HbA1c, the dose of metformin should be increased until the glycemic target is achieved or the maximum dose is reached. GLP-1RAs and SGLT-2 inhibitors are increasing in use as evidence accumulates for CVD and CKD benefits, in addition to weight loss and glucose-lowering effects.

Insulin secretagogues, biguanides, α-glucosidase inhibitors, thiazolidinediones, GLP-1RAs, DPP-4 inhibitors, SGLT-2 inhibitors, and insulin are approved for monotherapy of type 2 DM. Although each class of oral glucose-lowering agents has advantages and disadvantages (Table 416-6), certain generalizations apply: (1) insulin secretagogues, biguanides, GLP-1RAs, and thiazolidinediones improve glycemic control to a similar degree (1–2% reduction in HbA_{1c}) and are more effective than α-glucosidase inhibitors, DPP-4 inhibitors, and SGLT-2

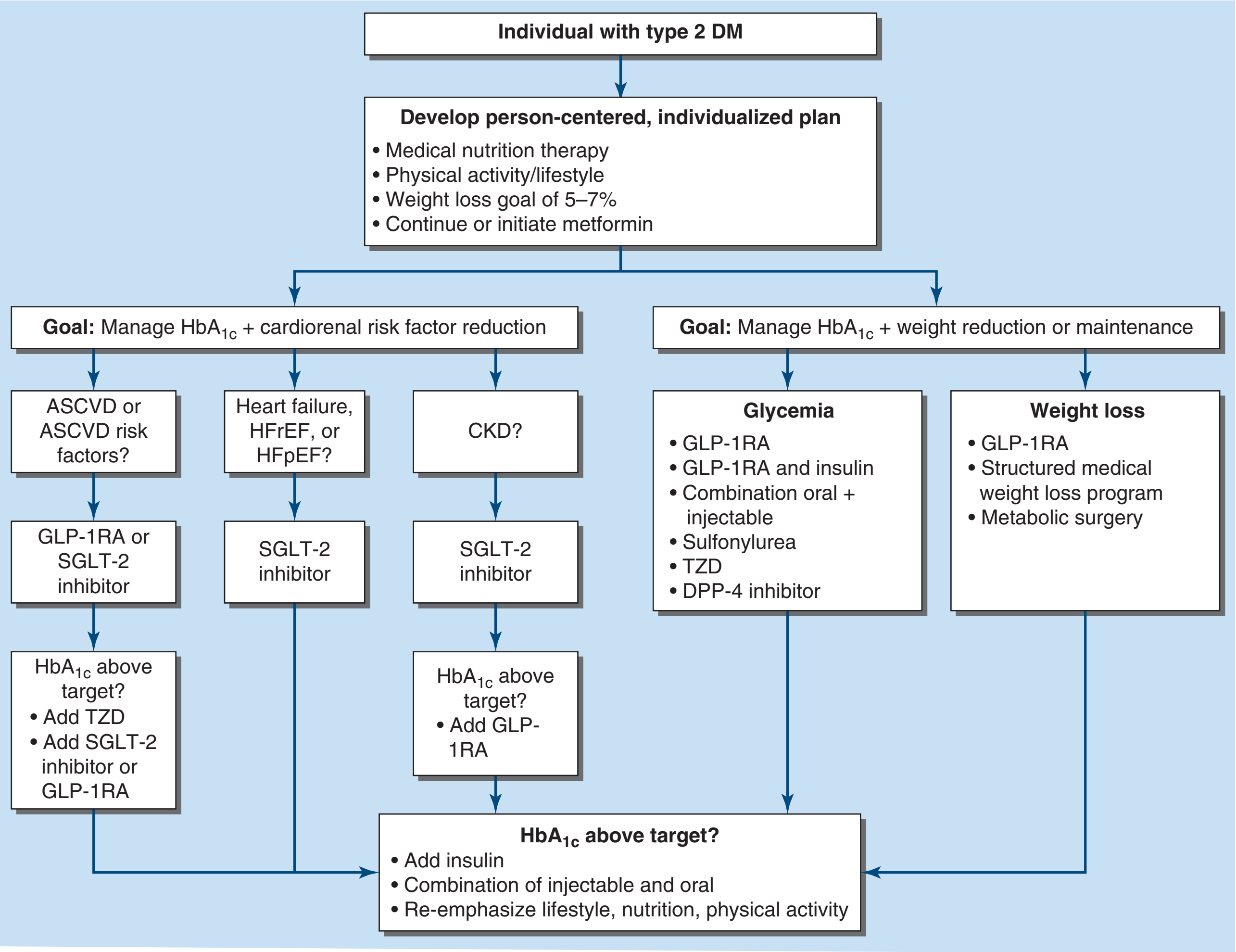
Individual with type 2 DM
Develop person-centered, individualized plan
• Medical nutrition therapy
• Physical activity/lifestyle
• Weight loss goal of 5–7%
• Continue or initiate metformin
Goal: Manage HbA1c + cardiorenal risk factor reduction
Goal: Manage HbA1c + weight reduction or maintenance
ASCVD or ASCVD risk factors?
Heart failure, HFrEF, or HFpEF?
CKD?
Glycemia
• GLP-1RA
• GLP-1RA and insulin
• Combination oral + injectable
• Sulfonylurea
• TZD
• DPP-4 inhibitor
Weight loss
• GLP-1RA
• Structured medical weight loss program
• Metabolic surgery
GLP-1RA or SGLT-2 inhibitor
SGLT-2 inhibitor
SGLT-2 inhibitor
HbA1c above target?
• Add TZD
• Add SGLT-2 inhibitor or GLP-1RA
HbA1c above target?
• Add GLP-1RA
HbA1c above target?
• Add insulin
• Combination of injectable and oral
• Re-emphasize lifestyle, nutrition, physical activity

FIGURE 416-3 Glycemic management of type 2 diabetes. See text for discussion of treatment of severe hyperglycemia or symptomatic hyperglycemia. In this Figure, the term glucagon-like peptide-1 receptor agonists (GLP-1RAs) refers to GLP-1 RAs and dual glucose-dependent insulinotropic polypeptide (GIP) and GLP-1 RAs. Agents that can be combined with metformin include insulin, GLP-1RAs, sodium-glucose cotransporter 2 (SGLT-2) inhibitors, insulin secretagogues, thiazolidinediones (TZD), α-glucosidase inhibitors, and dipeptidyl peptidase-4 (DPP-4) inhibitors. Injectable refers to insulin or GLP-RA. In individuals with type 2 DM and metabolic dysfunction–associated steatotic liver disease (see Chap. 354) or metabolic dysfunction–associated steatohepatitis (see Chap. 354), a GLP-1 RA or a dual glucose-dependent insulinotropic polypeptide (GIP) and GLP-1 RA and/or TZD (pioglitazone) should be considered. ASCVD, atherosclerotic cardiovascular disease; CKD, chronic kidney disease; DM, diabetes mellitus; HbA$_{1c}$, hemoglobin A$_{1c}$; HFrEF, heart failure with reduced ejection fraction; HFpEF, heart failure with preserved ejection fraction.

inhibitors; (2) insulin secretagogues, GLP-1RAs, DPP-4 inhibitors, α-glucosidase inhibitors, and SGLT-2 inhibitors begin to lower the plasma glucose immediately, whereas the glucose-lowering effects of the biguanides and thiazolidinediones takes several days; (3) not all agents are effective in all individuals with type 2 DM; (4) biguanides, α-glucosidase inhibitors, GLP-1RAs, DPP-4 inhibitors, thiazolidinediones, and SGLT-2 inhibitors do not directly cause hypoglycemia; (5) most individuals will eventually require treatment with more than one class of oral glucose-lowering agents or insulin, reflecting the progressive nature of type 2 DM; and (6) durability of glycemic control is slightly less for sulfonylureas compared to metformin or thiazolidinediones.

COMBINATION THERAPY WITH GLUCOSE-LOWERING AGENTS The approach to type 2 DM has changed dramatically with the demonstration that GLP-1RAs and SGLT-2 inhibitors reduce cardiovascular events and slow the progression of renal disease, indicating that a GLP-1RA or an SGLT-2 inhibitor should be used in most individuals with type 2 DM. Therapy should be dictated by whether ASCVD, heart failure, or CKD is present or whether weight loss is a major goal (Fig. 416-3). A number of combinations of therapeutic agents are useful in type 2 DM: metformin plus SGLT-2 inhibitor, metformin plus GLP-1RA, metformin plus insulin, or combinations of a long-acting insulin and a GLP-1RA. Because the mechanism of action of the first and second agents differs, the effect on glycemic control is usually additive. Recent results from the National Institutes of Health–funded Glycemia Reduction Approaches in Diabetes: A Comparative Effectiveness Study (GRADE) indicated that addition of liraglutide or basal insulin to metformin leads to better glycemic control than glimepiride or sitagliptin (SGLT-2 inhibitors were not studied). Medication costs vary considerably (Table 416-6), and this often factors into medication choice as drugs in some categories are very expensive (e.g., SGLT-2 inhibitors, GLP-1RAs). Several fixed-dose combinations of oral agents are available, but evidence that they are superior to titration of a single agent to a maximum dose and then addition of a second agent is lacking. If adequate control is not achieved with the combination of two agents (based on reassessment of the HbA$_{1c}$ every 3 months), a third oral agent, including basal insulin, should be considered (Fig. 416-3). Treatment approaches vary considerably from country to country. For example, α-glucosidase inhibitors are used commonly in South Asian patients (Indian) but infrequently in the United States or Europe. Whether this reflects an underlying difference in the disease or physician preference is not clear.

Treatment with insulin often becomes necessary as type 2 DM enters the phase of relative insulin deficiency and is signaled by inadequate glycemic control with one or two oral glucose-lowering agents. Insulin alone or in combination should be used in patients who fail to reach glycemic targets. For example, a single dose of long-acting insulin at bedtime is often effective in combination with metformin. As endogenous insulin production falls further, multiple injections of long-acting insulin together with rapid-acting insulin are necessary to control postprandial glucose excursions. These insulin regimens are identical to the long-acting and rapid-acting combination regimens discussed above for type 1 DM, although usually at higher doses given insulin resistance. Weight gain and hypoglycemia are the major adverse effects of insulin therapy. The addition of a GLP-1RA can limit this and reduce the dose of insulin needed. The daily insulin dose required can become quite large (1–2 units/kg per day) as endogenous insulin production falls and insulin resistance persists, especially in the setting of weight gain. Insulin plus a thiazolidinedione promotes weight gain and may be associated with peripheral edema. Addition of a GLP-1RA or a thiazolidinedione may necessitate a reduction in the insulin dose to avoid hypoglycemia. Patients requiring large doses of insulin (>200 units/day) can be treated with more concentrated forms of insulin to reduce the volume of injectate and improve absorption.

■ OTHER THERAPIES FOR DIABETES

Metabolic (also referred to as bariatric) surgery for obese individuals with type 2 DM has shown considerable effectiveness, sometimes with dramatic resolution of diabetes or major reductions in the needed dose of glucose-lowering therapies **(Chaps. 413 and 414)**. Several large, nonrandomized clinical trials have demonstrated a much greater efficacy of metabolic surgery compared to medical management in the treatment of type 2 DM, but these trials were conducted before the advent of recently available GLP-1RAs. The ADA clinical guidelines state that metabolic surgery should be considered in individuals with type 2 DM and a body mass index >30 kg/m^2 if hyperglycemia is inadequately controlled despite optimal medical therapy. Metabolic surgery is ideally performed in certified centers with experience with the procedures and associated nutritional support.

Short-term intense caloric restriction (very-low-calorie diet, typically 800–1000 calories/d) can dramatically improve type 2 DM, sometimes leading to resolution of the diabetes. Such an approach is more effective in recent-onset type 2 DM and should be supervised by a provider with expertise and accompanied by a long-term, weight-maintenance program.

Whole-pancreas transplantation can normalize glucose control in type 1 DM and when performed simultaneously with or after kidney transplantation can prolong the life of the kidney transplant by offering protection against recurrent diabetic nephropathy. However, the number of whole-pancreas transplants is declining, likely reflecting the success with CGM and AID. Pancreatic islet transplantation is a less invasive form of beta-cell replacement therapy for type 1 DM; an islet product has received FDA approval. Despite the risks associated with chronic immunosuppression, whole-pancreas and pancreatic islet transplantation may be considered for patients with severe metabolic instability or already requiring immunosuppression in support of a kidney or other organ transplant. Patients with chronic pancreatitis and preserved islet function who require pancreatectomy for pain relief may benefit from autologous islet transplantation as this may prevent or ameliorate postsurgical DM.

■ EMERGING THERAPIES

Recent clinical trials using transplantation of insulin-producing cells derived from human pluripotent stem cells have shown promise. Cost, durability, long-term safety, and patient selection remain to be determined. Many individuals with long-standing type 1 DM still produce very small amounts of insulin or have insulin-positive cells within the pancreas. This suggests that beta cells may slowly regenerate but are quickly destroyed by the autoimmune process. Efforts to suppress the autoimmune process, for example, with a monoclonal antibody that targets T lymphocytes, may preserve beta cells when given at the time of new-onset hyperglycemia in type 1 DM. This agent, teplizumab, a humanized monoclonal antibody to CD3 on T cells, has been approved by the FDA to delay the onset of clinical type 1 DM (stage 3) in patients 8 years of age or older with preclinical (stage 2) disease. Agents that target thioredoxin-interacting protein (TXNIP), especially Ca^{2+} channel blockers, have shown promise in recent-onset type 1 DM and in rodent models of diabetes.

ADVERSE EFFECTS OF THERAPY FOR DM

The benefits of efforts directed toward glycemic control must be balanced against the risks of treatment (Table 416-6). Side effects of intensive treatment include an increased frequency of serious hypoglycemia, weight gain, and greater demands on the individual with diabetes. The most serious complication of therapy for DM is hypoglycemia, and its prevention and treatment with oral glucose or glucagon administered intra-nasally or by injection are discussed in **Chap. 418**. Severe, recurrent, or unexplained hypoglycemia warrants reassessment of the treatment regimen and glycemic goal for the individual patient and possibly deintensification of insulin therapy (see categories of individuals for this would be appropriate in Table 416-4). Weight gain occurs with most (insulin, insulin secretagogues, thiazolidinediones) but not all (metformin, α-glucosidase inhibitors, GLP-1RAs, SGLT-1 inhibitors, DPP-4 inhibitors) therapies. The weight gain is partially due to the anabolic effects of insulin and the reduction in glucosuria.

ACUTE DISORDERS RELATED TO SEVERE HYPERGLYCEMIA

Individuals with type 1 or type 2 DM and severe hyperglycemia should be assessed for clinical stability, including mentation and hydration. The physician should determine if the individual with diabetes is stable or if DKA or a hyperglycemic hyperosmolar state (HHS) is present. In DKA, the hyperglycemia is accompanied by increased ketone concentration in blood and metabolic acidosis. In HHS, the hyperglycemia is usually greater, leading to hyperosmolality and marked dehydration but without ketosis or acidosis. Ketones bodies, an indicator of DKA, should be measured in individuals with type 1 DM when the plasma glucose is persistently >13.9 mmol/L (250 mg/dL). The possibility of DKA should always be considered in patients with type 1 DM during a concurrent illness or with symptoms such as nausea, vomiting, or abdominal pain. Measurement of β-hydroxybutyrate in the blood is preferred over urine testing with nitroprusside-based assays that measure only acetoacetate and acetone. Worldwide, the number of reported cases of DKA and HHS is increasing for unclear reasons. Both DKA and HHS are associated with potentially serious complications, including greater mortality.

Most cases of DKA are in individuals with type 1 DM, while HHS occurs mostly in individuals with type 2 DM. DKA, formerly considered a hallmark of type 1 DM, can also occur at diabetes diagnosis in obese young adults, often of Hispanic or African descent, whose laboratory values are similar to those seen in DKA associated with type 1 DM. However, after treatment of the DKA, these individuals recover their insulin secretory capacity, can gradually discontinue insulin treatment after a few weeks or months, and remain normoglycemic with only diet or oral medication. The cause of this atypical form of diabetes is unknown; it is often termed ketosis-prone diabetes. DKA can also occur in the setting of treatment of type 2 DM with an SGLT-2 inhibitor. Often the blood glucose is normal or just mildly elevated because of the glucosuria. **Table 416-7** compares the features of DKA, HHS, and euglycemic DKA associated with SGLT-2 inhibitors. DKA and HHS exist along a continuum of hyperglycemia, with up to one-third of patients having features of both.

■ DIABETIC KETOACIDOSIS

Clinical Features The symptoms and physical signs of DKA are listed in **Table 416-8** and usually develop over 24 h. DKA may be the initial symptom complex that leads to a diagnosis of type 1 DM, but more frequently, it occurs in individuals with established diabetes. Nausea and vomiting are often prominent, and their presence in an individual with diabetes warrants laboratory evaluation for DKA. Abdominal pain may be severe and can resemble acute pancreatitis or ruptured viscus. Hyperglycemia leads to glucosuria, volume depletion, and tachycardia. Hypotension can occur because of volume depletion in combination with peripheral vasodilatation. Kussmaul respirations and a fruity odor on the patient's breath (secondary to metabolic acidosis and increased acetone) are classic signs of the disorder. Lethargy and central nervous system depression may evolve into coma with severe DKA but should also prompt evaluation for other reasons for altered mental status (e.g., infection, hypoxemia). Cerebral edema, an extremely serious complication of DKA, is seen most frequently in children. Signs of infection, which may precipitate DKA, should be sought on physical examination, even in the absence of fever. Failure to augment insulin therapy during physiologic stress often compounds

TABLE 416-7 Laboratory Values in Diabetic Ketoacidosis (DKA), Hyperglycemic Hyperosmolar State (HHS), and Euglycemic DKA [Representative Ranges at Presentation; mmol/L (mg/dL)]

	DKA	HHS	EUGLYCEMIC DKA[c]
Glucose,[a] mmol/L (mg/dL)	11.1–33.3 (250–600)	33.3–66.6 (600–1200)	5.5-13.9 (100–250)[c]
Sodium, meq/L	125–135	135–145	Normal
Potassium[a,b]	Normal to ↑	Normal	Normal to ↑
Magnesium[a]	Normal	Normal	Normal
Chloride[a]	Normal	Normal	Normal
Phosphate[a,b]	Normal	Normal	Normal
Creatinine	Slightly to moderately ↑	Moderately ↑	Slightly ↑
Osmolality (mOsm/mL)	>300	>300	Normal
Serum/urine ketones[a]	>++	+/–	>++
Serum β-hydroxybutyrate, mmol/L	>3.0	<1.0	>3.0
Serum bicarbonate,[a] meq/L	<18	>18	<18
Arterial pH	6.8–7.3	>7.3	<7.3
Arterial P_{CO_2},[a] mmHg	20–30	Normal	20–30
Anion gap[a] (Na – [Cl + HCO_3])	↑	Normal to slightly ↑	↑

[a]Large changes occur during treatment of DKA; serum level may be normal initially but then require replacement. [b]Although plasma levels may be normal or high at presentation, total-body stores are usually depleted. [c]Sometimes occurs with sodium-glucose cotransporter 2 (SGLT-2) inhibitor treatment; disproportionate glucosuria is consistent with SGLT-2 inhibitor effect.

the problem. Tissue ischemia (heart, brain) can also be a precipitating factor. Omission of insulin because of an infusion pump delivery site occlusion or device malfunction, eating disorder, mental health disorders, or an unstable psychosocial environment may each be a factor precipitating DKA. Complete omission or inadequate administration of insulin by the patient or health care team (in a hospitalized patient with type 1 DM) may precipitate DKA. The rising cost of insulin has been a major challenge and has contributed to individuals with diabetes omitting or rationing their insulin, making them more vulnerable to DKA.

Pathophysiology DKA results from relative or absolute insulin deficiency combined with counterregulatory hormone excess (glucagon, catecholamines, cortisol, and growth hormone). The decreased ratio of insulin to glucagon promotes gluconeogenesis, glycogenolysis, and ketone body formation in the liver, as well as increases in substrate delivery from fat and muscle (free fatty acids, amino acids) to the liver. Ketosis results from a marked increase in free fatty acid release from adipocytes, with a resulting shift toward ketone body synthesis in the liver. Reduced insulin levels, in combination with elevations in catecholamines and growth hormone, also increase lipolysis and the release of free fatty acids. Markers of inflammation (cytokines, C-reactive protein) are elevated in both DKA and HHS.

Laboratory Abnormalities and Diagnosis The timely diagnosis of DKA is crucial and allows for prompt initiation of therapy. DKA is characterized by hyperglycemia (serum glucose >13.9 mmol/L [250 mg/dL], ketosis, and metabolic acidosis [serum bicarbonate <15–18 mmol/L with increased anion gap]) along with a number of secondary metabolic derangements (Table 416-7). Occasionally, the serum glucose is only minimally elevated and may even be normal (euglycemic DKA). This has been noted especially in individuals treated with SGLT-2 inhibitors. Arterial pH usually ranges between 6.8 and 7.3, depending on the severity of the acidosis. Despite a total-body potassium deficit, the serum potassium at presentation may be mildly elevated, secondary to the acidosis and volume depletion. Total-body stores of sodium, chloride, phosphorus, and magnesium are also reduced in DKA but are not accurately reflected by their levels in the serum because of hypovolemia and hyperglycemia. Elevated blood urea nitrogen (BUN) and serum creatinine levels reflect intravascular volume depletion. Leukocytosis, hypertriglyceridemia, and hyperlipoproteinemia are commonly found as well. Hyperamylasemia may suggest a diagnosis of pancreatitis, especially when accompanied by abdominal pain. However, in DKA the amylase is usually of salivary origin and thus is not diagnostic of pancreatitis. Serum lipase should be obtained if pancreatitis is suspected.

The measured serum sodium is reduced as a consequence of the hyperglycemia. An estimated correction is provided by the equation: (1.6-mmol/L [1.6-meq] reduction in serum sodium for each 5.6-mmol/L [100-mg/dL] rise in the serum glucose). A normal serum sodium in the setting of DKA indicates a more profound water deficit.

In DKA, the ketone body, β-hydroxybutyrate, is synthesized at a threefold greater rate than acetoacetate; however, acetoacetate is preferentially detected by a commonly used ketosis detection reagent (nitroprusside). The nitroprusside tablet, or stick, is often used to detect urine ketones; certain medications such as captopril, penicillamine, or valproic acid may cause false-positive reactions. Serum or plasma assays for β-hydroxybutyrate are preferred because they more accurately reflect the true ketone body level.

The degree of acidosis and hyperglycemia do not necessarily correlate closely because a variety of factors determine the level of hyperglycemia (oral intake, urinary glucose loss). Ketonemia is a consistent finding in DKA and distinguishes it from simple hyperglycemia. The differential diagnosis of DKA includes starvation ketosis, alcoholic ketoacidosis (bicarbonate usually >15 meq/L), and other forms of increased anion-gap acidosis (Chap. 55).

TABLE 416-8 Manifestations of Diabetic Ketoacidosis

Symptoms
- Nausea/vomiting
- Thirst/polyuria
- Abdominal pain
- Shortness of breath

Precipitating events
- Inadequate insulin administration
- Infection (pneumonia/UTI/gastroenteritis/sepsis)
- Infarction (cerebral, coronary, mesenteric, peripheral)
- Pancreatitis
- Drugs (cocaine)
- Pregnancy

Physical Findings
- Tachycardia
- Dehydration/hypotension
- Tachypnea/Kussmaul respirations/respiratory distress
- Abdominal tenderness (may resemble acute pancreatitis or surgical abdomen)
- Lethargy/obtundation/cerebral edema/possibly coma

Abbreviation: UTI, urinary tract infection.

TREATMENT

Diabetic Ketoacidosis

Based on laboratory values and clinical exam, DKA can be classified as mild (pH 7.25–7.3, serum bicarbonate 15–18 meq/L, mental status normal), moderate (pH 7.0–7.25, serum bicarbonate 10–15 meq/L, mildly reduced mental status), or severe (pH <7.0, serum bicarbonate <10–15 meq/L, reduced mental status, coma). The management of DKA is outlined in **Table 416-9**. After initiating IV fluid replacement and insulin therapy, the event that precipitated the episode of DKA should be sought and aggressively treated. If the patient is vomiting or has altered mental status, a nasogastric tube should be inserted to prevent aspiration of gastric contents. Central to successful treatment of DKA is careful monitoring and frequent reassessment to ensure that the patient and the metabolic derangements are improving. A comprehensive flow sheet should record chronologic changes in vital signs, fluid intake and output, and laboratory values as a function of insulin administered.

After the initial bolus of normal saline or lactated Ringer's, replacement of the sodium and free water deficit is carried out over the next 24 h (fluid deficit is often 3–5 L). When hemodynamic stability and adequate urine output are achieved, IV fluids should be switched to 0.45% saline or lactated Ringer's, depending on the calculated volume deficit. Ringer's lactate is associated with more rapid DKA resolution and a reduced trend toward hyperchloremia later in the course of DKA resolution.

A bolus of IV (0.1 units/kg) short-acting regular insulin is usually administered immediately (Table 416-9), and subsequent treatment should provide continuous and adequate levels of circulating insulin. IV administration is usually preferred (0.1 units/kg of regular insulin per h) but uncomplicated DKA can also be treated with SC short-acting insulin analogues. As the acidosis and insulin resistance associated with DKA resolve, the insulin infusion rate can be decreased (to 0.02–0.1 units/kg per h). Long-acting insulin, in combination with SC short-acting insulin, should be administered as soon as the patient resumes eating, because this facilitates transition to an outpatient insulin regimen and reduces length of hospital stay. Hyperglycemia usually improves at a rate of 4.2–5.6 mmol/L (50–100 mg/dL) per h as a result of insulin-mediated glucose disposal, reduced hepatic glucose release, and rehydration. Rehydration reduces catecholamines, increases urinary glucose loss, and expands the intravascular volume. The decline in the plasma glucose within the first 1–2 h may be more rapid and is mostly related to volume expansion. Ketoacidosis begins to resolve as insulin reduces lipolysis, increases peripheral ketone body use, suppresses hepatic ketone body formation, and promotes bicarbonate regeneration. However, the acidosis and ketosis resolve more slowly than hyperglycemia. Depending on the rise of serum chloride, the anion gap (but not bicarbonate) will normalize more quickly. A hyperchloremic acidosis (serum bicarbonate of 15–18 mmol/L [15–18 meq/L]) often follows successful treatment and gradually resolves as the kidneys regenerate bicarbonate and excrete chloride.

TABLE 416-9 Management of Diabetic Ketoacidosis (DKA)

1. Confirm diagnosis (↑ serum glucose, ↑ serum β-hydroxybutyrate, metabolic acidosis).
2. Admit to hospital; intensive care setting may be necessary for severe DKA (see text). Mild to moderate DKA can be treated in a step-down unit with close nursing and laboratory monitoring.
3. Assess:
    Serum electrolytes (K^+, Na^+, Mg^{2+}, Cl^-, bicarbonate, phosphate)
    Acid-base status—pH, HCO_3^-, PCO_2, β-hydroxybutyrate
    Renal function (creatinine, urine output)
4. Replace fluids: 2–3 L of 0.9% saline or lactated Ringer's over first 1–3 h (10–20 mL/kg per hour); subsequently, 0.45% saline at 250–500 mL/h; change to 5–10% glucose and 0.45% saline or lactated Ringer's at 150–250 mL/h when blood glucose reaches 250 mg/dL (13.9 mmol/L). For treatment of euglycemic DKA, start 5% or 10% dextrose infusion and insulin treatment when 0.9% saline is started; adjust dextrose infusion to prevent hypoglycemia.
5. Administer short-acting regular insulin: IV (0.1 units/kg), then 0.1 units/kg per hour by continuous IV infusion; increase two- to threefold if no response by 2–4 h. In mild to moderate DKA, subcutaneous rapid-acting insulin may be used with close monitoring (0.1 unit/kg rapid-acting insulin analogue subcutaneously and then 0.1 unit/kg every 1 h or 0.2 unit/kg every 2 h). Continue insulin treatment and 5% or 10% dextrose infusion to prevent hypoglycemia. If the initial serum potassium is <3.3 mmol/L (3.3 meq/L), do not administer insulin until the potassium is corrected.
6. Assess patient: What precipitated the episode (noncompliance, infection, trauma, pregnancy, infarction, cocaine)? Initiate appropriate workup for precipitating event (cultures, CXR, ECG, etc.).
7. Measure blood glucose every 1–2 h; measure electrolytes (especially K^+, bicarbonate, phosphate) and anion gap every 4 h for first 24 h.
8. Monitor blood pressure, pulse, respirations, mental status, fluid intake and output every 1–4 h.
9. Replace K^+ if ECG, urine flow, and creatinine are normal. If K^+ <3.5 mmol/L, administer 10–20 mmol/L per hour until K^+ >3.5 mmol/L. If K^+ 3.5–5 mmol/L, administer 10–20 mmol/L in each liter of IV fluid to keep serum K^+ between 4 and 5 mmol/L. If K^+ >5.0 mmol/L, start insulin but hold K^+. Recheck K^+ every 2 h to determine when to start K^+ replacement.
10. Continue above until patient is stable, glucose goal is 8.3–11.1 mmol/L (150–200 mg/dL), normal plasma ketone and pH, and bicarbonate ≥18 mmol/L. Insulin infusion may be decreased to 0.02–0.1 unit/kg per hour. Resolution of euglycemic DKA should be based on bicarbonate, not glucose, correction; see text.
11. Administer long-acting insulin as soon as patient is eating. Allow for a 2- to 4-h overlap in insulin infusion and SC long-acting insulin injection.

Abbreviations: CXR, chest x-ray; ECG, electrocardiogram.

Source: Adapted from multiple sources, including Nyenwe EA, Kitabchi AE: The evolution of diabetic ketoacidosis: An update of its etiology, pathogenesis and management. Metabolism 65:507, 2016; and Umpierrez GE et al: Hyperglycaemic crises in adults with diabetes: A consensus report. Diabetologia 67:1455, 2024.

Potassium stores are depleted in DKA (estimated deficit 3–5 mmol/kg [3–5 meq/kg]). During treatment with insulin and fluids, various factors contribute to the development of hypokalemia. These include insulin-mediated potassium transport into cells, resolution of the acidosis (which also promotes potassium entry into cells), and urinary loss of potassium salts of organic acids. Thus, potassium repletion should commence as soon as adequate urine output and a normal serum potassium are documented (Table 416-9).

Bicarbonate replacement has not been shown to improve outcomes. However, in the presence of severe acidosis (arterial pH <7.0), sodium bicarbonate (50 mmol [meq/L] in 200 mL of sterile water with 10 meq/L KCl per h) may be administered for the first 2 h until the pH is >7.0. Hypophosphatemia and hypomagnesemia may develop during DKA therapy, and if severe, may also require supplementation.

With appropriate therapy, the mortality rate of DKA is low (<1%) and is related more to the underlying or precipitating event, such as infection ((pneumonia, SARS-Co-V2, etc.) COVID-19), pregnancy, end-stage renal disease, or myocardial infarction. Venous thrombosis, upper GI bleeding, and acute respiratory distress syndrome occasionally complicate DKA. The major nonmetabolic complication of DKA therapy is cerebral edema, which most often develops in children as DKA is resolving. The etiology of and optimal therapy for cerebral edema are not well established.

Following treatment, the physician and patient should review the sequence of events that led to DKA to prevent future recurrences. Even a single episode of DKA is associated with a greatly increased 1-year mortality rate. Foremost is patient education about the symptoms of DKA, its precipitating factors, and the management of diabetes during a concurrent illness. In some individuals, DKA is recurrent and may indicate underlying mental health issues. The structural barriers to accessing care, insulin cost, and the social determinants of health often play a role.

HYPERGLYCEMIC HYPEROSMOLAR STATE

Clinical Features The most common presentation of HHS is an elderly individual with type 2 DM, with a several-week history of polyuria, weight loss, and diminished oral intake that culminates in mental confusion, lethargy, or coma. The physical examination reflects profound dehydration and hyperosmolality and reveals hypotension, tachycardia, and altered mental status. Notably absent are symptoms of nausea, vomiting, and abdominal pain and the Kussmaul respirations characteristic of DKA. HHS is often precipitated by a serious, concurrent illness such as myocardial infarction or stroke. Sepsis, pneumonia, and other serious infections are frequent precipitants and should be sought. In addition, a debilitating condition (prior stroke or dementia) or social situation that compromises water intake usually contributes to the development of the disorder.

Pathophysiology Relative insulin deficiency and inadequate fluid intake are the underlying causes of HHS. Insulin deficiency increases hepatic glucose production (through glycogenolysis and gluconeogenesis) and impairs glucose utilization in skeletal muscle (see above discussion of DKA). Hyperglycemia induces an osmotic diuresis that leads to intravascular volume depletion, which is exacerbated by inadequate fluid replacement. The absence of ketosis in HHS is not understood. Presumably, the insulin deficiency is only relative and less severe than in DKA. Lower levels of counterregulatory hormones and free fatty acids have been found in HHS than in DKA in some studies. It is also possible that the liver is less capable of ketone body synthesis or that the insulin/glucagon ratio does not favor ketogenesis.

Laboratory Abnormalities and Diagnosis The laboratory features in HHS are summarized in Table 416-7. Most notable are the marked hyperglycemia (plasma glucose may be >55.5 mmol/L [1000 mg/dL]), hyperosmolality (>300 mOsm/L), and prerenal azotemia. The measured serum sodium may be normal or slightly low despite the marked hyperglycemia. The corrected serum sodium is usually increased (add 1.6 meq to measured sodium for each 5.6-mmol/L [100-mg/dL] rise in the serum glucose). In contrast to DKA, acidosis and ketonemia are absent or mild. A small anion-gap metabolic acidosis may be present secondary to increased lactic acid. Moderate ketonuria, if present, is secondary to starvation.

TREATMENT

Hyperglycemic Hyperosmolar State

Volume depletion and hyperglycemia are prominent features of both HHS and DKA. Consequently, the therapy of these disorders shares several elements (Table 416-9). In both disorders, careful monitoring of the patient's fluid status, laboratory values, and insulin infusion rate is crucial. Underlying or precipitating problems should be aggressively sought and treated. In HHS, fluid losses and dehydration are usually more pronounced than in DKA due to the longer duration of the illness. The patient with HHS is usually older, more likely to have mental status changes, and more likely to have a life-threatening precipitating event with accompanying comorbidities. Even with proper treatment, HHS has a substantially higher mortality rate than DKA (up to 15% in some clinical series).

Fluid replacement should initially stabilize the hemodynamic status of the patient (1–3 L of 0.9% normal saline over the first 2–3 h). Because the fluid deficit in HHS is accumulated over a period of days to weeks, the rapidity of reversal of the hyperosmolar state must balance the need for free water repletion with the risk that too rapid a reversal may worsen neurologic function. If the serum sodium is >150 mmol/L (150 meq/L), 0.45% saline should be used. After hemodynamic stability is achieved, the IV fluid administration is directed at reversing the free water deficit using hypotonic fluids (0.45% saline initially, then 5% dextrose in water [D_5W]). The calculated free water deficit (which can be as great as 9–10 L) should be reversed over the next 1–2 days (infusion rates of 200–300 mL/h of hypotonic solution). Potassium repletion is usually necessary and should be dictated by repeated measurements of the serum potassium. In patients taking diuretics, the potassium deficit can be quite large and may be accompanied by magnesium deficiency. Hypophosphatemia may occur during therapy and can be improved by using KPO_4 and beginning nutrition.

As in DKA, rehydration and volume expansion lower the plasma glucose initially, but insulin is also required. A reasonable regimen for HHS begins with an IV insulin bolus of 0.1 unit/kg followed by IV insulin at a constant infusion rate of 0.1 unit/kg per h. If the serum glucose does not fall, increase the insulin infusion rate by twofold. As in DKA, glucose should be added to IV fluid when the plasma glucose falls to 11.1–13.9 mmol/L (200–250 mg/dL), and the insulin infusion rate should be decreased to 0.02–0.1 unit/kg per h. The insulin infusion should be continued until the patient has resumed eating and can be transferred to an SC insulin regimen. The patient should be discharged from the hospital on insulin. Some patients can later switch to oral glucose-lowering agents.

MANAGEMENT OF DIABETES IN A HOSPITAL OR FACILITY

Virtually all medical and surgical subspecialties are involved in the care of hospitalized patients with diabetes or individuals with diabetes in the perioperative setting. Hyperglycemia, whether in a patient with known diabetes or in someone without known diabetes, appears to be a predictor of poor outcome in hospitalized patients. General anesthesia, surgery, infection, or concurrent illness raises the levels of counterregulatory hormones (cortisol, growth hormone, catecholamines, and glucagon) and cytokines that may lead to transient insulin resistance and hyperglycemia. These factors increase insulin requirements by increasing glucose production and impairing glucose utilization and thus may worsen glycemic control. The concurrent illness or surgical procedure may lead to variable insulin absorption and also prevent the patient with DM from eating normally and, thus, may promote hypoglycemia. Glycemic control should be assessed on admission using the HbA_{1c}. Electrolytes, renal function, and intravascular volume status should be assessed as well. The high prevalence of ASCVD in individuals with DM (especially in type 2 DM) may necessitate preoperative cardiovascular evaluation **(Chap. 417)**. CGM in the hospital or intensive care unit (ICU) setting is not FDA approved. Individuals using CGM and/or AID prior to admission should continue to use these devices if the provider, the nursing staff, and the patient agree that this can be safely done within the context of the patient's current reason for hospitalization.

The goals of diabetes management during hospitalization or in the perioperative periods are near-normoglycemia, avoidance of hypoglycemia, and transition back to the outpatient diabetes treatment regimen. Upon hospital admission, frequent glycemic monitoring should begin, as should planning for diabetes management after discharge. Glycemic control appears to improve clinical outcomes in a variety of settings, but optimal glycemic goals for the hospitalized patient are incompletely defined. In a number of cross-sectional studies of patients with diabetes, a greater degree of hyperglycemia was associated with worse cardiac, neurologic, and infectious outcomes. In some studies, patients who do not have preexisting diabetes but who develop modest blood glucose elevations during their hospitalization appear to benefit from achieving near-normoglycemia using insulin treatment. However, a large randomized clinical trial (Normoglycemia in Intensive Care Evaluation Survival Using Glucose Algorithm Regulation [NICE-SUGAR]) of individuals in the ICU (most of whom were receiving mechanical ventilation) found an increased mortality rate and a greater number of episodes of severe hypoglycemia with very strict glycemic control (target blood glucose of 4.5–6 mmol/L or 81–108 mg/dL) compared to individuals with a more moderate glycemic goal (target blood glucose of <10 mmol/L or 180 mg/dL). Currently, most data suggest that very strict blood glucose control in acutely ill patients likely worsens outcomes and increases the frequency of hypoglycemia. The ADA suggests the following glycemic goals for hospitalized patients: (1) in critically or non-critically ill patients, glucose of 7.8–10.0 mmol/L or 140–180 mg/dL; (2) in selected patients, glucose of 6.1–7.8 mmol/L or 110–140 mg/dL with avoidance of hypoglycemia; and (3) the target range in the perioperative period should be 80–180 mg/dL (4.4–10.0 mmol/L).

Critical aspects for optimal diabetes care in the hospital include the following: (1) A hospital-wide system approach to treatment of hyperglycemia and prevention of hypoglycemia is needed. Inpatient diabetes management teams consisting of nurse practitioners and physicians are increasingly common. (2) Diabetes treatment plans should focus on the transition from the ICU and the transition from the inpatient to the outpatient setting. (3) Adjustment of the discharge treatment regimen of patients whose diabetes was poorly controlled on admission (as reflected by the HbA_{1c}) is important.

The physician caring for an individual with diabetes in the perioperative period, during times of infection or serious physical illness, or simply when the patient is fasting for a diagnostic procedure must monitor the plasma glucose vigilantly, adjust the diabetes treatment regimen, and provide glucose infusion as needed. Hypoglycemia is frequent in hospitalized patients, and many of these episodes are avoidable. Hospital systems should have a diabetes management protocol to avoid inpatient hypoglycemia. Measures to reduce or prevent hypoglycemia include frequent glucose monitoring, but it is also important to prevent hypoglycemia by anticipating drops in insulin requirement by factors such as decreasing renal function, decreasing glucocorticoid doses, or interruption of nutrition (parenteral or enteral or PO).

Depending on the severity of the patient's illness and the hospital setting, the physician can use either an insulin infusion or SC insulin. Insulin infusions are preferred in the ICU or in a clinically unstable setting because the half-life of the infused insulin is quite short (minutes). The absorption of SC insulin may be variable in such situations. Insulin infusions can also effectively control plasma glucose in the perioperative period and when the patient is unable to take anything by mouth, although for relatively short (<4 h) procedures, most patients can remain on SC insulin. Regular insulin is used rather than insulin analogues for IV insulin infusion because it is less expensive and equally effective. The physician must consider carefully the clinical setting in which an insulin infusion will be used, including whether adequate ancillary personnel are available to monitor the blood glucose frequently and whether they can adjust the insulin infusion rate to maintain the blood glucose within the optimal range. Insulin-infusion algorithms should integrate the insulin sensitivity of the patient, frequent blood glucose monitoring, and the trend of changes in the blood glucose to determine the insulin-infusion rate. Insulin-infusion algorithms jointly developed and implemented by nursing and physician staff are advised. Because of the short half-life of IV regular insulin, it is necessary to administer long-acting insulin prior to discontinuation of the insulin infusion (2–4 h before the infusion is stopped) to avoid a period of insulin deficiency.

In patients who are not critically ill or not in the ICU, basal or "scheduled" insulin is provided by SC, long-acting insulin supplemented by prandial and/or "corrective" insulin using a rapid-acting insulin. "Sliding scale," with short-acting or rapid-acting insulin alone, where no insulin is given unless the blood glucose is elevated, is inadequate for inpatient glucose management. The rapid-acting, preprandial insulin dose should include coverage for food consumption (based on anticipated carbohydrate intake) plus corrective insulin based on the patient's insulin sensitivity and the blood glucose. For example, if the patient is thin (and likely insulin-sensitive), an insulin correction factor might be 1 unit for each 2.7 mmol/L (50 mg/dL) over the glucose target. If the patient is obese and likely insulin-resistant, then the insulin correction factor might be 2 units for each 2.7 mmol/L (50 mg/dL) over the glucose target. It is critical to individualize the regimen and adjust the basal and prandial insulin doses frequently based on the corrective insulin required. A consistent carbohydrate-controlled diabetes meal plan for hospitalized patients provides a predictable amount of carbohydrate for a particular meal each day (but not necessarily the same amount for breakfast, lunch, and supper) and avoids concentrated sweets. Individuals with type 1 DM who are undergoing general anesthesia and surgery or who are seriously ill should receive continuous insulin, through an IV insulin infusion, their insulin infusion device, or by SC administration of a reduced dose of long-acting insulin. Rapid-acting insulin alone is insufficient. Prolongation of a surgical procedure or delay in the recovery room is common and may result in periods of insulin deficiency leading to DKA. Insulin infusion is the preferred method for managing patients with type 1 DM over a prolonged (several hours) perioperative period or when serious concurrent illness is present (0.5–1.0 units/h of regular insulin). If the diagnostic or surgical procedure is brief (<4 h), a reduced dose of SC insulin may suffice (20–50% basal reduction, with rapid-acting correctional dose insulin as needed). This approach prevents interruption of insulin infusion device therapy or, for MDI, facilitates the transition back to basal/bolus insulin after the procedure. The blood glucose should be monitored frequently during the illness or in the perioperative period.

Individuals with type 2 DM can be managed with either an insulin infusion or SC long-acting insulin (20–50% reduction depending on clinical setting) plus preprandial, rapid-acting insulin. Oral glucose-lowering agents should be discontinued upon admission (or up to a week prior to planned admission for SGLT-2 inhibitors) and are not useful in regulating the plasma glucose in clinical situations where the insulin requirements and glucose intake are changing rapidly. Moreover, these oral agents may be dangerous if the patient is fasting (e.g., hypoglycemia with sulfonylureas, euglycemic DKA with SGLT-2 inhibitors) or at risk for declining kidney function due to, for example, radiographic contrast media or unstable CHF (lactic acidosis with metformin). Once clinically stable, oral glucose-lowering agents may be resumed in anticipation of discharge. Each patient should receive an individualized, structured discharge plan for diabetes management.

The principles of the care of individuals with diabetes who are in a rehabilitation facility or a long-term care facility are similar to those in a hospitalized patient with the glycemic goals individualized based on the patient's overall clinical status (outlined in Table 416-4). Often the avoidance of hypoglycemia is the major goal with less intense glycemic targets.

SPECIAL CONSIDERATIONS IN DM

■ TOTAL PARENTERAL NUTRITION (TPN)/TOTAL ENTERAL NUTRITION (TEN)

(See also Chap. 335) TPN or TEN greatly increases insulin requirements. In addition, individuals not previously known to have DM may become hyperglycemic during TPN or TEN and require insulin treatment. For TPN, IV insulin infusion is the preferred treatment for hyperglycemia, and rapid titration to the required insulin dose is done most efficiently using a separate insulin infusion. After the total insulin dose has been determined, a proportion of this insulin may be added directly to the TPN solution to cover the nutritional requirements for insulin and adjusted based on the need for modified dosing of rapid-acting insulin. In TEN, hyperglycemia may be limited by using high-protein formulations but often requires insulin treatment. Individuals receiving enteral bolus feedings should receive SC, rapid-acting insulin prior to each bolus. As a start, 1 unit of insulin is given SC for each 10–15 g of carbohydrate in the bolus. Patients with insulin deficiency (type 1 DM and pancreatogenic DM) should also receive long-acting insulin (0.1–0.2 units/kg per day) to cover basal insulin requirements should the TPN or TEN be interrupted or cycled.

■ GLUCOCORTICOIDS

Glucocorticoids increase insulin resistance, decrease glucose utilization, increase hepatic glucose production, and impair insulin secretion. These changes lead to a worsening of glycemic control in individuals with DM and may precipitate hyperglycemia in other individuals. If new-onset hyperglycemia remains during chronic treatment with supraphysiologic doses of glucocorticoid (>5 mg of prednisone or equivalent), the DM may be called "steroid-induced diabetes." The effects of glucocorticoids on glucose homeostasis are dose-related, usually reversible, most pronounced in the postprandial period, and dependent on the timing and type of glucocorticoid. If the FPG is near the normal range, oral diabetes agents (e.g., sulfonylureas, metformin) may be sufficient to reduce hyperglycemia. If the FPG is >11.1 mmol/L (200 mg/dL), oral agents are usually not sufficient, and insulin therapy is required. If steroids are administered in the morning, then short-acting insulin and/or NPH in the morning may be sufficient to control postprandial glucose excursions.

■ DIABETES MANAGEMENT IN OLDER ADULTS

Diabetes is very common in older adults, being present in ~25% of individuals over the age of 65 years. Increasingly, individuals with many years of type 1 DM are part of this patient population. As discussed above (Table 416-4), individualized therapeutic goals and modalities in older adults should consider biologic age, other comorbidities and risk factors (e.g., hypertension, CVD), neurocognitive and physical functional status, living arrangements, social support, and other medications. For example, the HbA$_{1c}$ goal for a highly functional 80-year-old should be different from that for an individual with diabetes in long-term care (skilled nursing facilities). In the former, the HbA$_{1c}$ goal (<7.0–7.5%) and selected therapies may be similar to younger individuals, whereas in an individual with complex/poor health or cognitive impairment, an HbA$_{1c}$ goal of <8.0–8.5% would be reasonable. Critical to diabetes management in all older individuals is the avoidance of hypoglycemia, which can worsen underlying cognitive impairment or CVD. In choosing medications for diabetes, the adverse effects (Table 416-6) should be considered (especially heart failure, renal insufficiency, propensity for hypoglycemia, etc.). Hypertension and dyslipidemia should be treated in elderly individuals with diabetes because there is clear benefit of blood pressure control, with the benefit for lipid-lowering medications being less clearly demonstrated.

REPRODUCTIVE ISSUES

Reproductive capacity in either men or women with DM appears to be normal. Menstrual cycles may be associated with alterations in glycemic control in women with DM. Pregnancy is associated with marked insulin resistance; the increased insulin requirements often precipitate DM and lead to the diagnosis of gestational diabetes mellitus (GDM). Glucose, which at high levels is a teratogen to the developing fetus, readily crosses the placenta, but insulin does not. Thus, hyperglycemia from the maternal circulation may stimulate insulin secretion in the fetus. The anabolic and growth effects of insulin may result in macrosomia. GDM complicates ~7% (range 1–14%) of pregnancies. The incidence of GDM is greatly increased in certain racial and ethnic groups, including Black and Hispanic, consistent with a similar increased risk of type 2 DM. Current recommendations advise screening for glucose intolerance between weeks 24 and 28 of pregnancy in women not known to have diabetes. Therapy for GDM is similar to that for individuals with pregnancy-associated diabetes and involves MNT and insulin, if hyperglycemia persists. Oral glucose-lowering agents are not approved for use during pregnancy, but studies using metformin or glyburide have shown efficacy and have not found toxicity. With current practices, the morbidity and mortality rates of the mother with GDM and the fetus are not different from those in the nondiabetic population. Individuals who develop GDM are at marked increased risk for developing type 2 DM in the future and should be screened periodically for DM (see screening recommendations in Chap. 415). Most individuals with GDM revert to normal glucose tolerance after delivery, but some will continue to have overt diabetes or impairment of glucose tolerance after delivery. In addition, children of women with GDM appear to be at risk for obesity and glucose intolerance and have an increased risk of diabetes beginning in the later stages of adolescence.

Pregnancy in individuals with known DM requires meticulous planning and adherence to strict preconception treatment regimens. Intensive insulin therapy and near-normalization of the HbA$_{1c}$ (<6.5%) are essential for individuals with existing DM who are planning pregnancy. Consideration should be given to insulin infusion (e.g., AID) and CGM devices that may help to improve glycemic control prior to conception since the most crucial period of glycemic control is soon after fertilization. The risk of fetal malformations is increased 4–10 times in individuals with uncontrolled DM at the time of conception, and normal blood glucose during the preconception period and throughout the periods of organ development in the fetus should be the goal, with more frequent monitoring of HbA$_{1c}$ every 2 months throughout gestation. Maintenance of the HbA$_{1c}$ <6.0–6.5% reduces the incidence and severity of fetal macrosomia and neonatal hypoglycemia related to fetal hyperinsulinism driven by elevated maternal glucose.

FURTHER READING

American Diabetes Association: Comprehensive medical evaluation and assessment of comorbidities: Standards of Medical Care in Diabetes—2024. Diabetes Care 47:S52, 2024.

American Diabetes Association: Facilitating positive health behaviors and well-being to improve health outcomes: Diabetes—2024. Diabetes Care 44:S77, 2024.

American Diabetes Association: Pharmacologic approaches to glycemic treatment: Standards of medical care in diabetes—2024. Diabetes Care 47:S158, 2024.

American Diabetes Association: Obesity and weight management for the prevention and treatment of type 2 diabetes: Standards of care in diabetes—2024. Diabetes Care 47:S145, 2024.

American Diabetes Association: Older adults: Standards of medical care in diabetes—2024. Diabetes Care 47:S244, 2024.

Chow E et al: Euglycemic diabetic ketoacidosis in the era of SGLT2 inhibitors. BMJ Open Diabetes Res Care 11:e003666, 2023.

Hirsch IB et al: The evolution of insulin and how it informs therapy and treatment choices. Endocr Rev 41:733, 2020.

Holt RIG et al: The management of type 1 diabetes in adults. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 44:2589, 2021.

Kosiborod MN et al: Semaglutide in patients with obesity-related heart failure and type 2 diabetes. N Engl J Med 390:15, 2024.

Mallik R et al: The future is here: An overview of technology in diabetes. Diabetologia 67:2019, 2024.

Perkovic V et al: Effects of semaglutide on chronic kidney disease in patients with type 2 diabetes. N Engl J Med 391:2, 2024.

Qaseem A et al: Newer pharmacologic treatments in adults with type 2 diabetes: A clinical guideline from the American College of Physicians. Ann Intern Med 177:658, 2024.

Shaltout I et al: Risk stratification in people with diabetes for fasting during Ramadan: Consensus from Arabic Association for the Study of Diabetes and Metabolism. Curr Diabetes Rev 20:e201023222409, 2024.

Simmons D et al: Treatment of gestational diabetes mellitus diagnosed early in pregnancy. N Engl J Med 388:2132, 2023.

Umpierrez GE et al: Hyperglycaemic crises in adults with diabetes: a consensus report. Diabetologia 67:1455, 2024.

WEBSITES

Online resources for selection of diabetes technology:
Diabeteswise: *https://pro.diabeteswise.org/en/*
Diatribe: *https://diatribe.org/*
Panther: *https://www.pantherprogram.org/*

417 Diabetes Mellitus: Complications

Alvin C. Powers, John M. Stafford,
Michael R. Rickels

Diabetes-related complications affect many organ systems and are responsible for most of the morbidity and mortality associated with the disease. For many years in the United States, diabetes has been a leading cause of new blindness in adults, renal failure, and nontraumatic lower extremity amputation and is a leading contributor to coronary heart disease (CHD). Diabetes-associated microvascular complications usually do not appear until the second decade of hyperglycemia. In contrast, diabetes-associated atherosclerotic cardiovascular disease (ASCVD) risk, related in part to insulin resistance and its resultant dyslipidemia, may develop before hyperglycemia is established. Because type 2 diabetes mellitus (DM) often has a long asymptomatic period of hyperglycemia before diagnosis, many

TABLE 417-1 Diabetes-Related Complications

- Microvascular
  - Eye disease
    - Retinopathy (nonproliferative/proliferative)
    - Macular edema
  - Neuropathy
    - Sensory and motor (mono- and polyneuropathy)
    - Autonomic
  - Nephropathy (albuminuria and declining renal function)
- Macrovascular
  - Coronary heart disease
  - Peripheral arterial disease
  - Cerebrovascular disease
- Heart failure
- Other
  - Gastrointestinal (gastroparesis, diarrhea)
  - Genitourinary (uropathy/sexual dysfunction)
  - Dermatologic
  - Infectious
  - Cataracts
  - Glaucoma
  - Cheiroarthropathy[a]
  - Periodontal disease
  - Hearing loss
- Other comorbid conditions associated with type 1 or type 2 diabetes (relationship to hyperglycemia is uncertain): depression, obstructive sleep apnea, fatty liver disease, hip fracture, osteoporosis, cognitive impairment or dementia, low testosterone in men

[a] Thickened skin and reduced joint mobility.

individuals with type 2 DM have both glucose-related and insulin resistance–related complications at the time of diagnosis. Fortunately, many of the diabetes-related complications can be prevented or mitigated with aggressive glycemic, lipid, and blood pressure control, as well as efforts at early detection. Diagnosis of type 2 DM at younger age increases diabetes-related complications. One estimate indicated three to four years of reduced life expectancy for every decade of earlier diabetes diagnosis. This is emphasizes the critical role of diabetes prevention or delay.

Diabetes-related complications can be divided into vascular and nonvascular complications and are similar for type 1 and type 2 DM **(Table 417-1)**. The vascular complications of DM are further subdivided into microvascular (retinopathy, neuropathy, and nephropathy) and macrovascular complications (ASCVD, peripheral arterial disease [PAD], cerebrovascular disease, and heart failure). Microvascular complications are diabetes specific, whereas macrovascular complications have additional pathophysiologic features that are shared with the general population. Nonvascular complications include infections, skin changes, cheiroarthropathy, hearing loss, and increased risk of fractures, dementia, and impaired cognitive function.

■ GLYCEMIC CONTROL AND COMPLICATIONS

The microvascular complications of both type 1 and type 2 DM result from chronic hyperglycemia **(Fig. 417-1)**. Evidence implicating a causative role for chronic hyperglycemia in the development of macrovascular complications is less conclusive with other factors such as dyslipidemia and hypertension playing more important roles. ASCVD events and mortality rate are two to four times greater in patients with type 2 DM, correlate with fasting and postprandial plasma glucose levels as well as the hemoglobin A_{1c} (HbA_{1c}), and can be reduced by intensive diabetes management as demonstrated in patients with type 1 DM.

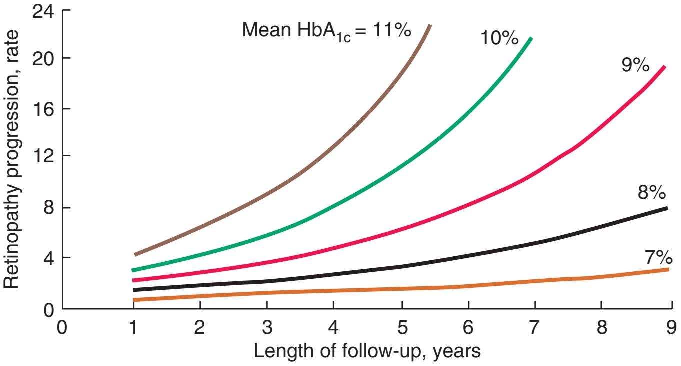

FIGURE 417-1 Relationship of glycemic control and diabetes duration to diabetic retinopathy. The progression of retinopathy in individuals in the Diabetes Control and Complications Trial is graphed as a function of the length of follow-up with different curves for different hemoglobin A_{1c} (HbA_{1c}) values. *(Reproduced with permission from The relationship of glycemic exposure (HbA1c) to the risk of development and progression of retinopathy in the diabetes control and complications trial. Diabetes 44:968, 1995.)*

The Diabetes Control and Complications Trial (DCCT) provided definitive proof that reduction in chronic hyperglycemia can prevent many complications of type 1 DM (Fig. 417-1). This large multicenter clinical trial randomized >1400 individuals with type 1 DM to either intensive or conventional diabetes management and prospectively evaluated the development of diabetes-related complications during a mean follow-up of 6.5 years. Individuals in the intensive diabetes management group received insulin by multiple daily injections or pump delivery along with extensive educational, psychological, and medical support, and achieved a substantially lower HbA_{1c} (7.3%) than individuals in the conventional diabetes management group (9.1%). After the DCCT results were reported in 1993, all study participants were offered intensive therapy and continue to be followed in the Epidemiology of Diabetes Intervention and Complications (EDIC) trial, which has completed >40 years of follow-up (DCCT + EDIC). When the DCCT phase ended at 6.5 years of follow-up, the initial separation in glycemic control disappeared with both arms maintaining a mean HbA_{1c} of 8.0%, allowing assessment of the legacy effect of 6.5 years of near-normoglycemia on the development of long-term complications.

The DCCT phase demonstrated that improvement of glycemic control reduced nonproliferative and proliferative retinopathy (47% reduction), albuminuria (39% reduction), clinical nephropathy (54% reduction), and neuropathy (60% reduction). Improved glycemic control also slowed the progression of early diabetic complications. During the DCCT phase, weight gain (4.6 kg) and severe hypoglycemia (requiring assistance of another person to treat) were more common in the intensive therapy group. The benefits of an improvement in glycemic control occurred over the entire range of elevated HbA_{1c} values (Fig. 417-1). The results of the DCCT predicted that individuals in the intensive diabetes management group would gain 7.7 additional years of vision, 5.8 additional years free from end-stage renal disease, and 5.6 years free from lower extremity amputations. If all complications of DM were combined, individuals in the intensive diabetes management group would experience >15.3 more years of life without significant microvascular complications of DM, compared to individuals who received standard therapy. This translates into an additional 5.1 years of life expectancy for individuals in the intensive diabetes management group. The 30-year follow-up data in the intensively treated group show a continued reduction in retinopathy, nephropathy, and cardiovascular disease. For example, individuals in the intensive therapy group had a 57% reduction in cardiovascular events (nonfatal myocardial infarction [MI], stroke, or death from a cardiovascular event) and a 33% reduction in the mortality rate, even though their subsequent glycemic control was the same as those in the conventional diabetes management group after the DCCT phase ended (year 6.5). During the EDIC phase, fewer in the intensely treated cohort became blind, lost a limb to amputation, or required dialysis. Other complications of diabetes, including autonomic neuropathy, bladder and sexual dysfunction, cardiac autonomic neuropathy, cheiroarthropathy and hearing loss, were reduced in the intensive therapy group. These results are even more impressive when one considers that initial DCCT results were reported in 1993 and diabetes therapy during the trial was quite different in terms of insulin formulations and delivery systems. Fingerstick blood glucose meters were used for glucose monitoring as this was prior to the advent of continuous glucose monitoring.

The United Kingdom Prospective Diabetes Study (UKPDS) studied the course of >5000 individuals with type 2 DM for >10 years. This study used multiple treatment regimens and monitored the effect of intensive glycemic control and risk factor treatment on the development of diabetes-related complications. Newly diagnosed individuals with type 2 DM were randomized to (1) intensive management using various combinations of insulin, a sulfonylurea, or metformin or (2) conventional therapy using dietary modification and pharmacotherapy with the goal of symptom prevention. In addition, individuals were randomly assigned to different antihypertensive regimens. Individuals in the intensive treatment arm achieved an HbA_{1c} of 7% compared to 7.9% in the standard treatment group. The UKPDS demonstrated that each percentage point reduction in HbA_{1c} was associated with a 35% reduction in microvascular complications. As in the DCCT, there was a continuous relationship between glycemic control and development of complications. Improved glycemic control also reduced the cardiovascular event rate in the follow-up period of >10 years.

One of the major findings of the UKPDS was that strict blood pressure control significantly reduced both macro- and microvascular complications. In fact, the beneficial effects of blood pressure control were greater than the beneficial effects of glycemic control. Lowering blood pressure to moderate goals (144/82 mmHg) reduced the risk of DM-related death, stroke, microvascular endpoints, retinopathy, and heart failure (risk reductions between 32 and 56%). The American Diabetes Association (ADA) recommends blood pressure control <130/80 mmHg. In the UKPDS, improved glycemic control early in the course of diabetes with a sulfonylurea or insulin, or with metformin, subsequently reduced risk of death and MI. Other trials such as the Action to Control Cardiovascular Risk in Diabetes (ACCORD) and Action in Diabetes and Vascular Disease: Preterax and Diamicron MR Controlled Evaluation (ADVANCE) trials also found that improved glycemic control reduced microvascular complications.

Thus, large clinical trials in type 1 and type 2 DM indicate that chronic hyperglycemia plays a causative role in the pathogenesis of diabetic micro- and macrovascular complications. In both the DCCT and the UKPDS, cardiovascular events were reduced at follow-up of >10 years, even though the improved glycemic control was not maintained. This legacy effect for a positive impact of a period of improved glycemic control on later diabetes complications has been termed *metabolic memory*, and this legacy effect was estimated to be 10 years or more. Of note, despite long-standing DM, some individuals never develop retinopathy or nephropathy, suggesting a genetic susceptibility for developing particular complications.

■ MECHANISMS OF COMPLICATIONS

Chronic hyperglycemia is the important etiologic factor leading to complications of DM, but the mechanism(s) by which it leads to such diverse cellular and organ dysfunction is unknown. The complications are likely multifactorial with an emerging hypothesis that hyperglycemia leads to epigenetic changes **(Chap. 479)** that influence gene expression in affected cells. Chronic hyperglycemia leads to formation of advanced glycosylation end products (AGEs; e.g., pentosidine, glucosepane, and carboxymethyllysine), which bind to specific cell surface receptor and/or the nonenzymatic glycosylation of intra- and extracellular proteins, leading to cross-linking of proteins, glomerular dysfunction, endothelial dysfunction, altered extracellular matrix composition, and accelerated atherosclerosis.

Growth factors may play an important role in some diabetes-related microvascular complications. For example, vascular endothelial growth factor A (VEGF-A) is increased locally in diabetic proliferative retinopathy, decreases after laser photocoagulation, and is the target inhibited by intravitreous injection therapy. A possible unifying mechanism is that hyperglycemia leads to increased production of reactive oxygen species or superoxide in the mitochondria and this may activate several pathways. Although hyperglycemia serves as the initial trigger for complications of diabetes, it is still unknown whether the same pathophysiologic processes are operative in all complications or whether some pathways predominate in certain organs.

The mechanisms of diabetes-related macrovascular complications including MI and stroke also include traditional cardiovascular risk factors (dyslipidemia, hypertension), insulin resistance, and inflammation. In T2DM, insulin resistance is present years prior to diagnosis and is associated with obesity and ectopic accumulation of lipids and fat in liver and muscle. Additionally, insulin fails to appropriately suppress lipolysis from adipose tissue, which results in increased delivery of fatty acids to liver, muscle, endothelial cells, and cardiac tissues, leading to tissue accumulation of triglycerides, diacylglycerol, and ceramides.

■ OPHTHALMOLOGIC COMPLICATIONS OF DIABETES MELLITUS

DM is the leading cause of new cases of blindness between the ages of 20 and 74 in the United States. Glaucoma and cataracts occur earlier and more frequently in individuals with diabetes. Severe vision loss is primarily the result of progressive diabetic retinopathy, which leads to significant macular edema and new blood vessel formation. Diabetic retinopathy is classified into two stages: nonproliferative and proliferative. Nonproliferative diabetic retinopathy usually appears late in the first decade or early in the second decade of hyperglycemia and is marked by retinal vascular microaneurysms, blot hemorrhages, and cotton-wool spots **(Fig. 417-2)**. Mild nonproliferative retinopathy may progress to more extensive disease, characterized by changes in venous vessel caliber, intraretinal microvascular abnormalities, and more numerous microaneurysms and hemorrhages. The pathophysiologic mechanisms invoked in nonproliferative retinopathy include loss of retinal pericytes, increased retinal vascular permeability, alterations in retinal blood flow, and abnormal retinal microvasculature, all of which can lead to retinal ischemia.

The appearance of neovascularization in response to retinal hypoxemia is the hallmark of proliferative diabetic retinopathy (Fig. 417-2). These newly formed vessels appear near the optic nerve and/or macula and rupture easily, leading to vitreous hemorrhage, fibrosis, and ultimately retinal detachment. Not all individuals with nonproliferative retinopathy go on to develop proliferative retinopathy, but the more severe the nonproliferative disease, the greater is the chance of evolution to proliferative retinopathy within 5 years. This creates an important opportunity for early detection and treatment of diabetic retinopathy. Clinically significant macular edema can occur in the context of nonproliferative or proliferative retinopathy. Fluorescein angiography and optical coherence tomography are useful to detect macular edema, which is associated with an increased chance of moderate visual loss over the next 3 years. Duration of DM and degree of glycemic control are the best predictors of the development of retinopathy; hypertension, nephropathy, and dyslipidemia are also risk factors. Although there is genetic susceptibility for retinopathy, it confers less influence than either the duration of DM or the degree of glycemic control.

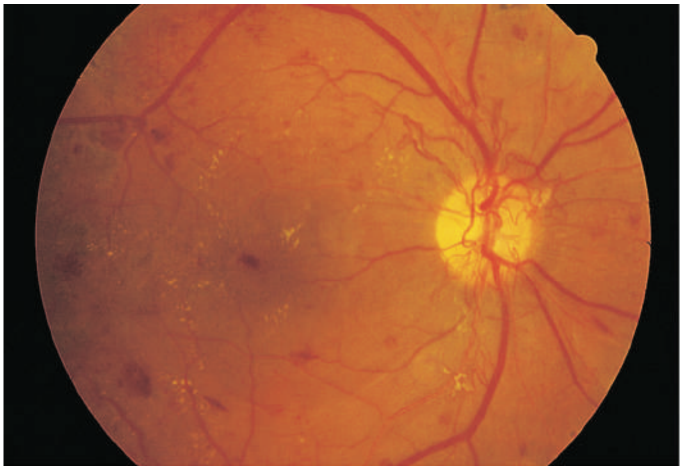

FIGURE 417-2 Diabetic retinopathy results in scattered hemorrhages, yellow exudates, and neovascularization. This patient has neovascular vessels proliferating from the optic disc, requiring urgent panretinal laser photocoagulation.

TREATMENT

Diabetic Retinopathy

The most effective therapy for diabetic retinopathy is prevention. Intensive glycemic and blood pressure control will delay the development and slow the progression of retinopathy in individuals with either type 1 or type 2 DM. Paradoxically, during the first 6–12 months of improved glycemic control, established diabetic retinopathy may transiently worsen. Fortunately, this progression is temporary, and in the long term, improved glycemic control is associated with less diabetic retinopathy. When associated with a marked glycemic improvement, glucagon-like peptide 1 (GLP-1) receptor agonists have been associated with an increased risk of worsening diabetic retinopathy; this should be considered when choosing agents to improve in glycemic control. Individuals with retinopathy may be candidates for prophylactic laser photocoagulation when initiating intensive therapy, and especially prior to pancreas or islet transplantation that can rapidly normalize glycemia. Women with type 1 or type 2 DM who are planning pregnancy should be screened prior to and during pregnancy. Once advanced retinopathy is present, improved glycemic control imparts less benefit. Appropriate ophthalmologic care can prevent most blindness. Lowering elevated levels of triglycerides with fenofibrate may also reduce the progression of retinopathy.

Regular, comprehensive eye examinations are essential for all individuals with DM **(see Table 416-1)**. Most diabetic eye disease can be successfully treated if detected early. Routine, nondilated eye examinations by the primary care provider or diabetes specialist are inadequate to detect diabetic eye disease, which requires a dilated eye exam performed by an optometrist or ophthalmologist or by retinal photography with remote reading. Subsequent management should be by a retinal specialist. Treatment of severe nonproliferative or proliferative retinopathy or macular edema with panretinal laser photocoagulation therapy and/or anti-VEGF therapy (intravitreous injection) usually is successful in preserving vision. Aspirin therapy does not appear to influence the natural history of diabetic retinopathy, and antiplatelet agents and anticoagulation may be continued in patients receiving intravitreal injections of anti-VEGF agents. Patients with severe proliferative retinopathy with vitreous hemorrhage and/or traction involving the macula often require surgical vitrectomy.

■ RENAL COMPLICATIONS OF DIABETES MELLITUS

Diabetic nephropathy is the leading cause of chronic kidney disease (CKD) and stage 5 CKD (e.g., end-stage renal disease; see **Chap. 322**) requiring renal replacement therapy. CKD in individuals with DM is associated with an increased risk of cardiovascular disease, and the prognosis of individuals with diabetes on dialysis is poor. Individuals with type 1 DM and diabetic nephropathy commonly also have diabetic retinopathy; this association is less pronounced in type 2 DM. The presence of CKD without retinopathy in type 1 DM should prompt investigation for alternative causes of kidney disease. Approximately 20–40% of patients with diabetes develop diabetic nephropathy. Known risk factors include a family history of diabetic nephropathy with additional genetic or environmental susceptibility factors likely contributing. Smoking accelerates the decline in renal function. Diabetic nephropathy and stage 5 CKD (e.g., end-stage renal disease; see **Chap. 322**) secondary to DM develop more commonly in Black, Native American, and Hispanic individuals.

Like other microvascular complications, the pathogenesis of diabetic nephropathy is related to chronic hyperglycemia. The mechanisms by which chronic hyperglycemia leads to diabetic nephropathy are incompletely defined but involve the effects of soluble factors (growth factors, angiotensin II, endothelin, AGEs), epigenetic changes, hemodynamic alterations in the renal microcirculation (glomerular hyperfiltration or hyperperfusion, increased glomerular capillary pressure), structural changes in the glomerulus (increased extracellular matrix, basement membrane thickening, mesangial expansion, fibrosis), and tubular dysfunction (tubulointerstitial damage, fibrosis). **See Chap. 322 for additional discussion.**

The natural history of diabetic nephropathy is characterized by a sequence of events that was initially defined for individuals with type 1 DM but appears similar in type 2 DM **(Fig. 417-3)**. Glomerular hyperperfusion and renal hypertrophy occur in the first years after the onset of DM and are associated with an increase of the estimated glomerular filtration rate (GFR). During the first 5 years of DM, thickening of the glomerular basement membrane, glomerular hypertrophy, and mesangial volume expansion occur as the GFR returns to normal. Once there is marked albuminuria and a reduction in GFR, these pathologic changes are likely irreversible.

As part of comprehensive diabetes care **(Chap. 416)**, diabetic nephropathy should be detected at an early stage when effective therapies can be instituted. Because some individuals with DM may have a decline in GFR in the absence of albuminuria, assessment should include both urinary albumin-to-creatinine ratio (UACR) on a spot specimen and an estimated GFR (eGFR). The urine protein measurement by routine urinalysis does not detect low levels of albumin excretion. Screening for albuminuria should commence 5 years after type 1 DM onset and at the time of diagnosis of type 2 DM and be performed annually. An elevated UACR should be confirmed on two to three occasions over a 3- to 6-month period since it can be falsely elevated by strenuous exercise at a time close to its measurement, infection, fever, congestive heart failure, marked hyperglycemia, marked hypertension, or prostate disease.

The ADA defines albuminuria as a persistently increased UACR >30 mg/g. Albuminuria should be quantified, with a moderate increase defined as 30–299 mg/g creatinine and severely elevated as >300 mg/g creatinine. The UACR is a continuous variable, but the greater the degree of albuminuria the more likely there is a reduced GFR. Elevations in the UACR are associated with an increased risk of cardiovascular disease. Once increased, the UACR should be measured more frequently (2–4 times/year).

Type IV renal tubular acidosis (hyporeninemic hypoaldosteronism) may occur in type 1 or 2 DM. These individuals develop a propensity to hyperkalemia and acidemia, which may be exacerbated by medications (especially angiotensin-converting enzyme [ACE] inhibitors, angiotensin receptor blockers [ARBs], and mineralocorticoid receptor antagonists). Patients with DM are predisposed to radiocontrast-induced nephrotoxicity. Risk factors for radiocontrast-induced nephrotoxicity are preexisting nephropathy and volume depletion. Individuals with DM undergoing radiographic procedures with iodinated contrast dye should be well hydrated before and after dye exposure, and the serum creatinine should be monitored for 24–48 h following the procedure.

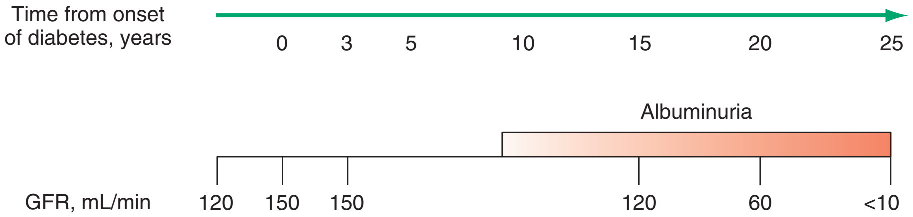

FIGURE 417-3 Time course of development of diabetic nephropathy. The relationship of time from onset of diabetes, albuminuria (urinary albumin-to-creatinine ratio [UACR]), and the glomerular filtration rate (GFR) are shown. This figure is typical for type 1 diabetes; individuals with type 2 diabetes may present with a lower GFR at the time of diagnosis.

Metformin should be held until postintervention confirmation of preserved kidney function.

TREATMENT

Diabetic Nephropathy

The optimal therapy for diabetic nephropathy is prevention by control of glycemia and blood pressure (blood pressure <130/80 mmHg) **(Chap. 416 outlines glycemic goals and approaches)**. Renin-angiotensin-aldosterone system inhibitors do not prevent the development of diabetic kidney disease if hypertension or albuminuria is not present. Interventions effective in slowing progression of albuminuria and the decline in kidney function include (1) improved glycemic control, (2) strict blood pressure control, (3) administration of an ACE inhibitor or ARB, (4) in individuals with type 2 DM, administration of a sodium-glucose cotransporter 2 (SGLT-2) inhibitor and (5) administration of a mineralocorticoid receptor antagonist (especially finerenone). Dyslipidemia should also be treated.

Improved glycemic control reduces the rate at which albuminuria appears and progresses in type 1 and type 2 DM. However, once there is a moderate level of albuminuria, it becomes more difficult for improved glycemic control to slow progression of renal disease, although 10 years of normoglycemia resulting from pancreas transplantation may lead to regression of mesangial glomerular lesions **(Fig. 417-4)**. During the late phase of declining renal function, insulin requirements may fall as the kidney is a site of insulin degradation. As the GFR decreases with progressive nephropathy, the use and dose of glucose-lowering agents should be reevaluated **(see Table 416-6)**. Some glucose-lowering medications (sulfonylureas and metformin) are contraindicated in advanced renal insufficiency, while others may require dose adjustment (glinides and DPP-4 inhibitors). SGLT2 inhibitors are not effective with eGFR < 20 mL/min/1.73 m^2.

HYPERTENSION

Many individuals with type 1 or type 2 DM develop hypertension. Hypertension accelerates complications of DM, particularly ASCVD, nephropathy, and retinopathy. Blood pressure should be measured at every clinic visit; individuals should also be encouraged to monitor their blood pressure at home. The blood pressure goal should be <130/80 mmHg in individuals with diabetes and possibly lower in individuals at increased risk for ASCVD or CKD progression. Because of the high prevalence of ASCVD disease in individuals with type 2 DM, the possibility of renovascular hypertension should be considered when the blood pressure is not readily controlled.

In addition to medications, therapy should include lifestyle modifications, including weight loss, exercise, stress management, sodium restriction, Dietary Approaches to Stop Hypertension (DASH)–style eating, and smoking cessation. In younger individuals or those with increased cardiovascular risk, the provider may recommend a lower target blood pressure. If the blood pressure is >150/90 mmHg, initial treatment should consist of two antihypertensive medications. Multiple agents are often required to control blood pressure. In pregnant individuals with diabetes and chronic hypertension, blood pressure control is associated with better pregnancy outcomes.

In the absence of kidney disease, ACE inhibitors or ARBs are effective antihypertensives but are no more effective than other antihypertensive classes such as thiazide-like diuretics and dihydropyridine calcium channel blockers. There is no benefit of intervention prior to onset of albuminuria or using a combination of an ACE inhibitor and an ARB. If use of either ACE inhibitors or ARBs is not possible or the blood pressure is not controlled, then diuretics, calcium channel blockers (nondihydropyridine class), or beta blockers (with caution in individuals at increased risk for experiencing hypoglycemia) may be used. Mineralocorticoid receptor antagonists can

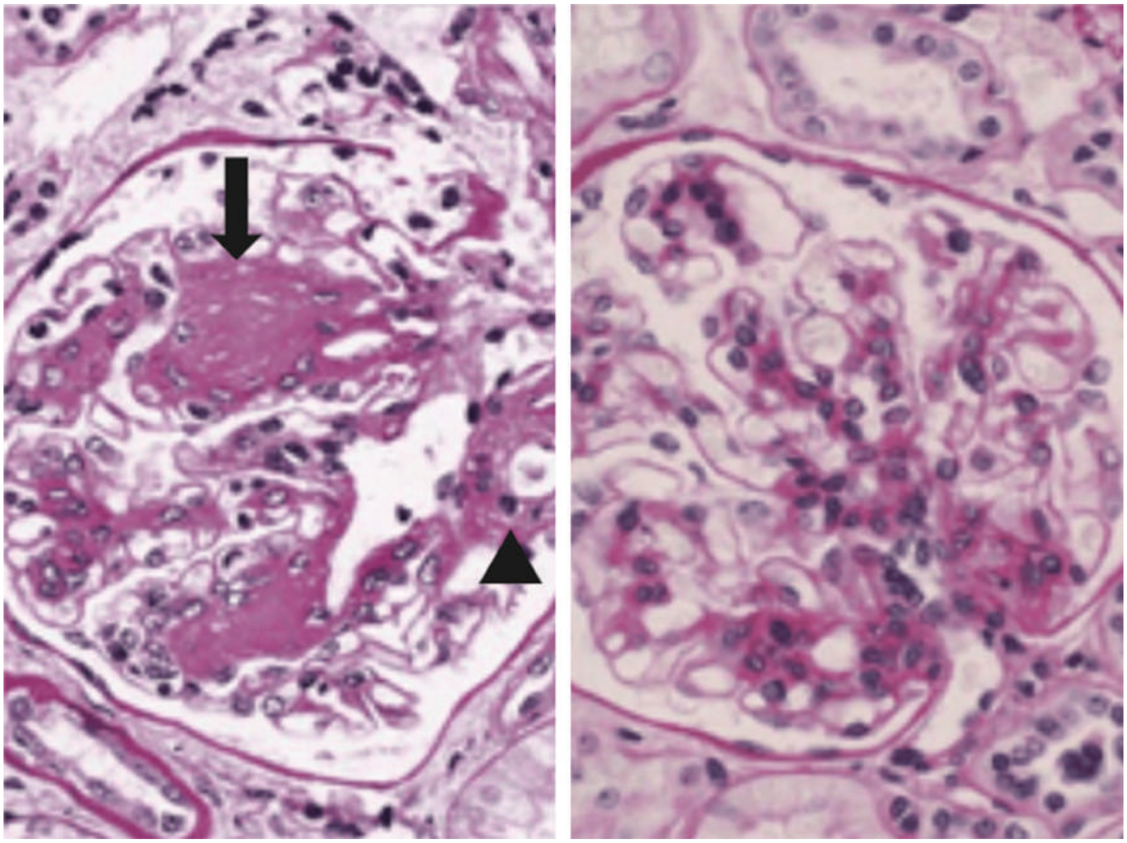

FIGURE 417-4 Diabetic glomerular changes in a patient with type 1 diabetes are reversed by 10 years of normoglycemia as a result of pancreas transplantation. Left panel shows diabetic glomerulosclerosis (*arrow*) and arteriolar hyalinosis (*arrowhead*) on kidney biopsy. Right panel shows a near-normal glomerulus in the same patient after 10 years of normoglycemia from pancreas transplantation. *(Reproduced with permission from P Fioretto et al: Reversal of lesions of diabetic nephropathy after pancreas transplantation. N Engl J Med 339:69, 1998.)*

help reduce blood pressure and albuminuria in refractory cases but require close monitoring of the serum potassium.

ALBUMINURIA OR CKD

Either ACE inhibitors or ARBs should be used to reduce albuminuria and slow the decline in GFR in individuals with type 1 or type 2 DM. Most experts consider the two classes of drugs to be equivalent in patients with diabetes. ARBs can be used as an alternative in patients who develop ACE inhibitor–associated cough or angioedema. After initiation of therapy, one should increase to the maximum tolerated dose while monitoring the serum creatinine and potassium and repeating the UACR one to four times per year. A rise in the serum creatinine up to 30% is acceptable. A goal for individuals with a UACR >300 mg/g creatinine is to reduce the UACR by 30%.

To reduce CKD progression and cardiovascular events in individuals with CKD, type 2 DM, and an eGFR >20 mL/min per 1.73 m^2, the addition of an SGLT-2 inhibitor, while continuing an ACE inhibitor or ARB, is recommended with any level of albuminuria. A GLP-1 agonist or a nonsteroidal mineralocorticoid receptor antagonist like finerenone will also reduce cardiovascular risk in individuals with type 2 DM and CKD. The GLP-1 receptor agonist semaglutide improves kidney outcomes and reduces death from cardiovascular causes in type 2 DM and CKD. SGLT-2 inhibitors are also discussed in **Chap. 265**, especially the use in heart failure treatment or prevention, and in **Chap. 322**, as related to CKD. Because of the elevated risk of euglycemic diabetic ketoacidosis, SGLT-2 inhibitors in individuals with type 1 DM and insulin-deficient type 2 DM should be used with caution and include patient education about ketone monitoring and recognizing diabetic ketoacidosis.

Nephrology consultation is indicated when the estimated GFR is <30 mL/min per 1.743 m^2, albuminuria is >300 mg/g creatinine, or if there are atypical features such as hematuria or rapidly declining renal function. The ADA suggests a protein intake of 0.8 g/kg of body weight per day in individuals with diabetic kidney disease. Complications of ASCVD are the leading cause of death in diabetic individuals with nephropathy; hyperlipidemia should be treated aggressively. Preemptive (before dialysis) kidney transplantation from a living donor should be considered in those nearing stage 5 CKD (e.g., end-stage renal disease; see **Chap. 333**) and for those with type 1 DM or insulin deficient type 2 DM, simultaneous pancreas-kidney transplantation from a deceased donor may be an option. As compared with nondiabetic individuals, hemodialysis in patients with DM is associated with more frequent complications, such as hypotension (due to autonomic neuropathy or loss of reflex tachycardia), more difficult vascular access, accelerated progression of retinopathy, and greater mortality.

NEUROPATHY AND DIABETES MELLITUS

Diabetic neuropathy, which occurs in ~50% of individuals with long-standing type 1 and type 2 DM, manifests as a diffuse neuropathy (distal symmetrical polyneuropathy and/or autonomic neuropathy), a mononeuropathy, and/or a radiculopathy/polyradiculopathy. As with other complications of DM, the development of neuropathy correlates with the duration of diabetes and glycemic control. Additional risk factors are body mass index (BMI) (the greater the BMI, the greater the risk of neuropathy) and smoking. The presence of ASCVD, elevated triglycerides, and hypertension is also associated with diabetic peripheral neuropathy. Both myelinated and unmyelinated nerve fibers are lost. Because the clinical features of diabetic neuropathy are similar to those of other neuropathies, the diagnosis of diabetic neuropathy should be made only after other possible etiologies are excluded **(Chap. 457)**.

Distal Symmetric Polyneuropathy (DSPN) DSPN, the most common form of diabetic neuropathy, most frequently presents with distal sensory loss and pain, but up to 50% of patients do not have symptoms of neuropathy. Symptoms may include a sensation of numbness, tingling, sharpness, or burning that begins in the feet and spreads proximally. Hyperesthesia, paresthesia, and dysesthesia also may occur. Pain typically involves the lower extremities, is usually present at rest, and worsens at night. Both an acute (lasting <12 months) and a chronic form of painful diabetic neuropathy may occur. The acute form is sometimes treatment-related, occurring in the context of improved glycemic control. As diabetic neuropathy progresses, the pain subsides and eventually disappears, but a sensory deficit persists, and motor defects may develop. Physical examination **(Chap. 415)** often reveals sensory loss (to 10-g monofilament and/or vibration), loss of ankle deep-tendon reflexes, abnormal position sense, and muscular atrophy or foot drop. Annual screening for DSPN should begin 5 years after diagnosis of type 1 DM and at the time of diagnosis of type 2 DM and is aimed at detecting loss of protective sensation (LOPS). LOPS and DSPN are major risk factors for foot ulceration and falls due to small and large nerve fiber dysfunction and predispose to lower extremity amputation.

Autonomic Neuropathy Individuals with long-standing type 1 or 2 DM may develop signs of autonomic dysfunction involving the parasympathetic (cholinergic) and sympathetic (adrenergic) systems. DM-related autonomic neuropathy can affect multiple organ systems, including the cardiovascular, gastrointestinal (GI), genitourinary, sudomotor, and metabolic systems. Cardiovascular autonomic neuropathy, reflected by decreased heart rate variability, resting tachycardia, and orthostatic hypotension, is associated with an increase in ASCVD. Orthostatic hypotension, a late and unusual complication of diabetes, is sometimes seen in patients with associated DSPN and severe parasympathetic dysfunction. Reports of sudden death in DM have also been attributed to autonomic neuropathy affecting the cardiovascular system and predisposing to severe hypoglycemia, both of which may prolong the QTc interval. Autonomic neuropathy may reduce counterregulatory hormone release (especially epinephrine), and contribute to an inability to sense hypoglycemia appropriately (hypoglycemia unawareness) **(Chap. 418)** that increases the risk of severe hypoglycemia. Gastroparesis and bladder-emptying abnormalities are often caused by the autonomic neuropathy seen in DM (discussed below). Hyperhidrosis of the upper extremities and anhidrosis of the lower extremities result from sympathetic nervous system dysfunction. Anhidrosis of the feet can promote dry skin with cracking, which increases the risk of foot ulceration.

Mononeuropathy and/or Radiculopathy/Polyradiculopathy Mononeuropathy (dysfunction of isolated cranial or peripheral nerves) is less common than polyneuropathy in DM and presents with pain and motor weakness in the distribution of a single nerve. Mononeuropathies can occur at entrapment sites such as carpal tunnel or be noncompressive. Involvement of the third cranial nerve is most common and is heralded by diplopia. Physical examination reveals ptosis and ophthalmoplegia with normal pupillary constriction to light. Sometimes other cranial nerves, such as IV, VI, or VII (Bell's palsy), are affected. Peripheral mononeuropathies or simultaneous involvement of more than one nerve (mononeuropathy multiplex) may also occur. Diabetic radiculopathy or polyradiculopathy is a syndrome characterized by severe disabling pain in the distribution of one or more nerve roots. It may be accompanied by motor weakness. Intercostal or truncal radiculopathy causes pain over the thorax or abdomen. Involvement of the lumbar plexus or femoral nerve may cause severe pain in the thigh or hip and may be associated with muscle weakness in the hip flexors or extensors (diabetic amyotrophy). Fortunately, diabetic polyradiculopathies are usually self-limited and resolve over 6–12 months.

TREATMENT

Diabetic Neuropathy

Prevention of diabetic neuropathy is critical through improved glycemic control. Treatment of diabetic neuropathy is less than satisfactory. Lifestyle modifications (exercise, diet) have some efficacy in DSPN in type 2 DM and hypertension and hypertriglyceridemia should be treated. Efforts to improve glycemic control in long-standing diabetes may be limited by hypoglycemia unawareness. Patients should avoid neurotoxins (including alcohol) and smoking and consider supplementation with vitamins for possible deficiencies (B_{12}, folate; **Chap. 344**). Metformin may reduce intestinal absorption of vitamin B_{12} in type 2 DM, and pernicious anemia is more common in type 1 DM where it is associated with anti–parietal cell autoantibodies and may require sublingual or parenteral B_{12} replacement. Patients should be educated that loss of sensation in the foot increases the risk for ulceration and its sequelae and that prevention of such problems is paramount. Patients with symptoms or signs of neuropathy or LOPS should check their feet daily and take precautions (footwear) aimed at preventing calluses or ulcerations. If foot deformities are present, a podiatrist should be involved.

Chronic, painful diabetic neuropathy is difficult to treat with only symptomatic treatment being available; evidence of the effectiveness of improved glycemic control in painful diabetic neuropathy is lacking. Sleep and mood disorders frequently accompany DPSN and should be treated. Symptomatic treatment of the pain using gabapentinoids (pregabalin, gabapentin), serotonin-norepinephrine reuptake inhibitors (duloxetine, venlafaxine, and desvenlafaxine), sodium channel blockers, tricyclic antidepressants, and a capsaicin patch have some efficacy for pain related to DPSN. Tapentadol, a centrally acting opioid, is also approved by the U.S. Food and Drug Administration (FDA) but has only modest efficacy and poses addiction risk, making it and other opioids less desirable and not first-line therapy. No direct comparisons of agents are available, and it is reasonable to switch agents if there is no response or if side effects develop. Referral to a pain management center may be necessary.

Therapy of orthostatic hypotension secondary to autonomic neuropathy is also difficult. Nonpharmacologic maneuvers (adequate salt intake, avoidance of dehydration and diuretics, lower extremity support hose, and physical activity) may offer some benefit. A variety of agents have limited success (midodrine and droxidopa are approved by the FDA for orthostatic hypotension of any etiology). Patients with resting tachycardia may be considered for beta blocker therapy with caution exercised if there is hypoglycemia unawareness. Patients with type 1 DM and orthostatic hypotension should be evaluated for primary adrenal insufficiency (Addison's disease) that may be associated with an autoimmune polyendocrine syndrome **(Chap. 401)**.

GASTROINTESTINAL/GENITOURINARY DYSFUNCTION

Long-standing type 1 and 2 DM may affect the motility and function of the GI and genitourinary systems. The most prominent GI symptoms are delayed gastric emptying (gastroparesis) and altered small- and large-bowel motility (constipation or diarrhea). Gastroparesis may present with symptoms of anorexia, nausea, vomiting, early satiety, and abdominal bloating. Microvascular complications (retinopathy and neuropathy) are usually present. Nuclear medicine scintigraphy after ingestion of a radiolabeled meal or digestible solids may document delayed gastric emptying but may not correlate well with the patient's symptoms. Although parasympathetic dysfunction secondary to chronic hyperglycemia is important in the development of gastroparesis, hyperglycemia itself also impairs gastric emptying. Nocturnal diarrhea, alternating with constipation, may be a feature of DM-related GI autonomic neuropathy. In type 1 DM, these symptoms should also prompt evaluation for celiac disease that is associated with anti-tissue transglutaminase autoantibodies because of its increased frequency.

Diabetic autonomic neuropathy may lead to genitourinary dysfunction, including cystopathy and female sexual dysfunction (reduced sexual desire, dyspareunia, reduced vaginal lubrication). Symptoms of diabetic cystopathy begin with an inability to sense a full bladder and a failure to void completely. As bladder contractility worsens, bladder capacity and the postvoid residual increase, leading to symptoms of urinary hesitancy, decreased voiding frequency, incontinence, and recurrent urinary tract infections.

Erectile dysfunction and retrograde ejaculation are very common in DM and may be one of the earliest signs of diabetic neuropathy **(Chap. 409)**. Erectile dysfunction, which increases in frequency with the age of the patient and the duration of diabetes, may occur in the absence of other signs of diabetic autonomic neuropathy.

TREATMENT

Gastrointestinal/Genitourinary Dysfunction

Current treatments for these complications of DM are inadequate and nonspecific. Improved glycemic control should be a goal but has not clearly shown benefit. Smaller, more frequent meals that are easier to digest (liquid) and low in fat and fiber may minimize symptoms of gastroparesis. Medications that slow gastric emptying (opioids, GLP-1 receptor agonists) should be avoided. Metoclopramide may be used with severe symptoms but is restricted to short-term treatment in both the United States and Europe. Symptoms of gastroesophageal reflux disease may require acid-blocking therapy with a histamine-2 receptor antagonist or proton pump inhibitor. Gastric electrical stimulatory devices are available. Diabetic diarrhea in the absence of bacterial overgrowth is treated symptomatically **(Chap. 336)**.

Diabetic cystopathy should be treated with scheduled voiding or self-catheterization. Drugs that inhibit type 5 phosphodiesterase are effective for erectile dysfunction, but their efficacy in individuals with DM is slightly lower than in the nondiabetic population **(Chap. 409)**.

CARDIOVASCULAR MORBIDITY AND MORTALITY

ASCVD, including PAD, CHD, heart failure, and cerebrovascular disease, occurs more frequently in individuals with type 1 or type 2 DM and is the major cause of mortality for individuals with diabetes. In addition, the prognosis for individuals with diabetes who have CHD is worse than for nondiabetics. CHD is more likely to involve multiple vessels in individuals with DM. The American Heart Association considers DM a controllable risk factor for cardiovascular disease; in some studies, type 2 DM patients without a prior MI have a similar risk for coronary artery-related events as nondiabetic individuals who have had a prior MI. Fortunately, the outcomes related to ASCVD have improved over the last decade for those without diabetes and individuals with diabetes as a result of modification of multiple risk factors.

Heart failure, which has not been recognized until recently as a diabetes-related complication, is twice as common in individuals with diabetes (type 1 or type 2). Heart failure is related to diabetes duration and hypertension and can present as heart failure with preserved ejection fraction (HFpEF), heart failure with mildly reduced ejection fraction (HFmEF), or heart failure with reduced ejection fraction (HFrEF) (see **Chap 276**). Some individuals with DM have reduced left ventricular function without CHD or hypertension, and this is sometimes termed "diabetic cardiomyopathy." The pathogenesis of this and heart failure associated with DM is not clear.

The prevention and management of ASCVD and heart failure in individuals with DM should focus on risk factors, including duration of diabetes, hypertension, dyslipidemia, CKD, albuminuria, obesity, and smoking. Many of these are modifiable and should prompt action by the patient and the provider. The foundation of prevention and management is the concurrent, integrated focus on four targets: glycemia, blood pressure, lipids, and the incorporation of therapies with cardiovascular and kidney outcome benefits. While these results are from observations and studies in type 2 DM, these strategies are also likely relevant to type 1 DM.

Cardiovascular risk assessment in type 2 DM should encompass a nuanced and individualized approach. For example, cardiovascular risk is lower and not equivalent in a younger individual with a brief duration of type 2 DM compared to an older individual with longstanding type 2 DM. Because of the high prevalence of underlying ASCVD in individuals with diabetes (especially in type 2 DM), evidence of ASCVD (e.g., cardiac stress test) should be sought in an individual with diabetes who has symptoms, even if atypical, suggestive of cardiac ischemia or peripheral or carotid arterial disease. However, the screening of asymptomatic individuals with diabetes for CHD is not recommended or cost-effective. The absence of chest pain ("silent ischemia") is common in individuals with diabetes, and a thorough cardiac evaluation should be considered prior to major surgical procedures.

TREATMENT

Cardiovascular Disease

Treatment of coronary disease in individuals with DM is similar to treatment in individuals without DM **(Chap. 284)**. Revascularization procedures for CHD, including percutaneous coronary interventions (PCIs) and coronary artery bypass grafting (CABG), may be less efficacious in individuals with DM. Initial success rates of PCI in individuals with DM are similar to those in the nondiabetic population, but higher rates of restenosis and lower long-term patency and survival rates have been reported. CABG plus optimal medical management likely has better outcomes than PCI for individuals with diabetes. Very strict glucose control has limited benefit on cardiovascular outcomes in individuals with established cardiovascular disease, indicating the importance of other factors such as insulin resistance, dyslipidemia, and inflammation.

In individuals with type 2 DM and ASCVD, the comprehensive effort to reduce cardiovascular risk (e.g., lifestyle management, blood pressure control, lipid management) should include an SGLT-2 inhibitor or a GLP-1 receptor agonist. If diabetic CKD or heart failure is present or likely, an SGLT-2 inhibitor is preferred. In individuals with type 2 DM and ASCVD or other ASCVD risk factors, a GLP-1 receptor agonist will reduce cardiovascular events. The combination of an SGLT-2 inhibitor and a GLP-1 receptor agonist likely provides additive risk reduction. In individuals with type 2 DM and CKD with albuminuria treated with maximum ACE inhibitor or ARB, addition of either a SGLT2 inhibitor or finerenone reduces CKD progression and improves cardiovascular outcomes. Combining a SGLT2 inhibitor with finerenone reduces the risk of hyperkalemia. Care of individuals with type 2 DM and heart failure or cardiovascular disease should involve a cardiovascular specialist and include treatment with an ACE inhibitor or ARB and a beta blocker. If an individual is already taking metformin and has an eGFR >30 mL/min per 1.73 m^2, reduced-dose metformin can be continued. Because of the elevated risk of euglycemic diabetic ketoacidosis with SGLT-2 inhibitors, patients treated with an SGLT-2 inhibitor should be counseled about the risk and symptoms of diabetic ketoacidosis and educated about the importance of measuring ketones if the clinical scenario suggests this possibility.

Antiplatelet therapy with aspirin (75–162 mg/d) as secondary prevention reduces cardiovascular events in individuals with DM who have ASCVD. Clopidogrel should be used in those with aspirin allergy or intolerance. The ADA recommends considering the use of aspirin for primary prevention of coronary events in individuals with diabetes with an increased cardiovascular risk (>50 years old with at least one risk factor such as hypertension, dyslipidemia, smoking, family history, or albuminuria). Aspirin is not recommended for primary prevention in those with a low cardiovascular risk (<50 years old with no risk factors).

Cardiovascular Risk Factors

DYSLIPIDEMIA Individuals with DM may have several forms of dyslipidemia **(Chap. 419)**. Because of the additive cardiovascular risk of hyperglycemia and hyperlipidemia, lipid abnormalities should be assessed and treated as part of comprehensive diabetes care **(Chap. 416)**. The most common pattern of dyslipidemia is hypertriglyceridemia and reduced high-density lipoprotein (HDL) cholesterol levels. DM itself does not increase levels of low-density lipoprotein (LDL), but the small dense LDL particles found in type 2 DM are more atherogenic because they are more easily glycated and susceptible to oxidation.

Almost all treatment studies of diabetic dyslipidemia have been performed in individuals with type 2 DM because of the greater frequency of dyslipidemia in this form of diabetes. Interventional studies have shown that the beneficial effects of LDL reduction with statins are similar in the diabetic and nondiabetic populations. No prospective studies have addressed similar questions in individuals with type 1 DM. Because the frequency of ASCVD is low in children and young adults with diabetes, assessment of cardiovascular risk should be incorporated into the guidelines discussed below. Statin usage is associated with a mild increase in the risk of developing type 2 DM. However, when appropriately indicated the cardiovascular benefits of statin use outweigh the mildly increased risk of diabetes.

Based on the guidelines provided by the ADA, all individuals with diabetes should be advised about lifestyle modification, including diet, weight loss, and increased physical activity **(Chap. 416)**. If individuals with diabetes have elevated triglyceride levels (>1.7 mmol/L [150 mg/dL]) or low HDL cholesterol (<1 mmol/L [40 mg/dL] in men and <1.3 mmol/L [50 mg/dL] in women), lifestyle modification and improved glycemic control should be further emphasized. If triglycerides are >5.7 mmol/L (500 mg/dL) on a statin, icosapent or fenofibrate can be considered to reduce ASCVD risk. The addition of fenofibrate may require reduction in statin dose to minimize the risk of myopathy.

In terms of pharmacologic therapy directed at LDL, the ADA recommends the following in addition to lifestyle: (1) all patients with diabetes and ASCVD should receive high-intensity statin therapy (atorvastatin 40–80 mg or rosuvastatin 20–40 mg) with a goal of >50% reduction in LDL and a cholesterol goal of <55 mg/dL. Adding ezetimibe or a PCSK9 inhibitor is advised if these goals are not met; (2) in patients aged 40–75 years without ASCVD, moderate-intensity statin therapy (other statins or lower dose of atorvastatin or rosuvastatin) should be used; (3) in patients aged 40–75 years with ASCVD risk factors, use high-intensity statin therapy with a goal to reduce LDL cholesterol by 50% and an LDL target of <70 mg/dL; (4) in patients aged 40–75 years with ASCVD risk factors on maximum statin therapy and a LDL ≥70 mg/dL, consider adding ezetimibe or PCSK9 inhibitor therapy; (5) in patients aged >75 years on a statin, continue statin or, if not on a statin, consider starting moderate-intensity statin therapy after discussion with the patient; and (6) in patients aged 20–39 years with additional risk factors, consider moderate-intensity statin therapy. If a patient with ASCVD cannot tolerate a statin, consider PCSK9 inhibitor therapy (monoclonal antibody or inclisiran, a small interfering RNA), or bempedoic acid. Statin therapy, when combined with fibrate or niacin, for reduction of LDL does not provide additional benefit.

HYPERTENSION Hypertension management is discussed above in the "Renal Complications of Diabetes Mellitus" section.

■ LOWER EXTREMITY COMPLICATIONS

DM is the leading cause of nontraumatic lower extremity amputation in the United States. Foot ulcers and infections are also a major source of morbidity in individuals with DM. The reasons for the increased incidence of these disorders in DM involve the interaction of several pathogenic factors: neuropathy, abnormal foot biomechanics, PAD, and poor wound healing. The peripheral sensory neuropathy interferes with normal protective mechanisms and allows the patient to sustain major or repeated minor trauma to the foot, often without knowledge of the injury. Disordered proprioception causes abnormal weight bearing while walking and subsequent formation of callus or ulceration. Motor and sensory neuropathy leads to abnormal foot muscle mechanics and to structural changes in the foot (hammer toe, claw toe deformity, prominent metatarsal heads, Charcot joint). Autonomic neuropathy results in anhidrosis and altered superficial blood flow in the foot, which promote drying of the skin and fissure formation. PAD and poor wound healing impede resolution of minor breaks in the skin, allowing them to enlarge and to become infected.

Some individuals with DM will develop a foot ulcer (great toe or metatarsophalangeal areas are most common), and a significant subset may ultimately undergo amputation (14–24% risk with that ulcer or subsequent ulceration). Risk factors for foot ulcers or amputation include male sex, diabetes for >10 years, peripheral neuropathy, abnormal structure of foot (bony abnormalities, callus, thickened nails), PAD, smoking, history of previous ulcer or amputation, visual impairment, poor glycemic control, and diabetic nephropathy, especially dialysis. Large calluses are often precursors to or overlie ulcerations.

TREATMENT

Lower Extremity Complications

The optimal therapy for foot ulcers and amputations is prevention through identification of high-risk patients, education of the patient, and institution of measures to prevent ulceration. High-risk patients should be identified during the routine, annual foot examination performed on all patients with DM (see "Ongoing Aspects of Comprehensive Diabetes Care" in **Chap. 416**). If the monofilament test or one of the other tests is abnormal, the patient is diagnosed with LOPS **(Chap. 415)**. Providers should consider screening for asymptomatic PAD in individuals >50 years of age who have diabetes and other risk factors using ankle-brachial index testing **(Chap. 292)**. Patient education should emphasize (1) careful selection of footwear, (2) daily inspection of the feet to detect early signs of poor-fitting footwear or minor trauma, (3) daily foot hygiene to keep the skin clean and moist, (4) avoidance of self-treatment of foot abnormalities and high-risk behavior (e.g., walking barefoot), and (5) prompt consultation with a health care provider if an abnormality arises. Involvement of a podiatrist is recommended for high-risk individuals (history of foot ulcers or amputation, those on dialysis, those with PAD, and those with foot deformities). Calluses and nail deformities should be treated by a podiatrist. Interventions directed at risk factor modification include orthotic shoes and devices, callus management, nail care, and prophylactic measures to reduce increased skin pressure from abnormal bony architecture. Attention to other risk factors for vascular disease (smoking, dyslipidemia, hypertension) and improved glycemic control are also important, especially LDL management as described above.

Despite preventive measures, foot ulceration and infection are common and represent a serious problem. Due to the multifactorial pathogenesis of lower extremity ulcers, management of these lesions is multidisciplinary and often demands expertise in orthopedics, vascular surgery, endocrinology, podiatry, and infectious diseases. The plantar surface of the foot is the most common site of ulceration. Ulcers may be primarily neuropathic (no accompanying infection) or may have surrounding cellulitis or osteomyelitis. Cellulitis without ulceration should be treated with antibiotics that provide appropriate empiric coverage (see below).

An infected ulcer is a clinical diagnosis, because superficial culture of any ulceration will likely find multiple bacterial species of unknown significance. The infection surrounding the foot ulcer may be due to multiple organisms, with aerobic gram-positive cocci (staphylococci including methicillin-resistant *Staphylococcus aureus* [MRSA], group A and B streptococci) being most common and with aerobic gram-negative bacilli and/or obligate anaerobes as co-pathogens.

Gas gangrene may develop in the absence of clostridial infection. Cultures should be obtained from the debrided ulcer base or from purulent drainage or aspiration of the wound. Wound depth should be determined by inspection and probing with a blunt-tipped sterile instrument. A wound that probes to the bone is highly likely to have underlying osteomyelitis. Plain radiographs of the foot should be performed to assess the possibility of osteomyelitis in chronic ulcers that have not responded to therapy. Magnetic resonance imaging (MRI) is the most specific modality, with nuclear medicine scans (PET, CT/SPECT) and labeled white cell studies as an alternative. Surgical debridement is often necessary.

Osteomyelitis is best treated by a combination of prolonged antibiotics and debridement of infected bone when possible. The possible contribution of vascular insufficiency should be considered in all patients. Peripheral arterial bypass procedures are often effective in promoting wound healing and in decreasing the need for amputation of the ischemic limb **(Chap. 292)**.

Interventions with demonstrated efficacy in diabetic foot ulcers or wounds include the following: (1) off-loading (complete avoidance of weight bearing on the ulcer, which removes the mechanical trauma that retards wound healing), (2) surgical debridement of nonviable tissue, (3) physiologic, topical wound dressings, (4) revascularization, and (5) treatment of infections with appropriate use of antibiotics. Amputation should be limited initially. If a wound fails to show significant improvement after 4 weeks of wound management with these five recommendations, one should consider advanced wound therapy that may include topical growth factors, acellular matrix tissues, bioengineered cellular therapies, negative-pressure wound therapy, electrical stimulation, pulsed radiofrequency, extracorporeal shockwave, hyperbaric oxygen therapy, and topical oxygen therapy. These modalities require interdisciplinary expertise and must be individualized to the patient and clinical setting. Antiseptic agents should be avoided. Topical antibiotics are of limited value. Referral for physical therapy, orthotic evaluation, and rehabilitation should occur once the infection is controlled.

Mild or non-limb-threatening infections can be treated with oral antibiotics directed predominantly at methicillin-susceptible staphylococci and streptococci (e.g., dicloxacillin, early-generation cephalosporins, amoxicillin-clavulanate). However, in patients with a prior history of MRSA or in locations with a high prevalence of MRSA, treatment with trimethoprim-sulfamethoxazole, doxycycline, linezolid, or clindamycin is preferred, depending on local antibiogram data. Surgical debridement of necrotic tissue, local wound care, and avoidance of weight bearing over the ulcer are crucial. Optimization of glycemic control should be a goal. More severe infections may require IV antibiotics as well as offloading and local wound care. IV antibiotics should provide broad-spectrum coverage directed toward *S. aureus*, including MRSA, streptococci, gram-negative aerobes, and anaerobic bacteria. Initial empiric antimicrobial regimens may include vancomycin plus a β-lactam/β-lactamase inhibitor or carbapenem, or vancomycin plus a quinolone with metronidazole. In some cases, daptomycin, ceftaroline, or linezolid may be substituted for vancomycin in consultation with an infectious diseases expert. If the infection surrounding the ulcer is not improving with antibiotics, reassessment of antibiotic coverage and reconsideration of the need for surgical debridement or revascularization are indicated. With clinical improvement, oral antibiotics and local wound care can be continued on an outpatient basis with close follow-up.

■ INFECTIONS

Individuals with DM have a greater frequency and severity of infection. The reasons for this include incompletely defined abnormalities in cell-mediated immunity and phagocyte function associated with hyperglycemia, as well as diminished vascularization. Hyperglycemia aids the colonization and growth of a variety of organisms (*Candida* and other fungal species). Many common infections are more frequent and severe in the diabetic population, whereas several rare infections are seen almost exclusively in the diabetic population. Examples of this latter category include rhinocerebral mucormycosis, emphysematous infections of the gallbladder and urinary tract, and "malignant" or invasive otitis externa. Invasive otitis externa is usually secondary to *Pseudomonas aeruginosa* infection in the soft tissue surrounding the external auditory canal, typically begins with pain and discharge, and may rapidly progress to osteomyelitis and meningitis. These infections should be considered, in particular, in patients presenting with severe hyperglycemia.

Pneumonia, urinary tract infections, and skin and soft tissue infections are all more common in the diabetic population. In general, the organisms that cause pulmonary infections are similar to those found in the nondiabetic population; however, gram-negative organisms, *S. aureus*, and *Mycobacterium tuberculosis* are more frequent pathogens. Adults with DM should receive vaccination against pneumococcus, respiratory syncytial virus, annually against influenza, and the coronavirus SARS-CoV-2, which causes increased morbidity and mortality in obese individuals and patients with DM **(Chap. 204)**. In addition to early antibiotic therapy for presumed bacterial infections, patients with DM should be considered for early intervention with antiviral agents (e.g., against influenza in flu, varicella-zoster virus in shingles) or SARS-CoV-2 in COVID. Urinary tract infections (either lower tract or pyelonephritis) are the result of common bacterial agents such as *Escherichia coli*, although several yeast species (e.g., *Candida albicans* and *C. glabrata*) are sometimes observed. Complications of urinary tract infections include emphysematous pyelonephritis and emphysematous cystitis. Bacteriuria occurs frequently in individuals with diabetic cystopathy and does not require antibiotic therapy except in specific circumstances such as pregnancy or a planned urologic procedure. Susceptibility to furunculosis, superficial candidal infections, and vulvovaginitis are increased. Poor glycemic control is a common denominator in individuals with these infections. Individuals with diabetes have an increased rate of colonization of *S. aureus* in the skinfolds and nares. Individuals with diabetes also have a greater risk of postoperative wound infections that may be mitigated by perioperative protocols for insulin administration to maintain glycemic control.

■ DERMATOLOGIC MANIFESTATIONS

The most common skin manifestations of DM are xerosis and pruritus and are usually relieved by skin moisturizers. Protracted wound healing and skin ulcerations are also frequent complications. Diabetic dermopathy, sometimes termed *pigmented pretibial papules*, or "diabetic skin spots," begins as an erythematous macule or papule that evolves into an area of circular hyperpigmentation. These lesions result from minor mechanical trauma in the pretibial region and are more common in elderly men with DM. Bullous diseases, such as bullosa diabeticorum (shallow ulcerations or erosions in the pretibial region), are also seen. *Necrobiosis lipoidica diabeticorum* is an uncommon disorder, accompanying diabetes in predominantly young women. This usually begins in the pretibial region as an erythematous plaque or papules that gradually enlarge, darken, and develop irregular margins, with atrophic centers and central ulceration. They are often painful. Vitiligo and alopecia areata occur at increased frequency in individuals with type 1 DM. *Acanthosis nigricans* (hyperpigmented velvety plaques seen on the neck, axilla, or extensor surfaces) is sometimes a feature of severe insulin resistance and accompanying diabetes. Generalized or localized *granuloma annulare* (erythematous plaques on the extremities or trunk), *lichen planus* (violaceous papules on the cutaneous surface with or without erosions in the mouth and genitalia), and *scleredema* (areas of skin thickening on the back or neck at the site of previous superficial infections) are more common in the diabetic population. *Lipoatrophy* and *lipohypertrophy* can occur at insulin injection sites but are now unusual with the use of human insulin and avoided by rotating injection sites.

■ FURTHER READING

Abel ED et al: Diabetes mellitus-progress and opportunities in the evolving epidemic. Cell 187:3789, 2024.

Adler AI et al: Post-trial monitoring of a randomised controlled trial of intensive glycaemic control in type 2 diabetes extended from 10 years to 24 years (UKPDS 91). Lancet 404:145, 2024.

American Diabetes Association: Cardiovascular disease and risk management: Standards of Medical Care in Diabetes-2024. Diabetes Care 47:S179, 2024.

American Diabetes Association: Chronic kidney disease and risk management: Standards of Care in Diabetes—2024. Diabetes Care 47:S219, 2024.

American Diabetes Association: Retinopathy, neuropathy, and foot care: Standards of Care in Diabetes—2024. Diabetes Care 2024; 47:S231, 2024.

Drucker DJ: Diabetes, obesity, metabolism, and SARS-CoV-2 infection: The end of the beginning. Cell Metab 33:479, 2021.

Mann JFE et al: Effects of semaglutide with and without concomitant SGLT2 inhibitor use in participants with type 2 diabetes and chronic kidney disease in the FLOW trial. Nat Med 30:2849, 2024.

Naaman S, Bakris G: Diabetic nephropathy: Update on pillars of therapy slowing progression. Diabetes Care 46:1574, 2023.

Nathan DM: Realising the long-term promise of insulin therapy: The DCCT/EDIC study. Diabetologia 64:1049, 2021.

Perkovic V et al: Effects of semaglutide on chronic kidney disease in patients with type 2 diabetes. N Engl J Med. 391:109, 2024.

Pop-Busui R et al: Heart failure: An underappreciated complication of diabetes. A consensus report of the American Diabetes Association. Diabetes Care 45:1670, 2022.

Senneville É et al: IWGDF/IDSA guidelines on the diagnosis and treatment of diabetes-related foot infections. Diabetes Metab Res Rev 40:e3687, 2024.

van Netten JJ et al: The International Working Group on the Diabetic Foot: Stories and numbers behind three decades of evidence-based guidelines for the management of diabetes-related foot disease. Diabetes Ther 15:19, 2024.

418

Hypoglycemia

Stephen N. Davis

Hypoglycemia is most commonly caused by insulin or insulin-producing drugs used to treat diabetes mellitus or by exposure to other drugs, including alcohol. However, a number of other disorders, including critical organ failure, sepsis and inanition, hormone deficiencies, non-β-cell tumors, insulinoma, inborn errors of metabolism, and prior gastric surgery, can cause hypoglycemia **(Table 418-1)**. Hypoglycemia may be documented by *Whipple's triad:* (1) symptoms consistent with hypoglycemia, (2) a low plasma glucose concentration measured with a precise method, and (3) relief of symptoms after the plasma glucose level is raised. The lower limit of the fasting plasma glucose concentration is normally ~70 mg/dL (~3.9 mmol/L), but lower venous glucose levels occur normally, late after a meal, during pregnancy, and during prolonged fasting (>24 h). Severe hypoglycemia can cause serious morbidity and increase the risk for serious cardiovascular events and mortality during and after the initial hypoglycemic episode. It should be considered in any patient with episodes of confusion, an altered level of consciousness, or a seizure.

■ SYSTEMIC GLUCOSE BALANCE AND GLUCOSE COUNTERREGULATION

Glucose is an obligate metabolic fuel for the brain under physiologic conditions. The brain cannot synthesize glucose or store more than a few minutes' supply as glycogen and therefore requires a continuous supply of glucose from the arterial circulation. As the arterial plasma glucose concentration falls below the physiologic range, blood-to-brain glucose transport becomes insufficient to support brain energy metabolism and function. However, multiple integrated glucose counterregulatory mechanisms normally prevent or rapidly correct hypoglycemia.

Plasma glucose concentrations are normally maintained within a relatively narrow range—roughly 70–110 mg/dL (3.9–6.1 mmol/L) in the fasting state, with transient higher excursions after a meal—despite wide variations in exogenous glucose delivery from meals and in endogenous glucose utilization by, for example, exercising muscle. Between meals and during fasting, plasma glucose levels are maintained

TABLE 418-1 Causes of Hypoglycemia Across the Life Span

Ill or Medicated Individual

1. Drugs
   - Insulin or insulin secretagogues
   - Alcohol
   - Others
2. Critical illness
   - Hepatic, renal, or cardiac failure
   - Sepsis
   - Inanition
3. Hormone deficiency
   - Cortisol
   - Growth hormone
   - Glucagon and epinephrine (in insulin-deficient diabetes)
4. Non–islet cell tumor (e.g., mesenchymal tumors)

Seemingly Well Individual

5. Endogenous hyperinsulinism
   - Insulinoma
   - Functional β-cell disorders (nesidioblastosis)
     - Noninsulinoma pancreatogenous hypoglycemia
     - Post–gastric bypass hypoglycemia
   - Insulin autoimmune hypoglycemia
     - Antibody to insulin
     - Antibody to insulin receptor
   - GLP-1 receptor agonists in combination with insulin and/or insulin secretagogues
   - Insulin secretagogues
   - Other
6. Disorders of gluconeogenesis and fatty acid oxidation
7. Exercise
8. Accidental, surreptitious, or malicious hypoglycemia
9. Prolonged fasting
10. Pregnancy

Source: Reproduced with permission from PE Cryer et al: Evaluation and management of adult hypoglycemic disorders: An Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 94:709, 2009.

by endogenous glucose production, hepatic glycogenolysis, and hepatic (and renal) gluconeogenesis (Fig. 418-1). Although hepatic glycogen stores are usually sufficient to maintain plasma glucose levels for ~8 h, this period can be shorter if glucose demand is increased by exercise or if glycogen stores are depleted by illness or starvation.

Gluconeogenesis normally requires low insulin levels and the presence of anti-insulin (counterregulatory) hormones together with a coordinated supply of precursors from muscle and adipose tissue to the liver and kidneys. Muscle provides lactate, pyruvate, alanine, glutamine, and other amino acids. Triglycerides in adipose tissue are broken down into fatty acids and glycerol, which is a gluconeogenic precursor. Fatty acids provide an alternative oxidative fuel to tissues other than the brain (which requires glucose).

Systemic glucose balance, maintenance of the normal plasma glucose concentration, is accomplished by a network of hormones, neural signals, and substrate effects that regulate endogenous glucose production and glucose utilization by tissues other than the brain (Chap. 415). Among the regulatory factors, insulin plays a dominant role (Table 418-2; Fig. 418-1). As plasma glucose levels decline within the physiologic range, pancreatic β-cell insulin secretion decreases, thereby increasing hepatic glycogenolysis and hepatic (and renal) gluconeogenesis. Low insulin levels also reduce glucose utilization in peripheral tissues, inducing lipolysis and proteolysis and consequently releasing gluconeogenic precursors. Thus, a decrease in insulin secretion is the first defense against hypoglycemia.

As plasma glucose levels decline just below the physiologic range, glucose counterregulatory (plasma glucose–raising) hormones are released (Table 418-2; Fig. 418-1). Among these, pancreatic α-cell glucagon and adrenomedullary epinephrine play a primary role. Glucagon stimulates hepatic glycogenolysis and gluconeogenesis. Adrenomedullary epinephrine also stimulates hepatic glycogenolysis and gluconeogenesis (and renal gluconeogenesis) but limits peripheral uptake of glucose and stimulates lipolysis with production of glycerol and fatty acids. Epinephrine becomes critical when glucagon is deficient. When hypoglycemia is prolonged beyond ~4 h, cortisol and growth hormone also support glucose production and restrict glucose utilization to a limited amount (both mechanisms are reduced by ~80% compared to epinephrine). Thus, cortisol and growth hormone play no role in defense against acute hypoglycemia.

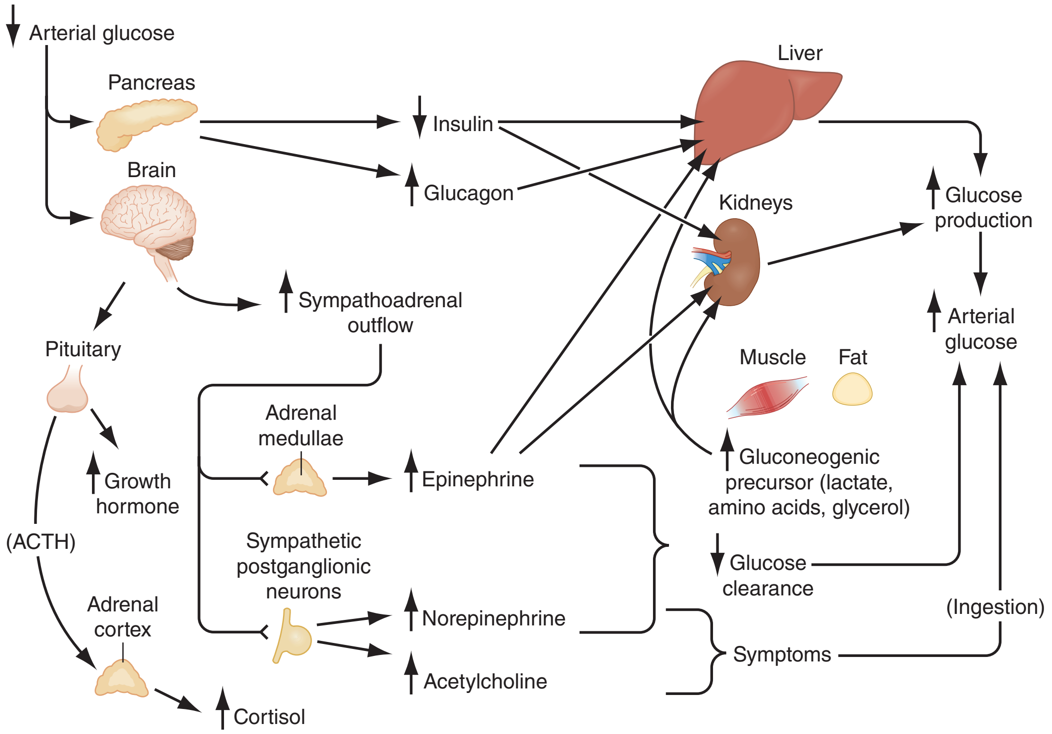

FIGURE 418-1 Physiology of glucose counterregulation: Mechanisms that normally prevent or rapidly correct hypoglycemia. In insulin-deficient diabetes, the key counterregulatory responses—suppression of insulin and increases in glucagon—are lost, and stimulation of sympathoadrenal outflow is attenuated. ACTH, adrenocorticotropic hormone.

TABLE 418-2 Physiologic Responses to Decreasing Plasma Glucose Concentrations

RESPONSE	GLYCEMIC THRESHOLD, mmoL/L (mg/dL)	PHYSIOLOGIC ↓ EFFECTS	ROLE IN PREVENTION OR CORRECTION OF HYPOGLYCEMIA (GLUCOSE COUNTERREGULATION)
↓ Insulin	4.4–4.7 (80–85)	$\uparrow R_a$ ($\downarrow R_d$), increased lipolysis; ↑ FFA ↑ Glycerol	Primary glucose regulatory factor/first defense against hypoglycemia
↑ Glucagon	3.6–3.9 (65–70)	$\uparrow R_a$	Primary glucose counterregulatory factor/second defense against hypoglycemia
↑ Epinephrine	3.6–3.9 (65–70)	$\uparrow R_a$, $\downarrow R_c$, increased lipolysis; ↑ FFA and glycerol	Third defense against hypoglycemia; critical when glucagon is deficient
↑ Cortisol and growth hormone	3.6–3.9 (65–70)	$\uparrow R_a$, $\downarrow R_c$	Involved in defense against prolonged hypoglycemia; not critical
Symptoms	2.8–3.1 (50–55)	Recognition of hypoglycemia	Prompt behavioral defense against hypoglycemia (food ingestion)
↓ Cognition	<2.8 (<50)	—	Compromises behavioral defense against hypoglycemia

Note: R_a, rate of glucose appearance, glucose production by the liver and kidneys; R_c, rate of glucose clearance, glucose utilization relative to the ambient plasma glucose by insulin-sensitive tissues; R_d, rate of glucose disappearance, glucose utilization by insulin-sensitive tissues such as skeletal muscle. R_d by the brain is not altered by insulin, glucagon, epinephrine, cortisol, or growth hormone.

Abbreviation: FFA, free fatty acids.

Source: Reproduced with permission from PE Cryer, in S Melmed et al: *Williams Textbook of Endocrinology*, 12th ed. New York, NY: Elsevier; 2012.

As plasma glucose levels fall further, symptoms prompt behavioral defense against hypoglycemia, including the ingestion of food (Table 418-2; Fig. 418-1). The normal glycemic thresholds for these responses to decreasing plasma glucose concentrations are shown in Table 418-2. However, these thresholds are dynamic. They shift to higher-than-normal glucose levels in people with poorly controlled diabetes, who can experience symptoms of hypoglycemia when their glucose levels decline toward the normal range. On the other hand, thresholds shift to lower-than-normal glucose levels in people with recurrent hypoglycemia; i.e., patients with intensively treated diabetes or an insulinoma have symptoms at glucose levels lower than those that cause symptoms in healthy individuals.

Clinical Manifestations Neuroglycopenic manifestations of hypoglycemia are the direct result of central nervous system glucose deprivation. These features include behavioral changes, confusion, fatigue, seizure, loss of consciousness, cardiac arrhythmias, and, if hypoglycemia is severe, death. Neurogenic (or autonomic) manifestations of hypoglycemia result from the perception of physiologic changes caused by the central nervous system–mediated sympathoadrenal discharge that is triggered by hypoglycemia. They include *adrenergic* symptoms (mediated largely by norepinephrine released from sympathetic postganglionic neurons but perhaps also by epinephrine released from the adrenal medullae), such as palpitations, tremor, and anxiety, as well as *cholinergic* symptoms (mediated by acetylcholine released from sympathetic postganglionic neurons), such as sweating, hunger, and paresthesias. Clearly, these are nonspecific symptoms. Their attribution to hypoglycemia requires that the corresponding plasma glucose concentration be low and that the symptoms resolve after the glucose level is raised (as delineated by Whipple's triad).

Common signs of hypoglycemia include diaphoresis and pallor. Heart rate and systolic blood pressure are typically increased but may not be raised in an individual who has experienced repeated, recent episodes of hypoglycemia. Neuroglycopenic manifestations are often observable. Transient focal neurologic deficits occur occasionally. Permanent neurologic deficits are rare.

Etiology and Pathophysiology Hypoglycemia activates proinflammatory, procoagulant, and proatherothrombotic responses in type 1 diabetes mellitus (T1DM), type 2 diabetes mellitus (T2DM), and nondiabetic individuals. These responses increase platelet aggregation, reduce fibrinolytic balance (increase plasminogen activator inhibitor-1), and increase intravascular coagulation. Hypoglycemia also reduces protective nitric oxide–mediated arterial vasodilator mechanisms in healthy, T1DM, and T2DM individuals.

■ HYPOGLYCEMIA IN DIABETES

Impact and Frequency Hypoglycemia is the limiting factor in the glycemic management of diabetes mellitus. First, it causes recurrent morbidity in most people with T1DM and in many with advanced T2DM, and it is sometimes fatal. Second, it precludes maintenance of euglycemia over a lifetime of diabetes and, thus, full realization of the well-established microvascular benefits of glycemic control. Third, it causes a vicious cycle of recurrent hypoglycemia by producing hypoglycemia-associated autonomic failure—i.e., the clinical syndromes of defective glucose counterregulation and of hypoglycemia unawareness.

Hypoglycemia is a fact of life for people with T1DM if treated with insulin, sulfonylurea, or glinides. They suffer an average of two episodes of symptomatic hypoglycemia per week and at least one episode of severe, at least temporarily disabling hypoglycemia each year. An estimated 6–10% of people with T1DM die as a result of hypoglycemia. The incidence of hypoglycemia is lower in T2DM than in T1DM. However, its prevalence in insulin-requiring T2DM is surprisingly high. Recent studies have revealed a hypoglycemia prevalence approaching 70%. In fact, as patients with T2DM outnumber those with T1DM by 10- to 20-fold, the prevalence of hypoglycemia is now greater in T2DM. Hypoglycemia can occur at any hemoglobin A_{1c} (HbA_{1c}) level. Although severe hypoglycemia occurs twice as frequently at lower HbA_{1c} levels in T1DM, it still occurs at HbA_{1c} levels >8%. In insulin-requiring T2DM, severe hypoglycemia can occur at lower HbA_{1c} values but also importantly at values of 8–10%. Severe hypoglycemia in T2DM carries an increased risk of severe cardiovascular and cerebrovascular morbidity and mortality for up to 1 year after the event. The risk of severe hypoglycemia and a subsequent cardiovascular adverse event is, in fact, relatively increased when trying to improve glucose control in some T2DM individuals with persistently raised HbA_{1c} values. Therefore, improvements in glycemic control in these individuals should be performed incrementally and carefully to avoid episodes of hypoglycemia. Insulin, sulfonylureas, or glinides can cause hypoglycemia in T2DM. Metformin, thiazolidinediones, α-glucosidase inhibitors, glucagon-like peptide 1 (GLP-1) receptor agonists, sodium-glucose cotransporter 2 inhibitors, and dipeptidyl peptidase IV (DPP-IV) inhibitors do not cause hypoglycemia. However, they increase the risk when combined with a sulfonylurea, glinide, or insulin. Notably, the frequency of hypoglycemia approaches that in T1DM as persons with T2DM develop absolute insulin deficiency and require more complex treatment with insulin.

Conventional Risk Factors The conventional risk factors for hypoglycemia in diabetes are identified on the basis of relative or absolute insulin excess. This occurs when (1) insulin (or insulin secretagogue) doses are excessive, ill-timed, or of the wrong type; (2) the influx of exogenous glucose is reduced (e.g., during an overnight fast, periods of temporary fasting, or after missed meals or snacks); (3) insulin-independent glucose utilization is increased (e.g., during exercise); (4) sensitivity to insulin is increased (e.g., with improved glycemic control, in the middle of the night, after exercise, or with increased fitness or weight loss); (5) endogenous glucose production is reduced (e.g., after alcohol ingestion); and (6) insulin clearance is reduced (e.g., in renal failure). However, these conventional risk factors alone explain a minority of episodes; other factors are typically involved.

Hypoglycemia-Associated Autonomic Failure (HAAF) While marked insulin excess alone can cause hypoglycemia, iatrogenic hypoglycemia in diabetes (T1DM and/or T2DM) is typically the result of the interplay of relative or absolute therapeutic insulin excess and compromised physiologic and behavioral defenses against falling plasma glucose concentrations (Table 418-2; **Fig. 418-2**). Defective glucose counterregulation compromises physiologic defense (particularly decrements in insulin and increments in glucagon and epinephrine), and hypoglycemia unawareness compromises behavioral defense (ingestion of carbohydrate).

DEFECTIVE GLUCOSE COUNTERREGULATION In the setting of absolute endogenous insulin deficiency, insulin levels do not decrease as plasma glucose levels fall; thus, the first defense against hypoglycemia is lost. After a few years of disease duration in T1DM, glucagon levels do not increase as plasma glucose levels fall; a second defense against hypoglycemia is lost. Reduced glucagon responses to hypoglycemia also occur in long-duration T2DM. However, pancreatic alpha cells that produce glucagon are present in the same number and size in T1DM as compared to age-matched nondiabetic individuals. Thus, the defect that restricts glucagon release during hypoglycemia in T1DM (and presumably in long-standing T2DM) appears to be a signaling defect, as glucagon responses to other physiologic stress in T1DM (e.g., exercise) are preserved. Finally, the increase in epinephrine levels, the third critical defense against acute hypoglycemia, is typically attenuated. The glycemic threshold for the sympathoadrenal (adrenomedullary epinephrine and sympathetic neural norepinephrine) response is shifted to lower plasma glucose concentrations. That shift is typically the result of recent antecedent iatrogenic hypoglycemia. In the setting of absent decrements in insulin and of absent increments in glucagon, the attenuated increment in epinephrine causes the clinical syndrome of defective glucose counterregulation. Affected patients are at ≥25-fold greater risk of severe iatrogenic hypoglycemia during intensive glycemic therapy for their diabetes than are patients with normal epinephrine responses. This functional—and potentially reversible—disorder is distinct from classic diabetic autonomic neuropathy, which also includes all of the above pathophysiologic defects, and is a structural and irreversible disorder.

HYPOGLYCEMIA UNAWARENESS The attenuated sympathoadrenal response (largely the reduced sympathetic neural response) to hypoglycemia causes the clinical syndrome of *hypoglycemia unawareness*—i.e., loss of the warning adrenergic and cholinergic symptoms that previously allowed the patient to recognize developing hypoglycemia and therefore to abort the episode by ingesting carbohydrates. Affected patients are at a sixfold increased risk of severe iatrogenic hypoglycemia during intensive glycemic therapy of their diabetes.

HAAF IN DIABETES The concept of HAAF in diabetes posits that recent antecedent iatrogenic hypoglycemia (or sleep or prior exercise) causes both defective glucose counterregulation (by reducing the epinephrine response to a given level of subsequent hypoglycemia in the setting of absent insulin and glucagon responses) and hypoglycemia unawareness (by reducing the sympathoadrenal response to a given level of subsequent hypoglycemia). These impaired responses, which can occur in individuals with either T1DM or T2DM, create a vicious cycle of recurrent iatrogenic hypoglycemia (Fig. 418-2). Hypoglycemia unawareness and, to some limited extent, the reduced epinephrine component of defective glucose counterregulation can be reversible by as little as 2–3 weeks of scrupulous avoidance of hypoglycemia in most affected patients.

On the basis of this pathophysiology, additional risk factors for hypoglycemia in diabetes include (1) absolute insulin deficiency, indicating that insulin levels will not decrease and glucagon levels will not increase as plasma glucose levels fall; (2) a history of severe hypoglycemia or of hypoglycemia unawareness, implying recent antecedent hypoglycemia, as well as prior exercise or sleep, indicating that the sympathoadrenal response will be attenuated; (3) impaired renal function resulting in reduced clearance of exogenous and endogenous insulin; (4) classical diabetic autonomic neuropathy; and (5) lower HbA_{1c} or lower glycemic goals even at elevated HbA_{1c} levels (8–10%), as they represent an increased probability of recent antecedent hypoglycemia.

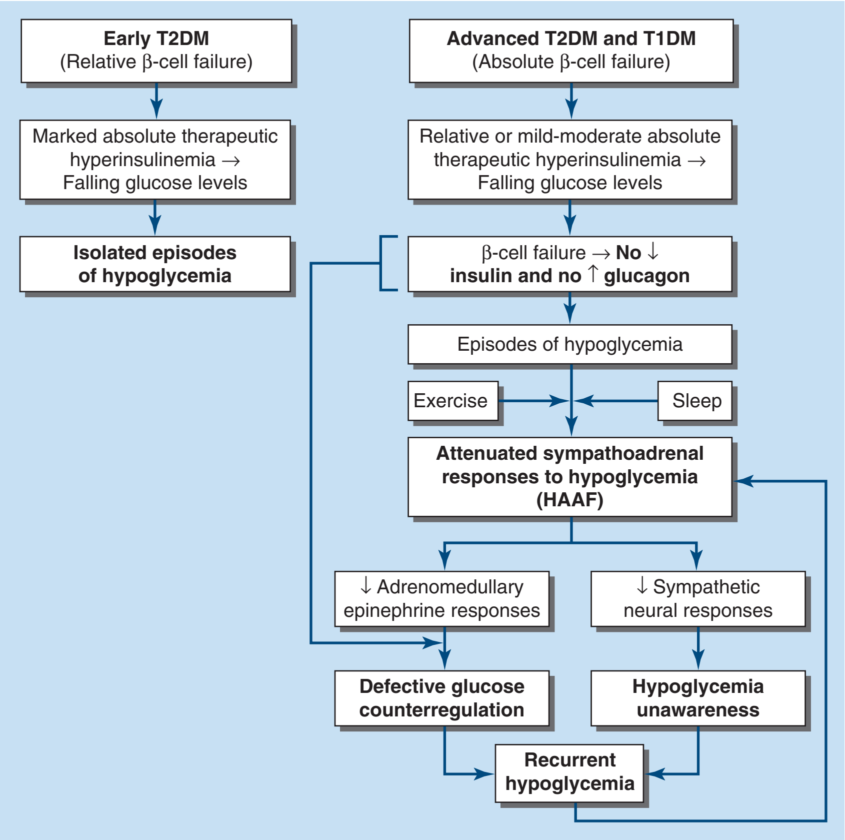
Early T2DM
(Relative β-cell failure)
Advanced T2DM and T1DM
(Absolute β-cell failure)
Marked absolute therapeutic hyperinsulinemia → Falling glucose levels
Relative or mild-moderate absolute therapeutic hyperinsulinemia → Falling glucose levels
Isolated episodes of hypoglycemia
β-cell failure → No ↓ insulin and no ↑ glucagon
Episodes of hypoglycemia
Exercise
Sleep
Attenuated sympathoadrenal responses to hypoglycemia (HAAF)
↓ Adrenomedullary epinephrine responses
↓ Sympathetic neural responses
Defective glucose counterregulation
Hypoglycemia unawareness
Recurrent hypoglycemia

FIGURE 418-2 Hypoglycemia-associated autonomic failure (HAAF) in insulin-deficient diabetes. T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus. *(Reprinted with permission from The American Diabetes Association. Copyright 2012 by the American Diabetes Association.)*

Hypoglycemia Risk Factor Reduction Several multicenter, randomized controlled trials investigating the potential benefits of tight glucose control in either inpatient or outpatient settings have reported a high prevalence of severe hypoglycemia. In the NICE-SUGAR study, attempts to control in-hospital plasma glucose values toward physiologic levels resulted in increased mortality risk. The ADVANCE and ACCORD studies and the Veterans Affairs Diabetes Trial (VADT) also found a significant incidence of severe hypoglycemia among T2DM patients. Severe hypoglycemia with accompanying serious cardiovascular morbidity and mortality also occurred in the standard (e.g., not receiving intensified treatment) control group in all of the above studies and in another large study in prediabetic and T2DM individuals (ORIGIN). Thus, as stated above, severe hypoglycemia can and does occur at HbA_{1c} values of 8–10% in both T1DM and T2DM. Somewhat surprisingly, all three studies found little or no benefit of intensive glucose control to reduce macrovascular events in T2DM. In fact, the ACCORD study was ended early because of the increased mortality rate in the intensive glucose control arm. Whether iatrogenic hypoglycemia was the cause of the increased mortality risk is not known. In light of these findings, some new recommendations and paradigms have been formulated. Whereas there is little debate regarding the need to reduce hyperglycemia in the hospital, the glycemic maintenance goals in critical care settings have been modified to stay between 140 and 180 mg/dL. Similar glycemic targets are also recommended in non–critically ill patients by a number of expert societies, although some recommend even more strict glucose control down to 108 mg/dL. Accordingly, the benefits of insulin therapy and reduced hyperglycemia can be obtained while the prevalence of hypoglycemia is reduced.

Similarly, evidence exists that intensive glucose control can reduce the prevalence of microvascular disease in both T1DM and T2DM. These benefits need to be weighed against the increased prevalence of hypoglycemia. Certainly, the level of glucose control (i.e., the HbA_{1c} value, symptoms of hyper- and hypoglycemia, and home glucose values) should be evaluated for each patient. Multicenter trials have demonstrated that individuals with recently diagnosed T1DM or T2DM can have better glycemic control with less hypoglycemia. In addition, there is still long-term benefit in reducing HbA_{1c} values from higher to lower, albeit still above recommended levels. Perhaps a reasonable therapeutic goal is the lowest HbA_{1c} level that does not cause severe hypoglycemia and that preserves awareness of hypoglycemia.

Recent studies have demonstrated the benefit of second-generation basal and prandial analogue insulins in reducing the risk of both nonsevere and severe hypoglycemia. The reduction of hypoglycemia occurred during both the day and night and was observed in T1DM and T2DM individuals. Addition of longer acting GLP-1 and dual GLP-1/gastric inhibitory polypeptide (GIP) receptor agonists to a basal insulin in the management of insulin-requiring T2DM has also resulted in lower hypoglycemic risk as compared to a basal insulin and a first-generation prandial insulin analogue. Pancreatic transplantation (both whole organ and islet cell) has been used in part as a treatment for severe hypoglycemia. Generally, rates of hypoglycemia are reduced after transplantation. This decrease appears to be due to increased physiologic insulin and glucagon responses during hypoglycemia.

The use of continuous glucose monitors (CGMs), either alone or in combination with continuous subcutaneous infusion via a wearable pump, offers promise as a method of reducing hypoglycemia while improving HbA_{1c}. Specifically, continuous glucose monitoring coupled with temporary discontinuation of subcutaneous insulin infusion when the monitor predicts a low glucose concentration is particularly promising. Studies investigating the use of CGM during inpatient care for both insulin-requiring pediatric and adult patients with diabetes are ongoing. Furthermore, progress utilizing a portable wearable closed-loop automated "artificial pancreas" or sensor-augmented pump therapy incorporating continuous glucose modulation of either insulin alone or bi-hormonal delivery of both insulin and glucagon has been established. Additionally, stem cell–derived β cells also offer promise of novel therapeutic interventions to reduce hypoglycemia.

Nonpharmacologic approaches of hypoglycemia risk reduction utilizing structured patient education have also been proven to be successful in T1DM and T2DM. Outpatient education consisting of adjustment of meal plans, exercise, and medications, combined with early recognition and treatment of hypoglycemia, have all been demonstrated to reduce hypoglycemic risk with even small improvements in HbA_{1c}. Other interventions to stimulate counterregulatory responses, such as selective serotonin reuptake inhibitors, β-adrenergic receptor antagonists, opiate receptor antagonists, and fructose, remain experimental and have not been assessed in large-scale clinical trials.

Thus, intensive glycemic therapy **(Chap. 416)** needs to be applied along with the patient's education and empowerment, frequent self-monitoring of blood glucose, flexible insulin (and other drug) regimens (including the use of insulin analogues, both short- and longer-acting), individualized glycemic goals, and ongoing professional guidance, support, and consideration of both the conventional risk factors and those indicative of compromised glucose counterregulation. Given a history of hypoglycemia unawareness, a 2- to 3-week period of scrupulous avoidance of hypoglycemia is indicated.

HYPOGLYCEMIA WITHOUT DIABETES

There are many causes of hypoglycemia (Table 418-1). Because hypoglycemia is common in insulin- or insulin secretagogue–treated diabetes, it is often reasonable to assume that a clinically suspicious episode is the result of hypoglycemia. On the other hand, because hypoglycemia is rare in the absence of relevant drug-treated diabetes (pregnancy and during severe episodes of morning sickness), it is reasonable to conclude that a hypoglycemic disorder is present only in patients in whom Whipple's triad can be demonstrated.

Particularly when patients are ill or medicated, the initial diagnostic focus should be on the possibility of drug involvement and then on critical illnesses, hormone deficiency, or non–islet cell tumor hypoglycemia. In the absence of any of these etiologic factors and in a seemingly well individual, the focus should shift to possible endogenous hyperinsulinism or accidental, surreptitious, or even malicious hypoglycemia.

Drugs Insulin and insulin secretagogues suppress glucose production and stimulate glucose utilization. Ethanol blocks gluconeogenesis but not glycogenolysis. Thus, alcohol-induced hypoglycemia typically occurs after a several-day ethanol binge during which the person eats little food, with consequent glycogen depletion. Ethanol is usually measurable in blood at the time of presentation, but its levels correlate poorly with plasma glucose concentrations. Because gluconeogenesis becomes the predominant route of glucose production during prolonged hypoglycemia, alcohol can contribute to the progression of hypoglycemia in patients with insulin-treated diabetes.

Many other drugs have been associated with hypoglycemia. These include commonly used drugs such as angiotensin-converting enzyme inhibitors and angiotensin receptor antagonists, β-adrenergic receptor antagonists, quinolone antibiotics, indomethacin, quinine, and sulfonamides.

Critical Illness Among hospitalized patients, serious illnesses such as renal, hepatic, or cardiac failure; sepsis; and inanition are second only to drugs as causes of hypoglycemia.

Rapid and extensive hepatic destruction (e.g., toxic hepatitis) causes fasting hypoglycemia because the liver is the major site of endogenous glucose production. The mechanism of hypoglycemia in patients with cardiac failure is unknown. Hepatic congestion and hypoxia may be involved. Although the kidneys are a source of glucose production, hypoglycemia in patients with renal failure is also caused by the reduced clearance of insulin (thereby inappropriately increasing insulin relative to the prevailing glucose levels) and the reduced mobilization of gluconeogenic precursors in renal failure.

Sepsis is a relatively common cause of hypoglycemia. Increased glucose utilization is induced by cytokine production in macrophage-rich tissues such as the liver, spleen, and lung. Hypoglycemia develops if glucose production fails to keep pace. Cytokine-induced inhibition of gluconeogenesis in the setting of nutritional glycogen depletion, in combination with hepatic and renal hypoperfusion, may also contribute to hypoglycemia.

Hypoglycemia can be seen with starvation. Due to brain conversion and utilization of alternative substrates, such as lactate, pyruvate, and ketone bodies, there is only a modest counterregulatory neuroendocrine and autonomic nervous system response. During periods of prolonged starvation (fasting), plasma glucose levels are lower in women as compared to men, perhaps because of loss of whole-body fat stores and subsequent depletion of gluconeogenic precursors (e.g., amino acids), necessitating increased glucose utilization.

Hormone Deficiencies Neither cortisol nor growth hormone is critical to the prevention of hypoglycemia, at least in adults. Nonetheless, hypoglycemia can occur with prolonged fasting in patients with primary adrenocortical failure (Addison's disease) or hypopituitarism. Anorexia and weight loss are typical features of chronic cortisol deficiency and likely result in glycogen depletion. Cortisol deficiency is associated with impaired gluconeogenesis and low levels of gluconeogenic precursors; these associations suggest that substrate-limited gluconeogenesis, in the setting of glycogen depletion, is the cause of hypoglycemia. Growth hormone deficiency can cause hypoglycemia in young children. In addition to extended fasting, high rates of glucose utilization (e.g., during exercise or in pregnancy) or low rates of glucose production (e.g., after alcohol ingestion) can precipitate hypoglycemia in adults with previously unrecognized hypopituitarism.

Hypoglycemia is not a feature of the epinephrine-deficient state that results from bilateral adrenalectomy when glucocorticoid replacement is adequate, nor does it occur during pharmacologic adrenergic blockade when other glucoregulatory systems are intact. Combined deficiencies of glucagon and epinephrine play a key role in the pathogenesis of iatrogenic hypoglycemia in people with insulin-deficient diabetes, as discussed earlier. Otherwise, deficiencies of these hormones are not usually considered in the differential diagnosis of a hypoglycemic disorder.

Non–β-Cell Tumors Fasting hypoglycemia, often termed *non–islet cell tumor hypoglycemia*, occurs occasionally in patients with large mesenchymal or epithelial tumors (e.g., hepatomas, adrenocortical carcinomas, carcinoids). The glucose kinetic patterns resemble those of hyperinsulinism (see next), but insulin secretion is suppressed appropriately during hypoglycemia. In most instances, hypoglycemia is due to overproduction of an incompletely processed form of insulin-like growth factor II ("big IGF-II") that does not complex normally with circulating binding proteins and thus more readily gains access to target tissues. The tumors are usually apparent clinically, plasma ratios of IGF-II to IGF-I are high, and free IGF-II levels (and levels of pro-IGF-II [1–21]) are elevated. Curative surgery is seldom possible, but reduction of tumor bulk may ameliorate hypoglycemia. Therapy with a glucocorticoid, growth hormone, or both has also been reported to alleviate hypoglycemia. Hypoglycemia attributed to ectopic IGF-I production has been reported but is rare.

Endogenous Hyperinsulinism Hypoglycemia due to endogenous hyperinsulinism can be caused by (1) a primary β-cell disorder—typically a β-cell tumor (*insulinoma*), sometimes multiple insulinomas, or a functional β-cell disorder with β-cell hypertrophy or hyperplasia; (2) an antibody to insulin or to the insulin receptor; (3) a β-cell secretagogue such as a sulfonylurea; or perhaps (4) ectopic insulin secretion, among other very rare mechanisms. None of these causes are common.

The fundamental pathophysiologic feature of endogenous hyperinsulinism caused by a primary β-cell disorder or an insulin secretagogue is the failure of insulin secretion to fall to very low levels during hypoglycemia. This feature is assessed by measurement of plasma insulin, C-peptide (the connecting peptide that is cleaved from proinsulin to produce insulin), proinsulin, and glucose concentrations during hypoglycemia. Insulin, C-peptide, and proinsulin levels need not be high relative to normal, euglycemic values; rather, they are inappropriately high in the setting of a low plasma glucose concentration. Critical diagnostic findings are a plasma insulin concentration ≥3 μU/mL (≥18 pmol/L), a plasma C-peptide concentration ≥0.6 ng/mL (≥0.2 nmol/L), and a plasma proinsulin concentration ≥5.0 pmol/L when the plasma glucose concentration is <55 mg/dL (<3.0 mmol/L) with symptoms of hypoglycemia. A low plasma β-hydroxybutyrate concentration (≤2.7 mmol/L) and an increment in plasma glucose level of >25 mg/dL (>1.4 mmol/L) after IV administration of glucagon (1.0 mg) indicate increased insulin (or IGF) actions.

The diagnostic strategy is (1) to measure plasma glucose, insulin, C-peptide, proinsulin, and β-hydroxybutyrate concentrations and to screen for circulating oral hypoglycemic agents during an episode of hypoglycemia and (2) to assess symptoms during the episode and seek their resolution following correction of hypoglycemia by glucose (either oral or parenteral) or by IV injection of glucagon (i.e., to document Whipple's triad). This is straightforward if the patient is hypoglycemic when seen. Since endogenous hyperinsulinemic disorders usually, but not invariably, cause fasting hypoglycemia, a diagnostic episode may develop after a relatively short outpatient fast. Serial sampling during an inpatient diagnostic fast of up to 72 h or after a mixed meal is more problematic. An alternative is to give patients a detailed list of the required measurements and ask them to present to an ambulatory care center or emergency room, with the list, during a symptomatic episode. Obviously, a normal plasma glucose concentration during a symptomatic episode indicates that the symptoms are not the result of hypoglycemia.

An *insulinoma*—an insulin-secreting pancreatic islet β-cell tumor—is the prototypical cause of endogenous hyperinsulinism and therefore should be sought in patients with a compatible clinical syndrome. However, insulinoma is not the only cause of endogenous hyperinsulinism. Some patients with fasting endogenous hyperinsulinemic hypoglycemia have diffuse islet involvement with β-cell hypertrophy and sometimes hyperplasia. This pattern is commonly referred to as *nesidioblastosis*, although β cells budding from ducts are not invariably found. Other patients have a similar islet pattern but with postprandial hypoglycemia, a disorder termed *noninsulinoma pancreatogenous hypoglycemia*. Post–gastric bypass postprandial hypoglycemia, which most often follows Roux-en-Y gastric bypass, is also characterized by diffuse islet involvement and endogenous hyperinsulinism. Multiple pathophysiologic mechanisms have been suggested including exaggerated GLP-1 responses to meals resulting in hyperinsulinemia, hypoglucagonemia, and hypoglycemia. However, other mechanisms may be responsible for the relative hyperinsulinemia, such as reduced insulin clearance and reduced glucagon responses to hypoglycemia. The relevant pathogenesis has not been clearly established. However, if medical treatment with agents such as an α-glucosidase inhibitor, diazoxide, or octreotide fails, partial pancreatectomy may be required. Autoimmune hypoglycemias include those caused by an antibody to insulin that binds postmeal insulin and then gradually disassociates, with consequent late postprandial hypoglycemia. Alternatively, an insulin receptor antibody can function as an agonist. The presence of an insulin secretagogue, such as a sulfonylurea or a glinide, results in a clinical and biochemical pattern similar to that of an insulinoma but can be distinguished by the presence of the circulating secretagogue. Finally, there are reports of very rare phenomena such as ectopic insulin secretion, a gain-of-function insulin receptor mutation, and exercise-induced hyperinsulinemia.

Insulinomas are uncommon, with an estimated yearly incidence of 1 in 250,000. Because >90% of insulinomas are benign, they are a treatable cause of potentially fatal hypoglycemia. The median age at presentation is 50 years in sporadic cases, but the tumor usually presents in the third decade when it is a component of multiple endocrine neoplasia type 1 **(Chap. 400)**. More than 99% of insulinomas are within the substance of the pancreas, and the tumors are usually small (<2.0 cm in diameter in 90% of cases). Therefore, they come to clinical attention because of hypoglycemia rather than mass effects. Computed tomography or magnetic resonance imaging detects ~70–80% of insulinomas. These methods detect metastases in the roughly 10% of patients with a malignant insulinoma. Transabdominal ultrasound often identifies insulinomas, and endoscopic ultrasound has a sensitivity of ~90%. Somatostatin receptor scintigraphy is thought to detect insulinomas in about half of patients. Selective pancreatic arterial calcium injections, with the endpoint of a sharp increase in hepatic venous insulin levels, regionalize insulinomas with high sensitivity, but this invasive procedure is seldom necessary except to confirm endogenous hyperinsulinism in the diffuse islet disorders. Intraoperative pancreatic ultrasonography almost invariably localizes insulinomas that are not readily palpable by the surgeon. Surgical resection of a solitary insulinoma is generally curative. Diazoxide, which inhibits insulin secretion, or the somatostatin analogue octreotide can be used to treat hypoglycemia in patients with unresectable tumors; everolimus, an mTOR (mammalian target of rapamycin) inhibitor, has also been successful in combination with the above approaches.

■ ACCIDENTAL, SURREPTITIOUS, OR MALICIOUS HYPOGLYCEMIA

Accidental ingestion of an insulin secretagogue (e.g., as the result of a pharmacy or other medical error) or even accidental administration of insulin can occur. Factitious hypoglycemia, caused by surreptitious or even malicious administration of insulin or an insulin secretagogue, shares many clinical and laboratory features with insulinoma. It is most common among health care workers, patients with diabetes or their relatives, and people with a history of other factitious illnesses. However, it should be considered in all patients being evaluated for hypoglycemia of obscure cause. Ingestion of an insulin secretagogue causes hypoglycemia with increased C-peptide levels, whereas exogenous insulin causes hypoglycemia with low C-peptide levels, reflecting suppression of insulin secretion.

Analytical error in the measurement of plasma glucose concentrations is rare. On the other hand, hand-held and continuous glucose monitors used to guide treatment of diabetes are not quantitative instruments, particularly at low glucose levels, and should not be used for the definitive diagnosis of hypoglycemia. Even with a quantitative method, low measured glucose concentrations can be artifactual—e.g., the result of continued glucose metabolism by the formed elements of the blood ex vivo, particularly in the presence of leukocytosis, erythrocytosis, or thrombocytosis or with delayed separation of the serum from the formed elements (pseudohypoglycemia).

INBORN ERRORS OF METABOLISM CAUSING HYPOGLYCEMIA

Nondiabetic hypoglycemia also results from inborn errors of metabolism. Such hypoglycemia most commonly occurs in infancy but can also occur in adulthood. Cases in adults can be classified into those resulting in fasting hypoglycemia, postprandial hypoglycemia, and exercise-induced hypoglycemia.

Fasting Hypoglycemia Although rare, disorders of glycogenolysis can result in fasting hypoglycemia. These disorders include glycogen storage disease (GSD) of types 0, I, III, and IV and Fanconi-Bickel syndrome **(Chap. 430)**. Patients with GSD types I and III characteristically have high blood lactate levels before and after meals, respectively. Both groups have hypertriglyceridemia, but ketones are high in GSD type III. Defects in fatty acid oxidation also result in fasting hypoglycemia. These defects can include (1) defects in the carnitine cycle; (2) fatty-acid β-oxidation disorders; (3) electron transfer disturbances; and (4) ketogenesis disorders. Finally, defects in gluconeogenesis (fructose-1,6-biphosphatase) have been reported to result in recurrent hypoglycemia and lactic acidosis.

Postprandial Hypoglycemia Inborn errors of metabolism resulting in postprandial hypoglycemia are also rare. These errors include (1) glucokinase, SUR1, and Kir6.2 potassium channel mutations; (2) congenital disorders of glycosylation; and (3) inherited fructose intolerance.

Exercise-Induced Hypoglycemia Exercise-induced hypoglycemia, by definition, follows exercise. It results in hyperinsulinemia caused by increased activity of monocarboxylate transporter 1 in β cells.

APPROACH TO THE PATIENT

Hypoglycemia

In addition to the recognition and documentation of hypoglycemia as well as its treatment (often on an urgent basis), diagnosis of the hypoglycemic mechanism is critical for the selection of therapy that prevents, or at least minimizes, recurrent hypoglycemia.

RECOGNITION AND DOCUMENTATION

Hypoglycemia is suspected in patients with typical symptoms; in the presence of confusion, an altered level of consciousness, or a seizure; or in a clinical setting in which hypoglycemia is known to occur. Blood should be drawn, whenever possible, before the administration of glucose to allow documentation of a low plasma glucose concentration. Convincing documentation of hypoglycemia requires the fulfillment of Whipple's triad. Thus, the ideal time to measure the plasma glucose level is during a symptomatic episode. A normal glucose level excludes hypoglycemia as the cause of the symptoms. A low glucose level confirms that hypoglycemia is the cause of the symptoms, provided the latter resolve after the glucose level is raised. When the cause of the hypoglycemic episode is obscure, additional measurements—made while the glucose level is low and before treatment—should include plasma insulin, C-peptide, proinsulin, and β-hydroxybutyrate levels; also critical are screening for circulating oral hypoglycemic agents and assessment of symptoms before and after the plasma glucose concentration is raised.

When the history suggests prior hypoglycemia and no potential mechanism is apparent, the diagnostic strategy is to evaluate the patient as just described and assess for Whipple's triad during and after an episode of hypoglycemia. On the other hand, while it cannot be ignored, a distinctly low plasma glucose concentration measured in a patient without corresponding symptoms raises the possibility of an artifact (pseudohypoglycemia).

DIAGNOSIS OF THE HYPOGLYCEMIC MECHANISM

In a patient with documented hypoglycemia, a plausible hypoglycemic mechanism can often be deduced from the history, physical examination, and available laboratory data (Table 418-1). Drugs, particularly alcohol or agents used to treat diabetes, should be the first consideration—even in the absence of known use of a relevant drug—given the possibility of surreptitious, accidental, or malicious drug administration. Other considerations include evidence of a relevant critical illness, hormone deficiencies (less commonly), and a non-β-cell tumor that can be pursued diagnostically (rarely). Absent one of these mechanisms in an otherwise seemingly well individual, the care provider should consider endogenous hyperinsulinism and proceed with measurements and assessment of symptoms during spontaneous hypoglycemia or under conditions that might elicit hypoglycemia.

URGENT TREATMENT

If the patient is able and willing, oral treatment with glucose tablets or glucose-containing fluids, candy, or food is appropriate. A reasonable initial dose is 15–20 g of glucose. If the patient is unable or unwilling (because of neuroglycopenia) to take carbohydrates orally, parenteral therapy is necessary. IV administration of glucose (25 g) should be followed by a glucose infusion guided by serial plasma glucose measurements. If IV therapy is not practical, SC or IM glucagon (1.0 mg in adults) can be used, particularly in patients with T1DM. Because it acts by stimulating glycogenolysis, glucagon is ineffective in glycogen-depleted individuals (e.g., those with alcohol-induced hypoglycemia). Glucagon also stimulates insulin secretion and is therefore less useful in T2DM. The somatostatin analogue octreotide can be used to suppress insulin secretion in sulfonylurea-induced hypoglycemia. These treatments raise plasma glucose concentrations only transiently, and patients should therefore be urged to eat as soon as is practical to replete glycogen stores.

PREVENTION OF RECURRENT HYPOGLYCEMIA

Prevention of recurrent hypoglycemia requires an understanding of the hypoglycemic mechanism. Offending drugs should be discontinued or their doses reduced. Hypoglycemia caused by a sulfonylurea can persist for hours or even days. Underlying critical illnesses can often be treated. Cortisol and growth hormone can be replaced if levels are deficient. Surgical, radiotherapeutic, or chemotherapeutic reduction of a non–islet cell tumor can alleviate hypoglycemia even if the tumor cannot be cured; glucocorticoid or growth hormone administration also may reduce hypoglycemic episodes in such patients. Surgical resection of an insulinoma is curative; medical therapy with diazoxide or octreotide can be used if complete resection is not possible and in patients with a nontumor β-cell disorder. Partial pancreatectomy may be necessary in the latter patients. The treatment of autoimmune hypoglycemia (e.g., with glucocorticoid or immunosuppressive drugs) is problematic, but these disorders are sometimes self-limited. Failing these treatments, frequent feedings and avoidance of fasting may be required. Administration of uncooked cornstarch at bedtime or even an overnight intragastric infusion of glucose may be necessary for some patients.

Acknowledgment

The author acknowledges contributions of the late Phil E. Cryer to this chapter in prior editions of Harrison's.

FURTHER READING

Black JE et al: Real-world effects of second-generation versus earlier intermediate/basal insulin analogues on rates of hypoglycemia in adults with type 1 and 2 diabetes (iNPHORM, US). Diabetes Ther 14:1299, 2023.

Cryer PE: *Hypoglycemia in Diabetes*, 3rd ed. Alexandria, VA, American Diabetes Association, 2016.

Cryer PE: Hypoglycemia, in *Williams Textbook of Endocrinology*, 13th ed, S Melmed et al (eds). Philadelphia, Saunders, 2016, pp. 1582–1607.

Lee AK et al: The association of severe hypoglycemia with incident cardiovascular events and mortality in adults with type 2 diabetes. Diabetes Care 41:104, 2018.

Nwokolo M, Hovorka R: The artificial pancreas and type 1 diabetes. J Clin Endocrinol Metab 108:1614, 2023.

Salehi M et al: Hypoglycemia after gastric bypass surgery: Current concepts and controversies. J Clin Endocrinol Metab 103:2815, 2018.

419

Disorders of Lipoprotein Metabolism

Daniel J. Rader

Lipoproteins are complexes of lipids and proteins that are essential for transport of cholesterol, triglycerides (TGs), and fat-soluble vitamins in the blood. Lipoproteins play essential roles in the transport of dietary cholesterol, long-chain fatty acids, and fat-soluble vitamins from the intestine to peripheral tissues and the liver; the transport of TGs, cholesterol, and fat-soluble vitamins from the liver to peripheral tissues; and the transport of cholesterol from peripheral tissues back to the liver and intestine for excretion. Disorders of lipoprotein metabolism can be primary (caused by genetic conditions) or secondary (to other medical conditions or environmental exposures) and involve either a substantial increase or decrease in specific circulating lipids or lipoproteins. Lipoprotein disorders can have a number of clinical consequences, most notably atherosclerotic cardiovascular disease (ASCVD), and are therefore important to appropriately diagnose and treat. This chapter reviews the etiology and pathophysiology of disorders of lipoprotein metabolism and clinical approaches to their diagnosis and management.

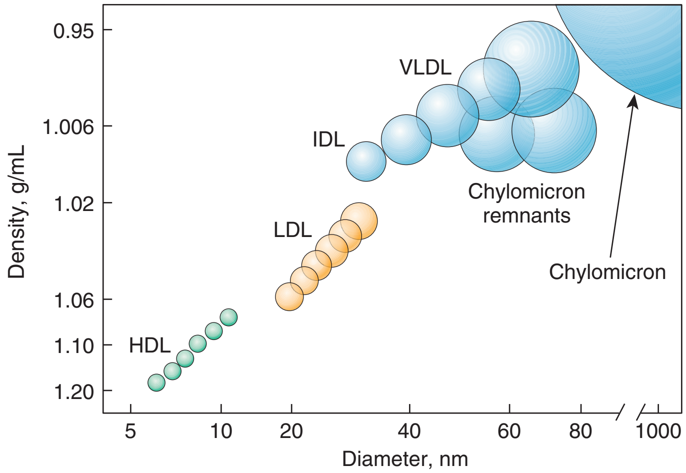

FIGURE 419-1 The density and size distribution of the major classes of lipoprotein particles. Lipoproteins are classified by density and size, which are inversely related. HDL, high-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; VLDL, very-low-density lipoprotein.

LIPOPROTEIN STRUCTURE AND METABOLISM

Lipoproteins contain an "oil droplet" core of hydrophobic lipids (TGs and cholesteryl esters) surrounded by a shell of hydrophilic lipids (phospholipids, unesterified cholesterol) and proteins (called apolipoproteins) that interact with body fluids **(Fig. 419-1)**. The plasma lipoproteins are divided into major classes based on their relative density: chylomicrons, very-low-density lipoproteins (VLDLs), intermediate-density lipoproteins (IDLs), low-density lipoproteins (LDLs), and high-density lipoproteins (HDLs). Each lipoprotein class comprises a family of particles that vary in density, size, and protein composition. Because lipid is less dense than water, the density of a lipoprotein particle is primarily determined by the amount of lipid per particle. Chylomicrons are the most lipid-rich and therefore least dense lipoprotein particles, whereas HDL have the least lipid and are therefore the most dense. Lipoprotein particles vary widely in size, with the largest particles (chylomicrons) being the most lipid-rich and the smallest particles (HDL) being the most dense.

The proteins associated with lipoproteins, called *apolipoproteins* **(Table 419-1)**, are required for the assembly, structure, function, metabolism, and catabolism of lipoproteins. Apolipoproteins provide a structural basis for lipoproteins, activate enzymes important in lipoprotein metabolism, and act as ligands for cell surface receptors. ApoB is the major structural protein of chylomicrons, VLDLs, IDLs, and LDLs (collectively known as apoB-containing lipoproteins). One

TABLE 419-1 Major Apolipoproteins

APOLIPOPROTEIN	PRIMARY SOURCE	LIPOPROTEIN ASSOCIATION	FUNCTION
ApoA-I	Intestine, liver	HDL, chylomicrons	Core structural protein for HDL, promotes cellular lipid efflux via ABCA1, activates LCAT
ApoA-II	Liver	HDL, chylomicrons	Structural protein for HDL
ApoA-V	Liver	VLDL, chylomicrons	Promotes LPL-mediated triglyceride lipolysis
Apo(a)	Liver	Lp(a)	Structural protein for Lp(a)
ApoB-48	Intestine	Chylomicrons, chylomicron remnants	Core structural protein for chylomicrons
ApoB-100	Liver	VLDL, IDL, LDL, Lp(a)	Core structural protein for VLDL, LDL, IDL, Lp(a); ligand for binding to LDL receptor (except for Lp(a))
ApoC-II	Liver	Chylomicrons, VLDL, HDL	Cofactor for LPL
ApoC-III	Liver, intestine	Chylomicrons, VLDL, HDL	Inhibitor of LPL activity and remnant lipoprotein binding to receptors
ApoE	Liver	Chylomicron remnants, IDL, HDL	Ligand for binding to LDL receptor and other receptors

Abbreviations: HDL, high-density lipoprotein; IDL, intermediate-density lipoprotein; LCAT, lecithin-cholesterol acyltransferase; LDL, low-density lipoprotein; Lp(a), lipoprotein(a); LPL, lipoprotein lipase; VLDL, very-low-density lipoprotein.

molecule of apoB, either apoB-48 (chylomicrons) or apoB-100 (VLDL, IDL, or LDL), is present on each lipoprotein particle. The human liver synthesizes the full-length apoB-100 (one of the largest proteins in humans), whereas the intestine makes the shorter apoB-48, which is derived from transcription of the same *APOB* gene after posttranscriptional mRNA editing. HDLs lack apoB and have different apolipoproteins that define this lipoprotein class, most importantly apoA-I, which is synthesized in both the liver and intestine and is found on virtually all HDL particles. ApoA-II is the second most abundant HDL apolipoprotein and is on approximately two-thirds of the HDL particles. ApoC-II, apoC-III, and apoA-V regulate the metabolism of TG-rich lipoproteins. ApoE plays a critical role in the metabolism and clearance of TG-rich particles. Most apolipoproteins, other than apoB, exchange actively among lipoprotein particles in the blood. Apolipoprotein(a) [apo(a)] is a distinctive apolipoprotein that results in the formation of a lipoprotein known as lipoprotein(a) [Lp(a)], which is discussed more below.

■ TRANSPORT OF INTESTINALLY DERIVED DIETARY LIPIDS BY CHYLOMICRONS

Chylomicrons play a critical role in the efficient transport of absorbed dietary lipids from the intestine to tissues that require fatty acids for energy or storage and then return of cholesterol, lipids, and fat-soluble vitamins to the liver **(Fig. 419-2)**. Dietary lipids are hydrolyzed by lipases within the intestinal lumen and emulsified with bile acids to form micelles. Dietary cholesterol, fatty acids, and fat-soluble vitamins are absorbed in the proximal small intestine. Cholesterol and retinol are esterified (by the addition of a fatty acid) in the enterocyte to form cholesteryl esters and retinyl esters, respectively. Longer-chain fatty acids (>12 carbons) are incorporated into TGs and packaged with apoB-48, phospholipids, cholesteryl esters, retinyl esters, and α-tocopherol (vitamin E) in a process that requires the action of the microsomal TG transfer protein (MTP) to form chylomicrons. Nascent chylomicrons are secreted into the intestinal lymph and delivered via the thoracic duct directly to the systemic circulation, where they are extensively processed by peripheral tissues before reaching the liver. The particles encounter lipoprotein lipase (LPL), which is transported from parenchymal cells and anchored to the luminal surface of the endothelium by the protein GPIHBP1 of capillaries in tissues, particularly adipose, heart, and skeletal muscle (Fig. 419-2). ApoC-II and apoA-V are apolipoproteins that are transferred to circulating chylomicrons from HDL in the postprandial state; apoC-II acts as a required cofactor for LPL activation, and apoA-V serves as a facilitator of LPL activity. The TGs in chylomicrons are hydrolyzed by LPL, and free fatty acids are released and taken up by adjacent myocytes or adipocytes and are either oxidized to generate energy or reesterified and stored as TG. Some of the released free fatty acids bind albumin before entering cells and are transported to other tissues, especially the liver. The chylomicron particle progressively shrinks in size as the hydrophobic TG core is hydrolyzed and the excess hydrophilic lipids (cholesterol and phospholipids) and apolipoproteins on the particle surface are transferred to HDL, ultimately creating chylomicron remnants.

Chylomicron remnants contain apoB-48, which lacks the region in apoB-100 that binds to the LDL receptor. Nevertheless, they are rapidly removed from the circulation by the liver through a process that critically requires apoE as a ligand for receptors in the liver. Few, if any, chylomicrons or chylomicron remnants are generally present in the blood after a 12-h fast, except in patients with certain disorders of lipoprotein metabolism as described below.

■ TRANSPORT OF HEPATICALLY DERIVED LIPIDS BY VLDL AND LDL

Another key role of lipoproteins is the transport of hepatic lipids from the liver to the periphery (Fig. 419-2) to provide an energy source during fasting and to deliver fat-soluble vitamins to key tissues. During the fasting state, lipolysis of adipose TGs generates fatty acids that are transported to the liver, and the liver is also capable of synthesizing fatty acids through de novo lipogenesis. These fatty acids are esterified by the liver into TGs, which are packaged into VLDL particles along with apoB-100, phospholipids, cholesteryl esters, and vitamin E in a process that also requires MTP. VLDL thus resemble chylomicrons in that they are "triglyceride-rich lipoproteins," but they contain apoB-100 rather than apoB-48, are smaller and less buoyant, and have a higher ratio of cholesterol to TG (~1 mg of cholesterol for every 5 mg of TG, whereas in chylomicrons, this ratio is closer to ~1:8). After secretion by the liver into the plasma, the circulating TGs in VLDL are hydrolyzed by LPL. After the relatively TG-depleted VLDL remnants dissociate from LPL, they are referred to as IDLs, which contain roughly similar amounts of cholesterol and TG by mass. The liver removes ~40–60% of IDL by receptor-mediated endocytosis via binding to apoE, which is acquired through transfer of this protein from HDL. The remainder of IDL is further remodeled by hepatic lipase (HL) to form LDL. During this process, phospholipids and TGs in the particle are hydrolyzed, and most of the remaining apolipoproteins except apoB-100 are transferred to other lipoproteins. LDL is primarily a by-product of fatty acid energy transport by VLDL with little true physiologic role; one exception is that LDL may be partially responsible for delivery of vitamin E to the retina and brain. LDL is ultimately removed from the circulation by receptor-mediated endocytosis (primarily via the LDL receptor) in the liver, with a region of apoB-100 serving as the specific ligand for binding to the LDL receptor. It should be noted that apoB-48 does not contain the LDL receptor-binding ligand region, and therefore, clearance of apoB-48-containing chylomicron remnants is dependent on apoE-mediated clearance as noted above. Some LDL particles are lipolytically processed to small dense LDL particles that are believed to be especially atherogenic.

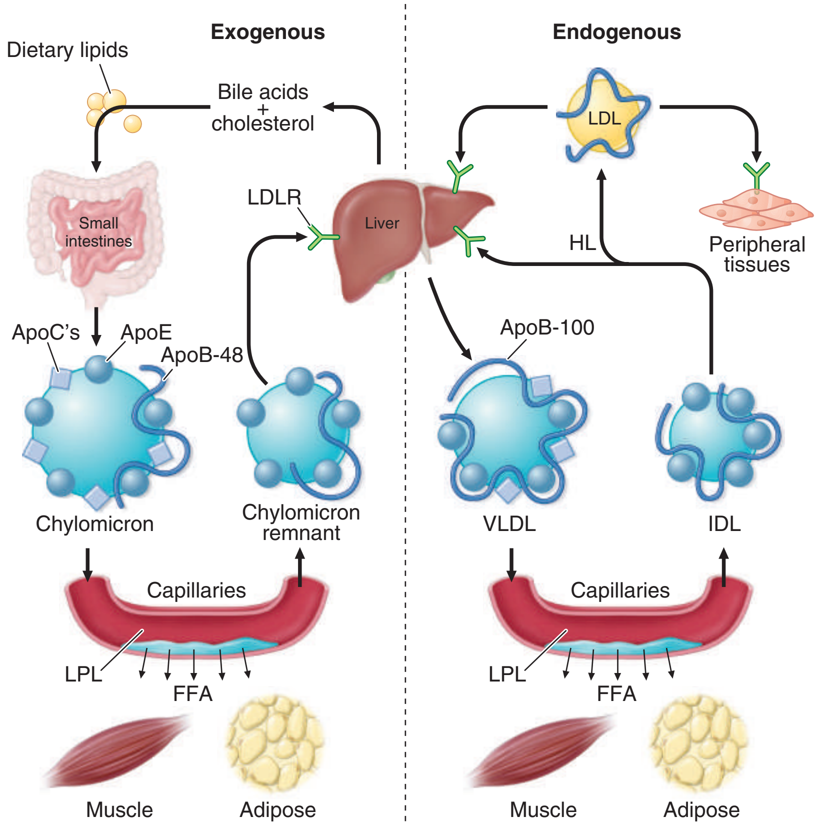

FIGURE 419-2 The exogenous and endogenous lipoprotein metabolic pathways. The exogenous pathway transports dietary lipids to the periphery and the liver. The endogenous pathway transports hepatic lipids to the periphery. FFA, free fatty acid; HL, hepatic lipase; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; LDLR, low-density lipoprotein receptor; LPL, lipoprotein lipase; VLDL, very-low-density lipoprotein.

Lp(a) is a lipoprotein similar to LDL in lipid and protein composition, but it contains an additional distinctive protein called apo(a). Apo(a) is synthesized in the liver and attached to apoB-100 by a disulfide linkage. The major site of clearance of Lp(a) is the liver, but the uptake pathway is not known. Lp(a) is now established as causal factor for ASCVD and aortic stenosis, and an elevated level of Lp(a) serves as an independent risk factor and merits more aggressive therapy to reduce LDL cholesterol levels (see below).

■ HDL METABOLISM AND REVERSE CHOLESTEROL TRANSPORT

All nucleated cells synthesize cholesterol, but only hepatocytes and enterocytes can effectively excrete cholesterol from the body, into either the bile or the gut lumen, respectively. In the liver, cholesterol is secreted into the bile, either directly or after conversion to bile acids. Cholesterol in peripheral cells is transported from the plasma membranes of peripheral cells to the liver and intestine by a process termed *reverse cholesterol transport* that is facilitated by HDL **(Fig. 419-3)**.

Nascent HDL particles are synthesized by the intestine and the liver. Newly secreted apoA-I rapidly acquires phospholipids and unesterified cholesterol from its site of synthesis (intestine or liver) via cellular efflux promoted by the membrane protein ATP-binding cassette protein A1 (ABCA1). This process results in the formation of discoidal HDL particles, which then recruit additional unesterified cholesterol from cells or circulating lipoproteins. Within the HDL particle, the cholesterol is esterified to cholesteryl ester (CE) through the addition of a fatty acid by lecithin-cholesterol acyltransferase (LCAT), a plasma enzyme associated with HDL; the hydrophobic CE forms the core of the mature HDL particle. As HDL acquires more CE, it becomes spherical, and additional apolipoproteins and lipids are transferred to the particles from the surfaces of chylomicrons and VLDLs during lipolysis.

HDL cholesterol in the blood is transported to hepatocytes by two major pathways. HDL CE can be "selectively" taken up by hepatocytes via the scavenger receptor class B1 (SR-B1), a cell surface HDL receptor that mediates the selective transfer of CE from HDL with subsequent dissociation and "recycling" of the HDL particle. In addition, HDL CE can be transferred to apoB-containing lipoproteins in exchange for TG by the cholesteryl ester transfer protein (CETP). The CE esters are then removed from the circulation by LDL receptor–mediated endocytosis. HDL-derived CE taken up by the hepatocyte through these pathways is hydrolyzed, and much of the cholesterol is ultimately excreted directly into the bile or converted to bile acids with excretion to bile, providing a biliary route into the intestinal lumen. There is also evidence that, under certain conditions, HDL cholesterol can be transported directly into the intestinal lumen without requiring a transhepatobiliary route, a process known as transintestinal cholesterol excretion.

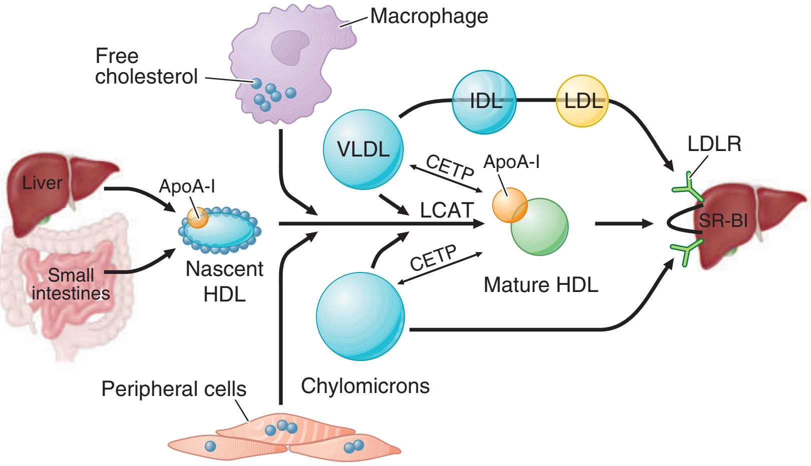

FIGURE 419-3 High-density lipoprotein (HDL) metabolism and reverse cholesterol transport. The HDL pathway transports excess cholesterol from the periphery back to the liver for excretion in the bile. The liver and the intestine produce nascent HDLs. Free cholesterol is acquired from macrophages and other peripheral cells and esterified by lecithin-cholesterol acyltransferase (LCAT), forming mature HDLs. HDL cholesterol can be selectively taken up by the liver via SR-BI (scavenger receptor class BI). Alternatively, HDL cholesteryl ester can be transferred by cholesteryl ester transfer protein (CETP) from HDLs to very-low-density lipoproteins (VLDLs) and chylomicrons, which can then be taken up by the liver. IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; LDLR, low-density lipoprotein receptor.

HDL particles undergo extensive remodeling within the plasma compartment by a variety of lipid transfer proteins and lipases. The phospholipid transfer protein (PLTP) transfers phospholipids from other lipoproteins to HDL or among different classes of HDL particles and is a regulator of HDL metabolism. After CETP- and PLTP-mediated lipid exchange, the TG-enriched HDL becomes a much better substrate for HL, which hydrolyzes the TGs and phospholipids to generate smaller HDL particles. A related enzyme called endothelial lipase (EL) hydrolyzes HDL phospholipids, generating smaller HDL particles that are catabolized faster. Remodeling of HDL influences the metabolism, function, and plasma concentrations of HDL.

SCREENING

Dyslipidemia is an important causal factor in ASCVD, and treatment has been proven to substantially reduce cardiovascular risk. Therefore, all adults (and many children) should be actively screened for plasma lipids. A lipid panel should be measured, preferably after an overnight fast. In most clinical laboratories, the total cholesterol and TGs in the plasma are measured enzymatically, and then after precipitation of apoB-containing lipoproteins, the cholesterol in the supernatant is measured to determine the HDL cholesterol (HDL-C). The LDL cholesterol (LDL-C) is then estimated using the following equation (the Friedewald formula):

$$\text{LDL-C} = \text{total cholesterol} - (\text{TG}/5) - \text{HDL-C}$$

(The VLDL cholesterol content is estimated by dividing the plasma TG by 5, reflecting the ratio of TG to cholesterol in VLDL particles.) This formula is reasonably accurate if test results are obtained on fasting plasma and if the TG level does not exceed ~200 mg/dL; by convention, it cannot be used if the TG level is >400 mg/dL. LDL-C can be directly measured by a number of methods. The non-HDL-C can be easily calculated by subtracting the HDL-C from the total cholesterol. It has the advantage of incorporating the cholesterol contained within VLDL and IDL, which in most cases is also atherogenic and associated with increased ASCVD risk. There is increasing evidence that measurement of plasma apoB levels may provide a better assessment of cardiovascular risk than the LDL-C level, and even the non-HDL-C level, and is recommended by some experts. While this has not yet become standard clinical practice, the data supporting the use of apoB as a risk marker and guide to therapeutic intervention are quite strong. There is also increasing interest in Lp(a), an independent ASCVD risk factor that is highly heritable and may be helpful in risk stratification. In patients with evidence of dyslipidemia, further evaluation and treatment are based on evidence of preexisting ASCVD and clinical assessment of cardiovascular risk using risk calculators such as the American Heart Association (AHA)/American College of Cardiology (ACC) risk calculator as well as, in some cases, based on additional approaches to risk assessment such as apoB and Lp(a) (see "Approach to the Patient" for more detailed discussion).

DISORDERS ASSOCIATED WITH ELEVATED APOB-CONTAINING LIPOPROTEINS

Disorders of lipoprotein metabolism that cause elevated levels of apoB-containing lipoproteins are among the most common and clinically important of the dyslipoproteinemias. They are generally characterized by increased plasma levels of total cholesterol, accompanied by increased TGs, LDL-C, or both. Many patients with hyperlipidemia have some combination of genetic predisposition (often polygenic) and medical or environmental contribution (medical condition, diet, lifestyle, or drug). Many, but not all, patients with hyperlipidemia are at increased risk for ASCVD, which is the primary reason for making the diagnosis, as intervention can substantially reduce this risk. In addition, patients with severe hypertriglyceridemia may be at risk for acute pancreatitis and require intervention to reduce this risk.

Although hundreds of proteins influence lipoprotein metabolism, and genetic variants in most of the genes that encode them interact with each other and the environment to produce dyslipidemia, there are a limited number of discrete "nodes" or pathways that regulate lipoprotein metabolism and are dysfunctional in specific dyslipidemias. These include (1) lipolysis of TG-rich lipoproteins by LPL; (2) receptor-mediated uptake of apoB-containing lipoproteins by the liver; (3) cellular cholesterol metabolism in the hepatocyte and the enterocyte; (4) assembly and secretion of VLDLs by the liver; and (5) neutral lipid transfer and phospholipid hydrolysis in the plasma. Primary genetic disorders of lipoprotein metabolism caused by single-gene mutations **(Table 419-2)** have taught us a great deal about the physiologic roles of specific proteins in these pathways in humans and are clinically important to diagnose and treat.

■ SEVERE HYPERTRIGLYCERIDEMIA

Severe hypertriglyceridemia (HTG) is defined by fasting TG levels >500 mg/dL and is usually accompanied by moderately elevated total cholesterol levels and reduced levels of HDL-C, usually without important elevation in LDL-C or apoB. It is medically important because it is associated with risk of acute pancreatitis and, in some cases, is also associated with increased risk of ASCVD. Severe HTG is usually caused by impaired lipolysis of TGs in TG-rich lipoproteins (TRLs) by the enzyme LPL. LPL is synthesized by adipocytes, skeletal myocytes, and cardiomyocytes, and its posttranslational maturation and folding require the action of lipase maturation factor 1 (LMF1). After secretion, it is transported from the subendothelial to the vascular endothelial surfaces by GPIHPB1, which docks it to the endothelial surface. ApoC-II is a required cofactor for LPL, and apoA-V promotes LPL activity, and both are transported to the bound LPL on the TRLs. Single-gene Mendelian disorders that reduce LPL have been described **(Table 419-3)** as reviewed below; the majority of patients with severe HTG have a polygenic predisposition to secondary factors like obesity or insulin resistance.

Primary (Genetic) Causes of Severe Hypertriglyceridemia •

FAMILIAL CHYLOMICRONEMIA SYNDROME (FCS) LPL is required for the hydrolysis of TGs in chylomicrons and VLDLs. Genetic deficiency or inactivity of LPL results in impaired lipolysis and profound elevations in plasma TGs, mostly in chylomicrons. While chylomicronemia predominates, in fact, these patients often have elevated plasma levels of VLDL as well. The fasting plasma is turbid, and if left undisturbed for several hours, the chylomicrons float to the top and form a creamy supernatant layer. Fasting TG levels are >500 mg/dL and usually >1000 mg/dL. Because chylomicrons contain cholesterol, fasting total cholesterol levels are also elevated. There are five genes in which mutations can result in FCS (Table 419-2). FCS has an estimated frequency of ~1 in 200,000–300,000, although its true prevalence is unknown. The most common molecular cause of FCS involves mutations in the *LPL* gene. *LPL deficiency* has autosomal recessive inheritance (loss-of-function mutations in both alleles). Heterozygotes with *LPL* mutations often have moderate elevations in plasma TG levels and increased risk for coronary heart disease (CHD). FCS can also be caused by mutations in genes that affect LPL processing or activity. For example, apoC-II is a required cofactor for LPL. *APOC2 deficiency* due to loss-of-function mutations in both *APOC2* alleles results in functional lack of LPL activity and severe hyperchylomicronemia that is indistinguishable from LPL deficiency. It is also recessive in inheritance pattern and much rarer than LPL deficiency. Another apolipoprotein, apoA-V, facilitates the association of TRLs with LPL and promotes hydrolysis of the TGs. Individuals harboring loss-of-function mutations in both *APOA5* alleles causing *APOA5 deficiency* develop a form of FCS. GPIHBP1 is required for transport and tethering of LPL to the endothelial luminal surface. Homozygosity for mutations in *GPIHBP1* that interfere with its synthesis or folding causes FCS. Autoantibodies to GPIHBP1 have also been reported to cause severe hyperchylomicronemia. Finally, *LMF1* is required for appropriate processing and folding of LPL, and biallelic loss-of-function mutations can cause FCS.

FCS can present in childhood or adulthood with severe abdominal pain due to acute pancreatitis. In this setting, the diagnosis should be suspected if a fasting TG level is >500 mg/dL. Eruptive xanthomas, which are small, yellowish-white papules, may appear in clusters on the back, buttocks, and extensor surfaces of the arms and legs. On funduscopic examination, the retinal blood vessels may be opalescent (lipemia retinalis). Hepatosplenomegaly is sometimes noted as a result of uptake of circulating chylomicrons by reticuloendothelial cells in the liver and spleen. Premature ASCVD is not generally a feature of FCS.

The diagnosis of FCS is a clinical diagnosis based on persistence and severity of HTG, with a history of acute pancreatitis or eruptive xanthomas increasing the suspicion. While LPL activity can be measured in "postheparin plasma" obtained after an IV heparin injection to release the endothelial-bound LPL, this assay is not widely available. Genetic testing of a panel of candidate FCS genes can be used to confirm the diagnosis but is not required for making the clinical diagnosis.

Because of the risk of pancreatitis, it is important to consider the diagnosis and institute therapeutic interventions in FCS. The goal is to prevent pancreatitis by reducing fasting TG levels to <500 mg/dL. Consultation with a registered dietician familiar with this disorder is essential. Dietary fat intake should be markedly restricted (to as little as 15 g/d), often with fat-soluble vitamin supplementation. Strict adherence to dietary fat restriction can be successful at controlling the chylomicronemia; fish oils or fibrates (such as fenofibrate) may be tried but are unlikely to be effective. Promising therapeutic approaches include the silencing of *APOC3* or ANGPTL3 in the liver. In patients with *APOC2* deficiency, apoC-II can be provided exogenously by infusing fresh-frozen plasma to resolve the chylomicronemia in the setting of severe acute pancreatitis. Management of patients with FCS is particularly challenging during pregnancy when VLDL production is increased.

FAMILIAL PARTIAL LIPODYSTROPHY (FPLD) FPLD is a genetic condition in which the generation of adipose tissue in certain fat depots is impaired and in others is excessive. FPLD is an underrecognized monogenic cause of severe HTG, which is likely due to both increased lipid synthesis and VLDL production, as well as reduced LPL-mediated clearance of TRLs. FPLD is typically a dominantly inherited disorder caused by mutations in several different genes, including lamin A/C (*LMNA*), PPARγ (*PPARG*), perilipin (*PLIN1*), *AKT2*, and *ADRA2A* (Table 419-2). FPLD is characterized by loss of subcutaneous fat in the extremities and buttocks, often accompanied by increased visceral fat. Because of the reduced or absent subcutaneous fat in the arms and legs, patients are often described as having a "muscular" appearance. In addition to severe HTG, FPLD patients usually have insulin resistance, often quite severe, accompanied by type 2 diabetes and hepatosteatosis. Pancreatitis secondary to HTG can be a complication; in addition, ASCVD risk is substantially increased in FPLD patients. The diagnosis of FPLD is a clinical diagnosis based on the constellation of metabolic findings accompanied by the distinctive distribution of adipose tissue. Genetic testing of a panel of candidate FPLD genes can be used to confirm the diagnosis but is not required for making the clinical diagnosis, and a negative result does not rule out the diagnosis. Because FPLD is a dominant disorder, the finding of a causal mutation should lead to family-based screening.

The dyslipidemia of FPLD can be difficult to manage clinically. Patients should be treated aggressively not only to reduce TG levels but also with statins and, if necessary, additional LDL-lowering therapies to reduce atherogenic lipoproteins. The insulin-resistant diabetes often requires aggressive management as well. Some patients have progression of fatty liver disease to metabolic-associated steatohepatitis, fibrosis, and cirrhosis. A different group of very rare patients have congenital generalized lipodystrophy, a recessive disorder caused by mutations in the *AGPAT2* and *BSCL2* genes. These patients have nearly complete absence of subcutaneous fat, accompanied by profound

TABLE 419-2 Primary Dyslipoproteinemias Caused by Known Single-Gene Mutations

GENETIC DISORDER	GENES MUTATED	LIPOPROTEINS AFFECTED	CLINICAL FINDINGS	GENETIC TRANSMISSION	ESTIMATED PREVALENCE
Severe Hypertriglyceridemia					
Familial chylomicronemia syndrome (FCS)	Biallelic LoF mutations in: *LPL, APOC2, APOA5, GPIHBP1, LMF1*	Elevated: Chylomicrons, VLDL Reduced: HDL	Pancreatitis, eruptive xanthomas, hepatosplenomegaly	AR	~1/200,000–300,000
Familial partial lipodystrophy (FPLD)	Heterozygous LoF mutations in: *LMNA, PPARG, PLIN1, AKT2, ADRA2A*	Elevated: Chylomicrons, VLDL, LDL Reduced: HDL	Insulin resistance, fatty liver disease, pancreatitis, central obesity, lack of subcutaneous adipose in extremities	AD	<1/1,000,000
Hypercholesterolemia					
Familial hypercholesterolemia (FH)	Heterozygous LoF mutations in *LDLR*	Elevated: LDL	Tendon xanthomas, premature atherosclerotic cardiovascular disease (ASCVD)	AD	~1/250
Familial defective apoB-100 (FDB)	Heterozygous LoF receptor binding region mutations in *APOB*	Elevated: LDL	Tendon xanthomas, premature ASCVD	AD	~1/1500
Autosomal dominant hypercholesterolemia (ADH), type 3	Heterozygous GoF mutations in *PCSK9*	Elevated: LDL	Tendon xanthomas, premature ASCVD	AD	<1/1,000,000
Autosomal recessive hypercholesterolemia (ARH)	Biallelic LoF mutations in *LDLRAP1*	Elevated: LDL	Tendon xanthomas, premature ASCVD	AR	<1/1,000,000
Sitosterolemia	Biallelic LoF mutations in *ABCG5, ABCG8*	Elevated: LDL	Tendon xanthomas, premature ASCVD	AR	<1/1,000,000
Lysosomal acid lipase deficiency	Biallelic LoF mutations in *LIPA*	Elevated: LDL Reduced: HDL	Fatty liver disease, micronodular cirrhosis	AR	<1/1,000,000
Mixed Dyslipidemia					
Familial dysbetalipoproteinemia (FDBL)	Biallelic carriers of the *APOE2* variant	Elevated: Chylomicron remnants, IDL	Palmar and tuberoeruptive xanthomas, premature ASCVD	AR	~1/10,000
Hepatic lipase deficiency	Biallelic LoF mutations in *LIPC*	Elevated: Chylomicron remnants, IDL, HDL	Premature ASCVD	AR	<1/1,000,000
Familial Hypolipidemia Syndromes					
Abetalipoproteinemia	Biallelic LoF mutations in *MTTP*	Absent: LDL Reduced: TG, HDL	Spinocerebellar degeneration, retinal degeneration	AR	<1/1,000,000
Familial hypobetalipoproteinemia	Heterozygous truncating mutations in *APOB*	Reduced: LDL	Fatty liver, reduced risk of ASCVD	AD	<1/1,000,000
Familial PCSK9 deficiency	Heterozygous LoF mutations in *PCSK9*	Reduced: LDL	Reduced risk of ASCVD	AD	~1/1,000
Familial combined hypolipidemia	Heterozygous LoF mutations in *ANGPTL3*	Reduced: TG, LDL, HDL	Reduced risk of ASCVD	AD	<1/1,000,000
Primary Low HDL Cholesterol Syndromes					
ApoA-I deletions/ mutations	Heterozygous structural mutations in *APOA1*	Reduced: HDL	Variable depending on mutation: premature ASCVD, systemic amyloidosis	AD	<1/1,000,000
Tangier disease	Biallelic LoF mutations in *ABCA1*	Nearly absent: HDL Reduced: LDL Elevated: TG	Peripheral neuropathy, hepatosplenomegaly	AR	<1/1,000,000
Familial LCAT deficiency (FLD); fish eye disease (FED)	Biallelic LoF mutations in *LCAT*	Markedly reduced: HDL	Corneal opacities (both FLD and FED), progressive chronic kidney disease (FLD only)	AR	<1/1,000,000

Abbreviations: AD, autosomal dominant; apo, apolipoprotein; AR, autosomal recessive; ARH, autosomal recessive hypercholesterolemia; CHD, coronary heart disease; GoF, gain of function; HDL, high-density lipoprotein; IDL, intermediate-density lipoprotein; LCAT, lecithin-cholesterol acyltransferase; LDL, low-density lipoprotein; LoF, loss of function; LPL, lipoprotein lipase; PVD, peripheral vascular disease; TG, triglyceride; VLDL, very-low density lipoprotein.

leptin deficiency, insulin resistance, severe HTG, and accumulation of TGs in multiple tissues including the liver. Patients with generalized lipodystrophy can be effectively treated with recombinant leptin administration, which often manages the multiple metabolic issues in these patients.

Multifactorial Severe Hypertriglyceridemia Most patients with severe HTG do not have a single-gene mutation but instead have a multifactorial etiology that includes genetics and environment. The prevalence of this phenotype is ~1 in 1000. HTG often runs in families, and the term *familial HTG* has been employed; however, except for the

TABLE 419-3 Secondary Causes of Altered Lipid and Lipoprotein Levels

	LDL-C		HDL-C		LP(a) ELEVATED
TG ELEVATED	**ELEVATED**	**REDUCED**	**ELEVATED**	**REDUCED**	
High-carbohydrate diet	Hypothyroidism	Vegan diet	High-fat diet	Hypertriglyceridemia	Chronic kidney disease
Alcohol	Cholestasis	Malabsorption	Alcohol	Vegan diet	Nephrotic syndrome
Obesity	Nephrotic syndrome	Malnutrition	Exercise	Malabsorption	Inflammation
Insulin resistance	Cushing's syndrome	Severe liver disease	Drugs: estrogen, phenytoin	Malnutrition	Menopause
Type 2 diabetes	Acute intermittent porphyria	Gaucher's disease		Sedentary lifestyle	Orchidectomy
Lipodystrophy	Drugs: corticosteroids, cyclosporin, sirolimus, carbamazepine	Chronic infectious disease		Smoking	Hypothyroidism
Chronic kidney disease		Hyperthyroidism		Obesity	Acromegaly
Nephrotic syndrome				Gaucher's disease	Drugs: growth hormone, isotretinoin
Viral hepatitis				LAL deficiency	
Sepsis				Drugs: anabolic steroids, testosterone, beta blockers	
Cushing's syndrome					
Acromegaly					
Glycogen storage disease					
Pregnancy					
Drugs: estrogen, glucocorticoids, isotretinoin, bexarotene, other retinoids, beta blockers, bile acid binding resins					

Abbreviations: HDL-C, high-density lipoprotein cholesterol; LAL, lysosomal acid lipase; LDL-C, low-density lipoprotein cholesterol; Lp(a), lipoprotein(a); TG, triglyceride.

genes in which mutations cause FCS or FPLD, reviewed above, no other classic Mendelian causes of HTG have been identified to date. Instead, extensive human genetic studies have clearly established a polygenic basis to this phenotype that consists of two categories: (1) rare heterozygous variants in the five genes discussed earlier that cause FCS in the homozygous state, and (2) a high burden of common variants that have small individual effects at raising TGs. Patients who inherit some combination of rare and common TG-raising alleles often have environmental factors that exacerbate their HTG. These "secondary" factors are reviewed in detail below, but the quantitatively most important factors promoting HTG include obesity, type 2 diabetes, insulin resistance, high-carbohydrate diet, and alcohol use. Multifactorial HTG is characterized by elevated fasting TGs but average to below average LDL-C levels and low HDL-C levels; apoB levels are not generally elevated. This condition is not generally associated with a significantly increased risk of ASCVD. However, if the HTG is exacerbated by environmental factors, medical conditions, or drugs, the TGs can rise to a level at which acute pancreatitis is a risk. Indeed, management of patients with this condition is mostly focused on reduction of TGs to prevent pancreatitis. It is important to consider and rule out secondary causes of the HTG. Patients who are at high risk for ASCVD due to other risk factors should be treated with statin therapy. In patients who are otherwise not at high risk for ASCVD, lipid-lowering drug therapy can frequently be avoided with appropriate dietary and lifestyle changes. Patients with plasma TG levels >500 mg/dL after a trial of diet and exercise should be considered for drug therapy with a fibrate or fish oil to reduce TGs in order to prevent pancreatitis. These patients should also be carefully evaluated for ASCVD risk and may be candidates for statin therapy to further reduce cholesterol and cardiovascular risk.

HYPERCHOLESTEROLEMIA (ELEVATED LDL-C)

Elevated LDL-C is common and is medically important because it is associated with risk of premature ASCVD. Elevated LDL-C is often caused by impaired uptake of LDL by the liver. As discussed above, the LDL receptor is the major receptor responsible for uptake of LDL, and most causes of elevated LDL-C converge on reduced expression or activity of the LDL receptor in the liver. One major environmental factor that reduces LDL receptor activity is a diet high in saturated and trans fats. Other medical conditions that reduce LDL receptor activity include hypothyroidism and estrogen deficiency. The diagnosis of familial hypercholesterolemia involving several genes that influence LDL clearance should be considered in patients with LDL-C levels >190 mg/dL (Table 419-2). However, the majority of patients with elevated LDL-C have a polygenic predisposition exacerbated by secondary factors like a diet high in saturated and trans fats.

Primary (Genetic) Causes of Elevated LDL-C • FAMILIAL HYPERCHOLESTEROLEMIA (FH) FH is an autosomal dominant disorder characterized by elevated plasma levels of LDL-C usually with relatively normal TG levels. FH is caused by mutations that lead to reduced function of the LDL receptor, with the most common being mutations in the *LDLR* gene itself. The reduction in LDL receptor activity in the liver results in a reduced rate of clearance of LDL from the circulation. The plasma level of LDL increases to a level such that the rate of LDL production equals the rate of LDL clearance by residual LDL receptor as well as non-LDL receptor mechanisms. Individuals with two mutated *LDLR* alleles (homozygotes or compound heterozygotes) have much higher LDL-C levels than those with one mutant allele, causing a condition known as *homozygous FH*.

Although mutations in *LDLR* are the most common cause of FH (and originally the term *FH* was used specifically for patients with *LDLR* mutations), mutations in at least two other genes, *APOB* and *PCSK9*, can also cause FH. ApoB-100 is the critical structural protein in LDL and contains a domain that serves as the ligand for binding to the LDL receptor. Mutations in the LDL receptor–binding domain of apoB-100 reduce the affinity of apoB/LDL binding to the LDL receptor, such that LDL is removed from the circulation at a reduced rate. This condition has also been termed *familial defective apoB* (FDB). Of note, truncating mutations in *APOB* cause *low* LDL-C levels (see below). The proprotein convertase subtilisin/kexin type 9 (PCSK9) is a secreted protein that binds to the LDL receptor and targets it for lysosomal degradation. Normally, after LDL binds to the LDL receptor, it is internalized along with the receptor, and in the low pH of the endosome, the LDL receptor dissociates from the LDL and recycles to the cell surface. When circulating PCSK9 binds the receptor, the complex is internalized and the receptor is directed to the lysosome, rather than to the cell surface, reducing the number of active LDL receptors. *Gain-of-function* mutations in *PCSK9* that enhance the activity of PCSK9 cause a form of FH, also known as ADH type 3. Of note, loss-of-function mutations in *PCSK9* reduce LDL-C levels (see below).

The population frequency of heterozygous FH was originally estimated to be 1 in 500 individuals, but recent data suggest it may be as high as ~1 in 250 individuals, making it one of the most common single-gene disorders in humans. FH has a much higher prevalence in certain founder populations, such as South African Afrikaners, Christian Lebanese, French Canadians, and Lancaster County Amish.

Heterozygous FH is characterized by elevated plasma levels of LDL-C (usually >190 mg/dL) and relatively normal levels of TGs. Patients with FH have hypercholesterolemia from birth, and FH diagnosis is often based on detection of hypercholesterolemia on routine lipid screening; this serves as the basis for the recommendation to screen children between the ages of 9 and 11. A family history of hypercholesterolemia or premature ASCVD should prompt targeted screening. Inheritance of FH is dominant, meaning that the condition is inherited from one parent, and ~50% of the patient's siblings and children can be expected to have FH. For this reason, family-based "cascade screening" can be very effective in identifying additional persons with FH. Physical findings in some, but not all, patients with FH may include corneal arcus and/or tendon xanthomas, particularly involving the dorsum of the hands and the Achilles tendons. Untreated heterozygous FH is associated with a markedly increased risk of cardiovascular disease; untreated men with heterozygous FH have an ~50% chance of having a myocardial infarction before age 60 years, and women with heterozygous FH are at substantially increased risk as well. The age of onset of cardiovascular disease is highly variable and depends on the specific molecular defect, the level of LDL-C, and coexisting cardiovascular risk factors.

The diagnosis of FH is generally a clinical diagnosis based on hypercholesterolemia with LDL-C >190 mg/dL in the absence of a secondary etiology and ideally with a family history of hypercholesterolemia and/or premature ASCVD. Secondary causes of significant hypercholesterolemia such as hypothyroidism, nephrotic syndrome, and obstructive liver disease should be excluded. Sequencing of an FH gene panel (*LDLR*, *APOB*, *PCSK9*) to confirm the diagnosis is widely available and worthy of consideration; persons with molecularly confirmed FH are at higher risk of ASCVD and therefore may benefit from more aggressive treatment, and the finding of a specific causal variant has implications for family-based cascade screening.

FH patients should be actively treated to lower plasma levels of LDL-C, preferably starting in childhood. Initiation of a diet low in saturated and trans fats is recommended, but heterozygous FH patients almost always require pharmacologic therapy for effective control of their LDL-C levels. Statins are the initial drug class of choice, and usually "high-intensity" statin therapy is needed. Many FH patients cannot achieve adequate control of their LDL-C levels even with high-intensity statin therapy, and a cholesterol absorption inhibitor (ezetimibe), a PCSK9 inhibitor, an ACL inhibitor (bempedoic acid), and a bile acid sequestrant are other classes of drugs that can be added to statins **(Table 419-4)**. Some patients with severe heterozygous FH cannot be adequately managed using existing therapies and are candidates for LDL apheresis, a physical method of purging the blood of LDL in which the LDL particles are selectively removed from the circulation. Other novel approaches for these patients are under development.

Homozygous FH (HoFH) is caused by loss-of-function mutations in both alleles of the LDL receptor or double heterozygosity for mutations in two FH genes. Patients with HoFH have been classified into those with virtually no detectable LDL receptor activity (*receptor negative*) and patients with markedly reduced but detectable LDL receptor activity (*receptor defective*). Untreated LDL-C levels in patients with HoFH range from ~400 to >1000 mg/dL, with receptor-defective patients at the lower end and receptor-negative patients at the higher end of the range. TGs are usually relatively normal. Some patients with HoFH, particularly receptor-negative patients, present in childhood with cutaneous planar xanthomas on the hands, wrists, elbows, knees, heels, or buttocks. The devastating consequence of HoFH is accelerated ASCVD, which often presents in childhood or early adulthood. Atherosclerosis often develops first in the aortic root, where it can cause aortic valvular or supravalvular stenosis, and typically extends into the coronary ostia, which become stenotic. Symptoms can be atypical, and sudden death is not uncommon. Untreated, receptor-negative patients with HoFH rarely survive beyond the second decade; patients with receptor-defective LDL receptor defects have a better prognosis but almost invariably develop clinically apparent atherosclerotic vascular disease by age 30 and often much sooner.

HoFH should be suspected in a child or young adult with LDL >400 mg/dL without secondary cause. Cutaneous xanthomas, evidence of ASCVD, and/or hypercholesterolemia in both parents all are supportive of the diagnosis. While the diagnosis is usually made on clinical grounds, genetic testing should be performed to identify specific causal variants. Patients with HoFH must be treated aggressively to delay the onset and progression of cardiovascular disease (CVD). Although receptor-negative patients have no response to statins and PCSK9 inhibitors, receptor-defective patients can have modest responses to these medicines, and they should be tried in patients with HoFH. Two drugs that reduce the hepatic production of VLDL and thus LDL, a small-molecule inhibitor of MTP and an antisense oligonucleotide to apoB, and an antibody that inhibits ANGPLT3 are approved for the treatment of patients with HoFH and should be considered in HoFH patients. LDL apheresis should be considered in HoFH patients who have persistently elevated LDL-C levels despite drug therapy. Liver transplantation is effective in decreasing plasma LDL-C levels in this disorder and is sometimes used as a last resort. Liver-directed gene therapy is under development for HoFH, as are other new therapeutic approaches intended to address the remaining unmet medical need.

FH is an autosomal dominant disorder. There are a few rare conditions that cause an FH-like phenotype in an autosomal recessive manner and should be considered in patients with severe hypercholesterolemia who do not report a family history of hypercholesterolemia or premature CHD.

AUTOSOMAL RECESSIVE HYPERCHOLESTEROLEMIA (ARH) ARH is a very rare autosomal recessive disorder that was originally reported in individuals of Sardinian descent. The disease is caused by mutations in the gene *LDLRAP1* encoding the protein LDLR adaptor protein (also called the ARH protein), which is required for LDL receptor–mediated endocytosis in the liver. LDLRAP1 binds to the cytoplasmic domain of the LDL receptor and links the receptor to the endocytic machinery. In the absence of LDLRAP1, LDL binds to the extracellular domain of the LDL receptor, but the lipoprotein-receptor complex fails to be internalized. ARH, like HoFH, is characterized by hypercholesterolemia, tendon xanthomas, and premature coronary artery disease (CAD). The levels of plasma LDL-C tend to be intermediate between the levels present in FH homozygotes and FH heterozygotes, and CAD is not usually symptomatic until the third decade. LDL receptor function in cultured fibroblasts is normal or only modestly reduced in ARH, whereas LDL receptor function in the liver is negligible. Unlike FH homozygotes, the hyperlipidemia responds to treatment with statins, but these patients often require additional therapy to lower plasma LDL-C to acceptable levels.

SITOSTEROLEMIA Sitosterolemia is a rare autosomal recessive disease that is caused by biallelic loss-of-function mutations in either of two members of the ATP-binding cassette (ABC) half transporter family, *ABCG5* and *ABCG8*. These genes are expressed in both enterocytes and hepatocytes. The proteins heterodimerize to form a functional complex that transports plant sterols such as sitosterol and campesterol, and animal sterols, predominantly cholesterol, across the apical biliary membrane of hepatocytes into the bile and across the apical luminal membrane of enterocytes into the gut lumen, thus reducing their (re)absorption and promoting their excretion. In normal individuals, <5% of dietary plant sterols are absorbed by the proximal small intestine. The small amounts of plant sterols that enter the circulation are preferentially excreted into the bile, and thus, levels of plant sterols are kept very low in tissues. In sitosterolemia, the intestinal absorption of sterols is increased and biliary and fecal excretion of the sterols is reduced, resulting in increased plasma and tissue levels of both plant sterols and cholesterol. The increase in hepatic sterol levels results in transcriptional suppression of the expression of the LDL receptor, resulting in reduced uptake of LDL and substantially increased LDL-C levels. In addition to the clinical picture of severe hypercholesterolemia, often accompanied by tendon xanthomas and premature ASCVD, these patients also have anisocytosis and poikilocytosis of erythrocytes and megathrombocytes due to the incorporation of plant sterols into cell membranes. Episodes of hemolysis and splenomegaly are a distinctive clinical feature of this disease compared

TABLE 419-4 Drugs Used to Treat Dyslipidemia

DRUG	MAJOR INDICATIONS	STARTING DOSE	MAXIMAL DOSE	MECHANISM	ADVERSE EFFECTS
LDL-Lowering Drugs					
HMG-CoA reductase inhibitors (statins)	Elevated LDL-C; increased CV risk			↓ Inhibition of cholesterol synthesis → ↑ Hepatic LDL receptors	Myalgias and myopathy, ↑ transaminases, ↑ diabetes risk
Lovastatin		20–40 mg daily	80 mg daily		
Pravastatin		40–80 mg daily	80 mg daily		
Simvastatin		20–40 mg daily	80 mg daily		
Fluvastatin		20–40 mg daily	80 mg daily		
Atorvastatin		20–40 mg daily	80 mg daily		
Rosuvastatin		5–20 mg daily	40 mg daily		
Pitavastatin		1–2 mg daily	4 mg daily		
Cholesterol absorption inhibitor	Elevated LDL-C			↓ Cholesterol absorption→ ↑ LDL receptors	Elevated transaminases
Ezetimibe		10 mg daily	10 mg daily		
Bile acid sequestrants	Elevated LDL-C			↑ Bile acid excretion → ↑ LDL receptors	Bloating, constipation, elevated triglycerides
Cholestyramine		4 g daily	32 g daily		
Colestipol		5 g daily	40 g daily		
Colesevelam		3750 mg daily	4375 mg daily		
PCSK9 inhibitors	Elevated LDL-C	140 mg SC every 2 weeks	420 mg SC every 1 month (HoFH)	↓ PCSK9 activity due to Ab inhibition → ↑ LDL receptors	Injection site reactions
Evolocumab (Ab)		75 mg SC every 2 weeks			
Alirocumab (Ab)			150 mg SC every 2 weeks		
Inclisiran (siRNA)		300 mg SC every 6 months	300 mg SC every 6 months	↓ PCSK9 synthesis due to siRNA silencing → ↑ LDL receptors	Injection site reactions
ATP citrate lyase inhibitor	Elevated LDL-C	180 mg daily	180 mg daily	↓ Inhibition of cholesterol synthesis → ↑ LDL receptors	↑ uric acid and gout ↑ cholelithiasis
Bempedoic acid					
MTP inhibitor	HoFH	5 mg daily	60 mg daily	MTP inhibition → ↓ VLDL assembly and secretion	Nausea, diarrhea, increased hepatic fat
Lomitapide					
ApoB inhibitor (ASO)	HoFH	200 mg SC weekly	200 mg SC weekly	↓ ApoB synthesis due to ASO silencing → ↓ ApoB/VLDL secretion	Injection site reactions, flu-like symptoms, increased hepatic fat
Mipomersen					
ANGPTL3 inhibitor (Ab)	HoFH	15 mg/kg IV q 4 weeks	15 mg/kg IV q 4 weeks	↓ ANGPTL3 activity due to Ab inhibition → ↑ LPL and EL activity, ↑ LDL catabolism	Reduced HDL-C levels
Evinacumab					
TG-Lowering Drugs					
Fibric acid derivatives (fibrates)	Elevated TG	600 mg bid	600 mg bid	↑ LPL, ↓ VLDL synthesis	Dyspepsia, myalgia, cholelithiasis, elevated transaminases
Gemfibrozil		40–160 mg daily depending on product	40–160 mg daily depending on product		
Fenofibrate					
Omega-3 fatty acids	Elevated TG	4 g daily	4 g daily	↑ TG catabolism	Dyspepsia
Acid ethyl esters					
Icosapent ethyl		4 g daily	4 g daily		

Abbreviations: Ab, antibody; GI, gastrointestinal; HDL-C, high-density lipoprotein cholesterol; HoFH, homozygous familial hypercholesterolemia; LDL, low-density lipoprotein; LDL-C, LDL cholesterol; LPL, lipoprotein lipase; TG, triglyceride; VLDL, very-low-density lipoprotein.

to other genetic forms of hypercholesterolemia and can be a clue to the diagnosis. Sitosterolemia should be suspected in a patient with severe hypercholesterolemia without a family history of such or who fails to respond to statin therapy. Sitosterolemia can be diagnosed by a laboratory finding of a substantial increase in plasma sitosterol and/or other plant sterols and should be confirmed by gene sequencing of *ABCG5* and *ABCG8*. It is important to make the diagnosis, because diet, bile acid sequestrants, and cholesterol-absorption inhibitors are the most effective agents to reduce LDL-C and plasma plant sterol levels in these patients. Of note, heterozygosity for mutations in *ABCG5* or *ABCG8* is now recognized to cause a moderate form of hypercholesterolemia.

LYSOSOMAL ACID LIPASE DEFICIENCY (LALD) LALD, also known as *cholesteryl ester storage disease*, is an autosomal recessive disorder caused by loss-of-function variants in both alleles of the gene *LIPA* encoding the enzyme lysosomal acid lipase (LAL). LAL is responsible for hydrolyzing neutral lipids, particularly TGs and CEs, after delivery to the lysosome by cell surface receptors such as the LDL receptor. It is particularly important in the liver, which clears large amounts of lipoproteins from the circulation. LALD is characterized by elevated LDL-C, usually in association with low HDL-C and with variably elevated TG levels, together with progressive fatty liver ultimately leading to hepatic fibrosis. Genetic deficiency of LAL results in accumulation of neutral lipid in the hepatocytes, leading to hepatosplenomegaly, microvesicular steatosis, and ultimately fibrosis and end-stage liver disease. The most severe form of this disorder, Wolman's disease, presents in infancy and is rapidly fatal. The etiology of the elevated LDL-C levels is primarily due to impaired LDL receptor–mediated clearance of LDL. LALD should be suspected in nonobese patients with elevated LDL-C, low HDL-C, and evidence of fatty liver in the absence of overt insulin resistance. The diagnosis can be made with a dried blood spot assay of LAL activity and confirmed by DNA genotyping for the most common mutation, followed if necessary by sequencing of the gene to find the second mutation. Liver biopsy is required to assess the degree of inflammation and fibrosis. LALD is underdiagnosed; it is critically important to suspect it and make the diagnosis because enzyme replacement therapy with sebelipase alfa is now available and is highly effective in treating this condition.

The above conditions primarily cause elevations in LDL due to impaired catabolism of LDL from the blood. There are a few forms of primary dyslipidemia that impair the catabolism of "remnant" TRLs (after their processing by LPL) and therefore cause elevations in both cholesterol and TGs due to remnant accumulation.

Multifactorial Hypercholesterolemia Most patients with elevated LDL-C do not have a single-gene disorder, as described above, but instead have a multifactorial etiology that includes genetics and environment. Genetic variation contributes substantially to elevated LDL-C levels in the general population. It has been estimated that at least 50% of variation in LDL-C is genetically determined. Many patients with elevated LDL-C have *polygenic hypercholesterolemia* due to multiple common genetic variants exerting modest LDL-raising effects. Individuals at the tail of the highest burden of polygenic risk score for LDL-C often have LDL-C levels that are similar to those with FH. In patients who are genetically predisposed to higher LDL-C levels, diet plays a key exacerbating role; indeed, increased saturated and trans fats in the diet shift the entire distribution of LDL-C levels in the population to the right. As described in more detail below, patients with elevated LDL-C should be carefully assessed for their risk of ASCVD and managed with lifestyle modification and LDL-lowering medications as needed to reduce LDL-C and risk of ASCVD.

■ MIXED HYPERLIPIDEMIA (ELEVATED TG AND LDL-C)

Mixed hyperlipidemia can be defined as fasting TGs >150 mg/dL and evidence of elevated cholesterol-containing lipoproteins (such as LDL-C >130 mg/dL or non-HDL-C >160 mg/dL). It is one of the most common types of lipid disorders seen in clinical practice, due both to genetic predisposition and influence of medical conditions and environmental factors (see below). It is generally associated with elevated risk of ASCVD, and therefore, patients with mixed hyperlipidemia should be carefully evaluated and managed to reduce this risk.

Primary (Genetic) Causes of Mixed Hyperlipidemia • **FAMILIAL DYSBETALIPOPROTEINEMIA (FDBL)** FDBL (also known as *type III hyperlipoproteinemia*) is a recessive disorder characterized by a mixed hyperlipidemia due to the accumulation of remnant lipoprotein particles (chylomicron remnants and VLDL remnants, or IDL). ApoE is present in multiple copies on chylomicron remnants and IDL and mediates their removal via hepatic lipoprotein receptors (Fig. 419-2). The *APOE* gene is polymorphic in sequence, resulting in the expression of three common isoforms: apoE3, which is the most common (~78% global allele frequency [AF]), apoE4 (~14% global AF), and apoE2 (~8% global AF). The apoE4 allele, which has an arginine instead of a cysteine at position 112, is widely known for being the major genetic risk factor for Alzheimer's disease. It is associated with slightly higher LDL-C levels and increased ASCVD risk but is not associated with FDBL. The apoE2 allele, which has a cysteine at position 158 instead of an arginine, is the cause of FDBL when present on both alleles. ApoE2 has a lower affinity for the LDL receptor; therefore, chylomicron remnants and IDL containing apoE2 are removed from plasma at a slower rate, leading to their accumulation in blood.

Approximately 0.5% of the general population are apoE2/E2 homozygotes, but only a small minority of these individuals actually develop hyperlipidemia characteristic of FDBL (which has a prevalence of ~1 in 10,000). Thus, an additional, sometimes identifiable, factor precipitates the development of overt dysbetalipoproteinemia in apoE2/E2 homozygotes. The most common precipitating factors are a high-fat diet, sedentary lifestyle, obesity, alcohol use, menopause, diabetes mellitus, hypothyroidism, renal disease, HIV infection, or certain drugs. Certain dominant-negative mutations in apoE can cause a dominant form of FDBL where the hyperlipidemia is fully manifest in the heterozygous state, but these mutations are very rare.

Patients with FDBL usually present in adulthood with hyperlipidemia, xanthomas, or premature coronary or peripheral vascular disease. In FDBL, in contrast to other disorders of elevated TGs, the plasma levels of cholesterol and TG are often elevated to a similar degree, and the level of HDL-C is usually normal. Two distinctive types of xanthomas, tuberoeruptive and palmar, are seen in FDBL patients. Tuberoeruptive xanthomas begin as clusters of small papules on the elbows, knees, or buttocks and can grow to the size of small grapes. Palmar xanthomas (alternatively called *xanthomata striata palmaris*) are orange-yellow discolorations of the creases in the palms and wrists. Both of these xanthoma types are virtually pathognomonic for FDBL. Subjects with FDBL have premature ASCVD and tend to have more peripheral vascular disease than is typically seen in FH.

The definitive diagnosis of FDBL can be made either by the documentation of very high levels of remnant lipoproteins or by identification of the apoE2/E2 genotype. A variety of methods are used to identify remnant lipoproteins in the plasma, including "β-quantification" by ultracentrifugation (ratio of directly measured VLDL cholesterol to total plasma TG >0.30), lipoprotein electrophoresis (broad β band), or nuclear magnetic resonance lipoprotein profiling. The Friedewald formula for calculation of LDL-C is not valid in FDBL because the VLDL particles are depleted in TG and enriched in cholesterol. The plasma levels of LDL-C are actually low in this disorder due to defective metabolism of VLDL to LDL. DNA-based apoE genotyping can be performed to confirm homozygosity for apoE2, which is diagnostic for FDBL. However, absence of the apoE2/E2 genotype does not strictly rule out the diagnosis of FDBL, because other mutations in apoE can (rarely) cause this condition.

Because FDBL is associated with increased risk of premature ASCVD, it should be treated aggressively. Other metabolic conditions that can exacerbate the hyperlipidemia (see above) should be managed. Patients with FDBL are typically diet-responsive and can respond favorably to low-cholesterol, low-fat diets and weight reduction. Alcohol intake should be curtailed. Pharmacologic therapy is often required, and statins are the first line in management. In the event of statin intolerance or insufficient control of hyperlipidemia, cholesterol absorption inhibitors, PCSK9 inhibitors, and fibrates are also effective in the treatment of FDBL.

HEPATIC LIPASE DEFICIENCY Hepatic lipase (HL; gene name *LIPC*) is a member of the same gene family as LPL and hydrolyzes TGs and phospholipids in remnant lipoproteins and HDL. Hydrolysis of lipids in remnant particles by HL contributes to their hepatic uptake via an apoE-mediated process. HL deficiency is a very rare autosomal recessive disorder caused by biallelic loss-of-function mutations in *LIPC*. It is characterized by elevated plasma levels of cholesterol and TGs (mixed hyperlipidemia) due to the accumulation of lipoprotein remnants, accompanied by elevated plasma level of HDL-C. The diagnosis is confirmed by confirmation of pathogenic mutations in both alleles of *LIPC*. Due to the small number of patients with HL deficiency, the association of this genetic defect with ASCVD is not entirely clear, although anecdotally, patients with HL deficiency who have premature CVD have been described. As with FDBL, statin therapy is recommended to reduce remnant lipoproteins and cardiovascular risk.

FAMILIAL COMBINED HYPERLIPIDEMIA (FCHL) FCHL is one of the most common familial lipid disorders; it is estimated to occur in ~1 in 100–200 individuals. FCHL is characterized by elevations in plasma levels of TGs (VLDL) and LDL-C (including especially a small dense form of LDL) and reduced plasma levels of HDL-C. This disorder is an important contributor to premature CHD; ~20% of patients who develop CHD under age 60 have FCHL. FCHL can manifest in childhood but is usually not fully expressed until adulthood. The disease clusters in families, and affected family members typically have one of three possible phenotypes: (1) elevated plasma levels of LDL-C, (2) elevated plasma levels of TGs due to elevation in VLDL, or (3) elevated plasma levels of both LDL-C and TG. The lipoprotein profile can switch among these three phenotypes in the same individual over time and may depend on factors such as diet, exercise, weight, and insulin sensitivity. Patients with FCHL have substantially elevated plasma levels of apoB, often disproportionately high relative to the plasma LDL-C concentration, indicating the presence of small dense LDL particles, which are characteristic of this syndrome.

Individuals with this phenotype generally share the same metabolic defect, namely overproduction of VLDL and apoB by the liver. The molecular etiology of this condition remains poorly understood, and no single gene has been identified in which mutations convincingly cause this disorder in a simple Mendelian fashion. It is likely that defects in a combination of genes can cause the condition, suggesting that a more appropriate term for the disorder might be *polygenic combined hyperlipidemia.*

The presence of a mixed dyslipidemia (plasma TG levels between 150 and 500 mg/dL and total cholesterol levels between 200 and 400 mg/dL, usually with HDL-C levels <40 mg/dL in men and <50 mg/dL in women) and a family history of mixed dyslipidemia and/or premature CHD suggests the diagnosis. Measurement of plasma apoB levels can help support the diagnosis if they are substantially elevated, particularly relative to the LDL-C level. Individuals with this disorder should be treated aggressively due to significantly increased risk of premature CHD, often disproportionate to the LDL-C level. Decreased dietary intake of simple carbohydrates, increased aerobic exercise, and weight loss can all have beneficial effects on the lipid profile. Patients with type 2 diabetes should be aggressively treated to maintain good glucose control. Virtually all patients with FCHL merit lipid-lowering drug therapy to reduce apoB-containing lipoprotein levels and lower the risk of ASCVD. High-intensity statins are first line, but many patients with FCHL require combination therapy that includes ezetimibe, a PCSK9 inhibitor, and/or bempedoic acid.

SECONDARY CONTRIBUTORS TO ELEVATED LEVELS OF APOB-CONTAINING LIPOPROTEINS

There are many "secondary" factors that contribute to dyslipidemia (Table 419-3), often acting in concert with polygenic predisposition as reviewed above. Some primarily affect TGs, some primarily affect LDL-C, and some influence both, with a great deal of variability. Here the major secondary contributors are reviewed.

Secondary Factors That Primarily Elevate TG Levels • **HIGH-CARBOHYDRATE DIET** Dietary carbohydrates are utilized as a substrate for fatty acid synthesis in the liver. Some of the newly synthesized fatty acids are esterified, forming TGs, and secreted in VLDL. Thus, excessive intake of calories as carbohydrates, which is frequent in Western societies, leads to increased hepatic VLDL-TG secretion and elevated TG levels. Reduction in carbohydrate consumption can have a substantial effect in reducing TG levels, although replacing carbohydrates with saturated fat can elevate LDL-C levels.

OBESITY, INSULIN RESISTANCE, AND TYPE 2 DIABETES **(See also Chaps. 413–415)** Obesity, insulin resistance, and type 2 diabetes mellitus are the most frequent contributors to dyslipidemia, primarily by influencing TGs. The increase in adipocyte mass and accompanying decreased insulin sensitivity associated with obesity have multiple effects on lipid metabolism, with one of the major effects being excessive hepatic VLDL production. More free fatty acids are delivered from the expanded and insulin-resistant adipose tissue to the liver, where they are reesterified in hepatocytes to form TGs, which are packaged into VLDLs for secretion into the circulation. In addition, the increased insulin levels promote increased fatty acid synthesis in the liver. In insulin-resistant patients who progress to type 2 diabetes mellitus, dyslipidemia remains common, even when the patient is under relatively good glycemic control. In addition to increased VLDL production, insulin resistance can also result in decreased LPL activity, resulting in reduced catabolism of chylomicrons and VLDLs and more severe HTG. This may be due in part to the effects of tissue insulin resistance leading to reduced transcription of LPL in skeletal muscle and adipose, as well as to increased production of the LPL inhibitor apoC-III by the liver. This reduction in LPL activity often exacerbates the effects of increased VLDL production and contributes to the dyslipidemia seen in these patients. The dyslipidemia in this setting is almost invariably associated with low HDL-C levels as well. A cluster of metabolic risk factors are often found together, including obesity, insulin resistance, hypertension, high TGs, and low HDL-C (the so-called "metabolic syndrome," **Chap. 420**).

ALCOHOL CONSUMPTION Excessive alcohol consumption inhibits hepatic oxidation of free fatty acids, thus promoting hepatic TG synthesis and VLDL secretion and leading to increased plasma TG levels. Regular alcohol use also raises plasma levels of HDL-C and should be considered in patients with the relatively unusual combination of elevated TGs and normal or elevated HDL-C. A careful history of alcohol use should be taken in patients with elevated TGs. Reduction in alcohol consumption can often have a substantial effect in reducing TG levels.

CHRONIC KIDNEY DISEASE **(See also Chap. 322)** Chronic kidney disease (CKD) is often associated with mild HTG (150–400 mg/dL) due to the accumulation of VLDLs and remnant lipoproteins in the circulation. TG lipolysis and remnant clearance are both reduced in patients with renal failure. Because the risk of ASCVD is increased in CKD, patients should usually be treated with lipid-lowering agents, particularly statins.

ESTROGEN AND OTHER DRUGS Many drugs have an impact on lipid metabolism and can result in significant alterations in the lipoprotein profile (Table 419-3). Estrogens often elevate TG levels, and TG levels can also increase during pregnancy. In women with HTG, plasma TG levels should be monitored when birth control pills or postmenopausal estrogen therapy is initiated and during pregnancy. Use of low-dose preparations of estrogen or the estrogen patch can minimize the effect of exogenous estrogen on lipids. Isotretinoin therapy for acne can cause substantial elevations in TGs, and TG levels should be checked at baseline and after initiation of therapy. Bexarotene therapy for cutaneous T-cell lymphoma often causes substantial increases in TGs, and patients should be monitored accordingly.

Secondary Factors That Elevate LDL-C Levels • **DIET HIGH IN SATURATED AND TRANS FATS** Dietary saturated and trans fats act to downregulate LDL receptor expression in the liver, leading to elevation in LDL-C levels and increased ASCVD risk. A careful dietary history should be taken in individuals with elevated LDL-C with a focus on sources of saturated and trans fats. Reduction in consumption of saturated and trans fats can sometimes have a substantial effect in reducing LDL-C levels and is a cornerstone of the initial nonpharmacologic management of hypercholesterolemia.

HYPOTHYROIDISM **(See also Chap. 394)** Hypothyroidism is the most important medical condition causing elevated LDL-C levels. It causes elevated plasma LDL-C levels due to downregulation of the hepatic LDL receptor, which is normally increased by the action of thyroid hormone. Because hypothyroidism is often subtle and therefore easily overlooked, all patients presenting with elevated plasma levels of LDL-C, especially if there has been an unexplained increase in LDL-C, should be screened for hypothyroidism by measuring thyroid-stimulating hormone (TSH). Thyroid replacement therapy usually reduces LDL-C levels; if not, the patient probably has a primary lipoprotein disorder and may require lipid-lowering drug therapy with a statin.

LIVER DISORDERS **(See also Chap. 347)** Cholestasis is almost invariably associated with hypercholesterolemia due to elevated LDL-C levels and, if severe, particles called Lp-X. A major pathway by which cholesterol is excreted from the body is via secretion into bile, either directly or after conversion to bile acids, and cholestasis blocks this critical excretory pathway. The increase in hepatocellular cholesterol results in downregulation of the LDL receptor, leading to increased plasma LDL-C levels. In severe cholestasis, excess free cholesterol, coupled with phospholipids, is shed into the plasma as a constituent of a lamellar particle called Lp-X. These unusual particles, which are not lipoproteins, lack apoB, and have an aqueous and not neutral lipid core, are rich in free cholesterol, and can deposit in the skin, producing xanthomas sometimes seen in patients with cholestasis. Some liver disorders can affect plasma lipid levels in other ways. Viral hepatitis can increase TGs, and liver failure can result in reduction in plasma cholesterol and TGs.

NEPHROTIC SYNDROME (See also Chap. 322) Nephrotic syndrome is a classic cause of excessive VLDL production leading to elevation in both TGs and LDL-C. The molecular mechanism of VLDL overproduction remains poorly understood but has been attributed to the effects of hypoalbuminemia leading to increased hepatic protein synthesis. Effective treatment of the underlying renal disease may normalize the lipid profile, but many patients with chronic nephrotic syndrome require lipid-lowering drug therapy with statins and sometimes additional drugs.

CUSHING'S SYNDROME (See also Chap. 398) Endogenous glucocorticoid excess in Cushing's syndrome is associated with increased VLDL synthesis and secretion leading to dyslipidemia characterized by HTG and elevated LDL-C. Treatment of the underlying cause is often sufficient to manage the dyslipidemia, but sometimes lipid-lowering drug therapy is needed.

IMMUNOSUPPRESSIVE THERAPY AND CORTICOSTEROIDS Several of the immunosuppressants used after solid organ transplantation, including cyclosporin and sirolimus, can cause substantial elevation in LDL-C and TG levels. These patients can present a difficult clinical management problem. Chronic corticosteroid use, whether after transplant or in other inflammatory conditions, can also result in elevations in LDL-C and TG levels, sometimes producing a substantial mixed dyslipidemia. When the immunosuppressant or steroid must be continued, which is often the case, drug therapy with statins may be indicated in certain patients, with careful attention to the potential for untoward muscle-related side effects.

DISORDERS ASSOCIATED WITH REDUCED APOB-CONTAINING LIPOPROTEINS

Plasma concentrations of LDL-C <60 mg/dL are unusual. Although in some cases, LDL-C levels in this range may be reflective of malnutrition or serious chronic illness, LDL-C <60 mg/dL in an otherwise healthy individual suggests an inherited condition. The major inherited causes of low LDL-C are reviewed here and listed in Table 419-2.

Abetalipoproteinemia The synthesis and secretion of apoB-containing lipoproteins in the enterocytes of the proximal small bowel and in the hepatocytes of the liver involve a complex series of events that coordinate the coupling of various lipids with apoB-48 and apoB-100, respectively. Abetalipoproteinemia is a rare autosomal recessive disease caused by loss-of-function mutations in the gene encoding MTP (gene name *MTTP*), a protein that transfers lipids to nascent chylomicrons and VLDLs in the intestine and liver, respectively. Plasma levels of cholesterol and TG are extremely low in this disorder, and chylomicrons, VLDLs, LDLs, and apoB are undetectable in plasma. The parents of patients with abetalipoproteinemia (obligate heterozygotes) have normal plasma lipid and apoB levels. Abetalipoproteinemia usually presents in early childhood with diarrhea and failure to thrive due to fat malabsorption. The initial neurologic manifestations are loss of deep tendon reflexes, followed by decreased distal lower extremity vibratory and proprioceptive sense, dysmetria, ataxia, and the development of a spastic gait, often by the third or fourth decade. Patients with abetalipoproteinemia also develop a progressive pigmented retinopathy presenting with decreased night and color vision, followed by reductions in daytime visual acuity and ultimately progressing to near-blindness. The presence of spinocerebellar degeneration and pigmented retinopathy in this disease has resulted in some patients with abetalipoproteinemia being misdiagnosed as having Friedreich's ataxia.

Most of the clinical manifestations of abetalipoproteinemia result from defects in the absorption and transport of fat-soluble vitamins. Vitamin E and retinyl esters are normally transported from enterocytes to the liver by chylomicrons, and vitamin E is dependent on VLDL for transport out of the liver and into the circulation. As a consequence of the inability of these patients to secrete apoB-containing particles, patients with abetalipoproteinemia are markedly deficient in vitamin E and are also mildly to moderately deficient in vitamins A and K. Patients with abetalipoproteinemia should be referred to specialized centers for confirmation of the diagnosis and appropriate therapy. Treatment consists of a low-fat, high-caloric, vitamin-enriched diet accompanied by large supplemental doses of vitamin E. It is imperative that treatment be initiated as soon as possible to prevent development of neurologic sequelae, which can progress even with high-dose vitamin E therapy. New therapies for this serious, albeit rare, disease are needed. The discovery that genetic loss of MTP causes absent LDL-C led to the development of an MTP inhibitor to treat homozygous FH (see below).

Familial Hypobetalipoproteinemia (FHBL) FHBL generally refers to a condition of low total cholesterol, LDL-C, and apoB due to mutations in the *APOB* gene. Most of the mutations causing FHBL result in a truncated apoB protein, resulting in impaired assembly and secretion of chylomicrons from enterocytes and VLDL from the liver. Any secreted VLDL particles containing a truncated apoB protein are cleared from the circulation at an accelerated rate, which also contributes to the low levels of LDL-C and apoB. Individuals heterozygous for these mutations usually have LDL-C levels <60–80 mg/dL and also tend to have low levels of plasma TG. Many FHBL patients have elevated levels of hepatic fat (due to reduced VLDL export) and sometimes have increased levels of liver transaminases, although it appears that these patients infrequently develop associated hepatic inflammation and fibrosis.

Truncating mutations in both apoB alleles cause homozygous FHBL, an extremely rare disorder resembling abetalipoproteinemia with nearly undetectable LDL-C and apoB. The neurologic defects in homozygous hypobetalipoproteinemia are similar to those seen in abetalipoproteinemia but tend to be less severe. Homozygous hypobetalipoproteinemia can be distinguished from abetalipoproteinemia by examining the inheritance pattern of the plasma LDL-C level. The levels of LDL-C and apoB are normal in the parents of patients with abetalipoproteinemia, a classic recessive condition, and low in those of patients with homozygous hypobetalipoproteinemia, a co-dominant condition. The discovery that truncating mutations in apoB reduce LDL-C led to the development of an antisense oligonucleotide to treat HoFH (see below).

Familial PCSK9 Deficiency Another inherited cause of low LDL-C results from loss-of-function mutations in *PCSK9*. PCSK9 is a secreted protein that binds to the extracellular domain of the LDL receptor in the liver and promotes the degradation of the receptor. Heterozygosity for nonsense mutations in *PCSK9* that interfere with the synthesis of the protein are associated with increased hepatic LDL receptor activity and reduced plasma levels of LDL-C. Such mutations are more frequent in individuals of African descent. Individuals who are heterozygous for a loss-of-function mutation in *PCSK9* have an ~30–40% reduction in plasma levels of LDL-C and have a substantial protection from CHD relative to those without a *PCSK9* mutation, presumably due to having lower plasma cholesterol levels since birth. Homozygotes for these nonsense mutations have been reported and have extremely low LDL-C levels (<20 mg/dL) but appear otherwise healthy. A sequence variation of somewhat higher frequency (R46L) is found predominantly in individuals of European descent. This mutation impairs, but does not completely destroy, PCSK9 function. As a consequence, the plasma levels of LDL-C in individuals carrying this mutation are more modestly reduced (~15–20%); individuals with these mutations have a 45% reduction in CHD risk. The discovery of this condition led to the development of therapies that antagonize or silence PCSK9, thus reducing LDL-C levels and risk of CHD (see below).

Familial Combined Hypolipidemia Nonsense mutations in both alleles of the gene angiopoietin-like 3 (*ANGPTL3*) lead to low plasma levels of all three major lipid fractions—TG, LDL-C, and HDL-C—a phenotype termed *familial combined hypolipidemia*.

ANGPTL3 is a protein synthesized by the liver and secreted into the bloodstream. It inhibits LPL, thus delaying clearance of TRLs from the blood and increasing TRL blood concentrations. Deficiency of ANGPTL3, therefore, raises LPL activity and lowers blood TG; it also lowers LDL-C and raises HDL-C levels apparently related to the effects of ANGPTL3 on endothelial lipase. ANGPTL3 deficiency is associated with a reduced risk for CHD. The discovery of this condition led to the development of therapies that antagonize or silence ANGPTL3 to reduce LDL-C and TG levels (see below).

DISORDERS ASSOCIATED WITH REDUCED HIGH-DENSITY LIPOPROTEINS

Low levels of HDL-C, generally defined as <50 mg/dL in women and <40 mg/dL in men, are very common in clinical practice. Low HDL-C is an important independent predictor of increased cardiovascular risk and has been used regularly in standardized risk calculators. As an independent risk factor, it has clinical value in the assessment of cardiovascular risk, and a patient with low HDL-C should generally be considered at higher risk of ASCVD. However, it is now considered doubtful that low HDL-C is directly *causal* for the development of ASCVD. Thus, while HDL-C remains an important biomarker for assessing cardiovascular risk, it is no longer considered a target for therapeutic intervention to raise HDL-C levels in order to reduce cardiovascular risk.

HDL metabolism is strongly influenced by TG metabolism, insulin resistance, and inflammation, among other environmental and medical factors. Thus, the HDL-C measurement integrates a number of cardiovascular risk factors, potentially explaining its strong inverse association with ASCVD. The majority of patients with low HDL-C have some combination of genetic predisposition and secondary factors. Variants in hundreds of genes have been shown to influence HDL-C levels. Even more important quantitatively, obesity and insulin resistance have strong suppressive effects on HDL-C, and low HDL-C in these conditions is widely observed. Furthermore, the vast majority of patients with elevated TGs have reduced levels of HDL-C due to the substantial interplay between the metabolism of TRLs and HDL (see above). Most patients with low HDL-C who have been studied in detail have accelerated catabolism of HDL and its associated apoA-I protein as the physiologic basis for the low HDL-C. Single-gene Mendelian disorders that reduce HDL-C have been described (Table 419-2) but are rare; the vast majority of patients with low HDL-C have a polygenic predisposition with secondary factors like obesity, insulin resistance, or HTG.

PRIMARY (GENETIC) CAUSES OF LOW HDL-C

Mutations in three key genes encoding proteins that play critical roles in HDL synthesis and catabolism result in hypoalphalipoproteinemia (primary low levels of HDL-C). Unlike the genetic forms of hypercholesterolemia, which are invariably associated with premature coronary atherosclerosis, genetic forms of hypoalphalipoproteinemia are usually not associated with clearly increased risk of ASCVD. Nevertheless, in the clinical setting of an HDL-C level <20 mg/dL without accompanying severe HTG, these rare conditions should be considered.

Gene Deletions and Missense Mutations in *APOA1* Complete genetic deficiency of apoA-I due to a complete deletion of the *APOA1* gene results in the virtual absence of circulating HDL, proving the critical role of apoA-I in HDL biogenesis. The *APOA1* gene is part of a gene cluster on chromosome 11 that includes *APOA5*, *APOC3*, and *APOA4*. Some patients with no apoA-I have large genomic deletions that include other genes in the cluster. The rare patient lacking apoA-I may have cholesterol deposits in the cornea and in the skin, and in contrast to the other genetic disorders of low HDL-C, premature CHD has been reported. Heterozygotes for apoA-I deletions have reduced HDL-C levels but no obvious clinical sequelae.

More common, but still rare, are heterozygous missense mutations in the *APOA1* gene associated with low plasma levels of HDL-C. The first example reported, and still the best known, is an Arg173Cys substitution in apoA-I (so-called apoA-I_{Milano}), found in multiple residents of a town in northern Italy. Heterozygotes for this mutation have very low plasma levels of HDL-C (<25 mg/dL) due to impaired LCAT activation and accelerated clearance of the HDL particles containing the abnormal apoA-I. Despite having very low plasma levels of HDL-C, these individuals do not appear to have an increased risk of premature CHD (neither are they protected against CHD as was initially believed). Multiple other rare *APOA1* missense mutations causing low HDL-C have been reported. A few of these mutations in *APOA1* (as well as some mutations in *APOA2*) promote the formation of amyloid fibrils, causing systemic amyloidosis.

Tangier Disease (*ABCA1* Deficiency) Tangier disease is a rare autosomal co-dominant form of extremely low plasma HDL-C levels that is caused by mutations in the *ABCA1* gene encoding ABCA1, a cellular transporter that facilitates efflux of unesterified cholesterol and phospholipids from cells to apoA-I as an acceptor (Fig. 419-3). Through transporting cellular lipids, ABCA1 in the hepatocytes and intestinal enterocytes promotes the extracellular lipidation of the apoA-I secreted from the basolateral membranes of these tissues. In the genetic absence of ABCA1, the nascent, poorly lipidated apoA-I is rapidly cleared from the circulation. Thus, patients with Tangier disease (both *ABCA1* alleles mutated) have extremely low circulating plasma levels of HDL-C (<5 mg/dL) and apoA-I (<5 mg/dL). Cholesterol accumulates in the reticuloendothelial system of these patients, resulting in hepatosplenomegaly and pathognomonic enlarged, grayish yellow or orange tonsils. An intermittent peripheral neuropathy (mononeuritis multiplex) or a sphingomyelia-like neurologic disorder can also be seen in this disorder. Tangier disease may be associated with some increased risk of ASCVD, although the association is not as robust as might have been anticipated, given the extremely low levels of HDL-C in these patients. Patients with Tangier disease also have low plasma levels of LDL-C, which may attenuate the atherosclerotic risk. Heterozygotes for *ABCA1* mutations have moderately reduced plasma HDL-C levels (~15–40 mg/dL), and the effect on risk of ASCVD remains uncertain.

Familial LCAT Deficiency This rare autosomal recessive disorder is caused by mutations in LCAT, an enzyme synthesized in the liver and secreted into the plasma, where it circulates associated with lipoproteins (Fig. 419-3). As reviewed above, the enzyme is activated by apoA-I and mediates the esterification of cholesterol to form CEs. Consequently, in familial LCAT deficiency, the proportion of free cholesterol in circulating lipoproteins is greatly increased (from ~25% to >70% of total plasma cholesterol). Deficiency in this enzyme interferes with the maturation of HDL particles and results in rapid catabolism of circulating apoA-I.

Two genetic forms of familial LCAT deficiency have been described in humans: complete deficiency (also called *classic LCAT deficiency*) and partial deficiency (also called *fish eye disease*). Progressive corneal opacification due to the deposition of free cholesterol in the cornea, very low plasma levels of HDL-C (usually <10 mg/dL), and variable HTG are characteristic of both disorders. In partial LCAT deficiency, there are no other known clinical sequelae. In contrast, patients with complete LCAT deficiency have hemolytic anemia and progressive renal insufficiency that eventually leads to end-stage renal disease. Remarkably, despite the extremely low plasma levels of HDL-C and apoA-I, premature ASCVD is not a consistent feature of either LCAT deficiency or fish eye disease. The diagnosis can be confirmed in a specialized laboratory by assaying plasma LCAT activity or by sequencing the *LCAT* gene.

Primary Hypoalphalipoproteinemia Primary hypoalphalipoproteinemia is defined as a plasma HDL-C level below the tenth percentile in the setting of relatively normal cholesterol and TG levels, no apparent secondary causes of low plasma HDL-C, and no clinical signs of LCAT deficiency or Tangier disease. A family history of low HDL-C suggests an inherited condition and may trigger an evaluation of one of the Mendelian causes of hypoalphalipoproteinemia. However, most patients with isolated low HDL do not have an identifiable single-gene disorder and likely have a polygenic etiology, possibly exacerbated by a secondary factor. The physiologic defect appears to be accelerated catabolism of HDL and its apolipoproteins. Several kindreds with primary hypoalphalipoproteinemia and an increased incidence of premature CHD have been described, although it is not clear if the low HDL-C level is the cause of the accelerated atherosclerosis in these families.

SECONDARY FACTORS THAT REDUCE HDL-C LEVELS

Hypertriglyceridemia Low HDL-C is very commonly found in association with elevated TG levels. The lipolysis of TRLs generates lipids that transfer to HDL, and therefore, any impairment of lipolysis (the most common cause of elevated TGs) leads to reduced HDL biosynthesis. In settings of elevated TGs, where the HDL-C is not reduced, alternative explanations (e.g., FDBL, alcohol, estrogens) should be considered. Conversely, an isolated low HDL-C in the presence of normal TGs should prompt consideration of a primary genetic etiology (as above) or specific secondary factors (see below).

Very-Low-Fat Diet Dietary fat is positively associated with HDL-C levels. Individuals who eat very-low-fat vegan diets or who have anorexia or severe fat malabsorption often have low levels of HDL-C that are secondary to low dietary fat. In this setting, LDL-C levels are also usually low as well. There is no known harm to low HDL-C levels in this setting and no indication for liberalizing the diet solely for the purpose of raising the HDL-C.

Sedentary Lifestyle and Obesity Physical activity is known to have a (generally modest) effect in raising HDL-C levels, and conversely, a sedentary lifestyle is often associated with low HDL-C levels. Concordant with that observation, obesity is frequently associated with low HDL-C levels even when overt insulin resistance or HTG is not present. Increased physical activity and weight loss usually have some effect in raising HDL-C, which is not the primary reason for recommending these interventions but can have a motivating influence on the patient.

ANABOLIC STEROIDS AND TESTOSTERONE Anabolic steroids have a well-established effect on lowering HDL-C levels, sometimes quite dramatically. Testosterone supplementation can also reduce HDL-C levels, although not to the degree caused by anabolic steroids. In a young male patient who presents with unexplained very low HDL-C, a careful history of medication and supplement use should be taken.

APPROACH TO THE PATIENT

Lipoprotein Disorders

The major goals in the diagnosis and clinical management of lipoprotein disorders are (1) prevention of CVD and related cardiovascular events and (2) prevention of acute pancreatitis in patients with severe HTG. Given the high prevalence of dyslipidemia and the proven clinical benefits of early diagnosis and initiation of therapy, it is essential that physicians screen lipids systematically, rule out secondary causes of dyslipidemia, suspect inherited disorders of lipoprotein metabolism where appropriate, actively promote family-based screening, carefully assess risk for ASCVD and consider additional risk stratification approaches, and be knowledgeable about the wide range of existing therapeutic options for dyslipidemia. The field of clinical lipidology has matured and is moving toward a more systematic clinical application of genomic medicine. Diagnostic DNA sequencing or genotyping in patients with suspected FCS, FPLD, FH, and FDBL has the potential to enhance molecular diagnosis, facilitate appropriate therapeutic interventions, and promote family-based cascade screening based on genetic diagnosis.

DIAGNOSIS

A critical first step in managing a lipoprotein disorder is to attempt to determine the class or classes of lipoproteins that are increased or decreased in the patient. Once the dyslipidemia is accurately classified, efforts should be directed to identify or rule out any possible secondary causes (Table 419-3). A careful social, medical, and family history should be obtained. In patients with elevated TG levels (>150 mg/dL), a fasting glucose and/or hemoglobin A_{1c} should be obtained to rule out diabetes. In patients with elevated LDL-C levels (>160 mg/dL), a TSH should be obtained to rule out hypothyroidism and consideration should be given to the possibility of liver or kidney disease. Once secondary causes have been ruled out, attempts should be made to diagnose a primary lipid disorder, because the underlying genetic defect can provide important prognostic information regarding the risk of pancreatitis in severe HTG and the risk of ASCVD in other dyslipidemias, as well as impact on the choice of drug therapy and the screening of other family members. Obtaining the correct diagnosis often requires a detailed family history, lipid analyses in family members, and sometimes specialized or genetic testing.

Severe Hypertriglyceridemia If the fasting plasma TG level is >500 mg/dL, the patient has severe HTG and may be at risk for pancreatitis. If the TG levels are persistently severely elevated, especially if they are >1000 mg/dL, and the total cholesterol-to-TG ratio is >8, FCS should be considered, and genetic testing of an FCS gene panel may be indicated (Table 419-2). If central obesity, insulin resistance, and/or fatty liver disease are also present, consideration should be given to the possibility of FPLD, and an FPLD gene panel may be indicated (Table 419-2). However, most individuals with severe HTG do not have a single-gene disorder but have increased polygenic risk for high TGs often exacerbated by secondary factors (e.g., diet, alcohol, obesity, insulin resistance, medications). Such patients are still at risk for acute pancreatitis and should be treated to reduce their TG levels and thus their risk of pancreatitis (see below).

Hypercholesterolemia If the LDL-C levels are >190 mg/dL, the patient has severe hypercholesterolemia and is at risk for premature ASCVD. In absence of secondary causes, FH should be considered, particularly if there is a family history of hypercholesterolemia and/or premature CHD, and genetic testing of an FH gene panel may be indicated (Table 419-2). The Centers for Disease Control and Prevention has identified FH as a Tier 1 condition for implementation of public health genomics. While FH is ultimately a clinical diagnosis, a finding of a causal mutation may appropriately result in earlier and more aggressive therapy to lower LDL-C and should also promote family-based cascade screening. Recessive forms of severe hypercholesterolemia are rare, but if a patient with severe hypercholesterolemia has parents with normal cholesterol levels, ARH, sitosterolemia, and LALD should be considered, and genetic testing may be indicated (Table 419-2). Patients without an identified genetic variant or who have more moderate hypercholesterolemia are likely to have polygenic hypercholesterolemia but should still be considered at risk and eligible for treatment (see below).

Mixed Hyperlipidemia Patients with elevations in fasting plasma levels of both TGs (>150 mg/dL) and LDL-C (>130 mg/dL), often accompanied by reduced levels of HDL-C (<40 mg/dL in men and <50 mg/dL in women), are common and such patients are often diagnosed as having "mixed hyperlipidemia." Most such patients are at increased risk of ASCVD and merit consideration of lifestyle and often pharmacologic interventions. Secondary factors, particularly obesity, insulin resistance, and type 2 diabetes, are common in such patients, who often also have increased polygenic risk for dyslipidemia. The presence of palmar or tuberous xanthomas or an unusual lipid profile of total cholesterol and TG levels in the same range with an HDL-C that is not reduced should prompt consideration of FDBL, or type III hyperlipidemia, and can be diagnosed by advanced lipoprotein testing or genetic testing for the *APOE2* genotype. FDBL patients should be managed aggressively due to substantially increased risk of ASCVD. More commonly, patients with mixed hyperlipidemia, particularly those with family histories of dyslipidemia or premature ASCVD, have familial combined hyperlipidemia (FCHL). ApoB should be measured in such patients, and the finding of substantially elevated apoB levels can help identify patients with FCHL, who are at especially increased risk of ASCVD and require more aggressive treatment.

TREATMENT

Severe Hypertriglyceridemia

There is a well-established observational relationship between severe HTG, particularly chylomicronemia, and acute pancreatitis; however, there has never been a clinical trial designed or powered to definitively prove that intervention to reduce TGs reduces the risk of pancreatitis. Nevertheless, it is generally considered appropriate medical practice to intervene in patients with TGs >500 mg/dL in order to reduce the risk of pancreatitis. It remains uncertain whether chylomicronemia increases risk for ASCVD per se. Importantly, moderate HTG (TG 150–500 mg/dL) is associated with increased ASCVD risk; management of these patients is focused on reducing risk of ASCVD and on reducing LDL-C, non-HDL-C, and apoB.

LIFESTYLE AND MODIFIABLE FACTORS

In patients with severe HTG, lifestyle modification can be associated with a significant reduction in plasma TG level. Importantly, certain medications can exacerbate HTG (Table 419-3). Patients who drink alcohol should be encouraged to decrease or preferably eliminate their intake. Patients with severe HTG often benefit from a formal dietary consultation with a dietician intimately familiar with counseling patients on the dietary management of high TGs. Dietary fat intake should be restricted to reduce the formation of chylomicrons in the intestine. The excessive intake of simple carbohydrates should be discouraged because insulin drives TG production in the liver. Aerobic exercise and even increase in regular physical activity can have a positive effect in reducing TG levels and should be strongly encouraged. For patients who are overweight, weight loss can help to reduce TG levels. In extreme cases, bariatric surgery has been shown to not only produce effective weight loss but also substantially reduce plasma TG levels. Many patients with diabetes have HTG, and better control of diabetes can result in lowering of TGs. GLP-1 agonists prescribed for diabetes or obesity can reduce TG levels.

PHARMACOLOGIC THERAPY

Despite lifestyle interventions, many patients with severe HTG require pharmacologic therapy (Table 419-4). Patients who persist in having fasting TG >500 mg/dL despite active lifestyle management are candidates for pharmacologic therapy. The two major classes of drugs used for management of these patients are fibrates and omega-3 fatty acids (fish oils). In addition, statins can reduce plasma TG levels and also reduce ASCVD risk and should be used in patients with severe HTG who are at increased risk of ASCVD.

Fibrates Fibric acid derivatives, or fibrates, are agonists of PPARα, a nuclear receptor involved in the regulation of lipid metabolism. Fibrates stimulate LPL activity (enhancing TG hydrolysis), reduce apoC-III synthesis (enhancing lipoprotein remnant clearance), promote β-oxidation of fatty acids, and may reduce VLDL TG production. Fibrates reduce TG levels by ~30% in individuals with severe HTG and are often used as first-line therapy. They do not reduce and sometimes modestly raise LDL-C levels. Fibrates are generally well tolerated but can cause myopathy, especially when combined with statins, can raise creatinine, and are associated with an increase in gallstones. Fibrates can potentiate the effect of warfarin and certain oral hypoglycemic agents.

Omega-3 Fatty Acids (Fish Oils) Omega-3 fatty acids, or omega-3 polyunsaturated fatty acids (n-3 PUFAs), commonly known as fish oils, are present in high concentration in fish and in flaxseed. The n-3 PUFAs used for the treatment of HTG are eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). n-3 PUFAs have been concentrated into capsules and, in doses of 3–4 g/d, are effective at lowering fasting TG levels by ~30%. Fish oils can cause an increase in plasma LDL-C levels in some patients. Icosapent ethyl is an EPA-only product that has been shown to reduce cardiovascular events in patients with HTG. In general, fish oils are well tolerated, with the major side effect being dyspepsia. They appear to be safe, at least at doses up to 3–4 g, but can be associated with a prolongation in the bleeding time. Fish oils can be first-line therapy for the treatment of severe HTG or can be used in combination with fibrates.

APOC3 Silencing ApoC-III inhibits LPL and TRL uptake, and genetic variants in the *APOC3* gene reduce TG levels and risk of ASCVD. Volanesorsen is an antisense oligonucleotide (ASO) targeted to the *APOC3* mRNA in the liver; it significantly reduces plasma apoC-III and TG levels and is approved in Europe for patients with FCS. It has been associated with severe thrombocytopenia. Additional therapeutic approaches to APOC3 and other targets for TG lowering (e.g., ANGPTL3) are in development.

Hypercholesterolemia (Elevated LDL-C with or without Elevated TG)

There are abundant and compelling data that intervention to reduce LDL-C substantially reduces the risk of ASCVD, including myocardial infarction and stroke, as well as total mortality. Thus, it is imperative that patients with hypercholesterolemia be carefully assessed for cardiovascular risk and need for intervention. It is also worth emphasizing that patients with or at high risk for ASCVD who have plasma LDL-C levels in the "normal" or average range also benefit from intervention to reduce LDL-C levels.

LIFESTYLE AND MODIFIABLE FACTORS

In patients with elevated LDL-C, lifestyle modifications can be effective but are often less effective than in HTG. Patients should receive dietary counseling to reduce the content of saturated fats and trans fats in the diet. Obese patients should make an effort to lose weight. Regular aerobic exercise has relatively little impact on reducing plasma LDL-C levels, although it has cardiovascular benefits independent of LDL-C lowering. Patients with hypothyroidism should be optimally controlled. Finally, certain medications can elevate LDL-C levels (Table 419-3).

PHARMACOLOGIC THERAPY

The decision to use LDL-lowering drug therapy (Table 419-4)—with a statin being first-line therapy—depends on the presence of ASCVD or, if absent, the level of LDL-C as well as the level of cardiovascular risk. In patients with established ASCVD, drug therapy to reduce LDL-C is well supported by clinical trial data to reduce LDL-C as long as it remains >70 mg/dL, using combination drug therapy if necessary. In the absence of ASCVD, patients with FH must be treated to reduce the very high lifetime risk of ASCVD, and treatment should be initiated as early as possible, ideally during childhood. Otherwise, the decision to initiate LDL-lowering drug therapy is generally determined by the level of cardiovascular risk. For patients >40 years old without clinical CVD, the ASCVD pooled cohort risk calculator can be used to determine the 10-year absolute risk for CVD, and current guidelines suggest that a 10-year risk >7.5% merits consideration of statin therapy regardless of plasma LDL-C level. For younger patients, the assessment of lifetime risk of CVD may help inform the decision to start a statin, as well as a careful assessment of family history of ASCVD. In patients for whom the decision to start a statin is uncertain due to borderline ASCVD risk and/or borderline LDL-C levels, additional risk stratification might be considered. Blood tests that predict ASCVD risk beyond traditional risk factors include apoB, Lp(a), and high-sensitivity C-reactive protein (hs-CRP). In patients who are of a sufficient age (men >40 years and women >50 years), a coronary artery calcium (CAC) score has been shown to provide independent information about risk of future CAD. Elevated levels of one or more of these biomarkers or an elevated CAC score might be used to justify initiation of statin therapy in primary prevention for patients who are in a borderline zone with regard to treatment. Finally, given the strong polygenic contribution to ASCVD, there is increasing interest in the concept that a polygenic risk score for CAD might eventually be of clinical utility in lifetime risk assessment and decision-making regarding statin therapy in primary prevention.

HMG-CoA Reductase Inhibitors (Statins) Statins inhibit HMG-CoA reductase, a key enzyme in cholesterol biosynthesis. By inhibiting cholesterol synthesis in the liver, statins lead to a counterregulatory increase in the expression of the LDL receptor and thus accelerated clearance of circulating LDL, resulting in a dose-dependent reduction in plasma levels of LDL-C. The magnitude of LDL-C lowering associated with statin treatment (~30–55%) varies by statin and among individuals, but once a patient is on a statin, the doubling of the statin dose produces a ~6% further reduction in the level of plasma LDL-C. An extensive body of randomized clinical trials has clearly established that statin therapy significantly reduces major cardiovascular events (and in some cases total mortality) in both primary and secondary prevention settings. The seven statins currently available differ in their LDL-C–reducing potency (Table 419-4). Current recommendations are to use high-intensity statin therapy in patients with ASCVD or deemed at high risk of ASCVD. Statins also modestly reduce plasma TGs in a dose-dependent fashion roughly proportional to their LDL-C–lowering effects.

Statins, taken in tablet form once a day, are remarkably safe and well tolerated. The most important side effect associated with statin therapy is muscle pain, or myalgia, which occurs in 3–5% of patients, some of whom are unable to tolerate any statin. Severe myopathy (associated with an increase in plasma creatine kinase [CK]) and even rhabdomyolysis can occur rarely with statin treatment. The risk of statin-associated myalgia or myopathy is increased by the presence of older age, frailty, renal insufficiency, and coadministration of drugs that interfere with the metabolism of certain statins, such as erythromycin and related antibiotics, antifungal agents, immunosuppressive drugs, and fibric acid derivatives (particularly gemfibrozil). In the event of muscle symptoms, a plasma CK level may be obtained to differentiate myopathy from myalgia. Serum CK levels need not be monitored on a routine basis in patients taking statins because an elevated CK in the absence of symptoms does not predict the development of myopathy and does not necessarily suggest the need for discontinuing the drug. Statins can result in elevation in liver transaminases (alanine aminotransferase [ALT] and aspartate aminotransferase [AST]), but it is usually mild and transient and generally does not require discontinuation. Finally, meta-analyses of large randomized controlled clinical trials with statins indicate a slight excess in incident type 2 diabetes, an observation as yet not fully understood. However, the cardiovascular benefits associated with statin therapy far outweigh the slight increase in incident diabetes. Based on their safety and extensively documented benefit with regard to cardiovascular outcomes, statins are the clear drug class of choice for LDL-C reduction and are by far the most widely used class of lipid-lowering drugs.

Cholesterol Absorption Inhibitor Cholesterol within the lumen of the small intestine is derived from the diet (about one-third) and the bile (about two-thirds) and is actively absorbed by the enterocyte through a process that involves the protein NPC1L1. Ezetimibe (Table 419-4) is a cholesterol absorption inhibitor that binds directly to and inhibits NPC1L1 and blocks the intestinal absorption of cholesterol. Ezetimibe (10 mg taken once daily) inhibits cholesterol absorption by almost 60%, resulting in a reduction in delivery of dietary sterols in the liver and a compensatory increase in hepatic LDL receptor expression. The mean reduction in plasma LDL-C on ezetimibe (10 mg) is 18%, and the effect is additive when used in combination with a statin. Effects on TG and HDL-C levels are negligible. Ezetimibe added to a statin has been shown to significantly reduce major cardiovascular events compared with statin alone. It is generally considered the second-line option for adding to a statin in order to achieve further LDL-C reduction. Ezetimibe is very safe and well-tolerated. When used in combination with a statin, monitoring of liver transaminases is recommended. The only roles for ezetimibe in monotherapy are in patients who do not tolerate statins and in some patients with sitosterolemia.

PCSK9 Inhibitors Circulating PCSK9 targets the LDL receptor for lysosomal degradation, thus reducing its recycling and abundance at the surface of the hepatocyte. Genetic loss of function of *PCSK9* results in low levels of LDL-C and protection from CAD. Antibodies to PCSK9 (Table 419-4) sequester it and functionally increase the number of LDL receptors available to remove LDL from the blood. They are highly effective in lowering LDL-C, with an ~60% reduction in LDL-C. They also reduce plasma levels of Lp(a) modestly. Both PCSK9 antibodies have been shown to significantly reduce cardiovascular events when added to a statin in patients with existing CAD. These antibodies are administered subcutaneously every 2 weeks. They are generally well tolerated, with a side effect being injection site reactions. They are generally indicated as second-line (added to statin) or third-line (added to statin plus ezetimibe) therapy in patients with FH or ASCVD in whom LDL-C is not reduced to acceptable levels with a statin (with or without ezetimibe) alone. An alternative approach to silencing *PCSK9*, inclisiran, is a therapeutic siRNA molecule that targets the *PCSK9* mRNA in the liver. In contrast to the antibodies, it is administered subcutaneously every 6 months. It is effective in reducing LDL-C by ~60% and appears to be well tolerated and safe; cardiovascular outcomes trials are ongoing.

ATP Citrate Lyase Inhibitor Bempedoic acid is a first-in-class competitive inhibitor of ATP citrate lyase (ACL), which acts on mitochondrial-derived citrate to generate production of acetyl-CoA, which is subsequently used for cholesterol synthesis. Thus, it reduces cholesterol synthesis through a different mechanism than statins, ultimately upregulating the hepatic LDL receptor. Bempedoic acid is a prodrug that requires activation by very-long-chain acyl-CoA synthetase-1 (ASCVL1), which is not expressed in skeletal muscle, potentially explaining why it has less association with myalgias than statins; indeed, it has been shown to be relatively well tolerated in patients with statin-induced myalgias. In phase 3 trials, bempedoic acid 180 mg daily reduced LDL-C by ~18% when added to a statin and by ~23% as monotherapy. In a large cardiovascular outcomes trial in statin-intolerant patients, bempedoic acid 180 mg daily was shown to significantly reduce (by 13%) a four-component composite of major adverse cardiovascular events. It is also available in a fixed-dose combination with ezetimibe, which reduced LDL-C by ~36%, for patients who are statin intolerant. It can be used in combination with statins but should not be used with simvastatin in a dose >20 mg. Bempedoic acid is associated with increased uric acid levels and gout as well as with increased liver enzymes and cholelithiasis. Unlike statins, it is not associated with increased incidence of diabetes.

Bile Acid Sequestrants (Resins) Bile acid sequestrants (BAS) bind bile acids in the intestine and promote their excretion rather than reabsorption in the ileum. To maintain the bile acid pool size, the liver diverts cholesterol to bile acid synthesis. The decreased hepatic intracellular cholesterol content results in upregulation of the LDL receptor and enhanced LDL clearance from the plasma. BAS, including cholestyramine, colestipol, and colesevelam (Table 419-4), primarily reduce plasma LDL-C levels but can cause an increase in plasma TGs. Therefore, patients with HTG generally should not be treated with bile acid–binding resins. Cholestyramine and colestipol are insoluble resins that must be suspended in liquids. Colesevelam is available as tablets but generally requires up to six to seven tablets per day for effective LDL-C lowering. BAS are effective in combination with statins and in combination with ezetimibe. Side effects of resins are limited to the gastrointestinal tract and include bloating and constipation. Because BAS are not systemically absorbed, they are very safe and are the cholesterol-lowering drug of choice in children and in women who are pregnant, lactating, or actively trying to conceive. However, they are otherwise fourth- or fifth-line drugs for LDL-C reduction in other settings.

Specialized Drugs for HoFH Three "orphan" drugs are approved specifically for the management of HoFH, a rare condition caused by biallelic mutations in the major genes causing FH in which patients respond poorly to traditional LDL-lowering medications. Lomitapide is a small-molecule inhibitor of MTP that reduces LDL-C by ~50%, and mipomersen is an antisense oligonucleotide against apoB that reduces LDL-C by ~25%. Both of these drugs reduce hepatic VLDL production and thus LDL-C levels; however, due to their mechanism of action, each drug causes an increase in hepatic fat, the long-term consequences of which are unknown. In addition, lomitapide is associated with gastrointestinal-related side effects, and mipomersen is associated with skin reactions and flulike symptoms. Finally, an antibody inhibitor of ANGPTL3, evinacumab, was approved in 2021 for the treatment of HoFH. In a phase 3 trial, an intravenous infusion every 4 weeks reduced LDL-C levels in patients with HoFH by ~50% and was well tolerated. One of these three drugs should be strongly considered in HoFH patients after a trial of a high-intensity statin, and possibly a PCSK9 inhibitor, is shown to be insufficient to reduce LDL-C levels.

LDL Apheresis Patients with severe hypercholesterolemia who cannot reduce their LDL-C to acceptable levels despite optimally tolerated combination drug therapy are candidates for LDL apheresis. In this process, the patient's plasma is passed over a column that selectively removes the LDL, and the LDL-depleted plasma is returned to the patient. LDL apheresis is indicated for patients on maximally tolerated combination drug therapy (including a PCSK9 inhibitor) who have CHD and a plasma LDL-C level >200 mg/dL or no CHD and a plasma LDL-C level >300 mg/dL; LDL apheresis could be considered in high-risk patients who have an LDL-C >160 mg/dL on maximal therapy.

FURTHER READING

Cholesterol Treatment Trialists' (CTT) Collaboration: Effects of statin therapy on diagnoses of new-onset diabetes and worsening glycaemia in large-scale randomised blinded statin trials: An individual participant data meta-analysis. Lancet Diabetes Endocrinol 12:306, 2024.

Hernandez P et al: Clinical management of hypertriglyceridemia in the prevention of cardiovascular disease and pancreatitis. Curr Atheroscler Reports 23:72, 2021.

Klarin D, Natarajan P: Clinical utility of polygenic risk scores for coronary artery disease. Nat Rev Cardiol 19:291, 2022.

Loh WJ, Watts GF: The inherited hypercholesterolemias. Endocrinol Metab Clin North Am 51:511, 2022.

Mangione CM et al: Statin use for the primary prevention of cardiovascular disease in adults: US Preventive Services Task Force recommendation statement. JAMA 328:746, 2022.

Piccirillo F et al: Novel antidiabetic agents and their effects on lipid profile: A single shot for several cardiovascular targets. Int J Mol Sci 24:10164, 2023.

Sakhuja S et al: Recurrent atherosclerotic cardiovascular disease events potentially prevented with guideline-recommended cholesterol-lowering therapy following myocardial infarction. Cardiovasc Drugs Ther 38:937, 2024.

Shamsudeen I, Hegele RA: Advances in the care of lipodystrophies. Curr Opin Endocrinol Diabetes Obesity 29:152, 2022.

Virani SS et al: 2023 AHA/ACC/ACCP/ASPC/NLA/PCNA guideline for the management of patients with chronic coronary disease: A report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines. Circulation 148:e9, 2023.

Watts GF et al: International Atherosclerosis Society guidance for implementing best practice in the care of familial hypercholesterolaemia. Nat Rev Cardiol 20:845, 2023.

420

The Metabolic Syndrome

Robert H. Eckel

The metabolic syndrome (syndrome X, insulin resistance syndrome) consists of a constellation of metabolic abnormalities that confer increased risk of cardiovascular disease (CVD) and diabetes mellitus. Evolution of the criteria for the metabolic syndrome since the original definition by the World Health Organization in 1998 reflects growing clinical evidence and analysis by a variety of consensus conferences and professional organizations. The major features of metabolic syndrome include central obesity, hypertriglyceridemia, low levels of high-density lipoprotein (HDL) cholesterol, hyperglycemia, and hypertension **(Table 420-1)**.

GLOBAL HEALTH/EPIDEMIOLOGY

The most challenging feature of the metabolic syndrome to define is waist circumference. Intraabdominal circumference (visceral adipose tissue) is most strongly related to insulin resistance and risk of diabetes and CVD, and for any given waist circumference, the distribution of adipose tissue between subcutaneous (SC) and visceral depots varies substantially. Thus, within and between populations, there is a lesser versus greater risk at the same waist circumference. These differences in populations reflect the range of waist circumferences considered to confer risk in different geographic locations (Table 420-1).

The prevalence of the metabolic syndrome varies around the world, in part reflecting the age and ethnicity of the populations studied and the diagnostic criteria applied. In general, the prevalence of metabolic syndrome increases with age. The prevalence of metabolic syndrome in the U.S. adult population meeting the criteria of the National Cholesterol Education Program (NCEP) and Adult Treatment Panel III (ATPIII) is ~35%. Greater global industrialization is associated with rising rates of obesity and related increase in the prevalence of the metabolic syndrome, especially as the population ages.

Using National Health and Nutrition Examination Survey (NHANES) data from 1999–2018, the prevalence of metabolic syndrome in 28,049 adults in the United States was 33.4%. The highest prevalence was age-dependent with reduction by age 80 among all subgroups, i.e., from 19.5% among those aged 20–39 years to 48.6% among those aged ≥60 years. Importantly, the rising prevalence and severity of obesity among children reflect features of the metabolic syndrome in a younger population, now estimated to be 12 and 30% among obese and overweight children, respectively.

The frequency distribution of different components of metabolic syndrome for the U.S. population (NHANES III) and the Guangdong Gut Microbiome Project of China is summarized in **Fig. 420-1**. Note the major differences in the U.S. population compared to Han Chinese. Moreover, within the United States, abdominal obesity appears equally prevalent in all U.S. races, whereas the prevalence of age-dependent other components differs as shown in Fig. 420-1. Increases in hyperglycemia were most evident in the 2017–2018 sample, whereas central obesity, low HDL cholesterol, and hypertension prevalence have been relatively constant, while levels of triglycerides (defined as >150 mg/dL) have progressively decreased.

RISK FACTORS

Overweight/Obesity Metabolic syndrome was first described in the early twentieth century; however, the worldwide overweight/obesity epidemic has recently been the force driving its increasing recognition. Central adiposity is a key feature of the syndrome, and the syndrome's prevalence reflects the strong relationship between waist circumference and increasing adiposity. However, despite the importance of obesity, patients who are of normal weight may also be insulin-resistant and may have metabolic syndrome. This phenotype is particularly evident for populations in India, Southeast Asia, and Central America.

TABLE 420-1 NCEP:ATPIII[a] 2001 and Harmonizing Definition Criteria for the Metabolic Syndrome

<table>
<tr><th>NCEP:ATPIII 2001</th><th colspan="3">HARMONIZING DEFINITION[b]</th></tr>
<tr><td rowspan="7">Three or more of the following:
• Central obesity: waist circumference >102 cm (males), >88 cm (females)
• Hypertriglyceridemia: triglyceride level ≥150 mg/dL or specific medication
• Low HDL[c] cholesterol: <40 mg/dL and <50 mg/dL for men and women, respectively, or specific medication
• Hypertension: blood pressure ≥130 mmHg systolic or ≥85 mmHg diastolic or specific medication
• Fasting plasma glucose level ≥100 mg/dL or specific medication or previously diagnosed type 2 diabetes</td><td colspan="3">Three of the following:</td></tr>
<tr><td colspan="3">Waist circumference (cm)</td></tr>
<tr><td>Men</td><td>Women</td><td>Ethnicity</td></tr>
<tr><td>≥94</td><td>≥80</td><td>Europid, sub-Saharan African, Eastern and Middle Eastern</td></tr>
<tr><td>≥90</td><td>≥80</td><td>South Asian, Chinese, and ethnic South and Central American</td></tr>
<tr><td>≥85</td><td>≥90</td><td>Japanese</td></tr>
<tr><td colspan="3">• Fasting triglyceride level >150 mg/dL or specific medication
• HDL cholesterol level <40 mg/dL and <50 mg/dL for men and women, respectively, or specific medication
• Blood pressure >130 mm systolic or >85 mm diastolic or previous diagnosis or specific medication
• Fasting plasma glucose level ≥100 mg/dL (alternative indication: drug treatment of elevated glucose levels)</td></tr>
</table>

[a]National Cholesterol Education Program and Adult Treatment Panel III. [b]In this analysis, the following thresholds for waist circumference were used: white men, ≥94 cm; African-American men, ≥94 cm; Mexican-American men, ≥90 cm; white women, ≥80 cm; African-American women, ≥80 cm; Mexican-American women, ≥80 cm. For participants whose designation was "other race—including multiracial," thresholds that were once based on Europid cutoffs (≥94 cm for men and ≥80 cm for women) and on South Asian cutoffs (≥90 cm for men and ≥80 cm for women) were used. For participants who were considered "other Hispanic," the International Diabetes Federation thresholds for ethnic South and Central Americans were used. [c]High-density lipoprotein.

Sedentary Lifestyle Physical inactivity and less cardiorespiratory fitness are predictors of CVD events and the related risk of death. Many components of the metabolic syndrome are associated with a sedentary lifestyle, including increased adipose tissue (predominantly central), reduced HDL cholesterol, and increased triglycerides, blood pressure, and glucose in genetically susceptible persons. Compared with individuals who watch television or videos or use the computer <1 h daily, those who do so for >4 h daily have a twofold increased risk of the metabolic syndrome.

Genetics No single gene explains the complex phenotype called metabolic syndrome. However, using genome-wide association and candidate gene approaches, several genetic variants are associated with metabolic syndrome. Although many of the loci have unknown function, many others relate to body weight and composition, insulin resistance, and unfavorable disturbances in lipid and lipoprotein metabolism. In general, heritability estimates for each of the metabolic traits exceed 50%.

Aging The metabolic syndrome affects nearly 50% of the U.S. population aged >60, and at >60 years of age, women are more often affected. The age dependency of the syndrome's prevalence is seen in most populations around the world.

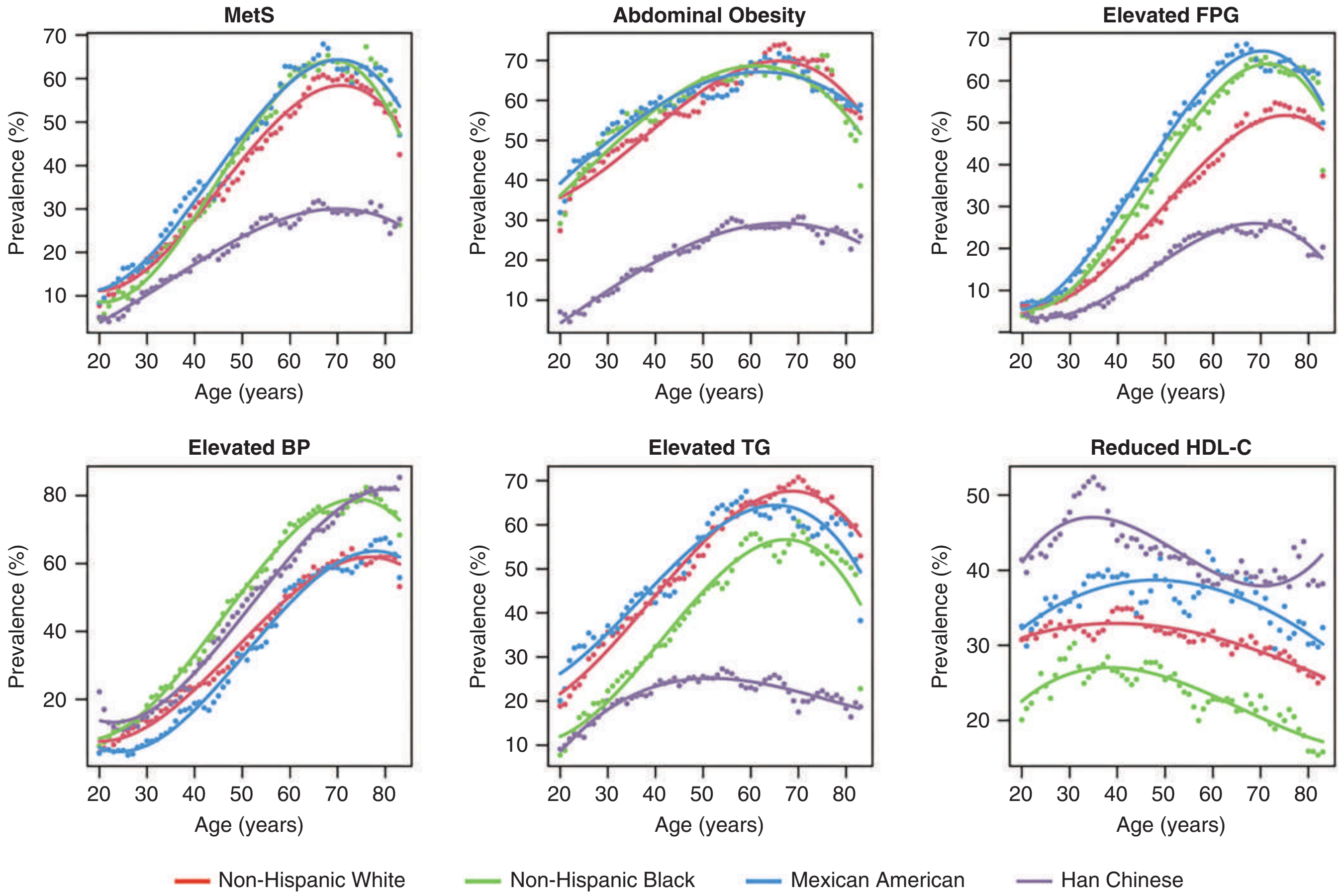

FIGURE 420-1 The frequency distribution of the metabolic syndrome for the U.S. population (National Health and Nutrition Examination Survey [NHANES] III) and the Guangdong Gut Microbiome Project of China. The prevalence of metabolic syndrome (MetS) and its components with age across different races. Prevalence was estimated using a SWAN algorithm (shown as dots). The trajectory of the prevalence of MetS with age was fitted by cubic regressions (shown as lines). BP, blood pressure; FPG, fasting plasma glucose; HDL-C, high-density lipid cholesterol; SWAN, sliding window–based algorithm; TG, triglycerides. *(Reproduced from R Zhang et al: The racial disparities in the epidemic of metabolic syndrome with increased age: A study from 28,049 Chinese and American adults. Front Public Health 9:797183, 2022.)*

Diabetes Mellitus Diabetes mellitus can be included in both the NCEP and the Harmonizing Definitions of metabolic syndrome, but the greatest value of metabolic syndrome, and especially fasting glucose, is predicting type 2 diabetes. The great majority (~75%) of patients with type 2 diabetes or impaired glucose tolerance have metabolic syndrome. The presence of metabolic syndrome in these populations relates to a higher prevalence of CVD than in patients who have type 2 diabetes or impaired glucose tolerance but do not have the syndrome.

Cardiovascular Disease Individuals with metabolic syndrome are twice as likely to die of CVD as those who do not, and their risk of acute myocardial infarction or stroke is threefold higher. The approximate prevalence of metabolic syndrome among patients with coronary heart disease (CHD) is up to 60% in persons >75 years, with a prevalence of ~35% among patients with premature coronary artery disease (age ≤45) and a particularly higher prevalence among women. With appropriate cardiac rehabilitation and changes in lifestyle (e.g., nutrition, physical activity, weight reduction, and—in some cases—pharmacologic therapy), the prevalence of the syndrome can be reduced.

Lipodystrophy Lipodystrophic disorders in general are associated with metabolic syndrome. Moreover, it is quite common for such patients to present with the metabolic syndrome. Both genetic lipodystrophy (e.g., Berardinelli-Seip congenital lipodystrophy, Dunnigan familial partial lipodystrophy) and acquired lipodystrophy (e.g., HIV-related lipodystrophy and in HIV patients receiving certain antiretroviral therapies) may give rise to severe insulin resistance and many of the components of metabolic syndrome.

ETIOLOGY

Insulin Resistance The most accepted and unifying hypothesis to describe the pathophysiology of metabolic syndrome is insulin resistance, caused systemically by an incompletely understood defect in insulin action (Chap. 415). The onset of insulin resistance is heralded by postprandial hyperinsulinemia, which is followed by fasting hyperinsulinemia and ultimately by hyperglycemia.

An early major contributor to the development of insulin resistance is an overabundance of circulating fatty acids (Fig. 420-2). Plasma albumin-bound free fatty acids are derived predominantly from adipose-tissue triglyceride stores released by intracellular lipolytic enzymes. The lipolysis of triglyceride-rich lipoproteins in tissues by lipoprotein lipase also produces free fatty acids. Insulin mediates both anti-lipolysis and the stimulation of lipoprotein lipase in adipose tissue. Of note, the inhibition of lipolysis in adipose tissue is the most sensitive pathway of insulin action. Thus, when insulin resistance develops, increased lipolysis produces more fatty acids, which further decreases the anti-lipolytic effect of insulin. Excessive fatty acids enhance substrate availability and create insulin resistance by modifying downstream signaling. Fatty acids impair insulin-mediated glucose uptake and are associated with accumulation of triglycerides in both skeletal and cardiac muscle, whereas increased fatty acid flux increases endogenous glucose production and triglyceride production, accumulation, and secretion in the liver.

Reductions in leptin action may also be a pathophysiologic mechanism to explain metabolic syndrome. Physiologically, leptin reduces appetite, promotes energy expenditure, and enhances insulin sensitivity. In addition, leptin may regulate cardiac and vascular function through a nitric oxide–dependent mechanism. However, when obesity

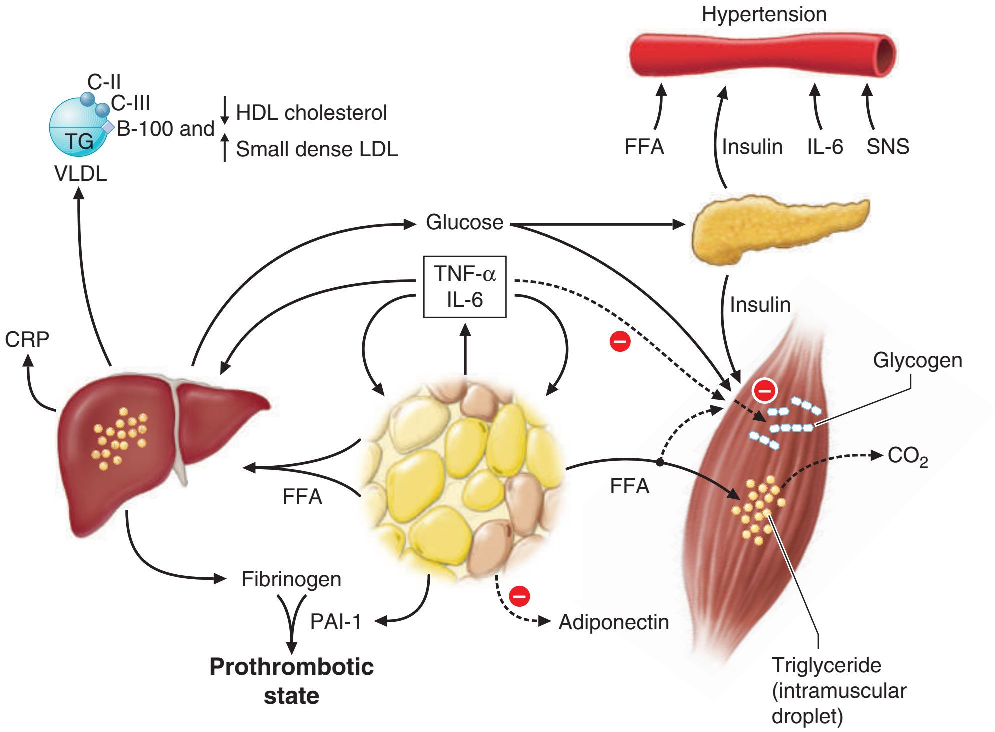

FIGURE 420-2 Pathophysiology of the metabolic syndrome. Free fatty acids (FFAs) are released in abundance from an expanded adipose tissue mass. In the liver, FFAs result in increased production of glucose and triglycerides and secretion of very-low-density lipoproteins (VLDLs). Associated lipid/lipoprotein abnormalities include reductions in high-density lipoprotein (HDL) cholesterol and an increased low-density lipoprotein (LDL) particle number. FFAs also reduce insulin sensitivity in muscle by inhibiting insulin-mediated glucose uptake. Associated defects include a reduction in glucose partitioning to glycogen and increased lipid accumulation in triglyceride (TG). The increase in circulating glucose, and to some extent FFAs, increases pancreatic insulin secretion, resulting in hyperinsulinemia. Hyperinsulinemia may result in enhanced sodium reabsorption and increased sympathetic nervous system (SNS) activity and contribute to hypertension, as might higher levels of circulating FFAs. The proinflammatory state is superimposed and contributory to the insulin resistance produced by excessive FFAs. The enhanced secretion of interleukin 6 (IL-6) and tumor necrosis factor α (TNF-α) produced by adipocytes and monocyte-derived macrophages results in more insulin resistance and lipolysis of adipose tissue triglyceride stores to circulating FFAs. IL-6 and other cytokines also enhance hepatic glucose production, VLDL production by the liver, hypertension, and insulin resistance in muscle. Insulin resistance also contributes to increased triglyceride accumulation in the liver (nonalcoholic fatty liver disease). Cytokines and FFAs also increase hepatic production of fibrinogen and adipocyte production of plasminogen activator inhibitor 1 (PAI-1), resulting in a pro-thrombotic state. Higher levels of circulating cytokines stimulate hepatic production of C-reactive protein (CRP). Reduced production of the anti-inflammatory and insulin-sensitizing cytokine adiponectin is also associated with the metabolic syndrome. *(Reproduced with permission from RH Eckel et al: The metabolic syndrome. Lancet 365:1415, 2005.)*

develops, hyperleptinemia ensues, with evidence of leptin resistance in the brain and other tissues resulting in insulin resistance and associated inflammation, hyperlipidemia, and a plethora of cardiovascular disorders, such as hypertension, atherosclerosis, CHD, and heart failure. Moreover, a series of adipokines relate to metabolic syndrome. Whereas adiponectin improves insulin sensitivity in adipose tissue and skeletal muscle, visfatin, fetuin-A, resistin, asprosin, and plasminogen activator inhibitor-1 contribute to insulin resistance and systemic glucose intolerance.

The oxidative stress hypothesis provides a unifying theory for aging and the predisposition to metabolic syndrome. In studies of insulin-resistant individuals with obesity or type 2 diabetes, the offspring of persons with type 2 diabetes, and the elderly, a defect in mitochondrial oxidative phosphorylation leads to the accumulation of triglycerides and related lipid molecules in muscle, liver, and other tissues, i.e., β-cells.

The gut microbiome has emerged as an important contributor to the development of obesity and related metabolic disorders, including inflammation and components of metabolic syndrome. Although the mechanisms remain uncertain, an increased ratio of Firmicutes/Bacteroidetes species in addition to genetic predisposition, diet, and bile acid metabolism are associated with and may play an etiologic role in metabolic syndrome.

Increased Waist Circumference Waist circumference is an important component of the most recent and frequently applied diagnostic criteria for metabolic syndrome. However, measuring waist circumference does not reliably distinguish increases in SC abdominal adipose tissue from that in intra-abdominal or visceral fat; this distinction requires dual X-ray absorptiometry (DEXA), computed tomography (CT), or magnetic resonance imaging (MRI) to discriminate. With increases in visceral adipose tissue, adipose tissue–derived free fatty acids reach the liver more readily. In contrast, increases in abdominal SC fat release lipolysis products into the systemic circulation and therefore have fewer direct effects on hepatic metabolism. Relative increases in visceral versus SC adipose tissue with increasing waist circumference in Asians and Asian Indians may explain the greater prevalence of metabolic syndrome in those populations than in African Americans, in whom SC fat predominates. It is also possible that visceral fat is a marker for—but not the source of—excess postprandial free fatty acids in obesity.

Dyslipidemia **(See also Chap. 419)** In general, free fatty acid flux from adipose tissue to the liver results in increased production of apolipoprotein (apo) B–containing, triglyceride-rich, very-low-density lipoproteins (VLDLs). The direct effect of insulin on this process is complex, but *hypertriglyceridemia* is an excellent marker of the insulin-resistant condition. Not only is hypertriglyceridemia a feature of metabolic syndrome, but patients with metabolic syndrome have elevated levels of apoC-III carried on VLDLs and other lipoproteins. This increase in apoC-III is inhibitory to lipoprotein lipase, reducing triglyceride-rich lipoprotein remnant removal, further contributing to hypertriglyceridemia, and confers more risk for atherosclerotic cardiovascular disease (ASCVD).

The other major lipoprotein disturbance in metabolic syndrome is a *reduction in HDL cholesterol*. This reduction is a consequence of changes in HDL composition and metabolism. In the presence of hypertriglyceridemia, a decrease in the cholesterol content of HDL is a consequence of reduced cholesteryl ester content of the lipoprotein core in combination with cholesteryl ester transfer protein–mediated alterations in triglycerides that make the HDL particle small and dense. This change in lipoprotein composition also results in increased clearance of HDL from the circulation. These changes in HDL have a relationship to insulin resistance that is probably indirect, occurring in concert with the changes in triglyceride-rich lipoprotein metabolism.

In addition to HDLs, low-density lipoproteins (LDLs) have alterations in composition in metabolic syndrome. With fasting serum triglycerides at >2.0 mM (~180 mg/dL), there is usually a predominance of small dense LDLs, which are thought to be more atherogenic, although their association with hypertriglyceridemia and low HDLs makes their independent contribution to ASCVD difficult to assess. Individuals with hypertriglyceridemia often have increases in cholesterol content of both VLDL1 and VLDL2 subfractions and in LDL particle number. Both lipoprotein changes may contribute to atherogenic risk in patients with metabolic syndrome.

Glucose Intolerance **(See also Chap. 415)** Defects in insulin action in metabolic syndrome lead to impaired suppression of endogenous glucose production by the liver (and kidney) and reduced glucose uptake and metabolism in insulin-sensitive tissues—i.e., muscle and adipose tissue. There is a strong relationship between impaired fasting glucose or impaired glucose tolerance and insulin resistance in studies of humans, nonhuman primates, and rodents. To compensate for defects in insulin action, insulin secretion and/or clearance increases or decreases, respectively, so that euglycemia remains. Ultimately, this compensatory mechanism fails because of defects in insulin secretion, resulting in progression from impaired fasting glucose and/or impaired glucose tolerance to type 2 diabetes mellitus.

Hypertension The relationship between insulin resistance and hypertension is well established. Paradoxically, under normal physiologic conditions, insulin-mediated increases in nitric oxide cause vasodilation with secondary effects on sodium reabsorption in the kidney. However, in the setting of insulin resistance, the vasodilatory effect of insulin is lost but the renal effect on sodium reabsorption is preserved. Sodium reabsorption is increased in Caucasians with metabolic syndrome but not in Africans or Asians. Insulin also increases the activity of the sympathetic nervous system, an effect that is preserved in the setting of insulin resistance. Insulin resistance is also associated with pathway-specific impairment in phosphatidylinositol-3-kinase signaling. In the endothelium, this impairment may cause an imbalance between the production of nitric oxide and the secretion of endothelin 1, with a consequent decrease in blood flow. In addition, increases in angiotensinogen gene expression in adipose tissue of obese subjects results in increases in circulating angiotensin II and vasoconstriction. Although these mechanisms are provocative, the inadequate evaluation of insulin action by measurement of fasting insulin levels or by homeostasis model assessment shows that insulin resistance contributes only partially to the increased prevalence of hypertension in metabolic syndrome.

Another possible mechanism underlying hypertension in metabolic syndrome is the vasoactive role of perivascular adipose tissue. Reactive oxygen species released by NADPH oxidase impair endothelial function and result in local vasoconstriction. Other paracrine effects such as leptin or other proinflammatory cytokines released from adipose tissue, such as tumor necrosis factor α (TNF-α), may also be important.

Hyperuricemia is another consequence of insulin resistance in metabolic syndrome. There is growing evidence not only that uric acid is associated with hypertension but also that reduction of uric acid normalizes blood pressure in hyperuricemic adolescents with hypertension. The mechanism appears to be in part related to an adverse effect of uric acid on nitric oxide synthase in the macula densa of the kidney and stimulation of the renin-angiotensin-aldosterone system.

Proinflammatory Cytokines The increases in proinflammatory cytokines—including interleukins 1, 6, and 18; resistin; TNF-α; and the systemic biomarker C-reactive protein—reflect overproduction by the expanded adipose tissue mass (Fig. 420-2). Adipose tissue–derived macrophages may be the primary source of proinflammatory cytokines locally and in the systemic circulation. It remains unclear, however, how much of the insulin resistance is caused by the paracrine effects of these cytokines and how much by the endocrine effects.

Adiponectin Adiponectin is an anti-inflammatory cytokine produced exclusively by adipocytes. Adiponectin enhances insulin sensitivity and inhibits many steps in the inflammatory process. In the liver, adiponectin inhibits the expression of gluconeogenic enzymes and the rate of glucose production. In muscle, adiponectin increases glucose transport and enhances fatty acid oxidation, partially through the activation of AMP kinase. Reductions in adiponectin levels are common in metabolic syndrome. The relative contributions of adiponectin deficiency and overabundance of the proinflammatory cytokines are unclear.

CLINICAL FEATURES

Symptoms and Signs Metabolic syndrome typically is not associated with symptoms. On physical examination, waist circumference and blood pressure are often elevated. The presence of either or both signs should prompt the clinician to search for other biochemical abnormalities that may be associated with metabolic syndrome. Much less frequently, lipoatrophy or acanthosis nigricans is present on examination. Because these physical findings characteristically are associated with severe insulin resistance, other components of metabolic syndrome are much more common.

Associated Diseases • **CARDIOVASCULAR DISEASE** The relative risk for new-onset CVD in patients with metabolic syndrome who do not have diabetes averages 1.5- to 3-fold. However, in INTERHEART, a study of 26,903 subjects from 52 countries, the risk for acute myocardial infarction in subjects with metabolic syndrome (World Health Organization or International Diabetes Federation definition) is comparable to that conferred by some, but not all, of the component risk factors. Diabetes mellitus (odds ratio [OR], 2.72) and hypertension (OR, 2.60) are stronger than other risk factors. Although congestive heart failure and metabolic syndrome can occur together, typically this consequence is secondary to metabolic syndrome–related ASCVD or hypertension. Metabolic syndrome is also associated with increases in the risk for stroke, peripheral vascular disease, and Alzheimer's disease. However, as for myocardial infarction, the risk beyond the additive role of the components of metabolic syndrome remains debatable. In the Reasons for Geographic and Racial Differences in Stroke (REGARDS) cohort, an observational study of black and white adults ≥45 years old across the United States, there were 9741 participants, and 41% had metabolic syndrome. After adjustment for multiple confounders, metabolic syndrome was associated with increases in high-sensitivity C-reactive protein (hsCRP), and this relationship was associated with a 1.34 relative risk for all-cause mortality, but <50% of deaths were from CVD. The population-attributable risk was 9.5% for metabolic syndrome alone and 14.7% for both metabolic syndrome and increased hsCRP. The relationship of metabolic syndrome and hsCRP to mortality was greater for whites than blacks.

TYPE 2 DIABETES Overall, the risk for type 2 diabetes among patients with metabolic syndrome is increased three- to fivefold. In the Framingham Offspring Study's 8-year follow-up of middle-aged participants, the population-attributable risk of metabolic syndrome for developing type 2 diabetes was 62% among men and 47% among women, yet increases in fasting plasma glucose explained most, if not all, of this increased risk.

Other Associated Conditions In addition to the features specifically used to define metabolic syndrome, other metabolic alterations are secondary to or accompany insulin resistance. Those alterations include increases in apoB and apoC-III, uric acid, prothrombotic factors (fibrinogen, plasminogen activator inhibitor 1), serum viscosity, asymmetric dimethylarginine, homocysteine, white blood cell count, proinflammatory cytokines, C-reactive protein, urine albumin/creatinine ratio, metabolic-associated fatty liver disease (MAFLD) and/or nonalcoholic steatohepatitis (NASH), polycystic ovary syndrome, and obstructive sleep apnea.

METABOLIC-ASSOCIATED FATTY LIVER DISEASE MAFLD has become the most common liver disease, in part a consequence of the insulin resistance of metabolic syndrome. The mechanism relates to increases in free fatty acid flux and reductions in intrahepatic fatty acid oxidation with resultant increases in triglyceride biosynthesis and hepatocellular accumulation, with variable inflammation and oxidative stress. The more serious Metabolic dysfunction-associated steatohepatitis (MASH), a consequence of MAFLD in some patients and a precursor of cirrhosis and end-stage liver disease, includes a more substantial proinflammatory contribution. MAFLD affects ~10% of the nonobese population and up to 65% of patients with metabolic syndrome; over half of these patients have metabolic dysfunction-associated steatohepatitis (MASH). As the prevalence of overweight/obesity and metabolic syndrome increases, NASH may become one of the more common causes of end-stage liver disease and hepatocellular carcinoma. Increasingly, studies have shown that MAFLD is related independently to CVD, especially coronary artery disease.

HYPERURICEMIA (See also Chap. 384) Hyperuricemia reflects defects in insulin action on the renal tubular reabsorption of uric acid and may contribute to hypertension through its effect on the endothelium. An increase in asymmetric dimethylarginine, an endogenous inhibitor of nitric oxide synthase, also relates to endothelial dysfunction. In addition, increases in the urine albumin/creatinine ratio may relate to altered endothelial pathophysiology in the insulin-resistant state.

POLYCYSTIC OVARY SYNDROME (See also Chap. 404) Polycystic ovary syndrome is highly associated with insulin resistance (50–80%) and metabolic syndrome, with a prevalence of the syndrome between 12 and 60% based on phenotypes D through A.

OBSTRUCTIVE SLEEP APNEA (See also Chap. 33) Obstructive sleep apnea is commonly associated with obesity, hypertension, increased circulating proinflammatory cytokines, impaired glucose tolerance, and insulin resistance. In fact, obstructive sleep apnea may predict metabolic syndrome, even in the absence of excess adiposity. Moreover, when biomarkers of insulin resistance are compared between patients with obstructive sleep apnea and weight-matched controls, insulin resistance is found to be more severe in those with apnea. Continuous positive airway pressure treatment improves insulin sensitivity in patients with obstructive sleep apnea.

DIAGNOSIS

The diagnosis of metabolic syndrome relies on fulfillment of the criteria listed in Table 420-1, as assessed using tools at the bedside and in the laboratory. The medical history should include evaluation of symptoms for obstructive sleep apnea in all patients and polycystic ovary syndrome in premenopausal women. Family history will help determine the risk for CVD and diabetes mellitus. Blood pressure and waist circumference measurements provide information necessary for the diagnosis.

Laboratory Tests Measurement of fasting lipids and glucose is needed in determining whether metabolic syndrome is present. The measurement of additional biomarkers associated with insulin resistance can be individualized. Such tests might include those for apoB, hsCRP, fibrinogen, uric acid, urinary albumin/creatinine ratio, and liver function. A sleep study should be performed if symptoms of obstructive sleep apnea are present. If polycystic ovary syndrome is suspected based on clinical features and anovulation, testosterone, luteinizing hormone, and follicle-stimulating hormone should be measured. MAFLD can be further assessed by the MAFLD fibrosis score (FIB4) or elastography.

TREATMENT

The Metabolic Syndrome

LIFESTYLE (SEE ALSO CHAP. 414)

Obesity, particularly abdominal, is the driving force behind metabolic syndrome. Thus, weight reduction is the primary approach to the disorder. With at least 5% and more so with 10% weight reduction, improvement in insulin sensitivity results in favorable modifications in many components of metabolic syndrome. In general, recommendations for weight loss include a combination of caloric restriction, increased physical activity, and behavior modification. Caloric restriction is the most important component, whereas increases in physical activity are important for maintenance of weight loss. Some but not all evidence suggests that the addition of exercise to caloric restriction may promote greater weight loss from the visceral depot. The tendency for weight regains after successful weight reduction underscores the need for long-lasting behavioral changes.

Diet Before prescribing a weight-loss diet, it is important to emphasize that it has taken the patient a long time to develop an expanded fat mass; thus, the correction need not occur quickly. Given, in general, that ~3500 kcal = 1 lb of adipose tissue, an ~500-kcal restriction daily equates to weight reduction of 1 lb per week. Diets restricted in carbohydrate typically provide a more rapid initial weight loss. However, after 1 year, the amount of weight reduction is minimally reduced or no different from that with caloric restriction alone. Thus, adherence to the diet is more important than the chosen diet. Moreover, there is concern about low-carbohydrate diets enriched in saturated fat, particularly for patients at risk for ASCVD. Therefore, a high-quality dietary pattern—i.e., a diet enriched in fruits, vegetables, whole grains, lean poultry, and fish—should be encouraged to maximize overall health benefit.

Physical Activity Before prescribing a physical activity program to patients with metabolic syndrome, it is important to ensure that the increased activity does not incur risk. Some high-risk patients should undergo formal cardiovascular evaluation before initiating an exercise program. For an inactive participant, gradual increases in physical activity should be encouraged to enhance adherence and avoid injury. Although increases in physical activity can lead to modest weight reduction, 60–90 min of moderate- to high-intensity daily activity is required to achieve this goal. Even if an overweight or obese adult is unable to undertake this level of activity, a health benefit will follow from at least 30 min of moderate-intensity activity daily. The caloric value of 30 min of a variety of activities can be found at *https://www.health.harvard.edu/diet-and-weight-loss/calories-burned-in-30-minutes-of-leisure-and-routine-activities*. Of note, a variety of routine activities, such as gardening, walking, and housecleaning, require moderate caloric expenditure. Thus, physical activity should not be defined solely in terms of formal exercise such as jogging, swimming, or tennis.

Behavior Modification Behavioral treatment typically includes recommendations for dietary restriction and more physical activity that predicts sufficient weight loss that benefits metabolic health. The subsequent challenge is the duration of the program because weight regain so often follows successful weight reduction. Improved long-term outcomes often follow a variety of methods, such as a personal or group counselor, the Internet, social media, and telephone follow-up to maintain contact between providers and patients.

Obesity **(See also Chap. 414)** In some patients with metabolic syndrome, treatment options need to extend beyond lifestyle intervention. Weight-loss drugs come in two major classes: appetite suppressants and absorption inhibitors. Appetite suppressants approved by the U.S. Food and Drug Administration (FDA) include phentermine (for short-term use [3 months] only) as well as phentermine/topiramate, naltrexone/bupropion, high-dose (3.0 mg) liraglutide (rather than 1.8 mg, the maximum for treatment of type 2 diabetes), and semaglutide (2.4 mg), which are approved without restrictions on the duration of therapy. In clinical trials, the phentermine/topiramate extended-release combination resulted in ~8% weight loss relative to placebo in 50% of patients. Side effects include palpitations, headache, paresthesias, constipation, and insomnia. Naltrexone/bupropion extended release reduces body weight by ≥10% in ~20% of patients; however, the drug combination is contraindicated in patients with seizure disorders or any condition that predisposes to seizures. Naltrexone/bupropion also increases pulse and blood pressure and should not be given to patients with uncontrolled hypertension. High-dose liraglutide, a glucagon-like peptide 1 (GLP-1) receptor agonist, results in ~6% weight loss relative to placebo with ~33% of patients with >10% weight loss. Common side effects are limited to the upper gastrointestinal tract, including nausea and, less frequently, emesis. Semaglutide (2.4 mg weekly) has been shown to produce an average weight loss of 14.9% over 68 weeks. Tirzepatide, a novel glucose-dependent insulinotropic polypeptide (GIP) and GLP-1 receptor agonist, has been tested for 72 weeks in participants with a mean body weight of 104.8 kg and mean body mass index (BMI) of 38.0 kg/m^2, with 94.5% of patients with a BMI of ≥30 kg/m^2. Participants experienced a dose-dependent reduction in weight ranging from –15.0% with 5 mg of tirzepatide weekly to 20.9% with the 15-mg dose. Benefits of GLP-1 receptor agonists on MAFLD are also noteworthy, but not yet FDA approved.

Orlistat inhibits fat absorption by ~30% and is moderately effective compared with placebo (~4% more weight loss). Moreover, orlistat reduced the incidence of type 2 diabetes, an effect that was especially evident among patients with impaired glucose tolerance at baseline. This drug is often difficult to take because of oily leakage per rectum. In general, for all weight-loss drugs, greater weight reduction leads to greater improvement in metabolic syndrome components, including the conversion from prediabetes to type 2 diabetes.

Metabolic or bariatric surgery is an important option for patients with metabolic syndrome who have a BMI >40 kg/m^2 or >35 kg/m^2 with comorbidities. An evolving application for metabolic surgery includes patients with a BMI as low as 30 kg/m^2 and type 2 diabetes. Gastric bypass or vertical sleeve gastrectomy results in dramatic weight reduction and improvement in most features of metabolic syndrome. A survival benefit with gastric bypass has also been realized.

LDL CHOLESTEROL (SEE ALSO CHAP. 419)

The rationale for the development of criteria for metabolic syndrome by NCEP was to go beyond LDL cholesterol in identifying and reducing the risk of ASCVD. The working assumption by the panel was that LDL cholesterol goals had already been achieved and that increasing evidence supports a linear reduction in ASCVD events because of progressive lowering of LDL cholesterol with statins with subsequent benefit using additional LDL cholesterol–lowering agents. The 2019 American College of Cardiology (ACC)/American Health Association (AHA) Cholesterol Guidelines have no specific recommendations for patients with metabolic syndrome; however, they recommend that patients aged 20–75 years with LDL cholesterol levels ≥190 mg/dL should use a high-intensity statin (e.g., atorvastatin 40–80 mg or rosuvastatin 20–40 mg daily) and those with type 2 diabetes aged 40–75 years should use a moderate-intensity statin and, if or when risk estimate is high, a high-intensity statin. For patients with metabolic syndrome but without diabetes, the 10-year ASCVD risk estimator should be employed, and patients with a risk ≥7.5% and ≤20% or persons aged 20–59 with elevated lifetime risk should have a discussion with their provider about initiating statin therapy for primary prevention of ASCVD. A coronary calcium score may help in making this decision.

Diets restricted in saturated fats (<6% of calories) and trans fats (as few as possible) should be applied aggressively. Although evidence is controversial, dietary cholesterol can also be restricted. If LDL cholesterol remains elevated, pharmacologic intervention is needed. Based on substantial evidence, treatment with statins, which lower LDL cholesterol by 15–60%, is the first-choice medication intervention. Of note, for each doubling of the statin dose, LDL cholesterol is further lowered by only ~6%. Hepatotoxicity (more than a threefold increase in hepatic aminotransferases) is rare, but myopathy occurs in ~10–20% of patients. The cholesterol absorption inhibitor ezetimibe is well tolerated and should be the second-choice medication intervention. Ezetimibe typically reduces LDL cholesterol by 15–20%. Bempedoic acid alone or in combination with ezetimibe is another option, with up to a 35% lowering of LDL cholesterol with the combination. Bempedoic acid can increase plasma uric acid. Proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors are potent LDL cholesterol–lowering drugs (~45–60%) but are not needed for most patients with metabolic syndrome. Of course, if these patients also have familial hypercholesterolemia or insufficient LDL cholesterol lowering on statins with or without ezetimibe, a PCSK9 inhibitor should be considered. The bile acid sequestrants cholestyramine, colestipol, and colesevelam may be more effective than ezetimibe alone, but because they can increase triglyceride levels, they must be used with caution in patients with metabolic syndrome when fasting triglycerides are >300 mg/dL. Side effects include gastrointestinal symptoms (palatability, bloating, belching, constipation, anal irritation). Nicotinic acid has similar LDL cholesterol–lowering capabilities (<20%); however, it may be associated with multiple adverse effects. Fibrates are best employed to lower LDL cholesterol when triglycerides are not elevated. Fenofibrate may be more effective than gemfibrozil in this setting.

TRIGLYCERIDES (SEE ALSO CHAP. 419)

The 2019 ACC/AHA Cholesterol Guidelines stated that fasting triglycerides >500 mg/dL should be treated to prevent more serious hypertriglyceridemia and pancreatitis. Although a fasting triglyceride value of >150 mg/dL is a component of metabolic syndrome, post hoc analyses of multiple fibrate trials have not suggested a triglyceride-related reduction in the primary ASCVD outcome in patients (with or without concomitant statin therapy) with fasting triglycerides >200 mg/dL, often in the setting of reduced levels of HDL cholesterol. It remains uncertain whether triglycerides cause ASCVD or if levels are just associated with increased ASCVD risk.

A fibrate (gemfibrozil or fenofibrate) is one drug class of choice to lower fasting triglyceride levels, which are typically reduced by 30–45%. Concomitant administration with drugs metabolized by the 3A4 cytochrome P450 system (including some statins) increases the risk of myopathy. In these cases, fenofibrate may be preferable to gemfibrozil. In the Veterans Affairs HDL Intervention Trial, gemfibrozil was administered to men with known CHD and levels of HDL cholesterol <40 mg/dL. A coronary disease event and mortality rate benefit was experienced predominantly among men with hyperinsulinemia and/or diabetes, many of whom were identified retrospectively as having metabolic syndrome. Of note, the degree of triglyceride lowering in this trial or other fibrate trials did not predict benefit.

Other drugs that lower triglyceride levels include statins, nicotinic acid, and prescription omega-3 fatty acids. For this purpose, an intermediate or high dose of the "more potent" statins (atorvastatin, rosuvastatin) is needed. The effect of nicotinic acid on fasting triglycerides is dose related and ~20–35%, an effect that is less pronounced than that of fibrates. In patients with metabolic syndrome and diabetes, nicotinic acid may increase fasting glucose levels, and clinical trials with nicotinic acid plus a statin have failed to reduce ASCVD events. Prescriptions of omega-3 fatty acid preparations that include high doses of eicosapentaenoic acid (EPA) with or without docosahexaenoic acid (DHA) (~1.5–4.5 g/d) lower fasting triglyceride levels by ~25–40%. The two omega-3 randomized controlled trials associated with ASCVD risk reduction, JELIS and REDUCE-IT, used EPA only, whereas STRENGTH, which was terminated prematurely because of futility, used EPA plus DHA. Here, no drug interactions with fibrates or statins occur, and the main side effect of their use is eructation with a fishy taste. Freezing the nutraceutical can partially block this unpleasant side effect. Importantly, lowering triglycerides with any of the pharmaceuticals has not been proven to be an independent predictor of CVD outcomes.

HDL CHOLESTEROL (SEE ALSO CHAP. 419)

Very few lipid-modifying compounds increase HDL cholesterol levels. Statins, fibrates, and bile acid sequestrants have modest effects (5–10%), whereas ezetimibe and omega-3 fatty acids have no effect. Nicotinic acid is the only currently available drug with predictable HDL cholesterol–raising properties. The response is dose related, and nicotinic acid can increase HDL cholesterol by up to 30% above baseline. After several trials of nicotinic acid versus placebo in statin-treated patients, there is no evidence that raising HDL cholesterol with nicotinic acid beneficially affects ASCVD events in patients with or without metabolic syndrome.

BLOOD PRESSURE (SEE ALSO CHAP. 288)

The direct relationship between blood pressure and all-cause mortality rate has been well established in studies comparing patients with hypertension (>140/90 mmHg), patients with prehypertension (>120/80 mmHg but <140/90 mmHg), and individuals with normal blood pressure (<120/80 mmHg). In patients who have metabolic syndrome without diabetes, the best choice for the initial antihypertensive medication is an angiotensin-converting enzyme (ACE) inhibitor or an angiotensin II receptor blocker, as these two classes of drugs are effective and well tolerated. Additional agents include a diuretic, calcium channel blocker, beta blocker, and mineralocorticoid inhibitor, such as the recent FDA-approved mineralocorticoid receptor antagonist finerenone. In all patients with hypertension, a sodium-restricted dietary pattern enriched in fruits and vegetables, whole grains, and low-fat dairy products should be advocated. Home monitoring of blood pressure may assist in maintaining good blood pressure control.

IMPAIRED FASTING GLUCOSE (SEE ALSO CHAP. 415)

In patients with metabolic syndrome and type 2 diabetes, aggressive glycemic control may favorably modify fasting levels of triglycerides and/or HDL cholesterol. In patients with impaired fasting glucose who do not have diabetes, a lifestyle intervention that includes weight reduction, dietary saturated fat restriction, and increased physical activity has been shown to reduce the incidence of type 2 diabetes. Metformin also reduces the incidence of diabetes, although the effect is less pronounced than that of lifestyle intervention.

INSULIN RESISTANCE (SEE ALSO CHAP. 416)

Several drug classes (biguanides, thiazolidinediones [TZDs]) increase insulin sensitivity. Because insulin resistance is the primary pathophysiologic mechanism for metabolic syndrome, representative drugs in these classes reduce its prevalence. Both metformin and TZDs enhance insulin action in the liver and suppress endogenous glucose production. TZDs, but not metformin, also improve insulin-mediated glucose uptake in muscle and adipose tissue. In a meta-analysis of nine trials involving 12,026 participants, the TZD pioglitazone versus placebo was associated with reduction in ASCVD events in patients with insulin resistance (metabolic syndrome), prediabetes, and type 2 diabetes. However, adverse effects including weight gain, bone fracture, and congestive heart failure with/or without edema were seen. Benefit of TZDs has been seen in patients with MAFLD, and with metformin in women with polycystic ovary syndrome, and both drug classes have been shown to reduce markers of inflammation. GLP-1 receptor agonists also improve insulin sensitivity, which is related to the amount of weight reduction.

FURTHER READING

Alberti KG et al: Harmonizing the metabolic syndrome: A joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation 120:1640, 2009.

Brown AE, Walker M: Genetics of insulin resistance and the metabolic syndrome. Curr Cardiol Rep 18:75, 2016.

Dobrowolski P et al: Metabolic syndrome: A new definition and management guidelines. Arch Med Sci 5:1, 2022.

Eckel RH et al: The metabolic syndrome. Lancet 365:1415, 2005.

Fahed G et al: Metabolic syndrome: Update on pathophysiology and management in 2021. Int J Mol Sci 23: 786, 2022.

Genser L et al: Obesity, type 2 diabetes, and the metabolic syndrome: Pathophysiologic relationships and guidelines for surgical intervention. Surg Clin North Am 96:681, 2016.

Lechner K et al: High-risk atherosclerosis and metabolic phenotype: The roles of ectopic adiposity, atherogenic dyslipidemia, and inflammation. Metab Syndr Relat Disord 18:176, 2020.

Neeland IJ et al: Visceral and ectopic fat, atherosclerosis, and cardiometabolic disease: A position statement. Lancet Diabetes Endocrinol 7:715, 2019.

Section 4 Disorders of Bone and Mineral Metabolism

421 Bone and Mineral Metabolism in Health and Disease

F. Richard Bringhurst, Henry M. Kronenberg, Eva S. Liu, Marc N. Wein

BONE STRUCTURE AND METABOLISM

Bone is a dynamic tissue that is remodeled constantly throughout life. The arrangement of compact and cancellous bone provides strength and density suitable for both mobility and protection. Compact or cortical bone forms the roughly cylindrical shell of long bones; cancellous or trabecular bone forms the plate-like meshwork that internally supports the cortical shell. In addition, bone provides a reservoir for calcium, magnesium, phosphorus, sodium, and other ions necessary for homeostatic functions. Bone also hosts and regulates hematopoiesis by providing niches for hematopoietic cell proliferation and differentiation. The skeleton is highly vascular and receives ~10% of the cardiac output. Remodeling of bone is accomplished by two distinct cell types: osteoblasts produce bone matrix, and osteoclasts resorb the matrix. The activities of these cells are coordinated by osteocytes, long-lived regulatory cells embedded within bone matrix.

The extracellular components of bone consist of a solid mineral phase in close association with an organic matrix, of which 90–95% is type I collagen **(Chap. 425)**. The noncollagenous portion of the organic matrix is heterogeneous and contains serum proteins such as albumin as well as many locally produced proteins, whose functions are incompletely understood. Those proteins include cell attachment/signaling proteins such as thrombospondin, osteopontin, and fibronectin; calcium-binding proteins such as matrix gla protein and osteocalcin; and proteoglycans such as biglycan and decorin. Some of the proteins organize collagen fibrils; others influence mineralization and binding of the mineral phase to the matrix.

The mineral phase is made up of calcium and phosphate and is best characterized as a poorly crystalline hydroxyapatite. The mineral phase of bone is deposited initially in intimate relation to the collagen fibrils and is laid down in specific locations in the "holes" between the collagen fibrils. This architectural arrangement of mineral and matrix results in a two-phase material well suited to withstand mechanical stresses. The organization of collagen influences the amount and type of mineral phase formed in bone. Although the primary structures of type I collagen in skin and bone tissues are similar, there are differences in posttranslational modifications and distribution of intermolecular cross-links. The holes in the packing structure of the collagen are larger in mineralized collagen of bone and dentin than in unmineralized collagens such as those in tendon. Single amino acid substitutions in the helical portion of either the α1 (*COL1A1*) or α2 (*COL1A2*) chains of type I collagen disrupt the organization of bone in the disease, osteogenesis imperfecta. The severe skeletal fragility associated with this group of disorders highlights the importance of the fibrillar matrix in the structure of bone **(Chap. 425)**.

Osteoblasts synthesize and secrete the organic matrix and regulate its mineralization. They are derived from cells of mesenchymal origin **(Fig. 421-1*A*)**. Osteoblast precursors derive from the periosteum, the bone marrow, or the hypertrophic chondrocytes at the end of the growth plate. Active osteoblasts are found on the surface of newly forming bone. As an osteoblast secretes matrix, which then is mineralized, the cell may become an *osteocyte*, still connected with its nutrient supply through a series of canaliculi. Osteocytes account for the vast majority of the cells in bone. They are thought to be the mechanosensors that communicate signals to surface osteoblasts and osteoclasts and their progenitors through the canalicular network and thereby serve as master regulators of bone formation and resorption. Osteocytes also secrete fibroblast growth factor 23 (FGF23), a major hormonal regulator of phosphate metabolism (see below). Mineralization of the matrix, both in trabecular bone and in osteones of compact cortical bone (*Haversian systems*), begins soon after the matrix is secreted (primary mineralization) but is not completed for several weeks or even longer (secondary mineralization). Although this mineralization takes advantage of the high concentrations of calcium and phosphate, already near saturation in serum, mineralization is a carefully regulated process that is dependent on the activity of osteoblast-derived alkaline phosphatase, which probably works by hydrolyzing inhibitors of mineralization, such as pyrophosphate.

Genetic studies in humans and mice have identified several key genes that control osteoblast development. *Runx2* is a transcription factor expressed specifically in chondrocyte (cartilage cells) and osteoblast progenitors as well as in hypertrophic chondrocytes and mature osteoblasts. *Runx2* regulates the expression of several important osteoblast proteins, including osterix (SP7) (another transcription factor needed for osteoblast maturation), osteopontin, bone sialoprotein, type I collagen, osteocalcin, and receptor-activator of nuclear factor (NF)-κB (RANK) ligand. *Runx2* expression is regulated in part by bone morphogenic proteins (BMPs). *Runx2*-deficient mice are devoid of osteoblasts, whereas mice with a deletion of only one allele (*Runx2* +/–) exhibit a delay in formation of the clavicles and some cranial bones. The latter abnormalities are similar to those in the human disorder *cleidocranial dysplasia*, which is also caused by heterozygous inactivating mutations in *Runx2*.

The paracrine signaling molecule, Indian hedgehog (Ihh), also plays a critical role in osteoblast development, as evidenced by Ihh-deficient mice that lack osteoblasts in the type of bone formed on a cartilage mold (endochondral ossification). Signals originating from members of the wnt family of paracrine factors are also important for osteoblast proliferation and differentiation. Osteocytes regulate osteoblasts partly by secreting a potent inhibitor of wnt signaling called sclerostin. Numerous other growth-regulatory factors affect osteoblast function, including the three closely related transforming growth factor βs, fibroblast growth factors (FGFs) 2 and 18, platelet-derived growth factor, and insulin-like growth factors (IGFs) I and II. Hormones such as parathyroid hormone (PTH) and 1,25-dihydroxyvitamin D [1,25$(OH)_2$D] activate receptors expressed by osteoblasts to assure mineral homeostasis and influence a variety of bone cell functions. Osteoclasts that resorb bone (see below) also regulate osteoblasts by releasing growth factors from bone matrix and by synthesizing proteins that can directly regulate osteoblastogenesis.

Resorption of bone is carried out mainly by *osteoclasts*, multinucleated cells that are formed by fusion of cells derived from the common precursor of macrophages and osteoclasts. Thus, these cells derive from the hematopoietic lineage, quite different from the mesenchymal lineage cells that become osteoblasts. Multiple factors that regulate osteoclast development have been identified **(Fig. 421-1*B*)**. Factors produced by osteocytes, osteoblasts, and marrow stromal cells allow cells of the osteoblast lineage to control osteoclast development and activity. Macrophage colony-stimulating factor (M-CSF) plays a critical role during several steps in the pathway and ultimately leads to fusion of osteoclast progenitor cells to form multinucleated, active osteoclasts. RANK ligand, a member of the tumor necrosis factor (TNF) family, is expressed on the surface of osteocytes, osteoblasts, and stromal fibroblasts. In a process involving cell-cell interactions, RANK ligand binds to the RANK receptor on osteoclast progenitors, stimulating osteoclast differentiation and activation. Alternatively, a soluble decoy receptor, referred to as osteoprotegerin (OPG), can bind RANK ligand and inhibit osteoclast differentiation. Several growth factors and cytokines (including interleukins 1, 6, and 11; TNF; and interferon γ) modulate osteoclast differentiation and function. Most hormones that influence osteoclast function do not target these cells directly but instead target cells of the osteoblast lineage to increase production of M-CSF and RANK. Both PTH and 1,25$(OH)_2$D increase osteoclast number and

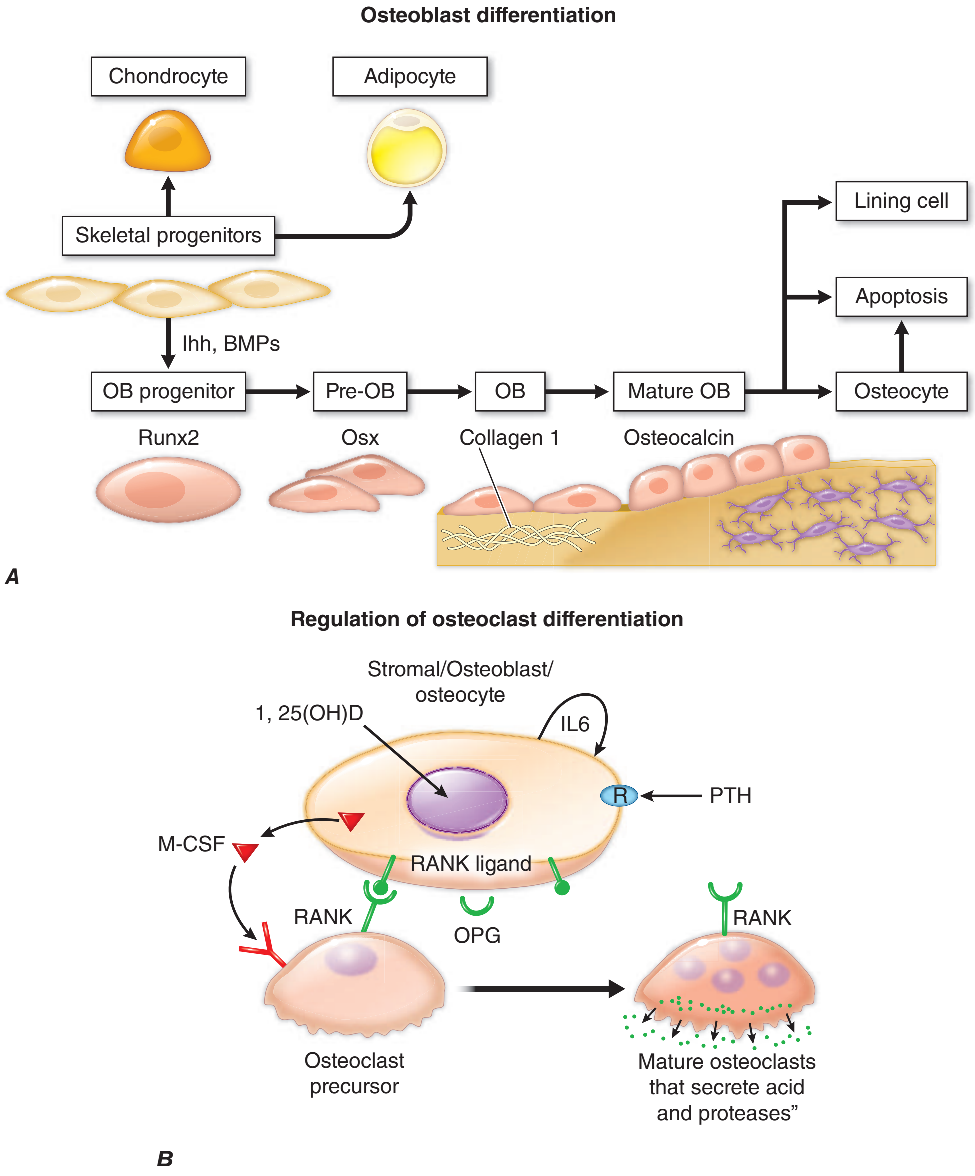

FIGURE 421-1 Pathways regulating development of (*A*) osteoblasts and (*B*) osteoclasts. ***A.*** Osteoblast lineage cells. Osteoblast progenitors are mesenchymal cells that are found in the perichondrium/periosteum, the bone marrow (where they also support hematopoietic stem cells and can become marrow adipocytes), and the growth plate (where late hypertrophic chondrocytes can die or become osteoblast precursors in the marrow). These progenitors can become chondrocytes early in development or in response to fracture postnatally. In response to Indian hedgehog (Ihh) and bone morphogenetic proteins (BMPs), these precursors become committed to osteoblast differentiation and are regulated by a series of transcription factors, with early essential factors, Runx2 and then osterix (SP7), shown here. On the bone surface, these cells express large amounts of collagen I and cell surface alkaline phosphatase, a crucial regulator of mineralization. Osteoblasts are relatively short-lived and subsequently either die or become osteocytes buried in bone matrix or quiescent bone-lining cells. To track the osteoblast (OB) lineage, we use CreERt mice driven by the Osx and collagen1 gene promotors. During OB differentiation, Osx is starting to be expressed early on in pre-OBs, while col1 is expressed later in OB development. It can therefore be expected that Osx- and col1-CreERt mouse lines will mark osteoblastic cells at different stages of differentiation. ***B.*** Regulation of osteoclast production. Osteoclast progenitors, derived from the hematopoietic lineage, respond to signals from cells of the osteoblast lineage to increase their number and activity. IL, interleukin; M-CSF, macrophage colony-stimulating factor; OPG, osteoprotegerin; PTH, parathyroid hormone.

activity by this indirect mechanism. Calcitonin, in contrast, binds to its receptor on the basal surface of osteoclasts and directly inhibits osteoclast function. Estradiol has multiple cellular targets in bone, including osteoclasts, immune cells, and osteoblasts; actions on all these cells serve to decrease osteoclast number and bone resorption.

Osteoclast-mediated resorption of bone takes place in scalloped spaces (*Howship's lacunae*) where the osteoclasts are attached through a specific αvβ3 integrin to components of the bone matrix. The osteoclast forms a tight seal to the underlying matrix and secretes protons, chloride, and proteinases into a confined space that has been likened to an extracellular lysosome. The active osteoclast surface forms a ruffled border that contains a specialized proton pump ATPase that secretes acid that solubilizes the mineral phase. Carbonic anhydrase (type II isoenzyme) within the osteoclast generates the needed protons. The bone matrix is resorbed in the acid environment adjacent to the ruffled border by proteases, such as cathepsin K, that act at low pH.

A distinct process, called osteocytic osteolysis, also causes bone resorption. Using a molecular machinery similar to that in osteoclasts, osteocytes dissolve mineral and matrix from the bone surrounding osteocytes. PTH and its relative, parathyroid hormone–related peptide (PTHrP), stimulate osteocytic osteolysis. The relative contributions of osteoclasts and osteocytes to bone resorption is uncertain.

In the embryo and the growing child, bone develops mostly by replacing previously calcified cartilage (endochondral bone formation) with subsequent remodeling or, in a few bones, is formed without a cartilage matrix (intramembranous bone formation). During endochondral bone formation, chondrocytes proliferate, secrete and mineralize a matrix, enlarge (hypertrophy), and then either die or differentiate into precursors of osteoblasts, lengthening bone and

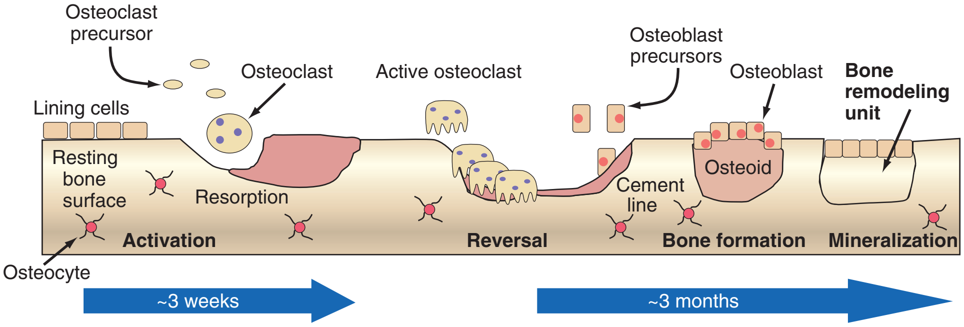

FIGURE 421-2 Schematic representation of bone remodeling. The cycle of bone remodeling is carried out by the basic multicellular unit (BMU), which consists of a group of osteoclasts and osteoblasts. In cortical bone, the BMUs tunnel through the tissue, whereas in cancellous bone, they move across the trabecular surface. The process of bone remodeling is initiated by the recruitment of osteoclast precursors, perhaps to sites of microdamage. These precursors fuse to form multinucleated, active osteoclasts that mediate bone resorption. Osteoclasts adhere to bone and subsequently remove it by acidification and proteolytic digestion. As the BMU advances, osteoclasts leave the resorption site, and osteoblasts, derived from marrow precursors and previously inactive bone lining cells, move in to cover the excavated area and begin the process of new bone formation by secreting osteoid, which eventually is mineralized into new bone. After osteoid mineralization, osteoblasts flatten and form a layer of lining cells over new bone, become osteocytes, or die.

providing the matrix and factors that stimulate endochondral bone formation. This program is regulated by both local factors, such as IGF-I and -II, Ihh, PTHrP, BMPs, and FGFs, and systemic hormones, such as growth hormone, glucocorticoids, and estrogen.

New bone, whether formed in infants or in adults during repair, has a relatively high ratio of cells to matrix and is characterized by coarse fiber bundles of collagen that are interlaced and randomly dispersed (woven bone). In adults, the more mature bone is organized with fiber bundles regularly arranged in parallel or concentric sheets (lamellar bone). In long bones, deposition of lamellar bone in a concentric arrangement around blood vessels forms the Haversian systems. Growth in length of bones is dependent on proliferation of cartilage cells and the endochondral sequence at the growth plate. Growth in width and thickness is accomplished by formation of bone at the periosteal surface and by resorption at the endosteal surface, with the rate of formation exceeding that of resorption. In adults, after the growth plates of cartilage close through the actions of estrogen, growth in length and endochondral bone formation cease. Even in adults, however, remodeling of bone (within Haversian systems as well as along the surfaces of trabecular bone) continues throughout life. In adults, ~4% of the surface of trabecular bone (such as iliac crest) is involved in active resorption, whereas 10–15% of trabecular surfaces are covered with osteoid, unmineralized new bone formed by osteoblasts. Radioisotope studies indicate that as much as 18% of the total skeletal calcium is deposited and removed each year. Thus, bone is an active metabolizing tissue that requires an intact blood supply. The cycle of bone resorption and formation is a highly orchestrated process, directed by osteocytes and carried out by the basic multicellular unit, which is composed of a group of osteoclasts and osteoblasts **(Fig. 421-2)**.

The response of bone to fractures, infection, and interruption of blood supply and to expanding lesions is relatively limited. Dead bone must be resorbed, and new bone must be formed, a process carried out in association with growth of new blood vessels into the involved area. In injuries that disrupt the organization of the tissue, such as a fracture, in which apposition of fragments is poor or when motion exists at the fracture site, progenitor stromal cells recapitulate the endochondral bone formation of early development and form cartilage that is replaced by bone and, variably, fibrous tissue. When there is good apposition with fixation and little motion at the fracture site, repair occurs predominantly by formation of new bone without other mediating tissue.

Remodeling of bone occurs along lines of force generated by mechanical stress. The signals from these mechanical stresses are sensed by osteocytes, which transmit signals to osteoclasts and osteoblasts or their precursors. One such signal made by osteocytes is sclerostin, an inhibitor of wnt signaling. Mechanical forces, as well as parathyroid hormone, suppress sclerostin production and thus increase bone formation by osteoblasts. Expanding lesions in bone such as tumors induce resorption at the surface in contact with the tumor by producing ligands such as PTHrP that stimulate osteoclast differentiation and function. Thus, bone plasticity reflects the interaction of cells with each other and with the environment.

Measurement of the products of osteoblast and osteoclast activity can assist in the diagnosis and management of bone diseases. Osteoblast activity can be assessed by measuring serum bone-specific alkaline phosphatase. Similarly, osteocalcin, a protein secreted from osteoblasts, is made virtually only by osteoblasts. Measurement of an amino-terminal fragment of procollagen I is also an effective index of bone formation. Osteoclast activity can be assessed by measurement of products of collagen degradation. Collagen molecules are covalently linked to each other in the extracellular matrix through the formation of hydroxypyridinium cross-links **(Chap. 425)**. After digestion by osteoclasts, these cross-linked peptides can be measured both in urine and in blood.

CALCIUM METABOLISM

Over 99% of the 1–2 kg of calcium present normally in the adult human body resides in the skeleton, where it provides mechanical stability and serves as a reservoir when needed to maintain extracellular fluid (ECF) calcium concentration **(Fig. 421-3)**. Skeletal calcium accretion first becomes significant during the third trimester of fetal life, accelerates throughout childhood and adolescence, reaches a peak in early adulthood, and gradually declines thereafter at rates that rarely exceed 1–2% per year. These slow changes in total skeletal calcium content contrast with relatively high daily rates of closely matched fluxes of calcium into and out of bone (~250–500 mg each), a process mediated by coupled activity of osteoblasts, osteoclasts, and osteocytes. Another 0.5–1% of skeletal calcium is freely exchangeable (e.g., in chemical equilibrium) with that in the ECF.

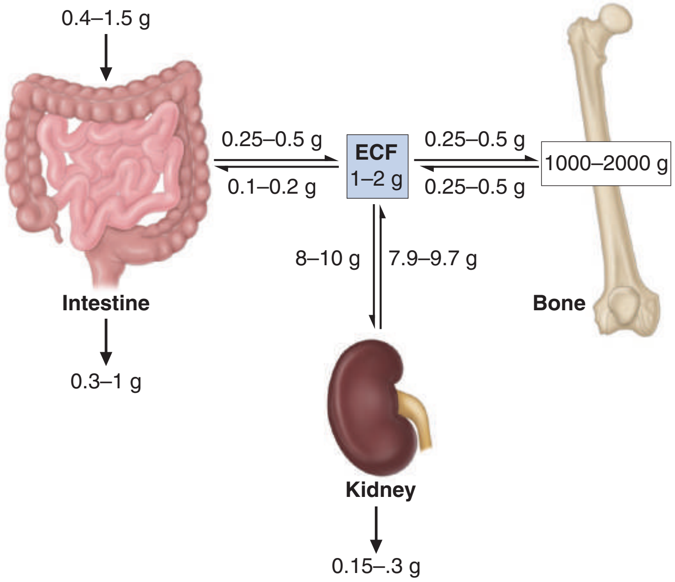

FIGURE 421-3 Calcium homeostasis. Schematic illustration of calcium content of extracellular fluid (ECF) and bone as well as of diet and feces; magnitude of calcium flux per day as calculated by various methods is shown at sites of transport in intestine, kidney, and bone. Ranges of values shown are approximate and were chosen to illustrate certain points discussed in the text. In conditions of calcium balance, rates of calcium release from and uptake into bone are equal.

The concentration of ionized calcium in the ECF must be maintained within a narrow range because of the critical role calcium plays in a wide array of cellular functions, especially those involved in neuromuscular activity, secretion, and signal transduction. Intracellular cytosolic free calcium levels are ~100 nmol/L and are 10,000-fold lower than ionized calcium concentration in the blood and ECF (1.1–1.3 mmol/L). Cytosolic calcium does not play the structural role played by extracellular calcium; instead, it serves a signaling function. The steep chemical gradient of calcium from outside to inside the cell promotes rapid calcium influx through various membrane calcium channels that can be activated by hormones, metabolites, or neurotransmitters, swiftly changing cellular function. In blood, total calcium concentration is normally 2.2–2.6 mM (8.5–10.5 mg/dL), of which ~50% is ionized. The remainder is bound ionically to negatively charged proteins (predominantly albumin and immunoglobulins) or loosely complexed with phosphate, citrate, sulfate, or other anions. Alterations in serum protein concentrations directly affect the total blood calcium concentration even if the ionized calcium concentration remains normal. An algorithm to correct for protein changes adjusts the total serum calcium (in mg/dL) upward by 0.8 times the deficit in serum albumin (g/dL) or by 0.5 times the deficit in serum immunoglobulin (in g/dL). Notably, such corrections provide only rough approximations of actual free calcium concentrations, however, and may be misleading, particularly during acute illness. Acidosis also alters ionized calcium by reducing its association with proteins. Accordingly, the best practice is to measure blood ionized calcium directly by a method that employs calcium-selective electrodes in acute settings during which calcium abnormalities might occur.

Control of the ionized calcium concentration in the ECF ordinarily is accomplished by adjusting the rates of calcium movement across intestinal and renal epithelia and into and out of bone. These adjustments are mediated mainly via changes in blood levels of the hormones PTH and $1,25(OH)_2D$. Acting via binding to calcium-sensing receptors (CaSRs) on the surface of parathyroid cells, blood ionized calcium suppresses PTH secretion by reducing levels of PTH mRNA, promoting the cleavage of PTH to inactive peptides, and suppressing release of PTH-containing granules from parathyroid cells. $1,25(OH)_2D$ inhibits PTH production in the parathyroid by an incompletely understood mechanism of negative feedback **(Chap. 422)**.

Normal dietary calcium intake in the United States varies widely, ranging from 10 to 37 mmol/d (400–1500 mg/d). A National Academy of Medicine (formerly, Institute of Medicine) analysis recommends a daily allowance of 25–30 mmol (1000–1200 mg) for most adults. Intestinal absorption of ingested calcium involves both active (transcellular) and passive (paracellular) mechanisms. Passive calcium absorption is nonsaturable and approximates 5% of daily calcium intake, whereas active absorption involves apical calcium entry via specific ion channels (TRPV5 in the kidney's distal tubule and TRPV6 in the intestine), whose expression is controlled principally by $1,25(OH)_2D$. This active transport mechanism normally accounts for absorption of 20–70% of dietary calcium. Active gastrointestinal calcium transport occurs mainly in the proximal small bowel (duodenum and proximal jejunum), although some active calcium absorption occurs in most segments of the small intestine. Optimal rates of calcium absorption require gastric acid. This is especially true for weakly dissociable calcium supplements such as calcium carbonate. In fact, large boluses of calcium carbonate are poorly absorbed because of their neutralizing effect on gastric acid. In achlorhydric subjects and for those individuals taking drugs that inhibit gastric acid secretion, or with diminished acid secretion following bariatric surgery, supplements should be taken with meals to optimize their absorption. Use of calcium citrate may be preferable in these circumstances. Calcium absorption may also be blunted in disease states such as pancreatic or biliary insufficiency, in which ingested calcium remains bound to unabsorbed fatty acids or other food constituents. At high levels of calcium intake, synthesis of $1,25(OH)_2D$ is reduced; this decreases the rate of active intestinal calcium absorption. The opposite occurs with dietary calcium restriction. Some calcium, ~2.5–5 mmol/d (100–200 mg/d), is excreted as an obligate component of intestinal secretions and is not regulated by calciotropic hormones.

The feedback-controlled hormonal regulation of intestinal absorptive efficiency results in a relatively constant daily net calcium absorption of ~5–10 mmol/d (200–400 mg/d) despite large changes in daily dietary calcium intake. This daily load of absorbed calcium is excreted by the kidneys in a manner that is also tightly regulated by the concentration of ionized calcium in the blood. Approximately 8–10 g/d of calcium is filtered by the glomeruli, of which only 2–3% appears in the urine. Most filtered calcium (65%) is reabsorbed in the proximal tubules via a passive, paracellular route that is coupled to concomitant NaCl reabsorption and not specifically regulated. The cortical thick ascending limb of Henle's loop (cTAL) reabsorbs roughly another 20% of filtered calcium, also via a paracellular mechanism. Calcium reabsorption in the cTAL requires a tight-junctional proteins called paracellin-1 and Claudin14 and is inhibited by increased blood concentrations of calcium or magnesium, acting via the CaSR, which is highly expressed on basolateral membranes in this nephron segment. Operation of the renal CaSR provides a mechanism, independent of those engaged directly by PTH or $1,25(OH)_2D$, by which serum ionized calcium can control renal calcium reabsorption. Finally, ~10% of filtered calcium is reabsorbed in the distal convoluted tubules (DCTs) by a highly regulated transcellular mechanism. Calcium enters the luminal surface of the cell through specific apical calcium channels (TRPV5). It then moves across the cell in association with a specific calcium-binding protein (calbindin-D28k) that buffers cytosolic calcium concentrations from the large mass of transported calcium. Basolateral Ca^{2+}-ATPases and Na^+/Ca^{2+} exchangers actively extrude calcium and thereby maintain the transcellular calcium gradient. All these processes are stimulated directly or indirectly by PTH and $1,25(OH)_2D$. The DCT is also the site of action of thiazide diuretics, which lower urinary calcium excretion by inducing sodium depletion and thereby augmenting proximal calcium reabsorption. Conversely, dietary sodium loads, or increased distal sodium delivery caused by loop diuretics or saline infusion, induce calciuresis.

The homeostatic mechanisms that normally maintain a constant serum ionized calcium concentration may fail at extremes of calcium intake or when the hormonal systems or organs involved are compromised. Thus, even with maximal activity of the vitamin D–dependent intestinal active transport system, sustained calcium intake <5 mmol/d (<200 mg/d) cannot provide enough net calcium absorption to replace obligate losses via the intestine, the kidney, sweat, and other secretions. In this case, increased blood levels of PTH and $1,25(OH)_2D$ activate osteoclastic bone resorption to obtain needed calcium from bone, which leads to progressive bone loss and negative calcium balance. Increased PTH and $1,25(OH)_2D$ also enhance renal calcium reabsorption, and $1,25(OH)_2D$ enhances calcium absorption in the gut. At very high calcium intakes (>100 mmol/d [>4 g/d]), passive intestinal absorption continues to deliver calcium into the ECF despite maximally downregulated intestinal active transport and renal tubular calcium reabsorption. This can cause severe hypercalciuria, nephrocalcinosis, progressive renal failure, and hypercalcemia (e.g., "milk-alkali syndrome"). Deficiency or excess of PTH or vitamin D, intestinal disease, and renal failure represent other commonly encountered challenges to normal calcium homeostasis **(Chap. 422)**.

PHOSPHORUS METABOLISM

Although 85% of the ~600 g of body phosphorus is present in bone mineral, phosphorus is also a major intracellular constituent both as the free anion(s) and as a component of numerous organophosphate compounds, including structural proteins, enzymes, transcription factors, carbohydrate and lipid intermediates, high-energy stores (ATP [adenosine triphosphate], creatine phosphate), and nucleic acids. Unlike calcium, phosphorus exists intracellularly at concentrations close to those present in ECF (e.g., 1–2 mmol/L). In cells and in the ECF, phosphorus exists in several forms, predominantly as $H_2PO_4^-$ or $NaHPO_4^-$, with perhaps 10% as HPO_4^{2-}. This mixture of anions will be referred to here as "phosphate." In serum, ~12% of phosphorus is bound to proteins. Concentrations of phosphates in blood and ECF generally are expressed in terms of elemental phosphorus, with the normal range in adults being 0.75–1.45 mmol/L (2.5–4.5 mg/dL).

Because the volume of the intracellular fluid compartment is twice that of the ECF, measurements of ECF phosphate may not accurately reflect phosphate availability within cells that follows even modest shifts of phosphate from one compartment to the other.

Phosphate is widely available in foods and is absorbed efficiently (65%) by the small intestine even in the absence of vitamin D. However, gut phosphate absorptive efficiency may be enhanced (to 85–90%) via active transport mechanisms that are stimulated by $1,25(\text{OH})_2\text{D}$. These mechanisms involve activation of $\text{Na}^+/\text{PO}_4^{2-}$ co-transporters, such as Npt2b, that move phosphate into intestinal cells against an unfavorable electrochemical gradient. Daily net intestinal phosphate absorption varies widely with the composition of the diet but is generally in the range of 500–1000 mg/d. Phosphate absorption can be inhibited by large doses of calcium salts or by sevelamer hydrochloride (Renagel), strategies commonly used to control levels of serum phosphate in chronic kidney disease. Aluminum hydroxide antacids also reduce phosphate absorption but are used less commonly because of the potential for aluminum toxicity. Low serum phosphate stimulates renal proximal tubular synthesis of $1,25(\text{OH})_2\text{D}$, perhaps by suppressing blood levels of FGF23 (see below).

Serum phosphate levels vary by as much as 50% on a normal day. This reflects the effect of food intake but also an underlying circadian rhythm that produces a nadir between 7 and 10 A.M. Carbohydrate administration, especially as IV dextrose solutions in fasting subjects, can decrease serum phosphate by >0.7 mmol/L (2 mg/dL) in common clinical settings including treatment of ketoacidosis or metabolic or respiratory alkalosis. Because of this wide variation in serum phosphate, it is best to perform measurements in the basal, fasting state.

Control of serum phosphate is determined mainly by the rate of renal tubular reabsorption of the filtered load, which is ~4–6 g/d. Because intestinal phosphate absorption is highly efficient, urinary excretion is not constant but varies directly with dietary intake. The fractional excretion of phosphate (ratio of phosphate to creatinine clearance) is generally in the range of 10–15%. The proximal tubule is the principal site at which renal phosphate reabsorption is regulated. This is accomplished by changes in the levels of apical expression and activity of specific $\text{Na}^+/\text{PO}_4^{2-}$ co-transporters (NaPi-2a and NaPi-2c) in the proximal tubule. Levels of these transporters at the apical surface of these cells are reduced rapidly by PTH, a major hormonal regulator of renal phosphate excretion. In addition, the circulating hormone FGF23 can inhibit phosphate reabsorption by a similar mechanism. FGF23 is synthesized by osteocytes. Activating *FGF23* mutations cause the rare disorder autosomal dominant hypophosphatemic rickets (ADHR). In contrast to PTH, FGF23 reduces synthesis of $1,25(\text{OH})_2\text{D}$, which may worsen the resulting hypophosphatemia by lowering intestinal phosphate absorption. Renal reabsorption of phosphate is responsive to changes in dietary intake such that experimental dietary phosphate restriction leads to a dramatic lowering of urinary phosphate within hours, preceding any decline in serum phosphate (e.g., filtered load). This physiologic renal adaptation to changes in dietary phosphate availability occurs independently of PTH and may be mediated in part by changes in levels of serum FGF23. Findings in *FGF23*-deficient mice suggest that FGF23 normally acts to lower blood phosphate and $1,25(\text{OH})_2\text{D}$ levels. In turn, elevation of blood phosphate and $1,25(\text{OH})_2\text{D}$ increases blood levels of FGF23.

Renal phosphate reabsorption is impaired by hypocalcemia, hypomagnesemia, and severe hypophosphatemia. Phosphate clearance is enhanced by ECF volume expansion and impaired by dehydration. Phosphate retention is an important pathophysiologic feature of renal insufficiency **(Chap. 322)**.

HYPOPHOSPHATEMIA

Causes Hypophosphatemia can occur by one or more of three primary mechanisms: (1) inadequate intestinal phosphate absorption, (2) excessive renal phosphate excretion, and (3) rapid redistribution of phosphate from the ECF into bone or soft tissue **(Table 421-1)**. Because phosphate is so abundant in foods, inadequate intestinal absorption is almost never observed now that aluminum hydroxide antacids, which

TABLE 421-1 Causes of Hypophosphatemia

- I. Reduced renal tubular phosphate reabsorption
  - A. PTH/PTHrP-dependent
    - 1. Primary hyperparathyroidism
    - 2. Secondary hyperparathyroidism
      - a. Vitamin D deficiency/resistance
      - b. Calcium starvation/malabsorption
      - c. Bartter's syndrome
      - d. Autosomal recessive renal hypercalciuria with hypomagnesemia
    - 3. PTHrP-dependent hypercalcemia of malignancy
    - 4. Familial hypocalciuric hypercalcemia
  - B. PTH/PTHrP-independent
    - 1. Excess FGF23 or other "phosphatonins"
      - a. X-linked hypophosphatemic rickets (XLH)
      - b. Autosomal recessive hypophosphatemia (ARHP)
      - c. Autosomal dominant hypophosphatemic rickets (ADHR) (DMP1, ENPP1 deficiency)
      - d. Tumor-induced osteomalacia syndrome (TIO)
      - e. McCune-Albright syndrome (fibrous dysplasia)
      - f. Epidermal nevus syndrome
    - 2. Intrinsic renal disease
      - a. Fanconi's syndrome(s)
      - b. Cystinosis
      - c. Wilson's disease
      - d. NaPi-2a or NaPi-2c mutations
    - 3. Other systemic disorders
      - a. Poorly controlled diabetes mellitus
      - b. Alcoholism
      - c. Hyperaldosteronism
      - d. Hypomagnesemia
      - e. Amyloidosis
      - f. Hemolytic-uremic syndrome
      - g. Renal transplantation or partial liver resection
      - h. Rewarming or induced hyperthermia
    - 4. Drugs or toxins
      - a. Ethanol
      - b. Acetazolamide, other diuretics
      - c. High-dose estrogens or glucocorticoids
      - d. Heavy metals (lead, cadmium, saccharated ferric oxide)
      - e. Toluene, *N*-methyl formamide
      - f. Cisplatin, ifosfamide, foscarnet, rapamycin
- II. Impaired intestinal phosphate absorption
  - A. Aluminum-containing antacids
  - B. Sevalamer
- III. Shifts of extracellular phosphate into cells
  - A. Intravenous glucose
  - B. Insulin therapy for prolonged hyperglycemia or diabetic ketoacidosis
  - C. Catecholamines (epinephrine, dopamine, albuterol)
  - D. Acute respiratory alkalosis
  - E. Gram-negative sepsis, toxic shock syndrome
  - F. Recovery from starvation or acidosis
  - G. Rapid cellular proliferation
    - 1. Leukemic blast crisis
    - 2. Intensive erythropoietin, other growth factor therapy
- IV. Accelerated net bone formation
  - A. After parathyroidectomy
  - B. Treatment of vitamin D deficiency, Paget's disease
  - C. Osteoblastic metastases

Abbreviations: PTH, parathyroid hormone; PTHrP, parathyroid hormone–related peptide.

bind phosphate in the gut, are no longer widely used. Fasting or starvation, however, may result in depletion of body phosphate and predispose to subsequent hypophosphatemia during refeeding, especially if this is accomplished with IV glucose alone.

Chronic hypophosphatemia usually signifies the presence of a persistent renal tubular phosphate-wasting disorder. Excessive activation of PTH/PTHrP receptors in the proximal tubule, as a result of primary or secondary hyperparathyroidism or because of the PTHrP-mediated hypercalcemia syndrome in malignancy **(Chap. 422)**, is a common cause of renal hypophosphatemia. Familial hypocalciuric hypercalcemia and Jansen's metaphyseal chondrodysplasia are rare examples of genetic disorders in this category **(Chap. 422)**.

Several genetic and acquired diseases cause PTH/PTHrP receptor-independent tubular phosphate wasting with associated rickets and osteomalacia. All these diseases manifest severe hypophosphatemia; renal phosphate wasting, sometimes accompanied by aminoaciduria; inappropriately low blood levels of 1,25(OH)$_2$D; low-normal serum levels of calcium; and evidence of impaired cartilage or bone mineralization (osteomalacia). Analysis of these diseases in patients *without* generalized proximal tubular defects (Fanconi syndrome) led to the discovery of the hormone FGF23, which is an important physiologic regulator of phosphate metabolism. FGF23 decreases phosphate reabsorption in the proximal tubule and also suppresses the 1α-hydroxylase responsible for synthesis of 1,25(OH)$_2$D. FGF23 is synthesized by cells of the osteoblast lineage, primarily osteocytes. High-phosphate diets increase FGF23 levels, and low-phosphate diets decrease them. ADHR was the first disease linked to abnormalities in FGF23. ADHR results from activating mutations in the gene that encodes FGF23. These mutations alter a cleavage site that ordinarily allows for inactivation of intact FGF23. Several other genetic disorders lead to elevated FGF23, hypophosphatemia, and osteomalacia. The most common of these is X-linked hypophosphatemic rickets (XLH), which results from inactivating mutations in an endopeptidase termed *PHEX* (*p*hosphate-regulating gene with *h*omologies to *e*ndopeptidases on the *X* chromosome) that is expressed most abundantly on the surface of osteocytes and mature osteoblasts. Patients with XLH usually have high FGF23 levels, and ablation of the *FGF23* gene reverses the hypophosphatemia found in the mouse version of XLH. How inactivation of PHEX leads to increased levels of FGF23 has not been determined. Two rare autosomal recessive hypophosphatemic syndromes associated with elevated FGF23 are due to inactivating mutations of dentin matrix protein 1 (DMP1) and ectonucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1), respectively, both of which normally are highly expressed in bone and presumably regulate FGF23 production. An unusual hypophosphatemic disorder, tumor-induced osteomalacia (TIO), is an acquired disorder in which tumors, usually of mesenchymal origin, secrete FGF23. The hypophosphatemic syndrome resolves completely within hours to days after successful resection of the responsible tumor. Such tumors typically express large amounts of FGF23 mRNA, and patients with TIO usually exhibit elevations of FGF23 in their blood. Neutralizing antibodies against FGF23 can be used to treat hypophosphatemia in patients with XLH and TIO.

Dent's disease is an X-linked recessive disorder caused by inactivating mutations in *CLCN5*, a chloride transporter expressed in endosomes of the proximal tubule; features include hypercalciuria, hypophosphatemia, and recurrent kidney stones. Renal phosphate wasting is common among poorly controlled diabetic patients and alcoholics, who therefore are at risk for iatrogenic hypophosphatemia when treated with insulin or IV glucose, respectively. Diuretics and certain other drugs and toxins can cause defective renal tubular phosphate reabsorption (Table 421-1).

In hospitalized patients, hypophosphatemia is often attributable to massive redistribution of phosphate from the ECF into cells. Insulin therapy for diabetic ketoacidosis is a paradigm for this phenomenon, in which the severity of the hypophosphatemia is related to the extent of antecedent depletion of phosphate and other electrolytes **(Chap. 416)**. The hypophosphatemia is usually greatest at a point many hours after initiation of insulin therapy and is difficult to predict from baseline measurements of serum phosphate at the time of presentation, when prerenal azotemia can obscure significant phosphate depletion. Other factors that may contribute to such acute redistributive hypophosphatemia include antecedent starvation or malnutrition, administration of IV glucose without other nutrients, elevated blood catecholamines (endogenous or exogenous), respiratory alkalosis, and recovery from metabolic acidosis.

Hypophosphatemia also can occur transiently (over weeks to months) during the phase of accelerated net bone formation that follows parathyroidectomy for severe primary hyperparathyroidism or during treatment of vitamin D deficiency or lytic Paget's disease. This is referred to as "hungry bone syndrome" and is most prominent in patients who preoperatively have evidence of very high bone turnover (e.g., elevated serum levels of alkaline phosphatase). Osteoblastic metastases can also lead to this syndrome.

Clinical and Laboratory Findings The clinical manifestations of severe hypophosphatemia reflect a generalized defect in cellular energy metabolism because of ATP depletion, a shift from oxidative phosphorylation toward glycolysis, and associated tissue or organ dysfunction. Acute, severe hypophosphatemia occurs mainly or exclusively in hospitalized patients with underlying serious medical or surgical illness and preexisting phosphate depletion due to excessive urinary losses, severe malabsorption, or malnutrition. Chronic hypophosphatemia tends to be less severe, with a clinical presentation dominated by musculoskeletal complaints such as bone pain, osteomalacia, pseudofractures, and proximal muscle weakness or, in children, rickets and short stature.

Neuromuscular manifestations of severe hypophosphatemia are variable but may include muscle weakness, lethargy, confusion, dysarthria, dysphagia, oculomotor palsies, nystagmus, ataxia, hyporeflexia, impaired sphincter control, paresthesia, generalized or Guillain-Barré–like ascending paralysis, seizures, coma, and even death. Serious sequelae such as paralysis and seizures are likely only at phosphate concentrations <0.25 mmol/L (<0.8 mg/dL). Rhabdomyolysis may develop during rapidly progressive hypophosphatemia. The diagnosis of hypophosphatemia-induced rhabdomyolysis may be overlooked, as up to 30% of patients with acute hypophosphatemia (<0.7 m*M*) have creatine phosphokinase elevations that peak 1–2 days after the nadir in serum phosphate, when the release of phosphate from injured myocytes may have led to a near normalization of circulating levels of phosphate.

Respiratory failure and cardiac dysfunction, which are reversible with phosphate treatment, may occur at serum phosphate levels of 0.5–0.8 mmol/L (1.5–2.5 mg/dL). Renal tubular defects, including tubular acidosis, glycosuria, and impaired reabsorption of sodium and calcium, may occur. Hematologic abnormalities correlate with reductions in intracellular ATP and 2,3-diphosphoglycerate and may include erythrocyte microspherocytosis and hemolysis; impaired oxyhemoglobin dissociation; defective leukocyte chemotaxis, phagocytosis, and bacterial killing; and platelet dysfunction.

TREATMENT

Hypophosphatemia

Severe hypophosphatemia (<0.75 mmol/L [<2 mg/dL]), particularly in the setting of underlying phosphate depletion, constitutes a dangerous electrolyte abnormality that should be corrected promptly. Unfortunately, the cumulative deficit in body phosphate cannot be predicted directly from knowledge of the circulating level of phosphate, and therapy must be approached empirically. The threshold for IV phosphate therapy and consequently the dose of phosphate to be administered should reflect consideration of renal function, the likely severity and duration of the underlying phosphate depletion, and the presence and severity of symptoms consistent with those of hypophosphatemia. In adults, phosphate may be safely administered IV as neutral mixtures of sodium or potassium phosphate salts at initial doses of 0.2–0.8 mmol/kg of elemental phosphorus over 6 h (e.g., 10–50 mmol over 6 h), with doses >20 mmol/6 h reserved for those who have serum levels

TABLE 421-2 Intravenous Therapy for Hypophosphatemia

CONSIDER

Likely severity of underlying phosphate depletion

Concurrent parenteral glucose administration

Presence of neuromuscular, cardiopulmonary, or hematologic complications of hypophosphatemia

Renal function (reduce dose by 50% if serum creatinine >220 μmol/L [>2.5 mg/dL])

Serum calcium level (correct hypocalcemia first; reduce dose by 50% in hypercalcemia)

Guidelines

SERUM PHOSPHORUS, MM (MG/DL)	RATE OF INFUSION, MMOL/H	DURATION, H	TOTAL ADMINISTERED, MMOL
<0.8 (<2.5)	2	6	12
<0.5 (<1.5)	4	6	24
<0.3 (<1)	8	6	48

Note: Rates shown are calculated for a 70-kg person; levels of serum calcium and phosphorus must be measured every 6–12 h during therapy; infusions can be repeated to achieve stable serum phosphorus levels >0.8 mmol/L (>2.5 mg/dL); most formulations available in the United States provide 3 mmol/mL of sodium or potassium phosphate.

<0.5 mmol/L (1.5 mg/dL) and normal renal function. A suggested approach is presented in Table 421-2. Serum levels of phosphate and calcium must be monitored closely (every 6–12 h) throughout treatment. It is important to avoid a serum calcium-phosphorus product >50 mg^2/dL^2 to reduce the risk of heterotopic calcification. Hypocalcemia, if present, should be corrected before administering IV phosphate. Less severe hypophosphatemia, in the range of 0.5–0.8 mmol/L (1.5–2.5 mg/dL), usually can be treated with oral phosphate in divided doses of 750–2000 mg/d as elemental phosphorus; higher doses can cause bloating and diarrhea.

Management of chronic hypophosphatemia requires knowledge of the cause(s) of the disorder. Hypophosphatemia related to the secondary hyperparathyroidism of vitamin D deficiency usually responds to treatment with vitamin D and calcium alone. XLH, ADHR, TIO, and related renal tubular disorders usually are managed with divided oral doses of phosphate, often with calcium and 1,25(OH)$_2$D supplements to bypass the block in renal 1,25(OH)$_2$D synthesis and prevent secondary hyperparathyroidism caused by suppression of ECF calcium levels. Care must be taken to be sure that oral calcium and phosphate are not administered at the same time, to avoid precipitation before absorption. Thiazide diuretics may be used to prevent nephrocalcinosis in patients who are managed this way. Complete normalization of hypophosphatemia is generally not possible in these conditions. Burosumab, a human monoclonal antibody that inhibits FGF23, has been approved for the treatment of XLH and TIO. It corrects hypophosphatemia, improves bone pain, and heals fractures in both adults and children.

Optimal therapy for TIO is surgical removal of the responsible tumor, which may be localized by radiographic skeletal survey or bone scan (many are located in bone) or by radionuclide scanning using sestamibi or labeled octreotide. Successful treatment of TIO-induced hypophosphatemia with octreotide has been reported in a small number of patients. Burosumab treatment can be used to treat hypophosphatemia in patients with TIO in whom tumors cannot be localized or removed.

HYPERPHOSPHATEMIA

Causes When the filtered load of phosphate and glomerular filtration rate (GFR) are normal, control of serum phosphate levels is achieved by adjusting the rate at which phosphate is reabsorbed by the proximal tubular NaPi-2 co-transporters. Hyperphosphatemia, defined in adults as a fasting serum phosphate concentration >1.8 mmol/L (5.5 mg/dL), usually results from impaired glomerular filtration, hypoparathyroidism, excessive delivery of phosphate into the ECF (from bone, gut, or parenteral phosphate therapy), or a combination of these factors (Table 421-3). The upper limit of normal serum phosphate concentrations is higher in children and neonates (2.4 mmol/L [7 mg/dL]). It is useful to distinguish hyperphosphatemia caused by impaired renal phosphate excretion from that which results from excessive delivery of phosphate into the ECF (Table 421-3).

TABLE 421-3 Causes of Hyperphosphatemia

- I. Impaired renal phosphate excretion
    - A. Renal insufficiency
    - B. Hypoparathyroidism
        - 1. Developmental
        - 2. Autoimmune
        - 3. After neck surgery or radiation
        - 4. Activating mutations of the calcium-sensing receptor
    - C. Parathyroid suppression
        - 1. Parathyroid-independent hypercalcemia
            - a. Vitamin D or vitamin A intoxication
            - b. Sarcoidosis, other granulomatous diseases
            - c. Immobilization, osteolytic metastases
            - d. Milk-alkali syndrome
        - 2. Severe hypermagnesemia or hypomagnesemia
    - D. Pseudohypoparathyroidism
    - E. Acromegaly
    - F. Tumoral calcinosis
    - G. Heparin therapy
- II. Massive extracellular fluid phosphate loads
    - A. Rapid administration of exogenous phosphate (intravenous, oral, rectal)
    - B. Extensive cellular injury or necrosis
        - 1. Crush injuries
        - 2. Rhabdomyolysis
        - 3. Hyperthermia
        - 4. Fulminant hepatitis
        - 5. Cytotoxic therapy
        - 6. Severe hemolytic anemia
    - C. Transcellular phosphate shifts
        - 1. Metabolic acidosis
        - 2. Respiratory acidosis

In chronic renal insufficiency, reduced GFR leads to phosphate retention. Hyperphosphatemia, and progressive loss of nephron function, in turn further impairs renal synthesis of 1,25(OH)$_2$D, increases FGF23 levels, and stimulates PTH secretion and parathyroid gland hypertrophy both directly and indirectly (by lowering blood ionized calcium levels). Thus, hyperphosphatemia is a major cause of the secondary hyperparathyroidism of renal failure (Chaps. 322 and 422).

Hypoparathyroidism leads to hyperphosphatemia via increased expression of NaPi-2 co-transporters in the proximal tubule. Hypoparathyroidism, or parathyroid suppression, has multiple potential causes, including autoimmune disease; developmental, surgical, or radiation-induced absence of functional parathyroid tissue; vitamin D intoxication or other causes of PTH-independent hypercalcemia; cellular PTH resistance (pseudohypoparathyroidism or hypomagnesemia); infiltrative disorders such as Wilson's disease and hemochromatosis; and impaired PTH secretion caused by hypermagnesemia, severe hypomagnesemia, or activating mutations in the CaSR. Hypocalcemia may also contribute directly to impaired phosphate clearance, as calcium infusion can induce phosphaturia in hypoparathyroid subjects. Increased tubular phosphate reabsorption also occurs in acromegaly, during heparin administration, and in tumoral calcinosis. Tumoral calcinosis is caused by a rare group of genetic disorders in which FGF23 is processed in a way that leads to low levels of active FGF23 in the bloodstream. This may result from mutations in the FGF23 sequence or via inactivating mutations in the *GALNT3* gene, which encodes a galactosaminyl transferase that normally adds sugar residues to FGF23 that slow its proteolysis. A similar syndrome results from FGF23 resistance due to inactivating mutations of the FGF23 co-receptor Klotho. These abnormalities cause elevated serum 1,25(OH)$_2$D, parathyroid suppression, increased intestinal calcium absorption, and focal hyperostosis with large, lobulated periarticular heterotopic ossifications (especially at shoulders or hips) and are accompanied by hyperphosphatemia. In some forms of tumoral calcinosis, serum phosphorus levels are normal.

When large amounts of phosphate are delivered rapidly into the ECF, hyperphosphatemia can occur despite normal renal function. Examples include overzealous IV phosphate therapy, oral or rectal administration of large amounts of phosphate-containing laxatives or enemas (especially in children), extensive soft tissue injury or necrosis (crush injuries, rhabdomyolysis, hyperthermia, fulminant hepatitis, cytotoxic chemotherapy), extensive hemolytic anemia, and transcellular phosphate shifts induced by severe metabolic or respiratory acidosis.

Clinical Findings The clinical consequences of acute, severe hyperphosphatemia are due mainly to the formation of widespread calcium phosphate precipitates and resulting hypocalcemia. Thus, tetany, seizures, accelerated nephrocalcinosis (with renal failure, hyperkalemia, hyperuricemia, and metabolic acidosis), and pulmonary or cardiac calcifications (including development of acute heart block) may occur. The severity of these complications relates to the elevation of serum phosphate levels, which can reach concentrations as high as 7 mmol/L (20 mg/dL) in instances of massive soft tissue injury or tumor lysis syndrome.

TREATMENT

Hyperphosphatemia

Therapeutic options for management of severe hyperphosphatemia are limited. Volume expansion may enhance renal phosphate clearance. Aluminum hydroxide antacids or sevelamer may be helpful in chelating and limiting absorption of offending phosphate salts present in the intestine. Hemodialysis is the most effective therapeutic strategy and should be considered early in the course of severe hyperphosphatemia, especially in the setting of renal failure and symptomatic hypocalcemia.

MAGNESIUM METABOLISM

Magnesium is the major intracellular divalent cation. Normal concentrations of extracellular magnesium and calcium are crucial for normal neuromuscular activity. Intracellular magnesium forms a key complex with ATP and is an important cofactor for a wide range of enzymes, transporters, and nucleic acids required for normal cellular function, replication, and energy metabolism. The concentration of magnesium in serum is closely regulated within the range of 0.7–1 mmol/L (1.5–2 meq/L; 1.7–2.4 mg/dL), of which 30% is protein-bound and another 15% is loosely complexed to phosphate and other anions. One-half of the 25 g (1000 mmol) of total body magnesium is located in bone, only one-half of which is insoluble in the mineral phase. Almost all extraskeletal magnesium is present within cells, where the total concentration is 5 m*M*, 95% of which is bound to proteins and other macromolecules. Because only 1% of body magnesium resides in the ECF, measurements of serum magnesium levels may not accurately reflect the level of total body magnesium stores.

Dietary magnesium content normally ranges from 6 to 15 mmol/d (140–360 mg/d), of which 30–40% is absorbed, mainly in the jejunum and ileum. Intestinal magnesium absorptive efficiency is stimulated by 1,25(OH)$_2$D and can reach 70% during magnesium deprivation. Urinary magnesium excretion normally matches net intestinal absorption and is ~4 mmol/d (100 mg/d). Regulation of serum magnesium concentrations is achieved mainly by control of renal magnesium reabsorption. Only 20% of filtered magnesium is reabsorbed in the proximal tubule, whereas 60% is reclaimed in the cTAL and another 5–10% in the DCT. Magnesium reabsorption in the cTAL occurs via a paracellular route that requires both a lumen-positive potential, created by NaCl reabsorption, and tight-junction proteins encoded by members of the Claudin gene family. Magnesium reabsorption in the cTAL is increased by PTH but inhibited by hypercalcemia or hypermagnesemia, both of which activate the CaSR in this nephron segment.

HYPOMAGNESEMIA

Causes Hypomagnesemia usually signifies substantial depletion of body magnesium stores (0.5–1 mmol/kg). Hypomagnesemia can result from intestinal malabsorption; protracted vomiting, diarrhea, or intestinal drainage; defective renal tubular magnesium reabsorption; or rapid shifts of magnesium from the ECF into cells, bone, or third spaces (Table 421-4). Dietary magnesium deficiency is unlikely except possibly in the setting of alcoholism. Rare genetic disorders that cause selective intestinal magnesium malabsorption have been described (primary infantile hypomagnesemia types 1 and 2). Another rare

TABLE 421-4 Causes of Hypomagnesemia

I. Impaired intestinal absorption
  A. Hypomagnesemia with secondary hypocalcemia (*TRPM6* mutations)
  B. Malabsorption syndromes
  C. Vitamin D deficiency
  D. Proton pump inhibitors
II. Increased intestinal losses
  A. Protracted vomiting/diarrhea
  B. Intestinal drainage, fistulas
III. Impaired renal tubular reabsorption
  A. Genetic magnesium-wasting syndromes
    1. Gitelman's syndrome
    2. Bartter's syndrome
    3. Claudin 16 or 19 mutations
    4. Potassium channel mutations (Kv1.1, Kir4.1)
    5. Na^+,K^+-ATPase γ-subunit mutations (FXYD2)
  B. Acquired renal disease
    1. Tubulointerstitial disease
    2. Postobstruction, ATN (diuretic phase)
    3. Renal transplantation
  C. Drugs and toxins
    1. Ethanol
    2. Diuretics (loop, thiazide, osmotic)
    3. Cisplatin
    4. Pentamidine, foscarnet
    5. Cyclosporine
    6. Aminoglycosides, amphotericin B
    7. Cetuximab
  D. Other
    1. Extracellular fluid volume expansion
    2. Hyperaldosteronism
    3. SIADH
    4. Diabetes mellitus
    5. Hypercalcemia
    6. Phosphate depletion
    7. Metabolic acidosis
    8. Hyperthyroidism
IV. Rapid shifts from extracellular fluid
  A. Intracellular redistribution
    1. Recovery from diabetic ketoacidosis
    2. Refeeding syndrome
    3. Correction of respiratory acidosis
    4. Catecholamines
  B. Accelerated bone formation
    1. Post-parathyroidectomy
    2. Treatment of vitamin D deficiency
    3. Osteoblastic metastases
  C. Other
    1. Pancreatitis, burns, excessive sweating
    2. Pregnancy (third trimester) and lactation

Abbreviations: ATN, acute tubular necrosis; SIADH, syndrome of inappropriate antidiuretic hormone.

inherited disorder (hypomagnesemia with secondary hypocalcemia) is caused by mutations in the gene encoding TRPM6, a protein that, with TRPM7, forms a channel important for both intestinal and distal-tubular renal transcellular magnesium transport. Malabsorptive states, often compounded by vitamin D deficiency, can critically limit magnesium absorption and produce hypomagnesemia despite the compensatory effects of secondary hyperparathyroidism and of hypocalcemia and hypomagnesemia to enhance cTAL magnesium reabsorption. Diarrhea or surgical drainage fluid may contain ≥5 mmol/L of magnesium. Proton pump inhibitors (omeprazole and others) may produce hypomagnesemia by an unknown mechanism that does not involve renal wasting of magnesium.

Several genetic magnesium-wasting syndromes have been described, including inactivating mutations of genes encoding the DCT NaCl co-transporter (Gitelman's syndrome), proteins required for cTAL Na-K-2Cl transport (Bartter's syndrome), claudin 16 or claudin 19 (autosomal recessive renal hypomagnesemia with hypercalciuria), a DCT Na^+,K^+-ATPase γ-subunit (autosomal dominant renal hypomagnesemia with hypocalciuria), DCT K^+ channels (Kv1.1, Kir4.1), and a mitochondrial gene encoding a tRNA. Activating mutations of the CaSR can cause hypomagnesemia as well as hypocalcemia. ECF expansion, hypercalcemia, and severe phosphate depletion may impair magnesium reabsorption, as can various forms of renal injury, including those caused by drugs such as cisplatin, cyclosporine, aminoglycosides, and pentamidine as well as the epidermal growth factor (EGF) receptor inhibitory antibody cetuximab (EGF action is required for normal DCT apical expression of TRPM6) (Table 421-4). A rising blood concentration of ethanol directly impairs tubular magnesium reabsorption, and persistent glycosuria with osmotic diuresis leads to magnesium wasting and probably contributes to the high frequency of hypomagnesemia in poorly controlled diabetic patients. Magnesium depletion is aggravated by metabolic acidosis, which causes intracellular losses as well.

Hypomagnesemia due to rapid shifts of magnesium from ECF into the intracellular compartment can occur during recovery from diabetic ketoacidosis, starvation, or respiratory acidosis. Less acute shifts may be seen during rapid bone formation after parathyroidectomy, with treatment of vitamin D deficiency, or with osteoblastic metastases. Large amounts of magnesium may be lost with acute pancreatitis, extensive burns, or protracted and severe sweating and during pregnancy and lactation.

Clinical and Laboratory Findings Hypomagnesemia may cause generalized alterations in neuromuscular function, including tetany, seizures, muscle weakness, ataxia, nystagmus, vertigo, depression, irritability, and psychosis. Patients are usually asymptomatic when serum magnesium concentrations are >0.5 mmol/L (1 meq/L; 1.2 mg/dL), although the severity of symptoms may not correlate well with serum magnesium levels. Cardiac arrhythmias may occur, including sinus tachycardia, other supraventricular tachycardias, and ventricular arrhythmias. Electrocardiographic abnormalities may include prolonged PR or QT intervals, T-wave flattening or inversion, and ST straightening. Sensitivity to digitalis toxicity may be enhanced.

Other electrolyte abnormalities often seen with hypomagnesemia, including hypocalcemia (with hypocalciuria) and hypokalemia, may not be easily corrected unless magnesium is administered as well. The hypocalcemia may be a result of concurrent vitamin D deficiency, although hypomagnesemia can cause impaired synthesis of $1,25(OH)_2D$, cellular resistance to PTH, and, at very low serum magnesium (<0.4 mmol/L [<0.8 meq/L; <1 mg/dL]), defects in PTH secretion; these abnormalities are reversible with therapy.

TREATMENT

Hypomagnesemia

Mild, asymptomatic hypomagnesemia may be treated with oral magnesium salts ($MgCl_2$, MgO, $Mg[OH]_2$) in divided doses totaling 20–30 mmol/d (40–60 meq/d). Diarrhea may occur with larger doses. More severe hypomagnesemia should be treated parenterally, preferably with IV $MgCl_2$, which can be administered safely as a continuous infusion of 50 mmol/d (100 meq Mg^{2+}/d) if renal function is normal. If GFR is reduced, the infusion rate should be lowered by 50–75%. Use of IM $MgSO_4$ is discouraged; the injections are painful and provide relatively little magnesium (2 mL of 50% $MgSO_4$ supplies only 4 mmol). $MgSO_4$ may be given IV instead of $MgCl_2$, although the sulfate anions may bind calcium in serum and urine and aggravate hypocalcemia. Serum magnesium should be monitored at intervals of 12–24 h during therapy, which may continue for several days because of impaired renal conservation of magnesium (only 50–70% of the daily IV magnesium dose is retained) and delayed repletion of intracellular deficits, which may be as high as 1–1.5 mmol/kg (2–3 meq/kg).

It is important to consider the need for calcium, potassium, and phosphate supplementation in patients with hypomagnesemia. Vitamin D deficiency frequently coexists and should be treated with oral or parenteral vitamin D or 25(OH)D (but not with $1,25(OH)_2D$, which may impair tubular magnesium reabsorption, possibly via PTH suppression). In severely hypomagnesemic patients with concomitant hypocalcemia and hypophosphatemia, administration of IV magnesium alone may worsen hypophosphatemia, provoking neuromuscular symptoms or rhabdomyolysis, due to rapid stimulation of previously suppressed PTH secretion. This is avoided by administering both calcium and magnesium.

HYPERMAGNESEMIA

Causes Hypermagnesemia is rarely seen in the absence of renal insufficiency as normal kidneys can excrete large amounts (250 mmol/d) of magnesium. Mild hypermagnesemia due to excessive reabsorption in the cTAL occurs with CaSR mutations in familial hypocalciuric hypercalcemia and has been described in some patients with adrenal insufficiency, hypothyroidism, or hypothermia. Massive exogenous magnesium exposures, usually via the gastrointestinal tract, can overwhelm renal excretory capacity and cause life-threatening hypermagnesemia **(Table 421-5)**. A notable example of this is prolonged retention of even normal amounts of magnesium-containing cathartics in patients with intestinal ileus, obstruction, or perforation. Extensive soft tissue injury or necrosis also can deliver large amounts of magnesium into the ECF in patients who have suffered trauma, shock, sepsis, cardiac arrest, or severe burns. Further, infusion of magnesium in pregnant women with eclampsia can lead to hypocalcemia.

Clinical and Laboratory Findings The most prominent clinical manifestations of hypermagnesemia are vasodilation and neuromuscular blockade, which may appear at serum magnesium concentrations >2 mmol/L (>4 meq/L; >4.8 mg/dL). Hypotension that is refractory to vasopressors or volume expansion may be an early sign. Nausea, lethargy, and weakness may progress to respiratory failure, paralysis,

TABLE 421-5 Causes of Hypermagnesemia

I. Excessive magnesium intake
    A. Cathartics, urologic irrigants
    B. Parenteral magnesium administration
II. Rapid mobilization from soft tissues
    A. Trauma, shock, sepsis
    B. Cardiac arrest
    C. Burns
III. Impaired magnesium excretion
    A. Renal failure
    B. Familial hypocalciuric hypercalcemia
IV. Other
    A. Adrenal insufficiency
    B. Hypothyroidism
    C. Hypothermia

and coma, with hypoactive tendon reflexes, at serum magnesium levels >4 mmol/L. Other findings may include gastrointestinal hypomotility or ileus; facial flushing; pupillary dilation; paradoxical bradycardia; prolongation of PR, QRS, and QT intervals; heart block; and, at serum magnesium levels approaching 10 mmol/L, asystole.

Hypermagnesemia, acting via the CaSR, causes hypocalcemia and hypercalciuria due to both parathyroid suppression and impaired cTAL calcium reabsorption.

TREATMENT

Hypermagnesemia

Successful treatment of hypermagnesemia generally involves identifying and interrupting the source(s) of magnesium and employing measures to increase magnesium clearance from the ECF. Use of magnesium-free cathartics or enemas may be helpful in clearing ingested magnesium from the gastrointestinal tract. Vigorous IV hydration should be attempted, if appropriate. Hemodialysis is effective and may be required in patients with significant renal insufficiency. Calcium, administered IV in doses of 100–200 mg over 1–2 h, has been reported to provide temporary improvement in signs and symptoms of hypermagnesemia.

VITAMIN D

SYNTHESIS AND METABOLISM

1,25-Dihydroxyvitamin D [$1,25(OH)_2D$] is the major steroid hormone involved in regulation of mineral ion homeostasis. Vitamin D and its metabolites are hormones and hormone precursors rather than vitamins, since in the proper biologic setting, they can be synthesized endogenously **(Fig. 421-4)**. In response to ultraviolet radiation of the skin, a photochemical cleavage results in the formation of vitamin D from 7-dehydrocholesterol. Cutaneous production of vitamin D is decreased by melanin and high solar protection factor sunblocks, which effectively impair skin penetration by ultraviolet light. The increased use of sunblocks in North America and Western Europe and a reduction in the magnitude of solar exposure of the general population over the past several decades has led to an increased reliance on dietary sources of vitamin D. In the United States and Canada, these sources largely consist of fortified cereals and dairy products, in addition to fish oils and egg yolks. Vitamin D from plant sources is in the form of vitamin D_2, whereas that from animal sources is vitamin D_3. These two forms have equivalent biologic activity and are activated equally well by the vitamin D hydroxylases in humans. Vitamin D enters the circulation, whether absorbed from the intestine or synthesized cutaneously, bound to vitamin D–binding protein, an α-globulin synthesized in the liver. Vitamin D is subsequently 25-hydroxylated in the liver by a cytochrome P450 oxidase in the mitochondria and microsomes. The activity of this hydroxylase is not tightly regulated, and the resultant metabolite, 25-hydroxyvitamin D [25(OH)D], is the major circulating and storage form of vitamin D. Approximately 88% of 25(OH)D circulates bound to the vitamin D–binding protein, 0.03% is free, and the rest circulates bound to albumin. The half-life of 25(OH)D is ~2–3 weeks, with that of $25(OH)D_2$ being shorter than that of $25(OH)D_3$ due to a lower affinity of vitamin D–binding protein for the former. The half-life of 25(OH)D is also greatly shortened when vitamin D–binding protein levels are reduced, as can occur with increased urinary losses in the nephrotic syndrome.

The second hydroxylation, required for the formation of the mature hormone, occurs in the kidney **(Fig. 421-5)**. The 25-hydroxyvitamin D-1α-hydroxylase (encoded by the *CYP27B1* gene) is a tightly regulated cytochrome P450–like mixed-function oxidase expressed in the proximal convoluted tubule cells of the kidney. PTH and hypophosphatemia are the major inducers of this microsomal enzyme in the kidney, whereas calcium, FGF23, and the enzyme's product, $1,25(OH)_2D$, repress it. The 25-hydroxyvitamin D-1α-hydroxylase is also present in numerous other cell types, where it is not subject to hormonal regulation. It is expressed in epidermal keratinocytes, but keratinocyte production of $1,25(OH)_2D$ is not thought to contribute to circulating levels of this hormone. In addition to being present in the trophoblastic layer of the placenta, the 1α-hydroxylase is produced by macrophages associated with granulomas and lymphomas. In these latter pathologic states, the activity of the enzyme is induced by interferon γ and TNF-α but is not regulated by calcium or $1,25(OH)_2D$; therefore, hypercalcemia, associated with elevated levels of $1,25(OH)_2D$, may be observed. Treatment of sarcoidosis-associated hypercalcemia with glucocorticoids, ketoconazole, or chloroquine reduces $1,25(OH)_2D$ production and effectively lowers serum calcium. In contrast, chloroquine has not been shown to lower the elevated serum $1,25(OH)_2D$ levels in patients with lymphoma.

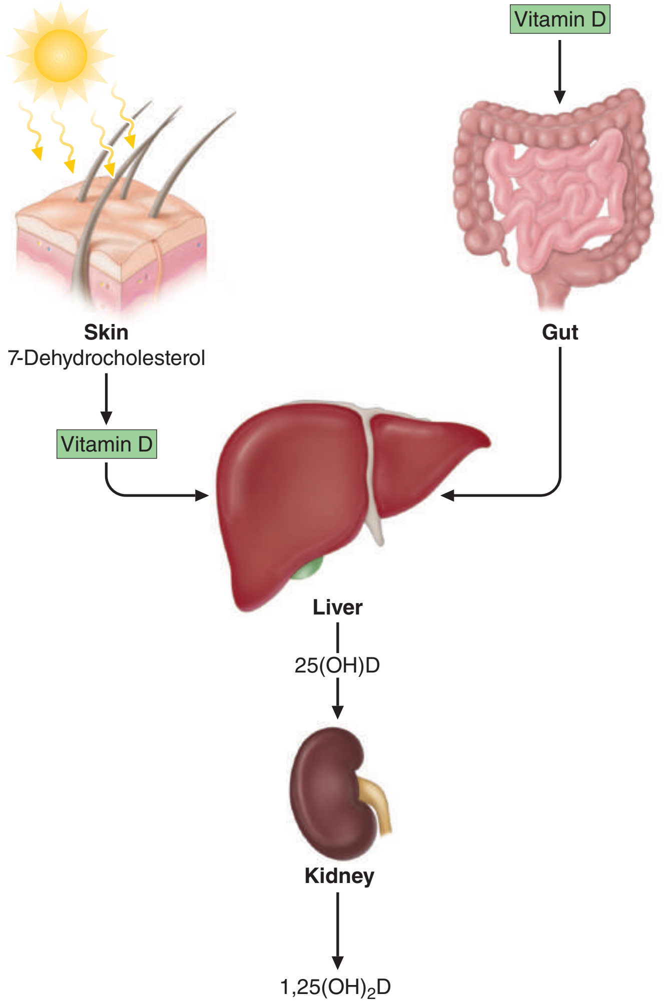

FIGURE 421-4 Vitamin D synthesis and activation. Vitamin D is synthesized in the skin in response to ultraviolet radiation and also is absorbed from the diet. It is then transported to the liver, where it undergoes 25-hydroxylation. This metabolite is the major circulating form of vitamin D. The final step in hormone activation, 1α-hydroxylation, occurs in the kidney.

The major pathway for inactivation of vitamin D metabolites is an additional hydroxylation step by the vitamin D 24-hydroxylase, an enzyme that is expressed in most tissues. $1,25(OH)_2D$ is the major inducer of this enzyme; therefore, this hormone promotes its own inactivation, thereby limiting its biologic effects. FGF23 also induces this hydroxylase, thereby reducing circulating $1,25(OH)_2D$ levels by increasing its inactivation, as well as by impairing its synthesis. Mutations of the gene encoding this enzyme (*CYP24A1*) can lead to infantile hypercalcemia, and in those less severely affected, long-standing hypercalciuria, nephrocalcinosis, and nephrolithiasis can occur.

Polar metabolites of $1,25(OH)_2D$ are secreted into the bile and reabsorbed via the enterohepatic circulation. Impairment of this recirculation, which is seen with diseases of the terminal ileum, leads to accelerated losses of vitamin D metabolites.

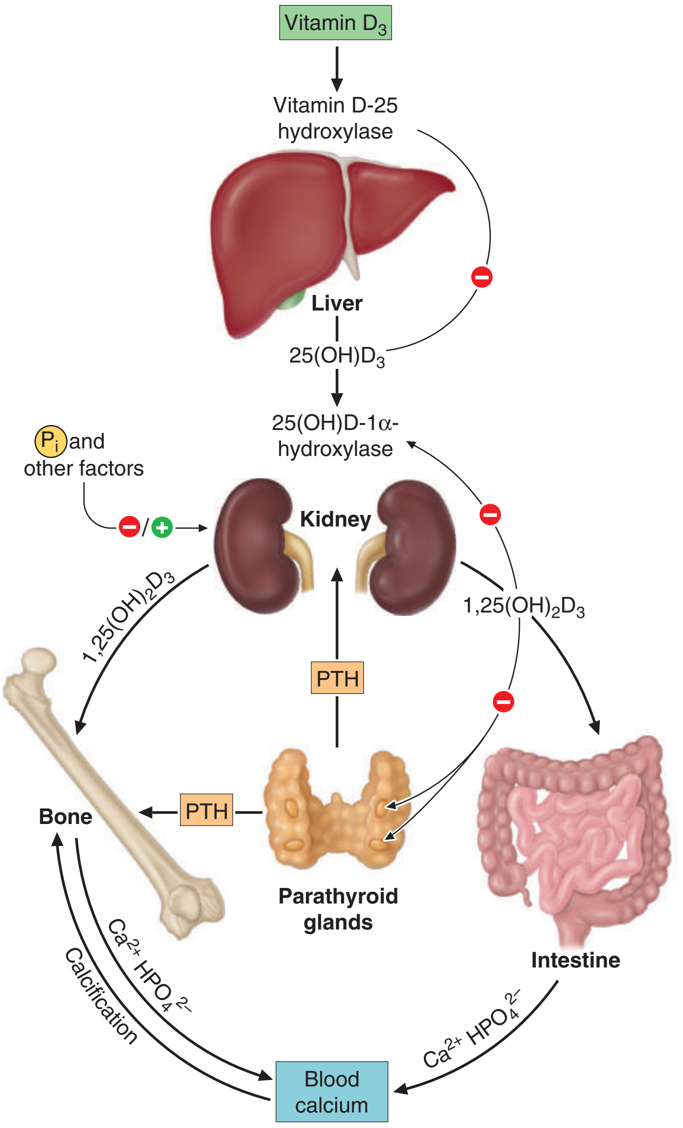

FIGURE 421-5 Schematic representation of the hormonal control loop for vitamin D metabolism and function. A reduction in the serum calcium below ~2.2 mmol/L (8.8 mg/dL) prompts a proportional increase in the secretion of parathyroid hormone (PTH) and so mobilizes additional calcium from the bone. PTH promotes the synthesis of 1,25(OH)$_2$D in the kidney, which in turn stimulates the mobilization of calcium from bone and intestine and regulates the synthesis of PTH by negative feedback.

ACTIONS OF 1,25(OH)$_2$D

1,25(OH)$_2$D mediates its biologic effects by binding to a member of the nuclear receptor superfamily, the vitamin D receptor (VDR). This receptor belongs to the subfamily that includes the thyroid hormone receptors, the retinoid receptors, and the peroxisome proliferator–activated receptors; however, in contrast to the other members of this subfamily, only one VDR isoform has been isolated. The VDR binds to target DNA sequences as a heterodimer with the retinoid X receptor, recruiting a series of coactivators that modify chromatin and approximate the VDR to the basal transcriptional apparatus, resulting in the induction of target gene expression. The mechanism of transcriptional repression by the VDR varies with different target genes but has been shown to involve either interference with the action of activating transcription factors or the recruitment of novel proteins to the VDR complex, resulting in transcriptional repression.

The affinity of the VDR for 1,25(OH)$_2$D is approximately three orders of magnitude higher than that for other vitamin D metabolites. Metabolites resulting from the 24-hydroxylation of vitamin D, including 1,24,25 trihydroxyvitamin D$_3$ (1,24,35(OH)$_3$D$_3$) and 24,25 dihydroxyvitamin D$_3$ (24,25(OH)D$_3$) have biologic effects that mimic 1,25(OH)$_2$D. In normal physiologic circumstances, these other metabolites are not thought to play significant roles in stimulating receptor-dependent actions. However, in states of vitamin D toxicity, the markedly elevated levels of 25(OH)D may lead to hypercalcemia by interacting directly with the VDR and by displacing 1,25(OH)$_2$D from vitamin D–binding protein, resulting in increased bioavailability of the active hormone.

The VDR is expressed in a wide range of cells and tissues. The molecular actions of 1,25(OH)$_2$D have been studied most extensively in tissues involved in the regulation of mineral ion homeostasis. This hormone is a major inducer of calbindin 9K, a calcium-binding protein expressed in the intestine, which is thought to play an important role in the active transport of calcium across the enterocyte. The two major calcium transporters expressed by intestinal epithelia, TRPV5 and TRPV6 (transient receptor potential vanilloid), are also vitamin D responsive. By inducing the expression of these and other genes in the small intestine, 1,25(OH)$_2$D increases the efficiency of intestinal calcium absorption, and it also has been shown to have several important actions in the skeleton. The VDR is expressed in osteoblasts and regulates the expression of several genes in this cell. These genes include the bone matrix proteins osteocalcin and osteopontin, which are upregulated by 1,25(OH)$_2$D, in addition to type I collagen, which is transcriptionally repressed by 1,25(OH)$_2$D. Both 1,25(OH)$_2$D and PTH induce the expression of RANK ligand in osteoblasts, which promotes osteoclast differentiation and increases osteoclast activity, by binding to RANK on osteoclast progenitors and mature osteoclasts. This is the mechanism by which 1,25(OH)$_2$D induces bone resorption. 1,25(OH)$_2$D regulates phosphate homeostasis, primarily by inducing the expression of FGF23 in osteocytes. The skeletal features associated with VDR-knockout mice (rickets, osteomalacia) are largely corrected by increasing calcium and phosphorus intake, underscoring the importance of vitamin D action in the gut.

The VDR is expressed in the parathyroid gland, and 1,25(OH)$_2$D has been shown to have antiproliferative effects on parathyroid cells and to suppress the transcription of the parathyroid hormone gene. These effects of 1,25(OH)$_2$D on the parathyroid gland are an important part of the rationale for current therapies directed at preventing and treating hyperparathyroidism associated with renal insufficiency and complications of chronic oral phosphate therapy.

The VDR is also expressed in tissues and organs that do not play a role in mineral ion homeostasis. Notable in this respect is the observation that 1,25(OH)$_2$D has an antiproliferative effect on several cell types, including keratinocytes, breast cancer cells, and prostate cancer cells. The effects of 1,25(OH)$_2$D and the VDR on keratinocytes are particularly intriguing, since the VDR is primarily a transcriptional repressor in these cells. Alopecia is seen in humans and mice with mutant VDRs but is not a feature of vitamin D deficiency; thus, the effects of the VDR on the hair follicle are ligand-independent. Vitamin D action is also important for regulating the normal maturation of the bone-tendon attachment site, called the enthesis.

VITAMIN D DEFICIENCY

The mounting concern about the relationship between solar exposure and the development of skin cancer has led to increased reliance on dietary sources of vitamin D. Although the prevalence of vitamin D deficiency varies, the third National Health and Nutrition Examination Survey (NHANES III) revealed that vitamin D deficiency is prevalent throughout the United States, with the prevalence being >29% in obese children. The clinical syndrome of vitamin D deficiency can be a result of deficient production of vitamin D in the skin, lack of dietary intake, accelerated losses of vitamin D, impaired vitamin D activation, or resistance to the biologic effects of 1,25(OH)$_2$D **(Table 421-6)**. The elderly and nursing home residents are particularly at risk for vitamin D deficiency, since both the efficiency of vitamin D synthesis in the skin and the absorption of vitamin D from the intestine decline with age. The presence of terminal ileal disease also results in impaired enterohepatic circulation of vitamin D metabolites. While intestinal malabsorption of dietary fats and short bowel syndrome, including that associated with intestinal bypass surgery, lead to vitamin D deficiency, the cause of vitamin D deficiency in obese individuals is poorly understood.

TABLE 421-6 Causes of Impaired Vitamin D Action

- Vitamin D deficiency
  - Impaired cutaneous production
  - Dietary absence
  - Malabsorption (short gut syndrome, gastric bypass)
- Accelerated loss of vitamin D
  - Increased metabolism (barbiturates, phenytoin, rifampin)
  - Impaired enterohepatic circulation
  - Nephrotic syndrome
  - *CYP3A4* mutation
- Impaired 25-hydroxylation
  - Liver disease, isoniazid
  - 25-Hydroxylase mutation
- Impaired 1α-hydroxylation
  - Hypoparathyroidism
  - Ketoconazole
  - 1α-Hydroxylase mutation
- FGF23 excess
  - Oncogenic osteomalacia
  - Hypophosphatemic rickets
  - Fibrous dysplasia
  - Chronic kidney disease
- Target organ resistance
  - Vitamin D receptor mutation
  - Phenytoin
- Other
  - Obesity

In addition to intestinal diseases, accelerated inactivation of vitamin D metabolites can be seen with drugs that induce hepatic cytochrome P450 mixed-function oxidases such as barbiturates, phenytoin, and rifampin. Gain-of-function mutations in *CYP3A4* accelerate the oxidation and inactivation of vitamin D metabolites, thus resulting in decreased serum levels of 25OHD and $1,25(\mathrm{OH})_2\mathrm{D}$. This form of rickets is autosomal recessive and presents during early childhood and can be treated with high doses of calcitriol or vitamin D. Impaired 25-hydroxylation, associated with severe liver disease or isoniazid, is an uncommon cause of vitamin D deficiency. A mutation in the gene responsible for 25-hydroxylation has been identified in a few kindreds. Increased circulating FGF23 levels impair 1α-hydroxylation, preventing the production of $1,25(\mathrm{OH})_2\mathrm{D}$. High levels of FGF23 are seen in those with genetic disorders associated with hypophosphatemic rickets, the most common of which is X-linked hypophosphatemia, and are prevalent in populations with profound renal dysfunction. Thus, therapeutic interventions should be considered in patients whose creatinine clearance is <0.5 mL/s (30 mL/min). Mutations in the renal 1α-hydroxylase are the basis for the genetic disorder pseudovitamin D–deficiency rickets (also called vitamin D–dependent rickets type I). This autosomal recessive disorder presents with the syndrome of vitamin D deficiency in the first year of life. Patients present with growth retardation, rickets, and hypocalcemic seizures. Serum $1,25(\mathrm{OH})_2\mathrm{D}$ levels are low despite normal 25(OH)D levels and elevated PTH levels. Treatment with vitamin D metabolites that do not require 1α-hydroxylation for activity results in disease remission, although lifelong therapy is required. A second autosomal recessive disorder, hereditary vitamin D–resistant rickets (also called vitamin D-dependent rickets type II), a consequence of vitamin D receptor mutations, is a greater therapeutic challenge. These patients present in a similar fashion during the first year of life, but alopecia often accompanies the disorder, demonstrating a functional role of the VDR in the keratinocyte stem cell population required for hair follicle regeneration. Serum levels of $1,25(\mathrm{OH})_2\mathrm{D}$ are dramatically elevated in these individuals both because of increased production due to stimulation of 1α-hydroxylase activity as a consequence of secondary hyperparathyroidism and because of impaired inactivation since induction of the 24-hydroxylase by $1,25(\mathrm{OH})_2\mathrm{D}$ requires an intact VDR. Since the receptor mutation results in hormone resistance, daily calcium and phosphate infusions may be required to bypass the defect in intestinal mineral ion absorption.

Regardless of the cause, the clinical manifestations of vitamin D deficiency are largely a consequence of impaired intestinal calcium absorption. Mild to moderate vitamin D deficiency is asymptomatic, whereas long-standing vitamin D deficiency results in hypocalcemia accompanied by secondary hyperparathyroidism, impaired mineralization of the skeleton (osteopenia on x-ray or decreased bone mineral density), and proximal myopathy. Vitamin D deficiency also has been shown to be associated with an increase in overall mortality, including cardiovascular causes. In the absence of an intercurrent illness, the hypocalcemia associated with long-standing vitamin D deficiency rarely presents with acute symptoms of hypocalcemia such as numbness, tingling, and seizures. However, the concurrent development of hypomagnesemia, which impairs parathyroid function, or the administration of potent bisphosphonates, which impair bone resorption, can lead to acute symptomatic hypocalcemia in vitamin D–deficient individuals.

Rickets and Osteomalacia In children, before epiphyseal fusion, vitamin D deficiency results in growth retardation associated with an expansion of the growth plate known as *rickets*. Three layers of chondrocytes are present in the normal growth plate: the reserve zone, the proliferating zone, and the hypertrophic zone. Rickets associated with impaired vitamin D action is characterized by expansion of the hypertrophic chondrocyte layer. The expansion of the growth plate is a consequence of impaired apoptosis of the late hypertrophic chondrocytes, an event that precedes replacement of these cells by osteoblasts during endochondral bone formation. Investigations in murine models demonstrate that hypophosphatemia, which in vitamin D deficiency is a consequence of secondary hyperparathyroidism, is a key etiologic factor in the development of the rachitic growth plate. Impaired actions specific to vitamin D also contribute to the expansion of the hypertrophic layer in the rachitic growth plate.

The hypocalcemia and hypophosphatemia that accompany vitamin D deficiency result in impaired mineralization of bone matrix proteins, a condition known as *osteomalacia*. Osteomalacia is also a feature of long-standing hypophosphatemia, which may result from renal phosphate wasting, or chronic use of etidronate or phosphate-binding antacids. This hypomineralized matrix is biomechanically inferior to normal bone; as a result, patients with osteomalacia are prone to bowing of weight-bearing extremities and skeletal fractures. Vitamin D and calcium supplementation have been shown to decrease the incidence of hip fracture among ambulatory nursing home residents in France, suggesting that undermineralization of bone contributes significantly to morbidity in the elderly. Proximal myopathy is a striking feature of severe vitamin D deficiency both in children and in adults. Rapid resolution of the myopathy is observed upon vitamin D treatment.

Although vitamin D deficiency is the most common cause of rickets and osteomalacia, many disorders lead to inadequate mineralization of the growth plate and bone. Calcium deficiency without vitamin D deficiency, the disorders of vitamin D metabolism previously discussed, and hypophosphatemia can all lead to inefficient mineralization. Even in the presence of normal calcium and phosphate levels, chronic acidosis and drugs such as bisphosphonates can lead to osteomalacia. The inorganic calcium/phosphate mineral phase of bone cannot form at low pH. Bisphosphonates bind to and prevent hydroxyapatite crystal growth. Since alkaline phosphatase is necessary for normal mineral deposition, probably because the enzyme can hydrolyze inhibitors of mineralization such as inorganic pyrophosphate, genetic inactivation of the alkaline phosphatase gene (hereditary hypophosphatasia) also can lead to osteomalacia in the setting of normal calcium and phosphate levels.

Diagnosis of Vitamin D Deficiency, Rickets, and Osteomalacia The most specific screening test for vitamin D deficiency in otherwise healthy individuals is a serum 25(OH)D level. Although the normal ranges vary, levels of 25(OH)D <37 nmol/L (<15 ng/mL) are associated with increasing PTH levels and lower bone density. The National Academy of Medicine has defined vitamin D sufficiency as a vitamin D level >50 nmol/L (>20 ng/mL), although higher levels may be required to optimize intestinal calcium absorption in the elderly and those with underlying disease states, including obesity. Vitamin D deficiency leads to impaired intestinal absorption of calcium, resulting in decreased serum total and ionized calcium values. This hypocalcemia results in secondary hyperparathyroidism, a homeostatic response that initially maintains serum calcium levels at the expense of the skeleton. Due to the PTH-induced increase in bone turnover, alkaline phosphatase levels are often increased. In addition to increasing bone resorption, PTH decreases urinary calcium excretion while promoting phosphaturia. This results in hypophosphatemia, which exacerbates the mineralization defect in the skeleton. With prolonged vitamin D deficiency resulting in osteomalacia, calcium stores in the skeleton become relatively inaccessible, since osteoclasts cannot resorb unmineralized osteoid, and frank hypocalcemia ensues. Since PTH is a major stimulus for the renal 25(OH)D 1α-hydroxylase, there is increased synthesis of the active hormone, 1,25(OH)$_2$D. Paradoxically, levels of this hormone are often normal in severe vitamin D deficiency. Therefore, measurements of 1,25(OH)$_2$D are not accurate reflections of vitamin D stores and should not be used to diagnose vitamin D deficiency in patients with normal renal function.

Radiologic features of vitamin D deficiency in children include a widened, expanded growth plate that is characteristic of rickets. These findings not only are apparent in the long bones but also are present at the costochondral junction, where the expansion of the growth plate leads to swellings known as the "rachitic rosary." Impairment of intramembranous bone mineralization leads to delayed fusion of the calvarial sutures and a decrease in the radiopacity of cortical bone in the long bones. If vitamin D deficiency occurs after epiphyseal fusion, the main radiologic finding is a decrease in cortical thickness and relative radiolucency of the skeleton. A specific radiologic feature of osteomalacia, whether associated with phosphate wasting or vitamin D deficiency, is pseudofractures, or Looser's zones. These are radiolucent lines that occur where large arteries are in contact with the underlying skeletal elements; it is thought that the arterial pulsations lead to the radiolucencies. As a result, these pseudofractures are usually a few millimeters wide, are several centimeters long, and are seen particularly in the scapula, the pelvis, and the femoral neck.

TREATMENT

Vitamin D Deficiency

Based on the National Academy of Medicine 2010 report, the recommended daily intake of vitamin D is 600 IU from 1 to 70 years of age, and 800 IU for those >70. Based on the observation that 800 IU of vitamin D, with calcium supplementation, decreases the risk of hip fractures in elderly women, this higher dose is thought to be an appropriate daily intake for prevention of vitamin D deficiency in adults. Multiple clinical trials, including the Vitamin D and Omega-3 Trial (VITAL), revealed that supplementation of vitamin D in older community-dwelling adults (>50 years of age) with adequate vitamin D levels, at doses at or above the recommended daily intake, does not further improve bone mineral density. The VITAL trial further showed that supraphysiologic doses of vitamin D in adults with normal vitamin D levels do not improve skeletal microarchitecture and do not prevent falls. Furthermore, treating older adults with daily small doses of vitamin D_3, such as 400 IU, can prevent fractures and falls, as compared with large intermittent bolus doses of vitamin D_3, which can result in increased incidence of fractures and falls. The safety margin for vitamin D is large, and vitamin D toxicity usually is observed only in patients taking doses in the range of 40,000 IU daily. Treatment of vitamin D deficiency should be directed at the underlying disorder, if possible, and also should be tailored to the severity of the condition. Vitamin D should always be repleted in conjunction with calcium supplementation since most of the consequences of vitamin D deficiency are a result of impaired mineral ion homeostasis. In patients in whom 1α-hydroxylation is impaired, metabolites that do not require this activation step are the treatment of choice. They include 1,25(OH)$_2$D$_3$ (calcitriol [Rocaltrol], 0.25–0.5 μg/d) and 1α-hydroxyvitamin D_2 (doxercalciferol [Hectorol], 2.5–5 μg/d). Outside the United States, 1α-hydroxyvitamin D_3 (alfacalcidol [One-Alpha], 0.25–1.0 μg/d) is also used. If the pathway required for activation of vitamin D is intact, severe vitamin D deficiency can be treated with pharmacologic repletion initially (50,000 IU weekly for 3–12 weeks), followed by maintenance therapy (800 IU daily). Pharmacologic doses may be required for maintenance therapy in patients who are taking medications such as barbiturates or phenytoin that accelerate metabolism of or cause resistance to 1,25(OH)$_2$D. Polymorphisms in the 25-hydroxylase and the 24-hydroxylase genes can also lead to different responses to the normal recommended daily intake of vitamin D. The hepatic enzyme cytochrome P450 3A4 (CYP3A4) is a strong inducer of the catabolism of vitamin D metabolites. Polymorphisms of the *CYP3A4* gene and certain drugs, such as phenytoin and rifampin, lead to strong induction of this enzyme; thus, those affected may also require higher doses of vitamin D supplementation. Calcium supplementation should include 1.5–2 g/d of elemental calcium. Normocalcemia is usually observed within 1 week of the institution of therapy, although increases in PTH and alkaline phosphatase levels may persist for 3–6 months. The most efficacious methods to monitor treatment and resolution of vitamin D deficiency are serum and urinary calcium measurements. In patients who are vitamin D replete and are taking adequate calcium supplementation, the 24-h urinary calcium excretion should be in the range of 100–250 mg/24 h. Lower levels suggest problems with adherence to the treatment regimen or with absorption of calcium or vitamin D supplements. Levels >250 mg/24 h predispose to nephrolithiasis and should lead to a reduction in vitamin D dosage and/or calcium supplementation.

ACKNOWLEDGMENT

The authors acknowledge Marie Demay (also a former author of this chapter) and Michael Mannstadt for their valuable input into this chapter.

FURTHER READING

Bikle D et al: Vitamin D metabolites in captivity? Should we measure free or total 25(OH)D to assess vitamin D status? J Steroid Biochem Mol Biol 173:105, 2017.

Bouillon R et al: Health effects of vitamin D supplementation: Lessons learned from randomized controlled trials and mendelian randomization studies. J Bone Miner Res 38:1391, 2023.

Carpenter TO et al: Burosumab therapy in children with X-linked hypophosphatemia. N Engl J Med 378:1987, 2018.

Christakos S et al: Vitamin D: Metabolism, molecular mechanism of action, and pleiotropic effects. Physiol Rev 96:365, 2016.

De Baaij JH et al: Magnesium in man: Implications for health and disease. Physiol Rev 95:1, 2015.

Kim JM et al: Osteoblast-osteoclast communication and bone homeostasis. Cells 10:2073, 2020.

Kovacs CS et al: The role of biomineralization in disorders of skeletal development and tooth formation. Nat Rev Endocrinol 17:336, 2021.

Robling AG, Bonewald LF: The osteocyte: New insights. Ann Rev Physiol 82:485, 2020.

Schlingmann KP, de Baaij JHF: The genetic spectrum of Gitelman (-like) syndromes. Curr Opin Nephrol Hypertens 31:508, 2022.

Zofkova I: Hypercalcemia. Pathophysiological aspects. Phys Res 65:1, 2016.

422

Disorders of the Parathyroid Gland and Calcium Homeostasis

Michael Mannstadt, Harald Jüppner

Four parathyroid glands are located posterior to the thyroid gland. They produce parathyroid hormone (PTH), which is the primary regulator of calcium homeostasis. PTH acts directly on bone, where it induces calcium (and phosphate) release, and on the kidney, where it enhances calcium reabsorption in the distal tubules. In the proximal renal tubules, PTH increases excretion of phosphate and the synthesis of 1,25-dihydroxyvitamin D (1,25(OH)$_2$D), a hormone that increases gastrointestinal calcium absorption. Serum PTH levels are tightly regulated by a negative feedback loop. Calcium, acting through the calcium-sensing receptor, and vitamin D, acting through its nuclear receptor, reduce PTH release and synthesis. Additional evidence indicates that fibroblast growth factor 23 (FGF23), a phosphaturic hormone, can suppress PTH secretion. Understanding the hormonal pathways that regulate calcium and phosphate levels as well as bone metabolism is essential for effective diagnosis and management of a wide array of hyper- and hypocalcemic disorders.

Primary hyperparathyroidism, characterized by excess production of PTH, is a common cause of hypercalcemia and is usually the result of autonomously functioning adenomas or hyperplasia. Surgery for this disorder is highly effective and has been shown to reverse some of the deleterious effects of long-standing PTH excess on bone density. Humoral hypercalcemia of malignancy (HHM) is also a common cause of hypercalcemia, which is usually due to the overproduction of parathyroid hormone–related peptide (PTHrP) by cancer cells. The similarities in the biochemical characteristics of hyperparathyroidism and HHM, first noted by Albright in 1941, are now known to reflect the actions of PTH and PTHrP through the same G protein–coupled PTH/PTHrP receptor (PH1R). The converse, namely hypocalcemia, can be caused by the lack of functional PTH, i.e., hypoparathyroidism, or by reduced PTH responsiveness of the proximal renal tubules, i.e., pseudohypoparathyroidism (PHP).

The genetic basis of numerous calcium and phosphate disorders, and the molecular characterization of parathyroid cell biology, have provided new insights into the regulation of calcium and phosphate homeostasis. In addition, PTH(1-34) and possibly some of its analogues are promising agents for the treatment of postmenopausal osteoporosis and as replacement therapy for hypoparathyroidism. Calcimimetic agents, which activate the calcium-sensing receptor (CaSR), have provided new approaches for PTH suppression, and calcilytics, which are negative allosteric modulators of the CaSR, show promising results in early clinical trials of patients with autosomal dominant hypocalcemia type 1, a rare dominant disease that is caused by activating CaSR mutations.

PTH

Structure and Physiology PTH is an 84-amino-acid single-chain peptide. The amino-terminal portion, PTH(1–34), is highly conserved and is critical for the biologic actions of the molecule. Modified synthetic fragments of the amino-terminal sequence as small as PTH(1–11) are sufficient to activate the PTH/PTHrP receptor, if provided at high enough concentrations (see below). C-terminal portions of full-length PTH(1–84) were shown to bind to a separate binding protein/receptor; however, the properties and biologic role(s), if any, of this presumed receptor for C-terminal PTH remain undefined.

The primary function of PTH is to maintain ionized calcium concentration in the extracellular fluid (ECF) within a narrow normal range. The hormone acts directly on bone and kidney and indirectly on the intestine through its effects on synthesis of $1,25(OH)_2D$ to increase serum calcium concentrations; in turn, PTH production is closely regulated by the concentration of serum ionized calcium. This feedback system is the critical homeostatic mechanism for maintenance of ECF calcium. Any tendency toward hypocalcemia, as might be induced by calcium- or vitamin D–deficient diets, is counteracted by increased PTH secretion. This in turn (1) increases bone turnover, thereby increasing the flow of calcium (and phosphate) from bone into blood; (2) increases calcium reabsorption in the distal tubules; and (3) indirectly increases the efficiency of calcium absorption in the intestine by stimulating the renal production of $1,25(OH)_2D$. Immediate control of blood calcium is due to PTH effects on bone and, to a lesser extent, on renal calcium clearance. Maintenance of calcium balance over a longer timescale, on the other hand, probably results from the effects of $1,25(OH)_2D$ on intestinal calcium absorption **(Chap. 421)**. The renal actions of PTH are exerted at multiple sites; in the proximal tubules, it increases urinary phosphate excretion, it augments calcium reabsorption in the distal tubules, and it enhances in the proximal tubules expression of CYP27B1, the enzyme that encodes the 25(OH) D-1α-hydroxylase. Every day, up to 12 mmol (500 mg) of calcium is transferred between the ECF and bone, which is a significant amount in relation to the total ECF calcium pool, and PTH plays a crucial role in regulating this transfer.

PTH has multiple actions on bone and integrates its calcemic actions (bone resorption to protect against hypocalcemia) with stimulation of bone formation. PTH-mediated changes in bone calcium release can be seen within minutes. The chronic effects of PTH are to increase the number of bone cells, both osteoblasts and osteoclasts, and to increase the remodeling of bone; these effects are apparent within hours after the hormone is given and persist for hours after PTH is withdrawn. Continuous exposure to elevated PTH (as in primary hyperparathyroidism or long-term PTH infusions in animals) leads to increased osteoclast-mediated bone resorption. However, the intermittent administration of relatively small amounts of PTH that elevate hormone levels minimally for 1–2 h each day lead to a net increase of bone mass rather than bone loss. Striking increases, especially in trabecular bone in the spine and hip, have been reported with the intermittent use of PTH for osteoporosis, and large clinical trials with PTH(1–34) as monotherapy revealed a highly significant increase in bone density and reduction in fracture incidence.

Osteoblasts (or their stromal cell precursors), which have PTH/PTHrP receptors, are crucial to this bone-forming effect of PTH. When PTH activates PTH/PTHrP receptors on osteocytes, release of calcium from the matrix surrounding these cells is enhanced; osteoclasts, which mediate bone breakdown, lack such receptors. PTH-mediated stimulation of osteoclasts is indirect, acting in part through RANKL released from osteoblasts to activate RANK on osteoclasts; in experimental studies of bone resorption in vitro, osteoblasts must be present for PTH to activate osteoclasts to resorb bone **(Chap. 421)**.

Synthesis, Secretion, and Metabolism • **SYNTHESIS** Parathyroid cells have multiple methods of adapting to increased needs for PTH production. Most rapid (within minutes) is secretion of preformed hormone in response to hypocalcemia. Second, within hours, PTH mRNA expression is induced by sustained hypocalcemia. Finally, protracted challenge leads within days to cellular replication to increase parathyroid gland mass.

PTH is initially synthesized as a larger molecule (preproPTH, consisting of 115 amino acids). After a first cleavage step to remove the "pre" sequence of 25 amino acid residues, a second cleavage step removes the "pro" sequence of 6 amino acid residues before secretion of the mature peptide comprising 84 residues. Homozygous or heterozygous mutations in the prepro-region can cause hypoparathyroidism by interfering with hormone synthesis, transport, or secretion. Thus far, only three homozygous mutations have been identified in the secreted PTH(1–84) that reduce its biological activity and are detected in some, but not all, PTH assays (see below).

Transcriptional suppression of the PTH gene by calcium is nearly maximal at physiologic calcium concentrations. Hypocalcemia increases transcriptional activity within hours. $1,25(OH)_2D$ also strongly suppresses PTH gene transcription. In patients with chronic kidney disease (CKD), administration of supraphysiologic doses of $1,25(OH)_2D$ or analogues of this active metabolite can dramatically suppress PTH overproduction and is thus used clinically to control severe secondary hyperparathyroidism. Regulation of proteolytic destruction of preformed PTH (posttranslational regulation of hormone production) is an important mechanism for mediating rapid (within minutes) changes in hormone availability. High calcium increases and low calcium inhibits the proteolytic destruction of stored hormone.

REGULATION OF PTH SECRETION PTH secretion increases steeply to a maximum value of about five times the basal rate of secretion as the calcium concentration falls from normal to 1.9–2.0 mmol/L (7.6–8.0 mg/dL; measured as total calcium). Severe intracellular magnesium deficiency impairs PTH secretion (see below).

ECF calcium controls PTH secretion by interaction with a CaSR, a G protein–coupled receptor (GPCR) for which Ca^{2+} ions act as the primary ligand (see below). This receptor, which also has phosphate binding sites, is a member of the class C GPCR superfamily and functions as an obligate homodimer. Characterized by a large extracellular domain that effectively clamps the small-molecule ligand, the CaSR is expressed in many tissues and cell types. The CaSR can couple to all four classes of G proteins in a cell-dependent context. In the parathyroids, the CaSR mediates its actions by coupling to two closely related G protein alpha-subunits, namely Gαq and Gα11, as well as Gαi. Activation of the CaSR by high calcium levels negatively regulates PTH secretion in the parathyroids and reduces calcium reabsorption in the distal renal tubules.

Genetic evidence further reinforced the essential role for the CaSR in maintaining calcium balance. Heterozygous loss-of-function CaSR mutations cause familial hypocalciuric hypercalcemia (FHH) type 1, a benign disease in which the blood calcium abnormality resembles that observed in hyperparathyroidism but with hypocalciuria; other more recently defined variants of FHH, namely FHH2 and FHH3, are caused either by heterozygous loss-of-function mutations in Gα11, the alpha-subunit of one of the signaling proteins downstream of the CaSR, or by heterozygous mutations in the adaptor protein AP2S1, which is key in the intracellular trafficking of the CaSR. Homozygous loss-of-function mutations in the CaSR are the cause of severe neonatal hyperparathyroidism, a disorder that is typically lethal if not treated within the first days of life. On the other hand, heterozygous gain-of-function mutations cause a form of hypocalcemia resembling hypoparathyroidism (see below).

METABOLISM PTH undergoes intraglandular proteolysis, which is regulated by extracellular calcium, and further degradation occurs after secretion into the circulation, mainly by liver and kidney. Removal of the critically important amino-terminus (as little as the first amino acid) produces biologically inactive PTH fragments, such as PTH(7–84), which can be detected equally well as PTH(1–84) by several commonly used immunometric PTH assays that employ two antibodies directed against portions of the N- and C-terminus, respectively. Earlier assays, now used only infrequently, measure predominantly middle and carboxyl-terminal fragments that have no or incompletely defined biological activity and are cleared more slowly from blood than the secreted PTH(1–84).

Although the problems inherent in PTH measurements have been largely circumvented by use of double-antibody immunometric assays, some evidence suggests that the PTH(7–84) (and probably related amino-terminally truncated fragments) can act, through yet undefined mechanisms, as an inhibitor of PTH action and may therefore be of clinical significance, particularly in patients with CKD. In this group of patients, efforts to prevent secondary hyperparathyroidism by a variety of measures (vitamin D analogues, higher calcium intake, higher dialysate calcium, phosphate-lowering strategies, and calcimimetic drugs) can lead to oversuppression of the parathyroid glands since some amino-terminally truncated PTH fragments can react in some immunometric PTH assays, thus overestimating the levels of biologically active PTH(1–84). Excessive parathyroid gland suppression in CKD can lead to adynamic bone disease (see below), which has been associated in children with further impaired growth and increased bone fracture rates in adults and can furthermore lead to significant hypercalcemia. The measurement of PTH with newer third-generation immunometric assays, which use detection antibodies directed against extreme amino-terminal PTH epitopes and thus detect only full-length PTH(1–84), has not yet clearly shown to be advantageous in the clinical setting.

PTHRP

Structure and Physiology PTHrP is responsible for most instances of HHM **(Chap. 98)**, a syndrome that resembles primary hyperparathyroidism but without elevated PTH levels. Most cell types normally produce PTHrP, including brain, pancreas, heart, lung, mammary tissue, placenta, endothelial cells, and smooth muscle. In fetal animals, PTHrP directs transplacental calcium transfer, and high concentrations of PTHrP are produced in mammary tissue and secreted into milk, but the biologic significance of this peptide in breast milk is unknown. PTHrP has paracrine and autocrine functions and it plays an essential role in diverse functions such as endochondral bone formation, branching morphogenesis of the breast, and possibly in uterine contraction.

PTH and PTHrP, although products of different genes, exhibit considerable functional and structural homology **(Fig. 422-1)** and have evolved from a shared ancestral gene. The structure of the gene encoding human PTHrP, however, is more complex than that of PTH, containing multiple additional exons, which can undergo alternate splicing patterns during formation of the mature mRNA. Protein products of 139, 141, and 173 amino acids are produced, and other molecular forms may result from tissue-specific degradation at accessible internal cleavage sites. The biologic roles of these various molecular species and the nature of the circulating forms of PTHrP are unclear. In fact, it is uncertain whether PTHrP circulates at any significant level in healthy children and adults. As a paracrine factor, PTHrP may be produced, act, and be destroyed locally within tissues. In adults, PTHrP appears to have little influence on calcium homeostasis, except in disease states, when large tumors, especially of the squamous cell type as well as renal cell carcinomas, lead to massive overproduction of the hormone and hypercalcemia.

Both PTH and PTHrP bind to and activate the PTH/PTHrP receptor. The PTH/PTHrP receptor (also known as the PTH-1 receptor [PTH1R]) belongs to the class B GPCRs that includes the receptors for calcitonin, glucagon, secretin, vasoactive intestinal peptide, and a few other peptides. Although both ligands activate the PTH1R, the two peptides induce distinct responses by the receptor, which explains how a single receptor without isoforms can serve different biologic roles. The extracellular regions of the receptor are involved in hormone binding, and the intracellular domains, after hormone activation, bind G protein subunits to transduce hormone signaling into cellular responses through the stimulation of second messenger formation. A second receptor that binds PTH, originally termed the PTH-2 receptor (PTH2R), is primarily expressed in brain, pancreas, and testis. Different mammalian PTH1Rs respond equivalently to PTH and PTHrP, at least when tested with traditional assays, whereas the human PTH2R responds efficiently only to PTH, but not to PTHrP. PTH2Rs from other species show little or no stimulation of second-messenger

	1				5					10					15					20					25					30
hPTH	SER	VAL	SER	GLU	ILE	GLN	LEU	MET	HIS	ASN	LEU	GLY	LYS	HIS	LEU	ASN	SER	MET	GLU	ARG	VAL	GLU	TRP	LEU	ARG	LYS	LYS	LEU	GLN	ASP
hPTHrp	ALA	–	–	–	HIS	–	–	LEU	–	ASP	LYS	–	–	SER	ILE	GLN	ASP	LEU	ARG	–	ARG	PHE	PHE	–	HIS	HIS	LEU	ILE	ALA	GLU

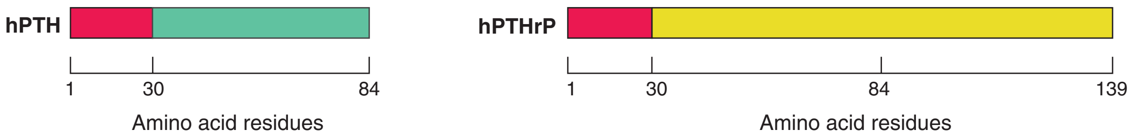

FIGURE 422-1 Schematic diagram to illustrate similarities and differences in structure of human parathyroid hormone (hPTH) and human PTH-related peptide (hPTHrP). Close structural (and functional) homology exists between the first 30 amino acids of hPTH and hPTHrP. The PTHrP sequence may be ≥139 amino acid residues in length. PTH is only 84 residues long; after residue 30, there is little structural homology between the two. Dashed lines in the PTHrP sequence indicate identity; underlined residues, although different from those of PTH, still represent conservative changes (charge or polarity preserved). Ten amino acids are identical, and a total of 20 of 30 are homologues.

formation in response to PTH or PTHrP. In fact, the endogenous ligand of the PTH2R was shown to be a hypothalamic peptide referred to as tubular infundibular peptide of 39 residues, TIP39, that is only distantly related to PTH and PTHrP. The PTH1R and the PTH2R can be traced backward in evolutionary time to fish, which express also a third receptor, the PTH3R, that is more closely related to the fish PTH1R than to the fish PTH2R. The evolutionary conservation of structure and function suggests important biologic roles for these receptors, even in fish, which lack discrete parathyroid glands but produce two molecules that are closely related to mammalian PTH.

Studies using the cloned PTH1R confirm that it can be coupled to more than one G protein and second-messenger pathway, thus contributing to the multiplicity of pathways stimulated by PTH. Activation of protein kinases (A and C) and calcium transport channels is associated with a variety of hormone-specific tissue responses. These responses include inhibition of phosphate and bicarbonate transport, stimulation of calcium transport, and activation of renal 1α-hydroxylase in the kidney. The responses in bone include effects on collagen synthesis, alkaline phosphatase, ornithine decarboxylase, citrate decarboxylase, and glucose-6-phosphate dehydrogenase activities; phospholipid synthesis; and calcium and phosphate transport. Ultimately, these biochemical events lead to an integrated hormonal response in bone turnover and calcium homeostasis. PTH also activates Na^+/Ca^{2+} exchangers at renal distal tubular sites and stimulates translocation of preformed calcium transport channels, moving them from the interior to the apical surface to increase tubular uptake of calcium. PTH-dependent stimulation of phosphate excretion involves reduced expression of two sodium-dependent phosphate co-transporters, NPT2a and NPT2c, at the apical membrane, thereby reducing phosphate reabsorption in the proximal renal tubules. Similar mechanisms may be involved in other renal tubular transporters that are influenced by PTH.

PTHrP exerts important developmental influences on fetal bone development and in adult physiology. Homozygous ablation of the gene encoding PTHrP (or disruption of the PTH1R gene) in mice causes a lethal phenotype in which animals are born with pronounced acceleration of chondrocyte maturation that resembles a human disease, Blomstrand lethal chondrodysplasia (BLC), that is caused by homozygous or compound heterozygous, inactivating PTH1R mutations **(Fig. 422-2)**. Heterozygous inactivating PTH1R mutations in humans furthermore can be a cause of delayed tooth eruption, while heterozygous inactivating PTHrP mutations lead to premature growth plate closure and reduced adult heights. Besides the lethal, biallelic PTH1R mutations that cause BLC, several homozygous mutations in this gene have now been identified in a rare recessive disease referred to as Eiken syndrome. Affected patients typically have normal mineral ion regulation, yet delayed growth plate maturation resulting in some bone deformities, reduced growth, and delayed tooth eruption; recently, a few cases were described with symptomatic hypocalcemia and considerably elevated PTH levels.

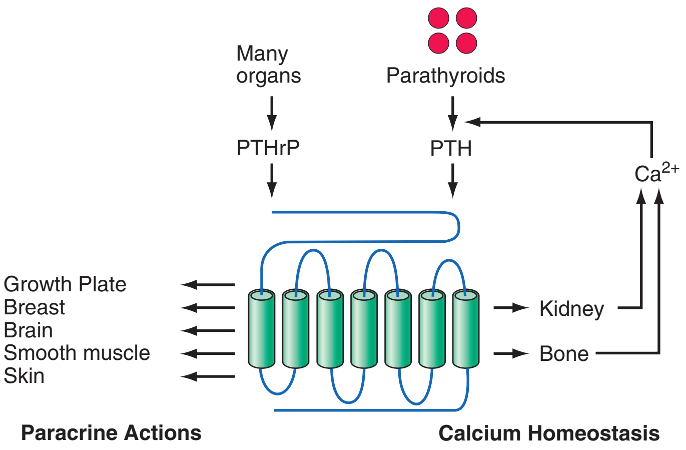

FIGURE 422-2 Dual role for the actions of the PTH/PTHrP receptor (PTH1R). Parathyroid hormone (PTH; endocrine-calcium homeostasis) and PTH-related peptide (PTHrP; paracrine–multiple tissue actions including growth plate cartilage in developing bone) use the single receptor for their disparate functions mediated by the amino-terminal 34 residues of either peptide. Other regions of both ligands interact with other receptors (not shown).

CALCITONIN

(See also Chap. 400) Calcitonin is a peptide hormone with hypocalcemic properties that in several mammalian species acts as an indirect antagonist to the calcemic actions of PTH. Calcitonin seems to be of limited physiologic significance in humans, at least with regard to calcium homeostasis. It is of medical significance because of its role as a tumor marker in sporadic and hereditary cases of medullary thyroid carcinoma and its medical use as an adjunctive treatment in severe hypercalcemia and in Paget's disease of bone at pharmacologic doses. Levels can also be elevated in patients with pseudohypoparathyroidism (PHP); the significance of this observation is unclear.

The hypocalcemic activity of calcitonin is accounted for primarily by inhibition of osteoclast-mediated bone resorption and secondarily by stimulation of renal calcium clearance. These effects are mediated by receptors on osteoclasts and renal tubular cells. Calcitonin exerts additional effects through receptors present in the brain, the gastrointestinal tract, and the immune system. The hormone, for example, exerts analgesic effects directly on cells in the hypothalamus and related structures, possibly by interacting with receptors for related peptide hormones such as calcitonin gene–related peptide (CGRP) or amylin. Both of these ligands have specific high-affinity receptors that share considerable structural similarity with the PH1R and can also bind to and activate calcitonin receptors. The calcitonin receptor shares considerable structural similarity with the PTH1R.

The naturally occurring calcitonins consist of a peptide chain of 32 amino acids. There is considerable sequence variability among species. Calcitonin from salmon, which is used therapeutically, is 10–100 times more potent than mammalian forms in lowering serum calcium.

The circulating level of calcitonin in humans is lower than that in many other species. In humans, even extreme variations in calcitonin production do not change calcium and phosphate metabolism; no definite effects are attributable to calcitonin deficiency (totally thyroidectomized patients receiving only replacement thyroxine) or excess (patients with medullary carcinoma of the thyroid, a calcitonin-secreting tumor) **(Chap. 400)**. Calcitonin has been a useful pharmacologic agent to suppress bone resorption in Paget's disease **(Chap. 424)** and osteoporosis **(Chap. 423)** and in the treatment of hypercalcemia of malignancy (see below). However, bisphosphates are usually more effective, and the physiologic role, if any, of calcitonin in humans is uncertain. On the other hand, ablation of the calcitonin gene (combined with ablation of the CGRP gene because both genes are in close proximity) in mice leads to reduced bone mineral density, suggesting that its biologic role in mammals is still not fully understood.

HYPERCALCEMIA

Introduction

(See also Chap. 57) Hypercalcemia can be a manifestation of a serious illness such as malignancy or can be detected coincidentally by laboratory testing in a patient with no obvious illness. The number of patients recognized with asymptomatic hypercalcemia, usually primary hyperparathyroidism, increased in the late twentieth century when wider testing became readily available.

Whenever hypercalcemia is confirmed, a definitive diagnosis must be established. Although hyperparathyroidism, a frequent cause of asymptomatic hypercalcemia, is a chronic disorder in which manifestations, if any, may be expressed only after months or years, hypercalcemia can also be the earliest manifestation of malignancy, the second most common cause of hypercalcemia in the adult. The causes of hypercalcemia are numerous **(Table 422-1)**, but hyperparathyroidism and cancer account for 90% of all cases.

Before initiating a diagnostic workup, confirm the presence of true hypercalcemia, not a false-positive laboratory test from factors like hemoconcentration during blood collection or elevation in serum proteins such as albumin. Since hypercalcemia is typically chronic, it is cost-effective to obtain several serum calcium and concomitant albumin measurements, which do not require fasting.

TABLE 422-1 Classification of Causes of Hypercalcemia

- **I. Parathyroid-Related**
  - A. Primary hyperparathyroidism
    - 1. Adenoma(s)
    - 2. Multiple endocrine neoplasia
    - 3. Parathyroid carcinoma
    - 4. Ectopic production of parathyroid hormone (PTH)
    - 5. Exogenous administration of PTH or analogues
  - B. Lithium therapy
  - C. Familial hypocalciuric hypercalcemia
- **II. Malignancy-Related**
  - A. Tumors with osteolytic metastases (breast, multiple myeloma, lymphoma, etc.)
  - B. Solid tumor with humoral mediation of hypercalcemia (squamous cell carcinoma of the lung, kidney, breast, and others)
  - C. $1,25(OH)_2D$-mediated hypercalcemia of malignancies (lymphoma, ovarian dysgerminoma, etc.)
- **III. Vitamin D–Related**
  - A. Vitamin D intoxication
  - B. ↑ $1,25(OH)_2D$; sarcoidosis and other granulomatous diseases, lymphoma
  - C. ↑ $1,25(OH)_2D$; impaired $1,25(OH)_2D$ metabolism due to biallelic 24-hydroxylase mutations or increased $1,25(OH)_2D$ synthesis due to inactivating biallelic mutations involving the renal sodium-dependent phosphate co-transporters
- **IV. Associated with High Bone Turnover**
  - A. Hyperthyroidism
  - B. Immobilization
  - C. Thiazides
  - D. Vitamin A intoxication
  - E. Fat necrosis
- **V. Associated with Renal Failure**
  - A. Tertiary hyperparathyroidism
  - B. Aluminum intoxication and adynamic bone disease
  - C. Milk-alkali syndrome

Clinical features are helpful in differential diagnosis. Hypercalcemia in an adult who is asymptomatic is usually due to primary hyperparathyroidism. In malignancy-associated hypercalcemia, the disease is usually not occult; rather, symptoms of malignancy bring the patient to the physician, and hypercalcemia is discovered during the evaluation. If asymptomatic hypercalcemia can be documented for more than a year, malignancy is unlikely. Nevertheless, differentiating primary hyperparathyroidism from occult malignancy can occasionally be difficult, and careful evaluation is required, particularly when the duration of the hypercalcemia is unknown. Other causes of hypercalcemia may include excessive intake of vitamin D or activated analogues, impaired metabolism of $1,25(OH)_2D$, high bone turnover from any of several causes, or renal failure (Table 422-1). Immunometric PTH assays serve as the principal laboratory test in establishing the diagnosis.

Hypercalcemia from any cause can result in fatigue, depression, mental confusion, anorexia, nausea, vomiting, constipation, reversible renal tubular defects, increased urine output, a short QT interval in the electrocardiogram, and, in some patients, cardiac arrhythmias. Generally, symptoms are more common at calcium levels >2.9–3.0 mmol/L (11.6–12.0 mg/dL), but some patients, even at this level, are asymptomatic. When the calcium level is >3.2 mmol/L (12.8 mg/dL), calcification in kidneys, skin, vessels, lungs, heart, and stomach occurs, and renal insufficiency may develop, particularly if blood phosphate levels are normal or elevated due to impaired renal excretion. Severe hypercalcemia, usually defined as ≥3.7–4.5 mmol/L (14.8–18.0 mg/dL), can be a medical emergency; coma and cardiac arrest can occur.

Acute management of the hypercalcemia is usually successful. The type of treatment is based on the severity of the hypercalcemia and the nature of associated symptoms, as outlined below.

■ PRIMARY HYPERPARATHYROIDISM

Pathophysiology • **NATURAL HISTORY AND INCIDENCE** Primary hyperparathyroidism, which is typically a disease of postmenopausal women, results from excessive secretion of PTH that is disproportionate to serum calcium levels. This typically causes hypercalcemia and hypophosphatemia. There is great variation in the manifestations. Patients may present with multiple signs and symptoms, including recurrent nephrolithiasis, peptic ulcers, mental changes, and, less frequently, extensive bone resorption. However, with greater awareness of the disease and wider use of multiphasic screening tests, including measurements of blood calcium, the diagnosis is frequently made in patients who have no symptoms and minimal, if any, signs of the disease other than hypercalcemia and elevated levels of PTH. The manifestations may be subtle, and the disease may have a benign course for many years or a lifetime. This milder form of the disease is usually termed *asymptomatic hyperparathyroidism* and can present with or without end-organ involvement. Rarely, hyperparathyroidism develops or worsens abruptly and causes severe complications such as marked dehydration and coma, so-called hypercalcemic parathyroid crisis.

The annual incidence of the disease varies globally and has changed over the decades likely reflecting changes in the usage of laboratory screening panels that include serum calcium. The current incidence in the United States is calculated to be about 50 per 100,000 person-years.

ETIOLOGY Parathyroid tumors are most often encountered as isolated monoclonal adenomas without other endocrinopathy. They may also arise in hereditary syndromes such as multiple endocrine neoplasia (MEN) syndromes. As many as 10% of patients with hyperparathyroidism have a genetic basis for the disease (see below). Parathyroid tumors may also arise as secondary to underlying disease (excessive stimulation in secondary hyperparathyroidism, especially chronic renal failure) or after other forms of excessive stimulation such as lithium therapy. These etiologies are discussed below.

Solitary Adenomas A single abnormal gland is the cause in ~80% of patients; the abnormality in the gland is usually a benign neoplasm or adenoma and extremely rarely a parathyroid carcinoma. More than one adenoma has been reported, and genetic causes of the disease often underlies chief cell hyperplasia of all four glands.

Hereditary Syndromes and Multiple Parathyroid Tumors Hereditary hyperparathyroidism can occur without other endocrine abnormalities but is usually part of a MEN syndrome **(Chap. 400)**. MEN 1 (Wermer's syndrome) consists of hyperparathyroidism and tumors of the pituitary and pancreas, often associated with gastric hypersecretion and peptic ulcer disease (Zollinger-Ellison syndrome). MEN 2A is characterized by pheochromocytoma and medullary carcinoma of the thyroid, as well as hyperparathyroidism; MEN 2B has additional associated features such as multiple neuromas but usually lacks hyperparathyroidism. MEN4, caused by mutations in the p27 cyclin-dependent kinase inhibitor (encoded for by *CDKN1B*), has similar clinical manifestations as MEN 1. Mutations in the *MAX* gene are associated with familial forms of pheochromocytoma/paraganglioma, and a few family members have also been reported to have primary hyperparathyroidism. Each of these MEN syndromes is transmitted in an apparent autosomal dominant manner.

The *hyperparathyroidism jaw tumor* (HPT-JT) syndrome occurs in families with parathyroid tumors (sometimes carcinomas) in association with benign jaw tumors. This disorder is caused by mutations in *CDC73* (*HRPT2*), and mutations in this gene are also observed in sporadic parathyroid cancers. Some kindreds exhibit hereditary hyperparathyroidism without other endocrinopathies, which has been referred to as *nonsyndromic familial isolated hyperparathyroidism* (FIHP). In some of these familial cases, the disease co-segregated with heterozygous mutations in *GCM2*. Inactivating or dominant-negative mutations in this parathyroid-specific transcription factor had initially been identified in familial forms of hypoparathyroidism. However, GCM2 variants that are predicted to cause a gain-of-function of GCM2 have been reported in a subset of patients with FIHP. Because the prevalence of these GCM2 variants is much higher than the prevalence of primary hyperparathyroidism, they might be a risk factor for developing the disease. Furthermore, there is speculation that some FIHP cases may be examples of variable expression of the other syndromes such as MEN 1, MEN 2, or the HPT-JT syndrome, but they may also have distinctive, still unidentified genetic causes.

Genetic Defects Associated with Hyperparathyroidism As in many other types of neoplasia, two fundamental types of genetic defects have been identified in parathyroid gland tumors: (1) overactivity of protooncogenes and (2) loss of function of tumor-suppressor genes. The former, by definition, can lead to uncontrolled cellular growth and function by activation (gain-of-function mutation) of a single allele of the responsible gene, whereas the latter requires loss of function of both allelic copies. Biallelic loss of function of a tumor-suppressor gene is usually characterized by a germline defect (all cells) of one allele (autosomal-dominant mode of inheritance) and an additional somatic deletion/mutation in the second allele of the tumor **(Fig. 422-3)**.

Mutations in the *MEN1* gene, which encodes the tumor suppressor MENIN, on chromosome 11q13 are responsible for causing MEN 1. Inheritance of one mutated allele in this hereditary syndrome, followed by loss of the other allele via somatic cell mutation, leads to monoclonal expansion and tumor development. Also, in ~15–20% of sporadic parathyroid adenomas, both alleles of the *MEN1* locus on chromosome 11 are somatically deleted, implying that the same defect responsible for MEN 1 can also cause the sporadic disease **(Fig. 422-3*A*)**. Consistent with the Knudson hypothesis for two-step neoplasia in certain inherited cancer syndromes **(Chap. 76)**, the earlier onset of hyperparathyroidism in the hereditary syndromes reflects the need for only one mutational event to trigger the monoclonal outgrowth. In sporadic adenomas, typically occurring later in life, two different somatic events must occur before the *MEN1* gene is silenced.

MEN 2 is an example for a mutation in a protooncogene and is associated with gain-of-function mutations in the *Ret* oncogene.

A more complex pattern, still incompletely resolved, arises with genetic defects and carcinoma of the parathyroids. This appears to be due to biallelic loss of a functioning copy of a gene, *CDC73* (or *HRPT2*), originally identified as the cause of the HPT-JT syndrome. Several inactivating mutations have been identified in *CDC73* (located on chromosome 1q21-31), which encodes a 531-amino-acid protein called parafibromin.

The discovery of the genetic mutations that lead to multiple endocrine neoplasias allows genetic testing of suspected probands and family members. Benefits of genetic testing include the ability to verify the clinical diagnosis, identify affected family members, and rule out the genetic variant in family members, who seem to be unaffected.

An important contribution from studies on the genetic origin of parathyroid carcinoma has been the realization that the mutations involve a different pathway than that involved with the benign gland enlargements. Unlike the pathogenesis of genetic alterations seen in colon cancer, where lesions evolve from benign adenomas to malignant disease by progressive genetic changes, the alterations commonly seen in most parathyroid cancers (*HRPT2* mutations) are infrequently seen in sporadic parathyroid adenomas.

Study of the parathyroid cancers found in some patients with the HPT-JT syndrome has led to identification of a much larger role for *HRPT2* mutations in most parathyroid carcinomas, including those that arise sporadically, without apparent association with the HPT-JT syndrome. Mutations in the coding region have been identified in 75–80% of all parathyroid cancers analyzed, leading to the conclusion that, with addition of presumed mutations in the noncoding regions, this genetic defect may be seen in essentially all parathyroid carcinomas. Of special importance was the discovery that, in some sporadic parathyroid cancers, germline mutations have been found; this, in turn, has led to careful investigation of the families of these patients and a new clinical indication for genetic testing in this setting.

Abnormalities at the *Rb* gene were the first to be noted in parathyroid cancer. The *Rb* gene, a tumor-suppressor gene located on

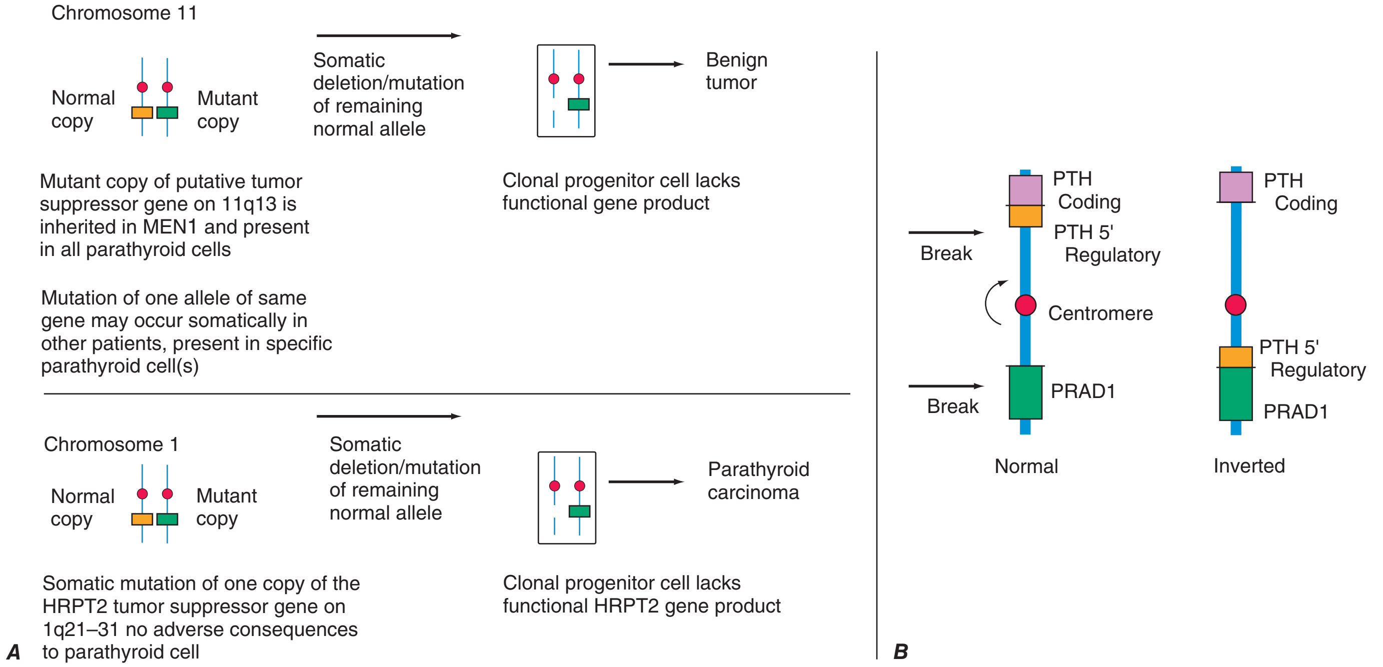
Chromosome 11
Normal copy
Mutant copy
Somatic deletion/mutation of remaining normal allele
Benign tumor
Mutant copy of putative tumor suppressor gene on 11q13 is inherited in MEN1 and present in all parathyroid cells
Clonal progenitor cell lacks functional gene product
Mutation of one allele of same gene may occur somatically in other patients, present in specific parathyroid cell(s)
Chromosome 1
Normal copy
Mutant copy
Somatic deletion/mutation of remaining normal allele
Parathyroid carcinoma
Somatic mutation of one copy of the HRPT2 tumor suppressor gene on 1q21–31 no adverse consequences to parathyroid cell
Clonal progenitor cell lacks functional HRPT2 gene product
A
PTH Coding
PTH 5' Regulatory
Break
Centromere
PRAD1
Break
Normal
PTH Coding
PTH 5' Regulatory
PRAD1
Inverted
B

FIGURE 422-3 ***A.*** Schematic diagram indicating molecular events in tumor susceptibility. The patient with the hereditary abnormality (multiple endocrine neoplasia [MEN]) is envisioned as having one defective gene inherited from the affected parent on chromosome 11, but one copy of the normal gene is present from the other parent. In the monoclonal tumor (benign tumor), a somatic event, here partial chromosomal deletion, removes the remaining normal gene from a cell. In nonhereditary tumors, two successive somatic mutations must occur, a process that takes a longer time. By either pathway, the cell, deprived of growth-regulating influence from this gene, has unregulated growth and becomes a tumor. A different genetic locus also involving loss of a tumor-suppressor gene termed *HRPT2* is involved in the pathogenesis of parathyroid carcinoma. *(Reproduced with permission from A Arnold: Genetic basis of endocrine disease 5. Molecular genetics of parathyroid gland neoplasia. J Clin Endocrinol Metab 77:1108, 1993.)* ***B.*** Schematic illustration of the mechanism and consequences of gene rearrangement and overexpression of the *PRAD1* protooncogene (pericentromeric inversion of chromosome 11) in parathyroid adenomas. The excessive expression of PRAD1 (a cell cycle control protein, cyclin D1) by the highly active PTH gene promoter in the parathyroid cell contributes to excess cellular proliferation. *(Reproduced with permission from J Habener, in L DeGroot, JL Jameson (eds): Endocrinology, 4th ed. Philadelphia, PA: Saunders; 2001.)*

chromosome 13q14, was initially associated with retinoblastoma but has since been implicated in other neoplasias, including parathyroid carcinoma. Early studies implicated allelic deletions of the *Rb* gene in many parathyroid carcinomas and decreased or absent expression of the Rb protein. However, because there are often large deletions in chromosome 13 that include many genes in addition to the *Rb* locus (with similar findings in some pituitary carcinomas), it remains possible that other tumor-suppressor genes on chromosome 13 may be playing a role in parathyroid carcinoma.

Overall, it seems there are multiple factors in parathyroid cancer, in addition to the *HRPT2* and *Rb* gene, although the *HRPT2* gene mutation is the most invariant abnormality. *RET* encodes a tyrosine kinase type receptor; specific inherited germline mutations lead to a constitutive activation of the receptor, thereby explaining the autosomal dominant mode of transmission and the relatively early onset of neoplasia. In the MEN 2 syndrome, the *RET* protooncogene may be responsible for the earliest disorder detected, the polyclonal disorder C-cell hyperplasia, which then is transformed into a clonal outgrowth—a medullary carcinoma with the participation of other, still uncharacterized genetic defects.

In some parathyroid adenomas, activation of a protooncogene occurs **(Fig. 422-3*B*)**. A reciprocal translocation involving chromosome 11 was identified as a somatic event that juxtaposes the *PTH* gene promoter upstream of *CCND1*, which encodes a cyclin D protein that plays a key role in normal cell division. This translocation plus other mechanisms that cause an equivalent overexpression of cyclin D1 are found in 20–40% of parathyroid adenomas.

Mouse models have confirmed the role of several of the major identified genetic defects in parathyroid disease and the MEN syndromes. Loss of the *MEN1* gene locus and overexpression of the *CCND1* protooncogene or the mutated *RET* protooncogene have been analyzed by genetic manipulation in mice, with the expected onset of parathyroid tumors or medullary thyroid carcinoma, respectively.

Pathology Adenomas are most often located in the inferior parathyroid glands, but in 6–10% of patients, parathyroid adenomas may be located in the thymus, the thyroid, the pericardium, or behind the esophagus. Adenomas are usually 0.5–5 g in size but may be as large as 10–20 g (normal glands weigh 25 mg on average). Chief cells are predominant in both hyperplasia and adenoma. With chief cell hyperplasia, the enlargement may be so asymmetric that some involved glands appear grossly normal. If generalized hyperplasia is present, however, histologic examination reveals a uniform pattern of chief cells and disappearance of fat even in the absence of an increase in gland weight. Thus, microscopic examination of biopsy specimens of several glands is essential to interpret findings at surgery.

Parathyroid carcinoma is often not aggressive. Long-term survival without recurrence is common if at initial surgery the entire gland is removed without rupture of the capsule. Recurrent parathyroid carcinoma is usually slow growing with local spread in the neck, and surgical correction of recurrent disease may be feasible. Occasionally, however, parathyroid carcinoma is more aggressive, with distant metastases (lung, liver, and bone) found at the time of initial operation. It may be difficult to appreciate initially that a primary tumor is carcinoma; increased numbers of mitotic figures and increased fibrosis of the gland stroma may precede invasion. The diagnosis of carcinoma is often made in retrospect when metastasis occur. Hyperparathyroidism from a parathyroid carcinoma may be indistinguishable from other forms of primary hyperparathyroidism but is usually more severe clinically. A potential clue to the diagnosis is offered by the degree of calcium elevation. Calcium values of 3.5–3.7 mmol/L (14–15 mg/dL) are frequent with carcinoma and may alert the surgeon to remove the abnormal gland with care to avoid capsular rupture. Recent findings concerning the genetic basis of some patients with parathyroid carcinoma (distinct from that of benign adenomas) indicate the need, in these kindreds, for family screening (see below).

Signs and Symptoms Many patients with primary hyperparathyroidism are asymptomatic. Manifestations of hyperparathyroidism involve primarily the kidneys and the skeletal system. Kidney involvement, due either to deposition of calcium in the renal parenchyma or to recurrent nephrolithiasis, was present in 60–70% of patients prior to 1970. With earlier detection, renal complications occur in <20% of patients in many large series. Renal stones are usually composed of either calcium oxalate or calcium phosphate. In occasional patients, repeated episodes of nephrolithiasis or the formation of large calculi may lead to urinary tract obstruction, infection, and loss of renal function. Nephrocalcinosis may also cause decreased renal function and phosphate retention.

The distinctive bone manifestation of hyperparathyroidism is *osteitis fibrosa cystica*, which occurred in 10–25% of patients in series reported 50 years ago. Histologically, the pathognomonic features are an increase in the giant multinucleated osteoclasts in scalloped areas on the surface of the bone (Howship's lacunae) and a replacement of the normal cellular and marrow elements by fibrous tissue. Radiographic changes include resorption of the phalangeal tufts and replacement of the usually sharp cortical outline of the bone in the digits by an irregular outline (subperiosteal resorption). In recent years, osteitis fibrosa cystica is very rare in primary hyperparathyroidism, probably due to the earlier detection and therefore milder form of the disease.

Dual-energy x-ray absorptiometry of the spine provides reproducible quantitative estimates (within a few percent) of spinal bone density. Similarly, bone density in the extremities can be quantified by densitometry of the hip or of the distal radius at a site chosen to be primarily cortical. Computed tomography (CT) is a very sensitive technique for estimating spinal bone density, but reproducibility of standard CT is no better than 5%. Newer CT techniques (spiral, "extreme" CT) are more reproducible but are currently available in a limited number of medical centers and used for research purposes. Cortical bone density is reduced, while cancellous bone density, especially in the spine, is relatively preserved.

In symptomatic patients, dysfunctions of the central nervous system (CNS), peripheral nerve and muscle, gastrointestinal tract, and joints also occur. It has been reported that severe neuropsychiatric manifestations may be reversed by parathyroidectomy. When present in symptomatic patients, neuromuscular manifestations may include proximal muscle weakness, easy fatigability, and atrophy of muscles and may be so striking as to suggest a primary neuromuscular disorder. The distinguishing feature is the complete regression of neuromuscular disease after surgical correction of the hyperparathyroidism.

Gastrointestinal manifestations are sometimes subtle and include vague abdominal complaints and disorders of the stomach and pancreas. Again, cause and effect are unclear. In MEN 1 patients with hyperparathyroidism, duodenal ulcer may be the result of associated pancreatic tumors that secrete excessive quantities of gastrin (Zollinger-Ellison syndrome). Pancreatitis has been reported in association with hyperparathyroidism, but the incidence and the mechanism are not established.

Much attention has been paid in recent years to the manifestations of and optimum management strategies for asymptomatic hyperparathyroidism. This is now the most prevalent form of the disease. *Asymptomatic primary hyperparathyroidism* is defined as biochemically confirmed hyperparathyroidism (elevated or inappropriately normal PTH levels despite hypercalcemia) with the absence of symptoms typically associated with more severe hyperparathyroidism and can occur with or without target organ involvement such as renal or bone disease.

Five conferences on the topic have been held in the United States over the past two decades, with the most recent in 2022. The published proceedings provide guidelines for diagnosis and treatment of primary hyperparathyroidism.

Issues of concern include the potential for cardiovascular deterioration, the presence of subtle neuropsychiatric symptoms, and the longer-term status of skeletal integrity in patients not treated surgically. The current consensus is that medical monitoring rather than surgical correction of hyperparathyroidism may be justified in certain patients. The current recommendation is that patients who show mild disease, as defined by the meeting guidelines **(Table 422-2)**, can be safely followed under management guidelines **(Table 422-3)**. There is, however, growing uncertainty about subtle disease manifestations and whether

TABLE 422-2 Guidelines for Surgery in Asymptomatic Primary Hyperparathyroidism

PARAMETER	GUIDELINE
Serum calcium	>1 mg/dL above normal
Renal	Creatinine clearance <60 mL/min 24-h urine for calcium >300 mg/d in men or >250 mg/d in women and increased stone risk by biochemical stone risk analysis Presence of nephrolithiasis or nephrocalcinosis by x-ray, ultrasound, or other imaging modalities
Skeletal	BMD by DXA: T score <–2.5 at any site Vertebral fracture by x-ray or VFA
Age	<50

Abbreviations: BMD, bone mineral density; DXA, dual-energy x-ray absorptiometry; VFA, vertebral fracture assessment.

Source: Data from JP Bilezikian et al: Evaluation and management of primary hyperparathyroidism: Summary statement and guidelines from the Fifth International Workshop. JBMR 37:2293, 2022.

surgery is therefore indicated in most patients. Among the issues is the evidence of eventual (>8 years) deterioration in bone mineral density after a decade of relative stability. There is concern that this late-onset deterioration in bone density in nonoperated patients could contribute significantly to the well-known age-dependent fracture risk (osteoporosis). Significant and sustained improvements in bone mineral density are seen after successful parathyroidectomy, and there is some evidence for reduction in fractures.

Cardiovascular disease, including left ventricular hypertrophy, cardiac functional defects, and endothelial dysfunction, has been reported as reversible in European patients with more severe symptomatic disease after surgery, leading to numerous studies of these cardiovascular features in those with milder disease. There are reports of endothelial dysfunction in patients with mild asymptomatic hyperparathyroidism, but the expert panels concluded that more observation is needed, especially regarding whether there is reversibility with surgery.

A topic of considerable interest and some debate is assessment of neuropsychiatric status and health-related quality of life status in hyperparathyroid patients both before surgery and in response to parathyroidectomy. Several observational studies have suggested improvements in symptom score after surgery. Randomized studies of surgery versus observation, however, have yielded inconclusive results, especially regarding benefits of surgery. Many studies report that hyperparathyroidism is associated with increased neuropsychiatric symptoms, but it is not possible at present to determine which patients might improve after surgery.

Diagnosis The diagnosis is typically made by detecting an elevated or inappropriately normal immunoreactive PTH level in a patient with asymptomatic hypercalcemia **(Fig. 422-4)** (see "Differential Diagnosis: Special Tests," below). Serum phosphate is usually low but may be normal, especially if renal failure has developed.

TABLE 422-3 Guidelines for Monitoring in Patients with Primary Hyperparathyroidism Who Do Not Undergo Parathyroidectomy

PARAMETER	GUIDELINE
Serum calcium and 25OHD	Annually
Renal	eGFR, annually; serum creatinine, annually. Abdominal imaging (x-ray, ultrasound, or CT), if clinically indicated. 24-h urine for calcium, if indicated
Creatinine clearance	Annually
Skeletal	DXA every 1–2 years (3 sites) (unless BMD is normal). X-ray or VFA of spine if clinically indicated (e.g., height loss, back pain)

Abbreviations: BMD, bone mineral density; CT, computed tomography; DXA, dual-energy x-ray absorptiometry; eGFR, estimated glomerular filtration rate; VFA, vertebral fracture assessment.

Source: Data from JP Bilezikian et al: Evaluation and management of primary hyperparathyroidism: Summary statement and guidelines from the Fifth International Workshop. JBMR 37:2293, 2022.

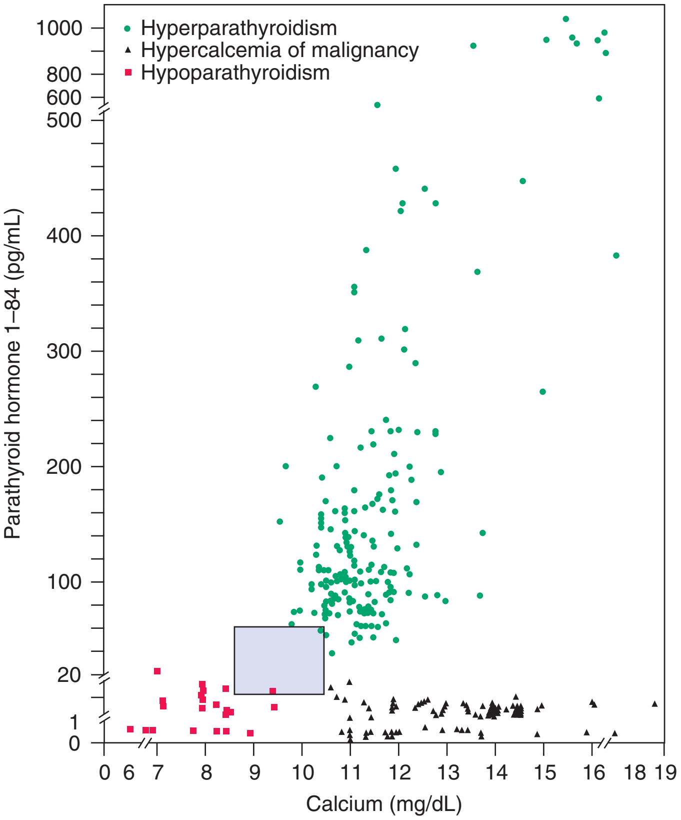

FIGURE 422-4 Levels of immunoreactive parathyroid hormone (PTH) detected in patients with primary hyperparathyroidism, hypercalcemia of malignancy, and hypoparathyroidism. Boxed area represents the upper and normal limits of blood calcium and/or immunoreactive PTH. *(Reproduced with permission from SR Nussbaum et al (eds): Endocrinology, 4th ed. Philadelphia, PA: Saunders; 2001.)*

Several modifications in PTH assays have been introduced in efforts to improve their utility in light of information about metabolism of PTH (as discussed above). First-generation assays were based on displacement of radiolabeled PTH from anti-PTH antibodies that often reacted with PTH fragments. Second-generation, double-antibody, or immunometric assays (one antibody that is usually directed against the carboxyl-terminal portion of intact PTH to capture the hormone and a second enzyme-labeled antibody that is usually directed against the amino-terminal portion of intact PTH) greatly improved the diagnostic discrimination of the tests by eliminating interference from circulating biologically inactive fragments, detected by the original first-generation assays. Third-generation assays, which detect even fewer inactive fragments, may be useful for clinical research studies as in management of chronic renal disease but have not replaced second-generation assays, which reliably help make the diagnosis of hypo- and hyperparathyroidism and differentiate this disease from other conditions of hypo- and hypercalcemia

TREATMENT

Primary Hyperparathyroidism

Surgical excision of the abnormal parathyroid tissue is the definitive therapy for this disease. As noted above, medical surveillance without operation for patients with mild, asymptomatic disease is, however, still preferred by some physicians and patients, particularly when the patients are more elderly. Evidence favoring surgery, if medically feasible, is growing because of concerns about skeletal, cardiovascular, and neuropsychiatric disease, even in mild hyperparathyroidism.

Two surgical approaches are generally practiced. The conventional parathyroidectomy procedure is neck exploration with general anesthesia; however, an outpatient procedure with local anesthesia, termed *minimally invasive parathyroidectomy,* is gaining traction though it has not yet replaced the traditional surgical approach.

Parathyroid exploration is challenging and should be undertaken by a surgeon experienced in this procedure. Certain features, like multiple abnormal glands in familial cases along with presurgical identification of one enlarged gland using several different imaging methods, can assist in determining the surgical approach. However, some critical decisions regarding management can be made only during the operation.

Preoperative imaging, such as neck ultrasounds, ^{99m}Tc sestamibi scans with single-photon emission CT (SPECT), C(11) choline PET/CT, and four-dimensional (4D) CT are used to predict the location of an abnormal gland. Intraoperative monitoring of PTH levels by rapid PTH immunoassays may be useful in guiding the surgery and a rapid fall (>50%) to normal levels of PTH is used in many centers to predict successful removal of the culprit gland(s).

Multiple-gland hyperplasia, as predicted in familial cases, and reoperation pose more difficult questions of surgical management and increase the risk of developing permanent hypoparathyroidism. Immediate transplantation of a portion of a removed, minced parathyroid gland into the muscles of the forearm is sometimes performed, with the view that surgical excision is easier from the ectopic site in the arm if there is recurrent hyperfunction.

In a minority of cases, if no abnormal parathyroid glands are found in the neck, the issue of further exploration must be decided. There are documented cases of five or six parathyroid glands and of unusual locations for adenomas such as in the mediastinum.

A decline in serum calcium often occurs within 24 h after successful surgery; usually, blood calcium falls to low-normal values for 3–5 days until the remaining parathyroid tissue resumes full hormone secretion. Risk factors for acute postoperative hypocalcemia is undermineralized bone matrix leading to "hungry bone" and vitamin D deficiency. With unexpected hypocalcemia, coexistent hypomagnesemia should be considered, as it interferes with PTH secretion and impairs response to PTH **(Chap. 421)**.

Transient hypoparathyroidism can occur but typically resolves within days; protracted recovery over several months can occur. Patients are unlikely to develop permanent hypoparathyroidism (often defined as hypoparathyroidism that persists 12 months after surgery) if their PTH levels tested within 12–24 h after surgery are >15 pg/mL.

Signs of hypocalcemia include symptoms such as muscle twitching, a general sense of anxiety, and positive Chvostek's and Trousseau's signs coupled with hypocalcemia. Therapy with oral calcium and sometimes low-dose calcitriol is often sufficient. Parenteral calcium therapy should be instituted when severe hypocalcemia is present. The rate and duration of IV therapy are determined by the severity of the symptoms and the response of the serum calcium to treatment. An infusion of 0.5–2 mg/kg per hour or 30–100 mL/h of a 1-mg/mL solution usually suffices to relieve symptoms. Usually, parenteral therapy is required for only a few days. If symptoms worsen or if parenteral calcium is needed for >2–3 days, therapy with a vitamin D analogue and/or oral calcium (2–4 g/d) should be started (see below). It is cost-effective to use calcitriol (doses of 0.5–1 μg/d) because of the rapidity of onset of effect and prompt cessation of action when stopped, in comparison to other forms of vitamin D. A rise in blood calcium after several months of treatment with calcium and calcitriol may indicate restoration of parathyroid function to normal. It is also appropriate to monitor serum PTH serially to estimate gland function in such patients.

If magnesium deficiency is present, it can complicate the postoperative course since significant magnesium deficiency impairs the secretion of PTH. Hypomagnesemia should be corrected whenever detected, typically with oral magnesium replacement, but parenteral repletion may be necessary.

MEDICAL MANAGEMENT

Medical monitoring rather than corrective surgery is still acceptable, but it is clear that surgical intervention is the more frequently recommended option for the reasons noted above. Despite the usefulness of the guidelines, the importance of individual patient and physician judgment and preference is clear in all recommendations.

There is no long-term experience regarding specific clinical outcomes such as fracture prevention, but it has been established that bisphosphonates increase bone mineral density significantly without changing serum calcium. Calcimimetics reduce PTH secretion and thus lower serum calcium but do not affect bone mineral density (BMD). A Scandinavian randomized clinical trial of parathyroidectomy versus observation in patients with mild primary hyperparathyroidism revealed no differences in morbidity or mortality after 10 years.

OTHER PARATHYROID CAUSES OF HYPERCALCEMIA

Lithium Therapy Lithium, used in the management of bipolar depression and other psychiatric disorders, causes hypercalcemia in ~10% of treated patients. The hypercalcemia is dependent on continued lithium treatment, remitting and recurring when lithium is stopped and restarted. The parathyroid adenomas reported in some hypercalcemic patients with lithium therapy may reflect the presence of an independently occurring parathyroid tumor; a permanent effect of lithium on parathyroid gland growth need not be implicated as most patients have complete reversal of hypercalcemia when lithium is stopped. However, long-standing stimulation of parathyroid cell replication by lithium may predispose to development of adenomas (as is documented in secondary hyperparathyroidism and renal failure).

At the levels achieved in blood in treated patients, lithium can be shown in vitro to shift the PTH secretion curve to the right in response to calcium; i.e., higher calcium levels are required to lower PTH secretion, probably acting at the calcium sensor (see below). This effect can cause elevated PTH levels and consequent hypercalcemia in otherwise normal individuals. Cinacalcet has been used successfully in these patients. Fortunately, there are usually alternative medications for the underlying psychiatric illness. Parathyroid surgery should not be recommended unless hypercalcemia and elevated PTH levels persist after lithium is discontinued.

GENETIC DISORDERS CAUSING HYPERPARATHYROIDISM-LIKE SYNDROMES

Familial Hypocalciuric Hypercalcemia FHH (also called *familial benign hypercalcemia*) is inherited as an autosomal dominant trait. Affected individuals are discovered because of asymptomatic hypercalcemia. Most cases of FHH are caused by an inactivating heterozygous CaSR mutations; this disorder, designated as FHH type 1, leads to inappropriately normal or even increased PTH secretion. Other forms of FHH are caused either by heterozygous loss-of-function mutations in *GNA11* (encoding Gα11), one of the signaling proteins at the CaSR (FHH2), or by heterozygous mutations in *AP2S1* (FHH3).

In FHH1, the primary defect is abnormal sensing of the blood calcium by the parathyroid gland and renal tubule, causing inappropriate secretion of PTH and excessive reabsorption of calcium in the distal renal tubules. Many different inactivating CaSR mutations have been identified in patients with FHH1. These mutations lower the capacity of the sensor to bind calcium and the mutant receptors function as though blood calcium levels were low; thus, inappropriate secretion of PTH occurs from an otherwise normal gland. Approximately two-thirds of patients with FHH have mutations within the protein-coding region of the CaSR gene (FHH1). Others have mutations in *GNA11*, the gene encoding the alpha-subunit of G11, a G-protein through which the CaSR signals (FHH2). These loss-of-function mutations have similar consequences to the loss-of-function mutations in the CaSR. FHH3 is caused by heterozygous mutations of the adaptor protein-2δ subunit (AP2S1), which is a component of clathrin-coated vesicles, and critical in clathrin-mediated endocytosis.

Since FHH is a life-long disease that cannot be cured by parathyroidectomy, it is critical to distinguish this rare disease from the more common primary hyperparathyroidism. One striking exception to the

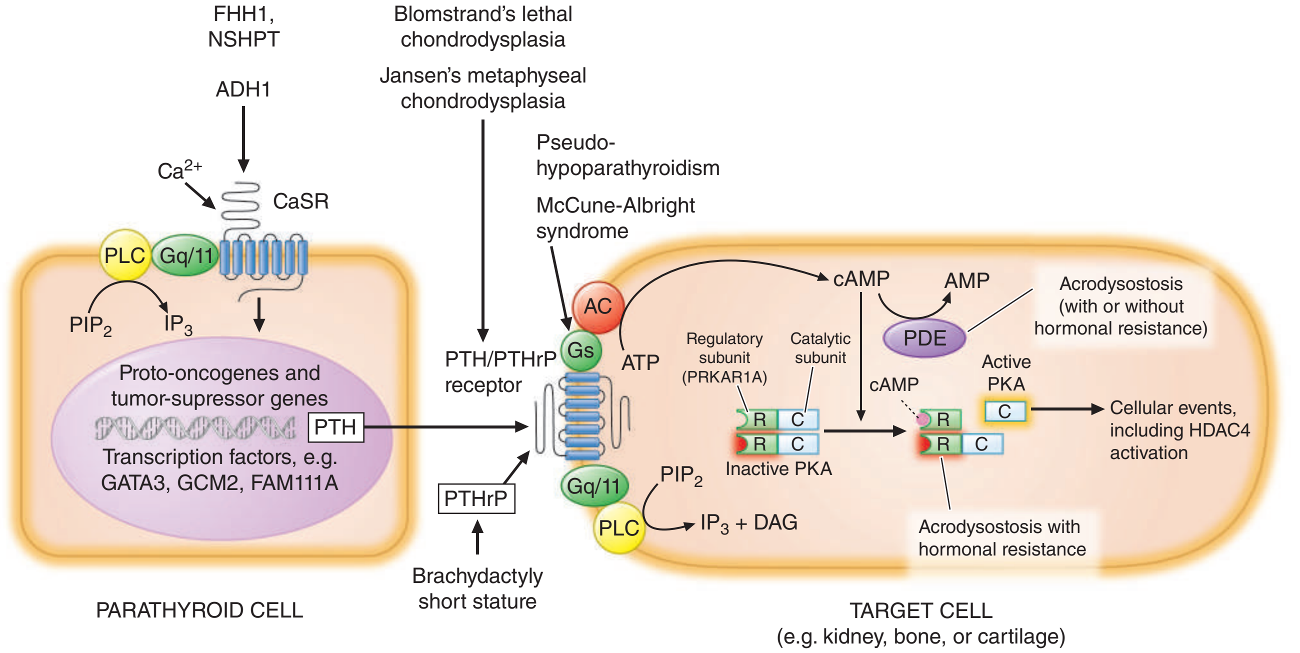
FHH1,
NSHPT
ADH1
Blomstrand's lethal chondrodysplasia
Jansen's metaphyseal chondrodysplasia
Pseudo-hypoparathyroidism
McCune-Albright syndrome
Ca2+
CaSR
PLC
Gq/11
PIP2
IP3
Proto-oncogenes and tumor-supressor genes
PTH
Transcription factors, e.g. GATA3, GCM2, FAM111A
PTH/PTHrP receptor
PTHrP
Brachydactyly short stature
AC
Gs
ATP
cAMP
AMP
PDE
Acrodysostosis (with or without hormonal resistance)
Regulatory subunit (PRKAR1A)
Catalytic subunit
R
C
Inactive PKA
cAMP
Active PKA
Cellular events, including HDAC4 activation
Gq/11
PLC
PIP2
IP3 + DAG
Acrodysostosis with hormonal resistance
PARATHYROID CELL
TARGET CELL
(e.g. kidney, bone, or cartilage)

FIGURE 422-5 Illustration of some genetic mutations that alter calcium metabolism by effects on the parathyroid cell or target cells of parathyroid hormone (PTH) action. Alterations in PTH production by the parathyroid cell can be caused by changes in the response to extracellular fluid calcium (Ca^{2+}) that are detected by the calcium-sensing receptor (CaSR). Furthermore, PTH (or PTH-related peptide [PTHrP]) can show altered efficacy in target cells such as in proximal tubular cells, by altered function of its receptor (PTH/PTHrP receptor) or the signal transduction proteins, G proteins such as Gsα that is linked to adenylate cyclase (AC), the enzyme responsible for producing cyclic AMP (cAMP) (also illustrated are Gαq/Gα11, which activate an alternate pathway of receptor signal transmission involving the generation of inositol triphosphate [IP_3] or diacylglycerol [DAG]). Heterozygous loss-of-function mutations in the CaSR cause familial benign hypocalciuric hypercalcemia (FBHH) and homozygous mutations (both alleles mutated) and neonatal severe hyperparathyroidism (NSHPT); heterozygous gain-of-function causes autosomal dominant hypercalciuric hypocalcemia (ADH1). Other defects in parathyroid cell function that occur at the level of gene regulation (oncogenes or tumor-suppressor genes) or transcription factors are discussed in the text. Blomstrand's lethal chondrodysplasia is due to homozygous or compound heterozygous loss-of-function mutations in the PTH/PTHrP receptor, a neonatally lethal disorder, while pseudohypoparathyroidism involves inactivation at the level of the G proteins, specifically mutations that eliminate or reduce Gsα activity in the kidney (see text for details). Acrodysostosis can occur with (mutant regulatory subunit of PKA) or without hormonal resistance (mutant *PDE4D* or *PDE3A*). Jansen's metaphyseal chondrodysplasia and McCune-Albright syndrome represent gain-of-function mutations in the PTH/PTHrP receptor and Gsα protein, respectively.

rule against parathyroid surgery in this syndrome is the occurrence, usually in consanguineous marriages (due to the rarity of the gene mutation), of a homozygous or compound heterozygote state, resulting in severe impairment of CaSR function. In this condition, neonatal severe hypercalcemia, total parathyroidectomy is mandatory, but calcimimetics have been used as a temporary measure.

Patients with FHH have lifelong hypercalcemia, which is typically mild and about 11 mg/dL. PTH values are often within normal laboratory range (which is inappropriate for hypercalcemia) or mildly elevated. Family history is often positive but can be negative because the patient has a de novo mutation or the penetrance of the disease is not 100%. Patients with primary hyperparathyroidism have <99% renal calcium reabsorption, whereas most patients with FHH have >99% reabsorption. The calcium–to–creatine clearance ratio (CCCR), determined on spot or 24-hour urine, can be used to help differentiate the two disorders, with the CCCR typically <0.01 in FHH and >0.02 in primary hyperparathyroidism. This distinction is far from perfect, and genetic testing is helpful in establishing the diagnosis. Panel testing for hypercalcemic disorders allows the simultaneous sequencing of all genes known to be implicated in hypercalcemia, including the genes causing FHH.

Rare but well-documented cases of acquired hypocalciuric hypercalcemia are reported due to antibodies against the CaSR. They appear to be a complication of an underlying autoimmune disorder and respond to therapies directed against the underlying disorder.

Jansen's Disease Activating mutations in the PTH/PTHrP receptor (PTH1R) have been identified as the cause of this rare autosomal dominant syndrome. Because the mutations lead to constitutive activation of receptor function, one abnormal copy of the mutant receptor is sufficient to cause the disease, thereby accounting for its dominant mode of transmission. Besides often severe hypercalcemia, patients affected by Jansen's disease have short-limbed dwarfism due to abnormal regulation of chondrocyte maturation in the growth plates of the bone that are formed through the endochondral process. In adult life, there are numerous abnormalities in bone, including multiple cystic resorptive areas resembling those seen in severe hyperparathyroidism. Hypercalcemia and hypophosphatemia with undetectable or low PTH levels are typically observed. The pathogenesis of the growth plate abnormalities in Jansen's disease has been confirmed by transgenic experiments in which targeted expression of the mutant PTH/PTHrP receptor to the proliferating chondrocyte layer of growth plate emulated several features of the human disorder. Other genetic mutations in the parathyroid gland or PTH target cells that affect Ca^{2+} metabolism are illustrated in Fig. 422-5.

MALIGNANCY-RELATED HYPERCALCEMIA

Clinical Syndromes and Mechanisms of Hypercalcemia

Hypercalcemia due to malignancy is common (occurring in as many as 20% of cancer patients, especially with certain types of tumors such as lung carcinoma), often severe and difficult to manage, and, typically easy to distinguish from primary hyperparathyroidism by a suppressed PTH. Although malignancy is usually clinically obvious or readily detectable by medical history, hypercalcemia can occasionally be due to an occult tumor.

Three main mechanisms of hypercalcemia are operative in cancer hypercalcemia. Humoral hypercalcemia of malignancy (HHM) is caused by tumors producing and secreting PTHrP that causes a clinical picture similar to primary hyperparathyroidism with increased bone resorption and hypercalcemia. However, PTH is suppressed. Patients with HHM may have low to normal levels of $1,25(OH)_2D$, instead of elevated levels as in true hyperparathyroidism probably reflecting subtle differences in the activation of the PTH1R by PTHrP versus PTH. Squamous cell carcinomas and renal, bladder, and colorectal cancer are examples of tumors that can cause HHM. Several different assays (single- or double-antibody, different epitopes) have been developed to detect PTHrP and do not cross-react with PTH. Most data indicate that circulating PTHrP levels are undetectable or low in normal individuals except in pregnancy (high in human milk) and elevated in most cancer patients with the humoral syndrome.

Alternatively, local osteolysis leading to release of cytokines (e.g., interleukin 1 and tumor necrosis factor) that activate osteoclasts occurs with hematologic malignancies such as leukemia, lymphoma, and multiple myeloma, but also with breast cancer.

A third mechanism leading to malignancy-associated hypercalcemia is an increased production and blood level of 1,25(OH)$_2$D, produced by abnormal lymphocytes or adjacent macrophages in lymphoma but also in ovarian dysgerminomas.

The etiologic mechanisms in cancer hypercalcemia may be multiple, even in the same patient. For example, in breast carcinoma (metastatic to bone) and in a distinctive type of T-cell lymphoma/leukemia initiated by human T-cell lymphotropic virus 1, hypercalcemia is caused by direct local lysis of bone as well as by a humoral mechanism involving excess production of PTHrP.

Hyperparathyroidism has been reported to coexist with the humoral cancer syndrome, and rarely, ectopic hyperparathyroidism due to tumor elaboration of true PTH is reported.

TREATMENT

Malignancy-Related Hypercalcemia

Treatment of the hypercalcemia of malignancy is first directed to control of tumor; reduction of tumor mass usually corrects hypercalcemia. If a patient has severe hypercalcemia yet has a good chance for effective tumor therapy, treatment of the hypercalcemia should be vigorous while awaiting the results of definitive therapy (see "General Approach to Hypercalcemic States" below). If hypercalcemia occurs in the late stages of a tumor that is resistant to antitumor therapy, the treatment of the hypercalcemia should be judicious as high calcium levels can have a mild sedating effect. Standard therapies for hypercalcemia (discussed below) are applicable to patients with malignancy.

■ VITAMIN D–RELATED HYPERCALCEMIA

Vitamin D–mediated hypercalcemia can be due to excessive ingestion of vitamin D analogues or abnormal metabolism of the vitamin. Abnormal metabolism of the vitamin is usually acquired in association with a widespread granulomatous disorder. Vitamin D metabolism is carefully regulated, particularly the activity of renal 1α-hydroxylase, the enzyme responsible for the production of 1,25(OH)$_2$D **(Chap. 421)**. The regulation of 1α-hydroxylase in sites other than the renal tubule differs and lacks the negative feedback regulations; these phenomena may explain the occurrence of hypercalcemia secondary to excessive 1,25(OH)$_2$D production in patients with sarcoidosis or lymphoma.

Vitamin D Intoxication Chronic ingestion of >10 times the normal physiologic requirement of vitamin D (amounts >10,000 U/d) is usually required to produce significant hypercalcemia in otherwise healthy individuals. The stated upper limit of safe dietary intake is 2000 U/d (50 μg/d) in adults because of concerns about potential toxic effects of cumulative supraphysiologic doses. These recommendations are now regarded as too restrictive, since some estimates are that, in elderly individuals in northern latitudes, ≥2000 U/d may be necessary to avoid vitamin D insufficiency.

Hypercalcemia in vitamin D intoxication is due to an excessive biologic action of the vitamin, perhaps the consequence of increased levels of 25(OH)D rather than merely increased levels of the active metabolite 1,25(OH)$_2$D (the latter may not be frankly elevated in vitamin D intoxication). These actions lead to both increased intestinal absorption of calcium and increased release of calcium from bone. 25(OH)D has definite, if low, biologic activity in the intestine and bone. The production of 25(OH)D is less tightly regulated than is the production of 1,25(OH)$_2$D. Hence, concentrations of 25(OH)D can be elevated severalfold in patients with excess vitamin D intake.

The diagnosis is substantiated by documenting elevated levels of 25(OH)D >100 ng/mL. Hypercalcemia is usually controlled by restriction of dietary calcium intake and appropriate attention to hydration. These measures, plus discontinuation of vitamin D, usually lead to resolution of hypercalcemia. However, because of the increased bone resorption caused by high levels of vitamin D, simple cessation of calcium intake is often insufficient therapy. Further, 25(OH)D stores in fat may be substantial, and vitamin D intoxication may persist for weeks after vitamin D ingestion is terminated. Such patients are responsive to glucocorticoids, which in doses of 40–100 mg/d of prednisone or its equivalent, usually return serum calcium levels to normal over several days; severe intoxication may require intensive therapy.

Sarcoidosis and Other Granulomatous Diseases In patients with sarcoidosis and other granulomatous diseases, such as tuberculosis and fungal infections, excess 1,25(OH)$_2$D is synthesized in macrophages or other cells in the granulomas. Indeed, increased 1,25(OH)$_2$D levels have been reported in anephric patients with sarcoidosis and hypercalcemia. Macrophages obtained from granulomatous tissue convert 25(OH)D to 1,25(OH)$_2$D at an increased rate. There is a positive correlation in patients with sarcoidosis between 25(OH)D levels (reflecting vitamin D intake) and the circulating concentrations of 1,25(OH)$_2$D, whereas normally, there is no increase in 1,25(OH)$_2$D with increasing 25(OH)D levels due to multiple feedback controls on renal 1α-hydroxylase **(Chap. 421)**. The usual regulation of active metabolite production by calcium and phosphate or by PTH does not operate in these patients. Instead, macrophages increase their production of the vitamin D receptor and of the 1α-hydroxylase in response to tumor necrosis factor and other inflammatory stimuli. PTH levels are usually low and 1,25(OH)$_2$D levels are elevated, but primary hyperparathyroidism and sarcoidosis may coexist in some patients.

Management of the hypercalcemia can often be accomplished by avoiding excessive sunlight exposure and limiting vitamin D and calcium intake. Presumably, however, the abnormal sensitivity to vitamin D and abnormal regulation of 1,25(OH)$_2$D synthesis will persist as long as the disease is active. Alternatively, glucocorticoids in the equivalent of 100 mg/d of hydrocortisone or equivalent doses of glucocorticoids may help control hypercalcemia. Glucocorticoids appear to act by blocking excessive production of 1,25(OH)$_2$D, as well as the response to it in target organs.

Hypercalcemia of Infancy Several variants of this rare abnormality of calcium homeostasis are now known. For example, *Williams' syndrome* is an autosomal dominant disorder characterized by multiple congenital development defects, including supravalvular aortic stenosis, intellectual disability, and an elfin facies, in association with hypercalcemia due to abnormal sensitivity to vitamin D. The hypercalcemia associated with the syndrome was first recognized in England, where it was thought, incorrectly, to be caused by the fortification of milk with vitamin D. The cardiac and developmental abnormalities were independently described, but the connection between these defects and hypercalcemia was not described until later. Levels of 1,25(OH)$_2$D can be elevated, ranging from 46 to 120 nmol/L (150–500 pg/mL). The mechanism of the abnormal sensitivity to vitamin D and of the increased circulating levels of 1,25(OH)$_2$D is still unclear. Studies suggest that genetic mutations involving microdeletions at the elastin locus and perhaps other genes on chromosome 7 may play a role in the pathogenesis.

Another genetic cause of hypercalcemia that starts in infancy is 24-hydroxylase deficiency that impairs catabolism of 1,25(OH)$_2$D caused by biallelic inactivating mutations in CYP24A1. Another rare cause of hypercalcemia involves mutation in the renal sodium-dependent phosphate transporters that lead to increased production of 1,25(OH)$_2$D leading to hypercalcemia (*NPT2a* mutations lead to more severe hypercalcemia than *NPT2c* mutations).

■ HIGH-BONE-TURNOVER STATES

Hyperthyroidism As many as 20% of hyperthyroid patients have high-normal or mildly elevated serum calcium concentrations;

hypercalciuria is even more common. The hypercalcemia is due to increased bone turnover, with bone resorption exceeding bone formation. Usually, the diagnosis is obvious, but signs of hyperthyroidism may occasionally be occult, particularly in the elderly (Chap. 396). Hypercalcemia is managed by treatment of the hyperthyroidism.

Immobilization Immobilization is a rare cause of hypercalcemia in adults in the absence of an associated disease but may cause hypercalcemia in children and adolescents, particularly after spinal cord injury and paraplegia or quadriplegia. With resumption of ambulation, the hypercalcemia in children usually returns to normal.

The mechanism appears to involve a disproportion between bone formation and bone resorption; the former decreased and the latter increased. Hypercalciuria and increased mobilization of skeletal calcium can develop in normal volunteers subjected to extensive bed rest, although hypercalcemia is unusual. Immobilization of an adult with a disease associated with high bone turnover, however, such as Paget's disease, may cause hypercalcemia.

Thiazides Administration of thiazides can cause hypercalcemia in patients with high rates of bone turnover. Commonly, thiazides are associated with aggravation of hypercalcemia in primary hyperparathyroidism, but this effect can be seen in other high-bone-turnover states as well. The mechanism of thiazide action is complex. Chronic thiazide administration leads to reduction in urinary calcium; the hypocalciuric effect appears to reflect the enhancement of proximal tubular resorption of sodium and calcium in response to sodium depletion. Some of this renal effect is due to augmentation of PTH action and is more pronounced in individuals with intact PTH secretion. However, thiazides cause hypocalciuria in hypoparathyroid patients if sodium intake is restricted. This finding is the rationale for the use of thiazides as an adjunct to therapy in hypoparathyroid patients, as discussed below. Thiazide administration to normal individuals causes a transient increase in blood calcium (usually within the high-normal range) that reverts to preexisting levels after a week or more of continued administration. If hormonal function and calcium and bone metabolism are normal, homeostatic controls are reset to counteract the mild calcium-elevating effect of the thiazides. In the presence of hyperparathyroidism or increased bone turnover from another cause, homeostatic mechanisms are ineffective. The abnormal effects of the thiazide on calcium metabolism disappear within days of cessation of the drug.

Vitamin A Intoxication Vitamin A intoxication is a rare cause of hypercalcemia and is most commonly a side effect of dietary faddism (Chap. 344). Calcium levels can be elevated into the 3- to 3.5-mmol/L (12–14 mg/dL) range after the ingestion of 50,000–100,000 units of vitamin A daily (10–20 times the minimum daily requirement). Typical features of severe hypercalcemia include fatigue, anorexia, and, in some, severe muscle and bone pain. Excess vitamin A intake is presumed to increase bone resorption.

The diagnosis can be established by history and by measurement of vitamin A levels in serum. Diagnostic challenges may arise in CKD where mildly elevated vitamin A levels are often seen with no association to hypercalcemia. Occasionally, skeletal x-rays reveal periosteal calcifications, particularly in the hands. Withdrawal of the vitamin is usually associated with prompt disappearance of the hypercalcemia and reversal of the skeletal changes. As in vitamin D intoxication, administration of 100 mg/d hydrocortisone or its equivalent leads to a rapid return of the serum calcium to normal.

HYPERCALCEMIA ASSOCIATED WITH CHRONIC KIDNEY DISEASE

The pathogenesis of secondary hyperparathyroidism in CKD is multifactorial and includes resistance to PTH and an increase in FGF23, which, in turn, inhibits the renal 1α-hydroxylase, thus reducing $1,25(OH)_2D$ levels and leading to further increases of PTH.

Occasional patients develop severe manifestations of secondary hyperparathyroidism, including hypercalcemia, pruritus, extraskeletal calcifications, and painful bones, despite aggressive medical efforts to suppress the hyperparathyroidism. It is now recognized that a true clonal outgrowth (irreversible) can arise in long-standing, inadequately treated CKD. PTH hypersecretion no longer responsive to medical therapy, a state of severe hyperparathyroidism in patients with CKD that requires surgery, has been referred to as *tertiary hyperparathyroidism*.

TREATMENT

Hypercalcemia in Tertiary Hyperparathyroidism

Medical therapy to reverse secondary hyperparathyroidism in CKD includes reduction of excessive blood phosphate by restriction of dietary phosphate, the use of nonabsorbable phosphate binders, and careful, selective addition of calcitriol (0.25–2 μg/d) or related analogues. Calcium carbonate became preferred over aluminum-containing antacids to prevent aluminum-induced bone disease. However, synthetic gels that also bind phosphate (such as sevelamer; Chap. 322) are now widely used, with the advantage of avoiding not only aluminum retention but also excess calcium loading, which may contribute to cardiovascular calcifications. Intravenous calcitriol (or related analogues), administered as several pulses each week, helps control secondary hyperparathyroidism. Aggressive but carefully administered medical therapy can often, but not always, reverse hyperparathyroidism and its symptoms and manifestations.

Parathyroid surgery is necessary to control tertiary hyperparathyroidism. Based on genetic evidence from examination of tumor samples in these patients, the emergence of autonomous parathyroid function is due to a monoclonal outgrowth of one or more previously hyperplastic parathyroid glands. The adaptive response has become an independent contributor to disease; this finding seems to emphasize the importance of optimal medical management to reduce the proliferative response of the parathyroid cells that enables the irreversible genetic change.

OTHER CAUSES OF HYPERCALCEMIA

Aluminum Intoxication Aluminum intoxication (and often hypercalcemia as a complication of medical treatment) in the past occurred in patients on chronic dialysis; manifestations included acute dementia and unresponsive and severe osteomalacia. Bone pain, multiple nonhealing fractures, particularly of the ribs and pelvis, and a proximal myopathy occur. Hypercalcemia develops when these patients are treated with vitamin D or calcitriol because of impaired skeletal responsiveness. Aluminum is present at the site of osteoid mineralization, osteoblastic activity is minimal, and calcium incorporation into the skeleton is impaired. The disorder is now rare because of the avoidance of aluminum-containing antacids or aluminum excess in the dialysis regimen.

Milk-Alkali Syndrome The milk-alkali syndrome is due to excessive ingestion of calcium and absorbable antacids such as milk or calcium carbonate. It is much less frequent since proton pump inhibitors and other treatments became available for peptic ulcer disease. For a time, the increased use of calcium carbonate in the management of secondary hyperparathyroidism led to reappearance of the syndrome. Several clinical presentations—acute, subacute, and chronic—have been described, all of which feature hypercalcemia, alkalosis, and renal failure. The chronic form of the disease, termed *Burnett's syndrome*, is associated with irreversible renal damage. The acute syndromes reverse if the excess calcium and absorbable alkali are stopped.

Individual susceptibility is important in the pathogenesis, as some patients are treated with calcium carbonate and alkali regimens without developing the syndrome. One variable is the fractional calcium absorption as a function of calcium intake. Some individuals absorb a high fraction of calcium, even with intakes ≥2 g of elemental calcium per day, instead of reducing calcium absorption with high intake, as occurs in most normal individuals. Resultant mild hypercalcemia after meals in such patients is postulated to contribute to the generation of alkalosis. Development of hypercalcemia causes increased sodium excretion and some depletion of total-body water. These phenomena and perhaps some suppression of endogenous PTH secretion due to mild hypercalcemia lead to increased bicarbonate resorption and to alkalosis in the face of continued calcium carbonate ingestion. Alkalosis per se selectively enhances calcium resorption in the distal nephron, thus aggravating the hypercalcemia. The cycle of mild hypercalcemia → bicarbonate retention → alkalosis → renal calcium retention → severe hypercalcemia perpetuates and aggravates hypercalcemia and alkalosis as long as calcium and absorbable alkali are ingested.

DIFFERENTIAL DIAGNOSIS OF HYPERCALCEMIA

Differential diagnosis of hypercalcemia is best achieved by using clinical criteria, but immunometric assays to measure PTH are especially useful in distinguishing among major causes **(Fig. 422-6)**. The clinical features that deserve emphasis are the presence or absence of symptoms or signs of disease and evidence of chronicity. If one discounts fatigue or depression, >90% of patients with primary hyperparathyroidism have *asymptomatic hypercalcemia*; symptoms of malignancy are usually present in cancer-associated hypercalcemia. Disorders other than hyperparathyroidism and malignancy cause <10% of cases of hypercalcemia, and some of the nonparathyroid causes are associated with clear-cut manifestations such as renal failure.

Hyperparathyroidism is the likely diagnosis in patients with *chronic hypercalcemia*. If hypercalcemia has been manifesting for >1 year, malignancy as the underlying cause is very unlikely. A striking feature of malignancy-associated hypercalcemia is the rapidity of the course, whereby signs and symptoms of the underlying malignancy are evident within months of the detection of hypercalcemia. Although clinical considerations are helpful in arriving at the correct diagnosis of the cause of hypercalcemia, appropriate laboratory testing is essential for definitive diagnosis. The immunoassay for PTH usually separates primary hyperparathyroidism from all other causes of hypercalcemia (exceptions are very rare reports of ectopic production of excess PTH by nonparathyroid tumors). Patients with hyperparathyroidism have elevated (or nonsuppressed) PTH levels despite hypercalcemia, whereas patients with malignancy and the other causes of hypercalcemia (except for disorders mediated by PTH such as lithium-induced hypercalcemia) have very low or undetectable levels. Intact PTH assays, based on the double-antibody method for PTH, exhibit very high sensitivity (especially if serum calcium is simultaneously evaluated) and specificity for the diagnosis of primary hyperparathyroidism (Fig. 422-4).

In summary, PTH values are elevated in >90% of parathyroid-related causes of hypercalcemia, undetectable or low in malignancy-related, vitamin D–related and high-bone-turnover causes of hypercalcemia. In view of the specificity of the PTH immunoassay and the high frequency of hyperparathyroidism in hypercalcemic patients, it is cost-effective to measure the PTH level in all hypercalcemic patients unless malignancy or a specific nonparathyroid disease is obvious. False-positive PTH assay results are rare but can be due to heterotopic antibodies. Immunoassays for PTHrP are helpful in diagnosing certain types of malignancy-associated hypercalcemia. Although FHH is parathyroid-related, the disease should be managed distinctively from primary hyperparathyroidism. Clinical features, family history, and the low urinary calcium excretion can help make the distinction, and genetic testing confirms the diagnosis. Because the incidence of malignancy and hyperparathyroidism both increase with age, they can coexist as two independent causes of hypercalcemia.

1,25(OH)$_2$D levels are in the upper range of normal or frankly elevated in many (but not all) patients with primary hyperparathyroidism. In other disorders associated with hypercalcemia, concentrations of 1,25(OH)$_2$D are low or, at the most, normal. However, this test is of low specificity and is not cost-effective, as not all patients with hyperparathyroidism have elevated 1,25(OH)$_2$D levels and not all nonparathyroid hypercalcemic patients have suppressed 1,25(OH)$_2$D. Measurement of 1,25(OH)$_2$D is, however, critically valuable in establishing the cause of hypercalcemia in sarcoidosis and certain lymphomas.

A useful general approach is outlined in Fig. 422-6. If the patient is *asymptomatic* and there is evidence of *chronicity* to the hypercalcemia, hyperparathyroidism is likely the cause and FHH needs to be excluded. If PTH levels (usually measured at least twice) are elevated, the clinical impression is confirmed, and little additional diagnostic evaluation is necessary. If there is only a short history or no data as to the duration of the hypercalcemia, *occult malignancy* must be considered; if the PTH levels are suppressed then a thorough workup must be undertaken for malignancy, which may include chest x-ray, CT of chest and abdomen, and bone scan. Immunoassays for PTHrP may be especially useful in such situations. Attention should also be paid to clues for underlying hematologic disorders such as anemia, increased plasma globulin, and abnormal serum immunoelectrophoresis; bone scans can be negative

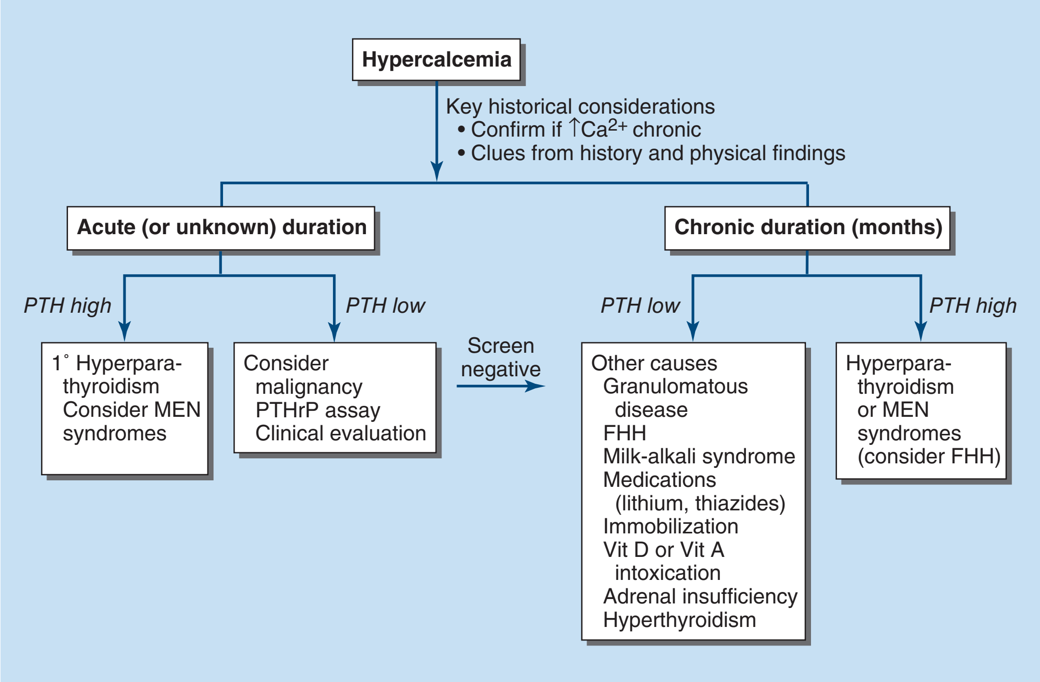
Hypercalcemia
Key historical considerations
• Confirm if ↑Ca2+ chronic
• Clues from history and physical findings
Acute (or unknown) duration
Chronic duration (months)
PTH high
PTH low
PTH low
PTH high
1° Hyperparathyroidism Consider MEN syndromes
Consider malignancy PTHrP assay Clinical evaluation
Screen negative
Other causes Granulomatous disease FHH Milk-alkali syndrome Medications (lithium, thiazides) Immobilization Vit D or Vit A intoxication Adrenal insufficiency Hyperthyroidism
Hyperparathyroidism or MEN syndromes (consider FHH)

FIGURE 422-6 Algorithm for the evaluation of patients with hypercalcemia. PTH levels (high or low) should be interpreted in the context of serum calcium levels, as they may be inappropriately high or low for the level of serum calcium. See text for details. FHH, familial hypocalciuric hypercalcemia; MEN, multiple endocrine neoplasia; PTH, parathyroid hormone; PTHrP, parathyroid hormone–related peptide; Vit, vitamin.

in some patients with metastases such as in multiple myeloma. Finally, if a patient with chronic hypercalcemia is asymptomatic and malignancy therefore seems unlikely on clinical grounds, but PTH values are not elevated, it is useful to search for other chronic causes of hypercalcemia such as occult sarcoidosis. A careful history of dietary supplements and drug use may suggest intoxication with vitamin D or vitamin A or the use of thiazides.

TREATMENT

General Approach to Hypercalcemic States

The approach to medical treatment of hypercalcemia varies with its severity. Mild hypercalcemia, <3 mmol/L (12 mg/dL), can usually be managed by hydration. More severe hypercalcemia (levels of 3.2–3.7 mmol/L [13–15 mg/dL]) must be managed more aggressively; above that level, hypercalcemia can be life-threatening and requires emergency measures **(Table 422-4)**. By using a combination of approaches in severe hypercalcemia, the serum calcium concentration can be decreased within 24–48 h in most patients, enough to relieve acute symptoms, prevent death from hypercalcemic crisis, and permit diagnostic evaluation. Therapy can then be directed at the underlying disorder—the second priority.

Hypercalcemia develops because of excessive skeletal calcium release, increased intestinal calcium absorption, or inadequate renal calcium excretion. Understanding the particular pathogenesis helps guide therapy. For example, hypercalcemia in patients with malignancy is primarily due to excessive skeletal calcium release and is, therefore, minimally improved by restriction of dietary calcium. On the other hand, patients with vitamin D hypersensitivity or vitamin D intoxication have excessive intestinal calcium absorption, and restriction of dietary calcium is beneficial. Decreased renal function or ECF depletion decreases urinary calcium excretion. In such situations, rehydration may rapidly reduce or reverse the hypercalcemia, even though increased bone resorption persists. As outlined below, the more severe the hypercalcemia, the greater the need for a combination of therapies. Rapid-acting (hours) approaches—rehydration, forced diuresis, and calcitonin—can be used with the most effective antiresorptive agents such as bisphosphonates (since severe hypercalcemia usually involves excessive bone resorption).

HYDRATION, INCREASED SALT INTAKE, AND MILD AND FORCED DIURESIS

The first principle of treatment is to restore normal hydration. Many hypercalcemic patients are dehydrated because of vomiting, inanition, and/or hypercalcemia-induced defects in urinary concentrating ability. The resultant drop in glomerular filtration rate is accompanied by an additional decrease in renal tubular sodium and calcium clearance. Restoring a normal ECF volume corrects these abnormalities and increases urine calcium excretion by 2.5–7.5 mmol/d (100–300 mg/d). Increasing urinary sodium excretion to 400–500 mmol/d increases urinary calcium excretion even further than simple rehydration. After rehydration has been achieved, saline can be administered, or furosemide or ethacrynic acid can be given to depress the tubular reabsorptive mechanism for calcium (care must be taken to prevent dehydration). The combined use of these therapies can increase urinary calcium excretion to ≥12.5 mmol/d (500 mg/d) in most hypercalcemic patients. Since this is a substantial percentage of the exchangeable calcium pool, the serum calcium concentration usually falls 0.25–0.75 mmol/L (1–3 mg/dL) within 24 h. Precautions should be taken to prevent potassium and magnesium depletion; calcium-containing renal calculi are a potential complication.

Under life-threatening circumstances, the preceding approach can be pursued more aggressively, but the availability of effective agents to block bone resorption (such as bisphosphonates) has reduced the need for extreme diuresis regimens (Table 422-4). Depletion of potassium and magnesium is inevitable unless replacements are given; pulmonary edema can be precipitated. The potential complications can be reduced by careful monitoring of central venous pressure and plasma or urine electrolytes; catheterization of the bladder may be necessary. Dialysis treatment may be needed when renal function is compromised.

BISPHOSPHONATES

The bisphosphonates are analogues of pyrophosphate, with high affinity for bone, especially in areas of increased bone turnover,

TABLE 422-4 Therapies for Severe Hypercalcemia

TREATMENT	ONSET OF ACTION	DURATION OF ACTION	ADVANTAGES	DISADVANTAGES
Most Useful Therapies				
IV hydration with normal saline	Hours	During infusion	Rehydration invariably needed	Volume overload
Forced diuresis; normal saline plus loop diuretic	Hours	During treatment	Rapid action	Volume overload, cardiac decompensation, intensive monitoring, electrolyte disturbance, inconvenience
Pamidronate	1–2 days	10–14 days to weeks	High potency; intermediate onset of action	Fever in 20%, hypophosphatemia, hypocalcemia, hypomagnesemia, rarely jaw necrosis
Zoledronate	1–2 days	>3 weeks	Same as for pamidronate (lasts longer)	Same as pamidronate above
Denosumab	1–2 days	>3 weeks	Strongest antiresorptive	Occasional severe hypocalcemia, rarely jaw necrosis, skin infections
Special Use Therapies				
Calcitonin	Hours	1–2 days	Rapid onset of action; useful as adjunct in severe hypercalcemia	Rapid tachyphylaxis
Phosphate oral	24 h	During use	Chronic management (with hypophosphatemia); low toxicity if P <4 mg/dL	Limited use except as adjuvant or chronic therapy
Glucocorticoids	Days	Days, weeks	Oral therapy, antitumor agent	Active only in certain malignancies, vitamin D excess, and sarcoidosis; glucocorticoid side effects
Dialysis	Hours	During use and 24–48 h afterward	Useful in renal failure; onset of effect in hours; can immediately reverse life-threatening hypercalcemia	Complex procedure, reserved for extreme or special circumstances

Source: Data from JP Bilezikian et al: Evaluation and management of primary hyperparathyroidism: Summary statement and guidelines from the Fifth International Workshop. JBMR 37:2293, 2022.

where they are powerful inhibitors of bone resorption. These bone-seeking compounds are stable in vivo because phosphatase enzymes cannot hydrolyze the central carbon-phosphorus-carbon bond. The bisphosphonates are concentrated in areas of high bone turnover and are taken up by and inhibit osteoclast action; the mechanism of action is complex. The bisphosphonate molecules that contain amino groups in the side chain structure (see below) interfere with prenylation of proteins and can lead to cellular apoptosis. The highly active non-amino-group–containing bisphosphonates are also metabolized to cytotoxic products.

A number of second- or third-generation compounds have become the mainstays of antiresorptive therapy for treatment of hypercalcemia and osteoporosis. The newer bisphosphonates have a highly favorable ratio of blocking resorption versus inhibiting bone formation; they inhibit osteoclast-mediated skeletal resorption yet do not cause mineralization defects at ordinary doses. Though the bisphosphonates have similar structures, the routes of administration, efficacy, toxicity, and side effects vary. Compounds commonly used are pamidronate, alendronate, and zoledronate. The IV use of pamidronate and zoledronate is approved for the treatment of hypercalcemia; between 30 and 90 mg pamidronate, given as a single IV dose over a few hours, returns serum calcium to normal within 24–48 h with an effect that lasts for weeks in 80–100% of patients. Zoledronate given as an infusion in doses of 5 mg has a more rapid and more sustained effect that lasts longer than pamidronate in direct comparison.

These drugs are used extensively in cancer patients. Absolute survival improvements are noted with pamidronate and zoledronate in multiple myeloma, for example. However, though rare, osteonecrosis of the jaw, especially after dental surgery, mainly in cancer patients treated with multiple doses of the more potent bisphosphonates, and atypical femoral fractures are potential side effects.

DENOSUMAB

Denosumab is the most recent antiresorptive therapy to be approved for the treatment of hypercalcemia, a monoclonal antibody that binds to RANK ligand (RANKL) and prevents it from binding to the receptor RANK on osteoclast precursors and mature osteoclasts. The inhibition of differentiation, activation, and function of osteoclasts leads to a reduction in bone resorption. It has a profound suppressive effect on biochemical markers of bone resorption and is the most powerful antiresorptive agent currently available. Repeated doses of denosumab, 120 mg given subcutaneously, may be effective in patients with hypercalcemia of malignancy who are not controlled by bisphosphonates. There are currently uncertainties how to manage rebound effects (fractures, hypercalcemia) after stopping denosumab, but antiresorptives such as alendronate and zoledronic acid can mitigate these effects.

OTHER THERAPIES

Calcitonin acts within a few hours of its administration, principally through receptors on osteoclasts, to block bone resorption. *Calcitonin*, after 24–48 h of use, is no longer effective in lowering calcium. Tachyphylaxis, a known phenomenon with this drug, seems to explain the results since the drug is initially often effective. Therefore, in life-threatening hypercalcemia, calcitonin can be used effectively within the first 24–48 h in combination with rehydration and saline diuresis while waiting for more sustained effects from simultaneously administered antiresorptives. Usual doses of calcitonin are 2–8 U/kg of body weight IV, SC, or IM every 6–12 h.

Glucocorticoids have utility, especially in hypercalcemia complicating certain malignancies. They increase urinary calcium excretion and decrease intestinal calcium absorption when given in pharmacologic doses, but they also cause negative skeletal calcium balance. In normal individuals and in patients with primary hyperparathyroidism, glucocorticoids neither increase nor decrease the serum calcium concentration. In patients with hypercalcemia due to certain osteolytic malignancies, however, glucocorticoids may be effective as a result of antitumor effects. The malignancies in which hypercalcemia responds to glucocorticoids include multiple myeloma, leukemia, Hodgkin's disease, other lymphomas, and carcinoma of the breast, at least early in the course of the disease. Glucocorticoids are also effective in treating hypercalcemia due to vitamin D intoxication and sarcoidosis. Glucocorticoids are also useful in the rare form of hypercalcemia, now recognized in certain autoimmune disorders in which inactivating antibodies against the receptor imitate FHH. Elevated PTH and calcium levels are effectively lowered by the glucocorticoids. In all the preceding situations, the hypocalcemic effect develops over several days, and the usual glucocorticoid dosage is 40–100 mg prednisone (or its equivalent) daily in divided doses. The side effects of chronic glucocorticoid therapy may be acceptable in some circumstances.

For patients with hypercalcemia due to CYP24A1 mutations, avoidance of sun, vitamin D intake, and reducing calcium intake are recommended. Fluconazole and ketoconazole, inhibitors of CYP27B1, and rifampin (an inducer of CYP3A4) have been reported to be beneficial; particular attention should be given to potential side effects of each of these medications.

Dialysis is often the treatment of choice for severe hypercalcemia complicated by renal failure, which is difficult to manage medically. Peritoneal dialysis with calcium-free dialysis fluid can remove 5–12.5 mmol (200–500 mg) of calcium in 24–48 h and lower the serum calcium concentration by 0.7–2.2 mmol/L (3–9 mg/dL). Large quantities of phosphate are lost during dialysis, and serum inorganic phosphate concentration usually falls, potentially aggravating hypercalcemia. Therefore, the serum inorganic phosphate concentration should be measured after dialysis, and phosphate supplements should be added to the diet or to dialysis fluids if necessary.

Phosphate therapy, PO or IV, has a limited role in certain circumstances **(Chap. 421)**. Correcting hypophosphatemia lowers the serum calcium concentration by several mechanisms, including bone/calcium exchange. The usual oral treatment is 1–1.5 g phosphorus per day for several days, given in divided doses. It is generally believed, but not established, that toxicity does not occur if therapy is limited to restoring serum inorganic phosphate concentrations to normal.

Raising the serum inorganic phosphate concentration above normal decreases serum calcium levels, sometimes strikingly. Intravenous phosphate is one of the most dramatically effective treatments available for severe hypercalcemia but is toxic and even dangerous (fatal hypocalcemia). For these reasons, it is used rarely and only in severely hypercalcemic patients with cardiac or renal failure where dialysis, the preferable alternative, is not feasible or is unavailable.

SUMMARY

The various therapies for hypercalcemia are listed in Table 422-4. The choice depends on the underlying disease, the severity of the hypercalcemia, the serum inorganic phosphate level, and the renal, hepatic, and bone marrow function. Mild hypercalcemia (≤3 mmol/L [12 mg/dL]) can usually be managed by hydration. Severe hypercalcemia (≥3.7 mmol/L [15 mg/dL]) requires rapid correction. IV pamidronate or zoledronate or subcutaneous denosumab should be administered. In addition, for the first 24–48 h, aggressive sodium-calcium diuresis with IV saline should be given. Following rehydration, furosemide or ethacrynic acid can be added, but only if appropriate monitoring is available and cardiac and renal function are adequate. For intermediate degrees of hypercalcemia between 3 and 3.7 mmol/L (12 and 15 mg/dL), vigorous hydration is recommended. Depending on symptoms and underlying causes, this may be combined with antiresorptives and other previously mentioned treatments.

■ HYPOCALCEMIA

(See also Chap. 57).

Pathophysiology *Chronic hypocalcemia* is less common than hypercalcemia; causes include CKD, hereditary and acquired hypoparathyroidism, vitamin D deficiency, PTH resistance, and hypomagnesemia.

Acute rather than chronic hypocalcemia is seen in critically ill patients or as a consequence of certain medications and often does not require specific treatment. Transient hypocalcemia is seen with severe sepsis, burns, acute kidney injury, and extensive transfusions with citrated blood. Although as many as one-half of patients in an intensive care setting are reported to have calcium concentrations of <2.1 mmol/L (8.5 mg/dL), most do not have a reduction in ionized calcium. Patients with severe sepsis may have a decrease in ionized calcium (true hypocalcemia), but in other severely ill individuals, hypoalbuminemia is the primary cause of the reduced total calcium concentration. Alkalosis increases calcium binding to proteins.

Medications such as protamine, heparin, and glucagon may cause transient hypocalcemia. These forms of hypocalcemia are usually not associated with tetany and resolve with improvement in the overall medical condition. The hypocalcemia after repeated transfusions of citrated blood usually resolves quickly.

Patients with *acute pancreatitis* have hypocalcemia that persists during the acute inflammation and varies in degree with disease severity. The cause of hypocalcemia remains unclear but may include saponification and systemic inflammation. PTH values are reported to be low, normal, or elevated, and both resistance to PTH and impaired PTH secretion have been postulated.

Chronic hypocalcemia, however, is usually symptomatic and requires treatment. Neuromuscular and neurologic manifestations of chronic hypocalcemia include tingling, muscle spasms, carpopedal spasm, and, in extreme cases, laryngeal spasm and convulsions. Increased intracranial pressure occurs in some patients with longstanding hypocalcemia, often in association with papilledema. Mental changes include irritability, depression, and psychosis. The QT interval on the electrocardiogram is prolonged, in contrast to its shortening with hypercalcemia. Arrhythmias occur, and digitalis effectiveness may be reduced. Intestinal cramps and chronic malabsorption may occur. Chvostek's or Trousseau's sign can be used to confirm latent tetany; the latter is more specific and sensitive for hypocalcemia.

Classification of Hypocalcemia The classification of hypocalcemia shown in **Table 422-5** is based on an organizationally useful premise that hypocalcemia may be primarily due to one of the two main calcium-regulating hormones, PTH and vitamin D, or other causes.

PTH-related Hereditary or acquired forms of hypoparathyroidism have a number of common components. The disease is rare with estimates from all causes to be ~25–35 patients/100,000 of the population (based on U.S. and Danish estimates). Symptoms of untreated hypocalcemia are shared by both types of hypoparathyroidism, although the onset of hereditary hypoparathyroidism can be more gradual and associated with other developmental defects. Basal ganglia calcification and extrapyramidal syndromes are more common and earlier in onset in hereditary hypoparathyroidism. Acquired hypoparathyroidism secondary to surgery in the neck is the most common cause of hypoparathyroidism, but the frequency of surgically induced parathyroid failure has diminished as a result of improved surgical techniques that spare the parathyroid glands and increased use of nonsurgical therapy for hyperthyroidism. PHP, an example of resistance to PTH action rather than a failure of production by the parathyroid gland, may share several features with hypoparathyroidism, including extraosseous calcification and extrapyramidal manifestations such as choreoathetotic movements and dystonia but circulating PTH is increased rather than decreased in the other conditions.

Papilledema, raised intracranial pressure, and lenticular cataracts may occur in both hereditary and acquired hypoparathyroidism, as do chronic changes in fingernails and hair, the latter usually reversible with treatment of hypocalcemia. Certain skin manifestations, including alopecia and candidiasis, are characteristic of hereditary hypoparathyroidism associated with autoimmune polyglandular failure **(Chap. 401)**.

Hypocalcemia associated with hypomagnesemia is associated with both deficient PTH release and impaired responsiveness to the hormone and is reversible with normalization of serum magnesium. Patients with hypocalcemia secondary to hypomagnesemia have low levels of circulating PTH, indicative of diminished hormone release despite a maximum physiologic stimulus by hypocalcemia. Hypermagnesemia, which can be iatrogenic, may inhibit PTH release leading to hypocalcemia.

TABLE 422-5 Functional Classification of Hypocalcemia (Excluding Neonatal Conditions)

1. **PTH-Related**
    - a. Absence of parathyroid glands or inactive PTH
        - i. Congenital: 22q11 deletion syndrome (DS), isolated hypoparathyroidism, PTH mutations, mutations in specific transcription factors (GCM2, GATA3)
        - ii. Destruction of glands: postsurgical, APECED, infiltrative disorders
    - b. Impaired secretion
        - i. Congenital: autosomal-dominant hypocalcemia
        - ii. Functional: hypomagnesemia, hypermagnesemia
    - c. Target organ resistance
        - i. Pseudohypoparathyroidism
        - ii. Hypomagnesemia
2. **Vitamin D–related**
    - a. Vitamin D deficiency: nutritional deficiency, impaired cutaneous production, malabsorption
    - b. Accelerated loss: impaired enterohepatic recirculation; increased metabolism due to anticonvulsants or antituberculosis therapy (e.g., rifampin)
    - c. Impaired 25-hydroxylation: severe liver disease, *CYP2R1* mutations
    - d. Impaired 1α-hydroxylation: renal insufficiency, azole antifungal medications that inhibit CYP27B1 (e.g., ketoconazole), genetic 1α hydroxylase deficiency, FGF23-related (TIO, XLH, CKD)
    - e. Target organ resistance: VDR mutations
3. **Others**
    - a. Impaired bone resorption: denosumab, antiresorptives
    - b. Excessive deposition into the skeleton: hungry bone syndrome (e.g., after parathyroidectomy for primary hyperparathyroidism), osteoblastic malignancies
    - c. Chelation: infusion of citrated blood products or EDTA, phosphate infusion
    - d. Critical illness: pancreatitis, ICU patients

Abbreviations: 22q11DS, 22q11 deletion syndrome; APECED, autoimmune polyendocrinopathy candidiasis ectodermal dystrophy; CKD, chronic kidney disease; EDTA, ethylenediaminetetraacetic acid; FGF23, fibroblast growth factor 23; ICU, intensive care unit; PTH, parathyroid hormone; TIO, tumor-induced osteomalacia; VDR, vitamin D receptor; XLH, X-linked hypophosphatemic rickets.

GENETIC CAUSES Hereditary hypoparathyroidism can occur as an isolated entity without other endocrine or dermatologic manifestations or in association with other abnormalities.

Hypoparathyroidism Associated with Other Abnormalities Hypoparathyroidism associated with defective development of both the thymus and the parathyroid glands is termed *DiGeorge syndrome*, *velocardiofacial syndrome, or 22q11 deletion syndrome*. Congenital cardiovascular, facial, and other developmental defects are present, and patients may die in early childhood with severe infections, hypocalcemia and seizures, or cardiovascular complications. Patients can survive into adulthood, and milder, incomplete forms may become manifest in childhood or adolescence. Most cases are sporadic, but autosomal dominant forms involving microdeletions of chromosome 22q11.2 or point mutations in the transcription factor TBX1 within that chromosomal region exist. Another autosomal dominant developmental defect with hypoparathyroidism, deafness, and renal dysplasia (HDR) is caused by mutations in the transcription factor GATA3 (chromosome 10p14), which is important in embryonic development and is expressed in developing kidney, ear structures, and the parathyroids. Autosomal recessive disorders comprising hypoparathyroidism include *Kenney-Caffey syndrome type 1*, which also features short stature, osteosclerosis, and thick cortical bones, and the related *Sanjad-Sakati syndrome*, which also exhibits growth failure and other dysmorphic features. Both syndromes involve mutations in a chaperone protein called *TBCE* (chromosome 1q42-q43), which is relevant to tubulin function. FAM111A defects (chromosome 11q12.1) were identified as the cause of *Kenney-Caffey syndrome type 2*.

Hypoparathyroidism that can occur in association with a monogenetic autoimmune syndrome involving failure of the adrenals, the ovaries, the immune system, and the parathyroids in association with recurrent mucocutaneous candidiasis, alopecia, vitiligo, and pernicious anemia is commonly referred to as *polyglandular autoimmune type 1 deficiency* or *autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED)* **(Chap. 401)**. This disorder is caused by mutations in the *AIRE* gene (chromosome 21q22.3). A stop codon mutation occurs in many Finnish families with the disorder, while another mutation (Y85C) is typically observed in Jews of Iraqi and Iranian descent.

Hypoparathyroidism is also seen in two disorders associated with mitochondrial dysfunction and myopathy, one termed *Kearns-Sayre syndrome* (KSS), with ophthalmoplegia and pigmentary retinopathy, and the other termed *MELAS syndrome* (*m*itochondrial *e*ncephalopathy, *l*actic *a*cidosis, and *s*troke-like episodes). Mutations or deletions in mitochondrial genes have been identified.

Isolated Hypoparathyroidism Several forms of hypoparathyroidism, each rare in frequency, are seen as isolated defects; the genetic mechanisms are varied. The inheritance includes autosomal dominant, autosomal recessive, and X-linked modes.

PTH Mutations Several autosomal defects involving the preproPTH sequence or the mature PTH have been recognized. The dominant forms are caused by point mutations in a critical region involved in intracellular transport of the hormone precursor. For example, an Arg for Cys mutation interferes with processing of the precursor and is believed to trigger an apoptotic cellular response, hence acting as a dominant negative. Recessive forms require both PTH alleles encoding the prepro sequence to be mutated. Only three homozygous mutations affecting the mature PTH have been described that lead to an autosomal recessive form of hypoparathyroidism. The defect for an X-linked recessive form of hypoparathyroidism has been localized to chromosome Xq26-q27, perhaps involving the *SOX3* gene.

CaSR Mutations Different gain-of-function mutations in the CaSR gene have been found in one form of hypocalcemia termed *autosomal dominant hypocalcemia (ADH) type 1*. The mutant receptor senses the ambient calcium level as excessive and suppresses PTH secretion, leading to hypocalcemia. The hypocalcemia is aggravated by constitutive receptor activity in the renal tubule causing excretion of inappropriate amounts of calcium. Recognition of the syndrome is important because efforts to treat the hypocalcemia with vitamin D analogues and increased oral calcium exacerbate the already excessive urinary calcium excretion leading to irreversible renal damage from stones and ectopic calcification. The orally available negative allosteric modulator (NAM) on the CaSR encaleret has been shown to normalize serum and urine calcium, as well as serum phosphate and magnesium in a phase 2 trial in patients with ADH1.

Other Causes of Isolated Hypoparathyroidism These include homozygous, inactivating mutations in the parathyroid-specific transcription factor GCM2 or heterozygous point mutations in this protein, which have a dominant-negative effect on the wild-type protein and thus lead to an autosomal dominant form of hypoparathyroidism. Furthermore, heterozygous mutations in Gα11, one of the two signaling proteins downstream of the CaSR, have been identified as a cause of autosomal dominant hypocalcemia, now referred to as ADH type 2. The *Bartter syndrome* is a group of disorders associated with disturbances in electrolyte and acid-base balance, sometimes with nephrocalcinosis and other features. Several types of ion channels or transporters are involved. Curiously, *Bartter syndrome type V* has electrolyte and pH disturbances but is caused by a gain-of-function mutation in the CaSR. The defect may be more severe than in ADH1 and explains the additional features seen beyond hypocalcemia and hypercalciuria. As with autoimmune disorders that block the CaSR (discussed above under hypercalcemic conditions), there are autoantibodies that at least transiently activate the CaSR, leading to suppressed PTH secretion and hypocalcemia.

Acquired chronic hypoparathyroidism is usually the result of inadvertent surgical removal of or damage to all the parathyroid glands; in some instances, not all the tissue is removed, but the remainder undergoes vascular supply compromise secondary to fibrotic changes in the neck after surgery. In the past, the most frequent cause of acquired hypoparathyroidism was surgery for hyperthyroidism. Hypoparathyroidism can also occur after surgery for hyperparathyroidism when the surgeon, facing the dilemma of removing too little tissue and thus not curing the hyperparathyroidism, removes too much. Parathyroid function may not be totally absent in all patients with postoperative hypoparathyroidism.

Very rare causes of acquired chronic hypoparathyroidism include radiation-induced damage subsequent to radioiodine therapy of hyperthyroidism and glandular damage in patients with hemochromatosis or hemosiderosis after repeated blood transfusions.

Transient hypoparathyroidism is frequent following surgery for hyperparathyroidism. After a variable period of hypoparathyroidism, normal parathyroid function may return due to hyperplasia or recovery of remaining tissue. Occasionally, recovery occurs months after surgery.

TREATMENT

Acquired and Hereditary Hypoparathyroidism

Conventional treatment has involved increasing serum calcium by administration of active vitamin D (calcitriol) combined with oral calcium supplementation. In many patients, blood calcium and phosphate levels are maintained satisfactorily, but some patients show a tendency to alternate between hypocalcemia and hypercalcemia, thus requiring close monitoring. Treatment with active vitamin D (calcitriol or alpha-calcidol) is preferred over high-dose plain vitamin D, which was standard in the past, particularly since calcitriol is cleared much more rapidly from the circulation than vitamin D. PTH analogues are being incorporated into the range of treatment options for this disease (see below).

Oral calcium and vitamin D increase serum calcium but do not address other functions of PTH. This current standard of care treatment exacerbates hypercalciuria, does not normalize hyperphosphatemia, and does not normalize the low bone turnover of hypoparathyroidism. Because of the hypercalciuria, blood calcium levels should be maintained at the lower end of the normal range or just below normal in these patients to avoid excessive urinary calcium excretion; otherwise, nephrocalcinosis and kidney stones can develop, and the risk of CKD is increased. Thiazide diuretics lower urine calcium by as much as 100 mg/d in hypoparathyroid patients on calcium and vitamin D, provided they are maintained on a low-sodium diet.

Until recently, hypoparathyroidism was the only endocrine disorder not being treated with the missing hormone. After the initial experimental use of PTH(1–34), the synthetic PTH fragment used in treatment of osteoporosis, showed promise, full-length PTH(1–84) has been shown to be effective and was approved by the U.S. Food and Drug Administration for therapy of hypoparathyroidism. The effective half-life of these PTH analogues is not long enough to achieve effective PTH effects over 24 h with one injection. Moreover, they are not approved or no longer available in the United States. Long-acting PTH molecules have been developed and are in clinical trials. Palopegteriparatide, an inactive prodrug that is given as an SC injection once daily and releases PTH(1–34) from an inert carrier in a sustained manner, showed normalization of blood and urine calcium in a phase 3 clinical trial and is approved in the US and Europe. Eneboparatide, a biased peptide agonist to the PTH receptor with a prolonged intracellular signaling, showed normalization of blood and urine calcium in patients with hypoparathyroidism in a phase 2 clinical trial and is currently undergoing a phase 3 trial.

HYPOMAGNESEMIA Severe hypomagnesemia (<0.4 mmol/L; <0.8 meq/L) is associated with hypocalcemia (see above). Restoration of the total-body magnesium deficit leads to rapid reversal of hypocalcemia. There are at least two causes of the hypocalcemia in severe hypomagnesemia—impaired PTH secretion and reduced responsiveness to PTH. **For further discussion of causes and treatment of hypomagnesemia, see Chap. 421.**

PTH levels are undetectable or inappropriately low in severe hypomagnesemia despite the stimulus of severe hypocalcemia, and acute repletion of magnesium leads to a rapid increase in PTH level. Serum phosphate levels are often not elevated, in contrast to the situation with acquired or idiopathic hypoparathyroidism, probably because phosphate deficiency is often seen in hypomagnesemia. In addition to diminished PTH secretion, some patients with low calcium and magnesium levels show a blunted peripheral response to exogenous PTH as documented by subnormal response in urinary phosphorus and urinary cyclic AMP excretion.

TREATMENT

Hypomagnesemia

Repletion of magnesium cures the condition. Repletion should be parenteral. Attention must be given to restoring the intracellular deficit, which may be considerable. After IV magnesium administration, serum magnesium may return transiently to the normal range, but unless replacement therapy is adequate, serum magnesium will again fall. If the cause of the hypomagnesemia is renal magnesium wasting, magnesium may have to be given long-term to prevent recurrence **(Chap. 421)**.

PTH TARGET ORGAN RESISTANCE PHP refers to a group of distinct inherited disorders that resemble hypoparathyroidism (in which PTH synthesis is deficient) and is manifested by hypocalcemia and hyperphosphatemia yet elevated PTH levels.

Patients affected by PHP type Ia (PHP1A) develop symptoms and signs of hypocalcemia in association with distinctive skeletal and developmental defects, referred to as Albright's hereditary osteodystrophy (AHO). The hypocalcemia is due to a deficient PTH response in the proximal renal tubules, probably leading to insufficient $1,25(OH)_2D$ production and thus impaired intestinal calcium absorption. Furthermore, PTH resistance in this portion of the kidney impairs urinary phosphate excretion, thus leading to elevated serum phosphate levels. Patients affected by PHP type Ib (PHP1B) also present with hypocalcemia and hyperphosphatemia but less frequently with obvious AHO features. In response to the hypocalcemia observed in either disorder, PTH levels increase, leading to parathyroid hyperplasia and, in some cases, to autonomous PTH secretion. Studies, both clinical and basic, have clarified some aspects of these disorders, including the variable clinical spectrum, the pathophysiology, the genetic defects, and their mode of inheritance.

A working classification of the various PHP forms is given in **Table 422-6**. The classification scheme is based on the signs of ineffective PTH action (low calcium and high phosphate), low or normal urinary cyclic AMP response to exogenous PTH, the presence or absence of AHO, and assays to measure the concentration of the Gsα subunit. Using these criteria, there are four types: PHP types Ia and Ib (PHP1A and PHP1B); pseudopseudohypoparathyroidism (PPHP), and PHP type II (PHP2). Another classification has been proposed recently, which is being debated.

PHP1A and PHP1B Individuals with PHP type I (PHP1), the most common of the disorders, show deficient urinary cyclic AMP excretion in response to administration of exogenous PTH. Patients with PHP1 are divided into PHP1A and PHP1B. Most patients with PHP1A show evidence for AHO and reduced amounts of Gsα protein/activity, as previously determined in readily accessible tissues such as erythrocytes, lymphocytes, or fibroblasts. Only some PHP1B patients show typical AHO features, but they usually have normal Gsα activity. PHP1C, sometimes listed as a third form of PHP1, is really a variant of PHP1A, although the mutant Gsα shows normal activity in certain in vitro assays.

Most patients who have PHP1A reveal characteristic features of AHO, which consist of short stature, early-onset obesity, round face, obesity, skeletal anomalies (brachydactyly), intellectual impairment, and/or heterotopic calcifications. Patients have low calcium and high phosphate levels, as with true hypoparathyroidism. PTH levels, however, are elevated, reflecting resistance to hormone action. In addition, hormonal resistance is observed at other Gsα-coupled receptors, particularly at the TSH receptor, leading to elevated levels of this hormone.

Calcium and phosphate deposits are frequently found in the basal ganglia. The typical shortening of metacarpal and metatarsal bones is caused by premature closing of the epiphyses and is probably a particularly sensitive sign of overall advanced skeletal maturation resulting in adult short stature.

INHERITANCE AND GENETIC DEFECTS Multiple defects at the *GNAS* locus have now been identified in PHP1A, PHP1B, and PPHP patients. This gene, which is located on chromosome 20q13.3, encodes the α-subunit of the stimulatory G protein (Gsα), among other products (see below). Mutations involving the *GNAS* exons encoding Gsα, which are the cause of PHP1A and PPHP, include abnormalities at splice junctions, point mutations, insertions, and/or deletions that all result in a Gsα protein with defective function, resulting in a 50% reduction of in vitro Gsα activity in erythrocytes or other cells. While PHP1A is caused by inactivating Gsα mutations on the maternal allele, PPHP is caused by the same or similar mutations on the paternal *GNAS* allele **(Fig. 422-7)**. The Gsα transcript is biallelically expressed in most tissues; however, expression from paternal allele is silenced through as-of-yet-unknown mechanisms in some tissues, including proximal renal tubules, thyroid, and pituitary. Consequently, inheritance of a molecular defect involving the paternal exons encoding Gsα has no implications with regard to hormone function, while inactivating Gsα mutations involving the maternal *GNAS* allele lead to little or no Gsα protein in these tissues **(Chap. 479)**. Thus, females affected by either PHP1A or PPHP will have offspring with PHP1A, if these children inherit the allele carrying the *GNAS* mutation; in contrast, if the mutant allele is inherited from a male affected by either disorder, the offspring will exhibit PPHP. However, patients affected by both disease subtypes develop some but not

TABLE 422-6 Classification of Pseudohypoparathyroidism (PHP) and Pseudopseudohypoparathyroidism (PPHP)

TYPE	HYPOCALCEMIA, HYPERPHOSPHATEMIA	RESPONSE OF URINARY cAMP TO PTH	SERUM PTH	Gsα SUBUNIT DEFICIENCY	AHO	RESISTANCE TO HORMONES OTHER THAN PTH
PHP1A	Yes	↓	↑	Yes	Yes	Yes
PPHP	No	Normal	Normal	Yes	Yes	No
PHP1B	Yes	↓	↑	No	Yes (less frequently and usually less severe)	Yes (in some patients)
PHP2	Yes	Normal (but ↓ phosphaturic response)	↑	No	No	No
Acrodysostosis due to *PRKAR1A* mutations with hormonal resistance	Yes	Normal (but ↓ phosphaturic response)	↑	No	Yes	Yes

Abbreviations: ↓, decreased; ↑, increased; AHO, Albright's hereditary osteodystrophy; cAMP, cyclic adenosine monophosphate; PTH, parathyroid hormone.

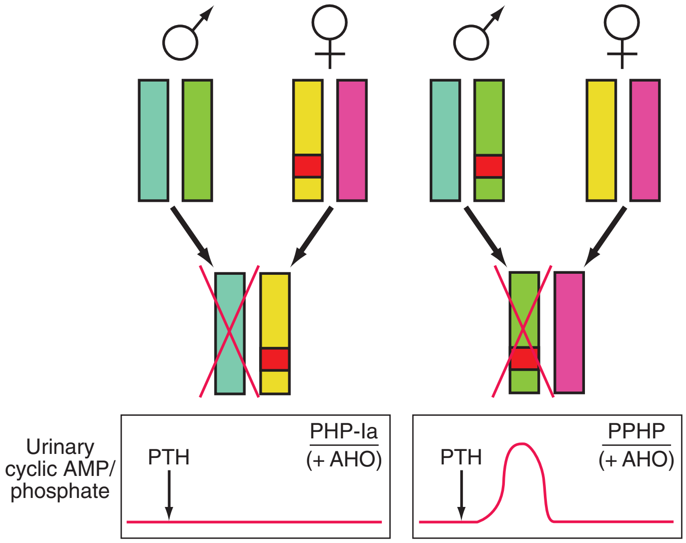

FIGURE 422-7 Paternal imprinting of renal parathyroid hormone (PTH) resistance (*GNAS* gene for Gsα subunit) in pseudohypoparathyroidism (PHP1A and PHP1B). An impaired excretion of urinary cyclic AMP and phosphate is observed in patients with PHP type I. In the renal cortex, there is selective silencing of paternal Gsα expression; consequently, mutations involving the maternal *GNAS* exons encoding Gsα or loss of methylation at *GNAS* exon A/B leads to reduced or completely absent Gsα protein in this portion of the kidney. The disease becomes manifest only in patients who inherit the defective gene from an obligate female carrier (*left*). If a genetic defect involving *GNAS* exons encoding Gsα is inherited from an obligate male carrier of the mutation (PHP1A or PPHP patient), no biochemical abnormality is encountered, and the administration of PTH causes an appropriate increase in the urinary cyclic AMP and phosphate concentration (pseudoPHP [PPHP]; *right*). Both patterns of inheritance lead to some but not all features of Albright's hereditary osteodystrophy (AHO), most likely because of haploinsufficiency; for example, Gsα protein derived from both parental *GNAS* alleles must be active for normal bone development. Maternal inheritance of a mutation (deletion, duplication, or inversion within or upstream of the *GNAS* locus) causes AD-PHP1B, while paternal inheritance does not lead to any detectable abnormality.

all AHO features, making it likely that Gsα haploinsufficiency becomes apparent during embryonic or postnatal development.

The complex mechanisms that control the *GNAS* gene contributed particularly to challenges involved in unraveling the pathogenesis of PHP1B. Analysis of families in which multiple members are affected by PHP1B, as well as studies of the complex parent-specific methylation of four regions within the complex *GNAS* locus, revealed that the autosomal dominant forms of PHP1B (AD-PHP1B) are caused by microdeletions, duplications, retrotransposon insertions, or inversions within or upstream of the *GNAS* locus. These genetic mutations are associated with a loss of DNA methylation at one or several loci on the maternal *GNAS* allele (Table 422-6). These abnormalities in methylation silence maternal Gsα expression, thus leading in the proximal renal tubules—where Gsα appears to be expressed predominantly from the maternal allele—to PTH resistance. While most cases of AD-PHP1B are by now resolved at the molecular level, the genetic defect responsible for the sporadic variant of PHP1B (sporPHP1B), the most frequent form of PHP1B, remains to be defined, except for those sporPHP1B cases that are caused by paternal uniparental isodisomy/heterodisomy of chromosome 20q (patUPD20q).

PHP1B patients, who rarely develop an AHO phenotype as severe as in PHP1A, develop hypocalcemia and hyperphosphatemia caused by PTH resistance and thus elevated PTH levels. The previously used Ellsworth-Howard test to assess the presence or absence of hormone resistance is used much less frequently, largely because of routinely available sensitive PTH assays (Table 422-6). As for PHP1A, these endocrine abnormalities become apparent only if disease-causing mutations are inherited maternally. Bone responsiveness may be excessive rather than blunted in PHP1B (and in PHP1A) patients, based on case reports that have emphasized an osteitis fibrosa–like pattern in several PHP1B patients. Some patients present with PTH resistance in the absence of AHO features and without *GNAS* methylation changes; it remains unclear why this PHP variant readily resolves upon treatment with vitamin D supplements.

PHP2 refers to patients with hypocalcemia and hyperphosphatemia, who have normal urinary cyclic AMP excretion, but an impaired urinary phosphaturic response to PTH. In one PHP2 variant, referred to as acrodysostosis with hormonal resistance, patients have a heterozygous defect in the regulatory subunit of PKA (PRKAR1A) that mediates the response to PTH distal to cyclic AMP production. Acrodysostosis without or with only mild hormonal resistance can be caused by heterozygous mutations in the cyclic AMP–selective phosphodiesterase 4D. Patients with one variant of acrodysostosis that is associated with hypertension, were shown to have heterozygous phosphodiesterase 3A mutations.

The diagnosis of these hormone-resistant states can usually be made when there is a positive family history for signs and symptoms of hypocalcemia with or without AHO features. In both categories—PHP1A and PHP1B—serum PTH levels are elevated, particularly when patients start to experience hypocalcemia during childhood. In PHP1A and PHP1B, the response of urinary cyclic AMP to the administration of exogenous PTH is blunted. The diagnosis of PHP2, in the absence of acrodysostosis, is more complex, and vitamin D deficiency must be excluded before such a diagnosis can be entertained.

TREATMENT

Pseudohypoparathyroidism

Treatment of PHP is similar to that of hypoparathyroidism, except that calcium and activated vitamin D analogues are usually given at higher doses to maintain blood calcium levels within the normal range and PTH levels in the upper end of normal or slightly elevated. Patients with PHP1 show no PTH resistance in the distal tubules—hence, urinary calcium clearance is typically not elevated, and these individuals are not at risk of developing nephrocalcinosis, as are patients with hypoparathyroidism, unless overtreatment occurs, for example, after the completion of pubertal development and skeletal mutation, when calcium and 1,25(OH)$_2$D treatment should be reduced. Variability in response makes it necessary to establish the optimal regimen for each patient.

Vitamin D Related • **VITAMIN D DEFICIENCY DUE TO INADEQUATE DIET AND/OR SUNLIGHT** Vitamin D deficiency due to inadequate intake of dairy products enriched with vitamin D, lack of vitamin supplementation, and reduced sunlight exposure in the elderly, particularly during winter in northern latitudes, is more common in the United States than previously recognized. Biopsies of bone in elderly patients with hip fracture (documenting osteomalacia) and abnormal levels of vitamin D metabolites, PTH, calcium, and phosphate indicate that vitamin D deficiency may occur in as many as 25% of elderly patients, particularly in northern latitudes in the United States. Concentrations of 25(OH)D are low or low-normal in these patients. Quantitative histomorphometric analysis of bone biopsy specimens from such individuals reveals widened osteoid seams consistent with osteomalacia **(Chap. 421)**. PTH hypersecretion compensates for the tendency for the blood calcium to fall but also increases renal phosphate excretion and thus causes osteomalacia.

Treatment involves adequate replacement with vitamin D and calcium until the deficiencies are corrected. Severe hypocalcemia rarely occurs in moderately severe vitamin D deficiency of the elderly, but vitamin D deficiency must be considered in the differential diagnosis of mild hypocalcemia.

Mild hypocalcemia, secondary hyperparathyroidism, severe hypophosphatemia, and a variety of nutritional deficiencies occur with gastrointestinal diseases. Hepatocellular dysfunction can lead to reduction in 25(OH)D levels, as in portal or biliary cirrhosis of the liver, and malabsorption of vitamin D and its metabolites, including 1,25(OH)$_2$D, may occur in a variety of bowel diseases, hereditary or acquired. Depending on the disorder, vitamin D or its metabolites can be given parenterally, guaranteeing adequate blood levels of active metabolites.

DEFECTIVE VITAMIN D METABOLISM • ***Anticonvulsant Therapy*** Anticonvulsant therapy with any of several agents induces acquired vitamin D deficiency by increasing the conversion of vitamin D to inactive compounds and/or causing resistance to its action. The more marginal the vitamin D intake in the diet, the more likely that anticonvulsant therapy will lead to abnormal mineral and bone metabolism.

Vitamin D–Dependent Rickets Type I Vitamin D–dependent rickets type I, previously termed *pseudo-vitamin D–resistant rickets*, is caused by homozygous or compound heterozygous mutations in the gene CYP27B1 encoding 25(OH)D-1α-hydroxylase. It differs from true vitamin D–resistant rickets (vitamin D–dependent rickets type II, see below) in that it is typically less severe and the biochemical and radiographic abnormalities can be readily reversed with physiologic doses of the vitamin's active metabolite, $1,25(OH)_2D$ **(Chap. 421)**. Clinical features include hypocalcemia, often with tetany or convulsions; hypophosphatemia due to secondary hyperparathyroidism; and thus, osteomalacia and increased levels of alkaline phosphatase.

Vitamin D–Dependent Rickets Type II Vitamin D–dependent rickets type II results from end-organ resistance to the active metabolite $1,25(OH)_2D$. The clinical features resemble those of the type I disorder and include hypocalcemia, hypophosphatemia, secondary hyperparathyroidism, and rickets but also partial or total alopecia. Plasma levels of $1,25(OH)_2D$ are elevated, in keeping with the refractoriness of the end organs. This disorder is caused by homozygous or compound heterozygous mutations in the gene encoding the vitamin D receptor; treatment requires regular, usually nocturnal calcium infusions, which normalize PTH levels, thus reducing urinary phosphate excretion and thereby improving rickets and thus growth, but do not restore hair growth **(Chap. 421)**.

CKD Improved medical management of CKD allows many patients to survive for decades and, hence, provides time enough to develop features of renal osteodystrophy, which must be controlled to avoid additional morbidity. Impaired production of $1,25(OH)_2D$ is a principal factor that causes calcium deficiency, secondary hyperparathyroidism, and bone disease; hyperphosphatemia, which lowers further blood calcium levels, typically occurs only in the later stages of the disease. Low levels of $1,25(OH)_2D$ due to increased FGF23 production in bone (and possibly other tissues) are critical in the development of hypocalcemia. It is notable that FGF23 levels are often dramatically elevated in end-stage kidney disease (ESKD). The uremic state also causes impairment of intestinal absorption by mechanisms other than defects in vitamin D metabolism. Nonetheless, treatment with supraphysiologic amounts of vitamin D or calcitriol can correct impaired calcium absorption. Increased FGF23 levels are seen already during the early CKD stages and have been reported to correlate with kidney disease progression, increased mortality, and left ventricular hypertrophy. Strategies involving different oral phosphate binders have therefore been pursued to lower intestinal phosphate absorption early during the course of kidney disease and to thereby lower FGF23 levels. However, these approaches have been largely disappointing. Furthermore, there is concern as to whether supplementation with activated vitamin D analogues increases further the circulating FGF23 levels and their "off-target" effects in CKD patients.

TREATMENT

Chronic Kidney Disease

Therapy of CKD **(Chap. 322)** involves appropriate management of patients prior to dialysis and adjustment of regimens once dialysis is initiated. Attention should be paid to restriction of phosphate in the diet; avoidance of aluminum-containing phosphate-binding antacids; provision of an adequate calcium intake by mouth, usually around 1 g/d; and supplementation with 0.25–1 μg/d calcitriol or other activated forms of vitamin D. The aim of therapy is to restore normal calcium balance to prevent osteomalacia and severe secondary hyperparathyroidism (it is usually recommended to maintain PTH levels between 100 and 300 pg/mL) and, in light of evidence of genetic changes and monoclonal outgrowths of parathyroid glands in CKD patients, to prevent secondary hyperparathyroidism from becoming autonomous hyperparathyroidism. Reduction of hyperphosphatemia and restoration of normal intestinal calcium absorption by calcitriol can improve blood calcium levels and reduce the manifestations of secondary hyperparathyroidism. Since adynamic bone disease can occur in association with low PTH levels, it is important to avoid excessive suppression of the parathyroid glands while recognizing the beneficial effects of controlling the secondary hyperparathyroidism. These patients should be closely monitored with intact PTH assays. Oral phosphate-binding agents such as sevelamer lower blood phosphate levels in ESKD, but their use in earlier CKD stages does not seem to be beneficial in lowering blood phosphate levels and to prevent the rise in FGF23.

Other Causes Treatment of patients using antiresorptives, such as bisphosphonates and denosumab, my result in hypocalcemia. Risk factors for this condition include vitamin D deficiency, low calcium intake, or advanced chronic kidney disease.

Increased enteric loss of 25-hydroxyvitamin D and 1,25(OH)2D due to intestinal disease and increased metabolism due to anticonvulsants or antituberculosis therapy can also lead to hypocalcemia.

Impaired 25-hydroxylation due to severe liver disease or mutations in CYP2R1, the 25-hydroxylase, are rare causes of hypocalcemia. Impaired 1α-hydroxylation is not uncommon and occurs in renal insufficiency, certain antifungals, and FGF23-related disorders.

Occasionally, loss of calcium from the ECF is so severe that PTH cannot compensate. Such situations include acute pancreatitis and severe, acute hyperphosphatemia, often in association with renal failure, conditions in which there is rapid efflux of calcium from the ECF. Severe hypocalcemia can occur quickly; PTH rises in response to hypocalcemia but does not return blood calcium to normal.

SEVERE, ACUTE HYPERPHOSPHATEMIA Severe hyperphosphatemia is associated with extensive tissue damage or cell destruction **(Chap. 421)**. The combination of increased release of phosphate from muscle and impaired ability to excrete phosphorus because of renal failure causes moderate to severe hyperphosphatemia, the latter causing calcium loss from the blood and mild to moderate hypocalcemia. Hypocalcemia is usually reversed with tissue repair and restoration of renal function as phosphorus and creatinine values return to normal. There may even be a mild hypercalcemic period in the oliguric phase of renal function recovery. This sequence, severe hypocalcemia followed by mild hypercalcemia, reflects widespread deposition of calcium in muscle and subsequent redistribution of some of the calcium to the ECF after phosphate levels return to normal.

Other causes of hyperphosphatemia include hypothermia, massive hepatic failure, and hematologic malignancies, either because of high cell turnover of malignancy or because of cell destruction by chemotherapy.

TREATMENT

Severe, Acute Hyperphosphatemia

Treatment is directed toward lowering of blood phosphate by the administration of phosphate-binding antacids or dialysis. Although calcium replacement may be necessary if hypocalcemia is severe and symptomatic, calcium administration during the hyperphosphatemic period tends to increase extraosseous calcium deposition and aggravate tissue damage. The levels of $1,25(OH)_2D$ may be low during the hyperphosphatemic phase and return to normal during the oliguric phase of recovery.

OSTEITIS FIBROSA AFTER PARATHYROIDECTOMY Severe hypocalcemia after parathyroid surgery is rare now that osteitis fibrosa cystica is an infrequent manifestation of hyperparathyroidism. When osteitis fibrosa cystica is severe, however, bone mineral deficits can be large. After parathyroidectomy, hypocalcemia can persist for days if calcium replacement is inadequate. Treatment may require parenteral administration of calcium; addition of calcitriol and oral calcium supplementation is sometimes needed for weeks to a month or two until bone defects are filled (which, of course, is of therapeutic benefit in the skeleton), making it possible to discontinue parenteral calcium and/or reduce the amount.

Differential Diagnosis Care must be taken to ensure that true hypocalcemia is present; in addition, acute transient hypocalcemia can be a manifestation of a variety of severe, acute illnesses, as discussed above. *Chronic hypocalcemia*, however, can usually be ascribed to a few disorders associated with absent or ineffective PTH. Important clinical criteria include the duration of the illness, signs or symptoms of associated disorders, and the presence of features that suggest a hereditary abnormality. A nutritional history can be helpful in recognizing a low intake of vitamin D and calcium in the elderly, and a history of excessive alcohol intake may suggest magnesium deficiency.

Differential Diagnosis Inherited hypoparathyroidism and PHP are lifelong illnesses, usually (but not always) appearing by adolescence; hence, a recent onset of hypocalcemia in an adult is more likely due to nutritional deficiencies, CKD, or intestinal disorders that result in deficient or ineffective vitamin D. Neck surgery, even long past, however, can be associated with a delayed onset of postoperative hypoparathyroidism. A history of seizure disorder raises the issue of anticonvulsive medication. Developmental defects may point to the diagnosis of PHP1A. Rickets and a variety of neuromuscular syndromes and deformities may indicate ineffective vitamin D action, either due to defects in vitamin D metabolism or to vitamin D deficiency.

A pattern of *low calcium with high phosphorus* in the absence of renal failure or massive tissue destruction almost invariably means hypoparathyroidism or PHP. A *low calcium and low phosphorus* pattern points to absent or ineffective vitamin D, thereby impairing the action of PTH on calcium metabolism (but not phosphate clearance). The relative ineffectiveness of PTH in calcium homeostasis in vitamin D deficiency, anticonvulsant therapy, gastrointestinal disorders, and hereditary defects in vitamin D metabolism leads to secondary hyperparathyroidism as a compensation. The excess PTH on renal tubule phosphate transport accounts for renal phosphate wasting and hypophosphatemia.

Acknowledgment

The authors are grateful to John T. Potts, Jr., for his contributions to this chapter over many previous editions of Harrison's.

■ FURTHER READING

Bastepe M, Jüppner H: Pseudohypoparathyroidism, Albright's hereditary osteodystrophy, and progressive osseous heteroplasia: Disorders caused by inactivating GNAS mutations, in *Endocrinology*, 8th ed, in *Endocrinology*, JL Jameson, LJ DeGroot (eds). Philadelphia, W.B. Saunders Company, 2023; pp 974–991.

Bilezikian JP et al: Evaluation and management of primary hyperparathyroidism: Summary statement and guidelines from the Fifth International Workshop. J Bone Miner Res 37:2293, 2022.

Thakker RV et al: Regulation of calcium homeostasis and Genetic disorders that affect calcium metabolism, in *DeGroot's Endocrinology*, 8th ed, RP Robertson (eds). Philadelphia, Elsevier 2023.

423

Osteoporosis

Peter R. Ebeling

Osteoporosis, a condition characterized by decreased bone strength and fragility fractures, is most common among postmenopausal women, but 30% of fragility fractures occur in men. Other underlying diseases can result in secondary osteoporosis. The clinical manifestations of osteoporosis are primarily vertebral, nonvertebral, and hip fractures. Osteoporosis affects >10 million individuals in the United States, but only a minority are currently diagnosed and treated.

DEFINITION

Osteoporosis is defined as a reduction in the strength of bone that leads to skeletal fragility and fractures. Despite bone mineral density being used to define osteoporosis, other important factors such as microarchitectural deterioration and misalignment of bone components also contribute. Thus, relying solely on the measure of bone mineral density may underestimate bone fragility. The World Health Organization (WHO) operationally defined osteoporosis as a bone density that falls 2.5 standard deviations (SDs) or more below the mean for young healthy adults (age 30 years) of the same sex and race—also referred to as a *T-score* of –2.5. Postmenopausal women in the lower end of the young normal range (a T-score <–1.0 to –2.5) are defined as having low bone density or osteopenia, a risk factor for developing osteoporosis. Despite a lower fracture risk in this group, >50% of fractures among postmenopausal women, including hip fractures, occur in individuals with osteopenia because the size of that population is much larger than the group defined as having osteoporosis by bone density T-score. This has led to a greater emphasis on absolute fracture risk, incorporating age, sex, and other major clinical risk factors with or without bone mineral density (BMD) to calculate the 10-year risk of hip or major osteoporotic fractures. The calculation of absolute fracture risk with tools such as FRAX® or the Garvan Fracture Risk Calculator has allowed the development of intervention thresholds for osteoporosis treatment that may be country specific and may differ from diagnostic thresholds (e.g., T-score <–2.5).

Fragility fractures are defined as fractures in adults occurring following a fall from standing height or less, but exclude finger, toes, face, and skull fractures. However, recent studies also indicate traumatic fractures should also be regarded as indicative of underlying skeletal fragility and require further evaluation.

EPIDEMIOLOGY

In the United States, as many as 10.8 million women and 2.5 million men have osteoporosis (BMD T-score <–2.5 at lumbar spine, total hip, or femoral neck). This does not include additional people who present with an osteoporosis-related fracture but with osteopenia (T-score <–1.0 to –2.5). It is estimated that 2 million osteoporosis-related fractures occur each year in the United States at a cost of $19 billion, a problem that will increase as the population ages. Globally, hip fractures are increasing and are associated with high costs. The failure to identify the first fragility fracture and intervene is estimated to cost $6 billion to Medicare alone for secondary fractures. Another 40 million Americans have osteopenia that potentially puts them at increased of fracture and of developing osteoporosis. Although osteoporosis is mostly age-related, some individuals appear more at risk. In women, the rapid loss of ovarian function during the perimenopause (on average around age 50 years) precipitates rapid bone loss over the next 5–7 years such that most women will have osteoporosis by age 70–80 years. As the population is aging, the number of individuals with osteoporosis and fractures is rising. As many fragility fractures due to osteoporosis occur in individuals with osteopenia, identification of those individuals with a high absolute fracture risk and their evaluation and treatment are important.

Most fractures, especially those of the hip and vertebrae, show exponential increases with advancing age **(Fig. 423-1)**. Lifetime osteoporotic fracture risk for a Caucasian woman who reaches the age of 50 years is ~50%, while the corresponding risk for a 50-year-old man is ~25%. Recent data suggest that fractures, including hip fractures, are increasing despite age-related fracture rates decreasing. This may be related to the aging of the population globally or to failure to evaluate and treat patients with a high absolute fracture risk.

About 300,000 hip fractures occur each year in the United States, almost all requiring hospital admission and surgical intervention. The lifetime probability that a 50-year-old white individual will have a hip fracture is 14% for women and 5% for men; the risk for African

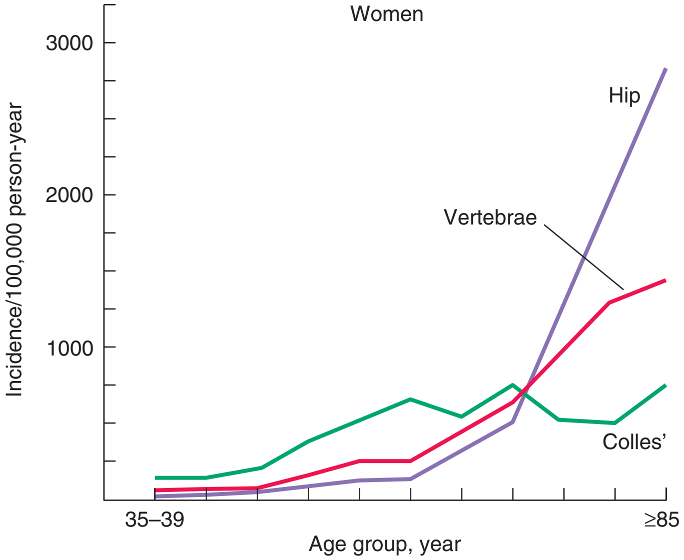

FIGURE 423-1 Epidemiology of vertebral, hip, and Colles' fractures with age. *(Reproduced with permission from C Cooper, LJ Melton 3rd: Epidemiology of osteoporosis. Trends Endocrinol Metab 3:224, 1992.)*

Americans is about half of those rates, and the risk for Asians and non-black Hispanics appears to be like that for Caucasians. Hip fractures are associated with a high incidence of mortality and morbidity, with 20–25% of patients dying in the year following the injury, with higher mortality rates among males and African Americans. Hip fracture rates and mortality also demonstrate high global variability. About 30% of survivors require long-term care (at least temporarily), and many never regain the independence that they had prior to the fracture. This sequela is the one most feared by patients.

There are ~500,000 symptomatic vertebral fractures per year in the United States, but >1,000,000 vertebral fractures may occur annually since only one-third are recognized clinically. The vast majority are clinically "silent" vertebral fractures identified incidentally during spinal radiography **(Fig. 423-2)** or may be suggested by significant height loss (>4 cm). However, even asymptomatic vertebral fractures are a major sign of skeletal fragility and increase the risk for subsequent fracture. Vertebral fractures, like other fragility fractures, are also associated with long-term morbidity and an increase in mortality. The occurrence of the first fracture greatly increases the risk of further fractures, especially in the first year. The consequence is height loss, kyphosis, and secondary pain and discomfort related to altered spinal biomechanics. Thoracic fractures can be associated with restrictive lung disease, whereas lumbar fractures are associated with abdominal symptoms including distention, early satiety, and constipation.

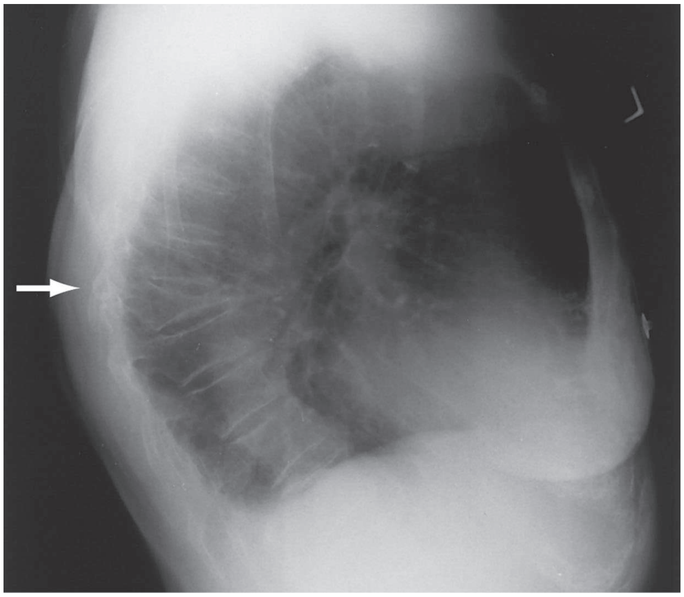

FIGURE 423-2 Lateral spine x-ray showing severe osteopenia and a severe wedge-type deformity (severe anterior compression).

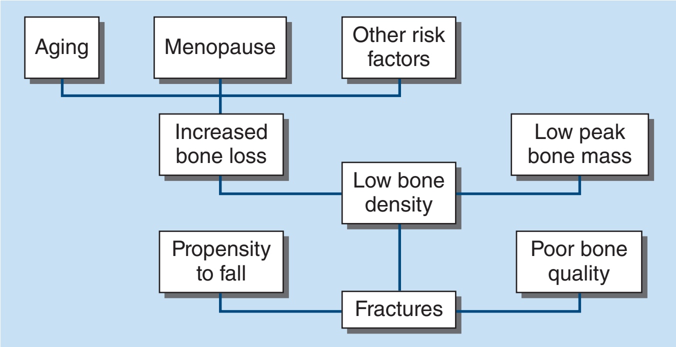

FIGURE 423-3 Factors leading to osteoporotic fractures.

Approximately 400,000 wrist fractures occur in the United States each year. Fractures of other bones (including ~150,000 pelvic fractures and >100,000 proximal humerus fractures) also occur due to osteoporosis. The threshold for fracture is reduced in osteoporotic bone **(Fig. 423-3)** due to increased skeletal fragility. Traumatic fractures are also increased in patients at risk of osteoporosis. Fewer than 20% of patients with a fracture are currently either investigated for osteoporosis or started on treatment within 6 months. A number of clinical risk factors for fracture exist; the common ones are summarized in **Table 423-1**. Prior fragility fractures, a family history of hip fracture, low body mass index, cigarette smoking, and excessive alcohol consumption are all independent predictors. Chronic inflammatory diseases, such as rheumatoid arthritis, increase the risk of osteoporosis, as do diseases associated with malabsorption (e.g., celiac disease) and male hypogonadism. Chronic diseases that increase the risk of falling or frailty, including dementia, Parkinson's disease, and multiple sclerosis, also increase fracture risk (Table 423-1). Many other risk factors for osteoporosis have been described including glucocorticoids, aromatase inhibitors, androgen deprivation therapy, air pollution, triclosan, bariatric surgery, diabetes mellitus, cerebrovascular accidents, dementia (including Alzheimer's), the death of a spouse, depression and its treatment with selective serotonin reuptake inhibitors (SSRIs), and proton pump inhibitors, to name a few.

Increasing frailty with age is a potent risk factor for fracture, as is sensory inattention (e.g., walking while looking at mobile phone).

Globally, fragility fractures are more common among women than men, presumably due to a lower peak bone mass as well as rapid post-menopausal bone loss in women. However, this sex difference in bone density and hip fracture incidence is not apparent in all countries, possibly due to genetics, physical activity levels, or diet.

Fragility fractures increase the risk for future fractures (Table 423-1). Vertebral fractures increase the risk of other vertebral fractures as well

TABLE 423-1 Risk Factors for Osteoporosis Fracture

NONMODIFIABLE	POTENTIALLY MODIFIABLE
Personal history of fracture as an adult	Current cigarette smoking
History of fracture in first-degree relative	Estrogen deficiency
Female gender	Early menopause (<45 years) or bilateral ovariectomy
Advanced age	Prolonged premenstrual amenorrhea (>1 year)
White race	Poor nutrition especially low calcium and vitamin D intake
Dementia	Alcoholism
	Impaired eyesight despite adequate correction
	Recurrent falls
	Inadequate physical activity
	Poor health/frailty

as fractures of the peripheral skeleton such as the hip and wrist. Wrist fractures also increase the risk of vertebral and hip fractures. Among individuals aged >50 years, any fracture (except those of the fingers, toes, face, and skull) should be considered as fragility fractures. Any fracture in a woman aged >50 years or a man aged >60 years should trigger investigations for osteoporosis. However, this does not occur in the majority as postfracture care is fragmented. Recent attempts to coordinate care using fracture liaison services to guide patients with fragility fractures through health care systems and ensure their investigation and initiate treatment for osteoporosis have been shown to improve outcomes but may be difficult to implement in countries without single-payor systems or closed health care systems.

The risk for future fracture after a first fracture is exponentially increased in the first 12–24 months, leading to the concept of imminent fracture risk. A recent large Medicare database study indicated that almost 20% of women will have a second fracture within 2 years after the first. Risk diminishes to less than half of that rate in the subsequent 3 years but remains persistently elevated after a vertebral or hip fracture.

PATHOPHYSIOLOGY

BONE REMODELING

A low peak bone mass may underlie the development of osteoporosis due to hormonal, genetic, or nutritional influences. Age-related changes in bone remodeling as well as extrinsic and intrinsic factors may then be superimposed. Consequently, an appreciation of bone remodeling is fundamental to developing an understanding of both the pathophysiology of bone loss **(Chap. 421)** and mechanisms of pharmacologic intervention. During growth, the skeleton increases in size by linear growth and by apposition of new bone tissue on the cortical periosteum **(Fig. 423-4)**. The latter process is called *modeling*, a process that also allows the long bones to adapt in shape to the stresses placed on them. Increased sex hormone production at puberty is required for skeletal maturation, with peak bone mass being achieved by early adulthood. Recent data suggest delayed puberty may be associated with low bone peak mass that persists into adulthood in both sexes. Sexual dimorphism in skeletal size occurs after puberty with larger bones in males, although true bone mineral density remains similar in both sexes. Nutrition and exercise also play an important role in growth, although genetic factors primarily determine peak bone mass.

Numerous genes control skeletal growth, peak bone mass, and body size, as well as skeletal structure and density. Heritability estimates of 50–80% for bone mineral density and size have been derived from twin studies. Though peak bone mass is often lower among individuals with a family history of osteoporosis, association studies of candidate genes (vitamin D receptors; type I collagen, estrogen receptors [ERs], and interleukin 6 [IL-6]; and insulin-like growth factor I [IGF-I]) and bone mass, bone turnover, and fracture prevalence have been inconsistent. There is no panel of genetic markers that can be used to diagnose osteoporosis. Linkage studies suggest that a genetic locus on chromosome 11 is associated with high bone mass. Families with high bone mass and lacking age-related bone loss have been shown to have an activating mutation in LRP5, low-density lipoprotein receptor–related protein 5. Conversely, an inactivating mutation results in osteoporosis-pseudoglioma syndrome, and LRP5 signaling is important in controlling bone formation. Genome-wide scans for low bone mass suggest multiple genes are involved, many of which are also implicated in control of body size.

In adults, bone *remodeling*, not modeling, is the principal metabolic skeletal process. Bone remodeling has two critical functions: (1) to repair bone microdamage to maintain skeletal strength and (2) to supply calcium from the skeleton when required to maintain serum calcium. Remodeling may be activated by bone microdamage due to excessive or accumulated mechanical stress. Acute demands for calcium involve osteoclast-mediated resorption as well as calcium transport by osteocytes. Chronic demands for calcium can result in secondary hyperparathyroidism, increased bone remodeling, and bone loss. Bone remodeling occurs through the well-coordinated activity of osteocytes, osteoblasts, and osteoclasts. Osteocytes are the terminal-differentiated cells derived from osteoblasts after incorporation into newly formed bone tissue. Osteoblasts derive from mesenchymal cell lineage and osteoclasts from monocyte/macrophage lineage. Remodeling sites are discrete units with osteoclasts initiating the process by removal of damaged bone tissue and osteoblasts synthesizing new organic bone that becomes gradually mineralized.

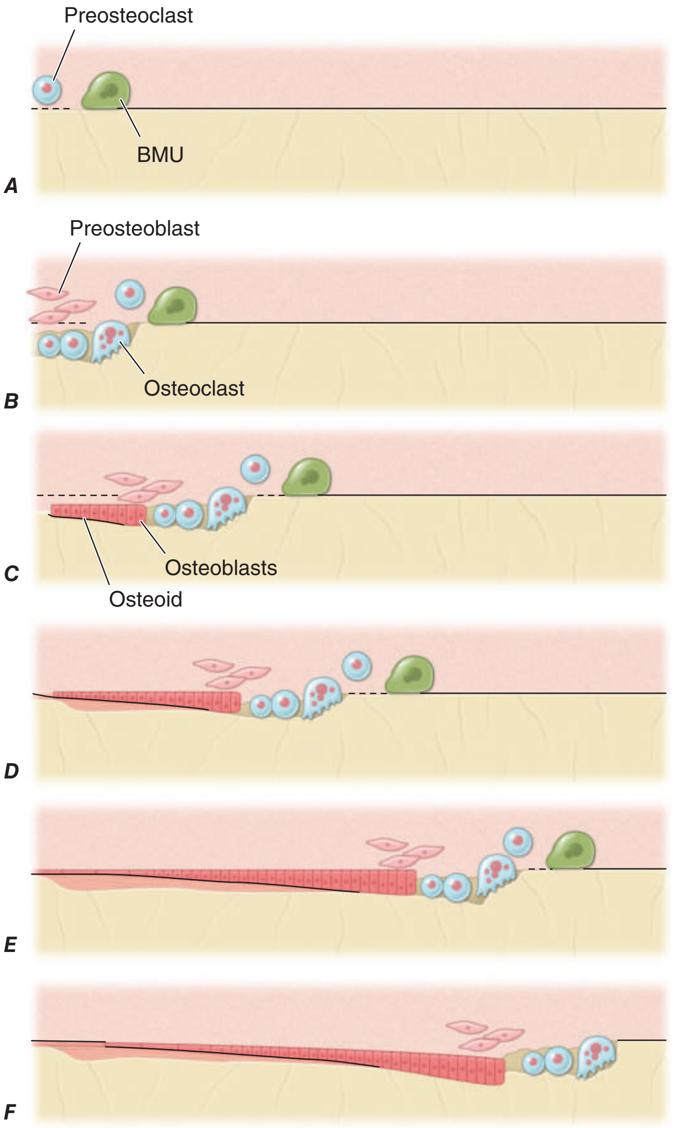

FIGURE 423-4 Mechanism of bone remodeling. The basic molecular unit (BMU) moves along the trabecular surface at a rate of ~10 μm/d. The figure depicts remodeling over ~120 days. ***A.*** Origination of BMU-lining cells contracts to expose collagen and attract preosteoclasts. ***B.*** Osteoclasts fuse into multinucleated cells that resorb a cavity. Mononuclear cells continue resorption, and preosteoblasts are stimulated to proliferate. ***C.*** Osteoblasts align at bottom of cavity and start forming osteoid (black). ***D.*** Osteoblasts continue formation and mineralization. Previous osteoid starts to mineralize (*horizontal lines*). ***E.*** Osteoblasts begin to flatten. ***F.*** Osteoblasts turn into lining cells; bone remodeling at initial surface (left of drawing) is now complete, but BMU is still advancing (to the right). *(Reproduced with permission from SM Ott, in JP Bilezikian, LG Raisz, GA Rodan: Principles of Bone Biology, vol. 18. San Diego, CA: Academic Press; 1996.)*

Bone remodeling is regulated by multiple hormones, including estrogens (in both sexes), androgens, vitamin D, and parathyroid hormone (PTH), as well as locally produced bone-derived growth factors, such as IGF-I, transforming growth factor β (TGF-β), PTH-related peptide (PTHrP), interleukins (ILs), prostaglandins, and members of the tumor necrosis factor (TNF) superfamily. These factors primarily modulate the rate at which new remodeling sites are activated, a process that results initially in bone resorption by osteoclasts, followed by a period of repair during which new bone tissue is synthesized by osteoblasts **(Chap. 421)**. The cytokine responsible for communication

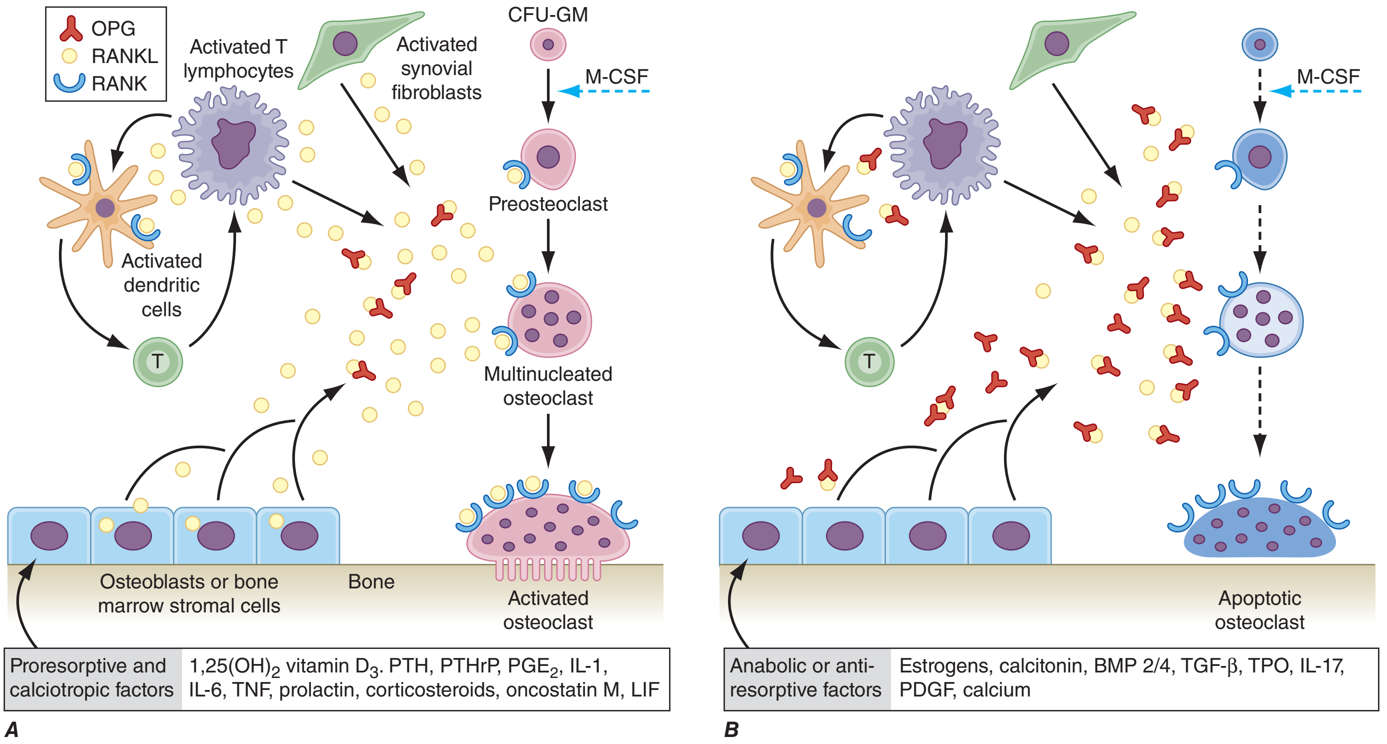
OPG
RANKL
RANK
Activated T lymphocytes
Activated synovial fibroblasts
CFU-GM
M-CSF
Preosteoclast
Activated dendritic cells
T
Multinucleated osteoclast
Osteoblasts or bone marrow stromal cells
Bone
Activated osteoclast
Proresorptive and calciotropic factors
1,25(OH)2 vitamin D3. PTH, PTHrP, PGE2, IL-1, IL-6, TNF, prolactin, corticosteroids, oncostatin M, LIF
A
M-CSF
T
Apoptotic osteoclast
Anabolic or anti-resorptive factors
Estrogens, calcitonin, BMP 2/4, TGF-β, TPO, IL-17, PDGF, calcium
B

FIGURE 423-5 Hormonal control of bone resorption. ***A.*** Proresorptive and calciotropic factors. ***B.*** Anabolic and antiosteoclastic factors. RANKL expression is induced in osteoblasts, activated T cells, synovial fibroblasts, and bone marrow stromal cells. It binds to membrane-bound receptor RANK to promote osteoclast differentiation, activation, and survival. Conversely, osteoprotegerin (OPG) expression is induced by factors that block bone catabolism and promote anabolic effects. OPG binds and neutralizes RANKL, leading to a block in osteoclastogenesis and decreased survival of preexisting osteoclasts. CFU-GM, colony-forming units, granulocyte macrophage; IL, interleukin; LIF, leukemia inhibitory factor; M-CSF, macrophage colony-stimulating factor; OPG-L, osteoprotegerin ligand; PDGF, platelet-derived growth factor; PGE_2, prostaglandin E_2; PTH, parathyroid hormone; RANKL, receptor activator of nuclear factor-κB; TGF-β, transforming growth factor β; TNF, tumor necrosis factor; TPO, thrombospondin. *(Reproduced with permission from WJ Boyle et al: Osteoclast differentiation and activation. Nature 423:337, 2003.)*

between the osteoblasts, other marrow cells, and osteoclasts is receptor activator of nuclear factor-κB (RANK) ligand (RANKL). RANKL, a member of the TNF family, is secreted by osteocytes, osteoblasts, and certain immune cells. The osteoclast receptor for this protein is referred to as *RANK*. Activation of RANK by RANKL is a final common path in osteoclast development and activation. A humoral decoy for RANKL, also secreted by osteoblasts, is referred to as *osteoprotegerin* **(Fig. 423-5)**. Modulation of osteoclast recruitment and activity appears to be related to the interplay among these three factors (RANKL, RANK, and osteoprotegerin). Additional influences include nutrition (particularly calcium intake) and physical activity level. RANKL production is in part regulated by the canonical Wnt signaling pathway. Wnt activation through mechanical loading or by hormonal or cytokine factors stimulates bone formation by increasing formation and activity of osteoblasts and decreases RANKL secretion, which inhibits osteoclast formation and activity. Sclerostin, also an osteocyte protein, is a major inhibitor of Wnt activation and bone formation. Its secretion is inhibited by weight-bearing physical activity and PTH. Both the RANKL and Wnt pathways have become major targets for osteoporosis treatment (see below).

In young adults, bone remodeling is balanced with resorbed bone being replaced by an equal amount of new bone. Thus, the mass of the skeleton remains constant after peak bone mass is achieved. However, after age 30–45 years, bone resorption exceeds formation, resulting in bone loss. This imbalance may begin at different ages, varies at different skeletal sites, and becomes exaggerated in women due to estrogen deficiency. Bone loss can be due to an increase in osteoclastic activity and/or a decrease in osteoblastic activity. In addition, an increase in remodeling activation frequency, and thus the number of remodeling sites, can magnify small imbalances seen at the individual remodeling unit. Increased bone remodeling can result in permanent bone loss and disrupted skeletal microarchitecture, with an imbalance between resorption and formation within each cycle. In trabecular bone, if the osteoclasts penetrate trabeculae, rapid bone loss ensues and cancellous connectivity is reduced. A higher number of remodeling sites increases the likelihood of this event. In cortical bone, increased activation of bone remodeling creates more porous bone. The effect of this increased cortical porosity on cortical bone strength may be modest if the overall diameter of the bone is not changed. However, decreased apposition of new bone on the periosteal surface coupled with increased endocortical resorption of bone decreases the biomechanical strength of long bones. Even a slight exaggeration in normal bone loss increases the risk of osteoporosis-related fractures because of microarchitectural changes, and osteoporosis is largely a disease of disordered skeletal microarchitecture. Currently the predominant clinical tool (dual-energy x-ray absorptiometry [DXA]) measures bone mass rather than bone microarchitecture. Several additional tools are available to estimate bone microarchitecture (including the trabecular bone score [TBS], a noninvasive addition to spinal DXA measurements, and high-resolution peripheral quantitative computerized tomography [HR-pQCT], which has limited availability).

CALCIUM NUTRITION

Peak bone mass may be impaired by inadequate dietary calcium intake during growth among other nutritional factors (calories, protein, and other minerals), leading to increased risk of osteoporosis later in life. During the adult phase of life, insufficient calcium intake contributes to secondary hyperparathyroidism and an increase in bone remodeling rate. PTH stimulates the 1-alpha hydroxylation of vitamin D in the kidney, leading to increased levels of 1,25-dihydroxyvitamin D [$1,25(OH)_2D$] and enhanced gastrointestinal calcium absorption. PTH also reduces renal calcium loss. Although these are appropriate compensatory homeostatic responses for improving calcium economy, the long-term effects are detrimental to the skeleton because the increased remodeling rates and the ongoing imbalance between resorption and formation at remodeling sites combine to accelerate bone loss.

Total daily calcium intakes <400 mg are detrimental to the skeleton, and intakes in the range of 600–800 mg, which is about the average intake among adults in the United States, are also probably suboptimal. The recommended daily required intake of 1000–1200 mg for adults accommodates population heterogeneity in controlling calcium balance. Such intakes should preferentially come from dietary sources, with supplements used only when dietary intakes fall short and cannot be modified easily. The supplement should be enough to bring total intake to ~1200 mg/d. Recent studies have suggested that there may be differences in safety based on calcium source so that increasing dietary calcium is preferred; higher intakes from supplement sources appear to result in a greater risk of renal stones and perhaps cardiovascular events (although the literature is inconsistent and controversial); however, supplement doses below about 700 mg/d have not been associated with cardiovascular events. Increasing calcium intake above this level does not have any benefit. Increasing calcium intake by itself will not prevent bone loss due to other factors (e.g., postmenopausal status)

■ VITAMIN D

(See also Chap. 421) Severe vitamin D deficiency causes rickets in children and osteomalacia in adults. However, vitamin D insufficiency (circulating levels of 25-hydroxyvitamin D [25(OH)D] that may be inadequate [<75 nmol/L or 30 ng/mL] but above the level that results in rickets) may be more prevalent than previously thought, particularly among individuals at increased risk such as the elderly; those living in northern or southern latitudes; and individuals with poor nutrition, obesity, malabsorption, or chronic liver or renal disease. Dark-skinned individuals are also at high risk of vitamin D in the insufficiency range or lower, but African Americans have a low risk of osteoporosis, with better calcium homeostasis than Caucasians.

Although there is considerable controversy about overall optimal health targets for serum 25(OH)D, there is evidence that for optimal skeletal health, serum 25(OH)D should be >75 nmol/L (30 ng/mL). To achieve this level for most adults requires limited skin exposure to sunlight (estimated to be exposure of face and arms for at least one-half hour each day) or an intake of at least 800–1000 units/d, or even higher in individuals with risk factors (as above), particularly obesity.

Vitamin D insufficiency leads to compensatory secondary hyperparathyroidism and is an important risk factor for osteoporosis and fractures. Some studies have shown that >50% of inpatients on a general medical service exhibit biochemical features of vitamin D deficiency, including increased levels of PTH and alkaline phosphatase and lower levels of ionized calcium. Among those living in northern and southern latitudes, vitamin D levels decline during the winter months without supplementation due to insufficient ultraviolet B radiation. This is associated with seasonal bone loss, reflecting increased bone turnover. Even among healthy ambulatory individuals, mild vitamin D deficiency is increasing in prevalence. In part, this is due to decreased exposure to sunlight. Treatment with vitamin D can return levels to normal (>75 nmol/L [30 ng/mL]) and prevent the associated increase in bone remodeling, bone loss, and fractures. Reduced falls and fracture rates also have been documented among individuals in northern latitudes who have greater vitamin D intake and have higher 25(OH)D levels (though one study suggested an increased fall risk with 25[OH]D levels >70 ng/mL or >175 nmol/L). Although vitamin D levels are suspected to affect risk and/or severity of other diseases, including cancers (colorectal, prostate, and breast), autoimmune diseases, multiple sclerosis, and cardiovascular disease, most controlled clinical trials have not confirmed these effects. However, vitamin D does prevent progression of prediabetes to diabetes in those with vitamin D deficiency (25[OH]D levels <12 ng/mL or <30 nmol/L). Treating vitamin D–sufficient individuals with vitamin D does not reduce fractures. For most adults, supplements of 1000–2000 IU/d are adequate and safe.

■ ESTROGEN STATUS

Estrogen deficiency causes bone loss by two distinct but interrelated mechanisms: (1) activation of new bone remodeling sites and (2) initiation or exacerbation of an imbalance between bone formation and resorption, in favor of the latter. The change in activation frequency causes a transient bone loss until a new steady state between resorption and formation is achieved. This commences in the perimenopause prior to cessation of periods. This remodeling imbalance results in a permanent decrement in mass. In addition, the increase in remodeling sites in the skeleton alone increases the probability of trabecular perforation, eliminating the template on which new bone can be formed and accelerating bone loss. The consequence is degraded skeletal microarchitecture, particularly affecting trabecular bone, so that at any given bone density, the risk of a fracture is greater in those who have experienced rapid bone loss than in those who have not. The addition of the TBS to spinal DXA measurements is an attempt to indirectly capture these microarchitectural changes, while HR-pQCT scans measure these directly.

The most common cause of estrogen deficiency is the menopause, which occurs on average at age 51 years **(Chap. 407)**. Thus, with current life expectancy, an average woman will spend ~35 years without an ovarian estrogen supply. Breast cancer treatment with either bilateral ovariectomies and/or aromatase inhibitors is an increasingly common cause of even more severe estrogen deficiency. The mechanism by which estrogen deficiency causes bone loss is summarized in Fig. 423-5. Bone marrow cells (macrophages, monocytes, osteoclast precursors, mast cells) as well as bone cells (osteoblasts, osteocytes, osteoclasts) express both ERs (α and β). Loss of estrogen increases RANKL production but also reduces osteoprotegerin production, increasing osteoclast formation and recruitment. Estrogen also may play a role in determining the life span of bone cells by controlling their rate of apoptosis. Thus, in situations of estrogen deprivation, the life span of osteoblasts may be decreased, whereas the longevity and activity of osteoclasts are increased. The rate and duration of bone loss after menopause are heterogeneous and unpredictable. Once surfaces are lost from trabecular bone, the rate of bone loss declines. In cortical bone, loss is slower but may continue for longer.

Since remodeling is initiated at the surface of bone, it follows that trabecular bone—which has a considerably larger surface area (80% of the total) than cortical bone—will be affected preferentially by estrogen deficiency. Fractures occur earliest at sites where trabecular bone contributes most to bone strength; consequently, vertebral fractures are the most common early skeletal consequence of estrogen deficiency.

In males, estrogen may have an important role in regulation of bone remodeling. In an experiment in which males were rendered estrogen and androgen deficient, restoring estrogen supply reduced remodeling rate more than restoring androgen.

■ PHYSICAL ACTIVITY

Inactivity, such as prolonged bed rest or paralysis, results in significant bone loss. Concordantly, athletes have higher bone mass than nonathletes. These increases in skeletal mass are most marked when the stimulus begins during growth in the years before puberty at the time of skeletal modeling and can result in a higher peak bone mass in both sexes. Adults are less able to increase bone mass after restoration of physical activity. Epidemiologic data support the beneficial effects on the skeleton of chronic high levels of physical activity. Fracture risk is lower in rural communities and in countries where physical activity is maintained into old age. However, when exercise is initiated during adult life, the effects of progressive resistance training on the skeleton are modest, with increases in spine and hip bone mass of 2–4% in short-term randomized trials of <2 years' duration. Muscle strength is also increased, while balance and functional exercises combined also reduce falls by about 25%. A Cochrane review showed balance and functional exercises reduced the number of people who experienced one or more fall-related fractures. Continuing physical activity into the later years may also slow cognitive decline, another major reason for including exercise programs for the aging population.

■ CHRONIC DISEASES

Various genetic and acquired diseases are associated with an increase in the risk of osteoporosis **(Table 423-2)**. Mechanisms that contribute to bone loss are unique for each disease and typically result from multiple factors, including nutrition, reduced physical activity levels, and factors that affect rates of bone remodeling or bone quality. In most, but not all circumstances, the primary diagnosis is made before osteoporosis presents clinically. Both type 1 and type 2 diabetes mellitus are associated with an increased fracture risk, with increased risk at higher

TABLE 423-2 Diseases Associated with an Increased Risk of Generalized Osteoporosis in Adults

Hypogonadal states
- Turner's syndrome
- Klinefelter's syndrome
- Anorexia nervosa
- Hypothalamic amenorrhea
- Hyperprolactinemia
- Other primary or secondary hypogonadal states

Endocrine disorders
- Cushing's syndrome
- Hyperparathyroidism
- Thyrotoxicosis
- Diabetes mellitus (both type 1 and 2)
- Acromegaly
- Adrenal insufficiency

Nutritional and gastrointestinal disorders
- Malnutrition
- Parenteral nutrition
- Malabsorption syndromes
- Gastrectomy
- Severe liver disease, especially biliary cirrhosis
- Pernicious anemia

Rheumatologic disorders
- Rheumatoid arthritis
- Ankylosing spondylitis

Hematologic disorders/malignancy
- Multiple myeloma
- Lymphoma and leukemia
- Malignancy-associated parathyroid hormone–related peptide (PTHrP) production
- Mastocytosis
- Hemophilia
- Thalassemia

Selected inherited disorders
- Osteogenesis imperfecta
- Marfan's syndrome
- Hemochromatosis
- Hypophosphatasia
- Glycogen storage diseases
- Homocystinuria
- Ehlers-Danlos syndrome
- Porphyria
- Menkes' syndrome
- Epidermolysis bullosa

Other disorders
- Immobilization
- Chronic obstructive pulmonary disease
- Pregnancy and lactation
- Scoliosis
- Multiple sclerosis
- Sarcoidosis
- Amyloidosis

bone density than in the nondiabetic population. This is due to differences in collagen cross-linking in bone tissue due to accumulation of advanced glycation end products, making it more brittle than normal, a predilection for conversion of precursors to adipose cells rather than osteoblasts, and the sequelae of diabetes that increase the risk of falls and injury.

Severe bone loss occurs in quadriplegic and paraplegic individuals below the level of the injury. The combination of loss of muscle function and innervation of both muscle and bone contributes to failure to recover mobility, which leads to a high fracture risk in those attempting to pursue athletic activities despite their primary diagnosis (e.g., wheelchair athletes). Bone loss also follows a stroke and is again dependent on the severity of the paralysis. The risk of fracture can be predicted by the FRAX (Fracture Risk Assessment) score and seems highest in the first year after stroke. The increasing prevalence of transgender and gender nonconforming individuals has prompted a guideline for evaluation of bone density in that population by the International Society of Clinical Densitometry published in 2019.

MEDICATIONS

Many medications used in clinical practice have potentially detrimental effects on the skeleton **(Table 423-3)**. *Glucocorticoids* are the most common cause of medication-induced osteoporosis. It is often not possible to determine the extent to which osteoporosis is related to glucocorticoid treatment or to other factors, as the effects of medication are superimposed on the effects of the primary disease, which may be associated with bone loss (e.g., rheumatoid arthritis). Excessive doses of thyroid hormone can accelerate bone remodeling and result in bone loss.

TABLE 423-3 Drugs Associated with an Increased Risk of Generalized Osteoporosis in Adults

- Glucocorticoids
- Cyclosporine
- Cytotoxic drugs
- Anticonvulsants
- Aromatase inhibitors
- Selective serotonin reuptake inhibitors
- Excessive thyroxine
- Aluminum
- Gonadotropin-releasing hormone agonists
- Heparin
- Lithium
- Protein pump inhibitors
- Thiazolidinediones
- Androgen deprivation therapies

Other medications have less detrimental effects on the skeleton than pharmacologic doses of glucocorticoids. *Anticonvulsants* are thought to increase the risk of osteoporosis, although many affected individuals have concomitant insufficiency of 1,25(OH)$_2$D, as some anticonvulsants induce the cytochrome P450 system and vitamin D metabolism. Patients undergoing transplantation are at high risk for rapid bone loss and fracture not only from glucocorticoids but also from treatment with other *immunosuppressants* such as cyclosporine and tacrolimus (FK506). In addition, these patients often have preexisting metabolic abnormalities such as hepatic or renal failure predisposing to bone loss. Long-term use of proton pump inhibitors and selective serotonin reuptake inhibitors has been shown to be associated with a higher risk of fracture in observational studies. Given their frequent long-term use, their skeletal effects are important from a public health perspective and for individual fracture risk.

Aromatase inhibitors, which potently block the aromatase enzyme that converts androgens and other adrenal precursors to estrogen, reduce circulating postmenopausal estrogen levels severely. These agents, used to treat breast cancer, also cause declines in bone density and rapidly increase fracture risk. Androgen deprivation therapy, used to treat men with prostate cancer, also results in rapid bone loss and increased fracture risk. The diabetes medications, thiazolidinediones and insulin, also increase the risk of fracture; however, metformin and glucagon-like peptide-1 receptor agonists decrease risk. It is difficult in some cases to separate the risk accrued by the underlying disease from that attributable to the medication. For example, both depression and diabetes are risk factors for fracture by themselves.

SMOKING

Smoking produces detrimental effects on bone mass mediated directly by toxic effects on osteoblasts or indirectly by modifying estrogen metabolism. On average, cigarette smokers reach menopause 1–2 years earlier than the general population. Cigarette smoking also results in intercurrent respiratory and other illnesses, frailty, decreased exercise, poor nutrition, and the need for additional medications (e.g., glucocorticoids for lung disease) that can all increase fracture risk.

OTHER POTENTIAL FACTORS

In the past few years, many potential risk factors for fracture have been identified. These include excessive alcohol intake and other drugs of abuse, pollution, use of triclosan, chronic obstructive pulmonary disease, excess vitamin B, and hormonal therapies utilized among the transgender population.

DIAGNOSIS

MEASUREMENT OF BONE MASS

Several noninvasive techniques are available for estimating skeletal mass or BMD. They include DXA, quantitative computed tomography (QCT), peripheral QCT, HR-pQCT, and ultrasound. DXA is a highly accurate x-ray technique that has become the standard for measuring bone density. Though it can be used for measurement in any skeletal site, clinical determinations usually are made of the lumbar spine and hip. DXA also can be used to measure the radius, total body bone mass, and body composition (lean mass, fat mass). Two x-ray energies are used to estimate mineralized tissue, allowing for correction for attenuation through soft tissue. The mineral content is divided by bone area, which partially corrects for body and bone size. However, this correction is only partial since DXA is a two-dimensional scanning technique and cannot estimate the depth of the bone. Thus, small slim people tend to have lower than average BMD, a feature that is important in interpreting BMD measurements. Bone spurs, which are common in osteoarthritis (spinal spondylosis), tend to falsely increase

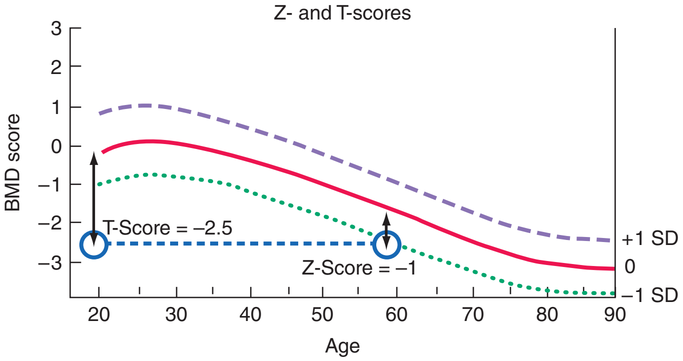

FIGURE 423-6 Relationship between Z-scores and T-scores in a 60-year-old woman. BMD, bone mineral density; SD, standard deviation.

bone density, mostly of the spine, which is a particular problem in measuring spine BMD in older individuals. Because DXA measurement devices are provided by different manufacturers, the output varies in absolute terms. Absolute bone density results are related to "normal" values by using T-scores (a T-score of 1 equals 1 SD), which compare individual results to those in a young adult population (age 30 years) matched for race and sex. The mean value is given a score of zero and the range is +2.5 to –2.5 (i.e., 2.5 SDs above or below the mean). Z-scores (also SDs) compare individual results to those of an age- and sex-matched reference population. Thus, a 60-year-old woman with a Z-score of –1 (1 SD below mean for age) has a T-score of –2.5 (2.5 SDs below mean for a young adult population) **(Fig. 423-6)**. A T-score ≤–2.5 in the lumbar spine, femoral neck, or total hip has been defined as osteoporosis, whereas a T-score of –1.0 to –2.5 has been defined as low bone density or osteopenia. Although the outputs from different instruments and, more importantly, different manufacturers correlate well, different machines used over time may show changes in BMD that may be attributed to biological changes or simply the result of differences between machines. This is particularly true with measurements of the femoral neck. Consequently, it is recommended that serial measurements be performed on the same machine and preferably by the same technician.

As noted above, since >50% of fractures occur in individuals with osteopenia (e.g., a T-score between –1.0 and –2.5), it is important to consider absolute fracture risk in addition to BMD. The most commonly used absolute fracture risk assessment tool is FRAX, which includes age, sex, height, weight, fracture history, hip fracture in a parent, glucocorticoid use, rheumatoid arthritis, and other secondary causes, with or without bone density of the femoral neck. The program calculates the estimated risk over a 10-year time frame for major osteoporosis-related fractures (clinical spine, hip, wrist, and proximal humerus) and for hip fracture.

Computed tomography (CT) can also be used to measure the spine and hip but is rarely used clinically, in part because the radiation exposure and cost are both much higher than with DXA. HR-pQCT measures cortical and trabecular bone variables in the forearm or tibia and provides information on skeletal microarchitecture noninvasively. Magnetic resonance imaging (MRI) can also be used to obtain some architectural information on the forearm and perhaps the hip but is primarily a research tool.

Ultrasound can be used to measure bone mass by calculating the attenuation of the signal as it passes through bone or the speed with which it traverses the bone. Although the ultrasound technique was purported to assess properties of bone other than mass (e.g., quality), this has not been confirmed. Because of its relatively low cost and mobility, ultrasound bone density measurement is amenable for use as a screening tool.

All these techniques for measuring BMD have been approved by the U.S. Food and Drug Administration (FDA) based on their capacity to predict fracture risk. The total hip is the preferred site of measurement, since it allows the best prediction of hip fracture risk. When hip measurements are performed by DXA, the spine is usually measured at the same time. In younger individuals such as perimenopausal or early postmenopausal women, spine measurements may be more sensitive. When the spine or hip is not measurable due to severe degenerative spine disease or scoliosis, or prior spine or hip surgery, wrist BMD is often measured.

TABLE 423-4 Indications for Bone Mineral Density Testing

- Women aged ≥65 and men aged ≥70; regardless of clinical risk factors
- Younger postmenopausal women, women in the menopausal transition, and men aged from 50 to 69 with clinical risk factors for fracture
- Adults who have a fracture at or after age 50
- Adults with a condition (e.g., rheumatoid arthritis) or taking a medication (e.g., glucocorticoids at a daily dose >5 mg prednisone or equivalent for >3 months associated with low bone mass or bone loss

INDICATIONS FOR BONE MASS MEASUREMENT

Several clinical guidelines have been developed for the use of bone densitometry in clinical practice **(Table 423-4)**. The National Osteoporosis Foundation (NOF) guidelines recommend bone mass measurements in postmenopausal women who have one or more risk factors for osteoporosis in addition to age, sex, and estrogen deficiency. The guidelines further recommend that bone mass measurement be considered in *all* women by age 65, a position ratified by the U.S. Preventive Health Services Task Force. In males, the use of bone density determination is not recommended until the age of 70 years in the absence of multiple risk factors or the occurrence of a fragility fracture.

The FRAX score incorporates clinical risk factors (age, prior fracture, family history of hip fracture, low body weight, tobacco use, excessive alcohol use, glucocorticoid use, and rheumatoid arthritis) with or without BMD to assess the 10-year fracture probabilities for major osteoporotic (hip, clinical spine, wrist, or humerus) and hip fractures. Fracture risk probability calculators are available as part of the report from all DXA machines and available online (*https://www.sheffield.ac.uk/FRAX/*) **(Fig. 423-7)**. In the United States, cost-effective treatment thresholds occur if the 10-year major osteoporotic fracture risk from FRAX is ≥20% and/or the 10-year risk of hip fracture is ≥3%. FRAX is an imperfect tool, as it does not include any assessment of fall risk, and secondary causes are excluded when BMD is entered. It also does not account for glucocorticoid dose or diabetes. Most importantly, it does not increase the absolute fracture risk conferred by a recent fracture versus the lesser risk conferred by a more remote fracture. Moreover, there is no mandate for vertebral fracture diagnosis and no additional fracture probability estimated for patients who have had multiple fragility fractures. Many of these deficiencies are likely to be corrected in the next version of FRAX (FRAX Plus and FRAX 2.0).

VERTEBRAL IMAGING

DXA equipment can also be used to obtain lateral images of the thoracic and lumbar spine, a technique called vertebral fracture assessment (VFA). While not as definitive as a radiograph, it is an excellent screening tool for vertebral abnormality in both women and men. It is an important assessment as the majority of vertebral fractures are asymptomatic (70%). Furthermore, VFA can detect vertebral abnormalities causing height loss or back pain that may be indicative of an undiagnosed vertebral fracture.

Because most vertebral fractures are asymptomatic, the diagnosis of vertebral fracture is rarely made when they first occur. However, vertebral fractures, whether symptomatic or asymptomatic, are associated with the same clinical sequelae, so it is critical they are identified. Vertebral fracture prevalence in the United States based on the National Health and Nutrition Evaluation Studies (NHANES) population appears to be 20% in the 1980s, when the strictest criteria for diagnosis are utilized. It is recommended that women aged 65 years or older and men aged 70 years or older undergo vertebral imaging if there is a T-score is ≤–1.5 at the spine, hip, or femoral neck. Vertebral imaging is also recommended for women by the age of 70 years and men by the age of 80 years if a T-score is <–1.0. For younger individuals, vertebral imaging is recommended for those with a fragility fracture, height loss, or glucocorticoid use. **(See Table 423-5.)**

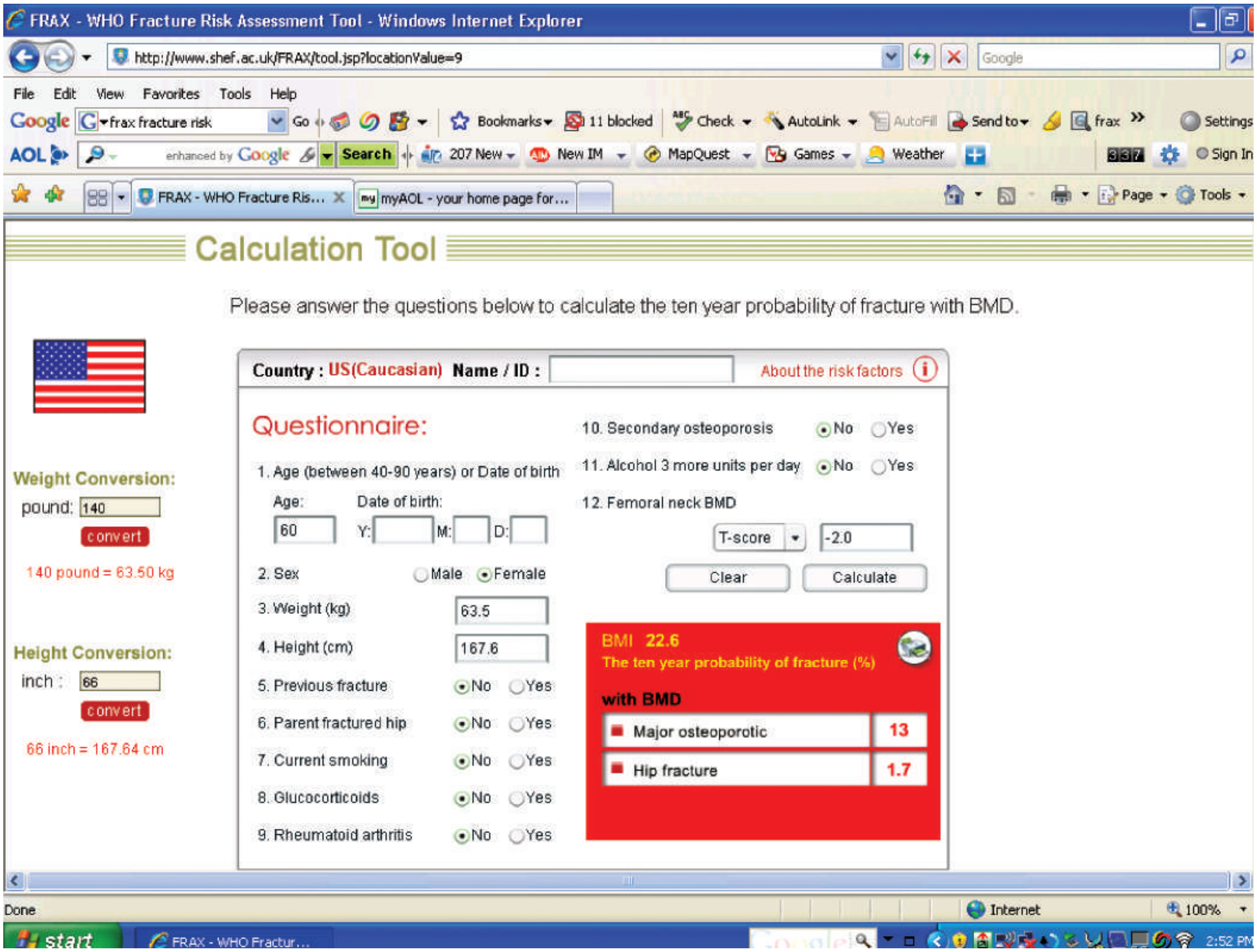
FRAX - WHO Fracture Risk Assessment Tool - Windows Internet Explorer
http://www.shef.ac.uk/FRAX/tool.jsp?locationValue=9
File Edit View Favorites Tools Help
frax fracture risk
Bookmarks
11 blocked
Check
AutoLink
AutoFill
Send to
frax
Settings
Search
207 New
New IM
MapQuest
Games
Weather
Sign In
FRAX - WHO Fracture Ris...
myAOL - your home page for...
Page
Tools
Done
Internet
100%
start
FRAX - WHO Fractur...
2:52 PM

Calculation Tool

Please answer the questions below to calculate the ten year probability of fracture with BMD.

Weight Conversion:
pound: 140
convert
140 pound = 63.50 kg

Height Conversion:
inch: 66
convert
66 inch = 167.64 cm

Country : US(Caucasian) Name / ID :
About the risk factors

Questionnaire:

1. Age (between 40-90 years) or Date of birth
Age: 60 Date of birth: Y: M: D:
2. Sex ○ Male ● Female
3. Weight (kg) 63.5
4. Height (cm) 167.6
5. Previous fracture ● No ○ Yes
6. Parent fractured hip ● No ○ Yes
7. Current smoking ● No ○ Yes
8. Glucocorticoids ● No ○ Yes
9. Rheumatoid arthritis ● No ○ Yes
10. Secondary osteoporosis ● No ○ Yes
11. Alcohol 3 more units per day ● No ○ Yes
12. Femoral neck BMD T-score -2.0

Clear Calculate

BMI 22.6
The ten year probability of fracture (%)
with BMD

Major osteoporotic	13
Hip fracture	1.7

FIGURE 423-7 FRAX calculation tool. When the answers to the indicated questions are filled in, the calculator can be used to assess the 10-year probability of fracture. The calculator (available online at *http://www.shef.ac.uk/FRAX/tool.jsp?locationValue=9*) also can risk adjust for various ethnic groups.

APPROACH TO THE PATIENT

Osteoporosis

The development of osteoporosis is a gradual process occurring under a variety of genetic and environmental influences throughout life. Just prior to and at the time of puberty are important times to affect accrual of peak bone mass while the perimenopause and menopause in women are important times to reduce bone loss. Recognition of these influences allows intervention at several points, although the type of intervention depends on a careful evaluation of each individual patient.

The menopausal transition affects all women by their late 50s and represents an opportunity to initiate a discussion about bone loss, the role of estrogen loss, and other clinical risk factors. Assessment of fracture risk with FRAX (with or without bone density) provides a 10-year estimate of hip and major osteoporosis-related fracture risk and opens the discussion about preventive steps including, if required, the use of medication. If risk is low, then nutrition and resistance training exercise are the focus. If menopausal symptoms are prominent and estrogen intervention is needed, then the added protection against bone loss and fragility fractures should be emphasized.

Among older women with a fracture, an evaluation of skeletal status including bone density testing should proceed. In this case, any fracture, whether traumatic or not, should trigger this assessment. Although osteoporosis is associated with a risk of fragility fracture, individuals with osteoporosis are also more likely to experience traumatic fractures, and such individuals should not be excluded from osteoporosis evaluation simply because of the level of trauma. This concept, while obvious and evidence-based, still needs emphasis with individual patients, physicians, and payors.

Patients who present with hip or spine fractures by definition have osteoporosis and will require treatment for both the fracture itself and the underlying skeletal disorder. Other long bone fractures (e.g., distal radius) are triggers for evaluation of the skeleton upon which treatment decisions can be based.

In all individuals presenting with a fracture as a result of a fall, fall prevention strategies including balance exercises are an important adjunct to other lifestyle and nutritional interventions.

TABLE 423-5 Indications for Vertebral Testing

Consider vertebral imaging tests for the following individuals[a]

- All women aged ≥70 and all men aged ≥80 if bone mineral density (BMD) T-score at the spine, total hip, or femoral neck is <1.0
- Women aged from 65 to 69 and men aged from 70 to 79 if BMD T-score at the spine, total hip, or femoral neck is <1.5
- Postmenopausal women and men aged ≥50 with specific risk factors:
  - Low-trauma fracture during adulthood (aged ≥50)
  - Historical height loss of ≥1.5 in. (4 cm)[b]
  - Prospective height loss of ≥0.8 in. (2 cm)[c]
  - Recent or ongoing long-term glucocorticoid treatment

[a] If bone density testing is not available, vertebral imaging may be considered based on age alone. [b] Current height compared to peak height during childhood. [c] Cumulative height loss measured during interval medical assessment.

ROUTINE LABORATORY EVALUATION

There is no established algorithm for the evaluation of women or men who present with osteoporosis. A general evaluation that includes complete blood count, serum and 24-h urine calcium, and renal and liver function tests is useful for identifying selected secondary causes of low bone mass, particularly for those with fractures or unexpectedly low Z-scores. An elevated serum calcium level suggests hyperparathyroidism or malignancy, whereas a reduced serum calcium level may reflect malnutrition or a malabsorption disease, such as celiac disease. In the presence of hypercalcemia, a serum PTH level differentiates between hyperparathyroidism (PTH↑) and malignancy (PTH↓), and a high PTHrP level can help document the presence of humoral hypercalcemia of malignancy **(Chap. 422)**. A low urine calcium (<50 mg/24 h) suggests malnutrition, or malabsorption; a high urine calcium (>300 mg/24 h) during normal calcium intake (excluding calcium supplements for at least a week before the urine collection) is indicative of hypercalciuria. Hypercalciuria occurs primarily in three situations: (1) a renal calcium leak; more common in males with osteoporosis; (2) absorptive hypercalciuria, which can be idiopathic or associated with increased $1{,}25(OH)_2D$ in granulomatous disease; or (3) hematologic malignancies or conditions associated with excessive bone turnover such as Paget's disease, hyperparathyroidism, hyperthyroidism, and iron chelation therapy (desferoxamine) for thalassemia. Renal hypercalciuria is treated with thiazide diuretics, which lower urine calcium and help improve calcium economy. In this setting, thiazides alone can improve bone mass and possibly reduce risk of fracture. They might also reduce renal stone risk.

Individuals who have fragility fractures or bone density in the osteoporotic range should have a measurement of serum 25(OH)D level since serum 25(OH)D levels <30 ng/mL (vitamin D insufficiency) are common, particularly in older adults. Hyperthyroidism should be evaluated by measuring thyroid-stimulating hormone (TSH).

When there is clinical suspicion of Cushing's syndrome, 24-h urinary free cortisol levels or a fasting serum cortisol should be measured after taking 1 mg dexamethasone at midnight. When bowel disease, malabsorption, or malnutrition is suspected, serum albumin, cholesterol, and a complete blood count should be checked. Asymptomatic malabsorption may be heralded by anemia (macrocytic—vitamin B_{12} or folate deficiency; microcytic—iron deficiency) or low serum cholesterol or urinary calcium levels. If these or other features suggest malabsorption, further evaluation is required. Asymptomatic celiac disease with selective malabsorption is being found with increasing frequency; the diagnosis can be made by testing for transglutaminase IgA antibodies but may require confirmation by endoscopic biopsy. A trial of a gluten-free diet can also be confirmatory **(Chap. 336)**. When osteoporosis is found associated with symptoms of rash, multiple allergies, diarrhea, or flushing, mastocytosis should be considered and excluded by using 24-h urine histamine collection or serum tryptase.

Myeloma can masquerade as generalized osteoporosis, although it more commonly presents with bone pain and characteristic "punched-out" lesions on radiography. Serum and urine electrophoresis and/or evaluation for serum free light chains in serum are required to exclude this diagnosis. More commonly, a monoclonal gammopathy of undetermined significance (MGUS) is found, and the patient is subsequently monitored to ensure that this is not an incipient myeloma. MGUS itself may be associated with an increased risk of osteoporosis. A bone marrow biopsy may be required to rule out myeloma (in patients with equivocal electrophoretic results) and can be used to exclude mastocytosis, leukemia, and other marrow infiltrative disorders, such as Gaucher's disease. Testosterone levels should be checked in men with osteoporosis.

An important cause of fracture among the aging population is diabetes, both type 1 and type 2. Patients with diabetes appear, at any given bone density, to be at higher risk of fracture than nondiabetics. The reasons include the effects on muscle and nerve that increase the risk of falls, but it can also be due to an underlying skeletal fragility as part of the metabolic consequences of diabetes with deposition of advanced glycation end products in bone.

BONE BIOPSY

Double tetracycline labeling of the skeleton allows determination of the rate of remodeling as well as evaluation for other metabolic bone diseases. However, the current use of BMD, in combination with hormonal evaluation and biochemical markers of bone remodeling, has largely replaced the clinical use of bone biopsy, although it remains an important tool in the diagnosis of chronic kidney disease–mineral bone disease (CKD-MBD), in evaluating the mechanism of action of osteoporosis pharmacology and in clinical research.

BIOCHEMICAL MARKERS

Several biochemical tests are available that provide an index of the overall rate of bone remodeling **(Table 423-6)**. Biochemical markers usually are characterized as those related primarily to *bone formation* or *bone resorption*. These tests measure the overall state of bone remodeling at a single point in time. Clinical use of these tests has been hampered by biologic variability (in part related to circadian rhythm and food consumption) as well as analytic variability, although the latter has improved.

For the most part, remodeling markers do not predict rates of bone loss well enough in individuals to make accurate assessment of potential future changes in bone density. However, they do provide adjunct information that assists in both evaluation of the patient and in assessment of treatment response. Markers of bone resorption may help in the prediction of fracture risk, independently of bone density, particularly in older individuals. In women ≥65 years, when bone density results are greater than the usual treatment thresholds noted above, a high level of bone resorption should prompt consideration of treatment. However, the primary use of biochemical markers is for monitoring treatment response. With the introduction of antiresorptive drugs, bone remodeling declines rapidly, with the fall in resorption occurring earlier than the fall in formation. Inhibition of bone resorption is maximal within 3 months or so. Thus, measurement of bone resorption (serum C-terminal telopeptide measured in a fasting specimen at 9 A.M. is the preferred marker) before initiating therapy and 3–6 months after starting therapy provides an earlier estimate of patient response than does bone densitometry. A decline in bone resorption markers can be ascertained after treatment with bisphosphonates, denosumab, or estrogen; this effect is less marked after treatment with weaker agents such as raloxifene or calcitonin. Bone turnover markers are also useful in monitoring the effects of anabolic drugs [hPTH(1–34) or teriparatide or romosozumab], which both rapidly increases bone formation (P1NP is the most sensitive, but osteocalcin is also a very good bone remodeling marker), while teriparatide increases bone resorption and romosozumab decreases bone resorption. The recent suggestion of "drug holidays" (see below) has opened another use for biochemical markers, allowing evaluation of the offset of effect of drugs such as oral and intravenous bisphosphonates. Drug holidays should be avoided for denosumab or anabolic drugs because of a more rapid offset effect observed with these drugs.

TABLE 423-6 Biochemical Markers of Bone Metabolism in Clinical Use

Bone formation
Serum bone-specific alkaline phosphatase
Serum osteocalcin
Serum propeptide of type I procollagen
Bone resorption
Urine and serum cross-linked N-telopeptide
Urine and serum cross-linked C-telopeptide

TREATMENT

Osteoporosis

MANAGEMENT OF PATIENTS WITH FRACTURES

Treatment of a patient with osteoporosis frequently involves management of acute fractures as well as treatment of the underlying disease. Hip fractures almost always require surgical repair. Depending on the location and severity of the fracture, condition of the neighboring joint, and general status of the patient, procedures may include open reduction and internal fixation with pins and plates, hemiarthroplasties, and total arthroplasties. These surgical procedures are followed by intense rehabilitation to return patients to their prefracture functional level. Long bone fractures often require either external or internal fixation. Other fractures (e.g., vertebral, rib, and pelvic fractures) can often be managed with supportive care, requiring no specific orthopedic treatment.

Only ~30% of vertebral compression fractures present with sudden-onset back pain. For acutely symptomatic fractures, treatment with analgesics is required, including nonsteroidal anti-inflammatory agents and/or acetaminophen, sometimes with the addition of a narcotic agent. (A few small, randomized clinical trials suggest that calcitonin may reduce pain related to acute vertebral compression fracture). Vertebral augmentation with the percutaneous injection of artificial cement (polymethylmethacrylate) into the vertebral body (vertebroplasty or kyphoplasty) may offer significant pain relief in a subset of patients with continuing pain; however, systematic reviews of controlled trials of both procedures have provided doubt about their efficacy. Furthermore, risks include acute extravasation of cement outside of the vertebral body with neurologic impairment and possibly an increased risk of vertebral fracture in adjacent vertebrae due to increased rigidity of the treated vertebral body. Short periods of bed rest may be helpful for pain management, but in general, early mobilization is recommended as it helps prevent further bone loss associated with immobilization. Occasionally, use of a soft elastic-style back brace may facilitate earlier mobilization. Muscle spasms often occur with acute compression fractures and can be treated with muscle relaxants and heat treatments. Severe pain usually resolves within 6–10 weeks. More chronic severe pain might suggest the possibility of multiple myeloma or other underlying bone conditions.

Vertebral fractures cause height loss because of the loss of vertebral body height during compression of the vertebral body. These fractures can produce a kyphosis, particularly when wedge shaped, or just loss of thoracic height. Chronic pain following vertebral fracture is probably not bony in origin; instead, it is related to abnormal strain on muscles, ligaments, and tendons and to secondary facet-joint arthritis associated with alterations in thoracic and/or abdominal shape. Chronic pain may also be the result of ribs sitting right on top of the iliac crest bones, particularly in patients who have had multiple vertebral compression fractures. Chronic pain is difficult to treat effectively and may require analgesics, sometimes including narcotic analgesics with the attendant risk of addiction. Frequent intermittent rest in a supine or semireclining position is often required to allow the soft tissues, which are under tension, to relax. Back and core-strengthening exercises may be beneficial. Heat treatments help relax muscles and reduce the muscular component of discomfort. Various physical modalities, such as ultrasound and transcutaneous nerve stimulation, may be beneficial in some patients. Pain also occurs in the neck region, not as a result of compression fractures (which almost never occur in the cervical spine as a result of osteoporosis) but because of chronic strain associated with trying to elevate the head in a person with a significant thoracic kyphosis.

Multiple vertebral fractures often are associated with often neglected psychological symptoms. Changes in body configuration and back pain can lead to marked loss of self-image and a secondary depression. Altered balance, precipitated by the kyphosis and the anterior movement of the body's center of gravity, leads to a fear of falling, a consequent tendency to remain indoors, and the onset of social isolation. These symptoms sometimes can be alleviated by family support and/or psychotherapy. Medication may be necessary when depressive features are present.

A MISSED OPPORTUNITY

Multiple studies show most patients presenting with fractures after age 50 years (even fractures traditionally linked to osteoporosis) are not screened or treated for osteoporosis. Only <25% of fracture patients receive follow-up care. Recently, several studies have demonstrated the effectiveness of a relatively simple and inexpensive program that reduces the risk of subsequent fractures. In the Kaiser system, it is estimated that a 20% decline in hip fracture occurrence was seen with the introduction of a fracture liaison service. This approach has also been successful in other non-U.S. health systems, including the United Kingdom and New Zealand. This involves a health care professional (usually a nurse or physician's assistant) whose job is to educate patients and coordinate evaluation and osteoporosis treatment as patients move through the emergency room, inpatient care in an acute care hospital, rehabilitation hospital care, and/or orthopedic practice to outpatient management. If the Kaiser experience can be repeated, there would not only be significant savings of health care dollars but also a dramatic drop in hip fracture incidence and a marked improvement in morbidity and mortality among the aging population. Initiatives to obtain Medicare funding for fracture liaison services in the United States are continuing but have not yet been successful.

MANAGEMENT OF THE UNDERLYING DISEASE

Risk Factor Reduction After risk assessment, patients should be thoroughly educated to reduce the impact of modifiable risk factors associated with bone loss and falling. Medications should be reviewed to ensure that all are necessary and taken at the lowest required dose. Glucocorticoid medication, if present, should be evaluated to determine that it is truly indicated and is being given in doses as low as possible. For those on thyroid hormone replacement, TSH testing should be performed to determine that an excessive dose is not being used, as iatrogenic thyrotoxicosis can be associated with increased bone loss. In patients who smoke, efforts should be made to facilitate smoking cessation. Reducing risk factors for falling also includes alcohol abuse treatment and a review of the medical regimen for any drugs that might be associated with orthostatic hypotension and/or sedation, including hypnotics, antipsychotics, and anxiolytics. If nocturia occurs, the frequency should be reduced, if possible (e.g., by decreasing or modifying diuretic use), as arising in the middle of sleep is a common precipitant of a fall. Patients should be instructed about environmental safety, such as eliminating exposed wires, curtain strings, slippery rugs, and mobile tables. Avoiding stocking feet on wood floors, checking carpet condition (particularly on stairs), and providing good light in paths to bathrooms and outside the home are important preventive measures. Treatment for impaired vision is recommended, particularly a problem with depth perception, which is specifically associated with increased falling risk. Patients with neurologic impairment (e.g., stroke, Parkinson's disease, Alzheimer's disease, multiple sclerosis) are particularly at risk of falling and require specialized supervision and care. In patients with risk factors for falls, especially those who live alone or spend significant time alone, medical alert systems should be prescribed.

Nutritional Recommendations • **Calcium** A large body of data indicates that less than optimal calcium intake results in bone loss. Consequently, an adequate intake suppresses bone turnover. Recommended intakes from an Institute of Medicine report are shown in **Table 423-7**. The NHANES have consistently documented that average calcium intakes fall considerably short of these recommendations. The preferred source of calcium is diet, but many patients require calcium supplementation to bring intake to ~1000 mg/d. Calcium supplement doses of ≤600 mg/d are likely to be safe. Best sources of

TABLE 423-7 Adequate Calcium Intake

LIFE STAGE GROUP	ESTIMATED ADEQUATE DAILY CALCIUM INTAKE, mg/d
Young children (1–3 years)	500
Older children (4–8 years)	800
Adolescents and young adults (9–18 years)	1300
Men and women (19–50 years)	1000
Men and women (51 years and older)	1200

Note: Pregnancy and lactation needs are the same as for nonpregnant women (e.g., 1300 mg/d for adolescents/young adults and 1000 mg/d for those ≥19 years old).

Source: Data from Institute of Medicine. Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride. Washington, DC: The National Academies Press; 1997.

calcium include dairy products (milk, yogurt, and cheese), nondairy milks (almond, rice, soy), nuts, and fortified foods such as certain cereals, waffles, snacks, juices, and crackers. Some of these fortified foods contain as much calcium per serving as milk. Various vegetables and fruits, such as kale, broccoli, and dried figs, contain reasonably high calcium content, although some of it may not be fully bioavailable. Calcium intake calculators are available at *NOF.org* or *NYSOPEP.org* and will give a rough idea of total calcium intake.

If calcium supplements are required, they should be taken in doses sufficient to bring total intake to the required level (~1000 mg/d). Doses of supplements should be ≤600 mg per single dose, as the calcium absorption fraction decreases at higher doses. Calcium supplements should be calculated on the basis of the elemental calcium content of the supplement, not the weight of the calcium salt **(Table 423-8)**. Calcium supplements containing carbonate are best taken with food since they require acid for solubility. Calcium citrate supplements can be taken at any time, although taking calcium supplements in the evening may reduce nocturnal increases in PTH.

Several controlled clinical trials of calcium, mostly with accompanying vitamin D, have confirmed reductions in clinical fractures, including fractures of the hip (~20–30% risk reduction), particularly in elderly individuals who are likely to be dietarily deficient. The majority of recent studies of pharmacologic agents have been conducted in the context of calcium replacement (± vitamin D). Thus, it is standard practice to ensure an adequate calcium and vitamin D intake in patients with osteoporosis whether they are receiving additional pharmacologic therapy or not. A systematic review confirmed a greater BMD response to antiresorptive therapy when calcium intake was adequate.

Although side effects from supplemental calcium are minimal (eructation and constipation mostly with carbonate salts), individuals with a history of kidney stones should have a 24-h urine calcium determination before starting increased calcium to avoid exacerbating hypercalciuria. A recent analysis of published data has suggested that high intakes of calcium from supplements are associated with an increase in the risk of renal stones, calcification in arteries, and potentially an increased risk of heart disease and stroke. This is controversial, with recent meta-analyses showing no effect of calcium supplements on cardiovascular events. Daily doses of <700 mg appear safe. However, since high calcium intakes also increase the risk of renal stones and confer no extra benefit to the skeleton, the recommendation that total intakes should be between 1000 and 1500 mg/d seems reasonable.

TABLE 423-8 Elemental Calcium Content of Various Oral Calcium Preparations

CALCIUM PREPARATION	ELEMENTAL CALCIUM CONTENT
Calcium citrate	60 mg/300 mg
Calcium lactate	80 mg/600 mg
Calcium gluconate	40 mg/500 mg
Calcium carbonate	400 mg/g
Calcium carbonate + 5 μg vitamin D_2 (OsCal 250)	250 mg/tablet
Calcium carbonate (Tums 500)	500 mg/tablet

Source: Adapted from SM Krane, MF Holick, in *Harrison's Principles of Internal Medicine*, 14th ed. New York, NY: McGraw Hill; 1998.

Vitamin D Diet alone is inadequate to maintain vitamin D sufficiency (serum 25[OH]D consistently >75 μmol/L [30 ng/mL]). Vitamin D is synthesized from a precursor in the skin under the influence of heat and ultraviolet light **(Chap. 421)**. Production is blocked by sunscreen and sun avoidance. However, large segments of the population do not obtain sufficient vitamin D from either skin production or dietary sources. Since vitamin D supplementation at doses that would achieve these serum levels is safe and inexpensive, the National Academy of Medicine recommends daily intakes of 200 IU for adults <50 years of age, 400 IU for those 50–70 years, and 600 IU for those >70 years (based on obtaining a serum level of 20 ng/mL, lower than the level recommended by most other guidelines). Multivitamin tablets usually contain 400 IU, and many calcium supplements also contain vitamin D. Some data suggest that higher doses (≥1000 IU) may be required, particularly in the obese, elderly, and chronically ill. A daily intake of up to 4000 IU/d is safe. For those with osteoporosis or those at risk, 1000–2000 IU/d can usually maintain serum 25(OH)D above 30 ng/mL. Recent large randomized controlled trials show vitamin D supplementation by itself at doses equivalent to 2000 IU/d does not appear to reduce fracture risk in older individuals with normal baseline vitamin D levels. However, the combination of adequate calcium intake and vitamin D does decrease fracture risk in those with vitamin D deficiency, particularly the institutionalized elderly population. Although low vitamin D levels appear associated with more serious outcomes in response to COVID-19, trials of vitamin D supplementation have been inconsistent and largely negative.

Other Nutrients Other nutrients such as salt, high animal protein intakes, and caffeine may have modest effects on calcium excretion or absorption. Adequate vitamin K status is required for optimal carboxylation of osteocalcin. States in which vitamin K nutrition or metabolism is impaired, such as with long-term warfarin therapy, have been associated with reduced bone mass. However, a recent trial showed no benefit of vitamin K_2 on BMD when added to calcium and vitamin D supplementation. Research concerning cola-based soda beverage intake is controversial but suggests a possible link to reduced bone mass through factors that appear independent of caffeine.

Magnesium is abundant in foods, and magnesium deficiency is quite rare in the absence of a serious chronic disease. Magnesium supplementation may be warranted in patients with inflammatory bowel disease, celiac disease, chemotherapy, severe diarrhea, malnutrition, or alcoholism. Dietary phytoestrogens, which are derived primarily from soy products and legumes (e.g., garbanzo beans [chickpeas] and lentils), exert some estrogenic activity but are insufficiently potent to justify their use in place of a pharmacologic agent in the treatment of osteoporosis.

Patients with hip fractures are often frail and relatively malnourished. Some data suggest an improved outcome in such patients when they are provided calorie and protein supplementation. Excessive protein intake can increase renal calcium excretion, but this can be corrected by an adequate calcium intake.

Exercise Exercise in young individuals prior to and at the time of puberty increases the likelihood that they will attain the maximal genetically determined peak bone mass. Meta-analyses of studies performed in postmenopausal women indicate that progressive high-intensity resistance training results in small gains in bone mass of 3–4%. However, this beneficial effect wanes if exercise is discontinued. Most of the studies are relatively short term, and a more substantial effect on bone mass is likely if exercise is continued over a long period. Exercise also has beneficial effects on neuromuscular function, and it improves coordination, balance, and strength, thereby reducing the risk of falling. A walking program is a practical way to start. Other activities such as dancing, racquet sports, cross-country skiing, and progressive high-intensity resistance training are also recommended, depending on the patient's personal preference and general condition. Even women who cannot walk benefit from swimming or water exercises, not so much for the effects on bone, which are quite minimal, but because of effects on muscle. Exercise habits should be consistent, optimally at least three times a week. For most patients, participation in exercise regimens that the patient enjoys is recommended in order to improve adherence. We also emphasize the importance of making exercise a social activity, again to improve adherence. Many individuals experience a fear of falling that can lead to social isolation and depression. Group exercise can help to alleviate this problem by providing a sense of social connectivity among participants.

Tai chi is a traditional Chinese martial art that utilizes a series of gentle, flowing movements to promote and maintain flexibility, balance, endurance, proprioception, and strength. It involves constant movement through all three spatial dimensions. As a three-dimensional exercise, tai chi may be incorporated as part of balance training program for individuals with osteoporosis. Tai chi is generally considered a safe activity. The evidence for fall prevention in tai chi has been evaluated in randomized controlled trials, and systematic reviews indicate tai chi can significantly reduce the risk of falls in older adults, particularly when it is practiced with increased frequency.

It is recommended that individuals with osteoporosis or osteoporotic vertebral fractures participate in exercise programs that involve both high-intensity progressive resistance and balance training. Slow, controlled movements are recommended in order to avoid injuries. Exercise modifications and avoidance of certain postures (such as spinal flexion) may be advised. Caution should be taken to avoid activities that can lead to potential fractures, such as performing activities on slippery surfaces, or twisting or bending the spine quickly while transitioning between different positions. Precautions should be taken to avoid injury when exercising with loads and performing exercises that challenge balance. For individuals with osteoporosis who have a high risk for fracture, vertebral fractures, sedentary lifestyle, or comorbid conditions that impact exercise tolerance, consultation with physical therapists to learn safe exercise programs is recommended.

PHARMACOLOGIC TREATMENT OF OSTEOPOROSIS

Several international osteoporosis management guidelines have been published. They recommend patients with fractures of the hip and spine should be evaluated for treatment. Patients with minimal trauma fractures in the setting of a BMD in the osteopenia or osteoporosis range should be treated with pharmacologic agents. Most guidelines also suggest that patients be considered for treatment when BMD T-score is ≤–2.5, a level consistent with the diagnosis of osteoporosis. Treatment should also be considered in postmenopausal women with a minimal trauma fracture or multiple risk factors even if BMD is not in the osteoporosis range. Treatment thresholds depend on cost-effectiveness analyses but, in the United States, are a >20% 10-year major fracture probability and >3% 10-year hip fracture probability. It must be emphasized, however, that as with other diseases, risk assessment is an inexact science when applied to the individual. As a result, patients often accept fracture risks that are higher than the physician might like out of concern for the usually considerably lower but better publicized risks of adverse events of drugs.

Pharmacologic therapies for osteoporosis are either antiresorptive or anabolic. The antiresorptive agents include medications that have broad effects such as hormone/estrogen therapy and selective estrogen receptor modulators (SERMS) as well as specific agents for osteoporosis treatment (bisphosphonates, denosumab, and calcitonin). The anabolic agents are teriparatide, abaloparatide, and romosozumab. Denosumab allows bone formation to continue, and thus, there is an increase in bone density beyond that occurring with agents that inhibit resorption directly leading to a reduction in bone formation like bisphosphonates.

Antiresorptive Agents • **Estrogens** A large body of clinical trial data indicates that various types of estrogens (conjugated equine estrogens, estradiol, estrone, esterified estrogens, ethinyl estradiol, and mestranol) reduce bone turnover, prevent bone loss, and induce small increases in bone mass of the spine, hip, and total body. The effects of estrogen are seen in women with natural or surgical menopause and in late postmenopausal women with or without established osteoporosis. Estrogens are efficacious when administered orally, transdermally, or by subcutaneous implant. For both oral and transdermal routes of administration, combined estrogen/progestin preparations are now available in many countries, obviating the problem of taking two tablets or using a patch and oral progestin.

For oral estrogens, the standard recommended doses have been 0.3 mg/d for esterified estrogens, 0.625 mg/d for conjugated equine estrogens, and 5 μg/d for ethinyl estradiol. For transdermal estrogen, the commonly used dose supplies 50 μg of estradiol per day, but a lower dose may be appropriate for some individuals. Dose-response data for conjugated equine estrogens indicate that lower doses (0.3 and 0.45 mg/d) are also effective. Doses even lower have also been shown to slow bone loss.

Fracture Data Epidemiologic databases indicate that women who take estrogen replacement have a 50% reduction, on average, of osteoporosis-related fractures, including hip fractures. The beneficial effect of estrogen is greatest among those who start replacement early and continue the treatment; the benefit declines after discontinuation to the extent that there is no residual protective effect against fracture by 10 years after discontinuation. The first clinical trial evaluating fractures as secondary outcomes, the Heart and Estrogen-Progestin Replacement Study (HERS) trial, showed no effect of hormone therapy on hip or other clinical fractures in women with established coronary artery disease. These data made the results of the Women's Health Initiative (WHI) exceedingly important **(Chap. 395)**. The estrogen-progestin arm of the WHI in >16,000 postmenopausal healthy women not selected based on low bone mass indicated that hormone therapy reduces the risk of hip and clinical spine fracture by 34% and that of all clinical fractures by 24%.

A few smaller clinical trials have evaluated spine fracture occurrence as an outcome with estrogen therapy. They have consistently shown that estrogen treatment reduces vertebral fracture incidence.

The WHI has provided a vast amount of data on the multisystemic effects of hormone therapy. Although earlier observational studies suggested that estrogen replacement might reduce heart disease, the WHI showed that combined estrogen-progestin treatment increased risk of fatal and nonfatal myocardial infarction by ~29%, confirming data from the HERS study. Other important relative risks included a 40% increase in stroke, a 100% increase in venous thromboembolic disease, and a 26% increase in risk of breast cancer. Subsequent analyses have confirmed the increased risk of stroke and, in a substudy, showed a twofold increase in dementia. Benefits other than the fracture reductions noted above included a 37% reduction in the risk of colon cancer. These relative risks must be interpreted in light of absolute risk **(Fig. 423-8)**. For example, out of 10,000 women treated with estrogen-progestin for 1 year, there will be 8 excess heart attacks, 8 excess breast cancers, 18 excess venous thromboembolic events, 5 fewer hip fractures, 44 fewer clinical fractures, and 6 fewer colorectal cancers. These numbers must be multiplied by the number of years of hormone treatment. There was no effect of combined hormone treatment on the risk of uterine cancer or total mortality.

It is important to note that these WHI findings apply specifically to hormone treatment in the form of conjugated equine estrogen plus medroxyprogesterone acetate. The relative benefits and risks of unopposed estrogen in women who had hysterectomies vary somewhat. They still show benefits against fracture occurrence

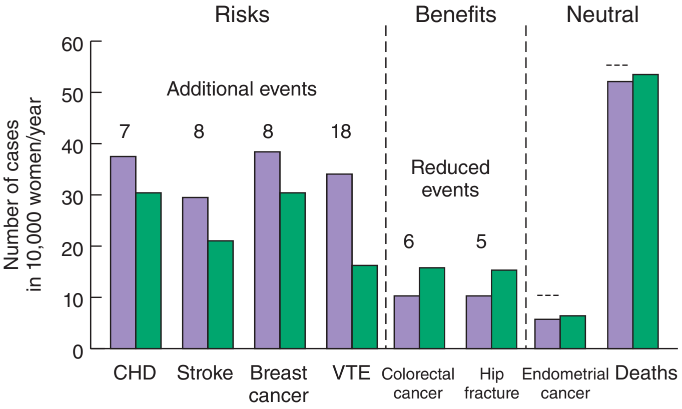

FIGURE 423-8 Effects of hormone therapy on event rates: green, placebo; purple, estrogen and progestin. CHD, coronary heart disease; VTE, venous thromboembolic events. *(Adapted from Women's Health Initiative. WHI HRT Update.)*

and increased risk of venous thrombosis and stroke, similar in magnitude to the risks for combined hormone therapy. In contrast, though, the estrogen-only arm of WHI indicated no increased risk of heart attack and a decreased risk of breast cancer. The data suggest that at least some of the detrimental effects of combined therapy are related to the progestin component. In addition, there is the possibility, suggested by primate data, that the risk accrues mainly to women who have some years of estrogen deficiency before initiating treatment. Therefore, the benefit/risk ratio of hormone therapy is improved if it is commenced before the age of 60 years or within 10 years of menopause. Nonetheless, there has been marked reluctance among women for estrogen therapy/hormone therapy, and the U.S. Preventive Services Task Force has specifically suggested that estrogen therapy/hormone therapy not be used for disease prevention.

Mode of Action Two subtypes of ERs, α and β, have been identified in bone and other tissues. Cells of monocyte lineage express both ERα and ERβ, as do osteoblasts. Estrogen-mediated effects vary with the receptor type. Using ER knockout mouse models, elimination of ERα produces a modest reduction in bone mass, whereas mutation of ERβ has less of an effect on bone. A man with a homozygous mutation of ERα had markedly decreased bone density as well as abnormalities in epiphyseal closure, confirming the important role of ERα in bone biology. The mechanism of estrogen action in bone is an area of active investigation (Fig. 423-5). Although data are conflicting, estrogens may inhibit osteoclasts directly. However, most estrogen (and androgen) effects on bone resorption are mediated indirectly through paracrine factors produced by osteoblasts. These actions include (1) increasing the inhibitor of RANK-L, osteoprotegerin, production by osteoblasts, (2) increasing IGF-I and TGF-β, and (3) suppressing IL-1 (α and β), IL-6, TNF-α, and osteocalcin synthesis. These indirect estrogen actions primarily decrease bone resorption.

Progestins In women with a uterus, daily progestin or cyclical progestins at least 12 days per month are prescribed in combination with estrogens to reduce the risk of uterine cancer. Medroxyprogesterone acetate and norethindrone acetate blunt the high-density lipoprotein response to estrogen, but micronized progesterone does not. Neither medroxyprogesterone acetate nor micronized progesterone appears to have an independent effect on bone; at lower doses of estrogen, norethindrone acetate may have an additive benefit. In breast tissue, progestins may account for the increased risk of breast cancer with combination treatment.

SERMs Two SERMs are used currently in postmenopausal women: raloxifene, which is approved by the FDA for the prevention and treatment of osteoporosis as well as the prevention of breast cancer, and tamoxifen, which is approved for the prevention and treatment of breast cancer. A third SERM, bazedoxifene, is marketed in combination with conjugated estrogen for treatment of menopausal symptoms and prevention of bone loss. Bazedoxifene protects the uterus and breast from effects of estrogen and makes the use of progestin unnecessary.

Tamoxifen reduces bone turnover and bone loss in postmenopausal women compared with placebo groups. These findings support the concept that tamoxifen acts as an estrogenic agent in bone. There are limited data on the effect of tamoxifen on fracture risk, but the Breast Cancer Prevention study indicated a possible reduction in clinical vertebral, hip, and Colles' fractures. Tamoxifen is not FDA approved for prevention or treatment of osteoporosis. The major benefit of tamoxifen is on breast cancer occurrence and recurrence in women with ER-positive tumors. The breast cancer prevention trial indicated that tamoxifen administration over 4–5 years reduced the incidence of new invasive and noninvasive breast cancer by ~45% in women at increased risk of breast cancer. The incidence of ER-positive breast cancers was reduced by 65%. Tamoxifen increases the risk of uterine cancer in postmenopausal women, limiting its use for breast cancer prevention in women at low or moderate risk.

Raloxifene (60 mg/d) has effects on bone turnover and bone mass that are very similar to those of tamoxifen, indicating that this agent is also estrogenic on the skeleton. The effect of raloxifene on bone density (+1.4–2.8% vs. placebo in the spine, hip, and total body) is somewhat less than that seen with standard doses of estrogens. Raloxifene reduces the occurrence of vertebral fracture by 30–50%, depending on the population; however, there are no data confirming that raloxifene can reduce the risk of nonvertebral fractures after 8 years of observation.

Raloxifene, like tamoxifen and estrogen, has effects on other organ systems. The most beneficial effect appears to be a reduction in invasive breast cancer (mainly decreased ER-positive) occurrence of ~65% in women who take raloxifene compared with placebo. In a head-to-head study, raloxifene was as effective as tamoxifen in preventing breast cancer in high-risk women, and raloxifene is FDA approved for this indication. In a further study, raloxifene had no effect on heart disease in women with increased risk for this outcome. In contrast to tamoxifen, raloxifene is not associated with an increase in the risk of uterine cancer or benign uterine disease. Raloxifene increases the occurrence of hot flashes but reduces serum total and low-density lipoprotein cholesterol, lipoprotein(a), and fibrinogen. Raloxifene, with its positive effects on breast cancer and vertebral fractures, has become a useful agent for the treatment of the younger asymptomatic postmenopausal woman. In some women, a recurrence of menopausal symptoms may occur. Usually this is evanescent but occasionally is sufficiently impactful on daily life and sleep that the drug must be withdrawn. Raloxifene increases the risk of deep-vein thrombosis and may increase the risk of death from stroke among older women. Consequently, it is not usually recommended for women over age 70 years.

MODE OF ACTION OF SERMS

All SERMs bind to the ER, but each agent produces a unique receptor-drug conformation. As a result, specific coactivator proteins or co-repressor proteins are bound to the receptor **(Chap. 389)**, resulting in differential effects on gene transcription that vary depending on other transcription factors present in the cell. Another aspect of selectivity is the affinity of each SERM for the different ERα and ERβ subtypes, which are expressed differentially in various tissues. These tissue-selective effects of SERMs offer the possibility of tailoring estrogen therapy to best meet the needs and risk factor profile of an individual patient.

Bisphosphonates Bisphosphonates have become the mainstay of osteoporosis treatment globally, in part related to cost as they have become generic. Alendronate, risedronate, ibandronate, and zoledronic acid are approved for the prevention and treatment of postmenopausal osteoporosis. Alendronate, risedronate, and zoledronic acid are also approved for the treatment of glucocorticoid-induced osteoporosis, and risedronate and zoledronic acid are approved for prevention of glucocorticoid-induced osteoporosis. Alendronate, risedronate, and zoledronic acid are also approved for treatment of osteoporosis in men.

Alendronate decreases bone turnover and increases bone mass by up to 8 and 6% versus placebo in the spine and hip, respectively. Multiple trials have evaluated its effect on fracture occurrence. The Fracture Intervention Trial provided evidence in >2000 women with prevalent vertebral fractures that daily alendronate treatment (5 mg/d for 2 years and 10 mg/d for 9 months afterward) reduces vertebral fracture risk by ~50%, multiple vertebral fractures by up to 90%, and hip fractures by up to 50%. Several subsequent trials have confirmed these findings **(Fig. 423-9)**. For example, in a study of >1900 women with low bone mass treated with alendronate (10 mg/d) versus placebo, the incidence of all nonvertebral fractures was reduced by ~47% after only 1 year. In the United States, the 70-mg weekly dose is approved for treatment of osteoporosis and the dose of 35 mg per week is approved for prevention, with those doses showing equivalence to daily dosing based on bone turnover and bone mass response.

Consequently, once-weekly therapy generally is preferred because of lower incidence of gastrointestinal side effects, ease of administration, and improved persistence with therapy. Alendronate should be taken with a full glass of water before breakfast after an overnight fast, as bisphosphonates are poorly absorbed. Because of the potential for esophageal irritation, alendronate is contraindicated in patients who have stricture, achalasia, or inadequate emptying of the esophagus. It is recommended that patients remain upright (standing or sitting) for at least 30 min after taking the medication to avoid esophageal irritation and that food and fluids (other than water) be avoided for the same duration. In clinical trials, overall gastrointestinal symptomatology was no different with alendronate than with placebo, but in practice, all oral bisphosphonates have been associated with esophageal irritation and inflammation.

Risedronate also reduces bone turnover and increases bone mass. Controlled clinical trials have demonstrated 40–50% reduction in vertebral fracture risk over 3 years, accompanied by a 40% reduction in clinical nonspine fractures. The only clinical trial specifically designed to evaluate hip fracture outcome (HIP) indicated that risedronate reduced hip fracture risk in women in their seventies with confirmed osteoporosis by 40%. In contrast, risedronate was not effective at reducing hip fracture occurrence in older women (80+ years) without proven osteoporosis. Studies have shown that 35 mg of risedronate administered once weekly is therapeutically equivalent to 5 mg/d. The instructions for oral administration noted for alendronate apply to all three oral bisphosphonates. There is also a preparation of risedronate with an enteric coating (35 mg) that can be taken after breakfast. Risedronate is the only bisphosphonate that has this dosing flexibility.

Ibandronate is the third amino-bisphosphonate approved in the United States. Ibandronate (2.5 mg/d) has been shown in clinical trials to reduce vertebral fracture risk by ~40% but with no overall effect on nonvertebral fractures. In a post hoc analysis of subjects with a femoral neck T-score of ≤–3, ibandronate reduced the risk of nonvertebral fractures by ~60%. In clinical trials, ibandronate doses of 150 mg per month PO or 3 mg every 3 months IV had greater effects on turnover and bone mass than did 2.5 mg/d. Patients should take oral ibandronate in the same way as other bisphosphonates but with 1 h elapsing before other food or drink (other than plain water).

Zoledronic acid is a potent bisphosphonate with a unique administration regimen (5 mg by 30-min IV infusion at most annually). Zoledronic acid data confirm that it is highly effective in fracture risk reduction. In a study of >7000 women followed for 3 years, zoledronic acid (5 mg IV annually) reduced the risk of vertebral fractures by 70%, nonvertebral fractures by 25%, and hip fractures by 40%. These results were associated with less height loss and disability. In the treated population, there was an increased risk of almost 25% of an acute phase response (APR) in patients with no prior bisphosphonate exposure (fever, myalgias, headache, malaise), but effects were short-lived (2–3 days). A recent study shows the APR can be eliminated by the administration of dexamethasone (4 mg) for 3 days commencing 90 min before the infusion. Although there was also an increased risk of atrial fibrillation in this trial, detailed evaluation of all bisphosphonates has failed to confirm a risk of atrial fibrillation. Zoledronic acid has also been studied in a placebo-controlled trial of older women and men within 3 months of an acute hip fracture. The risk of recurrent fracture was reduced by 35%, and there was a 28% reduction in mortality that was associated with a reduction in infections and cardiovascular events. In older postmenopausal women with osteopenia, the risk of nonvertebral, vertebral, or clinical fragility fractures was significantly lower in women with osteopenia who received zoledronate every 18 months for 6 years than in women who received placebo. In an extension study of the women who had received zoledronic acid infusions for 6 years, the risk of nonvertebral fractures remained low for up to 3.5 years after the last dose but increased after that time. This suggests less frequent dosing may be required after an initial course of three or four zoledronic acid infusions.

Common Bisphosphonate Adverse Events All bisphosphonates have been associated with some musculoskeletal and joint pains of unclear etiology, which are occasionally severe. There is potential for renal toxicity, particularly in women with stage 3 chronic kidney disease (CKD), and bisphosphonates are contraindicated in those with an estimated glomerular filtration rate <30–35 mL/min. Hypocalcemia can occur.

Two potential side effects have been associated with bisphosphonate use. The first is medication-related osteonecrosis of the jaw (MRONJ). MRONJ usually follows a dental procedure in which bone is exposed (dental extractions and implants). It is presumed that the exposed bone becomes infected and dies. MRONJ is more common among cancer patients receiving high doses of bisphosphonates for skeletal metastases. It is rare among persons with osteoporosis on usual doses of bisphosphonates. Additional risk factors for MRONJ are poor dental hygiene, diabetes, and glucocorticoids. Oral antibiotic rinses and oral systemic antibiotics may be useful to prevent this rare adverse event if risk is perceived to be particularly high. In addition, any invasive dental procedures should be performed prior to commencement of bisphosphonate therapy. Because bisphosphonates have a long skeletal half-life, it is unlikely that withholding therapy prior to an invasive procedure will reduce MRONJ risk; however, this is often recommended by dentists and faciomaxillary surgeons. Teriparatide treatment for 2 months has been shown to reduce bone lesions associated with MRONJ versus placebo. The second side effect is called atypical femoral fracture. These are stress fractures occurring in the subtrochanteric femoral region or across the femoral shaft distal to the lesser trochanter. They are often preceded by pain in the lateral thigh or groin that can be present for weeks, months, or even years before the fracture. The fractures occur following either no or trivial trauma, are predominantly horizontal with a medial beak, and are noncomminuted. An American Society for Bone and Mineral Research Task Force described the major and minor criteria for these fractures, which are related to duration of bisphosphonate therapy and increase after 5–7 years of treatment, but only after 3 years of exposure in Asian patients. The overall risk appears quite low, especially when compared with the number of hip fractures prevented by these therapies, but they often require surgical fixation and show delayed healing. The risk is also increased with Asian ethnicity, rheumatoid arthritis, glucocorticoids, proton pump inhibitors, and SSRIs. In some cases (10–30%), there is no bisphosphonate exposure, and in some of these cases, there may be an underlying rare monogenetic bone disease such as osteogenesis imperfecta, hypophosphatasia, X-linked hypophosphatemic rickets, X-linked osteoporosis, or pycnodysostosis. Other genetic risk factors for atypical femur fractures have been identified in families with bisphosphonate-associated atypical femur fractures. If the fractures are found early, when there is evidence of periosteal stress reaction or stress fracture, prior to the occurrence of overt fracture, surgical intervention may be

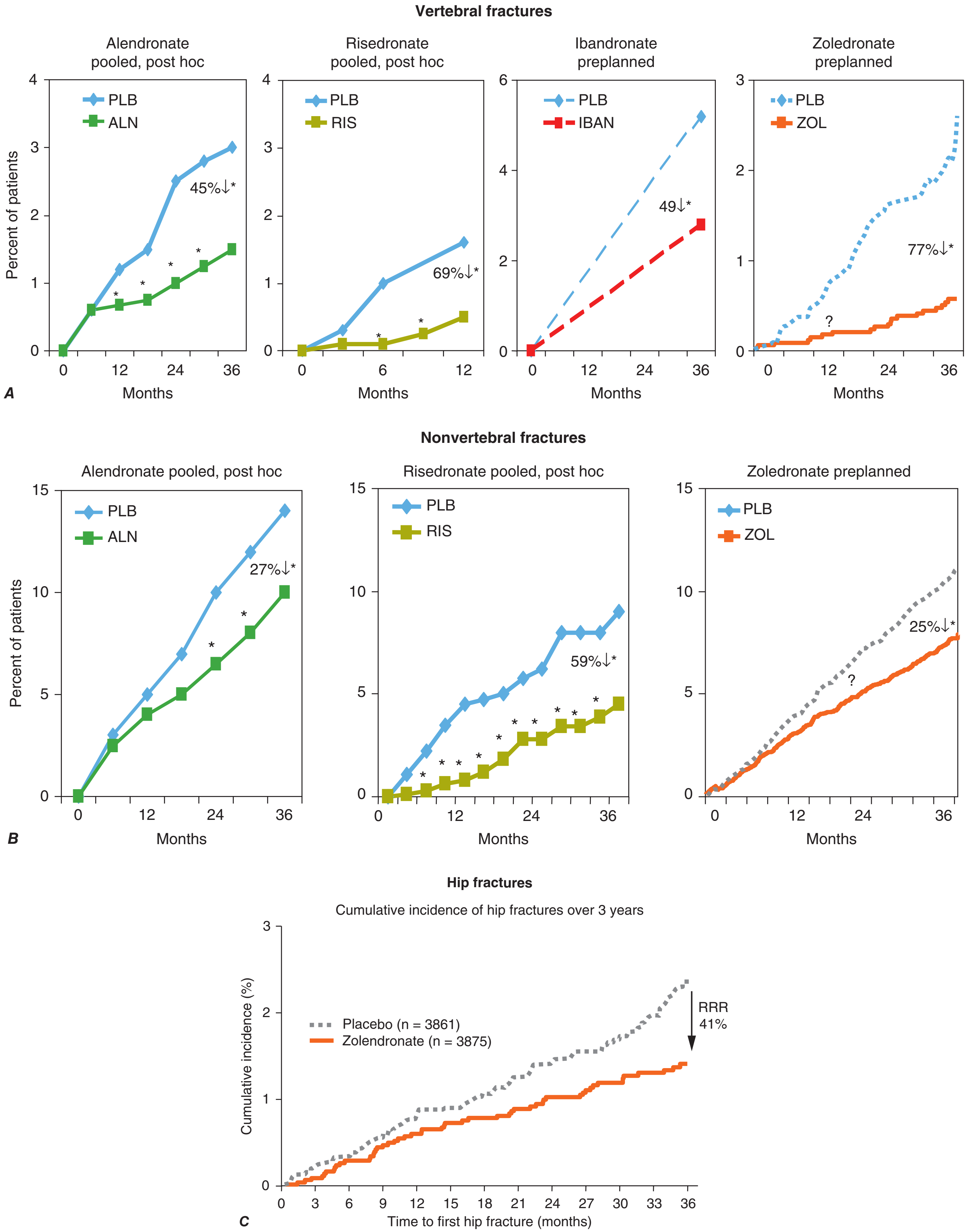

FIGURE 423-9 Effects of various bisphosphonates on fractures. ***A.*** Clinical vertebral fractures. ***B.*** Nonvertebral fractures. ***C.*** Hip fractures. Plb, placebo; RRR, relative risk reduction. *(Data from DM Black et al: J Clin Endocrinol Metab 85:4238, 2000; C Roux et al: Curr Med Res Opin 4:433, 2004; CH Chesnut et al: J Bone Miner Res 19: 1241, 2004; DM Black et al: N Engl J Med 356:1809, 2007; JT Harrington et al: Calcif Tissue Int 74:129, 2003.)*

required. The evidence that teriparatide can help heal the fracture and preclude the need for surgical repair from case series is mixed, and published trial data are not available. When patients initiate bisphosphonates, they should be informed that if they develop thigh or groin pain, they should inform their health care professional. Routine x-rays will sometimes detect cortical thickening or even a stress fracture, but MRI or technetium bone scans are more sensitive. The presence of an abnormality requires, at minimum, a period of modified weight-bearing and may need prophylactic rodding of the femur. It is important to realize that these may often be bilateral (~50% of the time), and when an abnormality is found, the other femur should be checked. It is unknown whether patients who have these atypical femur fractures can ever receive antiresorptive therapies again in the future, but it seems prudent to avoid their use for most of these individuals.

Mode of Action Bisphosphonates are structurally related to pyrophosphates, compounds that are incorporated into bone matrix. Bisphosphonates are taken up at sites of active bone remodeling and specifically impair osteoclast function and reduce osteoclast number, in part by inducing apoptosis. Recent evidence suggests that the nitrogen-containing bisphosphonates also inhibit protein prenylation, one of the end products in the mevalonic acid pathway, by inhibiting the enzyme farnesyl pyrophosphate synthase. This effect disrupts intracellular protein trafficking and ultimately may lead to apoptosis. Some bisphosphonates have very long retention in the skeleton and may exert long-term effects via recirculation of bisphosphonates retained in bone when bone remodeling recurs at that site.

Calcitonin Calcitonin is a polypeptide hormone produced in the thyroid gland **(Chap. 422)**. Its physiologic role is unclear as no skeletal disease has been described in association with calcitonin deficiency or excess. Calcitonin preparations are approved by the FDA for Paget's disease, hypercalcemia, and osteoporosis in women >5 years past menopause.

Injectable calcitonin produces small increments in bone mass of the lumbar spine. However, difficulty of administration and frequent reactions, including nausea and facial flushing, make general use limited. A nasal spray containing calcitonin (200 IU/d) is available for treatment of osteoporosis in postmenopausal women. One study suggests that nasal calcitonin produces small increments in bone mass and a small reduction in new vertebral fractures in calcitonin-treated patients (at one dose) versus those on calcium alone. There has been no proven effectiveness against nonvertebral fractures. Calcitonin is not indicated for prevention of osteoporosis and is not sufficiently potent to prevent bone loss in early postmenopausal women. Calcitonin might have an analgesic effect on bone pain, both in the subcutaneous and possibly in the nasal form. Concerns have been raised about an increase in the incidence of cancer associated with calcitonin use. Initially, the cancer noted was of the prostate, but an analysis of all data suggested a more general increase in cancer risk. In Europe, the European Medicines Agency has removed the osteoporosis indication, and an FDA Advisory Committee has voted for a similar change in the United States.

Mode of Action Calcitonin suppresses osteoclast activity by direct action on the osteoclast calcitonin receptor. Osteoclasts exposed to calcitonin cannot maintain their active ruffled border, which normally maintains close contact with underlying bone.

Denosumab Denosumab is a human monoclonal antibody that inhibits RANKL, a potent stimulus of osteoclast activity. It is administered twice yearly by subcutaneous injection. In a randomized controlled trial in postmenopausal women with osteoporosis, it has been shown to increase BMD at the spine, hip, and forearm and reduce vertebral, hip, and nonvertebral fractures over a 3-year period by 68, 40, and 20%, respectively **(Fig. 423-10)**. Unlike bisphosphonates in which BMD plateaus after 4–5 years of treatment, bone density continues to increase as long as denosumab treatment is continued. In a long-term extension of the pivotal fracture trial, BMD increased by 21.7 and 9.2% at the spine and hip, respectively, after 10 years of denosumab treatment. Over the 10 years, fracture rates also remained at least as low as those seen with denosumab during the active placebo-controlled portion of the trial. Other clinical trials indicate ability to increase bone mass in postmenopausal women with osteopenia and in postmenopausal women with breast cancer treated with aromatase inhibitors. In the oncology literature, denosumab reduces the risk of fractures in women on aromatase inhibitors. In a study of men with prostate cancer treated with androgen deprivation therapy, denosumab increased bone mass and reduced vertebral fracture incidence over 3 years. An analysis of five placebo-controlled studies has also

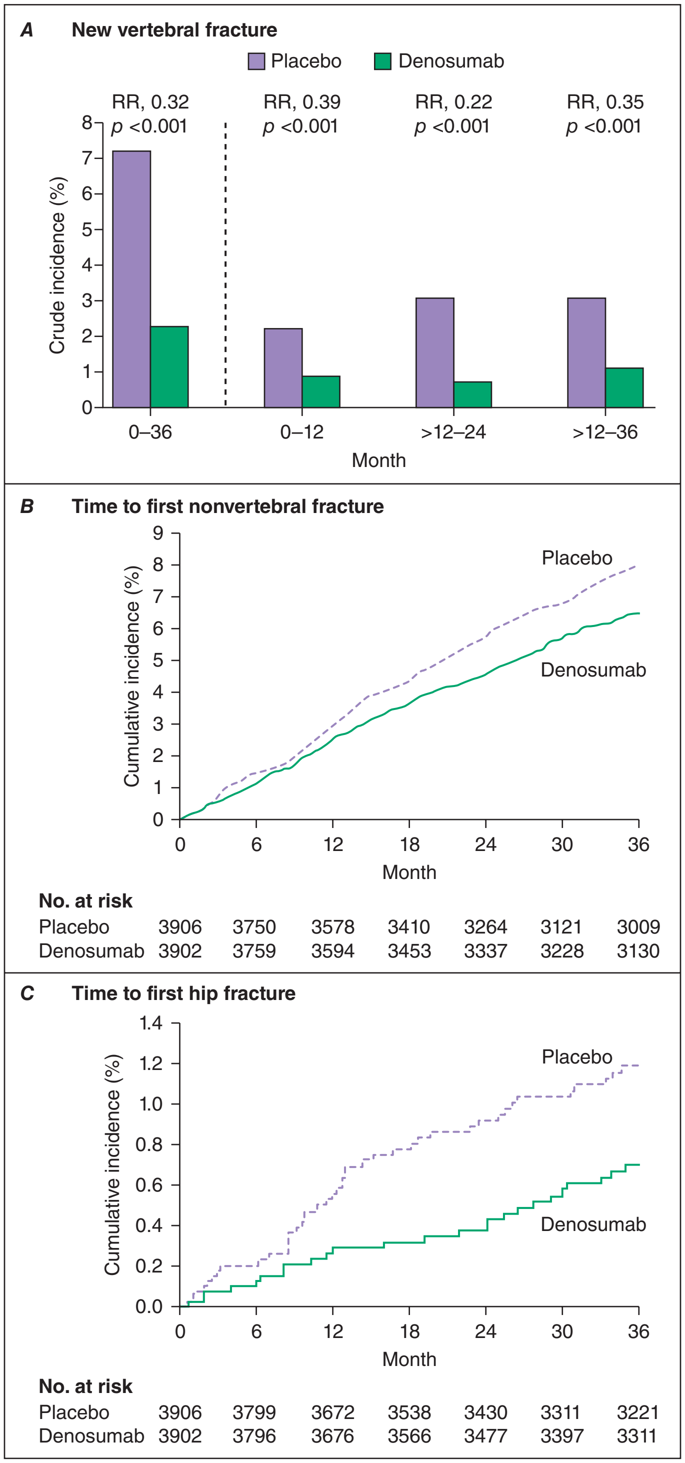

FIGURE 423-10 Effects of denosumab on the following: *A.* new vertebral fractures; and ***B.*** and ***C.*** times to nonvertebral and hip fracture. RR, relative risk. *(From SR Cummings et al: Denosumab for prevention of fractures in postmenopausal women with osteoporosis. N Engl J Med 361:756, 2009. Copyright © 2009 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

suggested reduced risk of falls in patients with osteoporosis treated with denosumab.

Denosumab was approved by the FDA in 2010 for the treatment of postmenopausal women who have a high risk for osteoporotic fractures, including those with a history of fracture or multiple risk factors for fracture, and those who have failed or are intolerant to other osteoporosis therapy. Denosumab is also approved for the treatment of osteoporosis in men at high risk for fracture, women with breast cancer on aromatase inhibitors, and men with prostate cancer on androgen deprivation treatment.

Denosumab may increase the risk of MRONJ and atypical femur fractures similarly to bisphosphonates. Estimated incidence is 5/10,000 patient-years for MRONJ and 1/10,000 patient-years for atypical femur fractures. Denosumab can cause hypersensitivity reactions, hypocalcemia, and skin reactions including dermatitis, rash, and eczema. Early concerns about an imbalance in infections with denosumab have largely been allayed. Severe hypocalcemia may occur in patients with CKD and an estimated glomerular filtration rate <30 mL/min, particularly in those on hemodialysis. Serum calcium levels should be monitored if denosumab is used in such patients, and active vitamin D analogues and calcium may be required for its treatment.

Unlike bisphosphonates, when denosumab is discontinued, there is a rebound increase in bone turnover and an acceleration of bone loss. This likely reflects the maturation of accumulated osteoclast precursors (osteomorphs) in bone marrow when the drug was administered that then coalesce to become mature bone resorbing cells once denosumab is withdrawn. Other mechanisms involving reduced osteoprotegerin levels and osteocytes may also be involved. The consequences of this rebound increase in remodeling-associated bone loss are trabecular perforation and a rapid increase in the risk of fracture, particularly vertebral fracture, with a specific increase in multiple vertebral fractures. In patients who need to stop denosumab or in patients in whom BMD and fracture risk reduction goals have been met, temporary use of bisphosphonate treatment may prevent the rebound increase in remodeling and rapid bone loss. In clinical practice, a single infusion of zoledronic acid at the time of the missed denosumab dose seems to maintain BMD for 1–2 years but may need to be repeated after 3 months, or more often, if bone turnover markers remain elevated. Oral bisphosphonates can also be used for 12–24 months to maintain bone density. In both cases, the required duration of bisphosphonate use to eliminate the rebound effect is not clear and may vary considerably among patients so bone density should continue to be monitored.

Mode of Action Denosumab is a fully human monoclonal antibody to RANKL, the final common effector of osteoclast formation, activity, and survival. Denosumab binds to RANKL, inhibiting its ability to initiate formation of mature osteoclasts from osteoclast precursors and to bring mature osteoclasts to the bone surface and initiate bone resorption. Denosumab also plays a role in reducing osteoclast survival. Through these actions on the osteoclast, denosumab induces a potent and rapid antiresorptive action, as assessed biochemically and histomorphometrically.

Anabolic Agents • **Parathyroid Hormone** Endogenous PTH is an 84-amino-acid peptide that is largely responsible for calcium homeostasis **(Chap. 422)**. Although chronic elevation of PTH, as occurs in hyperparathyroidism, is associated with bone loss (particularly at cortical bone sites such as the femoral neck and radius), PTH also can exert anabolic effects on bone if administered intermittently. Consistent with this, some observational studies have indicated that mild endogenous hyperparathyroidism is associated with maintenance of trabecular bone mass but loss of cortical bone. Based on these findings, early small-scale observational studies showed that PTH analogues could augment trabecular BMD. Subsequent controlled clinical trials have confirmed that PTH can increase bone mass and reduce fracture incidence. The first randomized controlled trial in postmenopausal women showed that PTH (1–34) (teriparatide), when superimposed on ongoing estrogen therapy, produced substantial increments in bone mass (13% over a 3-year period compared with estrogen alone) and reduced the risk of vertebral compression fractures. In the pivotal study (median, 19 months' duration), 20 μg PTH (1–34) daily by subcutaneous injection (with no additional therapy) reduced vertebral fractures by 65% and nonvertebral fractures by 53% **(Fig. 423-11)**. Teriparatide produces rapid and robust increases in bone formation and then bone remodeling overall, resulting in substantial increases in bone mass and improvements in microarchitecture, including cancellous connectivity and cortical width. The BMD effects, particularly in the hip, are lower when patients switch from bisphosphonates to teriparatide, possibly in proportion to the potency of the antiresorptive agent. The hip BMD effect is particularly impaired and may result in transient bone loss, when patients switch from

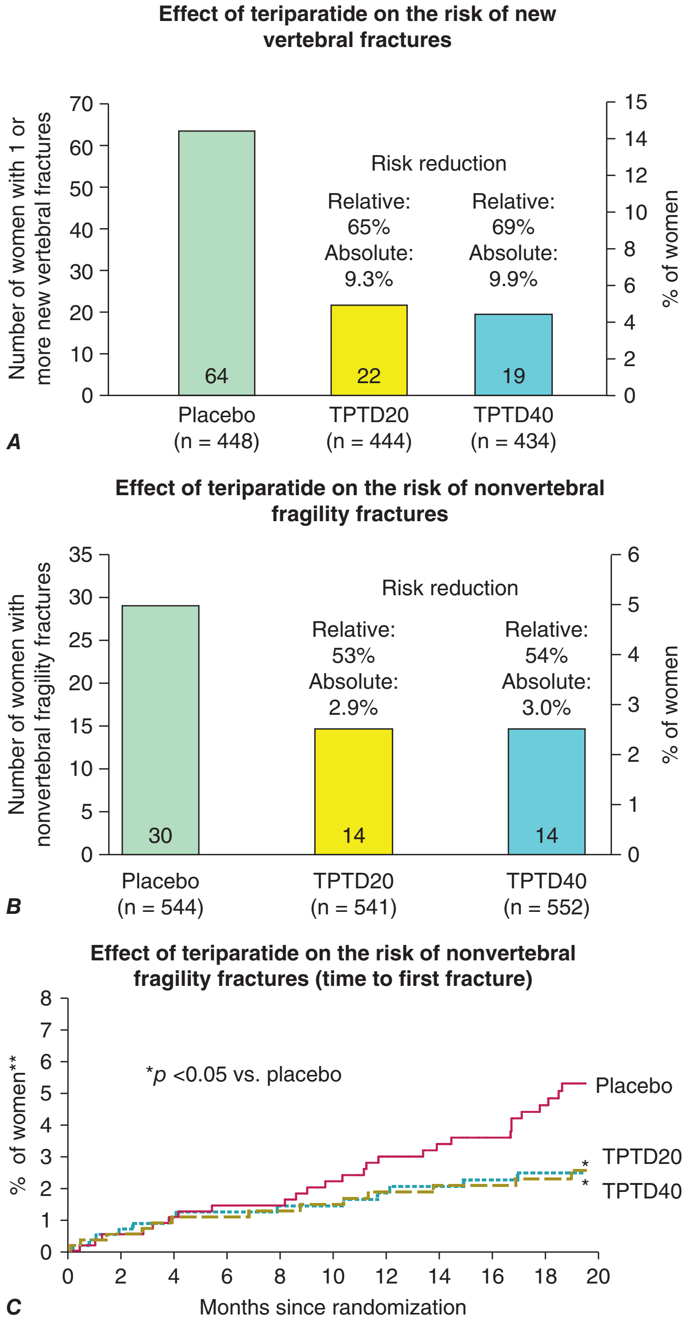

FIGURE 423-11 Effects of teriparatide (TPT) on the following: *A.* new vertebral fractures; and ***B.*** and ***C.*** nonvertebral fragility fractures. *(A and B are data from RM Neer et al: Effect of parathyroid hormone (1–34) on fractures and bone mineral density in postmenopausal women with osteoporosis. N Engl J Med May 344:1434, 2001. C From New England Journal of Medicine RM Neer et al: Effect of parathyroid hormone (1-34) on fractures and bone mineral density in postmenopausal women with osteoporosis, 344:1434-1441. Copyright © 2001 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

denosumab to teriparatide. In patients on denosumab who need teriparatide treatment, there may be a role for combination therapy. In previously untreated women, teriparatide is administered as monotherapy and followed by a potent antiresorptive agent such as denosumab or an oral or intravenous bisphosphonate. Combination therapy is generally avoided because of cost and potential inhibition of the anabolic activity of teriparatide.

Comparison with Antiresorptive Therapy In women with painful acute osteoporotic vertebral fractures, teriparatide reduced subsequent vertebral fractures by ~50% compared with risedronate. There was no difference in nonvertebral fracture outcome between the two medications. A study comparing teriparatide with risedronate in patients with prevalent vertebral fractures showed significant benefit for teriparatide against vertebral fractures **(Fig. 423-12)** and clinical fractures with a trend for a benefit of teriparatide against nonvertebral fractures.

Side effects of teriparatide are generally mild and can include muscle pain, weakness, dizziness, headache, and nausea. These tend to remit with continuing treatment but may be mitigated by giving injections at night. Rodents given prolonged treatment with PTH in high doses (3–60 times the human dose) developed osteogenic sarcomas after ~18 months of treatment. Rare cases of osteosarcoma have been described in patients treated with teriparatide consistent with the background incidence of osteosarcoma in adults. Long-term surveillance studies of a high proportion of patients diagnosed with osteosarcoma as adults in both the United States and Scandinavia reveal no prior exposure to teriparatide in any of the cases.

Teriparatide use may be limited by cost and its mode of administration (daily subcutaneous injection). Alternative modes of delivery have been investigated, but none have proven successful. Because of the rodent osteosarcoma data and the maximum duration of teriparatide in the pivotal trial of 2 years, the FDA previously limited teriparatide treatment to 2 years in a lifetime. However, registry data confirming no increased risk of osteogenic sarcoma in humans treated with teriparatide have allowed the FDA to drop this restriction on duration of use. As a result, consideration may be given to using intermittent courses of teriparatide therapy. In many other countries, however, teriparatide use is still restricted to 2 years in a lifetime.

Mode of Action Exogenously administered PTH appears to have direct actions on osteoblast activity, with biochemical and histomorphometric evidence of de novo bone formation within a week or two in response to teriparatide. There is subsequently resorption.

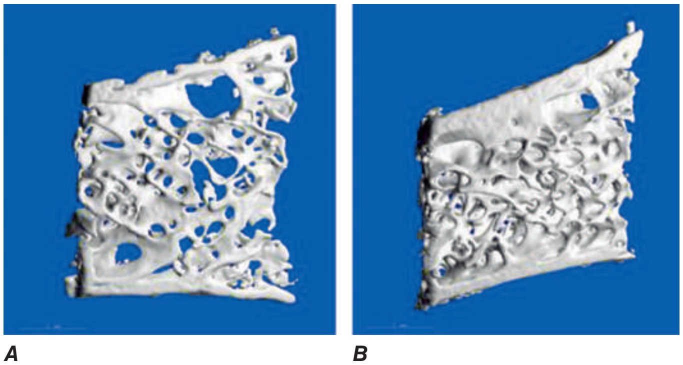

FIGURE 423-13 Effect of parathyroid hormone (PTH) treatment on bone microarchitecture. Paired biopsy specimens from a 64-year-old woman before (***A***) and after (***B***) treatment with PTH. *(Reproduced with permission from DW Dempster et al: Effects of daily treatment with parathyroid hormone on bone microarchitecture and turnover in patients with osteoporosis: A paired biopsy study. J Bone Miner Res 16:1846, 2001.)*

Subsequently, teriparatide primarily activates remodeling-based bone formation, favoring bone formation over bone resorption. Teriparatide given by daily injection stimulates osteoblast recruitment and activity through activation of Wnt signaling. Teriparatide produces a true increase in bone tissue and an apparent restoration of bone microarchitecture **(Fig. 423-13)**.

Abaloparatide Abaloparatide is a synthetic analogue of human PTHrP, which has significant homology to PTH and binds the PTH type 1 receptor. Abaloparatide and teriparatide exert different binding affinities to the two different receptor conformations, R^0 and RG. Compared with teriparatide, abaloparatide binds with similar high affinity to the RG conformation but with much lesser affinity to the R^0 conformation. These differences appear to result in a similar bone formation stimulus but lesser bone resorption stimulus, and abaloparatide was specifically chosen for development among many PTH and PTHrP analogues for what appeared to be an optimized anabolic profile.

In the phase 3 Abaloparatide Comparator Trial in Vertebral Endpoints (ACTIVE) study, 2463 postmenopausal women with osteoporosis were randomized to blinded daily subcutaneous abaloparatide versus placebo or open-label teriparatide. At 18 months, spine BMD increase was similar with abaloparatide and teriparatide (11.2% with abaloparatide and 10.5% with teriparatide); in the total hip, BMD increments were slightly larger with abaloparatide

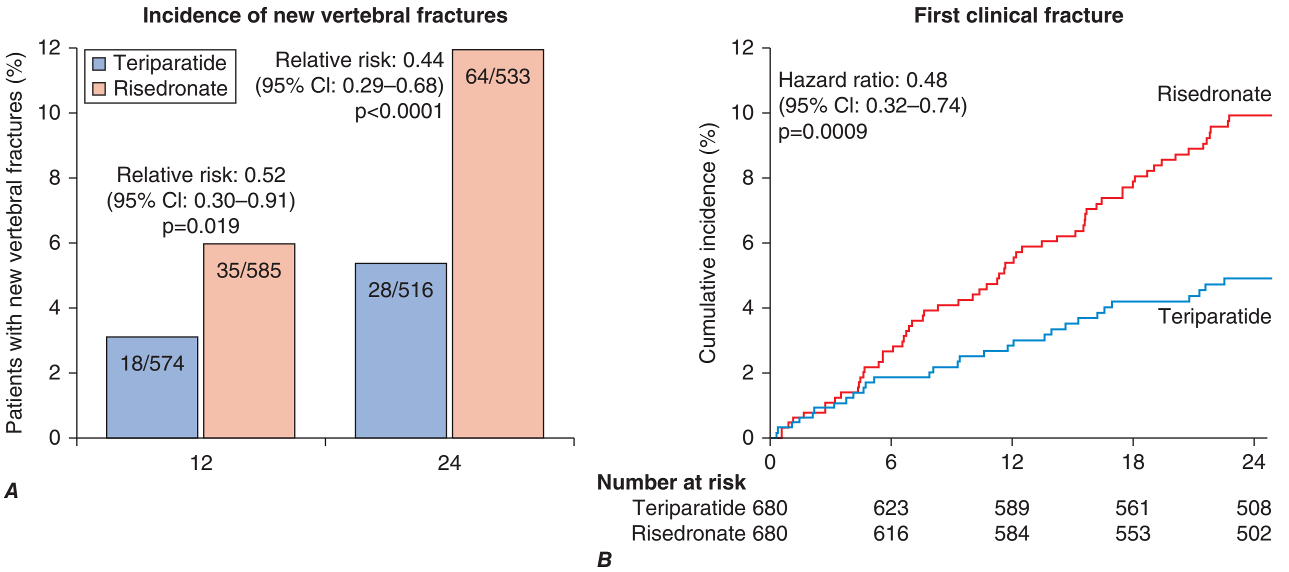

FIGURE 423-12 Effect of parathyroid hormone (PTH) treatment on vertebral fracture incidence at 12 and 24 months (***A***) and clinical fractures (***B***) compared with risedronate. *(Reproduced with permission from D Kendler et al: Effects of teriparatide and risedronate on new fractures in post-menopausal women with severe osteoporosis (VERO): A multicentre, double-blind, double-dummy, randomised controlled trial. Lancet 391:230, 2018.)*

(4.2 vs 3.3%). New vertebral fracture incidence was reduced by 86% with abaloparatide and 80% with teriparatide compared with placebo (both *p* <.001). The hazard ratio for abaloparatide versus teriparatide was not quoted. Nonvertebral fractures were reduced by 43% with abaloparatide (*p* = .05) and by 28% with teriparatide (not significant; *p* = .22). The ACTIVE study was extended, with 92% of eligible participants from the abaloparatide and placebo arms transitioned to open-label alendronate for a total treatment period of 24 months of alendronate. Both vertebral and nonvertebral fractures were less common in the group who transitioned from abaloparatide to alendronate, suggesting that the fracture benefit of abaloparatide can be maintained with antiresorptive treatment.

Romosozumab Romosozumab is a humanized antibody that blocks the osteocyte production of sclerostin, resulting in a unique action, with an increase in bone formation and decrease in bone resorption. In the pivotal trial (FRAME), 7180 postmenopausal women with osteoporosis were randomized to receive blinded monthly subcutaneous romosozumab (210 mg) or placebo for 1 year followed by transition to open-label subcutaneous denosumab (60 mg) every 6 months for an additional year. BMD increased by over 13% in the spine and almost 7% in the hip in 1 year with romosozumab. At 1 year, the incidence of new vertebral fractures in the romosozumab group was significantly reduced by 73% compared with placebo. Clinical fracture risk (nonvertebral fractures and clinical vertebral fractures combined) was significantly reduced by 36%. Nonvertebral fractures were also reduced, but the difference just missed statistical significance perhaps due to geographical differences; in the high-enrolling Latin American region, there was no significant reduction in nonvertebral fractures, probably due to a very low background incidence in that region. In the rest of the world, nonvertebral fractures were significantly reduced by >40%. During the second year of the FRAME study, both groups transitioned to denosumab. Over 24 months, women who had received romosozumab during the first 12 months and then denosumab had 75% fewer new vertebral fractures than those who had received placebo for a year followed by denosumab. There were also nearly significant trends toward reduced clinical and nonvertebral fractures in the romosozumab/denosumab group. Compared with baseline, BMD increased by 17.6% in the spine and 8.8% in the total hip in the romosozumab/denosumab group. Safety and tolerability of the two drugs were similar, with a slightly higher incidence of injection site reactions in the denosumab group. There was no increase of cardiovascular events in patients treated with romosozumab. The FRAME extension study where all participants received continued denosumab for an additional year (2 years in total) showed significant decreases in vertebral, clinical, and nonvertebral fractures with a trend for hip fracture reduction.

Comparison with Antiresorptive Therapy The ARCH trial of very-high-risk patients, all of whom have prevalent vertebral fractures, compared romosozumab with alendronate for 1 year, followed by transition to, or continuation of, alendronate for at least 2 additional years. This study showed increases in bone density that were comparable with the FRAME study and significant reductions in vertebral, nonvertebral, clinical, and hip fractures compared with alendronate **(Fig. 423-14)**. In this study, there was an increase in cardiovascular adverse events in patients treated with romosozumab, prompting a warning on the label and avoidance of this medication in patients with a past history of acute myocardial infarction or stroke.

Comparison with Other Anabolic Therapy No studies have compared romosozumab with other anabolic drugs with fractures as the primary outcome. One study compared the effects of romosozumab with teriparatide in postmenopausal women with osteoporosis who had been previously treated with bisphosphonates, including 12 months of alendronate immediately prior to the study. Effects of both anabolic drugs on BMD were attenuated by prior bisphosphonate therapy, but the increases in spine, total hip, and femoral neck BMD were greater with romosozumab than with teriparatide. Increases in cortical bone mineral content and bone strength at the hip were also greater with romosozumab than with teriparatide.

Sequence of Therapy Because teriparatide and romosozumab have greater efficacy at reducing vertebral and clinical fractures, and vertebral, clinical, nonvertebral, and hip fractures, respectively, compared with bisphosphonates, they are now being considered as first-line therapy for patients with severe osteoporosis. Increases in BMD also occur more rapidly and to a greater extent with anabolic drugs than with antiresorptive drugs, although the relative risk reduction for vertebral fractures is similar for zoledronic acid, denosumab, teriparatide, and romosozumab at about 70%. When their use is second-line following antiresorptive drugs, improvements in BMD are attenuated, particularly following denosumab therapy. In response to these clinical data, initiatives have been made to identify patients at very high risk for fracture who would derive the most benefit from these greater improvements in BMD and more rapid reductions in fracture risk. Both U.S. and international guidelines have therefore focused on defining a group of individuals at very high fracture risk who could be selected for anabolic therapy as first-line treatment. The definitions vary between these guidelines, but the criteria showing the most agreement are a recent fracture within the last 12–24 months, multiple fragility fractures (two or more), fractures while on antiresorptive therapy, or high absolute fracture risk (FRAX-derived 10-year risks of >4.5% or >30% for hip and major osteoporotic fractures, respectively).

There has also been a focus on a goal-directed approach to long-term management of fracture risk. This helps ensure the most appropriate initial treatment and treatment sequence is selected for individual patients. Importantly, the selection of initial treatment should focus on reducing fracture risk rapidly for patients at very high and/or imminent risk, such as in those with recent fractures. Initial treatment selection should also consider the likelihood that a BMD treatment target can be attained relatively rapidly and needs to consider differences in fracture risk reduction and BMD increases with anabolic versus antiresorptive therapy. In conclusion, for those patients with a very high and/or imminent fracture risk, the ideal treatment sequence would be anabolic followed by antiresorptive therapy. However, reimbursement considerations may limit the implementation of this approach.

OTHER PHARMACOLOGIC AGENTS NOT APPROVED IN THE UNITED STATES

Testosterone has been used to treat osteoporosis associated with low testosterone levels in men. There are data that indicate that testosterone can increase bone density, but there are no data indicating reductions in fractures. In fact, a recent large study showed fracture incidence was numerically higher among men who received testosterone than among those who received placebo. Since there are many other effects of testosterone, especially in older men (including prostate hypertrophy), decisions to use it for treatment of osteoporosis must take these multisystemic effects into account. In hypogonadal men with a high or very high fracture risk, anti-osteoporosis therapy should be used in addition to testosterone therapy.

Sodium fluoride was tested in two large parallel clinical trials in the late 1980s. Although BMD increased substantially, the increase was in part due to fluoride incorporation in the hydroxyapatite crystal. Fracture risk was not reduced and, in fact, was increased in nonvertebral sites. Therefore, fluoride is no longer considered a viable option for osteoporosis treatment.

Several small studies of growth hormone, alone or in combination with other agents, have not shown consistent or substantial positive effects on bone mass.

NONPHARMACOLOGIC APPROACHES

Protective pads worn around the outer thigh, which cover the trochanteric region of the hip, can prevent hip fractures in elderly residents in nursing homes. They are only effective when worn, and their use is limited largely by issues of compliance and comfort, but

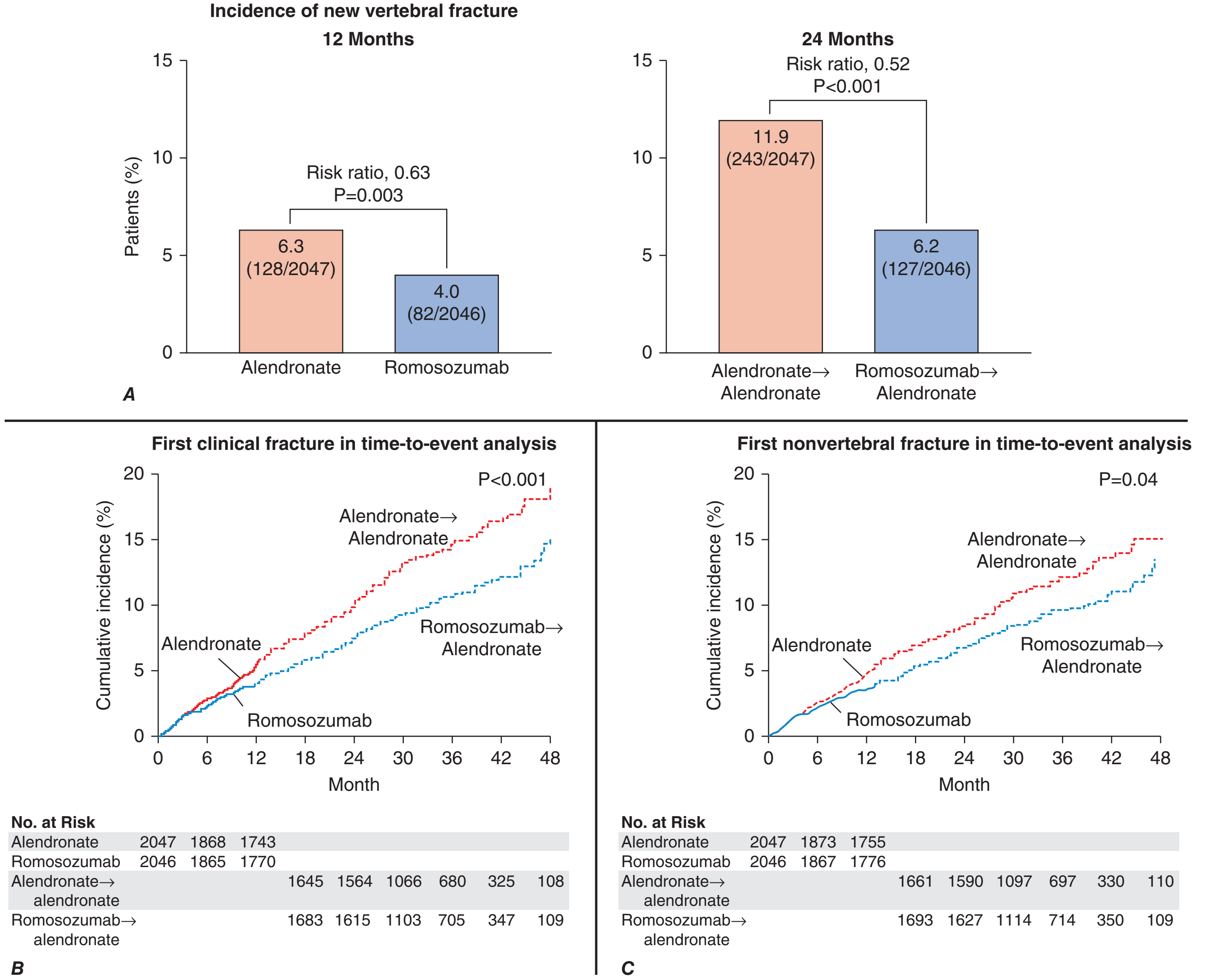

FIGURE 423-14 Effect of romosozumab versus alendronate for 12 months followed by alendronate on vertebral, clinical, and nonvertebral fracture incidence. *(From The New England Journal of Medicine, Romosozumab or Alendronate for Fracture Prevention in Women with Osteoporosis, KG Saag et al: 377:1417. Copyright @2017 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

new devices are being developed that may circumvent these problems and provide adjunctive treatments.

Kyphoplasty and *vertebroplasty* are nonpharmacologic approaches for the treatment of painful vertebral fractures. The overall data do not support routine surgical intervention for painful vertebral fractures since, while this can reduce pain, there is concern about long-term vertebral fracture risk. Vertebral augmentation should be avoided in patients with vertebral fractures associated with denosumab discontinuation as it can increase the risk of adjacent vertebral fractures in this setting.

TREATMENT MONITORING

There are currently no well-accepted guidelines for monitoring treatment of osteoporosis. Because most osteoporosis treatments produce small or moderate bone mass increments on average, it is reasonable to consider BMD as a monitoring tool. Changes must exceed ~3% in the spine and 6% in the hip to be considered significant in any individual. The total hip is the preferred site due to larger surface area and greater reproducibility. Medication-induced increments may require several years to produce changes of this magnitude. Consequently, it can be argued that BMD should be repeated at intervals of every 2 years. Only significant BMD reductions (>~3% or ~6% at the spine and or hip, respectively) should prompt a change in therapy, as it is expected that some individuals will not show responses greater than the detection limits of the current measurement techniques.

Biochemical markers of bone turnover can help in treatment monitoring, with significant changes seen within 3 months of initiating treatment with approved medications and the possible benefit of improving adherence. It remains unclear which endpoint is most useful. If bone turnover markers are used, a determination should be made before therapy is started and repeated ≥3–4 months after therapy is initiated. In general, a change in bone turnover markers must be 30–40% lower than the baseline, or in the lower part of the standardized premenopausal range, to be significant because of the biologic and analytic variability in these tests. Because markers change more rapidly than bone density, they are often early signs of treatment effect. Currently collagen C-telopeptide (CTX) measured on a fasting serum sample in the morning is the preferred marker of bone resorption, and the propeptide of type 1 collagen (P1NP) is the preferred marker for bone formation.

GLUCOCORTICOID-INDUCED OSTEOPOROSIS

Osteoporotic fractures are a well-characterized consequence of the hypercortisolism associated with Cushing's syndrome. However, the therapeutic use of glucocorticoids is by far the most common form of glucocorticoid-induced osteoporosis (GCIOP). Glucocorticoids are used widely in the treatment of a variety of disorders, including chronic lung disorders, rheumatoid arthritis and other connective tissue diseases, and inflammatory bowel disease, and after transplantation.

Osteoporosis and related fractures are serious side effects of chronic glucocorticoid therapy. Because the effects of glucocorticoids on the skeleton are often superimposed on the consequences of aging and menopause, it is not surprising that postmenopausal women and older men are most frequently affected. The skeletal response to glucocorticoids is remarkably heterogeneous, however, and even young, growing individuals treated with glucocorticoids can present with fractures.

The risk of fractures depends on the dose and duration of glucocorticoid therapy, although recent data suggest that there may be no completely safe dose. Bone loss is more rapid during the early months of treatment, and trabecular bone is affected more severely than cortical bone. As a result, fractures have been shown to increase within 3 months of glucocorticoid treatment. There is an increase in fracture risk in both the axial skeleton and the appendicular skeleton, including risk of hip fracture. Bone loss can occur with any route of glucocorticoid administration, including high-dose inhaled glucocorticoids and intra-articular injections. Alternate-day delivery does not appear to ameliorate the skeletal effects of glucocorticoids.

PATHOPHYSIOLOGY

Glucocorticoids increase bone loss by multiple mechanisms, including (1) inhibition of osteoblast function and an increase in osteoblast apoptosis, resulting in impaired synthesis of new bone; (2) stimulation of bone resorption, probably as a secondary effect; (3) impairment of the absorption of calcium across the intestine, probably by a vitamin D–independent effect; (4) increase of urinary calcium loss and perhaps induction of some degree of secondary hyperparathyroidism; (5) reduction of adrenal androgens and suppression of ovarian and testicular secretion of estrogens and androgens; and (6) induction of glucocorticoid myopathy, which may exacerbate effects on skeletal and calcium homeostasis as well as increase the risk of falls.

EVALUATION OF THE PATIENT

Because of the high prevalence of GCIOP, it is important to evaluate the status of the skeleton in all patients starting or already receiving long-term glucocorticoid therapy. Modifiable risk factors should be identified, including those for falls. Examination should include testing of height and muscle strength. Laboratory evaluation should include an assessment of 24-h urinary calcium. All patients on long-term (>3 months) glucocorticoids should have measurement of bone mass at both the spine and the hip using DXA. If only one skeletal site can be measured, it is best to assess the spine in individuals <60 years and the hip in those >60 years.

PREVENTION

Bone loss caused by glucocorticoids can be prevented, and the risk of fractures significantly reduced. Strategies must include using the lowest dose of glucocorticoid for disease management. Topical and inhaled routes of administration are preferred, where appropriate. Risk factor reduction is important, including smoking cessation, limitation of alcohol consumption, and participation in weight-bearing and resistance exercise, when appropriate. All patients should receive an adequate calcium and vitamin D intake from the diet or from supplements.

TREATMENT

Glucocorticoid-Induced Osteoporosis

Several bisphosphonates (alendronate, risedronate, and zoledronic acid) have been demonstrated in large clinical trials to reduce the risk of fractures in patients being treated with glucocorticoids and are FDA approved for the treatment of GCIOP. Teriparatide is also approved for treatment of GCIOP. In one trial comparing teriparatide with alendronate, BMD increases were much greater and vertebral fracture risk reduction was greater with teriparatide compared with alendronate. A study of denosumab indicates greater efficacy of denosumab compared with risedronate for treatment of GCIOP. There are no data for romosozumab in GCIOP. The American College of Rheumatology has published guidelines for the management of GCIOP.

Acknowledgment

The author gratefully acknowledges the foundational contributions of Drs. Robert Lindsay, Felicia Cosman, and Blossom Samuels to this chapter in previous editions of Harrison's.

FURTHER READING

Black DM et al: Atypical femur fracture risk versus fragility fracture prevention with bisphosphonates. N Engl J Med 383:743, 2020.

Compston J: Glucocorticoid-induced osteoporosis: An update. Endocrine 61:7, 2018.

Cosman F et al: Spine fracture prevalence in a nationally representative sample of US women and men aged >/=40 years: results from the National Health and Nutrition Examination Survey (NHANES) 2013–2014. Osteoporos Int 28:2319, 2017.

Cosman F et al: Treatment sequence for osteoporosis. Endocr Pract S1530, 2024.

Kendler DL et al: Effects of teriparatide and risedronate on new fractures in post-menopausal women with severe osteoporosis (VERO): A multicentre, double-blind, double-dummy, randomised controlled trial. Lancet 391:230, 2018.

Khosla S, Hofbauer LC: Osteoporosis treatment: Recent developments and ongoing challenges. Lancet Diabetes Endocrinol 5:898, 2017.

Reid IR, Billington EO: Drug therapy for osteoporosis in older adults. Lancet 399:1080, 2022.

Reid IR et al: Fracture prevention with zoledronate in older women with osteopenia. N Engl J Med 379:2407, 2018.

Roux C, Briot K: Imminent fracture risk. Osteoporos Int 28:1765, 2017.

Saag KG et al: Romosozumab or alendronate for fracture prevention in women with osteoporosis. N Engl J Med 377:1417, 2017.

Sing CW et al: Global epidemiology of hip fractures: Secular trends in incidence rate, post-fracture treatment, and all-cause mortality. J Bone Miner Res 38:1064, 2023.

Snyder PJ et al: Testosterone treatment and fractures in men with hypogonadism. N Engl J Med 390:203, 2024.

424

Paget's Disease and Other Dysplasias of Bone

Rajesh K. Jain, Tamara J. Vokes

PAGET'S DISEASE OF BONE

Paget's disease is a localized bone-remodeling disorder that affects widespread, noncontiguous areas of the skeleton. The pathologic process is initiated by overactive osteoclastic bone resorption followed by a compensatory increase in osteoblastic new bone formation, resulting in a structurally disorganized mosaic of woven and lamellar bone. Pagetic bone is expanded, less compact, and more vascular; thus, it is more susceptible to deformities and fractures. Although most patients are asymptomatic, symptoms resulting directly from bony involvement (bone pain, secondary arthritis, fractures) or secondarily from the expansion of bone causing compression of surrounding neural tissue are not uncommon.

Epidemiology There is a marked geographic variation in the frequency of Paget's disease, with high prevalence in Western Europe (Great Britain, France, and Germany, but not Switzerland or Scandinavia) and among those who have immigrated to Australia, New Zealand, South Africa, and North and South America. The disease is rare in native populations of the Americas, Africa, Asia, and the Middle East; when it does occur, the affected subjects usually have evidence of

European ancestry, supporting the migration theory. For unclear reasons, the prevalence and severity of Paget's disease are decreasing, and the age of diagnosis is increasing.

The prevalence is greater in males and increases with age. Autopsy series reveal Paget's disease in ~3% of those over age 40. Prevalence of positive skeletal radiographs in patients aged >55 years is 2.5% for men and 1.6% for women. Elevated alkaline phosphatase (ALP) levels in asymptomatic patients have an age-adjusted incidence of 12.7 and 7 per 100,000 person-years in men and women, respectively.

Etiology The etiology of Paget's disease of bone remains unknown, but evidence supports both genetic and viral etiologies. A positive family history is found in 15–25% of patients and, when present, raises the prevalence of the disease seven- to tenfold among first-degree relatives.

A clear genetic basis has been established for several rare familial bone disorders that clinically and radiographically resemble Paget's disease but have more severe presentation and earlier onset. A homozygous deletion of the *TNFRSF11B* gene, which encodes osteoprotegrin **(Fig. 424-1)**, causes *juvenile Paget's disease*, also known as *familial idiopathic hyperphosphatasia*, a disorder characterized by uncontrolled osteoclastic differentiation and resorption. Familial patterns of disease in several large kindred are consistent with an autosomal dominant pattern of inheritance with variable penetrance. *Familial expansile osteolysis, expansile skeletal hyperphosphatasia, and early-onset Paget's disease* are associated with mutations in the *TNFRSF11A* gene, which encodes RANK (receptor activator of nuclear factor-κB), a member of the tumor necrosis factor superfamily critical for osteoclast differentiation (Fig. 424-1). A mutation in profilin 1, a small actin protein that acts as a tumor suppressor, also causes early-onset Paget's disease with a predisposition for the development of osteosarcoma. Finally, mutations in the gene for valosin-containing protein cause a rare syndrome with autosomal dominant inheritance and variable penetrance known as *inclusion body myopathy with Paget's disease and frontotemporal dementia* (*IBMPFD*). The role of genetic factors is less clear in the more common form of late-onset Paget's disease. The most common mutations identified in familial and sporadic cases of Paget's disease have been in the *SQSTM1* gene (sequestasome-1 or p62 protein) in the C-terminal ubiquitin-binding domain. The other candidate genes include *CSF1* (1p13), which encodes macrophage colony-stimulating factor (M-CSF), a cytokine that is required for osteoclast differentiation; *RIN3* (14q32), which encodes a guanine exchange factor called Rab and Ras interactor 3; *OPTN* (10p13), which is involved in regulating nuclear factor (NF)-κB; *TNFRSF11A*, mentioned earlier; and *TM7SF4*, which encodes dendritic cell–specific transmembrane protein (DC-STAMP), a molecule that is essential for fusion of the osteoclast. The phenotypic variability in patients with *SQSTM1* mutations suggests that additional factors, such as other genetic influences or viral infection, may influence clinical expression of the disease.

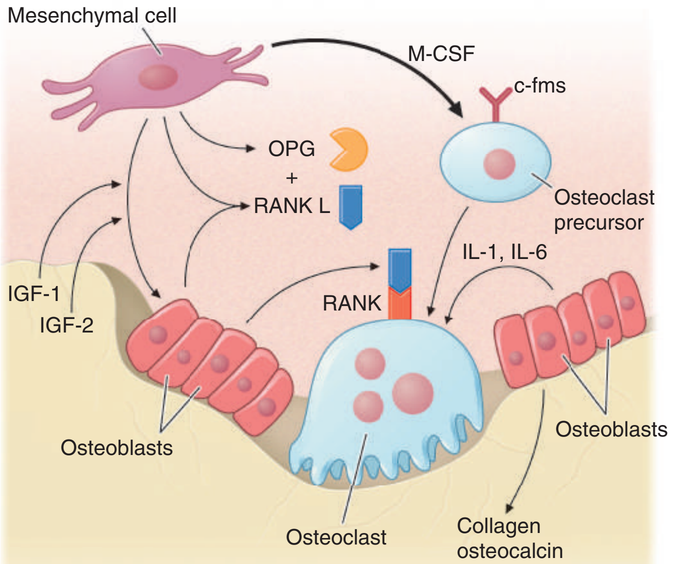

FIGURE 424-1 Diagram illustrating factors that promote differentiation and function of osteoclasts and osteoblasts and the role of the RANK pathway. Stromal bone marrow (mesenchymal) cells and differentiated osteoblasts produce multiple growth factors and cytokines, including macrophage colony-stimulating factor (M-CSF), to modulate osteoclastogenesis. RANKL (receptor activator of nuclear factor-κB [NF-κB] ligand) is produced by osteoblast progenitors and mature osteoblasts and can bind to a soluble decoy receptor known as osteoprotegerin (OPG) to inhibit RANKL action. Alternatively, a cell-cell interaction between osteoblast and osteoclast progenitors allows RANKL to bind to its membrane-bound receptor, RANK, thereby stimulating osteoclast differentiation and function. RANK binds intracellular proteins called tumor necrosis factor receptor–associated factors (TRAFs) that mediate receptor signaling through transcription factors such as NF-κB. M-CSF binds to its receptor, c-fms, which is the cellular homologue of the *fms* oncogene. See text for the potential role of these pathways in disorders of osteoclast function such as Paget's disease and osteopetrosis. IGF, insulin-like growth factor; IL, interleukin.

Several lines of evidence suggest that a viral infection may contribute to the clinical manifestations of Paget's disease, including (1) the presence of cytoplasmic and nuclear inclusions resembling paramyxoviruses (measles, respiratory syncytial virus, canine distemper virus) in pagetic osteoclasts and (2) viral mRNA in precursor and mature osteoclasts. The viral etiology is further supported by conversion of osteoclast precursors to pagetic-like osteoclasts by vectors containing the measles virus nucleocapsid or matrix genes. The decline in the incidence of Paget's disease coincides with the widespread vaccination against measles, also consistent with the potential role of virus in the development of the disease. However, the viral etiology has been questioned by the inability to culture a live virus from pagetic bone and by failure to clone the full-length viral genes from material obtained from patients with Paget's disease. Furthermore, patients with Paget's disease do not have higher antibody levels against paramyxoviruses or measles as compared to controls, nor do antibody levels correlate with disease severity in those with Paget's disease.

Pathophysiology The principal abnormality in Paget's disease is the increased number and activity of osteoclasts. Pagetic osteoclasts are large, increased 10- to 100-fold in number, and have a greater number of nuclei (as many as 100 compared to 3–5 nuclei in the normal osteoclast). The overactive osteoclasts may create a sevenfold increase in resorptive surfaces and an erosion rate of 9 μg/d (normal is 1 μg/d). Several causes for the increased number and activity of pagetic osteoclasts have been identified: (1) osteoclastic precursors are hypersensitive to $1{,}25(OH)_2D_3$; (2) osteoclasts are hyperresponsive to RANK ligand (RANKL), the osteoclast stimulatory factor that mediates the effects of most osteotropic factors on osteoclast formation; (3) marrow stromal cells from pagetic lesions have increased RANKL expression; (4) osteoclast precursor recruitment is increased by interleukin (IL) 6, which is increased in the blood of patients with active Paget's disease and is overexpressed in pagetic osteoclasts; (5) expression of the protooncogene c-*fos*, which increases osteoclastic activity, is increased; and (6) the antiapoptotic oncogene *Bcl*-2 in pagetic bone is overexpressed. Numerous osteoblasts are recruited to active resorption sites and produce large amounts of new bone matrix. As a result, bone turnover is high, and bone mass is normal or increased, not reduced, unless there is concomitant deficiency of calcium and/or vitamin D.

The characteristic feature of Paget's disease is increased bone resorption accompanied by accelerated bone formation. An initial osteolytic phase involves prominent bone resorption and marked hypervascularization. Radiographically, this manifests as an advancing lytic wedge, or "blade of grass" lesion. The second phase is a period of very active bone formation and resorption that replaces normal lamellar bone with haphazard (woven) bone. Fibrous connective tissue may replace normal bone marrow. In the final sclerotic phase, bone resorption declines progressively and leads to a hard, dense, less vascular pagetic or mosaic bone, which represents the so-called burned-out phase of Paget's disease. All three phases may be present at the same time at different skeletal sites.

Clinical Manifestations Diagnosis is often made in asymptomatic patients because they have elevated ALP levels on routine blood chemistry testing or an abnormality on a skeletal radiograph obtained for another indication. The skeletal sites most commonly involved are the pelvis, vertebral bodies, skull, femur, and tibia. Familial cases with an early presentation often have numerous active sites of skeletal involvement.

The most common presenting symptom is pain, which may result from increased bony vascularity, expanding lytic lesions, fractures, bowing, or other deformities. Bowing of the femur or tibia causes gait abnormalities and abnormal mechanical stresses with secondary osteoarthritis of the hip or knee joints. Long bone bowing also causes extremity pain by stretching the muscles attached to the bone softened by the pagetic process. Back pain results from enlarged pagetic vertebrae, vertebral compression fractures, spinal stenosis, degenerative changes of the joints, and altered body mechanics with kyphosis and forward tilt of the upper back. Rarely, spinal cord compression may result from bone enlargement or from the vascular steal syndrome. Skull involvement may cause headaches, symmetric or asymmetric enlargement of the parietal or frontal bones (frontal bossing), and increased head size. Cranial expansion may narrow cranial foramens and cause neurologic complications including hearing loss from cochlear nerve damage from temporal bone involvement, cranial nerve palsies, and softening of the base of the skull (*platybasia*) with the risk of brainstem compression. Pagetic involvement of the facial bones may cause facial deformity; loss of teeth and other dental conditions; and, rarely, airway compression.

Fractures are serious complications of Paget's disease and usually occur in long bones at areas of active or advancing lytic lesions. Common fracture sites are the femoral shaft and subtrochanteric regions. Neoplasms arising from pagetic bone are rare (<0.5%). The incidence of sarcoma appears to be decreasing, possibly because of earlier, more effective treatment with potent antiresorptive agents. The majority of tumors are osteosarcomas, which usually present with new pain in a long-standing pagetic lesion. Osteoclast-rich benign giant cell tumors may arise in areas adjacent to pagetic bone, and they respond to glucocorticoid therapy.

Cardiovascular complications may occur in patients with involvement of large (15–35%) portions of the skeleton and a high degree of disease activity (e.g., ALP four times above normal). The extensive arteriovenous shunting and marked increases in blood flow through the vascular pagetic bone lead to a high-output state and cardiac enlargement. However, high-output heart failure is relatively rare and usually develops in patients with concomitant cardiac pathology. In addition, calcific aortic stenosis and diffuse vascular calcifications have been associated with Paget's disease.

Diagnosis The diagnosis may be suggested on clinical examination by the presence of an enlarged skull with frontal bossing, bowing of an extremity, or short stature with simian posturing. An extremity with an area of warmth and tenderness to palpation may suggest an underlying pagetic lesion. Other findings include bony deformity of the pelvis, skull, spine, and extremities; arthritic involvement of the joints adjacent to lesions; and leg-length discrepancy resulting from deformities of the long bones.

Paget's disease is usually diagnosed from radiologic and biochemical abnormalities. Radiographic findings typical of Paget's disease include enlargement or expansion of an entire bone or area of a long bone, cortical thickening, coarsening of trabecular markings, and typical lytic and sclerotic changes. Skull radiographs **(Fig. 424-2)** reveal regions of "cotton wool," or osteoporosis circumscripta, thickening of diploic areas, and enlargement and sclerosis of a portion or all of one or more skull bones. Vertebral cortical thickening of the superior and inferior end plates creates a "picture frame" vertebra. Diffuse radiodense enlargement of a vertebra is referred to as "ivory vertebra." Pelvic radiographs may demonstrate disruption or fusion of the sacroiliac joints; porotic and radiodense lesions of the ilium with whorls of coarse trabeculation; thickened and sclerotic iliopectineal line (brim sign); and softening with protrusio acetabuli, with axial migration of the hips and functional flexion contracture. Radiographs of long bones reveal bowing deformity and typical pagetic changes of cortical thickening and expansion and areas of lucency and sclerosis **(Fig. 424-3)**. Radionuclide ^{99m}Tc bone scans are less specific but are more sensitive than standard radiographs for identifying sites of active skeletal lesions. Although computed tomography (CT) and magnetic resonance imaging (MRI) studies are not necessary in most cases, CT may be useful for the assessment of possible fracture, and MRI is necessary to assess the possibility of sarcoma, giant cell tumor, or metastatic disease in pagetic bone. Definitive diagnosis of malignancy often requires bone biopsy.

Biochemical evaluation is useful in the diagnosis and management of Paget's disease. The marked increase in bone turnover can be monitored using biochemical markers of bone formation and resorption. The parallel rise in markers of bone formation and resorption confirms the coupling of bone formation and resorption in Paget's disease. The degree of bone marker elevation reflects the extent and severity of the disease. For most patients, serum total ALP remains the test of choice both for diagnosis and assessing response to therapy. Occasionally, a symptomatic patient with evidence of progression at a single site may have a normal total ALP level but increased bone-specific ALP. For unclear reasons, serum osteocalcin, another marker of bone formation, is not always elevated and is not recommended for use in diagnosis or management of Paget's disease. In contrast, bone formation marker P1NP does reflect the activity of the disease and can be used instead of total ALP. Bone resorption markers (serum or urine N-telopeptide or C-telopeptide measured in the blood or urine) are also elevated in active Paget's disease and decrease more rapidly in response to therapy than does ALP.

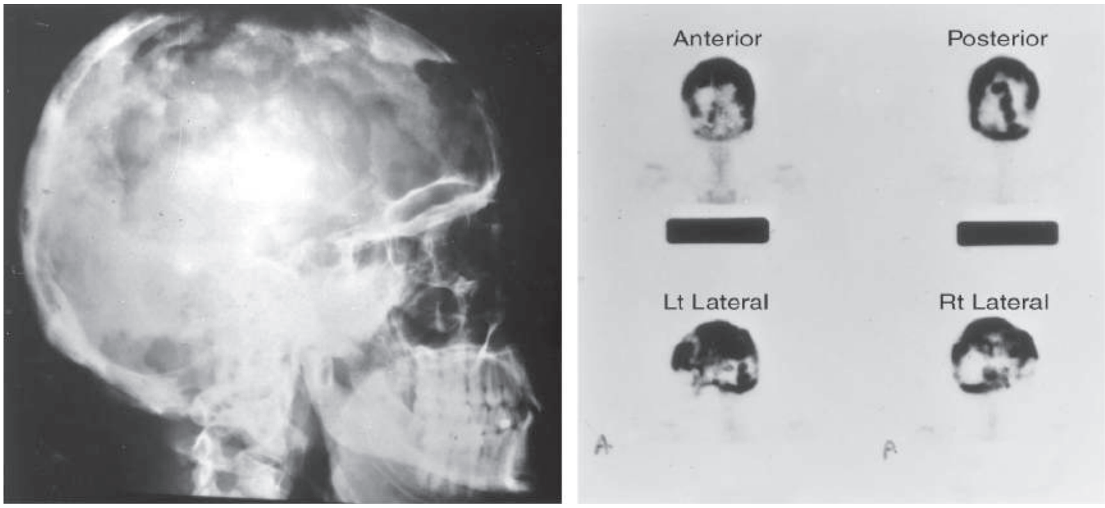

FIGURE 424-2 A 48-year-old woman with Paget's disease of the skull. ***Left.*** Lateral radiograph showing areas of both bone resorption and sclerosis. ***Right.*** ^{99m}Tc hydroxymethylene diphosphonate (HDP) bone scan with anterior, posterior, and lateral views of the skull showing diffuse isotope uptake by the frontal, parietal, occipital, and petrous bones.

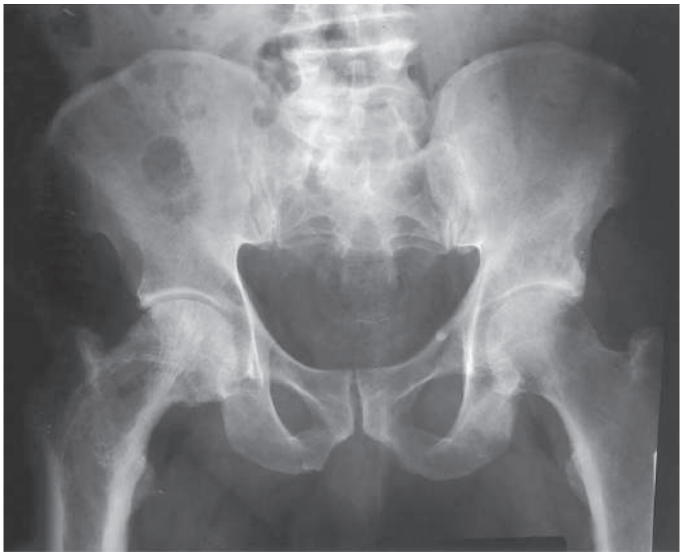

FIGURE 424-3 Radiograph of a 73-year-old man with Paget's disease of the right proximal femur. Note the coarsening of the trabecular pattern with marked cortical thickening and narrowing of the joint space consistent with osteoarthritis secondary to pagetic deformity of the right femur.

Serum calcium and phosphate levels are normal in Paget's disease. Immobilization of a patient with active Paget's disease may rarely cause hypercalcemia and hypercalciuria and increase the risk for nephrolithiasis. However, the discovery of hypercalcemia, even in the presence of immobilization, should prompt a search for another cause of hypercalcemia. In contrast, hypocalcemia or mild secondary hyperparathyroidism may develop in Paget's patients with very active bone formation and insufficient calcium and vitamin D intake, particularly during bisphosphonate therapy when bone resorption is rapidly suppressed and active bone formation continues. Therefore, adequate calcium and vitamin D intake should be instituted prior to administration of bisphosphonates.

TREATMENT

Paget's Disease of Bone

The development of effective and potent pharmacologic agents **(Table 424-1)** has changed the treatment philosophy from treating only symptomatic patients to treating asymptomatic patients who are at risk for complications. According to the Endocrine Society Clinical Practice Guidelines published in 2014, pharmacologic therapy is indicated for most patients with active Paget's disease who are at risk of complications. Treatment may be initiated to control symptoms caused by metabolically active Paget's disease such as bone pain, fracture, headache, pain from pagetic radiculopathy or arthropathy, or neurologic complications; to decrease local blood flow and minimize operative blood loss in patients who need surgery at an active pagetic site; to reduce hypercalciuria that may occur during immobilization; and to decrease the risk of complications when disease activity is high (elevated ALP) and when the site of involvement involves weight-bearing bones, areas adjacent to major joints, vertebral bodies, and the skull. Whether or not early therapy prevents late complications remains to be determined. Randomized studies from the United Kingdom showed no difference in bone pain, fracture rates, quality of life, and hearing loss between patients who received pharmacologic therapy to control symptoms (bone pain) and those receiving bisphosphonates to normalize serum ALP. However, the conclusions of these studies are debatable since the majority of subjects had already received bisphosphonate therapy in the past, perhaps limiting generalizability, and because the bone deformities that occur with Paget's disease may take many years to manifest. It seems likely that the restoration of normal bone architecture following suppression of pagetic activity will prevent further deformities and complications.

Agents approved for treatment of Paget's disease suppress the very high rates of bone resorption and secondarily decrease the high rates of bone formation (Table 424-1). As a result of decreasing bone turnover, pagetic structural patterns, including areas of poorly mineralized woven bone, are replaced by more normal cancellous or lamellar bone. Reduced bone turnover can be documented by a decline in serum formation markers (ALP and P1NP) and urine or serum resorption markers (N-telopeptide, C-telopeptide).

Bisphosphonates are the mainstay of pharmacologic therapy of Paget's disease. Among them, zoledronic acid is currently recommended as the first choice, particularly for those who have severe disease or need rapid normalization of bone turnover (neurologic symptoms, severe bone pain due to a lytic lesion, risk of an impending fracture, or pretreatment prior to elective surgery in an area of active disease). Zoledronic acid normalized bone turnover faster and in a high proportion of patients (>90%) than oral bisphosphonates with the therapeutic effect persisting for months or even years. It is given at a dose of 5 mg as an intravenous infusion over 20 min, although slower rates of infusion are recommended for elderly or those with mild impairment of renal function. More significant renal impairment (glomerular filtration rate <35 mL/min) is a contraindication for use of zoledronic acid due to higher risk of further deterioration of renal function. About 20–25% of patients experience a flulike syndrome after the first infusion, which can be partly ameliorated by pretreatment with acetaminophen, nonsteroidal anti-inflammatory drugs (NSAIDs), or glucocorticoids. Oral bisphosphonates, alendronate and risedronate, can be used in subjects who have mild disease or some degree of renal impairment. Oral bisphosphonates should be taken first thing in the morning on an empty stomach, followed by maintenance of upright posture with no food, drink, or other medications for 30–60 min. The efficacy of different agents, based on their ability to normalize or decrease ALP levels, is summarized in Table 424-1, although the response rates are not comparable because they are obtained from different studies.

The subcutaneous injectable form of salmon calcitonin is approved for the treatment of Paget's disease but is rarely used due to its low potency and should be reserved for patients who either do not tolerate bisphosphonates or have a contraindication to their use. For patients with contraindication to bisphosphonates, another alternative is denosumab, an antibody to RANKL, which has been reported to result in reduction in ALP. However, it has not been approved for this indication and has less complete and less durable effects than bisphosphonates.

TABLE 424-1 Pharmacologic Agents Approved for Treatment of Paget's Disease

NAME	DOSE AND MODE OF DELIVERY	NORMALIZATION OF ALKALINE PHOSPHATASE (ALP)
Zoledronic acid	5 mg IV over 15 min	90% of patients at 6 mo
Pamidronate	30 mg/d IV over 4 h on 3 days	~50% of patients
Risedronate	30 mg/d PO for 2 mo	73% of patients
Alendronate	40 mg/d PO for 6 mo	63% of patients
Tiludronate	800 mg/d PO for 3 mo	35% of patients
Etidronate	200–400 mg/d PO × 6 mo	15% of patients
Calcitonin (Miacalcin)	100 U SC daily for 6–18 mo (may reduce to 50 U 3× per week)	(Reduction of ALP by up to 50%)

SCLEROSING BONE DISORDERS

OSTEOPETROSIS

Osteopetrosis refers to a group of disorders caused by severe impairment of osteoclast-mediated bone resorption. Other terms that are often used include marble bone disease, which captures the solid x-ray appearance of the involved skeleton, and Albers-Schonberg disease, which refers to the milder, adult form of osteopetrosis also known as autosomal dominant osteopetrosis type II. The major types of osteopetrosis include malignant (severe, infantile, autosomal recessive) osteopetrosis and benign (adult, autosomal dominant) osteopetrosis types I and II. A rare autosomal recessive intermediate form has a more benign prognosis. Autosomal recessive carbonic anhydrase (CA) II deficiency produces osteopetrosis of intermediate severity associated with renal tubular acidosis and cerebral calcification.

Etiology and Genetics Naturally occurring and gene-knockout animal models with phenotypes similar to those of the human disorders have been used to explore the genetic basis of osteopetrosis. The primary defect in osteopetrosis is the loss of osteoclastic bone resorption and preservation of normal osteoblastic bone formation. Osteoprotegerin (OPG) is a soluble decoy receptor that binds osteoblast-derived RANK ligand, which mediates osteoclast differentiation and activation (Fig. 424-1). Transgenic mice that overexpress OPG develop osteopetrosis, presumably by blocking RANK ligand. Mice deficient in RANK lack osteoclasts and develop severe osteopetrosis.

Recessive mutations of CA II prevent osteoclasts from generating an acid environment in the clear zone between its ruffled border and the adjacent mineral surface. Absence of CA II, therefore, impairs osteoclastic bone resorption. Other forms of human disease have less clear genetic defects. About one-half of the patients with malignant infantile osteopetrosis have a mutation in the *TCIRG1* gene encoding the osteoclast-specific subunit of the vacuolar proton pump, which mediates the acidification of the interface between bone mineral and the osteoclast ruffled border. Mutations in the *CLCN7* chloride channel gene cause autosomal dominant osteopetrosis type II. A drug-induced version of osteopetrosis has been reported in children with osteogenesis imperfecta who receive repeated doses of bisphosphonates.

Clinical Presentation The incidence of autosomal recessive severe (malignant) osteopetrosis ranges from 1 in 200,000 to 1 in 500,000 live births. As bone and cartilage fail to undergo modeling, paralysis of one or more cranial nerves may occur due to narrowing of the cranial foramens. Failure of skeletal modeling also results in inadequate marrow space, leading to extramedullary hematopoiesis with hypersplenism and pancytopenia. Hypocalcemia due to lack of osteoclastic bone resorption may occur in infants and young children. The untreated infantile disease is fatal, often before age 5.

Adult (benign) osteopetrosis is an autosomal dominant disease that is usually diagnosed by the discovery of typical skeletal changes in young adults who undergo radiologic evaluation of a fracture. The prevalence is 1 in 100,000 to 1 in 500,000 adults. The course is not always benign, because fractures may be accompanied by loss of vision, deafness, psychomotor delay, mandibular osteomyelitis, and other complications usually associated with the juvenile form. In some kindred, nonpenetrance results in skip generations, while in other families, severely affected children are born into families with benign disease. The milder form of the disease does not usually require treatment.

Radiography Typically, there are generalized symmetric increases in bone mass with thickening of both cortical and trabecular bone. Diaphyses and metaphyses are broadened, and alternating sclerotic and lucent bands may be seen in the iliac crests, at the ends of long bones, and in vertebral bodies. The cranium is usually thickened, particularly at the base of the skull, and the paranasal and mastoid sinuses are underpneumatized.

Laboratory Findings The only significant laboratory findings are elevated serum levels of osteoclast-derived tartrate-resistant acid phosphatase (TRAP) and the brain isoenzyme of creatine kinase. Serum calcium may be low in severe disease, and parathyroid hormone and 1,25-dihydroxyvitamin D levels may be elevated in response to hypocalcemia.

TREATMENT

Osteopetrosis

Allogeneic human leukocyte antigen (HLA)–identical bone marrow transplantation has been successful in some children. Following transplantation, the marrow contains progenitor cells and normally functioning osteoclasts. With long-term follow-up after transplantation, radiographic improvements, such as improvements in Erlenmeyer flask deformities, are seen, although there is not complete normalization. A cure is most likely when children are transplanted before age 4. Marrow transplantation from nonidentical HLA-matched donors has a much higher failure rate. Limited studies in small numbers of patients have suggested variable benefits following treatment with interferon γ-1β, 1,25-dihydroxyvitamin D (which stimulates osteoclasts directly), methylprednisolone, and a low-calcium/high-phosphate diet.

Surgical intervention is indicated to decompress optic or auditory nerve compression. Orthopedic management is required for the surgical treatment of fractures and their complications, including malunion and postfracture deformity.

PYKNODYSOSTOSIS

This is an autosomal recessive form of osteosclerosis that is believed to have affected the French impressionist painter Henri de Toulouse-Lautrec. The molecular basis involves mutations in the gene that encodes cathepsin K, a lysosomal metalloproteinase highly expressed in osteoclasts and important for bone-matrix degradation. Osteoclasts are present but do not function normally. Pyknodysostosis is a form of short-limb dwarfism that presents with frequent fractures but usually a normal life span. Clinical features include short stature; kyphoscoliosis and deformities of the chest; high arched palate; proptosis; blue sclerae; dysmorphic features including small face and chin, frontooccipital prominence, pointed beaked nose, large cranium, and obtuse mandibular angle; and small, square hands with hypoplastic nails. Radiographs demonstrate a generalized increase in bone density, but in contrast to osteopetrosis, the long bones are normally shaped. Separated cranial sutures, including the persistent patency of the anterior fontanel, are characteristic of the disorder. There may also be hypoplasia of the sinuses, mandible, distal clavicles, and terminal phalanges. Persistence of deciduous teeth and sclerosis of the calvarium and base of the skull are also common. Histologic evaluation shows normal cortical bone architecture with decreased osteoblastic and osteoclastic activities. Serum chemistries are normal, and unlike osteopetrosis, there is no anemia. There is no known treatment for this condition, and there are no reports of attempted bone marrow transplant.

PROGRESSIVE DIAPHYSEAL DYSPLASIA

Also known as *Camurati-Engelmann disease*, progressive diaphyseal dysplasia is an autosomal dominant disorder that is characterized radiographically by diaphyseal hyperostosis and a symmetric thickening and increased diameter of the endosteal and periosteal surfaces of the diaphyses of the long bones, particularly the femur and tibia, and, less often, the fibula, radius, and ulna. The genetic defect responsible for the disease has been localized to the area of chromosome 19q13.2 encoding transforming growth factor (TGF)-β1. The mutation promotes activation of TGF-β1. The clinical severity is variable. The most common presenting symptoms are pain and tenderness of the involved areas, fatigue, muscle wasting, and gait disturbance. The weakness may be mistaken for muscular dystrophy. Characteristic body habitus includes thin limbs with little muscle mass yet prominent and palpable bones and, when the skull is involved, large head with prominent forehead and proptosis. Patients may also display signs of cranial nerve palsies, hydrocephalus, central hypogonadism, and Raynaud's phenomenon. Radiographically, patchy progressive endosteal and periosteal new bone formation is observed along the diaphyses of the long bones. Bone scintigraphy shows increased radiotracer uptake in involved areas.

Treatment with low-dose glucocorticoids relieves bone pain and may reverse the abnormal bone formation. Intermittent bisphosphonate therapy has produced clinical improvement in a limited number of patients. Disease activity may also attenuate as patients enter adulthood.

HYPEROSTOSIS CORTICALIS GENERALISATA

This is also known as *van Buchem's disease*; it is an autosomal recessive disorder characterized by endosteal hyperostosis in which osteosclerosis involves the skull, mandible, clavicles, and ribs. The major manifestations are due to narrowed cranial foramens with neural compressions that may result in optic atrophy, facial paralysis, and deafness. Adults may have an enlarged mandible. Serum ALP levels may be elevated, which reflect the uncoupled bone remodeling with high osteoblastic formation rates and low osteoclastic resorption. As a result, there is increased accumulation of normal bone. Endosteal hyperostosis with syndactyly, known as *sclerosteosis*, is a more severe form. The genetic defects for both sclerosteosis and van Buchem's disease have been associated with mutations in the *SOST* gene.

MELORHEOSTOSIS

Melorheostosis (Greek, "flowing hyperostosis") may occur sporadically or follow a pattern consistent with an autosomal recessive disorder. The major manifestation is progressive linear hyperostosis in one or more bones of one limb, usually a lower extremity. The name comes from the radiographic appearance of the involved bone, which resembles melted wax that has dripped down a candle. Symptoms appear during childhood as pain or stiffness in the area of sclerotic bone. There may be associated ectopic soft tissue masses, composed of cartilage or osseous tissue, and skin changes overlying the involved bone, consisting of scleroderma-like areas and hypertrichosis. The disease does not progress in adults, but pain and stiffness may persist. Laboratory tests are unremarkable. Somatic mutations in *MAP2K1*, which increases MEK1 activity downstream of the RAS pathway, and *SMAD3*, which upregulates the TGF-β/SMAD pathway, have been identified in affected bone in patients with melorheostosis. There is no specific treatment. Surgical interventions to correct contractures are often unsuccessful.

OSTEOPOIKILOSIS

The literal translation of osteopoikilosis is "spotted bones"; it is a benign autosomal dominant condition in which numerous small, variably shaped (usually round or oval) foci of bony sclerosis are seen in the epiphyses and adjacent metaphyses. The lesions may involve any bone except the skull, ribs, and vertebrae. They may be misidentified as metastatic lesions. The main differentiating points are that bony lesions of osteopoikilosis are stable over time and do not accumulate radionucleotide on bone scanning. In some kindred, osteopoikilosis is associated with connective tissue nevi known as *dermatofibrosis lenticularis disseminata*, also known as *Buschke-Ollendorff syndrome*. Most cases are caused by mutations in *LEMD3*, which is involved with bone morphogenetic protein (BMP) signaling. Histologic inspection reveals thickened but otherwise normal trabeculae and islands of normal cortical bone. No treatment is indicated.

HEPATITIS C–ASSOCIATED OSTEOSCLEROSIS

Hepatitis C–associated osteosclerosis (HCAO) is a rare acquired diffuse osteosclerosis in adults with prior hepatitis C infection. After a latent period of several years, patients develop diffuse appendicular bone pain and a generalized increase in bone mass with elevated serum ALP. Bone biopsy and histomorphometry reveal increased rates of bone formation, decreased bone resorption with a marked decrease in osteoclasts, and dense lamellar bone. One patient had increased serum OPG levels, and bone biopsy showed large numbers of osteoblasts positive for OPG and reduced osteoclast number. Empirical therapy includes pain control, and there may be beneficial response to bisphosphonate. Long-term antiviral therapy may reverse the bone disease.

DISORDERS ASSOCIATED WITH DEFECTIVE MINERALIZATION

HYPOPHOSPHATASIA

This is a rare inherited disorder that presents as rickets in infants and children or osteomalacia in adults with paradoxically low serum levels of ALP. The frequency of the severe neonatal and infantile forms is about 1 in 100,000 live births in Canada, where the disease is most common because of its high prevalence among Mennonites and Hutterites. It is rare in African Americans. The severity of the disease is remarkably variable, ranging from intrauterine death associated with profound skeletal hypomineralization at one extreme to premature tooth loss as the only manifestation in some adults. Severe cases are inherited in an autosomal recessive manner, but the genetic patterns are less clear for the milder forms. The disease is caused by a deficiency of tissue nonspecific (bone/liver/kidney) ALP (TNSALP), which, although ubiquitous, results only in bone abnormalities. Protein levels and functions of the other ALP isozymes (germ cell, intestinal, placental) are normal. Defective ALP permits accumulation of its major naturally occurring substrates including phosphoethanolamine (PEA), inorganic pyrophosphate (PPi), and pyridoxal 5′-phosphate (PLP). The accumulation of PPi interferes with mineralization through its action as a potent inhibitor of hydroxyapatite crystal growth.

Perinatal hypophosphatasia becomes manifest during pregnancy and is often complicated by polyhydramnios and intrauterine death. The infantile form becomes clinically apparent before the age of 6 months with failure to thrive, rachitic deformities, functional craniosynostosis despite widely open fontanels (which are actually hypomineralized areas of the calvarium), raised intracranial pressure, and flail chest with predisposition to pneumonia. Hypercalcemia and hypercalciuria are common. This form has a mortality rate of ~50%. Prognosis seems to improve for the children who survive infancy. Childhood hypophosphatasia has variable clinical presentation. Premature loss of deciduous teeth (before age 5) is the hallmark of the disease. Rickets causes delayed walking with waddling gait, short stature, and dolichocephalic skull with frontal bossing. The disease often improves during puberty but may recur in adult life. Adult hypophosphatasia presents during middle age with painful, poorly healing metatarsal stress fractures or thigh pain due to femoral pseudofractures. Presentation may be subtle with muscle pain or recurring headaches as the predominant symptoms. It is important to recognize hypophosphatasia in adults because treatment with bisphosphonates can result in increased rather than decreased bone fragility.

Laboratory investigation reveals low ALP levels and normal or elevated levels of serum calcium and phosphorus despite clinical and radiologic evidence of rickets or osteomalacia. Serum parathyroid hormone, 25-hydroxyvitamin D, and 1,25-dihydroxyvitamin D levels are normal. The elevation of PLP is specific for the disease and may even be present in asymptomatic parents of severely affected children. Because vitamin B_6 increases PLP levels, vitamin B_6 supplements should be discontinued 1 week before testing. Clinical testing is available to detect loss-of-function mutation(s) within the *ALPL* gene that encodes TNSALP.

In contrast to other forms of rickets and osteomalacia, calcium and vitamin D supplementation should be avoided because they may aggravate hypercalcemia and hypercalciuria. A low-calcium diet, glucocorticoids, and calcitonin have been used in a small number of patients with variable responses. Because fracture healing is poor, placement of intramedullary rods is best for acute fracture repair and for prophylactic prevention of fractures. In 2015, asfotase alfa, a TNSALP, was approved as enzyme replacement therapy for the perinatal/infantile- and juvenile-onset forms. With 7 years of therapy, children with perinatal/infantile forms showed sustained improvements in mineralization, along with improvements in other features, such as respiratory function and growth. In adolescents and adults, 5 years of therapy demonstrated improved functional abilities, such as increases in 6-min walk time.

AXIAL OSTEOMALACIA

This is a rare disorder characterized by defective skeletal mineralization despite normal serum calcium and phosphate levels. Clinically, the disorder presents in middle-aged or elderly men with chronic axial skeletal discomfort. Cervical spine pain may also be present. Radiographic findings are mainly osteosclerosis due to coarsened trabecular patterns typical of osteomalacia. Spine, pelvis, and ribs are most commonly affected. Histologic changes show defective mineralization and flat, inactive osteoblasts. The primary defect appears to be an acquired defect in osteoblast function. The course is benign, and there is no established treatment. Calcium and vitamin D therapies are not effective.

FIBROGENESIS IMPERFECTA OSSIUM

This is a rare condition of unknown etiology. It presents in both sexes; in middle age or later; and with progressive, intractable skeletal pain and fractures; worsening immobilization; and a debilitating course. The only biochemical abnormality is elevated ALP. Radiographic evaluation reveals generalized osteomalacia, osteopenia, and occasional pseudofractures. Histologic features include a tangled pattern of collagen fibrils with abundant osteoblasts and osteoclasts. Use of growth hormone led to substantial short-term clinical improvement in two adult patients, but long-term outcomes are unknown. No other effective treatment is known. Spontaneous remission has been reported in a small number of patients.

FIBROUS DYSPLASIA AND MCCUNE-ALBRIGHT SYNDROME

Fibrous dysplasia is a sporadic disorder characterized by the presence of one (monostotic) or more (polyostotic) expanding fibrous skeletal lesions composed of bone-forming mesenchyme. The association of the polyostotic form with café au lait spots and hyperfunction of an endocrine system such as pseudoprecocious puberty of ovarian origin is known as *McCune-Albright syndrome* (MAS). A spectrum of the phenotypes is caused by activating mutations in the *GNAS1* gene, which encodes the α subunit of the stimulatory G protein ($G_s\alpha$). As the postzygotic mutations occur at different stages of early development, the extent and type of tissue affected are variable and explain the mosaic pattern of skin and bone changes. GTP binding activates the $G_s\alpha$ regulatory protein and mutations in regions of $G_s\alpha$ that selectively inhibit GTPase activity, which results in constitutive stimulation of the cyclic AMP–protein kinase A signal transduction pathway. Such mutations of the $G_s\alpha$ protein–coupled receptor may cause autonomous function in bone (parathyroid hormone receptor); skin (melanocyte-stimulating hormone receptor); and various endocrine glands including ovary (follicle-stimulating hormone receptor), thyroid (thyroid-stimulating hormone receptor), adrenal (adrenocorticotropic hormone receptor), and pituitary (growth hormone–releasing hormone receptor). The skeletal lesions are composed largely of mesenchymal cells that do not differentiate into osteoblasts, resulting in the formation of imperfect bone. In some areas of bone, fibroblast-like cells develop features of osteoblasts in that they produce extracellular matrix that organizes into woven bone. Calcification may occur in some areas. In other areas, cells have features of chondrocytes and produce cartilage-like extracellular matrix.

Clinical Presentation Fibrous dysplasia occurs with equal frequency in both sexes, whereas MAS with precocious puberty is more common (10:1) in girls. The monostotic form is the most common and is usually diagnosed in patients between 20 and 30 years of age without associated skin lesions. The polyostotic form typically manifests in children <10 years old and may progress with age. Early-onset disease is generally more severe. Lesions may become quiescent in puberty and progress during pregnancy or with estrogen therapy. In polyostotic fibrous dysplasia, the lesions most commonly involve the maxilla and other craniofacial bones, ribs, and metaphyseal or diaphyseal portions of the proximal femur or tibia. Expanding bone lesions may cause pain, deformity, fractures, and nerve entrapment. Sarcomatous degeneration involving the facial bones or femur is infrequent (<1%). The risk of malignant transformation is increased by radiation, which has proven to be ineffective treatment. In rare patients with widespread lesions, renal phosphate wasting and hypophosphatemia may cause rickets or osteomalacia. Hypophosphatemia may be due to production of a phosphaturic factor by the abnormal fibrous tissue.

MAS patients may have café au lait spots, which are flat, hyperpigmented skin lesions that have rough borders ("coast of Maine") in contrast to the café au lait lesions of neurofibromatosis that have smooth borders ("coast of California"). The most common endocrinopathy is isosexual pseudoprecocious puberty in girls. Other less common endocrine disorders include thyrotoxicosis, Cushing's syndrome, acromegaly, hyperparathyroidism, hyperprolactinemia, and pseudoprecocious puberty in boys.

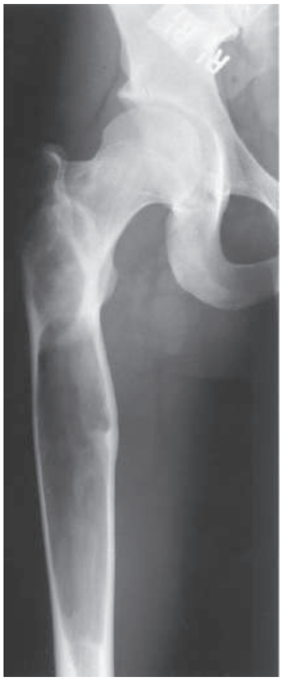

FIGURE 424-4 Radiograph of a 16-year-old male with fibrous dysplasia of the right proximal femur. Note the multiple cystic lesions, including the large lucent lesion in the proximal midshaft with scalloping of the interior surface. The femoral neck contains two lucent cystic lesions.

Radiographic Findings In long bones, the fibrous dysplastic lesions are typically well-defined, radiolucent areas with thin cortices and a ground-glass appearance. Lesions may be lobulated with trabeculated areas of radiolucency (Fig. 424-4). Involvement of facial bones usually presents as radiodense lesions, which may create a leonine appearance (leontiasis osea). Expansile cranial lesions may narrow foramens and cause optic lesions, reduce hearing, and create other manifestations of cranial nerve compression.

Laboratory Results Serum ALP is occasionally elevated, but calcium, parathyroid hormone, 25-hydroxyvitamin D, and 1,25-dihydroxyvitamin D levels are normal. Patients with extensive polyostotic lesions may have hypophosphatemia, hyperphosphaturia, and osteomalacia. The hypophosphatemia and phosphaturia are directly related to the levels of fibroblast growth factor 23 (FGF23). Biochemical markers of bone turnover may be elevated.

TREATMENT

Fibrous Dysplasia and MAS

Spontaneous healing of the lesions does not occur, and there is no established effective treatment. Improvement in bone pain and partial or complete resolution of radiographic lesions have been reported after IV bisphosphonate therapy. Denosumab given monthly or every 3 months is effective in reducing bone turnover markers and leads to some clinical improvement, though subsequent discontinuation of denosumab occasionally results in hypercalcemia. Surgical stabilization is used to prevent pathologic fracture or destruction of a major joint space and to relieve nerve root or cranial nerve compression or sinus obstruction.

OTHER DYSPLASIAS OF BONE AND CARTILAGE

PACHYDERMOPERIOSTOSIS

Pachydermoperiostosis, or hypertrophic osteoarthropathy (primary or idiopathic), is an autosomal dominant disorder characterized by periosteal new bone formation that involves the distal extremities. The lesions present as clubbing of the digits and hyperhidrosis and thickening of the skin, primarily of the face and forehead. The changes usually appear during adolescence, progress over the next decade, and then become quiescent. During the active phase, progressive enlargement of the hands and feet produces a paw-like appearance, which may be mistaken for acromegaly. Arthralgias, pseudogout, and limited mobility may also occur. The disorder must be differentiated from secondary hypertrophic osteopathy that develops during the course of serious pulmonary disorders. The two conditions can be differentiated by standard radiography of the digits in which secondary pachydermoperiostosis has exuberant periosteal new bone formation and a smooth and undulating surface. In contrast, primary hypertrophic osteopathy has an irregular periosteal surface.

The disease is genetically heterogeneous and is related to increases in prostaglandin E2. Synovial fluid does not have an inflammatory profile. There is no specific therapy, although a limited experience with colchicine suggests some benefit in controlling the arthralgias.

OSTEOCHONDRODYSPLASIAS

These include several hundred heritable disorders of connective tissue. These primary abnormalities of cartilage manifest as disturbances in cartilage and bone growth. Selected growth-plate chondrodysplasias are described here. **For discussion of chondrodysplasias, see Chap. 425.**

Achondrodysplasia This is a relatively common form of short-limb dwarfism that occurs in 1 in 15,000 to 1 in 40,000 live births. The disease is caused by a mutation of the fibroblast growth factor receptor 3 (*FGFR3*) gene that results in a gain-of-function state. Most cases are sporadic mutations. However, when the disorder appears in families, the inheritance pattern is consistent with an autosomal dominant disorder. The primary defect is abnormal chondrocyte proliferation at the growth plate that causes development of short, but proportionately thick, long bones. Other regions of the long bones may be relatively unaffected. The disorder is manifest by the presence of short limbs (particularly the proximal portions), normal trunk, large head, saddle nose, and an exaggerated lumbar lordosis. Severe spinal deformity may lead to cord compression. The homozygous disorder is more serious than the sporadic form and may cause neonatal death. Vosoritide, an analog of C-type natriuretic peptide, increased growth among children in phase 3 clinical trials and was approved in 2021. Treatment is controversial among patient support communities. Infigratinib, a selective FGFR1-3 tyrosine kinase inhibitor, is in clinical trials. Pseudoachondroplasia clinically resembles achondrodysplasia but has no skull abnormalities.

Enchondromatosis This is also called *dyschondroplasia* or *Ollier's disease*; it is also a disorder of the growth plate in which the primary cartilage is not resorbed. Cartilage ossification proceeds normally, but it is not resorbed normally, leading to cartilage accumulation. The changes are most marked at the ends of long bones, where the highest growth rates occur. Chondrosarcoma develops infrequently. The association of enchondromatosis and cavernous hemangiomas of the skin and soft tissues is known as *Maffucci's syndrome*. Both Ollier's disease and Maffucci's syndrome are associated with various malignancies, including granulosa cell tumor of the ovary and cerebral glioma.

Multiple Osteochondromas This is also called *multiple exostoses* or *diaphyseal aclasis*; it is a genetic disorder that follows an autosomal dominant pattern of inheritance. In this condition, areas of growth plates become displaced, presumably by growing through a defect in the perichondrium. The lesion begins with vascular invasion of the growth-plate cartilage, resulting in a characteristic radiographic finding of a mass that is in direct communication with the marrow cavity of the parent bone. The underlying cortex is resorbed. The disease is caused by inactivating mutations of the *EXT1* and *EXT2* genes, whose products normally synthesize heparan sulfate chains. The resulting heparan sulfate deficiency impacts signaling pathways and leads to ectopic chondrogenesis. Solitary or multiple lesions are located in the metaphyses of long bones. Although usually asymptomatic, the lesions may interfere with joint or tendon function or compress peripheral nerves. The lesions stop growing when growth ceases but may recur during pregnancy. There is a small risk for malignant transformation into chondrosarcoma. Palovarotene, a retinoic acid receptor agonist, is in clinical trials.

EXTRASKELETAL (ECTOPIC) CALCIFICATION AND OSSIFICATION

Deposition of calcium phosphate crystals (*calcification*) or formation of true bone (*ossification*) in nonosseous soft tissue may occur by one of three mechanisms: (1) metastatic calcification due to a supranormal calcium × phosphate concentration product in extracellular fluid; (2) dystrophic calcification due to mineral deposition into metabolically impaired or dead tissue despite normal serum levels of calcium and phosphate; and (3) ectopic ossification, or true bone formation. Disorders that may cause extraskeletal calcification or ossification are listed in **Table 424-2**.

METASTATIC CALCIFICATION

Soft tissue calcification may complicate diseases associated with significant hypercalcemia, hyperphosphatemia, or both. In addition, vitamin D and phosphate treatments or calcium administration in the presence of mild hyperphosphatemia, such as during hemodialysis, may induce ectopic calcification. Calcium phosphate precipitation may complicate any disorder when the serum calcium × phosphate concentration product is >75. The initial calcium phosphate deposition is in the form of small, poorly organized crystals, which subsequently organize into hydroxyapatite crystals. Calcifications that occur in hypercalcemic states with normal or low phosphate have a predilection for kidney, lungs, and gastric mucosa. Hyperphosphatemia with normal or low serum calcium may promote soft tissue calcification with predilection for the kidney and arteries. The disturbances of calcium and phosphate in renal failure and hemodialysis are common causes of soft tissue (metastatic) calcification.

TUMORAL CALCINOSIS

This is a rare genetic disorder characterized by masses of metastatic calcifications in soft tissues around major joints, most often shoulders, hips, and ankles. Tumoral calcinosis differs from other disorders in that the periarticular masses contain hydroxyapatite crystals or amorphous calcium phosphate complexes, whereas in fibrodysplasia ossificans progressiva (below), true bone is formed in soft tissues. About one-third of tumoral calcinosis cases are familial, with both autosomal recessive and autosomal dominant modes of inheritance reported. The disease is also associated with a variably expressed abnormality of dentition marked

TABLE 424-2 Diseases and Conditions Associated with Ectopic Calcification and Ossification

Metastatic calcification	Dystrophic calcification
Hypercalcemic states	Inflammatory disorders
Primary hyperparathyroidism	Scleroderma
Sarcoidosis	Dermatomyositis
Vitamin D intoxication	Systemic lupus erythematosus
Milk-alkali syndrome	Trauma-induced
Renal failure	Ectopic ossification
Hyperphosphatemia	Myositis ossificans
Tumoral calcinosis	Postsurgery
Secondary hyperparathyroidism	Burns
Pseudohypoparathyroidism	Neurologic injury
Renal failure	Other trauma
Hemodialysis	Fibrodysplasia ossificans progressiva
Cell lysis following chemotherapy	
Therapy with vitamin D and phosphate	

by short bulbous roots, pulp calcification, and radicular dentin deposited in swirls. The disorder is caused by gene mutations in *GALNT3*, *FGF23*, or *α-Klotho*, leading to FGF23 deficiency or resistance. The reduced activity of FGF23 leads to increased renal tubular reabsorption of phosphate, elevated serum phosphate, and spontaneous soft tissue calcification from elevated calcium-phosphate concentration product.

The disease usually presents in childhood and continues throughout the patient's life. The calcific masses are typically painless and grow at variable rates, sometimes becoming large and bulky. The masses are often located near major joints but remain extracapsular. Joint range of motion is not usually restricted unless the tumors are very large. Complications include compression of neural structures and ulceration of the overlying skin with drainage of chalky fluid and risk of secondary infection. Small deposits not detected by standard radiographs may be detected by ^{99m}Tc bone scanning. The most common laboratory findings are hyperphosphatemia and elevated serum 1,25-dihydroxyvitamin D levels. Serum calcium, parathyroid hormone, and ALP levels are usually normal. Renal function is also usually normal. Urine calcium and phosphate excretions are low, and calcium and phosphate balances are positive.

An acquired form of the disease may occur with other causes of hyperphosphatemia, such as secondary hyperparathyroidism associated with hemodialysis, hypoparathyroidism, pseudohypoparathyroidism, and massive cell lysis following chemotherapy for leukemia. Tissue trauma from joint movement may contribute to the periarticular calcifications. Metastatic calcifications are also seen in conditions associated with hypercalcemia, such as in sarcoidosis, vitamin D intoxication, milk-alkali syndrome, and primary hyperparathyroidism. In these conditions, however, mineral deposits are more likely to occur in proton-transporting organs such as kidney, lungs, and gastric mucosa in which an alkaline milieu is generated by the proton pumps.

TREATMENT

Tumoral Calcinosis

Therapeutic successes have been achieved with surgical removal of subcutaneous calcified masses, which tend not to recur if all calcification is removed from the site. Reduction of serum phosphate by chronic phosphorus restriction may be accomplished using low dietary phosphorus intake alone or in combination with oral phosphate binders. The addition of the phosphaturic agent acetazolamide may be useful. Limited experience using the phosphaturic action of calcitonin deserves further testing.

DYSTROPHIC CALCIFICATION

Posttraumatic calcification may occur with normal serum calcium and phosphate levels and normal ion-solubility product. The deposited mineral is either in the form of amorphous calcium phosphate or hydroxyapatite crystals. Soft tissue calcification complicating connective tissue disorders such as scleroderma, dermatomyositis, and systemic lupus erythematosus may involve localized areas of the skin or deeper subcutaneous tissue and is referred to as *calcinosis circumscripta*. Mineral deposition at sites of deeper tissue injury including periarticular sites is called *calcinosis universalis*.

ECTOPIC OSSIFICATION

True extraskeletal bone formation that begins in areas of fasciitis following surgery, trauma, burns, or neurologic injury is referred to as *myositis ossificans*. The bone formed is organized as lamellar or trabecular, with normal osteoblasts and osteoclasts conducting active remodeling. Well-developed haversian systems and marrow elements may be present. A second cause of ectopic bone formation occurs in an inherited disorder, *fibrodysplasia ossificans progressiva*.

FIBRODYSPLASIA OSSIFICANS PROGRESSIVA

This is also called *myositis ossificans progressiva*; it is a rare autosomal dominant disorder characterized by congenital deformities of the hands and feet and episodic soft tissue swellings that ossify. The disorder is caused by an activating mutation in activin receptor A type 1. Ectopic bone formation occurs in fascia, tendons, ligaments, and connective tissue within voluntary muscles. Tender, rubbery induration, sometimes precipitated by trauma, develops in the soft tissue and gradually calcifies. Eventually, heterotopic bone forms at these sites of soft tissue trauma. Morbidity results from heterotopic bone interfering with normal movement and function of muscle and other soft tissues. Mortality is usually related to restrictive lung disease caused by an inability of the chest to expand. Laboratory tests are unremarkable.

Until recently, there was no effective approved medical therapy. Bisphosphonates, glucocorticoids, and a low-calcium diet have largely been ineffective in halting progression of the ossification. Palovarotene has been shown to reduce new heterotopic ossification by 60% versus historical controls but increased premature epiphyseal closure in children. In 2023, the therapy was approved in the United States for females over age 8 and males over age 10. Another potential therapeutic option, REGN2477 (also known as garetosmab), an anti–activin A antibody, is in clinical trials. Surgical removal of ectopic bone is not recommended because the trauma of surgery may precipitate formation of new areas of heterotopic bone. Dental complications, including frozen jaw, may occur following injection of local anesthetics.

Acknowledgment

The authors acknowledge the contribution of Dr. Murray J. Favus to this chapter in previous editions of Harrison's.

FURTHER READING

Boyce AM, Collins MT: Fibrous dysplasia/McCune-Albright syndrome: A rare, mosaic disease of $G\alpha_s$ activation. Endocr Rev 41:345, 2020.

De Castro LF et al: Safety and efficacy of denosumab for fibrous dysplasia of bone. N Eng J Med 388:8, 2023.

Pognolo RJ et al: Reduction of new heterotopic ossification (HO) in the open-label, phase 3 MOVE trial of palovarotene for fibrodysplasia ossificans progressive (FOP). J Bone Miner Res 38:3, 2022.

Ralston SH et al: Diagnosis and management of Paget's disease of bone in adults: A clinical guideline. J Bone Miner Res 34:579, 2019.

Shapiro JR, Lewiecki EM: Hypophosphatasia in adults: Clinical assessment and treatment considerations. J Bone Miner Res 32:1977, 2017.

Singer FR et al: Paget's disease of bone: An endocrine society clinical practice guideline. J Clin Endocrinol Metab 99:4408, 2014.

Tan A et al: Long-term randomized trial of intensive versus symptomatic management in Paget's disease of the bone: The PRISM-EZ Study. J Bone Miner Res 32:1165, 2017.

Wu CC et al: Diagnosis and management of osteopetrosis: Consensus guidelines from the osteopetrosis working group. J Clin Endocrinol Metab 102:3111, 2017.

Section 5 Disorders of Intermediary Metabolism

425

Heritable Disorders of Connective Tissue

Joan C. Marini, Fransiska Malfait

CLASSIFICATION OF CONNECTIVE TISSUE DISORDERS

Some of the most common conditions that are transmitted genetically in families are disorders that produce clinically obvious changes in the bone, cartilage, skin, or relatively acellular tissues such as tendons that have been loosely defined as connective tissues. Because of their heritability, some of the disorders were recognized as potentially traceable to mutated genes soon after the principles of genetics were introduced into medicine by Garrod and others. About half a century later, McKusick emphasized the specificity of many of the diseases for selective connective tissues and suggested that they were probably caused by mutations in genes coding for the major proteins found in those tissues. In the past several decades, mutations in several hundred different genes expressed in connective tissues have been identified as the cause of many connective tissue disorders. However, classifying the disorders on the basis of either their clinical presentations or the mutations causing them continues to present a challenge for both the clinician and the molecular biologist.

Information on the disorders has continued to develop on two levels. The initial clinical classifications suggested by McKusick and many others had to be refined as more patients were examined. For example, some patients had skin changes similar to those commonly seen in Ehlers-Danlos syndrome (EDS), but this feature was overshadowed by other features such as extreme hypotonia or sudden rupture of large blood vessels. To account for the full spectrum of presentations in patients and families, many of the disorders have been reclassified several times, dividing each into a series of subtypes.

The identification of mutations causing the diseases has developed on a parallel track. The first genes cloned for connective tissues were the two genes coding for type I collagen (*COL1A1* and *COL1A2*), the most abundant protein in bones, skin, tendons, and several other tissues. This facilitated early studies in patients with osteogenesis imperfecta (OI) that revealed mutations in type I collagen genes. Biochemical data, developed primarily with cultures of skin fibroblasts from affected individuals, demonstrated that the mutations dramatically altered the synthesis of collagen α-chains or the structure of collagen fibers. The results stimulated efforts to identify additional mutations in genes coding for structural proteins. Genes for collagens provided an attractive paradigm to search for mutations, since a series of different types of collagens were found in different connective tissues and the collagen genes were readily isolated by their unique signature sequences. Also, the collagen genes were vulnerable to a large number of different mutations because of unusual structural requirements of the protein. The search for mutations in collagen genes proved fruitful in that mutations were found in most patients with OI, in many patients with hyperextensible skin and hypermobile joints, in some patients with dwarfism, and in patients with other disorders, including some such as Alport syndrome (AS) that were not initially classified as disorders of connective tissue. Also, mutations in collagen genes were found in subset of patients presenting with osteoarthritis (OA) or osteoporosis, likely representing the mildest end of the syndromic spectrum. However, the search for mutations quickly expanded to hundreds of other genes that included genes for other structural proteins, for the posttranslational modification and processing of the structural proteins, for chaperones, and for growth factors and their receptors and other genes whose functions are still not fully understood.

In many instances, the mutations helped to define the clinical subtype of the disorder, while in others, they revealed the genetic heterogeneity of the same clinical presentations. Conversely, some patients with different manifestations were found to have mutations in the same genes. In noncollagenous genes, it was sometimes difficult to establish whether a change in the structure of a gene caused the phenotypic changes in the patients or was simply a neutral polymorphism. Therefore, there has been a continuing debate as to whether the disorders should be classified by their clinical presentations or by the causative genes. As an illustration of the problems, mutations in 552 genes have now been found associated with 771 defined disorders of the skeleton. The latest (2023) nosology for the disorders, which adopted a dyadic naming system, systematically associating a phenotypic entity with the gene it arises from, remains "hybrid" in nature in the sense that the classification is not always based on the same criteria. Some diseases are grouped based on the causal gene, others are listed together, because they share common radiographic features, and still others are brought together because of a similar clinical course (lethality) or involvement of similar parts of the skeleton. A simpler system of classification proved feasible for one rare heritable disorder of skin, epidermolysis bullosa. The disorder was first defined clinically into subtypes based on the layers of the skin that were cleaved in friction-induced blisters. Most patients in each subtype were subsequently shown to have mutations in genes expressed in the corresponding layer of skin. Even with these patients, the strength of the genotype-phenotype correlation varies and mutations have not yet been found in every patient.

The best pathway through this maze of information is probably to begin by matching the signs and symptoms in a patient with the presentations that define each clinical classification. A major focus should be on the most common disorders, recognizing that the signs and symptoms may vary among different individuals and family members with the same diagnosis. Then, attempt to reach a decision, in consultation with the patient, parents, and specialist, as to whether a DNA analysis for the probable mutation is indicated. Among the considerations are the cost, the rigor with which the clinical classification has been linked to mutated genes, the reassurance the diagnosis can bring to patients and their families, the use of the genetic information for prenatal diagnosis, and the possibility that mutation-specific therapies may be developed in the future. For individuals affected with these disorders, consulting a specialist in the disease is highly recommended to determine a multidisciplinary program for management and therapy. Patient support groups have formed for many of the diseases and are an important source of information.

Patients with the most common forms of the disorders have mutations in a limited number of genes. This chapter will focus primarily on these. Also, it will provide a brief summary of biosynthesis and structure of connective tissues that may help guide the physician from the nature of the mutations to their clinical presentations.

■ COMPOSITION OF CONNECTIVE TISSUES

Connective tissues such as skin, bone, cartilage, ligaments, and tendons are the critical structural frameworks of the body. They consist of a complex interacting extracellular matrix network of collagens, proteoglycans, and a large number of noncollagenous glycoproteins and proteins. While these precise combinations of up to ~500 potential extracellular matrix building blocks, collectively called "the matrisome," provide tissue-specific function, there are many overarching similarities in composition such as the role of composite collagen fibrils in providing strength and form, elastin fibrils and proteoglycans and other interacting proteins, and glycoproteins that fine-tune function **(Table 425-1)**. The most abundant components of many connective tissues are three similar fibrillar collagens (types I, II, and III). They have a similar tensile strength that is comparable to that of steel wires. The three fibrillar collagens are distributed in a tissue-specific manner: type I collagen accounts for most of the protein of dermis, ligaments, tendons, and demineralized bone; type I and type III are the most abundant proteins of large blood vessels; and type II is the most abundant protein of cartilage.

■ BIOSYNTHESIS AND TURNOVER OF CONNECTIVE TISSUES

Connective tissues are among the most stable components in living organisms, but they are not inert. During embryonic development, connective tissue membranes appear as early as the four-cell blastocyst to provide a structural scaffold for the developing embryo. With the development of blood vessels and skeleton, there is a rapid increase in the synthesis, degradation, and resynthesis of connective tissues. The turnover continues at a slower, but still rapid pace throughout postnatal development and then spikes during the growth spurt of puberty. During adulthood, the metabolic turnover of most connective tissues is slow, but it continues at a moderate pace in bone. With age, malnutrition, physical inactivity, and low gravitational stress, the rate of degradation of most connective tissues, especially in bone and skin, begins to exceed the rate of synthesis and the tissues shrink. In starvation, a large fraction of the collagen in skin and other connective tissues is degraded and provides amino acids for gluconeogenesis **(Chap. 345)**. In both OA and rheumatoid arthritis, there is extensive degradation of

TABLE 425-1 Constituents of Connective Tissues and Their Associated Heritable Conditions

PROTEIN	TISSUE DISTRIBUTION	DISEASE	KEY MANIFESTATIONS
Collagen I	Bone, cornea, dermis, tendon	Osteogenesis imperfecta	Bone fragility with fractures and deformity; blue sclerae; dentinogenesis imperfecta; hearing loss
		EDS (various rare types)	Joint hypermobility; skin hyperextensibility; skin fragility; soft connective tissue fragility
		Caffey disease	Subperiosteal new bone formation; soft tissue swelling; fever and irritability
Collagen II	Cartilage, vitreous	Various chondrodysplasias	Skeletal dysplasia; ocular manifestations; hearing loss; orofacial findings
Collagen III	Dermis, aorta, uterus, intestine	Vascular EDS	Arterial, intestinal, and uterine fragility; thin translucent skin; easy bruising
Collagen IV	Basement membranes	Alport syndrome (*COL4A3/A4/A5*)	Hematuria; hearing loss; ocular abnormalities
		Brain small-vessel disease (*COL4A1/A2*)	Porencephaly; intracerebral hemorrhage; retinal arterial tortuosity; (congenital) cataract; Axenfeld-Rieger anomaly; hematuria; renal cysts; muscle cramps
Collagen V	Placental tissue, bone, dermis, cornea	Classic EDS	Joint hypermobility; skin hyperextensibility; atrophic scarring
Collagen VI	Uterus, dermis, cornea, cartilage	Bethlem myopathy and Ullrich congenital muscular dystrophy	Muscle weakness; joint contractures; joint hypermobility
Collagen VII	Skin, amniotic membrane, mucosal epithelium	Dystrophic epidermolysis bullosa	Skin blistering; oral and esophageal blistering; corneal erosions
Collagen VIII	Descemet's membrane, endothelial cells	Corneal dystrophy	Corneal endothelial dystrophy; stromal edema
Collagen IX	Cartilage, vitreous	Stickler syndrome	Spondyloepiphyseal dysplasia; early-onset osteoarthritis; high myopia; vitreoretinal abnormalities; hearing loss; cleft palate; midfacial hypoplasia
Collagen X	Calcifying cartilage	Multiple epiphyseal dysplasia	Epiphyseal dysplasia; early-onset osteoarthritis
Collagen XI	Cartilage, intervertebral disk	Various chondrodysplasias	Skeletal dysplasia; ocular manifestations; hearing loss; orofacial findings
Collagen XII	Dermis, tendon, cartilage	Myopathic EDS	Joint hypermobility; congenital muscle hypotonia and/or atrophy; proximal joint contractures
Collagen XVII	Corneal epithelial cells	Junctional epidermolysis bullosa	Blistering of the skin and mucosae (mild to severe)
Collagen XVIII	Pia, blood vessels of the developing human cerebral cortex	Knobloch syndrome	Early-onset severe myopia; vitreoretinal degeneration with retinal detachment; hydrocephalus; structural brain defects; epilepsy; cognitive dysfunction
Collagen XXVII	Chondrocytes, epithelial cell layers in developing tissues, including stomach, lung, gonad, skin, cochlea, and tooth	Steel syndrome	Osteochondrodysplasia with hip dislocations; dislocations of radial heads; carpal coalition; short stature; facial dysmorphism; scoliosis
Cartilage oligomeric matrix protein (COMP)	Cartilage, tendon, ligament, bone	Pseudoachondroplasia	Short-limb dwarfism; early-onset osteoarthritis
		Multiple epiphyseal dysplasia	Mildly short stature; early-onset osteoarthritis
Elastin	Dermis, arterial wall, lung	Cutis laxa	Wrinkled, redundant, sagging inelastic skin
		Williams syndrome	Cardiovascular disease (especially supravalvular aortic stenosis); orofacial features; intellectual deficit; connective tissue abnormalities; endocrine abnormalities
Fibrillin 1	Dermis, arterial wall, lung	Marfan syndrome	Aortic root aneurysm or dissection; ectopia lentis; marfanoid habitus
		Weill-Marchesani-syndrome	Short stature; joint stiffness; lens abnormalities; cardiovascular features
		Stiff skin syndrome	Progressive rock-hard skin; flexion contractures; hypertrichosis
		Geleophysic dysplasia	Short stature; joint stiffness; thickened skin; progressive cardiac valvular disease; orofacial features
Fibrillin 2	Bruch membrane	Congenital contractural arachnodactyly (CCA) or Beals-Hecht syndrome	Tall stature; arachnodactyly; (kypho)scoliosis; pectus deformities; contractures; muscle hypoplasia; mild cardiovascular involvement; long, narrow face, highly arched palate, micrognathia, crumpled external ears
		Acromelic dysplasia	Relative short stature; brachydactyly; toe walking; early onset carpal tunnel syndrome; short palpebral fissures
Fibronectin	Dermis, tendons, ligaments	Glomerulopathy with fibronectin deposits	Glomerulopathy with fibronectin deposits
		Spondylometaphyseal dysplasia, corner fracture type	Spondylometaphyseal dysplasia characterized by flake-like, triangular, or curvilinear ossification centers at the edges of irregular metaphyses that simulate fractures; short stature

(*Continued*)

TABLE 425-1 Constituents of Connective Tissues and Their Associated Heritable Conditions (Continued)

PROTEIN	TISSUE DISTRIBUTION	DISEASE	KEY MANIFESTATIONS
Aggrecan	Cartilage	Spondyloepiphyseal dysplasia, Kimberley type	Short stature; habitus; progressive osteoarthropathy; spondyloepiphyseal dysplasia
		Short stature; advanced bone age, with or without early-onset osteoarthritis and/or osteochondritis dissecans	Short stature and advanced bone age, with or without early-onset osteoarthritis and/or osteochondritis dissecans
		Spondyloepimetaphyseal dysplasia, aggrecan type	Severe short stature; spondyloepimetaphyseal dysplasia
Decorin	Dermis, tendons, ligaments, cornea	Congenital stromal corneal dystrophy	Corneal stromal opacification; visual loss; increased corneal thickness
Biglycan	Bone, cartilage, tendons	Meester-Loeys syndrome	Aortic aneurysm or dissection; orofacial features; joint hypermobility; ventricular dilatation on brain imaging; relative macrocephaly; hip dislocation; platyspondyly; phalangeal dysplasia; dysplastic epiphyses of the long bones
		X-linked spondyloepimetaphyseal dysplasia	Severe short-trunked dwarfism; brachydactyly; spondyloepimetaphyseal dysplasia

Abbreviation: EDS, Ehlers-Danlos syndromes.

articular cartilage collagen. Glucocorticoids weaken most tissues by decreasing collagen synthesis. In some pathologic states, however, collagen is deposited in excess. With most injuries to tissues, inflammatory and immune responses stimulate the deposition of collagen fibrils in the form of fibrotic scars. In humans, as distinct from many other species, the deposition of the fibrils is largely irreversible and prevents regeneration of normal tissues in diseases such as hepatic cirrhosis, pulmonary fibrosis, atherosclerosis, and nephrosclerosis.

Structure and Biosynthesis of Fibrillar Collagens The tensile strength of collagen fibers derives primarily from the self-assembly of protein monomers into large fibril structures in a process that resembles crystallization. The self-assembly requires monomers of highly uniform and relatively rigid structure. It also requires a complex series of posttranslational processing steps that maintain the solubility of the monomers until they are transported to the appropriate extracellular sites for fibril assembly. Because of the stringent requirements for correct self-assembly, it is not surprising that mutations in genes for fibrillar collagens cause many of the heritable diseases of connective tissues.

The monomers of the three fibrillar collagens are formed from three polypeptide chains, called α chains, that are wrapped around each other into a rope-like triple-helical conformation. The triple helix is a unique structure among proteins, and it provides rigidity to the molecule. It also orients the side chains of amino acids in an "inside out" manner relative to most other proteins so that the charged and hydrophobic residues on the surface can direct self-assembly of the monomers into fibrils. The triple-helical conformation of the monomer is generated because each of the α chains has a repetitive amino acid sequence in which glycine (Gly) appears as every third amino acid. Each α chain contains ~1000 amino acids. Therefore, the sequence of each α chain can be designated as $(\text{-Gly-X-Y-})_n$, where X and Y represent amino acids other than glycine and n is >338. The presence of glycine, the smallest amino acid, in every third position in the sequence is critical because this residue must fit into a sterically restricted space in the interior of the helix where the three chains come together. The requirement for a glycine residue at every third position explains the significant clinical effects of mutations that convert a glycine residue to an amino acid with a bulkier side chain (see below). Many of the X- and Y-position amino acids are proline and hydroxyproline, which, because of their ring structures, provide additional rigidity to the triple helix. Other X- and Y-positions are occupied by charged or hydrophobic amino acids that precisely direct lateral and longitudinal assembly of the monomers into highly ordered fibrils. Mutations that substitute amino acids in some X- and Y-positions, particularly arginine-to-cysteine substitutions, can also produce genetic diseases.

The fibers formed by the three fibrillar collagens differ in thickness and length, but they have a similar fine structure. As viewed by electron microscopy, they all have a characteristic pattern of cross-striations that are about one-quarter the length of the monomers and reflect the precise packing into fibrils. The three fibrillar collagens, however, differ in sequences found in the X- and Y-positions of the α chains and therefore in some of their physical properties. Type I collagen is a heterotrimer, composed of two identical α1(I) chains and a third α2(I) chain that differs slightly in its amino acid sequence. Types II and III collagen are homotrimers, each composed of three identical α chains distinct to that type of collagen.

To deliver a monomer of the correct structure to the appropriate site of fibril assembly, the biosynthesis of fibrillar collagens involves a large number of unique processing steps **(Fig. 425-1)**. The monomer, first synthesized as a soluble precursor called *procollagen*, contains an additional globular domain at each end. As the pre-proα chains of procollagen are synthesized on ribosomes, the free N-terminal ends move into the cisternae of the rough endoplasmic reticulum (ER). Signal peptides at the N-termini are cleaved, and additional posttranslational reactions begin. Proline and lysine residues in the Y-position of the Gly-X-Y repeating triplet are hydroxylated along the length of the helix by the enzymes prolyl 4-hydroxylase (P4H1) and lysyl hydroxylase (LH1), respectively. Hydroxyproline residues are essential for the three α chains of the monomer to fold into a triple helix at body temperature. P4H1 requires ascorbic acid as an essential cofactor, an observation that explains why wounds fail to heal in scurvy **(Chap. 344)**. In scurvy, some of the underhydroxylated and unfolded protein accumulates in the cisternae of the rough ER and is degraded. Many hydroxylysine residues are glycosylated with galactose or with galactose and glucose. Also, a large mannose-rich oligosaccharide is assembled on the C-terminal propeptide of each chain. The proα chains are assembled by interactions among these C-terminal propeptides that control the selection of the appropriate partner chains to form hetero- or homotrimers and provide the correct chain registration required for subsequent formation of the collagen triple helix. After the C-terminal propeptides assemble the three proα chains, a nucleus of triple helix is formed near the C-terminus, and the helical conformation is propagated toward the N-terminus in a zipper-like manner that resembles crystallization. The folding into the triple helix is spontaneous in solution, but as discussed below, identification of rare mutations causing OI demonstrated that the folding *in cellulo* is assisted by a number of ancillary proteins that also prevent collagen fibril formation within the ER. The fully folded procollagen is then transported to the Golgi via a specific COPII vesicle process. After further modifications in the Golgi stack, the procollagen is secreted into the pericellular space where distinct proteases remove the N- and C-propeptides at specific cleavage sites. The release of the propeptides decreases the solubility of the resulting collagen ~1000-fold. The entropic energy that is released drives the self-assembly of the collagen into fibrils. Self-assembled collagen fibers have considerable tensile strength, but their strength is increased further by cross-linking reactions that form covalent bonds between α chains in one molecule and α chains in adjacent molecules.

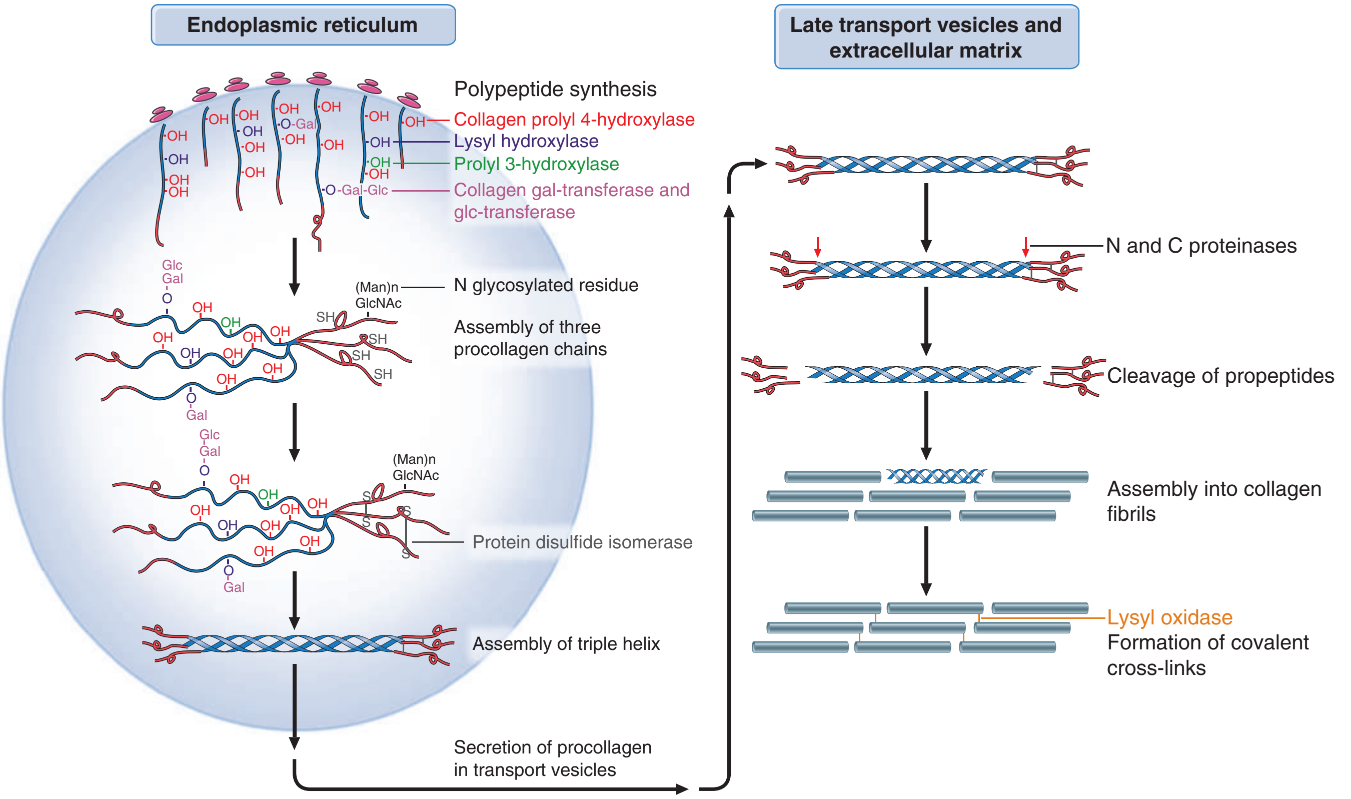
Endoplasmic reticulum
Late transport vesicles and extracellular matrix
Polypeptide synthesis
Collagen prolyl 4-hydroxylase
Lysyl hydroxylase
Prolyl 3-hydroxylase
Collagen gal-transferase and glc-transferase
OH
O-Gal
O-Gal-Glc
Glc
Gal
O
(Man)n
GlcNAc
N glycosylated residue
SH
Assembly of three procollagen chains
S
Protein disulfide isomerase
Assembly of triple helix
Secretion of procollagen in transport vesicles
N and C proteinases
Cleavage of propeptides
Assembly into collagen fibrils
Lysyl oxidase
Formation of covalent cross-links

FIGURE 425-1 Schematic summary of biosynthesis of fibrillar collagens. *(Reproduced with permission from J Myllyharju, KI Kivirikko: Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet 20:33, 2004.)*

The resulting fibers, composed of hundreds or thousands of triple-helical monomers, have some of the properties of a crystal but have innate imperfections that make them highly flexible.

Although the assembly of collagen monomers into fibers is largely a spontaneous reaction, the process in tissues is modulated by the presence of less abundant collagens (type V with type I, and type XI with type II) and by other components such as a series of small leucine-rich proteins (SLRPs). Some of the less abundant components alter the rate of fibril assembly, whereas others change the morphology of the fibers or their interactions with cells and other molecules. The presence of these other components is one explanation for why, in some tissues, the fibers are further assembled into large tendons; in others, into sheets; and in still others, into complex structures such as the hexagonal array of fibers that provide both the strength and transparency of the cornea.

Collagen fibers are resistant to most proteases, but during degradation of connective tissues, they are cleaved by specific matrix metalloproteinases (collagenases) that cause partial unfolding of the triple helices into gelatin-like structures that are further degraded by less specific proteinases.

OTHER COLLAGENS AND RELATED MOLECULES

The unique properties of the triple helix are used to define a family of at least 28 collagens that contain repetitive -Gly-X-Y- sequences and form triple helices of varying length and complexity. The proteins are heterogeneous both in structure and function, and many are the sites of mutations causing genetic diseases. For example, the type IV collagen found in basement membranes is composed of three α chains synthesized from any of six different genes. Mutations in the *COL4A3*, *COL4A4*, or *COL4A5* genes cause AS, while mutations in *COL4A1* and rarely *COL4A2* are associated with a spectrum of phenotypes including small-vessel brain disease of varying severity including porencephaly, variably associated with eye defects (retinal arterial tortuosity, Axenfeld-Rieger anomaly, cataract) and systemic findings (kidney involvement, muscle cramps, cerebral aneurysms, Raynaud phenomenon, cardiac arrhythmia, and hemolytic anemia).

Fibrillin Aggregates and Elastin In addition to tensile strength, many tissues such as the lung, large blood vessels, and ligaments require elasticity. The elasticity was originally ascribed to an amorphous rubber-like protein named elastin. Subsequent analyses, largely sparked by discoveries of mutations causing the Marfan syndrome (MFS), demonstrated that the elasticity resided in thin fibrils composed primarily of large glycoproteins named fibrillins. The fibrillins contain large numbers of epidermal growth factor–like domains interspersed with characteristic cysteine-rich domains that are also found in latent transforming growth factor β (TGF-β) binding proteins. The fibrillins assemble into long beadlike strands that also contain numerous other components including small and variable amounts of elastin, bone morphogenic proteins (BMPs), and microfibril-associated glycoproteins (MAGPs). Besides contributing to extracellular matrix structure, a major role for fibrillins in TGF-β signaling was emphasized by the discovery of mutations in genes coding for proteins involved in canonical TGF-β signaling in patients with Marfan-like manifestations, including thoracic aortic aneurysm.

Proteoglycans The resiliency to compression of connective tissues such as cartilage or the aorta is largely explained by the presence of proteoglycans. Proteoglycans are composed of a core protein to which are attached a large series of negatively charged polymers of disaccharides (largely chondroitin sulfates). At least 30 proteoglycans have been identified. They vary in their binding to collagens and other components of matrix, but specific functions have not been assigned to most. The major proteoglycan of cartilage, called aggrecan, has a core protein of 2000 amino acids that is decorated with ~100 side chains of chondroitin sulfate and keratin sulfate. The core protein, in turn, binds to long chains of the polymeric disaccharide hyaluronan to form proteoglycan aggregates, one of the largest soluble macromolecular structures in nature. Because of its highly negative charge and extended structure, the proteoglycan aggregate binds large amounts of water and small ions to distend the three-dimensional arcade of collagen fibers found in the same tissues. It thereby makes the cartilage resilient to pressure.

OSTEOGENESIS IMPERFECTA

OI is a phenotypically and genetically heterogeneous generalized connective tissue disorder. The hallmark features of OI are increased susceptibility to skeletal fractures, bone deformity, and growth deficiency. Bone fragility is based on decreased bone mass and increased bone brittleness due to defective mineralization. Secondary features of OI are highly variable even within a type and include blue sclerae, dentinogenesis imperfecta, hearing loss, basilar invagination, pulmonary function impairment, cardiac valve abnormalities, and ligamentous laxity. Most patients have defects in the structure or quantity of type I collagen.

Classification OI was originally classified into *congenita* and *tarda* subtypes depending on the age of symptom onset. Sillence proposed the classification that bears his name for four types based on clinical and radiologic findings and mode of inheritance. The extension of the Sillence classification was first based on distinctive bone histology (types V and VI OI) and subsequently on the discovery of new recessive genes (types VII–XXII). The debate between classification by phenotypic severity or gene defects has resulted in clinical and genetic classifications. The clinical classification can be useful for management but results in different type assignments in the same family or even in the same individual over their lifetime. The genetic classification **(Table 425-2)** groups patients by the causative gene. Because related causative genes were discovered close in time to each other, the genetic classification consequently groups types by overall mechanism and features OI as a collagen-related disorder.

Types I–IV OI are due to quantitative or structural defects in type I collagen itself. Type I is the mildest subtype, with reduced quantity of structurally normal collagen, and can produce mild or inapparent skeletal deformities. Most patients have distinctly blue sclerae. Types II, III, and IV are all caused by structural defects in one of the type I collagen α chains. Type II produces bone so brittle that infants have in utero fractures of ribs and long bones and die in the perinatal period. Type III is progressively deforming with moderate to severe bone deformity, and type IV has mild to moderate bone fragility and secondary features. Subsequent rare recessive OI types are all collagen-related. Types V and VI (*ITITM5* and *SERPINF1*) particularly compromise matrix mineralization. Types VII, VIII, and IX (*CRTAP, P3H1,* and *PPIB*) represent defects in the components of the procollagen prolyl 3-hydroxylation complex that modifies collagen posttranslationally. Types X–XII and XXI (*SERPINH1, FKBP10, BMP1,* and *KDELR1*) have compromised procollagen processing and cross-linking. The final grouping of types XIII–XVIII (*SP7, TMEM38B, WNT1, CREBL1, SPARC, MBTPS2, TEMTSA, MESD, CCDS134*) alter osteoblast differentiation and impair collagen matrix quality.

The clinical heterogeneity of affected individuals within a particular OI type and even with the same mutation is not understood, with unknown modifying factors presumably involved. Among adults with OI, women are prone to fracture during pregnancy and after menopause. Some variants of mild OI are first detected perimenopausally and must be distinguished from postmenopausal osteoporosis.

Incidence In North America and Europe, the estimated incidence of OI is 1 per 10,000–15,000 births, based on a combination of cases recognized at birth and population surveys for milder cases. In populations with a high level of consanguinity or a founder mutation, the incidence of the rare recessive forms of OI is a significant addition to the prevalence of dominant collagen defects.

Effects on Tissue Systems The phenotypic features of OI are highly variable, even within the types caused by defects in type I collagen. The following section generally focuses on dominant forms comprising the majority of cases, except as specified, but the descriptions can be generalized to a large extent.

Musculoskeletal Effects Bone in OI is both weak and brittle. At the mildest end of the spectrum (type I OI), individuals may have only several childhood fractures and be limited only from contact sports. More severe forms of OI require bone to be partially unloaded with assistive devices such as walkers or canes; many severe patients use electric chairs for both the weight bearing and the normal speed of mobility. In dominant OI, fragility fractures often decrease sharply after adequate bone mass is gained at puberty. Radiographs generally show osteopenia in all types, with disordered matrix organization detected most easily in lower long bones in moderate and severe forms. In lethal OI, radiographs show continuous beading of ribs from healing fractures and crumpled and undertubulated long bones. Lateral skull radiographs may show islands of Wormian bones, even in mild forms. The appearance of “popcorn” at the metaphyses of long bones occurs in many type III and IV children and coincides with increased growth deficiency. Often these bones are so soft that normal muscle pull can produce severe deformities. Kyphoscoliosis is associated with vertebral compressions but is not prevented by bisphosphonates, suggesting a contribution from ligamentous laxity.

OI bone is weak, in that it fractures with a lower load than normal, and brittle, in that it does not tolerate postyield displacement and snaps like chalk. The brittleness results from the paradoxical increased mineralization of OI bone. While dual-energy x-ray absorptiometry (DXA) bone density measurements uniformly return a reduced value for OI bone, it is performed with a phantom and detects mineral crystals that are in proper alignment. In contrast, quantitative backscattered electron imaging or three-dimensional (3D) computed tomography (CT), which detect all mineral in 3D, reveals that both dominant and recessive (except types XIV and XV) OI bone is hypermineralized. On histomorphometry, dominant OI bone has proper formation of lamellae but increased turnover, causing decreased bone volume. Type V OI has mesh-like bone lamellae, as well as a dislocated radial head, and may have hyperplastic callus formation, while type VI OI has distinctive fish scale lamellae on polarized light microscopy.

Many OI patients across the severity spectrum have increased ligamentous laxity. Patients with defects in processing the N-terminal propeptide of type I procollagen have large and small joint hypermobility similar to EDS. Muscle weakness of unknown etiology also occurs in OI, and the weakness and ligamentous laxity contribute to delayed motor development.

Pulmonary The leading cause of death in OI is pulmonary disease. Young children with severe OI often have repeated pneumonia; restrictive or obstructive disease develops in most older children and adults. Pulmonary function is impaired by marked scoliosis and chest wall deformity but also arises from intrinsic defects of lung parenchyma containing type I collagen, as shown by declining pulmonary function over time in children without scoliosis. More recent studies have reinforced the significant role of intrinsic lung abnormalities in OI, demonstrating reduced gas exchange, reduced airflow in small airways, and atelectasis in most individuals with collagen structural abnormalities. Bronchial thickening at the level of subsegmental bronchi in almost all patients further indicates the critical role of small airways. Mice with null *CRTAP* mutations (type VII OI) have abnormal alveolar development, and patients with recessive forms also have pulmonary complications. Evaluation of even asymptomatic moderate to severe OI patients by spirometry should initiate standard pulmonary interventions.

Cardiovascular Cardiovascular effects of OI manifest predominantly in adults. With type I collagen as a major component of matrix in cardiac valves and aortic wall, the most frequent manifestations are valvular, especially mitral regurgitation and aortic root dilatation. Impaired mechanical properties occasionally lead to aortic dissection. Echocardiography is appropriate with heart murmurs or cardiac symptoms and every 3–5 years in asymptomatic patients.

Dentinogenesis Imperfecta Dentinogenesis imperfecta (DI) is associated with types III and IV OI and recessive types with collagen processing defects. Tooth agenesis, especially of premolars, is also found in types III/IV OI. Teeth with disturbed formation of dentin during development may be translucent gray or have yellowish or brownish discoloration. Defects are manifest predominantly in primary teeth; detection in secondary teeth may require radiographs to identify

TABLE 425-2 Different Types of Osteogenesis Imperfecta (OI)

	OI TYPE	INHERITANCE	DEFECTIVE GENE	PROTEIN	OMIM	LOCUS	HYPERMINERALIZATION	DISTINGUISHING FEATURES
Defects in collagen structure and processing	I	AD	*COL1A1*	Collagen α1	166200	17q21.33	Yes	Loss of function of one of the COL1A1 alleles
	II–IV	AD	*COL1A1, COL1A2*	Collagen α1 or α2	166210, 259420, 166220	17q21.33, 7q21.3	Yes	Structural defects in collagen helix or C-propeptides
Procollagen processing defects	OI/EDS	AD	*COL1A1, COL1A2*	Procollagen α1 or α2	NA	17q21.33, 7q21.3	Yes	Defects in 90 residues at N-terminus of collagen helix that decrease pN-processing
	HBM	AD	*COL1A1, COL1A2*	Collagen α1 or α2	NA	17q21.33	Yes	Defects in C-propeptide cleavage site, DXA normal to increased
	XII	AR	*BMP1*	BMP1	614856	8p21.3	Yes	Deficiency of C-propeptidase; skeletal deformity severe plus rhizomelia, scoliosis; overlaps AD defects in type I collagen C-propeptide cleavage site
Bone mineralization defects	V	AD	*IFITM5*	BRIL (BRIL5′ MALEP)	610967	11p15.5	Yes	Calcification of interosseous membrane, dense metaphyseal band, hyperplastic callus, mesh-like pattern in lamellar bone
	Atypical VI	AD	*IFITM5*	BRIL (BRIL Ser40Leu)	610967	11p15.5	Yes	Increased osteoid, fish scale pattern in lamellar bone, increased ALP levels in childhood, symptom onset at birth
	VI	AR	*SERPINF1*	PEDF	613982	17p13.3	Yes	PEDF deficiency, increased osteoid, fish scale pattern in lamellar bone, increased ALP levels in childhood, onset after age 1 year
Defects in collagen modification	VII	AR	*CRTAP*	CRTAP	610682	3q22.3	Yes	Absent procollagen prolyl 3-hydroxylation; full OM, rhizomelia, white sclerae
	VIII	AR	*LERPE1*	P3H1	610915	1p34.2	Yes	Absent procollagen prolyl 3-hydroxylation; full OM, rhizomelia, “popcorn” metaphyses; white sclerae
	IX	AR	*PPIB*	CyPB	259440	15q22.31	Yes	Absent procollagen prolyl 3-hydroxylation; helix modification varies, without rhizomelia, white sclerae
	XIV	AR	*TMEM38B*	TRIC-B	615066	9q31.2	No	Decreased modification of collagen helix; Bedouin founder mutation; normal teeth, hearing
Defects in collagen folding and cross-linking	X	AR	*SERPINH1*	HSP47	613848	11q13.5	ND	Severe skeletal deformity, blue sclerae, DI, skin abnormalities, inguinal hernias
	XXI	AR	*KDELR2*	KDEL ER protein retention receptor; interacts with HSP47	619131	7p22.1	ND	Short stature, progressive skeletal deformation; dysmorphic facies, failure to thrive, hypotonia, joint hypermobility
	XI	AR	*FKBP10*	FKBP65	610968	17q21.2	Yes	May have congenital contractures
	NA	AR	*PLOD2*	LH2	609220	3q24	Yes	Progressive joint contractures
Osteoblast function and differentiation	XIII	AR	*SP7*	OSTERIX	613849	12q13.13	ND	Severe skeletal deformity, delayed tooth eruption, facial hypoplasia
	XV	AD/AR	*WNT1*	WNT1	615220	12q13.12	No	AR cases have severe progressive OI; may have neurologic defects
	XVI	AR	*CREB3L1*	OASIS	616215	11p11.2	Yes	Defect in RIP pathway; Oasis substrate of S1P/S2P; severe skeletal fragility and deformity
	XVII	AR	*SPARC*	SPARC	616507	5q33.1	Yes	Progressive severe bone fragility; hypotonia, joint laxity
	XVIII	XR	*MBTPS2*	S2P	301014	Xp22.12	Yes	X-linked OI, defect in RIP pathway; moderate to severe fragility, bowing; rhizomelia
Unclassified disorders	XIX	AR	*TEMT5A*	FAM46A	617952	6q14.1	ND	Defect in BMP/TGF-β signaling pathway; poly-adenylates transcripts of type I collagen, *SERPINF1*, and *SPARC*; severe to lethal OI; hyperlaxity, motor delay
	XX	AR	*MESD*	LRP chaperone MESD	618644	15q25.1	ND	Progressive deforming OI; severe to lethal; survivors have dental disorganization and intellectual disability; also classified with LRP5/6-related disorders
	XXII	AR	*CCDC134*	Coiled-coil domain-containing protein 134	618788	22q13.2	ND	Severe OI with pseudoarthrosis; could also be classified with MAPK/ERK skeletal dysplasias

Abbreviations: AD, autosomal dominant; ALP, alkaline phosphatase; AR, autosomal recessive; BMP, bone morphogenetic protein; DI, dentinogenesis imperfect; DXA, dual-energy x-ray absorptiometry; EDS, Ehlers-Danlos syndrome; HBM, high bone mass; NA, not applicable; ND, not determined; OI, osteogenesis imperfecta; OM, overmodification; OMIM, Online Mendelian Inheritance in Man; TGF, transforming growth factor.

characteristic narrow or obliterated pulp chambers. Crumbling at the dentin-enamel junction may require capping of teeth. Hypoplastic maxilla and relative mandibular prognathism in moderate to severe OI can result in type III malocclusion and impair normal chewing, requiring surgical correction.

Hearing Loss About half of patients with types I, III, and IV OI develop hearing loss, but its incidence in recessive types is unknown. Hearing loss usually begins in the second decade and progresses. The initial conductive loss, based on changes in the inner ear leading to stapes footplate fixation, can evolve into a mixed conductive and sensorineural loss. Regular screening allows referral for hearing aids, stapes surgery, or cochlear implants, as appropriate.

Other Features A variable intensity of blue or grayish sclerae is a well-known feature of OI. The color is most striking with collagen defects, especially types I and II OI and defects that affect N-terminal procollagen processing. Blue sclerae often occur in other connective tissue disorders such as EDS or MFS and may occur in individuals without connective tissue defects. Severe neonatal OI with white sclerae should prompt consideration of recessive forms, especially prolyl 3-hydroxylation defects. Abnormalities of the skull base, such as platybasia and basilar invagination, sometimes progress to clinically devasting basilar impression. Patients with height *Z*-scores of <–3 should be CT scanned at 3- to 5-year intervals. Significant growth deficiency is a cardinal feature of OI, ranging from minimally shorter than siblings in mild forms to greater extents in some severe cases, with adults shorter than 5-year-old children. There is both end-organ resistance to growth hormone (GH) and defective transition to bone at the growth plate. Types I and IV OI are often responsive to recombinant GH therapy.

Molecular Defects The great majority (80–85%) of cases of OI are caused by heterozygous mutations in either of the genes coding for the chains of type I procollagen, *COL1A1* or *COL1A2* (Table 425-2). Although thousands of unique mutations have been identified in type I collagen, they fall into several structural types. Null mutations in collagen chains are less detrimental than structural defects. Null mutations in *COL1A1* result in about half the normal level of collagen synthesis, but the collagen in matrix is structurally normal. These patients have mild type I OI. Null *COL1A2* mutations are rare, leading to an EDS-like condition with progressive cardiac-valvular defects when present in homozygous state.

Mutations that produce structural changes in type I collagen α chains cause types II, III, and IV OI. The most common of these are mutations resulting in substitutions for glycine residues required at every third residue along the helix. In effect, any of the 338 glycine residues in the helical domain of either the proα1 or proα2 chain of type I procollagen is a potential site for a disease-producing mutation. Other mutations affect the splicing of the exons encoding the α chains. Because each collagen exon encodes a discrete set of Gly-X-Y triplets, the abnormal splice products are most often in-frame and cause severe structural abnormalities. Use of alternative splice sites may lead to premature termination, mimicking null mutations, and a milder phenotype. Structural abnormalities in the procollagen helical region delay collagen folding and expose chains to posttranslational hydroxylation/glycosylation for a longer time. The abnormal procollagen triggers a cascade of intracellular and extracellular events including delayed collagen folding, ER stress, abnormal interaction with noncollagenous molecules, impaired osteoblast development and cross-talk with osteoclasts, and abnormal mineralization. There are some special sets of procollagen structural mutations with distinct mechanisms within types II, III, and IV. Mutations in the C-propeptide significantly delay chain assembly, and resulting procollagen is mislocalized to the ER lumen. Some of this procollagen is targeted for degradation by the ER-associated proteosomal pathway, while the secreted molecules delay pericellular processing of the C-propeptide. Mutations in the C-propeptide cleavage site itself prevent processing of the propeptide, leaving pC-collagen to be incorporated into matrix. This affects matrix mineralization, resulting in an unusual high bone mass form of OI that falls at the milder end of the type IV OI phenotype. Not surprisingly, null mutations in the C-propeptidase enzyme, BMP1, cause recessive type XII OI. Type XII OI is a severe condition because BMP1 is the cleavase for types I, II, and III procollagens and the glycoprotein decorin, which is a regulator of fibrillogenesis. Processing defects of the N-propeptide occur in the cleavage site itself or the 90 helix residues at the amino end. The persistence of the N-propeptide on a fraction of the molecules interferes with the self-assembly of normal collagen so that thin and irregular collagen fibrils are formed. They cause extreme laxity of large and small joints, intensely blue sclerae, and an OI severity comparable to type III/IV. Rare substitutions of charged amino acids (Asp, Arg) or a branched amino acid (Val) in X- or Y-positions produce lethal phenotypes, apparently because they are located at sites for lateral assembly of the monomers or binding of other components of the matrix.

Starting in 2006, a series of noncollagenous genes have been identified that cause (mostly) recessive OI. Importantly, all the genes have encoded proteins or cellular processes related to collagen, shifting the OI paradigm to dominant OI caused by collagen defects or *IFITM5* and recessive OI caused by proteins related to collagen modification, processing, folding, and cross-linking and osteoblast differentiation. The largest group of patients with OI not caused by collagen gene mutations have types V and VI OI, affecting bone mineralization. Type V OI, with dominant inheritance, is unusual in that all patients have the same recurrent mutation at the 5′-end of *IFITM5,* which generates a novel start codon in the transmembrane protein BRIL. The gain-of-function mutation causes distinctive radiologic (ossification of interosseus membrane and dense metaphyseal band) and phenotypic findings (hypertrophic callus). Osteoblasts with type V OI have increased mineralization and differentiation in culture. Type VI OI is a recessive form caused by null mutations in PEDF, a collagen-interacting molecule with a known antiangiogenic effect. A connection between types V and VI OI has been revealed by a set of patients with a BRIL p.S42L substitution who have clinical, histologic, serum marker, and phenotypic features of type VI OI. Both type VI OI osteoblasts and BRIL p.S42L osteoblasts have decreased cellular mineralization and SERPINF1 expression, while classic type V OI osteoblasts have the opposite findings. All three types decrease collagen production.

Types VII, VIII, and IX OI are severe recessive forms caused by deficiency of one of the components of the procollagen prolyl 3-hydroxylation complex, P3H1, CRTAP, or cyclophilin B (*PPIB*/CyPB). This complex 3-hydroxylates one proline residue per α chain, most critically α1(I)P986, in contrast to the proline 4-hydroxylation of multiple helical residues by P4H1. In murine models, loss of complex function results in a severe phenotype, while mutation of the P986 residue impairs collagen cross-linking and fine-tuning of collagen alignment in fibrils. The phenotype of these patients is distinctive for white sclerae, rhizomelia, and lack of relative macrocephaly; they share the bone fragility, high bone turnover, and elevated bone mineralization of classical OI.

Some recessive OI types that impair osteoblast function are caused by mutation in genes not previously understood to affect bone. Regulatory intramembrane proteolysis (RIP) is well known for its role in cholesterol synthesis, in which cells transport regulatory proteins from the ER membrane to the Golgi membrane in times of cell stress, where S1P and S2P Golgi proteases sequentially cleave the transcription factors, activating them to enter the nucleus. X-linked type XVIII OI with defective *MBTPS2*/S2P and type XVI OI with deficiency of an RIP substrate Oasis, a member of the ATF6 family of stress sensors, indicate the importance of RIP for bone formation (Table 425-2). For more recently identified OI types, the relationship of the causative gene to collagen has extended the spectrum of matrix abnormality. FAM46A (type XIX OI) polyadenylates collagen transcripts; defects lead to collagen deficiency approaching a null COL1A1 allele that does not fully explain the severe disorganization of collagen in bone matrix. Since FAM46A also polyadenylates the transcripts of other OI-causative genes, SERPINF1 and SPARC, the matrix defect may be multifactorial. For type XX OI, MESD is a direct cytosolic chaperone of pro-alpha1(I) chains, but several bone features that are not typical of OI, such as oligodontia and cartilage remnants, may reflect MESD's role as a chaperone for LRP5. KDELR2 defects (OI type XXI) have impaired retrograde transport of ER resident proteins, such as HSP47, because KDELR2 is a component of CopI vesicles. Collagen quantity in matrix is lower because of reduced collagen expression, but the impaired collagen fibrillogenesis may be caused by increased interaction of HSP47 with collagen monomers in the extracellular space. The causative gene for type XXII OI, *CCDC134*, is involved in the MAPK signaling pathway. The osteoblasts of affected individuals have increased ERK1/2 phosphorylation as well as reduced COL1A1 and osteopontin expression.

Inheritance and Mosaicism in Germline Cells and Somatic Cells Types I–V OI are inherited as autosomal dominant traits, while the rare forms are mostly recessive. Many patients with mild dominant OI represent familial traits, while sporadic new mutations are often responsible for dominant severe or lethal cases. Germline mosaicism in one parent may be the etiology of a severe dominant mutation in the child; in this circumstance, a second child may be affected with the same dominant mutation from unaffected parents. Recessive mutations in genes causing the rare forms of OI lead to more severe clinical outcomes; many of these offspring do not survive childhood, but moderately to severely affected young adults show us that these conditions must also be considered.

Diagnosis OI is usually diagnosed on the basis of clinical and radiographic criteria. The presence of fractures together with blue sclerae, DI, or family history of the disease is usually sufficient to make the diagnosis. X-rays reveal a decrease in bone density that can be verified by DXA bone densitometry, as well as characteristic deformities of long bones, thorax, and cranium. The differential diagnosis varies with age, including battered child syndrome, nutritional deficiencies, malignancies, and other inherited disorders such as chondrodysplasias and hypophosphatasia that can have overlapping presentations. A molecular diagnosis is now routinely obtained using targeted candidate gene sequencing, sometimes beginning with the dominant collagen and *IFITM5* panel. Skeletal disorder gene panels and whole-exome sequencing may be financially advantageous and pick up additional skeletal gene variants as well as the OI-causative gene. Although almost all cases can be diagnosed by sequencing, some may require bone histology and exome sequencing.

TREATMENT

Osteogenesis Imperfecta

Therapy should be directed toward maximizing the function of each individual, which includes decreasing fractures and deformity that interfere with function. Physical and occupational therapy are critical modalities. They are most commonly utilized after severe fractures or major surgery and should also be engaged consistently throughout the life span for maximizing mobility, functions of daily living, and the extent of physical conditioning possible. Water therapy is particularly useful at all ages. Diet should include adequate intake of calcium and vitamin D. Many patients are underweight for height as young children but overweight as adults, and nutritional management may be useful. Orthopedic procedures are required for deformities of long bone that interfere with standing or walking or when a bone has sustained repeated fractures. Intramedullary rods are often inserted when children are ready to stand and as needed thereafter to keep bone segments in good alignment and provide partial unloading of weight from bones. If scoliosis progresses, stabilization of the spine may be needed to maintain the curve at <60°. Medical management should also include presymptomatic screening for hearing loss, cardiac valve dysfunction, pulmonary function, and, in severe individuals, basilar invagination.

Drugs that have been developed for the therapy of postmenopausal osteoporosis are beneficial for some patients. Bisphosphonates, antiresorptive drugs that inhibit osteoclasts, increase DXA bone density and relieve vertebral compressions in most patients. They are regarded as a mainstay of care in many pediatric centers. However, several Cochran reports have not supported a clear reduction in fracture rate or bone pain from their use, and the dosing and duration of use are controversial. Currently, drugs with a bone-forming mechanism are in trials for OI, especially monoclonal antibodies to sclerostin that relieve its inhibition of osteoblast Wnt/β-catenin signaling, TGF-β inhibitors, and a PTH analogue that stimulates osteoblasts and is most beneficial for adults with milder OI. Potential therapies under investigation in animal models include chemical chaperones and mesenchymal stem cell therapy, and, currently in murine models, correction of specific defects in collagen by CRISPR.

EHLERS-DANLOS SYNDROMES

The Ehlers-Danlos syndromes (EDS) comprise a genetically heterogeneous group of heritable conditions that share several characteristics such as soft and hyperextensible skin, abnormal wound healing, easy bruising, and joint hypermobility. Additional clinical features that differ among the EDS types include fragility of soft tissues, blood vessels, and hollow organs and involvement of the musculoskeletal system and the eye. Mutations in genes coding for fibrillar collagens (type I, III, or V) are found in many patients, but other genes are affected in rare forms.

Classification Several types of EDS have been defined, based on clinical characteristics, mode of inheritance, and molecular defects (Table 425-3), and the classification of these types has been a dynamic process. The current classification defines 13 clinical EDS types that are caused by alterations in 19 different genes, but a recent study described another genetically distinct EDS type, bringing the total number of EDS-associated genes to 20. The EDS classification guides the clinical diagnosis, molecular confirmation, and genetic counseling of affected individuals and their family members.

Incidence An incidence of about 1 in 5000 individuals for all forms of EDS was proposed, with no apparent ethnic predisposition. The diagnosis of hypermobile EDS is more common in females than in males, but whether this is due to an increased incidence or more severe manifestation is unknown. The incidence for other types of EDS is similar in males and females. With incidences of 1 in 20,000 and 1 in 50,000–200,000 respectively, classic and vascular EDS are the most common genetically elucidated types of EDS. For the other types of EDS for which causative variants have been identified, there are no incidence estimates, but the numbers of people who have been reported worldwide with these disorders range between ~5 and ~100 individuals per EDS type. Patients with milder forms frequently do not seek medical attention.

Skin One of the principal features of EDS is skin hyperextensibility, that is, the skin stretches easily but snaps back after release. The skin often has a smooth, soft, or velvety feel to it and can be thin and translucent. It is fragile and tears easily, even after minor trauma, and heals slowly. Widened and thin atrophic scars are frequently observed in different types of EDS. Especially in classic EDS, atrophic scarring may be widespread, especially over pressure points and exposed areas such as the forehead, elbows, knees, and shins, with marked widening of the scars, which are covered by a very thin inelastic skin (papyraceous scars). Individuals with vascular EDS usually do not have a velvety hyperextensible skin, but skin can be thin and translucent with visible superficial veins. Easy bruising is common to most types of EDS and may manifest itself as spontaneous or recurring hematomas. These may cause discoloration of the skin due to deposition of hemosiderin, often referred to as "hemosiderotic" scars, especially in classic, vascular, and periodontic EDS.

Ligament and Joint Changes Joint hypermobility, another cardinal sign, is variable in severity and usually, but not always, generalized. While often an "asset" in childhood, it can become a serious burden over time, often complicated by repetitive subluxations, dislocations, sprains, and chronic joint pain that is difficult to treat. Other observed musculoskeletal features include congenital bilateral hip dislocation,

TABLE 425-3 Different Types of Ehlers-Danlos Syndrome (EDS)

	EDS TYPE	INHERITANCE	OMIM	LOCUS	GENE	PROTEIN	KEY MANIFESTATIONS
Defects in collagen primary structure and collagen processing	Classic EDS (cEDS)	AD	130000 130010	9q34.3 2q32.2	*COL5A1* *COL5A2*	Proα1(V) Proα2(V)	Skin hyperextensibility with atrophic scarring Generalized joint hypermobility
	Classic EDS (cEDS)	AD	/	17q21.33	*COL1A1*	Proα1(I) p.Arg312Cys	Skin hyperextensibility with atrophic scarring Generalized joint hypermobility Arterial rupture at young age
	Vascular EDS (vEDS)	AD	130050	2q32.2	*COL3A1*	Proα1(III)	Arterial rupture at young age Spontaneous sigmoid colon perforation in the absence of known colon disease Uterine rupture during third trimester of pregnancy Carotid-cavernous sinus fistula (in the absence of trauma)
	Arthrochalasia EDS (aEDS)	AD	130060 130060	17q21.33 7q21.3	*COL1A1* *COL1A2*	Proα1(I) Proα2(I)	Congenital bilateral hip dislocation Severe generalized joint hypermobility with multiple dislocations Skin hyperextensibility
	Dermatosparactic EDS (dEDS)	AR	225410	5q35.3	*ADAMTS2*	ADAMTS2	Extreme skin fragility with congenital or postnatal tears Craniofacial features Progressively redundant, lax skin with excessive skinfolds Increased palmar wrinkling Severe bruisability Umbilical hernia Postnatal growth retardation with short limbs Perinatal complications related to tissue fragility
	Cardiac-valvular EDS (cvEDS)	AR	225320	7q21.3	*COL1A2*	proα2(I)	Severe progressive cardiac-valvular insufficiency Skin involvement Joint hypermobility
Defects in collagen folding and collagen cross-linking	Kyphoscoliotic EDS (kEDS-PLOD1) Kyphoscoliotic EDS (kEDS-FKBP14)	AR AR	225400 614557	1p36.22 7p14.3	*PLOD1* *FKBP14*	Lysylhydroxylase 1 FKBP22	Congenital muscle hypotonia Congenital or early-onset kyphoscoliosis Generalized joint hypermobility with (sub)luxations
Defects in structure and function of myomatrix, the interface between muscle and ECM	Classic-like EDS type 1 (clEDS1)	AR	606408	6p21.33-p21.32	*TNXB*	Tenascin XB	Skin hyperextensibility with velvety skin texture and absence of atrophic scarring Generalized joint hypermobility Easily bruisable skin/spontaneous ecchymoses
	Myopathic EDS (mEDS)	AD/AR	616471	6q13-q14	*COL12A1*	Proα1(XII)	Congenital muscle hypotonia and/or muscle atrophy Joint contractures Joint hypermobility

Defects in glycosaminoglycan biosynthesis	Spondylodysplastic EDS (spEDS-B4GALT7)	AR	130070	5q35.3	*B4GALT7*	Galactosyltransferase I β4GalT7	Short stature (progressive in childhood) Muscle hypotonia (ranging from severe congenital to mild later-onset) Bowing of limbs Skeletal dysplasia
	Spondylodysplastic EDS (spEDS-B3GALT6)	AR	615349	1p36.33	*B3GALT6*	Galactosyltransferase II β3GalT6	
	Musculocontractural EDS (mcEDS-CHST14)	AR	601776	15q15.1	*CHST14*	Dermatan-4 sulfotransferase-1	Congenital multiple contractures (typically adduction/flexion contractures and talipes equinovarus) Craniofacial features Skin hyperextensibility, easy bruising, skin fragility with atrophic scars Increased palmar wrinkling
	Musculocontractural EDS (mcEDS-DSE)	AR	615539	6q22.1	*DSE*	Dermatan sulfate epimerase-1	
Defects in complement pathways	Periodontal EDS (pEDS)	AD	130080	12p13.31	*C1R* *C1S*	C1r C1s	Severe and intractable early-onset periodontitis Lack of attached gingiva Pretibial plaques
Defects in intracellular processes	Spondylodysplastic EDS (spEDS-SLC39A13)	AR	612350	11p11.2	*SLC39A13*	ZIP13	Short stature (progressive in childhood) Muscle hypotonia (ranging from severe congenital to mild later-onset) Bowing of limbs Skeletal dysplasia
	Brittle cornea syndrome (BCS)	AR	229200 614170	16q24 4q27	*ZNF469* *PRDM5*	ZNF469 PRDM5	Thin cornea with/without rupture Early-onset progressive keratoconus and/or keratoglobus Blue sclerae
Unclassified	Classic-like EDS type 2 (clEDS2)	AR	618000	7p13	*AEBP1*	AEBP1 (ACLP)	Skin hyperextensibility with atrophic scarring Generalized joint hypermobility Foot deformities Early-onset osteopenia
Unknown	Hypermobile EDS (hEDS)	? (AD)	130020	?	*?*	?	Generalized joint hypermobility Systemic manifestations of generalized connective tissue fragility Musculoskeletal complaints Positive family history Exclusion of other EDS types and other joint hypermobility-associated conditions

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; ECM, extracellular matrix; OMIM, Online Mendelian Inheritance in Man.

spine deformities (scoliosis, kyphosis), pectus deformities (pectus carinatum, pectus excavatum), club feet and other contractures, and in some rare types, a (mild) skeletal dysplasia. Muscle hypotonia is observed in a number of EDS types and, in combination with joint laxity, may cause floppy infant syndrome or a delay in motor development.

Other Features Signs of more generalized connective tissue weakness and fragility can be observed in varying degrees and may help to distinguish between the different EDS types. Rupture of medium and large-sized arteries is typical of vascular EDS but has been reported in a few other types as well, i.e., classic and kyphoscoliotic type. Vascular EDS patients are also at increased risk for rupture of the gastrointestinal tract, especially the sigmoid colon, the gravid uterus, and, more rarely, other internal organs such as liver or spleen. Valvular defects and aortic root dilatation are rare and are also restricted to some of the rarer types of EDS. Obstetrical and pelvic complications such as cervical insufficiency, premature rupture of membranes, vaginal lacerations, and organ prolapses (uterus, bladder, rectum) may occur. Sclerae may be blue, and more severe ophthalmologic complications, including keratoconus, keratoglobus, and scleral or corneal rupture, may be observed in some rare types.

Molecular Defects Subsets of patients with different types of EDS have mutations in the structural genes for fibrillar collagen types I, III, and V (Table 425-3). About 90% of classic EDS patients harbor a heterozygous mutation in *COL5A1* or *COL5A2* coding for type V collagen, a minor collagen found in association with type I collagen. Heterozygous mutations in the *COL3A1* gene for type III collagen, which is abundant in the blood vessel wall, are responsible for vascular EDS. Arthrochalasia EDS is caused by heterozygous mutations in either *COL1A1* or *COL1A2* that make type I procollagen resistant to cleavage by the procollagen N-proteinase ADAMTS2, whereas dermatosparaxis EDS is caused by biallelic mutations in the gene that codes for the ADAMTS2 itself, thereby reducing its enzyme activity. The persistence of the N-propeptide causes the formation of collagen fibrils that are thin and irregular. Other specific mutations in either *COL1A1* or *COL1A2* give rise to a few rare types of EDS. These include the cardiac-valvular type, which is caused by biallelic *COL1A2* mutations, leading to a complete absence of α2(I) chains. Patients with this condition are at risk for severe, progressive cardiac-valvular disease necessitating valve replacement. A specific arginine-to-cysteine substitution in the type I collagen α chain (p.Arg312Cys) is associated with an EDS phenotype that resembles that of classic EDS, but patients appear at increased risk for vascular rupture of medium-sized arteries. A few patients with a phenotype that couples EDS with signs of moderate to severe myopathy harbor heterozygous or homozygous mutations in *COL12A1*, coding for type XII collagen, a fibril-associated collagen with interrupted triple helices. Kyphoscoliotic EDS is caused by biallelic mutations either in the *PLOD1* gene, which encodes procollagen-lysine 5-dioxygenase (lysyl hydroxylase 1), an enzyme required for formation of stable cross-links in collagen fibers, or in the *FKBP14* gene, which encodes FKBP22, an endoplasmic resident molecular chaperone that acts as a quality control on the folded triple helix of type III collagen. Some patients with clinical characteristics that resemble those of classic EDS harbor biallelic mutations in either *TNXB*, encoding tenascin X, an extracellular matrix glycoprotein that appears to regulate the assembly of collagen fibers, or in *AEBP1*, which encodes the extracellular matrix–associated adipocyte enhancer-binding protein (AEBP1), which assists in collagen polymerization. Spondylodysplastic EDS is caused by biallelic mutations in *B3GALT7*, coding for galactosyltransferase I, or in *B3GALT6*, coding for galactosyltransferase II, both key enzymes in the biosynthesis of the linker region of glycosaminoglycans. Musculocontractural EDS results from mutations in genes coding enzymes responsible for dermatan biosynthesis: *CHST14*, dermatan 4-*O*-sulfotransferase 1, and *DSE*, dermatan sulfate epimerase. A rare spondylodysplastic type of EDS is caused by biallelic mutations in *SLC39A13*, encoding the intracellular zinc transporter ZIP13. Brittle cornea syndrome is caused by biallelic mutations in either *ZNF469* or *PRDM5*, both (putative) transcriptional regulators. Finally, periodontal EDS is caused by heterozygous mutations in *C1R* or *C1S*, coding for the complement pathway components C1q and C1s, respectively.

Diagnosis Diagnostic workup comprises clinical examination and should be followed by genetic testing in individuals who are suspected to have an EDS type. Genetic testing can include targeted mutation analysis in those with a family history of EDS caused by a known genetic variant or, more frequently, next-generation sequencing using multigene panels. Genetic diagnosis should lead to family testing. Of note, the genetic cause of hypermobile EDS has not been determined, and therefore, diagnosis of this condition is based on the presence of clinical manifestations and the exclusion of other types of EDS or other conditions associated with joint hypermobility. Correlations between genotype and phenotype are challenging and only starting to emerge, and as with other heritable diseases of connective tissue, there is a large degree of variability among members of the same family carrying the same mutation.

TREATMENT
Ehlers-Danlos Syndrome

All patients with EDS should receive multidisciplinary care and, if available, be part of a patient advocacy community. The precise treatment depends on the type of EDS and the clinical manifestations. Physiotherapy is essential for patients with musculoskeletal problems. Helmets and/or skin protections or joint protections, braces, or splints can be used to reduce the risk of injury in patients with skin fragility or joint hypermobility. Low-resistance exercises (such as walking or swimming) can improve joint stability, although exercises that place considerable strain on the joints (such as gymnastics or weightlifting) should be avoided. Monitoring for cardiovascular alterations using noninvasive procedures is recommended in patients at risk of adverse cardiovascular events only. Given the rarity of vascular EDS, referral to a center with EDS expertise is of vital importance. A clear protocol for emergency room evaluation in the case of major complications should be established, and patients should carry documentation of their genetic diagnosis, such as a MedicAlert. The psychosocial impact of a vascular EDS diagnosis often requires psychological care.

■ CHONDRODYSPLASIAS

(See also Chap. 424) Chondrodysplasias (CDs), also referred to as skeletal dysplasias or osteochondrodysplasias, encompass a heterogeneous group of disorders characterized by intrinsic abnormalities of cartilage and bone and are generally characterized by dwarfism and abnormal body proportions (disproportionate short stature). Many affected individuals develop degenerative joint changes, and mild CD in adults may be difficult to differentiate from primary generalized OA.

Classification The Nosology and Classification of Genetic Skeletal Disorders comprises 771 distinct disorders based on clinical, radiographic, and/or molecular phenotypes. Pathogenic variants have currently been found in 551 different genes. The conditions are divided into 41 groups based on gene/protein families (e.g., the type II collagen group), phenotypic presentation (e.g., spondylometaphyseal dysplasia), and pathophysiology (i.e., lysosomal storage disorders). One gene may be responsible for more than one condition (e.g., *COL2A1* mutations may cause a number of CDs including achondrogenesis, hypochondrogenesis, spondyloepiphyseal dysplasia congenita, Kniest and Stickler syndromes), or a condition may be due to mutations in more than one gene (e.g., geleophysic dysplasia can be caused by mutations in *ADAMTSL2, FBN1,* and *LTBP3*).

Incidence The overall incidence of all forms of CD ranges from 1 per 2500 to 1 per 4000 births. Data on the frequency of individual CDs are incomplete. The most common form of inherited disproportionate short stature is achondroplasia, with an estimated incidence of 1 per 26,000 to 1 per 28,000 live births.

Molecular Defects Mutations in the *COL2A1* gene, coding for the α chain of type II collagen of cartilage, are found in a group of patients with both mild and severe CDs. For example, a mutation in *COL2A1* substituting a cysteine residue for an arginine was found in a few unrelated families with spondyloepiphyseal dysplasia (SED) and precocious generalized OA. Mutations in the gene were also found in some lethal CDs characterized by gross deformities of bones and cartilage, such as those found in SED congenita, spondyloepimetaphyseal dysplasia congenita, hypochondrogenesis/achondrogenesis type II, and Kniest syndrome. The highest incidence of *COL2A1* mutations, however, occurs in patients with the distinctive features of the Stickler syndrome, which is characterized by skeletal changes, orofacial abnormalities, and ophthalmologic and auditory abnormalities. Most of the mutations in *COL2A1* are premature stop codons that produce haploinsufficiency. In addition, some of the patients with Stickler syndrome or a closely related syndrome have mutations in two genes specific for type XI collagen (*COL11A1* and *COL11A2*), which is an unusual heterotrimer formed from α chains encoded by *COL2A1, COL11A1,* and *COL11A2.* Mutations in the *COL11A1* gene are also found in patients with Marshall syndrome, which is similar to classic Stickler syndrome, but with more severe hearing loss and dysmorphic features, such as a flat or retracted midface with a flat nasal bridge, short nose, anteverted nostrils, long philtrum, and large-appearing eyes.

CDs are also caused by mutations in the less abundant collagens found in cartilage. For example, patients with Schmid metaphyseal CD have mutations in the gene for type X collagen, a short, network-forming collagen found in the hypertrophic zone of endochondral cartilage. The syndrome is characterized by short stature, coxa vara, flaring metaphyses, and waddling gait. As with other collagen genes, the most common mutations are of two types: nonsense mutations that lead to haploinsufficiency and structural mutations that compromise collagen assembly.

Some patients have mutations in genes for proteins that interact with collagens. Patients with pseudoachondroplasia or autosomal dominant multiple epiphyseal dysplasia have mutations in the gene for the cartilage oligomeric matrix protein (*COMP*), a protein that interacts with both collagens and proteoglycans in cartilage. However, some families with multiple epiphyseal dysplasia have a defect in one of the three genes for type IX collagen (*COL9A1*, *COL9A2*, and *COL9A3*) or in matrilin-3, another extracellular protein found in cartilage.

Some CDs are caused by mutations in genes that affect early development of cartilage and related structures. Achondroplasia is caused by mutations in the gene for a receptor for a fibroblastic growth factor (*FGFR3*). The mutations in the *FGFR3* gene causing achondroplasia are unusual in several respects. More than 99% of individuals with achondroplasia have one of two pathogenic variants (c.1138G> or c.1138G>C) in *FGFR3*, both resulting in the amino acid change p.Gly380Arg. Most patients harbor a sporadic new (de novo) mutation, and therefore, this nucleotide change is one of the most common recurring mutations in the human genome. The mutation causes unregulated signal transduction through the receptor and inappropriate development of cartilage. Mutations that alter other domains of *FGFR3* have been found in patients with the more severe disorders of hypochondroplasia and thanatophoric dysplasia and in a few families with a variant of craniosynostosis. However, most patients with craniosynostosis appear to have mutations in the related *FGFR2* gene. The similarities between the phenotypes produced by mutations in genes for fibroblast growth factor (FGF) receptors and mutations in structural proteins of cartilage are probably explained by the observation that the activity of FGFs is regulated in part by binding of FGFs to proteins sequestered in the extracellular matrix. Therefore, the situation parallels the interactions between transforming growth factors (TGFs) and fibrillin in MFS (see below).

Other mutations involve the proteoglycans of cartilage, aggrecan (*AGC1*) and perlecan (*HSPG2*), and in the proteoglycan posttranslational sulphation pathway (*DTDST*, *PAPSS2*, and *CHST3*).

Diagnosis The diagnosis of CDs is made on the basis of the physical appearance, slit-lamp eye examinations, x-ray findings, histologic changes, and clinical course. Targeted gene and exome sequencing or more global sequencing strategies are used for molecular diagnosis. Given the wide spectrum of CD phenotypes, these genetic tests are becoming critical diagnostic tools. For Stickler syndrome, more precise diagnostic criteria have made it possible to identify type I variants with mutations in the *COL2A1* gene with a high degree of accuracy. It has been suggested that the type II variant with mutations in the *COL11A1* gene can be identified on the basis of a "beaded" vitreous phenotype and that the type III variant with mutations in the *COL11A2* gene can be identified on the basis of the characteristic systemic features without the ocular involvement. Prenatal diagnosis based on analysis of DNA obtained from chorionic villus or amniotic fluid is possible.

TREATMENT

Chondrodysplasias

The treatment of CDs is symptomatic and is directed to secondary features such as degenerative arthritis. Many patients require joint replacement surgery and corrective surgery for cleft palate. The eyes should be monitored carefully for the development of cataracts and the need for laser therapy to prevent retinal detachment. In general, patients should be advised to avoid obesity and contact sports. Counseling for the psychological problems of short stature is critical. A randomized, double-blind, phase 3, placebo-controlled, multicenter trial with vosoritide, a biologic analogue of C-type natriuretic peptide, which is a potent stimulator of endochondral ossification, in children with achondroplasia showed that this is an effective and safe treatment to increase growth in children with achondroplasia.

■ HERITABLE THORACIC AORTIC ANEURYSM DISEASE

Heritable thoracic aortic aneurysm disease (HTAD) encompasses conditions in which aortic disease has a familial occurrence, due to an underlying genetic defect. HTAD is classified as syndromic or nonsyndromic. Syndromic HTAD may be associated with ocular, craniofacial, musculoskeletal, and skin features, with a recognizable, yet sometime subtle, phenotype. They are caused by mutations in genes that code for extracellular matrix proteins. Besides syndromic HTAD, there are several nonsyndromic forms of HTAD; patients with these conditions do not display an outward recognizable phenotype and are classified as having familial thoracic aortic aneurysm (FTAA). More extensive genetic screening in cohorts of patients with thoracic aortic aneurysm is, however, slowly revealing that there is no strict boundary between syndromic and nonsyndromic HTAD entities **(Table 425-4) (Chap. 291)**.

Classification The most common form of syndromic HTAD is MFS, caused by mutations in the gene for fibrillin-1 (*FBN1*). MFS was initially characterized by a triad of features: (1) skeletal changes that include long, thin extremities, frequently associated with loose joints; (2) reduced vision as the result of dislocations of the lenses (ectopia lentis); and (3) aortic aneurysms. An international panel has developed a series of revised Ghent criteria that are useful in classifying patients. Other major syndromic HTADs include the different genetic variants of Loeys-Dietz syndrome (LDS) (*TGFBR1, TGFBR2, TGFB2, TGFB3, SMAD2,* and *SMAD3*). Rare forms of syndromic HTAD include Shprintzen-Goldberg syndrome (*SKI*), Meester-Loeys syndrome (*BGN*), and arterial tortuosity syndrome (ATS) (*SLC2A10*).

Incidence and Inheritance The incidence of MFS is among the highest of any heritable disorder: ~1 in 3000–5000 births in most racial and ethnic groups. The related syndromes are less common. Mutations are generally inherited as autosomal dominant traits, but about one-fourth of patients have sporadic new mutations. The LDSs are less common, but their exact incidence is currently unknown.

Skeletal Effects Patients with MFS typically display a marfanoid habitus with tall stature and long limbs. The ratio of the upper segment (top of the head to the top of the pubic ramus) to the lower segment

TABLE 425-4 Heritable Thoracic Aortic Disease and Associated Genes and Proteins

	GENE	PROTEIN	CONDITION	OMIM	LOCUS
Extracellular matrix proteins	*COL3A1*	α1(III) collagen chain	Vascular EDS	130050	2q32
	FBN1	Fibrillin 1	Marfan syndrome	154700	15q21.1
	MFAP5	Microfibrillar associated protein 5	Familial thoracic aortic aneurysm 9	616166	12p13.31
	LOX	Lysyl oxidase	Familial thoracic aortic aneurysm 10	617168	5q23.1
TGF-β signaling	*TGFBR1*	Transforming growth factor receptor 1	Loeys-Dietz syndrome 1	609192	9q22.33
	TGFBR2	Transforming growth factor receptor 2	Loeys-Dietz syndrome 2	610168	3p24.1
	SMAD3	Mothers against decapentaplegic drosophila homolog 3	Loeys-Dietz syndrome 3	613795	15q22.33
	TGFB2	Transforming growth factor β2	Loeys-Dietz syndrome 4	614816	1q41
	TGFB3	Transforming growth factor β3	Loeys-Dietz syndrome 5	615582	14q23.3
	SMAD2	Mothers against decapentaplegic drosophila homolog 2	Arterial aneurysms and dissections	/	18q21.1
	ACTA2	Smooth muscle actin α2	Familial thoracic aortic aneurysm 6	611788	10q23.31
Smooth muscle contraction	*MYH11*	Smooth muscle myosin heavy chain 11	Familial thoracic aortic aneurysm 4	132900	16p13.11
	MYLK	Myosin light chain kinase	Familial thoracic aortic aneurysm 7	613780	3q21.1
	PRKG1	Protein kinase cGMP-dependent type 1	Familial thoracic aortic aneurysm 8	615436	10q11.2-q21.1

Abbreviations: EDS, Ehlers-Danlos syndromes; OMIM, Online Mendelian Inheritance in Man; TGF, transforming growth factor.

(top of the pubic ramus to the floor) is usually 2 standard deviations below mean for age, race, and sex. The fingers and hands are long and slender and have a spider-like appearance (arachnodactyly). Overlapping features in MFS and LDS include scoliosis or kyphoscoliosis; anterior chest deformities, including pectus excavatum, pectus carinatum, or asymmetry; pes planus; pneumothorax; and dural ectasia. A few patients have severe joint hypermobility similar to EDS. Clubfeet, joint contractures, and cervical spine instability are more frequently observed in LDS. Patients with *SMAD3* mutations are particularly prone to premature OA.

Cardiovascular Features Cardiovascular abnormalities are the major source of morbidity and mortality both in MFS and LDS (Chap. 291). Patients with MFS often have mitral valve prolapse that develops early in life and that progresses to mitral valve regurgitation of increasing severity in about one-quarter of patients. Dilation of the root of the aorta and the sinuses of Valsalva are characteristic and ominous features of MFS that can develop at any age. The rate of dilation is unpredictable, but it can lead to aortic regurgitation, dissection of the aorta, and rupture. Dilation is probably accelerated by physical and emotional stress as well as by pregnancy. Cardiovascular features of LDS also include dilatation of the aortic root at the level of the sinus of Valsalva, which can progress to dissection or rupture when left untreated. LDS is also known for its involvement of aneurysms affecting arterial branches of head, neck, thoracic and abdominal aorta, lung, and lower extremities and for the presence or tortuosity of these vessels. In contrast to MFS, congenital heart malformations are often noted.

Ocular Features Myopia is the most common ocular feature of MFS and often presents in early childhood. Displacement of the lens from the center of the pupil (ectopia lentis) occurs in ~60% of MFS patients. The ocular globe is frequently elongated. Retinal detachment, early cataract formation, and glaucoma can occur. Ectopia lentis does not usually occur in LDS, but other ocular features may be present, such as blue sclerae, strabismus, amblyopia, and myopia.

Other Features MFS patients typically have a high arched palate. Patients with LDS characteristically display hypertelorism (widely spaced eyes) and cleft palate or bifid (split) uvula. They may also have craniosynostosis. Shared mucocutaneous features include striae, typically over the shoulders and buttocks, and inguinal and incisional hernias. Patients with LDS may display more EDS-like skin features, such as thin translucent skin and widened scars.

Molecular Defects Approximately 95% of MFS patients are explained by *FBN1* defects, and so far, over 2000 different *FBN1* mutations have been described. Mutations in the same gene are found in a few patients who do not meet the Ghent criteria. Most *FBN1* gene mutations are unique and are scattered throughout its 65 coding exons. Approximately 10% are recurrent mutations that are largely located in CpG sequences known to be "hot spots." About one-third of the mutations introduce premature termination codons, and about two-thirds are missense mutations that alter calcium-binding domains in the repetitive epidermal growth factor–like domains of the protein. Rarer mutations alter the processing of the protein. As in many genetic diseases, the severity of the phenotype cannot be predicted from the nature of the mutation. In LDS, components of the TGF-β signaling pathway are mutated, including the cytokines (TGFβ2, TGFβ3), the receptors (TGFBR1, TGFBR2), and the downstream effectors (SMAD2, SMAD3).

The discovery that various conditions with pronounced clinical overlap with MFS were caused by mutations in genes coding for direct effectors and/or regulators of TGFβ signaling, including LDS, revealed that *FBN1* mutations not only lead to weakening of the extracellular matrix structure, but also influence the bioavailability of TGFβ. As a result, some of the manifestations of MFS have been shown to arise from alterations in binding sites that modulate TGFβ bioavailability during development of the skeleton and other tissues. In aortic tissues, reduced or altered forms of fibrillin-1 can stimulate the release of sequestered TGFβ and increase its activity. This results in altered transcription of target genes, including connective tissue growth factor and matrix metalloproteinases MMP2 and MMP9. The aortic phenotype is therefore caused by vascular remodeling due to a combination of structural microfibril changes, excess TGFβ, and overexpression of MMP-2 and MMP-9. The role of TGFβ in the pathophysiology of MFS has been further solidified by therapeutic use of angiotensin 2 receptor blockers (ARBs), proven to decrease TGFβ activity, to reduce the progression of aortic dilation. However, clinical trials with ARBs did not provide evidence of a dramatic decrease or prevention of aortic growth with ARBs.

Diagnosis When HTAD is present, genetic testing can confirm the diagnosis and allow identification of at-risk individuals. Referral to a specialty genetics service is critically important, and genetic counseling before testing is recommended. In view of phenotypic overlap between the syndromic HTAD, a multigene panel (usually including genes for syndromic and nonsyndromic HTAD) is recommended. All patients with a suspected diagnosis of MFS should have a slit-lamp examination and an echocardiogram. Also, homocystinuria should be ruled out by amino acid analysis of plasma (Chap. 431). The diagnosis of MFS according to the international Ghent standards places emphasis on two cardinal features, dilation of the ascending aorta with or without dissection and ectopia lentis. Other cardiovascular and ocular manifestations and findings in other organ systems such as the skeleton, dura, skin, and lungs contribute to a systemic score that guides diagnosis when aortic disease is present but ectopia lentis is not.

TREATMENT

Marfan and Loeys-Dietz Syndromes

Patients should be advised that vascular risks are increased by severe physical exertion, smoking, emotional stress, and pregnancy. Low-level moderate aerobic exercise and limits on isometric exercise are recommended. Prophylactic beta blocker and/or angiotensin II receptor blocker therapy are prescribed in normotensive individuals, and blood pressure control is important for those with hypertension. Surgical correction of the aorta, aortic valve, and mitral valve has been successful in many patients, but tissues are frequently friable. The scoliosis tends to be progressive, and surgical stabilization may be required. Dislocated lenses rarely require surgical removal, but patients should be followed closely for retinal detachment.

Acknowledgment

Darwin J. Prockop and John F. Bateman contributed to this chapter in the 20th edition and some material from that chapter has been retained here.

■ FURTHER READING

Besio R et al: Bone biology: Insights from osteogenesis imperfecta and related rare fragility syndrome. FEBS J 286:3033, 2019.

De Backer J et al: Genetic testing for aortopathies: Primer for the nongeneticist. Curr Opin Cardiol 34:585, 2019.

Jovanovic M et al: Osteogenesis imperfecta: Mechanisms and signaling pathways connecting classical and rare OI types. Endocr Rev 12:61, 2022.

Loeys BL et al: The revised Ghent nosology for the Marfan syndrome. J Med Genet 47:476, 2010.

Malfait F et al: The Ehlers-Danlos syndromes. Nat Rev Dis Primers 6:64, 2020.

Marini JC et al: Osteogenesis imperfecta. Nat Rev Dis Primers 3:17052, 2017.

Marzin P, Cormier-Daire V: New perspectives on the treatment of skeletal dysplasia. Ther Adv Endocrinol Metab 11:2042018820904016, 2020.

Unger S et al: Nosology and classification of genetic skeletal disorders: 2023 revision. Am J Med Genet A 191:1164, 2023.

426

Hemochromatosis

Darrell H. G. Crawford, David M. Frazer

■ DEFINITION

Hemochromatosis is a relatively common inherited disorder of iron metabolism more prevalent in European populations. It is now known to be an iron-storage disorder with underlying genetic heterogeneity that, in nearly all cases, causes inappropriately high cellular release of iron from iron exporting cells such as enterocytes and macrophages. The increase in intestinal iron absorption results in the deposition of excess iron in parenchymal cells with eventual cellular injury, tissue fibrosis, and organ failure. Cirrhosis of the liver, diabetes mellitus, arthritis, cardiomyopathy, and hypogonadotropic hypogonadism are the most important clinical consequences.

The term *hemochromatosis* describes the unique genetic clinical-pathologic condition with features of increased serum transferrin saturation, iron overload in the liver but not the spleen, prevalent involvement of periportal hepatocytes with iron sparing of Kupffer cells, and/or signs and symptoms of iron overload. These characteristics should occur in the absence of a primary/predominant red blood cell disorder to distinguish hemochromatosis from other iron overload conditions, which are also often genetic in origin.

The condition is most often caused by a mutation in the homeostatic iron regulator (*HFE*) gene, which is tightly linked to the HLA-A locus on chromosome 6p. Persons who are homozygous for the mutation are at increased risk of iron overload and account for 80–90% of hemochromatosis in persons of northern European descent. In such subjects, the presence of hepatic fibrosis, cirrhosis, arthropathy, or hepatocellular carcinoma constitutes iron overload–related disease. Rarer forms of non-*HFE* hemochromatosis are caused by mutations in other genes involved in iron metabolism **(Table 426-1)**. The disease is now often diagnosed during its early stages when iron overload and organ damage are minimal.

Secondary iron overload occurs as a result of an iron-loading anemia, such as thalassemia or sideroblastic anemia, in which erythropoiesis is ineffective and causes increased intestinal iron absorption. In the acquired iron-loading disorders, massive iron deposits in parenchymal tissues can lead to the same clinical and pathologic features as in hemochromatosis.

■ PREVALENCE

The prevalence of *HFE*-associated hemochromatosis mutations varies among different ethnic groups. It is most common in populations of northern European extraction in whom ~1 in 10 persons are heterozygous carriers and 0.3–0.5% are homozygotes, with even higher percentages in some Celtic populations such as those residing in Ireland and Brittany. The expression of the disease is variable and modified by several factors, especially alcohol consumption, dietary iron intake, blood loss associated with menstruation and pregnancy, and blood donation. Recent population studies indicate that ~40% of homozygous men develop iron overload–related complications and about 6% develop hepatic cirrhosis. For women, iron overload–related complications occur in ~10%. In addition, there are as yet unidentified modifying genes responsible for variable disease expression. Nearly 70% of untreated patients develop the first symptoms between ages 40 and 60. The disease is rarely evident before age 20, although with family screening (see “Screening for Hemochromatosis,” below) and periodic

TABLE 426-1 Classification of Iron Overload States

Hereditary Hemochromatosis	
Hemochromatosis, *HFE*-related (type 1)	
C282Y homozygosity	
C282Y/H63D compound heterozygosity	
Hemochromatosis, non-*HFE*-related	
Juvenile hemochromatosis (type 2A) (hemojuvelin mutations)	
Juvenile hemochromatosis (type 2B) (hepcidin mutation)	
Mutated transferrin receptor 2, *TFR2* (type 3)	
Mutated ferroportin 1 gene, *SLC40A1* (type 4)	
Acquired Iron Overload	
Iron-loading anemias	Chronic liver disease
Thalassemia major	Hepatitis C
Sideroblastic anemia	Alcoholic cirrhosis, especially when advanced
Chronic hemolytic anemias	Nonalcoholic steatohepatitis
Transfusional and parenteral iron overload	Porphyria cutanea tarda
Dietary iron overload	Dysmetabolic iron overload syndrome
	Post-portacaval shunting
Miscellaneous	
Iron overload in sub-Saharan Africa	
Neonatal iron overload	
Aceruloplasminemia	
Congenital atransferrinemia	

health examinations, asymptomatic subjects with iron overload can be identified, including young menstruating women.

Mutations in other genes involved in iron metabolism are responsible for non-*HFE*-associated hemochromatosis. In contrast to *HFE*-associated hemochromatosis, the non-*HFE*-associated forms of hemochromatosis (Table 426-1) are rare, but they affect all populations and may affect young people (juvenile hemochromatosis). These conditions result from mutations in one or more of the genes encoding proteins in the hepcidin pathway **(Fig. 426-1)**, including hepcidin, hemojuvelin, and transferrin receptor 2 (TFR2). The resultant clinical disease is very similar to HFE-related disease because they all lead to hepcidin deficiency via a final common pathway (Fig. 426-1).

A rare autosomal dominant form of hemochromatosis results from two types of mutations in the gene for the iron transporter ferroportin. Loss-of-function mutations decrease the cell surface localization of ferroportin in certain tissues, thereby reducing its ability to export iron ("ferroportin disease") and causing cellular iron retention, particularly in macrophages. A second mutation abolishes the hepcidin-induced ferroportin internalization and degradation resulting in a "gain of function." Here the tissue iron distribution is similar to that in HFE-related disease (e.g., in parenchymal cells).

GENETIC BASIS

The most common mutation in the *HFE* gene is a homozygous G to A transition that leads to a cysteine to tyrosine substitution at position 282 (C282Y) of the HFE protein. It has been identified in 85–90% of patients with hereditary hemochromatosis in populations of northern European descent but is found in only 60% of cases from Mediterranean populations. A second, relatively common *HFE* variant (H63D) results in a substitution of aspartic acid for histidine at residue 63 of the HFE protein. Homozygosity for H63D is not associated with clinically significant iron overload. Some compound heterozygotes (i.e., one copy each of C282Y and H63D) have mild to moderately increased body-iron stores but develop clinical disease only in association with

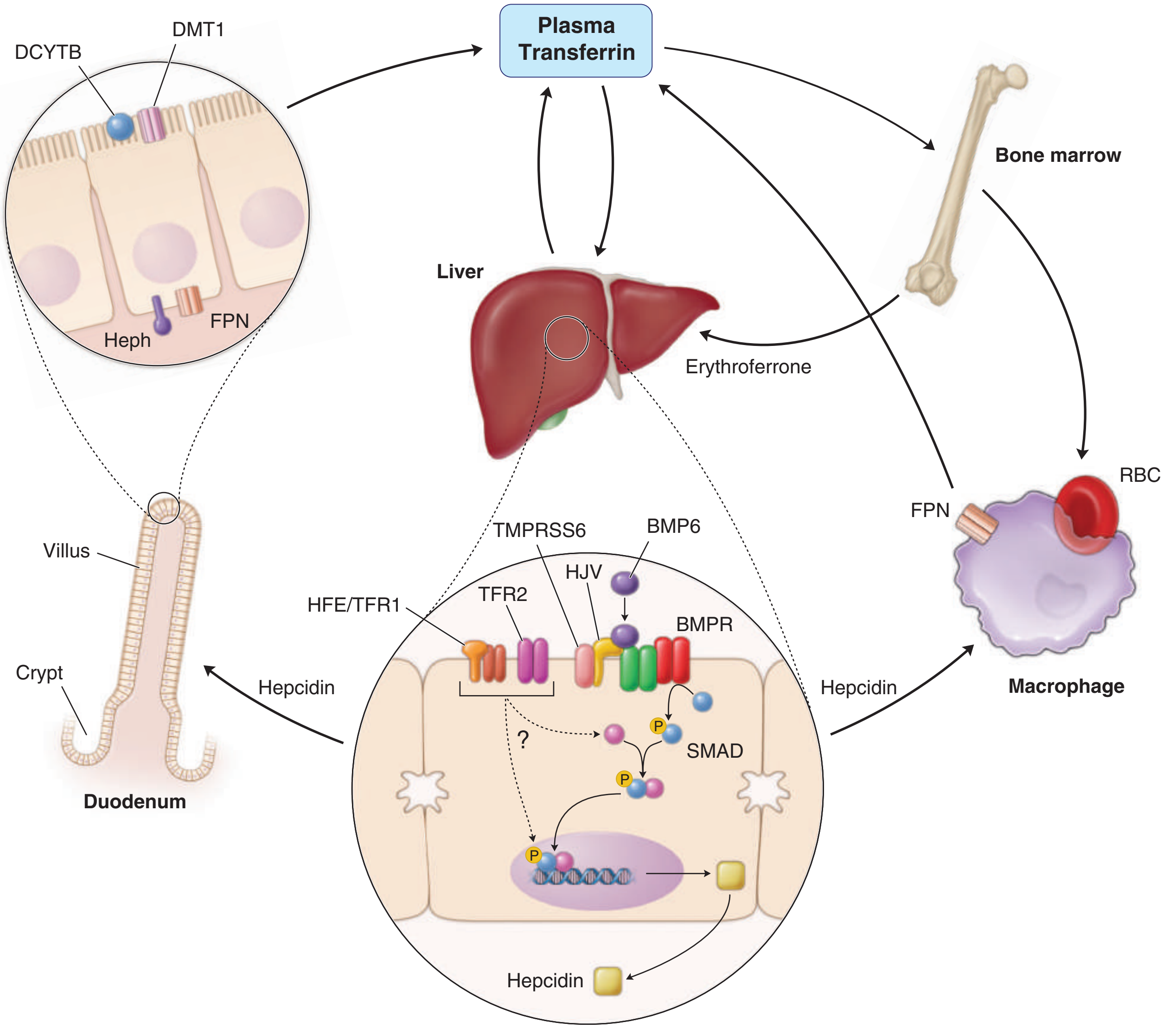

FIGURE 426-1 Pathways of normal iron homeostasis. Dietary inorganic iron traverses the brush border membrane of duodenal enterocytes via divalent metal-ion transporter 1 (DMT1) after reduction of ferric (Fe^{3+}) iron to the ferrous (Fe^{2+}) state by intestinal ferrireductases such as duodenal cytochrome B (DCYTB). Iron then moves from the enterocyte to the circulation via a process requiring the basolateral iron exporter ferroportin (FPN) and the iron oxidase hephaestin (Heph). In the circulation, iron binds to plasma transferrin and is thereby distributed to sites of iron utilization and storage. Much of the diferric transferrin supplies iron to immature erythrocyte cells in the bone marrow for hemoglobin synthesis. At the end of their life, senescent red blood cells (RBCs) are phagocytosed by macrophages, and iron is returned to the circulation after export through ferroportin. The liver-derived peptide hepcidin represses basolateral iron transport in the gut as well as iron released from macrophages and other cells and serves as a central regulator of body-iron traffic. At least two separate signals regulate hepcidin production in response to changes in body-iron requirements. The first involves the detection of circulating diferric transferrin by HFE and TFR2. A second relies on hepatic iron stores activating the hemojuvelin (HJV)-dependent bone morphogenetic protein (BMP)/SMAD pathway. This pathway is modified by erythroferrone released from erythroid precursor cells, which binds to BMP6 and inhibits its function. TMPRSS6 is a protease that regulates hepcidin production, possibly by modulating HJV activity. Heme is metabolized by heme oxygenase within the enterocytes, and the released iron then follows the same pathway. Mutations in the genes encoding HFE, TFR2, HJV, and hepcidin all lead to decreased hepcidin release and increased iron absorption, resulting in hemochromatosis (Table 426-1).

cofactors such as heavy alcohol intake or hepatic steatosis. *HFE*-associated hemochromatosis is inherited as an autosomal recessive trait, and heterozygotes have no, or minimal, increase in iron stores. However, this slight increase in hepatic iron can act as a cofactor that may modify the expression of other diseases such as porphyria cutanea tarda (PCT) or metabolic dysfunction associated steatohepatitis (MASH).

PATHOPHYSIOLOGY AND THE ROLE OF HEPCIDIN

Normally, the body-iron content of 3–4 g is maintained as a result of intestinal mucosal absorption of iron being equal to iron loss. This amount is ~1 mg/d in men and 1.5 mg/d in menstruating women. In hemochromatosis, intestinal iron absorption is greater than body requirements and amounts to ≥4 mg/d. The progressive accumulation of iron increases plasma iron concentration and saturation of transferrin and results in a progressive increase of plasma ferritin **(Fig. 426-2)**. Hepcidin is a key regulatory hormone that allows the bone marrow and other tissues to communicate their iron requirements. It was called hepcidin based upon its antibacterial activity ("HEPatic bacterioCIDal proteIN"). This liver-derived peptide represses basolateral iron export from intestinal enterocytes and iron release from macrophages and other cells by binding to ferroportin. Hepcidin, in turn, responds to signals in the liver mediated by HFE, TFR2, and hemojuvelin (Fig. 426-1). The development of hepcidin agonists represents a promising new therapeutic approach for iron overload disorders caused by low hepcidin levels.

The *HFE* gene encodes a 343-amino-acid protein that is structurally related to MHC class I proteins. The basic defect in *HFE*-associated hemochromatosis is a lack of cell surface expression of HFE (due to the C282Y mutation). The normal (wild-type) HFE protein forms a complex with β_2-microglobulin and transferrin receptor 1 (TFR1), and the C282Y mutation completely abrogates this interaction. As a result, the mutant HFE protein remains trapped intracellularly. Although the precise function of HFE at the cell surface is not known, mutations in this protein reduce hepcidin production leading to increased dietary iron absorption (Fig. 426-1). In advanced disease, the body may contain 20 g or more of iron, which is deposited mainly in parenchymal cells of the liver, pancreas, and heart. Iron deposition in the pituitary causes hypogonadotropic hypogonadism in both men and women. Tissue injury results from disruption of iron-laden lysosomes, lipid peroxidation of subcellular organelles by excess iron, and stimulation of collagen synthesis by activated hepatic stellate cells.

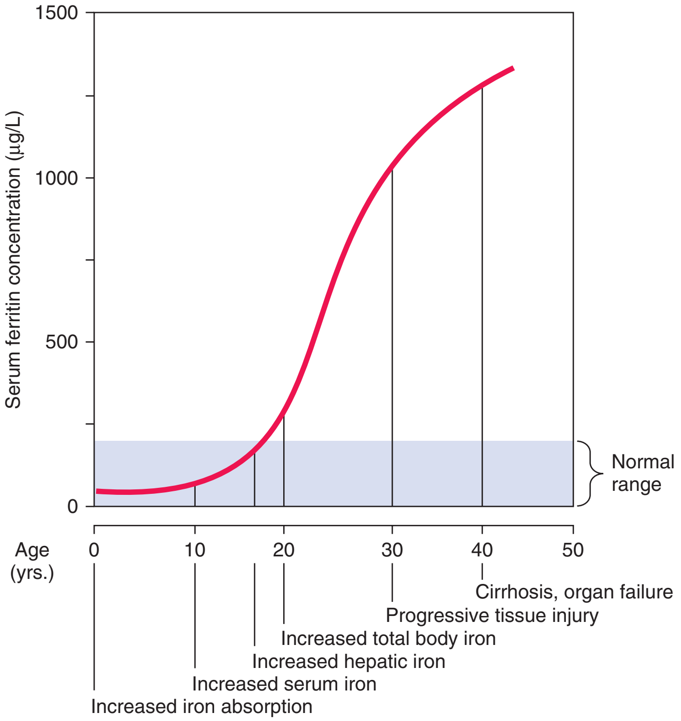

FIGURE 426-2 Sequence of events in genetic hemochromatosis and their correlation with the serum ferritin concentration. Increased iron absorption is present throughout life. Overt, symptomatic disease usually develops between ages 40 and 60, but latent disease can be detected long before this.

Secondary iron overload with iron deposition in parenchymal cells occurs in chronic disorders of erythropoiesis, particularly in those with defects in hemoglobin synthesis and ineffective erythropoiesis such as sideroblastic anemia and thalassemia **(Chap. 103)**. In these disorders, iron absorption is increased. Moreover, these patients require blood transfusions and are frequently treated inappropriately with iron. PCT, a disorder characterized by a defect in porphyrin biosynthesis **(Chap. 428)**, can also be associated with excessive parenchymal iron deposits. The magnitude of the iron load in PCT is usually insufficient to produce tissue damage. However, some patients with PCT also have mutations in the *HFE* gene, and some have associated hepatitis C virus (HCV) infection. Although the relationship between these disorders remains to be clarified, iron overload accentuates the inherited enzyme deficiency in PCT and should be avoided along with other agents (alcohol, estrogens, haloaromatic compounds) that may exacerbate PCT. Another cause of hepatic parenchymal iron overload is hereditary aceruloplasminemia. In this disorder, impairment of iron mobilization due to deficiency of ceruloplasmin (a ferroxidase) causes iron overload in hepatocytes and a range of other cell types.

Excessive iron ingestion over many years rarely results in hemochromatosis. An important exception has been reported in South Africa among groups who brew fermented beverages in vessels made of iron. Hemochromatosis has been described in apparently normal persons who have taken medicinal iron over many years, but such individuals probably had genetic disorders.

The common denominator in all patients with hemochromatosis is *excessive amounts of iron in parenchymal tissues*. Parenteral administration of iron in the form of blood transfusions or iron preparations results predominantly in reticuloendothelial cell iron overload. This appears to lead to less tissue damage than iron loading of parenchymal cells.

In the liver, parenchymal iron is in the form of ferritin and hemosiderin. In the early stages, these deposits are seen in the periportal parenchymal cells, especially within lysosomes in the pericanalicular cytoplasm of the hepatocytes. This stage progresses to perilobular fibrosis and the formation of fibrous septa due to activation of hepatic stellate cells. In the advanced stage, a macronodular or mixed macro- and micronodular cirrhosis develops. The underlying hepatic iron concentration is an important determinant of the risk of the development of hepatic fibrosis and cirrhosis.

Histologically, iron is increased in many organs, particularly in the liver, heart, and pancreas, and, to a lesser extent, in the endocrine glands. The epidermis of the skin is thin, and melanin is increased in the cells of the basal layer and dermis. Deposits of iron are present around the synovial lining cells of the joints.

CLINICAL MANIFESTATIONS

C282Y homozygotes can be characterized by their stage in the progression of disease as follows: (1) a genetic predisposition without biochemical or clinical abnormalities; (2) iron overload without symptoms; (3) iron overload with symptoms (e.g., arthritis and fatigue); and (4) iron overload with organ damage—in particular, cirrhosis. Many subjects with significant iron overload are asymptomatic. For example, in a study of 672 asymptomatic C282Y homozygous subjects (identified by either family screening or routine health examinations), there was hepatic iron overload (grades 2–4) in 56% and 34.5% of male and female subjects, respectively, hepatic fibrosis (stages 2–4) in 18.4% and 5.4%, respectively, and cirrhosis in 5.6% and 1.9%, respectively.

Initial symptoms of hemochromatosis are often nonspecific and include lethargy, arthralgia, skin pigmentation, loss of libido, and features of diabetes mellitus. Hepatomegaly, increased pigmentation, spider angiomas, splenomegaly, arthropathy, ascites, cardiac arrhythmias, congestive heart failure, loss of body hair, testicular atrophy, and jaundice are prominent in advanced disease.

Hepatocellular carcinoma develops in ~30% of patients with cirrhosis, and it is the most common cause of death in treated patients with cirrhosis—hence the importance of early diagnosis and therapy. The incidence increases with age, it is more common in men, and it occurs almost exclusively in cirrhotic patients.

Excessive skin pigmentation is present in patients with advanced disease. The characteristic metallic or slate-gray hue is sometimes referred to as *bronzing* and results from increased melanin and iron in the dermis. Pigmentation usually is diffuse and generalized.

Diabetes mellitus occurs in ~65% of patients with advanced disease and is more likely to develop in those with a family history of diabetes, suggesting that direct damage to the pancreatic islets by iron deposition occurs in combination with other risk factors. The management is similar to that of other forms of diabetes.

Arthropathy develops in 25–50% of symptomatic patients. It usually occurs after age 50 but may occur as a first manifestation or long after therapy. The joints of the hands, especially the second and third metacarpophalangeal joints, are usually the first joints involved, a feature that helps to distinguish the chondrocalcinosis associated with hemochromatosis from the idiopathic form **(Chap. 384)**. A progressive polyarthritis involving the wrists, hips, ankles, and knees may also ensue. Acute attacks of synovitis may be associated with deposition of calcium pyrophosphate (chondrocalcinosis or pseudogout), mainly in the knees. Radiologic manifestations include cystic changes of the subchondral bones, loss of articular cartilage with narrowing of the joint space, diffuse demineralization, hypertrophic bone proliferation, and calcification of the synovium. The arthropathy tends to progress despite removal of iron by phlebotomy. Although the relation of these abnormalities to iron metabolism is not known, the fact that similar changes occur in other forms of iron overload suggests that iron is directly involved.

Cardiac involvement is the presenting manifestation in ~15% of symptomatic patients. The most common manifestation is congestive heart failure, which occurs in ~10% of young adults with the disease, especially those with juvenile hemochromatosis. Symptoms of congestive heart failure may develop suddenly, with rapid progression to death if untreated. The heart is diffusely enlarged. This may be misdiagnosed as idiopathic cardiomyopathy if other overt manifestations are absent. Cardiac arrhythmias include premature supraventricular beats, paroxysmal tachyarrhythmias, atrial flutter, atrial fibrillation, and varying degrees of atrioventricular block.

Hypogonadism occurs in both sexes and may antedate other clinical features. Manifestations include loss of libido, impotence, amenorrhea, testicular atrophy, gynecomastia, and sparse body hair. These changes are primarily the result of decreased production of gonadotropins due to impairment of hypothalamic-pituitary function by iron deposition.

DIAGNOSIS

The association of (1) hepatomegaly, (2) skin pigmentation, (3) diabetes mellitus, (4) heart disease, (5) arthritis, and (6) hypogonadism should suggest the diagnosis. However, as stated above, significant iron overload may exist with none or only some of these manifestations. Therefore, a high index of suspicion is needed to make the diagnosis early. Treatment before permanent organ damage occurs can reverse the iron toxicity and restore life expectancy to normal.

The history should be particularly detailed in regard to disease in other family members and should include information on alcohol ingestion; iron intake; and ingestion of large doses of ascorbic acid, which promotes iron absorption **(Chap. 344)**. Appropriate tests should be performed to exclude iron deposition due to hematologic disease. The presence of liver, pancreatic, cardiac, and joint disease should be confirmed by physical examination, radiography, and standard function tests of these organs.

The degree of increase in total body iron stores can be assessed by (1) measurement of serum iron and the percent saturation of transferrin (or the unsaturated iron-binding capacity), (2) measurement of serum ferritin concentration, (3) liver biopsy with measurement of the iron concentration **(Table 426-2)**, and (4) magnetic resonance imaging (MRI) of the liver to quantify hepatic iron stores. In addition, a retrospective assessment of body-iron storage is also provided by performing weekly phlebotomy and calculating the amount of iron removed before iron stores are exhausted (1 mL blood = ~0.5 mg iron).

Each of these methods for assessing iron stores has advantages and limitations. The serum iron level and percent saturation of transferrin are elevated early in the course, but their specificity is reduced by significant false-positive and false-negative rates. For example, serum iron concentration may be increased in patients with alcoholic liver disease without iron overload (Table 426-2). In otherwise healthy persons, a fasting serum transferrin saturation >45% is abnormal and suggests homozygosity for hemochromatosis.

The serum ferritin concentration is used as an index of body-iron stores, whether decreased or increased. In fact, an increase of 1 μg/L in serum ferritin level reflects an increase of ~8–10 mg in body stores. In most untreated patients with hemochromatosis, the serum ferritin level is significantly increased (Fig. 426-2 and Table 426-2), and a serum ferritin level >1000 μg/L is a strong predictor of cirrhosis among individuals homozygous for the C282Y mutation. However, in patients with hepatic necroinflammatory conditions such as alcoholic liver disease and metabolic dysfunction–associated steatotic liver disease, acute hepatocellular necrosis, or systemic inflammatory conditions, serum ferritin levels may be elevated out of proportion to body iron stores due to increased release from tissues. Therefore, a repeat determination of serum ferritin should be carried out after acute hepatocellular damage has subsided (e.g., in alcoholic liver disease). Ordinarily, the combined measurements of the percent transferrin saturation and serum ferritin level provide a simple and reliable screening test for hemochromatosis, including the precirrhotic phase of the disease. If either of these tests is abnormal, genetic testing for hemochromatosis should be performed **(Fig. 426-3)**.

The role of liver biopsy in the diagnosis and management of hemochromatosis has been reassessed as a result of the widespread availability of genetic testing for the C282Y mutation. The absence of severe fibrosis can be accurately predicted in most patients using clinical and biochemical variables. Thus, there is virtually no risk of severe fibrosis in a C282Y homozygous subject with (1) serum ferritin level

TABLE 426-2 Representative Iron Values in Normal Subjects, Patients with Hemochromatosis, and Patients with Alcoholic Liver Disease

DETERMINATION	NORMAL	SYMPTOMATIC HEMOCHROMATOSIS	HOMOZYGOTES WITH EARLY, ASYMPTOMATIC HEMOCHROMATOSIS	HETEROZYGOTES	ALCOHOLIC LIVER DISEASE
Plasma iron, μmol/L (μg/dL)	9–27 (50–150)	32–54 (180–300)	Usually elevated	Elevated or normal	Often elevated
Total iron-binding capacity, μmol/L (μg/dL)	45–66 (250–370)	36–54 (200–300)	36–54 (200–300)	Normal	45–66 (250–370)
Transferrin saturation, %	22–45	50–100	50–100	Normal or elevated	27–60
Serum ferritin, μg/L		1000–6000	200–500	Usually <500	10–500
Men	20–250				
Women	15–150				
Liver iron, μg/g dry wt	300–1400	6000–18,000	2000–4000	300–3000	300–2000
Hepatic iron index	<1.0	>2	1.5–2	<2	<2

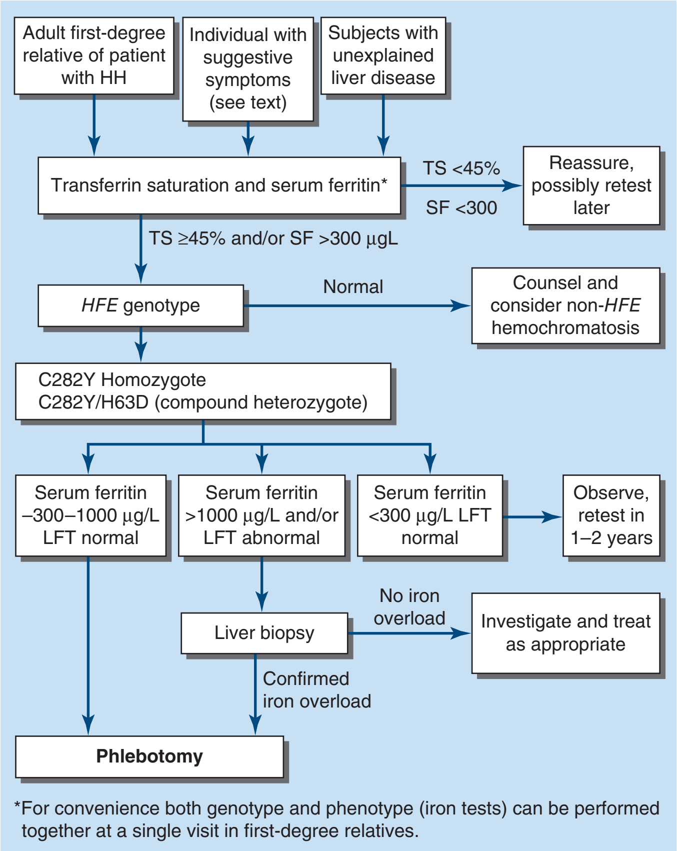
Adult first-degree relative of patient with HH
Individual with suggestive symptoms (see text)
Subjects with unexplained liver disease
Transferrin saturation and serum ferritin*
TS <45%
SF <300
Reassure, possibly retest later
TS ≥45% and/or SF >300 µgL
HFE genotype
Normal
Counsel and consider non-*HFE* hemochromatosis
C282Y Homozygote
C282Y/H63D (compound heterozygote)
Serum ferritin –300–1000 µg/L LFT normal
Serum ferritin >1000 µg/L and/or LFT abnormal
Serum ferritin <300 µg/L LFT normal
Observe, retest in 1–2 years
No iron overload
Liver biopsy
Investigate and treat as appropriate
Confirmed iron overload
Phlebotomy

*For convenience both genotype and phenotype (iron tests) can be performed together at a single visit in first-degree relatives.

FIGURE 426-3 Algorithm for screening for *HFE*-associated hemochromatosis. HH, hereditary hemochromatosis, homozygous subject (C282Y +/+); LFT, liver function test; SF, serum ferritin concentration; TS, transferrin saturation.

<1000 µg/L, (2) normal serum alanine aminotransferase values, (3) no hepatomegaly, and (4) no excess alcohol intake. However, it should be emphasized that liver biopsy is the most reliable method for establishing or excluding the presence of hepatic cirrhosis, which is the critical factor determining prognosis and the risk of developing hepatocellular carcinoma. Biopsy also permits histochemical estimation of tissue iron and measurement of hepatic iron concentration. Serum tests for liver fibrosis and transient elastography can guide decisions about the need for liver biopsy in affected patients. Increased density of the liver due to iron deposition can be demonstrated by computed tomography (CT) or MRI, and with improved technology, MRI can accurately determine hepatic iron concentration.

SCREENING FOR HEMOCHROMATOSIS

When the diagnosis of hemochromatosis is established, it is important to counsel and screen other family members **(Chap. 480)**. Asymptomatic and symptomatic family members with the disease usually have an increased saturation of transferrin and an increased serum ferritin concentration. These changes occur even before iron stores are greatly increased (Fig. 426-2). All adult first-degree relatives of patients with hemochromatosis should be tested for the C282Y and H63D mutations and counseled appropriately (Fig. 426-3). In affected individuals, it is important to confirm or exclude the presence of cirrhosis and begin therapy as early as possible. For children of an identified proband, testing for *HFE* mutations in the other parent is helpful because if normal, the child is an obligate heterozygote and at no risk. Otherwise, for practical purposes, children need not be checked before they are 18 years old.

The role of population screening for hemochromatosis is controversial. Recent studies indicate that it is highly effective for primary care physicians to screen subjects using transferrin saturation and serum ferritin levels. Such screening also detects iron deficiency. Genetic screening of the normal population is feasible but remains controversial in terms of cost-effectiveness.

TREATMENT

Hemochromatosis

The therapy of hemochromatosis involves removal of the excess body iron and supportive treatment of damaged organs. Iron removal is best accomplished by weekly or, with gross iron loading, twice-weekly phlebotomy of 500 mL. Although there is an initial modest decline in the volume of packed red blood cells to about 35 mL/dL, the level stabilizes after several weeks. The plasma transferrin saturation remains increased until the available iron stores are depleted. In contrast, the plasma ferritin concentration falls progressively, reflecting the gradual decrease in body-iron stores. One 500-mL unit of blood contains 200–250 mg of iron, and ≥25 g of iron may have to be removed. Therefore, in patients with advanced disease, weekly phlebotomy may be required for 1–2 years, and it should be continued until the serum ferritin level is ≤100 µg/L. Thereafter, phlebotomies are performed at appropriate intervals to maintain ferritin levels at ≤100 µg/L. The transferrin saturation fluctuates and may still be elevated but should not dictate further therapy unless it is persistently at 100% when free unbound iron may circulate. Usually, one phlebotomy every 3 months will suffice. It is important, however, not to overtreat and render the patient iron deficient.

Chelating agents such as deferoxamine, when given parenterally, remove 10–20 mg of iron per day, which is much less than that mobilized by once-weekly phlebotomy. Phlebotomy is also less expensive, more convenient, and safer for most patients. However, chelating agents may be indicated when anemia or hypoproteinemia is severe enough to preclude phlebotomy.

Effective oral iron chelating agents, deferasirox (Exjade) and deferiprone, are now available. These agents are effective in thalassemia and secondary iron overload but are expensive and carry the risk of significant side effects.

Alcohol consumption should be severely curtailed or eliminated because it increases the risk of cirrhosis in hereditary hemochromatosis nearly tenfold. Dietary adjustments are unnecessary, although vitamin C and iron supplements should be avoided. The management of hepatic failure, cardiac failure, and diabetes mellitus is similar to conventional therapy for these conditions. Loss of libido and change in secondary sex characteristics are managed with testosterone replacement or gonadotropin therapy **(Chap. 403)**.

End-stage liver disease may be an indication for liver transplantation, although results are improved if the excess iron can be removed beforehand. The available evidence indicates that the fundamental metabolic abnormality in hemochromatosis is reversed by successful liver transplantation.

PROGNOSIS

The principal causes of death are cardiac failure, hepatocellular failure, or portal hypertension and hepatocellular carcinoma.

Life expectancy is improved by removal of excessive iron stores and maintenance of these stores at near-normal levels. The 5-year survival rate with therapy increases from 33 to 89%. With repeated phlebotomy, the liver decreases in size, liver function improves, pigmentation of skin decreases, and cardiac failure may be reversed. Diabetes improves in ~40% of patients, but removal of excess iron has little effect on hypogonadism or arthropathy. Hepatic fibrosis may decrease, and cirrhosis may regress with adequate phlebotomy therapy. Hepatocellular carcinoma occurs as a late sequela in patients who are cirrhotic at presentation. The apparent increase in its incidence in treated patients is probably related to their increased life span. Hepatocellular carcinoma rarely develops if the disease is treated in the precirrhotic stage. Indeed, the life expectancy of homozygotes treated before the development of cirrhosis is normal.

The importance of family screening and early diagnosis and treatment cannot be overemphasized. Asymptomatic individuals detected by family studies should have phlebotomy therapy if iron stores are moderately to severely increased. Assessment of iron stores at appropriate intervals is also important. With this management approach, most manifestations of the disease can be prevented.

ROLE OF *HFE* MUTATIONS IN OTHER LIVER DISEASES

There is considerable interest in the role of *HFE* mutations and hepatic iron in several other liver diseases. Several studies have shown an increased prevalence of *HFE* mutations in PCT patients. Iron accentuates the inherited enzyme deficiency in PCT and clinical manifestations of PCT. The situation in metabolic dysfunction–associated steatohepatitis is less clear, but some studies have shown an increased prevalence of *HFE* mutations. Available evidence does not support a role for phlebotomy therapy unless there is a proven increase in hepatic iron stores.

HFE mutations are not increased in frequency in alcoholic liver disease. However, alcohol does reduce hepcidin expression, which accounts for increased iron absorption and hepatic iron sometimes seen in alcoholic liver disease. Hemochromatosis in a heavy drinker can be distinguished from alcoholic liver disease by the presence of the C282Y mutation.

End-stage liver disease may also be associated with iron overload of the degree seen in hemochromatosis. The mechanism is uncertain, although studies have shown reduced hepcidin and intestinal iron transporter expression. Hemolysis also plays a role. *HFE* mutations are uncommon.

GLOBAL CONSIDERATIONS

The *HFE* mutation is of northern European origin (Celtic or Nordic) with a heterozygous carrier rate of ~1 in 10 (1 in 8 in Ireland). Thus, *HFE*-associated hemochromatosis is quite rare in non-European populations, e.g., Asia. However, non-*HFE*-associated hemochromatosis resulting from mutations in other genes involved in iron metabolism (Fig. 426-1) is ubiquitous and should be considered when one encounters iron overload.

African iron overload occurs primarily in sub-Saharan Africa and was previously thought to be due to the consumption of an iron-rich fermented maize beverage. However, recent evidence suggests that it is primarily the result of a non-*HFE*-related genetic trait that is exacerbated by dietary iron loading. A similar form of iron overload has been described in African Americans.

Acknowledgment

The authors extend their gratitude to the late Professor Lawrie W. Powell and recognize his outstanding contributions to previous versions of this chapter.

FURTHER READING

Adams PC et al: Haemochromatosis. Lancet 401:1811, 2023.

Anderson GJ, Bardou-Jacquet E: Revisiting hemochromatosis: Genetic vs. phenotypic manifestations. Ann Transl Med 9:731, 2021.

Girelli D et al: Hemochromatosis classification: Update and recommendations by the BIOIRON Society. Blood 139:3018, 2022.

Olynyk JK, Ramm GA: Hemochromatosis. N Engl J Med 387:2159, 2022.

Powell LW et al: Haemochromatosis. Lancet 388:706, 2016.

427

Wilson's Disease

Stephen G. Kaler

Wilson's disease is an autosomal recessive inherited disorder of copper transport that primarily impacts the liver and brain. This reflects the critical need for homeostatic mechanisms to properly utilize this trace metal, both systemically and in the central nervous system. Since the initial detailed clinical description in 1912, Wilson's disease has emerged as arguably one of the best-characterized and most effectively managed human inborn errors of metabolism. The condition results from variants in *ATP7B*, a highly evolutionarily conserved P-type ion-motive ATPase that normally mediates copper removal from the liver via biliary excretion and prevents brain copper accumulation. Prompt diagnosis in the presymptomatic or early symptomatic phase of the illness and lifelong treatment are needed to prevent premature mortality in affected individuals.

HISTORY OF WILSON'S DISEASE

Wilson's disease (hepatolenticular degeneration) was first described in 1912 by neurologist S.A.K. Wilson, who recognized the heritable aspect of the condition. In 1948, the pathologist J.N. Cumings proposed an etiologic connection with copper overload. Several years later, a metal chelator developed to counteract an arsenic-based chemical warfare agent (lewisite) was used to successfully treat advanced Wilson's disease. In 1956, copper chelation by D-penicillamine was introduced and found preferable to anti-lewisite with respect to route of administration and side effect profile. In the early 1970s, an alternative copper chelator, triethylene tetramine, became the second U.S. Food and Drug Administration (FDA)–approved treatment for Wilson's disease. Also in the early 1970s, the first liver transplants were performed for Wilson's disease, with resultant correction of both hepatic failure and crippling neurologic impairments in patients unresponsive to medical therapies. The treatment potential of zinc salts to reduce gastrointestinal copper absorption in Wilson's disease was recognized in the early 1960s, eventually leading to FDA approval for this indication. Tetrathiomolybdate, which forms a tripartite complex with copper and albumin, and a bacterial peptide, methanobactin, which traverses mitochondrial membranes, are more recently proposed copper chelators with potential for treatment of Wilson's disease.

In 1993, the gene for Wilson's disease was identified and found to encode a copper-transporting ATPase, ATP7B, expressed primarily in liver and kidney. In addition to providing a molecular basis for diagnosis and genotype-phenotype correlations, the finding presents current opportunities for viral gene therapy that could impact future management of this illness.

PHENOTYPES

CLINICAL

Presenting clinical features of Wilson's disease include nonspecific liver disease, neurologic abnormalities, psychiatric illness, hemolytic anemia, renal tubular Fanconi syndrome, and various skeletal abnormalities. Age influences the specific presentation in Wilson's disease. Nearly all individuals who present with liver disease are <30 years of age, whereas those presenting with neurologic or psychiatric signs may range in age from the first to the fifth decade. This reflects the sequence of events in the pathogenesis of the illness. However, regardless of clinical presentation, some degree of liver disease is invariably present.

Hepatic Presentation With hepatic presentations, signs and symptoms include jaundice, hepatomegaly, edema, or ascites. Viral hepatitis and cirrhosis are often initial diagnostic considerations in individuals who, in fact, have Wilson's disease.

Neurologic Presentation In patients with neurologic presentations, abnormalities include distinctive speech difficulties (dysarthria), dystonia, rigidity, tremor (e.g., wing-beating) or choreiform movements, abnormal gait, and uncoordinated handwriting. Wilson's disease can properly be classified as a movement disorder. The neurologic signs and symptoms reflect the predilection for basal ganglia (e.g., caudate, putamen) involvement in the brains of affected persons. Wilson's disease may be mistakenly diagnosed as Parkinson disease or other movement disorders.

Psychiatric Presentation In psychiatric presentations, changes in personality (irritability, anger, poor self-control) or school performance, depression, and anxiety are common symptoms. Typically, patients presenting in this fashion are in their late teens or early

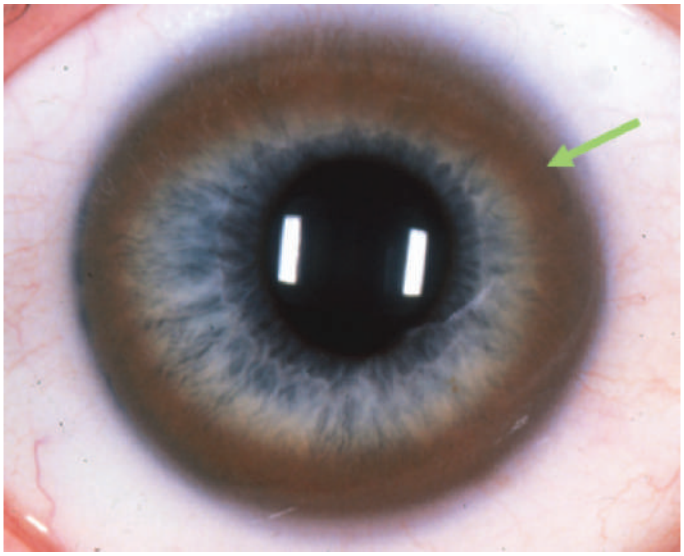

FIGURE 427-1 Kayser-Fleischer ring in Wilson's disease, representing copper deposition in Descemet membrane of the cornea. *(Image courtesy of Tjaard U. Hoogenraad MD, PhD, Department of Neurology, University Medical Centre Utrecht, Utrecht, The Netherlands.)*

twenties, a period during which substance abuse is also a diagnostic consideration. Wilson's disease should be formally excluded in all teenagers and young adults with new-onset psychiatric signs.

Ocular Manifestations The eye is a primary site of copper deposition in Wilson's disease, producing a pathognomonic sign, the Kayser-Fleischer ring **(Fig. 427-1)**, a golden to greenish-brown band in the peripheral cornea. This important diagnostic sign first appears as a superior crescent, then develops inferiorly, and ultimately becomes circumferential. Slit-lamp or optical coherent tomography examinations are required to detect rings in their early stage of formation. Copper can also accumulate in the lens and produce "sunflower" cataracts. Approximately 95% of Wilson's disease patients with neurologic signs manifest the Kayser-Fleischer ring compared to two-thirds of those with hepatic presentations. Copper chelation therapy causes fading and eventual disappearance of corneal copper.

Other Clinical Manifestations Secondary endocrine effects of Wilson-associated liver disease may include delayed puberty or amenorrhea. Renal tubular dysfunction in Wilson's disease leads to abnormal losses of amino acids, electrolytes, calcium, phosphorus, and glucose. Presumably, this effect is related to renal copper toxicity; high copper levels have been noted previously in the kidneys of patients with Wilson's disease. Treatment with copper chelation often improves the renal disturbances. There also can be skeletal effects of Wilson's disease, including osteoporosis and rickets, which may be attributable to renal losses of calcium and phosphorus. Osteoarthritis, primarily affecting the knees and wrists, may involve excess copper deposition in the bone and cartilage.

Hemolytic anemia due to the direct toxic effects of copper on red blood cell membranes is usually associated with release of massive quantities of hepatic copper into the circulation, a phenomenon that can be sudden and catastrophic.

■ BIOCHEMICAL

Laboratory findings that support the diagnosis of Wilson's disease include low levels of serum copper and serum ceruloplasmin, elevated hepatic transaminase levels, aminoaciduria, and hemolytic anemia. Incorporation of radiolabeled ^{64}Cu into serum ceruloplasmin, measured as the appearance of copper in the serum after an oral load, is a highly specific diagnostic test; patients with Wilson's disease incorporate very little ^{64}Cu into ceruloplasmin.

Increased urinary excretion of copper (>100 μg/24 h) is an easily performed and important diagnostic test for Wilson's disease. Acid-washed (copper-free) collection containers should be used. The penicillamine challenge is a variation using serial urine copper measurements in which 500 mg of penicillamine are administered orally after collecting a baseline 24-h urine. The penicillamine dose is repeated after 12 h, the midpoint of the second 24-h urine collection. A several-fold increase in copper excretion in the second collection suggests the diagnosis.

Although invasive, percutaneous needle liver biopsy for measurement of hepatic copper remains a gold standard technique for biochemical diagnosis of Wilson's disease. Hepatic copper values >200 μg per gram of dry weight (normal 20–50 μg) are characteristic of Wilson's disease. Inductively coupled plasma mass spectrometry and atomic absorption spectrometry are preferred quantitative methods; histochemical staining for copper in liver biopsy specimens is considered less reliable.

■ MOLECULAR

Wilson's disease is caused by loss-of-function variants in *ATP7B*. Despite similar genomic structures, large deletions are much less common in *ATP7B* than in *ATP7A*, the closely related X-linked gene responsible for Menkes disease. Several *ATP7B* missense variants are common (H1069Q, M645R, and R778L), with various allelic frequencies reflecting geographic, racial, and/or ethnic differences. Major *ATP7B* databases list >650 pathogenic or likely pathogenic variants. Population-based and genomic-based estimates of prevalence range from 1 in 7000 to 1 in 30,000, with genome-based ascertainments supporting the higher prevalence. This disparity may reflect incomplete penetrance, although there is little doubt that some affected individuals unfortunately escape medical attention. Advances in the application of whole genome sequencing (and/or measurement of ATP7B peptides) from newborn dried blood spots may transform presymptomatic diagnostic screening for Wilson disease in the future.

DIAGNOSIS

Currently, a formal diagnosis of Wilson's disease relies on a combination of clinical, biochemical, and molecular features **(Table 427-1)**. A scoring system (Leipzig) that weights and collates various signs and symptoms was produced by an international expert group in 2001 and remains a valuable guide to diagnosis that is endorsed by the European Association for the Study of the Liver (EASL).

TABLE 427-1 Main Diagnostic Features of Wilson's Disease

CLINICAL SIGNS AND SYMPTOMS	BIOCHEMICAL FINDINGS	MOLECULAR FINDINGS
Hepatic: Jaundice Anorexia Vomiting Ascites and/or edema Splenomegaly Neurologic: Dysarthria Facial grimace (risus sardonicus) Drooling Dysphagia Dysgraphia Dystonia Tremor ("wing-beating") Ataxia Seizures (rare) Ocular: Kayser-Fleischer ring Sunflower cataract (rare) Psychiatric: Decline in school performance Personality change Mood disorder Schizophrenia	Low serum copper Low serum ceruloplasmin Increased urinary copper excretion Elevated liver enzymes Hypoalbuminemia Increased liver copper Fatty liver Cirrhotic liver Hemolytic anemia Renal Fanconi syndrome	Variants in *ATP7B* on both chromosomes Variants or polymorphisms in other genes (e.g., *CAT, SOD2, MTHFR*) may influence clinical expression of Wilson's disease in some individuals

TREATMENT
Wilson's Disease

COPPER CHELATION

The era of successful treatment of Wilson's disease began with the use of British anti-lewisite (BAL) by a defined regimen of intramuscular injections. An orally administered alternative was D-penicillamine (Cuprimine), a free thiol that binds copper. This chelating drug does not formally correct the basic defect of impaired copper excretion in the bile. However, it greatly enhances urinary excretion of copper and thereby corrects and prevents copper overload and its effects. Pyridoxine (vitamin B$_6$) is usually prescribed concomitantly to counter the tendency for deficiency of this vitamin to develop during chronic penicillamine administration.

Certain individuals are intolerant of penicillamine, however, encountering significant side effects that include nephrotoxicity, hematologic abnormalities, and a distinctive rash, *elastosis perforans serpiginosa* (usually involving the neck and axillae). Furthermore, in some Wilson's disease patients with neurologic presentations, penicillamine treatment induces paradoxical worsening of neurologic status. Triethylenetetramine dihydrochloride (trientine hydrochloride [Syprine]) is a suitable alternative chelating agent with a somewhat better side effect profile.

Tetrathiomolybdate (TM) is another molecule in the Wilson's disease therapeutic armamentarium. TM forms stable tripartite complexes among albumin, copper, and itself. This drug functions both to decrease copper absorption and to reduce circulating free copper. It is fast-acting and can restore normal copper balance within several weeks compared to the several months required with other copper chelators or with zinc.

Copper Chelation Treatment During Pregnancy The rate of spontaneous miscarriage is increased in pregnant women with untreated Wilson's disease. From a benefit/risk perspective, it is important to maintain copper chelation treatment during pregnancy to prevent hepatic or neurologic relapse, as well as to lower risk of pregnancy loss. Some academic centers favor reduction of copper chelator dose during pregnancy, although if zinc monotherapy (see below) is in place at the time of conception, evidence suggests it is safe to maintain the usual daily dose. Since all anticopper medications enter breast milk, breastfeeding is not recommended for mothers with Wilson's disease.

REDUCTION OF COPPER ABSORPTION

Zinc acetate (Galzin) has proven highly effective for treatment of Wilson's disease. The mechanism involves induction of metallothionein synthesis in intestinal epithelial cells; increased metallothionein synthesis results in greater binding of dietary copper and thus decreased absorption. Zinc therapy has particular value in (1) young, presymptomatic patients; (2) patients who are pregnant given the possible fetal teratogenic effects of other compounds; and (3) as maintenance therapy for patients after their initial "de-coppering" is accomplished. Zinc acetate has minimal side effects. The sole drawback is the relatively long time (4–6 months) needed for restoration of proper copper balance when used as monotherapy in the initial stages of treatment.

LIVER TRANSPLANTATION

Liver transplantation is a consideration for Wilson's disease in advanced stages and/or when the condition is unresponsive to medical therapy. This is generally necessary only in cases where delayed diagnosis or poor medication compliance results in irreversible hepatic damage. A recently proposed alternative for this circumstance is methanobactin, a bacterial peptide that binds copper avidly and dramatically improves mitochondrial copper overload and restores normal mitochondrial morphology in a preclinical (rat) model of advanced Wilson's disease.

GENE THERAPY

In a preclinical (mouse) model of Wilson's disease, adeno-associated virus-mediated *ATP7B* introduction into hepatocytes was shown to be effective. Transduction of only 20% of hepatocytes was sufficient to normalize copper homeostasis in the animal model. Those results potentially pave the way for viral gene therapy in Wilson's disease patients, and clinical trials to evaluate this approach are underway.

FUTURE OUTLOOK

Wilson's disease is arguably one of the best-characterized human inborn errors of metabolism from combined clinical, biochemical, and molecular perspectives, related to the detailed attention devoted to this condition. As noted, novel copper chelators are still being evaluated, and generic formulations of established drugs are contributing to increased affordability for patients and their families. Viral gene therapy to provide working versions of *ATP7B* to the liver, kidney, and brain or that delivers gene-editing molecules to correct specific mutant alleles is now an emerging prospect. In addition, advances in newborn screening technology may eventually enable wider population-based screening for Wilson's disease, which would help address lingering questions about clinical penetrance. Such future progress in newborn screening would also avert the tragedy that missed diagnoses of this eminently treatable disorder of copper transport represent.

■ FURTHER READING

Bandmann O et al: Wilson's disease and other neurological copper disorders. Lancet Neurol 14:103, 2015.

Dev S et al: Wilson disease: Update on pathophysiology and treatment. Front Cell Dev Biol 10:871877, 2022.

Gao J et al: The global prevalence of Wilson disease from next-generation sequencing data. Genet Med 21:1155, 2019.

Kumar M et al: WilsonGen: A comprehensive clinically annotated genomic variant resource for Wilson's disease. Sci Rep 10:9037, 2020.

Lichtmannegger J et al: Methanobactin reverses acute liver failure in a rat model of Wilson disease. J Clin Invest 126:2721, 2016.

Murillo O et al: Liver expression of a MiniATP7B gene results in long-term restoration of copper homeostasis in a Wilson disease model in mice. Hepatology 70:108, 2019.

Sandahl TD et al: The prevalence of Wilson's disease: An update. Hepatology 71:722, 2020.

Wallace DF, Dooley JS: ATP7B variant penetrance explains differences between genetic and clinical prevalence estimates for Wilson disease. Hum Genet 139:1065, 2020.

428

The Porphyrias

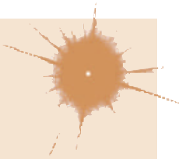

Robert J. Desnick, Manisha Balwani

The porphyrias are metabolic disorders, each resulting from the deficiency or increased activity of a specific enzyme in the heme biosynthetic pathway **(Fig. 428-1 and Table 428-1)**. These enzyme disorders are inherited as autosomal dominant, autosomal recessive, or X-linked traits, with the exception of porphyria cutanea tarda (PCT), which is usually sporadic (Table 428-1). The porphyrias are classified as either *hepatic* or *erythropoietic*, depending on the primary site of overproduction and accumulation of their respective porphyrin precursors or porphyrins **(Tables** 428-1 and **428-2)**, although some have overlapping features. For example, PCT is hepatic and presents with blistering cutaneous photosensitivity, which is typically characteristic of the erythropoietic porphyrias.

The major manifestations of the acute hepatic porphyrias are neurologic, including neuropathic abdominal pain, peripheral motor neuropathy, and mental disturbances, with attacks often precipitated by dieting, certain porphyrinogenic drugs, and hormonal changes. While hepatic porphyrias are symptomatic primarily in adults, rare homozygous variants of the autosomal dominant hepatic porphyrias

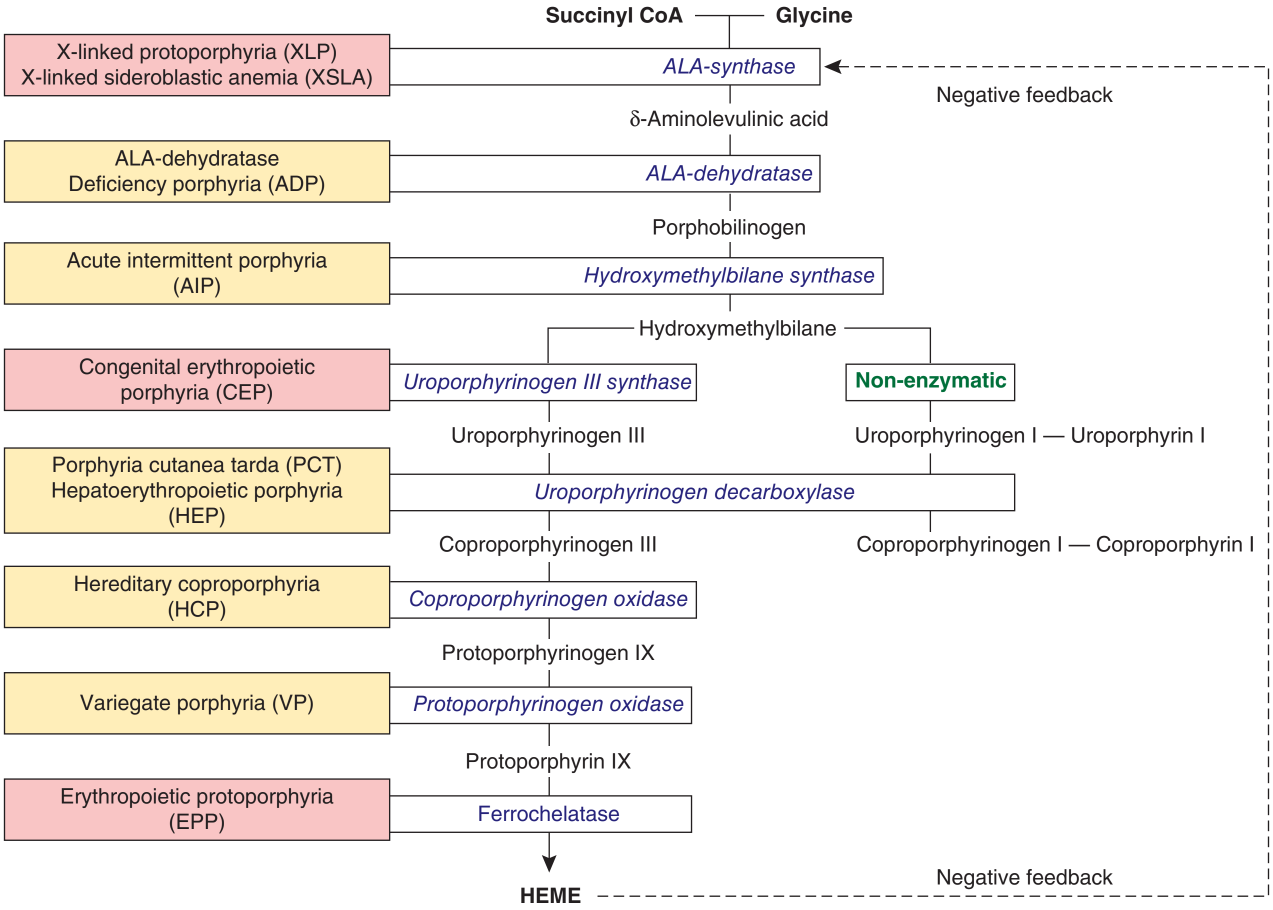
Succinyl CoA
Glycine
X-linked protoporphyria (XLP)
X-linked sideroblastic anemia (XSLA)
ALA-synthase
Negative feedback
δ-Aminolevulinic acid
ALA-dehydratase
Deficiency porphyria (ADP)
ALA-dehydratase
Porphobilinogen
Acute intermittent porphyria (AIP)
Hydroxymethylbilane synthase
Hydroxymethylbilane
Congenital erythropoietic porphyria (CEP)
Uroporphyrinogen III synthase
Non-enzymatic
Uroporphyrinogen III
Uroporphyrinogen I — Uroporphyrin I
Porphyria cutanea tarda (PCT)
Hepatoerythropoietic porphyria (HEP)
Uroporphyrinogen decarboxylase
Coproporphyrinogen III
Coproporphyrinogen I — Coproporphyrin I
Hereditary coproporphyria (HCP)
Coproporphyrinogen oxidase
Protoporphyrinogen IX
Variegate porphyria (VP)
Protoporphyrinogen oxidase
Protoporphyrin IX
Erythropoietic protoporphyria (EPP)
Ferrochelatase
HEME
Negative feedback

FIGURE 428-1 The human heme biosynthetic pathway indicating in *linked boxes* the enzyme that, when deficient or overexpressed, causes the respective porphyria. Hepatic porphyrias are shown in *yellow boxes* and erythropoietic porphyrias in *pink boxes.*

usually manifest clinically prior to puberty. In contrast, the erythropoietic porphyrias usually present at birth or in early childhood with cutaneous photosensitivity or, in the case of congenital erythropoietic porphyria (CEP), even in utero as nonimmune hydrops fetalis. Cutaneous sensitivity to sunlight results from excitation of excess porphyrins in the skin by long-wave ultraviolet light, leading to cell damage, scarring, and disfigurement. Thus, the porphyrias are metabolic disorders in which environmental, physiologic, and genetic factors interact to cause disease.

Because many symptoms of the porphyrias are nonspecific, diagnosis is often delayed. Laboratory measurement of porphyrin precursors (5′-aminolevulinic acid [ALA] and porphobilinogen [PBG]) in

TABLE 428-1 Human Porphyrias: Major Clinical and Laboratory Features

PORPHYRIA	DEFICIENT ENZYME	INHERITANCE	PRINCIPAL SYMPTOMS: NV OR CP+	ENZYME ACTIVITY % OF NORMAL	INCREASED PORPHYRIN PRECURSORS AND/OR PORPHYRINS		
					ERYTHROCYTES	URINE	STOOL
Hepatic Porphyrias							
5-ALA-dehydratase-deficient porphyria (ADP)	ALA-dehydratase	AR	NV	~5	Zn-protoporphyrin	ALA, coproporphyrin III	—
Acute intermittent porphyria (AIP)	HMB-synthase	AD	NV	~50	—	ALA, PBG, uroporphyrin	—
Porphyria cutanea tarda (PCT)	URO-decarboxylase	AD	CP	~20	—	Uroporphyrin, 7-carboxylate porphyrin	Isocoproporphyrin
Hereditary coproporphyria (HCP)	COPRO-oxidase	AD	NV and CP	~50	—	ALA, PBG, coproporphyrin III	Coproporphyrin III
Variegate porphyria (VP)	PROTO-oxidase	AD	NV and CP	~50	—	ALA, PBG, coproporphyrin III	Coproporphyrin III, protoporphyrin
Erythropoietic Porphyrias							
Congenital erythropoietic porphyria (CEP)	URO-synthase	AR	CP	1–5	Uroporphyrin I Coproporphyrin I	Uroporphyrin I[a] Coproporphyrin I[a]	Coproporphyrin I
Erythropoietic protoporphyria (EPP)	Ferrochelatase	AR	CP	~20–30	Protoporphyrin	—	Protoporphyrin
X-linked protoporphyria (XLP)	ALA-synthase 2	XL	CP	>100[b]	Protoporphyrin	—	Protoporphyrin

[a] Type I isomers. [b] Increased activity due to gain-of-function mutations in *ALAS2* exon 11.

Abbreviations: AD, autosomal dominant; ALA, 5-aminolevulinic acid; AR, autosomal recessive; COPRO, coproporphyrin; CP, cutaneous photosensitivity; NV, neurovisceral; PBG, porphobilinogen; PROTO, protoporphyrin; URO, uroporphyrin; XL, X-linked.

TABLE 428-2 Human HEME Biosynthetic Enzymes and Genes

ENZYME	GENE SYMBOL	CHROMOSOMAL LOCATION	CDNA (bp)	GENE		PROTEIN (aa)	SUBCELLULAR LOCATION	KNOWN MUTATIONS[b]	THREE-DIMENSIONAL STRUCTURE[c]
				SIZE (KB)	EXONS[a]				
ALA-synthase									
Housekeeping	*ALAS1*	3p21.1	2199	17	11	640	M	—	
Erythroid-specific	*ALAS2*	Xp11.2	1937	22	11	587	M	>30	—
ALA-dehydratase									
Housekeeping	*ALAD*	9q32	1149	15.9	12 (1A + 2 – 12)	330	C	12	Y
Erythroid-specific	*ALAD*	9q32	1154	15.9	12 (1B + 2 – 12)	330	C	—	
HMB-synthase									
Housekeeping	*HMBS*	11q23.3	1086	11	15 (1 + 3 – 15)	361	C	400	E
Erythroid-specific	*HMBS*	11q23.3	1035	11	15 (2 – 15)	344	C	10	
URO-synthase									
Housekeeping	*UROS*	10q26.2	1296	34	10 (1 + 2B – 10)	265	C	45	H
Erythroid-specific	*UROS*	10q26.2	1216	34	10 (2A + 2B – 10)	265	C	4	
URO-decarboxylase	*UROD*	1p34.1	1104	3	10	367	C	122	H
COPRO-oxidase	*CPOX*	3q12.1	1062	14	7	354	M	70	H
PROTO-oxidase	*PPOX*	1q23.3	1431	5.5	13	477	M	181	—
Ferrochelatase	*FECH*	18q21.31	1269	45	11	423	M	192	B

[a] Number of exons and those encoding separate housekeeping and erythroid-specific forms indicated in parentheses. [b] Number of known mutations from the Human Gene Mutation Database (*www.hgmd.org*). [c] Crystallized from human (H), murine (M), *Escherichia coli* (E), *Bacillus subtilis* (B), or yeast (Y) purified enzyme; references in Protein Data Bank (*www.rcsb.org*).

Abbreviations: ALA, 5-aminolevulinic acid; C, cytoplasm; COPRO, coproporphyrin; HMB, hydroxymethylbilane; M, mitochondria; PROTO, protoporphyrin; URO, uroporphyrin.

Source: Reproduced with permission from KE Anderson et al: Disorders of heme biosynthesis: X-linked sideroblastic anemia and the porphyrias, in CR Scriver: *The Metabolic and Molecular Bases of Inherited Diseases*. New York, NY: McGraw-Hill; 2001.

the urine or porphyrins in the urine, plasma, erythrocytes, or feces is required to confirm or exclude the various types of porphyria (see below). However, a definite diagnosis requires demonstration of the specific gene defect **(Table 428-3)**. The genes encoding all the heme biosynthetic enzymes have been characterized, permitting identification of the mutations causing each porphyria (Table 428-2). Molecular genetic analyses now make it possible to provide precise heterozygote or homozygote identification and prenatal diagnoses in families with known mutations.

In addition to recent reviews of the porphyrias, informative and up-to-date websites are sponsored by the United Porphyrias Association (*www.porphyria.org*) and the European Porphyria Network (*https://porphyria.eu/*). An extensive list of unsafe and safe drugs for individuals with acute porphyrias is provided at the Drug Database for Acute Porphyrias (*www.drugs-porphyria.org*).

GLOBAL CONSIDERATIONS

The porphyrias are panethnic metabolic diseases that affect individuals globally. The acute hepatic porphyrias—acute intermittent porphyria (AIP), hereditary coproporphyria (HCP), and variegate porphyria (VP)—are autosomal dominant disorders. The frequency of AIP, the most common acute hepatic porphyria, is ~1 in 20,000 among Caucasian individuals of Western European ancestry, and it is particularly frequent in Scandinavians, where the frequency in Sweden is ~1 in 10,000. VP is particularly frequent in South Africa, and its high prevalence (>10,000 affected patients) is in part due to a genetic "founder effect." The autosomal recessive acute hepatic porphyria, ALA-dehydratase-deficient porphyria (ADP), is extremely rare, and <20 patients have been identified worldwide.

The erythropoeitic protoporphyrias—CEP, erythropoietic protoporphyria (EPP), and X-linked protoporphyria (XLP)—also are panethnic. EPP is likely the most common porphyria, while CEP is very rare with about 200 reported cases worldwide. The frequency of EPP varies globally since most patients have the common low expression *FECH* allele, which ranges in frequency in different populations. This allele rarely occurs in Africans, is present in ~10% of the Caucasians, and is frequent (~30%) in the Japanese.

The autosomal recessive porphyrias, ADP, CEP, EPP, and hepatoerythropoietic porphyria (HEP), are more frequent in regions with high rates of consanguineous marriages. PCT, which is typically sporadic, occurs more frequently in countries in which its predisposing risk factors such as hepatitis C and HIV are more prevalent. The reported prevalence of EPP in the Caucasian population ranges from 1 in ~75,000 to 1 in ~150,000.

HEME BIOSYNTHESIS

Heme biosynthesis involves eight enzymatic steps in the conversion of glycine and succinyl-CoA to heme (**Fig. 428-2** and Table 428-2). These eight enzymes are encoded by nine genes, as the first enzyme in the pathway, ALA-synthase, has two genes that encode unique housekeeping (*ALAS1*) and erythroid-specific (*ALAS2*) isozymes. The first and last three enzymes in the pathway are located in the mitochondria, whereas the other four are in the cytosol. Heme is required for a variety of hemoproteins such as hemoglobin, myoglobin, respiratory cytochromes, and the cytochrome P450 (CYP) enzymes. Hemoglobin synthesis in erythroid precursor cells accounts for ~85% of daily heme synthesis in humans. Hepatocytes account for most of the rest, primarily for the synthesis of CYPs, which are especially abundant in the liver endoplasmic reticulum, and turn over more rapidly than many other hemoproteins, such as the mitochondrial respiratory cytochromes. As shown in Fig. 428-2, the pathway intermediates are the porphyrin precursors, ALA and PBG, and porphyrins (mostly in their reduced forms, known as *porphyrinogens*). At least in humans, these intermediates do not accumulate in significant amounts under normal conditions or have important physiologic functions.

The first enzyme, ALA-synthase, catalyzes the condensation of glycine, activated by pyridoxal phosphate and succinyl-coenzyme A, to form ALA. In the liver, this rate-limiting enzyme can be induced by a variety of drugs, steroids, and other chemicals. Distinct nonerythroid (e.g., housekeeping) and erythroid-specific forms of ALA-synthase are encoded by separate genes located on chromosome 3p21.1 (*ALAS1*) and Xp11.2 (*ALAS2*), respectively. Defects in the erythroid gene *ALAS2* that decrease its activity cause an X-linked sideroblastic anemia (XLSA). Gain-of-function mutations in the last exon (11) of *ALAS2* that increase its activity cause an X-linked form of EPP, known as XLP.

The second enzyme, ALA-dehydratase, catalyzes the condensation of two molecules of ALA to form PBG. Hydroxymethylbilane synthase (HMB-synthase; also known as PBG-deaminase) catalyzes

TABLE 428-3 Diagnosis of Acute and Cutaneous Porphyrias

SYMPTOMS	FIRST-LINE TEST: ABNORMALITY	POSSIBLE PORPHYRIA	SECOND-LINE TESTING IF FIRST-LINE TESTING IS POSITIVE: TO INCLUDE: URINE (U), PLASMA (P), AND FECAL (F) PORPHYRINS; FOR ACUTE PORPHYRIAS, ADD RED BLOOD CELL (RBC) HMB-SYNTHASE; FOR BLISTERING SKIN LESIONS, ADD P AND RBC PORPHYRINS	CONFIRMATORY TEST: ENZYME ASSAY AND/OR MUTATION ANALYSIS
Neurovisceral	**Spot U:** ↑↑ALA and normal PBG	ADP	**U porphyrins:** ↑↑, mostly COPRO III **P** and **F porphyrins:** normal or slightly ↑ **RBC HMB-synthase:** normal	Rule out other causes of elevated ALA; ↓↓RBC ALA-dehydratase activity (<10%); ALA-dehydratase mutation analysis
	Spot U: ↑↑PBG	AIP	**U porphyrins:** ↑↑, mostly URO and COPRO **P** and **F porphyrins:** normal or slightly ↑ **RBC HMB-synthase:** usually ↓	HMB-synthase mutation analysis
	“	HCP	**U porphyrins:** ↑↑, mostly COPRO III **P porphyrins:** normal or slightly ↑ (↑ if skin lesions present) **F porphyrins:** ↑↑, mostly COPRO III	Measure RBC HMB-synthase: normal activity COPRO-oxidase mutation analysis
	“	VP	**U porphyrins:** ↑↑, mostly COPRO III **P porphyrins:** ↑↑ (characteristic fluorescence peak at neutral pH) **F porphyrins:** ↑↑, mostly COPRO and PROTO	Measure RBC HMB-synthase: normal activity PROTO-oxidase mutation analysis
Blistering skin lesions	**P:** ↑ porphyrins	PCT and HEP	**U porphyrins:** ↑↑, mostly URO and heptacarboxylate porphyrin **P porphyrins:** ↑↑ **F porphyrins:** ↑↑, including increased isocoproporphyrin **RBC porphyrins:** ↑↑ zinc PROTO in HEP[a]	RBC URO-decarboxylase activity: half-normal in familial PCT (~20% of all PCT cases); substantially deficient in HEP URO-decarboxylase mutation analysis: mutation(s) present in familial PCT (heterozygous) and HEP (homozygous)
	“	HCP and VP	See HCP and VP above. Also, U ALA and PBG: may be ↑	
	“	CEP	**RBC and U porphyrins:** ↑↑, mostly URO I and COPRO I **F porphyrins:** ↑↑; mostly COPRO I	↓↓ RBC URO-synthase activity (<15%) URO-synthase mutation analysis
Nonblistering photosensitivity	**P:** porphyrins usually ↑	EPP	**RBC porphyrins:** ↓↓, mostly free PROTO **U porphyrins:** normal **F porphyrins:** normal or ↓, mostly PROTO	*FECH* mutation analysis
	P: porphyrins usually ↑	XLP	**RBC porphyrins:** ↑↑, approximately equal free and zinc PROTO **U porphyrins:** normal **F porphyrins:** normal or ↑, mostly PROTO	*ALAS2* mutation analysis

[a]Nonspecific increases in zinc protoporphyrins are common in other porphyrias.

Abbreviations: ADP, 5-ALA-dehydratase-deficient porphyria; AIP, acute intermittent porphyria; ALA, 5-aminolevulinic acid; CEP, congenital erythropoietic porphyria; COPRO I, coproporphyrin I; COPRO III, coproporphyrin III; EPP, erythropoietic protoporphyria; F, fecal; HCP, hereditary coporphyria; HEP, hepatoerythropoietic porphyria; ISOCOPRO, isocoproporphyrin; P, plasma; PBG, porphobilinogen; PCT, porphyria cutanea tarda; PROTO, protoporphyrin IX; RBC, erythrocytes; U, urine; URO I, uroporphyrin I; URO III, uroporphyrin III; VP, variegate porphyria; XLP, X-linked protoporphyria.

Source: Data from KE Anderson et al: Recommendations for the diagnosis and treatment of the acute porphyrias. Ann Intern Med 142:439, 2005.

the head-to-tail condensation of four PBG molecules by a series of deaminations to form the linear tetrapyrrole, HMB. Uroporphyrinogen III synthase (URO-synthase) catalyzes the rearrangement and rapid cyclization of HMB to form the asymmetric, physiologic, octacarboxylate porphyrinogen, uroporphyrinogen (URO'gen) III.

The fifth enzyme in the pathway, uroporphyrinogen decarboxylase (URO-decarboxylase), catalyzes the sequential removal of the four carboxyl groups from the acetic acid side chains of URO'gen III to form coproporphyrinogen (COPRO'gen) III, a tetracarboxylate porphyrinogen. This compound then enters the mitochondrion via a specific transporter, where COPRO-oxidase, the sixth enzyme, catalyzes the decarboxylation of two of the four propionic acid groups to form the two vinyl groups of protoporphyrinogen (PROTO'gen) IX, a decarboxylate porphyrinogen. Next, PROTO-oxidase oxidizes PROTO'gen to protoporphyrin IX by the removal of six hydrogen atoms. The product of the reaction is a porphyrin (oxidized form), in contrast to the preceding tetrapyrrole intermediates, which are porphyrinogens (reduced forms). Finally, ferrous iron is inserted into protoporphyrin IX to form heme, a reaction catalyzed by the eighth enzyme in the pathway, FECH (also known as heme synthase or protoheme ferrolyase).

REGULATION OF HEME BIOSYNTHESIS

Regulation of heme synthesis differs in the two major heme-forming tissues, the liver and erythron. In the liver, the concentration of "free" heme regulates the synthesis and mitochondrial translocation of the housekeeping form of ALA-synthase 1. Heme represses the synthesis of the ALA-synthase 1 messenger RNA (mRNA) and interferes with the transport of the enzyme from the cytosol into mitochondria. Hepatic ALA-synthase 1 is increased by many of the same chemicals that induce the CYP enzymes in the endoplasmic reticulum of the liver. Because most of the heme in the liver is used for the synthesis of CYP enzymes, hepatic ALA-synthase 1 and the CYPs are regulated in a coordinated fashion, and many drugs that induce hepatic ALA-synthase 1 also induce CYP gene expression. The other hepatic heme biosynthetic enzymes are presumably expressed at constant levels, although their relative activities and kinetic properties differ. For example, normal individuals have high activities of ALA-dehydratase but low activities of HMB-synthase, the latter being the second rate-limiting step in the pathway.

In the erythron, novel regulatory mechanisms allow for the production of the very large amounts of heme needed for hemoglobin synthesis. The response to stimuli for hemoglobin synthesis occurs during cell differentiation, leading to an increase in cell number. In contrast, the erythroid-specific ALA-synthase 2 is expressed at higher levels than the housekeeping enzyme, and erythroid-specific control mechanisms regulate other pathway enzymes as well as iron transport into erythroid cells. Separate erythroid-specific and nonerythroid or "housekeeping" transcripts are known for the first four enzymes in the pathway. As noted above, housekeeping- and erythroid-specific ALA-synthases are encoded by genes on different chromosomes, but for each of the

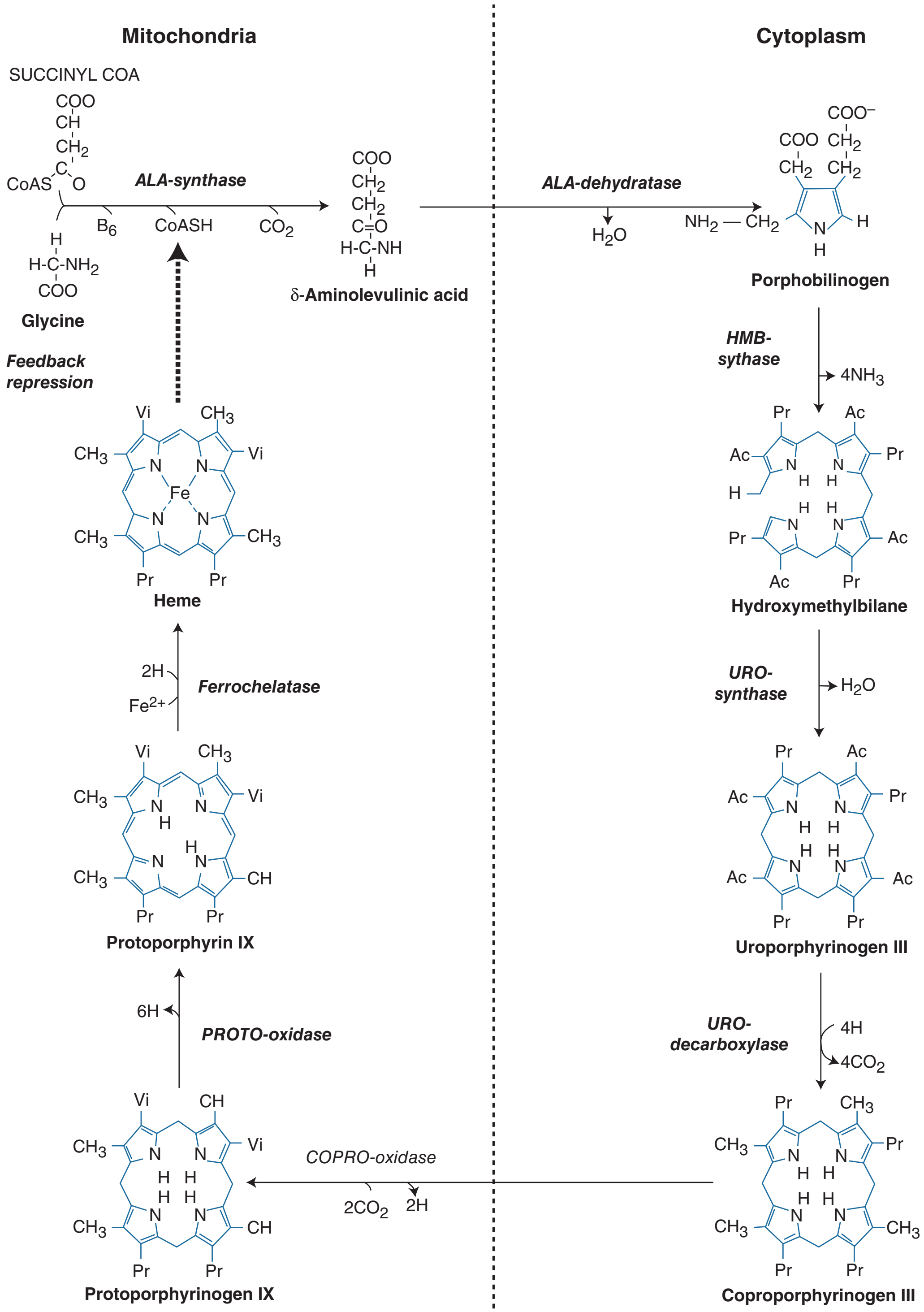
Mitochondria
Cytoplasm
SUCCINYL COA
COO
CH
CH2
C
CoAS
O
ALA-synthase
B6
CoASH
CO2
H
H-C-NH2
COO
Glycine
COO
CH2
CH2
C=O
H-C-NH
H
δ-Aminolevulinic acid
ALA-dehydratase
H2O
NH2 — CH2
COO
CH2
COO–
CH2
CH2
N
H
H
Porphobilinogen
Feedback repression
Vi
CH3
Vi
CH3
CH3
CH3
Fe
Pr
Pr
Heme
HMB-sythase
4NH3
Pr
Ac
Ac
Pr
H
Pr
Ac
Ac
Pr
Hydroxymethylbilane
2H
Fe2+
Ferrochelatase
URO-synthase
H2O
Protoporphyrin IX
Uroporphyrinogen III
6H
PROTO-oxidase
URO-decarboxylase
4H
4CO2
COPRO-oxidase
2CO2
2H
Protoporphyrinogen IX
Coproporphyrinogen III

FIGURE 428-2 The heme biosynthetic pathway showing the eight enzymes and their substrates and products. Four of the enzymes are localized in the mitochondria and four in the cytosol.

next three genes in the pathway, both erythroid and nonerythroid transcripts are transcribed by alternative promoters from their single respective genes (Table 428-2).

CLASSIFICATION OF THE PORPHYRIAS

As mentioned above, the porphyrias can be classified as either *hepatic* or *erythropoietic*, depending on whether the heme biosynthetic intermediates that accumulate arise initially from the liver or developing erythrocytes, or as *acute* or *cutaneous*, based on their clinical manifestations. Table 428-1 lists the porphyrias, their principal symptoms, and major biochemical abnormalities. Three of the five hepatic porphyrias—AIP, HCP, and VP—usually present during adult life with acute attacks of neurologic manifestations and elevated levels of one or both of the porphyrin precursors, ALA and PBG, and are thus classified as *acute hepatic porphyrias*. Patients with ADP have presented in infancy and adolescence and typically have elevated ALA with normal or slightly elevated PBG levels. The fifth hepatic disorder, PCT, presents with blistering skin lesions. HCP and VP also may have cutaneous manifestations similar to PCT.

The erythropoietic porphyrias—CEP, EPP, and XLP—are characterized by elevations of porphyrins in bone marrow and erythrocytes and present with cutaneous photosensitivity. The skin lesions in CEP resemble PCT but are usually much more severe, whereas EPP and XLP cause a more immediate, severe, painful, and nonblistering type of photosensitivity. EPP is the most common porphyria to cause symptoms before puberty. About 20% of EPP patients develop minor abnormalities of liver function, with up to ~5% developing hepatic complications that can lead to liver failure requiring liver transplantation. XLP has a clinical presentation similar to EPP causing photosensitivity and liver disease.

DIAGNOSIS OF PORPHYRIA

A few specific and sensitive first-line laboratory tests should be used whenever symptoms or signs suggest the diagnosis of porphyria (Table 428-3). If a first-line test is significantly abnormal, more comprehensive testing should follow to establish the type of porphyria, including the specific causative gene mutation.

Acute Hepatic Porphyrias An acute hepatic porphyria should be suspected in patients with neurovisceral symptoms after puberty. Symptoms include acute abdominal pain, nausea, vomiting, tachycardia, hypertension, and motor neuropathy. As these symptoms are common, other causes should be ruled out. The diagnosis is made by measuring urinary porphyrin precursors (ALA and PBG) in a spot sample of urine (Fig. 428-2). Urinary PBG is always increased during acute attacks of AIP, HCP, and VP and is not substantially increased in any other medical condition. Therefore, this measurement is both sensitive and specific. Results from spot (single-void) urine specimens are highly informative because very substantial increases in PBG are expected during acute attacks of porphyria. A 24-h collection is unnecessary. The same spot urine specimen should be saved for quantitative determination of ALA, PBG, and creatinine, in order to confirm the qualitative PBG result and also to detect patients with ADP. Urinary porphyrins may remain increased longer than porphyrin precursors in HCP and VP. Therefore, it is useful to measure total urinary porphyrins in the same sample, keeping in mind that urinary porphyrin increases are often nonspecific. Measurement of urinary porphyrins alone should be avoided for screening, because these may be increased in disorders other than porphyrias, such as chronic liver disease, and misdiagnoses of porphyria can result from minimal increases in urinary porphyrins that have no diagnostic significance. Measurement of erythrocyte HMB-synthase is not useful as a first-line test. Moreover, the enzyme activity is not decreased in all AIP patients, a borderline low normal value is not diagnostic, and the enzyme is not deficient in other acute porphyrias.

More extensive testing is justified when an initial test is positive. A substantial increase in PBG may be due to AIP, HCP, or VP. These acute porphyrias can be distinguished by measuring urinary porphyrins (using the same spot urine sample), fecal porphyrins, and plasma porphyrins. Assays for COPRO-oxidase or PROTO-oxidase are not available for clinical testing. More specifically, mutation analysis by sequencing the genes encoding HMB-synthase, COPRO-oxidase, and PROTO-oxidase will detect almost all disease-causing mutations and is diagnostic even when the levels of urinary ALA and PBG have returned to normal or near normal.

Cutaneous Porphyrias Blistering skin lesions due to porphyria are virtually always accompanied by increases in total plasma porphyrins. A fluorometric method is preferred, because the plasma porphyrins in VP are mostly covalently linked to plasma proteins and may be less readily detected by high-performance liquid chromatography (HPLC). The normal range for plasma porphyrins is somewhat increased in patients with end-stage renal disease.

Although a total plasma porphyrin determination will usually detect EPP and XLP, an erythrocyte protoporphyrin determination is more sensitive. Increases in erythrocyte protoporphyrin occur in many other conditions. Therefore, the diagnosis of EPP must be confirmed by showing a predominant increase in free protoporphyrin rather than zinc protoporphyrin. In XLP, both free and zinc protoporphyrin are markedly increased. Interpretation of laboratory reports can be difficult, because the term *free erythrocyte protoporphyrin* sometimes actually represents zinc protoporphyrin.

The various porphyrias that cause blistering skin lesions can be differentiated by measuring porphyrins in urine, feces, and plasma. The porphyrias should be confirmed by genetic testing and the demonstration of the causative pathogenic variant. It is often difficult to diagnose or "rule out" porphyria in patients who have had suggestive symptoms months or years in the past and in relatives of patients with acute porphyrias, because porphyrin precursors and porphyrins may be normal. In those situations, detection of the specific gene mutation in the index case can make the diagnosis and facilitate the diagnosis and genetic counseling of at-risk relatives. With the increased access and accuracy of genetic testing, this often precedes secondary biochemical testing in clinical practice. Consultation with a specialist laboratory and physician will assist in selecting the heme biosynthetic gene or genes to be sequenced.

THE HEPATIC PORPHYRIAS

Markedly elevated plasma and urinary concentrations of the porphyrin precursors, ALA and/or PBG, which originate from the liver, are especially evident during attacks of neurologic manifestations of the four acute porphyrias—ADP, AIP, HCP, and VP. In PCT, excess porphyrins also accumulate initially in the liver and cause chronic blistering of sun-exposed areas of the skin.

ALA-DEHYDRATASE-DEFICIENT PORPHYRIA

ADP is a rare, autosomal recessive, acute hepatic porphyria caused by a severe deficiency of ALA-dehydratase activity. To date, there are only a dozen documented cases, some in children or young adults, in which specific gene mutations have been identified. These affected homozygotes had <10% of normal ALA-dehydratase activity in erythrocytes, but their clinically asymptomatic parents and heterozygous relatives had about half-normal levels of activity and did not excrete increased levels of ALA. The frequency of ADP is unknown, but the frequency of heterozygous individuals with <50% normal ALA-dehydratase activity was ~2% in a screening study in Sweden. Because there are multiple causes for deficient ALA-dehydratase activity, it is important to confirm the diagnosis of ADP by mutation analysis.

Clinical Features The clinical presentation depends on the amount of residual ALA-dehydratase activity. Four of the documented patients were male adolescents with symptoms resembling those of AIP, including abdominal pain and neuropathy. One patient was an infant with more severe disease, including failure to thrive beginning at birth. The earlier age of onset and more severe manifestations in this patient reflect a more significant deficiency of ALA-dehydratase activity. Another patient developed an acute motor polyneuropathy at age 63 that was associated with a myeloproliferative disorder. He was heterozygous for an δ-aminolevulinic acid dehydratase (*ALAD*) mutation that presumably was present in erythroblasts that underwent clonal expansion due to the bone marrow malignancy.

Diagnosis All patients had significantly elevated levels of plasma and urinary ALA and urinary coproporphyrin (COPRO) III; ALAD activities in erythrocytes were <10% of normal. Hereditary tyrosinemia type 1 (fumarylacetoacetase deficiency) and lead intoxication should be considered in the differential diagnosis because either succinylacetone (which accumulates in hereditary tyrosinemia and is structurally similar to ALA) or lead can inhibit ALA-dehydratase, increase urinary excretion of ALA and COPRO III, and cause manifestations that resemble those of the acute porphyrias. Heterozygotes are clinically asymptomatic and do not excrete increased levels of ALA but can be detected by demonstration of intermediate levels of erythrocyte ALA-dehydratase activity or a specific mutation in the *ALAD* gene. To date, molecular studies of ADP patients have identified 12 pathogenic mutations, including missense mutations, splice-site mutations, and a two-base deletion in the *ALAD* gene (Human Gene Mutation Database; *www.hgmd.org*). The parents in each case were not consanguineous, and the index cases had inherited a different *ALAD* mutation from each parent. Prenatal diagnosis of this disorder is possible by determination of ALA-dehydratase activity and/or gene mutations in cultured chorionic villi or amniocytes.

Treatment The treatment of ADP acute attacks is similar to that of AIP (see below). The severely affected infant referred to above was supported by hyperalimentation and periodic blood transfusions but did not respond to intravenous hemin and died after liver transplantation.

ACUTE INTERMITTENT PORPHYRIA

This hepatic porphyria is an autosomal dominant condition resulting from the half-normal level of HMB-synthase activity. The disease is widespread but is especially common in Scandinavia and Great Britain. Clinical expression is highly variable, and activation of the disease is often related to environmental or hormonal factors, such as drugs, diet, and steroid hormones. Attacks can be prevented by avoiding known precipitating factors. Rare homozygous dominant AIP also has been described in children (see below).

Clinical Features Induction and increased expression of the rate-limiting hepatic gene *ALAS1* in heterozygotes who have half-normal HMB-synthase activity is thought to underlie the acute attacks in AIP. The disorder remains latent (or asymptomatic) in the great majority of those who are heterozygous for pathogenic *HMBS* mutations, and this is almost always the case prior to puberty. In patients with no history of acute symptoms, porphyrin precursor excretion is usually normal, suggesting that half-normal hepatic HMB-synthase activity is sufficient and that hepatic ALA-synthase activity is not increased. However, under conditions where heme synthesis is increased in the liver, half-normal HMB-synthase activity may become limiting, and ALA, PBG, and other heme pathway intermediates may accumulate and be excreted in the urine. Common precipitating factors include endogenous and exogenous steroids, porphyrinogenic drugs, alcohol ingestion, and low-calorie diets, usually instituted for weight loss.

The fact that AIP is almost always latent before puberty suggests that adult levels of steroid hormones are important for clinical expression. Symptoms are more common in women, suggesting a role for estrogens or progestins. Premenstrual attacks are probably due to increasing endogenous progesterone during the luteal phase of the menstrual cycle. Acute porphyrias are sometimes exacerbated by exogenous steroids, including oral contraceptive preparations containing progestins. Surprisingly, pregnancy is usually well tolerated, suggesting that beneficial metabolic changes may ameliorate the effects of high levels of progesterone. Extensive lists of unsafe and safe drugs are available on websites sponsored by the United Porphyria Association (*www.porphyria.org*) and the European Porphyria Network (*https://porphyria.eu/*), and at the Drug Database for Acute Porphyrias website (*www.drugs-porphyria.org*). Reduced intake of calories and carbohydrate, as may occur with illness or attempts to lose weight, can also increase porphyrin precursor excretion and induce attacks of porphyria. Studies in a knockout AIP mouse model indicate that the hepatic *ALAS1* gene is regulated, in part, by the peroxisome proliferator-activated receptor γ coactivator 1α (PGC-1α). Hepatic PGC-1α is induced by fasting, which in turn activates *ALAS1* transcription, resulting in increased heme biosynthesis. This finding suggests an important link between nutritional status and the attacks in acute porphyrias. Attacks also can be provoked by infections, surgery, and ethanol.

Because the neurovisceral symptoms rarely occur before puberty and are often nonspecific, a high index of suspicion is required to make the diagnosis. The disease can be disabling but is rarely fatal. Abdominal pain, the most common symptom, is poorly localized but may be associated with cramping, ileus, abdominal distention, and decreased bowel sounds. However, increased bowel sounds and diarrhea may occur. Abdominal tenderness, fever, and leukocytosis are usually absent or mild because the symptoms are neurologic rather than inflammatory. Nausea; vomiting; constipation; tachycardia; hypertension; mental symptoms; pain in the limbs, head, neck, or chest; muscle weakness; sensory loss; dysuria; and urinary retention are characteristic. Tachycardia, hypertension, restlessness, tremors, and excess sweating are due to sympathetic overactivity.

The peripheral neuropathy is due to axonal degeneration (rather than demyelinization) and primarily affects motor neurons. Significant neuropathy does not occur with all acute attacks; abdominal symptoms are usually more prominent. Motor neuropathy affects the proximal muscles initially, more often in the shoulders and arms. The course and degree of involvement are variable and sometimes may be focal and involve cranial nerves. Deep tendon reflexes initially may be normal or hyperactive but become decreased or absent as the neuropathy advances. Sensory changes such as paresthesia and loss of sensation are less prominent. Progression to respiratory and bulbar paralysis and death occurs especially when the diagnosis and treatment are delayed. Sudden death may result from sympathetic overactivity and cardiac arrhythmia.

Mental symptoms such as anxiety, insomnia, depression, disorientation, hallucinations, and paranoia can occur in acute attacks. Seizures can be due to neurologic effects or to hyponatremia. Treatment of seizures is difficult because most antiseizure drugs can exacerbate AIP (clonazepam may be safer than phenytoin or barbiturates). Hyponatremia results from hypothalamic involvement and inappropriate vasopressin secretion or from electrolyte depletion due to vomiting, diarrhea, poor intake, or excess renal sodium loss. When an attack resolves, abdominal pain may disappear within hours, and paresis begins to improve within days and may continue to improve over several years.

Homozygous dominant AIP (HD-AIP) is a rare form of AIP in which patients inherit *HMBS* mutations from each of their heterozygous parents and, therefore, have very low (<2%) enzyme activity. The disease has been described in a Dutch girl, two young British siblings, and a Spanish boy. In these homozygous affected patients, the disease presented in infancy with failure to thrive, developmental delay, bilateral cataracts, and/or hepatosplenomegaly. Urinary ALA and PBG concentrations were markedly elevated. All of these patients' *HMBS* mutations (R167W, R167Q, and R172Q) were in exon 10 within five bases of each other. Studies of the brain magnetic resonance images (MRIs) of children with homozygous AIP have suggested damage primarily in white matter that was myelinated postnatally, while tracks that myelinated prenatally were normal. Most children with homozygous AIP die at an early age. Recently, later-onset HD-AIP was described in an adult with leukoencephalopathy.

Diagnosis ALA and PBG levels are substantially increased in plasma and urine, especially during acute attacks. For example, urinary PBG excretion during an attack is usually 50–200 mg/24 h (220–880 μmol/24 h) (normal, 0–4 mg/24 h [0–18 μmol/24 h]), and urinary ALA excretion is 20–100 mg/24 h (150–760 μmol/24 h) (normal, 1–7 mg/24 h [8–53 μmol/24 h]). Because levels often remain high after symptoms resolve, the diagnosis of an acute attack in a patient with biochemically proven AIP is based primarily on clinical features. Excretion of ALA and PBG decreases over a few days after intravenous hemin administration or in a day after subcutaneous givosiran (see below). A normal urinary PBG level before hemin effectively excludes AIP as a cause for current symptoms. Fecal porphyrins are usually normal or minimally increased in AIP, in contrast to HCP and VP. Most AIP heterozygotes with no history of symptoms have normal urinary excretion of ALA and PBG and are classified as "latent" or "asymptomatic heterozygotes." Patients can also have high levels of urinary PBG and ALA with no clinical symptoms. These patients may have a previous history of an acute attack and are classified as asymptomatic high excretors (ASHE) or chronic high excretors (CHE). Therefore, the detection of the family's *HMBS* mutation will diagnose heterozygous asymptomatic family members. A urinary ALA and PBG will diagnosis CHE patients who may have a higher risk of an attack if they experience a precipitating factor such as administration of a porphyrinogenic drug.

Patients with *HMBS* mutations in the initiation of translation codon in exon 1 and in the intron 15′-splice donor site have normal enzyme levels in erythrocytes and deficient activity only in nonerythroid tissues. This occurs because the erythroid and housekeeping forms of HMB-synthase are encoded by a single gene, which has two promoters. Thus, the enzyme assay may not be diagnostic, and genetic testing should be used to confirm the diagnosis.

More than 550 *HMBS* mutations have been identified in AIP, including missense, nonsense, AND splicing lesions, insertions, and deletions, with most mutations found in only one or a few families (Human Gene Mutation Database, *www.hgmd.org*). The prenatal diagnosis of a fetus at risk can be made by analysis of the familial mutation in cultured amniotic cells or chorionic villi. However, this is seldom done because the prognosis of individuals with *HMBS* mutations is generally favorable.

TREATMENT

Acute Intermittent Porphyria

During acute attacks, narcotic analgesics may be required for abdominal pain, and phenothiazines are useful for nausea, vomiting, anxiety, and restlessness. Chloral hydrate can be given for insomnia, and benzodiazepines are probably safe in low doses if a minor tranquilizer is required. Carbohydrate loading, usually with intravenous glucose (at least 300 g daily), may be effective in milder acute attacks of porphyria (without paresis, hyponatremia, etc.) if hemin or givosiran (see below) is not available. Intravenous hemin was approved by the U.S. Food and Drug Administration (FDA) in 1984 to treat acute attacks and has been used effectively as a first-line therapy for acute attacks. The standard regimen is 3–4 mg/kg per d of heme, in the form of lyophilized hematin (Panhematin, Recordati Rare Diseases), heme albumin (hematin reconstituted with human albumin), or heme arginate (Orphan Europe), infused daily for 4 days or longer to reduce the pain. Heme arginate and heme albumin are chemically stable and are less likely than hematin to produce phlebitis or an anticoagulant effect. Recovery depends on the degree of neuronal damage and usually occurs in days, if therapy is started early. Recovery from severe motor neuropathy may require months or years. Identification and avoidance of inciting factors can hasten recovery from an attack and prevent future attacks. Inciting factors are usually multiple, and removal of one or more hastens recovery and helps prevent future attacks. Frequent attacks that occur during the luteal phase of the menstrual cycle may be prevented with a gonadotropin-releasing hormone analogue, which prevents ovulation and progesterone production, or by prophylactic hematin or givosiran administration.

In 2019, a hepatocyte-targeted RNA interference (RNAi) therapy, givosiran (Givlarri, Alnylam Pharmaceuticals), was approved by the FDA and the European Medicines Agency (EMA) for the treatment of the acute hepatic porphyrias. Givosiran, a monthly subcutaneous injection of 2.5 mg/kg, is designed to silence the expression of hepatic *ALAS1* mRNA and was initially shown in clinical trials to markedly reduce ALA and PBG levels in CHE patients and in patients with recurrent attacks. In a phase 3 trial in acute hepatic porphyria patients with recurrent attacks, the RNAi therapy significantly reduced the frequency of acute attacks, decreased hemin utilization, and improved daily pain scores.

The long-term risk of hypertension and chronic renal disease is increased in AIP; a number of patients have undergone successful renal transplantation. Studies have shown that up to 59% of symptomatic AIP patients will develop chronic kidney disease. The PEPT2 receptor polymorphic genotype affects the severity and prognosis of porphyria-associated kidney disease with the high affinity polymorphic PEPT2 *1 allele and the PEPT2 genotypes *1*1 and, to a lesser degree, *1*2 associated with decreasing kidney function. Chronic, low-grade abnormalities in liver function tests are common, and the risk of hepatocellular carcinoma is increased. Hepatic imaging is recommended at least every 6 months for early detection of these tumors. Other long-term complications include neuropathy, fatigue, chronic pain, nausea, depression, and/or anxiety.

Orthotopic liver transplantation (OLT) has been effective in patients with severe, disabling, intractable attacks that are refractory to hemin therapy. Reports from both the United Kingdom and the United States show a marked improvement with no subsequent attacks, an improvement in the neuropathic manifestations, and normalization of the urinary PBG and ALA levels after liver transplantation. OLT is associated with morbidity and mortality and should be considered a treatment of last resort in these patients. In addition, patients who already have advanced neuropathy are considered poor risks for transplantation. Of note, the effectiveness of givosiran in such patients has not been described. Some patients with both recurrent attacks and end-stage renal disease have benefitted from combined liver and kidney transplantation.

Liver-directed gene therapy has proven successful in the prevention of drug-induced biochemical attacks in a murine model of human AIP, and clinical trials of adeno-associated virus vector (AAV)-*HMBS* gene transfer were carried out over a decade ago. Although the therapy was safe, there was essentially no biochemical evidence of its effectiveness, nor did it prevent recurrent attacks in the treated patients. Recent advances have led to FDA-approved gene-editing/gene therapies for genetic disorders and may lead to future treatments and/or cures for the acute hepatic and erythropoietic porphyrias.

PORPHYRIA CUTANEA TARDA

PCT, the most common of the porphyrias, can be either sporadic (type 1) or familial (type 2) and can also develop after exposure to halogenated aromatic hydrocarbons. Hepatic URO-decarboxylase is deficient in all types of PCT, and for clinical symptoms to manifest, this enzyme deficiency must be substantial (~20% of normal activity or less). Its deficiency is currently attributed to generation of an URO-decarboxylase inhibitor in the liver, which forms a uroporphomethene in the presence of iron and under conditions of oxidative stress. The majority of PCT patients (~80%) have no *UROD* mutations and are said to have sporadic (type 1) disease. PCT patients heterozygous for *UROD* mutations have the familial (type 2) PCT. In these patients, inheritance of a *UROD* mutation from one parent results in half-normal enzymatic activity in liver and all other tissues, which is a significant predisposing factor, but is insufficient by itself to cause symptomatic PCT. As discussed below, other genetic and environmental factors contribute to susceptibility for both types of PCT. Because penetrance of the genetic trait is low, many patients with familial (type 2) PCT have no family history of the disease. HEP is an autosomal recessive form of porphyria due to the inheritance of two pathogenic *UROD* mutations resulting in the marked systemic deficiency of URO-decarboxylase activity with clinical symptoms in childhood.

Clinical Features Blistering skin lesions that appear most commonly on the backs of the hands are the major clinical feature (Fig. 428-3). These rupture and crust over, leaving areas of atrophy and scarring. Lesions may also occur on the forearms, face, legs, and feet. Skin friability and small white papules termed milia are common, especially on the backs of the hands and fingers. Hypertrichosis and hyperpigmentation, especially of the face, are especially troublesome in women. Occasionally, the skin over sun-exposed areas becomes severely thickened, with scarring and calcification that resembles systemic sclerosis. Neurologic features are absent.

A number of susceptibility factors, in addition to inherited *UROD* mutations in type 2 PCT, are recognized clinically and can affect management. These include hepatitis C, HIV, excess alcohol, elevated

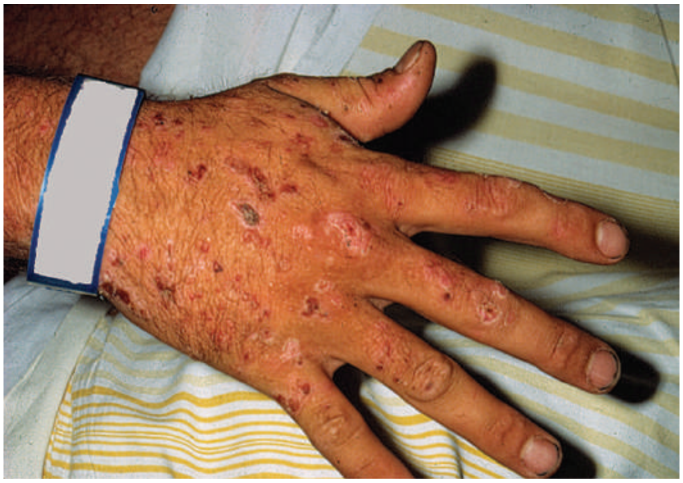

FIGURE 428-3 Typical cutaneous lesions in a patient with porphyria cutanea tarda. Chronic, crusted lesions resulting from blistering due to photosensitivity on the dorsum of the hand of a patient with porphyria cutanea tarda. *(Used with permission from Dr. Karl E. Anderson.)*

iron levels, and estrogens. The importance of excess hepatic iron as a precipitating factor is underscored by the finding that the incidence of the common hemochromatosis-causing mutations, hemochromatosis gene (*HFE*) mutations p.C282Y and p.H63D, are increased in patients with types 1 and 2 PCT **(Chap. 426)**. Excess alcohol is a long-recognized contributor, as is estrogen use in women. HIV is probably an independent but less common risk factor that, like hepatitis C, does not cause PCT in isolation. Multiple susceptibility factors that appear to act synergistically can be identified in individual patients. PCT patients characteristically have chronic liver disease and sometimes cirrhosis and are at risk for hepatocellular carcinoma. Various chemicals can also induce PCT; an epidemic of PCT occurred in eastern Turkey in the 1950s as a consequence of wheat contaminated with the fungicide hexachlorobenzene. PCT also occurs after exposure to other chemicals, including di- and trichlorophenols and 2,3,7,8-tetrachlorodibenzo-(*p*)-dioxin (TCDD, dioxin).

Diagnosis Porphyrins are increased in the liver, plasma, urine, and stool. The urinary ALA level may be slightly increased, but the PBG level is normal. Urinary porphyrins consist mostly of uroporphyrins and heptacarboxylate porphyrin, with lesser amounts of coproporphyrin and hexa- and pentacarboxylate porphyrins. Plasma porphyrins are also increased, and fluorometric scanning of diluted plasma at neutral pH can rapidly distinguish VP and PCT (Table 428-3). Isocoproporphyrins, which are increased in feces and sometimes in plasma and urine, are diagnostic for hepatic URO-decarboxylase deficiency.

Type 2 PCT and HEP can be distinguished from type 1 by finding decreased URO-decarboxylase in erythrocytes. URO-decarboxylase activity in liver, erythrocytes, and cultured skin fibroblasts in type 2 PCT is ~50% of normal in affected individuals and in asymptomatic heterozygous family members. In HEP, the URO-decarboxylase activity is markedly deficient, with typical levels of 3–10% of normal. Over 150 mutations have been identified in the *UROD* gene (Human Gene Mutation Database; *www.hgmd.org*). Of the mutations listed in the database, ~65% are missense or nonsense, and ~8% are splice-site mutations. Many *UROD* mutations have been identified in only one or two families.

TREATMENT

Porphyria Cutanea Tarda

Alcohol, estrogens, iron supplements, and, if possible, any drugs that may exacerbate the disease should be discontinued, but this step does not always lead to improvement. A complete response can almost always be achieved by the standard therapy, repeated phlebotomy, to reduce hepatic iron. A unit (450 mL) of blood can be removed every 1–2 weeks. The aim is to gradually reduce excess hepatic iron until the serum ferritin level reaches the lower limits of normal. Because iron overload is not marked in most cases, remission may occur after only five or six phlebotomies; however, PCT patients with hemochromatosis may require more treatments to bring their iron levels down to the normal range. To document improvement in PCT, it is most convenient to follow the total plasma porphyrin concentration, which becomes normal sometime after the target ferritin level is reached. Hemoglobin levels or hematocrits and serum ferritin should be followed closely to prevent development of iron deficiency and anemia. After remission, continued phlebotomy may not be needed. Plasma porphyrin levels are followed at 6- to 12-month intervals for early detection of recurrences, which are treated by additional phlebotomy.

An alternative when phlebotomy is contraindicated or poorly tolerated is a low-dose regimen of chloroquine or hydroxychloroquine, both of which complex with the excess porphyrins and promote their excretion. Small doses (e.g., 125 mg chloroquine phosphate twice weekly) should be given, because standard doses can induce transient, sometimes marked increases in photosensitivity and hepatocellular damage. Studies indicate that low-dose hydroxychloroquine is as safe and effective as phlebotomy in PCT. Hepatic imaging can diagnose or exclude complicating hepatocellular carcinoma. Treatment of PCT in patients with end-stage renal disease is facilitated by administration of erythropoietin.

Because hepatitis C virus (HCV) is a common precipitating factor causing PCT, the recent development of oral direct-acting antivirals for HCV has proven effective as a first primary treatment in HCV-infected PCT patients.

■ HEREDITARY COPROPORPHYRIA

HCP is an autosomal dominant hepatic porphyria that results from the half-normal activity of COPRO-oxidase. The disease presents with acute attacks, as in AIP. Cutaneous photosensitivity also may occur, but much less commonly than in VP. HCP patients may have acute attacks and cutaneous photosensitivity together or separately. HCP is less common than AIP and VP. Homozygous dominant HCP and harderoporphyria, a biochemically distinguishable variant of HCP, present with clinical symptoms in children (see below).

Clinical Features HCP is influenced by the same factors that cause attacks in AIP. The disease is latent before puberty, and symptoms, which are virtually identical to those of AIP, are more common in women. HCP is generally less severe than AIP. Blistering skin lesions are identical to PCT and VP and begin in childhood in rare homozygous cases.

Diagnosis COPRO III is markedly increased in the urine and feces in symptomatic patients and often persists, especially in feces, when there are no symptoms. Urinary ALA and PBG levels are increased (but less than in AIP) during acute attacks but may revert to normal more quickly than in AIP when symptoms resolve. Plasma porphyrins are usually normal or only slightly increased, but they may be higher in cases with skin lesions. The diagnosis of HCP is readily confirmed by increased fecal porphyrins consisting almost entirely of COPRO III, which distinguishes it from other porphyrias.

Although the diagnosis can be confirmed by measuring COPRO-oxidase activity, the assays for this mitochondrial enzyme are not available and require cells other than erythrocytes. To date, >95 mutations have been identified in the *CPOX* gene, ~70% of which are missense or nonsense (Human Gene Mutation Database; *www.hgmd.org*). Detection of a *CPOX* mutation in a symptomatic individual permits the identification of asymptomatic family members.

TREATMENT

Hereditary Coproporphyria

Neurologic symptoms are treated as in AIP (see above). Phlebotomy and chloroquine are not effective for the cutaneous lesions.

■ VARIEGATE PORPHYRIA

VP is an autosomal dominant hepatic porphyria that results from the deficient activity of PROTO-oxidase, the seventh enzyme in the heme biosynthetic pathway, and can present with neurologic symptoms, photosensitivity, or both. VP is particularly common in South Africa, where 3 of every 1000 whites have the disorder. Most are descendants of a couple who emigrated from the Netherlands to South Africa in 1688. In other countries, VP is less common than AIP. Rare cases of homozygous dominant VP, presenting in childhood with cutaneous symptoms, also have been reported.

Clinical Features VP can present with skin photosensitivity, acute neurovisceral crises, or both. In two large studies of VP patients, ~60% had only skin lesions, 20% had only acute attacks, and ~20% had both. Acute attacks are identical to those in AIP and are precipitated by the same factors as AIP (see above). Blistering skin manifestations are similar to those in PCT but are more difficult to treat and usually are of longer duration. Homozygous VP is associated with photosensitivity, neurologic symptoms, and developmental disturbances, including growth retardation, in infancy or childhood; all cases had increased erythrocyte levels of zinc protoporphyrin, a characteristic finding in all homozygous porphyrias so far described.

Diagnosis Urinary ALA and PBG levels are increased during acute attacks but may return to normal more quickly than in AIP. Increases in fecal protoporphyrin and COPRO III and in urinary COPRO III are more persistent. Plasma porphyrin levels also are increased, particularly when there are cutaneous lesions. VP can be distinguished rapidly from all other porphyrias by examining the fluorescence emission spectrum of porphyrins in plasma since VP has a unique fluorescence peak at neutral pH.

Assays of PROTO-oxidase activity in cultured fibroblasts or lymphocytes are not widely available. Over 215 mutations have been identified in the *PPOX* gene from unrelated VP patients (Human Gene Mutation Database; *www.hgmd.org*). The missense mutation encoding p.R59W is the common mutation in most South Africans with VP of Dutch descent. Five missense mutations were common in English and French VP patients; however, most mutations have been found in only one or a few families.

TREATMENT

Variegate Porphyria

Acute attacks are treated as in AIP, and hemin should be started early in most cases. Givosiran has proven effective in clinical trials for patients with recurrent attacks. Other than avoiding sun exposure, there are few effective measures for treating the skin lesions. β-Carotene, phlebotomy, and chloroquine are not helpful.

THE ERYTHROPOIETIC PORPHYRIAS

In the erythropoietic porphyrias, excess porphyrins from bone marrow erythrocyte precursors are transported via the plasma to the skin and lead to cutaneous photosensitivity.

X-LINKED SIDEROBLASTIC ANEMIA

XLSA results from the deficient activity of the erythroid form of ALA-synthase (ALA-synthase 2) and is associated with ineffective erythropoiesis, weakness, and pallor.

Clinical Features Typically, males with XLSA develop refractory hemolytic anemia, pallor, and weakness during infancy. They have secondary hypersplenism, become iron overloaded, and can develop hemosiderosis. The severity depends on the level of residual erythroid ALA-synthase activity and on the responsiveness of the specific mutation to pyridoxal 5′-phosphate supplementation (see below). Peripheral blood smears reveal a hypochromic, microcytic anemia with striking anisocytosis, poikilocytosis, and polychromasia; the leukocytes and platelets appear normal. Hemoglobin content is reduced, and the mean corpuscular volume and mean corpuscular hemoglobin concentration are decreased. Patients with milder, later-onset disease have been reported recently.

Diagnosis Bone marrow examination reveals hypercellularity with a left shift and megaloblastic erythropoiesis with an abnormal maturation. A variety of Prussian blue–staining sideroblasts are observed. Levels of urinary porphyrin precursors and of both urinary and fecal porphyrins are normal. The activity of erythroid ALA-synthase 2 is decreased in bone marrow, but this enzyme is difficult to measure in the presence of the normal ALA-synthase 1 housekeeping enzyme. Definitive diagnosis requires the demonstration of loss-of-function mutations in the erythroid *ALAS2* gene, of which >120 have been identified.

Treatment The severe anemia may respond to pyridoxine supplementation. This cofactor is essential for ALA-synthase activity, and mutations in the pyridoxine binding site of the enzyme have been found in several responsive patients. Cofactor supplementation may make it possible to eliminate or reduce the frequency of transfusions. Unresponsive patients may be transfusion dependent and require chelation therapy.

CONGENITAL ERYTHROPOIETIC PORPHYRIA

CEP, also known as Günther's disease, is an autosomal recessive disorder. It is due to the markedly deficient, but not absent, activity of URO-synthase and the resultant accumulation of URO I and COPRO I isomers. CEP is associated with hemolytic anemia and cutaneous lesions.

Clinical Features Severe cutaneous photosensitivity typically begins from birth. The skin over light-exposed areas is friable, and bullae and vesicles are prone to rupture and infection. Skin thickening, focal hypo- and hyperpigmentation, and hypertrichosis of the face and extremities are characteristic. Secondary infection of the cutaneous lesions can lead to disfigurement of the face and hands. Porphyrins are deposited in teeth and in bones. As a result, the teeth are brownish and fluoresce on exposure to long-wave ultraviolet light. Hemolysis is due to the marked increase in erythrocyte porphyrins and leads to splenomegaly. Adults with a milder later-onset form of the disease also have been described, including late-onset patients with myelodysplasias.

Diagnosis URO and COPRO (mostly type I isomers) accumulate in the bone marrow, erythrocytes, plasma, urine, and feces. The predominant porphyrin in feces is COPRO I. The diagnosis of CEP can be confirmed by demonstration of markedly deficient URO-synthase activity and/or by the identification of specific mutations in the *UROS* gene. The disease can be detected in utero by measuring porphyrins in amniotic fluid and URO-synthase activity in cultured amniotic cells or chorionic villi or by the detection of the family's specific gene mutations. Molecular analyses of the mutant alleles from unrelated patients have revealed the presence of >65 mutations in the *UROS* gene, including six in the erythroid-specific promoter of the *UROS* gene. Genotype/phenotype correlations can predict the severity of the disease. The CEP phenotype may be increased by the presence of exon 10 variants in the erythroid-specific ALA-synthase 2, mutations that typically cause XLP. One mutation (p.Arg216Trp) in *GATA1*, encoding the X-linked erythroid-specific transcription factor GATA binding protein 1 (GATA1), has been identified in two individuals with CEP who both also had other hematologic abnormalities.

TREATMENT

Congenital Erythropoietic Porphyria

Transfusion-dependent patients require periodic transfusions for anemia. Chronic transfusions of sufficient fresh packed erythrocytes to suppress erythropoiesis are effective in reducing porphyrin production but result in iron overload. Oral iron chelation is recommended. Splenectomy may reduce hemolysis and decrease transfusion requirements. Protection from sunlight and from minor skin trauma is essential to avoid/minimize cutaneous blistering. Complicating bacterial infections should be treated promptly. Recently, non-transfusion-dependent patients have been treated by periodic phlebotomies to decrease iron levels, thereby decreasing erythropoiesis and porphyrin accumulation. This approach has not been evaluated in clinical trials to date. Bone marrow and hematopoietic stem cell transplantation has proven curative in transfusion-dependent children, providing the rationale for future stem cell gene therapy.

ERYTHROPOIETIC PROTOPORPHYRIA

EPP is an autosomal recessive disorder resulting from the deficient activity of FECH, the last enzyme in the heme biosynthetic pathway. As noted above, EPP is likely the most common porphyria with onset typically in early childhood. EPP patients have FECH activities as low as 15–30% of normal in lymphocytes and cultured fibroblasts. Protoporphyrin IX accumulated in bone marrow reticulocytes and circulating erythrocytes is released into the plasma and then is taken up in the liver where it is excreted in the bile and feces. Plasma protoporphyrin IX taken up by the vascular cells in the skin is photoactivated on exposure to sunlight causing phototoxic cellular damage and excruciatingly painful nonblistering phototoxicity. In most symptomatic patients (>95%) with this disorder, a deleterious mutation in one *FECH* allele was inherited with the relatively common (~10% of Caucasians) intronic 3 (IVS3) variant (IVS3–48T>C) on the other allele; together, they result in the low expression of the normal enzyme. In ~2% of EPP families, two *FECH* deleterious mutations have been found.

XLP is a less common condition with the same phenotype in affected males, including increased erythrocyte protoporphyrin IX levels resulting from gain-of-function mutations in the last exon of the erythroid-specific form of 5-aminolevulinate-synthase 2 (*ALAS2*). These mutations delete or alter the ALAS2 C-terminal amino acids, resulting in its increased activity and the subsequent accumulation of protoporphyrin IX. Manifestations in female heterozygotes with XLP can range from asymptomatic to as severe as their affected male relatives. The variation in the presence and severity of manifestations in XLP heterozygotes results primarily from random X-chromosomal inactivation. XLP accounts for ~2–10% of cases with the EPP phenotype in Europe and North America. Rare patients with EPP symptoms and elevated erythrocyte protoporphyrin IX levels do not have mutations in *FECH* or *ALAS2* on genetic testing. In an affected family with EPP symptoms and accumulation of protoporphyrin IX, an autosomal dominant mutation was found in human CLPX, a modulator of heme biosynthesis.

Clinical Features In EPP and male XLP patients, skin photosensitivity, which differs from that in other cutaneous porphyrias, usually begins in early childhood. The initial symptoms on sun exposure consist of tingling, stinging, itching, or heat/burning sensations on the exposed skin occurring within <10 to 30 min of exposure in >60% of patients; most will have these prodromal symptoms within an hour of sun exposure. The prodromal symptoms are the "warning signal" to get out of the sun, thereby avoiding a severe incapacitating painful attack that can last from 2–5 days. Photosensitivity is associated with substantial elevations in erythrocyte protoporphyrin IX and occurs only in patients with genotypes that result in FECH activities below ~35% of normal. Vesicular lesions are uncommon. Redness and swelling develop after prolonged sun exposure and resemble angioedema **(Fig. 428-4)**. Pain symptoms may seem out of proportion to the visible skin involvement. Chronic skin changes may include lichenification, leathery pseudovesicles, labial grooving, and nail changes. Severe scarring is rare, as are pigment changes, friability, and hirsutism. Unless hepatic or other complications develop, protoporphyrin IX levels and symptoms of photosensitivity tend to remain remarkably stable over many years in most patients. Factors that exacerbate the hepatic porphyrias play no role in EPP or XLP.

The primary source of excess protoporphyrin is the bone marrow erythroid cells. In EPP patients, erythrocyte protoporphyrin IX is free (not complexed with zinc) and is mostly bound to hemoglobin. In plasma, protoporphyrin IX is bound to albumin. Hemolysis and anemia are absent or usually mild.

Although EPP is an erythropoietic porphyria, up to 27% of EPP patients may have minor abnormalities of liver function, and in ~2–5% of these patients, the accumulation of protoporphyrins causes chronic liver disease that can progress to liver failure requiring transplantation. Protoporphyrin IX is insoluble, and excess amounts form crystalline structures in liver cells (Fig. 428-4) and can decrease hepatic bile flow. Studies in the mouse model of EPP have shown that the bile duct epithelium may be damaged by toxic bile, leading to biliary fibrosis. Thus, rapidly progressive liver disease appears to be related to the cholestatic effects of protoporphyrins and is associated with increasing hepatic protoporphyrin IX levels due to impaired hepatobiliary excretion and increased photosensitivity. The hepatic complications also are often characterized by increasing levels of protoporphyrins in erythrocytes and plasma as well as severe abdominal and back pains, especially in the right upper quadrant. Gallstones composed at least in part of protoporphyrin IX occur in some patients. Hepatic complications appear to be higher in EPP due to two pathogenic *FECH* mutations and in males with XLP.

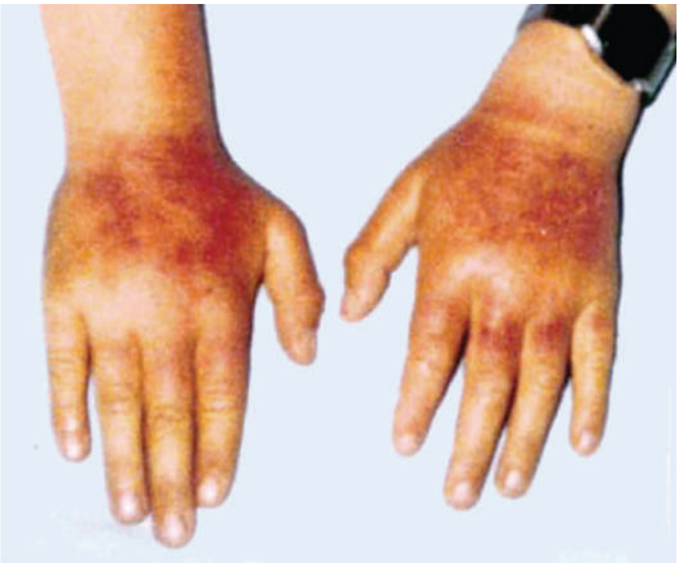

FIGURE 428-4 Erythema and edema of the hands due to acute photosensitivity in a 10-year-old boy with erythropoietic protoporphyria. *(Reproduced with permission from P Poblete-Gutiérrez et al: The porphyrias: clinical presentation, diagnosis and treatment. Eur J Dermatol 16:230, 2006.)*

Diagnosis A substantial increase in erythrocyte protoporphyrin IX, which is predominantly free and not complexed with zinc, is the hallmark of EPP. Protoporphyrin levels also are variably increased in bone marrow, plasma, bile, and feces. Erythrocyte protoporphyrin IX concentrations are increased in other conditions such as lead poisoning, iron deficiency, various hemolytic disorders, all homozygous forms of other porphyrias, and sometimes even in acute porphyrias. In all these conditions, however, in contrast to EPP, protoporphyrin IX is complexed with zinc. Therefore, after an increase in erythrocyte protoporphyrin IX is found in a suspected EPP patient, it is important to confirm the diagnosis by an assay that distinguishes free and zinc-complexed protoporphyrin. Erythrocytes in EPP also exhibit red fluorescence under fluorescence microscopy at 620 nm. Urinary levels of porphyrins and porphyrin precursors are normal. FECH activity in cultured lymphocytes or fibroblasts is decreased (<30% of normal mean). DNA diagnosis by mutation analysis is recommended to detect the causative *FECH* mutation(s) and/or the presence of the IVS3–48T>C low expression allele. To date, >235 mutations have been identified in the *FECH* gene, many of which result in an unstable or absent enzyme protein (null alleles) (Human Gene Mutation Database; *www.hgmd.org*).

In XLP, the erythrocyte protoporphyrin levels appear to be higher than in EPP, and the proportions of free and zinc protoporphyrin IX may reach 50%. XLP accounts for ~2% of patients with the EPP phenotype in Western Europe. Recent studies show that ~10% of North American patients with the EPP phenotype have XLP.

TREATMENT

Erythropoietic Protoporphyria

Avoiding sunlight exposure and wearing clothing designed to provide protection for conditions with chronic phototoxicity are essential. Various other treatments, including oral β-carotene and cimetidine, have proven of little benefit. Afamelanotide, an α-melanocyte-stimulating hormone (MSH) analogue that stimulates tanning, is administered subcutaneously every 2 months and has been approved for the treatment of EPP and XLP by the EMA and the FDA. Dersimelagon, an orally administered, small-molecule, selective melanocortin-1 receptor (MC1R) agonist that increases skin melanin without sun exposure, is currently in phase 3 clinical trials for EPP and XLP. Bitopertin, an orally administered glycine reuptake inhibitor, recently was shown to reduce blood protoporphyrin IX levels in clinical studies and is currently in phase 3 clinical trials.

Treatment of hepatic complications, which may be accompanied by motor neuropathy, is difficult. Cholestyramine and other porphyrin absorbents such as activated charcoal may interrupt the enterohepatic circulation of protoporphyrin and promote its fecal excretion, leading to some improvement. Plasmapheresis and intravenous hemin are sometimes beneficial.

Liver transplantation has been carried out in some EPP and XLP patients with severe liver complications and is often successful in the short term. However, the disease often recurs in the transplanted liver due to continued bone marrow production of excess protoporphyrin. In a retrospective study of 17 liver-transplanted EPP patients, 11 (65%) had recurrent EPP liver disease. Posttransplantation treatment with hematin and plasmapheresis should be considered to prevent the recurrence of liver disease. However, bone marrow transplantation, which has been successful in human EPP and which prevented liver disease in a mouse model, should be considered after liver transplantation, if a suitable donor can be found.

Acknowledgment

The authors thank Dr. Karl E. Anderson for his review of the manuscript and helpful comments and suggestions. This work is supported in part by the Porphyrias Consortium (U54 DK083909), a part of the National Institutes of Health (NIH) Rare Disease Clinical Research Network (RDCRN), supported through collaboration between the NIH Office of Rare Diseases Research (ORDR) at the National Center for Advancing Translational Science (NCATS) and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

FURTHER READING

Balwani M et al: Clinical, biochemical, and genetic characterization of North American patients with erythropoietic protoporphyria and X-linked protoporphyria. JAMA Dermatol 153:789, 2017.

Balwani M et al: Phase 3 trial of RNAi therapeutic givosiran for acute intermittent porphyria. N Engl J Med 382:2289, 2020.

Bonkovsky HL et al: Ledipasvir/sofosbuvir is effective as sole treatment of porphyria cuteanea tarda with chronic hepatitis C. Dig Dis Sci 68:2738,2023.

Chen B et al: Acute intermittent porphyria: Predicted pathogenicity of HMBS variants indicates extremely low penetrance of the autosomal dominant disease. Hum Mutat 37:1215, 2016.

Kazamel M et al: Pain in acute hepatic porphyrias: Update on pathophysiology and management. Front Neurol 13:1004125, 2022.

Levy C et al: Evidence-based consensus guidelines for the diagnosis and management of protoporphyria-related liver dysfunction in erythropoietic protoporphyria and X-linked protoporphyria. Hepatology 79:731, 2024.

Tchernitchko D et al: A variant of peptide transporter 2 predicts the severity of porphyria-associated kidney disease. J Am Soc Nephrol 28:1924, 2017.

Wang B et al: AGA clinical practice update on diagnosis and management of acute hepatic porphyrias: Expert review. Gastroenterology 164:481, 2023.

Yasuda M et al: Liver transplantation for acute intermittent porphyria: Biochemical and pathologic studies of the explanted liver. Mol Med 21:487, 2015.

429

Lysosomal Storage Diseases

Robert J. Hopkin, Gregory A. Grabowski

Lysosomes are heterogeneous subcellular organelles containing specific hydrolases that allow selective processing or degradation of proteins, nucleic acids, carbohydrates, and lipids. There are >50 different lysosomal storage diseases, now termed lysosomal system diseases (LSDs), classified according to the nature of the major accumulated materials (Table 429-1). Although all are rare diseases, an overview of several are provided here: Tay-Sachs disease, Fabry disease, Gaucher disease, Niemann-Pick disease, the mucopolysaccharidoses, Pompe disease, lysosomal acid lipase deficiency (LALD), Krabbe disease, CLN2-related Batten disease, and α-mannosidosis. LSDs should be considered in the differential diagnosis of patients with neurologic, renal, or muscular degeneration and/or unexplained hepatomegaly, splenomegaly, cardiomyopathy, or skeletal dysplasias and deformations. Physical findings are disease specific, and enzyme assays or genetic testing can be used to make a definitive diagnosis. Although the nosology of LSDs segregates the variants into distinct phenotypes, these are heuristic; in the clinic, each disease exhibits—to varying degrees—a spectrum of manifestations, from severe to attenuated variants.

PATHOGENESIS

Lysosomal biogenesis involves ongoing synthesis of lysosomal hydrolases, membrane constitutive proteins, and new membranes. Lysosomes originate from the fusion of trans-Golgi network vesicles with late endosomes. Progressive acidification accompanies the maturation of these vesicles; this gradient facilitates the pH-dependent dissociation of receptors and ligands and also activates lysosomal hydrolases. Lysosomes are components of the lysosome/autophagy/mitophagy system that are regulated by the mTORC1 modulation, via phosphorylation, of the transcription factors TFEB/TFE3 and the resultant control of the balance between anabolic and catabolic pathways. This regulation is disrupted to varying degrees in specific tissues affected by individual LSDs.

Abnormalities at any biosynthetic step can impair lysosomal enzyme function and lead to an LSD. After leader sequence clipping from the primary protein, remodeling of complex oligosaccharides (including the lysosomal targeting ligand mannose-6-phosphate as well as high-mannose oligosaccharide chains of many soluble lysosomal hydrolases) occurs during transit through the Golgi. Lysosomal integral or associated membrane proteins are sorted to the membrane or interior of the lysosome by several different peptide signals. Phosphorylation, sulfation, additional proteolytic processing, and macromolecular assembly of heteromers occur concurrently. Such posttranslational modifications are critical to enzyme function, and defects in these processes can result in multiple enzyme/protein deficiencies.

The final common pathway for LSDs is the accumulation of specific macromolecules within selected tissues and cells that normally have a high flux of these substrates. The majority of lysosomal enzyme deficiencies result from point mutations or genetic rearrangements at a locus that encodes a single lysosomal hydrolase or protein subunit. However, some mutations cause deficiencies of several different lysosomal hydrolases by alteration of the enzymes/proteins involved in targeting, active site modifications, macromolecular association, or trafficking. Nearly all LSDs are inherited as autosomal recessive disorders, except for Hunter (mucopolysaccharidosis type II), Danon, and Fabry diseases that are X-linked, and two autosomal dominant conditions causing Parry type neuronal ceroid lipofuscinosis (CLN) due to mutations in *DNAJC5* or frontotemporal dementia and CLN11 due to *GRN* (progranulin) mutations. Substrate accumulations lead to lysosomal distortions/dysfunctions that have significant pathophysiologic

TABLE 429-1 Selected Lysosomal System Diseases (also known as Lysosomal Storage Diseases)

					CLINICAL FEATURES					
DISORDER[a]	ENZYME DEFICIENCY (*GENE*) [SPECIFIC THERAPY]	PRIMARY ACCUMULATED MATERIALS	CLINICAL TYPES (ONSET)	INHERITANCE	NEUROLOGIC	LIVER, SPLEEN ENLARGEMENT	SKELETAL DYSPLASIA	OPHTHALMOLOGIC	HEMATOLOGIC	UNIQUE FEATURES
Mucopolysaccharidoses (MPS)										
MPS I H, Hurler	α-L-Iduronidase (*IDUA*) [ET, HSCT]	Dermatan sulfate Heparan sulfate	Infantile Intermediate	AR	Cognitive degeneration	+++	++++	Corneal clouding	Vacuolated lymphocytes	Coarse facies; cardiovascular involvement; joint stiffness
MPS I H/S, Hurler/Scheie			Childhood/adult		Cognitive degeneration					
MPS I S, Scheie					None					
MPS II, Hunter	Iduronate sulfatase (IDS) [ET]	Dermatan sulfate Heparan sulfate	Severe infantile Mild juvenile	X-linked	Cognitive degeneration, less in mild form	+++	++++	Retinal degeneration, no corneal clouding	Granulated lymphocytes	Coarse facies; cardiovascular involvement; joint stiffness; distinctive pebbly skin lesions
MPS III A, Sanfilippo A	Heparan-*N*-sulfatase (*SGSH*)	Heparan sulfate	Late infantile	AR	Severe cognitive degeneration	+	+	None	Granulated lymphocytes	Mild coarse facies
MPS III B, Sanfilippo B	*N*-Acetyl-α-glucosaminidase (*NGALU*)	Heparan sulfate	Late infantile	AR	Severe cognitive degeneration	+	+	None	Granulated lymphocytes	Mild coarse facies
MPS III C, Sanfilippo C	Acetyl-CoA: α-glucosaminide *N*-acetyltransferase (*HGSNAT*)	Heparan sulfate	Late infantile	AR	Severe cognitive degeneration	+	+	None	Granulated lymphocytes	Mild coarse facies
MPS III D, Sanfilippo D	*N*-Acetylglucosamine-6-sulfate sulfatase (*GNS*)	Heparan sulfate	Late infantile	AR	Severe cognitive degeneration	+	+	None	Granulated lymphocytes	Mild coarse facies
MPS IV A, Morquio A	*N*-Acetylgalactosamine-6-sulfate sulfatase (*GALNS*) [ET]	Keratan sulfate Chondroitin-6 sulfate	Childhood	AR	None	+	++++	Corneal clouding	Granulated neutrophils	Distinctive skeletal deformity; odontoid hypoplasia; aortic valve disease
MPS IV B, Morquio	β-Galactosidase (*GLB1*)		Childhood	AR	None	±	++++			
MPS VI, Maroteaux-Lamy	Arylsulfatase B (*ARSB*) [ET, HSCT]	Dermatan sulfate	Late infantile	AR	None	++	++++	Corneal clouding	Granulated neutrophils and lymphocytes	Coarse facies; valvular heart disease
MPS VII Sly	β-Glucuronidase (*GUS*) [ET]	Dermatan sulfate Heparan sulfate	Neonatal Infantile Adult	AR	Cognitive degeneration, absent in some adults	+++	+++	Corneal clouding	Granulated neutrophils	Coarse facies; vascular involvement; hydrops fetalis in neonatal form
GM_2 Gangliosidoses										
Tay-Sachs disease	β-Hexosaminidase A (*HEXA*)	GM_2 gangliosides	Infantile Juvenile	AR	Cognitive degeneration; seizures; later juvenile form	None	None	Cherry red spot in infantile form	None	Macrocephaly; hyperacusis in infantile form
Sandhoff disease	β-Hexosaminidases A and B (*HEXB*)	GM_2 gangliosides	Infantile	AR	Cognitive degeneration; seizures	++	±	Cherry red spot	None	Macrocephaly; hyperacusis

Neutral Glycosphingolipidoses										
Fabry disease	α-Galactosidase A (*GLA*) [ET, Chaperone]	Globotriaosylceramide	Childhood	X-linked	Painful acroparesthesias	None	None	Corneal dystrophy, vascular lesions	None	Cutaneous angiokeratomas; hypohydrosis
Gaucher disease	Acid β-glucosidase (*GBA*, a.k.a. *GBA1*) [ET, SRT]	Glucosylceramide, Glucosylsphingosine	Type 1 Type 2 Type 3	AR	None ++++ –/+++	++++ +++ ++++	++++ + ++++	None Strabismus Supranuclear gaze palsy	Gaucher cells in bone marrow; cytopenias	Types 1 and 3 highly variable
Niemann-Pick disease A and B	Acid sphingomyelinase (*SMPD1*) [ET]	Sphingomyelin	Neuronopathic, type A Nonneuronopathic, type B	AR	Cognitive degeneration; seizures	++++	None Osteoporosis	Macular degeneration	Foam cells in bone marrow	Pulmonary infiltrates Lung failure
Glycoproteinoses										
Fucosidosis	α-Fucosidase (*FUCA1*)	Glycopeptides; oligosaccharides	Infantile Juvenile	AR	Cognitive degeneration	++	++	None	Vacuolated lymphocytes; foam cells	Coarse facies; angiokeratomas in juvenile form
α-Mannosidosis	α-Mannosidase (*MAN2B1*) [ET]	Oligosaccharides	Infantile Milder variant	AR	Cognitive degeneration	+++	+++	Cataracts, corneal clouding	Vacuolated lymphocytes, granulated neutrophils	Coarse facies; enlarged tongue
β-Mannosidosis	β-Mannosidase (*MANBA*)	Oligosaccharides		AR	Seizures; cognitive degeneration		++	None	Vacuolated lymphocytes, foam cells	Angiokeratomas
Aspartylglucosaminuria	Aspartylglucosaminidase (*AGA*)	Aspartylglucosamine; glycopeptides	Young adult	AR	Cognitive degeneration	±	++	None	Vacuolated lymphocytes, foam cells	Coarse facies
Sialidosis	Neuraminidase (*NEU1*)	Sialyloligosaccharides	Type I, congenital Type II, infantile and juvenile	AR	Myoclonus; cognitive degeneration	++, less in type I	++, less in type I	Cherry red spot	Vacuolated lymphocytes	MPS phenotype in type II
Mucolipidoses (ML)										
ML-II, I-cell disease	UDP-*N*-Acetylglucosamine-1-phosphotransferase (*GNPTAB*)	Glycoprotein; glycolipids	Infantile	AR	Cognitive degeneration	+	++++	Corneal clouding	Vacuolated and granulated neutrophils	Coarse facies; absence of mucopolysacchariduria; gingival hypoplasia
ML-III, pseudo-Hurler polydystrophy	UDP-*N*-Acetylglucosamine-1-phosphotransferase (*GNPTAB*)	Glycoprotein; glycolipids	Late infantile	AR	Mild cognitive degeneration	None	+++	Corneal clouding, mild retinopathy, hyperopic astigmatism		Coarse facies; stiffness of hands and shoulders
Leukodystrophies										
Krabbe disease	Galactosylceramidase (*GALC*) [BMT/HSCT]	Galactosylceramide Galactosylsphingosine, a.k.a., psychosine	Infantile	AR	Cognitive degeneration, CNS, PNS dysmyelination	None	None	None	None	White matter globoid cells
Metachromatic leukodystrophy	Arylsulfatase A (*ARSA*)	Cerebroside sulfate	Infantile Juvenile Adult	AR	Cognitive degeneration; dementia; psychosis in adult	None	None	Optic atrophy	None	Gait abnormalities in late infantile form

(*Continued*)

TABLE 429-1 Selected Lysosomal System Diseases (also known as Lysosomal Storage Diseases) (Continued)

					CLINICAL FEATURES					
DISORDER[a]	ENZYME DEFICIENCY (GENE) [SPECIFIC THERAPY]	PRIMARY ACCUMULATED MATERIALS	CLINICAL TYPES (ONSET)	INHERITANCE	NEUROLOGIC	LIVER, SPLEEN ENLARGEMENT	SKELETAL DYSPLASIA	OPHTHALMOLOGIC	HEMATOLOGIC	UNIQUE FEATURES
Multiple sulfatase deficiency	Formylglycine-generating enzyme, a.k.a. FGE (*SUMF1*)	Sulfatides; mucopolysaccharides	Late infantile	AR	Cognitive degeneration	+	++	Retinal degeneration	Vacuolated and granulated cells	Absent activity of all known cellular sulfatases
Disorders of Neutral Lipids										
Infantile-onset LALD	Lysosomal acid lipase (*LIPA*) [ET, HSCT]	Cholesteryl esters; triglycerides	Infantile	AR	None	+++	None	None	None	Adrenal cortical calcification
Childhood/adult-onset LALD	Acid lysosomal lipase (*LIPA*) [ET]	Cholesteryl esters	Childhood	AR	None	Hepatomegaly	None	None	None	Fatty liver disease; cirrhosis
Farber disease	Acid ceramidase (*ASAH1*)	Ceramide	Infantile Juvenile	AR	Occasional cognitive degeneration	±	None	Macular degeneration	None	Arthropathy, subcutaneous nodules
Niemann-Pick disease C1 and C2	NPC1 (*NPC1*) NPC2 (*NPC2*)	Cholesterol	Childhood to adulthood	AR	Progressive CNS degeneration	Variable hepatosplenomegaly	None	Vertical supranuclear ophthalmoplegia	None	Rare fatal neonatal liver disease, cirrhosis
Disorders of Glycogen										
Pompe disease	Acid α-glucosidase (*GAA*) [ET]	Glycogen	Infantile, late onset	AR	Neuromuscular	±	None	None	None	Myocardiopathy
Late-onset GAA deficiency	Acid α-glucosidase (*GAA*) [ET]	Glycogen	Variable: juvenile to adulthood	AR	Neuromuscular	None	None	None	None	Respiratory insufficiency; neuromuscular disease
Danon disease	LAMP-2 (lysosomal associated membrane protein-2) (*LAMP2*)	Glycogen	Variable: childhood to adulthood	X-linked (?Dominant)	Cardiomyopathy Neuromuscular Inconsistent cognitive degeneration	None	None	None	None	Myocardial vacuolar degeneration
Neuronal Ceroid Lipofuscinoses										
CLN2 (a.k.a. NCL2)	TPP1 (tripeptidyl peptidase 1) (*TPP1*) [ICV ET]	Ceroid lipofuscin	Early childhood	AR	Neurodegenerative Loss of motor skills Myoclonus Loss of vision Cognitive loss Wheelchair bound by adolescence	None	None	Progressive vision loss	None	Symmetric retinal progressive degeneration by 4–6 years

[a]Comprehensive reviews of these lysosomal storage diseases can be found in DL Valle et al: *The Online Metabolic and Molecular Bases of Inherited Disease*, New York, McGraw-Hill, https://ommbid.mhmedical.com/book.*aspx?bookID=2709#225069419*.

Abbreviations: AR, autosomal recessive; BMT/HSCT, bone marrow or hematopoietic stem cell transplantation; ET, enzyme therapy; ICV ET, intracerebroventricular enzyme therapy; LALD, lysosomal acid lipase deficiency; SRT, substrate synthesis inhibition therapy (a.k.a., substrate reduction therapy).

consequences. In addition, abnormal amounts of metabolites may also have pharmacologic effects important to disease pathophysiology and propagation, particularly activation of the innate immune responses.

For many LSDs, the accumulated substrates are synthesized within particular tissue sites of pathology. Other diseases have greater exogenous substrate supplies. For example, substrates are delivered by low-density lipoprotein receptor–mediated uptake in Fabry and LALD or by phagocytosis in Gaucher disease type 1. The threshold hypothesis refers to a level of enzyme activity below which disease develops. Small changes in enzyme activity near that threshold can lead to or modify disease. A critical element of this model is that enzymatic activity can be challenged by changes in substrate flux based on genetic background, cell turnover, recycling, or metabolic demands. Thus, a set level of residual enzyme may be adequate for substrate in some tissues or cells but not in others. In addition, several variants of each LSD exist at the clinical level. These disorders therefore represent a spectrum of manifestations that are not easily dissociated into discrete entities. The molecular/genetic bases for such variations have not been elucidated in any detail.

Treatments approved by the European Medicines Agency (EMA) and U.S. Food and Drug Administration (FDA) are available for several LSDs. The first was enzyme replacement therapy (ET) for Gaucher disease; this has been followed by additional ETs, but subsequent developments have included modified enzyme infusion, substrate synthesis inhibition, hematopoietic stem cell transplant (HSCT), pharmacologic chaperone therapy (which use small molecules to stabilize or enhance enzyme function resulting from the mutated gene), intrathecal enzyme delivery, Trojan-horse strategies for blood-brain barrier penetration for central nervous system (CNS) delivery, and gene therapy. The technical ability to intervene for most LSDs now exists but with highly variable impact, i.e., therapeutic outcome, particularly in bones and the CNS. Significant additional research is needed to reach the goals of long-term survival with good function and quality of life.

SELECTED DISORDERS

TAY-SACHS DISEASE

About 1 in 30 Ashkenazi Jews is a carrier for Tay-Sachs disease (total hexosaminidase A [Hex A] deficiency), resulting from α-chain gene mutations. The infantile form is a neurodegenerative disease that results in death in infancy. It is characterized by macrocephaly, loss of motor skills, increased startle reaction, and a macular cherry red spot. The juvenile-onset form presents as ataxia and dementia, with death by age 10–15 years. The adult-onset disorder is characterized by clumsiness in childhood; progressive motor weakness in adolescence; and additional spinocerebellar and lower-motor-neuron signs and dysarthria in adulthood; intelligence declines slowly, and psychiatric disorders are common. Screening for Tay-Sachs disease carriers is recommended in the Ashkenazi Jewish population. Sandhoff disease, due to a deficiency in both Hex A and Hex B resulting from defective β-chains, is phenotypically similar to Tay-Sachs disease with the addition of hepatosplenomegaly and bony dysplasias.

FABRY DISEASE

Fabry disease, an X-linked disorder and likely the most prevalent LSD, results from mutations in *GLA*, which encodes α-galactosidase A. The estimated prevalence of hemizygous males ranges from 1 in 40,000 to 1 in 3500 in selected populations. Females have a higher prevalence of mutations, but more variable manifestations due to Lyonization effects. The disease manifests with angiokeratomas (telangiectatic skin lesions), hypohidrosis, corneal and lenticular opacities, acroparesthesia, and progressive disease of the kidney, heart, and brain vascular systems. Abdominal pain, recurrent diarrhea, and acroparesthesias (debilitating episodic burning pain of the hands and feet) may appear in childhood. In females, the overall manifestations vary greatly. Angiokeratomas are punctate, dark red to blue-black, and flat or slightly raised skin lesions; they do not blanch with pressure. They can be easily overlooked. They usually are most dense between the umbilicus and the knees—the "bathing suit area"—but may occur anywhere. Angiokeratomas also occur in several other very rare LSDs. Corneal and lenticular lesions, detectable on slit-lamp examination, may help in establishing a diagnosis of Fabry disease. Acroparesthesia can last from minutes to days and can be precipitated by changes in temperature, exercise, fatigue, or fever. Abdominal pain can resemble appendicitis or renal colic. Proteinuria, isosthenuria, and progressive renal dysfunction occur in the second to fourth decades; ~5% of male patients with "idiopathic renal failure" have *GLA* mutations. Hypertension, left ventricular hypertrophy, anginal chest pain, and congestive heart failure can occur in the third to fourth decades. About 1–3% of patients with "idiopathic hypertrophic myocardiopathy" have Fabry disease. Similarly, ~2–5% of patients with "idiopathic stroke" at 35–50 years of age have *GLA* mutations. Leg lymphedema occurs without hypoproteinemia. Death is due to cardiovascular, renal, or cerebrovascular disease. Variants with residual α-galactosidase A activity may have late-onset manifestations that are most prominent in the cardiovascular system and resemble hypertrophic cardiomyopathy. Cases with predominant cardiac, renal, or CNS manifestations have been reported. Up to 70% of heterozygous females exhibit clinical manifestations. However, in females, heart disease is the most common life-threatening manifestation. In males, it is renal disease followed by cardiovascular disease.

Gabapentin and carbamazepine diminish chronic and episodic acroparesthesias. Chronic hemodialysis or kidney transplantation can be lifesaving in patients with renal failure. Intravenous ET clears stored lipids from a variety of cells. More recently a chaperone therapy (migalastat) that stabilizes the residual enzyme made by the patient's body has allowed oral therapy for some patients with amenable mutations. Renal insufficiency, cardiac fibrosis, and stroke are irreversible; therefore, early institution of therapy provides the best opportunity to prevent or slow the progression of life-threatening complications.

GAUCHER DISEASE

Gaucher disease, a panethnic autosomal recessive disorder, results from defective activity of acid β-glucosidase; ~600 *GBA*, a.k.a. *GBA1*, mutations have been described in such patients. Clinically, disease variants are classified by the absence or presence and progression of primary CNS involvement.

Gaucher disease type 1 is a nonneuronopathic disease (i.e., absence of early-onset or progressive CNS disease) presenting in childhood to adulthood as slowly to rapidly progressive visceral disease. About 55–60% of patients are diagnosed at <20 years of age in white populations and at even younger ages in other groups. This pattern of presentation is distinctly bimodal, with peaks at <10–15 years and at ~25 years. Younger patients tend to have greater degrees of hepatosplenomegaly and accompanying blood cytopenias. In contrast, older patients have a greater tendency for chronic bone disease. Hepatosplenomegaly occurs in virtually all clinically identified patients and can be minor or massive. Accompanying anemia and thrombocytopenia are variable and are not directly related to liver or spleen volumes. Severe liver dysfunction is unusual, but progressive cirrhosis can occur. Splenic infarctions can resemble an acute abdomen. Pulmonary hypertension and alveolar Gaucher cell accumulation are uncommon but life-threatening and can occur at any age; this is more common in females who have been splenectomized. *GBA1* mutations in the heterozygous or homozygous states lead to a significantly increased lifetime risk for developing Parkinson disease. The complex mechanisms for this major risk require elucidation.

All patients with Gaucher disease have nonuniform infiltration of bone marrow by lipid-laden macrophages, termed Gaucher cells. This phenomenon can lead to marrow packing with subsequent infarction, ischemia, necrosis, and cortical bone destruction. Bone marrow involvement spreads from proximal to distal in the limbs and can involve the axial skeleton extensively, causing vertebral collapse. In addition to bone marrow involvement, bone remodeling is defective, with loss of total bone calcium leading to osteopenia, osteonecrosis, avascular infarction, and vertebral compression fractures with spinal cord involvement. Aseptic necrosis of the femoral head is common, as is fracture of the femoral neck. The mechanism by which diseased bone marrow macrophages interact with osteoclasts and/or osteoblasts to cause bone disease is not well understood. Chronic, ill-defined bone pain can be debilitating and poorly correlated with radiographic findings. "Bone crises" are associated with localized excruciating pain and, on occasion, local erythema, fever, and leukocytosis. These crises represent acute infarctions of bone, as evidenced in nuclear scans by localized absent uptake of pyrophosphate agents. Decreased acid β-glucosidase activity (0–20% of normal) in nucleated cells establishes the diagnosis. The enzyme is not normally present in bodily fluids. The sensitivity of enzyme testing is poor for heterozygote detection; molecular testing by whole *GBA1* sequencing is the standard. The disease frequency varies from ~1 in 1000 among Ashkenazi Jews to <1 in 100,000 in other populations; ~1 in 12–15 Ashkenazi Jews carries a Gaucher disease allele. Four common mutations account for ~85% of the mutations in that population of affected patients: p.N370S (a.k.a. p.N409S), 84GG (a G insertion at cDNA position 84), p.L444P (also known as p.L483P), and IVS-2^{+1} (an intron 2 splice junction mutation).

Genotype/phenotype studies indicate a significant, though not absolute, correlation between disease type and severity and the *GBA1* genotype. The most common mutation in the Ashkenazi Jewish population (p.N370S) shares, either homozygously or heteroallelically, a 100% association with nonneuronopathic or type 1 Gaucher disease. The N370S/N370S and N370S/other mutant allele genotypes are associated with later-onset/less severe disease and with earlier-onset/severe disease, respectively. As many as 40% of individuals with the N370S/N370S genotype do not present clinically. Other alleles include L444P (very low activity), 84GG (null), or IVS-2 (null) and rare/private or uncharacterized alleles. The L444P/L444P patients frequently have life-threatening to very severe/early-onset disease, and many, though not all, develop CNS involvement in the first two decades of life.

Symptom-based treatment of blood cytopenias and joint replacement surgeries continue to have important roles in management. However, regular intravenous ET has been the first-line treatment for significantly affected patients and is highly efficacious and safe in diminishing hepatosplenomegaly and improving hematologic values. An oral substrate reduction therapy (SRT; eliglustat tartrate), which inhibits glucosylceramide synthesis, is approved as a first-line therapy for adults. Bone disease is decreased and can be prevented, but irreversible damage cannot be reversed, by ET. Adult patients may benefit from adjunctive treatment with bisphosphonates or other interventions to improve bone density. Adults who cannot be treated with enzyme, either because it is not effective or because they have developed an allergy or other hypersensitivities to the enzyme, may receive SRT with either eliglustat tartrate or miglustat; the latter is approved as a second-line oral therapy.

Gaucher disease type 2 is a rare, severe, progressive CNS disease that leads to death by 2 years of age, depending on supportive care. Gaucher disease type 3 has highly variable manifestations in the CNS and viscera. It can present in early childhood with rapidly progressive, massive visceral disease and slowly progress to static CNS involvement that may not be evident by standard IQ evaluations; in adolescence with dementia; or in early adulthood with rapidly progressive, uncontrollable myoclonic seizures and mild visceral disease. Visceral disease in type 3 is nearly identical to that in type 1 but is generally more severe. Early CNS findings may be limited to defects in lateral gaze tracking (supranuclear lateral gaze palsy), which may remain static for decades. Cognitive degeneration can be slowly progressive or static. Type 3 is much more frequent among individuals of non–Western world descent. Visceral—but not CNS—involvement responds to ET.

NIEMANN-PICK DISEASES

Niemann-Pick diseases (acid sphingomyelinase deficiency [ASMD]) are autosomal recessive disorders that result from defects in acid sphingomyelinase (ASM). Types A and B are distinguished by the early age of onset and progressive CNS disease in type A. Type A typically has its onset in the first 6 months of life, with rapidly progressive CNS deterioration, spasticity, failure to thrive, and massive hepatosplenomegaly. Type B has a later, more variable onset and is characterized by a progression of hepatosplenomegaly, with eventual development of cirrhosis and hepatic parenchymal and Kupffer cell replacement by foam cells filled with sphingomyelin. Affected patients develop potentially lethal progressive pulmonary disease with dyspnea, hypoxemia, and a reticular infiltrative pattern on chest x-ray. Foam cells are present in alveoli, lymphatic vessels, and pulmonary arteries. Progressive hepatic or lung disease can lead to death in adolescence or early adulthood. The "type B" phenotype includes some patients with slowly progressive CNS involvement, termed A/B variant.

The diagnosis is established by markedly decreased (1–10% of normal) ASM activity in nucleated cells. Intravenous ET improves nonneurologic manifestations, including pulmonary disease. The efficacies of hepatic transplant (HT) or bone marrow transplantation (BMT/HSCT) are not established. More complications than expected have occurred with these interventions due to either (1) recurrence of hepatic disease in the transplant following HT due to repopulation of bone marrow–derived ASM-deficient myeloid cells or (2) lack of clearance of sphingomyelin in hepatocytes by ASM cross-correction following the BMT/HSCT of ASM-normal bone marrow stem cells.

Niemann-Pick C diseases are progressive CNS diseases due to mutations in either of the genes encoding NPC1 or NPC2, lysosomal proteins involved in free cholesterol and selected sphingolipid transport out of the lysosome. They can present with liver or splenic disease, but their major manifestations are progressive CNS disease over one to two decades. Treatment with substrate inhibition agents (e.g., miglustat) has shown minor CNS effects, and substrate depletion with cyclodextrin is in clinical trials for NPC1 disease.

MUCOPOLYSACCHARIDOSES

Mucopolysaccharidosis type I (MPS I) is an autosomal recessive disorder caused by deficiency of α-L-iduronidase (IDU). The spectrum of involvement traditionally has been divided into three categories: (1) Hurler disease (MPS I H) for severe deficiency with neurodegeneration, (2) Scheie disease (MPS I S) for later-onset disease without neurologic involvement and with relatively less severe disease in other organ systems, and (3) Hurler-Scheie syndrome (MPS I H/S) for patients intermediate between these extremes. MPS I H/S is characterized by severe somatic disease, usually without major overt neurologic deterioration. MPS I often presents in infancy or early childhood as chronic rhinitis, clouding of the corneas, hepatosplenomegaly, and progressive dysmorphia. As the disease progresses, nearly every organ system can be affected. In the more severe forms, cardiac and respiratory diseases become life threatening in childhood. Skeletal disease can be severe, resulting in very limited mobility. There are two current treatments for the MPS I diseases. HSCT is the standard treatment for patients presenting at <2 years of age who appear to have or are at risk for neurologic degeneration. Because early diagnosis and intervention are essential, MPS I has been added to the recommended newborn screen (NBS). HSCT results in stabilization of CNS disease, reverses hepatosplenomegaly, and improves cardiac and respiratory disease. HSCT does not eliminate corneal disease or result in the resolution of progressive skeletal disease. ET effectively addresses hepatosplenomegaly and alleviates cardiac and respiratory disease. The enzyme does not penetrate the blood-brain barrier and does not directly affect CNS disease. ET and HSCT appear to have similar effects on visceral signs and symptoms. ET poses a lower risk of life-threatening complications and may therefore be advantageous for patients without CNS disease. A combination of ET and HSCT has been used, with ET initiated prior to transplantation in an attempt to reduce the disease burden. The experience with this approach is not well documented, but it appears to have advantages over HSCT alone. It is clear that HSCT has benefited patients. However, late cardiac and respiratory complications of MPS I are being reported including obstructive breathing requiring pressure support, cardiomyopathy, and/or valve disease. Regular follow-up for patients with MPS I is required throughout their lives even after successful HSCT.

Hunter disease (MPS II) is an X-linked disorder due to deficiency in iduronate sulfate sulfatase (IDS) and has manifestations similar to those of MPS I, including some variants with neurologic degeneration. There is no corneal clouding or other eye disease. Like MPS I, MPS II is clinically variable, with CNS and non-CNS variants. HSCT has not been successful in treating CNS disease associated with MPS II. The FDA and EMA have approved ET for the visceral manifestations of MPS II.

MPS IV or Morquio syndrome is a rare autosomal recessive condition (1 in 200,000–300,000) and is different than the other mucopolysaccharidoses in presenting as a spondyloepiphyseal skeletal dysplasia and hyperextensibility of all joints. There are also major heart and respiratory complications. This disorder often presents in childhood, but the age of onset and rate of progression are quite variable. Two variants, type A and type B, are caused by deficiencies in *N*-acetyl-galactosamine-6-sulfatase (*GALNS*) and an acid β-galactosidase, respectively. A recombinant human GALNS ET (elosulfase alfa) is approved for the treatment of MPS IVA, making it essential to confirm the specific enzyme diagnosis. Treatment has been shown to improve ambulatory mobility and decrease pain. There is no current specific treatment for MPS IVB.

ET for Maroteaux-Lamy disease (MPS VI), arylsulfatase B (ARSB) deficiency, has received FDA approval as well as approval by similar agencies in other countries. This very rare autosomal recessive disorder is characterized by hepatosplenomegaly, bone disease, heart disease, and respiratory compromise. Short stature is also an important manifestation. Visceral signs and symptoms are similar to those in MPS I; however, MPS VI is not associated with neurologic degeneration.

MPS VII, Sly syndrome, is due to mutations in *GUSB*, which encodes β-glucuronidase. Severe deficiency in this enzyme may present with fetal hydrops, which can lead to stillbirth or perinatal demise. Other patients with MPS VII may present later with short stature, coarse facial features, and hepatosplenomegaly. There is ET for this disorder (vestronidase alfa-vjbk).

POMPE DISEASE

Acid maltase (acid α-glucosidase deficiency) due to *GAA* mutations, also called Pompe disease, is the only LSD leading to primary glycogen storage. The classic severe infantile form presents with hypotonia, myocardiopathy, and hepatosplenomegaly, as well as total deficiency of the enzyme. This variant is rapidly progressive and generally results in death in the first year of life. However, as with other LSDs, there are early- and late-onset forms of this disorder.

The late-onset variants may be as common as 1 in 40,000; patients typically present with a slowly progressive myopathy that may resemble limb-girdle muscular dystrophy. Respiratory insufficiency may be the presenting sign or may develop with advancing disease. In late stages of the disease, patients may require mechanical ventilation, report swallowing difficulties, and experience loss of bowel and bladder control. Myocardiopathy is not usually present in late-onset variants of Pompe disease.

The FDA, EMA, and similar agencies have approved ETs for Pompe disease patients of all ages. This treatment clearly prolongs life in the infantile form, consistently resulting in improved cardiac function. Respiratory function is also improved in most treated infants if instituted before age 6 months. Some infants demonstrate marked improvement in motor functions, while others have minor changes in muscle tone or strength. Many states have instituted NBS for Pompe disease. In addition, newer protocols for treatment with methotrexate and rituximab have greatly decreased antidrug antibody reactions, particularly in the enzyme-absent infantile variants. The combination of NBS and immunomodulation preceding ET has greatly improved therapeutic response and long-term survival. Prevention of deterioration has been shown with GAA ET in the late-onset forms. Early intervention with acid α-glucosidase ET in such patients may limit or prevent deterioration, but very advanced disease will have significant irreversible components.

LYSOSOMAL ACID LIPASE DEFICIENCY

Wolman syndrome (now infantile-onset LALD) and cholesterol ester storage disease (now childhood/adult-onset LALD) are caused by deficiency of lysosomal acid lipase (LAL) due to autosomal recessive mutations in *LIPA*. The diagnosis is established by enzyme or gene analyses of LAL or *LIPA* in serum/plasma or nucleated cells, respectively. LAL hydrolyzes cholesterol esters (CEs) and triglycerides (TGs) delivered to the lysosome via the LDLR pathway. Accumulation of these in the tissues leads to progressive organ dysfunction including liver disease, intestinal malabsorption, heart dysfunction, and other manifestations. The most severe form presents in early infancy as a medical emergency with severe failure to thrive, vomiting, and hepatosplenomegaly. The infantile-onset LALD patients die without specific treatment by age 1 year (median age of death, 3.7 months) from tissue accumulations of CEs and TGs and near total absence of LAL. Childhood/adult-onset LALD can have a variable age of initial presentation with nonspecific signs but often involves elevated liver enzymes, microvesicular fatty liver disease, cryptogenic cirrhosis, and varying severities of hepato-/splenomegaly, due to massive accumulations of CEs secondary to the LAL substrate preference for TGs by the low LAL levels. Importantly, neither clinical variant manifests primary CNS disease. Without treatment, disease progresses throughout life and may result in early (adolescence, mean ~13 years) liver cirrhosis and (early adulthood) atherosclerosis or early death. Importantly, statins can decrease the hypercholesterolemia but do not alter the basic progressive tissue (e.g., liver) pathology. The majority of the later onset patients are evaluated by hepatology or lipidology physicians. ET for LALD has major effects in reversing disease manifestations and is approved for patients at all ages by the EMA, FDA, and several other country agencies.

KRABBE DISEASE

Deficiency in β-galactocerebrosidase (GALC) causes Krabbe disease, an autosomal recessive neurodegenerative disorder due to mutations in *GALC*. Krabbe disease is panethnic but quite rare. The early infantile form presents at an average age of 4 months and progresses rapidly, with death at an average age of 18 months. Later onset forms also exist and have onsets and survival that are highly variable. The early-onset form presents with hyperirritability, feeding problems, fever, seizures, and neurodegeneration. Blindness, hypotonia, and loss of voluntary movement develop over time. Later onset forms present with spasticity, ataxia, vision loss, and behavioral problems and progress to dementia and early death. There is no FDA-approved treatment, but early, i.e., first or second month after birth, presymptomatic HSCT has been used. This results in improved survival, but neurologic problems are still common and slowly progressive in the CNS and more severe in the peripheral nervous system. More recently, studies in mouse and dog models have used gene therapy with dramatic improvement in both neurologic function and survival. Human studies are in clinical trials.

NEURONAL CEROID LIPOFUSCINOSIS TYPE 2 (NCL2 OR CLN2)

There are at least 13 mutant genes that have been associated with storage of neuronal ceroid lipofuscin. CLN2 is due to mutations in *TPP1* and deficiency in tripeptidyl peptidase 1. This autosomal recessive neurodegenerative disorder typically presents between age 2 and 4 years, most commonly with seizures, ataxia, myoclonus, and vision loss. Motor skill losses include sitting, walking, speech, and feeding and lead to severe disability and eventually death at an average age of 12 years. Intellectual disability and behavioral problems also become increasingly severe with age. Most affected children are wheelchair bound in late childhood, and survival beyond adolescence is rare. There are also later onset patients, and there is significant clinical overlap between CLN2 and other CLNs; confirmation of the diagnosis by gene sequencing is essential. FDA/EMA-approved treatment of CLN2 is with cerliponase alfa, an ET that is administered by intracerebroventricular infusions over several hours every 2 weeks. Administration of cerliponase alfa is facilitated by placement of an indwelling intracerebroventricular port to allow reliable access. Currently, this is the only approved ET that is intrathecally administered. CLN2 is the only neuronal ceroid lipofuscinosis that has a specific treatment. Several others are in preclinical development.

ALPHA MANNOSIDOSIS

Alpha mannosidosis is a very rare disorder caused by deficiency in α-ᴅ-mannosidase due to biallelic loss of function of MAN2B1. This leads to buildup of mannose-rich oligosaccharides in the cells, blood, and urine. The clinical manifestations include intellectual disability, skeletal changes that can be incapacitating, immunodeficiency, ataxia, poor balance, hearing loss, neurodegeneration, and psychosis. Early treatment with ET has been shown to improve mobility, immune function, and quality of life.

FURTHER READING

Aldenhoven M et al: Long-term outcome of Hurler syndrome patients after hematopoietic cell transplantation: An international multicenter study. Blood 125:2164, 2015.

Dimachkie MM et al: Long-term safety and efficacy of avalglucosidase alfa in patients with late-onset Pompe disease. Neurology 99:e536, 2022.

Grabowski GA et al: Gaucher disease: Basic and translational science needs for more complete therapy and management. Mol Genet Metab 132:59, 2021.

Schoser B et al: Survival and long-term outcomes in late-onset Pompe disease following alglucosidase alfa treatment: A systematic review and meta-analysis. J Neurol 264:621, 2017.

Schulz A et al: Study of intraventricular cerliponase alfa for CLN2 disease. N Engl J Med 378:1898, 2018.

Wanner C et al: Global reach of over 20 years of experience in the patient-centered Fabry Registry: Advancement of Fabry disease expertise and dissemination of real-world evidence to the Fabry community. Mol Genet Metab 139:107603, 2023.

430

Glycogen Storage Diseases and Other Inherited Disorders of Carbohydrate Metabolism

Rebecca L. Koch, Priya S. Kishnani

Carbohydrate metabolism plays a vital role in cellular function by providing the energy required for most metabolic processes. The relevant biochemical pathways involved in the metabolism of these carbohydrates are shown in **Fig. 430-1**. Glucose is the principal substrate of energy metabolism in humans. Metabolism of glucose generates ATP through glycolysis and mitochondrial oxidative phosphorylation. The body obtains glucose primarily through the ingestion of polysaccharides (e.g., starch) and disaccharides (e.g., lactose, maltose, and sucrose). Galactose and fructose are two other monosaccharides that serve as sources of fuel and precursors for cellular metabolism. However, their role as fuel sources is less significant than that of glucose. Galactose is primarily derived from lactose—a disaccharide (galactose + glucose) found in dairy products—and it is an important component of certain glycolipids, glycoproteins, and glycosaminoglycans. Fructose is naturally found in fruits, vegetables, and honey, as well as in foods containing sucrose (fructose + glucose).

The symptoms of the various disorders of carbohydrate metabolism range from minimally harmful to lethal. Clinical and biochemical manifestations of these disorders reflect the enzymatic defect, and as such, there are similarities being recognized in disorders falling along the same metabolic pathways, such as the similar phenotypes observed in disorders of glycogenolysis. Many disorders of carbohydrate metabolism are managed with specific dietary interventions. However, diet therapy alone does not prevent long-term complications, and there is a need for definitive therapies. Genes and corresponding pathogenic variants responsible for inherited defects of carbohydrate metabolism have been identified. With the use of tools such as DNA sequencing panels, whole exome sequencing, whole genome sequencing, and RNA-Seq, new disorders in carbohydrate metabolism continue to be identified and the phenotype of known disorders continues to expand. Advances in the molecular basis of these diseases are now being used to improve diagnosis and management. Some of these disorders are candidates for enzyme replacement therapy, substrate reduction therapy, mRNA therapy, gene replacement therapy, and other genomic tools, such as small interfering RNA (siRNA) technology and genome editing technology.

GLYCOGEN STORAGE DISEASES

Glycogen, the storage form of glucose in animal cells, is a tree-like molecule composed of glucose residues joined in straight chains by α-1,4 linkages and branched at intervals of 4–10 residues by α-1,6 linkages. Inherited defects in glycogen metabolism are referred to as *glycogen storage diseases* (GSDs), also referred to as *glycogenoses*. The glycogen in GSDs may be abnormal in quality (i.e., abnormal structure) and/or quantity. GSDs were historically classified in order of when the enzymatic defects were identified. GSDs are also classified based on the primary organs involved (liver or muscle) as well as clinical manifestations **(Table 430-1)**. The overall prevalence of all forms of GSDs is not known but is estimated to be 1 in 10,000 to 1 in 25,000 live births, with most inherited in an autosomal recessive manner and a few in an X-linked or autosomal dominant manner. With increased access to genetic testing and newborn screening programs, understanding the prevalence of all GSDs will improve. GSDs are pan-ethnic disorders but may have a relatively higher prevalence in certain populations, such as the adult-onset form of GSD IV (adult polyglucosan body disease [APBD]) in individuals of Ashkenazi Jewish ancestry or GSD VI in the Mennonite community. The most common childhood disorders are glucose-6-phosphatase deficiency (GSD type I), lysosomal acid α-glucosidase deficiency (GSD type II), debrancher enzyme deficiency (GSD type III), and liver phosphorylase kinase (PhK) deficiency (GSD type IX). The most common adult disorder is myophosphorylase deficiency (GSD type V).

LIVER GLYCOGEN STORAGE DISEASES

GSD Type I (Glucose-6-Phosphatase or Translocase Deficiency, Von Gierke Disease) GSD type I is an autosomal recessive disorder caused by glucose-6-phosphatase or translocase deficiency in liver, kidney, and intestinal mucosa. There are two subtypes of GSD I: type Ia, in which the glucose-6-phosphatase enzyme is defective (Online Mendelian Inheritance in Man [OMIM] #232200), and type Ib, in which the translocase that transports glucose-6-phosphate from the cytoplasm into the lumen of the endoplasmic reticulum is defective (OMIM #232220). The defects in both subtypes lead to inadequate conversion of glucose-6-phosphate to glucose and thus make affected individuals susceptible to fasting hypoglycemia.

Persons with GSD I typically present with hepatomegaly between 3 and 4 months of age but may present earlier with hypoglycemia and lactic acidosis during the neonatal period. Hypoglycemia and lactic acidosis can develop after a short fast, typically when infants start sleeping through the night or due to an intercurrent illness. Liver biopsy reveals hepatocytes distended by glycogen and fat with large and prominent lipid vacuoles. These children usually have doll-like facies with round cheeks, relatively thin extremities, short stature, and a protuberant abdomen that is due to massive hepatomegaly. The kidneys are enlarged, but the spleen and heart are of normal size. Easy bruising and epistaxis are associated with prolonged bleeding time as a result of impaired platelet aggregation/adhesion and/or an acquired von Willebrand–like disease. Laboratory abnormalities include hyperuricemia, hypertriglyceridemia, lactic acidosis, and lipid profile abnormalities including elevated total and low-density lipoprotein (LDL) cholesterol compared to the lower level of high-density lipoprotein (HDL) cholesterol. The serum of untreated patients is cloudy/milky due to the high concentrations of lipids. Type Ib patients have additional findings of neutropenia and impaired neutrophil function with associated recurrent bacterial infections and chronic oral and intestinal mucosal ulceration with risk for mouth sores, severe diarrhea, and

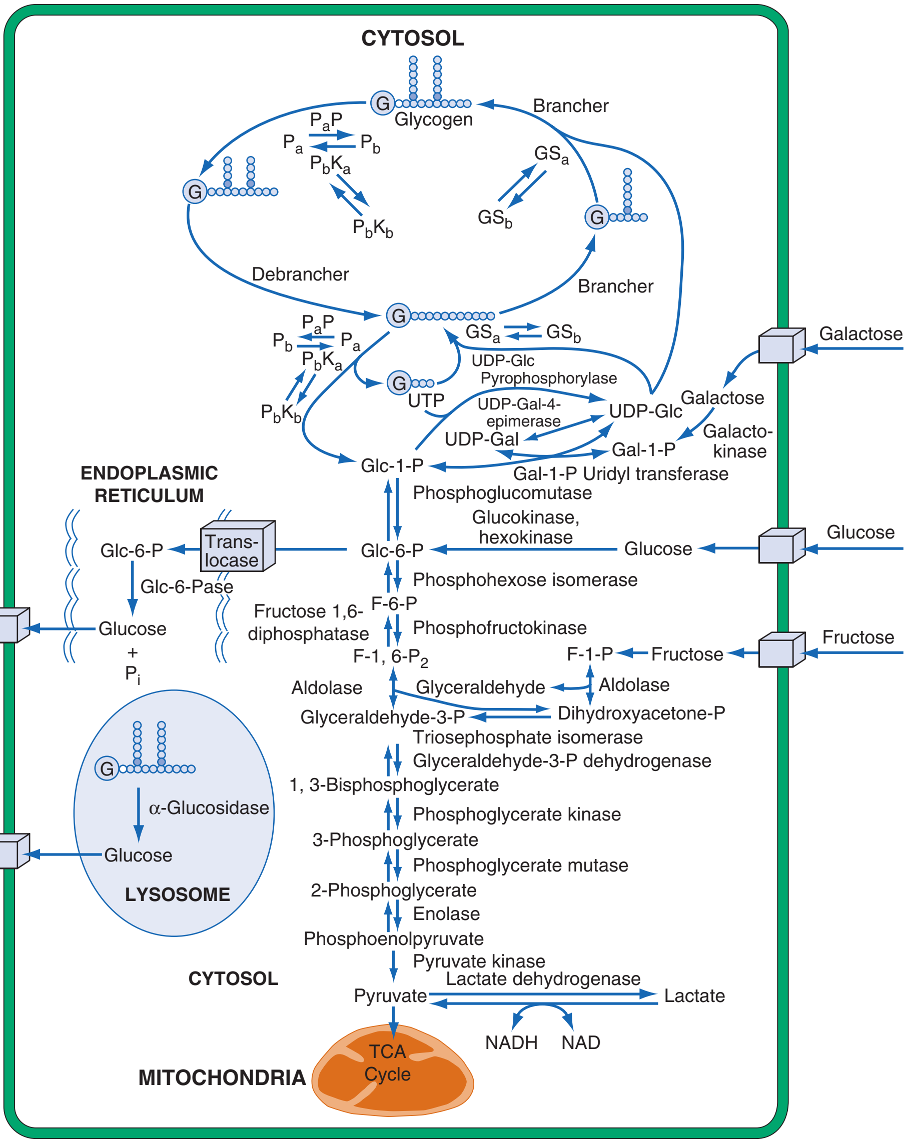
CYTOSOL
G
Glycogen
Brancher
P_aP
P_a
P_b
P_bK_a
P_bK_b
GS_a
GS_b
G
G
Debrancher
Brancher
G
P_aP
P_b
P_a
P_bK_a
P_bK_b
GS_a
GS_b
UDP-Glc Pyrophosphorylase
G
UTP
UDP-Gal-4-epimerase
UDP-Glc
Galactose
UDP-Gal
Gal-1-P
Galacto-kinase
Galactose
Glc-1-P
Gal-1-P Uridyl transferase
ENDOPLASMIC RETICULUM
Phosphoglucomutase
Glucokinase, hexokinase
Glc-6-P
Trans-locase
Glc-6-P
Glucose
Glucose
Glc-6-Pase
Phosphohexose isomerase
Fructose 1,6-diphosphatase
F-6-P
Glucose
+
P_i
Phosphofructokinase
F-1, 6-P_2
F-1-P
Fructose
Fructose
Aldolase
Glyceraldehyde
Aldolase
Glyceraldehyde-3-P
Dihydroxyacetone-P
Triosephosphate isomerase
Glyceraldehyde-3-P dehydrogenase
1, 3-Bisphosphoglycerate
G
α-Glucosidase
Phosphoglycerate kinase
3-Phosphoglycerate
Glucose
Phosphoglycerate mutase
LYSOSOME
2-Phosphoglycerate
Enolase
Phosphoenolpyruvate
Pyruvate kinase
CYTOSOL
Lactate dehydrogenase
Pyruvate
Lactate
NADH
NAD
TCA Cycle
MITOCHONDRIA

FIGURE 430-1 Metabolic pathways related to glycogen storage diseases and galactose and fructose disorders. Nonstandard abbreviations are as follows: GS_a, active glycogen synthase; GS_b, inactive glycogen synthase; P_a, active phosphorylase; P_b, inactive phosphorylase; P_aP, phosphorylase α phosphatase; P_bK_a, active phosphorylase β kinase; P_bK_b, inactive phosphorylase β kinase; G, glycogenin, the primer protein for glycogen synthesis. *(Modified from AR Beaudet, in KJ Isselbacher et al: Harrison's Principles of Internal Medicine, 13th ed. New York, NY: McGraw Hill; 1994.)*

malnutrition. Genetic testing provides a noninvasive means of reaching a definitive diagnosis for most patients with GSD types Ia and Ib.

Puberty and growth are often delayed in individuals with GSD I. Long-term complications include gout, which becomes symptomatic at puberty as a result of long-term hyperuricemia in untreated patients. Secondary to lipid abnormalities, there is an increased risk of pancreatitis. Patients with GSD I may be at increased risk for cardiovascular disease such as systemic hypertension and vascular dysfunction, though the latter remains debated. Pulmonary hypertension—although rare—has been reported. In adult patients, frequent fractures can occur, and radiographic evidence of osteopenia/osteoporosis can be found; in prepubertal patients, bone mineral content is significantly reduced. By the second or third decade of life, many patients with GSD I develop hepatocellular adenomas (HCA) that can hemorrhage and, in some cases, patients can develop hepatocellular carcinoma (HCC). End-stage renal disease is a serious late complication. Almost all patients aged >20 years have proteinuria, and many have hypertension, kidney stones, nephrocalcinosis, and altered creatinine clearance. Other complications include thyroid autoimmunity, particularly in GSD Ib, as well as hypercalcemia. Some women with GSD I have polycystic ovaries and menorrhagia; several reports of successful pregnancies suggest that fertility is not affected, although symptoms may be exacerbated due to pregnancy-related increases in renal perfusion and maternal blood volume.

The first line of treatment in GSD I is avoidance of fasting and frequent feedings. Due to fasting intolerance, continuous glucose monitoring is useful for monitoring blood glucose levels and trends. GSD I affects both glycogenolysis and gluconeogenesis, and therefore, a diet high in complex carbohydrates supplemented by uncooked cornstarch in small, frequent feedings provides a source of glucose and is used for treatment of both GSD Ia and GSD Ib. Modified, extended-release cornstarch products that are longer acting and better tolerated are available and may help extend the duration of euglycemia and improve metabolic control. Aminotransferase levels may be elevated at baseline, yet often quickly return to normal with treatment. Lipid profile abnormalities and levels of lactate and uric acid can also improve with good metabolic control yet remain a significant concern. Treatment of complications with medications may be necessary, such as citrate supplementation to prevent and/or treat nephrocalcinosis, allopurinol to control hyperuricemia, HMG-CoA reductase inhibitors and fibrate to reduce lipids, and angiotensin-converting enzyme inhibitors to treat microalbuminuria. Surgical resection, percutaneous ethanol injections, and radiofrequency ablation can be used to treat HCA. Liver transplantation can be lifesaving for those with HCA with the risk of malignant transformation to HCC, rapid growth in size or number, and/or severe/chronic poor metabolic control. Dialysis or kidney transplantation may be required in those with renal glomerular dysfunction progressing to end-stage renal disease. Gene replacement therapy and mRNA therapy are in clinical trials as treatment modalities for GSD Ia. Individuals with GSD Ib may require further intervention due to the consequences of neutropenia and neutrophil dysfunction. Historically, granulocyte colony-stimulating factor treatment was used, but empagliflozin, a renal sodium-glucose cotransporter 2 (SGLT2) inhibitor has recently been used to effectively treat the neutrophil dysfunction in GSD Ib. Though less common, there are reports of patients with GSD Ia homozygous for p.G188R variants developing neutropenia, neutrophil dysfunction, and recurrent infections traditionally characteristic of GSD Ib.

GSD Type III (Debranching Enzyme Deficiency, Limit Dextrinosis, Cori or Forbes Disease) GSD type III (OMIM #232400) is an autosomal recessive disorder caused by a deficiency of glycogen debranching enzyme. PhK, glycogen phosphorylase, and debranching enzyme are collectively responsible for the complete degradation of glycogen into glucose (glycogenolysis). When debranching enzyme is defective, glycogen breakdown is incomplete, resulting in abnormal glycogen accumulation with short outer chains, resembling limit dextrin. GSD III is mainly classified as (1) GSD IIIa, with liver, cardiac, and skeletal muscle involvement (~85% of cases), and (2) GSD IIIb, with primarily liver involvement (~15% of cases). The initial

TABLE 430-1 Features of Glycogen Storage Diseases, Galactose Metabolism Disorders, and Fructose Metabolism Disorders

TYPE	EPONYM	DEFICIENT ENZYME	CLINICAL FEATURES	COMMENTS
Liver Glycogen Storage Diseases				
0a		Glycogen synthase 2	Fasting ketotic hypoglycemia, growth retardation, postprandial hyperglycemia, lactic acidosis, and hyperalaninemia with normal insulin levels	Rare; no hepatomegaly
Ia	von Gierke	Glucose-6-phosphatase-α	Fasting hypoketotic hypoglycemia, lactic acidosis, hypercholesterolemia, hypertriglyceridemia, hyperuricemia, hepatomegaly and renomegaly, hepatic steatosis, short stature, doll-like facies with round cheeks, thin extremities	~80% of type I cases; complications in adolescence and adulthood include HCA with risk for HCC transformation in some, gout, osteoporosis, pulmonary hypertension, renal stones, hyperfiltration injury in early stages that can progress to renal failure, pancreatitis
Ib		Glucose-6-phosphate translocase	As for Ia, with additional findings of neutropenia and neutrophil dysfunction, inflammatory bowel disease, and increased risk for infections, mucosal ulceration, and periodontal disease	~20% of type I cases
IIIa	Cori or Forbes	Glycogen debranching enzyme (liver and muscle)	*Childhood:* Fasting ketotic/nonketotic hypoglycemia, hepatomegaly, short stature, muscle weakness, cardiomyopathy, cardiac arrhythmia, elevated aminotransferases, CK, and urine Glc$_4$ levels, hyperlipidemia *Adulthood:* Variable progression of liver fibrosis, cirrhosis with risk for liver failure, risk for HCC and HCA noted in some; progression of proximal and distal muscle atrophy and weakness, peripheral neuropathy with preferential median nerve involvement; variable cardiac arrhythmia and cardiomyopathy, PCOS	~85% of type III cases; muscle weakness is both proximal and distal and can progress to need for ambulatory aids such as wheelchair; risk of sudden death from cardiac arrhythmia
IIIb		Glycogen debranching enzyme (liver)	Liver symptoms same as in type IIIa; no skeletal muscle or cardiac symptoms	~15% of type III cases
IV	Andersen	Glycogen branching enzyme	Clinical continuum whereby hepatic, neurologic, cardiac, and muscular involvement can occur to varying degrees *Hepatic form:* Failure to thrive, hepatomegaly, elevated aminotransferase levels, progressive liver fibrosis and liver failure (death usually before age 5 years without liver transplantation); some do not progress to liver failure but may develop extrahepatic involvement such as myopathy and cardiomyopathy and are at risk for hepatic disease progression. Progression to APBD noted in some in adulthood. *Neuromuscular forms:* Perinatal and congenital forms lead to death in the neonatal period/infancy. Less severe involvement may present in early to late childhood with myopathy, neuromuscular dysfunction, and/or cardiomyopathy. *Adult-onset (APBD):* Progressive bilateral lower limb weakness and spasticity, neurogenic bladder (urinary incontinence and/or retention), peripheral neuropathy, cognitive impairment	Rare; APBD is prevalent in individuals of Ashkenazi Jewish ancestry
VI	Hers	Glycogen phosphorylase (liver)	Fasting ketotic/nonketotic hypoglycemia, hepatomegaly, liver fibrosis, hyperlipidemia, short stature	Often underdiagnosed; high prevalence in the Mennonite community; severe cases being recognized; cirrhosis, with progression to HCC has been reported in some; focal nodular hyperplasia and HCA also reported
IX α2 (previously known as IXa) IX γ2 (previously known as IXc) IX β (previously known as IXb)		Phosphorylase kinase (liver) Phosphorylase kinase (liver) Phosphorylase kinase (liver and muscle)	Fasting ketotic/nonketotic hypoglycemia, hepatomegaly, liver fibrosis, hyperlipidemia, elevated aminotransferase and urine Glc$_4$ levels, short stature In addition to features of liver disease observed in IX α2 and IX γ2, may have mild muscle weakness and hypotonia	GSD IX α2 is more common in males (X-linked, *PHKA2*); clinical phenotype of IX γ2 is often more severe than IX α2 and IX β, yet there is significant clinical variability within and between subtypes with severe cases being recognized in all subtypes; cirrhosis with progression to HCC has been reported; HCA also noted; catch-up growth is common
Fanconi-Bickel syndrome	Fanconi-Bickel	Glucose transporter 2 (GLUT2)	Rickets, short stature, failure to thrive, fasting hypoglycemia, postprandial hyperglycemia (impaired glucose and galactose utilization), hepatomegaly, hypophosphatemia, proximal tubular nephropathy, metabolic acidosis	Rare

(*Continued*)

TABLE 430-1 Features of Glycogen Storage Diseases, Galactose Metabolism Disorders, and Fructose Metabolism Disorders (Continued)

TYPE	EPONYM	DEFICIENT ENZYME	CLINICAL FEATURES	COMMENTS
Muscle Glycogen Storage Diseases				
0b		Glycogen synthase 1	Cardiomyopathy, cardiac arrythmia, muscle weakness, exercise intolerance	Rare
II	Pompe	Lysosomal acid α-glucosidase	*Classic infantile-onset:* Hypotonia, muscle weakness, cardiomyopathy, fatal within first 2 years of life *Nonclassic infantile-onset:* Presentation within first year of life with less severe cardiomyopathy and slower progression than the classic form *Late-onset:* Absence of cardiomyopathy in first year of life; kinematic/muscle weakness noted as early as first year of life, progressive skeletal muscle weakness, early involvement of proximal lower extremity muscles and respiratory muscles, especially the diaphragm	Higher frequency being reported since advent of newborn screening; enzyme (GAA) levels are undetectable or very low in infantile-onset form and variable in late-onset form when measured in skin or muscle
V	McArdle	Muscle phosphorylase	Exercise intolerance, myalgia, cramps, myoglobinuria with strenuous exercise, increased CK levels; "second wind" phenomenon	
VII	Tarui	Phosphofructokinase-M subunit	As for type V, with additional findings of compensated hemolysis, hyperuricemia; "out of wind" phenomenon rather than "second wind" phenomenon observed in GSD V	Rare, prevalent in individuals of Ashkenazi Jewish ancestry
IX α1 (previously known as IXd)		Phosphorylase kinase (muscle)	Exercise intolerance, myalgia, cramps, muscle weakness and atrophy, myoglobinuria with strenuous exercise	More common in males (X-linked, *PHKA1*)
X		Phosphoglycerate mutase 2	Exercise intolerance, cramps, myoglobinuria with strenuous exercise	Rare, most patients African American
XI – Lactate dehydrogenase A deficiency		Lactate dehydrogenase A	Exercise intolerance, cramps, myoglobinuria with strenuous exercise, dermatitis	Rare; reports of uterine stiffness and pain in pregnant patients
XII		Aldolase A	Muscle weakness, myoglobinuria with strenuous exercise, with additional finding of hemolytic anemia	Rare
XIII		β-enolase	Exercise intolerance, myalgia, cramps, myoglobinuria with strenuous exercise	Rare
XV (also known as polyglucosan body myopathy 2)		Glycogenin-1	Often classified as two subtypes: *GSD XV:* Progressive cardiac arrhythmia and/or cardiomyopathy causing heart failure and necessitating cardiac transplantation in some cases *Polyglucosan body myopathy 2:* Progressive myopathy with proximal and distal muscle weakness	Rare
Danon disease		Lysosomal-associated membrane protein 2	Severe progressive cardiomyopathy with WPW pattern progressing to heart failure, cardiac arrythmia, and heart failure in early childhood/adolescence; myopathy with muscle weakness, retinopathy, intellectual disability, and developmental delay	More common in males (X-linked, *LAMP2*)
PRKAG2 deficiency		AMP-activated gamma 2 protein kinase	Progressive cardiac arrhythmia and cardiomyopathy progressing to heart failure in childhood to adulthood; congenital/infantile form is rapidly fatal with hypertrophic cardiomyopathy with WPW pattern; other involvement includes myalgia and seizures	Autosomal dominant
Galactose Metabolism Disorders				
I (classical galactosemia)		Galactose-1-phosphate uridyltransferase (GALT)	Vomiting, jaundice, sepsis, cataracts, aminoaciduria, hepatomegaly, failure to thrive	Increased risk for developmental delays, speech problems, and motor dysfunction despite early diagnosis and treatment
Clinical variant galactosemia				Less severe disease course compared to classic galactosemia; caused by homozygous p.Ser135Leu variants, most common in African Americans and native South Africans
Biochemical variant galactosemia	Duarte	Galactose-1-phosphate uridyltransferase (GALT)	Asymptomatic	Caused by heterozygous or homozygous c.940A>G;-16_119delGTCA (Duarte variant)
II		Galactokinase (GALK)	Cataracts and pseudotumor cerebri	Clinical continuum with very mild to severe phenotype
III		Uridine diphosphate galactose-4-epimerase (GALE)	Similar to galactosemia type I; additional findings of hypotonia and hearing loss	Benign in some cases; more severe phenotype has been reported in others
IV		Galactose mutarotase (GALM)	Cataracts	Identified in Japan; considered to be mild

(*Continued*)

TABLE 430-1 Features of Glycogen Storage Diseases, Galactose Metabolism Disorders, and Fructose Metabolism Disorders (Continued)

TYPE	EPONYM	DEFICIENT ENZYME	CLINICAL FEATURES	COMMENTS
Fructose Metabolism Disorders				
Essential/benign fructosuria		Fructokinase	Asymptomatic	Benign, incidental finding made through detection of reducing substance in urine
Hereditary fructose intolerance		Aldolase B	Vomiting, jaundice, hepatomegaly, aversion to sweets, failure to thrive, convulsions, postprandial hypoglycemia	Severity of symptoms depending on age and quantity of fructose ingested; prognosis good with early diagnosis and fructose restriction

Abbreviations: APBD, adult polyglucosan body disease; CK, creatine kinase; GSD, glycogen storage disease; HCA, hepatocellular adenoma; HCC, hepatocellular carcinoma; PCOS, polycystic ovarian syndrome; PhK, phosphorylase kinase; WPW, Wolff-Parkinson-White.

presentation of GSD III is similar to that of GSD I with hypoglycemia (typically ketotic, yet some may present with nonketotic hypoglycemia), hepatomegaly, hyperlipidemia, and short stature occurring in infancy and early childhood. A distinguishing feature from GSD I is normal lactate and uric acid, and patients with GSD III have marked elevated aminotransferase levels in early stages that persist and then decrease over time. Levels of urine glucose tetrasaccharide (Glc_4), a breakdown product of glycogen, are typically elevated but follow the same pattern as aminotransferases and decrease over time. In most patients with GSD III, there is an apparent improvement in hepatomegaly and liver enzymes with age; however, there is now a recognition that patients develop progressive liver fibrosis and cirrhosis progressing to liver failure, and some can develop HCC. HCA may occur, although less commonly than in GSD I. Polycystic ovaries are reported in female patients with GSD III, and some patients develop features of polycystic ovarian syndrome, such as hirsutism and irregular menstrual cycles. Reports of successful pregnancy in women with GSD III suggest that fertility is normal.

Patients with GSD IIIa also have a variable skeletal myopathy and cardiomyopathy that can present in the first decade of life; these are often overlooked due to the significant hepatomegaly. Serum creatine kinase (CK) levels can be used to identify patients with muscle involvement, but normal levels are often noted in patients with muscle involvement. Left ventricular hypertrophy, significant scarring of the myocardium, and life-threatening arrhythmias resulting in death have been reported. Patients with GSD IIIa may experience exercise intolerance and muscle weakness in early childhood that can become severe after the third or fourth decade of life, resulting in use of assistive devices and wheelchair dependence. The pattern of muscle weakness is variable, and both proximal and distal muscle weakness are seen. Peripheral neuropathy with preferential median nerve involvement may become discernible later in life; however, there are opposing views concerning the existence of peripheral neuropathy in GSD III. Individuals with GSD IIIa are at an increased risk of osteoporosis.

In the past, diagnosis of GSD IIIa versus IIIb was confirmed by deficient or absent debranching enzyme activity in liver or skeletal muscle (enzyme activity is low in liver and normal in muscle in IIIb). However, with the availability of molecular genetic testing, reliance on invasive tissue biopsies is declining and DNA-based analyses now provide a noninvasive way of subtyping most patients. Specific variants in exon 2 of *AGL* are associated with the GSD IIIb phenotype and can occur with other variants in *AGL*. Liver histology reveals distended hepatocytes due to glycogen buildup and may show fat infiltration; fibrosis is present early on in the disease and has a variable rate of progression.

Debrancher enzyme deficiency prevents complete glycogenolysis in GSD III, but gluconeogenesis is intact. Hence, a high-protein diet with complex carbohydrates supplemented with uncooked cornstarch in small, frequent feedings is effective in preventing hypoglycemia. Individuals with GSD IIIa may benefit from dietary lipid manipulation, such as through the implementation of a high-fat, high-protein diet or use of medium-chain triglyceride supplementation. Careful monitoring of liver function, morphology, lipid profile, and growth is necessary given the potential impact on underlying liver disease and muscle disease progression. Diet therapy is not effective in preventing the overall progression of hepatic disease, cardiomyopathy, and myopathy. Just as in GSD I, continuous glucose monitoring is useful for monitoring blood glucose levels and trends. Blood ketones and glucose should be carefully evaluated during times of stress. Liver and heart transplantation may be considered in those with severe hepatic or cardiac involvement, respectively. Even with liver or heart transplantation, muscle disease continues to progress and is a significant unmet need for these patients.

GSD Type IV (Glycogen Branching Enzyme Deficiency, Andersen Disease, Adult Polyglucosan Body Disease, Amylopectinosis) GSD type IV (OMIM #232500) is caused by deficiency of branching enzyme, leading to accumulation of abnormal glycogen with poor solubility (polyglucosan). The disease is clinically heterogeneous, with multisystem organ involvement, yet the primary presentation was traditionally characterized by manifestations in either liver or muscle; thus, two main types—hepatic and neuromuscular—have been recognized. Individuals with the "classic progressive" hepatic subtype typically present in the first 18 months of life with failure to thrive, hepatosplenomegaly, and progressive liver cirrhosis leading to liver failure and death in the first 5 years of life without liver transplantation. Extrahepatic manifestations including cardiac and nervous system involvement may present and progress after liver transplantation. A less severe hepatic form referred to as the "nonprogressive" hepatic subtype has been observed; in these individuals, liver disease is slowly progressive or attenuated and liver transplantation is not considered urgent. Lifelong monitoring is imperative as there are reports of individuals with hepatic GSD IV that developed HCC, progressed to liver failure, and developed extrahepatic disease manifestations. Some patients have gone on to develop symptoms consistent with APBD.

The neuromuscular forms of the disease have four recognized subtypes: perinatal, congenital, childhood, and adult forms. The perinatal and congenital forms are lethal, and death occurs in the neonatal period. The childhood form presents with myopathy or cardiomyopathy, with typical systemic findings. The adult form is known as APBD (OMIM #263570) and includes central and peripheral nervous system involvement characterized by gait abnormalities due to spastic paraplegia, neurogenic bladder, peripheral neuropathy, leukodystrophy, autonomic dysfunction, and, in the advanced stages, cognitive impairment. Life expectancy is considered to be shortened. APBD has a high prevalence in the Ashkenazi Jewish population (estimated carrier rate 1 in 48).

Definitive diagnosis of GSD IV requires demonstration of pathogenic variants in the *GBE1* gene. Enzyme activity demonstrating reduced glycogen branching enzyme activity in cultured skin fibroblasts, leukocytes, liver, or muscle is supportive of a GSD IV diagnosis. Treatment for all forms of GSD IV is symptomatic. Recently, our understanding of the natural history has improved, and patients with GSD IV are no longer classified by the traditional discrete hepatic/neuromuscular subtypes; rather, GSD IV is conceptualized as a multidimensional clinical continuum whereby hepatic, neuromuscular, and cardiac involvement occurs to varying degrees in different patients.

GSD Type VI (Liver Phosphorylase Deficiency) GSD type VI (OMIM #232700) is an autosomal recessive disorder caused by liver glycogen phosphorylase deficiency. PhK regulates liver glycogen phosphorylase activity by converting the inactive glycogen phosphorylase b form to the active a form (phosphorylated glycogen phosphorylase). Therefore, the clinical presentation of GSD VI is very similar to liver

GSD IX (PhK deficiency) and includes variable fasting hypoglycemia (ketotic or nonketotic), hepatomegaly, elevated aminotransferases, and hyperlipidemia. Historically, GSD VI was thought to have a benign course, but clinical heterogeneity in symptom severity has been observed. Liver fibrosis, cirrhosis, focal nodular hyperplasia, HCA, and HCC have been reported. Diagnosis is confirmed by identifying two pathogenic variants in the *PYGL* gene. Glycogen phosphorylase enzyme levels will be reduced or deficient in liver tissue. It is thought that GSD VI is generally underdiagnosed and has a high prevalence in the Mennonite community (1 in 1000). Just like in other disorders of glycogenolysis (GSD III and IX), gluconeogenesis is intact, and therefore, a high-protein diet is recommended along with small frequent meals with complex carbohydrates and avoidance of fasting to prevent hypoglycemia. Uncooked cornstarch supplementation may be needed to maintain euglycemia, and continuous glucose monitoring is useful for monitoring blood glucose levels and trends. Blood ketones and glucose should be carefully evaluated during times of stress.

GSD Type IX (Liver Phosphorylase Kinase Deficiency)

Deficiency of the PhK causes GSD IX, a heterogeneous group of disorders. Similar to patients with GSD VI, patients with hepatic GSD IX typically present with fasting hypoglycemia (ketotic or nonketotic), hepatomegaly, and elevated aminotransferase, cholesterol, triglyceride, and urine Glc_4 levels. The PhK enzyme complex consists of four subunits (α, β, γ, and δ), and each subunit is encoded by different genes that are differentially expressed in various tissues. GSD IX can thus be divided into several subtypes on the basis of the gene/subunit involved.

The most common subtype is GSD IX α2 (also known as GSD IXa; OMIM #306000), an X-linked liver PhK deficiency caused by pathogenic variants in the *PHKA2* gene. PhK activity may also be deficient in erythrocytes and leukocytes but is normal in skeletal muscle. Much like GSD VI, GSD IX α2 has historically been considered to have a benign course, but clinical heterogeneity in symptom severity is being increasingly recognized, with severe disease involvement in some cases. Liver histology shows distention of hepatocytes due to excess glycogen accumulation as well as fibrosis ranging from mild to severe. Other subtypes of type IX liver GSD include GSD IX β (also known as GSD IXb; OMIM #261750) and GSD IX γ2 (also known as IXc; OMIM #613027). GSD IX β is an autosomal recessive form of liver and muscle PhK deficiency caused by *PHKB* pathogenic variants. The affected gene *PHKB* is expressed in the liver and muscle so myopathy may be observed but is typically less severe compared to the liver disease. GSD IX γ2 accounts for 25% of GSD IX and is an autosomal recessive form of liver PhK deficiency due to pathogenic variants in *PHKG2*. GSD IX γ2 is typically a more severe phenotype when compared to GSD IX α2 and GSD IX β due to defects in the catalytic domain of PhK. Affected patients often present with severe liver fibrosis and may progress to cirrhosis. Previously, infants with severe isolated cardiomyopathy and low PhK activity in the heart and muscle were considered to have a subtype of GSD IX; however, there were no pathogenic variants in the genes encoding for the PhK subunits. This presentation was later classified as *PRKAG2 syndrome*, which causes a secondary decrease in PhK activity. Details about PRKAG2 syndrome are described later in the "Muscle Glycogen Storage Diseases" section.

Treatment of liver GSD IX (types α2, β, and γ2) is symptom based. Like in other disorders of glycogenolysis (GSD III and VI), gluconeogenesis is intact, and therefore, a high-protein diet with complex carbohydrates in small, frequent feedings is needed to prevent hypoglycemia. Uncooked cornstarch supplementation may be needed to maintain euglycemia, and continuous glucose monitoring is useful for monitoring blood glucose levels and trends. Blood ketones and glucose should be carefully evaluated during times of stress. Liver transplantation may be considered in those with liver failure. In addition to fibrosis and cirrhosis, HCA and HCC have been observed in adolescents and adults with hepatic GSD IX, and therefore, it is recommended that patients be monitored for hepatic complications with regular computed tomography (CT) or magnetic resonance imaging (MRI) scans.

Though previously thought to be a mild disease, a broad clinical spectrum of presentations is now being recognized in hepatic GSD IX, with more severe cases coming to light, even in the X-linked form. Further research is needed to completely understand the natural history and long-term complications of liver GSD IX.

Other Liver GSDs GSD 0a is a rare disorder caused by deficiency in glycogen synthase 2 (GYS2; OMIM #240600) resulting in abnormal glycogen synthesis in the liver. Clinical features of GSD 0a include fasting ketotic hypoglycemia, short stature, postprandial hyperglycemia, lactic acidosis, and hyperalaninemia with normal insulin levels. Unlike many liver GSDs, hepatomegaly is not a feature of GSD 0a. Fanconi-Bickel syndrome (OMIM #227810), previously referred to as GSD XI-Fanconi-Bickel syndrome, is caused by defects in the facilitative glucose transporter 2 (GLUT-2), which transports glucose in and out of hepatocytes, pancreatic cells, and the basolateral membranes of intestinal and renal epithelial cells. Patients with Fanconi-Bickel syndrome may have increased renal clearance of glucose, amino acids, calcium, phosphate, bicarbonate, and uric acid due to proximal renal tubular dysfunction, impaired glucose and other monosaccharide utilization, and accumulation of glycogen in liver and kidney. Clinical presentations include rickets, short stature, failure to thrive, postprandial hyperglycemia, fasting hypoglycemia, hepatomegaly, hypophosphatemia, proximal tubular nephropathy, and metabolic acidosis.

MUSCLE GLYCOGEN STORAGE DISEASES

GSD Type II (Acid α-1,4 Glucosidase Deficiency, Pompe Disease) Pompe disease (OMIM #232300) is an autosomal recessive lysosomal and glycogen storage disorder caused by a deficiency of lysosomal acid α-glucosidase, an enzyme responsible for the degradation of glycogen to glucose in the lysosomes. This disease is characterized by the accumulation of glycogen in the lysosomes as opposed to primary accumulation in cytoplasm (as in the other GSDs).

At the most severe end of the disease spectrum is classic infantile-onset Pompe disease (IOPD) in which infants present with cardiomyopathy at birth or within the first few weeks of life. Patients develop a generalized muscle weakness with feeding difficulties, macroglossia, and congestive heart failure due to the rapidly progressive hypertrophic cardiomyopathy. Without treatment, patients with IOPD do not survive beyond 2 years of life. A variant form, known as nonclassic IOPD, also presents in the first year of life with less severe cardiomyopathy and slower disease progression. Patients with late-onset Pompe disease (LOPD) can present as early as the first few months of life to as late as the sixth decade. The distinguishing feature of LOPD from IOPD is the absence of cardiomyopathy in the first year of life. The clinical picture in LOPD is dominated by slowly progressive, predominantly proximal limb girdle muscle weakness. The pelvic girdle, paraspinal muscles, and diaphragm are most seriously affected. Respiratory symptoms include sleep apnea, sleep disordered breathing, decreased forced vital capacity, somnolence, morning headache, orthopnea, and exertional dyspnea. Respiratory failure causes significant morbidity and mortality in LOPD. In rare instances, patients present with respiratory insufficiency as the initial symptom. Young children with LOPD may have delayed motor milestones and difficulty in walking. With the advent of newborn screening for Pompe disease, delayed motor milestones and other kinematic/musculoskeletal findings such as scapular winging and weakness of pelvic girdle and core are being recognized as early as the first year of life in some babies with LOPD.

As the disease advances in LOPD, patients may require some form of ventilatory support and become dependent on a walking aid or wheelchair. Basilar artery aneurysms and dilation of the ascending aorta have been observed in some patients. Ptosis, lingual weakness, hypernasality, speech and articulation difficulties, gastrointestinal dysmotility, and bladder/bowel incontinence due to poor sphincter tone are now being recognized as part of the clinical spectrum. There is increasing recognition of central and peripheral nervous system involvement in IOPD with white matter hyperintensities, learning disabilities, slow processing speeds, sensorineural hearing loss, foot-slapping gait, bulbar weakness with dysarthria and dysphagia, and small-fiber neuropathy.

Laboratory findings include elevated levels of serum CK and aminotransferases as well as urine Glc_4, especially at the severe end of the disease spectrum, and they can be used as biomarkers to monitor disease progression and treatment responsiveness. In IOPD, chest x-ray shows massive cardiomegaly, echocardiogram shows severely elevated left ventricular mass index, and electrocardiographic findings include a high-voltage QRS complex and a shortened PR interval. Electromyography reveals myopathic features, with irritability of muscle fibers and pseudomyotonic discharges, which appears early in the paraspinal muscles. Muscle biopsy typically shows vacuoles that stain positive for glycogen, rupture of lysosomal membranes, and accumulation of glycogen in the cytoplasm from lysosomal spillover; the muscle acid phosphatase level is increased, presumably from a compensatory increase of lysosomal enzymes. Note that serum CK is not always elevated in LOPD, and depending on the muscle biopsied or tested, muscle histology or electromyography results may appear normal due to heterogeneity in muscle disease severity, and therefore, normal results do not rule out a diagnosis. There is an increasing role of muscle ultrasound and MRI as noninvasive biomarkers to assess extent and progression of disease involvement.

The confirmatory step for a diagnosis of Pompe disease is enzyme testing demonstrating deficient acid α-glucosidase (GAA), which can be measured in muscle, cultured skin fibroblasts, or blood; the latter is increasingly being used and is very reliable when performed in laboratories with experience. Enzyme levels when measured in skin or muscle are undetectable or very low in IOPD (typically <1%), whereas they are more variable in LOPD (typically 2–35%). In addition, gene sequencing can identify two pathogenic variants in the *GAA* gene and confirms the diagnosis.

The approval of enzyme replacement therapy (ERT) with alglucosidase alfa in 2006 has been lifesaving and has changed the clinical course of Pompe disease with substantial improvements in prognosis of IOPD and LOPD. Two next-generation ERT therapies have recently received approval by the U.S. Food and Drug Administration and the European Medicines Agency (avalglucosidase alfa and cipaglucosidase alfa with miglustat); both have improved targeting to the skeletal muscle. Patients with IOPD who make no endogenous GAA protein are classified as cross-reactive immunologic material (CRIM) negative. As a result of their CRIM-negative status, their immune system recognizes the recombinant GAA in ERT as a foreign protein and they typically develop high and sustained antibody titers to ERT that renders the treatment ineffective. Immunomodulation at the time of ERT initiation is used to induce immune tolerance in CRIM-negative patients. It is also used in some CRIM-positive patients who are considered at risk of mounting an immune response. Other adjunctive treatment options for Pompe disease include use of β_2 agonists, dietary modifications, submaximal aerobic exercise, and respiratory muscle strength training. Early diagnosis with early ERT initiation is the key to treatment efficacy. Gene replacement therapy, substrate reduction therapy (GYS1 inhibition), and therapies that can cross the blood brain barrier are being studied in early-phase clinical studies as additional treatment modalities.

Pompe disease is now part of the recommended uniform screening panel (RUSP) for newborns in the United States. In Taiwan where newborn screening for Pompe disease is performed routinely for all infants, early disease detection and treatment initiation have led to better treatment outcomes in patients with Pompe disease. Similar evidence is also emerging in the United States.

GSD Type V (Muscle Phosphorylase Deficiency, McArdle Disease) GSD type V (OMIM #232600) is an autosomal recessive disorder caused by deficiency of muscle phosphorylase. McArdle disease is a prototypical muscle-energy disorder, as the enzyme deficiency limits ATP generation by glycogenolysis and results in glycogen accumulation.

There can be a broad, heterogeneous spectrum of clinical presentations in the neonatal period with rapidly fatal at one extreme and the classical form (myalgia, cramps, and myoglobinuria) at the other. Symptom onset as late as the eighth decade has been reported. Patients typically develop muscle stiffness, pain, and weakness induced by exercise. The degree of muscle involvement is variable among the symptomatic patients; however, the exercise intolerance typically worsens over time due to inability to break down glycogen to glucose and thus a defect in the ability to convert glucose to lactate. Symptoms can be precipitated by (1) brief, high-intensity activity, such as sprinting or carrying heavy loads; and/or (2) less intense but sustained activity, such as climbing stairs or walking uphill. Most patients can engage in moderate exercise, such as walking on level ground, for extended periods. Patients often exhibit the "second wind" phenomenon in which they are able to continue the activity without pain after taking a short break from the initiation of strenuous physical effort; this phenomenon is unique to GSD V and is due to fatty acid oxidation becoming the main energy source, rather than glycogenolysis and anaerobic glycolysis, as exercise progresses. Burgundy-colored urine resulting from myoglobinuria secondary to rhabdomyolysis is reported after strenuous exercise. Acute renal failure and acute compartment syndrome can result from intense myoglobinuria after vigorous exercise. Although most patients experience episodic muscle pain and cramping because of exercise, 35% report permanent pain that seriously affects sleep and other activities. In rare cases, electromyographic findings may suggest inflammatory myopathy, a diagnosis that may be confused with polymyositis. Patients with GSD V are at risk for statin-induced myopathy and rhabdomyolysis as well as malignant hyperthermia warranting precautions for halogenated anesthetics and muscle relaxants.

At rest, the serum CK level is usually elevated; after exercise, the CK level increases even more. Exercise leads to an increase in levels of blood ammonia, inosine, hypoxanthine, and uric acid; these abnormalities reflect residues of accelerated muscle purine nucleotide recycling as a result of insufficient ATP production. Lack of increase in blood lactate suggests a defect in the conversion of glycogen/glucose to lactate. NADH is also underproduced during physical exertion. Historically, exaggerated blood ammonia elevations after an ischemic exercise test were indicative of a muscle GSD and suggested a defect in the conversion of glycogen or glucose to lactate. An ischemic exercise test is typically no longer used; instead, a noninvasive, nonischemic forearm exercise test has been developed and is easy to perform, has high sensitivity, and is cost-effective. However, the abnormal exercise response does not exclude other muscle GSDs and includes some risk. The cycle test detects the hallmark heart rate observed during the second-wind phenomenon.

A diagnostic confirmation is established by demonstration of pathogenic variants in the myophosphorylase gene (*PYGM*) or by enzymatic assay in muscle tissue. Clinically asymptomatic individuals with absent muscle phosphorylase activity have been identified due to elevated serum CK levels.

Treatment for GSD V consists of pre-exercise consumption of simple carbohydrates (e.g., sucrose or sports drinks) to protect muscles and improve exercise tolerance prior to the onset of the second wind. Regular exercise at moderate intensity is recommended to improve exercise capacity. Compared to patients who are physically inactive, those who are physically active are known to have improved cardiorespiratory fitness and a better long-term clinical course. Patients with GSD V are considered to have an increased risk of coronary artery disease. Additionally, poor bone health, reduced lean body mass, and increased body mass index (BMI) have been observed in inactive patients. Ocular involvement, including ptosis and retinal abnormalities, is becoming increasingly recognized.

GSD VII (Phosphofructokinase Deficiency, Tarui Disease) GSD type VII (Tarui disease; OMIM #232800) is an autosomal recessive disorder caused by phosphofructokinase M subunit deficiency, which results in impaired glycolysis in the skeletal muscle and red blood cells. Clinical presentations are similar to GSD V, with exercise intolerance, fatigue, myalgia, cramps, and increased CK and myoglobinuria with strenuous exercise, increasing risk for life-threatening complications including acute renal failure and compartment syndrome. However, there are some key differences that are important to recognize and consider when managing the disease. Unlike GSD V, an "out of wind" phenomenon occurs where exercise intolerance is particularly worse after meals rich in carbohydrates because increased glucose levels inhibit lipolysis and deprive skeletal muscle of fatty acid and ketones. Additionally, unlike in GSD V, in GSD VII, there is no spontaneous "second wind" phenomenon because of the inability of muscle to metabolize glucose efficiently. Therefore, consumption of simple carbohydrates prior to or during exercise can be detrimental and worsen exercise capability. Compensatory hemolysis and hyperuricemia are also observed in GSD VII. Gouty arthritis associated with hyperuricemia is more prominent in GSD VII compared to GSD V. Like GSD V, patients with GSD VII are at risk for statin-induced myopathy and rhabdomyolysis as well as malignant hyperthermia. Diagnosis is confirmed through genetic testing of the *PFKM* gene. Unlike in GSD V, muscle glycogen content is not markedly elevated in GSD VII. Treatment includes avoidance of strenuous exercise, compressive devices, and tourniquets. The use of IV dextrose in the clinical setting should be avoided. The use of dietary therapy remains unclear.

GSD Type IX (Muscle Phosphorylase Kinase Deficiency) GSD IX α1 (also known as GSD IXd; OMIM #300559) is an X-linked disorder due to muscle PhK deficiency caused by pathogenic variants in the *PHKA1*. There are limited cases of GSD IX α1 reported in the literature. Patients can present at any time from early childhood to adulthood with symptoms including exercise intolerance, fatigue, cramps, hypotonia, and progressive muscle weakness and atrophy. They are at risk for rhabdomyolysis. Electromyographic findings may be normal. Treatment for muscle PhK deficiency is symptomatic. Long-term natural history data are needed to better understand GSD IX α1.

GSD Type XV (Polyglucosan Body Myopathy 2) GSD type XV (OMIM #613507) is an autosomal recessive myopathy caused by variants in the *GYG1* gene impairing glycogenin-1 function. *GYG1* pathogenic variants result in a reduced or complete absence of glycogenin-1, which impacts its autoglycosylation and/or its interaction with glycogen synthase, resulting in impaired glycogen synthesis and formation of polyglucosan. Historically, polyglucosan body myopathy 2 (OMIM #616199) was termed for individuals with pathogenic variants in the *GYG1* gene who have progressive skeletal myopathy. Limited cases have been reported, and affected individuals commonly present with adult-onset proximal and distal muscle weakness predominantly affecting the hip and shoulder girdles. The disease course is often progressive, with the most disabling muscle weakness found at older age. Alternatively, GSD XV is often termed for individuals with pathogenic variants in the *GYG1* gene who have severe cardiac involvement without musculoskeletal manifestations. In these cases, cardiomyopathy, cardiac arrhythmia, and cardiac failure have necessitated cardiac transplantation. Further natural history data on GSD XV are needed. There is increasing awareness of GSD XV as a spectrum disorder with different degrees of both cardiac and skeletal muscle involvement.

Danon Disease Danon disease (OMIM #300257) is an X-linked lysosomal and glycogen storage disorder caused by pathogenic variants in the *LAMP2* gene. Deficiency of lysosomal-associated membrane protein 2 (LAMP2) results in defective autophagosomal-lysosomal fusion and excessive accumulation of autophagosomes in the heart and skeletal muscle. Patients with Danon disease present primarily with hypertrophic cardiomyopathy but can be distinguished by electrophysiologic abnormalities, particularly ventricular preexcitation and conduction defects. In Danon disease, Wolff-Parkinson-White (WPW) syndrome pattern is five times greater in prevalence than in idiopathic and familial hypertrophic cardiomyopathy. Therefore, in a young male with hypertrophic cardiomyopathy, the presence of WPW pattern on electrocardiogram strongly suggests Danon disease. The onset of cardiac symptoms in males typically occurs in childhood or adolescence and includes chest pain, palpitations, syncope, and cardiac arrest. Ocular manifestations are often underrecognized and include peripheral pigmentary retinopathy, lens changes, and abnormal electroretinograms. Elevated aminotransferase and CK levels and respiratory weakness have been observed due to underlying muscle involvement as well as sequelae of progressive cardiac disease. Mild learning disability and cognitive deficits have been noted, as well as speech and language delays, attention deficits, behavioral problems, and dysmetria. The prognosis for Danon disease is poor, with progressive end-stage heart failure in early adulthood. Female carriers can also be symptomatic; although the disease is less severe compared to males, cardiomyopathy, skeletal myopathy, retinopathy, and cognitive dysfunction have been described. Treatment is symptomatic and involves management of heart failure, correction of conduction abnormalities, and physical therapy, among others. Heart transplantation can be considered for cases with rapid progression or end-stage heart failure. Neuropsychological evaluations and special education support may be required for those with intellectual disabilities. Gene replacement therapy is under early-phase clinical studies.

PRKAG2 Syndrome AMP-activated protein kinase gamma 2 (PRKAG2; OMIM #261740) deficiency, known as PRKAG2 syndrome, is an autosomal dominant disorder caused by pathogenic variants in the *PRKAG2* gene. PRKAG2 is important in many cellular ATP metabolic pathways, and affected individuals present with cardiac abnormalities including hypertrophic cardiomyopathy and conduction system abnormalities, particularly WPW syndrome. The extent of cardiac involvement in PRKAG2 syndrome is variable and includes supraventricular tachycardia, sinus bradycardia, left ventricular dysfunction, or even sudden cardiac death in some cases. In addition to cardiac involvement, there is a broad spectrum of phenotypic presentations including myopathy, myalgia, and seizures. Other manifestations include developmental delays, hypotonia, areflexia, tremors, feeding difficulties, frequent respiratory infections, and failure to thrive. Unlike Danon disease, cardiomyopathy due to *PRKAG2* pathogenic variants is compatible with longer-term survival except for a congenital form that presents in early infancy with a rapid fatal course due to massive glycogen deposition in the myocardium. PRKAG2 syndrome should be considered as a differential diagnosis in infants presenting with severe hypertrophic cardiomyopathy. Cardiac biopsy may reveal the presence of glycogen and polyglucosan in vacuoles in cardiomyocytes. In rare instances, patients with PRKAG2 syndrome may be misdiagnosed as having IOPD due to phenotypical similarity. Treatment is symptomatic, and heart transplantation is indicated for those in end-stage heart failure.

Other Muscle GSDs GSD IIIa, IV, and IX β have muscle involvement as mentioned in each respective subsection under "Liver Glycogen Storage Diseases." GSD 0b is a rare disorder caused by deficiency in muscle glycogen synthase (GYS1; OMIM #138570). The skeletal and cardiac muscles are affected. Only a handful of cases with GSD 0b have been described, and clinical presentations include sudden cardiac death in middle childhood. Additional enzyme defects produce muscle glycogenosis with muscle energy impairment, namely deficiencies in phosphoglycerate mutase (GSD X; OMIM #261670), lactate dehydrogenase A (GSD XI-lactate dehydrogenase A deficiency; OMIM #612933), aldolase A (GSD XII; OMIM #611881), and β-enolase (GSD XIII; OMIM #612932). Clinical presentations include myopathy and risk for rhabdomyolysis similar to GSD V and VII. Deficiency of aldolase A is also associated with hemolytic anemia.

DISORDERS OF GALACTOSE METABOLISM

Galactose is found in high concentrations in foods that contain lactose, such as dairy products. The term *galactosemia* refers to four distinct inborn errors of galactose metabolism; however, *galactosemia* has historically been designated galactose-1-phosphate uridyltransferase (GALT) deficiency (galactosemia type I) because it is the most prevalent form.

■ GALACTOSEMIA TYPE I (GALACTOSE-1-PHOSPHATE URIDYLTRANSFERASE DEFICIENCY)

GALT deficiency (OMIM #230400) encompasses three subtypes: classical, clinical variant, and biochemical variant galactosemia. *Classic galactosemia* is caused by GALT deficiency with a GALT enzyme activity that is absent or barely detectable. It is a severe disease with an incidence of 1 in 16,000 to 60,000 and an early onset of symptoms. A newborn infant typically receives up to 40% of total daily energy intake as lactose (glucose plus galactose) through breastmilk or milk-based formulas. Without the functional GALT enzyme, the infant is unable to metabolize galactose-1-phosphate (Fig. 430-1), which consequently accumulates and results in injury to the kidneys, liver, central nervous system, eyes, and, in the long term, the ovaries. After the first feeding, infants can present with vomiting, diarrhea, jaundice, hepatomegaly, aminoaciduria, lethargy, and feeding difficulties. There is an increased risk for susceptibility to infection with gram-negative organisms, such as *Escherichia coli*, often with the onset of sepsis preceding the diagnosis of galactosemia. Additional findings may include convulsions/seizures, failure to thrive, cataracts, bleeding diathesis, hypoglycemia, and cerebral edema. Widespread newborn screening (NBS) for galactosemia has identified these infants early and permitted prompt placement on dietary restriction. Elimination of galactose from the diet reverses growth failure as well as renal and hepatic dysfunction, improving the prognosis. However, even with early dietary treatment, long-term developmental outcomes in classic galactosemia are poor, with most patients having motor speech disorders and learning disabilities that may increase in severity with age. The development of cataracts, tremors, and low bone density is common in adults. Impaired motor function and balance (with or without overt ataxia) is frequently seen. Ovarian dysfunction manifesting as primary or secondary amenorrhea is seen in females, with 80–90% or more of women reporting hypergonadotropic hypogonadism. While most female patients are subfertile due to primary ovarian insufficiency, a few successful pregnancies have been reported. The treatment of galactosemia to prevent long-term complications remains a challenge.

Genotype and phenotype relationship is well established in galactosemia with the Q188R mutation in homozygosity causing classic galactosemia. The p.Ser135Leu variant causes *clinical variant galactosemia*, which is more common in the African American and native South African populations and is associated with a less severe disease course due to residual GALT activity (typically absent or barely detectable in erythrocytes but ~10% in liver and intestinal tissue). Elimination of galactose from the diet is recommended for clinical variant galactosemia. African Americans with clinical variant galactosemia (homozygous p.Ser135Leu variants) and adequate early treatment do not typically develop the long-term complications observed in classic galactosemia, including primary ovarian insufficiency. The *Duarte variant* (c.940A>G;-16_119delGTCA), also known as the *D variant* or *D_2 variant*, is considered to be protective and causes the *biochemical variant galactosemia*, which is asymptomatic despite reduced GALT enzyme activity (typically ~25%) and elevated galactose-1-phosphate levels in red blood cells. Dietary therapy for the biochemical variant galactosemia is no longer recommended.

GALACTOSEMIA TYPE II (GALACTOKINASE DEFICIENCY)

Galactokinase (GALK) deficiency (galactosemia type II; OMIM #230200) has an estimated prevalence of 1 in 50,000 to 2,200,000. A higher incidence is observed in individuals of Romanian origin (estimated carrier rate 1 in 47). Unlike the multisystemic symptoms associated with galactosemia type I, the only consistently reported abnormalities in GALK deficiency are cataracts without visual impairment and, rarely, pseudotumor cerebri. Other findings include hyperbilirubinemia, bleeding diathesis, and elevated aminotransferases. There have been reports of intellectual impairments, which can be severe in some cases. Since many NBS programs measure GALT activity and not galactose levels directly, diagnosing GALK deficiency often entails ruling out GALT deficiency. Elimination of galactose in the diet is recommended.

GALACTOSEMIA TYPE III (URIDINE DIPHOSPHATE GALACTOSE-4-EPIMERASE DEFICIENCY)

Deficiency of uridine diphosphate galactose-4-epimerase (GALE) causes galactosemia type III (OMIM #606953), which is considered to be rare but more common than GALK deficiency. The benign form (also known as peripheral GALE deficiency) is incidentally discovered through NBS; the enzyme deficiency is limited to white and red blood cells, and no dietary treatment is required. The intermediate form is associated with reduced GALE activity in white and red blood cells as well as in other tissue cells. Neonates with the intermediate form may be asymptomatic despite consuming galactose-containing foods. A more severe form has also been recognized and is due to generalized GALE deficiency in all tested tissues. Clinical manifestations resemble GALT deficiency with cataracts and damage to the liver, kidney, and brain. Hypotonia and sensorineural hearing loss have also been observed. Dietary restriction—not elimination—of galactose is indicated for both the intermediate and severe forms. Galactose is an essential component of many structural proteins in the nervous system, and patients with GALE deficiency cannot synthesize UDP-galactose from glucose; therefore, complete elimination of galactose is not recommended.

GALACTOSEMIA TYPE IV (GALACTOSE MUTAROTASE DEFICIENCY)

Galactosemia type IV (OMIM #618881) is a newly recognized autosomal recessive conditionm and there are few documented cases. Deficiency of galactose mutarotase (GALM) was incidentally identified through NBS in Japan. GALM deficiency is considered to be mild, and cataracts are the only reported persistent symptom.

DISORDERS OF FRUCTOSE METABOLISM

Fructose naturally occurs in high concentrations in fruits and foods containing sucrose. Two inborn errors of the metabolism are associated with the fructose metabolism pathway: essential/benign fructosuria and hereditary fructose intolerance as a part of glycolysis. Fructolysis occurs predominantly in the liver.

ESSENTIAL/BENIGN FRUCTOSURIA

Essential/benign fructosuria (OMIM #229800) is an autosomal recessive condition affecting the *KHK* gene causing deficiency of ketohexokinase, commonly referred to as fructokinase (Fig. 430-1). It is a benign condition incidentally diagnosed from the presence of fructose as a reducing substance in the urine. No treatment is necessary.

HEREDITARY FRUCTOSE INTOLERANCE

Deficiency of fructose-1,6-bisphosphate aldolase (aldolase B; hereditary fructose intolerance; OMIM #229600) is a severe disease in infants. These patients are healthy and asymptomatic until fructose is ingested, usually from fruit, fruit juice, sweetened cereal, or sucrose-containing formula. Aldolase B deficiency causes fructose-1-phosphate accumulation and ATP depletion, which inhibits gluconeogenesis and glycogenolysis and can lead to a rapid drop in blood glucose shortly after consumption of fructose or related sugars. Clinical manifestations resemble galactosemia and include jaundice, hepatomegaly, vomiting, lethargy, irritability, and convulsions. Ingestion of large amounts of fructose can result in progressive coma. Laboratory findings show prolonged clotting time, hypoalbuminemia, elevation of bilirubin and aminotransferase levels, and proximal renal tubular dysfunction. If the disease goes undiagnosed and the intake of fructose continues, postprandial hypoglycemic episodes recur, and eventually death can occur from progressive liver and renal failure. Hepatomegaly, liver fibrosis, Fanconi-like syndrome, and nephrocalcinosis have been observed. The mainstay of treatment is the elimination of all sources of sucrose, fructose, and sorbitol from the diet and medications. Older patients tend to develop an aversion to fructose-containing foods. With early and consistent dietary treatment, liver and kidney dysfunction improve and catch-up growth is common. Hepatic steatosis has been observed even in treated individuals. The long-term prognosis of treated individuals is good. The incidence of celiac disease is higher among patients with hereditary fructose intolerance (>10%) than in the general population (1–3%).

GLOBAL CONSIDERATIONS

GSDs and other inherited disorders of carbohydrate metabolism, although individually rare, are reported in most ethnic populations. The prevalent genetic variants for each disease may vary in different ethnic populations, but clinical symptoms are remarkably similar, and treatment guidelines apply to all. Symptomatic treatment is available for these disorders, and today, advances in the field including newborn screening have resulted in more definitive diagnosis and better treatment approaches. There are many promising therapies on the horizon with several ongoing clinical trials, including those investigating the use of ERT, mRNA therapy, gene replacement therapy, gene editing, and substrate reduction therapy. In the past, prognosis for many disorders of carbohydrate metabolism was guarded, but with early diagnosis and better management, survival rates have improved and many affected children are surviving into adulthood.

■ FURTHER READING

Fernandes SA et al: Benign or not benign? Deep phenotyping of liver glycogen storage disease IX. Mol Genet Metab 131:299, 2020.

Heinemann JB et al: Features and outcome of galactokinase deficiency in children diagnosed by newborn screening. J Inherit Metab Dis 34:399, 2011.

Hong KN et al: International consensus on differential diagnosis and management of patients with Danon disease: JACC state-of-the-art review. J Am Coll Cardiol 82:1628, 2023.

Grünert SC et al: Improved inflammatory bowel disease, wound healing and normal oxidative burst under treatment with empagliflozin in glycogen storage disease type Ib. Orphanet J Rare Dis 15:218, 2020.

Hannah WB et al: Glycogen storage diseases. Nat Rev Dis Primers 9:46, 2023.

Hedberg-Oldfors C et al: Cardiomyopathy as presenting sign of glycogenin-1 deficiency—report of three cases and review of the literature. J Inherit Metab Dis 40:139, 2017.

Herbert M et al: Role of continuous glucose monitoring in the management of glycogen storage disorders. J Inherit Metab Dis 41:917, 2018.

Katler QS et al: A multinational study of acute and long-term outcomes of Type 1 galactosemia patients who carry the S135L (c.404C > T) variant of GALT. J Inherit Metab Dis 45:1106, 2022.

Kiely et al: A novel approach to characterize phenotypic variation in GSD IV: Reconceptualizing the clinical continuum. Front Genet 13:992406, 2022.

Koch RL et al: Natural history study of hepatic glycogen storage disease type IV and comparison to Gbe1ys/ys model. JCI Insight 9:e177722, 2024.

Kronn DF et al: Management of confirmed newborn-screened patients with Pompe disease across the disease spectrum. Pediatrics 40:S24, 2017.

Li N et al: Clinical and molecular characterization of patients with fructose 1,6-bisphosphatase deficiency. Int J Mol Sci 18:857, 2017.

Musumeci O et al: Recurrent rhabdomyolysis due to muscle β-enolase deficiency: Very rare or underestimated? J Neurol 261:2424, 2014.

Papadopoulos C et al: Aldolase A deficiency: Report of new cases and literature review. Mol Genet Metab Rep 23:100730, 2021.

Porto AG et al: Clinical spectrum of PRKAG2 syndrome. Circ Arrhythm Electrophysiol 9:e003121, 2016.

Quinlivan R et al: Pharmacological and nutritional treatment for McArdle disease (glycogen storage disease type V). Cochrane Database Syst Rev 2014:CD003458, 2014.

Rubio-Gozalbo ME et al: The natural history of classic galactosemia: Lessons from the GalNet registry. Orphanet J Rare Dis 14:86, 2019.

Steinmann B et al: Disorders of fructose metabolism. *The Online Metabolic and Molecular Bases of Inherited Disease*. New York, McGraw-Hill, 2013.

Timson DJ: The structural and molecular biology of type III galactosemia. IUBMB Life 58:83, 2006.

Welling L et al: International clinical guideline for the management of classical galactosemia: Diagnosis, treatment, and follow-up. J Inherit Metab Dis 40:171, 2017.

431

Inherited Disorders of Amino Acid Metabolism in Adults

Nicola Longo

Amino acids are the building blocks of proteins and serve as neurotransmitters (glycine, glutamate, γ-aminobutyric acid) or as precursors of hormones, coenzymes, pigments, purines, or pyrimidines. Eight amino acids, referred to as *essential* (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine), cannot be synthesized by humans and must be obtained from dietary sources. The others can be formed endogenously. Each amino acid has a unique degradative pathway by which its nitrogen and carbon components are used for the synthesis of other amino acids, carbohydrates, and lipids. Disorders of amino acid metabolism and transport **(Chap. 432)** are individually rare—the incidences range from 1 in 10,000 for cystinuria or phenylketonuria to 1 in 200,000 for homocystinuria or alkaptonuria—but collectively, they affect perhaps 1 in 4000 newborns. Almost all are transmitted as autosomal recessive traits.

The features of inherited disorders of amino acid catabolism are summarized in **Table 431-1**. In general, these disorders are named for the compound that accumulates to highest concentration in blood (*-emias*) or urine (*-urias*). In the aminoacidopathies, the parent amino acid is found in excess, whereas products in the catabolic pathway accumulate in organic acidemias. Which compound(s) accumulates depends on the site of the enzymatic block, the reversibility of the reactions proximal to the lesion, and the availability of alternative pathways of metabolic "runoff." Biochemical and genetic heterogeneity are common. Six distinct forms of hyperphenylalaninemia and nine forms of homocystinuria (with or without methylmalonic acidemia) are recognized. Such heterogeneity reflects the complexity of amino acid metabolism requiring multiple enzymes (gene products) for proper functioning.

The manifestations of these conditions differ widely (Table 431-1). Some, such as sarcosinemia, produce no clinical consequences. At the other extreme, complete deficiency of ornithine transcarbamylase is lethal in the untreated neonate. Central nervous system (CNS) dysfunction, in the form of delays in development/intellectual disability, seizures, or behavioral disturbances, is present in more than half the disorders. Protein-induced vomiting, neurologic dysfunction, and hyperammonemia occur in many disorders of the urea cycle. Metabolic ketoacidosis, often accompanied by hyperammonemia, is frequent in organic acidemias. Some disorders produce focal tissue or organ involvement such as liver disease, renal failure, cutaneous abnormalities, or ocular lesions.

Defects in the synthesis of nonessential amino acids (asparagine, glutamine, proline, serine) involve predominantly the brain with neurologic symptoms, with other organs occasionally affected. Dominant mutations in at least one of these genes can cause tremor or spastic paraplegia in adults.

The analysis of plasma amino acids (by ion-exchange chromatography or liquid chromatography/tandem mass spectrometry), urine organic acids (by gas chromatography/mass spectrometry), and plasma acylcarnitine profile (by tandem mass spectrometry) is commonly used to diagnose and monitor most of these disorders. The diagnosis is confirmed by enzyme assay on cells or tissues from the patients or, more commonly, by DNA testing. The clinical manifestations in many of these conditions can be prevented or mitigated if a diagnosis is achieved early and appropriate treatment (e.g., dietary protein or amino acid restriction or vitamin supplementation) is instituted promptly. For this reason, newborn screening programs seek to identify several of these disorders. Infants with a positive screening test need additional metabolic testing (usually suggested by the newborn screening program) to confirm or exclude the diagnosis. Confirmed

TABLE 431-1 Inherited Disorders of Amino Acid Metabolism

AMINO ACID(S)	CONDITION	ENZYME DEFECT	CLINICAL FINDINGS	INHERITANCE
Phenylalanine	Phenylketonuria	Phenylalanine hydroxylase	Intellectual disability, microcephaly, hypopigmented skin and hairs, eczema, “mousy” odor	AR
	DHPR deficiency	Dihydropteridine reductase	Intellectual disability, hypotonia, spasticity, myoclonus	AR
	PTPS deficiency	6-Pyruvoyl-tetrahydropterin synthase	Dystonia, neurologic deterioration, seizures, intellectual disability	AR
	GTP cyclohydrolase 1 deficiency	GTP cyclohydrolase 1	Intellectual disability, seizures, dystonia, temperature instability	AR
	Carbinolamine dehydratase deficiency	Pterin-4α-carbinolamine dehydratase	Transient hyperphenylalaninemia (benign)	AR
	DNAJC12 deficiency	Hydroxylase co-chaperone	Dystonia, parkinsonism, intellectual disability	AR
Tyrosine	Tyrosinemia type 1 (hepatorenal)	Fumarylacetoacetate hydrolase	Liver failure, cirrhosis, rickets, failure to thrive, peripheral neuropathy, “boiled cabbage” odor	AR
	Tyrosinemia type 2 (oculocutaneous)	Tyrosine transaminase	Palmoplantar keratosis, painful corneal erosions with photophobia, learning disability	AR
	Tyrosinemia type 3	4-Hydroxyphenylpyruvate dioxygenase	Hypertyrosinemia with normal liver function, occasional mental delay	AR
	Hawkinsinuria	4-Hydroxyphenylpyruvate dioxygenase	Transient failure to thrive, metabolic acidosis in infancy	AD
	Alkaptonuria	Homogentisic acid oxidase	Ochronosis, arthritis, cardiac valve involvement, coronary artery calcification	AR
	Maleylacetoacetate isomerase deficiency	Maleylacetoacetate isomerase	No clinical symptoms, elevated succinylacetone in blood and urine	AR
	Albinism (oculocutaneous)	Tyrosinase	Hypopigmentation of hair, skin, and optic fundus; visual loss; photophobia	AR
	Albinism (ocular)	Different enzymes or transporters	Hypopigmentation of optic fundus, visual loss	AR, XL
	DOPA-responsive dystonia	Tyrosine hydroxylase	Rigidity, truncal hypotonia, tremor, intellectual disability	AR
GABA	4-Hydroxybutyric aciduria	Succinic semialdehyde dehydrogenase	Seizures, intellectual disability, hypotonia	AR
	ABAT deficiency	GABA transaminase	Seizures, intellectual disability, hypotonia	AR
Tryptophan	Hydroxykynureninuria	Kynureninase	Intellectual disability, spasticity	AR
Histidine	Histidinemia	Histidine-ammonia lyase	Benign	AR
	Urocanic aciduria	Urocanase	Occasional intellectual disability	AR
	Formiminoglutamic aciduria	Formiminotransferase	Occasional intellectual disability	AR
Glycine	Glycine encephalopathy	Glycine cleavage (4 enzymes)	Infantile seizures, lethargy, apnea, profound intellectual disability	AR
	Sarcosinemia	Sarcosine dehydrogenase	Benign	AR
	Hyperoxaluria type I	Alanine:glyoxylate aminotransferase	Calcium oxalate nephrolithiasis, renal failure	AR
	Hyperoxaluria type II	D-Glyceric acid dehydrogenase/glyoxylate reductase	Calcium oxalate nephrolithiasis, renal failure	AR
Serine	3-PGDH deficiency	Phosphoglycerate dehydrogenase	Seizures, microcephaly, intellectual disability	AR
	PSAT1 deficiency	Phosphoserine aminotransferase	Seizures, microcephaly, intellectual disability	AR
	PSP deficiency	Phosphoserine phosphatase	Seizures, microcephaly, intellectual disability	AR
Proline	Hyperprolinemia type 1	Proline oxidase	Benign	AR
	Hyperprolinemia type 2	Δ^1-Pyrroline-5-carboxylate dehydrogenase	Febrile seizures, intellectual disability	AR
	Hyperhydroxyprolinemia	Hydroxyproline oxidase	Benign	AR
	Prolidase deficiency	Prolidase	Mild intellectual disability, chronic dermatitis, autoimmunity	AR
	PYCR1 deficiency	Pyrroline-5-carboxylate reductase 1	Wrinkly skin, joint laxity, typical facial features, intellectual disability, osteopenia, intrauterine growth retardation, hypotonia	AR
	PYCR2 deficiency	Pyrroline-5-carboxylate reductase 2	Microcephaly, hypomyelination, and reduced cerebral white matter volume, failure to thrive, intellectual disability, movement disorders, seizures	AR
Proline (ornithine, arginine, citrulline)	Δ^1-Pyrroline-5-carboxylate synthase deficiency	Δ^1-Pyrroline-5-carboxylate synthase	Hypotonia, seizures, neurodegeneration, peripheral neuropathy, joint laxity, skin hyperelasticity, subcapsular cataracts, hyperammonemia, adult spastic paraparesis (AD)	AR, AD
Methionine	Hypermethioninemia	Methionine adenosyltransferase	Usually benign	AR
	S-Adenosylhomocysteine hydrolase deficiency	*S*-Adenosylhomocysteine hydrolase	Hypotonia, intellectual disability, absent tendon reflexes, delayed myelination	AR
	Glycine *N*-methyltransferase deficiency	Glycine *N*-methyltransferase	Elevated liver transaminases	AR
	Adenosine kinase deficiency	Adenosine kinase	Intellectual disability, seizures, liver dysfunction	AR

(*Continued*)

TABLE 431-1 Inherited Disorders of Amino Acid Metabolism (Continued)

AMINO ACID(S)	CONDITION	ENZYME DEFECT	CLINICAL FINDINGS	INHERITANCE
Homocysteine	Homocystinuria	Cystathionine β-synthase	Lens dislocation, thrombotic vascular disease, intellectual disability, osteoporosis	AR
	Homocystinuria	5,10-Methylenetetrahydrofolate reductase	Intellectual disability, gait and psychiatric abnormalities, recurrent strokes	AR
	Homocystinuria	Methionine synthase and Methionine synthase reductase (cblE, G)	Intellectual disability, hypotonia, seizures, megaloblastic anemia	AR
	Homocystinuria and methylmalonic acidemia	Vitamin B_{12} lysosomal efflux and metabolism (cblC, -epiC, -D, -F, -J, -X)	Intellectual disability, lethargy, failure to thrive, hypotonia, seizures, megaloblastic anemia	AR, XL
Cystathionine	Cystathioninuria	β-Cystathioninase	Benign	AR
Cysteine	Sulfocystinuria	Sulfite oxidase or molybdenum cofactor deficiency	Seizures, intellectual disability, dislocated lenses	AR
Lysine	Hyperlysinemia, saccharopinuria	α-Aminoadipic semialdehyde synthase	Benign	AR
	Pyridoxine-dependent seizures	L-Δ^1-Piperideine-6-carboxilate dehydrogenase	Seizures, intellectual disability	AR
Lysine, tryptophan	α-Ketoadipic acidemia	α-Ketoadipic acid dehydrogenase DHTKD1	Benign	AR
Lysine, tryptophan	Glutaric acidemia type 1	Glutaryl-CoA dehydrogenase	Progressive severe dystonia and athetosis, motor delays	AR
Ornithine	Gyrate atrophy of the choroid and retina	Ornithine-Δ-aminotransferase	Myopia, night blindness, loss of peripheral vision, cataracts, chorioretinal degeneration	AR
Urea cycle	Carbamoylphosphate synthase-1 deficiency	Carbamoylphosphate synthase-1	Lethargy progressing to coma, protein aversion, intellectual disability, hyperammonemia	AR
	N-Acetylglutamate synthase deficiency	*N*-Acetylglutamate synthase	Lethargy progressing to coma, protein aversion, intellectual disability, hyperammonemia	AR
	Ornithine transcarbamylase deficiency	Ornithine transcarbamylase	Lethargy progressing to coma, protein aversion, intellectual disability, hyperammonemia	XL
	Citrullinemia type 1	Argininosuccinate synthase	Lethargy progressing to coma, protein aversion, intellectual disability, hyperammonemia, liver failure	AR
	Argininosuccinic acidemia	Argininosuccinate lyase	Lethargy progressing to coma, protein aversion, intellectual disability, hyperammonemia, trichorrhexis nodosa, liver failure	AR
	Arginase deficiency	Arginase	Spastic tetraparesis, microcephaly, intellectual disability, mild hyperammonemia	AR
	Hyperornithinemia, hyperammonemia, homocitrullinuria	Mitochondrial ornithine carrier ORNT1	Vomiting, lethargy, failure to thrive, intellectual disability, episodic confusion, hyperammonemia, protein intolerance	AR
	Citrullinemia type 2	Mitochondrial aspartate/glutamate carrier CTLN2	Neonatal intrahepatic cholestasis, adult presentation with sudden behavioral changes and stupor, coma, hyperammonemia, liver failure	AR
Glutamine	Glutamine synthetase deficiency	Glutamine synthase	Brain malformations, pachygyria, seizures, hypotonia, intellectual disability, dysmorphic features, low glutamine	AR
	Glutaminase deficiency	Glutaminase	Epileptic encephalopathy, intellectual disability, ataxia, elevated glutamine	AR
Asparagine	Asparagine synthetase deficiency	Asparagine synthase	Epileptic encephalopathy, seizures, microcephaly, simplified gyration pattern, hypotonia, tetraplegia, intellectual disability	
Valine	Isobutyryl-CoA dehydrogenase deficiency	Isobutyryl-CoA dehydrogenase	Benign	AR
Isoleucine, leucine, valine	Maple syrup urine disease	Branched chain ketoacid dehydrogenase (E1α, E1β, E2, E3 deficiency)	Lethargy, vomiting, encephalopathy, seizures, intellectual disability, "maple syrup" odor, protein intolerance	AR
Isoleucine, leucine, valine	Hypervalinemia	Branched-chain amino acid transferase 2 (BCAT2)	Autism, headaches, intellectual disability	AR
Isoleucine, leucine, valine	Branched-chain amino acid deficiency	Branched chain ketoacid dehydrogenase kinase (BCHDK)	Autism, epilepsy, intellectual disability, microcephaly	
Leucine	Isovaleric acidemia	Isovaleryl-CoA dehydrogenase	Acidosis, ketosis, vomiting, coma, hyperammonemia, "sweaty feet" odor, protein intolerance	AR
	3-Methylcrotonyl glycinuria	3-Methylcrotonyl-CoA carboxylase	Stress-induced metabolic acidosis, hypotonia, hypoglycemia, "cat's urine" odor	AR
	3-Methylglutaconic aciduria type I	3-Methylglutaconyl-CoA hydratase deficiency	Stress-induced acidosis, leukodystrophy, hypotonia, hepatomegaly	AR
	3-Hydroxy-3-methylglutaric aciduria	3-Hydroxy-3-methylglutaryl-CoA lyase	Stress-induced hypoketotic hypoglycemia and acidosis, encephalopathy, hyperammonemia	AR

(*Continued*)

TABLE 431-1 Inherited Disorders of Amino Acid Metabolism (Continued)

AMINO ACID(S)	CONDITION	ENZYME DEFECT	CLINICAL FINDINGS	INHERITANCE
Isoleucine	2-Methylbutyryl-glycinuria	2-Methylbutyryl-CoA dehydrogenase	Benign	AR
	2-Methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency	2-Methyl-3-hydroxybutyryl-CoA dehydrogenase	Developmental regression, seizures, and rigidity sometimes triggered by illnesses	XL
	3-Oxothiolase deficiency	3-Oxothiolase	Fasting-induced acidosis and ketosis, vomiting, lethargy	AR
Isoleucine, methionine, threonine, valine	Propionic acidemia (pccA, -B)	Propionyl-CoA carboxylase	Metabolic ketoacidosis, hyperammonemia, hypotonia, lethargy, coma, protein intolerance, intellectual disability, hyperglycinemia	AR
	Multiple carboxylase/ biotinidase deficiency	Holocarboxylase synthase or biotinidase	Metabolic ketoacidosis, diffuse rash, alopecia, seizures, intellectual disability	AR
	Methylmalonic acidemia (mutase, cblA, B, racemase)	Methylmalonyl-CoA mutase/ racemase or cobalamin reductase/ adenosyltransferase	Metabolic ketoacidosis, hyperammonemia, hypertonia, lethargy, coma, protein intolerance, intellectual disability, hyperglycinemia	AR

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; Cbl, cobalamin; DOPA, dihydroxyphenylalanine; GABA, γ-aminobutyric acid; GTP, guanosine 5′-triphosphate; XL, X-linked.

cases should be referred to a metabolic center for initiation of therapy. The parents need to be counseled about the natural history of the disease and its recurrence risk in future pregnancies. In some cases, parents need testing because they might have a disorder themselves (such as glutaric acidemia type 1, methylcrotonyl coenzyme A carboxylase deficiency, primary carnitine deficiency, or fatty acid oxidation defects) since mothers with these conditions can sometimes be identified by abnormal newborn screening results in their offspring. Some metabolic disorders can remain asymptomatic until adult age, presenting only when fasting or severe stress requires full activity of affected metabolic pathways to provide energy.

Selected disorders that illustrate the principles, properties, and problems presented by the disorders of amino acid metabolism are discussed in this chapter.

THE HYPERPHENYLALANINEMIAS

The hyperphenylalaninemias (Table 431-1) result from impaired conversion of phenylalanine to tyrosine. The most common and clinically important is *phenylketonuria* (frequency 1:16,500), which is an autosomal recessive disorder characterized by an increased concentration of phenylalanine and its by-products in body fluids and by severe intellectual disability if untreated in infancy. It results from reduced activity of phenylalanine hydroxylase. The accumulation of phenylalanine inhibits the transport of other amino acids required for protein or neurotransmitter synthesis, reduces synthesis and increases degradation of myelin, and leads to inadequate formation of norepinephrine and serotonin. Phenylalanine is a competitive inhibitor of tyrosinase, a key enzyme in the pathway of melanin synthesis, resulting in hypopigmentation of hair and skin. Untreated children with classic phenylketonuria are normal at birth but fail to attain early developmental milestones, develop microcephaly, and demonstrate progressive impairment of cerebral function. Hyperactivity, seizures, and severe intellectual disability are major clinical problems later in life. Electroencephalographic abnormalities; "mousy" odor of skin, hair, and urine (due to phenylacetate accumulation); and a tendency to develop hypopigmentation (compared to the family background) and eczema complete the devastating clinical picture. In contrast, affected children who are detected and treated at birth show none of these abnormalities.

TREATMENT

Phenylketonuria

To prevent intellectual disability, diagnosis and initiation of dietary treatment of classic phenylketonuria must occur before the child is 2 weeks of age. For this reason, newborns in North America, Australia, and Europe are screened by determinations of blood phenylalanine levels. Abnormal values are confirmed using quantitative analysis of plasma amino acids. Dietary phenylalanine restriction is usually instituted if blood phenylalanine levels are >360 μmol/L. Treatment consists of a special diet low in phenylalanine and supplemented with tyrosine since tyrosine becomes an essential amino acid in phenylalanine hydroxylase deficiency. With therapy, plasma phenylalanine concentrations should be maintained between 120 and 360 μmol/L for life. Compliance with the strict diet is often difficult as patients become older; increased levels of phenylalanine in adults can cause deficits in executive function or psychiatric symptoms. Oral tetrahydrobiopterin (5–20 mg/kg per d), an essential cofactor of phenylalanine hydroxylase, can reduce phenylalanine levels in some patients with phenylketonuria in conjunction with a low-protein diet. Pegvaliase is a pegylated form of phenylalanine ammonia lyase, a bacterial enzyme that converts phenylalanine to *trans*-cinnamic acid and ammonia. This injectable drug can substantially reduce phenylalanine levels, allowing a normal diet. The bacterial origin of pegvaliase can cause immune reactions that limit its use in some patients with phenylketonuria.

Women with phenylketonuria can become pregnant. If maternal phenylalanine levels are not strictly controlled before and during pregnancy, their offspring are at increased risk for congenital defects and microcephaly (*maternal phenylketonuria*). After birth, these children have severe intellectual disability and growth retardation. Pregnancy risks can be minimized by continuing lifelong phenylalanine-restricted diets and assuring strict phenylalanine restriction 2 months prior to conception and throughout gestation.

■ THE HOMOCYSTINURIAS (HYPERHOMOCYSTEINEMIAS)

The homocystinurias include 10 biochemically and clinically distinct disorders (Table 431-1) characterized by increased concentration of the sulfur-containing amino acid homocysteine in blood and urine.

Classic homocystinuria, the most common (frequency 1:450,000), results from reduced activity of cystathionine β-synthase **(Fig. 431-1)**, the pyridoxal phosphate–dependent enzyme that condenses homocysteine with serine to form cystathionine. Most patients present between 3 and 5 years of age with dislocated optic lenses and intellectual disability (in about half of cases). Some patients develop a marfanoid habitus and radiologic evidence of osteoporosis.

Life-threatening vascular complications (affecting coronary, renal, and cerebral arteries) can occur during the first decade of life and are the major cause of morbidity and mortality. Classic homocystinuria can be diagnosed with analysis of plasma amino acids, showing elevated methionine and presence of free homocystine. Total plasma homocysteine is also extremely elevated (usually >100 μ*M*). Elevated levels of methionine can be also detected by neonatal screening, but milder variants can be missed by this approach. Treatment consists of a special diet restricted in protein and methionine. In approximately half of patients, oral pyridoxine (25–500 mg/d) produces a fall in plasma methionine and homocysteine concentration in body fluids. Folate and vitamin B_{12} deficiency should be prevented by adequate supplementation. Betaine is also effective in reducing homocysteine levels by favoring its remethylation to methionine.

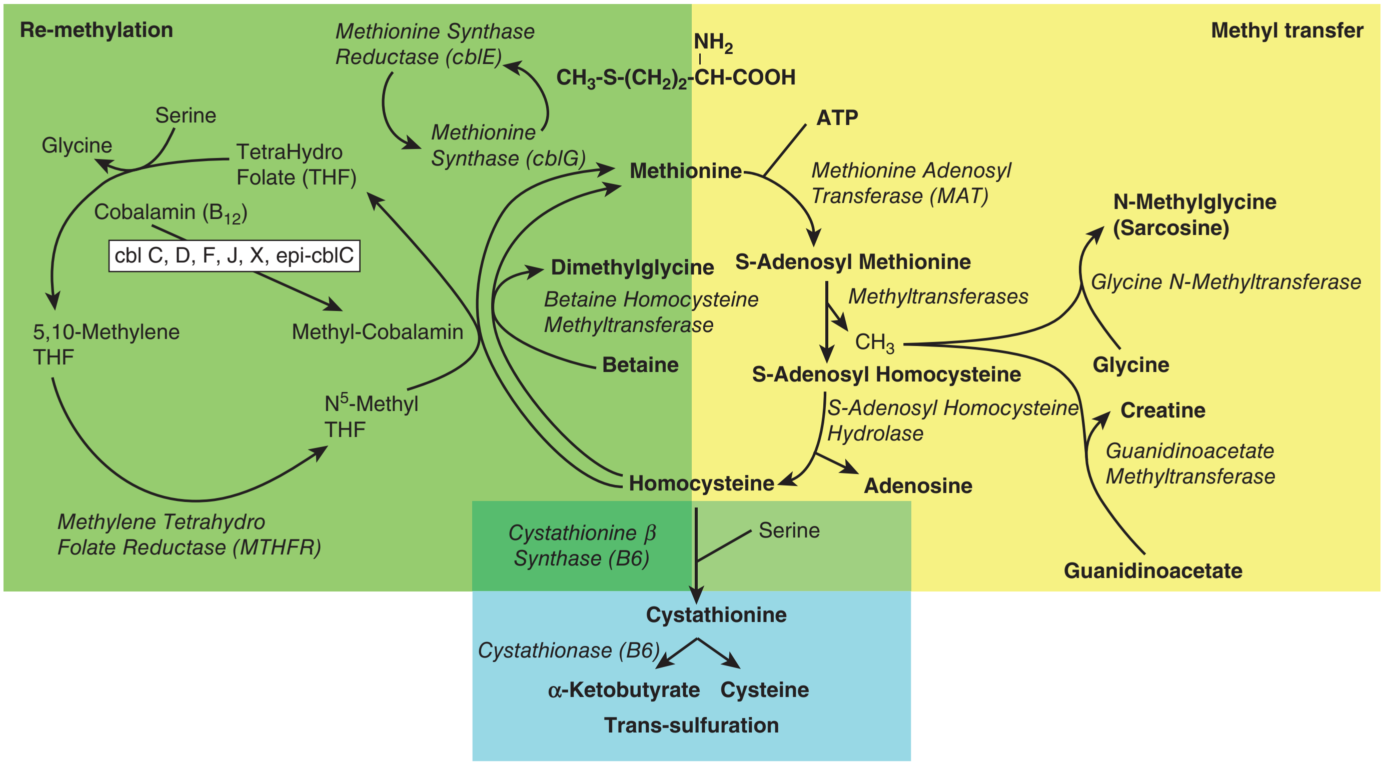
Re-methylation
Methionine Synthase Reductase (cblE)
NH2
CH3-S-(CH2)2-CH-COOH
Methyl transfer
Serine
Glycine
TetraHydro Folate (THF)
Methionine Synthase (cblG)
Methionine
ATP
Methionine Adenosyl Transferase (MAT)
Cobalamin (B12)
N-Methylglycine (Sarcosine)
cbl C, D, F, J, X, epi-cblC
Dimethylglycine
S-Adenosyl Methionine
Betaine Homocysteine Methyltransferase
Methyltransferases
Glycine N-Methyltransferase
5,10-Methylene THF
Methyl-Cobalamin
CH3
Betaine
Glycine
S-Adenosyl Homocysteine
N5-Methyl THF
S-Adenosyl Homocysteine Hydrolase
Creatine
Guanidinoacetate Methyltransferase
Homocysteine
Adenosine
Methylene Tetrahydro Folate Reductase (MTHFR)
Cystathionine β Synthase (B6)
Serine
Guanidinoacetate
Cystathionine
Cystathionase (B6)
α-Ketobutyrate
Cysteine
Trans-sulfuration

FIGURE 431-1 Pathways, enzymes, and coenzymes involved in the homocystinurias. Methionine transfers a methyl group during its conversion to homocysteine. Defects in methyl transfer or in the subsequent metabolism of homocysteine by the pyridoxal phosphate (vitamin B_6)-dependent cystathionine β-synthase increase plasma methionine levels. Homocysteine is transformed into methionine via remethylation. This occurs through methionine synthase, a reaction requiring methylcobalamin and folic acid. Deficiencies in these enzymes or lack of cofactors is associated with decreased or normal methionine levels. In an alternative pathway, homocysteine can be remethylated by betaine:homocysteine methyl transferase.

The other forms of homocystinuria are the result of impaired remethylation of homocysteine to methionine. This can be caused by defective methionine synthase or reduced availability of two essential cofactors, 5-methyltetrahydrofolate and methylcobalamin (methylvitamin B_{12}). In contrast to cystathionine β-synthase, elevated levels of free homocysteine are associated with low levels of methionine in the plasma amino acid profile in remethylation defects. Most of these conditions present with delays in development and some with megaloblastic anemia (methionine synthase-cblG and methionine synthase reductase-cblE deficiency, in addition to combined methylmalonic acidemia-homocystinuria- cblC, cblD, cblF, cblJ, see **Chap. 104**). Therapy in these cases requires administration of methylfolate, hydroxycobalamin (an activated form of vitamin B_{12}), and betaine.

Hyperhomocysteinemia refers to increased total plasma concentration of homocysteine with or without an increase in free homocysteine (disulfide form). Hyperhomocysteinemia, in the absence of significant homocystinuria, is found in some heterozygotes for the genetic defects noted above or in homozygotes for milder variants. Changes of homocysteine levels are also observed with deficiency of pyridoxine, folic acid, or vitamin B_{12}; with increasing age; with smoking; in postmenopausal women; in patients with renal failure, hypothyroidism, leukemias, autoinflammatory disorders; and during therapy with drugs such as methotrexate, nitrous oxide, givosiran, isoniazid, and some antiepileptic agents. Elevated homocysteine produces endothelial dysfunction, acting as an atherogenic and thrombophilic agent. Increased total plasma homocysteine has been associated with an increased risk for coronary, cerebrovascular, and peripheral arterial disease as well as for deep-vein thrombosis. In addition, hyperhomocysteinemia and folate and vitamin B_{12} deficiencies have been associated with an increased risk of neural tube defects in pregnant women and dementia (Alzheimer's type), as well as Parkinson's disease in the general population. Vitamin B_{12}, folic acids, and pyridoxine supplements can reduce total plasma homocysteine levels in these cases, with reduction of the risk of stroke when levels are more severely increased (>30 μ*M*).

ALKAPTONURIA

Alkaptonuria is a rare (frequency 1:250,000) disorder of tyrosine catabolism in which deficiency of homogentisate 1,2-dioxygenase (also known as *homogentisic acid oxidase*) leads to excretion of large amounts of homogentisic acid in urine and accumulation of oxidized homogentisic acid pigment in connective tissues (*ochronosis*). Alkaptonuria may go unrecognized until middle life, when degenerative joint disease develops. Prior to this time, about half of patients might be diagnosed for the presence of urine that becomes dark with standing or addition of alkali. Foci of gray-brown scleral pigment and generalized darkening of the concha, antihelix, and, finally, helix of the ear usually develop after age 30. Low back pain usually starts between 30 and 40 years of age. *Ochronotic arthritis* is heralded by pain, stiffness, and some limitation of motion of the hips, knees, and shoulders. Acute arthritis may resemble rheumatoid arthritis, but small joints are usually spared. Pigmentation of heart valves, larynx, tympanic membranes, and skin occurs, and occasional patients develop pigmented renal or prostatic calculi. Pigment deposition in the heart and blood vessels leads to aortic stenosis necessitating valve replacement, especially after 60 years of age. The diagnosis should be suspected in a patient whose urine darkens to blackness. Homogentisic acid in urine is identified by urine organic acid analysis. Ochronotic arthritis is treated symptomatically with pain medications, spinal surgery, and arthroplasty **(Chap. 383)**. Nitisinone (2-[2-nitro-4-trifluoromethylbenzoyl]-1,3-cyclohexanedione), a drug used in tyrosinemia type 1, at low dose (10 mg/d) reduces urinary excretion of homogentisic acid, delays progression, and improves clinical signs of alkaptonuria.

UREA CYCLE DEFECTS

Excess ammonia generated from protein nitrogen is removed by the urea cycle, a process mediated by several enzymes and transporters (**Fig. 431-2**, Table 431-1). Complete absence of any of these enzymes usually causes severe hyperammonemia in newborns, while milder variants can be seen in adults. The accumulation of ammonia and glutamine leads to direct neuronal toxicity and brain edema. Deficiencies

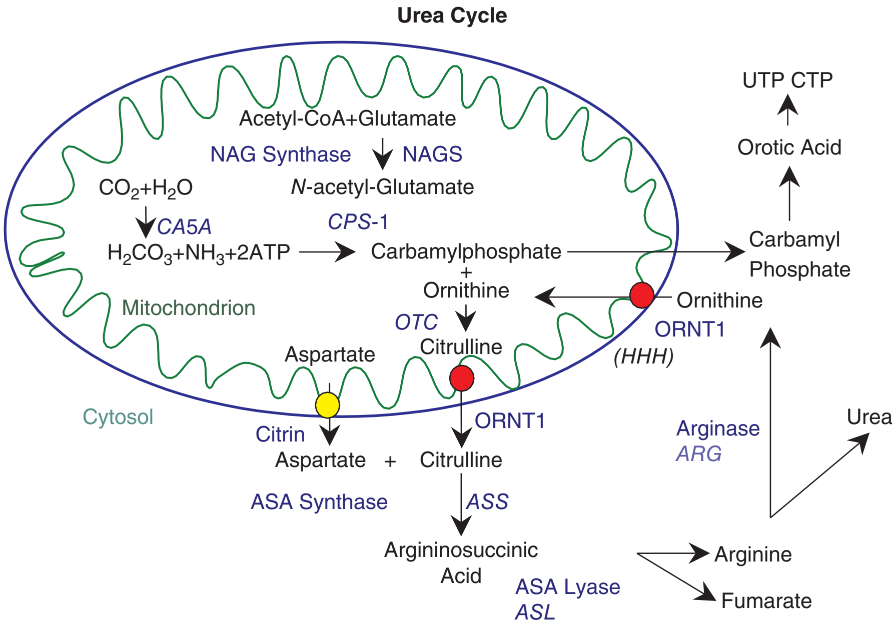

FIGURE 431-2 The urea cycle. This cycle, which is fully expressed only in the liver, forms urea starting from ammonia (NH_3) derived from the nitrogen group of all amino acids. It requires many enzymes and mitochondrial transporters, any of which can be defective and may impair the function of the urea cycle. Ammonia escaping the urea cycle in periportal hepatocytes is conjugated with glutamate by glutamine synthase in perivenous hepatocytes to generate glutamine. ARG, arginase; ASA, argininosuccinic acid; ASL, argininosuccinate lyase; ASS, argininosuccinate synthase; CA5A, carbonic anhydrase 5a; citrin (*SLC25A13*), aspartate/glutamate exchanger; CP, carbamylphosphate; CPS-1, carbamylphosphate synthase 1; CTP, cytidine triphosphate; HHH, hyperammonemia, hyperornithinemia, homocitrullinuria syndrome; NAG, *N*-acetylglutamate; NAGS, *N*-acetylglutamate synthase; ORNT1 (*SLC25A15*), ornithine/citrulline mitochondrial transporter; OTC, ornithine transcarbamylase; UTP, uridine triphosphate.

in urea cycle enzymes are individually rare, but as a group, they affect ~1:35,000 individuals. They are all transmitted as autosomal recessive traits, with the exception of ornithine transcarbamylase deficiency, which is X-linked and the most frequent urea cycle defect. Hepatocytes of females with ornithine transcarbamylase deficiency express either the normal or the mutant allele due to random X-inactivation and may be unable to remove excess ammonia if mutant cells are predominant.

Infants with classic urea cycle defects present at 1–4 days of life with refusal to eat and lethargy progressing to coma and death. Milder enzyme deficiencies present with protein avoidance, recurrent vomiting, migraine, mood swings, chronic fatigue, irritability, and disorientation that can progress to coma. Some cases have presented with acute or chronic hepatic dysfunction. Females with ornithine transcarbamylase deficiency can present at time of childbirth due to the combination of involuntary fasting and stress that favors catabolism. Administration of systemic corticosteroids or chemotherapy can precipitate hyperammonemia and can be fatal in previously asymptomatic individuals of any age. These patients may be misdiagnosed as having gastrointestinal disorders, food allergies, behavioral problems, or nonspecific hepatitis. The diagnosis requires measurement of plasma ammonia, plasma amino acids, and urine orotic acid, useful for differentiating ornithine transcarbamylase deficiency from carbamyl phosphate synthase-1 and *N*-acetylglutamate synthase deficiency. Increased plasma glutamine is seen with all urea cycle defects since ammonia not removed by the urea cycle in periportal hepatocytes is conjugated to glutamate by glutamine synthase in perivenous hepatocytes. Citrulline is low or undetectable in proximal defects of the urea cycle (*N*-acetylglutamate synthase, carbamylphosphate synthase 1, and ornithine transcarbamylase deficiency), with urine orotic acid being increased only in ornithine transcarbamylase deficiency. Plasma citrulline is markedly increased in argininosuccinic acid synthase deficiency (citrullinemia type 1), with a milder elevation in argininosuccinic acid lyase deficiency in the presence of argininosuccinic acid (argininosuccinic aciduria). Arginine levels are usually normal to low in these conditions and become markedly elevated only in patients with arginase deficiency. In addition to urea cycle defects, hyperammonemia can also be caused by liver disease from any cause and several organic acidemias and fatty acid oxidation defects (the latter two excluded by the analysis of urine organic acids and plasma acylcarnitine profile).

TREATMENT

Urea Cycle Defects

Therapy is aimed at stopping catabolism and ammonia production by providing adequate calories (as IV glucose and lipids in the comatose patient) and, if needed, insulin. Excess nitrogen is removed by IV phenylacetate and benzoate (0.25 g/kg for the priming dose and subsequently as an infusion over 24 h) that conjugate with glutamine and glycine, respectively, to form phenylacetylglutamine and hippuric acid, water-soluble molecules efficiently excreted in urine. Arginine (200 mg/kg per d) becomes an essential amino acid (except in arginase deficiency) and should be provided intravenously to resume protein synthesis. If these measures fail to reduce ammonia, hemodialysis should be initiated promptly. Chronic therapy consists of a protein-restricted diet, phenylbutyrate, glycerol phenylbutyrate (a liquid drug better tolerated by most patients), arginine, or citrulline supplements, depending on the specific diagnosis. Oral carglumic acid can restore a functional urea cycle in patients with *N*-acetylglutamate synthase deficiency and renders other therapies unnecessary. Liver transplantation should be considered in patients with severe urea cycle defects that are difficult to control medically.

Hyperammonemia due to a functional deficiency of glutamine synthase can occur in patients receiving chemotherapy for different malignancies or undergoing solid organ transplants. It can also be seen with hepatic cirrhosis. Several of these patients have been successfully rescued from hyperammonemia using the protocol described above for urea cycle defects.

FURTHER READING

Guéant JL et al: Hyperhomocysteinemia in cardiovascular diseases: Revisiting observational studies and clinical trials. Thromb Haemost 123:270, 2023.

Ranganath LR et al: Efficacy and safety of once-daily nitisinone for patients with alkaptonuria (SONIA 2): An international, multicentre, open-label, randomised controlled trial. Lancet Diabetes Endocrinol 8:762, 2020.

Van Spronsen FJ et al: Phenylketonuria. Nat Rev Dis Primers 7:36, 2021.

432 Inherited Defects of Membrane Transport

Nicola Longo

Membrane transporters mediate the passage of amino acids, oligopeptides, sugars, cations, anions, vitamins, water, and other molecules across cellular membranes and are encoded by members of the solute-carrier gene (SLC) superfamily. These transporters are located on the plasma membrane or intracellular organelles, and their cellular and tissue distribution in addition to the presence (or absence) of redundant transporters explains organ involvement and possible metabolic disturbances. Transport processes are essential for the normal function of every organ, but especially the brain and sensory organs **(Table 432-1)**. Inherited defects impairing the transport of selected amino acids that can present in adults are discussed here as examples of the abnormalities encountered; others are considered elsewhere in this text.

CYSTINURIA

Cystinuria (worldwide frequency of 1 in 7000) is an autosomal recessive disorder caused by defective transporters in the apical brush border of proximal renal tubule and small intestinal cells. It is characterized by impaired reabsorption and excessive urinary excretion of the amino acids lysine, arginine, ornithine, and cystine that are dibasic in the physiologically acidic pH of urine. Because cystine is poorly soluble, its excess predisposes to the formation of renal, ureteral, and bladder stones. Such stones are responsible for the signs and symptoms of the disorder.

There are two variants of cystinuria. Homozygotes for both variants have high urinary excretion of cystine, lysine, arginine, and ornithine. Type A heterozygotes usually have normal urinary amino acid excretion, whereas most type B heterozygotes have moderately increased urinary excretion of cystine that, in some circumstances, can result in the formation of kidney stones. The gene for type A cystinuria (*SLC3A1*, chromosome 2p16.3) encodes a membrane glycoprotein. Type B cystinuria is caused by mutations in *SLC7A9* (chromosome 19q13) that encodes the $b^{0,+}$ amino acid transporter. The glycoprotein encoded by *SLC3A1* favors the correct processing of the $b^{0,+}$ membrane transporter and explains why mutations in two different genes cause a similar disease.

Cystine stones account for 1–2% of all urinary tract calculi and for ~4–5% of stones in children. Cystinuria homozygotes regularly excrete 2400–7200 μmol (600–1800 mg) of cystine daily. Since the maximum solubility of cystine in the physiologic urinary pH range of 4.5–7.0 is ~1200 μmol/L (300 mg/L), cystine needs to be diluted to 2.5–7 L of water to prevent crystalluria. Stone formation usually manifests in the second or third decade but may occur in the first year of life. Symptoms and signs are those typical of urolithiasis: hematuria, flank pain, renal colic, obstructive uropathy, and infection **(Chap. 330)**. Recurrent urolithiasis may lead to progressive renal insufficiency.

Cystinuria is suspected after observing typical hexagonal crystals in the sediment of acidified, concentrated, chilled urine or after performing a urinary nitroprusside test. Quantitative urine amino acid analysis shows selective overexcretion of cystine, lysine, arginine, and ornithine. Quantitative measurements are important for differentiating heterozygotes from homozygotes and for following free cystine excretion during therapy.

Management is aimed at preventing cystine crystal formation by increasing urinary volume and by maintaining an alkaline urine pH. Fluid ingestion in excess of 4 L/d is essential, and 5–7 L/d is optimal. Urinary cystine concentration should be <1000 μmol/L (250 mg/L). The daily fluid ingestion necessary to maintain this dilution of excreted cystine should be spaced over 24 h, with one-third of the total volume ingested between bedtime and 3 A.M. Cystine solubility rises sharply above pH 7.5, and urinary alkalinization (with potassium citrate) can be therapeutic. Penicillamine (1–3 g/d) and tiopronin (α-mercaptopropionylglycine, 800–1200 mg/d in four divided doses) undergo sulfhydryl-disulfide exchange with cystine to form mixed disulfides. Because these disulfides are much more soluble than cystine, pharmacologic therapy can prevent and promote dissolution of calculi. Penicillamine can have significant side effects and should be reserved for patients who fail to respond to hydration alone or who are in a high-risk category (e.g., one remaining kidney, renal insufficiency). When medical management fails, shock wave lithotripsy, ureteroscopy, and percutaneous nephrolithotomy are effective for most stones. Open urologic surgery is considered only for complex staghorn stones or when the patient has concomitant renal or ureteral abnormalities. Occasional patients progress to renal failure and require kidney transplantation.

LYSINURIC PROTEIN INTOLERANCE

Lysinuric protein intolerance is characterized by a defect in renal tubular reabsorption and intestinal transport of the three dibasic amino acids lysine, arginine, and ornithine but *not* cystine. It is most common in Finland (1 in 60,000), southern Italy, and Japan, but is rare elsewhere. The transport defect affects basolateral rather than luminal membrane transport and causes secondary impairment of the urea cycle. The defective gene (*SLC7A7*, chromosome 14q11.2) encodes the y^+LAT membrane transporter, which associates with the cell-surface glycoprotein 4F2 heavy chain to form the complete sodium-independent transporter y^+L.

Manifestations are related to impairment of the urea cycle and to immune dysfunction likely attributable to nitric oxide overproduction secondary to arginine intracellular trapping within white blood cells. Affected patients present in childhood with hepatosplenomegaly, protein intolerance, and episodic ammonia intoxication. Adults may present with severe osteoporosis, pancreatitis, impaired renal function, pulmonary alveolar proteinosis, various autoimmune disorders, and an incompletely characterized immune deficiency. Plasma concentrations of lysine, arginine, and ornithine are reduced, whereas urinary excretion of lysine, arginine, ornithine, and orotic acid is increased. Hyperammonemia may develop after the ingestion of protein loads or with infections, probably because of insufficient amounts of ornithine to maintain proper function of the urea cycle. Diagnosis is confirmed by sequencing of the *SLC7A7* gene that is included in most hyperammonemia panels. Therapy consists of dietary protein restriction, supplementation of citrulline (2–8 g/d), a neutral amino acid that fuels the urea cycle when metabolized to arginine and ornithine, and nitrogen scavengers (phenylbutyrate, benzoate) in case of persistent hyperammonemia. Pulmonary disease can respond to glucocorticoids or recombinant human granulocyte-macrophage colony-stimulating factor but might require broncho-alveolar or whole lung lavage in some patients. Women with lysinuric protein intolerance who become pregnant have an increased risk of anemia, toxemia, and bleeding complications during delivery. These can be minimized by aggressive nutritional therapy and control of blood pressure. Their infants can have intrauterine growth restriction but have normal neurologic function.

CITRULLINEMIA TYPE 2 (CITRIN DEFICIENCY)

Citrullinemia type 2 is a recessive condition caused by deficiency of the mitochondrial aspartate-glutamate carrier AGC2 (citrin). A defect in this transporter reduces the availability of cytoplasmic aspartate to combine with citrulline to form argininosuccinate **(see Fig. 431-2)**, impairing the urea cycle and decreasing the transfer of reducing equivalents from the cytosol to the mitochondria through the malate-aspartate NADH shuttle. Mutations in the *SLC25A13* gene on chromosome 7q21.3 that encodes for this transporter are rare in Caucasians but affect ~1:20,000 people with ancestry from Japan, China, and Southeast Asia with variable penetrance.

The disease can present in children with neonatal intrahepatic cholestasis, failure to thrive, and dyslipidemia but usually presents with sudden onset between 20 and 50 years of age with recurring episodes of hyperammonemia with associated neuropsychiatric symptoms such as altered mental status, irritability, seizures, or coma-resembling hepatic encephalopathy. Some patients might come to medical attention for

TABLE 432-1 Genetic Disorders of Amino Acid Transport

DISORDER	SUBSTRATES	TISSUES MANIFESTING TRANSPORT DEFECT	MOLECULAR DEFECT	MAJOR CLINICAL MANIFESTATIONS	INHERITANCE
Cystinuria	Cystine, lysine, arginine, ornithine	Proximal renal tubule, jejunal mucosa	Shared dibasic-cystine transporter *SLC3A1, SLC7A9*	Cystine nephrolithiasis	AR
Lysinuric protein intolerance	Lysine, arginine, ornithine	Proximal renal tubule, jejunal mucosa	Dibasic transporter *SLC7A7*	Protein intolerance, hyperammonemia, intellectual disability	AR
Hartnup disease	Neutral amino acids	Proximal renal tubule, jejunal mucosa	Neutral amino acid transporter *SLC6A19*	Constant neutral aminoaciduria, intermittent symptoms of pellagra	AR
Histidinuria	Histidine	Proximal renal tubule, jejunal mucosa	Histidine transporter	Intellectual disability	AR
Iminoglycinuria	Glycine, proline, hydroxyproline	Proximal renal tubule, jejunal mucosa	Shared glycine–amino acid transporter *SLC6A20, SLC6A18, SLC36A2*	None	AR
Dicarboxylic aminoaciduria	Glutamic acid, aspartic acid	Proximal renal tubule, jejunal mucosa	Shared dicarboxylic amino acid transporter *SLC1A1*	None	AR
Hyperargininemia	Arginine, lysine, ornithine	Ubiquitous	CAT2 cationic amino acid transporter *SLC7A2*	Hyperargininemia, Hyperammonemia (?)	AR
Brain branched-chain amino acid deficiency	Leucine, isoleucine, valine	Plasma membrane of blood-brain barrier	Branched-chain amino acid transporter *SLC7A5*	Microcephaly, intellectual disability, seizures, autism	AR
Citrullinemia type 2	Aspartate, glutamate, malate	Inner mitochondrial membrane	Mitochondrial aspartate/glutamate carrier 2 *SLC25A13*	Sudden behavioral changes with stupor, coma, hyperammonemia	AR
Hyperornithinemia, hyperammonemia, homocitrullinuria	Ornithine, citrulline	Inner mitochondrial membrane	Mitochondrial ornithine carrier *SLC25A15*	Lethargy, failure to thrive, intellectual disability, episodic confusion, hyperammonemia, protein intolerance	AR
Epileptic encephalopathy	Aspartate, glutamate, malate	Inner mitochondrial membrane	Mitochondrial aspartate/glutamate carrier 1 *SLC25A12*	Intellectual disability, epilepsy, hypotonia, cerebral atrophy, and hypomyelination	AR
Epileptic encephalopathy	Glutamate	Inner mitochondrial membrane	Mitochondrial glutamate carrier *SLC25A22*	Intellectual disability, epilepsy	AR
Epileptic encephalopathy	Glutamic acid, aspartic acid	Presynaptic glutamatergic nerve endings	EEAT2 Neuronal dicarboxylic amino acid transporter *SLC1A2*	Developmental and Epileptic Encephalopathy	AD
Episodic ataxia	Glutamic acid, aspartic acid	Presynaptic glutamatergic nerve endings	EEAT1 Neuronal dicarboxylic amino acid transporter *SLC1A3*	Episodic ataxia	AD
Brain serine deficiency	Alanine, serine, cysteine, threonine	Neuronal cells	ASCT neutral amino acid transporter *SLC1A4*	Progressive microcephaly, intellectual disability, spasticity	AR
Glycine encephalopathy with normal serum glycine	Glycine	Astrocytes and neuronal cells	GLYT1 astrocyte glycine transporter *SLC6A9*	Arthrogryposis, apnea, axial hypotonia, spasticity, intellectual disability	AR
Hyperekplexia-3	Glycine	Neuronal cells	GLYT2 Presynaptic glycine transporter *SLC6A5*	Exaggerated startle response, hypertonia, apnea	AR
Intellectual disability	Proline, glycine, leucine, and alanine, glutamine	Neuronal cells synaptic vesicles	NTT4 synaptic vesicle neutral amino acid transporter *SLC6A17*	Intellectual disability, tremor	AR
Deafness	Glutamic acid	Neuronal cortical synaptic vesicles	VGLUT3 vesicular glutamate transporter *SLC17A8*	Deafness	AD
Foveal hypoplasia	Glutamine	Retinal photoreceptors	*SLC38A8*	Foveal hypoplasia, optic nerve decussation defects, anterior segment dysgenesis	AR
Retinitis pigmentosa	Arginine, lysine, ornithine	Retinal photoreceptors	Cationic amino acid transporter *SLC7A14*	Retinitis pigmentosa, blindness	AR
Early retinal degeneration	Taurine	Retinal cells	TAUT taurine transporter *SLC6A6*	Nystagmus, vision loss, retinal degeneration	AR
Cystinosis	Cystine	Lysosomal membranes	Lysosomal cystine transporter	Renal failure, hypothyroidism, blindness	AR

Abbreviations: AD, autosomal dominant; AR, autosomal recessive.

hypertriglyceridemia, pancreatitis, hepatoma, or fatty liver histologically similar to nonalcoholic steatohepatitis. Without therapy, most symptomatic patients die with cerebral edema within a few years. Episodes are usually triggered by medications (such as acetaminophen), surgery, alcohol, or high sugar intake, with the latter conditions causing NADH production in the cytoplasm. NADH is not generated by the metabolism of proteins or fats, and individuals with citrullinemia type 2 spontaneously prefer foods such as meat, eggs, and fish and avoid carbohydrates.

Laboratory studies during an acute attack can show elevated ammonia, citrulline, and arginine with low or normal levels of glutamine (the latter is usually increased in classic urea cycle defects). Levels of galactose-1-phosphate in red blood cells are also increased, reflecting defective transfer of reducing equivalents from the cytosol to mitochondria. The diagnosis is confirmed by demonstrating pathogenic variants in the *SLC25A13* gene. Liver transplantation prevents progression of the disease and normalizes biochemical parameters. A ketogenic diet high in fats and proteins and low in carbohydrates with supplements of medium-chain triglycerides, arginine, and pyruvate is also effective in preventing or delaying disease progression.

HARTNUP DISEASE

Hartnup disease (frequency 1 in 24,000) is an autosomal recessive disorder characterized by pellagra-like skin lesions, variable neurologic manifestations, and neutral and aromatic aminoaciduria. Alanine, serine, threonine, valine, leucine, isoleucine, phenylalanine, tyrosine, tryptophan, glutamine, asparagine, and histidine are excreted in urine in quantities 5–10 times greater than normal, and intestinal transport of these same amino acids is defective. The defective neutral amino acid transporter, B°AT1 encoded by the *SLC6A19* gene on chromosome 5p15, requires either collectrin or angiotensin-converting enzyme 2 (one of the binding sites for SARS-CoV-2) for surface expression in the kidney and intestine, respectively.

The clinical manifestations result from nutritional deficiency of the essential amino acid tryptophan, caused by its intestinal and renal malabsorption, and of niacin, which derives in part from tryptophan metabolism. Only a small fraction of patients with Hartnup disease develop symptoms, implying that manifestations depend on other factors in addition to the transport defect. The diagnosis of Hartnup disease should be suspected in any patient with clinical features of pellagra, recurrent diarrhea, and/or neurologic symptoms who does not have a history of dietary niacin deficiency **(Chap. 345)**. The neurologic and psychiatric manifestations range from attacks of spastic paraplegia to cerebellar ataxia to mild emotional lability to frank delirium, and they are usually accompanied by exacerbations of the erythematous, eczematoid skin rash. Fever, sunlight, stress, and sulfonamide therapy provoke clinical relapses. Diagnosis is made by detection of the neutral aminoaciduria (which does not occur in dietary niacin deficiency) and is confirmed by genetic testing of the *SLC6A19* or the *CTLRN* gene (coding for collectrin), whose deficiency produces a biochemical phenocopy. Treatment includes a high-protein diet and daily nicotinamide supplementation (50–250 mg).

CYSTINOSIS

Cystinosis (frequency 1 in 100,000–200,000) is an autosomal recessive disorder caused by mutations in the *CTNS* gene encoding the lysosomal cystine/proton transporter (cystinosin). In this condition, cystine derived from protein degradation accumulates inside lysosomes and forms crystals due to its poor solubility. Depending on the degree of impairment of transporter function, three clinical forms are recognized. The most severe form, classic nephropathic cystinosis, causes renal Fanconi syndrome with rickets during the first year of life and, without treatment, evolves to renal failure usually by 10 years of age. Juvenile nephropathic cystinosis presents with proteinuria slowly leading to kidney failure, whereas photophobia, caused by deposition of cystine crystals in the cornea, is the only manifestation of ocular nonnephropathic cystinosis. Cystinosis is suspected by the identification of cystine crystals in the cornea by slit lamp examination and diagnosed by measuring cystine content in white blood cells and/or DNA testing (including deletion analysis) of the *CTNS* gene. Therapy consists in the administration of extended release cysteamine bitartrate that enters lysosomes and forms a mixed disulfide with cysteine that is exported from the lysosome using a cationic amino acid transporter. This drug is given orally and should be slowly increased to the maintenance dose of 1.3 g/m^2 divided into two daily administrations while monitoring white blood cell (WBC) cystine levels for efficacy. This therapy delays renal failure and is more effective if started early in the course of the disease. Cysteamine eye drops can relieve photophobia. Renal replacement therapy with salts, alkali, and activated vitamin D is necessary for renal Fanconi syndrome. Cystine accumulation occurs in all organs and tissues, causing additional complications such as hypothyroidism, hypohydrosis, diabetes, and delayed puberty in both males and females with primary hypogonadism in males. Growth hormone replacement, L-thyroxine for hypothyroidism, insulin for diabetes mellitus, and testosterone for hypogonadism in males may be necessary. Despite therapy, many patients with cystinosis progress to end-stage renal failure and require kidney transplantation. Late-onset complications include hepatomegaly and splenomegaly that occur in approximately one-third of subjects and a vacuolar myopathy causing weakness (initially involving the distal extremities), swallowing difficulties, gastrointestinal dysmotility, and pulmonary insufficiency. Before the availability of cystine-depleting therapy and renal transplantation, the life span in nephropathic cystinosis was <10 years. With current therapies, affected individuals can survive into the late forties with satisfactory quality of life.

FURTHER READING

Bölsterli BK et al: Ketogenic diet treatment of defects in the mitochondrial malate aspartate shuttle and pyruvate carrier. Nutrients 14:3605, 2022.

Levtchenko E et al: Expert guidance on the multidisciplinary management of cystinosis in adolescent and adult patients. Clin Kidney J 15:1675, 2022.

Servais A et al: Cystinuria: Clinical practice recommendation. Kidney Int 99:48, 2021.

Yahyaoui R, Pérez-Frías J: Amino acid transport defects in human inherited metabolic disorders. Int J Mol Sci 21:119, 2019.

This page intentionally left blank

Section 1 Diagnosis of Neurologic Disorders

433

Approach to the Patient with Neurologic Disease

Daniel H. Lowenstein, S. Andrew Josephson, Stephen L. Hauser

Neurologic diseases are common and costly. According to estimates by the World Health Organization, neurologic disorders affect nearly one in three individuals and represent the leading cause of disability and the second leading cause of death worldwide **(Fig. 433-1)**. These numbers have increased significantly over the past 30 years and are expected to double by 2050, due to multiple factors, including aging of the population, population growth, and a rising prevalence of metabolic risk conditions.

Because therapies now exist for many neurologic disorders, a skillful approach to diagnosis is essential. Errors commonly result from an overreliance on costly neuroimaging procedures and laboratory tests, which, while useful, do not substitute for an adequate history and examination. The proper approach begins with the patient and focuses the clinical problem, first in anatomic and then in pathophysiologic terms; only then should a specific neurologic diagnosis be entertained. This method ensures that technology is judiciously applied, a correct diagnosis is established in an efficient manner, and treatment is promptly initiated.

THE NEUROLOGIC METHOD

DEFINE THE ANATOMY

The first priority is to identify the region of the nervous system that is likely to be responsible for the symptoms. Can the disorder be mapped to one specific location, is it multifocal, or is a diffuse process present? Are the symptoms restricted to the nervous system, or do they arise in the context of a systemic illness? Is the problem in the central nervous system (CNS), the peripheral nervous system (PNS), or both? In the CNS, is the cerebral cortex, basal ganglia, brainstem, cerebellum, or spinal cord responsible? Are the pain-sensitive meninges involved? In the PNS, could the disorder be located in peripheral nerves and, if

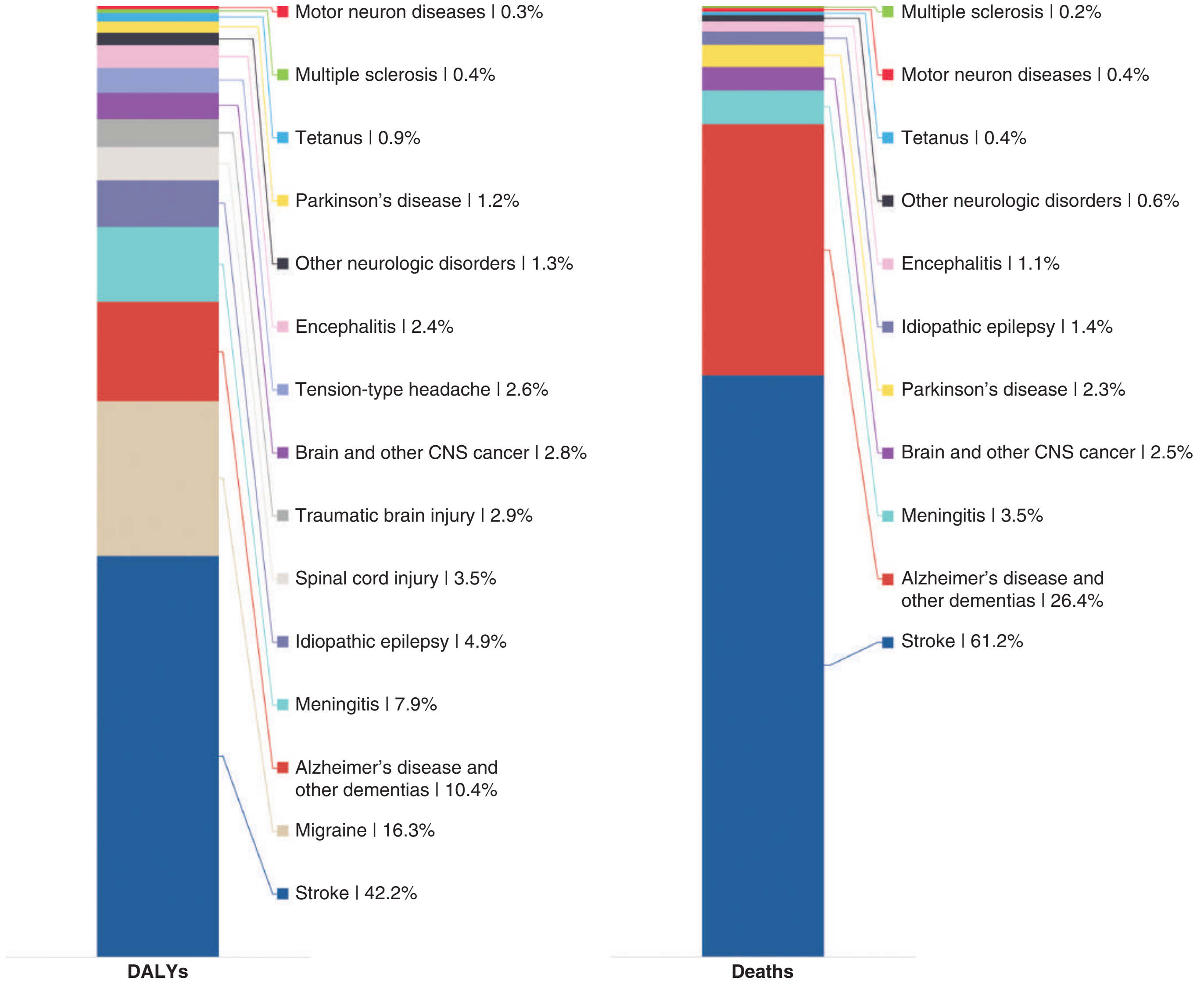

FIGURE 433-1 Proportional contributions to the overall global burden of neurologic disorders. Proportions (%) of disability-adjusted life-years (DALYs) and deaths. *(Based on data from GBD 2016 Neurology Collaborators: Lancet Neurol 18:459, 2019.)*

so, are motor or sensory nerves primarily affected, or is a lesion in the neuromuscular junction or muscle more likely?

The first clues to defining the anatomic area of involvement appear in the history, and the examination is then directed to confirm or rule out these impressions and to clarify uncertainties. A more detailed examination of a particular region of the CNS or PNS is often indicated. For example, the examination of a patient who presents with a history of ascending paresthesias and weakness should be directed toward deciding, among other things, if the lesion is in the spinal cord or peripheral nerves. Focal back pain, a spinal cord sensory level, and incontinence suggest a spinal cord origin, whereas a stocking-glove pattern of sensory loss suggests peripheral nerve disease; areflexia usually indicates peripheral neuropathy but may also be present, typically transiently, with acute spinal cord disorders.

Deciding "where the lesion is" accomplishes the task of limiting the possible etiologies to a manageable, finite number. In addition, this strategy safeguards against making serious errors. Symptoms of recurrent vertigo, diplopia, and nystagmus should not trigger "multiple sclerosis" as an answer (etiology) but "brainstem" or "pons" (location); then a diagnosis of brainstem arteriovenous malformation will not be missed for lack of consideration. Similarly, the combination of visual blurring and spastic ataxic paraparesis suggests possible optic nerve and spinal cord disease; multiple sclerosis (MS), CNS syphilis, and vitamin B_{12} deficiency are treatable disorders that can produce this syndrome. Once the question, "Where is the lesion?" is answered, then the question "What is the lesion?" can be addressed.

IDENTIFY THE PATHOPHYSIOLOGY

Clues to the pathophysiology of the disease process may also be present in the history. Primary neuronal (gray matter) disorders often present as early cognitive disturbances, movement disorders, or seizures, whereas disorders of the connecting pathways (white matter involvement) produce "long tract" motor, sensory, visual, and/or cerebellar manifestations. Progressive and symmetric symptoms often have a metabolic or degenerative origin; in such cases, lesions are usually not sharply circumscribed. Thus, a patient with paraparesis and a clear spinal cord sensory level is unlikely to have vitamin B_{12} deficiency as the explanation. A Lhermitte symptom (electric shock–like sensations evoked by neck flexion) is due to ectopic impulses that arise in white matter pathways and occurs with demyelination in the cervical spinal cord; among many possible causes, this symptom may indicate MS in a young adult or compressive cervical spondylosis in an older person. Symptoms that worsen after exposure to heat or exercise may indicate conduction block in demyelinated axons, as occurs in MS. A patient with recurrent episodes of diplopia and dysarthria associated with exercise or fatigue may have a disorder of neuromuscular transmission such as myasthenia gravis. Slowly advancing visual scotoma with luminous edges, termed *fortification spectra*, indicates spreading cortical depression, typically with migraine.

THE NEUROLOGIC HISTORY

Attention to the description of symptoms experienced by the patient and substantiated by family members and others often permits an accurate localization and determination of the probable cause, even before the neurologic examination is performed. The history also helps focus the neurologic examination that follows. Each complaint should be pursued as far as possible to identify the location of the lesion, the likely underlying pathophysiology, and potential etiologies. For example, a patient complains of weakness of the right arm. What are the associated features? Does the patient have difficulty with brushing hair or reaching upward (proximal) or fastening buttons or opening a plastic bottle (distal)? Negative associations may also be crucial. A right-handed patient with a right hemiparesis without a language deficit likely has a lesion (internal capsule, brainstem, or spinal cord) different from that of a patient with a right hemiparesis and aphasia (left hemisphere). Other pertinent features of the history include the following:

1. *Temporal course of the illness*. It is important to determine the precise time of appearance and rate of progression of the symptoms experienced by the patient. The rapid onset of a neurologic complaint, occurring within seconds or minutes, usually indicates a vascular event, a seizure, or migraine. The onset of sensory symptoms located in one extremity that spread over a few seconds to adjacent portions of that extremity and then to the other regions of the body suggests a seizure. A similar but slower temporal march of symptoms accompanied by headache, nausea, or visual disturbance suggests migraine. Less well-localized symptoms that are maximum at onset and persist for seconds, minutes, or much less commonly hours, point to the possibility of a transient ischemic attack (TIA). The presence of "positive" sensory symptoms (e.g., tingling or sensations that are difficult to describe) or involuntary motor movements suggests a seizure; in contrast, transient loss of function (negative symptoms) suggests a TIA. A stuttering onset where symptoms appear, stabilize, and then progress over hours or days also suggests cerebrovascular disease; an additional history of transient remission or regression indicates that the process is more likely due to ischemia rather than hemorrhage. A gradual evolution of symptoms over hours or days suggests a toxic, metabolic, infectious, or inflammatory process. Progressing symptoms associated with the systemic manifestations of fever, stiff neck, and altered level of consciousness imply an infectious process. Relapsing and remitting symptoms involving different levels of the nervous system suggest MS or other inflammatory processes. Slowly progressive symptoms without remissions are characteristic of neurodegenerative disorders, chronic infections, gradual intoxications, and neoplasms.
2. *Patients' descriptions of the complaint*. The same words often mean different things to different patients. "Dizziness" may imply impending syncope, a sense of disequilibrium, or true spinning vertigo. "Numbness" may mean a complete loss of feeling, a positive sensation such as tingling, or even weakness. "Blurred vision" may be used to describe unilateral visual loss, as in transient monocular blindness, or diplopia. The interpretation of the true meaning of the words used by patients to describe symptoms obviously becomes even more complex when there are differences in primary languages and cultures.
3. *Corroboration of the history by others*. It is almost always helpful to obtain additional information from family, friends, or other observers to corroborate or expand the patient's description. Memory loss, aphasia, loss of insight, intoxication, and other factors may impair the patient's capacity to communicate normally with the examiner or prevent openness about factors that have contributed to the illness. Episodes of loss of consciousness necessitate that details be sought from observers to ascertain precisely what has happened during the event.
4. *Family history*. Many neurologic disorders have an underlying genetic component. The presence of a Mendelian disorder, such as Huntington's disease or Charcot-Marie-Tooth neuropathy, is often obvious if family data are available. Clearly note if the family history has been fully ascertained or if some branches of the family could not be traced. More detailed questions about family history are often necessary in polygenic disorders such as MS, migraine, and many types of epilepsy. It is important to elicit family history about all illnesses, in addition to neurologic and psychiatric disorders. A familial propensity to hypertension or heart disease is relevant in a patient who presents with a stroke. Numerous inherited neurologic diseases are associated with multisystem manifestations that may provide clues to the correct diagnosis (e.g., neurofibromatosis, Wilson's disease, mitochondrial disorders).
5. *Medical illnesses*. Many neurologic diseases occur in the context of systemic disorders. Diabetes mellitus, hypertension, and abnormalities of blood lipids predispose to cerebrovascular disease. A solitary mass lesion in the brain may be an abscess in a patient with valvular heart disease, a primary hemorrhage in a patient with a coagulopathy, a lymphoma or toxoplasmosis in a patient with AIDS, or a metastasis in a patient with underlying cancer. Patients with malignancy may also present with a neurologic paraneoplastic syndrome **(Chap. 99)** or complications from chemotherapy, immunotherapy,

or radiotherapy. Marfan's syndrome and related collagen disorders predispose to dissection of the cranial arteries and aneurysmal subarachnoid hemorrhage; the latter may also occur with polycystic kidney disease and fibromuscular dysplasia. Various neurologic disorders occur with dysthyroid states or other endocrinopathies. It is especially important to look for the presence of systemic diseases in patients with peripheral neuropathy. Most patients with coma in a hospital setting have a metabolic, toxic, or infectious cause.

6. *Drug use and abuse and toxin exposure*. It is essential to inquire about the history of drug use, both prescribed and illicit. Sedatives, antidepressants, and other psychoactive medications are frequently associated with acute confusional states, especially in the elderly. Aminoglycoside antibiotics may exacerbate symptoms of weakness in patients with disorders of neuromuscular transmission, such as myasthenia gravis, and may cause dizziness secondary to ototoxicity. Vincristine and other antineoplastic drugs can cause peripheral neuropathy, immunosuppressive agents such as cyclosporine can produce encephalopathy, and immunotherapies with checkpoint inhibitors and chimeric antigen receptor (CAR) T cells are associated with various neurotoxicities including brain edema, inflammation, and demyelination **(Chap. 318)**. Excessive vitamin ingestion can lead to disease; examples include vitamin A and intracranial hypertension ("pseudotumor cerebri") or pyridoxine and peripheral neuropathy. Many patients are unaware that over-the-counter sleeping pills, cold preparations, and diet pills are actually drugs. Alcohol, the most prevalent neurotoxin, is often not recognized as such by patients, and other drugs of abuse such as cocaine, methamphetamine, and heroin can cause a wide range of neurologic abnormalities. A history of environmental or industrial exposure to neurotoxins may provide an essential clue; consultation with the patient's coworkers or employer may be required.
7. *Formulating an impression of the patient*. Use the opportunity while taking the history to form an impression of the patient. Is the information forthcoming, or does it take a circuitous course? Is there evidence of anxiety, depression, or hypochondriasis? Are there any clues to problems with language, memory, insight, comportment, or behavior? The neurologic assessment begins as soon as the patient comes into the room and the first introduction is made.

THE NEUROLOGIC EXAMINATION

The neurologic examination can be challenging and complex; it has many components and includes a number of skills that can be mastered only through repeated use of the same techniques on a large number of individuals with and without neurologic disease. Mastery of the complete neurologic examination is usually important only for physicians in neurology and associated specialties. However, knowledge of the basics of the examination, especially those components that are effective in screening for neurologic dysfunction, is essential for all clinicians, especially generalists.

There is no single, universally accepted sequence of the examination that must be followed, but most clinicians begin with assessment of mental status followed by the cranial nerves (CN), motor system, reflexes, sensory system, coordination, and gait. Whether the examination is basic or comprehensive, it is essential that it is performed in an orderly and systematic fashion to avoid errors and serious omissions. Thus, the best way to learn and gain expertise in the examination is to choose one's own approach and practice it frequently and do it in the same exact sequence each time. Essential tools for the examination should be accessible whenever the exam goes beyond screening, including a Snellen visual chart, portable ophthalmoscope, reflex hammer, a 128-Hz tuning fork, and disposable cotton wisps, tongue depressors, and pins.

The detailed description that follows describes the more commonly used parts of the neurologic examination, with a particular emphasis on the components that are considered most helpful for the assessment of common neurologic problems. Each section also includes a brief description of the minimal examination necessary to adequately screen for abnormalities in a patient who has no symptoms suggesting neurologic dysfunction. A screening examination done in this way can be completed in 3–5 min. *Video demonstrations of the neurologic screening examination (V6) and the detailed neurologic examination (V7) can be found in the Harrison's Video Collection included in this textbook.*

Several additional points about the examination are worth noting. First, in recording observations, it is important to describe what is found rather than to apply a poorly defined medical term (e.g., "patient groans to sternal rub" rather than "obtunded"). Second, subtle CNS abnormalities are best detected by carefully comparing a patient's performance on tasks that require simultaneous activation of both cerebral hemispheres (e.g., eliciting a pronator drift of an outstretched arm with the eyes closed; extinction on one side of bilaterally applied light touch, also with eyes closed; or decreased arm swing or a slight asymmetry when walking). Third, if the patient's complaint is brought on by some activity, reproduce the activity in the office. If the complaint is of dizziness when the head is turned in one direction, have the patient do this and also look for associated signs on examination (e.g., nystagmus or dysmetria). If pain or weakness occurs after walking two blocks, have the patient leave the office and walk this distance and immediately return, and repeat the relevant parts of the examination. Finally, the use of tests that are individually tailored to the patient's problem can be of value in assessing changes over time. Tests of walking a 7.5-m (25-ft) distance (normal, 5–6 s; note assistance, if any), repetitive finger or toe tapping (normal, 20–25 taps in 5 s), or handwriting are examples.

MENTAL STATUS EXAMINATION

- *The bare minimum: During the interview, look for difficulties with communication and determine whether the patient has recall and insight into recent and past events.*

The mental status examination is underway as soon as the physician begins observing and speaking with the patient. If the history raises any concern for abnormalities of higher cortical function or if cognitive problems are observed during the interview, then detailed testing of the mental status is indicated. The patient's ability to understand the language used for the examination, cultural background, educational experience, sensory or motor problems, or comorbid conditions must be factored into the applicability of the tests and interpretation of results.

The Mini-Mental State Examination (MMSE) is a standardized screening examination of cognitive function that is extremely easy to administer and takes <10 min to complete **(Chap. 31)**. Using age-adjusted values for defining normal performance, the test is ~85% sensitive and 85% specific for making the diagnosis of dementia that is moderate or severe, especially in educated patients. When there is sufficient time available in the outpatient setting, the MMSE is one of the best methods for documenting the current mental status of the patient, and this is especially useful as a baseline assessment to which future scores of the MMSE can be compared.

Individual elements of the mental status examination can be subdivided into level of consciousness, orientation, speech and language, memory, fund of information, insight and judgment, abstract thought, and calculations.

Level of consciousness is the patient's relative state of awareness of self and the environment, and ranges from fully awake to comatose. When the patient is not fully awake, the examiner should describe the responses to the minimum stimulus necessary to elicit a reaction, ranging from verbal commands to a brief, painful stimulus such as a squeeze of the trapezius muscle. Responses that are directed toward the stimulus and signify some degree of intact cerebral function (e.g., opening the eyes and looking at the examiner or reaching to push away a painful stimulus) must be distinguished from reflex responses of a spinal origin (e.g., triple flexion response—flexion at the ankle, knee, and hip in response to a painful stimulus to the foot).

Orientation is tested by asking the person to state their name, location, and time (day of the week and date); time is usually the first to be affected in a variety of conditions.

Speech is assessed by observing articulation, rate, rhythm, and prosody (i.e., the changes in pitch and accentuation of syllables and words).

Language is assessed by observing the content of the patient's verbal and written output, response to spoken commands, and ability to read. A typical testing sequence is to ask the patient to name successively more detailed components of clothing, a watch, or a pen; repeat the phrase "No ifs, ands, or buts"; follow a three-step, verbal command; write a sentence; and read and respond to a written command.

Memory should be analyzed according to three main time scales: (1) immediate memory is assessed by saying a list of three items and having the patient repeat the list immediately; (2) short-term memory is tested by asking the patient to recall the same three items 5 and 15 min later; and (3) long-term memory is evaluated by determining how well the patient is able to provide a coherent chronologic history of their illness or personal events.

Fund of information is assessed by asking questions about major historic or current events, with special attention to educational level and life experiences.

Abnormalities of *insight and judgment* are usually detected during the interview; a more detailed assessment can be elicited by asking the patient to describe how they would respond to situations having a variety of potential outcomes (e.g., "What would you do if you found a wallet on the sidewalk?").

Abstract thought can be tested by asking the patient to describe similarities between various objects or concepts (e.g., apple and orange, desk and chair, poetry and sculpture) or to list items having the same attributes (e.g., a list of four-legged animals).

Calculation ability is assessed by having the patient carry out a computation that is appropriate to the patient's age and education (e.g., serial subtraction of 7 from 100 or 3 from 20; or word problems involving simple arithmetic).

CRANIAL NERVE EXAMINATION

- *The bare minimum: Check the fundi, visual fields, pupil size and reactivity, extraocular movements, and facial movements.*

The CNs are best examined in numerical order, except for grouping together CN III, IV, and VI because of their similar function.

CN I (Olfactory) Testing is often omitted unless there is suspicion for inferior frontal lobe disease (e.g., meningioma). With eyes closed, ask the patient to sniff a mild stimulus such as toothpaste or coffee and identify the odorant.

CN II (Optic) Check visual acuity (with eyeglasses or contact lens correction) using a Snellen chart or similar tool. Test the visual fields by confrontation, i.e., by comparing the patient's visual fields to your own. As a screening test, it is usually sufficient to examine the visual fields of both eyes simultaneously; individual eye fields should be tested if there is any reason to suspect a problem of vision by the history or other elements of the examination, or if the screening test reveals an abnormality. Face the patient at a distance of ~0.6–1.0 m (2–3 ft) and place your hands at the periphery of your visual fields in the plane that is equidistant between you and the patient. Instruct the patient to look directly at the center of your face and to indicate when and where they see one of your fingers moving. Beginning with the two inferior quadrants and then the two superior quadrants, move your index finger of the right hand, left hand, or both hands simultaneously and observe whether the patient detects the movements. A single small-amplitude movement of the finger is sufficient for a normal response. Focal perimetry and tangent screen examinations should be used to map out visual field defects fully or to search for subtle abnormalities. Optic fundi should be examined with an ophthalmoscope, and the color, size, and degree of swelling or elevation of the optic disc noted, as well as the color and texture of the retina. The retinal vessels should be checked for size, regularity, arteriovenous nicking at crossing points, hemorrhage, and exudates.

CN III, IV, VI (Oculomotor, Trochlear, Abducens) Describe the size and shape of the pupils and reaction to light and accommodation (i.e., as the eyes converge while following your finger as it moves toward the bridge of the nose). To check extraocular movements, ask the patient to keep their head still while tracking the movement of the tip of your finger. Move the target slowly in the horizontal and vertical planes; observe any paresis, nystagmus, or abnormalities of smooth pursuit (saccades, oculomotor ataxia, etc.). If necessary, the relative position of the two eyes, both in primary and multidirectional gaze, can be assessed by comparing the reflections of a bright light off both pupils. However, in practice it is typically more useful to determine whether the patient describes diplopia in any direction of gaze; true diplopia should almost always resolve with one eye closed. Horizontal nystagmus is best assessed at 45° and not at extreme lateral gaze (which is uncomfortable for the patient); the target must often be held at the lateral position for at least a few seconds to detect an abnormality.

CN V (Trigeminal) Examine sensation within the three territories of the branches of the trigeminal nerve (ophthalmic, maxillary, and mandibular) on each side of the face. As with other parts of the sensory examination, testing of two sensory modalities derived from different anatomic pathways (e.g., light touch and temperature) is sufficient for a screening examination. Testing of other modalities, the corneal reflex, and the motor component of CN V (jaw clench—masseter muscle) is indicated when suggested by the history.

CN VII (Facial) Look for facial asymmetry at rest and with spontaneous movements. Test eyebrow elevation, forehead wrinkling, eye closure, smiling, and cheek puff. Look in particular for differences in the lower versus upper facial muscles; weakness of the lower two-thirds of the face with preservation of the upper third suggests an upper motor neuron lesion, whereas weakness of an entire side suggests a lower motor neuron lesion.

CN VIII (Vestibulocochlear) Check the patient's ability to hear a finger rub or whispered voice with each ear. Further testing for air versus mastoid bone conduction (Rinne) and lateralization of a 512-Hz tuning fork placed at the center of the forehead (Weber) should be done if an abnormality is detected by history or examination. Any suspected problem should be followed up with formal audiometry. **For further discussion of assessing vestibular nerve function in the setting of dizziness, coma, or hearing loss, see Chaps. 24, 30, and 36, respectively**.

CN IX, X (Glossopharyngeal, Vagus) Observe the position and symmetry of the palate and uvula at rest and with phonation ("aah"). The pharyngeal ("gag") reflex is evaluated by stimulating the posterior pharyngeal wall on each side with a sterile, blunt object (e.g., tongue blade), but the reflex may be absent in normal individuals.

CN XI (Spinal Accessory) Check shoulder shrug (trapezius muscle) and head rotation to each side (sternocleidomastoid) against resistance.

CN XII (Hypoglossal) Inspect the tongue for atrophy or fasciculations, position with protrusion, and strength when extended against the inner surface of the cheeks on each side.

MOTOR EXAMINATION

- *The bare minimum: Look for muscle atrophy and check extremity tone. Assess upper extremity strength by checking for pronator drift and strength of wrist or finger extensors. Assess lower extremity strength by checking strength of the toe extensors.*

The motor examination includes observations of muscle appearance, tone, and strength. Although gait is in part a test of motor function, it is usually evaluated separately at the end of the examination.

Appearance Inspect and palpate muscle groups under good light and with the patient in a comfortable and symmetric position. Check for muscle fasciculations, tenderness, and atrophy or hypertrophy. Involuntary movements may be present at rest (e.g., tics, myoclonus, choreoathetosis, pill-rolling tremor of Parkinson's disease), during maintained posture (essential tremor), or with voluntary movements (intention tremor of cerebellar disease or familial tremor).

Tone Muscle tone is tested by measuring the resistance to passive movement of a relaxed limb. Patients often have difficulty relaxing during this procedure, so it is useful to distract the patient to minimize active movements. In the upper limbs, tone is assessed by rapid pronation and supination of the forearm and flexion and extension at the wrist. In the lower limbs, while the patient is supine, the examiner's hands are placed behind the knees and rapidly raised; with normal tone, the ankles drag along the table surface for a variable distance before rising, whereas increased tone results in an immediate lift of the heel off the surface. Decreased tone is most commonly due to lower motor neuron or peripheral nerve disorders. Increased tone may be evident as spasticity (resistance determined by the angle and velocity of motion; corticospinal tract disease), rigidity (similar resistance in all angles of motion; extrapyramidal disease), or paratonia (fluctuating changes in resistance; frontal lobe pathways; or normal difficulty in relaxing). Cogwheel rigidity, in which passive motion elicits jerky interruptions in resistance, is seen in parkinsonism.

Strength Testing for pronator drift is an extremely useful method for screening upper limb weakness. The patient is asked to hold both arms fully extended and parallel to the ground with eyes closed. This position should be maintained for ~10 s; any flexion at the elbow or fingers or pronation of the forearm, especially if asymmetric, is a sign of potential weakness. Patients with shoulder pain or a limited range of motion may have an apparent pronator drift that is not due to true weakness. Muscle strength is further assessed by having the patient exert maximal effort for the particular muscle or muscle group being tested. It is important to isolate the muscles as much as possible, i.e., hold the limb so that only the muscles of interest are active. It is also helpful to palpate accessible muscles as they contract. Grading muscle strength and evaluating the patient's effort are an art that takes time and practice. Muscle strength is traditionally graded using the following scale:

0 = no movement
1 = flicker or trace of contraction but no associated movement at a joint
2 = movement with gravity eliminated
3 = movement against gravity but not against resistance
4– = movement against a mild degree of resistance
4 = movement against moderate resistance
4+ = movement against strong resistance
5 = full power

However, in many cases, it is more practical to use the following terms:

Paralysis = no movement
Severe weakness = movement with gravity eliminated
Moderate weakness = movement against gravity but not against mild resistance
Mild weakness = movement against moderate resistance
Full strength

Noting the pattern of weakness is as important as assessing the magnitude of weakness. Unilateral or bilateral weakness of the upper limb extensors and lower limb flexors ("pyramidal weakness") suggests a lesion of the pyramidal tract, bilateral proximal weakness suggests myopathy, and bilateral distal weakness suggests peripheral neuropathy.

REFLEX EXAMINATION

- *The bare minimum: Check the biceps, patellar, and Achilles reflexes.*

Muscle Stretch Reflexes Those that are typically assessed include the biceps (C5, C6), brachioradialis (C5, C6), triceps (C6, C7), and sometimes finger flexor (C8, T1) reflexes in the upper limbs and the patellar or quadriceps (L3, L4) and Achilles (S1, S2) reflexes in the lower limbs. The patient should be relaxed and the muscle positioned midway between full contraction and extension. Reflexes may be enhanced by asking the patient to voluntarily contract other, distant muscle groups (Jendrassik maneuver). For example, upper limb reflexes may be reinforced by voluntary teeth-clenching, and the Achilles reflex by hooking the flexed fingers of the two hands together and attempting to pull them apart. For each reflex tested, the two sides should be tested sequentially, and it is important to determine the smallest stimulus required to elicit a reflex rather than the maximum response. Reflexes are graded according to the following scale:

0 = absent
1 = present but diminished
2 = normoactive
3 = increased
4 = clonus

Cutaneous Reflexes The plantar reflex is elicited by stroking, with a noxious stimulus such as a tongue blade, the lateral surface of the sole of the foot beginning near the heel and moving across the ball of the foot to the great toe. The normal reflex consists of plantar flexion of the toes. With upper motor neuron lesions above the S1 level of the spinal cord, a paradoxical extension of the toe is observed, associated with fanning and extension of the other toes (termed an *extensor plantar response*, or *Babinski sign*). However, despite its popularity, the reliability and validity of the Babinski sign for identifying upper motor neuron weakness are limited—it is far more useful to rely on tests of tone, strength, stretch reflexes, and coordination. Superficial abdominal reflexes are elicited by gently stroking the abdominal surface near the umbilicus in a diagonal fashion with a sharp object (e.g., the wooden end of a cotton-tipped swab) and observing the movement of the umbilicus. Normally, the umbilicus will pull toward the stimulated quadrant. With upper motor neuron lesions, these reflexes are absent. They are most helpful when there is preservation of the upper (spinal cord level T9) but not lower (T12) abdominal reflexes, indicating a spinal lesion between T9 and T12, or when the response is asymmetric. Other useful cutaneous reflexes include the cremasteric (ipsilateral elevation of the testicle following stroking of the medial thigh; mediated by L1 and L2) and anal (contraction of the anal sphincter when the perianal skin is scratched; mediated by S2, S3, S4) reflexes. It is particularly important to test for these reflexes in any patient with suspected injury to the spinal cord or lumbosacral roots.

Primitive Reflexes With disease of the frontal lobe pathways, several primitive reflexes not normally present in the adult may appear. The suck response is elicited by lightly touching with a tongue blade the center of the lips, and the root response the corner of the lips; the patient will move the lips to suck or root in the direction of the stimulus. The grasp reflex is elicited by touching the palm between the thumb and index finger with the examiner's fingers; a positive response is a forced grasp of the examiner's hand. In many instances, stroking the back of the hand will lead to its release. The palmomental response is contraction of the mentalis muscle (chin) ipsilateral to a scratch stimulus diagonally applied to the palm.

SENSORY EXAMINATION

- *The bare minimum: Ask whether the patient can feel light touch and the temperature of a cool object in each distal extremity. Check double simultaneous stimulation using light touch on the hands. Perform the Romberg maneuver.*

Evaluating sensation is usually the most unreliable part of the examination because it is subjective and is difficult to quantify. In the compliant and discerning patient, the sensory examination can be extremely helpful for the precise localization of a lesion. With patients who are uncooperative or lack an understanding of the tests, it may be useless. The examination should be focused on the suspected lesion. For example, in spinal cord, spinal root, or peripheral nerve abnormalities, all major sensory modalities should be tested while looking for a pattern consistent with a spinal level and dermatomal or nerve distribution. In patients with lesions at or above the brainstem, screening the primary sensory modalities in the distal extremities along with tests of "cortical" sensation is usually sufficient.

The five primary sensory modalities—light touch, pain, temperature, vibration, and joint position—are tested in each limb. Light touch is assessed by stimulating the skin with single, very gentle touches of the examiner's finger or a wisp of cotton. Pain is tested using a new pin, and temperature is assessed using a metal object (e.g., tuning fork) that has been immersed in cold and warm water. Vibration is tested using a 128-Hz tuning fork applied to the distal phalanx of the great toe or index finger just below the nail bed. By placing a finger on the opposite side of the joint being tested, the examiner compares the patient's threshold of vibration perception with their own. For joint position testing, the examiner grasps the digit or limb laterally and distal to the joint being assessed; small 1- to 2-mm excursions can usually be sensed. The Romberg maneuver is primarily a test of proprioception. The patient is asked to stand with the feet as close together as necessary to maintain balance while the eyes are open, and the eyes are then closed. A loss of balance with the eyes closed is an abnormal response.

"Cortical" sensation is mediated by the parietal lobes and represents an integration of the primary sensory modalities; testing cortical sensation is only meaningful when primary sensation is intact. Double simultaneous stimulation is especially useful as a screening test for cortical function; with the patient's eyes closed, the examiner lightly touches one or both hands and asks the patient to identify the stimuli. With a parietal lobe lesion, the patient may be unable to identify the stimulus on the contralateral side when both hands are touched. Other modalities relying on the parietal cortex include the discrimination of two closely placed stimuli as separate (two-point discrimination), identification of an object by touch and manipulation alone (stereognosis), and the identification of numbers or letters written on the skin surface (graphesthesia).

COORDINATION EXAMINATION

- *The bare minimum: Observe the patient at rest and during spontaneous movements. Test rapid alternating movements of the hands and feet and finger to nose.*

Coordination refers to the orchestration and fluidity of movements. Even simple acts require cooperation of agonist and antagonist muscles, maintenance of posture, and complex servomechanisms to control the rate and range of movements. Part of this integration relies on normal function of the cerebellar and basal ganglia systems. However, coordination also requires intact muscle strength and kinesthetic and proprioceptive information. Thus, if the examination has disclosed abnormalities of the motor or sensory systems, the patient's coordination should be assessed with these limitations in mind.

Rapid alternating movements in the upper limbs are tested separately on each side by having the patient make a fist, partially extend the index finger, and then tap the index finger on the distal thumb as quickly as possible. In the lower limb, the patient rapidly taps the foot against the floor or the examiner's hand. If these rapid alternating movements are imprecise or vary in amplitude or rhythm, a cerebellar lesion is suspected; if, however, they are slow compared with the other side, a lesion of the pyramidal tract is most likely. Finger-to-nose testing is primarily a test of cerebellar function; the patient is asked to touch their index finger repetitively to the nose and then to the examiner's outstretched finger, which moves with each repetition. A similar test in the lower extremity is to have the patient raise the leg and touch the examiner's finger with the great toe. Another coordination test in the lower limbs is the heel-knee-shin maneuver; in the supine position, the patient is asked to slide the heel of each foot from the knee down the shin of the other leg. For all these movements, the accuracy, speed, and rhythm are noted.

GAIT EXAMINATION

- *The bare minimum: Observe the patient while walking normally, on the heels and toes, and along a straight line.*

Watching the patient walk is the most important part of the neurologic examination. Normal gait requires that multiple systems—including strength, sensation, and coordination—function in a highly integrated fashion. Unexpected abnormalities may be detected that prompt the examiner to return in more detail to other aspects of the examination. The patient should be observed while walking and turning normally, walking on the heels, walking on the toes, and walking heel-to-toe along a straight line. The examination may reveal decreased arm swing on one side (corticospinal tract disease), a stooped posture and short-stepped gait (parkinsonism), a broad-based unstable gait (ataxia), scissoring (spasticity), or a high-stepped, slapping gait (posterior column or peripheral nerve disease), or the patient may appear to be stuck in place (apraxia with frontal lobe disease).

NEUROLOGIC DIAGNOSIS

The clinical data obtained from the history and examination are interpreted to arrive at an anatomic localization that best explains the clinical findings **(Table 433-1)**, to narrow the list of diagnostic possibilities, and to select the laboratory tests most likely to be informative. The laboratory assessment may include (1) serum electrolytes; complete blood count; and renal, hepatic, endocrine, and immune studies; (2) cerebrospinal fluid examination; (3) focused neuroimaging studies **(Chap. 434)**; or (4) electrophysiologic studies. The anatomic localization, mode of onset and course of illness, other medical data, and laboratory findings are then integrated to establish an etiologic diagnosis.

The neurologic examination may be normal even in patients with a serious neurologic disease, such as seizures, chronic meningitis, or a TIA. A comatose patient may arrive with no available history, and in such cases, the approach is as described in **Chap. 30**. In other patients, an inadequate history may be overcome by a succession of examinations from which the course of the illness can be inferred. In perplexing cases, it is useful to remember that uncommon presentations of common diseases are more likely than rare etiologies. Thus, even in

TABLE 433-1 Findings Helpful for Localizations within the Nervous System

	SIGNS
Cerebrum	Abnormal mental status or cognitive impairment Seizures Unilateral weakness[a] and sensory abnormalities including head and limbs Visual field abnormalities Movement abnormalities (e.g., diffuse incoordination, tremor, chorea)
Brainstem	Isolated cranial nerve abnormalities (single or multiple) "Crossed" weakness[a] and sensory abnormalities of head and limbs, e.g., weakness of right face and left arm and leg
Spinal cord	Back pain or tenderness Weakness[a] and sensory abnormalities sparing the head Mixed upper and lower motor neuron findings Sensory level Sphincter dysfunction
Spinal roots	Radiating limb pain Weakness[b] or sensory abnormalities following root distribution **(see Figs. 27-2 and 27-3)** Loss of reflexes
Peripheral nerve	Mid or distal limb pain Weakness[b] or sensory abnormalities following nerve distribution **(see Figs. 27-2 and 27-3)** "Stocking or glove" distribution of sensory loss Loss of reflexes
Neuromuscular junction	Bilateral weakness including face (ptosis, diplopia, dysphagia) and proximal limbs Increasing weakness with exertion Sparing of sensation
Muscle	Bilateral proximal or distal weakness Sparing of sensation

[a] Weakness along with other abnormalities having an "upper motor neuron" pattern, i.e., spasticity, weakness of extensors > flexors in the upper extremity and flexors > extensors in the lower extremity, and hyperreflexia. [b] Weakness along with other abnormalities having a "lower motor neuron" pattern, i.e., flaccidity and hyporeflexia.

tertiary care settings, multiple strokes are usually due to emboli and not vasculitis, and dementia with myoclonus is usually Alzheimer's disease and not a prion disorder or a paraneoplastic illness. Finally, the most important task of a primary care physician faced with a patient who has a new neurologic complaint is to assess the urgency of referral to a specialist. Here, the imperative is to rapidly identify patients likely to have nervous system infections, acute strokes, and spinal cord compression or other treatable mass lesions and arrange for immediate care.

■ FURTHER READING

Brazis P et al: *Localization in Clinical Neurology*, 8th ed. Philadelphia, Lippincott William & Wilkins, 2021.

Campbell WW, Barohn RJ: *DeJong's The Neurological Examination*, 8th ed. Philadelphia, Lippincott William & Wilkins, 2019.

Feigin VL et al: The global burden of neurological disorders: Translating evidence into policy. Lancet Neurol 19:255, 2020.

GBD 2019 Diseases and Injuries Collaborators: Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. Lancet 396:1204, 2020.

O'Brien M: *Aids to the Examination of the Peripheral Nervous System*, 6th ed. Elsevier, Amsterdam, 2023.

434 Neuroimaging in Neurologic Disorders

William P. Dillon

Numerous noninvasive imaging options are available to clinicians evaluating patients with neurologic disorders. These include computed tomography (CT) and magnetic resonance (MR) imaging (MRI), plus their variations, including CT angiography (CTA); perfusion CT (pCT); dual-energy CT; photon counting CT; MR angiography (MRA); MR vessel wall imaging; functional MRI (fMRI); MR spectroscopy (MRS); MR neurography (MRN); diffusion-weighted MR imaging (DWI); diffusion tensor MR imaging (DTI); susceptibility-weighted MR imaging (SWI); arterial spin label imaging (ASL); and perfusion MRI (pMRI). Furthermore, a number of interventional neuroradiologic techniques have matured including catheter embolization, stent retrieval thrombectomy, aneurysm coiling and stenting, as well as numerous techniques for spine disorders including CT and fluoroscopic myelography and CT-guided spine procedures for diagnosing and treating pain and oncology, including dynamic CT myelography, radiofrequency and cold ablation techniques, image-guided blood and fibrin glue patches and transvenous embolization of cerebrospinal fluid (CSF) venous fistulae. Multidetector CTA (MDCTA) and gadolinium-enhanced MRA techniques have reduced the need for catheter-based angiography, which is now reserved for patients in whom small-vessel detail is essential for diagnosis or for whom concurrent interventional therapy is planned **(Table 434-1)**.

TABLE 434-1 Guidelines for the Use of CT, Ultrasound, and MRI

CONDITION	RECOMMENDED TECHNIQUE
Hemorrhage	
Acute parenchymal	CT, MR
Subacute/chronic	MRI
Subarachnoid hemorrhage	CT, CTA, lumbar puncture → angiography
Aneurysm	Angiography > CTA, MRA
Chronic subarachnoid blood	MR with SWI
Ischemic infarction	
Hemorrhagic infarction	CT or MRI
Bland infarction	MRI with diffusion > CT, CTA, angiography
Carotid or vertebral dissection	MRI/MRA> CTA
Vertebral basilar insufficiency	CTA, MRI with DWI MRA
Carotid stenosis	CTA, MRA > US
Suspected mass lesion	
Neoplasm, primary or metastatic	MRI ± contrast
Infection/abscess	MRI ± contrast
Immunosuppressed with focal findings	MRI ± contrast
Vascular malformation	MRI ± angiography
White matter disorders	MRI
Acute demyelinating disease	MRI ± contrast
Dementia	MRI > CT, contrast if mass
Trauma	
Acute trauma	CT
Shear injury/chronic hemorrhage	MRI + SWI
Headache/migraine	MRI > CT
Seizure	
First time, no focal neurologic deficits	MRI > CT
With neurologic deficit, or immunocompromised or cancer	MRI ± contrast > CT
Partial complex/refractory	MRI
Cranial neuropathy	MRI ± contrast
Meningeal disease	MRI ± contrast
Spine	
Low-back pain	
No neurologic deficits	MRI or CT after >6 weeks
With focal deficits	MRI > CT
Spinal stenosis	MRI or CT
Cervical spondylosis	MRI, CT, CT myelography
Infection	MRI ± contrast CT
Myelopathy	MRI ± contrast
Arteriovenous malformation	MRI ± contrast angiography

Abbreviations: CT, computed tomography; CTA, CT angiography; MRA, magnetic resonance angiography; MRI, magnetic resonance imaging; SWI, susceptibility-weighted imaging.

In general, MRI is more sensitive than CT for the detection of lesions affecting the peripheral and central nervous system (CNS). Diffusion MR, a sequence sensitive to the microscopic motion of water, is the most sensitive technique for detecting acute ischemic stroke of the brain or spinal cord and is also useful in the detection and characterization of encephalitis, abscess, Creutzfeldt-Jacob disease, cerebral tumors, and acute demyelinating lesions. CT, however, is acquired more quickly, making it a pragmatic choice for uncooperative patients, or those with suspected acute stroke, hemorrhage, and acute intracranial or spinal trauma. CT is also more sensitive than MRI for visualizing fine osseous detail and thus is appropriate for the initial imaging evaluation of conductive hearing loss, and lesions affecting the osseous skull and spine. MR may, however, add important diagnostic information regarding bone marrow infiltrative processes that can be difficult to detect on CT.

COMPUTED TOMOGRAPHY

■ TECHNIQUE

The CT image is a cross-sectional representation of anatomy created by a computer-generated analysis of the attenuation of x-ray beams passed through a section of the body. The x-ray beam, collimated to the desired slice width, rotates around the patient and passes through selected regions in the body. X-rays are variably attenuated by body structures and converted into light photons by ceramic scintillators as part of the detectors aligned 180° from the x-ray tube. These photons

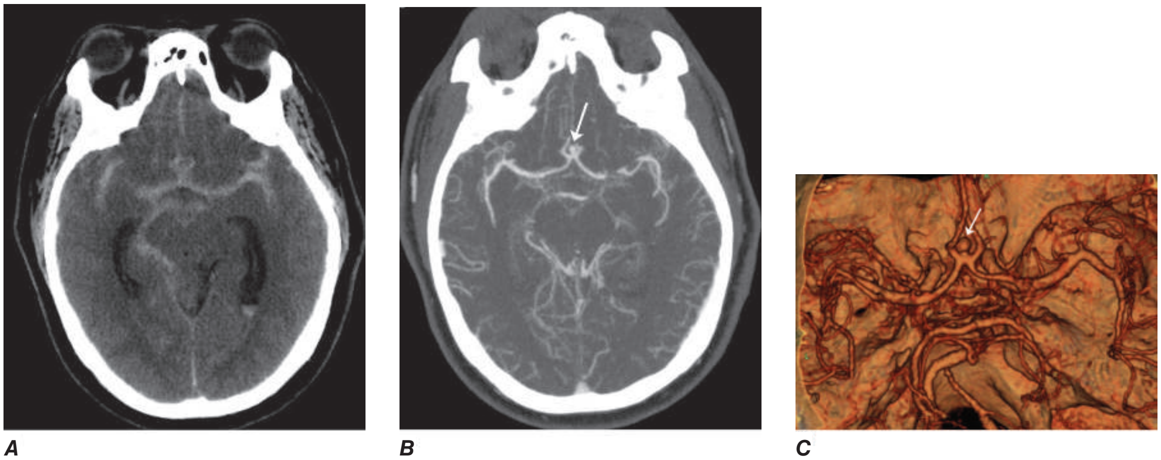

FIGURE 434-1 Computed tomography (CT) angiography (CTA) of ruptured anterior cerebral artery aneurysm in a patient presenting with acute headache. ***A.*** Noncontrast CT demonstrates subarachnoid and intraventricular hemorrhage and mild obstructive hydrocephalus. ***B.*** Axial maximum-intensity projection from CTA demonstrates enlargement of the anterior cerebral artery (*arrow*). ***C.*** Three-dimensional surface reconstruction using a workstation confirms the anterior cerebral aneurysm and demonstrates its orientation and relationship to nearby vessels (*arrow*). CTA image is produced by 0.5- to 1-mm helical CT scans performed during a rapid bolus infusion of IV contrast medium.

are then converted into electric signals. A computer calculates a "back projection" image from the 360° x-ray attenuation profile. Greater x-ray attenuation (e.g., as caused by bone) results in areas of high "density" (whiter) on the scan, whereas soft-tissue structures that have poor attenuation of x-rays, such as organs and air-filled cavities, are lower (gray-black) in density. The resolution of an image depends on the radiation dose, the detector size, collimation (slice thickness), the field of view, and the matrix size of the display. A modern CT scanner can obtain sections as thin as 0.5–1 mm with 0.4-mm in-plane resolution at a speed of 0.3 s per rotation; complete studies of the brain are completed in 1–10 s.

Multidetector CT (MDCT) is now standard. Single or multiple (from 4 to 320) solid-state ceramic detectors positioned opposite to the x-ray source result in multiple slices per revolution of the beam around the patient. In helical mode, the table moves continuously through the rotating x-ray beam, generating a continuous "helix" of x-ray information that is reformatted into various slice thicknesses and planes. Advantages of MDCT include shorter scan times, reduced patient and organ motion, and the ability to acquire images dynamically during the infusion of intravenous (IV) contrast, the basis of CTA and CT perfusion **(Figs. 434-1*B* and *C*)**. CTA is displayed in three dimensions to yield angiogram-like images **(Figs. 434-1*C*, 434-2*E* and *F*, and see Fig. 438-3)**. However, the detectors, while numerous, have inefficiencies based on their physical architecture. New so-called photon-counting CT scanners use different technology for detectors, converting x-ray directly into an electric signal, rather than photon light, resulting in improved resolution, reduced electronic noise and thus dose reduction, improved spectral energy detection that reduces calcium and bone artifact, and improved contrast-to-noise ratio, which permits a reduction in the dose of contrast material required. Photon counting scanners are now able to produce CT images with 0.2 mm resolution often at reduced radiation dose **(Fig. 434-3)**.

IV iodinated contrast is used to identify vascular structures and to detect defects in the blood-brain barrier (BBB) that are caused by tumors, infarcts, and infections. In the normal CNS, only vessels and structures lacking a BBB (e.g., the pituitary gland, choroid plexus, and dura) enhance after contrast administration. While helpful in characterizing mass lesions as well as essential for the acquisition of CTA studies, the decision to use contrast material should always be considered carefully as it carries a small risk of allergic reaction and adds additional expense.

■ INDICATIONS

CT is the primary study of choice in the evaluation of an acute change in mental status, focal neurologic findings, acute trauma to the brain and spine, suspected subarachnoid hemorrhage, and conductive hearing loss (Table 434-1). CT often is complementary to MR in the evaluation of the skull base, orbit, and osseous structures of the spine. In the spine, CT is useful in evaluating patients with osseous spinal stenosis and spondylosis, as well as in patients with failed back surgery; however, MRI is often preferred when neurologic deficits are present. CT is often acquired following intrathecal contrast injection to evaluate for spinal and intracranial CSF fistula, as well as the spinal subarachnoid space (*CT myelography*) in failed back surgery syndromes.

■ COMPLICATIONS

CT is safe, fast, and reliable. Radiation exposure depends on the dose used but is normally 2–5 mSv (millisievert) for a routine brain CT study. In comparison, the average American receives about 6.2 mSv per year from all sources of radiation, including 3.1 mSv from naturally occurring sources and a similar amount from man-made sources and applications such as jet plane trips, smoke detectors, and medical imaging. While ionizing radiation can potentially induce damage to DNA, such risks are felt to be minimal at normal background radiation levels, especially at the low-dose levels related to diagnostic imaging. The risk of harm from radiation depends on the amount of dose, the delivery rate, the type of radiation, sensitivity of the tissue exposed, and the gender and age and health of the person exposed. Cancers that might develop from a radiation exposure usually have a latency period of 2–10 years or longer, after exposure. Nevertheless, for all patients, especially children, it is important to use as low a radiation dose as possible for diagnostic imaging purposes. Where feasible, MR or ultrasound is preferred. With the advent of MDCT, CTA, and CT perfusion, the diagnostic benefit must always be weighed against any increased radiation exposure. Advances in postprocessing software now permit acceptable diagnostic CT scans at 30–40% lower radiation doses compared with prior technology.

The most frequent CT-related complications are those associated with use of IV contrast agents. Ionic contrast agents have been largely replaced by safer nonionic compounds.

Contrast-associated acute kidney injury (CA-AKI) is a general term used to describe a sudden deterioration in renal function that occurs within 48 h following the intravascular administration of iodinated contrast medium. This rare complication appears to be more frequent with intraarterial injections during coronary angiography than with IV administration. CA-AKI from gadolinium-based contrast media probably does not occur or is exceptionally rare. CA-AKI is a correlative diagnosis and may result from hemodynamic changes, renal tubular obstruction and cell damage, or immunologic reactions to contrast agents. Although there is no accepted definition of CA-AKI, the diagnosis of AKI is made according to the Kidney Disease Improving Global Outcomes (KDIGO) criteria if one of the following occurs

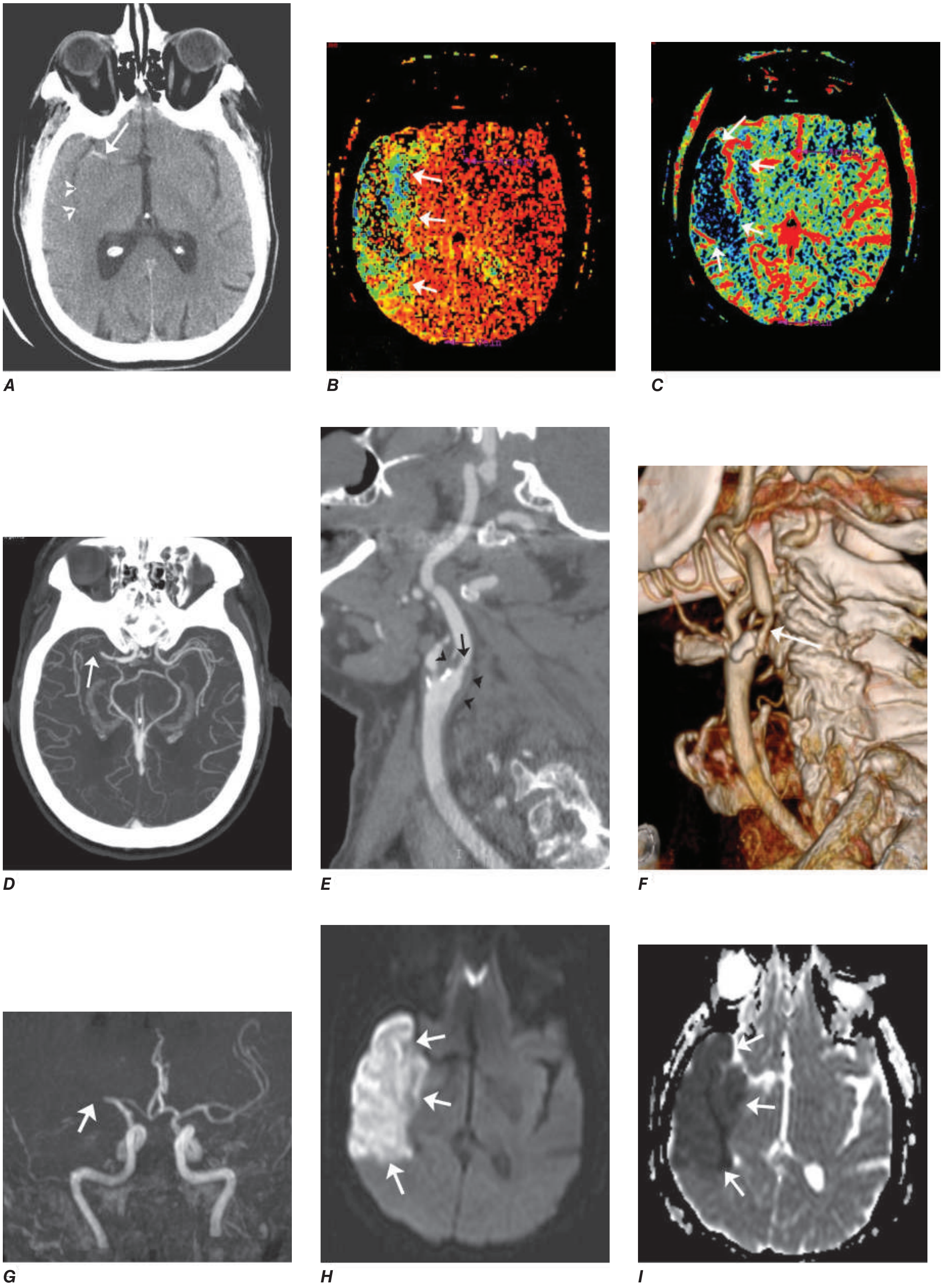

FIGURE 434-2 Acute left hemiparesis due to right middle cerebral artery occlusion. ***A.*** Axial noncontrast computed tomography (CT) scan demonstrates high density within the right middle cerebral artery (*arrow*) associated with subtle low density involving the right putamen (*arrowheads*). ***B.*** Mean transit time CT perfusion parametric map indicating prolonged mean transit time involving the right middle cerebral territory (*arrows*). ***C.*** Cerebral blood volume (CBV) map shows reduced CBV involving an area within the defect shown in ***B***, indicating a high likelihood of infarction (*arrows*). ***D.*** Axial maximum-intensity projection from a CT angiography (CTA) study through the circle of Willis demonstrates an abrupt occlusion of the proximal right middle cerebral artery (*arrow*). ***E.*** Sagittal reformation through the right internal carotid artery demonstrates a low-density lipid-laden plaque (*arrowheads*) narrowing the lumen (*black arrow*). ***F.*** Three-dimensional surface-rendered CTA image demonstrates calcification and narrowing of the right internal carotid artery (*arrow*), consistent with atherosclerotic disease. ***G.*** Coronal maximum-intensity projection from magnetic resonance angiography shows right middle cerebral artery (MCA) occlusion (*arrow*). ***H.*** and ***I.*** Axial diffusion-weighted image (***H***) and apparent diffusion-coefficient image (***I***) document the presence of a right middle cerebral artery infarction.

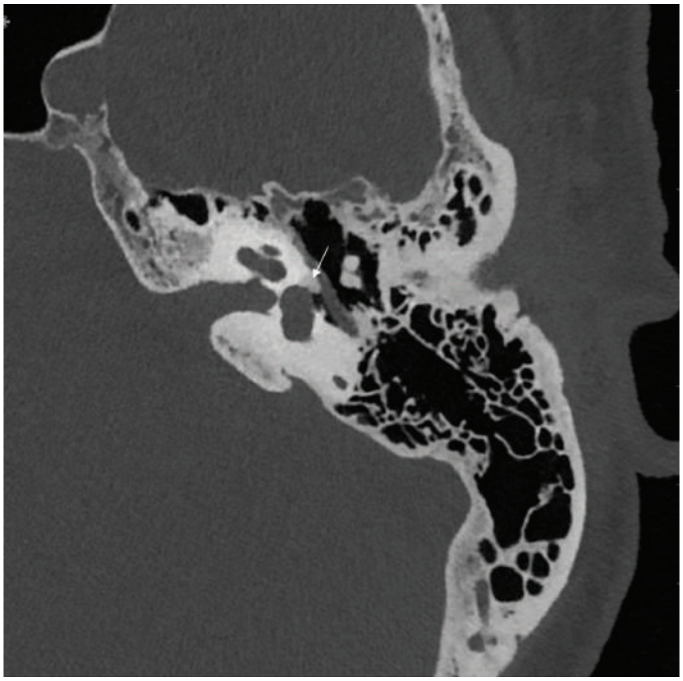

FIGURE 434-3 Photon-counting CT of the temporal bone. This 0.2-mm-thick image was obtained on a photon-counting CT scanner. It demonstrates an otosclerosis lesion (*arrow*) at the oval window in a patient with mixed hearing loss. The resolution of this image exceeds that of conventional 64-slice CT scanners (0.625 mm) due to improved detectors (see explanation in chapter). *(Courtesy of Ian Mark, MD, and the Mayo Clinic.)*

within 48 h after a nephrotoxic event (e.g., intravascular iodinated contrast medium exposure):

1. Absolute serum creatinine increase ≥0.3 mg/dL (>26.4 μmol/L)
2. A percentage increase in serum creatinine ≥50% (≥1.5-fold above baseline)
3. Urine output reduced to ≤0.5 mL/kg per h for at least 6 h

However, other causes of acute renal failure must be excluded. The prognosis is usually favorable, with serum creatinine levels returning to baseline within 1–2 weeks. Risk factors for contrast nephropathy are also controversial, but consensus is that the most important risk factor is preexisting severe renal insufficiency. Other proposed risk factors include diabetes mellitus, dehydration, cardiovascular disease, diuretic use, advanced age, hypertension, hyperuricemia, and multiple iodinated contrast medium doses in a short time interval (<24 h), but these have not been rigorously confirmed. Patients with diabetes and those with mild renal failure should be well hydrated prior to the administration of contrast agents; careful consideration should be given to alternative imaging techniques. Estimated glomerular filtration rate (eGFR) is a more reliable indicator of renal function compared to creatinine alone because it considers age and sex in the calculation. In one study, 15% of outpatients with a normal serum creatinine had an eGFR of ≤50 mL/min per 1.73 m^2 (normal is ≥90 mL/min per 1.73 m^2). The exact eGFR threshold, below which withholding IV contrast should be considered, is also controversial. The risk of contrast nephropathy is minimal in patients with eGFR >30 mL/min per 1.73 m^2; however, most of these patients will have only a temporary rise in creatinine. The risk of dialysis after receiving contrast significantly increases in patients with eGFR <30 mL/min per 1.73 m^2. Little evidence exists that IV iodinated contrast material is an independent risk factor for acute kidney injury in patients with eGFR ≥30 mL/min per 1.73 m^2. The American College of Radiology suggests, if a threshold for risk is used at all, an eGFR of <30 mL/min per 1.73 m^2 seems to have the greatest level of evidence. If contrast must be administered to a patient with an eGFR <30 mL/min per 1.73 m^2, the patient should be well hydrated. The decision to administer closely spaced contrast-enhanced studies is a clinical one, as there is insufficient data on this topic. High-risk patients should, however, be treated with greater caution than the general population. Use of other agents such as bicarbonate and acetylcysteine were previously thought to be protective against CA-AKI; however, recent meta-analyses have failed to show protection over normal saline. Patients with renal failure who require contrast administration should not have acute dialysis or continuous renal replacement therapy initiated, or alter their schedule due to the risks, costs, and lack of benefit.

Below are suggested guidelines for creatinine testing prior to contrast administration. If a recent serum creatinine (within 60 days in most clinical settings) is not available, creatinine testing should be performed *if* the patient has *any* of the following risk factors:

- Age >60 years or hypertension (however, this results in a large false-positive rate for those with eGFR <30 mL/min per 1.73 m^2)
- Personal history of "kidney disease" as an adult, including kidney surgery, ablation, transplant, or prior dialysis
- Diabetes mellitus treated with insulin or other prescribed medications
- Metformin or metformin-containing drug combinations
- Collagen vascular disease (e.g., systemic lupus erythematosus [SLE], scleroderma, rheumatoid arthritis)

Allergy Immediate reactions following IV contrast media occur through several mechanisms. The most severe are related to allergic hypersensitivity (anaphylaxis) and range from mild hives to bronchospasm and death. The pathogenesis of allergic hypersensitivity reactions is thought to include the release of mediators such as histamine, antibody-antigen reactions, and complement activation. Severe allergic reactions occur in ~0.04% of patients receiving nonionic media, sixfold lower than with ionic media. Risk factors include a history of prior contrast reaction (fivefold increased likelihood), food and or drug allergies, and atopy (asthma and hay fever). The predictive value of specific allergies, such as those to shellfish, once thought important, is now recognized to be unreliable. Nonetheless, in patients with a history worrisome for potential allergic reaction, a noncontrast CT or MRI procedure should be considered as an alternative to contrast administration. If iodinated contrast is absolutely required, a nonionic agent should be used in conjunction with pretreatment with glucocorticoids and antihistamines **(Table 434-2)**; however, pretreatment does not guarantee safety. Patients with allergic reactions to iodinated contrast material do not usually react to gadolinium-based MR contrast material, although such reactions can occur. It would be wise to pretreat patients with a prior allergic history to MR contrast administration in a similar fashion. Subacute (>1 h after injection) reactions are frequent and probably related to T cell–mediated immune reactions. These are typically urticarial but can occasionally be more severe. Additives or contaminants, such as calcium-chelating substances or substances eluted from rubber stoppers in bottles or syringes, may be contributory in some allergy-like contrast reactions. Drug provocation and skin testing may be required to determine both the culprit agent involved and a safe alternative.

Other side effects of CT contrast include a sensation of warmth throughout the body and a metallic taste during IV administration. Extravasation of contrast media, although rare, can be painful and lead to a compartment syndrome. When this occurs, immediate consultation with plastic surgery is indicated. Patients with significant cardiac disease may be at increased risk for contrast reactions, and in these patients, limits to the volume and osmolality of the contrast media should be considered. Patients who may undergo systemic radioactive iodine therapy for thyroid disease or cancer should not receive iodinated contrast media, if possible, because this will decrease the uptake of the radioisotope into the tumor or thyroid (see the *American College of Radiology Manual on Contrast Media*, 2023; *https://www.acr.org/-/media/ACR/Files/Clinical-Resources/Contrast_Media.pdf*).

MAGNETIC RESONANCE IMAGING

TECHNIQUE

MRI is a complex interaction between protons in biologic tissues, a static magnetic field (the magnet), and energy in the form of radiofrequency (Rf) waves of a specific frequency introduced by coils placed next to the body part of interest. Images are made by computerized processing of resonance information received from protons (typically hydrogen)

TABLE 434-2 Guidelines for Premedication of Patients with Prior Contrast Allergy

13 h prior to examination:

Prednisone, 50 mg PO *or* methylprednisolone, 32 mg PO

7 h prior to examination:

Prednisone, 50 mg PO *or* methylprednisolone, 32 mg PO

1 h prior to examination:

Prednisone, 50 mg PO

Diphenhydramine, 50 mg intravenously, intramuscularly, or by mouth (optional)

Immediately prior to examination:

Benadryl, 50 mg IV (alternatively, can be given PO 2 h prior to exam)

If a patient is unable to take oral medication, 200 mg hydrocortisone IV for each dose of oral prednisone may be used

If a patient is allergic to diphenhydramine in a situation where diphenhydramine would otherwise be considered, an alternate antihistamine without cross-reactivity may be considered, or the antihistamine portion of the regimen may be removed

Accelerated IV Premedication

1. Methylprednisolone sodium succinate (e.g., Solu-Medrol) 40 mg IV

   or

   Hydrocortisone sodium succinate (e.g., Solu-Cortef) 200 mg IV immediately, and then every 4 h until contrast medium administration.

   plus

   Diphenhydramine 50 mg IV 1 h before contrast medium administration. This regimen is usually 4–5 h in duration.
2. Dexamethasone sodium sulfate (e.g., Decadron) 7.5 mg IV immediately, and then every 4 h until contrast medium administration

   plus

   Diphenhydramine 50 mg IV 1 h before contrast medium administration. This regimen may be useful in patients with an allergy to methylprednisolone and is also usually 4–5 h in duration.
3. Methylprednisolone sodium succinate (e.g., Solu-Medrol) 40 mg IV or hydrocortisone sodium succinate (e.g., Solu-Cortef) 200 mg IV

   plus

   Diphenhydramine 50 mg IV, each 1 h before contrast medium administration. This regimen, and all other regimens with a duration <4–5 h, has no evidence of efficacy. It may be considered in emergent situations when there are no alternatives.

Note: Premedication regimens <4–5 h in duration (oral or IV) have not been shown to be effective.

in the body. Field strength of the magnet is directly related to signal-to-noise ratio. While 1.5-Tesla (T) and 3-T magnets are now widely available and have distinct advantages in the brain and musculoskeletal systems, even higher field magnets (7 and 11+ T) and positron emission tomography (PET)-MR machines promise increased resolution and anatomic-functional information on a variety of disorders. Lower field strength magnets (0.55 T and lower) are available, having advantages of smaller size and weight, as well as improvement in signal-to-noise and susceptibility artifacts due to de-noising software and machine learning algorithms. Spatial localization of proton signal is achieved by magnetic gradients surrounding the main magnet, which impart slight changes in magnetic field throughout the imaging volume. Rf pulses transiently excite the energy state of the hydrogen protons in the body. Rf is administered at a frequency specific for the proton element (hydrogen) and the field strength of the magnet. The subsequent return to equilibrium energy state (*relaxation*) of the hydrogen protons results in a release of specific Rf energy (the *echo*), which is detected by the coils that delivered the Rf pulses. Fourier analysis is used to transform the echo into the information used to form an MR image. The MR image thus consists of a map of the distribution of hydrogen protons, with signal intensity imparted by both density of hydrogen protons and differences in the relaxation times and phase (see below) of hydrogen protons on different molecules. Although clinical MRI currently makes use of the ubiquitous hydrogen proton, sodium and carbon imaging and spectroscopy are also possible, but have yet to be integrated into mainstream practice.

TABLE 434-3 Some Common Intensities on T1- and T2-Weighted MRI Sequences

			SIGNAL INTENSITY			
IMAGE	**TR**	**TE**	**CSF**	**FAT**	**BRAIN**	**EDEMA**
T1W	Short	Short	Low	High	Low	Low
T2W	Long	Long	High	High	Medium	High
FLAIR (T2)	Long	Long	Low	High	Medium	High

Abbreviations: CSF, cerebrospinal fluid; FLAIR, fluid-attenuated inversion recovery; TE, interval between radiofrequency pulse and signal reception; TR, interval between radiofrequency pulses; T1W and T2W, T1- and T2-weighted.

T1 and T2 Relaxation Times The rate of return to equilibrium of perturbed protons is called the *relaxation rate*. The relaxation rate varies among normal and pathologic tissues. The relaxation rate of a hydrogen proton in a tissue is influenced by local interactions with surrounding molecules and atomic neighbors. Two relaxation rates, T1 and T2, influence the signal intensity of the image. The T1 relaxation time is the time, measured in milliseconds, for 63% of the hydrogen protons to return to their normal equilibrium state, whereas the T2 relaxation is the time for 63% of the protons to become dephased owing to interactions among nearby protons. The intensity and image contrast of the signal within various tissues can be modulated by altering acquisition parameters such as the interval between Rf pulses (TR) and the time between the Rf pulse and the signal reception (TE). T1-weighted (T1W) images are produced by keeping the TR and TE relatively short, whereas using longer TR and TE times produces T2-weighted (T2W) images. Fat and subacute hemorrhage have relatively shorter T1 relaxation rates and thus higher signal intensity than brain on T1W images. Structures containing more water, such as CSF and edema, have long T1 and T2 relaxation rates, resulting in relatively lower signal intensity on T1W images and higher signal intensity on T2W images **(Table 434-3)**. Gray matter contains 10–15% more water than white matter, which accounts for much of the intrinsic contrast between the two on MRI **(Fig. 434-5*A*)**. T2W images are more sensitive than T1W images to edema, demyelination, infarction, and chronic hemorrhage, whereas T1W imaging is more sensitive to subacute hemorrhage and fat-containing structures.

Many different MR pulse sequences exist, and each can be obtained with two-dimensional or three-dimensional techniques and in various planes **(Figs.** 434-2, **434-4, and 434-5)**. The selection of a proper protocol that will best answer a clinical question depends on an accurate clinical history and indication for the examination. Fluid-attenuated inversion recovery (FLAIR) is a very useful pulse sequence in which the normally high signal intensity of CSF on T2W images is suppressed, improving the conspicuity of edematous lesions **(Fig. 434-5*B*)**. FLAIR images are more sensitive than standard spin echo images for water-containing lesions or edema, especially those close to CSF-filled cisterns and sulci. Diffusion-weighted imaging is also routinely obtained in most brain protocols. This sequence interrogates the microscopic motion of water, which is reduced in areas of infarction, abscess, and some tumors. Susceptibility-weighted imaging (SWI) is a gradient echo sequence that is very sensitive to alterations in local magnetic field generated by blood, calcium, and air. SWI is routinely obtained to detect microhemorrhages, such as is typical of amyloid angiopathy, hypertension, hemorrhagic metastases, traumatic brain injury, and thrombotic states **(Fig. 434-6*C*)**. MR images can be generated in any plane without changing the patient's position. Each sequence, however, is currently obtained separately and takes 1–10 min on average to complete. Three-dimensional volumetric imaging is routine, resulting in a volume of data that can be reformatted in any orientation to highlight certain disease processes. Perfusion techniques such as arterial spin labeling also provide quantitative imaging information regarding cerebral blood flow, and fat-suppressed imaging obtained with T1W and T2W imaging is useful for detection of fat-containing structures as well as improving contrast with other structures such as nerves in the case of peripheral nerve imaging, skull base imaging, and spinal cord imaging.

MR Contrast Material The heavy-metal element gadolinium forms the basis of all currently approved IV MR contrast agents.

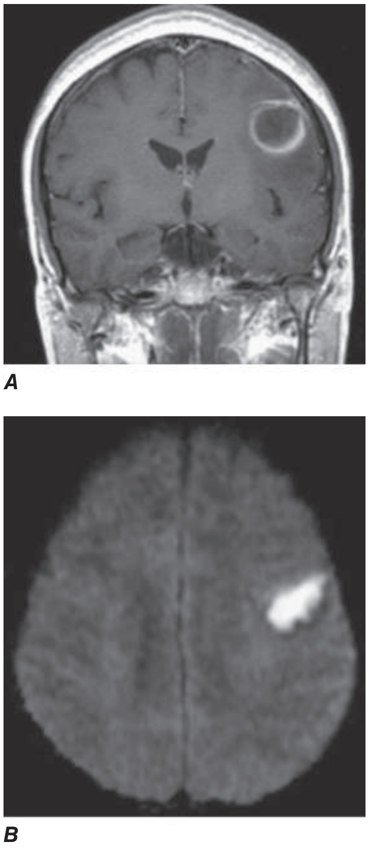

FIGURE 434-4 Cerebral abscess in a patient with fever and a right hemiparesis. ***A.*** Coronal postcontrast T1-weighted image demonstrates a ring-enhancing mass in the left frontal lobe. ***B.*** Axial diffusion-weighted image demonstrates restricted diffusion (high signal intensity) within the lesion, which in this setting is highly suggestive of cerebral abscess.

Gadolinium reduces the T1 and T2 relaxation times of nearby water protons in the presence of a magnetic field, resulting in a contrast enhancement on T1W images and a low signal on T2W images (the latter require a sufficient local concentration, usually in the form of an IV bolus). Unlike iodinated contrast agents, the effect of MR contrast agents depends on the presence of local hydrogen protons on which it must act to achieve the desired effect. There are multiple gadolinium agents approved in the United States for use with MRI. These differ according to the attached chelated moiety, which also affects the strength of chelation of the otherwise toxic gadolinium element. The chelating carrier molecule for gadolinium can be classified by whether it is macrocyclic or has linear geometry and whether it is ionic or nonionic. Macrocyclic ligands (group 2 agents) are considered more stable as the gadolinium ion is "caged" in the cavity of the ligand, and thus the rate of dissociation of gadolinium is slower compared to linear ligands (group 1 agents). Most agents are excreted by the renal system.

BRAIN ACCUMULATION OF GADOLINIUM It has been shown that gadolinium can accumulate in certain areas of the brain, primarily the dentate nuclei and globus pallidus, after serial administration of all types of gadolinium-based contrast agents (GBCAs). Autopsy studies have shown all GBCAs can lead to gadolinium deposition in brain; however, most clinical studies have demonstrated that linear GBCAs have more detectable gadolinium deposition than macrocyclic GBCAs. Gadolinium deposition in the brain appears to be dose dependent and occurs in patients with no clinical evidence of kidney or liver disease. To date, there have been no reports to suggest that these deposits are associated with histologic changes that would suggest neurotoxicity, even among agents with the highest rates of deposition. GBCAs can not only deposit in the brain but also in the skin, bone, liver, and other organs. This had importance in patients with renal failure who were exposed to linear gadolinium agents in the past, resulting in a rare but serious illness of the skin and organs secondary to accumulation of toxic gadolinium (nephrogenic systemic sclerosis).

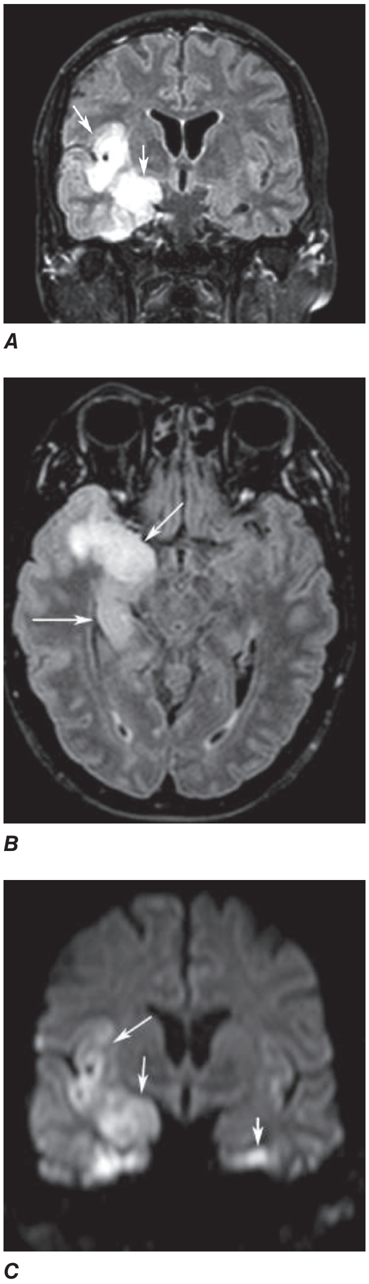

FIGURE 434-5 Herpes simplex encephalitis in a patient presenting with altered mental status and fever. ***A.*** and ***B.*** Coronal (***A***) and axial (***B***) T2-weighted fluid-attenuated inversion recovery images demonstrate expansion and high signal intensity involving the right medial temporal lobe and insular cortex (*arrows*). ***C.*** Coronal diffusion-weighted image demonstrates high signal intensity indicating restricted diffusion involving the right medial temporal lobe and hippocampus (*arrows*) as well as subtle involvement of the left inferior temporal lobe (*arrowhead*). This is most consistent with neuronal death and can be seen in acute infarction as well as encephalitis and other inflammatory conditions. The suspected diagnosis of herpes simplex encephalitis was confirmed by cerebrospinal fluid polymerase chain reaction analysis.

ALLERGIC HYPERSENSITIVITY Gadolinium-DTPA (diethylenetriaminepentaacetic acid) does not normally cross the intact BBB immediately but will enhance lesions lacking a BBB **(Fig. 434-4*A*)** as well as areas of the brain that normally are devoid of the BBB (pituitary, dura, choroid plexus). However, gadolinium contrast slowly crosses an intact BBB over time and especially in the setting of reduced renal clearance or inflamed meninges. The agents are generally well tolerated; overall adverse events after injection range from 0.07–2.4%. True allergic reactions are rare (0.004–0.7%) but have been reported. Severe life-threatening reactions are exceedingly rare; in one report, only 55 reactions out of 20 million doses occurred. However, the adverse reaction rate in patients with a prior history of reaction to gadolinium is eight

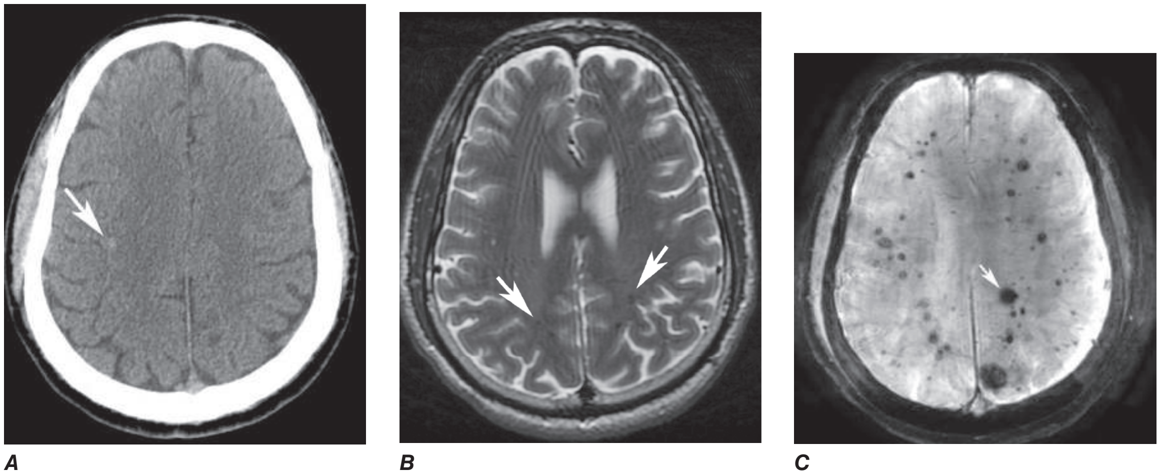

FIGURE 434-6 Susceptibility-weighted imaging in a patient with familial cavernous malformations. ***A.*** Noncontrast computed tomography scan shows one hyperdense lesion in the right hemisphere (*arrow*). ***B.*** T2-weighted fast-spin echo image shows subtle low-intensity lesions (*arrows*). ***C.*** Susceptibility-weighted image shows numerous low-intensity lesions consistent with hemosiderin-laden cavernous malformations (*arrow*).

times higher than normal. Other risk factors include atopy or asthma (3.7%). There is rare cross-reactivity between different classes of contrast media; a prior reaction to gadolinium-based contrast does not predict a future reaction to iodinated contrast medium, or vice versa, more than any other unrelated allergy. Gadolinium contrast material can be administered safely to children as well as adults, although these agents are generally avoided in those aged <6 months.

NEPHROTOXICITY Contrast-induced renal failure does not occur with gadolinium agents. A rare complication, nephrogenic systemic fibrosis (NSF), has occurred in patients with severe renal insufficiency who have been exposed to linear (group 1 and 3) gadolinium contrast agents. The onset of NSF has been reported between 5 and 75 days following exposure; histologic features include thickened collagen bundles with surrounding clefts, mucin deposition, and increased numbers of fibrocytes and elastic fibers in skin. In addition to dermatologic symptoms, other manifestations include widespread fibrosis of the skeletal muscle, bone, lungs, pleura, pericardium, myocardium, kidney, muscle, bone, testes, and dura. The American College of Radiology recommends that a glomerular filtration rate (GFR) assessment be obtained within 6 weeks prior to elective gadolinium-based MR contrast agent administration in patients with:

1. A history of renal disease (including solitary kidney, renal transplant, renal tumor)
2. Age >60 years
3. History of hypertension
4. History of diabetes
5. History of severe hepatic disease, liver transplantation, or pending liver transplantation; for these patients, it is recommended that the patient's GFR assessment be nearly contemporaneous with the MR examination.

The incidence of NSF in patients with severe renal dysfunction (GFR <30 mL/min per 1.73 m^2) varies from 0.19 to 4%. Other risk factors for NSF include acute kidney injury, the use of nonmacrocyclic agents, and repeated or high-dose exposure to gadolinium. The American College of Radiology Committee on Drugs and Contrast Media considers the risk of NSF among patients exposed to standard or lower doses of group 2 gadolinium agents (macrocyclic agents) to be sufficiently low or possibly nonexistent such that the assessment of renal function is optional prior to administration. Group 2 agents are strongly preferred in patients at risk for NSF. Renal function, dialysis status, or informed consent are not recommended prior to injection of group 2 agents, but deference is made to local practice preferences. Patients receiving any group 1 (linear) or 3 gadolinium-containing agents should be considered at risk of NSF if they are on dialysis (of any form); have severe or end-stage chronic renal disease (eGFR <30 mL/min per 1.73 m^2) without dialysis; have eGFR of 30–40 mL/min per 1.73 m^2 without dialysis (as the GFR may fluctuate); or have acute renal insufficiency. The use of gadolinium in young children and infants is discouraged due to the unknown risks and their immature renal systems.

■ COMPLICATIONS AND CONTRAINDICATIONS

From the patient's perspective, an MRI examination can be intimidating, and a higher level of cooperation is required than with CT. The patient lies on a table that is moved into a long, narrow gap within the magnet. Approximately 5% of the population experiences severe claustrophobia in the MR environment. This can be reduced by mild sedation but remains a problem for some. Movement of the patient during an MR examination may distort all the images in sequence; therefore, uncooperative patients should either be sedated for the MR study or scanned with CT. Generally, children aged <8 years usually require anesthesia monitored sedation to complete the MR examination without motion degradation.

MRI is considered safe for patients, even at 3- to 7-T field strengths. Serious injuries have been caused, however, by attraction of external ferromagnetic objects into the magnet, which act as missiles if brought too close to the magnet. Likewise, ferromagnetic implants, such as aneurysm clips, may torque within the magnet, causing damage to vessels and even death. Metallic foreign bodies in the eye have moved and caused intraocular hemorrhage; screening for ocular metallic fragments is indicated in those with a history of metal work or ocular metallic foreign bodies. Implanted cardiac pacemakers are generally a contraindication to MRI owing to the risk of induced arrhythmias; however, some newer pacemakers have been shown to be safe, and if necessary, MR may be performed if the pacemaker can be safely turned off during the scan. All health care personnel and patients must be screened and educated thoroughly to prevent such disasters because the magnet is always "on." Table 434-4 lists common contraindications for MRI.

MAGNETIC RESONANCE ANGIOGRAPHY

On routine spin echo MR sequences, moving protons (e.g., flowing blood, CSF) exhibit complex MR signals that range from high to low signal intensity relative to background stationary tissue. Fast-flowing blood returns no signal (flow void) on routine T1W or T2W spin echo MR images. Slower-flowing blood, as occurs in veins or distal to arterial stenosis, may appear high in signal. *MR angiography* makes use of pulse sequences called *gradient echo sequences* that increase the signal intensity of moving protons in contrast to suppressed low signal background intensity of stationary tissue. This results in a stack of images, which can be reformatted in any plane to highlight vascular anatomy and relationships.

Several types of MRA techniques exist. *Time-of-flight* (TOF) MRA is normally done without contrast administration and relies on the suppression of nonmoving tissue to provide a low-intensity background

TABLE 434-4 Common Contraindications to Magnetic Resonance Imaging

Cardiac pacemaker or permanent pacemaker leads
Internal defibrillatory device
Cochlear prostheses
Bone growth stimulators
Spinal cord stimulators
Electronic infusion devices
Intracranial aneurysm clips (some but not all)
Ocular implants (some) or ocular metallic foreign body
McGee stapedectomy piston prosthesis
DuraPhase penile implant
Swan-Ganz catheter
Magnetic stoma plugs
Magnetic dental implants
Magnetic sphincters
Ferromagnetic inferior vena cava filters, coils, stents—safe 6 weeks after implantation
Tattooed eyeliner (contains ferromagnetic material and may irritate eyes)

Note: See also *http://www.mrisafety.com*.

for the high signal intensity of flowing blood entering the section. A typical TOF MRA sequence results in a series of contiguous, thin MR sections (0.6–0.9 mm thick), which can be viewed as a stack and manipulated to create an angiographic image data set that can be reformatted and viewed in various planes and angles, much like that seen with conventional angiography **(Fig. 434-2*G*)**.

Phase-contrast MRA has a longer acquisition time than TOF MRA, but in addition to providing anatomic information like that of TOF imaging, it can be used to reveal the velocity and direction of blood flow in each vessel.

MRA is also often acquired during infusion of IV gadolinium contrast material. Advantages include faster imaging times (1–2 min vs 10 min), fewer flow-related artifacts, and four-dimensional temporal imaging resulting in arterial and venous phases. Recently, contrast-enhanced MRA has become the standard for assessment of the extracranial vascular structures. This technique entails rapid imaging using coronal three-dimensional TOF sequences during a bolus infusion of gadolinium contrast agent.

MRA has lower spatial resolution compared with conventional film-based angiography and, therefore, is inherently less sensitive to detection of small-vessel abnormalities, such as vasculitis and distal vasospasm. MRA is also less sensitive to slowly flowing blood and thus may not reliably differentiate complete from near-complete occlusions. Motion, either by the patient or by anatomic structures, may distort the MRA images, creating artifacts. These limitations notwithstanding, MRA has proved useful in evaluation of the extracranial carotid and vertebral circulation as well as of larger-caliber intracranial arteries and dural sinuses. It has also proved useful in the noninvasive detection of intracranial aneurysms and vascular malformations.

Vessel wall MR imaging (VWI) is an MR technique that relies on suppression of all moving protons within vessels and CSF, combined with IV contrast administration **(Fig. 434-7)**. Unlike MRA, VWI is a high spatial resolution, three-dimensional, T1W technique used to assess pathology of the vessel wall itself. This technique can be used to detect, characterize, and differentiate such pathologies as atherosclerosis, vasculitis (e.g., primary angiitis of the central nervous system [PACNS]), and vasculopathies such as reversible cerebral vasoconstriction syndrome (RCVS) and has been used to assess the wall of aneurysms.

ECHO-PLANAR MRI

Echo-planar MRI (EPI) forms the basis of several important MR imaging sequences. EPI uses fast gradients that are switched on and off at high speeds to create the information used to form an image. With EPI, all the information required for processing an image is accumulated in milliseconds, and the information for the entire brain can be obtained in <1–2 min, depending on the degree of resolution required or desired. Fast MRI reduces patient and organ motion and is the basis of perfusion imaging during contrast infusion and kinematic motion studies. EPI is also the sequence used to obtain diffusion-weighted imaging (DWI) and tractography (DTI), as well as functional MRI (fMRI) and arterial spin-labeled (ASL) perfusion studies **(Figs. 434-2*H*,** 434-4, **434-5*C*,** and 434-7; and **Fig. 437-13)**.

Perfusion and diffusion imaging are EPI techniques that are useful in early detection of ischemic injury of the brain and may be useful together to demonstrate infarcted tissue as well as ischemic but potentially viable tissue at risk of infarction (e.g., the ischemic penumbra). DWI assesses microscopic motion of water; water protons that move reduce signal intensity on diffusion-weighted images. Pathology that reduces microscopic water motion results in relatively higher signal. Infarcted tissue reduces the water motion within cells and in the interstitial tissues, resulting in high signal on DWI. DWI is the most sensitive technique for detection of acute cerebral infarction of <7 days in duration (Fig. 434-2*H*). It is also quite sensitive for detecting dying or dead brain tissue secondary to encephalitis, as well as abscess and purulent formations (Fig. 434-4*B*).

Perfusion MRI can be performed by the acquisition of fast EPI during a rapid IV bolus of gadolinium contrast material or by noncontrast ASL techniques. With contrast perfusion imaging, parametric maps of relative cerebral blood volume, mean transit time (MTT), time to maximum (tMAX), and cerebral blood flow can be derived. Prolonged MTT and tMAX and reduction in cerebral blood volume and cerebral blood flow are typical of infarction. In the setting of reduced blood flow, a prolonged MTT of contrast but normal or elevated cerebral blood volume may indicate tissue supplied by slower collateral flow that is at risk of infarction. Perfusion MRI imaging can also be used in the assessment of brain tumors to differentiate intraaxial primary tumors, whose BBB is relatively intact, from extraaxial tumors or metastases, which demonstrate a relatively more permeable BBB.

DTI is derived from diffusion MRI sequences. This technique assesses the direction and integrity of protons flowing within white matter architecture. It has proven valuable in the assessment of subcortical white matter tract anatomy prior to brain tumor surgery, as well as in determining normal and abnormal white matter architecture in congenital and acquired pathologies such as traumatic brain injury and assessing the integrity of peripheral nerves and spinal cord lesions **(Fig. 434-8)**.

fMRI is an EPI technique that localizes regions of activity in the brain following task activation or at rest (so-called resting-state fMRI). Neuronal activity elicits a slight increase in the delivery of oxygenated blood flow to a specific region of activated brain. This results in an alteration in the balance of oxyhemoglobin and deoxyhemoglobin, which yields a 2–3% increase in signal intensity within veins and local capillaries. Currently, preoperative somatosensory, motor, and auditory cortex localization is performed, and methods to assess motor and language function are in development. This technique has proved useful to neuroscientists interested in interrogating the localization of specific brain functions.

ARTERIAL SPIN LABELING

ASL is a quantitative noninvasive MR technique that measures cerebral blood flow (Fig. 434-7). Blood traversing in the neck is labeled by an MR pulse and then imaged in the brain after a short (2 s) delay. The signal is reflective of blood flow. ASL has become almost standard in many MR protocols because it is relatively fast to acquire and does not require contrast administration. Increased cerebral flow is more easily identified than slow flow, which can be sometimes difficult to quantify. This technique has also been useful in detecting shunting in arteriovenous malformations and fistulas, as well as increased blood flow in brain tumors, and in patients after transient ischemic attack, seizure, or migraine.

MAGNETIC RESONANCE NEUROGRAPHY

MRN is an MR technique that shows promise in detecting increased signal in irritated, inflamed, or infiltrated peripheral nerves. T1W and T2W imaging are obtained with fat-suppressed fast-spin echo imaging or short inversion recovery sequences. Inflamed peripheral nerves will demonstrate high signal on T2W imaging. MRN is indicated in

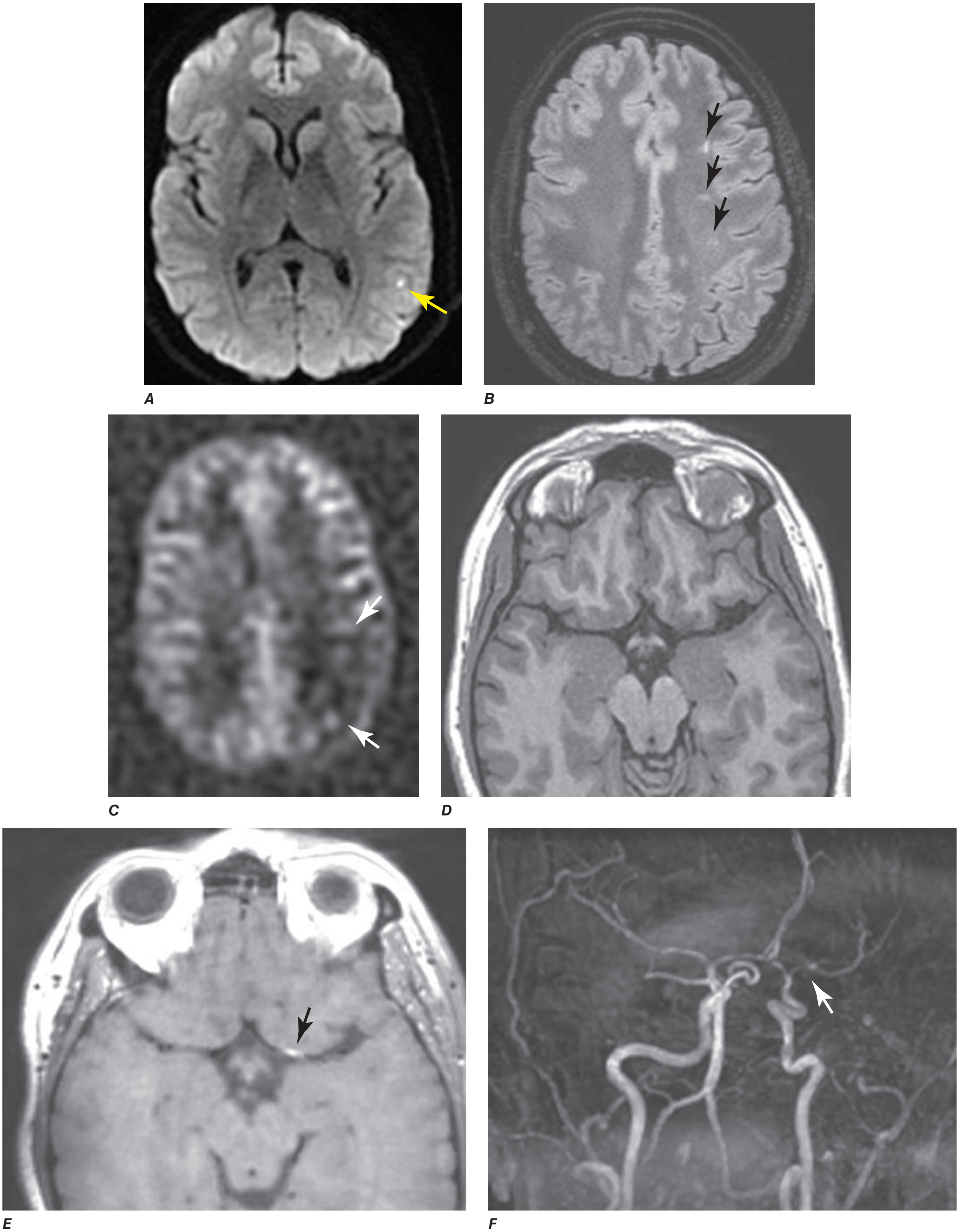

FIGURE 434-7 Arterial spin label and vessel wall imaging in a 25-year-old woman with focal cerebral arteriopathy. The patient had an 8-month history of intermittent weakness of the right side with spasms. Imaging shows evidence of cerebral ischemia. Cerebrospinal fluid was transiently inflammatory. ***A.*** Diffusion-weighted image shows focal region of reduced diffusion in left parietal lobe. ***B.*** T2 fluid-attenuated inversion recovery images show several foci of high signal in left deep subcortical white matter. ***C.*** Arterial spin label image demonstrates reduced cerebral blood flow in left parietal lobe (*arrows*). ***D.*** Three-dimensional (3D) T1 image without contrast administration. ***E.*** 3D T1-weighted cube vessel wall image following gadolinium contrast shows focal enhancement of the left proximal middle cerebral artery (*arrow*). ***F.*** 3D time-of-flight magnetic resonance angiography shows focal narrowing of the left supraclinoid internal carotid artery and proximal middle cerebral artery (*arrow*).

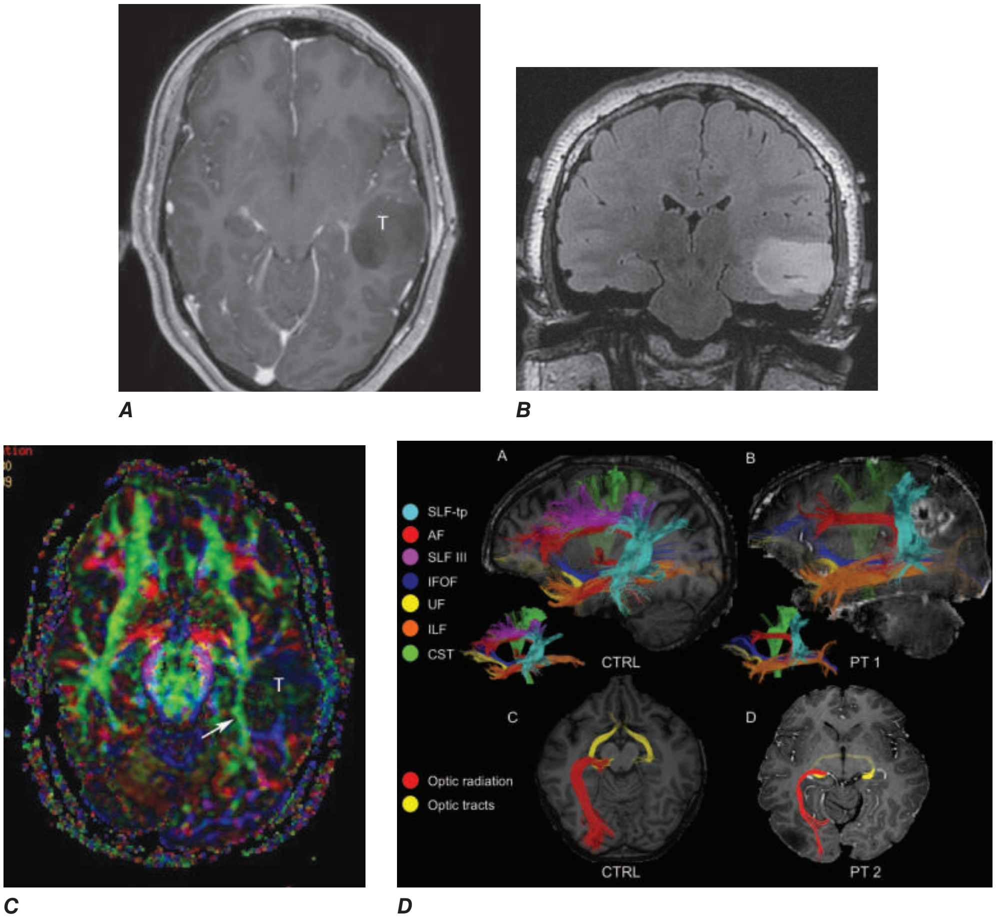

FIGURE 434-8 Diffusion tractography in cerebral glioma. A, B and C are images showing a left temporal mass lesion (T) that medially displaces the inferior longitudinal fasciculus (arrow). D. Different patient showing: Associative and descending pathways in a healthy subject (***A***) and in a patient with parietal lobe glioblastoma (***B***) presenting with a language deficit: the mass causes a disruption of the arcuate-SLF complex, in particular of its anterior portion (SLF III). Also shown are bilateral optic tract and left optic radiation pathways in a healthy subject (***C***) and in a patient with left occipital grade II oligoastrocytoma (***D***): the mass causes a disruption of the left optic radiation. *Shown in neurologic orientation, i.e., the left brain appears on the left side of the image.* AF, long segment of the arcuate fascicle; CST, corticospinal tract; IFOF, inferior fronto-occipital fascicle; ILF, inferior longitudinal fascicle; SLF III, superior longitudinal fascicle III or anterior segment of the arcuate fascicle; SLF-tp, temporo-parietal portion of the superior longitudinal fascicle or posterior segment of the arcuate fascicle; T, tumor; UF, uncinated fascicle. *(Part D used with permission from Eduardo Caverzasi and Roland Henry.)*

patients with radiculopathy whose conventional MR studies of the spine (cervical or lumbar) are normal or in those suspected of peripheral nerve entrapment or trauma. This technique is now also being used to assess peripheral nerve damage after trauma or from compressive and autoimmune neuropathies.

POSITRON EMISSION TOMOGRAPHY

PET relies on the detection of positrons emitted during the decay of a radionuclide that has been injected into a patient. The most frequently used moiety is 2-[^{18}F] fluoro-2-deoxy-D-glucose (FDG), which is an analogue of glucose and is taken up by cells competitively with 2-deoxyglucose. Many other radioisotopes are used in other indications. With FDG, multiple images of glucose uptake activity are formed 45–60 min after IV administration of FDG. Images reveal differences in regional glucose activity among normal and pathologic brain structures. FDG-PET is used primarily for the detection of extracranial metastatic disease; however, a lower activity of FDG in the parietal lobes is associated with Alzheimer's disease, a finding that may simply reflect atrophy that occurs in the later stages of the disease. Combination PET-CT scanners, in which both CT and PET are obtained at one sitting, have largely replaced PET scans alone. MR-PET scanners have also been developed and may prove useful for imaging the brain and other organs without the radiation exposure of CT. More recent PET ligand developments include beta-amyloid and tau PET tracers **(Chap. 442)**. Studies have shown an increased percentage of amyloid deposition in patients with Alzheimer's disease compared with mild cognitive impairment and healthy controls; however, up to 25% of cognitively "normal" older patients show abnormalities on amyloid PET imaging. This may either reflect subclinical disease processes or a variation of normal. Tau imaging may be more specific for Alzheimer's disease, and clinical studies are in progress.

MYELOGRAPHY

TECHNIQUE

Myelography involves the intrathecal instillation of specially formulated water-soluble iodinated contrast medium into the lumbar or cervical subarachnoid space. CT scanning is typically performed after myelography to better demonstrate the spinal cord and roots, which appear as filling defects in the opacified subarachnoid space. CT myelography, in which CT is performed after the subarachnoid injection of a small amount of contrast material, has replaced conventional myelography for many indications, thereby reducing exposure to radiation and contrast media. CT is obtained at a slice thickness of ~2.5 mm and reconstructed at 0.625-mm thick slices, which can quickly be reformatted in sagittal and coronal planes, equivalent to traditional myelography projections.

INDICATIONS

CT myelography and MRI have largely replaced conventional myelography for the diagnosis of diseases of the spinal canal and cord (Table 434-1). Remaining indications for conventional plain film myelography include the evaluation of suspected meningeal or arachnoid cysts and the localization of CSF fistulas. Conventional myelography and CT myelography provide the most precise information in patients with failed back syndrome following spinal fusion procedures.

CONTRAINDICATIONS

Myelography is relatively safe; however, it should be performed with caution in any patient with elevated intracranial pressure, evidence of a spinal block, or a history of allergic reaction to intrathecal contrast media. In patients with a suspected spinal block, MR is the preferred imaging technique. If myelography is necessary, only a small amount of contrast medium should be instilled below the block to minimize the risk of neurologic deterioration. Lumbar puncture (LP) is to be avoided in patients with bleeding disorders and those with infections of the overlying soft tissues. Anticoagulant therapy should be withheld prior to elective LP to avoid epidural or intradural hemorrhage, unless required in emergent situations **(Chap. S3)**.

COMPLICATIONS

Headache is the most frequent complication of myelography and is reported to occur in 5–30% of patients. Nausea and vomiting may also occur rarely. Postural headache (post-LP headache) is generally due to continued epidural leakage of CSF from the dural puncture site. A higher incidence is noted among younger women and with the use of larger gauge cutting-type spinal needles. If significant headache persists for >48 h, placement of an epidural blood patch should be considered. Vasovagal syncope may occur during LP; it is accentuated by the upright position used during conventional lumbar myelography. Adequate hydration before and after myelography will reduce the incidence of this complication. **Management of LP headache is discussed in Chap. 17.**

Hearing loss is a rare complication of myelography. It may result from a direct toxic effect of the contrast medium or from an alteration of the pressure equilibrium between CSF and perilymph in the inner ear. Puncture of the spinal cord is a rare but serious complication of cervical (C1–2) or high LP. The risk of cord puncture is greatest in patients with spinal stenosis, Chiari malformations, or conditions that reduce CSF volume. CT myelography following a lumbar injection and MRI are safer alternatives to cervical puncture. Reactions to intrathecal contrast administration are rare; aseptic meningitis and encephalopathy are reported rare complications. The latter is usually dose related and associated with contrast entering the intracranial subarachnoid space. Seizures rarely occur following myelography, historically reported in 0.1–0.3% of patients. Risk factors include a preexisting seizure disorder and the use of a total iodine dose of >4500 mg. Other reported complications include hyperthermia, hallucinations, depression, and anxiety states. These side effects have been reduced by the development of nonionic, water-soluble contrast agents as well as by head elevation and generous hydration following myelography.

SPINE INTERVENTIONS

DISKOGRAPHY

The evaluation of back pain and radiculopathy **(Chap. 18)** may require diagnostic procedures that attempt either to reproduce the patient's pain or relieve it, indicating its correct source prior to lumbar fusion. Diskography is now rarely indicated. It is performed by CT or fluoroscopic placement of a 22- to 25-gauge needle into the intervertebral disk and subsequent injection of 1–3 mL of contrast media. The intradiscal pressure is recorded, as is an assessment of the patient's response to the injection of contrast material. Little or no pain is felt during injection of a normal disk, which does not accept much more than 1 mL of contrast material, even at pressures as high as 415–690 kPa (60–100 lb./in^2). CT and plain films are obtained following the procedure. Concerns have been raised that diskography may contribute to an accelerated rate of disk degeneration; furthermore, patients who suffer from depression or anxiety are more likely to find diskography painful, and in some cases, the procedure-associated pain became persistent, lasting a year or longer. Thus, it is rarely used as a reliable biomarker of pain generation. Newer spectroscopic disc techniques are being explored for the detection of painful degenerative disks.

SELECTIVE NERVE ROOT AND EPIDURAL SPINAL INJECTIONS

Percutaneous selective nerve root and epidural administration of glucocorticoid and anesthetic mixtures may be both therapeutic and diagnostic. Typically, 1–2 mL of an equal mixture of a long-acting glucocorticoid such as betamethasone or dexamethasone combined with a long-acting anesthetic such as bupivacaine is instilled under CT or fluoroscopic guidance in the intraspinal epidural space or adjacent to an existing nerve root in question as a pain source. This can also be performed into the facet joints, or around the medial nerve branches that supply innervation to the facet joints as well. Radiofrequency ablation of the medial branches that supply sensation to the facet joints is commonly performed, and ablation techniques of the basivertebral nerves that conduct the sensation from degenerative disks have proven useful in certain cases of painful disk degeneration. Cement placement into compression fractures of the vertebral bodies, so-called vertebroplasty and kyphoplasty techniques, are commonly performed for pain-generating fractures, especially in elderly patients or those with pathologic fractures.

ANGIOGRAPHY

Catheter angiography is indicated for evaluating intracranial small-vessel pathology (e.g., vasculitis), for assessing vascular malformations and aneurysms, and in endovascular therapeutic procedures (Table 434-1). As noted above, angiography has been replaced for many indications by CT/CTA or MRI/MRA.

Angiography carries the greatest risk of morbidity of all diagnostic imaging procedures, owing to the necessity of inserting a catheter into a blood vessel, directing the catheter to the required location, injecting contrast material to visualize the vessel, and removing the catheter while maintaining hemostasis. Therapeutic transcatheter procedures (see below) have become important options for the treatment of some cerebrovascular diseases. The decision to undertake a diagnostic or therapeutic angiographic procedure requires careful assessment of the goals of the investigation and its attendant risks.

Patients undergoing angiography should be well hydrated before and after the procedure. Because the femoral route is used most, the femoral artery must be compressed after the procedure to prevent a hematoma from developing. The puncture site and distal pulses should be evaluated carefully after the procedure; complications can include thigh hematoma or lower-extremity emboli.

COMPLICATIONS

A common femoral arterial puncture provides retrograde access via the aorta to the aortic arch and great vessels. The most feared complication of cerebral angiography is stroke. Thrombus can form on or inside the tip of the catheter, rarely arterial dissection or perforation can occur, and atherosclerotic thrombus or plaque can be dislodged by the catheter or guide wire or by the force of injection and can embolize distally in the cerebral circulation. Risk factors for ischemic complications include limited experience on the part of the angiographer, atherosclerosis, vasospasm, low cardiac output, decreased oxygen-carrying capacity, advanced age, and prior history of migraine. The risk of a neurologic complication varies but is ~4% for transient ischemic attack and stroke, 0.5% for permanent deficit, and <0.1% for death.

Nonionic contrast material is used exclusively in cerebral angiography. Nonionic contrast injected into the cerebral vasculature can be neurotoxic if the BBB is breached, either by an underlying disease or by the injection of hyperosmolar contrast agent. Patients with dolichoectasia of the basilar artery can suffer reversible brainstem dysfunction and acute short-term memory loss during angiography, owing to the slow percolation of the contrast material and the consequent prolonged exposure of the brain. Rarely, an intracranial aneurysm ruptures during an angiographic contrast injection, causing subarachnoid hemorrhage, perhaps because of injection under high pressure.

SPINAL ANGIOGRAPHY

Spinal angiography is indicated to evaluate the location of vascular malformations and to identify the artery of Adamkiewicz **(Chap. 453)** prior to aortic aneurysm repair. The procedure is lengthy and requires the use of relatively large volumes of contrast; the incidence of serious complications, including paraparesis, subjective visual blurring, and altered speech, is <1%. Gadolinium-enhanced MRA has been used successfully in this setting, as has iodinated contrast CTA, which has promise for replacing diagnostic spinal angiography for some indications.

INTERVENTIONAL NEURORADIOLOGY

This rapidly developing field is providing new therapeutic options for patients with challenging neurovascular problems. Available procedures include detachable coil therapy for aneurysms, particulate or liquid adhesive embolization of arteriovenous malformations, stent retrieval systems for embolectomy in acute stroke, balloon angioplasty and stenting of arterial stenosis or vasospasm, transarterial or transvenous embolization of dural arteriovenous fistulas and CSF-venous fistulas of the spine, balloon occlusion of carotid-cavernous and vertebral fistulas, endovascular treatment of vein-of-Galen malformations, preoperative embolization of tumors, and thrombolysis of acute arterial or venous thrombosis. Many of these disorders place the patient at high risk of cerebral hemorrhage, stroke, or death.

The highest complication rates are found with the therapies designed to treat the highest risk diseases. The advent of electrolytically detachable coils ushered in a new era in the treatment of cerebral aneurysms **(Chap. 440)**. Two randomized trials found reductions of morbidity and mortality at 1 year among those treated for aneurysm with detachable coils compared with neurosurgical clipping. In many centers, coiling has become standard therapy for many proximal circle of Willis aneurysms.

Finally, recent studies of stent retrieval systems used to withdraw emboli have shown improved clinical outcomes in patients presenting with large-vessel occlusions and signs of acute stroke **(Chap. 438)**.

FURTHER READING

Bambach S et al: Arterial spin labeling applications in pediatric and adult neurologic disorders. J Magn Reson Imaging 55:698, 2020.

Choi JW, Moon WJ: Gadolinium deposition in the brain: Current updates. Korean J Radiol 20:134, 2019.

Mandell DM et al: Intracranial vessel wall MRI: Principles and expert consensus recommendations of the American Society of Neuroradiology. AJNR Am J Neuroradiol 38:218, 2017.

Nadjir R et al: *Neuroradiology: The Core Requisites*, 5th ed. Philadelphia, Elsevier, 2024.

Pelz DM et al: Interventional neuroradiology: A review. Can J Neurol Sci 16:1, 2020.

Schönmann C, Brockow K: Adverse reactions during procedures: Hypersensitivity to contrast agents and dyes. Ann Allergy Asthma Immunol 124:156, 2020.

Tournier JD: Diffusion MRI in the brain—theory and concepts. Prog Nucl Magn Reson Spectrosc 112-113:1, 2019.

Watson RE et al: MR imaging safety events: Analysis and improvement. Magn Reson Imaging Clin N Am 28:593, 2020.

435

Pathobiology of Neurologic Diseases

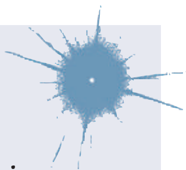

Stephen L. Hauser, Arnold R. Kriegstein, Stanley B. Prusiner

The human nervous system is the organ of consciousness, cognition, ethics, and behavior; as such, it is the most intricate structure known to exist. More than one-third of the 23,000 genes encoded in the human genome are expressed in the nervous system, and for many, different isoforms are expressed that further increase the resulting specificities and potential functionality. Each mature brain is composed of 100 billion neurons, several million miles of axons and dendrites, and $>10^{15}$ synapses. Neurons exist within a dense parenchyma of multifunctional glial cells that synthesize myelin, preserve homeostasis, and regulate immune responses. Measured against this background of complexity, the achievements of molecular neuroscience have been extraordinary. Advances have occurred in parallel with the development of new enabling technologies—in bioengineering and computational sciences; imaging; and cell, molecular, and chemical biology—and moving forward it is likely that the pace of new discoveries will only increase. This chapter reviews a few of the most dynamic areas in neuroscience, specifically highlighting advances in immunology and inflammation, neurodegeneration, and stem cell biology. In each of these areas, recent discoveries are providing context for an understanding of the triggers and mechanisms of disease and offering new hope for prevention, treatment, and repair of nervous system injuries. Discussions of the neurogenetics of behavior, advances in addiction science, and diseases caused by network dysfunction can be found in **Chap. 462** (Biology of Psychiatric Disorders); and new approaches to rehabilitation via harnessing of neuroplasticity, neurostimulation, and computer–brain interfaces are presented in **Chap. 500** (Emerging Neurotherapeutic Technologies).

NEUROIMMUNOLOGY AND NEUROINFLAMMATION

Neuroimmunology traditionally comprises the science of immune-mediated diseases of the nervous system, especially autoimmune diseases such as multiple sclerosis, myasthenia gravis, and Guillain-Barré syndrome, as well as disorders in which immune-mediated neurologic damage occurs in the context of infection or neoplasia. More recently, recognition that neuroinflammation and the innate immune system play key roles in an expanded category of neurologic disorders, and neurodegenerative disorders in particular, has focused renewed attention on the intrinsic cellular components in the central nervous system (CNS) that mediate tissue damage, especially microglia/macrophages and astrocytes, and also on oligodendrocytes, which are now recognized as central players across a wide range of brain disorders. It is also increasingly recognized that extensive networks of communication exist between each of these cell types and that resulting *non-cell-autonomous* pathologies are likely to underlie many human CNS disorders.

MICROGLIA AND MACROPHAGES

These represent the most abundant cell types in the nervous system responsible for antigen presentation and innate immunity. Brain microglia ("small glue") are derived from a primitive macrophage population in the yolk sac that migrates to the nervous system early in embryogenesis before the blood-brain barrier is formed. Once in the CNS, a variety of cell signals mediate microglial proliferation, migration, and differentiation. Microglia maintain their cell numbers primarily through self-renewal and not repopulation from the circulation; however, it is now also clear that under some conditions peripherally derived macrophages that become microglia-like can replace damaged or defective microglia throughout the life span. Most microglia receive survival signals through colony-stimulating factor 1 receptor (Csf1r), via its natural ligands Csf1 produced by astrocytes and oligodendrocytes, and interleukin (IL) 34 produced by neurons. Depletion of microglia by administration of a selective inhibitor of Csf1r (PLX5622) was followed by rapid repopulation, which led to identification of a second population of ramified microglial precursor cells that do not require Csf1r signaling. Microglia have traditionally been divided into "resting" and "activated" states, the former characterized by an extensively ramified appearance and the latter by a globular amoeboid phenotype. A variety of factors and cues, including type 1 interferons, can prime resting microglia toward an activated state, and these "primed" microglia are then hyperresponsive to a variety of secondary immune challenges including infection. Recently, single-cell transcriptome sequencing of microglia has also revealed a high degree of heterogeneity that was previously unappreciated, the functional consequences of which are largely unknown.

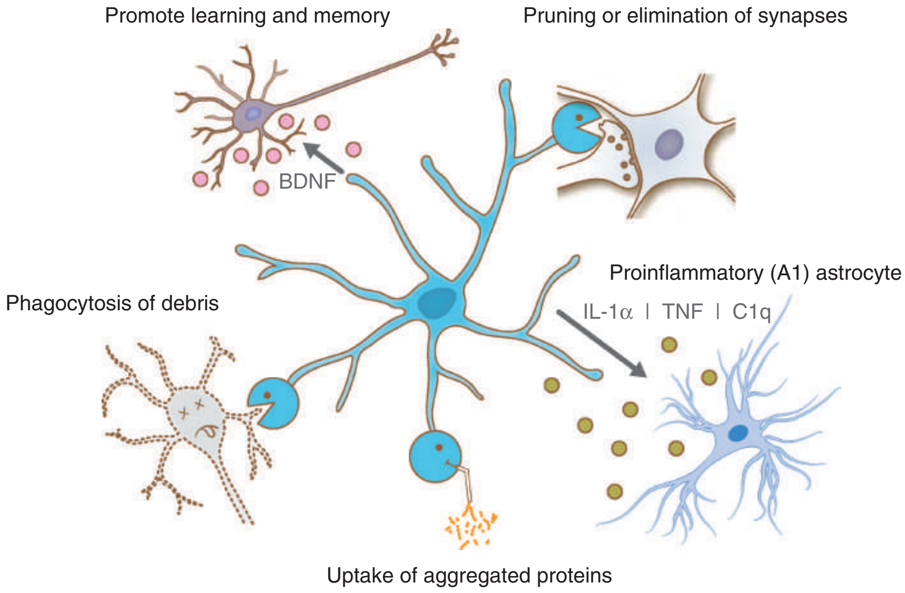
Promote learning and memory
Pruning or elimination of synapses
BDNF
Proinflammatory (A1) astrocyte
Phagocytosis of debris
IL-1α | TNF | C1q
Uptake of aggregated proteins

FIGURE 435-1 The multifunctional microglial cell. Microglia have diverse functions that can support healthy development and maintain homeostasis or contribute to tissue damage in pathologic conditions. Homeostatic functions include promotion of learning and memory through secretion of soluble proteins such as brain-derived neurotrophic factor (BDNF); participation in normal synaptic pruning; and clearing cellular debris and protein aggregates via phagocytosis. However, in pathologic states, activated microglia also contribute to tissue damage by targeting normal healthy neurons and synapses; by promoting formation of β-amyloid or other misfolded proteins deposited in neurodegenerative diseases; and by secreting cytokines (such as interleukin 1α, tumor necrosis factor, and the complement component C1q) incriminated in induction of neurotoxic A1 astrocytes. In addition, microglia have diverse functions in adaptive immunity, including roles in antigen presentation and immune regulation (Fig. 435-2). *(Reproduced with permission from J Herz et al: Myeloid cells in the central nervous system. Immunity 46:943, 2017.)*

Microglia play critical roles in sculpting neuronal populations during development and across the life span, through secretion of brain-derived neurotrophic factor (BDNF) and other trophic factors that promote neuronal survival, and also via production of reactive oxygen species (ROS) and other molecules that mediate cell death. Microglia regulate development and maintenance of neural circuits through pruning of excitatory synapses and control of dendritic spine densities **(Fig. 435-1)**. Mice depleted of microglia during development exhibit a variety of cognitive, learning, and behavioral deficits, including abnormal social behaviors. These processes are dependent on classical complement pathway molecules, including secretion of C1q and expression of complement receptor 3 (CR3) and CR5. Abnormalities in these microglial-dependent networks are now recognized as critical to the pathogenesis of most neurodegenerative and age-related pathologies.

Activation of the complement cascade has assumed a major place in current concepts of pathogenesis of Alzheimer's disease (AD) and other dementias, as follows. Synapses targeted for elimination are tagged by the complement proteins C1q and C3, the levels of which increase in the presence of excess β-amyloid. C3-bearing synapses are targeted for elimination by microglia that express CR3, and knockout of C1q or C3 can rescue the clinical and pathologic abnormalities associated with neurodegeneration. Modification of aberrant complement activity represents an attractive approach for treatment of neurodegeneration from multiple causes, not only AD, but also Parkinson's and Huntington's disease, among others, in which microglia are responsible for clearing pathologic protein aggregates. An ideal therapeutic would target the deleterious microglial functions while preserving ones that are central to homeostasis such as synaptic pruning. Another consideration, especially given the diverse roles of the complement system in all organs of the body, is potential off-target toxicities. Clinical trials are now underway with an inhibitory monoclonal antibody against C1q for Huntington's disease and amyotrophic lateral sclerosis (ALS). While the specific mechanisms of complement-dependent neurodegeneration will likely differ in distinct neurodegenerative conditions, these data provide hope that complement-pathway interventions could represent an approach to control of neurodegenerative pathologies mediated at least in part through the innate immune system.

Genetic evidence also supports a primary role for microglia in numerous neurodegenerative conditions and disease states, in contrast to earlier views in which their role was seen as largely secondary and involving phagocytosis of cell debris. More than half of all genes implicated in genome-wide association studies in AD implicate innate immune processes and microglia. One direct genetic link is the phagocytosis-associated gene *TREM2* (triggering receptor expressed on myeloid cells 2). TREM2 is a microglial receptor that can bind amyloid, induce proliferation and migration of microglia, and possibly limit the spread of disease-associated AD aggregates. Moreover, a soluble TREM2 cleavage product promotes neuroinflammation and inhibits aggregation of β-amyloid. Loss-of-function mutations in *TREM2* increase AD risk up to fourfold. In one mouse model of AD, overexpression of TREM2 blocked AD pathology and rescued performance on tests of learning and memory; however, in other models, including tau models, the effects of TREM2 targeting were found to be stage specific or inconsistent. A clinical trial of an agonist monoclonal antibody against TREM2 is underway.

Immune system genes implicated in susceptibility to other late-life dementias also represent promising targets for therapy. For example, ~10% of all cases of frontotemporal degeneration (FTD) are due to heterozygous mutations of the gene granulin (*GRN*), which encodes the protein progranulin expressed in neurons and microglial cells **(Chap. 443)**. Progranulin is a neurotrophic factor essential for lysosomal function and microglial homeostasis. *GRN* mutations with resulting haploinsufficiency of progranulin are highly penetrant for FTD. When progranulin is deleted in mice, an age-dependent microglial activation phenotype results, associated with upregulation of proinflammatory neurotoxic cytokines, complement components, and other genes associated with innate immunity, along with enhanced pruning of inhibitory synapses and behavioral manifestations reminiscent of human FTD. Remarkably, inhibition of complement activation can rescue all of these deficits. These data indicate a primary role for microglial activation in FTD caused by mutations in *GRN*, likely mediated through increased production of C3 complement, enhanced lysosomal trafficking, and excessive synaptic pruning in affected brain regions affected. Clinical trials underway for neurodegeneration due to progranulin deficiency include a monoclonal antibody to block lysosomal degradation of progranulin, as well as gene and protein replacement therapies.

Microglia are generally considered the most important source of antigen-presenting cells in the brain, as they express class II major histocompatibility complex molecules enabling antigen presentation to CD4 T cells. In Parkinson's disease (PD), neurotoxic T cells reactive against epitopes of α-synuclein are commonly found, and it has been postulated that antigen presentation by microglia may have initiated this autoimmune response. Beyond a traditional role as antigen-presenting cells, microglia are now understood to also have a range of other interactions with T and B lymphocytes. For example, secretion of CCL3 by microglia, which is enhanced with aging, promotes the recruitment of memory CD8 T cells, which might account for the presence of CD8 T cells in a wide range of neuropathologic conditions including multiple sclerosis (MS).

Microglia are located throughout the brain parenchyma, while the closely related brain macrophages occur primarily in perivascular regions, including the meninges and choroid plexus. Like microglia, brain macrophages are derived from yolk sac precursors that appear to enter the brain at an early developmental stage and propagate locally, although some choroid plexus macrophages may also be replenished at low levels from the bloodstream on a continuing basis. Under inflammatory conditions, however, large numbers of hematogenously derived monocytes readily enter the brain parenchyma. In the disease model, experimental autoimmune encephalomyelitis (EAE), macrophages derived from bone marrow monocytes are the critical population that initiates inflammatory demyelination at paraxonal regions near nodes of Ranvier **(Fig. 435-2)**. Brain macrophages have multiple proinflammatory functions, including promoting adhesion, attraction, and activation of B and T lymphocytes; providing antigen-specific activation of T cells via antigen presentation of immunogenic peptides, including autoantigens, complexed to surface class II major histocompatibility complex (MHC II) molecules; and contributing to cell injury through generation of oxidative stress and cytotoxicity. By contrast, microglia have been traditionally thought to downregulate inflammatory responses and promote tissue repair in EAE. This model of the relative roles of macrophage and microglial cells is certainly an oversimplification, and more nuanced functions of these cell types have been revealed by single-cell sequencing methods, demonstrating previously unappreciated heterogeneity influenced by location, context, and environmental cues.

ASTROCYTES

Astrocytes represent half or more of all cells in the CNS. Thought earlier to function as simple interstitial supporting cells that provide scaffolds for neuronal migration and contribute to homeostasis, emerging data indicate far more pleiotropic functions for this cell type. Astrocytes, like microglia, play profound roles in the life of synapses by secreting factors (such as apolipoprotein E, thrombospondins, and glypicans) that regulate development, maintenance, and pruning of presynaptic and postsynaptic structures. Influenced by local neuronal activity, astrocytes actively phagocytose synapses. Pruning of synapses and clearance of apoptotic cells by astrocytes are mediated through the scavenger receptor multiple EGF-like domains 10 (Megf10), a high-affinity receptor for C1q. Astrocytes also participate in dynamic regulation of vascular tone, in part through astrocyte-astrocyte communication mediated through gap junctions and calcium waves modulated by neuronal activity; support blood-brain barrier and glymphatic (see below) integrity through extension of foot processes to vascular structures and expression of aquaporin-4 water channels; and carry out additional metabolic functions essential for maintenance of neuron health.

Recent work has highlighted the transcriptional and functional heterogeneity of reactive astrocytes that, depending on the context, could regulate inflammation, promote neurotoxicity, or aid in

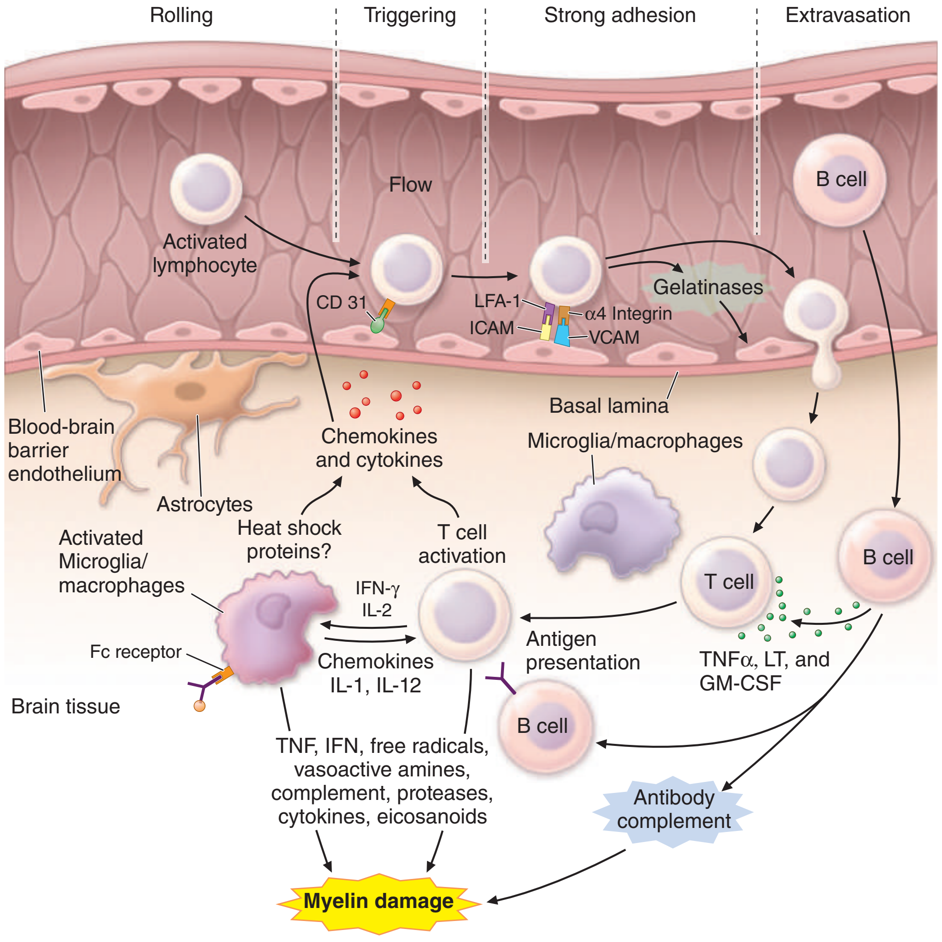
Rolling
Triggering
Strong adhesion
Extravasation
Flow
B cell
Activated lymphocyte
CD 31
LFA-1
ICAM
α4 Integrin
VCAM
Gelatinases
Basal lamina
Blood-brain barrier endothelium
Astrocytes
Chemokines and cytokines
Microglia/macrophages
Activated Microglia/ macrophages
Heat shock proteins?
T cell activation
T cell
B cell
IFN-γ
IL-2
Chemokines
IL-1, IL-12
Antigen presentation
TNFα, LT, and GM-CSF
Fc receptor
Brain tissue
B cell
TNF, IFN, free radicals, vasoactive amines, complement, proteases, cytokines, eicosanoids
Antibody complement
Myelin damage

FIGURE 435-2 A model for experimental allergic encephalomyelitis (EAE). Crucial steps for disease initiation and progression include peripheral activation of preexisting autoreactive T cells; homing to the central nervous system (CNS) and extravasation across the blood-brain barrier; reactivation of T cells by exposed autoantigens; secretion of cytokines; activation of microglia and astrocytes and recruitment of a secondary inflammatory wave; and immune-mediated myelin destruction. ICAM, intercellular adhesion molecule; IFN, interferon; IL, interleukin; LFA-1, leukocyte function-associated antigen-1; TNF, tumor necrosis factor; VCAM, vascular cell adhesion molecule.

protection and repair. As examples, in response to cytokines such as granulocyte-macrophage colony-stimulating factor (GM-CSF) produced by infiltrating lymphocytes, astrocytes can augment neuroinflammation via secretion of chemokines such as CCL2 and CXCL10 that attract lymphocytes and monocytes to the CNS. In MS and other neuroimmune disorders, production of B-cell activation factor (BAFF) by astrocytes within lesions promotes the survival and activation of pathogenic autoreactive B cells. Under other conditions, astrocytes appear to limit inflammation, through expression of tumor necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL) induced by interferon γ produced by natural killer cells, or by secretion of the regulatory cytokine IL-27. And in still other settings, astrocytes can secrete factors that promote neuronal survival and axon regeneration, for example, by secretion of wingless-type MMTV integration site 1 (Wnt1), a neuroprotective factor for dopaminergic neurons.

It is also now clear that bidirectional crosstalk between immune system cells and astrocytes can profoundly influence the brain's response to injury. In the best-studied example, activated microglia, through secretion of IL-1a, TNF, and C1q, can induce astrocytes to transform to a reactive disease-promoting phenotype. Such cells lose the capacity to phagocytose synapses and myelin debris and become toxic in vitro to neurons and mature oligodendrocytes, via different mechanisms including complement-mediated damage and production of nitric oxide (NO) and free radicals. Interestingly, oligodendrocyte progenitor cells (OPCs), abundant in active lesions of MS **(Chap. 455)** within the inflammatory milieu, are resistant to astrocyte-mediated killing. In ALS models modeled on human ALS-causing mutations in *SOD1* and *C9orf72* genes, astrocyte expression of the mutant genes promoted neurotoxicity, indicating a non-cell-autonomous disease mechanism mediated through astrocytes. Similarly, in AD, astrocytes have been implicated in several genetic forms of AD, including amyloid accumulation mediated through apolipoprotein E (ApoE), which in the nervous system is expressed primarily in astrocytes. Thus, reactive astrocytes could promote damage in disorders as varied as MS **(Chap. 455)**, AD **(Chap. 442)**, PD **(Chap. 446)**, and ALS **(Chap. 448)**, despite the distinct etiologies and pathologies of these conditions.

OLIGODENDROCYTES AND MYELIN

Myelin is the multilayered insulating substance that surrounds axons and speeds impulse conduction by permitting action potentials to jump between naked regions of axons (nodes of Ranvier) and across myelinated segments. Oligodendrocytes contact axons at paranodes, where sodium and potassium channels essential for saltatory conduction are clustered. Molecular interactions between the myelin membrane and axon are required to maintain the stability, function, and normal life span of both structures. The process of myelination is directed both by axon-derived cues as well as the physical properties of the axon-membrane curvature. Importantly, ongoing neuronal activity influences both the differentiation of oligodendrocytes as well as the extent of myelination, a process referred to as *adaptive myelination*. A single oligodendrocyte usually ensheaths multiple axons in the CNS, whereas in the peripheral nervous system (PNS), each Schwann cell typically myelinates a single axon. Oligodendrocytes can increase their surface length by as much as 2000-fold, the maintenance of which places significant metabolic demands on the cell.

Myelin is a lipid-rich material formed by a spiraling process of the membrane of the myelinating cell around the axon, creating multiple membrane bilayers that are tightly apposed (compact myelin) by charged protein interactions. From an evolutionary perspective, an increase in the efficiency of nerve conduction enabled by CNS myelin has permitted expanded neuronal connectivity and brain complexity without the need to dramatically increase brain size. CNS myelin has also evolved to provide critical neuroprotective support to axons. Not surprisingly, myelin pathologies cause or contribute to many neurologic conditions. Several clinically important neurologic disorders result from inherited mutations in myelin proteins **(Chap. 457)**. Constituents of myelin have a propensity to be targeted as autoantigens in autoimmune demyelinating disorders, in part due to molecular similarities between myelin and microbial cell wall constituents **(Chaps. 455 and 456)**. Alterations in myelin are increasingly recognized as a contributor to cognitive changes associated with aging, including dementia. For example, in AD models, oligodendrocyte damage can promote the deposition of disease-associated amyloid plaques. Oligodendrocytes can also be the site of disease-associated protein deposition in some neurodegenerative disorders, for example, with toxic accumulation of α-synuclein aggregates in multiple system atrophy **(Chap. 451)**. Evidence also supports a role for oligodendrocyte and myelin pathology in disorders as varied as ALS **(Chap. 448)**, traumatic brain injury **(Chap. 454)**, and stroke **(Chap. 437)**, among other conditions.

Premyelinating OPCs are highly motile cells that migrate extensively during development and in the adult brain following injuries to the myelin sheath. OPCs migrate along the inner (or abluminal) surface of endothelial cells, a process regulated by *Wnt* pathway signaling and upregulation of the chemokine receptor Cxcr4 that drives their attachment and retention to the vasculature. In the normal adult brain, large numbers of OPCs are widely distributed. Following demyelination, remyelination is largely dependent on OPCs that differentiate into myelin-producing oligodendrocytes and produce characteristic thinly remyelinated fibers. In some situations, a second population of regenerating oligodendrocytes derived from neural stem cells can mediate more effective remyelination, with thicker lamellae and greater functional preservation of axons. A C^{14} labeling study from human MS lesions indicated that a third population of nonmitotic preexisting oligodendrocytes may represent an additional source of remyelinating cells.

Both acquired demyelinating disorders, such as MS, and inherited ones, such as Pelizaeus-Merzbacher disease (duplication or deletion of the CNS proteolipid protein gene) and adrenoleukodystrophy (mutations in the *ABCD1* gene responsible for transport of very-long-chain fatty acids into the peroxisome for degradation), are associated with progressive axonal loss. Loss of oligodendrocyte support can produce axonal damage through a variety of mechanisms, including reductions in the supply of glucose and other essential nutrients; an increased axonal workload; impaired glutamate and calcium buffering; mitochondrial damage; loss of neurotrophins; enhanced susceptibility to reactive oxygen species including nitric oxide; and failure to maintain normal synapses.

A number of molecules have been identified that regulate oligodendrocyte differentiation and myelination, including LINGO-1, hyaluronan, chondroitin sulfate proteoglycan, the *Wnt* pathway, Notch (and its receptor Jagged), fibrinogen, and the M1 muscarinic receptor Chrm1, all of which are inhibitory. Other targets are the retinoic acid receptor RXRγ, vitamin D, and thyroid hormone, all of which promote oligodendrocyte maturation. All are also potential targets for myelin repair therapies. In the EAE model of autoimmune demyelination (Fig. 435-2), oligodendrocyte-specific knockout of *Chrm1* improved remyelination, protected axons, and restored function, directly demonstrating that remyelination can be neuroprotective following injury. A pivotal trial of a monoclonal antibody against LINGO-1 in patients with acute optic neuritis failed to improve clinical outcomes, a disappointing result given that the antibody appeared to have promising clinical effects in an earlier phase 2 trial. In a preliminary trial of chronic optic neuritis, a promising result was reported with clemastine, an antihistamine and M1 muscarinic receptor antagonist, raising hope that clinically effective remyelination might be achievable even in a chronic demyelinating condition.

LYMPHATICS OF THE CENTRAL NERVOUS SYSTEM

Two recently identified lymphatic structures of the CNS are the glymphatic and deep dural lymphoid systems, responsible for clearance of debris in the CNS, and likely also serving roles in immune surveillance. The brain has traditionally been considered to lack a classical lymphatic system, and immune responses against antigens are less effectively generated in the CNS than in other organ systems, a concept termed *immune privilege*. However, the immune privilege status of the brain is relative and not absolute. Also, given the high metabolic demands of the brain, some mechanism for efficient removal of solute and debris must be present. One well-established pathway involves the passive flow of solutes from the brain parenchyma into the cerebrospinal fluid (CSF), and their exit via the arachnoid granulations, as well as along cranial and spinal nerve roots to a series of lymphoid structures located in the cribriform plate, nasal mucosa, and elsewhere.

The *glymphatic system* derives its name from a distinctive architecture involving lymphoid-like structures and astroglial cells. CSF synthesized in the arachnoid villi circulates through the ventricles and subarachnoid space surrounding the convexities of the brain and spinal cord and exits through conduits surrounding arterioles penetrating into the brain parenchyma. These spaces are lined by endothelial cells internally and by astrocyte foot processes that form the external walls. Aided by arterial propulsion, CSF moves out of these specialized conduits and into astrocytes via foot processes rich in aquaporin-4 water channels, and then in the interstitium of brain parenchyma, picks up solutes and particulate debris that are then carried to perivenous spaces where they passage to exit the brain and drain into the lymphatic system. In mice, knockout of aquaporin-4 markedly reduced the flow of interstitial fluids in the brain, underscoring the critical role of astrocyte uptake of CSF in this process. Interstitial flow in the CNS is also impaired with aging, possibly related to changes in astrocytic aquaporin-4 expression. A fascinating aspect of the glymphatic system is that the transport of fluids and solutes accelerates with sleep, arguing for a critical role for sleep in promoting clearance of debris needed to meet the high metabolic demands of the nervous system. Furthermore, in disease models, aggregated proteins associated with neurodegenerative disease, such as β-amyloid associated with AD **(Chap. 442)**, were also more efficiently cleared during sleep. Indeed, in mice genetically engineered to produce excess β-amyloid and develop AD-like cognitive decline, sleep deprivation increased accumulation of amyloid plaques. Glymphatic pathways are also likely to represent an important egress pathway for lymphocytes in the CNS and a route for lymphocyte encounters with CNS antigens in cervical lymph nodes. In this regard, deep cervical lymph nodes may be a site for antigen-specific stimulation of B cells in MS **(Chap. 455)**.

A second pathway consists of a plexus of small lymphatic-like vessels located on the external surface of meningeal arteries and deep dural sinuses (including the sagittal and transverse sinuses), structures that exit the brain along the surface of veins and arteries and drain to the deep cervical lymph nodes. These conduits appear to represent a lymphoid drainage system distinct from vascular endothelium. These sinus-associated lymphoid structures may be most important in clearing solutes from the CSF, in contrast to the glymphatic system that likely functions to remove waste products from the brain interstitium; however, the exact functions of these two systems and their interrelationships are only beginning to be understood.

MICROBIOTA AND NEUROLOGIC DISEASE

The human microbiome **(Chap. 484)** represents the collective set of genes from the 10^{14} organisms living in our gut, skin, mucosa, and other sites. The aggregate number of genes encoded by bacteria living in and around us (i.e., the microbiome) outnumber our own genome by a factor of 100, and these can encode a wide variety of molecules that directly or indirectly affect nervous system development, maintenance, and function. Distinctive microbial communities have been found in individuals with different genetic backgrounds, ethnicities, diets, and environments. While progressively evolving with age, in any individual, the predominant gut microbiota can be remarkably stable over decades, but also can be altered by exposure to certain microbial species (e.g., by ingestion of probiotics) or the frequent use of antibiotics.

Gut microbes can shape immune responses through the interaction of their metabolism with that of humans. These gut–brain interactions are likely to be important in understanding the pathogenesis of many autoimmune neurologic diseases. For example, mice raised in a germ-free environment (or treated with broad-spectrum antibiotics) are resistant to EAE, an effect associated with decreased production of proinflammatory cytokines and conversely more production of the immunosuppressive cytokines IL-10 and IL-13 as well as an increase in regulatory T and B lymphocytes. Intestinal microbiota from patients with MS were found to promote EAE when transferred to germ-free mice, possibly due to imbalances between bacterial species that promote inflammation (e.g., *Akkermansia muciniphila* and *Acinetobacter calcoaceticus*) and those that induce regulatory immune responses (e.g., *Parabacteroides distasonis, Prevotella copri,* and certain species of *Clostridium*). Recent studies have shown that the transcription factor aryl hydrocarbon receptor (AHR) is an important regulator of the differentiation of murine and human regulatory T cells and microglia. AHR is activated by endogenous physiologic ligands, but also by metabolic products of the commensal flora.

In addition to nonspecific effects on immune homeostasis mediated by cytokines and regulatory lymphocytes, some microbial proteins could trigger a cross-reactive immune response against a homologous protein in the nervous system, a mechanism termed *molecular mimicry*. Examples include cross-reactivity between the astrocyte water channel aquaporin-4 and an ABC transporter permease from *Clostridia perfringens* in neuromyelitis optica **(Chap. 456)**; human leukocyte antigen (HLA) molecules with *A. muciniphila* peptides of *A. muciniphila* and Epstein-Barr virus in MS **(Chap. 455)**; the neural ganglioside Gm1 and similar sialic acid–containing structures from *Campylobacter jejuni* in Guillain-Barré syndrome **(Chap. 458)**; and the sleep-promoting protein hypocretin and hemagglutinin from an H1N1 influenza virus in narcolepsy **(Chap. 33)**.

Microbial genes also encode molecules that can affect development of neurons and glia, and influence myelination and plasticity. Bacterial-derived short-chain fatty acids, for example, regulate production of brain-derived neurotrophic factor (BDNF). Bacteria produce a variety of neurotransmitters including γ-aminobutyric acid (GABA) and serotonin and other neuroactive peptides that can modulate the hypothalamic-pituitary axis. Gut microbiota influence development and activity of the enteric nervous system, which communicates bidirectionally with the CNS via the vagus nerve that innervates the upper gut and proximal colon. As these gut–brain relationships become better defined, a role for the microbial environment in the pathogenesis of a much wider spectrum of neurologic conditions and behaviors seems likely, extending well beyond the traditional boundaries of immune-mediated pathologies. In this regard, it has long been known that gut bacteria can influence brain function, based mostly on classic studies demonstrating that products of gut microbes can worsen hepatic encephalopathy, forming the basis of treatment with antibiotics for this condition.

Mice that develop in a germ-free environment display less anxiety, lower responses to stressful situations, more exploratory locomotive behaviors, and impaired memory formation compared with non-germ-free counterparts. These behaviors were related to changes in gene expression in pathways related to neural signaling, synaptic function, and modulation of neurotransmitters. Moreover, this behavior could be reversed when the germ-free mice were co-housed with non-germ-free mice. As noted above, intestinal microbiota were found to be required for the normal development and function of brain microglia, potentially linking these behavioral effects to specific cellular targets in the CNS. Some actions of gut microbial species on microglia also appear to be sex- and age-specific.

The vagus nerve has been implicated in anxiety- and depression-like behaviors in mice. Ingestion of *Lactobacillus rhamnosus* induced changes in expression of the inhibitory neurotransmitter GABA1b in neurons of the limbic cortex, hippocampus, and amygdala, associated with reduced levels of corticosteroids and reduced anxiety- and depression-like behaviors. Remarkably, these changes could be blocked by vagotomy.

A related area of emerging interest is in a possible contribution of the gut microbiome to autism and related disorders. Children with autistic spectrum disorders (ASD) have long been known to have gastrointestinal disturbances, and the severity of dysbiosis appears to correlate with the severity of autism. In several murine models of autism, manipulation of the gut microbiome ameliorated the behavioral abnormalities. A role for the proinflammatory cytokine IL-17 was implicated as a possible mediator in producing the ASD-like changes. In mice, an ASD-like disorder could be induced in offspring after injecting the pregnant mother with the viral RNA mimic, polyinosinic:polycytidylic acid (poly I:C); oral treatment of offspring with *Bacteroides fragilis* corrected a range of autistic behaviors in these mice and also improved gastrointestinal dysfunction. These preclinical data led to a small uncontrolled study of fecal gut transplantation in children with ASD that reported encouraging results but will need to be confirmed in rigorous controlled trials.

There has been considerable interest in the possible role of the microbiome in a variety of vascular, traumatic, and neurodegenerative diseases, possibly mediated through actions on innate immunity and microglia. In SOD1 transgenic ALS-prone mice, a germ-free environment exacerbated disease progression, and symptoms could be ameliorated by increasing levels of *A. muciniphila* or its nicotinamide (vitamin B_3) metabolite; a small preliminary clinical trial of nicotinamide supplementation subsequently reported encouraging results in ALS patients.

In a PD model, injection of misfolded α-synuclein into the gut triggered deposition of α-synuclein in the brain, an effect that was blocked when the vagus nerve was severed. This supported a prion mechanism (see below) for PD pathogenesis, in which vagal transport of aggregated α-synuclein might seed the CNS via the vagus nerve. The concept of a gut origin of PD is also consistent with clinical and pathologic studies, and by some epidemiologic data suggesting that vagotomy may be protective against PD. In transgenic mice that overexpress human α-synuclein, transplantation of intestinal microbiota from PD patients worsened motor deficits, α-synuclein deposition, and neuroinflammation, and conversely, pathology could be ameliorated with germ-free housing conditions or antibiotic treatment. In other work, a protein of *Escherichia coli*, named Curli, has been shown to misfold and potentially serve as a template for subsequent propagation of misfolded α-synuclein. The possibility that a bacterial protein could initiate the cascade of events leading to PD is an extraordinary, but still unproven, hypothesis.

Thus, there are a variety of mechanisms whereby the microbiome—in gut, skin, and other bodily surfaces—can modulate healthy brain function or influence susceptibility to and expression of a variety of brain diseases. The specific mechanisms are likely to vary with each condition but are likely to include promoting autoimmunity through molecular mimicry; disrupting immune homeostasis; modulating brain function via bacterial metabolites that travel in the circulation and cross the blood-brain barrier; influencing neural activity through signals generated in the enteric nervous system; or even the possibility that prion-like aggregates are formed in the gut and passage to the CNS via the vagus nerve. Although no proven benefit of modulating the microbiome exists for any brain disease today, this situation is likely to change as information about specific disease-associated microbial communities rapidly increases, along with improved methods to fine-tune their proportions and overall diversity in humans.

PATHOLOGIC PROTEINS, PRIONS, AND NEURODEGENERATION (FIG. 435-3)

■ PROTEIN AGGREGATION AND CELL DEATH

The term *protein aggregation* has become widely used to describe easily recognizable hallmarks of neurodegeneration. While such neuropathologic hallmarks including plaques, neurofibrillary tangles (NFTs), and inclusion bodies are often thought to cause neurologic dysfunction, numerous new discoveries over the past several decades have rendered this view increasingly unlikely. Instead, protein aggregates represent

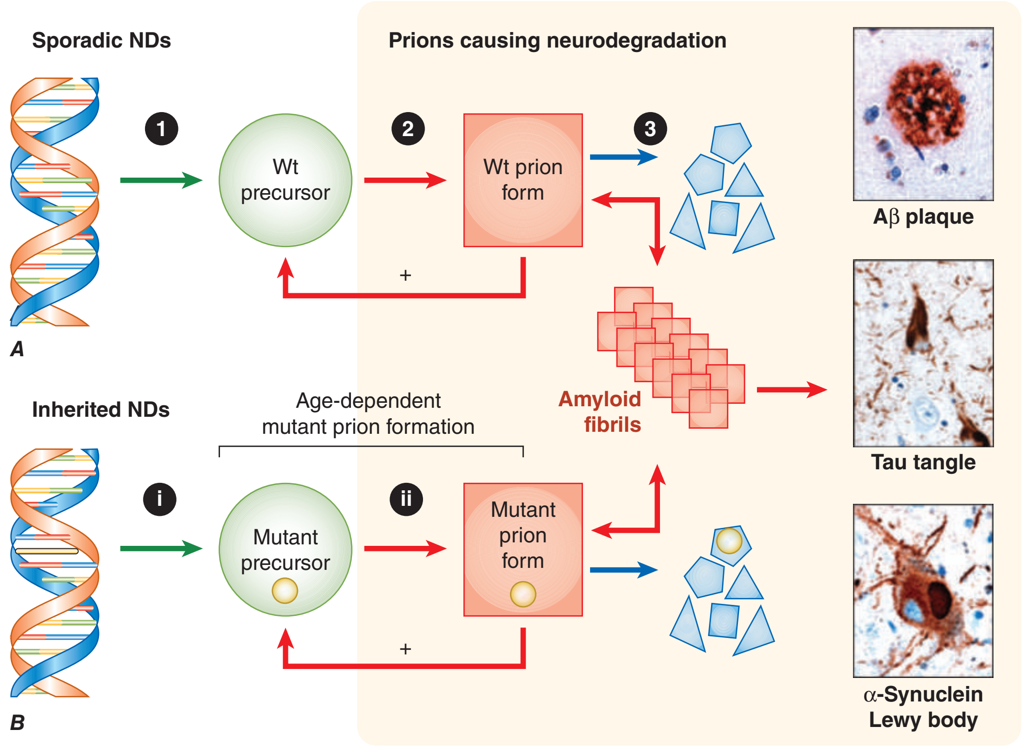

FIGURE 435-3 Neurodegeneration caused by prions. *A.* In sporadic neurodegenerative diseases (NDs), wild-type (Wt) prions multiply through self-propagating cycles of posttranslational modification, during which the precursor protein (*green circle*) is converted into the prion form (*red square*), which generally is high in β-sheet content. Pathogenic prions are most toxic as oligomers and less toxic after polymerization into amyloid fibrils. The small polygons (*blue*) represent proteolytic cleavage products of the prion. Depending on the protein, the fibrils coalesce into Aβ amyloid plaques in Alzheimer's disease (AD), neurofibrillary tangles in AD and other tauopathies, or Lewy bodies in Parkinson's disease (PD) and dementia with Lewy bodies. Drug targets for the development of therapeutics include: (1) lowering the precursor protein, (2) inhibiting prion formation, and (3) enhancing prion clearance. ***B.*** Late-onset heritable neurodegeneration argues for two discrete events: The (*i*) first event is the synthesis of mutant precursor protein (*green circle*), and the (*ii*) second event is the age-dependent formation of mutant prions (*red square*). The *highlighted yellow bar* in the DNA structure represents mutation of a base pair within an exon, and the *small yellow circles* signify the corresponding mutant amino acid substitution. *Green arrows* represent a normal process; *red arrows*, a pathogenic process; and *blue arrows*, a process that is known to occur but unknown whether it is normal or pathogenic. *(Used with permission of Annual Reviews, from [Biology and genetics of prions causing neurodegeneration, SB Prusiner, 47:601, 2013 permission conveyed through Copyright Clearance Center, Inc.)*

accumulations of toxic proteins that may become less harmful when they are sequestered into plaques, NFTs, and inclusion bodies.

Most mutations in the amyloid precursor protein (APP) gene causing familial AD are concentrated within the Aβ peptide. Many of these mutations increase production of the Aβ42 peptide composed of β-amyloid with 42 amino acids, which has an increased propensity to adopt a prion conformation, as compared to β-amyloid with 40 amino acids. In contrast, mutations in the APP that reduce the production of β-amyloid protect against the development of AD and are associated with preserved cognition in the elderly. The most common cause of NFTs is AD, but the precise molecular events that produce tangles are unknown. Mutations in the *MAPT* gene encoding tau stimulate NFT formation in familial frontotemporal dementia, inherited progressive supranuclear palsy, and other familial tauopathies. Like AD, the majority of most tauopathies as well as PD are sporadic.

The second most common neurodegenerative disease is PD. The saga of α-synuclein and PD begins in 1996 with the identification of a mutation in a family of Greek descent. With this family and others, there were sufficient patients to establish genetic linkage. Soon thereafter, immunostaining showed that α-synuclein was present in Lewy bodies, and the following year staining of glial cytoplasmic inclusions (GCIs) was identified in the brains of deceased multiple-system atrophy (MSA) patients. Subsequently, brains from deceased MSA patients transmitted the disease to transgenic mice, establishing that the α-synucleinopathies are prion diseases. Before the α-synuclein (*SCNA*) gene was found to cause familial PD, other genes such as the leucine-rich repeat kinase 2 (*LRRK2*) were found to modify the onset of PD; other similar PD modifier genes include *parkin, PINK1,* and *DJ-1*. PINK1 is a mitochondrial kinase (see below), and DJ-1 is a protein involved in protection from oxidative stress. Parkin, which causes autosomal recessive early-onset PD-like illness, is a ubiquitin ligase. The characteristic histopathologic feature of PD is the Lewy body, an eosinophilic cytoplasmic inclusion that contains both neurofilaments and α-synuclein. Huntington's disease (HD) and cerebellar degenerations are associated with expansions of polyglutamine repeats in proteins, which aggregate to produce neuronal intranuclear inclusions. Familial ALS is associated with superoxide dismutase (SOD1) mutations and cytoplasmic inclusions containing superoxide dismutase. An important finding was the discovery that ubiquitinated inclusions observed in most cases of ALS and the most common form of frontotemporal dementia are composed of TAR DNA-binding protein 43 (*TDP-43*). Subsequently, mutations in the *TDP-43* gene, and in the fused in sarcoma gene (*FUS*), were found in familial ALS. Both of these proteins are involved in transcription regulation as well as RNA metabolism.

Another key mechanism linked to cell death is mitochondrial dynamics, which refers to the processes involved in movement of mitochondria, as well as in mitochondrial fission and fusion, which play a critical role in mitochondrial turnover and in replenishment of damaged mitochondria. Mitochondrial dysfunction is strongly linked to the pathogenesis of a number of neurodegenerative diseases such as Friedreich's ataxia, which is caused by mutations in an iron-binding protein that plays an important role in transferring iron to iron-sulfur clusters in aconitase and complex I and II of the electron transport chain. Mitochondrial fission is dependent on the dynamin-related proteins (Drp1), which bind to its receptor Fis, whereas mitofusins 1 and 2 (MFN 1/2) and optic atrophy protein 1 (OPA1) are responsible for fusion of the outer and inner mitochondrial membrane, respectively. Mutations in *MFN2* cause Charcot-Marie-Tooth neuropathy type 2A, and mutations in *OPA1* cause autosomal dominant optic atrophy. Both β-amyloid and mutant huntingtin protein induce mitochondrial fragmentation and neuronal cell death associated with increased activity of Drp1. In addition, mutations in genes causing autosomal recessive PD, *parkin* and *PINK1*, cause abnormal mitochondrial morphology and result in impairment of the ability of the cell to remove damaged mitochondria by autophagy.

As noted above, one major scientific question is whether protein aggregates directly contribute to neuronal death or whether they are merely secondary bystanders. A focus in all the neurodegenerative diseases is on small-protein aggregates termed *oligomers*. How many monomers polymerize into a particular disease-specific oligomer has been elusive. Whether oligomers are the toxic species of β-amyloid, α-synuclein, or proteins with expanded polyglutamines such as the one causing HD remains to be established. Protein aggregates are usually ubiquitinated, which targets them for degradation by the 26S component of the proteasome. An inability to degrade protein aggregates could lead to cellular dysfunction, impaired axonal transport, and cell death by apoptotic mechanisms.

Autophagy is the degradation of cystolic components in lysosomes. There is increasing evidence that autophagy plays an important role in degradation of protein aggregates in the neurodegenerative diseases, and it is impaired in AD, PD, FTD, and HD. Autophagy is particularly important to the health of neurons, and failure of autophagy contributes to cell death. In HD, a failure of cargo recognition occurs, contributing to protein aggregates and cell death.

There is other evidence for lysosomal dysfunction and impaired autophagy in PD. Mutations in glucocerebrosidase (*GBA*) are associated with 5% of all PD cases as well as 8–9% of patients with dementia with Lewy bodies. Notably, glucocerebrosidase and enzymatic activity are reduced in the substantia nigra of sporadic PD patients. α-Synuclein is degraded by chaperone-mediated and macro autophagy. The degradation of α-synuclein has been shown to be impaired in transgenic mice deficient in glucocerebrosidase, and α-synuclein inhibits the activity of glucocerebrosidase; thus, there appears to be bidirectional feedback between α-synuclein and glucocerebrosidase.

The retromer complex is a conserved membrane-associated protein complex that functions in endosome to Golgi transport. The retromer complex contains a cargo selective complex consisting of VPS35, VPS26, and VPS29, along with a sorting nexin dimer. Mutations in VPS35 were shown to be a cause of late-onset autosomal dominant PD. The retromer also traffics APP away from endosomes, where it is cleaved to generate β-amyloid. Deficiencies of VPS35 and VPS26 were also identified in hippocampal brain tissue from AD. A potential therapeutic approach to these diseases might therefore be to use chaperones to stabilize the retromer and reduce the generation of β-amyloid and α-synuclein.

PRIONS AND NEURODEGENERATIVE DISEASES

As we have learned more about the etiology and pathogenesis of the neurodegenerative diseases, it has become clear that the histologic abnormalities that were once curiosities, in fact, are likely to reflect the etiologies. For example, the amyloid plaques in kuru and Creutzfeldt-Jakob disease (CJD) are filled with the PrP^{Sc} prions that have assembled into fibrils. The past three decades have witnessed an explosion of new knowledge about prions. For many years, kuru, CJD, and scrapie of sheep were thought to be caused by slow-acting viruses, but a large body of experimental evidence argues that the infectious pathogens causing these diseases are devoid of nucleic acid. Such pathogens are called prions, which are composed of host-encoded proteins that adopt alternative conformations that undergo self-propagation. Prions impose their conformations on the normal, precursor proteins, which in turn become self-templating, resulting in faithful copies; most prions are enriched for β-sheet and can assemble into amyloid fibrils.

Similar to the plaques in kuru and CJD that are composed of PrP prions, the amyloid plaques in AD are filled with Aβ prions that have polymerized into fibrils. This relationship between the neuropathologic findings and the etiologic prion was strengthened by the genetic linkage between familial CJD and mutations in the PrP gene, as well as (as noted above) between familial AD and mutations in the *APP* gene. Moreover, a mutation in the *APP* gene that prevents Aβ peptide formation was correlated with a decreased incidence of AD in Iceland.

The heritable neurodegenerative diseases offer an important insight into the pathogenesis of the more common sporadic ones. Although the mutant proteins that cause these disorders are expressed in the brains of people early in life, the diseases do not occur for many decades. Many explanations for the late onset of familial neurodegenerative diseases have been offered, but none is supported by substantial experimental evidence. The late onset might be due to a second event

TABLE 435-1 Prion-Based Classification of Neurodegenerative Diseases

NEURODEGENERATIVE DISEASE	CAUSATIVE PRION PROTEINS
Creutzfeldt-Jakob disease (CJD)	PrP^{Sc}
Kuru	PrP^{Sc}
Gerstmann-Sträussler-Scheinker disease (GSS)	PrP^{Sc}
Fatal insomnia	PrP^{Sc}
Bovine spongiform encephalopathy (BSE)	PrP^{Sc}
Scrapie	PrP^{Sc}
Chronic wasting disease (CWD)	PrP^{Sc}
Feline spongiform encephalopathy	PrP^{Sc}
Transmissible mink encephalopathy	PrP^{Sc}
Alzheimer's disease (AD)	$A\beta \rightarrow$ tau
Down syndrome	$A\beta \rightarrow$ tau
ALS-parkinsonism dementia complex (PDC) of Guam	$A\beta \rightarrow$ tau
Parkinson's disease (PD)	α-Synuclein
Dementia with Lewy bodies	α-Synuclein
Multiple-system atrophy	α-Synuclein
Frontotemporal dementias (FTDs)	Tau, TDP43, FUS (C9orf72, progranulin)
Posttraumatic FTD	
Chronic traumatic encephalopathy (CTE)	Tau
Amyotrophic lateral sclerosis (ALS)	SOD1, TDP43, FUS (C9orf72)
Huntington's disease (HD)	Huntingtin

in which a mutant protein, after its conversion into a prion, begins to accumulate at some rather advanced age. Such a formulation is also consistent with data showing that the protein quality-control mechanisms diminish in efficiency with age. Thus, the prion forms of both wild-type and mutant proteins are likely to be efficiently degraded in younger people but are less well handled in older individuals. This explanation is consistent with the view that neurodegenerative diseases are disorders of the aging nervous system.

A new classification for neurodegenerative diseases can be proposed based on not only the traditional phenotypic presentation and neuropathology but also the prion etiology **(Table 435-1)**. An expanding body of experimental data has accumulated connecting prions in each of these illnesses. In addition to kuru and CJD, Gerstmann-Sträussler-Scheinker disease (GSS) and fatal insomnia in humans are caused by PrP^{Sc} prions. In animals, PrP^{Sc} prions cause scrapie of sheep and goats, bovine spongiform encephalopathy (BSE), chronic wasting disease (CWD) of deer and elk, feline spongiform encephalopathy, and transmissible mink encephalopathy (TME). Similar to PrP, Aβ, tau, α-synuclein, superoxide dismutase 1 (SOD1), and possibly huntingtin all adopt alternative conformations that become self-propagating, and thus, each protein can become a prion and be transferred to synaptically connected neurons. Moreover, each of these prions causes a distinct constellation of neurodegenerative diseases.

Evidence for a prion etiology of AD comes from a series of transmission experiments initially performed in marmosets and subsequently in transgenic mice expressing the mutant APP from which the Aβ peptide is derived (Table 435-1). Synthetic mutant Aβ peptides folded into a β-sheet-rich conformation exhibited prion infectivity in cultured cells. Studies of the tau protein have shown that it not only features in the pathogenesis of AD but also causes the frontotemporal dementias including chronic traumatic encephalopathy, which has been reported in both contact sport athletes and military personnel who have suffered traumatic brain injuries. A series of incisive studies using cultured cells and Tg mice have demonstrated that both tau and Aβ prions are found together in the brains of AD patients. These findings indicated that AD is a double-prion disease (Table 435-1); unexpectedly, two more double-prion diseases have been identified recently. Patients with Down syndrome, from 6–72 years of age, all had both Aβ and tau prions in their brains with the frequent diagnosis of AD. The third double-prion disease has been found in the Chamorro people on Guam as well as Japanese living on the Kii peninsula: both groups of people develop ALS with dementia and both have Aβ and tau prions in their brains.

In contrast to Aβ and tau prions, α-synuclein prions cause very different illnesses, i.e., PD, dementia with Lewy bodies (DLB), and MSA. Brains from MSA patients inoculated into Tg(*SCNA**A53T) mice died ~90 days after intracerebral inoculation, whereas mutant α-synuclein (A53T) prions formed spontaneously in Tg mouse brains that killed recipient Tg mice in ~200 days (Table 435-1).

For many years, the most frequently cited argument against prions was the existence of strains that produced distinct clinical presentations and different patterns of neuropathologic lesions. Some investigators argued that the biologic information carried in different prion strains could be encoded only within a nucleic acid. Subsequently, many studies demonstrated that strain-specified variation is enciphered in the conformation of PrP^{Sc}, but the molecular mechanisms responsible for the storage of this biologic information remains enigmatic. The neuroanatomical patterns of prion deposition have been shown to be dependent on the particular strain of prion. Convincing evidence in support of this proposition has been accumulated for PrP, Aβ, tau, and α-synuclein prions. The most persuasive information on prion strains comes from studies in yeast where the tools of yeast genetics allowed inciteful investigations to be performed in ways that could not be accomplished in mammals.

Although the number of prions identified in mammals and in fungi continues to expand, the existence of prions in other phylogeny remains undetermined. Some mammalian prions perform vital functions and do not cause disease; such nonpathogenic prions include the cytoplasmic polyadenylation element-binding (CPEB) protein, the mitochondrial antiviral-signaling (MAVS) protein, and T cell–restricted intracellular antigen 1 (TIA-1).

Many but not all prion proteins adopt a β-sheet-rich conformation and appear to readily oligomerize as this process becomes self-propagating. Control of the self-propagating state of benign mammalian prions is less well understood than that of pathogenic mammalian prions, which appear to multiply exponentially. We do not know if prions multiply as monomers or as oligomers; notably, the ionizing radiation target size of PrP^{Sc} prions suggests it is a trimer. The oligomeric states of pathogenic mammalian prions are thought to be toxic; larger polymers, such as amyloid fibrils, seem to be a mechanism for minimizing toxicity.

The development of drugs designed to inhibit the conversion of the normal precursor proteins into prions or to enhance the degradation of prions focuses on the initial step in prion accumulation. Although a dozen drugs that cross the blood-brain barrier have been identified that prolong the lives of mice infected with scrapie prions, none has been identified that extends the lives of Tg mice that replicate human CJD prions. Despite doubling or tripling the length of incubation times in mice inoculated with scrapie prions, all of the mice eventually succumb to illness. Because all of the treated mice develop neurologic dysfunction at the same time, the mutation rate as judged by drug resistance is likely to approach 100%, which is much higher than mutation rates recorded for bacteria and viruses. Mutations in prions seem likely to represent conformational variants that are selected for in mammals where survival becomes limited by the fastest-replicating prions. The results of these studies make it likely that cocktails of drugs that attack a variety of prion conformers will be required for the development of effective therapeutics.

NEURAL STEM CELL BIOLOGY

Normal and genetically modified ("transgenic") mice are the most widely used model systems to study features of human nervous system diseases. However, modeling genetic diseases in rodents is limited to the relatively small number of monogenic human diseases where the specific gene mutations are known and is further limited by species differences. The latter can be particularly important in brain regions such as the cerebral cortex that have undergone significant evolutionary expansion in humans. These shortcomings, which likely contribute to the low probability that therapeutic efficacy translates from animal models to humans, can potentially be overcome through stem cell models that enable the use of human cells and tissues to model human diseases. The advent of new stem cell technologies is transforming our understanding of the pathobiology of human neurologic diseases. Stem cell platforms are being used to screen for therapeutic agents, to uncover adverse drug effects, and to discover novel therapeutic targets.

Among the most exciting recent advances in stem cell technology is the ability to convert somatic cells, either skin fibroblasts or blood cells, into pluripotent stem cells known as induced pluripotent stem cells (iPSCs). This technology has introduced an entirely new and powerful approach to study the pathobiology of heritable diseases. Pluripotent stem cells can be easily obtained through minimally invasive procedures such as a skin biopsy or blood sample and converted to pluripotency through application of a cocktail of reprogramming factors to create iPSCs. Initially, a set of four programming factors, Oct3/4, Klf-4, Sox2, and c-Myc, was delivered to cells using lentiviruses that stably integrated the reprogramming factor genes into the iPSC genome, potentially altering disease phenotypes and also abrogating expression of native genes at the DNA sites where the factors integrated. Newer techniques have been developed that use nonintegrating approaches such as through the use of Sendai virus, messenger RNA (mRNA), or episomal vectors that circumvent these problems. Once created, iPSC lines can be expanded indefinitely to produce a limitless supply of stem cells. These cells are the starting material for the derivation of specific cell types based on protocols that use small molecules, proteins, or direct gene induction to recapitulate developmental programs. Most current protocols derive neuronal progenitors through dual-SMAD inhibition, a step that involves the use of small-molecule inhibitors to block endoderm and mesodermal cell fates, thereby creating neural cells by default. Multiple protocols have been developed over the last decade for creating large numbers of human neuron progenitor cell types and directing them toward specific nervous system cell fates, including neuron subtypes from multiple regions of brain and spinal cord as well as retinal cells, glial cells including astrocytes and oligodendrocytes, immune cells, and peripheral nervous system cells.

The primary medical benefit of iPSC technology is that it enables the creation of patient-specific cells or tissues that are genetically matched to individual subjects. This approach enables the study of not only monogenetic disorders but also sporadic forms of disease and complex polygenic disorders including those with unidentified risk loci. Furthermore, by deriving iPSC cell lines from multiple patients, it is possible to explore how disease phenotypes may vary according to genetic background. Another approach that has been used to generate specific neuron and glial cell types from somatic cells such as fibroblasts is through direct reprogramming. This approach relies on a cocktail of specific transcription factors to directly convert somatic cells into the alternate desired cell type. This approach bypasses the epigenetic reset that accompanies cells as they are reprogrammed to a pluripotent state. The advantage of this approach is that age-related epigenetic signatures are not erased, so that derived neurons may more readily reflect diseases that manifest in older cells.

Despite the advantages of using in vitro models of nervous system diseases derived from patient-specific iPSCs, several potential roadblocks remain. There are no standard reprogramming or derivation protocols, and the different methods can result in considerable variability in the disease phenotypes reported by different laboratories. Confidence in the specificity of a particular phenotype is therefore increased if it has been validated across multiple laboratories. There is also the problem of inherent variability between patient lines that may result from their different genetic backgrounds. One solution, available only in the case of monogenic disorders, is to use isogenic controls generated using gene editing, such as with CRISPR-Cas9 technology, to create disease and control lines on an identical genetic background. However, because differences in genetic background can influence the penetrance of a particular trait, it will still be necessary to compare disease lines from multiple patients to discern a true disease phenotype. For polygenic disorders where the causative mutations are unknown, it will not be possible to create isogenic controls, and in these situations, the best strategy for improving reliability and sensitivity is to compare lines from multiple patients.

ORGANOIDS

Most nervous system disorders, including ASD, schizophrenia, PD, AD, and ALS, are complex disorders, resulting from an unknown combination of gene mutations, and manifest not only in specific cell types but also in alterations of the local tissue environment. These disorders are difficult to model in animals, but they are approachable using three-dimensional human iPSC stem cell models, often referred to as "organoids." Organoids are derived from pluripotent stem cells that are directed along a tissue-specific lineage through the timed application of growth factors, genes, or small-molecule activators or inhibitors, and allowed to aggregate into three-dimensional structures. With time, cell intrinsic programs are spontaneously engaged and the cellular aggregates begin to self-organize and develop into structures that recapitulate the complex topographical and cellular diversity of normal organ development. In this way, it has been possible to create, at least in part, in vitro brainlike organoids that resemble parts of the human brain at early stages of development. When allowed to develop from an anterior neural tube stage, these structures can become heterogeneous, containing regions with forebrain, midbrain, and/or hindbrain identity and can often include retina-like structures. The high degree of variability in such "whole-brain organoids" can be a liability for controlled studies and can be reduced by the use of directed protocols that restrict outcomes to more defined brain regions, such as forebrain, cortex, or ganglionic eminence. A variety of protocols have now been developed to generate organoids with specific regional identity, and fusing organoids of different regional identity with each other has been used to reproduce cellular interactions such as neuronal migration across regions. Many protocols are focused on modeling cortical development, and they can reproduce developmental features including a diversity of progenitor and neuronal cell types topographically distributed within ventricular and subventricular progenitor regions and rudimentary cortical layers. However, the organoids follow a human developmental timetable and still remain at stages roughly comparable to late fetal development even after 6–9 months of culture. Moreover, they lack key cell types such as endothelial cells, pericytes, and microglia, and have few if any astrocytes or oligodendrocytes. Recently, it has become possible to derive these cell types independently from stem cells and then combine them with brain organoids to create tissue-diverse brain organoids that also contain, for example, vascular or immune cells. Nonetheless, while still only reflecting rudimentary organizational and compositional features, organoids have become attractive models to study human brain development and the pathophysiology of human nervous system diseases in the context of a partially organized brainlike structure.

BRAIN DEVELOPMENT AND DEVELOPMENTAL DISORDERS: MICROCEPHALY AND LISSENCEPHALY

Transcriptional analysis has suggested that the neurons produced by most stem cell protocols resemble early- to mid-gestational stages of human brain development. The immaturity of stem cell–derived human neurons may limit their utility for modeling adult diseases, but it does make them ideally suited for the study of brain development and the pathophysiology of neurodevelopmental disorders.

Primary autosomal recessive microcephaly (MCPH) is a rare neurodevelopmental disorder producing severe microcephaly with simplified cortical gyration and intellectual disability. MCPH was one of the first disorders to be studied using cerebral organoids. Mutations in genes encoding microtubule spindle components and spindle-associated proteins are the most frequent causes of congenital microcephaly. Among them is cyclin-dependent kinase 5 related activator protein 2 (CDK5RAP2). Skin fibroblasts derived from a single microcephalic patient carrying a mutation in CDK5RAP2 were used to generate four iPSC lines. Cerebral organoids grown from these cell lines contained fewer proliferating progenitor cells and showed premature neural differentiation compared to wild-type controls. Introducing functional CDK5RAP2 by electroporation partially rescued the disease phenotype, supporting the notion that failure of the founder population of neural progenitors to properly expand underlies the smaller brain. This study demonstrated that brain organoids derived from patients with microcephaly can be used to reproduce features of the disease but did not reveal new insights or disease features of CDK5RAP2 microcephaly that had not already been described in mouse models.

Organoids, however, are well-suited to model human microcephaly because the phenotype manifests at early prenatal ages and can be assessed by organoid size, which can serve as a proxy for a microcephaly phenotype. For example, a loss-of-function screen of candidate microcephaly human genes was carried out in organoids using CRISPR-lineage tracing; 173 potential microcephaly genes were tested, and novel mechanisms involved in brain size control were discovered including a pathway involved in endoplasmic reticulum function and extracellular matrix production.

Cortical organoids have also been used to model lissencephaly or "smooth brain." Miller-Dieker syndrome (MDS), a severe congenital form of lissencephaly, was modeled in organoids and features of the human disease were observed that had not been noted in murine models. Classical lissencephaly is a genetic neurologic disorder associated with intellectual disability and intractable epilepsy, and MDS is a severe form of the disorder. Cortical folding in humans begins toward the end of the second trimester, but gyrencephaly depends upon earlier events such as neural progenitor cell proliferation and neuronal migration, which can be modeled in organoids. The human organoid model of MDS exhibited several neural progenitor cell phenotypes that had already been reported in mouse models, including altered mitotic spindle orientation and neuronal migration defects. But the organoids also displayed a mitotic defect in a specific neural stem cell subtype, the outer radial glia cell (oRG), that had not been observed in mice. oRG cells are enriched in the outer subventricular zone, a proliferative region that is large in primates and not present in rodents. These cells are particularly numerous in the developing human cortex and are thought to underlie the developmental and evolutionary expansion of the human cortex. oRG cells from MDS patients behaved abnormally and had arrested or delayed mitoses. MDS organoids also identified noncell autonomous defects in *Wnt* signaling as an underlying mechanism. These insights into mechanistic and cell-type-specific features of human disease highlight how organoid technology can provide new and valuable perspectives on the pathophysiology of disorders of in utero development.

■ ACQUIRED NEURODEVELOPMENTAL DISORDERS: ZIKA

The outbreak of Zika virus (ZIKV) and associated microcephaly cases in the Americas provided a test case for the utility of brain organoids to model acquired human microcephaly. Despite a correlation between Zika infection rates and the incidence of congenital microcephaly, compelling evidence that ZIKV caused microcephaly was lacking in the early phases of the epidemic. The causal link between ZIKV and congenital microcephaly was buttressed by two studies in 2016 that used human iPSC-derived neural progenitor cells and organoids to demonstrate ZIKV tropism for human neural progenitor cells. Neural progenitor cells (radial glia) were readily infected in vitro with subsequent progenitor cell death and involution of organoid size. Forebrain organoids were further used to highlight the role of the flavivirus entry factor, AXL, in determining viral tropism, and were also used to explore the disease mechanism by demonstrating upregulation of the innate immune receptor toll-like receptor 3 (TLR) in response to ZIKV infection. Stem cell–derived models of human brain development have also demonstrated centrosomal abnormalities in radial glia and alteration in the cleavage plane of mitotic radial glia associated with premature neural differentiation. Mouse models are also being used to study the pathophysiology of congenital ZIKV syndrome, but the availability of unlimited numbers of human neural cells produced using stem cell technology has enabled high-throughput screening assays to test libraries of clinically approved compounds for potential therapeutic agents. This strategy has already highlighted several compounds that could potentially help protect against ZIKV microcephaly.

■ NEURODEVELOPMENTAL DISORDERS: AUTISM AND SCHIZOPHRENIA

ASDs are complex and heterogeneous neurodevelopmental disorders usually manifesting in childhood with difficulties in social interaction, verbal and nonverbal communication, and repetitive behaviors. The cellular and molecular mechanisms underlying ASD are thought to arise at stages of fetal brain development, making them well-suited for exploration using human iPSC-derived disease models. iPSC-derived neurons have been used to study the pathophysiology of disorders associated with ASD that are caused by monogenic mutations, including fragile X, Rett, and Timothy syndromes.

Fragile X is the most common heritable cause of intellectual disability, affecting 1 in 4000 males and 1 in 8000 females, and is a leading genetic cause of ASD. Patients also have speech delay, growth and motor abnormalities, hyperactivity, and anxiety. The causative mutation lies in the *FMR1* gene and produces a CGG triplet repeat expansion from a normal number of 5–20 to >200, leading to epigenetic silencing of the *FMR1* gene and loss of the fragile X mental retardation protein. The epigenetic mechanism means that unlike a simple gene deletion that would lead to ubiquitous loss of expression, the FMR1 locus becomes hypermethylated and epigenetically silenced during differentiation; thus, FMR1 protein is expressed by the early embryo and becomes absent only around the beginning of the second trimester. Interestingly, this expression pattern is recapitulated during cellular differentiation in stem cell models. Pluripotent fragile X stem cell lines have been derived from embryos identified through preimplantation genetic diagnosis and by reprogramming skin fibroblasts from fragile X patients to create iPSC lines. In both cases, *FMR1* was expressed by the pluripotent stem cells but underwent transcriptional silencing following differentiation. Fragile X stem cell lines can therefore be used to study the mechanism of FMR1 silencing, an effort that is ongoing. Neurons generated from fragile X iPSC cells reproduce features observed in neurons from transgenic *FMR1* mouse models and patients, including stunted neurites with decreased branching, increasing confidence in the iPSC model. In addition to providing a model that can be used to study disease pathogenesis, fragile X iPSC-derived neurons could be used to screen for potential therapeutic agents or gene-editing strategies to remove the repressive epigenetic marks induced by the mutation and rescue the phenotype.

Rett syndrome is an X-linked neurodevelopmental disorder with dominant inheritance caused by a mutation in the *MECP2* gene. Because males carrying one copy of the defective gene usually die in infancy, most patients are girls. Random inactivation of the X chromosome in girls results in mosaic cellular expression of the mutation that circumvents fatality and produces a variable phenotype. The symptoms are present in early childhood and include microcephaly associated with developmental delay, autistic-like behaviors and cognitive dysfunction, seizures, and repetitive motor actions; these then progress to include difficulties with gait, swallowing, and breathing before usually stabilizing with patients surviving to adulthood. The pathophysiology of Rett syndrome is presumed to involve abnormal epigenetic regulation leading to decreased transcriptional repression of genes whose overexpression produces the disease phenotype, although this concept has been contested. In one of the first studies to use iPSC modeling to study Rett syndrome, it was discovered that when fibroblasts from patients were reprogrammed to pluripotent stem cells, X inactivation was erased. In apparent recapitulation of endogenous events, X chromosome inactivation re-occurred during neuronal differentiation, producing a mosaic of cells carrying the mutant gene intermingled with normal cells. Rett neurons had fewer dendritic spines and synapses, smaller cell bodies, and reduced network activity. Another iPSC model of Rett syndrome highlighted the potential role of altered inhibitory function. Rett neurons were found to have a deficit of potassium/chloride cotransporter (KCC2) that is developmentally regulated and normally leads to a switch in GABA signaling from excitatory at embryonic ages to inhibitory by birth. In Rett neurons, KCC2 expression level was low, and the functional switch in GABA effects was delayed, contributing to some of the disease features and possibly accounting for the developmental onset of the disease. One curious feature of some iPSC Rett lines was that despite the mosaic expression of the mutation, disease phenotypes were observed in all cells. Possibly, this could reflect a noncell autonomous effect, but as in all iPSC disease models, confidence in disease-specific features will be increased when similar phenotypes are seen across multiple independent studies.

Timothy syndrome, another severe neurodevelopmental disease associated with ASD, has been modeled using iPSC-derived organoids. Timothy syndrome is caused by a mutation in the *CACNA1C* gene coding for a voltage-gated calcium channel, and neuron defects in Timothy syndrome organoids were rescued by selectively altering calcium channel activity. In one study, two separate organoids were produced with different regional identity, one represented neocortex and one a more ventral structure known as the medial ganglionic eminence, which is the source of most cortical interneurons. The two organoids were then fused together to allow the interneurons to migrate into the cortex, mimicking their endogenous behavior. The ability to model interneuron migration led to the discovery of a cell-autonomous migration defect in the disease-carrying neurons.

The majority of nervous system diseases, including ASD, are polygenic and cannot be modeled in animals but can be modeled using patient-derived iPSCs. For example, a subset of patients with ASD have large head size, and a cohort of patients with this phenotype was used to generate iPSCs that were converted to neural progenitor cells and forebrain neurons. The progenitors had an accelerated cell cycle and produced an excess of inhibitory interneurons and had exuberant cellular overgrowth of neurites and synapses. This last feature is in contrast to the decrease in spines and synapses observed in other iPSC models of ASD such as fragile X and Rett syndrome and underscores the need for replication and validation of purported disease phenotypes given the high variability based on differences between stem cell lines, protocols, patient genetic background, and other factors. Moreover, the clinical features of most neuropsychiatric diseases reflect disorders in processes such as circuit formation and refinement that occur after birth and may be difficult to capture at the fetal stage of development reflected in stem cell models.

Patient stem cells have also been used by multiple groups to study the pathophysiology of schizophrenia, producing a variety of diverse and sometimes contradictory results. Reports claim obvious phenotypes such as disruptions in the adherens junctions of forebrain radial glia or aberrant neuronal migration, although such gross abnormalities observed at the equivalent of in utero stages of development seem very unlikely to underlie a disease that usually manifests at adolescence or young adulthood. Other studies report abnormalities related to abnormal microRNA expression, disordered cyclic AMP and *Wnt* signaling, abnormal stress responses, diminished neuronal connectivity, fewer neuronal processes, problems with neuronal differentiation, and mitochondrial abnormalities, among others. While the pathophysiology of as complex a neurodevelopmental disorder as schizophrenia may be multidimensional, it is unclear which, if any, of the reported findings in iPSC models reflect the true pathology of schizophrenia. Progress will likely depend on the adoption of more standard and reproducible protocols, more rigorous identification of cell types, markers of regional identity, and indicators of maturity.

ALZHEIMER'S DISEASE

As noted above, the leading concept of AD pathogenesis, the amyloid hypothesis, suggests that an imbalance between production and clearance of β-amyloid leads to excessive accumulation of β-amyloid peptide and the formation of NFTs within neurons, composed of aggregated hyperphosphorylated tau proteins. Additionally, aggregates of amyloid fibrils are deposited outside neurons in the form of neuritic plaques. Among the causes of familial AD are mutations in genes involved in β-amyloid production, including amyloid precursor protein (APP) and presenilin 1 and 2. Shortly after the introduction of iPSC technology, human stem cell–derived neurons were generated from patients carrying mutations in AD-causative genes as well as from sporadic AD cases. The disease neurons developed hallmarks of AD including intracellular accumulation of β-amyloid and phosphorylated tau, as well as secretion of APP cleavage products, features that could be reduced by adding β- or γ-secretase inhibitors or β-amyloid-specific antibodies. The neurons also demonstrated other disease features observed in postmortem AD tissues. However, extracellular β-amyloid aggregation and NFTs were not robustly modeled in these two-dimensional systems, presumably because secreted factors were able to readily diffuse away. The use of three-dimensional organoids to model AD overcame this limitation, presumably by recreating a more faithful extracellular matrix. Organoid models promoted the aggregation of β-amyloid, and more readily recapitulated the pathologic features of AD, including the formation of NFTs and neuritic plaques.

It is hoped that the new stem cell models, particularly organoid models, will accelerate our understanding of AD by enabling the study of human disease-carrying cells in a quasi in situ setting. These new models may lead to discovery of novel druggable targets and new diagnostic and prognostic biomarkers. One concern is that the pathogenic features of AD usually appear in the sixth or seventh decade of life and progress slowly over years, while most protocols for the derivation of human cortical neurons generate cells over weeks or months and most remain comparable to immature neurons at fetal stages of development. Nonetheless, these young cells have been used to model neurodegenerative diseases such as AD and HD that strike patients in middle to late adulthood. Possibly the onset of disease phenotype is accelerated in stem cell models due to increased cellular stress, which appears to be a feature of stem cell culture, or disease features may actually have a subtle onset at earlier stages than generally suspected. Indeed, 3-year-old children at genetic risk of developing early-onset AD appear to have smaller hippocampal size and lower scores on memory tests than children in a nonrisk group. The phenotypes of adult neurodegenerative diseases that are visible at fetal stages may or may not correspond to those manifest at later, adult stages, but they may offer the possibility of devising preventative strategies effective at very early stages of the disease.

CELL TYPE DISORDERS: ALS AND HUNTINGTON'S DISEASE

In diseases such as ALS, PD, and HD, that mostly target specific neuron subtypes, stem cells provide an ideal means to study the vulnerable human cell populations. By enabling the production of unlimited numbers of normal and diseased human midbrain dopaminergic neurons for the study of PD, medium spiny striatal neurons for HD, and spinal and cortical motor neurons for ALS, iPSC approaches have the potential to transform our understanding and management of these diseases. Stem cell–derived neurons serve as platforms to explore mechanisms of cell vulnerability, to screen drugs for neural protection, and potentially to derive neurons for replacement therapy.

AMYOTROPHIC LATERAL SCLEROSIS

One of the first protocols for producing neurons of a specific subtype from embryonic stem cells recapitulated normal developmental programs to generate mouse spinal motor neurons. Pluripotent mouse stem cells underwent neural induction and adopted a caudal identity through the application of retinoic acid, and subsequently adopted motor neuron fate through the action of Sonic hedgehog (Shh), a ventralizing factor. Generating human motor neurons proved more complex, requiring additional steps, such as early exposure to the growth factor, FGF2. The first application of stem cell–derived motor neurons to study ALS involved the use of mouse motor neurons generated from transgenic mice expressing a mutation in the superoxide dismutase 1 (*SOD1*) gene, the most common mutation responsible for familial ALS. Only 5–10% of ALS cases are familial, but the known mutations provide a useful entry point to tease apart the causative pathophysiology. Mutations in *SOD1* produce ALS through a toxic gain of function for which the mechanism remains unclear, despite the use of multiple transgenic animal and iPSC models. The use of mouse embryonic stem cell–derived motor neurons, however, demonstrated that toxic factors secreted by SOD1 astrocytes contribute to the death of motor neurons. Interestingly, stem cell–derived interneurons were spared, indicating a specific vulnerability of motor neurons. These findings helped establish the notion that a non-cell-autonomous toxic mechanism contributes to ALS pathogenesis and may ultimately lead to novel treatment strategies. These findings also highlight that modeling the full pathophysiology of ALS may require the reproduction of a complex environment including motor neurons, astrocytes, and possibly additional cell types such as microglia. A variety of approaches including co-culture of specific cell types, three-dimensional spinal cord organoids, and microfluidic organ-on-chip models are being explored to achieve a more complete facsimile of spinal cord organization. Similar to other neurologic disorders where a clearly defined phenotype has been observed in human stem cell–derived models, there is hope that drug screening using human disease-expressing cells will identify a potential therapeutic compound.

HUNTINGTON'S DISEASE

HD is caused by an expansion in CAG triplet repeats in the huntingtin gene, which leads to an expanded polyglutamine tract in the huntingtin protein. HD is dominantly inherited, with symptoms of cognitive decline and uncontrollable gait and limb motions beginning in the third to fifth decade of life with progression to dementia and death approximately 20 years later. Mutant huntingtin causes a toxic gain of function, with the degree of effect related to the CAG repeat length. For example, a CAG length of 40–60 repeats produces adult-onset HD, whereas repeats of 60 or more produce juvenile-onset disease. Although it has been 25 years since the discovery of this causative mutation, the disease mechanism remains poorly understood. Excess huntingtin protein and protein fragments accumulate in specific subtypes of neurons where they misfold and form aggregates that are visible as cellular inclusions. Affected cells eventually die, possibly as a result of metabolic toxicity. The medium spiny neurons of the striatum are the most vulnerable neurons, spurring ongoing attempts to produce replacement cells derived from stem cells, but neuron loss is widespread including in the cortex, complicating a cell replacement approach for this disease. HD iPSCs have been generated from patients with various CAG repeat lengths, but those from juvenile-onset disease with the longest repeat lengths have been favored as being most likely to express robust disease phenotypes at an early stage. This is particularly important given the immature stage of maturation of stem cell–derived human neurons. This approach has been able to produce disease phenotypes observed in patients including huntingtin protein aggregation, decreased metabolic capacity, increased oxidative stress with mitochondrial fragmentation, and apoptosis enhanced by withdrawal of growth factor support. However, many of these phenotypes were observed in pluripotent cells prior to neural differentiation and in neural progenitors and a broad array of CNS neurons in contrast to the cell type–specific features of the disease. Nonetheless, neurons that assumed striatal fate appear to be more vulnerable to stress and apoptosis than other cell types. As with other iPSC models of nervous system diseases, there have so far been few efforts to validate results in multiple iPSC lines having different genetic backgrounds but with similar CAG repeat lengths. An HD consortium has been formed to address this problem by generating a series of iPSC lines from multiple patients. An alternative strategy to validate disease phenotypes has been to use gene editing to create isogenic iPSC lines that are corrected to produce wild-type control and HD iPSC lines against the same genetic background.

FUTURE PERSPECTIVES

Despite early successes, it may prove difficult to reconstitute neurodegenerative disease conditions in human cells in vitro over a short course of time because the pathogenic changes of degenerative diseases progress slowly and commence in the later stages of life. The differentiation and maturation of human neurons from stem cell lines occur over a span of months, which may not be long enough to establish the aged-brain conditions under which patients develop robust neurodegenerative pathology. Possible manipulation through gene editing or by application of aging-associated stresses, such as DNA-damaging agents or proteasome inhibitors, may accelerate the expression of degenerative phenotypes in human iPSC-derived cellular models. Stem cell–derived organoid models are also ideal platforms to apply methods for cellular-level visualization such as CLARITY and multielectrode recording techniques to better evaluate three-dimensional organoid structures and explore early-forming circuits. These applications are only just beginning.

Two-dimensional cell cultures are ideal for production and evaluation of large numbers of specific cells of a particular identity but may not provide the complex extracellular environment necessary to model certain disease processes, such as extracellular protein aggregation. These features can be best modeled using three-dimensional organoids, but current methods do not reproduce all the relevant features of brain tissue. Optimization will be needed to better reproduce the cellular composition of brain, including endothelial cells, astrocytes, microglia, and oligodendrocytes. It may also be necessary to combine different brain regions generated separately, possibly by fusion of tissues such as dorsal cortex, subpallium, thalamus, retina, and others in so-called "assembloids." One roadblock has been the limited ability to recreate tissues or neurons with regional brain identity, such as hippocampus, thalamus, or cerebellum. However, recently regionalized organoids have been created using innovative microfluidic chambers and the imposition of morphogen gradients such as bone morphogenetic proteins (BMPs) and WNTs. More faithful organoid models could also emerge through the application of bioengineered scaffolds, matrices, or perfusion systems that might allow the growth of larger structures, a feature currently limited due to the emergence of a necrotic core when nonperfused organoids exceed a certain size. Of course, not all aspects of mature brain architecture and function will be modeled by these tissue structures, particularly as they generally represent prenatal stages of development, but perhaps the most precocious events in disease etiology can be captured and investigated, and these may share mechanistic pathways with disease features that manifest at later stages.

The current excitement surrounding human stem cells has more to do with their promise to improve on animal models of disease, for which their potential appears unlimited, rather than on their use as a source for cell-based therapies, where the potential has thus far been relatively limited. Even without new insights into disease pathogenesis, it is likely that iPSC models such as brain organoids will serve as drug-screening platforms for discovery of novel therapeutics and for detection of off-target and toxic effects. The failure of many neurotherapeutic approaches to translate from animal models to clinical practice underscores the need for better predictive models, and stem cell models and brain organoids based on human cells may be ideally suited to bridge this divide.

A CURRENT PERSPECTIVE ON NEURAL STEM CELLS IN THE CLINIC

The prospect of stem cell therapies to treat diseases or injuries of the nervous system has captured the attention of researchers, clinicians, and the public. The pace of research is usually slow and deliberate, but in the stem cell arena, there has been enormous pressure to accelerate the pace of progress in order to bring cell-based therapies to the clinic. Expectations have been raised, and clinics have already begun offering unproven or dangerous treatments to a public that can be, in some situations, ill-informed and vulnerable to exploitation. Nonetheless, there has been remarkable progress over the past few years toward legitimate stem cell–based therapies for a number of nervous system disorders. There are multiple clinical trials underway fueling cautious optimism that stem cells will eventually realize the promise of regenerative therapy for at least some currently untreatable or incurable nervous system diseases.

PARKINSON'S DISEASE

Pursuit of a cell-based therapy for PD using fetal-derived midbrain cells has been ongoing for many decades but with disappointing results. Following anecdotal success in a handful of patients who appeared to improve following striatal grafts of fetal midbrain dopaminergic cells, two National Institutes of Health (NIH)-sponsored double-blind control studies were launched in the 1990s. However, only a small number of younger patients showed some benefit, and several patients developed spontaneous dyskinetic movements related to the therapy. These efforts constituted a failed trial as the treated patients who did not experience side effects failed to improve significantly. However, techniques to extract dopaminergic cells from fetal tissue have improved, and on the basis of encouraging results in individual transplanted patients, some of whom have managed to go off their Parkinson's medication, a new trial of fetal cell transplantation for PD was conducted in Europe. Unfortunately, enrollment was unexpectedly slow and the trial was terminated after several years without reaching the anticipated number of patients. There was no clear clinical improvement among the treated patients, possibly due to difficulty harvesting sufficient fetal cells to reach the targeted number per patient. Despite the disappointing clinical outcome, enthusiasm remains high for a better outcome with stem cell–derived dopamine neurons where, among other differences, cell number is not a limitation.

The dyskinesias that curtailed the NIH trials in the 1990s were eventually ascribed to an abundance of serotonergic neurons that were inadvertently included in some of the cell grafts. Protocols for deriving dopaminergic neurons from stem cells would presumably avoid this complication by providing a more purified cell population, and several groups around the world have been aggressively pursuing a stem cell–based approach to PD. In 2018, researchers from Kyoto University in Japan started a phase 1/2 clinical trial to treat PD using stem cells. The investigators chose to use iPSCs derived from a healthy person who had the most common HLA haplotype in Japan. The iPSCs were used to make dopamine-secreting neurons. Seven patients have had the reprogrammed stem cells surgically delivered into the brain and have been followed for two years postinjection to assess safety and possible efficacy. The U.S. Food and Drug Administration (FDA) recently approved the first clinical trial of a stem cell–derived dopamine neuron for the treatment of PD in the United States. These cells, derived from an embryonic stem cell line, have been delivered to 10 patients in a phase 1 clinical trial to assess safety, tolerability, and preliminary efficacy. A European trial of embryonic stem cell–derived dopaminergic neurons led by scientists in Sweden and the United Kingdom began in 2023. Four patients given a low dose experienced no acute adverse effects, and four more will be treated with a higher dose, and all will be followed for 36 months to evaluate safety and tolerability.

A clinical trial using iPSC-generated dopaminergic neurons to treat patients with PD was begun in 2018 in Japan. The cell line was derived from a healthy individual with the most common HLA haplotype in the Japanese population. The clinical product consisted of 80% dopamine neuron progenitor cells and a population of glial cells. Because the cell line was genetically altered, the karyotype, plasmid survival, and genomic and epigenomic abnormalities were evaluated and known carcinogenic gene mutations were screened. A single-arm, nonrandomized and open-phase 1/2 study was begun in 2018 with ~5 million cells transplanted stereotactically to the putamen bilaterally. Seven patients have been treated thus far.

BATTEN'S DISEASE AND PELIZAEUS-MERZBACHER DISEASE

One of the first cell-based clinical trials for a neurologic disease targeted patients suffering from an untreatable childhood disorder, Batten disease. Batten disease is an autosomal recessive metabolic disorder resulting from an inability to synthesize a lysosomal enzyme critical to brain function. The phase 1 trial involved six patients with infantile and late-infantile forms of the disease who received neural stem cells rather than any specific postmitotic cell type. Neural stem cells derived from donated fetal tissue were expanded in vitro prior to surgical grafting into the brain. This approach was not without risk, as the neural stem cells were proliferating and could potentially form an abnormal growth. The rationale was that the cells would be capable of synthesizing and secreting the missing lysosomal enzyme and would therefore serve as a delivery device. Animal studies using a transgenic mouse model of Batten disease demonstrated rescue, and this promising result led to a small phase 1 trial. The phase 1 study was considered a success as no adverse events were reported and the cells appeared to be safe, though there was no clinical improvement and no clear evidence of whether the cells had dispersed, transformed into neurons or glia, or indeed survived at all. Despite clearing the phase 1 trial, the company did not pursue further trials for Batten disease, but instead initiated clinical trials using the same cell product for several other indications, including an inherited fatal dysmyelination syndrome known as Pelizaeus-Merzbacher disease (PMD). The human neural stem cells have both neurogenic and gliogenic potential, and when delivered to white matter regions in experimental animals, most persisting cells had become oligodendrocytes. This supported use of the cells to promote myelin formation in conditions such as PMD. The company also initiated trials in spinal cord injury. However, the spinal cord trial failed to achieve sufficient benefit in phase 2 and the company ceased its work on stem cell therapies.

SPINAL CORD INJURY

There is a pressing need for novel therapies for spinal cord injury, with >1 million patients worldwide suffering from spinal cord injuries and no effective treatment options. Not surprisingly, there has been intense interest in achieving a stem cell treatment for this condition and dozens of early-stage clinical trials, and anecdotal treatment results have been reported by investigators around the globe. The vast majority have not been blinded controlled trials, but rather individual reports treating a handful of patients, and somewhat surprisingly, most are using mesenchymal stem cells (MSCs) or hematopoietic stem cells that normally generate either bone, cartilage, fat, or blood cells. As described below, the rationale for the use of MSCs for neurologic conditions is based on vague and poorly understood mechanisms of action.

A series of stem cell trials designed to treat subacute spinal cord injury is underway in the United States and Europe using neural stem cells or their derivatives as potential therapeutic agents. The first to enter clinical trials in the United States was based on a protocol designed to generate oligodendrocytes from pluripotent embryonic stem cells. Evidence of efficacy was obtained in animal models following surgical grafting of cells to sites of spinal cord injury. However, evidence of myelination of host axons was minimal, and other mechanisms were invoked for improvement in gait, including trophic support and immune modulation. Regulatory permission for a phase 1 trial for subacute midthoracic injury was initially stalled by concern over abnormal growths at sites of cell deposit in some animals, but this was satisfactorily addressed and patient trials commenced. However, following a change in leadership, the stem cell program was terminated. The program was acquired by another company that has resumed the spinal cord injury trial and received regulatory approval to advance to include cervical-level injuries. The current phase 1/2a multicenter clinical trial is an open-label, single-arm trial testing three sequential escalating doses administered 21–42 days after injury in 25 patients with subacute severe cervical spinal cord injuries. No adverse events have been reported for 21 patients at 2 years posttreatment. A later-stage comparative clinical trial is now planned to probe for possible efficacy.

In 2020, an open-label, single-arm clinical trial for subacute spinal cord injury patients was begun in Japan using iPSC-derived neural stem and progenitor cells injected directly into the damaged spinal cord. The rationale for injecting neuronal precursor cells into the injured spinal cord is not clear as no specific mechanism of action has been demonstrated. Trophic support and neuroimmune modulation have been proposed, and while studies in mice show that newborn neurons can form synapses onto host spinal neurons, how this could lead to improved locomotor or sensory function remains unclear. Results of the trial have yet to be reported.

Initial clinical trials for stem cell therapies that address grave medical conditions such as spinal cord injury often involve small numbers of patients, and larger pivotal trials that can confirm clinical benefit can take a long time to conduct. In a controversial move designed to accelerate the regulatory approval process, in 2014 the Ministry of Health of Japan authorized the use of conditional and time-limited (CTL) approval for regenerative medical products. This designation can be applied when safety has been demonstrated but efficacy has not yet been fully established and enables the manufacturer to market and sell the product. CTL approval converts to full approval if postmarketing clinical data demonstrate clinical benefit, but if no convincing clinical benefit can be demonstrated within 7 years, the product must be withdrawn.

The effort to bring MSC treatments to patients with currently untreatable neurologic conditions has consequently had a recent boost in Japan. Autologous bone marrow–derived stem cells were evaluated in an open-label trial involving 13 patients who had experienced severe spinal cord injury 1 month prior to treatment. Mesenchymal stem cells were taken from the patient's own bone marrow, expanded in number, and delivered back to the patient by intravenous infusion. Six months after treatment, 12 of the patients improved by at least one level on the American Spinal Injury Association impairment scale that ranks muscle contraction and touch, although patients with the same or even greater degree of injury can show spontaneous improvement. As is the case for most MSC therapies, the proposed mechanism of action was quite vague, including reducing inflammation, protecting existing neurons, or replacing damaged neurons. It is also unclear how intravenous delivery could accomplish any of the proposed actions. In 2018, on the basis of the results, unpublished at the time, Japan's health ministry gave conditional (CTL) approval for the treatment, called Stemirac. This became the first stem cell therapy for spinal cord injury to receive government approval for sale to patients. But the approval of a therapy that may carry risk following a small, unblinded, and uncontrolled study without actual proof of efficacy raised considerable concern among scientists in the stem cell community. Charging patients for such an unproven therapy raises even more ethical concerns. Patients in Japan can now be charged for their treatment while trials to test efficacy are still proceeding.

AMYOTROPHIC LATERAL SCLEROSIS

The possibility of treating ALS by replacing dying motor neurons with stem cell–derived substitutes has excited interest, but this prospect seems very remote. Even if new neurons are able to integrate into spinal cord circuits and become properly innervated, they would have to grow long axons that would take many months to years to project to appropriate targets and attract myelinating Schwann cells. Furthermore, cells would need to be grafted at multiple spinal cord and brainstem levels, and the upper motor neuron deficit would need to be treated by replacing projecting neurons in the motor cortex. An additional complication is the recent finding that spinal motor neurons have unique segmental identity, and replacement cells might need to be generated with a range of molecular identities in order to integrate at multiple spinal levels. This would still leave unaddressed the toxic effects recently shown to be produced in ALS by diseased astrocytes and microglia that could attack the replacement cells. A more tractable near-term solution would be to graft support cells that could rescue or protect endogenous motor neurons from damage. This approach was tried in a mouse model of ALS. Human fetal stem cell–derived neural progenitor cells engineered to express glial cell line–derived neurotrophic factor (GDNF), a growth factor known to provide trophic support for neurons, were grafted to the spinal cord of young ALS mice. The cells dispersed and were able to rescue motor neurons, a very promising result, but disappointingly, the animals became weak and died at the same rate as untreated control animals. ALS is a deadly disease with no known treatment; thus in the hope that patients will respond differently from mice, a phase 1/2a clinical trial based on this approach was approved by the FDA in 2016 and completed in 2019. Patients had the GDNF-producing progenitor cells surgically grafted unilaterally into the lumbar spinal cord. Although disease progression appeared to slow in some treated patients, this was not statistically significant. Of note, several patients developed painful schwannomas, a result of off-targeting deposit of the GDNF-secreting cells close to the dorsal horn root entry zone. Autopsy study of 13 of the 18 treated patients who died within 2 years of treatment showed that the grafted cells had survived. Based on these results, a new trial has begun. With the notion that rescuing lower motor neurons alone may not be sufficient to arrest progression of a disease that affects both upper and lower motor neurons, this trial aims to enroll 16 patients who will have the GDNF-secreting progenitor cells grafted to the motor cortex in an area controlling hand movement. The goal is to try to slow the progression of the upper motor deficit associated with ALS.

Among the many MSC-based clinical trials for ALS, two are particularly notable. Corestem, a stem cell company in South Korea, launched a phase 1 open-label study demonstrating the safety and feasibility of intrathecal injections of autologous bone marrow–derived MSCs in seven patients with ALS. This was followed by a phase 2 trial that demonstrated safety and efficacy for slowing disease progression. On the basis of these results, Corestem received conditional approval in South Korea in 2014 to market the first stem cell therapy for ALS. By 2021, >300 patients had received this cell treatment. However, full approval is contingent on the results of a randomized, double-blind, placebo-controlled, multicenter phase 3 study, which began in 2021. The trial aims to enroll 115 patients who will receive repeated intrathecal injections of the autologous cells. The goal is to slow or prevent ALS progression and delay death for up to 3 years.

The importance of conducting proper phase 3 clinical trials to determine therapeutic efficacy in ALS is underscored by the recent experience of BrainStorm Cell Therapeutics. In 2016, the company reported preliminary positive results for its bone marrow MSC cell therapy in an uncontrolled study of nearly 50 ALS patients. Based on those results, the company launched a phase 3, multicenter, placebo-controlled, randomized, double-blind trial of 189 ALS patients. In 2020, the company reported that there was no significant clinical improvement in the treatment group. Interestingly, despite the failed clinical trial, a public campaign led by ALS patients and advocates called on the FDA to approve the stem cell treatment. The social media response prompted the FDA to take the unusual step of releasing a public statement underscoring the lack of efficacy, and in 2023, an FDA advisory committee voted against approval. The company responded by withdrawing its Biologics License Application and announcing a phase 3b trial that will focus on early-stage patients.

EPILEPSY

A phase 1/2 clinical trial was launched in 2022 using stem cell–derived inhibitory interneurons to treat medically intractable temporal lobe epilepsy. Somatostatin (SST)-expressing inhibitory interneurons, matching the predominant cortical interneuron subtype, were derived from an embryonic stem cell line. Following demonstration of efficacy in a mouse model of temporal lobe epilepsy, 10 patients who had been determined to have a seizure focus in the hippocampus had stem cell–derived interneurons surgically delivered to the epileptic hippocampus. The first five patients received a low dose followed by five patients who received a higher dose. All 10 patients will be followed for 2 years after treatment. As in the PD clinical trials, all patients received antirejection therapy for the first 12 months. The current standard of care for patients with intractable temporal lobe epilepsy is surgical ablation or resection of the epileptic hippocampus. Surgically treated patients can have seizure remission but often have cognitive decline, particularly if the focus is in the dominant hemisphere. Early indications are that the cell-based therapy can not only reduce seizures but also spare or improve cognitive function.

MACULAR DEGENERATION

Because iPSC technology enables the generation of pluripotent stem cells from adult somatic cells, it has enabled the production of patient-specific autologous cells for cell replacement therapy. Following Shinya Yamanaka's discovery of iPSCs, the Japanese government has invested in bringing iPSC-derived cell therapy to the clinic. Banks of iPSC lines selected to capture the diversity of HLA haplotypes found in the Japanese population have been produced in the hope that these will allow cell therapies to be matched to individual patient haplotypes in order to avoid immune rejection. While these stem cell banks were still being produced, the first Japanese study to use stem cells was approved in 2013 and involved patients who were to receive customized therapy using cells derived from their own skin fibroblasts. The targeted disease was age-related macular degeneration, a common cause of blindness in the elderly that results from loss of retinal pigment epithelial (RPE) cells. RPE cells are relatively easy to generate from pluripotent stem cells, making replacement therapy an attractive target in this condition. A challenge is to coax the replacement cells to recreate an epithelium in the subretinal space. The Japanese approach involves surgical insertion of a biofilm seeded with RPE cells into the retina. One patient was treated with their own stem cell–derived RPE cells, but prior to treating a second patient, the genome of the RPE cell line was sequenced, and a mutation was discovered in a known oncogene. The trial was halted and a decision made to discontinue the effort for customized cell therapy in favor of using RPE cells derived from the national repository of banked iPSC lines which undergo extensive gene sequencing and quality controls. This outcome serves as a caution for the challenges involved in bringing a customized cell therapy to the clinic.

■ MESENCHYMAL STEM CELLS

By far the largest number of human trials have been performed using MSCs sourced from a variety of sites including bone marrow, peripheral blood, adipose tissue, umbilical cord, and other sites. Interest in the potential utility of MSCs for regenerative therapy began with the optimistic report that bone marrow stem cells were pluripotent and capable of generating nerve and heart muscle as well as blood cells. The possibility that easily obtainable MSCs could be used to regenerate injured or diseased cells or organs to treat diseases ranging from stroke, neurodegenerative disease, myocardial infarct, and even diabetes, generated enormous enthusiasm. The enthusiasm proved irresistible to many, and even after the initial reports were discredited—MSCs turned out not to be pluripotent stem cells as initially thought—a veritable flood of papers began to appear claiming disease-modifying activity of MSCs in mouse models of a wide range of degenerative disease and injury models. But when it became clear that the MSCs were not transforming into or generating new neurons or cardiac myocytes, alternative mechanisms of action were invoked, including the release of trophic factors, cytokines, or inflammatory modulators that were credited with producing their remarkable restorative effects. The relative ease with which blood or adipose tissue can be harvested from patients or donors and MSCs extracted has led to a rapidly expanding number of clinical trials for conditions ranging from stroke and MS to AD, ALS, and PD. Furthermore, a loophole in the regulatory framework of the FDA allows autologous cell therapy to escape regulation provided that the cells have not been significantly processed. This lax regulation has spawned a veritable industry of stem cell clinics making unsubstantiated claims of success in treating nervous system diseases. Patients have died from treatments in unregulated clinics operating in countries around the world, and three patients became blind in a well-publicized incident following stem cell treatments delivered by a Florida clinic. The "stem cells" were derived from the patients' own fat tissue and blood. These activities represent the dark side of the stem cell revolution perpetrated by practitioners who exploit the desperation of patients and their families. Legitimate and effective stem cell therapies will emerge over time, but given the prevalence and abundance of misleading information available on the Internet and elsewhere, a trusted and well-informed physician can play a key role in helping patients navigate the current cell therapy minefield.

■ MSCS FOR TRAUMATIC BRAIN INJURY

An allogeneic bone marrow–derived MSC line received conditional marketing approval in Japan in 2024 for the indication of improving chronic motor paralysis resulting from traumatic brain injury. The MSCs were transiently transfecting with the human Notch-1 intracellular domain gene to promote FDF-2 secretion in order to "enhance their ability to regenerate nerve cells" according to the pharmaceutical company that developed the cell-based therapy. The approval followed results of a phase 2 clinical trial conducted in Japan and the United States. Forty-six patients with moderate to severe traumatic brain injury and chronic motor deficits had MSCs stereotactically infused into an area of encephalomalacia identified on MRI scan while a sham group of 15 patients had burr holes only. The trial met the primary endpoint showing significant improvement in motor function at 24 weeks on the Fugle-Meyer Motor Scale (FMMS) (p = .04). Interestingly, a small improvement was noted in the sham-treated group as well, indicating the presence of a placebo effect. A larger, double-blind, randomized, sham-controlled study is now planned.

■ PERSPECTIVE

The premise that stem cell biology would herald an era of regenerative medicine has fueled exaggerated claims, false starts, and a proliferation of bogus clinics. But now we may be on the threshold of a new era of stem cell-based therapies for neurologic diseases and disorders including PD, spinal cord injury, ALS, and epilepsy. Whether this promise becomes reality will depend on the outcome of the first wave of pivotal double-blind controlled trials that are now being conducted or planned.

■ FURTHER READING

Ayers JI et al: Different α-synuclein prion strains cause dementia with Lewy bodies and multiple system atrophy. Proc Natl Acad Sci USA 119:e2113489119, 2022.

Batista AF et al: The importance of complement-mediated immune signaling in Alzheimer's disease pathogenesis. Int J Mol Sci 25:817, 2024.

Carlson GA, Prusiner SB: How an infection of sheep revealed prion mechanisms in Alzheimer's disease and other neurodegenerative disorders. Int J Mol Sci 22:4861, 2021.

Condello C et al: Expanding the prion paradigm to include Parkinson and Alzheimer diseases. JAMA Neurol 81:1023, 2024.

Eichmüller OL, Knoblich JA: Human cerebral organoids: A new tool for clinical neurology research. Nat Rev Neurol 18:661, 2022.

Garton T et al: Neurodegeneration and demyelination in multiple sclerosis. Neuron 112:3231, 2024.

Kandel ER et al (eds): *Principles of Neural Science*, 6th ed. McGraw Hill, New York, 2021.

Kim TW et al: Pluripotent stem cell therapies for Parkinson disease: Present challenges and future opportunities. Front Cell Dev Biol 8:729, 2020.

Lee HG et al: Neuroinflammation: An astrocyte perspective. Sci Transl Med 15:eadi7828, 2023.

Li Q, Barres BA: Microglia and macrophages in brain homeostasis and disease. Nat Rev Immunol 18:225, 2018.

Liu L et al: Microbiota and the gut-brain-axis: Implications for new therapeutic design in the CNS. EBioMedicine 77:103908, 2022.

Pallarés-Moratalla C, Bergers G: The ins and outs of microglial cells in brain health and disease. Front Immunol 15:1305087, 2024.

Pease-Raissi SE, Chan JR: Building a (w)rapport between neurons and oligodendroglia: Reciprocal interactions underlying adaptive myelination. Neuron 109:1258, 2021.

Section 2 Diseases of the Central Nervous System

436

Seizures and Epilepsy

Patricia Dugan, Vikram R. Rao

A *seizure* (from the Latin *sacire*, "to take possession of") is a transient occurrence of signs or symptoms due to abnormal excessive or synchronous neuronal activity in the brain. Depending on the distribution of discharges, this abnormal brain activity can have various manifestations, ranging from dramatic convulsive activity to experiential phenomena not readily discernible by an observer. Although a variety of factors influence the incidence and prevalence of seizures, ~5–10% of the population will have at least one seizure, with the highest incidence occurring in early childhood and late adulthood.

The meaning of the term *seizure* needs to be carefully distinguished from that of epilepsy. *Epilepsy* describes a condition in which a person has a risk of *recurrent* seizures due to a chronic, underlying process. This definition implies that a person with a single seizure, or recurrent seizures due to correctable or avoidable circumstances, does not necessarily have epilepsy (although a single seizure associated with clinical or electroencephalographic features portending high risk of recurrence may establish the diagnosis of epilepsy). Epilepsy refers to a clinical phenomenon rather than a single disease entity because many forms and causes exist. However, among the many causes of epilepsy, there are various *epilepsy syndromes* in which the clinical and pathologic characteristics are distinctive and suggest a specific underlying etiology.

Using the definition of epilepsy as two or more unprovoked seizures, the incidence of epilepsy is ~0.3–0.5% in different populations throughout the world, and the prevalence of epilepsy has been estimated at 5–30 persons per 1000.

CLASSIFICATION OF SEIZURES

Determining the type of seizure that has occurred is essential for focusing the diagnostic approach on particular etiologies, selecting appropriate therapy, and providing information regarding prognosis. The International League Against Epilepsy (ILAE) Commission on Classification and Terminology updated their approach to classification of seizures in 2017 **(Table 436-1)**. This system is based on the clinical features of seizures and associated electroencephalographic findings. Other potentially distinctive features such as etiology or cellular substrate are not considered in this classification system, although this will undoubtedly change in the future as more is learned about the pathophysiologic mechanisms that underlie specific seizure types.

A fundamental principle is that seizures may be either focal or generalized. *Focal seizures* originate within networks limited to one brain region (note that the term *partial seizures* is no longer used). *Generalized seizures* arise within and rapidly engage networks distributed across both cerebral hemispheres. Focal seizures are often associated with structural abnormalities of the brain. In contrast, generalized seizures may result from cellular, biochemical, or structural abnormalities with a more widespread distribution. There are exceptions in both cases, however.

■ FOCAL ONSET SEIZURES

Focal seizures arise from a neuronal network either discretely localized within one brain region or more broadly distributed but still within a cerebral hemisphere. With the new classification system, the subcategories of "simple focal seizures" and "complex focal seizures" have been eliminated. Instead, the classification emphasizes the effect on awareness (intact or impaired) and nature of the onset (motor or nonmotor). Focal seizures can also evolve into generalized seizures. In the past, this was referred to as *focal seizures with secondary generalization*, but the new system relies on descriptions of the type of generalized seizures that evolve from the focal seizure.

The routine interictal (i.e., between seizures) electroencephalogram (EEG) in patients with focal seizures is often normal or may show brief discharges termed *epileptiform spikes*, or *sharp waves*. Because focal seizures can arise from the medial temporal lobe or inferior frontal lobe (i.e., regions distant from the scalp), the EEG recorded during the seizure may be nonlocalizing. However, the region of seizure onset may be defined using surgically placed intracranial electrodes.

TABLE 436-1 Classification of Seizures[a]

1. **Focal Onset**
   (Can be further described as having intact or impaired awareness, motor or nonmotor onset, or evolve from focal to bilateral tonic clonic)
2. **Generalized Onset**
   a. Motor
      Tonic-clonic
      Other motor (e.g., atonic, myoclonic)
   b. Nonmotor (absence)
3. **Unknown Onset**
   Motor, nonmotor, or unclassified

[a] Based on the 2017 International League Against Epilepsy classification of seizure types. (Data from RS Fisher et al: Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. Epilepsia 58:522, 2017.)

Focal Seizures with Intact Awareness Focal seizures can have motor manifestations (such as tonic, clonic, or myoclonic movements) or nonmotor manifestations (such as sensory, autonomic, or emotional symptoms) without impairment of awareness. For example, a patient having a focal motor seizure arising from the right primary motor cortex near the area controlling hand movement will experience involuntary movements of the contralateral left hand. Since the cortical region controlling hand movement is immediately adjacent to the region for facial expression, the seizure may also cause abnormal movements of the face synchronous with the movements of the hand. The EEG recorded with scalp electrodes during the seizure (i.e., an ictal EEG) may show abnormal focal discharges over the corresponding area of cerebral cortex.

Three additional features of focal motor seizures are worth noting. First, in some patients, the abnormal motor movements may begin in a very restricted region, such as the fingers, and gradually progress (over seconds to minutes) to include a larger portion of the extremity. This phenomenon, described by Hughlings Jackson and known as a "Jacksonian march," represents the spread of seizure activity over a progressively larger region of motor cortex. Second, patients may experience a localized paresis (Todd's paralysis) for minutes to many hours in the involved region following the seizure. Third, in rare instances, the seizure may continue for hours or days. This condition, termed *epilepsia partialis continua*, is often refractory to medical therapy.

Focal seizures may also manifest as changes in somatic sensation (e.g., paresthesias), vision (flashing lights or formed hallucinations), equilibrium (sensation of falling or vertigo), or autonomic function (flushing, sweating, piloerection). Focal seizures arising from the temporal lobe may also cause the sensation of acrid odors (e.g., bleach or burning rubber) or tastes (e.g., bitter or metallic), or hearing sounds (simple noise or complex sounds), or an epigastric sensation that rises to the head. Some patients describe intense internal feelings such as fear, a dreamlike sense, depersonalization, déjá vu, or illusions that objects are growing smaller (micropsia) or larger (macropsia). These subjective, "experiential" events that are not directly observable by someone else are referred to as *auras*.

Focal Seizures with Impaired Awareness Focal seizures may also be accompanied by a transient impairment of the patient's ability to maintain normal contact with the environment. The patient is unable to respond appropriately to visual or verbal commands during the seizure and has impaired recollection or awareness of the ictal phase. The seizures frequently begin with an aura (i.e., a focal seizure without cognitive disturbance) that is stereotypic for the patient. The start of the ictal phase is often a motionless stare, which marks the onset of the period of impaired awareness. The impaired awareness may be accompanied by *automatisms*, involuntary movements that can involve basic behaviors, such as chewing, lip smacking, swallowing, or "picking" hand movements, or more elaborate behaviors, such as a display of emotion or running. The patient is typically disoriented following the seizure, and the transition to full recovery of consciousness may range from seconds to hours. Examination immediately following the seizure may show an anterograde amnesia or transient neurologic deficits (such as aphasia, hemi-neglect, or visual loss) caused by postictal inhibition of the cortical regions most involved in the seizure.

The range of potential clinical behaviors linked to focal seizures is so broad that extreme caution is advised before concluding that paroxysmal, stereotypic episodes of bizarre or atypical behavior are not due to seizure activity. In such cases, additional detailed EEG studies may be helpful.

■ EVOLUTION OF FOCAL SEIZURES TO GENERALIZED SEIZURES

Focal seizures can spread to involve both cerebral hemispheres and produce a generalized seizure, usually of the tonic-clonic variety (discussed below). This evolution is observed frequently following focal seizures arising from a region in the frontal lobe but may also be associated with focal seizures occurring elsewhere in the brain. A focal seizure that evolves into a generalized seizure is often difficult to distinguish from a primary generalized onset tonic-clonic seizure because bystanders tend to emphasize the more dramatic, generalized convulsive phase of the seizure and overlook the more subtle, focal symptoms present at onset. In some cases, the focal onset of the seizure becomes apparent only when a careful history identifies a preceding aura. Often, however, the focal onset is not clinically evident and may be established only through careful EEG analysis. Nonetheless, distinguishing between these two entities is extremely important, because there are substantial differences in the evaluation and treatment of epilepsies characterized by focal versus generalized onset seizures.

GENERALIZED ONSET SEIZURES

Generalized seizures arise at some point in the brain but immediately and rapidly engage neuronal networks in both cerebral hemispheres. Several types of generalized seizures have features that place them in distinctive categories and facilitate clinical diagnosis.

Typical Absence Seizures Typical absence seizures are characterized by sudden, brief lapses of consciousness without loss of postural control. The seizure usually lasts for only seconds, consciousness returns as suddenly as it was lost, and there is no postictal confusion. Although the brief loss of consciousness may be clinically inapparent or the sole manifestation of the seizure, absence seizures are usually accompanied by subtle, bilateral motor signs such as rapid blinking, chewing movements, or small-amplitude, clonic movements of the hands.

Typical absence seizures are associated with a group of genetically determined epilepsies with onset usually in childhood (ages 4–10 years) or early adolescence and are the main seizure type in 15–20% of children with epilepsy. The seizures can occur hundreds of times per day, but the child may be unaware of or unable to convey their existence. Because the clinical signs of the seizures are subtle, especially to parents who may not have had previous experience with seizures, it is not surprising that the first clue to absence epilepsy is often unexplained “daydreaming” and a decline in school performance recognized by a teacher. Indeed, absence epilepsy is often misdiagnosed as an attention deficit disorder.

The electrophysiologic hallmark of typical absence seizures is a burst of generalized, symmetric, 3-Hz, spike-and-slow-wave discharges that begins and ends suddenly, superimposed on a normal EEG background. Periods of spike-and-slow-wave discharges lasting more than a few seconds usually correlate with clinical signs, but the EEG often shows many more brief bursts of abnormal cortical activity than suspected clinically. Hyperventilation tends to provoke these electrographic discharges and even the seizures themselves and is routinely used when recording the EEG.

Atypical Absence Seizures Atypical absence seizures have features that deviate both clinically and electrophysiologically from typical absence seizures. For example, the lapse of consciousness is usually longer and less abrupt in onset and cessation, and the seizure is accompanied by more obvious motor signs that may include focal or lateralizing features. The EEG shows a generalized, slow spike-and-slow-wave pattern with a frequency of ≤ 2.5 Hz, as well as other abnormal activity. Atypical absence seizures are usually associated with diffuse or multifocal structural abnormalities of the brain and therefore may accompany other signs of neurologic dysfunction, such as developmental delay. Furthermore, the seizures are less responsive to anticonvulsants compared to typical absence seizures.

Generalized, Tonic-Clonic Seizures Generalized onset tonic-clonic seizures are the main seizure type in ~10% of all persons with epilepsy. They are also the most common seizure type resulting from metabolic derangements and are therefore frequently encountered in many different clinical settings. The seizure usually begins abruptly without warning, although some patients describe vague premonitory symptoms in the hours leading up to the seizure. This prodrome is distinct from the stereotypic auras associated with focal seizures that generalize. The initial phase of the seizure is usually tonic contraction of muscles throughout the body, accounting for several classic features of the event. Tonic contraction of the muscles of expiration and the larynx at the onset will produce a loud moan or “ictal cry.” Respirations are impaired, secretions pool in the oropharynx, and cyanosis develops. Contraction of the jaw muscles may cause biting of the tongue. A marked enhancement of sympathetic tone leads to increases in heart rate, blood pressure, and pupillary size. After 10–20 s, the tonic phase of the seizure typically evolves into the clonic phase, produced by the superimposition of periods of muscle relaxation on the tonic muscle contraction. The periods of relaxation progressively increase until the end of the ictal phase, which usually lasts no more than 1 min. The postictal phase is characterized by unresponsiveness, muscular flaccidity, and excessive salivation that can cause stridorous breathing and partial airway obstruction. Bladder or bowel incontinence may occur at this point. Patients gradually regain consciousness over minutes to hours, and, during this transition, there is typically a period of postictal confusion. Patients subsequently complain of headache, fatigue, and muscle ache that can last for many hours. The duration of impaired consciousness in the postictal phase can be extremely long (i.e., many hours) in patients with prolonged seizures or underlying central nervous system (CNS) disease.

The EEG during the tonic phase of the seizure shows a progressive increase in generalized low-voltage fast activity, followed by generalized high-amplitude, polyspike discharges. In the clonic phase, the high-amplitude activity is typically interrupted by slow waves to create a spike-and-slow-wave pattern. Generalized seizures tend to terminate synchronously over widespread brain regions. The postictal EEG shows diffuse suppression of all cerebral activity, followed by slowing that gradually recovers as the patient awakens.

There are several variants of generalized motor seizures, including pure tonic and pure clonic seizures. Brief tonic seizures lasting only a few seconds are especially noteworthy since they are usually associated with specific epilepsy syndromes having mixed seizure phenotypes, such as the Lennox-Gastaut syndrome (discussed below).

Atonic Seizures Atonic seizures are characterized by sudden loss of postural muscle tone lasting 1–2 s. Consciousness is briefly impaired, but there is usually no postictal confusion. A very brief seizure may cause only a quick head drop or nodding movement, whereas a longer seizure will cause the patient to collapse (hence, the less formal term, *drop attacks*). This can be extremely dangerous because there is a substantial risk of direct head injury with the fall. The EEG shows brief, generalized spike-and-wave discharges followed immediately by diffuse slow waves that correlate with the loss of muscle tone. Like pure tonic seizures, atonic seizures are usually seen in association with known epilepsy syndromes.

Myoclonic Seizures Myoclonus is a sudden and brief muscle contraction that may involve one part of the body or the entire body. A normal, common physiologic form of myoclonus is the sudden jerking movement observed while falling asleep. Pathologic myoclonus is most often seen in association with metabolic disorders, degenerative CNS diseases, or anoxic brain injury **(Chap. 318)**. Although the distinction from other forms of myoclonus is imprecise, myoclonic seizures are true epileptic events because they are caused by cortical (vs subcortical or spinal) dysfunction. The EEG shows bilaterally synchronous spike-and-slow-wave discharges immediately prior to the movement and muscle artifact associated with the myoclonus. Myoclonic seizures usually coexist with other forms of generalized seizures but are the predominant feature of juvenile myoclonic epilepsy (JME) (discussed below).

Epileptic Spasms Epileptic spasms are characterized by a briefly sustained flexion or extension of predominantly proximal muscles, including truncal muscles. The EEG usually shows *hypsarrhythmia*, a chaotic background of diffuse, large-amplitude slow waves and irregular, multifocal spikes and sharp waves. During the clinical spasm, there is a marked suppression of the EEG background (the “electrodecremental response”). The electromyogram (EMG) also reveals a characteristic rhomboid pattern that may help distinguish spasms from brief tonic and myoclonic seizures. Epileptic spasms occur predominantly in infants and likely result from differences in neuronal function and connectivity in the immature versus mature CNS.

EPILEPSY SYNDROMES

Epilepsy syndromes are disorders in which epilepsy is a predominant feature, and there is sufficient evidence (e.g., through clinical, EEG, radiologic, or genetic observations) to suggest a common underlying mechanism. Three important epilepsy syndromes are listed below; additional examples with a known genetic basis are shown in **Table 436-2**.

JUVENILE MYOCLONIC EPILEPSY

JME is a generalized seizure disorder of unknown cause that appears in early adolescence and is usually characterized by bilateral myoclonic jerks that may be single or repetitive. The myoclonic seizures are most frequent in the morning after awakening and can be provoked by sleep deprivation. Awareness is preserved unless the myoclonus is especially severe. Many patients also experience generalized tonic-clonic seizures, and up to one-third have absence seizures. Although complete remission is uncommon, the seizures usually respond well to appropriate antiseizure medication. There is often a family history of epilepsy, and genetic studies suggest a polygenic cause.

LENNOX-GASTAUT SYNDROME

Lennox-Gastaut syndrome occurs in children and is defined by the following triad: (1) multiple seizure types (usually including generalized tonic-clonic, atonic, and atypical absence seizures); (2) an EEG with slow (<3 Hz) spike-and-wave discharges and a variety of other abnormalities; and (3) developmental delay. Lennox-Gastaut syndrome is associated with CNS disease or dysfunction from a variety of causes, including de novo mutations, developmental abnormalities, perinatal hypoxia/ischemia, trauma, infection, and other acquired lesions. The multifactorial nature of this syndrome suggests that it is a nonspecific response of the brain to diffuse neuronal dysfunction. Unfortunately, many patients have a poor prognosis due to the underlying CNS disease and the physical and psychosocial consequences of severe, poorly controlled epilepsy. Implanted neurostimulation devices (discussed below) are now being investigated for treatment of seizures in Lennox-Gastaut syndrome and other generalized epilepsies.

MESIAL TEMPORAL LOBE EPILEPSY SYNDROME

Mesial temporal lobe epilepsy (MTLE) is the most common syndrome associated with focal seizures with impairment of consciousness and is an example of an epilepsy syndrome with distinctive clinical, EEG, and pathologic features **(Table 436-3)**. High-resolution magnetic resonance imaging (MRI) can detect the characteristic hippocampal sclerosis that appears to be essential in the pathophysiology of MTLE for many patients **(Fig. 436-1)**. Recognition of this syndrome is especially important because it tends to be refractory to treatment with anticonvulsants but responds well to surgical intervention. Advances in the understanding of basic mechanisms of epilepsy have come through studies of experimental models of MTLE, discussed below.

THE CAUSES OF SEIZURES AND EPILEPSY

Seizures are a result of a shift in the normal balance of excitation and inhibition within the CNS. Given the numerous properties that control neuronal excitability, it is not surprising that there are many ways to perturb this normal balance and, therefore, many different causes of both seizures and epilepsy. Three clinical observations emphasize how a variety of factors determine why certain conditions may cause seizures or epilepsy in a given patient.

1. *The normal brain can have a seizure under the appropriate circumstances, and there are differences between individuals in the susceptibility or threshold for seizures*. For example, seizures may be induced by high fever in children who are otherwise normal and who never develop other neurologic problems, including epilepsy. However, febrile seizures occur only in a relatively small proportion of children. This implies there are various underlying *endogenous factors* that influence the threshold for having a seizure. Some of these factors are genetic, as a family history of epilepsy has a clear influence on the likelihood of seizures occurring in otherwise normal individuals. Normal development also plays an important role, because the brain appears to have different seizure thresholds at different maturational stages.
2. *There are a variety of conditions that have an extremely high likelihood of resulting in a chronic seizure disorder*. One of the best examples of this is severe, penetrating head trauma, which is associated with up to a 45% risk of subsequent epilepsy. The high propensity for severe traumatic brain injury to lead to epilepsy suggests that the injury results in a long-lasting pathologic change in the CNS that transforms a presumably normal neuronal network into one that is abnormally hyperexcitable. This process is known as *epileptogenesis*, and the specific changes that result in a lowered seizure threshold can be considered *epileptogenic factors*. Other processes associated with epileptogenesis include stroke, infection, neurodegeneration, and abnormalities of CNS development.
3. *Seizures are episodic*. Seizures occur intermittently, and, depending on the underlying cause, people with epilepsy may feel completely normal for months or years between seizures. This implies there are important provocative or *precipitating factors* that induce seizures in people with epilepsy. Similarly, precipitating factors are responsible for causing the single seizure in someone without epilepsy. Precipitants include those due to intrinsic physiologic processes, such as psychological or physical stress, sleep deprivation, or hormonal changes. They also include exogenous factors such as exposure to toxic substances, certain medications, and intermittent photic stimulation from strobe lights or some video games.

These observations emphasize the concept that the many causes of seizures and epilepsy result from a dynamic interplay between endogenous factors, epileptogenic factors, and precipitating factors. The potential role of each needs to be considered when determining the appropriate management of a patient with seizures. For example, the identification of predisposing factors (e.g., family history of epilepsy) in a patient with febrile seizures may increase the necessity for closer follow-up and a more aggressive diagnostic evaluation. Finding an epileptogenic lesion may help in the estimation of seizure recurrence and duration of therapy. Removal or modification of a precipitating factor may be an effective and safer method for preventing further seizures than the prophylactic use of anticonvulsant drugs. An emerging concept holds that underlying seizure risk itself fluctuates cyclically, potentially explaining why the same precipitating factor (e.g., sleep deprivation) can be well tolerated on some occasions but result in a seizure on others.

CAUSES ACCORDING TO AGE

In practice, it is useful to consider the etiologies of seizures based on the age of the patient, because age is one of the most important factors determining both the incidence and the likely causes of seizures or epilepsy **(Table 436-4)**. During the *neonatal period and early infancy*, potential causes include hypoxic-ischemic encephalopathy, trauma, CNS infection, congenital CNS abnormalities, and metabolic disorders. Babies born to mothers using neurotoxic drugs such as cocaine, heroin, or ethanol are susceptible to drug-withdrawal seizures in the first few days after delivery. Hypoglycemia and hypocalcemia, which can occur as secondary complications of perinatal injury, are also causes of early postnatal seizures. Seizures due to inborn errors of metabolism usually present once regular feeding begins, typically 2–3 days after birth. Pyridoxine (vitamin B_6) deficiency, an important cause of neonatal seizures, is treated with pyridoxine replacement. Idiopathic or familial forms of benign neonatal seizures are also seen during this time.

The most common seizures arising in *late infancy and early childhood* are febrile seizures, which are seizures associated with fevers but without evidence of CNS infection or other defined causes. The overall prevalence is 3–5% and even higher in some parts of the world such as Asia. Patients often have a family history of febrile seizures or epilepsy. Febrile seizures usually occur between 3 months and 5 years of age and have a peak incidence between 18 and 24 months. The typical scenario is a child who has a generalized, tonic-clonic seizure during a febrile illness in the setting of a common childhood infection such as otitis

TABLE 436-2 Examples of Genes Associated with Epilepsy Syndromes[a]

GENE (LOCUS)	FUNCTION OF GENE	CLINICAL SYNDROME	COMMENTS
CHRNA4 (20q13.2)	Nicotinic acetylcholine receptor subunit; mutations cause alterations in Ca^{2+} flux through the receptor; this may reduce the amount of GABA release in presynaptic terminals	Sleep-related hypermotor epilepsy (SHE); childhood onset; brief, nighttime seizures with prominent motor movements; often misdiagnosed as primary sleep disorder	Rare; first identified in a large Australian family; other families found to have mutations in *CHRNA2* or *CHRNB2*, and some families appear to have mutations at other loci
KCNQ2 (20q13.3)	Voltage-gated potassium channel subunits; mutation in pore regions may cause a 20–40% reduction of potassium currents, which will lead to impaired repolarization	Self-limited familial neonatal epilepsy; autosomal dominant inheritance; onset in first week of life in infants who are otherwise normal; remission usually within weeks to months; long-term epilepsy in 10–15%	Rare; other families found to have mutations in *KCNQ3*; sequence and functional homology to *KCNQ1*, mutations of which cause long QT syndrome and a cardiac-auditory syndrome
SCN1A (2q24.3)	α-Subunit of a voltage-gated sodium channel; numerous mutations affecting sodium currents that cause either gain or loss of function; network effects appear related to expression in excitatory or inhibitory cells	Very common cause of Dravet syndrome (severe myoclonic epilepsy of infancy) and some cases of Lennox-Gastaut syndrome. Also found in other syndromes, including genetic epilepsy with febrile seizures plus (GEFS+); autosomal dominant inheritance; presents with febrile seizures at median 1 year, which may persist >6 years, then variable seizure types not associated with fever	Incidence of Dravet syndrome is 1 in 20,000 births, and de novo *SCN1A* mutation is found in ~80% of cases. Incidence in GEFS+ uncertain; identified in other families with mutations in other sodium channel subunits (*SCN2B* and *SCN2A*) and $GABA_A$ receptor subunit (*GABRG2* and *GABRA1*); significant phenotypic heterogeneity within same family, including members with febrile seizures only. Avoid sodium channel–blocking antiseizure medications
LGI1 (10q24)	Leucine-rich glioma-inactivated 1 gene; previous evidence for role in glial tumor progression; recent studies suggest an influence in the postnatal development of glutamatergic circuits in the hippocampus	Autosomal dominant epilepsy with auditory features (ADEAF); a form of lateral temporal lobe epilepsy with auditory symptoms or aphasia as a major focal seizure manifestation; age of onset usually between 10 and 25 years	Mutations found in up to 50% of families containing two or more subjects with focal epilepsy with ictal auditory symptoms, suggesting that at least one other gene may underlie this syndrome
DEPDC5 (22q12.2)	Disheveled, Egl-10, and pleckstrin domain containing protein 5; exerts an inhibitory effect on mammalian target of rapamycin (mTOR)–mediated processes, such as cell growth and proliferation	Autosomal dominant familial focal epilepsy with variable foci (FFEVF); family members have seizures originating from different cortical regions; neuroimaging usually normal but may harbor subtle malformations; recent studies also suggest association with benign epilepsy with centrotemporal spikes	Study of families with the limited number of affected members revealed mutations in ~12% of families; thus, may be a relatively common cause of lesion-negative focal epilepsies with suspected genetic basis. Also associated with mutations in the GATOR1 genes *NPRL2* and *NPRL3*
GRIN2A (16p13.2)	Encodes NMDA receptor (NMDAR) subunit GluN2A	Spectrum of phenotypes ranging from benign childhood epilepsy with centrotemporal spikes (BECTS) to epilepsy-aphasia syndromes such as Landau-Kleffner syndrome (LKS) and other epileptic encephalopathies	Complex inheritance is implicated, and studies have shown considerable inter- and intrafamilial phenotypic variability and incomplete penetrance
CDKL-5 (Xp22.13)	Encodes cyclin-dependent kinase-like 5 (CDKL-5), a serine-threonine kinase involved in neural maturation and synaptogenesis	CDKL-5 deficiency disorder (CDD) results from pathogenic mutation in the *CDKL5* gene that causes absence or nonfunctional CDKL-5 protein. CDD is a severe developmental epileptic encephalopathy characterized by very-early-onset seizures. X-linked, affects females more than males	Ganaxolone is a recently approved antiseizure drug that has been shown to significantly reduce CDD-associated seizures
SLC2A1 (1p34.2)	Glucose transporter protein type 1 (GLUT1); transports glucose across the blood-brain barrier	Loss of function of one allele leads to GLUT1 deficiency, a severe metabolic encephalopathy including intractable epilepsy, complex motor dysfunction, and intellectual disability. Milder GLUT1 deficiency causes a combination of movement disorder (paroxysmal exertional dyskinesia) and epilepsy with prominent absence seizures, though intellect is often normal	Milder forms of epilepsy due to GLUT1 deficiency may respond to standard antiseizure medications, but the gold standard treatment for refractory forms is the ketogenic diet, which bypasses defective glucose transport to provide an alternative energy supply to the brain
CSTB (21q22.3)	Cystatin B, a noncaspase cysteine protease inhibitor; normal protein may block neuronal apoptosis by inhibiting caspases directly or indirectly (via cathepsins), or controlling proteolysis	Progressive myoclonus epilepsy (PME) (Unverricht-Lundborg disease); autosomal recessive inheritance; age of onset between 6 and 15 years, myoclonic seizures, ataxia, and progressive cognitive decline; brain shows neuronal degeneration	Overall rare, but relatively common in Finland and western Mediterranean (>1 in 20,000); precise role of cystatin B in human disease unknown, although mice with null mutations of cystatin B have similar syndrome
EPM2A (6q24)	Laforin, a protein tyrosine phosphatase (PTP); involved in glycogen metabolism and may have antiapoptotic activity	Progressive myoclonus epilepsy (Lafora's disease); autosomal recessive inheritance; age of onset 6–19 years, death within 10 years; brain degeneration associated with polyglucosan intracellular inclusion bodies in numerous organs	Most common PME in southern Europe, Middle East, northern Africa, and Indian subcontinent; genetic heterogeneity; unknown whether seizure phenotype due to degeneration or direct effects of abnormal laforin expression
Doublecortin (Xq21-24)	Doublecortin, expressed primarily in frontal lobes; directly regulates microtubule polymerization and bundling	Classic lissencephaly associated with severe mental retardation and seizures in males; subcortical band heterotopia with more subtle findings in females (presumably due to random X inactivation); X-linked dominant	Relatively rare but of uncertain incidence; recent increased ascertainment due to improved imaging techniques; relationship between migration defect and seizure phenotype unknown

[a] The first seven syndromes listed in the table (SHE, benign familial neonatal convulsions, GEFS+, ADEAF, FFEVF, BECTS, LKS, CDD) are examples of genetic epilepsies associated with identified gene mutations. The last three syndromes are examples of the numerous Mendelian disorders in which seizures are one part of the phenotype.

Abbreviation: GABA, γ-aminobutyric acid.

TABLE 436-3 Characteristics of the Mesial Temporal Lobe Epilepsy Syndrome

History	
History of febrile seizures	Rare generalized seizures
Family history of epilepsy	Seizures may remit and reappear
Early onset	Seizures often intractable
Clinical Observations	
Aura common	Postictal disorientation
Behavioral arrest/stare	Memory loss
Complex automatisms	Dysphasia (with focus in dominant hemisphere)
Unilateral posturing	
Laboratory Studies	
Unilateral or bilateral anterior temporal spikes on EEG	
Hypometabolism on interictal PET	
Hyperperfusion on ictal SPECT	
Material-specific memory deficits on intracarotid amobarbital (Wada) test	
MRI Findings	
Small hippocampus with increased signal on T2-weighted sequences and loss of trilaminar hippocampal internal architecture	
Small temporal lobe	
Enlarged temporal horn	
Pathologic Findings	
Highly selective loss of specific cell populations within hippocampus in most cases, granule cell layer dispersion, gliosis	

Abbreviations: EEG, electroencephalogram; MRI, magnetic resonance imaging; PET, positron emission tomography; SPECT, single-photon emission computed tomography.

TABLE 436-4 Causes of Seizures

Neonates (<1 month)	Perinatal hypoxia and ischemia Intracranial hemorrhage and trauma CNS infection Metabolic disturbances (hypoglycemia, hypocalcemia, hypomagnesemia, pyridoxine deficiency) Drug withdrawal Developmental disorders Genetic disorders
Infants and children (>1 month and <12 years)	Febrile seizures Genetic disorders (metabolic, degenerative, primary epilepsy syndromes) CNS infection Developmental disorders Trauma
Adolescents (12–18 years)	Trauma Genetic disorders Infection Illicit drug use Brain tumor
Young adults (18–35 years)	Trauma Alcohol withdrawal Illicit drug use Brain tumor Autoantibodies
Older adults (>35 years)	Cerebrovascular disease Brain tumor Alcohol withdrawal Metabolic disorders (uremia, hepatic failure, electrolyte abnormalities, hypoglycemia, hyperglycemia) Alzheimer's disease and other degenerative CNS diseases Autoantibodies

Abbreviation: CNS, central nervous system.

media, respiratory infection, or gastroenteritis. The seizure is likely to occur during the rising phase of the temperature curve (i.e., during the first day) rather than well into the course of the illness. A *simple* febrile seizure is a single, isolated event, brief, and symmetric in appearance. *Complex* febrile seizures are characterized by repeated seizure activity, a duration >15 minutes, or by focal features. Approximately one-third of patients with febrile seizures will have a recurrence, but <10% have three or more episodes. Recurrences are much more likely when the febrile seizure occurs in the first year of life. Simple febrile seizures are not associated with increased risk of developing epilepsy, while complex febrile seizures have a risk of 2–5%; other risk factors include the presence of preexisting neurologic deficits and a family history of nonfebrile seizures.

Childhood marks the age at which many of the well-defined epilepsy syndromes present. Some children who are otherwise normal develop idiopathic, generalized tonic-clonic seizures without other features that fit into specific syndromes. Temporal lobe epilepsy usually presents in childhood and may be related to mesial temporal lobe sclerosis (as part of the MTLE syndrome) or other focal abnormalities, such as cortical dysgenesis. Other types of focal seizures, including those that evolve into generalized seizures, may be late manifestations of a developmental disorder, an acquired lesion such as head trauma, CNS infection (e.g., viral encephalitis), or a CNS tumor.

The period of *adolescence and early adulthood* is one of transition during which the idiopathic or genetically based epilepsy syndromes, including JME and juvenile absence epilepsy, become less common, while epilepsies secondary to acquired CNS lesions begin to predominate. Seizures that arise in patients in this age range may be associated with head trauma, CNS infections (including parasitic infections such as cysticercosis), brain tumors, congenital CNS abnormalities, illicit drug use, or alcohol withdrawal. Autoantibodies directed against CNS antigens such as potassium channels or glutamate receptors are a cause of epilepsy that also begins to appear in this age group (although cases of autoimmunity are being increasingly described in the pediatric population), including patients without an identifiable cancer. This etiology should be suspected when a previously normal individual presents

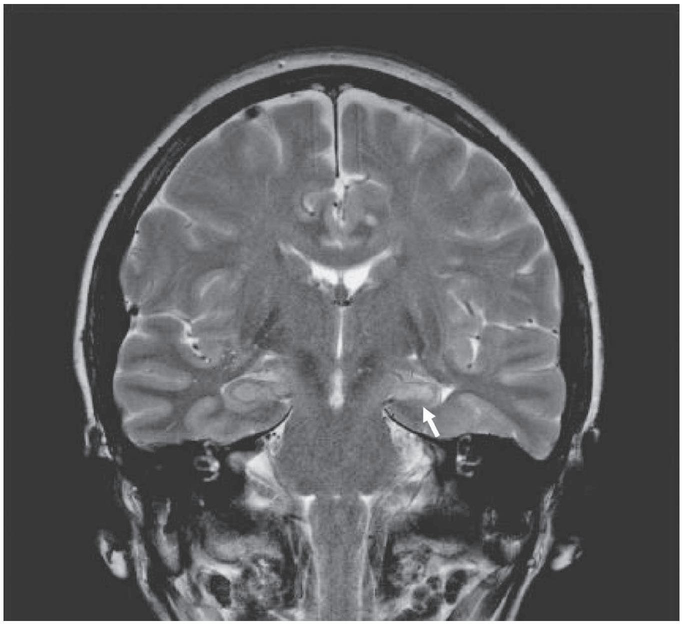

FIGURE 436-1 Mesial temporal lobe epilepsy. The electroencephalogram and seizure semiology were consistent with a left temporal lobe focus. This coronal high-resolution T2-weighted fast spin echo magnetic resonance image obtained at 3 Tesla is at the level of the hippocampal bodies and shows abnormal high signal intensity, blurring of internal laminar architecture, and reduced size of the left hippocampus (arrow) relative to the right. This triad of imaging findings is consistent with hippocampal sclerosis.

with fulminant seizures, especially when combined with psychiatric symptoms and changes in cognitive function.

Head trauma is a common cause of epilepsy in adolescents and adults. The head injury can be caused by a variety of mechanisms, and the likelihood of developing epilepsy is strongly correlated with the severity of the injury. A patient with a penetrating head wound, depressed skull fracture, intracranial hemorrhage, or prolonged posttraumatic coma or amnesia has a 30–50% risk of developing epilepsy, whereas a patient with a closed head injury and cerebral contusion has a 5–25% risk. Recurrent seizures usually develop within 1 year after head trauma, although intervals of >10 years are well known. In controlled studies, mild head injury, defined as a concussion with amnesia or loss of consciousness of <30 min, was found to be associated with only a slightly increased likelihood of epilepsy. Nonetheless, some patients experience focal seizures within hours or days of a mild head injury and subsequently develop chronic seizures of the same type; such cases may represent rare examples of epilepsy resulting from mild head injury.

The causes of seizures in *older adults* include cerebrovascular disease, trauma (including subdural hematoma), CNS tumors, and degenerative diseases. Cerebrovascular disease may account for ~50% of new cases of epilepsy in patients >65 years. Acute poststroke seizures (i.e., occurring within the first 24 h after acute stroke) are more common after hemorrhagic strokes compared to ischemic strokes. Chronic seizures typically appear months to years after the initial event and are associated with all forms of stroke.

Metabolic disturbances such as electrolyte imbalance, hypo- or hyperglycemia, renal failure, and hepatic failure may cause seizures at any age. Similarly, endocrine disorders, hematologic disorders, vasculitides, and many other systemic diseases may cause seizures over a broad age range. Many medications and abused substances can precipitate seizures as well **(Table 436-5)**.

TABLE 436-5 Drugs and Other Substances That Can Cause Seizures

- **Alkylating agents** (e.g., busulfan, chlorambucil)
- **Antimalarials** (chloroquine, mefloquine)
- **Antimicrobials/antivirals**
  - β-Lactam and related compounds
  - Quinolones
  - Acyclovir
  - Isoniazid
  - Ganciclovir
- **Anesthetics and analgesics**
  - Meperidine
  - Fentanyl
  - Tramadol
  - Local anesthetics
- **Dietary supplements**
  - Ephedra (ma huang)
  - Gingko
- **Immunomodulatory drugs**
  - Cyclosporine
  - OKT3 (monoclonal antibodies to T cells)
  - Tacrolimus
  - Interferons
- **Psychotropics**
  - Antidepressants (e.g., bupropion)
  - Antipsychotics (e.g., clozapine)
  - Lithium
- **Radiographic contrast agents**
- **Drug withdrawal**
  - Alcohol
  - Baclofen
  - Barbiturates (short-acting)
  - Benzodiazepines (short-acting)
  - Zolpidem
- **Drugs of abuse**
  - Amphetamine
  - Cocaine
  - Phencyclidine
  - Methylphenidate
- **Flumazenil**[a]

[a]In benzodiazepine-dependent patients.

BASIC MECHANISMS

MECHANISMS OF SEIZURE INITIATION AND PROPAGATION

Focal seizure activity can begin in a discrete region of cortex and then slowly invade the surrounding regions. The hallmark of an established seizure is typically an electrographic "spike" due to intense near-simultaneous firing of many local excitatory neurons, resulting in an apparent hypersynchronization of the excitatory bursts across a relatively large cortical region. The bursting activity in individual neurons (the "paroxysmal depolarization shift") is caused by a relatively long-lasting depolarization of the neuronal membrane due to influx of extracellular calcium (Ca^{2+}), which leads to the opening of voltage-dependent sodium (Na^+) channels, influx of Na^+, and generation of repetitive action potentials. This is followed by a hyperpolarizing afterpotential mediated by γ-aminobutyric acid (GABA) receptors or potassium (K^+) channels, depending on the cell type. The synchronized bursts from enough neurons result in summation of field potentials producing a spike discharge on the EEG.

The spreading seizure wavefront is thought to slow and ultimately halt by intact hyperpolarization and a "surround" inhibition created by feedforward activation of inhibitory neurons. With sufficient activation, there is a recruitment of surrounding neurons via a number of synaptic and nonsynaptic mechanisms, including (1) an increase in extracellular K^+, which blunts hyperpolarization and depolarizes neighboring neurons; (2) accumulation of Ca^{2+} in presynaptic terminals, leading to enhanced neurotransmitter release; (3) depolarization-induced activation of the *N*-methyl-D-aspartate (NMDA) subtype of the excitatory amino acid receptor, which causes additional Ca^{2+} influx and neuronal activation; and (4) ephaptic interactions related to changes in tissue osmolarity and cell swelling. The recruitment of a sufficient number of neurons leads to the propagation of excitatory currents into contiguous areas via local cortical connections and to more distant areas via long commissural pathways such as the corpus callosum.

Many factors control neuronal excitability, and thus, there are many potential mechanisms for altering a neuron's propensity to have bursting activity. Mechanisms *intrinsic* to the neuron include changes in the conductance of ion channels, response characteristics of membrane receptors, cytoplasmic buffering, second-messenger systems, and protein expression as determined by gene transcription, translation, and posttranslational modification. Mechanisms *extrinsic* to the neuron include changes in the amount or type of neurotransmitters present at the synapse, modulation of receptors by extracellular ions and other molecules, and temporal and spatial properties of synaptic and nonsynaptic input. Glial cells, such as astrocytes and oligodendrocytes, have an important role in many of these mechanisms as well.

Certain recognized causes of seizures are explained by these mechanisms. For example, accidental ingestion of domoic acid, an analogue of glutamate (the principal excitatory neurotransmitter in the brain) produced by naturally occurring microscopic algae, causes profound seizures via direct activation of excitatory amino acid receptors throughout the CNS. Penicillin, which can lower the seizure threshold in humans and is a potent convulsant in experimental models, reduces inhibition by antagonizing the effects of GABA at its receptor. The basic mechanisms of other precipitating factors of seizures, such as sleep deprivation, fever, alcohol withdrawal, hypoxia, and infection, are not as well understood but presumably involve analogous perturbations in neuronal excitability. Similarly, the endogenous factors that determine an individual's seizure threshold may relate to these properties as well.

Knowledge of the mechanisms responsible for initiation and propagation of most generalized seizures (including tonic-clonic, myoclonic, and atonic types) remains rudimentary and reflects the limited understanding of the connectivity of the brain at a systems level. Much more is understood about the origin of generalized spike-and-wave discharges in absence seizures. These appear to be related to oscillatory rhythms normally generated during sleep by circuits connecting the thalamus and cortex. This oscillatory behavior involves an interaction between $GABA_B$ receptors, T-type Ca^{2+} channels, and K^+ channels located within the thalamus. Pharmacologic studies indicate that modulation of these receptors and channels can induce absence seizures, and there is good evidence that the genetic forms of absence epilepsy may be associated with mutations of components of this system.

MECHANISMS OF EPILEPTOGENESIS

Epileptogenesis refers to the transformation of a normal neuronal network into one that is chronically hyperexcitable. There is often a delay of months to years between an initial CNS injury such as trauma, stroke, or infection and the first clinically evident seizure. The injury appears to initiate a process that gradually lowers the seizure threshold in the affected region until a spontaneous seizure occurs. In many genetic and idiopathic forms of epilepsy, epileptogenesis is presumably determined by developmentally regulated events.

Pathologic studies of the hippocampus from patients with temporal lobe epilepsy suggest that some forms of epileptogenesis are related to *structural changes in neuronal networks*. For example, many patients with MTLE have a highly selective loss of neurons that normally contribute to inhibition of the main excitatory neurons within the dentate gyrus. There is also evidence that, in response to the loss of neurons, there is reorganization of surviving neurons in a way that affects the excitability of the network. Some of these changes can be seen in experimental models of prolonged electrical seizures or traumatic brain injury. Thus, an initial injury such as head injury may lead to a focal region of structural change that causes local hyperexcitability. The local hyperexcitability leads to further structural changes that evolve over time until the focal lesion produces clinically evident seizures. Similar models have provided strong evidence for long-term alterations in *intrinsic, biochemical properties of cells* within the network such as chronic changes in glutamate or GABA receptor function. Induction of inflammatory cascades may be a critical factor in these processes as well.

GENETIC CAUSES OF EPILEPSY

An area of ongoing progress in epilepsy research has been the identification of genetic mutations associated with a variety of epilepsy syndromes (Table 436-2). Although most of the mutations identified to date cause rare forms of epilepsy, their discovery has led to extremely important conceptual advances. For example, it appears that many of the inherited epilepsies are due to mutations affecting ion channel function. These syndromes are therefore part of the larger group of channelopathies causing paroxysmal disorders such as cardiac arrhythmias, episodic ataxia, periodic weakness, and familial hemiplegic migraine. Other gene mutations are proving to be associated with pathways influencing CNS development, synaptic physiology, or neuronal homeostasis. De novo and somatic mutations may explain a significant proportion of these syndromes, especially those with onset in early childhood. A current challenge is to identify the multiple susceptibility genes that underlie the more common forms of idiopathic epilepsies. Ion channel mutations and copy number variants may be pathogenic in a subset of these patients.

MECHANISMS OF ACTION OF ANTISEIZURE DRUGS

Antiseizure drugs appear to act primarily by blocking the initiation or spread of seizures. This occurs through a variety of mechanisms that modify the activity of ion channels or neurotransmitters, and in most cases, the drugs have pleiotropic effects. The mechanisms include inhibition of Na^+-dependent action potentials in a frequency-dependent manner (e.g., phenytoin, carbamazepine, lamotrigine, topiramate, zonisamide, lacosamide, rufinamide, cenobamate), inhibition of voltage-gated Ca^{2+} channels (phenytoin, gabapentin, pregabalin), facilitating the opening of potassium channels (ezogabine), attenuation of glutamate activity (lamotrigine, topiramate, felbamate, perampanel), potentiation of GABA receptor function (benzodiazepines, barbiturates), increase in the availability of GABA (valproic acid, gabapentin, tiagabine), and modulation of release of synaptic vesicles (levetiracetam, brivaracetam). Two of the effective drugs for absence seizures, ethosuximide and valproic acid, probably act by inhibiting T-type Ca^{2+} channels in thalamic neurons. Cannabidiol (CBD), a derivative of cannabis plants, is effective for reducing seizures in children with Dravet syndrome, Lennox-Gastaut syndrome, and tuberous sclerosis but does not act through endogenous cannabinoid receptors. Rather, CBD has a multimodal mechanism of action involving modulation of intracellular calcium via G protein–coupled receptor 55, extracellular calcium influx via transient receptor potential vanilloid type 1 (TRPV1) channels, and adenosine-mediated signaling. Fenfluramine is an amphetamine derivative that probably exerts its antiseizure effect by inhibiting serotonin reuptake and increasing extracellular levels. Ganaxolone is a synthetic analogue of allopregnanolone and an allosteric modulator of $GABA_A$ receptors that reduces seizures associated with cyclin-dependent kinase-like 5 (CDKL-5) deficiency disorder.

In contrast to the relatively large number of antiseizure drugs that can attenuate seizure activity, there are currently no drugs known to prevent the formation of a seizure focus following CNS injury. Eventual development of such "antiepileptogenic" drugs will provide means of preventing the emergence of epilepsy following injuries such as head trauma, stroke, and CNS infection.

APPROACH TO THE PATIENT

Seizure

When a patient presents shortly after a seizure, the first priorities are attention to vital signs, respiratory and cardiovascular support, and treatment of seizures if they resume (see "Treatment: Seizures and Epilepsy"). Life-threatening conditions such as CNS infection, metabolic derangement, or drug toxicity must be recognized and managed appropriately.

When the patient is not acutely ill, the evaluation will initially focus on whether there is a history of earlier seizures **(Fig. 436-2)**. If this is the first seizure, then the emphasis will be to (1) establish whether the reported episode was a seizure rather than another paroxysmal event, (2) determine the cause of the seizure by identifying risk factors and precipitating events, and (3) decide whether antiseizure drug therapy is required in addition to treatment for any underlying illness.

In the patient with prior seizures or a known history of epilepsy, the evaluation is directed toward (1) identification of the underlying cause and precipitating factors, and (2) determination of the adequacy of the patient's current therapy.

HISTORY AND EXAMINATION

The first goal is to determine whether the event was truly a seizure. A detailed history is essential, because *in many cases the diagnosis of a seizure is based solely on clinical grounds—the examination and laboratory studies are often normal*. Questions should focus on the symptoms before, during, and after the episode to differentiate a seizure from other paroxysmal events (see "Differential Diagnosis of Seizures" below). Seizures frequently occur out-of-hospital, and the patient may be unaware of the ictal and immediate postictal phases; thus, witnesses to the event should be interviewed carefully.

The history should also focus on risk factors and predisposing events. Clues for a predisposition to seizures include a history of febrile seizures, a family history of seizures, and, of particular importance, earlier auras or brief seizures not recognized as such. Epileptogenic factors such as prior head trauma, stroke, tumor, or CNS infection should be identified. In children, a careful assessment of developmental

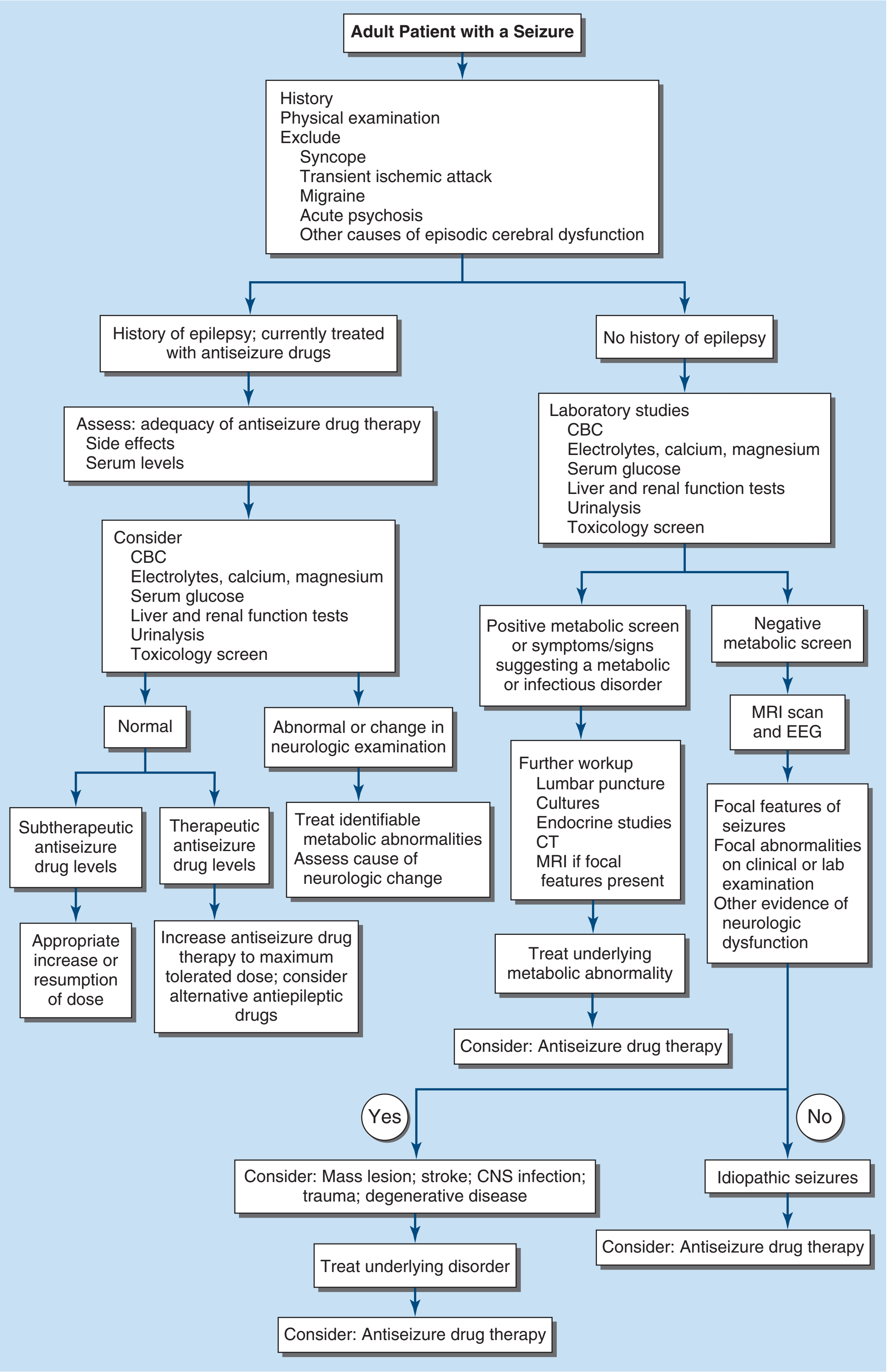
Adult Patient with a Seizure
History
Physical examination
Exclude
Syncope
Transient ischemic attack
Migraine
Acute psychosis
Other causes of episodic cerebral dysfunction
History of epilepsy; currently treated with antiseizure drugs
No history of epilepsy
Assess: adequacy of antiseizure drug therapy
Side effects
Serum levels
Laboratory studies
CBC
Electrolytes, calcium, magnesium
Serum glucose
Liver and renal function tests
Urinalysis
Toxicology screen
Consider
CBC
Electrolytes, calcium, magnesium
Serum glucose
Liver and renal function tests
Urinalysis
Toxicology screen
Positive metabolic screen or symptoms/signs suggesting a metabolic or infectious disorder
Negative metabolic screen
Normal
Abnormal or change in neurologic examination
MRI scan and EEG
Further workup
Lumbar puncture
Cultures
Endocrine studies
CT
MRI if focal features present
Subtherapeutic antiseizure drug levels
Therapeutic antiseizure drug levels
Treat identifiable metabolic abnormalities
Assess cause of neurologic change
Focal features of seizures
Focal abnormalities on clinical or lab examination
Other evidence of neurologic dysfunction
Appropriate increase or resumption of dose
Increase antiseizure drug therapy to maximum tolerated dose; consider alternative antiepileptic drugs
Treat underlying metabolic abnormality
Consider: Antiseizure drug therapy
Yes
No
Consider: Mass lesion; stroke; CNS infection; trauma; degenerative disease
Idiopathic seizures
Treat underlying disorder
Consider: Antiseizure drug therapy
Consider: Antiseizure drug therapy

FIGURE 436-2 Evaluation of the adult patient with a seizure. CBC, complete blood count; CNS, central nervous system; CT, computed tomography; EEG, electroencephalogram; MRI, magnetic resonance imaging.

milestones may provide evidence for underlying CNS disease. Precipitating factors such as sleep deprivation, systemic diseases, electrolyte or metabolic derangements, acute infection, drugs that lower the seizure threshold (Table 436-5), or alcohol or illicit drug use should also be identified.

The general physical examination includes a search for signs of infection or systemic illness. Careful examination of the skin may reveal signs of neurocutaneous disorders, such as tuberous sclerosis or neurofibromatosis, or chronic liver or renal disease. A finding of organomegaly may indicate a metabolic storage disease, and limb asymmetry may provide a clue to brain injury early in development. Signs of head trauma and use of alcohol or illicit drugs should be sought. Auscultation of the heart and carotid arteries may identify an abnormality that predisposes to cerebrovascular disease.

All patients require a complete neurologic examination, with particular emphasis on eliciting signs of cerebral hemispheric disease **(Chap. 433)**. Careful assessment of mental status (including memory, language function, and abstract thinking) may suggest lesions in the anterior frontal, parietal, or temporal lobes. Testing of visual fields will help screen for lesions in the optic pathways and occipital lobes. Screening tests of motor function such as pronator drift, deep tendon reflexes, gait, and coordination may suggest lesions in motor (frontal) cortex, and cortical sensory testing (e.g., double simultaneous stimulation) may detect lesions in the parietal cortex.

LABORATORY STUDIES

Routine blood studies are indicated to identify the more common metabolic causes of seizures such as abnormalities in electrolytes, glucose, calcium, or magnesium, and hepatic or renal disease. A screen for toxins in blood and urine should also be obtained from all patients in appropriate risk groups, especially when no clear precipitating factor has been identified. A lumbar puncture is indicated if there is any suspicion of meningitis or encephalitis, and it is mandatory in all patients infected with HIV, even in the absence of symptoms or signs suggesting infection. Testing for autoantibodies in the serum and cerebrospinal fluid (CSF) should be considered in patients presenting with fulminant onset of epilepsy associated with other abnormalities such as psychiatric symptoms or cognitive disturbances **(Chap. 99)**.

ELECTROPHYSIOLOGIC STUDIES

The electrical activity of the brain (the EEG) is easily recorded from electrodes placed on the scalp. The potential difference between pairs of electrodes on the scalp (bipolar derivation) or between individual scalp electrodes and a relatively inactive common reference point (referential derivation) is amplified and displayed on a computer monitor. Digital systems allow the EEG to be displayed with any desired format and to be manipulated for more detailed analysis, and they also enable computational algorithms that can automatically detect certain abnormalities. The characteristics of the normal EEG depend on the patient's age and level of arousal. The rhythmic activity normally recorded represents the postsynaptic potentials of vertically oriented pyramidal cells of the cerebral cortex and is characterized by its frequency. In normal awake adults lying quietly with the eyes closed, an 8- to 13-Hz alpha rhythm is seen posteriorly in the EEG, intermixed with a variable amount of generalized faster (beta) activity (>13 Hz); the alpha rhythm is attenuated when the eyes are opened **(Fig. 436-3)**. During drowsiness and light sleep, slower activity in the theta (4–7 Hz) and delta (<4 Hz) ranges becomes more conspicuous.

All patients who have a possible seizure disorder should be evaluated with an EEG as soon as possible. In the evaluation of a patient with suspected epilepsy, the presence of *electrographic seizure activity* during the clinical event (i.e., abnormal, repetitive, rhythmic activity having a discrete onset and termination) clearly establishes the diagnosis. The EEG is always abnormal during generalized tonic-clonic seizures. The absence of electrographic seizure activity does not exclude a seizure disorder, however, because focal seizures may originate from a region of the cortex that cannot be detected by standard scalp electrodes. Because seizures are typically infrequent and unpredictable, it is often not possible to obtain the EEG during a clinical event. In such situations, activating procedures are generally undertaken to provoke abnormalities while the EEG is recorded. These procedures commonly include hyperventilation, photic stimulation, sleep, and sleep deprivation on the night prior to the recording. Continuous monitoring for prolonged periods in video-EEG telemetry units for hospitalized patients or the use of portable equipment to record the EEG continuously for ≥24 h in ambulatory patients has made it easier to capture the electrophysiologic correlates of clinical events. Video-EEG telemetry is now a routine approach for the accurate diagnosis of epilepsy in patients with poorly characterized events or seizures that are difficult to control. A variety of technologies, including sub-scalp EEG devices, are being developed to enable ultra-long-term ambulatory recordings of brain activity, analogous to implanted loop recorders used to capture infrequent cardiac arrhythmias.

The EEG may also be helpful in the interictal period by showing certain abnormalities that are highly supportive of the diagnosis of epilepsy. Such *epileptiform activity* consists of bursts of abnormal discharges containing spikes or sharp waves. The presence of epileptiform activity is not entirely specific for epilepsy, but it has a much greater prevalence in patients with epilepsy than in other individuals. However, even in an individual who is known to have epilepsy, the initial routine interictal EEG may be normal 50–80% of the time. Thus, the EEG has limited sensitivity and cannot establish the diagnosis of epilepsy in many cases.

The EEG is also used for classifying seizure disorders and aiding in the selection of antiseizure medications **(Fig. 436-4)**. For example, episodic generalized spike-wave activity is usually seen in patients with typical absence epilepsy and may be seen with other generalized epilepsy syndromes. Focal interictal epileptiform discharges would support the diagnosis of a focal seizure disorder such as temporal lobe epilepsy or frontal lobe seizures, depending on the location of the discharges.

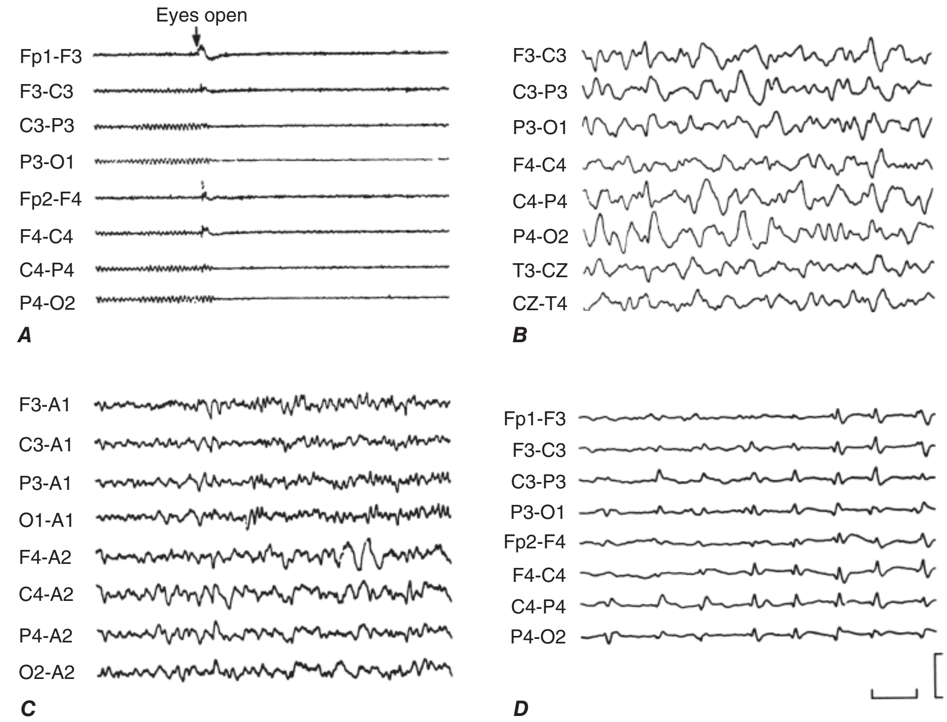

FIGURE 436-3 Electroencephalograms. ***A.*** Normal electroencephalogram (EEG) showing a posteriorly situated 9-Hz alpha rhythm that attenuates with eye opening. ***B.*** Abnormal EEG showing irregular diffuse slow activity in an obtunded patient with encephalitis. ***C.*** Irregular slow activity in the right central region, on a diffusely slowed background, in a patient with a right parietal glioma. ***D.*** Periodic complexes occurring once every second in a patient with Creutzfeldt-Jakob disease. Horizontal calibration: 1 s; vertical calibration: 200 μV in A, 300 μV in other panels. In this and the following figure, electrode placements are indicated at the left of each panel and accord with the international 10–20 system. A, earlobe; C, central; F, frontal; Fp, frontal polar; O, occipital; P, parietal; T, temporal. Right-sided placements are indicated by even numbers, left-sided placements by odd numbers, and midline placements by Z. *(Reproduced with permission from MJ Aminoff: Aminoff's Electrodiagnosis in Clinical Neurology, 6th ed. Oxford: Elsevier Saunders, 2012.)*

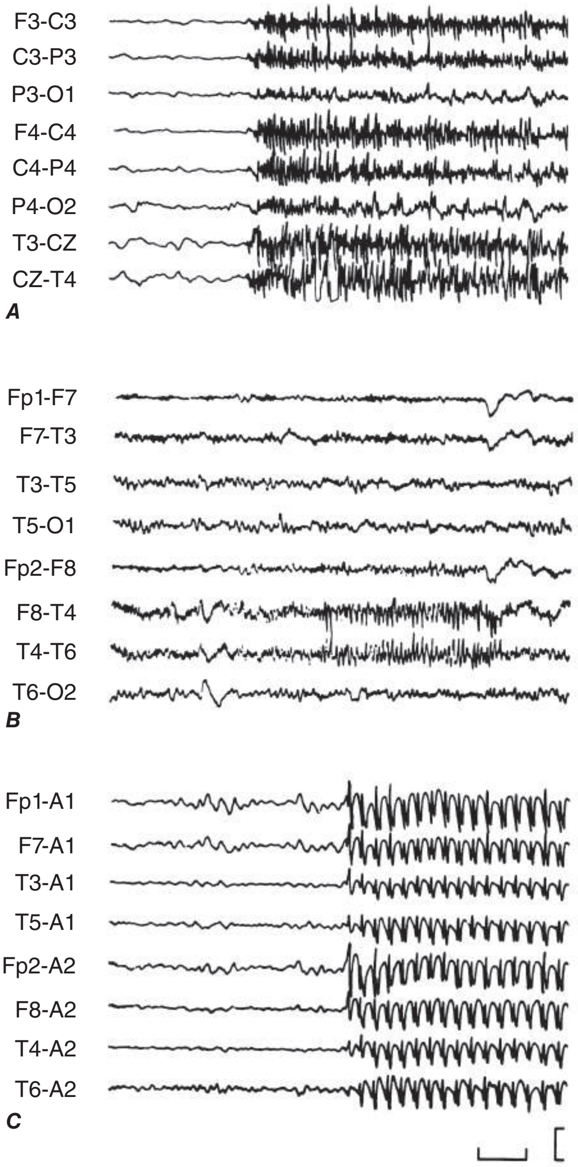

FIGURE 436-4 Electrographic seizures. ***A.*** Onset of a tonic seizure showing generalized repetitive sharp activity with synchronous onset over both hemispheres. ***B.*** Burst of repetitive spikes occurring with sudden onset in the right temporal region during a clinical spell characterized by transient impairment of awareness. ***C.*** Generalized 3-Hz spike-wave activity occurring synchronously over both hemispheres during an absence seizure. Horizontal calibration: 1 s; vertical calibration: 400 μV in *A*, 200 μV in *B*, and 750 μV in *C*. *(Reproduced with permission from MJ Aminoff: Aminoff's Electrodiagnosis in Clinical Neurology, 6th ed. Oxford: Elsevier Saunders, 2012.)*

The routine scalp EEG may also be used to assess the prognosis of seizure disorders; in general, a normal EEG implies a better prognosis, whereas an abnormal background or frequent epileptiform activity suggests a worse outcome. Unfortunately, the EEG has not proved to be useful in predicting which patients with predisposing conditions such as head injury or brain tumor will go on to develop epilepsy, because in such circumstances, epileptiform activity is commonly encountered regardless of whether seizures occur.

High-density EEG provides higher resolution localization by utilizing up to 257 scalp electrodes, as opposed to the 21 electrodes used in a standard EEG recording. This makes use of scalp voltage field data to determine dipole location, orientation, and propagation. EEG source imaging (ESI) permits visualization of these dipoles by mapping them onto a three-dimensional model that is co-registered on brain MRI.

Magnetoencephalography (MEG) provides another way of looking noninvasively at cortical activity. Instead of measuring electrical activity of the brain, it measures the small magnetic fields that are generated by this activity. The epileptiform activity seen on MEG can be analyzed, and its source in the brain can be estimated using a variety of mathematical techniques. These source estimates can then be plotted on an anatomic image of the brain such as an MRI (discussed below) to generate a magnetic source image (MSI). MSI can be useful to localize potential seizure foci.

■ BRAIN IMAGING

Almost all patients with new-onset seizures should have a brain imaging study to determine whether there is an underlying structural abnormality that is responsible. The only potential exception to this rule is children who have an unambiguous history and examination suggestive of a benign, generalized seizure disorder such as absence epilepsy. MRI has been shown to be superior to computed tomography (CT) for the detection of cerebral lesions associated with epilepsy. In some cases, MRI will identify lesions such as tumors, vascular malformations, or other pathologies that need urgent therapy. The availability of newer MRI methods, such as three-dimensional structural imaging at submillimeter resolution, has increased the sensitivity for detection of abnormalities of cortical architecture, including hippocampal atrophy associated with mesial temporal sclerosis, as well as abnormalities of neuronal migration. In such cases, the findings provide an explanation for the patient's seizures and point to the need for chronic antiseizure drug therapy or possible surgical resection.

In the patient with a suspected CNS infection or mass lesion, CT scanning should be performed emergently when MRI is not immediately available. Otherwise, it is usually appropriate to obtain an MRI study within a few days of the initial evaluation. Functional imaging procedures such as positron emission tomography (PET) and single-photon emission computed tomography (SPECT) are also used to evaluate certain patients with medically refractory seizures (Table 436-3).

■ GENETIC TESTING

With the increasing recognition of specific gene mutations causing epilepsy (Table 436-2), genetic testing is beginning to emerge as part of the diagnostic evaluation of patients with epilepsy. In addition to providing a definitive diagnosis (which may be of great benefit to the patient and family members and curtail the pursuit of additional, unrevealing laboratory testing), genetic testing may offer a guide for therapeutic options (see section "Selection of Antiseizure Drugs" below). Genetic testing is currently being done mainly in infants and children with epilepsy syndromes thought to have a genetic cause but should also be considered in older patients with a history suggesting an undiagnosed genetic epilepsy syndrome that began early in life.

DIFFERENTIAL DIAGNOSIS OF SEIZURES

Disorders that may mimic seizures are listed in **Table 436-6**. In most cases, seizures can be distinguished from other conditions by meticulous attention to the history and relevant laboratory studies. On occasion, additional studies such as video-EEG monitoring, sleep studies, tilt-table analysis, or cardiac electrophysiology may be required to reach a correct diagnosis. Two of the more common nonepileptic syndromes in the differential diagnosis are discussed below.

■ SYNCOPE

The diagnostic dilemma encountered most frequently is the distinction between a generalized seizure and syncope. Observations by the patient and bystanders that can help differentiate between the two are listed in **Table 436-7**. Characteristics of a seizure include the presence of an aura, cyanosis, unconsciousness, motor manifestations lasting >15 s, postictal disorientation, muscle soreness, and sleepiness. In contrast, a syncopal episode is more likely if the event was provoked by acute pain or emotional stress or occurred immediately after arising from the lying or sitting position. Patients with syncope often describe a stereotyped transition from consciousness to unconsciousness that includes tiredness, sweating, nausea, and tunneling of vision, and they experience a relatively brief loss of consciousness with rapid recovery of normal mentation. Headache and incontinence are unreliable

TABLE 436-6 Differential Diagnosis of Seizures

Syncope
Vasovagal syncope
Cardiac arrhythmia
Valvular heart disease
Cardiac failure
Orthostatic hypotension
Psychological disorders
Psychogenic seizure
Hyperventilation
Panic attack
Metabolic disturbances
Alcoholic blackouts
Delirium tremens
Hypoglycemia
Hypoxia
Psychoactive drugs (e.g., hallucinogens)
Migraine
Confusional migraine
Basilar migraine
Transient ischemic attack (TIA)
Basilar artery TIA
Sleep disorders
Narcolepsy/cataplexy
Benign sleep myoclonus
Movement disorders
Tics
Nonepileptic myoclonus
Paroxysmal choreoathetosis
Special considerations in children
Breath-holding spells
Migraine with recurrent abdominal pain and cyclic vomiting
Benign paroxysmal vertigo
Apnea
Night terrors
Sleepwalking

features in differentiating between syncope and seizure. A brief period (i.e., 1–10 s) of convulsive motor activity is frequently seen immediately at the onset of a syncopal episode, especially if the patient remains in an upright posture after fainting (e.g., in a dentist's chair) and therefore has a sustained decrease in cerebral perfusion. Rarely, a syncopal episode can induce a full tonic-clonic seizure. In such cases, the evaluation must focus on both the cause of the syncopal event as well as the possibility that the patient has a propensity for recurrent seizures. Postictal symptoms can be helpful when differentiating convulsive syncope from seizure, as confusion and disorientation are typically much less prominent following syncope.

PSYCHOGENIC SEIZURES

Psychogenic seizures are nonepileptic behaviors that resemble seizures. They are often part of a conversion reaction precipitated by underlying psychological distress. Certain behaviors such as side-to-side turning of the head, ictal eye closure, asynchronous and large-amplitude shaking movements of the limbs, twitching of all four extremities without loss of consciousness, and pelvic thrusting are more commonly associated with psychogenic rather than epileptic seizures. Psychogenic seizures often last longer than epileptic seizures and may wax and wane over minutes to hours. However, the distinction is sometimes difficult on clinical grounds alone, and there are many examples of diagnostic errors made by experienced epileptologists. This is especially true for psychogenic seizures that resemble focal seizures, because the behavioral manifestations of focal seizures (especially of frontal lobe origin) can be extremely unusual, and, in both cases, the routine surface EEG may be normal. Video-EEG monitoring is very useful when historic features are nondiagnostic. Generalized tonic-clonic seizures always produce marked EEG abnormalities during and after the seizure. For suspected focal seizures, the use of additional electrodes may help to localize a seizure focus. Most generalized seizures and some focal seizures are accompanied by a rise in serum prolactin during the immediate 30-min postictal period, whereas psychogenic seizures are not, though this is not always reliable because baseline prolactin levels are rarely available and certain medications can elevate prolactin levels. The diagnosis of psychogenic seizures also does not exclude a concurrent diagnosis of epilepsy, because the two often coexist.

TABLE 436-7 Features That Distinguish Generalized Tonic-Clonic Seizure from Syncope

FEATURES	SEIZURE	SYNCOPE
Immediate precipitating factors	Usually none	Emotional stress, Valsalva, orthostatic hypotension, cardiac etiologies
Premonitory symptoms	None or aura (e.g., odd odor)	Tiredness, nausea, diaphoresis, tunneling of vision
Posture at onset	Variable	Usually erect
Transition to unconsciousness	Often immediate	Gradual over seconds[a]
Duration of unconsciousness	Minutes	Seconds
Duration of tonic or clonic movements	30–60 s	Never >15 s
Facial appearance during event	Cyanosis, frothing at mouth	Pallor
Disorientation and sleepiness after event	Many minutes to hours	<5 min
Aching of muscles after event	Often	Sometimes
Biting of tongue	Sometimes	Rarely
Incontinence	Sometimes	Sometimes
Headache	Sometimes	Rarely

[a] May be sudden with certain cardiac arrhythmias.

TREATMENT

Seizures and Epilepsy

Therapy for a patient with a seizure disorder is almost always multimodal and includes treatment of underlying conditions that cause or contribute to the seizures, avoidance of precipitating factors, suppression of recurrent seizures by prophylactic therapy with antiseizure medications or surgery, and addressing a variety of psychological and social issues. Treatment plans must be individualized, given the many different types and causes of seizures as well as the differences in efficacy and toxicity of antiseizure medications for each patient. In almost all cases, a neurologist with experience in the treatment of epilepsy should design and oversee implementation of the treatment strategy. Furthermore, patients with refractory epilepsy or those who require polypharmacy with antiseizure drugs should remain under the regular care of a neurologist.

TREATMENT OF UNDERLYING CONDITIONS

If the sole cause of a seizure is a metabolic disturbance such as an abnormality of serum electrolytes or glucose, then treatment is aimed at reversing the metabolic problem and preventing its recurrence. Therapy with antiseizure drugs is usually unnecessary unless the metabolic disorder cannot be corrected promptly and the patient is at risk of having further seizures. If the apparent cause of a seizure was a medication (e.g., bupropion) or illicit drug use (e.g., cocaine), then appropriate therapy is avoidance of the drug; there is usually no need for antiseizure medications unless subsequent seizures occur in the absence of these precipitants.

Seizures caused by a structural CNS lesion such as a brain tumor, vascular malformation, or brain abscess may not recur after appropriate treatment of the underlying lesion. However, despite removal of the structural lesion, there is a risk that the seizure focus will remain in the surrounding tissue or develop de novo as a result of gliosis and other processes induced by surgery, radiation, or other therapies. Most patients are therefore maintained on an antiseizure medication for 6–12 months, and an attempt is made to withdraw medications only if the patient has been completely seizure free. If seizures are refractory to medication, the patient may benefit from surgical removal of the seizure-producing brain tissue (see below).

AVOIDANCE OF PRECIPITATING FACTORS

Unfortunately, little is known about the specific factors that determine precisely when a seizure will occur in a patient with epilepsy. An almost universal precipitating factor for seizures is sleep deprivation, so patients should do everything possible to optimize their sleep quality. Many patients can identify other avoidable situations that appear to lower their seizure threshold. For example, patients may note an association between alcohol intake and seizures, and they should be encouraged to modify their drinking habits accordingly. There are also relatively rare cases of patients with seizures that are induced by highly specific stimuli such as a video game monitor, music, startling sounds, or an individual's voice ("reflex epilepsy"). Because there is often an association between stress and seizures, stress reduction techniques such as physical exercise, meditation, or counseling may be helpful.

ANTISEIZURE DRUG THERAPY

Antiseizure drug therapy is the mainstay of treatment for most people with epilepsy. The overall goal is to completely prevent seizures without causing any untoward side effects, preferably with a single medication and a dosing schedule that is easy for the patient to follow. Seizure classification is an important element in designing the treatment plan, because some antiseizure drugs have different activities against various seizure types. However, there is considerable overlap between antiseizure drugs, and choice of therapy is often determined by anticipated side effects, drug-drug interactions, medical comorbidities, dosing frequency, and cost.

When to Initiate Antiseizure Drug Therapy Antiseizure drug therapy should be started in any patient with recurrent seizures of unknown etiology or a known cause that cannot be reversed. Whether to initiate therapy in a patient with a single seizure is controversial. Patients with a single seizure due to an identified lesion such as a CNS tumor, infection, or trauma, in which there is strong evidence that the lesion is epileptogenic, should be treated. The risk of seizure recurrence in a patient with an apparently unprovoked seizure is uncertain, with estimates ranging from 31 to 71% in the first 12 months after the initial seizure. This uncertainty arises from differences in the underlying seizure types and etiologies in various published epidemiologic studies. Generally accepted risk factors associated with recurrent seizures include the following: (1) prior brain insult such as a stroke or trauma, (2) an EEG with epileptiform abnormalities, (3) a significant brain imaging abnormality, or (4) a nocturnal seizure. Most patients with one or more of these risk factors should be treated. Issues such as employment or driving may influence the decision regarding whether to start medications as well. For example, a patient with a single unprovoked seizure whose job depends on driving may prefer taking an antiseizure drug to reduce risk of seizure recurrence and the potential loss of driving privileges.

Selection of Antiseizure Drugs Antiseizure drugs available in the United States are shown in **Table 436-8**, and the main pharmacologic characteristics of commonly used drugs are listed in **Table 436-9**. Worldwide, older medications such as phenytoin, valproic acid, carbamazepine, phenobarbital, and ethosuximide are generally used as first-line therapy for most seizure disorders because, overall, they are as effective as more recent drugs and significantly less expensive overall. Most of the new drugs that have become available in the past decade are used as adjunctive therapy, although many are now also being used as first-line monotherapy.

TABLE 436-8 Selection of Antiseizure Drugs

GENERALIZED-ONSET TONIC-CLONIC	FOCAL	TYPICAL ABSENCE	ATYPICAL ABSENCE, MYOCLONIC, ATONIC
First-Line			
Lamotrigine	Lamotrigine	Valproic acid	Valproic acid
Valproic acid	Carbamazepine	Ethosuximide	Lamotrigine
	Oxcarbazepine	Lamotrigine	Topiramate
	Eslicarbazepine		
	Phenytoin		
	Levetiracetam		
Alternatives			
Zonisamide[a]	Zonisamide[a]	Clonazepam	Clonazepam
Phenytoin	Brivaracetam	Zonisamide	Felbamate
Levetiracetam	Topiramate	Levetiracetam	Clobazam
Carbamazepine	Valproic acid		Rufinamide
Oxcarbazepine	Tiagabine[a]		Fenfluramine
Topiramate	Gabapentin[a]		
Phenobarbital	Lacosamide[a]		
Primidone	Phenobarbital		
Felbamate	Primidone		
Perampanel	Felbamate		
	Perampanel		
	Cenobamate[a]		

[a]As adjunctive therapy.

In addition to efficacy, factors influencing the choice of an initial medication include the convenience of dosing (e.g., once daily vs three times daily) and potential side effects. In this regard, many of the newer drugs have the advantage of reduced drug-drug interactions and easier dosing. Almost all the commonly used antiseizure drugs can cause similar, dose-related side effects such as sedation, ataxia, dizziness, and diplopia. Long-term use of some agents in adults, especially the elderly, can lead to osteoporosis. Close follow-up is required to ensure these side effects are promptly recognized and reversed. Most of the older drugs and some of the newer ones can also cause idiosyncratic toxicity such as rash, bone marrow suppression, or hepatotoxicity. Although rare, these side effects should be considered during drug selection, and patients must be instructed about symptoms or signs that should signal the need to alert their health care provider. For some drugs, laboratory tests (e.g., complete blood count and liver function tests) are recommended prior to the institution of therapy (to establish baseline values) and during initial dosing and titration of the agent. Monitoring serum concentrations of antiseizure medications can help determine when a therapeutic dose has been reached, though clinical response is paramount (see below).

An important advance in the care of people with epilepsy has been the application of genetic testing to help guide the choice of therapy (as well as establishing the underlying cause of a patient's syndrome; Table 436-2). For example, the identification of a mutation in the *SLC2A1* gene, which encodes the glucose type 1 transporter (GLUT-1) and is a cause of GLUT-1 deficiency, should prompt immediate treatment with the ketogenic diet. Mutations of the *ALDH7A1* gene, which encodes antiquitin, can cause alterations in pyridoxine metabolism that are reversed by treatment with pyridoxine. There is also mounting evidence that certain gene mutations may indicate better or worse response to specific antiseizure drugs. For example, patients with mutations in the sodium channel subunit *SCN1A* should generally avoid taking phenytoin or lamotrigine, whereas patients with mutations in the *SCN2A* or *SCN8A* sodium channel subunits appear to respond favorably to high-dose phenytoin. Genetic testing may also help predict antiseizure drug toxicity. Studies have shown that individuals of Asian descent who carry the human leukocyte antigen (HLA) allele HLA-B*1502 are at particularly high risk of developing serious skin reactions from carbamazepine, phenytoin, oxcarbazepine, and lamotrigine. HLA-A*31:01 has also been found to be associated with carbamazepine-induced hypersensitivity reactions in patients of European or Japanese ancestry.

Antiseizure Drug Selection for Focal Seizures Carbamazepine (or related drugs, oxcarbazepine and eslicarbazepine), lamotrigine, phenytoin, and levetiracetam are currently the drugs of choice approved for the initial treatment of focal seizures, including those that evolve into generalized seizures. Overall, they have very similar efficacy, but differences in pharmacokinetics and toxicity are the main determinants for use in a given patient. For example, an advantage of carbamazepine (which is also available in an extended-release form) is that its metabolism follows first-order pharmacokinetics, which allows for a linear relationship between drug dose, serum levels, and toxicity. Carbamazepine can cause leukopenia, aplastic anemia, or hepatotoxicity and would therefore be

TABLE 436-9 Dosage and Adverse Effects of Commonly Used Antiepileptic Drugs

GENERIC NAME	TRADE NAME	PRINCIPAL USES	TYPICAL DOSE; DOSE INTERVAL	HALF-LIFE	THERAPEUTIC RANGE	ADVERSE EFFECTS		DRUG INTERACTIONS[a]
						NEUROLOGIC	SYSTEMIC	
Brivaracetam	Briviact	Focal onset	100–200 mg/d; bid	7–10 h	Not established	Fatigue Dizziness Weakness Ataxia Mood changes	Gastrointestinal irritation	May increase carbamazepine-epoxide causing decreased tolerability May increase phenytoin
Cannabidiol	Epidiolex	Dravet and Lennox-Gastaut syndromes	10–20 mg/kg per d; bid	18–32 h	Not established	Sedation	Elevated transaminases Anorexia Weight loss Diarrhea	Increases clobazam causing somnolence
		Tuberous sclerosis complex-associated seizures						
Carbamazepine	Tegretol[c]	Tonic-clonic Focal onset	600–1800 mg/d (15–35 mg/kg, child); bid (capsules or tablets), tid-qid (oral suspension)	10–17 h (variable due to autoinduction: complete 3–5 wk after initiation)	4–12 μg/mL	Ataxia Dizziness Diplopia Vertigo	Aplastic anemia Leukopenia Gastrointestinal irritation Hepatotoxicity Hyponatremia Rash	Level decreased by enzyme-inducing drugs[b] Level increased by erythromycin, propoxyphene, isoniazid, cimetidine, fluoxetine
Cenobamate	Xcopri	Focal onset	100–400 mg/d; daily (tablets)	50–60 h	Not established	Cognitive dysfunction Dizziness Disequilibrium Gait disturbance Headache	Anorexia Constipation Diarrhea Dyspepsia Nausea	Major CYP3A4 inducer; moderate CYP2C19 inhibitor
Clobazam	Onfi	Lennox-Gastaut syndrome	10–40 mg/d (5–20 mg/d for patients <30 kg body weight); bid	36–42 h (71–82 h for less active metabolite)	Not established	Fatigue Sedation Ataxia Aggression Insomnia	Constipation Anorexia Skin rash	Level increased by CYP2C19 inhibitors
Clonazepam	Klonopin	Absence Atypical absence Myoclonic	1–12 mg/d; qd-tid	24–48 h	10–70 ng/mL	Ataxia Sedation Lethargy	Anorexia	Level decreased by enzyme-inducing drugs[b]
Eslicarbazepine	Aptiom	Focal onset	400–1600 mg/d; qd	20–24 h	10–35 μg/mL (as oxcarbazepine mono-hydroxy derivative)	Sedation Ataxia Dizziness Diplopia Vertigo	See carbamazepine	Level decreased by enzyme-inducing drugs[b]
Ethosuximide	Zarontin	Absence	750–1250 mg/d (20–40 mg/kg); qd-bid	60 h, adult 30 h, child	40–100 μg/mL	Ataxia Lethargy Headache	Gastrointestinal irritation Skin rash Bone marrow suppression	Level decreased by enzyme-inducing drugs[b] Level increased by valproic acid
Felbamate	Felbatol	Focal onset Lennox-Gastaut syndrome Tonic-clonic	2400–3600 mg/d, tid-qid	16–22 h	30–60 μg/mL	Insomnia Dizziness Sedation Headache	Aplastic anemia Hepatic failure Weight loss Gastrointestinal irritation	Increases phenytoin, valproic acid, active carbamazepine metabolite
Fenfluramine	Fintepla	Dravet and Lennox-Gastaut syndromes	0.1–0.35 mg/kg/dose bid (oral solution); dosage depends on coadministration with stiripentol and/or clobazam	20 h	Not established	Ataxia Behavioral disturbance Headache Somnolence	Anorexia Constipation Hypertension Serotonin syndrome Weight loss	CYP1A2, CYP2B6, CYP2D6, CYPDA4 substrate

(*Continued*)

TABLE 436-9 Dosage and Adverse Effects of Commonly Used Antiepileptic Drugs (Continued)

GENERIC NAME	TRADE NAME	PRINCIPAL USES	TYPICAL DOSE; DOSE INTERVAL	HALF-LIFE	THERAPEUTIC RANGE	ADVERSE EFFECTS		DRUG INTERACTIONS[a]
						NEUROLOGIC	SYSTEMIC	
Gabapentin	Neurontin	Focal onset	900–2400 mg/d; tid-qid	5–9 h	2–20 μg/mL	Sedation Dizziness Ataxia Fatigue	Gastrointestinal irritation Weight gain Edema	No known significant interactions
Ganaxolone	Ztalmy	CDKL-5 deficiency disorder–associated seizures	450–1800 mg/d; tid (oral suspension)	34 h	Not established	Gait disturbance Somnolence	Bronchitis Fever Nasal congestion	CYP2B6, CYP2C19, CYP3A4 substrate
Lacosamide	Vimpat	Focal onset	200–400 mg/d; bid	13 h	Not established	Dizziness Ataxia Diplopia Vertigo	Gastrointestinal irritation Cardiac conduction (PR interval prolongation)	Level decreased by enzyme-inducing drugs[b]
Lamotrigine	Lamictal[c]	Focal onset Tonic-clonic Atypical absence Myoclonic Lennox-Gastaut syndrome	150–500 mg/d; bid (immediate release), daily (extended release) (lower daily dose for regimens with valproic acid; higher daily dose for regimens with an enzyme inducer)	25 h 14 h (with enzyme inducers), 59 h (with valproic acid)	2.5–20 μg/mL	Dizziness Diplopia Sedation Ataxia Headache	Skin rash Stevens-Johnson syndrome	Level decreased by enzyme-inducing drugs[b] and oral contraceptives Level increased by valproic acid
Levetiracetam	Keppra	Focal onset	1000–3000 mg/d; bid (immediate release), daily (extended release)	6–8 h	5–45 μg/mL	Sedation Fatigue Incoordination Mood changes	Anemia Leukopenia	No known significant interactions
Oxcarbazepine[c]	Trileptal	Focal onset Tonic-clonic	900–2400 mg/d (30–45 mg/kg, child); bid	10–17 h (for active metabolite)	10–35 μg/mL	Fatigue Ataxia Dizziness Diplopia Vertigo Headache	See carbamazepine	Level decreased by enzyme-inducing drugs[b] May increase phenytoin
Perampanel	Fycompa	Focal onset Tonic-clonic	4–12 mg; qd	105 h	Not established	Dizziness Somnolence Aggression Ataxia Anxiety Paranoia	Headache Nausea	Level decreased by enzyme-inducing drugs[b]
Phenobarbital	Luminal	Tonic-clonic Focal onset	60–180 mg/d; qd-tid	90 h	10–40 μg/mL	Sedation Ataxia Confusion Dizziness Decreased libido Depression	Skin rash	Level increased by valproic acid, phenytoin
Phenytoin[c] (diphenylhydantoin)	Dilantin	Tonic-clonic Focal onset	300–400 mg/d (3–6 mg/kg, adult; 4–8 mg/kg, child); qd-tid	24 h (wide variation, dose-dependent)	10–20 μg/mL	Dizziness Diplopia Ataxia Incoordination Confusion	Gingival hyperplasia Lymphadenopathy Hirsutism Osteomalacia Facial coarsening Skin rash	Level increased by isoniazid, sulfonamides, fluoxetine Level decreased by enzyme-inducing drugs[b] Altered folate metabolism
Primidone	Mysoline	Tonic-clonic Focal onset	750–1000 mg/d; bid-tid	Primidone, 8–15 h Phenobarbital, 90 h	Primidone, 4–12 μg/mL Phenobarbital, 10–40 μg/mL	Same as phenobarbital		Level increased by valproic acid Level decreased by phenytoin (increased conversion to phenobarbital)

(*Continued*)

TABLE 436-9 Dosage and Adverse Effects of Commonly Used Antiepileptic Drugs (Continued)

GENERIC NAME	TRADE NAME	PRINCIPAL USES	TYPICAL DOSE; DOSE INTERVAL	HALF-LIFE	THERAPEUTIC RANGE	ADVERSE EFFECTS		DRUG INTERACTIONS[a]
						NEUROLOGIC	SYSTEMIC	
Rufinamide	Banzel	Lennox-Gastaut syndrome	3200 mg/d (45 mg/kg, child); bid	6–10 h	Not established	Sedation Fatigue Dizziness Ataxia Headache Diplopia	Gastrointestinal irritation Leukopenia Cardiac conduction (QT interval shortening)	Level decreased by enzyme-inducing drugs[b] Level increased by valproic acid May increase phenytoin
Tiagabine	Gabitril	Focal onset	32–56 mg/d; bid-qid (as adjunct to enzyme-inducing antiepileptic drug regimen)	2–5 h (with enzyme inducer), 7–9 h (without enzyme inducer)	Not established	Confusion Sedation Depression Dizziness Speech or language problems Paresthesias Psychosis	Gastrointestinal irritation	Level decreased by enzyme-inducing drugs[b]
Topiramate[c]	Topamax	Focal onset Tonic-clonic Lennox-Gastaut syndrome	200–400 mg/d; bid (immediate release), daily (extended release)	20 h (immediate release), 30 h (extended release)	2–20 μg/mL	Psychomotor slowing Sedation Speech or language problems Fatigue Paresthesias	Renal stones (avoid use with other carbonic anhydrase inhibitors) Glaucoma Weight loss Hypohidrosis	Level decreased by enzyme-inducing drugs[b]
Valproic acid (valproate sodium, divalproex sodium)	Depakene Depakote	Tonic-clonic Absence Atypical absence Myoclonic Focal onset Atonic	750–2000 mg/d (20–60 mg/kg); bid-qid (immediate and delayed release), daily (extended release)	15 h	50–125 μg/mL	Ataxia Sedation Tremor	Hepatotoxicity Thrombocytopenia Gastrointestinal irritation Weight gain Transient alopecia Hyperammonemia	Level decreased by enzyme-inducing drugs[b]
Zonisamide	Zonegran	Focal onset Tonic-clonic	200–400 mg/d; qd-bid	50–68 h	10–40 μg/mL	Sedation Dizziness Confusion Headache Psychosis	Anorexia Renal stones Hypohidrosis	Level decreased by enzyme-inducing drugs[b]

[a] Examples only; please refer to other sources for comprehensive listings of all potential drug-drug interactions. [b] Phenytoin, carbamazepine, phenobarbital. [c] Extended-release product available.

contraindicated in patients with predispositions to these problems. Oxcarbazepine has the advantage of being metabolized in a way that avoids an intermediate metabolite ("toxic epoxide") associated with some of the side effects of carbamazepine. Oxcarbazepine also has fewer drug interactions than carbamazepine. Eslicarbazepine has a long serum half-life and is dosed once daily.

Lamotrigine tends to be well tolerated in terms of side effects and has mood-stabilizing properties that can be beneficial. However, patients need to be particularly vigilant about the possibility of a skin rash during the initiation of therapy. This can be extremely severe and lead to Stevens-Johnson syndrome if unrecognized and if the medication is not discontinued immediately. This risk can be reduced by using low initial doses and slow titration. Lamotrigine must be started at even lower initial doses when used as add-on therapy with valproic acid, because valproic acid inhibits lamotrigine metabolism and substantially prolongs its half-life.

Phenytoin has a relatively long half-life and offers the advantage of once- or twice-daily dosing compared to twice- or thrice-daily dosing for many of the other drugs. However, phenytoin shows properties of nonlinear kinetics, such that small increases in phenytoin doses above a standard maintenance dose can precipitate marked side effects. This is one of the main causes of acute phenytoin toxicity (dizziness, diplopia, ataxia). Long-term use of phenytoin is associated with untoward cosmetic effects (e.g., hirsutism, coarsening of facial features, gingival hypertrophy) and osteoporosis. Due to these side effects, phenytoin is often avoided in young patients who are likely to require the drug for many years.

Levetiracetam has the advantage of having no known clinically relevant drug-drug interactions, making it especially useful in the elderly and patients on other medications. However, a significant number of patients taking levetiracetam complain of irritability, anxiety, and other psychiatric symptoms.

Topiramate can be used for both focal and generalized seizures. Like some of the other antiseizure drugs, topiramate can cause significant psychomotor slowing and other cognitive problems. Additionally, it should not be used in patients at risk for renal stones.

Valproic acid is an effective alternative for some patients with focal seizures, especially when the seizures generalize. Gastrointestinal side effects are fewer when using the delayed-release formulation. Laboratory testing is required to monitor toxicity because valproic acid can rarely cause reversible bone marrow suppression and hepatotoxicity. This drug should generally be avoided in patients with preexisting bone marrow or liver disease. Valproic acid also has relatively high risks of unacceptable adverse effects for women of childbearing age, including hyperandrogenism, that may affect fertility and teratogenesis (e.g., neural tube defects) in offspring. Irreversible, fatal hepatic failure appearing as an idiosyncratic rather than dose-related side effect is a relatively rare complication; its risk is highest in children <2 years old, especially those taking other antiseizure drugs or with inborn errors of metabolism.

Zonisamide, brivaracetam, tiagabine, gabapentin, perampanel, and lacosamide are additional drugs currently used for the treatment of focal seizures with or without evolution into generalized seizures. Phenobarbital and other barbiturate compounds were commonly used in the past as first-line therapy for many forms of epilepsy. However, the barbiturates frequently cause sedation in adults, hyperactivity in children, and other more subtle cognitive changes; thus, their use should be limited to situations in which no other suitable treatment alternatives exist. Cenobamate is a recently approved antiseizure drug that has been shown to significantly improve seizure control in patients with focal epilepsy who were not adequately treated with up to three medications.

Antiseizure Drug Selection for Generalized Seizures Lamotrigine, valproic acid, and levetiracetam are currently considered the best initial choice for the treatment of primary generalized, tonic-clonic seizures. Topiramate, zonisamide, perampanel, phenytoin, carbamazepine, and oxcarbazepine are suitable alternatives, although carbamazepine, oxcarbazepine, and phenytoin can worsen certain types of generalized seizures. Valproic acid is particularly effective in absence, myoclonic, and atonic seizures. It is therefore commonly used in patients with generalized epilepsy syndromes having mixed seizure types. However, levetiracetam, rather than valproic acid, is increasingly considered the initial drug of choice for women with epilepsies having mixed seizure types given the adverse effects of valproic acid for women of childbearing age (discussed below). Lamotrigine is also an alternative to valproate, especially for absence epilepsies. Ethosuximide is a particularly effective drug for the treatment of absence seizures, but it is not useful for tonic-clonic or focal seizures. Periodic monitoring of blood cell counts is required since ethosuximide rarely causes bone marrow suppression.

INITIATION AND MONITORING OF THERAPY

Because the response to any antiseizure drug is unpredictable, patients should be carefully educated about the approach to therapy. The goal is to prevent seizures and minimize the side effects of treatment; determination of the optimal medication and the optimal dose typically involves trial and error. This process may take months or longer if the baseline seizure frequency is low. Most antiseizure drugs need to be introduced relatively slowly to minimize side effects. Patients should expect that minor side effects such as mild sedation, slight changes in cognition, or imbalance will typically resolve within a few days. Starting doses are usually the lowest value listed under the dosage column in Table 436-9. Subsequent increases should be made only after achieving a steady state with the previous dose (i.e., after an interval of five or more half-lives).

Monitoring of serum antiseizure drug levels can be very useful for establishing the initial dosing schedule. However, the published therapeutic ranges of serum drug concentrations are only an approximate guide for determining the proper dose for a given patient. The key determinants are the clinical measures of seizure frequency and presence of side effects, not the laboratory values. Conventional assays of serum drug levels measure the total drug (i.e., both free and protein bound). However, it is the concentration of free drug that reflects extracellular levels in the brain and correlates best with efficacy. Thus, patients with decreased levels of serum proteins (e.g., decreased serum albumin due to impaired liver or renal function) may have an increased ratio of free to bound drug, yet the concentration of free drug may be adequate for seizure control. These patients may have a "subtherapeutic" drug level, but the dose should be changed only if seizures remain uncontrolled, not just to achieve a "therapeutic" level. It is also useful to monitor free drug levels in such patients. In practice, other than during the initiation or modification of therapy, monitoring of antiseizure drug levels is most useful for documenting adherence, assessing clinical suspicion of toxicity, or establishing baseline serum concentrations prior to pregnancy, when clearance of many antiseizure drugs increases significantly.

If seizures continue despite gradual increases to the maximum tolerated dose and documented compliance, then it becomes necessary to switch to another antiseizure drug. This is usually done by maintaining the patient on the first drug while a second drug is added. The dose of the second drug should be adjusted to decrease seizure frequency without causing toxicity. Once this is achieved, the first drug can be gradually withdrawn (usually over weeks unless there is significant toxicity). The dose of the second drug is then further optimized based on seizure response and side effects. Monotherapy should be the goal whenever possible.

WHEN TO DISCONTINUE THERAPY

Some patients who have their seizures completely controlled with antiseizure drugs can eventually discontinue therapy. The following patient profile yields the greatest chance of remaining seizure free after drug withdrawal: (1) complete medical control of seizures for 1–5 years; (2) single seizure type, with generalized seizures having a better prognosis than focal seizures; (3) normal neurologic examination, including intelligence; (4) no family history of epilepsy; and (5) normal EEG. The appropriate seizure-free interval is unknown and depends on the form of epilepsy and whether or not the causal factor is still present (e.g., resection of a brain tumor causing seizures). However, it seems reasonable to attempt withdrawal of therapy after 2 years in a patient who meets all the above criteria, is motivated to discontinue the medication, and clearly understands the potential risks and benefits. In most cases, it is preferable to reduce the dose of the drug gradually over 2–3 months. Most recurrences occur in the first 3 months after discontinuing therapy, and patients should be advised to avoid potentially dangerous situations such as driving or swimming during this period. Rarely, seizure-free patients who discontinue antiseizure medications but then have a recurrent seizure may not regain full control when these medications are resumed.

TREATMENT OF REFRACTORY EPILEPSY

Approximately half of people with epilepsy do not respond to treatment with the first antiseizure drug, and it becomes necessary to try additional drugs alone or in combination. Patients who have focal epilepsy related to an underlying structural lesion or those with multiple seizure types and developmental delay are particularly likely to require multiple drugs. There are currently no clear guidelines for rational polypharmacy, although in theory, a combination of drugs with different mechanisms of action may be most useful. In most cases, the initial combination therapy combines first-line drugs (i.e., carbamazepine, oxcarbazepine, lamotrigine, valproic acid, levetiracetam, and phenytoin). If these drugs are unsuccessful, then the addition of other drugs such as cenobamate, zonisamide, brivaracetam, topiramate, or lacosamide is indicated. Patients with myoclonic seizures resistant to valproic acid may benefit from the addition of levetiracetam, zonisamide, clonazepam, or clobazam, and those with absence seizures may respond to a combination of valproic acid and ethosuximide. The same principles concerning the monitoring of therapeutic response, toxicity, and serum levels for monotherapy apply to polypharmacy, and potential drug interactions need to be recognized. If there is no improvement, a third drug can be added while the first two are maintained. If there is a response, the less effective or less well tolerated of the first two drugs should be gradually withdrawn.

SURGICAL TREATMENT OF REFRACTORY EPILEPSY

Approximately 30% of patients with epilepsy continue to have seizures despite efforts to find an effective combination of antiseizure drugs. For some patients with focal epilepsy, surgery can be extremely effective in substantially reducing seizure frequency and even providing complete seizure control. Understanding the potential value of surgery is especially important when a patient's seizures are not controlled with initial treatment, as such patients often do not respond to subsequent medication trials. Rather than submitting the patient to years of unsuccessful medical therapy and the psychosocial trauma and increased mortality associated with ongoing seizures, the patient should have an efficient but relatively brief attempt at medical therapy and then be referred for surgical evaluation.

The most common surgical procedure for patients with temporal lobe epilepsy involves resection of the anteromedial temporal lobe (temporal lobectomy) or a more limited removal of the underlying hippocampus and amygdala (amygdalohippocampectomy). Focal seizures arising from extratemporal regions may be abolished by a focal neocortical resection with precise removal of an identified lesion (lesionectomy). Localized neocortical resection without a clear lesion identified on MRI is also possible when other tests (e.g., MEG, PET, SPECT) implicate a focal cortical region as a seizure onset zone. When the cortical region cannot be removed, multiple subpial transection, which disrupts intracortical connections, is sometimes used to prevent seizure spread. Hemispherectomy or multilobar resection is useful for some patients with severe seizures due to hemispheric abnormalities such as hemimegalencephaly or other dysplastic abnormalities, and corpus callosotomy has been shown to be effective for disabling tonic or atonic seizures, usually when they are part of a mixed-seizure syndrome (e.g., Lennox-Gastaut syndrome).

Presurgical evaluation is designed to identify the functional and structural basis of the patient's seizure disorder. Inpatient video-EEG monitoring is used to localize the seizure focus and to correlate the abnormal electrophysiologic activity with neuroimaging and behavioral manifestations of the seizure. Routine scalp EEG recordings and a high-resolution MRI scan are often sufficient for localization of the epileptogenic focus, especially when the findings are concordant. Functional imaging studies such as SPECT, PET, and MEG are adjunctive tests that may help to reveal or verify the localization of an apparent epileptogenic region. Once the presumed location of the seizure onset is identified, additional studies, including neuropsychological testing, the intracarotid amobarbital (Wada) test, and functional MRI may be used to assess language and memory localization and to determine the possible functional consequences of surgical removal of the epileptogenic region. In some cases, standard noninvasive evaluation is not sufficient to localize the seizure onset zone, and invasive electrophysiologic monitoring is required for more definitive localization. Traditionally, this required open craniotomy and subdural electrode placement. Robot-assisted stereotactic EEG (stereo-EEG) has surged in use as a less invasive surgical option that involves placing depth electrodes through burr holes in the skull and into the brain parenchyma to record from regions suspected of generating seizures. Although subdural electrodes provide extensive cortical sampling, they are unable to record from deep structures; depth electrodes permit sampling of both cortical and subcortical tissue, albeit with more restricted spatial resolution. Stereo-EEG studies are typically associated with shorter hospital stays and lower complication rates than subdural electrode studies.

The exact extent of the resection to be undertaken can also be determined by performing cortical mapping at the time of the surgical procedure, allowing for a tailored resection. This involves electrocorticographic recordings made with electrodes on the surface of the brain to identify the extent of epileptiform disturbances. If the region to be resected is within or near brain regions suspected of having sensorimotor or language function, electrical cortical stimulation mapping is performed on the awake patient to determine the function of cortical regions in question to avoid resection of so-called eloquent cortex and thereby minimize postsurgical deficits.

Advances in presurgical evaluation and microsurgical techniques have led to a steady increase in the success of epilepsy surgery. Clinically significant complications of surgery are <5%, and the use of functional mapping procedures has markedly reduced the neurologic sequelae due to removal or sectioning of brain tissue. For example, ~70% of well-selected patients treated with temporal lobectomy will become seizure free, and another 15–25% will have at least a 90% reduction in seizure frequency. Marked improvement is also usually seen in patients treated with hemispherectomy for catastrophic seizure disorders due to large hemispheric abnormalities. Postoperatively, patients generally need to remain on antiseizure drug therapy, but the marked reduction of seizures following resective surgery can have a very beneficial effect on quality of life. Recently, catheter-based stereotactic laser thermal ablation has been developed as a less invasive means for destroying the seizure focus in select patients.

Not all medically refractory patients are suitable candidates for resective surgery or laser ablation. For example, some patients have seizures arising from more than one brain region or from a single "eloquent" region that mediates a critical function (e.g., vision, movement, language), such that the potential harm from removal is unacceptably high. In these patients, implanted neurostimulation devices that deliver electrical energy to the brain to reduce seizures represent palliative treatment options. Vagus nerve stimulation (VNS) involves an extracranial device that works through scheduled intermittent ("open loop") stimulation of the left vagus nerve. Efficacy of VNS is limited, and side effects related to recurrent laryngeal nerve activation (e.g., hoarseness, throat pain, dyspnea) can be significant and dose-limiting. By contrast, responsive neurostimulation (RNS) involves an implanted device connected to two lead wires that are placed intracranially at the site(s) from where seizures arise. The neurostimulator detects the onset of a seizure (often before the seizure becomes clinically apparent) and delivers electrical stimulation—typically imperceptible—directly to the brain to reduce seizures over time, a form of "closed loop" neurostimulation. RNS is the only device that provides chronic EEG, which has a growing number of clinical applications, such as quantifying the lateralization of seizures arising from both sides of the brain, characterizing clinical spells, assessing effects of medications and other therapeutic interventions, and revealing cyclical patterns of epileptic brain activity that may help anticipate future events. A third modality, thalamic deep brain stimulation (DBS), involves open loop stimulation of deep, bilateral cerebral structures, the anterior thalamic nuclei, which are key nodes in limbic circuits mediating certain types of seizures. Whereas precise seizure localization is necessary for RNS, it is not required for VNS or DBS. Stimulation of thalamic nuclei, which project broadly to different cortical regions, with RNS or DBS is an appealing approach to treating poorly localized, spatially extensive, or multifocal seizure foci.

Long-term clinical trials of all three neurostimulation devices demonstrate significant reductions in frequency with outcomes improving over time, but only a minority of patients treated with these devices achieve seizure freedom (e.g., ~15% with RNS). Furthermore, no head-to-head device trials exist to establish relative superiority, so choice of a device is guided by patient-specific factors and by the strengths and limitations of each technology.

STATUS EPILEPTICUS

Status epilepticus refers to continuous seizures or repetitive, discrete seizures with impaired consciousness in the interictal period. Status epilepticus has numerous subtypes, including generalized convulsive status epilepticus (GCSE) (e.g., persistent, generalized electrographic seizures, coma, and tonic-clonic movements) and nonconvulsive status epilepticus (NCSE; e.g., persistent absence seizures or focal seizures with confusion or partially impaired consciousness, and minimal motor abnormalities). The duration of seizure activity sufficient to meet the definition of status epilepticus has traditionally been specified as 15–30 min. However, a more practical definition is to consider status epilepticus as a situation in which the duration of seizures prompts the acute use of anticonvulsant therapy. For GCSE, this is typically when seizures last beyond 5 min.

GCSE is an emergency and must be treated immediately, because cardiorespiratory dysfunction, hyperthermia, and metabolic derangements can develop as a consequence of prolonged seizures, and these can lead to irreversible CNS injury. Furthermore, CNS injury can occur even when the patient is paralyzed with neuromuscular blockade but continues to have electrographic seizures. The most common causes of GCSE are anticonvulsant withdrawal or noncompliance, metabolic disturbances, drug toxicity, CNS infection, CNS tumors, refractory epilepsy, and head trauma.

GCSE is obvious when the patient is having overt seizures. However, after 30–45 min of uninterrupted seizures, the signs may become increasingly subtle. Patients may have mild clonic movements of only the fingers or fine, rapid movements of the eyes. There may be paroxysmal episodes of tachycardia, hypertension, and pupillary dilation. In such cases, the EEG may be the only method of establishing the diagnosis. Thus, if the patient stops having overt seizures, yet remains comatose, an EEG should be performed to rule out ongoing status epilepticus. This is obviously also essential when a patient with GCSE has been paralyzed with neuromuscular blockade in the process of protecting the airway.

The first steps in the management of a patient in GCSE are to attend to any acute cardiorespiratory problems or hyperthermia, perform a brief medical and neurologic examination, establish venous access, and send samples for laboratory studies to identify metabolic abnormalities. Anticonvulsant therapy should then begin without delay (Fig. 436-5).

The treatment of NCSE is thought to be less urgent than GCSE, because the ongoing seizures are not accompanied by the severe metabolic disturbances seen with GCSE. However, evidence suggests that NCSE, especially that caused by ongoing, focal seizure activity, is associated with cellular injury in the region of the seizure focus; therefore, this condition should be treated as promptly as possible using the general approach described for GCSE. Portable headband devices that provide a limited form of EEG can be rapidly applied without skilled technicians to help rule out NCSE in acute care settings.

BEYOND SEIZURES: OTHER MANAGEMENT ISSUES

EPILEPSY COMORBIDITIES

The adverse effects of epilepsy often go beyond clinical seizures. Many people with epilepsy feel completely normal between seizures and live highly successful and productive lives. However, a significant proportion of patients suffer from varying degrees of cognitive dysfunction, including psychiatric disease, and it has become increasingly clear that the network dysfunction underlying epilepsy can have effects well beyond the occurrence of seizures. For example, patients with seizures secondary to developmental abnormalities or acquired brain injury may have impaired cognitive function and other neurologic deficits due to abnormal brain structure. Frequent interictal EEG abnormalities are associated with subtle dysfunction of memory and attention. Patients with many seizures, especially those emanating from the temporal lobe, often note an impairment of short-term memory that may progress over time.

The psychiatric problems associated with epilepsy include depression, anxiety, and psychosis. This risk varies considerably depending on many factors, including the etiology, frequency, and severity of seizures and the patient's age and previous personal or family history of psychiatric disorder. Depression occurs in ~20–30% of patients, and the incidence of suicide is higher in people with epilepsy than in the general population. Depression should be treated through counseling and/or medication. The selective serotonin reuptake inhibitors (SSRIs) typically have minimal effect on seizures, whereas tricyclic antidepressants may lower the seizure threshold. Anxiety can be a seizure symptom, and anxious or psychotic behavior can occur during a postictal delirium. Postictal psychosis is a rare phenomenon that typically occurs after a period of increased seizure frequency. There is

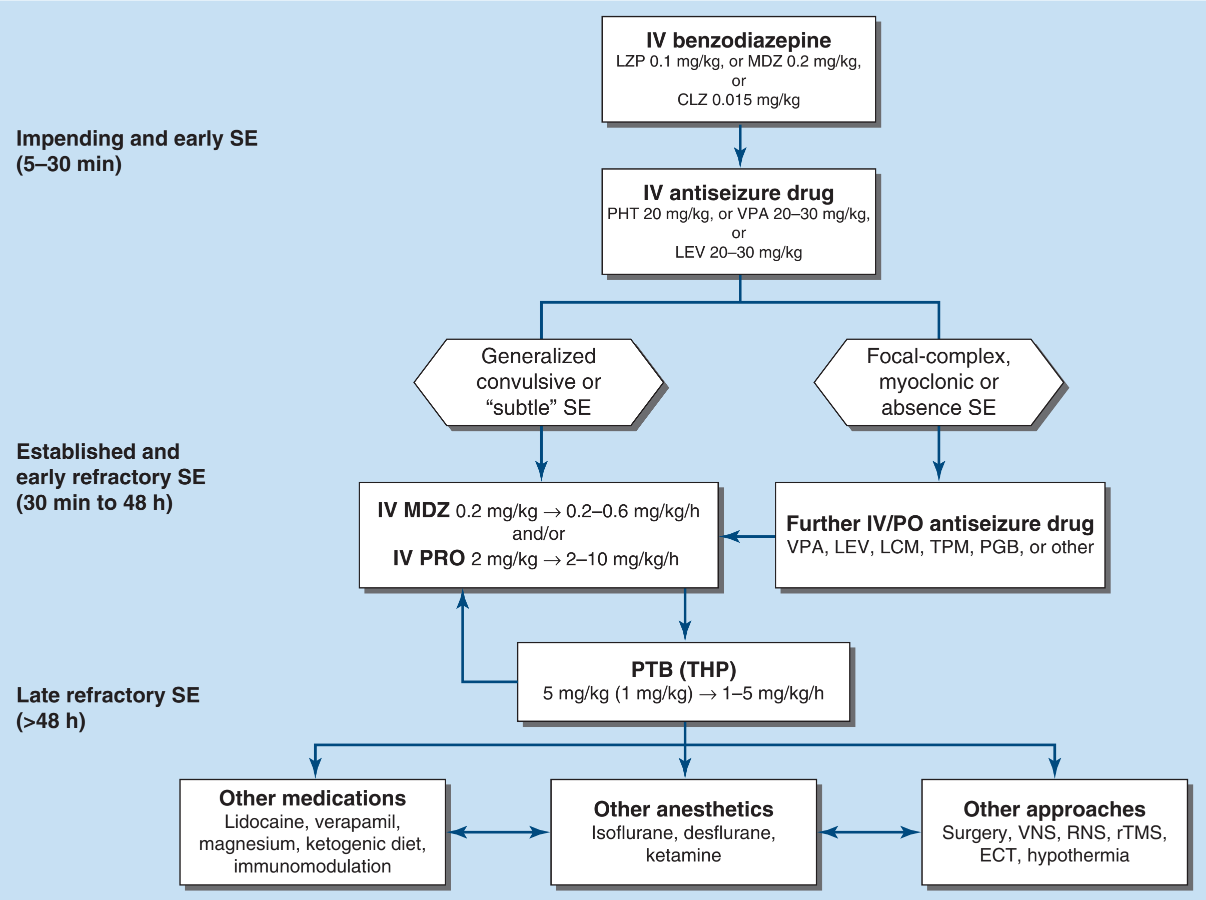
Impending and early SE
(5–30 min)
IV benzodiazepine
LZP 0.1 mg/kg, or MDZ 0.2 mg/kg, or CLZ 0.015 mg/kg
IV antiseizure drug
PHT 20 mg/kg, or VPA 20–30 mg/kg, or LEV 20–30 mg/kg
Generalized convulsive or "subtle" SE
Focal-complex, myoclonic or absence SE
Established and early refractory SE
(30 min to 48 h)
IV MDZ 0.2 mg/kg → 0.2–0.6 mg/kg/h and/or IV PRO 2 mg/kg → 2–10 mg/kg/h
Further IV/PO antiseizure drug
VPA, LEV, LCM, TPM, PGB, or other
Late refractory SE
(>48 h)
PTB (THP)
5 mg/kg (1 mg/kg) → 1–5 mg/kg/h
Other medications
Lidocaine, verapamil, magnesium, ketogenic diet, immunomodulation
Other anesthetics
Isoflurane, desflurane, ketamine
Other approaches
Surgery, VNS, RNS, rTMS, ECT, hypothermia

FIGURE 436-5 Pharmacologic treatment of generalized tonic-clonic status epilepticus (SE) in adults. CLZ, clonazepam; ECT, electroconvulsive therapy; LCM, lacosamide; LEV, levetiracetam; LZP, lorazepam; MDZ, midazolam; PGB, pregabalin; PHT, phenytoin or fosphenytoin; PRO, propofol; PTB, pentobarbital; RNS, responsive neurostimulation; rTMS, repetitive transcranial magnetic stimulation; THP, thiopental; TPM, topiramate; VNS, vagus nerve stimulation; VPA, valproic acid. *(Data from AO Rossetti, DH Lowenstein: Management of refractory status epilepticus in adults: still more questions than answers. Lancet Neurol 10:922, 2011.)*

usually a brief lucid interval lasting up to a week, followed by days to weeks of agitated, psychotic behavior. The psychosis usually resolves spontaneously but frequently will require short-term treatment with antipsychotic or anxiolytic medications.

MORTALITY OF EPILEPSY

People with epilepsy have a risk of death that is roughly two to three times greater than expected in a matched population without epilepsy. Most of the increased mortality is due to the underlying etiology of epilepsy (e.g., tumors or strokes in older adults). However, a significant number of patients die from accidents, status epilepticus, and a syndrome known as *sudden unexpected death in epilepsy* (SUDEP), which usually affects young people with convulsive seizures and tends to occur at night. The cause of SUDEP is unknown; it may result from brainstem-mediated effects of seizures on pulmonary, cardiac, and arousal functions. Genetic mutations may be the cause of both epilepsy and a cardiac conduction defect that gives rise to sudden death.

PSYCHOSOCIAL ISSUES

There continues to be a cultural stigma about epilepsy, although it is slowly declining in societies with effective health education programs. Many people with epilepsy harbor fear of progressive cognitive decline or dying during a seizure. These issues need to be carefully addressed by educating the patient about epilepsy and by ensuring that family members, teachers, fellow employees, and other associates are equally well informed. A useful source of educational material is the website *www.epilepsy.com*.

EMPLOYMENT, DRIVING, AND OTHER ACTIVITIES

Many patients with epilepsy face difficulty in obtaining or maintaining employment, even when their seizures are well controlled. Federal and state legislation is designed to prevent employers from discriminating against people with epilepsy, and patients should be encouraged to understand and claim their legal rights. Patients in these circumstances also benefit greatly from the assistance of health providers who act as strong patient advocates.

Loss of driving privileges is one of the most disruptive social consequences of epilepsy. Physicians should be very clear about local regulations concerning driving and epilepsy because the laws vary considerably among states and countries. In all cases, it is the physician's responsibility to warn patients of the danger imposed on themselves and others while driving if their seizures are uncontrolled (unless the seizures are not associated with impairment of consciousness or motor control). In general, most states in the United States allow patients to drive after a seizure-free interval (on or off medications) of 3–18 months.

Patients with incompletely controlled seizures must also contend with the risk of being in other situations where an impairment of consciousness or loss of motor control could lead to major injury or death. Thus, depending on the type and frequency of seizures, many patients need to be instructed to avoid working at heights or with machinery or to have someone close by for activities such as bathing and swimming.

The importance of quantifying seizures in people living with epilepsy has catalyzed a burgeoning industry of wearable sensors, such as wristwatches, that can detect seizures through noninvasive measurement of physiologic variables. Generally, non-EEG devices either have low sensitivity or high false alarm rate, but reliability is highest for detection of tonic-clonic seizures.

SPECIAL ISSUES RELATED TO WOMEN AND EPILEPSY

CATAMENIAL EPILEPSY

Some women experience a marked increase in seizure frequency around the time of menses. This is believed to be mediated by either the effects of estrogen and progesterone on neuronal excitability or changes in antiseizure drug levels due to altered protein binding or metabolism. Vulnerability to seizures is typically highest just before and during menses and during ovulation due to relatively high estrogen and low progesterone levels. Some women with epilepsy may benefit from increases in antiseizure drug dosages during menses. Natural progestins or intramuscular medroxyprogesterone may be of benefit to a subset of women.

PREGNANCY

Most women with epilepsy who become pregnant will have an uncomplicated gestation and deliver a normal baby. However, epilepsy poses some important risks to a pregnancy. Seizure frequency during pregnancy will remain unchanged in ~50% of women, increase in ~30%, and decrease in ~20%. Changes in seizure frequency are attributed to endocrine effects on the CNS, variations in antiseizure drug pharmacokinetics (such as acceleration of hepatic drug metabolism, increased renal blood flow, increased volume of distribution, or effects on plasma protein binding), or decreased antiseizure medication absorption due to nausea and vomiting. It is useful to see patients at frequent intervals during pregnancy and monitor serum antiseizure drug levels monthly; the risk of seizure exacerbation increases when serum levels decrease >35% from prepregnancy levels. Measurement of the unbound drug concentrations may be useful if there is an increase in seizure frequency or worsening of side effects of antiseizure drugs.

The overall incidence of fetal abnormalities in children born to mothers with epilepsy is 5–6%, compared to 2–3% in healthy women. Part of the higher incidence is due to teratogenic effects of antiseizure drugs, and the risk increases with the number of medications used (e.g., 10–20% risk of malformations with three drugs) and possibly with higher doses. A meta-analysis of published pregnancy registries and cohorts found that the most common malformations were defects in the cardiovascular and musculoskeletal system (1.4–1.8%). Valproic acid is strongly associated with an increased risk of adverse fetal outcomes (7–20%). Findings from a large pregnancy registry suggest that the newer antiseizure drugs are far safer than valproic acid.

Because the potential harm of uncontrolled convulsive seizures on the mother and fetus is considered greater than the teratogenic effects of antiseizure drugs, it is currently recommended that pregnant women be maintained on effective drug therapy. When possible, it seems prudent to have the patient on monotherapy at the lowest effective dose, especially during the first trimester. For some women, however, the type and frequency of their seizures may allow for them to safely wean off antiseizure drugs prior to conception. Patients should also take folate (1–4 mg/d), because the antifolate effects of anticonvulsants are thought to play a role in the development of neural tube defects, although the benefits of this treatment remain unproven in this setting.

Enzyme-inducing drugs such as phenytoin, carbamazepine, oxcarbazepine, topiramate, phenobarbital, and primidone cause a transient and reversible deficiency of vitamin K–dependent clotting factors in ~50% of newborn infants. Although neonatal hemorrhage is uncommon, the mother should be treated with oral vitamin K (20 mg/d, phylloquinone) in the last 2 weeks of pregnancy, and the infant should receive intramuscular vitamin K (1 mg) at birth.

CONTRACEPTION

Special care should be taken when prescribing antiseizure medications for women who are taking oral contraceptive agents. Drugs such as carbamazepine, phenytoin, phenobarbital, and topiramate can significantly decrease the efficacy of oral contraceptives via enzyme induction and other mechanisms. Patients should be advised to consider alternative forms of contraception, including intrauterine devices and other long-acting reversible contraceptives, or their oral contraceptive medications should be modified to offset the effects of the antiseizure medications. Estrogen-containing oral contraceptive agents induce glucuronidation of lamotrigine and can decrease lamotrigine serum levels by >50%.

BREAST-FEEDING

Antiseizure medications are excreted into breast milk, and the ratio of drug concentration in breast milk relative to serum ranges from ~5% (valproic acid) to 300% (levetiracetam). Given the overall benefits of breast-feeding and the lack of evidence for long-term harm to the infant by being exposed to antiseizure drugs, mothers with epilepsy should be encouraged to breast-feed unless there is evidence of drug effects on the infant, such as lethargy or poor feeding.

SPECIAL ISSUES RELATED TO EPILEPSY IN THE ELDERLY

Epilepsy has a bimodal distribution according to age, with the highest incidence in the very young and in the elderly. The increased incidence in the elderly may be attributed to age- and aging-related epileptogenic factors. The most common cause is stroke, followed by neoplasm, and dementia, resulting in a preponderance of extratemporal epilepsy. Antiseizure medication selection requires careful consideration of medical and psychiatric comorbidities, side effect profiles, effects on mood and cognition, and drug-drug interactions.

■ FURTHER READING

Bui E: Women's issues in epilepsy. Continuum (Minneap Minn) 28:399, 2022.

Cornes SB, Shih T: Evaluation of the patient with spells. Continuum (Minneap Minn) 17:984, 2011.

Crepeau AZ, Sirven JI: Management of adult onset seizures. Mayo Clin Proc 92:306, 2017.

Ellis CA et al: Epilepsy genetics: Clinical impacts and biological insights. Lancet Neurol 19:93, 2020.

Fisher RS et al: Operational classification of seizure types by the International League Against Epilepsy: Position paper of the ILAE Commission for Classification and Terminology. Epilepsia 58:522, 2017.

Gavvala JR, Schuele SU: New-onset seizure in adults and adolescents: A review. JAMA 316:2657, 2016.

Jetté N et al: Surgical treatment for epilepsy: The potential gap between evidence and practice. Lancet Neurol 15:982, 2016.

Kanner AM: Management of psychiatric and neurological comorbidities in epilepsy. Nat Rev Neurol 12:106, 2016.

Krumholz A et al: Evidence-based guideline: Management of an unprovoked first seizure in adults: Report of the Guideline Development Subcommittee of the American Academy of Neurology and the American Epilepsy Society. Neurology 84:1705, 2015.

Kwan P, Brodie MJ: Early identification of refractory epilepsy. N Engl J Med 342:314, 2000.

Markert MS, Fisher RS: Neuromodulation: Science and practice in epilepsy: Vagus nerve simulation, thalamic deep brain stimulation, and responsive neurostimulation. Expert Rev Neurother 19:17, 2019.

Rao VR et al: Cues for seizure timing. Epilepsia 62:S15, 2021.

Sen A et al: Epilepsy in older people. Lancet 395:735, 2020.

437

Introduction to Cerebrovascular Diseases

Wade S. Smith, J. Claude Hemphill, III

Cerebrovascular diseases include some of the most common and devastating disorders: ischemic stroke and hemorrhagic stroke. Stroke is the second leading cause of death worldwide, with 7.1 million dying from stroke in 2020. Nearly 7 million Americans age 20 or older report having had a stroke, and the prevalence is estimated to rise by 3.4 million adults in the next decade, representing 4% of the entire adult population. Although mortality is increased from 6.2 million in 2010, the age standardized death rate has fallen by 15% in this decade, likely due to better prevention and treatment. However, overall disease burden will continue to climb as the population ages, and stroke is likely to remain the second most common disabling condition in individuals aged 50 or older worldwide.

A stroke, or cerebrovascular accident, is defined as an abrupt onset of a neurologic deficit that is attributable to a vascular cause. Thus, the definition of stroke is clinical, and laboratory studies including brain imaging are used to support the diagnosis. The clinical manifestations of stroke are highly variable because of the complex anatomy of the brain and its vasculature. *Cerebral ischemia* is caused by a reduction in blood flow that lasts longer than several seconds. Neurologic symptoms are manifest within seconds because neurons lack glycogen, so energy failure is rapid. If the cessation of flow lasts for more than a few minutes, *infarction* or death of brain tissue results. When blood flow is quickly restored, brain tissue can recover fully and the patient's symptoms are only transient: this is called a *transient ischemic attack* (TIA). The definition of TIA requires that all neurologic signs and symptoms resolve within 24 h without evidence of brain infarction on brain imaging. Stroke has occurred if the neurologic signs and symptoms last for >24 h or brain infarction is demonstrated. A generalized reduction in cerebral blood flow due to systemic hypotension (e.g., cardiac arrhythmia, sepsis, or hemorrhagic shock) usually produces syncope **(Chap. 23)**. If low cerebral blood flow persists for a longer duration, then infarction in the border zones between the major cerebral artery distributions may develop. In more severe instances as in cardiac arrest, *global hypoxia-ischemia* causes widespread brain injury; the constellation of cognitive sequelae that ensues is called *hypoxic-ischemic encephalopathy* **(Chap. 318)**. *Focal ischemia* or infarction, conversely, is usually caused by thrombosis of the cerebral vessels themselves or by emboli from a proximal arterial source or the heart **(Chap. 438)**. *Intracranial hemorrhage* is caused by bleeding directly into or around the brain; it produces neurologic symptoms by producing a mass effect on neural structures, from the toxic effects of blood itself, or by increasing intracranial pressure **(Chap. 439)**.

APPROACH TO THE PATIENT

Cerebrovascular Disease

Rapid evaluation is essential for use of acute treatments such as thrombolysis or thrombectomy. However, patients with acute stroke often do not seek medical assistance on their own because they may lose the appreciation that something is wrong (anosognosia) or lack the knowledge that acute treatment is beneficial; it is often a family member or a bystander who calls for help. Therefore, patients and their family members should be counseled to call emergency medical services immediately if they experience or witness the sudden onset of any of the following: loss of sensory and/or motor function on one side of the body (nearly 85% of ischemic stroke patients have hemiparesis); change in vision, gait, or ability to speak or understand; or a sudden, severe headache. The acronym FAST (*f*acial weakness, *a*rm weakness, *s*peech abnormality, and *t*ime) is simple and helpful to teach to the lay public about the common physical symptoms of stroke and to underscore that treatments are highly time sensitive.

Other causes of sudden-onset neurologic symptoms that may mimic stroke include seizure, intracranial tumor, migraine, and metabolic encephalopathy. An adequate history from an observer that no convulsive activity occurred at the onset usually excludes seizure **(Chap. 436)**, although ongoing complex partial seizures without tonic-clonic activity can on occasion mimic stroke. Tumors **(Chap. 95)** may present with acute neurologic symptoms due to hemorrhage, seizure, or hydrocephalus. Surprisingly, migraine **(Chap. 441)** can mimic stroke, even in patients without a significant migraine history. When migraine develops without head pain (*acephalgic migraine*), the diagnosis can be especially difficult. Patients without any prior history of migraine may develop acephalgic migraine even after age 65. A sensory disturbance is often prominent, and the sensory deficit, as well as any motor deficits, tends to migrate slowly across a limb, over minutes rather than seconds as with stroke. The diagnosis of migraine becomes more secure as the cortical disturbance begins to cross vascular boundaries or if classic visual symptoms are present such as scintillating scotomata. At times, it may be impossible to make the diagnosis of migraine until there have been multiple episodes with no residual symptoms or signs and no changes on brain magnetic resonance imaging (MRI). Metabolic encephalopathies typically produce fluctuating mental status changes without focal neurologic findings. However, in the setting of prior stroke or brain injury, a patient with

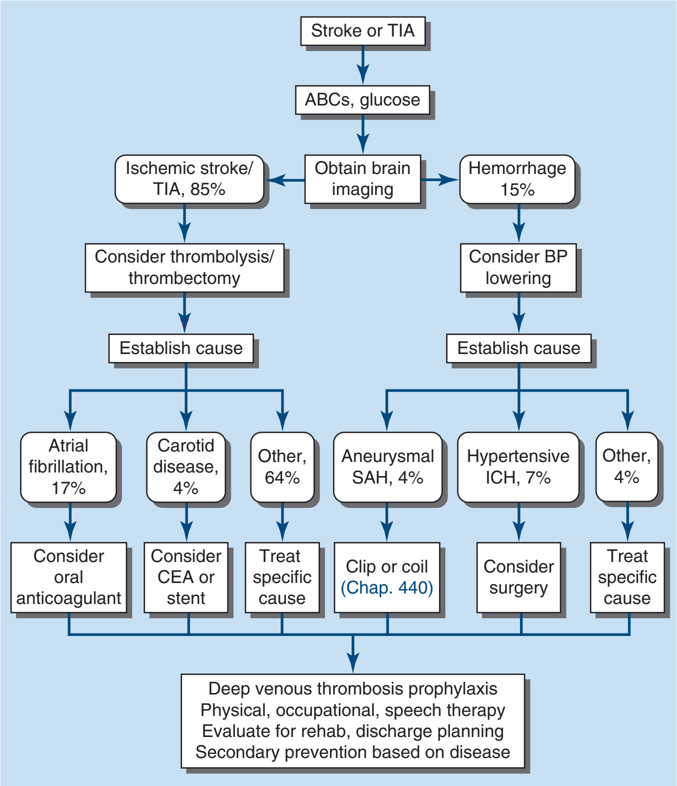

FIGURE 437-1 Medical management of stroke and TIA. *Rounded boxes* are diagnoses; *rectangles* are interventions. Numbers are percentages of stroke overall. ABCs, airway, breathing, circulation; BP, blood pressure; CEA, carotid endarterectomy; ICH, intracerebral hemorrhage; SAH, subarachnoid hemorrhage; TIA, transient ischemic attack.

fever or sepsis may manifest a recurrent hemiparesis, which clears rapidly when the infection is treated. The metabolic process serves to "unmask" a prior deficit and is termed "stroke recrudescence."

Once the diagnosis of stroke is made, a brain imaging study is necessary to determine if the cause of stroke is ischemia or hemorrhage **(Fig. 437-1)**. Computed tomography (CT) imaging of the brain is the standard imaging modality to detect the presence or absence of intracranial hemorrhage (see "Imaging Studies," below). If the stroke is ischemic, administration of recombinant tissue plasminogen activator (rtPA) or endovascular mechanical thrombectomy may be beneficial in restoring cerebral perfusion **(Chap. 438)**. Medical management to reduce the risk of complications becomes the next priority, followed by plans for secondary prevention. For ischemic stroke, several strategies can reduce the risk of subsequent stroke in all patients, while other strategies are effective for patients with specific causes of stroke such as cardiac embolus and carotid atherosclerosis. For hemorrhagic stroke, aneurysmal subarachnoid hemorrhage (SAH) and hypertensive intracerebral hemorrhage are two important causes. **The treatment and prevention of hypertensive intracerebral hemorrhage are discussed in Chap. 439. SAH is discussed in Chap. 440.**

■ STROKE SYNDROMES

A careful history and neurologic examination can often localize the region of brain dysfunction; if this region corresponds to an arterial distribution, the possible causes responsible for the syndrome can be narrowed. This is of particular importance when the patient presents with a TIA and a normal examination. For example, if a patient develops language loss and a right homonymous hemianopia, a search for causes of left middle cerebral emboli should be performed. A finding of an isolated stenosis of the right internal carotid artery in that patient, for example, suggests an asymptomatic carotid stenosis, and the search for other causes of stroke should continue. The following sections describe the clinical findings of cerebral ischemia associated with cerebral vascular territories depicted in **Figs. 437-2 through 437-11**. Stroke syndromes are divided into (1) large-vessel stroke within the anterior circulation, (2) large-vessel stroke within the posterior circulation, and (3) small-vessel disease of either vascular bed.

Stroke within the Anterior Circulation The internal carotid artery and its branches compose the anterior circulation of the brain. These vessels can be occluded by intrinsic disease of the vessel (e.g., atherosclerosis or dissection) or by embolic occlusion from a proximal source as discussed above. Occlusion of each major intracranial vessel has distinct clinical manifestations.

MIDDLE CEREBRAL ARTERY Occlusion of the proximal middle cerebral artery (MCA) or one of its major branches is most often due to an embolus (artery-to-artery, cardiac, or of unknown source) rather than intracranial atherothrombosis. Atherosclerosis of the proximal MCA may cause distal emboli to the middle cerebral territory or, less commonly, may produce low-flow TIAs. Collateral formation via leptomeningeal vessels often prevents MCA stenosis from becoming symptomatic.

The cortical branches of the MCA supply the lateral surface of the hemisphere except for (1) the frontal pole and a strip along the superomedial border of the frontal and parietal lobes supplied by the anterior cerebral artery (ACA) and (2) the lower temporal and occipital pole convolutions supplied by the posterior cerebral artery (PCA) (Figs. 437-2–437-5).

The proximal MCA (M1 segment) gives rise to penetrating branches (termed *lenticulostriate arteries*) that supply the putamen, outer globus pallidus, posterior limb of the internal capsule, adjacent corona radiata, and most of the caudate nucleus (Fig. 437-2). In the sylvian fissure, the MCA in most patients divides into *superior* and *inferior* divisions (M2 branches). Branches of the inferior division supply the inferior parietal and temporal cortex, and those from the superior division supply the frontal and superior parietal cortex (Fig. 437-3).

If the entire MCA is occluded at its origin (blocking both its penetrating and cortical branches) and the distal collaterals are limited, the clinical findings are contralateral hemiplegia, hemianesthesia, homonymous hemianopia, and a day or two of gaze preference to the ipsilateral side. Dysarthria is common because of facial weakness. When the dominant hemisphere is involved, global aphasia is present also, and when the nondominant hemisphere is affected, anosognosia, constructional apraxia, and neglect are found **(Chap. 32)**.

Complete MCA syndromes occur most often when an embolus occludes the stem of the artery. Cortical collateral blood flow and differing arterial configurations are probably responsible for the development of many partial syndromes. Partial syndromes may also be due to emboli that enter the proximal MCA without complete occlusion, occlude distal MCA branches, or fragment and move distally.

Partial syndromes due to embolic occlusion of a single branch include hand, or arm and hand, weakness alone (brachial syndrome) or facial weakness with nonfluent (Broca) aphasia **(Chap. 32)**, with or without arm weakness (frontal opercular syndrome). A combination of sensory disturbance, motor weakness, and nonfluent aphasia suggests that an embolus has occluded the proximal superior division and infarcted large portions of the frontal and parietal cortices (Fig. 437-3). If a fluent (Wernicke's) aphasia occurs without weakness, the inferior division of the MCA supplying the posterior part (temporal cortex) of the dominant hemisphere is probably involved. Jargon speech and an inability to comprehend written and spoken language are prominent features, often accompanied by a contralateral, homonymous superior quadrantanopia. Hemineglect or spatial agnosia without weakness indicates that the inferior division of the MCA in the nondominant hemisphere is involved.

Occlusion of a lenticulostriate vessel produces small-vessel (lacunar) stroke within the internal capsule (Fig. 437-2). This produces pure motor stroke or sensory-motor stroke contralateral to the lesion. Ischemia within the genu of the internal capsule causes primarily facial weakness followed by arm and then leg weakness as the ischemia moves posterior within the capsule. Alternatively, the contralateral hand may become ataxic, and dysarthria will be prominent (clumsy hand, dysarthria lacunar syndrome). Lacunar infarction affecting the globus pallidus and putamen often has few clinical signs, but parkinsonism and hemiballismus have been reported.

ANTERIOR CEREBRAL ARTERY The ACA is divided into two segments: the precommunal (A1) circle of Willis, or stem, which connects

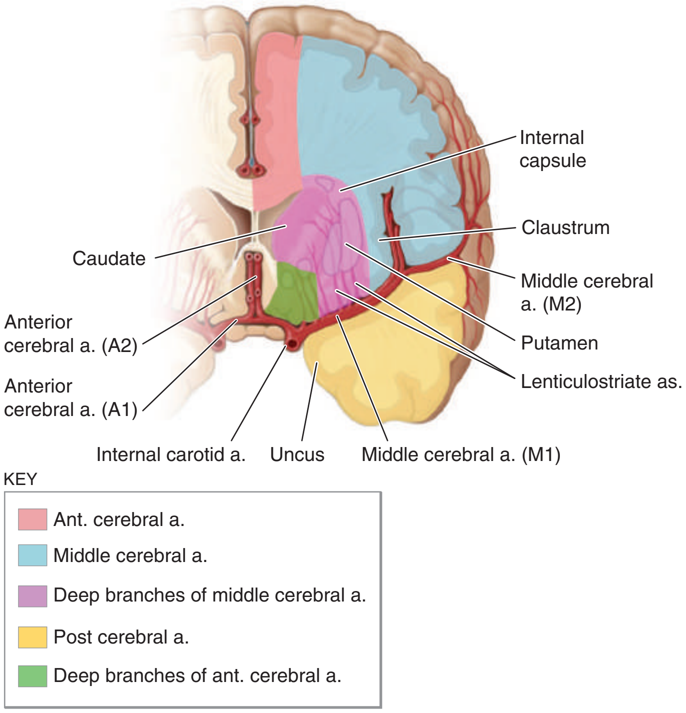
Internal capsule
Claustrum
Caudate
Middle cerebral a. (M2)
Anterior cerebral a. (A2)
Putamen
Anterior cerebral a. (A1)
Lenticulostriate as.
Internal carotid a.
Uncus
Middle cerebral a. (M1)
KEY
Ant. cerebral a.
Middle cerebral a.
Deep branches of middle cerebral a.
Post cerebral a.
Deep branches of ant. cerebral a.

FIGURE 437-2 Diagram of a cerebral hemisphere in coronal section showing the territories of the major cerebral vessels that branch from the internal carotid arteries.

the internal carotid artery to the anterior communicating artery, and the postcommunal (A2) segment distal to the anterior communicating artery (Figs. 437-2 and 437-4). The A1 segment gives rise to several deep penetrating branches that supply the anterior limb of the internal capsule, the anterior perforate substance, amygdala, anterior hypothalamus, and the inferior part of the head of the caudate nucleus.

Occlusion of the proximal ACA is usually well tolerated because of collateral flow through the anterior communicating artery and collaterals through the MCA and PCA. Occlusion of a single A2 segment results in the contralateral symptoms noted in Fig. 437-4. If both A2 segments arise from a single anterior cerebral stem (contralateral A1 segment atresia), the occlusion may affect both hemispheres. Profound abulia (a delay in verbal and motor response) and bilateral pyramidal signs with paraparesis or quadriparesis and urinary incontinence result.

ANTERIOR CHOROIDAL ARTERY This artery arises from the internal carotid artery and supplies the posterior limb of the internal capsule and the white matter posterolateral to it, through which pass some of the geniculocalcarine fibers (Fig. 437-5). The complete syndrome of anterior choroidal artery occlusion consists of contralateral hemiplegia, hemianesthesia (hypesthesia), and homonymous hemianopia. However, because this territory is also supplied by penetrating vessels of the proximal MCA and the posterior communicating and posterior choroidal arteries, minimal deficits may occur, and patients frequently recover substantially. Anterior choroidal strokes are usually the result of in situ thrombosis of the vessel, and the vessel is particularly vulnerable to iatrogenic occlusion during surgical clipping of aneurysms arising from the internal carotid artery.

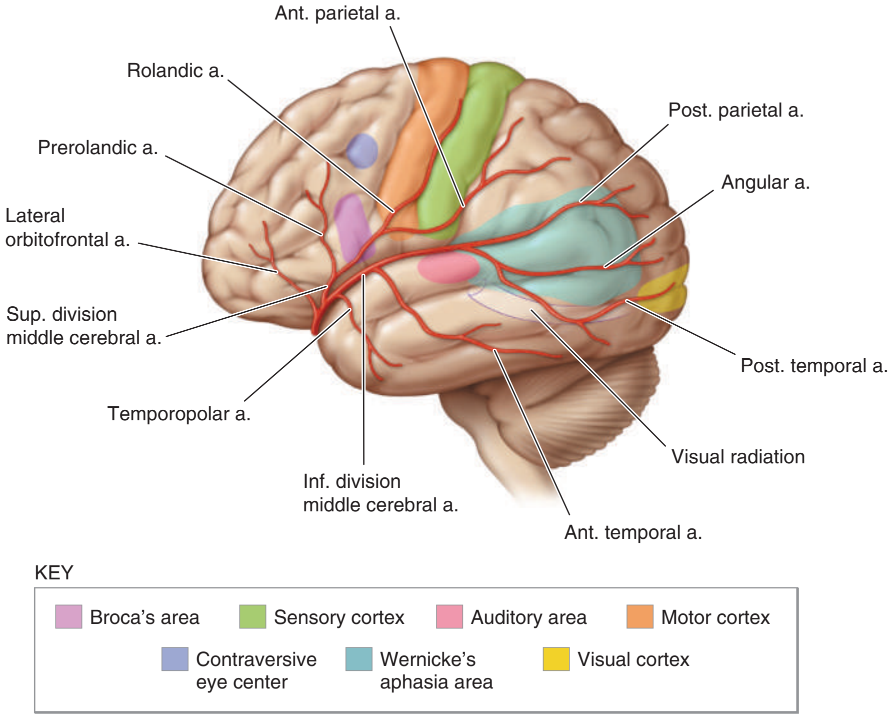
Ant. parietal a.
Rolandic a.
Post. parietal a.
Prerolandic a.
Angular a.
Lateral orbitofrontal a.
Sup. division middle cerebral a.
Post. temporal a.
Temporopolar a.
Visual radiation
Inf. division middle cerebral a.
Ant. temporal a.
KEY
Broca's area
Sensory cortex
Auditory area
Motor cortex
Contraversive eye center
Wernicke's aphasia area
Visual cortex

FIGURE 437-3 Diagram of a cerebral hemisphere, lateral aspect, showing the branches and distribution of the middle cerebral artery (MCA) and the principal regions of cerebral localization. Note the bifurcation of the MCA into a superior and inferior division.

Signs and symptoms: *Structures involved*

Paralysis of the contralateral face, arm, and leg; sensory impairment over the same area (pinprick, cotton touch, vibration, position, two-point discrimination, stereognosis, tactile localization, barognosis, cutaneographia): *Somatic motor area for face and arm and the fibers descending from the leg area to enter the corona radiata and corresponding somatic sensory system*

Motor aphasia: Motor speech area of the dominant hemisphere

Central aphasia, word deafness, anomia, jargon speech, sensory agraphia, acalculia, alexia, finger agnosia, right-left confusion (the last four comprise the Gerstmann syndrome): Central, suprasylvian speech area and parietooccipital cortex of the dominant hemisphere

Conduction aphasia: Central speech area (parietal operculum)

Apractagnosia of the nondominant hemisphere, anosognosia, hemiasomatognosia, unilateral neglect, agnosia for the left half of external space, dressing "apraxia," constructional "apraxia," distortion of visual coordinates, inaccurate localization in the half field, impaired ability to judge distance, upside-down reading, visual illusions (e.g., it may appear that another person walks through a table): *Nondominant parietal lobe (area corresponding to speech area in dominant hemisphere); loss of topographic memory is usually due to a nondominant lesion, occasionally to a dominant one*

Homonymous hemianopsia (or less frequently a superior quadrantanopsia due to isolated anterior temporal lobe infarction or inferior quadrantanopsia due to isolated parietal lobe infarction)

Paralysis of conjugate gaze to the opposite side: *Frontal contraversive eye field or projecting fibers.*

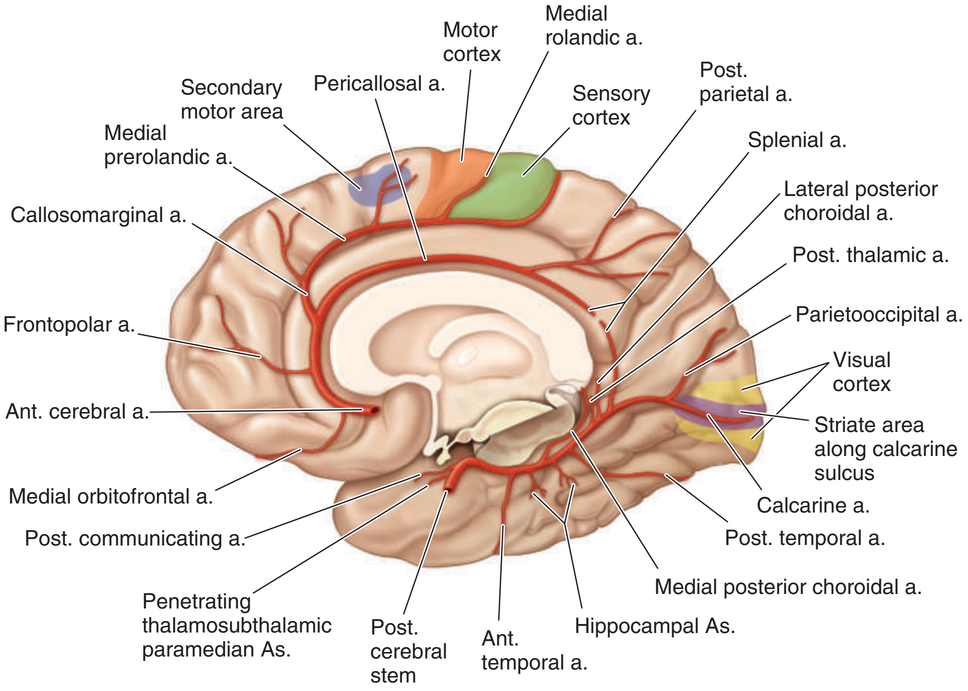

FIGURE 437-4 Diagram of a cerebral hemisphere, medial aspect, showing the branches and distribution of the anterior cerebral artery and the principal regions of cerebral localization.

Signs and symptoms: *Structures involved*

Paralysis of opposite foot and leg: *Motor leg area*

A lesser degree of paresis of opposite arm: *Arm area of cortex or fibers descending to corona radiata.*

Cortical sensory loss over toes, foot, and leg: *Sensory area for foot and leg*

Urinary incontinence: *Sensorimotor area in paracentral lobule*

Contralateral grasp reflex, sucking reflex, gegenhalten (paratonic rigidity): *Medial surface of the posterior frontal lobe; likely supplemental motor area*

Abulia (akinetic mutism), slowness, delay, intermittent interruption, lack of spontaneity, whispering, reflex distraction to sights and sounds: *Uncertain localization—probably cingulate gyrus and medial inferior portion of frontal, parietal, and temporal lobes*

Impairment of gait and stance (gait apraxia): *Frontal cortex near leg motor area*

Dyspraxia of left limbs, tactile aphasia in left limbs: *Corpus callosum*

INTERNAL CAROTID ARTERY The clinical picture of internal carotid occlusion varies depending on whether the cause of ischemia is propagated thrombus, embolism, or low flow. The cortex supplied by the MCA territory is affected most often. With a competent circle of Willis, occlusion may go unnoticed. If the thrombus propagates up the internal carotid artery into the MCA or embolizes it, symptoms are identical to proximal MCA occlusion (see above). Sometimes there is massive infarction of the entire deep white matter and cortical surface. When the origins of both the ACA and MCA are occluded at the top of the carotid artery, abulia or stupor occurs with hemiplegia, hemianesthesia, and aphasia or anosognosia. When the PCA arises from the internal carotid artery (a configuration called a *fetal PCA*), it may also become occluded and give rise to symptoms referable to its peripheral territory (Figs. 437-4 and 437-5).

In addition to supplying the ipsilateral brain, the internal carotid artery perfuses the optic nerve and retina via the ophthalmic artery. In ~25% of symptomatic internal carotid disease, recurrent transient monocular blindness (amaurosis fugax) warns of the lesion. Patients typically describe a horizontal shade that sweeps down or up across the field of vision. They may also complain that their vision was blurred in that eye or that the upper or lower half of vision disappeared. In most cases, these symptoms last only a few minutes. Rarely, ischemia or infarction of the ophthalmic artery or central retinal arteries occurs at the time of cerebral TIA or infarction.

A high-pitched prolonged carotid bruit fading into diastole is often associated with tightly stenotic lesions. As the stenosis grows tighter and flow distal to the stenosis becomes reduced, the bruit becomes fainter and may disappear when occlusion is imminent.

COMMON CAROTID ARTERY All symptoms and signs of internal carotid occlusion may also be present with occlusion of the common carotid artery. Jaw claudication may result from low flow in the external carotid branches. Bilateral common carotid artery occlusions at their origin may occur in Takayasu's arteritis **(Chap. 375)**.

Stroke within the Posterior Circulation The posterior circulation is composed of the paired vertebral arteries, the basilar artery, and the paired PCAs. The vertebral arteries join to form the basilar artery at the pontomedullary junction. The basilar artery divides into two PCAs in the interpeduncular fossa (Figs. 437-4–437-6). These major arteries give rise to long and short circumferential branches and to smaller deep penetrating branches that supply the cerebellum, medulla, pons, midbrain, subthalamus, thalamus, hippocampus, and medial temporal and occipital lobes. Occlusion of each vessel produces its own distinctive syndrome.

POSTERIOR CEREBRAL ARTERY In 75% of cases, both PCAs arise from the bifurcation of the basilar artery; in 20%, one has its origin from the ipsilateral internal carotid artery via the posterior communicating artery; in 5%, both originate from the respective ipsilateral internal carotid arteries (Figs. 437-4–437-6). The precommunal, or P1, segment of the true PCA is atretic in such cases.

PCA syndromes usually result from atheroma formation or emboli that lodge at the top of the basilar artery; posterior circulation disease may also be caused by dissection of either vertebral artery or fibromuscular dysplasia.

Two clinical syndromes are commonly observed with occlusion of the PCA: (1) *P1 syndrome*: midbrain, subthalamic, and thalamic signs, which are due to disease of the proximal P1 segment of the PCA or its penetrating branches (thalamogeniculate, Percheron, and posterior choroidal arteries); and (2) *P2 syndrome*: cortical temporal and occipital lobe signs, due to occlusion of the P2 segment distal to the junction of the PCA with the posterior communicating artery.

P1 SYNDROMES Infarction usually occurs in the ipsilateral subthalamus and medial thalamus and in the ipsilateral cerebral peduncle and midbrain (Figs. 437-5 and 437-11). A third nerve palsy with contralateral ataxia (Claude's syndrome) or with contralateral hemiplegia (Weber's syndrome) may result. The ataxia indicates involvement of

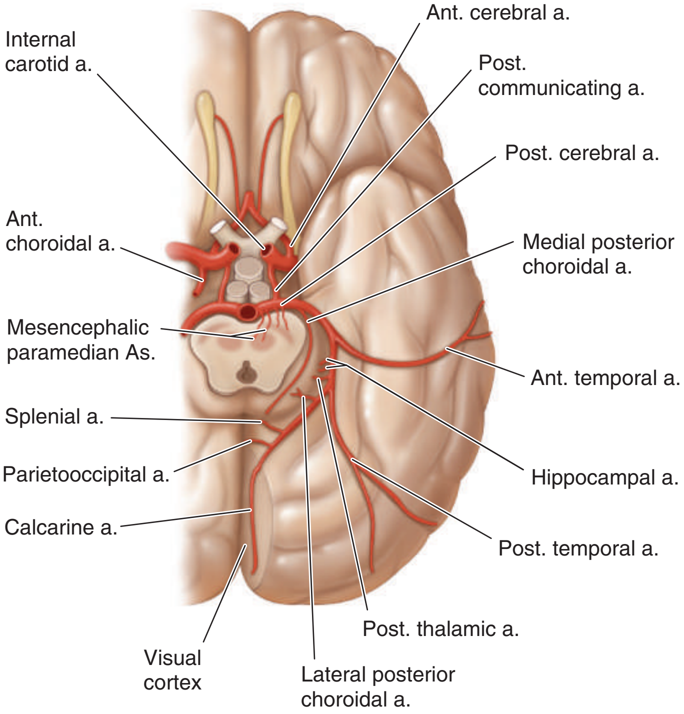

FIGURE 437-5 Inferior aspect of the brain with the branches and distribution of the posterior cerebral artery and the principal anatomic structures shown.

Signs and symptoms: *Structures involved*

Peripheral territory (see also Fig. 437-9). Homonymous hemianopia (often upper quadrantic): *Calcarine cortex or optic radiation nearby.* Bilateral homonymous hemianopia, cortical blindness, awareness or denial of blindness; tactile naming, achromatopia (color blindness), failure to see to-and-fro movements, inability to perceive objects not centrally located, apraxia of ocular movements, inability to count or enumerate objects, tendency to run into things that the patient sees and tries to avoid: *Bilateral occipital lobe with possibly the parietal lobe involved.* Verbal dyslexia without agraphia, color anomia: *Dominant calcarine lesion and posterior part of corpus callosum.* Memory defect: *Hippocampal lesion bilaterally or on the dominant side only.* Topographic disorientation and prosopagnosia: *Usually with lesions of nondominant, calcarine, and lingual gyrus.* Simultanagnosia, hemivisual neglect: *Dominant visual cortex, contralateral hemisphere.* Unformed visual hallucinations, peduncular hallucinosis, metamorphopsia, teleopsia, illusory visual spread, palinopsia, distortion of outlines, central photophobia: *Calcarine cortex.* Complex hallucinations: *Usually nondominant hemisphere.*

Central territory. Thalamic syndrome: sensory loss (all modalities), spontaneous pain and dysesthesias, choreoathetosis, intention tremor, spasms of hand, mild hemiparesis: *Posteroventral nucleus of thalamus; involvement of the adjacent subthalamus body or its afferent tracts.* Thalamoperforate syndrome: crossed cerebellar ataxia with ipsilateral third nerve palsy (Claude's syndrome): *Dentatothalamic tract and issuing third nerve.* Weber's syndrome: third nerve palsy and contralateral hemiplegia: *Third nerve and cerebral peduncle.* Contralateral hemiplegia: *Cerebral peduncle.* Paralysis or paresis of vertical eye movement, skew deviation, sluggish pupillary responses to light, slight miosis and ptosis (retraction nystagmus and "tucking" of the eyelids may be associated): *Supranuclear fibers to third nerve, interstitial nucleus of Cajal, nucleus of Darkschewitsch, and posterior commissure.* Contralateral rhythmic, ataxic action tremor; rhythmic postural or "holding" tremor (rubral tremor): *Dentatothalamic tract.*

the red nucleus or dentatorubrothalamic tract; the hemiplegia is localized to the cerebral peduncle (Fig. 437-11). If the subthalamic nucleus is involved, contralateral hemiballismus may occur. Occlusion of the artery of Percheron produces paresis of upward gaze and drowsiness and often abulia. Extensive infarction in the midbrain and subthalamus occurring with bilateral proximal PCA occlusion presents as coma, unreactive pupils, bilateral pyramidal signs, and decerebrate rigidity.

Occlusion of the penetrating branches of thalamic and thalamogeniculate arteries produces less extensive thalamic and thalamocapsular lacunar syndromes. The *thalamic Déjérine-Roussy syndrome* consists of contralateral hemisensory loss followed later by an agonizing, searing, or burning pain in the affected areas. It is persistent and responds poorly to analgesics. Anticonvulsants (carbamazepine or gabapentin) or tricyclic antidepressants may be beneficial.

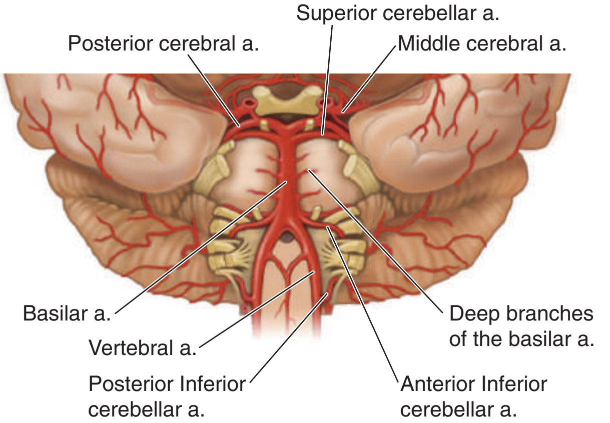
Superior cerebellar a.
Posterior cerebral a.
Middle cerebral a.
Basilar a.
Deep branches of the basilar a.
Vertebral a.
Posterior Inferior cerebellar a.
Anterior Inferior cerebellar a.

FIGURE 437-6 Diagram of the posterior circulation, showing the intracranial vertebral arteries forming the basilar artery that gives off the anterior inferior cerebellar, superior cerebellar, and posterior cerebral arteries. The posterior inferior cerebellar artery arises from each of the vertebral segments. The majority of brainstem blood flow arises from numerous deep branches of the basilar artery that penetrate directly into the brainstem.

P2 SYNDROMES (Figs. 437-4 and 437-5) Occlusion of the distal PCA causes infarction of the medial temporal and occipital lobes. Contralateral homonymous hemianopia without macula sparing is the usual manifestation. (MCA strokes often produce hemianopia but typically spare the macula as calcarine cortex is perfused by the P2 segment.) Occasionally, only the upper quadrant of visual field is involved or the macula vision is spared. If the visual association areas are spared and only the calcarine cortex is involved, the patient may be aware of visual defects. Medial temporal lobe and hippocampal involvement may cause an acute disturbance in memory, particularly if it occurs in the dominant hemisphere. The defect usually clears because memory has bilateral representation. If the dominant hemisphere is affected and the infarct extends to involve the splenium of the corpus callosum, the patient may demonstrate alexia without agraphia. Visual agnosia for faces, objects, mathematical symbols, and colors and anomia with paraphasic errors (amnestic aphasia) may also occur, even without callosal involvement. Occlusion of the PCA can produce *peduncular hallucinosis* (visual hallucinations of brightly colored scenes and objects).

Bilateral infarction in the distal PCAs produces cortical blindness (blindness with preserved pupillary light reaction). The patient is often unaware of the blindness or may even deny it (*Anton's syndrome*). Tiny islands of vision may persist, and the patient may report that vision fluctuates as images are captured in the preserved portions. Rarely, only peripheral vision is lost and central vision is spared, resulting in "gun-barrel" vision. Bilateral visual association area lesions may result in *Balint's syndrome*, a disorder of the orderly visual scanning of the environment **(Chap. 32)**, usually resulting from infarctions secondary to low flow in the "watershed" between the distal PCA and MCA territories, as occurs after cardiac arrest. Patients may experience persistence of a visual image for several minutes despite gazing at another scene (*palinopsia*) or an inability to synthesize the whole of an image (*asimultanagnosia*). Embolic occlusion of the top of the basilar artery can produce any or all the central or peripheral territory symptoms. The hallmark is the sudden onset of bilateral signs, including ptosis, pupillary asymmetry or lack of reaction to light, and somnolence. Patients will often have posturing and myoclonic jerking that simulates seizure. Interrogation of the noncontrast CT scan for a hyperdense basilar artery sign (indicating thrombus in the basilar artery) or CT angiography (CTA) establishes this diagnosis. Physicians should be suspicious of this rare but potentially treatable stroke syndrome in the setting of presumed new-onset seizure and cranial nerve deficits.

VERTEBRAL AND POSTERIOR INFERIOR CEREBELLAR ARTERIES The vertebral artery, which arises from the innominate artery on the right and the subclavian artery on the left, consists of four segments. The first (V1) extends from its origin to its entrance into the sixth or fifth transverse vertebral foramen. The second segment (V2) traverses the vertebral foramina from C6 to C2. The third (V3) passes through the transverse foramen and circles around the arch of the atlas to pierce the dura at the foramen magnum. The fourth (V4) segment courses

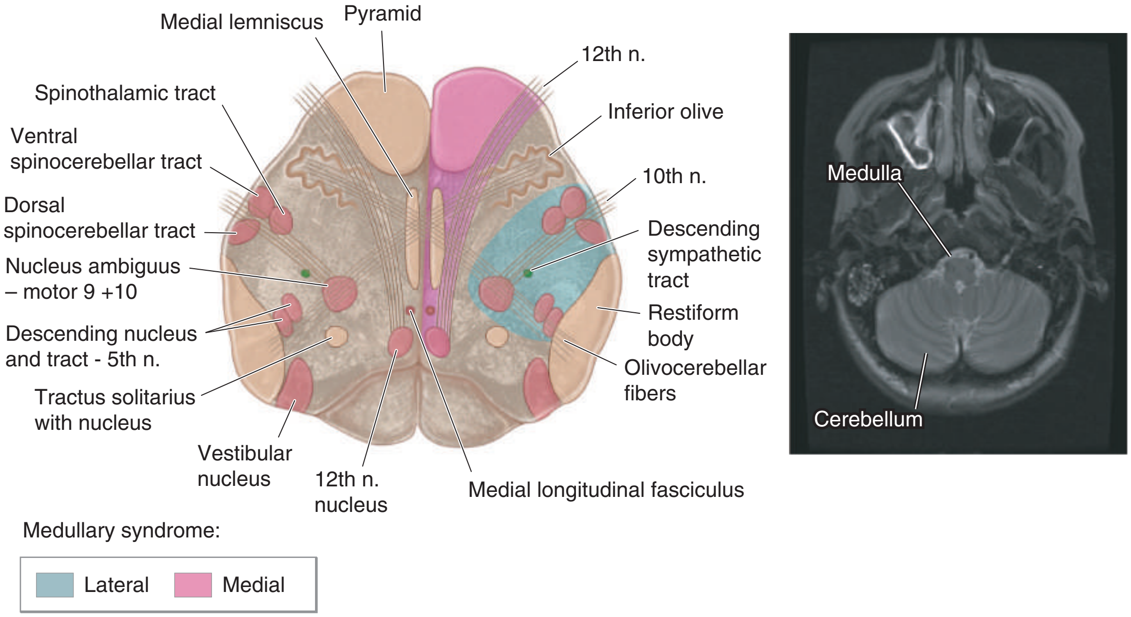
Medial lemniscus
Pyramid
12th n.
Spinothalamic tract
Inferior olive
Ventral spinocerebellar tract
10th n.
Dorsal spinocerebellar tract
Descending sympathetic tract
Nucleus ambiguus – motor 9 +10
Restiform body
Descending nucleus and tract - 5th n.
Olivocerebellar fibers
Tractus solitarius with nucleus
Vestibular nucleus
12th n. nucleus
Medial longitudinal fasciculus
Medulla
Cerebellum
Medullary syndrome:
Lateral
Medial

FIGURE 437-7 Axial section at the level of the medulla, depicted schematically on the left, with a corresponding magnetic resonance image on the right. Note that in Figs. 437-7 through 437-11, all drawings are oriented with the dorsal surface at the bottom, matching the orientation of the brainstem that is commonly seen in all modern neuroimaging studies. Approximate regions involved in medial and lateral medullary stroke syndromes are shown.

Signs and symptoms: *Structures involved*

1. Medial medullary syndrome (occlusion of vertebral artery or of branch of vertebral or lower basilar artery)

On side of lesion

Paralysis with atrophy of one-half half the tongue: *Ipsilateral twelfth nerve*

Paralysis of arm and leg, sparing face; impaired tactile and proprioceptive sense over one-half the body: *Contralateral pyramidal tract and medial lemniscus*

Pain, numbness, impaired sensation over one-half the face: *Descending tract and nucleus fifth nerve*

Ataxia of limbs, falling to side of lesion: *Uncertain—restiform body, cerebellar hemisphere, cerebellar fibers, spinocerebellar tract (?)*

Nystagmus, diplopia, oscillopsia, vertigo, nausea, vomiting: *Vestibular nucleus*

Horner's syndrome (miosis, ptosis, decreased sweating): *Descending sympathetic tract*

Dysphagia, hoarseness, paralysis of palate, paralysis of vocal cord, diminished gag reflex: *Issuing fibers ninth and tenth nerves*

Loss of taste: *Nucleus and tractus solitarius*

Numbness of ipsilateral arm, trunk, or leg: *Cuneate and gracile nuclei*

Weakness of lower face: *Genuflected upper motor neuron fibers to ipsilateral facial nucleus*

Impaired pain and thermal sense over half the body, sometimes face: *Spinothalamic tract*

5. Basilar artery syndrome (the syndrome of the lone vertebral artery is equivalent): A combination of the various brainstem syndromes plus those arising in the posterior cerebral artery distribution.

Bilateral long tract signs (sensory and motor; cerebellar and peripheral cranial nerve abnormalities): *Bilateral long tract; cerebellar and peripheral cranial nerves*

Paralysis or weakness of all extremities, plus all bulbar musculature: *Corticobulbar and corticospinal tracts bilaterally*

upward to join the other vertebral artery to form the basilar artery (Fig. 437-6); only the fourth segment gives rise to branches that supply the brainstem and cerebellum. The posterior inferior cerebellar artery (PICA) in its proximal segment supplies the lateral medulla and, in its distal branches, the inferior surface of the cerebellum.

Atherothrombotic lesions have a predilection for V1 and V4 segments of the vertebral artery. The first segment may become diseased at the origin of the vessel and may produce posterior circulation emboli; collateral flow from the contralateral vertebral artery or the ascending cervical, thyrocervical, or occipital arteries is usually sufficient to prevent low-flow TIAs or stroke. When one vertebral artery is atretic and an atherothrombotic lesion threatens the origin of the other, the collateral circulation, which may also include retrograde flow down the basilar artery, is often insufficient (Figs. 437-5 and 437-6). In this setting, low-flow TIAs may occur, consisting of syncope, vertigo, and alternating hemiplegia; this state also sets the stage for thrombosis. Disease of the distal fourth segment of the vertebral artery can promote thrombus formation manifest as embolism or with propagation as basilar artery thrombosis. Stenosis proximal to the origin of the PICA can threaten the lateral medulla and posterior inferior surface of the cerebellum.

If the subclavian artery is occluded proximal to the origin of the vertebral artery, there is a reversal in the direction of blood flow in the ipsilateral vertebral artery. Exercise of the ipsilateral arm may increase demand on vertebral flow, producing posterior circulation TIAs, or "subclavian steal."

Although atheromatous disease rarely narrows the second and third segments of the vertebral artery, this region is subject to dissection, fibromuscular dysplasia, and, rarely, encroachment by osteophytic spurs within the vertebral foramina.

Embolic occlusion or thrombosis of a V4 segment causes ischemia of the lateral medulla. The constellation of vertigo, numbness of the ipsilateral face and contralateral limbs, diplopia, hoarseness, dysarthria, dysphagia, and ipsilateral Horner's syndrome is called the *lateral medullary (or Wallenberg's) syndrome* (Fig. 437-7). Ipsilateral upper motor neuron facial weakness can also occur. Most cases result from ipsilateral vertebral artery occlusion; in the remainder, PICA occlusion is responsible. Occlusion of the medullary penetrating branches of the vertebral artery or PICA results in partial syndromes. *Hemiparesis is not a typical feature of vertebral artery occlusion; however, quadriparesis may result from occlusion of the anterior spinal artery.*

Rarely, a *medial medullary syndrome* occurs with infarction of the pyramid and contralateral hemiparesis of the arm and leg, sparing the face. If the medial lemniscus and emerging hypoglossal nerve fibers are involved, contralateral loss of joint position sense and ipsilateral tongue weakness occur.

Cerebellar infarction can lead to *respiratory arrest* due to brainstem herniation from cerebellar swelling, closure of the aqueduct of Silvius or fourth ventricle, followed by hydrocephalus and central herniation. This added downward displacement of the brainstem from hydrocephalus will exacerbate respiratory and hemodynamic instability. Drowsiness, Babinski signs, dysarthria, and bifacial weakness may be absent, or present only briefly, before respiratory arrest ensues. Gait unsteadiness, headache, dizziness, nausea, and vomiting may be the only early symptoms and signs and should arouse suspicion of this impending complication, which may require neurosurgical decompression, often with an excellent outcome. Separating these symptoms from those of

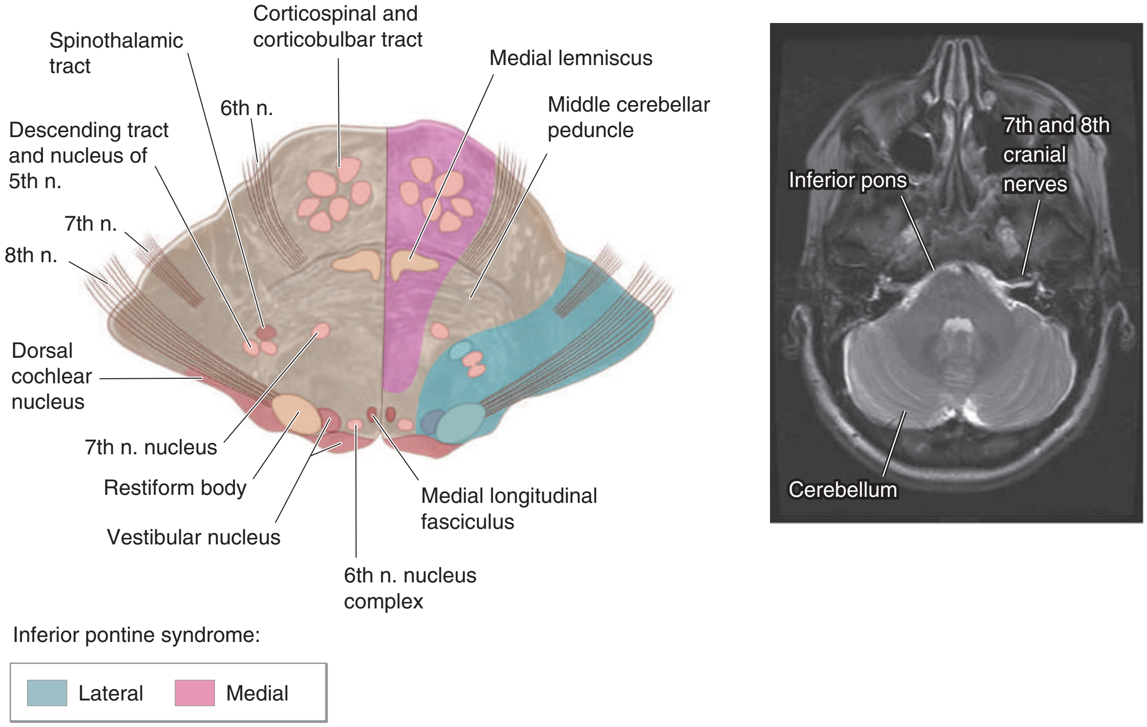
Corticospinal and corticobulbar tract
Spinothalamic tract
Medial lemniscus
Middle cerebellar peduncle
6th n.
Descending tract and nucleus of 5th n.
7th n.
8th n.
Dorsal cochlear nucleus
7th n. nucleus
Restiform body
Vestibular nucleus
Medial longitudinal fasciculus
6th n. nucleus complex
Inferior pontine syndrome:
Lateral
Medial
7th and 8th cranial nerves
Inferior pons
Cerebellum

FIGURE 437-8 Axial section at the level of the inferior pons, depicted schematically on the left, with a corresponding magnetic resonance image on the right. Approximate regions involved in medial and lateral inferior pontine stroke syndromes are shown.

Signs and symptoms: *Structures involved*
Paralysis of conjugate gaze to side of lesion (preservation of convergence): *Center for conjugate lateral gaze*
Nystagmus: *Vestibular nucleus*
Ataxia of limbs and gait: Likely *middle cerebellar peduncle*
Diplopia on lateral gaze: *Abducens nerve*
Paralysis of face, arm, and leg: *Corticobulbar and corticospinal tract in lower pons*
Impaired tactile and proprioceptive sense over one-half of the body: *Medial lemniscus*
Horizontal and vertical nystagmus, vertigo, nausea, vomiting, oscillopsia: *Vestibular nerve or nucleus*
Facial paralysis: *Seventh nerve*
Paralysis of conjugate gaze to side of lesion: *Center for conjugate lateral gaze*
Deafness, tinnitus: *Auditory nerve or cochlear nucleus*
Ataxia: *Middle cerebellar peduncle and cerebellar hemisphere*
Impaired sensation over face: *Descending tract and nucleus fifth nerve*
Impaired pain and thermal sense over one-half the body (may include face): *Spinothalamic tract*

viral labyrinthitis can be a challenge, but headache, neck stiffness, and unilateral dysmetria favor stroke.

BASILAR ARTERY Branches of the basilar artery (Fig. 437-6) supply the base of the pons and superior cerebellum and fall into three groups: (1) paramedian, 7–10 in number, which supply a wedge of pons on either side of the midline; (2) short circumferential, 5–7 in number, that supply the lateral two-thirds of the pons and middle and superior cerebellar peduncles; and (3) bilateral long circumferential (superior cerebellar and anterior inferior cerebellar arteries), which course around the pons to supply the cerebellar hemispheres.

Atheromatous lesions can occur anywhere along the basilar trunk but are most frequent in the proximal basilar and distal vertebral segments. Typically, lesions occlude either the proximal basilar and one or both vertebral arteries. The clinical picture varies depending on the availability of retrograde collateral flow from the posterior communicating arteries. Rarely, dissection of a vertebral artery may involve the basilar artery and, depending on the location of true and false lumen, may produce multiple penetrating artery strokes.

Although atherothrombosis occasionally occludes the distal portion of the basilar artery, emboli from the heart or proximal vertebral or basilar segments are more commonly responsible for "top of the basilar" syndromes.

Because the brainstem contains many structures in close apposition, a diversity of clinical syndromes may emerge with ischemia, reflecting involvement of the corticospinal and corticobulbar tracts, ascending sensory tracts, and cranial nerve nuclei (Figs. 437-7 to 437-11).

The symptoms of transient ischemia or infarction in the territory of the basilar artery often do not indicate whether the basilar artery itself or one of its branches is diseased, yet this distinction has important implications for therapy. *The picture of complete basilar occlusion, however, is easy to recognize as a constellation of bilateral long tract signs (sensory and motor) with signs of cranial nerve and cerebellar dysfunction*. Patients may have spontaneous posturing movements that are myoclonic in nature and simulate seizure activity. These movements are brief, repetitive, and multifocal and often confused with status epilepticus. CT or magnetic resonance angiography can rapidly detect basilar thrombosis, and rapid treatment (thrombectomy) can be lifesaving. A "locked-in" state of preserved consciousness with quadriplegia and cranial nerve signs suggests complete pontine and lower midbrain infarction. The therapeutic goal is to identify *impending* basilar occlusion before devastating infarction occurs. A series of TIAs and a slowly progressive, fluctuating stroke are extremely significant, because they often herald an atherothrombotic occlusion of the distal vertebral or proximal basilar artery.

TIAs in the proximal basilar distribution may produce vertigo (often described by patients as "swimming," "swaying," "moving," "unsteadiness," or "light-headedness"). Other symptoms that warn of basilar thrombosis include diplopia, dysarthria, facial or circumoral numbness, and hemisensory symptoms. In general, symptoms of basilar branch TIAs affect one side of the brainstem, whereas symptoms of basilar artery TIAs usually affect both sides, although a "herald" hemiparesis has been emphasized as an initial symptom of basilar occlusion. Most often, TIAs, whether due to impending occlusion of the basilar artery or a basilar branch, are short lived (5–30 min) and repetitive, occurring several times a day. The pattern suggests intermittent reduction of flow. Although treatment with intravenous heparin or various

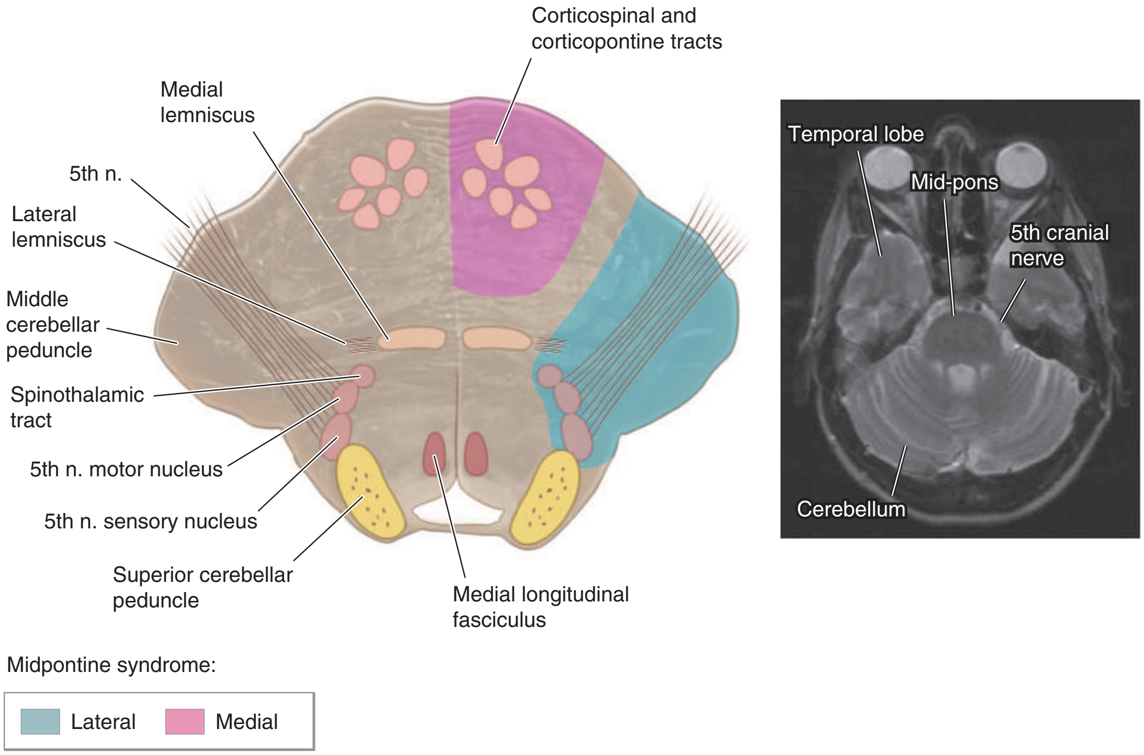

FIGURE 437-9 Axial section at the level of the midpons, depicted schematically on the left, with a corresponding magnetic resonance image on the right. Approximate regions involved in medial and lateral midpontine stroke syndromes are shown.

Signs and symptoms: *Structures involved*
Ataxia of limbs and gait (more prominent in bilateral involvement): *Pontine nuclei*
Paralysis of face, arm, and leg: *Corticobulbar and corticospinal tract*
Variable impaired touch and proprioception when lesion extends posteriorly: *Medial lemniscus*
Ataxia of limbs: *Middle cerebellar peduncle*
Paralysis of muscles of mastication: *Motor fibers or nucleus of fifth nerve*
Impaired sensation over side of face: *Sensory fibers or nucleus of fifth nerve*
Impaired pain and thermal sense on limbs and trunk: *Spinothalamic tract*

combinations of antiplatelet agents has been used to prevent clot propagation, there is no specific evidence to support any one approach, and endovascular intervention is also an option.

Atherothrombotic occlusion of the basilar artery with infarction usually causes *bilateral* brainstem signs. A gaze paresis or internuclear ophthalmoplegia associated with ipsilateral hemiparesis may be the only manifestation of bilateral brainstem ischemia. More often, unequivocal signs of bilateral pontine disease are present. Complete basilar thrombosis carries a high mortality.

Occlusion of a branch of the basilar artery usually causes *unilateral* symptoms and signs involving motor, sensory, and cranial nerves. If symptoms remain unilateral, concern over pending basilar occlusion should be reduced.

Occlusion of the superior cerebellar artery results in severe ipsilateral cerebellar ataxia, nausea and vomiting, dysarthria, and contralateral loss of pain and temperature sensation over the extremities, body, and face (spino- and trigeminothalamic tract). Partial deafness, ataxic tremor of the ipsilateral upper extremity, Horner's syndrome, and palatal myoclonus may occur rarely. Partial syndromes occur frequently (Fig. 437-10). With large strokes, swelling and mass effects may compress the midbrain or produce hydrocephalus; these symptoms may evolve rapidly. Neurosurgical intervention may be lifesaving in such cases.

Occlusion of the anterior inferior cerebellar artery produces variable degrees of infarction because the size of this artery and the territory it supplies vary inversely with those of the PICA. The principal symptoms include (1) ipsilateral deafness, facial weakness, vertigo, nausea and vomiting, nystagmus, tinnitus, cerebellar ataxia, Horner's syndrome, and paresis of conjugate lateral gaze; and (2) contralateral loss of pain and temperature sensation. An occlusion close to the origin of the artery may cause corticospinal tract signs (Fig. 437-8).

Occlusion of one of the short circumferential branches of the basilar artery affects the lateral two-thirds of the pons and middle or superior cerebellar peduncle, whereas occlusion of one of the paramedian branches affects a wedge-shaped area on either side of the medial pons (Figs. 437-8–437-10).

■ IMAGING STUDIES

See also Chap. 434.

CT Scans CT radiographic images identify or exclude hemorrhage as the cause of stroke, and they identify extraparenchymal hemorrhages, neoplasms, abscesses, and other conditions masquerading as stroke. Brain CT scans obtained in the first several hours after an infarction generally show no abnormality **(Fig. 437-12*A*)**, and the infarct may not be seen reliably for 24–48 h. The decision to treat patients with IV plasminogen activators is based on the clinical diagnosis of stroke and a CT scan showing no hemorrhage. CT may fail to show small ischemic strokes in the posterior fossa because of bone artifact; small infarcts on the cortical surface may also be missed.

Contrast-enhanced CT scans add specificity by showing contrast enhancement of subacute infarcts and allow visualization of venous structures. Coupled with multidetector scanners, CT angiography can be performed with administration of IV iodinated contrast allowing visualization of the cervical and intracranial arteries, intracranial veins, and aortic arch in one imaging session. Carotid disease and intracranial vascular occlusions are readily identified with this method **(see Fig. 438-2)**. After an IV bolus of contrast, deficits in brain perfusion produced by vascular occlusion can also be demonstrated **(Fig. 437-12*D*)** and used to predict the region of infarcted brain and the brain at risk of further infarction (i.e., the ischemic penumbra, see "Pathophysiology of Ischemic Stroke" in **Chap. 438**). CT imaging is also sensitive for detecting SAH (although by itself does not rule it out), and CTA can readily

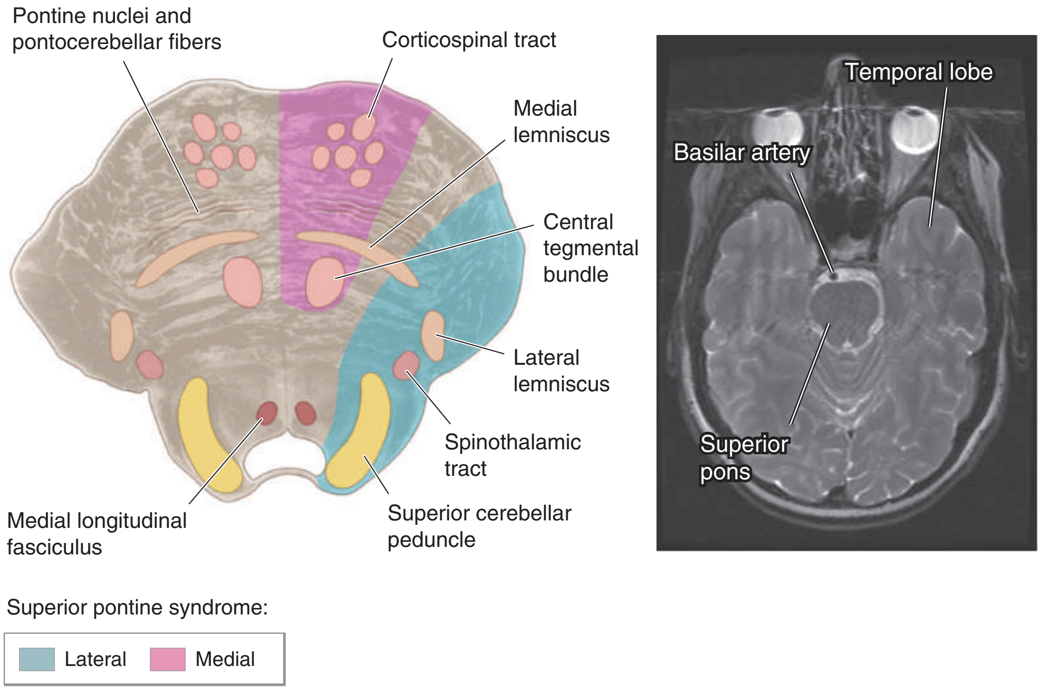
Pontine nuclei and pontocerebellar fibers
Corticospinal tract
Medial lemniscus
Central tegmental bundle
Lateral lemniscus
Spinothalamic tract
Superior cerebellar peduncle
Medial longitudinal fasciculus
Temporal lobe
Basilar artery
Superior pons
Superior pontine syndrome:
Lateral
Medial

FIGURE 437-10 Axial section at the level of the superior pons, depicted schematically on the left, with a corresponding magnetic resonance image on the right. Approximate regions involved in medial and lateral superior pontine stroke syndromes are shown.

Signs and symptoms: *Structures involved*

Cerebellar ataxia (probably): *Superior and/or middle cerebellar peduncle*

Internuclear ophthalmoplegia: *Medial longitudinal fasciculus*

Myoclonic syndrome, palate, pharynx, vocal cords, respiratory apparatus, face, oculomotor apparatus, etc.: *Localization uncertain—central tegmental bundle, dentate projection, inferior olivary nucleus*

Paralysis of face, arm, and leg: *Corticobulbar and corticospinal tract*

Rarely touch, vibration, and position are affected: *Medial lemniscus*

Ataxia of limbs and gait, falling to side of lesion: *Middle and superior cerebellar peduncles, superior surface of cerebellum, dentate nucleus*

Dizziness, nausea, vomiting; horizontal nystagmus: *Vestibular nucleus*

Paresis of conjugate gaze (ipsilateral): *Pontine contralateral gaze*

Skew deviation: *Uncertain*

Miosis, ptosis, decreased sweating over face (Horner's syndrome): *Descending sympathetic fibers*

Tremor: Localization unclear—*Dentate nucleus, superior cerebellar peduncle*

Impaired pain and thermal sense on face, limbs, and trunk: *Spinothalamic tract*

Impaired touch, vibration, and position sense, more in leg than arm (there is a tendency to incongruity of pain and touch deficits): *Medial lemniscus (lateral portion)*

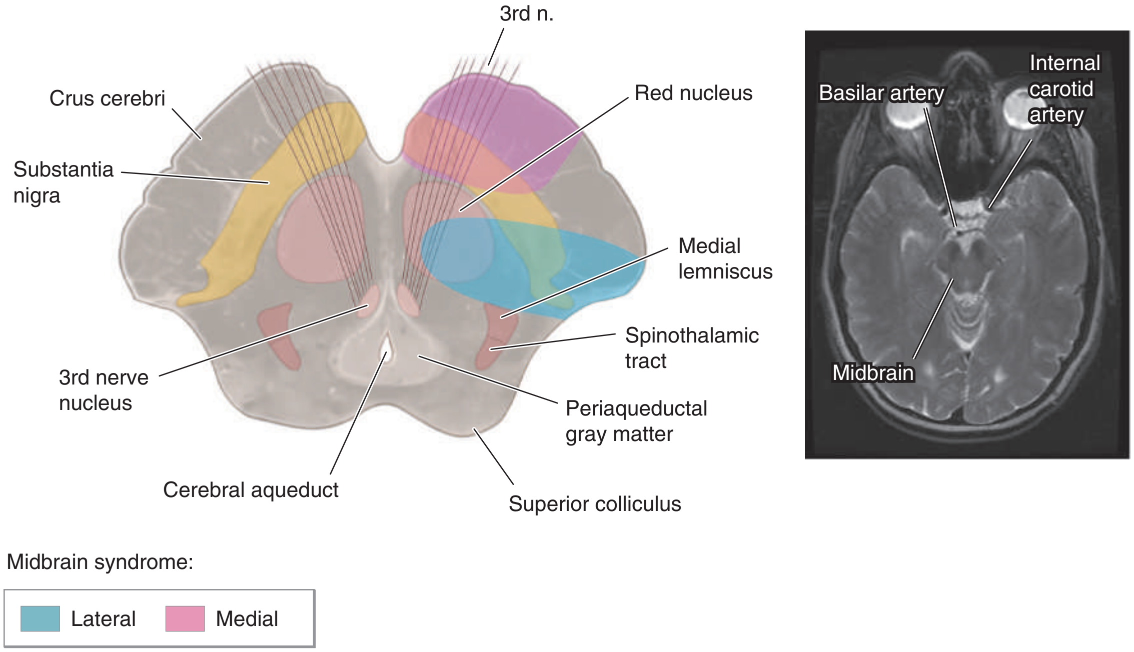
3rd n.
Crus cerebri
Red nucleus
Substantia nigra
Medial lemniscus
Spinothalamic tract
3rd nerve nucleus
Periaqueductal gray matter
Cerebral aqueduct
Superior colliculus
Basilar artery
Internal carotid artery
Midbrain
Midbrain syndrome:
Lateral
Medial

FIGURE 437-11 Axial section at the level of the midbrain, depicted schematically on the left, with a corresponding magnetic resonance image on the right. Approximate regions involved in medial and lateral midbrain stroke syndromes are shown.

Signs and symptoms: *Structures involved*

Eye "down and out" secondary to unopposed action of fourth and sixth cranial nerves, with dilated and unresponsive pupil: *Third nerve fibers*

Paralysis of face, arm, and leg: *Corticobulbar and corticospinal tract descending in crus cerebri*

Eye "down and out" secondary to unopposed action of fourth and sixth cranial nerves, with dilated and unresponsive pupil: *Third nerve fibers and/or third nerve nucleus*

Hemiataxia, hyperkinesias, tremor: *Red nucleus, dentatorubrothalamic pathway*

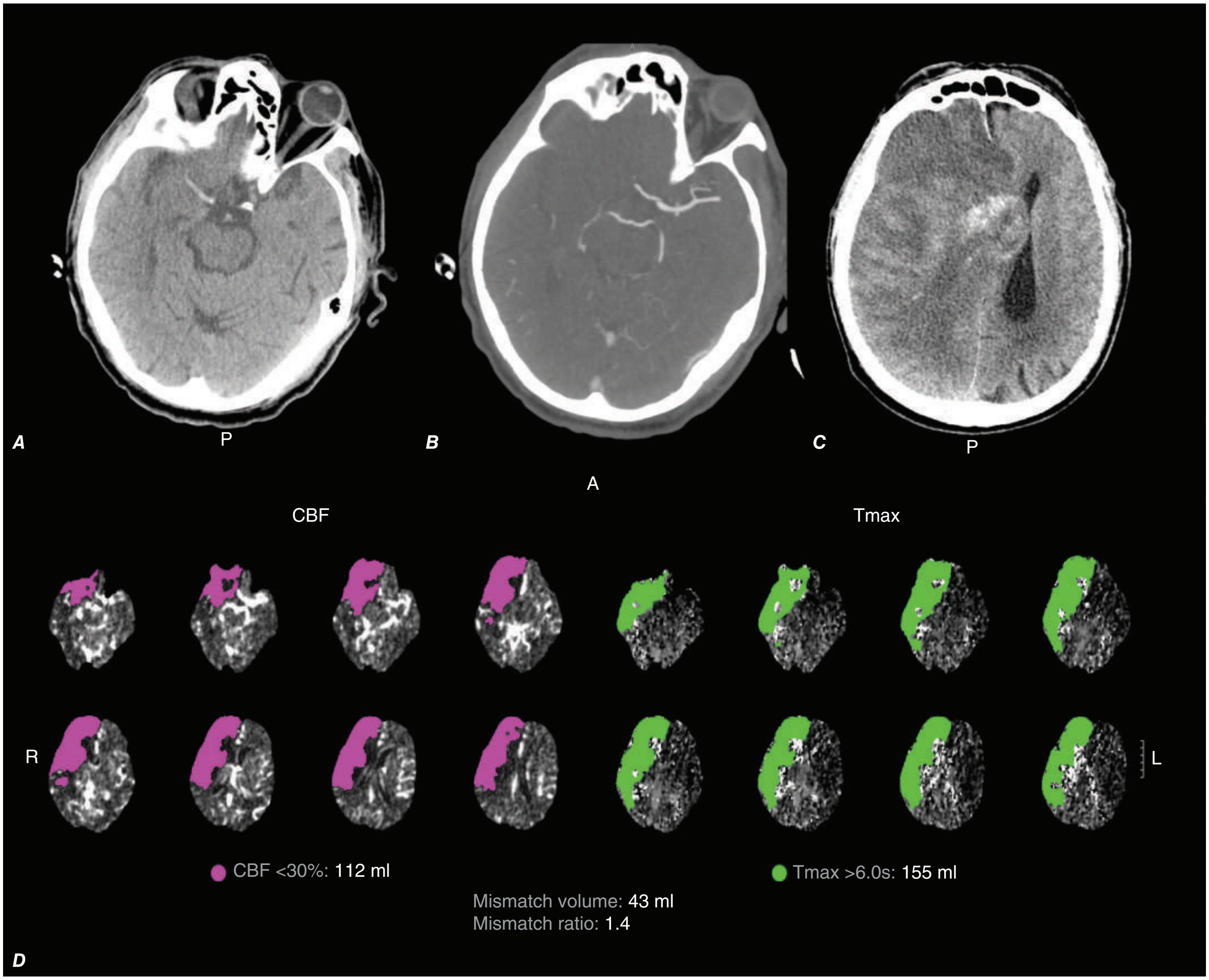

FIGURE 437-12 ***(A)*** Noncontrast head computed tomography (CT) image of an 83-year-old man with sudden onset left hemiplegia, left homonymous hemianopia, rightward gaze deviation and left hemineglect showing no clear brain infarction and hyperdensity within the right middle cerebral artery (MCA) suggestive of clot. ***(B)*** CT angiography performed at the same time as the CT showing absence of the right MCA and anterior cerebral artery (ACA) vessels consistent with occlusion of the bifurcation of the right intracranial internal carotid artery. ***(C)*** Follow-up head CT 1 day later showing extensive infarction of the right frontal lobe with brain herniation. ***(D)*** CT perfusion performed with studies shown in ***A*** and ***B***. This predicts a large core infarction (pink regions) despite attempts at revascularization with thrombectomy. CBF, cerebral blood flow.

identify intracranial aneurysms **(Chap. 440)**. Because of its speed and wide availability, noncontrast head CT is the imaging modality of choice in patients with acute stroke (Fig. 437-1), and CTA and CT perfusion imaging may also be useful and convenient adjuncts.

MRI

MRI reliably documents the extent and location of infarction in all areas of the brain, including the posterior fossa and cortical surface. Diffusion-weighted imaging (DWI) identifies regions of brain infarction within minutes of the stroke onset **(Fig. 437-13*A, B*)**, while fluid-attenuated inversion recovery (FLAIR) imaging reliably reveals areas of prior brain infarction from a few days to years later **(Fig. 437-13C)**. CT is poorly sensitive to brain infarction in the posterior fossae compared to MRI DWI images **(Fig. 437-13*D, E*)**. MRI also identifies intracranial hemorrhage and other abnormalities and, using special sequences, can be as sensitive as CT for detecting acute intracerebral hemorrhage. MRI scanners with magnets of higher field strength produce more reliable and precise images. Using IV administration of gadolinium contrast, magnetic resonance (MR) perfusion studies can be performed. Brain regions showing poor perfusion but no abnormality on diffusion provide, compared to CT, an equivalent measure of the ischemic penumbra. MR angiography is highly sensitive for stenosis of extracranial internal carotid arteries and of large intracranial vessels. With higher degrees of stenosis, MR angiography tends to overestimate the degree of stenosis when compared to conventional x-ray angiography. MRI with fat saturation is an imaging sequence used to visualize extra- or intracranial arterial dissection. This sensitive technique images clotted blood within the dissected vessel wall. Iron-sensitive imaging (ISI) is helpful to detect cerebral microbleeds that may be present in cerebral amyloid angiopathy and other hemorrhagic disorders.

MRI is more expensive and time consuming than CT and less readily available. Claustrophobia and the logistics of imaging acutely critically ill patients also limit its application. Most acute stroke protocols use CT because of these limitations. However, MRI is useful outside the acute period by more clearly defining the extent of tissue injury and discriminating new from old regions of brain infarction. MRI may have utility in patients with TIA, because it is also more likely to identify new infarction, which is a strong predictor of subsequent stroke.

Cerebral Angiography Conventional x-ray cerebral angiography is the gold standard for identifying and quantifying atherosclerotic stenoses of the cerebral arteries and for identifying and characterizing other pathologies, including aneurysms, vasospasm, intraluminal thrombi, fibromuscular dysplasia, arteriovenous fistulae, vasculitis, and collateral channels of blood flow. Conventional angiography carries risks of arterial damage, groin hemorrhage, embolic stroke, and

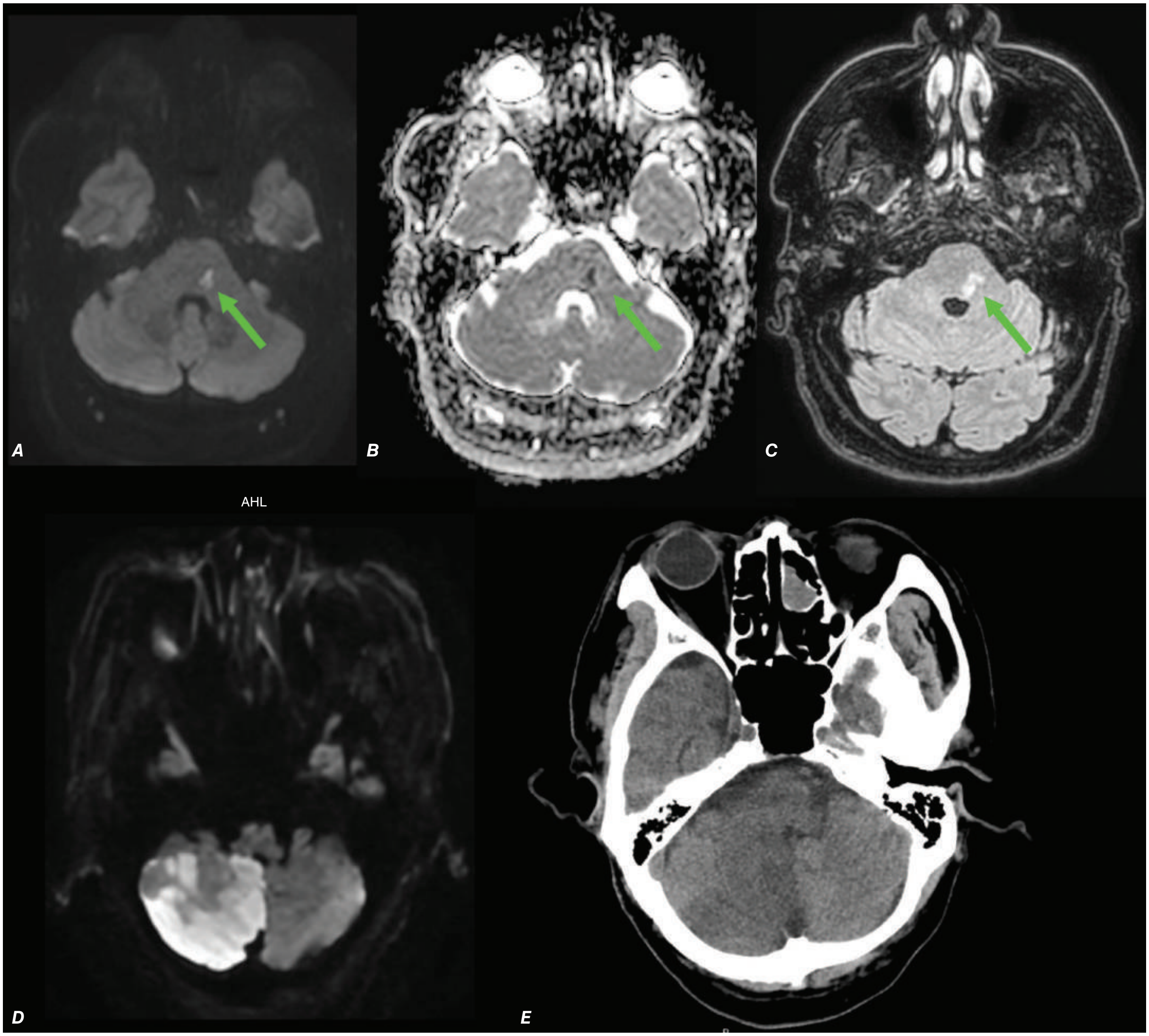

FIGURE 437-13 Examples of magnetic resonance imaging (MRI) imaging of acute ischemic infarcts. ***(A)*** Diffusion-weighted image (DWI) revealing bright region in the left pons (*arrow*), and ***(B)*** dark region on apparent diffusion coefficient (ADC) in the same region. Bright regions on DWI and corresponding dark regions on ADC are consistent with acute brain ischemia (within minutes to hours of stroke onset). (***C***) Same region of the brain 2 days later showing bright region in the infarcted tissue on fluid-attenuated inversion recovery (FLAIR) images. The DWI and ADC values will return to normal after 10–14 days while the FLAIR abnormality will persist long term. ***(D)*** DWI MRI and ***(E)*** computed tomography (CT) images of a patient with acute-onset vertigo. The DWI image shows an acute ischemic infarction of the right posterior inferior cerebellar artery territory within the cerebellum, whereas the CT scan is only subtly hypodense in the same region. MRI is superior to CT imaging for identifying ischemic infarction especially within the posterior fossa.

renal failure from contrast nephropathy, so it should be reserved for situations where less invasive means are inadequate. Acute stroke treatment with endovascular thrombectomy has proven effective in ischemic strokes caused by internal carotid terminus or MCA occlusions and is now part of routine clinical practice at centers that have this capability **(see Chap. 438)**.

Ultrasound Techniques Stenosis at the origin of the internal carotid artery can be identified and quantified reliably by ultrasonography that combines a B-mode ultrasound image with a Doppler ultrasound assessment of flow velocity ("duplex" ultrasound). Transcranial Doppler (TCD) assessment of MCA, ACA, and PCA flow and of vertebrobasilar flow is also useful. This latter technique can detect stenotic lesions in the large intracranial arteries because such lesions increase systolic flow velocity. TCD can also detect microemboli from otherwise asymptomatic carotid plaques. In many cases, MR angiography combined with carotid and transcranial ultrasound studies eliminates the need for conventional x-ray angiography in evaluating vascular stenosis. Alternatively, CTA of the entire head and neck can be performed during the initial imaging of acute stroke. Because this images the entire arterial system relevant to stroke, with the exception of the heart, much of the clinician's stroke workup can be completed with this single imaging study.

Radionuclide Perfusion Techniques Both xenon techniques (principally xenon-CT) and positron emission tomography (PET) can quantify cerebral blood flow. These tools are generally used for research **(Chap. 434)** but can be useful for determining the significance of arterial stenosis and planning for revascularization surgery. Single-photon emission computed tomography (SPECT) and CT or MR perfusion techniques report relative cerebral blood flow. As noted above, CT imaging is used as the initial imaging modality for acute stroke, and some centers combine both CTA and CT perfusion imaging together with the noncontrast CT scan. CT perfusion imaging increases the sensitivity for detecting ischemia and can measure the ischemic penumbra (Fig. 437-12). Alternatively, MR perfusion can be combined with MR diffusion imaging to identify the ischemic penumbra as the mismatch between these two imaging sequences.

FURTHER READING

Blumenfeld H: *Neuroanatomy Through Clinical Cases*, 3rd ed. New York, Sinauer Associates, 2020.

Tsao CW: Heart disease and stroke statistics-2023 update: A report from the American Heart Association. Circulation 147:e93, 2023.

438

Ischemic Stroke

Wade S. Smith, Anthony S. Kim, J. Claude Hemphill, III

The clinical diagnosis of stroke is discussed in **Chap. 437**. Once this diagnosis is made and either a noncontrast computed tomography (CT) scan or magnetic resonance imaging (MRI) scan has been performed, rapid reversal of ischemia is paramount. This chapter will focus on the stroke treatment timeline and subsequent secondary stroke prevention.

PATHOPHYSIOLOGY OF ISCHEMIC STROKE

Acute occlusion of an intracranial vessel causes reduction in blood flow to the brain region it supplies **(Fig. 438-1)**. The magnitude of flow reduction is a function of collateral blood flow, and this depends on an individual's vascular anatomy (which may be altered by disease), the site of occlusion, and systemic blood pressure. A decrease in cerebral blood flow to zero causes death of brain tissue (neuron cell bodies, dendrites, axons, and glial cells) within 4–10 min; values <16–18 mL/100 g tissue per minute cause infarction within an hour; and values <20 mL/100 g tissue per minute cause ischemia without infarction unless prolonged for several hours or days. If blood flow is restored to ischemic tissue before significant infarction develops, the patient may experience only transient symptoms, and the clinical syndrome is called a transient ischemic attack (TIA). Another important concept is the *ischemic penumbra*, defined as the ischemic but reversibly dysfunctional tissue surrounding a core area of infarction. The penumbra can be imaged by perfusion imaging using MRI or CT (see Fig. 438-3 and **Figs. 437-12** and **437-13**). The ischemic penumbra will eventually progress to infarction if no change in flow occurs, and hence, saving the ischemic penumbra is the goal of revascularization therapy. Restoration of blood flow provides oxygen and glucose to the penumbral tissue, preventing infarction not only by supplying fuel for metabolism but by reversing tissue acidosis, clearing glutamate and toxic oxygen species, and halting waves of cortical spreading depression emanating from the ischemic core that add metabolic stress to the tissue.

Ischemia causes a reduction in glucose and oxygen delivery, which in turn results in reduced capacity of cells to generate ATP. Without ATP, membrane ion pumps stop functioning and cells depolarize, allowing intracellular sodium and calcium to rise. Cellular depolarization also causes glutamate release from synaptic terminals and a failure of glutamate uptake by glial cells. The resulting sustained elevation in extracellular glutamate produces neurotoxicity by activating postsynaptic glutamate receptors that increase neuronal calcium influx and the production of reactive oxygen species. Reactive oxygen species damage DNA, lipid membranes, and likely other vital functions of cells. An innate immune response becomes apparent within a few hours after stroke, consisting of activation of proinflammatory microglia (the resident immune cells in brain) and infiltration of immune cells from the circulation. While important in repairing stroke damage, this acute inflammatory response may also contribute to tissue injury after stroke by release of proteases and reactive oxygen species. Ischemia and the postischemic inflammatory response also injure or destroy axons and dendrites at some distance from the infarct itself. Fever dramatically worsens brain injury during ischemia, as does hyperglycemia (glucose >11.1 mmol/L [200 mg/dL]), so it is reasonable to suppress fever and prevent hyperglycemia during and after brain ischemia. The value of induced mild hypothermia to improve stroke outcomes has not been clearly demonstrated and remains the subject of continuing clinical research.

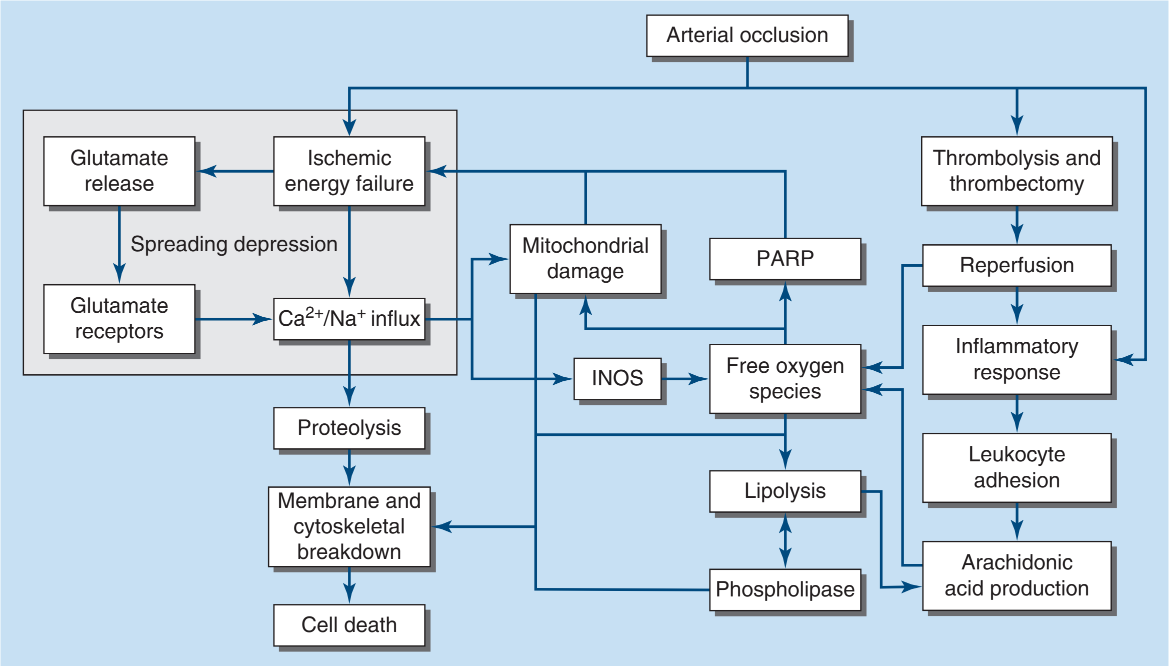

FIGURE 438-1 Major steps in the cascade of cerebral ischemia. See text for details. iNOS, inducible nitric oxide synthase; PARP, poly-A ribose polymerase.

TREATMENT

Acute Ischemic Stroke (Fig. 438-2)

After the clinical diagnosis of stroke is made **(Chap. 437)**, an orderly and prompt process of evaluation and treatment should follow. The first goal is to prevent or reverse brain injury. Attend to the patient's airway, breathing, and circulation (ABCs), and treat hypoglycemia or hyperglycemia if identified by finger stick testing. Perform an emergency noncontrast head CT scan to differentiate between ischemic stroke and hemorrhagic stroke **(Chap. 439)**; there are no reliable clinical findings that conclusively separate ischemia from hemorrhage, although a more depressed level of consciousness, higher initial blood pressure, or worsening of symptoms after onset favor hemorrhage, and a deficit that is maximal at onset, or remits, suggests ischemia. Treatments designed to reverse or lessen the amount of tissue infarction and improve clinical outcome fall within six categories: (1) medical support, (2) IV thrombolysis, (3) endovascular revascularization, (4) antithrombotic treatment, (5) neuroprotection, and (6) stroke centers and rehabilitation.

MEDICAL SUPPORT

When ischemic stroke occurs, the immediate goal is to optimize cerebral perfusion in the surrounding ischemic penumbra. Attention is also directed toward preventing the common complications of bedridden patients—infections (pneumonia, urinary, and skin) and deep-venous thrombosis (DVT) with pulmonary embolism. Subcutaneous heparin (unfractionated and low-molecular-weight) is safe and can be used concomitantly. Use of pneumatic compression stockings is of proven benefit in reducing risk of DVT and is a safe alternative to heparin.

Because collateral blood flow within the ischemic brain may be blood pressure dependent, there is controversy about whether blood pressure should be lowered acutely. Blood pressure should be reduced if it exceeds 220/120 mmHg, if there is malignant hypertension **(Chap. 288)** or concomitant myocardial ischemia, or if blood pressure is >185/110 mmHg and thrombolytic therapy is anticipated. When faced with the competing demands of myocardium and brain, lowering the heart rate with a β_1-adrenergic blocker (such as esmolol) can be a first step to decrease cardiac work and maintain blood pressure. Routine lowering of blood pressure below the limits listed above has the potential to worsen outcomes. Fever is detrimental and should be treated with antipyretics and surface cooling. Serum glucose should be monitored and kept <10.0 mmol/L (180 mg/dL), and above at least 3.3 mmol/L (60 mg/dL); a more intensive glucose control strategy does not improve outcome.

Between 5 and 10% of patients develop enough cerebral edema to cause obtundation and brain herniation. Edema peaks on the second or third day but can cause mass effect for ~10 days. The larger the infarct, the greater the likelihood that clinically significant edema will develop. Water restriction and IV mannitol

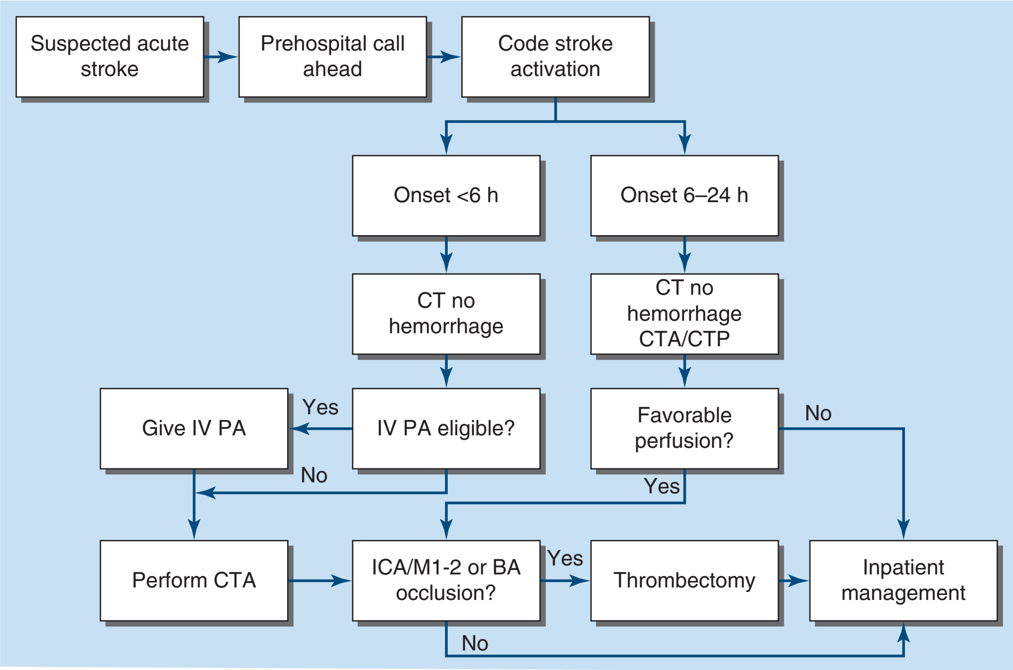

FIGURE 438-2 Management of acute stroke (pathway followed by the authors). For suspected stroke identified by prehospital professionals, we encourage calling ahead to the destination hospital. This allows early "stroke code" activation to prepare for an emergent computed tomography (CT) on arrival. For patients with onset <6 h from last time seen normal, we expedite a noncontrast head CT scan, and if free of hemorrhage and the patient is IV thrombolysis eligible (typically <4.5 h of last seen well time), this is administered in the CT scanner. (For IV tissue PA [tPA], the bolus is given and infusion initiated; for tenecteplase, the full dose is given as a bolus.) Then CT angiography (CTA) from left atrium to skull vertex is performed to identify an eligible target lesion for thrombectomy. For a patient presenting in the 6- to 24-h time window, thrombolysis is not considered, and the decision to perform thrombectomy is based on perfusion imaging.

Priorities of Acute Stroke Consultation: Once stroke is suspected, the first priorities are to assess airway and blood pressure, followed by establishing the time last seen normal. Patients with disabling neurologic deficits (particularly with National Institutes of Health Stroke Scale >5) may be eligible for thrombolytic and/or endovascular therapy. Based on the onset time, we follow the protocol shown in the figure. Following acute treatments, if any, we proceed with establishing the cause of the ischemic stroke. If atrial fibrillation is established or newly discovered, we favor use of apixaban 5 mg twice daily (or a reduced dose of 2.5 mg twice daily for impaired glomerular filtration rate) lifelong. If atrial fibrillation is not detected during the hospital encounter, we obtain an ambulatory electrocardiogram monitor to surveil for intermittent atrial fibrillation while treating with antiplatelet agents, then convert to oral anticoagulation if intermittent atrial fibrillation is detected. If we identify significant internal carotid stenosis, we refer for carotid endarterectomy during the same hospitalization regardless of infarct size. For all else, we use the dual antiplatelet agents aspirin (81 mg) and either clopidogrel (600 mg-load, followed by 75 mg daily) or ticagrelor (180-mg load, followed by 90 mg twice daily) daily for 21–30 days then continue aspirin at 81 mg daily. Ticagrelor has the advantage of not being affected by common polymorphisms of CYP2C19 that limit efficacy of clopidogrel in significant proportions of patients, particularly those of Asian descent. If the CTA revealed significant intracranial atherosclerosis or other precranial vessel stenosis within the vascular territory of the infarct (lumen caliber reduced by >50%), we continue dual antiplatelet agents for at least 3 months and then convert to a single agent. Unless contraindicated, all patients receive a high-intensity statin such as atorvastatin 80 mg or rosuvastatin 40 mg, with goal low-density lipoprotein level of <70 mg/dL unless the stroke has a nonatherothrombotic cause. Patients who are statin intolerant can receive PCSK9 inhibitors. Blood pressure control should target systolic blood pressure <120 mmHg long term, but we allow permissive hypertension for the first few days or weeks to help with collateral flow to the brain. BA, basilar artery; CTP, computed tomography perfusion; ICA, internal carotid artery; IV, intravenous; M1, middle cerebral artery first division; M2, middle cerebral artery second division; PA, plasminogen activator (tissue plasminogen activator or tenecteplase).

or hypertonic saline may be used to raise the serum osmolarity, but hypovolemia should be avoided because this may contribute to hypotension and worsening infarction. Combined analysis of three randomized European trials of hemicraniectomy (craniotomy and temporary removal of part of the skull) shows that hemicraniectomy reduces mortality by 50%, and the clinical outcomes of survivors are significantly improved. Older patients (age >60 years) benefit less but still significantly. The size of the diffusion-weighted imaging volume of brain infarction during the acute stroke is a predictor of future deterioration requiring hemicraniectomy.

Special vigilance is warranted for patients with cerebellar infarction. These strokes may mimic labyrinthitis because of prominent vertigo and vomiting; the presence of head or neck pain should alert the physician to consider cerebellar stroke due to vertebral artery dissection. Even small amounts of cerebellar edema can acutely increase intracranial pressure (ICP) by obstructing cerebrospinal fluid (CSF) flow leading to hydrocephalus or by directly compressing the brainstem. The resulting brainstem compression can manifest as coma and respiratory arrest and require emergency surgical decompression. Suboccipital decompression is recommended in patients with cerebellar infarcts who demonstrate neurologic deterioration and should be performed before significant brainstem compression occurs.

INTRAVENOUS THROMBOLYSIS

The National Institute of Neurological Disorders and Stroke (NINDS) Recombinant Tissue Plasminogen Activator (rtPA) Stroke Study showed a clear benefit for IV rtPA in selected patients with acute stroke. The NINDS study used IV rtPA (0.9 mg/kg to a 90-mg maximum; 10% as a bolus, then the remainder over 60 min) versus placebo in ischemic stroke within 3 h of onset. One-half of the patients were treated within 90 min. Symptomatic intracranial hemorrhage occurred in 6.4% of patients on rtPA and 0.6% on placebo. In the rtPA group, there was a significant 12% absolute increase in the number of patients with only minimal disability (32% on placebo and 44% on rtPA) and a nonsignificant 4% reduction in mortality (21% on placebo and 17% on rtPA). Thus, despite an increased incidence of symptomatic intracranial hemorrhage, treatment with IV rtPA within 3 h of the onset of ischemic stroke improved clinical outcome.

Three subsequent trials of IV rtPA did not confirm this benefit, perhaps because of the dose of rtPA used, the timing of its delivery, and small sample size. When data from all randomized IV rtPA trials were combined, however, efficacy was confirmed in the <3-h time window, and efficacy likely extended to 4.5 h and possibly to 6 h. Based on these combined results, the European Cooperative Acute Stroke Study (ECASS) III explored the safety and efficacy of rtPA in the 3- to 4.5-h time window. Unlike the NINDS study, patients aged >80 years and diabetic patients with a previous stroke were excluded. In this 821-patient randomized study, efficacy was again confirmed, although the treatment effect was less robust than in the 0- to 3-h time window. In the rtPA group, 52.4% of patients achieved a good outcome at 90 days, compared to 45.2% of the placebo group (odds ratio [OR] 1.34, $p = .04$). The symptomatic intracranial hemorrhage rate was 2.4% in the rtPA group and 0.2% in the placebo group ($p = .008$).

Based on these data, rtPA is approved in the 3- to 4.5-h window in Europe and Canada but is only approved for 0–3 h in the United States. A dose of 0.6 mg/kg is typically used in Japan and other Asian countries based on observation of >600 patients given this lower dose and observing similar outcomes to historical controls and a lower rate of intracranial hemorrhage. This dose also mitigates concerns that patients of Asian descent have a higher propensity to bleed from most antithrombotic and thrombolytic medications. The infrastructure to efficiently administer IV rtPA to eligible patients is a central component of primary stroke centers (see below). It represents the first treatment proven to improve clinical outcomes in ischemic stroke and is cost-effective and cost-saving. The time of stroke onset is defined as the time the patient's symptoms were witnessed to begin or the time the patient was last seen as normal. Patients who awaken with stroke have the onset defined as when they went to bed. Advanced neuroimaging techniques **(see Chap. 437)** may help to select patients beyond the 4.5-h window who will benefit from thrombolysis. Two trials using MRI selection beyond 4.5 h have shown clinical benefit from IV rtPA. Patients with minor stroke (nondisabling deficit and National Institutes of Health Stroke Scale [NIHSS] 0–5) appear to respond to acute aspirin or short-term dual antiplatelet therapy using the combination of aspirin and clopidogrel as well as IV rtPA. **Table 438-1** summarizes eligibility criteria and instructions for administration of IV rtPA.

TABLE 438-1 Administration of Intravenous Recombinant Tissue Plasminogen Activator (rtPA) or Tenecteplase for Acute Ischemic Stroke (AIS)[a]

INDICATION	CONTRAINDICATION
Clinical diagnosis of stroke Onset of symptoms to time of drug administration ≤4.5 h[b] CT scan showing no hemorrhage, and no edema > 1/3 of the MCA territory Age ≥18 years	Sustained BP >185/110 mmHg despite treatment Bleeding diathesis Recent head injury or intracerebral hemorrhage Major surgery in preceding 14 days Gastrointestinal bleeding in preceding 21 days Recent myocardial infarction

Administration of stroke thrombolysis
IV access with two peripheral IV lines (avoid arterial or central line placement)
Review eligibility for stroke thrombolysis
Administer 0.9 mg/kg IV (maximum 90 mg) rtPA IV as 10% of total dose by bolus, followed by remainder of total dose over 1 h[c]
Or Administer 0.25 mg/kg IV (maximum 25 mg) tenecteplase IV push over 5 s[d]
Frequent cuff BP monitoring
No other antithrombotic treatment for 24 h
For decline in neurologic status or uncontrolled BP, stop infusion, give cryoprecipitate, and reimage brain emergently
Avoid urethral catheterization for ≥2 h

[a]See Activase (tissue plasminogen activator) package insert for complete list of contraindications and dosing. [b]Depending on the country, IV rtPA may be approved for up to 4.5 h with additional restrictions. [c]An rtPA dose of 0.6 mg/kg is commonly used in Asia (Japan and China) based on randomized data indicating less hemorrhage and similar efficacy using this lower dose. [d]Use of tenecteplase for acute ischemic stroke is off-label.

Abbreviations: BP, blood pressure; CT, computed tomography; MCA, middle cerebral artery.

The plasminogen activator tenecteplase (0.25 mg/kg IV bolus over 5 s with a maximum dose of 25 mg) has been directly compared to rtPA and is being adopted by many centers because it is given without need for a 1-h infusion. This improves the efficiency of transferring patients to comprehensive stroke centers for thrombectomy because the IV infusion required for IV rtPA is not required for tenecteplase, thus obviating need for critical care transport. Several trials using tenecteplase prior to endovascular therapy have found it to be safe.

ENDOVASCULAR REVASCULARIZATION

Ischemic stroke from large-vessel intracranial occlusion results in high rates of mortality and morbidity. Occlusions in such large vessels (middle cerebral artery [MCA], intracranial internal carotid artery, and the basilar artery) generally involve a large clot burden and often fail to open with IV thrombolysis alone.

Endovascular mechanical thrombectomy has been studied as an alternative or adjunctive treatment of acute stroke in patients who are ineligible for, or have contraindications to, thrombolytics or in those who failed to achieve vascular recanalization with IV thrombolytics **(Fig. 438-3)**. In 2015, the results of six randomized trials were published, all demonstrating that endovascular therapy improved clinical outcomes for internal carotid and MCA occlusions proven by CT angiography (CTA), under 6 h from stroke

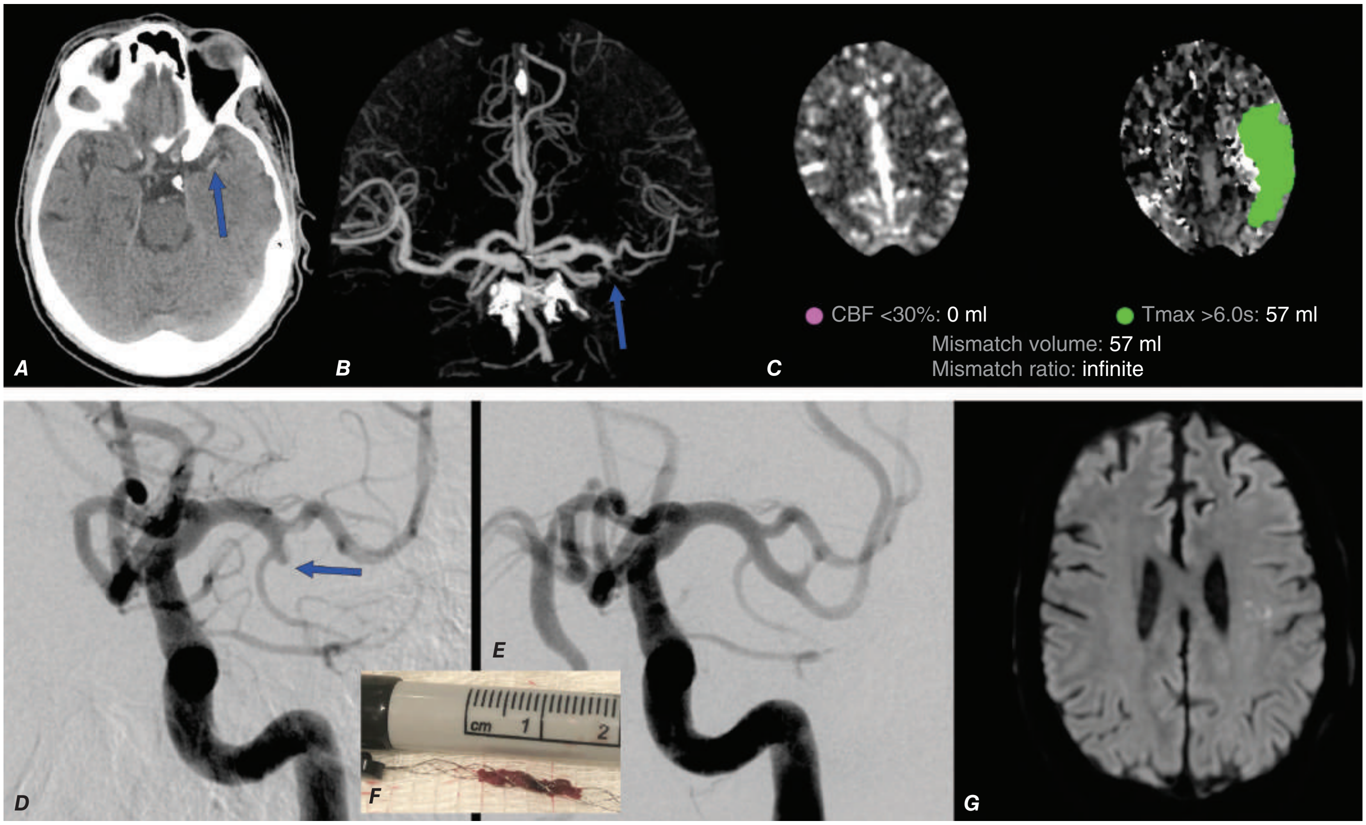

FIGURE 438-3 (***A***) Noncontrast head computed tomography (CT) scan of a 78-year-old man with atrial fibrillation and hypertension who was not taking oral anticoagulants and awoke with right hemiparesis and expressive aphasia. The head CT shows no intracerebral hemorrhage. He was not treated with plasminogen activators because his last seen normal time was 8 h prior. Head CT also shows hyperdensity in the left middle cerebral artery (MCA, *arrow*); this finding is highly specific for MCA occlusion but is poorly sensitive, as only 20% of patients with MCA occlusion show hyperdensity. (***B***) To confirm a large-vessel occlusion, CT angiography (CTA) performed in the same session reveals an occlusion of a secondary branch of the left MCA (*arrow*). (***C***) CT perfusion performed immediately following the CTA shows no core infarct (no pink signal in the left image) but a large region (green shading in the right image) of ischemic tissue that will die if revascularization is not achieved. (***D***) Catheter angiography shows the occluded branch of the left MCA (*arrow*) and (***E***) restored flow after successful clot removal (***F***). (***G***) A subsequent diffusion-weighted imaging scan shows a very small residual brain infarction. CBF, cerebral blood flow.

onset, with or without pretreatment with IV tissue plasminogen activator (tPA). One study concluded that patients were home nearly 2 months earlier if they received endovascular therapy. A combined meta-analysis of all patients in these trials confirmed a large benefit with endovascular therapy (OR, 2.49; 95% confidence interval [CI], 1.76–3.53; p <.001). The percentage of patients who achieved modified Rankin scores of 0–2 (normal or symptomatic but independent) was 46% in the endovascular group and 26.5% in the medical arm. A more recent meta-analysis reveals a mortality benefit with thrombectomy as well. As with IV rtPA treatment, clinical outcome is dependent on time to effective therapy. The odds of a good outcome exceed 3 if groin puncture occurs within 2 h of symptom onset but is only 2 if 8 h elapse. Over 80% of patients who had vessel opening within 1 h of arrival to the emergency department had a good outcome, whereas only one-third had a good outcome if 6 h elapsed.

The outcomes from endovascular therapy are likely improved with IV rtPA treatment prior to thrombectomy if the patient is eligible for rtPA and it is safe to administer. Recent data support replacing IV rtPA with IV tenecteplase because its simple bolus administration makes transporting the patient to an endovascular center less cumbersome.

Extending the time window beyond 6 h appears to be effective if the patient has specific imaging findings demonstrating good vascular collaterals (CT perfusion or magnetic resonance [MR] perfusion techniques, **see Chap. 437**) and can be treated within 24 h. The Clinical Mismatch in the Triage of Wake Up and Late Presenting Strokes Undergoing Neurointervention with Trevo (DAWN) trial reported good outcomes more frequently with endovascular therapy than with medical care alone (47 vs 13%, p <.0001). The Endovascular Therapy Following Imaging Evaluation for Ischemic Stroke 3 (DEFUSE-3) trial confirmed these results (45 vs 17%, p <.001) if treated up to 16 h from stroke onset. Nonrandomized data of thrombectomy for basilar occlusion have found this treatment to be safe up to 24 h from symptom onset and associated with lower 3-month Rankin scores. An example of how advanced imaging techniques (CT, CTA, CT perfusion, catheter-based angiography), endovascular thrombectomy with clot removal, and follow-up MRI can lead to a better than predicted stroke outcome is shown in Fig 438-3.

Now that endovascular stroke therapy is proven to be effective, the creation of comprehensive stroke centers designed to rapidly identify and treat patients with large-vessel cerebral ischemia has been a major focus internationally. Creating regional systems of care whereby stroke patients are first evaluated at acute stroke ready hospitals or primary stroke centers (which can administer IV rtPA or tenecteplase) then transferred to thrombectomy-capable or comprehensive stroke centers if needed, or directly triaged to thrombectomy-capable or comprehensive centers based on field assessment, appears to be an effective strategy to improve outcomes.

ANTITHROMBOTIC TREATMENT

Platelet Inhibition Aspirin is the only antiplatelet agent that has been proven to be effective for the acute treatment of ischemic stroke; there are several antiplatelet agents proven for the secondary prevention of stroke (see below). Two large trials, the International Stroke Trial (IST) and the Chinese Acute Stroke Trial (CAST), found that the use of aspirin within 48 h of stroke onset reduced both stroke recurrence risk and mortality minimally. Among 19,435 patients in IST, those allocated to aspirin, 300 mg/d, had slightly fewer deaths within 14 days (9.0 vs 9.4%), significantly fewer recurrent ischemic strokes (2.8 vs 3.9%), no excess of hemorrhagic strokes (0.9 vs 0.8%), and a trend toward a reduction in death or dependence at 6 months (61.2 vs 63.5%). In CAST, 21,106 patients with ischemic stroke received 160 mg/d of aspirin or a placebo for up to 4 weeks. In the aspirin group, there were very small reductions in early mortality (3.3 vs 3.9%), recurrent ischemic strokes (1.6 vs 2.1%), and dependency at discharge or death (30.5 vs 31.6%). These trials demonstrate that the use of aspirin in the treatment of acute ischemic stroke is safe and produces a small net benefit. For every 1000 acute strokes treated with aspirin, ~9 deaths or nonfatal stroke recurrences will be prevented in the first few weeks, and ~13 fewer patients will be dead or dependent at 6 months. The short-term combination of aspirin with clopidogrel or with ticagrelor following minor stroke or TIA is effective at preventing early second stroke (see below).

Anticoagulation Numerous clinical trials have failed to demonstrate any benefit of routine anticoagulation in the primary treatment of atherothrombotic cerebral ischemia and have also shown an increase in the risk of brain and systemic hemorrhage. Therefore, the routine use of heparin or other anticoagulants for patients with atherothrombotic stroke is not warranted. Heparin and oral anticoagulation are likely no more effective than aspirin for stroke associated with arterial dissection. However, there may be benefit of anticoagulation for halting progression of dural sinus thrombosis.

NEUROPROTECTION

Neuroprotection is the concept of providing a treatment that prolongs the brain's tolerance to ischemia. Drugs that block the excitatory amino acid pathways have been shown to protect neurons and glia in animals, but despite multiple clinical trials, they have not yet been proven to be beneficial in humans. Hypothermia is a powerful neuroprotective treatment in patients with cardiac arrest **(Chap. 318)** and is neuroprotective in animal models of stroke, but it has not been adequately studied in patients with ischemic stroke and is associated with an increase in pneumonia rates that could adversely impact stroke outcomes. Hypothermia combined with hemicraniectomy is no more effective than hemicraniectomy with euthermia.

STROKE CENTERS AND REHABILITATION

Patient care in stroke units followed by rehabilitation services improves neurologic outcomes and reduces mortality. Use of clinical pathways and staff dedicated to the stroke patient can improve care. This includes use of standardized stroke order sets. Stroke teams that provide emergency 24-h evaluation of acute stroke patients for acute medical management and consideration of thrombolysis or endovascular treatments, potentially provided using telemedicine/telestroke services, are essential components of primary and comprehensive stroke centers, respectively.

Proper rehabilitation of the stroke patient includes early physical, occupational, and speech therapy. It is directed toward educating the patient and family about the patient's neurologic deficit, preventing the complications of immobility (e.g., pneumonia, DVT and pulmonary embolism, pressure sores of the skin, and muscle contractures), and providing encouragement and instruction in overcoming the deficit. Use of pneumatic compression stockings is of proven benefit in reducing risk of DVT and is a safe alternative to heparin. The goal of rehabilitation is to return the patient home and to maximize recovery by providing a safe, progressive regimen suited to the individual patient. Additionally, the use of constrained movement therapy (immobilizing the unaffected side) has been shown to improve hemiparesis following stroke, even years after the stroke, suggesting that physical therapy can recruit unused neural pathways. Controversy exists regarding whether selective serotonin uptake inhibitors improve motor recovery, but they may be helpful in preventing poststroke depression. Newer robotic therapies and neuromodulation approaches using transcranial magnetic stimulation or transcranial direct current stimulation are under active investigation **(Chap. 500)**. The human nervous system is more adaptable than previously thought, and developing physical and pharmacologic strategies to enhance long-term neural recovery are the subject of ongoing research.

ETIOLOGY OF ISCHEMIC STROKE

(Fig. 438-4 and Table 438-2) Although the initial management of acute ischemic stroke often does not depend on the etiology, establishing a cause is essential to reduce the risk of recurrence. Focus should be on atrial fibrillation and carotid atherosclerosis, because these etiologies

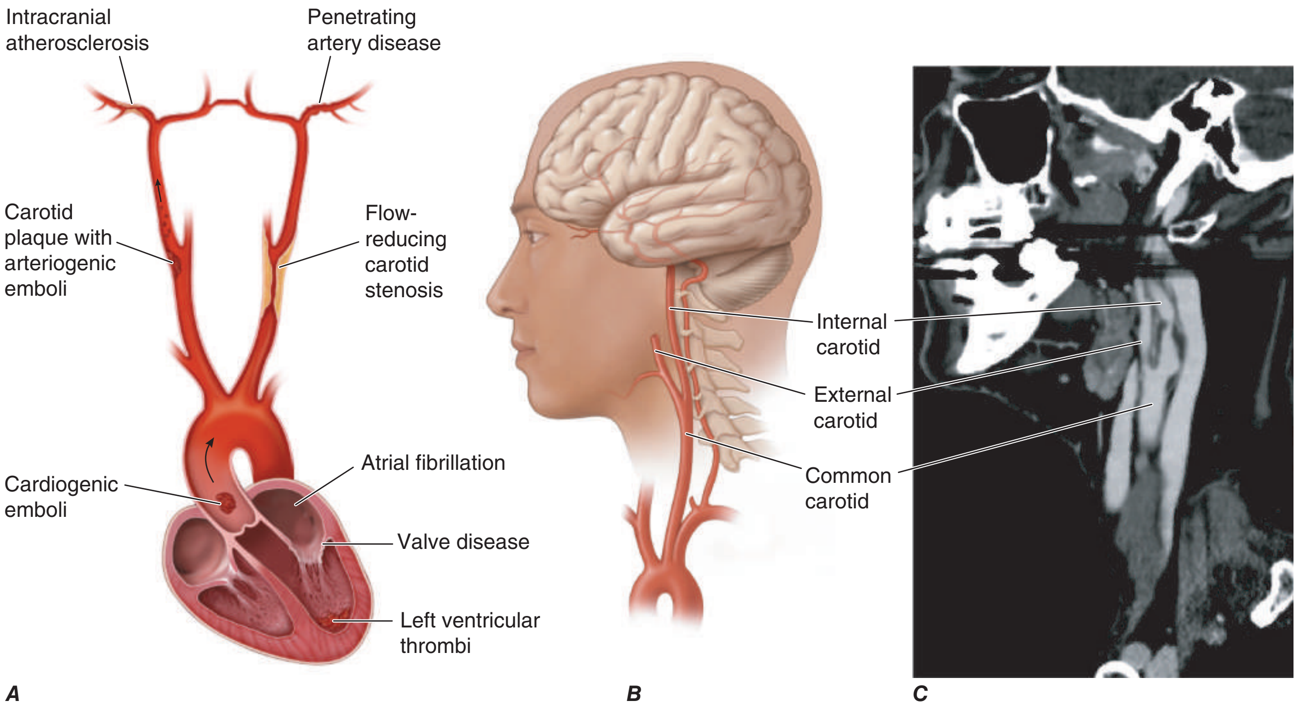
Intracranial atherosclerosis
Penetrating artery disease
Carotid plaque with arteriogenic emboli
Flow-reducing carotid stenosis
Cardiogenic emboli
Atrial fibrillation
Valve disease
Left ventricular thrombi
Internal carotid
External carotid
Common carotid
A
B
C

FIGURE 438-4 Pathophysiology of ischemic stroke. ***A.*** Diagram illustrating the three major mechanisms that underlie ischemic stroke: (1) occlusion of an intracranial vessel by an embolus (e.g., cardiogenic sources such as atrial fibrillation or artery-to-artery emboli from carotid atherosclerotic plaque), often affecting the large intracranial vessels; (2) in situ thrombosis of an intracranial vessel, typically affecting the small penetrating arteries that arise from the major intracranial arteries; and (3) hypoperfusion caused by flow-limiting stenosis of a major extracranial (e.g., internal carotid) or intracranial vessel, often producing "watershed" ischemia. ***B.*** and ***C.*** Diagram and reformatted computed tomography angiogram of the common, internal, and external carotid arteries. High-grade stenosis of the internal carotid artery, which may be associated with either cerebral emboli or flow-limiting ischemia, was identified in this patient.

TABLE 438-2 Causes of Ischemic Stroke

COMMON CAUSES	UNCOMMON CAUSES
Thrombosis	Hypercoagulable disorders
Lacunar stroke (small vessel)	Protein C deficiency[a]
Large-vessel thrombosis	Protein S deficiency[a]
Dehydration	Antithrombin III deficiency[a]
Embolic occlusion	Antiphospholipid syndrome
Artery-to-artery	Factor V Leiden mutation[a]
Carotid bifurcation	Prothrombin G20210A mutation[a]
Aortic arch	Systemic malignancy
Arterial dissection	Sickle cell anemia
Cardioembolic	β Thalassemia
Atrial fibrillation	Polycythemia vera
Mural thrombus	Systemic lupus erythematosus
Myocardial infarction	Homocysteinemia
Dilated cardiomyopathy	Thrombotic thrombocytopenic purpura
Valvular lesions	Disseminated intravascular coagulation
Mitral stenosis	Dysproteinemias[a]
Mechanical valve	Nephrotic syndrome[a]
Bacterial endocarditis	Inflammatory bowel disease[a]
Paradoxical embolus	Oral contraceptives
Atrial septal defect	COVID-19 infection
Patent foramen ovale	Venous sinus thrombosis[b]
Atrial septal aneurysm	Fibromuscular dysplasia
Spontaneous echo contrast	Vasculitis
Stimulant drugs: cocaine, amphetamine	Systemic vasculitis (PAN, granulomatosis with polyangiitis, Takayasu's, giant cell arteritis)
	Primary CNS vasculitis
	Meningitis (syphilis, tuberculosis, fungal, bacterial, zoster)
	Noninflammatory vasculopathy
	Reversible vasoconstriction syndrome
	Fabry's disease
	Angiocentric lymphoma
	Cardiogenic
	Mitral valve calcification
	Atrial myxoma
	Intracardiac tumor
	Marantic endocarditis
	Libman-Sacks endocarditis
	Subarachnoid hemorrhage vasospasm
	Moyamoya disease
	Eclampsia

[a]Chiefly cause venous sinus thrombosis. [b]May be associated with any hypercoagulable disorder.

Abbreviations: CNS, central nervous system; PAN, polyarteritis nodosa.

have proven secondary prevention strategies. The clinical presentation and examination findings often establish the cause of stroke or narrow the possibilities to a few. Judicious use of laboratory testing and imaging studies completes the initial evaluation. Nevertheless, nearly 30% of strokes remain unexplained despite extensive evaluation.

Clinical examination should focus on the peripheral and cervical vascular system (measuring blood pressure), the heart (dysrhythmia, murmurs), extremities (peripheral emboli), and retina (effects of hypertension and cholesterol emboli [Hollenhorst plaques]). A complete neurologic examination is performed to localize the anatomic site of stroke **(Chap. 437)**. An imaging study of the brain is nearly always indicated and is required for patients being considered for thrombolysis; it may be combined with CT- or MRI-based angiography to visualize the vasculature of the neck and intracranial vessels **(see "Imaging Studies," Chap. 437)**. A chest x-ray, electrocardiogram (ECG), urinalysis, complete blood count, erythrocyte sedimentation rate (ESR), serum electrolytes, blood urea nitrogen (BUN), creatinine, blood glucose, serum lipid profile, prothrombin time (PT), and partial thromboplastin time (PTT) are often useful and should be considered in all patients. An ECG, and subsequent cardiac telemetry, may demonstrate arrhythmias or reveal evidence of recent myocardial infarction (MI). Of all these studies, only brain imaging and finger stick blood glucose are necessary prior to IV thrombolysis; the results of other studies should not delay the rapid administration of IV thrombolysis if the patient is eligible.

Cardioembolic Stroke Cardioembolism is responsible for ~20% of all ischemic strokes. Stroke caused by heart disease is primarily due to embolism of thrombotic material forming on the atrial or ventricular wall or the left heart valves. These thrombi then detach and embolize into the arterial circulation. The thrombus may fragment or lyse quickly, producing only a TIA. Alternatively, the arterial occlusion may last longer, producing stroke. Embolic strokes tend to occur suddenly with maximum neurologic deficit present at onset. With reperfusion following more prolonged ischemia, petechial hemorrhages can occur within the ischemic territory. These are usually of no clinical significance and should be distinguished from frank intracranial hemorrhage into a region of ischemic stroke where the mass effect from the hemorrhage can cause a significant decline in neurologic function.

Emboli from the heart most often lodge in the intracranial internal carotid artery, the MCA, the posterior cerebral artery (PCA), or one of their branches; infrequently, the anterior cerebral artery (ACA) is involved. Emboli large enough to occlude the stem of the MCA (3–4 mm) or internal carotid terminus lead to large infarcts that involve both deep gray and white matter and some portions of the cortical surface and its underlying white matter. A smaller embolus may occlude a small cortical or penetrating arterial branch. The location and size of an infarct within a vascular territory depend on the extent of the collateral circulation.

The most significant cause of cardioembolic stroke in most of the world is nonrheumatic (often called nonvalvular) atrial fibrillation. MI, prosthetic valves, rheumatic heart disease, and ischemic cardiomyopathy are other considerations (Table 438-2). Cardiac disorders causing brain embolism are discussed in the chapters on heart diseases, but a few pertinent aspects are highlighted here.

Nonrheumatic atrial fibrillation is the most common cause of cerebral embolism overall. The presumed stroke mechanism is thrombus formation in the fibrillating atrium or atrial appendage, with subsequent embolization. Patients with atrial fibrillation have an average annual risk of stroke of ~5%. The risk of stroke can be estimated by calculating the CHA_2DS_2-VASc score **(Table 438-3)**. Left atrial enlargement is an additional risk factor for formation of atrial thrombi. Rheumatic heart disease usually causes ischemic stroke when there is prominent mitral stenosis or atrial fibrillation. Recent MI may be a source of emboli, especially when transmural and involving the anteroapical ventricular wall, and prophylactic anticoagulation following MI with left ventricular thrombus has been shown to reduce ischemic stroke risk. Mitral valve prolapse is not usually a source of emboli unless the prolapse is severe.

Paradoxical embolization occurs when venous thrombi migrate to the arterial circulation, usually via a patent foramen ovale (PFO) or atrial septal defect. Bubble-contrast echocardiography (IV injection of agitated saline coupled with either transthoracic or transesophageal echocardiography) can demonstrate a right-to-left cardiac shunt, revealing the conduit for paradoxical embolization. Alternatively, a right-to-left shunt is implied if immediately following IV injection of agitated saline, the ultrasound signature of bubbles is observed during transcranial Doppler insonation of the MCA; pulmonary arteriovenous malformations should be considered if this test is positive yet an echocardiogram fails to reveal an intracardiac shunt. Both techniques are highly sensitive for detection of right-to-left shunts. Besides venous clot, fat and tumor emboli, bacterial endocarditis, IV air, and amniotic fluid emboli at childbirth may occasionally be responsible for paradoxical embolization. The importance of a PFO as a cause of stroke is debated, particularly because they are present in ~15% of the general population. The presence of a venous source of embolus, most commonly a deep-venous thrombus, may provide confirmation of the importance of a PFO with an accompanying right-to-left shunt in a particular case. Meta-analysis of three recent randomized trials reported a hazard ratio of 0.41 for recurrent stroke (about a 1% per year absolute reduction) using percutaneous occlusion devices in patients with larger PFOs and no other explanation for their stroke. Guidelines now endorse PFO closure with percutaneous devices after consultation with a neurologist and a cardiologist.

Bacterial endocarditis can be a source of valvular vegetations that give rise to septic emboli. The appearance of multifocal symptoms and signs in a patient with stroke makes bacterial endocarditis more likely. Infarcts of microscopic size occur, and large septic infarcts may evolve into brain abscesses or cause hemorrhage into the infarct, which generally precludes use of anticoagulation or thrombolytics. Mycotic aneurysms caused by septic emboli may also present as subarachnoid hemorrhage (SAH) or intracerebral hemorrhage.

Artery-to-Artery Embolic Stroke Thrombus formation on atherosclerotic plaques may embolize to intracranial arteries producing an artery-to-artery embolic stroke. Less commonly, a diseased vessel may acutely thrombose. Unlike the myocardial vessels, artery-to-artery embolism, rather than local thrombosis, appears to be the dominant vascular mechanism causing large-vessel brain ischemia. Any diseased vessel may be an embolic source, including the aortic arch, common carotid, internal carotid, vertebral, and basilar arteries.

CAROTID ATHEROSCLEROSIS Atherosclerosis within the carotid artery occurs most frequently within the common carotid bifurcation and proximal internal carotid artery; the carotid siphon (portion within the cavernous sinus) is also vulnerable to atherosclerosis. Male gender, older age, smoking, hypertension, diabetes, and hypercholesterolemia are risk factors for carotid disease, as they are for stroke in general **(Table 438-4)**. Carotid atherosclerosis produces an estimated 10% of ischemic stroke. **For further discussion of the pathogenesis of atherosclerosis, see Chap. 244.**

Carotid disease can be classified by whether the stenosis is symptomatic or asymptomatic and by the degree of stenosis (percent narrowing of the narrowest segment compared to a nondiseased segment). Symptomatic carotid disease implies that the patient has experienced a recent (within 6 months) stroke or TIA within the vascular distribution of the artery, and it is associated with a greater risk of subsequent stroke than asymptomatic stenosis, in which the patient is symptom free and the stenosis is detected through screening. Greater degrees of arterial narrowing are generally associated with a higher risk of stroke, except that those with near occlusions are at lower risk of stroke.

OTHER CAUSES OF ARTERY-TO-ARTERY EMBOLIC STROKE *Intracranial atherosclerosis* produces stroke either by an embolic mechanism or by in situ thrombosis of a diseased vessel. It is more common in patients of Asian and African-American descent. Recurrent stroke risk is ~15% per year, similar to untreated symptomatic carotid atherosclerosis.

Dissection of the internal carotid or vertebral arteries or even vessels beyond the circle of Willis is a common source of embolic stroke in young (age <60 years) patients. The dissection is usually painful and precedes the stroke by several hours or days. Extracranial dissections do not cause hemorrhage, presumably because of the tough adventitia of these vessels. Intracranial dissections, conversely, may produce SAH because the adventitia of intracranial vessels is thin and pseudoaneurysms may form, requiring urgent treatment to prevent rerupture. Treating asymptomatic pseudoaneurysms following extracranial dissection is likely not necessary. The cause of dissection is usually unknown, and recurrence is rare. Ehlers-Danlos type IV, Marfan's disease and related disorders of connective tissue **(Chap. 425)**, and fibromuscular dysplasia **(Chap. 292)** are associated with dissections. Trauma (usually a motor vehicle accident or a sports injury) can cause carotid and vertebral artery dissections. Spinal manipulative therapy is associated with vertebral artery dissection and stroke. Most dissections heal spontaneously, and stroke or TIA is uncommon beyond 2 weeks. One trial showed no difference in stroke prevention with antiplatelet regimens compared to anticoagulation, with a low recurrent stroke rate of 2%.

TABLE 438-3 Recommendations on Chronic Use of Antithrombotics for Various Cardiac Conditions

CONDITION	RECOMMENDATION
Nonvalvular atrial fibrillation	Calculate CHA_2DS_2-VASc score[a]
• CHA_2DS_2-VASc score of 0	Aspirin or no antithrombotic
• CHA_2DS_2-VASc score of 1	Aspirin or OAC
• CHA_2DS_2-VASc score of ≥2	OAC
Rheumatic mitral valve disease	
• With atrial fibrillation, previous embolization, or atrial appendage thrombus, or left atrial diameter >55 mm	OAC
• Embolization or appendage clot despite OAC	OAC plus aspirin
• Mitral valve prolapse	
• Asymptomatic	No therapy
• With otherwise cryptogenic stroke or TIA	Aspirin
• Atrial fibrillation	OAC
Mitral annular calcification	
• Without atrial fibrillation but systemic embolization, or otherwise cryptogenic stroke or TIA	Aspirin
• Recurrent embolization despite aspirin	OAC
• With atrial fibrillation	OAC
Aortic valve calcification	
• Asymptomatic	No therapy
• Otherwise cryptogenic stroke or TIA	Aspirin
Aortic arch mobile atheroma	
• Otherwise cryptogenic stroke or TIA	Aspirin or OAC
Patent foramen ovale	
• Otherwise cryptogenic ischemic stroke or TIA	Aspirin or closure with device
• Indication for OAC (deep-venous thrombosis or hypercoagulable state)	OAC
Mechanical heart value	
• Aortic position, bileaflet or Medtronic Hall tilting disk with normal left atrial size and sinus rhythm	VKA INR 2.5, range 2–3
• Mitral position tilting disk or bileaflet valve	VKA INR 3.0, range 2.5–3.5
• Mitral or aortic position, anterior-apical myocardial infarct or left atrial enlargement	VKA INR 3.0, range 2.5–3.5
• Mitral or aortic position, with atrial fibrillation, or hypercoagulable state, or low ejection fraction, or atherosclerotic vascular disease	Aspirin plus VKA INR 3.0, range 2.5–3.5
• Systemic embolization despite target INR	Add aspirin and/or increase INR: prior target was 2.5, increase to 3.0, range 2.5–3.5; prior target was 3.0, increase to 3.5, range 3–4
Bioprosthetic valve	
• No other indication for VKA therapy	Aspirin
Infective endocarditis	Avoid antithrombotic agents
Nonbacterial thrombotic endocarditis	
• With systemic embolization	Full-dose, unfractionated heparin or SC LMWH, or Xa inhibitor

[a] CHA_2DS_2-VASc score is calculated as follows: 1 point for congestive heart failure, 1 point for hypertension, 2 points for age ≥75 years, 1 point for diabetes mellitus, 2 points for stroke or TIA, 1 point for vascular disease (prior myocardial infarction, peripheral vascular disease, or aortic plaque), 1 point for age 65–74 years, 1 point for female sex category; sum of points is the total CHA_2DS_2-VASc score.

Note: Dose of aspirin is 50–325 mg/d; target INR for VKA is between 2 and 3 unless otherwise specified.

Abbreviations: INR, international normalized ratio; LMWH, low-molecular-weight heparin; OAC, oral anticoagulant (VKA, thrombin inhibitor, or oral factor Xa inhibitors); TIA, transient ischemic attack; VKA, vitamin K antagonist.

Sources: Data from DE Singer et al: Chest 133:546S, 2008; DN Salem et al: Chest 133:593S, 2008; CT January et al: JACC 64:2246, 2014.

TABLE 438-4 Risk Factors for Stroke

RISK FACTOR	RELATIVE RISK	RELATIVE RISK REDUCTION WITH TREATMENT	NUMBER NEEDED TO TREAT[a]	
			PRIMARY PREVENTION	SECONDARY PREVENTION
Hypertension	2–5	38%	100–300	50–100
Atrial fibrillation	1.8–2.9	68% warfarin, 21% aspirin	20–83	13
Diabetes	1.8–6	No proven effect		
Smoking	1.8	50% at 1 year, baseline risk at 5 years postcessation		
Hyperlipidemia	1.8–2.6	16–30%	560	230
Asymptomatic carotid stenosis	2.0	53%	85	N/A
Symptomatic carotid stenosis (70–99%)		65% at 2 years	N/A	12
Symptomatic carotid stenosis (50–69%)		29% at 5 years	N/A	77

[a] Number needed to treat to prevent one stroke annually. Prevention of other cardiovascular outcomes is not considered here.

Abbreviation: N/A, not applicable.

SMALL-VESSEL STROKE

The term *lacunar infarction* refers to infarction following atherothrombotic or lipohyalinotic occlusion of a small artery in the brain. The term *small-vessel stroke* denotes occlusion of such a small penetrating artery and is now the preferred term. Small-vessel strokes account for ~20% of all strokes.

Pathophysiology The MCA stem, the arteries comprising the circle of Willis (A1 segment, anterior and posterior communicating arteries, and P1 segment), and the basilar and vertebral arteries all give rise to 30- to 300-μm branches that penetrate the deep gray and white matter of the cerebrum or brainstem **(Fig. 438-5)**. Each of these small branches can occlude either by atherothrombotic disease at its origin or by the development of lipohyalinotic thickening. Thrombosis of these vessels causes small infarcts that are referred to as *lacunes* (Latin for "lake" of fluid noted at autopsy). These infarcts range in size from 3 mm to 2 cm in diameter. Hypertension and age are the principal risk factors.

Clinical Manifestations The most common small-vessel stroke syndromes are the following: (1) *pure motor hemiparesis* from an infarct in the posterior limb of the internal capsule or the pons; the face, arm, and leg are almost always involved; (2) *pure sensory stroke* from an infarct in the ventral thalamus; (3) *ataxic hemiparesis* from an infarct in the ventral pons or internal capsule; (4) and *dysarthria and a clumsy hand* or arm due to infarction in the ventral pons or in the genu of the internal capsule.

Transient symptoms (small-vessel TIAs) may herald a small-vessel infarct; they may occur several times a day and last only a few minutes. Recovery from small-vessel strokes tends to be more rapid and complete than recovery from large-vessel strokes; in some cases, however, there is early worsening of symptoms or a stuttering course and severe permanent disability may result.

A large-vessel source (either thrombosis or embolism) may manifest initially as a small-vessel infarction. Therefore, the search for embolic sources (carotid and heart) should not be completely abandoned in the evaluation of these patients. Secondary prevention of small-vessel stroke involves risk factor modification, specifically reduction in blood pressure (see "Treatment: Primary and Secondary Prevention of Stroke and TIA," below).

LESS COMMON CAUSES OF STROKE

(Table 438-2) *Hypercoagulable disorders* **(Chap. 69)** primarily increase the risk of cortical vein or cerebral venous sinus thrombosis. Systemic

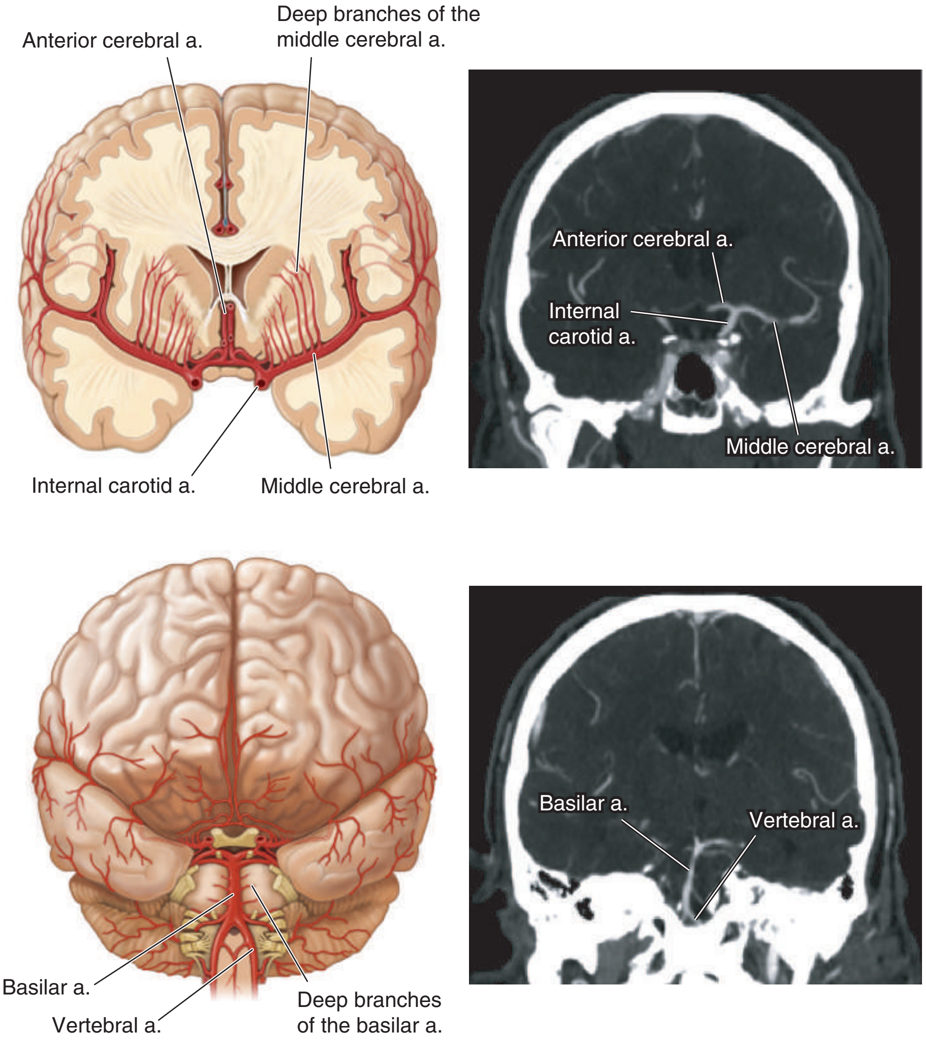

FIGURE 438-5 Diagrams and reformatted computed tomography (CT) angiograms in the coronal section illustrating the deep penetrating arteries involved in small-vessel strokes. In the anterior circulation, small penetrating arteries called *lenticulostriates* arise from the proximal portion of the anterior and middle cerebral arteries and supply deep subcortical structures (***upper panels***). In the posterior circulation, similar arteries arise directly from the vertebral and basilar arteries to supply the brainstem (***lower panels***). Occlusion of a single penetrating artery gives rise to a discrete area of infarct (pathologically termed a "lacune," or lake). Note that these vessels are too small to be visualized on CT angiography.

lupus erythematosus **(Chap. 368)** with Libman-Sacks endocarditis can be a cause of embolic stroke. These conditions overlap with the antiphospholipid syndrome **(Chap. 369)**, which probably requires long-term anticoagulation to prevent further stroke. Homocysteinemia may cause arterial thromboses as well; this disorder is caused by various mutations in the homocysteine pathways and responds to different forms of cobalamin depending on the mutation. Disseminated intravascular coagulopathy can cause both venous and arterial occlusive events; COVID-19 infection may predispose for acute ischemic stroke due to large-vessel occlusion.

Venous sinus thrombosis of the lateral or sagittal sinus or of small cortical veins (cortical vein thrombosis) occurs as a complication of oral contraceptive use, pregnancy and the postpartum period, inflammatory bowel disease, intracranial infections (meningitis), and dehydration. It is also seen in patients with laboratory-confirmed thrombophilia including antiphospholipid syndrome, polycythemia, sickle cell anemia, deficiencies of proteins C and S, factor V Leiden mutation (resistance to activated protein C), antithrombin III deficiency, homocysteinemia, and the prothrombin G20210A mutation. Women who take oral contraceptives and have the prothrombin G20210A mutation may be at particularly high risk for sinus thrombosis. Patients present with headache and may also have focal neurologic signs (especially paraparesis) and seizures. Often, CT imaging is normal unless an intracranial venous hemorrhage has occurred, but the venous sinus occlusion is readily visualized using MR or CT venography or conventional x-ray angiography. With greater degrees of sinus thrombosis, the patient may develop signs of increased ICP and coma. Intravenous heparin, regardless of the presence of intracranial hemorrhage, reduces morbidity and mortality, and the long-term outcome is generally good. Heparin prevents further thrombosis and reduces venous hypertension and ischemia. If an underlying hypercoagulable state is not found, many physicians treat with oral anticoagulants for 3–6 months and then convert to aspirin, depending on the degree of resolution of the venous sinus thrombus. Anticoagulation is often continued indefinitely if thrombophilia is diagnosed.

Sickle cell anemia (SS disease) is a common cause of stroke in children. A subset of homozygous carriers of this hemoglobin mutation develop stroke in childhood, and this may be predicted by documenting high-velocity blood flow within the MCAs using transcranial Doppler ultrasonography. In children who are identified to have high velocities, treatment with aggressive exchange transfusion dramatically reduces risk of stroke, and if exchange transfusion is ceased, their stroke rate increases again along with MCA velocities.

Fibromuscular dysplasia **(Chap. 292)** affects the cervical arteries and occurs mainly in women. The carotid or vertebral arteries show multiple rings of segmental narrowing alternating with dilatation. Vascular occlusion is usually incomplete. The process is often asymptomatic but occasionally is associated with an audible bruit, TIAs, or stroke. Involvement of the renal arteries is common and may cause hypertension. The cause and natural history of fibromuscular dysplasia are unknown. TIA or stroke generally occurs only when the artery is severely narrowed or dissects. Anticoagulation or antiplatelet therapy may be helpful.

Temporal (giant cell) arteritis **(Chap. 375)** is a relatively common affliction of elderly individuals in which the external carotid system, particularly the temporal arteries, undergoes subacute granulomatous inflammation with giant cells. Occlusion of posterior ciliary arteries derived from the ophthalmic artery results in blindness in one or both eyes and can be prevented with glucocorticoids. It rarely causes stroke because the internal carotid artery is usually not inflamed. Idiopathic giant cell arteritis involving the great vessels arising from the aortic arch (*Takayasu's arteritis*) may cause carotid or vertebral thrombosis; it is rare in the Western Hemisphere.

Necrotizing (or granulomatous) arteritis **(Chap. 375)**, occurring alone or in association with generalized polyarteritis nodosa or granulomatosis with polyangiitis, involves the distal small branches (<2 mm diameter) of the main intracranial arteries and produces small ischemic infarcts in the brain, optic nerve, and spinal cord. The CSF often shows pleocytosis, and the protein level is elevated. *Primary central nervous system vasculitis* is rare; small or medium-sized vessels are usually affected, without apparent systemic vasculitis. The differential diagnosis includes other inflammatory vasculopathies including infection (tuberculous, fungal), sarcoidosis, angiocentric lymphoma, carcinomatous meningitis, and noninflammatory causes such as atherosclerosis, emboli, connective tissue disease, vasospasm, migraine-associated vasculopathy, and drug-associated causes. Some cases develop in the postpartum period and are self-limited.

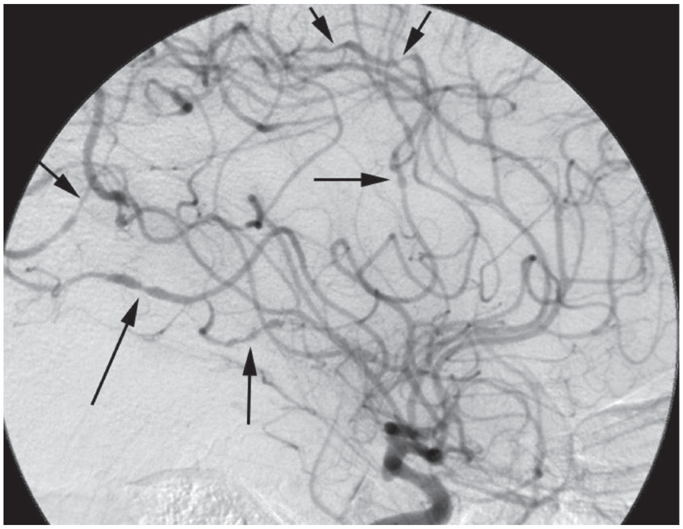

FIGURE 438-6 Cerebral angiogram from a 32-year-old male with central nervous system vasculopathy. Dramatic beading (*arrows*) typical of vasculopathy is shown.

Patients with any form of vasculopathy may present with insidious progression of combined white and gray matter infarctions, prominent headache, and cognitive decline. Brain biopsy or high-resolution conventional x-ray angiography is usually required to make the diagnosis **(Fig. 438-6)**. A lumbar puncture (elevated white blood cells, elevated IgG index, bands on electrophoresis) can provide support for an inflammatory etiology of a neurovascular problem. When inflammation is confirmed, aggressive immunosuppression with glucocorticoids, and often cyclophosphamide, is usually necessary to prevent progression; a diligent investigation for infectious causes such as tuberculosis is essential prior to immunosuppression. With prompt recognition and treatment, many patients can make an excellent recovery.

Drugs, in particular amphetamines and perhaps cocaine, may cause stroke on the basis of acute hypertension or drug-induced vasculopathy. This vasculopathy is commonly due to vasospasm or atherosclerosis, but cases of inflammatory vasculitis have also been reported. No data exist on the value of any treatment, but cessation of stimulants is prudent. Phenylpropanolamine has been linked with intracranial hemorrhage, as has cocaine and methamphetamine, perhaps related to a vasculopathy. *Moyamoya disease* is a poorly understood occlusive disease involving large intracranial arteries, especially the distal internal carotid artery and the stem of the MCA and ACA. Vascular inflammation is absent. The lenticulostriate arteries develop a rich collateral circulation around the occlusive lesion, which gives the impression of a "puff of smoke" (*moyamoya* in Japanese) on conventional x-ray angiography. Other collaterals include transdural anastomoses between the cortical surface branches of the meningeal and scalp arteries. The disease occurs mainly in Asian children or young adults, but the appearance may be identical in adults who have atherosclerosis, particularly in association with diabetes. Intracranial hemorrhage may result from rupture of the moyamoya collaterals; thus, anticoagulation is risky. Progressive occlusion of large surface arteries can occur, producing large-artery distribution strokes. Surgical bypass of extracranial carotid arteries to the dura or MCAs may prevent stroke and hemorrhage.

Posterior reversible encephalopathy syndrome (PRES) can occur with head injury, seizure, migraine, sympathomimetic drug use, and eclampsia and in the postpartum period. The pathophysiology is uncertain but likely involves a hyperperfusion state where blood pressure exceeds the upper limit of cerebral autoregulation resulting in cerebral edema **(Chap. 318)**. Patients complain of headache and manifest fluctuating neurologic symptoms and signs, especially visual symptoms. Sometimes cerebral infarction ensues, but typically, the clinical and imaging findings reverse completely. MRI findings are characteristic with edema present within the occipital lobes but also can be generalized and do not respect any single vascular territory. A closely related *reversible cerebral vasoconstriction syndrome* (RCVS) typically presents with sudden, severe headache closely mimicking SAH. Patients may experience ischemic infarction and intracerebral hemorrhage and typically have new-onset, severe hypertension. Conventional x-ray angiography reveals changes in the vascular caliber throughout the hemispheres resembling vasculitis, but the process is noninflammatory. Oral calcium channel blockers may be effective in producing remission, and recurrence is rare.

Leukoaraiosis, or *periventricular white matter disease*, is the result of multiple small-vessel infarcts within the subcortical white matter. It is readily seen on CT or MRI scans as areas of white matter injury surrounding the ventricles and within the corona radiata. The pathophysiologic basis of the disease is lipohyalinosis of small penetrating arteries within the white matter, likely produced by chronic hypertension. Patients with periventricular white matter disease may develop a subcortical dementia syndrome, and it is likely that this common form of dementia may be delayed or prevented with antihypertensive medications **(Chap. 444)**.

CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy) is an inherited disorder that presents as small-vessel strokes, progressive dementia, and extensive symmetric white matter changes often including the anterior temporal lobes visualized by MRI. Approximately 40% of patients have migraine with aura, often manifest as transient motor or sensory deficits. Onset is usually in the fourth or fifth decade of life. This autosomal dominant condition is caused by one of several mutations in *Notch-3*, a member of a highly conserved gene family characterized by epidermal growth factor repeats in its extracellular domain. Other monogenic ischemic stroke syndromes include cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL) and hereditary endotheliopathy, retinopathy, nephropathy, and stroke (HERNS). Fabry's disease also produces both a large-vessel arteriopathy and small-vessel infarctions. The *COL4A1* mutation is associated with multiple small-vessel strokes with hemorrhagic transformation.

■ TRANSIENT ISCHEMIC ATTACKS

TIAs are episodes of stroke symptoms that last only briefly; the standard definition of duration is <24 h, but most TIAs last <1 h. If a relevant brain infarction is identified on brain imaging, the clinical entity is now classified as stroke regardless of the duration of symptoms. A normal brain imaging study following a TIA does not rule out TIA; rather, the clinical syndrome is diagnostic. The causes of TIA are similar to those of ischemic stroke, but because TIAs may herald stroke, they are an important risk factor that should be considered urgently. TIAs may arise from emboli to the brain or from in situ thrombosis of an intracranial vessel. With a TIA, the occluded blood vessel reopens and neurologic function is restored.

The risk of stroke after a TIA is ~10–15% in the first 3 months, with most events occurring in the first 2 days. This risk can be directly estimated using the well-validated $ABCD^2$ score **(Table 438-5)**. Therefore, urgent evaluation and treatment are justified. Because etiologies for stroke and TIA are identical, evaluation for TIA should parallel that of stroke.

TREATMENT

Transient Ischemic Attack

The improvement characteristic of TIA is a contraindication to thrombolysis. However, because the risk of subsequent stroke in the first few hours and days following TIA is high, some physicians admit the patient to the hospital so a plasminogen activator can be rapidly administered if symptoms return. The combination of aspirin and clopidogrel was found to prevent stroke following TIA better than aspirin alone in a large Chinese randomized trial and the National Institutes of Health (NIH)–sponsored POINT trial. Failure to respond to the combination of aspirin and clopidogrel is linked to carriage of a common CYP2C19 polymorphism that leads to poor metabolism of clopidogrel into its active form. This mutation is common, particularly in Asians. Recently, ticagrelor, 180-mg loading dose and then 90 mg twice daily, was tested in combination with aspirin compared to aspirin alone, and this also showed benefit in preventing stroke; this dual antiplatelet regimen may be favored because of the lack of genetic heterogeneity in platelet inhibition.

TABLE 438-5 Risk of Stroke Following Transient Ischemic Attack: The $ABCD^2$ Score

CLINICAL FACTOR	SCORE
A: Age ≥60 years	1
B: SBP >140 mmHg or DBP >90 mmHg	1
C: Clinical symptoms	
Unilateral weakness	2
Speech disturbance without weakness	1
D: Duration	
>60 min	2
10–59 min	1
D: Diabetes (oral medications or insulin)	1
TOTAL SCORE	**SUM EACH CATEGORY**
$ABCD^2$ Score Total	3-Month Rate of Stroke (%)[a]
0	0
1	2
2	3
3	3
4	8
5	12
6	17
7	22

[a] Data ranges are from five cohorts.

Abbreviations: DBP, diastolic blood pressure; SBP, systolic blood pressure.

Source: Data from SC Johnston et al: Validation and refinement of scores to predict very early stroke risk after transient ischaemic attack. Lancet 369:283, 2007.

Primary and Secondary Prevention of Stroke and TIA

GENERAL PRINCIPLES

Many medical and surgical interventions, as well as lifestyle modifications, are available for preventing stroke. Some of these can be widely applied because of their low cost and minimal risk; others are expensive and carry substantial risk but may be valuable for selected high-risk patients. Identification and control of modifiable risk factors, and especially hypertension, is the best strategy to reduce the burden of stroke, and the total number of strokes could be reduced substantially by these means (Table 438-4).

ATHEROSCLEROSIS RISK FACTORS

The relationship of various factors to the risk of atherosclerosis is described in **Chaps. 244 and 245**. Older age, diabetes mellitus, hypertension, tobacco smoking, abnormal blood cholesterol (particularly, low high-density lipoprotein [HDL] and/or elevated low-density lipoprotein [LDL]), lipoprotein (a) excess, and other factors are either proven or probable risk factors for ischemic stroke, largely by their link to atherosclerosis. Risk of stroke is much greater in those with prior stroke or TIA. Many cardiac conditions predispose to stroke, including atrial fibrillation and recent MI. Oral contraceptives and hormone replacement therapy increase stroke risk, and although rare, certain inherited and acquired hypercoagulable states predispose to stroke.

Hypertension is the most significant of the risk factors; in general, all hypertension should be treated to a target of <130/80 mmHg. Recent data (the Systolic Blood Pressure Intervention

Trial—SPRINT) suggest that lowering systolic blood pressure <120 mmHg reduces stroke and heart attack by 43% compared to systolic blood pressure <140 mmHg, without an increased risk of syncope or falls, although patients with a history of stroke were specifically excluded from this study. The presence of known cerebrovascular disease is not a contraindication to treatment aimed at achieving normotension. Data are particularly strong in support of thiazide diuretics and angiotensin-converting enzyme inhibitors.

Several trials have confirmed that statin drugs reduce the risk of stroke even in patients without elevated LDL or low HDL. The Stroke Prevention by Aggressive Reduction in Cholesterol Levels (SPARCL) trial showed benefit in secondary stroke reduction for patients with recent stroke or TIA who were prescribed atorvastatin, 80 mg/d. The primary prevention trial, Justification for the Use of Statins in Prevention: An Intervention Trial Evaluating Rosuvastatin (JUPITER), found that patients with an elevated C-reactive protein benefitted by daily use of this statin, despite LDL <130 mg/dL. Primary stroke occurrence was reduced by 51% (hazard ratio, 0.49; $p = .004$), and there was no increase in the rates of intracranial hemorrhage. Meta-analysis has also supported a primary treatment effect for statins given acutely for ischemic stroke. A serum LDL <70 mg/dL lowers recurrent stroke risk better than an LDL of 90–110 mg/dL. Therefore, a statin should be considered in all patients with prior ischemic stroke. Tobacco smoking should be discouraged in all patients **(Chap. 465)**. The use of pioglitazone (an agonist of peroxisome proliferator-activated receptor gamma) in patients with type 2 diabetes and previous stroke does not lower stroke, MI, or vascular death rates but is effective in lowering vascular events in patients with stroke and prediabetes or insulin resistance alone. Diabetes prevention is likely the most effective strategy for primary and secondary stroke prevention.

ANTIPLATELET AGENTS FOR STROKE PREVENTION

Platelet antiaggregation agents can prevent atherothrombotic events, including TIA and stroke, by inhibiting the formation of intra-arterial platelet aggregates. These can form on diseased arteries, induce thrombus formation, and occlude or embolize into the distal circulation. Aspirin, clopidogrel, the combination of aspirin plus extended-release dipyridamole, and recently ticagrelor are the antiplatelet agents most commonly used for this purpose. Ticagrelor has not been found to be better than aspirin for stroke prevention except in combination with aspirin following TIA.

Aspirin is the most widely studied antiplatelet agent. Aspirin acetylates platelet cyclooxygenase, which irreversibly inhibits the formation in platelets of thromboxane A_2, a platelet aggregating and vasoconstricting prostaglandin. This effect is permanent and lasts for the usual 8-day life of the platelet. Paradoxically, aspirin also inhibits the formation in endothelial cells of prostacyclin, an antiaggregating and vasodilating prostaglandin. This effect is transient. As soon as aspirin is cleared from the blood, the nucleated endothelial cells again produce prostacyclin. Aspirin in low doses given once daily inhibits the production of thromboxane A_2 in platelets without substantially inhibiting prostacyclin formation. Higher doses of aspirin have not been proven to be more effective than lower doses.

Clopidogrel and ticagrelor block the adenosine diphosphate (ADP) receptor on platelets and thus prevent the cascade resulting in activation of the glycoprotein IIb/IIIa receptor that leads to fibrinogen binding to the platelet and consequent platelet aggregation. Clopidogrel can cause rash and, in rare instances, thrombotic thrombocytopenic purpura. The Clopidogrel versus Aspirin in Patients at Risk of Ischemic Events (CAPRIE) trial, which led to U.S. Food and Drug Administration (FDA) approval, found that it was only marginally more effective than aspirin in reducing risk of stroke. The Management of Atherothrombosis with Clopidogrel in High-Risk Patients (MATCH) trial was a large multicenter, randomized, double-blind study that compared clopidogrel in combination with aspirin to clopidogrel alone in the secondary prevention of TIA or stroke. The MATCH trial found no difference in TIA or stroke prevention with this combination but did show a small but significant increase in major bleeding complications (3 vs 1%). In the Clopidogrel for High Atherothrombotic Risk and Ischemic Stabilization, Management, and Avoidance (CHARISMA) trial, which included a subgroup of patients with prior stroke or TIA along with other groups at high risk of cardiovascular events, there was no benefit of clopidogrel combined with aspirin compared to aspirin alone. Lastly, the SPS3 trial looked at the long-term combination of clopidogrel and aspirin versus clopidogrel alone in small-vessel stroke and found no improvement in stroke prevention and a significant increase in both hemorrhage and death. Thus, the long-term use of clopidogrel in combination with aspirin is not recommended for stroke prevention.

The short-term combination of clopidogrel with aspirin may be effective in preventing second stroke, however. A large trial of Chinese patients enrolled within 24 h of TIA or minor ischemic stroke found that a clopidogrel-aspirin regimen (clopidogrel 300 mg load then 75 mg/d with aspirin 75 mg for the first 21 days) was superior to aspirin (75 mg/d) alone, with 90-day stroke risk decreased from 11.7 to 8.2% ($p < .001$) and no increase in major hemorrhage. This benefit was limited to those not carrying the CYP2C19 polymorphism associated with clopidogrel hypometabolism. An international NIH-sponsored trial demonstrated similar results; therefore, the combination of aspirin and clopidogrel should be administered for TIA or minor ischemic stroke for the first 21 days before switching to monotherapy.

A recent study of oral ticagrelor plus aspirin versus aspirin alone has shown similar benefits in secondary stroke reduction and carries the likely advantage that ticagrelor's antiplatelet effect is not genetically variable, as is the case with clopidogrel.

Dipyridamole is an antiplatelet agent that inhibits the uptake of adenosine by a variety of cells, including those of the vascular endothelium. The accumulated adenosine is an inhibitor of aggregation. At least in part through its effects on platelet and vessel wall phosphodiesterases, dipyridamole also potentiates the antiaggregatory effects of prostacyclin and nitric oxide produced by the endothelium and acts by inhibiting platelet phosphodiesterase, which is responsible for the breakdown of cyclic AMP. The resulting elevation in cyclic AMP inhibits aggregation of platelets. Dipyridamole is erratically absorbed depending on stomach pH, but a newer formulation combines timed-release dipyridamole, 200 mg, with aspirin, 25 mg, and has better oral bioavailability. This combination drug was studied in three trials. The European Stroke Prevention Study (ESPS) II showed efficacy of both 50 mg/d of aspirin and extended-release dipyridamole in preventing stroke and a significantly better risk reduction when the two agents were combined. The open-label ESPRIT (European/Australasian Stroke Prevention in Reversible Ischaemia Trial) trial confirmed the ESPS-II results. After 3.5 years of follow-up, 13% of patients on aspirin and dipyridamole and 16% on aspirin alone (hazard ratio, 0.80; 95% CI, 0.66–0.98) met the primary outcome of death from all vascular causes. In the Prevention Regimen for Effectively Avoiding Second Strokes (PRoFESS) trial, the combination of extended-release dipyridamole and aspirin was compared directly with clopidogrel with and without the angiotensin receptor blocker telmisartan; there were no differences in the rates of second stroke (9% each) or degree of disability in patients with median follow-up of 2.4 years. Telmisartan also had no effect on these outcomes. This suggests that these antiplatelet regimens are similar and raises questions about default prescription of agents to block the angiotensin pathway in all stroke patients. The principal side effect of dipyridamole is headache. The combination capsule of extended-release dipyridamole and aspirin is approved for prevention of stroke.

Many large clinical trials have demonstrated clearly that most antiplatelet agents reduce the risk of all important vascular atherothrombotic events (i.e., ischemic stroke, MI, and death due to all vascular causes) in patients at risk for these events. The overall *relative* reduction in risk of nonfatal stroke is ~25–30% and of all vascular events is ~25%. The *absolute* reduction varies considerably, depending on the patient's risk. Individuals at very low risk for stroke seem to experience the same relative reduction, but their risks may be so low that the "benefit" is meaningless. Conversely, individuals with a 10–15% risk of vascular events per year experience a reduction to ~7.5–11%.

Aspirin is inexpensive, can be given in low doses, and could be recommended for all adults to prevent both stroke and MI. However, it causes epigastric discomfort, gastric ulceration, and gastrointestinal hemorrhage, which may be asymptomatic or life threatening. Consequently, not every 40- or 50-year-old should be advised to take aspirin regularly because the risk of atherothrombotic stroke is extremely low and is outweighed by the risk of adverse side effects. Conversely, every patient who has experienced an atherothrombotic stroke or TIA and has no contraindication to antiplatelet therapy (or indication for anticoagulation) should be taking an antiplatelet agent regularly because the average annual risk of another stroke is 8–10%; another few percent will experience an MI or vascular death. Clearly, the likelihood of benefit far outweighs the risks of treatment.

The choice of antiplatelet agent and dose must balance the risk of stroke, the expected benefit, and the risk and cost of treatment. However, there are no definitive data, and opinions vary. Many authorities believe low-dose (30–75 mg/d) and high-dose (650–1300 mg/d) aspirin are about equally effective. Some advocate very low doses to avoid adverse effects, and still others advocate very high doses to be sure the benefit is maximal. Most physicians in North America recommend 81–325 mg/d, whereas most Europeans recommend 50–100 mg. Clopidogrel and extended-release dipyridamole plus aspirin are being increasingly recommended as first-line drugs for secondary prevention. Similarly, the choice of aspirin, clopidogrel, or dipyridamole plus aspirin must balance the fact that the latter are more effective than aspirin but the cost is higher, and this is likely to affect long-term patient adherence. The use of platelet aggregation studies in individual patients taking aspirin is controversial because of limited data.

In our practices, when considering antithrombotic therapy for secondary stroke prevention for noncardioembolic strokes and TIAs, we prescribe aspirin 81 mg/d in aspirin-I patients after an initial load of 325 mg. We add either clopidogrel (600-mg load, then 75 mg daily) or ticagrelor (180-mg load, then 90 mg twice daily) for TIA or minor stroke (NIHSS <5) for 21–30 days, followed by monotherapy with aspirin alone at 81 mg daily. We treat stroke due to intracranial atherosclerosis with aspirin 81 mg plus clopidogrel 75 mg daily for 3 months, after which time treatment is continued with aspirin alone.

ANTICOAGULATION THERAPY AND EMBOLIC STROKE PREVENTION

Several trials have shown that anticoagulation (international normalized ratio [INR] range, 2–3) in patients with chronic nonvalvular (nonrheumatic) atrial fibrillation (NVAF) prevents cerebral embolism and stroke and is safe. For primary prevention and for patients who have experienced stroke or TIA, anticoagulation with a vitamin K antagonist (VKA) reduces the risk by ~67%, which clearly outweighs the 1–3% risk per year of a major bleeding complication. VKAs are difficult to dose, their effects vary with dietary intake of vitamin K, and they require frequent blood monitoring of the PTT/INR. Several direct oral anticoagulants (DOACs) have recently been shown to be more convenient and efficacious for stroke prevention in NVAF. A randomized trial compared the oral thrombin inhibitor dabigatran to VKAs in a noninferiority trial to prevent stroke or systemic embolization in NVAF. Two doses of dabigatran were used: 110 mg/d and 150 mg/d. Both dose tiers of dabigatran were noninferior to VKAs in preventing second stroke and systemic embolization, and the higher dose tier was superior (relative risk, 0.66; 95% CI, 0.53–0.82; *p* <.001) and the rate of major bleeding was lower in the lower dose tier of dabigatran compared to VKAs. Dabigatran requires no blood monitoring to titrate the dose, and its effect is independent of oral intake of vitamin K. Newer oral factor Xa inhibitors have also been found to be equivalent or safer and more effective than VKAs in NVAF stroke prevention. In the Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE) trial, patients were randomized between apixaban, 5 mg twice daily, and dose-adjusted warfarin (INR 2–3). The combined endpoint of ischemic or hemorrhagic stroke or system embolism occurred in 1.27% of patients in the apixaban group and in 1.6% in the warfarin group (*p* <.001 for noninferiority and *p* <.01 for superiority). Major bleeding was 1% less, favoring apixaban (*p* <.001). Similar results were obtained in the Rivaroxaban Once Daily Oral Direct Factor Xa Inhibition Compared with Vitamin K Antagonism for Prevention of Stroke and Embolism Trial in Atrial Fibrillation (ROCKET-AF). In this trial, patients with NVAF were randomized to rivaroxaban versus warfarin: 1.7% of the factor Xa group and 2.2% of the warfarin group reached the endpoint of stroke and systemic embolism (*p* <.001 for noninferiority); intracranial hemorrhage was also lower with rivaroxaban. Finally, the factor Xa inhibitor edoxaban was also found to be noninferior to warfarin. Thus, oral factor Xa inhibitors are at least a suitable alternative to VKAs, for both primary and secondary prevention, and likely are superior both in efficacy and perhaps compliance. Recent FDA approval of a reversal agent for the Xa inhibitors apixaban and rivaroxaban (andexanet alfa) provides an antidote in the case of major bleeding. Idarucizumab has been available for reversal of dabigatran. Randomized trials have not demonstrated the superiority of anticoagulants over antiplatelet medications for strokes that appear embolic without a clear source, even when limited to the subset with evidence of an atrial cardiopathy.

For patients who cannot take anticoagulant medications, clopidogrel plus aspirin was compared to aspirin alone in the Atrial Fibrillation Clopidogrel Trial with Irbesartan for Prevention of Vascular Events (ACTIVE-A). Clopidogrel combined with aspirin was more effective than aspirin alone in preventing vascular events, principally stroke, but increased the risk of major bleeding (relative risk, 1.57; *p* <.001). Left atrial appendage occlusion followed by antiplatelet therapy was found to be noninferior to oral Xa inhibitors in patients at moderate to high risk of bleeding in a single trial. If confirmed, this may be a safer strategy than management with aspirin alone for these patients at high risk of atrial fibrillation–related stroke.

The decision to use anticoagulation for primary prevention is based primarily on risk factors (Table 438-3). The history of a TIA or stroke tips the balance in favor of anticoagulation regardless of other risk factors. Intermittent atrial fibrillation carries the same risk of stroke as chronic atrial fibrillation, and several ambulatory studies of seemingly "cryptogenic" stroke have found evidence of intermittent atrial fibrillation in nearly 20% of patients monitored for a few weeks. Interrogation of implanted pacemakers also confirms an association between subclinical atrial fibrillation and stroke risk. Therefore, for patients with otherwise cryptogenic embolic stroke (no evidence of any other cause for stroke), ambulatory monitoring for at least 30 days is a reasonable strategy to determine the best prophylactic therapy, and some patients may benefit from placement of longer-term implantable loop recorders.

Because of the high annual stroke risk in untreated rheumatic heart disease with atrial fibrillation, primary prophylaxis against stroke has not been studied in a double-blind fashion. These patients generally should receive long-term anticoagulation. Dabigatran and the oral Xa inhibitors have not been studied in this population.

Anticoagulation also reduces the risk of embolism in acute MI. Most clinicians recommend a 3-month course of anticoagulation when there is anterior Q-wave infarction, substantial left ventricular dysfunction, congestive heart failure, mural thrombosis, or atrial fibrillation. Oral anticoagulants are recommended long term if atrial fibrillation persists.

Stroke secondary to thromboembolism is one of the most serious complications of prosthetic heart valve implantation. The intensity of anticoagulation and/or antiplatelet therapy is dictated by the type of prosthetic valve and its location. Dabigatran may be less effective than warfarin, and the oral Xa inhibitors have not been studied in this population.

If the embolic source cannot be eliminated, anticoagulation should in most cases be continued indefinitely. Many neurologists recommend combining antiplatelet agents with anticoagulants for patients who "fail" anticoagulation (i.e., have another stroke or TIA), but the evidence basis for this is lacking.

It is our practice to prescribe apixaban 5 mg twice daily (adjusted to 2.5 mg twice daily if age, weight, and renal function criteria are met) for nonvalvular atrial fibrillation with CHA_2DS_2-VASc score of ≥2, aspirin 81 mg plus clopidogrel 75 mg daily for patients who cannot take oral anticoagulation, and VKAs for valvular atrial fibrillation or mechanical heart valve.

ANTICOAGULATION THERAPY AND NONCARDIOGENIC STROKE

Data do not support the use of long-term VKAs for preventing atherothrombotic stroke for either intracranial or extracranial cerebrovascular disease. The Warfarin-Aspirin Recurrent Stroke Study (WARSS) found no benefit of warfarin sodium (INR 1.4–2.8) over aspirin, 325 mg, for secondary prevention of stroke but did find a slightly higher bleeding rate in the warfarin group; a European study confirmed this finding. The Warfarin and Aspirin for Symptomatic Intracranial Disease (WASID) study (see below) demonstrated no benefit of warfarin (INR 2–3) over aspirin in patients with symptomatic intracranial atherosclerosis and found a higher rate of bleeding complications. Two trials testing factor Xa medications for prevention of embolic stroke of undetermined source (ESUS) failed to show benefit compared to treatment with antiplatelet medications and a third trial limited to ESUS patients with atrial cardiopathy had similar results. The oral factor Xa inhibitor apixaban was found to be noninferior to subcutaneous dalteparin for patients with cancer and venous thromboembolism; many oncologists are using Xa inhibitors to prevent second stroke in patients with malignancy.

It is our practice to prescribe aspirin for secondary stroke prevention in noncardiogenic cerebral embolism except for stroke associated with cancer (apixaban 5 mg twice daily) and the antiphospholipid syndrome (warfarin with target INR 2–3).

TREATMENT

Carotid Atherosclerosis

Carotid atherosclerosis can be removed surgically (endarterectomy), mitigated with endovascular stenting with or without balloon angioplasty, or using the transcarotid artery revascularization (TCAR) approach. Anticoagulation has not been directly compared with antiplatelet therapy for carotid disease.

SURGICAL THERAPY

Symptomatic carotid stenosis was studied in the North American Symptomatic Carotid Endarterectomy Trial (NASCET) and the European Carotid Surgery Trial (ECST). Both showed a substantial benefit for surgery in patients with stenosis of ≥70%. In NASCET, the average cumulative ipsilateral stroke risk at 2 years was 26% for patients treated medically and 9% for those receiving the same medical treatment plus a carotid endarterectomy. This 17% *absolute* reduction in the surgical group is a 65% *relative* risk reduction favoring surgery (Table 438-4). NASCET also showed a significant, although less robust, benefit for patients with 50–70% stenosis. ECST found harm for patients with stenosis <30% treated surgically.

A patient's risk of stroke and possible benefit from surgery are related to the presence of retinal versus hemispheric symptoms, degree of arterial stenosis, extent of associated medical conditions (of note, NASCET and ECST excluded "high-risk" patients with significant cardiac, pulmonary, or renal disease), institutional surgical morbidity and mortality, timing of surgery relative to symptoms, and other factors. A recent meta-analysis of the NASCET and ECST trials demonstrated that endarterectomy is most beneficial when performed within 2 weeks of symptom onset. In addition, benefit is more pronounced in patients >75 years, and men appear to benefit more than women.

In summary, a patient with recent symptomatic hemispheric ischemia, high-grade stenosis in the appropriate internal carotid artery, and an institutional perioperative morbidity and mortality rate of ≤6% generally should undergo carotid endarterectomy. If the perioperative stroke rate is >6% for any particular surgeon, however, the benefits of carotid endarterectomy are questionable.

The indications for surgical treatment of *asymptomatic carotid disease* have been clarified by the results of the Asymptomatic Carotid Atherosclerosis Study (ACAS) and the Asymptomatic Carotid Surgery Trial (ACST). ACAS randomized asymptomatic patients with ≥60% stenosis to medical treatment with aspirin or the same medical treatment plus carotid endarterectomy. The surgical group had a risk over 5 years for ipsilateral stroke (and any perioperative stroke or death) of 5.1%, compared to a risk in the medical group of 11%. Although this demonstrates a 53% *relative* risk reduction, the *absolute* risk reduction is only 5.9% over 5 years, or 1.2% annually (Table 438-4). Nearly one-half of the strokes in the surgery group were caused by preoperative angiograms. ACST randomized asymptomatic patients with >60% carotid stenosis to endarterectomy or medical therapy. The 5-year risk of stroke in the surgical group (including perioperative stroke or death) was 6.4%, compared to 11.8% in the medically treated group (46% relative risk reduction and 5.4% absolute risk reduction).

In both ACAS and ACST, the perioperative complication rate was higher in women, perhaps negating any benefit in the reduction of stroke risk within 5 years. It is possible that with longer follow-up, a clear benefit in women will emerge. At present, carotid endarterectomy in asymptomatic women remains particularly controversial.

In summary, the natural history of asymptomatic stenosis is an ~2% per year stroke rate, whereas symptomatic patients experience a 13% per year risk of stroke. Whether to recommend carotid revascularization for an asymptomatic patient is somewhat controversial and depends on many factors, including patient preference, degree of stenosis, age, gender, and comorbidities. Medical therapy for reduction of atherosclerosis risk factors, including cholesterol-lowering agents and antiplatelet medications, is generally recommended for patients with asymptomatic carotid stenosis. As with atrial fibrillation, it is imperative to counsel the patient about TIAs so that therapy can be revised if symptoms develop.

ENDOVASCULAR THERAPY

Balloon angioplasty coupled with stenting is one option to open stenotic carotid arteries and maintain their patency. These techniques can treat carotid stenosis not only at the bifurcation but also near the skull base and in the intracranial segments. The Stenting and Angioplasty with Protection in Patients at High Risk for Endarterectomy (SAPPHIRE) trial randomized high-risk patients (defined as patients with clinically significant coronary or pulmonary disease, contralateral carotid occlusion, restenosis after endarterectomy, contralateral laryngeal-nerve palsy, prior radical neck surgery or radiation, or age >80) with symptomatic carotid stenosis >50% or asymptomatic stenosis >80% to either stenting combined with a distal emboli-protection device or endarterectomy. The risk of death, stroke, or MI within 30 days and ipsilateral stroke or death within 1 year was 12.2% in the stenting group and 20.1% in the endarterectomy group ($p = .055$), suggesting that stenting is at the very least comparable to endarterectomy as a treatment option for this patient group at high risk of surgery. However, the outcomes with both interventions may not have been better than leaving the carotid stenoses untreated, particularly for the asymptomatic patients, and much of the benefit seen in the stenting group was due to a reduction in periprocedure MI. Two randomized trials comparing stents to endarterectomy in lower-risk patients have been published. The Carotid Revascularization Endarterectomy versus Stenting Trial (CREST) enrolled patients with either asymptomatic or symptomatic stenosis. The 30-day risk of stroke was 4.1% in the stent group and 2.3% in the surgical group, but the 30-day risk of MI was 1.1% in the stent group and 2.3% in the surgery group, suggesting relative equivalence of risk between the procedures. At median follow-up of 2.5 years, the combined endpoint of stroke, MI, and death was the same (7.2% stent vs 6.8% surgery) and remained so at 10-year follow-up. The rate of restenosis at 2 years was also similar in both groups. The International Carotid Stenting Study (ICSS) randomized symptomatic patients to stents versus endarterectomy and found a different result: at 120 days, the incidence of stroke, MI, or death was 8.5% in the stenting group versus 5.2% in the endarterectomy group ($p = .006$). At median follow-up of 5 years, these differences were no longer significant except a small increase in nondisabling stroke in the stenting group but no change in the average disability. In meta-analysis, carotid endarterectomy (CEA) is less morbid in older patients (aged ≥70) than is stenting. Recently, transcarotid artery revascularization (TCAR), which involves the reversal of blood flow during an angioplasty and stenting procedure, has been offered as an alternative to transfemoral carotid artery stenting or when CEA presents high risks. Investigation is ongoing in asymptomatic patients to compare medical therapy to stenting and CEA. This will likely answer how well medical patients do with more modern medical therapy (statins, close blood pressure control, and lifestyle modification).

BYPASS SURGERY

Extracranial-to-intracranial (EC-IC) bypass surgery has been proven ineffective for atherosclerotic stenoses that are inaccessible to conventional CEA. In patients with recent stroke, an associated carotid occlusion, and evidence of inadequate perfusion of the brain as measured with positron emission tomography, no benefit from EC-IC bypass was found in a trial stopped for futility.

PATENT FORAMEN OVALE

In patients with PFO and/or atrial septal aneurysm with an embolic stroke and no other cause identified, three randomized trials using various endovascular closure devices individually and in meta-analysis reported a significant (1% per year) reduction in second stroke compared to antiplatelet agents. If the neurologic opinion is that no other source of stroke is identified and consultation with a cardiologist knowledgeable about PFO closure supports intervention, we recommend endovascular PFO closure.

INTRACRANIAL ATHEROSCLEROSIS

The WASID trial randomized patients with symptomatic stenosis (50–99%) of a major intracranial vessel to either high-dose aspirin (1300 mg/d) or warfarin (target INR, 2.0–3.0), with a combined primary endpoint of ischemic stroke, brain hemorrhage, or death from vascular cause other than stroke. The trial was terminated early because of an increased risk of adverse events related to warfarin anticoagulation. With a mean follow-up of 1.8 years, the primary endpoint was seen in 22.1% of patients in the aspirin group and 21.8% of the warfarin group. Death from any cause was seen in 4.3% of the aspirin group and 9.7% of the warfarin group; 3.2% of patients on aspirin experienced major hemorrhage, compared to 8.3% of patients taking warfarin.

Intracranial stenting of intracranial atherosclerosis was found to be dramatically harmful compared to aspirin in the Stenting and Aggressive Medical Management for Preventing Recurrent Stroke in Intracranial Stenosis (SAMMPRIS) trial. This trial enrolled newly symptomatic TIA or minor stroke patients with associated 70–99% intracranial stenosis to primary stenting with a self-expanding stent or to medical management. Both groups received clopidogrel, aspirin, statin, and aggressive control of blood pressure. The endpoint of stroke or death occurred in 14.7% of the stented group and 5.8% of the medically treated groups ($p = .002$). This low rate of second stroke was significantly lower than in the WASID trial and suggests that aggressive medical management had a marked influence on secondary stroke risk. A concomitant study of balloon-expandable stenting was halted early at 125 patients because of the negative SAMMPRIS results and due to harm. Therefore, routine use of intracranial stenting is harmful, and medical therapy is superior for intracranial atherosclerosis.

Dural Sinus Thrombosis Limited evidence exists to support short-term use of anticoagulants, regardless of the presence of intracranial hemorrhage, for venous infarction following sinus thrombosis. The long-term outcome for most patients, even those with intracerebral hemorrhage, is excellent.

Acknowledgment
The authors acknowledge the contributions of S. Claiborne Johnston to earlier editions of this chapter.

FURTHER READING

Goyal M et al: Endovascular thrombectomy after large-vessel ischaemic stroke: A meta-analysis of individual patient data from five randomised trials. Lancet 387:1723, 2016.

Grotta JC et al: Prospective, multicenter, controlled trial of mobile stroke units. N Engl J Med 385:971, 2021.

Jiang H et al: An updated meta-analysis on the clinical outcomes of percutaneous left atrial appendage closure versus direct oral anticoagulation in patients with atrial fibrillation. J Am Coll Cardiol 200:135, 2023.

Joglar JA et al: 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 149:e1, 2024.

Powers WJ et al: Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 50:e344, 2019.

Saver JL et al: Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: A meta-analysis. JAMA 316:1279, 2016.

Sprint Research Group et al: A randomized trial of intensive versus standard blood-pressure control. N Engl J Med 373:2103, 2015.

Torbey MT et al: Evidence-based guidelines for the management of large hemispheric infarction: A statement for health care professionals from the Neurocritical Care Society and the German Society for Neuro-intensive Care and Emergency Medicine. Neurocrit Care 22:146, 2015.

439

Intracerebral Hemorrhage

J. Claude Hemphill, III,
Edilberto Amorim, Wade S. Smith

Intracerebral hemorrhage (ICH) is a form of stroke **(see Chap. 437)**. Compared to ischemic stroke, patients with ICH are more likely to present with headache; however, brain imaging is required to distinguish these entities. CT imaging of the head is highly sensitive and specific for intracranial hemorrhage and determines the location(s) of bleeding. Hemorrhages are classified by their location and the underlying vascular pathology. ICH is defined as spontaneous hemorrhage directly into the brain parenchyma and will be considered here along with intracranial vascular anomalies such as arteriovenous malformations (AVMs) of the brain. Other categories of intracranial hemorrhage include bleeding into subarachnoid, subdural, or epidural spaces, usually caused by trauma **(Chap. 454)**, and subarachnoid hemorrhage due to trauma or the rupture of an intracranial aneurysm **(Chap. 440)**.

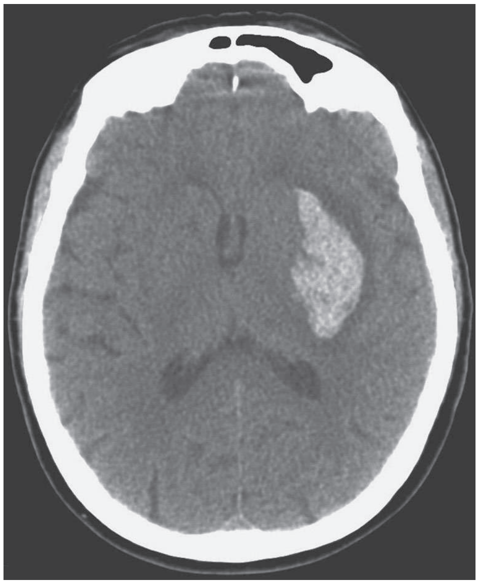

FIGURE 439-1 Hypertensive intracerebral hemorrhage. Transaxial noncontrast computed tomography scan through the region of the basal ganglia reveals a hematoma involving the left putamen in a patient with rapidly progressive onset of right hemiparesis.

DIAGNOSIS

Intracranial hemorrhage is often identified on noncontrast computed tomography (CT) imaging of the head during the acute evaluation of stroke. Because CT is more widely available and may be logistically easier to perform than magnetic resonance imaging (MRI), CT imaging is generally the preferred method for acute stroke evaluation **(Fig. 439-1)**. The location of the hemorrhage narrows the differential diagnosis to a few entities. **Table 439-1** lists the causes and anatomic spaces involved in intracranial hemorrhages.

TABLE 439-1 Causes of Intracerebral Hemorrhage (ICH)

CAUSE	LOCATION	COMMENTS
Primary ICH		
Cerebral amyloid angiopathy	Lobar	Degenerative disease of intracranial vessels; associated with dementia, rare in patients <60 years
Coagulopathy	Any	Risk for hematoma expansion
Drug	Any, lobar, subarachnoid	Cocaine, amphetamine
Hypertension	Putamen, globus pallidus, thalamus, cerebellar hemisphere, pons	Chronic hypertension produces hemorrhage from small (~30–100 μm) vessels in these regions
Secondary ICH		
Aneurysm	Subarachnoid, intraparenchymal, rarely subdural	Mycotic and nonmycotic forms of aneurysms
Arteriovenous malformation	Lobar, intraventricular, subarachnoid	Risk is ~2–4% per year for bleeding if previously unruptured
Capillary telangiectasias	Usually brainstem	Rare cause of hemorrhage
Cavernous angioma	Intraparenchymal	Multiple cavernous angiomas linked to mutations in *KRIT1*, *CCM2*, and *PDCD10* genes
Dural arteriovenous fistula	Lobar, subarachnoid	Produces bleeding from venous hypertension
Dural sinus thrombosis	Along sagittal sinus, posterior temporal/inferior parietal	Sagittal sinus thrombosis can cause hemispheric parasagittal hemorrhage with edema; vein of Labbé occlusion from transverse sinus occlusion produces posterior temporal and/or inferior parietal hemorrhage
Metastatic or primary brain tumors	Lobar	Lung, choriocarcinoma, melanoma, renal cell carcinoma, thyroid, hepatocellular carcinoma, and pilocytic astrocytoma are more commonly associated with bleeding complications
Transformation of prior ischemic infarction	Basal ganglion, subcortical regions, lobar	Occurs in a significant proportion of ischemic strokes, more commonly in large hemispheric infarctions; is symptomatic in 3–9% of patients undergoing acute intervention

EMERGENCY MANAGEMENT

Close attention should be paid to airway management because deterioration in the level of consciousness is common and often progressive. The initial blood pressure should be maintained until the results of the CT scan are reviewed and demonstrate ICH. A higher blood pressure may promote hematoma expansion, but it remains unclear if lowering of blood pressure reduces hematoma growth. Recent clinical trials have shown that systolic blood pressure (SBP) can be safely lowered acutely and rapidly to <140 mmHg in patients with spontaneous ICH whose initial SBP was 150–220 mmHg. The INTERACT2 trial was a large phase 3 clinical trial to address the effect of acute blood pressure lowering on ICH functional outcome. INTERACT2 randomized patients with spontaneous ICH within 6 h of onset and a baseline SBP of 150–220 mmHg to two different SBP targets (<140 and <180 mmHg). In those with the target SBP <140 mmHg, 52% had an outcome of death or major disability at 90 days compared with 55.6% of those with a target SBP <180 mmHg (p = .06). There was a significant shift to improved outcomes in the lower blood pressure arm, whereas both groups had a similar mortality. ATACH2 was a similarly designed clinical trial that assessed the same blood pressure targets but demonstrated no difference in outcome between groups; however, aggressive blood pressure lowering did increase renal adverse events. Current U.S. and European guidelines emphasize that blood pressure lowering to a target SBP is likely safe and possibly beneficial. While the specific optimal target remains a point of debate, the most recent American Heart Association/American Stroke Association guidelines for the management of spontaneous ICH endorse achieving and maintaining a target SBP of 130–150 mmHg in these patients to avoid unintended hypoperfusion. It is unclear whether these clinical trial results apply to patients who have higher SBP on presentation or who are deeply comatose with possible elevated intracranial pressure (ICP). In patients who have ICP monitors in place, maintaining the cerebral perfusion pressure (mean arterial pressure [MAP] minus ICP) of 60 to ≥70 mmHg is reasonable, depending on the individual patient's cerebral autoregulation status **(Chap. 318)**. Blood pressure should be lowered using IV drugs with less cerebral vasodilating action such as nicardipine, clevidipine, labetalol, or esmolol. Patients with radiographic evidence of hydrocephalus or cerebellar ICH with depressed mental status should undergo urgent neurosurgical evaluation; these patients require close monitoring because they can deteriorate rapidly. Based on the clinical examination and CT findings, further imaging studies may be necessary, including MRI or conventional x-ray angiography. Stuporous or comatose patients with clinical and imaging signs of herniation can be presumptively treated for elevated ICP with tracheal intubation and sedation, administration of osmotic diuretics such as mannitol or hypertonic saline, and elevation of the head of the bed while surgical consultation is obtained **(Chap. 318)**. Rapid reversal of coagulopathy ideally within 1 h of presentation and consideration of surgical evacuation of the hematoma (detailed below) are two other principal aspects of initial emergency management.

INTRACEREBRAL HEMORRHAGE

ICH accounts for ~10% of all strokes, and ~35–45% of patients die within the first month. Incidence rates are particularly high in Asian and Black patient groups. Hypertension, coagulopathy, sympathomimetic drugs (cocaine, methamphetamine), and cerebral amyloid angiopathy (CAA) cause most of these hemorrhages. Advanced age, heavy alcohol, and low-dose aspirin use in those without symptomatic cardiovascular disease increase ICH risk, and cocaine or methamphetamine use is one of the most important causes in the young.

Hypertensive ICH • PATHOPHYSIOLOGY Hypertensive ICH usually results from spontaneous rupture of a small penetrating artery deep in the brain. The most common sites are the basal ganglia (especially the putamen), thalamus, cerebellum, and pons. The small arteries in these areas seem most prone to hypertension-induced vascular injury. When hemorrhages occur in other brain areas or in nonhypertensive patients, greater consideration should be given to other causes such as hemorrhagic disorders, neoplasms, vascular malformations, vasculitis, and CAA. The hemorrhage may be small, or a large clot may form and compress adjacent tissue, causing herniation and death. Blood may also dissect into the ventricular space, which substantially increases morbidity and may cause hydrocephalus.

Most hypertensive ICHs initially develop over 30–90 min, whereas those associated with anticoagulant therapy may evolve for as long as 24–48 h. It is now recognized that about a third of patients even with no coagulopathy may have significant hematoma expansion within the first day. Within 48 h, macrophages begin to phagocytize the hemorrhage at its outer surface. After 1–6 months, the hemorrhage is generally resolved to a slitlike cavity lined with a glial scar and hemosiderin-laden macrophages.

CLINICAL MANIFESTATIONS ICH generally presents as the abrupt onset of a focal neurologic deficit. Seizures are uncommon on presentation but may occur in 6–15% of patients within the first 3 days. Although clinical symptoms may be maximal at onset, more commonly, the focal deficit worsens over 30–90 min and is associated with a diminishing level of consciousness and signs of increased ICP such as headache and vomiting.

The putamen is the most common site for hypertensive hemorrhage, and the adjacent internal capsule is usually damaged (Fig. 439-1). Contralateral hemiparesis is therefore the sentinel sign. When mild, the face sags on one side over 5–30 min, speech becomes slurred, the arm and leg gradually weaken, and the eyes deviate away from the side of the hemiparesis. The paralysis may worsen until the affected limbs become flaccid or extend rigidly. When hemorrhages are large, drowsiness gives way to stupor as signs of upper brainstem compression appear. Coma ensues, accompanied by deep, irregular, or intermittent respiration, a dilated and fixed ipsilateral pupil, and decerebrate rigidity. Edema in adjacent brain tissue may cause progressive deterioration over 24–96 h.

Thalamic hemorrhages may also produce a contralateral hemiplegia or hemiparesis from pressure on, or dissection into, the adjacent internal capsule. A prominent sensory deficit involving all modalities is usually present. Aphasia, often with preserved verbal repetition, may occur after hemorrhage into the dominant thalamus, and constructional apraxia or mutism occurs in some cases of nondominant hemorrhage. There may also be a homonymous visual field defect. Thalamic hemorrhages cause several typical ocular disturbances by extension inferiorly into the upper midbrain. These include deviation of the eyes downward and inward so that they appear to be looking at the nose, unequal pupils with absence of light reaction, skew deviation with the eye opposite the hemorrhage displaced downward and medially, ipsilateral Horner's syndrome, absence of convergence, paralysis of vertical gaze, and retraction nystagmus. Patients may later develop a chronic, contralateral pain syndrome (Déjérine-Roussy syndrome).

In pontine hemorrhages, deep coma with quadriplegia often occurs over a few minutes. Typically, there is prominent decerebrate rigidity and "pinpoint" (1 mm) pupils that react to light. There is impairment of reflex horizontal eye movements evoked by head turning (doll's-head or oculocephalic maneuver) or by irrigation of the ears with ice water **(Chap. 30)**. Hyperpnea, severe hypertension, and hyperhidrosis are common. Most patients with deep coma from pontine hemorrhage ultimately die or develop a locked-in state, but small hemorrhages are compatible with survival and significant recovery.

Cerebellar hemorrhages usually develop over several hours and are characterized by occipital headache, repeated vomiting, and ataxia of gait. In mild cases, there may be no other neurologic signs except for gait ataxia. Dizziness or vertigo may be prominent. There is often paresis of conjugate lateral gaze toward the side of the hemorrhage, forced deviation of the eyes to the opposite side, or an ipsilateral sixth nerve palsy. Less frequent ocular signs include blepharospasm, involuntary closure of one eye, ocular bobbing, and skew deviation. Dysarthria and dysphagia may occur. As the hours pass, the patient often becomes stuporous and then comatose from brainstem compression or obstructive hydrocephalus; immediate surgical evacuation before severe brainstem compression occurs may be lifesaving. Hydrocephalus from fourth ventricle compression can be relieved by external ventricular drainage; however, in this situation, definitive hematoma evacuation is recommended rather than treatment with ventricular drainage alone. If the deep cerebellar nuclei are spared, full recovery is common.

Lobar hemorrhages usually present with symptoms related to the specific site of origin. The major neurologic deficit with an occipital hemorrhage is hemianopsia; with a left temporal hemorrhage, aphasia and confusion; with a parietal hemorrhage, hemisensory loss; and with frontal hemorrhage, arm weakness. Large hemorrhages may be associated with stupor or coma if they compress the thalamus or midbrain. Most patients with lobar hemorrhages have focal headaches, and more than one-half vomit or are drowsy. Seizures may occur.

Other Causes of ICH and Intracranial Hemorrhage CAA is a disease of the elderly in which arteriolar degeneration occurs and amyloid is deposited in the walls of the cerebral arteries. Amyloid angiopathy causes both single and recurrent lobar hemorrhages and is probably the most common cause of lobar hemorrhage in the elderly. It accounts for some intracranial hemorrhages associated with IV thrombolysis given for myocardial infarction. This disorder can be suspected in patients who present with multiple hemorrhages (and infarcts) over several months or years or in patients with "microbleeds" in the cortex, seen on brain MRI sequences sensitive for hemosiderin (susceptibility weighted imaging), but it is definitively diagnosed by pathologic demonstration of Congo red staining of amyloid in cerebral vessels. The ε2 and ε4 allelic variations of the apolipoprotein E gene are associated with increased risk of recurrent lobar hemorrhage and may therefore be markers of amyloid angiopathy. Positron emission tomography imaging can image amyloid-beta deposits in CAA using specific antibody labels and may be helpful in diagnosing CAA noninvasively. Although cerebral biopsy is the most definitive method of diagnosis, evidence of inflammation on lumbar puncture should prompt consideration of CAA-associated vasculitis as an underlying cause, and oral glucocorticoids may be beneficial. Noninflammatory CAA has no specific treatment. Oral anticoagulants are typically avoided.

Cocaine and *methamphetamine* are frequent causes of stroke in young (age <45 years) patients. ICH, ischemic stroke, and subarachnoid hemorrhage (SAH) are all associated with stimulant use. Angiographic findings vary from completely normal arteries to large-vessel occlusion or stenosis, vasospasm, or changes consistent with vasculopathy. The mechanism of sympathomimetic-related stroke is not known, but cocaine enhances sympathetic activity causing acute, sometimes severe, hypertension, and this may lead to hemorrhage. Slightly more than one-half of stimulant-related intracranial hemorrhages are intracerebral and the rest are subarachnoid. In cases of SAH, a saccular aneurysm is usually identified. Presumably, acute hypertension causes aneurysmal rupture.

Head injury often causes intracranial hemorrhage. The common sites are intraparenchymal (especially temporal and inferior frontal lobes) and into the subarachnoid, subdural, and epidural spaces. Trauma must be considered in any patient with an unexplained acute neurologic deficit (hemiparesis, stupor, or confusion), particularly if the deficit occurred in the context of a fall **(Chap. 454)**.

Intracranial hemorrhages associated with *anticoagulant therapy* can occur at any location; they are often lobar or subdural. Anticoagulant-related ICHs may continue to evolve over 24–48 h, especially if coagulopathy is insufficiently reversed. Coagulopathy and thrombocytopenia should be reversed rapidly, as discussed below. ICH associated with *hematologic disorders* (leukemia, aplastic anemia, thrombocytopenic purpura) can occur at any site and may present as multiple ICHs. Skin and mucous membrane bleeding may be evident and offers a diagnostic clue.

Hemorrhage into a *brain tumor* may be the first manifestation of neoplasm. Choriocarcinoma, malignant melanoma, renal cell carcinoma, and thyroid, lung, and hepatocellular carcinoma are among the most common metastatic tumors associated with ICH. Pilocytic astrocytoma and glioblastoma multiforme in adults and medulloblastoma in children may also have areas of ICH.

Hypertensive encephalopathy is a complication of malignant hypertension. In this acute syndrome, severe hypertension is associated with headache, nausea, vomiting, convulsions, confusion, stupor, and coma. Focal or lateralizing neurologic signs, either transitory or permanent, may occur but are infrequent and therefore suggest some other vascular disease (hemorrhage, embolism, or atherosclerotic thrombosis). There may be retinal hemorrhages, exudates, papilledema (hypertensive retinopathy), and evidence of renal and cardiac disease. MRI brain imaging shows a pattern of typically posterior (occipital > frontal) brain edema. The hypertension may be essential or due to chronic renal disease, acute glomerulonephritis, acute toxemia of pregnancy, pheochromocytoma, or other causes. Lowering the blood pressure reverses the process, but stroke can occur, especially if blood pressure is lowered too rapidly. Neuropathologic examination reveals multifocal to diffuse cerebral edema and hemorrhages of various sizes from petechial to massive. Microscopically, there is necrosis of arterioles, minute cerebral infarcts, and hemorrhages. The terms *hypertensive encephalopathy* and *posterior reversible encephalopathy syndrome* **(Chap. 318)** should be reserved for this syndrome and not for chronic recurrent headaches, dizziness, recurrent transient ischemic attacks, or small strokes that often occur in association with high blood pressure. Distinguishing hypertensive encephalopathy with ICH from hypertensive ICH is important since aggressive lowering of SBP to 130–150 mmHg acutely is often considered in hypertensive ICH, but less aggressive measures should be used in hypertensive encephalopathy. Having no alteration in mental status or other prodrome prior to the ICH favors hypertensive ICH as the disease.

Primary intraventricular hemorrhage is rare and should prompt investigation for an underlying vascular anomaly. Sometimes bleeding begins within the periventricular substance of the brain and dissects into the ventricular system without leaving signs of intraparenchymal hemorrhage. Alternatively, bleeding can arise from periependymal veins. Vasculitis, usually polyarteritis nodosa or lupus erythematosus, can produce hemorrhage in any region of the central nervous system; most hemorrhages are associated with hypertension, but the arteritis itself may cause bleeding by disrupting the vessel wall. Nearly one-half of patients with primary intraventricular hemorrhage have identifiable bleeding sources seen using conventional angiography.

Venous sinus thrombosis **(Chap. 438)** causes cortical vein hypertension, cerebral edema, and venous infarction. This may progress to cause intracranial hemorrhage surrounding the region of the occluded cerebral venous sinus or within the drainage region of the vein of Labbé, producing a posterior temporal or inferior parietal hematoma. Despite the presence of hemorrhage, IV anticoagulation is indicated to reduce the venous hypertension and limit venous ischemia and further bleeding.

Sepsis can cause small petechial hemorrhages throughout the cerebral white matter. *Moyamoya disease* **(Chap. 438)**, mainly an occlusive arterial disease that causes ischemic symptoms, may on occasion produce ICH. Hemorrhages into the spinal cord are usually the result of an AVM, cavernous malformation, or metastatic tumor. *Epidural spinal hemorrhage* produces a rapidly evolving syndrome of spinal cord or nerve root compression **(Chap. 453)**. Spinal hemorrhages usually present with sudden back pain and some manifestation of myelopathy.

Laboratory and Imaging Evaluation Patients should have routine blood chemistries and hematologic studies. Specific attention to the platelet count, prothrombin time, partial thromboplastin time, and international normalized ratio is important to identify coagulopathy. CT imaging reliably detects acute focal hemorrhages in the supratentorial space. Rarely, very small pontine or medullary hemorrhages may not be well delineated because of motion and bone-induced artifact that obscure structures in the posterior fossa. After the first 2 weeks, x-ray attenuation values of clotted blood diminish until they become isodense with surrounding brain. Mass effect and edema may remain. In some cases, a surrounding rim of contrast enhancement appears after 2–4 weeks and may persist for months. MRI, although more sensitive for delineating posterior fossa lesions, is generally not necessary for primary diagnosis. MR angiography (MRA), CT angiography (CTA), and conventional x-ray angiography are used when the cause of intracranial hemorrhage is uncertain, particularly if the patient is young or not hypertensive and the hematoma is not in one of the usual sites for hypertensive hemorrhage. CTA or postcontrast CT imaging may reveal one or more small areas of enhancement within a hematoma; this "spot sign" is thought to represent ongoing bleeding. The presence of a spot sign is associated with an increased risk of hematoma expansion, increased mortality, and lower likelihood of favorable functional outcome. Because patients typically have focal neurologic signs and obtundation and often show signs of increased ICP, a lumbar puncture is generally unnecessary and should usually be avoided because it may induce cerebral herniation.

TREATMENT

Intracerebral Hemorrhage

ACUTE MANAGEMENT

After immediate attention to blood pressure and airway protection (see above), focus can switch to medical and surgical management. Approximately 40% of patients with a hypertensive ICH die, but survivors can have a good to complete recovery. The ICH Score **(Table 439-2)** is a validated clinical grading scale that is useful for stratification of mortality risk and clinical outcome. However, a specific ICH clinical grading scale should not be used to precisely prognosticate outcome because of the concern of creating a self-fulfilling prophecy of poor outcome if early aggressive care is withheld. Any identified coagulopathy should be corrected as soon as possible. For patients taking vitamin K antagonists (VKAs), rapid correction of coagulopathy can be achieved by infusing prothrombin complex concentrates (PCCs), which can be administered quickly, with

TABLE 439-2 The Intracerebral Hemorrhage Score

CLINICAL OR IMAGING FACTOR	POINT SCORE
Age	
<80 years	0
≥80 years	1
Hematoma Volume	
<30 cc	0
≥30 cc	1
Intraventricular Hemorrhage Present	
No	0
Yes	1
Infratentorial Origin of Hemorrhage	
No	0
Yes	1
Glasgow Coma Scale Score	
13–15	0
5–12	1
3–4	2
Total Score	0–6 Sum of each category above

Source: Reproduced with permission from JC Hemphill 3rd et al: The ICH score: A simple, reliable grading scale for intracerebral hemorrhage. Stroke 32:891, 2001.

vitamin K administered concurrently. Fresh frozen plasma (FFP) is an alternative, but since it requires larger fluid volumes and longer time to achieve adequate reversal than PCC, it is not recommended if PCC is available. Idarucizumab is a monoclonal antibody to dabigatran, and the administration of two doses reverses the anticoagulation effect of dabigatran quickly. The oral Xa inhibitors apixaban and rivaroxaban can be reversed with andexanet alfa. PCC may partially reverse the effects of oral factor Xa inhibitors and are reasonable to administer if andexanet alfa is not available. When ICH is associated with thrombocytopenia (platelet count <50,000/μL), transfusion of fresh platelets is indicated. A clinical trial of platelet transfusions in patients with ICH and without thrombocytopenia who were taking antiplatelet drugs showed no benefit and possible harm.

Hematomas may expand for several hours following the initial hemorrhage, even in patients without coagulopathy. The precise mechanism is unclear. A phase 3 trial of treatment with recombinant factor VIIa reduced hematoma expansion; however, clinical outcomes were not improved, so use of this drug is not recommended. The administration of tranexamic acid was not found to alter outcome in a large randomized trial. Blood pressure lowering has been considered due to the theoretical risk of acutely elevated blood pressure on hematoma expansion, although clinical trials did not find a difference in hematoma expansion between the SBP targets of 140–180 mmHg. In deep hemorrhages that involve the basal ganglia, more intensive blood pressure lowering reduced hematoma expansion but had no effect on functional outcome.

Initial clinical trials of evacuation of supratentorial hematomas, primarily via standard craniotomy, did not demonstrate clear benefit; however, recent focus on minimally-invasive surgical techniques holds promise. The International Surgical Trial in Intracerebral Haemorrhage (STICH) randomized patients with supratentorial ICH to either early surgical evacuation or initial medical management. No benefit was found in the early surgery arm, although analysis was complicated by the fact that 26% of patients in the initial medical management group ultimately had surgery for neurologic deterioration. The follow-up study, STICH-II, found that craniotomy and hematoma evacuation within 24 h of lobar supratentorial hemorrhage did not improve overall outcome but might have a role in select severely affected patients. However, many centers still consider surgery for patients deemed salvageable and who are experiencing progressive neurologic deterioration due to herniation. Surgical techniques continue to evolve. In a clinical trial of minimally invasive hematoma evacuation using instillation of the thrombolytic agent alteplase into the clot, mortality was decreased but there was not an improvement in functional outcome. In 2024, the first randomized trial demonstrating improvement in functional outcome after surgical hematoma evacuation was published. The Early MiNimally-invasive Removal of IntraCerebral Hemorrhage (ENRICH) trial found that surgical removal of lobar hematomas within 24 hours of onset in selected patients (hematoma volume 30–80 mL; Glasgow Coma Score 5–14 [Table 454-1]; pre-ICH functionally independent) was beneficial compared with medical management alone. Several clinical trials testing other minimally invasive surgical hematoma evacuation techniques are ongoing.

For cerebellar hemorrhages in patients with decreased level of consciousness or obstructive hydrocephalus, a neurosurgeon should be consulted immediately to assist with the evaluation. If the patient is alert without focal brainstem signs and the hematoma is small, surgical removal is usually unnecessary. Patients with hematomas >1 cm in diameter require careful observation for signs of impaired consciousness, progressive hydrocephalus, and precipitous respiratory failure. Hydrocephalus due to cerebellar hematoma generally requires surgical evacuation and should usually not be treated solely with ventricular drainage.

Tissue surrounding hematomas is displaced and compressed but not necessarily infarcted. Hence, major functional improvement often occurs as the hematoma is reabsorbed and the adjacent tissue regains its function over several months following acute injury. Careful management of the patient during the acute phase of the hemorrhage can lead to considerable recovery.

Bundles of care that incorporate multiple interventions may provide more value in the management of ICH patients than separating out different singular interventions. A hospital in the United Kingdom found that an ICH care bundle consisting of coagulopathy reversal, blood pressure lowering, and neurosurgical referral decreased patient mortality when implemented as a quality assurance project. In the INTERACT-3 randomized clinical trial performed in low- to middle-income countries, a bundle of care that included vitamin K coagulopathy reversal, acute blood pressure lowering, glucose control, and temperature control was associated with improved functional outcome and fewer adverse events.

Surprisingly, ICP is often normal even with large ICHs. However, if the hematoma causes marked midline shift of structures with consequent obtundation, coma, or hydrocephalus, osmotic agents can be instituted in preparation for placement of a ventriculostomy or parenchymal ICP monitor **(Chap. 318)**. Once ICP is recorded, CSF drainage (if available), osmotic therapy, and blood pressure management can be tailored to maintain cerebral perfusion pressure (MAP minus ICP) of 60 to ≥70 mmHg. For example, if ICP is found to be high, CSF can be drained from the ventricular space and osmotic therapy continued; persistent or progressive elevation in ICP may prompt surgical evacuation of the clot. Alternately, if ICP is normal, interventions such as osmotic therapy may be tapered. Because hyperventilation may produce ischemia due to cerebral vasoconstriction, induced hyperventilation should be limited to acute resuscitation of the patient with presumptive high ICP and eliminated once osmotic therapy or surgical treatments have been instituted. Glucocorticoids are not recommended for the treatment of intracerebral hemorrhage.

PREVENTION

Hypertension is the leading cause of primary ICH. Prevention is aimed at reducing chronic hypertension, eliminating excessive alcohol use, and discontinuing use of illicit drugs such as cocaine and amphetamines. Oral anticoagulant medications should generally be avoided in patients with high-risk features for CAA, but antiplatelet agents may be administered if there is an indication based on atherothrombotic vascular disease. Ongoing studies are investigating the risk-benefit ratio of reinitiation of anticoagulation in patients with recent ICH who have atrial fibrillation.

VASCULAR ANOMALIES

ARTERIOVENOUS MALFORMATIONS

True AVMs are shunts between the arterial and venous systems that may present with headache, seizures, and intracranial hemorrhage. AVMs consist of a tangle of abnormal vessels across the cortical surface or deep within the brain substance. AVMs vary in size from a small blemish a few millimeters in diameter to a large mass of tortuous channels composing an arteriovenous shunt of sufficient magnitude to raise cardiac output and precipitate heart failure. Blood vessels forming the tangle interposed between arteries and veins are usually abnormally thin and histologically resemble both arteries and veins. AVMs occur in all parts of the cerebral hemispheres, brainstem, and spinal cord, but the largest ones are most frequently located in the posterior half of the hemispheres, commonly forming a wedge-shaped lesion extending from the cortex to the ventricle. Most AVMs are congenital, but cases of acquired lesions have been reported.

Bleeding, headache, and seizures are most common between the ages of 10 and 30, occasionally as late as the fifties. AVMs are more frequent in men, and rare familial cases have been described. Familial AVM may be a part of the autosomal dominant syndrome of hereditary hemorrhagic telangiectasia (Osler-Rendu-Weber) syndrome due to mutations in either endoglin or activin receptor-like kinase 1, both involved in transforming growth factor (TGF) signaling and angiogenesis.

Headache (without bleeding) may be hemicranial and throbbing, like migraine, or diffuse. Focal seizures, with or without generalization, occur in ~30% of cases. One-half of AVMs become evident as ICHs. In most, the hemorrhage is mainly intraparenchymal with extension into the subarachnoid space in some cases. Unlike primary SAHs **(Chap. 440)**, blood from a ruptured AVM is usually not deposited in the basal cisterns, and symptomatic cerebral vasospasm is rare. The risk of AVM rupture is strongly influenced by a history of prior rupture. Although unruptured AVMs have a hemorrhage rate of ~2–4% per year, previously ruptured AVMs may have a rate as high as 17% a year, at least for the first year. Hemorrhages may be massive, leading to death, or may be as small as 1 cm in diameter, leading to minor focal symptoms or no deficit. The AVM may be large enough to steal blood away from adjacent normal brain tissue or to increase venous pressure significantly to produce venous ischemia locally and in remote areas of the brain. This is seen most often with large AVMs in the territory of the middle cerebral artery.

Large AVMs of the anterior circulation may be associated with a systolic and diastolic bruit (sometimes self-audible) over the eye, forehead, or neck and a bounding carotid pulse. Headache at the onset of AVM rupture is generally not as explosive as with aneurysmal rupture. MRI is better than CT for diagnosis, although noncontrast CT scanning sometimes detects calcification of the AVM and contrast may demonstrate the abnormal blood vessels. Once identified, conventional x-ray angiography is the gold standard for evaluating the precise anatomy of the AVM.

Surgical treatment of AVMs presenting with hemorrhage, often done in conjunction with preoperative embolization to reduce operative bleeding, is usually indicated for accessible lesions. Stereotactic radiosurgery, an alternative to conventional surgery, can produce a slow sclerosis of the AVM over 2–3 years.

Several angiographic features can be used to help predict future bleeding risk. Paradoxically, smaller lesions seem to have a higher hemorrhage rate. The presence of deep venous drainage, venous outflow stenosis, and intranidal aneurysms may increase rupture risk. Because of the relatively low annual rate of hemorrhage and the risk of complications due to surgical or endovascular treatment, the indications for surgery in asymptomatic AVMs are uncertain. The ARUBA (A Randomized Trial of Unruptured Brain Arteriovenous Malformations) trial randomized patients to medical management versus intervention (surgery, endovascular embolization, combination embolization and surgery, or gamma-knife). The trial was stopped prematurely for harm, with the medical arm achieving the combined endpoint of death or symptomatic stroke in 10% of patients compared to 31% in the intervention group at a mean follow-up time of 33 months. This highly significant finding argues against routine intervention for patients presenting without hemorrhage, although debate ensues regarding the generalizability of these results.

CAVERNOUS ANGIOMAS

Cavernous angiomas (cavernous malformations) are tufts of capillary sinusoids that form within the deep hemispheric white matter and brainstem with no normal intervening neural structures. The pathogenesis is unclear. Most cavernous angiomas are congenital, but they may occur during life as well. Familial cavernous angiomas have been mapped to several different genes: *KRIT1, CCM2*, and *PDCD10*. Both *KRIT1* and *CCM2* have roles in blood vessel formation, whereas *PDCD10* is an apoptotic gene. Cavernous angiomas are typically <1 cm in diameter and are often associated with a venous anomaly. Bleeding is usually of small volume, causing slight mass effect only. The bleeding risk for single cavernous malformations is 0.7–1.5% per year and may be higher for patients with prior clinical hemorrhage or multiple malformations. Seizures may occur if the malformation is located near the cerebral cortex. Surgical resection eliminates bleeding risk and may reduce seizure risk but is usually reserved for those malformations that form near the brain surface in patients with prior clinical episodes of bleeding or with medically refractory seizures. Stereotactic radiosurgery has been considered as a secondary treatment, but risks may outweigh benefits. Retrospective data show that intracranial hemorrhage from cavernous malformations is likely not increased with administration of antiplatelet and anticoagulant medications prescribed for other medical conditions.

Developmental venous anomalies are the result of development of anomalous cerebral, cerebellar, or brainstem venous drainage. These structures, unlike AVMs, are functional venous channels. They are of little clinical significance and should be ignored if found incidentally on brain imaging studies. Surgical resection of these anomalies may result in venous infarction and hemorrhage. Venous anomalies may be associated with cavernous malformations, which do carry some bleeding risk.

Capillary telangiectasias are true capillary malformations that often form extensive vascular networks through an otherwise normal brain structure. The pons and deep cerebral white matter are typical locations, and these capillary malformations can be seen in patients with hereditary hemorrhagic telangiectasia (Osler-Rendu-Weber) syndrome. If bleeding does occur, it rarely produces mass effect or significant symptoms. No treatment options exist.

Dural arteriovenous fistulas are acquired connections usually from a dural artery to a dural sinus. Patients may complain of a pulse-synchronous cephalic bruit ("pulsatile tinnitus") and headache. Depending on the magnitude of the shunt, venous pressures may rise high enough to cause cortical ischemia or venous hypertension and hemorrhage, particularly SAH. Surgical and endovascular techniques are usually curative. These fistulas may form because of trauma, but most are idiopathic. There is an association between fistulas and dural sinus thrombosis. Fistulas have been observed to appear months to years following venous sinus thrombosis, suggesting that angiogenesis factors elaborated from the thrombotic process may cause these anomalous connections to form. Alternatively, dural arteriovenous fistulas can produce venous sinus occlusion over time, perhaps from the high pressure and high flow through a venous structure.

FURTHER READING

Anderson CS et al: Rapid blood-pressure lowering in patients with acute intracerebral hemorrhage. N Engl J Med 368:2355, 2013.

Christensen H et al: European stroke organization guideline on reversal of oral anticoagulants in acute intracerebral hemorrhage. Euro Stroke J 4:294, 2019.

Greenberg SM et al: 2022 Guideline for the Management of Patients With Spontaneous Intracerebral Hemorrhage: A Guideline From the American Heart Association/American Stroke Association. Stroke 53:e282, 2022.

Ma L et al: The third Intensive Care Bundle with Blood Pressure Reduction in Acute Cerebral Haemorrhage Trial (INTERACT3): An international, stepped wedge cluster randomised controlled trial. Lancet 402:27, 2023.

Mohr JP et al: Medical management with or without interventional therapy for unruptured brain arteriovenous malformations (ARUBA): A multicentre, non-blinded, randomised trial. Lancet 383:614, 2014.

Pradilla G et al: Trial of early minimally invasive removal of intracerebral hemorrhage. N Engl J Med 390:1277, 2024.

440

Subarachnoid Hemorrhage

Wade S. Smith, Nerissa U. Ko,
J. Claude Hemphill, III

Subarachnoid hemorrhage (SAH) renders the brain critically ill from both primary and secondary brain insults. Excluding head trauma, the most common cause of SAH is rupture of a saccular aneurysm. Other causes include bleeding from a vascular malformation (arteriovenous malformation or dural arteriovenous fistula) and extension into the subarachnoid space from a primary intracerebral hemorrhage. Some idiopathic SAHs are localized to the perimesencephalic cisterns and are benign; they probably have a venous or capillary source, and angiography is unrevealing.

SACCULAR (“BERRY”) ANEURYSM

Autopsy and angiography studies have found that ~2% of adults harbor intracranial aneurysms, for a prevalence of 4 million persons in the United States. The incidence of SAH from aneurysmal rupture is estimated at between 6 and 11 per 100,000 person-years, resulting in 25,000–30,000 cases annually in the United States, with a 1.3 relative risk in women. Although most affected patients are under age 55, there is also an increasing incidence with age. The overall mortality rate for aneurysmal SAH is ~35%, with approximately one-third of patients dying immediately and prior to hospital admission. Of those who survive, more than half are left with clinically significant neurologic deficits because of the initial hemorrhage, delayed cerebral ischemia, or hydrocephalus. If the patient survives but the aneurysm is not obliterated, the rate of rebleeding is ~20% in the first 2 weeks, 30% in the first month, and ~3% per year afterward. Given these alarming figures, the major therapeutic emphasis is on preventing the predictable early complications of the SAH.

Unruptured, asymptomatic aneurysms are much less dangerous than recently ruptured ones. A large international observational study found that the annual risk of rupture for unruptured aneurysms <7 mm in size was 0% over 5 years. However, subsequent studies from Japan and Finland found that the majority of ruptured aneurysms were <6 mm in size. Aneurysms of <3 mm in size rarely, if ever, bleed. As the size of an aneurysm increases, so does the risk for rupture, but growth is not linear and appears to occur in phases, making surveillance for growing aneurysms problematic. The location of an unruptured aneurysm is also important in assessment, as basilar bifurcation and origin posterior communicating artery aneurysms appear to have a higher risk for rupture than other sites. Because of their longer length of exposure to risk of rupture, younger patients with aneurysms >10 mm in size may benefit from prophylactic treatment (see below). As with the treatment of asymptomatic carotid stenosis **(Chap. 438)**, this risk-benefit ratio strongly depends on the rate of procedural complications.

Giant aneurysms, those >25 mm in diameter, occur at the same sites (see below) as small aneurysms, and account for 5% of cases. The three most common locations are the terminal internal carotid artery, middle cerebral artery (MCA) bifurcation, and top of the basilar artery. Their risk of rupture is ~8–10% annually after identification and may remain high indefinitely. They often cause symptoms by compressing the adjacent brain or cranial nerves.

Mycotic aneurysms are usually located distal to the first bifurcation of major arteries of the circle of Willis. Most result from infected emboli due to bacterial endocarditis causing septic degeneration of arteries and subsequent dilation and rupture. Whether these lesions should be sought and repaired prior to rupture or left to heal spontaneously with antibiotic treatment remains controversial.

Pathophysiology Saccular aneurysms occur at the bifurcations of the large- to medium-sized intracranial arteries; rupture is into the subarachnoid space in the basal cisterns and sometimes into the parenchyma of the adjacent brain.

Approximately 89% of aneurysms occur in the anterior circulation, mostly on the circle of Willis **(Fig. 440-1)**. About 20% of patients have multiple aneurysms, many at mirror sites bilaterally. As an aneurysm develops, it typically forms a neck with a dome. The length of the neck and the size of the dome vary greatly and are important factors in planning neurosurgical obliteration or endovascular embolization. The arterial internal elastic lamina disappears at the base of the neck. The media thins, and connective tissue replaces smooth-muscle cells. At the site of rupture (most often the dome), the wall thins, and the tear that allows bleeding is often ≤0.5 mm long. Fusiform aneurysms, where the locations of inflow and outflow are different, typically occur in the basilar artery and are very challenging to treat.

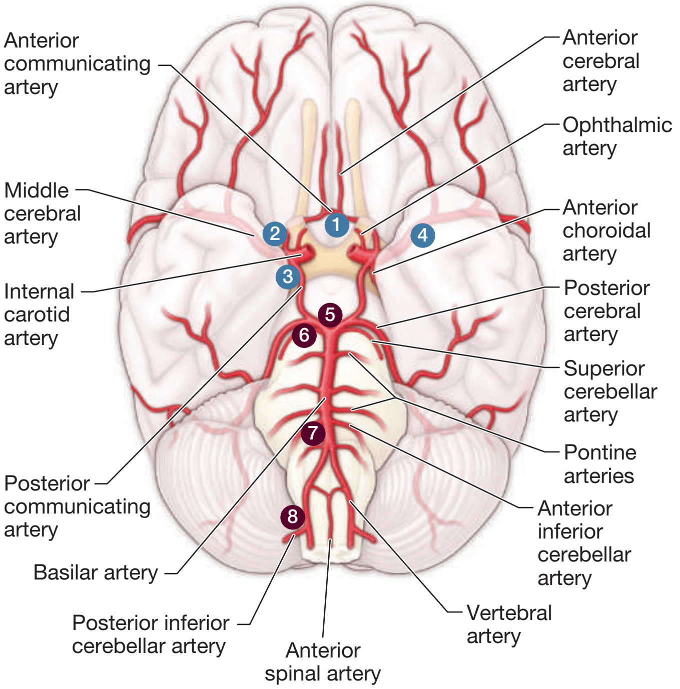
Anterior communicating artery
Middle cerebral artery
Internal carotid artery
Posterior communicating artery
Basilar artery
Posterior inferior cerebellar artery
Anterior spinal artery
Vertebral artery
Anterior inferior cerebellar artery
Pontine arteries
Superior cerebellar artery
Posterior cerebral artery
Anterior choroidal artery
Ophthalmic artery
Anterior cerebral artery
1
2
3
4
5
6
7
8

FIGURE 440-1 View of the major blood vessels supplying the brain and common locations of saccular aneurysms: (1) Anterior communicating (12%), (2) internal carotid (30%), (3) posterior communicating (12%), (4) middle cerebral (34%), (5) basilar terminus, (6) superior cerebellar, (7) anterior inferior cerebellar, and (8) posterior inferior cerebellar aneurysm. Locations 1–4 are considered anterior circulation aneurysms totaling 89% overall, while locations 5–8 total 11%.

Clinical Manifestations Most unruptured intracranial aneurysms are completely asymptomatic. Symptoms are usually due to rupture and resultant SAH, although some unruptured aneurysms present with mass effect on cranial nerves or brain parenchyma. At the moment of aneurysmal rupture with a major SAH, the intracranial pressure (ICP) suddenly rises. This may account for the sudden transient loss of consciousness that occurs in nearly half of patients. Sudden loss of consciousness may be preceded by a brief moment of excruciating headache, but most patients first complain of headache upon regaining consciousness. In 10% of cases, aneurysmal bleeding is severe enough to cause loss of consciousness for several days. In ~45% of cases, severe headache associated with exertion is the presenting complaint. The patient often calls the headache “the worst headache of my life”; however, the most important characteristic is sudden onset. Occasionally, these ruptures may present as headache of only moderate intensity or as a change in the patient’s usual headache pattern. The headache is usually generalized, often with neck stiffness, and vomiting is common.

Although sudden headache in the absence of focal neurologic symptoms is the hallmark of aneurysmal rupture, focal neurologic deficits may occur. Anterior communicating artery or MCA bifurcation aneurysms may rupture into the adjacent brain or subdural space and form a hematoma large enough to produce mass effect. The deficits that result can include hemiparesis, aphasia, and mental slowness (abulia).

Occasionally, prodromal symptoms suggest the location of a progressively enlarging unruptured aneurysm. A third cranial nerve palsy, particularly when associated with pupillary dilation, loss of ipsilateral (but retained contralateral) light reflex, and focal pain above or behind the eye, may occur with an expanding aneurysm at the junction of the posterior communicating artery and the internal carotid artery. A sixth nerve palsy may indicate an aneurysm in the cavernous sinus, and visual field defects can occur with an expanding supraclinoid carotid or anterior cerebral artery (ACA) aneurysm. Occipital and posterior cervical pain may signal a posterior inferior cerebellar artery or anterior inferior cerebellar artery aneurysm **(Chap. 438)**. Pain in or behind the eye and in the low temple can occur with an expanding MCA aneurysm. Thunderclap headache is a variant of migraine that simulates an SAH. Before concluding that a patient with sudden, severe headache has thunderclap migraine, a definitive workup for aneurysm or other intracranial pathology is required.

TABLE 440-1 Grading Scales for Subarachnoid Hemorrhage

GRADE	HUNT-HESS SCALE	WORLD FEDERATION OF NEUROSURGICAL SOCIETIES (WFNS) SCALE
1	Asymptomatic, or minimal headache and slight nuchal rigidity. Normal mental status, no cranial nerve or motor findings	GCS[a] score 15, no motor deficits
2	Moderate to severe headache, nuchal rigidity, normal mental status and motor function, may have cranial nerve deficit	GCS score 13–14, no motor deficits
3	Somnolent, confused, may have cranial nerve or mild motor deficit	GCS score 13–14, with motor deficits
4	Stupor, moderate to severe motor deficit, may have intermittent reflex posturing	GCS score 7–12, with or without motor deficits
5	Coma, reflex posturing or flaccid	GCS score 3–6, with or without motor deficits

[a] Glasgow Coma Scale; see Table 454-1.

Source: Reproduced with permission from WFNS Scale: Report of World Federation of Neurological Surgeons Committee on a Universal Subarachnoid hemorrhage Grading Scale. J Neurosurg 68:985, 1988.

Aneurysms can undergo small ruptures and leaks of blood into the subarachnoid space, so-called *sentinel bleeds*. Sudden unexplained headache at any location should raise suspicion of SAH and be investigated because a major hemorrhage may be imminent.

The initial clinical manifestations of SAH can be graded using the Hunt-Hess or World Federation of Neurosurgical Societies classification schemes (Table 440-1). For ruptured aneurysms, prognosis for a good outcome falls as the grade increases. For example, it is unusual for a Hunt-Hess grade 1 patient to die if the aneurysm is treated, but the mortality rate for grade 4 and 5 patients may be as high as 60%.

Delayed Neurologic Deficits There are four major causes of delayed neurologic deficits: rerupture, hydrocephalus, delayed cerebral ischemia, and hyponatremia.

1. *Rerupture*. The incidence of rerupture of an untreated aneurysm in the first month following SAH is ~30%, with the peak in the first 7 days. Rerupture is associated with a 50% mortality rate and poor outcome. Early treatment eliminates this risk, and advances in endovascular and surgical techniques contribute to better outcomes.
2. *Hydrocephalus*. Acute hydrocephalus can cause stupor and coma and can be mitigated by placement of an external ventricular drain. More often, subacute hydrocephalus may develop over a few days or weeks and cause progressive drowsiness or slowed mentation with incontinence. Hydrocephalus may clear spontaneously, require temporary ventricular drainage, or in some cases require placement of a ventriculoperitoneal shunt. Chronic hydrocephalus may develop weeks to months after SAH and manifest as gait difficulty, incontinence, or impaired mentation. Subtle signs may be a lack of initiative in conversation or a failure to recover independence.
3. *Delayed cerebral ischemia*. Vasospasm is the narrowing of the arteries at the base of the brain following SAH and has been associated with delayed cerebral ischemia and infarction. Delayed cerebral ischemia occurs in ~30% of patients and is the major cause of delayed morbidity and death. Signs first appear 4–14 days after the hemorrhage, most often at 7 days. While therapies can be targeted for vasospasm, there are other complex mechanisms associated with delayed cerebral ischemia and progression to infarction independent of vasospasm alone.

   Narrowing of the arteries causing vasospasm and thickening of the vessel wall is believed to result from direct effects of clotted blood and its breakdown products on the arteries within the subarachnoid space. In general, the more blood that surrounds the arteries, the greater is the chance of symptomatic vasospasm. Focal narrowing of major arteries produces symptoms referable to the appropriate vascular territory (Chap. 437). All of these focal symptoms may present abruptly, fluctuate, or develop over a few days. The clinical syndrome often manifests as a decline in mental status and worsening headache.

   Vasospasm of the large arteries can be detected reliably with conventional x-ray angiography, but this procedure is invasive and carries the risk of stroke and other complications. Transcutaneous Doppler ultrasound is based on the principle that the velocity of blood flow within an artery will rise as the lumen diameter is narrowed. By directing the probe along the MCA and proximal ACA, carotid terminus, and vertebral and basilar arteries on a daily or every-other-day basis, vasospasm can be reliably detected and treatments initiated to prevent cerebral ischemia (see below). Computed tomography (CT) angiography is another method that can detect vasospasm. The addition of CT perfusion imaging may help identify reversible ischemic deficits. In high-grade patients, invasive neuromonitoring techniques can also be considered.
4. *Hyponatremia*. Hyponatremia may be profound and can develop quickly in the first 2 weeks following SAH. There is both natriuresis and volume depletion with SAH, so that patients become both hyponatremic and hypovolemic. Both atrial natriuretic peptide and brain natriuretic peptide have a role in producing this "cerebral salt-wasting syndrome." Typically, it clears over the course of 1–2 weeks and, in the setting of SAH, should not be treated with free-water restriction as this may increase the risk of stroke (see below).

Laboratory Evaluation and Imaging (Fig. 440-2) The hallmark of aneurysmal rupture is blood in the CSF. More than 95% of cases have enough blood to be visualized on a high-quality noncontrast CT scan obtained within 72 h. If the scan fails to establish the diagnosis of SAH and no mass lesion or obstructive hydrocephalus is found, a lumbar puncture should be performed to establish the presence of subarachnoid blood. Lysis of the red blood cells and subsequent conversion of hemoglobin to bilirubin stains the spinal fluid yellow within 6–12 h. This xanthochromic spinal fluid peaks in intensity at 48 h and lasts for 1–4 weeks, depending on the amount of subarachnoid blood present.

The extent and location of subarachnoid blood on a noncontrast CT scan help locate the underlying aneurysm, identify the cause of any neurologic deficit, and predict the occurrence of vasospasm. The likelihood of symptomatic vasospasm in the MCA and ACA can be predicted based on the size and location of clotted blood (Table 440-2). CT scans less reliably predict vasospasm in the vertebral, basilar, or posterior cerebral arteries.

Lumbar puncture prior to an imaging procedure is indicated only if a CT scan is not available at the time of the suspected SAH. Once the diagnosis of hemorrhage from a ruptured saccular aneurysm is confirmed, four-vessel conventional x-ray angiography (both carotids and both vertebrals) is generally performed to localize and define the anatomic details of the aneurysm and to determine if other unruptured aneurysms exist (Fig. 440-2*C*). The ruptured aneurysm can be treated using endovascular techniques at the time of the initial angiogram to expedite treatment and minimize the number of invasive procedures. CT angiography is an alternative method for locating the aneurysm and may be sufficient for planning definitive therapy.

Close monitoring (daily or twice daily) of electrolytes is important because hyponatremia can occur precipitously during the first 2 weeks following SAH (see above).

The electrocardiogram (ECG) frequently shows ST-segment and T-wave changes similar to those associated with cardiac ischemia. A prolonged QRS complex, increased QT interval, and prominent "peaked" or deeply inverted symmetric T waves are usually secondary to the intracranial hemorrhage. Structural myocardial lesions produced by circulating catecholamines and excessive discharge of sympathetic neurons may occur after SAH, causing these ECG changes and a reversible cardiomyopathy sufficient to cause shock. Echocardiography often reveals a pattern of regional wall motion abnormalities that follow the distribution of sympathetic nerves rather than the major coronary arteries, with relative sparing of the ventricular wall apex. The sympathetic nerves themselves appear to be injured by direct toxicity

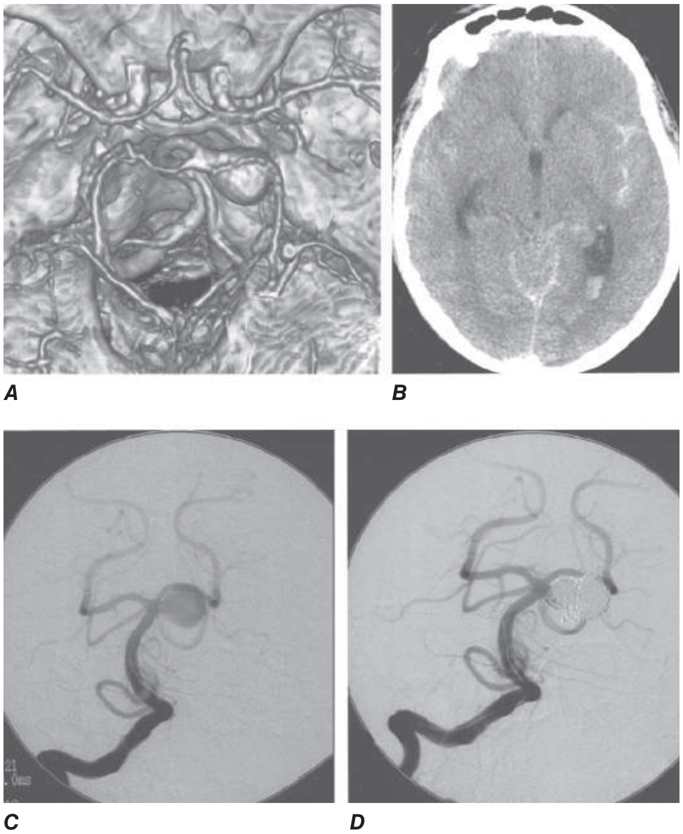

FIGURE 440-2 Subarachnoid hemorrhage. ***A.*** Computed tomography (CT) angiography revealing an aneurysm of the left superior cerebellar artery. ***B.*** Noncontrast CT scan at the level of the third ventricle revealing subarachnoid blood (*bright*) in the left sylvian fissure and within the left lateral ventricle. ***C.*** Conventional anteroposterior x-ray angiogram of the right vertebral and basilar artery showing the large aneurysm. ***D.*** Conventional angiogram following coil embolization of the aneurysm, whereby the aneurysm body is filled with platinum coils delivered through a microcatheter navigated from the femoral artery into the aneurysm neck.

from the excessive catecholamine release. An asymptomatic troponin elevation is common. While arrythmias can occur after SAH, serious ventricular dysrhythmias occurring in-hospital are unusual.

TREATMENT

Subarachnoid Hemorrhage

Early aneurysm repair prevents rerupture and allows the safe application of techniques to improve blood flow (e.g., induced hypertension) should vasospasm and delayed cerebral ischemia develop. At many centers, definitive repair is carried out within 24 h of the bleed in all patients who are stable enough to tolerate the procedure.

TABLE 440-2 Modified Fisher Grading System for Prediction of Vasospasm Risk

GRADE	CT SCAN FINDINGS	RISK OF SYMPTOMATIC VASOSPASM
0	No subarachnoid or intraventicular blood	0%
1	Focal or diffuse thin subarachnoid blood without intraventricular blood	24%
2	Focal or diffuse thin subarachnoid blood with intraventricular blood	33%
3	Focal or diffuse thick subarachnoid blood without intraventricular blood	33%
4	Focal or diffuse thick subarachnoid blood with intraventricular blood	40%

Note: "Thin" is <1 mm, whereas "thick" is ≥1 mm.

Abbreviation: CT, computed tomography.

Source: JA Frontera et al: Prediction of symptomatic vasospasm after subarachnoid hemorrhage: The modified Fisher scale. Neurosurgery 59:21, 2006.

An aneurysm can be "clipped" by a neurosurgeon or "coiled" by an endovascular surgeon. Surgical repair involves placing a metal clip across the aneurysm neck, thereby immediately eliminating the risk of rebleeding. This approach requires craniotomy and brain retraction, which is associated with neurologic morbidity. Endovascular techniques involve placing coils, or other embolic material, within the aneurysm via a catheter that is passed from the femoral or radial artery. The aneurysm is packed tightly to enhance thrombosis and over time is walled off from the circulation **(Fig. 440-2*D*)**. There have been two prospective randomized trials of surgery versus endovascular treatment for ruptured aneurysms: the first was the International Subarachnoid Aneurysm Trial (ISAT), which was terminated early when 24% of patients treated with endovascular therapy were dead or dependent at 1 year compared to 31% treated with surgery, a significant 23% relative reduction. After 5 years, the risk of death was still lower in the coiling group, although the proportion of survivors who were independent was the same in both groups. The risk of rebleeding was generally low, but was more common in the coiling group. These results favoring coiling at 1 year were confirmed in a second trial, although the differences in functional outcome of survivors were no longer significant at 3 years. Because some aneurysms have a morphology that is not amenable to endovascular treatment, surgery remains an important treatment option, especially in cases where evacuation of a parenchymal hematoma could be beneficial. Newer endovascular devices and techniques using balloon-assisted coiling or placement of flow-diverting stents are increasing the types of aneurysms amenable to endovascular intervention. Centers that combine both endovascular and neurosurgical expertise likely offer the best outcomes for patients, and transfer to centers that specialize in aneurysm treatment are associated with improved outcomes and lower mortality rates.

The early medical management of SAH focuses on protecting the airway, managing blood pressure before and after repair of the aneurysm, preventing rebleeding prior to the intervention, and treating hydrocephalus. Subsequent management is focused on preventing late neurologic injury, including managing vasospasm and delayed cerebral ischemia, treating hyponatremia, limiting secondary brain insults from medical comorbidities, and preventing pulmonary emboli.

Intracranial hypertension following aneurysmal rupture occurs secondary to subarachnoid blood, parenchymal hematoma, acute hydrocephalus, and/or loss of vascular autoregulation. Patients who are stuporous should undergo emergent ventriculostomy to measure and treat high ICP in order to prevent cerebral herniation and ischemia. Medical therapies designed to combat raised ICP (e.g., osmotic therapy and sedation) can also be used as needed. High ICP refractory to treatment is a poor prognostic sign. Drainage of CSF via a lumbar route has been shown to decrease the rate of delayed cerebral injury, vasospasm, and mortality but confounds measurement of true ICP.

Prior to definitive treatment of the ruptured aneurysm, care is required to maintain adequate cerebral perfusion pressure while avoiding excessive elevation of arterial pressure. If the patient is alert, it is reasonable to lower the systolic blood pressure to below 160 mmHg using agents such as nicardipine, clevidipine, or labetalol to limit blood pressure variability. If the patient has a depressed level of consciousness, ICP should be measured and the cerebral perfusion pressure targeted to 60–70 mmHg. If headache or neck pain is severe, mild sedation and analgesia are prescribed. Extreme sedation is avoided if possible because it can obscure the ability to clinically detect changes in neurologic status. Goal-directed therapy to target euvolemia is recommended, while avoiding hypervolemia, which has been associated with a greater risk for complications.

Seizures are uncommon at the onset of aneurysmal rupture, but in patients who present with seizures, treatment with a 7-day course of an anticonvulsant such as levetiracetam is reasonable. The quivering, jerking, and extensor posturing that often accompany loss of consciousness with SAH are probably related to the sharp rise in ICP rather than seizures. Monitoring with electroencephalogram

(EEG) can help detect seizures in patients with poor mental status or fluctuating exam findings. Anticonvulsants are sometimes administered as prophylactic therapy in high-risk patients, but if used, the treatment course should also not exceed >7 days. Phenytoin should be avoided because of its association with increased morbidity and mortality in this setting.

Glucocorticoids may reduce the headache and neckache caused by the irritative effect of the subarachnoid blood; however, there is no good evidence that steroids reduce cerebral edema, are neuroprotective, or reduce vascular injury, and their routine use therefore is not recommended.

Antifibrinolytic agents are not routinely prescribed but have been considered in patients in whom aneurysm repair cannot proceed immediately. They are associated with a reduced incidence of aneurysmal rerupture but may also increase the risk of delayed cerebral ischemia and deep-vein thrombosis (DVT). More recent studies showed no improvement in functional outcomes, and use of these agents is not currently recommended.

Delayed cerebral ischemia remains the leading cause of morbidity and mortality following aneurysmal SAH in patients who survive the initial hemorrhage. Treatment with the calcium channel antagonist nimodipine (60 mg PO every 4 h) has been shown to improve outcomes, perhaps by preventing ischemic injury rather than reducing the risk of vasospasm. Nimodipine can cause significant hypotension in some patients, which may worsen cerebral ischemia in patients with vasospasm. Symptomatic cerebral vasospasm can also be treated by increasing the cerebral perfusion pressure by raising mean arterial pressure through plasma volume expansion and the judicious use of IV vasopressor agents, usually phenylephrine or norepinephrine. Increasing perfusion pressure has been associated with clinical improvement in many patients, but high arterial pressure may also promote rebleeding in unrepaired aneurysms. Treatment with induced hypertension in symptomatic patients requires close monitoring of arterial pressures. Prophylactic use of induced hypertension is not recommended, as it is associated with worse outcomes. Euvolemia should be targeted as significant hypervolemia may lead to cardiopulmonary complications. Hypovolemia should be strictly avoided.

If delayed cerebral ischemia due to vasospasm persists despite optimal medical therapy, endovascular rescue therapies with intra-arterial vasodilators and percutaneous transluminal angioplasty can be considered **(Fig. 440-3)**. Vasodilatation by direct angioplasty appears to be permanent, allowing hypertensive therapy to be tapered sooner. In contrast, the pharmacologic vasodilators (verapamil and nicardipine) do not last more than about 24 h, and therefore, multiple treatments may be required until the subarachnoid blood is reabsorbed. Newer therapies including milrinone, neural ganglia blocks, antithrombotic agents, and intrathecal and cisternal agents are currently under investigation.

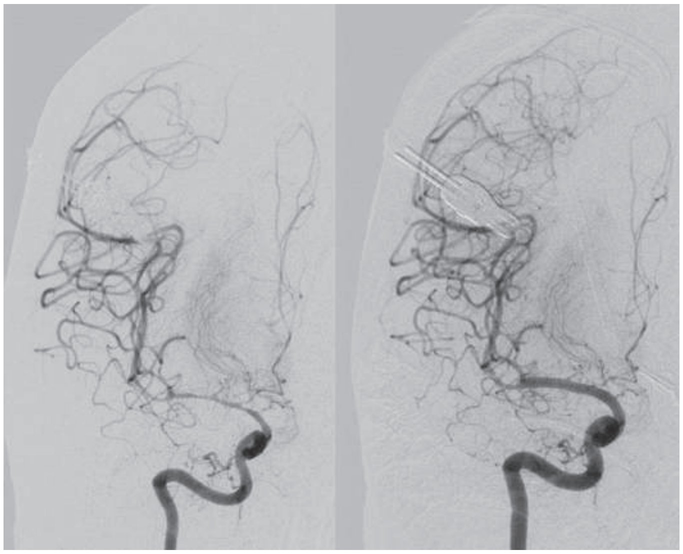

FIGURE 440-3 Vasospasm of the right middle cerebral artery. ***A.*** Catheter angiography demonstrates significant narrowing of the right middle cerebral artery (MCA). ***B.*** Because of symptomatic delayed cerebral ischemia, soft-balloon angioplasty was used to dilate the proximal portion of the main MCA stem.

Severe cerebral edema in patients with infarction from vasospasm may increase the ICP to levels that reduce cerebral perfusion pressure. Treatment may include cerebrospinal fluid (CSF) drainage, mannitol, or hypertonic saline; for intractable cases, hemicraniectomy, deep sedation, paralysis, and moderate hypothermia may be considered.

Delayed cerebral ischemia may also occur in the absence of significant large-vessel vasospasm. Potential mechanisms include microthrombosis, activation of the inflammatory cascade, microvascular dysregulation and constriction, and cortical spreading depolarization. Targeted treatments for these mechanisms are under investigation.

Acute hydrocephalus can cause stupor or coma. It may clear spontaneously or require temporary ventricular drainage. When chronic hydrocephalus develops, ventricular shunting is the treatment of choice.

Free-water restriction is contraindicated in patients with SAH at risk for delayed cerebral ischemia because hypovolemia and hypotension may occur and precipitate cerebral ischemia. Many patients continue to experience a decline in serum sodium due to cerebral salt wasting despite receiving parenteral fluids containing normal saline. Frequently, supplemental oral salt coupled with normal saline will mitigate hyponatremia, but often patients also require intravenous hypertonic (3%) saline. Care must be taken not to correct serum sodium too quickly in patients with marked hyponatremia of several days' duration, as the osmotic demyelination syndrome **(Chap. 318)** may occur.

All patients should have pneumatic compression stockings applied to prevent pulmonary emboli. Unfractionated heparin and low-molecular-weight heparinoids administered subcutaneously for DVT prophylaxis can be initiated within 1–2 days following endovascular treatment or craniotomy with surgical clipping; this approach is more effective than use of pneumatic compression stockings alone. Treatment of pulmonary emboli depends on whether the aneurysm has been treated and whether or not the patient has had a craniotomy. Continuous systemic anticoagulation with heparin is contraindicated in patients with ruptured and untreated aneurysms. It is a relative contraindication following craniotomy for several days, and it may delay thrombosis of a coiled aneurysm. If DVT or pulmonary emboli occur within the first days following craniotomy, use of an inferior vena cava filter may be considered to prevent additional emboli, whereas systemic anticoagulation with heparin is preferred following successful endovascular treatment.

In patients who survive their acute hospitalization, follow-up care is important to address the high prevalence of cognitive and behavioral deficits that greatly impact quality of life. Efficient recognition and treatment of these disorders can improve both short-term and long-term outcomes. In addition, some patients require ongoing follow-up to manage unruptured aneurysms that may require further care.

FURTHER READING

Hoh BL et al: 2023 Guideline for the management of patients with aneurysmal subarachnoid hemorrhage: A guideline from the American Heart Association/American Stroke Association. Stroke 54:e314, 2023.

Molyneux AJ et al: The durability of endovascular coiling versus neurosurgical clipping of ruptured cerebral aneurysms: 18 year follow-up of the UK cohort of the International Subarachnoid Aneurysm Trial (ISAT). Lancet 385:691, 2015.

Treggiari MM et al: Guidelines for the neurocritical care management of aneurysmal subarachnoid hemorrhage. Neurocrit Care 39:1, 2023.

441 Migraine and Other Primary Headache Disorders

Peter J. Goadsby

The general approach to headache as a cardinal symptom is covered elsewhere **(Chap. 17)**; here, disorders in which headache and associated features occur in the absence of any exogenous cause are discussed. The most common are migraine, tension-type headache (TTH), and the trigeminal autonomic cephalalgias (TACs), notably cluster headache; the complete list is summarized in **Table 441-1**.

MIGRAINE

Migraine, the second most common cause of headache, and the most common headache-related, and indeed neurologic, cause of disability in the world, afflicts ~15% of women and 6% of men over a 1-year period. It is usually an episodic headache associated with certain features, such as sensitivity to light, sound, or movement; nausea and vomiting often accompany the headache. A useful description of migraine is a recurring syndrome of headache associated with other symptoms of neurologic dysfunction in varying admixtures **(Table 441-2)**. A migraine attack has three phases: premonitory (prodrome), headache phase, and postdrome; each has distinct and sometimes disabling symptoms, which may overlap. About 20–25% of migraine patients have a fourth phase: aura. Migraine can often be recognized by its activators, referred to as *triggers*.

Migraineurs are particularly sensitive to environmental and sensory stimuli; migraine-prone patients do not habituate easily to sensory stimuli. This sensitivity is amplified in women during the menstrual cycle. Headache can be initiated or amplified by various triggers, including altered sleep patterns; hunger; let-down from stress; physical exertion; stormy weather or barometric pressure changes; hormonal fluctuations during menses; and alcohol or other chemical stimulation, such as with nitrates. Knowledge of a patient's susceptibility to specific triggers can be useful in management strategies involving lifestyle adjustments, although it is becoming recognized that some apparent triggers, such as light sensitivity, may be part of the initial phase of the attack; i.e., the premonitory phase or prodrome.

Pathogenesis The sensory sensitivity that is characteristic of migraine is probably due to dysfunction of monoaminergic and other sensory control systems located in the brainstem and hypothalamus **(Fig. 441-1)**.

Activation of cells in the trigeminal nucleus results in the release of vasoactive neuropeptides, notably calcitonin gene–related peptide (CGRP) and pituitary adenylate cyclase activating polypeptide (PACAP), at vascular terminals of the trigeminal nerve and within the trigeminal nucleus. CGRP receptor antagonists, *gepants*, have now been shown to be effective in the acute and preventive treatment of migraine, and four monoclonal antibodies to CGRP, or its receptor, have been shown to be effective in migraine prevention, as has one PACAP monoclonal antibody in a phase 2 study. Centrally, the second-order trigeminal neurons cross the midline and project to ventrobasal and posterior nuclei of the thalamus for further processing. Additionally, there are projections to the periaqueductal gray and hypothalamus, from which reciprocal descending systems have established antinociceptive effects. Other brainstem regions likely to be involved in descending modulation of trigeminal pain include the locus coeruleus and parabrachial nucleus in the pons and the rostroventromedial medulla.

Pharmacologic and other data point to the involvement of the neurotransmitter 5-hydroxytryptamine (5-HT; also known as *serotonin*) in migraine. In the late 1950s, methysergide was suggested to antagonize certain peripheral actions of 5-HT and was introduced, based on its anti-inflammatory properties, as a migraine preventive. The *triptans* were designed to stimulate selectively subpopulations of 5-HT receptors; at least 14 different 5-HT receptors exist in humans. The triptans are potent agonists of 5-HT$_{1B}$ and 5-HT$_{1D}$ receptors, and some are active at the 5-HT$_{1F}$ receptor; the latter's exclusive agonists are called *ditans*. Triptans arrest nerve signaling in the nociceptive pathways of the trigeminovascular system, at least in the trigeminal nucleus caudalis and trigeminal sensory thalamus, in addition to promoting cranial vasoconstriction, whereas *ditans*, which are also effective in acute migraine, act only at neural targets. A range of other neural targets are currently under investigation for the acute and preventive management of migraine.

Data also support a role for dopamine in the pathophysiology of migraine. Many migraine premonitory symptoms can be induced by dopaminergic stimulation. Moreover, there is dopamine receptor hypersensitivity in migraineurs, as demonstrated by the induction of yawning, nausea, vomiting, hypotension, and other symptoms of a migraine attack by dopaminergic agonists at doses that do not affect nonmigraineurs. Dopamine receptor antagonists are effective therapeutic agents in migraine, especially when given parenterally or concurrently with other antimigraine agents. Moreover, hypothalamic activation, anterior to that seen in cluster headache, has been shown in the premonitory (prodromal) phase of migraine using functional imaging, and this may hold a key to understanding some part of the role of dopamine in the disorder.

Migraine genes identified by studying families with familial hemiplegic migraine (FHM) reveal involvement of ion channels, suggesting that alterations in membrane excitability can predispose to migraine aura. Mutations involving the Ca$_v$2.1 (P/Q)–type voltage-gated calcium channel *CACNA1A* gene are now known to cause FHM 1; this mutation is responsible for about 50% of FHM cases. Mutations in the Na$^+$-K$^+$ ATPase *ATP1A2* gene, designated FHM 2, are responsible for about 20% of FHMs. Mutations in the neuronal voltage-gated sodium channel *SCN1A* cause FHM 3. Functional neuroimaging has suggested that brainstem regions in migraine **(Fig. 441-2)** and the posterior hypothalamic gray matter region close to the human circadian pacemaker cells of the suprachiasmatic nucleus in cluster headache **(Fig. 441-3)** are good candidates for specific involvement in these primary headache disorders.

Diagnosis and Clinical Features Classic diagnostic criteria for migraine headache are listed in **Table 441-3** and should be considered together with the extended features in Table 441-2. A high index of suspicion is required to diagnose migraine: the migraine aura, consisting of visual disturbances with flashing lights or zigzag lines moving across the visual field or of other neurologic symptoms, is reported in only 20–25% of patients. It should be distinguished from the pan-field television static-like disturbance now recognized as *visual snow*. The first phase of a migraine attack for most patients is the premonitory (prodromal) phase consisting of some or all of the following: yawning, sleepiness, fatigue, cognitive dysfunction, mood change, neck discomfort, polyuria, and food cravings; this can last from a few hours to days. Typically, the headache phase follows with its associated features, such as nausea, photophobia, and phonophobia as well as allodynia or vertigo. When questioned, these typical migraine symptoms also emerge in the premonitory phase, and typical premonitory symptoms also continue into the headache phase. As the headache lessens, many patients enter a postdrome, most commonly feeling tired/weary, having problems concentrating, and experiencing mild neck discomfort that can last for hours and sometimes up to a day. A headache diary can often be helpful in making the diagnosis and in assessing disability and the frequency of acute attacks. Patients with episodes of migraine on 8 or more days per month and with at least 15 total days of headache per month are considered to have chronic migraine **(see "Chronic Daily Headache" in Chap. 17)**. Migraine must be differentiated from TTH (discussed below), which is reported to be the most common primary headache disorder. Migraine has several forms that have been defined (Table 441-1): migraine with and without aura and chronic migraine are the most important. *Migraine at its most*

TABLE 441-1 Primary Headache Disorders, Modified from International Classification of Headache Disorders-III (Headache Classification Committee of the International Headache Society, 2018)

1. Migraine	1.1 Migraine without aura 1.2 Migraine with aura 1.2.1 Migraine with typical aura 1.2.1.1 Typical aura with headache 1.2.1.2 Typical aura without headache 1.2.2 Migraine with brainstem aura 1.2.3 Hemiplegic migraine 1.2.3.1 Familial hemiplegic migraine (FHM) 1.2.3.1.1 Familial hemiplegic migraine type 1 1.2.3.1.2 Familial hemiplegic migraine type 2 1.2.3.1.3 Familial hemiplegic migraine type 3 1.2.3.1.4 Familial hemiplegic migraine, other loci 1.2.3.2 Sporadic hemiplegic migraine 1.2.4 Retinal migraine 1.3 Chronic migraine 1.4 Complications of migraine 1.4.1 Status migrainosus 1.4.2 Persistent aura without infarction 1.4.3 Migrainous infarction 1.4.4 Migraine aura-triggered seizure 1.5 Probable migraine 1.5.1 Probable migraine without aura 1.5.2 Probable migraine with aura 1.6 Episodic syndromes that may be associated with migraine 1.6.1 Recurrent gastrointestinal disturbance 1.6.1.1 Cyclical vomiting syndrome 1.6.1.2 Abdominal migraine 1.6.2 Benign paroxysmal vertigo 1.6.3 Benign paroxysmal torticollis A 1.6.4 Infantile colic A 1.6.6 Vestibular migraine
2. Tension-type headache	2.1 Infrequent episodic tension-type headache 2.2 Frequent episodic tension-type headache 2.3 Chronic tension-type headache 2.4 Probable tension-type headache
3. Trigeminal autonomic cephalalgias	3.1 Cluster headache 3.1.1 Episodic cluster headache 3.1.2 Chronic cluster headache 3.2 Paroxysmal hemicrania 3.2.1 Episodic paroxysmal hemicrania 3.2.2 Chronic paroxysmal hemicrania 3.3 Short-lasting unilateral neuralgiform headache attacks 3.3.1 Short-lasting unilateral neuralgiform headache attacks with conjunctival injection and tearing (SUNCT) 3.3.1.1 Episodic SUNCT 3.3.1.2 Chronic SUNCT 3.3.2 Short-lasting unilateral neuralgiform headache attacks with cranial autonomic symptoms (SUNA) 3.3.2.1 Episodic SUNA 3.3.2.2 Chronic SUNA 3.4 Hemicrania continua 3.5 Probable trigeminal autonomic cephalalgia
4. Other primary headache disorders	4.1 Primary cough headache 4.2 Primary exercise headache 4.3 Primary headache associated with sexual activity 4.4 Primary thunderclap headache 4.5 Cold-stimulus headache 4.5.1 Headache attributed to external application of a cold stimulus 4.5.2 Headache attributed to ingestion or inhalation of a cold stimulus 4.6 External-pressure headache 4.6.1 External-compression headache 4.6.2 External-traction headache 4.7 Primary stabbing headache 4.8 Nummular headache 4.9 Hypnic headache 4.10 New daily persistent headache (NDPH)

Headache Classification Committee of the International Headache Society (IHS) 38(1), pp. 1-211. Copyright © 2018 by (International Headache Society) Reprinted by Permission of SAGE Publications.

TABLE 441-2 Migraine Symptoms by Attack Phase

Premonitory (prodromal)
- Neck discomfort
- Higher center
    - Cognitive impairment (brain "fog")
    - Mood change
    - Fatigue
- Homeostatic
    - Yawning/sleepiness
    - Polyuria/polydipsia
    - Food cravings

Aura
- Neurologic disturbance, such as scintillating scotoma

Headache phase
- Pain
- Nausea/vomiting
- Sensory sensitivity
    - Photophobia
    - Phonophobia
    - Osmophobia
    - Allodynia
    - Vertigo

Postdrome
- Tiredness
- Weariness
- Concentration impairment

Source: Adapted from PJ Goadsby et al: Pathophysiology of migraine: A disorder of sensory processing. Physiol Rev 97:553, 2017.

basic level is headache with associated features, and TTH is headache that is featureless. Most patients with disabling headache probably have migraine.

Patients with acephalgic migraine (typical aura without headache, 1.2.1.2 in Table 441-1) experience recurrent neurologic symptoms, often with nausea or vomiting, but with little or no headache. Vertigo can be prominent; it has been estimated that one-third of patients referred for vertigo or dizziness have a primary diagnosis of migraine; the term vestibular migraine is often used in this setting (Table 441-1 A1.6.6). Migraine aura can have prominent brainstem symptoms, and the terms *basilar artery* and *basilar-type migraine* have now been replaced by *migraine with brainstem aura* (Table 441-1).

TREATMENT

Migraine Headache

Once a diagnosis of migraine has been established, it is important to assess the extent of a patient's disease and disability. The Migraine Disability Assessment Score (MIDAS) is a well-validated, easy-to-use tool **(Fig. 441-4)**.

Patient education is an important aspect of migraine management. Information for patients is available at websites such as the American Migraine Foundation (*www.americanmigrainefoundation.org*) and the Migraine Trust (*www.migrainetrust.org*). It is helpful for patients to understand that migraine is an inherited tendency to headache; that migraine can be modified and controlled by lifestyle adjustments and medications, but it cannot be eradicated; and that, except on some occasions in women on oral estrogens or contraceptives, migraine is not associated with serious or life-threatening illnesses.

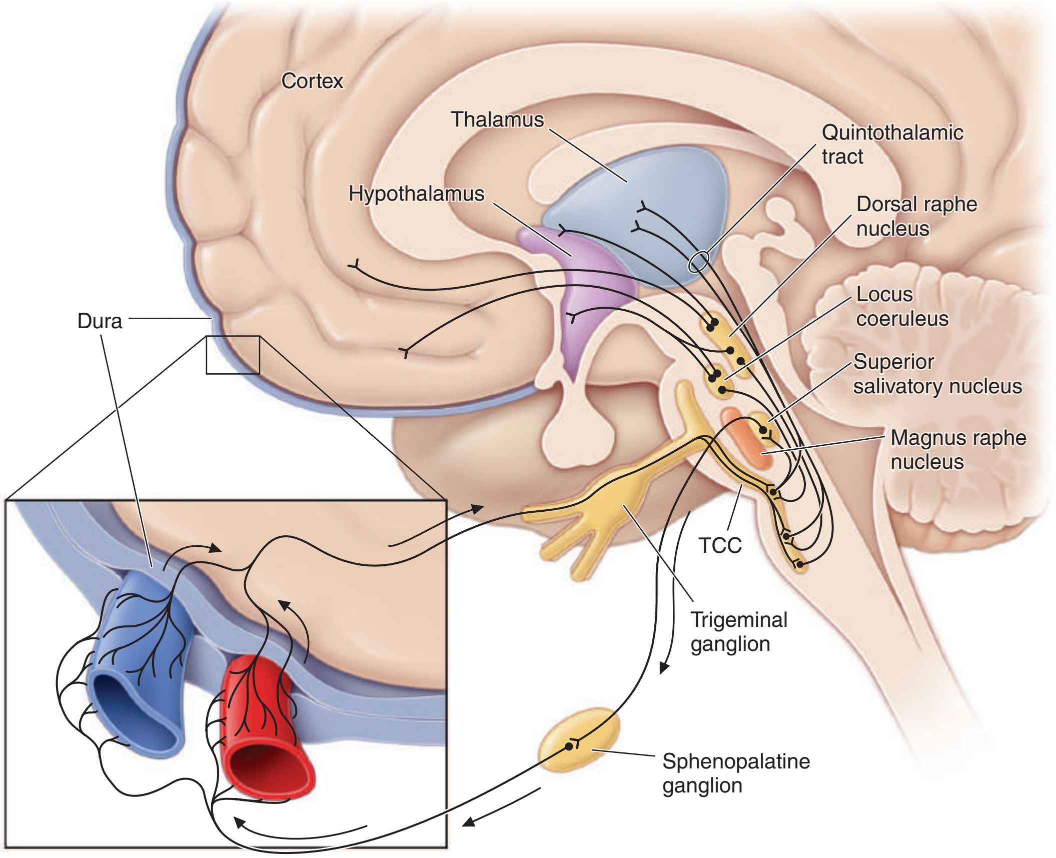

FIGURE 441-1 Brainstem pathways that modulate sensory input. The key pathway for pain in migraine is the trigeminovascular input from the meningeal vessels, which passes through the trigeminal ganglion and synapses on second-order neurons in the trigeminocervical complex (TCC). These neurons in turn project in the quintothalamic tract and, after decussating in the brainstem, synapse on neurons in the thalamus. Important modulation of the trigeminovascular nociceptive input comes from the dorsal raphe nucleus, locus coeruleus, and nucleus raphe magnus.

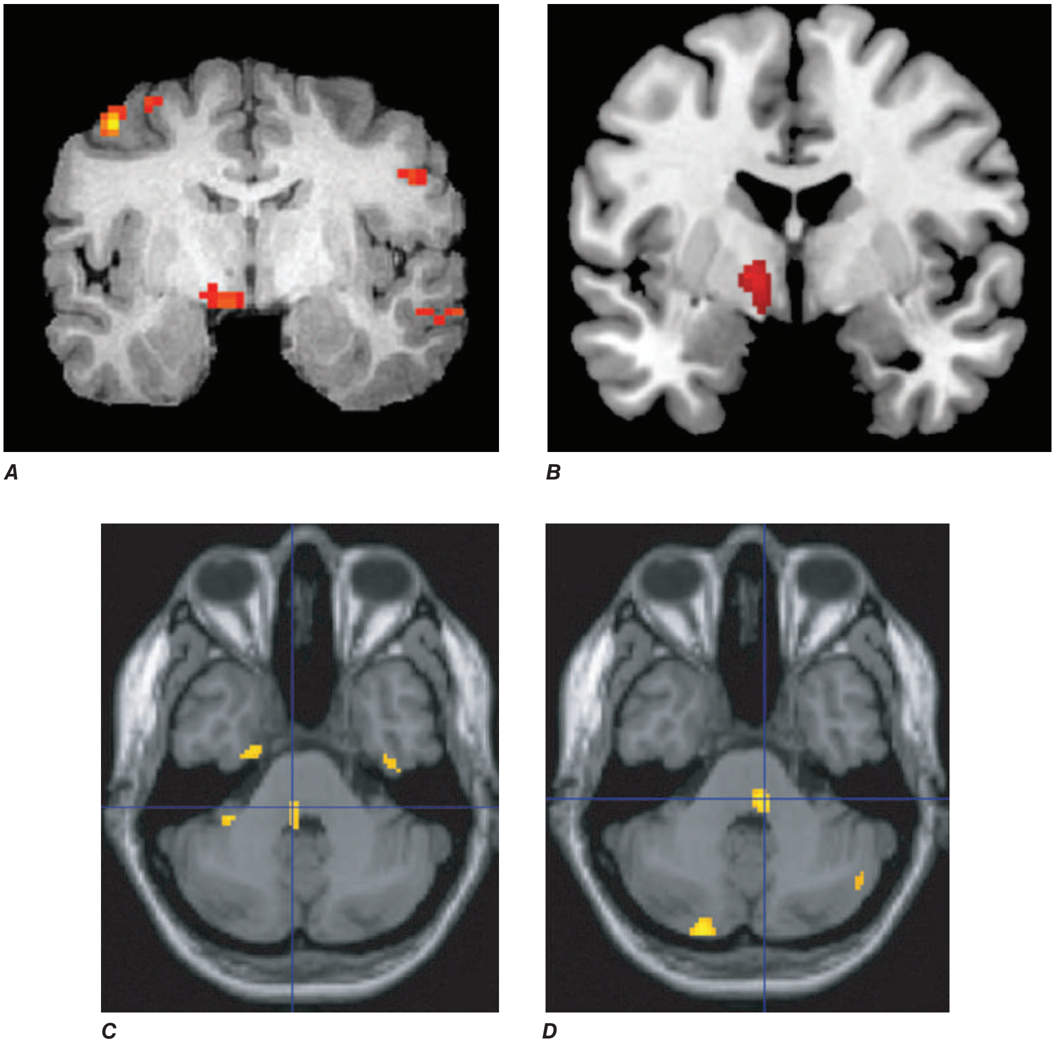

FIGURE 441-2 Positron emission tomography (PET) and arterial spin labelled activations in migraine. Hypothalamic, dorsal midbrain, and dorsolateral pontine activation are seen in triggered attacks in the premonitory phase before pain, whereas in migraine attacks, dorsolateral pontine activation persists, as it does in chronic migraine (not shown). The dorsolateral pontine area, which includes the noradrenergic locus coeruleus, is fundamental to the expression of migraine. Moreover, lateralization of changes in this region of the brainstem correlates with lateralization of the head pain in hemicranial migraine; the scans shown in panels ***C*** and ***D*** are of patients with acute migraine headache on the right and left side, respectively. *(Panel A from FH Maniyar et al: Brain activations in the premonitory phase of nitroglycerin-triggered migraine attacks. Brain 137:232, 2014. Panel B from unpublished data, Karsan and Goadsby. Panels C and D from SK Afridi et al: A PET study exploring the laterality of brainstem activation in migraine using glyceryl trinitrate. Brain 128:932, 2005.)*

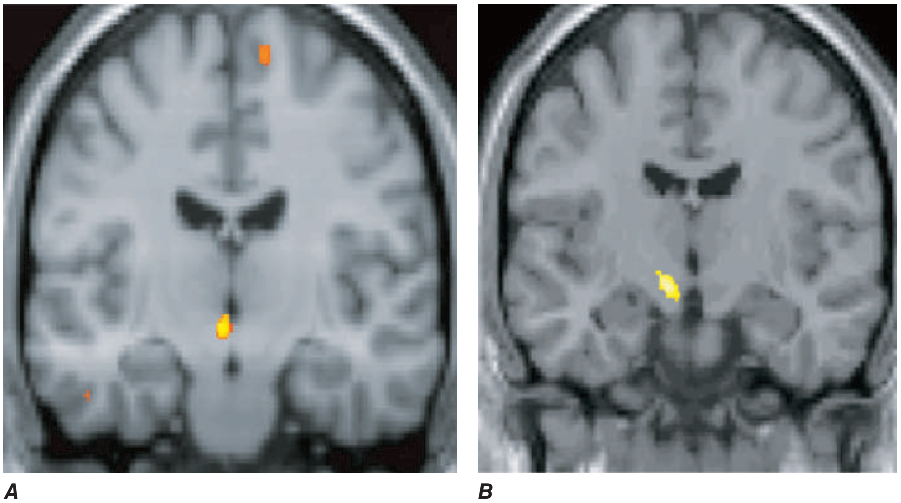

FIGURE 441-3 ***A.*** Posterior hypothalamic gray matter region activation demonstrated by positron emission tomography in a patient with acute cluster headache. ***B.*** High-resolution T1-weighted magnetic resonance image obtained using voxel-based morphometry demonstrates increased gray matter activity, lateralized to the side of pain in a patient with cluster headache. *(Panel A from A May et al: Hypothalamic activation in cluster headache attacks. Lancet 352:275, 1998. Panel B from A May et al: Correlation between structural and functional changes in brain in an idiopathic headache syndrome. Nat Med 5:836, 1999.)*

TABLE 441-3 Simplified Diagnostic Criteria for Migraine

REPEATED ATTACKS OF HEADACHE LASTING 4–72 H IN PATIENTS WITH A NORMAL PHYSICAL EXAMINATION, NO OTHER REASONABLE CAUSE FOR THE HEADACHE, AND:

AT LEAST 2 OF THE FOLLOWING FEATURES:	PLUS AT LEAST 1 OF THE FOLLOWING FEATURES:
Unilateral pain	Nausea/vomiting
Throbbing pain	Photophobia and phonophobia
Aggravation by movement	
Moderate or severe intensity	

Source: Adapted from the International Headache Society Classification (Headache Classification Committee of the International Headache Society, Cephalalgia 38:1, 2018).

NONPHARMACOLOGIC MANAGEMENT

Migraine can often be managed to some degree by a variety of nonpharmacologic approaches. When patients can identify reliable triggers, their avoidance can be useful. A regulated lifestyle is helpful, including a healthy diet, regular exercise, regular sleep patterns, avoidance of excess caffeine and alcohol, and avoidance of acute changes in stress levels, being particularly wary of the letdown effect.

The measures that benefit a given individual should be used routinely because they provide a simple, cost-effective approach to migraine management. Patients with migraine do not encounter more stress than headache-free individuals; overresponsiveness to changes in stress appears to be the issue. Because the stresses of everyday living cannot be eliminated, lessening one's response to stress by various techniques is helpful for many patients. These may include yoga, transcendental meditation, hypnosis, and conditioning techniques such as biofeedback. For most patients seen in clinical practice, this approach is, at best, an adjunct to pharmacotherapy. Nonpharmacologic measures are unlikely to prevent all migraine attacks, and pharmacologic approaches are often needed.

ACUTE ATTACK THERAPIES FOR MIGRAINE

The mainstay of pharmacologic therapy is the judicious use of one or more of the many medicines that are effective in migraine **(Table 441-4)**. The selection of the optimal regimen for a given patient depends on a number of factors, the most important of which is the severity of the attack. Mild migraine attacks can usually be managed by oral agents; the average efficacy (pain relief) rate is 50–70%. Severe migraine attacks may require parenteral therapy. Most drugs effective in the treatment of migraine are members of one of five major pharmacologic classes: nonsteroidal anti-inflammatory drugs; 5-HT$_{1B/1D}$ receptor agonists—triptans; CGRP receptor antagonists—gepants; 5-HT$_{1F}$ receptor agonists—ditans; and dopamine receptor antagonists.

In general, an adequate dose of whichever agent is chosen should be used as soon as possible after the onset of an attack. If additional medication is required within 60 min because symptoms return or have not abated, the initial dose should be increased for subsequent attacks or a different class of drug tried as first-line treatment. Repeat dosing of the same medicine at 2 h, while safe, is ineffective for triptans, and in contrast effective for gepants. Migraine therapy must be individualized; a standard approach for all patients is not possible. A therapeutic regimen may need to be refined until one is identified that provides the patient with rapid, complete, and consistent relief with minimal side effects **(Table 441-5)**.

Nonsteroidal Anti-Inflammatory Drugs (NSAIDs) Both the severity and duration of a migraine attack can be reduced significantly by NSAIDs (Table 441-4). Indeed, many undiagnosed migraineurs self-treat with nonprescription NSAIDs. A general consensus is that NSAIDs are most effective when taken early in the migraine attack. However, the effectiveness of these agents in migraine is usually less than optimal in moderate or severe migraine attacks. The combination of acetaminophen (paracetamol), aspirin, and caffeine has been approved for use by the U.S. Food and Drug Administration (FDA) for the treatment of mild to moderate migraine. The combination of aspirin and metoclopramide has been shown to be

***MIDAS Questionnaire**

INSTRUCTIONS: Please answer the following questions about ALL headaches you have had over the last 3 months. Write zero if you did not do the activity in the last 3 months.

1. On how many days in the last 3 months did you miss work or school because of your headaches?.. ____ days
2. How many days in the last 3 months was your productivity at work or school reduced by half or more because of your headaches *(do not include days you counted in question 1 where you missed work or school)*?.......................... ____ days
3. On how many days in the last 3 months did you **not** do household work because of your headaches? .... ____ days
4. How many days in the last 3 months was your productivity in household work reduced by half or more because of your headaches *(do not include days you counted in question 3 where you did not do household work)*?.................... ____ days
5. On how many days in the last 3 months did you miss family, social, or leisure activities because of your headaches? .... ____ days

A. On how many days in the last 3 months did you have a headache? *(If a headache lasted more than one day, count each day.)*....................................... ____ days

B. On a scale of 0–10, on average how painful were these headaches? *(Where 0 = no pain at all, and 10 = pain as bad as it can be.)*.. ____

*Migraine Disability Assessment Score
(Questions 1–5 are used to calculate the MIDAS score.)
Grade I—Minimal or Infrequent Disability: 0–5
Grade II—Mild or Infrequent Disability: 6–10
Grade III—Moderate Disability: 11–20
Grade IV—Severe Disability: > 20

© Innovative Medical Research 1997

FIGURE 441-4 The Migraine Disability Assessment Score (MIDAS) Questionnaire. *(Courtesy of Dr. Richard Lipton.)*

TABLE 441-4 TREATMENT of Acute Migraine

DRUG	TRADE NAME	DOSAGE
Simple Analgesics		
Acetaminophen, aspirin, caffeine	Excedrin Migraine	Two tablets or caplets q6h (max 8 per day)
NSAIDs		
Naproxen	Aleve, Anaprox, generic	220–550 mg PO bid
Ibuprofen	Advil, Motrin, Nuprin, generic	400 mg PO q3–4h
Tolfenamic acid	Clotam Rapid	200 mg PO; may repeat ×1 after 1–2 h
Diclofenac K	Cambia	50 mg PO with water
$5\text{-HT}_{1B/1D}$ Receptor Agonists—Triptans		
Oral		
Ergotamine 1 mg, caffeine 100 mg	Cafergot	One or two tablets at onset, then one tablet q½h (max 6 per day, 10 per week)
Naratriptan	Amerge	2.5-mg tablet at onset
Rizatriptan	Maxalt Maxalt-MLT	5–10-mg tablet at onset
Sumatriptan	Imitrex	50–100-mg tablet at onset
Frovatriptan	Frova	2.5-mg tablet at onset
Almotriptan	Axert	12.5-mg tablet at onset
Eletriptan	Relpax	40 or 80 mg at onset
Zolmitriptan	Zomig Zomig Rapimelt	2.5-mg tablet at onset
Nasal		
Dihydroergotamine	Migranal Nasal Spray Trudhesa Nasal Spray	Prior to nasal spray, the pump must be primed 4 times; 1 spray (0.5 mg) is administered, followed in 15 min by a second spray One spray into each nostril
Sumatriptan	Imitrex Nasal Spray	5–20 mg intranasal spray as 4 sprays of 5 mg or a single 20 mg spray
Zolmitriptan	Zomig	5 mg intranasal spray as one spray
Parenteral		
Dihydroergotamine	DHE-45	1 mg IV, IM, or SC at onset and q1h (max 3 mg/d, 6 mg per week)
Sumatriptan	Imitrex Injection Alsuma Sumavel DosePro	3, 4, or 6 mg SC at onset (may repeat once after 1 h for max of 2 doses in 24 h)
CGRP Receptor Antagonists—Gepants		
Oral		
Rimegepant	Nurtec	75 mg ODT PO
Ubrogepant	Ubrelvy	50 or 100 mg PO; a second dose may be taken 2 h after the first, if needed
Nasal		
Zavegepant	Zavzpret	10 mg intranasal, single spray to one nostril once in 24 h
5-HT_{1F} Receptor Agonist—Ditans		
Oral		
Lasmiditan	Reyvow	50, 100, or 200 mg PO
Dopamine Receptor Antagonists		
Oral		
Metoclopramide	Reglan,[a] generic[a]	5–10 mg/d
Prochlorperazine	Compazine,[a] generic[a]	1–25 mg/d
Parenteral		
Chlorpromazine	Generic[a]	0.1 mg/kg IV at 2 mg/min; max 35 mg/d
Metoclopramide	Reglan,[a] generic	10 mg IV
Prochlorperazine	Compazine,[a] generic[a]	10 mg IV
Other		
Parenteral		
Opioids	Generic[a]	Multiple preparations and dosages; **see Table 14-1**
Other		
Neuromodulation		
Single-pulse transcranial magnetic stimulation (sTMS)	Savi Dual	Two pulses at onset followed by two further pulses
Noninvasive vagus nerve stimulation (nVNS)	gammaCore	Two doses each of 120 s
Remote electrical neuromodulation (REN)	Nerivio	30- to 45-min stimulation to the upper arm
Transcutaneous supraorbital nerve stimulation	Cefaly	60-min stimulation
External concurrent occipital and trigeminal neurostimulation (eCOT-NS)	Relivion	30- to 60-min stimulation

[a] Not all drugs are specifically indicated by the U.S. Food and Drug Administration for migraine. Local regulations and guidelines should be consulted.

Note: Antiemetics (e.g., domperidone 10 mg or ondansetron 4 or 8 mg) or prokinetics (e.g., metoclopramide 10 mg) are sometimes useful adjuncts.

Abbreviations: 5-HT, 5-hydroxytryptamine; NSAIDs, nonsteroidal anti-inflammatory drugs; ODT, orally disintegrating tablets.

TABLE 441-5 Clinical Stratification of Acute Specific Migraine Treatments

CLINICAL SITUATION	TREATMENT OPTIONS
Failed NSAIDs/ analgesics	**First tier**
	Sumatriptan 50 mg or 100 mg PO
	Almotriptan 12.5 mg PO
	Rizatriptan 10 mg PO
	Eletriptan 40 mg PO
	Zolmitriptan 2.5 mg PO
	Rimegepant 75 mg
	Ubrogepant 50 or 100 mg
	Lasmiditan 50, 100, or 200 mg
	Slower effect/better tolerability
	Naratriptan 2.5 mg PO
	Frovatriptan 2.5 mg PO
	Infrequent headache
	Ergotamine/caffeine 1–2/100 mg PO
	Dihydroergotamine nasal spray 2 mg
Early nausea or difficulties taking tablets	Zolmitriptan 5 mg nasal spray
	Sumatriptan 20 mg nasal spray
	Rizatriptan 10 mg MLT wafer
	Zavegepant 10 mg nasal spray
Headache recurrence	Ergotamine 2 mg (most effective PR/usually with caffeine)
	Naratriptan 2.5 mg PO
	Almotriptan 12.5 mg PO
	Eletriptan 40 mg
	Rimegepant 75 mg
	Ubrogepant 50 or 100 mg
Tolerating acute treatments poorly	Naratriptan 2.5 mg
	Almotriptan 12.5 mg
	Rimegepant 75 mg
	Ubrogepant 50, 100 mg
	Single-pulse transcranial magnetic stimulation
	Noninvasive vagus nerve stimulation
	Remote electrical neuromodulation
Early vomiting	Zolmitriptan 5 mg nasal spray
	Zavegepant 10 mg nasal spray
	Sumatriptan 25 mg PR
	Sumatriptan 6 mg SC
Menses-related headache	**Prevention**
	Ergotamine PO at night
	Estrogen patches
	Rimegepant 75 mg PO taken during the menses
	Treatment
	Triptans
	Dihydroergotamine nasal spray
Very rapidly developing symptoms	Zolmitriptan 5 mg nasal spray
	Zavegepant 10 mg nasal spray
	Sumatriptan 6 mg SC
	Dihydroergotamine 1 mg IM

Abbreviation: NSAIDs, nonsteroidal anti-inflammatory drugs.

comparable to a single dose of oral sumatriptan. Important side effects of NSAIDs include dyspepsia and gastrointestinal irritation.

5-HT$_{1B/1D}$ RECEPTOR AGONISTS—TRIPTANS AND ERGOTS

Oral Stimulation of 5-HT$_{1B/1D}$ receptors can stop an acute migraine attack. Ergotamine and dihydroergotamine are nonselective receptor agonists, whereas the triptans are selective 5-HT$_{1B/1D}$ receptor agonists. A variety of triptans—sumatriptan, almotriptan, eletriptan, frovatriptan, naratriptan, rizatriptan, and zolmitriptan—are available for the treatment of migraine.

Each drug in the triptan class has similar pharmacologic properties, varying slightly in terms of clinical efficacy. Rizatriptan and eletriptan are, on a population basis, the most efficacious of the triptans. Sumatriptan and zolmitriptan have similar rates of efficacy as well as time to onset, with an advantage of having multiple formulations, whereas almotriptan has a similar rate of efficacy to sumatriptan and is better tolerated, and frovatriptan and naratriptan are somewhat slower in onset and are also well tolerated. Clinical efficacy appears to be related more to the t_{max} (time to peak plasma level) than to the potency, half-life, or bioavailability. This observation is consistent with a large body of data indicating that faster-acting analgesics are more effective than slower-acting ones.

Unfortunately, monotherapy with a selective oral 5-HT$_{1B/1D}$ receptor agonist does not result in rapid, consistent, and complete relief of migraine in all patients. Triptans are generally not effective in migraine with aura unless given after the aura is completed and the headache initiated. Side effects are common, although often mild and transient. Moreover, 5-HT$_{1B/1D}$ receptor agonists are contraindicated in individuals with a history, symptoms, or signs of ischemic cardiac, cerebrovascular, or peripheral vascular syndromes. Recurrence of headache, within the usual time course of an attack, is another important limitation of triptan use and occurs at least occasionally in most patients. Evidence from randomized controlled trials shows that coadministration of a longer-acting NSAID, naproxen 500 mg, with sumatriptan will augment the initial effect of sumatriptan and, importantly, reduce rates of headache recurrence.

Ergotamine preparations offer a nonselective means of stimulating 5-HT$_1$ receptors. A nonnauseating dose of ergotamine should be sought because a dose that provokes nausea is too high and may intensify head pain. Oral (excluding sublingual) formulations of ergotamine also contain 100 mg caffeine (theoretically to enhance ergotamine absorption and possibly to add additional analgesic activity). The average oral ergotamine dose for a migraine attack is 2 mg. Because the clinical studies demonstrating the efficacy of ergotamine in migraine predated the clinical trial methodologies used with the triptans, it is difficult to assess the comparative efficacy of ergotamine versus the triptans. In general, with use of ergotamine there appears to be a much higher incidence of nausea than with triptans but less headache recurrence.

Nasal Nasal formulations of dihydroergotamine, zolmitriptan, or sumatriptan can be useful in patients requiring a nonoral route of administration. The nasal sprays result in substantial blood levels within 30–60 min. Although in theory nasal sprays might provide faster and more effective relief of a migraine attack than oral formulations, their reported pain relief rate is only ~50–60%. Studies with a new inhalational formulation of dihydroergotamine indicate that its absorption problems can be overcome to produce rapid onset of action with good tolerability.

Parenteral Administration of drugs by injection, such as dihydroergotamine and sumatriptan, is approved by the FDA for the rapid relief of a migraine attack. Peak plasma levels of dihydroergotamine are achieved 3 min after IV dosing, 30 min after intramuscular (IM) dosing, and 45 min after subcutaneous (SC) dosing. If an attack has not already peaked, SC or IM administration of 1 mg of dihydroergotamine is adequate for about 80–90% of patients. Sumatriptan, 3, 4, or 6 mg SC, depending on local availability, is effective in ~50–80% of patients and can now be administered by a needle-free device.

CALCITONIN GENE-RELATED PEPTIDE (CGRP) RECEPTOR ANTAGONISTS—GEPANTS

Gepants are small-molecule CGRP receptor antagonists that are effective in the acute treatment of migraine. Three are currently approved by the FDA: rimegepant, and ubrogepant are oral, and zavegepant is a nasal spray (Table 441-4). They are more likely to render patients pain-free at 2 h and most bothersome symptom–free when compared with placebo in large phase 3 clinical trials. The most bothersome symptom is derived by asking patients to identify which symptom—of nausea, photophobia, or phonophobia—was most bothersome during the treated attack; success required that this symptom was eliminated at 2 h. Gepants are extremely well tolerated with only a small percentage of patients reporting troublesome side effects, such as mild nausea.

5-HT$_{1F}$ RECEPTOR AGONISTS—DITANS

Lasmiditan, a highly selective, orally available, 5-HT$_{1F}$ receptor agonist, has been approved by the FDA for the acute treatment of migraine based on large phase 3 studies where it was superior to placebo (Table 441-4). Ditans have no vascular effects because the 5-HT$_{1F}$ receptor is located in the central and peripheral nervous system but not vasculature; the class thus unequivocally fills a gap in therapy for patients with cardiovascular and cerebrovascular disease. The major side effect is dizziness, occurring in ~15% of patients in clinical trials, and somnolence in 6%. Patients are advised not to drive for 8 h after treatment.

DOPAMINE RECEPTOR ANTAGONISTS

Oral Oral dopamine receptor antagonists can be considered as adjunctive therapy in migraine. Drug absorption is impaired during migraine because of reduced gastrointestinal motility. Delayed absorption occurs even in the absence of nausea and is related to the severity of the attack and not its duration. Therefore, when oral NSAIDs and/or triptan agents fail, the addition of a dopamine receptor antagonist, such as metoclopramide 10 mg, prochlorperazine 10 mg, or domperidone 10 mg (not available in the United States), should be considered to enhance gastric absorption. In addition, dopamine receptor antagonists decrease nausea/vomiting and restore normal gastric motility.

Parenteral Dopamine receptor antagonists (e.g., chlorpromazine, prochlorperazine, metoclopramide) by injection can also provide significant acute relief of migraine; they can be used in combination with parenteral 5-HT$_{1B/1D}$ receptor agonists. A common IV protocol used for the treatment of severe migraine is the administration over 2 min of a mixture of 5 mg of prochlorperazine and 0.5 mg of dihydroergotamine.

OTHER OPTIONS FOR ACUTE MIGRAINE

Oral For milder attacks of migraine, the combination of acetaminophen, aspirin, and caffeine is FDA approved (Table 441-4).

Parenteral Opioids are modestly effective in the acute treatment of migraine. For example, IV meperidine (50–100 mg) is given frequently in the emergency department (ED). This regimen "works" in the sense that the pain of migraine is eliminated. Importantly, it is clear from a randomized controlled trial that prochlorperazine is superior to hydromorphone in the ED setting. However, opioids are clearly suboptimal for patients with recurrent headache. Opioids do not treat the underlying headache mechanism; rather, they act to alter the pain sensation, and there is evidence their use may decrease the likelihood of a response to triptans in the future. Moreover, in patients taking oral opioids, such as oxycodone or hydrocodone, habituation or addiction can greatly confuse the treatment of migraine. Opioid craving and/or withdrawal can aggravate and accentuate migraine. Therefore, it is recommended that opioid use in migraine be limited to patients with severe, but infrequent, headaches that are unresponsive to other pharmacologic approaches or who have contraindications to other therapies.

Neuromodulation Single-pulse transcranial magnetic stimulation (sTMS) is FDA cleared for the acute treatment of migraine. Two pulses can be applied at the onset of an attack, and this can be repeated. The use of sTMS is safe where there is no cranial metal implant and offers an option to patients seeking nonpharmaceutical approaches to treatment. Similarly, a noninvasive vagus nerve stimulator (nVNS) is FDA cleared for the treatment of migraine attacks in adults. One to two 120-s doses may be applied for attack treatment. Remote electrical neuromodulation using a smartphone app that stimulates the upper arm for 30–45 min is also effective for treatment of acute migraine. Transcutaneous supraorbital nerve stimulation for 60 min and external concurrent occipital and trigeminal neurostimulation (eCOT-NS) for 30–60 min are both FDA cleared.

MEDICATION-OVERUSE HEADACHE

Acute attack medications, particularly opioid or barbiturate-containing compound analgesics, have a propensity to aggravate headache frequency and induce a state of refractory daily or near-daily headache called *medication-overuse headache*. This condition is likely not a separate headache entity but a reaction of the patient's underlying migraine biology to a particular medicine. Migraine patients who have two or more headache days a week should be cautioned about frequent analgesic use **(see "Chronic Daily Headache" in Chap. 17)**.

PREVENTIVE TREATMENTS FOR MIGRAINE

Patients with an increasing frequency of migraine attacks or with attacks that are either unresponsive or poorly responsive to abortive treatments are good candidates for preventive agents. In general, a preventive medication should be considered in patients with four or more migraine days a month. Significant side effects are associated with the use of many agents; furthermore, determination of dose can be difficult because the recommended doses have been derived for conditions other than migraine. The mechanism of action of older medicines is unclear; it seems likely that the brain sensitivity that underlies migraine is modified. Patients are usually started on a low dose of a chosen treatment; the dose is then gradually increased, up to a reasonable maximum, to achieve clinical benefit.

Treatments that have the capacity to stabilize migraine are listed in **Table 441-6**. Most treatments must be taken daily, and there is usually a lag of 2–12 weeks before an effect is seen. The drugs that have been approved by the FDA for the preventive treatment of migraine include propranolol, timolol, sodium valproate, topiramate, eptinezumab, erenumab, fremanezumab, galcanezumab, rimegepant, and atogepant. In addition, a number of other drugs appear to display preventive efficacy. This group includes amitriptyline, candesartan, nortriptyline, flunarizine, phenelzine, and cyproheptadine. Placebo-controlled trials of onabotulinum toxin type A in episodic migraine were negative, whereas, overall, placebo-controlled trials in chronic migraine were positive. The FDA has approved a range of neuromodulation approaches for the preventive treatment of migraine (Table 441-6). They offer a well-tolerated, effective option for patients. Phenelzine is a monoamine oxidase inhibitor (MAOI); therefore, tyramine-containing foods, decongestants, and meperidine are contraindicated, and it is reserved for only very recalcitrant cases. Methysergide is now of historical interest only because it is no longer manufactured. Melatonin has been reported to be useful, with controlled trial evidence, but is not approved for this indication in the United States.

The probability of success with any one of the antimigraine drugs is ~40–50%. Many patients are managed adequately with well-tolerated doses of candesartan, propranolol, amitriptyline, topiramate, or valproate. As data on fetal developmental issues have arisen, both topiramate and valproate are now considered less attractive for use in females during reproductive years. If these agents fail or produce unacceptable side effects, neuromodulation approaches can be used (Table 441-6). Once effective stabilization is achieved, the drug is continued for ~6–12 months and then slowly tapered, assuming the patient agrees, to assess the continued need. Many patients are able to discontinue medication and experience fewer and milder attacks for long periods. The advent of CGRP monoclonal antibodies and CGRP receptor antagonists has significantly changed the landscape of preventive treatment; with the combination of efficacy that is often within the first month and excellent tolerability, expectations of outcomes have changed.

TABLE 441-6 Preventive Treatments in Migraine[a]

DRUG	DOSE	SELECTED SIDE EFFECTS
Beta blocker		
Propranolol	40–120 mg bid	Reduced energy
Metoprolol	25–100 mg bid	Tiredness
		Postural symptoms
		Contraindicated in asthma
Antidepressants		
Amitriptyline	10–75 mg at night	Drowsiness
Dosulepin	25–75 mg at night	
Nortriptyline	25–75 mg at night	*Note:* Some patients may only need a total dose of 10 mg, although generally 1–1.5 mg/kg body weight is required.
Venlafaxine	75–150 mg/d	
Anticonvulsants		
Topiramate	25–200 mg/d	Paresthesias
		Cognitive symptoms
		Weight loss
		Glaucoma
		Caution with nephrolithiasis
Valproate	400–600 mg bid	Drowsiness
		Weight gain
		Tremor
		Hair loss
		Fetal abnormalities
		Hematologic or liver abnormalities
Serotonergic drugs	0.5–2 mg qd	Weight gain
Pizotifen[b]		
CGRP pathway blockers		
Eptinezumab	100 or 300 mg IV every 12 weeks	Nasopharyngitis
Erenumab	70 or 140 mg SC monthly	Nasopharyngitis, constipation
Fremanezumab	225 mg monthly or 675 mg q3 months, SC	Injection site reactions
Galcanezumab	240 mg loading then 120 mg monthly, SC	Nasopharyngitis
Rimegepant	75 mg every other day	Nausea abdominal pain/dyspepsia
Atogepant	10, 30, or 60 mg once daily	Constipation, nausea
Other classes		
Flunarizine[b]	5–15 mg qd	Drowsiness
		Weight gain
		Depression
		Parkinsonism
Candesartan	4–24 mg daily	Dizziness
Memantine	5–20 mg daily	Dizziness, tiredness
Melatonin	3–12 mg nightly	Drowsiness
Neuromodulation		Lightheadedness
Single-pulse transcranial magnetic stimulation (sTMS)	4–24 pulses per day	Tingling
Noninvasive vagus nerve stimulation (nVNS)	120-s treatments 2–3 times daily	Tinnitus
Remote electrical neuromodulation (REN)	45 min every other day	Site discomfort, irritation or pain
Transcutaneous supraorbital nerve stimulation	20 min daily	Muscle twitching
		Well-tolerated; some local sensory symptoms
		Local paresthesia
Chronic migraine		
Onabotulinum toxin type A	155 U	Loss of brow furrow
No convincing evidence from controlled trials		
Verapamil		
Controlled trials demonstrate *no effect*		
Nimodipine		
Clonidine		
Selective serotonin reuptake inhibitors: fluoxetine		

[a] Commonly used preventives are listed with typical doses and common side effects. Not all listed medicines are approved by the U.S. Food and Drug Administration; local regulations and guidelines should be consulted. [b] Not available in the United States.

TENSION-TYPE HEADACHE

Clinical Features The term *tension-type headache* is commonly used to describe a chronic head-pain syndrome characterized by bilateral tight, bandlike discomfort. The pain typically builds slowly, fluctuates in severity, and may persist more or less continuously for many days. The headache may be episodic or chronic (present >15 days per month).

A useful clinical approach is to diagnose TTH in patients whose headaches are completely without accompanying features such as nausea, vomiting, photophobia, phonophobia, osmophobia, throbbing, and aggravation with movement. Such an approach neatly separates migraine, which has one or more of these features and is the main differential diagnosis, from TTH. The International Headache Society's main definition of TTH allows an admixture of nausea, photophobia, or phonophobia in various combinations, although the appendix definition does not; this illustrates the difficulty in distinguishing these two clinical entities. In clinical practice, using the appendix definition to dichotomize patients on the basis of the presence of associated features (migraine) and the absence of associated features (TTH) is highly recommended. Indeed, patients whose headaches fit the TTH phenotype and who have migraine at other times, along with a family history of migraine, migrainous illnesses of childhood, or typical migraine triggers to their migraine attacks, may be biologically different from those who have TTH headache with none of the features. TTH may be infrequent (episodic) or occur on 15 days or more a month (chronic).

Pathophysiology The pathophysiology of TTH is incompletely understood. It seems likely that TTH is due to a primary disorder of central nervous system pain modulation alone, unlike migraine, which involves a more generalized disturbance of sensory modulation. Data suggest a genetic contribution to TTH, but this may not be a valid finding: given the current diagnostic criteria, the studies undoubtedly included many migraine patients. The name *tension-type headache* implies that pain is a product of *nervous tension*, but there is no clear evidence for tension as an etiology. Muscle contraction has been considered to be a feature that distinguishes TTH from migraine, but there appear to be no differences in contraction between the two headache types. In primary care, <10% of patients presenting with headache over 3 months or more without any neurologic symptoms or signs have tension-type headache; >90% have migraine.

TREATMENT

Tension-Type Headache

The pain of TTH can generally be managed with simple analgesics such as acetaminophen, aspirin, or NSAIDs. Behavioral approaches including relaxation can also be effective. Clinical studies have demonstrated that triptans in pure TTH are not helpful, although triptans are effective in apparent TTH when the patient also has migraine. For chronic TTH, amitriptyline is the only proven treatment (Table 441-6); other tricyclics, selective serotonin reuptake inhibitors, and the benzodiazepines have not been shown to be effective. There is no evidence for the efficacy of acupuncture. Placebo-controlled trials of onabotulinum toxin type A in chronic TTH were negative.

TRIGEMINAL AUTONOMIC CEPHALALGIAS, INCLUDING CLUSTER HEADACHE

The TACs describe a grouping of primary headache disorders including cluster headache, paroxysmal hemicrania (PH), SUNCT (*s*hort-lasting *u*nilateral *n*euralgiform headache attacks with *c*onjunctival injection and *t*earing)/SUNA (*s*hort-lasting *u*nilateral *n*euralgiform headache attacks with cranial *a*utonomic symptoms), and hemicrania continua (Table 441-1). TACs are characterized by relatively short-lasting attacks of head pain associated with lateralized cranial autonomic symptoms, such as lacrimation, conjunctival injection, aural fullness, or nasal congestion **(Table 441-7)**. Pain is usually severe and may occur more than once a day. Because of the associated nasal congestion or rhinorrhea, patients are often misdiagnosed with "sinus headache" and treated with decongestants, which are ineffective.

TABLE 441-7 Clinical Features of the Trigeminal Autonomic Cephalalgias

	CLUSTER HEADACHE	PAROXYSMAL HEMICRANIA	SUNCT/SUNA
Gender	M > F	F = M	F ~ M
Pain			
Type	Stabbing, boring	Throbbing, boring, stabbing	Burning, stabbing, sharp
Severity	Excruciating	Excruciating	Severe to excruciating
Site	Orbit, temple	Orbit, temple	Periorbital
Attack frequency	1/alternate day–8/d	1–20/d (>5/d for more than half the time)	3–200/d
Duration of attack	15–180 min	2–30 min	5–240 s
Autonomic features	Yes	Yes	Yes (prominent conjunctival injection and lacrimation)[a]
Migrainous features[b]	Yes	Yes	Yes
Alcohol trigger	Yes	No	No
Cutaneous triggers	No	No	Yes
Indomethacin effect	—	Yes[c]	—
Abortive treatment	Sumatriptan injection or nasal spray Zolmitriptan nasal spray Oxygen nVNS[c]	No effective treatment	Lidocaine (IV)
Preventive treatment	Verapamil	Indomethacin[d]	Lamotrigine
	Galcanezumab	nVNS	
	Topiramate		Topiramate
	Melatonin		
	Lithium		Gabapentin

[a] If conjunctival injection and tearing are not present, consider SUNA. [b] Nausea, photophobia, or phonophobia; photophobia and phonophobia are typically unilateral on the side of the pain. [c] Noninvasive vagus nerve stimulation is U.S. Food and Drug Administration approved in episodic cluster headache [d] Indicates complete response to indomethacin.

Abbreviations: nVNS, non-invasive vagus nerve stimulation; SUNA, *s*hort-lasting *u*nilateral *n*euralgiform headache attacks with cranial *a*utonomic features; SUNCT, *s*hort-lasting *u*nilateral *n*euralgiform headache attacks with *c*onjunctival injection and *t*earing.

TACs must be differentiated from short-lasting headaches that do not have prominent cranial autonomic syndromes, notably trigeminal neuralgia (TN), primary stabbing headache, and hypnic headache. The cycling pattern and length, frequency, and timing of attacks are useful in classifying patients. Patients with TACs should be considered, if clinically indicated, to undergo pituitary imaging and pituitary function tests because there is an excess of TAC presentations in clinical practice in patients with pituitary tumor–related headache, particularly prolactin and growth hormone secreting tumors.

Cluster Headache Cluster headache is a relatively rare form of primary headache, although nonetheless a common condition, with a population frequency of ~0.1%. The pain is deep, usually retroorbital, often excruciating in intensity, nonfluctuating, and explosive in quality. A core feature of cluster headache is periodicity. Usually one of the daily attacks of pain recurs at about the same hour each day for the duration of a cluster bout. The typical cluster headache patient has daily bouts of one to two attacks of relatively short-duration unilateral pain for 8–10 weeks a year; this is usually followed by a pain-free interval that averages a little less than 1 year. Cluster headache is characterized as chronic when there is <3 months of sustained remission without treatment. Patients are generally perfectly well between episodes. Onset of attacks is nocturnal in about 50% of patients, and men are affected three times more often than women. Patients with cluster headache tend to move about during attacks, pacing, rocking, or rubbing their head for relief; some may even become aggressive during attacks. This is in sharp contrast to patients with migraine, who prefer to remain motionless during attacks.

Cluster headache is associated with ipsilateral symptoms of cranial parasympathetic autonomic activation: conjunctival injection or lacrimation, aural fullness, rhinorrhea or nasal congestion, or cranial sympathetic dysfunction such as ptosis. The sympathetic deficit is peripheral and likely to be due to parasympathetic activation with injury to ascending sympathetic fibers surrounding a dilated carotid artery as it passes into the cranial cavity. When present, photophobia and phonophobia are more likely to be unilateral and on the same side of the pain, rather than bilateral, in contrast to migraine. This phenomenon of unilateral photophobia/phonophobia is characteristic of TACs. Cluster headache is likely to be a disorder involving central pacemaker neurons and neurons in the posterior hypothalamic region (Fig. 441-3).

TREATMENT

Cluster Headache

The most satisfactory treatment is the administration of drugs to prevent cluster attacks until the bout is over. However, treatment of acute attacks is required for all cluster headache patients at some time.

ACUTE ATTACK TREATMENT

Cluster headache attacks peak rapidly, and thus a treatment with rapid onset is required. Many patients with acute cluster headache respond very well to oxygen inhalation. This should be given as 100% oxygen at 10–12 L/min for 15–20 min. It appears that high flow and high oxygen content are important, thus ultra-high flow with a demand valve may be a better option for some patients. Sumatriptan 6 mg SC is rapid in onset and will usually shorten an attack to 10–15 min; where available, 3 or 4 mg SC can also be effective; there is no evidence of tachyphylaxis. Sumatriptan (20 mg) and zolmitriptan (5 mg) nasal sprays are both effective in acute cluster headache, offering a useful option for patients who may not wish to self-inject daily. nVNS is FDA cleared for the acute treatment of attacks in episodic cluster headache using three 2-min stimulation cycles applied consecutively at the onset of headache on the side of pain; this may be repeated after 9 min. Oral sumatriptan is not effective for prevention or for acute treatment of cluster headache.

PREVENTIVE TREATMENTS (TABLE 441-8)

The choice of a preventive treatment in cluster headache depends in part on the length of the bout. Patients with long bouts or those with chronic cluster headache require medicines that are safe when taken for long periods. For patients with relatively short bouts, limited courses of oral glucocorticoids can be very useful. A 10-day course of prednisone, beginning at 60 mg daily for 7 days and followed by a rapid taper, may interrupt the pain bout for many patients. Greater occipital nerve injection with lidocaine and corticosteroids has been shown to be effective in randomized controlled trials, with a benefit that lasts up to 6–8 weeks. The CGRP monoclonal antibody galcanezumab has been approved by the FDA for treatment of episodic cluster headache; it reduces attack frequency, is well tolerated, and is often an effective option.

TABLE 441-8 Preventive Management of Cluster Headache

SHORT-TERM PREVENTION	LONG-TERM PREVENTION
EPISODIC CLUSTER HEADACHE	EPISODIC CLUSTER HEADACHE AND PROLONGED CHRONIC CLUSTER HEADACHE
Prednisone 1 mg/kg up to 60 mg qd, tapering over 21 days Verapamil 160–960 mg/d Galcanezumab 300 mg SC Greater occipital nerve injection (local anesthetic and corticosteroids)	Verapamil 160–960 mg/d nVNS 6–24 stimulations/d Melatonin[a] 9–12 mg/d Topiramate[a] 100–400 mg/d Lithium 400–800 mg/d

Abbreviations: nVNS, noninvasive vagus nerve stimulation.

Most experts favor verapamil as the first-line preventive treatment for patients with chronic cluster headache or with prolonged bouts. While verapamil compares favorably with lithium in practice, some patients require verapamil doses far in excess of those administered for cardiac disorders. The initial dose range is 40–80 mg twice daily; effective doses may be as high as 960 mg/d. Side effects such as constipation, leg swelling, or gingival hyperplasia can be problematic. Of paramount concern, however, is the cardiovascular safety of verapamil, particularly at high doses. Verapamil can cause heart block by slowing conduction in the atrioventricular node, a condition that can be monitored by following the PR interval on a standard electrocardiogram (ECG). Approximately 20% of patients treated with verapamil develop ECG abnormalities, which can be observed with doses as low as 240 mg/d; these abnormalities can worsen over time in patients on stable doses. A baseline ECG is recommended for all patients. The ECG is repeated 10 days after a dose change in patients whose dose is being increased above 240 mg daily. Dose increases are usually made in 80-mg increments. For patients on long-term verapamil, ECG monitoring every 6 months is advised.

NEUROMODULATION THERAPY

When medical therapies fail in chronic cluster headache, neuromodulation strategies can be used. Sphenopalatine ganglion (SPG) stimulation with an implanted battery-free stimulator has been shown in randomized controlled trials to be effective in aborting attacks and reducing their frequency over time. nVNS compares favorably with standard-of-care in open-label experience. Similarly, occipital nerve stimulation has been used open label and appears to be beneficial. Deep-brain stimulation of the region of the posterior hypothalamic gray matter is successful in about 50% of patients treated, although its risk-versus-benefit ratio makes it inappropriate before all other less invasive options have been explored.

PAROXYSMAL HEMICRANIA

Paroxysmal hemicrania (PH) is characterized by frequent unilateral, severe, short-lasting episodes of headache. Like cluster headache, the pain tends to be retroorbital but may be experienced all over the head and is associated with cranial autonomic phenomena such as lacrimation and nasal congestion. Patients with remissions are said to have episodic PH, whereas those with the nonremitting form are said to have chronic PH. The essential features of PH are unilateral very severe pain; short-lasting attacks (2–45 min); very frequent attacks (usually >5 a day); marked autonomic features ipsilateral to the pain; rapid course (<72 h); and excellent response to indomethacin. In contrast to cluster headache, which predominantly affects males, the male-to-female ratio in PH is close to 1:1.

Indomethacin (25–75 mg tid), which can completely suppress attacks of PH, is the treatment of choice. Although therapy may be complicated by indomethacin-induced gastrointestinal side effects, currently there are no consistently effective alternatives. Topiramate is helpful in some cases. nVNS can be very effective in PH. Melatonin may be indomethacin-sparing in some patients. Verapamil, an effective treatment for cluster headache, does not appear to be useful for PH. In occasional patients, PH can coexist with TN (PH-tic syndrome); similar to cluster-tic syndrome, each component may require separate treatment.

Secondary PH has been reported with lesions in the region of the sella turcica, including arteriovenous malformation, cavernous sinus meningioma, pituitary pathology, and epidermoid tumors. Secondary PH is more likely if the patient requires high doses (>200 mg/d) of indomethacin. In patients with apparent bilateral PH, raised cerebrospinal fluid (CSF) pressure should be suspected. It is important to note that indomethacin reduces CSF pressure. When a diagnosis of PH is considered, magnetic resonance imaging (MRI) is indicated to exclude a pituitary lesion.

■ SUNCT/SUNA

SUNCT is a rare primary headache syndrome characterized by severe, unilateral orbital or temporal pain that is stabbing or throbbing in quality. Diagnosis requires at least 20 attacks, lasting for 5–240 s; ipsilateral conjunctival injection and lacrimation should be present. In some patients, conjunctival injection or lacrimation is missing, and the diagnosis of SUNA can be made.

Diagnosis The pain of SUNCT/SUNA is unilateral and may be located anywhere in the head. Three basic patterns can be seen: single stabs, which are usually short-lived; groups of stabs; or a longer attack comprising many stabs between which the pain does not completely resolve, thus giving a "saw-tooth" phenomenon with attacks lasting many minutes. Each pattern may be seen in the context of an underlying continuous head pain. Characteristics that lead to a suspected diagnosis of SUNCT are the cutaneous (or other) triggers of attacks, a lack of refractory period to triggering between attacks, and the lack of a response to indomethacin. Apart from trigeminal sensory disturbance, the neurologic examination is normal in primary SUNCT/SUNA.

The diagnosis of SUNCT/SUNA is often confused with TN, particularly in first-division TN **(Chap. 452)**. Minimal or no cranial autonomic symptoms and a clear refractory period to triggering indicate a diagnosis of TN.

Secondary (Symptomatic) SUNCT SUNCT can be seen with posterior fossa or pituitary lesions. All patients with SUNCT/SUNA should be evaluated with pituitary function tests and a brain MRI with pituitary views.

TREATMENT
SUNCT/SUNA

ABORTIVE THERAPY

Therapy of acute attacks is not a useful concept in SUNCT/SUNA because the attacks are of such short duration. However, IV lidocaine, which arrests the symptoms, can be used in hospitalized patients.

PREVENTIVE THERAPY

Long-term prevention to minimize disability and hospitalization is the goal of treatment. The most effective treatment for prevention is lamotrigine, 200–400 mg/d. Topiramate and gabapentin may also be effective. Carbamazepine, 400–500 mg/d, has been reported by patients to offer modest benefit.

Surgical approaches such as microvascular decompression or destructive trigeminal procedures are seldom useful and often produce long-term complications. Greater occipital nerve injection has produced limited benefit in some patients. nVNS may be useful. Occipital nerve stimulation is probably helpful in a subgroup of these patients. For intractable cases, short-term prevention with IV lidocaine can be effective.

■ HEMICRANIA CONTINUA

The essential features of hemicrania continua are moderate and continuous unilateral pain associated with fluctuations of severe pain; complete resolution of pain with indomethacin; and exacerbations that may be associated with cranial autonomic features, including conjunctival injection, lacrimation, and photophobia on the affected side. The age of onset ranges from 10 to 70 years; women are affected twice as often as men. The cause is unknown.

TREATMENT
Hemicrania Continua

Treatment consists of indomethacin; other NSAIDs appear to be of little or no benefit. The IM injection of 100 mg of indomethacin has been proposed as a diagnostic tool, and administration with a placebo injection in a blinded fashion can be very useful diagnostically. Alternatively, a trial of oral indomethacin, starting with 25 mg tid, then 50 mg tid, and then 75 mg tid, can be given. Up to 2 weeks at the maximal dose may be necessary to assess whether a dose has a useful effect. Topiramate can be helpful in some patients. nVNS can be very useful in these patients. Melatonin can be useful as an indomethacin-sparing agent. Occipital nerve stimulation probably has a role in patients with hemicrania continua who are unable to tolerate indomethacin.

■ OTHER PRIMARY HEADACHE DISORDERS

Primary Cough Headache Primary cough (Valsalva maneuver) headache is a generalized headache that begins suddenly, lasts for seconds or several minutes, sometimes up to a few hours, and is precipitated by coughing; it is preventable by avoiding coughing or other precipitating events, which can include sneezing, straining, laughing, or stooping. In all patients with this syndrome, serious etiologies must be excluded before a diagnosis of "benign" primary cough headache can be established. A Chiari malformation or any lesion causing obstruction of CSF pathways or displacing cerebral structures can be the cause of the head pain. Other conditions that can present with cough or exertional headache as the initial symptom include cerebral aneurysm, carotid stenosis, and vertebrobasilar disease. Benign cough headache can resemble benign exertional headache (below); patients with the former condition are typically older.

TREATMENT
Primary Cough Headache

Indomethacin 25–50 mg two to three times daily is the treatment of choice. Some patients with cough headache obtain complete cessation of their attacks with lumbar puncture; this is a simple option when compared to prolonged use of indomethacin, and it is effective in about one-third of patients. The mechanism of this response is unclear.

Primary Exercise Headache Primary exercise headache has features resembling both cough headache and migraine. It may be precipitated by any form of exercise; it often has the pulsatile quality of migraine. The pain lasts <48 h, is bilateral, and is often throbbing at onset; migrainous features may develop in patients susceptible to migraine. The duration tends to be shorter in adolescents than in older adults. Primary exercise headache can be prevented by avoiding excessive exertion, particularly in hot weather or at high altitude.

The mechanism of primary exercise headache is unclear. Acute venous distension likely explains one syndrome—the acute onset of headache with straining and breath holding, as in weightlifter's headache. Because exercise can trigger headache in a number of serious underlying conditions (Chap. 17), these must be considered in patients with exercise headache. Pain from angina may be referred to the head, probably by central connections of vagal afferents, and may present as exercise headache (cardiac cephalgia). The link to exercise is the main clinical clue that headache is of cardiac origin. Pheochromocytoma may occasionally cause exercise headache. Intracranial lesions and stenosis of the carotid arteries are other possible etiologies.

TREATMENT

Primary Exercise Headache

Exercise regimens should begin modestly and progress gradually to higher levels of intensity. Indomethacin at daily doses from 25–150 mg is generally effective in benign exertional headache. Indomethacin (50 mg), a gepant-rimegepant (75 mg orally) or ubrogepant (100 mg orally), ergotamine (1 mg orally), and dihydroergotamine (2 mg by nasal spray) are useful short-term preventive measures.

Primary Headache Associated with Sexual Activity Three types of sex headache are reported: a dull bilateral ache in the head and neck that intensifies as sexual excitement increases; a sudden, severe, explosive headache occurring at orgasm; and a postural headache developing after coitus. The latter arises from vigorous sexual activity and is a form of low CSF pressure headache and thus not a primary headache disorder (Chap. 17). Headaches developing at the time of orgasm are not always benign; 5–12% of cases of subarachnoid hemorrhage are precipitated by sexual intercourse. Sex headache is reported by men more often than women and may occur at any time during the years of sexual activity. It may appear on several occasions in succession and then not trouble the patient again, even without an obvious change in sexual activity. In patients who stop sexual activity when headache is first noticed, the pain may subside within a period of 5 min to 2 h. In about half of patients, sex headache will subside within 6 months. Most patients with sex headache do not have exercise or cough headache; this clinical paradox is generally a marker of primary sex headache. Migraine is probably more common in patients with sex headache.

TREATMENT

Primary Sex Headache

Benign sex headaches recur irregularly and infrequently. Management can often be limited to reassurance and advice about ceasing sexual activity if a mild, warning headache develops. Propranolol can be used to prevent headache that recurs regularly or frequently, but the dosage required varies from 40–200 mg/d. An alternative is the calcium channel–blocking agent diltiazem, 60 mg tid. Indomethacin (25–50 mg), a gepant-rimegepant (75 mg orally) or ubrogepant (100 mg orally), or frovatriptan (2.5 mg), taken 30–45 min prior to sexual activity can also be helpful.

Primary Thunderclap Headache Sudden onset of severe headache may occur in the absence of any known provocation. The differential diagnosis includes the sentinel bleed of an intracranial aneurysm, reversible cerebral vasoconstriction syndrome (RCVS), cervicocephalic arterial dissection, and cerebral venous thrombosis. Headaches of explosive onset may also be caused by the ingestion of sympathomimetic drugs or of tyramine-containing foods in a patient who is taking MAOIs, or they may be a symptom of pheochromocytoma. Whether thunderclap headache can be the presentation of an unruptured cerebral aneurysm is uncertain; some experts believe that RCVS is the cause of most or all cases of otherwise undiagnosed thunderclap headache. When neuroimaging studies and lumbar puncture exclude subarachnoid hemorrhage, patients with thunderclap headache usually do very well over the long term. In one study of patients whose computed tomography (CT) scans and CSF findings were negative, ~15% had recurrent episodes of thunderclap headache, and nearly half subsequently developed migraine or TTH.

The first presentation of any sudden-onset severe headache should be diligently investigated with neuroimaging (CT or, when possible, MRI with MR angiography) and CSF examination. In the presence of posterior leukoencephalopathy, the differential diagnosis includes cerebral angiitis, posterior reversible encephalopathy syndrome (PRES), drug toxicity (cyclosporine, intrathecal methotrexate/cytarabine, pseudoephedrine, or cocaine), posttransfusion effects, and postpartum angiopathy. Treatment with nimodipine may be helpful, although the vasoconstriction often resolves spontaneously.

Cold-Stimulus Headache This refers to head pain triggered by application or ingestion/inhalation of something cold. It is brought on quickly and typically resolves within 10–30 min of the stimulus being removed. It is best recognized as "brain-freeze" headache or ice-cream headache when due to ingestion. Although cold may be uncomfortable at some level for many people, it is the reliable, severe, and somewhat prolonged nature of these pains that set them apart. The transient receptor potential cation subfamily M member 8 (TRPM8) channel, a known cold-temperature sensor, may be a mediator of this syndrome. Naproxen 500 mg taken 30 min prior to exposure can be helpful for this problem.

External Pressure Headache External pressure from compression or traction on the head can produce a pain that may have some generalized component, although the pain is largely focused around the site of the pressure. It typically resolves within an hour of the stimulus being removed. Examples of stimuli include helmets, swimming goggles, or very long ponytails. Treatment is to recognize the problem and remove the stimulus.

Primary Stabbing Headache The essential features of primary stabbing headache are stabbing pain confined to the head or, rarely, the face, lasting from 1 to many seconds and occurring as a single stab or a series of stabs; absence of associated cranial autonomic features; absence of cutaneous triggering of attacks; and a pattern of recurrence at irregular intervals (hours to days). When present in adolescents, primary stabbing headache may be a presenting and very troublesome problem for the patient. The pains have been variously described as "ice-pick pains" or "jabs and jolts." They are more common in patients with other primary headaches, such as migraine, the TACs, and hemicrania continua. A key clinical feature is an irregular cadence compared to the regular cadence of the throbbing or pounding that characterizes migraine.

TREATMENT

Primary Stabbing Headache

The response of primary stabbing headache to indomethacin (25–50 mg two to three times daily) is usually excellent. As a general rule, the symptoms wax and wane, and after a period of control on indomethacin, it is appropriate to withdraw treatment and observe the outcome.

Nummular Headache Nummular headache is felt as a round or elliptical discomfort that is fixed in place, ranges in size from 1–6 cm, and may be continuous or intermittent. Uncommonly, it may be multifocal. It may be episodic but is more often continuous during exacerbations. Accompanying the pain there may be a local sensory disturbance, such as allodynia or hypesthesia. Local dermatologic or bony lesions need to be excluded by examination and investigation. This condition can be difficult to treat when present in isolation;

tricyclics, such as amitriptyline, or anticonvulsants, such as topiramate or valproate, are most often tried. This phenotype can be seen in combination with migraine and the TACs, in which cases treatment of the primary headache disorder is often effective for the nummular headache as well.

Hypnic Headache This headache syndrome typically begins a few hours after sleep onset. The headaches last from 15–30 min and are typically moderately severe and generalized, although they may be unilateral and can be throbbing. Patients may report falling back to sleep only to be awakened by a further attack a few hours later; up to three repetitions of this pattern occur through the night. Daytime naps can also precipitate head pain. Most patients are female, and the onset is usually after age 60 years. Headaches are typically bilateral but may be unilateral. Photophobia, phonophobia, and nausea are usually absent. The major secondary consideration in this headache type is poorly controlled hypertension; 24-h blood pressure monitoring is recommended to detect this treatable condition.

TREATMENT

Hypnic Headache

Patients with hypnic headache generally respond to a bedtime dose of lithium carbonate (200–600 mg). One to two cups of coffee, or caffeine 60 mg orally, at bedtime may be effective in approximately one-third of patients. Reports suggest that verapamil, 160 mg; flunarizine, 5 mg nightly; or indomethacin, 25–75 mg nightly, can be effective.

New Daily Persistent Headache Primary new daily persistent headache (NDPH) occurs in both men and women. It can be of the migrainous type, with features of migraine, or it can be featureless, appearing as new-onset TTH. Those with migrainous features are the most common form and include unilateral headache and throbbing pain; each feature is present in about one-third of patients. Nausea, photophobia, and/or phonophobia occur in about half of patients. Some patients have a previous history of migraine. NDPH may be more common in adolescents. Treatment of migrainous-type primary NDPH consists of using the preventive therapies effective in migraine (see above). Featureless NDPH is one of the primary headache disorders most refractory to treatment. Standard preventive therapies can be offered but are often ineffective. The secondary NDPHs are discussed elsewhere **(Chap. 17)**.

FURTHER READING

Buse DC et al: Demographics, headache features, and comorbidity profiles in relation to headache frequency in people with migraine: Results of the American Migraine Prevalence and Prevention (AMPP) Study. Headache 60:2340, 2020.

Charles A, Pozo-Rosich P: Targeting calcitonin gene-related peptide: A new era in migraine therapy. Lancet 394:1765, 2019.

Cittadini E, Goadsby PJ: Hemicrania continua: A clinical study of 39 patients with diagnostic implications. Brain 133:1973, 2010.

Cittadini E et al: Paroxysmal hemicrania: A prospective clinical study of thirty-one cases. Brain 131:1142, 2008.

de Boer I et al: Advance in genetics of migraine. Curr Opin Neurol 32:413, 2019.

Ferrari MD et al: Migraine. Nat Prim 8:2, 2022.

Goadsby PJ et al: Pathophysiology of migraine: A disorder of sensory processing. Physiol Rev 97:553, 2017.

Schankin CJ et al: "Visual snow": A disorder distinct from persistent migraine aura. Brain 137:1419, 2014.

Wei DY, Goadsby PJ: Cluster headache pathophysiology: Insights from current and emerging treatments. Nat Rev Neurol 17:308, 2021.

Weng H et al: Phenotypic and treatment outcome data on SUNCT and SUNA, including a randomised placebo-controlled trial. Cephalalgia 38:1554, 2018.

442

Alzheimer's Disease

Gil D. Rabinovici, Peter A. Ljubenkov,
William W. Seeley, Bruce L. Miller

ALZHEIMER'S DISEASE

Approximately 55 million people across the world are living with dementia. Alzheimer's disease (AD) is the most common cause of dementia, contributing to an estimated 60–70% of all cases. Total U.S. health care costs related to dementia care are estimated at $360 billion in 2024, translating into an average of ~$25,000 per patient. Furthermore, the emotional toll for family members and caregivers is immeasurable. AD can manifest as early as the third decade of life, but it is the most common neuropathology contributing to dementia in the elderly. Patients most often present with an insidious loss of episodic memory followed by a slowly progressive dementia. In typical amnestic AD, brain atrophy begins in the medial temporal lobes before spreading to inferior temporal, lateral and medial parietal, and dorsolateral frontal cortices. Microscopically, there are widespread neuritic plaques containing amyloid beta (Aβ), neurofibrillary tangles (NFTs) composed of hyperphosphorylated tau filaments, and Aβ accumulation in blood vessel walls in cortex and leptomeninges (see "Pathology," below). The identification of causative mutations and susceptibility genes for AD has provided a foundation for rapid progress in understanding the biological basis of the disorder. The major genetic risk factor for AD is the ε4 allele of the apolipoprotein E (*ApoE*) gene. Carrying one ε4 allele increases the risk for AD by two- to threefold in women, whereas carrying two alleles increases the risk 10- to 15-fold in both sexes. Rapid progress in the development of imaging, cerebrospinal fluid (CSF), and plasma biomarkers of Aβ and phosphorylated tau has enabled detection of AD pathologic hallmarks in living people, opening the door to early detection and intervention with biologically specific therapies.

CLINICAL MANIFESTATIONS

The cognitive changes of AD tend to follow a characteristic pattern, beginning with memory impairment and progressing to deficits in executive, language, and visuospatial functions. Yet, ~20% of patients with AD present with nonmemory complaints such as word-finding, organizational, or navigational difficulty. In other patients, visual processing dysfunction (referred to as posterior cortical atrophy syndrome) or a progressive "logopenic" aphasia characterized by difficulties with naming and repetition is the primary manifestation of AD for years before progressing to involve memory and other cognitive domains. Still other patients may present with an asymmetric akinetic-rigid-dystonic ("corticobasal") syndrome or a dysexecutive/behavioral (i.e., "frontal" variant of AD). Depression, social withdrawal, and anxiety occur in early disease stages and may represent a prodrome before cognitive symptoms are apparent.

In early stages of typical amnestic AD, the memory loss may go unrecognized or be ascribed to benign forgetfulness of aging. The term *subjective cognitive decline* refers to self-perceived worsening in memory or other cognitive abilities that may not be noticeable to others or apparent on formal neuropsychological testing. Once the memory loss becomes noticeable to the patient and spouse and is confirmed on standardized memory tests, the term *mild cognitive impairment* (MCI) is often used. This construct provides useful prognostic information because ~50% of patients with MCI (roughly 12% per year) will progress to the dementia stage over 4 years. Increasingly, the MCI construct is being replaced by the notion of "early symptomatic AD" to signify that AD is considered the underlying disease (based on clinical or biomarker evidence) in a patient who remains functionally compensated. Even earlier in the course, "preclinical AD" refers to a person with biomarker evidence of amyloid pathology (with or without tau pathology) in the absence of symptoms. It is estimated that preclinical biomarker changes may precede clinical symptoms by 20 years or more, creating a window of opportunity for early-stage treatment and prevention trials. Emerging evidence suggests that partial and sometimes generalized seizures herald AD and can occur even prior to dementia onset, especially in younger patients and those with autosomal dominant AD-causing mutations.

Eventually with AD, the cognitive problems begin to interfere with daily activities, such as keeping track of finances, following instructions on the job, driving, shopping, and housekeeping. Some patients are unaware of these difficulties (*anosognosia*), but most remain acutely attuned to their deficits in early disease stages. Changes in environment (travel, relocation, hospitalization) tend to destabilize the patient. Over time, patients become lost on walks or while driving. Social graces, routine behavior, and superficial conversation may be surprisingly intact, even into the later stages of the illness.

In the middle stages of AD, the patient is unable to work, is easily lost and confused, and requires daily supervision. Language becomes impaired—first naming, then comprehension, and finally fluency. Word-finding difficulties and circumlocution can be evident in the early stages, even when formal testing demonstrates intact naming and fluency. *Apraxia* emerges, manifesting as trouble performing learned sequential motor tasks such as using utensils or appliances. Visuospatial deficits begin to interfere with dressing, eating, or even walking, and patients fail to solve simple puzzles or copy geometric figures. Simple calculations and clock reading become difficult in parallel.

In the late stages, some persons remain ambulatory, wandering aimlessly. Loss of judgment and reasoning is inevitable. Delusions are prevalent and usually simple, with common themes of theft, infidelity, or misidentification. Disinhibition and uncharacteristic belligerence may occur and alternate with passivity and withdrawal. Sleep-wake patterns are disrupted, and nighttime wandering becomes disturbing to the household. Some patients develop a shuffling gait with generalized muscle rigidity associated with slowness and awkwardness of movement. Patients often look parkinsonian **(Chap. 446)** but rarely have a high-amplitude, low-frequency tremor at rest. There is a strong overlap between dementia with Lewy bodies (DLB) **(Chap. 445)** and AD, and some AD patients develop more classical parkinsonian features.

In the end stages, patients with AD become rigid, mute, incontinent, and bedridden, and need help with eating, dressing, and toileting. Hyperactive tendon reflexes and myoclonic jerks (sudden brief contractions of various muscles or the whole body) may occur spontaneously or in response to physical or auditory stimulation. Often death results from malnutrition, secondary infections, pulmonary emboli, heart disease, or, most commonly, aspiration. The typical duration of symptomatic AD is 8–10 years, but the course ranges from 1 to 25 years. For unknown reasons, some patients with AD show a steady decline in function while others have prolonged plateaus without major deterioration.

An increased risk for seizures is also increasingly recognized as a feature of AD; different seizure types (both generalized and focal onset) have been reported, and these occur more frequently as the disease progresses.

DIAGNOSIS

See also "Other Causes of Dementia," below, and the general discussion of dementia presented in Chap. 31. Early in the disease course, other etiologies of dementia should be excluded **(see Tables 31-1, 31-3, and 31-4)**. *Slowly progressive decline in memory and orientation, normal results on systemic laboratory tests, and a magnetic resonance imaging (MRI) or computed tomography (CT) scan showing only distributed or posteriorly predominant cortical and hippocampal atrophy (see below) are suggestive of AD.* A clinical diagnosis of AD reached after careful evaluation is confirmed at autopsy ~70–80% of the time, with misdiagnosed cases usually resulting from limbic-predominant age-related TDP-43 encephalopathy (LATE) with or without hippocampal sclerosis, primary age-related tauopathy (PART), Lewy body disease (LBD), vascular pathology, or frontotemporal lobar degeneration (FTLD; **Chap. 443**).

Simple clinical clues are useful in the differential diagnosis. Early prominent gait disturbance with only mild memory loss suggests vascular dementia or, rarely, normal pressure hydrocephalus (NPH), discussed later. Resting tremor with stooped posture, bradykinesia, and masked facies suggest PD **(Chap. 446)** or DLB **(Chap. 445)**. When dementia occurs after a well-established diagnosis of PD, PD dementia (PDD) is usually the correct diagnosis, but many patients with this diagnosis will show a mixture of AD and LBD at autopsy. The early appearance of parkinsonian features in association with fluctuating alertness, visual hallucinations, or delusional misidentification suggests DLB. Chronic alcoholism should prompt the search for vitamin deficiency. Loss of joint position and vibration sensibility accompanied by Babinski signs suggests vitamin B_{12} deficiency, especially in a patient with a history of autoimmune disease, small bowel resection or irradiation, or veganism **(Chap. 104)**. Early onset of a focal seizure suggests a metastatic or primary brain neoplasm **(Chap. 95)**. Previous or ongoing depression raises suspicion for depression-related cognitive impairment, although significant cognitive changes with depression is common and AD and DLB can feature a prodrome of depression or anxiety. A history of treatment for insomnia, anxiety, psychiatric disturbance, or epilepsy suggests chronic drug intoxication. Rapid progression over a few weeks or months associated with rigidity and myoclonus suggests Creutzfeldt-Jakob disease (CJD) **(Chap. 449)**. Prominent behavioral changes with intact navigation and focal anterior-predominant atrophy on brain imaging are typical of FTD. A positive family history of dementia suggests either one of the familial forms of AD or one of the other genetic disorders associated with dementia, such as FTD **(Chap. 443)**, Huntington's disease (HD) **(Chap. 447)**, prion disease **(Chap. 449)**, or rare hereditary ataxias **(Chap. 450)**.

Electroencephalogram (EEG) is usually normal or shows nonspecific slowing; prolonged EEG can be used to seek out intermittent nonconvulsive seizures.

Structural neuroimaging studies (CT and MRI) do not show a single specific pattern with AD and may be normal early in the disease. As AD progresses, more distributed but usually posterior-predominant cortical atrophy becomes apparent, along with atrophy of the medial temporal memory structures **(see Fig. 31-1)**. The main purpose of structural imaging is to exclude other disorders, such as primary and secondary neoplasms, vascular dementia, diffuse white matter disease, and normal-pressure hydrocephalus (NPH). Imaging also helps to distinguish AD from other degenerative disorders, such as frontotemporal dementia (FTD) **(Chap. 443)** or the prion disorder CJD **(Chap. 449)**, which feature imaging patterns that are different from AD. Functional imaging studies, such as fluorodeoxyglucose (FDG) positron emission tomography (PET), reveal hypometabolism in the posterior temporal-parietal cortex in AD **(see Fig. 31-1)**.

Amyloid PET imaging (e.g., with radiotracers [^{11}C]PIB, [^{18}F]florbetapir, [^{18}F]florbetaben, or [^{18}F]flutemetamol) confirms the presence of neuritic and diffuse Aβ plaques throughout the neocortex **(Fig. 442-1)**. [^{18}F]florbetapir, [^{18}F]florbetaben, and [^{18}F]flutemetamol are approved for clinical use in the United States and other countries. Although amyloid PET binding is detected in AD, ~25% of cognitively unimpaired older individuals also have positive scans, thought to represent preclinical disease and an increase in the risk of converting to clinical AD. Similarly, dementia due to a non-AD disorder can be the underlying etiology in a patient who tests positively on amyloid PET due to comorbid AD pathology. Amyloid PET ligands also bind to vascular Aβ deposits in cerebral amyloid angiopathy (CAA) **(Chap. 439)**. Therefore, clinical use of amyloid PET should be restricted to specific scenarios in which knowledge of amyloid status is expected to impact diagnosis and change management. For example, a negative amyloid PET scan in a patient with dementia makes an AD diagnosis unlikely. Conversely, a positive PET scan can be used to establish eligibility for novel amyloid-lowering therapies in a patient who meets the clinical criteria for treatment (see below).

Tau PET radiotracers (e.g., [^{18}F]flortaucipir, [^{18}F]MK-6240, [^{18}F]PI-2620) bind to the paired helical filaments that form neurofibrillary tangles and are primarily available in the research setting. [^{18}F]Flortaucipir is also approved for clinical use in the United States "to estimate the density and distribution of aggregated tau neurofibrillary tangles in adult patients with cognitive impairment who are being evaluated for AD." The pattern of binding is largely consistent with

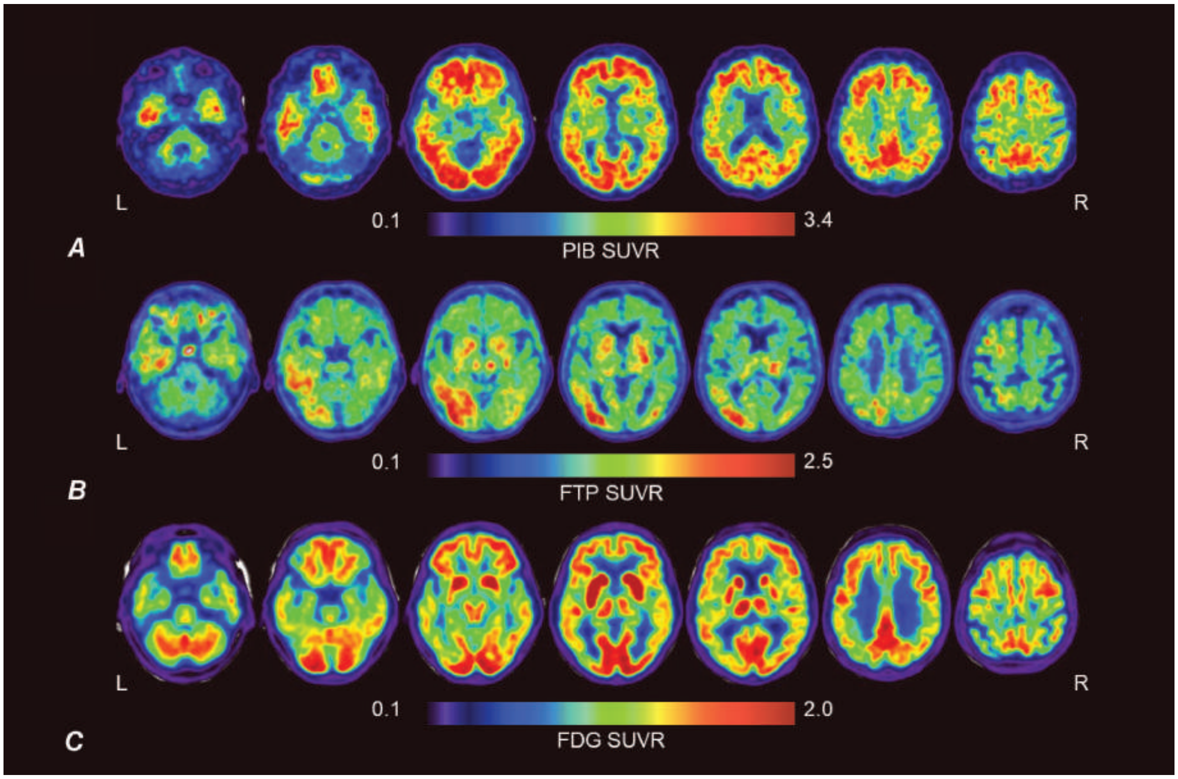

FIGURE 442-1 Molecular imaging of Alzheimer's disease pathophysiology in an 81-year-old with mild Alzheimer's disease. ***A.*** Aβ positron emission tomography (PET) with [^{11}C]PIB reveals extensive radiotracer retention in neocortex, consistent with the known distribution of amyloid plaques. ***B.*** Tau PET with [^{18}F]FTP shows asymmetric uptake predominantly in the left temporal cortex, consistent with intermediate-stage neurofibrillary tangles. Tracer uptake in midbrain and basal ganglia represents "off-target" (non-tau-related) tracer retention. ***C.*** Fluorodeoxyglucose (FDG)-PET reveals reduced tracer uptake in left greater than right temporal and parietal cortex, indicative of decreased synaptic activity. The pattern of hypometabolism corresponds more closely to the pattern of tau than amyloid deposition. ***A–C.*** Axial brain slices are shown in neurologic orientation. L, left; R, right; SUVR, standardized uptake value ratio, a quantitative measure of PET radiotracer retention.

Braak neuropathologic staging of neurofibrillary tangles, with early retention in medial temporal regions, followed by spread into temporoparietal and cingulate cortices, dorsolateral prefrontal regions, and ultimately, primary sensory and motor areas. However, tau PET signal lags behind neuropathologic staging of tangles due to limited sensitivity, and completely negative tau PET does not rule out early neurofibrillary pathology (Braak stages I–III). Notably, tau PET radiotracers developed to detect AD-related tau aggregation have limited utility in detecting aggregated tau in non-AD tauopathies (e.g., progressive supranuclear palsy [PSP], cortical basal degeneration [CBD], chronic traumatic encephalopathy [CTE]).

Routine spinal fluid examination is generally normal, but CSF reductions in Aβ_{42} levels and the Aβ_{42}/Aβ_{40} ratio correlate with amyloid deposition, and increases in phosphorylated tau (at residue 181 or 217; p-Tau181 or p-Tau217) detect AD-related changes in tau phosphorylation and secretion. On the other hand, increases in total tau levels represent a nonspecific finding seen in AD but also in other causes of neurodegeneration. Several CSF AD biomarker assays are now approved for clinical use in the United States and other countries. Decreases in the Aβ_{42}/Aβ_{40} ratio or increases in the p-Tau181/Aβ_{42} ratio show higher agreement with amyloid PET results or AD neuropathology than any single CSF analyte. Increase in the microtubule-binding region (MTBR) of tau containing the residue 243 (MTBR-tau243) is an experimental biomarker (not yet available in clinical practice) that better correlates with neurofibrillary tangles and tau PET binding.

In recent years, there has been significant progress in the development of plasma measurements of Aβ and phosphorylated tau with ultra-sensitive immunoassays or mass spectrometry. These blood-based AD biomarkers are entering the clinical arena and will undoubtedly improve access, scalability, and cost-effectiveness of AD biomarker testing in diverse practice settings. Similar to CSF, decreases in the plasma Aβ_{42}/Aβ_{40} ratio and increases in plasma p-Tau181, p-Tau217, or p-Tau231 show high concordance with positive amyloid PET scans and AD neuropathology at autopsy. In conjunction with a clinical evaluation, plasma assays may be sufficient to establish the diagnosis of AD for certain use cases, with the highest-performing assays showing comparable diagnostic performance to CSF biomarkers in detecting AD neuropathology. It is important to note that concentrations of Aβ and p-Tau in blood can be impacted by medical comorbidities (e.g., reduced creatinine clearance, elevated body mass index), which may lead to false-positive or false-negative results, particularly if values are near the test's threshold.

While positive AD imaging or fluid biomarkers are highly predictive of underlying AD neuropathology, the question of whether AD represents a primary or contributing cause to an individual patient's clinical presentation is not always straightforward. For example, in a patient who presents at a relatively young age with classical AD symptoms, compatible MRI changes, and positive CSF or amyloid PET, it is highly likely that AD neuropathology is the primary cause of cognitive decline. In an elderly patient with the same clinical profile, however, AD is still likely to be contributing to impairment, although a contribution of other non-AD co-pathologies is also likely. In the case of a patient who presents with classical FTD and positive AD biomarkers, FTLD pathology may be the primary cause of impairment, with the positive AD biomarkers indicating incidental preclinical AD. Conversely, a patient who presents with a clinical syndrome suggesting AD but with *negative* AD biomarkers is likely to have a non-AD neuropathology as the primary cause of impairment. In older patients, age-related conditions that selectively target the medial temporal lobes, such as LATE or PART, can mimic AD clinically, and these should be high on the list of diagnostic possibilities in biomarker-negative individuals presenting with a progressive amnestic dementia.

The proliferation of AD biomarkers has led to a movement to redefine AD on purely biological grounds, based solely on positivity of AD biomarkers and independent of the clinical syndrome. In the Revised Criteria for Diagnosis and Staging of AD proposed by the Alzheimer's Association, amyloid PET, CSF/plasma Aβ, and p-Tau measures are considered "core 1 biomarkers" that are necessary and sufficient for the diagnosis of AD, irrespective of clinical symptoms. Tau PET is designated a "core 2" biomarker to be used for biomarker-based disease staging. Clinical staging is performed independent of biomarkers based on the level of functional impairment, spanning the range from asymptomatic "preclinical" AD to severe dementia.

EPIDEMIOLOGY

The most important risk factors for AD are increasing age and a positive family history. In the United States, ~10% of people over age 65 have AD, including 3% of people age 65–74, 17% of people age 75–84, and 32% of people age 85 and older. A positive family history of dementia suggests a genetic contribution to AD, which is usually attributable to the apolipoprotein E (ApoE) ε4 risk allele. Autosomal dominant inheritance occurs in only 1–2% of patients and is typically accompanied by a multigenerational history of early-onset dementia. Female sex is a risk factor independent of the greater longevity of females, and women who carry a single Apo ε4 allele are more susceptible than are male ε4 carriers. A history of mild-to-severe traumatic brain injury increases the risk for AD. AD is more common in groups with low educational attainment, but education influences test-taking ability, and it is clear that AD can affect persons of all intellectual levels. One study found that the capacity to express complex written language in early adulthood correlated with a decreased risk for AD. Similarly, illiteracy and low educational attainment are risk factors for dementia. Numerous environmental factors, including aluminum, mercury, and viruses, have been proposed as causes of AD, but rigorous studies have failed to demonstrate a significant role for any of these exposures. Similarly, several studies suggest that the use of nonsteroidal anti-inflammatory agents is associated with a decreased risk of AD, but this risk has not been confirmed in large prospective studies. Vascular disease, and stroke in particular, seems to lower the threshold for the clinical expression of AD. Also, in many patients with AD, amyloid angiopathy can lead to microhemorrhages, large lobar hemorrhages, ischemic infarctions most often in the subcortical white matter, or in rare cases an inflammatory leukoencephalopathy. Diabetes increases the risk of AD threefold. Elevated homocysteine and cholesterol levels; hypertension; obesity; hearing loss; tobacco use; diminished serum levels of folic acid; low dietary intake of fruits, vegetables, and red wine; sleep disorders; low levels of exercise; and air pollution exposure are all being explored as potential risk factors for dementia in general and AD in particular.

PATHOLOGY

At autopsy, the earliest and most severe degeneration is usually found in the medial temporal lobe (entorhinal/perirhinal cortex and hippocampus), inferolateral temporal cortex, and nucleus basalis of Meynert. The characteristic microscopic findings are neuritic plaques and NFTs **(Fig. 442-2)**. These lesions accumulate in small numbers during normal brain aging but dominate the picture in AD. The overall burden of AD neuropathologic changes can be graded based on the topography of Aβ plaques, the density of neuritic plaques, and the spatial extent of NFTs present. Increasing evidence suggests that soluble amyloid species called *oligomers* may cause cellular dysfunction and represent the early toxic molecule in AD. Eventually, further amyloid polymerization and fibril formation lead to neuritic plaques, which contain a central core of amyloid, proteoglycans, ApoE, α-antichymotrypsin, and other proteins. Aβ is a protein of 39–42 amino acids that is derived proteolytically from a larger transmembrane protein, *amyloid precursor protein* (APP), when APP is cleaved by β and γ secretases **(Fig. 442-3)**. The normal function of the Aβ peptides is uncertain. APP has neurotrophic and neuroprotective properties. The plaque core is surrounded by a halo, which contains dystrophic, tau-immunoreactive neurites and activated microglia. The accumulation of Aβ in cerebral arterioles is termed *amyloid angiopathy*. NFTs are composed of silver-staining neuronal cytoplasmic fibrils composed of abnormally phosphorylated tau protein; they appear as paired helical filaments by electron microscopy. Tau binds to and stabilizes microtubules, supporting axonal transport of organelles, glycoproteins, neurotransmitters, and other important cargoes throughout the neuron. Once hyperphosphorylated, tau can no longer bind properly to microtubules and redistributes from the axon to throughout the neuronal cytoplasm and distal dendrites, compromising function. Other theories emphasize that abnormal conformations of tau induce misfolding of native (unfolded) tau into pathologic conformations and that this prion-like templating process is responsible for tau spreading **(Chap. 435)**. Finally, patients with AD often show comorbid LBD, TDP-43, or vascular pathology. Most prevailing rodent models of AD involve expression of mutant transgenes that leads to $A\beta_{42}$ accumulation in the absence of tauopathy. Even in these models, diminishing neuronal tau ameliorates cognitive deficits and nonconvulsive seizures while $A\beta_{42}$ continues to accumulate,

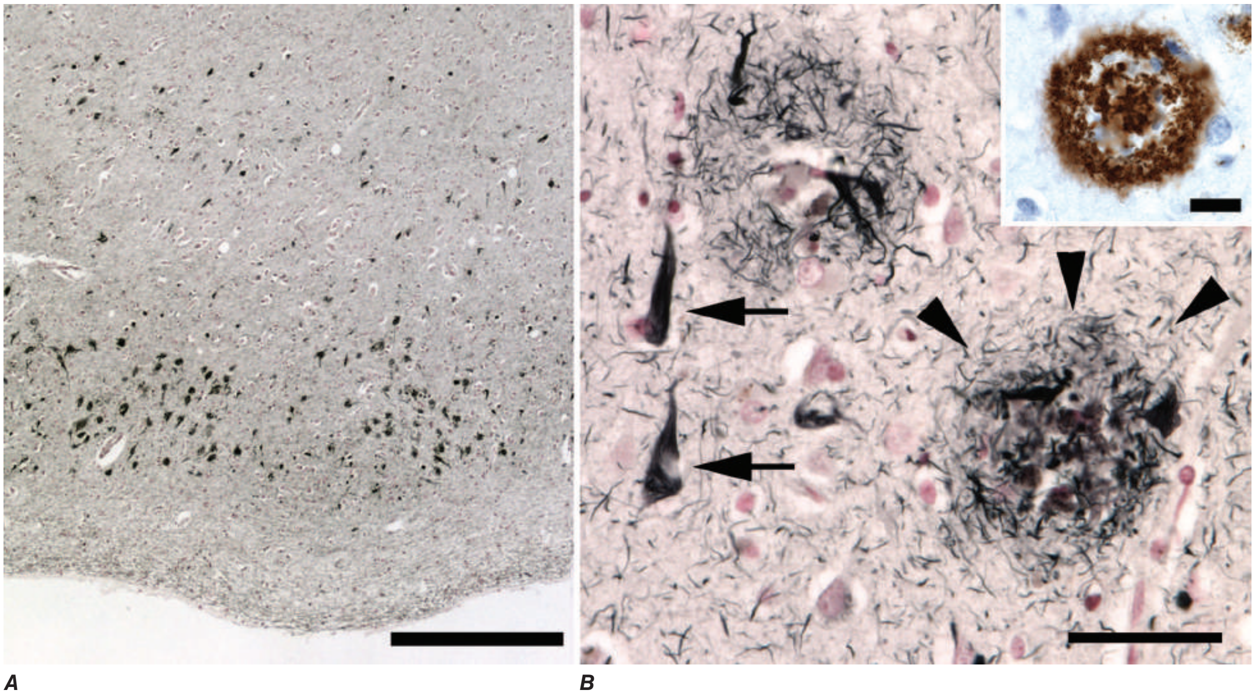

FIGURE 442-2 Neuropathology of Alzheimer's disease. ***A.*** Early neurofibrillary degeneration, consisting of neurofibrillary tangles and neuropil threads, preferentially affects the medial temporal lobes, especially the stellate pyramidal neurons that compose the layer 2 islands of entorhinal cortex, as shown using Gallyas silver staining. ***B.*** Higher magnification view reveals the fibrillar nature of tangles (*arrows*) and the complex structure of neuritic plaques (*arrowheads*), whose major component is Aβ (*inset* shows immunohistochemistry for Aβ). Scale bars are 500 μM in ***A***, 50 μM in ***B***, and 20 μM in ***B*** inset.

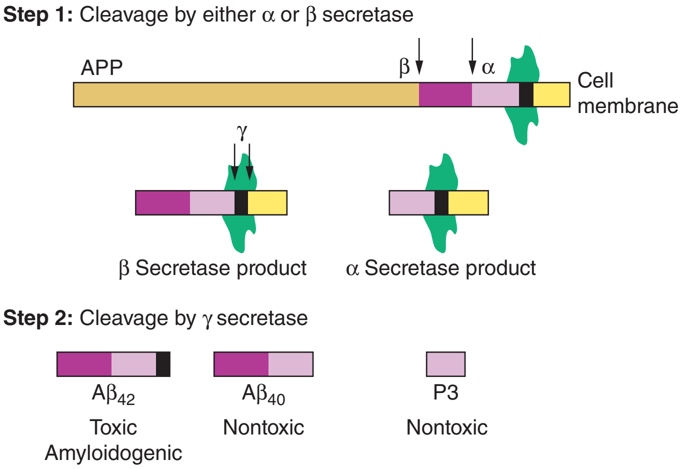

FIGURE 442-3 Amyloid precursor protein (APP) is catabolized by α, β, and γ secretases. A key initial step is the digestion by either β secretase (BASE) or α secretase (ADAM10 or ADAM17 [TACE]), producing smaller nontoxic products. Cleavage of the β secretase product by γ secretase (Step 2) results in either the toxic $A\beta_{42}$ or the nontoxic $A\beta_{40}$ peptide; cleavage of the α secretase product by γ secretase produces the nontoxic P3 peptide. Excess production of $A\beta_{42}$ is a key initiator of cellular damage in Alzheimer's disease (AD). Therapeutics for AD have focused on attempts to reduce accumulation of $A\beta_{42}$ by antagonizing β or γ secretases, promoting α secretase, or clearing $A\beta_{42}$ that has already formed by use of specific antibodies.

raising hope for tau-lowering therapies in humans. Biochemically, AD is associated with a decrease in the cortical levels of several proteins and neurotransmitters, especially acetylcholine, its synthetic enzyme choline acetyltransferase, and nicotinic cholinergic receptors. Reduction of acetylcholine reflects degeneration of cholinergic neurons in the nucleus basalis of Meynert, located just below the thalamus and adjacent to the third ventricle, that project throughout the cortex. There is also noradrenergic and serotonergic depletion due to degeneration of upper brainstem nuclei such as the locus coeruleus (norepinephrine) and dorsal raphe (serotonin), where tau-immunoreactive neuronal cytoplasmic inclusions can be identified in early adult life, even in individuals lacking entorhinal cortex NFTs.

GENETIC CONSIDERATIONS

Several genes play an important role in the pathogenesis of AD. One is the *APP* gene on chromosome 21. Adults with trisomy 21 (Down's syndrome) consistently develop the typical neuropathologic hallmarks of AD if they survive beyond age 40 years, and many develop a progressive dementia superimposed on their baseline deficits. The extra dose of the *APP* gene on chromosome 21 is the initiating cause of AD in adult Down's syndrome and results in excess cerebral amyloid production. Supporting this hypothesis, some families with early-age-of-onset familial AD (FAD) have point mutations in *APP*. Although very rare, these families were the first examples of single-gene autosomal dominant transmission of AD.

Investigation of large families with multigenerational FAD led to the discovery of two additional AD-causing genes, the *presenilins*. Presenilin-1 (*PSEN-1*) is on chromosome 14 and encodes presenilin-1 protein (also known as S182). Mutations in this gene cause an early-age-of-onset AD, with onset typically before age 60 and often before age 50, transmitted in an autosomal dominant, highly penetrant fashion. More than 100 different mutations have been found in the *PSEN-1* gene in families from a wide range of ethnic backgrounds. Presenilin-2 (*PSEN-2*) is on chromosome 1 and encodes the presenilin-2 protein (also known as STM2). A mutation in the *PSEN-2* gene was first found in a group of American families with Volga German ethnic background. Mutations in *PSEN-1* are much more common than those in *PSEN-2*. The presenilins are highly homologous and encode similar proteins that at first appeared to have seven transmembrane domains (hence the designation *STM*), but subsequent studies have suggested eight such domains, with a ninth submembrane region. Both presenilins are cytoplasmic neuronal proteins that are widely expressed throughout the nervous system. They are homologous to a cell-trafficking protein, sel 12, found in the nematode *Caenorhabditis elegans*. Prior to symptom onset, patients with mutations in the presenilin genes have elevated CSF levels of $A\beta_{42}$, and in cell culture, *PSEN-1* mutations produce increased $A\beta_{42}$. *PSEN-1* is involved in the cleavage of APP at the γ secretase site, and mutations in either gene (*PSEN-1* or *APP*) may disturb γ secretase cleavage. Mutations in *PSEN-1* are the most common cause of early-age-of-onset FAD, representing 40–70% of all cases. Mutations in *PSEN-1* tend to produce AD with an earlier age of onset (mean onset 45 years) and a shorter, more rapidly progressive course (mean duration 6–7 years) than mutations in *PSEN-2* (mean onset 53 years; duration 11 years). Although some carriers of *PSEN-2* mutations have had onset of dementia after the age of 70, mutations in the presenilins rarely lead to late-age-of-onset AD. Clinical genetic testing for these uncommon mutations is available but likely to be revealing only in early-age-of-onset FAD and should be performed in association with formal genetic counseling.

The *Apoε* gene on chromosome 19 is involved in the pathogenesis of AD. The protein product, ApoE, participates in cholesterol transport **(Chap. 419)**, and the gene has three alleles: ε2, ε3, and ε4. The Apo ε4 allele confers increased risk of AD in the general population, including sporadic and late-age-of-onset familial forms. Approximately 24–30% of the nondemented white population has at least one ε4 allele (12–15% allele frequency), and ~2% are ε4/ε4 homozygotes. Among patients with AD, 40–65% have at least one ε4 allele, a highly significant elevation compared with controls. The increased risk of AD associated with a single ε4 allele is approximately three times higher than in ε4 noncarriers (and higher in female than male heterozygotes), while the risk in ε4 homozygotes is 10–15 times higher than in ε4 noncarriers. The risk of AD in Apo ε4 carriers also varies by racial and ethnic background, with increased risk in East Asians and decreased risk in African Americans and Hispanics compared to whites. Additionally, many patients with AD have no ε4 allele, and ε4 carriers may never develop AD. Therefore, ε4 is neither necessary nor sufficient to cause AD. Nevertheless, the Apo ε4 allele represents the most important genetic risk factor for sporadic AD and acts as a dose-dependent disease modifier, with each Apo ε4 allele associated with an approximately 10-year earlier age of onset. The association between Apo ε4 and AD is strongest in patients between the ages of 60 and 85 and is weaker in younger patients and in the very old. The precise mechanisms through which Apo ε4 confers AD risk or hastens onset remain unclear, but ε4 leads to less efficient amyloid clearance and production of toxic fragments from cleavage of the molecule. Apo ε can be identified in neuritic plaques and may also be involved in neurofibrillary tangle formation, because it binds to tau protein. Interestingly, carriers of a point mutation in Apo ε3, termed the *Christchurch* mutation, may be protective against tau aggregation and dementia despite having amyloid plaques. Apo ε4 decreases neurite outgrowth in dorsal root ganglion neuronal cultures, perhaps indicating a deleterious role in the brain's response to injury. Increasing evidence suggests that the ε2 allele may reduce AD risk. The ε4 allele is also associated with increased risk for CAA, DLB, and vascular dementia, while its association with FTD is uncertain. Some evidence suggests that ε4 may worsen the expression of non-AD neurodegenerative disorders, as well as following head trauma and other brain injuries. Use of Apo ε4 testing in AD diagnosis remains controversial because its predictive value remains unclear and many individuals with the ε4 allele will never develop dementia. However, clinical Apo ε4 testing is recommended for patients who are being evaluated for treatment with anti-Aβ monoclonal antibodies in order to determine their risk of adverse effects (see below). ApoE genotyping is available in some straight-to-consumer genetic testing platforms.

Additional genes are also likely to be involved in AD, especially as minor risk alleles for sporadic forms of the disease. Genome-wide association studies have identified >40 additional common genetic variants that, individually, have small (i.e., odds ratios ~1.1–1.2 or 0.8–0.9) impact on the risk of AD. Implicated genes converge in biological pathways related to innate immunity, lipid metabolism, and synaptic function. Examples include the clusterin (*CLU*), phosphatidylinositol-binding clathrin assembly protein (*PICALM*), and complement component

(3b/4b) receptor 1 (*CR1*) genes, among others. *CLU* may play a role in synapse turnover, *PICALM* participates in clathrin-mediated endocytosis, and *CR1* may be involved in amyloid clearance or synapse loss through the complement pathway. *TREM2* is a gene involved with inflammation that increases the likelihood of dementia. Homozygous mutation carriers develop a frontal dementia with bone cysts (Nasu-Hakola disease), whereas heterozygotes are predisposed to the development of AD. *TREM2* risk alleles are rare but have strong effects, with odds ratios estimated at 3–4 for developing clinical AD. Polygenic hazard scores that integrate the presence of multiple risk and protective alleles may be useful in predicting an individual's lifetime risk of developing AD. The vast majority of AD genetic studies have focused on white populations of European descent, and much less is known about the genetics of AD in nonwhite populations.

TREATMENT

Alzheimer's Disease

The management of AD requires a multidomain approach that includes neurotransmitter-based therapies to manage symptoms, emerging molecular therapies that target AD pathophysiology with the goal of slowing disease progression and clinical decline, and ongoing patient and caregiver education.

NEUROTRANSMITTER-BASED THERAPIES

Donepezil (target dose, 10 mg daily), rivastigmine (target dose, 6 mg twice daily or 9.5-mg patch daily), galantamine (target dose, 24 mg daily, extended-release), and memantine (target dose, 10 mg twice daily) are approved by the U.S. Food and Drug Administration (FDA) for the treatment of AD. Dose escalations for each of these medications must be carried out over 4–6 weeks to minimize side effects. The pharmacologic action of donepezil, rivastigmine, and galantamine is inhibition of the cholinesterases, primarily acetylcholinesterase, with a resulting increase in cerebral acetylcholine levels. Memantine appears to act by blocking overexcited *N*-methyl-D-aspartate (NMDA) glutamate receptors. Double-blind, placebo-controlled, crossover studies with cholinesterase inhibitors (in mild-to-severe AD dementia) and memantine (in moderate-to-severe AD dementia) have shown them to be associated with modestly improved caregiver ratings of patients' functioning and with an apparent decreased rate of decline in cognitive test scores over periods of up to 3 years. The average patient on an anticholinesterase inhibitor maintains their mini-mental state examination (MMSE) score for close to a year, whereas a placebo-treated patient declines 2–3 points over the same time period. Memantine, used in conjunction with cholinesterase inhibitors or by itself, slows cognitive deterioration and decreases caregiver burden for patients with moderate to severe AD, but is not approved for mild AD. Neither cholinesterase inhibitors nor memantine have proven efficacious in patients with MCI, though the clinical trials lacked biomarkers at the time, and in retrospect likely included a mix of patients with AD and non-AD-related memory impairment. Cholinesterase inhibitors are easy to administer, and their major side effects are gastrointestinal symptoms (nausea, diarrhea, cramps), altered sleep with unpleasant or vivid dreams, bradycardia (usually benign), and muscle cramps. Potential side effects associated with memantine include constipation, dizziness, headache, and somnolence. A common approach to AD drug therapy is to initiate a cholinesterase inhibitor for a patient diagnosed with mild AD dementia and to add memantine when patients enter the moderate stage of disease. Cholinesterase inhibitors may also be effective in treating delusions and hallucinations, while memantine can reduce agitation.

THERAPIES TARGETING AMYLOID-β

A novel class of antiamyloid monoclonal antibodies has recently been approved for treatment of early clinical stages of AD. These drugs promote clearance of target Aβ epitopes, substantially lower amyloid plaque burden on amyloid PET, and comprise the first class of disease-modifying therapies to receive FDA approval for treatment of AD. Lecanemab (which targets Aβ protofibrils) and donanemab (which targets a pyroglutamate form of Aβ in plaques) are fully FDA approved for clinical use in early-stage AD based on evidence of clinical efficacy in phase 3, double-blinded, randomized, placebo-controlled clinical trials. Lecanemab is delivered as an intravenous infusion every 2 weeks, while donanemab is administered intravenously monthly. A third antibody, aducanumab, was the first to receive accelerated FDA approval based on reduction of amyloid PET signal, but evidence of clinical efficacy was ambiguous, and ultimately this drug has been removed from clinical use. All three of these antibodies lead to robust reductions in amyloid burden as measured by PET. Interestingly, drug trials of Aβ-targeting antibodies that showed less robust amyloid lowering on PET (e.g., solanezumab, crenezumab, gantenerumab) did not find a clinical benefit. Antiamyloid antibody therapy is currently restricted to patients with MCI or mild dementia and biomarker confirmation of Aβ pathology by PET or CSF. Clinical trials are underway for cognitively unimpaired individuals with biomarker evidence of Aβ pathology (i.e., preclinical AD). Patients with clinical features suggestive of non-AD causes of cognitive decline were excluded from the pivotal clinical trials and should not be treated, regardless of biomarker status.

In Clarity-AD, a phase 3, randomized, placebo-controlled trial of lecanemab in patients with early symptomatic AD, patients who received lecanemab for 18 months experienced an average 27% slowing in their rate of clinical decline compared to placebo, as measured by change in the Clinical Dementia Rating Scale sum of boxes score (CDR-SB), a clinical scale that measures cognition and function and was the study's primary outcome. Similar slowing of decline was reported on all secondary cognitive and functional outcomes. In TRAILBLAZER-ALZ2, a phase 3, randomized, placebo-controlled trial of donanemab in early symptomatic AD, patients on active treatment showed 22% slowing on the integrated Alzheimer Disease Rating Scale (iADRS; primary outcome) and 28% slowing on the CDR-SB (secondary outcome) over 76 weeks compared to placebo, with similar effects on all other secondary clinical endpoints. Patients in TRAILBLAZER-ALZ2 were stratified by tau PET at baseline, with patients with low-medium tau PET at trial entry showing the greatest clinical benefit (35% slowing on iADRS, 36% slowing on CDR-SB), suggesting that amyloid lowering may particularly benefit patients in earlier stages of tau spread. The duration of treatment was titrated based on amyloid PET response, with patients switched from donanemab to placebo when complete amyloid clearance was seen on PET. These findings suggest that limited-duration treatment until PET clearance may be sufficient to derive a robust clinical response. Given these results, future clinical practice may leverage baseline biomarkers of tau burden to identify patients more likely to benefit from antiamyloid therapy, whereas longitudinal biomarkers of Aβ burden may help to define the necessary duration of therapy.

Antiamyloid antibodies are associated with significant potential side effects, including infusion reactions and a form of secondary brain inflammation and hemorrhage, collectively referred to as amyloid-related imaging abnormalities (ARIA) **(Fig. 442-4)**. Rates of infusion reactions in the active treatment arms of the phase 3 trials were 26% for lecanemab and 9% for donanemab. ARIA-E manifests as focal areas of edema or sulcal effusions, detected by T2/fluid-attenuated inversion recovery (FLAIR) hyperintensities on MRI. ARIA-H typically manifests as microbleeds or superficial siderosis detected by gradient echo (GRE) or susceptibility-weighted imaging (SWI) MRI sequences, although lobar hemorrhages can rarely occur. Both ARIA-E and ARIA-H are thought to be related to CAA. Spontaneous ARIA-H occurs frequently in AD, but spontaneous ARIA-E (also known as inflammatory CAA) occurs very rarely in lieu of antiamyloid therapies. In the phase 3 trials, overall rates of ARIA-E were 13% for lecanemab and 24% for donanemab, whereas ARIA-H rates were 17% for lecanemab and 31% for donanemab. Approximately 75% of ARIA cases were asymptomatic and detected on safety MRI scans conducted throughout the study. When present,

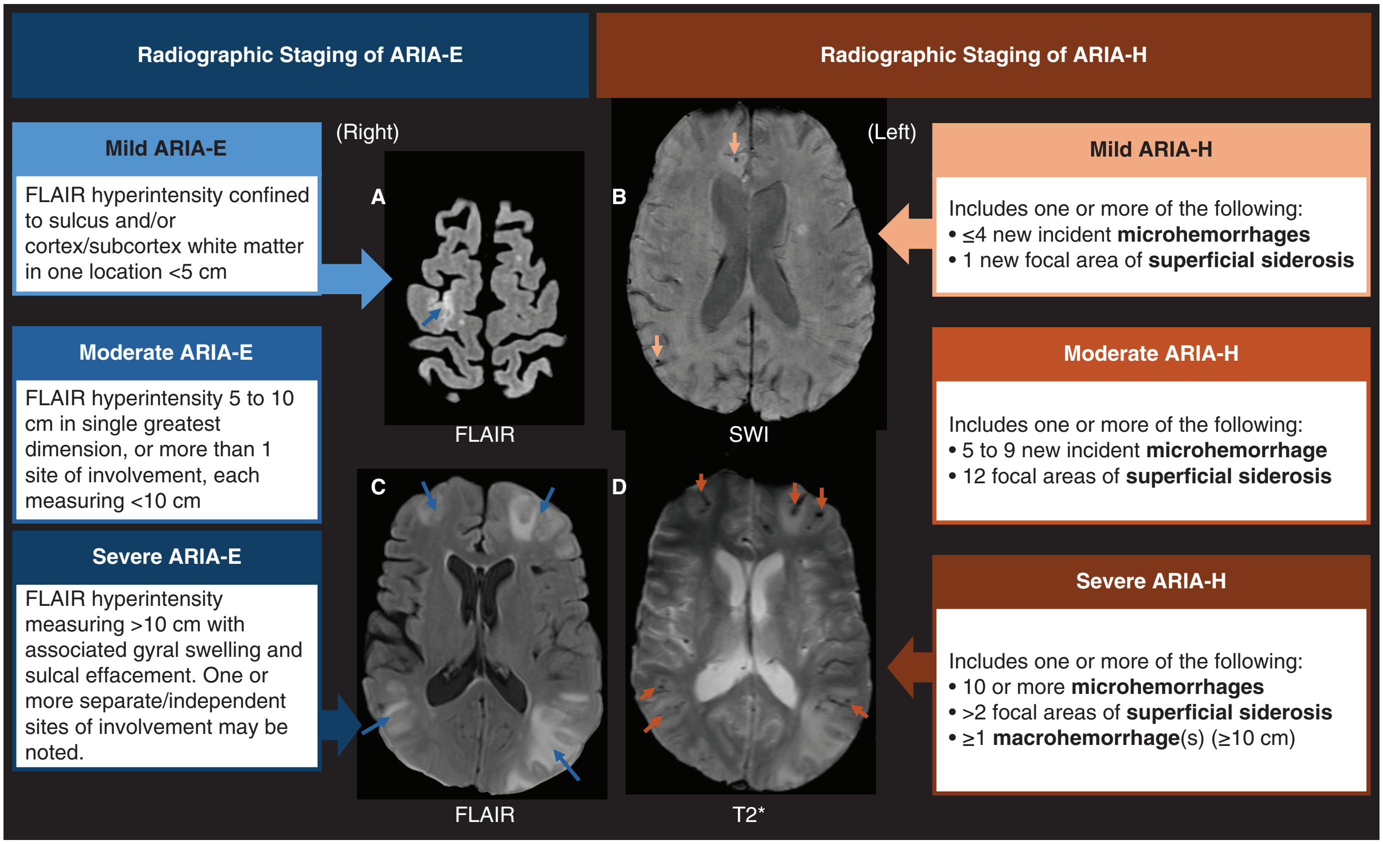

Radiographic Staging of ARIA-E

Mild ARIA-E

FLAIR hyperintensity confined to sulcus and/or cortex/subcortex white matter in one location <5 cm

Moderate ARIA-E

FLAIR hyperintensity 5 to 10 cm in single greatest dimension, or more than 1 site of involvement, each measuring <10 cm

Severe ARIA-E

FLAIR hyperintensity measuring >10 cm with associated gyral swelling and sulcal effacement. One or more separate/independent sites of involvement may be noted.

Radiographic Staging of ARIA-H

Mild ARIA-H

Includes one or more of the following:
- ≤4 new incident **microhemorrhages**
- 1 new focal area of **superficial siderosis**

Moderate ARIA-H

Includes one or more of the following:
- 5 to 9 new incident **microhemorrhage**
- 12 focal areas of **superficial siderosis**

Severe ARIA-H

Includes one or more of the following:
- 10 or more **microhemorrhages**
- >2 focal areas of **superficial siderosis**
- ≥1 **macrohemorrhage**(s) (≥10 cm)

FIGURE 442-4 Radiographic staging of amyloid-related imaging abnormalities (ARIAs). ***A.*** Axial fluid-attenuated inversion recovery (FLAIR) image depicting radiographically mild ARIA-E in an asymptomatic patient receiving lecanemab treatment. Cortical edema and adjacent sulcal effacement are present in a right superior frontal region with <5 cm of involvement. ***B.*** Axial susceptibility-weighted imaging (SWI) depicting radiographically mild ARIA-H (including right anterior cingulate and right parietal lobar microhemorrhages) in an asymptomatic patient receiving lecanemab. ***C–D.*** Axial FLAIR and T2* sequence images depicting radiographically severe ARIA-E and radiographically severe ARIA-H, respectively, in a patient presenting with focal neurologic symptoms localized to the left occipital lobe during aducanumab treatment. FLAIR imaging reveals multiple regions of cortical edema, with adjacent subcortical edema and sulcal effacement. T2* reveals multiple regions of cortical siderosis, including in bilateral frontal and posterior regions of concomitant ARIA-E, as well as numerous lobar microhemorrhages.

ARIA symptoms are typically mild and nonspecific (e.g., headache, dizziness, greater confusion). However, severe symptoms, including seizures, stroke-like episodes, malignant hypertension, and (very rarely) death can occur. The risk of ARIA increases with each copy of the Apo ε4 allele and is highest in Apo ε4 homozygotes. Apo ε4 homozygotes are also at increased risk of symptomatic, severe, and recurrent ARIA, leading the FDA to issue a box warning for ε4 homozygotes in the prescribing information for both lecanemab and donanemab. Clinical Apo ε genotyping is recommended prior to initiating anti-Aβ antibody therapy to adequately inform shared decision-making about treatment risks and benefits.

In a clinical setting, uncontrolled hypertension, radiographic evidence of extensive CAA (>4 brain microhemorrhages or ≥1 region of cortical siderosis), and other factors that may increase ARIA risk are regarded as contraindications for antiamyloid antibodies. Limited data are available about the safety of anticoagulants in the context of anti-Aβ therapy, with the concern that anticoagulation may increase the risk of ARIA-H. Expert recommendations suggest excluding patients on anticoagulants from treatment until more safety data are available. Patients treated with a single antiplatelet agent do not appear to be at higher risk for ARIA.

Patients treated with lecanemab or donanemab should be monitored for ARIA with multiple surveillance MRIs during the first 6–12 months of treatment, following the schedule recommended by the FDA in the prescribing information. T2/FLAIR and GRE/SWI sequences should be performed, and patients should ideally be imaged longitudinally on the same MRI scanner and sequences to allow optimal comparisons between time points. An urgent MRI should be performed whenever ARIA is clinically suspected based on symptoms. The severity of ARIA is graded based on radiographic criteria, as well as the presence and severity of clinical symptoms. Patients can be treated safely through asymptomatic and radiographically mild ARIA, provided MRI scans are performed monthly to monitor evolution. Symptomatic or radiographically moderate to severe ARIA should lead to temporary suspension of treatment until ARIA-E resolves and ARIA-H stabilizes. Permanent cessation of treatment is recommended if symptoms are severe, ARIA has recurred more than twice, ARIA-H has led to macrohemorrhage or >10 additional microhemorrhages since treatment initiation, or more than one area of superficial siderosis has emerged.

Cases of catastrophic, multifocal brain hemorrhages due to severe ARIA-E and ARIA-H leading to death have been reported in patients receiving antiamyloid antibody therapy while being treated with thrombolytic medications (e.g., tissue plasminogen activator) for acute stroke. When presenting with stroke-like symptoms, patients receiving anti-Aβ monoclonal antibodies should receive an immediate MRI to differentiate acute stroke from ARIA. Until more data are available, treatment with lecanemab or donanemab should therefore be considered a contraindication to peripheral thrombolysis.

It is current practice at our institution to restrict use of antiamyloid monoclonal antibodies to patients with biomarker-confirmed Aβ pathology and early symptomatic disease, including MCI or mild AD dementia, who are not expected to require anticoagulant therapy. We corroborate interpretation of Aβ biomarkers and scrutinize baseline brain MRI for disqualifying lesions that may increase risk of ARIA and intracerebral hemorrhage, such as radiographic evidence of extensive CAA. All patients must know their ApoE allele status and understand their associated ARIA risk before starting treatment. Apo ε4 homozygous patients may elect to pursue therapy at our institution after careful consideration of their ARIA risk, although several institutions do prohibit or carefully restrict use in Apo ε4 homozygous patients. Finally, and consistent with consensus clinical practice, all patients who receive therapy must undergo regular MRI safety screening to rule out ARIA.

PATIENT AND CAREGIVER EDUCATION

Building rapport with the patient, family members, and other caregivers is essential. In the early stages of AD, memory aids such as notebooks and posted daily reminders can be helpful. Family members should emphasize activities that are pleasant while curtailing those that increase stress on the patient. Kitchens, bathrooms, stairways, and bedrooms need to be made safe, and eventually, patients will need to stop driving. Patients can be encouraged to engage in lifestyle modifications that may be protective against aging and neurodegeneration, such as physical, cognitive, and social activity; vascular risk factor modification; healthy sleep habits; and dietary modifications (e.g., Mediterranean or Dietary Approaches to Stop Hypertension diets). Loss of independence and change of environment may worsen confusion, agitation, and anger. Communication and repeated calm reassurance are necessary. Caregiver "burnout" is common, often resulting in nursing home placement of the patient or new health problems for the caregiver. Respite breaks for the caregiver help to maintain a successful long-term therapeutic milieu. Use of adult day care centers can be helpful. Local and national support groups, such as the Alzheimer's Association and the Family Caregiver Alliance, are valuable resources. Internet access to these resources has become available to clinicians and families in recent years.

ADDITIONAL THERAPIES

Mild to moderate depression is common in the early stages of AD and may respond to antidepressants or cholinesterase inhibitors. Selective serotonin reuptake inhibitors (SSRIs) are commonly used due to their low anticholinergic side effects (e.g., escitalopram, target dose, 5–10 mg daily). Seizures can be treated with levetiracetam unless the patient had a different regimen that was effective prior to the onset of AD. Agitation, insomnia, hallucinations, and belligerence are especially troublesome characteristics of some AD patients, and these behaviors can lead to nursing home placement. The newer generation of atypical antipsychotics, such as risperidone, quetiapine, and olanzapine, are being used in low doses to treat these neuropsychiatric symptoms. Brexpiprazole, an atypical antipsychotic that acts on noradrenergic, serotonergic, and dopaminergic neurotransmitter systems, is the only drug to receive FDA approval for treatment of agitation in AD, based on a 12-week, double-blinded, randomized, placebo-controlled trial. The few controlled studies comparing drugs against behavioral intervention in the treatment of agitation suggest mild efficacy with significant side effects related to sleep, gait, and cardiovascular complications, including an increased risk of death. Antipsychotics carry a black box FDA warning for use in elderly patients with dementia and thus should be prescribed only with caution; however, careful, daily, nonpharmacologic behavior management is often not available, rendering medications necessary for some patients. Medications with strong anticholinergic effects should be vigilantly avoided, including prescription and over-the-counter sleep aids (e.g., diphenhydramine) or incontinence therapies (e.g., oxybutynin).

Several commonly used medications and supplements, including estrogen hormone replacement therapy, statins, vitamin E, and ginkgo biloba, appeared to be associated with a decreased risk of AD in epidemiologic or observational studies but did not show efficacy in prospective, randomized, double-blinded, placebo-controlled trials. Many vitamins and dietary supplements are marketed directly to consumers as "memory enhancing" or protective against AD without clinical evidence. Patients and families may come across anecdotal reports of "miraculous" responses to aggressive treatments such as anti-interferon intrathecal infusions, intravenous immunoglobulin, antibiotics (purportedly to treat Lyme disease or another questionable infection), metal chelation, and stem cell therapies, but there is no scientific evidence to support use of any of these approaches to treating AD and significant concern for harm.

EXPERIMENTAL THERAPIES

The design of AD clinical trials has been transformed by the availability of PET, CSF, and more recently blood-based biomarkers of Aβ and tau. Many trials now require biomarker evidence of AD for trial inclusion. Biomarkers help assess target engagement (e.g., changes in Aβ biomarkers in an antiamyloid trial) or modification of downstream disease pathophysiology (e.g., changes in tau biomarkers in an antiamyloid trial), with the pivotal trials leading to approval of lecanemab and donanemab being emblematic of this novel approach. Increasingly, many trials have shifted toward enrolling patients in the asymptomatic (preclinical) or very early symptomatic stages of AD, using positive biomarkers as the primary inclusion criterion. Primary (biomarker negative) and secondary (biomarker positive but no symptoms) prevention trials are underway in autosomal dominant mutation carriers, Apo ε4 homozygotes, and even in the normally aging population.

Active vaccination against Aβ is another approach that aims to promote immune-mediated clearance of amyloid pathology. The first $A\beta_{42}$ vaccine trial in humans was aborted after a minority of patients developed meningoencephalitis, but subsequent trials with less immunogenic formulations have shown more favorable safety profiles.

Oral drugs that inhibit β and γ secretase reduce the cleavage of APP to $A\beta_{42}$ and showed promise in ameliorating pathology and behavioral changes in AD transgenic mice. Unfortunately, placebo-controlled trials failed to show clinical efficacy, and trials of β secretase inhibitors, in particular, found significant *worsening* of cognition in treated patients versus placebo, although fortunately, this effect proved transient after discontinuing the drug. It is unclear whether toxicity of β and γ secretase inhibitors was directly related to changes in Aβ metabolism or to "off-target" drug effects.

Monoclonal antibodies directed against phosphorylated tau are in earlier stages of development. These antibodies aim to prevent the transsynaptic spread of tau and have proven effective in tau transgenic mice. Safety profiles in human studies have proven favorable thus far, but clinical results have been lacking. Lowering of tau expression via antisense oligonucleotides (ASOs) or small interfering RNA is a compelling therapeutic strategy that rescues most elements of the AD phenotype in AD transgenic mice. A recent phase 1 trial of the tau-targeting ASO MAPT_{RX}, delivered intrathecally, showed the treatment to be well tolerated, with significant lowering of tau biomarkers in CSF and PET. A phase 2 clinical trial is currently underway. Additional therapeutic approaches targeting tau include active immunization; inhibition of tau phosphorylation, acetylation, and aggregation; and microtubule stabilization. Other druggable pathways represented in the AD drug development pipeline include those targeting neuroinflammation, metabolism/bioenergetics, synaptic plasticity, neuroprotection, and neurotransmitter-based cognitive enhancement or treatment of neuropsychiatric symptoms.

A general approach to the symptomatic management of dementia is presented in Chap. 31.

OTHER CAUSES OF DEMENTIA

FTD **(Chap. 443)**, vascular dementia **(Chap. 444)**, DLB **(Chap. 445)**, and prion diseases **(Chap. 449)** are covered in dedicated chapters.

Prion diseases such as CJD are rare neurodegenerative conditions (prevalence ~1 per million) that produce dementia. CJD is a rapidly progressive disorder associated with dementia, focal cortical signs, rigidity, and myoclonus, causing death <1 year after first symptoms appear. The rapidity of progression seen with CJD is uncommon in AD so that the distinction between the two disorders is usually straightforward, although AD can on occasion present as a rapidly progressive dementia. In general, CBD **(Chap. 443)** and DLB **(Chap. 445)**, more rapid degenerative dementias with prominent movement abnormalities, are more likely to be mistaken for CJD. The differential diagnosis for CJD includes other rapidly progressive dementing conditions such as viral or bacterial encephalitides, Hashimoto's encephalopathy, central nervous system (CNS) vasculitis, lymphoma, or paraneoplastic/autoimmune syndromes **(Chap. 99)**. The markedly abnormal periodic complexes on EEG and cortical ribboning and basal ganglia hyperintensities on diffusion-weighted imaging or FLAIR MRI are diagnostic features of CJD, although rarely, prolonged focal or generalized seizures can produce a similar imaging appearance.

Huntington's disease (HD) **(Chap. 447)** is an autosomal dominant degenerative brain disorder. HD clinical hallmarks include chorea, behavioral disturbance, and executive impairment. Symptoms typically begin in the fourth or fifth decade, but there is a wide range, from childhood to >70 years. Memory is frequently not impaired until late in the disease, but attention, judgment, self-awareness, and executive functions are often deficient at an early stage. Depression, apathy, social withdrawal, irritability, and intermittent disinhibition are common. Delusions and obsessive-compulsive behavior may occur. Disease duration is variable but typically lasts ~15 years.

NPH is a relatively uncommon but treatable syndrome. The clinical, physiologic, and neuroimaging characteristics of NPH must be carefully distinguished from those of other dementias associated with gait impairment. Historically, many patients treated for NPH have suffered from other dementias, particularly AD, vascular dementia, DLB, and PSP **(Chap. 443)**. For NPH, the clinical triad includes an abnormal gait (ataxic or apractic), dementia (usually mild to moderate, with an emphasis on executive impairment), and urinary urgency or incontinence. Neuroimaging reveals enlarged lateral ventricles (hydrocephalus) with little or no cortical atrophy, although the sylvian fissures may appear propped open (so-called "boxcarring"), which can be mistaken for perisylvian atrophy. Crowding of dorsal frontal-parietal gyri helps distinguish NPH from other movement disorders, such as PSP and CBD, in which dorsal atrophy with sulcal widening is common. NPH is a communicating hydrocephalus with a patent aqueduct of Sylvius **(see Fig. 31-3)**, in contrast to aqueductal stenosis, in which the aqueduct is small. Lumbar puncture opening pressure falls in the high-normal range, and the CSF protein, glucose, and cell counts are normal. NPH may be caused by obstruction to normal CSF flow over the cerebral convexities and delayed resorption into the venous system. The indolent nature of the process results in enlarged lateral ventricles with relatively little increase in CSF pressure. Presumed edema, stretching, and distortion of subfrontal white matter tracts may lead to clinical symptoms, but the precise underlying pathophysiology remains unclear. Some patients provide a history of conditions that produce meningeal scarring (blocking CSF resorption) such as previous meningitis, subarachnoid hemorrhage, or head trauma. Others with longstanding but asymptomatic congenital hydrocephalus may have adult-onset deterioration in gait or memory that is confused with NPH. In contrast to AD, the patient with NPH complains of an early and prominent gait disturbance without cortical atrophy on CT or MRI.

Numerous attempts to improve NPH diagnosis with various special studies and predict the success of ventricular shunting have been undertaken. These tests include radionuclide cisternography (showing a delay in CSF absorption over the convexity) and various efforts to monitor and alter CSF flow dynamics, including a constant-pressure infusion test. None has proven to be specific or consistently useful. A transient improvement in gait or cognition may follow lumbar puncture (or serial punctures) with removal of 30–50 mL of CSF, but this finding has also not proved to be consistently predictive of postshunt improvement. Perhaps the most reliable strategy is a period of close inpatient evaluation before, during, and after lumbar CSF drainage. Occasionally, when a patient with AD presents with gait impairment (at times due to comorbid subfrontal vascular injury) and absent or only mild cortical atrophy on CT or MRI, distinguishing NPH from AD can be challenging. Hippocampal atrophy on MRI favors AD, whereas a characteristic "magnetic" gait with external hip rotation, low foot clearance, and short strides, along with prominent truncal sway or instability, favors NPH. The diagnosis of NPH should be avoided when hydrocephalus is not detected on imaging studies, even if the symptoms otherwise fit. Thirty to fifty percent of patients identified by careful diagnosis as having NPH will improve with ventricular shunting. Gait may improve more than cognition, but many reported failures to improve cognitively may have resulted from comorbid AD. Importantly, the presence of positive CSF AD biomarkers or amyloid PET is associated with lower likelihood of response to shunting. Short-lasting improvement is common. Patients should be carefully selected for shunting, because subdural hematoma, infection, and shunt failure are known complications and can be a cause for early nursing home placement in an elderly patient with previously mild dementia.

Intracranial hypotension, sometimes called sagging brain syndrome, is a disorder caused by low CSF pressure, leading to downward pressure on the subcortical structures and disruption of cerebral function. It presents in a variable manner with headache, often exacerbated by coughing or a Valsalva maneuver or by moving from lying to standing. Other common symptoms include dizziness, vomiting, disruption of sleep-wake cycles, and sometimes a progressive behavioral variant FTD-like syndrome **(Chap. 443)**. Although sometimes idiopathic, this syndrome can be caused by CSF leaks secondary to lumbar puncture, head trauma, or spinal cord arachnoid cysts. Treatment consists of finding and patching the CSF leak.

Dementia can accompany *chronic alcoholism* **(Chap. 464)** and may result from associated malnutrition, especially of B vitamins, particularly thiamine. Other poorly defined aspects of chronic alcoholism may, however, also produce cerebral damage. A rare idiopathic syndrome of dementia and seizures with degeneration of the corpus callosum has been reported primarily in male Italian red wine drinkers (Marchiafava-Bignami disease).

Thiamine (vitamin B_1) deficiency causes Wernicke's encephalopathy **(Chap. 318)**. The clinical presentation is usually a malnourished patient (frequently but not necessarily alcoholic) with confusion, ataxia, and diplopia resulting from inflammation and necrosis of periventricular midline structures, including dorsomedial thalamus, mammillary bodies, midline cerebellum, periaqueductal gray matter, and trochlear and abducens nuclei. Damage to the dorsomedial thalamus correlates most closely with the memory loss. Prompt administration of parenteral thiamine (100 mg intravenously for 3 days followed by daily oral dosage) may reverse the disease if given within the first days of symptom onset. Prolonged untreated thiamine deficiency can result in an irreversible and profound amnestic syndrome (Korsakoff's syndrome) or even death.

In *Korsakoff's syndrome*, the patient is unable to recall new information despite normal immediate memory, attention span, and level of consciousness. Memory for new events is seriously impaired, whereas knowledge acquired prior to the illness remains relatively intact. Patients are easily confused, disoriented, and cannot store information for more than a few minutes. Superficially, they may be conversant, engaging, and able to perform simple tasks and follow immediate commands. Confabulation is common, although not always present. There is no specific treatment because the previous thiamine deficiency has produced irreversible damage to the medial thalamic nuclei and mammillary bodies. Mammillary body atrophy may be visible on MRI in the chronic phase **(see Fig. 318-6)**.

Vitamin B_{12} deficiency, as can occur in pernicious anemia, causes a megaloblastic anemia and may also damage the nervous system **(Chaps. 104 and 453)**. Neurologically, it most commonly produces a spinal cord syndrome (myelopathy) affecting the posterior columns (loss of vibration and position sense) and corticospinal tracts (hyperactive tendon reflexes with Babinski signs); it also damages peripheral nerves (neuropathy), resulting in sensory loss with depressed tendon reflexes. Damage to myelinated axons may also cause dementia. The mechanism of neurologic damage is unclear but may be related to a deficiency of *S*-adenosyl methionine (required for methylation of myelin phospholipids) due to reduced methionine synthase activity or accumulation of methylmalonate, homocysteine, and propionate, providing abnormal substrates for fatty acid synthesis in myelin. Use of histamine blockers or metformin, vegan diets, autoimmunity against gastric parietal cells, and various causes of malabsorption are the typical causes for vitamin B_{12} deficiency. The neurologic sequelae of vitamin B_{12} deficiency may occur in the absence of hematologic manifestations, making it critical to avoid using the complete blood count (CBC) and blood smear as a substitute for measuring B_{12} blood levels. Treatment with parenteral vitamin B_{12} (1000 μg intramuscularly daily for a week, weekly for a month, and monthly for life for pernicious anemia) stops progression of the disease if instituted promptly, but complete reversal of advanced nervous system damage will not occur.

Deficiency of nicotinic acid (*pellagra*) is associated with skin rash over sun-exposed areas, glossitis, and angular stomatitis **(Chap. 344)**. Severe dietary deficiency of nicotinic acid along with other B vitamins such as pyridoxine may result in spastic paraparesis, peripheral neuropathy, fatigue, irritability, and dementia. This syndrome has been seen in prisoners of war and in concentration camps but should be considered in any malnourished individual. Low serum folate levels appear to be a rough index of malnutrition, but isolated folate deficiency has not been proved as a specific cause of dementia.

CNS infections usually cause delirium and other acute neurologic syndromes. However, some chronic CNS infections, particularly those associated with chronic meningitis **(Chap. 144)**, may produce a dementing illness. The possibility of chronic infectious meningitis should be suspected in patients presenting with a dementia or behavioral syndrome, who also have headache, meningismus, cranial neuropathy, and/or radiculopathy. Between 20 and 30% of patients in the advanced stages of HIV infection become demented **(Chap. 208)**. Cardinal features include psychomotor retardation, apathy, and impaired memory. This syndrome may result from secondary opportunistic infections but can also be caused by direct infection of CNS neurons with HIV. Neurosyphilis **(Chap. 187)** was a common cause of dementia in the preantibiotic era; it is now uncommon but can still be encountered in patients with multiple sex partners, particularly among patients with HIV. Characteristic CSF changes consist of pleocytosis, increased protein, and a positive Venereal Disease Research Laboratory (VDRL) test. The recent SARS-CoV-2 pandemic was associated in some individuals with persistent postrecovery changes in memory, executive, and other cognitive functions; responsible mechanisms might include effects of inflammation, multiorgan system failure, or virus-associated vascular injury.

Primary and metastatic *neoplasms of the CNS* **(Chap. 95)** usually produce focal neurologic findings and seizures rather than dementia, but if tumor growth begins in the frontal or temporal lobes, the initial manifestations may be memory loss or behavioral changes. An autoimmune, sometimes paraneoplastic, syndrome of dementia associated with occult carcinoma (often small-cell lung cancer) is termed *limbic encephalitis*. In this syndrome, confusion, agitation, seizures, poor memory, emotional changes, and frank dementia may occur. Paraneoplastic *encephalitis associated with NMDA receptor antibodies* presents as a progressive psychiatric disorder with memory loss and seizures; affected patients are often young women with ovarian teratoma. Autoimmune etiologies also include antibodies targeting leucine-rich glioma-inactivated 1 (LGI1; faciobrachial dystonic seizures); contactin-associated protein-like 2 (Caspr2; insomnia, ataxia, myotonia); and α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) receptor (limbic encephalitis with relapses), among others **(Chap. 99)**.

A *nonconvulsive seizure disorder* **(Chap. 436)** may underlie a syndrome of confusion, clouding of consciousness, and garbled speech. Often, psychiatric disease is suspected, but an EEG demonstrates the epileptic nature of the illness. If recurrent or persistent, the condition may be termed *complex partial status epilepticus*. The cognitive disturbance often responds to anticonvulsant therapy. The etiology may be previous small strokes or head trauma; some cases are idiopathic. Nonconvulsive temporal lobe seizures can also emerge early in the course of AD.

It is important to recognize *systemic diseases* that indirectly affect the brain and produce chronic confusion or dementia. Such conditions include hypothyroidism; vasculitis; and hepatic, renal, or pulmonary disease. Hepatic encephalopathy may begin with irritability and confusion and slowly progress to agitation, lethargy, and coma.

Isolated vasculitis of the CNS (CNS granulomatous angiitis) **(Chaps. 375 and 438)** occasionally causes a chronic encephalopathy associated with confusion, disorientation, and clouding of consciousness. Headache is common, and strokes and cranial neuropathies may occur. Brain imaging studies may be normal or nonspecifically abnormal. CSF analysis reveals a mild pleocytosis or protein elevation. Cerebral angiography can show multifocal stenoses involving medium-caliber vessels, but some patients have only small-vessel disease that is not revealed on angiography. The angiographic appearance is not specific and may be mimicked by atherosclerosis, infection, or other causes of vascular disease. Brain or meningeal biopsy demonstrates endothelial cell proliferation and mononuclear infiltrates within blood vessel walls. The prognosis is often poor, although the disorder may remit spontaneously. Some patients respond to glucocorticoids or chemotherapy.

Chronic metal exposure represents a rare cause of dementia. The key to diagnosis is to elicit a history of exposure at work or home. Chronic lead poisoning from inadequately fire-glazed pottery has been reported. Fatigue, depression, and confusion may be associated with episodic abdominal pain and peripheral neuropathy. Gray lead lines appear in the gums, usually accompanied by an anemia with basophilic stippling of red blood cells. The clinical presentation can resemble that of acute intermittent porphyria **(Chap. 428)**, including elevated levels of urine porphyrins as a result of the inhibition of δ-aminolevulinic acid dehydrase. The treatment is chelation therapy with agents such as ethylenediamine tetraacetic acid (EDTA). Chronic mercury poisoning produces dementia, peripheral neuropathy, ataxia, and tremulousness that may progress to a cerebellar intention tremor or choreoathetosis. The confusion and memory loss of chronic arsenic intoxication is also associated with nausea, weight loss, peripheral neuropathy, pigmentation and scaling of the skin, and transverse white lines of the fingernails (Mees' lines). Treatment is chelation therapy with dimercaprol (BAL). Aluminum poisoning is rare but was documented with the dialysis dementia syndrome, in which water used during renal dialysis was contaminated with excessive amounts of aluminum. This poisoning resulted in a progressive encephalopathy associated with confusion, nonfluent aphasia, memory loss, agitation, and, later, lethargy and stupor. Speech arrest and myoclonic jerks were common and associated with severe and generalized EEG changes. The condition has been eliminated by the use of deionized water for dialysis.

Recurrent head trauma in professional athletes may lead to a dementia previously referred to as "punch-drunk" syndrome or *dementia pugilistica* but now known as chronic traumatic encephalopathy (CTE) to signify its relevance to contact sport athletes other than boxers **(Chap. 454)**. The symptoms can be progressive, beginning late in an athlete's career or, more often, after retirement. Early in the course, a personality change occurs, associated with social instability, explosive rage, and sometimes paranoia and delusions. Later, memory loss progresses to full-blown dementia, often associated with parkinsonian signs and ataxia or intention tremor. At autopsy, the cerebral cortex shows tau-immunoreactive NFTs that are more prominent than amyloid plaques (which are usually diffuse or absent rather than neuritic). NFTs and tau-positive reactive astrocytes are often clustered in the depths of cortical sulci and in a perivascular distribution. TDP-43 inclusions have also been reported, highlighting the overlap with the FTD spectrum **(Chap. 443)**. Loss of neurons in the substantia nigra is a variable feature, and some with TDP-43 inclusions also develop motor neuron disease (MND) **(Chap. 448)**.

Chronic subdural hematoma **(Chap. 454)** is also occasionally associated with dementia, often in the context of underlying cortical atrophy from conditions such as AD or HD.

Transient global amnesia (TGA) is characterized by the sudden onset of a severe episodic memory deficit, usually occurring in persons aged >50 years. Often the amnesia occurs in the setting of an emotional stimulus or physical exertion. During the attack, the individual is alert and communicative, general cognition seems intact, and there are no other neurologic signs or symptoms. The patient may seem confused and repeatedly ask about their location in place and time. The ability to form new memories returns after a period of hours, and the individual returns to normal with no recall for the period of the attack. Frequently no cause is determined, but cerebrovascular disease, epilepsy (7% in one study), migraine, or cardiac arrhythmias have all been implicated. Approximately one-quarter of patients experience recurrent attacks. Rare instances of permanent memory loss have been reported in patients with TGA-like spells, usually representing ischemic infarction of the hippocampus or dorsomedial thalamic nucleus bilaterally. Seizure activity due to AD should always be suspected in this syndrome.

The *ALS/parkinsonian/dementia complex of Guam* is a rare degenerative disease that has occurred in the Chamorro natives on the island of Guam. Individuals may have any combination of parkinsonian features, dementia, and MND. The most characteristic pathologic features are the presence of NFTs in degenerating neurons of the cortex and substantia nigra and loss of motor neurons in the spinal cord, although recent reanalysis has shown that some patients with this illness also show coexisting TDP-43 pathology. Epidemiologic evidence supports a possible environmental cause, such as exposure to a neurotoxin or an infectious agent with a long latency period. One interesting but unproven candidate neurotoxin is the seed of the false palm tree, which Guamanians traditionally used to make flour. The amyotrophic lateral sclerosis (ALS) syndrome is no longer present in Guam, but a dementing illness with rigidity continues to be seen.

Rarely, adult-onset leukodystrophies, lysosomal storage diseases, and other genetic disorders can present as a dementia in middle to late life. Metachromatic leukodystrophy (MLD) causes a progressive psychiatric or dementia syndrome associated with an extensive, confluent frontal white matter abnormality. MLD is diagnosed by measuring reduced arylsulfatase A enzyme activity in peripheral white blood cells. Adult-onset presentations of adrenoleukodystrophy have been reported in female carriers, and these patients often feature spinal cord and posterior white matter involvement. Adrenoleukodystrophy is diagnosed by demonstrating increased levels of plasma very-long-chain fatty acids. CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy) is another genetic syndrome associated with white matter disease, often frontally and temporally predominant. Diagnosis is made with skin biopsy, which shows osmophilic granules in arterioles, or increasingly through genetic testing for mutations in Notch 3. The neuronal ceroid lipofuscinoses are a genetically heterogeneous group of disorders associated with myoclonus, seizures, vision loss, and progressive dementia. Diagnosis is made by finding eosinophilic curvilinear inclusions within white blood cells or neuronal tissue.

Psychogenic amnesia for personally important memories can be seen. Whether this results from deliberate avoidance of unpleasant memories, outright malingering, or unconscious repression remains unknown and probably depends on the patient. Event-specific amnesia is more likely to occur after violent crimes such as homicide of a close relative or friend or sexual abuse. It may develop in association with severe drug or alcohol intoxication and sometimes with schizophrenia. More prolonged psychogenic amnesia occurs in fugue states that also commonly follow severe emotional stress. The patient with a fugue state suffers from a sudden loss of personal identity and may be found wandering far from home. *In contrast to neurologic amnesia, fugue states are associated with amnesia for personal identity and events closely associated with the personal past.* At the same time, memory for other recent events and the ability to learn and use new information are preserved. The episodes usually last hours or days and occasionally weeks or months while the patient takes on a new identity. On recovery, there is a residual amnesia gap for the period of the fugue. Very rarely does selective loss of autobiographic information reflect a focal injury to the brain areas involved with these functions.

Psychiatric diseases may mimic dementia. Severely depressed or anxious individuals may appear demented, a phenomenon sometimes called *pseudodementia*. Memory and language are usually intact when carefully tested, and a significant memory disturbance usually suggests an underlying dementia, even if the patient is depressed. Patients in this condition may feel confused and unable to accomplish routine tasks. Vegetative symptoms, such as insomnia, lack of energy, poor appetite, and concern with bowel function, are common. Onset is often more abrupt, and the psychosocial milieu may suggest prominent reasons for depression. Such patients respond to treatment of the underlying psychiatric illness. Schizophrenia is usually not difficult to distinguish from dementia, but occasionally the distinction can be problematic. Schizophrenia generally has a much earlier age of onset (second and third decades) than most dementing illnesses and is associated with intact memory. The delusions and hallucinations of schizophrenia are usually more complex, bizarre, and threatening than those of dementia. Some chronic schizophrenics develop an unexplained progressive dementia late in life that is not related to AD. Conversely, FTD, HD, vascular dementia, DLB, AD, or leukoencephalopathy can begin with schizophrenia-like features, leading to the misdiagnosis of a psychiatric condition. Later age of onset, significant deficits on cognitive testing, or the presence of abnormal neuroimaging suggest a degenerative condition. Memory loss may also be part of a *conversion disorder*. In this situation, patients commonly complain bitterly of memory loss, but careful cognitive testing either does not confirm the deficits or demonstrates inconsistent or unusual patterns of cognitive problems. The patient's behavior and "wrong" answers to questions often indicate that they understand the question and know the correct answer.

Clouding of cognition by *chronic drug or medication use*, often prescribed by physicians, is an important cause of dementia. Sedatives, tranquilizers, and analgesics used to treat insomnia, pain, anxiety, or agitation may cause confusion, memory loss, and lethargy, especially in the elderly. Discontinuation of the offending medication often improves mentation.

■ FURTHER READING

Andrews SJ et al: Interpretation of risk loci from genome-wide association studies of Alzheimer's disease. Lancet Neurol 19:326, 2020.

Belloy ME et al: A quarter century of APOE and Alzheimer's disease: Progress to date and the path forward. Neuron 101:820, 2019.

Cummings J et al: Progress in pharmacologic management of neuropsychiatric syndromes in neurodegenerative disorders: A review. JAMA Neurol 81:645, 2024.

Graff-Radford J et al: New insights into atypical Alzheimer's disease in the era of biomarkers. Lancet Neurol 20:222, 2021.

Jack CR et al: Revised criteria for the diagnosis and staging of Alzheimer's disease. Nat Med 30:2121, 2024.

Schindler SE et al: Acceptable performance of blood biomarker tests of amyloid pathology: Recommendations from the Global CEO Initiative on Alzheimer's Disease. Nat Rev Neurol 20:426, 2024.

Sims JR et al: Donanemab in early symptomatic Alzheimer disease: The TRAILBLAZER-ALZ 2 randomized clinical trial. JAMA 330:512, 2023.

Van Dyck CH et al: Lecanemab in early Alzheimer's disease. N Engl J Med 388:9, 2023.

443

Frontotemporal Dementia

William W. Seeley, Bruce L. Miller

Frontotemporal dementia (FTD) refers to a group of clinical syndromes united by their links to underlying frontotemporal lobar degeneration (FTLD) pathology. FTD, like the other major neurodegenerative diseases, is considered a disease of abnormal protein aggregation, with either tau or transactive response DNA-binding protein of 43 kDa (TDP-43) implicated in most cases. FTD most often begins in the fifth to seventh decades of life and is nearly as prevalent as Alzheimer's disease (AD) in this age group. Early studies suggested that FTD may be more common in men than women; however, more recent reports cast doubt on this finding. Although a family history of dementia is common, autosomal dominant inheritance is seen in only 10–20% of all FTD cases.

CLINICAL MANIFESTATIONS

Familial and sporadic forms of FTLD present with remarkable clinical heterogeneity. Three core clinical syndromes have been described **(Fig. 443-1)**. In the behavioral variant (bvFTD), the most common FTD syndrome, social and emotional dysfunction manifests as apathy, disinhibition, compulsivity, loss of empathy, and overeating, often but not always accompanied by deficits in executive control. Two forms of primary progressive aphasia (PPA), the semantic and nonfluent/agrammatic variants, are commonly due to FTLD and are included under the FTD umbrella. In the semantic variant, patients slowly lose the ability to decode word, object, person-specific, and emotion meaning, whereas patients with the nonfluent/agrammatic variant develop profound inability to produce words, often with prominent motor speech impairment. Any of these three clinical syndromes, but most often bvFTD, may be accompanied by motor neuron disease (MND) **(Chap. 448)**, in which case the term FTD-MND is applied. In addition, the corticobasal syndrome (CBS) and progressive supranuclear palsy–Richardson syndrome (PSP-RS) can be considered part of the FTD clinical spectrum. Furthermore, patients may evolve from any of the major syndromes described above to have prominent features of another syndrome.

Findings at the bedside are dictated by the anatomic localization of the disorder. Degeneration with atrophy occurs in the medial and orbital frontal cortex and anterior insula in bvFTD; the anterior temporal region in semantic variant PPA; and the opercular frontal and precentral gyrus of the dominant hemisphere in nonfluent/agrammatic PPA. Typically, parietal functions such as visuospatial processing and arithmetic calculations are unaffected even late in the FTD syndromes. Many patients with nonfluent aphasia or bvFTD later develop aspects of PSP-RS as disease spreads into subcortical or brainstem structures or CBS-like features appear as disease moves into perirolandic cortices.

GENETIC CONSIDERATIONS

Autosomal dominant forms of FTD can result from mutations in *C9orf72* (chromosome 9), *GRN* (chromosome 17), and *MAPT* (chromosome 17) genes. A hexanucleotide (GGGGCC) expansion in a noncoding exon of *C9ORF72* is the most common genetic cause of familial or sporadic FTD (usually presenting as bvFTD with or without MND) and amyotrophic lateral sclerosis (ALS). The expansion is associated with *C9orf72* haploinsufficiency, nuclear mRNA foci containing transcribed portions of the expansion and other mRNAs, neuronal cytoplasmic inclusions containing dipeptide repeat proteins translated from the repeat mRNA, and TDP-43 neuronal cytoplasmic and glial inclusions. The pathogenic significance of these various features is a topic of vigorous investigation. *MAPT* mutations lead to a change in the alternate splicing of tau or cause loss of function in the tau molecule, thereby altering microtubule binding. With *GRN*, mutations in the coding sequence of the gene encoding progranulin protein result in mRNA degradation due to nonsense-mediated decay, leading to a ~50% reduction in circulating progranulin protein levels. Intriguingly, homozygous *GRN* mutations cause neuronal ceroid lipofuscinosis, focusing investigators on the lysosome as a site of molecular dysfunction in *GRN*-related FTD. Progranulin is a growth factor that binds to tumor necrosis factor (TNF) and sortilin receptors and participates in tissue repair and tumor growth. How progranulin mutations lead to FTD remains unknown, but the most likely mechanisms include lysosomal dysfunction and neuroinflammation. Often, *MAPT* and *GRN* mutations are associated with parkinsonian features, whereas ALS is rare. Infrequently, mutations in the valosin-containing protein (*VCP*, chromosome 9), TANK binding kinase 1 (TBK-1), T cell–restricted intracellular antigen-1 (TIA1), and charged multivesicular body protein 2b (*CHMP2b,* chromosome 3) genes also lead to autosomal dominant familial FTD. Mutations in the *TARDBP* (encoding TDP-43) and *FUS* (encoding fused in sarcoma [FUS]) genes (see below) cause familial ALS, sometimes in association with an FTD syndrome, although a few patients presenting with FTD alone have been reported.

NEUROPATHOLOGY

The pathological hallmark of FTLD is a focal atrophy of frontal, insular, and/or temporal cortex, which can be visualized with neuroimaging studies (Fig. 443-1) and is often profound at autopsy. Neuroimaging studies suggest that atrophy often begins focally in one hemisphere before spreading to anatomically interconnected cortical and subcortical regions. Loss of cortical serotonergic innervation is seen in many patients. In contrast to AD, the cholinergic system is relatively spared in FTD, which accounts for the poor efficacy of acetylcholinesterase inhibitors in this group.

Although early studies suggested that 15–30% of patients with FTD showed underlying AD at autopsy, progressive refinement in clinical diagnosis has improved prediction accuracy, and most patients diagnosed with FTD at a dementia clinic will show underlying FTLD pathology. Microscopic findings seen across all patients with FTLD include gliosis, microvacuolation, and neuronal loss, but the disease is subtyped according to the protein composition of neuronal and glial inclusions, which contain either tau or TDP-43 in ~90% of patients, with the remaining ~10% showing inclusions containing the FET family of proteins (**F**US, **E**wing sarcoma protein, **T**AF-15) **(Fig. 443-2)**.

PATHOGENESIS

In FTLD-tau, the toxicity and spreading capacity of misfolded tau are critical for the pathogenesis of inherited and sporadic tauopathies, although loss of tau microtubule stabilizing function may also play a role. In recent years, the distinctive structures of the misfolded tau in each FTLD tauopathy have been resolved using cryo-electron microscopy, opening up new approaches to diagnosis and treatment. TDP-43 and FET family proteins in contrast, are RNA/DNA binding proteins whose roles in neuronal function are still being actively investigated. TDP-43 is a master regulator of gene expression, and loss of TDP-43 function results in mis-splicing events leading to mRNA degradation (via nonsense-mediated decay) or aberrant transcripts that give rise to stable but dysfunctional peptides. One key role of TDP-43 and FET family proteins may be the chaperoning of mRNAs to the distal neuron for activity-dependent translation within dendritic spines. Because these proteins also form intracellular aggregates and produce similar anatomic progression, protein toxicity and spreading may also factor heavily

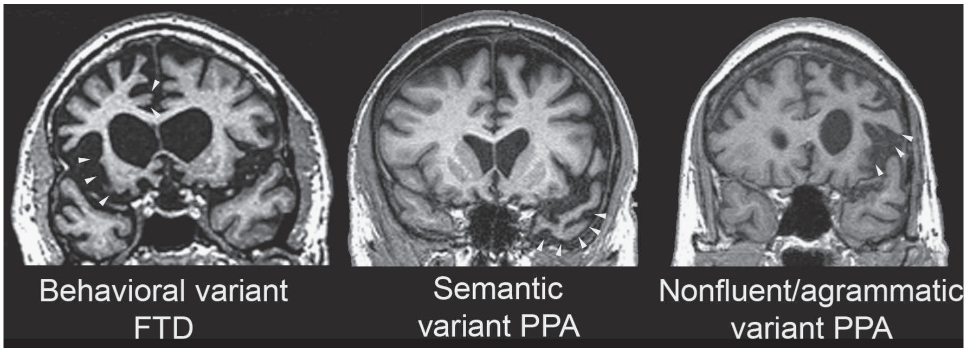

FIGURE 443-1 Three major frontotemporal dementia (FTD) clinical syndromes. Coronal magnetic resonance imaging sections from representative patients with behavioral variant FTD (*left*) and the semantic (*center*) and nonfluent/agrammatic (*right*) variants of primary progressive aphasia (PPA). Areas of early and severe atrophy in each syndrome are highlighted (*white arrowheads*). The behavioral variant features anterior cingulate and frontoinsular atrophy, spreading to orbital and dorsolateral prefrontal cortex. Semantic variant PPA shows prominent temporopolar atrophy, more often on the left. Nonfluent/agrammatic variant PPA is associated with dominant frontal opercular and dorsal insula degeneration.

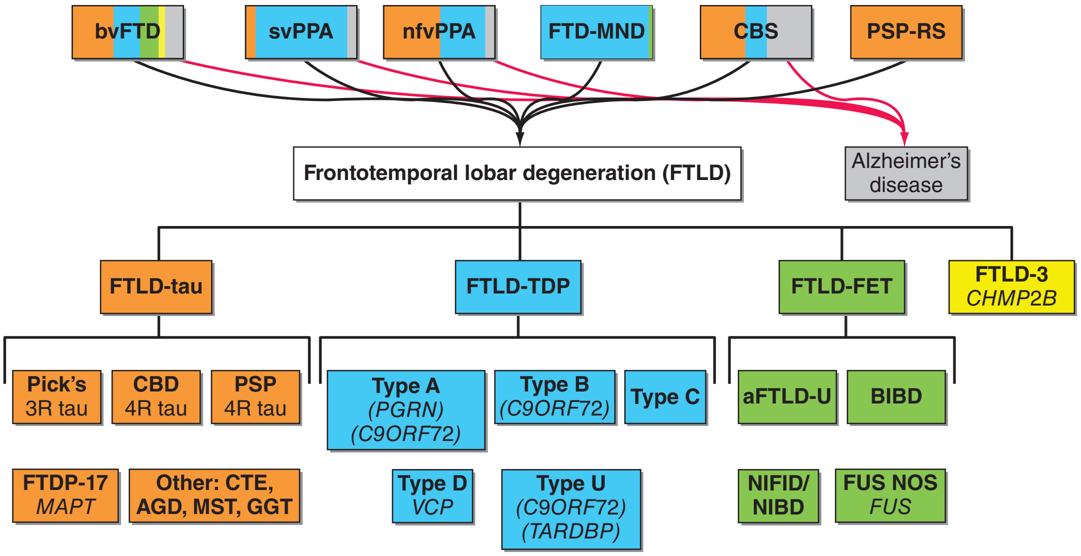

FIGURE 443-2 Frontotemporal dementia syndromes are united by underlying frontotemporal lobar degeneration pathology, which can be divided according to the presence of tau, TDP-43, or FUS-containing inclusions in neurons and glia. Correlations between clinical syndromes and major molecular classes are shown with colored shading. Despite improvements in clinical syndromic diagnosis, a small percentage of patients with some frontotemporal dementia syndromes will show Alzheimer's disease neuropathology at autopsy (*gray shading*). aFTLD-U, atypical frontotemporal lobar degeneration with ubiquitin-positive inclusions; AGD, argyrophilic grain disease; BIBD, basophilic inclusion body disease; bvFTD, behavioral variant frontotemporal dementia; CBD, corticobasal degeneration; CBS, corticobasal syndrome; CTE, chronic traumatic encephalopathy; FET, FUS, Ewing sarcoma protein, TAF-15 family of proteins; FTD-MND, frontotemporal dementia with motor neuron disease; FTDP-17, frontotemporal dementia with parkinsonism linked to chromosome 17; FUS, fused in sarcoma; GGT, globular glial tauopathy; MST, multisystem tauopathy; nfvPPA, nonfluent/agrammatic variant primary progressive aphasia; NIBD, neurofilament inclusion body disease; NIFID, neuronal intermediate filament inclusion disease; PSP, progressive supranuclear palsy; PSP-RS, progressive supranuclear palsy–Richardson syndrome; svPPA, semantic variant primary progressive aphasia; Type U, unclassifiable type.

in the pathogenesis of FTLD-TDP and FTLD-FET. As with tau, the ultrastructural characteristics of the TDP-43 and FET family protein misfolding events are now being actively characterized, with each pathologically recognized morphological subtype corresponding to a disease-specific fold.

Increasingly, misfolded proteins in neurodegenerative disease are recognized as having "prion-like" or "corruptive" properties in that they can template the misfolding of their natively folded (or unfolded) protein counterparts, a process that creates exponential amplification of protein misfolding within a cell and may promote transcellular and even transsynaptic protein propagation between cells. This hypothesis could provide a unifying explanation for the stereotypical and network-rooted patterns of disease spread observed in each syndrome **(Chap. 435)**.

Although the term *Pick's disease* was once used to describe a progressive degenerative disorder characterized by selective involvement of the anterior frontal and temporal neocortex and pathologically by intraneuronal cytoplasmic inclusions (*Pick bodies*), it is now used only in reference to a specific FTLD-tau histopathologic subtype. Classical Pick bodies are argyrophilic, staining positively with the Bielschowsky silver method (but not with the Gallyas method) and also with immunostaining for hyperphosphorylated tau. Recognition of the three FTLD major molecular classes has allowed delineation of distinct FTLD subtypes within each class. These subtypes, based on the morphology and distribution of the neuronal and glial inclusions **(Fig. 443-3)**, account for the vast majority of patients, and some subtypes show strong clinical or genetic associations (Fig. 443-2). Despite this progress, clinical features do not allow reliable prediction of the underlying FTLD subtype, or even the major molecular class, for all clinical syndromes. Molecular positron emission tomography (PET) imaging with ligands chosen to bind misfolded tau protein shows promise, but to date, these ligands only show robust and specific binding to AD-related misfolded tau. Because FTLD-tau and FTLD-TDP account for 90% of FTLD patients, the ability to detect pathologic tau (or TDP-43) protein deposition in vivo would greatly improve prediction accuracy, especially when amyloid PET imaging is negative.

TREATMENT

Caregivers for patients with FTD carry a heavy burden, especially when the illness disrupts core emotional and personality functions of the loved one. Treatment is symptomatic, and there are currently no therapies known to slow progression or improve symptoms. Many of the behaviors that may accompany FTD, such as depression, hyperorality, compulsions, and irritability, can be ameliorated with antidepressants, especially selective serotonin reuptake inhibitors (SSRIs). Because FTD is often accompanied by parkinsonism, antipsychotics, which can exacerbate this problem, must be used with caution. Experimental therapeutics, most targeting genetic forms of FTD, have just begun to enter clinical trials, but to date, no disease-modifying treatments have shown efficacy. **A general approach to the symptomatic management of dementia is presented in Chap. 31.**

PROGRESSIVE SUPRANUCLEAR PALSY SYNDROME

PSP-RS is a degenerative disorder that involves the brainstem, basal ganglia, diencephalon, and selected areas of cortex. Clinically, PSP-RS begins with falls and executive dysfunction or subtle personality changes (such as mental rigidity, impulsivity, or apathy). Shortly thereafter, a progressive oculomotor syndrome ensues that begins with square wave jerks, followed by slowed saccades (vertical worse than horizontal) before resulting in progressive supranuclear ophthalmoparesis. Dysarthria, dysphagia, and symmetric axial rigidity can be prominent features that emerge at any point in the illness. A stiff, unstable posture with hyperextension of the neck and a slow, jerky, toppling gait are characteristic. Frequent unexplained and sometimes spectacular falls are common secondary to a combination of axial rigidity, inability to look down, and impaired judgment. Even once patients have severely limited voluntary eye movements, they retain oculocephalic reflexes (demonstrated using a vertical doll's head maneuver); thus, the oculomotor disorder is supranuclear. The dementia overlaps with bvFTD, featuring apathy, frontal-executive dysfunction, poor judgment, slowed thought processes, impaired verbal fluency, and difficulty with sequential actions and shifting from one task to another. These features are common at presentation and often precede the motor syndrome. Some patients with a pathologic diagnosis of PSP begin with a nonfluent aphasia or motor speech disorder and progress to classical PSP-RS. Response to L-dopa is limited or absent; no other treatments exist. Death occurs within 5–10 years of onset. Like Pick's disease, increasingly the term *PSP* is used to refer to a specific histopathologic entity within the FTLD-tau class. In PSP, accumulation of hyperphosphorylated 4-repeat tau is seen within neurons and glia. Tau neuronal inclusions often appear tangle-like and may be large, spherical ("globose"), and coarse in subcortical and brainstem

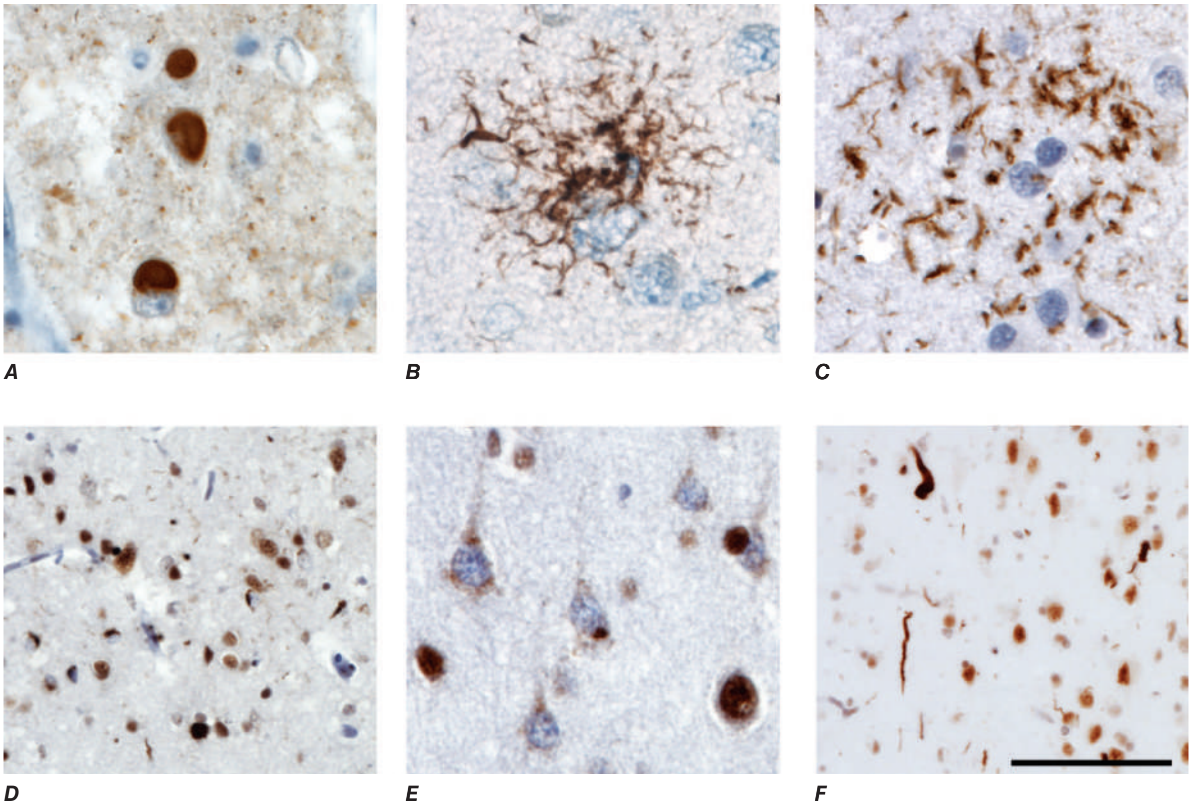

FIGURE 443-3 Neuropathology in frontotemporal lobar degeneration (FTLD). FTLD-tau (***A–C***) and FTLD-TDP (***D–F***) account for >90% of patients with FTLD, and immunohistochemistry reveals characteristic lesions in each of the major histopathologic subtypes within each class: ***A.*** Pick bodies in Pick's disease; ***B.*** a tufted astrocyte in progressive supranuclear palsy; ***C.*** an astrocytic plaque in corticobasal degeneration; ***D.*** small compact or crescentic neuronal cytoplasmic inclusions and short, thin neuropil threads in FTLD-TDP, type A; ***E.*** diffuse/granular neuronal cytoplasmic inclusions (with a relative paucity of neuropil threads) in FTLD-TDP, type B; and ***F.*** long, tortuous dystrophic neurites in FTLD-TDP, type C. TDP can be seen within the nucleus in neurons lacking inclusions but mislocalizes to the cytoplasm and forms inclusions in FTLD-TDP. Immunostains are 3-repeat tau (***A***), phospho-tau (***B*** and ***C***), and TDP-43 (***D–F***). Sections are counterstained with hematoxylin. Scale bar applies to all panels and represents 50 μm in ***A***, ***B***, ***C***, and ***E*** and 100 μm in ***D*** and ***F***.

structures. The most prominent involvement is in the subthalamic nucleus, globus pallidus, substantia nigra, periaqueductal gray, tectum, oculomotor nuclei, pontine nuclei, and dentate nucleus of cerebellum. Neocortical tangle-like inclusions, like those in AD, often take on a more flame-shaped morphology, but the tau folds in AD and PSP are distinct. Prominent tau-positive glial inclusions are an essential feature of PSP. Tufted astrocytes in the neocortex or striatum are the signature lesion (Fig. 443-3). Coiled oligodendroglial inclusions ("coiled bodies") are common but nonspecific. Most patients with PSP-RS show PSP at autopsy, although small numbers will show another tauopathy (corticobasal degeneration [CBD] or globular glial tauopathy, or FTLD with a *MAPT* mutation; Fig. 443-2).

In addition to its overlap with FTD and CBS (see below), PSP is often confused with idiopathic *Parkinson's disease* (PD). Although elderly patients with PD may have restricted upgaze, they do not develop downgaze paresis or other abnormalities of voluntary eye movements typical of PSP. Dementia ultimately occurs in most patients with PD, often due to the emergence of a full-blown dementia with Lewy bodies (DLB)-like syndrome or comorbid AD-type dementia. Furthermore, the behavioral syndromes seen with DLB differ from PSP (see below). Dementia in PD becomes more likely with increasing age, increasing severity of extrapyramidal signs, long disease duration, and the presence of depression. Patients with PD who develop dementia also show cortical atrophy on brain imaging. Neuropathologically, there may be AD-related changes in the cortex or Lewy body disease (LBD)-related α-synuclein inclusions in both the limbic system and cerebral cortex. **DLB and PD are discussed in Chaps. 445 and 446, respectively.**

■ CORTICOBASAL SYNDROME

CBS is a slowly progressive dementia-movement disorder associated with severe degeneration in the perirolandic cortex and basal ganglia (substantia nigra and striatopallidum). Patients typically present with asymmetric rigidity, dystonia, myoclonus, and apraxia that render a progressively incapacitated limb, at times associated with *alien limb* phenomena in which the limb exhibits unintended motor actions such as grasping, groping, drifting, or undoing. Eventually CBS becomes bilateral and leads to dysarthria, slow gait, action tremor, and a frontal-predominant dementia. Whereas CBS refers to the clinical syndrome, CBD refers to a specific histopathologic FTLD-tau entity (Fig. 443-2). Although CBS was once thought to be pathognomonic for CBD, increasingly it has been recognized that CBS can be due to CBD, PSP, FTLD-TDP, and AD, with the latter accounting for up to 30% of CBS in some series. In CBD, the microscopic features include ballooned, achromatic, tau-positive neurons; astrocytic plaques (Fig. 443-3); and other dystrophic glial tau pathomorphologies that overlap with those seen in PSP. Most specifically, CBD features a severe tauopathy burden in the subcortical white matter, consisting of axonal threads and oligodendroglial coiled bodies. As shown in Fig. 443-2, patients with bvFTD, nonfluent/agrammatic PPA, and PSP-RS may also show CBD at autopsy, emphasizing the importance of distinguishing clinical and pathologic constructs and terminology. Treatment of CBS remains symptomatic; no disease-modifying therapies are available.

■ FURTHER READING

Boeve BF et al: Advances and controversies in frontotemporal dementia: Diagnosis, biomarkers, and therapeutic considerations. Lancet Neurol 21:258, 2022.

Creekmore BC et al: Neurodegenerative disease tauopathies. Annu Rev Pathol 19:345, 2024.

Irwin DJ et al: Frontotemporal lobar degeneration: Defining phenotypic diversity through personalized medicine. Acta Neuropathol 129:469, 2015.

Roberson ED: Mouse models of frontotemporal dementia. Ann Neurol 72:837, 2012.

Seeley WW: Behavioral variant frontotemporal dementia. Continuum (Minneap Minn) 25:76, 2019.

444 Vascular Dementia

Steven M. Greenberg, William W. Seeley

The term *vascular dementia* has traditionally been used to describe a subset of dementia cases due primarily to one or more symptomatic strokes. Considered as such, vascular dementia is usually ranked the second most frequent cause of dementia, exceeded only by Alzheimer's disease **(Chap. 442)**, and is especially common in populations with limited access to medical care, where vascular risk factors are undertreated. More recently, this relatively narrow definition of vascular dementia has been substantially broadened to encompass the full impact of cerebrovascular disease on age-related cognitive decline. The term *vascular contributions to cognitive impairment and dementia* (VCID) reflects the observation that pathologic changes involving the cerebral vasculature are highly prevalent in the elderly and contribute to cognitive impairment, whether occurring in isolation or—more commonly—in conjunction with other neurodegenerative processes. The concept of VCID is one facet of the contemporary understanding of age-related cognitive decline as due to cumulative effects of distinct and overlapping neuropathologic changes. Multifactorial or "mixed" dementias appear to be more prevalent than single-etiology dementias and thus represent the rule rather than the exception.

Symptomatic stroke and asymptomatic vascular lesions, most commonly detected with brain magnetic resonance imaging (MRI) scans, both contribute importantly to cognitive impairment. At least some cognitive impairment is present in approximately half of stroke survivors and progressively increases with longer periods of follow-up. Population-based studies also demonstrate substantially increased risk of cognitive impairment among individuals without symptomatic stroke but with MRI evidence of cerebrovascular disease. The high risk for subsequent cognitive impairment or dementia conferred by MRI markers of otherwise silent vascular brain injury highlights the cumulative impact of small distributed brain injuries—often associated with small-vessel brain disease—on compromising brain function. Further support for this framework comes from the correlation of cognitive performance during life with postmortem neuropathology. Analysis of large community-based samples demonstrates independent contributions to cognitive dysfunction and decline from both grossly visible infarcts and pathologic markers of overall cerebrovascular disease severity such as atherosclerosis, arteriolosclerosis, and cerebral amyloid angiopathy scores. The Religious Orders Study and Memory and Aging Project analysis of 1079 community-based participants, for example, found each of these cerebrovascular entities to be moderate to severe in >30% of postmortem brains and, when present, to each account for ~20% of an individual's premortem cognitive decline.

Recent epidemiologic evidence of a decline in age-adjusted dementia incidence hints at the potential public impact of improving vascular health. The population-based Framingham Study reported 5-year age- and sex-adjusted cumulative hazard rates for dementia of 3.6 per 100 persons during the late 1970s to early 1980s, 2.8 in the late 1980s to early 1990s, 2.2 in the late 1990s to early 2000s, and 2.0 in the late 2000s to early 2010s. These time intervals coincide with parallel trends in hypertension control and stroke prevention, though the associations do not prove causation. Evidence supporting a potential causative effect of blood pressure control came from the SPRINT-MIND trial targeting systolic blood pressure (SBP) of <120 mmHg versus 140 mmHg in hypertensive individuals aged ≥50 years. The study ended prematurely because of effective prevention of cardiovascular outcomes in the lower SBP target group but nonetheless demonstrated that SBP reduction reduced rates of mild cognitive impairment (hazard ratio [HR], 0.81; 95% confidence interval [CI], 0.69–0.95) and combined mild cognitive impairment or probable dementia (HR, 0.85; 95% CI, 0.74–0.97), although not dementia alone (HR, 0.83; 95% CI, 0.67–1.04). It is notable that both these studies measured all-cause cognitive impairment rather than just a vascular dementia subset, underlining the potential importance of VCID as a target for dementia prevention.

GLOBAL CONSIDERATIONS

A review of data from across the globe indicates good evidence for variability in vascular dementia. Intracranial atherosclerosis, for example, is higher in Asians, Hispanics, and American blacks than it is in European and American whites, while whites may have more extracranial disease. The causes of these disparities remain under investigation but likely include access to health care, lifestyle factors such as diet, and possible genetic influences.

SUBTYPES OF CEREBROVASCULAR DISEASE ASSOCIATED WITH VCID

Large Cerebral Strokes Symptomatic strokes, whether ischemic **(Chap. 438)** or hemorrhagic **(Chap. 439)**, reflect irreversible injury to discrete areas of cerebral cortex, subcortical white matter, or other subcortical and infratentorial structures and produce cognitive impairment as a function of their size and location. Rare individual infarcts in specific strategic locations such as thalamus, medial temporal cortex, anterior corpus callosum, or dominant-side angular gyrus can sufficiently impair episodic memory and functional skills to meet memory-based criteria for dementia. More commonly, strokes occur outside these strategic territories and affect various other aspects of cognition such as executive function, processing speed, and visuospatial performance that fall under the broader VCID concept. Multiple strokes and larger volumes of infarcted territory are associated with a higher likelihood of poststroke cognitive dysfunction.

Stroke patients who make good cognitive recovery nonetheless demonstrate accelerated poststroke cognitive decline. Community-based individuals in the longitudinal Reasons for Geographic and Racial Differences in Stroke study, for example, changed trajectory from an average prestroke cognitive gain of 0.021 points/year to poststroke cognitive loss of –0.035 points/year on the six-item screener global cognitive function scale. Mechanisms for poststroke cognitive decline likely include ongoing effects of the cerebrovascular disease that gave rise to the index stroke as well as loss of cognitive reserve that makes the brain less resilient to any additional age-related disorders.

Cerebral Small-Vessel Disease Diseases of the brain's small vessels **(Chap. 438)** can also cause symptomatic ischemic or hemorrhagic stroke but are more often clinically asymptomatic and recognized only during evaluation for cognitive decline or other symptoms. The two common age-related cerebral small-vessel pathologies are arteriolosclerosis and cerebral amyloid angiopathy. *Arteriolosclerosis* represents thickening of arterioles due to infiltration of plasma proteins into the vessel wall. The primary risk factors for this process are age, hypertension, and diabetes mellitus. Cerebrovascular arteriolosclerosis can present as a cause of ischemic or hemorrhagic symptomatic stroke, both most commonly centered in territories supplied by deep penetrating vessels such as thalamus, basal ganglia, or brainstem. *Cerebral amyloid angiopathy* is defined by deposition of the β-amyloid peptide in the walls of small cerebral arteries, arterioles, and capillaries, with consequent loss of normal wall structure. Its primary risk factor is advancing age. Cerebral amyloid angiopathy is most often recognized symptomatically as a cause of intracerebral hemorrhage **(Chap. 439)**, commonly located in cerebral cortex, subcortical white matter (collectively known as lobar hemorrhages), or the cerebral convexity subarachnoid space. This small-vessel pathology also appears to confer increased risk for the adverse amyloid-related imaging abnormalities (ARIA) associated with recently approved immunotherapies for Alzheimer's disease **(Chap. 442)**. The distinction between the deep penetrating territories most commonly affected by arteriolosclerosis and superficial lobar brain regions affected by cerebral amyloid angiopathy often allows the two small-vessel diseases to be radiographically distinguished.

Despite differences in their underlying pathogenic mechanisms, the two cerebral small-vessel diseases produce a similar range of ischemic and hemorrhagic brain lesions detectable by histopathology at autopsy or MRI scan during life **(Fig. 444-1)**. *Small (lacunar) infarcts* are a common feature of arteriolosclerosis and less commonly of cerebral amyloid angiopathy. Chronic lacunar infarcts can appear on MRI fluid-attenuated inversion recovery (FLAIR) sequences as a hyperintense rim surrounding a hypointense cavitated core with diameters typically

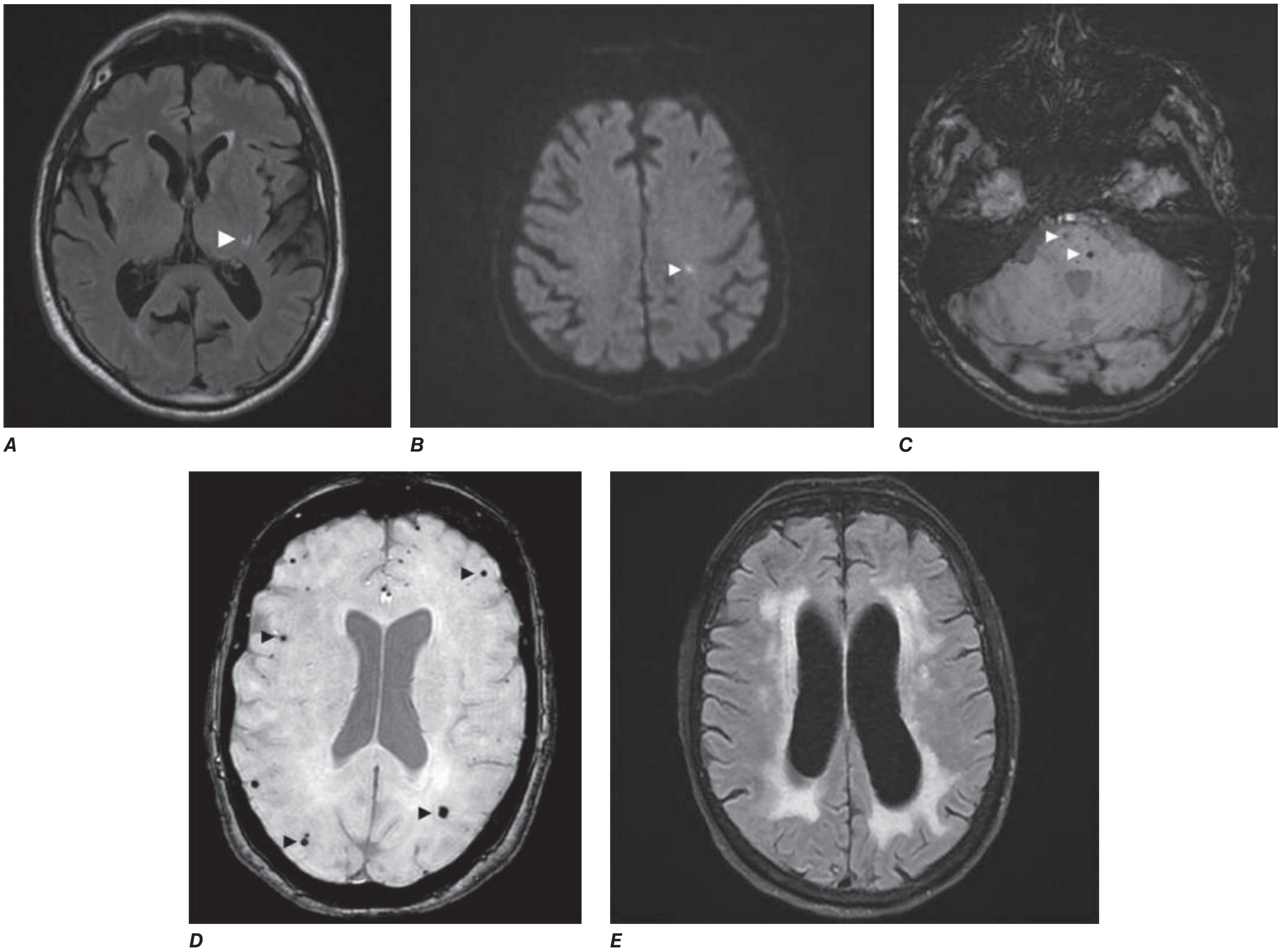

FIGURE 444-1 Magnetic resonance imaging (MRI) markers of cerebral small vessel disease. ***A.*** Lacunar infarct: fluid-attenuated inversion recovery (FLAIR) sequence showing hyperintense rim surrounding a hypointense cavitated core in the left thalamus (arrowhead). ***B.*** Acute microinfarct: diffusion-weighted sequence showing small hyperintense lesion in the left centrum semiovale (*arrowhead*). ***C.*** Cerebral microbleeds in deep penetrating brain region: $T2^*$-weighted sequence showing multiple small hypointense lesions in the pons (*arrowheads*). ***D.*** Cerebral microbleeds in lobar brain regions: $T2^*$-weighted sequence showing multiple small hypointense lesions lobar brain regions (*arrowheads*). ***E.*** White matter hyperintensities: FLAIR sequence showing confluent diffuse hyperintensities in white matter.

3–15 mm **(Fig. 444-1*A*)**, but this characteristic appearance evolves in only a subset of small infarctions, and many cannot be readily identified in the chronic stage. *Microinfarcts* <3 mm are characteristic of both small-vessel diseases. They are substantially more numerous than lacunar infarcts but less easily visualized. Acute microinfarcts may be visible as punctate hyperintensities on diffusion-weighted MRI images **(Fig. 433-1*B*)**, whereas a small subset of chronic microinfarcts is detectable on high-resolution T2-weighted MRI sequences as hyperintense lesions in the cerebral cortex. *Cerebral microbleeds* are less numerous than lacunes or microinfarcts but readily detected in their chronic stage because of the paramagnetic effects of iron products. These appear as round hypointense lesions on $T2^*$-weighted MRI, primarily in deep penetrating brain regions if caused by arteriolosclerosis **(Fig. 444-1*C*)** or lobar regions if caused by cerebral amyloid angiopathy **(Fig. 444-1*D*)**.

Other MRI markers of small-vessel disease identify diffuse injury of the white matter. *White matter hyperintensities* on T2-weighted or FLAIR MRI sequences **(Fig. 444-1*E*)** are an almost ubiquitous feature of aging. Although these lesions are readily visible on clinical MRI, they represent a nonspecific marker of white matter gliosis, demyelination, or increased water content. Extremely severe diffuse white matter vascular injury is commonly referred to as *Binswanger's disease* or subcortical arteriosclerotic encephalopathy, recognized as a clinical syndrome with gradual cognitive deterioration and notable white matter changes of small-vessel ischemic disease. On neuroimaging, a progressive confluent subcortical and periventricular white matter disease is seen **(see Fig. 31-2)**, with hypoperfusion and hypometabolism. More subtle alterations in white matter structure can be sensitively and quantitatively detected by *diffusion tensor* MRI **(Chap. 434)** as increased water diffusivity or decreased diffusion directionality. Diffusion tensor measures of white matter structural integrity show a consistent association with cognitive performance and gait speed, reflecting the central role of disconnection of key brain networks in mediating the effects of cerebral small-vessel disease. These diffusion tensor–based methods often require complex processing and are typically used in research rather than clinical settings. A relatively simple diffusion tensor–based metric defined by the peak width of the skeletonized mean diffusivity (PSMD) histogram has emerged as a candidate method for quantifying white matter disconnection. Functional MRI measurement of cerebrovascular reactivity to physiologic stimuli is generally not performed in clinical practice, but it may become abnormal decades before appearance of structural brain injury and therefore represents a promising outcome marker for identifying disease-modifying interventions aimed to slow or prevent vascular brain injury.

Role of Accompanying Brain Pathologies The concept of VCID posits that large strokes and small-vessel disease often occur in combination with neurodegenerative brain diseases, most commonly Alzheimer's disease **(Chap. 442)**. Many clinicopathologic correlation studies have established that the co-occurrence of cerebrovascular and neurodegenerative lesions produces more cognitive and functional impairment than expected from the effects of each disease mechanism considered independently. Interactions between cerebrovascular and neurodegenerative processes may also contribute to dementia. Such interactions might involve loss of blood-brain barrier integrity (possibly allowing brain penetration of neurotoxic or inflammatory agents) and impaired clearance of β-amyloid or other pathogenic molecules from the brain (postulated to occur along perivascular drainage pathways driven by physiologic vascular motion).

APPROACH TO THE PATIENT

Vascular Dementia

Identifying vascular contributors to a patient's cognitive impairment can clarify the etiologic diagnosis and point to specific interventions aimed at slowing progression. Clinical evaluation is focused on identifying vascular risk factors (hypertension, diabetes mellitus, dyslipidemia, tobacco use, atrial fibrillation, coronary artery disease, or peripheral vascular disease), history of prior symptoms of stroke or transient ischemic attack, and family history of early stroke or vascular disease. Although stepwise progression and certain cognitive deficits such as loss of executive function are particularly suggestive, most individuals with VCID follow the more typical pattern of gradual progression of impaired episodic memory.

The mainstay for detection and subtyping of cerebrovascular disease is brain MRI. The MRI should include FLAIR, diffusion-weighed, and T2*-weighted sequences to detect the range of lesions noted above: large and small chronic infarcts, acute microinfarcts, microbleeds, and white matter hyperintensities. Vessel imaging studies such as computed tomography or magnetic resonance angiography are not required for initial evaluation of cognitive impairment, though they may be useful for determining the cause of any macroscopic infarcts that are identified. Genetic testing for rare hereditary forms of VCID such as cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) **(Chap. 438)** or hereditary cerebral amyloid angiopathy can be considered for cases in which there is a particularly young onset, positive family history, or suggestive neuroimaging, but is otherwise unnecessary.

TREATMENT

Vascular Dementia

Very few trials have addressed the optimal treatment for individuals with asymptomatic large- or small-vessel cerebrovascular disease, leaving uncertainty as to whether to follow primary or secondary stroke prevention guidelines. At a minimum, treatment should assiduously follow primary stroke prevention guidelines. The American Heart Association recommends the prudent approach for vascular health of managing blood pressure, controlling cholesterol, reducing blood sugar, maintaining an active lifestyle, adhering to a heart-healthy diet, losing weight, discontinuing tobacco, and getting healthy sleep (Life's Essential 8, *https://www.heart.org/en/healthy-living/healthy-lifestyle/lifes-essential-8*). Blood pressure targets are <140/90 mmHg for all individuals and <130/80 mmHg for those with estimated 10-year cardiovascular disease risk ≥10%, which likely applies to many individuals with imaging evidence of asymptomatic brain infarcts or advanced small-vessel disease. The usefulness of other treatments for secondary stroke prevention such as antiplatelet or statin therapy has not been established for asymptomatic infarcts. These agents are reasonable to consider, however, when the imaging appearance suggests embolic or large-vessel-related strokes. All individuals with asymptomatic infarcts should be screened for atrial fibrillation, and those with embolic-appearing infarcts can be considered for prolonged cardiac monitoring. Similarly, patients with infarcts in the territories of large arteries should be considered for vascular imaging.

The few trials of symptomatic medications for cognitive impairment due to vascular etiologies have suggested modest cognitive benefits comparable to those found in Alzheimer's disease patients. Therefore, it may be reasonable in VCID to consider agents such as the cholinesterase inhibitors donepezil, rivastigmine, or galantamine for mild to moderate cognitive impairment and high-dose donepezil or the *N*-methyl-D-aspartate receptor antagonist memantine for moderate to severe impairment **(Chap. 442)**. A shared decision-making approach in considering these medications is useful, given their relatively small impact on daily function.

FURTHER READING

Boyle PA et al: Person-specific contribution of neuropathologies to cognitive loss in old age. Ann Neurol 83:74, 2018.

Corriveau RA et al: The science of vascular contributions to cognitive impairment and dementia (VCID): A framework for advancing research priorities in the cerebrovascular biology of cognitive decline. Cell Mol Neurobiol 36:281, 2016.

Dichgans M, Leys D: Vascular cognitive impairment. Circ Res 120:573, 2017.

Düering M et al: Neuroimaging standards for research into small vessel disease-advances since 2013. Lancet Neurol 22:602, 2023.

Greenberg SM et al: Cerebral amyloid angiopathy and Alzheimer disease: One peptide, two pathways. Nat Rev Neurol 16:30, 2020.

Levine DA et al: Trajectory of cognitive decline after incident stroke. JAMA 314:41, 2015.

Smith EE et al: Prevention of stroke in patients with silent cerebrovascular disease: A scientific statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 48:e44, 2017.

Snowdon DA et al: Brain infarction and the clinical expression of Alzheimer disease. The Nun Study. JAMA 277:813, 1997.

Vermeer SE et al: Silent brain infarcts and the risk of dementia and cognitive decline. N Engl J Med 348:1215, 2003.

445

Dementia with Lewy Bodies

Irene Litvan, William W. Seeley, Bruce L. Miller

Lewy body disease (LBD), manifesting as Parkinson's disease dementia (PDD) or dementia with Lewy bodies (DLB), is the second most common cause of neurodegenerative dementia, after Alzheimer's disease (AD) **(Chap. 442)**. Approximately 10% of patients with Parkinson's disease (PD) develop PDD per year, with the majority of PD patients developing PDD over time. The incidence of DLB is ~7 per 100,000 person-years. The prevalence of both PDD and DLB increases with aging, and both affect men more often than women. The development of increasingly useful biomarkers for PD and DLB is making possible new operational definitions, classifications, and staging for these disorders, and these are likely to continue to evolve over time.

CLINICAL MANIFESTATIONS

Most investigators conceptualize PDD and DLB as points on a spectrum of LBD pathology. Cognitively, PDD and DLB usually manifest with severe executive, attentional, and visuospatial deficits but preserved episodic memory. Cognitive decline in LBD affects performance of daily living activities beyond other PD symptoms. Early psychosis including well-formed visual hallucinations, fluctuating cognition, rapid eye movement sleep behavior disorder (RBD), and parkinsonism are the main diagnostic features in DLB. The sense of a presence behind the person may precede well-formed hallucinations. Delusions are less frequent than hallucinations and are usually related to misidentification, infidelity, theft, or persecution. Fluctuating attention and concentration are other characteristic features. Minor day-to-day variation in cognitive functioning is common across dementias, but in DLB, these fluctuations can be marked, with short periods of confusion or severe lethargy that may rapidly resolve. Patients with PDD and DLB are highly sensitive to infectious or metabolic disturbances. The first manifestation of DLB in some patients is delirium, often precipitated by an infection, new medicine, or other systemic disturbance. Parkinsonism in DLB is usually associated with early postural instability and can present early or later in the course. RBD is a characteristic, often prodromal, feature. Normally, dreaming is accompanied by skeletal muscle paralysis, but patients with RBD enact dreams, often violently, leading to injuries to themselves or their bed partners. Both PDD and DLB may be accompanied or preceded by anosmia, constipation, RBD, depression, and anxiety.

The symptom profile in DLB and PDD can provide clues for the differential diagnosis at the clinic. Clinically, the time interval between parkinsonism and dementia differentiates PDD and DLB. PDD presents in patients with long-standing PD, who manifest dementia often with visual hallucinations, fluctuating attention or alertness, and RBD. On the other hand, when the dementia and the neuropsychiatric symptoms precede or co-emerge with the parkinsonism, the patient is diagnosed with DLB. Patients with DLB, more frequently than those with PDD, also have AD co-pathology, making the prediction of underlying pathology challenging for clinicians. Episodic memory disturbance points to the diagnosis of comorbid AD. Orthostatic hypotension that can lead to syncopal events, erectile dysfunction, and constipation can be present early in DLB, at times making it challenging to differentiate DLB from multiple system atrophy (MSA). In MSA, the autonomic disturbances occur early and are usually more severe than in DLB, and cognition is relatively preserved. Anosmia is also more characteristic of LBD than MSA. Skin biopsy and serum with biomarkers for α-synuclein oligomers, a major component of Lewy bodies, have shown potential for differentiating PD from MSA, and if validated for clinical use, this type of test may also differentiate DLB or PDD from MSA in the future.

■ PRODROMAL PHASE

Both DLB and PDD have a prodromal phase where patients have a mild cognitive impairment (MCI), with cognitive deficits that do not have a substantial impact on daily life. PD-MCI is characterized by deficits in executive, attention, and visuospatial disturbances, but can also present with an amnestic or multiple-domain MCI. Prodromal DLB is characterized by similar cognitive disturbances but is also associated with either hallucinations unrelated to medications, RBD, fluctuations in attention, or parkinsonism. It is at times challenging to differentiate prodromal MCI-DLB and PD-MCI when the major features are RBD and parkinsonism, for which the term *prodromal MCI–Lewy body* (MCI-LB) was recently proposed. RBD may precede the development of an LBD-related syndrome by many years, usually evolving into either PD or DLB. The clinical profile and several biomarkers can help differentiate MCI due to LBD versus AD pathology **(Table 445-1)**.

TABLE 445-1 Distinguishing MCI Due to Lewy Body Disease or Alzheimer's Disease

CLINICAL FEATURES	PRODROMAL MCI-LB PATHOLOGY	PRODROMAL MCI-AD PATHOLOGY
MCI	MCI usually affecting executive, attention, and/or visuospatial functions	MCI with impaired memory and semantic naming
Fluctuating cognition with variations in attention	Frequent and severe	Rare or not severe
Sleep	REM sleep behavior disorder	Insomnia, frequent awakenings
Recurrent visual hallucinations	Frequent	Rare
Biomarkers		
Polysomnogram	REM sleep behavior disorder without atonia	Normal
CSF	Decreased CSF α-synuclein by RT-QuIC	Decreased CSF β-amyloid and increased phospho-tau. This can now be performed in blood.
MRI	Atrophy of the amygdala	Atrophy of the parahippocampal/hippocampal areas
^{18}F-deoxyglucose PET scan	Hypometabolism in occipital lobe and increased in posterior cingulate (cingulate island sign)	Hypometabolism in parietotemporal lobes
Amyloid PET scan	Normal, unless associated with AD	Abnormal parietotemporal areas
MIBG myocardial scintigraphy	Postganglionic sympathetic denervation	Normal
DAT scan or PET dopamine scan	Reduced dopamine transporter in the basal ganglia, particularly putamen	Normal

Abbreviations: AD, Alzheimer's disease; CSF, cerebrospinal fluid; DAT, dopamine transporter; LB, Lewy bodies; MCI, mild cognitive impairment; MIBG, meta-iodobenzylguanidine; MRI, magnetic resonance imaging; PET, positron emission tomography; REM, rapid eye movement; RT-QuIC, real-time quaking-induced conversion.

PATHOLOGY

The key neuropathologic feature in LBD is the presence of Lewy bodies and Lewy neurites throughout specific brainstem nuclei, substantia nigra, amygdala, cingulate gyrus, and, ultimately, the neocortex. Lewy bodies are intraneuronal cytoplasmic inclusions that stain with periodic acid–Schiff (PAS) and ubiquitin but are now identified with antibodies to the presynaptic protein α-synuclein. Lewy bodies are composed of straight neurofilaments 7–20 nm long with surrounding amorphous material and contain epitopes recognized by antibodies against phosphorylated and nonphosphorylated neurofilament proteins, ubiquitin, and α-synuclein. The presence of α-synuclein aggregates in neurons and glia in PDD and DLB molecularly classifies these diseases as synucleinopathies. In general, neuronal and synaptic loss, rather than Lewy pathology per se, best predicts the clinical deficits.

Formal criteria identify three stages of progression: (1) brainstem predominant; (2) transitional limbic; and (3) diffuse neocortical. Importantly, healthy older individuals may also show isolated scattered Lewy body pathology in the substantia nigra, amygdala, or olfactory bulb. Pathologic studies have shown that PD usually starts in the enteric nervous system and spreads through the vagus nerve to the heart, lower brainstem, substantia nigra, limbic system, and lastly the cerebral cortex. PD may also begin in the olfactory bulb and spread through olfactory system connections or start independently in enteric and olfactory bulb areas. Evidence from human anatomic pathology and animal models suggests that LBD may similarly propagate via a prion-like mechanism. Abnormally folded α-synuclein aggregates propagate transneuronally following connection pathways of the nervous system. This pathologic propagation from the periphery to the brain correlates with the evolution of clinical symptoms; PD usually manifests first with nonmotor features characterized by constipation and/or hyposmia, followed by anxiety, depression, RBD, parkinsonism, and lastly dementia. PDD is manifested clinically when limbic and cortical areas are involved.

A profound cholinergic deficit, owing to basal forebrain and pedunculopontine nucleus involvement, is present in most patients with DLB and may be associated with the characteristic fluctuations, inattention, and visual hallucinations. Adrenergic deficits from locus coeruleus involvement further undermine arousal and alerting.

PATHOGENESIS

Both genes and environmental factors are thought to contribute to the development of LBD. The presence of α-synuclein aggregates in Lewy bodies led to the discovery of α-synuclein duplications and triplications that manifest clinically as PD or DLB. There are multiple genes associated with PD, but mutations of glucocerebrosidase (GBA) particularly lead to PDD or DLB presentations **(Chap. 446)**.

The origins of LBD in gastrointestinal and olfactory areas suggest that environmental toxins acting on a susceptible genetic background may contribute to LBD pathogenesis (a "double-hit" hypothesis). Several toxins have been associated with PD **(Chap. 446)**, but epidemiologic studies of risk factors in DLB remain inconclusive.

LABORATORY FEATURES

In patients presenting with cognitive disturbances, it is always necessary to rule out treatable causes of dementia such as drugs, infections, or metabolic disturbances **(Chap. 31)**. Magnetic resonance imaging (MRI) of the brain can be helpful to rule out vascular parkinsonism or subdural hematomas, or support the diagnosis of other disorders such as MSA (i.e., pontine "hot-cross buns" sign; see **Fig. 451-6**).

The biomarkers that can help diagnose LBD include the following: a polysomnogram showing RBD without atonia, seed amplification assays (SAAs) to detect αSyn in cerebrospinal fluid (CSF), demonstrating skin deposition of α-synuclein, iodine-123-meta-iodobenzylguanidine (MIBG) cardiac scintigraphy showing cardiac postganglionic sympathetic denervation, and dopamine transporter imaging using single-photon emission computed tomography (SPECT) or positron emission tomography (PET) or, if associated with AD, increased CSF or blood levels of phospho-tau217 or phospho-tau181 (Table 445-1).

TREATMENT

Dementia with Lewy Bodies

Although there are currently no disease-modifying agents to prevent, slow, or cure LBD-related dementias, several symptomatic treatments are available. By addressing the substantial cholinergic deficit in DLB, cholinesterase inhibitors such as rivastigmine (target dose 6 mg twice daily or 9.5 mg patch daily) or donepezil (target dose 10 mg daily) often improve cognition, reduce hallucinosis, and stabilize delusional symptoms. The atypical antipsychotic pimavanserin is frequently helpful to treat the psychosis and does not worsen parkinsonism; it is approved by the U.S. Food and Drug Administration (FDA) for patients with PDD and is often used off-label for DLB. Pimavanserin (34 mg daily) is a selective inverse agonist of the serotonin 5-HT$_{2A}$ receptor that does not block dopamine receptors but carries an FDA warning regarding an increase in risk of death, especially in older patients. Low-dose clozapine (begin at 6.25 mg, increasing up to 25 mg, daily) is also effective for treating hallucinations and delusions, but requires frequent blood draws due to the risk of agranulocytosis. Patients with LBD are sensitive to dopaminergic medications, which must be carefully titrated; tolerability may be improved with concomitant use of a cholinesterase inhibitor. Patients with DLB should not be exposed to typical neuroleptics, which can lead to a neuroleptic malignant syndrome and death, or anticholinergics or dopamine agonists that can exacerbate their symptoms.

RBD usually responds to melatonin, requiring at times 20 mg/d. If melatonin is not effective, clonazepam, gabapentin, or codeine can be used with caution due to the possibility of worsening cognition or falls. Antidepressants, especially those with strong anxiolytic properties (escitalopram, paroxetine, duloxetine, or venlafaxine; see **Chap. 463**), are often necessary for mood and anxiety symptoms. Orthostatic hypotension may require treatment with nonpharmacologic measures (diet high in salt and liquids, a 30° elevation of the head of the bed) or pharmacologic therapies (i.e., fludrocortisone, midodrine, droxidopa). Physical therapy can maximize motor function and protect against fall-related injury. Home safety assessments and transfer instruction should also be provided. Education for patients and caregivers and social worker support are also important. Therefore, the care of patients with LBD requires a multidisciplinary approach.

An experimental treatment that aims to slow down the progression of the disease is neflamapimod, which targets the synapse and showed promising results in initial phase 2a studies, findings that await confirmation.

The majority of caregivers for individuals with DLB are women, often spouses, who frequently experience high levels of burden and depression. The severity of behavioral symptoms, sleep disturbances, and autonomic symptoms in the person with DLB is associated with higher caregiver burden, leading to a poorer quality of life for the caregiver. The most commonly reported caregiver concerns include the inability to plan for the future, prioritizing the needs of the person with DLB over their own, and worry about the person with DLB becoming too dependent on the caregiver, among others. Overall, caregivers expressed satisfaction with the support provided by the medical team, but they reported the lowest satisfaction with information about disease progression and the sharing of information among medical team members. Clinicians can address caregiver needs by providing support resources, educating caregivers about DLB, and developing management strategies for the range of troubling symptoms experienced by patients.

■ FURTHER READING

Diaz-Galvan P et al: Plasma biomarkers of Alzheimer's disease in the continuum of dementia with Lewy bodies. Alzheimers Dement 20:2485, 2024.

Emre M et al: Clinical diagnostic criteria for dementia associated with Parkinson's disease. Mov Disord 22:1689, 2007.

Litvan I et al: Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. Mov Disord 27:349, 2012.

Mckeith IG et al: Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium. Neurology 89:88, 2017.

Mckeith IG et al: Research criteria for the diagnosis of prodromal dementia with Lewy bodies. Neurology 94:743, 2020.

Okuzumi A et al: Propagative α-synuclein seeds as serum biomarkers for synucleinopathies. Nat Med 29:1448, 2023.

Rossi M et al: Ultrasensitive RT-QuIC assay with high sensitivity and specificity for Lewy body-associated synucleinopathies. Acta Neuropathol 140:49, 2020.

Sonni I et al: Clinical validity of presynaptic dopaminergic imaging with 123I-ioflupane and noradrenergic imaging with 123I-MIBG in the differential diagnosis between Alzheimer's disease and dementia with Lewy bodies in the context of a structured 5-phase development framework. Neurobiol Aging 52:228, 2017.

446

Parkinson's Disease

C. Warren Olanow[*],
Anthony H. V. Schapira, Christine Klein

PARKINSON'S DISEASE AND RELATED DISORDERS

Parkinson's disease (PD) is the second most common age-related neurodegenerative disease, exceeded only by Alzheimer's disease (AD). Its cardinal clinical features were first described by the English physician James Parkinson in 1817. James Parkinson was a general physician who captured the essence of this condition based on a visual inspection of a mere handful of patients, several of whom he only observed walking on the street and did not formally examine. It is estimated that the number of people with PD worldwide is ~10.8 million, and this number is expected to double within 20 years based on the aging of the population. The mean age of onset of PD is about 60 years, and the lifetime risk is ~3% for men and 2% for women. The frequency of PD increases with age, but cases can be seen in individuals in their twenties and even younger, particularly when associated with a pathogenic gene mutation.

---

[*] Deceased.

Clinically, PD is characterized by bradykinesia (slowing), rest tremor, rigidity (stiffness), and gait dysfunction with postural instability. These are known as the classical or "cardinal" features of PD. Additional clinical features can include freezing of gait, speech difficulty, swallowing impairment, and a series of nonmotor features that include autonomic disturbances, sensory alterations, mood disorders, sleep disorders, and cognitive impairment/dementia (see **Table 446-1** and discussion below).

Pathologically, the hallmark features of PD are degeneration of dopaminergic neurons in the substantia nigra pars compacta (SNc), reduced striatal dopamine, and intraneuronal proteinaceous inclusions in cell bodies and axons that stain for α-synuclein (known as Lewy bodies and Lewy neurites; collectively as Lewy pathology) **(Fig. 446-1)**. While interest has focused on the dopamine system, neuronal degeneration with Lewy pathology can also affect cholinergic neurons of the nucleus basalis of Meynert (NBM), norepinephrine neurons of the locus coeruleus (LC), serotonin neurons in the raphe nuclei of the brainstem, and neurons of the olfactory system, cerebral hemispheres, spinal cord, and peripheral autonomic nervous system. This "nondopaminergic" pathology is likely responsible for the nonmotor clinical features listed above and in Table 446-1. It has been postulated that in some cases Lewy pathology can begin in the peripheral autonomic nervous system, gastrointestinal (GI) tract, olfactory system, or dorsal motor nucleus of the vagus nerve and then spread in a predictable and sequential manner to affect the SNc and cerebral hemispheres (Braak staging). These studies suggest that the classic degeneration of SNc dopamine neurons and the cardinal motor features of PD may develop at a mid-stage of the illness. Indeed, epidemiologic

TABLE 446-1 Clinical Features of Parkinson's Disease

CARDINAL MOTOR FEATURES	OTHER MOTOR FEATURES	NONMOTOR FEATURES
Bradykinesia	Micrographia	Anosmia
Rest tremor	Masked facies (hypomimia)	Sensory disturbances (e.g., pain, hyposmia)
Rigidity	Reduced eye blinking	Mood disorders (e.g., depression, anxiety, apathy)
Postural instability	Drooling	Sleep disturbances (e.g., fragmented sleep, RBD)
	Soft voice (hypophonia)	Autonomic disturbances
	Dysphagia	Orthostatic hypotension
	Freezing	Gastrointestinal disturbances
	Falling	Genitourinal disturbances
		Sexual dysfunction
		Cognitive impairment/dementia

Abbreviation: RBD, rapid eye movement sleep behavior disorder.

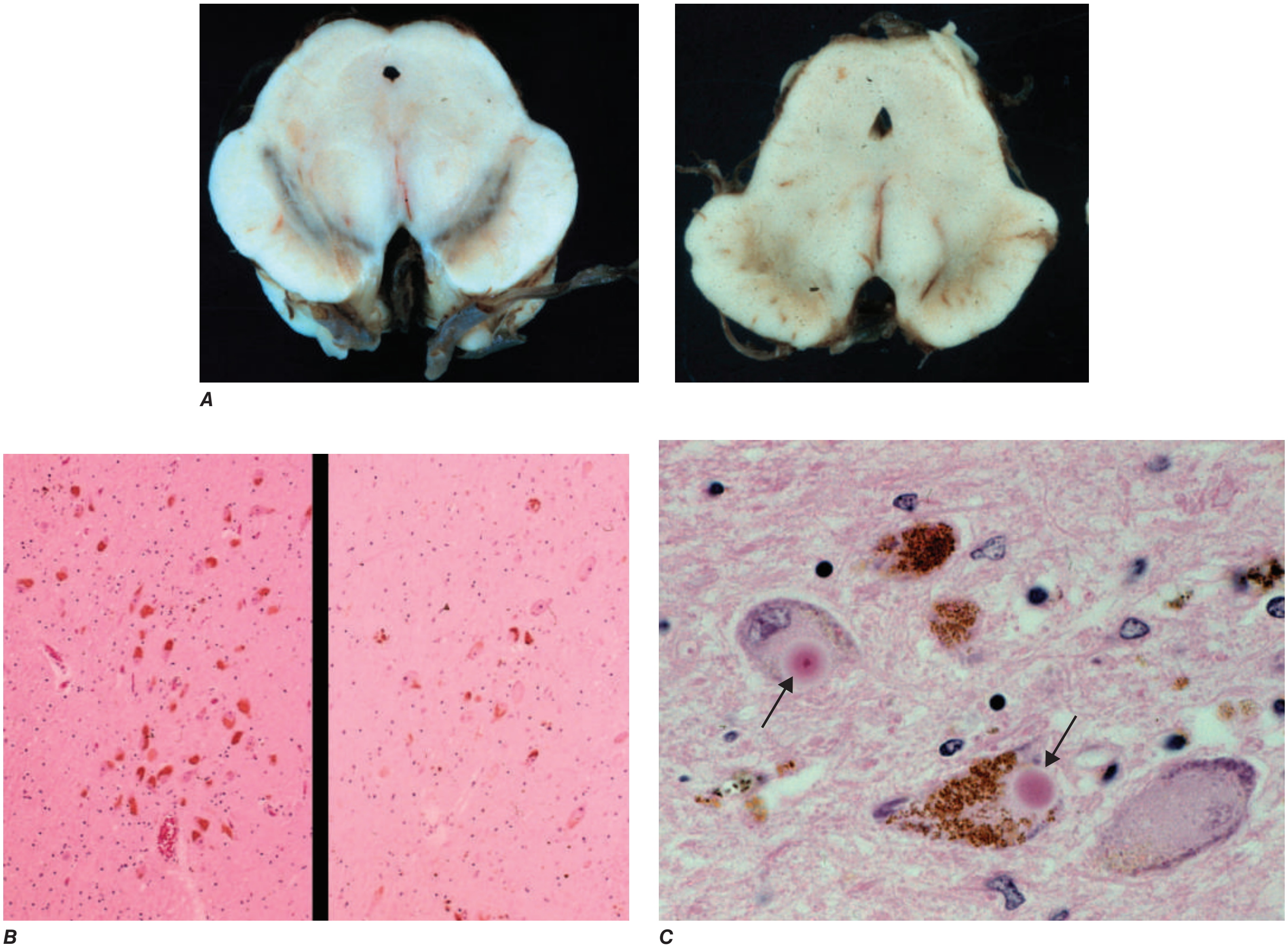

FIGURE 446-1 Pathologic specimens from a patient with Parkinson's disease (PD) compared to a normal control demonstrating (***A***) reduction of pigment in SNc in PD (*right*) versus control (*left*), (***B***) reduced numbers of cells in SNc in PD (*right*) compared to control (*left*), and (***C***) Lewy bodies (*arrows*) within melanized dopamine neurons in PD. SNc, substantia nigra pars compacta.

TABLE 446-2 Differential Diagnosis of Parkinsonism

Parkinson's disease	Atypical parkinsonism	Secondary parkinsonism	Other neurodegenerative disorders associated with parkinsonism
Sporadic	Multiple-system atrophy (MSA)	Drug-induced	Wilson's disease
Genetic	Cerebellar type (MSA-c)	Tumor	Huntington's disease
PD with dementia/dementia with Lewy bodies	Parkinson type (MSA-p)	Infection	Neurodegeneration with brain iron accumulation
	Progressive supranuclear palsy	Vascular	SCA 3 (spinocerebellar ataxia)
	Parkinsonian variant	Normal-pressure hydrocephalus	Fragile X–associated ataxia-tremor-parkinsonism
	Richardson variant	Trauma	Prion diseases
	Corticobasal syndrome	Liver failure	X-linked dystonia-parkinsonism
		Toxins (e.g., carbon monoxide, manganese, MPTP, cyanide, hexane, methanol, carbon disulfide)	Alzheimer's disease with parkinsonism
			Dopa-responsive dystonia

Abbreviation: MPTP, 1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine.

studies suggest that clinical symptoms reflecting early involvement of nondopaminergic neurons such as constipation, anosmia, rapid eye movement (REM) behavior sleep disorder, and cardiac denervation can precede the onset of the classic motor features of PD by several years if not decades. Originally it was considered that these represent risk factors for developing PD, but based on pathological findings, it is now considered likely that they represent an early premotor form of the disease. These observations have led to the notion of "body-first" and "brain-first" forms of PD based on whether pathology initially develops in the brain or periphery. Efforts are underway to accurately define the premotor stage of PD with high sensitivity and specificity. This will be of particular importance when a neuroprotective therapy becomes available as it will be desirable to initiate a disease-modifying treatment at the earliest stage of the disease possible.

Recently, two new classifications have been developed aimed at defining the early stages of PD based on biological research criteria. The first stages PD based on neuronal α-synuclein accumulation. The second takes into account α-synuclein deposition, but also the distribution of neurodegeneration and pathogenic variants in known PD-causative genes.

DIAGNOSIS AND DIFFERENTIAL DIAGNOSIS

Parkinsonism is a term that is used to define a syndrome manifest by bradykinesia with rigidity and/or tremor. The differential diagnosis includes PD, atypical parkinsonisms such as *multiple-system atrophy* (MSA) and *progressive supranuclear palsy* (PSP), secondary parkinsonism, and parkinsonism associated with other neurodegenerative conditions in which parkinsonian features are present (see **Table 446-2** and discussion below). These conditions affect the basal ganglia, a group of subcortical nuclei that include the striatum (putamen and caudate nucleus), subthalamic nucleus (STN), globus pallidus pars externa (GPe), globus pallidus pars interna (GPi), and the SNc **(Fig. 446-2)**. They differ, however, in the precise site of involvement within the basal ganglia, the specific pathologic characteristics, and the clinical picture. Among the different forms of parkinsonism, PD is the most common (~75% of cases). Historically, PD was diagnosed based on the presence of two of three parkinsonian features (tremor, rigidity, bradykinesia). However, postmortem studies found a 24% error rate when diagnosis was based solely on these criteria. Clinicopathologic correlation studies subsequently determined that parkinsonism (bradykinesia and rigidity) associated with rest tremor, asymmetry of motor impairment, and a good response to levodopa is much more likely to predict the correct pathologic diagnosis. With these revised criteria (known as the U.K. Brain Bank Criteria), a clinical diagnosis of PD could be confirmed pathologically in >90% of patients. Imaging of the dopamine system and new biomarkers (see below) further increase diagnostic accuracy. The International Parkinson's Disease and Movement Disorder Society (MDS) has proposed revised clinical criteria for PD (known as the MDS Clinical Diagnostic Criteria for Parkinson's disease), which are thought to increase diagnostic accuracy even further, particularly in early cases where levodopa has not yet been introduced. While motor parkinsonism has been retained as the core feature of the disease, in these criteria, the specific diagnosis of PD relies on three additional categories of diagnostic features: supportive criteria (features that increase confidence in the diagnosis of PD), absolute exclusion criteria, and red flags (which must be counterbalanced by supportive criteria to permit a diagnosis of PD). Utilizing these criteria, two levels of certainty have been delineated: clinically established PD and clinically probable PD (see Berg et al. in "Further Reading").

Imaging of the brain dopamine system can be helpful in diagnosing PD and is performed using positron emission tomography (PET) or single-photon emission computed tomography (SPECT). These studies typically show reduced and asymmetric uptake in the striatum, particularly in the posterior putamen with relative sparing of the caudate nucleus **(Fig. 446-3)**. These findings reflect the degeneration

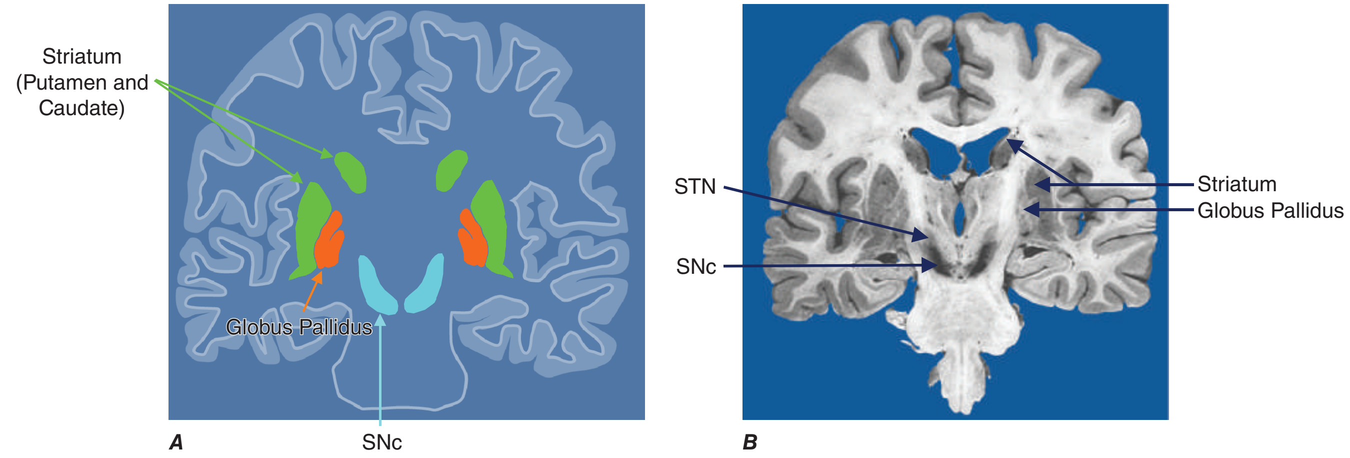

FIGURE 446-2 Basal ganglia nuclei. Schematic (***A***) and postmortem (***B***) coronal sections illustrating the various components of the basal ganglia. SNc, substantia nigra pars compacta; STN, subthalamic nucleus.

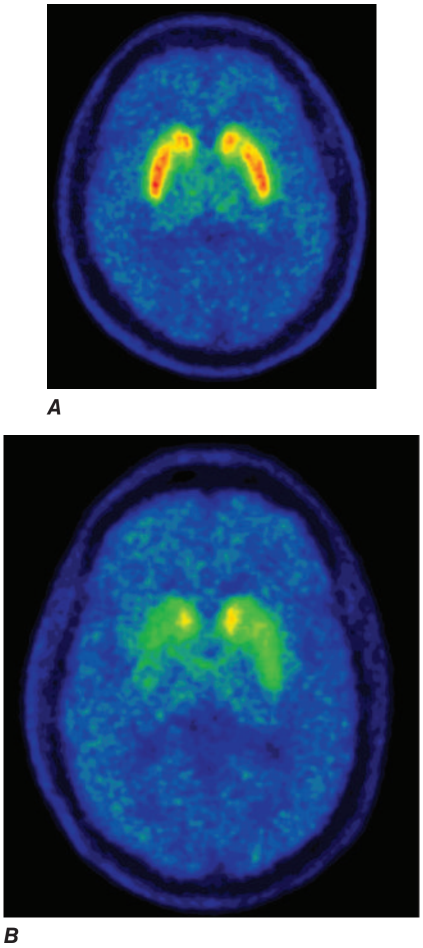

FIGURE 446-3 [^{11}C]Dihydrotetrabenazine positron emission tomography (a marker of VMAT2) in healthy control (*A*) and Parkinson's disease (*B*) patient. Note the reduced striatal uptake of tracer, which is most pronounced in the posterior putamen and tends to be asymmetric. *(Courtesy of Dr. Jon Stoessl.)*

of nigrostriatal dopaminergic neurons and the loss of their striatal terminals. Imaging is useful in patients where there is diagnostic uncertainty (e.g., early-stage disease, essential tremor, dystonic tremor, psychogenic tremor) or in research studies in order to ensure diagnostic accuracy, but it is not routinely required in clinical practice. This may change in the future when a disease-modifying therapy becomes available and it becomes critically important to make a correct diagnosis as early as possible. There is also some evidence suggesting that a diagnosis of PD, and even prodromal PD, may be made based on the presence of increased iron in the SNc using transcranial sonography or special magnetic resonance imaging (MRI) protocols. There have been intensive efforts to image α-synuclein in the brain but, in contrast to beta-amyloid or tau imaging in Alzheimer's disease, this has proven difficult as most of the abnormal α-synuclein protein is located within cells. This makes it difficult to develop a marker that binds to α-synuclein and that can be detected with imaging.

There has been a longstanding interest in developing a biomarker for PD that could aid in diagnosis, differentiate PD from other parkinsonian conditions, potentially assess the effects of a putative disease-modifying therapy, and be used as an endpoint in clinical trials. Considerable interest has focused on detecting abnormal α-synuclein deposits in cerebrospinal fluid (CSF), blood, muscle, and other tissues, but results to date have been inconsistent. The development of the α-synuclein seeding amplification assay (SAA) has provided a novel means to support a clinical diagnosis of PD. The SAA was developed for use on CSF and skin and provides a binary result indicating the presence or absence of endogenous α-synuclein sufficient to result in aggregation upon addition of α-synuclein "seeds." This assay has very high sensitivity and specificity and is able to distinguish PD from other parkinsonisms. At present, the test has been primarily applied in a research setting, but the development of a blood-based assay may extend its use into a clinical role. This assay also has the potential to permit diagnosis in early-stage and even prodromal PD.

Genetic testing can be helpful for establishing a diagnosis but is not routinely employed as monogenic forms of PD are relatively uncommon and account for only 5% of cases, although this increases to 15% when pathogenic variants in the strongest known risk gene, *glucocerebrosidase* (*GBA1*), are included (see discussion below), and this number may increase as more knowledge is acquired. A genetic form of PD should be considered in patients with a strong positive family history, early age of onset (<40 years), and a particular ethnic background (see below), and in research studies. Genetic variants of *GBA1* are the most common genetic association with PD. They are present in ~10% of PD patients and in 25% of Ashkenazi PD patients. However, only ~20–30% of people with *GBA1* variants will develop PD, and PD risk is correlated with the severity of the variant effect. Pathogenic variants in the *LRRK2* gene have also attracted particular interest as they are responsible for ~3% of typical sporadic cases of the disease. *LRRK2* mutations are a particularly common cause of PD (~25%) in Ashkenazi Jews and North African Berber Arabs; however, there is considerable variability in penetrance, and ~40–50% of carriers never develop clinical features of PD. Interestingly, some PD cases associated with *LRRK2* mutations and other genetic causes have been described without Lewy bodies. Genetic testing is of particular interest for identifying at-risk individuals in a research setting and for defining enriched populations for clinical trials of therapies directed at a pathogenic mutation or pathway.

Atypical, Secondary, and Other Forms of Parkinsonism

Atypical parkinsonism refers to a group of neurodegenerative conditions that are usually associated with more widespread pathology than found in PD (e.g., degeneration potentially involving the striatum, globus pallidus, cerebellum, and brainstem as well as the SNc). These conditions include MSA **(Chap. 451)**, PSP **(Chap. 443)**, and *corticobasal syndrome* (CBS) **(Chap. 443)**. As a group, they tend to present with parkinsonism (rigidity and bradykinesia) but manifest clinical differences from PD, reflecting their different pathologies. Clinical features that typically differ from classical PD include early involvement of speech and gait, absence of rest tremor, lack of motor asymmetry, poor or no response to levodopa, and a more aggressive clinical course. They can be difficult to distinguish from PD in the early stages where levodopa has not yet been tried and in some cases that show a modest benefit from levodopa, but the diagnosis usually becomes clear as the disease evolves over time.

Neuroimaging of the dopamine system is usually not helpful, as striatal dopamine depletion can be seen in both PD and atypical parkinsonism. By contrast, metabolic imaging of the basal ganglia/thalamus network (using 2-F-deoxyglucose) may be helpful, showing a pattern of decreased activity in the GPi with increased activity in the thalamus, the reverse of what is seen in PD.

MSA manifests as a combination of the atypical parkinsonian features described above, as well as varying degrees of cerebellar and autonomic features. Clinical syndromes can be divided into a predominantly parkinsonian (MSA-p), cerebellar (MSA-c), and more rarely, a primary autonomic form. Clinically, MSA is suspected when a patient has features of atypical parkinsonism in conjunction with cerebellar signs and/or prominent autonomic dysfunction, usually orthostatic hypotension and a poor or absent response to levodopa **(Chap. 451)**. The use of biomarkers (e.g., SAA) has increased the accuracy of diagnosis in the early stages of the disease, and the rate of progression is typically more aggressive than in classic PD. Pathologically, MSA is characterized by degeneration of the SNc, striatum, cerebellum, and inferior olivary nuclei coupled with characteristic glial cytoplasmic inclusions (GCIs) that stain positively for α-synuclein aggregates (Lewy bodies), which accumulate in oligodendrocytes rather than in SNc neurons as in PD. MRI can show pathologic iron accumulation in the striatum on T2-weighted scans, high signal change in the region of the external surface of the putamen (putaminal rim) in MSA-p, or cerebellar and brainstem atrophy (the pontine "hot cross bun" sign **[Fig. 451-6]**) in MSA-c. There is currently no established evidence for any gene mutation or genetic risk factor for MSA, and no specific treatment exists.

PSP is characterized by the features noted above coupled with slow ocular saccades, eyelid apraxia, and restricted vertical eye movements with impairment of downward gaze. Patients frequently experience hyperextension of the neck with early gait disturbance and falls. In later stages, speech and swallowing difficulty and cognitive impairment may become evident. Two clinical forms of PSP have been identified: a "Parkinson" form that can closely resembles PD in the early stages and can include a positive response to levodopa, and the more classic "Richardson" form that is characterized by the features described above with little or no response to levodopa. MRI may reveal a characteristic atrophy of the midbrain with relative preservation of the pons on midsagittal images (the so-called "hummingbird sign"). Pathologically, PSP is characterized by degeneration of the SNc, striatum, STN, midline thalamic nuclei, and pallidum, coupled with neurofibrillary tangles and inclusions that stain for the tau protein. Mutations in the *MAPT* gene encoding the tau protein have been detected in some familial cases.

CBS is a relatively uncommon condition that usually presents with asymmetric dystonic contractions, and clumsiness of one hand coupled with cortical sensory disturbances manifest as apraxia, agnosia, focal limb myoclonus, or alien limb phenomenon (where the limb assumes a position in space without the patient being aware of its location or recognizing that the limb belongs to them). Dementia may occur at any stage of the disease. Both cortical and basal ganglia features are required to make this diagnosis. MRI frequently shows asymmetric cortical atrophy, but this must be carefully sought and may not be obvious on casual inspection. Pathologic findings include achromatic neuronal degeneration with tau deposits. Considerable overlap may occur both clinically and pathologically between CBS and PSP, and they may be difficult to distinguish without pathologic confirmation.

Secondary parkinsonisms occur as a consequence of other etiologic factors such as drugs, stroke, tumor, infection, or toxins (e.g., carbon monoxide, manganese) that cause basal ganglia dysfunction. Clinical features reflect the region of the basal ganglia that has been damaged. For example, strokes or tumors that affect the SNc may have a clinical picture that is very similar to PD, whereas toxins such as carbon monoxide or manganese that damage the globus pallidus more closely resemble atypical parkinsonism and have a poor response to levodopa. Dopamine-blocking agents such as neuroleptics are the most common cause of secondary parkinsonism. These drugs are most widely used in psychiatry, but physicians should be aware that drugs such as metoclopramide, which are primarily used to treat GI problems, are also neuroleptic *agents* and may induce secondary parkinsonism. These drugs can also cause acute and tardive dyskinesias **(see Chap. 447)**. Other drugs that can cause secondary parkinsonism include tetrabenazine, calcium channel blockers (flunarizine, cinnarizine), amiodarone, and lithium.

Parkinsonism can also be seen as a feature of *dopa-responsive dystonia* (DRD), a condition that typically results from pathogenic variants in the *GTP-cyclohydrolase 1* gene, which lead to a defect in a cofactor for tyrosine hydroxylase with impairment in the manufacture of dopa and dopamine. While it typically presents as dystonia **(Chap. 447)**, it can present as a biochemically based form of parkinsonism (due to reduced synthesis of dopamine) closely resembling PD. DRD patients respond to levodopa, but abnormalities on fluorodopa PET (FD-PET) are typically not seen, nor are drug-induced dyskinesias, reflecting a biochemical abnormality without degeneration of the underlying anatomic structures. DRD should be considered in individuals aged <20 years who present with parkinsonism, particularly if there are dystonic features.

Finally, parkinsonism can be seen as a feature of a variety of other neurodegenerative disorders such as Wilson's disease **(Chaps. 427 and 447)**, Huntington's disease (especially the juvenile form known as the Westphal variant) **(Chap. 447)**, certain spinocerebellar ataxias **(Chap. 450)**, and neurodegenerative disorders with brain iron accumulation such as pantothenate kinase (PANK)–associated neurodegeneration (formerly known as Hallervorden-Spatz disease). It is particularly important to rule out Wilson's disease, as progression can be prevented with the use of copper chelators.

TABLE 446-3 Features Suggesting an Atypical or Secondary Cause of Parkinsonism

SYMPTOMS/SIGNS	ALTERNATIVE DIAGNOSIS TO CONSIDER
History	
Early speech and gait impairment (lack of tremor, lack of motor asymmetry, early falls)	Atypical parkinsonism
Exposure to neuroleptics	Drug-induced parkinsonism
Onset prior to age 40 years	Genetic form of PD, Wilson's disease, DRD
Liver disease	Wilson's disease, non-Wilsonian hepatolenticular degeneration
Hallucinations and dementia which precede the development of PD features	Dementia with Lewy bodies
Diplopia, impaired vertical gaze	PSP
Poor or no response to an adequate trial of levodopa	Atypical or secondary parkinsonism
Physical Examination	
Dementia as first or early feature	Dementia with Lewy bodies
Prominent orthostatic hypotension	MSA
Prominent cerebellar signs	MSA-c
Slow saccades with impaired downgaze	PSP
High-frequency (6–10 Hz) symmetric postural tremor with a prominent kinetic component	Essential tremor

Abbreviations: DRD, dopa-responsive dystonia; MSA-c, multiple-system atrophy–cerebellar type; MSA-p, multiple-system atrophy–Parkinson's type; PD, Parkinson's disease; PSP, progressive supranuclear palsy.

Some features that suggest that parkinsonism might be due to a condition other than classic PD are shown in **Table 446-3**.

ETIOLOGY AND PATHOGENESIS

Most PD cases occur sporadically and are of unknown cause. Gene mutations (see below) are the only known causes of PD and may be found even in seemingly sporadic cases. Twin studies performed several decades ago suggested that environmental factors may play an important role in patients with an age of onset ≥50 years, with genetic factors being more important in younger-onset patients. However, the demonstration of genetic variants (e.g., *LRRK2* and *GBA1*) causing later onset PD shows that certain monogenic forms can manifest as late as in the eighth or ninth decade. With the advent of new sequencing technologies (long-read sequencing), numerous monogenic causes of late-onset neurodegenerative diseases have recently been identified, such as intronic repeat expansions in the *FGF14* gene causing late-onset ataxia **(Chap. 450)**, and it is conceivable that additional monogenic forms of PD will also be identified with this technology. In addition, it is likely that genetic factors could modify age at onset and severity of both genetic and nongenetic forms of PD.

The environmental hypothesis received some support in the 1980s with the demonstration that MPTP (1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine), a by-product of the illicit manufacture of a heroin-like drug, caused a PD syndrome in addicts in northern California. MPTP is transported into the central nervous system, where it is oxidized to form MPP^+, a mitochondrial toxin that is selectively taken up by, and damages, dopamine neurons, but typically without the formation of Lewy bodies. Importantly, MPTP or MPTP-like compounds have not been linked to sporadic PD. Epidemiologic studies have reported an increased risk of developing PD in association with exposure to pesticides, solvents, rural living, farming, and drinking well water, but study results have been inconsistent. Additionally, dozens of other associations have also been reported in individual studies. To date, no environmental factor has yet been proven to be a cause of PD. Some possible protective factors have also been identified in epidemiologic studies, including caffeine, cigarette smoking, intake of nonsteroidal anti-inflammatory drugs, and calcium channel blockers.

The validity of these findings and the responsible mechanism remain to be established.

Large studies show that about 15% of PD cases are familial in origin, and mutations in several PD-linked genes have been identified **(Table 446-4)**. While uncommon pathogenic variants in PD genes (i.e., mutations) have been shown to be causative of PD or to contribute to PD risk, a plethora of common genetic variants—alone or in combination as part of polygenic risk scores—are associated with an increased risk of developing PD. These include variants in the *SNCA*, *LRRK2*, *MAPT*, and *GBA1* genes and may be ethnicity-specific, such as a strong risk variant confined to the African or African-admixed population. It has been proposed that many cases of PD may be due to a "double hit" involving an interaction between (1) one or more genetic risk factors that induce susceptibility and (2) exposure to a toxic environmental factor that may induce epigenetic or somatic DNA alterations or has the potential to directly damage the dopaminergic system. In this scenario, two factors (or more) are required for PD to ensue, while the presence of either one alone is not sufficient to cause the disease. While the "double-hit" hypothesis is of interest, there is no direct evidence for its support at this time. Furthermore, even if a genetic or environmental risk factor doubles the risk of developing PD, this only results in a lifetime risk of 4–6% or lower, and thus cannot presently be used for individual patient counseling.

Thus, the bulk of accumulating evidence suggests that genetic factors play an important role in both familial and "sporadic" forms of PD, while the role of environmental factors remains unsettled. Although mutations in PD genes identified to date cause only a minority of cases of PD, they have been very helpful in pointing to specific targets, pathways, and molecular mechanisms that are likely to be central to the neurodegenerative process in the sporadic form of the disease. Detailed clinical and genetic features of monogenic PD are available in the MDSGene database (*www.mdsgene.org*).

The α-synuclein gene (*SNCA*) was the first to be linked to PD and is also the most intensely investigated with respect to causative mutations, risk variants, function, and role in the etiopathogenesis of PD. Shared clinical features of patients with *SNCA* mutations include earlier age of disease onset than in nongenetic PD, a faster progression of motor signs that are mostly levodopa-responsive, early occurrence of motor fluctuations, and presence of prominent nonmotor features, particularly cognitive impairment. Importantly, duplication or triplication of the wild-type *SNCA* gene also causes PD, with triplication carriers being more severely affected than carriers of duplications. These findings indicate that increased production of the normal protein alone can cause PD. Intriguingly, α-synuclein constitutes the major component of Lewy bodies, implicating the protein in sporadic forms of PD as well (Fig. 446-1). In a remarkable study, Lewy pathology was discovered to have developed in healthy embryonic dopamine neurons that had been implanted into the striatum of PD patients, suggesting that the abnormal protein had transferred from affected cells to healthy unaffected dopamine neurons. Based on these findings, it has been proposed that the α-synuclein protein may be a prion and PD a prion disorder **(Chaps. 435 and 449)**. In this model, α-synuclein can misfold

TABLE 446-4 Confirmed Genetic Causes of Parkinson's Disease (PD) with a Clinical Presentation Similar to Idiopathic PD[a]

DESIGNATION[a] AND REFERENCE	GENEREVIEWS AND OMIM REFERENCE	CLINICAL CLUES[B]	COMMENTS
Dominantly Inherited PD			
PARK-*SNCA*	GeneReviews *http://www.ncbi.nlm.nih.gov/books/NBK1223/* OMIM 168601	Median AAO: 46 years (range 19–77 years); 25th/75th percentile: 36/54 years. Gene duplications cause classical PD. Most missense mutations and triplications cause early-onset, severe parkinsonism with prominent cognitive dysfunction	Very rare form of PD, α-synuclein protein main component of Lewy bodies, the pathological hallmark of PD
PARK-*LRRK2*	GeneReviews *http://www.ncbi.nlm.nih.gov/books/NBK1208/* OMIM 607060	Median AAO: 56 years (range 20–95 years); 25th/75th percentile: 47/64 years. Clinically typical PD with slightly slower progression	Most common known genetic form of PD
PARK-*VPS35*	GeneReviews *http://www.ncbi.nlm.nih.gov/books/NBK1223/* OMIM 616710	Median AAO: 52 years (range 26–75 years); 25th/75th percentile: 45/61 years. Clinically typical PD	Very rare form of PD
PARK-*CHCHD2*	GeneReviews N/A OMIM 614203	Likely clinically typical PD. Systematic MDSGene review not yet available	Very rare form of PD, predominantly found in Asia
PARK-*RAB32*	GeneReviews *N/A* OMIM 612906 (disease link not yet included)	Likely clinically typical PD, possibly more frequent dementia. Systematic MDSGene review not yet available	Most recently found form of PD. All currently identified patients and families carry the same founder pathogenic variant
PARK-*GBA1*	GeneReviews *http://www.ncbi.nlm.nih.gov/books/NBK1223/* OMIM 168600/606463	Clinically overall typical PD; however, faster progression and greater risk of cognitive impairment. Systematic MDSGene review not yet available	Strongest known genetic risk factor for PD; incomplete penetrance
Recessively Inherited PD			
PARK-*PRKN*	GeneReviews *http://www.ncbi.nlm.nih.gov/books/NBK1155/* OMIM 600116	Median AAO: 31 years (range 3–81 years); 25th/75th percentile: 23/38 years. Often presents with dystonia, typically in a leg	Most common early-onset form of genetic PD. Protein name: Parkin
PARK-*PINK1*	GeneReviews *http://www.ncbi.nlm.nih.gov/books/NBK1223/* OMIM 605909	Median AAO: 32 years (range 9–67 years); 25th/75th percentile: 24/40 years. Prominent psychiatric features have been described in several families	Clinically very similar to PARK-*PRKN* but much rarer
PARK-*PARK7*	GeneReviews *http://www.ncbi.nlm.nih.gov/books/NBK1223/* OMIM 606324	Median AAO: 27 years (range 15–40 years); 25th/75th percentile: 22/34	Clinically very similar to PARK-*PRKN* and PARK-*PINK1*, but rarest of all forms. Protein name: DJ-1

[a]According to the recommendations of the International Parkinson's and Movement Disorder Society (C Marras et al: Mov Disord 31:436, 2016; and L Lange et al: Mov Disord 37:905, 2022). [b]Adapted from MDSGene (*www.mdsgene.org*).

Abbreviations: AAO, age at onset; N/A, not applicable; OMIM, Online Mendelian Inheritance in Man; PD, Parkinson's disease.

to form β-rich sheets, join to form toxic oligomers and aggregates, polymerize to form amyloid plaques (i.e., Lewy bodies), and cause neurodegeneration with spread to unaffected neurons. Indeed, injection of purified α-synuclein fibrils into the striatum of both transgenic and wild-type rodents produced Lewy pathology in host neurons, neurodegeneration, behavioral abnormalities, and spread of α-synuclein pathology to anatomically connected sites. Further support for this hypothesis comes from the demonstration that inoculation into the striatum of homogenates derived from human Lewy bodies induces dopamine cell degeneration and widespread Lewy pathology in mice and primates. Evidence also suggests that in some cases α-synuclein pathology might begin peripherally within the GI tract and spread by way of the vagus nerve to the lower brainstem (dorsal motor nucleus of the vagus) and ultimately to the SNc to cause the motor features of PD (the Braak hypothesis). There is also interest in the possibility that the gut microbiome in PD patients can cause inflammatory changes that promote α-synuclein misfolding with spread to the brain via the vagus nerve. The gut-brain axis might therefore offer a mechanism by which α-synuclein pathology could spread to the brain and cause PD. The prion hypothesis for PD represents an exciting, although still unproven, line of investigation.

Multiple lines of evidence support the concept that neuroprotective therapies for PD might be developed based on inhibiting the accumulation or accelerating the removal of toxic forms of α-synuclein, knocking down levels of host *SNCA* to prevent their misfolding, preventing the spread of misfolded *SNCA*, or blocking the templating phenomenon whereby misfolded α-synuclein promotes misfolding of the native protein in a prion-like chain reaction. Numerous studies testing different approaches to targeting α-synuclein are ongoing. Interestingly, postmortem studies in PD patients who had undergone a transplant procedure observed that inflammation with activated microglia at the transplant site preceded the development of α-synuclein aggregates by many years. This suggests the possibility that a chronic inflammatory milieu could promote misfolding of host α-synuclein, leading to neurodegeneration; to date, however, immune-based approaches to clearing α-synuclein have not been successful in PD.

Pathogenic variants of the *GBA1* gene represent the most important risk factor in terms of both the development of PD and its severity. GBA encodes the enzyme glucocerebrosidase (GCase), which promotes lysosomal function and enhances the clearance of misfolded α-synuclein. The identification of *GBA1* as a risk for PD resulted from the clinical observation that patients with Gaucher's disease (GD) and their relatives show features of parkinsonism more frequently than would be expected. This clinical observation led to the discovery that literally hundreds of variants in *GBA1* confer risk for the development of PD. Experimentally, it has been shown that reduced levels of GCase activity due to GBA variants impair lysosomal function, resulting in the accumulation of α-synuclein. Conversely, the accumulation of α-synuclein leads to inhibition of lysosomal function and a further reduction in levels of GCase by interfering with endoplasmic reticulum-to-Golgi trafficking. Thus, there is a vicious cycle in which decreased GCase activity leads to the accumulation of α-synuclein, and increased levels of α-synuclein lead to impairment in lysosomal function. In this regard, it is noteworthy that lysosomal function is impaired and levels of GCase are reduced in patients with sporadic PD, and not just in those with *GBA1* variants. These findings suggest that this molecular pathway may not only apply to patients with a *GBA1* variants but also to patients with sporadic PD or other synucleinopathies who have normal wild-type *GBA1* alleles. Some studies suggest that patients with certain GBA1 variants (e.g., L444P) have a faster rate of progression and an increased frequency of cognitive impairment. Drug and gene-based therapies that enhance GCase activity and promote lysosomal function are currently being tested in the clinic as putative neuroprotective therapies.

Multiple *LRRK2* pathogenic mutations have also been clearly linked to PD; p.G2019S is the most common, possibly due to a founder effect in the Ashkenazi Jewish and North African Arab populations. Pathogenic variants in *LRRK2* account for 2–41% of familial PD cases (depending on the specific population) and are also found in apparently sporadic cases, albeit at a lower rate. More than 200 variants have been reported. Recently developed functional assays testing activation of kinase activity as a gain-of-function effect help distinguish causative variants from those of uncertain significance. The phenotype of *LRRK2* p.G2019S mutations is largely indistinguishable from that of sporadic PD, although tremor appears to be more common and disease progression is slightly slower than in idiopathic PD. The penetrance of *LRRK2* pathogenic variants is incomplete (30–74% depending on the ethnic group), and patients tend to run a more benign course, with less cognitive impairment than seen in idiopathic PD. The mechanism responsible for cell death with this mutation is likely due to enhanced kinase activity with altered phosphorylation of target proteins (including autophosphorylation) with possible impairment of lysosomal function. In laboratory models, kinase inhibitors can block toxicity associated with *LRRK2* pathogenic variants. Accordingly, there has been interest in developing drugs directed at this target. However, nonselective kinase inhibitors are potentially toxic to the lungs and kidneys. Fortunately, LRRK2 inhibitors have now been developed that have good preclinical safety and are currently being tested in PD populations with and without pathogenic *LRRK2* variants. There has been particular interest in c-Abl inhibitors, which target the tyrosine residue on α-synuclein protein and potentially prevent conversion to a toxic species. A pathogenic variant in *RAB32* has been identified as a novel form of dominantly inherited PD with incomplete penetrance. Interestingly, *LRRK2* and *RAB32* interact directly, thus representing another potentially druggable target.

Pathogenic variants in *PRKN, PINK1,* and *PARK7* have also been identified as a cause of PD. *PRKN* mutations are the most common and the major cause of autosomal recessive early-onset PD, accounting for up to 77% of juvenile PD patients, with an age of onset <20 years, and for 10–20% of early-onset PD patients in general. The disease is slowly progressive, responds well to antiparkinsonian treatment, and is commonly complicated by dystonia, but rarely by dementia. Pathologically, neurodegeneration tends to be restricted to the SNc and LC in patients with *PRKN* mutations, and Lewy bodies are only present in ~20% of the brains. The reason for these differences from classic PD is not known but may be related to the fact that parkin is a ubiquitin ligase and ubiquitination of damaged proteins is required for their clearance and possibly for their incorporation into Lewy bodies. The clinical phenotypes linked to pathogenic variants in *PRKN, PINK1,* and *PRK7* are indistinguishable from one another. Parkin and PINK1 proteins are involved in cell protection mechanisms related to the turnover and clearance of damaged mitochondria (mitophagy). Indeed, mutations in *Parkin* and *PINK1* cause mitochondrial dysfunction in transgenic animals that can be corrected with overexpression of parkin. Improving mitochondrial function is another attractive therapeutic target as postmortem studies in PD patients show a defect in complex I of the respiratory chain in SNc neurons.

Several factors have been implicated in the pathogenesis of cell death in PD, including oxidative stress, inflammation, excitotoxicity, mitochondrial dysfunction, and lysosomal/proteasomal dysfunction. Inflammation and altered immunity have also been implicated as potentially key factors in the degenerative process. Genetic studies demonstrate an association of PD with the class II human leukocyte antigen (HLA) gene *DRB1* (variants of which are associated with either protection or risk for PD) and findings in monogenic forms of PD demonstrating a role of inflammation and the immune system. As noted above, this is supported by pathologic studies demonstrating that inflammation occurs years before the accumulation of α-synuclein aggregates in transplanted patients, suggesting that inflammation plays a triggering role. Altered immunity has also been suggested by studies showing that autoreactive T cells recognizing peptides derived from α-synuclein are present in PD patients. Further, drugs such as sargramostim that upregulate T-regulatory cells have shown positive results in studies in animal models, and early studies in PD patients are underway. Whatever the pathogenic mechanism, cell death appears to occur, at least in part, by way of a signal-mediated apoptotic or "suicidal" process. Each of these mechanisms offers a potential target for putative neuroprotective drugs; however, clinical

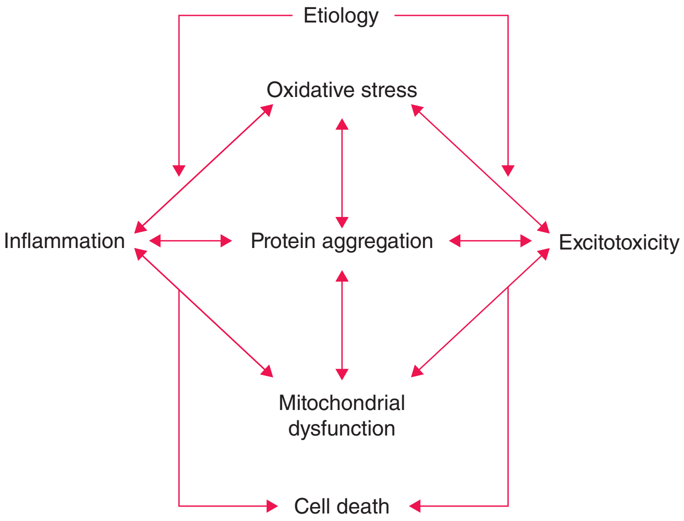

FIGURE 446-4 Schematic representation of how pathogenetic factors implicated in Parkinson's disease interact in a network manner, ultimately leading to cell death. This figure illustrates how interference with any one of these factors may not necessarily stop the cell death cascade. *(Reproduced with permission from CW Olanow: The pathogenesis of cell death in Parkinson's disease. Movement Disorders 22:S-335, 2007.)*

studies to date have not conclusively demonstrated a benefit using therapies directed against any of these targets. Moreover, it is not clear which of these factors is primary, if they are the same in all cases or specific to individual subgroups, if they act by way of a network such that multiple insults are required for neurodegeneration to ensue, or if the findings discovered to date merely represent epiphenomena unrelated to the true cause of cell death that still remains undiscovered (Fig. 446-4).

It is anticipated that a better understanding of the pathways involved in the etiology and pathogenesis of cell death in PD will permit the development of more relevant animal models and better-defined targets for the development of neuroprotective drugs.

PATHOPHYSIOLOGY OF PD

The classic model of the organization of the basal ganglia in the normal and PD states is provided in Fig. 446-5. With respect to motor function, a series of neuronal circuits with multiple feedback and feedforward loops link the basal ganglia nuclei with corresponding cortical and brainstem motor regions in a somatotopic manner. The striatum is the major input region of the basal ganglia, whereas the GPi and substantia nigra (SNr) are the major output regions. The input and output regions are connected via direct and indirect pathways that have reciprocal effects on basal ganglia activity and motor function. The output of the basal ganglia provides inhibitory (GABAergic) tone to modulate excitatory thalamic and brainstem neurons that, in turn, connect to motor systems in the cerebral cortex and spinal cord that control motor function. An increase in neuronal activity in the output regions of the basal ganglia (GPi/SNr) is associated with reduced thalamic activity and poverty of movement or parkinsonism, while decreased output results in movement facilitation. Dopaminergic projections from SNc neurons serve to modulate neuronal firing (in both directions) and thus to stabilize the basal ganglia network. Normal dopamine innervation thus serves to facilitate the selection of the desired movement and suppression or rejection of unwanted movements. Cortical loops integrating the cortex and the basal ganglia are thought to also play an important role in regulating other systems, such as behavioral, emotional, and cognitive functions.

In PD, dopamine denervation with loss of dopaminergic tone leads to increased firing of neurons in the STN and GPi, excessive inhibition of the thalamus, reduced activation of cortical motor systems, and the development of parkinsonian features (Fig. 446-5). The current role of surgery in the treatment of PD is based on this model, which predicted that lesions or high-frequency stimulation of the STN or GPi might reduce their inhibition of thalamocortical pathways and thus improve PD features. This model has not proven to be as valuable in understanding dyskinesia where benefits are also seen with lesions in these regions (see below) and where it is now thought that that dyskinesia arises from altered firing patterns and not just firing frequency.

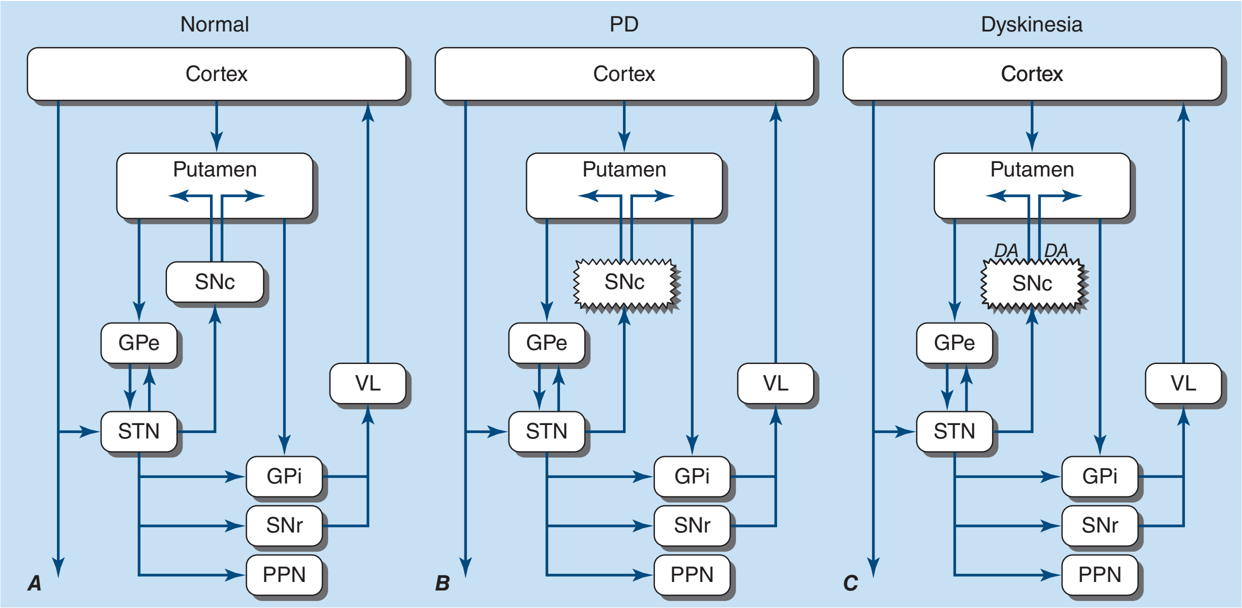

FIGURE 446-5 Basal ganglia organization. Classic model of the organization of the basal ganglia in the normal (***A***), Parkinson's disease (PD) (***B***), and levodopa-induced dyskinesia (***C***) state. Inhibitory connections are shown as *blue arrows* and excitatory connections as *red arrows*. The striatum is the major input region and receives its input from the motor regions of the cerebral cortex. The GPi and SNr are the major output regions, and they project to the thalamocortical and brainstem motor regions. The striatum and GPi/SNr are connected by direct and indirect striatal pathways whose neurons are D1- and D2-bearing, respectively. This model predicts that parkinsonism results from decreased dopamine inhibition of the indirect pathway, leading to increased neuronal firing in the STN and GPi with inhibition of thalamocortical firing. These observations suggested that lesions or DBS of these targets might provide antiparkinsonian benefit. The model also predicts that dyskinesia results from decreased firing of the output regions, resulting in excessive cortical activation by the thalamus. This component of the model is not completely correct because lesions of the GPi ameliorate rather than increase dyskinesia in PD, suggesting that firing frequency is just one of the components that lead to the development of dyskinesia and that other components of the neuronal firing pattern such as pauses and bursts are also important. DBS, deep brain stimulation; GPe, external segment of the globus pallidus; GPi, internal segment of the globus pallidus; PPN, pedunculopontine nucleus; SNc, substantia nigra, pars compacta; SNr, substantia nigra, pars reticulata; STN, subthalamic nucleus; VL, ventrolateral thalamus. *(Derived from JA Obeso et al: Trends Neurosci 23:S8, 2000.)*

COVID-19 AND PD

SARS-CoV-2 viral infection can worsen PD features and off time. In addition, having PD increases the risks of complications and death rate associated with having a SARS-CoV-2 infection. Interestingly, it has been shown that the SARS-CoV-2 virus can enter the brain and cause inflammation with microglial activation, and new-onset cases of PD have been reported following infection. In this regard, it raises similarities to PD cases associated with the influenza A epidemic in 1918. Home confinement due to the risks of acquiring SARS-CoV-2 infection has also altered conduct of clinical trials in PD patients and promoted "remote" clinical trials in which patients are evaluated online rather than in person. With validation of the reliability of this approach, it is likely that remote clinical trials will be increasingly employed in routine clinical trials of PD patients.

TREATMENT

Parkinson's Disease

LEVODOPA

Since its introduction in the late 1960s, levodopa has been the mainstay of therapy for PD. Experiments in the late 1950s by Carlsson and colleagues demonstrated that blocking dopamine uptake with reserpine caused rabbits to become parkinsonian; this could be reversed with the dopamine precursor levodopa. Subsequently, Hornykiewicz demonstrated a dopamine deficiency in the striatum of PD patients and suggested the potential benefit of dopamine replacement therapy. Dopamine does not cross the blood-brain barrier (BBB), so clinical trials were initiated with levodopa, the precursor of dopamine. Studies over the course of the next decade confirmed the value of levodopa and revolutionized the treatment of PD.

Levodopa is routinely administered in combination with a peripheral decarboxylase inhibitor to prevent its peripheral metabolism to dopamine and the development of nausea, vomiting, and orthostatic hypotension due to activation of dopamine receptors in the area postrema (the nausea and vomiting center) that are not protected by the BBB. In the United States, levodopa is combined with the decarboxylase inhibitor carbidopa (Sinemet), whereas in many other countries, it is combined with benserazide (Madopar). Levodopa plus a decarboxylase inhibitor is also available in a methylated formulation, a controlled-release formulation (Sinemet CR or Madopar HP), and in combination with a catechol-*O*-methyltransferase (COMT) inhibitor (Stalevo). A long-acting formulation of levodopa (Rytary), a levodopa-carbidopa intestinal gel administered continuously by intra-intestinal infusion, continuous subcutaneous infusions of a levodopa formulation, and an inhaled form of levodopa that is absorbed through the pulmonary alveoli are also now available (see below).

Levodopa remains the most effective symptomatic treatment for PD and the gold standard against which new therapies are compared. No current medical or surgical treatment provides antiparkinsonian benefits superior to what can be achieved with levodopa. Levodopa benefits the classic motor features of PD, prolongs independence and employability, improves quality of life, and increases life span. Indeed, levodopa also benefits some "nondopaminergic" features such as anxiety, depression, and sweating. Almost all PD patients experience improvement, and failure to respond to an adequate trial of levodopa should cause the diagnosis to be questioned.

There are important limitations of levodopa therapy. Acute dopaminergic side effects include nausea, vomiting, and orthostatic hypotension. These are usually transient and can generally be avoided by starting with low doses and gradual up-titration. If they persist, they can be treated with additional doses of a peripheral decarboxylase inhibitor (e.g., carbidopa) or administered with food or a peripheral dopamine-blocking agent such as domperidone (not available in the United States). As the disease continues to progress, features such as falling, freezing, autonomic dysfunction, sleep disorders, and dementia may emerge that are not adequately controlled by levodopa. Indeed, these nondopaminergic features (especially falls and dementia) are the primary source of disability and the main reason for hospitalization and nursing home placement for patients with advanced PD in the levodopa era.

The major concern with levodopa is that chronic treatment with ongoing disease progression is associated in most patients with the development of motor complications. Motor complications consist of fluctuations in motor response ("on" episodes when the drug is working and "off" episodes when Parkinsonian features return as the drug wears off) and involuntary movements known as dyskinesias, which typically complicate "on" periods **(Fig. 446-6)**. When patients initially take levodopa, benefits are long-lasting (many hours and to some degree even weeks—the "long-duration" response) even though the drug has a relatively short half-life (60–90 min). With continued treatment, however, the duration of benefit following an individual dose becomes progressively shorter until benefits approach the half-life of the drug. This loss of benefit is known as the *wearing-off effect*. Some patients may also experience a rapid and unpredictable switch from the on to the off state known as the *on-off phenomenon*. In advanced cases, because of variability in the bioavailability of standard oral levodopa, the response to an individual dose of levodopa may be variable and unpredictable with the patient experiencing a full-on response, a partial-on response, a delay in turning on (delayed-on), or no response at all (no-on). Peak-dose dyskinesias occur at the time of levodopa peak plasma concentration and maximal clinical benefit. They are usually choreiform but can manifest as dystonic movements, myoclonus, or other movement disorders. They are not troublesome when mild but can be disabling when severe and can limit the ability to use higher doses of levodopa to better control PD motor features. In more advanced states, patients may cycle between "on" periods complicated by disabling dyskinesias

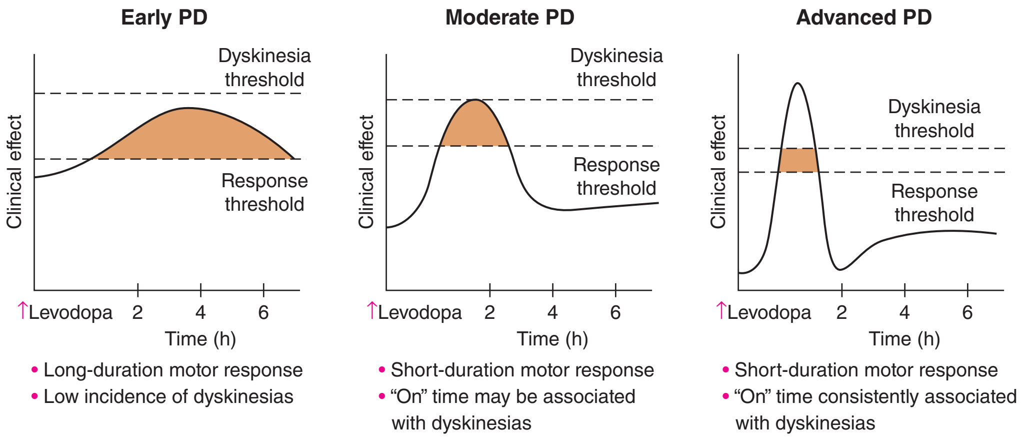

FIGURE 446-6 Changes in motor response associated with chronic levodopa treatment. Levodopa-induced motor complications. Schematic illustration of the gradual shortening of the duration of a beneficial motor response to levodopa (wearing off) and the appearance of dyskinesias complicating "on" time. PD, Parkinson's disease.

and "off" periods in which they suffer from severe parkinsonism and painful dystonic postures. Patients may also experience "diphasic dyskinesias," which occur with lower plasma levodopa levels and manifest as the levodopa dose begins to take effect and again as it wears off. These dyskinesias typically consist of transient, stereotypic, rhythmic movements that predominantly involve the lower extremities asymmetrically and are frequently associated with parkinsonism in other body regions. They can be relieved by increasing the dose of levodopa (although higher doses may induce peak-dose dyskinesia) and disappear as the concentration declines. Long-term double-blind studies show that the risk of developing motor complications can be minimized by using the lowest dose of levodopa that provides satisfactory benefit and through the use of polypharmacy to avoid the need for raising the dose of levodopa.

The precise cause of levodopa-induced motor complications is not known. They are more likely to occur in younger individuals, with the use of higher doses of levodopa, in females, and in those with more severe disease. The classic model of the basal ganglia has been useful for understanding the origin of motor features in PD, as noted above, but has proven less valuable for understanding levodopa-induced dyskinesias (Fig. 446-5). The model predicts that dopamine replacement might excessively inhibit the pallidal output system, thereby leading to increased thalamocortical activity, enhanced stimulation of cortical motor regions, and the development of dyskinesia. However, lesions of the pallidum that dramatically reduce its output are associated with amelioration rather than induction of dyskinesia as would be suggested by the classic model. It is now thought that dyskinesia results from alterations in the GPi/SNr neuronal firing pattern (pauses, bursts, synchrony, etc.) and not simply the firing frequency alone. This leads to the transmission of "misinformation" from pallidum to thalamus/cortex, which along with firing frequency contributes to the development of dyskinesia. Surgical or ultrasound lesions or high-frequency stimulation targeted at the GPi or STN presumably ameliorate dyskinesia by interfering with (blocking or masking) this abnormal neuronal activity and preventing the transfer of misinformation to motor systems.

A number of studies suggest that motor complications develop in response to nonphysiologic levodopa replacement. Striatal dopamine levels are normally maintained at a relatively constant level. In PD, where dopamine neurons and terminals have degenerated, striatal dopamine levels are dependent on the peripheral availability of levodopa. Intermittent oral doses of levodopa result in fluctuating plasma levels because of the short half-life of the drug and variability in the transit of the drug from the stomach to the jejunum where it is absorbed. These fluctuations are also reflected in the brain and result in striatal dopamine receptors being exposed to alternating pathologically high and low concentrations of dopamine. This in turn has been shown to induce molecular alterations in striatal neurons, neurophysiologic changes in pallidal output neurons, and ultimately the development of motor complications. It has been hypothesized that more continuous delivery of levodopa might be more physiologic and prevent the development or reduce the frequency of motor complications. Indeed, double-blind studies in PD patients have demonstrated that continuous intraintestinal infusion of levodopa/carbidopa and continuous subcutaneous infusion of apomorphine or levodopa are associated with significant improvement in "off" time and in "on" time without troublesome dyskinesia, compared with intermittent doses of standard oral levodopa. These benefits are superior to those observed in placebo-controlled studies with other dopaminergic agents. Intestinal infusion of levodopa is approved in the United States and Europe (Duodopa, Duopa). The treatment can, however, be complicated by potentially serious adverse events related to the surgical procedure, problems related to the tubing, and the inconvenience of having to wear an infusion system. Continuous subcutaneous delivery of levodopa or apomorphine avoids the need for a surgical procedure but is associated with a high frequency of cutaneous lesions and still requires wearing the inconvenient pump system. These are approved in Europe but not yet in the United States.

Behavioral complications can also be associated with levodopa treatment. A dopamine dysregulation syndrome has been described where patients have a craving for levodopa and take frequent and unnecessary doses of the drug in an addictive manner. (In this regard, it is noteworthy that cocaine binds to the dopamine uptake receptor.) PD patients taking high doses of levodopa can also develop purposeless, stereotyped behaviors such as the assembly and disassembly or collection and sorting of objects. This is known as punding, a term taken from the Swedish description of the meaningless behaviors seen in chronic amphetamine users. Hypersexuality and other impulse-control disorders are occasionally encountered with levodopa but are more commonly seen with dopamine agonists.

Finally, because levodopa undergoes oxidative metabolism and has the potential to generate toxic free radicals, there has been concern that independent of the drug's ability to provide symptomatic benefits, it might accelerate neuronal degeneration. Alternatively, as levodopa improves long-term outcomes in comparison to the pre-levodopa era, it has been suggested that by restoring striatal dopamine, levodopa has the potential to have a disease-modifying or neuroprotective effect. Neither of these hypotheses has been established. A recent delayed start study (explained below) showed neither beneficial nor deleterious effects of levodopa on the rate of clinical progression. Thus, it is generally recommended that levodopa be used solely based on its potential to provide symptomatic benefits balanced by the risk of inducing motor complications and other side effects.

DOPAMINE AGONISTS

Dopamine agonists are a diverse group of drugs that act directly on dopamine receptors. Unlike levodopa, they do not require metabolic conversion to an active product and do not undergo oxidative metabolism. Initial dopamine agonists were ergot derivatives (e.g., bromocriptine, pergolide) and were associated with potentially serious ergot-related side effects such as cardiac valvular damage and pulmonary fibrosis. They have largely been replaced by a second generation of non-ergot dopamine agonists (e.g., pramipexole, ropinirole, rotigotine). In general, dopamine agonists do not have comparable efficacy to levodopa. They were initially introduced as adjuncts to levodopa to enhance motor function and reduce "off" time in fluctuating patients. Subsequently, it was shown that dopamine agonists are less prone than levodopa to induce dyskinesia, possibly because they are relatively long acting in comparison to levodopa. For this reason, many physicians initiated therapy with a dopamine agonist, particularly in younger patients who are more prone to develop motor complications, although supplemental levodopa is eventually required in virtually all patients. This view has been tempered by the recognition that dopamine agonists are associated with potentially serious adverse effects such as unwanted sleep episodes and impulse control disorders (see below). Both ropinirole and pramipexole are available as orally administered immediate (three times a day) and extended-release (once a day) formulations. Rotigotine is administered as a once-daily transdermal patch and may be useful in managing surgical patients who are not able to be treated with an oral therapy. Apomorphine is the one dopamine agonist with efficacy thought to be comparable to levodopa, but it must be administered parenterally as it is rapidly and extensively metabolized if taken orally. It has a short half-life and duration of activity (45 min). It can be administered by subcutaneous injection as a rescue agent for the treatment of severe "off" episodes but can also be administered by continuous subcutaneous infusion where it has been shown to reduce both "off" time and dyskinesia in advanced patients. A sublingual bilayer formulation of apomorphine has been approved as a rapid and reliable therapy for individual "off" periods that avoids the need for a subcutaneous (SC) injection (see below).

Dopamine agonist use is associated with a variety of side effects. Acute side effects are primarily dopaminergic and include nausea, vomiting, and orthostatic hypotension. These can usually be avoided or minimized by starting with low doses and slowly up-titrating over weeks. Side effects associated with chronic use include hallucinations, cognitive impairment, and leg edema. Sedation with sudden unintended episodes of falling asleep that can occur in dangerous situations, such as while driving a motor vehicle, have been reported. Patients should be informed about this potential problem and should not drive when tired. Dopamine agonists can also be associated with impulse-control disorders, including pathologic gambling, hypersexuality, and compulsive eating and shopping. Patients should be advised of these risks and specifically questioned for their occurrence at follow-up examinations. The precise cause of these problems, and why they appear to occur more frequently with dopamine agonists than levodopa, remains to be resolved, but differential effects on reward systems associated with dopamine and alterations in the ventral striatum and orbitofrontal regions have been implicated. In general, chronic side effects are dose-related and can be avoided or minimized with lower doses. Injections of apomorphine can be complicated by skin lesions at sites of administration, which can be minimized by proper cleaning and alternating the injection sites. The sublingual bilayer formulation of apomorphine can be associated with oropharyngeal side effects, but these are generally mild and resolve either spontaneously or with treatment withdrawal. A selective D1 agonist has been developed and shown to have mild antiparkinsonian effects but not greater than those seen with other available dopamine agonists.

MAO-B INHIBITORS

Inhibitors of monoamine oxidase type B (MAO-B) block central dopamine MAO-B-oxidative metabolism and thereby increase synaptic concentrations of the neurotransmitter. Selegiline and rasagiline are relatively selective suicide inhibitors of the MAO-B isoform of the enzyme. Clinically, these agents provide antiparkinsonian benefits when used as monotherapy in early disease stages and reduced "off" time when used as an adjunct to levodopa in patients with motor fluctuations. MAO-B inhibitors are generally safe and well tolerated. They may increase dyskinesia in levodopa-treated patients, but this can usually be controlled by down-titrating the dose of levodopa. Inhibition of the MAO-A isoform prevents metabolism of tyramine in the gut, leading to a potentially fatal hypertensive reaction known as a "cheese effect" because it can be precipitated by foods rich in tyramine such as some cheeses, aged meats, and red wine. Currently available MAO-B inhibitors are selective, do not functionally inhibit the MAO-A enzyme, and are not associated with a cheese effect with doses used in clinical practice. There are theoretical risks of a serotonin reaction in patients receiving concomitant selective serotonin reuptake inhibitor (SSRI) antidepressants, but these are rarely encountered. Safinamide is a reversible and selective MAO-B inhibitor that has been approved as an adjunct to levodopa for treating advanced PD patients with motor fluctuations. The drug also acts to block activated sodium channels and inhibit glutamate release and therefore has the potential to provide antidyskinetic as well as ant-parkinsonian effects.

Interest in MAO-B inhibitors has also focused on their potential to have disease-modifying effects (see below).

COMT INHIBITORS

When levodopa is administered with a decarboxylase inhibitor, it is primarily metabolized in the periphery by the COMT enzyme. Inhibitors of COMT block its peripheral metabolism, increase the elimination half-life of levodopa, and enhance its brain availability. Combining levodopa with a COMT inhibitor reduces "off" time and prolongs "on" time in fluctuating patients while enhancing motor scores. The COMT inhibitors tolcapone and entacapone have been available for more than a decade; tolcapone is administered three times daily, while entacapone is administered in combination with each dose of levodopa. Opicapone, a long-acting COMT inhibitor that only requires once-daily administration, has more recently been approved in both Europe and the United States. A combination tablet of levodopa, carbidopa, and entacapone (Stalevo) is also available.

Side effects of COMT inhibitors are primarily dopaminergic (nausea, vomiting, increased dyskinesia) and can usually be controlled by down-titrating the dose of levodopa by 20–30% if required. Severe diarrhea has been described with tolcapone, and to a lesser degree with entacapone, and necessitates stopping the medication in 5–10% of individuals. Rare cases of fatal hepatic toxicity have been reported with tolcapone. It is still used because it is the most effective of the COMT inhibitors, but periodic monitoring of liver function is required. Liver problems have not been encountered with entacapone or opicapone. Discoloration of urine can be seen with COMT inhibitors due to accumulation of a metabolite, but it is of no clinical concern.

It has been proposed that initiating levodopa in combination with a COMT inhibitor to enhance its elimination half-life could provide more continuous levodopa delivery and reduce the risk of motor complications (see below). While this result has been demonstrated in a preclinical MPTP model of PD, and continuous infusion reduces both "off" time and dyskinesia in advanced PD patients, no benefit of initiating levodopa with a COMT inhibitor compared to levodopa alone was detected in early PD patients in the STRIDE-PD study. This may have been because the combination was not administered at frequent enough intervals to provide continuous levodopa availability. For now, the main value of COMT inhibitors continues to be as an adjunct to levodopa.

OTHER MEDICAL THERAPIES

Adenosine $A2_A$ receptor antagonists are a class of drugs that inhibit $A2_A$ receptors that form heterodimers with D2 dopamine receptors on medium spiny striatal D2-bearing neurons of the indirect pathway. Blockade of $A2_A$ receptors decreases the excessive activation of the indirect pathway in PD and theoretically restores balance in the basal ganglia–thalamocortical circuit, providing a dopaminergic effect without the need to increase levodopa doses and activate D1-receptor-bearing neurons that comprise the direct pathway. Three $A2_A$ antagonists have been studied in PD, but development in two has been discontinued: preladenant because it failed in phase 3 studies and tozadenant because of agranulocytosis in a few patients. Istradefylline is the only agent that is currently approved for use. Clinical trials in advanced PD patients showed improvement in "off" time comparable to other available agents but not in dyskinesia. The drug is generally well tolerated, with adverse events similar to dopaminergic agents. Interestingly, caffeine is a potent $A2_A$ antagonist, and epidemiologic studies suggest that drinking coffee is associated with a reduced frequency of PD. This has raised the question as to whether this class of agent might be neuroprotective, but this has not been established in clinical trials.

Amantadine was originally introduced as an antiviral agent, but the drug was observed to also have antiparkinsonian effects, likely due to antagonism of the *N*-methyl-D-aspartate (NMDA) receptor. While some physicians use amantadine in patients with early disease for its mild symptomatic effects, it is most widely used as an antidyskinesia agent in patients with advanced PD. Indeed, it is the only oral agent demonstrated in controlled studies to reduce dyskinesia without worsening parkinsonian features (indeed, motor benefits have been reported to be improved). Cognitive impairment is a major concern, particularly with high doses. Other side effects include livedo reticularis and weight gain. Amantadine should always be discontinued gradually because patients can experience withdrawal-like symptoms. An extended-release formulation of amantadine has also been developed.

Central-acting anticholinergic drugs such as trihexyphenidyl and benztropine were used historically for the treatment of PD, but they lost favor with the introduction of levodopa. Their major clinical effect is on tremor, although it is not certain that this benefit is superior to what can be obtained with agents such as levodopa and dopamine agonists. Still, they can be helpful in individual patients with severe tremor. Their use is limited particularly in the elderly, due to their propensity to induce a variety of side effects, including urinary dysfunction, glaucoma, and particularly cognitive impairment.

TABLE 446-5 Drugs Commonly Used for Treatment of Parkinson's Disease[a]

AGENT	AVAILABLE DOSAGES	TYPICAL DOSING
Levodopa[a]		
Carbidopa/levodopa	10/100, 25/100, 25/250 mg	200–1000 mg levodopa/day
Benserazide/levodopa	25/100, 50/200 mg	
Carbidopa/levodopa CR	25/100, 50/200 mg	
Benserazide/levodopa MDS	25/200, 25/250 mg	
Parcopa	10/100, 25/100, 25/250 mg	
Rytary (carbidopa/levodopa)	23.75/95, 36.25/145, 48.75/195, 61.25/245	See conversion tables
Carbidopa/levodopa/entacapone	12.5/50/200, 18.75/75/200, 25/100/200, 31.25/125/200, 37.5/150/200, 50/200/200 mg	
Dopamine agonists		
Pramipexole	0.125, 0.25, 0.5, 1.0, 1.5 mg	0.25–1.0 mg tid
Pramipexole ER	0.375, 0.75, 1.5. 3.0, 4.5 mg	1–3 mg/d
Ropinirole	0.25, 0.5, 1.0, 3.0 mg	6–24 mg/d
Ropinirole XL	2, 4, 6, 8 mg	6–24 mg/d
Rotigotine patch	2-, 4-, 6-, 8-mg patches	4–24 mg/d
Apomorphine SC	2–8 mg	2–8 mg
COMT inhibitors		
Entacapone	200 mg	200 mg with each levodopa dose
Tolcapone	100, 200 mg	100–200 mg tid
Opicapone	50 mg	50 mg HS
MAO-B inhibitors		
Selegiline	5 mg	5 mg bid
Rasagiline	0.5, 1.0 mg	1 mg QAM
Safinamide	100 mg	100 mg QAM
On-demand therapy for off periods		
Inhaled levodopa	5–40 mg	Up to 5 doses per day
Apomorphine sublingual strip		Up to 5 doses per day
Others		
$A2_A$ antagonist—Istradefylline	20, 40 mg	20 or 40 mg/d
Amantadine—immediate, extended-release	100–400 mg	

[a]Treatment should be individualized. Generally, drugs should be started in low doses and titrated to optimal dose.

Note: Drugs should not be withdrawn abruptly but should be gradually lowered or removed as appropriate.

Abbreviations: COMT, catechol-*O*-methyltransferase; MAO-B, monoamine oxidase type B; QAM, every morning.

The anticonvulsant zonisamide has also been shown to have mild antiparkinsonian effects and is approved for use in Japan. Its mechanism of action is unknown. Several classes of drugs are currently being investigated in an attempt to enhance antiparkinsonian effects, reduce "off" time, and treat or prevent dyskinesia. These include a selective inhibitor of the GPR6 receptor, which is localized to D2-bearing striatal neurons, and a selective antagonist of the D3 receptor.

A list of the major drugs and available dosage strengths currently available to treat PD is provided in **Table 446-5**.

CONTINUOUS DOPAMINERGIC DELIVERY

As noted above, there is evidence suggesting that motor complications are related to nonphysiologic restoration of brain DA with intermittent oral doses of short-acting levodopa formulations. To overcome these problems, several approaches have been developed to deliver levodopa in a more continuous manner. These include continuous intraintestinal and continuous subcutaneous delivery. Each of these has been shown to provide more stable plasma levodopa levels than intermittent doses of standard levodopa and to be associated with reduced "off" time and increased "on" time without troublesome dyskinesia. Similar results have also been seen with continuous subcutaneous delivery of apomorphine, as well as continuous oral delivery using a small intraoral micropump attached to a retainer. Attempts continue to develop an oral formulation of levodopa that can provide relatively continuous plasma levodopa levels and avoid a surgical procedure with resulting risk for cutaneous nodules and abscesses and the need to wear an inconvenient and cumbersome infusion pump.

ON-DEMAND THERAPIES FOR OFF PERIODS

Despite all available therapies including continuous delivery, many patients still experience "off" periods. Off periods can be disabling for patients, placing them at risk for falling and choking. As noted above, taking an additional levodopa tablet does not reliably treat individual off episodes, and some patients may continue in the off state for hours despite a levodopa dose. This inability to reliably and rapidly treat off episodes causes many patients to become depressed, withdrawn, and unwilling to participate in social or business activities. Three therapies have now been approved as specific on-demand treatments for off periods: inhaled levodopa, subcutaneous injection of apomorphine, and sublingual apomorphine. Each of these is fast acting, avoids the variable bioavailability seen with standard oral levodopa, and provides a predictable return to the on state.

NEUROPROTECTION

Despite the many therapeutic agents available for the symptomatic treatment of PD, patients continue to progress and to develop intolerable disability. A neuroprotective or disease-modifying therapy that slows or stops disease progression remains the major unmet therapeutic need. Some trials have shown positive results (e.g., selegiline, rasagiline, pramipexole, ropinirole) consistent with a disease-modifying effect. However, it has not been possible to determine with certainty if the positive results were due to neuroprotection with slowing of disease progression or confounding symptomatic or pharmacologic effects that mask disease progression. Interest has focused on selegiline and rasagiline, as MPTP toxicity can be prevented experimentally by coadministration of an MAO-B inhibitor. These agents block the oxidative conversion of MPTP to the pyridinium ion MPP^+ that is taken up by and selectively damages dopamine neurons. MAO-B inhibitors also have the potential to block the oxidative metabolism of dopamine and prevent oxidative stress. In addition, both selegiline and rasagiline incorporate a propargyl ring within their molecular structure that provides antiapoptotic effects in laboratory models.

In the classic DATATOP study, selegiline delayed the time until the emergence of disability necessitating the introduction of levodopa in untreated PD patients. However, it could not be definitively determined whether this benefit was due to a neuroprotective effect that slowed disease progression or a symptomatic effect that merely masked ongoing neurodegeneration. The ADAGIO study tested the putative neuroprotective effects of rasagiline using a two-period delayed-start design. In the first period, patients are randomized to treatment with the active drug or placebo. In the second period, patients in both groups receive the active treatment. If early treatment provides an enduring benefit that cannot be achieved with delayed treatment, the result is consistent with a disease-modifying effect. In ADAGIO, early treatment with rasagiline 1 mg/d provided significant benefits that could not be achieved when treatment with the same drug was initiated at a later time point, consistent with a disease-modifying effect. However, this benefit was not seen with the 2-mg dose, and it did not receive regulatory approval for this indication. The reason the 2-mg dose failed remains uncertain, but many physicians use rasagiline in early-stage patients based on its potential to have neuroprotective effects.

Neuroprotective therapies that prevent the formation or accumulation of toxic α-synuclein species, inhibit LRRK2, or enhance GCase are currently being studied. GLP-1 agonists, developed for use in diabetes, have also shown some promise based on anti-inflammatory and pro-mitochondrial actions, but results in double-blind studies have been inconsistent.

SURGICAL TREATMENT

Surgical treatments for PD have been used for more than a century. Lesions were initially placed in the motor cortex and improved tremor but were associated with motor deficits, and this approach was abandoned. Subsequently, it was appreciated that lesions placed into the ventral intermediate (VIM) nucleus of the thalamus reduced contralateral tremor without inducing hemiparesis, but these lesions did not meaningfully help other more disabling features of PD. In the 1990s, it was shown that lesions placed in the posteroventral portion of the GPi (motor territory) improved rigidity and bradykinesia as well as tremor. Importantly, pallidotomy was also associated with marked improvement in contralateral dyskinesia. This procedure gained favor with greater understanding of the pathophysiology of PD (see above). However, this procedure is not optimal because PD affects both sides of the body and bilateral lesions are associated with side effects such as dysphagia, dysarthria, and impaired cognition. Lesions of the STN are also associated with antiparkinsonian benefit and reduced levodopa requirement, but there is a concern about the risk of hemiballismus, and this procedure is not commonly performed.

Most surgical procedures for PD performed today use deep brain stimulation (DBS). Here, an electrode is placed into the target area and connected to a stimulator inserted subcutaneously over the chest wall. DBS simulates the effects of a lesion without needing to make a brain lesion. The precise mechanism whereby DBS works is not fully understood but may act by disrupting the abnormal neurophysiological signals that are associated with PD and motor complications. The stimulation variables can be adjusted with respect to electrode configuration, voltage, frequency, and pulse duration in order to maximize benefit and minimize adverse side effects. The procedure does not require making a lesion in the brain and is thus suitable for performing bilateral procedures with relative safety. In cases where there are no benefits or with intolerable side effects, stimulation can be stopped and the system removed.

DBS for PD is primarily used to target the STN or the GPi. It provides antiparkinsonian benefits, particularly with respect to tremor, and reduces both "off" time and dyskinesias but does not provide antiparkinsonian benefits that are superior to levodopa. The procedure is thus primarily indicated for patients who suffer disability from severe tremor or from levodopa-induced motor complications that cannot be satisfactorily controlled with drug adjustments. Side effects can result from the surgical procedure (hemorrhage, infarction, infection), the DBS system (infection, lead break, lead displacement, skin ulceration), or the stimulation itself (ocular and speech abnormalities, muscle twitches, paresthesias, depression, and rarely suicide). Results of DBS of the STN and GPi are comparable, but GPi stimulation may be associated with a reduced frequency of depression. Although not all PD patients are candidates, the procedure can be profoundly beneficial for the appropriate patient. Long-term studies demonstrate continued benefits with respect to the dopaminergic features of PD, but DBS does not prevent the development of nondopaminergic features, which continue to evolve as the disease progresses and are a source of disability. Studies continue to evaluate the optimal way to use DBS (e.g., low- vs high-frequency stimulation, closed loop systems, adaptive approaches). Trials of DBS in early PD patients show benefits that may be superior to best medical therapy, but this must be weighed against the cost of the procedure and the risk of side effects in patients who might otherwise be well controlled with relatively safe medical therapies for many years especially if used correctly (see below). Additionally, the PD landscape is changing with the availability of on-demand therapies for treating off periods and the likelihood that future therapies may provide continuous levodopa availability with a reduced risk of motor complications. Controlled studies comparing DBS to other therapies aimed at improving motor function without causing dyskinesia, such as continuous intraintestinal or SC levodopa infusions, remain to be performed. The utility of DBS may also be reduced in the future if new medical therapies are developed that provide the benefits of levodopa without motor complications. New targets for DBS are also being actively explored directed at gait dysfunction, depression, and cognitive impairment, as well as "smart" closed-loop devices that sense the patient's need for stimulation **(Chap. 500)**.

MRI-guided ultrasound is also now being used to target critical regions such as the GPi or STN in PD patients with motor complications in a relatively noninvasive manner that avoids the needs for a surgical procedure. Preliminary results suggest good target localization and safety. Ultrasound has also been used to interrupt the BBB in a specific location, which might facilitate access to the brain for therapies that otherwise might not cross the BBB.

OTHER EXPERIMENTAL THERAPIES FOR PD

These include cell-based therapies (e.g., transplantation of dopamine neurons derived from stem cells), gene therapies, and trophic factors. Transplant strategies are based on the concept of implanting dopaminergic cells into the striatum to replace degenerating SNc dopamine neurons. Fetal nigral mesencephalic cells have been demonstrated to survive implantation, re-innervate the striatum in an organotypic manner, and restore motor function in PD models. However, two double-blind studies failed to show significant benefit of fetal nigral transplantation in comparison to a sham operation. Grafting of fetal nigral cells is associated with a previously unrecognized form of dyskinesia (graft-induced dyskinesia) that persists after lowering or even stopping levodopa. This has been postulated to be related to suboptimal release of dopamine from grafted cells, leading to a sustained form of diphasic dyskinesia. In addition, there is evidence that, after many years, transplanted healthy embryonic dopamine neurons from unrelated donors develop PD pathology and become dysfunctional, suggesting transfer of α-synuclein from affected to unaffected neurons in a prion-like manner (see discussion above). There are also concerns about immune reactions to the injection of foreign tissue. Stem cells, and specifically autologous induced pluripotent stem (IPS) cells derived from the recipient, may overcome problems related to immune reactions and physiologic integration, but many of the concerns listed above still apply. To date, stem cells have not yet been properly tested in double-blind studies and bear the additional theoretical concern of malignant transformation and other unanticipated side effects. Importantly, it is not clear how replacing dopamine cells alone will improve the nondopaminergic features of PD such as falling and dementia, which are the major sources of disability for patients with advanced disease. While there remains a need for scientifically based studies to evaluate the potential role of cell-based therapies in PD, there is no basis for treating PD patients with stem cells in nonresearch studies, as is being marketed in some countries.

Trophic factors are a series of proteins that enhance neuronal growth and potentially could restore function to damaged neurons. Based on laboratory studies, several different trophic factors appear to have beneficial effects on dopamine neurons, and glial-derived neurotrophic factor (GDNF) and neurturin have attracted particular attention as possible therapies for PD. However, double-blind trials of intraventricular and intraputaminal infusions of GDNF failed to benefit PD patients, possibly because of inadequate delivery of the trophic molecule to the target region. Gene therapy offers the potential of providing long-term expression of a therapeutic protein with a single procedure. Gene therapy involves placing the nucleic acid of a therapeutic protein into a viral vector that can then be taken up and incorporated into the genome of host cells and then synthesized and released on a continual basis. The AAV2 virus has been most often used as the vector because it does not promote an inflammatory response, is not incorporated into the host genome, does not induce insertional mutagenesis, and is associated with long-lasting transgene expression. AAV2 delivery of the trophic factor neurturin (a member of the GDNF family) showed promising results in open-label trials but also failed in double-blind trials, even when injected into both the putamen and the SNc. Nonetheless, long-term postmortem studies have demonstrated transgene survival with biological effects as long as 10 years after treatment. However, the degree of putaminal coverage was very small, and it is likely that much higher gene doses will be required if this type of therapy is to provide clinically meaningful results.

Gene delivery is also being explored as a means of delivering aromatic amino acid decarboxylase with or without tyrosine hydroxylase into the striatum to facilitate the conversion of orally administered levodopa to dopamine. Animal studies suggest that this approach can provide antiparkinsonian benefits with reduced motor complications; clinical trials in PD patients are underway. Gene therapy is also being studied as a way to enhance *GBA1* and the gene product GCase in an attempt to promote lysosomal clearance of misfolded α-synuclein protein.

Importantly, no clinically significant adverse events have been encountered in gene therapy studies directed at the central nervous system to date, but there remains a risk of unanticipated side effects including mutagenesis. Further, it is not clear how current approaches directed at the dopamine system, even if successful, will address the nondopaminergic features of the illness.

MANAGEMENT OF NONMOTOR AND NONDOPAMINERGIC FEATURES OF PD

Although PD treatment has primarily focused on the dopaminergic features of the illness, management of the nondopaminergic features should not be ignored. Some nonmotor features benefit from dopaminergic drugs. For example, problems such as anxiety, panic attacks, depression, pain, sweating, sensory problems, freezing, and constipation all tend to be worse during "off" periods and have been reported to improve with better dopaminergic control. Recent studies with light therapy suggest that exposure to the specific light frequencies can restore a more normal circadian rhythm (which is altered in PD) and provide both motor and nonmotor benefits, particularly with respect to sleep and mood.

Approximately 50% of PD patients suffer depression during the course of the disease, and depression is frequently underdiagnosed and undertreated. Antidepressants should not be withheld, particularly for patients with major depression, although dopaminergic agents such as pramipexole may prove helpful for treating both depression and PD motor features. Anxiety is also a common problem, and if not adequately controlled with antiparkinsonian therapies, it can be treated with short-acting benzodiazepines.

Psychosis can be a problem for some PD patients and is often a harbinger of developing dementia. In contrast to AD, hallucinations in PD patients are typically visual, formed, and nonthreatening. Importantly, they can be associated with the use of dopaminergic drugs and may limit the use of these agents required for satisfactory motor control. Initial management is to withdraw agents that are less effective than levodopa, such as anticholinergics, amantadine, and dopamine agonists, followed by lowering the dose of levodopa if possible. Psychosis in PD often responds to low doses of atypical neuroleptics and may permit higher doses of levodopa to be tolerated. Clozapine is an effective drug, but it can be associated with agranulocytosis, and regular monitoring is required. Quetiapine avoids these problems, but it has not been established to be effective in placebo-controlled trials. Pimavanserin (Nuplazid) differs from other atypical neuroleptics in that it is an inverse agonist and antagonist of the serotonin 5-HT_{2A} receptor. It has been shown to be effective in short-term double-blind trials but has only mild efficacy (although it can be very effective in individual patients) and has been reported to be associated with QT prolongation and death in elderly patients.

Dementia in PD (PDD) is common, ultimately affecting as many as 80% of patients. Its frequency increases with aging and, in contrast to AD, primarily affects executive functions and attention, with relative sparing of language, memory, and calculation domains. When dementia precedes or develops within 1 year after onset of motor dysfunction, it is by convention referred to as dementia with Lewy bodies (DLB; **Chap. 445**). Interestingly, if dementia develops in a PD patient after 12 months, it is referred to as PD dementia, although it is not clear that these represent different disease entities. These patients are particularly prone to experience hallucinations and diurnal fluctuations. Pathologically, DLB is characterized by Lewy bodies distributed throughout the cerebral cortex (especially the hippocampus and amygdala) and is more likely to be associated with AD pathology. It is notable that variants of the *GBA1* gene are a significant risk factor for both PD and DLB. Mild cognitive impairment (MCI) frequently precedes the onset of dementia and is a more reliable index of impending dementia than in the general population as it occurs in the setting of a neurodegenerative disorder. Indeed, many PD patients demonstrate abnormalities in cognitive testing even at the earliest stages of the disease despite having no overt clinical dysfunction. Drugs used to treat PD can worsen cognitive function and should be stopped or reduced to try and provide a compromise between antiparkinsonian benefit and preserved cognitive function. Drugs are usually discontinued in the following sequence: anticholinergics, amantadine, dopamine agonists, COMT inhibitors, and MAO-B inhibitors. Eventually, patients with cognitive impairment should be managed with the lowest dose of standard levodopa that provides meaningful antiparkinsonian effects and does not worsen mental function. Anticholinesterase agents such as memantine and cholinesterase inhibitors such as rivastigmine improve measures of cognitive function and can improve attention in PD, but do not improve cognition or quality of life in any meaningful way. More effective therapies that treat or prevent dementia are a critical unmet need in the therapy of PD.

Autonomic disturbances are common and frequently require attention. Orthostatic hypotension can be problematic and contribute to falling. Initial treatment should include adding salt to the diet and elevating the head of the bed to prevent overnight sodium natriuresis. Low doses of fludrocortisone (Florinef) or midodrine provide control for most cases. The norepinephrine precursor 3-0-methylDOPS (Droxidopa) has been shown to provide mild but transient benefits for patients with orthostatic hypotension. Vasopressin and erythropoietin can be used in more severe or refractory cases. If orthostatic hypotension is prominent in early parkinsonian cases, a diagnosis of MSA should be considered **(Chap. 451)**. Sexual dysfunction may be helped with sildenafil or tadalafil. Urinary problems, especially in males, should be treated in consultation with a urologist to exclude prostate problems. Anticholinergic agents, such as oxybutynin (Ditropan), may be helpful. Constipation can be a very important problem for PD patients. Mild laxatives or enemas can be useful, but physicians should first ensure that patients are drinking adequate amounts of fluid and consuming a diet rich in bulk with green leafy vegetables and bran. Agents that promote GI motility can also be helpful. Several studies are evaluating the effect on constipation of agents that interfere with inflammation and α-synuclein misfolding in the GI tract.

Sleep disturbances are common in PD patients, with many experiencing fragmented sleep with excess daytime sleepiness. These can be severe and result in sudden-onset sleep episodes that may occur in dangerous situations such as while driving a car. These problems tend to be exaggerated by dopamine agonists, particularly in high doses. These problems may relate to alterations in circadian rhythm associated with degeneration in melanopsin-containing neurons in the retina and cells of the suprachiasmatic nucleus, which occur in PD patients. Recent studies suggest that both motor and nonmotor features may be improved with light therapy using specific wavelengths that restore circadian rhythm in PD patients.

Restless leg syndrome, sleep apnea, and other sleep disorders also occur with increased frequency in PD and should be treated as appropriate. REM behavior disorder (RBD) is a syndrome composed of violent movements and vocalizations during REM sleep, possibly representing acting out of dreams due to a failure of motor inhibition that typically accompanies REM sleep **(Chap. 33)**. Many PD patients have a history of RBD preceding the onset of the classic motor features of PD by many years, and most cases of RBD eventually go on to develop an α-synucleinopathy (PD or MSA). Low doses of clonazepam (0.5–1 mg at bedtime) are usually effective in controlling this problem. Consultation with a sleep specialist and polysomnography may be necessary to identify and optimally treat sleep problems. Excess daytime sleepiness can be problematic for PD patients, and therapies such as sodium oxybate (Xyrem) that are effective in narcolepsy are currently being evaluated in PD

NONPHARMACOLOGIC THERAPY

Gait dysfunction with falling is an important cause of disability in PD. Dopaminergic therapies may be of help for patients whose gait is worse in "off" time, but there are currently no specific therapies for gait dysfunction. Canes and walkers may become necessary to increase stability and reduce the risk of falling. An effective therapy for gait impairment is an important unmet need in PD.

Freezing, where patients suddenly become stuck in place for seconds to minutes as if their feet were glued to the ground, is another important problem and a major cause of falling. Freezing may occur during "on" or "off" periods. Freezing during "off" periods may respond to dopaminergic therapies, but there are no specific treatments for "on" period freezing and the mechanism is not well understood. Some patients will respond to sensory cues such as marching in place, singing a song, or stepping over an imaginary line or obstacle.

Speech impairment is another source of disability for many advanced PD patients. Speech therapy programs may be helpful, but benefits are generally limited and transient.

Exercise has been shown to help maintain and even improve function for PD patients, and active and passive exercises with full range of motion reduce the risk of arthritis and frozen joints. Some laboratory studies suggest the possibility that exercise might also have neuroprotective effects, but this has not been confirmed in PD patients. Exercise is generally recommended for all PD patients. It is less clear that any specific type of physical therapy or exercise program, such as tai chi or dance, offers any specific advantage. It is important for patients to maintain social and intellectual activities to the extent possible. Education, assistance with financial planning, social services, and attention to home safety are important elements of the overall care plan. Information is available through numerous PD foundations and on the web but should be reviewed with physicians to ensure accuracy. The needs of the caregiver should not be neglected. Caring for a person with PD involves a substantial work effort, and there is an increased incidence of depression among caregivers. Support groups for patients and caregivers may be useful.

CURRENT MANAGEMENT OF PD

The management of PD should be tailored to the needs of the individual patient, and there is no single treatment approach that is universally accepted and applicable to all individuals. Clearly, if an agent could be demonstrated to have disease-modifying effects, it should be initiated at the time of diagnosis or even in the premotor stage once that can be diagnosed with confidence. Recent studies suggest that striatal dopamine terminal degeneration may be complete within 4 years of diagnosis and thus limit the potential benefit of a therapy started after that time, even if it has been shown to have protective effects. Epidemiologic and pathologic studies suggest that constipation, RBD, and anosmia may represent premotor features of PD and, along with imaging of the dopamine system and biomarkers (see above), could permit diagnosis and the initiation of a disease-modifying therapy prior to the onset of the classical motor features of the disease. However, no therapy has yet been conclusively proven to be a disease-modifying agent, although as noted above, rasagiline 1 mg/d met all three prespecified primary endpoints consistent with such a benefit. For now, physicians must use their judgment in deciding whether or not to introduce a drug such as rasagiline for its possible disease-modifying effects based on available preclinical and clinical information.

The next important issue to address is when to initiate symptomatic therapy and which agent to use. Several studies suggest that it may be best to start therapy at the time of diagnosis in order to preserve beneficial compensatory mechanisms and possibly provide functional benefits with improved quality of life even in the early stage of the disease. Levodopa remains the most effective symptomatic therapy for PD, and the American Academy of Neurology recommends starting it immediately using low doses (≤400 mg/d), as motor complications have now clearly been shown to be dose-related. Other experts, however, prefer to delay introduction of levodopa treatment, particularly in younger patients, in order to reduce the risk of inducing motor complications. An alternate approach is to begin with am MAO-B inhibitor and/or a dopamine agonist and reserve levodopa for later stages when these drugs no longer provide satisfactory control. In making this decision, the patient's age, degree of disability, and the side effect profile of the drug must all be considered. In patients with more severe disability, the elderly, and those with cognitive impairment, significant comorbidities, or uncertain diagnosis, most physicians would initiate therapy with levodopa. A new pill that combines very low doses of rasagiline and pramipexole in extended-release formulations has been developed that provides clinical benefits comparable to high-dose pramipexole in higher doses but without sleep-related and dopaminergic side effects. As such, it could represent an alternative to levodopa as initial therapy for PD.

Regardless of initial choice, most patients ultimately benefit from polypharmacy (a combination of levodopa, an MAO-B inhibitor, and a dopamine agonist) in order to minimize the total daily levodopa dose and reduce the risk of motor complications. While it is important to use low doses of each agent to reduce the risk of side effects, patients should not be denied levodopa when they cannot be adequately controlled with alternative medications. It is also important to discuss the risks and benefits of the different therapeutic options with patients so that they have informed opinions as to whether they wish to start therapy early and, if so, which drug to start.

If motor complications develop, patients can initially be treated by adjusting the frequency and dose of levodopa or by combining lower doses of levodopa with a dopamine agonist, a COMT inhibitor, or an MAO-B inhibitor. An $A2_A$ antagonist such as istradefylline is an additional therapy that can be used for treating off periods. Amantadine is the only drug that has been demonstrated to treat dyskinesia without worsening parkinsonism, but benefits may decline over time and there are important side effects related to cognitive function particularly with higher doses. On-demand therapies such as subcutaneous apomorphine, inhaled levodopa, and sublingual apomorphine can be used to treat individual off periods and can delay the need for surgery in some patients. In advanced cases where patients suffer motor complications that cannot be adequately controlled with medical therapies, it may be necessary to consider a surgical procedure such as DBS or a continuous dopaminergic therapy such as Duodopa or subcutaneous infusion of levodopa or apomorphine, but as described above, these procedures have their own set of complications. The use of DBS in early PD patients has been advocated by some, but there is considerable skepticism about this approach considering the costs and potential side effects, when inexpensive, well-tolerated, and effective medical alternatives are available. Continuous intraintestinal infusion of levodopa/carbidopa intestinal gel (Duodopa) offers similar benefits to DBS, but also requires a surgical intervention with potentially serious complications. Continuous subcutaneous infusion of levodopa or apomorphine does not require surgery but is associated with potentially troublesome skin nodules and abscesses and requires wearing an inconvenient infusion pump during the course of the day and potentially around the clock. Comparative studies of these approaches are awaited. There are ongoing efforts aimed at developing a long-acting formulation of levodopa that mirrors the pharmacokinetic properties of a levodopa infusion. Such a formulation might provide all of the benefits of levodopa without motor

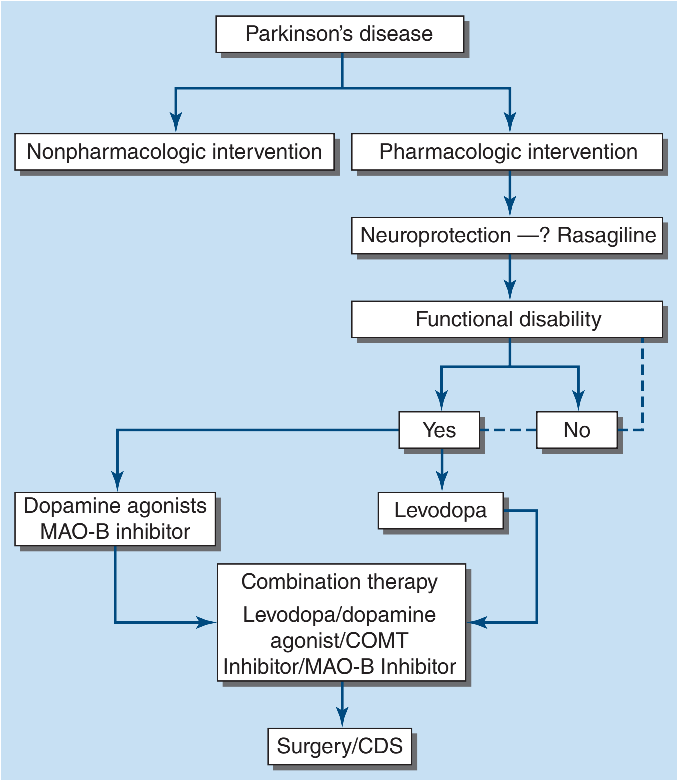
Parkinson's disease
Nonpharmacologic intervention
Pharmacologic intervention
Neuroprotection —? Rasagiline
Functional disability
Yes
No
Dopamine agonists
MAO-B inhibitor
Levodopa
Combination therapy
Levodopa/dopamine agonist/COMT Inhibitor/MAO-B Inhibitor
Surgery/CDS

FIGURE 446-7 Treatment options for the management of Parkinson's disease (PD). Decision points include: (1) Introduction of a neuroprotective therapy: no drug has been established to have or is currently approved for neuroprotection or disease modification, but there are several agents that have this potential based on laboratory and preliminary clinical studies (e.g., rasagiline 1 mg/d). (2) When to initiate symptomatic therapy: There is a trend toward initiating therapy at the time of diagnosis or early in the course of the disease because patients may have some disability even at an early stage, and there is the possibility that early treatment may preserve beneficial compensatory mechanisms; however, some experts recommend waiting until there is functional disability before initiating therapy. (3) What therapy to initiate: many experts favor starting with low doses of levodopa particularly in the elderly and those with more advanced disease. A monoamine oxidase type B (MAO-B) inhibitor may be preferred in mildly affected patients because of their good safety profile and the potential for a disease-modifying effect. Some prefer dopamine agonists for younger patients with functionally significant disability as they have a reduced risk of inducing motor complications. All patients eventually require levodopa, but it is generally recommended to employ polypharmacy using low doses of multiple drugs to avoid side effects associated with high doses of any one agent and minimize the risks of levodopa-induced motor complications. (4) Management of motor complications: motor complications are typically approached with combination therapy to try and reduce dyskinesia and enhance the "on" time. When medical therapies cannot provide satisfactory control, surgical therapies such as deep brain stimulation (DBS) or continuous infusion of levodopa/carbidopa or apomorphine can be considered. (5) Nonpharmacologic approaches: interventions such as exercise, education, and support should be considered throughout the course of the disease. CDS, continuous dopaminergic stimulation; COMT, catechol-*O*-methyltransferase. *(Reproduced with permission CW Olanow et al: Neurology 72:S1, 2009.)*

complications and avoid the need for polypharmacy and surgical intervention. Treatment for the nonmotor features of PD should be instituted as deemed appropriate, and exercise therapy is recommended for all patients.

A decision tree that considers the various treatment options and decision points for the management of PD is provided in **Fig. 446-7**.

FURTHER READING

Balestrino R, Schapira AHV: Parkinson disease. Eur J Neurol 27:27, 2020.

Ben-Shlomo Y et al: The epidemiology of Parkinson's disease. Lancet 403:283, 2024.

Berg D et al: MDS research criteria for prodromal Parkinson's disease. Mov Disord 12:1600, 2015.

Blauwendraat C et al: The genetic architecture of Parkinson's disease. Lancet Neurol 19:170, 2020.

Bloem BR et al: Parkinson's disease. Lancet 397:2284, 2021.

Marras C et al: Nomenclature of genetic movement disorders: Recommendations of the International Parkinson and Movement Disorder Society task force. Mov Disord 32:724, 2017.

Morris HR et al: The pathogenesis of Parkinson's disease. Lancet 403:293, 2024.

Obeso JA et al: Past, present and future of Parkinson's disease: A special essay on the 200th Anniversary of the Shaking Palsy. Mov Disord 32:1264, 2017.

Postuma RB et al: MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord 12:1591, 2015.

Schapira AHV et al: Non-motor features of Parkinson disease. Nat Rev Neurosci 18:446, 2017.

Siderowf A et al: Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using alpha-synuclein seed amplification: A cross-sectional study. Lancet Neurol 22:407, 2023.

447

Tremor, Chorea, and Other Movement Disorders

C. Warren Olanow[*], Christine Klein

HYPERKINETIC MOVEMENT DISORDERS

Hyperkinetic movement disorders are characterized by involuntary movements unaccompanied by weakness. The major clinical features are summarized in **Table 447-1**. The term is somewhat arbitrary and potentially misleading as hypokinetic disorders such as Parkinson's disease (PD) are often accompanied by tremor, while hyperkinetic disorders such as dystonia may be manifest as slow or restricted movement because of severe muscle contractions. Nonetheless, the terms continue to be used by convention. The major hyperkinetic movement disorders and the diseases with which they are associated are considered in this section.

TREMOR

CLINICAL FEATURES

Tremor is defined as an involuntary, rhythmic, oscillatory movement of a body part with alternating contraction of agonist and antagonist muscles. It can be most prominent at rest (rest tremor), on assuming a posture (postural tremor), on actively reaching for a target (kinetic or intention tremor), or on carrying out a movement (action tremor). Tremor may also be characterized based on its distribution, frequency, amplitude, and related neurologic dysfunction. Tremor is classified along two axes: Axis 1 covers the clinical characteristics and historical features (age at onset, family history, temporal evolution), tremor characteristics (body distribution, activation condition), associated signs (systemic, neurologic), and laboratory tests (electrophysiology, imaging). Axis 2 relates to the etiology of the tremor and distinguishes genetic, secondary, or idiopathic origins.

Essential tremor (ET) is characterized by a tremor that typically occurs while trying to sustain a posture and/or an action tremor that is noted when reaching toward a target. This contrasts with the resting tremor of PD (**Chap. 446**), which is characterized by a predominant resting tremor and is less pronounced with action. Cerebellar dysfunction is characterized by a kinetic tremor (brought out by trying to touch an object) and is usually associated with hypotonia and past pointing. Healthy individuals can have a physiologic tremor that typically manifests as a mild, high-frequency (10–12 Hz), postural, or

---

[*] Deceased.

TABLE 447-1 Hyperkinetic Movement Disorders

Tremor	Rhythmic oscillation of a body part due to intermittent muscle contractions
Dystonia	Involuntary, patterned, sustained, or repeated muscle contractions often associated with twisting movements and abnormal posture
Athetosis	Slow, distal, writhing, involuntary movements with a propensity to affect the arms and hands (this represents a form of dystonia with increased mobility)
Chorea	Rapid, semi-purposeful, graceful, dance-like nonpatterned involuntary movements involving distal or proximal muscle groups. When the movements are of large amplitude and predominant proximal distribution, the term *ballism* is used.
Myoclonus	Sudden, brief (<100 ms), jerk-like, arrhythmic muscle twitches
Tic	Brief, repeated, stereotyped muscle contractions that can often be suppressed for a short time. These can be simple and involve a single muscle group or complex and affect a range of motor activities.

action tremor typically affecting the upper extremities. This tremor is usually of no clinical consequence and often is only appreciated with an accelerometer or under stress. An enhanced physiologic tremor (EPT) can be seen in up to 10% of the population and tends to occur in association with lifting a weight, anxiety, fatigue, a metabolic disturbance (e.g., hyperthyroidism, electrolyte abnormalities), drugs (e.g., valproate, lithium), or toxins (e.g., caffeine, smoking, alcohol). Treatment is initially directed at control of any underlying disorder, and if necessary, it can often be improved with a beta blocker.

■ ESSENTIAL TREMOR

ET is the most common movement disorder, affecting ~1% of the population and 5% of those over 60 years (an estimated 5–10 million persons in the United States or Western Europe). It can present in childhood but dramatically increases in prevalence in those aged >70 years. ET is characterized by a high-frequency tremor (6–10 Hz) that predominantly affects the upper extremities. The tremor is most often manifest as a postural or action tremor and, in severe cases, can interfere with functions such as eating and drinking. It is typically bilateral and symmetric but may begin on one side and remain asymmetric. Patients with severe ET can have an intention tremor with overshoot and slowness of movement, along with mild ataxia, suggesting the possibility of a cerebellar origin. Tremor involves the head in ~30% of cases, voice in ~20%, tongue in ~20%, face/jaw in ~10%, and lower limbs in ~10%. Multiple body parts are involved in ~50% of cases. The tremor is characteristically improved by alcohol and worsened by stress. Usually, the neurologic examination is normal aside from tremor, but subtle impairment of coordination or tandem walking may be present, and disturbances of hearing, cognition, personality, mood, and olfaction have been described. The differential diagnosis includes dystonic tremor (see below) or PD. PD can usually be differentiated from ET because the former tends to be present primarily at rest and to be suppressed by a voluntary action. Further, PD is typically associated with bradykinesia with progressive slowing of sequential movements (sequence effect), rigidity, gait, postural instability, and other parkinsonian features. However, the examiner should be aware that PD patients may have a postural tremor and ET patients may develop a rest tremor, but these typically only begin after a latency of a few seconds (emergent tremor). In contrast to the micrographia of PD, ET patients have relatively large handwriting with evidence of tremor in writing samples. The examiner must be careful to identify tremor when assessing tone in order to distinguish the interruption of movement associated with tremor from the cogwheel rigidity found in PD.

■ ETIOLOGY AND PATHOPHYSIOLOGY

The etiology and pathophysiology of ET are not known. Approximately 50% of ET patients have a positive family history with an autosomal dominant pattern of inheritance. Linkage studies have detected possible loci in large ET families. Expansion of a GGC repeat in the human-specific *NOTCH2NLC* gene has been found to be associated with ET, but no independently confirmed causative gene has been identified to date. It is likely that there are several as yet undiscovered genes underlying ET that have thus far escaped detection because of the heterogeneity of the syndrome and the high frequency of ET in the population, likely resulting in a large number of phenocopies (i.e., family members with a similar clinical syndrome but not carrying the same causative mutation). The cerebellum and inferior olives have been implicated as possible sites of an altered "tremor pacemaker" based on the presence of cerebellar signs in ~10% of ET patients, as well as increased metabolic activity and blood flow in these regions in some patients. Some pathologic studies have described cerebellar pathology with a loss of Purkinje cells and axonal torpedoes, suggesting a neurodegenerative disease, but these findings remain controversial, and the precise pathologic correlate of ET remains to be defined. Interest has also focused on the possibility that ET is caused by degeneration of GABAergic cerebellar neurons with defects in neurotransmission. It is likely that multiple causes of ET will ultimately be identified.

■ TREATMENT

Many cases are mild, do not cause any functional impairment, and require no treatment other than reassurance. Occasionally, tremor can be severe and interfere with eating, writing, and activities of daily living. This is more likely to occur as the patient ages and is often associated with a reduction in tremor frequency. Beta blockers and primidone are the standard drug therapies for ET and are useful in ~50% of cases. Propranolol (20–120 mg daily, given in divided doses) is usually effective at relatively low doses, but higher doses may be needed in some patients. The drug is contraindicated in patients with bradycardia or asthma. Hand tremor tends to be most improved, while head tremor is often refractory. Primidone can be helpful but should be started at low doses (12.5 mg) and gradually increased as necessary (125–250 mg three times daily) to avoid sedation, nausea, and dizziness. Benefits have also been reported with gabapentin and topiramate, but these drugs have not been widely employed. Botulinum toxin injections may be helpful for limb or voice tremor, but treatment can be associated with muscle weakness. Surgical therapies targeting the ventro-intermediate (VIM) nucleus of the thalamus can be very effective for severe and drug-resistant cases. More recently, focal ultrasound (a procedure that does not require surgery) has also been shown to be an effective therapy against tremor in some cases of ET.

DYSTONIA

■ CLINICAL FEATURES

Dystonia is a movement disorder characterized by sustained or intermittent synchronous muscle contractions of agonist and antagonist muscles causing abnormal, often repetitive, painful movements and postures. Dystonic movements are typically patterned and twisting and may be associated with a "dystonic" tremor. This tremor can usually be distinguished from ET as the tremor is most pronounced when the body part is moved in the direction of the dystonia and relieved when the body part is moved in the direction opposite to the dystonia. Dystonia can range from minor contractions affecting only an individual muscle group (focal) to severe and disabling contractions with involvement of multiple muscle groups (i.e., multifocal, segmental, or generalized). Nonmotor features such as pain, depression, anxiety, and impaired sleep can be associated with, or even precede onset of, the dystonia. The frequency of dystonia is estimated to be about 30 per 100,000 but is likely to be higher because many cases are not recognized or correctly diagnosed. Dystonia is often brought out by voluntary movements (action dystonia) and can extend to involve other muscle groups and body regions not required for the intended action (overflow contractions). Dystonia can be aggravated by stress and fatigue and attenuated by relaxation and sensory tricks such as touching the affected body part (geste antagoniste).

Historically, dystonia has been described as primary or secondary. However, because of a confusing and not always congruent combination of phenotypic and etiologic features, the older terms are no longer

TABLE 447-2 Monogenic Forms of Isolated and Combined Dystonia

FORM OF DYSTONIA		GENE	DESIGNATION AND PHENOTYPIC SUBGROUP	ADDITIONAL DISTINGUISHING FEATURES	MODE OF INHERITANCE
Isolated[a]		*TOR1A*	DYT-TOR1A	Childhood or adolescent onset, generalized	AD
		KMT2B	DYT-KMT2B	Early onset, generalized, mild syndromic features	AD
		THAP1	DYT-THAP1	Adolescent onset, cranial or generalized	AD
		ANO3	DYT-ANO3	Adult onset, focal or segmental	AD
		GNAL	DYT-GNAL	Mostly adult onset, focal or segmental	AD
		VPS16	DYT-VPS16	Frequent cervical and laryngeal dystonia	AD or AR
		EIF2AK2	DYT-EIF2AK2	Childhood or adolescent onset, focal to generalized	AD or AR
		PRKRA	DYT-PRKRA	Generalized	AR
		HPCA	DYT-HPCA	Childhood onset	AR
		AOPEP	DYT-AOPEP	Frequent cervical and laryngeal dystonia	AR
Combined[b]	**Dystonia plus parkinsonism**	*GCH1*	DYT-GCH1	Dopa-responsive	AD
		TAF1	DYT-TAF1	Neurodegeneration	XL
		ATP1A3	DYT-ATP1A3	Rapid onset	AD
	Dystonia plus myoclonus	*SGCE*	DYT-SGCE	Alcohol responsive	AD
		KCTD17	DYT-KCTD17	Childhood onset	AD

[a]*KMT2B* pathogenic variants may present with mild syndromic features. [b]Selected examples.

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; XL, X-linked.

recommended. A Movement Disorder Society Task Force recommended classifying dystonia along the same axes as ET: clinical and etiologic. On clinical grounds, dystonia can be categorized by age of onset (infancy, childhood, adolescence, early and late adulthood), body distribution (focal, segmental, multifocal, and generalized), temporal pattern (static or progressive, action-specific [diurnal and paroxysmal]), and association with additional features. Clinical description along these lines enables formulating specific dystonia syndromes (e.g., early-onset generalized isolated dystonia). From an etiologic point of view, dystonia primarily reflects genetic abnormalities, although occasionally it may be secondary to other causes, such as trauma and stroke. Genetic features used for classification include mode of inheritance or identification of a specific pathogenic gene variant. More than 200 genes have been linked to different types of dystonia, primarily childhood-onset and generalized forms. These include forms in which dystonia is the only disease manifestation with the exception of tremor ("isolated dystonia") and forms in which dystonia co-occurs with another movement disorder such as parkinsonism, myoclonus, or other neurologic and/or nonneurologic manifestations ("combined dystonia") and may not even be the dominant clinical feature. This group represents the most heterogeneous class in terms of clinical expression. A list of confirmed monogenic mutations associated with isolated or combined dystonias is provided in Table 447-2.

■ ISOLATED DYSTONIAS

Focal, Multifocal, and Segmental Dystonia Adult-onset, focal dystonia is by far the most frequent form of isolated dystonia. Women are affected about twice as often as men, with the exception of writer's cramp, which occurs more frequently in men than in women. Focal dystonia typically presents in the fourth to sixth decade. The major clinical phenotypes are: (1) *Cervical dystonia*—dystonic contractions of neck muscles causing the head to deviate to one side (*laterocollis*), twist (*torticollis*), move in a forward direction (*anterocollis*), or move in a backward direction (*retrocollis*). Muscle contractions can be painful and occasionally can be complicated by a secondary cervical radiculopathy and even myelopathy. (2) *Blepharospasm*—dystonic contraction of the eyelids resulting in increased blinking and eye closure that can interfere with reading, watching television, working on a computer, and driving. This can sometimes be so severe as to cause functional blindness. (3) *Oromandibular dystonia* (OMD)—contractions of muscles of the lower face, lips, tongue, and jaw (opening or closing). Meige's syndrome is a combination of OMD and blepharospasm that predominantly affects women aged >60 years. (4) *Spasmodic dysphonia*—dystonic contractions of the vocal cords during phonation, causing impaired speech. Most cases affect the adductor muscles and cause speech to have a choking or strained quality. Less commonly, the abductors are affected, leading to speech with a breathy or whispering quality. (5) *Limb dystonias*—these can be present in either arms or legs and are often brought out by task-specific activities such as handwriting (writer's cramp), playing a musical instrument (musician's cramp), or putting in golf (the yips). The vast majority of patients with focal dystonia have cervical dystonia (~40%) or blepharospasm (~15%). Focal hand or leg dystonia (~10%), musician's dystonia (~3%), spasmodic dysphonia (~2%), and OMD (~1%) are much less common. Focal dystonias can extend to involve other body regions (~30% of cases) and are frequently misdiagnosed as psychiatric or orthopedic in origin. Their cause is usually not known. They are rarely monogenic (~1%); autoimmunity and trauma have been suggested as other possible etiologies. Focal dystonias are often associated with a high-frequency tremor that can resemble ET. Dystonic tremor can usually be distinguished from ET because it tends to occur in the direction of the dystonic contraction and disappears when the dystonia is relieved (i.e., turning the head in the opposite direction of the dystonia).

Generalized Dystonia Generalized dystonia is often hereditary and, unlike focal dystonia, typically has an age of onset in childhood or adolescence. There are currently at least 10 well-established genes that, when mutated, can cause mostly isolated, segmental or generalized dystonia: *ANO3, AOPEP, EIF2AK2, GNAL, HPCA, KMT2B, PRKRA, THAP1, TOR1A,* and *VPS16*. The *AOPEP, HPCA*, and *PRKRA* pathogenic mutations are recessively inherited, while others (e.g., *EIF2AK2* and *VPS16*) can be either dominantly or recessively inherited. According to the recommendations of the International Parkinson's Disease and Movement Disorder Society, monogenic forms of dystonia are classified according to the absence or presence of accompanying additional clinical features and preceded by a "DYT" prefix, e.g., DYT-TOR1A.

Mutations in the *TOR1A* gene (torsin family 1 member A—formerly known as the *DYT1* gene) are the most common cause of early-onset generalized dystonia. The first, and currently the only, clearly established mutation is a 3-base pair deletion in the *TOR1A* gene. The mutation is frequently found among Ashkenazi Jewish patients due to a founder effect. Mutation carriers usually present with dystonia in an extremity in childhood that later progresses to affect other body parts, but the face and neck are typically spared. Rare carriers of two mutated alleles have been described and are characterized by a severe neurodevelopmental syndrome and arthrogryposis.

Missense mutations in *KMT2B* (*lysine methyltransferase 2B*) are another relatively frequent cause of early-onset generalized dystonia, which may be accompanied by other syndromic features, including intellectual disability, microcephaly, psychiatric features, dysmorphia, or skin lesions. The majority of the mutations occur de novo. *KMT2B* mutations may account for up to 10% of early-onset generalized dystonia, but further validation is warranted, and placement into the group of isolated versus complex dystonias is currently under debate.

Mutations in the *THAP1* gene (*THAP domain containing apoptosis-associated protein 1*) have been linked to adolescent-onset dystonia with mixed phenotype. About 100 different mutations have been reported in *THAP1*. Mutations typically manifest with dysphonia or writer's cramp beginning in late childhood or adolescence. Over the course of the disease, dystonia can spread to other body parts with prominent craniocervical involvement. While DYT-Tor1A and DYT-KMT2B typically respond well to deep brain stimulation (DBS) of the globus pallidus internus bilaterally, the DBS response is much more variable in carriers of pathogenic variants in the *THAP1* gene.

Mutations in the *ANO3* gene (*anoctamin 3*) were first reported in patients with predominantly craniocervical dystonia with a broad range of ages of onset. While a large number of missense variants can be found in healthy individuals, a pathogenic role of *ANO3* mutations has been confirmed by the description of additional families with dystonia and myoclonic jerks.

Mutations in the *GNAL* gene (*guanine nucleotide-binding protein subunit alpha L*) are a rare cause of cervical or cranial dystonia, with a few patients developing generalized dystonia. The mean age of onset is in the thirties.

Pathogenic variants in the *VPS16* gene can be inherited in a recessive or dominant fashion, with the latter mode of inheritance being more common. Currently, >30 carriers of ~20 different, often truncating, heterozygous pathogenic variants have been described. The median age of onset is 14 years. Dystonia tends to generalize and is typically isolated, although more complex phenotypes have also been described.

Pathogenic missense variants in *EIF2AK2*, coding for the eukaryotic translation initiation factor 2-alpha kinase 2, cause dystonia with a median age at onset of 6 years and onset often in the limbs followed by generalization. There is a recurrent missense variant (p.Gly130Arg) in most patients. The EIF2AK2 protein is one of the kinases responsible for eIF2a phosphorylation and is thus linked to the same pathway as PRKRA.

The vast majority of *PRKRA* mutation carriers develop a generalized dystonia, frequently with laryngeal involvement. Likewise, all patients described to carry *HPCA* mutations are characterized by generalized dystonia with childhood onset.

The median age at onset of carriers of recessively inherited, biallelic, and typically truncating pathogenic variants in the *AOPEP* gene is 20 years, with frequent onset in the hands. Most patients progress to isolated, generalized dystonia.

COMBINED DYSTONIAS

A number of other well-established genes have been described that are associated with combined forms of dystonia in which dystonia occurs in conjunction with a different movement disorder (e.g., parkinsonism or myoclonus) or with other neurologic and/or nonneurologic features.

Dopa-responsive dystonia (DRD; also known as Segawa syndrome) is caused by mutations in the *GCH1* gene (*GTP cyclohydrolase-1*) that encodes for the rate-limiting enzyme in the biosynthesis of dopamine via the biopterin pathway. It manifests as a childhood-onset form of dystonia with diurnal fluctuations, and it is important to recognize as the condition dramatically responds to low doses of levodopa. Parkinsonism can be a major or even the only finding, and there may be a presynaptic dopaminergic deficit as evidenced by single-photon emission computed tomography. Younger patients are frequently misdiagnosed as having cerebral palsy, mistaking dystonia for spasticity, and it is important that young-onset forms of dystonia should be tested with levodopa to exclude the possibility of DRD. To date, >100 different mutations have been reported with a penetrance of ~50% and incidence considerably higher in women compared to men. Recessively inherited (biallelic) mutations in *GCH1* result in a much more severe clinical phenotype with developmental delay and infantile onset. Due to the enzymatic defect in the levodopa biosynthesis, there is a lifelong and dramatic response to levodopa therapy. Importantly, since dopamine terminals do not degenerate and the dopamine neuronal network is anatomically preserved, fluctuations in dopamine levels can be avoided, and accordingly, these patients do not develop dyskinesia with chronic levodopa treatment.

X-linked dystonia-parkinsonism (Lubag) presents with a combined form of dystonia and parkinsonism that is found exclusively in patients of Filipino origin due to a founder effect that seems to be fully penetrant. The typical presentation is a focal (cranial) dystonia that rapidly generalizes and, after 5–10 years, is gradually replaced by a form of L-dopa-unresponsive parkinsonism. A retrotransposon insertion in the *TAF1* (*TATA-box binding protein associated factor 1*) gene is the cause of the disease. Sixty-five percent of the age-at-onset variability is explained by the variable length of a hexameric repeat expansion within the retrotransposon and genotypes at three single-nucleotide polymorphisms in the *MSH3* and *PMS1* genes acting as age-at-onset modifiers.

Mutations in the *ATP1A3* (*ATPase Na*$^+$/*K*$^+$ *transporting subunit alpha 3*) gene present with a characteristic, sudden-onset dystonia, usually in adolescence or young adulthood, often triggered by high fever, physical exertion, or emotional stress. Dystonic symptoms frequently show a rostrocaudal gradient with a strong involvement of the bulbar region, often accompanied by parkinsonian features such as bradykinesia. In addition, mutations in *ATP1A3* have been linked to a variety of clinical syndromes (pleiotropy), including epileptic or hemiplegic attacks, ataxia, cognitive decline, and other neurologic disorders, often with a more severe course and an earlier age at onset.

Myoclonic-dystonia is characterized by action-induced, alcohol-responsive myoclonic jerks predominantly involving the upper body half. Onset is usually in childhood or adolescence. Many individuals also develop psychiatric features such as depression, anxiety-related disorders, and alcohol dependence. The disorder is primarily related to mutations in the *SGCE* gene (*sarcoglycan epsilon*), which codes for the ε member of the sarcoglycan family. About 80 different mutations have been reported in *SGCE*, including deletions of the entire gene. The latter type of mutation often also involves loss of adjacent genes, leading to additional clinical features such as joint problems. *SGCE* mutations are incompletely penetrant and only manifest when inherited from the father due to the epigenetic effect of maternal imprinting of *SGCE*. *KCTD17* mutations are another recently identified cause of myoclonus-dystonia.

A number of additional monogenic causes have been suggested for isolated and combined forms of dystonia but still await independent confirmation. Table 447-2 provides a list of the confirmed monogenic forms of isolated and combined dystonias.

Diagnostic Considerations In the largest group of combined dystonias, dystonia is a part of a more complex syndrome that is characterized by multiple different clinical manifestations of the disease. Most frequently, they are hereditary, such as Wilson's disease (WD), Lesch-Nyhan syndrome, corticobasal ganglionic disorders, and a variety of other neurologic, neurometabolic, neurodevelopmental, and mitochondrial disorders. Dystonia may also develop as a consequence of drugs or toxins. Drug-induced dystonia may be acute or chronic and is most commonly seen with neuroleptic drugs or after chronic levodopa treatment in PD patients. Dystonia can also be observed following discrete lesions in the striatum (e.g., caudate/putamen) and occasionally in the globus pallidus, thalamus, cortex, or brainstem due to infarction, hemorrhage anoxia, trauma, tumor, infection, or toxins such as manganese or carbon monoxide. In these cases, dystonia often assumes a segmental distribution but may be generalized when lesions are bilateral or widespread. More rarely, dystonia can develop following peripheral nerve injury and be associated with features of complex regional pain syndrome **(Chap. 19)**. A psychogenic origin is responsible for some cases of dystonia; these typically present with fixed, immobile dystonic postures (see below).

PATHOPHYSIOLOGY OF DYSTONIA

Even in cases with a known dystonia gene mutation, the pathophysiologic basis of dystonia is not completely known. Physiologically, dystonia is characterized by co-contracting synchronous bursts of agonist and antagonist muscle groups with recruitment of muscle groups that are not required for a given movement (overflow). Dystonia is characterized by derangement of the basic physiologic principle of action selection, leading to abnormal recruitment of inappropriate muscles for a given action with inadequate inhibition of this undesired motor activity. Loss of surround inhibition is observed at multiple levels of the motor system (e.g., cortex, brainstem, spinal cord), accompanied by increased cortical excitability and reorganization. Attention has focused on the basal ganglia as the site of origin of most types of dystonia, as there are alterations in blood flow and metabolism in these structures. Further, lesions of the basal ganglia (particularly the putamen) can induce dystonia, and surgical ablation or DBS of specific regions of the globus pallidus may ameliorate dystonia. The dopamine system has been specifically implicated because dopaminergic therapies can both induce and treat some forms of dystonia in different circumstances. However, no specific pathology has been consistently identified that underlies dystonia.

TREATMENT

Dystonia

Treatment of acute dystonia should include the immediate withdrawal of any precipitating agent. A variety of drug therapies may be beneficial including diphenhydramine, benztropine, benzodiazepines, or dopamine agonists.

Treatment of chronic dystonia is for the most part symptomatic except in rare cases where correction of a primary underlying condition is possible. WD should be ruled out, particularly in young patients with dystonia. Levodopa should be tried in all cases of childhood-onset dystonia to test for DRD. High-dose anticholinergics (e.g., trihexyphenidyl 20–120 mg/d) may be beneficial in children, but adults can rarely tolerate high doses because of side effects related to cognitive impairment and hallucinations. Oral baclofen (20–120 mg) may also be helpful, but benefits, if present at all, are usually modest, and side effects of sedation, weakness, and memory loss can be problematic. Intrathecal infusion of baclofen is more likely to be useful, particularly for leg and trunk dystonia, but benefits are frequently not sustained, and complications can be serious and include infection, seizures, and coma. Tetrabenazine is another consideration; the usual starting dose is 12.5 mg/d and the average treating dose is 25–75 mg/d, but its use may be limited by sedation and the development of parkinsonism. Parkinsonian side effects can be minimized with deuterated tetrabenazine (discussed below). Neuroleptics can both improve and induce dystonia, but they are typically not recommended because of their potential to induce parkinsonism and other movement disorders, including tardive dystonia. Clonazepam and diazepam are sometimes effective.

Botulinum toxin is the preferred treatment for patients with focal and segmental dystonia, particularly where involvement is limited to small muscle groups such as in blepharospasm, torticollis, and spasmodic dysphonia. Botulinum toxin acts by blocking the release of acetylcholine at the neuromuscular junction, leading to reduced dystonic muscle contractions. However, treatment with botulinum toxin can be complicated by excessive weakness that can be troublesome, particularly if injections involve the neck and swallowing muscles. No systemic side effects are encountered with the doses typically used, but benefits are transient, and repeat injections are typically required at 2- to 4-month intervals. Some patients fail to respond after having experienced an initial benefit. This has been attributed to the development of neutralizing antibodies, but improper muscle selection, injection technique, and inadequate dose should be excluded. A new long-acting formulation of botulinum toxin (daxibotulinumtoxinA-l) that provides benefits for up to 6 months has recently been approved in the United States for the treatment of cervical dystonia.

Surgical therapy is a consideration for patients with severe generalized dystonia who are not responsive to other treatments. Peripheral procedures such as rhizotomy and myotomy were used in the past to treat cervical dystonia but are now rarely employed. DBS of the pallidum can provide dramatic benefits for some patients with various forms of hereditary and nonhereditary generalized dystonia. This represents a major therapeutic advance because previously there was no consistently effective therapy, especially for patients with generalized dystonia and severe disability. Benefits tend to be obtained with a lower frequency of stimulation than used in PD or ET and often occur only after a relatively long latency (weeks to months). Better results are typically obtained in younger patients with shorter disease duration and in those with certain monogenic forms, such as DYT-Tor1A. DBS may also be valuable for patients with focal and secondary dystonias, although results are less consistent. Neurophysiologic studies suggest that DBS acts by suppressing theta oscillations in the basal ganglia network that correlate with the dystonia. Adverse effects of DBS in dystonia patients include paresthesia, dysarthria, gait disturbance, and mood change. Dyskinesia can occur with stimulation of the subthalamic nucleus (STN), while bradykinesia and impaired coordination have been reported with stimulation of the globus pallidus pars interna (GPi). Focal ultrasound is being assessed as a possible alternative to surgical therapy. Supportive treatments such as physical therapy and education should be a part of the treatment regimen for all types of dystonia.

Physicians should be aware of dystonic storm, a rare but potentially fatal condition that can occur in response to a stress situation such as a surgical procedure or a systemic infection in patients with preexisting dystonia. It consists of the acute onset of generalized and persistent dystonic contractions that can involve the vocal cords or laryngeal muscles leading to airway obstruction. Patients may experience rhabdomyolysis with renal failure and should be managed in an intensive care unit with airway protection if required. Treatment can be instituted with one or a combination of anticholinergics, diphenhydramine, baclofen, benzodiazepines, and dopaminergic agents. In severe cases, anesthesia with muscle paralysis may be required.

CHOREAS

HUNTINGTON'S DISEASE

Huntington's disease (HD) is a progressive, fatal, highly penetrant autosomal dominant disorder characterized by motor, behavioral, oculomotor, and cognitive dysfunction. The disease is named for George Huntington, a family physician who described cases on Long Island, New York, in the nineteenth century. Onset is typically between the ages of 25 and 45 years (range, 3–70 years) with a prevalence of 2–8 cases per 100,000 and an average age at death of 60 years. It is prevalent in Europe, North America, South America, and Australia but is rare in African blacks and Asians. HD is characterized by rapid, nonpatterned, semi-purposeful, involuntary choreiform movements, and for this reason was formerly referred to as Huntington's chorea. However, dysarthria, gait disturbance, parkinsonism, oculomotor abnormalities, behavioral disturbance, and cognitive impairment with dementia are also common clinical features. Thus, the condition is currently referred to as Huntington's disease. In the early stages, chorea tends to be focal or segmental but progresses over time to involve multiple body regions. With advancing disease, there tends to be a reduction in chorea and the emergence of dystonia, rigidity, bradykinesia, and myoclonus. Functional decline is often predicted by progressive weight loss despite adequate calorie intake. In younger patients (~10% of cases), HD can present as an akinetic-rigid parkinsonian syndrome known as the Westphal variant. Eye movement abnormalities may be an early manifestation of HD. These include slowed and reduced amplitude saccades with intrusions in smooth pursuit movements and impaired convergence. HD patients eventually develop behavioral and cognitive disturbances, and the majority progress to dementia. Depression with suicidal tendencies, aggressive behavior, and psychosis can be prominent features. HD patients may also develop non-insulin-dependent diabetes mellitus

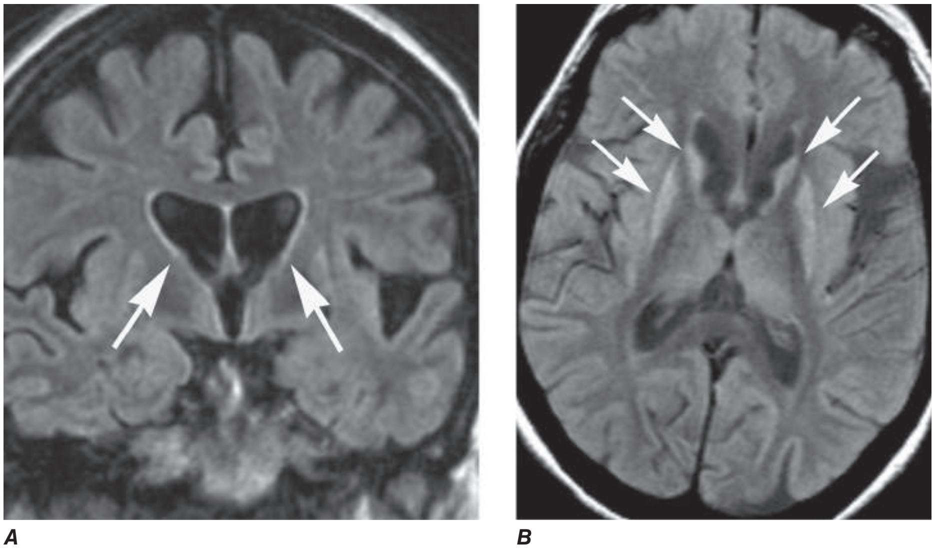

FIGURE 447-1 Huntington's disease. ***A.*** Coronal fluid-attenuated inversion recovery (FLAIR) magnetic resonance imaging shows enlargement of the lateral ventricles reflecting typical atrophy (*arrows*). ***B.*** Axial FLAIR image demonstrates abnormal high signal in the caudate and putamen (*arrows*).

and neuroendocrine abnormalities (e.g., hypothalamic dysfunction). A clinical diagnosis of HD can be strongly suspected in cases of chorea with a positive family history, but genetic testing provides the ultimate confirmation of the diagnosis.

The disease predominantly affects the striatum but progresses to involve the cerebral cortex and other brain regions. Progressive atrophy of the head of the caudate nucleus, which forms the lateral margin of the lateral ventricles, can be readily visualized on magnetic resonance imaging (MRI) **(Fig. 447-1)**, but the putamen can be equally or even more severely affected. More diffuse cortical atrophy can be seen in the middle and late stages of the disease. Supportive studies include reduced metabolic activity in the caudate nucleus and putamen and reduced brain metabolites on magnetic resonance spectroscopy. Genetic testing can be used to confirm the diagnosis and to detect at-risk individuals in the family, but it must be performed in conjunction with trained counselors because advising patients of positive results can worsen depression and even generate suicidal reactions. Indeed, genetic counseling is a requirement in some regions. The neuropathology of HD consists of prominent neuronal loss and gliosis in the caudate nucleus and putamen; similar changes can also be widespread in the cerebral cortex. Intraneuronal inclusions containing aggregates of ubiquitin and the mutant protein huntingtin are found in the nuclei of some affected neurons.

In anticipation of developing neuroprotective therapies, an intensive effort has been made to define the earliest stage of HD. Subtle motor impairment, cognitive alterations, and imaging changes can be detected in at-risk individuals who later develop the manifest form of the disease. Defining the rate of progression of these features is paramount for future studies of putative disease-modifying therapies designed to slow the rate of disease progression and the development of cumulative disability.

ETIOLOGY

HD is caused by a mutation in which there is an increase in the number of polyglutamine (CAG) repeats (>40) in the coding sequence of the *Huntingtin* gene located on the short arm of chromosome 4. The larger the number of repeats, the earlier the disease is manifest. Intermediate forms of the disease with 36–39 repeats are described in some patients, typically with less severe clinical involvement and reduced penetrance (i.e., not every mutation carrier develops the disease). These clinic-genetic observations have recently been incorporated into a new research classification of HD. Carriers of a fully penetrant (>40 repeats) allele are defined as having stage 0 HD, irrespective of their affected status. These developments are meant to facilitate the inclusion of at-risk individuals or those in the earliest disease stages into clinical trials.

Expansion of repeat length tends to occur, particularly in males, with subsequent generations having larger numbers of repeats and earlier age of disease onset, a phenomenon referred to as anticipation. There is also evidence of postnatal somatic gene expansion that occurs over time as well as genetic modifiers of disease progression, for example, in the *FAN1* and *MSH3* genes, the latter overlapping with X-linked dystonia-parkinsonism. New-onset gene expansion in patients with no family history is rare.

The *Huntingtin* gene encodes the highly conserved cytoplasmic protein huntingtin (HTT), which is widely distributed in neurons throughout the central nervous system (CNS). Mutated HTT RNA is toxic and disrupts transcription, impairs immune and mitochondrial function, and is aberrantly modified posttranslationally. Genome-wide association studies have nominated DNA repair pathways as modifiers of somatic instability and disease course in HD. Fragments of the mutant HTT can also be toxic, possibly by translocating into the nucleus and interfering with transcriptional regulation of proteins. Neuronal inclusions found in affected regions in HD may represent a protective mechanism aimed at segregating and facilitating the clearance of these toxic proteins. There is also interest in the possibility that the accumulation and aggregation of toxic proteins in HD, like Alzheimer's disease **(Chap. 442)** and PD **(Chap. 446)**, may be critical to the disease process and reflect a prion-like disorder **(Chap. 449; see also Chap. 435)**. Models of HD with striatal pathology can be induced in multiple transgenic animals that express the mutant gene and by excitotoxic agents such as kainic acid and 3-nitropropionic acid, which promote calcium entry into the cell and cytotoxicity. These relevant animal models can be helpful in assessing potential therapeutic agents. Interestingly, when correcting the neurotransmitter deficit present in HD mice in the very first week of life, disease development can be prevented.

TREATMENT

Huntington's Disease

Although the gene for HD was identified >30 years ago, there is still no disease-modifying therapy for this disorder, and clinically meaningful symptomatic treatment is limited. Current treatment involves a multidisciplinary approach, with medical, neuropsychiatric, and social approaches, as well as genetic counseling for patients and their families. Dopamine-blocking agents such as tetrabenazine and valbenazine have been approved to treat the choreiform movements but can be associated with secondary parkinsonism as an adverse event. Deuterated tetrabenazine (Austedo) and a long-acting version of deuterated tetrabenazine have also been approved as treatments for chorea in HD. Deuteration interferes with the metabolism of tetrabenazine and avoids the high maximum concentration (C_{max}), which is thought to contribute to adverse effects. In clinical trials, deuterated tetrabenazine has been shown to have fewer dose-related side effects and less parkinsonism than tetrabenazine and, therefore, can be administered in higher doses with potentially superior clinical benefits. Neuroleptics are generally not recommended because of their potential to induce other troubling movement disorders and because HD chorea tends to be self-limited and is usually not disabling. These drugs may be used, however, in patients with severe and disabling chorea. There are currently no therapies approved for treating the more disabling motor features of HD, but large-scale platform studies are ongoing, which are assessing a variety of different therapeutic approaches using a common placebo group. Depression and anxiety can be major problems, and patients should be treated with appropriate antidepressant and antianxiety drugs and monitored for mania and suicidal ideations. Psychosis can be treated with atypical antipsychotics such as clozapine (50–600 mg/d), quetiapine (50–600 mg/d), and risperidone (2–8 mg/d).

A neuroprotective therapy that slows or stops disease progression is the major unmet medical need. Some strategies are designed to reduce the formation or accumulation of mutant HTT. These largely focus on inhibiting mRNA synthesis either by blocking transcription (zinc finger motif protein), preventing posttranscriptional processes, promoting early mRNA degradation (antisense oligonucleotides [ASO]), or inhibiting translation with short interfering RNA. The most advanced of these experimental therapeutic approaches investigated intrathecal administration of an ASO in patients with early HD in a randomized, placebo-controlled, double-blind clinical trial. While a dose-dependent reduction in concentrations of mutant HTT was observed and there were no side effects, the study was terminated early because no clinical benefit was detected. Drugs that enhance mitochondrial function and increase the clearance of defective mitochondria are also being tested as possible disease-modifying therapies. Other investigative approaches include immunotherapy, dietary supplements (resveratrol), lipid-lowering medication (fenofibrate), anaplerotic therapy (triheptanoin), and DBS of the GPi. A promising therapy is the sigma 1 receptor agonist pridopidine. Various clinical trials have suggested that the drug may provide benefit with respect to total motor function and total functional capacity, particularly in patients with relatively mild disease. Double-blind studies are currently underway. Preliminary clinical trials testing cell-based therapies (stem cells and fetal striatal cells) have been initiated, aimed at replacing damaged striatal neurons, but efficacy and safety of these procedures have not yet been determined. Experimentally, there is great interest in the potential of using CRISPR (gene editing) techniques to target and destroy or prevent the formation of mutant *Huntingtin* RNA and to reduce accumulation of the abnormal protein. Numerous other molecular and gene-based approaches are being evaluated to interfere with the formation and promote the clearance of the toxic protein, and many clinical studies are anticipated to begin within the next few years.

HUNTINGTON'S DISEASE–LIKE DISORDERS

A group of rare inherited conditions that can mimic HD, designated HD-like (HDL) disorders, have also been identified. HDL-1, -2, and -4 are autosomal dominant conditions that typically present in adulthood. HDL-3 is recessively inherited, presents in early childhood, and differs markedly from HD and the other HDLs. HDL-1 is due to expansion of an octapeptide repeat in *PRNP*, the gene encoding the prion protein **(Chap. 449)**. Thus, HDL-1 is properly considered a prion disease. Patients exhibit onset of personality change in the third or fourth decade, followed by chorea, rigidity, myoclonus, ataxia, and epilepsy. HDL-2 manifests in the third or fourth decade with a variety of movement disorders, including chorea, dystonia, or parkinsonism, and dementia. Most patients are of African descent. Acanthocytosis can sometimes be seen in these patients, and this condition must be distinguished from neuroacanthocytosis (below). HDL-2 is caused by an abnormally expanded CTG/CAG trinucleotide repeat expansion in the *junctophilin-3* (*JPH3*) gene. The pathology of HDL-2 consists of intranuclear inclusions immunoreactive for ubiquitin and expanded polyglutamine repeats. HDL-4, the most common condition in this group, is caused by expansion of trinucleotide repeats in *TBP*, the gene that encodes the TATA box-binding protein involved in regulating transcription; this condition is identical to spinocerebellar ataxia (SCA) 17 **(Chap. 450)**, and most patients present with ataxia rather than chorea. Like in HD, a certain range of repeat expansions in the *TBP* gene is associated with reduced penetrance, whereby penetrance was recently identified to be full when a variant in the *Stub1* gene is present in conjunction with the repeat expansion. Mutations of the *C9Orf72* gene associated with frontotemporal dementia and amyotrophic lateral sclerosis **(Chaps. 443 and 448)** have also been reported in some individuals with an HDL phenotype.

■ OTHER CHOREAS

Chorea can be seen in a number of other disorders related to genetic mutations or other disease states.

Among the hereditary forms of childhood-onset chorea, mutations in the *NKX2-1* gene cause a benign hereditary chorea. Mutations in the *ADCY5* (adenylate cyclase 5) gene are an increasingly recognized and relatively common cause of childhood-onset chorea, often in combination with dystonia and developmental delay. Characteristic perioral movements are a hallmark of the disorder. Notably, patients respond well to caffeine.

Chorea-acanthocytosis (neuroacanthocytosis) is a progressive and typically fatal autosomal recessive disorder that is characterized by chorea coupled with red cell abnormalities on peripheral blood smear (acanthocytes). The chorea can be severe and associated with self-mutilating behavior, dystonia, tics, seizures, and a polyneuropathy. Mutations in the *VPS13A* gene encoding chorein have been described. A phenotypically similar X-linked form of the disorder has been described in older individuals who have reactivity with Kell blood group antigens (McLeod syndrome). A benign hereditary chorea of childhood (BHC1) due to mutations in the gene for thyroid transcription factor 1 and a late-onset benign senile chorea (BHC2) have also been reported. It is important to do genetic testing in these patients to ensure that they do not have HD.

Chorea may also occur in association with a variety of infections and degenerative disorders as well as vascular diseases and hypo- and hyperglycemia. *Sydenham's chorea* (originally called St. Vitus's dance) is more common in females and is typically seen in childhood (5–15 years). It often develops in association with prior exposure to group A streptococcal infection **(Chap. 153)** and is thought to be autoimmune in nature. It is characterized by the acute onset of choreiform movements and behavioral disturbances. With the reduction in the incidence of rheumatic fever, the incidence of Sydenham's chorea has fallen, but it can still be seen in developing countries. The chorea generally responds to dopamine-blocking agents, valproic acid, and carbamazepine, and is usually self-limited. Treatment is generally restricted to those with severe chorea. Chorea may recur in later life, particularly in association with pregnancy (chorea gravidarum) or treatment with sex hormones. Several reports have documented cases of chorea associated with *N*-methyl-ᴅ-aspartate (NMDA) receptor antibody–positive encephalitis, following herpes simplex virus encephalitis, and in paraneoplastic syndromes associated with anti-CRMP-5 or anti-Hu antibodies **(Chap. 99)**. *Systemic lupus erythematosus* **(Chap. 368)** is the most common systemic disorder that is associated with chorea. The chorea may last for only a few days but can be long-lasting and persist for years. Chorea can also be seen with hyperthyroidism, autoimmune disorders including Sjögren's syndrome, infectious disorders including HIV, metabolic alterations, and polycythemia rubra vera. Chorea has been described following open-heart surgery in the pediatric population. It may also occur in association with many medications (especially anticonvulsants, cocaine, CNS stimulants, estrogens, and lithium). In particular, chorea is commonly seen as a side effect of chronic levodopa treatment in patients with PD **(Chap. 446)**.

BALLISM/HEMIBALLISMUS

Ballism is a violent form of choreiform movement composed of wild, flinging, large-amplitude movements most frequently affecting proximal limb muscles on one side of the body (hemiballism). The movements may only affect one limb (monoballism) or, more exceptionally, both upper or lower limbs (paraballism). The movements may be so severe as to cause exhaustion, dehydration, local injury, and, in extreme cases, death. Fortunately, dopamine-blocking drugs can be very helpful, and importantly, hemiballismus is usually self-limiting and tends to resolve spontaneously after weeks or months. The most common cause is a partial lesion (infarct or hemorrhage) in the STN, but in 30–40% of cases, the lesion is found in the putamen, thalamus, or parietal cortex. In extreme cases, pallidotomy or DBS of the GPi can be effective and abolish the involuntary movements. Interestingly, surgically induced lesions and DBS of the STN in PD patients are usually not associated with hemiballismus.

TICS

A *tic* is a brief, rapid, recurrent, stereotyped and seemingly purposeless motor contraction. Motor tics can be simple, with movement only affecting an individual muscle group (e.g., blinking, twitching of the nose, jerking of the neck), or complex, with coordinated involvement of multiple muscle groups (e.g., jumping, sniffing, head banging, and echopraxia [mimicking movements]). Phonic (or vocal) tics can also be simple (e.g., grunting) or complex (e.g., echolalia [repeating other people's words], palilalia [repeating one's own words], and coprolalia [expression of obscene words]). Patients may also experience sensory tics, composed of unpleasant focal sensations in the face, head, or neck. These can be mild and of little clinical consequence or severe and disabling. Tics may present in adulthood and can be seen in association with a variety of disorders, including PD, HD, trauma, dystonia, drugs (e.g., levodopa, neuroleptics), and toxins.

TOURETTE'S SYNDROME

Tourette's syndrome (TS) is a neurobehavioral disorder named after the French neurologist Georges Gilles de la Tourette. It predominantly affects males, and the prevalence is estimated to be 0.03–1.6%, but it is likely that many mild cases do not come to medical attention. TS is characterized by multiple motor tics, often accompanied by vocalizations (phonic tics). Patients characteristically can voluntarily suppress tics for short periods of time but then experience an irresistible urge to express them. Tics vary in intensity and may be absent for days or weeks only to recur, occasionally in a different pattern. Tics tend to present between ages 2 and 15 years (mean 7 years) and often lessen or even disappear in adulthood, particularly in males. Associated behavioral disturbances include anxiety, depression, attention-deficit hyperactivity disorder (ADHD), and obsessive-compulsive disorder. Patients may experience personality disorders, self-destructive behaviors, difficulties in school, and impaired interpersonal relationships.

Etiology and Pathophysiology TS has a high heritability and is thus thought to be a genetic disorder, but no specific monogenic cause has yet been identified. Current evidence supports a complex inheritance pattern with an important contribution of de novo, likely gene-disrupting variants. Genome-wide linkage studies have suggested *Slit*, *Trk-like 1,* and *histidine decarboxylase* (*HDC*) genes as conferring genetic risk for TS. The risk of a family with one affected child having a second one with TS is ~25%. The pathophysiology of TS is not known, but alterations in dopamine neurotransmission, opioids, and second-messenger systems have been proposed.

TREATMENT

Tics and Tourette's Syndrome

There is no cure for tics or TS. Patients with mild disease often only require education and counseling (for themselves and family members). In a high proportion of patients, the severity of tics wanes in adult life, becoming less of a medical problem, thus arguing for conservative management if possible during the first decades of life. Drug treatment to help control tics is indicated when the tics are disabling and interfere with quality of life and social interactions. Therapy is individualized, and few treatment regimens have been properly evaluated in double-blind trials. Some physicians use the α-agonist clonidine, starting at low doses and gradually increasing the dose and frequency until satisfactory control is achieved. Guanfacine (0.5–2 mg/d) is an α-agonist that is preferred by some because it only requires once-daily dosing. Other physicians prefer to use neuroleptics, but treatment can be complicated by tardive dyskinesia and weight gain. Atypical neuroleptics are usually used initially (risperidone, olanzapine, ziprasidone) because they are thought to be associated with a reduced risk of tardive dyskinesia. If they are not effective, low doses of classical neuroleptics such as haloperidol, fluphenazine, pimozide, or tiapride can be tried because the risk of tardive dyskinesia in young people is relatively low. Tetrabenazine and deuterated tetrabenazine may be recommended but are associated with depression. The dopamine D1 antagonist ecopipam was reported to provide benefits without serious side effects in a controlled trial of short duration in TS. Antiepileptic drugs such as topiramate may provide benefit for some patients. Botulinum toxin injections can be effective in controlling focal tics that involve small muscle groups. There is also interest in the potential value of a wrist device that delivers electrical pulses and has been reported to reduce the frequency and severity of tics in an open-label study. The potential value of closed-loop DBS targeting the anterior portion of the internal capsule, the GPi, or the thalamus is currently being explored for severely affected patients. A large-scale public database and registry for DBS in TS has been established. Behavioral features, and particularly anxiety and compulsions, can be a disabling feature of TS and should be treated as appropriate. Behavioral and speech therapies are also occasionally employed but have not been formally tested. ADHD medications such as methylphenidate are sometimes used to increase attention and concentration but may also exacerbate tics.

MYOCLONUS

Myoclonus is a brief, rapid (<100 ms), shock-like, jerky movement consisting of single or repetitive muscle discharges that can occur with or without a pattern and with a variable frequency. Myoclonic jerks can be focal, multifocal, segmental, or generalized and can occur spontaneously, in association with voluntary movement (action myoclonus), or in response to an external stimulus (reflex myoclonus). Myoclonic jerks can be severe and interfere with normal movement or benign and of no clinical consequence, as is commonly observed in normal people who can experience myoclonic jerks when waking up or falling asleep (hypnagogic jerks). Negative myoclonus consists of a brief loss of muscle activity (e.g., asterixis in hepatic failure). Palatal myoclonus (or palatal tremor) involves contractions of the soft palate and may be associated with an audible click that can be disturbing to the patient and family members. This is usually idiopathic and benign but can be related to a lesion in the cerebellum or brainstem. Myoclonus may also occur consequent to injury to a peripheral or cranial nerve (e.g., hemifacial spasm).

Myoclonic jerks differ from tics in that they are typically not repetitive, are not suppressible, and can severely interfere with normal voluntary movement. They can be associated with abnormal neuronal discharges in cortical, subcortical, brainstem, or spinal cord regions, particularly in cases related to hypoxemia (especially following cardiac arrest), encephalopathy, and neurodegeneration. Reversible myoclonus can be seen with metabolic disturbances (renal failure, electrolyte imbalance, hypocalcemia), toxins, and many medications. Hereditary myoclonus syndromes can be grouped into three classes based on clinical features: prominent myoclonus, prominent myoclonus combined with another prominent movement disorder, and disorders that usually present with other phenotypes but can also manifest as a prominent myoclonus syndrome. An additional movement disorder is seen in nearly all myoclonus syndromes, most commonly ataxia or dystonia. Furthermore, cognitive decline and epilepsy are present in the majority of patients. The frequent association with epilepsy suggests that a brief epileptic-like discharge could underlie myoclonus in some situations.

Myoclonic epilepsy is a disorder comprised of myoclonus and epilepsy. It can be associated with other focal neurologic deficits, has a variable but progressive course, and may ultimately be fatal. The most common form of action myoclonus of cortical origin with generalized epilepsy is myoclonic epilepsy or Unverricht-Lundborg disease (EPM-1). Ataxia may also be a feature. This is an autosomal recessive disease caused by pathogenic variants in the *CSBT* gene. Other causes are Lafora body epilepsy or progressive myoclonic epilepsy (PME-2) caused by mutations in the *EPM2A* or *NHLRC1* genes. Neuronal ceroid lipofuscinosis (Batten's disease) is another consideration. In patients with less severe or absent epilepsy, mitochondrial disorders and neurodegenerative disorders affecting the cerebellum (i.e., SCAs) should be considered. Essential myoclonus is a relatively benign familial condition characterized by multifocal, very brief, lightning-like movements that are frequently alcohol sensitive. Mutations in the *epsilon-sarcoglycan* gene have been associated with myoclonus seen in association with dystonia (myoclonus-dystonia).

The precise cause of myoclonus is not known but, in some cases, is thought to be due to overexcitability or impaired inhibition of cortical or peripheral nerve stimuli related to a particular movement. Imaging studies are seeking to define the altered connectivity in neuronal circuits that underlies myoclonus.

TREATMENT
Myoclonus

Treatment primarily consists of managing the underlying condition or removing an offending agent. Pharmacologic therapy involves one or a combination of GABAergic agents such as valproic acid (800–3000 mg/d), piracetam (8–20 g/d), clonazepam (2–15 mg/d), levetiracetam (1000–3000 mg/d), or primidone (500–1000 mg/d). Treatment may be associated with striking clinical improvement in chronic cases in which a cortical origin for the myoclonic discharges has been identified (e.g., postanoxic myoclonus, progressive myoclonic epilepsy). In some cases, combinations of drugs may prove helpful. The serotonin precursor 5-hydroxytryptophan (plus carbidopa) may be useful in cases of postanoxic myoclonus. DBS can be highly effective in myoclonus-dystonia. Botulinum toxin has been used successfully in some patients with focal myoclonus, palatal myoclonus, and hemifacial spasm. Some patients with hemifacial spasm have also been reported to benefit from neurosurgical decompression of the involved facial nerve.

DRUG-INDUCED MOVEMENT DISORDERS

This important group of movement disorders is primarily associated with drugs that block dopamine receptors (neuroleptics) or central dopaminergic transmission. These drugs are widely used in psychiatry, but it is important to appreciate that drugs used in the treatment of nausea or vomiting (e.g., prochlorperazine [Compazine]) or gastroesophageal disorders (e.g., metoclopramide [Reglan]) are neuroleptic agents and can cause these disorders. Hyperkinetic movement disorders secondary to neuroleptic drugs can be divided into those that present acutely, subacutely, or after prolonged exposure (tardive syndromes). Dopamine-blocking drugs can also be associated with a reversible parkinsonian syndrome for which anticholinergics are often concomitantly prescribed, but these drugs are not effective antiparkinsonian agents and are associated with cognitive side effects, and there is concern that such treatment might actually increase the risk of developing a tardive syndrome.

ACUTE

Dystonia is the most common acute hyperkinetic drug reaction (see above). It is typically generalized in children and focal in adults (e.g., blepharospasm, torticollis, or OMD). The reaction can develop within minutes of exposure and can be successfully treated in most cases with parenteral administration of anticholinergics (benztropine), diphenhydramine, benzodiazepines (lorazepam, clonazepam, or diazepam), or dopamine agonists. The abrupt onset of severe spasms may occasionally be confused with a seizure; however, there is no loss of consciousness, and no automatisms, electroencephalogram abnormalities, or postictal features typical of epilepsy. The acute onset of chorea, stereotypic behavior, and tics may also be seen, particularly following exposure to CNS stimulants such as methylphenidate, cocaine, or amphetamines. In rare cases, the airway may be affected and must be protected.

SUBACUTE

Akathisia is the most common reaction in this category. Akathisia consists of motor restlessness with a need to move that is alleviated by movement. It is most frequently associated with use of neuroleptic drugs and generally starts within 2 weeks of initiating therapy. It can also be seen with calcium channel blockers, antiemetics, cocaine, and sedatives. The cause is not known but is thought to relate to blocking the dopaminergic system. Therapy consists of lowering the dose or removing the offending agent. When this is not possible, symptoms may be ameliorated with benzodiazepines, anticholinergics, beta blockers, or dopamine agonists. Treatment is generally effective, but in chronic cases, it may become associated with anxiety, depression, and even suicide.

TARDIVE SYNDROMES

These disorders develop months to years after initiation of a neuroleptic agent. Tardive dyskinesias (TD) are most common and typically present with choreiform movements involving the mouth, lips, and tongue. In severe cases, the trunk, limbs, and respiratory muscles may also be affected. In approximately one-third of patients, TDs remit within 3 months of stopping the drug, and most patients gradually improve over the course of several years. However, abnormal movements may also develop, persist, or worsen after stopping the offending agent. The movements are often mild and more upsetting to the family than to the patient, but in some cases, they can be severe and disabling, particularly in the context of an underlying psychiatric disorder. Second-generation or atypical antipsychotics (e.g., clozapine, risperidone, olanzapine, quetiapine, ziprasidone, and aripiprazole) are thought to be associated with a lower risk of causing TD in comparison to traditional antipsychotics, although they do not eliminate this risk. Younger patients have a lower risk of developing neuroleptic-induced TD, whereas the elderly, females, and those with underlying organic cerebral dysfunction are at greater risk. Chronic neuroleptic use is associated with increased risk of TD, and the U.S. Food and Drug Administration has specifically warned that use of metoclopramide for >12 weeks increases the risk of TD. Because TD can be permanent and resistant to treatment, antipsychotics should be used judiciously and atypical neuroleptics should be the preferred agent whenever possible, although there are now questions as to the risk of TD with atypical neuroleptics as well. In all patients on these agents, the need for continued use should be regularly evaluated. The cause of TD is not known with certainty, but it is thought to be related to hypersensitivity of dopamine receptors following the use of dopamine D2–blocking agents. This concept is based on observations that acute discontinuation can lead to accentuation of TD, while higher doses of these agents or the introduction of more potent neuroleptics can alleviate symptoms (at least transiently). Another hypothesis is that structural changes at the receptor level due to toxic effects of the neuroleptic may be causative. It has also been suggested that there may be a genetic predisposition in individuals who develop TD.

Treatment primarily consists of tapering and withdrawal of the offending agent. If the patient is receiving a traditional antipsychotic, and withdrawal is not possible, replacement with an atypical antipsychotic (e.g., clozapine) should be tried. Abrupt cessation of a neuroleptic should be avoided because acute withdrawal can induce worsening. TD can persist after withdrawal of antipsychotics and can be difficult to treat. Tetrabenazine, a vesicular monoamine transporter type 2 (VMAT-2) inhibitor that blocks storage of dopamine, has been used to treat TD but is short-acting and is associated with a dose-related new onset or worsening of parkinsonian features. Valbenazine, an ester of tetrabenazine, has been approved for the treatment of tardive dyskinesia in a dose of 80 mg/d based on efficacy in double-blind trials, but it is associated with sleepiness and QT prolongation. Deuterated tetrabenazine has also been approved for this indication. Deuteration provides a longer half-life with lower C_{max}, reducing risk of parkinsonian side effects. It can be individually titrated and permits the use of higher doses with lower risk of side effects. In open-label studies, benefits have also been reported with valproic acid (750–3000 mg/d), anticholinergics, and botulinum toxin injections. Other approaches include baclofen (40–80 mg/d) and clonazepam (1–8 mg/d). In some refractory cases, pallidal DBS may be an option.

Chronic neuroleptic exposure can also be associated with a tardive dystonia, with preferential involvement of axial muscles and characteristic rocking movements of the trunk and pelvis. Gray coloration of skin can be a clue that the patient is receiving a neuroleptic in patients for whom the cause of the dystonia is not obvious. Tardive dystonia can be more troublesome than tardive dyskinesia and frequently persists despite stopping medication. Valproic acid, anticholinergics, and botulinum toxin may occasionally be beneficial, but patients are frequently refractory to medical therapy. Tardive akathisia, tardive TS, and tardive tremor syndromes are rare but may also occur after chronic neuroleptic exposure.

Neuroleptic medications can also be associated with a neuroleptic malignant syndrome (NMS). NMS is characterized by the acute or subacute onset of muscle rigidity, elevated temperature, altered mental status, hyperthermia, tachycardia, labile blood pressure, and renal failure with markedly elevated creatine kinase levels. Symptoms typically evolve within days or weeks after initiating the drug. NMS can also be precipitated by the abrupt withdrawal of dopaminergic medications in PD patients. Treatment involves immediate cessation of the offending antipsychotic drug and the introduction of a dopaminergic agent (e.g., a dopamine agonist or levodopa), dantrolene, or a benzodiazepine. In very severe cases, when oral intake is not possible, a patch (delivering rotigotine subcutaneously) or an infusion pump (delivering apomorphine or levodopa subcutaneously) may be required. Treatment may need to be undertaken in an intensive care setting and include supportive measures such as control of body temperature (antipyretics and cooling blankets), hydration, electrolyte replacement, and control of renal function and blood pressure.

Drugs that have serotonin-like activity (tryptophan, MDMA or "ecstasy," meperidine) or that block serotonin reuptake can induce a rare, but potentially fatal, serotonin syndrome that is characterized by confusion, hyperthermia, tachycardia, and coma, as well as rigidity, ataxia, and tremor. Myoclonus is often a prominent feature, in contrast to NMS, which it resembles in other respects. Patients can be managed with propranolol, diazepam, diphenhydramine, chlorpromazine, or cyproheptadine, as well as supportive measures.

A variety of other drugs can be associated with hyperkinetic movement disorders. Some examples include phenytoin (chorea, dystonia, tremor, myoclonus), carbamazepine (tics and dystonia), tricyclic antidepressants (dyskinesias, tremor, myoclonus), fluoxetine (myoclonus, chorea, dystonia), oral contraceptives (dyskinesia), β-adrenergics (tremor), buspirone (akathisia, dyskinesias, myoclonus), digoxin, cimetidine, diazoxide, lithium, methadone, and fentanyl (dyskinesias). And as described in an earlier chapter **(Chap. 446)**, treatment of PD with levodopa can be associated with dyskinetic movements; these are typically choreiform, but dystonia and myoclonus may also occur.

PAROXYSMAL DYSKINESIAS

Paroxysmal dyskinesias are a group of rare disorders characterized by episodic, brief involuntary movements that can manifest as various types of hyperkinetic movements, including chorea, dystonia, tremor, myoclonus, and ballism. There are three main types: (1) *paroxysmal kinesigenic dyskinesia (PKD)*, where the involuntary movements are triggered by sudden movement; (2) *paroxysmal nonkinesigenic dyskinesias (PNKD)*, where the attacks are not induced by movement; and (3) rare cases of *paroxysmal exertion-induced dyskinesia (PED)*, where attacks are induced by prolonged exercise.

PKDs are characterized by brief, self-limited attacks induced by the onset of movement such as running but also occasionally by unexpected sound or photic stimulation. Attacks may affect one side of the body, last seconds to minutes at a time, and recur several times a day. They usually manifest as a mixed hyperkinetic movement disorder with dystonic posturing of a limb, ballismus, and chorea, which may also become generalized. PKD is most commonly familial with an autosomal dominant pattern of inheritance and mutations in the *proline-rich transmembrane protein 2* (*PRRT2*) gene but may also occur secondary to various brain disorders such as multiple sclerosis or hyperglycemia. PKD is more frequent in males (4:1), and the onset is typically in the first or second decade of life. About 70% report sensory symptoms such as tingling or numbness of the affected limb preceding the attack by a few milliseconds. The evolution is relatively benign, and there is a trend toward resolution of the attacks over time. Treatment with low-dose anticonvulsant therapy such as carbamazepine or phenytoin is advised when the attacks are frequent and interfere with daily life activities and is effective in ~80% of patients. Some clinical features of PKD (abrupt and short-lasting attacks preceded by an "aura"), the association with true seizure episodes, and its favorable response to anticonvulsant drugs have led to speculation that it is epileptic in origin, but this has not been established.

PNKD involves attacks of generalized dyskinesias precipitated by alcohol, caffeine, stress, or fatigue. In comparison to PKD, the episodes have a relatively longer duration (minutes to hours) and are less frequent (one to three per day). PNKD is inherited as an autosomal dominant condition with high (~80%) but incomplete penetrance. A missense mutation in the *myofibrillogenesis regulator* (*PNKD*) gene has been identified in several families. Recognition of the condition and elimination of the underlying precipitating factors, where possible, are the first priorities. Tetrabenazine, neuroleptics, dopamine-blocking agents, propranolol, clonazepam, and baclofen may be helpful. Treatment may not be required if the condition is mild and self-limited. Most patients with PNKD do not benefit from anticonvulsant drugs, but these should be tried, and some may respond to clonazepam or other benzodiazepines.

PED is characterized by a combination of chorea, athetosis, and dystonia in excessively exercised body regions, with the legs being most frequently affected. They are frequently familial. A single attack lasts from a few minutes to an hour and occurs after prolonged physical exercise. In addition to the movement disorder, several patients have other disease manifestations between episodes such as epilepsy, hemolytic anemia, and migraine. The *SLC2A1* (solute carrier family 2 member 1) gene, previously linked to GLUT1 (glucose transporter of the blood-brain barrier) deficiency syndrome, can also cause paroxysmal PED. Treatment includes avoiding prolonged physical exercise. Whereas anticonvulsants and most medications are typically not effective, a ketogenic diet may be an effective therapeutic option.

Other (rare) forms of paroxysmal dyskinesia are caused by pathogenic variants in the *ECHS1*, *GLDC*, *KCNMA1, SCN8A*, and *TMEM151A* genes.

RESTLESS LEGS SYNDROME

Restless legs syndrome (RLS) is a neurologic disorder that affects ~10% of the adult population (it is rare in Asians). It was first described in the seventeenth century by the English physician Thomas Willis but has only recently been appreciated to be a bona fide movement disorder. The four core symptoms required for diagnosis are an urge to move the legs usually caused or accompanied by an unpleasant sensation in the legs; symptoms that begin or worsen with rest; partial or complete relief by movement; and worsening during the evening or night. Symptoms are often mild but can cause significant morbidity in some individuals.

Symptoms most commonly begin in the legs but can spread to, or even begin in, the upper limbs. The unpleasant sensation is often described as a creepy-crawly feeling, paresthesia, or burning. In ~80% of patients, RLS is associated with periodic leg movements (PLMs) during sleep and occasionally while awake. These involuntary movements are usually brief, lasting no more than a few seconds, and recur every 5–90 s. The restlessness and PLMs are a major cause of sleep disturbance, leading to poor-quality sleep and daytime sleepiness. RLS is also commonly associated with depression, anxiety, and hypertension.

The mean age of onset in familial forms is in the third decade, although pediatric cases are recognized. The severity of symptoms is variable. Secondary RLS may be associated with pregnancy or a range of underlying disorders, including anemia, ferritin deficiency, renal failure, and peripheral neuropathy. There is an association with abnormalities of iron metabolism, possibly because low iron can result in reduced dopamine levels. Diagnosis is made on clinical grounds but can be supported by polysomnography and the demonstration of PLMs. Recent studies suggest that age, sex, and genetic markers can be used to accurately predict who is likely to develop RLS in 90% of cases. The neurologic examination is normal. Secondary causes of RLS should be excluded, and ferritin levels, glucose, and renal function should be measured. The pathogenesis is thought to be associated with an alteration in dopamine function, which may be peripheral or central, but this has not been specifically defined. Primary RLS is often familial and has a strong genetic component; however, no causative gene has yet been identified. Genome association studies have identified >150 variants associated with RLS risk, with the strongest candidates in the *PTPRD*, *BTBD9*, and *MEIS1* genes. Interestingly no genetic linkage to iron has been identified.

Most RLS sufferers have mild symptoms that do not require specific treatment. General measures to improve sleep hygiene and quality should be attempted first. If symptoms remain intrusive, low doses of dopamine agonists, e.g., pramipexole (0.25–0.5 mg), ropinirole (1–2 mg), or patch rotigotine (2–3 mg), taken 1–2 h before bedtime are generally effective. Levodopa may also be effective but is more likely to be associated with augmentation (spread and worsening of symptoms and emergence during the day) or rebound (reappearance sometimes with worsening of symptoms at a time related to the drug's short half-life). Augmentation can also be seen with chronic use of drugs such as dopamine agonists, particularly if higher doses are employed. Other drugs that have been reported to be effective in individual cases include anticonvulsants, analgesics, and opiates, but these are not commonly employed. Tonic motor activation (TOMAC) is a nonpharmacologic approach to RLS that has not responded to drug therapy and has recently been approved in the United States. The treatment involves electrical stimulation of the peroneal nerves during the night and is reported to be effective and to improve sleep quality. Management of secondary RLS should be directed to correcting the underlying disorder (e.g., iron replacement for anemia).

OTHER DISORDERS THAT MAY PRESENT WITH A COMBINATION OF PARKINSONISM AND HYPERKINETIC MOVEMENTS

■ WILSON'S DISEASE (SEE ALSO CHAP. 427)

Wilson's disease (WD) is an inherited autosomal recessive disorder of copper metabolism that produces neurologic, psychiatric, and liver manifestations, alone or in combination. It is caused by mutations in the *ATP7B* gene encoding a P-type ATPase. The disease was first described by the English neurologist Kinnier Wilson at the beginning of the twentieth century, although at around the same time, the German physicians Kayser and Fleischer separately noted the characteristic association of corneal pigmentation (Kayser-Fleischer rings) along with hepatic and neurologic features. WD has a worldwide prevalence of ~1 in 30,000, with a mutation carrier frequency of 1 in 90. About half of WD patients (especially younger patients) present with liver abnormalities. The remainder present with neurologic disease (with or without underlying liver abnormalities), and a small proportion have hematologic or psychiatric problems at disease onset.

Neurologic onset usually manifests in the second decade with tremor, rigidity, and dystonia. The tremor is usually in the upper limbs, bilateral, and asymmetric. Tremor can be on intention or occasionally at rest, and in advanced disease can take on a wing-beating characteristic (a flapping movement when the arms are held outstretched with the fingers opposed). Other features can include parkinsonism with bradykinesia, dystonia (particularly facial grimacing), dysarthria, and dysphagia. More than half of those with neurologic features have a history of psychiatric disturbances, including depression, mood swings, and overt psychosis. Kayser-Fleischer (KF) rings are seen in virtually all patients with neurologic features and 80% of those with hepatic presentations. KF rings represent the deposition of copper in Descemet's membrane around the cornea. They consist of a characteristic grayish rim or circle at the limbus of the cornea and are best detected by slit-lamp examination. Neuropathologic examination is characterized by neurodegeneration and astrogliosis in the basal ganglia, particularly in the striatum.

WD should always be considered in the differential diagnosis of a movement disorder, particularly when arising in the first decades of life. Low levels of blood copper and ceruloplasmin and high levels of urinary copper may be present, but normal levels do not exclude the diagnosis. Brain imaging usually reveals generalized brain atrophy in established cases, and ~50% have signal hypointensity in the caudate head, putamen, globus pallidus, substantia nigra, and red nucleus on T2-weighted MRI scans. However, the correlation of imaging changes with clinical features is not good. Liver biopsy with demonstration of high copper levels and genetic testing remain the gold standard for diagnosis.

In the absence of treatment, the course is progressive and leads to severe neurologic dysfunction and early death in most patients, although a small proportion experience a relatively benign course. Treatment is directed at reducing tissue copper levels and maintenance therapy to prevent reaccumulation. There is no clear consensus on optimal treatment, and patients should be managed in a unit with expertise in treating this disease. Penicillamine is frequently used to increase copper excretion but may lead to a worsening of symptoms in the initial stages of therapy. Side effects are common and can to some degree be attenuated by co-administration of pyridoxine. Tetrathiomolybdate blocks the absorption of copper and can be used instead of penicillamine. Trientine tetrahydrochloride and zinc are useful drugs for maintenance therapy. Effective treatment can reverse the neurologic features in most patients, particularly when started early. However, some patients may still progress, especially those with hepatocerebral disease. KF rings tend to decrease after 3–6 months and disappear by 2 years. Adherence to maintenance therapy is a major challenge in long-term care. Patients with advanced hepatic disease may require a liver transplant, and the potential role of organ-specific chelation therapy is under investigation. Gene therapy studies that involve an infusion of a working copy of the *ATP7B* gene into the liver are being investigated clinically, while preclinical studies are testing the novel chelator methanobactin.

■ NEURODEGENERATION WITH BRAIN IRON ACCUMULATION

Neurodegeneration with brain iron accumulation (NBIA) represents a group of inherited disorders characterized by iron accumulation in the basal ganglia. Clinically, they can manifest as a progressive neurologic disorder with a variety of clinical features including parkinsonism, dystonia, neuropsychiatric abnormalities, and retinal degeneration. Cognitive disorders and cerebellar dysfunction may also be seen. Presentation is usually in childhood, but adult cases have been described. Multiple genes have been identified. Pantothenate kinase-associated neurodegeneration (PKAN), formerly known as Hallervordan-Spatz disease, is caused by a mutation in the *PANK2* gene and is the most common form of NBIA, accounting for ~50% of cases. Onset is usually in early childhood and is manifest as a combination of dystonia, parkinsonism, and spasticity. MRI shows a characteristic low signal abnormality in the center of the globus pallidus on T2-weighted scans caused by iron accumulation and known as the "eye of the tiger" sign. Numerous other gene mutations have been described associated with iron accumulation, including *PLA2G6, C19orf12, FA2H, ATP13A2, WDR45, FTL, CP, COASY*, and *DCAF17*. One must be cautious, however, not to assume that all cases with iron accumulation in the basal ganglia represent an NBIA because iron accumulation in some basal ganglia regions is normal, and excess iron accumulation may occur in the basal ganglia as a nonspecific secondary consequence of neurodegeneration unrelated to a defect in iron metabolism. There are no specific treatments; iron binding may help slow progression, but this has not been established.

FUNCTIONAL (PSYCHOGENIC) DISORDERS

Virtually all movement disorders including tremor, tics, dystonia, myoclonus, chorea, ballism, and parkinsonism can be psychogenic in origin. The term *functional neurologic symptom disorder (FND)/conversion disorder* has been suggested to replace the term *psychogenic disorder* in order to remove the criterion of psychological stress as a prerequisite for diagnosis; however, the terminology remains controversial and both terms are used. A diagnosis can be made by identifying neurologic signs that are specific to FNDs without reliance on psychological stressors or suggestive historical clues. Tremor affecting the upper limbs is the most common psychogenic movement disorder. Psychogenic movements can result from a somatoform or conversion disorder, malingering (e.g., seeking financial gain), or a factitious disorder (e.g., seeking psychological gain) **(Chap. 463)**. Functional movement disorders are relatively common (estimated at 2–3% of patients seen in a movement disorder clinic), more frequent in women, disabling for the patient and family, and expensive for society. Clinical features suggesting a functional or psychogenic movement disorder include an acute onset with a pattern of abnormal movement that is inconsistent with the phenotype of a known movement disorder. Diagnosis is based on the nonorganic quality of the movement, the absence of findings of an organic disease process, and positive features that specifically point to a functional illness such as variability and distractibility. For example, in a functional tremor disorder, the magnitude of the tremor is increased with attention and diminishes or even disappears when the patient is distracted by being asked to perform a different task or is unaware that he or she is being observed. This is the opposite of what is seen with an organic tremor where the magnitude of tremor is increased with distraction and tends to be reduced when observed. Other positive features suggesting a psychogenic problem include variable tremor frequency, entrainment of the tremor frequency with the frequency of a designated movement in the contralateral limb such as tapping, and a response to placebo interventions. Associated features can include nonanatomic sensory findings, give-way weakness, astasia-abasia (an odd, gyrating gait or posture) **(Chap. 28)**, and multiple somatic complaints with no underlying pathology (somatoform disorder). Comorbid psychiatric problems such as anxiety, depression, and emotional trauma may be present but are not necessary for the diagnosis, which is why some prefer to call the movement disorder functional rather than psychogenic. Functional movement disorders typically occur as an isolated entity but may be seen in association with an underlying organic problem. The diagnosis can usually be made based on history and clinical features alone, and unnecessary tests or medications can be avoided. If there are underlying psychiatric problems, they should be identified and treated, but as noted, many patients with functional movement disorders have no obvious psychiatric pathology. Treatment of FND starts with explaining the diagnosis to the patient in a nonthreatening manner, but many are resistant to accepting this diagnosis. Psychological therapies (especially cognitive-behavioral) are the method of choice. An increasing role of physiotherapy has also recently been recognized, and a recent trial of physiotherapy and cognitive-behavioral therapy in combination was found to effectively improve symptoms in nearly half of patients. Comorbid depression, anxiety, and pain may be treated pharmacologically. Patients with hypochondriasis, factitious disorders, and malingering have a poor prognosis.

■ FURTHER READING

Baumgartner AJ et al: Novel targets in deep brain stimulation for movement disorders. Neurosurg Rev 45:2593, 2022.

Bhatia KP et al: Tremor Task Force of the International Parkinson and Movement Disorder Society. Mov Disord 33:75, 2018.

Billnitzer A, Jankovic J: Current management of tics and Tourette syndrome: Behavioral, pharmacologic, and surgical treatments. Neurotherapeutics 17:1681, 2020.

Elias WJ, Shah BB: Essential tremor. JAMA 332:418, 2024.

Espay AJ et al: Current concepts in diagnosis and treatment of functional neurological disorders. JAMA Neurol 75:1132, 2018.

Lange LM et al: Nomenclature of Genetic Movement Disorders: Recommendations of the International Parkinson and Movement Disorder Society Task Force—An update. Mov Disord 37:905, 2022.

Macias-Garcia D et al: Combined physiotherapy and cognitive behavioral therapy for functional movement disorders: A randomized clinical trial. JAMA Neurol 81:966, 2024.

Mestre TA: Recent advances in the therapeutic development for Huntington disease. Parkinsonism Relat Disord 59:125, 2019.

Tabrizi SJ et al: Potential disease-modifying therapies for Huntington's disease: Lessons learned and future opportunities. Lancet Neurol 21:645, 2022.

Thomsen M et al: Genetics and pathogenesis of dystonia. Annu Rev Pathol 19:99, 2024.

448

Amyotrophic Lateral Sclerosis and Other Motor Neuron Diseases

Robert H. Brown, Jr.

AMYOTROPHIC LATERAL SCLEROSIS

Amyotrophic later sclerosis (ALS) is the most common progressive motor neuron disease. It is a prime example of a neurodegenerative disease and is arguably the most devastating of the neurodegenerative disorders.

■ PATHOLOGY

The pathologic hallmark of motor neuron degenerative disorders is death of lower motor neurons (consisting of anterior horn cells in the spinal cord and their brainstem homologues innervating bulbar muscles) and upper, or corticospinal, motor neurons (originating in layer five of the motor cortex and descending via the pyramidal tract to synapse with lower motor neurons, either directly or indirectly via interneurons) **(Chap. 26)**. Although at its onset ALS may involve selective loss of function of only upper or lower motor neurons, it ultimately causes progressive loss of both categories of motor neurons. Indeed, in the absence of clear involvement of both motor neuron types, the diagnosis of ALS is questionable. In a subset of cases, ALS arises concurrently with frontotemporal dementia **(Chap. 443)**; in these instances, there is degeneration of frontotemporal cortical neurons and corresponding cortical atrophy.

Other motor neuron diseases involve only particular subsets of motor neurons **(Tables 448-1 and 448-2)**. Thus, in bulbar palsy and spinal muscular atrophy (SMA, predominantly in children) and progressive muscular atrophy (PMA, in adults), the lower motor neurons of brainstem and spinal cord, respectively, are most severely involved. By contrast, pseudobulbar palsy, primary lateral sclerosis (PLS), and hereditary spastic paraplegia (HSP) affect only upper motor neurons innervating the brainstem and spinal cord.

In each of these diseases, the affected motor neurons undergo shrinkage, often with accumulation of the pigmented lipid (lipofuscin) that normally develops in these cells with advancing age. In ALS, the motor neuron cytoskeleton is typically affected early in the illness. Focal enlargements are frequent in proximal motor axons; ultrastructurally, these "spheroids" are composed of accumulations of neurofilaments and other proteins. Commonly in both sporadic and familial ALS, the affected neurons demonstrate ubiquitin-positive aggregates, often associated with the protein TDP43 (see below). Also seen is

TABLE 448-1 Etiology of Motor Neuron Disorders

DIAGNOSTIC CATEGORY	INVESTIGATION
Structural lesions	MRI scan of head (including foramen magnum and cervical spine)
Parasagittal or foramen magnum tumors	
Cervical spondylosis	
Chiari malformation of syrinx	
Spinal cord arteriovenous malformation	
Infections	CSF exam, culture
Bacterial—tetanus, Lyme	Lyme titer
Viral—poliomyelitis, herpes zoster	Antiviral antibody
Retroviral—myelopathy	HTLV-1 titers
Intoxications, physical agents	24-h urine for heavy metals
Toxins—lead, aluminum, others	Serum lead level
Drugs—strychnine, phenytoin	
Electric short, x-irradiation	
Immunologic mechanisms	Complete blood count[a]
Plasma cell dyscrasias	Sedimentation rate[a]
Autoimmune polyradiculopathy	Total protein[a]
Motor neuropathy with conduction block	Anti-GM1 antibodies[a]
	Anti-Hu antibody
Paraneoplastic	MRI scan, bone marrow biopsy
Paracarcinomatous	
Metabolic	Fasting blood sugar[a]
Hypoglycemia	Routine chemistries including calcium[a]
Hyperparathyroidism	PTH
Hyperthyroidism	Thyroid function[a]
Deficiency of folate, vitamin B_{12}, vitamin E	Vitamin B_{12}, vitamin E, folate[a]
	Serum zinc, copper[a]
Malabsorption	24-h stool fat, carotene, prothrombin time
Deficiency of copper, zinc	Fasting lactate, pyruvate, ammonia
Mitochondrial dysfunction	Consider mtDNA
Hyperlipidemia	Lipid electrophoresis
Hyperglycinuria	Urine and serum amino acids
	CSF amino acids
Hereditary disorders	WBC DNA for mutational analysis
C9orf72	
Superoxide dismutase	
TDP43	
FUS/TLS	
Androgen receptor defect (Kennedy's disease)	

[a] Should be obtained in all cases.

Abbreviations: CSF, cerebrospinal fluid; FUS/TLS, fused in sarcoma/translocated in liposarcoma; HTLV-1, human T-cell lymphotropic virus; MRI, magnetic resonance imaging; PTH, parathyroid; WBC, white blood cell.

TABLE 448-2 Sporadic Motor Neuron Diseases

CHRONIC	ENTITY
Upper and lower motor neuron	Amyotrophic lateral sclerosis
Predominantly upper motor neuron	Primary lateral sclerosis
Predominantly lower motor neuron	Multifocal motor neuropathy with conduction block Motor neuropathy with paraproteinemia or cancer Motor predominant peripheral neuropathies
Other	
Associated with other neurodegenerative disorders	
Secondary motor neuron disorders (see Table 448-1)	
Acute	
Poliomyelitis	
Herpes zoster	
Coxsackie virus	
West Nile virus	

proliferation of astroglia and microglia, the inevitable accompaniment of all degenerative processes in the central nervous system (CNS).

The death of the peripheral motor neurons in the brainstem and spinal cord leads to denervation and atrophy of the corresponding muscle fibers. Histochemical and electrophysiologic evidence indicates that in the early phases of the illness denervated muscle can be reinnervated by sprouting of nearby distal motor nerve terminals, although reinnervation in this disease is considerably less extensive than in most other disorders affecting motor neurons (e.g., poliomyelitis, peripheral neuropathy). As denervation progresses, muscle atrophy is readily recognized in muscle biopsies and on clinical examination. This is the basis for the term *amyotrophy*. The loss of cortical motor neurons results in thinning of the corticospinal tracts that travel via the internal capsule **(Fig. 448-1)** and pyramidal tracts in the brainstem to the lateral and anterior white matter columns of the spinal cord. The loss of fibers in the lateral columns and resulting fibrillary gliosis impart a particular firmness (*lateral sclerosis*). A remarkable feature of the disease is the selectivity of neuronal cell death. By light microscopy, the entire sensory apparatus and cerebellar structures that control the coordination of movement remain intact. Except in cases of frontotemporal dementia, the components of the brain required for cognitive processing are also preserved. However, immunostaining indicates that neurons bearing ubiquitin, a marker for degeneration, are also detected in nonmotor systems. Moreover, studies of glucose metabolism in the illness also indicate that there is neuronal dysfunction outside of the motor system. Pathologic studies reveal proliferation of microglial cells and astrocytes in affected regions; in some cases, this phenomenon, designated neuroinflammation, can be visualized using positron emission tomography (PET) scanning for ligands that are recognized by activated microglia. Within the motor system, there is some selectivity

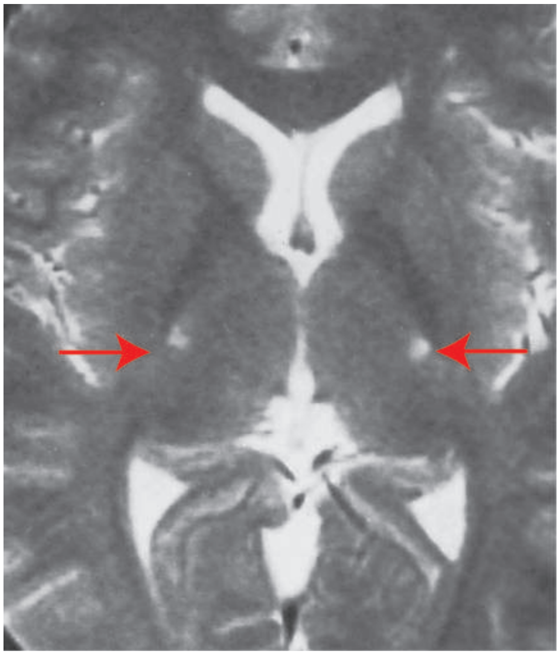

FIGURE 448-1 Amyotrophic lateral sclerosis. Axial T2-weighted magnetic resonance imaging (MRI) scan through the lateral ventricles of the brain reveals abnormal high signal intensity within the corticospinal tracts (*arrows*). This MRI feature represents an increase in water content in myelin tracts undergoing Wallerian degeneration secondary to cortical motor neuronal loss. This finding is commonly present in ALS but can also be seen in AIDS-related encephalopathy, infarction, or other disease processes that produce corticospinal neuronal loss in a symmetric fashion.

of involvement. Thus, motor neurons required for ocular motility remain unaffected, as do the parasympathetic neurons in the sacral spinal cord (the nucleus of Onufrowicz, or Onuf) that innervate the sphincters of the bowel and bladder.

CLINICAL MANIFESTATIONS

The manifestations of ALS are somewhat variable depending on whether corticospinal neurons or lower motor neurons in the brainstem and spinal cord are more prominently involved. With lower motor neuron dysfunction and early denervation, typically the first evidence of the disease is insidiously developing asymmetric weakness, usually first evident distally in one of the limbs. A detailed history often discloses recent development of cramping with volitional movements, typically in the early hours of the morning (e.g., while stretching in bed). Weakness caused by denervation is associated with progressive wasting and atrophy of muscles and, particularly early in the illness, spontaneous twitching of motor units, or fasciculations. In the hands, a preponderance of extensor over flexor weakness is common. When the initial denervation involves bulbar rather than limb muscles, the problem at onset is difficulty with chewing, swallowing, and movements of the face and tongue. Rarely, early involvement of the muscles of respiration may lead to death before the disease is far advanced elsewhere. With prominent corticospinal involvement, there is hyperactivity of the muscle-stretch reflexes (tendon jerks) and, often, spastic resistance to passive movements of the affected limbs. Patients with significant reflex hyperactivity complain of muscle stiffness often out of proportion to weakness. Degeneration of the corticobulbar projections innervating the brainstem results in dysarthria and exaggeration of the motor expressions of emotion. The latter leads to involuntary excess in weeping or laughing (pseudobulbar affect).

Virtually any muscle group may be the first to show signs of disease, but, as time passes, more and more muscles become involved until ultimately the disorder takes on a symmetric distribution in all regions. It is characteristic of ALS that, regardless of whether the initial disease involves upper or lower motor neurons, both will eventually be implicated. Even in the late stages of the illness, sensory, bowel and bladder, and cognitive functions are preserved. Even when there is severe brainstem disease, ocular motility is spared until the very late stages of the illness. As noted, in some cases (particularly those that are familial), ALS develops concurrently with frontotemporal dementia, characterized by early behavioral abnormalities with prominent behavioral features indicative of frontal lobe dysfunction.

A committee of the World Federation of Neurology has established diagnostic guidelines for ALS. Essential for the diagnosis is simultaneous upper and lower motor neuron involvement with progressive weakness and the exclusion of all alternative diagnoses. The disorder is ranked as “definite” ALS when three or four of the following are involved: bulbar, cervical, thoracic, and lumbosacral motor neurons. When two sites are involved, the diagnosis is “probable,” and when only one site is implicated, the diagnosis is “possible.” An exception is made for those who have progressive upper and lower motor neuron signs at only one site and a mutation in the gene encoding superoxide dismutase (SOD1; see below).

It is now recognized that another clinical manifestation in most cases of ALS is the presence in cerebrospinal fluid (CSF) and serum of markers of neurodegeneration, such as elevated levels of neurofilament light chains (Nfl) or phosphorylated neurofilament heavy chains; some markers of inflammation (e.g., monocyte chemoattractant protein 1) are also elevated. Higher levels of serum or CSF Nfl are correlated with more aggressive disease and more rapid disease progression. Accordingly, these CSF biomarkers are increasingly used as endpoints in clinical trials.

EPIDEMIOLOGY

The illness is relentlessly progressive, leading to death from respiratory paralysis; the median survival is from 3 to 5 years. There are very rare reports of stabilization or even regression of ALS. In most societies, there is an incidence of 1–3 per 100,000 and a prevalence of 3–5 per 100,000. It is striking that about 1 in 500 deaths in North America and Western Europe (and probably elsewhere) are due to ALS; this finding predicts that >500,000 individuals now alive in the United States will die of ALS. Several endemic foci of higher prevalence exist in the western Pacific (e.g., in specific regions of Guam or Papua New Guinea). In the United States and Europe, men are somewhat more frequently affected than women. Epidemiologic studies have incriminated risk factors for this disease including exposure to pesticides and insecticides, silica, smoking, and possibly service in the military. Although ALS is overwhelmingly a sporadic disorder, some 10% of cases are inherited as an autosomal dominant trait.

FAMILIAL ALS

Several forms of selective motor neuron disease are inheritable **(Table 448-3)**. Familial ALS (FALS) involves both corticospinal and lower motor neurons. Apart from its inheritance as an autosomal dominant trait, it is clinically indistinguishable from sporadic ALS. Genetic studies have identified mutations in multiple genes, including those encoding the protein C9orf72 (open reading frame 72 on chromosome 9), cytosolic enzyme SOD1 (superoxide dismutase), the RNA binding proteins TDP43 (encoded by the TAR DNA binding protein gene), and fused in sarcoma/translocated in liposarcoma (FUS/TLS), as the most common causes of FALS. Mutations in *C9orf72* account for ~45–50% of FALS and perhaps 5–10% of sporadic ALS cases. Mutations in SOD1 explain another 20% of cases of FALS, whereas TDP43 and FUS/TLS each represent about 5% of familial cases. Mutations in several other genes (e.g., *NEK1*, *optineurin*, *TBK1*, *KIF5A*, *TUBA4*, and *PFN11*) each cause ~1% of cases.

Rare mutations in other genes are also clearly implicated in ALS-like diseases. Thus, a familial, dominantly inherited motor disorder that in some individuals closely mimics the ALS phenotype arises from mutations in a gene that encodes a vesicle-binding protein. Mutations in *senataxin*, a helicase, cause an early-adult-onset, slowly evolving ALS variant. Kennedy’s syndrome is an X-linked, adult-onset disorder that may mimic ALS, as described below. *Tau* gene mutations usually underlie frontotemporal dementia but in some instances may be associated with prominent motor neuron findings.

Genetic analyses are also beginning to illuminate the pathogenesis of some childhood-onset motor neuron diseases. For example, a slowly disabling degenerative, predominantly upper motor neuron disease that starts in the first decade is caused by mutations in a gene that expresses a novel signaling molecule with properties of a guanine-exchange factor, termed *alsin*.

DIFFERENTIAL DIAGNOSIS

Because ALS is currently untreatable, it is imperative that potentially remediable causes of motor neuron dysfunction be excluded (Table 448-1). This is particularly true in cases that are atypical by virtue of (1) restriction to either upper or lower motor neurons, (2) involvement of neurons other than motor neurons, and (3) evidence of motor neuronal conduction block on electrophysiologic testing. Compression of the cervical spinal cord or cervicomedullary junction from tumors in the cervical regions or at the foramen magnum or from cervical spondylosis with osteophytes projecting into the vertebral canal can produce weakness, wasting, and fasciculations in the upper limbs and spasticity in the legs, closely resembling ALS. The absence of cranial nerve involvement may be helpful in differentiation, although some foramen magnum lesions may compress the twelfth cranial (hypoglossal) nerve, with resulting paralysis of the tongue. Absence of pain or of sensory changes, normal bowel and bladder function, normal radiologic studies of the spine, and normal CSF all favor ALS. Where doubt exists, magnetic resonance imaging (MRI) scans and possibly contrast myelography should be performed to visualize the cervical spinal cord.

Another important entity in the differential diagnosis of ALS is *multifocal motor neuropathy with conduction block* (MMCB), discussed below. A diffuse, lower motor axonal neuropathy mimicking ALS sometimes evolves in association with hematopoietic disorders such as lymphoma or multiple myeloma. In this clinical setting, the presence of an M-component in serum should prompt consideration of a bone

TABLE 448-3 Genetic Motor Neuron Diseases

DISEASE	GENE SYMBOL	GENE NAME	INHERITANCE	U.S. FREQUENCY % FALS	USUAL ONSET	PROTEIN FUNCTION	UNUSUAL FEATURES
I. Selected Upper and Lower Motor Neurons (Familial ALS) + Frontotemporal Dementia (FTD)							
ALS/ALS-FTD	*C9ORF72*	Chromosome 9 open reading frame 72	AD	45% (6–10% SALS)	Adult	Regulates vesicle trafficking	May also be associated with parkinsonism, PLS
ALS	*SOD1*	Cu/Zn superoxide dismutase 1	AD	20% (2% SALS)	Adult	Protein antioxidant	
ALS/ALS-FTD	*TARDBP*	TAR DNA binding protein	AD	5%	Adult	DNA, RNA binding	
ALS/ALS-FTD	*FUS/TLS*	Fused in sarcoma/ translocated in liposarcoma	AD	5%	Adult	DNA, RNA binding	
ALS/ALS-FTD	*CCNF*	E3 ubiquitin ligase cyclin F	AD	2%	Adult	Mediates ubiquitination	
ALS	*NEK1*	NMA-related kinase	AR	2%	Adult	Microtubules, nuclear transport	
ALS/ALS-FTD	*TBK1*	Tank binding kinase 1	AD	2%	Adult	Regulates autophagy, inflammation	
ALS	*KIF5A*	Kinesin family member 5A	AD	1–2%	Early adult	Microtubule motor	CMT
ALS	*PFN1*	Profilin 1	AD	~1%	Adult	Involved in actin polymerization	
ALS/ALS-FTD	*OPTN*	Optineurin	AD/AR	~1%	Adult	Attenuates NF-κB	
ALS	*SPG11*	Spastic paraplegia 11	AR	~1%	Adult	Vesicle trafficking	Spastic paraplegia
ALS	*SETX*	Senataxin	AD	~1%	Late juvenile	DNA helicase	Late childhood onset
ALS/ALS-FTD	*VCP*	Valosin-containing protein	AD	~ 1%	Adult	ATPase	Paget's, myopathy
ALS-FTD	*UBQLN2*	Ubiquilin 2	XR	<1%	Adult or juvenile	Protein degradation	
ALS-FTD	*CHMP2B*	Chromatin modifying protein 2B	AD	<1%	Adult	Chromatin binding protein	
ALS-FTD	*MAPT*	Microtubule Associated Protein Tau	AD	<1%	Adult	Cytoskeletal protein	Usually causes only FTD
ALS2	*ALS2*	Alsin	AR	<1%	Juvenile	GEF signaling	Corticobulbar/ corticospinal may mimic PLS
ALS-FTD	*CHMP2B*	Chromatin modifying protein 2B	AD	<1%	Adult	Chromatin binding protein	
II. Lower Motor Neurons							
Spinal muscular atrophies	*SMN*	Survival motor neuron	AR	1/10,000 live births	Infancy	RNA metabolism	
GM2-gangliosidosis							
1. Sandhoff's disease	*HEXB*	Hexosaminidase B	AR		Childhood	Ganglioside recycling	
2. AB variant	*GM2A*	GM2-activator protein	AR		Childhood	Ganglioside recycling	
3. Adult Tay-Sachs disease	*HEXA*	Hexosaminidase A	AR		Childhood	Ganglioside recycling	
X-linked spinobulbar muscular atrophy	*AR*	Androgen receptor	XR		Adult	Nuclear signaling	
III. Upper Motor Neuron (Selected HSPs)							
SPG3A	*ATL1*	Atlastin	AD	10% AD FSP	Childhood	GTPase—vesicle recycling	
SPG4	*SPAST*	Spastin	AD	50–60% AD FSP	Early adulthood	ATPase family—microtubule associate	Some sensory loss
SPG10	*KIF5A*	Kinesin heavy chain isoform 5A	AD	10% AD FSP	Second–third decade	Motor-associated protein	± Peripheral neuropathy, retardation
SPG31	*REEP1*	Receptor Expression Enhancing Protein 1	AD	10% AD FSP	Early	Mitochondrial protein	Rarely, amyotrophy
SPG5	*CYP7B1*	Cytochrome P450	AR	5–10% AR FSP	Variable	Degrades endogenous substances	Sensory loss
SPG7	*SPG7*	Paraplegin	AR	5–10% AR FSP	Variable	Mitochondrial protein	Rarely, optic atrophy, ataxia, rarely PLS

(*Continued*)

TABLE 448-3 Genetic Motor Neuron Diseases (Continued)

DISEASE	GENE SYMBOL	GENE NAME	INHERITANCE	U.S. FREQUENCY % FALS	USUAL ONSET	PROTEIN FUNCTION	UNUSUAL FEATURES
SPG11	*SPG11*	Spatacsin	AR	20–70% AR FSP depends on ethnicity	Predominantly childhood	Cytosolic,? membrane-associated	Some sensory loss, thin corpus callosum; may mimic ALS (ALS5)
SPG2	*PLP*	Proteolipid protein	XR	<1%	Early childhood	Myelin protein	Sometimes multiple CNS features
Adrenoleukodystrophy	*ALDP*	Adrenoleukodystrophy protein	XR	<1%	Early adulthood	ATP binding transporter protein	Possible adrenal insufficiency, CNS inflammation

Abbreviations: AD, autosomal dominant; ALS, amyotrophic lateral sclerosis; AR, autosomal recessive; CNS, central nervous system; CMT, Charcot-Marie-Tooth; BSCL2, Bernadelli-Seip congenital lipodystrophy 2B; FALS, familial amyotrophic lateral sclerosis; FSP, familial spastic paraplegia; FUS/TLS, fused in sarcoma/translocated in liposarcoma; GEF, guanidine nucleotide exchange factor; HSP, hereditary spastic paraplegia; NF-κB, nuclear factor-κB; PLS, primary lateral sclerosis; SALS, sporadic amyotrophic lateral sclerosis; TDP43, Tar DNA binding protein 43 kd; XR, X-linked recessive.

marrow biopsy. Lyme disease **(Chap. 191)** may also cause an axonal, lower motor neuropathy, although typically with intense proximal limb pain and a CSF pleocytosis.

Other treatable disorders that occasionally mimic ALS are chronic lead poisoning and thyrotoxicosis. These disorders may be suggested by the patient's social or occupational history or by unusual clinical features. When the family history is positive, disorders involving the genes encoding C9orf72, cytosolic SOD1, TDP43, FUS/TLS, and adult hexosaminidase A or α-glucosidase deficiency **(Chap. 429)** must be excluded. These are readily identified by appropriate laboratory tests; importantly, panels for simultaneous analysis of multiple ALS and frontotemporal dementia (FTD) genes are now commercially available. Benign fasciculations are occasionally a source of concern because on inspection they resemble the fascicular twitchings that accompany motor neuron degeneration. The absence of weakness, atrophy, or denervation phenomena on electrophysiologic examination usually excludes ALS or other serious neurologic disease. Patients who have recovered from poliomyelitis may experience a delayed deterioration of motor neurons that presents clinically with progressive weakness, atrophy, and fasciculations. Its cause is unknown, but it is thought to reflect sublethal prior injury to motor neurons by poliovirus **(Chap. 210)**.

Rarely, ALS develops concurrently with features indicative of more widespread neurodegeneration. Neuropsychological testing may detect subtle cognitive impairment in ~15% of cases that clinically are purely ALS; these cognitive deficits worsen with disease progression. Importantly, one often encounters the combination of ALS and FTD in individuals who harbor *C9orf72* mutations. The simultaneous occurrence of these disorders reflects shared embryologic origins and transcription factor expression in corticospinal motor neurons and neurons implicated in FTD (von Economo neurons). Overall, up to 40% of FTD cases harbor mutations in the *C9orf72* gene. Beyond *C9orf72*, several other ALS genes can trigger both ALS and FTD (see Table 448-3). As another example of an atypical phenotype, prominent amyotrophy has been described as a dominantly inherited disorder in individuals with bizarre behavior and a movement disorder suggestive of parkinsonism; many such cases have now been ascribed to mutations that alter the expression of tau protein in the brain **(Chap. 443)**. An ALS-like disorder has also been described in some individuals with chronic traumatic encephalopathy **(Chap. 454)**, associated with deposition of TDP43 and neurofibrillary tangles in motor neurons.

PATHOGENESIS

The cause of sporadic ALS is not well defined, in part because there is no animal model for this form of ALS. Strikingly, motor neurons derived from stem cells of individuals with sporadic ALS can display diminished viability, suggesting that heritable factors play a role. Several mechanisms that impair motor neuron viability have been elucidated in rodents that harbor transgenes with mutant *SOD1, profilin-1*, or *C9orf72*. One may loosely group the genetic causes of ALS into three categories. In one group, the primary problem is inherent instability of the mutant proteins, with subsequent perturbations in protein degradation (SOD1, ubiquilin-1 and 2, p62). In the second category, the causative mutant genes perturb RNA processing, transport, and metabolism (C9orf73, TDP43, FUS). In the case of C9orf72, the molecular pathology is an expansion of an intronic hexanucleotide repeat (-GGGGCC-) beyond an upper normal of 30 repeats to hundreds or even thousands of repeats. As observed in other intronic repeat disorders such as myotonic dystrophy **(Chap. 460)** and spinocerebellar atrophy type 8 **(Chap. 450)**, the expanded intronic repeats generate expanded RNA repeats that form intranuclear foci and may confer toxicity by sequestering transcription factors or by undergoing noncanonical protein translation across all possible reading frames of the expanded RNA tracts. Importantly, the latter process generates lengthy dipeptides that are detected in the spinal fluid and are a unique biomarker for C9orf72 ALS. TDP43 and FUS are multifunctional proteins that bind RNA and DNA and shuttle between the nucleus and the cytoplasm, playing multiple roles in the control of cell proliferation, DNA repair and transcription, and gene translation, both in the cytoplasm and locally in dendritic spines in response to electrical activity. How mutations in *FUS/TLS* provoke motor neuron cell death is not clear, although this may represent loss of function of FUS/TLS in the nucleus or an acquired, toxic function of the mutant proteins in the cytosol. In the third group of ALS genes, the primary problem is defective axonal cytoskeleton and transport (dynactin, profilin-1). It is striking that variants in other genes influence survival in ALS but not ALS susceptibility. Intermediate-length polyglutamine-coding expansions (-CAG-) in the gene *ataxin-2* confer increased ALS susceptibility; suppression of ataxin-2 expression extends survival in transgenic ALS mice. Beyond the upstream, primary defects, it is also evident that the ultimate neuronal cell death process is complex, involving multiple cellular processes acting in diverse components of the motor neuron (dendrites, cell body, axons, neuromuscular junction) to accelerate cell death. These include but are not limited to excitotoxicity, defective autophagy, impairment of axonal transport, oxidative stress, activation of endoplasmic reticulum stress and the unfolded protein response, and mitochondrial dysfunction. In addition, the hexanucleotide expansions that cause C9orf72 ALS disrupt nucleocytoplasmic transport; the importance of this observation is underscored by the finding that mutations in the gene encoding GLE1, a protein that mediates mRNA export, cause an aggressive, infantile motor neuron disease.

Multiple studies have convincingly demonstrated that proliferating, activated nonneuronal cells such as microglia and astrocytes importantly influence the disease course, at least in ALS-transgenic mice. A striking additional finding in ALS and most neurodegenerative disorders is that miscreant proteins arising from gene defects in familial forms of these diseases are often implicated in sporadic forms of the same disorder. For example, some reports propose that nonheritable, posttranslational modifications in SOD1 are pathogenic in sporadic ALS; indeed, SOD1 aggregates are sometimes observed in spinal cord in sporadic ALS without SOD1 mutations.

TREATMENT

Amyotrophic Lateral Sclerosis

No treatment arrests the underlying pathologic process in ALS. The drug riluzole (100 mg/d) was approved for ALS because it produces a modest lengthening of survival. In one trial, the survival rate at 18 months with riluzole was similar to placebo at 15 months. The mechanism of this effect is not known with certainty; riluzole may reduce excitotoxicity by diminishing glutamate release. Riluzole is generally well tolerated; nausea, dizziness, weight loss, and elevated liver enzymes occur occasionally. A second drug, edaravone, has also been approved by the U.S. Food and Drug Administration (FDA) based on a single 6-month study in a highly selected ALS population that demonstrated a modest reduction in the trajectory of worsening on an ALS disability scale; survival was not included as an endpoint. This drug, which is believed to act as an antioxidant, was initially administered via recurring monthly 10-day series of daily intravenous infusions. A formulation for oral use is now available.

Interventions such as antisense oligonucleotides (ASO) and microRNAs that diminish expression of mutant SOD1 protein prolong survival in transgenic-ALS rodent models and are also now under investigation in SOD1-mediated ALS. Tofersen, an ASO that suppresses SOD1 expression following intrathecal delivery, is now FDA approved for SOD1-mediated ALS. Pilot studies of an ASO targeting FUS/TLS have also been promising. Pathophysiologic studies of cell lines and animal models incorporating mutant SOD1, C9orf72, and other ALS genes have disclosed diverse targets for therapy; consequently, multiple therapies are presently in clinical trials for ALS including experimental trials of small molecules, mesenchymal stem cells, and immunosuppression.

In the absence of a primary therapy for ALS, a variety of rehabilitative aids may substantially assist ALS patients. Foot-drop splints facilitate ambulation by obviating the need for excessive hip flexion and by preventing tripping on a floppy foot. Finger-extension splints can potentiate grip. Respiratory support may be life-sustaining. For patients electing against long-term ventilation by tracheostomy, positive-pressure ventilation by mouth or nose provides transient (weeks to months) relief from hypercarbia and hypoxia. Also extremely beneficial for some patients is a respiratory device (cough assist machine) that produces an artificial cough. This is highly effective in clearing airways and preventing aspiration pneumonia. When bulbar disease prevents normal chewing and swallowing, gastrostomy is uniformly helpful, restoring normal nutrition and hydration. Fortunately, an increasing variety of speech synthesizers are now available to augment speech when there is advanced bulbar palsy. These facilitate oral communication and may be effective for telephone use.

In contrast to ALS, several of the disorders (Tables 448-1 and 448-3) that bear some clinical resemblance to ALS are treatable. For this reason, a careful search for causes of secondary motor neuron disease is warranted.

OTHER MOTOR NEURON DISEASES

SELECTED LOWER MOTOR NEURON DISORDERS

In these motor neuron diseases, the peripheral motor neurons are affected without evidence of involvement of the corticospinal motor system (Tables 448-1, 448-2, and 448-3).

X-Linked Spinobulbar Muscular Atrophy (Kennedy's Disease) This is an X-linked lower motor neuron disorder in which progressive weakness and wasting of limb and bulbar muscles begins in males in mid-adult life and is conjoined with androgen insensitivity manifested by gynecomastia and reduced fertility **(Chap. 403)**. In addition to gynecomastia, which may be subtle, two findings distinguishing this disorder from ALS are the absence of signs of pyramidal tract disease (spasticity) and the presence of a subtle sensory neuropathy in some patients. The underlying molecular defect is an expanded trinucleotide repeat (CAG) in the first exon of the androgen receptor gene on the X chromosome. An inverse correlation appears to exist between the number of CAG repeats and the age of onset of the disease.

Adult Tay-Sachs Disease Several reports have described adult-onset, predominantly lower motor neuropathies arising from deficiency of the enzyme β-hexosaminidase (hex A). These tend to be distinguishable from ALS because they are very slowly progressive and in some cases may have been symptomatic for years; dysarthria and radiographically evident cerebellar atrophy may be prominent. In rare cases, spasticity may also be present, although it is generally absent **(Chap. 429)**.

Spinal Muscular Atrophy The SMAs are a family of selective lower motor neuron diseases of early onset. Despite some phenotypic variability (largely in age of onset), the defect in the majority of families with SMA is loss of a protein (SMN, for survival motor neuron) that is important in the formation and trafficking of RNA complexes across the nuclear membrane. Neuropathologically, these disorders are characterized by extensive loss of large motor neurons; muscle biopsy reveals evidence of denervation atrophy. Several clinical forms exist.

Infantile SMA (SMA I, Werdnig-Hoffmann disease) has the earliest onset and most rapidly fatal course. In some instances, it is apparent even before birth, as indicated by decreased fetal movements late in the third trimester. Though alert, afflicted infants are weak and floppy (hypotonic) and lack muscle-stretch reflexes. Death generally ensues within the first year of life. *Chronic childhood SMA* (SMA II) begins later in childhood and evolves with a more slowly progressive course. *Juvenile SMA* (SMA III, Kugelberg-Welander disease) manifests during late childhood and runs a slow, indolent course. Unlike most denervating diseases, in this chronic disorder, weakness is greatest in the proximal muscles; indeed, the pattern of clinical weakness can suggest a primary myopathy such as limb-girdle dystrophy. Electrophysiologic and muscle biopsy evidence of denervation distinguish SMA III from the myopathic syndromes. Remarkably, two treatments have shown dramatic benefit in infantile SMA. One, nusinersen, now an approved therapy, entails administering small oligonucleotides that alter mRNA splicing of one of the SMN genes, generating sufficient normal SMN protein to provide clinical benefit (including prolonged survival). The other treatment uses systemically administered adeno-associated virus (AAV) to deliver the missing SMN gene to motor neurons and other cells.

Multifocal Motor Neuropathy with Conduction Block In this disorder, lower motor neuron function is regionally and chronically disrupted by focal blocks in conduction. Many cases have elevated serum titers of mono- and polyclonal antibodies to ganglioside GM1; it is hypothesized that the antibodies produce selective, focal, paranodal demyelination of motor neurons. MMCB is not typically associated with corticospinal signs. In contrast with ALS, MMCB may respond dramatically to therapy such as IV immunoglobulin or chemotherapy; thus, it is imperative that MMCB be excluded when considering a diagnosis of ALS.

Other Forms of Lower Motor Neuron Disease In individual families, other syndromes characterized by selective lower motor neuron dysfunction in an SMA-like pattern have been described. There are rare X-linked and autosomal dominant forms of apparent SMA. There is an ALS variant of juvenile onset, the Fazio-Londe syndrome, that involves mainly the musculature innervated by the brainstem. A component of lower motor neuron dysfunction is also found in degenerative disorders such as Machado-Joseph disease and the related olivopontocerebellar degenerations **(Chap. 450)**. Finally, a group of lower motor neuron disorders, sometimes mimicking Charcot-Marie-Tooth disease, and sometimes spinal muscular atrophy, are caused by mutations in the enzymes (tRNA synthetases) that charge tRNA with specific amino acids, an early step in protein synthesis.

SELECTED DISORDERS OF THE UPPER MOTOR NEURON

Primary Lateral Sclerosis This rare disorder arises sporadically in adults in mid-to-late life. Clinically, PLS is characterized by progressive spastic weakness of the limbs, preceded or followed by spastic dysarthria and dysphagia, indicating combined involvement of the corticospinal and corticobulbar tracts. Fasciculations, amyotrophy, and sensory changes are absent; neither electromyography nor muscle biopsy shows denervation. On neuropathologic examination, there is selective loss of the large pyramidal cells in the precentral gyrus and degeneration of the corticospinal and corticobulbar projections. The peripheral motor neurons and other neuronal systems are spared. The course of PLS is usually indolent; infrequently, there is conversion to a more aggressive course with lower motor neuron degeneration as in ALS. Early in its course, PLS raises the question of multiple sclerosis, other demyelinating diseases, or adult-onset spastic paraplegia as diagnostic considerations **(Chap. 455)**. A myelopathy suggestive of PLS is infrequently seen with infection with the retrovirus human T-cell lymphotropic virus 1 (HTLV-1) **(Chap. 453)**. The clinical course and laboratory testing will distinguish these possibilities.

Hereditary Spastic Paraplegia In its pure form, HSP is usually transmitted as an autosomal trait; most adult-onset cases are dominantly inherited. There are >80 genetic types of HSP for which causative mutations in >60 genes have been identified. Table 448-3 lists more commonly identified genetic types of HSP. Symptoms usually begin in the third or fourth decade of life, presenting as progressive spastic weakness beginning in the lower extremities; however, there are variants with onset so early that the differential diagnosis includes cerebral palsy. HSP typically has a long survival, presumably because respiratory function is spared. Late in the illness, there may be urinary urgency and incontinence and sometimes fecal incontinence; sexual function tends to be preserved.

In pure forms of HSP, the spastic leg weakness is often accompanied by posterior column (vibration and position) abnormalities and disturbance of bowel and bladder function. Some family members may have spasticity without clinical symptoms.

By contrast, particularly when recessively inherited, HSP may have complex or complicated forms in which altered corticospinal and dorsal column function is accompanied by significant involvement of other regions of the nervous system, including amyotrophy, intellectual disability, optic atrophy, and sensory neuropathy.

Neuropathologically, in HSP, there is degeneration of the corticospinal tracts, which appear nearly normal in the brainstem but show increasing atrophy at more caudal levels in the spinal cord; in effect, this pathologic picture is of a dying-back or distal axonopathy of long neuronal fibers within the CNS.

Defects at numerous loci underlie both dominantly and recessively inherited forms of HSP (Table 448-3). The gene most commonly implicated in dominantly inherited HSP is *spastin*, which encodes a microtubule interacting protein. The most common childhood-onset dominant form arises from mutations in the *atlastin* gene.

An infantile-onset form of X-linked, recessive HSP arises from mutations in the gene for myelin proteolipid protein. This is an example of rather striking allelic variation, as most other mutations in the same gene cause not HSP but Pelizaeus-Merzbacher disease, a widespread disorder of CNS myelin. Another recessive variant is caused by defects in the *paraplegin* gene. Paraplegin has homology to metalloproteases that are important in mitochondrial function in yeast. A slowly progressive, adult-onset X-linked progressive spastic paralysis designated adrenomyeloneuropathy is caused by mutations in the *ABCD1* gene; these cases are associated with elevated serum levels of very-long-chain fatty acids **(Chap. 453)**.

FURTHER READING

Akçimen F et al: Amyotrophic lateral sclerosis: Translating genetic discoveries into therapies. Nat Rev Genetics 44:642, 2023.

Baryshnikov VA et al: Antisense oligonucleotide silencing of FUS expression as as therapeutic approach in amyotrophic lateral sclerosis. Nat Med 28(1):104, 2022.

Brown RH, Al-Chalabi A: Review article: Amyotrophic lateral sclerosis. N Engl J Med 377:162, 2017.

Chio A et al: Cognitive impairment across ALS clinical stages in a population cohort. Neurology 93:e984, 2018.

Finkel RS et al: Treatment of infantile-onset spinal muscular atrophy with nusinersin: A phase 2, open-label, dose-escalation study. Lancet 388:3017, 2016.

Gendron TF et al: Poly(GP) proteins are a useful pharmacodynamic marker for C9ORF72-associated amyotrophic lateral sclerosis. Sci Transl Med 9:pii:eaai7866, 2017.

Miller TM et al: Trial of antisense oligonucleotide tofersen for ALS. N Engl J Med 387:1099, 2022.

Mueller C et al: SOD1 suppression with adeno-associated virus and microRNA in familial ALS. N Engl J Med 383:151, 2020.

Schüle R et al: Hereditary spastic paraplegia: Clinicogenetic lessons from 608 patients. Ann Neurol 79:646, 2016.

Shahim P et al: Neurofilaments in sporadic and familial amyotrophic lateral sclerosis: A systematic review and meta-analysis. Genes 15:496, 2024.

Taylor JP et al: Decoding ALS: From genes to mechanism. Nature 539:197, 2016.

Van Damme P, Robberecht W: STING-induced inflammation—A novel therapeutic target in ALS? N Engl J Med 384:765, 2021.

Visser AE et al: Multicentre, population-based, case-control study of particulates, combustion products and amyotrophic lateral sclerosis risk. J Neurol Neurosurg Psychiatry 90:854, 2019.

WEBSITES

Several websites provide valuable information on ALS including those offered by the Muscular Dystrophy Association (*www.mdausa.org*), the Amyotrophic Lateral Sclerosis Association (*www.alsa.org*), the World Federation of Neurology and the Neuromuscular Unit at Washington University in St. Louis (*www.neuro.wustl.edu*), and the Northeast Amyotrophic Lateral Sclerosis Consortium (*www.neals.org*).

449

Prion Diseases

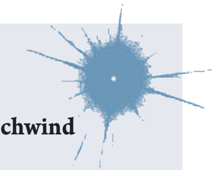

Stanley B. Prusiner, Michael D. Geschwind

Prions are proteins that adopt alternative conformations, which become self-propagating. Some prions cause degeneration of the central nervous system (CNS). Once relegated to causing a group of rare CNS disorders, such as Creutzfeldt-Jakob disease (CJD), increasing evidence argues that prions also cause more common neurodegenerative diseases (NDs) including Alzheimer's disease (AD) and Parkinson's disease (PD). While CJD is caused by the accumulation of PrP^{Sc} prions **(Table 449-1)**, α-synuclein prions cause multiple system atrophy (MSA) **(Chap. 451)**. Infectious MSA prions have been recovered from human brain samples stored in formalin for up to 20 years. Similar resistance to formalin was demonstrated for brain samples from sheep with scrapie. Increasing data suggest that Aβ and tau prions together may cause AD, α-synuclein prions PD in addition to MSA, and tau prions frontotemporal lobar degeneration (FTLD). CJD typically presents as a rapidly progressive dementia accompanied by other motor abnormalities and behavioral changes. The illness is relentlessly progressive and generally causes death within ~7 months from onset. Most patients with sporadic CJD (sCJD) are between 50 and 75 years of age, although patients as young as 12 and as old as 96 have been described.

TABLE 449-1 Glossary of PrP Prion Terminology

Prion	*Pro*teinaceous *in*fectious particle that lacks nucleic acid. Prions are composed entirely of alternatively folded proteins that undergo self-propagation. Distinct strains of prions exhibit different biologic properties, which are epigenetically heritable. PrP prions cause scrapie in sheep and goats, mad cow disease, and related neurodegenerative diseases of humans such as Creutzfeldt-Jakob disease (CJD).
PrP^{Sc}	Disease-causing *S*crapie isoform of the prion protein. This protein is the only identifiable macromolecule in purified preparations of scrapie prions.
PrP^{C}	*C*ellular isoform of the prion protein. PrP^{C} is the precursor of PrP^{Sc}.
PrP 27-30	A fragment of PrP^{Sc}, generated by truncation of the NH$_2$-terminus by limited digestion with proteinase K. PrP 27-30 retains prion infectivity and polymerizes into amyloid.
PRNP	PrP gene located on human chromosome 20.
Prion rod	An aggregate of prions composed largely of PrP 27-30 molecules. Created by detergent extraction and limited proteolysis of PrP^{Sc}. Morphologically and histochemically indistinguishable from many amyloids.
PrP amyloid	Amyloid containing PrP in the brains of animals or humans with prion disease; often accumulates as plaques.

TABLE 449-2 The PrP Prion Diseases

DISEASE	HOST	MECHANISM OF PATHOGENESIS
Human		
Kuru	Fore people	Infection through ritualistic cannibalism
iCJD	Humans	Infection from prion-contaminated hGH, dura mater grafts, etc.
vCJD	Humans	Infection from bovine prions
fCJD	Humans	Germline mutations in *PRNP*
GSS	Humans	Germline mutations in *PRNP*
FFI	Humans	Germline mutation in *PRNP* (D178N, M129)
sCJD	Humans	Somatic mutation or spontaneous conversion of PrP^{C} into PrP^{Sc}?
sFI	Humans	Somatic mutation or spontaneous conversion of PrP^{C} into PrP^{Sc}?
Animal		
Scrapie	Sheep, goats	Infection in genetically susceptible sheep and goats
BSE	Cattle	Infection with prion-contaminated MBM
TME	Mink	Infection with prions from sheep or cattle
CWD	Mule deer, elk, or moose	Unknown
FSE	Cats	Infection with prion-contaminated beef
Exotic ungulate encephalopathy	Greater kudu, nyala, or oryx	Infection with prion-contaminated MBM

Abbreviations: BSE, bovine spongiform encephalopathy; CJD, Creutzfeldt-Jakob disease; CWD, chronic wasting disease; fCJD, familial Creutzfeldt-Jakob disease; FFI, fatal familial insomnia; FSE, feline spongiform encephalopathy; GSS, Gerstmann-Sträussler-Scheinker disease; hGH, human growth hormone; iCJD, iatrogenic Creutzfeldt-Jakob disease; MBM, meat and bone meal; sCJD, sporadic Creutzfeldt-Jakob disease; sFI, sporadic fatal insomnia; TME, transmissible mink encephalopathy; vCJD, variant Creutzfeldt-Jakob disease.

The potential role of prions in the pathogenesis of NDs is reviewed in **Chap. 435**.

CJD is one malady in a group of disorders caused by prions composed of the human prion protein (PrP). PrP prions reproduce by binding to the normal, *c*ellular isoform of the *p*rion *p*rotein (PrP^{C}) and stimulating conversion of PrP^{C} into the disease-causing isoform PrP^{Sc}. PrP^{C} is rich in α-helix and has little β-structure, whereas PrP^{Sc} has less α-helix and a high amount of β-structure. The α-to-β structural transition in PrP is the fundamental event underlying this group of prion diseases.

Four new concepts have emerged from studies of PrP prions: (1) Prions are the only known transmissible pathogens that are devoid of nucleic acid; all other infectious agents possess genomes composed of either RNA or DNA that direct the synthesis of their progeny. (2) Prion diseases may manifest as infectious, genetic, or sporadic disorders; no other group of illnesses with a single etiology presents with such a wide spectrum of clinical manifestations. (3) Prion diseases result from the accumulation of PrP^{Sc}, the conformation of which differs substantially from that of its precursor, PrP^{C}. (4) Distinct strains of prions exhibit different biologic properties, which are epigenetically inherited. In other words, PrP^{Sc} can exist in a variety of different conformations, many of which seem to specify disease phenotypes.

How a specific conformation of a PrP^{Sc} molecule is imparted to PrP^{C} during prion replication to produce nascent PrP^{Sc} with the same conformation is not well understood. Additionally, it is unclear what factors determine where in the CNS a particular PrP^{Sc} molecule will be created.

SPECTRUM OF PrP PRION DISEASES

The sporadic form of CJD is the most common PrP prion disorder in humans. sCJD accounts for ~85% of all cases of human PrP prion disease, and genetic prion diseases account for 10–15% of all cases **(Table 449-2)**. Genetic prion diseases were historically divided into three forms: familial CJD (fCJD), Gerstmann-Sträussler-Scheinker (GSS) disease, and fatal familial insomnia (FFI). All dominantly inherited PrP prion diseases are caused by mutations in the PrP gene.

Although infectious PrP prion diseases account for <1% of all cases and infection does not seem to play an important role in the natural history of these illnesses, the transmissibility of PrP prions is an important biologic feature. Kuru of the Fore people of Papua New Guinea resulted from the consumption of brains from dead relatives during ritualistic cannibalism. After the cessation of this practice in the late 1950s, kuru nearly disappeared, with the exception of a few recent patients exhibiting incubation periods of >50 years. Iatrogenic CJD (iCJD) results from the accidental inoculation of patients with prions through medical procedures such as cadaver-derived dura mater grafts and human pituitary hormones. Variant CJD (vCJD) that mostly occurs in teenagers and young adults in Europe, predominantly the United Kingdom and France, is the result of exposure to tainted beef from cattle with bovine spongiform encephalopathy (BSE). Although occasional cases of iCJD still occur, this form of CJD is currently on the decline due to public health measures aimed at preventing the spread of PrP prions.

More than seven diseases of animals are caused by prions (Table 449-2). Scrapie of sheep and goats is the prototypic PrP prion disease. Mink encephalopathy, BSE, feline spongiform encephalopathy, exotic ungulate encephalopathy, and nonhuman primate prion disease are all thought to occur after the consumption of prion-infected foodstuffs. The BSE epidemic emerged in Britain in the late 1980s and was shown to be due to industrial cannibalism. Whether BSE began as a sporadic case of BSE in a cow or started with scrapie in sheep is unknown. The origin of chronic wasting disease (CWD), a prion disease endemic in deer and elk in regions of North America, and more recently identified in isolated populations in Scandinavia and Korea, is uncertain. In contrast to other prion diseases, CWD is highly transmissible among cervids. Bodily excretions, such as feces, urine, and saliva, from asymptomatic, infected cervids contain prions that are likely to be responsible for the spread of CWD. Notably, mink are carnivores and mink encephalopathy is spread from one animal to another.

■ EPIDEMIOLOGY

CJD is found throughout the world. The incidence of sCJD is ~1–2 cases per million population, although a person's lifetime risk of dying from CJD is ~1 in 5000 to 6000 deaths. Because sCJD is an age-dependent ND, its incidence is expected to increase steadily as older segments of populations in developed and developing countries continue to expand. Although many geographic clusters of CJD have been reported, each has been shown to segregate with a PrP gene mutation and/or included misdiagnoses. Attempts to identify common exposure to some etiologic agent have been unsuccessful for both the sporadic and familial cases. Ingestion of scrapie-infected sheep or goats as a cause of CJD in humans has not been demonstrated, and epidemiologic studies do not support this, although speculation about this potential route of infection continues. Whether PrP prion disease in deer, elk, or moose has passed to cows, sheep, or directly to humans remains unknown. Studies with mice modified to carry the human *PRNP* gene demonstrate that oral infection with CWD prions can occur, but the process is inefficient compared to intracerebral inoculation. The U.S. Centers for Disease Control and Prevention (CDC) conducts surveillance of CJD in the United States to ascertain the number and type of cases annually. Because up to 90% of culled deer in some game herds have been shown to harbor CWD prions, the CDC also has a study following deer hunters to determine if they have an increased rate of prion disease and whether it is a novel prion disorder.

PATHOGENESIS

The human PrP prion diseases were initially classified as NDs of unknown etiology. Even though the familial nature of GSS and a subset of CJD cases was well described, the significance of this observation became more obscure with the transmission of GSS and CJD to animals since genetic NDs were not considered transmissible. With the transmission of kuru and CJD to nonhuman primates, investigators began to view these diseases as infectious CNS illnesses caused by slow viruses. Eventually, the true cause of GSS and a minority of CJD cases became clear with the discovery in 1989 of mutations in the *PRNP* gene of these familial patients. The prion concept explains how a single disease can manifest as sporadic, heritable (i.e., genetic), and infectious. Moreover, the hallmark of all PrP prion diseases, whether sporadic, dominantly inherited, or acquired by infection, is that they involve the aberrant folding of the PrP protein.

A major feature that distinguishes PrP prions from viruses is the finding that both the normal and disease-causing PrP isoforms are encoded by a chromosomal gene. In humans, the PrP gene is designated *PRNP* and is located on the short arm of chromosome 20. Limited proteolysis of PrP^{Sc} produces a smaller, protease-resistant molecule of ~142 amino acids designated PrP 27-30, whereas PrP^{C} is completely hydrolyzed under the same conditions **(Fig. 449-1)**. PrP 27-30 polymerizes into prion rods that are morphologically indistinguishable from the filaments that aggregate to form PrP amyloid plaques in the CNS. This discovery raised the possibility that many other NDs might be caused by different proteins, all of which can fold into prions.

Prion Strains Distinct strains of PrP prions exhibit different biologic properties, which are epigenetically heritable. The existence of prion strains raised the question of how heritable biologic information can be enciphered in a molecule other than nucleic acid. Various strains of PrP prions have been defined by incubation times, distribution of neuronal vacuolation (i.e., spongiform change) on neuropathology, and stabilities of PrP^{Sc} to denaturation. Subsequently, the patterns of PrP^{Sc} deposition were found to correlate with the neuroanatomic location and pattern of vacuolation, and these patterns were also used to characterize prion strains.

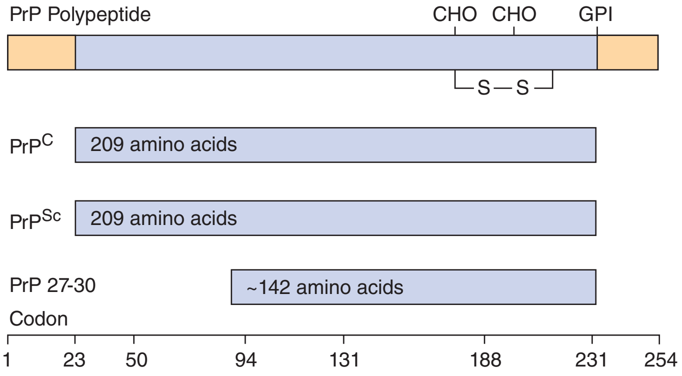

FIGURE 449-1 PrP prion protein isoforms. Bar diagram of Syrian hamster PrP, which consists of 254 amino acids. After processing of the NH$_2$ and COOH termini, both PrP^{C} and PrP^{Sc} consist of 209 residues. After limited proteolysis, the NH$_2$ terminus of PrP^{Sc} is truncated to form PrP 27-30 composed of ~142 amino acids. CHO, N-linked sugars; GPI, glycosylphosphatidylinositol anchor attachment site; S–S, disulfide bond.

Persuasive evidence that strain-specific information is enciphered in the tertiary structure of PrP^{Sc} comes from transmission of two different inherited human prion diseases to mice expressing a chimeric human–mouse PrP transgene. In most forms of fCJD and the majority of sCJD cases, the protease-resistant fragment of PrP^{Sc} after deglycosylation has a molecular mass of 21 kDa (i.e., type 1 prions), whereas in FFI, and a minority of sCJD cases, it is 19 kDa (type 2 prions) **(Table 449-3)**. This difference in molecular mass was shown to be due to different sites of proteolytic cleavage at the NH$_2$ termini of the two human PrP^{Sc} molecules, reflecting different tertiary structures. These distinct conformations were not unexpected because the amino acid sequences of the PrP fragments differ. Extracts from the brains of patients with FFI transmitted disease to the mice expressing the chimeric human–mouse PrP transgene and resulted in the formation of 19-kDa PrP^{Sc}, whereas brain extracts from patients with fCJD and sCJD harboring 21-kDa PrP^{Sc} resulted in 21-kDa PrP^{Sc} in mice expressing the same transgene. On second passage, these differences were maintained, demonstrating that chimeric PrP^{Sc} can exist in two different conformations as demonstrated by the sizes of the protease-resistant fragments, even though the amino acid sequence of PrP^{Sc} is invariant.

This analysis was extended when patients with sporadic fatal insomnia (sFI) were identified. Although they did not carry a *PRNP* mutation, the patients demonstrated a clinical and pathologic phenotype that was indistinguishable from that of patients with FFI. Furthermore, 19-kDa PrP^{Sc} was found in their brains, and on passage of sFI prion disease to mice expressing the chimeric human–mouse PrP transgene, 19-kDa PrP^{Sc} was also found. These findings indicate that the disease phenotype is dictated by the conformation of PrP^{Sc} and not the amino acid sequence. PrP^{Sc} acts as a template for the conversion of PrP^{C} into nascent PrP^{Sc}. On the passage of prions into mice expressing a chimeric hamster–mouse PrP transgene, a change in the conformation of PrP^{Sc} was accompanied by the emergence of a new strain of prions.

Many new strains of prions were generated using recombinant PrP (recPrP) produced in bacteria; recPrP was polymerized into amyloid fibrils to make "synthetic prions," which were inoculated into transgenic mice overexpressing high levels of wild-type mouse PrP^{C}. Approximately 500 days later, the mice died of prion disease. The incubation times (i.e., time to clinical disease onset) of the "synthetic prions" in mice were dependent on the conditions used for polymerization of the amyloid fibrils, which affected the stability of those amyloid fibrils. Highly stable amyloids gave rise to stable prions with long incubation times; low-stability amyloids led to prions with short incubation times. Amyloids of intermediate stability gave rise to prions with intermediate stabilities and intermediate incubation times. Such findings are consistent with earlier studies showing that the incubation times of synthetic and naturally occurring prions are directly proportional to the stability of the prion.

Species Barrier Studies on the role of the primary and tertiary structures of PrP in the transmission of prion disease have provided new insights into the pathogenesis of these maladies. The amino acid sequence of PrP encodes the species of the prion, and the prion derives its PrP^{Sc} sequence from the last mammal in which it was passaged. While the primary structure (i.e., amino acid sequence) of PrP is likely to be the most important or even the sole determinant of the tertiary structure of PrP^{C}, PrP^{Sc} seems to function as a template in determining the tertiary structure of nascent PrP^{Sc} molecules as they are formed from PrP^{C}. In turn, prion diversity appears to be enciphered in the conformation of PrP^{Sc}, and thus prion strains seem to represent different conformers of PrP^{Sc}.

In general, transmission of PrP prion disease from one species to another is inefficient, in that not all intracerebrally inoculated animals develop disease, and those that fall ill do so only after long incubation times that can approach the natural life span of the animal. This "species barrier" to transmission is correlated with the degree of similarity between the amino acid sequences of PrP^{C} in the inoculated host and of PrP^{Sc} in the inoculum. The importance of sequence similarity between

TABLE 449-3 Distinct Prion Strains Generated in Humans with Inherited Prion Diseases and Transmitted to Transgenic Mice[a]

INOCULUM	HOST SPECIES	HOST PrP GENOTYPE	INCUBATION TIME [DAYS ± SEM] (n/n_0)	PrP^{Sc} (kDa)
None	Human	FFI(D178N, M129)		19
FFI	Mouse	Tg(MHu2M)	206 ± 7 (7/7)	19
FFI → Tg(MHu2M)	Mouse	Tg(MHu2M)	136 ± 1 (6/6)	19
None	Human	fCJD(E200K)		21
fCJD	Mouse	Tg(MHu2M)	170 ± 2 (10/10)	21
fCJD → Tg(MHu2M)	Mouse	Tg(MHu2M)	167 ± 3 (15/15)	21

[a] Tg(MHu2M) mice express a chimeric mouse–human PrP gene.

Notes: Clinicopathologic phenotype is determined by the conformation of PrP^{Sc} in accord with the results of the transmission of human prions from patients with FFI to transgenic mice.

Abbreviations: fCJD, familial Creutzfeldt-Jakob disease; FFI, fatal familial insomnia; SEM, standard error of the mean.

the host and donor PrP argues that PrP^{C} directly interacts with PrP^{Sc} in the prion conversion process.

SPORADIC AND INHERITED PRP PRION DISEASES

Several different scenarios might explain the initiation of sporadic prion disease: (1) A somatic mutation in a single cell may be the cause and thus follow a path similar to that for germline mutations in inherited disease. In this situation, the primary structure of PrP^{C} made from the mutated gene would be more susceptible to misfolding into PrP^{Sc}. This mutant PrP^{Sc} then must be capable of targeting wild-type PrP^{C}, a process known to be possible for some mutations (i.e., high penetrance) but less likely for others (low penetrance). (2) The activation energy barrier separating wild-type PrP^{C} from PrP^{Sc}, preventing conversion to PrP^{Sc}, could be crossed on rare occasions in the context of a population. Most individuals would be spared, but presentations in older persons who have had more time for this conversion to occur would be seen. (3) PrP^{Sc} may be present at low levels in some normal cells, where it performs an important, but yet unknown, function. The level of PrP^{Sc} in such cells is hypothesized to be sufficiently low as not to be detected by routine bioassay. In some altered metabolic states, the cellular mechanisms for clearing PrP^{Sc} might become compromised, and the rate of PrP^{Sc} formation would then begin to exceed the capacity of the cell to clear it. The third possible mechanism is attractive because it suggests that PrP^{Sc} is not simply a misfolded protein, as proposed for the first and second mechanisms, but that it is an alternatively folded molecule with a function. Moreover, the multitude of conformational states that PrP^{Sc} can adopt, as described above, raises the possibility that PrP^{Sc} or another protein might function in a process such as short-term memory where information storage is thought to occur in the absence of new protein synthesis.

More than 40 different mutations resulting in nonconservative substitutions in the human *PRNP* gene have been found to segregate with inherited human prion diseases. Missense mutations, a deletion, expansions in the octapeptide repeat region of the gene, called octapeptide repeat insertions (OPRIs), and stop codon mutations are responsible for genetic forms of prion disease.

Although phenotypes may vary dramatically, even within families, specific phenotypes observed with certain mutations appear to cause fCJD. More than 20 missense variants—including substitutions at codons 102, 105, 117, 198, and 217, and mid to longer OPRIs—cause the GSS form of PrP prion disease, with prominent parkinsonism and/or cerebellar features and typically onset of dementia later in the course. Regarding OPRI mutations, the normal human PrP sequence contains an unstable section in the N-terminal region comprised of five repeats—a nine-amino-acid sequence or nonapeptide (R1) followed by four octapeptide repeats (R2, R2, R3, R4), which includes two tandem R2 domains. Insertions from 2 to 12 extra octapeptide repeats cause variable phenotypes including conditions indistinguishable from sCJD, GSS-like presentations, and even a slowly progressive dementing illness of many years' duration to an early-age-of-onset disorder that is similar to AD. A mutation at codon 178 that results in substitution of asparagine for aspartic acid generally causes FFI if methionine is encoded at codon 129 on the same allele. In contrast, a typical CJD phenotype generally occurs when there is a valine at codon 129 of the same allele. Stop codon (nonsense) mutations are rare and cause a range of phenotypes, including some with a prolonged course of years to decades, GSS- or AD-like presentations, autonomic and sensory peripheral nervous system involvement, chronic gastrointestinal upset, and extensive PrP^{Sc} amyloid deposits.

■ HUMAN *PRNP* GENE POLYMORPHISMS

Polymorphisms influence the susceptibility to sporadic, genetic, and acquired forms of PrP prion disease. The methionine [M] or valine [V] polymorphism at codon 129 of human *PRNP* not only modulates the age of onset of some genetic prion diseases but also can affect the clinical phenotype. Sporadic CJD can be divided into six different molecular subtypes, based on the combination of codon 129 polymorphism (MM, MV, or VV) and the prion type (1 or 2), with each subtype having a particular clinical and pathological presentation. MM1/MV1 subtypes are the most common (~40–70% of cases) and usually have the most prototypic form of sCJD with rapidly progressive dementia, ataxia, and myoclonus and a mean survival of ~4–7 months. The VV2 subtype represents ~15% of sCJD cases, usually starts with ataxia, and has a similar survival as MM1/MV1. The MV2 subtype represents ~10% of cases and has a longer mean survival of ~17 months. The vast majority of MV2 cases are a form with kuru plaques in the cerebellum, called MV2K, which are clinically similar to VV2 (i.e., early ataxia) but have a slower progression and longer survival. A minority of MV2 cases are of a cortical subtype called MV2-cortical (MV2C) with significant vacuolation (spongiform change) surrounded by perivacuolar PrP^{Sc} staining in all cortical layers and are usually without kuru plaques or cerebellar involvement. The MV2C cases present as a slowly progressive cognitive/dementia syndrome with motor symptoms occurring late in the disease course. The MM2 subtype represents ~4% of sCJD cases, has a mean survival of ~15.5 months, and is divided about equally into two subtypes: MM2-thalamic (MM2T, also called sFI) and MM2-cortical (MM2C). MM2T is clinicopathologically nearly identical to FFI (see below), whereas MM2C presents similarly to MV2C with a relatively slow progressive dementia and has a mean age of onset in the 50s, about a decade younger than most sCJD. VV1 is the least common subtype, representing ~1% of cases, presenting as a progressive dementia with a mean age of onset in the mid to late 40s, about two decades earlier than most other sCJD subtypes.

Substitution of the basic residue lysine for glutamine at position 218 in mouse PrP produced dominant-negative inhibition of prion replication in transgenic mice. This same lysine substituted for glutamine at position 219 in human PrP has been found in 12% of the Japanese population, a group that appears to be resistant to prion disease. Dominant-negative inhibition of prion replication was also found with substitution of the basic residue arginine at position 171; sheep with arginine were resistant to scrapie prions but were susceptible to BSE prions that were inoculated intracerebrally. A very interesting polymorphism at codon 127 in *PRNP* was identified among longtime survivors of the kuru epidemic in the Fore people of Papua New Guinea, which when expressed in transgenic mice with humanized *PRNP* prevented the animals from acquiring prion disease.

ACQUIRED (TRANSMITTED) PrP PRION DISEASES

IATROGENIC CJD

Accidental transmission of CJD to humans through medical procedures (i.e., iatrogenic) appears to have occurred with cadaver-derived human pituitary hormones, dura mater grafts, and corneal transplants, as well as through contaminated electroencephalogram (EEG) electrode implantation and possibly through other neurosurgical procedures. Corneas from donors with unsuspected CJD have been transplanted to apparently healthy recipients who developed CJD after variable incubation periods. Two other cases arose due to contamination during epilepsy surgery from depth EEG electrodes previously used in a patient who unknowingly had CJD; these electrodes were subsequently implanted in a chimpanzee, causing CJD 18 months later.

Surgical procedures may have resulted in other accidental inoculations of patients with prions, presumably because some instrument or apparatus in the operating theater became contaminated when a CJD patient underwent surgery. Although the epidemiology of these studies is highly suggestive, no proof for such episodes exists.

Dura Mater Grafts More than 200 cases of CJD after implantation of dura mater grafts (dCJD) have been recorded. Dura mater is normally collected from cadavers, mass sterilized in a heated vat, freeze-dried, and prepared for use in a variety of surgical procedures. Unfortunately, some of the donor cadavers unknowingly had prion disease. All but possibly two of the grafts appear to have been acquired from a single manufacturer. More than two-thirds of the cases occurred in Japan. Patients with dCJD usually present with cerebellar ataxia, visual symptoms, and dementia and have a mean incubation period of 12 years (range 1.3–30 years). Two subtypes of dCJD have been identified in Japan, a type with PrP^{Sc} plaques and a type without plaques (synaptic PrP^{Sc}).

Human Growth Hormone and Pituitary Gonadotropin Therapy The transmission of CJD prions from contaminated human growth hormone (hGH) preparations derived from human pituitaries has been responsible for fatal cerebellar disorders with dementia in >200 patients ranging in age from 5 to 42 years, most occurring in France, the United Kingdom, and the United States. These patients received injections of hGH every 2–4 days for ~2–12 years. If it is thought that these patients developed CJD from injections of prion-contaminated hGH preparations, the possible incubation periods range from 4 to 30 years. Recombinant hGH is now exclusively used therapeutically so that possible contamination with prions is no longer an issue. Four cases of CJD also occurred in women in Australia receiving human pituitary gonadotropin, with incubation periods of 12–16 years.

Notably, there is some evidence that deceased patients who received hGH early in life may have inadvertently received Aβ prions also, which can lead to amyloid and even tau pathology. Whether iatrogenic propagation of Aβ or tau prions in the human CNS led to an ND, such as AD or cerebral amyloid angiopathy (CAA), in these patients is still controversial.

VARIANT CJD

The restricted geographic occurrence and chronology of vCJD raised the possibility that BSE prions had been transmitted to humans through the consumption of tainted beef. More than 200 cases of vCJD have occurred, with >90% of these in Britain. Variant CJD has also been reported in people either living in or originating from France, Ireland, Italy, the Netherlands, Portugal, Spain, Saudi Arabia, the United States, Canada, and Japan. For some of these patients, such as those from North America, evidence suggests they acquired the disease while living or traveling outside their home country.

The steady decline in the number of vCJD cases over the past decade argues that there will not be a prion disease epidemic in Europe, similar to those seen for BSE and kuru. What is certain is that PrP-prion-tainted meat should be prevented from entering the human food supply.

The most compelling evidence that vCJD is caused by BSE prions was obtained from experiments in mice expressing the bovine PrP transgene. Both BSE and vCJD prions were efficiently transmitted to these transgenic mice and with similar incubation periods. In contrast to sCJD prions, vCJD prions did not transmit disease efficiently to mice expressing a chimeric human–mouse PrP transgene. Earlier studies with nontransgenic mice suggested that vCJD and BSE might be derived from the same source because both inocula transmitted disease with similar but very long incubation periods.

Attempts to determine the origin of BSE and vCJD prions have relied on passaging studies in mice, some of which are described above, as well as studies of the conformation and glycosylation of PrP^{Sc}. One scenario suggests that a particular conformation of bovine PrP^{Sc} was selected for heat resistance during the rendering process and was then reselected multiple times as cattle infected by ingesting prion-contaminated meat and bone meal (MBM) were slaughtered and their offal rendered into more MBM. Variant CJD cases have virtually disappeared with protection of the beef supply in Europe. Interestingly, almost all of the ~238 cases of vCJD reported as of 2024 have been homozygous for methionine (MM) at codon 129 in *PRNP*. However, two cases (one probable and one definite) were codon 129 MV, which is the most common codon 129 polymorphism in most of the world. This finding raises the concern that persons with this polymorphism might have a longer incubation period and that another rise in cases might still occur. Of particular concern is that four known (and a fifth possible) secondary cases of vCJD infection occurred from blood product transfusions. These persons received blood components (non-leukodepleted red blood cells [RBCs] in the four known cases and factor X in the fifth case) from asymptomatic donors who later developed vCJD infection. The second of four RBC recipients did not die from vCJD but was found to have vCJD prions in the lymphoreticular system. The RBC donors did not develop vCJD infection until ~1.5–3.3 years after donation, and the incubation period for recipients of RBCs ranged from 5 to 8.5 years. Thus, vCJD is the only form of human prion disease proven to be transmissible by blood. Further evidence of the transmissibility of vCJD is that among the two most recent cases of vCJD identified, one in France and one in Italy, both had laboratory exposure to BSE-infected brain tissue. The French patient accidentally stabbed herself with forceps being used on frozen brain sections from a transgenic mouse overexpressing human PrP and inoculated with BSE. She developed symptoms 7.5 years later at age 31 and died from definite vCJD after 19 months in 2019. The last reported case of vCJD worldwide was in France in 2021, and the prior two cases occurred in the United Kingdom in 2013 and 2016.

NEUROPATHOLOGY

Frequently, the brains of patients with CJD have no recognizable abnormalities on gross examination. Patients who survive for several years have variable degrees of cerebral atrophy.

On light microscopy, the pathologic hallmarks of CJD are spongiform degeneration (vacuolation), neuronal loss, and astrocytic gliosis. The lack of an inflammatory response in CJD and other prion diseases is an important pathologic feature of these degenerative disorders. Spongiform degeneration is characterized by many 1- to 5-μm vacuoles in the neuropil between nerve cell bodies. Generally, the vacuolation occurs in the cerebral cortex, putamen, caudate nucleus, thalamus, and molecular layer of the cerebellum. Astrocytic gliosis is a constant but nonspecific feature of PrP prion diseases. Widespread proliferation of fibrous astrocytes is found throughout the gray matter of brains infected with CJD prions. Astrocytic processes filled with glial filaments form extensive networks. The degree and location of these pathologic hallmarks vary depending on the type of human prion disease, including between sCJD subtypes described above.

Amyloid plaques have been found in ~10% of CJD cases. Purified CJD prions from humans and animals exhibit the ultrastructural and histochemical characteristics of amyloid when treated with detergents during limited proteolysis. On first passage of samples from some human Japanese CJD cases into mice, amyloid plaques were found. These plaques stain with antibodies raised against PrP, demonstrating that the amyloid is composed of PrP.

The amyloid plaques of GSS disease are morphologically distinct from those seen in kuru or scrapie. GSS plaques consist of a central dense core of amyloid surrounded by smaller globules of amyloid. Ultrastructurally, they consist of a radiating fibrillar network of amyloid fibrils, with scant or no neuritic degeneration. The plaques can be distributed throughout the brain but are most frequently found in the cerebellum. They are often located adjacent to blood vessels. Congophilic angiopathy has been noted in some cases of GSS disease.

In vCJD, a characteristic feature is the presence of "florid plaques." These are composed of a central core of PrP amyloid, surrounded by vacuoles in a pattern suggesting petals on a flower.

CLINICAL FEATURES

Nonspecific prodromal symptoms occur in approximately a third of patients with CJD and may include fatigue, sleep disturbance, weight loss, headache, anxiety, vertigo, malaise, and ill-defined pain. Most patients with CJD present with cognitive and/or motor deficits. Behavioral and psychiatric symptoms, such as depression, anxiety, irritability, apathy, insomnia, appetite changes, psychosis, and visual hallucinations, are very common and often early features. These deficits usually progress over weeks or months to a state of profound dementia characterized by memory loss, impaired judgment, and a decline in virtually all aspects of intellectual function. A minority of patients present early with either isolated visual impairment or cerebellar gait and coordination deficits, referred to as the Heidenhain and Brownell-Oppenheim variants, respectively. Frequently, the cerebellar deficits are rapidly followed by progressive dementia. Visual problems often begin with blurred vision and diminished acuity, rapidly followed by dementia. Patients with early visual deficits often have a faster decline overall.

Motor symptoms and signs other than cerebellar ataxia include extrapyramidal dysfunction manifested as rigidity, masklike facies, dystonia, myoclonus, or less commonly choreoathetoid movements and pyramidal signs (usually mild and not actual weakness). Some uncommon features include seizures (usually major motor), hypoesthesia, supranuclear gaze palsy, motor neuron disease, or dysautonomic signs such as changes in body temperature and sweating.

Most patients with the most common subtype of CJD will eventually develop myoclonus. Unlike other involuntary movements, myoclonus usually persists during sleep. Startle myoclonus elicited by loud sounds, bright lights, or a person or object suddenly appearing in a patient's visual field is frequent. It is important to stress that myoclonus is neither specific nor confined to CJD and tends to occur later in the course of CJD. Dementia with myoclonus can also be due to AD **(Chap. 442)**, dementia with Lewy bodies **(Chap. 445)**, corticobasal degeneration **(Chap. 443)**, cryptococcal encephalitis **(Chap. 221)**, or the myoclonic epilepsy disorder Unverricht-Lundborg disease **(Chap. 436)**.

Clinical Course Most patients with sCJD and most types of fCJD live ~6–12 months after the onset of clinical signs and symptoms. Life expectancies can be longer, up to a few years, for less common sCJD subtypes, and some mutations causing genetic prion disease can have durations of a decade or longer.

DIAGNOSIS

The constellation of a rapid onset of cognitive impairment over weeks to months, with myoclonus and other motor symptoms, and typical magnetic resonance imaging (MRI) findings (see below) in an afebrile 60- to 70-year-old patient generally indicates CJD—most commonly sCJD.

Variations in the typical course appear in genetic and transmitted (i.e., acquired) forms of the disease. As noted above, most mutations causing fCJD have a slightly earlier mean age of onset, although usually an otherwise similar clinical and radiologic presentation as sCJD. In GSS, cerebellar ataxia is usually a prominent and presenting feature, with dementia occurring late in the disease course. GSS also presents earlier than sCJD (mean age ~43 years) and usually progresses more slowly, leading to death ~5 years after symptom onset. FFI is typically characterized by insomnia and dysautonomia; dementia occurs only in the terminal phase of the illness; survival is typically <2 years and sometimes just a few months.

Variant CJD has a different clinical course from most other prion diseases, with an early psychiatric prodrome (most commonly depression, anxiety, apathy, withdrawal, and/or delusions) that persists for several months prior to the appearance of other neurologic symptoms including cerebellar ataxia, painful sensory symptoms, a movement disorder (often myoclonus, dystonia, and/or chorea), and cognitive impairment progressing to dementia. The mean age of onset for vCJD is 28 years (median 26, range 12–74), with the majority of patients being <55 years old.

LABORATORY TESTS

The only highly specific diagnostic tests for CJD and other human PrP prion diseases measure PrP^{Sc}. The most widely used method involves limited proteolysis that generates PrP 27-30, which is detected by immunoassay after denaturation. In humans, the diagnosis of CJD can be established by brain biopsy if PrP^{Sc} is detected, although with current ancillary testing available, biopsy is rarely indicated. Because PrP^{Sc} is not uniformly distributed throughout the CNS, the absence of PrP^{Sc} in a limited sample such as a biopsy does not rule out prion disease. The use of reverse templated quake-induced conversion assay (RT-QuIC; see below), a method for amplifying prions into amyloid fibrils and detecting them with thioflavin fluorescence, has greatly increased the sensitivity of brain biopsy. If no attempt is made to measure or detect PrP^{Sc} but pathologic changes typical of CJD are seen in a brain biopsy, then the diagnosis is reasonably secure.

Brain MRI has become an important diagnostic tool for prion diseases, especially CJD, and can help distinguish CJD from most other conditions. The first reported MRI findings were on T2-weighted MRI and were hyperintensities in the striatum (caudate and putamen) and less commonly in the thalamus (depending on the type of prion disease). Subsequently, with fluid-attenuated inversion recovery (FLAIR) sequences, hyperintensity of the cortex (cortical ribboning) could also be seen. Diffusion MRI, with a combination of diffusion-weighted imaging (DWI) and attenuation deficient coefficient (ADC) sequences, greatly improved sensitivity and specificity (to mid to high 90th percentile) of MRI for prion disease, showing hyperintensity (i.e., brightness or high signal) on DWI with corresponding hypointensity (i.e., darkness or low signal) on ADC in the cortex (cortical ribboning) and striatum and less commonly in the thalamus and/or cerebellum **(Fig. 449-2)**. This abnormal signal on DWI and ADC is due to reduced or restricted diffusion of water molecules in these brain regions secondary to vacuolation. DWI and FLAIR MRI, however, often show areas of artifactual hyperintensity, particularly in regions where there is air adjacent to brain tissue, such as near sinuses. The true abnormal signal of reduced or restricted diffusion can be distinguished from artifact by (1) acquiring the diffusion MRI in multiple planes (e.g., axial and coronal), (2) looking for corresponding hypointensity in the cortex or deep nuclei on ADC, and/or increasing the degree of diffusion weighting. T2, FLAIR, and diffusion MRI are often normal, however, particularly in GSS, FFI, sFI, and in some rare genetic prion diseases such as those due to stop-codon mutations. Any prion disease with a long duration (e.g., >1–2 years) often will show nonspecific atrophy on brain MRI or head computed tomography (CT). To a limited extent, the pattern of these MRI abnormalities can also help determine the subtype of sCJD present. For example, MRI findings in the MM1/MV1 subtype are typically located in the cortex (i.e., cortical ribboning) and deep nuclei (striatum +/– thalamus), MV2K and VV2 involve deep nuclei, and MM2/MV2C are predominantly cortical. These abnormalities may be unilateral or bilateral; when they are bilateral, they may be symmetric or asymmetric. The pattern and type of MRI abnormalities in CJD are not seen with other NDs but can overlap with viral encephalitis, paraneoplastic/autoimmune encephalopathy syndromes, metabolic disorders, or seizures.

To establish the diagnosis of either sCJD or familial prion disease, sequencing the *PRNP* gene must be performed. Finding the wild-type *PRNP* gene sequence permits the diagnosis of sCJD if there is no history to suggest infection from an exogenous source of prions. The identification of a mutation in the *PRNP* gene sequence that encodes a nonconservative amino acid substitution argues for familial prion disease.

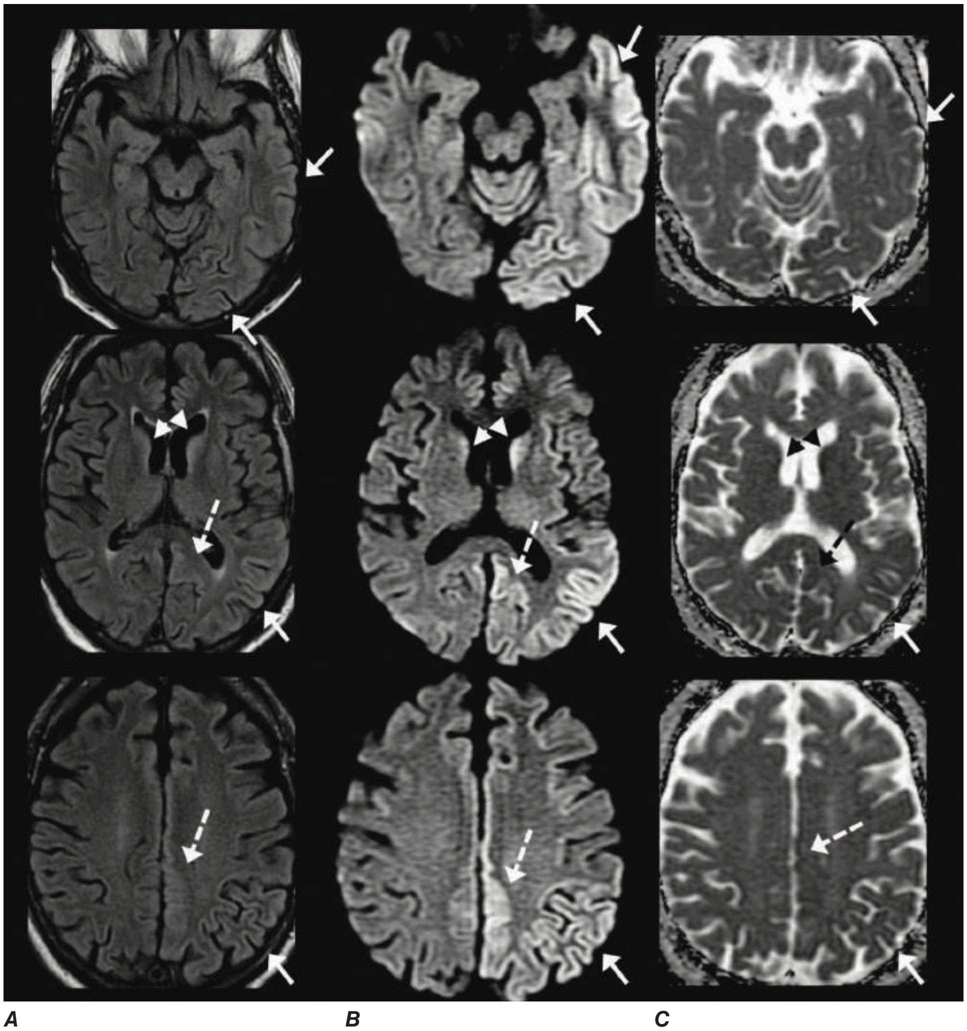

FIGURE 449-2 Brain magnetic resonance imaging (MRI) in a 72-year-old patient with sporadic Creutzfeldt-Jakob disease, MM1 subtype (classic subtype), showing axial (***A***) fluid-attenuated inversion recovery (FLAIR), (***B***) diffusion-weighted imaging (DWI), and (***C***) attenuation deficient coefficient (ADC) sequences at three different levels. There is cortical ribboning indicating restricted diffusion in the left much greater than right temporal (*solid white arrow*), occipital (*solid white arrow*), parietal (*solid white arrow*), insular (*no arrow*), and posterior cingulate cortices (*dashed arrows*). There is also restricted diffusion in the left greater than the right caudate head (*arrowheads*). The corresponding hypointensity (very dark) areas on the ADC sequence confirm that the hyperintensity (bright) areas on DWI and FLAIR are regions of reduced diffusion, not artifacts. Note that the abnormalities are best seen on DWI sequences. Images are radiologic orientation (right side of image is left side of brain).

General cerebrospinal fluid (CSF) laboratory testing (i.e., cell count, protein, glucose) is nearly always normal, except that mild and nonspecific protein elevation and, rarely, mild pleocytosis can be seen in a minority of cases. The first CSF surrogate biomarker used in clinical practice and still incorporated in several diagnostic criteria for sCJD is the 14-3-3 protein, which is elevated in many forms of brain cell injury. Although newer enzyme-linked immunosorbent assays (ELISAs) quantitatively measuring the 14-3-3γ isoform have improved sensitivity and specificity, this test is still less accurate diagnostically than some other biomarkers. The sensitivity of 14-3-3 ELISAs in sCJD ranges from ~60% to almost 90% depending on the sCJD subtype (highest in MM1/MV1), whereas the range of specificity has been reported to be as low as 40% to as high as 96% depending on the controls used, making its utility questionable. The level of total tau protein (t-tau), a microtubule-associated protein expressed in neurons and glia, is elevated in the CSF in many conditions associated with brain cell injury. ELISAs measuring t-tau appear to have a sensitivity and a specificity in the low to mid 90%, which is clinically superior to 14-3-3; however, there is no consensus at this time regarding the best t-tau ELISA or cutoff that should be used to support a diagnosis of sCJD.

Assays that amplify PrP^{Sc} into amyloid PrP^{Sc} fibrils and detect the fibrils by adding thioflavin, which binds amyloid and fluoresces, have greatly enhanced premortem diagnosis of human prion diseases. The first such assay developed was the protein misfolding cyclic amplification (PMCA), which was modified into the more user-friendly RT-QuIC assay that is now widely used in clinical practice but still primarily performed at national prion surveillance centers. The sensitivity and specificity of the current CSF RT-QuIC assay (i.e., second generation) for sCJD are about 90–95% and 98–99%, respectively. The advantage of PMCA and RT-QuIC is that they detect prions; however, false positives for RT-QuIC do occur, and as noted by the sensitivity, upward of 10% of sCJD cases are negative for RT-QuIC. As with other surrogate biomarkers, RT-QuIC is less sensitive for some uncommon sCJD subtypes, for several genetic prion diseases (i.e., GSS and other slower progressing forms), and particularly for vCJD. In contrast, PMCA seems to be more sensitive when applied to CSF and possibly blood and urine from patients with vCJD, but this test is not available for clinical practice in most countries. Improved methods to detect prions in blood, urine, skin, and even tears will be welcomed.

EEG can be useful in the diagnosis of CJD, particularly when other ancillary tests are unrevealing, although only ~60% of patients (mostly with the MM1/MV1 subtype of sCJD) show the typical pattern of periodic sharp wave complexes (PSWCs). PSWCs are usually repetitive, high-voltage, bi- or triphasic sharp discharges that normally appear quite late in the clinical course. Even when PSWCs are seen, they may be transient and might require serial EEGs to detect. During the early phase of CJD, the EEG is usually normal or shows only scattered theta and even delta slow wave activity. The presence of these stereotyped periodic bursts of PSWCs, <200 ms in duration and occurring every

1–2 s, makes the diagnosis of CJD very likely. These discharges are frequently but not always symmetric; there may be a one-sided predominance in amplitude. As CJD progresses, normal background rhythms become fragmentary and slower.

DIFFERENTIAL DIAGNOSIS

Many conditions mimic CJD, including various NDs, autoimmune/paraneoplastic encephalopathies or ataxias, infections, and even psychiatric conditions. Dementia with Lewy bodies (DLB) **(Chap. 445)** is one of the most common disorders to be mistaken for CJD, particularly when there is a phase of the illness with a fast decline. It can present rarely in a subacute fashion with delirium, myoclonus, and extrapyramidal features. Other NDs to consider include AD, frontotemporal dementia, corticobasal degeneration, progressive supranuclear palsy, ceroid lipofuscinosis, and myoclonic epilepsy with Lafora bodies. Usually when these diseases are mistaken for CJD, they have a more slowly progressive and perhaps subtle onset over a few years and then a sudden decline, which makes clinicians consider CJD. A thorough history of the earliest features of the illness—obtained through speaking with friends, family members, or coworkers—often reveals a slower onset over a few years with a more recent rapid decline, suggesting a non-CJD etiology. There have been, however, rapidly progressive cases of AD with a course of <3 years from first symptom onset, often with ataxia, myoclonus, and other symptoms similar to those seen in CJD. Many of these cases have elevated CSF biomarkers such as 14-3-3. The absence of abnormalities on diffusion MRI (i.e., DWI and ADC) will almost always distinguish these conditions from CJD. CSF RT-QuIC, if positive, can also be helpful.

Several autoantibody-mediated autoimmune encephalopathies (AEs) **(Chap. 99)**, such as anti-LGI1 (leucine-rich glioma inactivated 1), anti-Crmp5 (collapsin response-mediator protein-5 or anti-CV2), and anti-AMPAR (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor) antibody-mediated AEs, can have significant clinical overlap with CJD. Detection of autoantibodies in the serum and/or CSF, depending on the specific antibody-mediated syndrome, and the absence of brain MRI and CSF prion-specific biomarkers can help distinguish these cases from CJD. In AE with seizures, however, brain MRI may show diffusion MRI abnormalities similar to those in CJD, though this is rare, and these abnormalities disappear soon after seizures are treated. In contrast, these MRI abnormalities generally persist in CJD, except in very long-lived cases.

Intracranial vasculitides **(Chap. 375)** may produce nearly all the symptoms and signs associated with CJD, sometimes without systemic abnormalities. Myoclonus is uncommon with cerebral vasculitis, but focal seizures may confuse the diagnosis. Prominent headache, absence of myoclonus, stepwise change in deficits, abnormal CSF, and focal white matter change on MRI or angiographic abnormalities all favor vasculitis.

Other diseases that can simulate CJD include neurosyphilis **(Chap. 187)**, AIDS dementia complex **(Chap. 208)**, progressive multifocal leukoencephalopathy **(Chap. 142)**, subacute sclerosing panencephalitis, progressive rubella panencephalitis, herpes simplex encephalitis **(Chap. 142)**, diffuse intracranial tumor (gliomatosis cerebri; **Chap. 95**), anoxic encephalopathy, dialysis dementia, uremia, hepatic encephalopathy, and lithium or bismuth intoxication.

Fever, elevated sedimentation rate, leukocytosis in blood, or a pleocytosis in CSF should alert the physician to another etiology that explains the patient's CNS dysfunction, although there are rare cases of CJD in which mild CSF pleocytosis or mild elevation in IgG index or oligoclonal bands are observed.

CARE OF CJD PATIENTS

There are no disease-modifying treatments for prion diseases, and treatment is symptomatic.

Although CJD is communicable, the likelihood of transmission from one patient to another is remote. The risk of accidental inoculation by aerosols is minuscule; nonetheless, procedures producing aerosols should be performed in certified biosafety cabinets. Biosafety level 2 practices, containment equipment, and facilities are recommended by the CDC and the National Institutes of Health. The primary concern in caring for patients with CJD is the inadvertent infection of health-care workers by needle and stab wounds, although with the possible exception of vCJD (as noted above) in which blood transfusions appear to carry some minimal risk for transmission. When caring for patients with prion disease, standard universal precautions used in the clinical setting (e.g., gloves, gowns, and/or eye protection) are recommended when handling bodily fluids (e.g., blood, urine, and feces). Electroencephalographic and electromyographic needles should not be reused after studies on patients with CJD have been performed.

Autopsies on patients whose clinical diagnosis is CJD can be performed with minimal risk to pathologists or other morgue employees if proper prion-specific precautions are followed. Standard microbiologic practices outlined here, along with specific recommendations for decontamination, are generally adequate precautions for the care of patients with CJD and the handling of infected specimens.

DECONTAMINATION OF CJD PRIONS

Prions are generally resistant to commonly used inactivation procedures, and there is some disagreement about the optimal conditions for sterilization. Some investigators recommend treating CJD-contaminated materials once with 1 *N* NaOH at room temperature, but we believe this procedure may be inadequate for sterilization. Autoclaving at 134°C for 5 h or treatment with 2 *N* NaOH for several hours is recommended for sterilization of prions. The term *sterilization* implies complete destruction of prions; any residual infectivity can be hazardous. Transgenic mouse studies show that sCJD prions bound to stainless-steel surfaces are resistant to inactivation by autoclaving at 134°C for 2 h; exposure of bound prions to an acidic detergent solution prior to autoclaving rendered prions susceptible to inactivation. Recent studies show that α-synuclein prions in brain homogenates prepared from MSA patients bind to stainless-steel wires and that the bound prions can be transmitted to transgenic mice expressing mutant human α-synuclein.

Prion precaution protocols should be used for any patient with known or suspected CJD who is undergoing a surgical procedure that has a high risk of exposure to prions (e.g., neurosurgery). In such protocols, procedures should be implemented to reduce exposure of operating room staff and to isolate surgical equipment until the diagnosis has been definitively determined. If the patient is known to have prion disease, the equipment should be destroyed if possible or, if not possible, then thoroughly cleansed to eliminate risk of prion exposure. Importantly, as human prions appear to be more resistant than many animal prions to denaturation, particularly when bound to metal, prion removal methods used in the clinical setting should be based on data from studies using human prions.

PREVENTION AND THERAPEUTICS

There is no known effective therapy for preventing or treating CJD. The finding that phenothiazines and acridines inhibit PrP^{Sc} formation in cultured cells led to clinical studies of quinacrine in CJD patients. Unfortunately, quinacrine failed to slow the rate of cognitive decline in CJD, possibly because therapeutic concentrations of quinacrine were not achieved in the brain. Although inhibition of the P-glycoprotein (Pgp) transport system resulted in substantially increased quinacrine levels in the brains of mice, the prion incubation times were not extended by treatment with the drug. Whether such an approach can be used to treat CJD remains to be established.

Like the acridines, anti-PrP antibodies have been shown to eliminate PrP^{Sc} from cultured cells. Additionally, such antibodies in mice, either administered by injection or produced from a transgene, have been shown to prevent prion disease when prions are introduced by a peripheral route, such as intraperitoneal inoculation. Unfortunately, the antibodies were ineffective in mice inoculated intracerebrally with prions. Several drugs, including pentosan polysulfate as well as porphyrin and phenylhydrazine derivatives, delay the onset of disease in animals inoculated intracerebrally with prions if the drugs are given intracerebrally beginning soon after inoculation.

New treatment trials are underway and, based on animal models, hold promise even when treatment is begun close to symptom onset.

The best method to treat human prion diseases may be by reducing PrP, the substrate for PrP^{Sc}, rather than trying to directly reduce or eliminate PrP^{Sc}. The lifespan of mice is normal when PrP^{C} is reduced or even eliminated; when *PRNP* is knocked out, mice live a normal span of time with minimal deficits detected, aside from a mild neuropathy. Identification of adults hemizygous for *PRNP* suggests that humans can also live with reduced levels of PrP^{C}. Currently, antisense oligonucleotides (ASOs) against PrP^{C} are being tested in symptomatic prion disease. One advantage of treatment methods reducing PrP^{C} is that if they are effective, they should work for all strains and types of human prion disease, as the method is independent of the form of PrP^{Sc} that is pathogenic in a given individual.

■ FURTHER READING

Aoyagi A et al: Aβ and tau prion-like activities decline with longevity in the Alzheimer's disease human brain. Sci Transl Med 11:eaat8462, 2019.

Bizzi A et al: Subtype diagnosis of sporadic Creutzfeldt-Jakob disease with diffusion magnetic resonance imaging. Ann Neurol 89:560, 2021.

Collinge J: Mammalian prions and their wider relevance in neurodegenerative diseases. Nature 539:217, 2016.

Geschwind MD: Prion diseases. Continuum (Minneap Minn) 21:1612, 2015.

Hermann P et al: Biomarkers and diagnostic guidelines for sporadic Creutzfeldt-Jakob disease. Lancet Neurol 20:235, 2021.

Jucker M, Walker LC: Evidence for iatrogenic transmission of Alzheimer's disease. Nat Med 30:344, 2024.

Kraus A et al: High-resolution structure and strain comparison of infectious mammalian prions. Mol Cell 81:4540, 2021.

Prusiner SB (ed): *Prion Biology*. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press, 2017.

Prusiner SB (ed): *Prion Diseases*. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press, 2017.

Prusiner SB et al: Evidence for α-synuclein prions causing multiple system atrophy in humans with parkinsonism. Proc Natl Acad Sci USA 112:E5308, 2015.

Staffaroni AM et al: Neuroimaging in dementia. Semin Neurol 37:510, 2017.

450

Ataxic Disorders

Roger N. Rosenberg, Vikram G. Shakkottai

APPROACH TO THE PATIENT

Ataxic Disorders

Symptoms and signs of ataxia consist of gait impairment, unclear ("scanning") speech, visual blurring due to nystagmus, hand incoordination, and tremor with movement. These result from the involvement of the cerebellum and its afferent and efferent pathways, including the spinocerebellar pathways, and the frontopontocerebellar pathway originating in the rostral frontal lobe. True cerebellar ataxia must be distinguished from ataxia associated with vestibular nerve or labyrinthine disease, as the latter results in a disorder of gait associated with a significant degree of dizziness, light-headedness, or the perception of movement **(Chap. 24)**. True cerebellar ataxia is devoid of these vertiginous complaints and is clearly an unsteady gait due to imbalance. Sensory disturbances can also on occasion simulate the imbalance of cerebellar disease; with sensory ataxia, imbalance dramatically worsens when visual input is removed (Romberg sign). Rarely, weakness of proximal leg muscles mimics cerebellar disease. In the patient who presents with ataxia, the rate and pattern of the development of cerebellar symptoms help to narrow the diagnostic possibilities **(Table 450-1)**. A gradual and progressive increase in symptoms with bilateral and symmetric involvement suggests a genetic, metabolic, immune, or toxic etiology. Conversely, focal, unilateral symptoms with headache and impaired level of consciousness accompanied by ipsilateral cranial nerve palsies and contralateral weakness imply a space-occupying cerebellar lesion.

SYMMETRIC ATAXIA

Progressive and symmetric ataxia can be classified with respect to onset as acute (over hours or days), subacute (weeks or months), or chronic (months to years). Acute and reversible ataxias include those caused by intoxication with alcohol, phenytoin, lithium, barbiturates, and other drugs. Intoxication caused by toluene exposure, gasoline sniffing, glue sniffing, spray painting, or exposure to methyl mercury or bismuth are additional causes of acute or subacute ataxia, as is treatment with cytotoxic chemotherapeutic drugs such as fluorouracil and paclitaxel. Patients with a postinfectious syndrome (especially after varicella) may develop gait ataxia and mild dysarthria, both of which are reversible **(Chap. 456)**. Rare infectious causes of acquired ataxia include poliovirus, coxsackievirus, echovirus, Epstein-Barr virus, toxoplasmosis, *Legionella*, and Lyme disease.

The subacute development of ataxia of gait over weeks to months (degeneration of the cerebellar vermis) may be due to the combined effects of alcoholism and malnutrition, particularly with deficiencies of vitamins B$_1$ and B$_{12}$. Hyponatremia has also been associated with ataxia. Paraneoplastic cerebellar ataxia is associated with a number of different tumors (and autoantibodies) such as breast and ovarian cancers (anti-Yo), small-cell lung cancer (anti-PQ-type voltage-gated calcium channel), and Hodgkin's disease (anti-Tr) **(Chap. 99)**. Another paraneoplastic syndrome associated with myoclonus and opsoclonus occurs with breast (anti-Ri) and lung cancers and neuroblastoma. For all of these paraneoplastic ataxias, the neurologic syndrome may be the presenting symptom of the cancer. Autoantibody-associated cerebellar syndromes also occur without a cancer association. The most common is a progressive ataxic syndrome affecting speech and gait associated with serum anti-glutamic acid decarboxylase (GAD65) antibodies **(Chap. 99)**. Another immune-mediated progressive ataxia is associated with antigliadin (and antiendomysium) antibodies and the human leukocyte antigen (HLA) DQB1*0201 haplotype; in some affected patients, biopsy of the small intestine reveals villus atrophy consistent with gluten-sensitive enteropathy **(Chap. 336)**. Finally, subacute progressive ataxia may be caused by a prion disorder, especially when an infectious etiology, such as transmission from contaminated human growth hormone, is responsible **(Chap. 449)**.

Chronic symmetric gait ataxia suggests an inherited ataxia (discussed below), a metabolic disorder, or a chronic infection. Hypothyroidism must always be considered as a readily treatable and reversible form of gait ataxia. Infectious diseases that can present with ataxia are meningovascular syphilis and tabes dorsalis due to degeneration of the posterior columns and spinocerebellar pathways in the spinal cord.

FOCAL ATAXIA

Acute focal ataxia commonly results from cerebrovascular disease, usually ischemic infarction or cerebellar hemorrhage. These lesions typically produce cerebellar symptoms ipsilateral to the injured cerebellum and may be associated with an impaired level of consciousness due to brainstem compression and increased intracranial pressure; ipsilateral pontine signs, including sixth and seventh nerve palsies, may be present. Focal and worsening signs of acute ataxia should also prompt consideration of a posterior fossa subdural hematoma, bacterial abscess, or primary or metastatic cerebellar

TABLE 450-1 Etiology of Cerebellar Ataxia

SYMMETRIC AND PROGRESSIVE SIGNS			FOCAL AND IPSILATERAL CEREBELLAR SIGNS		
ACUTE (HOURS TO DAYS)	**SUBACUTE (DAYS TO WEEKS)**	**CHRONIC (MONTHS TO YEARS)**	**ACUTE (HOURS TO DAYS)**	**SUBACUTE (DAYS TO WEEKS)**	**CHRONIC (MONTHS TO YEARS)**
Intoxication: alcohol, lithium, phenytoin, barbiturates (positive history and toxicology screen) Acute viral cerebellitis (CSF supportive of acute viral infection) Postinfection syndrome	Intoxication: mercury, solvents, gasoline, glue Cytotoxic chemotherapeutic drugs Alcoholic-nutritional (vitamin B_1 and B_{12} deficiency) Lyme disease	Paraneoplastic syndrome Antigliadin antibody syndrome Hypothyroidism Inherited diseases Tabes dorsalis (tertiary syphilis) Phenytoin toxicity Amiodarone	Vascular: cerebellar infarction, hemorrhage, or subdural hematoma Infectious: cerebellar abscess (mass lesion on MRI/CT, history in support of lesion)	Neoplastic: cerebellar glioma or metastatic tumor (positive for neoplasm on MRI/CT) Demyelinating: multiple sclerosis (history, CSF, and MRI are consistent) AIDS-related multifocal leukoencephalopathy (positive HIV test and CD4+ cell count for AIDS)	Stable gliosis secondary to vascular lesion or demyelinating plaque (stable lesion on MRI/CT older than several months) Congenital lesion: Chiari or Dandy-Walker malformations (malformation noted on MRI/CT)

Abbreviations: CSF, cerebrospinal fluid; CT, computed tomography; MRI, magnetic resonance imaging.

tumor. Computed tomography (CT) or magnetic resonance imaging (MRI) studies will reveal clinically significant processes of this type. Many of these lesions represent true neurologic emergencies, as sudden herniation, either rostrally through the tentorium or caudal herniation of cerebellar tonsils through the foramen magnum, can occur and is usually devastating. Acute surgical decompression may be required **(Chap. 318)**. Lymphoma or progressive multifocal leukoencephalopathy (PML) in a patient with AIDS may present with an acute or subacute focal cerebellar syndrome. Chronic etiologies of progressive ataxia include multiple sclerosis **(Chap. 455)** and congenital lesions such as a Chiari malformation **(Chap. 453)** or a congenital cyst of the posterior fossa (Dandy-Walker syndrome).

THE INHERITED ATAXIAS

Inherited ataxias may show autosomal dominant, autosomal recessive, or maternal (mitochondrial) modes of inheritance. A genomic classification **(Table 450-2)**[1] has now largely superseded previous ones based on clinical expression alone.

Although the clinical manifestations and neuropathologic findings of cerebellar disease dominate the clinical picture, there may also be characteristic changes in the basal ganglia, brainstem, spinal cord, optic nerves, retina, and peripheral nerves. In large families with dominantly inherited ataxias, many gradations are observed from purely cerebellar manifestations to mixed cerebellar and brainstem disorders, cerebellar and basal ganglia syndromes, and spinal cord or peripheral nerve disease. Rarely, dementia is present as well. The clinical picture may be homogeneous within a family with dominantly inherited ataxia, but sometimes most affected family members show one characteristic syndrome, while one or several members have an entirely different phenotype.

AUTOSOMAL DOMINANT ATAXIAS

The autosomal spinocerebellar ataxias (SCAs) include SCA types 1 through 50, dentatorubropallidoluysian atrophy (DRPLA), and episodic ataxia (EA) types 1 to 7 (Table 450-2). SCA1, SCA2, SCA3 (Machado-Joseph disease [MJD]), SCA6, SCA7, and SCA17 are caused by CAG triplet repeat expansions in different genes. SCA8 is due to an untranslated CTG repeat expansion, SCA12 is linked to an untranslated CAG repeat, and SCA10 is caused by an untranslated pentanucleotide repeat. The clinical phenotypes of these SCAs overlap. The genotype has become the gold standard for diagnosis and classification. CAG encodes glutamine, and these expanded CAG triplet repeat expansions result in expanded polyglutamine proteins, termed *ataxins*, that produce a toxic gain of function with autosomal dominant inheritance. Although the phenotype is variable for any given disease gene, a pattern of neuronal loss with gliosis is produced that is relatively unique for each ataxia. Immunohistochemical and biochemical studies have shown cytoplasmic (SCA2), neuronal (SCA1, MJD, SCA7), and nucleolar (SCA7) accumulation of the specific mutant polyglutamine-containing ataxin proteins. Expanded polyglutamine ataxins with more than ~40 glutamines are potentially toxic to neurons for a variety of reasons including the following: high levels of gene expression for the mutant polyglutamine ataxin in affected neurons; conformational change of the aggregated protein to a β-pleated structure; abnormal transport of the ataxin into the nucleus (SCA1, MJD, SCA7); binding to other polyglutamine proteins, including the TATA-binding transcription protein and the CREB-binding protein, impairing their functions; altering the efficiency of the ubiquitin-proteasome system of protein turnover; and inducing neuronal apoptosis. An earlier age of onset (anticipation) and more aggressive disease in subsequent generations are due to further expansion of the CAG triplet repeat and increased polyglutamine number in the mutant ataxin. The most common disorders are discussed below.

SCA1

SCA1 was previously referred to as *olivopontocerebellar atrophy*, but genomic data have shown that that entity represents several different genotypes with overlapping clinical features.

Symptoms and Signs SCA1 is characterized by the development in early- or middle-adult life of progressive cerebellar ataxia of the trunk and limbs, impairment of equilibrium and gait, slowness of voluntary movements, scanning speech, nystagmoid eye movements, and oscillatory tremor of the head and trunk. Dysarthria, dysphagia, and oculomotor and facial palsies may also occur. Extrapyramidal symptoms include rigidity, an immobile face, and parkinsonian tremor. The reflexes are usually normal, but knee and ankle jerks may be lost, and extensor plantar responses may occur. Dementia may be noted but is usually mild. Impairment of sphincter function is common, with urinary and sometimes fecal incontinence. Cerebellar and brainstem atrophy are evident on MRI **(Fig. 450-1)**.

Marked shrinkage of the ventral half of the pons, disappearance of the olivary eminence on the ventral surface of the medulla, and atrophy of the cerebellum are evident on gross postmortem inspection of the brain. Variable loss of Purkinje cells, reduced numbers of cells in the molecular and granular layer, demyelination of the middle cerebellar peduncle and the cerebellar hemispheres, and severe loss of cells in the pontine nuclei and olives are found on histologic examination. Degenerative changes in the striatum, especially the putamen, and loss of the pigmented cells of the substantia nigra may be found in cases with extrapyramidal features. More widespread degeneration in the central nervous system (CNS), including involvement of the posterior columns and the spinocerebellar fibers, is often present.

GENETIC CONSIDERATIONS

SCA1 encodes a gene product, called *ataxin-1*, that regulates transcriptional repression with various nuclear factors. As a protein that can bind RNA, ataxin-1 may also regulate gene transcription posttranslationally. The mutant allele has 40 CAG repeats located

---

[1] Table 450-2 can be found online at *accessmedicine.com*.

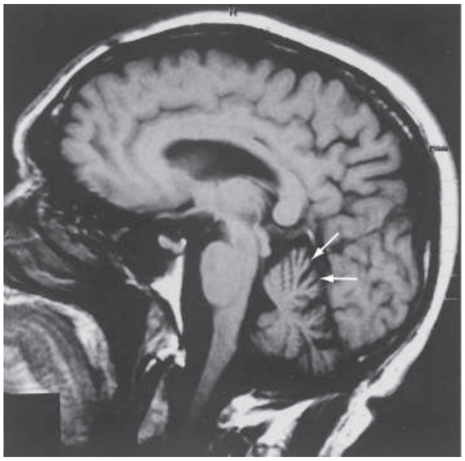

FIGURE 450-1 Sagittal magnetic resonance imaging (MRI) of the brain of a 60-year-old man with gait ataxia and dysarthria due to spinocerebellar ataxia type 1 (SCA1), illustrating cerebellar atrophy (*arrows*). *(Reproduced with permission from RN Rosenberg, P Khemani, in RN Rosenberg, JM Pascual [eds]: Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease, 5th ed. London, Elsevier, 2015.)*

within the coding region, whereas alleles from unaffected individuals have ≤36 repeats. A few patients with 38–40 CAG repeats have been described. There is a direct correlation between a larger number of repeats and a younger age of onset for SCA1. Juvenile patients have higher numbers of repeats, and anticipation is present in subsequent generations. Transgenic mice carrying SCA1 developed ataxia and Purkinje cell pathology. Leucine-rich acidic nuclear protein localization, but not aggregation, of ataxin-1 appears to be required for cell death initiated by the mutant protein.

■ SCA2

Symptoms and Signs Another clinical phenotype, SCA2, has been described in patients from Cuba and India. Cuban patients probably are descendants of a common ancestor, and the population may be the largest homogeneous group of patients with ataxia described. The age of onset ranges from 2 to 65 years, and there is considerable clinical variability within families. Although neuropathologic and clinical findings are compatible with a diagnosis of SCA1, including slow saccadic eye movements, ataxia, dysarthria, parkinsonian rigidity, optic disc pallor, mild spasticity, and retinal degeneration, SCA2 is a unique form of cerebellar degenerative disease.

■ GENETIC CONSIDERATIONS

The gene in SCA2 families also contains CAG repeat expansions coding for a polyglutamine-containing protein, ataxin-2. Normal alleles contain 15–32 repeats; mutant alleles have 35–77 repeats. Ataxin-2 has recently been shown to assemble with polyribosomes. Ataxin-2 is also an important risk factor for sporadic amyotrophic lateral sclerosis (ALS).

■ MACHADO-JOSEPH DISEASE/SCA3

MJD was first described among the Portuguese and their descendants in New England and California. Subsequently, MJD has been found in families from Portugal, Australia, Brazil, Canada, China, England, France, India, Israel, Italy, Japan, Spain, Taiwan, and the United States. In most populations, it is the most common autosomal dominant ataxia.

Symptoms and Signs MJD has been classified into three clinical types. In type I MJD (ALS-parkinsonism-dystonia type), neurologic deficits appear in the first two decades and involve weakness and spasticity of extremities, especially the legs, often with dystonia of the face, neck, trunk, and extremities. Patellar and ankle clonus are common, as are extensor plantar responses. The gait is slow and stiff, with a slightly broadened base and lurching from side to side; this gait results from spasticity, not true ataxia. There is no truncal titubation. Pharyngeal weakness and spasticity cause difficulty with speech and swallowing. Of note is the prominence of horizontal and vertical nystagmus, loss of fast saccadic eye movements, hypermetric and hypometric saccades, and impairment of upward vertical gaze. Facial fasciculations, facial myokymia, lingual fasciculations without atrophy, ophthalmoparesis, and ocular prominence are common early manifestations.

In type II MJD (ataxic type), true cerebellar deficits of dysarthria and gait and extremity ataxia begin in the second to fourth decades along with corticospinal and extrapyramidal deficits of spasticity, rigidity, and dystonia. Type II is the most common form of MJD. Ophthalmoparesis, upward vertical gaze deficits, and facial and lingual fasciculations are also present. Type II MJD can be distinguished from the clinically similar disorders SCA1 and SCA2.

Type III MJD (ataxic-amyotrophic type) presents in the fifth to seventh decades with a pancerebellar disorder that includes dysarthria and gait and extremity ataxia. Distal sensory loss involving pain, touch, vibration, and position senses and distal atrophy are prominent, indicating the presence of peripheral neuropathy. The deep tendon reflexes are depressed to absent, and there are no corticospinal or extrapyramidal findings.

The mean age of onset of symptoms in MJD is 25 years. Neurologic deficits invariably progress and lead to death from debilitation within 15 years of onset, especially in patients with types I and II disease. Usually, patients retain full intellectual function.

The major pathologic findings are variable loss of neurons and glial replacement in the corpus striatum and severe loss of neurons in the pars compacta of the substantia nigra. A moderate loss of neurons occurs in the dentate nucleus of the cerebellum and in the red nucleus. Purkinje cell loss and granule cell loss occur in the cerebellar cortex. Cell loss also occurs in the dentate nucleus and in the cranial nerve motor nuclei. Sparing of the inferior olives distinguishes MJD from other dominantly inherited ataxias.

■ GENETIC CONSIDERATIONS

The gene for MJD maps to 14q24.3-q32. Unstable CAG repeat expansions are present in the MJD gene coding for a polyglutamine-containing protein named ataxin-3, or MJD-ataxin. An earlier age of onset is associated with longer repeats. Alleles from normal individuals have between 12 and 37 CAG repeats, whereas MJD alleles have 60–84 CAG repeats. Polyglutamine-containing aggregates of ataxin-3 (MJD-ataxin) have been described in neuronal nuclei undergoing degeneration. MJD-ataxin codes for a ubiquitin protease, which is inactive due to expanded polyglutamines. Proteosome function is impaired, resulting in altered clearance of proteins and cerebellar neuronal loss.

■ SCA6

Genomic screening for CAG repeats in other families with autosomal dominant ataxia and vibratory and proprioceptive sensory loss have yielded another locus. Of interest is that different mutations in the same gene for the α_{1A} voltage-dependent calcium channel subunit (CACNLIA4; also referred to as the *CACNA1A* gene) at 19p13 result in different clinical disorders. CAG repeat expansions (21–27 in patients; 4–16 triplets in normal individuals) result in late-onset progressive ataxia with cerebellar degeneration. Missense mutations in this gene result in familial hemiplegic migraine. Nonsense mutations resulting in termination of protein synthesis of the gene product yield hereditary paroxysmal cerebellar ataxia or EA. Some patients with familial hemiplegic migraine develop progressive ataxia and also have cerebellar atrophy.

■ SCA7

This disorder is distinguished from all other SCAs by the presence of retinal pigmentary degeneration. The visual abnormalities first appear as blue-yellow color blindness and proceed to frank visual loss with macular degeneration. In almost all other respects, SCA7 resembles several other SCAs in which ataxia is accompanied by various noncerebellar findings, including ophthalmoparesis and extensor plantar responses. The genetic defect is an expanded CAG repeat in the SCA7 gene at 3p14-p21.1. The expanded repeat size in SCA7 is highly variable. Consistent with this, the severity of clinical findings varies from essentially asymptomatic to mild late-onset symptoms to severe, aggressive disease in childhood with rapid progression. Marked anticipation has been recorded, especially with paternal transmission. The disease protein, ataxin-7, forms aggregates in nuclei of affected neurons, as has also been described for SCA1 and SCA3/MJD. Ataxin-7 is a subunit of GCN5, a histone acetyltransferase-containing complex.

SCA8

This form of ataxia is caused by a CTG repeat expansion in an untranslated region of a gene on chromosome 13q21. There is marked maternal bias in transmission, perhaps reflecting contractions of the repeat during spermatogenesis. The mutation is not fully penetrant. Symptoms include slowly progressive dysarthria and gait ataxia beginning at ~40 years of age with a range between 20 and 65 years. Other features include nystagmus, leg spasticity, and reduced vibratory sensation. Severely affected individuals are nonambulatory by the fourth to sixth decades. MRI shows cerebellar atrophy. The mechanism of disease may involve a dominant "toxic" effect occurring at the RNA level, as occurs in myotonic dystrophy.

SCA27B

SCA27B is a recently discovered entity, resulting from an intronic GAA repeat expansion in the *FGF14* gene, and is one of the most common late-onset inherited ataxias. SCA27B occurs with a median age of onset of 60 years and presents as a relatively pure cerebellar ataxia with episodic symptoms at the disease onset. Other clinical features include afferent sensory deficits and dysautonomia. Cognitive impairment is infrequent. Both the episodic symptoms and symptom severity of ataxia appear to improve with 4-aminopyridine, although randomized clinical trials have not yet been performed in this population.

DENTATORUBROPALLIDOLUYSIAN ATROPHY

DRPLA has a variable presentation that may include progressive ataxia, choreoathetosis, dystonia, seizures, myoclonus, and dementia. DRPLA is due to unstable CAG triplet repeats in the open reading frame of a gene named *atrophin* located on chromosome 12p12-ter. Larger expansions are found in patients with earlier onset. The number of repeats is 49 in patients with DRPLA and ≤26 in normal individuals. Anticipation occurs in successive generations, with earlier onset of disease in association with an increasing CAG repeat number in children who inherit the disease from their father. One well-characterized family in North Carolina has a phenotypic variant known as the *Haw River syndrome*, now recognized to be due to the DRPLA mutation.

EPISODIC ATAXIA

EA types 1 and 2 are two rare dominantly inherited disorders that have been mapped to chromosomes 12p (a potassium channel gene, *KCNA1*, Phe249Leu mutation) for type 1 and 19p for type 2. Patients with EA-1 have brief episodes of ataxia with myokymia and nystagmus that last only minutes. Startle, sudden change in posture, and exercise can induce episodes. Acetazolamide or anticonvulsants may be therapeutic. Patients with EA-2 have episodes of ataxia with nystagmus that can last for hours or days. Stress, exercise, or excessive fatigue may be precipitants. Acetazolamide may be therapeutic and can reverse the relative intracellular alkalosis detected by magnetic resonance spectroscopy. Stop codon, nonsense mutations causing EA-2 have been found in the *CACNA1A* gene, encoding the α_{1A} voltage-dependent calcium channel subunit (see "SCA6," above).

AUTOSOMAL RECESSIVE ATAXIAS

Friedreich's Ataxia This is the most common form of inherited ataxia, composing one-half of all hereditary ataxias. It can occur in a classic form or in association with a genetically determined vitamin E deficiency syndrome; the two forms are clinically indistinguishable.

SYMPTOMS AND SIGNS Friedreich's ataxia presents before 25 years of age with progressive staggering gait, frequent falling, and titubation. The lower extremities are more severely involved than the upper ones. Dysarthria occasionally is the presenting symptom; rarely, progressive scoliosis, foot deformity, nystagmus, or cardiopathy is the initial sign.

The neurologic examination reveals nystagmus, loss of fast saccadic eye movements, truncal titubation, dysarthria, dysmetria, and ataxia of trunk and limb movements. Extensor plantar responses (with normal tone in trunk and extremities), absence of deep tendon reflexes, and weakness (greater distally than proximally) are usually found. Loss of vibratory and proprioceptive sensation occurs. The median age of death is 35 years. Women have a significantly better prognosis than men.

Cardiac involvement occurs in 90% of patients. Cardiomegaly, symmetric hypertrophy, murmurs, and conduction defects are reported. Moderate intellectual disability or psychiatric syndromes are present in a small percentage of patients. A high incidence (20%) of diabetes mellitus is found and is associated with insulin resistance and pancreatic β-cell dysfunction. Musculoskeletal deformities are common and include pes cavus, pes equinovarus, and scoliosis. MRI of the spinal cord shows atrophy **(Fig. 450-2)**.

The primary sites of pathology are the spinal cord, dorsal root ganglion cells, and the peripheral nerves. Slight atrophy of the cerebellum and cerebral gyri may occur. Sclerosis and degeneration occur predominantly in the spinocerebellar tracts, lateral corticospinal tracts, and posterior columns. Degeneration of the glossopharyngeal, vagus, hypoglossal, and deep cerebellar nuclei is described. The cerebral cortex is histologically normal except for loss of Betz cells in the precentral gyri. The peripheral nerves are extensively involved, with a loss of large myelinated fibers. Cardiac pathology consists of myocytic hypertrophy and fibrosis, focal vascular fibromuscular dysplasia with subintimal or medial deposition of periodic acid-Schiff (PAS)-positive material, myocytopathy with unusual pleomorphic nuclei, and focal degeneration of nerves and cardiac ganglia.

GENETIC CONSIDERATIONS

The classic form of Friedreich's ataxia has been mapped to 9q13-q21.1, and the mutant gene, *frataxin*, contains expanded GAA triplet repeats in the first intron. There is homozygosity for expanded GAA repeats in >95% of patients. Normal persons have 7–22 GAA repeats, and patients have 200–900 GAA repeats. A more varied clinical syndrome has been described in compound heterozygotes who have one copy of the GAA expansion and the other copy a point mutation in the *frataxin* gene. When the point mutation is located in the region of the gene that encodes the amino-terminal half of frataxin, the phenotype is milder, often consisting of a spastic gait, retained or exaggerated reflexes, no dysarthria, and mild or absent ataxia.

Patients with Friedreich's ataxia have undetectable or extremely low levels of *frataxin* mRNA, as compared with carriers and unrelated

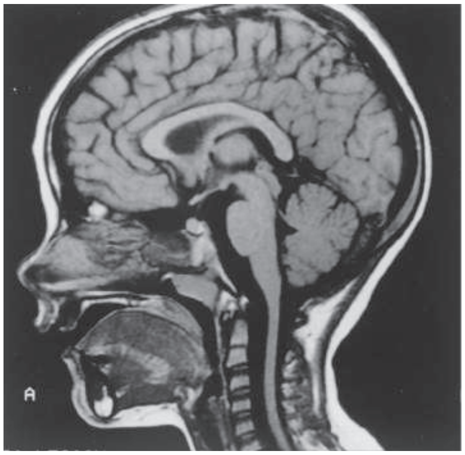

FIGURE 450-2 Sagittal magnetic resonance imaging (MRI) of the brain and spinal cord of a patient with Friedreich's ataxia, demonstrating spinal cord atrophy. *(Reproduced with permission from RN Rosenberg, P Khemani, in RN Rosenberg, JM Pascual [eds]: Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease, 5th ed. London, Elsevier, 2015.)*

individuals; thus, disease appears to be caused by a loss of expression of the frataxin protein. Frataxin is a mitochondrial protein involved in iron homeostasis. Mitochondrial iron accumulation due to loss of the iron transporter coded by the mutant *frataxin* gene results in a deficiency in iron/sulfur clusters containing mitochondrial enzymes, decreased ATP production, and accumulation of iron in the heart. Excess oxidized iron results in turn in the oxidation of cellular components and irreversible cell injury.

Two forms of hereditary ataxia associated with abnormalities in the interactions of vitamin E (α-tocopherol) with very-low-density lipoprotein (VLDL) have been delineated. These are abetalipoproteinemia (Bassen-Kornzweig syndrome) and ataxia with vitamin E deficiency (AVED). Abetalipoproteinemia is caused by mutations in the gene coding for the larger subunit of the microsomal triglyceride transfer protein (MTP). Defects in MTP result in impairment of formation and secretion of VLDL in liver. This defect results in a deficiency of delivery of vitamin E to tissues, including the central and peripheral nervous system, as VLDL is the transport molecule for vitamin E and other fat-soluble substitutes. AVED is due to mutations in the gene for α-tocopherol transfer protein (α-TTP). These patients have an impaired ability to bind vitamin E into the VLDL produced and secreted by the liver, resulting in a deficiency of vitamin E in peripheral tissues. Hence, either absence of VLDL (abetalipoproteinemia) or impaired binding of vitamin E to VLDL (AVED) causes an ataxic syndrome. Once again, a genotype classification has proved to be essential in sorting out the various forms of the Friedreich's disease syndrome, which may be clinically indistinguishable.

RFC1-Related CANVAS Syndrome Biallelic intronic AAGGG repeat expansions in the replication factor C subunit 1 (*RFC1*) gene are the cause of late-onset ataxia, particularly if associated with sensory neuronopathy and bilateral vestibular areflexia (CANVAS syndrome). A chronic unexplained cough is often associated with and may precede the onset of neurologic symptoms.

Ataxia Telangiectasia • SYMPTOMS AND SIGNS Patients with ataxia telangiectasia (AT) present in the first decade of life with progressive telangiectatic lesions associated with deficits in cerebellar function and nystagmus. The neurologic manifestations correspond to those in Friedreich's disease, which should be included in the differential diagnosis. Truncal and limb ataxia, dysarthria, extensor plantar responses, myoclonic jerks, areflexia, and distal sensory deficits may develop. There is a high incidence of recurrent pulmonary infections and neoplasms of the lymphatic and reticuloendothelial system in patients with AT. Thymic hypoplasia with cellular and humoral (IgA and IgG2) immunodeficiencies, premature aging, and endocrine disorders such as type 1 diabetes mellitus are described. There is an increased incidence of lymphomas, Hodgkin's disease, acute T-cell leukemias, and breast cancer.

The most striking neuropathologic changes include loss of Purkinje, granule, and basket cells in the cerebellar cortex as well as of neurons in the deep cerebellar nuclei. The inferior olives of the medulla may also have neuronal loss. There is a loss of anterior horn neurons in the spinal cord and of dorsal root ganglion cells associated with posterior column spinal cord demyelination. A poorly developed or absent thymus gland is the most consistent defect of the lymphoid system.

GENETIC CONSIDERATIONS

The gene for AT (the *ATM* gene) at 11q22-23 encodes a protein that is similar to several yeast and mammalian phosphatidylinositol-3′ kinases involved in mitogenic signal transduction, meiotic recombination, and cell cycle control. Defective DNA repair in AT fibroblasts exposed to ultraviolet light has been demonstrated. The discovery of *ATM* permits early diagnosis and identification of heterozygotes who are at risk for cancer (e.g., breast cancer). Elevated serum alphafetoprotein and immunoglobulin deficiency are noted.

MITOCHONDRIAL ATAXIAS

Spinocerebellar syndromes have been identified with mutations in mitochondrial DNA (mtDNA). Thirty pathogenic mtDNA point mutations and 60 different types of mtDNA deletions are known, several of which cause or are associated with ataxia **(Chap. 460)**.

TREATMENT

Ataxic Disorders

The most important goal in management of patients with ataxia is to identify treatable disease entities. Mass lesions must be recognized promptly and treated appropriately. Autoimmune paraneoplastic disorders can often be identified by the clinical patterns of disease that they produce, measurement of specific autoantibodies, and uncovering the primary cancer; these disorders are often refractory to therapy, but some patients improve following removal of the tumor or immunotherapy **(Chap. 99)**. Ataxia with antigliadin antibodies and gluten-sensitive enteropathy may improve with a gluten-free diet. Malabsorption syndromes leading to vitamin E deficiency may lead to ataxia. The vitamin E deficiency form of Friedreich's ataxia must be considered, and serum vitamin E levels measured. Vitamin E therapy is indicated for these rare patients. Vitamin B_1 and B_{12} levels in serum should be measured, and the vitamins administered to patients having deficient levels. Hypothyroidism is easily treated. The cerebrospinal fluid should be tested for a syphilitic infection in patients with progressive ataxia and other features of tabes dorsalis. Similarly, antibody titers for Lyme disease and *Legionella* should be measured and appropriate antibiotic therapy should be instituted in antibody-positive patients. Aminoacidopathies, leukodystrophies, urea-cycle abnormalities, and mitochondrial encephalomyopathies may produce ataxia, and some dietary or metabolic therapies are available for these disorders. The deleterious effects of phenytoin and alcohol on the cerebellum are well known, and these exposures should be avoided in patients with ataxia of any cause.

There is no proven therapy for any of the autosomal dominant ataxias (SCA1 to SCA43). Omaveloxolone, a NRF2 agonist, is the only U.S. Food and Drug Administration–approved agent for Friedreich's ataxia. NRF2 is a transcription factor that regulates gene transcripts involved in mitochondrial energy production and addresses the root cause of mitochondrial dysfunction in Friedreich's ataxia. Iron chelators and antioxidant drugs are potentially harmful in Friedreich's patients because they may increase heart muscle injury. Acetazolamide can reduce the duration of symptoms of EA. At present, identification of an at-risk person's genotype, together with appropriate family and genetic counseling, can reduce the incidence of these cerebellar syndromes in future generations **(Chap. 480)**.

GENETIC DIAGNOSTIC LABORATORIES

1. Baylor College of Medicine; Houston, Texas, 1-713-798-6522
   http://www.bcm.edu/genetics/index.cfm?pmid=21387
2. The University of Chicago Genetic Services Laboratories
   https://dnatesting.uchicago.edu
3. GeneDx
   http://www.genedx.com
4. Transgenomic, 1-877-274-9432
   http://www.transgenomic.com/labs/neurology

GLOBAL FEATURES

Ataxias with autosomal dominant, autosomal recessive, X-linked, or mitochondrial forms of inheritance are present on a worldwide basis. MJD (SCA3) (autosomal dominant) and Friedreich's ataxia (autosomal recessive) are the most common types in most populations. Genetic markers are now commercially available to precisely identify the genetic mutation for correct diagnosis and also for family planning. Early detection of asymptomatic preclinical disease can reduce or eliminate the inherited form of ataxia in some families on a global, worldwide basis.

FURTHER READING

Anheim M et al: The autosomal recessive cerebellar ataxias. N Engl J Med 16:636, 2012.

Cortese A et al: Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. Nat Genet 51:649, 2019.

Jacobi H et al: Long-term disease progression in spinocerebellar ataxia types 1, 2, 3, and 6: A longitudinal cohort study. Lancet Neurol 14:1101, 2015.

Lynch DR et al: Omaveloxolone for the treatment of Friedreich ataxia: Clinical trial results and practical considerations. Expert Rev Neurother 24:251, 2024.

Paulson HL et al: Polyglutamine spinocerebellar ataxias—from genes to potential therapy. Nat Rev Neurosci 18:613; 2017.

Pellerin D et al: Deep intronic *FGF14* GAA repeat expansion in late-onset cerebellar ataxia. N Engl J Med 388:128, 2023.

451

Disorders of the Autonomic Nervous System

Safwan S. Jaradeh

The autonomic nervous system (ANS) has a large central nervous system (CNS) organization and an extensive peripheral nervous system (PNS) network that innervates and influences all organ systems. As such, it is important in securing normal homeostasis. The ANS has three major components: parasympathetic, sympathetic, and enteric. The parasympathetic component has mainly a craniosacral localization (brainstem and sacral cord), while the sympathetic component has mainly a thoracolumbar localization (first thoracic to second lumbar segments). ANS neurons of these components project thinly myelinated fibers to autonomic ganglia, where they synapse with postganglionic, mainly unmyelinated fibers that innervate the respective target organs. The enteric nervous system (ENS) has its own set of autonomic neurons and neurotransmitters but is heavily regulated by the central and peripheral autonomic nervous systems. The ANS also sensitizes endocrine organs to the effect of hormones, influencing how hormonal balances may affect neurologic diseases.

The classification of autonomic dysfunction (or dysautonomia) can be based on the impaired component (sympathetic, parasympathetic, or enteric), but a more practical clinical approach is based on organ specificity: heart, blood vessels, sweat glands, pupils, gastrointestinal, and genitourinary, either in isolation or in any potential combination. The time course of the development of dysautonomia (acute, subacute, or chronic; progressive or static) is especially useful in the differential etiologic diagnosis.

The ANS functions as a reflex circuit, using the sensory and sensorial feedback from the organs to modulate the ANS output and maintain homeostasis. This requires exquisite integration of the central ANS, peripheral ANS, and target organs. In the CNS, autonomic circuits operate at several levels: cerebral supratentorial, cerebral infratentorial, and spinal. Supratentorial areas include the insular cortex, prefrontal cortex, limbic lobe located on the medial surface of the cerebral hemispheres, hypothalamus, and hypothalamic-pituitary axis. Infratentorial areas are mainly the nucleus tractus solitaries (NTS), dorsal motor nucleus of the vagus (DMV) and nucleus ambiguous (NA). Spinal areas are in the intermediate lateral horns (IML) of the spinal cord gray matter between the first thoracic segment and the fourth sacral segment. Descending CNS information to the brainstem and IML allows the CNS to modulate autonomic activity. Peripheral, visceral, spinal, and brainstem afferent information conveyed to the CNS is also important in influencing autonomic function in health and disease. ANS dysfunction may explain many symptoms related to the aging process and various degenerative neurologic disorders, such as Parkinson's disease, multiple system atrophy, Lewy body disease, and pure autonomic failure. It features prominently in various autoimmune disorders, particularly autoimmune autonomic ganglionopathy and certain paraneoplastic disorders. ANS dysfunction in a more restricted form is present in various cyclical disorders such as migraine, cyclic vomiting, and other circadian rhythm disorders. It plays an important role in the genesis of various conditions of orthostatic intolerance (OI) including orthostatic hypotension (OH), postural orthostatic tachycardia syndrome (POTS), and syncope. **A detailed discussion of syncope can be found in Chap. 23.**

ANATOMIC, BIOCHEMICAL, AND PHARMACOLOGIC ORGANIZATION

Connections between ANS centers in the cerebral cortex and brainstem coordinate autonomic outflow and their integration with emotional and higher mental functions.

The preganglionic parasympathetic neurons leave the CNS in the third, seventh, ninth, and tenth cranial nerves as well as the second to fourth sacral nerves. The preganglionic sympathetic neurons exit the spinal cord between the first thoracic and the second lumbar segments **(Fig. 451-1)**. The autonomic ganglia reside outside the CNS. The sympathetic ganglia are located in paravertebral and prevertebral ganglia, while the parasympathetic ganglia are located near, or sometimes within, the target organs. Responses to sympathetic and parasympathetic stimulation are frequently antagonistic **(Table 451-1)**, allowing fine-tuning of autonomic responses. In general, increased sympathetic activity leads to the "fight or flight" response, while parasympathetic activity leads to the "rest and digest" response.

Biochemically, acetylcholine (ACh) is the preganglionic neurotransmitter for both sympathetic and parasympathetic divisions of the ANS as well as the postganglionic neurotransmitter of the parasympathetic neurons; all preganglionic receptors are nicotinic, while postganglionic receptors are muscarinic in type. Norepinephrine (NE) is the neurotransmitter of postganglionic sympathetic neurons, except for sympathetic cholinergic neurons innervating the eccrine sweat glands where the neurotransmitter is also ACh.

As stated above, the ENS has a large number of neurons organized in plexuses. Meissner's (submucosal) plexus, Auerbach's (myenteric) and interstitial cells of Cajal (circular muscular) compose the majority of the ENS; they contain numerous neurotransmitters. Parasympathetic control of the gastrointestinal system is mediated through the craniospinal nerves (vagus and S2–S4 nerves), while sympathetic control is mediated through the thoracolumbar (T1–L2) nerves.

CLINICAL EVALUATION

Disorders of the ANS are rather common and can occur with either CNS or PNS pathology that disrupts the afferent limb, CNS processing center, efferent limb, or any combination of these. It is also important to recognize that clinical manifestations can result from decreased function, overactivity, or dysregulation of autonomic circuits.

The evaluation begins by taking an accurate and complete autonomic medical history. The chief complaint and history of present illness should be sought from the patient, assisted by family members or significant others when necessary. Records from previous hospitalizations and test results are useful but cannot substitute for direct history taking. It is important to obtain the sequence of symptoms as accurately as possible. Antecedent events such as a viral illness or surgical procedure followed by subacute OI suggests an autoimmune cause, while a history of frequent fainting since childhood or adolescence favors a congenital or genetic condition. Therefore, a detailed family history is often essential, particularly in younger patients. In a patient with POTS, which occurs mainly in young women, it is important to ask privately about emotional or physical abuse in childhood. Physical deconditioning often exacerbates chronic OI and POTS. A prior history of irradiation to the neck in a patient with labile blood pressure and OI suggests arterial baroreflex failure due to injury to the carotid sinus baroreceptors.

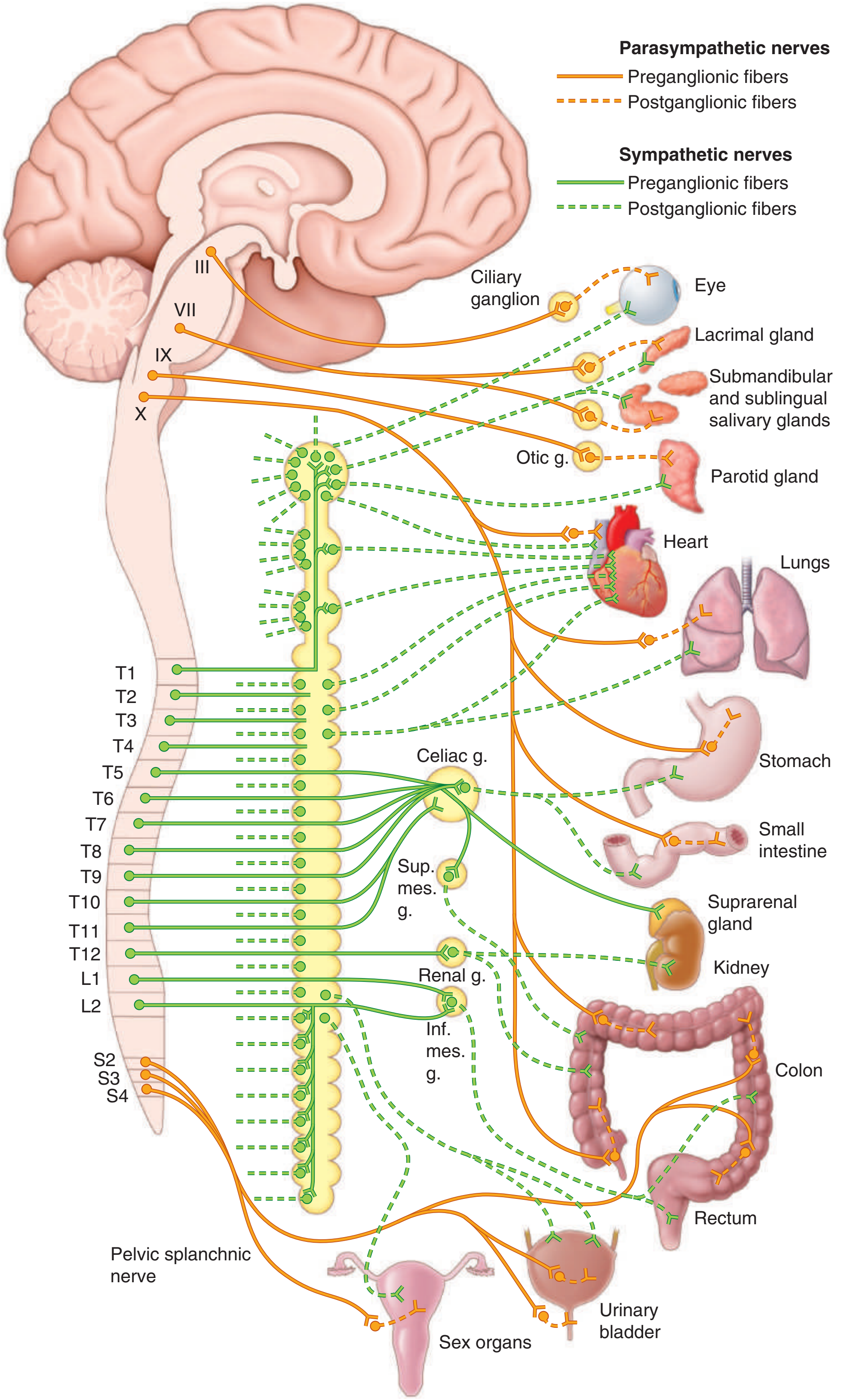
Parasympathetic nerves
Preganglionic fibers
Postganglionic fibers
Sympathetic nerves
Preganglionic fibers
Postganglionic fibers
III
VII
IX
X
Ciliary ganglion
Eye
Lacrimal gland
Submandibular and sublingual salivary glands
Otic g.
Parotid gland
Heart
Lungs
T1
T2
T3
T4
T5
T6
T7
T8
T9
T10
T11
T12
L1
L2
Celiac g.
Stomach
Small intestine
Sup. mes. g.
Suprarenal gland
Kidney
Renal g.
Inf. mes. g.
Colon
S2
S3
S4
Rectum
Pelvic splanchnic nerve
Urinary bladder
Sex organs

FIGURE 451-1 Schematic representation of the autonomic nervous system. *(Reproduced with permission from R Snell: Clinical Neuroanatomy, 7th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2009.)*

Like other medical histories, it is important to determine exacerbating and alleviating factors. In patients with OI, the positional nature of symptoms (standing vs supine or sitting) can be a key clue to diagnosis. Patients with neurogenic OH are often more symptomatic in the morning, after exercise particularly in the heat, or after eating a large meal, especially with alcohol. Patients with POTS may feel worse when standing in a warm shower and notice reddish or purplish feet in that situation. However, given the widespread autonomic network underlying many dysautonomic states, these conditions are often multisystemic, and specific symptoms can vary from day to day.

A review of autonomic symptoms begins with questions about OI (lightheadedness, dizziness, orthostatic weakness or dyspnea, orthostatic neck and shoulder discomfort, orthostatic headache), a history of supine hypertension, unpredictable blood pressure swings,

TABLE 451-1 Effects of Sympathetic and Parasympathetic Systems on Various Effector Organs

	SYMPATHETIC	PARASYMPATHETIC
Pupil	Pupillodilation (alpha)	Pupilloconstriction
Accommodation	Decreased	Increased
Heart	Positive chronotropic effect (beta) Positive inotropic effect (beta)	Negative chronotropic effect Negative inotropic effect
Arteries	Vasoconstriction (alpha) Vasodilation (beta)	Vasodilation
Veins	Vasoconstriction (alpha) Vasoconstriction (beta)	
Tracheobronchial tree	Bronchodilation (beta)	Bronchoconstriction Increased bronchial gland secretions
Gastrointestinal tract	Decreased motility (beta) Contraction of sphincters (alpha)	Increased motility Relaxation of sphincter
Bladder	Detrusor relaxation (beta) Contraction of sphincter (alpha)	Detrusor contraction Relaxation of sphincter
Salivary glands	Scant, thick, viscid saliva (alpha)	Copious, thin, watery saliva
Skin	Piloerection (cutis anserina)	No piloerection
Sweat glands	Increased secretion (cholinergic)	Decreased secretion
Genitalia	Ejaculation	Ejaculation/Erection
Adrenal medulla	Catecholamine release	
Glycogen	Glycogenolysis (alpha and beta) Lipolysis (alpha and beta)	Glycogen synthesis

Source: Reproduced with permission from WW Campbell: The autonomic nervous system, in *DeJong's The Neurologic Examination*, 8th ed. Wolters Kluwer, 2020.

or syncopal episodes. Gastrointestinal symptoms may resemble those of irritable bowel syndrome (constipation, diarrhea, or both), but also constant nausea and bloating, unexplained vomiting, or dumping syndrome may be present, both in diabetics and in others. Genitourinary symptoms include dysuria, incontinence, urgency, and male erectile dysfunction. Sweating abnormalities (dry skin, excessive sweating) and body temperature changes reflect disordered thermoregulation and heat dissipation. Sicca syndrome may resemble that of Sjögren syndrome. Blurred vision from dark to light suggests slow, tonic pupillary dilation.

Pertinent nonautonomic symptoms such as acroparesthesias or burning feet (and hands) suggest an autonomic small-fiber neuropathy. Impaired olfaction, impaired balance, abnormal movements, and abnormal sleep behavior (apnea, stridor, or excessive moving and talking during sleep known as dream enactment behavior) are red flags for a degenerative neurologic extrapyramidal disorder.

A complete listing of all prescribed drugs, over-the-counter medications, herbal remedies, and dietary supplements is important because they can affect autonomic function and may produce unexpected adverse events. For instance, certain diet supplements have sympathomimetic activity and cause unexpected hypertension, tachycardia, or both.

Physical examination begins by obtaining full vital signs, including supine blood pressure and heart rate, followed by orthostatic blood pressure values at 1 and 3 min after rising from a seated position. An abnormal body temperature, either hyperthermia or hypothermia, may be a clue to a systemic or endocrine illness. Next comes the inspection of the limbs, looking for changes in the feet (pes cavus or planus), vasomotor color changes in the distal limbs, neuropathic foot or toe ulcers, scoliosis, chest deformity (pectus excavatum or carinii), hypermobility, skin elasticity, arachnodactyly, an abnormal pattern of sweating or hair distribution, or angiokeratoma (Fabry's disease). Head and neck evaluation begins with a careful pupillary examination in a dark room, looking for asymmetry, miosis, mydriasis, reaction to light and accommodation, and rate and symmetry of redilation. A high-arched palate is present in some patients with OI and Ehlers-Danlos syndrome **(Chap. 425)**. The tongue's appearance (smooth vs normal) and size (macroglossia suggests amyloidosis) should be noted.

For patients in whom a neurologic cause is likely, a complete evaluation of the cranial nerves, strength and muscle tone, reflexes, sensory function including small-fiber modalities (pinprick, temperature), cerebellar function, and gait is important **(Chap. 433)**. This may require a neurologic referral if the primary care physician is uncomfortable with the assessment.

LOCALIZATION

Once the history and examination are complete, an initial localization is usually possible, based on a working knowledge of the anatomy and pharmacology of ANS components. For example, parasympathetic dysfunction is mainly cholinergic. Patients will report dry mouth and constipation, urinary retention, and erectile failure in men. On the other hand, sympathetic dysfunction is mainly adrenergic (norepinephrine, epinephrine, dopamine) and symptoms and signs of OI predominate, given the prominent effect of these neurotransmitters on cardiac function and vasomotor tone. These patients will report lightheadedness, dizziness, faintness, visual changes, and a need to sit down.

Muscarinic dysfunction leads to reduced sweating, hyperthermia, reduced salivation and lacrimation (sicca syndrome), mydriasis with blurred vision, slow gastrointestinal motility, and urinary retention. Tachycardia and arrhythmias are less common. Some CNS symptoms such as delirium and hallucinations may occur.

Adrenergic dysfunction presents with excessive sweating, vasomotor changes, irritable bowel syndrome, dysuria and stress incontinence. Tachycardia and arrhythmias are common. Miosis or mydriasis may occur, usually without blurred vision. Some CNS symptoms are panic, anxiety, and mood disorders.

Certain features favor preganglionic versus postganglionic localization. Preganglionic disorders are more likely to cause body temperature dysregulation, syncope and POTS, sleep disturbances, and autonomic storms. Postganglionic disorders are more likely in the presence of acral pain, acral anhidrosis, severe OH, and severe gastrointestinal or urinary dysmotility. However, overlap may occur as some pathologic conditions involve both preganglionic and postganglionic systems; examples include autoimmune, paraneoplastic, and neurodegenerative disorders.

LABORATORY EVALUATION

Testing begins by obtaining basic blood tests (complete blood count with differential, comprehensive metabolic panel, erythrocyte sedimentation rate). It is important to check iron, ferritin, vitamin B_{12}, and vitamin D levels as well as folic acid. In patients with gastrointestinal distress, particularly with significant weight loss, measurement of various nutritional factors and vitamins is important, including less common ones such as thiamine, vitamin B_6, niacin, riboflavin, vitamin C, vitamin E, serum copper, coenzyme Q10, and carnitine levels. A search for endocrine disorders such as diabetes, thyroid disease, or disorders of adrenal cortical (cortisol) and medullary (catecholamine) function may be necessary, as well as 5-hydroxyindoleacetic acid (5-HIAA) to look for carcinoid tumors. In patients with small-fiber neuropathy and normal hemoglobin A_{1c}, a 2-h glucose tolerance test is necessary, as glucose intolerance is a rather frequent cause of isolated small-fiber neuropathy in the absence of overt diabetes. In patients with OI, supine and upright catecholamine levels can provide useful information. In patients with central or preganglionic adrenergic dysautonomia, supine catecholamine levels tend to be normal but fail to rise significantly in the upright position; the normal healthy response is a doubling or greater. In patients with peripheral or postganglionic adrenergic dysautonomia, supine catecholamine levels tend to be low and also fail to rise significantly in the upright position.

In some acquired autonomic disorders, rheumatologic test screening, particularly for Sjögren's syndrome **(Chap. 373)**, scleroderma **(Chap. 372)**, and mixed connective tissue disease **(Chap. 372)**, is useful. If the bedside examination or autonomic testing shows evidence for a peripheral autonomic neuropathy, serum free light chains and protein electrophoresis to look for paraproteinemia is important, as it may be the sign of early primary amyloidosis (Chap. 117). In probable acquired autonomic disorders, serologic tests to check for autoimmune markers are indicated, including antibodies to the ganglionic ACh nicotinic receptor (see "Autoimmune Autonomic Ganglionopathy," below), voltage-gated potassium and calcium P/Q channel, plexin, and glutamic acid decarboxylase (GAD). If a paraneoplastic etiology is suspected, testing for anti-Hu, anti-CRMP5, anti-amphiphysin, or other autoantibodies may also be indicated (Chap. 99). If there is a history of infection, appropriate serologic tests can be requested (e.g., HIV, Epstein-Barr virus, herpes simplex virus, varicella-zoster virus, cytomegalovirus). In young patients or those with a positive family history, testing for porphyria (Chap. 428) and Fabry's disease (Chap. 429) is important, as these conditions are now treatable (Table 451-2).

AUTONOMIC EVALUATION TESTS

The most commonly employed autonomic tests evaluate cardiovascular function.

■ HEART RATE VARIATION TO DEEP BREATHING

This tests the parasympathetic cardiovascular reflexes mediated via the vagus nerve. The patient is instructed to breathe deeply at a rate of 6 breaths/min, and the acceleration and deceleration of the heart rate accompanying the inspiratory/expiratory cycle is calculated and averaged. The score is the average of the heart rate variability, or the expiratory to inspiratory ratio (Fig. 451-2). The ratio decreases gradually with age and is therefore age-corrected.

■ VALSALVA MANEUVER

This test is performed in the supine position. The patient blows into a closed mouthpiece to increase the intrathoracic pressure by

TABLE 451-2 Classification of Clinical Autonomic Disorders

I. Autonomic Disorders with Brain Involvement

- A. Associated with multisystem degeneration
  - 1. Multisystem degeneration: autonomic failure clinically prominent
    - a. Multiple-system atrophy (MSA)
    - b. Parkinson's disease with autonomic failure
    - c. Diffuse Lewy body disease with autonomic failure
  - 2. Multisystem degeneration: autonomic failure clinically not usually prominent
    - a. Parkinson's disease without autonomic failure
    - b. Other extrapyramidal disorders (inherited spinocerebellar atrophies, progressive supranuclear palsy, corticobasal degeneration, Machado-Joseph disease, fragile X syndrome [FXTAS])
- B. Unassociated with multisystem degeneration (focal CNS disorders)
  - 1. Disorders mainly due to cerebral cortex involvement
    - a. Frontal cortex lesions causing urinary/bowel incontinence
    - b. Focal seizures (temporal lobe or anterior cingulate)
    - c. Cerebral infarction of the insula
  - 2. Disorders of the limbic and paralimbic circuits
    - a. Limbic encephalitis
    - b. Autonomic seizures
    - c. Shapiro's syndrome (agenesis of corpus callosum, hyperhidrosis, hypothermia)
  - 3. Disorders of the hypothalamus
    - a. Thiamine deficiency (Wernicke-Korsakoff syndrome)
    - b. Diencephalic syndrome
    - c. Neuroleptic malignant syndrome
    - d. Serotonin syndrome
    - e. Fatal familial insomnia
    - f. Antidiuretic hormone (ADH) syndromes (diabetes insipidus, inappropriate ADH secretion)
    - g. Disturbances of temperature regulation (hyperthermia, hypothermia)
    - h. Disturbances of sexual function
    - i. Disturbances of appetite
    - j. Disturbances of BP/HR and gastric function
    - k. Horner's syndrome
  - 4. Disorders of the brainstem and cerebellum
    - a. Posterior fossa tumors
    - b. Syringobulbia and Arnold-Chiari malformation
    - c. Disorders of BP control (hypertension, hypotension)
    - d. Cardiac arrhythmias
    - e. Central sleep apnea
    - f. Baroreflex failure
    - g. Horner's syndrome
    - h. Vertebrobasilar and lateral medullary (Wallenberg's) syndromes
    - i. Brainstem encephalitis

II. Autonomic Disorders with Spinal Cord Involvement

- A. Traumatic quadriplegia
- B. Syringomyelia
- C. Subacute combined degeneration
- D. Multiple sclerosis and neuromyelitis optica
- E. Amyotrophic lateral sclerosis (mild, late)
- F. Tetanus
- G. Stiff-person syndrome
- H. Spinal cord tumors

III. Autonomic Neuropathies

- A. Acute/subacute autonomic neuropathies
  - a. Subacute autoimmune autonomic ganglionopathy (AAG)
  - b. Subacute paraneoplastic autonomic neuropathy
  - c. Guillain-Barré syndrome
  - d. Botulism
  - e. Porphyria
  - f. Drug-induced autonomic neuropathies—stimulants, drug withdrawal, vasoconstrictor, vasodilators, beta-receptor antagonists, beta-agonists
  - g. Toxin-induced autonomic neuropathies
  - h. Subacute cholinergic neuropathy
- B. Chronic peripheral autonomic neuropathies
  - 1. Small fiber neuropathy, including cryptogenic sensory polyneuropathy (CSPN)
  - 2. Combined sympathetic and parasympathetic failure
    - a. Amyloid
    - b. Diabetic autonomic neuropathy
    - c. AAG (paraneoplastic and idiopathic)
    - d. Sensory neuronopathy with autonomic failure
    - e. Fabry's disease
    - f. Diabetic, uremic, or nutritional deficiency
    - g. Familial dysautonomia (Riley-Day syndrome)
    - h. Hereditary sensory and autonomic neuropathy
    - i. HIV-related autonomic neuropathy
    - j. Geriatric dysautonomia (age >80 years)
  - 3. Disorders of orthostatic intolerance: OH; POTS; reflex syncope; prolonged bed rest; space flight; chronic fatigue

Abbreviations: BP, blood pressure; CNS, central nervous system; HR, heart rate; OH, orthostatic hypotension; POTS, postural orthostatic tachycardia syndrome.

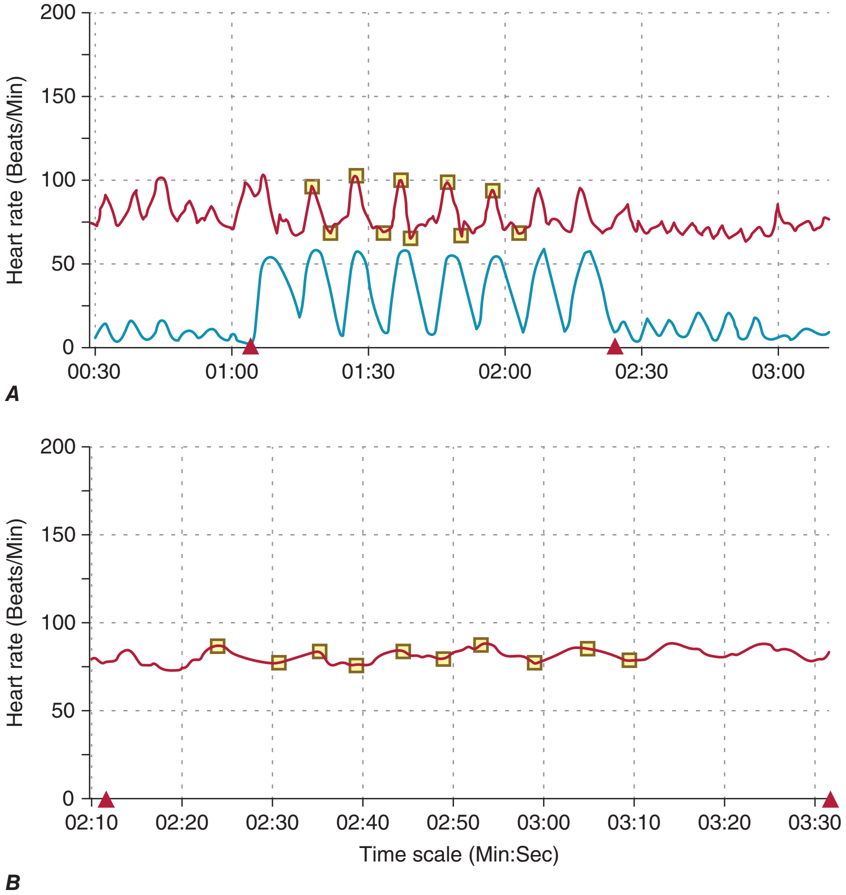

FIGURE 451-2 Examples of normal (*A*) and reduced (*B*) heart rate variability to deep breathing.

>40 mmHg for 15 s. During blowing, there is a decrease in venous return to the heart, leading to a fall in blood pressure and an associated compensatory tachycardia. Reduced blood pressure induces a compensatory peripheral sympathetic vasoconstriction in the lower extremities, which stops the blood pressure fall and reverses it close to baseline. Following release of the effort, blood pressure rises due to increased cardiac filling from the venous return, along with persistent sympathetic vasoconstriction. This leads to transient hypertension (an overshoot response) that stimulates the aortic and carotid baroreceptors leading to reflex bradycardia. The Valsalva ratio is the ratio of the higher to the lower heart rate and is a measure of the parasympathetic arm of the baroreflex. The magnitude of the drop in blood pressure, and of recovery in the overshoot phase, produces an adrenergic index score that, along with the blood pressure recovery time, is a marker of the sympathetic arm of the baroreflex (Fig. 451-3).

The heart rate variation to deep breathing and the Valsalva ratio are diminished in most patients with autonomic failure, including diabetes and α-synuclein disorders and also in patients who take anticholinergic medications.

■ TILT TABLE TESTING

This test of great value for patients with all syndromes of OI, with or without syncope. Many autonomic laboratories are equipped with special devices and a finger cuff that enables the accurate reproduction of beat-to-beat arterial waveforms. The tilt test is an excellent test of sympathetic vasomotor and cardiovascular innervation, as it detects OH in all its forms, as well as excessive orthostatic tachycardia (Fig. 451-4). When syncope occurs, testing allows differentiation between the various types (i.e., vasovagal, vasodepressor, and cardiodepressor syncope).

Tilt Table Testing for Recurrent Syncope When an initial phase of tilt at an angle of 60–70° for 30–40 min is negative, pharmacologic provocation of syncope (with intravenous, sublingual, or spray nitroglycerin, or intravenous isoproterenol) can be used. This increases the sensitivity of testing, at the expense of specificity (i.e., some patients with normal sympathetic function may also faint after administration of these drugs).

■ SUDOMOTOR FUNCTION TESTING

Sweating is secondary to the release of ACh from sympathetic postganglionic fibers. The quantitative sudomotor axon reflex test (QSART) measures regional ACh-induced sweating. The test is usually performed at four different sites (forearm, proximal leg, distal leg, and foot) to provide information on both the extent and distribution of postganglionic sudomotor impairment (Fig. 451-5). A low or absent response indicates a lesion of the postganglionic sudomotor axon. If the equipment necessary to perform the test is not available, sympathetic skin responses recorded from the foot and the hand using standard nerve conduction equipment can be used, although results obtained in this manner may be less sensitive and specific. As a general rule, autonomic or small-fiber neuropathies often lead to reduced sweating in the feet. The thermoregulatory sweat test (TST) is a qualitative measure of global sweat production in response to an elevation of body temperature under controlled conditions. An indicator powder dusted on the anterior surface of the forehead and body changes color with sweat production in a hot and humid chamber. The pattern of color change measures the integrity of both preganglionic and postganglionic sudomotor functions. A postganglionic lesion is present if both QSART and TST show absent sweating. In a preganglionic lesion, the QSART is preserved but TST shows anhidrosis.

■ SKIN BIOPSY

Skin biopsies are commonly performed to diagnose small-fiber neuropathies. These consist of small punch biopsy specimens, often 2–4 mm in diameter, stained for special axonal markers. This allows the identification of epidermal nerve fiber density as well as sudomotor innervation density. Patients often undergo biopsies from two to four sites, usually from the distal and proximal lower extremity as well

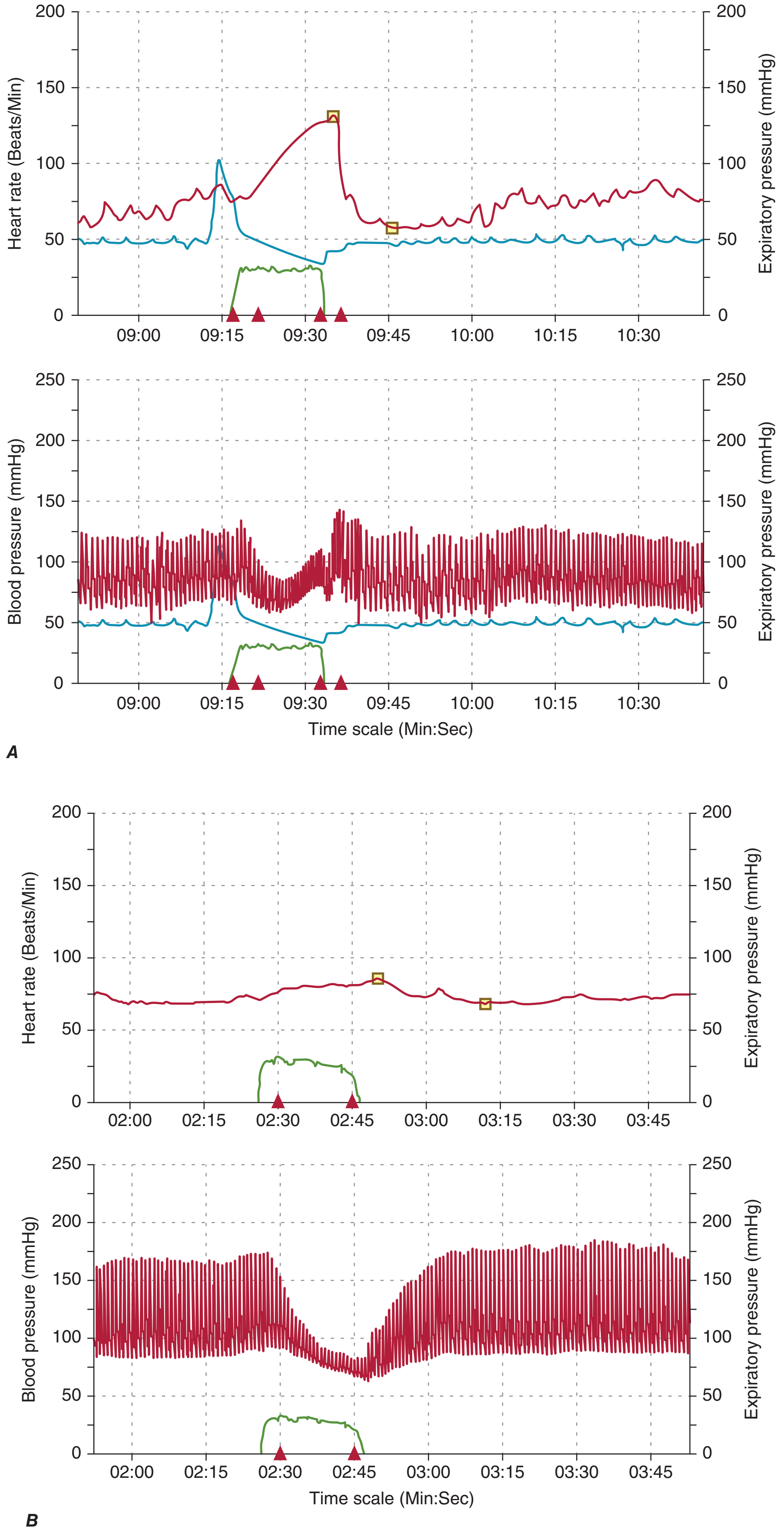

FIGURE 451-3 ***A.*** The tracing shows a normal Valsalva ratio and blood pressure response. ***B.*** A low Valsalva ratio and abnormal blood pressure response are illustrated.

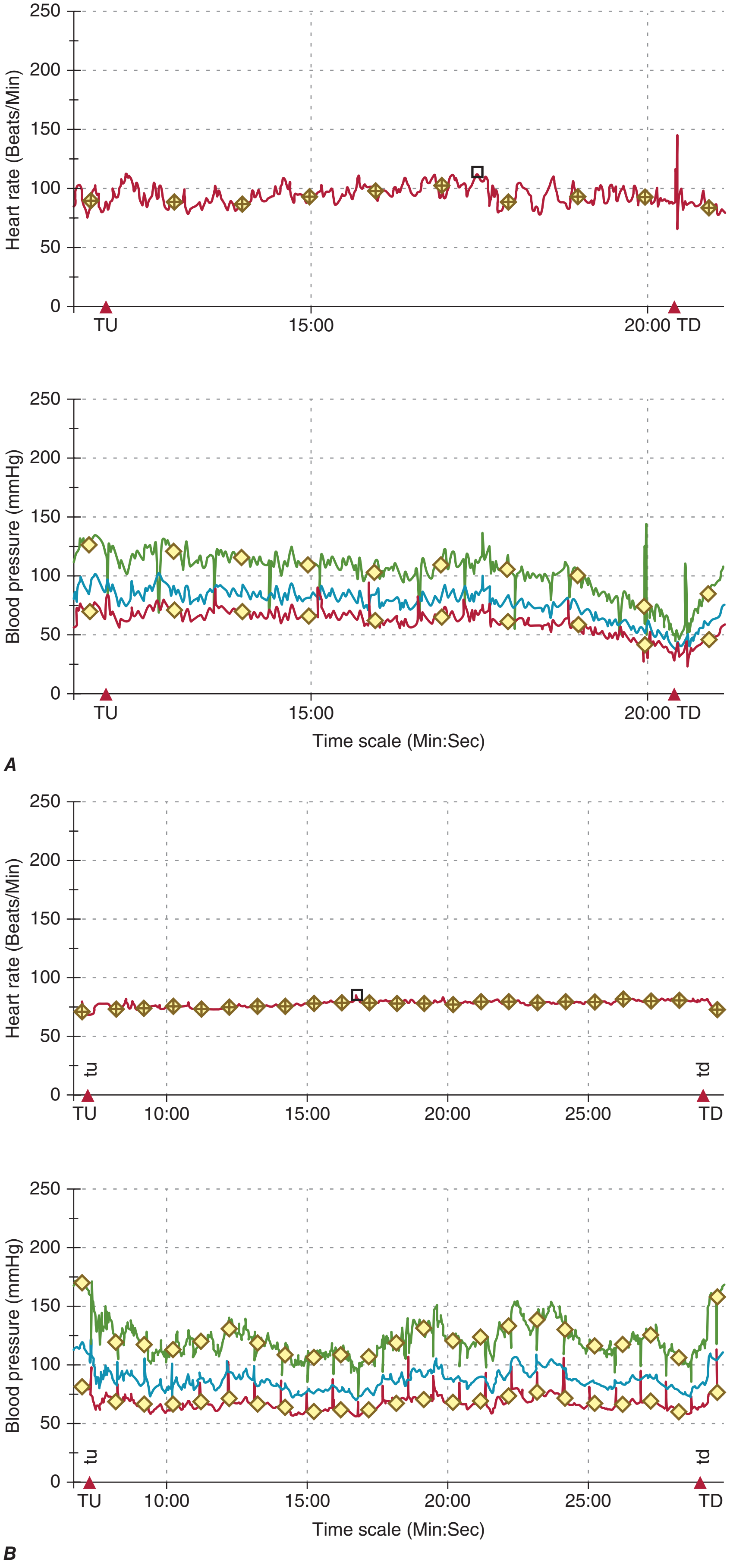

FIGURE 451-4 Examples of abnormal tilt responses. In each panel, the first tracing represents the heart rate, while the second represents the systolic, mean, and diastolic blood pressure responses. ***A.*** Progressive orthostatic hypotension leading to syncope. ***B.*** Neurogenic orthostatic hypotension (orthostatic hypotension without an adequate increase in heart rate). Notice the baseline supine hypertension (systolic blood pressure of 175 mmHg) due to associated baroreflex impairment. ***C.*** Postural orthostatic tachycardia syndrome (POTS). Notice the dramatic increase of the heart rate by >50 beats/min after tilt upward. TD, tilt downward; TU, tilt upward.

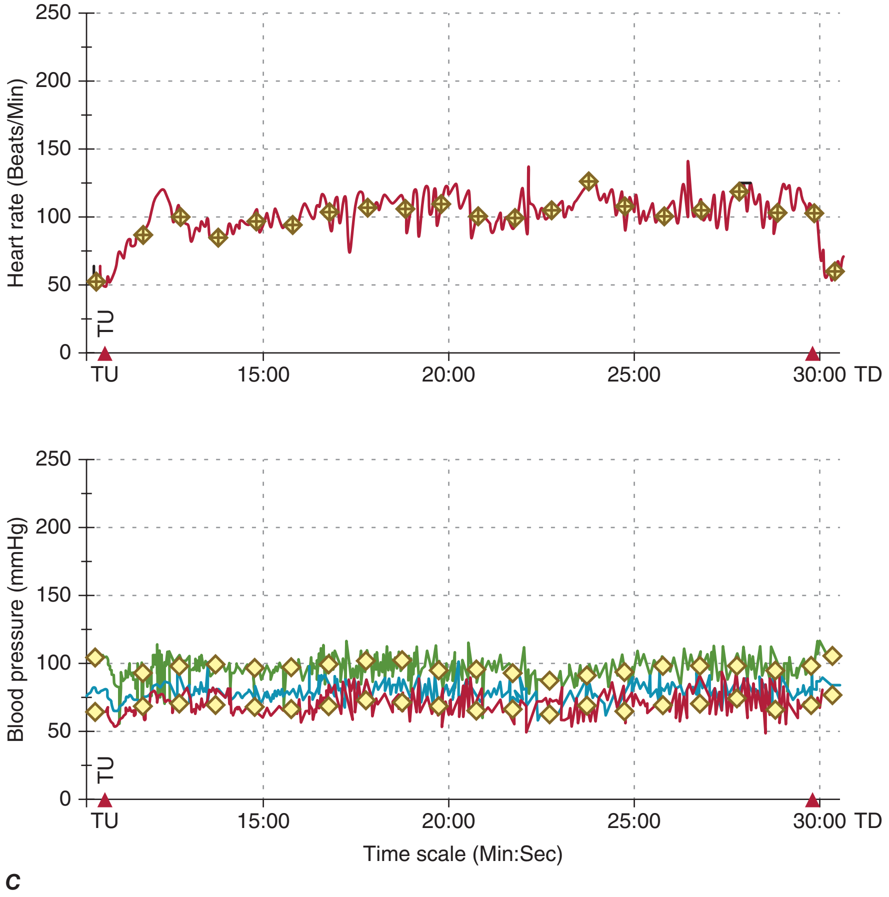

FIGURE 451-4 (*Continued*)

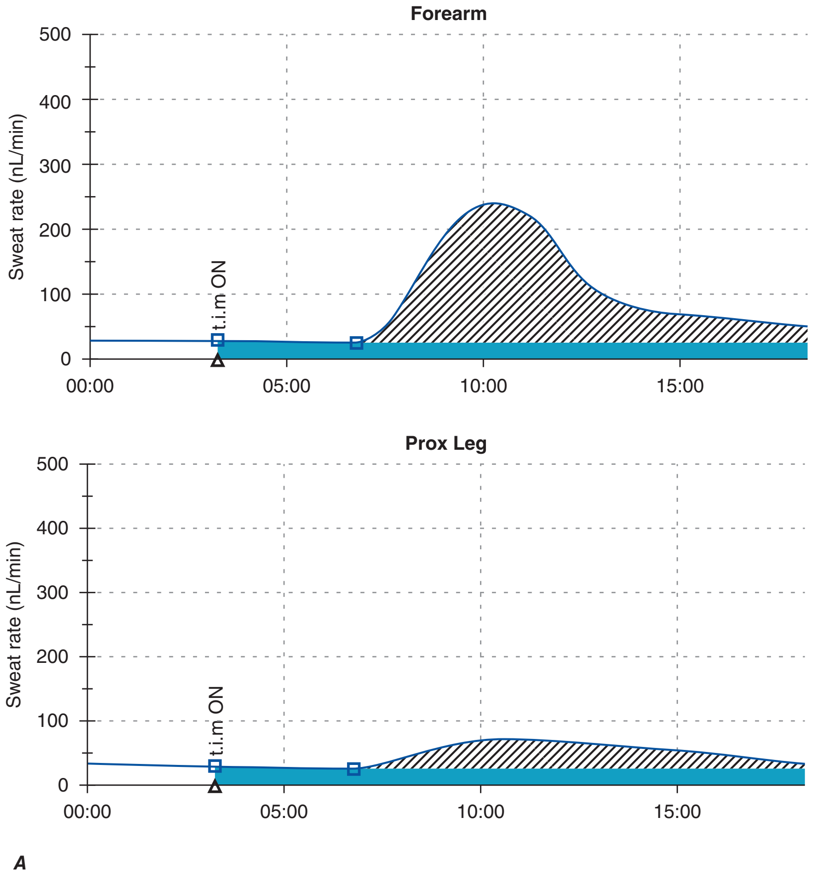

FIGURE 451-5 ***A.*** Examples of normal (forearm, top tracing) and reduced (proximal leg, bottom tracing) quantitative sweat responses consistent with a mild, length-dependent autonomic neuropathy. ***B.*** Normal (forearm, top tracing), reduced (proximal leg, middle tracing), and quasi-absent (distal leg, bottom tracing) quantitative sweat responses consistent with a length-dependent autonomic and small-fiber neuropathy. Tim ON, time when electrical stimulation is turned on to start iontophoresis of acetylcholine and induce sweating.

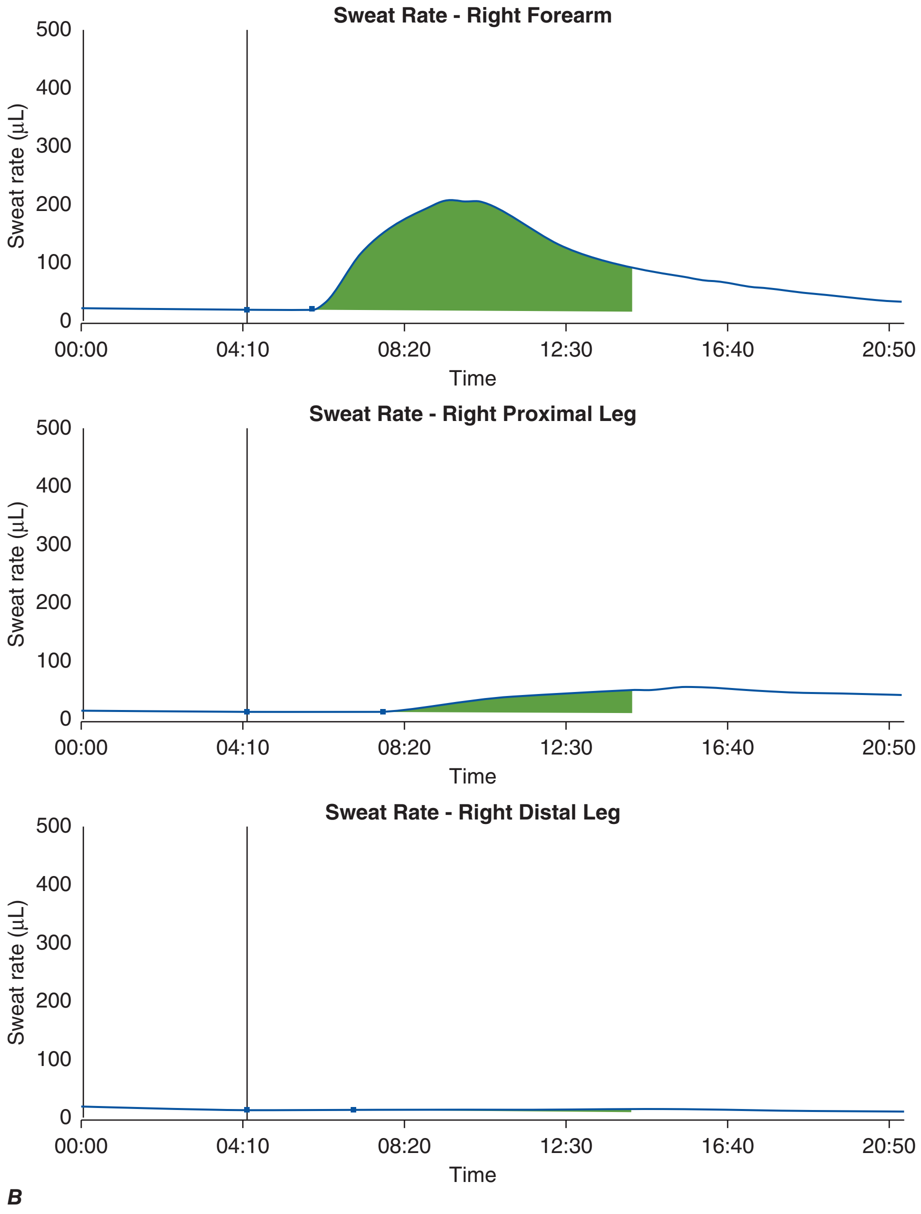

FIGURE 451-5 (*Continued*)

as the forearm. The reduction in the number of nerve fibers and the pattern of reduction assist in establishing the diagnosis but also provide a clue to the underlying etiology, particularly when the fiber loss is not length-dependent. Skin biopsies are also useful in diagnosing autonomic dysfunction secondary to synucleinopathies, as at least one specimen may demonstrate phosphorylated α-synuclein inclusions in postganglionic sympathetic adrenergic and cholinergic nerve fibers. In these cases, a more proximal biopsy obtained near the cervical paraspinal region can be a helpful diagnostic tool if it shows accumulation of α-synuclein (see specific disorders below).

If all testing remains unrevealing, it is reasonable to consider a fat aspirate biopsy to check for amyloidosis.

SPECIFIC CONDITIONS OF ANS DYSFUNCTION

■ ORTHOSTATIC HYPOTENSION

OH is a common disorder with many underlying conditions. It can be debilitating, leading to fatigue, falls, syncope, cognitive impairment, and end-organ damage. OH increases with advancing age and is commonly associated with a variety of general medical, neurologic, and primary autonomic disorders including diabetes, heart failure, kidney failure, and neurodegenerative and autoimmune diseases.

Upon standing or with tilt testing, ~500–700 mL of blood shifts to peripheral venous vessels in the lower limbs and splanchnic bed, reducing venous return. In a healthy person, there is slight increase in heart rate and peripheral venoconstriction and arteriolar constriction that maintains systolic blood pressure (SBP) and increases slightly diastolic blood pressure (DBP). This normal response depends on interactions between multiple homeostatic systems, including baroreceptors, autonomic afferents, central processing centers, autonomic efferents, and peripheral vascular sympathetic receptors.

Classical OH is defined as a sustained reduction of SBP ≥20 mmHg or DBP ≥10 mmHg within 3 min of active standing or on a head-up tilt (HUT) test ≥60°. Variants of OH include smaller but symptomatic reductions in SBP when the supine SBP is low (90–100 mmHg) but drops well below this level with standing or tilt. In patients with hypertension, OH is defined as an orthostatic drop of SBP/DBP ≥30/15 mmHg.

The prevalence of OH in the general population ranges from <5% below age 50 up to 20% above age 70. The diagnosis can be overlooked, as most patients are asymptomatic or have only minor symptoms. OH is associated with increased risk of falls, coronary heart disease, heart failure, stroke, and death.

When the orthostatic compensatory heart rate (HR) increase exceeds 15 beats/min, OH is nonneurogenic. On the contrary, neurogenic OH is associated with a blunted compensatory HR increase of <15 beats/min, provided there are no other possible explanations, such as cardioactive medications or intrinsic cardiac rhythm disturbances preventing an orthostatic HR increase. A more accurate way to diagnose neurogenic OH (NOH) is by dividing the HR increase by the fall in SBP (ΔHR/ΔSBP), as this metric provides a marker of cardiac baroreflex gain. A ratio <0.5 beats/min/mmHg after 3 min of standing (or tilt) more accurately indicates NOH. NOH is characterized by inadequate release of peripheral norepinephrine leading to impaired systemic vascular tone. The distinction is important because patients with NOH have a greater risk of mortality, approaching 44% over 30 months of follow-up, and exceeding 60% over 10 years. Furthermore, NOH can be associated with supine hypertension in patients with baroreflex impairment or autonomic failure, particularly in neurodegenerative conditions.

Clinical Presentation OH symptoms result from decreased blood flow to the brain. The most common symptom is lightheadedness on standing. Patients may report other orthostatic symptoms, such as palpitations, flushing, or pallor. More severe OH may lead to fainting or near-fainting. Triggers include bed rest (explaining the common occurrence upon waking up in the morning), food ingestion (particularly a large meal with alcohol), fever, heat, exercise, hyperventilation, medications, sepsis, or surgery. Aging, long periods of exercise, and intensification of antihypertensive treatment sometimes trigger the symptoms. Patients with underlying cardiac conditions (heart failure, chronotropic insufficiency) often have an added cardiogenic component to their OH. However, in the absence of triggers, many patients with mild OH remain asymptomatic, as cerebral autoregulation can compensate for mild swings in blood pressure and blood flow. Remarkably, in cases of chronic OH, particularly NOH, symptoms may remain quite mild in the face of significant blood pressure decrements.

Many patients with chronic OH have indirect symptoms. Instead of lightheadedness and near fainting, they may report dizziness (usually feeling wobbly, falling, or rarely vertigo), wooziness, or orthostatic imbalance. They may report fatigue and tiredness. Some may complain of generalized weakness after prolonged standing or walking. Some report neck and shoulder discomfort (coat hanger syndrome) secondary to hypoperfusion of the cervical paraspinal and trapezius muscles. In others, orthostatic dyspnea, known as platypnea, develops secondary to hypoperfusion of the lung apex. Myocardial hypoperfusion may cause orthostatic angina. A few patients also report orthostatic headache that may be mistaken as low-pressure cerebrospinal fluid headache.

OH can be classified into four functional classes. In functional class I, patients are asymptomatic but may show symptoms of OH including syncope and falls. In functional class II, symptoms occur weekly or monthly, with mild to moderate limitation of daily living. In functional class III, symptoms are more severe and frequent and markedly limit daily activities. In functional class IV, severe symptoms are daily, leading to significant disability.

Other OH variants include initial, delayed, and postprandial OH. Initial OH is an exaggerated, abrupt decrease in blood pressure (>40 mmHg SBP and >20 mmHg DBP) within 15–30 s upon standing due to transient mismatch between cardiac output and systemic vascular resistance. This condition occurs almost exclusively in young subjects (adolescents and young adults) but may also occur in older patients treated with antihypertensive drugs. The magnitude of the initial OH is often less with tilt than it is with active standing.

In delayed OH, SBP often remains steady for the first 3–5 min and then begins to decrease. In some cases, delayed OH does not appear until after 10 min of tilt. While the primary etiology in such cases may be secondary to medications, it often denotes some type of autonomic dysfunction, such as autonomic neuropathy.

Postprandial OH is a blood pressure drop after eating. OH often appears within 2 h of a larger meal, particularly if there is a high-carbohydrate load or alcohol ingestion. In these conditions, the meal produces gastric and esophageal distension and an increase in the splanchnic blood volume. It is more frequent in elderly patients with underlying neurologic conditions, particularly autonomic disorders. These patients sometimes faint after getting up from the dining table.

When OH is accompanied by baroreflex impairment, as seen with autonomic neuropathies or degenerative autonomic disorders, it can be associated with supine hypertension. In these patients, supine SBP during sleep may be higher than when they wake up in the morning, a pattern known as reverse dipping. Supine hypertension is a particular risk factor for end-organ damage (kidney, cardiac, or cerebrovascular). Over time, affected individuals develop alterations in the renin-angiotensin-aldosterone axis, leading to worse daytime OH by inducing pressure diuresis and volume loss during sleep.

Management Management of OH includes nonpharmacologic and pharmacologic measures. Nonpharmacologic measures are essential and may be sufficient in earlier stage I OH. However, the physician needs to ascertain whether the patient is actually asymptomatic and safe. This begins by addressing possible identifiable confounders, such as medications interfering with cardiovascular reflexes (e.g., diuretics, antihypertensives, tricyclic antidepressants, antipsychotics, dopamine agonists). Comorbidities can include iron deficiency, vitamin B_{12} and D deficiencies, and frailty. Increasing fluid intake up to 3 L/d and salt intake to 2 g up to 3 times a day can expand plasma and blood volume and often improve orthostatic tolerance. It is important to monitor the patient's weight and blood pressure on this regimen. In some patients, it would be reasonable to check the 24-h urine sodium before and after initiation of the fluid and salt load. Other simple remedies include lower limb stockings applied before getting out of bed. An abdominal binder is also quite effective in reducing the amount of splanchnic blood pooling. Physical countermaneuvers, such as crossing the legs while standing, are also helpful. It is also prudent to recommend elevating the head of bed by up to 30° to prevent the recurrence of supine hypertension and decrease nocturia.

Pharmacologic agents are quite effective. Fludrocortisone is a mineralocorticoid that expands the plasma volume, in addition to having a mild α-agonist effect. The dose is 0.1–0.2 mg/d (sometimes up to 0.4 mg/d) taken usually once in the morning. Because of its long duration of effect, it is important to monitor for supine hypertension. By increasing fluid retention, it can predispose to pedal edema, a low potassium level, headaches, and occasional mood changes. For older patients who require treatment for several years or longer, it is also prudent to check their bone density intermittently.

Midodrine is a direct agonist of α_1-adrenoceptors. It is prudent to start at a dose of 2.5 mg 3 times daily and increase gradually up to 15 mg 3 times daily. Most patients respond to 7.5–10 mg TID. The drug's onset of action is between 30 and 60 min, and its effect lasts up to 4 h. A usual schedule is to take it before getting out of bed, before lunch, and in mid-to-late afternoon, with the last dose usually before 6:00 P.M. to avoid nocturia and supine hypertension. Side effects include scalp tingling and goosebumps because of its stimulatory effect on pilomotor nerve fibers and urinary urgency because of its stimulatory effect at the bladder neck.

Pyridostigmine, a drug commonly used in myasthenia gravis, acts as a cholinesterase inhibitor and increases ganglionic transmission. Patients may be sensitive to the drug, so it is practical to start with 30 mg TID and increase gradually to 60 mg TID. It potentiates the effect of midodrine. It may cause abdominal cramps and diarrhea in susceptible patients. It is particularly beneficial for autonomic patients with OH and slow gastrointestinal motility. It does not promote supine hypertension, an advantage for this drug.

Atomoxetine is a norepinephrine reuptake inhibitor that is useful as an adjuvant therapy. Most patients benefit from 18 mg in the morning. While some patients may need it midday to early afternoon, it is important to monitor for symptoms and signs of CNS stimulation such as anxiety and tremor.

Droxidopa is the newest agent approved for treatment of NOH. It is a norepinephrine precursor. Treatment should start with 100 mg TID and be gradually increased to 400–600 mg TID. The medication improved orthostatic symptoms and decreased falls in PD patients. Supine hypertension remains an issue in most patients.

Desmopressin increases water reabsorption and reduces urination. Low-dose oral administration or nasal spray can be useful in reducing the frequency of nocturia that often interrupts sleep. Patients should be monitored for hyponatremia.

When patients with OH or NOH develop supine hypertension, it is important to revise their medications, monitor end-organ function, ensure elevation of the head of bed during sleep (30–45°), and use short-acting blood pressure–lowering drugs at bedtime.

POSTURAL ORTHOSTATIC TACHYCARDIA SYNDROME

POTS is a common syndrome of chronic (>3 months) OI without sustained OH, accompanied by an increase in HR of ≥30 beats/min in adults (≥40 beats/min if under the age of 19 years) or an orthostatic HR >120 beats/min within 10 min of standing that subsides on sitting or lying down. The symptoms reflect a combination of cerebral hypoperfusion (lightheadedness, blurred vision, cognitive decline, and brain fog) and sympathetic activation (palpitations, chest pain, and tremulousness). Near-syncope is common, but complete syncope is less frequent. The condition is four to five times more common in women, and most develop the syndrome between the ages of 15 and 45 years. Symptoms may flare up around menstrual cycles.

POTS is often multifactorial, as the pathophysiology is heterogeneous: distal sympathetic denervation in the feet and legs with preserved cardiovascular sympathetic function (so-called neuropathic POTS), venous pooling, reduced cardiac function due to deconditioning, hypovolemia, altered baroreceptor regulation, increased sympathetic activity, and rarely reduced parasympathetic activity may play a role. About half the cases follow a viral illness. Most prominent among these are Epstein-Barr virus, coronavirus, and enteroviral illnesses. The condition has increased since the COVID-19 pandemic; it is the leading manifestation of long COVID. Other POTS triggers include surgery, the postpartum period, puberty, pregnancy, concussion, and chronic emotional distress. Ehlers-Danlos syndrome is a significant predisposing condition, present in up to 25% of patients. When symptoms of POTS are prominent, patients become sedentary, exacerbating their deconditioning and leading to a vicious cycle with prolongation of the syndrome. In the chronic state, cardiovascular deconditioning becomes the common pathway.

POTS comorbidities include migraine and various headache disorders, symptoms of gastrointestinal dysmotility, chronic fatigue, fibromyalgia, gastrointestinal and bladder visceral pain, and sleep disturbances. Some patients report temperature intolerance, more to heat than cold, with impaired sweating (either increased or decreased). Patients often appear anxious in the clinic and have a tendency to be hypervigilant with respect to somatic symptoms, but the incidence of significant psychiatric disorders does not differ from that in the general population. Some patients report a new onset of various allergic symptoms (flushing, hives, and environmental and food allergies) suggestive of some type of mast cell activation syndrome.

While there is no standardized approach to the diagnosis, it is important to rule out primary cardiac causes, including arrhythmias (POTS is a sinus rhythm disorder), as well as endocrine conditions that lead to sympathetic activation, such as pheochromocytoma and hyperthyroidism. In addition to routine complete blood count and electrolyte panel, it is important also to measure ferritin and vitamin D levels and correct these if low, as this helps the condition respond better to other treatments.

Management has some similarities to the management of OH. Typically, it encourages the expansion of fluid volume with water (8–10 cups daily) and salt (6–8 g/d). Elastic stockings, waist high with 20–30 mmHg counterpressure, reduce venous pooling and improve venous return. Pharmacotherapy with low-dose fludrocortisone can be helpful. When symptoms of sympathetic overactivity are prominent, low doses of a beta blocker, increased gradually, can be helpful. These patients can be sensitive to beta blockers, so it is important to measure their blood pressure while on treatment; if it decreases, they would need a higher dose of fludrocortisone or the addition of a vasoactive drug such as midodrine. The latter is particularly useful in patients with neuropathic POTS, as it counteracts the significant peripheral vasodilation that is present. In patients with gastrointestinal hypomotility, low-dose pyridostigmine can be helpful and may improve the patient's fatigue as well. When tachycardia proves refractory to the above measures, low-dose ivabradine can also be tried. A graduated program of exercise to improve cardiac deconditioning is essential and often needs at least 6 months before significant benefit is obtained. Because of the orthostatic nature of the condition, exercise is better tolerated in the supine or sitting position (recumbent bike, rowing, swimming) and should involve a combination of aerobic cardiovascular training and resistance training primarily to the leg muscles. If the patient's symptoms and tolerance improve after 3 months, exercise in the upright position (treadmill, elliptical) can be gradually advanced.

SYNUCLEINOPATHIES

This group of disorders is secondary to the accumulation of phosphorylated α-synuclein in various parts of the ANS. They include multiple system atrophy (MSA), Parkinson's disease (PD; **Chap. 446**) with autonomic failure, Dementia with Lewy bodies (DLB; **Chap. 445**), and primary autonomic failure (PAF). Pathologically, there are aggregates of misfolded α-synuclein in the CNS and PNS. Cardiovascular autonomic denervation is more common in PD and PAF, whereas it tends to occur later in MSA and DLB.

PAF is a sporadic peripheral autonomic degenerative disease that has clinical manifestations similar to those in other autonomic neuropathies, such as diabetic autonomic neuropathy, and the autoimmune autonomic neuropathies. NOH is the prominent clinical feature and frequently associated with supine hypertension. After 5–10 years, a significant number of PAF patients develop a CNS synucleinopathy, particularly PD or DLB but also MSA. In addition to the presence of autonomic dysfunction, skin punch biopsy with immunohistochemistry staining can demonstrate small nerve fibers and intraneural α-synuclein deposition.

MSA, or *Shy-Drager syndrome*, is a disorder that combines autonomic failure *and* either parkinsonism (MSA-P) or a cerebellar syndrome (MSA-C). MSA is uncommon, with a prevalence estimated at 2–5 per 100,000 individuals. Onset is typically in the mid-fifties, men are slightly more affected than women, and most cases are sporadic. MSA generally progresses relentlessly to death 7–10 years after onset, but some patients may survive beyond 10 years. MSA-C is more common in Asian population, particularly in Japan, while MSA-P is the more common form in the rest of the world, including Western countries. A characteristic of the extrapyramidal parkinsonian phenotype is the relatively poor response to levodopa; when a response occurs, it tends to be transient. Rigidity dominates, while rest tremor is not prominent. Other distinguishing features that favor MSA-P over idiopathic PD is the early impairment of urinary control in both genders, early erectile dysfunction in men, symptomatic OH within 2 years of onset, preserved olfaction, and a lower frequency of gastrointestinal symptomatology, especially colonic hypomotility and constipation. Early autonomic dysfunction, early bladder dysfunction, female gender, and rapid progression of disability imply a worse prognosis.

Brain magnetic resonance imaging (MRI) can show iron deposition in the striatum in MSA-P, and cerebellar atrophy with a characteristic T2 hyperintense signal ("hot cross bun" sign) in the pons in MSA-C **(Fig. 451-6)**. However, the typical MRI findings may be present only with advanced disease. Cardiac postganglionic adrenergic innervation, measured by the uptake of meta-iodobenzylguanidine (MIBG) on a radionuclear scan or fluorodopamine on positron emission tomography, is usually intact in MSA, whereas it is low in the dysautonomias of PD, LBD, and PAF. The neuropathology of MSA shows neuronal loss and gliosis in many CNS regions, including the brainstem, cerebellum, striatum, and intermediolateral cell column of the thoracolumbar spinal cord. The characteristic pathologic feature is the presence of glial cytoplasmic inclusions that stain positively for α-synuclein (early Lewy bodies) primarily in oligodendrocytes, in contrast to their neuronal localization in PD or LBD. Research has shown some transfer of these cytoplasmic inclusions from cell to cell, akin to that of a prion **(Chap. 449)**.

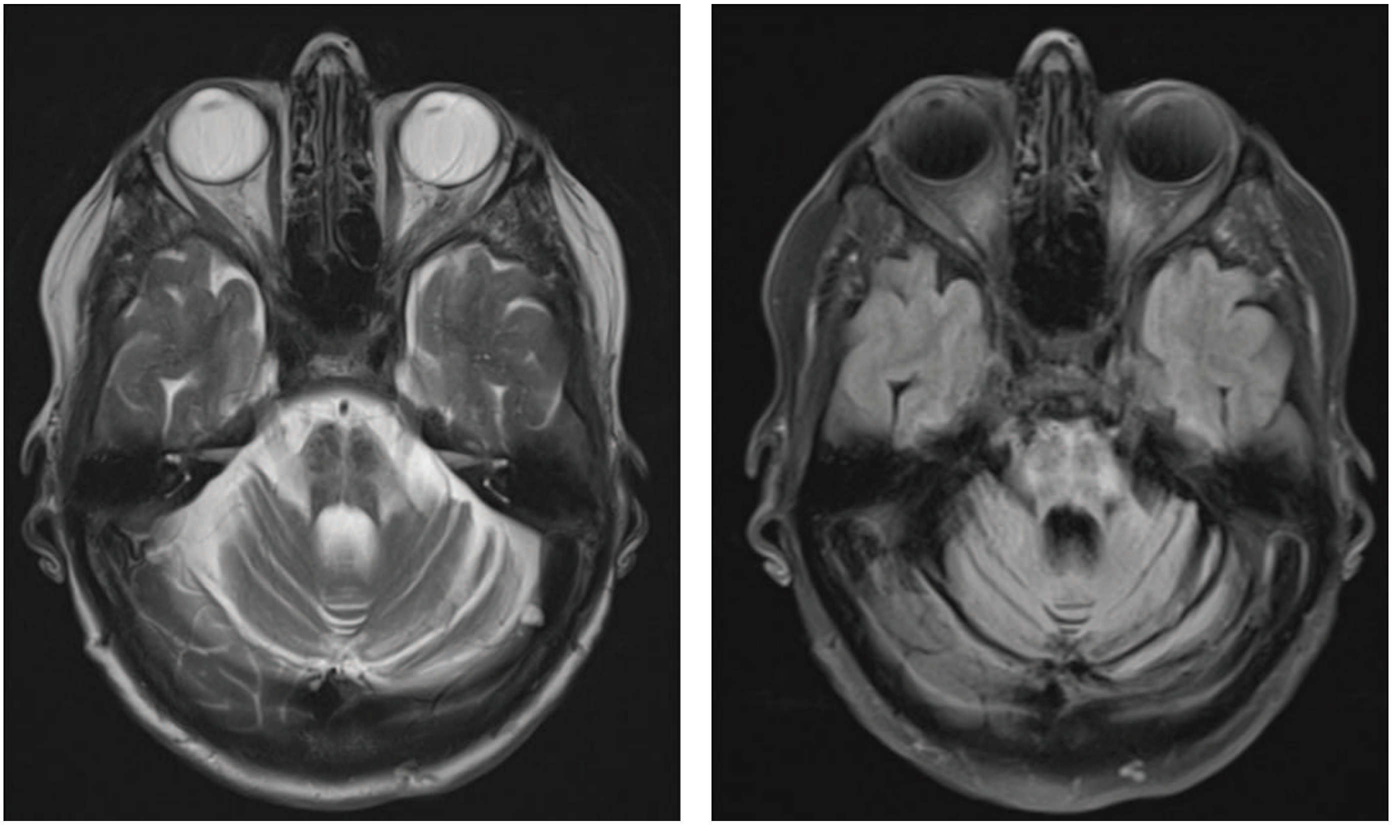

FIGURE 451-6 Multiple system atrophy, cerebellar type (MSA-C). Axial T2-weighted and FLAIR (fluid-attenuated inversion recovery) magnetic resonance images at the level of the pons demonstrate a characteristic cruciform hyperintense signal in the pons, the "hot cross bun" sign. This appearance is characteristic but not pathognomonic of MSA-C, as it can also be seen in some spinocerebellar atrophies and other neurodegenerative conditions affecting the brainstem. The images additionally demonstrate atrophy of the pons, bilateral middle cerebellar peduncles, and cerebellum. *(Courtesy of Dr. Nancy Fischbein, Stanford University School of Medicine.)*

Management is symptomatic for NOH (see earlier section on OH), sleep disorders including laryngeal stridor, and gastrointestinal and urinary dysfunction. Gastrointestinal symptoms can be managed with frequent small meals, a soft diet, stool softeners, and bulk agents. Persistent significant gastric and colonic hypomotility requires the administration of the newer motility stimulation drugs, such as prucalopride, linaclotide, plecanatide, or lubiprostone. Metoclopramide stimulates gastric emptying but should be used only as a short-term measure, if at all, because it worsens parkinsonism by blocking central dopamine receptors. Domperidone, a peripheral dopamine (D_2 and D_3) receptor antagonist, is also effective but is not U.S. Food and Drug Administration approved for use in the United States because of its potential arrhythmogenic risk. Pharmacologic management of neurogenic bladder relies usually on muscarinic blockers to reduce bladder urgency. Oxybutynin is often a first-line agent, but increasing the dose may cause other antimuscarinic side effect such as dry mouth, dry eyes, and anhidrosis. More selective blockers (tolterodine, darifenacin, or solifenacin) produced fewer systemic side effects. Mirabegron and vibegron are selective β_3-adrenergic agonists that reduces detrusor overactivity and are quite effective. A potential beneficial side effect is their gentle hypertensive effect that may help with OH management but may also require adjusting doses of other OH medications to avoid significant supine hypertension.

Autonomic dysfunction also occurs in DLB **(Chap. 445)** with the severity usually intermediate between that of MSA and PD. In multiple sclerosis (MS; **Chap. 455**), autonomic complications reflect the CNS location of MS lesions (more common with brainstem and spinal cord demyelination) and generally worsen with disease duration and disability but are generally less severe than with synucleinopathies.

SPINAL CORD DISORDERS

Spinal cord lesions may cause focal autonomic deficits dominated by OH, bowel dysfunction, and genitourinary dysfunction, as well as temperature dysregulation secondary to anhidrosis below the level of the lesion. Quadriparesis predisposes to supine hypertension and OH after upward tilting. *Autonomic dysreflexia* is a dramatic blood pressure increase that occurs in response to irritation of the bladder, skin, or muscles. It usually follows traumatic spinal cord lesions above the T6 level, as lesions below T6 allow for compensatory splanchnic vasodilation and a lesser likelihood of autonomic dysreflexia. The condition sometimes interferes with patient care, particularly when patients require urinary catheterization for distended bladder or rectal management for constipation and fecal impaction. In addition to blood pressure surges that may induce intracranial vasospasm or hemorrhage, other symptoms include facial flushing, headache, hypertension, piloerection, and even cardiac arrhythmia. Sitting the patient up reduces excessive supine hypertension. Prophylactic clonidine can reduce the hypertension resulting from bladder or rectal stimulation. **Disorders of the spinal cord are discussed in Chap. 453.**

PERIPHERAL NERVE AND NEUROMUSCULAR TRANSMISSION DISORDERS

Peripheral autonomic neuropathies are the most common cause of autonomic insufficiency. These complex disorders can impact cardiovascular, gastrointestinal, urogenital, and sudomotor systems. They affect small myelinated and unmyelinated fibers of the sympathetic and parasympathetic nerves and commonly occur in diabetes mellitus, amyloidosis, chronic alcoholism, porphyria, idiopathic small-fiber polyneuropathy, and Guillain-Barré syndrome. Neuromuscular junction disorders with autonomic involvement include botulism and Lambert-Eaton myasthenic syndrome **(Chaps. 457–459)**.

Diabetes Mellitus Diabetes mellitus remains one of the leading causes of peripheral autonomic neuropathy. The presence of autonomic neuropathy increases the mortality rate of diabetes up to threefold, even after adjusting for other cardiovascular risk factors. The neuropathy risk correlates with the product of the patient's hemoglobin A_{1c} values and the duration of the diabetes. Early autonomic impairment is most often manifest as cardiovascular parasympathetic hypofunction. As diabetes progresses, sympathetic hypofunction follows, dominated by OH. The autonomic involvement also predicts other complications including coronary artery disease, renal disease, stroke, and sleep apnea. It also plays a significant role in the gastrointestinal dysmotility, which may further complicate glycemic control. Improved glycemic control significantly reduces the long-term risk of autonomic cardiovascular neuropathy. **Diabetes mellitus is discussed in Chaps. 415–417.**

Amyloidosis Amyloidosis, both acquired primary (AL) and hereditary transthyretin (hATTR) types, leads to autonomic neuropathy. Patients usually present with a distal sensorimotor polyneuropathy accompanied by autonomic insufficiency, but an autonomic neuropathy in isolation may also occur. A history of carpal tunnel syndrome is common. Postmortem studies reveal amyloid deposition in many organs, including two sites that contribute to autonomic failure: intraneural blood vessels and autonomic ganglia. For AL amyloidosis, the diagnosis is made by blood tests showing monoclonal gammopathy and elevated free light chains (with lambda more often than kappa being the culprit). For hATTR, If genetic testing for *TTR* mutations is negative, confirmation usually requires histologic diagnosis (abdominal fat pad, rectal mucosa, nerve/muscle biopsy, or myocardial biopsy in cases with cardiomyopathy) to search for amyloid deposits. Gene-based treatments (either RNA silencing or antisense oligonucleotides) are effective for hATTR amyloidosis. Treatment for AL amyloidosis usually relies on chemotherapeutic drugs. Death is usually due to cardiac or renal involvement. **Amyloidosis is discussed in Chap. 117.**

Alcoholic Neuropathy Abnormalities in parasympathetic vagal and efferent sympathetic function are usually mild in alcoholic polyneuropathy. OH is usually due to brainstem involvement, rather than injury to the PNS. Impotence is a common symptom, but concurrent gonadal hormone abnormalities may play a role in this symptom. Clinical symptoms of autonomic failure generally appear only when the stocking-glove polyneuropathy is severe, and there may be coexisting ataxia or Wernicke's encephalopathy **(Chap. 318)**. Autonomic involvement may contribute to the high mortality rates associated with alcoholism. **Alcohol use disorders are discussed in Chap. 464.**

Porphyria Autonomic dysfunction occurs only in the hepatic porphyrias. It is documented in acute intermittent porphyria, but may also occur with variegate porphyria and hereditary coproporphyria. Sympathetic overactivity dominates the clinical presentation. Autonomic symptoms include tachycardia, sweating, urinary retention, abdominal pain, nausea and vomiting, insomnia, hypertension, and (less commonly) hypotension. Often, patients have psychiatric symptoms dominated by anxiety, but depression and bipolar-like presentations may occur. Abnormal autonomic function can occur both during acute attacks and during remissions. Elevated catecholamine levels during acute attacks correlate with the degree of tachycardia and hypertension that is present. **Porphyria is discussed in Chap. 428.**

Guillain-Barré Syndrome (GBS) GBS often causes autonomic instability, more often but not exclusively in more severe cases. Cardiovascular dysautonomia can be life-threatening, and gastrointestinal autonomic involvement, abnormal sweating, and pupillary dysfunction may also occur. The dysautonomia is secondary to the demyelination of the vagus and glossopharyngeal nerves and the preganglionic sympathetic white rami communicantes. In some cases, autonomic involvement outweighs other manifestations of the motor or sensory neuropathy. *Acute autonomic and sensory neuropathy* is a variant that spares the motor system and presents with NOH and varying degrees of sensory loss. The treatment is similar to that for GBS (IV immunoglobulin or plasma exchange), but the prognosis is less favorable, with persistent sensory deficits and variable degrees of OH present in many patients. **GBS is discussed in Chap. 458.**

Autoimmune Autonomic Ganglionopathy (AAG) AAG is a rare form of dysautonomia. Approximately 100 Americans are diagnosed each year. It affects people of both sexes and all ages, although it is more common in women and in adults. Other names for this condition are acute pandysautonomia, idiopathic subacute autonomic neuropathy, and autoimmune autonomic neuropathy. It is typically associated with high titers of autoantibodies against the α3 subunit of the ganglionic acetylcholine nicotinic receptor (g-AChR α3). AAG often follows a triggering event: infection, surgery, autoimmunity, or malignancy in most patients. This condition is important to recognize because it is a treatable disorder. Symptoms mimic those caused by ganglion-blocking drugs *without* neuromuscular block, depending upon on which ANS division predominates. Ocular involvement leads to mydriasis and cycloplegia with blurred vision. Bronchial involvement leads to bronchodilation. Gastrointestinal involvement leads to reduced motility and secretion. Genitourinary involvement leads to urinary retention and impaired erection and ejaculation. Cardiac involvement leads to mild tachycardia. Vascular involvement reduces sympathetic tone and produces postural hypotension. Involvement of salivary and sweat glands leads to dry mouth and impaired sweating. Symptoms of autonomic failure occur in various combinations: syncope and OH; labile blood pressure; gastrointestinal hypomotility; bladder hypomotility; pupillary dysfunction with Adie's tonic pupil (slow reaction to light, bitter reaction to accommodation, slow relaxation afterward); dry mouth and eyes; and anhidrosis. Optic neuritis has also been reported in smokers with positive g-AChR α3 antibodies.

Some patients with higher titers of AAG have a paraneoplastic form of the disease associated with small cell lung carcinoma, thymoma, or lymphoma **(Chap. 99)**. Therefore, it is prudent to search for malignancy in the initial workup of AAG. Although no large studies have evaluated the efficacy of various treatments in AAG, IV immunoglobulin or plasmapheresis is generally used as first-line therapy, with glucocorticoids, rituximab, or mycophenolate mofetil, alone or in combination, used for nonresponders. Symptomatic treatment is of utmost importance, with attention to OI, gastrointestinal dysmotility, and genitourinary disturbances. Nonpharmacologic treatments such as exercise, increasing salt and fluid intake, compression stockings, and good sleep habits can also help.

Seronegative Autoimmune Autonomic Neuropathy (SAAN) SAAN is likely a seronegative variant of AAG, as this condition presents with subacute autonomic disturbances similar to AAG. Pathology shows preferential involvement of small unmyelinated nerve fibers, with sparing of larger myelinated ones. The neuropathy follows a viral infection in about half of cases. Several cases have been associated with use of checkpoint inhibitors. About one-third of untreated patients improve over time. Management is similar to that for AAG, although some patients may respond better to glucocorticoids as first-line therapy.

Botulism Botulinum toxin binds presynaptically to cholinergic nerve terminals and blocks ACh release. This condition presents with motor paralysis and signs of cholinergic failure that include blurred vision, sluggishly reactive pupils, dry mouth, dry skin, reduced gastric and colonic motility, and urinary retention **(Chap. 138)**.

Chronic Idiopathic Anhidrosis Chronic idiopathic anhidrosis is a rare condition in which patients present with heat intolerance but preserved vasomotor function. There is no associated somatic neuropathy. Some patients may have Adie's tonic pupils as well (Ross syndrome).

Acknowledgment

Richard J. Barohn and John W. Engstrom contributed to this chapter in the prior edition and material from that chapter has been retained here.

FURTHER READING

Campbell WW, Baron RJ: The autonomic nervous system, in *DeJong's The Neurologic Examination*, 8th ed. WW Campbell, RJ Barohn (eds). Philadelphia, Wolters Kluwer, 2020.

Cheshire WP: Autonomic history, examination and laboratory evaluation. Continuum (Minneap Minn) 26:25, 2020.

Donadio V et al: Phosphorylated α-synuclein in skin Schwann cells: A new biomarker for multiple system atrophy. Brain 146:1065, 2023.

Fedorowski A, Sutton R: Autonomic dysfunction and postural orthostatic tachycardia syndrome in postacute Covid 19 syndrome. Nat Rev Cardiol 20:281, 2023.

Gibbons C et al: Cutaneous α-synuclein alpha signatures in patients with multiple system atrophy and Parkinson disease. Neurology 100:e1529, 2023.

Gibbons CH et al: Skin biopsy in evaluation of autonomic disorders. Continuum (Minneap Minn) 26:200, 2020.

Jaradeh SS, Prieto TE: Evaluation of the autonomic nervous system. Phys Med Rehabil Clin N Am 14:287, 2003.

Lamotte G, Sandroni P: Updates from the diagnosis and treatment of peripheral autonomic neuropathies. Curr Neurol Neurosci Rep 12:823, 2022.

Vernino S: Autoimmune autonomic disorders. Continuum (Minneap Minn) 26:44, 2020.

452

Trigeminal Neuralgia, Bell's Palsy, and Other Cranial Nerve Disorders

Vanja C. Douglas, Stephen L. Hauser

The cranial nerves consist of 12 paired nerves that mediate variable combinations of motor, sensory, and autonomic functions. They are considered as a group because of their close anatomic relationship to the brainstem **(Fig. 452-1)** and to one another, and tendency to be involved together in a variety of disease states. Nine cranial nerves connect directly with brainstem nuclei; the exceptions are cranial nerves 1 (olfactory) and 2 (optic) that are more accurately considered fiber tracts of the brain, and cranial nerve 11 (spinal accessory) whose motor neurons reside largely in the upper cervical cord. Analogous to spinal nerves **(Chap. 453)**, motor fibers of the cranial nerves have their origin in the brainstem or upper cervical cord, while sensory nerves are pseudounipolar, with ganglia outside the central nervous system and a synapse with second-order fibers in the brainstem.

Symptoms and signs of cranial nerve pathology are common in internal medicine. They may develop in the context of a widespread neurologic disturbance, and in such situations, cranial nerve involvement may represent the initial manifestation of the illness. In other disorders, involvement is largely restricted to one or several cranial nerves; these distinctive disorders are reviewed in this chapter. **Disorders of olfaction are discussed in Chap. 35, vision and ocular movement in Chap. 34, hearing in Chap. 36, and vestibular function in Chap. 24.**

FACIAL PAIN OR NUMBNESS

ANATOMIC CONSIDERATIONS

The trigeminal (fifth cranial) nerve supplies sensation to the skin of the face, anterior half of the head, and the nasal and oral mucosa **(Fig. 452-2)**. The motor part innervates the muscles involved in chewing (including masseter, temporalis, and pterygoids) as well as the anterior belly of the digastric, mylohyoid, tensor veli palatini, and the tensor tympani (hearing especially for high-pitched tones). It is the largest of the cranial nerves. It exits in the lateral midpons and traverses the middle cranial fossa to the semilunar (gasserian, trigeminal) ganglion in Meckel's cave, where the nerve splits into three divisions (ophthalmic [V1], maxillary [V2], and mandibular [V3]). V1 and V2 traverse the cavernous sinus to exit in the superior orbital fissure and foramen rotundum; V3 exits through the foramen ovale. The trigeminal nerve is predominantly sensory, and motor innervation is exclusively carried in V3. The cornea is primarily innervated by V1, although an inferior crescent may be V2. Upon entering the pons, pain and temperature fibers descend ipsilaterally as far as the upper cervical spinal cord as the spinal tract of V, before synapsing with the spinal nucleus of V; this accounts for the facial numbness that can occur with spinal cord lesions above C2. In the brainstem, the spinal tract of V is also located adjacent to crossed ascending fibers of the spinothalamic tract, producing a "crossed" sensory loss for pain and temperature (ipsilateral face, contralateral arm/trunk/leg) with lesions of the lateral lower brainstem. CN V is also ensheathed by oligodendrocyte-derived, rather than Schwann cell–derived, myelin for up to 7 mm after it leaves the brainstem, unlike just a few millimeters for other cranial and spinal nerves; this may explain some instances of trigeminal neuralgia in multiple sclerosis (MS) **(Chap. 455)**, a disorder of oligodendrocyte myelin.

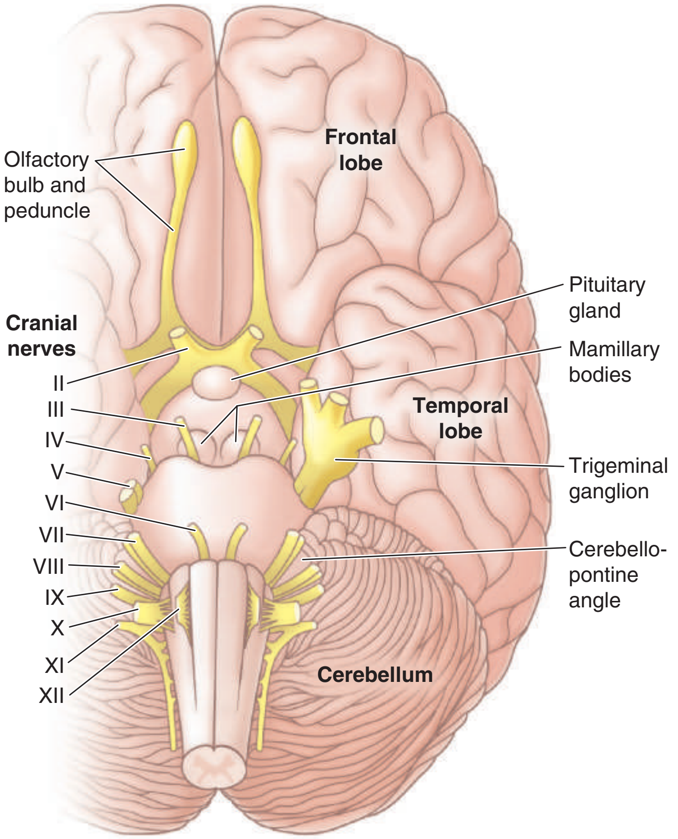

FIGURE 452-1 Ventral view of the brain, illustrating relationships between the 12 pairs of cranial nerves and the brainstem. *(Adapted from SG Waxman: Clinical Neuroanatomy, 29th ed. http://www.accessmedicine.com.)*

TRIGEMINAL NEURALGIA (TIC DOULOUREUX)

Clinical Manifestations Trigeminal neuralgia is characterized by excruciating paroxysms of pain in the lips, gums, cheek, or chin and, very rarely, in the distribution of the ophthalmic division of the fifth nerve. The pain seldom lasts more than a few seconds or a minute or two but may be so intense that the patient winces, hence the term tic. The paroxysms, experienced as single jabs or clusters, tend to recur frequently, both day and night, for several weeks at a time. They may occur spontaneously or be brought on with movements of affected areas by speaking, chewing, or smiling. Another characteristic feature is the presence of trigger zones, typically on the face, lips, or tongue, that provoke attacks; patients may report that tactile stimuli—e.g., washing the face, brushing the teeth, or exposure to a draft of air—generate excruciating pain. An essential feature of trigeminal neuralgia is that objective signs of sensory loss cannot be demonstrated on examination.

Trigeminal neuralgia is relatively common, with an estimated annual incidence of 4–8 per 100,000 individuals. Middle-aged and

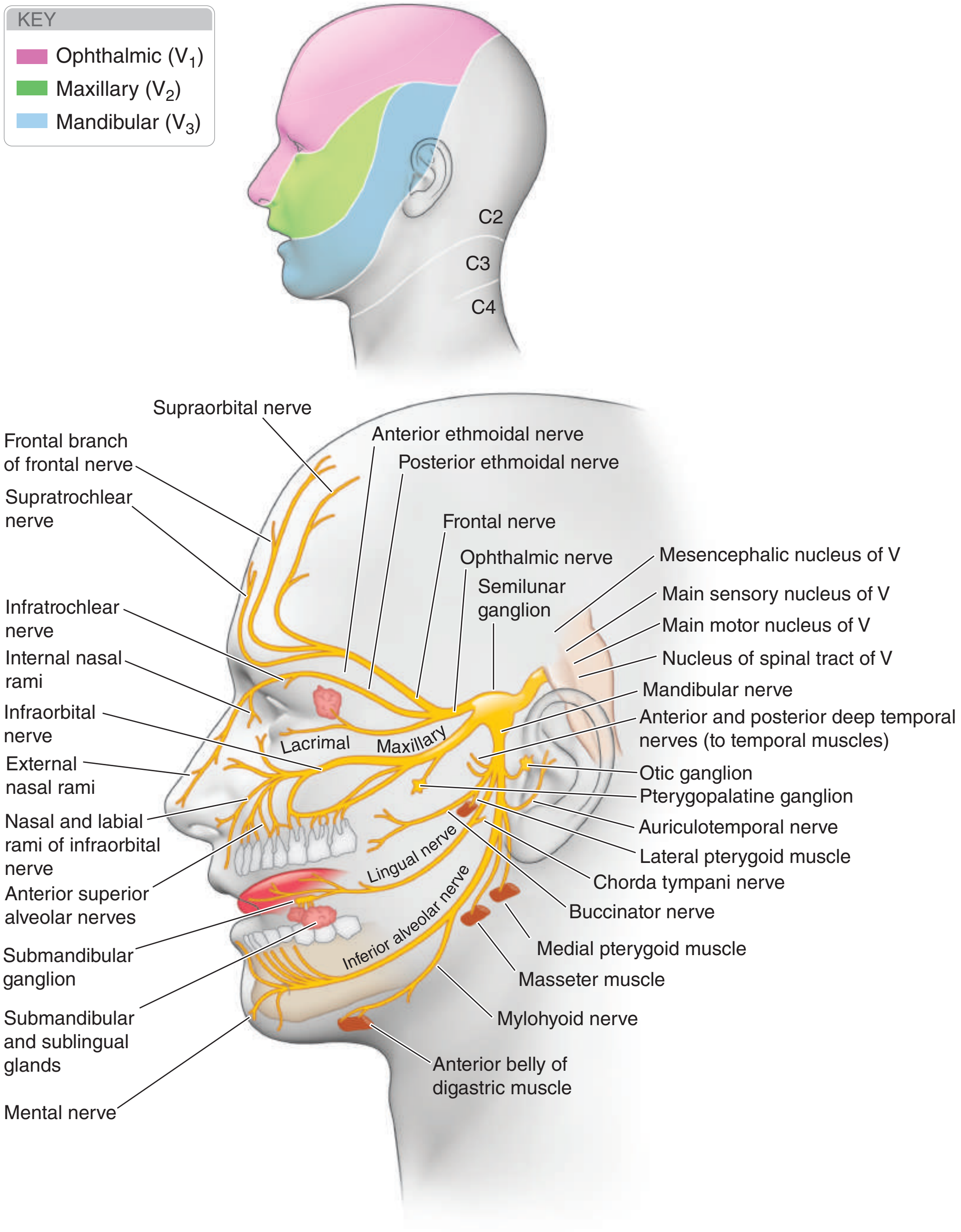
KEY
Ophthalmic (V_1)
Maxillary (V_2)
Mandibular (V_3)
C2
C3
C4
Supraorbital nerve
Frontal branch of frontal nerve
Supratrochlear nerve
Infratrochlear nerve
Internal nasal rami
Infraorbital nerve
External nasal rami
Nasal and labial rami of infraorbital nerve
Anterior superior alveolar nerves
Submandibular ganglion
Submandibular and sublingual glands
Mental nerve
Anterior ethmoidal nerve
Posterior ethmoidal nerve
Frontal nerve
Ophthalmic nerve
Semilunar ganglion
Mesencephalic nucleus of V
Main sensory nucleus of V
Main motor nucleus of V
Nucleus of spinal tract of V
Mandibular nerve
Anterior and posterior deep temporal nerves (to temporal muscles)
Otic ganglion
Pterygopalatine ganglion
Auriculotemporal nerve
Lateral pterygoid muscle
Chorda tympani nerve
Buccinator nerve
Medial pterygoid muscle
Masseter muscle
Mylohyoid nerve
Anterior belly of digastric muscle
Lacrimal
Maxillary
Lingual nerve
Inferior alveolar nerve

FIGURE 452-2 The trigeminal nerve and its branches and sensory distribution on the face. The three major sensory divisions of the trigeminal nerve consist of the ophthalmic, maxillary, and mandibular nerves. *(Reproduced with permission from Waxman SG: Clinical Neuroanatomy, 26th ed. New York, McGraw-Hill, 2009.)*

elderly persons are affected primarily, and ~60% of cases occur in women. Onset is typically sudden, and bouts tend to persist for weeks or months before remitting spontaneously. Remissions may be long-lasting, but in most patients, the disorder ultimately recurs.

Pathophysiology Symptoms result from ectopic generation of action potentials in pain-sensitive afferent fibers of the fifth cranial nerve root just before it enters the lateral surface of the pons. Compression or other pathology in the nerve leads to demyelination of large myelinated fibers that do not themselves carry pain sensation but become hyperexcitable and electrically coupled with smaller unmyelinated or poorly myelinated pain fibers in close proximity; this may explain why tactile stimuli, conveyed via the large myelinated fibers, can stimulate paroxysms of pain. Compression of the trigeminal nerve root by a blood vessel, most often the superior cerebellar artery or on occasion a tortuous vein, is believed to be the source of trigeminal neuralgia in most patients. In cases of vascular compression, age-related brain sagging and increased vascular thickness and tortuosity may explain the prevalence of trigeminal neuralgia in later life.

Differential Diagnosis Trigeminal neuralgia must be distinguished from other causes of face and head pain **(Chap. 17)** and from pain arising from diseases of the jaw, teeth, or sinuses. Pain from migraine or cluster headache tends to be deep-seated and steady, unlike the superficial stabbing quality of trigeminal neuralgia; rarely, cluster headache is associated with trigeminal neuralgia, a syndrome known as *cluster-tic*. Other rare, paroxysmal headache disorders such as short-lasting unilateral headache attacks with conjunctival injection and tearing, short-lasting unilateral headache attacks with cranial autonomic symptoms, and paroxysmal hemicrania **(Chap. 441)** are distinguished by the frequency and duration of attacks and associated autonomic symptoms. In temporal arteritis, superficial facial pain is present but is not typically shocklike, the patient frequently complains of myalgias and other systemic symptoms, and an elevated erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP) is usually present (Chap. 375). When trigeminal neuralgia develops in a young adult or is bilateral, MS is a key consideration, and in such cases, the cause is often a demyelinating plaque near the root entry zone of the fifth nerve in the pons, especially when there is evidence of superimposed facial sensory loss, which can be subtle. Cases that are secondary to mass lesions—such as aneurysms, neurofibromas, acoustic schwannomas, or meningiomas—usually produce objective signs of sensory loss in the trigeminal nerve distribution (trigeminal neuropathy, see below).

Laboratory Evaluation An ESR or CRP is indicated if temporal arteritis is suspected. Neuroimaging studies are often necessary to exclude secondary causes and help assess overlying vascular lesions in order to plan for decompression surgery.

TREATMENT

Trigeminal Neuralgia

Drug therapy with carbamazepine is effective in ~50–75% of patients. Carbamazepine should be started as a single daily dose of 100 mg taken with food and increased gradually (by 100 mg daily in divided doses every 1–2 days) until substantial (>50%) pain relief is achieved. Most patients require a maintenance dose of 200 mg four times daily. Doses >1200 mg daily provide no additional benefit. Dizziness, imbalance, sedation, and rare cases of agranulocytosis are the most important side effects of carbamazepine. If treatment is effective, it is usually continued for 1 month and then tapered as tolerated. Oxcarbazepine (300–1200 mg bid) is an alternative to carbamazepine that has less bone marrow toxicity and probably is equally efficacious. If these agents are not well tolerated or are ineffective, phenytoin (300–400 mg daily) is another option. Lamotrigine (400 mg daily), baclofen (10–20 mg tid), or topiramate (50 mg bid) may also be tried. Gabapentin, up to 3600 mg daily in divided doses, may occasionally provide relief.

If drug treatment fails, surgical therapy should be offered. The most widely used method is currently microvascular decompression to relieve pressure on the trigeminal nerve as it exits the pons. This procedure requires a suboccipital craniotomy. This procedure appears to have a >70% efficacy rate and a low rate of pain recurrence in responders; the response is better for classic ticlike symptoms than for nonlancinating facial pains. High-resolution magnetic resonance angiography is useful preoperatively to visualize the relationships between the fifth cranial nerve root and nearby blood vessels.

Gamma knife radiosurgery of the trigeminal nerve root is also used for treatment and results in complete pain relief, sometimes delayed in onset, in approximately one-half of patients and a low risk of persistent facial numbness; the response is sometimes long-lasting, but recurrent pain develops over 2–3 years in one-third of patients. Compared with surgical decompression, gamma knife surgery appears to be somewhat less effective but has few serious complications.

Another procedure, radiofrequency thermal rhizotomy, creates a heat lesion of the trigeminal ganglion or nerve. Short-term relief is experienced by >95% of patients; long-term studies indicate that pain recurs in up to one-third of treated patients. Postoperatively, partial numbness of the face is common, masseter (jaw) weakness may occur especially following bilateral procedures, and corneal denervation with secondary keratitis can follow rhizotomy for first-division trigeminal neuralgia. Percutaneous balloon compression of the trigeminal ganglion is an alternative approach performed under general anesthesia that results in similar rates of short- and long-term pain relief and is also commonly complicated by ipsilateral facial numbness.

TRIGEMINAL NEUROPATHY

A variety of diseases can affect the trigeminal nerve (Table 452-1). Most present with sensory loss on the face or with weakness of the jaw muscles. Deviation of the jaw on opening indicates weakness of the pterygoids on the side to which the jaw deviates. Some cases are due to Sjögren's syndrome or a collagen-vascular disease such as systemic lupus erythematosus, scleroderma, or mixed connective tissue disease. Among infectious causes, herpes zoster (acute or postherpetic) and leprosy should be considered. Tumors of the middle cranial fossa (meningiomas), of the trigeminal nerve (schwannomas), or of the base of the skull (metastatic tumors) may cause a combination of motor and sensory signs. Lesions in the cavernous sinus can affect the first and second divisions of the trigeminal nerve, and lesions of the superior orbital fissure can affect the first (ophthalmic) division; the accompanying corneal anesthesia increases the risk of ulceration (neurokeratitis).

TABLE 452-1 Trigeminal Nerve Disorders

Nuclear (Brainstem) Lesions
Multiple sclerosis
Stroke
Syringobulbia
Glioma
Lymphoma
Preganglionic Lesions
Acoustic neuroma
Meningioma
Metastasis
Chronic meningitis
Cavernous carotid aneurysm
Semilunar Ganglion Lesions
Trigeminal neuroma
Herpes zoster
Infection (spread from otitis media or mastoiditis)
Cavernous Sinus Lesions (see Table 452-2)
Peripheral Nerve Lesions
Tumor (e.g., nasopharyngeal carcinoma, squamous cell carcinoma, lymphoma)
Trauma
Guillain-Barré syndrome
Sjögren's syndrome
Collagen-vascular diseases
Sarcoidosis
Leprosy
Drugs (stilbamidine, trichloroethylene)
Idiopathic trigeminal neuropathy

Isolated sensory loss over the chin (mental neuropathy) can be the only manifestation of systemic malignancy. Rarely, an idiopathic form of trigeminal neuropathy is observed. It is characterized by numbness and paresthesias, sometimes bilaterally, with loss of sensation in the territory of the trigeminal nerve but without weakness of the jaw. Gradual recovery is the rule. Tonic spasm of the masticatory muscles, known as trismus, is symptomatic of tetanus (Chap. 157) or may occur in patients treated with phenothiazines.

FACIAL WEAKNESS

ANATOMIC CONSIDERATIONS

(Fig. 452-3) The seventh cranial nerve supplies all the muscles concerned with facial expression, as well as the stapedius, stylohyoid, and posterior belly of the digastric. The sensory and parasympathetic components (the nervus intermedius) convey taste sensation from the anterior two-thirds of the tongue, cutaneous impulses from the anterior wall of the external auditory canal, and preganglionic parasympathetic signals to the pterygopalatine and submandibular ganglia, stimulating lacrimation, rhinorrhea, and salivation. The cell bodies of pseudounipolar sensory neurons lie in the geniculate ganglion. The motor nucleus of the seventh nerve lies anterior and lateral to the abducens nucleus. After leaving the pons, the seventh nerve enters the internal auditory meatus with the acoustic nerve. The nerve continues its course

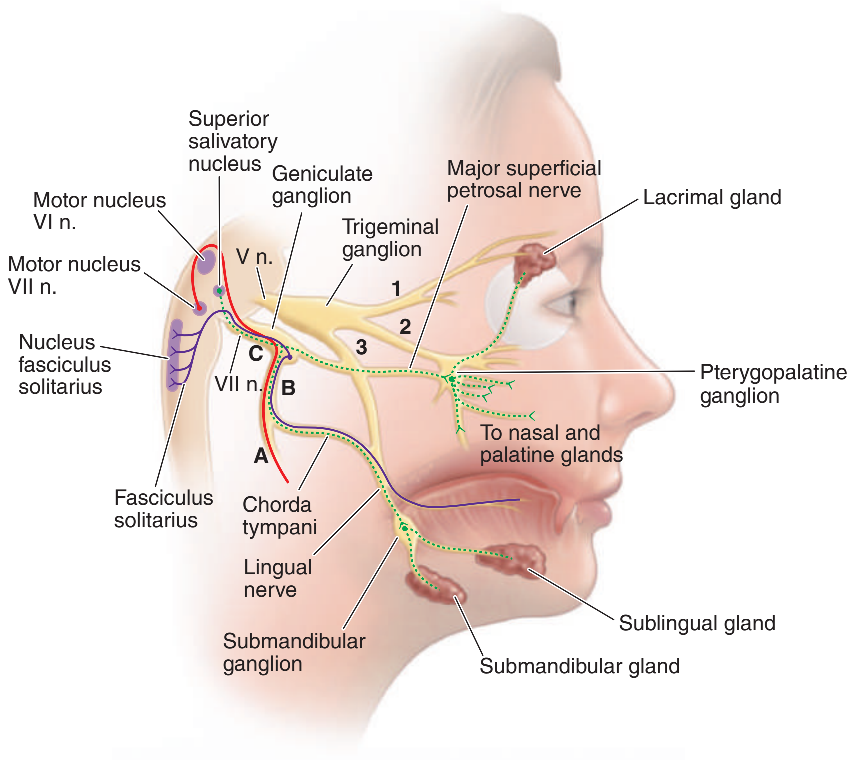

FIGURE 452-3 The facial nerve. A, B, and C denote lesions of the facial nerve at the stylomastoid foramen, distal and proximal to the geniculate ganglion, respectively. *Green lines* indicate the parasympathetic fibers, *red line* indicates motor fibers, and *purple lines* indicate visceral afferent fibers (taste). *(Reproduced with permission from MB Carpenter: Core Text of Neuroanatomy, 2nd ed. Williams & Wilkins, 1978.)*

in its own bony channel, the facial canal, and exits from the skull via the stylomastoid foramen. It then passes through the parotid gland and subdivides to supply the facial muscles.

A complete interruption of the facial nerve at the stylomastoid foramen paralyzes all muscles of facial expression. The corner of the mouth droops, the creases and skinfolds are effaced, the forehead is unfurrowed, and the eyelids will not close. Upon attempted closure of the lids, the eye on the paralyzed side rolls upward (Bell's phenomenon). The lower lid sags and falls away from the conjunctiva, permitting tears to spill over the cheek. Food collects between the teeth and lips, and saliva may dribble from the corner of the mouth. The patient complains of a heaviness or numbness in the face, but sensory loss is rarely demonstrable and taste is intact.

If the lesion is in the middle-ear portion, taste is lost over the anterior two-thirds of the tongue on the same side. If the nerve to the stapedius is interrupted, there is hyperacusis (sensitivity to loud sounds). Lesions in the internal auditory meatus may affect the adjacent auditory and vestibular nerves, causing deafness, tinnitus, or dizziness. Intrapontine lesions that paralyze the face usually affect the abducens nucleus as well, and often the corticospinal and sensory tracts.

If the peripheral facial paralysis has existed for some time and recovery of motor function is incomplete, a continuous diffuse contraction of facial muscles may appear. The palpebral fissure becomes narrowed, and the nasolabial fold deepens. Facial spasms, initiated by movements of the face, may develop (hemifacial spasm). Anomalous regeneration of seventh nerve fibers may result in other troublesome phenomena. If fibers originally connected with the orbicularis oculi come to innervate the orbicularis oris, closure of the lids may cause a retraction of the mouth (synkinesis), or if parasympathetic fibers originally connected with salivary glands later innervate the lacrimal gland, anomalous tearing ("crocodile tears") may occur with eating. Another facial synkinesia is triggered by jaw opening, causing closure of the eyelids on the side of the facial palsy (jaw-winking).

■ BELL'S PALSY

The most common form of facial paralysis is Bell's palsy. The annual incidence of this idiopathic disorder is ~25 per 100,000 annually, or about 1 in 60 persons in a lifetime. Risk factors include pregnancy and diabetes mellitus.

Clinical Manifestations The onset of Bell's palsy is fairly abrupt, with maximal weakness being attained by 48 h as a general rule. Pain behind the ear may precede the paralysis for a day or two. Taste sensation may be lost unilaterally, and hyperacusis may be present. In some cases, there is mild cerebrospinal fluid lymphocytosis. MRI may reveal swelling and uniform enhancement of the geniculate ganglion and facial nerve and, in some cases, entrapment of the swollen nerve in the temporal bone. Approximately 80% of patients recover within a few weeks or months. Electromyography may be of some prognostic value; evidence of denervation after 10 days indicates there has been axonal degeneration, that there will be a long delay (3 months as a rule) before regeneration occurs, and that it may be incomplete. The presence of incomplete paralysis in the first week is the most favorable prognostic sign. Recurrences are reported in ~7% of cases.

Pathophysiology In acute Bell's palsy, there is inflammation of the facial nerve with mononuclear cells, consistent with an infectious or immune cause. Herpes simplex virus (HSV) type 1 DNA was frequently detected in endoneurial fluid and posterior auricular muscle, suggesting that a reactivation of this virus in the geniculate ganglion may be responsible for most cases. Reactivation of varicella-zoster virus is associated with Bell's palsy in up to one-third of cases and may represent the second most frequent cause. A variety of other viruses including SARS-CoV-2 have also been implicated less commonly, and Bell's palsy can be observed in the setting of human immunodeficiency virus (HIV) seroconversion.

Differential Diagnosis There are many other causes of acute facial palsy that must be considered in the differential diagnosis of Bell's palsy. Lyme disease can cause unilateral or bilateral facial palsies; in endemic areas, ≥10% of cases of facial palsy are likely due to infection with *Borrelia burgdorferi* **(Chap. 191)**. Ramsay Hunt syndrome, caused by reactivation of herpes zoster in the geniculate ganglion, consists of a severe facial palsy associated with a vesicular eruption in the external auditory canal and sometimes in the pharynx and other parts of the

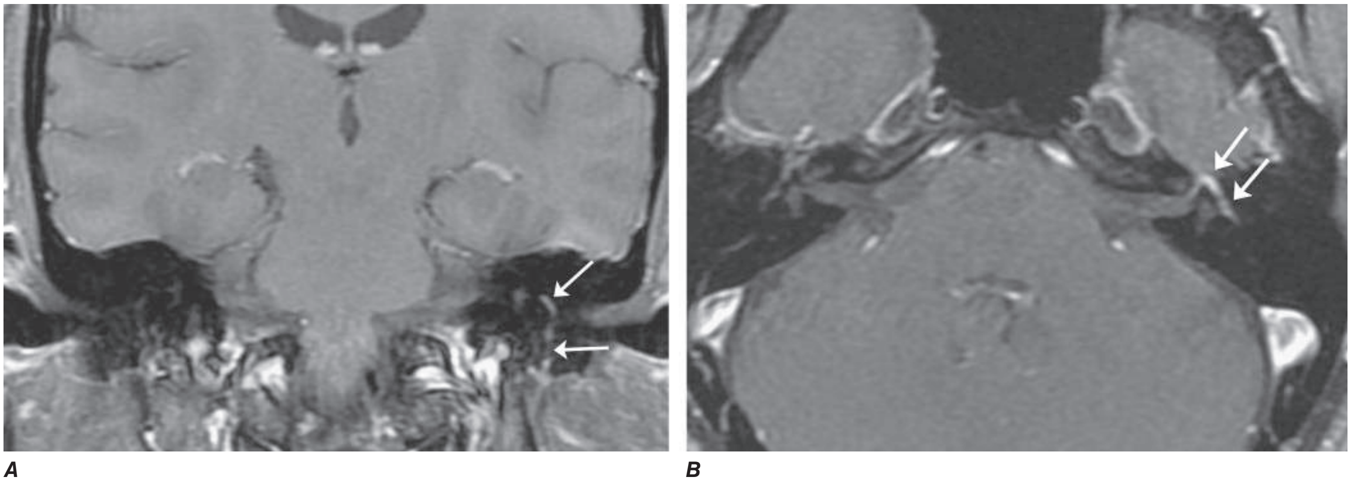

FIGURE 452-4 Axial and coronal T1-weighted images after gadolinium with fat suppression demonstrate diffuse smooth linear enhancement of the left facial nerve, involving the genu, tympanic, and mastoid segments within the temporal bone (*arrows*), without evidence of mass lesion. Although highly suggestive of Bell's palsy, similar findings may be seen with other etiologies such as Lyme disease, sarcoidosis, and perineural malignant spread.

cranial integument; often the eighth cranial nerve is affected as well. Facial palsy that is often bilateral occurs in sarcoidosis **(Chap. 379)** and in Guillain-Barré syndrome **(Chap. 458)**. Leprosy frequently involves the facial nerve, and facial neuropathy may also occur in diabetes mellitus, connective tissue diseases including Sjögren's syndrome, and amyloidosis. The rare Melkersson-Rosenthal syndrome consists of recurrent facial paralysis; recurrent—and eventually permanent—facial (particularly labial) edema; and, less constantly, plication of the tongue. Its cause is unknown. Acoustic neuromas frequently involve the facial nerve by local compression. Infarcts, demyelinating lesions of MS, and tumors are common pontine lesions that interrupt the facial nerve fibers; other signs of brainstem involvement are usually present. Tumors that invade the temporal bone (carotid body, cholesteatoma, dermoid) may produce a facial palsy, but the onset is insidious and the course progressive. Facial palsy after temporal bone fracture can present acutely or after a delay of several days; blunt head injury without temporal bone fracture may also trigger facial palsy.

All these forms of nuclear or peripheral facial palsy must be distinguished from the supranuclear type. In the latter, the frontalis and orbicularis oculi muscles of the forehead are involved less than those of the lower part of the face, since the upper facial muscles are innervated by corticobulbar pathways from both motor cortices, whereas the lower facial muscles are innervated only by the opposite hemisphere. In supranuclear lesions, there may be a dissociation of emotional and voluntary facial movements, and often some degree of paralysis of the arm and leg or an aphasia (in dominant-hemisphere lesions) is present.

Laboratory Evaluation The diagnosis of Bell's palsy can usually be made clinically in patients with (1) a typical presentation, (2) no risk factors or preexisting symptoms for other causes of facial paralysis, (3) absence of cutaneous lesions of herpes zoster in the external ear canal, and (4) a normal neurologic examination apart from the facial nerve. Particular attention to the eighth cranial nerve, which courses near to the facial nerve in the pontomedullary junction and in the temporal bone, and to other cranial nerves is essential. In atypical or uncertain cases, an ESR or CRP, testing for diabetes mellitus, a Lyme titer, HIV serologies, angiotensin-converting enzyme and chest imaging studies for possible sarcoidosis, a lumbar puncture for possible Guillain-Barré syndrome, or MRI scanning may be indicated. MRI often shows swelling and enhancement of the facial nerve in idiopathic Bell's palsy **(Fig. 452-4)**.

TREATMENT

Bell's Palsy

Symptomatic measures include (1) the use of paper tape to depress the upper eyelid during sleep and prevent corneal drying, (2) artificial tears, and (3) massage of the weakened muscles. A course of glucocorticoids, given as prednisone 60–80 mg daily during the first 5 days and then tapered over the next 5 days, modestly shortens recovery and improves the functional outcome. Although large and well-controlled randomized trials found no added benefit of the antiviral agents valacyclovir (1000 mg daily for 5–7 days) or acyclovir (400 mg five times daily for 10 days) compared to glucocorticoids alone, either of these agents should be used if vesicular lesions are observed in the palate or external auditor canal. For patients with permanent paralysis from Bell's palsy, a number of cosmetic surgical procedures have been used to restore a relatively symmetric appearance to the face.

OTHER MOTOR DISORDERS OF THE FACE

Hemifacial spasm consists of painless irregular involuntary contractions on one side of the face. Most cases appear related to vascular compression of the exiting facial nerve in the pons. Other cases develop as a sequela to Bell's palsy or are secondary to compression and/or demyelination of the nerve by tumor, infection, or MS. Local injections of botulinum toxin into affected muscles can relieve spasms for 3–4 months, and the injections can be repeated. Refractory cases due to vascular compression usually respond to surgical decompression of the facial nerve. Anecdotal reports describe success using carbamazepine, gabapentin, or baclofen. Blepharospasm is an involuntary recurrent spasm of both eyelids that usually occurs in elderly persons as an isolated phenomenon or with varying degrees of spasm of other facial muscles. Severe, persistent cases of blepharospasm can be treated by local injection of botulinum toxin into the orbicularis oculi. Clonazepam, baclofen, and trihexyphenidyl have also been used to treat this disorder. Facial myokymia refers to a fine rippling activity of the facial muscles; it may be caused by MS or follow Guillain-Barré syndrome **(Chap. 458)**.

OTHER CRANIAL NERVE DISORDERS

GLOSSOPHARYNGEAL NEURALGIA

The ninth cranial (glossopharyngeal) nerve **(Fig. 452-5)** conveys somatic sensation from the pharynx, middle ear, tympanic membrane, eustachian tube, and posterior third of the tongue to the spinal trigeminal nucleus. It also relays taste from the posterior third of the tongue and information about blood pressure from baroreceptors in the carotid sinus to the nucleus solitarius, which also serves as the sensory nucleus for the vagus nerve. Motor function originates in the nucleus ambiguus and is limited to the stylopharyngeus muscle. Parasympathetic fibers from the medullary inferior salivatory nucleus synapse in the otic ganglion with postganglionic fibers that innervate the parotid gland. Glossopharyngeal neuralgia resembles trigeminal neuralgia in many respects but is much less common. Sometimes it involves portions of the tenth (vagus) nerve. The pain is intense and paroxysmal; it originates on one side of the throat, approximately in the tonsillar fossa.

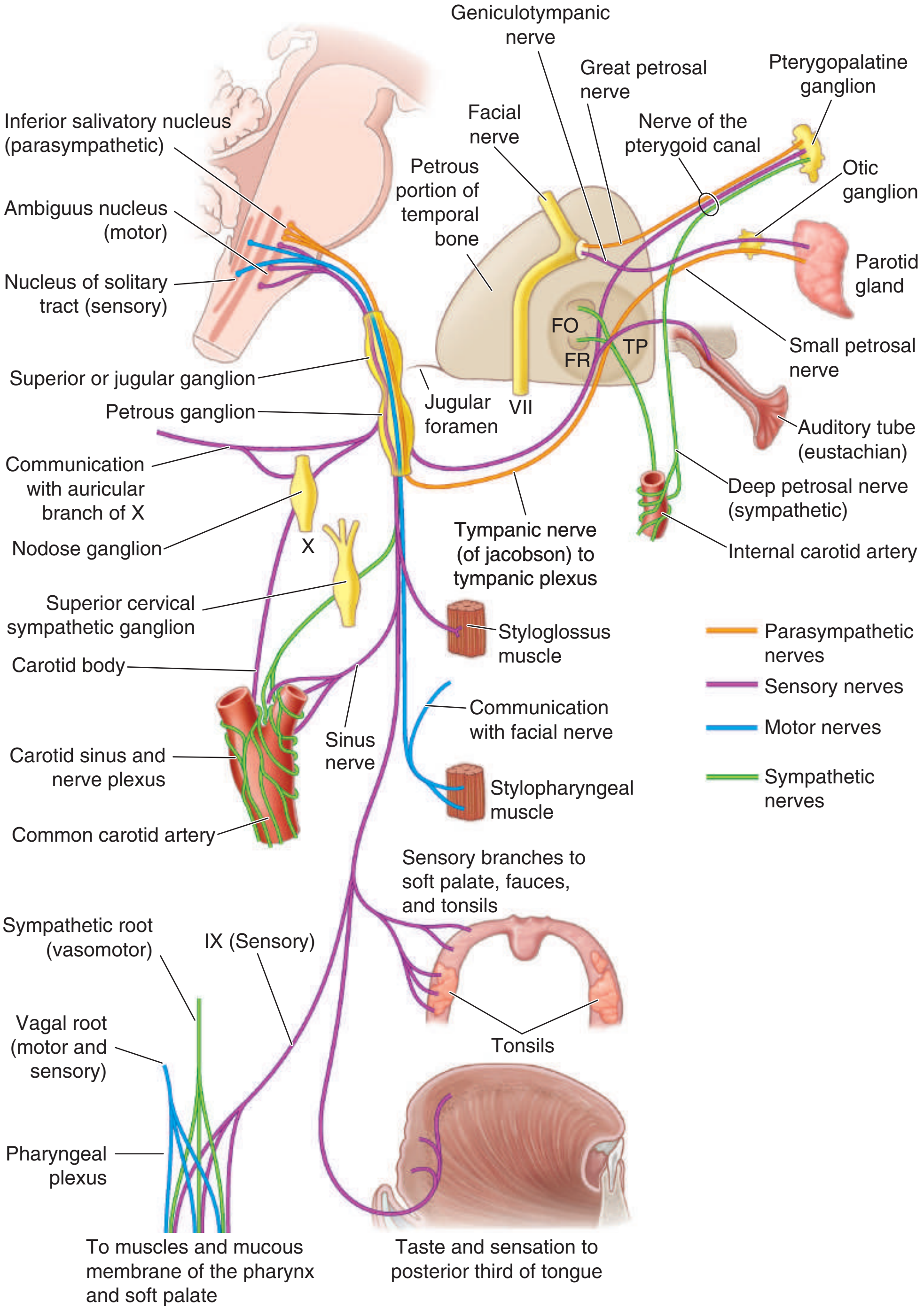
Geniculotympanic nerve
Great petrosal nerve
Pteryogopalatine ganglion
Inferior salivatory nucleus (parasympathetic)
Facial nerve
Nerve of the pterygoid canal
Petrous portion of temporal bone
Otic ganglion
Ambiguus nucleus (motor)
Parotid gland
Nucleus of solitary tract (sensory)
FO
TP
FR
Small petrosal nerve
Superior or jugular ganglion
Jugular foramen
VII
Petrous ganglion
Auditory tube (eustachian)
Communication with auricular branch of X
Deep petrosal nerve (sympathetic)
X
Nodose ganglion
Tympanic nerve (of jacobson) to tympanic plexus
Internal carotid artery
Superior cervical sympathetic ganglion
Styloglossus muscle
Parasympathetic nerves
Sensory nerves
Carotid body
Motor nerves
Communication with facial nerve
Carotid sinus and nerve plexus
Sinus nerve
Sympathetic nerves
Stylopharyngeal muscle
Common carotid artery
Sensory branches to soft palate, fauces, and tonsils
Sympathetic root (vasomotor)
IX (Sensory)
Tonsils
Vagal root (motor and sensory)
Pharyngeal plexus
To muscles and mucous membrane of the pharynx and soft palate
Taste and sensation to posterior third of tongue

FIGURE 452-5 The ninth cranial (glossopharyngeal) nerve. FO, foramen ovale; FR, foramen rotundum; TP, tympanum plexus. *(Reproduced with permission from SG Waxman: Clinical Neuroanatomy, 29th ed. New York, McGraw Hill, 2020.)*

In some cases, the pain is localized in the ear or may radiate from the throat to the ear because of involvement of the tympanic branch of the glossopharyngeal nerve. Spasms of pain may be initiated by swallowing or coughing. There is no demonstrable motor or sensory deficit. Cardiac symptoms—bradycardia or asystole, hypotension, and fainting—have been reported. Glossopharyngeal neuralgia can result from vascular compression, MS, or tumors, but many cases are idiopathic. Medical therapy is similar to that for trigeminal neuralgia, and carbamazepine is generally the first choice. If drug therapy is unsuccessful, surgical procedures—including microvascular decompression if vascular compression is evident—or rhizotomy of glossopharyngeal and vagal fibers in the jugular bulb is frequently successful.

■ DYSPHAGIA AND DYSPHONIA

The tenth cranial (vagus) nerve **(Fig. 452-6)** carries somatic sensation from the posterior aspect of the external auditory canal, laryngopharynx, superior larynx, and meninges of the posterior fossa to the spinal trigeminal nucleus, as well as taste from the epiglottis and pharynx and visceral sensation from chemoreceptors and baroreceptors in the aortic arch, heart, and gastrointestinal tract to the splenic flexure to the nucleus solitarius. The motor part originates in the nucleus ambiguous and innervates most muscles of the oropharynx and soft palate as well as all laryngeal muscles. Parasympathetic fibers originate in the dorsal motor nucleus of the vagus nerve and decrease the heart rate through action at the sinoatrial and atrioventricular nodes; others promote peristalsis and secretion of the alimentary tract from the esophagus to the splenic flexure. When the intracranial portion of one vagus (tenth cranial) nerve is interrupted, the soft palate droops ipsilaterally and does not rise in phonation. There is loss of the gag reflex on the affected side, as well as of the "curtain movement" of the lateral wall of the pharynx, whereby the faucial pillars move medially as the palate rises in saying "ah." The voice is hoarse and slightly

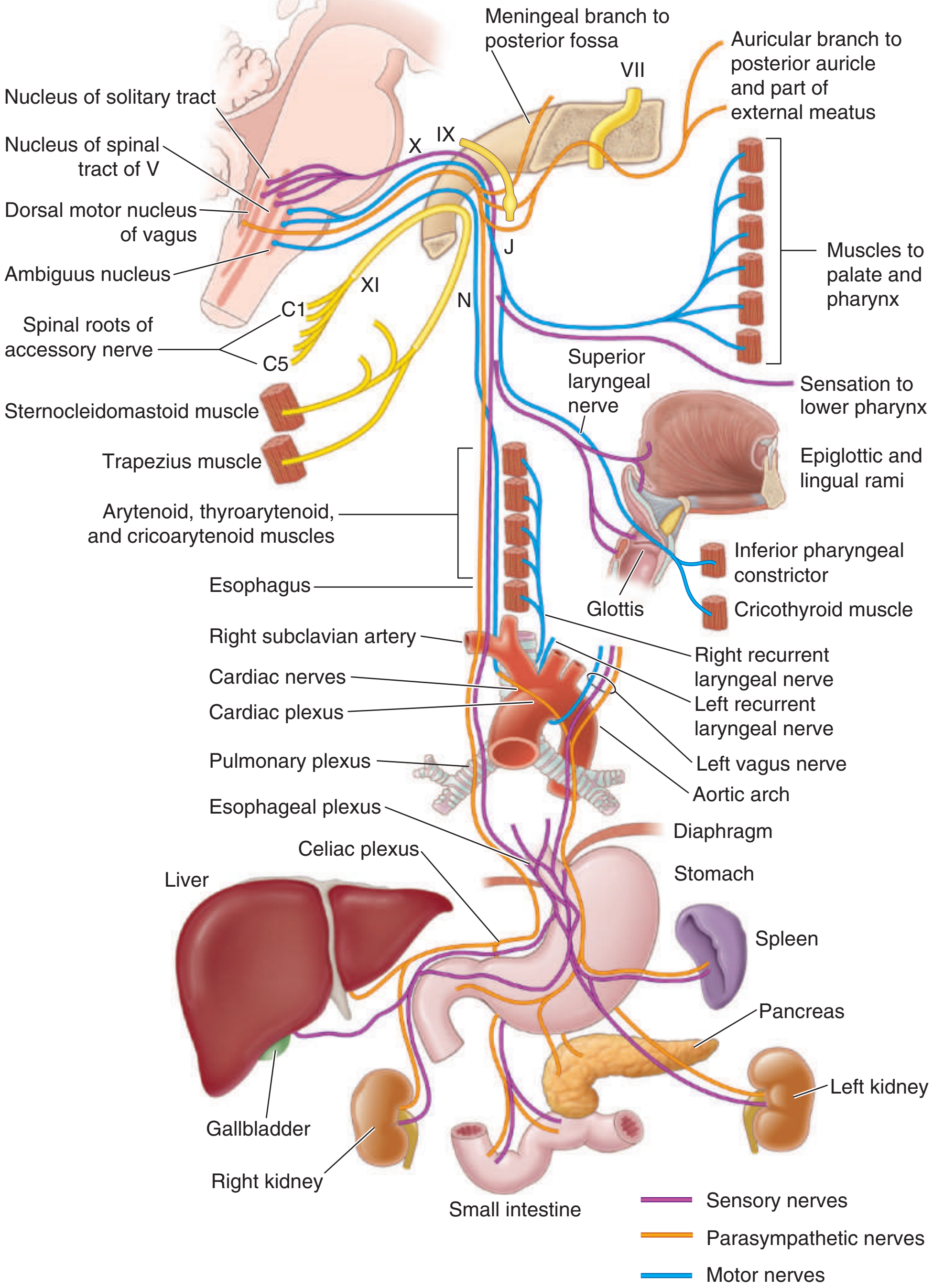
Meningeal branch to posterior fossa
VII
Auricular branch to posterior auricle and part of external meatus
Nucleus of solitary tract
Nucleus of spinal tract of V
X
IX
Dorsal motor nucleus of vagus
J
Muscles to palate and pharynx
Ambiguus nucleus
XI
C1
N
Spinal roots of accessory nerve
C5
Superior laryngeal nerve
Sensation to lower pharynx
Sternocleidomastoid muscle
Epiglottic and lingual rami
Trapezius muscle
Arytenoid, thyroarytenoid, and cricoarytenoid muscles
Inferior pharyngeal constrictor
Esophagus
Glottis
Cricothyroid muscle
Right subclavian artery
Right recurrent laryngeal nerve
Cardiac nerves
Left recurrent laryngeal nerve
Cardiac plexus
Left vagus nerve
Pulmonary plexus
Aortic arch
Esophageal plexus
Diaphragm
Celiac plexus
Stomach
Liver
Spleen
Pancreas
Left kidney
Gallbladder
Right kidney
Small intestine
Sensory nerves
Parasympathetic nerves
Motor nerves

FIGURE 452-6 The vagus nerve. J, jugular (superior) ganglion; N, nodose (inferior) ganglion. *(Reproduced with permission from SG Waxman: Clinical Neuroanatomy, 29th ed. New York, McGraw Hill, 2020.)*

nasal, and the vocal cord lies immobile midway between abduction and adduction. Loss of sensation at the external auditory meatus and the posterior pinna may also be present.

The vagus nerve may be involved at the meningeal level by neoplastic and infectious processes and within the medulla by tumors, vascular lesions (e.g., the lateral medullary syndrome), and motor neuron disease. The nerve may be involved by infection with varicella-zoster virus. Injury to the vagus nerve in the carotid sheath can occur with carotid dissection or following endarterectomy. The pharyngeal branches of both vagal nerves may be affected in diphtheria; the voice has a nasal quality, and regurgitation of liquids through the nose occurs during swallowing. Polymyositis and dermatomyositis, which cause hoarseness and dysphagia by direct involvement of laryngeal and pharyngeal muscles, may be confused with diseases of the vagus nerves. Dysphagia is also a symptom in some patients with myotonic dystrophy. **Nonneurologic causes of dysphagia are discussed in Chap. 47.**

The recurrent laryngeal nerves, especially the left, are most often damaged as a result of intrathoracic disease. Aneurysm of the aortic arch, an enlarged left atrium, and tumors of the mediastinum and bronchi are much more frequent causes of an isolated vocal cord palsy than are intracranial disorders. However, a substantial number of cases of recurrent laryngeal palsy remain idiopathic.

When confronted with a case of laryngeal palsy, the physician must attempt to determine the site of the lesion. If it is intramedullary, there are usually other signs, such as ipsilateral cerebellar dysfunction, loss of pain and temperature sensation over the ipsilateral face and contralateral arm and leg, and an ipsilateral Horner's syndrome. If the lesion is extramedullary, the glossopharyngeal and spinal accessory nerves are frequently involved (jugular foramen syndrome). If it is extracranial in the posterior laterocondylar or retroparotid space, there may be a combination of ninth, tenth, eleventh, and twelfth cranial nerve palsies and Horner's syndrome **(Table 452-2)**. If there is no sensory loss over

TABLE 452-2 Cranial Nerve Syndromes

SITE	CRANIAL NERVES	USUAL CAUSE
Orbital apex	II, III, IV, first division V, VI	Invasive fungal infections, amyloidosis, granulomatous disease
Sphenoid fissure (superior orbital)	III, IV, first division V, VI	Invasive tumors of sphenoid bone; aneurysms
Lateral wall of cavernous sinus	III, IV, first division V, VI, often with proptosis	Infection, thrombosis, aneurysm or fistula of cavernous sinus; invasive tumors from sinuses and sella turcica; benign granuloma responsive to glucocorticoids
Retrosphenoid space	II, III, IV, V, VI	Large tumors of middle cranial fossa
Apex of petrous bone	V, VI	Petrositis; tumors of petrous bone
Internal auditory meatus	VII, VIII	Tumors of petrous bone (dermoids, etc.); infectious processes; acoustic neuroma
Pontocerebellar angle	V, VI, VII, VIII, and sometimes IX	Acoustic neuroma; meningioma
Jugular foramen	IX, X, XI	Tumors and aneurysms
Posterior laterocondylar space	IX, X, XI, XII	Tumors of parotid gland and carotid body and metastatic tumors
Posterior retroparotid space	IX, X, XI, XII, and Horner's syndrome	Tumors of parotid gland, carotid body, lymph nodes; metastatic tumor; tuberculous adenitis

the palate and pharynx and no palatal weakness or dysphagia, the lesion is below the origin of the pharyngeal branches, which leave the vagus nerve high in the cervical region; the usual site of disease is then the mediastinum.

■ NECK WEAKNESS

The eleventh cranial nerve (spinal accessory) is a pure motor nerve arising from the nucleus ambiguus and the ventral horn of the spinal cord from C1–C6. The nerve travels superiorly through the foramen magnum and exits through the jugular foramen to innervate the ipsilateral sternocleidomastoid and trapezius muscles. Isolated involvement of the accessory (eleventh cranial) nerve can occur anywhere along its route, resulting in partial or complete paralysis of the sternocleidomastoid and trapezius muscles. Spinal accessory nerve palsy does not result in significant neck weakness because several other muscles also turn the head and flex the neck; therefore, detection of accessory nerve injury relies on palpating the absence of sternocleidomastoid contraction during head turning. Similarly, shoulder shrug is only slightly impacted by trapezius weakness, although the affected shoulder is lower at rest, scapular winging occurs, and the arm cannot abduct beyond 90°. Isolated spinal accessory nerve palsy is often iatrogenic due to neck surgery or jugular vein cannulation, or traumatic. An idiopathic form of accessory neuropathy, akin to Bell's palsy, has been described, and it may be recurrent in some cases. Most but not all patients recover.

■ TONGUE PARALYSIS

The twelfth cranial nerve (hypoglossal) supplies the ipsilateral muscles of the tongue. Nerve lesions cause the tongue to deviate toward the ipsilateral side during protrusion due to ipsilateral genioglossus weakness, in addition to weakness of tongue movements toward the affected side to weakness of ipsilateral intrinsic tongue musculature. Atrophy and fasciculations of the tongue develop weeks to months after interruption of the nerve. The nucleus of the nerve or its fibers of exit may be involved by intramedullary lesions such as tumor, poliomyelitis, or most often motor neuron disease. Lesions of the basal meninges and the occipital bones (platybasia, invagination of occipital condyles, Paget's disease) may compress the nerve in its extramedullary course or as it exits the skull in the hypoglossal canal. Isolated lesions of unknown cause can occur.

MULTIPLE CRANIAL NERVE PALSIES

When multiple cranial nerves are affected by a disease process, the clinical approach begins by determining whether the pathology lies within or outside of the brainstem. Lesions that lie on the surface of the brainstem are characterized by involvement of adjacent cranial nerves (often occurring in succession) and late and rather slight involvement of the long sensory and motor pathways and segmental structures lying within the brainstem. The opposite is true of primary lesions within the brainstem. Extramedullary lesions are more likely to cause bone erosion or enlargement of the foramens of exit of cranial nerves. By contrast, intramedullary lesions involving cranial nerves often produce a crossed sensory or motor paralysis (cranial nerve signs on one side of the body and tract signs on the opposite side).

Involvement of multiple cranial nerves outside the brainstem is frequently the result of trauma, localized infections including varicella-zoster virus, infectious and noninfectious (especially carcinomatous) causes of meningitis **(Chaps. 143 and 144)**, granulomatous diseases such as granulomatosis with polyangiitis **(Chap. 375)**, Behçet's disease **(Chap. 376)**, vascular disorders including those associated with diabetes, enlarging aneurysms, or locally infiltrating tumors. Among the tumors, nasopharyngeal cancers, lymphomas, neurofibromas, meningiomas, chordomas, cholesteatomas, carcinomas, and sarcomas have all been observed to involve a succession of lower cranial nerves. Owing to their anatomic relationships, the multiple cranial nerve palsies form a number of distinctive syndromes, listed in Table 452-2. Sarcoidosis **(Chap. 379)** is the cause of some cases of multiple cranial neuropathy; tuberculosis, the Chiari malformation, platybasia, and basilar invagination of the skull are additional causes.

Cavernous sinus syndrome **(Fig. 452-7)** is a distinctive and frequently life-threatening disorder. It often presents as orbital or facial pain; orbital swelling, chemosis due to occlusion of the ophthalmic veins; fever; oculomotor neuropathy affecting the third, fourth, and sixth cranial nerves; and trigeminal neuropathy affecting the ophthalmic (V1) and occasionally the maxillary (V2) divisions of the trigeminal nerve. Cavernous sinus thrombosis, often secondary to infection from orbital cellulitis (frequently *Staphylococcus aureus*), a cutaneous source on the face, or sinusitis (especially with mucormycosis in diabetic patients), is the most frequent cause; other etiologies include aneurysm of the carotid artery, a carotid-cavernous fistula (orbital bruit may be present), meningioma, nasopharyngeal carcinoma, other tumors, or an idiopathic granulomatous disorder (Tolosa-Hunt syndrome, discussed below). The two cavernous sinuses directly communicate via intercavernous channels; thus, involvement on one side may extend to become bilateral. Early diagnosis is essential, especially when due to infection, and treatment depends on the underlying etiology.

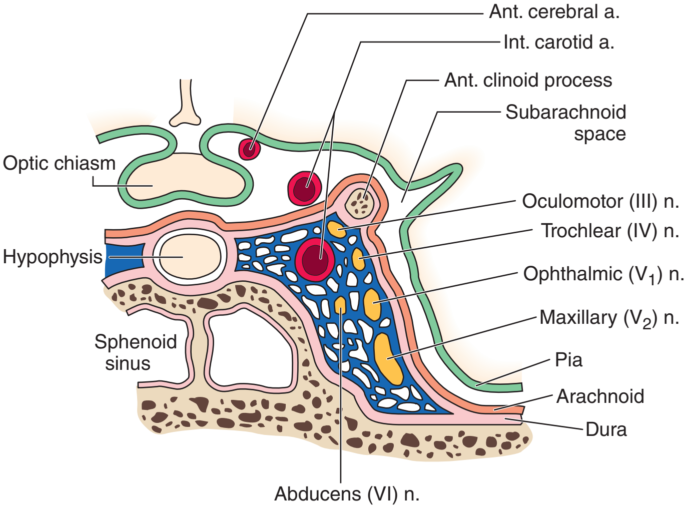

FIGURE 452-7 Anatomy of the cavernous sinus in coronal section, illustrating the location of the cranial nerves in relation to the vascular sinus, internal carotid artery (which loops anteriorly to the section), and surrounding structures.

In infectious cases, prompt administration of broad-spectrum antibiotics, drainage of any abscess cavities, and identification of the offending organism are essential. Anticoagulant therapy may benefit cases of primary thrombosis. Repair or occlusion of the carotid artery may be required for treatment of fistulas or aneurysms.

Tolosa-Hunt syndrome is characterized by onset, over days or a few weeks, of severe orbital pain with variable ophthalmoparesis and numbness of the upper face (V1 and V2 divisions of the trigeminal nerve) and by a dramatic clinical response to glucocorticoids. Although distinctive, the presentation can be mimicked by numerous other conditions that involve the cavernous sinus and orbit, including sarcoid **(Chap. 379)**, vasculitis **(Chap. 375)**, IgG4-related disease **(Chap. 380)**, lymphoma, and fungal infections. MRI can suggest the presence of granulomatous inflammation, but biopsy is sometimes required for diagnosis. A dramatic improvement in pain is usually evident within a few days; oral prednisone (60 mg daily) is usually continued for 2 weeks and then gradually tapered over a month, or longer if pain recurs. Occasionally an immunosuppressive medication, such as azathioprine or methotrexate, needs to be added to maintain an initial response to glucocorticoids.

Lesions in the superior orbital fissure and orbital apex cause more prominent vision loss than those in the cavernous sinus due to compression of the optic nerve; the second branch of the trigeminal nerve is usually spared. The cause is often an invasive fungal infection, frequently due to osseous erosion through the wall of the maxillary, sphenoid, or ethmoid sinuses. Infiltrative processes such as amyloidosis, granulomatosis with polyarteritis, an idiopathic granulomatous inflammation similar to Tolosa-Hunt, and IgG4-related disease are additional causes. Biopsy is often necessary for diagnosis.

As noted above, Guillain-Barré syndrome commonly affects the facial nerves bilaterally. In the Fisher variant of Guillain-Barré syndrome, oculomotor paresis occurs with ataxia and areflexia in the limbs **(Chap. 458)**. Wernicke's encephalopathy can cause a severe ophthalmoplegia combined with other brainstem signs **(Chap. 318)**.

Progressive bulbar palsy is a slowly progressive purely motor disorder affecting multiple cranial nerve nuclei. Weakness of the face, jaw, pharynx, neck, and tongue is usually present accompanied by atrophy and fasciculations. It is a form of motor neuron disease **(Chap. 448)**. Pure motor syndromes without atrophy raise the question of myasthenia gravis **(Chap. 459)**, and when rapidly evolving, Guillain-Barré syndrome, diphtheria, and poliomyelitis are additional considerations.

Glossopharyngeal neuropathy in conjunction with vagus and accessory nerve palsies may occur with herpes zoster infection or with a tumor or aneurysm in the posterior fossa or in the jugular foramen, through which all three nerves exit the skull. Hoarseness due to vocal cord paralysis, some difficulty in swallowing, deviation of the soft palate to the intact side, anesthesia of the posterior wall of the pharynx, and weakness of the upper part of the trapezius and sternocleidomastoid muscles make up jugular foramen syndrome.

Paralysis of the vagus and hypoglossal nerves (Tapia syndrome) can rarely follow endotracheal intubation and has been reported during the COVID-19 pandemic; symptoms consist of dysphonia and tongue deviation and usually resolve within a few months.

An idiopathic form of multiple cranial nerve involvement on one or both sides of the face is occasionally seen. The syndrome consists of a subacute onset of boring facial pain, followed by paralysis of motor cranial nerves. The clinical features overlap those of Tolosa-Hunt syndrome and appear to be due to idiopathic inflammation of the dura mater, which may be visualized by MRI. The syndrome is usually responsive to glucocorticoids.

■ FURTHER READING

Abad S et al: IgG4-related disease in patients with idiopathic orbital inflammation syndrome: Data from the French SIOI prospective cohort. Acta Ophthalmol 97:e648, 2019.

Bendsten L et al: European Academy of Neurology guideline on trigeminal neuralgia. Eur J Neurol 26:831, 2019.

Gagyor I et al: Antiviral treatment of Bell's palsy (idiopathic facial paralysis). Cochrane Database Syst Rev 9:CD001869, 2019.

Gutierrez S et al: Lower cranial nerve syndromes: A review. Neurosurg Rev 44:1345, 2020.

Madhok VB et al: Corticosteroids for Bell's palsy (idiopathic facial paralysis). Cochrane Database Syst Rev 7:CD001942, 2016.

Mullen E et al: Reappraising the Tolosa-Hunt syndrome diagnostic criteria: A case series. Headache 60:259, 2020.

Rafati A et al: Association of SARS-CoV-2 vaccination or infection with Bell palsy: A systematic review and meta-analysis. JAMA Otolaryngol Head Neck Surg 149:493, 2023.

453

Diseases of the Spinal Cord

Stephen L. Hauser

Diseases of the spinal cord are frequently devastating. They produce quadriplegia, paraplegia, and sensory deficits far beyond the damage they would inflict elsewhere in the nervous system because the spinal cord contains, in a small cross-sectional area, almost the entire motor output and sensory input of the trunk and limbs. Many spinal cord diseases are reversible if recognized and treated at an early stage **(Table 453-1)**; thus, they are among the most critical of neurologic emergencies. Proper management requires the efficient use of diagnostic procedures, guided by knowledge of the anatomy and clinical features of spinal cord diseases.

APPROACH TO THE PATIENT

Spinal Cord Disease

SPINAL CORD ANATOMY RELEVANT TO CLINICAL SIGNS

The spinal cord is a thin, tubular extension of the central nervous system contained within the bony spinal canal. It originates at the medulla and continues caudally to the conus medullaris at the lumbar level; its fibrous extension, the filum terminale, terminates at the coccyx. The adult spinal cord is ~46 cm (18 in.) long, oval in shape, and enlarged in the cervical and lumbar regions, where neurons that innervate the upper and lower extremities, respectively, are located. The white matter tracts containing ascending sensory and descending motor pathways are located peripherally, whereas nerve cell bodies are clustered in an inner region of gray matter shaped like a four-leaf clover that surrounds the central canal (anatomically an extension of the fourth ventricle). The membranes that cover the spinal cord—the pia, arachnoid, and dura—are continuous with those of the brain, and the cerebrospinal fluid is contained within the subarachnoid space between the pia and arachnoid.

The spinal cord has 31 segments, each defined by an exiting ventral motor root and entering dorsal sensory root. During embryologic development, growth of the cord lags behind that of the vertebral column, and the mature spinal cord ends at approximately the first lumbar vertebral body. The lower spinal nerves take an increasingly downward course to exit via intervertebral foramina. The first seven pairs of cervical spinal nerves exit above the same-numbered vertebral bodies, whereas all the subsequent nerves exit below the same-numbered vertebral bodies; this is because there are eight cervical spinal cord segments but only seven cervical vertebrae. The relationship between spinal cord segments and the corresponding vertebral bodies is shown in **Table 453-2**. These relationships assume particular importance for localization of lesions that cause spinal cord compression. Sensory loss below

TABLE 453-1 Treatable Spinal Cord Disorders

Compressive
Epidural, intradural, or intramedullary neoplasm
Epidural abscess
Epidural hemorrhage
Cervical spondylosis
Herniated disk
Posttraumatic compression by fractured or displaced vertebra or hemorrhage
Vascular
Arteriovenous malformation and dural fistula
Antiphospholipid syndrome and other hypercoagulable states
Inflammatory
Multiple sclerosis
Neuromyelitis optica
Sarcoidosis
Systemic immune-mediated disorders: SLE, Sjögren's, Behcet's disease, APL antibody syndrome, others vasculitis
Other CNS disorders: anti-MOG, anti-GFAP, paraneoplastic,[a] CLIPPERS, Erdheim-Chester
Infectious
Viral: VZV, HSV-1 and -2, CMV, HIV, HTLV-1, others
Bacterial and mycobacterial: *Borrelia*, *Listeria*, syphilis, others
Mycoplasma pneumoniae
Parasitic: schistosomiasis, toxoplasmosis, cysticercosis
Developmental
Syringomyelia
Meningomyelocele
Tethered cord syndrome
Metabolic
Vitamin B_{12} deficiency (subacute combined degeneration)
Folate deficiency
Copper deficiency

[a] Including anti-amphiphysin, CRMP-5, Hu.

Abbreviations: CLIPPERS, chronic lymphocytic inflammation with pontine perivascular enhancement responsive to steroids; CMV, cytomegalovirus; CNS, central nervous system; CRMP5, collapsin response mediator 5-IgG; GFAP, glial fibrillary acidic protein; HSV, herpes simplex virus; HTLV, human T-cell lymphotropic virus; MOG, myelin oligodendrocyte glycoprotein; SLE, systemic lupus erythematosus; VZV, varicella-zoster virus.

the level of the umbilicus, for example, corresponds to pathology at the T10 cord segment, which is located adjacent to the seventh or eighth thoracic vertebral body **(see Figs. 27-2 and 27-3)**. In addition, at every level, the main ascending and descending tracts are somatotopically organized with a laminated distribution that reflects the origin or destination of nerve fibers.

Determining the Level of the Lesion The presence of a horizontally defined level below which sensory, motor, and autonomic function is impaired is the hallmark of a spinal cord lesion. This *sensory level* is sought by asking the patient to identify a pinprick or cold stimulus applied to the proximal legs and lower trunk and successively moved up toward the neck on each side. Sensory loss below this level is the result of damage to the spinothalamic tract on the opposite side, one to two segments higher in the case of a unilateral spinal cord lesion, and at the level of a bilateral lesion. The discrepancy in the level of a unilateral lesion is the result of the course of the second-order sensory fibers, which originate in the dorsal horn and ascend for one or two levels as they cross anterior to the central canal to join the opposite spinothalamic tract. Lesions that transect the descending corticospinal and other motor tracts cause paraplegia or quadriplegia with heightened deep tendon reflexes, Babinski signs, and eventual spasticity (upper motor neuron syndrome). Transverse damage to the cord also produces autonomic disturbances consisting of absent sweating below the implicated cord level and bladder, bowel, and sexual dysfunction.

TABLE 453-2 Spinal Cord Levels Relative to the Vertebral Bodies

SPINAL CORD LEVEL	CORRESPONDING VERTEBRAL BODY
Upper cervical	Same as cord level
Lower cervical	1 level higher
Upper thoracic	2 levels higher
Lower thoracic	2–3 levels higher
Lumbar	T10–T12
Sacral	T12–L1

The uppermost level of a spinal cord lesion can also be localized by attention to the *segmental signs* corresponding to disturbed motor or sensory innervation by an individual cord segment. A band of altered sensation (hyperalgesia or hyperpathia) at the upper end of the sensory disturbance, fasciculations or atrophy in muscles innervated by one or several segments, or a muted or absent deep tendon reflex may be noted at this level. These signs also can occur with focal root or peripheral nerve disorders; thus, they are most useful when they occur together with signs of long-tract damage. With severe and acute transverse lesions, the limbs initially may be flaccid rather than spastic. This state of "spinal shock" lasts for several days, rarely for weeks, and may be mistaken for extensive damage to the anterior horn cells over many segments of the cord or for an acute polyneuropathy.

The main features of transverse damage at each level of the spinal cord are summarized below.

Cervical Cord Upper cervical cord lesions produce quadriplegia and weakness of the diaphragm. The uppermost level of weakness and reflex loss with lesions at C5–C6 is in the biceps; at C7, in finger and wrist extensors and triceps; and at C8, finger and wrist flexion. Horner's syndrome (miosis, ptosis, and facial hypohidrosis) may accompany a cervical cord lesion at any level.

Thoracic Cord Lesions here are localized by the sensory level on the trunk and, if present, by the site of midline back pain. Useful markers of the sensory level on the trunk are the nipples (T4) and umbilicus (T10). Leg weakness and disturbances of bladder and bowel function accompany the paralysis. Lesions at T9–T10 paralyze the lower—but not the upper—abdominal muscles, resulting in upward movement of the umbilicus when the abdominal wall contracts (*Beevor's sign*).

Lumbar Cord Lesions at the L2–L4 spinal cord levels paralyze flexion and adduction of the thigh, weaken leg extension at the knee, and abolish the patellar reflex. Lesions at L5–S1 paralyze only movements of the foot and ankle, flexion at the knee, and extension of the thigh, and abolish the ankle jerks (S1).

Sacral Cord/Conus Medullaris The conus medullaris is the tapered caudal termination of the spinal cord, comprising the sacral and single coccygeal segments. The distinctive conus syndrome consists of bilateral saddle anesthesia (S3–S5), prominent bladder and bowel dysfunction (urinary retention and incontinence with lax anal tone), and impotence. The bulbocavernosus (S2–S4) and anal (S4–S5) reflexes are absent **(Chap. 433)**. Muscle strength is largely preserved. By contrast, lesions of the cauda equina, the nerve roots derived from the lower cord, are characterized by low back and radicular pain, asymmetric leg weakness and sensory loss, variable areflexia in the lower extremities, and relative sparing of bowel and bladder function. Mass lesions in the lower spinal canal often produce a mixed clinical picture with elements of both cauda equina and conus medullaris syndromes.

Special Patterns of Spinal Cord Disease The location of the major ascending and descending pathways of the spinal cord are shown in **Fig. 453-1**. Most fiber tracts—including the posterior columns and the spinocerebellar and pyramidal tracts—are situated on the side of the body they innervate. However, afferent fibers mediating pain and temperature sensation ascend in the spinothalamic

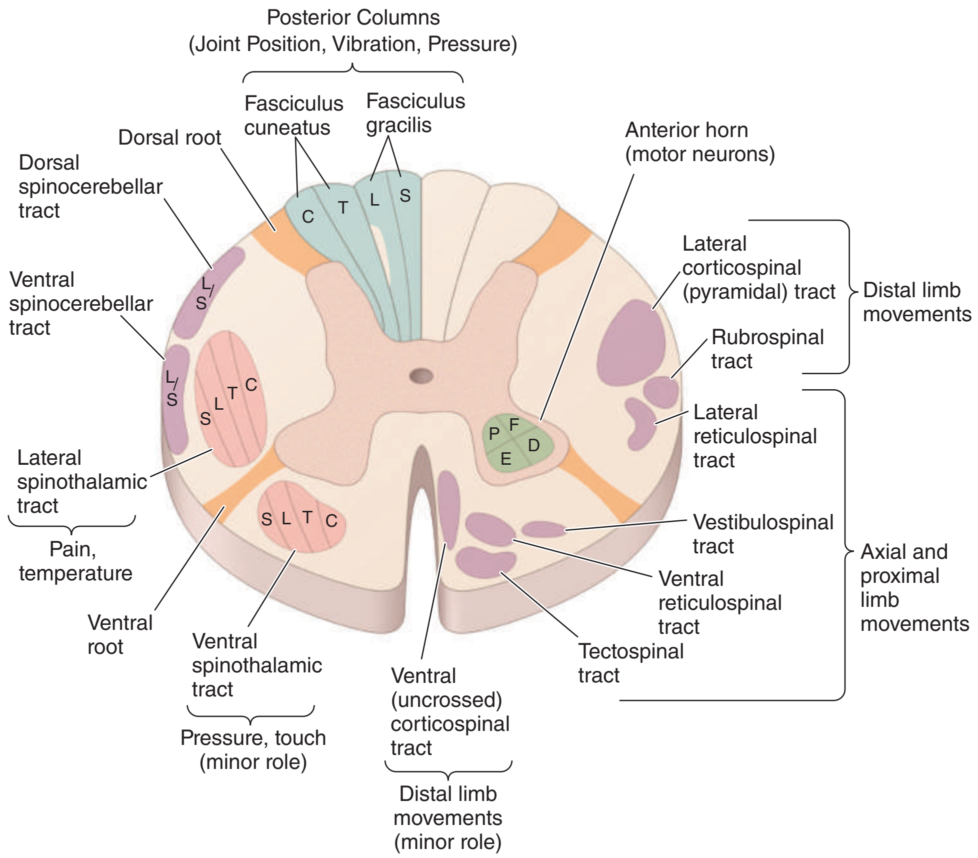
Posterior Columns
(Joint Position, Vibration, Pressure)
Fasciculus cuneatus
Fasciculus gracilis
C
T
L
S
Dorsal root
Dorsal spinocerebellar tract
L/S
Ventral spinocerebellar tract
L/S
C
T
L
S
Lateral spinothalamic tract
Pain, temperature
S L T C
Ventral root
Ventral spinothalamic tract
Pressure, touch (minor role)
Ventral (uncrossed) corticospinal tract
Distal limb movements (minor role)
Anterior horn (motor neurons)
P F E D
Lateral corticospinal (pyramidal) tract
Rubrospinal tract
Distal limb movements
Lateral reticulospinal tract
Vestibulospinal tract
Ventral reticulospinal tract
Tectospinal tract
Axial and proximal limb movements

FIGURE 453-1 Transverse section through the spinal cord, composite representation, illustrating the principal ascending (*left*) and descending (*right*) pathways. The lateral and ventral spinothalamic tracts ascend contralateral to the side of the body that is innervated. In humans, the lateral corticospinal (pyramidal) tract is thought to lack strict somatotopic organization in the spinal cord. C, cervical; D, distal; E, extensors; F, flexors; L, lumbar; P, proximal; S, sacral; T, thoracic.

tract contralateral to the side they supply. The anatomic configurations of these tracts produce characteristic syndromes that provide clues to the underlying disease process.

Brown-Sequard Hemicord Syndrome This consists of ipsilateral weakness (corticospinal tract) and loss of joint position and vibratory sense (posterior column), with contralateral loss of pain and temperature sense (spinothalamic tract) one or two levels below the lesion. Segmental signs, such as radicular pain, muscle atrophy, or loss of a deep tendon reflex, are unilateral. Partial forms are more common than the fully developed syndrome.

Central Cord Syndrome This syndrome results from selective damage to the gray matter nerve cells and crossing spinothalamic tracts surrounding the central canal. In the cervical cord, the central cord syndrome produces arm weakness out of proportion to leg weakness and a "dissociated" sensory loss, meaning loss of pain and temperature sensations over the shoulders, lower neck, and upper trunk (cape distribution), in contrast to preservation of light touch, joint position, and vibration sense in these regions. Spinal trauma, syringomyelia, and intrinsic cord tumors are the main causes.

Anterior Cord Syndrome Infarction of the cord is generally the result of occlusion or diminished flow in the anterior spinal artery. The result is bilateral tissue destruction at several contiguous levels that spares the posterior columns. All spinal cord functions—motor, sensory, and autonomic—are lost below the level of the lesion, with the striking exception of retained vibration and position sensation.

Foramen Magnum Syndrome Lesions in this area interrupt decussating pyramidal tract fibers destined for the legs, which cross caudal to those of the arms, resulting in weakness of the legs (*crural paresis*). Compressive lesions near the foramen magnum may produce weakness of the ipsilateral shoulder and arm followed by weakness of the ipsilateral leg, then the contralateral leg, and finally the contralateral arm, an "around-the-clock" pattern that may begin in any of the four limbs. There is typically suboccipital pain spreading to the neck and shoulders.

Intramedullary and Extramedullary Syndromes It is useful to differentiate *intramedullary* processes, arising within the substance of the cord, from *extramedullary* ones that lie outside the cord and compress the spinal cord or its vascular supply. The differentiating features are only relative and serve as clinical guides. With extramedullary lesions, radicular pain is often prominent, and there is early sacral sensory loss and spastic weakness in the legs with incontinence due to injury to the corresponding sensory and motor fibers in the spinothalamic and corticospinal tracts (Fig. 453-1). Intramedullary lesions tend to produce poorly localized burning pain rather than radicular pain and to spare sensation in the perineal and sacral areas ("sacral sparing"), reflecting the laminated configuration of the spinothalamic tract with sacral fibers outermost; corticospinal tract signs appear later. Regarding extramedullary lesions, a further distinction is made between extradural and intradural masses, as the former are generally malignant and the latter benign (neurofibroma being a common cause). Consequently, a long duration of symptoms favors an intradural origin.

ACUTE AND SUBACUTE SPINAL CORD DISEASES

Symptoms of the cord diseases that evolve over days or weeks typically present as focal neck or back pain, followed by various combinations of paresthesias, sensory loss, motor weakness, and sphincter disturbance. There may be mild sensory symptoms only or a devastating functional transection of the cord. When paresthesias begin in the feet and then ascend, a polyneuropathy is often considered, and in such cases, the presence of bladder disturbances and a sharply demarcated spinal cord level provide important clues to the spinal cord origin of the disease.

In severe and abrupt cases, areflexia reflecting spinal shock may be present, but hyperreflexia supervenes over days or weeks; persistent areflexic paralysis with a sensory level usually indicates necrosis over multiple segments of the spinal cord.

APPROACH TO THE PATIENT

Compressive and Noncompressive Myelopathy

DISTINGUISHING COMPRESSIVE FROM NONCOMPRESSIVE MYELOPATHY

The first priority is to exclude treatable compression of the cord by a mass lesion. The common causes are tumor, epidural abscess or hematoma, herniated disk, and spondylitic vertebral pathology. Epidural compression due to malignancy or abscess often causes warning signs of neck or back pain, bladder disturbances, and sensory symptoms that precede the development of paralysis. Spinal subluxation, hemorrhage, and noncompressive etiologies such as infarction are more likely to produce myelopathy without antecedent symptoms. Magnetic resonance imaging (MRI) with gadolinium, centered on the clinically suspected level, is the initial diagnostic procedure if it is available; it is often appropriate to image the entire spine (cervical through sacral regions) to search for additional clinically silent lesions. Once compressive lesions have been excluded, noncompressive causes of acute myelopathy that are intrinsic to the cord are considered, primarily vascular, inflammatory, and infectious etiologies.

COMPRESSIVE MYELOPATHIES

Neoplastic Spinal Cord Compression In adults, most neoplasms are epidural in origin, resulting from metastases to the adjacent vertebral column. The propensity of solid tumors to metastasize to the vertebral column probably reflects the high proportion of bone marrow located in the axial skeleton. Almost any malignant tumor can metastasize to the spinal column, with breast, lung, prostate, kidney, lymphoma, and myeloma being particularly frequent. The thoracic spinal column is most commonly involved; exceptions are metastases from prostate and ovarian cancer, which occur disproportionately in the sacral and lumbar vertebrae, probably from spread through Batson's plexus, a network of veins along the anterior epidural space. Retroperitoneal neoplasms (especially lymphomas or sarcomas) enter the spinal canal laterally through the intervertebral foramina and produce radicular pain with signs of weakness that corresponds to the level of involved nerve roots.

Pain is usually the initial symptom of spinal metastasis; it may be aching and localized or sharp and radiating in quality and typically worsens with movement, coughing, or sneezing and characteristically awakens patients at night. A recent onset of persistent back pain, particularly if in the thoracic spine (which is uncommonly involved by spondylosis), should prompt consideration of vertebral metastasis. Rarely, pain is mild or absent. Plain radiographs of the spine and radionuclide bone scans have a limited role in diagnosis because they do not identify 15–20% of metastatic vertebral lesions and fail to detect paravertebral masses that reach the epidural space through the intervertebral foramina. MRI provides excellent anatomic resolution of the extent of spinal tumors **(Fig. 453-2)** and is able to distinguish between malignant lesions and other masses—epidural abscess, tuberculoma, lipoma, or epidural hemorrhage, among others—that present in a similar fashion. Vertebral metastases are usually hypointense relative to a normal bone marrow signal on T1-weighted MRI; after the administration of gadolinium, contrast enhancement may deceptively "normalize" the appearance of the tumor by increasing its intensity to that of normal bone marrow. Infections of the spinal column (osteomyelitis and related disorders) are distinctive in that, unlike tumor, they often cross the disk space to involve the adjacent vertebral body.

If spinal cord compression is suspected, imaging should be obtained promptly. If there are radicular symptoms but no evidence of myelopathy, it may be safe to defer imaging for 24–48 h. Up to 40% of patients who present with cord compression at one level are found to have asymptomatic epidural metastases elsewhere; thus, imaging of the entire length of the spine is important to define the extent of disease.

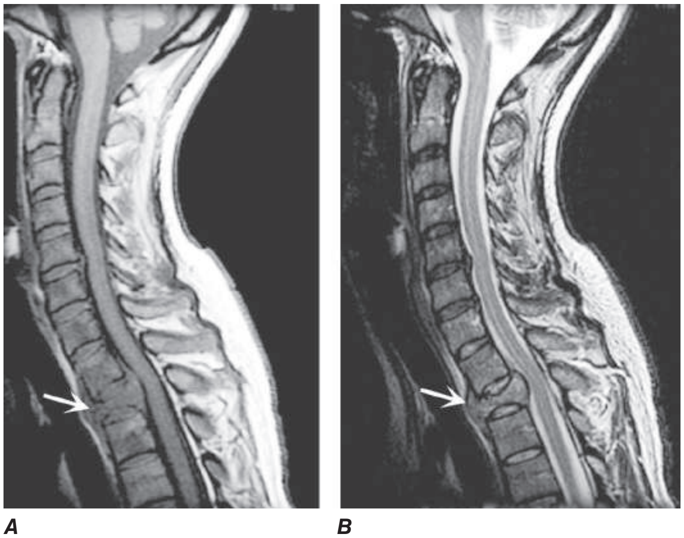

FIGURE 453-2 Epidural spinal cord compression due to breast carcinoma. Sagittal T1-weighted (***A***) and T2-weighted (***B***) magnetic resonance imaging scans through the cervicothoracic junction reveal an infiltrated and collapsed second thoracic vertebral body with posterior displacement and compression of the upper thoracic spinal cord. The low-intensity bone marrow signal in ***A*** signifies replacement by tumor.

TREATMENT

Neoplastic Spinal Cord Compression

Proper management is based on multiple considerations, including radiosensitivity of the primary tumor, extent of compression, prior therapy to the site, and stability of the spine. Treatment includes glucocorticoids to reduce cord edema, surgery and/or local radiotherapy (initiated as early as possible) to the symptomatic lesion, and specific therapy for the underlying tumor type. Glucocorticoids (typically dexamethasone, 10 mg intravenously) can be administered before an imaging study if there is clinical suspicion of cord compression and continued at a lower dose (4 mg every 6 h orally) until definitive treatment with radiotherapy and/or surgical decompression is completed. In one trial, initial management with surgery followed by radiotherapy was more effective than radiotherapy alone for patients with a single area of spinal cord compression by extradural tumor; however, patients with recurrent cord compression, brain metastases, radiosensitive tumors, or severe motor symptoms of >48 h in duration were excluded from this study. Stereotactic body radiotherapy, which delivers high doses of focused radiation, is preferred for radioresistant tumor types and for patients requiring re-irradiation.

Biopsy of the epidural mass is unnecessary in patients with known primary cancer, but it is indicated if a history of underlying cancer is lacking. Surgical treatment, either decompression by laminectomy or a spinal fixation procedure, is also indicated when signs of cord compression worsen despite radiotherapy; the maximum-tolerated dose of radiotherapy has been delivered previously to the site; a vertebral compression fracture or spinal instability contributes to cord compression; or in cases of high-grade spinal cord compression from a radioresistant tumor.

A good response to therapy can be expected in individuals who are ambulatory at presentation. Treatment usually prevents new weakness, and some recovery of motor function occurs in up to one-third of patients. Motor deficits (paraplegia or quadriplegia), once established for >12 h, do not usually improve, and beyond 48 h, the prognosis for substantial motor recovery is poor. Although most patients do not experience recurrences in the months following radiotherapy, with survival beyond 2 years, recurrence becomes increasingly likely and can be managed with additional radiotherapy.

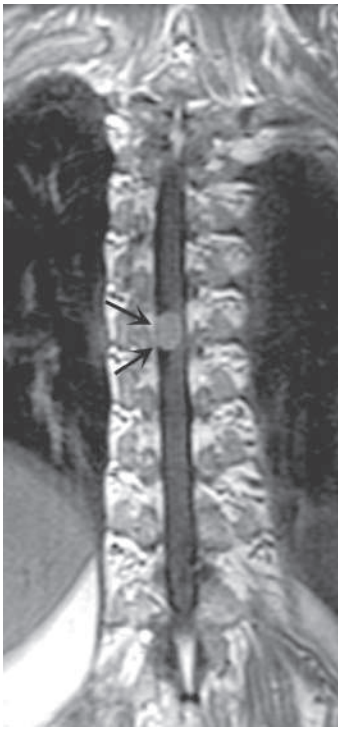

FIGURE 453-3 Magnetic resonance imaging of a thoracic meningioma. Coronal T1-weighted postcontrast image through the thoracic spinal cord demonstrates intense and uniform enhancement of a well-circumscribed extramedullary mass (*arrows*) that displaces the spinal cord to the left.

In contrast to tumors of the epidural space, most intradural mass lesions are slow-growing and benign. Meningiomas and neurofibromas account for most of these, with occasional cases caused by chordoma, lipoma, dermoid, or sarcoma. Meningiomas **(Fig. 453-3)** are often located posterior to the thoracic cord or near the foramen magnum, although they can arise from the meninges anywhere along the spinal canal. Neurofibromas are benign tumors of the nerve sheath that typically arise from the posterior root; when multiple, neurofibromatosis is the likely etiology. Symptoms usually begin with radicular sensory symptoms followed by an asymmetric, progressive spinal cord syndrome. Therapy is surgical resection.

Primary intramedullary tumors of the spinal cord are uncommon. They present as central cord or hemicord syndromes, often in the cervical region. There may be poorly localized burning pain in the extremities and sparing of sacral sensation. In adults, these lesions are ependymomas, hemangioblastomas, or low-grade astrocytomas **(Fig. 453-4)**. Complete resection of an intramedullary ependymoma is often possible with microsurgical techniques. Debulking of an intramedullary astrocytoma can also be helpful, as these are often slowly growing lesions; the value of adjunctive radiotherapy and chemotherapy is uncertain. Secondary (metastatic) intramedullary tumors also occur, especially in patients with advanced metastatic disease **(Chap. 95)**, although these are not nearly as frequent as brain metastases.

Spinal Epidural Abscess Spinal epidural abscess presents with midline back or neck pain, fever, and progressive limb weakness. Prompt recognition of this distinctive process may prevent permanent sequelae. Aching pain is almost always present, either over the spine or in a radicular pattern. The duration of pain prior to presentation is generally ≤2 weeks but may on occasion be several months or longer. Fever is typically but not invariably present, accompanied by elevated white blood cell count, sedimentation rate, and C-reactive protein. As the abscess expands, further spinal cord damage results from venous congestion and thrombosis. Once weakness and other signs of myelopathy appear, progression may be rapid and irreversible. A more chronic sterile granulomatous form of abscess is also known, usually after treatment of an acute epidural infection.

Risk factors include an impaired immune status (HIV, diabetes mellitus, renal failure, alcoholism, malignancy), intravenous drug abuse, and infections of the skin or other tissues. Two-thirds of epidural infections result from hematogenous spread of bacteria from the skin (furunculosis), soft tissue (pharyngeal or dental abscesses; sinusitis), or deep viscera (bacterial endocarditis). The remainder arises from direct extension of a local infection to the subdural space; examples of local predisposing conditions are vertebral osteomyelitis, decubitus ulcers, lumbar puncture, epidural anesthesia, or spinal surgery. Most cases are due to *Staphylococcus aureus*; gram-negative bacilli, *Streptococcus*, anaerobes, and fungi can also cause epidural abscesses. Methicillin-resistant *Staphylococcus aureus* (MRSA) is an important consideration, and therapy should be tailored to this possibility. Tuberculosis from an adjacent vertebral source (Pott's disease) remains an important cause in the developing world.

MRI **(Fig. 453-5)** localizes the abscess and excludes other causes of myelopathy. Blood cultures are positive in more than half of cases, but direct aspiration of the abscess at surgery is often required for a

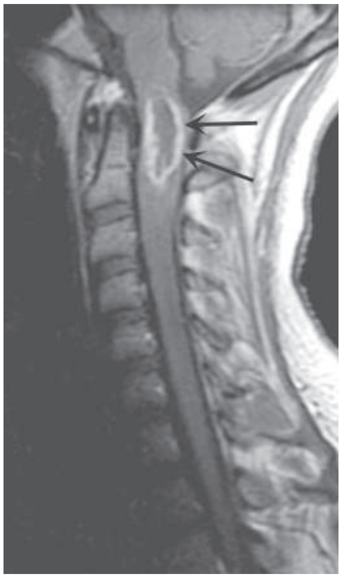

FIGURE 453-4 Magnetic resonance imaging of an intramedullary astrocytoma. Sagittal T1-weighted postcontrast image through the cervical spine demonstrates expansion of the upper cervical spine by a mass lesion emanating from within the spinal cord at the cervicomedullary junction. Irregular peripheral enhancement occurs within the mass (*arrows*).

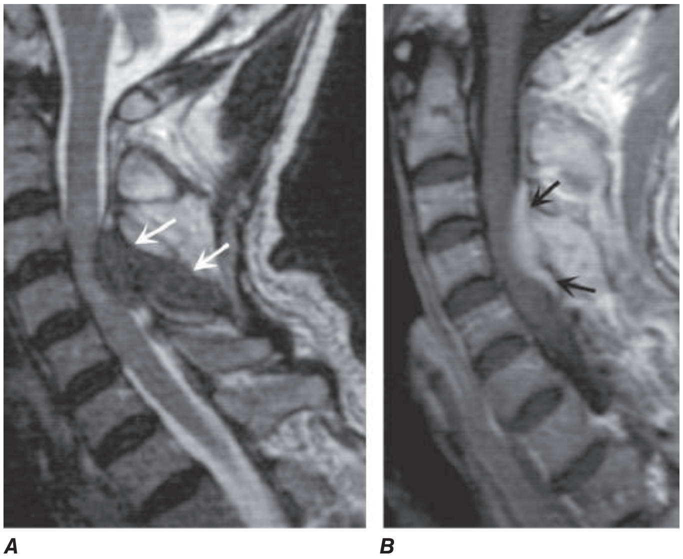

FIGURE 453-5 Magnetic resonance (MR) imaging of a spinal epidural abscess due to tuberculosis. ***A.*** Sagittal T2-weighted free spin-echo MR sequence. A hypointense mass replaces the posterior elements of C3 and extends epidurally to compress the spinal cord (*arrows*). ***B.*** Sagittal T1-weighted image after contrast administration reveals a diffuse enhancement of the epidural process (*arrows*) with extension into the epidural space.

microbiologic diagnosis. Lumbar puncture is only required if encephalopathy or other clinical signs raise the question of associated meningitis, a feature that is found in <25% of cases. The level of the puncture should be planned to minimize the risk of meningitis due to passage of the needle through infected tissue. A high cervical tap is sometimes the safest approach. Cerebrospinal fluid (CSF) abnormalities in epidural and subdural abscesses consist of pleocytosis with a preponderance of polymorphonuclear cells, an elevated protein level, and a reduced glucose level, but the responsible organism is not cultured unless there is associated meningitis.

TREATMENT

Spinal Epidural Abscess

Treatment is by decompressive laminectomy with debridement combined with long-term antibiotic treatment. Surgical evacuation prevents development of paralysis and may improve or reverse paralysis in evolution, but it is unlikely to improve fixed deficits more than several days in duration. Broad-spectrum antibiotics, typically vancomycin 15–20 mg/kg q12h (*Staphylococcus* including MRSA, *Streptococcus*), ceftriaxone 2 g q12h (gram-negative bacilli), and when indicated, metronidazole 30 mg/kg per day divided into q6h intervals (anaerobes), should be started empirically before surgery and then modified on the basis of culture results; medication is generally continued for 6–8 weeks. If surgery is contraindicated or if there is a fixed paraplegia or quadriplegia that is unlikely to improve following surgery, long-term administration of systemic and oral antibiotics can be used; in such cases, the choice of antibiotics may be guided by results of blood cultures. Surgical management remains the treatment of choice unless the abscess is limited in size and causes few or no neurologic signs.

With prompt diagnosis and treatment, up to two-thirds of patients experience significant recovery.

Spinal Epidural Hematoma Hemorrhage into the epidural (or subdural) space causes acute focal or radicular pain followed by variable signs of a spinal cord or conus medullaris disorder. Therapeutic anticoagulation, trauma, tumor, or blood dyscrasias are predisposing conditions. Rare cases complicate lumbar puncture or epidural anesthesia. MRI and computed tomography (CT) confirm the clinical suspicion and can delineate the extent of the bleeding. Treatment consists of prompt reversal of any underlying clotting disorder and surgical decompression. Surgery may be followed by substantial recovery, especially in patients with some preservation of motor function preoperatively. Because of the risk of hemorrhage, lumbar puncture should be avoided whenever possible in patients with severe thrombocytopenia or other coagulopathies.

Hematomyelia Hemorrhage into the substance of the spinal cord is a rare result of trauma, intraparenchymal vascular malformation (see below), vasculitis due to polyarteritis nodosa or systemic lupus erythematosus (SLE), bleeding disorders, or a spinal cord neoplasm. Hematomyelia presents as an acute painful transverse myelopathy. With large lesions, extension into the subarachnoid space results in subarachnoid hemorrhage **(Chap. 440)**. Diagnosis is by MRI or CT. Therapy is supportive, and surgical intervention is generally not useful. An exception is hematomyelia due to an underlying vascular malformation, for which spinal angiography and endovascular occlusion may be indicated, or surgery to evacuate the clot and remove the underlying vascular lesion.

Acute Spondylitic Myelopathy Of particular concern are hyperextension injuries in patients with underlying degenerative cervical spine disease **(Chap. 19)**. The provoking stimulus may be obvious, such as a forward fall, or occur after seemingly innocuous low-impact movements of the neck. A preexisting stenotic spinal canal is often present, and “buckling” of the posterior ligamentum flavum (less commonly acute disk herniation or subluxation) is believed to produce the cord compression, sometimes with a central cord syndrome (see above) and involvement of the upper, more than lower, limbs. Deficits can be transient, resulting in a “concussion” of the spinal cord, or permanent. *The more common syndrome of chronic spondylitic myelopathy is discussed below.*

NONCOMPRESSIVE MYELOPATHIES

Once a compressive etiology has been excluded as the cause of an acute myelopathy, the principal challenge is to distinguish vascular/ischemic from inflammatory/infectious causes. This is often not straightforward because clinical presentations can overlap. Moreover, findings that usually point to an inflammatory etiology—such as focal gadolinium enhancement on MRI scans or pleocytosis in the CSF—can also occur with spinal cord ischemia. Ischemia is likely in hyperacute presentations with back or neck pain and when an anterior pattern of spinal cord injury is identified on clinical examination or by MRI. By contrast, inflammation is more likely in cases that develop subacutely or when systemic symptoms, CSF oligoclonal bands, or multiple discrete spinal cord MRI lesions are present. The most frequent inflammatory causes of acute myelopathy are multiple sclerosis (MS); neuromyelitis optica (NMO); sarcoidosis; systemic inflammatory diseases such as SLE and Behcet’s disease; postinfectious or idiopathic transverse myelitis, which is presumed to be an immune condition related to acute disseminated encephalomyelitis **(Chap. 456)**; and infectious (primarily viral) causes.

The evaluation generally requires a lumbar puncture and a search for underlying systemic disease **(Table 453-3)**.

Spinal Cord Infarction The cord is supplied by three arteries that course vertically over its surface: a single anterior spinal artery and paired posterior spinal arteries. The anterior spinal artery originates in paired branches of the vertebral arteries at the craniocervical junction and is fed by additional radicular vessels that arise at C6, at an upper thoracic level, and, most consistently, at T11–L2 (artery of Adamkiewicz). At each spinal cord segment, paired penetrating vessels branch from the anterior spinal artery to supply the anterior two-thirds of the

TABLE 453-3 Considerations in the Evaluation of Myelopathy

1. MRI of spinal cord with and without contrast (exclude compressive causes).
2. CSF studies: Cell count, protein, glucose, IgG index/synthesis rate, oligoclonal bands, VDRL; Gram’s stain, acid-fast bacilli, and India ink stains; PCR for VZV, HSV-2, HSV-1, EBV, CMV, HHV-6, enteroviruses, HIV; antibody for HTLV-1, *Borrelia burgdorferi*, *Mycoplasma pneumoniae*, and *Chlamydia pneumoniae*; viral, bacterial, mycobacterial, and fungal cultures.
3. Blood studies for infection: HIV; RPR; IgG and IgM enterovirus antibody; IgM WNV, group B arbovirus, mumps, measles, rubella, *Brucella melitensis*, *Chlamydia psittaci*, *Bartonella henselae*, schistosomal antibody; PCR and antigen tests for SARS-CoV-2; cultures for *B. melitensis*. Also consider nasal/pharyngeal/anal cultures for enteroviruses; stool O&P for *Schistosoma* ova.
4. Vascular causes: MRI, CT myelogram; spinal angiogram.
5. Multiple sclerosis: Brain MRI scan; evoked potentials.
6. Neuromyelitis optica and related disorders: Serum anti-aquaporin-4 antibody, anti-MOG antibody, anti-GFAP antibody.
7. Sarcoidosis: Serum angiotensin-converting enzyme; serum Ca; 24-h urine Ca; chest x-ray; chest CT; slit-lamp eye examination; total-body gallium scan; lymph node biopsy.
8. Systemic immune-mediated disorders: ESR; ANA; ENA; dsDNA; rheumatoid factor; anti-SSA; anti-SSB, complement levels; antiphospholipid and anticardiolipin antibodies; pANCA; antimicrosomal and antithyroglobulin antibodies; if Sjögren’s syndrome suspected, Schirmer test, salivary gland scintigraphy, and salivary/lacrimal gland biopsy.
9. Paraneoplastic disorders: Antibody for amphiphysin, CRMP5, Hu, others.
10. Other: vitamin B_{12}, copper, zinc.

Abbreviations: ANA, antinuclear antibodies; Ca, calcium; CMV, cytomegalovirus; CRMP5, collapsin response mediator 5-IgG; CSF, cerebrospinal fluid; CT, computed tomography; EBV, Epstein-Barr virus; ENA, epithelial neutrophil-activating peptide; ESR, erythrocyte sedimentation rate; GFAP, glial fibrillary acidic protein; HHV, human herpes virus; HSV, herpes simplex virus; HTLV, human T-cell leukemia/lymphoma virus; MOG, myelin oligodendrocyte glycoprotein; MRI, magnetic resonance imaging; O&P, ova and parasites; pANCA, perinuclear antineutrophilic cytoplasmic antibodies; PCR, polymerase chain reaction; RPR, rapid plasma reagin (test); VDRL, Venereal Disease Research Laboratory; VZV, varicella-zoster virus; WNV West Nile virus.

cord; the posterior spinal arteries, which often become less distinct below the midthoracic level, supply the posterior columns.

Spinal cord ischemia can occur at any level; however, the presence of the artery of Adamkiewicz below, and the anterior spinal artery circulation above, creates a region of marginal blood flow in the upper thoracic segments. With hypotension or cross-clamping of the aorta, cord infarction typically occurs at the level of T3–T4 and also at boundary zones between the anterior and posterior spinal artery territories. The latter may result in a rapidly progressive syndrome over hours of weakness and spasticity with little sensory change.

Acute infarction in the territory of the *anterior spinal artery* produces paraplegia or quadriplegia, dissociated sensory loss affecting pain and temperature sense but sparing vibration and position sense, and loss of sphincter control (anterior cord syndrome). Onset may be sudden but more typically is progressive over minutes or a few hours, unlike stroke in the cerebral hemispheres. Sharp midline or radiating back pain localized to the area of ischemia is frequent. Areflexia due to spinal shock is often present initially; with time, hyperreflexia and spasticity appear. Less common is infarction in the territory of the *posterior spinal arteries*, resulting in loss of posterior column function either on one side or bilaterally.

Causes of spinal cord infarction include aortic atherosclerosis, dissecting aortic aneurysm, vertebral artery occlusion or dissection in the neck, aortic surgery, or profound hypotension from any cause. Cardiogenic emboli, vasculitis **(Chap. 375)**, and collagen vascular disease (particularly SLE **[Chap. 368]**, Sjögren's syndrome **[Chap. 373]**, and the antiphospholipid antibody syndrome **[Chap. 369]**) are other etiologies. Occasional cases develop from *embolism of nucleus pulposus* material into spinal vessels, usually from local spine trauma. A *surfer's myelopathy,* usually in the thoracic region, has been associated with prolonged back extension due to lifting the upper body off the board while waiting for waves; it typically manifests as back pain followed by an anterior cord syndrome with progressive paralysis and loss of sphincter control and is likely vascular in origin. A few reports have also been associated with cocaine use, as well as with heroin. In a substantial number of cases, no cause can be found, and thromboembolism in arterial feeders is suspected.

MRI may fail to demonstrate infarctions of the cord, especially in the first day, but often the imaging becomes abnormal at the affected level. MRI features suggestive of cord infarction include diffusion-weighted restriction; longitudinally extensive anterior T2 signal brightness on sagittal images ("pencil-like sign"); focal enhancement in the anterior horns; and paired areas of focal T2 hyperintensity in the anterior medial cord on axial images ("owl's eyes"). When present, infarction of a vertebral body adjacent to the area of cord involvement is diagnostically helpful.

With cord infarction due to presumed thromboembolism, acute anticoagulation is not indicated, with the possible exception of the unusual transient ischemic attack or incomplete infarction with a stuttering or progressive course. The antiphospholipid antibody syndrome is treated with anticoagulation **(Chap. 369)**. Increasing systemic blood pressure to a mean arterial pressure of >90 mmHg, or lumbar drainage of spinal fluid, was reportedly helpful in a few published cases of cord infarction, but neither of these approaches has been studied systematically. Prognosis following spinal cord infarction is influenced by the severity of the deficits at presentation; patients with severe motor weakness and those with persistent areflexia usually do poorly, but in one large series, some improvement over time occurred in many patients, with more than half ultimately regaining some ambulation.

Inflammatory and Immune Myelopathies (Myelitis) This broad category includes MS, NMO, and postinfectious myelitis, as well as sarcoidosis, systemic autoimmune disease, and infections. In approximately one-quarter of cases of myelitis, no underlying cause can be identified. Some will later manifest additional symptoms of an immune-mediated disease. *Transverse myelitis* refers to a pattern of extensive spinal cord injury due to inflammation, clinically manifest as bilateral sensory symptoms, unilateral or bilateral weakness, and bladder and/or bowel disturbance. In most of the developed world, MS is the most common inflammatory cause of an acute myelitis, but involvement is usually partial and not transverse. *Recurrent episodes of myelitis* are usually due to one of the immune-mediated diseases or to infection with herpes simplex virus (HSV) type 2 (below).

MULTIPLE SCLEROSIS MS may present with acute myelitis, particularly in individuals of Asian or African ancestry. In whites, MS attacks rarely cause a transverse myelopathy (i.e., attacks of bilateral sensory disturbances, unilateral or bilateral weakness, and bladder or bowel symptoms), but MS is among the most common causes of a partial cord syndrome. MRI findings in MS-associated myelitis typically consist of mild swelling of the cord and diffuse or multifocal "shoddy" areas of abnormal signal on T2-weighted sequences. Contrast enhancement, indicating disruption in the blood-brain barrier associated with inflammation, is present in many acute cases. In one study 68% of patients presenting with partial myelitis developed MS after a mean follow-up of 4 years; risk factors for conversion to MS included age <40 years, inflammatory CSF, and more than three periventricular lesions on brain MRI.

Treatment of acute episodes of MS-associated myelitis consists of intravenous methylprednisolone (500 mg qd for 3 days) followed by oral prednisone (1 mg/kg per day for several weeks, then a gradual taper). A course of plasma exchange may be indicated for severe cases if glucocorticoids are ineffective. **MS is discussed in Chap. 455.**

NEUROMYELITIS OPTICA NMO is an immune-mediated disorder consisting of a severe myelopathy that is typically longitudinally extensive, meaning that the lesion typically spans three or more vertebral segments. NMO is associated with optic neuritis that is often bilateral and may precede or follow myelitis by weeks or months and also by brainstem and, in some cases, hypothalamic or focal cerebral white matter involvement. Recurrent myelitis without optic nerve or other involvement can also occur in NMO. CSF studies reveal a variable mononuclear pleocytosis of up to several hundred cells per microliter (higher than in typical MS) with occasional cases showing polymorphonuclear predominant pattern; oligoclonal bands are present in <20% of NMO cases. Diagnostic serum autoantibodies against the water channel protein aquaporin-4 (AQP-4) are present in 90% of patients with NMO; in some AQP-4-negative cases, autoantibodies against the central nervous system (CNS) myelin protein myelin oligodendrocyte glycoprotein (MOG) are found. NMO is also associated with SLE (see below) as well as with other systemic autoimmune diseases; rare cases are paraneoplastic. Acute relapses of NMO are treated with glucocorticoids and, for severe or refractory cases, plasma exchange. Three monoclonal antibodies are now available for prophylactic treatment: eculizumab, a terminal complement inhibitor; inebilizumab, a B-cell depleter; and satralizumab, an interleukin (IL) 6 receptor blocker. Other options include off-label use of azathioprine, mycophenolate, or rituximab. Treatment for 5 years or longer is generally recommended. **NMO is discussed in Chap. 456.**

SARCOIDOSIS Sarcoid myelopathy may present as a slowly progressive or relapsing disorder. Clinically, sensory involvement often predominates. MRI may show edematous swelling of the spinal cord mimicking tumor and in some cases longitudinally extensive involvement resembling NMO. Subpial gadolinium enhancement of active lesions, which may appear nodular, are typically located along the dorsal surface of the cord; on axial images, these dorsal lesions combined with enhancement of the central canal can produce a characteristic "trident sign." The typical CSF profile consists of a mild lymphocyte-predominant pleocytosis and elevated protein level; in a minority of cases, reduced glucose and oligoclonal bands are found. When present, the hypoglycorrhachia can be helpful in distinguishing neurosarcoid from other noninfectious causes of myelitis. The diagnosis is particularly difficult when systemic manifestations of sarcoid are minor or absent (nearly 50% of cases) or when other typical neurologic manifestations of the disease, such as cranial neuropathy, hypothalamic involvement, or meningeal enhancement visualized by MRI, are lacking. A slit-lamp examination of the eye to search for uveitis, chest x-ray and CT to assess pulmonary involvement and mediastinal lymphadenopathy, serum or CSF angiotensin-converting enzyme (ACE; lacks specificity and values are elevated in only a minority of cases), serum calcium, and a gallium scan may assist in the diagnosis. Initial treatment is with high doses of glucocorticoids, which need to be administered long term and tapered slowly while monitoring resolution of clinical and MRI signs of active disease; relapses are managed with high-dose glucocorticoids plus a steroid-sparing immunosuppressant drug (typically mycophenolate mofetil, azathioprine, or methotrexate) or with the tumor necrosis factor α-inhibitor infliximab. **Sarcoidosis is discussed in Chap. 379.**

SYSTEMIC IMMUNE-MEDIATED DISORDERS Myelitis occurs in a small number of patients with SLE, many cases of which are associated with antibodies to AQP-4 and satisfy diagnostic criteria for NMO (discussed above). These patients are at high risk of developing future episodes of myelitis and/or optic neuritis. In others, the etiology of SLE-associated myelitis is uncertain. Antiphospholipid antibodies have been suggested to play a role; however, the frequency of these antibodies is similar in SLE patients with and without myelitis. NMO-associated myelitis typically produces severe clinical disease, CSF pleocytosis with polymorphonuclear leukocytes, and an MRI pattern of central gray matter spinal cord involvement; in cases not due to NMO, less severe and more subacutely evolving clinical findings are often present, with milder CSF lymphocytic pleocytosis and MRI changes consistent with white matter involvement of the cord. In both forms, CSF oligoclonal bands are variable findings. There are no systematic trials of therapy for SLE myelitis, but based on limited data, patients with AQP-4 antibodies should be treated as for NMO (above), and in others, high-dose glucocorticoids followed by cyclophosphamide have been recommended. Severe episodes that do not initially respond to glucocorticoids are often treated with a course of plasma exchange. Sjögren's syndrome **(Chap. 373)** can also be associated with NMO and also with cases of chronic progressive myelopathy. Other immune-mediated myelitides include Behçet's disease **(Chap. 376)**, antiphospholipid antibody syndrome **(Chap. 369)**, mixed connective tissue disease **(Chap. 372)**, and vasculitis related to polyarteritis nodosa, perinuclear antineutrophilic cytoplasmic (pANCA) antibodies, or primary CNS vasculitis **(Chap. 375)**. Occasional cases of myelitis, often accompanied by other manifestations that can include encephalitis or optic neuritis, have been associated with autoantibodies against glial fibrillary acidic protein (GFAP) **(Chap. 456)**. Other rare etiologies are chronic lymphocytic inflammation with pontine perivascular enhancement responsive to steroids (CLIPPERS) and Erdheim-Chester disease producing inflammatory mass-like lesions that can be intramedullary or extra-axial and compressive.

POSTINFECTIOUS MYELITIS Many cases of myelitis, termed *postinfectious* or *postvaccinal*, follow an infection or vaccination. Numerous organisms have been implicated, including Epstein-Barr virus (EBV), cytomegalovirus (CMV), and mycoplasma most frequently, with many others including influenza, measles, varicella, mumps, and yellow fever also described. Recently, vaccination (or active infection) with SARS-CoV-2 virus has been associated with a small number of cases of myelitis and NMO. As in the related disorder acute disseminated encephalomyelitis **(Chap. 456)**, postinfectious myelitis often begins as the patient appears to be recovering from an acute febrile infection or in the subsequent days or weeks, but an infectious agent cannot be isolated from the nervous system or CSF. Serum anti-MOG antibodies are present acutely in some cases, especially in children. The presumption is that the myelitis represents an autoimmune disorder triggered by infection and is not due to direct infection of the spinal cord. No randomized controlled trials of therapy exist; treatment is usually with glucocorticoids or, in fulminant cases, plasma exchange.

ACUTE INFECTIOUS MYELITIS Many viruses have been associated with an acute myelitis that is infectious in nature rather than postinfectious. Nonetheless, the two processes are often difficult to distinguish. Herpes zoster is a well characterized agent of viral myelitis, with direct spread to the spinal cord from dorsal root ganglia, and HSV types 1 and 2, EBV, CMV, and rabies virus are other well-described causes. Zika virus has also been recognized as a cause of infectious myelitis, as has a rare association with monkeypox. HSV-2 (and less commonly HSV-1) produces a distinctive syndrome of recurrent sacral and cauda equina neuritis in association with outbreaks of genital herpes (Elsberg's syndrome). Poliomyelitis is a prototypic viral myelitis that is more or less restricted to the anterior gray matter of the cord containing the spinal motoneurons. A polio-like syndrome can also be caused by a large number of enteroviruses (including enterovirus A-71 and coxsackie) and, importantly, by West Nile virus and other flaviviruses such as Japanese encephalitis. Beginning in 2012, cases of acute flaccid paralysis in children and adolescents have appeared associated with enterovirus A-71 and D-68 infection. Chronic viral myelitic infections, such as those due to HIV or human T-cell lymphotropic virus type 1 (HTLV-1), are discussed below.

Bacterial and mycobacterial myelitis (most are essentially abscesses) are less common than viral causes and much less frequent than cerebral bacterial abscess. Almost any pathogenic species may be responsible, including *Borrelia burgdorferi* (Lyme disease), *Listeria monocytogenes, Mycobacterium tuberculosis,* and *Treponema pallidum* (syphilis). *Mycoplasma pneumoniae* may be a cause of myelitis, but its status is uncertain because many cases are more properly classified as postinfectious.

Schistosomiasis **(Chap. 241)** is an important cause of parasitic myelitis in endemic areas. The process is intensely inflammatory and granulomatous, caused by a local response to tissue-digesting enzymes from the ova of the parasite, typically *Schistosoma haematobium* or *Schistosoma mansoni*. Toxoplasmosis **(Chap. 235)** can occasionally cause a focal myelopathy, and this diagnosis should especially be considered in patients with AIDS **(Chap. 208)**. *Cysticercosis* **(Chap. 242)** is another consideration, although myelitis from this helminth is far less common than parenchymal brain or meningeal involvement.

In cases of suspected viral myelitis, it may be appropriate to begin specific therapy pending laboratory confirmation. Herpes zoster, HSV, and EBV myelitis are treated with intravenous acyclovir (10 mg/kg q8h) or oral valacyclovir (2 g tid) for 10–14 days; CMV is treated with ganciclovir (5 mg/kg IV bid) plus foscarnet (60 mg/kg IV tid) or cidofovir (5 mg/kg per week for 2 weeks, then biweekly for two additional doses).

High-Voltage Electrical Injury Spinal cord injuries are prominent following electrocution from lightning strikes or other accidental electrical exposures. The syndrome consists of transient weakness acutely (often with an altered sensorium and focal cerebral disturbances), sometimes followed several days or even weeks later by a myelopathy that can be severe and permanent. This is a rare injury type, and limited data incriminate a vascular pathology involving the anterior spinal artery and its branches in some cases. Therapy is supportive.

CHRONIC MYELOPATHIES

SPONDYLOTIC MYELOPATHY

Spondylotic myelopathy is the most common cause of myelopathy and of gait difficulty in the elderly, accounting for more than half of nontraumatic spinal cord injuries in some series. Neck and shoulder pain with stiffness are early symptoms; impingement of bone and soft tissue overgrowth on nerve roots results in radicular arm pain, most often in a C5 or C6 distribution. Compression of the cervical cord, which occurs in fewer than one-third of cases, produces a slowly progressive spastic paraparesis, at times asymmetric and often accompanied by paresthesias in the feet and hands. Vibratory sense is diminished in the legs, there is a Romberg sign, and occasionally there is a sensory level for vibration or pinprick on the upper thorax. In some cases, coughing or straining produces leg weakness or radiating arm or shoulder pain. Dermatomal sensory loss in the arms, atrophy of intrinsic hand muscles, increased deep-tendon reflexes in the legs, and extensor plantar responses are common. Urinary urgency or incontinence occurs in advanced cases, but there are many alternative causes of these problems in older individuals. A tendon reflex in the arms is often diminished at some level, most often at the biceps (C5–C6). In individual cases, radicular, myelopathic, or combined signs may predominate. The diagnosis should always be considered in cases of progressive cervical myelopathy, paresthesias of the feet and hands, or wasting of the hands.

Diagnosis is usually made by MRI and may be suspected from CT images; plain x-rays are less helpful. Extrinsic cord compression and deformation are appreciated on axial MRI views, and T2-weighted sequences may reveal areas of high signal intensity within the cord adjacent to the site of compression. A cervical collar may be helpful in milder cases, but the likelihood of progression of medically treated myelopathy is high, estimated at 8% over 1 year. Definitive therapy consists of surgical decompression, either posterior laminectomy or an anterior approach with resection of the protruded disk and bony material. **Cervical spondylosis and related degenerative diseases of the upper spine are discussed in Chap. 19.**

VASCULAR MALFORMATIONS OF THE CORD AND DURA

Vascular malformations, comprising ~4% of all mass lesions of the cord and overlying dura, are treatable causes of progressive myelopathy. Most common are fistulas located within the dura or posteriorly along the surface of the cord. Most dural arteriovenous (AV) fistulas are located at or below the midthoracic level, usually consisting of a direct connection between a radicular feeding artery in the nerve root sleeve with dural veins. The typical presentation is a middle-aged man with a progressive myelopathy that worsens slowly or intermittently and may have periods of remission, sometimes mimicking MS. Acute deterioration due to hemorrhage into the spinal cord (hematomyelia) or subarachnoid space may also occur but is rare. In many cases, progression results from local ischemia and edema due to venous congestion. Most patients have incomplete sensory, motor, and bladder disturbances. The motor disorder may predominate and produce a mixture of upper and restricted lower motor neuron signs, simulating amyotrophic lateral sclerosis (ALS). Pain over the dorsal spine, dysesthesias, or radicular pain may be present. Other symptoms suggestive of AV malformation (AVM) or dural fistula include intermittent claudication; symptoms that change with posture, exertion, Valsalva maneuver, or menses; and fever.

Less commonly, AVM disorders are intramedullary rather than dural. One unusual disorder is a progressive thoracic myelopathy with paraparesis developing over weeks or months, characterized pathologically by abnormally thick, hyalinized vessels within the cord (subacute necrotic myelopathy or Foix-Alajouanine syndrome).

Spinal bruits are infrequent but may be sought at rest and after exercise in suspected cases. A vascular nevus on the overlying skin may indicate an underlying vascular malformation as occurs with Klippel-Trenaunay-Weber syndrome. MR angiography and CT angiography can detect the draining vessels of many AVMs **(Fig. 453-6)**. Definitive diagnosis requires selective spinal angiography, which defines the feeding vessels and the extent of the malformation. Treatment is tailored to the anatomy and location of the lesion and generally consists of microsurgical resection, endovascular embolization of the major feeding vessels, or a combination of the two approaches.

RETROVIRUS-ASSOCIATED MYELOPATHIES

The myelopathy associated with HTLV-1, formerly called tropical spastic paraparesis, is a slowly progressive spastic syndrome with variable sensory and bladder disturbance. Approximately half of patients have mild back or leg pain. The neurologic signs may be asymmetric, often lacking a well-defined sensory level; the only sign in the arms may be hyperreflexia after several years of illness. The onset is usually insidious, and the tempo of clinical progression occurs at a variable rate; in one study, median time for progression to cane-, walker-, or wheelchair-dependent state was 6, 13, and 21 years, respectively. Progression appears to be more rapid in older patients and those with higher viral loads. Diagnosis is made by demonstration of HTLV-1-specific antibody in serum by enzyme-linked immunosorbent assay (ELISA), confirmed by radioimmunoprecipitation or Western blot analysis. Especially in endemic areas, a finding of HTLV-1 seropositivity in a patient with myelopathy does not necessarily prove that HTLV-1 is causative. The CSF/serum antibody index may provide support by establishing intrathecal synthesis of antibodies, including oligoclonal antibodies, favoring HTVL-1 myelopathy over asymptomatic carriage. Measuring proviral DNA by polymerase chain reaction (PCR) in peripheral blood and CSF cells can be useful as an ancillary part of diagnosis. The pathogenesis of the myelopathy is uncertain. It could result from an immune response directed against HTLV-1 antigens in the nervous system or, alternatively, to secondary autoimmunity triggered by the viral infection. There is no proven effective treatment. Based on limited evidence, systemic glucocorticoids, pulsed high-dose induction followed by low-dose chronic maintenance, can be tried, and mogamulizumab, a monoclonal antibody directed against CCR4, has been reported in one preliminary trial to slow progression and reduce neurologic disability in some recipients.

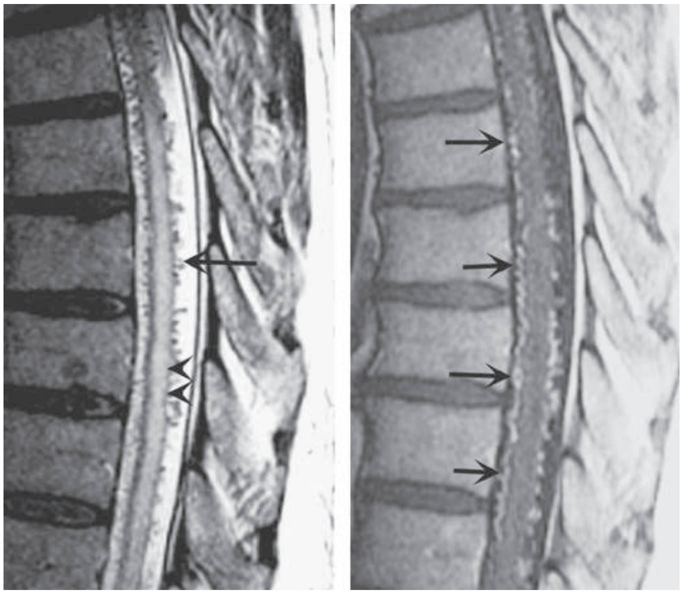

FIGURE 453-6 Arteriovenous malformation. Sagittal magnetic resonance scans of the thoracic spinal cord: T2 fast spin-echo technique (*left*) and T1 postcontrast image (*right*). On the T2-weighted image (*left*), abnormally high signal intensity is noted in the central aspect of the spinal cord (*arrowheads*). Numerous punctate flow voids indent the dorsal and ventral spinal cord (*arrow*). These represent the abnormally dilated venous plexus supplied by a dural arteriovenous fistula. After contrast administration (*right*), multiple, serpentine, enhancing veins (*arrows*) on the ventral and dorsal aspect of the thoracic spinal cord are visualized, diagnostic of arteriovenous malformation. This patient was a 54-year-old man with a 4-year history of progressive paraparesis.

A progressive myelopathy can also result from HIV infection **(Chap. 208)**. It is characterized by vacuolar degeneration of the posterior and lateral tracts, resembling subacute combined degeneration (see below).

SYRINGOMYELIA

Syringomyelia is a developmental cavity in the cervical cord that may enlarge and produce progressive myelopathy or may remain asymptomatic. Symptoms begin insidiously in adolescence or early adulthood, progress irregularly, and may undergo spontaneous arrest for several years. Many young patients acquire a cervical-thoracic scoliosis. More than half of all cases are associated with Chiari type 1 malformations in which the cerebellar tonsils protrude through the foramen magnum and into the cervical spinal canal. The pathophysiology of syrinx expansion is controversial, but some interference with the normal flow of CSF seems likely, perhaps by the Chiari malformation. Acquired cavitations of the cord in areas of necrosis are also termed *syrinx cavities*; these follow trauma, myelitis, necrotic spinal cord tumors, and chronic arachnoiditis due to tuberculosis and other etiologies.

The presentation is a central cord syndrome consisting of a regional dissociated sensory loss (loss of pain and temperature sensation with sparing of touch and vibration) and areflexic weakness in the upper limbs. The sensory deficit has a distribution that is "suspended" over the nape of the neck, shoulders, and upper arms (cape distribution) or in the hands. Most cases begin asymmetrically with unilateral sensory loss in the hands that leads to injuries and burns that are not

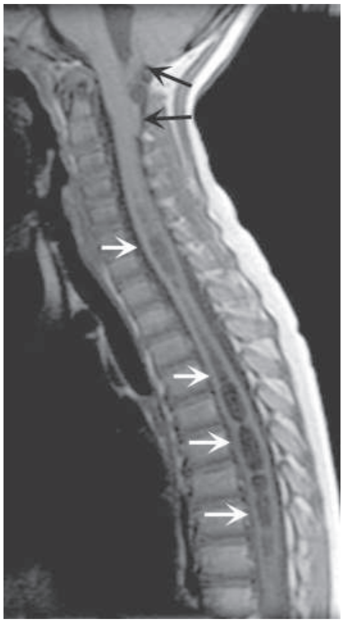

FIGURE 453-7 Magnetic resonance imaging of syringomyelia associated with a Chiari malformation. Sagittal T1-weighted image through the cervical and upper thoracic spine demonstrates descent of the cerebellar tonsils below the level of the foramen magnum (*black arrows*). Within the substance of the cervical and thoracic spinal cord, a cerebrospinal fluid collection dilates the central canal (*white arrows*).

appreciated by the patient. Muscle wasting in the lower neck, shoulders, arms, and hands with asymmetric or absent reflexes in the arms reflects expansion of the cavity in the gray matter of the cord. As the cavity enlarges and compresses the long tracts, spasticity and weakness of the legs, bladder and bowel dysfunction, and Horner's syndrome appear. Some patients develop facial numbness and sensory loss from damage to the descending tract of the trigeminal nerve (C2 level or above). In cases with Chiari malformations, cough-induced headache and neck, arm, or facial pain may be reported. Extension of the syrinx into the medulla, syringobulbia, causes palatal or vocal cord paralysis, dysarthria, horizontal or vertical nystagmus, episodic dizziness or vertigo, and tongue weakness with atrophy.

MRI accurately identifies developmental and acquired syrinx cavities and their associated spinal cord enlargement **(Fig. 453-7)**. Images of the brain and the entire spinal cord should be obtained to delineate the full longitudinal extent of the syrinx, assess posterior fossa structures for the Chiari malformation, and determine whether hydrocephalus is present.

TREATMENT

Syringomyelia

Surgical decompression is the treatment of choice, with mixed results reported in most series. The Chiari tonsillar herniation may be decompressed, generally by suboccipital craniectomy, upper cervical laminectomy, and placement of a dural graft. Fourth ventricular outflow is reestablished by this procedure. If the syrinx cavity is large, some surgeons recommend direct decompression or drainage, but an added benefit of this procedure has not been demonstrated, and complications are common. Shunting of hydrocephalus, when present, generally precedes any attempt to correct the syrinx. Surgery may stabilize the neurologic deficit, and some patients improve. Patients with few symptoms and signs from the syrinx do not require surgery and are followed by serial clinical and imaging examinations.

Syrinx cavities secondary to trauma or infection, if symptomatic, can be treated with a decompression and drainage procedure in which a small shunt is inserted between the cavity and subarachnoid space; alternatively, the cavity can be fenestrated. Cases due to intramedullary spinal cord tumor are generally managed by resection of the tumor.

CHRONIC MYELOPATHY OF MULTIPLE SCLEROSIS

A chronic progressive myelopathy is the most frequent cause of disability in both primary progressive and secondary progressive forms of MS. Involvement is typically bilateral but asymmetric and produces motor, sensory, and bladder/bowel disturbances. Fixed motor disability appears to result from extensive loss of axons in the corticospinal tracts. Diagnosis is facilitated by identification of earlier attacks such as optic neuritis. MRI, CSF, and evoked-response testing are confirmatory. Treatment with ocrelizumab, an anti-CD20 B-cell monoclonal antibody, is effective in patients with primary progressive MS, and disease-modifying therapy is also indicated in patients with secondary progressive MS who have clinical or MRI evidence of active disease. **MS is discussed in Chap. 455.**

SUBACUTE COMBINED DEGENERATION (VITAMIN B_{12} DEFICIENCY)

This treatable myelopathy presents with subacute paresthesias in the hands and feet, loss of vibration and position sensation, and a progressive spastic and ataxic weakness. Loss of reflexes due to an associated peripheral neuropathy in a patient who also has Babinski signs is a helpful diagnostic clue. Optic atrophy and irritability or other cognitive changes may be prominent in advanced cases and are occasionally the presenting symptoms. The myelopathy of subacute combined degeneration tends to be diffuse rather than focal; signs are generally symmetric and reflect predominant involvement of the posterior and lateral tracts, including Romberg sign. Causes include dietary deficiency, especially in vegans, and gastric malabsorption syndromes including pernicious anemia **(Chap. 104)**. The diagnosis is confirmed by the finding of macrocytic red blood cells, a low serum B_{12} concentration, and elevated serum levels of homocysteine and methylmalonic acid. Treatment is by replacement therapy, beginning with 1000 μg of intramuscular vitamin B_{12} daily for 5–7 days and then continued as a once-weekly dose for 4–8 weeks and then as a monthly maintenance dose; oral maintenance with high doses of cyanocobalamin (1–2 mg daily) can also be used for maintenance, as small amounts of vitamin B_{12} are absorbed passively by the gut even in pernicious anemia.

Two closely related conditions deserve mention here. The first is folate deficiency–associated myelopathy, now only rarely seen since widespread programs of dietary fortification with folate have been implemented. A second is due to inhalation with nitrous oxide (laughing gas), an irreversible inhibitor of vitamin B_{12}, which also produces a myelopathy identical to subacute combined degeneration. Exposure to nitrous oxide may occur during dental or surgical procedures or from recreational inhalation ("doing whippets").

HYPOCUPRIC MYELOPATHY

This myelopathy is similar to subacute combined degeneration, except serum levels of B_{12} are normal. Low levels of serum copper are found, and often there is also a low level of serum ceruloplasmin. Some cases follow gastrointestinal procedures, particularly bariatric surgery, that result in impaired copper absorption; others have been associated with excess zinc from health food supplements or, in the past, zinc-containing denture creams, all of which impair copper absorption via induction of metallothionein, a copper-binding protein. Many cases are idiopathic. There is often a coexisting anemia. Improvement or at least stabilization may be expected with reconstitution of copper stores by oral supplementation (2 mg/d).

TABES DORSALIS

The classic syphilitic syndromes of tabes dorsalis and meningovascular inflammation of the spinal cord are now less frequent than in the past but must be considered in the differential diagnosis of spinal cord disorders. The characteristic symptoms of tabes are fleeting and repetitive lancinating pains, primarily in the legs or less often in the back, thorax, abdomen, arms, and face. Ataxia of the legs and gait due to loss of position sense occurs in half of patients. Paresthesias, bladder disturbances, and acute abdominal pain with vomiting (visceral crisis) occur in 15–30% of patients. The cardinal signs of tabes are loss of reflexes in the legs; impaired position and vibratory sense; Romberg sign; and, in almost all cases, bilateral Argyll Robertson pupils, which fail to constrict to light but accommodate. Diabetic polyradiculopathy may simulate this condition. Treatment of tabes dorsalis and other forms of neurosyphilis consists of penicillin G administered intravenously, or intramuscularly in combination with oral probenecid **(Chap. 187)**.

HEREDITARY SPASTIC PARAPLEGIA

Many cases of slowly progressive myelopathy are genetic in origin **(Chap. 448)**. More than 90 different causative loci have been identified, including autosomal dominant, autosomal recessive, and X-linked forms. Especially for the recessive and X-linked forms, a family history of myelopathy may be lacking. Most patients present with almost imperceptibly progressive spasticity and weakness in the legs, usually but not always symmetrical. Sensory symptoms and signs are absent or mild, but sphincter disturbances may be present. In some families, additional neurologic signs are prominent, including nystagmus, ataxia, or optic atrophy. The onset may be as early as the first year of life or as late as middle adulthood. Only symptomatic therapies are available.

PRIMARY LATERAL SCLEROSIS

This is a mid- to late-life-onset degenerative disorder characterized by progressive spasticity with weakness, eventually accompanied by dysarthria and dysphonia; bladder symptoms occur in approximately half of patients. Sensory function is spared. The disorder resembles ALS and is considered a variant of the motor neuron degenerations, but without the characteristic lower motor neuron disturbance and with typically a slower progression. Some cases may represent late-onset cases of hereditary spastic paraplegia, particularly autosomal recessive or X-linked varieties in which a family history may be absent. **(See also Chap. 448.)**

ADRENOMYELONEUROPATHY

This X-linked peroxisomal disorder is a variant of adrenoleukodystrophy (ALD). Most affected males have a history of adrenal insufficiency and then develop a progressive spastic (or ataxic) paraparesis beginning in early or sometimes middle adulthood; some patients also have cerebral involvement and/or a mild peripheral neuropathy. Female heterozygotes may develop a slower, insidiously progressive spastic myelopathy beginning later in adulthood and without adrenal insufficiency. Diagnosis is usually made by demonstration of elevated levels of very-long-chain fatty acids in plasma and in cultured fibroblasts. The responsible gene encodes the adrenoleukodystrophy protein (ALDP), a peroxisomal membrane transporter involved in carrying long-chain fatty acids to peroxisomes for degradation. Corticosteroid replacement is indicated if hypoadrenalism is present. Allogeneic bone marrow transplantation has been successful in slowing progression of cognitive decline in some patients with ALD treated early in their disease but appears to be ineffective for the myelopathy. A preliminary study of leriglitazone, a peroxisome proliferator-activated receptor gamma (PPARγ) agonist, reported suggestive slowing of myelopathic signs in some patients, but the primary endpoint was not met. Nutritional supplements (Lorenzo's oil) have also been attempted for this condition without evidence of efficacy.

CANCER-RELATED SYNDROMES

Cancer-related causes of chronic myelopathy, besides the common neoplastic compressive myelopathy discussed earlier, include radiation injury **(Chap. 95)** and a myelopathy resembling subacute combined degeneration that can follow intrathecal administration of methotrexate (a folate antagonist). Rare paraneoplastic myelopathies are most often associated with lung cancer and anti-amphiphysin (also breast), anti-collapsin response mediator 5 (CRMP5) (also lymphoma), or anti-Hu antibodies **(Chap. 99)**. Another uncommon lymphoma-associated paraneoplastic syndrome is a progressive flaccid paresis with destruction of anterior horn cells. NMO with AQP-4 antibodies **(Chap. 456)** can also rarely be paraneoplastic in origin. Several series have reported cases of myelopathy associated with use of checkpoint inhibitors in cancer treatment; some are associated with development of paraneoplastic or NMO antibodies. Metastases to the cord are probably more common than any of these disorders in patients with cancer.

OTHER CHRONIC MYELOPATHIES

Tethered cord syndrome is a developmental disorder of the lower spinal cord and nerve roots that rarely presents in adulthood as low back pain accompanied by a progressive lower spinal cord and/or nerve root syndrome. Some patients have a leg or foot deformity indicating a longstanding process, and in others, a dimple, patch of hair, or sinus tract on the skin overlying the lower back is the clue to a congenital lesion. Diagnosis is made by MRI, which demonstrates a low-lying conus medullaris and thickened filum terminale. The MRI may also reveal diastematomyelia (division of the lower spinal cord into two halves), lipomas, cysts, or other congenital abnormalities of the lower spine coexisting with the tethered cord. Treatment is with surgical release.

There are a number of rare toxic causes of spastic myelopathy, including lathyrism due to ingestion of chickpeas containing the excitotoxin β-*N*-oxalylamino-L-alanine (BOAA), seen primarily in the developing world or during famines, and Konzo due to ingestion of the cyanogen-containing casava plant found in sub-Saharan Africa.

Often, a cause of intrinsic myelopathy can be identified only through periodic reassessment.

REHABILITATION OF SPINAL CORD DISORDERS

The prospects for recovery from an acute destructive spinal cord lesion fade after ~6 months. There are currently no effective means to promote repair of injured spinal cord tissue; promising but entirely experimental approaches include the use of factors that influence reinnervation by axons of the corticospinal tract, nerve and neural sheath graft bridges, forms of electrical stimulation at the site of injury, and the local introduction of stem cells. The disability associated with irreversible spinal cord damage is determined primarily by the level of the lesion and by whether the disturbance in function is complete or incomplete **(Table 453-4)**. Even a complete high cervical cord lesion may be compatible with a productive life. The primary goals are development of a rehabilitation plan framed by realistic expectations and attention to the neurologic, medical, and psychological complications that commonly arise.

Many of the usual symptoms associated with medical illnesses, especially somatic and visceral pain, may be lacking because of the destruction of afferent pain pathways. Unexplained fever, worsening of spasticity, or deterioration in neurologic function should prompt a search for infection, thrombophlebitis, or an intraabdominal pathology. The loss of normal thermoregulation and inability to maintain normal body temperature can produce recurrent fever (*quadriplegic fever*), although most episodes of fever are due to infection of the urinary tract, lung, skin, or bone.

Bladder dysfunction generally results from loss of supraspinal innervation of the detrusor muscle of the bladder wall and the

TABLE 453-4 Expected Neurologic Function Following Complete Cord Lesions

LEVEL	SELF-CARE	TRANSFERS	MAXIMUM MOBILITY
High quadriplegia (C1–C4)	Dependent on others; requires respiratory support	Dependent on others	Motorized wheelchair
Low quadriplegia (C5–C8)	Partially independent with adaptive equipment	May be dependent or independent	May use manual wheelchair, drive an automobile with adaptive equipment
Paraplegia (below T1)	Independent	Independent	Ambulates short distances with aids

Source: Adapted from JF Ditunno, CS Formal: Chronic spinal cord injury. N Engl J Med 330:550, 1994.

sphincter musculature. Detrusor spasticity is treated with anticholinergic drugs (oxybutynin, 2.5–5 mg qid) or tricyclic antidepressants with anticholinergic properties (imipramine, 25–200 mg/d). Failure of the sphincter muscle to relax during bladder emptying (urinary dyssynergia) may be managed with the α-adrenergic blocking agent terazosin hydrochloride (1–2 mg tid or qid), with intermittent catheterization, or, if that is not feasible, by use of a condom catheter in men or a permanent indwelling catheter. Surgical options include the creation of an artificial bladder by isolating a segment of intestine that can be catheterized intermittently (enterocystoplasty) or can drain continuously to an external appliance (urinary conduit). Bladder areflexia due to acute spinal shock or conus lesions is best treated by catheterization. Bowel regimens and disimpaction are necessary in many patients to ensure at least biweekly evacuation and avoid colonic distention or obstruction.

Patients with acute cord injury are at risk for venous thrombosis and pulmonary embolism. Use of calf-compression devices and anticoagulation with low-molecular-weight heparin are recommended. In cases of persistent paralysis, anticoagulation should probably be continued for 3 months.

Prophylaxis against decubitus ulcers should involve frequent changes in position in a chair or bed, the use of special mattresses, and cushioning of areas where pressure sores often develop, such as the sacral prominence and heels. Early treatment of ulcers with careful cleansing, surgical or enzyme debridement of necrotic tissue, and appropriate dressing and drainage may prevent infection of adjacent soft tissue or bone.

Spasticity is aided by stretching exercises to maintain mobility of joints. Drug treatment is effective but may result in reduced function, as some patients depend on spasticity as an aid to stand, transfer, or walk. Baclofen (up to 240 mg/d in divided doses) is effective; it acts by facilitating γ-aminobutyric acid–mediated inhibition of motor reflex arcs. Diazepam acts by a similar mechanism and is useful for leg spasms that interrupt sleep (2–4 mg at bedtime). Tizanidine (2–8 mg tid), an α_2 adrenergic agonist that increases presynaptic inhibition of motor neurons, is another option. For nonambulatory patients, the direct muscle inhibitor dantrolene (25–100 mg qid) may be used, but it is potentially hepatotoxic. In refractory cases, intrathecal baclofen administered via an implanted pump, botulinum toxin injections, or dorsal rhizotomy may be required to control spasticity.

Despite the loss of sensory function, many patients with spinal cord injury experience chronic pain sufficient to diminish their quality of life. Randomized controlled studies indicate that gabapentin or pregabalin is useful in this setting. Epidural electrical stimulation and intrathecal infusion of pain medications have been tried with some success. **Management of chronic pain is discussed in Chap. 14.**

A paroxysmal autonomic hyperreflexia may occur following lesions above the major splanchnic sympathetic outflow at T6. Headache, flushing, and diaphoresis above the level of the lesion, as well as hypertension with bradycardia or tachycardia, are the major symptoms. The trigger is typically a noxious stimulus—for example, bladder or bowel distention, a urinary tract infection, or a decubitus ulcer—below the level of the cord lesion. Treatment consists of removal of offending stimuli; ganglionic blocking agents (mecamylamine, 2.5–5 mg) or other short-acting antihypertensive drugs are useful in some patients.

Emerging neuro-therapeutic technologies, including rehabilitation robotics and brain-machine interfaces, offer real hope to advance prospects for full and productive lives in patients with disabling myelopathies **(Chap. 500)**.

FURTHER READING

Badhiwala JH et al: Degenerative cervical myelopathy—update and future directions. Nat Rev Neurol 16:108, 2020.

Bradshaw MJ et al: Neurosarcoidosis: Pathophysiology, diagnosis, and treatment. Neurol Neuroimmunol Neuroinflamm 8:e1084, 2021.

Garg RK et al: Spinal cord involvement in COVID-19: A review. J Spinal Cord Med 46:390, 2023.

Gritsch D, Valencia-Sanchez C: Drug-related immune-mediated myelopathies. Front Neurol 13:1003270, 2022.

Hassing LT et al: Nitrous-oxide-induced polyneuropathy and subacute combined degeneration of the spine: Clinical and diagnostic characteristics in 70 patients, with focus on electrodiagnostic studies. Eur J Neurol 31:e16076, 2024.

Holroyd KB, Berkowitz AL: Metabolic and toxic myelopathies. Continuum (Minneap Minn) 30:199, 2024.

Jain S et al: Transverse myelitis associated with systemic lupus erythematosus (SLE-TM): A review article. Lupus 32:1033, 2023.

Kühl JS et al: Long-term outcomes of allogeneic haematopoietic stem cell transplantation for adult cerebral X-linked adrenoleukodystrophy. Brain 140:953, 2017.

Montalvo M, Flanagan EP: Paraneoplastic/autoimmune myelopathies. Handb Clin Neurol 200:193, 2024.

Ropper AE, Ropper AH: Acute spinal cord compression. N Engl J Med 376:1358, 2017.

Stenimahitis V et al: Long-term outcomes after periprocedural and spontaneous spinal cord infarctions: A population-based cohort study. Neurology 101:e114, 2023.

Yáñez ML et al: Diagnosis and treatment of epidural metastases. Cancer 123:1106, 2017.

454 Concussion and Other Traumatic Brain Injuries

Geoffrey T. Manley, Benjamin L. Brett, Michael McCrea

Traumatic brain injury (TBI) represents a significant global public health problem. In the United States, estimates of the frequency of TBI range between 2.5 and 4.8 million cases per year, depending on the study and methods used to define and include cases. Age-specific rates show a bimodal distribution, with highest risk in younger individuals and older adults. The most common mechanism of injury in the young is motor vehicle accidents and is more common in men, whereas in older adults, falls are the major cause of injury and are more likely to occur in women. Nonfatal TBI-related hospitalization rates are comparable across non-Hispanic white, non-Hispanic black, and Hispanic persons with age adjustment.

TBI imposes substantial demands on health care systems. Worldwide, at least 10 million TBIs are serious enough to result in death or hospitalization, producing a global economic burden of $400 billion annually. In the United States, the estimated annual cost is >$76 billion. Due to advances in medical care and other factors, more people are surviving TBI than ever before. Brain injury accounts for more lost productivity at work among Americans than any other form of injury. An estimated 5.3 million Americans are living with significant disabilities resulting from TBI that complicate their return to a full and productive life. Increased media attention to military and sports-related TBI has highlighted the growing concern that injuries that were previously dismissed can have lifelong consequences for some individuals.

Head injuries are so common that almost all physicians will be called upon to provide some aspect of immediate care or to see patients who are suffering from various sequelae. Patients and their families initially need education regarding the natural history of TBI along with treatment of acute symptoms such as headache. Continued follow-up is important to ensure that the sequelae experienced by some patients—such as the persistent postconcussion symptoms (PPCS) of headache, disturbances in balance, depression, and sleep disorder—are identified and treated appropriately. Effective management of TBI and its consequences often requires a coordinated multidisciplinary care team.

DEFINITION AND CLASSIFICATION

TBI is commonly defined as an alteration in brain function, or other evidence of brain pathology, caused by an external force, and characterized by the following: (1) any period of loss or decreased level of consciousness (LOC), (2) any loss of memory for events immediately before (retrograde) or after (posttraumatic) the injury, (3) any neurologic deficits, and/or (4) any alteration in mental state at the time of injury.

Evidence of TBI can include visual, neuroradiologic, or laboratory confirmation of damage to the brain, but TBI is more often diagnosed on the basis of acute clinical criteria. In addition to standard computed tomography (CT) imaging and conventional clinical magnetic resonance imaging (MRI), advanced MRI imaging (functional MRI, cerebral blood flow, diffusion MRI) techniques show increasing sensitivity, and it is likely that sensitive blood-based biomarkers will play an increasingly important role in the diagnosis and treatment of these patients (described below).

Mechanisms of TBI Common mechanisms of TBI include the head being struck by an object, the head striking an object, the brain undergoing an acceleration/deceleration movement, a foreign body penetrating the brain, or forces generated from events such as a blast or explosion. Unintentional falls and motor vehicle crashes have historically been cited as the most common cause of TBI. All forms of transportation, however, are common causes of TBI, including motorcycle crashes, bicycle accidents, skateboarding, and pedestrian injuries. The other leading causes of TBI are falls, assaults, and sports, with varied frequency across the lifespan. Certainly, there has been an increased focus on the high frequency of mild TBI (mTBI), often referred to as concussion, encountered by athletes participating in contact and collision sports at all competitive levels, as well as the potential short-term effects and long-term risks associated with sport-related concussion.

Classification of TBI Severity Numerous systems have been developed over the years to define and classify TBI severity along a continuum from mild to moderate to severe. These systems are usually most applicable to closed head injuries. In nearly all classification systems, TBI severity is graded based on acute injury characteristics rather than postacute injury status, as other factors can intervene to influence functional outcome. This can be problematic, as some patients with severe TBI will have a full recovery and some with mild TBI will be left with lifetime disability. Historically, the presence and duration of unconsciousness and amnesia have been the main points of distinction along the gradient of TBI severity. Efforts to develop more advanced classification systems that incorporate patho-anatomical features reflecting biological response to injury, such as blood-based biomarkers, as well as socioenvironmental factors that may influence ultimate prognosis are currently underway. A recent working group convened by the National Institutes of Health has proposed a new framework for TBI characterization that includes clinical features, blood-based biomarkers, and imaging features, along with modifiers. As this more precise nomenclature is validated, it is anticipated that it will replace the current approach of characterizing patients based solely on their clinical exam and symptoms.

The Glasgow Coma Scale (GCS) is the most recognized and widely used method for grading TBI severity. The GCS provides a practical indicator of gross neurologic status by assessing motor function, verbal responses, and the patient's ability to open his or her eyes voluntarily or in response to external commands and stimuli. The grading is applied to the best response that can be elicited from the patient at the time of assessment, preferably before any paralyzing or sedating medication is administered or the patient is intubated, as these interventions confound interpretation of the score. The GCS assessment produces scores ranging from 3 to 15 **(Table 454-1)**.

Upon the 40th anniversary of the GCS in 2014, the wording for responses was revised, and recommendations were made to improve its utility. Importantly, individual patients are best described by the three components of the coma scale (eye, verbal, motor, e.g., E3V4M6); the derived total coma score (e.g., 13) is less informative and should only be used to characterize groups of patients.

Several injury-classification systems have been developed to go beyond GCS score or acute injury characteristics and incorporate chief signs and symptoms in defining mTBI. The use of multiple severity indicators is intended to improve sensitivity in the detection of mTBI (GCS 13–15), while also taking into consideration traditional acute injury characteristics that have been presumed to predict outcome following mild and moderate brain injury. LOC and posttraumatic amnesia (PTA) remain the most common injury characteristics referenced in these classification systems. In the case of moderate (GCS 9–12) and severe (GCS 3–8) TBI, GCS score and the duration of LOC and PTA can be robust predictors of long-term outcome and morbidity. In cases of mTBI, however, while PTA and LOC are important indicators of acute injury, they are less predictive of eventual recovery time and outcome, particularly within sport-related concussion.

TABLE 454-1 Glasgow Coma Scale

EYE OPENING (E)		VERBAL RESPONSE (V)	
Spontaneous	4	Oriented	5
To speech	3	Confused	4
To pressure	2	Words	3
None	1	Sounds	2
		None	1
Best Motor Response (M)			
Obeying commands	6		
Localizing	5		
Normal flexion	4		
Abnormal flexion	3		
Extension	2		
None	1		

Note: Revised GCS (2014).

Source: Reproduced with permission from G Teasdale et al: The Glasgow Coma Scale at 40 years: Standing the test of time. Lancet Neurol 13:844, 2014.

TBI TYPES AND PATHOLOGIES

Mild TBI (Concussion) It is estimated that 70–90% of all treated TBIs are mild in severity based on traditional case definitions and acute injury characteristics, with most reported estimates in the order of 85%. The published figures likely underrepresent the true incidence of mTBI because of variable case definitions and heterogeneous methods. Moreover, because a subgroup of individuals with milder brain injuries does not seek medical attention, epidemiologic studies that depend on hospital-based data also underestimate the true incidence. In fact, it has been estimated that current data collection sources and methods may only capture one of every nine mTBI/concussions in the United States.

The term *concussion*, while popular, is vague and is not based on widely accepted objective criteria, resulting in multiple definitions from various groups. There has been debate as to whether concussion is part of the TBI spectrum or a separate entity. The Concussion in Sports Group has concluded that "concussion is a traumatic brain injury" as part of the consensus statement definition of the injury. By firmly placing concussion in the spectrum of TBI, the underlying pathophysiologic processes common to all TBI presentations can now be considered together.

CT imaging is often normal in this population. However, emerging evidence indicates that 3-tesla (3T) MRI scans with greater image resolution can identify pathology consistent with acute brain injury such as contusion and microhemorrhage. When patients with mTBI have CT and/or MRI abnormalities, they are often referred to as having complicated mTBI and are more likely to have an unfavorable outcome.

SKULL FRACTURE, EXTRA-AXIAL HEMATOMA, CONTUSION, AND AXONAL INJURY

SKULL FRACTURE

A blow to the skull that exceeds the elastic tolerance of the bone causes a fracture. Intracranial lesions accompany roughly two-thirds of skull fractures, and the presence of a fracture increases many-fold the chances of an underlying subdural or epidural hematoma. Consequently, fractures are primarily markers of the site and severity of injury. If the underlying arachnoid membrane has been torn, fractures also provide potential pathways for entry of bacteria to the cerebrospinal fluid (CSF) with a risk of meningitis and for leakage of CSF outward through the dura. If there is leakage of CSF, severe orthostatic headache results from lowered pressure in the spinal fluid compartment.

Most fractures are linear and extend from the point of impact toward the base of the skull. Basilar skull fractures are often extensions of adjacent linear fractures over the convexity of the skull but may occur independently owing to stresses on the floor of the middle cranial fossa or occiput. Basilar fractures are usually parallel to the petrous bone or along the sphenoid bone and directed toward the sella turcica and ethmoidal groove. Although most basilar fractures are uncomplicated, they can cause CSF leakage, pneumocephalus, and delayed cavernous-carotid fistulas. Hemotympanum (blood behind the tympanic membrane), ecchymosis over the mastoid process (Battle sign), and periorbital ecchymosis ("raccoon sign") are clinical signs associated with basilar fractures.

EPIDURAL AND SUBDURAL HEMATOMAS

Hemorrhages between the dura and skull (epidural) or beneath the dura (subdural) have characteristic clinical and imaging features. They are sometimes associated with underlying brain contusions and other injuries, often making it difficult to determine the relative contribution of each component to the clinical state. The mass effect of raised intracranial pressure (ICP) caused by these hematomas can be life threatening, making it imperative to identify them rapidly by CT or MRI scan and to surgically remove them when appropriate.

Epidural Hematoma (Fig. 454-1) These highly dangerous lesions usually arise from an injury to a meningeal arterial vessel and evolve rapidly. They are often accompanied by a "lucid interval" of several minutes to hours prior to neurologic deterioration. They occur in up to 10% of cases of severe head injury but are less often associated with underlying cortical damage compared to subdural hematomas. Rapid surgical evacuation and ligation or cautery of the damaged vessel, usually the middle meningeal artery that has been lacerated by an overlying skull fracture, is indicated. If recognized and treated rapidly, patients often have a favorable outcome.

Acute Subdural Hematoma (Fig. 454-2) Direct cranial trauma may be minor and is not always required for acute subdural hemorrhage to occur, especially in the elderly and those taking anticoagulant medications. Acceleration forces alone, as from whiplash, are sometimes sufficient to produce subdural hematoma. Up to one-third of patients have a lucid interval lasting minutes to hours before coma supervenes, but most are drowsy or comatose from the moment of injury. A unilateral headache and slightly enlarged pupil on the side of the hematoma are frequently, but not invariably, present. Small subdural hematomas may be asymptomatic and usually do not require surgical evacuation if they do not enlarge. Stupor or coma, hemiparesis, and unilateral pupillary enlargement are signs of larger hematomas. The bleeding that causes larger subdural hematomas is primarily venous in origin, although arterial bleeding sites are sometimes found at operation, and a few large hematomas have a purely arterial origin. In an acutely deteriorating patient, an emergency craniotomy is required. In contrast to epidural hematomas, there is significant morbidity and mortality associated with acute subdural hematomas that require surgery.

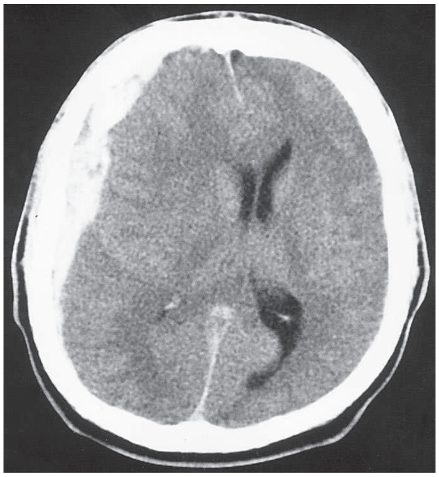

FIGURE 454-2 Acute subdural hematoma. Noncontrast computed tomography scan reveals a hyperdense clot that has an irregular border with the brain and causes more horizontal displacement (mass effect) than might be expected from its thickness. The disproportionate mass effect is the result of the large rostral-caudal extent of these hematomas. Compare to Fig. 454-1.

Chronic Subdural Hematoma A subacutely evolving syndrome due to subdural hematoma occurs days or weeks after injury with drowsiness, headache, confusion, or mild hemiparesis, usually in the elderly with age-related atrophy and often after only minor or unnoticed trauma. On imaging studies, chronic subdural hematomas appear as crescentic clots over the convexity of one or both hemispheres, most commonly in the frontotemporal region (Fig. 454-3). A history of trauma may or may not be elicited in relation to chronic subdural hematoma; the injury may have been trivial and forgotten, particularly in the elderly and those with clotting disorders. Headache is common

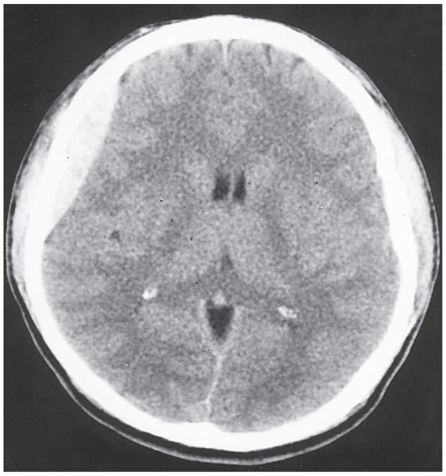

FIGURE 454-1 Acute epidural hematoma. The tightly attached dura is stripped from the inner table of the skull, producing a characteristic lenticular-shaped hemorrhage on noncontrast computed tomography scan. Epidural hematomas are usually caused by tearing of the middle meningeal artery following fracture of the temporal bone.

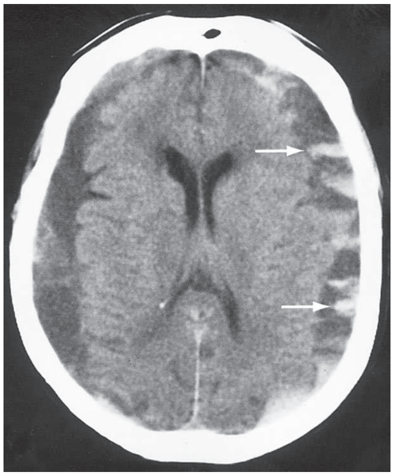

FIGURE 454-3 Computed tomography scan of chronic bilateral subdural hematomas of different ages. The collections began as acute hematomas and have become hypodense in comparison to the adjacent brain after a period during which they were isodense and difficult to appreciate. Some areas of resolving blood are contained on the more recently formed collection on the left (*arrows*).

but not invariable. Additional features that may appear weeks later include slowed thinking, vague change in personality, seizure, or a mild hemiparesis. The headache typically fluctuates in severity, sometimes with changes in head position. Drowsiness, inattentiveness, and incoherence of thought are generally more prominent than focal signs such as hemiparesis. Rarely, chronic hematomas cause brief episodes of hemiparesis or aphasia that are indistinguishable from transient ischemic attacks.

CT without contrast initially shows a low-density mass over the convexity of the hemisphere. Between 2 and 6 weeks after the initial bleeding, the clot becomes isodense compared to adjacent brain and may be inapparent. Many subdural hematomas that are several weeks in age contain areas of blood and intermixed serous fluid. Infusion of contrast material demonstrates enhancement of the vascular fibrous capsule surrounding the collection. MRI reliably identifies both subacute and chronic hematomas.

Clinical observation coupled with serial imaging is a reasonable approach to patients with few symptoms and small chronic subdural collections that do not cause mass effect. Treatment with surgical evacuation through burr holes is usually successful, if a cranial drain is used postoperatively. The fibrous membranes that grow from the dura and encapsulate the collection may require removal with a craniotomy to prevent recurrent fluid accumulation.

TRAUMATIC SUBARACHNOID HEMORRHAGE

Subarachnoid hemorrhage (SAH) is common in TBI. Rupture of small cortical arteries or veins can cause bleeding into the subarachnoid space. Traumatic SAH is often seen in the sulci and is frequently the only radiographic finding on CT following mild TBI. SAH occurs diffusely after severe TBI and confers an increase in mortality. In mild TBI, SAH provides an objective imaging biomarker for TBI and, in some patients, is associated with unfavorable outcomes.

CONTUSION (FIG. 454-4)

A surface bruise of the brain, or contusion, consists of varying degrees of petechial hemorrhage, edema, and tissue destruction. Contusions and deeper hemorrhages result from mechanical forces that displace and compress the hemispheres forcefully and by deceleration of the brain against the inner skull, either under a point of impact (coup lesion) or, as the brain swings back, in the antipolar area (contrecoup lesion). Trauma sufficient to cause prolonged unconsciousness usually produces some degree of contusion. Blunt deceleration impact, as occurs against an automobile dashboard or from falling forward onto a hard surface, causes contusions on the orbital surfaces of the frontal lobes and the anterior and basal portions of the temporal lobes. With lateral forces, as from impact on an automobile door frame, contusions are situated on the lateral convexity of the hemisphere. The clinical signs of contusion are determined by the location and size of the lesion; often, there are no focal abnormalities with a routine neurologic exam, but these injured regions are later the sites of gliotic scars that may produce seizures. A hemiparesis or gaze preference is fairly typical of moderately sized contusions. Large bilateral contusions produce stupor with extensor posturing, while those limited to the frontal lobes cause a taciturn state. Contusions in the temporal lobe may cause delirium or an aggressive, combative syndrome. Torsional or shearing forces within the brain can cause hemorrhages of the basal ganglia and other deep regions. Large contusions and hemorrhages after minor trauma should raise concerns for coagulopathy due to an underlying disease or more commonly anticoagulant therapy.

Acute contusions are easily visible on CT and MRI scans, appearing as inhomogeneous hyperdensities on CT and as hyperintensities on T2 and fluid-attenuated inversion recovery (FLAIR) MRI sequences; there is usually surrounding localized brain edema and some subarachnoid bleeding. Blood in the CSF due to trauma may provoke a mild inflammatory reaction. Over a few days, contusions acquire a surrounding contrast enhancement and edema that may be mistaken for tumor or abscess.

AXONAL INJURY (FIG. 454-5)

Traumatic axonal injury (TAI) is one of the most common injuries after TBI. There is disruption, or shearing, of axons at the time of impact, and this is associated with microhemorrhages. It occurs following high-speed deceleration injuries, such as motor vehicle collisions (see Johnson et al, 2013, in Further Readings). The presence of four or more areas of TAI is called diffuse axonal injury (DAI) and, when widespread, has been proposed to explain persistent coma and the vegetative state after TBI (Chap. 30). Only severe TAI lesions that contain substantial blood are visualized by CT, usually in the corpus callosum and centrum semiovale. More commonly, the CT will be negative for TAI, but subsequent MRI, particularly gradient-echo or susceptibility-weighted imaging, will show hemosiderin deposits reflective of microhemorrhages in addition to the axonal damage on diffusion sequences. Traditionally, TAI and DAI have been considered as sequelae much more likely to result from moderate and severe injuries. Accumulating evidence has demonstrated that diffuse white matter abnormalities purportedly reflective of axonal injury, such as changes in microstructure and neurite density, are quite common in mild TBI as well. The degree of these changes correlates with metrics of injury severity (e.g., symptom burden) and recovery duration.

CRANIAL NERVE INJURIES

The cranial nerves most often injured with TBI are the olfactory, optic, oculomotor, and trochlear nerves; the first and second branches of the trigeminal nerve; and the facial and auditory nerves. Anosmia and an apparent loss of taste (actually a loss of perception of aromatic flavors,

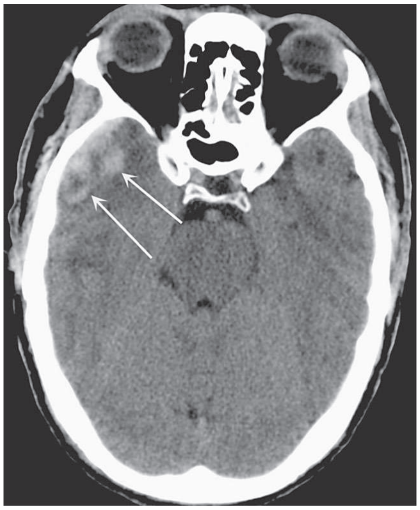

FIGURE 454-4 Traumatic cerebral contusion. Noncontrast computed tomography scan demonstrating a hyperdense hemorrhagic region in the anterior temporal lobe.

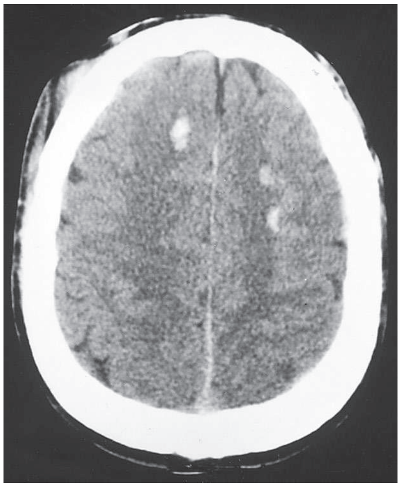

FIGURE 454-5 Multiple small areas of hemorrhage and tissue disruption in the white matter of the frontal lobes on noncontrast computed tomography scan. These appear to reflect an extreme type of the diffuse axonal shearing lesions that occur with closed head injury.

with retained elementary taste perception) occur in ~10% of persons with serious head injuries, particularly from falls on the back of the head. This is the result of displacement of the brain and shearing of the fine olfactory nerve filaments that course through the cribriform bone. At least partial recovery of olfactory and gustatory function is expected, but if bilateral anosmia persists for several months, the prognosis is poor. Partial optic nerve injuries from closed trauma result in blurring of vision, central or paracentral scotomas, or sector defects. Direct orbital injury may cause short-lived blurred vision for close objects due to reversible iridoplegia. Diplopia limited to downward gaze and corrected when the head is tilted away from the side of the affected eye indicates trochlear (fourth nerve) nerve damage. It occurs frequently as an isolated problem after minor head injury or may develop for unknown reasons after a delay of several days. Facial nerve injury caused by a basilar fracture is present immediately in up to 3% of severe injuries; it may also be delayed for 5–7 days. Fractures through the petrous bone, particularly the less common transverse type, are liable to produce facial palsy. Delayed facial palsy occurring up to a week after injury, the mechanism of which is unknown, has a good prognosis. Injury to the eighth cranial nerve from a fracture of the petrous bone causes loss of hearing, vertigo, and nystagmus immediately after injury. Deafness from eighth nerve injury is rare and must be distinguished from blood in the middle ear or disruption of the middle ear ossicles. Dizziness, tinnitus, and high-tone hearing loss occur from cochlear concussion.

SEIZURES

Convulsions are surprisingly uncommon immediately after TBI, but a brief period of tonic extensor posturing or a few clonic movements of the limbs just after the moment of impact can occur. However, the cortical scars that evolve from contusions are highly epileptogenic and may later manifest as seizures, even after many months or years **(Chap. 436)**. The severity of injury roughly determines the risk of future seizures. It has been estimated that 17% of individuals with brain contusion, subdural hematoma, or prolonged LOC will develop a seizure disorder and that this risk extends for an indefinite period of time, whereas the risk is ≤2% after mild injury. The majority of convulsions in the latter group occur within 5 years of injury but may be delayed for decades. Penetrating injuries have a much higher rate of subsequent epilepsy.

CLINICAL SYNDROMES AND TREATMENT OF HEAD INJURY

CONCUSSION/MILD TBI

The patient who has briefly lost consciousness or been stunned after a minor head injury usually becomes fully alert and attentive within minutes but may complain of headache, dizziness, faintness, nausea, a single episode of emesis, difficulty with concentration, a brief amnestic period, or slight blurring of vision. This typical concussion syndrome has a good prognosis with little risk of subsequent deterioration. Children are particularly prone to drowsiness, vomiting, and irritability, symptoms that are sometimes delayed for several hours after apparently minor injuries. Vasovagal syncope that follows injury may cause undue concern. Generalized or frontal headache is common in the following days. It may be migrainous (throbbing and hemicranial) in nature or aching and bilateral. After several hours of observation, patients with minor injury may be accompanied home and observed for a day by a family member or friend, with written instructions to return if symptoms worsen.

Persistent severe headache and repeated vomiting in the context of normal alertness and no focal neurologic signs is usually benign, but CT should be obtained and a longer period of observation is appropriate. The decision to perform imaging tests also depends on clinical signs that indicate that the impact was severe (e.g., persistent confusion, repeated vomiting, palpable skull fracture); the presence of other serious bodily injuries, an underlying coagulopathy, or age >65 years; and on the degree of surveillance that can be anticipated after discharge. Guidelines have also indicated that older age (>65 years), two or more episodes of vomiting, >30 min of retrograde or persistent anterograde amnesia, seizure, and concurrent drug or alcohol intoxication are sensitive (but not specific) indicators of intracranial hemorrhage that justify CT scanning.

Though not incorporated into conventional clinical practice guidelines, growing evidence suggests that MRI improves sensitivity for detection of small intracranial hemorrhages and other lesions in mild TBI patients, particular among those with negative findings on CT. Specifically, intracranial abnormalities are fairly common on MRI (27%) in CT-negative patients. Further, acute MRI findings have prognostic utility in predicting recovery and outcome after mTBI/concussion (e.g., risk of functional impairment, time to return to activity).

Blood-based (serum and plasma) biomarkers of astrocyte damage/astrogliosis (glial fibrillary acidic protein [GFAP]) and neuronal injury (ubiquitin carboxy-terminal hydrolase L1 [UCHL1]) also hold promise in improving detection and outcome prediction across the full spectrum of TBI. With development and regulatory approval of new rapid assay systems, these biomarkers can now be used for real-time point-of-care assessment; GFAP in particular has high discriminant ability to detect intracranial abnormalities, as well as potential to differentiate CT+, CT–/MRI+, and CT–/MRI– patients. Similar to MRI, emerging biomarkers appear to have not only diagnostic but also prognostic utility in predicting the trajectory of recovery and functional impairments weeks and months after TBI.

SPORT-RELATED CONCUSSION

Concussion is a frequent injury in contact and collision sports (e.g., football, hockey, wrestling) at all levels of participation, including youth sports. Head injury associated with sport and recreational activity accounts for 45% of TBI-related emergency department visits in children age 17 years and under. Over the last decade, data from the Centers for Disease Control and Prevention indicate a 27% decrease in emergency department visits for sport- and recreation-related TBI in the United States between 2012 and 2018, with a specific decline in contact sport–related visits by 32%. Given that national and state surveillance systems continue to report increased sport-related concussion rates over the same time period, it could be inferred that diagnosis and management of sport-related concussion outside of the emergency department have increased.

The natural history of clinical recovery following sport-related concussion has been a subject of substantial ongoing research. Recent large prospective studies have reinforced earlier indications that the acute recovery is favorable. For example, a longitudinal study of over 34,000 collegiate athletes (1751 who experienced concussion) from 22 institutions found that median time to symptom resolution was 6.4 days and median time for return to sport was 12.8 days. By 1 month after injury, 92% of athletes had experienced resolution of symptoms and 85% had been cleared to return to sport. Across several studies, postinjury symptom burden is the most robust predictor of recovery and risk of prolonged symptoms. Other injury/postinjury factors associated with longer return to sport time were continued play following injury and access to health care providers. Several other prospective studies have replicated that the overwhelming majority of athletes across ages (pediatric to adult) achieve a complete recovery in symptoms, cognitive functioning, postural stability, and other functional impairments over a period of 1–4 weeks following concussion.

There continues to be a growing focus toward a more rapid return to activity and early rehabilitation following injury. Specifically, while experts agree that initial rest after injury is beneficial for recovery, extended inactivity beyond 5 days can be detrimental and increase risk for protracted recovery. Rather, active rehabilitation involving supervised subthreshold exercise has been shown to decrease the duration of symptoms and reduce risk of a protracted recovery. Cervicovestibular rehabilitation has been recently recommended for those who experience neck pain, dizziness, or headaches for >10 days after injury.

Preliminary evidence suggests that earlier return to learn/school is associated with faster injury recovery, and there has been an increased focus on facilitating a faster and optimal return to learn/school through accommodations. This could include, but is not limited to, a shortened schedule or built-in breaks for fatigue, opportunity for a quiet work environment to prevent headaches or attentional difficulties, and setting expectations for course work to reduce anxiety, among other accommodations. Most patients (>90%) experience a full return to learn without accommodations by 10 days.

There are a small, select percentage of athletes who remain symptomatic or impaired on functional testing well beyond the window of recovery commonly reported in group studies. The greatest challenge arguably still facing sport medicine clinicians and public health experts is how to most effectively manage and reduce risk in this subset of athletes who do not follow the "typical" course of recovery. The precise likelihood that an athlete will not follow the typical course of rapid, spontaneous recovery and instead exhibit prolonged postconcussive symptoms or other functional impairments after concussion remains unclear. In addition to the injury-related or postinjury factors identified above, a preinjury mental health disorder, migraine, and prior concussion have been consistently associated with the potential for prolonged recovery.

Following acute concussion, multimodal advanced neuroimaging has demonstrated a variety of changes, including decreased cerebral blood flow, increased global and local functional connectivity, and alterations in white matter microstructure reflecting axonal organization. In general, these metrics correlate with measures of injury severity, and resolution of these changes tends to parallel clinical recovery. However, a number of studies have shown that slight changes on advanced multimodal imaging can persist even after symptoms have fully resolved, supporting the concept that the "tail" of neurobiologic recovery may extend beyond the time course of apparent clinical recovery.

In the current absence of adequate data, a commonsense approach to athletic concussion has been to remove the individual from play immediately and avoid contact sports for at least several days after a mild injury, and for a longer period if there are more severe injuries or if there are protracted neurologic symptoms such as headache and difficulty concentrating. No individual should return to play unless all concussion-related symptoms have resolved and an assessment has been made by a health care professional who has experience with treatment of concussion. Validated symptom inventories, such as the Rivermead Post-Concussion Symptom Questionnaire **(Table 454-2)**, have been developed to aid clinicians with recording and quantifying the diverse range of physical, cognitive, and behavioral symptoms that can occur following concussion. In addition to characterizing the constellation of acute symptoms and their severity, symptom inventories can be beneficial to track the course and resolution of symptoms through recovery. Differentiating concussion-related symptoms from factors that may be also influencing endorsement (e.g., preinjury mood disorders) is an important component of managing recovery from sport-related concussion. Once cleared, the individual can then begin a graduated program of increasing activity. These guidelines are designed in part to avoid a perpetuation of symptoms but also to prevent the rare *second-impact syndrome*, in which diffuse and fatal cerebral swelling follows a second minor head injury.

TABLE 454-2 Review of Concussion Symptoms

PHYSICAL	COGNITIVE	BEHAVIORAL
Headaches	Forgetfulness or poor memory	Being irritable, easily angered
Dizziness	Poor concentration	Feeling depressed or tearful
Nausea and/or vomiting	Taking longer to think	Feeling frustrated or impatient
Noise sensitivity		Restlessness
Sleep disturbance		
Fatigue		
Blurred vision		
Light sensitivity		
Double vision		

Note: Items were adapted from the Rivermead Post-Concussion Symptom Questionnaire. Each item is rated on a 5-point Likert scale (0–4), as follows: 0 = Not experienced at all; 1 = No more of a problem now than preinjury; 2 = A mild problem; 3 = A moderate problem; 4 = A severe problem. Total scores can range from 0–64.

POSTCONCUSSIVE STATES

There has been a recent paradigm shift from the term *postconcussion syndrome* (PCS) to *persistent postconcussion symptoms* (PPCS) in an effort to avoid classifying prolonged sequelae under a nonspecific diagnostic label and a shifting focus toward identifying and targeting treatment toward direct areas of persistent difficulty. PPCS following mild TBI could include symptoms of fatigue, dizziness, headache, and/or difficulty in concentration. Management is difficult and generally requires identification and management of the specific problem or problems that are most troubling to the individual. A clear explanation and education around the symptoms that may follow concussion have been shown to reduce subsequent complaints. Care is taken to avoid prolonged use of drugs that produce dependence. Headache may initially be treated with acetaminophen and small doses of amitriptyline. Vestibular exercises **(Chap. 24)** and small doses of vestibular suppressants such as promethazine (Phenergan) may be helpful when dizziness is the main problem. After mild or moderate injury, patients who have difficulty with memory or with complex cognitive tasks at work may be reassured to know that these problems usually improve over several months, and a reduced workload or other accommodations may be prescribed in the interim. For select cases, speech and language therapy intervention may be appropriate.

For the vast majority of individuals with mTBI, symptoms of PPCS subside and resolve within a few weeks of injury. For a subset of individuals with mTBI, however, complaints of postconcussion symptoms persist beyond the expectation derived from TBI severity markers. Subtypes of PPCS have been proposed to improve characterization of specific symptoms or types of sequelae following mTBI. These include neurologic, cognitive, behavioral, or somatic complaints that continue beyond the acute and subacute periods, becoming chronic and often operationalized as persisting beyond 3 months. Although the overall risk of experiencing PPCS following mTBI is low, the frequency of mTBI patients who present in a clinical setting is believed to be higher.

mTBI patients with PPCS frequently present to the outpatient clinics of primary care physicians, physiatrists, or neurologists seeking relief for lingering related symptoms. While some patients will have already received an initial medical workup to rule out a more serious brain injury during the acute phase, many patients will have had no prior contact with health care specialists. A medical workup ordered in the outpatient setting for PPCS-related complaints is typically unremarkable for any identifiable neurologic cause to account for the persisting symptoms reported by the patient. The development of uniform decision trees or "standard of care" treatment regimens for PPCS has been limited by the diversity of symptoms that patients experience, even within mTBI subgroups that have sustained very similar injury patterns. While some patients experience somatic symptoms, others complain of subjective cognitive or behavioral changes. Symptom inventories (Table 454-2) can be helpful in documenting the broad range of these symptoms and serve as a metric for improvement following symptom-based treatment.

Active rehabilitation for the treatment of PPCS involving subthreshold exercise has increased in popularity over recent years and has gained empirical support for its effectiveness as a useful intervention for protracted recovery.

PPCS are often influenced by diverse cognitive, emotional, medical, psychosocial, and motivational factors. Because of this complexity, treatments targeting persistent and refractory symptoms should be tailored to the needs and expectations of the individual patient, with referrals to specialists as needed for assistance with management of headache, neck and back pain, dizziness and vertigo, and other persisting symptoms. A comprehensive review of concussion and persisting symptoms, presented in Table 454-2, allows for development of an individualized approach that leverages currently available treatment for those sequelae that are most bothersome to the patient (e.g., vestibular or cervicovestibular rehabilitation therapy for vertigo, melatonin for sleep disturbance). Patients are frequently referred to behavioral health providers such as neuropsychologists, rehabilitation psychologists, health psychologists, and/or psychiatrists for a variety of reasons, but particularly when they are experiencing persistent cognitive, emotional, or behavioral changes. Patients with mood disorders (e.g., depression), anxiety disorders (e.g., posttraumatic stress disorder), or adjustment reactions may benefit from psychiatric consultation for appropriate medication trials or from time-limited psychotherapy such as cognitive behavioral therapy.

Due to the complexity of presentation and varying diagnostic criteria, there are limited studies regarding overall prognosis of PPCS. However, treatment targeted to the individual's specific persisting difficulties can improve functional outcomes and patient-rated quality of life. Further, collaborative care has been shown to improve outcomes among patients experiencing PPCS. These improved outcomes are likely due to a multidisciplinary team's ability to simultaneously address the diverse set of difficulties that can occur with PPCS.

INJURY OF INTERMEDIATE SEVERITY

Patients who are not fully alert or have persistent confusion, behavioral changes, extreme dizziness, or focal neurologic signs such as hemiparesis should be admitted to the hospital and undergo a cerebral imaging study. A cerebral contusion or hematoma will usually be found. Common syndromes include (1) delirium with a disinclination to be examined or moved, expletive speech, and resistance if disturbed (anterior temporal lobe contusions); (2) a quiet, disinterested, slowed mental state (abulia) alternating with irascibility (inferior frontal and frontopolar contusions); (3) a focal deficit such as aphasia or mild hemiparesis (due to subdural hematoma or convexity contusion or, less often, carotid artery dissection); (4) confusion and inattention, poor performance on simple mental tasks, and fluctuating orientation (associated with several types of injuries, including those described above, and with medial frontal contusions and interhemispheric subdural hematoma); (5) repetitive vomiting, nystagmus, drowsiness, and unsteadiness (labyrinthine concussion, but occasionally due to a posterior fossa subdural hematoma or vertebral artery dissection); and (6) diabetes insipidus (damage to the median eminence or pituitary stalk). Injuries of this degree can be complicated by drug or alcohol intoxication, and clinically inapparent cervical spine injury may be present. Blast injuries are often accompanied by rupture of the tympanic membranes.

After surgical removal of hematomas, patients in this category improve over weeks to months. During the first week, the state of alertness, memory, and other cognitive functions often fluctuate, and agitation and somnolence are common. Behavioral changes tend to be worse at night, as with many other encephalopathies, and may be treated with small doses of antipsychotic medications. Subtle abnormalities of attention, intellect, spontaneity, and memory return toward normal weeks or months after the injury, sometimes abruptly. However, the full extent of recovery may not be realized for several years. Persistent cognitive problems are discussed below.

SEVERE INJURY

Patients who are comatose from the moment of injury require immediate neurologic attention and resuscitation. After intubation, with care taken to immobilize the cervical spine, the depth of coma, pupillary size and reactivity, limb movements, and Babinski responses are assessed. As soon as vital functions permit and cervical spine x-rays and a CT scan have been obtained, the patient should be transported to a critical care unit. Hypoxia should be reversed, and normal saline used as the resuscitation fluid in preference to albumin. The finding of an epidural or subdural hematoma or large intracerebral hemorrhage is usually an indication for prompt surgery and intracranial decompression in an otherwise salvageable patient. Measurement of ICP with a ventricular catheter or fiberoptic device in order to guide treatment has been favored by many units but has not improved outcome. Similarly, induced hypothermia has shown no benefit. Hyperosmolar intravenous solutions are used in various regimens to limit intracranial pressure. Prophylactic antiepileptic medications are recommended for 7 days and should be discontinued unless there are multiple seizures after injury. Management of raised ICP, a frequent feature of severe head injury, is discussed in **Chap. 318**.

Despite the improvement in mortality for severe TBI over the past few decades, a great deal of therapeutic nihilism persists in TBI. The common use of a 6-month outcome for TBI clinical studies reinforces this misconception. The recovery from severe TBI can take years. Furthermore, the ability to predict long-term outcome is limited and frequently incorrect. Best-practice guidelines recommend, in the absence of brain death, that aggressive therapy be instituted for at least 72 h in the acute injury period.

LONG-TERM OUTCOMES IN TBI

Continued follow-up of prospective studies has increased awareness of TBI as a chronic condition with evolving changes and needs over several years after injury. While the majority of individuals remain broadly stable, clinically meaningful improvement and decline across multiple domains (psychiatric, cognitive, and functional outcomes) have been observed from 2 to 7 years after injury, regardless of initial TBI severity. The direction of the changes beyond 1 year after injury can be variable (may decline and improve later, or vice versa). Collectively, this indicates that functional status remains dynamic beyond 1 year after injury, a conventionally considered plateau of recovery. Future investigation is required to better understand which factors are associated with improvement and decline beyond 1 year after injury. Regardless, there is growing awareness that TBI is an evolving condition that requires ongoing monitoring, rehabilitation, and support to address individual patient needs.

Chronic difficulties associated with TBI (characterized above), as opposed to level of risk for Alzheimer's disease and related dementias (ADRD), is an important distinction in determining long-term outcomes of TBI. TBI (aggregated mild to severe) is associated with a 63–96% increased risk of all-cause dementia. The degree of risk for dementia ranges along the gradient of TBI severity (i.e., greatest risk among severe injuries). There is some evidence that repeated mTBI or sport-related concussions may be associated with elevated risk as well. To date, however, investigations have less reliably established mTBI as a robust risk factor for dementia, likely due to methodologic heterogeneity (e.g., use of different diagnostic criteria, exposure misclassification, self-report vs physician diagnoses of TBI or dementia).

Though an identified risk factor for all-cause dementia, pathophysiologic and epidemiologic factors that underlie the association between TBI and risk of specific neurodegenerative pathologies and dementia subtypes are not well understood. As a result, associations between TBI and clinical syndromes (e.g., Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis) or distinct neuropathologies (e.g., beta-amyloid, Lewy bodies, transactive response DNA-binding protein 43) have been inconsistently reported in the literature. In a large study involving clinical and neuropathologic data from three pooled prospective studies of community-based cohorts, a significant relationship was found between TBI with LOC >1 h and subsequent Parkinson's disease diagnosis, progression rate of parkinsonism, and Lewy body accumulation at postmortem examination. Positron emission tomography (PET) studies have allowed for in vivo investigation of neuropathologic deposition and have failed to consistently observe associations between remote TBI (mild through severe) and amyloid or tau deposition.

Among former contact and collision sport athletes, exposure to repetitive head impacts (RHI) involving external blows that do not produce signs and symptoms of mTBI/concussion is common. The brains of these patients with substantial RHI exposure may display a characteristic deposition of tau protein in neurons located in the superficial cortical layers and perivascular regions and particularly in the depths of sulci. This pattern has been defined as the pathognomonic lesion of *chronic traumatic encephalopathy* (CTE). The degree to which this neuropathologic finding is present in nonathlete populations with potential RHI exposure (e.g., former military service members and veterans, victims of intimate partner violence) is uncertain, although the prevalence was found to be low in a large autopsy study of military service members.

A variety of neurodegenerative pathologies are commonly found in the presence of CTE, adding to the complexity of diagnosis. Furthermore, the dynamic interplay between RHI and mTBI history is not well understood. While staging criteria for this neuropathologic entity have yet to be established, a consensus meeting to define the neuropathologic criteria for CTE proposed an algorithm assessing CTE as "low" or "high" in severity. Overall, its contribution, if any, to late-life dementia and parkinsonism in former athletes, soldiers, or others who have sustained repeated concussive injuries is unknown.

Research criteria for the clinical diagnosis of CTE have been proposed. The criteria generally require substantial exposure to RHI, cognitive impairment (primarily in the domains of episodic memory and executive function) and/or neurobehavioral dysregulation, progressive course, and the absence of an alternative explanation for symptoms. Multiple studies have suggested that these proposed criteria lack specificity (i.e., they are frequent in other conditions and non-CTE cases). As such, CTE remains a postmortem diagnosis.

Investigations have not observed robust or consistent in vivo brain-related changes associated with years of contact sport/football exposure (a commonly used proxy measure for RHI) using advanced MRI, PET imaging, or blood-based biomarkers. For example, associations between years of participation and amyloid deposition or white matter hyperintensity volume have not been observed. Studies of brain morphometry (volumetric and structural changes) and tau deposition have been more variable, though evidence suggests that conventionally employed PET tracers and blood biomarkers may be limited in their specificity for CTE p-tau. Impairment of neuropsychological and neuropsychiatric function is most commonly observed in those with polypathology, particularly amyloid. Taken together, further study is required to better refine the clinical and postmortem diagnostic criteria of CTE, enhance clinicopathologic correlation, and ultimately improve patient care and management. **CTE is also discussed in Chap. 435.**

FURTHER READING

Brett BL et al: Long-term multidomain patterns of change after traumatic brain injury: A TRACK-TBI LONG Study. Neurol 101:7, 2023.

Johnson VE et al: Axonal pathology in traumatic brain injury. Exp Neurol 246:35, 2013.

Kowalski R et al: Recovery of consciousness and functional outcome in moderate and severe traumatic brain injury. JAMA Neurol 78:548, 2021.

McCrory P et al: Consensus statement on concussion in sport—the 5th international conference on concussion in sport held in Berlin, October 2016. Br J Sports Med 51:838, 2017.

Mez J et al: Clinicopathological evaluation of chronic traumatic encephalopathy in players of American football. JAMA 318:360, 2017.

Nelson L et al: Recovery after mild traumatic brain injury in patients presenting to US level I trauma centers: A Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI) study. JAMA Neurol 76:1049, 2019.

Taylor CA et al: Traumatic brain injury-related emergency department visits, hospitalizations, and deaths—United States, 2007 and 2013. MMWR Surveill Summ 66:1, 2017.

455

Multiple Sclerosis

Bruce A. C. Cree, Stephen L. Hauser

Multiple sclerosis (MS) is an autoimmune disease of the central nervous system (CNS) characterized by chronic inflammation, demyelination, gliosis (plaques or scarring), and neuronal loss; the course can be relapsing or progressive. MS plaques typically develop at different times and in different CNS locations (i.e., MS is said to be disseminated in time and space). One million individuals in the United States, and millions worldwide, are affected. The clinical course is extremely variable, ranging from a relatively benign condition to a rapidly evolving and incapacitating disease requiring profound lifestyle adjustments. The past decade has seen tremendous progress in understanding basic disease mechanisms underlying MS and in developing highly effective therapies especially for the relapsing form of the disease. These advances have dramatically improved the long-term outcome for patients.

CLINICAL MANIFESTATIONS

Onset is typically between 20 and 40 years (slightly later in men than in women), but the disease can present across the lifespan. Women are affected approximately three times more often than men. Early symptoms may be severe or seem so trivial that a patient may not seek medical attention for months or years. On occasion, MS lesions are located exclusively in noneloquent regions of the nervous system, and in such instances, clinical manifestations can be largely or entirely absent. Autopsy series identified MS in some individuals (~0.1% of cases) who were seemingly asymptomatic during life, and magnetic resonance imaging (MRI) scans obtained for unrelated reasons also showed evidence of asymptomatic MS, an incidental finding termed a *radiologically isolated syndrome* (RIS; see below).

Specific symptoms of MS are varied and reflect the location and severity of lesions within the CNS **(Table 455-1)**. Moreover, neurologic examination often reveals unexpected findings in addition to the anticipated ones. For example, a patient may present with symptoms in one leg but signs in both.

Sensory symptoms include both paresthesias (e.g., tingling, prickling sensations, "pins and needles," formications, or painful burning) and hypesthesia (e.g., reduced sensation, numbness, or a "dead" feeling). Unpleasant sensations (e.g., feelings that body parts are swollen, wet, raw, or tightly wrapped) are also common. Sensory impairment of the trunk and legs below a horizontal line on the torso (a sensory level) indicates that the spinal cord is the site of the disturbance. It is often accompanied by a bandlike sensation of tightness around the torso. Pain is a common symptom of MS, experienced by >50% of patients. Pain can occur anywhere on the body and can change locations over time.

Optic neuritis (ON) presents as diminished visual acuity, dimness, or decreased color perception (desaturation) in the central field of vision. These symptoms can be mild or may progress to severe visual loss. Rarely, there is complete loss of light perception. Visual symptoms are generally monocular but may be bilateral. Periorbital pain (aggravated by eye movement) typically precedes or accompanies the visual loss. An afferent pupillary defect **(Chap. 34)** is usually present. Fundoscopic examination may be normal or reveal optic disc swelling (papillitis). Pallor of the optic disc (optic atrophy) commonly follows ON. Uveitis is uncommon and should raise the possibility of alternative diagnoses such as sarcoidosis or lymphoma.

TABLE 455-1 Initial Symptoms of Multiple Sclerosis (MS)

SYMPTOM	PERCENTAGE OF CASES	SYMPTOM	PERCENTAGE OF CASES
Sensory loss	37	Lhermitte	3
Optic neuritis	36	Pain	3
Weakness	35	Dementia	2
Paresthesias	24	Visual loss	2
Diplopia	15	Facial palsy	1
Ataxia	11	Impotence	1
Vertigo	6	Myokymia	1
Paroxysmal attacks	4	Epilepsy	1
Bladder	4	Falling	1

Source: Data from RJ Swingler, DA Compston: The morbidity of multiple sclerosis. Q J Med 83:325, 1992.

Weakness of the limbs can manifest as loss of strength, speed, or dexterity; as fatigue; or as a disturbance of gait. Exercise-induced weakness is a characteristic symptom of MS. The weakness is of the upper motor neuron type **(Chap. 26)** and is usually accompanied by other pyramidal signs such as spasticity, hyperreflexia, and extensor plantar responses. Occasionally, a tendon reflex may be lost (simulating a peripheral nerve lesion) if an MS lesion disrupts the afferent reflex fibers in the spinal cord **(see Fig. 26-2)**.

Facial weakness due to a lesion in the pons may resemble idiopathic Bell's palsy **(Chap. 452)**. Unlike Bell's palsy, facial weakness in MS is usually not associated with ipsilateral loss of taste sensation or retroauricular pain.

Spasticity **(Chap. 26)** is commonly associated with spontaneous and movement-induced muscle spasms, especially in the legs. This can be accompanied by painful spasms interfering with ambulation, work, or self-care. Occasionally, spasticity provides support for the body weight during ambulation, and in these cases, treatment of spasticity may actually do more harm than good.

Visual blurring in MS may result from ON or diplopia (double vision); if the symptom resolves when either eye is covered, the cause is diplopia. *Diplopia* may be caused by internuclear ophthalmoplegia (INO) or palsy of the sixth cranial nerve (rarely the third or fourth). An INO consists of impaired adduction of one eye due to a lesion in the ipsilateral medial longitudinal fasciculus **(Chaps. 34 and V3)**. Prominent nystagmus is often observed in the abducting eye, along with a small skew deviation. A bilateral INO is particularly suggestive of MS. Other common gaze disturbances in MS include (1) a horizontal gaze palsy, (2) a "one and a half" syndrome (horizontal gaze palsy plus an INO), and (3) acquired pendular nystagmus.

Ataxia usually manifests as cerebellar tremors **(Chap. 450)**. Ataxia may also involve the head and trunk or the voice, producing a characteristic cerebellar dysarthria (scanning speech).

Vertigo may appear suddenly from a brainstem lesion, superficially resembling acute labyrinthitis **(Chap. 24)**. *Hearing loss* **(Chap. 36)** may also occur in MS but is uncommon.

ANCILLARY SYMPTOMS

Paroxysmal symptoms are distinguished by their brief duration (10 s to 2 min), high frequency (5–40 episodes per day), lack of any alteration of consciousness or change in background electroencephalogram during episodes, and a self-limited course (generally lasting weeks to months). They may be precipitated by hyperventilation or movement. Manifestations can include Lhermitte's symptom; tonic contractions of a limb, face, or trunk (tonic seizures); paroxysmal dysarthria and ataxia; paroxysmal sensory disturbances; and several other less well-characterized syndromes. Paroxysmal symptoms probably result from spontaneous discharges arising at the edges of demyelinated plaques and spreading to adjacent white matter tracts.

Lhermitte's symptom is an electric shock–like sensation (typically induced by flexion or other movements of the neck) that radiates down the back into the legs. Rarely, it radiates into the arms. It is generally self-limited but may persist for years. Lhermitte's symptom can also occur with other disorders of the cervical spinal cord (e.g., cervical spondylosis).

Trigeminal neuralgia, hemifacial spasm, and glossopharyngeal neuralgia **(Chap. 452)** can occur when the demyelinating lesion involves the root entry (or exit) zone of the fifth, seventh, and ninth cranial nerve, respectively. Trigeminal neuralgia (tic douloureux) is a very brief lancinating facial pain often triggered by an afferent input from the face or teeth. Most cases of trigeminal neuralgia are not MS related; however, atypical features such as onset before age 50 years, bilateral symptoms, objective sensory loss, or nonparoxysmal pain should raise the possibility that MS could be responsible.

Facial myokymia consists of either persistent rapid flickering contractions of the facial musculature (especially the lower portion of the orbicularis oculus) or a contraction that slowly spreads across the face. It results from lesions of the corticobulbar tracts or brainstem course of the facial nerve.

Heat sensitivity refers to neurologic symptoms produced by an elevation of the body's core temperature. For example, unilateral visual blurring may occur during a hot shower or with physical exercise (*Uhthoff's symptom*). It is also common for MS symptoms to worsen transiently, sometimes dramatically, during febrile illnesses. Such heat-related symptoms probably result from transient conduction block.

Bladder dysfunction is ultimately present in most MS patients. During normal reflex voiding, relaxation of the bladder sphincter (α-adrenergic innervation) is coordinated with contraction of the detrusor muscle in the bladder wall (muscarinic cholinergic innervation). *Detrusor hyperreflexia*, due to impairment of suprasegmental inhibition, causes urinary frequency, urgency, nocturia, and uncontrolled bladder emptying. *Detrusor sphincter dyssynergia*, due to loss of synchronization between detrusor and sphincter muscles, causes difficulty in initiating and/or stopping the urinary stream, producing hesitancy, urinary retention, overflow incontinence, and recurrent infection.

Constipation occurs in some patients, especially with advanced disease. Fecal urgency or *bowel incontinence* is less common than urinary symptoms but can be socially debilitating.

Sexual dysfunction may manifest as decreased libido, impaired genital sensation, impotence in men, and diminished vaginal lubrication or adductor spasms in women.

Cognitive dysfunction is often mild when present, but can include memory loss; impaired attention; difficulties in executive functioning, memory, and problem solving; slowed information processing; and problems shifting between cognitive tasks. Euphoria (elevated mood) or emotional lability (pseudobulbar palsy) was once thought to be characteristic of MS but is actually relatively uncommon. Cognitive dysfunction sufficient to impair activities of daily living is rare.

Depression, experienced by approximately half of patients, can be reactive, endogenous, or part of the illness itself and can contribute to fatigue.

Fatigue **(Chap. 25)** is experienced by most MS patients and is the most common reason for work-related disability in MS. Fatigue can be exacerbated by elevated temperatures, depression, expending exceptional effort to accomplish basic activities of daily living, or sleep disturbances (e.g., from frequent nocturnal awakenings to urinate).

DISEASE COURSE

In the traditional model of MS, the disease was considered to have three principal clinical forms, designated relapsing-remitting, secondary progressive, and primary progressive. Relapses were thought to be caused by inflammation, while progression was the consequence of neurodegeneration. More recently, these categories were supplanted by a unitary view of the disease, in which inflammation and neurodegeneration are present in most patients throughout the disease course. The concept that all MS is a single disease is also supported by findings from genetics, epidemiology, immunology, and pathology. Nonetheless, from a clinical perspective, it is still often useful to apply the classical subtype scheme to assessment and management of patients.

1. *Relapsing-remitting or bout onset MS* (RRMS) accounts for 90% of MS cases and is characterized by discrete attacks of neurologic dysfunction that generally evolve over days to weeks (rarely over hours). In early MS, there is often substantial or complete recovery over the ensuing weeks to months. However, as attacks continue, recovery may be less evident. Between attacks, patients were earlier thought to be neurologically stable; however, it is now clear that most if not all patients with RRMS experience subtle "silent" progression even when relapse-free **(Fig. 455-1)**. The category *relapsing MS* (RMS) is used to identify all relapsing patients, both RRMS as well as secondary progressive patients who continue to experience attacks.
2. *Secondary progressive MS* (SPMS) always begins as RRMS. At some point, however, the clinical course changes so that the patient experiences progressive deterioration in function unassociated with acute attacks. SPMS produces a greater amount of fixed neurologic disability than RRMS. A practical definition for SPMS is a patient who has developed some level of permanent walking disability not due exclusively to relapses. The Extended Disability Status Score (EDSS) is a widely used measure of neurologic impairment in MS

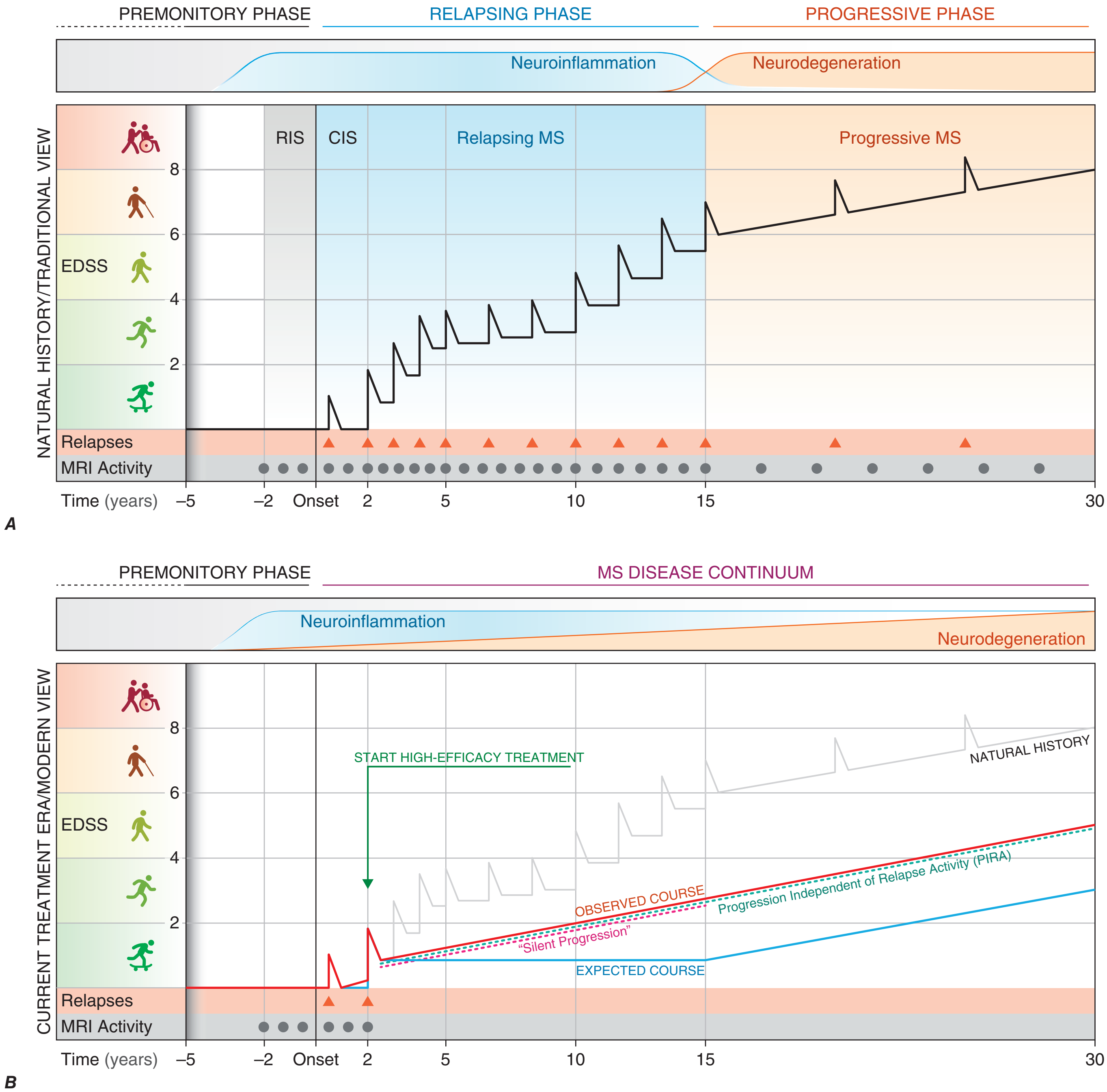

FIGURE 455-1 The clinical course of multiple sclerosis (MS) in the current treatment era. The top half of the figure illustrates the traditional view of the natural history of relapse-onset MS in the pretreatment era. During the relapsing phase, disability accumulation was thought to result from incomplete recovery from relapses, until relapse-independent disability, designated SPMS, supervened. In the bottom half of the figure, the "new" natural history of MS in the current treatment era is shown. With use of highly effective therapies, attacks are abolished in most patients, but insidious progression independent of relapse activity, termed "silent progression," is now evident during the relapsing phase. CIS, clinically isolated syndrome; EDSS, Extended Disability Status Score; MRI, magnetic resonance imaging; RIS, radiologically isolated syndrome; SPMS, secondary progressive multiple sclerosis.

(Table 455-2); an EDSS of 4 or greater, plus a Functional Status Scale (FSS) motor system score of 2 or greater, can support a diagnosis of SPMS. For a patient with RRMS, in the pretreatment era, the risk of developing SPMS was ~3% each year, meaning that the great majority of RRMS would ultimately evolve into SPMS. However, more recent case series have indicated a much lower rate of evolution to SPMS, estimated at <1% each year, likely due to widespread use of increasingly effective therapies for MS.

3. *Primary progressive MS* (PPMS) accounts for ~10% of cases. These patients do not experience attacks but rather steadily decline in function from disease onset. Compared to RRMS, the sex distribution is more even, the disease begins later in life (mean age ~40 years), and disability develops faster relative to the onset of the first clinical symptom. As noted above, despite these differences PPMS appears to represent the same underlying illness as RRMS and SPMS, and some PPMS patients experience relapses over the course of their illness. The term *active progressive MS* is used to categorize progressive MS patients (both SPMS and PPMS) who experience relapses or are found to have new lesions on serial MRI scans.

Disability in MS is thought to accumulate as either a consequence of limited recovery following an acute relapse, a process also known as *relapse associated worsening* (RAW), or from presumed underlying neurodegeneration in the absence of clinical relapse, a process termed *progression independent of relapsing activity* (PIRA). Although RAW was once thought to be the primary driver of disability accumulation in RRMS, it is now clear that PIRA is the cause of disability accumulation in RRMS, SPMS, and PPMS. That PIRA events can occur "silently,"

TABLE 455-2 Scoring Systems for Multiple Sclerosis (MS)

Expanded Disability Status Scale (EDSS)

0.0 = Normal neurologic examination (all grade 0 in functional status [FS])
1.0 = No disability, minimal signs in one FS (i.e., grade 1)
1.5 = No disability, minimal signs in more than one FS (more than one grade 1)
2.0 = Minimal disability in one FS (one FS grade 2, others 0 or 1)
2.5 = Minimal disability in two FS (two FS grade 2, others 0 or 1)
3.0 = Moderate disability in one FS (one FS grade 3, others 0 or 1) or mild disability in three or four FS (three/four FS grade 2, others 0 or 1) although fully ambulatory
3.5 = Fully ambulatory but with moderate disability in one FS (one grade 3) and one or two FS grade 2; or two FS grade 3; or five FS grade 2 (others 0 or 1)
4.0 = Ambulatory without aid or rest for ~500 m
4.5 = Ambulatory without aid or rest for ~300 m
5.0 = Ambulatory without aid or rest for ~200 m
5.5 = Ambulatory without aid or rest for ~100 m
6.0 = Unilateral assistance required to walk about 100 m with or without resting
6.5 = Constant bilateral assistance required to walk about 20 m without resting
7.0 = Unable to walk beyond about 5 m even with aid; essentially restricted to wheelchair; wheels self and transfers alone
7.5 = Unable to take more than a few steps; restricted to wheelchair; may need aid to transfer
8.0 = Essentially restricted to bed or chair or perambulated in wheelchair, but out of bed most of day; retains many self-care functions; generally has effective use of arms
8.5 = Essentially restricted to bed much of the day; has some effective use of arm(s); retains some self-care functions
9.0 = Helpless bed patient; can communicate and eat
9.5 = Totally helpless bed patient; unable to communicate or eat
10.0 = Death due to MS

Functional Status (FS) Score

A. Pyramidal functions
0 = Normal
1 = Abnormal signs without disability
2 = Minimal disability
3 = Mild or moderate paraparesis or hemiparesis, or severe monoparesis
4 = Marked paraparesis or hemiparesis, moderate quadriparesis, or monoplegia
5 = Paraplegia, hemiplegia, or marked quadriparesis
6 = Quadriplegia

B. Cerebellar functions
0 = Normal
1 = Abnormal signs without disability
2 = Mild ataxia
3 = Moderate truncal or limb ataxia
4 = Severe ataxia all limbs
5 = Unable to perform coordinated movements due to ataxia

C. Brainstem functions
0 = Normal
1 = Signs only
2 = Moderate nystagmus or other mild disability
3 = Severe nystagmus, marked extraocular weakness, or moderate disability of other cranial nerves
4 = Marked dysarthria or other marked disability
5 = Inability to swallow or speak

D. Sensory functions
0 = Normal
1 = Vibration or figure-writing decrease only, in 1 or 2 limbs
2 = Mild decrease in touch or pain or position sense, and/or moderate decrease in vibration in 1 or 2 limbs, or vibratory decrease alone in 3 or 4 limbs
3 = Moderate decrease in touch or pain or position sense, and/or essentially lost vibration in 1 or 2 limbs, or mild decrease in touch or pain, and/or moderate decrease in all proprioceptive tests in 3 or 4 limbs
4 = Marked decrease in touch or pain or loss of proprioception, alone or combined, in 1 or 2 limbs or moderate decrease in touch or pain and/or severe proprioceptive decrease in >2 limbs
5 = Loss (essentially) of sensation in 1 or 2 limbs or moderate decrease in touch or pain and/or loss of proprioception for most of the body below the head
6 = Sensation essentially lost below the head

E. Bowel and bladder functions
0 = Normal
1 = Mild urinary hesitancy, urgency, or retention
2 = Moderate hesitancy, urgency, retention of bowel or bladder, or rare urinary incontinence
3 = Frequent urinary incontinence
4 = In need of almost constant catheterization
5 = Loss of bladder function
6 = Loss of bowel and bladder function

F. Visual (or optic) functions
0 = Normal
1 = Scotoma with visual acuity (corrected) better than 20/30
2 = Worse eye with scotoma with maximal visual acuity (corrected) of 20/30 to 20/59
3 = Worse eye with large scotoma, or moderate decrease in fields, but with maximal visual acuity (corrected) of 20/60 to 20/99
4 = Worse eye with marked decrease of fields and maximal acuity (corrected) of 20/100 to 20/200; grade 3 plus maximal acuity of better eye of 20/60 or less
5 = Worse eye with maximal visual acuity (corrected) <20/200; grade 4 plus maximal acuity of better eye of ≤20/60
6 = Grade 5 plus maximal visual acuity of better eye of ≤20/60

G. Cerebral (or mental) functions
0 = Normal
1 = Mood alteration only (does not affect EDSS score)
2 = Mild decrease in mentation
3 = Moderate decrease in mentation
4 = Marked decrease in mentation
5 = Chronic brain syndrome—severe or incompetent

Source: Adapted from JF Kurtzke: Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). Neurology 33:1444, 1983.

meaning so insidiously that neither the patient nor the provider recognizes their occurrence at the time of gradual worsening, raises the important question as to whether there is a meaningful distinction between RRMS and SPMS. If any confirmed PIRA event is considered to be indicative of SPMS, then the SPMS onset begins much earlier in the disease course when patients still experience relapses but have only accumulated relative minor disability.

■ EPIDEMIOLOGY

Geographic gradients are consistently observed in MS, with the highest prevalence generally found in temperate zones; in tropical regions, the prevalence is often 10-fold to 20-fold less. In addition, a north-south gradient was observed in numerous national and regional studies, with decreasing rates as one moves equatorially.

The prevalence of MS also increased steadily in several regions around the world over the past half-century, presumably reflecting the impact of some environmental shift, improved diagnosis, and/or a longer lifespan. Moreover, this increase appears to have occurred to a greater degree in women than men and in nonwhite populations. In the United States, there is a slightly higher prevalence in white compared with black individuals, with lower estimates in Hispanics, followed by Asians.

Multiple lines of evidence incriminate a role for infection with the Epstein-Barr virus (EBV) in MS. Individuals who have never been EBV infected (~5% of the population globally) have a very low MS risk, ~20-fold lower than in EBV-positive individuals, and a history of infectious mononucleosis (associated with initial exposure to EBV during adolescence or later in life) increases risk more than twofold higher yet. Higher antibody titers to EBV nuclear antigens were repeatedly associated with MS risk, and studies from longitudinal biobank collections showed that serologic conversion to EBV is a near-universal prerequisite for development of MS. Following primary EBV infection, a lifelong infection is established in most individuals, with latent EBV exclusively present in very small numbers (~1:10^{-6}) of B lymphocytes. EBV-infected B cells were not consistently identified in the nervous system of MS patients. It is possible that ongoing lytic cycles by very few infected B cells residing within the CNS could produce bursts of inflammation and MS lesions; however, it is more likely that pathology could be triggered by B cell–mediated antigen presentation of EBV peptides that cross-react with MS autoantigens via molecular mimicry (see "Immunology," below).

A history of cigarette smoking is also associated with MS risk. Interestingly, in an animal model of MS, the lung was identified as a critical site for activation of pathogenic T lymphocytes responsible for autoimmune demyelination.

Finally, vitamin D deficiency has been repeatedly associated with MS. Immunoregulatory effects of vitamin D could explain these apparent relationships. Exposure of the skin to ultraviolet B (UVB) radiation from the sun is essential for the biosynthesis of vitamin D, and this endogenous production is the most important source of vitamin D in most individuals. A diet rich in fatty fish represents another source of vitamin D. At higher latitudes, the amount of UVB radiation reaching the earth's surface is often insufficient, particularly during winter months, and consequently, low serum levels of vitamin D are frequent in temperate zones. The common practice to avoid direct sun exposure and the widespread use of sunblock would be expected to exacerbate any population-wide vitamin D deficiency.

GENETIC CONSIDERATIONS

MS aggregates within some families, and adoption, half-sibling, twin, and spousal studies indicate that familial aggregation is primarily due to genetic factors. Importantly, family studies also support a contribution of environment, as fraternal twins of MS patients are at higher risk than nontwin siblings **(Table 455-3)**.

Susceptibility to MS is polygenic, with each gene contributing a relatively small amount to overall risk. The strongest susceptibility signal genome-wide maps to the human leukocyte antigen (HLA)-DRB1 gene in the class II region of the major histocompatibility complex (MHC) and specifically to HLA-DRB1*1501 (formerly designated DR2), and this association accounts for ~10% of the disease risk. This HLA association, first described in the early 1970s, suggests that MS, at its core, is an autoimmune disease. Whole-genome association studies have identified >230 other MS susceptibility variants, each of which individually has only a very small effect on MS risk. Many of these MS-associated genes have known roles in the adaptive and innate immune system, for example, the genes for the interleukin (IL) 7 receptor (CD127), IL-2 receptor (CD25), and T-cell costimulatory molecule LFA-3 (CD58); some variants also influence susceptibility to other autoimmune diseases in addition to MS. The variants identified so far all lack specificity and sensitivity for MS; thus, at present, they are not useful for diagnosis and have no meaningful effect on the clinical course of MS once it begins. For many years, identification of genes that influence disease expression was elusive, but recently, the first loci for MS severity were identified; unlike risk genes, these variants appear to operate in the nervous system rather than immune system, and one signal reaching genome-wide significance, located in the region of dysferlin and a zinc finger gene (ZNF638), confers a 7-year acceleration of progression to wheelchair-dependent status.

TABLE 455-3 Risk of Developing Multiple Sclerosis (MS)

1 in 3	If an identical twin has MS
1 in 15	If a fraternal twin has MS
1 in 25	If a sibling has MS
1 in 50	If a parent or half-sibling has MS
1 in 100	If a first cousin has MS
1 in 1000	If a spouse has MS
1 in 1000	If no one in the family has MS

PATHOGENESIS

PATHOLOGY

Demyelination New MS lesions begin with perivenular cuffing by inflammatory mononuclear cells, predominantly T cells and macrophages, which also infiltrate the surrounding white matter. At sites of inflammation, the blood-brain barrier (BBB) is disrupted, but unlike vasculitis, the vessel wall is preserved. At the leading edge of lesions, cytotoxic CD8 cells are found. Involvement of the humoral immune system is also evident; B lymphocytes infiltrate the nervous system, myelin-specific autoantibodies are present on degenerating myelin sheaths, and complement is activated.

Sharply demarcated areas of demyelination are the pathologic hallmark of MS lesions, and evidence of myelin degeneration is found at the earliest time points of tissue injury. Although relative sparing of axons is typical, partial or total axonal destruction can also occur, especially within highly inflammatory lesions. In some lesions, surviving oligodendrocytes or those that differentiate from precursor cells partially remyelinate the surviving axons, producing so-called *shadow plaques*. However, in many lesions, oligodendrocyte precursor cells are present but fail to differentiate into mature myelin-producing cells. Therefore, promoting remyelination to protect axons remains an important therapeutic goal. As lesions evolve, there is prominent astrocytic proliferation (gliosis), and the term *sclerosis* refers to these gliotic plaques that have a rubbery or hardened texture at autopsy.

Neurodegeneration Cumulative axonal and neuronal loss is the most important contributor to irreversible neurologic disability and progressive symptoms. With paraplegia due to MS, as many as 70% of axons are ultimately lost from the lateral corticospinal (e.g., motor) tracts. Demyelination can reduce trophic support for axons, redistribute ion channels, and destabilize action potential membrane potentials. Axons can adapt initially to these injuries, but over time, distal and retrograde degeneration ("dying-back" axonopathy) occurs.

Multiple pathologies appear to contribute to progressive symptoms. *Chronic active plaques* are preexisting white matter lesions that show evidence of persistent inflammation, progressive axonal loss, and gradual concentric expansion, with large numbers of microglial cells at the leading edge of enlarging lesions without BBB disruption. Also important is a primary injury to the cerebral cortex. Cortical plaques are frequent in MS but are generally not well visualized by MRI; these can extend upward from adjacent white matter lesions or may be located entirely within the cortex or underneath the pia. *Ectopic lymphoid follicles* are aggregates of B, T, and plasma cells located in the superficial meninges, especially overlying deep cortical sulci; similar clusters are also present in perivascular spaces. Ectopic lymphoid follicles are associated with underlying demyelination and neuronal loss in the cerebral cortex, and diffusible factors from these lymphoid cells appear to mediate subpial cortical demyelination and neurodegeneration. Cux2-positive neurons in layer 2 and 3 of the neocortex appear to be particularly vulnerable. Neuronal and axonal death may result from glutamate-mediated excitotoxicity, oxidative injury, iron accumulation, and/or mitochondrial failure.

In relapsing MS, inflammation is characterized by focal perivenular infiltration of lymphocytes and monocytes, BBB disruption, and active demyelination. By contrast, inflammation in progressive MS is more diffuse, with widespread microglial proliferation across large areas of white matter, accompanied by infiltration of CD8 T cells and plasmablasts/plasma cells. Reduced myelin staining and axonal injury ("dirty white matter") are associated with these chronic pathologies. Astrogliosis has long been known to be a prominent feature of MS pathology, and activated astrocytes likely contribute directly to neuronal and myelin injury **(Chap. 435)**. Ongoing inflammation occurs behind an intact BBB in many patients with progressive MS, possibly accounting for the failure of immunotherapies not capable of crossing the BBB to benefit patients with progressive MS.

IMMUNOLOGY

An autoimmune response directed against components of CNS myelin, and perhaps other neural elements as well, remains the cornerstone of current concepts of MS pathogenesis. However, specific antigenic targets in MS have never been conclusively identified.

B Lymphocytes and Antibodies B cells are centrally involved in the development of demyelinating lesions, as evidenced by the efficacy of B cell–based treatments in all forms of MS (see "Treatment" below). Clonally restricted populations of activated, antigen-experienced, memory B cells and plasma cells are present in MS lesions, in meningeal lymphoid follicle-like structures overlying the cerebral cortex, and in cerebrospinal fluid (CSF). They produce the oligoclonal immunoglobulins and increased antibody synthesis rates in the CSF long useful in the diagnosis of MS. Myelin-specific autoantibodies, some directed against an extracellular myelin protein, myelin oligodendrocyte glycoprotein (MOG), have been detected bound to degenerating myelin in MS plaques. However, many more antibodies derived from these B cells appear to be directed against a diverse array of ubiquitous intracellular proteins seemingly unrelated to MS pathogenesis. Furthermore, the specific targets are different in each patient. Therefore, although these highly restricted CNS antibodies are characteristic of MS, their role in disease remains uncertain.

More likely, the antigen-presenting cell (APC) function of B cells explains their role in MS pathogenesis. Fragments of self-peptides derived from HLA-DR2 proteins themselves were found to bind intact DRB1*1501 molecules on B cells and serve as antigens for presentation to T cells. Memory CD4+ T cells derived from CSF responded to these self-peptides bound to DR2 molecules, and in some cases, these self-peptides were cross-reactive with several myelin antigens, as well as proteins derived from EBV, *Akkermansia muciniphila* (a commensal gut bacterium associated with dysbiosis in MS patients), and RAS guanyl-releasing protein 2 (RASGRP2), previously found to be a possible T-cell autoantigen in MS. Thus, MS-associated HLA proteins contain fragments that might trigger autoimmunity through molecular mimicry with viral, bacterial, or normal host antigens.

Autoreactive T Lymphocytes Autoreactive T cells may be triggered and sustained via B-cell antigen presentation. Myelin basic protein (MBP), an intracellular protein involved in myelin compaction, is an important T-cell antigen in experimental allergic encephalomyelitis (EAE), a laboratory model for MS. Activated MBP-reactive T cells have been identified in the blood, in CSF, and within MS lesions. The MS-associated HLA-DRB1*1501 protein binds with high affinity to a fragment of MBP (spanning amino acids 89–96), potentially stimulating T-cell responses to this self-protein. Several different populations of proinflammatory T cells are likely to mediate autoimmunity in MS. T-helper type 1 (T_H1) cells producing interferon γ (IFN-γ) are one key effector population; T_H1 cytokines, including IL-2, tumor necrosis factor (TNF)-α, and IFN-γ, play key roles in activating and maintaining autoimmune responses, and TNF-α and IFN-γ may directly injure oligodendrocytes or the myelin membrane. B cells from MS patients are also known to be high producers of TNF-α. As noted above, CD8 cytotoxic T cells are present at the active edges of expanding MS lesions, and activated CD8 cells also appear to be enriched for reactivity against myelin antigens in MS patients.

Microglial Activation Widespread microglial activation is a hallmark of progressive MS pathology. Activated microglia are found in cortical plaques in the absence of macrophage and leukocyte infiltrates. Some cortical plaques are found adjacent to sites of meningeal inflammation in which tertiary lymphoid follicles are found. As discussed above, these meningeal lymphoid structures are a hallmark of MS pathology. Microglial activation in MS is thought to be triggered by proinflammatory B and T lymphocytes or in response to tissue injury signals via toll-like receptor signaling. Although once thought to exist in either proinflammatory or anti-inflammatory states, microglia are now understood to have varied and context-dependent transcriptional states.

PHYSIOLOGY

Nerve conduction in myelinated axons occurs in a saltatory manner, with the nerve impulse jumping from one node of Ranvier to the next without depolarization of the axonal membrane underlying the myelin sheath between nodes **(Fig. 455-2*A*)**. This produces faster conduction velocities (~70 m/s) than the slow velocities (~1 m/s) produced by continuous propagation in unmyelinated nerves. Conduction block occurs when the nerve impulse is unable to traverse the demyelinated segment. This can happen when the resting axon membrane becomes hyperpolarized due to exposure of voltage-dependent potassium channels that are normally buried underneath the myelin sheath. A temporary conduction block often follows a demyelinating event before sodium channels (originally concentrated at the nodes) redistribute along the naked axon **(Fig. 455-2*B*)**. This redistribution ultimately allows continuous propagation of nerve action potentials through the demyelinated segment. Conduction block may be incomplete, affecting high- but not low-frequency volleys of impulses. Variable conduction block can also occur with raised body temperature or metabolic alterations. These factors may explain clinical fluctuations that vary from hour to hour or appear with fever or exercise. Conduction slowing occurs when the demyelinated segments of the axonal membrane are reorganized to support continuous (slow) nerve impulse propagation.

DIAGNOSIS

There is no single diagnostic test for MS. Diagnostic criteria for clinically definite MS require two or more episodes of symptoms and two or more signs that reflect pathology in anatomically noncontiguous white matter tracts of the CNS **(Table 455-4)**. Symptoms must last for >24 h and occur as distinct episodes separated by a month or more. In patients who have only one of the two required signs on neurologic examination, the second may be documented by abnormal tests such as MRI or evoked potentials (EPs). Similarly, in the most recent diagnostic scheme, the second clinical event (in time) may be supported solely by MRI findings, consisting of either the development of new focal white matter lesions on MRI or the simultaneous presence of both an enhancing lesion and a nonenhancing lesion in an asymptomatic location. In patients whose course is progressive from onset for ≥6 months

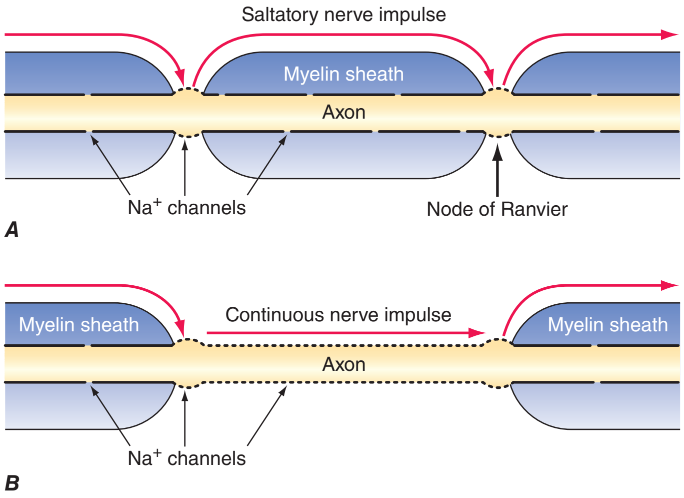

FIGURE 455-2 Nerve conduction in myelinated and demyelinated axons. *A.* Saltatory nerve conduction in myelinated axons occurs with the nerve impulse jumping from one node of Ranvier to the next. Sodium channels (*shown as breaks in the solid black line*) are concentrated at the nodes where axonal depolarization occurs. ***B.*** Following demyelination, additional sodium channels are redistributed along the axon itself, thereby allowing continuous propagation of the nerve action potential despite the absence of myelin.

TABLE 455-4 Diagnostic Criteria for Multiple Sclerosis (MS)

CLINICAL PRESENTATION	ADDITIONAL DATA NEEDED FOR MS DIAGNOSIS
2 or more attacks; objective clinical evidence of 2 or more lesions or objective clinical evidence of 1 lesion with reasonable historical evidence of a prior attack	None
2 or more attacks; objective clinical evidence of 1 lesion	Dissemination in space, demonstrated by ≥1 T2 lesion on MRI in at least 2 out of 4 MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial, or spinal cord) OR • Await a further clinical attack implicating a different CNS site
1 attack; objective clinical evidence of 2 or more lesions	Dissemination in time, demonstrated by • Simultaneous presence of asymptomatic gadolinium-enhancing and nonenhancing lesions at any time OR • A new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI, irrespective of its timing with reference to a baseline scan OR • Await a second clinical attack
1 attack; objective clinical evidence of 1 lesion (clinically isolated syndrome)	Dissemination in space and time, demonstrated by: For dissemination in space • ≥1 T2 lesion in at least 2 out of 4 MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial, or spinal cord) OR • Await a second clinical attack implicating a different CNS site AND • For dissemination in time • Simultaneous presence of asymptomatic gadolinium-enhancing and nonenhancing lesions at any time OR • A new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI, irrespective of its timing with reference to a baseline scan OR • Await a second clinical attack
Insidious neurologic progression suggestive of MS (PPMS)	1 year of disease progression (retrospectively or prospectively determined) PLUS 2 out of the 3 following criteria: • Evidence for dissemination in space in the brain based on ≥1 T2+ lesions in the MS-characteristic periventricular, juxtacortical, or infratentorial regions • Evidence for dissemination in space in the spinal cord based on ≥2 T2+ lesions in the cord • Positive CSF (isoelectric focusing evidence of oligoclonal bands and/or elevated IgG index)

Abbreviations: CNS, central nervous system; CSF, cerebrospinal fluid; MRI, magnetic resonance imaging; PPMS, primary progressive multiple sclerosis.

Source: Reproduced with permission from AJ Thompson et al: Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 17:162, 2018.

without superimposed relapses, documentation of intrathecal IgG synthesis may be used to support a diagnosis of PPMS.

DIAGNOSTIC TESTS

■ MAGNETIC RESONANCE IMAGING

MRI has revolutionized the diagnosis and management of MS **(Fig. 455-3)**; characteristic abnormalities are found in >95% of patients, although the majority of lesions visualized by MRI are asymptomatic. An increase in vascular permeability from a breakdown of the BBB is detected by leakage of intravenous gadolinium (Gd) into the parenchyma. Such leakage occurs early in the development of an MS lesion and serves as a marker of inflammation. Gd enhancement typically persists for <1 month, and the residual MS plaque remains visible indefinitely as a focal area of hyperintensity (a lesion) on T2-weighted images. Lesions are frequently oriented perpendicular to the ventricular surface, corresponding to a pattern of perivenous demyelination (Dawson's fingers; **Fig. 455-3*B***). Lesions are multifocal within the brain, brainstem, and spinal cord. Lesions >6 mm located in the corpus callosum, periventricular white matter, brainstem, cerebellum, or spinal cord are particularly helpful diagnostically. Also useful diagnostically is a *central vein sign* within plaques visualized with susceptibility-weighted (such as T2*) sequences **(Chap. 434)**. Criteria for the use of MRI in diagnosis of MS are shown in Table 455-4.

Serial MRI studies in RRMS reveal that bursts of focal inflammatory disease activity occur far more frequently than would have been predicted by the frequency of relapses. Thus, early in MS, most disease activity is clinically silent.

The total volume of T2-weighted signal abnormality (the "burden of disease") shows a significant (albeit weak) correlation with clinical disability. Quantitative measures of brain and especially spinal cord atrophy provide evidence of diffuse tissue injury and correlate more strongly with measures of disability or progressive MS. Serial MRI studies also indicate that progressive brain atrophy occurs even in very early MS and continues throughout the disease course. Approximately one-third of T2-weighted lesions appear as hypointense lesions (*black holes*) on T1-weighted imaging. Black holes are markers of irreversible demyelination and axonal loss, although even this measure depends on the timing of the image acquisition (e.g., most acute Gd-enhancing T2 lesions are T1 dark, and in chronic lesions, there is progressive T1 darkening over time).

■ CEREBROSPINAL FLUID

CSF changes in MS include a mononuclear cell pleocytosis and an increased level of intrathecally synthesized IgG. The total CSF protein is usually normal or only mildly elevated. Various formulas distinguish intrathecally synthesized IgG from IgG that entered the CNS passively from the serum. The CSF IgG index expresses the ratio of IgG to albumin in the CSF divided by the same ratio in the serum. The IgG synthesis rate uses serum and CSF IgG and albumin measurements to calculate the rate of CNS IgG synthesis. The measurement of oligoclonal bands (OCBs) by agarose gel electrophoresis of the CSF assesses intrathecal production of specific IgG clones separated by differences in charge. Two or more discrete OCBs, not present in a paired serum sample, are found in >90% of patients with MS. OCBs may be absent at the onset of MS, and in individual patients, the number of bands may increase over time.

A mild CSF pleocytosis (>5 cells/μL) is present in ~25% of cases, usually in young patients with RMS. A pleocytosis of >75 cells/μL, the presence of polymorphonuclear leukocytes, or a protein concentration >1 g/L (>100 mg/dL) in CSF should raise concern that the patient may not have MS. Because of its utility to rule in and also rule out MS, CSF examination is highly recommended as part of the routine MS workup, and especially when the diagnosis is uncertain.

■ EVOKED POTENTIALS

EP testing assesses function in afferent (visual, auditory, and somatosensory) or efferent (motor) CNS pathways. EPs use computer averaging to measure CNS electric potentials evoked by repetitive stimulation of selected peripheral nerves or of the brain. These tests provide the most information when the pathways studied are clinically uninvolved. For example, in a patient with a relapsing spinal cord syndrome with sensory deficits in the legs, an abnormal somatosensory EP following posterior tibial nerve stimulation provides little new information. By contrast, an abnormal visual EP in this circumstance would permit a diagnosis of clinically definite MS (Table 455-4). Abnormalities on one or more EP modalities occur in 80–90% of MS patients. EP abnormalities are not specific to MS, although a marked delay in the latency of a

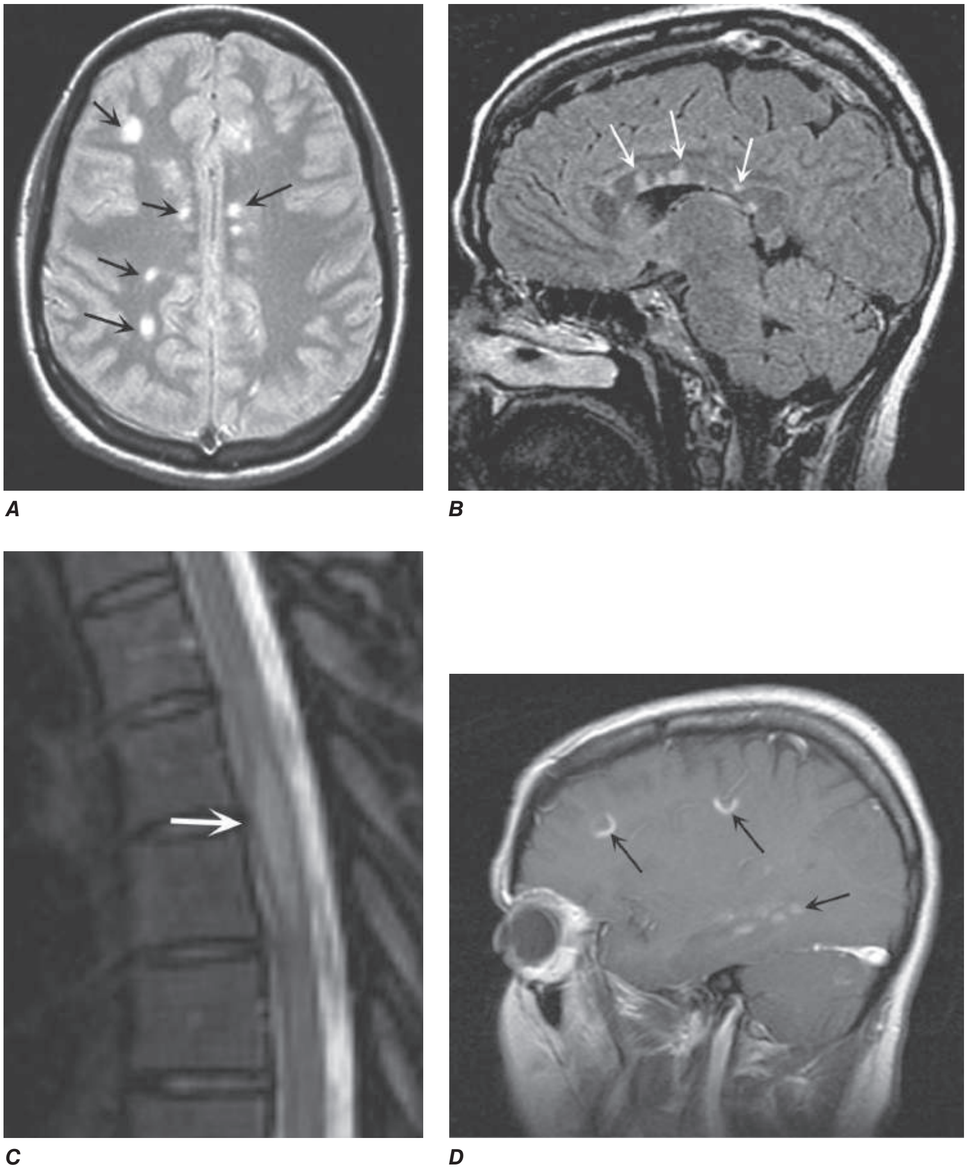

FIGURE 455-3 Magnetic resonance imaging findings in multiple sclerosis (MS). ***A.*** Axial first-echo image from T2-weighted sequence demonstrates multiple bright signal abnormalities in white matter, typical for MS. ***B.*** Sagittal T2-weighted fluid-attenuated inversion recovery (FLAIR) image in which the high signal of cerebrospinal fluid (CSF) has been suppressed. CSF appears dark, whereas areas of brain edema or demyelination appear high in signal, as shown here in the corpus callosum (*arrows*). Lesions in the anterior corpus callosum are frequent in MS and rare in vascular disease. ***C.*** Sagittal T2-weighted fast spin echo image of the thoracic spine demonstrates a fusiform high-signal-intensity lesion in the midthoracic spinal cord. ***D.*** Sagittal T1-weighted image obtained after the intravenous administration of gadolinium diethylene triamine pentaacetic acid (DTPA) reveals focal areas of blood-brain barrier disruption, identified as high-signal-intensity regions (*arrows*).

specific EP component (as opposed to a reduced amplitude or distorted wave shape) suggests demyelination.

DIFFERENTIAL DIAGNOSIS

The possibility of an alternative diagnosis should always be considered **(Table 455-5)**, particularly when (1) symptoms are localized exclusively to the posterior fossa, craniocervical junction, or spinal cord; (2) the patient is <15 or >60 years of age; (3) the clinical course is progressive from onset; (4) the patient has never experienced visual, sensory, or bladder symptoms; or (5) laboratory findings (e.g., MRI, CSF, or EPs) are atypical. Similarly, symptoms that are uncommon or rare in MS (e.g., aphasia, parkinsonism, chorea, isolated dementia, severe muscular atrophy, peripheral neuropathy, episodic loss of consciousness, fever, headache, seizures, or coma) favor an alternative diagnosis. Diagnosis can be particularly difficult in patients with a rapid or explosive (stroke-like) onset or those with mild symptoms and a normal neurologic examination. Rarely, intense inflammation and swelling may produce a mass lesion that mimics a primary or metastatic tumor. Disorders possibly mistaken for MS include neuromyelitis optica (NMO) and the more recently identified myelin oligodendrocyte protein-associated disease (MOGAD) and glial fibrillary acid protein (GFAP) disorders **(Chap. 456)**; these should be considered in patients who present with bilateral and/or severe optic neuritis or severe transverse myelitis. With hyperacute or postinfectious presentations, another consideration is acute disseminated encephalomyelitis (ADEM; **Chap. 456**). Other possibilities include Sjögren's syndrome, sarcoidosis, vascular disorders (antiphospholipid syndrome and vasculitis), rarely CNS lymphoma, and still more rarely infections such as syphilis or Lyme disease. The specific tests required to exclude alternative diagnoses will vary with each clinical situation; however, an erythrocyte sedimentation rate, serum B_{12} level, antinuclear antibodies, and treponemal antibody should probably be obtained in all patients with suspected MS.

TREATMENT

Therapy for MS can be divided into several categories: (1) treatment of acute attacks, (2) treatment with disease-modifying agents that reduce the biologic activity of MS, and (3) symptomatic therapy. Treatments that promote remyelination or neural repair do not currently exist, but several promising approaches are being actively investigated.

TABLE 455-5 Disorders That Can Mimic Multiple Sclerosis (MS)

- Acute disseminated encephalomyelitis (ADEM)
- Antiphospholipid antibody syndrome
- Behçet's disease
- Cerebral autosomal-dominant arteriopathy, subcortical infarcts, and leukoencephalopathy (CADASIL)
- Congenital leukodystrophies (e.g., adrenoleukodystrophy, metachromatic leukodystrophy)
- Glial Fibrillary Acidic Protein (GFAP) Autoimmunity
- Human immunodeficiency virus (HIV) infection
- Ischemic optic neuropathy (arteritic and nonarteritic)
- Lyme disease
- Mitochondrial encephalopathy with lactic acidosis and stroke (MELAS)
- Myelin oligodendrocyte glycoprotein-associated disease (MOGAD)
- Neoplasms (e.g., lymphoma, glioma, meningioma)
- Neuromyelitis optica
- Sarcoidosis
- Sjögren's syndrome
- Stroke and ischemic cerebrovascular disease
- Syphilis
- Systemic lupus erythematosus and related collagen vascular disorders
- Tropical spastic paraparesis (HTLV-1/2 infection)
- Vascular malformations (especially spinal dural AV fistulas)
- Vasculitis (primary CNS or other)
- Vitamin B_{12} deficiency

Abbreviations: AV, arteriovenous; CNS, central nervous system; HTLV, human T-cell lymphotropic virus.

As noted above, the EDSS is a widely used measure of neurologic impairment in MS (Table 455-2). Most patients with EDSS scores <3.5 walk normally and are generally not disabled; by contrast, patients with EDSS scores >4.0 have progressive MS (SPMS or PPMS), are gait-impaired, and often are occupationally disabled.

ACUTE ATTACKS OR INITIAL DEMYELINATING EPISODES

When patients experience acute deterioration, it is important to consider whether this change reflects new disease activity or a "pseudoexacerbation" resulting from an increase in ambient temperature, fever, or an infection. When the clinical change is thought to reflect a pseudoexacerbation, glucocorticoid treatment is inappropriate. Glucocorticoids are used to manage first attacks and exacerbations that are moderate to severe in severity. They provide short-term clinical benefit by reducing the degree and duration of attacks. Whether treatment provides any long-term benefit on the course of the illness is less clear. Therefore, mild attacks are often not treated. Physical and occupational therapy can help with mobility and manual dexterity.

Glucocorticoid treatment is usually administered as intravenous methylprednisolone, 500–1000 mg/d for 3–5 days, either without a taper or followed by a course of oral prednisone beginning at a dose of 60–80 mg/d and gradually tapered over 2 weeks. Orally administered methylprednisolone, prednisone, or dexamethasone (in equivalent dosages) can be substituted for the intravenous portion of the therapy. Outpatient treatment is almost always possible.

Side effects of short-term glucocorticoid therapy include fluid retention, potassium loss, weight gain, gastric disturbances, acne, and emotional lability. Concurrent use of a low-salt, potassium-rich diet and avoidance of potassium-wasting diuretics are advisable. Lithium carbonate (300 mg orally bid) may help manage emotional lability and insomnia associated with glucocorticoid therapy. Patients with a history of peptic ulcer disease may require cimetidine (400 mg bid) or ranitidine (150 mg bid). Proton pump inhibitors such as pantoprazole (40 mg orally bid) may reduce the likelihood of gastritis, especially when large doses are administered orally. Plasma exchange (five to seven exchanges: 40–60 mL/kg per exchange, every other day for 14 days) may benefit patients with fulminant attacks of demyelination that are unresponsive to glucocorticoids.

DISEASE-MODIFYING THERAPIES FOR MS

RMS More than a dozen immunomodulatory and immunosuppressive agents are in use for treatment of RMS (Table 455-6). In phase 3 clinical trials, each was shown to reduce the frequency of clinical relapses and evolution of new brain MRI lesions. Each can also be used in SPMS patients who continue to experience attacks, both because SPMS can be difficult to distinguish from RRMS and because the available clinical trials, although not all definitive, suggest that such patients may sometimes derive therapeutic benefit. Moreover, regulators now consider patients with recent relapses to be a "relapsing form of MS" (i.e., RMS), regardless of whether these patients previously had progressive disability independent from relapses. It is important to note that the relative efficacy of the different agents has not been directly tested in head-to-head studies and that cross-trial comparisons are inaccurate. However, given the increasingly complex landscape of therapeutics for MS, for convenience, the presentation of these agents was divided by an estimate of their relative (high, moderate, or modest) perceived level of efficacy. These are meant to serve as a general guide, with the caveat that considerable variance exists in practice patterns, as well as availability of these agents, in different parts of the world.

Therapy should be initiated in all patients diagnosed with RMS and those presenting with a first demyelinating event who are at high risk for MS (sometimes termed as a clinically isolated syndrome [CIS]). We favor use of the most highly effective disease-modifying therapies as first-line options for most patients. This recommendation is based on evidence from long-term prospective trials and real-world data indicating that initial treatment with highly effective agents provides outstanding control against relapsing disease, maximum protection against relapse-independent progression, and long-term outcomes, and is safe. We typically begin with an anti-CD20 B-cell-targeting drug (ocrelizumab, ofatumumab, or ublituximab) or, if these approved treatments are unavailable, with rituximab or a biosimilar. In JCV-negative patients, we begin with the cell-trafficking inhibitor natalizumab. Anti-CD20 agents are particularly attractive given their high level of efficacy, relative ease of use, favorable safety profile, and absence of rebound following discontinuation. For patients who prefer oral treatment, either an S1P modulator or a fumarate is also reasonable for first-line therapy.

First-line treatment with high-efficacy therapy has supplanted the alternative approach in which a treatment of modest or moderate effectiveness was first used and therapy advanced to a more effective agent when breakthrough disease (evident clinically or by MRI) occurred. Older first-generation therapies, such as IFN-βor glatiramer acetate, are often continued in patients who are doing well on these agents but are less commonly used today for patients with new-onset MS. Irrespective of the agent used, a change in therapy may be required in patients with suboptimal responses, such as those experiencing relapses and/or active MRI scans while on treatment, or for adverse events that may be drug-related. Pregnancy-related management is discussed later in this chapter.

Some patients, especially those with a mild initial RRMS course—e.g., a normal examination or minimal impairment (EDSS ≤2.5) and low disease activity by MRI—may initially decline therapy with a potent immunosuppressive drug. In these situations, either an oral (fumarates, S1P modulators, or teriflunomide) or injectable (IFN-β or glatiramer acetate) agent can be considered. The injectable agents (IFN-β and glatiramer acetate) have a superb long-term track record for safety but have a high nuisance factor due to the need for frequent injections, as well as bothersome side effects that reduce adherence.

As noted above, multiple lines of evidence indicate that institution of effective therapy can improve the long-term outcome of MS, including a prolongation of the time to reach disability outcomes (e.g., SPMS and requiring assistance to ambulate) and reduction in MS-related mortality. These benefits seem most conspicuous when treatment is begun early in the relapsing stage of the illness. It may be reasonable to delay initiating treatment in some patients with (1) normal neurologic

TABLE 455-6 Disease-Modifying Therapies for Multiple Sclerosis

CATEGORY AND MECHANISM OF ACTION	GENERIC NAME (TRADE NAME)	DOSE AND INTERVAL	CHARACTERISTICS	COMMENTS (USE, ADVERSE EFFECTS, ETC.)
Highly Effective				
Anti-CD20 B cell MAbs: Depletes B lymphocytes, especially motile B cells in peripheral blood; B cells in lymphoid organs variably protected; plasma cells preserved	Ocrelizumab (Ocrevus)	600-mg infusion q6 months (first dose given as two 300-mg infusions 14 days apart)	Humanized ADCC > complement	Infusion reactions usually mild; outstanding efficacy and safety in long-term RMS extension trials; also approved for PPMS
	Ofatumumab (Kesimpta)	20-mg subcutaneous injections monthly (after 3 weekly 20-mg loading doses)	Fully human Complement > ADCC	Advantage of home-based treatment; only very minor injection- related reactions
	Ublituximab (Briumvi)	450-mg infusion q6 months (first dose given as 150-mg, followed 14 days later by 450-mg, infusions)	Chimeric ADCC > Complement	
	Rituximab (Rituxan and biosimilars)	1000-mg infusion q6 months (dose used in phase 2 trial in RMS); some clinicians use 500 mg IV q6 months	Chimeric Complement > ADCC	Formally tested only in preliminary (phase 2) study
Anti-α4 subunit of $\alpha4\beta1$ integrin (adhesion molecule) MAb: Prevents lymphocytes from binding to endothelial cells and entering the CNS	Natalizumab (Tysabri)	300-mg monthly infusion	Humanized	Hypersensitivity Rxns (<10%) including anaphylaxis; NAbs in ~6%; major risk is PML (0.4%); can be given safely only if serum antibodies to JC virus are absent (~50% of patients); repeat testing q6 months with ongoing treatment; risk of rebound disease activity after cessation
Anti-CD52 MAb: Depletes lymphocytes and monocytes	Alemtuzumab (Lemtrada)	12 mg/m^2 infusion for 5 consecutive days; a second 3-day course administered 1 year later	Long-lasting benefits but serious risks limit use; approval in United States only for patients who have failed at least two other drugs	Multiple autoimmune complications including thyroid (~25%) and ITP (1–3%), malignancies, infection risk
Moderately Effective				
Sphingosine-1-phosphate (S1P) modulators: Prevents egress of lymphocytes from secondary lymphoid organs Pretreatment CBC, LFTs, ECG, eye exam required; vaccinate for VZV in seronegative patients	Fingolimod (Gilenya)	0.5 mg oral once daily	Binds to S1P1, S1P3, S1P4, and S1P5 receptors	Heart block or bradycardia can occur with first dose; a 6-h period of initial observation with ECG monitoring required; LFT abnormalities, macular edema; rare VZV or cryptococcal infections; risk of rebound disease activity after cessation for all agents in this class
	Ozanimod (Zeposia)	1 mg oral once daily	S1P1- and S1P5-selective inhibitor (cardiac receptors are mostly S1P3 and only weakly engaged by ozanimod)	Up-titration regimen used to begin treatment; first-dose monitoring not required for most patients
	Ponesimod (Ponvory)	20 mg oral once daily	S1P1-selective modulator	Up-titration regimen used to begin treatment; initial dose requires 4-h cardiac monitoring for patients with heart rate <55 beats/min
	Siponimod (Mayzent)	Based on CYP2C9 genotype. 1 mg oral daily for pts with CYP2C9 1/*3 or 2/*3 Dose reduced in patients with the CYP2C9 *3/*3 genotype (<0.5% of the population) due to substantially elevated drug levels	S1P1- and S1P5-selective modulator	Approved for SPMS with active disease (relapses or new focal MRI lesions); first-dose monitoring only for patients with sinus bradycardia, heart block, or prior myocardial infarction or heart failure
Fumarate: Immunomodulator; reduces proinflammatory and increases regulatory cytokines; inhibits degradation of Nrf2, increasing natural antioxidants	Dimethyl fumarate (Tecfidera)	240 mg oral twice daily (half-dose for first 7 days)	Metabolized to active compound monomethyl fumarate	Gastrointestinal side effects, flushing; these may improve over time; monitor for LFT abnormalities and for lymphopenia (which can persist after drug cessation); rare PML cases
	Diroximel fumarate (Vumerity)	262 mg oral twice daily	Metabolized to active compound monomethyl fumarate	Similar side effect profile as dimethyl fumarate
2-Chlorodeoxyadenosine: Lymphocytotoxic; possibly followed by reconstitution by nonpathogenic immune cells	Cladribine (Mavenclad)	Weight-based oral dosing (3.5 mg/kg) divided over 4–5 days, repeated 23–27 days later; a second identical course is administered 1 year later	Purine analogue prodrug phosphorylated in lymphocytes and incorporated into DNA, triggering apoptosis; long-lasting	Long-lasting benefits but use limited by risks of malignancy, teratogenicity, and infection including PML

(*Continued*)

TABLE 455-6 Disease-Modifying Therapies for Multiple Sclerosis (Continued)

CATEGORY AND MECHANISM OF ACTION	GENERIC NAME (TRADE NAME)	DOSE AND INTERVAL	CHARACTERISTICS	COMMENTS (USE, ADVERSE EFFECTS, ETC.)
Modestly Effective				
Glatiramer acetate: Immunomodulator; reduces proinflammatory and increases regulatory cytokines; induces antigen-specific suppressor T cells; binds MHC molecules	Glatiramer acetate (Copaxone)	Subcutaneous injection 20 mg daily, or alternatively, 40 mg three times weekly	Synthetic random polypeptide of four amino acids (L-glutamic acid, L-lysine, L-alanine, L-tyrosine)	Well tolerated; injection site reactions; ~15% of patients experience one (or less often more than one) episode of flushing, chest tightness, dyspnea, palpitations, anxiety
Interferon (IFN)-β: Immunomodulator; reduces proinflammatory and increases regulatory cytokines; interferes with antigen presentation, T-cell proliferation, lymphocyte trafficking	Interferon-β-1a (Avonex)	Intramuscular injections 30 mg once weekly	Neutralizing antibodies in 2–10% (can decrease over time with all IFN preparations)	With all IFN preparations: flu-like symptoms (fever, chills, myalgias) common; managed with NSAIDs; mild lab abnormalities (LFTs, lymphopenia); rare cases of severe hepatotoxicity
	Interferon-β-1a (Rebif)	Subcutaneous injections 44 mg three times per week	Neutralizing antibodies in 15–25%	
	Interferon-β-1b (Betaseron or Extavia)	Subcutaneous injections 250 mg every other day	Neutralizing antibodies in 30–40%	
	Pegylated interferon-β-1a (Plegridy)	Subcutaneous injections 125 mg every 14 days	Pegylation increases half-life; neutralizing antibodies in <1%	
Teriflunomide: Anti-inflammatory; limits proliferation of rapidly dividing B and T lymphocytes	Teriflunomide (Aubagio)	14 mg oral daily	Inhibits mitochondrial enzyme dihydro-orotate dehydrogenase involved in de novo pyrimidine synthesis; cytostatic rather than cytotoxic	Hair thinning, gastrointestinal toxicity (nausea and diarrhea), rarely toxic epidermal necrolysis or Stevens-Johnson syndrome; long-lasting teratogenicity (elimination protocol with cholestyramine or activated charcoal)

Abbreviations: ADCC, antibody-dependent cell-mediated cytotoxicity; CBC, complete blood count; CNS, central nervous system; ECG, electrocardiogram; ITP, immune thrombocytopenia; LFT, liver function test; MAbs, monoclonal antibodies; MHC, major histocompatibility complex; MRI, magnetic resonance imaging; NAb, neutralizing antibody; NSAID, nonsteroidal anti-inflammatory drug; PML, progressive multifocal leukoencephalopathy; PPMS, primary progressive multiple sclerosis; RMS, relapsing multiple sclerosis; Rxn, reaction; SPMS, secondary progressive multiple sclerosis; VZV, varicella-zoster virus.

examinations, (2) a single attack or a low attack frequency, and (3) a low burden of disease as assessed by brain MRI. Untreated patients, however, should be followed closely with periodic brain MRI scans; the need for therapy is reassessed if scans reveal evidence of ongoing, subclinical disease. Finally, vitamin D deficiency should be corrected in all patients with MS, and generally this requires oral supplementation with vitamin D_3, 4000 IU daily. Several clinical trials showed that supplementation with vitamin D in relapsing MS patients reduces MRI measures of disease activity and may also reduce the relapse frequency in patients actively treated with either IFN or glatiramer acetate.

SPMS For patients with active SPMS, either ocrelizumab or siponimod is a reasonable first-line option. Ocrelizumab is approved for active SPMS despite not having been specifically studied in this patient population. Siponimod in a single pivotal study reduced the risk of progression in SPMS; however, subgroup analysis showed that patients with a relapse in the 2 years prior to treatment and those with contrast-enhancing lesions on brain MRI received the most therapeutic benefit. Regulatory bodies also approved cladribine and ponesimod for active SPMS despite neither having been specifically studied in this MS subgroup.

PPMS Ocrelizumab was shown to reduce progression of clinical disability in PPMS by 25% and also improve other clinical and MRI markers of inflammatory and degenerative disease activity. Although the magnitude of the effect in PPMS is lower than in RMS, for the average patient with PPMS, these data translate to the expectation that >7 years of wheelchair-independent function is gained on average. Ocrelizumab is the only agent convincingly shown to modify the course of PPMS.

■ OTHER OFF-LABEL TREATMENT OPTIONS

Autologous hematopoietic stem cell transplantation appears to be highly effective in reducing relapses and may improve disability in relapsing MS. It appears to be largely ineffective for patients with progressive MS. Stem cell transplantation also carries significant risk, however, including toxicities from chemotherapy-conditioning regimens. Ongoing clinical trials should help to better position this procedure with respect to available pharmacologic interventions.

Intravenous immunoglobulin (IVIg), administered in monthly pulses (up to 1 g/kg) for up to 2 years, appears to reduce annual exacerbation rates. However, its use is limited because of its high cost, questions about optimal dose, and uncertainty about any impact on long-term disability. It can be considered when the risks of immunosuppression preclude use of other MS agents.

Numerous clinical trials of promising experimental therapies are currently underway. These include studies testing higher doses of ocrelizumab; Bruton's tyrosine kinase (BTK) inhibitors to selectively deplete B cells, plasma cells, and microglia; CD19-targeted chimeric antigen receptor (CAR) T cells; and molecules to promote remyelination.

THERAPIES TO AVOID

Many purported treatments for MS have never been subjected to scientific scrutiny. These include dietary therapies (e.g., the Swank diet, the Paleo diet, the Wahls diet), megadose vitamins, calcium orotate, bee stings, cow colostrum, hyperbaric oxygen, procarin (a combination of histamine and caffeine), chelation, acupuncture, acupressure, various Chinese herbal remedies, and removal of mercury-amalgam tooth fillings, among others. Although infections with EBV, human herpesvirus (HHV) 6, or other agents are plausibly involved in MS, treatment with antiviral agents or antibiotics is not recommended. A chronic cerebrospinal insufficiency (CCSVI) was proposed as a cause of MS, and surgical intervention with vascular repair was recommended; however, multiple studies failed to confirm the initial claims. A double-blind trial of high-dose biotin to improve disability in progressive forms of MS also found no benefit. Patients should avoid costly or potentially hazardous unproven treatments, many of which also lack biologic plausibility.

SYMPTOMATIC THERAPIES

For all patients, it is important to encourage attention to a healthy lifestyle, including maintaining an optimistic outlook, a healthy diet, and regular exercise as tolerated (swimming is often well-tolerated because of the cooling effect of cold water). It is reasonable also to correct vitamin D deficiency with oral vitamin D and consider dietary supplementation with long-chain fatty acids (such as omega-3 oil tablets) due to their mild immunomodulatory effects.

Bladder dysfunction management is best guided by urodynamic testing because symptoms correlate poorly with the specific pathophysiology, which can also change over time as the disease evolves. The underlying cause can be bladder hyperreflexia, atony, or dyssynergia between the detrusor and the external sphincter muscle.

Detrusor hyperreflexia can be initially managed with evening fluid restriction or frequent voluntary voiding. If these methods fail, beta-3 adrenergic agonists such mirabegron (25–50 mg/d) and vibegron (75 mg/d) that relax bladder smooth muscle should be tried. Beta-3 adrenergic agonists are preferred over anticholinergic agents such as oxybutynin (5–15 mg/d), propantheline bromide (10–15 mg/d), tolterodine tartrate (2–4 mg/d), or solifenacin (5–10 mg/d) because anticholinergic side effects can worsen other MS symptoms including cognitive dysfunction. Co-administration of pseudoephedrine (30–60 mg) with anticholinergics is sometimes beneficial. Detrusor muscle injections of botulinum toxin (e.g., onabotulinumtoxinA 200 IU) can be useful when anticholinergics are ineffective or produce side effects such as cognitive dysfunction or fatigue.

An *atonic bladder* due to loss of reflex bladder wall contraction may respond to bethanechol (30–150 mg/d), and *detrusor/sphincter dyssynergia* may respond to phenoxybenzamine (10–20 mg/d) or terazosin hydrochloride (1–20 mg/d). However, both conditions often require catheterization.

Urinary tract infections should be treated promptly. Patients with postvoid residual urine volumes >200 mL are predisposed to infections. Prevention by urine acidification (with cranberry juice or vitamin C) inhibits some bacteria. Prophylactic administration of antibiotics is sometimes necessary but may lead to colonization by resistant organisms. Intermittent catheterization may help to prevent recurrent infections and reduce overflow incontinence.

Treatment of *constipation* includes high-fiber diets and fluids. Natural or other laxatives may help. *Fecal incontinence* may respond to a reduction in dietary fiber.

Spasticity and *spasms* may improve with physical therapy, regular exercise, and stretching. Avoidance of triggers (e.g., infections, fecal impactions, bed sores) is extremely important. Effective medications include baclofen (20–120 mg/d), diazepam (2–40 mg/d), tizanidine (8–32 mg/d), dantrolene (25–400 mg/d), and cyclobenzaprine hydrochloride (10–60 mg/d). For severe spasticity, a baclofen pump (delivering medication directly into the CSF) can provide substantial relief.

Weakness can sometimes be improved with the use of potassium channel blockers such as 4-aminopyridine (20 mg/d) and 3,4-diaminopyridine (40–80 mg/d), particularly in the setting where lower-extremity weakness interferes with the patient's ability to ambulate. Extended-release 4-aminopyridine (10 mg twice daily) can be obtained either as dalfampridine (Ampyra) or through a compounding pharmacy. The principal concern with the use of these agents is the possibility of inducing seizures at high doses.

Ataxia/tremor is often intractable. Clonazepam (0.5–2 mg/d), primidone (50–250 mg/d), propranolol (40–200 mg/d), or ondansetron (8–16 mg/d) may help. Wrist weights occasionally reduce tremor in the arm or hand. Thalamotomy and deep-brain stimulation have been tried with mixed success.

Pain is treated with anticonvulsants (gabapentin [300–3600 mg/d]; pregabalin [50–300 mg/d]; carbamazepine [100–1000 mg/d]; phenytoin [300–600 mg/d]); tricyclic antidepressants (amitriptyline [25–100 mg/d], nortriptyline [25–100 mg/d], desipramine [100–300 mg/d]); serotonin-norepinephrine reuptake inhibitors (duloxetine [20–120 mg/d] or venlafaxine [75–225 mg/d]); or antiarrhythmics (mexiletine, 300–900 mg/d). If these approaches fail, patients should be referred to a comprehensive pain-management program.

Depression should be actively treated. Useful drugs include the selective serotonin reuptake inhibitors (escitalopram [10–20 mg/d], fluoxetine [20–80 mg/d], or sertraline [50–200 mg/d]), tricyclic antidepressants (amitriptyline [25–150 mg/d], nortriptyline [25–150 mg/d], or desipramine [100–300 mg/d]) and mixed norepinephrine/serotonin reuptake inhibitors (duloxetine [20–120 mg/d] or venlafaxine [75–225 mg/d]).

Fatigue may improve with assistive devices, help in the home, or successful management of spasticity. Careful attention to medications that could contribute to fatigue is often helpful. For example, patients who require anticholinergic medication for nocturia may benefit from dosing at bedtime only. Excessive daytime somnolence caused by MS may respond to methylphenidate (5–25 mg/d), modafinil (100–400 mg/d), or armodafinil (150–250 mg/d).

Cognitive problems may respond marginally to lisdexamfetamine (40 mg/d).

Paroxysmal symptoms respond dramatically to low-dose anticonvulsants (acetazolamide [200–600 mg/d], carbamazepine [50–400 mg/d], phenytoin [50–300 mg/d], or gabapentin [600–1800 mg/d]).

Heat sensitivity may respond to heat avoidance, air-conditioning, or cooling garments.

Sexual dysfunction may be helped by lubricants to aid in genital stimulation and sexual arousal. Management of pain, spasticity, fatigue, and bladder/bowel dysfunction may also help. Sildenafil (50–100 mg), tadalafil (5–20 mg), or vardenafil (5–20 mg), taken 1–2 h before sex, are standard treatments for erectile dysfunction.

PROGNOSIS

Historically, most patients with MS ultimately experienced progressive neurologic disability. In older studies conducted before disease-modifying therapies for MS were available, 15 years after onset, only 20% of patients had no functional limitation, and between one-third and one-half of RMS patients progressed to SPMS and required assistance with ambulation; furthermore, 25 years after onset, ~80% of MS patients reached this level of disability. Long-term studies from the early treatment era indicated clearly that prognosis had improved substantially, with a two- to threefold slowing of transition from RMS to SPMS. And currently with high-efficacy therapy widely available, the prognosis continues to improve; relapses are largely eliminated and relapse-independent progression has been further slowed. For example, the ocrelizumab extension trials revealed that after 10 years of continuous treatment, nearly 80% of patients with RMS experienced no disease worsening and more patients had improved than worsened, and in PPMS, more than one-third had experienced no worsening and more than 80% were still ambulatory. As noted above, the long-term course may improve further as highly efficacious agents are increasingly employed early in the disease course. However, many patients with progressive MS still continue to worsen despite best available therapy, and for these patients, more effective therapies are sorely needed.

Natural history studies from the pretreatment era indicated that certain clinical features suggest a more favorable prognosis. These include ON or sensory symptoms at onset; fewer than two relapses in the first year of illness; and minimal impairment after 5 years. Predictors of an early aggressive course of the illness include an older age at symptom onset, male gender, greater disability, and the appearance of motor signs during the first year of the illness. Importantly, some MS patients, estimated at <10%, have a benign variant of MS and never develop neurologic disability even when untreated.

PREGNANCY-RELATED ISSUES

Pregnant MS patients experience fewer attacks during gestation (especially in the last trimester) but more attacks in the first 3 months postpartum. When considering the pregnancy year as a whole (i.e., 9 months of pregnancy plus 3 months postpartum), the overall disease course is unaffected. Disease-modifying therapy is generally discontinued during pregnancy, and special care needs to be taken with agents (natalizumab and the S1P drugs) that have risk of rebound disease activity following discontinuation. Replacement of these agents with an anti-CD20 B-cell therapy such as ocrelizumab before conception appears to mitigate this risk and also provide long-lasting protection that extends throughout the pregnancy and immediate postpartum period. Intravenously administered B-cell therapies also provide an attractive management option for many patients contemplating pregnancy, and our practice is to advise patients to stop receiving infusions 3–4 months prior to attempting to conceive. Earlier studies raised concerns that drugs used for fertility treatments might worsen MS disease

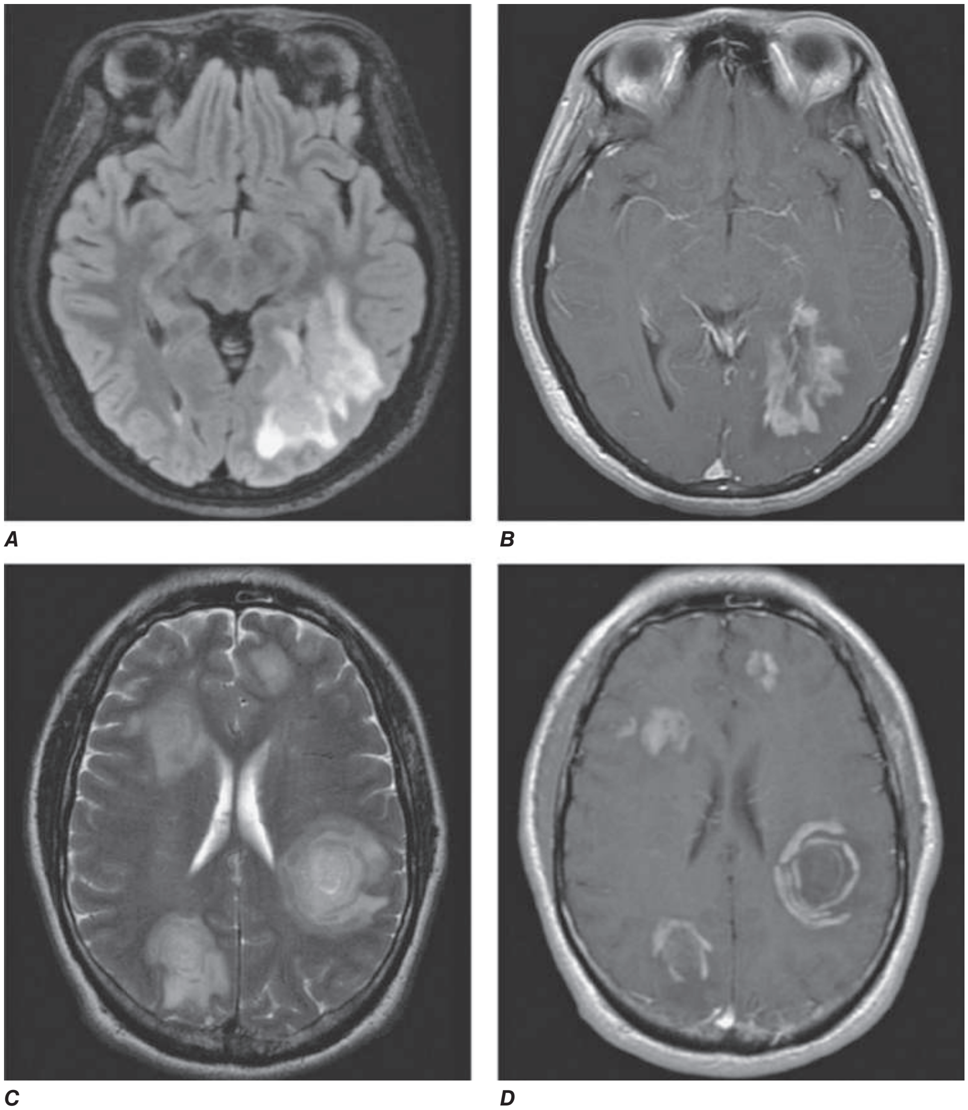

FIGURE 455-4 Magnetic resonance imaging findings in variants of MS. ***A*** and ***B***. Acute tumefactive MS. In ***A***, a sagittal T2-weighted fluid-attenuated inversion recovery (FLAIR) image of a large solitary right parieto-occipital white matter lesion is shown, with effacement of overlying cortical sulci consistent with mass effect. In ***B***, T1-weighted image obtained after the intravenous administration of gadolinium diethylene triamine pentaacetic acid (DTPA) reveals a large serpiginous area of blood-brain barrier disruption consistent with acute inflammation. ***C*** and ***D***. Balo's concentric sclerosis. In ***C***, an axial T2-weighted sequence shows multiple areas of abnormal ovoid bright signal in the supratentorial white matter bilaterally; some lesions reveal concentric layers, typical of Balo's concentric sclerosis. In ***D***, T1-weighted magnetic resonance images after gadolinium demonstrate abnormal enhancement of all lesions with some lesions demonstrating concentric ring enhancement.

activity, but more recent studies indicate that these produce little if any additional risk.

CLINICAL VARIANTS OF MS

Acute or fulminant MS (*Marburg's variant*) is an aggressive demyelinating process that in some cases progresses inexorably to death within 1–2 years. Typically, there are no remissions. Marburg's variant does not seem to follow infection or vaccination, and it is unclear whether this syndrome represents an extreme form of MS or another disease altogether. When an acute demyelinating syndrome presents as a solitary expansile lesion, a brain tumor is often suspected **(Fig. 455-4*A*, *B*)**. Such cases are designated tumefactive MS, and a brain biopsy may be required to establish the diagnosis. *Balo's concentric sclerosis* is another fulminant demyelinating syndrome characterized by concentric brain or spinal cord lesions with alternating spheres of demyelination and remyelination **(Fig. 455-4*C*, *D*)**. For these fulminant demyelinating states, no controlled trials of therapy exist; high-dose glucocorticoids and plasma exchange are often used, with uncertain benefit.

FURTHER READING

Absinta M et al: Mechanisms underlying progression in multiple sclerosis. Curr Opin Neurol 33:277, 2020.

Bjornevik K et al: Epstein-Barr virus as a leading cause of multiple sclerosis: Mechanisms and implications. Nat Rev Neurol 19:160, 2023.

Brown JWL et al: Association of initial disease-modifying therapy with later conversion to secondary progressive multiple sclerosis. JAMA 321:175, 2019.

Cree BAC et al: Silent progression in disease activity-free relapsing multiple sclerosis. Ann Neurol 85:653, 2019.

Graham EL et al: Practical considerations for managing pregnancy in patients with multiple sclerosis: Dispelling the myths. Neurol Clin Pract 14:e200253, 2024.

Hauser SL et al: Safety of ocrelizumab in patients with relapsing and primary progressive multiple sclerosis. Neurology 97:e1546, 2021.

International Multiple Sclerosis Genetics Consortium; Multiple MS Consortium: Locus for severity implicates CNS resilience in progression of multiple sclerosis. Nature 619:323, 2023.

Kuhlmann T et al: Multiple sclerosis progression: Time for a new mechanism-driven framework. Lancet Neurol 22:78, 2023.

Portaccio E et al: Progression is independent of relapse activity in early multiple sclerosis: A real-life cohort study. Brain 145:2796, 2022.

Thompson AJ et al: Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 17:162, 2018.

Woo MS et al: The neuropathobiology of multiple sclerosis. Nat Rev Neurosci 25:493, 2024.

456 Neuromyelitis Optica

Bruce A. C. Cree, Stephen L. Hauser

Neuromyelitis optica (NMO) is an aggressive, antibody-mediated, inflammatory disorder characterized by recurrent attacks of optic neuritis (ON) and myelitis; the more inclusive term *NMO spectrum disorder* (NMOSD) was proposed to incorporate individuals with partial forms and those with involvement of additional structures in the central nervous system **(Table 456-1)**. NMO is more frequent in women than men (9:1) and typically begins in adulthood (40-year mean age of onset) but can arise at any age. An important consideration, especially early in its presentation, is distinguishing between NMO and multiple sclerosis (MS; **Chap 455**). In patients with NMO, attacks of ON can be bilateral and produce severe vision loss (uncommon in MS); myelitis can be severe and transverse (rare in MS) and is typically longitudinally extensive **(Fig. 456-1)** involving three or more contiguous vertebral segments. In contrast to MS, progressive symptoms only rarely occur in NMO, and accumulation of disability is caused by attack-related injury. The brain magnetic resonance imaging (MRI) was earlier thought to be normal in NMO, but it is now recognized that in many cases brain lesions are present, including areas of nonspecific signal change as well as lesions associated with specific syndromes such as the area postrema in the lower medulla presenting as intractable hiccoughs or vomiting; the hypothalamus causing a sleep disorder or endocrinopathy; or the cerebral hemispheres producing focal symptoms, encephalopathy, or seizures. Large MRI lesions in the cerebral hemispheres can be asymptomatic, sometimes have a "cloud-like" appearance, and unlike MS lesions, are often not destructive and can resolve completely. Spinal cord MRI lesions typically consist of focal enhancing areas of swelling and tissue destruction, extending over three or more spinal cord segments, and on axial sequences, these are centered on the gray matter of the cord. Cerebrospinal fluid (CSF) findings include pleocytosis greater than that observed in MS, with neutrophils and eosinophils present in many acute cases; oligoclonal bands (OCBs) are uncommon, occurring in <20% of NMO patients. The pathology of NMO is a distinctive astrocytopathy with inflammation, loss of astrocytes, and an absence of staining of the water channel protein aquaporin-4 (AQP4) by immunohistochemistry, plus thickened blood vessel walls, demyelination, and deposition of antibody and complement.

TABLE 456-1 Diagnostic Criteria for Neuromyelitis Optica Spectrum Disorder (NMOSD)

Diagnostic Criteria for NMOSD with AQP4-IgG
1. At least one core clinical characteristic
2. Positive test for AQP4-IgG using best available detection method (cell-based assay strongly recommended)
3. Exclusion of alternative diagnoses
Diagnostic Criteria for NMOSD without AQP4-IgG or NMOSD with Unknown AQP4-IgG Status
1. At least two core clinical characteristics occurring as a result of one or more clinical attacks and meeting all of the following requirements:
a. At least one core clinical characteristic must be optic neuritis, acute myelitis with LETM, or area postrema syndrome
b. Dissemination in space (two or more different clinical characteristics)
c. Fulfillment of additional MRI requirements, as applicable
2. Negative test for AQP4-IgG using best available detection method or testing unavailable
3. Exclusion of alternative diagnoses
Core Clinical Characteristics
1. Optic neuritis
2. Acute myelitis
3. Area postrema syndrome: episode of otherwise unexplained hiccups or nausea or vomiting
4. Acute brainstem syndrome
5. Symptomatic narcolepsy or acute diencephalic clinical syndrome (hypothalamic dysfunction) with NMOSD-typical diencephalic MRI lesions
6. Symptomatic cerebral syndrome with NMOSD-typical brain lesions
Additional MRI Requirements for NMOSD without AQP4-IgG and NMOSD with Unknown AQP4-IgG Status
1. Acute optic neuritis: requires brain MRI showing (a) normal findings or only nonspecific white matter lesions, OR (b) optic nerve MRI with T2-hyperintense lesion of T1-weighted gadolinium-enhancing lesion extending over >1/2 optic nerve length or involving optic chiasm
2. Acute myelitis: requires associated intramedullary MRI lesion extending ≥3 contiguous segments (LETM) OR ≥3 contiguous segments of focal spinal cord atrophy in patients with history compatible with acute myelitis
3. Area postrema syndrome requires associated dorsal medulla/area postrema lesions
4. Acute brainstem syndrome requires periependymal brainstem lesions

Abbreviations: AQP4, aquaporin-4; LETM, longitudinally extensive transverse myelitis; MRI, magnetic resonance imaging.

Source: Reproduced with permission from DM Wingerchuk et al: International consensus diagnostic criteria for neuromyelitis optica spectrum disorders. Neurology 85:177, 2015.

NMO is an autoimmune disease associated with a highly specific autoantibody directed against the water channel protein AQP4 that is present in the sera of ~90% of affected patients. AQP4 is localized to the foot processes of astrocytes in close apposition to endothelial surfaces, as well as at paranodal regions near nodes of Ranvier. Pathologic studies in NMO show loss of AQP4 associated with antibody and complement deposition and cytolysis of astrocytes within NMO lesions. It is likely that AQP4 antibodies are pathogenic because passive transfer of AQP4 antibodies into laboratory animals can reproduce histologic features of the disease. Both complement fixation and antibody-dependent cell-mediated cytotoxicity (ADCC) are thought to contribute to astrocyte injury. Proinflammatory T lymphocytes of the T_H17 type recognize an immunodominant epitope of AQP4 and may also contribute to pathogenesis. Recent data have identified a population of B lymphocytes in the thymus that constitutively express AQP4 and interact with AQP4-specific T cells to produce antigen-specific tolerance; this interaction might be impaired in NMO. During acute attacks of myelitis, CSF levels of interleukin-6 (IL-6; a proinflammatory cytokine) and glial fibrillary acidic protein (GFAP; an astrocyte protein) are markedly elevated, consistent with active inflammation and astrocyte injury. Because of the high specificity of the antibody, its presence is diagnostic when found in conjunction with a typical clinical presentation. Anti-AQP4 seropositive patients have a high risk for future attacks; more than half will relapse within 1 year if untreated.

CLINICAL COURSE

NMO is typically a recurrent disease; the course is monophasic in <10% of patients. Individuals who test negative for AQP4 antibodies are somewhat more likely to have a monophasic course. Untreated NMO is usually severely disabling over time; in one series, respiratory failure from cervical myelitis was present in one-third of patients, and 8 years after onset, 60% of patients were blind and more than half had permanent paralysis of one or more limbs. Observational data suggest that the long-term course of NMO has been substantially improved since the development of therapies to treat acute attacks and prevent relapses. The 5-year survival rate appears to have increased from 68–75% around 1999 to 91–98% as of 2017, a change presumably due to improved diagnosis and widespread use of immune-suppressing therapies.

GLOBAL CONSIDERATIONS

The incidence and prevalence of NMO shows considerable variation between populations and geographic regions, with prevalence estimates that range from <1 to >4 per 100,000. Although NMO can occur in people of any ethnic background, individuals of Asian and African origin are disproportionately affected, with the highest reported prevalence in Martinique. Among white populations, MS **(Chap. 455)** is far more common than NMO.

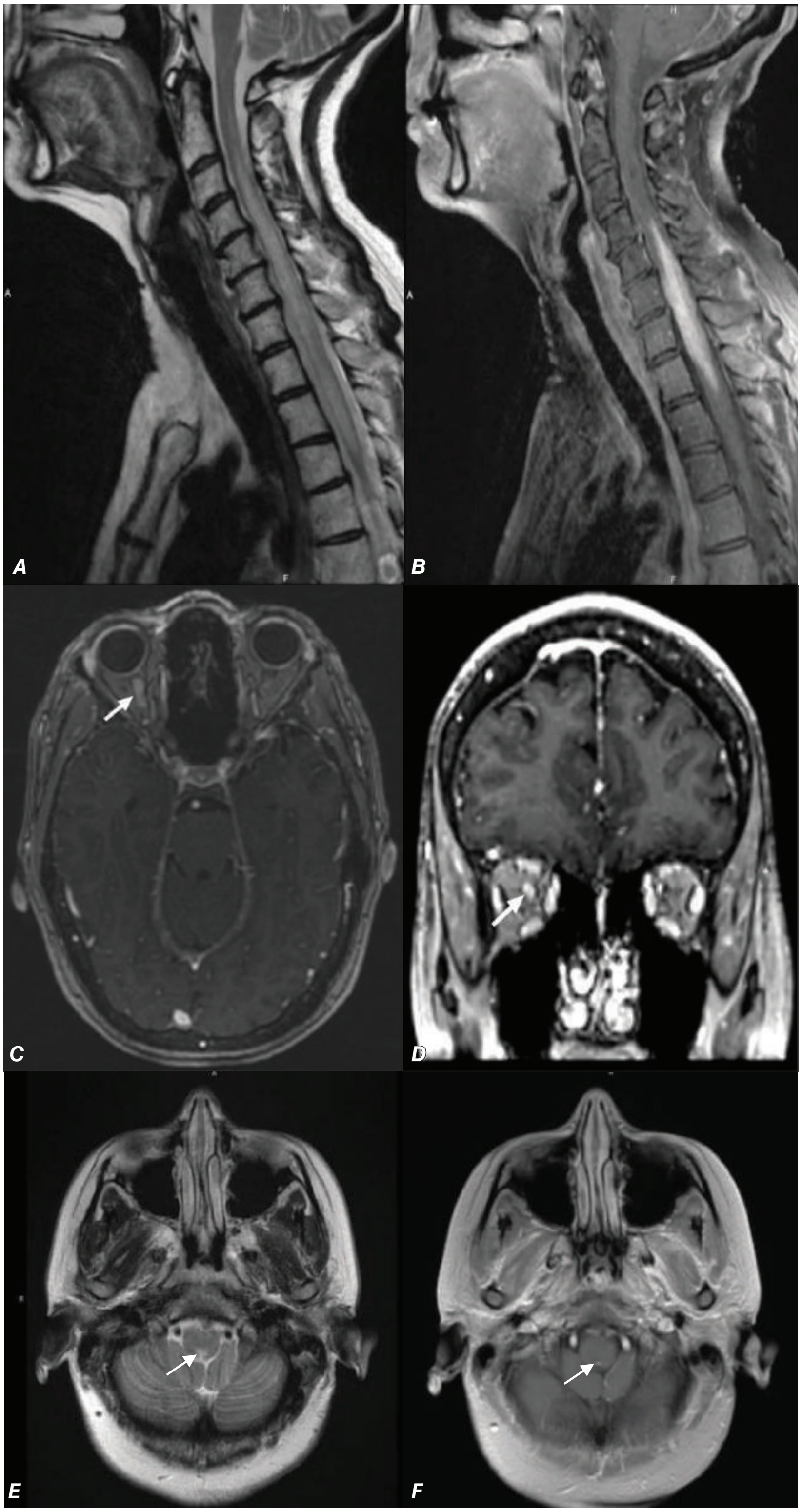

FIGURE 456-1 Imaging findings in neuromyelitis optica: longitudinally extensive transverse myelitis, optic neuritis, and brainstem involvement. ***A.*** Sagittal fluid-attenuated inversion recovery (FLAIR) cervical spine MRI showing an area of increased signal change on T2-weighted imaging spanning more than three vertebral segments in length. ***B.*** Sagittal T1-weighted cervical spine MRI following gadolinium-diethylene triamine pentaacetic acid (DPTA) infusion showing enhancement. ***C and D.*** Axial and coronal brain MRI following gadolinium DPTA infusion shows enhancement of the right optic nerve. ***E.*** Axial brain MRI shows an area of hyperintense signal on T2-weighted imaging within the area postrema (*arrow*). ***F.*** Axial T1-weighted brain MRI following gadolinium-DPTA infusion shows punctate enhancement of the area postrema (arrow).

Interestingly, when MS affects individuals of African or Asian ancestry, there is a propensity for demyelinating lesions in some individuals to involve predominantly the optic nerve and spinal cord, an MS subtype termed *opticospinal MS*. Some patients with opticospinal MS are seropositive for AQP4 antibodies, indicating that they represent NMOSD.

ASSOCIATED CONDITIONS

Up to 40% of NMO patients have a systemic autoimmune disorder, such as systemic lupus erythematosus, Sjögren's syndrome, perinuclear antineutrophil cytoplasmic antibody (p-ANCA)–associated vasculitis, myasthenia gravis, Hashimoto's thyroiditis, or mixed connective tissue disease. This is another feature distinct from MS; MS patients rarely have other comorbid autoimmune diseases other than hypothyroidism. In some NMO cases, onset may be associated with acute infection with varicella-zoster virus, Epstein-Barr virus, HIV, or tuberculosis. Rare cases appear to be paraneoplastic and associated with breast, lung, or other cancers.

TREATMENT

Neuromyelitis Optica

ACUTE ATTACKS

Acute attacks are usually treated with high-dose glucocorticoids (e.g., methylprednisolone 1 g/d for 5–10 days followed by a prednisone taper). Plasma exchange (typically 5–7 exchanges of 1.5 plasma volumes/exchange) is used empirically for acute episodes that do not respond to glucocorticoids.

PROPHYLAXIS AGAINST RELAPSES

Given the unfavorable natural history of untreated NMO, prophylaxis against relapses is recommended for nearly all patients. Four monoclonal antibody medications are approved for attack prevention in NMO: an anti-CD19 B-cell depleter (inebilizumab), an IL-6 receptor blocker (satralizumab), and two terminal complement inhibitors (eculizumab and ravulizumab) **(Table 456-2)**. It is our general practice to begin therapy with either inebilizumab or satralizumab and to use complement inhibitors as second-line treatment for nonresponders. None of the U.S. Food and Drug Administration–approved medications are approved for AQP4-negative patients, and in practice, obtaining these medications for seronegative patients is challenging from a payer perspective. In AQP4-seronegative patients, the risk of a relapse is lower (approximately half of patients have only a single attack), and in these patients, our practice is to begin treatment with empiric therapies such as rituximab or mycophenolate mofetil. Because AQP4-seronegative patients require involvement of both optic neuritis and longitudinally extensive myelitis to meet criteria for seronegative NMOSD, the majority of AQP4-seronegative patients will have had more than one attack.

Inebilizumab (Uplizna) Inebilizumab is a humanized affinity-optimized, afucosylated monoclonal antibody that binds to the B-cell surface antigen CD19 and depletes a wide range of B cells including pre-B cells and some plasmablasts. Expression of CD19 starts at an earlier stage of B-cell development than CD20, and in contrast to CD20, which is absent from plasmablasts and mature plasma cells, CD19 is expressed on plasmablasts, as well as a proportion of plasma cells in secondary lymphoid organs and bone marrow. In the pivotal clinical trial that included individuals both positive and negative for AQP4, inebilizumab reduced the time to the first attack by 77% compared to placebo (hazard ratio, 0.23; *p* <.0001). Inebilizumab-treated participants also had reduced rates of hospitalizations, disability worsening, and new MRI lesions.

Inebilizumab is dosed as an initial 300-mg IV infusion followed 2 weeks later by a second 300-mg IV infusion, with subsequent doses of 300-mg infusions every 6 months thereafter. Inebilizumab is associated with a dose dependent decline in serum IgG levels and with neutropenia in some patients.

TABLE 456-2 Attack Risk Reductions of Approved Neuromyelitis Optica Spectrum Disorder (NMOSD) Treatments

	RISK REDUCTION IN AQUAPORIN-4–SEROPOSITIVE NMOSD
Eculizumab (add-on to immune suppression)	94%, *p* <.001
Ravulizumab (add-on to immune suppression)	100%, *p* <.001
Inebilizumab (monotherapy)	78%, *p* = .01
Satralizumab (monotherapy)	77%, *p* <.001
Satralizumab (add-on to immune suppression)	74%, *p* = .001

Satralizumab (Enspryng) Satralizumab is a monoclonal antibody that binds to the IL-6 receptor, blocking engagement of IL-6, a proinflammatory cytokine that is upregulated in the CSF during acute NMO attacks. Satralizumab was investigated in NMOSD in two registration trials: one as monotherapy and the other as add-on therapy. Both AQP4-seropositive and AQP4-seronegative participants were enrolled. In the monotherapy study, the risk of attack was reduced by 74% with satralizumab (*p* = .014), and in the add-on study, the risk of attack was reduced by 78% (*p* = .014). Although both studies recruited substantial numbers of AQP4-seronegative participants, there was no clinically meaningful impact of satralizumab in seronegative participants.

Satralizumab is administered as a loading dosage of 120 mg by subcutaneous injection at weeks 0, 2, and 4, followed by a maintenance dose of 120 mg every 4 weeks. Screening for hepatitis B virus, tuberculosis, and liver transaminase elevations is required before starting satralizumab. Aspartate aminotransferase (AST) and alanine aminotransferase (ALT) should be monitored during treatment for transaminase elevations and complete blood count (CBC) monitored for neutropenia. Satralizumab is also associated with weight gain; body weight increases of at least 7% from baseline occurred in 30% of satralizumab-treated participants compared to 8% of those treated with placebo.

Eculizumab (Soliris) The monoclonal antibody eculizumab binds with high affinity to the complement protein C5, inhibiting its cleavage into C5a and C5b and thereby preventing generation of the terminal complement attack complex C5b-9. Eculizumab was investigated as an add-on therapy in AQP4-seropositive NMO. The time to first attack was longer in patients treated with eculizumab compared to placebo (relative risk reduction, 94%; hazard ratio, 0.058; *p* <.0001) Eculizumab-treated patients also had a 96% relative reduction in the annualized attack rate, as well as dramatically reduced rates of hospitalizations and need for glucocorticoids or plasma exchange for acute attacks. Eculizumab is dosed as follows: 900 mg weekly for the first 4 weeks, followed by 1200 mg for the fifth dose 1 week later, and then 1200 mg every 2 weeks thereafter.

Because of its toxicity profile, eculizumab is available only through a restricted program under a Risk Evaluation and Mitigation Strategy (REMS). Life-threatening and fatal meningococcal infections have occurred (boxed warning). Eculizumab-treated patients must be immunized with meningococcal vaccines (MenACWY two doses at least 8 weeks apart plus either MenB-4C [two doses at least 1 month apart] or MenB-FHbp [three doses over 6 months]) at least 2 weeks prior to administering the first dose unless the risks of delaying eculizumab therapy outweigh the risks of developing a meningococcal infection. The MenACWY vaccine requires revaccination every 5 years, and the MenB-4C/MenB-FHbp requires a booster 1 year after vaccination and then every 2–3 years with ongoing treatment. Vaccination reduces, but does not eliminate, the risk of meningococcal infections. Eculizumab-treated patients must be monitored closely for early signs of meningococcal infections and evaluated immediately if infection is suspected.

Ravulizumab (Ultomiris) Like eculizumab, ravulizumab is a terminal complement inhibitor. Ravulizumab was developed using antibody recycling technology to increase its half-life, thereby allowing less frequent administration than its parent compound eculizumab. In a single-arm, open-label study of ravulizumab in NMOSD, ravulizumab-treated participants did not experience any attacks during 50 weeks of observation. A historical control arm from the eculizumab PREVENT clinical trial was used for comparison. Ravulizumab-treated patients also require vaccination for meningococcal infections. It seems likely that ravulizumab produces even more potent complement inhibition than eculizumab, resulting in apparently greater efficacy but also a higher risk of meningococcal infection. Ravulizumab is administered intravenously with a weight-based loading dose followed by maintenance dosing every 8 weeks.

OTHER TREATMENTS

Prior to the approval of disease-modifying therapies for NMO, several empiric regimens were commonly used including mycophenolate mofetil (1000 mg bid); rituximab, a B-cell-depleting anti-CD20 monoclonal antibody (2 g IV every 6 months); or a combination of glucocorticoids (500 mg IV methylprednisolone daily for 5 days, then oral prednisone 1 mg/kg per day for 2 months, followed by slow taper) plus azathioprine (2 mg/kg per day started on week 3).

Plasma cell–targeted therapy with anti–B-cell maturation antigen (BCMA) chimeric antigen receptor (CAR) T cells has shown promise in an early clinical study of patients with NMOSD who failed to respond to other treatments.

Importantly, some therapies with proven efficacy in MS do not appear to be useful for NMO. Available evidence suggests that interferon beta is ineffective and paradoxically may increase the risk of NMO relapses, and based on limited data, glatiramer acetate, fingolimod, natalizumab, and alemtuzumab also appear to be ineffective. These distinctions highlight the need for efficient diagnosis of NMO.

MYELIN OLIGODENDROCYTE GLYCOPROTEIN ANTIBODY–ASSOCIATED DISEASE

Although long considered to be a likely target for antibody-mediated demyelination, anti–myelin oligodendrocyte glycoprotein (MOG) antibodies detected by a cell-based assay that enables recognition of MOG epitopes in a lipid bilayer were only recently found to be associated with cases of acute disseminated encephalomyelitis (ADEM) in children and then with cases of AQP4-seronegative NMO. Further studies showed that patients who are seropositive for anti-MOG antibodies are at risk for bilateral, synchronous optic neuritis and myelitis and meningoencephalitis **(Table 456-3)**. A clinical feature that can help distinguish ON associated with MOG antibody–associated disease (MOGAD) from NMO or MS is the presence of papillitis seen by fundoscopy or orbital MRI, a finding that is common in MOGAD, rare in NMO, and variable but usually unilateral in MS. ON associated with MOGAD is typically longitudinally extensive on MRI, and brain MRI can be normal or show fluffy areas of increased signal change in white or gray matter structures, similar to NMO. MRI lesions that are typical for MS, including finger-like lesions oriented perpendicular to the ventricular surface (Dawson fingers) and T1 hypointense lesions, are uncommon. Spinal cord lesions can be longitudinally extensive or short and sometimes involve the conus medullaris. Demyelination associated with MOGAD is sometimes monophasic, as in ADEM, but can also be recurrent. The CSF may show a pleocytosis with occasional neutrophils. Elevated intrathecal synthesis of gamma globulins is atypical: oligoclonal bands are present in only ~6–13% of cases. A small percentage of patients who present with syndromes suggestive of MOGAD but test negative (or only weakly positive) for serum anti-MOG antibodies have detectable antibodies restricted to the CSF; thus, in these seronegative cases, CSF anti-MOG antibody testing should be pursued. The mechanism of central nervous system (CNS) injury in MOGAD is not established and could involve MOG-reactive cytotoxic T cells in addition to anti-MOG antibodies. Studies in MOG-induced experimental autoimmune encephalomyelitis, an animal model, suggest that anti-MOG antibodies may opsonize traces of MOG protein in secondary lymphoid tissues, triggering a peripheral immune response against MOG.

TABLE 456-3 Diagnostic Criteria for Myelin Oligodendrocyte Glycoprotein Antibody–Associated Disease (MOGAD)

Diagnostic Criteria for MOGAD
1. At least one core clinical event:
Optic neuritis
Myelitis
ADEM
Cerebral monofocal or polyfocal deficits
Brainstem or cerebellar deficits
Cerebral or cortical encephalitis often with seizures
2. Positive MOG IgG test (cell-based assay)
Clear positive – no additional requirements
Low positive, or positive without titer or CSF positive only – Must be AQP4 seronegative and have one or more of the following supporting features
Supporting clinical or radiographic features:
Optic neuritis – bilateral synchronous, >50% of length of optic nerve involvement, perineural optic nerve sheath enhancement, disc edema
Myelitis – longitudinally extensive, central cord involvement or "H sign," conus involvement
Brain/brainstem/cerebellar – ill-defined lesions, deep gray matter involvement, cortical lesions with or without overlying meningeal enhancement
3. Exclusion of better diagnoses including MS

Abbreviations: ADEM, acute disseminated encephalomyelitis; AQP4, aquaporin 4; CSF, cerebrospinal fluid; MOG, myelin oligodendrocyte glycoprotein; MS, multiple sclerosis.

Source: Adapted from B Banwell et al: Lancet Neurol 22:268, 2023.

Acute episodes are managed with high-dose glucocorticoids followed by a prednisone taper and sometimes by plasmapheresis, as with NMO. Brain lesions associated with MOGAD often respond rapidly to treatment with glucocorticoids and may resolve entirely. Some patients experience disease recurrence following discontinuation of prednisone and can become glucocorticoid dependent. Clinical trials in MOGAD are underway with satralizumab (the IL-6 receptor blocker indicated for AQP4-seropositive NMO) and rozanolixizumab (Rystiggo; a humanized monoclonal antibody that binds the neonatal Fc receptor). There are limited data on use of other immune-suppressing medications typically used in NMO. Off-label empiric treatments include daily prednisone, IV immunoglobulin, rituximab, and mycophenolate mofetil. Anti-MOG antibody titers appear to decline either spontaneously or in the setting of treatment.

GLIAL FIBRILLARY ACIDIC PROTEIN (GFAP) AUTOIMMUNITY

Autoimmunity against the astrocyte protein GFAP presents with a range of symptoms referable to meningismus, encephalitis, myelitis, and optic neuritis. Some cases follow a viral prodrome, and fever and headache are commonly present. The clinical syndrome can also include a movement disorder reflecting involvement of deep gray matter structures. MRI shows characteristic patterns of gadolinium enhancement localized to GFAP-enriched CNS regions including venous structures in a periventricular radial orientation, the leptomeninges, the periependymal spinal cord, and a striking serpiginous pattern involving brain parenchyma. Longitudinally extensive spinal cord involvement can also be present. These patterns share similarities with those observed in neurosarcoidosis, and their presence should prompt consideration for either condition. A lymphocytic pleocytosis is commonly present in the CSF. Antibodies against GFAP can be measured in the serum or CSF. GFAP autoimmunity is found as a paraneoplastic syndrome in ~25% of cases, most commonly associated with ovarian teratoma, and can coexist with anti-*N*-methyl-D-aspartate receptor (NMDAR) encephalitis or NMO. T cells are implicated in pathophysiology based on histopathology and association with checkpoint inhibitor treatment for cancer or in the setting of HIV. GFAP autoimmunity is generally glucocorticoid responsive. Early recognition with prompt intervention is associated with more favorable outcomes. Relapses occur in ~20% of patients and require use of immune suppression therapy.

ACUTE DISSEMINATED ENCEPHALOMYELITIS

ADEM has a monophasic course and is often associated with an antecedent infection (postinfectious encephalomyelitis); historically, a small number of ADEM cases have followed immunization (postvaccinal encephalomyelitis). ADEM is far more common in children than adults, and many adult cases initially thought to represent ADEM subsequently experience late relapses qualifying as either MS or another chronic inflammatory disorder such as vasculitis, sarcoidosis, or lymphoma. Many cases previously thought to represent ADEM are now recognized as MOGAD. The hallmark of ADEM is the presence of widely scattered foci of perivenular inflammation and demyelination that can involve both white matter and gray matter structures, in contrast to larger confluent white matter lesions typical of MS. In the most explosive form of ADEM, acute hemorrhagic leukoencephalitis, the lesions are vasculitic and hemorrhagic, and the clinical course is devastating.

Postinfectious encephalomyelitis is most frequently associated with the viral exanthems of childhood. Infection with measles virus is the most common antecedent (1 in 1000 cases). Worldwide, measles encephalomyelitis is still common, although use of the live measles vaccine has dramatically reduced its incidence. In developed countries, ADEM is now most frequently associated with varicella (chickenpox) infections (1 in 4000–10,000 cases). It may also follow infection with rubella, mumps, influenza, parainfluenza, Epstein-Barr virus, human herpesvirus-6, HIV, dengue, Zika, other viruses, and *Mycoplasma pneumoniae*. Cases have also been described in association with SARS-CoV-2 infection. Some patients may have a nonspecific upper respiratory infection or no known antecedent illness. Modern vaccines appear to pose no meaningful risk for ADEM; one large study (Vaccine Safety Datalink) of 24 different vaccines in >9 million individuals (64 million doses in total) revealed no excess risk for ADEM, with the possible exception of Tdap (tetanus, diphtheria, acellular pertussis) vaccine estimated at less than one case per million doses.

All forms of ADEM presumably result from a cross-reactive immune response to the infectious agent that then triggers an inflammatory demyelinating response. Autoantibodies to MBP and other myelin antigens have been detected in the CSF from some patients with ADEM, and as noted above, ADEM cases with serum or CSF antibodies against MOG are now considered to be MOGAD.

CLINICAL MANIFESTATIONS

In severe cases, onset is abrupt and progression rapid (hours to days). In postinfectious ADEM, the neurologic syndrome generally begins late in the course of the viral illness as the exanthem is fading. Fever reappears, and headache, meningismus, and lethargy progressing to coma may develop. Seizures are common. Signs of disseminated neurologic disease are consistently present (e.g., hemiparesis or quadriparesis, extensor plantar responses, lost or hyperactive tendon reflexes, sensory loss, and brainstem involvement). In ADEM due to chickenpox, cerebellar involvement is often conspicuous. CSF protein is modestly elevated (0.5–1.5 g/L [50–150 mg/dL]). Lymphocytic pleocytosis, generally ≥200 cells/μL, occurs in 80% of patients. Occasional patients have higher counts or a mixed polymorphonuclear-lymphocytic pattern during the initial days of the illness. Transient CSF oligoclonal banding was reported in a minority of cases. MRI usually reveals extensive changes in the brain and spinal cord, consisting of white matter hyperintensities on T2 and fluid-attenuated inversion recovery (FLAIR) sequences with gadolinium enhancement on T1-weighted sequences.

DIAGNOSIS

The diagnosis is most reliably established when there is a history of a recent infectious illness. In severe cases with predominantly cerebral involvement, acute encephalitis due to infection with herpes simplex or other viruses including HIV may be difficult to exclude; other considerations include hypercoagulable states including the antiphospholipid antibody syndrome, autoimmune (paraneoplastic) limbic encephalitis, vasculitis, sarcoidosis, primary CNS lymphoma, or metastatic cancer. An explosive presentation of MS can mimic ADEM, and especially in adults, it may not be possible to distinguish these conditions acutely. The simultaneous onset of disseminated symptoms and signs is common in ADEM and rare in MS. Similarly, meningismus, encephalopathy (drowsiness, stupor or coma), and seizures suggest ADEM rather than MS. Unlike MS, in ADEM, optic nerve involvement is generally bilateral and transverse myelopathy complete. MRI findings that favor ADEM include extensive and relatively symmetric white matter abnormalities, basal ganglia or cortical gray matter lesions, and gadolinium enhancement of all abnormal areas. In contrast, OCBs in the CSF are more common in MS. In one study of adult patients initially thought to have ADEM, 30% experienced additional relapses over a follow-up period of 3 years, and they were reclassified as having MS. Other patients initially classified as ADEM are subsequently found to have NMO, MOGAD, or GFAP autoimmunity. Occasional patients with “recurrent ADEM” have also been reported, especially children; however, it is not possible to distinguish this entity from atypical MS. Because of the clinical overlap at presentation between ADEM and MS, it is important that routine surveillance imaging be performed following recovery from ADEM so that subclinical disease activity due to MS can be recognized and treatment for MS initiated.

TREATMENT

Initial therapy is with high-dose glucocorticoids; depending on the response, treatment may need to be continued for 8 weeks. Patients who fail to respond within a few days may benefit from a course of plasma exchange or IV immunoglobulin. The prognosis reflects the severity of the underlying acute illness. In modern case series of presumptive ADEM in adults, mortality rates of 5–20% are reported, and many survivors have permanent neurologic sequelae.

FURTHER READING

Banwell B et al: Diagnosis of myelin oligodendrocyte glycoprotein antibody-associated disease: International MOGAD Panel proposed criteria. Lancet Neurol 22:268, 2023.

Baxter R et al: Acute demyelinating events following vaccines: A case-centered analysis. Clin Infect Dis 63:1456, 2016.

Cacciaguerra L et al: Updates in NMOSD and MOGAD diagnosis and treatment: A tale of two central nervous system autoimmune inflammatory disorders. Neurol Clin 42:77, 2024.

Cree BAC et al: Inebilizumab for the treatment of neuromyelitis optica spectrum disorder (N-MOmentum): A double-blind, randomised placebo-controlled phase 2/3 trial. Lancet 394:1352, 2019.

Hagbohm C et al: Clinical and neuroimaging phenotypes of autoimmune glial fibrillary acidic protein astrocytopathy: A systematic review and meta-analysis. Eur J Neurol 20:e16284, 2024.

Pittock SJ et al: Eculizumab in aquaporin-4-positive neuromyelitis optica spectrum disorder. N Engl J Med 381:614, 2019.

Qin C et al: Single-cell analysis of anti-BCMA CAR T cell therapy in patients with central nervous system autoimmunity. Sci Immunol 9:eadj9730, 2024.

Traboulsee A, et al. Safety and efficacy of satralizumab monotherapy in neuromyelitis optica spectrum disorder: A randomised, double-blind, multicentre, placebo-controlled phase 3 trial. Lancet Neurol 19:402, 2020.

Wingerchuk DM et al: International consensus diagnostic criteria for neuromyelitis optica spectrum disorders. Neurology 85:177, 2015.

Section 3 Nerve and Muscle Disorders

457

Peripheral Neuropathy

Anthony A. Amato, Richard J. Barohn

Peripheral nerves are composed of sensory, motor, and autonomic elements. Diseases can affect the cell body of a neuron or its peripheral processes, namely the axons or the encasing myelin sheaths. Most peripheral nerves are mixed and contain sensory and motor as well as autonomic fibers. Nerves can be subdivided into three major classes: large myelinated, small myelinated, and small unmyelinated. Motor axons are usually large myelinated fibers that conduct rapidly (~50 m/s). Sensory fibers may be any of the three types. Large-diameter sensory fibers conduct proprioception and vibratory sensation to the brain, while the smaller-diameter myelinated and unmyelinated fibers transmit pain and temperature sensation. Autonomic nerves are also small in diameter. Thus, peripheral neuropathies can impair sensory, motor, or autonomic function, either singly or in combination. Peripheral neuropathies are further classified into those that primarily affect the cell body (e.g., neuronopathy or ganglionopathy), myelin (myelinopathy), and the axon (axonopathy). These different classes of peripheral neuropathies have distinct clinical and electrophysiologic features. This chapter discusses the clinical approach to a patient suspected of having a peripheral neuropathy, as well as specific neuropathies, including hereditary and acquired neuropathies. **The inflammatory neuropathies are discussed in Chap. 458.**

GENERAL APPROACH

In approaching a patient with a neuropathy, the clinician has three main goals: (1) identify where the lesion is, (2) identify the cause, and (3) determine the proper treatment. The first goal is accomplished by obtaining a thorough history, neurologic examination, and electrodiagnostic and other laboratory studies **(Fig. 457-1)**. While gathering this information, seven key questions are asked **(Table 457-1)**, the answers to which help identify the pattern of involvement and the cause of the neuropathy **(Table 457-2)**. Despite an extensive evaluation, in approximately half of patients, no etiology is ever found; these patients typically have a predominately sensory polyneuropathy and have been labeled as having idiopathic or cryptogenic sensory and sensorimotor polyneuropathy (CSPN).

■ INFORMATION FROM THE HISTORY AND PHYSICAL EXAMINATION: SEVEN KEY QUESTIONS (TABLE 457-1)

1. What Systems Are Involved? It is important to determine if the patient's symptoms and signs are motor, sensory, autonomic, or a combination of these. If the patient has only weakness without any evidence of sensory or autonomic dysfunction, a motor neuropathy, neuromuscular junction abnormality, or myopathy should be considered. Some peripheral neuropathies are associated with significant autonomic nervous system dysfunction. Symptoms of autonomic involvement include fainting spells or orthostatic lightheadedness; heat intolerance; or any bowel, bladder, or sexual dysfunction **(Chap. 451)**. There will typically be an orthostatic fall in blood pressure without an appropriate increase in heart rate. Autonomic dysfunction in the absence of diabetes should alert the clinician to the possibility of amyloid polyneuropathy. Rarely, a pandysautonomic syndrome can be the only manifestation of a peripheral neuropathy without other motor or sensory findings. The majority of neuropathies are predominantly sensory in nature.

2. What Is the Distribution of Weakness? Delineating the pattern of weakness, if present, is essential for diagnosis, and in this regard, two additional questions should be answered: (1) Does the weakness only involve the distal extremity, or is it both proximal and distal? and (2) Is the weakness focal and asymmetric, or is it symmetric? Symmetric proximal and distal weakness is the hallmark of acquired immune demyelinating polyneuropathies, both the acute form (Guillain-Barré syndrome [GBS]) and the chronic form (chronic inflammatory demyelinating polyneuropathy [CIDP]) **(Chap. 458)**. The importance of finding symmetric proximal and distal weakness in a patient who presents with both motor and sensory symptoms cannot

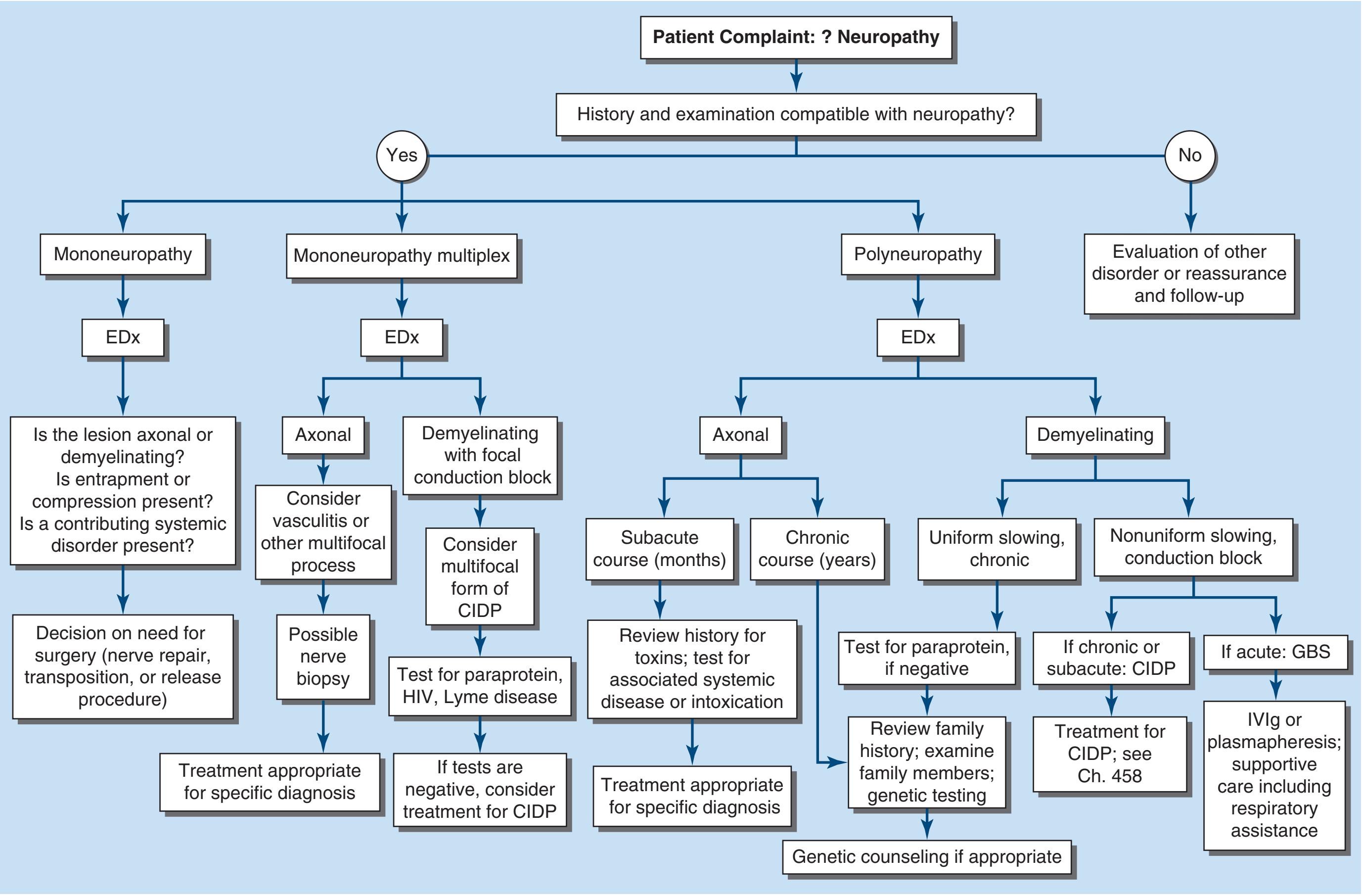
Patient Complaint: ? Neuropathy
History and examination compatible with neuropathy?
Yes
No
Mononeuropathy
Mononeuropathy multiplex
Polyneuropathy
Evaluation of other disorder or reassurance and follow-up
EDx
EDx
EDx
Is the lesion axonal or demyelinating? Is entrapment or compression present? Is a contributing systemic disorder present?
Axonal
Demyelinating with focal conduction block
Axonal
Demyelinating
Consider vasculitis or other multifocal process
Consider multifocal form of CIDP
Subacute course (months)
Chronic course (years)
Uniform slowing, chronic
Nonuniform slowing, conduction block
Decision on need for surgery (nerve repair, transposition, or release procedure)
Possible nerve biopsy
Test for paraprotein, HIV, Lyme disease
Review history for toxins; test for associated systemic disease or intoxication
Test for paraprotein, if negative
If chronic or subacute: CIDP
If acute: GBS
Treatment appropriate for specific diagnosis
If tests are negative, consider treatment for CIDP
Treatment appropriate for specific diagnosis
Review family history; examine family members; genetic testing
Treatment for CIDP; see Ch. 458
IVIg or plasmapheresis; supportive care including respiratory assistance
Genetic counseling if appropriate

FIGURE 457-1 Approach to the evaluation of peripheral neuropathies. CIDP, chronic inflammatory demyelinating polyradiculoneuropathy; EDx, electrodiagnostic; GBS, Guillain-Barré syndrome; IVIg, intravenous immunoglobulin.

TABLE 457-1 Approach to Neuropathic Disorders: Seven Key Questions

1. **What systems are involved?**
   - Motor, sensory, autonomic, or combinations
2. **What is the distribution of weakness?**
   - Only distal versus proximal and distal
   - Focal/asymmetric versus symmetric
3. **What is the nature of the sensory involvement?**
   - Temperature loss or burning or stabbing pain (e.g., small fiber)
   - Vibratory or proprioceptive loss (e.g., large fiber)
4. **Is there evidence of upper motor neuron involvement?**
   - Without sensory loss
   - With sensory loss
5. **What is the temporal evolution?**
   - Acute (days to 4 weeks)
   - Subacute (4–8 weeks)
   - Chronic (>8 weeks)
   - Monophasic, progressive, or relapsing-remitting
6. **Is there evidence for a hereditary neuropathy?**
   - Family history of neuropathy
   - Lack of sensory symptoms despite sensory signs
7. **Are there any associated medical conditions?**
   - Cancer, diabetes mellitus, connective tissue disease or other autoimmune diseases, infection (e.g., HIV, Lyme disease, leprosy)
   - Medications including over-the-counter drugs that may cause a toxic neuropathy
   - Preceding events, drugs, toxins

be overemphasized because this identifies the important subset of patients who may have a treatable acquired demyelinating neuropathic disorder (i.e., GBS or CIDP).

Findings of an asymmetric or multifocal pattern of weakness narrow the differential diagnosis. Some neuropathic disorders may present with unilateral extremity weakness. In the absence of sensory symptoms and signs, such weakness evolving over weeks or months would be worrisome for motor neuron disease (e.g., amyotrophic lateral sclerosis [ALS]), but it would be important to exclude multifocal motor neuropathy that may be treatable **(Chap. 458)**. In a patient presenting with asymmetric subacute or acute sensory and motor symptoms and signs, radiculopathies, plexopathies, compressive mononeuropathies, or multiple mononeuropathies (e.g., mononeuropathy multiplex) must be considered.

ALS **(Chap. 448)** can produce prominent neck extensor weakness (head drop), tongue and pharyngeal weakness (dysarthria and dysphagia), or shortness of breath. These focal symmetric weakness patterns can also be seen in neuromuscular junction disorders (myasthenia gravis, Lambert-Eaton myasthenic syndrome [LEMS] **[Chap. 459]**) and some myopathies, particularly isolated neck extensor myopathy **(Chap. 460)**.

3. What Is the Nature of the Sensory Involvement? The patient may have loss of sensation (numbness), altered sensation to touch (hyperpathia or allodynia), or uncomfortable spontaneous sensations (tingling, burning, or aching) **(Chap. 27)**. Neuropathic pain can be burning, dull, and poorly localized (protopathic pain), presumably transmitted by polymodal C nociceptor fibers, or sharp and lancinating (epicritic pain), relayed by A-delta fibers. If pain and temperature perception are lost, while vibratory and position sense are preserved along with muscle strength, deep tendon reflexes, and normal nerve conduction studies (NCS), a small-fiber neuropathy is likely. The most likely causes of small-fiber neuropathies, when one is identified, are diabetes mellitus (DM) or glucose intolerance. Amyloid neuropathy should be considered as well in such cases, but most of these small-fiber neuropathies remain idiopathic despite extensive evaluation.

Severe proprioceptive loss also narrows the differential diagnosis. Affected patients will note imbalance, especially in the

TABLE 457-2 Patterns of Neuropathic Disorders

Pattern 1: Symmetric proximal and distal weakness with sensory loss
- Consider: inflammatory demyelinating polyneuropathy (GBS and CIDP)

Pattern 2: Symmetric distal sensory loss with or without distal weakness
- Consider: cryptogenic or idiopathic sensory polyneuropathy (CSPN), diabetes mellitus and other metabolic disorders, drugs, toxins, familial (HSAN), CMT, amyloidosis, CANVAS, SORD neuropathy, and others

Pattern 3: Asymmetric distal weakness with sensory loss
- With involvement of multiple nerves
  - Consider: multifocal CIDP, vasculitis, cryoglobulinemia, amyloidosis, sarcoid, infectious (leprosy, Lyme, hepatitis B, C, or E, HIV, CMV), HNPP, tumor infiltration
- With involvement of single nerves/regions
  - Consider: may be any of the above but also could be compressive mononeuropathy, plexopathy, or radiculopathy

Pattern 4: Asymmetric proximal and distal weakness with sensory loss
- Consider: polyradiculopathy or plexopathy due to diabetes mellitus, meningeal carcinomatosis or lymphomatosis, sarcoid, amyloid, hereditary plexopathy (HNPP, HNA), idiopathic

Pattern 5: Asymmetric distal weakness without sensory loss
- With upper motor neuron findings
  - Consider: motor neuron disease
- Without upper motor neuron findings
  - Consider: progressive muscular atrophy, juvenile monomelic amyotrophy (Hirayama's disease), multifocal motor neuropathy, multifocal acquired motor axonopathy

Pattern 6: Symmetric sensory loss and distal areflexia with upper motor neuron findings
- Consider: vitamin B_{12}, vitamin E, and copper deficiency with combined system degeneration with peripheral neuropathy, chronic liver disease, hereditary leukodystrophies (e.g., adrenomyeloneuropathy), HSP-plus

Pattern 7: Symmetric weakness without sensory loss
- With proximal and distal weakness
  - Consider: SMA
- With distal weakness
  - Consider: hereditary motor neuropathy ("distal" SMA) or atypical CMT

Pattern 8: Focal midline proximal symmetric weakness
- Neck extensor weakness
  - Consider: ALS
- Bulbar weakness
  - Consider: ALS/PLS, isolated bulbar ALS (IBALS), Kennedy's syndrome (X-linked, bulbospinal SMA), bulbar presentation GBS
- Diaphragm weakness (SOB)
  - Consider: ALS

Pattern 9: Asymmetric proprioceptive sensory loss without weakness
- Consider causes of a sensory neuronopathy (ganglionopathy):
- Cancer (paraneoplastic)
- CANVAS
- Sjögren's syndrome
- Idiopathic sensory neuronopathy (possible GBS variant)
- Cisplatin and other chemotherapeutic agents
- Vitamin B_6 toxicity
- HIV-related sensory neuronopathy

Pattern 10: Autonomic symptoms and signs
- Consider neuropathies associated with prominent autonomic dysfunction:
- Hereditary sensory and autonomic neuropathy
- Amyloidosis (familial and acquired)
- Diabetes mellitus
- GBS
- Idiopathic pandysautonomia (may be a variant of GBS)
- Porphyria
- HIV-related autonomic neuropathy
- Vincristine and other chemotherapeutic agents

Abbreviations: ALS, amyotrophic lateral sclerosis; CIDP, chronic inflammatory demyelinating polyneuropathy; CANVAS, cerebellar ataxia, neuropathy, and vestibular areflexia syndrome; CMT, Charcot-Marie-Tooth disease; CMV, cytomegalovirus; GBS, Guillain-Barré syndrome; HIV, human immunodeficiency virus; HNA, hereditary neuralgic amyotrophy; HNPP, hereditary neuropathy with liability to pressure palsies; HSAN, hereditary sensory and autonomic neuropathy; HSP-plus, hereditary spastic paraplegia plus neuropathy; PLS, primary lateral sclerosis; SMA, spinal muscular atrophy; SOB, shortness of breath; SORD, sorbitol dehydrogenase deficiency.

dark. A neurologic examination revealing a dramatic loss of proprioception with vibration loss and normal strength should alert the clinician to consider a sensory neuronopathy/ganglionopathy (Pattern 9, Table 457-2). In particular, if this loss is asymmetric or affects the arms more than the legs, this pattern suggests a non-length-dependent process as seen in sensory neuronopathies.

4. Is There Evidence of Upper Motor Neuron Involvement? If the patient presents with symmetric distal sensory symptoms and signs suggestive of a distal sensory neuropathy, but there is additional evidence of symmetric upper motor neuron involvement **(Chap. 26)**, the physician should consider a combined system degeneration with neuropathy. The most common cause for this pattern is vitamin B_{12} deficiency, but other etiologies should also be considered (e.g., copper deficiency, human immunodeficiency virus [HIV] infection, severe hepatic disease, adrenomyeloneuropathy [AMN]), and hereditary spastic paraplegia plus a neuropathy.

5. What Is the Temporal Evolution? It is important to determine the onset, duration, and evolution of symptoms and signs. Does the disease have an acute (days to 4 weeks), subacute (4–8 weeks), or chronic (>8 weeks) course? Is the course monophasic, progressive, or relapsing? Most neuropathies are insidious and slowly progressive in nature. Neuropathies with acute and subacute presentations include GBS, vasculitis, and radiculopathies related to diabetes or Lyme disease. A relapsing course can be present in CIDP and porphyria.

6. Is There Evidence for a Hereditary Neuropathy? In patients with slowly progressive distal weakness over many years with few sensory symptoms yet significant sensory deficits on clinical examination, the clinician should consider a hereditary neuropathy (e.g., Charcot-Marie-Tooth disease [CMT]). On examination, the feet may show high or flat arches or hammer toes, and scoliosis may be present. In suspected cases, it may be necessary to perform neurologic and electrophysiologic studies on family members in addition to the patient.

7. Does the Patient Have Any Other Medical Conditions? It is important to inquire about associated medical conditions (e.g., DM, systemic lupus erythematosus [SLE]); preceding or concurrent infections (e.g. diarrheal illness preceding GBS); surgeries (e.g., gastric bypass and nutritional neuropathies); medications (toxic neuropathy), including over-the-counter vitamin preparations (B_6); alcohol; dietary habits; and use of dentures (e.g., fixatives contain zinc that can lead to copper deficiency).

■ PATTERN RECOGNITION APPROACH TO NEUROPATHIC DISORDERS

Based on the answers to the seven key questions, neuropathic disorders can be classified into several patterns based on the distribution or pattern of sensory, motor, and autonomic involvement (Table 457-2). Each pattern has a limited differential diagnosis, and information from laboratory studies usually permits a final diagnosis to be established.

■ ELECTRODIAGNOSTIC STUDIES

The electrodiagnostic (EDx) evaluation of patients with a suspected peripheral neuropathy consists of NCS and needle electromyography (EMG). In addition, studies of autonomic function can be valuable. The electrophysiologic data can confirm whether the neuropathic disorder is a mononeuropathy, multiple mononeuropathy (mononeuropathy multiplex), radiculopathy, plexopathy, or generalized polyneuropathy. Similarly, EDx evaluation can ascertain whether the process involves only sensory fibers, motor fibers, autonomic fibers, or a combination of these. Finally, the electrophysiologic data can help distinguish axonopathies from myelinopathies as well as axonal degeneration secondary to ganglionopathies from the more common length-dependent axonopathies.

NCS are most helpful in classifying a neuropathy as due to axonal degeneration or segmental demyelination **(Table 457-3)**. In general, low-amplitude potentials with relatively preserved distal latencies, conduction velocities, and late potentials, along with fibrillations on needle EMG, suggest an axonal neuropathy. On the other hand, slow conduction velocities, prolonged distal latencies and late potentials, relatively preserved amplitudes, and the absence of fibrillations on needle EMG imply a primary demyelinating neuropathy. The presence of nonuniform slowing of conduction velocity, conduction block, or temporal dispersion further suggests an acquired demyelinating neuropathy (e.g., GBS or CIDP) as opposed to a hereditary demyelinating neuropathy (e.g., CMT type 1).

TABLE 457-3 Electrophysiologic Features: Axonal Degeneration versus Segmental Demyelination

	AXONAL DEGENERATION	SEGMENTAL DEMYELINATION
Motor Nerve Conduction Studies		
CMAP amplitude	Decreased	Normal (except with CB or distal dispersion)
Distal latency	Normal	Prolonged
Conduction velocity	Normal	Slow
Conduction block	Absent	Present
Temporal dispersion	Absent	Present
F wave	Normal or absent	Prolonged or absent
H reflex	Normal or absent	Prolonged or absent
Sensory Nerve Conduction Studies		
SNAP amplitude	Decreased	Normal or decreased
Distal latency	Normal	Prolonged
Conduction velocity	Normal	Slow
Needle EMG		
Spontaneous activity		
Fibrillations	Present	Absent
Fasciculations	Present	Absent
Motor unit potentials		
Recruitment	Decreased	Decreased
Morphology	Long duration, large amplitude, polyphasic (if there is reinnervation)	Normal

Abbreviations: CB, conduction block; CMAP, compound motor action potential; EMG, electromyography; SNAP, sensory nerve action potential.

Autonomic studies are used to assess small myelinated (A-delta) or unmyelinated (C) nerve fiber involvement. Such testing includes heart rate response to deep breathing, heart rate and blood pressure response to both the Valsalva maneuver and tilt-table testing, and quantitative sudomotor axon reflex testing **(Chap. 451)**. These studies are particularly useful in patients who have pure small-fiber neuropathy or autonomic neuropathy in which routine NCS are normal.

■ OTHER IMPORTANT LABORATORY INFORMATION

In patients with generalized symmetric peripheral neuropathy, a standard laboratory evaluation should include a complete blood count, basic chemistries including serum electrolytes and tests of renal and hepatic function, fasting blood glucose (FBS), hemoglobin (Hb) A_{1c}, thyroid function tests, B_{12}, folate, erythrocyte sedimentation rate (ESR), rheumatoid factor, antinuclear antibodies (ANA), serum protein electrophoresis (SPEP) and immunoelectrophoresis or immunofixation, and free light chains in serum and urine. Quantification of the concentration of serum-free light chains and the kappa/lambda ratio is more sensitive than SPEP, immunoelectrophoresis, or immunofixation to detect a monoclonal gammopathy and therefore should be done if amyloidosis is suspected. A skeletal survey should be performed in patients with acquired demyelinating neuropathies and M-spikes to look for osteosclerotic or lytic lesions. Patients with monoclonal gammopathy should also be referred to a hematologist for consideration of a bone marrow biopsy. An oral glucose tolerance test is indicated in patients with painful sensory neuropathies even if FBS and HbA_{1c} are normal, as the test is abnormal in about one-third of such patients. In addition to the above tests, patients with a mononeuropathy multiplex pattern of involvement should have a vasculitis workup, including antineutrophil cytoplasmic antibodies (ANCAs), cryoglobulins, hepatitis serology, Western blot for Lyme disease, HIV, and occasionally a cytomegalovirus (CMV) titer.

There are many autoantibody panels (various antiganglioside antibodies) marketed for screening routine neuropathy patients for a treatable condition. These autoantibodies have no proven clinical utility or added benefit beyond the information obtained from a complete clinical examination and detailed EDx. A heavy metal screen is also not necessary as a screening procedure, unless there is a history of possible exposure or suggestive features on examination (e.g., severe painful sensorimotor and autonomic neuropathy and alopecia—thallium; severe painful sensorimotor neuropathy with or without gastrointestinal [GI] disturbance and Mee's lines—arsenic; wrist or finger extensor weakness and anemia with basophilic stippling of red blood cells—lead).

In patients with suspected GBS or CIDP, a lumbar puncture is indicated to look for an elevated cerebrospinal fluid (CSF) protein. In idiopathic cases of GBS and CIDP, CSF pleocytosis is usually absent. If cells are present, one should consider HIV infection, Lyme disease, sarcoidosis, or lymphomatous or leukemic infiltration of nerve roots. Recently, serum IgG4 antibodies to neurofascin and contactin-2 have been discovered in CIDP with severe sensory ataxia, tremor, and distal weakness **(Chap. 458)**. These cases are difficult to treat with standard immunotherapies but may respond to rituximab. Some patients with GBS and CIDP have abnormal liver function tests. In these cases, it is important to also check for hepatitis B and C, HIV, CMV, and Epstein-Barr virus (EBV) infection. In patients with an axonal GBS (by EMG/NCS) or those with a suspicious coinciding history (e.g., unexplained abdominal pain, psychiatric illness, significant autonomic dysfunction), it is reasonable to screen for porphyria.

In patients with a severe sensory ataxia, a sensory ganglionopathy or neuronopathy should be considered. The most common causes of sensory ganglionopathies are Sjögren's syndrome **(Chap. 373)** and a paraneoplastic neuropathy **(Chap. 99)**. Neuropathy can be the initial manifestation of Sjögren's syndrome. Thus, one should always inquire about dry eyes and mouth in patients with sensory signs and symptoms. Further, some patients can manifest sicca complex without other manifestations of Sjögren's syndrome. Thus, patients with sensory ataxia should be tested for antibodies to SS-A/Ro and SS-B/La, in addition to the routine ANA. To evaluate a possible paraneoplastic sensory ganglionopathy, antineuronal nuclear antibodies (e.g., anti-Hu antibodies) should be obtained. These antibodies are most commonly seen in patients with small-cell carcinoma of the lung but are also present with breast, ovarian, lymphoma, and other cancers. Importantly, the paraneoplastic neuropathy can precede the detection of the cancer, and detection of these autoantibodies should lead to a search for malignancy.

■ NERVE BIOPSIES

Nerve biopsies are now rarely performed in the evaluation of neuropathies. The primary indication for nerve biopsy is suspicion for amyloid neuropathy or vasculitis. In most instances, the abnormalities present on biopsies do not help distinguish one form of peripheral neuropathy from another (beyond what is already apparent by clinical examination and the NCS). Nerve biopsies should only be performed when the NCS are abnormal. The sural nerve is most commonly biopsied because it is a pure sensory nerve and biopsy will not result in loss of motor function. In suspected vasculitis, a combination biopsy of a superficial peroneal nerve (pure sensory) and the underlying peroneus brevis muscle obtained from a single small incision increases the diagnostic yield. Tissue can be analyzed to assess for evidence of inflammation, vasculitis, or amyloid deposition. Semithin plastic sections, teased fiber preparations, and electron microscopy are used to assess the morphology of the nerve fibers and to distinguish axonopathies from myelinopathies.

■ SKIN BIOPSIES

Skin biopsies are sometimes used to diagnose a small-fiber neuropathy. Following a punch biopsy of the skin in the distal lower extremity, immunologic staining can be used to measure the density of small unmyelinated fibers. The density of these nerve fibers is reduced in patients with small-fiber neuropathies in whom NCS and routine nerve biopsies are often normal. This technique may allow for an objective measurement in patients with mainly subjective symptoms. However, it often adds little to what one already knows from the clinical examination and EDx.

SPECIFIC DISORDERS

■ HEREDITARY NEUROPATHIES

CMT disease is the most common type of hereditary neuropathy (Pattern 2, Table 457-2). Rather than one disease, CMT is a syndrome of many genetically distinct disorders **(Table 457-4)**. The various subtypes of CMT are classified according to the nerve conduction velocities (NCVs) and predominant pathology (e.g., demyelination or axonal degeneration), inheritance pattern (autosomal dominant, autosomal recessive, or X-linked), and the specific mutated genes. Type 1 CMT (or CMT1) refers to inherited demyelinating sensorimotor neuropathies, whereas the axonal sensory neuropathies are classified as CMT2. By definition, motor conduction velocities in the arms are slowed to <38 m/s in CMT1 and are >38 m/s in CMT2. However, most cases of CMT1 actually have motor NCVs between 20 and 25 m/s. CMT1 and CMT2 usually begin in childhood or early adult life; however, onset later in life can occur, particularly in CMT2. Both are inherited in an autosomal dominant fashion, with a few exceptions. There are no medical therapies for any of the CMTs, but physical and occupational therapy can be beneficial, as can bracing (e.g., ankle-foot orthotics for foot drop) and other orthotic devices.

■ CMT1

CMT1 is the most common form of hereditary neuropathy. Affected individuals usually present in the first to third decade of life with distal leg weakness (e.g., foot drop), although patients may remain asymptomatic even late in life. People with CMT generally do not complain of numbness or tingling, which can be helpful in distinguishing CMT from acquired forms of neuropathy in which sensory symptoms usually predominate. Although usually asymptomatic, reduced sensation to all modalities is apparent on examination. Muscle stretch reflexes are unobtainable or reduced throughout. There is often atrophy of the muscles below the knee (particularly the anterior compartment), leading to so-called inverted champagne bottle legs.

Motor NCVs are generally in the 20–25 m/s range. Nerve biopsies usually are not performed on patients suspected of having CMT1, because the diagnosis usually can be made by less invasive testing (e.g., NCS and genetic studies). However, when done, the biopsies reveal reduced numbers of myelinated nerve fibers with a predilection for loss of large-diameter fibers and Schwann cell proliferation around thinly or demyelinated fibers, forming so-called onion bulbs.

CMT1A is the most common subtype of CMT1, representing 70% of cases, and is caused by a 1.5-megabase (Mb) duplication within chromosome 17p11.2-12 encoding the gene for peripheral myelin protein-22 (PMP-22). This results in patients having three copies of the PMP-22 gene rather than two. This protein accounts for 2–5% of myelin protein and is expressed in compact regions of the peripheral myelin sheath. Approximately 20% of patients with CMT1 have CMT1B, caused by mutations in the myelin protein zero (MPZ). CMT1B is for the most part clinically, electrophysiologically, and histologically indistinguishable from CMT1A. MPZ is an integral myelin protein and accounts for more than half of the myelin protein in peripheral nerves. Other forms of CMT1 are much less common and also indistinguishable from one another clinically and electrophysiologically (Table 457-4).

■ CMT2

CMT2 occurs approximately half as frequently as CMT1, and CMT2 tends to present later in life. Affected individuals usually become symptomatic in the second decade; some cases present earlier in childhood, whereas others remain asymptomatic into late adult life. Clinically, CMT2 is for the most part indistinguishable from CMT1. NCS are helpful in this regard; in contrast to CMT1, the velocities are normal or

TABLE 457-4 Classification of Charcot-Marie-Tooth Disease and Related Neuropathies

NAME	INHERITANCE	GENE LOCATION	GENE
CMT1			
CMT1A	AD	17p11.2	*PMP22* (usually duplication of gene)
CMT1B	AD	1q21-23	*MPZ*
CMT1C	AD	16p13.1-p12.3	*LITAF*
CMT1D	AD	10q21.1-22.1	*ERG2*
CMT1E (with deafness)	AD	17p11.2	*PMP22* gene (usually point mutations)
CMT1F	AD	8p13-21	*NEFL*
CMT1G	AD	8q21	*PMP22*
HNPP	AD	17p11.2	*PMP22* (deletion of gene)
		1q21-23	*MPZ*
CMT dominant-intermediate (CMTDI)			
CMT-DIA	AD	10q24.1-25.1	?
CMT-DIB	AD	19.p12-13.2	*DNM2*
CMT-DIC	AD	1p35	*YARS*
CMT-DID	AD	1q22	*MPZ*
CMT-DIE	AD	14q32.33	*IFN-2*
CMT-DIF	AD	3q26	*GNB4*
CMT-DIG	AD	8p31	*NEFL*
CMT recessive-intermediate (CMT-RI)			
CMT-RIA	AR	8q21.1	*GDAP1*
CMT-RIB	AR	6q23	*KARS5*
CMT-RIC	AR	1p36	*PLEKHG5*
CMT-RI D	AR	12q24	*COX6A1*
CMT2			
CMT2A2 (allelic to HMSN VI with optic atrophy)	AD	1p36.2	*MFN2*
CMT2B	AD	3q13-q22	*RAB7*
CMT2B1 (allelic to LGMD 1B)	AR	1q21.2	*LMNA*
CMT2B2	AR	19q13	*PNKP*
CMT2C (allelic to scapuloperoneal neuropathy)	AD	12q23-24	*TRPV4*
CMT2D (allelic to distal SMA5)	AD	7p14	*GARS1*
CMT2DD	AD	1p13	*ATP1A1*
CMT2E (allelic to CMT1F)	AD	8p21	NEFL
CMT2EE	AD	2p23	*MPV17*
CMT2F	AD	7q11-q21	*HSPB1*
CMT2G (allelic to CMT2P)	AD	9q31.3-34.2	*LRSAM1*
CMT2I (allelic to CMT1B)	AD	1q22	*MPZ*
CMT2J	AD	1q22	*MPZ*
CMT2H, CMT2K (allelic to CMT4A)	AD	8q13-q21	*GDAP1*
CMT2L (allelic to distal hereditary motor neuropathy type 2)	AD	12q24	*HSPB8*
CMT2M	AD	16q22	*DNM2*
CMT2N	AD	16q22.1	*AARS*
CMT2O	AD	14q32.31	*DYNC1H1*
CMT2P	AD and AR	9q31.3-34.2	*LRSAM1*
CMT2P-Okinawa (allelic to HSMN2P)	AD	3q13-q14	*TFG*
CMT2Q	AD	10p14	*DHTKD1*
CMT2RCMT2S	AD	4q	*TRIM2*
CMT2T	AD	11q13.3	*IGHMBP2*
CMT2U	ADAD	3q25.2	*MME*
CMT2V	AD	12q13	*MARS1*
CMT2W	AD	17q11	*NAGLU*
CMT2X	AD	5q31	*HARS1*
CMT2Y	AD	15q21.1	*SPB11*
CMT2Z	AD	9p13	*VCP*
		22q12	*MORC2*

(*Continued*)

TABLE 457-4 Classification of Charcot-Marie-Tooth Disease and Related Neuropathies (Continued)

NAME	INHERITANCE	GENE LOCATION	GENE
CMT3	AD	17p11.2	*PMP22*
(Dejerine-Sottas disease, congenital hypomyelinating neuropathy)	AD	1q21-23	*MPZ*
	AR	10q21.1-22.1	*ERG2*
	AR	19q13	*PRX*
CMT4			
CMT4A	AR	8q13-21.1	*GDAP1*
CMT4B1	AR	11q23	*MTMR2*
CMT4B2	AR	11p15	*SBF2*
CMT4B3	AR	22q13.33	*SBF1*
CMT4C	AR	5q23-33	*SH3TC2*
CMT4D (HMSN-Lom)	AR	8q24	*NDRG1*
CMT4E (congenital hypomyelinating neuropathy)	AR	10q21.3	*ERG2*
CMT4F	AR	19q13.1-13.3	*PRX*
CMT4G	AR	10q23.2	*HK1*
CMT4H	AR	12p11.21	*FGD4*
CMT4J	AR	6q21	*FIG4*
CMT4K	AR	9q34	*SURF1*
CMTX (X-linked)			
CMTX1	X-linked dominant	Xq13	*GJB1*
CMTX4	X-linked recessive	Xq26.1	*AIFM1*
CMTX5	X-linked recessive	Xq22.3	*PRPS1*
CMTX6	X-linked dominant	Xp22.11	*PDK3*
HSAN1A	AD	9q22	*SPTLC1*
HSAN1C	AD	14q24.3	*SPTLC2*
HSAN1D	AD	14q21.3	*ATL1*
HSAN1E	AD	19p13.2	*DNMT1*
HSAN1F	AD	11q13.1	*ATL3*
HSAN2A	AR	12p13.33	*WNK1*
HSAN2B	AR	5p15.1	*RETREG1* (*FAM134B*)
HSAN2C	AR	12q13.13	*KIF1A*
HSAN2D	AR	2q24.3	*SCN9A*
HSAN3A (Riley-Day syndrome; hereditary dysautonomia)	AR	9q21	*ELP1* (*IKBKAP*)
HSAN4	AR	3q	*NTRK1*
HSAN5	AR	1p13.2	*NGF*
HSAN6	AR	6p12.1	*DST*
HSAN7	AD	3p22.2	*SCN11A*
HSAN8	AR	9q34.12	*PRDM12*
HSAN9	AR	14q32.31	*TECPR2*
Others			
HNA	AD	17q24	*SEPT9*
SORD neuropathy (allelic to distal HMN8)	AR	15q21.1	*SORD*
Hereditary neuropathy with neuromyotonia	AR	5q23.3	*HINT1*
CANVAS	AR	4p14	*RFC1*

Abbreviations: AARS, alanyl-tRNA synthetase; AD, autosomal dominant; AR, autosomal recessive; ATL, atlastin; CANVAS, cerebellar ataxia, neuropathy, and vestibular areflexia syndrome; CMT, Charcot-Marie-Tooth; DNMT1, DNA methyltransferase 1; DYS, dystonin; DYNC1HI, cytoplasmic dynein 1 heavy chain 1; ELP1, elongator complex protein 1; ERG2, early growth response-2 protein; FAM134B, family with sequence similarity 134, member B; FIG4, FDG1-related F actin-binding protein; GDAP1, ganglioside-induced differentiation-associated protein-1; HK1, hexokinase 1; HMSN-P, hereditary motor and sensory neuropathy proximal; HNA, hereditary neuralgic amyotrophy; HNPP, hereditary neuropathy with liability to pressure palsies; HSAN, hereditary sensory and autonomic neuropathy; IFN2, inverted formin-2; IKBKAP, $_k$B kinase complex-associated protein; LGMD, limb girdle muscular dystrophy; LITAF, lipopolysaccharide-induced tumor necrosis factor α factor; LRSAM1, E3 ubiquitin-protein ligase; MED25, mediator 25; MFN2, mitochondrial fusion protein mitofusin 2 gene; MPZ, myelin protein zero protein; MTMR2, myotubularin-related protein-2; NDRG1, N-myc downstream regulated 1; NGF, Beta-nerve growth factor; NTRK, 1trkA/NGF receptor; PMP-22, peripheral myelin protein-22; PRKWNK1, protein kinase, lysine deficient 1; PRPS1, phosphoribosylpyrophosphate synthetase 1; RAB7, Ras-related protein 7; RFC1, replication factor C subunit 1; SEPT9, septin 9; SH3TC2, SH3 domain and tetratricopeptide repeats 2; SMA, spinal muscular atrophy; SORD, sorbitol dehydrogenase; SPTLC, serine palmitoyltransferase long-chain base; TFG, TRK-fused gene; TrkA/NGF, tyrosine kinase A/nerve growth factor; tRNA, transfer ribonucleic acid; TRPV4, transient receptor potential cation channel, subfamily V, member 4; WNK1, WNK lysine deficient; YARS, tyrosyl-tRNA synthetase.

Source: Modified from AA Amato, J Russell: *Neuromuscular Disorders,* 2nd ed. New York, McGraw-Hill, 2016, Table 11-1, pp. 265–266.

only slightly slowed. The most common cause of CMT2 is a mutation in the gene for mitofusin 2 (MFN2), which accounts for ~20–30% of CMT2 cases overall. MFN2 localizes to the outer mitochondrial membrane, where it regulates the mitochondrial network architecture by participating in mitochondrial fusion. The other genes associated with CMT2 are much less common (Table 457-4).

CMT DOMINANT AND RECESSIVE INTERMEDIATE

In CMT dominant-intermediate (CMT-DI) and CMT recessive-intermediate (CMT-RI), the NCVs are faster than usually seen in CMT1 (e.g., >38 m/s) but slower than in CMT2 (Table 457-4).

CMT3

CMT3 was originally described by Dejerine and Sottas as a hereditary demyelinating sensorimotor polyneuropathy presenting in infancy or early childhood. Affected children are severely weak. Motor NCVs are markedly slowed, typically ≤5–10 m/s. Most cases of CMT3 are caused by point mutations in the genes for PMP-22, MPZ, or ERG-2, which are also the genes responsible for CMT1. The term *CMT3* is no longer recommended, but rather, the neuropathy is classified as CMT1 if autosomal dominant or as CMT4 in cases of autosomal recessive inheritance.

CMT4

CMT4 is extremely rare and is characterized by a severe, childhood-onset sensorimotor polyneuropathy that is usually inherited in an autosomal recessive fashion. Electrophysiologic and histologic evaluations can show demyelinating or axonal features. CMT4 is genetically heterogeneous (Table 457-4).

CMTX

There are several forms of X-linked CMT, the most common type is CMTX1 (Table 457-4). This shares clinical features similar to CMT1 and CMT2, except that the neuropathy is much more severe in males than in females. CMT1X accounts for ~10–15% of CMT overall. Males usually present in the first two decades of life with atrophy and weakness of the distal arms and legs, areflexia, pes cavus, and hammer toes. Obligate female carriers are frequently asymptomatic but can develop signs and symptoms of CMT. Onset in females is usually after the second decade of life, and the neuropathy is milder in severity.

NCS reveal features of both demyelination and axonal degeneration. In males, motor NCVs in the arms and legs are moderately slowed (in the low to mid 30-m/s range). About 50% of males with CMTX1 have motor NCVs between 15 and 35 m/s with ~80% of these falling between 25 and 35 m/s (intermediate slowing). In contrast, ~80% of females with CMTX1 have NCVs in the normal range and 20% have NCVs in the intermediate range. CMT1X is caused by mutations in *GJB1*, which encodes for gap junction protein-beta or connexin-32. Connexins are gap junction structural proteins that are important in cell-to-cell communication.

Hereditary Neuropathy with Liability to Pressure Palsies (HNPP) HNPP is an autosomal dominant disorder related to CMT1A. While CMT1A is usually associated with a 1.5-Mb duplication in chromosome 17p11.2 that results in an extra copy of the PMP-22 gene, HNPP is caused by inheritance of the chromosome with the corresponding 1.5-Mb deletion of this segment, and thus, affected individuals have only one copy of the PMP-22 gene. Patients usually manifest in the second or third decade of life with painless numbness and weakness in the distribution of single peripheral nerves, although multiple mononeuropathies can occur (Pattern 3, Table 457-2). Symptomatic mononeuropathy or multiple mononeuropathies are often precipitated by trivial compression of nerve(s) as can occur with wearing a backpack, leaning on the elbows, or crossing one's legs for even a short period of time. These pressure-related mononeuropathies may take weeks or months to resolve. In addition, some affected individuals manifest with a progressive or relapsing, generalized and symmetric, sensorimotor peripheral neuropathy that resembles CMT.

Hereditary Neuralgic Amyotrophy (HNA) HNA is an autosomal dominant disorder characterized by recurrent attacks of pain, weakness, and sensory loss in the distribution of the brachial plexus often beginning in childhood (Pattern 4, Table 457-2). These attacks are similar to those seen with idiopathic brachial plexitis (see below). Attacks may occur in the postpartum period, following surgery, or at other times of stress. Most patients recover over several weeks or months. Slightly dysmorphic features, including hypotelorism, epicanthal folds, cleft palate, syndactyly, micrognathia, and facial asymmetry, are evident in some individuals. EDx demonstrate an axonal process. HNA is genetically heterogeneous but can be caused by mutations in septin 9 (*SEPT9*). Septins may be important in formation of the neuronal cytoskeleton and have a role in cell division, but it is not known how mutations in *SEPT9* lead to HNA.

Hereditary Sensory and Autonomic Neuropathy (HSAN) The HSANs are a very rare group of hereditary neuropathies in which sensory and autonomic dysfunction predominates over muscle weakness, unlike CMT, in which motor findings are most prominent (Pattern 2, Table 457-2; Table 457-4). Nevertheless, affected individuals can develop motor weakness, and there can be overlap with CMT. There are no medical therapies available to treat these neuropathies, other than prevention and treatment of mutilating skin and bone lesions.

Of the HSANs, only HSAN1 typically presents in adults. HSAN1 is the most common of the HSANs and is inherited in an autosomal dominant fashion. Affected individuals usually manifest in the second through fourth decades of life. HSAN1 is associated with the degeneration of small myelinated and unmyelinated nerve fibers leading to severe loss of pain and temperature sensation, deep dermal ulcerations, recurrent osteomyelitis, Charcot joints, bone loss, gross foot and hand deformities, and amputated digits. Although most people with HSAN1 do not complain of numbness, they often describe burning, aching, or lancinating pains. Autonomic neuropathy is not a prominent feature, but bladder dysfunction and reduced sweating in the feet may occur. HSAN1A, which is most common, is caused by mutations in the serine palmitoyltransferase long-chain base 1 (*SPTLC1*) gene.

OTHER HEREDITARY NEUROPATHIES
(TABLE 457-5)

SORBITAL DEHYDROGENASE DEFICIENCY WITH PERIPHERAL NEUROPATHY

Sorbitol dehydrogenase deficiency with peripheral neuropathy (SORD) is a newly reported entity that is very important as it appears to be the most common autosomal recessive inherited form of neuropathy. It presents as a slowly progressive, length-dependent, axonal motor greater than sensory polyneuropathy or pure motor neuropathy. Age of onset is usually in the late teens. It is caused by pathogenic mutations in the *SORD* gene. SORD is the second enzyme of the two-step polyol pathway whereby glucose is metabolized into sorbitol, and then SORD oxidizes sorbitol into fructose. Sorbitol is a relatively nonmetabolizable sugar, and levels are markedly increased.

FABRY'S DISEASE

Fabry's disease (angiokeratoma corporis diffusum) is an X-linked dominant disorder. Although men are more commonly and severely affected, women can also manifest symptoms and signs of the disease. Angiokeratomas are reddish-purple maculopapular lesions that are usually found around the umbilicus, scrotum, inguinal region, and perineum. Burning or lancinating pain in the hands and feet often develops in males in late childhood or early adult life (Pattern 2, Table 457-2). However, the neuropathy is usually overshadowed by complications arising from an associated premature atherosclerosis (e.g., hypertension, renal failure, cardiac disease, and stroke) that often lead to death by the fifth decade of life. Some patients also manifest primarily with a dilated cardiomyopathy.

Fabry's disease is caused by mutations in the α-galactosidase gene that lead to the accumulation of ceramide trihexoside in nerves and

TABLE 457-5 Rare Hereditary Neuropathies

Hereditary Disorders of Lipid Metabolism
Metachromatic leukodystrophy
Krabbe's disease (globoid cell leukodystrophy)
Fabry's disease
Adrenoleukodystrophy/adrenomyeloneuropathy
Refsum's disease
Tangier disease
Cerebrotendinous xanthomatosis
Hereditary Ataxias with Neuropathy
CANVAS (cerebellar ataxia, neuropathy, and vestibular areflexia syndrome)
Friedreich's ataxia
Vitamin E deficiency
Spinocerebellar ataxia
Abetalipoproteinemia (Bassen-Kornzweig disease)
Disorders of Defective DNA Repair
CANVAS
Ataxia-telangiectasia
Cockayne's syndrome
Giant Axonal Neuropathy
Porphyria
Acute intermittent porphyria (AIP)
Hereditary coproporphyria (HCP)
Variegate porphyria (VP)
Familial Amyloid Polyneuropathy (FAP)
Transthyretin-related
Gelsolin-related
Apolipoprotein A1-related

Source: Modified from AA Amato, J Russell: *Neuromuscular Disorders,* 2nd ed. New York, McGraw-Hill, 2016, Table 12-1, p. 299.

blood vessels. A decrease in α-galactosidase activity is evident in leukocytes and cultured fibroblasts. Glycolipid granules may be appreciated in ganglion cells of the peripheral and sympathetic nervous systems and in perineurial cells. Enzyme replacement therapy (ERT) may improve the neuropathy if patients are treated early, before irreversible nerve fiber loss develops. Current U.S. Food and Drug Administration (FDA)-approved recombinant ERTs are agalsidase-α (Replagal; 0.2 mg/kg body weight) and agalsidase-β (Fabrazyme; 1 mg/kg body weight) and pegunigalsidase (Elfabrio), which are each given intravenously every 2 weeks. In addition, migalastat is an oral pharmacologic chaperone that increases the enzyme activity of "amenable" mutations (defined as those mutations in the catalytic domain of the enzyme that lead to misfolding of the enzyme but otherwise would not significantly impair its function). Such mutations occur in ~50% of patients. Migalastat had been shown to reduce left ventricular mass and stabilize kidney function, but studies have not assessed if the neuropathy improves or stabilizes.

■ ADRENOLEUKODYSTROPHY/ ADRENOMYELONEUROPATHY

Adrenoleukodystrophy (ALD) and AMN are allelic X-linked dominant disorders caused by mutations in the peroxisomal transmembrane adenosine triphosphate-binding cassette (ABC) transporter gene. Patients with ALD manifest with central nervous system (CNS) abnormalities. However, ~30% of patients with mutations in this gene present with the AMN phenotype that typically manifests in the third to fifth decade of life as mild to moderate peripheral neuropathy combined with progressive spastic paraplegia (Pattern 6, Table 457-2) **(Chap. 453)**. Rare patients present with an adult-onset spinocerebellar ataxia or only with adrenal insufficiency.

EDx is suggestive of a primary axonopathy with secondary demyelination. Nerve biopsies demonstrate a loss of myelinated and unmyelinated nerve fibers with lamellar inclusions in the cytoplasm of Schwann cells. Very-long-chain fatty acid (VLCFA) levels (C24, C25, and C26) are increased in the urine. Laboratory evidence of adrenal insufficiency is evident in approximately two-thirds of patients. The diagnosis can be confirmed by genetic testing.

Adrenal insufficiency is managed by replacement therapy; however, there is no proven effective therapy for the neurologic manifestations of ALD/AMN. Diets low in VLCFAs and supplemented with Lorenzo's oil (erucic and oleic acids) reduce the levels of VLCFAs and increase the levels of C22 in serum, fibroblasts, and liver; however, several large, open-label trials of Lorenzo's oil failed to demonstrate efficacy. Although allogeneic bone marrow transplantation and gene therapy have been successful in slowing progression of cognitive decline in some patients with ALD treated early in their disease, these approaches are ineffective for the myelopathy or neuropathy.

■ REFSUM'S DISEASE

Refsum's disease can manifest in infancy to early adulthood with the classic tetrad of (1) peripheral neuropathy, (2) retinitis pigmentosa, (3) cerebellar ataxia, and (4) elevated CSF protein concentration. Most affected individuals develop progressive distal sensory loss and weakness in the legs leading to foot drop by their twenties (Pattern 2, Table 457-2). Subsequently, the proximal leg and arm muscles may become weak. Patients may also develop sensorineural hearing loss, cardiac conduction abnormalities, ichthyosis, and anosmia.

Serum phytanic acid levels are elevated. Sensory and motor NCS reveal reduced amplitudes, prolonged latencies, and slowed conduction velocities. Nerve biopsy demonstrates a loss of myelinated nerve fibers, with remaining axons often thinly myelinated and associated with onion bulb formation.

Refsum's disease is genetically heterogeneous but autosomal recessive in nature. Classical Refsum's disease with childhood or early adult onset is caused by mutations in the gene that encodes for phytanoyl-CoA α-hydroxylase (*PAHX*). Less commonly, mutations in the gene encoding peroxin 7 receptor protein (PRX7) are responsible. These mutations lead to the accumulation of phytanic acid in the central and peripheral nervous systems. Treatment is removal of phytanic precursors (phytols: fish oils, dairy products, and ruminant fats) from the diet.

■ TANGIER DISEASE

Tangier disease is a rare autosomal recessive disorder that can present as (1) asymmetric multiple mononeuropathies, (2) a slowly progressive symmetric polyneuropathy predominantly in the legs, or (3) a pseudo-syringomyelia pattern with dissociated sensory loss (i.e., abnormal pain/temperature perception but preserved position/vibration in the arms **[Chap. 453]**). The tonsils may appear swollen and yellowish-orange in color, and there may also be splenomegaly and lymphadenopathy.

Tangier disease is caused by mutations in the ATP-binding cassette transporter 1 (*ABC1*) gene, which leads to markedly reduced levels of high-density lipoprotein (HDL) cholesterol levels, whereas triacylglycerol levels are increased. Nerve biopsies reveal axonal degeneration with demyelination and remyelination. Electron microscopy demonstrates abnormal accumulation of lipid in Schwann cells, particularly those encompassing unmyelinated and small myelinated nerves. There is no specific treatment.

■ PORPHYRIA

Porphyria is a group of inherited disorders caused by defects in heme biosynthesis **(Chap. 428)**. Three forms of porphyria are associated with peripheral neuropathy: acute intermittent porphyria (AIP), hereditary coproporphyria (HCP), and variegate porphyria (VP). The acute neurologic manifestations are similar in each, with the exception that a photosensitive rash is seen with HCP and VP but not in AIP. Attacks of porphyria can be precipitated by certain drugs (usually those metabolized by the P450 system), hormonal changes (e.g., pregnancy, menstrual cycle), and dietary restrictions.

An acute attack of porphyria may begin with sharp abdominal pain. Subsequently, patients may develop agitation, hallucinations, or seizures. Several days later, back and extremity pain followed by weakness ensues, mimicking GBS (Pattern 1, Table 457-2). Weakness can involve the arms or the legs and can be asymmetric, proximal, or distal in distribution, as well as affecting the face and bulbar musculature. Dysautonomia and signs of sympathetic overactivity are common (e.g., pupillary dilation, tachycardia, and hypertension). Constipation, urinary retention, and incontinence can also be seen.

The CSF protein is typically normal or mildly elevated. Liver function tests and hematologic parameters are usually normal. Some patients are hyponatremic due to inappropriate secretion of antidiuretic hormone **(Chap. 390)**. The urine may appear brownish in color secondary to the high concentration of porphyrin metabolites. Accumulation of intermediary precursors of heme (i.e., d-aminolevulinic acid, porphobilinogen, uroporphobilinogen, coproporphyrinogen, and protoporphyrinogen) is found in urine. Specific enzyme activities can also be measured in erythrocytes and leukocytes. The primary abnormalities on EDx are marked reductions in compound motor action potential (CMAP) amplitudes and signs of active axonal degeneration on needle EMG.

The porphyrias are inherited in an autosomal dominant fashion. AIP is associated with porphobilinogen deaminase deficiency, HCP is caused by defects in coproporphyrin oxidase, and VP is associated with protoporphyrinogen oxidase deficiency. The pathogenesis of the neuropathy is not completely understood. Treatment with glucose and hematin may reduce the accumulation of heme precursors. Intravenous glucose is started at a rate of 10–20 g/h. If there is no improvement within 24 h, intravenous hematin 2–5 mg/kg per day for 3–14 days should be administered. Givosiran is a small interfering RNA (siRNA) that neutralizes excess aminolevulinic acid (ALA) mRNA in hepatocytes for patients with recurrent attacks of acute intermittent porphyria. In clinical trials, givosiran 2.5 mg/kg subcutaneously per month led to reduced attack frequency, better daily pain scores for pain, improved quality of life, lower levels of urinary ALA and porphobilinogen, and fewer days of hematin compared with placebo.

■ CEREBELLAR ATAXIA, NEUROPATHY, AND VESTIBULAR AREFLEXIA SYNDROME (CANVAS)

Cerebellar ataxia, neuropathy, and vestibular areflexia syndrome (CANVAS) appears to be the most common cause of autosomal recessive ataxia. It usually manifests in middle adult life with a sensory neuropathy/neuronopathy that progresses over the course of 10–15 years to cerebellar and vestibular dysfunction, as well as a dry cough. Examination reveals loss of large-fiber sensory modalities with a sensory ataxia as well as cerebellar ataxia. The clinical spectrum is quite broad, however, and some patients manifest with upper and low motor neuron involvement (spasticity, brisk reflexes, muscle atrophy, and fasciculations) similar to amyotrophic lateral sclerosis. CANVAS can also present with dysautonomia and features of parkinsonism. NCS reveal low-amplitude or absent sensory responses that are in a non-length-dependent pattern, and EMG can show signs of active denervation and chronic reinnervation. Brain magnetic resonance imaging (MRI) scans can reveal cerebellar atrophy. Sural nerve biopsies have shown loss of large myelinated axons, and autopsy studies demonstrate degeneration of the dorsal root ganglia and posterior columns. In most cases, CANVAS is associated with biallelic $(AAGGG)_n$ repeat expansions in the second intron of the replication factor complex subunit 1 (*RFC1*). The is a DNA polymerase accessory protein required for the coordinated synthesis of both DNA strands during replication and after DNA damage.

■ FAMILIAL AMYLOID POLYNEUROPATHY

Familial amyloid polyneuropathy (FAP) is phenotypically and genetically heterogeneous and is caused by mutations in the genes for transthyretin (TTR), apolipoprotein A1, or gelsolin **(Chap. 117)**. The majority of patients with FAP have mutations in the TTR gene. Amyloid deposition may be evident in abdominal fat pad, rectal, or nerve biopsies. The clinical features, histopathology, and EDx reveal abnormalities consistent with a generalized or multifocal, predominantly axonal but occasionally demyelinating, polyneuropathy.

Patients with TTR-related FAP usually develop insidious onset of numbness and painful paresthesias in the distal lower limbs in the third to fourth decade of life, although some patients develop the disorder later in life (Pattern 2, Table 457-2). Carpal tunnel syndrome (CTS) is common. Autonomic involvement can be severe, leading to postural hypotension, constipation or persistent diarrhea, erectile dysfunction, and impaired sweating (Pattern 10, Table 457-2). Amyloid deposition also occurs in the heart, kidneys, liver, and corneas. Patients usually die 10–15 years after the onset of symptoms from cardiac failure or complications from malnutrition. Because the liver produces much of the body's TTR, liver transplantation has been used to treat FAP related to TTR mutations. Serum TTR levels decrease after transplantation, and improvement in clinical and EDx features has been reported. Both tafamidis meglumine (20 mg daily) and diflunisal (250 mg twice daily), which prevent misfolding and deposition of mutated TTR, appear to slow the rate of deterioration in patients with TTR-related FAP. Several forms of gene therapy are also now available. Randomized, placebo-controlled trials of antisense oligonucleotides, patisiran 0.3 mg/kg intravenous every 3 weeks, vutrisiran 25 mg subcutaneous every 3 months, and eplontersen 45 mg subcutaneous every 4 weeks, as well as the siRNA inotersen 300 mg subcutaneous weekly, have been shown to be effective in FAP related to TTR mutations. These drugs block expression of both mutant and wild-type TTR, reducing amyloid precursor protein synthesis.

Patients with apolipoprotein A1–related FAP (Van Allen type) usually present in the fourth decade with numbness and painful dysesthesias in the distal limbs. Gradually, the symptoms progress, leading to proximal and distal weakness and atrophy. Although autonomic neuropathy is not severe, some patients develop diarrhea, constipation, or gastroparesis. Most patients die from systemic complications of amyloidosis (e.g., renal failure) 12–15 years after the onset of the neuropathy.

Gelsolin-related amyloidosis (Finnish type) is characterized by the combination of lattice corneal dystrophy and multiple cranial neuropathies that usually begin in the third decade of life. Over time, a mild generalized sensorimotor polyneuropathy develops. Autonomic dysfunction does not occur.

ACQUIRED NEUROPATHIES

■ PRIMARY OR AL AMYLOIDOSIS (SEE CHAP. 117)

Besides FAP, amyloidosis can also be acquired. In primary or AL amyloidosis, the abnormal protein deposition is composed of immunoglobulin light chains. AL amyloidosis occurs in the setting of multiple myeloma (MM), Waldenström's macroglobulinemia, lymphoma, other plasmacytomas, or lymphoproliferative disorders, or without any other identifiable disease.

Approximately 30% of patients with AL primary amyloidosis present with a polyneuropathy, most typically painful dysesthesias and burning sensations in the feet (Pattern 2, Table 457-2). However, the trunk can be involved, and some patients manifest with a mononeuropathy multiplex pattern. CTS occurs in 25% of patients and may be the initial manifestation. The neuropathy is slowly progressive, and eventually, weakness develops along with large-fiber sensory loss. Most patients develop autonomic involvement with postural hypertension, syncope, bowel and bladder incontinence, constipation, impotence, and impaired sweating (Pattern 10, Table 457-2). Patients generally die from their systemic illness (renal failure, cardiac disease).

The monoclonal protein may be composed of IgG, IgA, IgM, or only free light chain. Lambda (λ) is more common than κ light chain (>2:1) in AL amyloidosis. The CSF protein is often increased (with normal cell count), and thus, the neuropathy may be mistaken for CIDP **(Chap. 458)**. Nerve biopsies reveal axonal degeneration and amyloid deposition in either a globular or diffuse pattern infiltrating the perineurial, epineurial, and endoneurial connected tissue and in blood vessel walls.

The median survival of patients with primary amyloidosis is <2 years, with death usually from progressive congestive heart failure or renal failure. Chemotherapy with melphalan, prednisone, and colchicine, to reduce the concentration of monoclonal proteins, and autologous stem cell transplantation may prolong survival, but whether the neuropathy improves is controversial.

DIABETIC NEUROPATHY

DM is the most common cause of peripheral neuropathy in developed countries. DM is associated with several types of polyneuropathy: distal symmetric sensory or sensorimotor polyneuropathy, autonomic neuropathy, diabetic neuropathic cachexia, polyradiculoneuropathies, cranial neuropathies, and other mononeuropathies. Risk factors for the development of neuropathy include long-standing, poorly controlled DM and the presence of retinopathy and nephropathy.

Diabetic Distal Symmetric Sensory and Sensorimotor Polyneuropathy (DSPN) DSPN is the most common form of diabetic neuropathy and manifests as sensory loss beginning in the toes that gradually progresses over time up the legs and into the fingers and arms (Pattern 2, Table 457-2). When severe, a patient may develop sensory loss in the trunk (chest and abdomen), initially in the midline anteriorly and later extending laterally. Tingling, burning, deep aching pains may also be apparent. NCS usually show reduced amplitudes and mild to moderate slowing of conduction velocities. Nerve biopsy reveals axonal degeneration, endothelial hyperplasia, and, occasionally, perivascular inflammation. Tight control of glucose may reduce the risk of developing neuropathy or improve the underlying neuropathy. A variety of medications have been used with variable success to treat painful symptoms associated with DSPN, including anticonvulsants, antidepressants, sodium channel blockers, and other analgesics (Table 457-6).

Diabetic Autonomic Neuropathy Autonomic neuropathy is typically seen in combination with DSPN. The autonomic neuropathy can manifest as abnormal sweating, dysfunctional thermoregulation, dry eyes and mouth, pupillary abnormalities, cardiac arrhythmias, postural hypotension, GI abnormalities (e.g., gastroparesis, postprandial bloating, chronic diarrhea, or constipation), and genitourinary dysfunction (e.g., impotence, retrograde ejaculation, incontinence) (Pattern 10, Table 457-2). Tests of autonomic function are generally abnormal, including sympathetic skin responses and quantitative sudomotor axon reflex testing. Sensory and motor NCS generally demonstrate features described above with DSPN.

Diabetic Radiculoplexus Neuropathy (Diabetic Amyotrophy or Bruns-Garland Syndrome) Diabetic radiculoplexus neuropathy is the presenting manifestation of DM in approximately one-third of patients. Typically, patients present with severe pain in the low back, hip, and thigh in one leg. Rarely, the diabetic polyradiculoneuropathy begins in both legs at the same time (Pattern 4, Table 457-2). Atrophy and weakness of proximal and distal muscles in the affected leg become apparent within a few days or weeks. The neuropathy is often accompanied or heralded by severe weight loss. Weakness usually progresses over several weeks or months but can continue to progress for 18 months or more. Subsequently, there is slow recovery, but many are left with residual weakness, sensory loss, and pain. In contrast to the more typical lumbosacral radiculoplexus neuropathy, some patients develop thoracic radiculopathy or, even less commonly, a cervical polyradiculoneuropathy. CSF protein is usually elevated, while the cell count is normal. ESR is often increased. EDx reveals evidence of active denervation in affected proximal and distal muscles in the limbs and in paraspinal muscles. Nerve biopsies may demonstrate axonal degeneration along with perivascular inflammation. Patients with severe pain are sometimes treated in the acute period with glucocorticoids, although a randomized controlled trial has yet to be performed, and the natural history of this neuropathy is gradual improvement.

Diabetic Mononeuropathies or Multiple Mononeuropathies The most common mononeuropathies are median neuropathy at the wrist and ulnar neuropathy at the elbow, but peroneal neuropathy at the fibular head and sciatic, lateral femoral, cutaneous, or cranial neuropathies also occur (Pattern 3, Table 457-2). In regard to cranial mononeuropathies, seventh nerve palsies are relatively common but may have other, nondiabetic etiologies. In diabetics, a third nerve palsy is most common, followed by sixth nerve and, less frequently, fourth nerve palsies. Diabetic third nerve palsies are characteristically pupil-sparing (Chap. 34).

HYPOTHYROIDISM

Hypothyroidism is more commonly associated with a proximal myopathy, but some patients develop a neuropathy, most typically CTS. Rarely, a generalized sensory polyneuropathy characterized by painful paresthesias and numbness in both the legs and hands can occur. Treatment is correction of the hypothyroidism.

TABLE 457-6 Treatment of Painful Sensory Neuropathies

THERAPY	ROUTE	DOSE	SIDE EFFECTS
First-Line			
Lidoderm 5% patch	Apply to painful area	Up to 3 patches qd	Skin irritation
Tricyclic antidepressants (e.g., amitriptyline, nortriptyline)	PO	10–100 mg qhs	Cognitive changes, sedation, dry eyes and mouth, urinary retention, constipation
Gabapentin	PO	300–1200 mg tid	Cognitive changes, sedation, peripheral edema
Pregabalin	PO	50–100 mg tid	Cognitive changes, sedation, peripheral edema
Duloxetine	PO	30–60 mg qd	Cognitive changes, sedation, dry eyes, diaphoresis, nausea, diarrhea, constipation
Second-Line			
Carbamazepine	PO	200–400 mg q 6–8 h	Cognitive changes, dizziness, leukopenia, liver dysfunction
Phenytoin	PO	200–400 mg qhs	Cognitive changes, dizziness, liver dysfunction
Venlafaxine	PO	37.5–150 mg/d	Asthenia, sweating, nausea, constipation, anorexia, vomiting, somnolence, dry mouth, dizziness, nervousness, anxiety, tremor, and blurred vision as well as abnormal ejaculation/orgasm and impotence
Tramadol	PO	50 mg qid	Cognitive changes, gastrointestinal upset
Third-Line			
Mexiletine	PO	200–300 mg tid	Arrhythmias
Other Agents			
EMLA cream	Apply cutaneously	qid	Local erythema
2.5% lidocaine			
2.5% prilocaine			
Capsaicin 0.025–0.075% cream	Apply cutaneously	qid	Painful burning skin

Source: Modified from AA Amato, J Russell: *Neuromuscular Disorders,* 2nd ed. New York, McGraw-Hill, 2016, Table 22-3, p. 485.

SJÖGREN'S SYNDROME

Sjögren's syndrome, characterized by the sicca complex of xerophthalmia, xerostomia, and dryness of other mucous membranes, can be complicated by neuropathy **(Chap. 373)**. Most common is a length-dependent axonal sensorimotor neuropathy characterized mainly by sensory loss in the distal extremities (Pattern 2, Table 457-2). A pure small-fiber neuropathy or a cranial neuropathy, particularly involving the trigeminal nerve, can also be seen. Sjögren's syndrome is also associated with sensory neuronopathy/ganglionopathy. Patients with sensory ganglionopathies develop progressive numbness and tingling of the limbs, trunk, and face in a non-length-dependent manner such that symptoms can involve the face or arms more than the legs. The onset can be acute or insidious. Sensory examination demonstrates severe vibratory and proprioceptive loss leading to sensory ataxia.

Patients with neuropathy due to Sjögren's syndrome may have ANAs, SS-A/Ro, and SS-B/La antibodies in the serum, but most do not. NCS demonstrate reduced amplitudes of sensory studies in the affected limbs. Nerve biopsy demonstrates axonal degeneration. Nonspecific perivascular inflammation may be present, but only rarely is there necrotizing vasculitis. There is no specific treatment for neuropathies related to Sjögren's syndrome. When vasculitis is suspected, immunosuppressive agents may be beneficial. Occasionally, the sensory neuronopathy/ganglionopathy stabilizes or improves with immunotherapy, such as intravenous immunoglobulin.

RHEUMATOID ARTHRITIS

Peripheral neuropathy occurs in at least 50% of patients with rheumatoid arthritis (RA) and may be vasculitic in nature **(Chap. 370)**. Vasculitic neuropathy can present with a mononeuropathy multiplex (Pattern 3, Table 457-2), a generalized symmetric pattern of involvement (Pattern 2, Table 457-2), or a combination of these patterns **(Chap. 375)**. Neuropathies may also result from drugs used to treat RA (e.g., tumor necrosis blockers, leflunomide). Nerve biopsy often reveals thickening of the epineurial and endoneurial blood vessels as well as perivascular inflammation or vasculitis, with transmural inflammatory cell infiltration and fibrinoid necrosis of vessel walls. The neuropathy is usually responsive to immunomodulating therapies.

SYSTEMIC LUPUS ERYTHEMATOSUS

Between 2 and 27% of individuals with SLE develop a peripheral neuropathy **(Chap. 368)**. Affected patients typically present with a slowly progressive sensory loss beginning in the feet. Some patients develop burning pain and paresthesias with normal reflexes, and NCS suggest a pure small-fiber neuropathy (Pattern 2, Table 457-2). Less common are multiple mononeuropathies presumably secondary to necrotizing vasculitis (Pattern 3, Table 457-2). Rarely, a generalized sensorimotor polyneuropathy meeting clinical, laboratory, electrophysiologic, and histologic criteria for either GBS or CIDP may occur. Immunosuppressive therapy may be beneficial in SLE patients with neuropathy due to vasculitis. Immunosuppressive agents are less likely to be effective in patients with a generalized sensory or sensorimotor polyneuropathy without evidence of vasculitis. Patients with a GBS or CIDP-like neuropathy should be treated accordingly **(Chap. 458)**.

SYSTEMIC SCLEROSIS (SCLERODERMA)

A distal symmetric, mainly sensory polyneuropathy complicates 5–67% of scleroderma cases (Pattern 2, Table 457-2) **(Chap. 372)**. Cranial mononeuropathies can also develop, most commonly of the trigeminal nerve, producing numbness and dysesthesias in the face. Multiple mononeuropathies also occur (Pattern 3, Table 457-2). The EDx and histologic features of nerve biopsy are those of an axonal sensory greater than motor polyneuropathy.

MIXED CONNECTIVE TISSUE DISEASE

A mild distal axonal sensorimotor polyneuropathy occurs in ~10% of patients with mixed connective tissue disease.

SARCOIDOSIS

The peripheral nervous system or CNS is involved in ~5% of patients with sarcoidosis **(Chap. 379)**. The most common cranial nerve involved is the seventh nerve, which can be affected bilaterally. Some patients develop radiculopathy or polyradiculopathy (Pattern 4, Table 457-2). With a generalized root involvement, the clinical presentation can mimic GBS or CIDP. Patients can also present with multiple mononeuropathies (Pattern 3, Table 457-2) or a generalized, slowly progressive, sensory greater than motor polyneuropathy (Pattern 2, Table 457-2). Some have features of a pure small-fiber neuropathy. EDx reveals an axonal neuropathy. Nerve biopsy can reveal noncaseating granulomas infiltrating the endoneurium, perineurium, or epineurium along with lymphocytic necrotizing angiitis. Neurosarcoidosis may respond to treatment with glucocorticoids or other immunosuppressive agents.

HYPEREOSINOPHILIC SYNDROME

Hypereosinophilic syndrome is characterized by eosinophilia associated with various skin, cardiac, hematologic, and neurologic abnormalities. A generalized peripheral neuropathy or a mononeuropathy multiplex occurs in 6–14% of patients (Pattern 2, Table 457-2).

CELIAC DISEASE (GLUTEN-INDUCED ENTEROPATHY OR NONTROPICAL SPRUE)

Neurologic complications, particularly ataxia and peripheral neuropathy, are estimated to occur in 10% of patients with celiac disease **(Chap. 336)**. A generalized sensorimotor polyneuropathy, pure motor neuropathy, multiple mononeuropathies, autonomic neuropathy, small-fiber neuropathy, and neuromyotonia have all been reported in association with celiac disease or antigliadin/antiendomysial antibodies (Patterns 2, 3, and 9; Table 457-2). Nerve biopsy may reveal a loss of large myelinated fibers. The neuropathy may be secondary to malabsorption of vitamins B_{12} and E. However, some patients have no appreciable vitamin deficiencies. The pathogenic basis for the neuropathy in these patients is unclear but may be autoimmune in etiology. The neuropathy does not appear to respond to a gluten-free diet. In patients with vitamin B_{12} or vitamin E deficiency, replacement therapy may improve or stabilize the neuropathy.

INFLAMMATORY BOWEL DISEASE

Ulcerative colitis and Crohn's disease may be complicated by GBS, CIDP, generalized axonal sensory or sensorimotor polyneuropathy, small-fiber neuropathy, or mononeuropathy (Patterns 2 and 3, Table 457-2) **(Chap. 337)**. These neuropathies may be autoimmune, nutritional (e.g., vitamin B_{12} deficiency), treatment related (e.g., metronidazole), or idiopathic in nature. An acute neuropathy with demyelination resembling GBS, CIDP, or multifocal motor neuropathy may occur in patients treated with tumor necrosis factor α blockers.

UREMIC NEUROPATHY

Approximately 60% of patients with renal failure develop a polyneuropathy characterized by length-dependent numbness, tingling, allodynia, and mild distal weakness (Pattern 2, Table 457-2). Rarely, a rapidly progressive weakness and sensory loss very similar to GBS can occur that improves with an increase in the intensity of renal dialysis or with transplantation (Pattern 1, Table 457-2). Mononeuropathies can also occur, the most common of which is CTS. Ischemic monomelic neuropathy (see below) can complicate arteriovenous shunts created in the arm for dialysis (Pattern 3, Table 457-2). EDx in uremic patients reveals features of a length-dependent, primarily axonal, sensorimotor polyneuropathy. Sural nerve biopsies demonstrate a loss of nerve fibers (particularly large myelinated nerve fibers), active axonal degeneration, and segmental and paranodal demyelination. The sensorimotor polyneuropathy can be stabilized by hemodialysis and improved with successful renal transplantation.

CHRONIC LIVER DISEASE

A generalized sensorimotor neuropathy characterized by numbness, tingling, and minor weakness in the distal aspects of primarily the lower limbs commonly occurs in patients with chronic liver failure. EDx studies are consistent with a sensory greater than motor axonopathy. Occasionally patients with severe liver disease develop a combined neuropathy and myopathy. Sural nerve biopsy reveals both segmental demyelination and axonal loss. It is not known if hepatic failure in isolation can cause peripheral neuropathy, as the majority of patients have liver disease secondary to other disorders, such as alcoholism or viral hepatitis, which can also cause neuropathy.

■ CRITICAL ILLNESS POLYNEUROPATHY

The most common causes of acute generalized weakness leading to admission to a medical intensive care unit (ICU) are GBS and myasthenia gravis (Pattern 1, Table 457-2) **(Chaps. 458 and 459)**. However, weakness developing in critically ill patients while in the ICU is usually caused by critical illness polyneuropathy (CIP) or critical illness myopathy (CIM) or, much less commonly, by prolonged neuromuscular blockade. From a clinical and EDx standpoint, it can be quite difficult to distinguish these disorders. Most specialists believe that CIM is more common. Both CIM and CIP develop as a complication of sepsis and multiple organ failure. They usually present as an inability to wean a patient from a ventilator. A coexisting encephalopathy may limit the neurologic examination, in particular the sensory examination. Muscle stretch reflexes are absent or reduced.

Serum creatine kinase (CK) is usually normal; an elevated serum CK would point to CIM as opposed to CIP. NCS reveal absent or markedly reduced amplitudes of motor and sensory studies in CIP, whereas sensory studies are relatively preserved in CIM. Needle EMG usually reveals profuse positive sharp waves and fibrillation potentials, and it is not unusual in patients with severe weakness to be unable to recruit motor unit action potentials. The pathogenic basis of CIP is not known. Perhaps circulating toxins and metabolic abnormalities associated with sepsis and multiorgan failure impair axonal transport or mitochondrial function, leading to axonal degeneration.

■ LEPROSY (HANSEN'S DISEASE)

Leprosy, caused by the acid-fast bacteria *Mycobacterium leprae*, is the most common cause of peripheral neuropathy in Southeast Asia, Africa, and South America **(Chap. 184)**. Clinical manifestations range from tuberculoid leprosy at one end of the spectrum to lepromatous leprosy at the other end, with borderline leprosy in between. Neuropathies are most common in patients with borderline leprosy. Superficial cutaneous nerves of the ears and distal limbs are commonly affected. Mononeuropathies, multiple mononeuropathies, or a slowly progressive symmetric sensorimotor polyneuropathy may develop (Patterns 2 and 3, Table 457-2). Sensory NCS are usually absent in the lower limb and are reduced in amplitude in the arms. Motor NCS may demonstrate reduced amplitudes in affected nerves but occasionally can reveal demyelinating features. Leprosy is usually diagnosed by skin lesion biopsy. Nerve biopsy can also be diagnostic, particularly when there are no apparent skin lesions. The tuberculoid form is characterized by granulomas, and bacilli are not seen. In contrast, with lepromatous leprosy, large numbers of infiltrating bacilli, T_H2 lymphocytes, and organism-laden, foamy macrophages with minimal granulomatous infiltration are evident. The bacilli are best appreciated using the Fite stain, where they can be seen as red-staining rods often in clusters free in the endoneurium, within macrophages, or within Schwann cells.

Patients are generally treated with multiple drugs: dapsone, rifampin, and clofazimine. Other medications that are used include thalidomide, pefloxacin, ofloxacin, sparfloxacin, minocycline, and clarithromycin. Patients are generally treated for 2 years. Treatment is sometimes complicated by the so-called reversal reaction, particularly in borderline leprosy. The reversal reaction can occur at any time during treatment and develops because of a shift to the tuberculoid end of the spectrum, with an increase in cellular immunity during treatment. The cellular response is upregulated as evidenced by an increased release of tumor necrosis factor α, interferon γ, and interleukin 2, with new granuloma formation. This can result in an exacerbation of the rash and the neuropathy as well as in appearance of new lesions. High-dose glucocorticoids blunt this adverse reaction and may be used prophylactically at treatment onset in high-risk patients. Erythema nodosum leprosum (ENL) is also treated with glucocorticoids or thalidomide.

■ LYME DISEASE

Lyme disease is caused by infection with *Borrelia burgdorferi*, a spirochete usually transmitted by the deer tick *Ixodes dammini* **(Chap. 191)**. Neurologic complications may develop during the second and third stages of infection. Facial neuropathy is most common and is bilateral in about half of cases, which is rare for idiopathic Bell's palsy. Involvement of nerves is frequently asymmetric. Some patients present with a polyradiculoneuropathy or multiple mononeuropathies (Pattern 3 or 4, Table 457-2). EDx is suggestive of a primary axonopathy. Nerve biopsies can reveal axonal degeneration with perivascular inflammation. Treatment is with antibiotics.

■ DIPHTHERITIC NEUROPATHY

Diphtheria is caused by the bacteria *Corynebacterium diphtheriae* **(Chap. 155)**. Infected individuals present with flu-like symptoms of generalized myalgias, headache, fatigue, low-grade fever, and irritability within a week to 10 days of the exposure. Between 20 and 70% of patients develop a peripheral neuropathy caused by a toxin released by the bacteria. Three to 4 weeks after infection, patients may note decreased sensation in their throat and begin to develop dysphagia, dysarthria, hoarseness, and blurred vision due to impaired accommodation. A generalized polyneuropathy may manifest 2 or 3 months following the initial infection, characterized by numbness, paresthesias, and weakness of the arms and legs and occasionally ventilatory failure (Pattern 1, Table 457-2). CSF protein can be elevated with or without lymphocytic pleocytosis. EDx suggests a diffuse axonal sensorimotor polyneuropathy. Antitoxin and antibiotics should be given within 48 h of symptom onset. Although early treatment reduces the incidence and severity of some complications (i.e., cardiomyopathy), it does not appear to alter the natural history of the associated peripheral neuropathy. The neuropathy usually resolves after several months.

■ COVID-19

GBS **(Chap. 458)** has been reported in the setting of acute COVID-19 infection though a causal relationship has not been clearly established. There does appear to be an increased risk of GBS with adenovirus-vector vaccines but not the messenger RNA vaccines.

■ HUMAN IMMUNODEFICIENCY VIRUS

HIV infection can result in a variety of neurologic complications, including peripheral neuropathies **(Chap. 208)**. Approximately 20% of HIV-infected individuals develop a neuropathy as a direct result of the virus itself or as a result of other associated viral infections (e.g., CMV) or neurotoxicity secondary to antiviral medications (see below). The major presentations of peripheral neuropathy associated with HIV infection include (1) distal symmetric polyneuropathy (DSP), (2) inflammatory demyelinating polyneuropathy (including both GBS and CIDP), (3) multiple mononeuropathies (e.g., vasculitis, CMV-related), (4) polyradiculopathy (usually CMV-related), (5) autonomic neuropathy, and (6) sensory ganglionitis.

HIV-Related Distal Symmetric Polyneuropathy DSP is the most common form of peripheral neuropathy associated with HIV infection and usually is seen in patients with AIDS. It is characterized by numbness and painful paresthesias involving the distal extremities (Pattern 2, Table 457-2). The pathogenic basis for DSP is unknown but is not due to actual infection of the peripheral nerves. The neuropathy may be immune mediated, perhaps caused by the release of cytokines from surrounding inflammatory cells. Vitamin B_{12} deficiency may contribute in some instances but is not a major cause of most cases of DSP. Older antiretroviral agents (e.g., dideoxycytidine, dideoxyinosine, stavudine) are also neurotoxic and can cause a painful sensory neuropathy.

HIV-Related Inflammatory Demyelinating Polyradiculoneuropathy Both acute inflammatory demyelinating polyneuropathy (AIDP) and CIDP can occur as a complication of HIV infection (Pattern 1, Table 457-2). AIDP usually develops at the time of seroconversion, whereas CIDP can occur any time in the course of the infection. Clinical and EDx features are indistinguishable from idiopathic AIDP or CIDP (Chap. 458). In addition to elevated protein levels, lymphocytic pleocytosis is evident in the CSF, a finding that helps distinguish this HIV-associated polyradiculoneuropathy from idiopathic AIDP/CIDP.

HIV-Related Progressive Polyradiculopathy An acute, progressive lumbosacral polyradiculoneuropathy usually secondary to CMV infection can develop in patients with AIDS (Pattern 4, Table 457-2). Patients present with severe radicular pain, numbness, and weakness in the legs, which is usually asymmetric. CSF is abnormal, demonstrating a high protein level, along with a reduced glucose concentration and notably a neutrophilic pleocytosis. EDx studies reveal features of active axonal degeneration. The polyradiculoneuropathy may improve with antiviral therapy.

HIV-Related Multiple Mononeuropathies Multiple mononeuropathies can also develop in patients with HIV infection, usually in the context of AIDS. Weakness, numbness, paresthesias, and pain occur in the distribution of affected nerves (Pattern 3, Table 457-2). Nerve biopsies can reveal axonal degeneration with necrotizing vasculitis or perivascular inflammation. Glucocorticoid treatment is indicated for vasculitis directly due to HIV infection.

HIV-Related Sensory Neuronopathy/Ganglionopathy Dorsal root ganglionitis is a very rare complication of HIV infection, and neuronopathy can be the presenting manifestation. Patients develop sensory ataxia similar to idiopathic sensory neuronopathy/ganglionopathy (Pattern 9, Table 457-2). NCS reveal reduced amplitudes or absence of sensory nerve action potentials (SNAPs).

■ HERPES VARICELLA-ZOSTER VIRUS

Peripheral neuropathy from herpes varicella-zoster (HVZ) infection results from reactivation of latent virus or from a primary infection (Chap. 198). Two-thirds of infections in adults are characterized by dermal zoster in which severe pain and paresthesias develop in a dermatomal region followed within a week or two by a vesicular rash in the same distribution (Pattern 3, Table 457-2). Weakness in muscles innervated by roots corresponding to the dermatomal distribution of skin lesions occurs in 5–30% of patients. Approximately 25% of affected patients have continued pain (postherpetic neuralgia [PHN]). A large clinical trial demonstrated that vaccination against zoster reduces the incidence of HVZ among vaccine recipients by 51% and reduces the incidence of PHN by 67%. Treatment of PHN is symptomatic (Table 457-6).

■ CYTOMEGALOVIRUS

CMV can cause an acute lumbosacral polyradiculopathy and multiple mononeuropathies in patients with HIV infection and in other immune deficiency conditions (Pattern 4, Table 457-2) (Chap. 200).

■ EPSTEIN-BARR VIRUS

EBV infection has been associated with GBS, cranial neuropathies, mononeuropathy multiplex, brachial plexopathy, lumbosacral radiculoplexopathy, and sensory neuronopathies (Patterns 1, 3, 4, and 9, Table 457-2) (Chap. 199).

■ HEPATITIS VIRUSES

Hepatitis B and C can cause multiple mononeuropathies related to vasculitis, AIDP, or CIDP (Patterns 1 and 3, Table 457-2) (Chap. 352).

NEUROPATHIES ASSOCIATED WITH MALIGNANCY

Patients with malignancy can develop neuropathies due to (1) a direct effect of the cancer by invasion or compression of the nerves, (2) remote or paraneoplastic effect, (3) a toxic effect of treatment, or (4) as a consequence of immune compromise caused by immunosuppressive medications. The most common associated malignancy is lung cancer, but neuropathies also complicate carcinoma of the breast, ovaries, stomach, colon, rectum, and other organs, including the lymphoproliferative system.

■ PARANEOPLASTIC SENSORY NEURONOPATHY/GANGLIONOPATHY

Paraneoplastic encephalomyelitis/sensory neuronopathy (PEM/SN) usually complicates small-cell lung carcinoma (Chap. 99). Patients usually present with numbness and paresthesias in the distal extremities that are often asymmetric. The onset can be acute or insidiously progressive. Prominent loss of proprioception leads to sensory ataxia (Pattern 9; Table 457-2). Weakness can be present, usually secondary to an associated myelitis, motor neuronopathy, or concurrent LEMS. Many patients also develop confusion, memory loss, depression, hallucinations or seizures, or cerebellar ataxia. Polyclonal antineuronal antibodies (IgG) directed against a 35- to 40-kDa protein or complex of proteins, the so-called Hu antigen, are found in the sera or CSF in the majority of patients with paraneoplastic PEM/SN. CSF may be normal or may demonstrate mild lymphocytic pleocytosis and elevated protein. PEM/SN is probably the result of antigenic similarity between proteins expressed in the tumor cells and neuronal cells, leading to an immune response directed against both cell types. Treatment of the underlying cancer generally does not affect the course of PEM/SN. However, occasional patients may improve following treatment of the tumor. Unfortunately, plasmapheresis, intravenous immunoglobulin, and immunosuppressive agents have not shown benefit.

■ NEUROPATHY SECONDARY TO TUMOR INFILTRATION

Malignant cells, in particular leukemia and lymphoma, can infiltrate cranial and peripheral nerves, leading to mononeuropathy, mononeuropathy multiplex, polyradiculopathy, plexopathy, or even a generalized symmetric distal or proximal and distal polyneuropathy (Patterns 1, 2, 3, and 4; Table 457-2). Neuropathy related to tumor infiltration is often painful; it can be the presenting manifestation of the cancer or the heralding symptom of a relapse. The neuropathy may improve with treatment of the underlying leukemia or lymphoma or with glucocorticoids.

■ NEUROPATHY AS A COMPLICATION OF BONE MARROW TRANSPLANTATION

Neuropathies may develop in patients who undergo bone marrow transplantation (BMT) because of the toxic effects of chemotherapy, radiation, infection, or an autoimmune response directed against the peripheral nerves. Peripheral neuropathy in BMT is often associated with graft-versus-host disease (GVHD). Chronic GVHD shares many features with a variety of autoimmune disorders, and it is possible that an immune-mediated response directed against peripheral nerves is responsible. Patients with chronic GVHD may develop cranial neuropathies, sensorimotor polyneuropathies, multiple mononeuropathies, and severe generalized peripheral neuropathies resembling AIDP or CIDP (Patterns 1, 2, and 3; Table 457-2). The neuropathy may improve by increasing the intensity of immunosuppressive or immunomodulating therapy and resolution of the GVHD.

■ LYMPHOMA

Lymphomas may cause neuropathy by infiltration or direct compression of nerves or by a paraneoplastic process. The neuropathy can be purely sensory or motor but most commonly is sensorimotor. The pattern of involvement may be symmetric, asymmetric, or multifocal, and the course may be acute, gradually progressive, or relapsing and remitting (Patterns 1, 2, and 3; Table 457-2). EDx can be compatible with either an axonal or demyelinating process. CSF may reveal lymphocytic pleocytosis and an elevated protein. Nerve biopsy may demonstrate endoneurial inflammatory cells in both the infiltrative and the paraneoplastic etiologies. A monoclonal population of cells favors lymphomatous invasion. The neuropathy may respond to treatment of the underlying lymphoma or immunomodulating therapies.

■ MULTIPLE MYELOMA

MM usually presents in the fifth to seventh decade of life with fatigue, bone pain, anemia, and hypercalcemia (Chap. 116). Clinical and EDx features of neuropathy occur in as many as 40% of patients. The most common pattern is that of a distal, axonal, sensory, or sensorimotor polyneuropathy (Pattern 2; Table 457-2). Less frequently, a chronic demyelinating polyradiculoneuropathy may develop (Pattern 1; Table 457-2) (see POEMS, Chap. 458). MM can be complicated by amyloid polyneuropathy and should be considered in patients with painful paresthesias, loss of pinprick and temperature discrimination, and autonomic dysfunction (suggestive of a small-fiber neuropathy) and CTS. Expanding plasmacytomas can compress cranial nerves and spinal roots as well. A monoclonal protein, usually composed of γ or μ heavy chains or κ light chains, may be identified in the serum or urine. EDx usually shows reduced amplitudes with normal or only mildly abnormal distal latencies and conduction velocities. A superimposed median neuropathy at the wrist is common. Abdominal fat pad, rectal, or sural nerve biopsy can be performed to look for amyloid deposition. Unfortunately, the treatment of the underlying MM does not usually affect the course of the neuropathy.

■ NEUROPATHIES ASSOCIATED WITH MONOCLONAL GAMMOPATHY OF UNCERTAIN SIGNIFICANCE (SEE CHAP. 458)

Toxic Neuropathies Secondary to Chemotherapy Many of the commonly used chemotherapy agents can cause a toxic neuropathy (Table 457-7). The mechanisms by which these agents cause toxic neuropathies vary, as does the specific type of neuropathy produced. The risk of developing a toxic neuropathy or more severe neuropathy appears to be greater in patients with a preexisting neuropathy (e.g., CMT disease, diabetic neuropathy) and those who also take other potentially neurotoxic drugs (e.g., nitrofurantoin, isoniazid, disulfiram, pyridoxine). Chemotherapeutic agents usually cause a sensory greater than motor length-dependent axonal neuropathy or neuronopathy/ganglionopathy (Patterns 2 and 9; Table 457-2).

OTHER TOXIC NEUROPATHIES

Neuropathies can develop as complications of toxic effects of various drugs and other environmental exposures (Table 457-8). The more common neuropathies associated with these agents are discussed here.

■ CHLOROQUINE AND HYDROXYCHLOROQUINE

Chloroquine and hydroxychloroquine can cause a toxic myopathy characterized by slowly progressive, painless, proximal weakness and atrophy, which is worse in the legs than the arms. In addition, neuropathy can also develop with or without the myopathy leading to sensory loss and distal weakness. The "neuromyopathy" usually appears in patients taking 500 mg daily for a year or more but has been reported with doses as low as 200 mg/d. Serum CK levels are usually elevated due to the superimposed myopathy. NCS reveal mild slowing of motor and sensory NCVs with a mild to moderate reduction in the amplitudes, although NCS may be normal in patients with only the myopathy. EMG demonstrates myopathic muscle action potentials (MUAPs), increased insertional activity in the form of positive sharp waves, fibrillation potentials, and occasionally myotonic potentials, particularly in the proximal muscles. Neurogenic MUAPs and reduced recruitment are found in more distal muscles. Nerve biopsy demonstrates autophagic vacuoles within Schwann cells. Vacuoles may also be evident in muscle biopsies. The pathogenic basis

TABLE 457-7 Toxic Neuropathies Secondary to Chemotherapy

DRUG	MECHANISM OF NEUROTOXICITY	CLINICAL FEATURES	NERVE HISTOPATHOLOGY	EMG/NCS
Vinca alkaloids (vincristine, vinblastine, vindesine, vinorelbine)	Interfere with axonal microtubule assembly; impairs axonal transport	Symmetric, S-M, large-/small-fiber PN; autonomic symptoms common; infrequent cranial neuropathies	Axonal degeneration of myelinated and unmyelinated fibers; regenerating clusters, minimal segmental demyelination	Axonal sensorimotor PN; distal denervation on EMG; abnormal QST, particularly vibratory perception
Cisplatin	Preferential damage to dorsal root ganglia: ? binds to and cross-links DNA ? inhibits protein synthesis ? impairs axonal transport	Predominant large-fiber sensory neuronopathy; sensory ataxia	Loss of large > small myelinated and unmyelinated fibers; axonal degeneration with small clusters of regenerating fibers; secondary segmental demyelination	Low-amplitude or unobtainable SNAPs with normal CMAPs and EMG; abnormal QST, particularly vibratory perception
Taxanes (paclitaxel, docetaxel)	Promotes axonal microtubule assembly; interferes with axonal transport	Symmetric, predominantly sensory PN; large-fiber modalities affected more than small-fiber	Loss of large > small myelinated and unmyelinated fibers; axonal degeneration with small clusters of regenerating fibers; secondary segmental demyelination	Axonal sensorimotor PN; distal denervation on EMG; abnormal QST, particularly vibratory perception
Suramin				
Axonal PN	Unknown;? inhibition of neurotrophic growth factor binding;? neuronal lysosomal storage	Symmetric, length-dependent, sensory-predominant PN	None described	Abnormalities consistent with an axonal S-M PN
Demyelinating PN	Unknown;? immunomodulating effects	Subacute, S-M PN with diffuse proximal and distal weakness; areflexia; increased CSF protein	Loss of large and small myelinated fibers with primary demyelination and secondary axonal degeneration; occasional epi- and endoneurial inflammatory cell infiltrates	Features suggestive of an acquired demyelinating sensorimotor PN (e.g., slow CVs, prolonged distal latencies and F-wave latencies, conduction block, temporal dispersion)
Cytarabine (ARA-C)	Unknown;? selective Schwann cell toxicity;? immunomodulating effects	GBS-like syndrome; pure sensory neuropathy; brachial plexopathy	Loss of myelinated nerve fibers; axonal degeneration; segmental demyelination; no inflammation	Axonal, demyelinating, or mixed S-M PN; denervation on EMG
Etoposide (VP-16)	Unknown;? selective dorsal root ganglia toxicity	Length-dependent, sensory-predominant PN; autonomic neuropathy	None described	Abnormalities consistent with an axonal S-M PN
Bortezomib (Velcade)	Unknown	Length-dependent, sensory, predominantly small-fiber PN	Not reported	Abnormalities consistent with an axonal sensory neuropathy with early small-fiber involvement (abnormal autonomic studies)

Abbreviations: CMAP, compound motor action potential; CSF, cerebrospinal fluid; CVs, conduction velocities; EMG, electromyography; GBS, Guillain-Barré syndrome; NCS, nerve conduction studies; PN, polyneuropathy; QST, quantitative sensory testing; S-M, sensorimotor; SNAP, sensory nerve action potential.

Source: Reproduced with permission from AA Amato, J Russell: Neuromuscular Disorders, 2nd ed. New York: McGraw-Hill; 2016.

TABLE 457-8 Toxic Neuropathies

DRUG	MECHANISM OF NEUROTOXICITY	CLINICAL FEATURES	NERVE HISTOPATHOLOGY	EMG/NCS
Misonidazole	Unknown	Painful paresthesias and loss of large- and small-fiber sensory modalities and sometimes distal weakness in length-dependent pattern	Axonal degeneration of large, myelinated fibers; axonal swellings; segmental demyelination	Low-amplitude or unobtainable SNAPs with normal or only slightly reduced CMAP amplitudes
Metronidazole	Unknown	Painful paresthesias and loss of large- and small-fiber sensory modalities and sometimes distal weakness in length-dependent pattern	Axonal degeneration	Low-amplitude or unobtainable SNAPs with normal CMAPs
Chloroquine and hydroxychloroquine	Amphiphilic properties may lead to drug-lipid complexes that are indigestible and result in accumulation of autophagic vacuoles	Loss of large- and small-fiber sensory modalities and distal weakness in length-dependent pattern; superimposed myopathy may lead to proximal weakness	Axonal degeneration with autophagic vacuoles in nerves as well as muscle fibers	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes; distal denervation on EMG; irritability and myopathic-appearing MUAPs proximally in patients with superimposed toxic myopathy
Amiodarone	Amphiphilic properties may lead to drug-lipid complexes that are indigestible and result in accumulation of autophagic vacuoles	Paresthesias and pain with loss of large- and small-fiber sensory modalities and distal weakness in length-dependent pattern; superimposed myopathy may lead to proximal weakness	Axonal degeneration and segmental demyelination with myeloid inclusions in nerves and muscle fibers	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes; can also have prominent slowing of CVs; distal denervation on EMG; irritability and myopathic-appearing MUAPs proximally in patients with superimposed toxic myopathy
Colchicine	Inhibits polymerization of tubulin in microtubules and impairs axoplasmic flow	Numbness and paresthesias with loss of large-fiber modalities in a length-dependent fashion; superimposed myopathy may lead to proximal in addition to distal weakness	Nerve biopsy demonstrates axonal degeneration; muscle biopsy reveals fibers with vacuoles	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes; irritability and myopathic-appearing MUAPs proximally in patients with superimposed toxic myopathy
Podophyllin	Binds to microtubules and impairs axoplasmic flow	Sensory loss, tingling, muscle weakness, and diminished muscle stretch reflexes in length-dependent pattern; autonomic neuropathy	Axonal degeneration	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes
Thalidomide	Unknown	Numbness, tingling, and burning pain and weakness in a length-dependent pattern	Axonal degeneration; autopsy studies reveal degeneration of dorsal root ganglia	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes
Disulfiram	Accumulation of neurofilaments and impaired axoplasmic flow	Numbness, tingling, and burning pain in a length-dependent pattern	Axonal degeneration with accumulation of neurofilaments in the axons	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes
Dapsone	Unknown	Distal weakness that may progress to proximal muscles; sensory loss	Axonal degeneration and segmental demyelination	Low-amplitude or unobtainable CMAPs with normal or reduced SNAP amplitudes
Leflunomide	Unknown	Paresthesias and numbness in a length-dependent pattern	Unknown	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes
Nitrofurantoin	Unknown	Numbness, painful paresthesias, and severe weakness that may resemble GBS	Axonal degeneration; autopsy studies reveal degeneration of dorsal root ganglia and anterior horn cells	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes
Pyridoxine (vitamin B_6)	Unknown	Dysesthesias and sensory ataxia; impaired large-fiber sensory modalities on examination	Marked loss of sensory axons and cell bodies in dorsal root ganglia	Reduced amplitudes or absent SNAPs
Isoniazid	Inhibits pyridoxal phosphokinase leading to pyridoxine deficiency	Dysesthesias and sensory ataxia; impaired large-fiber sensory modalities on examination	Marked loss of sensory axons and cell bodies in dorsal root ganglia and degeneration of the dorsal columns	Reduced amplitudes or absent SNAPs and, to a lesser extent, CMAPs
Ethambutol	Unknown	Numbness with loss of large-fiber modalities on examination	Axonal degeneration	Reduced amplitudes or absent SNAPs
Antinucleosides	Unknown	Dysesthesia and sensory ataxia; impaired large-fiber sensory modalities on examination	Axonal degeneration	Reduced amplitudes or absent SNAPs
Phenytoin	Unknown	Numbness with loss of large-fiber modalities on examination	Axonal degeneration and segmental demyelination	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes
Lithium	Unknown	Numbness with loss of large-fiber modalities on examination	Axonal degeneration	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes

(*Continued*)

TABLE 457-8 Toxic Neuropathies (Continued)

DRUG	MECHANISM OF NEUROTOXICITY	CLINICAL FEATURES	NERVE HISTOPATHOLOGY	EMG/NCS
Acrylamide	Unknown; may be caused by impaired axonal transport	Numbness with loss of large-fiber modalities on examination; sensory ataxia; mild distal weakness	Degeneration of sensory axons in peripheral nerves and posterior columns, spinocerebellar tracts, mammillary bodies, optic tracts, and corticospinal tracts in the CNS	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes
Carbon disulfide	Unknown	Length-dependent numbness and tingling with mild distal weakness	Axonal swellings with accumulation of neurofilaments	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes
Ethylene oxide	Unknown; may act as alkylating agent and bind DNA	Length-dependent numbness and tingling; may have mild distal weakness	Axonal degeneration	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes
Organophosphates	Bind and inhibit neuropathy target esterase	Early features are those of neuromuscular blockade with generalized weakness; later axonal sensorimotor PN ensues	Axonal degeneration along with degeneration of gracile fasciculus and corticospinal tracts	Early: repetitive firing of CMAPs and decrement with repetitive nerve stimulation; late: axonal sensorimotor PN
Hexacarbons	Unknown; may lead to covalent cross-linking between neurofilaments	Acute, severe sensorimotor PN that may resemble GBS	Axonal degeneration and giant axons swollen with neurofilaments	Features of a mixed axonal and/or demyelinating sensorimotor axonal PN—reduced amplitudes, prolonged distal latencies, conduction block, and slowing of CVs
Lead	Unknown; may interfere with mitochondria	Encephalopathy; motor neuropathy (often resembles radial neuropathy with wrist and finger drop); autonomic neuropathy; bluish-black discoloration of gums	Axonal degeneration of motor axons	Reduction of CMAP amplitudes with active denervation on EMG
Mercury	Unknown; may combine with sulfhydryl groups	Abdominal pain and nephrotic syndrome; encephalopathy; ataxia; paresthesias	Axonal degeneration; degeneration of dorsal root ganglia, calcarine, and cerebellar cortex	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes
Thallium	Unknown	Encephalopathy; painful sensory symptoms; mild loss of vibration; distal or generalized weakness may also develop; autonomic neuropathy; alopecia	Axonal degeneration	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes
Arsenic	Unknown; may combine with sulfhydryl groups	Abdominal discomfort, burning pain, and paresthesias; generalized weakness; autonomic insufficiency; can resemble GBS	Axonal degeneration	Low-amplitude or unobtainable SNAPs with normal or reduced CMAP amplitudes; may have demyelinating features: prolonged distal latencies and slowing of CVs
Gold	Unknown	Distal paresthesias and reduction of all sensory modalities	Axonal degeneration	Low-amplitude or unobtainable SNAPs

Abbreviations: CMAP, compound motor action potential; CVs, conduction velocities; EMG, electromyography; GBS, Guillain-Barré syndrome; MUAP, muscle action potential; NCS, nerve conduction studies; PN, polyneuropathy; S-M, sensorimotor; SNAP, sensory nerve action potential.

Source: Reproduced with permission from AA Amato, J Russell: Neuromuscular Disorders, 2nd ed. New York: McGraw-Hill; 2016.

of the neuropathy is not known but may be related to the amphiphilic properties of the drug. These agents contain both hydrophobic and hydrophilic regions that allow them to interact with the anionic phospholipids of cell membranes and organelles. The drug-lipid complexes may be resistant to digestion by lysosomal enzymes, leading to the formation of autophagic vacuoles filled with myeloid debris that may in turn cause degeneration of nerves and muscle fibers. The signs and symptoms of the neuropathy and myopathy are usually reversible following discontinuation of medication.

■ AMIODARONE

Amiodarone can cause a neuromyopathy similar to chloroquine and hydroxychloroquine. The neuromyopathy typically appears after patients have taken the medication for 2–3 years. Nerve biopsy demonstrates a combination of segmental demyelination and axonal loss. Electron microscopy reveals lamellar or dense inclusions in Schwann cells, pericytes, and endothelial cells. The inclusions in muscle and nerve biopsies have persisted as long as 2 years following discontinuation of the medication.

■ COLCHICINE

Colchicine can also cause a neuromyopathy. Patients usually present with proximal weakness and numbness and tingling in the distal extremities. EDx reveals features of an axonal polyneuropathy. Muscle biopsy reveals a vacuolar myopathy, whereas sensory nerves demonstrate axonal degeneration. Colchicine inhibits the polymerization of tubulin into microtubules. The disruption of the microtubules probably leads to defective intracellular movement of important proteins, nutrients, and waste products in muscle and nerves.

■ THALIDOMIDE

Thalidomide is an immunomodulating agent used to treat MM, GVHD, leprosy, and other autoimmune disorders. Thalidomide is associated with severe teratogenic effects as well as peripheral neuropathy that can be dose-limiting. Patients develop numbness, painful tingling, and burning discomfort in the feet and hands and less commonly muscle weakness and atrophy. Even after stopping the drug for 4–6 years, as many as 50% patients continue to have significant symptoms. NCS demonstrate reduced amplitudes or complete absence of SNAPs, with preserved conduction velocities when obtainable. Motor

NCS are usually normal. Nerve biopsy reveals a loss of large-diameter myelinated fibers and axonal degeneration. Degeneration of dorsal root ganglion cells has been reported at autopsy.

PYRIDOXINE (VITAMIN B_6) TOXICITY

Pyridoxine is an essential vitamin that serves as a coenzyme for transamination and decarboxylation. However, at high doses (116 mg/d), patients can develop a severe sensory neuropathy with dysesthesias and sensory ataxia. NCS reveal absent or markedly reduced SNAP amplitudes with relatively preserved CMAPs. Nerve biopsy reveals axonal loss of fiber at all diameters. Loss of dorsal root ganglion cells with subsequent degeneration of both the peripheral and central sensory tracts have been reported in animal models.

■ ISONIAZID

One of the most common side effects of isoniazid (INH) is peripheral neuropathy. Standard doses of INH (3–5 mg/kg per day) are associated with a 2% incidence of neuropathy, whereas neuropathy develops in at least 17% of patients taking in excess of 6 mg/kg per d. The elderly, malnourished, and "slow acetylators" are at increased risk for developing the neuropathy. INH inhibits pyridoxal phosphokinase, resulting in pyridoxine deficiency and the neuropathy. Prophylactic administration of pyridoxine 100 mg/d can prevent the neuropathy from developing.

■ ANTIRETROVIRAL AGENTS

The nucleoside analogues zalcitabine (dideoxycytidine or ddC), didanosine (dideoxyinosine or ddI), stavudine (d4T), lamivudine (3TC), and antiretroviral nucleoside reverse transcriptase inhibitor (NRTI) are used to treat HIV infection. One of the major dose-limiting side effects of these medications is a predominantly sensory, length-dependent, symmetrically painful neuropathy (Pattern 2; Table 457-2). Zalcitabine (ddC) is the most extensively studied of the nucleoside analogues, and at doses >0.18 mg/kg per d, it is associated with a subacute onset of severe burning and lancinating pains in the feet and hands. NCS reveal decreased amplitudes of the SNAPs with normal motor studies. The nucleoside analogues inhibit mitochondrial DNA polymerase, which is the suspected pathogenic basis for the neuropathy. Because of a "coasting effect," patients can continue to worsen even 2–3 weeks after stopping the medication. Following dose reduction, improvement in the neuropathy is seen in most patients after several months (mean time ~10 weeks).

■ HEXACARBONS (*n*-HEXANE, METHYL *n*-BUTYL KETONE)/GLUE SNIFFER'S NEUROPATHY

n-Hexane and methyl *n*-butyl ketone are water-insoluble industrial organic solvents that are also present in some glues. Exposure through inhalation, accidentally or intentionally (glue sniffing), or through skin absorption can lead to a profound subacute sensory and motor polyneuropathy (Pattern 2; Table 457-2). NCS demonstrate decreased amplitudes of the SNAPs and CMAPs with slightly slow conduction velocities. Nerve biopsy reveals a loss of myelinated fibers and giant axons that are filled with 10-nm neurofilaments. Hexacarbon exposure leads to covalent cross-linking between axonal neurofilaments that results in their aggregation, impaired axonal transport, swelling of the axons, and eventual axonal degeneration.

■ LEAD

Lead neuropathy is uncommon, but it can be seen in children who accidentally ingest lead-based paints in older buildings and in industrial workers exposed to lead-containing products. The most common presentation of lead poisoning is an encephalopathy; however, symptoms and signs of a primarily motor neuropathy can also occur. The neuropathy is characterized by an insidious and progressive onset of weakness usually beginning in the arms, in particular involving the wrist and finger extensors, resembling a radial neuropathy. Sensation is generally preserved; however, the autonomic nervous system can be affected (Patterns 2, 3, and 10; Table 457-2). Laboratory investigation can reveal a microcytic hypochromic anemia with basophilic stippling of erythrocytes, an elevated serum lead level, and an elevated serum coproporphyrin level. A 24-h urine collection demonstrates elevated levels of lead excretion. The NCS may reveal reduced CMAP amplitudes, while the SNAPs are typically normal. The pathogenic basis may be related to abnormal porphyrin metabolism. The most important principle of management is to remove the source of the exposure. Chelation therapy with calcium disodium ethylene-diaminetetraacetic acid (EDTA), British anti-Lewisite (BAL), and penicillamine also demonstrates variable efficacy.

■ MERCURY

Mercury toxicity may occur as a result of exposure to either organic or inorganic mercurials. Mercury poisoning presents with paresthesias in hands and feet that progress proximally and may involve the face and tongue. Motor weakness can also develop. CNS symptoms often overshadow the neuropathy. EDx shows features of a primarily axonal sensorimotor polyneuropathy. The primary site of neuromuscular pathology appears to be the dorsal root ganglia. The mainstay of treatment is removing the source of exposure.

■ THALLIUM

Thallium can exist in a monovalent or trivalent form and is primarily used as a rodenticide. The toxic neuropathy usually manifests as burning paresthesias of the feet, abdominal pain, and vomiting. Increased thirst, sleep disturbances, and psychotic behavior may be noted. Within the first week, patients develop pigmentation of the hair, an acne-like rash in the malar area of the face, and hyperreflexia. By the second and third weeks, autonomic instability with labile heart rate and blood pressure may be seen. Hyporeflexia and alopecia also occur but may not be evident until the third or fourth week following exposure. With severe intoxication, proximal weakness and involvement of the cranial nerves can occur. Some patients require mechanical ventilation due to respiratory muscle involvement. The lethal dose of thallium is variable, ranging from 8 to 15 mg/kg body weight. Death can result in <48 h following a particularly large dose. NCS demonstrate features of a primarily axonal sensorimotor polyneuropathy. With acute intoxication, potassium ferric ferrocyanide II may be effective in preventing absorption of thallium from the gut. However, there may be no benefit once thallium has been absorbed. Unfortunately, chelating agents are not very efficacious. Adequate diuresis is essential to help eliminate thallium from the body without increasing tissue availability from the serum.

■ ARSENIC

Arsenic is another heavy metal that can cause a toxic sensorimotor polyneuropathy. The neuropathy manifests 5–10 days after ingestion of arsenic and progresses for several weeks, sometimes mimicking GBS. The presenting symptoms are typically an abrupt onset of abdominal discomfort, nausea, vomiting, pain, and diarrhea followed within several days by burning pain in the feet and hands. Examination of the skin can be helpful in the diagnosis as the loss of the superficial epidermal layer results in patchy regions of increased or decreased pigmentation on the skin several weeks after an acute exposure or with chronic low levels of ingestion. Mee's lines, which are transverse lines at the base of the fingernails and toenails, do not become evident until 1 or 2 months after the exposure. Multiple Mee's lines may be seen in patients with long fingernails who have had chronic exposure to arsenic. Mee's lines are not specific for arsenic toxicity as they can also be seen following thallium poisoning. Because arsenic is cleared from blood rapidly, the serum concentration of arsenic is not diagnostically helpful. However, arsenic levels are increased in the urine, hair, and fingernails of patients exposed to arsenic. Anemia with stippling of erythrocytes is common, and occasionally, pancytopenia and aplastic anemia can develop. Increased CSF protein levels without pleocytosis can be seen; this can lead to misdiagnosis as GBS. NCS are usually suggestive of an axonal sensorimotor polyneuropathy; however, demyelinating features can be present. Chelation therapy with BAL has yielded inconsistent results; therefore, it is not generally recommended.

NUTRITIONAL NEUROPATHIES

■ COBALAMIN (VITAMIN B_{12})

Pernicious anemia is the most common cause of cobalamin deficiency. Other causes include dietary avoidance (vegetarians), gastrectomy, gastric bypass surgery, inflammatory bowel disease, pancreatic insufficiency, bacterial overgrowth, and possibly histamine-2 blockers and proton pump inhibitors. An underappreciated cause of cobalamin deficiency is food-cobalamin malabsorption. This typically occurs in older individuals and results from an inability to adequately absorb cobalamin in food protein. No apparent cause of deficiency is identified in a significant number of patients with cobalamin deficiency. The use of nitrous oxide as an anesthetic agent or as a recreational drug can produce acute cobalamin deficiency neuropathy and subacute combined degeneration.

Complaints of numb hands typically appear before lower extremity paresthesias are noted. A preferential large-fiber sensory loss affecting proprioception and vibration with sparing of small-fiber modalities is present; an unsteady gait reflects sensory ataxia. These features, coupled with diffuse hyperreflexia and absent Achilles reflexes, should always focus attention on the possibility of cobalamin deficiency (Patterns 2 and 6; Table 457-2). Optic atrophy and, in severe cases, behavioral changes ranging from mild irritability and forgetfulness to severe dementia and frank psychosis may appear. The full clinical picture of subacute combined degeneration is uncommon. CNS manifestations, especially pyramidal tract signs, may be missing, and in fact, some patients may only exhibit symptoms of peripheral neuropathy.

EDx shows an axonal sensorimotor neuropathy. CNS involvement produces abnormal somatosensory and visual evoked potential latencies. The diagnosis is confirmed by finding reduced serum cobalamin levels. In up to 40% of patients, anemia and macrocytosis are lacking. Serum methylmalonic acid and homocysteine, the metabolites that accumulate when cobalamin-dependent reactions are blocked, are elevated. Antibodies to intrinsic factor are present in ~60% and antiparietal cell antibodies in ~90% of individuals with pernicious anemia.

Cobalamin deficiency can be treated with various regimens of cobalamin. One typical regimen consists of 1000 μg cyanocobalamin IM weekly for 1 month and monthly thereafter. Patients with food cobalamin malabsorption can absorb free cobalamin and therefore can be treated with oral cobalamin supplementation. An oral cobalamin dose of 1000 μg/d should be sufficient. Treatment for cobalamin deficiency usually does not completely reverse the clinical manifestations, and at least 50% of patients exhibit some permanent neurologic deficit.

THIAMINE DEFICIENCY

Thiamine (vitamin B_1) deficiency is an uncommon cause of peripheral neuropathy in developed countries. It is now most often seen as a consequence of chronic alcohol abuse, recurrent vomiting, total parenteral nutrition, and bariatric surgery. Thiamine deficiency polyneuropathy can occur in normal, healthy young adults who do not abuse alcohol but who engage in inappropriately restrictive diets. Thiamine is water-soluble. It is present in most animal and plant tissues, but the greatest sources are unrefined cereal grains, wheat germ, yeast, soybean flour, and pork. Beriberi means "I can't, I can't" in Singhalese, the language of natives of what was once part of the Dutch East Indies (now Sri Lanka). *Dry beriberi* refers to neuropathic symptoms. The term *wet beriberi* is used when cardiac manifestations predominate (in reference to edema). Beriberi was relatively uncommon until the late 1800s when it became widespread among people for whom rice was a dietary mainstay. This epidemic was due to a new technique of processing rice that removed the germ from the rice shaft, rendering the so-called polished rice deficient in thiamine and other essential nutrients.

Symptoms of neuropathy follow prolonged deficiency. These begin with mild sensory loss and/or burning dysesthesias in the toes and feet and aching and cramping in the lower legs. Pain may be the predominant symptom. With progression, patients develop features of a nonspecific generalized polyneuropathy, with distal sensory loss in the feet and hands.

Blood and urine assays for thiamine are not reliable for diagnosis of deficiency. Erythrocyte transketolase activity and the percentage increase in activity (in vitro) following the addition of thiamine pyrophosphate (TPP) may be more accurate and reliable. EDx shows nonspecific findings of an axonal sensorimotor polyneuropathy. When a diagnosis of thiamine deficiency is made or suspected, thiamine replacement should be provided until proper nutrition is restored. Thiamine is usually given intravenously or intramuscularly at a dose of 100 mg/d. Although cardiac manifestations show a striking response to thiamine replacement, neurologic improvement is usually more variable and less dramatic.

VITAMIN E DEFICIENCY

The term *vitamin E* is usually used for α-tocopherol, the most active of the four main types of vitamin E. Because vitamin E is present in animal fat, vegetable oils, and various grains, deficiency is usually due to factors other than insufficient intake. Vitamin E deficiency usually occurs secondary to lipid malabsorption or in uncommon disorders of vitamin E transport. One hereditary disorder is abetalipoproteinemia, a rare autosomal dominant disorder characterized by steatorrhea, pigmentary retinopathy, acanthocytosis, and progressive ataxia. Patients with cystic fibrosis may also have vitamin E deficiency secondary to steatorrhea. There are genetic forms of isolated vitamin E deficiency not associated with lipid malabsorption. Vitamin E deficiency may also occur as a consequence of various cholestatic and hepatobiliary disorders as well as short-bowel syndromes resulting from the surgical treatment of intestinal disorders.

Clinical features may not appear until many years after the onset of deficiency. The onset of symptoms tends to be insidious, and progression is slow. The main clinical features are spinocerebellar ataxia and polyneuropathy, thus resembling Friedreich's ataxia or other spinocerebellar ataxias. Patients manifest progressive ataxia and signs of posterior column dysfunction, such as impaired joint position and vibratory sensation. Because of the polyneuropathy, there is hyporeflexia, but plantar responses may be extensor as a result of the spinal cord involvement (Patterns 2 and 6; Table 457-2). Other neurologic manifestations may include ophthalmoplegia, pigmented retinopathy, night blindness, dysarthria, pseudoathetosis, dystonia, and tremor. Vitamin E deficiency may present as an isolated polyneuropathy, but this is very rare. The yield of checking serum vitamin E levels in patients with isolated polyneuropathy is extremely low, and this test should not be part of routine practice.

Diagnosis is made by measuring α-tocopherol levels in the serum. EDx shows features of an axonal neuropathy. Treatment is replacement with oral vitamin E, but high doses are not needed. For patients with isolated vitamin E deficiency, treatment consists of 1500–6000 IU/d in divided doses.

VITAMIN B_6 DEFICIENCY

Vitamin B_6, or pyridoxine, can produce neuropathic manifestations from both deficiency and toxicity. Vitamin B_6 toxicity was discussed above. Vitamin B_6 deficiency is most commonly seen in patients treated with isoniazid or hydralazine. The polyneuropathy of vitamin B_6 is nonspecific, manifesting as a generalized axonal sensorimotor polyneuropathy. Vitamin B_6 deficiency can be detected by direct assay. Vitamin B_6 supplementation with 50–100 mg/d is suggested for patients being treated with isoniazid or hydralazine. This same dose is appropriate for replacement in cases of nutritional deficiency.

PELLAGRA (NIACIN DEFICIENCY)

Pellagra is produced by deficiency of niacin. Although pellagra may be seen in alcoholics, this disorder has essentially been eradicated in most Western countries by means of enriching bread with niacin. Nevertheless, pellagra continues to be a problem in a number of underdeveloped regions, particularly in Asia and Africa, where corn is the main source of carbohydrate. Neurologic manifestations are variable; abnormalities can develop in the brain and spinal cord as well as peripheral nerves. When peripheral nerves are involved, the neuropathy is usually mild and resembles beriberi. Treatment is with niacin 40–250 mg/d.

COPPER DEFICIENCY

A syndrome that has only recently been described is myeloneuropathy secondary to copper deficiency **(see also Chap. 453)**. Most patients present with lower limb paresthesias, weakness, spasticity, and gait difficulties (Pattern 6; Table 457-2). Large-fiber sensory function is impaired, reflexes are brisk, and plantar responses are extensor. In some cases, light touch and pinprick sensation are affected, and NCS indicate sensorimotor axonal polyneuropathy in addition to myelopathy.

Hematologic abnormalities are a known complication of copper deficiency; these can include microcytic anemia, neutropenia, and occasionally pancytopenia. Because copper is absorbed in the stomach and proximal jejunum, many cases of copper deficiency occur in the setting of prior gastric surgery. Excess zinc is an established cause of copper deficiency. Zinc upregulates enterocyte production of metallothionine, which results in decreased absorption of copper. Excessive dietary zinc supplements or denture cream containing zinc can produce this clinical picture. Other potential causes of copper deficiency include malnutrition, prematurity, total parenteral nutrition, and ingestion of copper-chelating agents.

Following oral or IV copper replacement, some patients show neurologic improvement, but this may take many months or not occur at all. Replacement consists of oral copper sulfate or gluconate 2 mg one to three times a day. If oral copper replacement is not effective, elemental copper in the copper sulfate or copper chloride forms can be given as 2 mg IV daily for 3–5 days, then weekly for 1–2 months until copper levels normalize. Thereafter, oral daily copper therapy can be resumed. In contrast to the neurologic manifestations, most of the hematologic indices normalize in response to copper replacement therapy.

■ NEUROPATHY ASSOCIATED WITH GASTRIC SURGERY

Polyneuropathy may occur following gastric surgery for ulcer, cancer, or weight reduction. This usually occurs in the context of rapid, significant weight loss and recurrent, protracted vomiting. The clinical picture is one of acute or subacute sensory loss and weakness. Neuropathy following weight loss surgery usually occurs in the first several months after surgery. Weight reduction surgical procedures include gastrojejunostomy, gastric stapling, vertical banded gastroplasty, and gastrectomy with Roux-en-Y anastomosis. The initial manifestations are usually numbness and paresthesias in the feet (Pattern 2; Table 457-2). In many cases, no specific nutritional deficiency factor is identified.

Management consists of parenteral vitamin supplementation, especially including thiamine. Improvement has been observed following supplementation, parenteral nutritional support, and reversal of the surgical bypass. The duration and severity of deficits before identification and treatment of neuropathy are important predictors of final outcome.

CRYPTOGENIC (IDIOPATHIC) SENSORY AND SENSORIMOTOR POLYNEUROPATHY

Cryptogenic (idiopathic) sensory and sensorimotor polyneuropathy (CSPN) is a diagnosis of exclusion, established after a careful medical, family, and social history; neurologic examination; and directed laboratory testing. Despite extensive evaluation, the cause of polyneuropathy in as many as 50% of all patients is idiopathic. CSPN should be considered a distinct diagnostic subset of peripheral neuropathy. The onset of CSPN is predominantly in the sixth and seventh decades. Patients complain of distal numbness, tingling, and often burning pain that invariably begins in the feet and may eventually involve the fingers and hands ("burning feet syndrome"). Patients exhibit a distal sensory loss to pinprick, touch, and vibration in the toes and feet, and occasionally in the fingers (Pattern 2; Table 457-2). It is uncommon to see significant proprioception deficits, even though patients may complain of gait unsteadiness. However, tandem gait may be abnormal in a minority of cases. Neither subjective nor objective evidence of weakness is a prominent feature. Most patients have evidence of both large- and small-fiber loss on neurologic examination and EDx. Approximately 10% of patients have only evidence of small-fiber involvement. The ankle muscle stretch reflex is frequently absent, but in cases with predominantly small-fiber loss, this may be preserved. The EDx findings range from isolated SNAP abnormalities (usually with loss of amplitude), to evidence for an axonal sensorimotor neuropathy, to a completely normal study (if primarily small fibers are involved). Therapy primarily involves the control of neuropathic pain (Table 457-6) if present. A large comparative effectiveness study in CSPN showed that the drugs nortriptyline and duloxetine outperformed pregabalin and mexiletine. These drugs should not be used if the patient has only numbness and tingling but no pain.

Although no treatment is available that can reverse an idiopathic distal peripheral neuropathy, the prognosis is good. Progression often does not occur or is minimal, with sensory symptoms and signs progressing proximally up to the knees and elbows. The disorder does not lead to significant motor disability over time. The relatively benign course of this disorder should be explained to patients.

MONONEUROPATHIES/PLEXOPATHIES/RADICULOPATHIES (PATTERN 3; TABLE 457-2)

■ MEDIAN NEUROPATHY

CTS is a compression of the median nerve in the carpal tunnel at the wrist. The median nerve enters the hand through the carpal tunnel by coursing under the transverse carpal ligament. The symptoms of CTS consist of numbness and paresthesias variably in the thumb, index, middle, and half of the ring finger. At times, the paresthesias can include the entire hand and extend into the forearm or upper arm or can be isolated to one or two fingers. Pain is another common symptom and can be located in the hand and forearm and, at times, in the proximal arm. CTS is common and often misdiagnosed as thoracic outlet syndrome. The signs of CTS are decreased sensation in the median nerve distribution; reproduction of the sensation of tingling when a percussion hammer is tapped over the wrist (Tinel sign) or the wrist is flexed for 30–60 s (Phalen sign); and weakness of thumb opposition and abduction. EDx is extremely sensitive and shows slowing of sensory and, to a lesser extent, motor median potentials across the wrist. Ultrasound can show focal swelling of the median nerve at the wrist. Treatment options consist of avoidance of precipitating activities; control of underlying systemic-associated conditions if present; nonsteroidal anti-inflammatory medications; neutral (volar) position wrist splints, especially for night use; glucocorticoid/anesthetic injection into the carpal tunnel; and surgical decompression by dividing the transverse carpal ligament. The surgical option should be considered if there is a poor response to nonsurgical treatments; if there is thenar muscle atrophy and/or weakness; and if there are significant denervation potentials on EMG.

Other proximal median neuropathies are very uncommon and include the pronator teres syndrome and anterior interosseous neuropathy. These often occur as a partial form of brachial plexitis.

■ ULNAR NEUROPATHY AT THE ELBOW—"CUBITAL TUNNEL SYNDROME"

The ulnar nerve passes through the condylar groove between the medial epicondyle and the olecranon. Symptoms consist of paresthesias, tingling, and numbness in the medial hand and half of the fourth and the entire fifth fingers, pain at the elbow or forearm, and weakness. Signs consist of decreased sensation in an ulnar distribution, Tinel's sign at the elbow, and weakness and atrophy of ulnar-innervated hand muscles. The Froment sign indicates thumb adductor weakness and consists of flexion of the thumb at the interphalangeal joint when attempting to oppose the thumb against the lateral border of the second digit. EDx may show slowing of ulnar motor NCV across the elbow with prolonged ulnar sensory latencies. Ultrasound can show swelling of the ulnar nerve around the elbow as well. Treatment consists of avoiding aggravating factors, using elbow pads, and surgery to decompress the nerve in the cubital tunnel. Ulnar neuropathies can also rarely occur at the wrist in the ulnar (Guyon) canal or in the hand, usually after trauma.

■ RADIAL NEUROPATHY

The radial nerve winds around the proximal humerus in the spiral groove and proceeds down the lateral arm and enters the forearm, dividing into the posterior interosseous nerve and superficial nerve. The symptoms and signs consist of wrist drop; finger extension weakness; thumb abduction weakness; and sensory loss in the dorsal web between the thumb and index finger. Triceps and brachioradialis strength is often normal, and triceps reflex is often intact. Most cases of radial neuropathy are transient compressive (neuropraxic) injuries that recover spontaneously in 6–8 weeks. If there has been prolonged compression and severe axonal damage, it may take several months to recover. Treatment consists of cock-up wrist and finger splints, avoiding further compression, and physical therapy to avoid flexion contracture. If there is no improvement in 2–3 weeks, an EDx study is recommended to confirm the clinical diagnosis and determine the degree of severity.

LATERAL FEMORAL CUTANEOUS NEUROPATHY (MERALGIA PARESTHETICA)

The lateral femoral cutaneous nerve arises from the upper lumbar plexus (spinal levels L2/3), crosses through the inguinal ligament near its attachment to the iliac bone, and supplies sensation to the anterior lateral thigh. The neuropathy affecting this nerve is also known as meralgia paresthetica. Symptoms and signs consist of paresthesias, numbness, and occasionally pain in the lateral thigh. Symptoms are increased by standing or walking and are relieved by sitting. There is normal strength, and knee reflexes are intact. The diagnosis is clinical, and further tests usually are not performed. EDx is only needed to rule out lumbar plexopathy, radiculopathy, or femoral neuropathy. If the symptoms and signs are classic, EMG is not necessary. Symptoms often resolve spontaneously over weeks or months, but the patient may be left with permanent numbness. Treatment consists of weight loss and avoiding tight belts. Analgesics in the form of a lidocaine patch, nonsteroidal agents, and occasionally medications for neuropathic pain can be used (Table 457-6). Rarely, locally injecting the nerve with an anesthetic can be tried. There is no role for surgery.

FEMORAL NEUROPATHY

Femoral neuropathies can arise as complications of retroperitoneal hematoma, lithotomy positioning, hip arthroplasty or dislocation, iliac artery occlusion, femoral arterial procedures, infiltration by hematogenous malignancy, penetrating groin trauma, pelvic surgery including hysterectomy and renal transplantation, and diabetes (a partial form of lumbosacral diabetic plexopathy); some cases are idiopathic. Patients with femoral neuropathy have difficulty extending their knee and flexing the hip. Sensory symptoms occurring either on the anterior thigh and/or medial leg occur in only half of reported cases. A prominent painful component is the exception rather than the rule, may be delayed, and is often self-limited in nature. The quadriceps (patellar) reflex is diminished.

SCIATIC NEUROPATHY

Sciatic neuropathies commonly complicate hip arthroplasty, pelvic procedures in which patients are placed in a prolonged lithotomy position, trauma, hematomas, tumor infiltration, and vasculitis. In addition, many sciatic neuropathies are idiopathic. Weakness may involve all motions of the ankles and toes as well as flexion of the leg at the knee; abduction and extension of the thigh at the hip are spared. Sensory loss occurs in the entire foot and the distal lateral leg. The ankle jerk and, on occasion, the internal hamstring reflex are diminished or more typically absent on the affected side. The peroneal subdivision of the sciatic nerve is typically involved disproportionately to the tibial counterpart. Thus, patients may have only ankle dorsiflexion and eversion weakness with sparing of knee flexion, ankle inversion, and plantar flexion; these features can lead to misdiagnosis of a common peroneal neuropathy.

PERONEAL NEUROPATHY

The sciatic nerve divides at the distal femur into the tibial and peroneal nerve. The common peroneal nerve passes posterior and laterally around the fibular head, under the fibular tunnel. It then divides into the superficial peroneal nerve, which supplies the ankle evertor muscles and sensation over the anterolateral distal leg and dorsum of the foot, and the deep peroneal nerve, which supplies ankle dorsiflexors and toe extensor muscles and a small area of sensation dorsally in the area of the first and second toes.

Symptoms and signs consist of foot drop (ankle dorsiflexion, toe extension, and ankle eversion weakness) and variable sensory loss, which may involve the superficial and deep peroneal pattern. There is usually no pain. Onset may be on awakening in the morning. Peroneal neuropathy needs to be distinguished from L5 radiculopathy. In L5 radiculopathy, ankle invertors and evertors are weak and needle EMG reveals denervation. EDx can help localize the lesion. Peroneal motor conduction velocity shows slowing and amplitude drop across the fibular head. Management consists of rapid weight loss and avoiding leg crossing. Foot drop is treated with an ankle brace. A knee pad can be worn over the lateral knee to avoid further compression. Most cases spontaneously resolve over weeks or months.

RADICULOPATHIES

Radiculopathies are most often due to compression from degenerative joint disease and herniated disks, but there are a number of unusual etiologies (Table 457-9). **Degenerative spine disease affects a number of different structures, which narrow the diameter of the neural foramen or canal of the spinal column and compromise nerve root integrity; these are discussed in detail in Chaps. 18 and 19.**

PLEXOPATHIES (PATTERN 4; TABLE 457-2)

BRACHIAL PLEXUS

The brachial plexus is composed of three trunks (upper, middle, and lower), with two divisions (anterior and posterior) per trunk (Fig. 457-2). Subsequently, the trunks divide into three cords (medial, lateral, and posterior), and from these, arise the multiple terminal nerves innervating the arm. The anterior primary rami of C5 and C6 fuse to form the upper trunk; the anterior primary ramus of C7 continues as the middle trunk, while the anterior rami of C8 and T1 join to form the lower trunk. There are several disorders commonly associated with brachial plexopathy.

Immune-Mediated Brachial Plexus Neuropathy Immune-mediated brachial plexus neuropathy (IBPN) goes by various terms, including *acute brachial plexitis, neuralgic amyotrophy*, and *Parsonage-Turner syndrome*. IBPN usually presents with an acute onset of severe pain in the shoulder region. The intense pain usually lasts several days to a few weeks, but a dull ache can persist. Individuals who are affected may not appreciate weakness of the arm early in the course because the pain limits movement. However, as the pain dissipates, weakness and often sensory loss are appreciated. Attacks can occasionally recur.

Clinical findings are dependent on the distribution of involvement (e.g., specific trunk, divisions, cords, or terminal nerves). The most

TABLE 457-9 Causes of Radiculopathy

- Herniated nucleus pulposus
- Degenerative joint disease
- Rheumatoid arthritis
- Trauma
- Vertebral body compression fracture
- Pott's disease (tuberculosis)
- Compression by extradural mass (e.g., meningioma, metastatic tumor, hematoma, abscess)
- Primary nerve tumor (e.g., neurofibroma, schwannoma, neurinoma)
- Carcinomatous meningitis
- Perineurial spread of tumor (e.g., prostate cancer)
- Acute inflammatory demyelinating polyradiculopathy
- Chronic inflammatory demyelinating polyradiculopathy
- Sarcoidosis
- Amyloidoma
- Diabetic radiculopathy
- Infection (Lyme disease, herpes zoster, HIV, cytomegalovirus, syphilis, schistosomiasis, *Strongyloides*)
- Arachnoiditis (e.g., postsurgical)
- Radiation

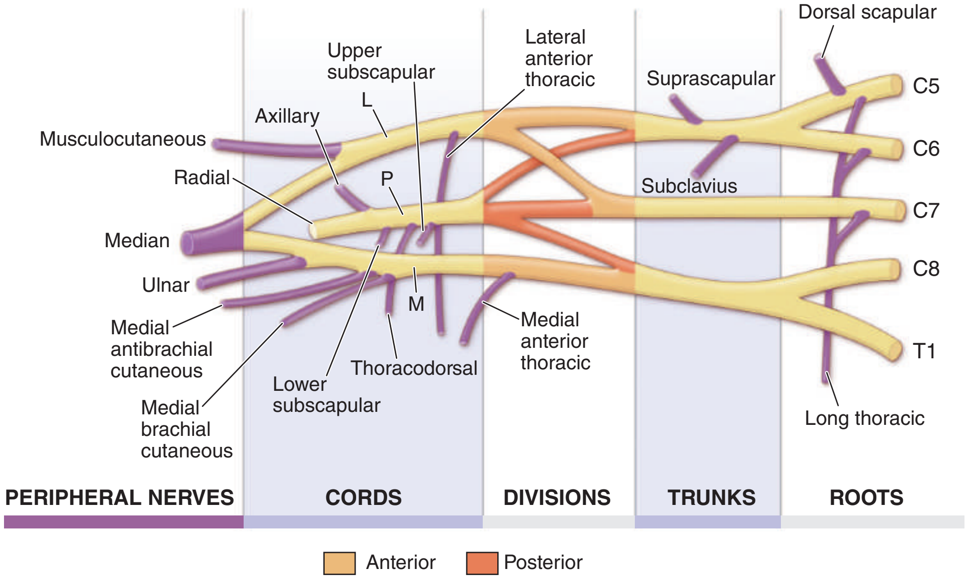

FIGURE 457-2 Brachial plexus anatomy. L, lateral; M, medial; P, posterior. *(Reproduced with permission J Goodgold: Anatomical Correlates of Clinical Electromyography. Baltimore, Williams and Wilkins, 1974.)*

common pattern of IBPN involves the upper trunk or a single or multiple mononeuropathies primarily involving the suprascapular, long thoracic, or axillary nerves. Additionally, the phrenic and anterior interosseous nerves may be concomitantly affected. Any of these nerves may also be affected in isolation. EDx is useful to confirm and localize the site(s) of involvement. Empirical treatment of severe pain with glucocorticoids is often used in the acute period.

Brachial Plexopathies Associated with Neoplasms Neoplasms involving the brachial plexus may be primary nerve tumors, local cancers expanding into the plexus (e.g., Pancoast lung tumor or lymphoma), and metastatic tumors. Primary brachial plexus tumors are less common than the secondary tumors and include schwannomas, neurinomas, and neurofibromas. Secondary tumors affecting the brachial plexus are more common and are always malignant. These may arise from local tumors, expanding into the plexus. For example, a Pancoast tumor of the upper lobe of the lung may invade or compress the lower trunk, whereas a primary lymphoma arising from the cervical or axillary lymph nodes may also infiltrate the plexus. Pancoast tumors typically present as an insidious onset of pain in the upper arm, sensory disturbance in the medial aspect of the forearm and hand, and weakness and atrophy of the intrinsic hand muscles along with an ipsilateral Horner's syndrome. Chest computed tomography (CT) scans or MRI can demonstrate extension of the tumor into the plexus. Metastatic involvement of the brachial plexus may occur with spread of breast cancer into the axillary lymph nodes and local spread into the nearby nerves.

Perioperative Plexopathies (Median Sternotomy) The most common surgical procedures associated with brachial plexopathy as a complication are those that involve median sternotomies (e.g., open-heart surgeries and thoracotomies). Brachial plexopathies occur in as many as 5% of patients following a median sternotomy and typically affect the lower trunk. Thus, individuals manifest with sensory disturbance affecting the medial aspect of forearm and hand along with weakness of the intrinsic hand muscles. The mechanism is related to the stretch of the lower trunk, so most individuals who are affected recover within a few months.

Lumbosacral Plexus The lumbar plexus arises from the ventral primary rami of the first to the fourth lumbar spinal nerves (Fig. 457-3). These nerves pass downward and laterally from the vertebral column within the psoas major muscle. The femoral nerve derives from the dorsal branches of the second to the fourth lumbar ventral rami. The obturator nerve arises from the ventral branches of the same lumbar rami. The lumbar plexus communicates with the sacral plexus by the lumbosacral trunk, which contains some fibers from the fourth and all of the fibers from the fifth lumbar ventral rami (Fig. 457-4).

The sacral plexus is the part of the lumbosacral plexus that is formed by the union of the lumbosacral trunk with the ventral rami of the first to fourth sacral nerves. The plexus lies on the posterior and posterolateral wall of the pelvis with its components converging toward the sciatic notch. The lateral trunk of the sciatic nerve (which forms the common peroneal nerve) arises from the union of the dorsal branches of the lumbosacral trunk (L4, L5) and the dorsal branches of the S1 and S2 spinal nerve ventral rami. The medial trunk of the sciatic nerve (which forms the tibial nerve) derives from the ventral branches of the same ventral rami (L4-S2).

LUMBOSACRAL PLEXOPATHIES

Plexopathies are typically recognized when motor, sensory, and if applicable, reflex deficits occur in multiple nerve and segmental distributions confined to one extremity. If localization within the lumbosacral plexus can be accomplished, designation as a lumbar plexopathy, a sacral plexopathy, a lumbosacral trunk lesion, or a panplexopathy is the best localization that can be expected. Although lumbar plexopathies may be bilateral, usually occurring in a stepwise and chronologically dissociated manner, sacral plexopathies are more likely to behave in this manner due to their closer anatomic proximity. The differential diagnosis of plexopathy includes disorders of the conus medullaris and cauda equina (polyradiculopathy). If there is a paucity of pain and sensory involvement, motor neuron disease should be considered as well.

The causes of lumbosacral plexopathies are listed in Table 457-10. Diabetic radiculopathy (discussed above) is a fairly common cause of painful leg weakness. Lumbosacral plexopathies are a well-recognized complication of retroperitoneal hemorrhage. Various primary and metastatic malignancies can affect the lumbosacral plexus as well; these include carcinoma of the cervix, endometrium, and ovary; osteosarcoma; testicular cancer; MM; lymphoma; acute myelogenous leukemia; colon cancer; squamous cell carcinoma of the rectum; adenocarcinoma of unknown origin; and intraneural spread of prostate cancer.

RECURRENT NEOPLASTIC DISEASE OR RADIATION-INDUCED PLEXOPATHY

The treatment for various malignancies is often radiation therapy, the field of which may include parts of the brachial plexus. It can be difficult in such situations to determine if a new brachial or lumbosacral plexopathy is related to tumor within the plexus or from

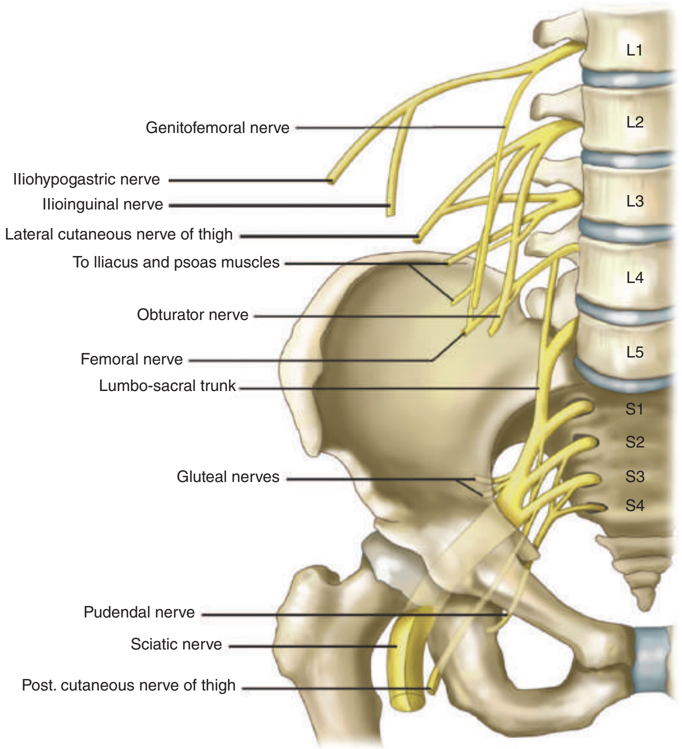
L1
L2
L3
L4
L5
S1
S2
S3
S4
Genitofemoral nerve
Iliohypogastric nerve
Ilioinguinal nerve
Lateral cutaneous nerve of thigh
To Iliacus and psoas muscles
Obturator nerve
Femoral nerve
Lumbo-sacral trunk
Gluteal nerves
Pudendal nerve
Sciatic nerve
Post. cutaneous nerve of thigh

FIGURE 457-3 Lumbosacral plexus. *(Reproduced with permission from AA Amato, JA Russell (eds): Neuromuscular Disorders, 2nd ed. New York: McGraw-Hill Education; 2016.)*

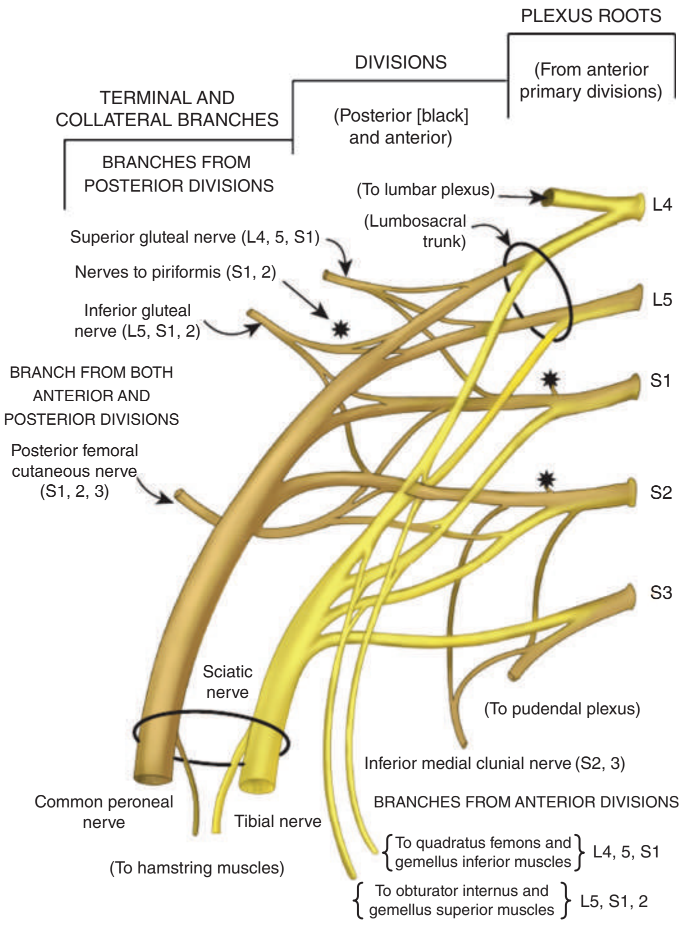
PLEXUS ROOTS
(From anterior primary divisions)
DIVISIONS
(Posterior [black] and anterior)
TERMINAL AND COLLATERAL BRANCHES
BRANCHES FROM POSTERIOR DIVISIONS
(To lumbar plexus)
L4
(Lumbosacral trunk)
Superior gluteal nerve (L4, 5, S1)
Nerves to piriformis (S1, 2)
L5
Inferior gluteal nerve (L5, S1, 2)
BRANCH FROM BOTH ANTERIOR AND POSTERIOR DIVISIONS
S1
Posterior femoral cutaneous nerve (S1, 2, 3)
S2
S3
Sciatic nerve
(To pudendal plexus)
Inferior medial clunial nerve (S2, 3)
Common peroneal nerve
Tibial nerve
BRANCHES FROM ANTERIOR DIVISIONS
(To hamstring muscles)
To quadratus femons and gemellus inferior muscles
L4, 5, S1
To obturator internus and gemellus superior muscles
L5, S1, 2

FIGURE 457-4 Lumbosacral trunk sacral plexus and sciatic nerve. *(Reproduced with permission from AA Amato, JA Russell (eds): Neuromuscular Disorders, 2nd ed. New York: McGraw-Hill Education; 2016.)*

TABLE 457-10 Lumbosacral Plexopathies: Etiologies

- Retroperitoneal hematoma
- Psoas abscess
- Malignant neoplasm
- Benign neoplasm
- Radiation
- Amyloid
- Diabetic radiculoplexus neuropathy
- Idiopathic radiculoplexus neuropathy
- Sarcoidosis
- Aortic occlusion/surgery
- Lithotomy positioning
- Hip arthroplasty
- Pelvic fracture
- Obstetric injury

radiation-induced nerve damage. Radiation can be associated with microvascular abnormalities and fibrosis of surrounding tissues, which can damage the axons and the Schwann cells. Radiation-induced plexopathy can develop months or years following therapy and is dose dependent.

Tumor invasion is usually painful and more commonly affects the lower trunk, whereas radiation injury is often painless and affects the upper trunk. Imaging studies such as MRI and CT scans are useful but can be misleading, especially when there is small microscopic invasion of the plexus. EMG can be informative if myokymic discharges are appreciated, as this finding strongly suggests radiation-induced damage.

EVALUATION AND TREATMENT OF PLEXOPATHIES

Most patients with plexopathies will undergo both imaging with MRI and EDx evaluations. Severe pain from acute idiopathic lumbosacral plexopathy may respond to a short course of glucocorticoids.

FURTHER READING

Amato AA, Ropper AH: Sensory ganglionopathy. N Engl J Med 383:1657, 2020.

Amato AA, Russell J: *Neuromuscular Disorders*, 2nd ed. New York, McGraw-Hill, 2016.

Barohn RJ, Amato AA: Pattern-recognition approach to neuropathy and neuronopathy. Neurol Clin 31:343, 2013.

Barohn RJ et al: Patient Assisted Intervention for Neuropathy: Comparison of Treatment in Real Life Situations (PAIN-CONTRoLS) Bayesian adaptive comparative effectiveness randomized trial. JAMA Neurol 78:68, 2021.

Cortese A et al: Biallelic mutations in SORD cause a common and potentially treatable hereditary neuropathy with implications for diabetes. Nat Genet 52:473, 2020. [Published correction appears in Nat Genet 52:640, 2020.]

Cortese A et al: Cerebellar ataxia, neuropathy and vestibular areflexia syndrome (CANVAS): Genetic and clinical aspects. Pract Neurol 22:14, 2022.

Elafros MA et al: Towards prevention of diabetic peripheral neuropathy: Clinical presentation, pathogenesis, and new treatments. Lancet Neurol 21:922, 2022.

Hobson-Webb LD, Juel VC: Common entrapment neuropathies. Continuum (Minneap Minn) 23:487, 2017.

Ioannou A et al: RNA Targeting and gene editing strategies for transthyretin amyloidosis. BioDrugs 37:127, 2023.

Jin PH, Shin SC: Neuropathy of connective tissue diseases and other systemic diseases. Semin Neurol 39:651, 2019.

Klein CJ: Charcot-Marie-Tooth disease and other hereditary neuropathies. Continuum (Minneap Minn) 26:1224, 2020.

458

Guillain-Barré Syndrome and Other Immune-Mediated Neuropathies

Stephen L. Hauser, Anthony A. Amato

GUILLAIN-BARRÉ SYNDROME

Guillain-Barré syndrome (GBS) is an acute, frequently severe, and fulminant polyradiculoneuropathy that is autoimmune in nature. It occurs year-round at a rate of between 10 and 20 cases per million annually; in the United States, ~5000–6000 cases occur per year. Males are at slightly higher risk for GBS than females, and in Western countries, adults are more frequently affected than children.

Clinical Manifestations GBS manifests as a rapidly evolving areflexic motor paralysis with or without sensory disturbance. The usual pattern is an ascending paralysis that may be first noticed as rubbery legs. Weakness typically evolves over hours to a few days and is frequently accompanied by tingling dysesthesias in the extremities. The legs are usually more affected than the arms, and facial diparesis is present in 50% of affected individuals. The lower cranial nerves are also frequently involved, causing bulbar weakness with difficulty handling secretions and maintaining an airway; the diagnosis in these patients may initially be mistaken for brainstem ischemia. Pain in the neck, shoulder, back, or diffusely over the spine is also common in the early stages of GBS, occurring in ~50% of patients. Most patients require hospitalization, and in different series, up to 30% require ventilatory assistance at some time during the illness. The need for mechanical ventilation is associated with more severe weakness on admission, a rapid tempo of progression, and the presence of facial and/or bulbar weakness during the first week of symptoms. Fever and constitutional symptoms are absent at the onset and, if present, cast doubt on the diagnosis. Deep tendon reflexes attenuate or disappear within the first few days of onset. Cutaneous sensory deficits (e.g., loss of pain and temperature sensation) are usually relatively mild, but functions subserved by large sensory fibers, such as deep tendon reflexes and proprioception, are more severely affected. Bladder dysfunction may occur in severe cases but is usually transient. If bladder dysfunction is a prominent feature and comes early in the course or there is a sensory level on examination, diagnostic possibilities other than GBS should be considered, particularly spinal cord disease (Chap. 453). Once clinical worsening stops and the patient reaches a plateau (almost always within 4 weeks of onset), further progression is unlikely.

Autonomic involvement is common and may occur even in patients whose GBS is otherwise mild. The usual manifestations are loss of vasomotor control with wide fluctuations in blood pressure, postural hypotension, and cardiac dysrhythmias. These features require close monitoring and management and can be fatal. Pain is another common feature of GBS; in addition to the acute pain described above, a deep aching pain may be present in weakened muscles that patients liken to having overexercised the previous day. Other pains in GBS include dysesthetic pain in the extremities as a manifestation of sensory nerve fiber involvement. These pains are self-limited and often respond to standard analgesics (Chap. 14).

Several subtypes of GBS are recognized, as determined primarily by electrodiagnostic (EDx) and pathologic distinctions (Table 458-1). The most common variant is acute inflammatory demyelinating polyneuropathy (AIDP). Additionally, there are two "axonal" or "nodal/paranodal" variants, which are often clinically severe: the acute motor axonal neuropathy (AMAN) and acute motor sensory axonal neuropathy (AMSAN) subtypes. In addition, a range of limited or regional GBS syndromes are also encountered. Notable among these is the Miller Fisher syndrome (MFS), which presents as rapidly evolving ataxia and areflexia of limbs without weakness, and ophthalmoplegia, often with pupillary paralysis. The MFS variant accounts for ~5% of all cases and is strongly associated with antibodies to the ganglioside GQ1b (see "Immunopathogenesis," below). Other regional variants of GBS include (1) pure sensory forms; (2) ophthalmoplegia with anti-GQ1b antibodies as part of severe motor-sensory GBS; (3) GBS with severe bulbar and facial paralysis, sometimes associated with antecedent cytomegalovirus (CMV) infection and anti-GM2 antibodies; and (4) acute pandysautonomia (Chap. 451).

Antecedent Events Approximately 70% of cases of GBS occur 1–3 weeks after an acute infectious process, usually respiratory or gastrointestinal. Culture and seroepidemiologic techniques show that 20–30% of all cases occurring in North America, Europe, and Australia are preceded by infection or reinfection with *Campylobacter jejuni*. A similar proportion is preceded by a human herpes virus infection, often CMV or Epstein-Barr virus. Other viruses (e.g., HIV, hepatitis E, Zika) and also *Mycoplasma pneumoniae* have been identified as agents involved in antecedent infections. Cases of GBS have been reported in association with SARS-CoV-2 infection during the COVID-19 pandemic, but a causal relationship has not been established.

C. jejuni has also been implicated in summer outbreaks of AMAN among children and young adults exposed to chickens in rural China, as has infection by Zika virus in the increased incidence of GBS in Brazil and other endemic regions.

Recent immunizations have also been implicated in GBS. The swine influenza vaccine, administered widely in the United States in 1976, is the most notable example. Influenza vaccines in use from 1992 to 1994, however, resulted in only one additional case of GBS per million persons vaccinated, and the more recent seasonal influenza vaccines appear to confer a GBS risk of <1 per million. Epidemiologic studies looking at H1N1 vaccination demonstrated at most only a slight increased risk of GBS. There appears to be an increased risk of GBS with SARS-CoV-2 vaccines using adenovirus vectors, but not the messenger RNA vaccines. Meningococcal vaccinations (Menactra) do not appear to carry an increased risk. Older-type rabies vaccine, prepared in nervous system tissue, is implicated as a trigger of GBS in developing countries where it is still used; the mechanism is presumably immunization against neural antigens.

TABLE 458-1 Subtypes of Guillain-Barré Syndrome (GBS)

SUBTYPE	FEATURES	ELECTRODIAGNOSIS	PATHOLOGY
Acute inflammatory demyelinating polyneuropathy (AIDP)	Adults affected more than children; 90% of cases in Western world; recovery rapid; anti-GM1 antibodies (<50%)	Demyelinating	First attack on Schwann cell surface; widespread myelin damage, macrophage activation, and lymphocytic infiltration; variable secondary axonal damage
Acute motor axonal neuropathy (AMAN)	Children and young adults; prevalent in China and Mexico; may be seasonal; recovery rapid; anti-GD1a antibodies	Axonal	First attack at motor nodes of Ranvier; macrophage activation, few lymphocytes, frequent periaxonal macrophages; extent of axonal damage highly variable
Acute motor sensory axonal neuropathy (AMSAN)	Mostly adults; uncommon; recovery slow, often incomplete; closely related to AMAN	Axonal	Same as AMAN, but also affects sensory nerves and roots; axonal damage usually severe
Miller Fisher syndrome (MFS)	Adults and children; ophthalmoplegia, ataxia, and areflexia; anti-GQ1b antibodies (90%)	Axonal or demyelinating	Few cases examined; resembles AIDP

GBS also occurs more frequently than can be attributed to chance alone in patients with lymphoma (including Hodgkin's disease), in HIV-seropositive individuals, and in patients with systemic lupus erythematosus (SLE) and possibly Sjogren's syndrome. GBS, other inflammatory neuropathies, and myositis can also occur as a complication of immune checkpoint inhibitors used to treat various cancers.

Immunopathogenesis Several lines of evidence support an autoimmune basis for AIDP, the most common and best-studied type of GBS; the concept extends to all of the subtypes of GBS (Table 458-1).

It is likely that both cellular and humoral immune mechanisms contribute to tissue damage in AIDP. AIDP is also closely analogous to an experimental T cell–mediated immunopathy designated *experimental allergic neuritis* (EAN). EAN is induced in laboratory animals by immune sensitization against protein fragments derived from peripheral nerve proteins and, in particular, against the P2 protein. Based on analogy to EAN, it was initially thought that AIDP was also likely to have a T cell–mediated pathogenesis, and consistent with this concept, autoreactive T cells against several peripheral myelin proteins have recently been identified in peripheral blood, cerebrospinal fluid (CSF), and infiltrating nerves from AIDP patients. However, abundant data also indicate that autoantibodies directed against T cell–independent nonprotein determinants may be the central mediators in many cases. Involvement of the humoral arm of the immune system in AIDP is supported by the demonstration of terminal complement complex on Schwann cells in autopsy series and induction of complement-dependent demyelination and conduction block following injection of serum from patients with GBS into nerves of animals. In AMAN, there is deposition of IgG and complement activation products on the nodal and internodal axolemma of motor fibers.

Circumstantial evidence suggests that all GBS results from immune responses to nonself antigens (infectious agents, vaccines) that misdirect to host nerve tissue through a resemblance-of-epitope (molecular mimicry) mechanism **(Fig. 458-1)**. The neural targets are likely to be glycoconjugates, specifically gangliosides **(Table 458-2; Fig. 458-2)**. Gangliosides are complex glycosphingolipids that contain one or more sialic acid residues; various gangliosides participate in cell-cell interactions (including those between axons and glia), modulation of receptors, and regulation of growth. They are typically exposed on the plasma membrane of cells, rendering them susceptible to an antibody-mediated attack. Gangliosides and other glycoconjugates are present in large quantity in human nervous tissues and in key sites, such as nodes of Ranvier. Antiganglioside antibodies, most frequently to GM1, are common in GBS (20–50% of cases), particularly in AMAN and AMSAN, and in those cases, they are preceded by *C. jejuni* infection. Some AIDP autoantibodies may recognize glycolipid heterocomplexes, rather than single species, present on cell membranes. Furthermore, isolates of *C. jejuni* from stool cultures of patients with GBS have surface glycolipid structures that antigenically cross react with gangliosides, including GM1, concentrated in human nerves. Sialic acid residues from pathogenic *C. jejuni* strains can also trigger activation of dendritic cells via signaling through Toll-like receptor 4 (TLR4), promoting B-cell differentiation and further amplifying humoral autoimmunity. Another line of evidence implicating humoral autoimmunity is derived from cases of GBS that followed intravenous administration of bovine brain gangliosides for treatment of various neuropathies; 5–15 days after injection, some recipients developed AMAN with high titers of anti-GM1 antibodies that recognized epitopes at nodes of Ranvier and motor endplates. Experimentally, anti-GM1 antibodies can trigger complement-mediated injury at paranodal axon-glial junctions, disrupting the clustering of sodium channels and likely contributing to conduction block (see "Pathophysiology," below).

Anti-GQ1b IgG antibodies are found in >90% of patients with MFS (Table 458-2; Fig. 458-2), and titers of IgG are highest early in

Subtypes and variants	IgG autoantibodies to
Guillain-Barré syndrome	
Acute inflammatory demyelinating polyneuropathy	None
Facial variant: Facial diplegia and paresthesia	None
Acute motor axonal neuropathy	GM1, GD1a
More and less extensive forms	
Acute motor-sensory axonal neuropathy	GM1, GD1a
Acute motor-conduction-block neuropathy	GM1, GD1a
Pharyngeal-cervical-brachial weakness	GT1a>GQ1b>>GD1a
Miller Fisher syndrome	GQ1b, GT1a
Incomplete forms	
Acute ophthalmoparesis (without ataxia)	GQ1b, GT1a
Acute ataxic neuropathy (without ophthalmoplegia)	GQ1b, GT1a
CNS variant: Bickerstaff's brainstem encephalitis	GQ1b, GT1a

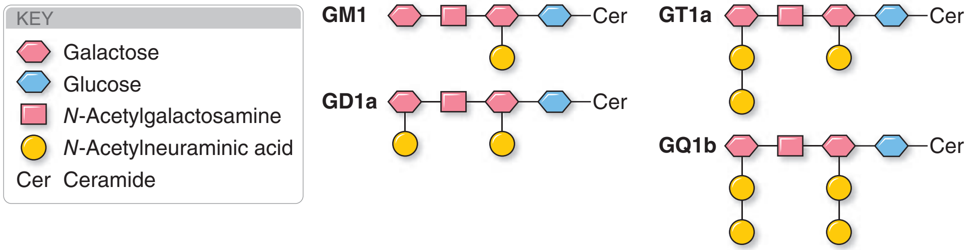

FIGURE 458-1 Spectrum of disorders in Guillain-Barré syndrome and associated antiganglioside antibodies. IgG autoantibodies against GM1 or GD1a are strongly associated with acute motor axonal neuropathy (AMAN), as well as the more extensive acute motor-sensory axonal neuropathy (AMSAN), and the less extensive acute motor-conduction-block neuropathy. IgG anti-GQ1b antibodies, which cross-react with GT1a, are strongly associated with Miller Fisher syndrome, its incomplete forms (acute ophthalmoparesis [without ataxia] and acute ataxic neuropathy [without ophthalmoplegia]), and its more extensive form, Bickerstaff's brainstem encephalitis. Pharyngeal-cervical-brachial weakness is categorized as a localized form of acute motor axonal neuropathy or an extensive form of Miller Fisher syndrome. Half of patients with pharyngeal-cervical-brachial weakness have IgG anti-GT1a antibodies, which often cross-react with GQ1b. IgG anti-GD1a antibodies have also been detected in a small percentage of patients. The anti-GQ1b antibody syndrome includes Miller Fisher syndrome, acute ophthalmoparesis, acute ataxic neuropathy, Bickerstaff's brainstem encephalitis, and pharyngeal-cervical-brachial weakness. The presence of clinical overlap also indicates that Miller Fisher syndrome is part of a continuous spectrum with these conditions. Patients who have had Guillain-Barré syndrome overlapped with Miller Fisher syndrome or with its related conditions have IgG antibodies against GM1 or GD1a as well as against GQ1b or GT1a, supporting a link between AMAN and anti-GQ1b syndrome. *(From N Yuki, H-P Hartung: Guillain-Barré syndrome. N Engl J Med 366:2294, 2012. Copyright © 2012 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

TABLE 458-2 Principal Antiglycolipid Antibodies Implicated in Immune Neuropathies

CLINICAL PRESENTATION	ANTIBODY TARGET	USUAL ISOTYPE
Acute Immune Neuropathies (Guillain-Barré Syndrome)		
Acute inflammatory demyelinating polyneuropathy (AIDP)	No clear patterns GM1 most common	IgG (polyclonal)
Acute motor axonal neuropathy (AMAN)	GD1a, GM1, GM1b, GalNAc–GD1a (<50% for any)	IgG (polyclonal)
Miller Fisher syndrome (MFS)	GQ1b (>90%)	IgG (polyclonal)
Acute pharyngeal cervicobrachial neuropathy (APCBN)	GT1a (? most)	IgG (polyclonal)
Chronic Immune Neuropathies		
Chronic inflammatory demyelinating polyneuropathy (CIDP) (75%)	Rarely to P0, myelin P2 protein, or PMP22	IgG, IgA
CIDP-M (MGUS associated) (25%)	Neural binding sites	IgG, IgA (monoclonal)
Anti-MAG neuropathy	SGPG, SGLPG (on MAG) (50%) Uncertain (50%)	IgM (monoclonal) IgM (monoclonal)
Nodal/paranodal neuropathies	Approximately 10% to CNTN1 or NF155, less often to NF140/186 and Caspr1	IgG4 with CNTN1, NF155, NF140/186, Caspr1 Rare IgM with NF155
Multifocal motor neuropathy (MMN)	GM1, GalNAc–GD1a, others (25–50%)	IgM (polyclonal, monoclonal)
Chronic sensory ataxic neuropathy	GD1b, GQ1b, and other b-series gangliosides	IgM (monoclonal)

Abbreviations: CIDP-M, CIDP with a monoclonal gammopathy; Caspr1, contactin associated protein-1; CNTN1, contactin-1; MAG, myelin-associated glycoprotein; MGUS, monoclonal gammopathy of undetermined significance; NF140/186, neurofascin 140/186; NF155, neurofascin 155; SGPG, sulfoglucuronyl paragloboside; SGLPG, sulfoglucuronyl lactosaminyl paragloboside.

Source: Modified with permission from HJ Willison, N Yuki: Peripheral neuropathies and anti-glycolipid antibodies. Brain 125:2591, 2002.

the course. Anti-GQ1b antibodies are not found in other forms of GBS unless there is extraocular motor nerve involvement. A possible explanation for this association is that extraocular motor nerves are enriched in GQ1b gangliosides in comparison to limb nerves. In addition, a monoclonal anti-GQ1b antibody raised against *C. jejuni* isolated from a patient with MFS blocked neuromuscular transmission experimentally.

Taken together, these observations provide strong but still inconclusive evidence that autoantibodies play an important pathogenic role in GBS. Although antiganglioside antibodies have been studied most intensively, other antigenic targets may also be important. Proof that these antibodies are pathogenic requires that they be capable of mediating disease following direct passive transfer to naïve hosts; this has not yet been demonstrated, although one case of possible maternal-fetal transplacental transfer of GBS has been described.

In AIDP, an early step in the induction of tissue damage appears to be complement deposition along the outer surface of the Schwann cell. Activation of complement initiates a characteristic vesicular disintegration of the myelin sheath and also leads to recruitment of activated macrophages, which participate in damage to myelin and axons. In AMAN, the pattern is different in that complement is deposited along with IgG at the nodes of Ranvier along large motor axons. Interestingly, in cases of AMAN, antibodies against GD1a appear to have a fine specificity that favors binding to motor rather than sensory nerve roots, even though this ganglioside is expressed on both fiber types.

Pathophysiology In the demyelinating forms of GBS (AIDP), the basis for flaccid paralysis and sensory disturbance is conduction block. This finding, demonstrable electrophysiologically, implies that the axonal connections remain intact. Hence, recovery can take place rapidly as remyelination occurs. In severe cases of demyelinating GBS, secondary axonal degeneration usually occurs; its extent can be estimated electrophysiologically. More secondary axonal degeneration correlates with a slower rate of recovery and a greater degree of residual disability. With AMAN and AMSAN, a primary axonal pattern is encountered electrophysiologically (low-amplitude compound muscle action potentials). The implication has been that axons have degenerated and become disconnected from their targets, specifically the neuromuscular junctions, and must therefore regenerate for recovery to take place. However, the rapid recovery in many cases suggests the low amplitudes are often from reversible conduction block due to binding of antibodies to ion channel proteins in the nodes and paranodes. In severe cases, axonal degeneration can occur, and it is in these cases that recovery is much slower.

Laboratory Features CSF findings are distinctive, consisting of an elevated CSF protein level (1–10 g/L [100–1000 mg/dL]) without accompanying pleocytosis. The CSF is often normal when symptoms have been present for ≤48 h; by the end of the first week, the level of protein is usually elevated. A transient increase in the CSF white cell count (10–100/μL) occurs on occasion in otherwise typical GBS; however, a sustained CSF pleocytosis suggests an alternative diagnosis (viral myelitis) or a concurrent diagnosis such as unrecognized HIV infection, leukemia or lymphoma with infiltration of nerves, or neurosarcoidosis. EDx features are mild or absent in the early stages of GBS and lag behind the clinical evolution. In AIDP, the earliest features are prolonged F-wave latencies, prolonged distal latencies, and reduced amplitudes of compound muscle action potentials (CMAPs), probably owing to the predilection for involvement of nerve roots and distal motor nerve terminals early in the course. Later, slowing of conduction velocity, conduction block, and temporal dispersion may be appreciated (Table 458-1). Occasionally, sensory nerve action potentials (SNAPs) may be normal in the feet (e.g., sural nerve) when abnormal in the arms. This is also a sign that the patient does not have one of the more typical "length-dependent" polyneuropathies. As mentioned, in AMAN and AMSAN, the principal EDx finding is reduced amplitude of CMAPs (and also SNAPs with AMSAN) without conduction slowing or prolongation of distal latencies, which early on is caused by conduction block but later can be due to axonal degeneration.

Diagnosis GBS is a descriptive entity. The diagnosis of AIDP is made by recognizing the pattern of rapidly evolving paralysis with areflexia, absence of fever or other systemic symptoms, and characteristic antecedent events. In 2011, the Brighton Collaboration developed a new set of case definitions for GBS in response to needs of epidemiologic studies of vaccination and assessing risks of GBS **(Table 458-3)**. These criteria have subsequently been validated. Other disorders that may enter into the differential diagnosis include acute myelopathies (especially with prolonged back pain and sphincter disturbances); diphtheria (early oropharyngeal disturbances); Lyme polyradiculitis and other tick-borne paralyses; porphyria (abdominal pain, seizures, psychosis); vasculitic neuropathy (check erythrocyte sedimentation rate, described below); poliomyelitis and acute flaccid myelitis (wild-type poliovirus, West Nile virus, enterovirus D68, enterovirus A71, Japanese encephalitis virus, and the wild-type poliovirus); CMV polyradiculitis (in immunocompromised patients); critical illness neuropathy or myopathy; neuromuscular junction disorders such as myasthenia gravis and botulism (pupillary reactivity lost early); poisonings with organophosphates, thallium, or arsenic; paralytic shellfish poisoning; or severe hypophosphatemia (rare). Cases of acute flaccid myelitis may pose particular challenges in distinguishing these from GBS because sphincter disturbances may be absent.

Laboratory tests are helpful primarily to exclude mimics of GBS. CSF pleocytosis is seen with poliomyelitis, acute flaccid myelitis, and Lyme and CMV polyradiculitis. EDx features may be minimal early in GBS, and the CSF protein level may not rise until the end of the first

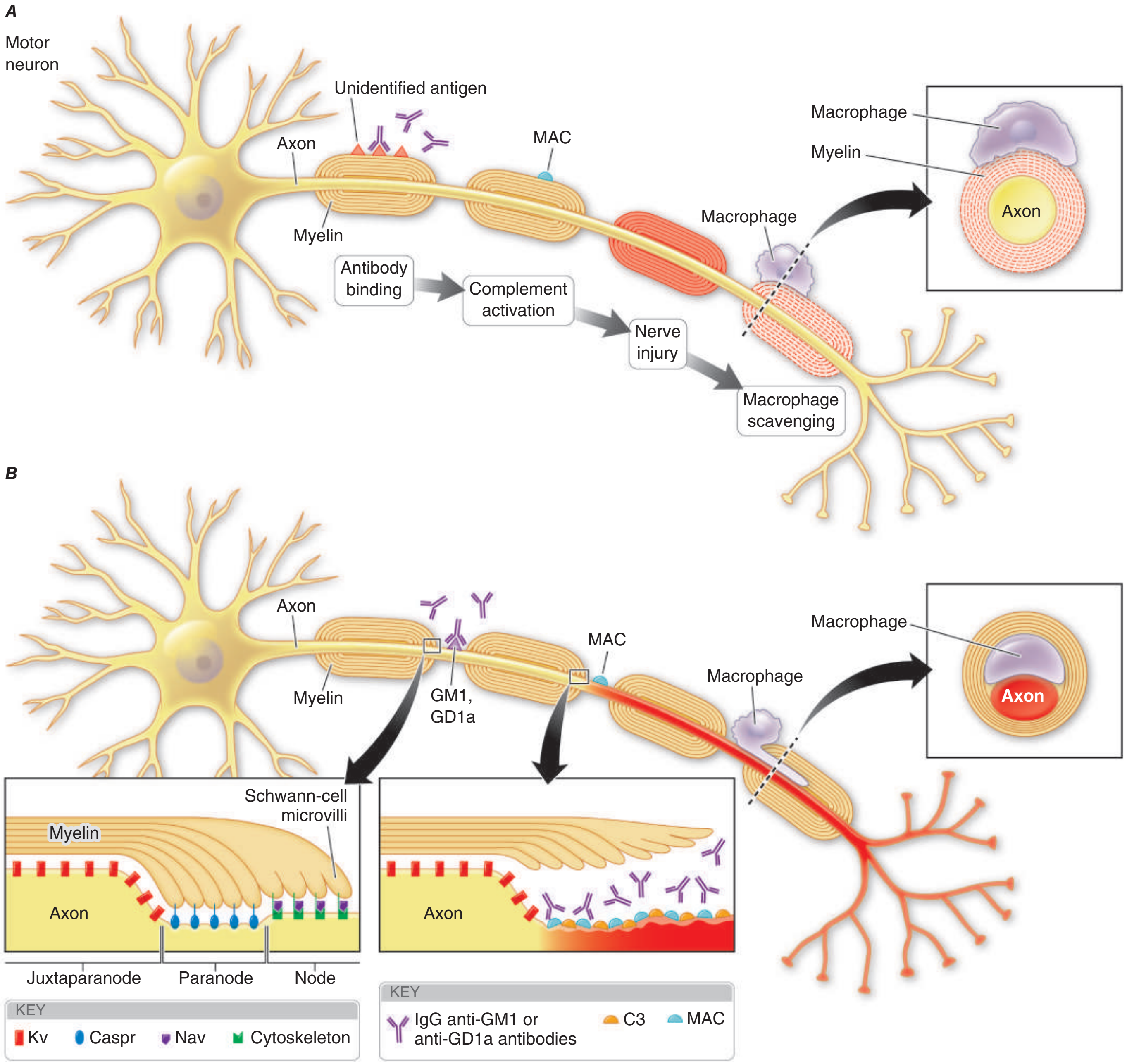
A
Motor neuron
Unidentified antigen
Axon
MAC
Macrophage
Myelin
Axon
Myelin
Macrophage
Antibody binding
Complement activation
Nerve injury
Macrophage scavenging
B
Axon
Myelin
GM1, GD1a
MAC
Macrophage
Macrophage
Axon
Schwann-cell microvilli
Myelin
Axon
Axon
Juxtaparanode
Paranode
Node
KEY
Kv
Caspr
Nav
Cytoskeleton
KEY
IgG anti-GM1 or anti-GD1a antibodies
C3
MAC

FIGURE 458-2 Possible immune mechanisms in Guillain-Barré syndrome (GBS). Panel ***A*** shows the immunopathogenesis of AIDP. Although autoantigens have yet to be unequivocally identified, autoantibodies may bind to myelin antigens and activate complement. This is followed by the formation of membrane-attack complex (MAC) on the outer surface of Schwann cells and the initiation of vesicular degeneration. Macrophages subsequently invade myelin and act as scavengers to remove myelin debris. Panel ***B*** shows the immunopathogenesis of acute axonal forms of GBS (acute motor axonal neuropathy [AMAN] and acute motor-sensory axonal neuropathy [AMSAN]). Myelinated axons are divided into four functional regions: the nodes of Ranvier, paranodes, juxtaparanodes, and internodes. Gangliosides GM1 and GD1a are strongly expressed at the nodes of Ranvier, where the voltage-gated sodium (Nav) channels are localized. Contactin-associated protein (Caspr) and voltage-gated potassium (Kv) channels are respectively present at the paranodes and juxtaparanodes. IgG anti-GM1 or anti-GD1a autoantibodies bind to the nodal axolemma, leading to MAC formation. This results in the disappearance of Nav clusters and the detachment of paranodal myelin, which can lead to nerve-conduction failure and muscle weakness. Axonal degeneration may follow at a later stage. Macrophages subsequently invade from the nodes into the periaxonal space, scavenging the injured axons. *(From N Yuki, H-P Hartung: Guillain-Barré syndrome. N Engl J Med 366:2294, 2012. Copyright © 2012 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

week. If the diagnosis is strongly suspected, treatment should be initiated without waiting for evolution of the characteristic EDx and CSF findings to occur. GBS patients with risk factors for HIV or with CSF pleocytosis should have a serologic test for HIV.

TREATMENT

Guillain-Barré Syndrome

In the vast majority of patients with GBS, treatment should be initiated as soon after diagnosis as possible. Each day counts; ~2 weeks after the first motor symptoms, it is not known whether immunotherapy is still effective. If the patient has already reached the plateau stage, then treatment probably is no longer indicated, unless the patient has severe motor weakness and one cannot exclude the possibility that an immunologic attack is still ongoing. Either high-dose intravenous immune globulin (IVIg) or plasmapheresis (PLEX) can be initiated, as they are equally effective for typical GBS. A combination of the two therapies is not significantly better than either alone. IVIg is often the initial therapy chosen because of its ease of administration and good safety record. IVIg is usually administered as five daily infusions for a total dose of 2 g/kg body weight. There is some evidence that GBS autoantibodies are neutralized by anti-idiotypic antibodies present in IVIg

TABLE 458-3 Brighton Criteria for Diagnosis of Guillain-Barré Syndrome (GBS) and Miller Fisher Syndrome

Clinical case definitions for diagnosis of GBS

Level 1 of diagnostic certainty

Bilateral AND flaccid weakness of the limbs

AND

Decreased or absent deep tendon reflexes in weak limbs

AND

Monophasic illness pattern and interval between onset and nadir of weakness between 12 h and 28 days and subsequent clinical plateau

AND

Electrophysiologic findings consistent with GBS

AND

Cytoalbuminologic dissociation (i.e., elevation of CSF protein level above laboratory normal value AND CSF total white cell count <50 cells/μL)

AND

Absence of an identified alternative diagnosis for weakness

Level 2 of diagnostic certainty

Bilateral AND flaccid weakness of the limbs

AND

Decreased or absent deep tendon reflexes in weak limbs

AND

Monophasic illness pattern and interval between onset and nadir of weakness between 12 h and 28 days and subsequent clinical plateau

AND

CSF total white cell count <50 cells/μL (with or without CSF protein elevation above laboratory normal value)

OR

If CSF not collected or results not available, electrophysiologic studies consistent with GBS

AND

Absence of identified alternative diagnosis for weakness

Level 3 of diagnostic certainty

Bilateral and flaccid weakness of the limbs

AND

Decreased or absent deep tendon reflexes in weak limbs

AND

Monophasic illness pattern and interval between onset and nadir of weakness between 12 h and 28 days and subsequent clinical plateau

AND

Absence of identified alternative diagnosis for weakness

Clinical case definitions for diagnosis of Miller Fisher syndrome

Level 1 of diagnostic certainty

Bilateral ophthalmoparesis and bilateral reduced or absent tendon reflexes, and ataxia

AND

Absence of limb weakness

AND

Monophasic illness pattern and interval between onset and nadir of weakness between 12 h and 28 days and subsequent clinical plateau

AND

Cytoalbuminologic dissociation (i.e., elevation of cerebrospinal protein above the laboratory normal and total CSF white cell count <50 cells/μL)

AND

Nerve conduction studies are normal, OR indicate involvement of sensory nerves only

AND

No alterations in consciousness or corticospinal tract signs

AND

Absence of identified alternative diagnosis

Level 2 of diagnostic certainty

Bilateral ophthalmoparesis and bilateral reduced or absent tendon reflexes and ataxia

AND

Absence of limb weakness

AND

Monophasic illness pattern and interval between onset and nadir of weakness between 12 h and 28 days and subsequent clinical plateau

AND

CSF with a total white cell count <50 cells/μL) (with or without CSF protein elevation above laboratory normal value)

OR

Nerve conduction studies are normal, OR indicate involvement of sensory nerves only

AND

No alterations in consciousness or corticospinal tract signs

AND

Absence of identified alternative diagnosis

Level 3 of diagnostic certainty

Bilateral ophthalmoparesis and bilateral reduced or absent tendon reflexes and ataxia

AND

Absence of limb weakness

AND

Monophasic illness pattern and interval between onset and nadir of weakness between 12 h and 28 days and subsequent clinical plateau

AND

No alterations in consciousness or corticospinal tract signs

AND

Absence of identified alternative diagnosis

Abbreviation: CSF, cerebrospinal fluid.

Source: Reproduced with permission from JJ Sejvar et al: Guillain-Barré syndrome and Fisher syndrome: Case definitions and guidelines for collection, analysis, and presentation of immunization safety data. Vaccine 29:599, 2011.

preparations, perhaps accounting for the therapeutic effect. A course of PLEX usually consists of ~40–50 mL/kg plasma exchange (PE) 4–6 times over 7–12 days. Meta-analysis of randomized clinical trials indicates that treatment reduces the need for mechanical ventilation by nearly half (from 27 to 14% with PLEX) and increases the likelihood of full recovery at 1 year (from 55 to 68%). Functionally significant improvement may occur toward the end of the first week of treatment or may be delayed for several weeks. The lack of noticeable improvement following a course of IVIg or PLEX is not an indication to treat with the alternate treatment. However, there are occasional patients who are treated early in the course of GBS and improve, who then relapse within a month. Brief retreatment with the original therapy is usually effective in such cases. Glucocorticoids have not been found to be effective in GBS. Occasional patients with very mild forms of GBS, especially those who appear to have already reached a plateau when initially seen, may be managed conservatively without IVIg or PLEX.

In the worsening phase of GBS, most patients require monitoring in a critical care setting, with particular attention to vital capacity, heart rhythm, blood pressure, nutrition, deep-vein thrombosis prophylaxis, cardiovascular status, early consideration (after 2 weeks of intubation) of tracheotomy, and chest physiotherapy. As noted, ~30% of patients with GBS require ventilatory assistance, sometimes for prolonged periods of time (several weeks or longer). Frequent turning and assiduous skin care are important, as are daily range-of-motion exercises to avoid joint contractures and daily reassurance as to the generally good outlook for recovery.

Prognosis and Recovery Approximately 85% of patients with GBS achieve a full functional recovery within several months to a year, although minor findings on examination (such as areflexia) may persist and patients often complain of continued symptoms, including fatigue. The mortality rate is <5% in optimal settings; death usually results from secondary pulmonary complications. The outlook is worst in patients with severe proximal motor and sensory axonal damage. Such axonal damage may be either primary or secondary in nature (see "Pathophysiology," above), but in either case, successful regeneration cannot occur. Other factors that worsen the outlook for recovery are advanced age, a fulminant or severe attack, and a delay in the onset of treatment. Elevated serum levels of neurofilament light (Nfl) chains and high titers of serum anti-GM1 antibodies are both associated with more axonal involvement in GBS and poor recovery. Between 5 and 10% of patients with typical GBS have one or more late relapses; many of these cases are then classified as chronic inflammatory demyelinating polyneuropathy (CIDP).

CHRONIC INFLAMMATORY DEMYELINATING POLYNEUROPATHY

CIDP is distinguished from GBS by its chronic course. In other respects, this neuropathy shares many features with the common demyelinating form of GBS, including elevated CSF protein levels and the EDx findings of acquired demyelination. Most cases occur in adults, and males are affected slightly more often than females. The incidence of CIDP is lower than that of GBS, but due to the protracted course, the prevalence is greater. As with GBS, CIDP and its variants can be triggered by use of immune checkpoint inhibitors used to treat various cancers.

Clinical Manifestations Onset is usually gradual over a few months or longer, but in a few cases, the initial attack is indistinguishable from that of GBS. An acute-onset form of CIDP may mimic GBS but should be considered if it deteriorates >9 weeks after onset or relapses at least three times. Symptoms are both motor and sensory in most cases. Weakness of the limbs is usually symmetric but can be strikingly asymmetric in multifocal acquired demyelinating sensory and motor (MADSAM) neuropathy variant (Lewis-Sumner syndrome) in which discrete peripheral nerves are involved. There is considerable variability from case to case. Some patients experience a chronic progressive course, whereas others, usually younger patients, have a relapsing and remitting course. A small proportion have cranial nerve findings, including external ophthalmoplegia. Some have only motor findings, and a small proportion present with a relatively pure syndrome of sensory ataxia. The latter can be seen in the chronic inflammatory sensory polyradiculopathy (CISP) variant of CIDP in which demyelination predominantly occurs at the sensory roots or with the distal acquired demyelinating symmetric (DADS) variant. Some patients with CISP have mild motor involvement, and these cases are termed CISP-plus. New European Academy of Neurology/Peripheral Nerve Society (EAN/PNS) criteria for CIDP considers CISP as a separate entity, but we and others still feel it belongs as a subcategory of CIDP as the histopathology and response to treatment are similar. CIDP tends to ameliorate over time with treatment; the result is that many years after onset, nearly 75% of patients have reasonable functional status. Death from CIDP is uncommon.

Diagnosis The diagnosis rests on characteristic clinical, CSF, and electrophysiologic findings. The CSF is usually acellular with an elevated protein level, sometimes several times normal. As with GBS, a CSF pleocytosis should lead to the consideration of HIV infection, leukemia or lymphoma, and neurosarcoidosis. EDx findings reveal variable degrees of conduction slowing, prolonged distal latencies, distal and temporal dispersion of CMAPs, and conduction block as the principal features. In particular, the presence of conduction block is a certain sign of an acquired demyelinating process. Evidence of axonal loss, presumably secondary to demyelination, is present in >50% of patients. Serum protein electrophoresis with immunofixation is indicated to search for monoclonal gammopathy and associated conditions (see "Monoclonal Gammopathy of Undetermined Significance," below). Magnetic resonance imaging (MRI) can demonstrate enlarged nerves, clumping of cauda equina, and enhancement. Ultrasound is cheaper and often more readily available and can likewise show enlargement of nerves at the roots or more distally. Studies have shown that imaging complements EDx findings and increases sensitivity. In all patients with presumptive CIDP, it is also reasonable to exclude vasculitis, collagen vascular disease (especially SLE), chronic hepatitis, HIV infection, amyloidosis, and diabetes mellitus. Other associated conditions include inflammatory bowel disease and lymphoma.

Pathogenesis Biopsy in typical CIDP reveals little inflammation and onion-bulb changes (imbricated layers of attenuated Schwann cell processes surrounding an axon) that result from recurrent demyelination and remyelination (Fig. 458-1). The response to therapy suggests that CIDP is immune-mediated; CIDP responds to glucocorticoids, whereas GBS does not. Passive transfer of demyelination into experimental animals has been accomplished using IgG purified from the serum of some patients with CIDP, lending support for a humoral autoimmune pathogenesis. A minority of patients have serum antibodies against P0, myelin P2 protein, or PMP22 (proteins whose genes are mutated in certain forms of hereditary Charcot-Marie-Tooth neuropathy).

As many as 25% of patients with clinical features of CIDP also have a monoclonal gammopathy of undetermined significance (MGUS), discussed below. Cases associated with monoclonal IgA or IgG kappa usually respond to treatment as favorably as cases without a monoclonal gammopathy. Patients with IgM-kappa monoclonal gammopathy and antibodies directed against myelin-associated glycoprotein (MAG) have a distinct demyelinating polyneuropathy with more sensory findings, usually only distal weakness, and a poor response to immunotherapy.

TREATMENT

Chronic Inflammatory Demyelinating Polyneuropathy

Most authorities initiate treatment for CIDP when progression is rapid or walking is compromised. If the disorder is mild, management can be expectant, awaiting spontaneous remission. Controlled studies have shown that high-dose IVIg, subcutaneous Ig (scIg), PLEX, and glucocorticoids are all more effective than placebo. Initial therapy is usually with IVIg, administered as 2.0 g/kg body weight given in divided doses over 2–5 days; three monthly courses are generally recommended before concluding a patient has failed treatment. If the patient responds, the infusion intervals can be gradually increased or the dosage decreased (e.g., starting at 1 g/kg every 3–4 weeks). Patients who require more frequent IVIg, experience side effects with IVIg (headaches), have poor venous access, or find it more convenient are treated with scIg (2–3 times a week such that the total dosage per month is the same or slightly higher than the monthly dosage of IVIg). PLEX, which appears to be as effective as IVIg, is initiated at 2–3 treatments per week for 6 weeks; periodic retreatment may also be required. Treatment with glucocorticoids is another option (60–80 mg prednisone PO daily for 1–2 months, followed by a gradual dose reduction of 10 mg per month as tolerated), but long-term adverse effects including bone demineralization, gastrointestinal bleeding, and cushingoid changes are problematic. As many as one-third of patients with CIDP fail to respond adequately to the initial therapy chosen; a different treatment should then be tried. Patients who fail therapy with IVIg, scIg, PLEX, and glucocorticoids may benefit from treatment with immunosuppressive agents such as azathioprine, methotrexate, cyclosporine, and cyclophosphamide, either alone or as adjunctive therapy. Use of these therapies requires periodic reassessment of their risks and benefits. A trial of efgartigimod alfa, a human neonatal Fc antibody fragment approved for myasthenia gravis (Chap. 457), demonstrated effectiveness in preventing relapses in CIDP; however, the medication will need to be compared with other more established options to determine its place in the CIDP treatment algorithm. In patients with a CIDP-like neuropathy who fail to respond to treatment, it is important to evaluate for a nodopathy, paranodopathy, or POEMS syndrome (*p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathy, *m*onoclonal gammopathy, *s*kin changes; see below).

NODOPATHIES AND PARANODOPATHIES

Approximately 10% of patients previously considered to have CIDP have autoantibodies targeting antigens residing in nodal and paranodal regions that are responsible for the positioning and anchoring of ion channels and myelin folds in strategic locations along the axolemma. The EAN/PNS criteria now consider these separate from CIDP, and they are called nodopathies and paranodopathies. These neuropathies are associated with IgG4 isotype antibodies directed against nodal or paranodal antibodies including contactin-1 (CNTN1) or neurofascin 155 (NF155) and, less commonly, IgM anti-NF140/186. Patients typically manifest with progressive symmetric, distally predominant weakness, sensory ataxia, and postural and intention tremor. Renal failure and nephrotic syndrome from membranous glomerulonephritis are associated with CNTN1 neuropathy. Of note, the CNTN1 protein is also present on podocytes on kidneys, and antibodies that deposit along the glomerular basement membrane are visible on renal biopsy. Other antibodies have less clinical specificity. Anti-contactin associated protein-1 (Caspr1) antibodies are associated with severe neuropathic pain. Passive transfer of IgG4 CNTN1 antibodies produces paranodal damage and ataxia in rodents. Electrophysiology is indistinguishable from typical CIDP.

Importantly, as these nodopathies and paranodopathies are usually associated with IgG4 antibodies, they are less responsive to IVIg. However, they can respond to rituximab.

MULTIFOCAL MOTOR NEUROPATHY

Multifocal motor neuropathy (MMN) is a distinctive but uncommon neuropathy that presents with slowly progressive motor weakness and atrophy evolving over years in the distribution of selected nerve trunks, associated with sites of persistent focal motor conduction block in the same nerve trunks. Sensory fibers are relatively spared. The arms are affected more frequently than the legs, and >75% of all patients are male. Some cases have been confused with lower motor neuron forms of amyotrophic lateral sclerosis **(Chap. 448)**. Less than 50% of patients present with high titers of polyclonal IgM antibody to the ganglioside GM1. It is uncertain how this finding relates to the discrete foci of persistent motor conduction block, but high concentrations of GM1 gangliosides are normal constituents of nodes of Ranvier in peripheral nerve fibers. Pathology reveals demyelination and mild inflammatory changes at the sites of conduction block.

Most patients with MMN respond to high-dose IVIg or scIg (dosages as for CIDP, above); periodic retreatment is required (usually at least monthly) to maintain the benefit. Some refractory patients have responded to rituximab or cyclophosphamide. Glucocorticoids and PE are not effective.

NEUROPATHIES WITH MONOCLONAL GAMMOPATHY

MULTIPLE MYELOMA

Clinically overt polyneuropathy occurs in ~5% of patients with the commonly encountered type of multiple myeloma, which exhibits either lytic or diffuse osteoporotic bone lesions. These neuropathies are sensorimotor, are usually mild and slowly progressive but may be severe, and generally do not reverse with successful suppression of the myeloma. In most cases, EDx and pathologic features are consistent with a process of axonal degeneration.

In contrast, myeloma with osteosclerotic features, although representing only 3% of all myelomas, is associated with polyneuropathy in one-half of cases. These neuropathies, which may also occur with solitary plasmacytoma, are distinct because they (1) are demyelinating or mixed axonal and demyelinating by EDx, have elevated CSF protein, and clinically resemble CIDP; (2) often respond to radiation therapy or removal of the primary lesion; (3) are associated with different monoclonal proteins and light chains (almost always lambda as opposed to primarily kappa in the lytic type of multiple myeloma); (4) are typically refractory to standard treatments of CIDP; and (5) may occur in association with other systemic findings including thickening of the skin, hyperpigmentation, hypertrichosis, organomegaly, endocrinopathy, anasarca, and clubbing of fingers. These are features of POEMS syndrome. Levels of vascular endothelial growth factor (VEGF) are increased in the serum, and this factor is thought to somehow play a pathogenic role in this syndrome. Treatment of the neuropathy is best directed at the osteosclerotic myeloma using surgery, radiotherapy, chemotherapy, or autologous peripheral blood stem cell transplantation.

Neuropathies are also encountered in other systemic conditions with gammopathy, including Waldenström macroglobulinemia, primary systemic amyloidosis, and cryoglobulinemic states (mixed essential cryoglobulinemia, some cases of hepatitis C).

MONOCLONAL GAMMOPATHY OF UNDETERMINED SIGNIFICANCE

Chronic polyneuropathies occurring in association with MGUS are usually associated with the immunoglobulin isotypes IgG, IgA, and IgM. Most patients present with isolated sensory symptoms in their distal extremities and have EDx features of an axonal sensory or sensorimotor polyneuropathy. These patients otherwise resemble idiopathic sensory polyneuropathy, and the MGUS might just be coincidental. They usually do not respond to immunotherapies designed to reduce the concentration of the monoclonal protein. Some patients, however, present with generalized weakness and sensory loss and EDx studies indistinguishable from CIDP without monoclonal gammopathy (see "Chronic Inflammatory Demyelinating Polyneuropathy," above), and their response to immunosuppressive agents is also similar. An exception is the syndrome of IgM-kappa monoclonal gammopathy associated with an indolent, long-standing, sometimes static sensory neuropathy, frequently with tremor and sensory ataxia. Most patients are men and aged >50 years. In the majority, the monoclonal IgM immunoglobulin binds to a normal peripheral nerve constituent, MAG, found in the paranodal regions of Schwann cells. Binding appears to be specific for a polysaccharide epitope that is also found in other normal peripheral nerve myelin glycoproteins, P0 and PMP22, and also in other normal nerve-related glycosphingolipids (Fig. 458-1). In the MAG-positive cases, IgM paraprotein is incorporated into the myelin sheaths of affected patients and widens the spacing of the myelin lamellae, thus producing a distinctive ultrastructural pattern. Demyelination and remyelination are the hallmarks of the lesions, but axonal loss develops over time. These anti-MAG polyneuropathies are typical refractory to immunotherapy. In a small proportion of patients (30% at 10 years), MGUS will in time evolve into frankly malignant conditions such as multiple myeloma or lymphoma.

VASCULITIC NEUROPATHY

Peripheral nerve involvement is common in polyarteritis nodosa (PAN), appearing in half of all cases clinically and in 100% of cases at postmortem studies **(Chap. 375)**. The most common pattern is multifocal (asymmetric) motor-sensory neuropathy (mononeuropathy multiplex) due to ischemic lesions of nerve trunks and roots; however, some cases of vasculitic neuropathy present as a distal, symmetric sensorimotor polyneuropathy. Symptoms of neuropathy are a common presenting complaint in patients with PAN. The EDx findings are those of an axonal process. Small- to medium-sized arteries of the vasa nervorum, particularly the epineural vessels, are affected in PAN, resulting in a widespread ischemic neuropathy. A high frequency of neuropathy occurs in eosinophilic granulomatosis with polyangiitis (Churg-Strauss syndrome [CSS]).

Systemic vasculitis should always be considered when a subacute or chronically evolving mononeuropathy multiplex occurs in conjunction with constitutional symptoms (fever, anorexia, weight loss, loss of energy, malaise, and nonspecific pains). Diagnosis of suspected vasculitic neuropathy is made by a combined nerve and muscle biopsy, with serial section or skip-serial techniques.

Approximately one-third of biopsy-proven cases of vasculitic neuropathy are "nonsystemic" in that the vasculitis appears to affect only peripheral nerves. Constitutional symptoms are absent, and the course is more indolent than that of PAN. The erythrocyte sedimentation rate may be elevated, but other tests for systemic disease are negative.

Nevertheless, clinically silent involvement of other organs is likely, and vasculitis is frequently found in muscle biopsied at the same time as nerve.

Vasculitic neuropathy may also be seen as part of the vasculitis syndrome occurring in the course of other connective tissue disorders. The most frequent is rheumatoid arthritis, but ischemic neuropathy due to involvement of vasa nervorum may also occur in mixed cryoglobulinemia, Sjögren's syndrome, granulomatosis with polyangiitis (formerly known as Wegener's), hypersensitivity angiitis, SLE, and progressive systemic sclerosis.

Some vasculitides are associated with antineutrophil cytoplasmic antibodies (ANCAs), which in turn are subclassified as cytoplasmic (cANCA) or perinuclear (pANCA). cANCAs are directed against proteinase 3 (PR3), whereas pANCAs target myeloperoxidase (MPO). PR3/cANCAs are associated with eosinophilic granulomatosis with polyangiitis, whereas MPO/pANCAs are typically associated with microscopic polyangiitis, CSS, and less commonly PAN. Of note, MPO/pANCA has also been seen in minocycline-induced vasculitis.

Management of these neuropathies, including the "nonsystemic" vasculitic neuropathy, consists of treatment of the underlying condition as well as the aggressive use of glucocorticoids and cyclophosphamide. Use of these immunosuppressive agents has resulted in dramatic improvements in outcome, with 5-year survival rates now >80%. Clinical trials found that the combination of rituximab and glucocorticoids is not inferior to cyclophosphamide and glucocorticoids. Thus, combination therapy with glucocorticoids and rituximab is recommended as the standard initial treatment, particularly for ANCA-associated vasculitis. Mepolizumab, an anti–interleukin 5 monoclonal antibody, when added to standard care, is also effective for treatment of eosinophilic granulomatosis with polyangiitis.

ANTI-Hu PARANEOPLASTIC NEUROPATHY

(Chap. 99) This uncommon immune-mediated disorder manifests as a sensory neuronopathy (i.e., selective damage to sensory nerve bodies in dorsal root ganglia). The onset is often asymmetric with dysesthesias and sensory loss in the limbs that soon progress to affect all limbs, the torso, and the face. Marked sensory ataxia, pseudoathetosis, and inability to walk, stand, or even sit unsupported are frequent features and are secondary to the extensive deafferentation. Subacute sensory neuronopathy may be idiopathic, but more than half of cases are paraneoplastic, primarily related to lung cancer, and most of those are small-cell lung cancer (SCLC). Diagnosis of the underlying SCLC requires awareness of the association, testing for the paraneoplastic antibody, and often positron emission tomography (PET) scanning for the tumor. The target antigens are a family of RNA-binding proteins (HuD, HuC, and Hel-N1) that in normal tissues are only expressed by neurons. The same proteins are usually expressed by SCLC, triggering in some patients an immune response characterized by antibodies and cytotoxic T cells that cross-react with the Hu proteins of the dorsal root ganglion neurons, resulting in immune-mediated neuronal destruction. An encephalomyelitis may accompany the sensory neuronopathy and presumably has the same pathogenesis. Neurologic symptoms usually precede, by ≤6 months, the identification of SCLC. The sensory neuronopathy runs its course in a few weeks or months and stabilizes, leaving the patient disabled. Most cases are unresponsive to treatment with glucocorticoids, IVIg, PE, or immunosuppressant drugs.

■ FURTHER READING

Amato AA, Ropper AH: Sensory ganglionopathy. N Engl J Med 383:1657, 2020.

Cortese A et al: Antibodies to neurofascin, contactin-1, and contactin-associated protein 1 in CIDP: Clinical relevance of IgG isotype. Neurol Neuroimmunol Neuroinflamm 7:E639, 2020.

Gwathmey KG et al: Peripheral nerve vasculitis: Classification and disease associations. Neurol Clin 37:303, 2019.

Keh RYS et al: COVID-19 vaccination and Guillain-Barré syndrome: Analyses using the National Immunoglobulin Database. Brain 146:739, 2023.

Koike H et al: ANCA-associated vasculitic neuropathies: A review. Neurol Ther 11:21, 2022.

Puwanant A et al: Clinical spectrum of neuromuscular complications after immune checkpoint inhibition. Neuromuscul Disord 29:127, 2019.

Súkeníková L et al: Autoreactive T cells target peripheral nerves in Guillain-Barré syndrome. Nature 626:160, 2024.

Van den Bergh PYK et al: European Academy of Neurology/Peripheral Nerve Society guideline on diagnosis and treatment of chronic inflammatory demyelinating polyradiculoneuropathy: Report of a joint Task Force—Second revision. Eur J Neurol 28:3556, 2021.

van Doorn PA et al: European Academy of Neurology/Peripheral Nerve Society Guideline on diagnosis and treatment of Guillain-Barré syndrome. Eur J Neurol 30:3646, 2023.

Willison AG et al: SARS-CoV-2 vaccination and neuroimmunological disease: A review. JAMA Neurol 81:179, 2024.

459

Myasthenia Gravis and Other Diseases of the Neuromuscular Junction

Anthony A. Amato, Amanda C. Guidon

Myasthenia gravis (MG) is a neuromuscular junction (NMJ) disorder characterized by weakness and fatigability of skeletal muscles. The underlying defect is a decrease in the number of available acetylcholine receptors (AChRs) at NMJs due to an antibody-mediated autoimmune attack. Available treatments for MG are highly effective, although side effects can limit their use and a cure has remained elusive.

■ PATHOPHYSIOLOGY

At the NMJ **(Fig. 459-1, Video 459-1)**, acetylcholine (ACh) is synthesized in the motor nerve terminal and stored in vesicles (quanta). When an action potential travels down a motor nerve and reaches the nerve terminal, ACh from 150 to 200 vesicles is released and combines with AChRs that are densely packed at the crests of postsynaptic folds on skeletal muscle. The AChR consists of five subunits (2α, 1β, 1δ, 1γ, or ε) arranged around a central pore. When ACh combines with the binding sites on α subunits of the AChR, the channel in the AChR opens, permitting the rapid entry of cations, chiefly sodium, which produces depolarization at the end-plate region of the muscle fiber. If the depolarization is sufficiently large, it initiates an action potential that is propagated along the muscle fiber, triggering muscle contraction. This process is rapidly terminated by hydrolysis of ACh by acetylcholinesterase, which is present within the synaptic cleft, and by diffusion of ACh away from the receptor.

Muscle-specific tyrosine kinase (MuSK) is a postsynaptic transmembrane protein that helps stabilize postsynaptic clustering of AChRs. Agrin is released from the presynaptic motor nerve terminal and binds low-density lipoprotein receptor-related protein 4 (LRP4). This agrin-LRP4 complex activates MuSK. This facilitates recruitment of cytoplasmic proteins, including downstream of tyrosine kinase 7 (DOK7) and rapsyn, to assist in clustering AChR. These various proteins are important in the pathogenesis of not only MG but also some of the hereditary congenital myasthenic syndromes.

In MG, the fundamental defect is a decrease in the number of available AChRs at the postsynaptic muscle membrane. In addition, the postsynaptic folds are flattened, or "simplified." These changes result in decreased efficiency of neuromuscular transmission. Therefore, although ACh is released normally, it produces small end-plate potentials that may fail to trigger muscle action potentials. Failure of transmission results in muscle weakness.

The amount of ACh released per impulse normally declines on repeated activity (termed *presynaptic rundown*). In myasthenic patients, reduced efficiency of neuromuscular transmission, combined with this normal rundown, results in activation of fewer and fewer muscle fibers by successive nerve impulses, and hence increasing weakness, or *myasthenic fatigue*. This mechanism also accounts for the decremental response to repetitive nerve stimulation seen on electrodiagnostic testing.

MG is an autoimmune disorder most commonly caused by anti-AChR antibodies. The anti-AChR antibodies reduce the number of available AChRs at NMJs by three distinct mechanisms: (1) accelerated turnover of AChRs by a mechanism involving cross-linking and rapid endocytosis of the receptors; (2) damage to the postsynaptic muscle membrane through antibody-mediated complement activation; and (3) blockade of the active site of the AChR (i.e., the site that normally binds ACh). An immune response to MuSK, a protein involved in AChR clustering at the NMJ (as noted above), also results in MG, with reduction of AChRs demonstrated experimentally. Anti-MuSK antibody occurs in ~10% of patients (~40% of AChR antibody–negative patients with generalized MG), whereas 1–3% have antibodies to another protein at the NMJ—LRP4—that, as mentioned, is also important for clustering of AChRs. These pathogenic antibodies are IgG and are T-cell dependent. Thus, immunotherapeutic strategies directed against either the antibody-producing B cells or helper T cells, directly reducing the pathogenic antibodies, or blocking complement-mediated destruction of the AChRs are all effective in anti-AChR-positive MG. MuSK antibodies exert their pathogenic effect by directly inhibiting binding between MuSK and LRP4, leading to loss of AChRs and other functions of MuSK. Of note, MuSK antibodies are of the IgG4 subtype and, as such, do not activate complement. As a result, anti-MuSK-positive MG does not respond to complement inhibition. LRP4 antibodies are of the IgG1 subclass and also cause complement-mediated destruction, similar to AChR antibodies, and possibly interrupt agrin-induced MuSK activation.

Although MG is caused by autoantibodies, a significant contribution exists from T cells, including T regulatory (Treg) cells. These Tregs are critical in suppressing activation of other immune cells that have escaped negative selection in the thymus. Because these other cells have not been deleted by negative selection or suppressed in the periphery, they attack "self" antigens. Deficiency or dysfunction of Tregs contributes to the pathogenesis not only of MG but of many other autoimmune diseases. The primary source of Treg cells is the thymus, which

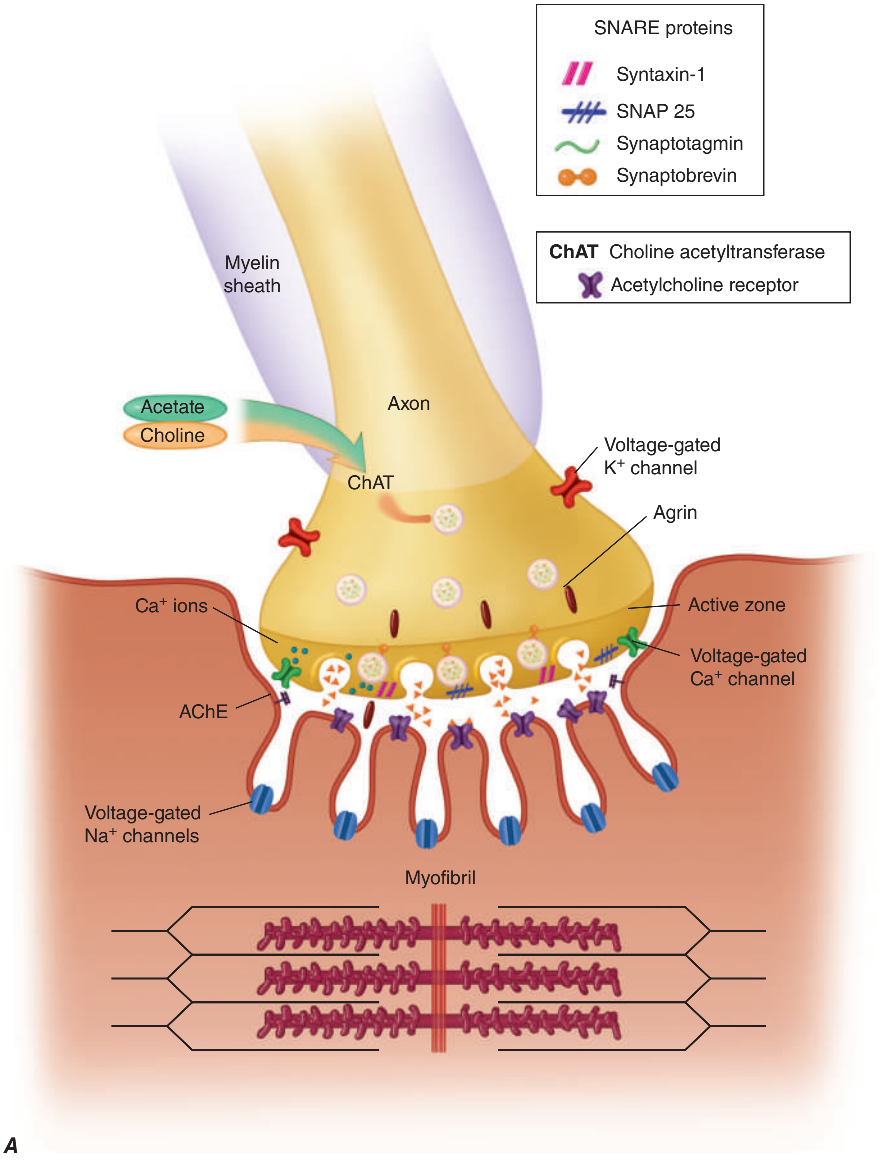

FIGURE 459-1 Illustrations of (*A*) a normal presynaptic neuromuscular junction, (*B*) a normal postsynaptic terminal, and (*C*) a myasthenic neuromuscular junction. AChE, acetylcholinesterase. See text for description of normal neuromuscular transmission. The myasthenia gravis (MG) junction demonstrates a reduced number of acetylcholine receptors (AChRs); flattened, simplified postsynaptic folds; and a widened synaptic space. See Video 459-1 also. *(Reproduced with permission from AA Amato, J Russell: Neuromuscular Disorders, 2nd ed. New York, McGraw-Hill; 2016.)*

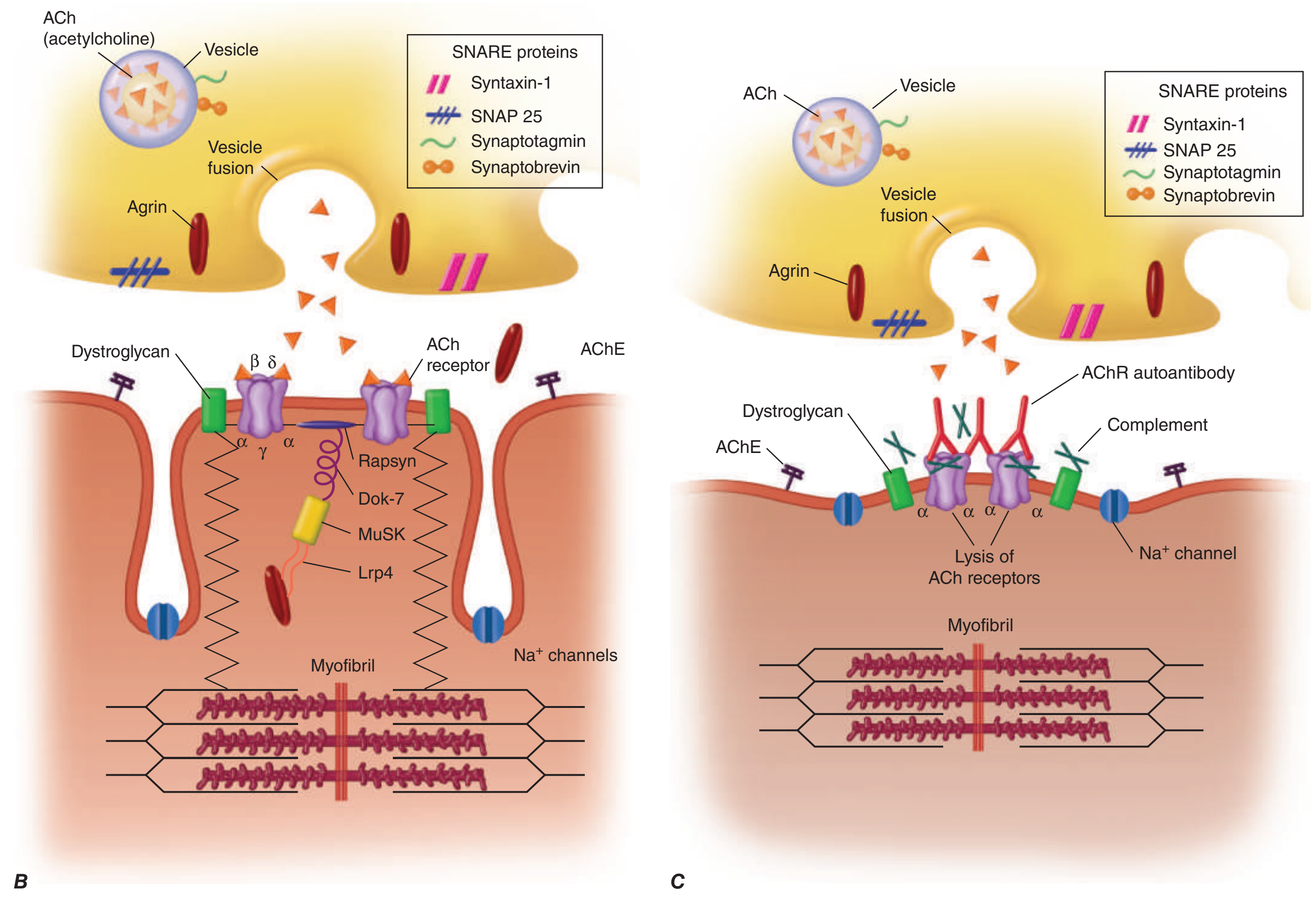

FIGURE 459-1 (*Continued*)

is abnormal in ~75% of patients with AChR antibody–positive MG. In ~65%, the thymus is "hyperplastic," with the presence of active germinal centers detected histologically. The hyperplastic thymus may be but is not necessarily enlarged. An additional 10% of patients have thymic tumors (thymomas). Muscle-like cells within the thymus (myoid cells), which express AChRs on their surface, may serve as a source of autoantigen and trigger the autoimmune reaction within the thymus gland.

CLINICAL FEATURES

MG has an incidence ranging from 6.3 to 29 per million and a prevalence ranging from 100 to 361 per million. It affects individuals in all age groups, but peak incidence occurs in women in their twenties and thirties and in men in their fifties and sixties. Overall, women are affected more frequently than men, in a ratio of ~3:2. Cardinal features are *weakness* and *fatigability* of muscles. Myasthenic weakness often worsens during repeated use (fatigue) and/or late in the day and may improve following rest or sleep. The course of MG is variable. Exacerbations and remissions may occur, particularly during the first 1–3 years after disease onset. In ~85% of patients, myasthenic weakness becomes generalized, affecting facial, bulbar, axial, or limb muscles in addition to ocular muscles. If weakness remains restricted to ocular muscles for 3 years, future generalization is unlikely, and these patients are said to have *ocular MG*. However, we have seen rare patients generalize >5 years after onset of ocular MG. Unrelated infections, systemic disorders, or tapering of MG therapies can lead to increased myasthenic weakness and may precipitate myasthenic exacerbation or "crisis" (see below). Some exacerbations occur without any identifiable precipitating factors.

The distribution of muscle weakness often has a characteristic pattern. Cranial muscles, particularly extraocular and eyelid muscles, are frequently involved early in the course of MG; diplopia and ptosis are common initial symptoms. Facial weakness produces a "snarling" expression when the patient attempts to smile. Weakness in chewing is most noticeable after prolonged effort or chewing hard or tough foods like meat. Speech may have a nasal timbre caused by weakness of the palate or a dysarthric "mushy" quality due to tongue weakness. Hoarseness can occur from laryngeal weakness. Difficulty in swallowing (dysphagia) may occur as a result of weakness of the palate, tongue, or pharynx, giving rise to nasal regurgitation or aspiration of liquids or food. Bulbar, neck, and ventilatory weakness can be especially prominent in MuSK antibody–positive MG. Weakness in neck extensor muscles can lead to head drop. Limb weakness in MG is often proximal and may be asymmetric. Nonetheless, some patients manifest with mainly distal weakness (finger and wrist drop or foot drop). Deep tendon reflexes are typically preserved. Sensory symptoms, sensory loss, and pain are absent. If ventilatory weakness necessitates intubation or noninvasive ventilation to avoid intubation, the patient is said to be in *MG crisis*. All other worsening is termed *exacerbation*.

DIAGNOSIS AND EVALUATION (TABLE 459-1)

The diagnosis is suspected based on weakness and fatigability in the typical distribution described above, without loss of deep tendon reflexes or sensory signs or symptoms or abnormality of other neurologic functions. The suspected diagnosis should be confirmed definitively before treatment is undertaken; this is essential because (1) other treatable conditions may closely resemble MG and (2) the treatment of MG may involve surgery and the prolonged use of drugs with potential side effects.

Ice-Pack Test If a patient has ptosis, application of a pack of ice over a ptotic eye for 2 min often results in improvement if the ptosis is due to an NMJ defect. A lid rise of 2 mm following this cooling is considered a positive result. This is hypothesized to be due to less depletion of quanta of AChR in the cold and reduced activity of

TABLE 459-1 Diagnosis of Myasthenia Gravis (MG)

- History
  - Diplopia, ptosis, dysarthria, dysphagia, dyspnea
  - Weakness in characteristic distribution: proximal limbs, neck extensors, generalized
  - Fluctuation and fatigue: worse with repeated activity, improved by rest
  - Effects of previous treatments
- Physical examination
  - Evaluation for ptosis at rest and following 1 min of exercise, extraocular muscles and subjective diplopia, orbicularis oculi and oris strength, jaw opening and closure
  - Assessment of muscle strength in neck and extremities
  - Weakness following repeated shoulder abduction
  - Vital capacity measurement
  - Absence of other neurologic signs
- Laboratory testing
  - Anti-AChR radioimmunoassay: ~85% positive in generalized MG; 50% in ocular MG; definite diagnosis if positive; negative result does not exclude MG; ~40% of AChR antibody–negative patients with generalized MG have anti-MuSK antibodies and ~2% have LRP4 antibodies
  - Repetitive nerve stimulation: decrement of >10% at 3 Hz: highly probable
  - Single-fiber electromyography: blocking and jitter, with normal fiber density; confirmatory, but not specific
  - Edrophonium chloride (Enlon) 2 mg + 8 mg IV; highly probable diagnosis if unequivocally positive
- Ice-pack test looking for improvement in ptosis is very sensitive
- For ocular or cranial MG: exclude intracranial lesions by CT or MRI

Abbreviations: AChR, acetylcholine receptor; CT, computed tomography; LRP4, lipoprotein receptor-related protein 4; MRI, magnetic resonance imaging; MuSK, muscle-specific tyrosine kinase.

acetylcholinesterase at the NMJ. It is a quick and easy test to do in the clinic or at the bedside of a hospitalized patient.

Autoantibodies Associated with MG As previously mentioned, anti-AChR antibodies are detectable in the serum of ~85% of all myasthenic patients but in only ~50% of patients with weakness confined to the ocular muscles. The presence of anti-AChR antibodies is virtually diagnostic of MG, but a negative test does not exclude the disease. The measured level of anti-AChR antibody does not correspond well with the severity of MG in different patients. Antibodies to MuSK are present in up to 40% of AChR antibody–negative patients with generalized MG depending on the population. MuSK antibodies are rarely present in AChR antibody–positive patients or in patients with MG limited to ocular muscles. A small proportion of MG patients without antibodies to AChR or MuSK have antibodies to LRP4. Sending LRP4 antibodies has a low specificity. As such, we only check them in patients with clear MG by phenotype and electrodiagnostic testing but absent AChR and MuSK antibodies. Additionally, antibodies against agrin also have been found in rare patients with MG, but it is unclear if they are pathogenic, and they are not currently tested in clinical practice. Additionally, anti-striated muscle antibodies directed against titin and other skeletal muscle components have been identified in some patients. However, they are not pathogenic, and their presence does not confirm the diagnosis of MG or the presence of a thymoma. Given their limited utility and potential for misinterpretation, we do not order them. Furthermore, antibodies directed against Caspr2 (contactin-associated protein-like 2) may coexist primarily in patients with thymoma who have MG and neuromyotonia or Morvan's syndrome. The presence of these antibodies can help confirm the diagnosis of a second paraneoplastic syndrome in these clinical situations.

Electrodiagnostic Testing Repetitive nerve stimulation may provide helpful diagnostic evidence of MG. Medications that inhibit acetylcholinesterase should be stopped 12–24 h or for as long as possible before testing. It is best to test weak muscles or proximal muscle groups. Electrical stimulation is delivered at a rate of two or three per second to the appropriate nerves, and action potentials are recorded from the muscles. In normal individuals, the amplitude of the evoked muscle action potentials does not change by >10% at these rates of stimulation. However, in myasthenic patients, there is a rapid reduction of >10% in the amplitude of the evoked responses. If repetitive nerve stimulation is normal and/or symptoms are exclusively ocular, single-fiber electromyography (EMG), a specialized more sensitive test typically done at MG referral centers, is performed.

Anticholinesterase Test Drugs that inhibit the enzyme acetylcholinesterase allow ACh to interact repeatedly with the limited number of AChRs in MG, producing improvement in muscle strength. Edrophonium was most commonly used historically for diagnostic testing because of the rapid onset (30 s) and short duration (~5 min) of its effect, with an objective endpoint such as ptosis typically measured. Edrophonium is no longer used due to potential for side effects and lack of availability.

Pulmonary Function Tests (Chap. 295) Measurements of ventilatory function are valuable because of the frequency and seriousness of respiratory impairment in myasthenic patients.

Differential Diagnosis Other conditions that cause weakness of the cranial and/or somatic musculature include nonautoimmune congenital myasthenia, drug-induced myasthenia, Lambert-Eaton myasthenic syndrome (LEMS), hyperthyroidism (Graves' disease), botulism, intracranial mass lesions, oculopharyngeal dystrophy, and mitochondrial myopathy (Kearns-Sayre syndrome, progressive external ophthalmoplegia). Treatment with immune checkpoint inhibitors (ICIs) for cancer may also result in autoimmune MG. Myositis and myocarditis are also often found in combination with MG as a complication of ICIs (Chap. 377). Symptoms typically begin after the first or second cycle of treatment, with ptosis, diplopia, bulbar, neck, extremity weakness, and respiratory weakness. ICI-related myositis without disordered neuromuscular transmission can mimic MG, itself causing a similar pattern of weakness including ocular and bulbar weakness, which is uncommon in other autoimmune myopathies. Patients usually improve when the ICI is discontinued and a short course of glucocorticoids is given, with intravenous immunoglobulin (IVIg) or plasma exchange depending on severity; however, with fulminant disease, the fatality rate remains high, mainly due to the concurrent myocarditis. Treatment with penicillamine (used for scleroderma or rheumatoid arthritis) has also been associated with MG. Aminoglycoside, quinolone and macrolide antibiotics, intravenous magnesium, or procainamide can also cause exacerbation of weakness in myasthenic patients; very large doses can cause neuromuscular weakness in normal individuals. Botulinum toxin injections should be avoided in MG patients.

The *congenital myasthenic syndromes (CMS)* comprise a rare heterogeneous group of disorders of the NMJ that are not autoimmune but rather are due to mutations in >30 identified genes. Virtually any component of the NMJ may be affected. Alterations in function of the presynaptic nerve terminal, in the various subunits of the AChR, acetylcholinesterase, or the other molecules involved in end-plate development or maintenance, have been identified in the different forms of CMS. These disorders share many of the clinical features of autoimmune MG, including weakness and fatigability of proximal or distal extremity muscles and often involving extraocular and eyelid muscles similar to the distribution in autoimmune MG. CMS is most often suspected when symptoms of myasthenia began in infancy or childhood; however, some patients initially present in adulthood. As in autoimmune MG, repetitive nerve stimulation is often associated with a decremental response. Some forms of CMS (e.g., acetylcholinesterase deficiency, prolonged open channel syndrome) have a feature of after-discharges that are not seen in MG. An additional clue is the absence of AChR and MuSK antibodies, although these are absent in ~10% of generalized MG patients (so-called double seronegative MG).

The prevalence of CMS is estimated at ~3.8 per 100,000. The most common genetic defects occur in the ε subunit of the AChR, accounting for ~50% of CMS cases, with mutations in the genes encoding for rapsin, COLQ, DOK7, agrin, and GFPT together accounting for ~40%.

In most of the recessively inherited forms of CMS, the mutations are heteroallelic; that is, different mutations affecting each of the two alleles are present. Features of the most common forms of CMS are summarized in **Table 459-2**. Molecular analysis is required for precise elucidation of the defect; this may lead to helpful treatment as well as genetic counseling. Some forms of CMS improve with acetylcholinesterase inhibitors, while others (e.g., slow channel syndrome, acetylcholinesterase deficiency, DOK7-related CMS) actually worsen. Fluoxetine and quinidine can be useful for slow channel syndrome, and albuterol for mutations affecting acetylcholinesterase, DOK7, rapsyn, and agrin. Additionally, ephedrine and 3,4-diaminopyridine (3,4-DAP) may be of benefit in some forms of CMS.

LEMS is a presynaptic disorder of the NMJ that causes skeletal muscle weakness; however, the pattern of involvement differs from that in MG. The proximal muscles of the lower limbs are most commonly affected, although other muscles may be involved as well. Cranial and bulbar weakness, including ptosis of the eyelids, diplopia, dysarthria, and dysphagia may occur but are not typically the presenting or prominent symptoms. However, LEMS can be further distinguished from MG because patients with LEMS often have depressed or absent reflexes and experience autonomic symptoms such as dry mouth, orthostasis, and impotence **(Chap. 451)**. Nerve stimulation produces an initial low-amplitude compound muscle action potential and, at low rates of repetitive stimulation (2–3 Hz), a decremental response as seen in MG; however, at high rates (20–50 Hz) or following brief exercise, incremental responses occur. LEMS is caused by autoantibodies directed against P/Q-type calcium channels at the presynaptic motor nerve terminals detected in ~85% of LEMS patients. These autoantibodies impair the release of ACh from nerve terminals. In young adults, particularly women, LEMS is less commonly associated with an underlying cancer. However, in older adults, LEMS is associated with malignancy, most commonly small-cell lung cancer (SCLC), and virtually all of these patients have P/Q-type calcium channel autoantibodies. The tumor cells may express calcium channels that stimulate the autoimmune response. Initial management requires comprehensive evaluation for malignancy and reassessment if the initial malignancy evaluation is negative. Treatment of LEMS symptoms involves therapy first with 3,4-DAP and pyridostigmine. 3,4-DAP acts by blocking potassium channels, which results in prolonged depolarization of the motor nerve terminals, thus enhancing ACh release. Pyridostigmine prolongs the action of ACh, allowing repeated interactions with AChRs. If symptoms are severe or life-threatening or if symptomatic therapy is insufficient, immunomodulatory therapy including IVIg or plasma exchange can be used.

Botulism **(Chap. 158)** is due to potent bacterial toxins produced by any of eight different strains of *Clostridium botulinum*. The toxins enzymatically cleave specific proteins essential for the release of ACh from the motor nerve terminal, thereby interfering with neuromuscular transmission. Most commonly, botulism is caused by ingestion of improperly prepared food containing toxin. Rarely, the nearly ubiquitous spores of *C. botulinum* may germinate in wounds. In infants, the spores may germinate in the gastrointestinal (GI) tract and release toxin, causing muscle weakness. Patients present with myasthenia-like bulbar weakness (e.g., diplopia, dysarthria, dysphagia) and lack sensory symptoms and signs. Weakness may generalize to the limbs and may result in respiratory failure. Reflexes are present early, but they may be diminished as the disease progresses. Mentation is normal. Autonomic findings include paralytic ileus, constipation, urinary retention, dilated or poorly reactive pupils, and dry mouth. The demonstration of toxin in serum by bioassay is definitive, but the results usually take a relatively long time to be completed and may be negative. Nerve stimulation studies reveal reduced compound muscle action potential (CMAP) amplitudes that increase following high-frequency repetitive stimulation. Treatment includes ventilatory support and aggressive inpatient supportive care (e.g., nutrition, deep-vein thrombosis prophylaxis) as needed. Antitoxin should be given as early as possible to be effective and can be obtained through the Centers for Disease Control and Prevention. A preventive vaccine is available for laboratory workers or other highly exposed individuals.

Hyperthyroidism is readily diagnosed or excluded by tests of thyroid function, which should be carried out routinely in patients with suspected MG. Abnormalities of thyroid function (hyper- or hypothyroidism) may increase myasthenic weakness. Diplopia resembling that in MG may occasionally be due to an intracranial mass lesion that compresses nerves to the extraocular muscles (e.g., sphenoid ridge meningioma), but magnetic resonance imaging (MRI) of the head and orbits usually reveals the lesion.

Progressive external ophthalmoplegia is a rare condition resulting in weakness of the extraocular muscles and often symmetric ptosis, which may be accompanied by weakness of the proximal muscles of the limbs and other systemic features. Most patients with this condition have mitochondrial disorders that can be detected by genetic testing or with muscle biopsy **(Chap. 460)**.

Search for Associated Conditions (Table 459-3) Myasthenic patients have an increased incidence of several associated disorders. Thymic abnormalities occur in ~75% of AChR antibody–positive patients, as noted above. Neoplastic change (thymoma) may produce enlargement of the thymus, which is detected by chest computed tomography (CT) or MRI. A thymic shadow on CT scan may normally be present through young adulthood, but enlargement of the thymus in a patient age >40 years is highly suspicious for thymoma. Approximately 10–15% of patients with MG have thymoma, and therefore, chest imaging to evaluate this possibility is performed at diagnosis.

Hyperthyroidism occurs in 3–8% of patients and may aggravate the myasthenic weakness. Thyroid function tests should be obtained in all patients with suspected MG. Other autoimmune disorders, most commonly systemic lupus erythematosus and rheumatoid arthritis, can coexist with MG; associations also occur with neuromyelitis optica, multiple sclerosis, neuromyotonia, Morvan's syndrome (encephalitis, insomnia, confusion, hallucinations, autonomic dysfunction, and neuromyotonia), rippling muscle disease, granulomatous myositis/myocarditis, and chronic inflammatory demyelinating polyneuropathy.

An infection of any kind can exacerbate typical MG and should be sought carefully in patients with relapses. Because of the side effects of glucocorticoids and other immunotherapies used in the treatment of MG, a thorough medical investigation should be undertaken, searching specifically for evidence of chronic or latent infection (such as tuberculosis or hepatitis), hypertension, diabetes, renal disease, and glaucoma.

TREATMENT

Myasthenia Gravis

The prognosis of MG has improved strikingly as a result of advances in treatment. Nearly all myasthenic patients can be returned to full productive lives with proper therapy. Common treatments for MG include anticholinesterase medications, glucocorticoids and other immunosuppressive agents, thymectomy, plasmapheresis, IVIg, rituximab, and the recently approved complement inhibitors and neonatal Fc receptor (FcRn) antagonists **(Fig. 459-2)**.

ANTICHOLINESTERASE MEDICATIONS

Anticholinesterase medication produces at least partial improvement in most myasthenic patients, although improvement is complete in only a few. Patients with anti-MuSK MG generally obtain less benefit from anticholinesterase agents than those with AChR antibodies and may actually worsen. Pyridostigmine is the most widely used anticholinesterase drug and is initiated at a dosage of 30–60 mg three to four times daily. The beneficial action of oral pyridostigmine begins within 15–30 min and lasts for 3–4 h, but individual responses vary. The frequency and amount of the dose should be tailored to the patient's individual requirements throughout the day. For example, patients with weakness in chewing and swallowing may benefit by taking the medication before meals so that peak strength coincides with mealtimes. Long-acting pyridostigmine may occasionally be useful to get the patient through the night but should not be used for daytime medication because of variable absorption. The maximum useful dose of pyridostigmine

TABLE 459-2 Congenital Myasthenic Syndromes (CMS)

CMS SUBTYPE	GENE	CLINICAL FEATURES	ELECTROPHYSIOLOGIC FEATURES	RESPONSE TO ACHE INHIBITORS	TREATMENT
Presynaptic Disorders					
CMS with paucity of ACh release	*CHAT; CHT*	AR; early onset, respiratory failure at birth, episodic apnea, improvement with age	Decremental response to RNS	Improve	AChE inhibitors; 3,4-DAP
Synaptic Disorders					
AChE deficiency	*COLQ*	AR; early onset; variable severity; axial weakness with scoliosis; apnea; +/– EOM involvement, slow or absent pupillary responses	After discharges on nerve stimulation and decrement on RNS	Worsen	Albuterol; ephedrine; 3,4-DAP; avoid AChE inhibitors
Postsynaptic Disorders Involving AChR Deficiency or Kinetics					
Primary AChR deficiency	AChR subunit genes	AR; early onset; variable severity; fatigue; typical MG features	Decremental response to RNS	Improve	AChE inhibitors; 3,4-DAP
AChR kinetic disorder: slow channel syndrome	AChR subunit genes	AD; onset childhood to early adult; weak forearm extensors and neck; respiratory weakness; variable severity	After discharges on nerve stimulation and decrement on RNS	Worsen	Fluoxetine and quinidine; avoid AChE inhibitors
AChR kinetic disorder: fast channel syndrome	AChR subunit genes	AR; early onset; mild to severe; ptosis, EOM involvement; weakness and fatigue	Decremental response to RNS	Improve	AChE inhibitors; caution with 3,4-DAP
Postsynaptic Disorders Involving Abnormal Clustering/Function of AChR					
	DOK 7	AR; limb girdle weakness with ptosis but no EOM involvement	Decremental response to RNS	Variable	Albuterol; ephedrine; may worsen with AChE inhibitors
	Rapsyn	AR; early onset with hypotonia, respiratory failure, and arthrogryposis at birth to early adult onset resembling MG	Decremental response to RNS	Variable	Albuterol
	Agrin	AR; limb girdle or distal weakness, apnea	Decremental response to RNS	Variable	Albuterol; may worsen with AChE inhibitors
	MuSK	AR; congenital or childhood onset of ptosis, EOM and progressive limb girdle weakness	Decremental response to RNS	Variable	Variable response to AChE inhibitors and 3,4-DAP Positive response to albuterol
	LRP4	AR; congenital onset with hypotonia; ventilatory failure, mild ptosis, and EOM weakness	Decremental response to RNS	Worsen	Worsen with AChE inhibitors
Other Postsynaptic Disorders					
Limb-girdle CMS with tubular aggregates	*GFPT1; DPAGT1; ALG2; ALG14; DPAGT1*	AR; limb-girdle weakness usually without ptosis or EOM weakness; onset in infancy or early adult	Decremental response to RNS	Variable	Albuterol; ephedrine; variable response to AChE inhibitors and 3,4-DAP; albuterol
Congenital muscular dystrophy with myasthenia	*Plectin*	AR; infantile or childhood onset of generalized weakness including ptosis and EOM; epidermolysis bullosa simplex; elevated CK	Decremental response to RNS	Variable	No response to AChE and 3,4-DAP

Abbreviations: ACh, acetylcholine; AChE, acetylcholinesterase; AChR, acetylcholine receptor; AD, autosomal dominant; AR, autosomal recessive; CHAT, choline acetyl transferase; CHT, sodium-dependent high-affinity choline transport 1; CK, creatine kinase; CMA, congenital myasthenic syndrome; COLQ, collaganic tail of endplate acetylcholinesterase; 3,4-DAP, 3,4-diaminopyridine; Dok7, downstream of tyrosine kinase 7; DPAGT1, UDP-N-acetylglucosamine-dolichyl-phosphate N-acetylglucosamine phosphotransferase; EOM, extraocular muscle; GFPT1, glutamine-fructose-6-phosphate aminotransferase 1; LRP4, lipoprotein receptor-related protein 4; MG, myasthenia gravis; MuSK, muscle specific kinase; RNS, repetitive nerve stimulation.

Source: Reproduced with permission from AA Amato, et al (eds): *Amato and Russell's Neuromuscular Disorders*, 3rd ed. New York: McGraw Hill; 2025.

rarely exceeds 360–480 mg daily. Overdosage with anticholinesterase medication may cause increased weakness and other side effects. In some patients, muscarinic side effects of the anticholinesterase medication (diarrhea, abdominal cramps, excess salivation, nausea) may limit the dose tolerated. Atropine/diphenoxylate or loperamide is useful for the treatment of gastrointestinal symptoms.

THYMECTOMY

Two separate issues should be distinguished: (1) surgical removal of thymoma, and (2) thymectomy as a treatment for MG. Surgical removal of a thymoma is necessary because of the possibility of local tumor spread, although most thymomas are histologically benign. A large international study (the MGTX trial) of extended transsternal thymectomy in nonthymomatous, AChR antibody–positive, generalized MG demonstrated that participants who underwent thymectomy had improved strength and function, required less prednisone and fewer additions of second-line agents (e.g., azathioprine), and had fewer hospitalizations for exacerbations lasting at least 5 years. Whether or not less invasive thymectomy provides identical benefit is unknown; however, less invasive techniques are now used in most thymectomies at many institutions. Importantly, patients with ocular myasthenia, MuSK-positive, and seronegative MG were all excluded from the MGTX study; retrospective and anecdotal evidence suggests that these patients may not benefit from thymectomy. Thymectomy should never be carried out as an emergency procedure, but only when the patient is adequately prepared. If necessary, treatment with IVIg or plasmapheresis may be used before surgery to maximize strength in weak patients.

IMMUNOTHERAPY

The choice of immunotherapy should be guided by the relative benefits and risks for the individual patient and the urgency of treatment. It is helpful to develop a treatment plan based on short-term, intermediate-term, and long-term objectives. For example,

TABLE 459-3 Disorders Associated with Myasthenia Gravis and Recommended Laboratory Tests

Associated disorders

- *Disorders of the thymus*: thymoma, hyperplasia
- *Other autoimmune neurologic disorders:* chronic inflammatory demyelinating polyneuropathy, neuromyelitis optica
- *Other autoimmune disorders*: Hashimoto's thyroiditis, Graves' disease, rheumatoid arthritis, systemic lupus erythematosus, skin disorders, family history of autoimmune disorder
- *Disorders or circumstances that may exacerbate myasthenia gravis*: hyperthyroidism or hypothyroidism, occult infection, medical treatment for other conditions (see Table 459-5)
- *Disorders that may interfere with therapy*: tuberculosis, diabetes, peptic ulcer, gastrointestinal bleeding, renal disease, hypertension, asthma, osteoporosis, obesity

Recommended laboratory tests or procedures

- CT or MRI of chest
- Tests for antinuclear antibodies, rheumatoid factor
- Thyroid function tests
- Testing for tuberculosis
- Fasting blood glucose, hemoglobin A_{1c}
- Pulmonary function tests
- Bone densitometry

Abbreviations: CT, computed tomography; MRI, magnetic resonance imaging.

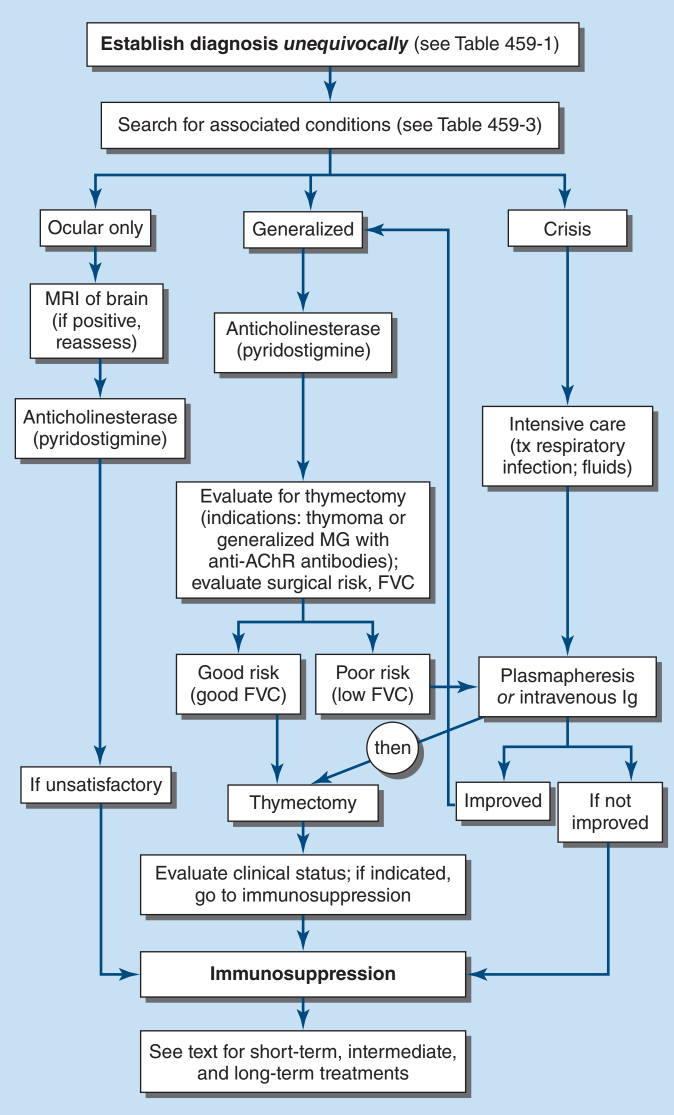
Establish diagnosis ***unequivocally*** (see Table 459-1)
Search for associated conditions (see Table 459-3)
Ocular only
Generalized
Crisis
MRI of brain (if positive, reassess)
Anticholinesterase (pyridostigmine)
Anticholinesterase (pyridostigmine)
Intensive care (tx respiratory infection; fluids)
Evaluate for thymectomy (indications: thymoma or generalized MG with anti-AChR antibodies); evaluate surgical risk, FVC
Good risk (good FVC)
Poor risk (low FVC)
Plasmapheresis *or* intravenous Ig
then
If unsatisfactory
Thymectomy
Improved
If not improved
Evaluate clinical status; if indicated, go to immunosuppression
Immunosuppression
See text for short-term, intermediate, and long-term treatments

FIGURE 459-2 Algorithm for the management of myasthenia gravis. FVC, forced vital capacity; MRI, magnetic resonance imaging.

if immediate improvement is essential typically because of the severity of weakness, IVIg should be administered or plasmapheresis should be undertaken as "rescue" therapy. For the intermediate term, glucocorticoids, cyclosporine or tacrolimus, rituximab, and the newer complement inhibitors and FcRn antagonists generally produce clinical improvement within a period of 1–3 months. They can be used for bridging until other immunotherapies become effective or in refractory patients. The beneficial effects of other nonsteroidal immunosuppressive therapies, azathioprine and mycophenolate mofetil, usually begin after many months (and as long as 1–1.5 years), However, these drugs have advantages over glucocorticoids for the long-term treatment of patients with MG. Rituximab is highly effective in patients with MuSK antibody–positive MG.

Glucocorticoid Therapy Glucocorticoids, when used properly, produce improvement in myasthenic weakness in the great majority of patients. To minimize adverse side effects, prednisone should be given in a single morning dose rather than in divided doses throughout the day. In patients with only ocular or mild generalized weakness, the initial dose can be relatively low (15–25 mg/d). The dose is increased stepwise, as tolerated by the patient (usually by 5 mg/d at 7- to 14-day intervals), until there is marked clinical improvement or a dose of 50–60 mg/d is reached. The full effect of a particular dose of prednisone often takes 2–3 weeks to observe. In patients with more severe weakness and those already in the hospital and/or intubated, starting at a high dose is reasonable, typically after pretreatment with IVIg or plasma exchange to protect against early steroid-associated worsening. Patients are maintained for about a month on the dose that controls their symptoms, and then the dosage is slowly tapered (no faster than 10 mg a month until on 20 mg daily and then by 2.5–5 mg every 1–3 months until on 10 mg daily, and more slowly thereafter) to determine the minimum effective dose. Close monitoring both for side effects and for efficacy is essential. Some patients can be managed without the addition of other immunotherapies. Patients on long-term glucocorticoid therapy must be followed carefully to prevent or treat adverse side effects. The most common errors include (1) an insufficient duration or dose of prednisone—improvement may be delayed and gradual; (2) tapering the dosage too early, too rapidly, or excessively; and (3) lack of attention to prevention and treatment of side effects.

Other Immunotherapies Mycophenolate mofetil, azathioprine, cyclosporine, tacrolimus, rituximab, and rarely, cyclophosphamide are effective in many patients, either alone or in various combinations.

Mycophenolate mofetil is widely used because of its presumed effectiveness and relative lack of side effects. A dose of 1–1.5 g bid is recommended. Its mechanism of action involves inhibition of purine synthesis by the de novo pathway. Since lymphocytes have only the de novo pathway, but lack the alternative salvage pathway that is present in all other cells, mycophenolate inhibits proliferation of lymphocytes but not proliferation of other cells. It does not kill or eliminate preexisting autoreactive lymphocytes, and therefore, clinical improvement may be delayed for many months to a year, until the preexisting autoreactive lymphocytes die spontaneously. The advantage of mycophenolate lies in its relative paucity of adverse side effects. The primary side effect is diarrhea or other GI symptoms. Rare side effects are development of leukopenia and very small risks of malignancy or progressive multifocal leukoencephalopathy inherent in nearly all immunosuppressive treatments. Although two published studies did not show positive outcomes, most experts attribute the negative results to flaws in the trial designs, and mycophenolate is widely used and supported in many guidelines for long-term treatment of myasthenic patients.

Azathioprine has long been used for MG, and a randomized, clinical trial demonstrated that it was effective in reducing the dosage of prednisone necessary to control MG symptoms. However, the beneficial effect can take a year or more to become evident.

Approximately 10–15% of patients are unable to tolerate azathioprine because of idiosyncratic reactions consisting of flulike symptoms (e.g., fever and malaise, abdominal pain), bone marrow suppression, or abnormalities of liver function. An initial dose of 50 mg/d is given for about a week to test for these side effects. If this dose is tolerated, it is increased by 50 mg weekly to 150 mg daily. Some patients require additional increases to reach a dose of ~2–3 mg/kg of total body weight or until the white blood count falls to 3000–4000/μL. Allopurinol should never be used in combination with azathioprine because the two drugs share a common degradation pathway; the result may be severe bone marrow suppression due to increased effects of the azathioprine.

The calcineurin inhibitors cyclosporine and tacrolimus are effective in MG and work more rapidly than azathioprine and mycophenolate. However, both, and cyclosporin in particular, are associated with more frequent severe side effects including hypertension, nephrotoxicity, and drug interactions. The usual dose of cyclosporine is 4–5 mg/kg per d, and the average dose of tacrolimus is 0.07–0.1 mg/kg per d, given in two equally divided doses. “Trough” blood levels are measured 12 h after the evening dose. The therapeutic range for the trough level of cyclosporine is 150–200 ng/L, and for tacrolimus, it is 5–15 ng/L.

Rituximab is a monoclonal antibody that binds to the CD20 molecule on B lymphocytes. It is widely used for the treatment of B-cell lymphomas and has also proven successful in the treatment of several autoimmune diseases including rheumatoid arthritis, pemphigus, and some IgM-related neuropathies. Rituximab can induce prolonged remissions in MuSK antibody–positive MG, which was previously more difficult to treat than anti-AChR–positive MG. We treat MuSK antibody–positive MG patients with 1 g IV on two occasions 2 weeks apart. Periodically, a repeat course needs to be administered; some MuSK patients can go up to 2–3 years between infusions.

A large National Institutes of Health–sponsored randomized trial of rituximab in AChR antibody–positive generalized MG failed to demonstrate efficacy, but many of the participants had long-standing MG that failed other therapies. However, a more recent randomized, placebo-controlled trial from Sweden of new-onset MG (<1 year) reported that a single infusion of 500 mg IV rituximab resulted in greater likelihood of participants achieving minimal MG manifestations and reduced need for rescue medications compared with placebo at 48 weeks. Further studies are needed, however, to determine how long this improvement may last and the need for retreatment.

For the rare refractory MG patient, a course of high-dose cyclophosphamide may induce long-lasting benefit. At high doses, cyclophosphamide eliminates mature lymphocytes but spares hematopoietic precursors (stem cells), because they express the enzyme aldehyde dehydrogenase, which hydrolyzes cyclophosphamide. This procedure is reserved for refractory patients and should be administered only in a facility fully familiar with this approach. Maintenance immunotherapy after treatment is usually required to sustain the beneficial effect.

NEWLY APPROVED TREATMENTS

Special attention needs to be given to the newly approved therapies for MG. Complement inhibitors and FcRn inhibitors have revolutionized treatment of patients with MG **(Table 459-4)**. Because they work quickly in most patients, they may be used as bridge therapies until other immunotherapies can “kick in” or in those who are refractory to standard treatments.

Complement Inhibitors Currently, three complement inhibitors are U.S. Food and Drug Administration (FDA) approved for AChR antibody–positive generalized MG based on positive clinical trial results. Most patients who will improve on these agents will do so within the first 12 weeks, and improvement is appreciated in many within the first 1–4 weeks. These drugs each work by inhibiting the cleavage of C5 in the terminal complement cascade. Eculizumab was shown to be effective in a positive phase 3 study, which led to FDA approval in 2017. Subsequently, ravulizumab was approved in 2021. Both eculizumab and ravulizumab are monoclonal antibodies given intravenously; ravulizumab has a longer effect. Zilucoplan is the latest complement inhibitor and was approved in 2023. Unlike eculizumab and ravulizumab, zilucoplan is a subcutaneously administered macrocyclic peptide inhibitor of C5. Because it is not a monoclonal antibody like eculizumab and ravulizumab, it can be coadministered with plasma exchange, IVIg, or FcRn antagonists. An additional benefit is that patients can self-administer zilucoplan with a prefilled syringe. In practice, we typically reassess efficacy at 12 weeks in patients treated with C5 inhibitors and decide whether or not to continue treatment.

Complement inhibition increases the risk of meningococcal infection. Therefore, a first series of vaccinations with both quadrivalent and MenB vaccines is given at least 14 days prior to initiation of treatment and then again 1–2 months later. Those patients in whom treatment needs to be started sooner than this initial vaccination series is complete should receive antibiotic prophylaxis (penicillin). Vaccination reduces, but does not eliminate, the risk of meningitis. Physicians must enroll in drug-specific risk evaluation and mitigation strategy programs for all complement inhibitors and counsel patients regarding the risk and signs and symptoms of meningitis. Patients are recommended to carry a safety/alert wallet card.

Neonatal Fc Receptor (FcRn) Antagonists FcRns on endothelial cells salvage IgG and albumin from degradation by lysosomes, leading to longer IgG half-lives. Blocking the FcRn results in increased catabolism of IgG, thereby reducing IgG (and pathogenic antibody) levels. The potential benefits over plasma exchange include the ease of administration, increased availability, and reduced risk in patients with coagulopathies or limited peripheral venous access. Efgartigimod and rozanolixizumab are now approved for clinical use based on their efficacy in clinical trials (Table 459-4). Efgartigimod can be given intravenously or subcutaneously, whereas rozanolixizumab is given via a subcutaneous infusion. Importantly, only rozanolixizumab is approved for both anti-AChR and anti-MuSK generalized myasthenia. As with complement inhibitors, the FcRn inhibitors usually are effective within the first 3 months of treatment, again often within the first month. Side effects are somewhat variable and include increased risk of respiratory and urinary infections, headaches (including aseptic meningitis with rozanolixizumab), and hypoalbuminemia (FcRn also prevents lysosomes from degrading albumin). Comparative efficacy with one another or with complement inhibitors cannot be readily ascertained with existing clinical trial data. One potential benefit of FcRn antagonists over complement inhibitors is that guidance from clinical trials exists on how to ramp up or ramp down administration frequency based on clinical response.

PLASMAPHERESIS AND INTRAVENOUS IMMUNOGLOBULIN

Plasmapheresis has long been used therapeutically in MG. Plasma, which contains the pathogenic antibodies, is mechanically separated from the blood cells, which are then returned to the patient. A course of five or six exchanges (3–4 L per exchange) is generally administered over a 10- to 14-day period. Plasmapheresis produces a short-term reduction in anti-AChR antibodies, with clinical improvement in many patients. It is most useful as a temporary treatment in severely affected patients or to improve the patient’s condition prior to surgery (e.g., thymectomy).

The indications for the use of IVIg are the same as those for plasma exchange: to produce rapid improvement to help the patient through a difficult period of myasthenic weakness or prior to surgery. This treatment has the advantages of not requiring special equipment or large-bore venous access. The usual dose is 2 g/kg, which is typically administered over 2–5 days. Improvement occurs in ~70% of patients, beginning during treatment or within a week and continuing for weeks to months. The exact mechanism of

TABLE 459-4 Comparison of New Complement Inhibitors and FcRn Inhibitors for Generalized Myasthenia

DRUG/MECHANISM	TRIAL(S)	FDA APPROVED DOSING	CLINICAL TRIAL POPULATION	NOTES
Approved Complement Inhibitors				
Eculizumab (humanized monoclonal Ab anti-C5, inhibits terminal complement/MAC activation)	Phase 2 REGAIN : 26 weeks REGAIN open-label extension: 22.7 months (median), up to 3 years	Loading: 900 mg IV weekly × 4 Maintenance: 1200 mg IV on week 5 then q2 weeks	AChR ab + gMG (class II–IV) Refractory (at least 2 NSISTs or at least 1 NSIST and PLEX/IVIg) MGADL score ≥6	Did not reach statistical significance for primary MGADL endpoint Reached significance for multiple secondary endpoints Requires meningococcal vaccination
Ravulizumab (humanized monoclonal Ab anti-C5, inhibits terminal complement/MAC activation)	Phase 2 CHAMPION MG: 26 weeks CHAMPION MG open-label extension	Actual body weight–based dosing Loading: 40 to <60 kg: 2400 mg IV; 60 to <100 kg: 2700 mg IV; ≥100 kg: 3000 mg IV Maintenance (14 days after loading and then q8 weeks): 40 to <60 kg: 3000 mg IV; 60 to <100 kg: 3300 mg IV; ≥100 kg: 3600 mg IV	Adults with AChR ab + gMG (class II–IV) MGADL score ≥6	Requires meningococcal vaccination
Zilucoplan (synthetic macrocyclic peptide targeting C5/C5b, inhibits terminal complement/MAC activation)	Phase 2 RAISE: 12 weeks RAISE-XT open-label extension	Phase 3 dosing 0.3 mg/kg SC daily Label dosing (prefilled syringes): Actual body weight–based daily SC injections <56 kg: 16.6 mg daily; 56 kg to <77 kg: 23 mg; ≥77 kg: 32.4 mg	Adults with AChR ab + gMG (class II–IV) MGADL score ≥6 QMG ≥12	Requires meningococcal vaccination Self-administered SC
Approved FcRn Inhibitors				
Efgartigimod IV/SC (human anti-FcRn IgG1 Fc fragment; reduces autoantibody levels and IgG recycling)	Phase 2 ADAPT: 26 weeks ADAPT open-label extension: up to 3 years ADAPT-SC noninferiority study, open-label parallel-group: 12 weeks with open-label extension	Weight-based IV: 10 mg/kg IV (up to 1200 mg) weekly × 4 = 1 cycle Fixed dose SC: 1,008 mg SC weekly × 4 = 1 cycle	Adults with gMG regardless of Ab status, MGADL at least 5 with 50% nonocular	ADAPT was designed to observe wearing off – cycles repeated at return of symptoms and no sooner than every 8 weeks Number of infusions per cycle and time between cycles can be individualized (as was done in the open-label extension) Efgartigimod SC is not currently approved for self-injection; health care provider administered; refrigeration required
Rozanolixizumab (human anti-FcRn IgG4 monoclonal antibody; reduces autoantibody levels and IgG recycling)	Phase 2 MycarinG: 18 weeks Open-label extension: completed	Phase 3 included 7 and 10 mg/kg given as SC infusion weekly for 6 weeks followed by 8 weeks off. Patients averaged 4 treatment cycles per year (range 1–7) Clinical dosing: <50 kg: 420 mg; 50 to <100 kg: 560 mg; ≥100 kg: 840 mg given as a weekly health care provider–administered SC infusion for 6 weeks (1 cycle)	Adults with AChR or MuSK Ab + gMG (11% of participants) MGADL score ≥3 QMG score ≥11	Headache occurred in 38–45% of treatment group and 19% of placebo, including rare aseptic meningitis Infection rate higher in 10 mg/kg dosing group; efficacy equivalent Shorter mean disease duration than other phase 3 trials (5–6 years) Hypoalbuminemia and peripheral edema

Abbreviations: AChR, acetylcholine receptor; Ab, antibody; FcRn, neonatal Fc receptor; FDA, Food and Drug Administration; gMG, generalized myasthenia gravis; IVIg, intravenous immunoglobulin; MAC, membrane attack complex; MGADL, Myasthenia Gravis Activities of Daily Living; MuSK, muscle-specific tyrosine kinase; NSIST, nonsteroidal immunosuppressant therapy; PLEX, plasma exchange; QMG, quantitative myasthenia gravis.

Source: Reproduced with permission from AA Amato et al (eds): *Amato and Russell's Neuromuscular Disorders*, 3rd ed. New York: McGraw Hill; 2025.

action of IVIg in MG is unknown; the treatment has no consistent long-term effect on the measurable amount of circulating AChR antibody. Adverse reactions are generally not serious but may include headache, fluid overload, and rarely aseptic meningitis, renal failure, hemolytic anemia, and embolic or thrombotic events. IVIg or plasma exchange is occasionally used in combination with other immunosuppressive therapy for maintenance treatment of difficult MG, though this is less common in the contemporary era since the advent of C5 inhibitors and FcRn antagonists.

INVESTIGATIONAL TREATMENTS

Several trials of different complement and FcRn inhibitors are underway. Inhibitors of interleukin 6 and CD19 targets on B cells are also being studied. Notably, CD19-targeting chimeric antigen receptor (CAR) therapy and chimeric autoantibody receptor T (CAART) therapy targeting the antibody ligand on T cells are also in clinical trials for MG.

MANAGEMENT OF MYASTHENIC CRISIS

Myasthenic crisis is defined as an exacerbation of weakness sufficient to endanger life; it usually includes ventilatory failure caused by diaphragmatic and intercostal muscle weakness. Treatment should be carried out in intensive care units staffed with teams experienced in the management of MG. The possibility that deterioration could be due to excessive anticholinesterase medication ("cholinergic crisis") is unlikely given that very high doses of cholinesterase inhibitors are rarely used but is considered in the differential. The most common cause of crisis is intercurrent infection. This should be treated immediately because the mechanical and immunologic defenses of the patient can be assumed to be compromised. The myasthenic patient with fever and early infection should be treated like other immunocompromised patients. Early and effective antibiotic therapy, ventilatory assistance, and pulmonary physiotherapy are essentials of the treatment program. As discussed above, plasmapheresis or IVIg is frequently helpful in hastening recovery.

MANAGEMENT OF MYASTHENIA ASSOCIATED WITH IMMUNE CHECKPOINT INHIBITOR THERAPY

MG is a rare complication of ICI therapy for cancer. It can develop de novo or as an exacerbation of preexisting diagnosed or undiagnosed disease. Patients usually manifest with ocular, bulbar, neck, and respiratory weakness within the first one to four cycles of ICI therapy. Compared to idiopathic MG, ICI-associated MG is more likely to be seronegative and overlap with myositis and myocarditis. The mortality rate is 20–50%, most often because of severe myocarditis. Importantly, ICI-associated myositis is more common than immune-related MG. It can resemble MG clinically, with prominent or exclusively ocular weakness but without evidence for a decrementing response on repetitive nerve stimulation. The mainstay of treatment for ICI-associated MG is glucocorticoids, including IV solumedrol (which differs from idiopathic MG), and with plasma exchange or IVIg added for severe weakness. Complement inhibitors have recently been reported as effective in AChR antibody–positive ICI-associated MG with myositis and myocarditis.

DRUGS TO AVOID IN MYASTHENIC PATIENTS

Many drugs can potentially exacerbate weakness in patients with MG **(Table 459-5)**. As a rule, the listed drugs should be avoided whenever possible.

TABLE 459-5 Drugs with Interactions in Myasthenia Gravis (MG)

Drugs That May Exacerbate Weakness in Patients with MG
Antibiotics
Aminoglycosides: e.g., streptomycin, tobramycin, kanamycin
Quinolones: e.g., ciprofloxacin, levofloxacin, ofloxacin, gatifloxacin
Macrolides: e.g., erythromycin, azithromycin
Nondepolarizing muscle relaxants for surgery
D-Tubocurarine (curare), pancuronium, vecuronium, atracurium
Beta-blocking agents
Propranolol, atenolol, metoprolol
Local anesthetics and related agents
Procaine, Xylocaine in large amounts
Procainamide (for arrhythmias)
Botulinum toxin
Botox exacerbates weakness
Quinine derivatives
Quinine, quinidine, chloroquine, mefloquine (Lariam)
Magnesium
Decreases acetylcholine release
Penicillamine
May cause MG
Checkpoint inhibitors
May cause MG and other autoimmune neuromuscular disorders (e.g., myositis, inflammatory neuropathy)
Drugs with Important Interactions in MG
Cyclosporine and tacrolimus
Broad range of drug interactions, which may raise or lower levels
Azathioprine
Avoid allopurinol—combination may result in myelosuppression

PATIENT ASSESSMENT

To evaluate the effectiveness of treatment as well as drug-induced side effects, it is important to assess the patient's clinical status systematically at baseline and on repeated interval examinations. Following the patient with spirometry with determination of forced vital capacity and mean inspiratory and expiratory pressures is important.

PROGNOSIS

Approximately 20% of patients with MG achieve a sustained remission and can be tapered off all immunotherapies. There does not appear to be a correlation between disease severity and likelihood of remission. Thymectomy may increase the chance of achieving remission in anti-AChR MG, but the large, randomized MGTX trial was too short in duration to examine this endpoint; rather, the results revealed only that thymectomy was efficacious and led to less use of glucocorticoids and second-line agents. Mortality from MG diminished greatly during the twentieth century, changing from a "grave" illness with mortality of nearly 70% a century ago, to 2–30% by the 1950s, with contemporary estimates in the 1–5% range. Anti-MuSK patients generally were more difficult to treat than anti-AChR MG in the past. However, recent series suggest that rituximab is effective in this subgroup, thereby reducing these risks and improving the prognosis. Nonparaneoplastic LEMS is usually responsive to immunotherapy and symptomatic treatment with pyridostigmine and 3,4-DAP. In older adults, LEMS is most often paraneoplastic, and screening for an underlying tumor is indicated **(Chap. 99)**. Recent studies suggest that survival in patients with LEMS has improved, for uncertain reasons and likely not due to earlier diagnosis and treatment of the tumor. There is wide variability in age of onset, severity, and prognosis of the many types of CMS.

GLOBAL ISSUES

The incidence of MG and its subtypes varies in different populations, for example, occurring in ~2–10/10^6 individuals in the United States and the Netherlands and up to 20/10^6 individuals in Spain. Estimates of prevalence in different parts of the world range widely from 2–360/10^6. The age of onset may also be influenced by geographic and/or ethnic differences. Juvenile-onset MG is uncommon in Western populations but may represent more than half of cases in Asians. MuSK MG appears to be more common in the Mediterranean area of Europe than in northern Europe and is also more common in the northern regions of East Asia than in the southern regions. A concern during the COVID-19 pandemic is whether MG patients on immunosuppressive therapies might be at increased risk of infection or developing a more severe course. Furthermore, flares of MG can be triggered by infection, and contracting COVID-19 may lead to an exacerbation, including MG crisis. We have not reduced the dosage of immunosuppressive medications in MG patients who are doing well but have been more likely to manage worsening disease by treating with IVIg rather than increasing the dosage of, or adding new, immunosuppressive agents. Patients are strongly advised to receive the COVID-19 vaccine, wear masks, and maintain social distancing, particularly when infection levels are high in their communities. An international panel published guidelines for management of MG patients during the pandemic.

FURTHER READING

Amato AA et al: *Amato and Russell's Neuromuscular Disorders*, 3rd ed. New York, McGraw Hill, 2025.

Guidon AC: Lambert-Eaton myasthenic syndrome, botulism, and immune checkpoint inhibitor-related myasthenia gravis. Continuum (Minneap Minn) 25:1785, 2019.

Gwathmey KG et al: How should newer therapeutic agents be incorporated into the treatment of patients with myasthenia gravis? Muscle Nerve 69:389, 2024.

Hehir MK 2nd, Li Y: Diagnosis and management of myasthenia gravis. Continuum (Minneap Minn) 28:1615, 2022.

International MG/COVID-19 Working Group et al: Guidance for the management of myasthenia gravis (MG) and Lambert-Eaton myasthenic syndrome (LEMS) during the COVID-19 pandemic. J Neurol Sci 412:116803, 2020.

Narayanaswami P et al: International consensus guidance for management of myasthenia gravis: 2020 update. Neurology 96:114, 2021.

Ohno K et al: Clinical and pathologic features of congenital myasthenic syndromes caused by 35 genes: A comprehensive review. Int J Mol Sci 24:3730, 2023.

Piehl F et al: Efficacy and safety of rituximab for new-onset generalized myasthenia gravis: The RINOMAX randomized clinical trial. JAMA Neurol 79:1105, 2022.

Sacca F et al: Efficacy of innovative therapies in myasthenia gravis: A systematic review, meta-analysis and network meta-analysis. Eur J Neurol 30:3854, 2023.

Salari N et al: Global prevalence of myasthenia gravis and the effectiveness of common drugs in its treatment: A systematic review and meta-analysis. J Transl Med 19:516, 2021.

Wolfe GI et al: Long-term effect of thymectomy plus prednisone versus prednisone alone in patients with non-thymomatous myasthenia gravis: 2-year extension of the MGTX randomised trial. Lancet Neurol 18:259, 2019.

VIDEO 459-1 Myasthenia gravis and other diseases of the neuromuscular junction.

460

Muscular Dystrophies and Other Muscle Diseases

Anthony A. Amato, Robert H. Brown, Jr.

Myopathies are disorders with structural changes or functional impairment of muscle and can be differentiated from other diseases of the motor unit (e.g., lower motor neuron or neuromuscular junction pathologies) by characteristic clinical and laboratory findings. **Myasthenia gravis and related disorders are discussed in Chap. 459; inflammatory myopathies are discussed in Chap. 377.**

CLINICAL FEATURES

The most important aspect of assessing individuals with neuromuscular disorders is taking a thorough history of the patient's symptoms, disease progression, and past medical and family history, as well as performing a detailed neurologic examination. Based on this and additional laboratory workup (e.g., serum creatine kinase [CK], electromyography [EMG]), one can usually localize the site of the lesion to muscle (as opposed to motor neurons, peripheral nerves, or neuromuscular junction) and the pattern of muscle involvement. It is this pattern of muscle involvement that is most useful in narrowing the differential diagnosis **(Table 460-1)**. Most myopathies present with proximal, symmetric limb weakness with preserved reflexes and sensation. However, asymmetric and predominantly distal weakness can be seen in some myopathies. An associated sensory loss suggests a peripheral neuropathy or a central nervous system (CNS) abnormality (e.g., myelopathy) rather than a myopathy. On occasion, disorders affecting the motor nerve cell bodies in the spinal cord (anterior horn cell disease), the neuromuscular junction, or peripheral nerves can mimic findings of myopathy.

Muscle Weakness Symptoms of muscle weakness can be either intermittent or persistent. Disorders causing *intermittent weakness* (Table 460-1 and **Fig. 460-1**) include myasthenia gravis, periodic paralyses (hypokalemic or hyperkalemic), and metabolic energy deficiencies of glycolysis (especially myophosphorylase deficiency), fatty acid utilization (carnitine palmitoyltransferase [CPT] deficiency), and some mitochondrial myopathies. The states of energy deficiency cause activity-related muscle breakdown accompanied by myoglobinuria.

Most muscle disorders cause *persistent weakness* (Table 460-1 and **Fig. 460-2**). In the majority of these, including most types of muscular dystrophy and inflammatory myopathies, the proximal muscles are weaker than the distal and are symmetrically affected, and the facial muscles are spared, a pattern referred to as *limb-girdle weakness*. The differential diagnosis is more restricted for other patterns of weakness. Facial weakness (difficulty with eye closure and impaired smile) and scapular winging **(Fig. 460-3)** are characteristic of facioscapulohumeral dystrophy (FSHD). Facial and distal limb weakness associated with hand grip myotonia is virtually diagnostic of myotonic dystrophy type 1. When other cranial nerve muscles are weak, causing ptosis or extraocular muscle weakness, the most important disorders to consider include neuromuscular junction disorders, oculopharyngeal muscular dystrophy, mitochondrial myopathies, or some of the congenital myopathies (Table 460-1). A pathognomonic pattern characteristic of inclusion body myositis is atrophy and weakness of the flexor forearm (e.g., wrist and finger flexors) and quadriceps muscles that is often asymmetric. Less frequently seen, but important diagnostically, are the axial myopathies that predominantly affect the paraspinal muscles and include dropped head syndrome indicative of selective neck extensor muscle weakness. The most important neuromuscular diseases associated with this axial muscle weakness include myasthenia gravis, amyotrophic lateral sclerosis, sporadic late-onset nemaline rod myopathy (SLONM), late-onset ryanodine receptor 1 (RyR1) myopathies, hyperparathyroidism, focal myositis, and some forms of inclusion body myopathy. A final pattern, recognized because of preferential distal extremity weakness, is seen in the distal myopathies.

It is important to examine functional capabilities to help disclose certain patterns of weakness (Table 460-1 and **Table 460-2**). The Gower sign **(Fig. 460-4)** is particularly useful. Observing the gait of an individual may disclose a hyperlordotic posture caused by combined trunk and hip weakness, frequently exaggerated by toe walking **(Fig. 460-5)**. A waddling gait is caused by the inability of weak hip muscles to prevent hip drop or hip dip. Hyperextension of the knee (genu recurvatum or back-kneeing) is characteristic of quadriceps muscle weakness, and a steppage gait, due to foot drop, accompanies distal weakness.

Any disorder causing muscle weakness may be accompanied by *fatigue*, referring to an inability to maintain or sustain a force (pathologic fatigability). This condition must be differentiated from asthenia, a type of fatigue caused by excess tiredness or lack of energy. Associated symptoms may help differentiate asthenia and pathologic fatigability. Asthenia is often accompanied by a tendency to avoid physical activities, complaints of daytime sleepiness, necessity for frequent naps, and difficulty concentrating on activities such as reading. There may be feelings of overwhelming stress and depression. In contrast, pathologic fatigability occurs in disorders of neuromuscular transmission and in disorders altering energy production, including defects in glycolysis, lipid metabolism, or mitochondrial energy production. Pathologic fatigability also occurs in chronic myopathies because of difficulty accomplishing a task with less muscle. Pathologic fatigability is accompanied by abnormal clinical or laboratory findings. Fatigue without those supportive features almost never indicates a primary muscle disease.

Muscle Pain (Myalgias), Cramps, and Stiffness Some myopathies can be associated with muscle pain, cramps, contractures, stiff or rigid muscles, or inability to relax the muscles (e.g., myotonia) (Table 460-1). *Muscle cramps* are abrupt in onset, short in duration, triggered by voluntary muscle contraction, and may cause abnormal posturing of the joint. Muscle cramps often occur in neurogenic disorders, especially motor neuron disease **(Chap. 448)**, radiculopathies, and polyneuropathies **(Chap. 457)**, but are not a feature of most primary muscle diseases.

A *muscle contracture* is different from a muscle cramp. In both conditions, the muscle becomes hard, but a contracture is associated with energy failure in glycolytic disorders. The muscle is unable to relax after an active muscle contraction. The EMG shows electrical silence. Confusion is created because contracture also refers to a muscle that cannot be passively stretched to its proper length (fixed contracture)

TABLE 460-1 Myopathies by Pattern of Weakness/Muscle Involvement

Proximal (Limb-Girdle) Weakness

- Most dystrophies (e.g., dystrophinophies, limb-girdle, myofibrillar myopathy, myotonic dystrophy type 2, rare FSHD)
- Congenital myopathies (e.g., central core, multiminicore, centronuclear, nemaline rod)
- Metabolic myopathies (e.g., glycogen and lipid storage diseases)
- Mitochondrial myopathies
- Inflammatory myopathies (DM, PM, IMNM, anti-synthetase syndrome)
- Toxic myopathies (see Table 460-6)
- Endocrine myopathies
- Neuromuscular junction disorders (myasthenia gravis, LEMS, congenital myasthenia, botulism, **see Chap. 459**)
- SLONM
- Late-onset central core (*RYR1* mutations)
- SLONM
- Metabolic (late-onset Pompe, McArdle disease, lipid storage, mitochondrial)
- Hyperparathyroidism/osteomalacia/vitamin D deficiency
- Myasthenia gravis

Distal Weakness

- Distal muscular dystrophies/myofibrillar myopathy (see Table 460-5)
- Congenital myopathies (e.g., late-onset centronuclear and nemaline rod myopathies)
- Oculopharyngeal distal myopathy
- Metabolic
  - Glycogen storage disease (e.g., brancher and debrancher deficiency, rarely McArdle disease)
  - Lipid storage disease (e.g., neutral lipid storage myopathy, multiacyldehydrogenase deficiency)
- NMJ disorders (e.g., rare myasthenia gravis and congenital myasthenia)

Proximal Arm/Distal Leg Weakness (Scapuloperoneal or Humeroperonal) Weakness

- Facioscapulohumeral muscular dystrophy (FSHD)
- Scapuloperoneal myopathy and neuropathy
- Myofibrillar myopathies
- Emery-Dreifuss muscular dystrophy (EDMD)
- Bethlem myopathy

Distal Arm/Proximal Leg Weakness

- Inclusion body myositis (usually wrist and finger flexors in arms, hip flexors and knee extensors in legs, and asymmetric)
- Myotonic dystrophy (uncommon presentation)

Axial Muscle Weakness

- Inflammatory (cervicobrachial myositis)
- sIBM and hIBM
- Myotonic dystrophy 2
- Isolated neck extensor myopathy/bent spine syndrome
- FSHD

Eye Muscle Weakness (Ptosis/Ophthalmoparesis)

- Ptosis without ophthalmoparesis
  - Myotonic dystrophy
  - Congenital myopathies
  - Neuromuscular junction disorders
- Ptosis with ophthalmoparesis
  - Oculopharyngeal dystrophy
  - Oculopharygeal distal myopathy
  - Mitochondrial myopathy
  - hIBM type 3
  - Neuromuscular junction disorders

Episodic Weakness or Myoglobinuria

- Related to exercise
  - Glycogenoses (e.g., McArdle disease, etc.)
  - Lipid disorders (e.g., CPT2 deficiency)
  - Mitochondrial myopathies (e.g., cytochrome B deficiency)
- Not related to exercise
  - *RYR1* mutations can cause malignant hyperthermia, episodic rhabdomyolysis/myoglobinuria, and atypical periodic paralysis
  - Other causes of malignant hyperthermia
- Drugs/toxins (e.g., statins)
  - Prolonged/intensive eccentric exercise
  - Inflammatory (e.g., PM/DM—rare, viral/bacterial infections)
- Delayed or unrelated to exercise
  - Periodic paralysis (e.g., hereditary hyper- or hypokalemic, thyrotoxic, associated renal tubular acidosis, acquired electrolyte imbalance)
  - NMJ disorders

Muscle Stiffness/Decreased Ability to Relax

- Myotonic dystrophy 1 and 2
- Myotonia congenita
- Paramyotonia congenita
- Hyperkalemic periodic paralysis with myotonia
- Potassium aggravated myotonia
- Schwartz-Jampel syndrome
- Other: rippling muscle disease (acquired and hereditary), acquired neuromyotonia (Isaacs' syndrome), stiff-person syndrome, Brody's disease

Abbreviations: DM, dermatomyositis; hIBM, hereditary inclusion body myopathy; IMNM, immune-mediated necrotizing myopathy; LEMS, Lambert-Eaton myasthenic syndrome; NMJ, neuromuscular junction; PM, polymyositis; sIBM, sporadic inclusion body myositis; SLONM, sporadic late-onset nemaline myopathy.

because of fibrosis. In some muscle disorders, especially in Emery-Dreifuss muscular dystrophy (EDMD) and Bethlem myopathy, fixed contractures occur early and represent distinctive features of the disease.

Myotonia is a condition of prolonged muscle contraction followed by slow muscle relaxation. It always follows muscle activation (action myotonia), usually voluntary, but may be elicited by mechanical stimulation (percussion myotonia) of the muscle. Myotonia typically causes difficulty in releasing objects after a firm grasp. In myotonic muscular dystrophy type 1 (DM1), distal weakness usually accompanies myotonia, whereas in DM2, proximal muscles are more affected. Myotonia also occurs with *myotonia congenita* (a chloride channel disorder), but in this condition, muscle weakness is usually not prominent. Myotonia may also be seen in individuals with sodium channel mutations (*hyperkalemic periodic paralysis* or *potassium-sensitive myotonia*). Another sodium channelopathy, *paramyotonia congenita* (PC), also is associated with muscle stiffness. In contrast to other disorders associated with myotonia in which the myotonia is eased by repetitive activity, PC is named for a paradoxical phenomenon whereby the myotonia worsens with repetitive activity. Potassium-aggravated myotonia is an allelic disorder in which myotonia is brought on by consumption of too much potassium-containing foods.

Muscle stiffness can refer to different phenomena. Some patients with inflammation of joints and periarticular surfaces feel stiff. This condition is different from the disorders of hyperexcitable motor nerves causing stiff or rigid muscles. In *stiff-person syndrome*, spontaneous discharges of the motor neurons of the spinal cord cause involuntary muscle contractions mainly involving the axial (trunk) and proximal lower extremity muscles. The gait becomes stiff and labored, with hyperlordosis of the lumbar spine. Superimposed episodic muscle spasms are precipitated by sudden movements, unexpected noises, and emotional upset. The muscles relax during sleep. Serum antibodies against glutamic acid decarboxylase are present in approximately two-thirds of cases. In *acquired neuromyotonia* (*Isaacs' syndrome*), there is hyperexcitability of the peripheral nerves manifesting as continuous muscle fiber activity in the form of widespread fasciculations and myokymia with impaired muscle relaxation. Muscles of the leg are stiff,

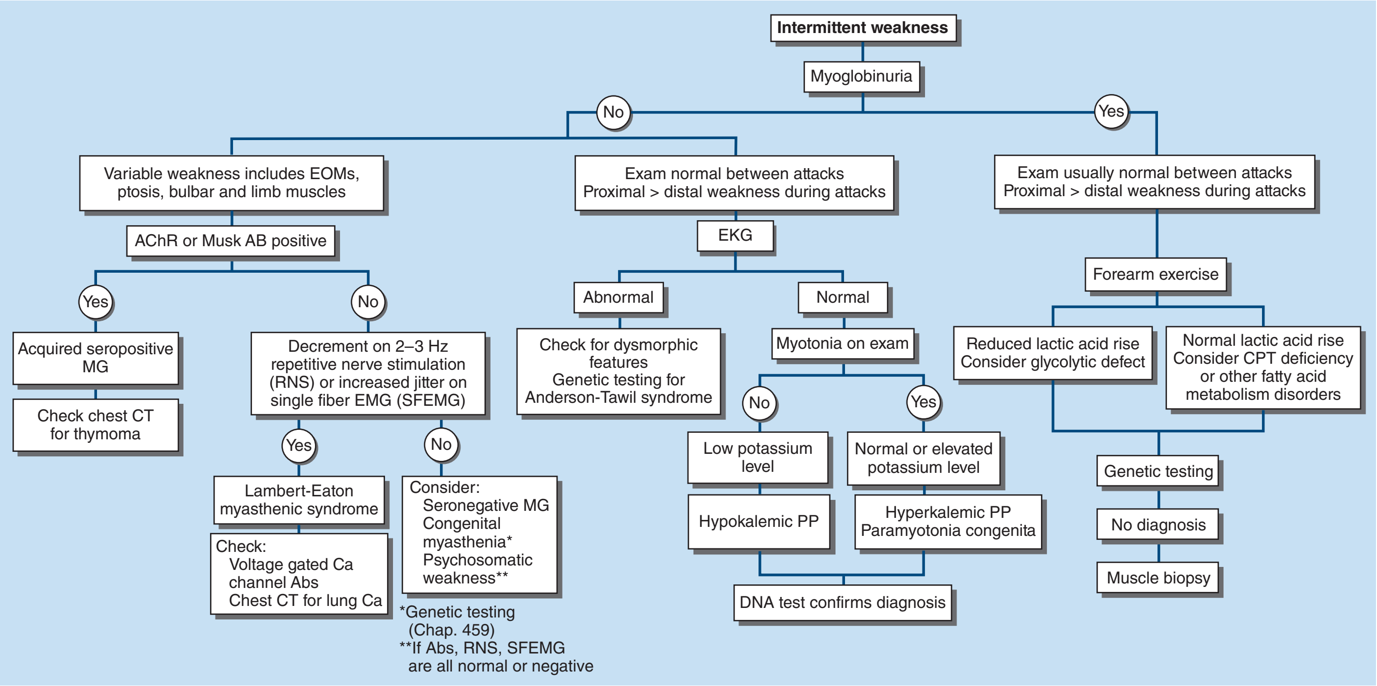
Intermittent weakness
Myoglobinuria
No
Yes
Variable weakness includes EOMs, ptosis, bulbar and limb muscles
AChR or Musk AB positive
Yes
No
Acquired seropositive MG
Check chest CT for thymoma
Decrement on 2–3 Hz repetitive nerve stimulation (RNS) or increased jitter on single fiber EMG (SFEMG)
Yes
No
Lambert-Eaton myasthenic syndrome
Check: Voltage gated Ca channel Abs Chest CT for lung Ca
Consider: Seronegative MG Congenital myasthenia* Psychosomatic weakness**
*Genetic testing (Chap. 459)
**If Abs, RNS, SFEMG are all normal or negative
Exam normal between attacks Proximal > distal weakness during attacks
EKG
Abnormal
Normal
Check for dysmorphic features Genetic testing for Anderson-Tawil syndrome
Myotonia on exam
No
Yes
Low potassium level
Normal or elevated potassium level
Hypokalemic PP
Hyperkalemic PP Paramyotonia congenita
DNA test confirms diagnosis
Exam usually normal between attacks Proximal > distal weakness during attacks
Forearm exercise
Reduced lactic acid rise Consider glycolytic defect
Normal lactic acid rise Consider CPT deficiency or other fatty acid metabolism disorders
Genetic testing
No diagnosis
Muscle biopsy

FIGURE 460-1 Diagnostic evaluation of intermittent weakness. AChR AB, acetylcholine receptor antibody; CPT, carnitine palmitoyltransferase; EKG, electrocardiogram; EMG, electromyogram; EOMs, extraocular muscles; MG, myasthenia gravis; PP, periodic paralysis.

and the constant contractions of the muscle cause increased sweating of the extremities. This peripheral nerve hyperexcitability is mediated by antibodies that target voltage-gated potassium channels.

There are two painful muscle conditions of particular importance, neither of which is associated with muscle weakness. *Fibromyalgia* is a common, yet poorly understood myofascial pain syndrome in which patients complain of severe muscle pain and tenderness, severe fatigue, and often poor sleep. Serum CK, erythrocyte sedimentation rate (ESR), EMG, and muscle biopsy are normal **(Chap. 385)**. *Polymyalgia rheumatica* occurs mainly in patients aged >50 years and is characterized by stiffness and pain in the shoulders, lower back, hips, and thighs **(Chap. 375)**. The ESR and CRP are elevated, while serum CK, EMG, and muscle biopsy are normal.

Muscle Enlargement and Atrophy In most myopathies, muscle tissue is replaced by fat and connective tissue, but the size of the muscle is usually not affected. However, in many limb-girdle muscular dystrophies, enlarged calf muscles are typical. The enlargement represents true muscle hypertrophy; thus, the term *pseudohypertrophy* should be avoided when referring to these patients. The calf muscles remain very strong even late in the course of these disorders. Muscle enlargement can also result from infiltration by sarcoid granulomas, amyloid deposits, bacterial and parasitic infections, and focal myositis.

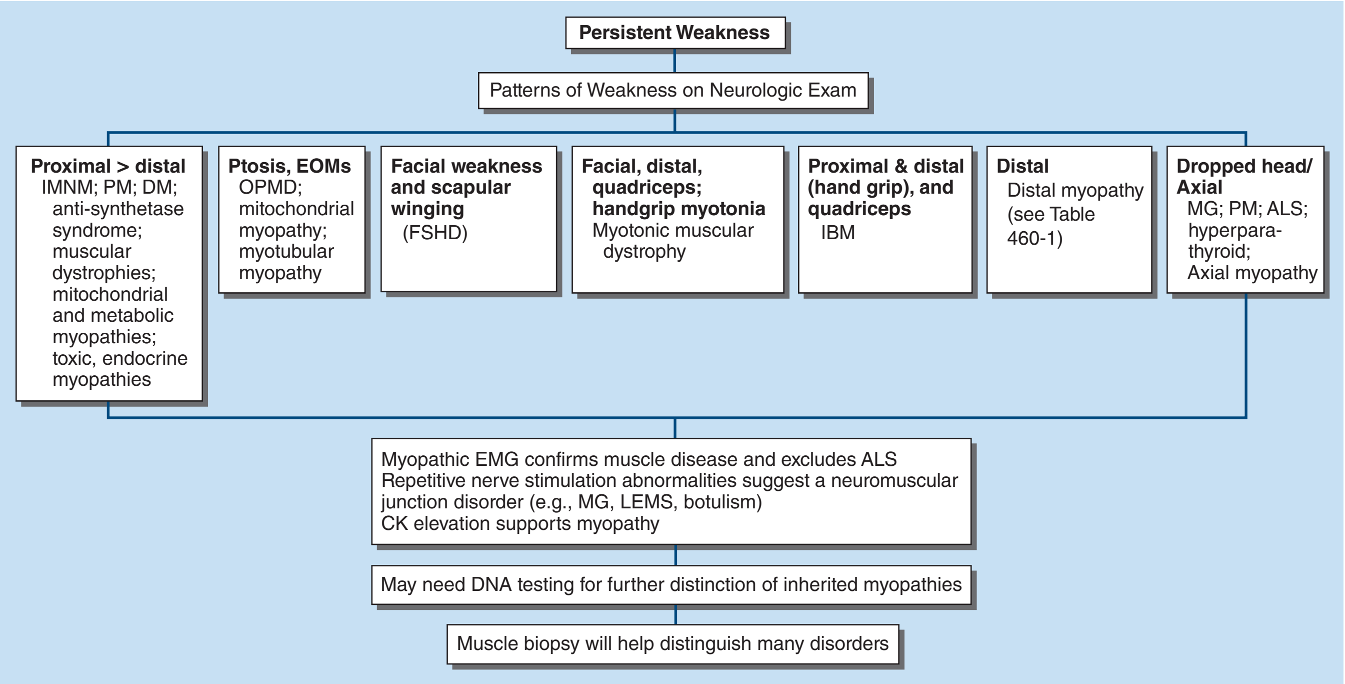
Persistent Weakness
Patterns of Weakness on Neurologic Exam
Proximal > distal IMNM; PM; DM; anti-synthetase syndrome; muscular dystrophies; mitochondrial and metabolic myopathies; toxic, endocrine myopathies
Ptosis, EOMs OPMD; mitochondrial myopathy; myotubular myopathy
Facial weakness and scapular winging (FSHD)
Facial, distal, quadriceps; handgrip myotonia Myotonic muscular dystrophy
Proximal & distal (hand grip), and quadriceps IBM
Distal Distal myopathy (see Table 460-1)
Dropped head/ Axial MG; PM; ALS; hyperpara-thyroid; Axial myopathy
Myopathic EMG confirms muscle disease and excludes ALS Repetitive nerve stimulation abnormalities suggest a neuromuscular junction disorder (e.g., MG, LEMS, botulism) CK elevation supports myopathy
May need DNA testing for further distinction of inherited myopathies
Muscle biopsy will help distinguish many disorders

FIGURE 460-2 Diagnostic evaluation of persistent weakness. Examination reveals one of seven patterns of weakness. The pattern of weakness in combination with the laboratory evaluation leads to a diagnosis. ALS, amyotrophic lateral sclerosis; CK, creatine kinase; DM, dermatomyositis; EMG, electromyography; EOMs, extraocular muscles; FSHD, facioscapulohumeral dystrophy; IBM, inclusion body myositis; IMNM, immune-mediated necrotizing myopathy; MG, myasthenia gravis; OPMD, oculopharyngeal muscular dystrophy; PM, polymyositis.

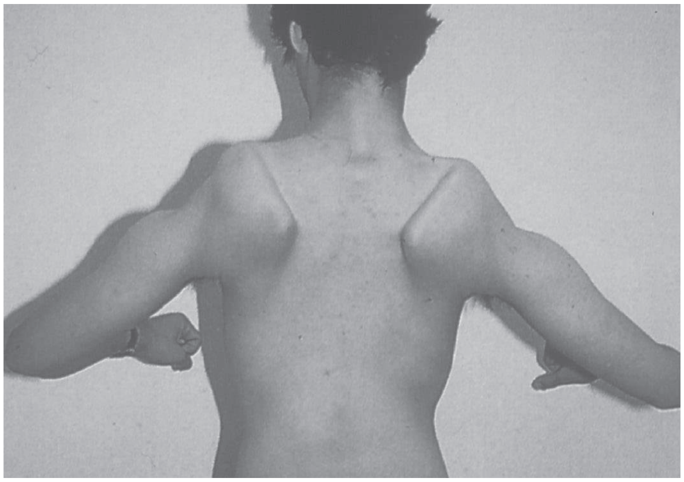

FIGURE 460-3 Facioscapulohumeral dystrophy with prominent scapular winging.

In contrast, muscle atrophy is characteristic of other myopathies. In Miyoshi myopathy, which can be caused by mutations in the genes that encode for dysferlin and anoctamin 5, there is a predilection for early atrophy of the gastrocnemius muscles, particularly the medial aspect. Atrophy of the humeral muscles is characteristic of FSHD and EDMD.

LABORATORY EVALUATION

Various tests can be used to evaluate a suspected myopathy, including CK levels, endocrine studies (e.g., thyroid function tests, parathyroid hormone and vitamin D levels), autoantibodies (associated with myositis and systemic disorders), forearm exercise test, muscle biopsy, and genetic testing. Electrodiagnostic studies can be useful to differentiate myopathies from other neuromuscular disorders (motor neuron disease, peripheral neuropathies, neuromuscular junction disorders) but, in most instances, do not help distinguish the specific type of myopathy.

Serum Enzymes CK is the most sensitive measure of muscle damage. The MM isoenzyme predominates in skeletal muscle, whereas CK-myocardial bound (CK-MB) is the marker for cardiac muscle. Serum CK can be elevated in normal individuals without provocation, presumably on a genetic basis or after strenuous activity, trauma, a prolonged muscle cramp, or a generalized seizure. Aspartate aminotransferase (AST), alanine aminotransferase (ALT), aldolase, and lactate dehydrogenase (LDH) are enzymes sharing an origin in both muscle and liver. Problems arise when the levels of these enzymes are found to be elevated in a routine screening battery, leading to the erroneous assumption that liver disease is present when in fact muscle could be the cause. An elevated γ-glutamyl transferase (GGT) helps to establish a liver origin because this enzyme is not found in muscle. Rarely, aldose can be elevated in an inflammatory myopathy when CK, AST, and ALT are normal, signifying that the inflammation predominantly affects the perimysium (dermatomyositis, graft-versus-host disease) or the surrounding fascia (fasciitis).

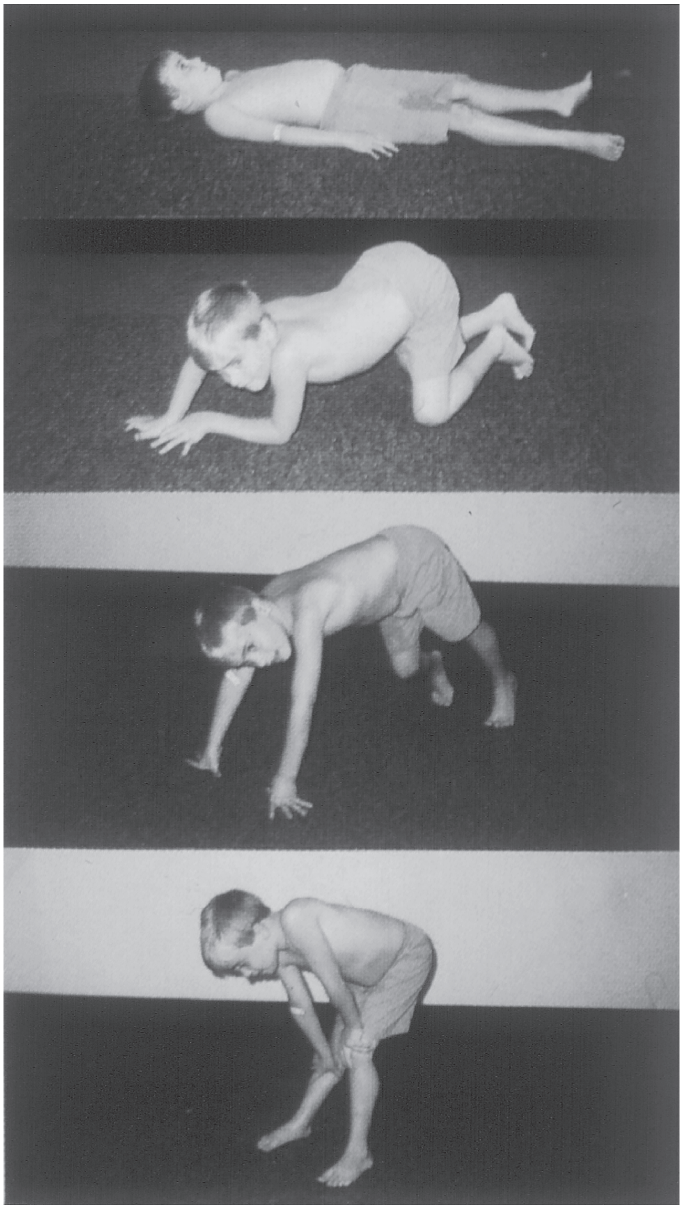

FIGURE 460-4 Gower sign showing a patient using his arms to climb up the legs in attempting to get up from the floor.

TABLE 460-2 Observations on Examination That Disclose Muscle Weakness

FUNCTIONAL IMPAIRMENT	MUSCLE WEAKNESS
Inability to forcibly close eyes	Upper facial muscles
Impaired pucker	Lower facial muscles
Inability to raise head from prone position	Neck extensor muscles
Inability to raise head from supine position	Neck flexor muscles
Inability to raise arms above head	Proximal arm muscles (may be only scapular stabilizing muscles)
Inability to walk without hyperextending knee (back-kneeing or genu recurvatum)	Knee extensor muscles
Inability to walk with heels touching the floor (toe walking)	Shortening of the Achilles tendon
Inability to lift foot while walking (steppage gait or foot drop)	Anterior compartment of leg
Inability to walk without a waddling gait	Hip muscles
Inability to get up from the floor without climbing up the extremities (Gowers' sign)	Hip, thigh, and trunk muscles
Inability to get up from a chair without using arms	Hip muscles

Electrodiagnostic Studies EMG, repetitive nerve stimulation, and nerve conduction studies (NCS) **(Chap. 457)** are helpful in differentiating myopathies from motor neuron disease, neuropathies, and neuromuscular junction diseases. Routine NCS are typically normal in myopathies, but reduced amplitudes of compound muscle action potentials may be seen in atrophied muscles. The needle EMG may reveal irritability on needle insertion and spontaneously that is suggestive of a myopathy with active necrosis or muscle membrane instability (inflammatory myopathies, dystrophies, toxic myopathies, myotonic myopathies), whereas a lack of irritability is characteristic of long-standing myopathic disorders (muscular dystrophies with severe fibrofatty replacement, endocrine myopathies, disuse atrophy, and many of the metabolic myopathies between bouts of rhabdomyolysis). In addition, the EMG may demonstrate myotonic discharges that will narrow the differential diagnosis (Table 460-1). Another important

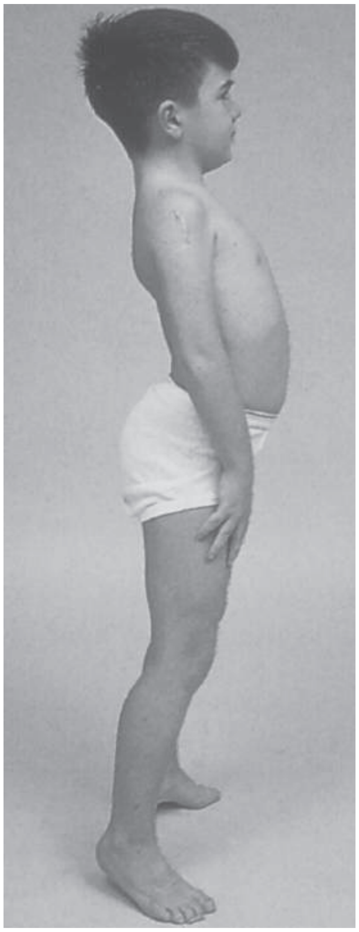

FIGURE 460-5 Hyperlordotic posture, exaggerated by standing on toes, associated with trunk and hip weakness.

EMG finding is the presence of short-duration, small-amplitude, polyphasic motor unit action potentials (MUAPs). In myopathies, the MUAPs fire early but at a normal rate to compensate for the loss of individual muscle fibers, whereas in neurogenic disorders, the MUAPs fire faster. An EMG is usually normal in steroid or disuse myopathy, both of which are associated with type 2 fiber atrophy; this is because the EMG preferentially assesses the physiologic function of type 1 fibers. The EMG can supplement the clinical examination in choosing an appropriately affected muscle to biopsy.

Imaging Studies Skeletal magnetic resonance imaging (MRI) and ultrasound are increasingly utilized to assess the pattern of muscle involvement, which can help in narrowing the diagnosis, and are often more sensitive than the clinical examination and EMG, particularly early in a disease course. For example, there is early predilection of the vastus lateralis and medialis muscles with relative sparing of the rectus femoris muscles on imaging of thigh muscles in patients with inclusion body myositis, and this can be appreciated on imaging prior to weakness being detected on manual muscle testing. MRI can also demonstrate fasciitis when the clinical examination and EMG are normal. Imaging can also be used to help guide what muscle to biopsy in patients with weakness on manual muscle testing and EMG abnormalities only in muscles that are not typically biopsied (e.g., paraspinal or hip girdle). We have found imaging helpful in patients with presumed muscular dystrophy when the muscle biopsy is not diagnostic and genetic testing shows only a variation of unclear significance. In this situation, the pattern of muscle involvement on imaging can support the known pattern of muscle involvement of a specific hereditary myopathy. The cost and availability of MRI preclude routine use in some settings, but ultrasound is more readily available and less expensive.

Genetic Testing This is increasingly available and is the gold standard for diagnosing patients with hereditary myopathies. Next-generation sequencing panels are increasing utilized, but clinicians need to know their limitations; large deletions and duplications can be missed, as can mutations in noncoding (intronic) regions. Furthermore, testing often reveals sequence alterations of unclear significance.

Forearm Exercise Test With exercise-induced muscle pain and myoglobinuria, there may be a defect in glycolysis. For safety, the test should not be performed under ischemic conditions to avoid an unnecessary insult to the muscle, causing rhabdomyolysis. The test is performed by placing a small indwelling catheter into an antecubital vein. A baseline blood sample is obtained for lactic acid and ammonia. The forearm muscles are exercised by asking the patient to vigorously open and close the hand for 1 min. Blood is then obtained at intervals of 1, 2, 4, 6, and 10 min for comparison with the baseline sample. A three- to fourfold rise of lactic acid is typical. The simultaneous measurement of ammonia serves as a control because it should also rise with exercise. In patients with myophosphorylase deficiency and certain other glycolytic defects, the lactic acid rise will be absent or below normal, while the rise in ammonia will reach control values. If there is lack of effort, neither lactic acid nor ammonia will rise. Patients with selective failure to increase ammonia may have myoadenylate deaminase deficiency. This condition has been reported to be a cause of myoglobinuria, but deficiency of this enzyme in asymptomatic individuals makes interpretation controversial.

Muscle Biopsy Muscle biopsy is extremely helpful in evaluation of acquired myopathies but is performed less frequently in suspected hereditary myopathies as genetic testing has become more widely available. However, muscle biopsy can be helpful in cases of suspected hereditary myopathy in which genetic testing was nondiagnostic. Almost any superficial muscle can be biopsied, but it is important to biopsy one that is affected clinically but not too severely (for example, grade 4 out of 5 strength or movement against moderate resistance by manual muscle testing) (Chap. 433). A specific diagnosis can be established in many disorders.

HEREDITARY MYOPATHIES

Muscular dystrophy refers to a group of hereditary progressive diseases, each with unique phenotypic and genetic features (Tables 460-3 through 460-6 and Fig. 460-6). The prognosis of dystrophies is slow progressive weakness, though the severity and course are variable between and even within subtypes. Some are associated with cardiac and ventilatory muscle involvement, which are the leading causes of mortality. Unfortunately, there are no specific medical therapies for most of the muscular dystrophies, and treatment is aimed at maintaining function with physical and occupational therapy. Noninvasive ventilation and tracheostomy may be warranted. Those with cardiomyopathy may require afterload reduction, antiarrhythmic agents, pacemakers or intracardiac defibrillators, and occasionally cardiac transplantation. We will focus primarily on those that manifest in adulthood.

DUCHENNE AND BECKER MUSCULAR DYSTROPHY (DMD AND BMD)

DMD and BMD are X-linked recessive muscular dystrophies caused by mutations in the *dystrophin* gene. Affecting 1 in 3000 male births, DMD is the most common mutational disease affecting boys. The incidence of BMD is ~5 per 100,000.

Clinical Features Proximal muscles, especially of the lower extremities, are prominently involved in both disorders. This becomes evident in DMD very early; boys with DMD have difficulty climbing stairs and never run well. As the disease progresses, weakness becomes more generalized. Hypertrophy of muscles, particularly in the calves, is an early and prominent finding. Most patients with BMD first experience difficulties between ages 5 and 15 years, although onset in the third or fourth decade or even later can occur. Life expectancy for DMD and BMD is reduced, but most BMD cases survive into the fourth or fifth decade. Intellectual disability may occur in both disorders but is less common in BMD. Cardiac involvement is common in both DMD and BMD and may result in heart failure; some BMD

TABLE 460-3 Autosomal Dominant (AD) Limb-Girdle Muscular Dystrophies (LGMDs)

OLD / NEW NOMENCLATURE	INHERITANCE	GENE	AFFECTED PROTEIN
LGMD1A / MFM3	AD	*MYOT*	Myotilin
LGMD1B / EDMD	AD	*LMNA*	Lamin A and C
LGMD1C / Rippling muscle disease	AD	*CAV3*	Caveolin-3
LGMD1D / LMGDD1	AD	*DNAJB6*	DNAJ heat shock protein family (Hsp40) member B6
LGMD1E / MFM1	AD	*DES*	Desmin
LGMD1F / LGMDD2	AD	*TNPO3*	Transportin 3
LGMD1G / LGMDD3	AD	*HNRNPDL*	Heterogeneous nuclear ribonucleoprotein D like protein
LGMD1H / Discarded due to false linkage			
LGMD1I / LGMDD4	AD	*CAPN3*	Calpain 3
Bethlem myopathy / LGMDD5	AD	*COL6A1/2/3*	Collagen type VI alpha

patients manifest with only heart failure. Other less common presentations of dystrophinopathy are asymptomatic hyper-CK-emia, myalgias without weakness, and myoglobinuria.

Laboratory Features Serum CK levels are usually elevated. Muscle biopsies demonstrate dystrophic features. Western blot analysis of muscle biopsy samples demonstrates absent dystrophin in DMD or reduction in levels or size of dystrophin in BMD. In both disorders, mutations can be established using DNA from peripheral blood leukocytes. In most cases, muscle biopsies are no longer performed when DMD or BMD is suspected, as genetic testing is less invasive, less costly, and routinely available. Deletions within or duplications of the *dystrophin* gene are common in both DMD and BMD; in ~95% of cases, the mutation does not alter the translational reading frame of messenger RNA. These "in-frame" mutations allow for production of some dystrophin, which accounts for the presence of altered rather than absent dystrophin on Western blot analysis and a milder clinical phenotype.

TABLE 460-4 Autosomal Recessive (AR) Limb-Girdle Muscular Dystrophies (LGMDs)

OLD / NEW NOMENCLATURE	INHERITANCE	GENE	AFFECTED PROTEIN
LGMD2A / LGMDR1	AR	*CAPN3*	Calpain 3
LGMD2B / LGMDR2	AR	*DYSF*	Dysferlin
LGMD2C / LGMDR5	AR	*SGCG*	γ-Sarcoglycan
LGMD2D / LGMDR3	AR	*SGCA*	α-Sarcoglycan
LGMD2E / LGMDR4	AR	*SCGB*	β-Sarcoglycan
LGMD2F / LGMDR6	AR	*SCGD*	δ-Sarcoglycan
LGMD2G / LGMDR7	AR	*TCAP*	Telethonin
LGMD2H / LGMDR8	AR	*TRIM32*	Tripartite motif-containing 32
LGMD2I / LGMDR9	AR	*FKRP*	Fukutin-related protein
LGMD2J / LGMDR10	AR	*TTN*	Titin
LGMD2K / LGMDR11	AR	*POMT1*	Protein O-mannosyltransferase 1
LGMD2L / LGMDR12	AR	*ANO5*	Anoctamin 5
LGMD2M / LGMDR13	AR	*FKTN*	Fukutin
LGMD2N / LGMDR14	AR	*POMT2*	Protein O-mannosyltransferase 2
LGMD2O / LGMDR15	AR	*POMGnT1*	Protein O-linked mannose Beta-1,2-N-acetyl glucosaminyltranferase-1
LGMD2P / LGMDR16	AR	*DAG1*	α-Dystroglycan
LGMD2Q / LGMDR17	AR	*PLEC1*	Plectin 1
LGMD2R / MFM1	AR	*DES*	Desmin
LGMD2S / LGMDR18	AR	*TRAPPC11*	Trafficking protein particle complex 11
LMGD2T / LGMDDR19	AR	*GMPPB*	DP-mannose pyrophosphorylase B
LGMD2U / LGMDR20	AR	*CRPPA*	CDP-L-ribitol pyrophosphorylase A (also known as ISPD)
LGMD2V / Pompe disease	AR	*GAA*	A Alpha-glucosidase
LGMD2W / PINCH-2-related myopathy	AR	*LIMS2*	PINCH-2
LGMD2X / LGMDR25	AR	*BVES*	Blood vessel endothelial substance
LGMD2Y / TOR1AIP1-related myopathy	AR	*TOR1AIP1*	Torsin A interacting protein 1
LGMD2Z / LGMDR21	AR	*POGLUT1*	Protein O-glucosyltransferase 1
Bethlem myopathy /LGMDR22	AR	*COL6A1/2/3*	Collagen VI subunits A1, A2, or A3
Laminin α2-related dystrophy / LGMDR23	AR	*LAMA2*	Laminin subunit alpha 2
POMGNT2-related dystrophy/ LGMDR24	AR	*POMGNT2*	Protein O-linked mannose beta 1,4-N-acetyl-glucosaminyltransferase 2
NA / LGMDR26	AR	*POPDC3*	Popeye domain-containing protein 3
NA / LGMDR27	AR	*JAG2*	Jagged2

Abbreviation: NA, not applicable.

TABLE 460-5 Hereditary Distal Myopathies/Dystrophies

DISORDER	INHERITANCE	GENE	AFFECTED PROTEIN
Welander	AD	*TIA1*	T-cell restricted intracellular antigen
Udd	AD	*TTN*	Titin
Markesbery-Griggs	AD	*LDB3*	ZASP
GNE myopathy (Nonaka; hIBM2)	AR	*GNE*	UDP-*N*-acetylglucosamine 2-epimerase/ *n*-acetylmannosamine kinase
Miyoshi 1	AR	*DYSF*	Dysferlin
Miyoshi 3	AR	*ANO5*	Anoctamin 5
Laing	AD	*MYH7*	Myosin heavy chain 7
Williams	AD	*FLNC*	Filamin C
Distal myopathy with vocal cord and pharyngeal weakness (VCPDM)	AD	*MTR3*	Matrin 3
KLHL9 myopathy	AD	*KLH9*	KELCH-like homologue 9
ADSSL myopathy	AR	*ADSSL*	Adenylosuccinate synthase
PLIN4 myopathy	AD	*PLIN4*	Perilipin-4

Abbreviations: AD, autosomal dominant; AR, autosomal recessive.

TREATMENT

Duchenne and Becker Muscular Dystrophy

Glucocorticoids slow progression in DMD, but their use has not been adequately studied in BMD. Physical and occupational therapy are important in helping maintain function. As death is often from the associated cardiomyopathy, it is important to follow patients with a cardiologist and treat appropriately. Small studies suggest that there may be a clinical benefit in selected cases of DMD from short oligonucleotides that permit skipping of mutant exons, leading to expression of a short but nonetheless functional dystrophin protein. In parallel, other studies suggest that small molecules may permit read-through of protein-truncating mutations in some DMD cases. Gene therapy studies have not as yet been conducted in BMD.

LIMB-GIRDLE MUSCULAR DYSTROPHY

The limb-girdle muscular dystrophies (LGMDs) are a genetically heterogenous group of dystrophies in which males and females are affected equally, with typical onset ranging from late in the first decade to the fourth decade. The LGMDs usually manifest with progressive weakness of pelvic and shoulder girdle musculature and are often clinically indistinguishable from DMD and BMD. Respiratory insufficiency from weakness of the diaphragm may occur, as may cardiomyopathy. Serum CKs are elevated, and the EMG is myopathic. Muscle biopsies reveal dystrophic features, but the findings are not specific to differentiate subtypes from one another unless immunohistochemistry is employed (e.g., immunostaining for various sarcoglycans, dysferlin, alpha-dystroglycan) or there are features to suggest one of the myofibrillar myopathies. Nonetheless, definitive diagnosis requires genetic testing.

The traditional classification of LGMD is based on autosomal dominant (LGMD1) and autosomal recessive (LGMD2) inheritance. Superimposed on the backbone of LGMD1 and LGMD2, the classification uses a sequential alphabetical lettering system (LGMD1A, LGMD2A, etc.) based on genotype. However, ever-expanding discoveries of new genes have outgrown the alphabet. The European Neuromuscular Centre (ENMC) proposed a new nomenclature in which autosomal dominant cases are termed LGMD "D" and autosomal recessive as LGMD "R," followed by a numerical number based on genotype. Furthermore, this new classification only includes cases in which at least two unrelated families have been reported, the predominant weakness at onset was proximal, independent ambulation was achieved at some time, CK is elevated, and muscle biopsies or imaging revealed dystrophic features. Thus, mutations in the *CPN3* gene leading to a deficiency in calpain-3, which traditionally were classified as LGMD2A, are classified as LGMDR1 by this new system. In contrast, mutations in myotilin (LGMD1A) and desmin (LGMD1E and LGMD2R) and that often have more distal weakness and have biopsy features of a myofibrillar myopathy are not classified as a LGMD in this new scheme but rather as subtypes of myofibrillar myopathy. Likewise, laminopathies (LGMD1B) are considered a subtype of EDMD rather than an LGMD. The myofibrillar myopathies are now considered as being separate from LGMD. This new classification of LGMD and distal muscular dystrophies is summarized in Tables 460-3 and 460-4.

A recent metanalysis reported the prevalence of LGMD to be 1.63 per 100,000 (range, 0.56–5.75 per 100,000), while estimated prevalences

TABLE 460-6 Myofibrillary Myopathies (MFM)

MYOFIBRILLAR MYOPATHY	INHERITANCE	GENE	AFFECTED PROTEIN
MFM1	AD/AR	*DES*	Desmin
MFM2	AD	*CRYAB*	Alpha-B crystallin
MFM3	AD	*MYOT*	Myotolin
MFM4	AD	*LDP3*	ZASP
MFM5	AD	*FLNC*	Filamin C
MFM6	AD	*BAG3*	Bcl-2-binding protein
MFM7	AD	*KY*	Kyphoscoliosis peptidase
MFM8	AD	*PYROXD1*	Pyridine nucleotide-disulfide oxidoreductase domain-containing protein 1
MFM9	AD	*TTN*	Titin
MFM10	AD	*SVIL*	Supervillin
MFM11	AD	*UNC45B*	UNC45 myosin chaperone B
MFM12	AD	*MYL2*	Myosin light chain 2

Abbreviations: AD, autosomal dominant; AR, autosomal recessive.

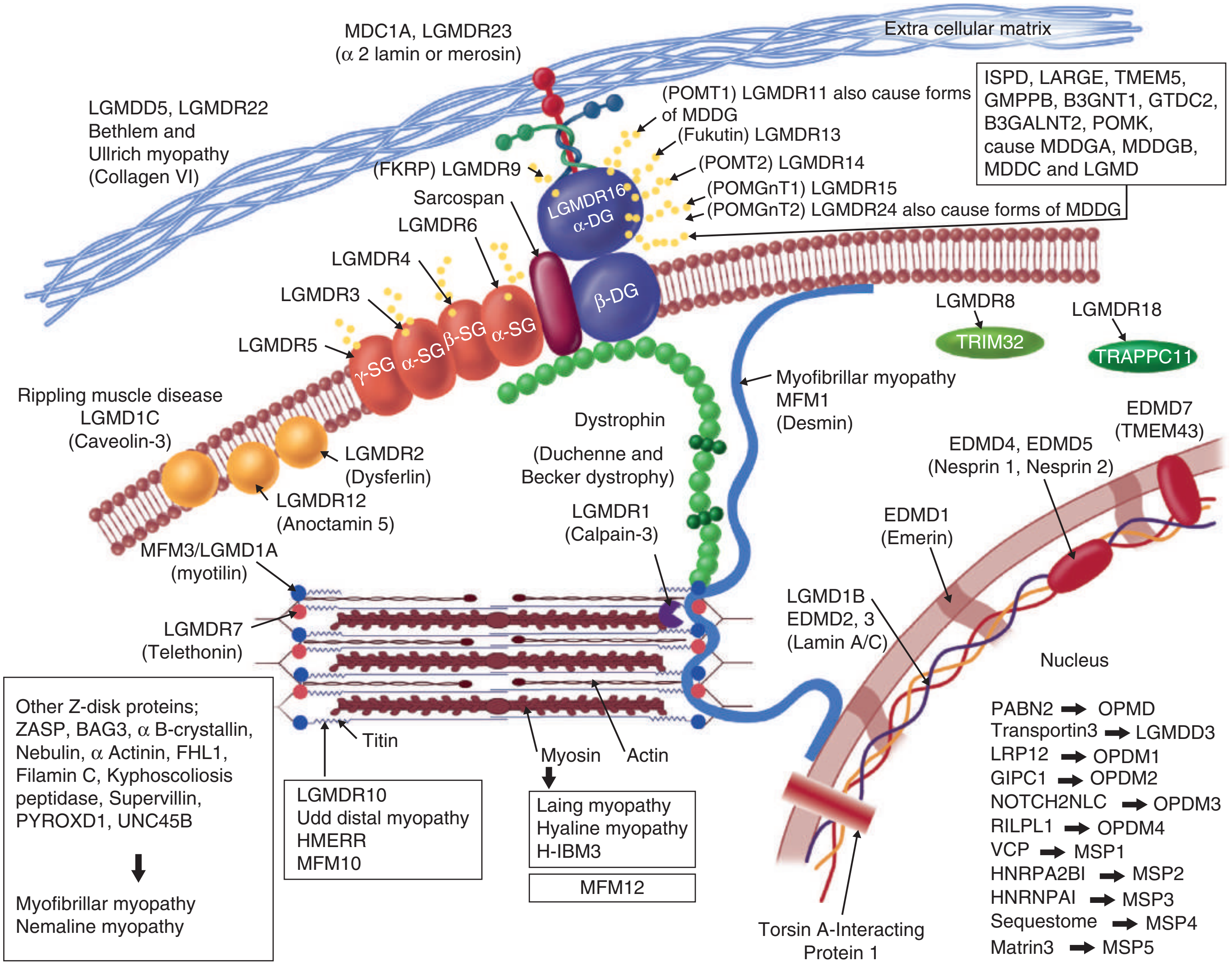
MDC1A, LGMDR23
(α 2 lamin or merosin)
Extra cellular matrix
LGMDD5, LGMDR22
Bethlem and
Ullrich myopathy
(Collagen VI)
(POMT1) LGMDR11 also cause forms of MDDG
(Fukutin) LGMDR13
(POMT2) LGMDR14
(POMGnT1) LGMDR15
(POMGnT2) LGMDR24 also cause forms of MDDG
ISPD, LARGE, TMEM5, GMPPB, B3GNT1, GTDC2, B3GALNT2, POMK, cause MDDGA, MDDGB, MDDC and LGMD
(FKRP) LGMDR9
Sarcospan
LGMDR16
α-DG
LGMDR6
LGMDR4
LGMDR3
β-DG
LGMDR5
γ-SG
α-SG
β-SG
α-SG
LGMDR8
TRIM32
LGMDR18
TRAPPC11
Myofibrillar myopathy
MFM1
(Desmin)
Rippling muscle disease
LGMD1C
(Caveolin-3)
LGMDR2
(Dysferlin)
LGMDR12
(Anoctamin 5)
Dystrophin
(Duchenne and
Becker dystrophy)
LGMDR1
(Calpain-3)
EDMD7
(TMEM43)
EDMD4, EDMD5
(Nesprin 1, Nesprin 2)
EDMD1
(Emerin)
MFM3/LGMD1A
(myotilin)
LGMDR7
(Telethonin)
LGMD1B
EDMD2, 3
(Lamin A/C)
Nucleus
Other Z-disk proteins; ZASP, BAG3, α B-crystallin, Nebulin, α Actinin, FHL1, Filamin C, Kyphoscoliosis peptidase, Supervillin, PYROXD1, UNC45B
Myofibrillar myopathy
Nemaline myopathy
Titin
LGMDR10
Udd distal myopathy
HMERR
MFM10
Myosin
Actin
Laing myopathy
Hyaline myopathy
H-IBM3
MFM12
Torsin A-Interacting
Protein 1
PABN2 → OPMD
Transportin3 → LGMDD3
LRP12 → OPDM1
GIPC1 → OPDM2
NOTCH2NLC → OPDM3
RILPL1 → OPDM4
VCP → MSP1
HNRPA2BI → MSP2
HNRNPAI → MSP3
Sequestome → MSP4
Matrin3 → MSP5

FIGURE 460-6 Proteins involved in the muscular dystrophies. This schematic shows the location of various sarcolemmal, sarcomeric, nuclear, and enzymatic proteins associated with muscular dystrophies. The diseases associated with mutations in the genes responsible for encoding these proteins are shown in boxes. Dystrophin, via its interaction with the dystroglycan complex, connects the actin cytoskeleton to the extracellular matrix. Extracellularly, the sarcoglycan complex interacts with biglycan, which connects this complex to the dystroglycan complex and the extracellular matrix collagen. Various enzymes are important in the glycosylation of the α-dystroglycan and mediate its binding to the extracellular matrix and usually cause a congenital muscular dystrophy with severe brain and eye abnormalities but may cause milder limb-girdle muscular dystrophy (LGMD) phenotype. Mutations in genes that encode for sarcomeric and Z-disk proteins cause forms of LGMD and distal myopathies (including myofibrillar myopathy, forms of hereditary inclusion body myopathy) as well as nemaline rod myopathy and other "congenital" myopathies. Mutations affecting nuclear membrane proteins are responsible for most forms of Emery-Dreifuss muscular dystrophy (EDMD). Mutations in other nuclear genes cause other forms of dystrophy. *(Reproduced with permission from AA Amato et al (eds): Amato and Russell's Neuromuscular disorders, 3rd ed. New York: McGraw Hill; 2025.)*

of individual specific subtypes of LGMDs vary. The most common types of adult-onset LGMD are calpainopathy (LGMD2A/LGMDR1), fukutin-related protein (FKRP) deficiency (LGMD2I/LGMDR9), and anoctaminopathy (LGMD2L/LGMDR12). Calpainopathy (LGMD2A/LGMDR1), the most common cause of LGMD in those with ancestry from Spain, France, Italy, and Great Britain, is associated with marked scapular winging, lack of calf muscle hypertrophy, and lack of cardiac and lung involvement. Of note, autosomal dominant mutations in an intron of the calpain-3 gene is responsible for LGMD1I/LGMDD4. LGMD2I/LGMDR9 is more common in individuals with northern European ancestry, is associated with calf muscle hypertrophy, and can have cardiac and lung involvement out of proportion to extremity weakness. LGMD2L/LGMDR12 accounts for ~7% of LGMD in the United States, and the prevalence is higher in northern Europe; as seen in dysferlinopathies (LGMD2B/LGMDR2 and Miyoshi myopathy type 1), anoctaminopathy has an early predilection for medial calf atrophy and weakness.

Importantly, immune-mediated necrotizing myopathies can mimic LGMD clinically and histopathologically (Chap. 377). Anyone suspected of having an LGMD but without definite pathogenic mutation(s) identified on genetic testing should be screened for the presence of serum antibodies against HMGCR and SRP to assess for a treatable autoimmune cause.

EMERY-DREIFUSS MUSCULAR DYSTROPHY

There are at least seven subtypes of EDMD that have been associated with mutations in *EMD* (EDMD1), *LMNA* (EDMD2 and EDMD3), *SYNE1* (EDMD4), *SYNE2* (EDMD5), *FHL1* (EDMD6), and *TMEM43* (EDMD7), encoding emerin, lamin A/C, nesprin-1, nesprin-2, FHL1, and LUMA, respectively. Mutations in *EMD* and *FHL* produce X-linked inheritance, whereas the others can be autosomal dominant (*LMNA, SYNE1, SYNE2, LUMA*) or autosomal recessive (*LMNA1*). The clinical phenotypes are quite similar.

Clinical Features Prominent contractures can be recognized in early childhood and teenage years, often preceding muscle weakness. The contractures persist throughout the course of the disease and are present at the elbows, ankles, and neck. Muscle weakness affects humeral and peroneal muscles at first and later spreads to a limb-girdle distribution (Table 460-1). The cardiomyopathy is potentially life threatening and may result in sudden death. A spectrum of atrial rhythm and conduction defects includes atrial fibrillation, atrial standstill, and atrioventricular heart block. Some patients have a dilated cardiomyopathy. Female carriers of the X-linked variant may manifest with a cardiomyopathy.

Laboratory Features Serum CK is usually slightly elevated, and the EMG is myopathic. Muscle biopsy usually shows nonspecific dystrophic features, although cases associated with *FHL1* mutations have features of myofibrillar myopathy. Immunohistochemistry reveals absent emerin staining of myonuclei in X-linked EDMD due to *emerin* mutations. Electrocardiograms (ECGs) demonstrate atrial and atrioventricular rhythm disturbances.

X-linked EDMD usually arises from defects in the *emerin* gene encoding a nuclear envelope protein. *FHL1* mutations are also a cause of X-linked scapuloperoneal dystrophy but can also present with an X-linked form of EDMD. The autosomal dominant disease can be caused by mutations in the *LMNA* gene encoding Lamin A/C; in the synaptic nuclear envelope protein 1 (*SYNE1*) or 2 (*SYNE2*) encoding nesprin-1 and nesprin-2, respectively; and in *TMEM43* encoding LUMA. These proteins are essential components of the filamentous network underlying the inner nuclear membrane. Loss of structural integrity of the nuclear envelope from defects in emerin, Lamin A/C, nesprin-1, nesprin-2, and LUMA accounts for overlapping phenotypes.

TREATMENT

Emery-Dreifuss Muscular Dystrophy

Supportive care should be offered for neuromuscular disability, including ambulatory aids, if necessary. Stretching of contractures is difficult. Management of cardiomyopathy and arrhythmias (e.g., early use of a defibrillator or cardiac pacemaker) may be lifesaving.

MYOTONIC DYSTROPHY

There are two distinct forms of myotonic dystrophy (*dystrophia myotonica* [DM]), namely myotonic dystrophy type 1 (DM1) and myotonic dystrophy type 2 (DM2), also called *proximal myotonic myopathy* (PROMM).

Clinical Features The clinical expression of DM1 varies widely and involves many systems other than muscle. Affected patients may have a “hatchet-faced” appearance due to temporalis, masseter, and facial muscle atrophy and weakness. Frontal baldness is frequent. Weakness of wrist and fingers occurs early, as does foot drop. Proximal muscles are less affected. Palatal, pharyngeal, and tongue involvement can lead to dysarthria and dysphagia. Some patients have diaphragm and intercostal muscle weakness, resulting in ventilatory insufficiency. Myotonia is usually apparent by the age of 5 years and is best demonstrable by percussion of the thenar eminence or asking patients to close their fingers very tightly and then relax.

ECG abnormalities include first-degree heart block and more extensive conduction system involvement. Complete heart block and sudden death can occur. Congestive heart failure occurs infrequently but may result from cor pulmonale secondary to respiratory failure. Other associated features include intellectual impairment, hypersomnia, posterior subcapsular cataracts, gonadal atrophy, insulin resistance, and decreased esophageal and colonic motility.

Congenital myotonic dystrophy is a more severe form of DM1 and occurs in ~25% of infants of affected mothers. It is characterized by severe facial and bulbar weakness, transient neonatal respiratory insufficiency, and intellectual disability.

DM2 or PROMM involves mainly proximal muscles. Other features of the disease overlap with DM1, including cataracts, testicular atrophy, insulin resistance, constipation, hypersomnia, and cognitive defects. Cardiac conduction defects occur but are less common. The hatchet face and frontal baldness are also less consistent features. A very striking difference is the failure to clearly identify a congenital form of DM2.

Laboratory Features The diagnosis of myotonic dystrophy can usually be made on the basis of clinical findings. Serum CK levels may be normal or mildly elevated. EMG evidence of myotonia is present in most cases of DM1 but is more patchy in DM2. Muscle biopsy is not typically performed for diagnosis but is sometimes done when the clinical features and electrophysiologic features are not recognized. The major histopathologic features in both DM1 and DM2 are numerous internalized nuclei in individual muscle fibers combined with many atrophic fibers with pyknotic nuclear clumps.

DM1 and DM2 are autosomal dominant disorders. DM1 is transmitted by an intronic mutation consisting of an unstable expansion of a CTG trinucleotide repeat in a serine-threonine protein kinase gene (named *DMPK*). An increase in the severity of the disease phenotype in successive generations (genetic anticipation) is accompanied by an increase in the number of trinucleotide repeats. The unstable triplet repeat in myotonic dystrophy can be used for prenatal diagnosis. Congenital disease occurs almost exclusively in infants born to affected mothers.

DM2 is caused by a DNA expansion mutation consisting of a CCTG repeat in intron 1 of the *CNBP* gene encoding the CCHC-type zinc finger nucleic acid binding protein. The DNA expansions in DM1 and DM2 impair muscle function by a toxic gain of function of the mutant mRNA. In both DM1 and DM2, the mutant RNA appears to form intranuclear inclusions composed of aberrant RNA. These RNA inclusions sequester RNA-binding proteins essential for proper splicing of a variety of other mRNAs. This leads to abnormal transcription of multiple proteins in a variety of tissues/organ systems, in turn causing the systemic manifestations of DM1 and DM2.

TREATMENT

Myotonic Dystrophy

The myotonia in DM1 and DM2 is usually not so bothersome to warrant treatment, but when it is, mexiletine may be helpful. A cardiac pacemaker or implantable cardioverter defibrillator should be considered for patients with significant arrhythmia. Molded ankle-foot orthoses help stabilize gait in patients with foot drop. Excessive daytime somnolence with or without sleep apnea is not uncommon. Sleep studies, noninvasive respiratory support (biphasic positive airway pressure [BiPAP]), and treatment with modafinil may be beneficial.

FACIOSCAPULOHUMERAL (FSHD) MUSCULAR DYSTROPHY

There are two forms of FSHD that have similar pathogenesis. Most patients have FSHD type 1 (95%), whereas ~5% have FSHD2. Both forms are clinically and histopathologically identical. The prevalence FSHD is ~5 per 100,000 individuals.

Clinical Features FSHD typically presents in childhood or young adulthood. In most cases, facial weakness is the initial manifestation, appearing as an inability to smile, whistle, or fully close the eyes. Loss of scapular stabilizer muscles makes arm elevation difficult. Scapular winging (Fig. 460-3) becomes apparent with attempts at abduction and forward movement of the arms. Biceps and triceps muscles may be severely affected, with relative sparing of the deltoid muscles. Weakness is invariably worse for wrist extension than for wrist flexion, and weakness of the anterior compartment muscles of the legs may lead to foot drop. In 20% of patients, weakness progresses to involve the pelvic muscles, and severe functional impairment and possible wheelchair dependency result. The heart is not involved, but there can be ventilatory muscle weakness in 5% of affected individuals. There is an increased incidence of nerve deafness. *Coats' disease*, a disorder consisting of telangiectasia, exudation, and retinal detachment, also occurs.

Laboratory Features The serum CK level may be normal or mildly elevated. EMG and muscle biopsy show nonspecific abnormalities but on occasion can reveal a prominent inflammatory infiltrate leading to an incorrect diagnosis of myositis **(Chap. 377)**.

FSHD1 is associated with deletions of tandem 3.3-kb repeats at 4q35. The deletion reduces the number of repeats to a fragment of <35 kb in most patients. Within these repeats lies the *DUX4* gene, which usually is not expressed after early muscle development. In patients with FSHD1, these deletions in the setting of a specific polymorphism lead to hypomethylation of the region and toxic expression of the *DUX4* gene. In cases of FSHD2, there is no deletion, but rather mutations in three different genes have been identified, each of which interestingly leads to hypomethylation of the DUX4 region and the permissive expression of the *DUX4* gene. Dominant mutations in the structural maintenance of chromosomes hinge domain 1 (*SMCHD1*) gene are the most common cause of FSHD2, but heterozygous mutations in the DNA methyltransferase 3B (*DNMT3B*) gene and homozygous mutations in the ligand-dependent nuclear receptor-interacting factor 1 (*LRIF1*) gene also cause autosomal recessive FSHD2. These proteins normally interact with SMCHD1, and mutations lead to hypomethylation of DUX4. As in FSHD1, this leads to an overexpression of the *DUX4* transcript that encodes for double homeobox 4, which itself is a transcription factor controlling the expression of other genes. In turn, this likely results in the altered expression of additional genes.

TREATMENT

Facioscapulohumeral Muscular Dystrophy

No specific treatment is available, though clinical trials assessing the safety and efficacy of reducing DUX4 expression are ongoing. Physical and occupational therapy are the current mainstays of treatment. Ankle-foot orthoses are helpful for foot drop. Scapular stabilization procedures improve scapular winging and function.

■ OCULOPHARYNGEAL DYSTROPHY (OPMD)

OPMD represents one of several disorders characterized by progressive external ophthalmoplegia, which consists of slowly progressive ptosis and limitation of eye movements with sparing of pupillary reactions for light and accommodation. Patients usually do not complain of diplopia, in contrast to patients having conditions with a more acute onset of ocular muscle weakness (e.g., myasthenia gravis).

Clinical Features OPMD has a late onset; it usually presents in the fourth to sixth decade with ptosis or dysphagia. The extraocular muscle impairment is less prominent in the early phase but may become severe over time. The swallowing problem may lead to aspiration. Weakness of the neck and proximal extremities can develop but is usually mild in degree.

Laboratory Features The serum CK level may be two to three times normal. EMG can identify myopathic changes in weak muscles. Muscle biopsies are no longer necessary for diagnosis in most cases but, when performed, demonstrate muscle fibers with rimmed vacuoles. On electron microscopy, a distinctive feature of OPMD is the presence of 8.5-nm tubular filaments in some muscle cell nuclei.

OPMD is an autosomal dominant disorder that has a high incidence in certain populations (e.g., French-Canadians, individuals of Spanish ancestry, and Ashkenazi Jews). The molecular defect in OPMD is an expansion of a polyalanine repeat tract in a poly-RNA-binding protein (*PABP2*) gene. *PABP2* is involved in polyadenylation of mRNAs and their transport through the nuclei pores into the cytoplasm. The expansion of the GCG repeats results in abnormal folding of the polyalanine domains of *PABP2* and its resistance to nuclear proteasomal degradation. This in turn may result in (1) direct toxicity of the intranuclear aggregates; (2) intranuclear sequestration of essential transcription factors, molecular chaperones, RNA binding proteins, and RNAs by these intranuclear aggregates; or (3) suppression of the normal function of the wild-type protein.

TREATMENT

Oculopharyngeal Dystrophy

Dysphagia can lead to significant undernourishment and aspiration. Cricopharyngeal myotomy may improve swallowing. Eyelid crutches can improve vision when obstructed by ptosis; candidates for ptosis surgery must be carefully selected—those with severe facial weakness are not suitable.

■ OCULOPHARYNGEAL DISTAL MYOPATHY (OPDM)

Clinical Features OPDM is characterized by adult-onset ptosis, external ophthalmoplegia, facial muscle weakness, distal limb muscle weakness and atrophy, and pharyngeal involvement, resulting in dysphagia and dysarthria. Some patients manifest with only ptosis without pharyngeal or distal weakness.

Laboratory Features Serum CK levels are normal or only mildly elevated. EMG is myopathic. Muscle biopsies reveal dystrophic features including muscle fibers with rimmed vacuoles. Intramyonuclear inclusions immunostaining with anti-phospho-p62/SQSTM1 antibodies are evident. Similar intranuclear inclusions are found on skin biopsies.

OPDM is a genetically heterogeneous autosomal disorder caused by trinucleotide repeat expansions (CTG) in the 5′ untranslated region (UTR) regions of *LRP12* (OPDM1), *G1PC1* (OPDM2), *NOTCH2NLC* (OPDM3), and *RILPL1* (OPMD4). Notably, the CGG repeat expansion in *NOTCH2NLC* is also the cause of neuronal intranuclear hyaline inclusion disease and other neurodegenerative diseases affecting the brain. These repeat expansion disorders lead to RNA-mediated sequestration of RNA-binding proteins and altered translation of proteins.

TREATMENT

Oculopharyngeal Distal Myopathy

Treatment of dysphagia and ptosis is similar to that noted with OPMD.

■ DISTAL MYOPATHIES/DYSTROPHIES

The distal myopathies are notable for their preferential distal distribution of muscle weakness in contrast to most muscle conditions associated with proximal weakness. The major distal myopathies are summarized in Tables 460-1, 460-5, and 460-6.

Clinical Features *Welander, Udd*, and *Markesbery-Griggs type distal myopathies* are all late-onset, dominantly inherited disorders of distal limb muscles, usually beginning after age 40 years. Welander distal myopathy preferentially involves the wrist and finger extensors, whereas the others are associated with anterior tibial weakness leading to progressive foot drop. *Laing distal myopathy* is also a dominantly inherited disorder heralded by tibial weakness; however, it is distinguished by onset in childhood or early adult life. GNE myopathy (previously known as *Nonaka distal myopathy* and autosomal recessive hereditary inclusion body myopathy) and *Miyoshi myopathy* are distinguished by autosomal recessive inheritance and onset in the late teens or twenties. GNE and Williams myopathy produce prominent anterior tibial weakness, whereas Miyoshi myopathy is unique in that gastrocnemius muscles are preferentially affected at onset. Finally, the *myofibrillar myopathies* (MFMs) are a clinically and genetically heterogeneous group of muscular dystrophies that can be associated with prominent distal or proximal weakness; they can be inherited in an autosomal dominant or recessive pattern (Table 460-6).

Laboratory Features Serum CK levels are markedly elevated in Miyoshi myopathy, but in the other conditions, serum CK is only slightly increased. EMGs are myopathic and can be irritable with myotonic discharges in MFM. Muscle biopsy shows nonspecific dystrophic features and, with the exception of Laing and Miyoshi myopathies, often shows rimmed vacuoles. MFM is associated with the accumulation of dense inclusions and amorphous material best seen on Gomori trichrome staining along with myofibrillar disruption on electron microscopy. Immune staining sometimes demonstrates accumulation of desmin and other proteins in MFM, large deposits of myosin heavy chain in the subsarcolemmal region of type 1 muscle fibers in Laing myopathy, and reduced or absent dysferlin in Miyoshi myopathy type 1.

TREATMENT

Distal Myopathies

Occupational therapy is offered for loss of hand function; ankle-foot orthoses can support distal lower limb muscles. The MFMs can be associated with cardiomyopathy (congestive heart failure or arrhythmias) and respiratory failure that may require medical management. Laing-type distal myopathy can also be associated with a cardiomyopathy.

MULTISYSTEM PROTEINOPATHIES (MSP)

The multisystem proteinopathies (MSPs) are genetically heterogenous disorders featured by hereditary inclusion body myopathy (IBM), amyotrophic lateral sclerosis, parkinsonism, frontotemporal dementia, and Paget disease of bone. Some forms have also been referred to as IBMPFD for some of the above major clinical features. Patients present in adulthood with progressive proximal or distal weakness. Serum CK is usually mildly elevated. EMG shows features of an irritable myopathy but also neurogenic features as well. Muscle biopsies in patients with myopathy show rimmed vacuoles, inclusions that immunostain with ubiquitin, and TDP-43 extrusion from myonuclei. Most are caused by mutations in genes that encode for RNA-binding proteins or proteins involved in the elimination of other aged proteins. There are at least five types of MSP (Table 460-7).

SPORADIC LATE-ONSET NEMALINE MYOPATHY

Clinical Features Sporadic late onset nemaline myopathy (SLONM) should not be confused with congenital forms of nemaline myopathy, which usually are congenital and/or hereditary in nature. SLONM is not a genetic disorder and usually presents after the age of 40 years with proximal extremity weakness. Some patients may present with an axial myopathy, isolated head drop, or bent spine syndrome from paraspinal muscle weakness. Ventilatory muscle involvement and cardiomyopathy may develop. Additionally, SLONM can complicate HIV infection.

Laboratory Features Serum CK is usually normal or mildly elevated and can be lower than normal. EMG reveals signs of an irritable myopathy. About 50% of cases are associated with a monoclonal gammopathy of undetermined significance (IgG or IgA). Muscle biopsies can reveal inflammatory cell infiltrates, trabeculated or lobulated fibers, many atrophic muscle fibers, and fibers with nemaline rods. The rods are often smaller than ones seen in the hereditary nemaline myopathies and may be missed on routine light microscopy if thickness of the sections is >3 μm. However, the rods are almost always appreciated on electron microscopy, and on immunohistochemistry, the rods are usually immunoreactive to anti-α-actinin antibody.

TREATMENT

SLONM

Some patients with SLONM respond to intravenous immunoglobulin or other immunosuppressive therapies. Autologous stem cell transplantation has been beneficial in some patients with SLONM and a monoclonal gammopathy.

TABLE 460-7 Multisystem Proteinopathies

MULTISYSTEM PROTEINOPATHY	INHERITANCE	GENE	AFFECTED PROTEIN
MSP1 / IBMPFD1	AD	*VCP*	Valosin-containing protein
MSP2 / IBMPFD2	AD	*HNRPA2B1*	HNRPA2B1
MSP3 / IBMPFD3	AD	*HNRNPA1*	HNRNPA1
MSP4	AD	*SQTM1*	Sequestome
MSP5	AD	*MTR3*	Matrin 3

Abbreviations: AD, autosomal dominant; HNRNPA1, heterogeneous nuclear ribonucleoprotein A1; HNRPA2B1, heterogeneous nuclear ribonucleoprotein A2/B1; IBMPFD, inclusion body myopathy, Paget disease, frontotemporal dementia; MSP, multisystem proteinopathy.

DISORDERS OF MUSCLE ENERGY METABOLISM

There are two principal sources of energy for skeletal muscle—fatty acids and glucose. Abnormalities in either glucose or lipid utilization can be associated with distinct clinical presentations that can range from an acute, painful syndrome with rhabdomyolysis and myoglobinuria to a chronic, progressive muscle weakness simulating muscular dystrophy (Table 460-1). As with the muscular dystrophies, there are no specific medical treatments available.

GLYCOGEN STORAGE AND GLYCOLYTIC DEFECTS

Disorders of Glycolysis Causing Exercise Intolerance Several glycolytic defects are associated with recurrent myoglobinuria. The most common is *McArdle disease* caused by mutations in the *PYGM* gene leading to *myophosphorylase deficiency*. Symptoms of muscle pain and stiffness usually begin in adolescence. With severe episodes, myoglobinuria can occur.

Certain features help distinguish some enzyme defects. In McArdle disease, exercise tolerance can be enhanced by a slow induction phase (warm-up) or brief periods of rest, allowing for the start of the "second-wind" phenomenon (switching to utilization of fatty acids). Varying degrees of hemolytic anemia accompany deficiencies of both phosphofructokinase (mild) and phosphoglycerate kinase (severe). In phosphoglycerate kinase deficiency, the usual clinical presentation is a seizure disorder associated with intellectual disability; exercise intolerance is an infrequent manifestation.

In all of these conditions, the serum CK levels fluctuate widely and may be elevated even during symptom-free periods. CK levels >100 times normal are expected accompanying myoglobinuria. A forearm exercise test reveals a blunted rise in venous lactate with a normal rise in ammonia. A definitive diagnosis of glycolytic disease can be made by muscle biopsy with appropriate staining and enzyme assays, but genetic testing is now done in lieu of biopsy in most cases.

Training may enhance exercise tolerance, perhaps by increasing perfusion to muscle. Dietary intake of free glucose or fructose prior to activity may improve function, but care must be taken to avoid obesity from ingesting too many calories.

Disorders of Glycogen Storage Causing Progressive Weakness • α-GLUCOSIDASE, OR ACID MALTASE, DEFICIENCY (POMPE DISEASE) Three clinical forms of α-glucosidase, or acid maltase, deficiency (*type II glycogenosis*) can be distinguished. The infantile form is the most common, with onset of symptoms in the first 3 months of life. Infants develop severe muscle weakness, cardiomegaly, hepatomegaly, and respiratory insufficiency. Glycogen accumulation in motor neurons of the spinal cord and brainstem contributes to muscle weakness. Death usually occurs by 1.5 years of age. In the childhood form, the picture resembles DMD with delayed motor milestones resulting from proximal limb muscle weakness and involvement of respiratory muscles. The heart may be involved, but the liver and brain are unaffected. The adult form usually begins in the third or fourth decade but can present as late as the seventh decade. Ventilatory weakness can be the initial and only manifestation in 20–30% of late-onset cases.

The serum CK level is 2–10 times normal in infantile or childhood-onset Pompe disease but can be normal in adult-onset cases. EMG can demonstrate muscle membrane irritability, particularly in the paraspinal muscles. The muscle biopsy in infants typically reveals vacuoles containing glycogen and the lysosomal enzyme acid phosphatase. Electron microscopy reveals membrane-bound and free tissue glycogen. However, muscle biopsies in late-onset Pompe disease may demonstrate only nonspecific abnormalities. Enzyme analysis of dried blood spots is a sensitive technique to screen for Pompe disease. A definitive diagnosis is established by genetic testing.

Pompe disease is inherited as an autosomal recessive disorder caused by mutations of the *α-glucosidase* gene. Enzyme replacement therapy (ERT) with IV recombinant human α-glucosidase is beneficial in infantile-onset Pompe disease. In late-onset cases, ERT has a more modest benefit.

OTHER GLYCOGEN STORAGE DISEASES WITH PROGRESSIVE WEAKNESS In *debranching enzyme deficiency* (*type III glycogenosis*), a slowly progressive form of muscle weakness can develop after puberty. Rarely, myoglobinuria may be seen. Patients are usually diagnosed in infancy, however, because of hypotonia and delayed motor milestones; hepatomegaly, growth retardation, and hypoglycemia are other manifestations. *Branching enzyme deficiency* (*type IV glycogenosis*) is a rare and fatal glycogen storage disease characterized by failure to thrive and hepatomegaly. Hypotonia and muscle wasting may be present, but the skeletal muscle manifestations are minor compared to liver failure. An autosomal dominant glycogen storage disease was reported in a single family that was due to a mutation in the *PYGM* gene that typically causes autosomal recessive McArdle disease. Affected individuals presented with progressive proximal weakness, no exercise intolerance, normal CK, and a normal lactic acid increase with exercise.

LIPID AS AN ENERGY SOURCE AND ASSOCIATED DEFECTS

Lipid is an important muscle energy source during rest and during prolonged, submaximal exercise. Oxidation of fatty acids occurs in the mitochondria. To enter the mitochondria, a fatty acid must first be converted to an "activated fatty acid," acyl-CoA. The acyl-CoA must be linked with carnitine by the enzyme CPT for transport into the mitochondria.

Carnitine Palmitoyltransferase 2 (CPT2) Deficiency CPT2 deficiency is the most common recognizable cause of recurrent myoglobinuria. Onset is usually in the teenage years or early twenties. Muscle pain and myoglobinuria typically occur after prolonged exercise but can also be precipitated by fasting or infections; up to 20% of patients do not exhibit myoglobinuria, however. Strength is normal between attacks. In contrast to disorders caused by defects in glycolysis, in which muscle cramps follow short, intense bursts of exercise, the muscle pain in CPT2 deficiency does not occur until the limits of utilization have been exceeded and muscle breakdown has already begun.

Serum CK levels and EMG findings are both usually normal between episodes. A normal rise of venous lactate during forearm exercise distinguishes this condition from glycolytic defects. Muscle biopsy does not show lipid accumulation and is usually normal between attacks. The diagnosis requires direct measurement of muscle CPT or genetic testing. Attempts to improve exercise tolerance with frequent meals and a low-fat, high-carbohydrate diet, or by substituting medium-chain triglycerides in the diet, have not proven to be beneficial.

MITOCHONDRIAL MYOPATHIES

Mitochondria play a key role in energy production. Oxidation of the major nutrients derived from carbohydrate, fat, and protein leads to the generation of reducing equivalents. The latter are transported through the respiratory chain in the process known as *oxidative phosphorylation*. The energy generated by the oxidation-reduction reactions of the respiratory chain is stored in an electrochemical gradient coupled to ATP synthesis.

A novel feature of mitochondria is their genetic composition. Each mitochondrion possesses a DNA genome that is distinct from that of the nuclear DNA. Human mitochondrial DNA (mtDNA) consists of a double-strand, circular molecule comprising 16,569 base pairs (bp). It codes for 22 transfer RNAs, 2 ribosomal RNAs, and 13 polypeptides of the respiratory chain enzymes. The genetics of mitochondrial diseases differ from the genetics of chromosomal disorders. The DNA of mitochondria is directly inherited from the cytoplasm of the gametes, mainly from the oocyte. The sperm contributes very little of its mitochondria to the offspring at the time of fertilization. Thus, mitochondrial genes are derived almost exclusively from the mother, accounting for maternal inheritance of some mitochondrial disorders.

Patients with mitochondrial myopathies have clinical manifestations that usually fall into three groups: chronic progressive external ophthalmoplegia (CPEO), skeletal muscle–CNS syndromes, and pure myopathy simulating muscular dystrophy or metabolic myopathy. Unfortunately, no specific medical therapies are clearly beneficial, although coenzyme Q10 supplements are often prescribed.

Kearns-Sayre Syndrome (KSS) This is a widespread multiorgan system disorder with a defined triad of clinical findings: onset before age 20, CPEO, and pigmentary retinopathy, plus one or more of the following features: complete heart block, cerebrospinal fluid (CSF) protein >1 g/L (100 mg/dL), or cerebellar ataxia. The cardiac disease includes syncopal attacks and cardiac arrest related to the abnormalities in the cardiac conduction system: prolonged intraventricular conduction time, bundle branch block, and complete atrioventricular block. Death attributed to heart block occurs in ~20% of the patients. Varying degrees of progressive limb muscle weakness and easy fatigability affect activities of daily living. Many affected individuals have intellectual disabilities. Endocrine abnormalities are also common, including gonadal dysfunction in both sexes with delayed puberty, short stature, and infertility. Diabetes mellitus occurs in ~13% of KSS patients. Other less common endocrine disorders include thyroid disease, hyperaldosteronism, Addison's disease, and hypoparathyroidism.

Serum CK and lactate levels are normal or slightly elevated. Serum levels of fibroblast growth factor 21 (FGF-21) and growth and differentiation factor 15 (GDF-15) are often elevated in mitochondrial disorders with muscle weakness. EMG is often myopathic. NCS may be abnormal related to an associated neuropathy. Muscle biopsies reveal ragged red fibers and cytochrome oxidase (COX)–negative fibers. By electron microscopy, there are increased numbers of mitochondria that often appear enlarged and contain paracrystalline inclusions.

KSS is a sporadic disorder caused by single mtDNA deletions that are presumed to arise spontaneously in the ovum or zygote. The most common deletion, occurring in about one-third of patients, removes 4977 bp of contiguous mtDNA. Monitoring for cardiac conduction defects is critical. Prophylactic pacemaker implantation is indicated when ECGs demonstrate a bifascicular block.

Progressive External Ophthalmoplegia (PEO) PEO can be caused by nuclear DNA mutations affecting mtDNA and thus inherited in a Mendelian fashion or by mutations in mtDNA. Onset is usually after puberty. Fatigue, exercise intolerance, dysphagia, and complaints of muscle weakness are typical. The neurologic examination confirms the ptosis and ophthalmoplegia, usually asymmetric in distribution. Patients do not complain of diplopia. Mild facial, neck flexor, and proximal weakness is typical. Rarely, respiratory muscles may be progressively affected and may be the direct cause of death.

Serum CK and lactate can be normal or mildly elevated. The EMG can be myopathic. Ragged red and COX-negative fibers are prominently displayed in the muscle biopsy.

This autosomal dominant form of CPEO is most commonly caused by mutations in the genes encoding adenine nucleotide translocator 1 (*ANT1*), twinkle gene (*C10orf2*), and mtDNA polymerase 1 (*POLG1*). Autosomal recessive PEO can also be caused by mutations in *POLG1*. Point mutations have been identified within various mitochondrial tRNA (Leu, Ile, Asn, Trp) genes in families with maternal inheritance of PEO.

There is no specific medical treatment available; exercise may improve function, but this will depend on the patient's ability to participate.

Myoclonic Epilepsy with Ragged Red Fibers (MERRF) The onset of MERRF is variable, ranging from late childhood to middle adult life. Characteristic features include myoclonic epilepsy, cerebellar ataxia, and progressive proximal muscle weakness. The seizure disorder is an integral part of the disease and may be the initial symptom. Cerebellar ataxia precedes or accompanies epilepsy. Other more variable features include dementia, peripheral neuropathy, optic atrophy, hearing loss, and diabetes mellitus.

Serum CK levels and lactate may be normal or elevated. EMG is myopathic, and in some patients, NCS show a neuropathy. The electroencephalogram is abnormal, corroborating clinical findings of epilepsy. Typical ragged red fibers are seen on muscle biopsy. MERRF is caused by maternally inherited point mutations of mitochondrial tRNA genes. The most common mutation found in 80% of MERRF patients is an A to G substitution at nucleotide 8344 of tRNA lysine (A8344G tRNA^{lys}). Only supportive treatment is possible, with special attention to epilepsy.

Mitochondrial Myopathy, Encephalopathy, Lactic Acidosis, and Stroke-like Episodes (MELAS) MELAS is the most common mitochondrial encephalomyopathy. The term *stroke-like* is appropriate because the cerebral lesions do not conform to a strictly vascular distribution. The onset in the majority of patients is before age 20. Seizures, usually partial motor or generalized, are common and may represent the first clearly recognizable sign of disease. The cerebral insults that resemble strokes cause hemiparesis, hemianopia, and cortical blindness. A presumptive stroke occurring before age 40 should place this mitochondrial encephalomyopathy high in the differential diagnosis. Associated conditions include hearing loss, diabetes mellitus, hypothalamic pituitary dysfunction causing growth hormone deficiency, hypothyroidism, and absence of secondary sexual characteristics. In its full expression, MELAS leads to dementia, a bedridden state, and a fatal outcome. Serum lactic acid is typically elevated.

The CSF protein is also increased but is usually ≤1 g/L (100 mg/dL). Muscle biopsies show ragged red fibers. Neuroimaging demonstrates basal ganglia calcification in a high percentage of cases. Focal lesions that mimic infarction are present predominantly in the occipital and parietal lobes. Strict vascular territories are not respected, and cerebral angiography fails to demonstrate lesions of the major cerebral blood vessels.

MELAS is usually caused by maternally inherited point mutations of mitochondrial tRNA genes. The A3243G point mutation in tRNA$^{Leu(UUR)}$ is the most common, occurring in ~80% of MELAS cases. No specific treatment is available. Supportive treatment is essential for the stroke-like episodes, seizures, and endocrinopathies.

Mitochondrial DNA Depletion Syndromes Mitochondrial DNA depletion syndrome (MDS) is a heterogeneous group of disorders that are inherited in an autosomal recessive fashion and can present in infancy or in adults. MDS can be caused by mutations in several genes (*TK2*, *DGUOK*, *RRM2B*, *TYMP*, *SUCLA1*, and *SUCLA2*) that lead to depletion of mitochondrial deoxyribonucleotides (dNTP) necessary for mtDNA replication. The other major cause of MDS is a set of mutations in genes essential for mtDNA replication (e.g., *POLG1* and *C10orf2*). The clinical phenotypes associated with MDS vary. Patients may develop a severe encephalopathy (e.g., Leigh's syndrome), PEO, an isolated myopathy, myo-neuro-gastrointestinal encephalopathy (MNGIE), and a sensory neuropathy with ataxia.

DISORDERS OF MUSCLE MEMBRANE EXCITABILITY

Muscle membrane excitability is affected in a group of disorders referred to as *channelopathies*. These disorders usually present with episodic muscle weakness (periodic paralysis) and sometimes myotonia or paramyotonia (Table 460-1).

CALCIUM CHANNEL DISORDERS OF MUSCLE

Hypokalemic Periodic Paralysis (HypoKPP) This is an autosomal dominant disorder with onset in adolescence. Males are more often affected because of decreased penetrance in females. Episodic weakness with onset after age 25 is almost never due to periodic paralyses, with the exception of thyrotoxic periodic paralysis. Attacks are often provoked by meals high in carbohydrates or sodium and may accompany rest following prolonged exercise. Weakness usually affects proximal limb muscles more than distal. Ocular and bulbar muscles are less likely to be affected. Respiratory muscles are usually spared, but when they are involved, the condition may prove fatal. Weakness may take as long as 24 h to resolve. Life-threatening cardiac arrhythmias related to hypokalemia may occur during attacks. As a late complication, patients commonly develop severe, disabling proximal lower extremity weakness.

Attacks of thyrotoxic periodic paralysis resemble those of primary HypoKPP. Despite a higher incidence of thyrotoxicosis in women, men, particularly those of Asian descent, are more likely to manifest this complication. Attacks abate with treatment of the underlying thyroid condition.

A low serum potassium level during an attack, excluding secondary causes, establishes the diagnosis. In the midst of an attack of weakness, motor conduction studies may demonstrate reduced amplitudes, whereas EMG may show electrical silence in severely weak muscles. In between attacks, the EMG and routine NCS are normal. However, a long exercise NCS test may demonstrate decrementing amplitudes.

HypoKPP type 1 is the most common form and is caused by mutations in the voltage-sensitive, skeletal muscle calcium channel gene, *CALCL1A3*. Approximately 10% of cases are HypoKPP type 2, arising from mutations in the voltage-sensitive sodium channel gene (*SCN4A*). In both forms, the mutations lead to an abnormal gating pore current that predisposes the muscle cell to depolarize when potassium levels are low.

TREATMENT

Hypokalemic Periodic Paralysis

Mild attacks usually do not require medical treatment. However, severe attacks of weakness can be improved by the administration of potassium. Oral KCl (0.2–0.4 mmol/kg) can be given every 30 min. Only rarely is IV therapy necessary (e.g., when swallowing problems or vomiting is present). The long-term goal of therapy is to avoid attacks. Patients should be made aware of the importance of a low-carbohydrate, low-sodium diet and consequences of intense exercise. Prophylactic administration of acetazolamide or dichlorphenamide can reduce attacks of periodic weakness. However, in patients with HypoKPP type 2, attacks of weakness can be exacerbated with these medications.

SODIUM CHANNEL DISORDERS OF MUSCLE

Hyperkalemic Periodic Paralysis (HyperKPP) The term *hyperkalemic* is misleading because patients are often normokalemic during attacks. That attacks are precipitated by potassium administration best defines the disease. The onset is usually in the first decade; males and females are affected equally. Attacks are brief and mild, usually lasting 30 min to several hours. Weakness affects proximal muscles, sparing bulbar muscles. Attacks are precipitated by rest following exercise and fasting.

Potassium may be slightly elevated or normal during an attack. As in HypoKPP, NCS in HyperKPP muscle may demonstrate reduced motor amplitudes and the EMG may be silent in very weak muscles. A long exercise NCS test can reveal diminished amplitudes as well. The EMG may demonstrate myotonic discharges. HyperKPP is caused by mutations of the voltage-gated sodium channel *SCN4A* gene. Acetazolamide or dichlorphenamide can reduce the frequency and severity of attacks. Mexiletine may be helpful in patients with significant clinical myotonia.

Paramyotonia Congenita In PC, the attacks of weakness are cold-induced or occur spontaneously and are mild. Myotonia is a prominent feature but worsens with muscle activity (paradoxical myotonia). This is in contrast to classic myotonia in which exercise alleviates the condition. Attacks of weakness are seldom severe enough to require emergency room treatment. Over time, patients develop inter-attack weakness as they do in other forms of periodic paralysis.

Serum CK is usually mildly elevated. Routine NCS are normal. Short exercise NCS tests may be abnormal, however, and cooling of the muscle often dramatically reduces the amplitude of the compound muscle action potentials. EMG reveals diffuse myotonic potentials in

PC. Upon local cooling of the muscle, the myotonic discharges disappear as the patient becomes unable to activate MUAPs.

PC is inherited as an autosomal dominant condition; voltage-gated sodium channel mutations are responsible, and thus, this disorder is allelic with HyperKPP. Mexiletine is reported to be helpful in reducing the myotonia.

POTASSIUM CHANNEL DISORDERS

Andersen-Tawil Syndrome This rare disease is characterized by episodic weakness, cardiac arrhythmias, and dysmorphic features (short stature, scoliosis, clinodactyly, hypertelorism, small or prominent low-set ears, micrognathia, and broad forehead). The cardiac arrhythmias are potentially serious and life threatening. They include long QT, ventricular ectopy, bidirectional ventricular arrhythmias, and tachycardia. The disease is most commonly caused by mutations of the inwardly rectifying potassium channel (*Kir 2.1*) gene that heighten muscle cell excitability. The episodes of weakness may differ between patients because of potassium variability. Acetazolamide may decrease the attack frequency and severity.

CHLORIDE CHANNEL DISORDERS

Two forms of this disorder, autosomal dominant (*Thomsen disease*) and autosomal recessive (*Becker disease*), are both caused by mutations in the chloride channel 1 gene (*CLCN1*). Symptoms are noted in infancy and early childhood. The severity lessens in the third to fourth decade. Myotonia is worsened by cold and improved by activity. The gait may appear slow and labored at first but improves with walking. In Thomsen disease, muscle strength is normal, but in Becker disease, which is usually more severe, there may be muscle weakness. Muscle hypertrophy is usually present. Myotonic discharges are prominently displayed by EMG recordings. Serum CK is normal or mildly elevated. Mexiletine is helpful in relieving the myotonia.

ENDOCRINE AND METABOLIC MYOPATHIES

Endocrinopathies can cause weakness, but fatigue is more common than true weakness. The serum CK level is often normal (except in hypothyroidism), and the muscle histology is characterized by atrophy rather than destruction of muscle fibers. Nearly all endocrine myopathies respond to treatment.

THYROID DISORDERS

Hypothyroidism (Chap. 395) Patients with hypothyroidism have frequent muscle complaints, and about one-third have proximal muscle weakness. Muscle cramps, pain, and stiffness are common. Some patients have enlarged muscles. Features of slow muscle contraction and relaxation occur in 25% of patients; the relaxation phase of muscle stretch reflexes is characteristically prolonged and best observed at the ankle or biceps brachii reflexes. The serum CK level is often elevated (up to 10 times normal). EMG is typically normal. Muscle biopsy shows no distinctive morphologic abnormalities.

Hyperthyroidism (Chap. 396) Patients who are thyrotoxic commonly have proximal muscle weakness, but they rarely complain of myopathic symptoms. Activity of deep tendon reflexes may be enhanced. Fasciculations may be apparent and, when coupled with increased muscle stretch reflexes, may lead to an erroneous diagnosis of amyotrophic lateral sclerosis. A form of hypokalemic periodic paralysis can occur in patients who are thyrotoxic. Mutations in the *KCNJ18* gene that encodes for the inwardly rectifying potassium channel, Kir 2.6, have been discovered in up to a third of cases.

PARATHYROID DISORDERS (SEE ALSO CHAP. 422)

Hyperparathyroidism Proximal muscle weakness, muscle wasting, and brisk muscle stretch reflexes are the main features of this endocrinopathy. Some patients develop neck extensor weakness (part of the dropped head syndrome). Serum CK levels are usually normal or slightly elevated. Serum parathyroid hormone levels are elevated, while vitamin D and calcium levels are usually reduced. Muscle biopsies show only mild type 2 fiber atrophy.

Hypoparathyroidism An overt myopathy due to hypocalcemia rarely occurs. Neuromuscular symptoms are usually related to localized or generalized tetany. Serum CK levels may be increased secondary to muscle damage from sustained tetany. Hyporeflexia or areflexia is usually present and contrasts with the hyperreflexia in hyperparathyroidism.

ADRENAL DISORDERS (SEE ALSO CHAP. 398)

Conditions associated with glucocorticoid excess cause a myopathy; steroid myopathy is the most commonly diagnosed endocrine muscle disease. Proximal muscle weakness combined with a cushingoid appearance are the key clinical features. Serum CK and EMG are normal. Muscle biopsy, not typically done for diagnostic purposes, reveals type 2b muscle fiber atrophy. In primary hyperaldosteronism (*Conn's syndrome*), neuromuscular complications are due to potassium depletion. The clinical picture is one of persistent muscle weakness. Long-standing hyperaldosteronism may lead to proximal limb weakness and wasting. Serum CK levels may be elevated, and a muscle biopsy may demonstrate necrotic fibers. These changes relate to hypokalemia and are not a direct effect of aldosterone on skeletal muscle.

PITUITARY DISORDERS (SEE ALSO CHAP. 392)

Patients with acromegaly usually have mild proximal weakness. Muscles often appear enlarged but exhibit decreased force generation. The duration of acromegaly, rather than the serum growth hormone levels, correlates with the degree of myopathy.

DIABETES MELLITUS (SEE ALSO CHAP. 417)

Neuromuscular complications of diabetes mellitus are most often related to neuropathy. The only notable myopathy is ischemic infarction of leg muscles, usually involving one of the thigh muscles but on occasion affecting the distal leg. This condition occurs in patients with poorly controlled diabetes and presents with the abrupt onset of pain, tenderness, and edema of a thigh or calf. The area of muscle infarction is hard and indurated. The muscles most often affected include the vastus lateralis, thigh adductors, and biceps femoris. Computed tomography (CT) or MRI can demonstrate focal abnormalities in the affected muscle. Diagnosis by imaging is preferable to muscle biopsy, if possible, as hemorrhage into the biopsy site can occur.

MYOPATHIES OF SYSTEMIC ILLNESS

Systemic illnesses such as chronic respiratory, cardiac, or hepatic failure are frequently associated with severe muscle wasting and complaints of weakness. Fatigue is usually a more significant problem than weakness, which is typically mild.

DRUG-INDUCED OR TOXIC MYOPATHIES

The most common toxic myopathies are caused by the cholesterol-lowering agents and glucocorticoids. Others impact practice to a lesser degree but are important to consider in specific situations. Table 460-8 provides a comprehensive list of drug-induced myopathies with their distinguishing features.

MYOPATHY FROM LIPID-LOWERING AGENTS

All classes of lipid-lowering agents have been implicated in muscle toxicity, including HMG-CoA reductase inhibitors (statins) and, to a much lesser extent, fibrates, niacin, and ezetimibe. Myalgia and elevated CKs are the most common manifestations. Rarely, patients exhibit proximal weakness or myoglobinuria. Concomitant use of statins with fibrates and cyclosporine increases the risk of severe

TABLE 460-8 Drug-Induced Myopathies

DRUGS	MAJOR TOXIC REACTION
Lipid-lowering agents HMG-CoA reductase inhibitors Fibric acid derivatives Niacin (nicotinic acid)	Drugs belonging to all three of the major classes of lipid-lowering agents can produce a spectrum of toxicity: asymptomatic serum creatine kinase elevation, myalgias, exercise-induced pain, rhabdomyolysis, and myoglobinuria.
Glucocorticoids	Acute, high-dose glucocorticoid treatment can cause acute quadriplegic myopathy. These high doses of steroids are often combined with nondepolarizing neuromuscular blocking agents, but the weakness can occur without their use. Chronic steroid administration produces predominantly proximal weakness.
Nondepolarizing neuromuscular blocking agents	Acute quadriplegic myopathy can occur with or without concomitant glucocorticoids.
Zidovudine	Mitochondrial myopathy with ragged red fibers.
Drugs of abuse Alcohol Amphetamines Cocaine Heroin Phencyclidine Meperidine	All drugs in this group can lead to widespread muscle breakdown, rhabdomyolysis, and myoglobinuria. Local injections cause muscle necrosis, skin induration, and limb contractures.
Autoimmune myopathy Statins Checkpoint inhibitors D-Penicillamine	Use of statins may cause an immune-mediated necrotizing myopathy associated with HMG-CoA reductase antibodies. Checkpoint inhibitors can be complicated by myositis, myocarditis, myasthenia gravis, and immune-mediated neuropathies. Myasthenia gravis has also been reported with penicillamine.
Amphophilic cationic drugs Amiodarone Chloroquine Hydroxychloroquine	All amphophilic drugs have the potential to produce painless, proximal weakness associated with necrosis and autophagic vacuoles in the muscle biopsy.
Antimicrotubular drugs Colchicine	This drug produces painless, proximal weakness especially in the setting of renal failure. Muscle biopsy shows necrosis and fibers with autophagic vacuoles.

myotoxicity. EMG demonstrates irritability, and myopathic units and muscle biopsies reveal necrotic muscle fibers in weak muscles. Severe myalgia, weakness, marked elevations in serum CK (>3–5 times baseline), and myoglobinuria are indications for stopping the drug. Patients usually improve with drug cessation, although this may take several weeks. Rare cases continue to progress after the offending agent is discontinued. It is possible that in such cases the statin may have triggered an immune-mediated necrotizing myopathy, as these individuals require immunotherapy (e.g., intravenous immunoglobulin or immunosuppressive agents) to improve and often relapse when these therapies are discontinued **(Chap. 377)**. Autoantibodies directed against HMG-CoA reductase have been identified in many of these cases.

GLUCOCORTICOID-RELATED MYOPATHIES

Glucocorticoid myopathy occurs with chronic treatment or as "acute quadriplegic" myopathy secondary to high-dose IV glucocorticoid use. Chronic administration produces proximal weakness accompanied by cushingoid manifestations, which can be quite debilitating; the chronic use of prednisone at a daily dose of ≥30 mg/d is most often associated with toxicity. Patients taking fluorinated glucocorticoids (triamcinolone, betamethasone, dexamethasone) appear to be at especially high risk for myopathy. In chronic steroid myopathy, the serum CK is usually normal. Serum potassium may be low. The muscle biopsy in chronic cases shows preferential type 2 muscle fiber atrophy; this is not reflected in the EMG, which is usually normal.

Patients receiving high-dose IV glucocorticoids for status asthmaticus, chronic obstructive pulmonary disease, organ transplantation, or other indications may develop severe generalized weakness (critical illness myopathy). This myopathy, also known as acute quadriplegic myopathy, can also occur in the setting of sepsis. Involvement of the diaphragm and intercostal muscles causes ventilatory muscle weakness and is usually appreciated when patients are unable to be weaned off a ventilatory in the intensive care unit. NCS demonstrate reduced compound muscle action potentials in the setting of relatively preserved sensory potentials. EMG can demonstrate abnormal insertional and spontaneous activity and early recruitment of myopathic appearing units in those muscles that can be activated. Muscle biopsy can show a distinctive loss of thick filaments (myosin) by electron microscopy. Treatment is withdrawal of glucocorticoids and physical therapy, but the recovery is slow. Patients require supportive care and rehabilitation.

OTHER DRUG-INDUCED MYOPATHIES

Certain drugs produce painless, largely proximal muscle weakness. These drugs include the amphophilic cationic drugs (amiodarone, chloroquine, hydroxychloroquine) and antimicrotubular drugs (colchicine) (Table 460-6). Muscle biopsy can be useful in the identification of toxicity because autophagic vacuoles are prominent pathologic features of these toxins.

GLOBAL ISSUES

As previously discussed, certain dystrophies have an increased prevalence in different parts of the world. LGMD2A/LGMDR1 is the most common LGMD in individuals from Spain, France, Italy, and Great Britain; LGMD2I/LGMDR9 is more common in those with northern European ancestry. GNE myopathy is the most common form of distal myopathy in Japan but is also prevalent in the Ashkenazi population. OPMD is most common in those with ancestry from Spain and French-Canada as well as among Ashkenazi. Epidemiologic studies are lacking regarding other forms of myopathy and their prevalence in different areas of the world.

FURTHER READING

Amato AA et al: *Amato and Russell's Neuromuscular Disorders*, 3rd ed. McGraw Hill, 2025.

Chin HL et al: A clinical approach to diagnosis and management of mitochondrial myopathies. Neurotherapeutics 21:e00304, 2024.

Doughty CT, Amato AA: Toxic myopathies. Continuum (Minneap Minn) 25:1712, 2019.

Heller SA et al: Emery-Dreifuss muscular dystrophy. Muscle Nerve 61:436, 2020.

Johnson NE: Myotonic muscular dystrophies. Continuum (Minneap Minn) 25:1682, 2019.

Johnson NE, Statland JM: The limb-girdle muscular dystrophies. Continuum (Minneap Minn) 28:1698, 2022.

Mah JK et al: A systematic review and meta-analysis on the epidemiology of the muscular dystrophies. Can J Neurol Sci 43:163, 2016.

Mul K: Facioscapulohumeral Muscular Dystrophy. Continuum (Minneap Minn) 28:1735, 2022.

Rodolico C et al: Endocrine myopathies: Clinical and histopathological features of the major forms. Acta Myol 39:130, 2020.

Rosow LK, Amato AA: The role of electrodiagnostic testing, imaging, and muscle biopsy in the investigation of muscle disease. Continuum (Minneap Minn) 22:1787, 2016.

Straub V et al: LGMD Workshop Study Group. 229th ENMC international workshop: Limb girdle muscular dystrophies—Nomenclature and reformed classification Naarden, the Netherlands, 17-19 March 2017. Neuromuscul Disord 28:702, 2018.

Section 4 Syndromes Associated with Chronic Fatigue

461 Myalgic Encephalomyelitis/ Chronic Fatigue Syndrome

Elizabeth R. Unger, Jin-Mann S. Lin, Jeanne Bertolli

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a chronic complex illness with multisystem manifestations and long-term impact on functional impairment comparable to multiple sclerosis, rheumatoid arthritis, and congestive heart failure. The hallmark of ME/CFS is persistent and unexplained fatigue resulting in significant impairment in daily functioning, along with worsening symptoms following physical or mental exertion that would have been tolerated before illness (postexertional malaise). Besides intense fatigue, many patients report concomitant symptoms such as pain, cognitive dysfunction, and unrefreshing sleep. Additional symptoms can include headache, sore throat, tender lymph nodes, muscle aches, joint aches, feverishness, difficulty sleeping, psychiatric problems, allergies, and abdominal cramps. The recognition that ME/CFS is one diagnosable condition in Long COVID has raised clinical awareness about this poorly understood illness, although patients still face stigma and misunderstanding among health care providers.

The condition has been known by many names, and debate about the name and case definition continues. The composite name ME/CFS was adopted by the U.S. Department of Health and Human Services in recognition of the limitations of either ME (absence of definitive inflammation in brain and spinal cord) or CFS (trivializes an often devastating illness through confusion with fatigue that everyone experiences).

EPIDEMIOLOGY

Determining how frequently ME/CFS occurs and characteristics of those affected has been complicated by variability in study design and application of case definitions. In the absence of a simple diagnostic test, evaluation by an experienced clinician is required for case identification. Clinic-based studies most accurately identify patients with ME/CFS but overrepresent higher socioeconomic groups with access to ME/CFS clinics. Population-based studies with or without a clinical evaluation estimated that between 836,000 and 3.3 million Americans have ME/CFS. However, studies indicate that ≥80% of those meeting criteria for ME/CFS had not been diagnosed by a health care provider. The illness costs the U.S. economy between $18 and $51 billion annually in medical costs and lost income. ME/CFS is three to four times more common in women than men. The highest prevalence is among those 40–50 years of age, but the age range is broad and includes children and adolescents. Persons of all races and ethnicities are affected, and there is some evidence that socioeconomically disadvantaged groups are at increased risk.

RISK FACTORS AND PATHOPHYSIOLOGY

A wide variety of infectious agents have been reported to be associated with a postinfectious fatiguing illness resembling ME/CFS. These include both viral and nonviral pathogens, such as Epstein-Barr virus, Ross River virus, *Coxiella burnetti* (Q fever), Ebola virus, SARS-CoV-1, and *Giardia*. While recovery from these infections is the rule, ~10% of those infected remain ill for ≥6 months. Most recently, published reports suggest that SARS-CoV-2 infection is also associated with prolonged fatiguing illness. Host and pathogen factors associated with recovery versus persistent disease remain elusive. In addition to infectious insults, a variety of stressors, including toxins, physical trauma, adverse events, and allostatic load (or "wear and tear" on the body), have been found to be associated with ME/CFS. Twin studies and family histories suggest a role for shared environment as well as genetic factors.

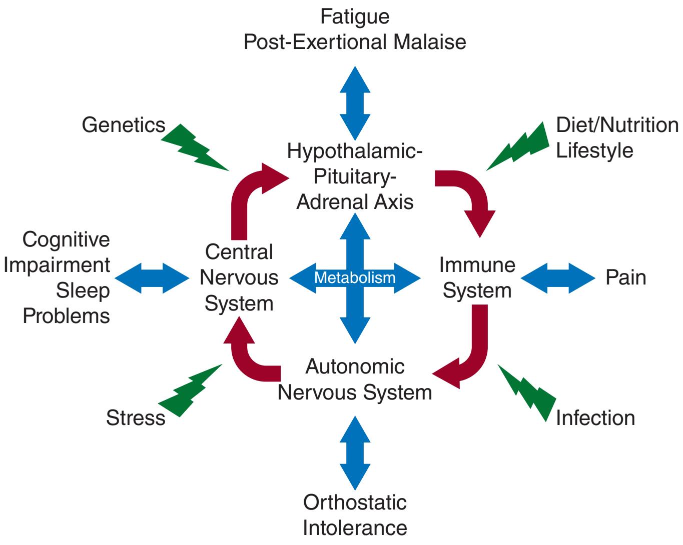

FIGURE 461-1 A multisystem model for myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). An example of a unifying model for ME/CFS demonstrating the interactions of multiple organ systems and environmental, genetic, and behavioral factors contributing to symptoms.

Evidence for immunologic dysfunction is inconsistent. Modest elevations in titers of antinuclear antibodies, reductions in immunoglobulin subclasses, deficiencies in mitogen-driven lymphocyte proliferation, reductions in natural killer cell activity, disturbances in cytokine production, and altered T-cell metabolism have been described. None of these immune findings has been firmly established and none of these changes appear in most patients. In theory, symptoms of ME/CFS could result from excessive production of a cytokine, such as interleukin 1 or interferon α, which induces fatigue and other flulike symptoms; however, compelling data in support of this hypothesis are lacking.

Other studies have reported various nonspecific changes in regional brain structures estimated by magnetic resonance imaging; dysfunction of the autonomic nervous system; abnormalities in the hypothalamic-pituitary-adrenal (HPA) axis; altered metabolism; and dysbiosis of the intestinal microbiome. Confirmatory studies are needed, and none of the findings are consistent enough to be used for diagnosis. It is clear that ME/CFS represents a complex disorder with alterations in multiple interrelated homeostatic systems. A variety of unifying models for the illness have been proposed, and discoveries about the pathophysiology of ME/CFS hold promise for elucidating novel mechanisms and interactions important in other illnesses **(Fig. 461-1)**.

APPROACH TO THE PATIENT

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome

DIAGNOSIS

A diagnosis of ME/CFS is made based on patient-reported symptoms that fit a characteristic profile. After a careful review of the literature and symptom-based case definitions for ME, CFS, or ME/CFS, the Institute of Medicine (IOM) committee recommended in 2015 straightforward diagnostic criteria **(Table 461-1)**. This includes the symptoms consistently noted in prior consensus case definitions: fatigue limiting the patient's ability to participate in their usual pre-illness activities, sleep problems, and postexertional malaise (PEM). PEM is a relapse in symptoms triggered by physical, emotional, or mental exertion that would not have been problematic for the patient before onset of ME/CFS. The relapse lasts more than a day and sometimes weeks. In addition, either difficulty

TABLE 461-1 2015 Institute of Medicine Diagnostic Criteria for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome

Substantial reduction or impairment in the ability to engage in pre-illness levels of activity (occupational, educational, social, or personal life) that: • lasts for >6 months • is accompanied by fatigue that is often profound, of new or definite onset (not lifelong), not the result of ongoing excessive exertion, and is not substantially alleviated by rest
Postexertional malaise (PEM)[a]—worsening of symptoms after physical, mental, or emotional exertion that would not have caused a problem before the illness
Unrefreshing sleep[a]
Cognitive impairment or **orthostatic intolerance**[a]

[a]Frequency and severity of symptoms should be assessed; should be present at least half of the time and with at least moderate intensity.

thinking and concentrating (often referred to by patients as "brain fog") or orthostatic intolerance should be present.

Patients with ME/CFS may experience a wide range of other symptoms not specified in the IOM diagnostic criteria **(Table 461-2)**. As a result, patients meeting ME/CFS criteria could have very different clinical features based on the type, frequency, and severity of their symptoms. Patients may describe a precipitating cause for their illness, such as a known or presumed infection, but frequently no initiating factor is recognized. The symptoms may occur suddenly within a day or week or may occur gradually.

While the diagnostic criteria specifies that illness must be present at least 6 months, the possibility of ME/CFS should be considered for patients with consistent symptoms persisting >1 month, and evaluation and supportive care can begin as early as 4–6 weeks after onset. Listening to patients' descriptions of what they are experiencing is important. Asking questions can help patients accurately describe their experience with fatigue and PEM. These include asking about current activity levels compared with before they became ill, what happens when they are as active as they were pre-illness, and how long it takes to recover after exertion. Whereas patients recognize relapses, the relation of relapse to activity level may not be apparent, and as a result, PEM may not be recognized. Patients may also appear well during an office visit, only to relapse afterward from exertion surrounding the consultation.

Although the 2015 IOM ME/CFS criteria do not list medical or psychological conditions that exclude the diagnosis of ME/CFS, a careful clinical evaluation is required to identify and treat other illnesses that could explain or contribute to the patient's symptoms. The initial evaluation also requires reviewing family history, medical history (including infections, traumas/surgeries, and occupational exposure to environmental toxins), and medications and supplements; performing a physical examination, including lean test for postural orthostatic tachycardia syndrome (POTS; **Chap. 451**); a mental health assessment (screen for depression and anxiety); and routine screening laboratory tests (if recent results are not on record). As routine laboratory tests are usually within normal limits, their role is in identifying other illnesses, and the specific panel of tests should be adjusted based on the patient's presentation. Typically the tests include complete blood count, erythrocyte sedimentation rate, electrolytes, fasting glucose, renal function tests (blood urea nitrogen, glomerular filtration rate), calcium, phosphate, liver function (bilirubin, alanine aminotransferase, alkaline phosphatase, aspartate aminotransferase, gamma-glutamyl transferase, total protein, albumin/globulin ratio), C-reactive protein, thyroid function (thyroid-stimulating hormone, free thyroxine), iron studies to assess both iron overload and iron deficiency (serum iron, transferrin saturation, ferritin), celiac disease screening tests, and urinalysis.

DIFFERENTIAL DIAGNOSIS AND COMORBID CONDITIONS

While the differential diagnosis for fatigue is quite broad **(Chap. 25)**, further workups and referrals should be chosen carefully based on the patient's history, symptoms (particularly those that are new, worsening, or unusual), and results of initial laboratory tests. Conditions reported to occur in association with ME/CFS **(Table 461-3)** should be kept in mind during the evaluation and follow-up, as management and treatment modalities for these comorbidities could contribute to an improved quality of life.

MANAGEMENT

While there are no approved drugs to treat or cure ME/CFS, patients benefit from receiving a diagnosis and an individualized plan that addresses the symptoms that are most problematic for the patient. Some symptoms, in particular, disturbed sleep **(Chap. 33)** and pain **(Chap. 14)**, may improve with nonpharmacologic therapies (e.g., sleep hygiene, massage, acupuncture, hot or cold packs) or medications. Any medications should be started at lower doses than usual and only slowly increased. Patients with ME/CFS have been reported to be more sensitive to medications than the general population, and benefits with fewer toxicities may be achieved at lower doses. Narcotics should be avoided, and referral to sleep centers or other specialists may be required.

Controlled therapeutic trials have not established significant benefit for patients with ME/CFS from acyclovir, fludrocortisone, galantamine, modafinil, and IV immunoglobulin, among other agents. These studies have been limited by small numbers and lack power to investigate benefit in patient subgroups. Preliminary small studies reported the possible effectiveness of the B-cell-targeting anti-CD20 monoclonal antibody rituximab in ME/CFS, but a subsequent large, well-designed, prospective, double-blind study found no benefit. Numerous anecdotes circulate regarding other traditional and nontraditional therapies. It is important to guide patients away from therapeutic modalities that are toxic, expensive, or unreasonable.

Educating the patient and family about PEM can be helpful in avoiding the harmful cycle of overexertion during "good days" followed by relapse that can negate any functional gains. This is often referred to as "push and crash." Recognizing limits and using activity management (pacing) can help limit PEM. It is important to maintain tolerated activity levels to minimize deconditioning. Activity may be advanced very gradually as tolerated.

TABLE 461-2 Additional Symptoms Experienced by Patients with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome

Joint pain without swelling or redness
Muscle aches
New headaches
Tender lymph nodes
Sensitivity to sensory stimuli (e.g., light, noise, smells)
Sore throat
Shortness of breath
Irregular heartbeat
Alcohol intolerance
Difficulties with temperature regulation (feeling feverish or chilled)

TABLE 461-3 Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Comorbid Conditions

Chronic overlapping pain conditions: fibromyalgia (FM), chronic migraine, temporomandibular joint disease (TMJ), irritable bowel syndrome (IBS), endometriosis, vulvodynia, urologic chronic pelvic pain syndromes (UCPPS)
Postural orthostatic tachycardia syndrome (POTS)
Allergies
Sjögren's syndrome
Ehlers-Danlos syndrome
Mast cell activation syndrome (MCAS)
Dysautonomia
Multiple chemical sensitivities

Counseling may help patients and their families cope with the long-term consequences of living with a chronic illness. Consultation with a physical or occupational therapist may identify energy-saving strategies for activities of daily living as well as needed accommodations, such as a wheelchair for activities that require walking longer distances or prolonged standing.

COURSE AND PROGNOSIS

The illness severity varies from mild or moderate, with patients retaining varying degrees of pre-illness function, to severe, with patients essentially homebound. Most patients experience some improvement and stabilize, although return to their prior level of function is unusual. A continued decline in function should prompt evaluation for other illnesses. Patients should be re-evaluated at scheduled intervals to adjust treatments and detect any intercurrent disease. New or changing symptoms should be worked up to identify any new illnesses. Given the social isolation and loss of hope associated with a debilitating chronic illness, serious depression and an increased risk of suicide are reported for patients with ME/CFS. Clinicians should be prepared to screen for this and refer patients as needed.

FURTHER READING

Centers for Disease Control and Prevention: Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). Available from *https://www.cdc.gov/me-cfs/about/index.html*. Accessed June 4, 2024.

Choutka J et al: Unexplained post-acute infection syndromes. Nat Med 28:911, 2022.

Grach SL et al: Diagnosis and management of myalgic encephalomyelitis/chronic fatigue syndrome. Mayo Clin Proc 98:1544, 2023.

Institute of Medicine: *Beyond Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Redefining an Illness*. Washington, DC: The National Academies Press, 2015.

Komaroff AL et al: ME/CFS and Long COVID share similar symptoms and biological abnormalities: Road map to the literature. Front Med (Lausanne) 10:1187163, 2023.

Lapp CW: Initiating care of a patient with myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). Front Pediatr 6:415, 2019.

Rowe PC et al: Myalgic encephalomyelitis/chronic fatigue syndrome diagnosis and management in young people: A primer. Front Pediatr 5:121, 2017.

Vahratian A et al: Myalgic encephalomyelitis/chronic fatigue syndrome in adults: United States, 2021-2022. NCHS Data Brief 488:1, 2023.

Walitt B et al: Deep phenotyping of post-infectious myalgic encephalomyelitis/chronic fatigue syndrome. Nat Commun 15:907, 2024.

Section 5 Psychiatric and Addiction Disorders

462

Biology of Psychiatric Disorders

Robert O. Messing, Eric J. Nestler,
Matthew W. State

Psychiatric disorders are central nervous system diseases characterized by disturbances in emotion, cognition, motivation, and socialization. They are highly heritable, with genetic risk comprising 20–90% of disease vulnerability depending on the syndrome. As a result of their prevalence, early onset, and persistence, they contribute substantially to the burden of illness worldwide. All psychiatric disorders are broad heterogeneous syndromes that currently lack well-defined neuropathology and bona fide biologic markers. Therefore, diagnoses continue to be made solely from clinical observations using criteria in the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (*DSM-5*), of the American Psychiatric Association **(see Chap. 463)**.

There is increasing agreement that the classification of psychiatric illnesses in the *DSM* does not accurately reflect their underlying biology. Uncertainties in diagnosis complicate efforts to study the genetic basis and attendant neurobiological mechanisms underlying mental illness, though recent technologic advances along with the consolidation of very large patient cohorts have, for multiple disorders, led to significant progress in these realms. In addition, there have been efforts to address the limitations of a categorical nosology directly through the development of an alternative diagnostic scheme, termed Research Domain Criteria (RDoC). This system classifies mental illness on the basis of core behavioral abnormalities shared across several syndromes—such as psychosis (loss of reality) or anhedonia (decreased ability to experience pleasure)—and the associated brain circuitry that controls these behavioral domains. Such classifications may assist in defining the biologic basis of key symptoms. Other factors that have impeded progress in understanding mental illness include the lack of access to pathologic brain tissue except upon death and inherent limitations of animal models for disorders defined largely by behavioral abnormalities (e.g., hallucinations, delusions, guilt, suicidality) that are inaccessible in animals.

Despite these limitations, the past decade has been marked by real progress. Neuroimaging methods are beginning to provide evidence of brain pathology; genome-wide association studies and high-throughput sequencing are reliably identifying genes and genomic loci that confer risk for severe forms of mental illness; and investigations of better-validated animal models, leveraging a host of new methods to study molecular, cellular, and circuit-level processes, are offering new insight into disease pathogenesis. There is also excitement in the utility of neurons, glia, and brain organoids induced in vitro from patient-derived pluripotent stem cells, providing novel ways to study disease pathophysiology and screen for new treatments. There is consequently justified optimism that the field of psychiatry will better integrate behaviorally defined syndromes with an understanding of biological substrates in a way that will drive the development of improved treatments and eventually cures and preventive measures. This chapter describes several examples of recent discoveries in basic neuroscience and genetics that have informed our current understanding of disease mechanisms in psychiatry.

NEUROGENETICS

Because the human brain can only be examined indirectly during life, genome analyses have been extremely important for obtaining molecular clues about the pathogenesis of psychiatric disorders. Moreover, the identification of germline risk alleles and mutations provides potential traction on the question of cause versus effect. In other types of cross-sectional studies, it may be impossible to determine whether a phenotype or biomarker observed in affected humans or model systems reflects an etiologic factor or a compensatory response. In contrast, germline genetic risk is present before the brain develops—at least theoretically allowing for experiments to address temporal sequencing.

A wealth of new information has been made possible by two decades of advances subsequent to the sequencing of the human genome. These have enabled affordable, very large-scale genome-wide association and high-throughput sequencing studies. A striking example of the impact of these developments has been progress in the genetics of autism spectrum disorders (ASDs), a phenotypically heterogeneous neurodevelopmental syndrome characterized by impaired social communication and restricted, repetitive patterns of behavior. ASDs are highly heritable. Concordance rates in monozygotic twins range from 60–90%, a four- to sixfold increase compared to dizygotic twins and siblings. ASDs are also highly genetically heterogeneous and, like many psychiatric conditions, are mainly inherited in a polygenic fashion, conferred by a conspiracy of alleles that are common in the population and carry small individual effects. Indeed, the increments of increased risk for any common ASD risk allele are so modest that studies of tens of thousands of individuals have identified only a handful of associations meeting gold-standard genome-wide statistical thresholds. However, with increasingly large cohorts and an associated increase in statistical power, the number of identified risk loci is destined to continue to grow.

At the same time, the development of next-generation DNA sequencing has identified a substantial minority of ASD patients who carry rare, often spontaneous (de novo), heterozygous, protein-damaging mutations. This latter group has served as a powerful resource for the identification of specific ASD risk genes of very large effect **(Fig. 462-1)**. To date, about 70–250 ASD genes have been identified by high-throughput sequencing analyses based on false discovery rates ranging from <0.001 to <0.1, respectively. In addition, approximately 10% of ASD-affected individuals carry rare, typically de novo, submicroscopic gains or losses of chromosomal material, known as copy number variations (CNVs), that also confer very large risks. All told, these mutations in individual genes or gene-rich genomic regions account for ~20–30% of cases of ASD seen in clinic, although none individually account for >1%. Importantly, as gene discovery in ASD has progressed, considerable overlap, both phenotypically and genetically, has been found among ASD, epilepsy, and well-established intellectual disability syndromes. For example, individuals with *fragile X syndrome* or *tuberous sclerosis* **(Chap. 95)** show elevated rates of ASD, and mutations in the causal genes may be found in patients who present with otherwise idiopathic ASD.

The discovery of very large-effect ASD risk genes that are vulnerable to protein-damaging mutations has provided important opportunities to delve into pathologic mechanisms. From the earliest successes in gene discovery, several common biological themes have emerged. For instance, many of the identified pathogenic rare mutations are in genes that encode proteins involved in synaptic structure or function or in transcriptional and chromatin regulation (Fig. 462-1). More recently, studies of high-confidence ASD risk mutations in model systems have confirmed these predictions and pointed to neurogenesis and neuronal migration as additional shared pathological mechanisms. Moreover, the availability of increasingly comprehensive maps of human brain gene expression has enabled studies of when and where ASD risk genes converge. These transcriptomic studies have repeatedly pointed to glutamatergic neurons in the mid-fetal human cortex (Fig. 462-1) as one of several regions and cell types enriched for ASD genetic vulnerability. Given that many autism risk genes are biologically pleiotropic—that is, they serve multiple different functions—the identification of anatomic and temporal dimensions of risk should help narrow in on pathophysiologic mechanisms and potential therapeutic targets. Moreover, as the number of large-effect ASD risk genes grows and the scale of transcriptomic, epigenomic, and proteomic mapping of human brain development expands, points of pathogenic convergence promise to be identifiable at single-cell resolution.

A deeper understanding of disease pathogenesis and the identification of genetic subtypes is ultimately aimed at developing more effective, rational, and personalized therapies, particularly for those who are most severely affected. In this regard, in humans, increasing attention has turned to nucleic acid targeting to treat severe phenotypes in cases in which a highly penetrant coding mutation is present, for example, with CRISPR-based therapies or the use of antisense oligonucleotides (ASOs). Remarkable success with very early intervention in spinal muscular atrophy using these strategies has piqued interest in their utility in a range of brain-based conditions. Currently, among neurodevelopmental disorders, these approaches are being actively pursued for well-known intellectual disability syndromes that may also manifest core features of or elevated risk for ASD, such as Angelman syndrome and SHANK3 deletion/Phelan-McDermid syndrome. If successful, such efforts would be transformational, and potentially not only for the individuals carrying mutations that may be amenable to nucleic acid–targeting approaches.

The ability to catalog common genetic variants and assay them on array-based platforms and carry out whole exome sequencing has also allowed investigators to leverage large patient cohorts to reliably detect risk loci for schizophrenia and bipolar disorder. In contrast to ASD, where the lion's share of early success resulted from the study of rare, large-effect, de novo mutations, much of gene discovery for these syndromes has resulted from genome-wide association studies of common inherited polymorphisms. To date, several hundred distinct genomic regions, marked by associated single nucleotide polymorphisms, have been identified in schizophrenia, some of which show risk as well for bipolar disorder. Several identified genes are parts of molecular complexes, such as voltage-gated calcium channels (in particular, *CACNA1C* and *CACNB2*) and the postsynaptic density of excitatory synapses. Notably, as the scale of high-throughput sequencing studies has expanded, rare large-effect mutations have also been identified in schizophrenia. To date, 10 high-confidence genes have been identified.

Genes that promote risk for addiction and depression have also begun to emerge from large studies. One susceptibility locus for addiction is the *CHRNA5-A3-B4* nicotinic acetylcholine receptor gene cluster on chromosome 15 associated with nicotine and alcohol addiction. Genome-wide association studies of depression and addiction have required hundreds of thousands of cases and controls to identify the first statistically significant loci using state-of-the-art approaches. For example, a meta-analysis of >1 million individuals was able to identify 110 unique genetic variants associated with problematic alcohol use. Such findings collectively point to the tremendous heterogeneity of these disorders as well as the very small biological effects conferred by any individual common allele.

A recurrent theme in genetic studies of psychiatric disorders is phenotypic pleiotropy, namely, that individual risk genes may be associated with multiple psychiatric and neurodevelopmental syndromes. For example, functionally identical heterozygous deletions of the gene *NRXN1* are associated with ASD, schizophrenia, intellection disability, epilepsy, and other neurodevelopmental phenotypes. Common polymorphisms in *CACNA1C* are associated with both schizophrenia and bipolar disorder. Rare mutations in this same gene may lead to severe neurodevelopmental syndromes and congenital heart disease, including Timothy syndrome, which may include autistic features. Likewise, there is striking overlap among the phenotypes associated with large-effect CNVs, including ASD, schizophrenia, and bipolar disorder, as well as epilepsy and intellectual disability. For example, duplication of chromosome 16p is associated with both schizophrenia and autism, whereas deletions in the DiGeorge's (velocardiofacial) syndrome region are associated with schizophrenia, autism, and bipolar disorder. These findings attest to the complexity of psychiatric disorders, the very large gap between molecular mechanisms and the current categorical diagnostic schemes, and the influence of additional factors that combine to specify the ultimate phenotype. The latter might include polygenic "background," stochastic events, epigenetic effects, and environmental factors. This pleiotropy of consequences for a given genetic mutation in psychiatry is akin to the pleiotropy seen for neurodegenerative disorders as well as for many cancer-causing mutations, where the same mutation can lead to very different disorders across the population.

SIGNAL TRANSDUCTION

Studies of signal transduction disturbances in psychiatric disorders have provided insight into development of new therapeutic agents. For example, lithium is a highly effective drug for bipolar disorder and competes with magnesium to inhibit numerous magnesium-dependent enzymes, including GSK3β and several enzymes involved in phosphoinositide signaling that lead to activation of protein kinase C. These findings have led to discovery programs focused on developing GSK3β or protein kinase C inhibitors as potential novel treatments for mood disorders, although none have demonstrated clinical efficacy to date.

The observations that tricyclic antidepressants (e.g., imipramine) inhibit serotonin and/or norepinephrine reuptake and that monoamine oxidase inhibitors (e.g., tranylcypromine) are effective antidepressants initially led to the view that depression is caused by a deficiency of these monoamines. However, this hypothesis has not

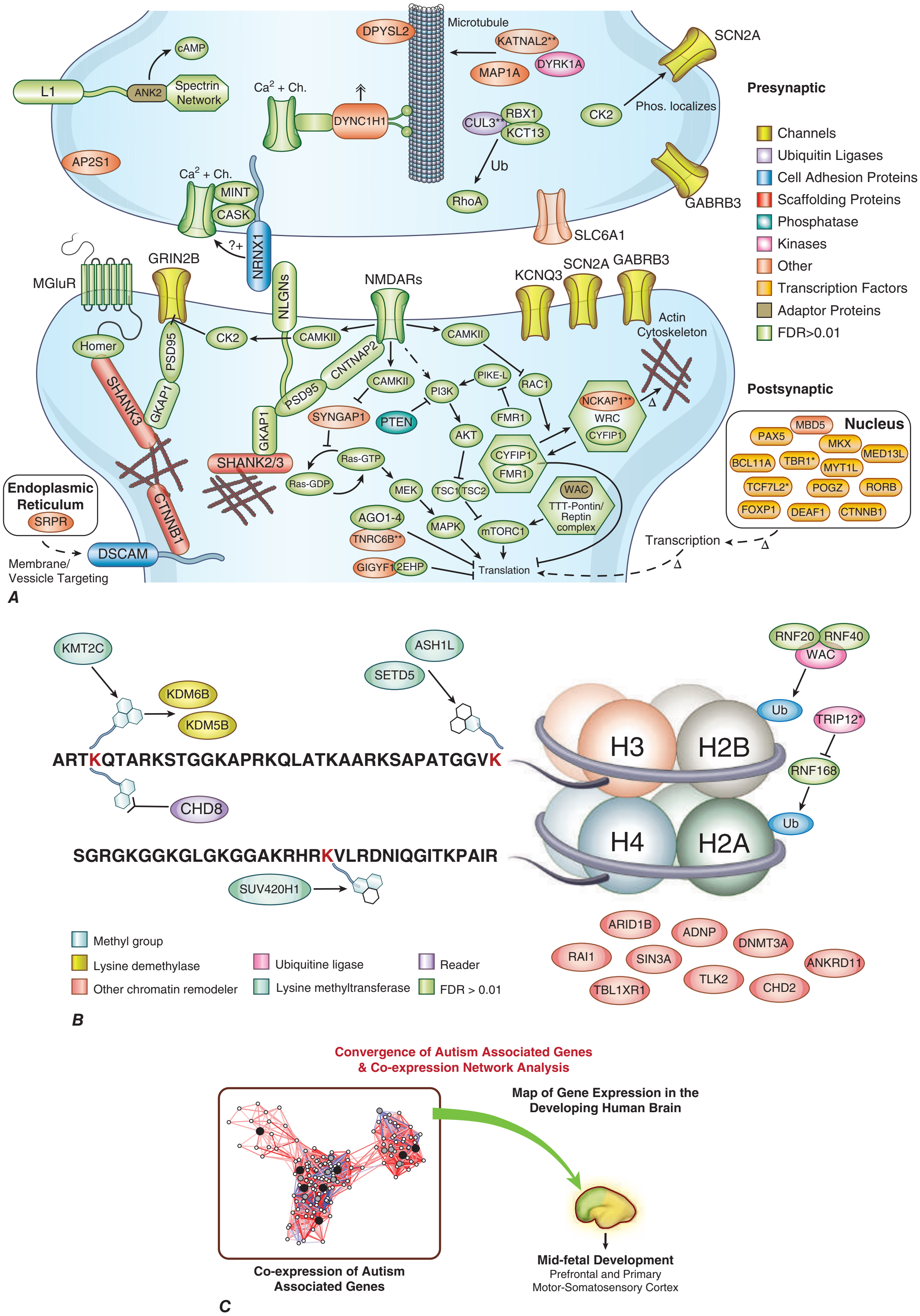
Microtubule
DPYSL2
KATNAL2**
DYRK1A
MAP1A
cAMP
L1
ANK2
Spectrin Network
Ca2 + Ch.
DYNC1H1
CUL3**
RBX1
KCT13
CK2
Phos. localizes
SCN2A
Ub
RhoA
AP2S1
Ca2 + Ch.
MINT
CASK
NRNX1
?+
GABRB3
SLC6A1
Presynaptic
Channels
Ubiquitin Ligases
Cell Adhesion Proteins
Scaffolding Proteins
Phosphatase
Kinases
Other
Transcription Factors
Adaptor Proteins
FDR>0.01
MGluR
GRIN2B
NLGNs
NMDARs
KCNQ3
SCN2A
GABRB3
Actin Cytoskeleton
Homer
PSD95
CK2
CAMKII
CAMKII
CNTNAP2
SHANK3
GKAP1
PSD95
CAMKII
PI3K
PIKE-L
RAC1
NCKAP1**
WRC
CYFIP1
SYNGAP1
PTEN
FMR1
AKT
GKAP1
SHANK2/3
Ras-GTP
Ras-GDP
CYFIP1
FMR1
MEK
TSC1
TSC2
WAC
TTT-Pontin/Reptin complex
AGO1-4
MAPK
mTORC1
TNRC6B**
GIGYF1
2EHP
Translation
Transcription
Endoplasmic Reticulum
SRPR
CTNNB1
DSCAM
Membrane/ Vessicle Targeting
Postsynaptic
Nucleus
MBD5
PAX5
MKX
BCL11A
TBR1*
MYT1L
MED13L
TCF7L2*
POGZ
RORB
FOXP1
DEAF1
CTNNB1
A
KMT2C
KDM6B
KDM5B
ASH1L
SETD5
ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVK
CHD8
SGRGKGGKGLGKGGAKRHRKVLRDNIQGGITKPAIR
SUV420H1
RNF20
RNF40
WAC
Ub
TRIP12*
RNF168
Ub
H3
H2B
H4
H2A
Methyl group
Lysine demethylase
Other chromatin remodeler
Ubiquitine ligase
Lysine methyltransferase
Reader
FDR > 0.01
ARID1B
ADNP
DNMT3A
RAI1
SIN3A
ANKRD11
TLK2
CHD2
TBL1XR1
B
Convergence of Autism Associated Genes & Co-expression Network Analysis
Map of Gene Expression in the Developing Human Brain
Co-expression of Autism Associated Genes
Mid-fetal Development
Prefrontal and Primary Motor-Somatosensory Cortex
C

FIGURE 462-1 Functional characteristics and developmental convergence of autism spectrum disorder (ASD) associated genes. A representative selection of genes associated with ASD based on recurrent rare coding mutations is shown in ***A*** and ***B***. Those genes encoding proteins with a false discovery rate (FDR) <0.01 in Sanders et al, *Neuron* 2015, and Satterstrom et al, *Cell* 2020, are highlighted with respect to their putative functions. Genes meeting the highest confidence criteria in Sanders et al 2015 and showing either an FDR >0.01 or an FDR >0.3 in Satterstrom are noted (* and **, respectively). Additional interacting and functionally related molecules that do not meet the above criteria are shown in *green*. *FMR1*, *TSC1*, and *TSC2* are syndromic ASD genes included in the figure (***A***). Multiple gene ontology analyses of ASD genes have highlighted both pre- and postsynaptic molecules (***A***) and chromatin modifiers (***B***) as points of enrichment. In ***C***, an alternative strategy for grouping ASD risk genes is highlighted (Willsey et al, *Cell* 2013), based on their spatiotemporal expression patterns as opposed to putative functions. One analytic strategy, illustrated in ***C***, leveraged only high-confidence ASD genes and examined their developmental expression patterns using the BrainSpan data set. Convergence for ASD risk was identified in deep layer (V and VI) excitatory neurons in mid-fetal human cortex. Multiple analyses have similarly found glutamatergic neurons in mid-fetal prefrontal cortex as one point of convergence, with somewhat less agreement on layer specificity and potential additional spatiotemporal points of convergence.

been substantiated. A cardinal feature of these drugs is that long-term (weeks to months) administration is needed for their antidepressant effects. This means that their short-term actions, namely promotion of serotonin or norepinephrine function, are not per se antidepressant but rather induce a cascade of adaptations in the brain that underlie their slowly developing clinical effects. The nature of these therapeutic drug-induced adaptations has not been identified with certainty. A subset of depressed patients display upregulation of the hypothalamic-pituitary-adrenal (HPA) axis characterized by increased secretion of corticotropin-releasing factor (CRF) and glucocorticoids. One hypothesis posits that in these patients excessive glucocorticoids cause atrophy of hippocampal neurons, which is associated with reduced hippocampal volumes seen clinically. Chronic antidepressant administration might reverse this atrophy by increasing brain-derived neurotrophic factor (BDNF) or a host of other neurotrophic factors in the hippocampus. A role for stress-induced decreases in the generation of newly born hippocampal granule cell neurons, and its reversal by antidepressants through BDNF or other growth factors, has also been suggested.

A major advance in recent years has been the identification of several rapidly acting antidepressants with non–monoamine-based mechanisms of action. The best established is ketamine, a noncompetitive antagonist of *N*-methyl-D-aspartate (NMDA) glutamate receptors among other actions, which exerts rapid (hours) and robust antidepressant effects in severely depressed patients who have not responded to other treatments. Ketamine, which at higher doses is psychotomimetic and anesthetic, exerts these antidepressant effects at lower doses with minimal side effects. However, the response to ketamine is transient, which has led to several approaches to maintain treatment response, such as repeated ketamine delivery. The mechanism underlying ketamine's antidepressant action is not known, and its action as an NMDA receptor antagonist has recently been called into question. Nevertheless, ketamine's striking clinical efficacy has stimulated animal research on the role of glutamate neurotransmission and synaptic plasticity in key limbic regions. Recent evidence supports a role for TORC1 or BDNF activation, as blockade of either blocks the antidepressant-like effects of ketamine in animal models. Mechanisms by which ketamine activates these signaling cascades are currently an active area of investigation.

Another area of great interest is the potential clinical utility of psychedelic drugs and 3,4-methylenedioxy-methamphetamine (MDMA or ecstasy). Psychedelic drugs are thought to act as partial agonists of serotonin 5-HT$_{2A}$ receptors, whereas MDMA promotes serotonin release from nerve terminals. Both are being studied for treatment of depression and posttraumatic stress disorder, although much additional clinical research is needed to establish their efficacy, safety, and mechanism of action.

A major goal in the field of substance use disorders has been to identify neuroadaptive mechanisms that lead from recreational use to addiction. Such research has determined that repeated intake of abused drugs induces specific changes in cellular signal transduction, leading to changes in synaptic strength (long-term potentiation or depression) and neuronal structure (altered dendritic branching or cell soma size) within the brain's reward circuitry. These drug-induced modifications are mediated in part by changes in gene expression, achieved by regulation of transcription factors (e.g., CREB [cAMP response element-binding protein] and ΔFOSB [a FOS family protein]) and their target genes. Such alterations in gene expression are associated with lasting alterations in epigenetic modifications, including histone acetylation and methylation and DNA methylation. These adaptations provide opportunities for developing treatments targeted to drug-addicted individuals. The fact that the spectrum of these adaptations differs in part depending on the particular addictive substance used raises hope that treatments could be developed that are specific for different classes of addictive drugs and less likely to disturb basic mechanisms that govern normal motivation and reward.

Increasingly, causal relationships are being established between individual molecular and cellular adaptations and specific behavioral abnormalities that characterize the addicted state. For example, acute activation of μ-opioid receptors by morphine or other opioids activates G$_{i/o}$ proteins, leading to inhibition of adenylyl cyclase (AC), resulting in reduced cyclic AMP (cAMP) production, protein kinase A (PKA) activation, and activation of the transcription factor CREB. Repeated administration of these drugs **(Fig. 462-2)** evokes a homeostatic response involving upregulation of ACs and PKA and increased activation of CREB. Such upregulation of cAMP-CREB signaling has been identified in the locus coeruleus (LC), periaqueductal gray, ventral tegmental area (VTA), nucleus accumbens (NAc), and several other central nervous system (CNS) regions and contributes to opioid craving and signs of opioid withdrawal. The fact that endogenous opioid peptides do not produce tolerance and dependence, while morphine and related drugs do, may relate to the observation that, unlike endogenous opioids, morphine and like drugs are weak inducers of μ-opioid receptor desensitization and endocytosis. Therefore, these drugs cause prolonged receptor activation and inhibition of ACs, which provides a powerful stimulus for the upregulation of cAMP-CREB signaling that characterizes the opioid-dependent state.

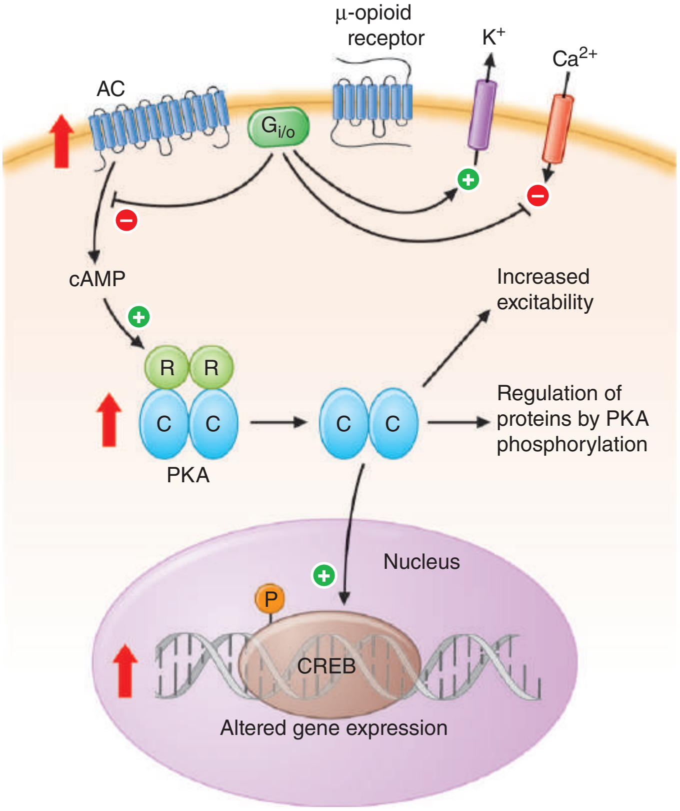

FIGURE 462-2 Opioid action in the locus coeruleus (LC). Binding of opioid agonists to μ-opioid receptors on LC neurons catalyzes nucleotide exchange on G$_i$ and G$_o$ proteins, leading to inhibition of adenylyl cyclase (AC), neuronal hyperpolarization via activation of K$^+$ channels, and perhaps inhibition of Ca^{2+} channels. Inhibition of AC reduces protein kinase A (PKA) activity and phosphorylation of several PKA substrate proteins, thereby altering their function. For example, opioids reduce phosphorylation of the cAMP response element-binding protein (CREB), which initiates longer term changes in neuronal function. Chronic administration of opioids increases levels of AC isoforms, PKA catalytic (C) and regulatory (R) subunits, and the phosphorylation of several proteins, including CREB (indicated by *red arrows*). These changes contribute to the altered phenotype of the drug-addicted state. For example, the excitability of LC neurons is increased by enhanced cAMP signaling. Activation of CREB causes upregulation of AC isoforms and tyrosine hydroxylase, the rate-limiting enzyme in catecholamine biosynthesis.

SYSTEMS NEUROSCIENCE

The study of interconnected brain circuits that drive behavior has been greatly advanced through newer methods in brain imaging that have documented abnormalities in neural function and connectivity in psychiatric disorders. Electroceutical devices, which use electrical or magnetic stimulation to control neuronal activity, have had some success in depression, obsessive-compulsive disorder, pain, and addiction. The past decade has also witnessed the development of revolutionary new

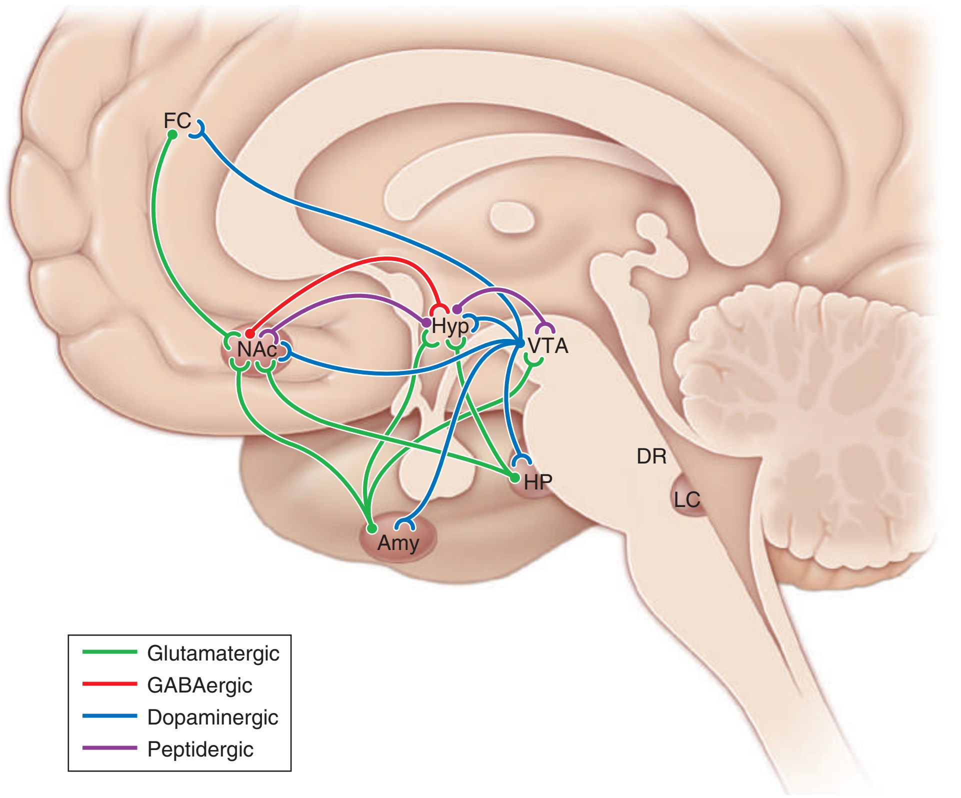

FIGURE 462-3 Neural circuitry of depression and addiction. The figure shows a simplified summary of a series of limbic circuits in the brain that regulate mood and motivation and are implicated in depression and addiction. Shown in the figure are the hippocampus (HP) and amygdala (Amy) in the temporal lobe, regions of prefrontal cortex, nucleus accumbens (NAc), and hypothalamus (Hyp). Only a subset of the known interconnections among these brain regions is shown. Also shown is the innervation of several of these brain regions by monoaminergic neurons. The ventral tegmental area (VTA) provides dopaminergic input to each of the limbic structures. Norepinephrine (from the locus coeruleus [LC]) and serotonin (from the dorsal raphe [DR] and other raphe nuclei) innervate all of the regions shown. In addition, there are strong connections between the hypothalamus and the VTA-NAc pathway. Important peptidergic projections from the hypothalamus include those from the arcuate nucleus that release β-endorphin and melanocortin and from the lateral hypothalamus that release orexin.

techniques—optogenetics, designer receptors, and ligands—that provide unprecedented temporal and spatial control of neural circuits. The development of genetically encoded calcium detectors and of high-density electrode arrays has allowed in vivo monitoring of thousands of neurons in multiple brain regions simultaneously. Advances in histology and microscopy now permit three-dimensional imaging of specific proteins in the intact brain, while advances in endoscopic microscopy allow imaging of hundreds of neurons within deep brain structures in awake, freely moving animals. Together with recent advances in machine learning and artificial intelligence for analysis of large complex datasets, these new methods are revolutionizing our ability to understand the circuit basis of brain function.

Positron emission tomography (PET), diffusion tensor imaging (DTI), and functional magnetic resonance imaging (fMRI) have identified neural circuits that contribute to psychiatric disorders, for example, defining the neural circuitry of mood within the brain's limbic system **(Fig. 462-3)**. Integral to this system are the NAc (important also for brain reward—see below), amygdala, hippocampus, and regions of prefrontal cortex. Recent optogenetic research in animals, where the activity of specific types of neurons in defined circuits can be controlled with light, has confirmed the importance of this limbic circuitry in controlling depression-related behavioral abnormalities. Given that many symptoms of depression (so-called neurovegetative symptoms) involve physiologic functions, a key role for the hypothalamus is presumed as well. A subset of depressed individuals shows a small reduction in hippocampal size, as noted above. In addition, brain imaging investigations have revealed increased activation of the amygdala by negative stimuli and reduced activation of the NAc by rewarding stimuli. There is also evidence for altered activity in prefrontal cortex, such as hyperactivity of subgenual area 25 in anterior cingulate cortex. Such findings have led to trials of deep brain stimulation (DBS) of either the NAc or subgenual area 25 **(see Fig. 32-1)**, which appears to be therapeutic in some severely depressed individuals.

In schizophrenia, structural and functional imaging studies have confirmed earlier pathologic studies that show enlargement of the ventricular system and reduction of cortical and subcortical gray matter in frontal and temporal lobes and in the limbic system. Functional imaging studies show reduced metabolic (presumably neural) activity in the dorsolateral prefrontal cortex at rest and when performing tests of executive function, including working memory. There is also evidence for impaired structural and task-related functional connectivity, mainly in frontal and temporal lobes. The reduction in cortical thickness seen in schizophrenia is associated with increased cell packing density and reduced neuropil (defined as axons, dendrites, and glial cell processes) without an apparent change in neuronal cell number. Specific classes of interneurons in prefrontal cortex consistently show reduced expression of the gene encoding the enzyme glutamic acid decarboxylase 1 (*GAD1*), which synthesizes γ-aminobutyric acid (GABA), the principal inhibitory neurotransmitter in the brain. Recently, results from well-powered genome-wide association studies point to synaptic pruning, including the involvement of microglia, as a potential contributing mechanism. In the region of the genome most strongly associated with schizophrenia risk, variations in the relative expression of two isotypes of complement component 4, C4A and C4B, have been found to account for a significant proportion of this genetic signal. Studies of loss of C4 in mice show deficient synaptic pruning, leading to the hypothesis that increased expression of C4A in humans may result in excessive synaptic pruning. Such results point to the potential for a gene-driven understanding of pathophysiology; however, the findings also leave some important questions unanswered. The strongest effect haplotype in humans still only accounts for a very small increase in risk, with an odds ratio of <1.3. In contrast, having a sibling with schizophrenia increases risk approximately tenfold. In short, whether this allele reflects a driving pathophysiologic mechanism remains to be determined. Moreover, humans have diverged at the C4 locus compared with rodents such that only a single C4 isotype is present in the mouse, preventing any analysis of the putative effects of changing the ratio of C4A to C4B—the phenomenon associated with disease risk in humans. Nonetheless, all the aforementioned findings support the notion that schizophrenia is a developmental neurodegenerative disorder with some evidence pointing to loss of cortical interneurons in frontal and temporal lobes.

Work in rodent and nonhuman primate models of addiction has established the brain's reward regions as key neural substrates for the acute actions of drugs of abuse and for addiction induced in vulnerable individuals by repeated drug administration (Fig. 462-3). Midbrain dopamine neurons in the VTA function normally as rheostats of reward: they are activated by natural rewards (food, sex, social interaction) or even by the expectation of such rewards, and many are suppressed by the absence of an expected reward or by aversive

TABLE 462-1 Initial Actions of Drugs of Abuse

DRUG	NEUROTRANSMITTER AFFECTED	DRUG TARGET (ACTION)
Opioids	Endorphins, enkephalins	μ- and δ-opioid receptors (agonist)
Psychostimulants (cocaine, amphetamine, methamphetamine)	Dopamine	Dopamine transporter (antagonist—cocaine; reverse transport—amphetamine, methamphetamine)
Nicotine	Acetylcholine	Nicotinic cholinergic receptors (agonist)
Ethanol	GABA	$GABA_A$ receptors (positive allosteric modulator)
	Glutamate	NMDA glutamate receptors (antagonist)
	Acetylcholine	Nicotinic cholinergic receptors (allosteric modulator)
	Serotonin	$5\text{-}HT_3$ receptor (positive allosteric modulator)
	Others	Calcium-activated K^+ channel (activator)
Marijuana	Endocannabinoids (anandamide, 2-arachidonoylglycerol)	CB_1 receptor (agonist)
Phencyclidine	Glutamate	NMDA glutamate receptor (antagonist)

Abbreviations: GABA, γ-aminobutyric acid; NMDA, *N*-methyl-D-aspartate.

stimuli. These neurons thereby transmit crucial survival signals to the rest of the limbic brain to promote reward-related behavior, including motor responses to seek and obtain the rewards (NAc), memories of reward-related cues and contexts (amygdala, hippocampus), and executive control of obtaining rewards (prefrontal cortex).

Drugs of abuse alter neurotransmission through initial actions at different classes of ion channels, neurotransmitter receptors, or neurotransmitter transporters **(Table 462-1)**. Studies in animal models have demonstrated that although the initial targets differ, the actions of these drugs converge on the brain's reward circuitry by promoting dopamine neurotransmission in the NAc and other limbic targets of the VTA. In addition, some drugs promote activation of opioid and cannabinoid receptors, which modulate this reward circuitry. By these mechanisms, drugs of abuse produce powerful rewarding signals, which, after repeated drug administration, corrupt a vulnerable brain's reward circuitry in ways that promote addiction. Three major pathologic adaptations have been described. First, drugs produce tolerance in reward circuits and increased activity in limbic stress circuits, which promote escalating drug intake and a negative emotional state during drug withdrawal that promotes relapse. Second, sensitization to the rewarding effects of the drugs and associated cues is seen during prolonged abstinence and also triggers relapse. Third, executive function is impaired in such a way as to increase impulsivity and compulsivity, both of which promote relapse.

Imaging studies in humans confirm that addictive drugs, as well as craving for them, activate the brain's reward circuitry. In addition, patients who abuse alcohol or psychostimulants show reduced gray matter in the prefrontal cortex as well as reduced activity in anterior cingulate and orbitofrontal cortex during tasks of attention and inhibitory control. It is thought that damage to these cortical areas contributes to addiction by impairing decision-making and increasing impulsivity.

NEUROINFLAMMATION

There is increasing evidence for the involvement of nonneuronal cell types and inflammatory mechanisms in a wide range of psychiatric syndromes. For example, a subset of depressed patients displays elevated blood levels of interleukin 6 (IL-6), tumor necrosis factor α (TNF-α), and other proinflammatory cytokines. Moreover, rodents exposed to chronic stress exhibit similar increases in peripheral levels of these cytokines, and peripheral or central delivery of those cytokines to normal rodents increases their susceptibility to chronic stress. These findings have led to the novel idea of using peripheral cytokines as biomarkers of a subtype of depression and the potential utility of developing new antidepressants that oppose the actions of specific cytokines.

Recent evidence has also linked proinflammatory signaling in the brain to addiction, particularly to alcohol. Alcohol use disorder is associated with impaired innate immunity, increases in circulating proinflammatory cytokines, and increases in brain expression of several immune-related genes. Many of these genes are expressed by astrocytes and microglia and by neurons under certain pathologic conditions, where they play important roles in modifying neuronal function and plasticity. For example, cytokine monocyte chemotactic protein-1 (MCP-1) modulates the release of certain neurotransmitters and, when administered into the VTA, increases neuronal excitability, promotes dopamine release, and increases locomotor activity. Gene expression studies of alcohol drinking in mice have identified a network of regulated neuroimmune proteins in brain, and a role in regulation of alcohol consumption has been validated for several, including chemokines MCP-1 and chemokine (C-C motif) ligand 3 (CCL3), beta-2 microglobulin, CD14, IL-1 receptor antagonist, toll-like receptors 3 (TLR3) and 7 (TLR7), and cathepsins S and F. This work has led to discovery of anti-inflammatory medications that reduce alcohol intake in animals, such as antagonists of phosphodiesterase 4, which regulates cAMP availability, or agonists of peroxisome proliferator-activated receptors (PPARs), which are transcription factors that repress key inflammatory signaling molecules such as nuclear factor-κB (NF-κB) and nuclear factor of activated T cells (NFAT). A major focus of current research is to define the sites and mechanisms by which proinflammatory cytokines impair brain function to elicit a depressive episode or promote drug abuse, including a role for astrocytes and microglia.

CONCLUSIONS

This brief narrative illustrates the substantial progress that is being made in understanding the genetic and neurobiological basis of mental illness. It is anticipated that biologic measures will be used increasingly to more accurately diagnose and subtype psychiatric disorders and that targeted therapeutics will become available for these complex conditions.

FURTHER READING

Fu MJ et al: Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. Nat Genet 54:1320, 2022.

Gandal MJ et al: The road to precision psychiatry: Translating genetics into disease mechanisms. Nat Neurosci 19:1397, 2016.

Howes OD et al: Schizophrenia: from neurochemistry to circuits, symptoms and treatments. Nat Rev Neurol 20:22, 2024.

Koob GF, Volkow ND: Neurobiology of addiction: A neurocircuitry analysis. Lancet Psychiatry 3:760, 2016.

McClellan JM et al: An evolutionary perspective on complex neuropsychiatric disease. Neuron 112:7, 2024.

Rajasethupathy P et al: Targeting neural circuits. Cell 165:524, 2016.

Ron D, Barak S: Molecular mechanisms underlying alcohol-drinking behaviours. Nat Rev Neurosci 17:576, 2016.

Satterstrom FK et al: Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell 180:568, 2020.

Tian R et al: Whole-exome sequencing in UK Biobank reveals rare genetic architecture for depression. Nat Commun 15:1755, 2024.

Wohleb ES et al: Integrating neuroimmune systems in the neurobiology of depression. Nat Rev Neurosci 17:497, 2016.

Zhou H et al: Multi-ancestry study of the genetics of problematic alcohol use in over 1 million individuals. Nat Med 29:231, 2023.

463

Psychiatric Disorders

Victor I. Reus

Psychiatric disorders are common in medical practice and may present either as a primary disorder or as a comorbid condition. The prevalence of mental or substance use disorders in the United States is ~30%, but only one-third of affected individuals are currently receiving treatment. Global burden of disease statistics indicates that 4 of the 10 most important causes of morbidity and attendant health care costs worldwide are psychiatric in origin.

Changes in health care delivery underscore the need for primary care physicians to assume responsibility for the initial diagnosis and treatment of the most common mental disorders. Prompt diagnosis is essential to ensure that patients have access to appropriate medical services and to maximize the clinical outcome. Validated patient-based questionnaires have been developed that systematically probe for signs and symptoms associated with the most prevalent psychiatric diagnoses and guide the clinician into targeted assessment. The Primary Care Evaluation of Mental Disorders (PRIME-MD; and a self-report form, the Patient Health Questionnaire) and the Symptom-Driven Diagnostic System for Primary Care (SDDS-PC) are inventories that require only 10 min to complete and link patient responses to the formal diagnostic criteria of anxiety, mood, somatoform, and eating disorders and to alcohol abuse or dependence. A variety of smart phone apps for assessment and monitoring of psychiatric conditions and for psychological and pharmacologic treatment interventions are also available.

A physician who refers patients to a psychiatrist should know not only when doing so is appropriate but also how to refer because societal misconceptions and the stigma of mental illness impede the process. Primary care physicians should base referrals to a psychiatrist on the presence of signs and symptoms of a mental disorder and not simply on the absence of a physical explanation for a patient's complaint. The physician should discuss with the patient the reasons for requesting the referral or consultation and provide reassurance that they will continue to provide medical care and work collaboratively with the mental health professional. Consultation with a psychiatrist or transfer of care is appropriate when physicians encounter evidence of psychotic symptoms, mania, severe depression, or anxiety; symptoms of posttraumatic stress disorder (PTSD); suicidal or homicidal preoccupation; or a failure to respond to first-order treatment. This chapter reviews the clinical assessment and treatment of some of the most common mental disorders presenting in primary care and is based on the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision (DSM-5-TR), the framework for categorizing psychiatric illness used in the United States. **Eating disorders are discussed later in this chapter, and the biology of psychiatric and addictive disorders is discussed in Chap. 462.**

GLOBAL CONSIDERATIONS

The DSM-5-TR and the tenth revision of the International Classification of Diseases (ICD-10-CM), which is used more commonly worldwide, have taken somewhat differing approaches to the diagnosis of mental illness, but considerable effort has been expended to provide an operational translation between the two nosologies. Both systems are in essence purely descriptive and emphasize clinical pragmatism, in distinction to the Research Domain Criteria (RDOC) proposed by the National Institute of Mental Health, which aspires to provide a causal framework for classification of behavioral disturbance. More recently, an alternative approach, based on a hierarchy of dimensions generated by factor analytic techniques, HiTOP (Hierarchical Taxonomy of Psychopathology) has been put forward as a possible improvement over DSM-5-TR for clinical usage, although it too lacks validity and is prone to diagnostic error. None of these diagnostic systems has as yet achieved adequate validation, and large-scale genetic investigations have revealed that differing psychiatric disorders overlap in their genetic risk variants and phenotypic symptoms. The Global Burden of Disease Study (2019), using available epidemiologic data, nevertheless has reinforced the conclusion that, regardless of nosologic differences, mental and substance abuse disorders are the major cause of life-years lost to disability among all medical illnesses, affecting >300 million individuals worldwide. There is general agreement that high-income countries will need to build capacity in professional training in low- and middle-income countries in order to provide an adequate balanced care model for the delivery of evidence-based therapies for mental disorders. Recent surveys that indicate a dramatic increase in mental disorder prevalence in rapidly developing countries, such as China, may reflect both an increased recognition of the issue and also the consequence of social turmoil, stigma, and historically inadequate resources. A salient example of the ways in which societal disruption and isolation may contribute to exacerbating already unmet mental health needs can be seen in the COVID-19 pandemic, which resulted in an increased incidence of diagnosed psychiatric disorders in both affected and unaffected individuals, as well as caregivers. The need for improved prevention strategies and for more definitive and effective interventional treatments remains a global concern.

ANXIETY DISORDERS

Anxiety disorders, the most prevalent psychiatric illnesses in the general community, are present in 15–20% of medical clinic patients. Anxiety, defined as a subjective sense of unease, dread, or foreboding, can indicate a primary psychiatric condition or can be a component of, or reaction to, a primary medical disease. The primary anxiety disorders are classified according to their duration and course and the existence and nature of precipitants.

When evaluating the anxious patient, the clinician must first determine whether the anxiety antedates or postdates a medical illness or is due to a medication side effect. Approximately one-third of patients presenting with anxiety have a medical etiology for their psychiatric symptoms, but an anxiety disorder can also present with somatic symptoms in the absence of a diagnosable medical condition.

PANIC DISORDER

Clinical Manifestations Panic disorder is defined by the presence of recurrent and unpredictable panic attacks, which are distinct episodes of intense fear and discomfort associated with a variety of physical symptoms, including palpitations, sweating, trembling, shortness of breath, chest pain, dizziness, and a fear of impending doom or death. Paresthesias, gastrointestinal distress, and feelings of unreality are also common. Diagnostic criteria require at least 1 month of concern or worry about the attacks or a change in behavior related to them. The lifetime prevalence of panic disorder is 2–3%. Panic attacks have a sudden onset, developing within 10 min and usually resolving over the course of an hour, and can occur in an unexpected fashion, such as when waking from sleep. The frequency and severity of panic attacks vary, ranging from once a week to clusters of attacks separated by months of well-being. The first attack is usually outside the home, and onset is typically in late adolescence to early adulthood. In some individuals, anticipatory anxiety develops over time and results in a generalized fear and a progressive avoidance of places or situations in which a panic attack might recur. *Agoraphobia*, which occurs commonly in patients with panic disorder, is an acquired irrational fear of being in places where one might feel trapped or unable to escape. It may, however, be diagnosed even if panic disorder is not present. Typically, it leads the patient into a progressive restriction in lifestyle and, in a literal sense, in geography. Frequently, patients are embarrassed that they are housebound and dependent on the company of others to go out into the world and do not volunteer this information; thus, physicians will fail to recognize the syndrome if direct questioning is not pursued.

Differential Diagnosis A diagnosis of panic disorder is made after a medical etiology for the panic attacks has been ruled out. A variety of cardiovascular, respiratory, endocrine, and neurologic conditions can present with anxiety as the chief complaint. Patients with true panic disorder will often focus on one specific feature to the exclusion of others. For example, 20% of patients who present with syncope as a primary medical complaint have a primary diagnosis of a mood, anxiety, or substance abuse disorder, the most common being panic disorder. The differential diagnosis of panic disorder is complicated by a high rate of comorbidity with other psychiatric conditions, especially alcohol and benzodiazepine abuse, which patients initially use in an attempt at self-medication. Some 75% of panic disorder patients will also satisfy criteria for major depression at some point in their illness.

When the history is nonspecific, physical examination and focused laboratory testing must be used to rule out anxiety states resulting from medical disorders such as pheochromocytoma, thyrotoxicosis, or hypoglycemia. Electrocardiogram (ECG) and echocardiogram may detect some cardiovascular conditions associated with panic, such as paroxysmal atrial tachycardia and mitral valve prolapse. In two studies, panic disorder was the primary diagnosis in 43% of patients with chest pain who had normal coronary angiograms and was present in 9% of all outpatients referred for cardiac evaluation. Panic disorder has also been diagnosed in many patients referred for pulmonary function testing or with symptoms of irritable bowel syndrome.

Etiology and Pathophysiology The etiology of panic disorder is unknown but appears to involve a genetic predisposition, altered autonomic responsivity, and social learning. Panic disorder shows familial aggregation; the disorder is concordant in 30–45% of monozygotic twins, and genome-wide screens have identified suggestive risk loci. Acute panic attacks appear to be associated with increased noradrenergic discharges in the locus coeruleus. Intravenous infusion of sodium lactate evokes an attack in two-thirds of panic disorder patients, as do the α_2-adrenergic antagonist yohimbine, cholecystokinin tetrapeptide (CCK-4), and carbon dioxide inhalation. It is hypothesized that each of these stimuli activates a pathway involving noradrenergic neurons in the locus coeruleus and serotonergic neurons in the dorsal raphe. Resting-state functional magnetic resonance imaging (fMRI) has identified abnormalities in the default mode network involving the medial temporal lobe, with greater activation in the sensorimotor cortex in panic disorder and in amygdala-frontal connectivity in social anxiety disorder.

TREATMENT

Panic Disorder

Patients with panic disorder have a heightened sensitivity to somatic symptoms, which triggers increasing arousal, setting off the panic attack; accordingly, therapeutic intervention involves altering the patient's cognitive interpretation of anxiety-producing experiences, as well as preventing the attack itself.

Achievable goals of treatment are to decrease the frequency of panic attacks and to reduce their intensity. The cornerstone of drug therapy is antidepressant medication **(Tables 463-1 through 463-3)**. Selective serotonin reuptake inhibitors (SSRIs) benefit the majority of panic disorder patients and do not have the adverse effects of tricyclic antidepressants (TCAs). Fluoxetine, paroxetine, sertraline, and the selective serotonin-norepinephrine reuptake inhibitor (SNRI) venlafaxine have received approval from the U.S. Food and Drug Administration (FDA) for this indication. These drugs should be started at one-third to one-half of their usual antidepressant dose (e.g., 5–10 mg fluoxetine, 25–50 mg sertraline, 10 mg paroxetine, venlafaxine 37.5 mg). Monoamine oxidase inhibitors (MAOIs) are also effective and may specifically benefit patients who have comorbid features of atypical depression (i.e., hypersomnia and weight gain). Insomnia, orthostatic hypotension, and the need to maintain a low-tyramine diet (avoidance of cheese and wine) have limited their use, however. Antidepressants typically take 2–6 weeks to become effective, and doses may need to be adjusted based on the clinical response.

Because of anticipatory anxiety and the need for immediate relief of panic symptoms, benzodiazepines are useful early in the course of treatment and sporadically thereafter **(Table 463-4)**. FDA-approved agents include alprazolam and clonazepam. A recent Cochrane review found no difference between antidepressants and benzodiazepines in response rate, with desipramine and alprazolam ranked highest in achieving remission and clonazepam and alprazolam in reducing frequency of panic attacks. In treatment-resistant cases, short-term augmentation with aripiprazole, divalproex sodium, or pindolol has some evidence for efficacy. There also is no clear difference in short-term efficacy between psychological therapies and antidepressant or benzodiazepine treatment, alone or in combination.

Early psychotherapeutic intervention and education aimed at symptom control enhance the effectiveness of drug treatment. Patients can be taught breathing techniques, be educated about physiologic changes that occur with panic, and learn to expose themselves voluntarily to precipitating events in a treatment program spanning 12–15 sessions. Homework assignments and monitored compliance are important components of successful treatment. Once patients have achieved a satisfactory response, drug treatment should be maintained for 1–2 years to prevent relapse. Controlled trials indicate a success rate of 75–85%, although the likelihood of complete remission is somewhat lower.

■ GENERALIZED ANXIETY DISORDER

Clinical Manifestations Patients with generalized anxiety disorder (GAD) have persistent, excessive, and/or unrealistic worry associated with muscle tension, impaired concentration, autonomic arousal, feeling "on edge" or restless, and insomnia **(Table 463-5)**. Onset is usually before age 20 years, and a history of childhood fears and social inhibition may be present. The lifetime prevalence of GAD is 5–6%; the risk is higher in first-degree relatives of patients with the diagnosis. Interestingly, family studies indicate that GAD and panic disorder segregate independently. More than 80% of patients with GAD also suffer from major depression, dysthymia, or social phobia. Comorbid substance abuse is common in these patients, particularly alcohol and/or sedative/hypnotic abuse. Patients with GAD worry excessively over minor matters, with life-disrupting effects; unlike panic disorder, complaints of shortness of breath, palpitations, and tachycardia are relatively rare.

Etiology and Pathophysiology Most anxiogenic and anxiolytic agents act on the γ-aminobutyric acid $(GABA)_A$ receptor/chloride ion channel complex, implicating this neurotransmitter system in pathogenesis. Benzodiazepines are thought to bind two separate $GABA_A$ receptor sites: type I, which has a broad neuroanatomic distribution, and type II, which is concentrated in the hippocampus, striatum, and neocortex. The antianxiety effects of the various benzodiazepines are influenced by their relative binding to alpha 2 and 3 subunits of the $GABA_A$ receptor, and sedation and memory impairment to the alpha 1 subunit. Serotonin (5-hydroxytryptamine [5-HT]) and 3α-reduced neuroactive steroids (allosteric modulators of $GABA_A$) also appear to have a role in anxiety, and buspirone, a partial 5-HT_{1A} receptor agonist, and certain 5-HT_{2A} and 5-HT_{2C} receptor antagonists (e.g., mirtazapine and nefazodone) may have beneficial effects.

TREATMENT

Generalized Anxiety Disorder

A combination of pharmacologic and psychotherapeutic interventions is most effective in GAD, but complete symptomatic relief is rare. A short course of a benzodiazepine is usually indicated, preferably lorazepam, oxazepam, clonazepam, or alprazolam, although only the last two are FDA approved. (The first two of these agents are metabolized via conjugation rather than oxidation and thus do not accumulate if hepatic function is impaired; the latter also has limited active metabolites.) Treatment should be initiated at the lowest dose possible and prescribed on an as-needed basis as symptoms warrant. Benzodiazepines differ in their milligram per kilogram potency, half-life, lipid solubility, metabolic pathways, and presence of active metabolites. Agents that are absorbed rapidly and

TABLE 463-1 Antidepressants

NAME	USUAL DAILY DOSE (mg)	SIDE EFFECTS	COMMENTS
SSRIs			
Fluoxetine (Prozac)	10–80	Headache; nausea and other GI effects; jitteriness; insomnia; sexual dysfunction; can affect plasma levels of other medicines (except sertraline); akathisia rare	Once-daily dosing, usually in the morning; fluoxetine has very long half-life; must not be combined with MAOIs
Sertraline (Zoloft)	50–200		
Paroxetine (Paxil)	20–60		
Fluvoxamine (Luvox)	100–300		
Citalopram (Celexa)	20–60		
Escitalopram (Lexapro)	10–30		
TCAs and Tetracyclics			
Amitriptyline (Elavil)	150–300	Anticholinergic (dry mouth, tachycardia, constipation, urinary retention, blurred vision); sweating; tremor; postural hypotension; cardiac conduction delay; sedation; weight gain	Once-daily dosing, usually qhs; blood levels of most TCAs available; can be lethal in overdose (lethal dose = 2 g); nortriptyline best tolerated, especially by elderly
Nortriptyline (Pamelor)	50–200		
Imipramine (Tofranil)	150–300		
Desipramine (Norpramin)	150–300		
Doxepin (Sinequan)	150–300	Nausea, anxiety, dry mouth	FDA-approved for OCD
Clomipramine (Anafranil)	150–300		
Maprotiline (Ludiomil)	25–150		
Protriptyline (Vivactil)	15–40		
Trimipramine (Surmontil)	75–200	Drowsiness, constipation, dry mouth	tid or qid dosing required
Amoxapine (Asendin)	100–300		Lethality in OD, EPS possible
Mixed Serotonin-Norepinephrine Reuptake Inhibitors (SNRI) and Receptor Blockers			
Venlafaxine (Effexor), XR	75–375	Nausea; dizziness; dry mouth; headaches; increased blood pressure; anxiety and insomnia	bid–tid dosing (extended-release available); lower potential for drug interactions than SSRIs; contraindicated with MAOIs
Desvenlafaxine (Pristiq)	50–400	Nausea, dizziness, insomnia	Primary metabolite of venlafaxine; no increased efficacy with higher dosing
Duloxetine (Cymbalta)	40–60	Nausea, dizziness, headache, insomnia, constipation	May have utility in treatment of neuropathic pain and stress incontinence
Mirtazapine (Remeron)	15–45	Somnolence, weight gain; neutropenia rare	Once-a-day dosing; 5-HT$_3$ antagonist
Vilazodone (Viibryd)	40	Nausea, diarrhea, headache; dosage adjustment if given with CYP3A4 inhibitor/stimulator	Also 5-HT$_{1A}$ receptor partial agonist
Vortioxetine (Trintellix)	5–20	Nausea, diarrhea, sweating, headache; low incidence of sedation or weight gain	No specific p450 effects; 5-HT$_{3A}$ and 5-HT$_7$ receptor antagonist, 5-HT$_{1B}$ partial agonist, and 5-HT$_{1A}$ agonist
Levomilnacipran (Fetzima)	40–120	Nausea, constipation, sweating; rare increase in blood pressure/pulse	Most noradrenergic of SNRIs
Mixed-Action Drugs			
Bupropion (Wellbutrin), CR, XR	250–450	Jitteriness; flushing; seizures in at-risk patients; anorexia; tachycardia; psychosis	tid dosing, but sustained-release also available; fewer sexual side effects than SSRIs or TCAs; may be useful for adult ADD
Trazodone (Desyrel)	200–600	Sedation; dry mouth; ventricular irritability; postural hypotension; priapism rare	Useful in low doses for sleep because of sedating effects with no anticholinergic side effects
Trazodone extended-release (Oleptro)	150–375	Daytime somnolence, dizziness, nausea	
Nefazodone	300–600	Headache, nausea, dizziness	Rare risk of liver failure, priapism
Gepirone extended-release (Exxua)	18.2–72.6	Dizziness, nausea, insomnia	Partial agonist/5-HT$_{1A}$; 5-HT$_{2A}$ antagonist
Esketamine (Spravato)	56–84 1–2 times a week	Sedation, dissociation, respiratory depression	Ketamine isomer for treatment-resistant depression and/or suicidal risk
Zuranolone (Zurzuvae)	50 qhs for 14 days	Somnolence, confusion, dizziness	Oral neuroactive steroid for postpartum depression
Dextromethorphan-bupropion (Auvelity)	45/105 bid	Dizziness, headache, diarrhea	Possible increased speed of response
MAOIs			
Phenelzine (Nardil)	45–90	Insomnia; hypotension; edema; anorgasmia; weight gain; neuropathy; hypertensive crisis; toxic reactions with SSRIs; narcotics	May be more effective in patients with atypical features or treatment-refractory depression
Tranylcypromine (Parnate)	20–50		
Isocarboxazid (Marplan)	20–60		Less weight gain and hypotension than phenelzine
Transdermal selegiline (Emsam)	6–12	Local skin reaction, hypertension	No dietary restrictions with 6-mg dose

Abbreviations: ADD, attention-deficit disorder; EPS, extrapyramidal symptoms; FDA, U.S. Food and Drug Administration; GI, gastrointestinal; MAOIs, monoamine oxidase inhibitors; OCD, obsessive-compulsive disorder; OD, overdose; SSRIs, selective serotonin reuptake inhibitors; TCAs, tricyclic antidepressants.

TABLE 463-2 Management of Antidepressant Side Effects

SYMPTOMS	COMMENTS AND MANAGEMENT STRATEGIES
Gastrointestinal	
Nausea, loss of appetite	Usually short-lived and dose-related; consider temporary dose reduction or administration with food and antacids
Diarrhea	Famotidine, 20–40 mg/d
Constipation	Wait for tolerance; try diet change, stool softener, exercise; avoid laxatives
Sexual dysfunction	Consider dose reduction; drug holiday
Anorgasmia/ impotence; impaired ejaculation	Bethanechol, 10–20 mg, 2 h before activity, or cyproheptadine, 4–8 mg, 2 h before activity, or bupropion, 100 mg bid, or amantadine, 100 mg bid/tid
Orthostasis	Tolerance unlikely; increase fluid intake, use calf exercises/support hose; fludrocortisone, 0.025 mg/d
Anticholinergic	Wait for tolerance
Dry mouth, eyes	Maintain good oral hygiene; use artificial tears, sugar-free gum
Tremor/jitteriness	Antiparkinsonian drugs not effective; use dose reduction/ slow increase; lorazepam, 0.5 mg bid, or propranolol, 10–20 mg bid
Insomnia	Schedule all doses for the morning; trazodone, 50–100 mg qhs
Sedation	Caffeine; schedule all dosing for bedtime; bupropion, 75–100 mg in afternoon
Headache	Evaluate diet, stress, other drugs; try dose reduction; amitriptyline, 50 mg/d
Weight gain	Decrease carbohydrates; exercise; consider fluoxetine
Loss of therapeutic benefit over time	Related to tolerance? Increase dose or drug holiday; add amantadine, 100 mg bid, buspirone, 10 mg tid, or pindolol, 2.5 mg bid

TABLE 463-3 Possible Drug Interactions with Selective Serotonin Reuptake Inhibitors

AGENT	EFFECT
Monoamine oxidase inhibitors	Serotonin syndrome—absolute contraindication
Serotonergic agonists, e.g., tryptophan, fenfluramine, triptans	Potential serotonin syndrome
Drugs that are metabolized by P450 isoenzymes: tricyclics, other SSRIs, antipsychotics, beta blockers, codeine, triazolobenzodiazepines, calcium channel blockers	Delayed metabolism resulting in increased blood levels and potential toxicity
Drugs that are bound tightly to plasma proteins, e.g., warfarin	Increased bleeding secondary to displacement
Drugs that inhibit the metabolism of SSRIs by P450 isoenzymes, e.g., quinidine	Increased SSRI side effects

Abbreviation: SSRIs, selective serotonin reuptake inhibitors.

are lipid soluble, such as diazepam, have a rapid onset of action and a higher abuse potential. Benzodiazepines should generally not be prescribed for >4–6 weeks because of the development of tolerance and the serious risk of abuse and dependence. Withdrawal must be closely monitored as relapses can occur. It is important to warn patients that concomitant use of alcohol or other sedating drugs may exacerbate side effects and impair their ability to function. An optimistic approach that encourages the patient to clarify environmental precipitants, anticipate their reactions, and plan effective response strategies is an essential element of therapy.

Adverse effects of benzodiazepines generally parallel their relative half-lives. Longer-acting agents, such as diazepam, chlordiazepoxide, flurazepam, and clonazepam, tend to accumulate active metabolites, with resultant sedation, impairment of cognition, and poor psychomotor performance. Shorter-acting compounds, such as alprazolam, lorazepam, and oxazepam, can produce daytime anxiety, early-morning insomnia, and, with discontinuation, rebound anxiety and insomnia. Although patients develop tolerance to the sedative effects of benzodiazepines, they are less likely to habituate to the adverse psychomotor effects. Withdrawal from the longer half-life benzodiazepines can be accomplished through gradual, stepwise dose reduction (by 10% every 1–2 weeks) over 6–12 weeks. It is usually more difficult to taper patients off shorter-acting benzodiazepines. Physicians may need to switch the patient to a benzodiazepine with a longer half-life or use an adjunctive medication such as a beta blocker or carbamazepine, before attempting to discontinue the benzodiazepine. Withdrawal reactions vary in severity and duration; they can include depression, anxiety, lethargy, diaphoresis, autonomic arousal, and, rarely, seizures.

Buspirone is a nonbenzodiazepine anxiolytic agent. It is nonsedating, does not produce tolerance or dependence, does not interact

TABLE 463-4 Anxiolytics

NAME	EQUIVALENT PO DOSE (mg)	ONSET OF ACTION	HALF-LIFE (h)	COMMENTS
Benzodiazepines				
Diazepam (Valium)	5	Fast	20–70	Active metabolites; quite sedating
Flurazepam (Dalmane)	15	Fast	30–100	Flurazepam is a prodrug; metabolites are active; quite sedating
Triazolam (Halcion)	0.25	Intermediate	1.5–5	No active metabolites; can induce confusion and delirium, especially in elderly
Lorazepam (Ativan)	1	Intermediate	10–20	No active metabolites; direct hepatic glucuronide conjugation; quite sedating; FDA-approved for anxiety with depression
Alprazolam (Xanax)	0.5	Intermediate	12–15	Active metabolites; not too sedating; FDA-approved for panic disorder and anxiety with depression; tolerance and dependence develop easily; difficult to withdraw
Chlordiazepoxide (Librium)	10	Intermediate	5–30	Active metabolites; moderately sedating
Oxazepam (Serax)	15	Slow	5–15	No active metabolites; direct glucuronide conjugation; not too sedating
Temazepam (Restoril)	15	Slow	9–12	No active metabolites; moderately sedating
Clonazepam (Klonopin)	0.5	Slow	18–50	No active metabolites; moderately sedating; FDA-approved for panic disorder
Clorazepate (Tranxene)	15	Fast	40–200	Low sedation; unreliable absorption
Prazepam (Centrax)	10–60	Fast	29–224	Less sedating than diazepam
Nonbenzodiazepines				
Buspirone (BuSpar)	7.5	2 weeks	2–3	Active metabolites; tid dosing—usual daily dose 10–20 mg tid; nonsedating; no additive effects with alcohol; useful for controlling agitation in demented or brain-injured patients

Abbreviation: FDA, U.S. Food and Drug Administration.

TABLE 463-5 Diagnostic Criteria for Generalized Anxiety Disorder

A. Excessive anxiety and worry (apprehensive expectation), occurring more days than not for at least 6 months, about a number of events or activities (such as work or school performance).

B. The individual finds it difficult to control the worry.

C. The anxiety and worry are associated with three (or more) of the following six symptoms (with at least some symptoms having been present for more days than not for the past 6 months):

Note: Only one item is required in children.

1. Restlessness or feeling keyed up or on edge.
2. Being easily fatigued.
3. Difficulty concentrating or mind going blank.
4. Irritability.
5. Muscle tension.
6. Sleep disturbance (difficulty falling or staying asleep, or restless, unsatisfying sleep).

D. The anxiety, worry, or physical symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

E. The disturbance is not attributable to the physiological effects of a substance (e.g., a drug of abuse, a medication) or another medical condition (e.g., hyperthyroidism).

F. The disturbance is not better explained by another mental disorder (e.g., anxiety or worry about having panic attacks in panic disorder, negative evaluation in social anxiety disorder, contamination or other obsessions in obsessive-compulsive disorder, separation from attachment figures in separation anxiety disorder, reminders of traumatic events in posttraumatic stress disorder, gaining weight in anorexia nervosa, physical complaints in somatic symptom disorder, perceived appearance flaws in body dysmorphic disorder, having a serious illness in illness anxiety disorder, or the content of delusional beliefs in schizophrenia or delusional disorder).

Source: Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*-Text Revision, 5th ed. (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

with benzodiazepine receptors or alcohol, and has no abuse or disinhibition potential. However, it requires several weeks to take effect and requires thrice-daily dosing. Patients who were previously responsive to a benzodiazepine are unlikely to rate buspirone as equally effective, but patients with head injury or dementia who have symptoms of anxiety and/or agitation may do well with this agent. Escitalopram, paroxetine, duloxetine, and venlafaxine are FDA approved for the treatment of GAD, usually at doses that are comparable to their efficacy in major depression, and may be preferable to usage of benzodiazepines in the treatment of chronic anxiety. Benzodiazepines are contraindicated during pregnancy and breast-feeding.

Anticonvulsants with GABAergic properties may also be effective against anxiety. Gabapentin, oxcarbazepine, tiagabine, pregabalin, and divalproex have all shown some degree of benefit in a variety of anxiety-related syndromes in off-label usage.

PHOBIC DISORDERS

Clinical Manifestations The cardinal feature of phobic disorders is a marked and persistent fear of objects or situations, exposure to which results in an immediate anxiety reaction. The patient avoids the phobic stimulus, and this avoidance usually impairs occupational or social functioning. Panic attacks may be triggered by the phobic stimulus or may occur spontaneously. Unlike patients with other anxiety disorders, individuals with phobias usually experience anxiety only in specific situations. Common phobias include fear of closed spaces (claustrophobia), fear of blood, and fear of flying. Social phobia is distinguished by a specific fear of social or performance situations in which the individual is exposed to unfamiliar individuals or to possible examination and evaluation by others. Examples include having to converse at a party, using public restrooms, or meeting strangers. In each case, the affected individual is aware that the experienced fear is excessive and unreasonable given the circumstance. The specific content of a phobia may vary across gender, ethnic, and cultural boundaries.

Phobic disorders are common, affecting ~7–9% of the population. Twice as many females are affected than males. Full criteria for diagnosis are usually satisfied first in early adulthood, but behavioral avoidance of unfamiliar people, situations, or objects dating from early childhood is common.

In one study of female twins, concordance rates for agoraphobia, social phobia, and animal phobia were found to be 23% for monozygotic twins and 15% for dizygotic twins. A twin study of fear conditioning, a model for the acquisition of phobias, demonstrated a heritability of 35–45%. Animal studies of fear conditioning have indicated that processing of the fear stimulus occurs through the lateral nucleus of the amygdala, extending through the central nucleus and projecting to the periaqueductal gray region, lateral hypothalamus, and paraventricular hypothalamus.

TREATMENT

Phobic Disorders

Beta blockers (e.g., propranolol, 20–40 mg orally 2 h before the event) are particularly effective in the treatment of "performance anxiety" (but not general social phobia) and appear to work by blocking the peripheral manifestations of anxiety such as perspiration, tachycardia, palpitations, and tremor. MAOIs alleviate social phobia independently of their antidepressant activity, and paroxetine, sertraline, fluvoxamine CR, and venlafaxine XR have received FDA approval for treatment of social anxiety. Benzodiazepines can be helpful in reducing fearful avoidance, but the chronic nature of phobic disorders limits their usefulness.

Behaviorally focused psychotherapy is an important component of treatment because relapse rates are high when medication is used as the sole treatment. Cognitive-behavioral strategies are based on the finding that distorted perceptions and interpretations of fear-producing stimuli play a major role in perpetuation of phobias. Individual and group therapy sessions teach the patient to identify specific negative thoughts associated with the anxiety-producing situation and help to reduce the patient's fear of loss of control. In desensitization therapy, hierarchies of feared situations are constructed, and the patient is encouraged to pursue and master gradual exposure to the anxiety-producing stimuli.

Patients with social phobia, in particular, have a high rate of comorbid alcohol abuse, as well as of other psychiatric conditions (e.g., eating disorders), necessitating the need for parallel management of each disorder if anxiety reduction is to be achieved.

STRESS DISORDERS

Clinical Manifestations Patients may develop anxiety after exposure to extreme traumatic events such as the threat of personal death or injury or the death of a loved one. The reaction may occur shortly after the trauma (*acute stress disorder*) or be delayed and subject to recurrence (PTSD) **(Table 463-6)**. In both syndromes, individuals experience associated symptoms of detachment and loss of emotional responsivity. The patient may feel depersonalized and unable to recall specific aspects of the trauma, although typically, it is reexperienced through intrusions in thought, dreams, or flashbacks, particularly when cues of the original event are present. Patients often actively avoid stimuli that precipitate recollections of the trauma and demonstrate a resulting increase in vigilance, arousal, and startle response. Patients with stress disorders are at risk for the development of other disorders related to anxiety, mood, and substance abuse (especially alcohol). Between 5 and 10% of Americans will at some time in their life satisfy criteria for PTSD, with women more likely to be affected than men. A validated four-item screen for PTSD (PC-PTSD) is available.

Risk factors for the development of PTSD include a past psychiatric history and personality characteristics of high neuroticism and extroversion. Twin studies show a substantial genetic influence on all symptoms associated with PTSD, with less evidence for an environmental effect.

TABLE 463-6 Diagnostic Criteria for Posttraumatic Stress Disorder

Posttraumatic Stress Disorder in Individuals Older Than 6 Years

Note: The following criteria apply to adults, adolescents, and children older than 6 years.

For children 6 years and younger, see corresponding criteria below.

A. Exposure to actual or threatened death, serious injury, or sexual violence in one (or more) of the following ways:

1. Directly experiencing the traumatic event(s). 2. Witnessing, in person, the event(s) as it occurred to others. 3. Learning that the traumatic event(s) occurred to a close family member or close friend. In cases of actual or threatened death of a family member or friend, the event(s) must have been violent or accidental. 4. Experiencing repeated or extreme exposure to aversive details of the traumatic event(s) (e.g., first responders collecting human remains; police officers repeatedly exposed to details of child abuse).

Note: Criterion A4 does not apply to exposure through electronic media, television, movies, or pictures, unless this exposure is work related.

B. Presence of one (or more) of the following intrusion symptoms associated with the traumatic event(s), beginning after the traumatic event(s) occurred:

1. Recurrent, involuntary, and intrusive distressing memories of the traumatic event(s).

Note: In children older than 6 years, repetitive play may occur in which themes or aspects of the traumatic event(s) are expressed.

2. Recurrent distressing dreams in which the content and/or affect of the dream are related to the traumatic event(s).

Note: In children, there may be frightening dreams without recognizable content.

3. Dissociative reactions (e.g., flashbacks) in which the individual feels or acts as if the traumatic event(s) were recurring. (Such reactions may occur on a continuum, with the most extreme expression being a complete loss of awareness of present surroundings.)

Note: In children, trauma-specific reenactment may occur in play.

4. Intense or prolonged psychological distress at exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event(s). 5. Marked physiological reactions to internal or external cues that symbolize or resemble an aspect of the traumatic event(s).

C. Persistent avoidance of stimuli associated with the traumatic event(s), beginning after the traumatic event(s) occurred, as evidenced by one or both of the following:

1. Avoidance of or efforts to avoid distressing memories, thoughts, or feelings about or closely associated with the traumatic event(s). 2. Avoidance of or efforts to avoid external reminders (people, places, conversations, activities, objects, situations) that arouse distressing memories, thoughts, or feelings about or closely associated with the traumatic event(s).

D. Negative alterations in cognitions and mood associated with the traumatic event(s), beginning or worsening after the traumatic event(s) occurred, as evidenced by two (or more) of the following:

1. Inability to remember an important aspect of the traumatic event(s) (typically due to dissociative amnesia and not to other factors such as head injury, alcohol, or drugs). 2. Persistent and exaggerated negative beliefs or expectations about oneself, others, or the world (e.g., "I am bad," "No one can be trusted," "The world is completely dangerous," "My whole nervous system is permanently ruined"). 3. Persistent, distorted cognitions about the cause or consequences of the traumatic event(s) that lead the individual to blame himself/herself or others. 4. Persistent negative emotional state (e.g., fear, horror, anger, guilt, or shame). 5. Markedly diminished interest or participation in significant activities. 6. Feelings of detachment or estrangement from others. 7. Persistent inability to experience positive emotions (e.g., inability to experience happiness, satisfaction, or loving feelings).

E. Marked alterations in arousal and reactivity associated with the traumatic event(s), beginning or worsening after the traumatic event(s) occurred, as evidenced by two (or more) of the following:

1. Irritable behavior and angry outbursts (with little or no provocation) typically expressed as verbal or physical aggression toward people or objects. 2. Reckless or self-destructive behavior. 3. Hypervigilance. 4. Exaggerated startle response. 5. Problems with concentration. 6. Sleep disturbance (e.g., difficulty falling or staying asleep or restless sleep).

F. Duration of the disturbance (Criteria B, C, D, and E) is more than 1 month.

G. The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

H. The disturbance is not attributable to the physiological effects of a substance (e.g., medication, alcohol) or another medical condition.

Specify whether:

With dissociative symptoms: The individual's symptoms meet the criteria for posttraumatic stress disorder, and in addition, in response to the stressor, the individual experiences persistent or recurrent symptoms of either of the following:

1. 1. **Depersonalization:** Persistent or recurrent experiences of feeling detached from, and as if one were an outside observer of, one's mental processes or body (e.g., feeling as though one were in a dream; feeling a sense of unreality of self or body or of time moving slowly).

   2. **Derealization:** Persistent or recurrent experiences of unreality of surroundings (e.g., the world around the individual is experienced as unreal, dreamlike, distant, or distorted).

2. **Note:** To use this subtype, the dissociative symptoms must not be attributable to the physiological effects of a substance (e.g., blackouts, behavior during alcohol intoxication) or another medical condition (e.g., complex partial seizures).

Specify if:

With delayed expression: If the full diagnostic criteria are not met until at least 6 months after the event (although the onset and expression of some symptoms may be immediate).

Posttraumatic Stress Disorder in Children 6 Years and Younger

A. In children 6 years and younger, exposure to actual or threatened death, serious injury, or sexual violence in one (or more) of the following ways:

1. Directly experiencing the traumatic event(s). 2. Witnessing, in person, the event(s) as it occurred to others, especially primary caregivers. 3. Learning that the traumatic event(s) occurred to a parent or caregiving figure.

B. Presence of one (or more) of the following intrusion symptoms associated with the traumatic event(s), beginning after the traumatic event(s) occurred:

1. Recurrent, involuntary, and intrusive distressing memories of the traumatic event(s).

Note: Spontaneous and intrusive memories may not necessarily appear distressing and may be expressed as play reenactment.

2. Recurrent distressing dreams in which the content and/or affect of the dream are related to the traumatic event(s).

Note: It may not be possible to ascertain that the frightening content is related to the traumatic event.

3. Dissociative reactions (e.g., flashbacks) in which the child feels or acts as if the traumatic event(s) were recurring. (Such reactions may occur on a continuum, with the most extreme expression being a complete loss of awareness of present surroundings.) Such trauma-specific reenactment may occur in play. 4. Intense or prolonged psychological distress at exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event(s). 5. Marked physiological reactions to reminders of the traumatic event(s).

C. One (or more) of the following symptoms, representing either persistent avoidance of stimuli associated with the traumatic event(s) or negative alterations in cognitions and mood associated with the traumatic event(s), must be present, beginning after the event(s) or worsening after the event(s):

Persistent Avoidance of Stimuli

1. Avoidance of or efforts to avoid activities, places, or physical reminders that arouse recollections of the traumatic event(s). 2. Avoidance of or efforts to avoid people, conversations, or interpersonal situations that arouse recollections of the traumatic event(s).

(*Continued*)

TABLE 463-6 Diagnostic Criteria for Posttraumatic Stress Disorder (Continued)

Negative Alterations in Cognitions

3. Substantially increased frequency of negative emotional states (e.g., fear, guilt, sadness, shame, confusion). 4. Markedly diminished interest or participation in significant activities, including constriction of play. 5. Socially withdrawn behavior. 6. Persistent reduction in expression of positive emotions.

D. Alterations in arousal and reactivity associated with the traumatic event(s), beginning or worsening after the traumatic event(s) occurred, as evidenced by two (or more) of the following:

1. Irritable behavior and angry outbursts (with little or no provocation) typically expressed as verbal or physical aggression toward people or objects (including extreme temper tantrums). 2. Hypervigilance. 3. Exaggerated startle response. 4. Problems with concentration. 5. Sleep disturbance (e.g., difficulty falling or staying asleep or restless sleep).

E. The duration of the disturbance is more than 1 month.

F. The disturbance causes clinically significant distress or impairment in relationships with parents, siblings, peers, or other caregivers or with school behavior.

G. The disturbance is not attributable to the physiological effects of a substance (e.g., medication or alcohol) or another medical condition.

Specify whether:

With dissociative symptoms: The individual's symptoms meet the criteria for posttraumatic stress disorder, and the individual experiences persistent or recurrent symptoms of either of the following:

1. Depersonalization: Persistent or recurrent experiences of feeling detached from, and as if one were an outside observer of, one's mental processes or body (e.g., feeling as though one were in a dream; feeling a sense of unreality of self or body or of time moving slowly). 2. Derealization: Persistent or recurrent experiences of unreality of surroundings (e.g., the world around the individual is experienced as unreal, dreamlike, distant, or distorted).

Note: To use this subtype, the dissociative symptoms must not be attributable to the physiological effects of a substance (e.g., blackouts) or another medical condition (e.g., complex partial seizures).

Specify if:

With delayed expression: If the full diagnostic criteria are not met until at least 6 months after the event (although the onset and expression of some symptoms may be immediate).

Source: Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*-Text Revision, 5th ed. (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

Etiology and Pathophysiology It is hypothesized that in PTSD there is excessive release of norepinephrine from the locus coeruleus in response to stress and increased noradrenergic activity at projection sites in the hippocampus and amygdala. These changes theoretically facilitate the encoding of fear-based memories. Greater sympathetic responses to cues associated with the traumatic event occur in PTSD, although pituitary adrenal responses are blunted. In addition to fear learning, changes in threat detection (insula overactivity), executive function, emotional regulation, and contextual learning have been documented. Predictive biomarkers include increased heart rate and serum lactate, decreased coagulation, insulin resistance, and alterations in glycolysis and fatty acid uptake.

TREATMENT

Stress Disorders

Acute stress reactions are usually self-limited, and treatment typically involves the short-term use of benzodiazepines and supportive/expressive psychotherapy. The chronic and recurrent nature of PTSD, however, requires a more complex approach using drug and behavioral treatments. PTSD is highly correlated with peritraumatic dissociative symptoms and the development of an acute stress disorder at the time of the trauma. Attempts to prevent or ameliorate PTSD through usage of agents such as escitalopram, hydrocortisone, and intranasal oxytocin in the acute stress period have proven equivocal. The SSRIs (paroxetine and sertraline are FDA approved for PTSD), venlafaxine, fluoxetine, and topiramate can all reduce anxiety, symptoms of intrusion, and avoidance behaviors. Recently, the psychedelic agent MDMA (3,4-methylenedioxymethamphetamine) demonstrated efficacy as an adjunct to intensive psychotherapeutic intervention, as did stellate ganglion block. Low-dose trazodone and mirtazapine, sedating antidepressants, are frequently used at night to help with insomnia. Benzodiazepines and SSRIs, however, should not be given in the early aftermath of trauma. Psychotherapeutic strategies for PTSD help the patient overcome avoidance behaviors and demoralization and master fear of recurrence of the trauma; therapies that encourage the patient to dismantle avoidance behaviors through stepwise focusing on the experience of the traumatic event, such as trauma-focused cognitive-behavioral and eye movement desensitization and reprocessing (EMDR) therapies and prolonged exposure therapy utilizing augmented or virtual reality are the most effective. Debriefing after the traumatic event does not prevent PTSD and may exacerbate symptoms.

OBSESSIVE-COMPULSIVE DISORDER

Clinical Manifestations Obsessive-compulsive disorder (OCD) is characterized by obsessive thoughts and compulsive behaviors that impair everyday functioning. Fears of contamination and germs are common, as are handwashing, counting behaviors, and having to check and recheck such actions as whether a door is locked. The degree to which the disorder is disruptive for the individual varies, but in all cases, obsessive-compulsive activities take up >1 h per day and are undertaken to relieve the anxiety triggered by the core fear. Patients often conceal their symptoms, usually because they are embarrassed by the content of their thoughts or the nature of their actions. Physicians must ask specific questions regarding recurrent thoughts and behaviors, particularly if physical clues such as chafed and reddened hands or patchy hair loss (from repetitive hair pulling, or trichotillomania) are present. Comorbid conditions are common, the most frequent being depression, other anxiety disorders, eating disorders, and tics. OCD has a lifetime prevalence of 2–3% worldwide. Onset is usually gradual, beginning in early adulthood, but childhood onset is not rare. The disorder usually has a waxing and waning course, but some cases may show a steady deterioration in psychosocial functioning.

Etiology and Pathophysiology A genetic contribution to OCD is suggested by twin studies, but no susceptibility gene for OCD has been identified to date. Insulin signaling has been implicated in some recent reports. Family studies show an aggregation of OCD with Tourette's disorder, and both are more common in males and in first-born children.

The anatomy of obsessive-compulsive behavior is thought to include the orbital frontal cortex, caudate nucleus, and globus pallidus. The caudate nucleus appears to be involved in the acquisition and maintenance of habit and skill learning, and interventions that are successful in reducing obsessive-compulsive behaviors also decrease metabolic activity in the caudate.

TREATMENT

Obsessive-Compulsive Disorder

Clomipramine, fluoxetine, fluvoxamine, paroxetine, and sertraline are approved for the treatment of OCD in adults (and all but paroxetine are also approved for children). Clomipramine is a TCA that is often tolerated poorly owing to anticholinergic and sedative side effects at the doses required to treat the illness (25–250 mg/d); its efficacy in OCD is unrelated to its antidepressant activity.

Fluoxetine (5–60 mg/d), fluvoxamine (25–300 mg/d), paroxetine (40–60 mg/d), and sertraline (50–150 mg/d) are as effective as clomipramine and have a more benign side effect profile. Venlafaxine and duloxetine also have shown efficacy but are not FDA approved. Only 50–60% of patients with OCD show adequate improvement with pharmacotherapy alone. In treatment-resistant cases, augmentation with other serotonergic agents such as buspirone, or with a neuroleptic or benzodiazepine, may be beneficial, or alternatively, high-dose theta burst repetitive transcranial magnetic stimulation (rTMS). In severe cases, closed loop deep-brain stimulation has been found to be effective. When a therapeutic response is achieved, long-duration maintenance therapy is usually indicated.

For many individuals, particularly those with time-consuming compulsions, behavior therapy and exposure response prevention will result in as much improvement as that afforded by medication. Effective techniques include the gradual increase in exposure to stressful situations, maintenance of a diary to clarify stressors, and homework assignments that substitute new activities for compulsive behaviors.

MOOD DISORDERS

Mood disorders are characterized by a disturbance in the regulation of mood, behavior, and affect. Mood disorders are subdivided into (1) depressive disorders, (2) bipolar disorders, and (3) depression in association with medical illness or alcohol and substance abuse **(Chaps. 464 through 468)**. Major depressive disorder (MDD) is differentiated from bipolar disorder by the absence of a manic or hypomanic episode. The relationship between pure depressive syndromes and bipolar disorders is not well understood; MDD is more frequent in families of bipolar individuals, but the reverse is not true. In the most recent Global Burden of Disease Study conducted by the World Health Organization (2019), depression was the single largest factor contributing to disability, which had increased 61% as measured by disability-adjusted life-years (DALYs) since 1990. In the United States, lost productivity directly related to mood disorders has been estimated at $55.1 billion per year.

DEPRESSION IN ASSOCIATION WITH MEDICAL ILLNESS

Depression occurring in the context of medical illness is difficult to evaluate. Depressive symptomatology may reflect the psychological stress of coping with the disease, may be caused by the disease process itself or by the medications used to treat it, or may simply coexist in time with the medical diagnosis.

Virtually every class of *medication* includes some agent that can induce depression. Antihypertensive drugs, anticholesterolemic agents, and antiarrhythmic agents are common triggers of depressive symptoms. Iatrogenic depression should also be considered in patients receiving glucocorticoids, antimicrobials, systemic analgesics, antiparkinsonian medications, and anticonvulsants. To decide whether a causal relationship exists between pharmacologic therapy and a patient's change in mood, it may sometimes be necessary to undertake an empirical trial of an alternative medication.

Between 20 and 30% of *cardiac* patients manifest a depressive disorder; an even higher percentage experience depressive symptomatology when self-reporting scales are used. Depressive symptoms following unstable angina, myocardial infarction, cardiac bypass surgery, or heart transplant impair rehabilitation and are associated with higher rates of mortality and medical morbidity. Depressed patients often show decreased variability in heart rate (an index of reduced parasympathetic nervous system activity), which may predispose individuals to ventricular arrhythmia and increased morbidity. Depression also appears to increase the risk of coronary heart disease, possibly through increased platelet aggregation. TCAs are contraindicated in patients with bundle branch block, and TCA-induced tachycardia is an additional concern in patients with congestive heart failure. SSRIs appear not to induce ECG changes or adverse cardiac events and thus are reasonable first-line drugs for patients at risk for TCA-related complications. SSRIs may interfere with hepatic metabolism of anticoagulants, however, causing increased anticoagulation.

In patients with *cancer*, the mean prevalence of depression is 25%, but depression occurs in 40–50% of patients with cancers of the pancreas or oropharynx. This association is not due to the effect of cachexia alone, as the higher prevalence of depression in patients with pancreatic cancer persists when compared to those with advanced gastric cancer. Initiation of antidepressant medication in cancer patients has been shown to improve quality of life as well as mood. Psychotherapeutic approaches, particularly group therapy, may have some effect on short-term depression, anxiety, and pain symptoms.

Depression occurs frequently in patients with *neurologic disorders*, particularly cerebrovascular disorders, Parkinson's disease, dementia, multiple sclerosis, and traumatic brain injury. One in five patients with left-hemisphere stroke involving the dorsolateral frontal cortex experiences major depression. Late-onset depression in otherwise cognitively normal individuals increases the risk of a subsequent diagnosis of Alzheimer's disease. All classes of antidepressant agents are effective against these depressions, as are, in some cases, stimulant compounds. SNRIs such as duloxetine or levomilnacipran may be more effective in depression associated with chronic pain.

The reported prevalence of depression in patients with *diabetes mellitus* varies from 8 to 27%, with the severity of the mood state correlating with the level of hyperglycemia and the presence of diabetic complications. Treatment of depression may be complicated by effects of antidepressive agents on glycemic control. MAOIs can induce hypoglycemia and weight gain, whereas TCAs can produce hyperglycemia and carbohydrate craving. SSRIs and SNRIs, like MAOIs, may reduce fasting plasma glucose but are easier to use and may also improve dietary and medication compliance.

Hypothyroidism is frequently associated with features of depression, most commonly depressed mood and memory impairment. Hyperthyroid states may also present in a similar fashion, usually in geriatric populations. Improvement in mood usually follows normalization of thyroid function, but adjunctive antidepressant medication is sometimes required. Patients with subclinical hypothyroidism can also experience symptoms of depression and cognitive difficulty that respond to thyroid replacement.

The lifetime prevalence of depression in *HIV-positive* individuals has been estimated at 22–45%. The relationship between depression and disease progression is multifactorial and likely to involve psychological and social factors, alterations in immune function, and central nervous system (CNS) disease. Chronic hepatitis C infection is also associated with depression, which may worsen with interferon-α treatment.

Some chronic disorders of uncertain etiology, such as chronic fatigue syndrome **(Chap. 461)** and fibromyalgia **(Chap. 385)**, are strongly associated with depression and anxiety; patients may benefit from antidepressant treatment or anticonvulsant agents such as pregabalin.

DEPRESSIVE DISORDERS

Clinical Manifestations Major depression is defined as depressed mood on a daily basis for a minimum duration of 2 weeks **(Table 463-7)**. An episode may be characterized by sadness, indifference, apathy, or irritability and is usually associated with changes in sleep patterns, appetite, and weight; motor agitation or retardation; fatigue; impaired concentration and decision-making; feelings of shame or guilt; and thoughts of death or dying. Patients with depression have a profound loss of pleasure in all enjoyable activities, exhibit early-morning awakening, feel that the dysphoric mood state is qualitatively different from sadness, and often notice a diurnal variation in mood (worse in morning hours). Patients experiencing bereavement or grief may exhibit many of the same signs and symptoms of major depression, although the emphasis is usually on feelings of emptiness and loss, rather than anhedonia and loss of self-esteem, and the duration is usually limited. In certain cases, however, the diagnosis of major depression may be warranted even in the context of a significant loss.

Approximately 15% of the population experiences a major depressive episode at some point in life, and 6–8% of all outpatients in

TABLE 463-7 Criteria for a Major Depressive Episode

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

Note: Do not include symptoms that are clearly attributable to another medical condition.

1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (**Note:** In children and adolescents, can be irritable mood.)
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (Note: In children, consider failure to make expected weight gain.)
4. Insomnia or hypersomnia nearly every day.
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

B. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

C. The episode is not attributable to the physiological effects of a substance or another medical condition.

Note: Criteria A–C represent a major depressive episode.

Note: Responses to a significant loss (e.g., bereavement, financial ruin, losses from a natural disaster, a serious medical illness or disability) may include the feelings of intense sadness, rumination about the loss, insomnia, poor appetite, and weight loss noted in Criterion A, which may resemble a depressive episode.

Although such symptoms may be understandable or considered appropriate to the loss, the presence of a major depressive episode in addition to the normal response to a significant loss should also be carefully considered. This decision inevitably requires the exercise of clinical judgment based on the individual's history and the cultural norms for the expression of distress in the context of loss.[1]

D. At least one major depressive episode is not better explained by schizoaffective disorder and is not superimposed on schizophrenia, schizophreniform disorder, delusional disorder, or other specified and unspecified schizophrenia spectrum and other psychotic disorders.

E. There has never been a manic episode or a hypomanic episode.

Note: This exclusion does not apply if all of the manic-like or hypomanic-like episodes are substance-induced or are attributable to the physiological effects of another medical condition.

[1] In distinguishing grief from a major depressive episode (MDE), it is useful to consider that in grief the predominant affect is feelings of emptiness and loss, while in an MDE it is persistent depressed mood and the inability to anticipate happiness or pleasure. The dysphoria in grief is likely to decrease in intensity over days to weeks and occurs in waves, the so-called pangs of grief. These waves tend to be associated with thoughts or reminders of the deceased. The depressed mood of an MDE is more persistent and not tied to specific thoughts or preoccupations. The pain of grief may be accompanied by positive emotions and humor that are uncharacteristic of the pervasive unhappiness and misery characteristic of an MDE. The thought content associated with grief generally features a preoccupation with thoughts and memories of the deceased, rather than the self-critical or pessimistic ruminations seen in an MDE. In grief, self-esteem is generally preserved, whereas in an MDE feelings of worthlessness and self-loathing are common. If self-derogatory ideation is present in grief, it typically involves perceived failings vis-à-vis the deceased (e.g., not visiting frequently enough, not telling the deceased how much he or she was loved). If a bereaved individual thinks about death and dying, such thoughts are generally focused on the deceased and possibly about "joining" the deceased, whereas in an MDE such thoughts are focused on ending one's own life because of feeling worthless, undeserving of life, or unable to cope with the pain of depression.

Source: Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*-Text Revision, 5th ed. (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

primary care settings satisfy diagnostic criteria for the disorder. Depression is often undiagnosed and, even more frequently, is treated inadequately. If a physician suspects the presence of a major depressive episode, the initial task is to determine whether it represents unipolar or bipolar depression or is one of the 10–15% of cases that are secondary to general medical illness or substance abuse. Physicians should also assess the risk of suicide by direct questioning, as patients are often reluctant to verbalize such thoughts without prompting. If specific plans are uncovered or if significant risk factors exist (e.g., a past history of suicide attempts, profound hopelessness, concurrent medical illness, substance abuse, or social isolation), the patient must be referred to a mental health specialist for immediate care. The physician should specifically probe each of these areas in an empathic and hopeful manner, being sensitive to denial and possible minimization of distress. The presence of anxiety, panic, or agitation significantly increases near-term suicidal risk. Approximately 4–5% of all depressed patients will commit suicide; most will have sought help from physicians within 1 month of their deaths.

In some depressed patients, the mood disorder does not appear to be episodic and is not clearly associated with either psychosocial dysfunction or change from the individual's usual experience in life. Persistent depressive disorder (*dysthymic disorder*) consists of a pattern of chronic (at least 2 years), ongoing depressive symptoms that are usually less severe and/or less numerous than those found in major depression, but the functional consequences may be equivalent to or even greater; the two conditions are sometimes difficult to separate and can occur together ("double depression"). Many patients who exhibit a profile of pessimism, disinterest, and low self-esteem respond to antidepressant treatment. Persistent and chronic depressive disorders occur in ~2% of the general population.

Depression is approximately twice as common in women as in men, and the incidence increases with age in both sexes. Twin studies indicate that the liability to major depression of early onset (before age 25 years) is largely genetic in origin. Negative life events can precipitate and contribute to depression, but genetic factors influence the sensitivity of individuals to these stressful events. In most cases, both biologic and psychosocial factors are involved in the precipitation and unfolding of depressive episodes. The most potent stressors appear to involve death of a relative, assault, or severe marital or relationship problems.

Unipolar depressive disorders usually begin in early adulthood and recur episodically over the course of a lifetime. The best predictor of future risk is the number of past episodes; 50–60% of patients who have a first episode have at least one or two recurrences. Some patients experience multiple episodes that become more severe and frequent over time. The duration of an untreated episode varies greatly, ranging from a few months to ≥1 year. The pattern of recurrence and clinical progression in a developing episode are also variable. Within an individual, the nature of episodes (e.g., specific presenting symptoms, frequency, and duration) may be similar over time. In a minority of patients, a severe depressive episode can progress to a psychotic state, and in elderly patients, depressive symptoms can be associated with cognitive deficits mimicking dementia ("pseudodementia"). A seasonal pattern of depression, called *seasonal affective disorder*, may manifest with onset and remission of episodes at predictable times of the year. This disorder is more common in women, with symptoms of anergy, fatigue, weight gain, hypersomnia, and episodic carbohydrate craving. The prevalence increases with distance from the equator, and improvement may occur by altering light exposure.

Etiology and Pathophysiology Although evidence for genetic transmission of unipolar depression is not as strong as in bipolar disorder, monozygotic twins have a higher concordance rate (46%) than dizygotic siblings (20%), with little support for any effect of a shared family environment. Large-scale genome-wide association studies (GWAS) involving hundreds of thousands of cases and controls have identified several hundred loci across the genome, some of which are unique to major depression, but others of which overlap with findings from disparate psychiatric disorders, indicating possible pleiotropy. Epigenetic changes are also likely to contribute to risk.

Neuroendocrine abnormalities that reflect the neurovegetative signs and symptoms of depression include increased cortisol and corticotropin-releasing hormone (CRH) secretion, a decreased inhibitory response of glucocorticoids to dexamethasone, and a blunted response of thyroid-stimulating hormone (TSH) level to infusion of thyroid-releasing hormone (TRH). Antidepressant treatment leads to normalization of these abnormalities. Major depression is also associated with changes in levels of proinflammatory cytokines and neurotrophins, an increase in measures of oxidative stress and cellular aging, telomere shortening, epigenetic changes, and mitochondrial dysfunction. Alterations in the gut microbiome may also be involved.

Diurnal variations in symptom severity and alterations in circadian rhythmicity of a number of neurochemical and neurohumoral factors suggest that a primary defect may be present in regulation of biologic rhythms. Patients with major depression show consistent findings of a decrease in rapid eye movement (REM)–sleep onset (REM latency), an increase in REM density, and, in some subjects, a decrease in stage IV delta slow-wave sleep.

Although antidepressant drugs inhibit neurotransmitter uptake within hours, their therapeutic effects typically emerge over several weeks, implicating adaptive changes in second messenger systems and neurotrophic and transcription factors as possible mechanisms of action.

TREATMENT

Depressive Disorders

Treatment planning requires coordination of short-term strategies to induce remission combined with longer-term maintenance designed to prevent recurrence. The most effective intervention for achieving remission and preventing relapse is medication, but combined treatments, incorporating psychotherapy to help the patient cope with decreased self-esteem and demoralization, improve outcomes, as do self-help strategies such as exercise **(Fig. 463-1)**. Approximately 40% of primary care patients with depression drop out of treatment and discontinue medication if symptomatic improvement is not noted within a month, unless additional support is provided. Outcome improves with (1) increased intensity and frequency of visits during the first 4–6 weeks of treatment, (2) supplemental educational materials, and (3) psychiatric consultation as indicated. Despite the widespread use of SSRIs and other second-generation antidepressant drugs, there is no convincing evidence that these classes of antidepressants are more efficacious than TCAs. Between 60 and 70% of all depressed patients respond to any drug chosen, if it is given in a sufficient dose for 6–8 weeks.

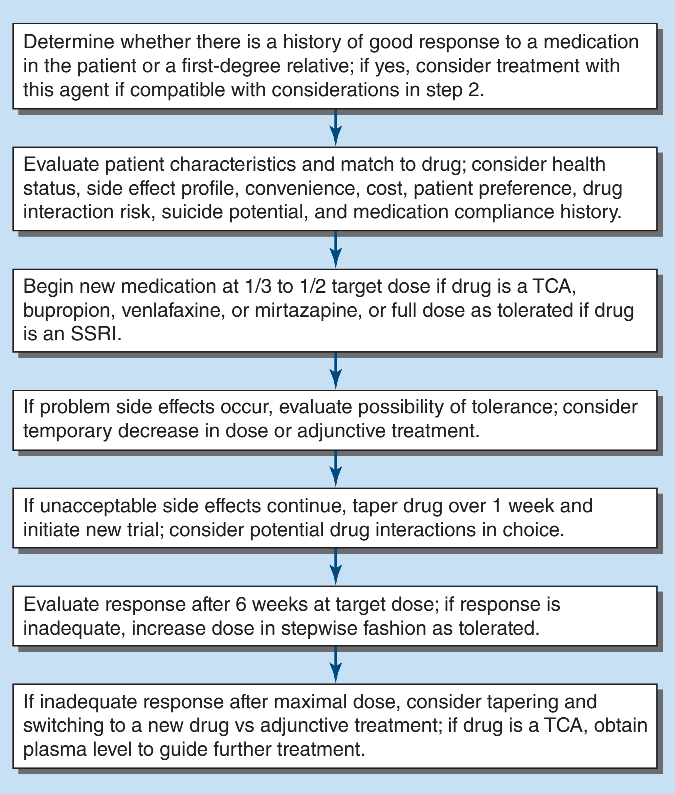

Determine whether there is a history of good response to a medication in the patient or a first-degree relative; if yes, consider treatment with this agent if compatible with considerations in step 2.

Evaluate patient characteristics and match to drug; consider health status, side effect profile, convenience, cost, patient preference, drug interaction risk, suicide potential, and medication compliance history.

Begin new medication at 1/3 to 1/2 target dose if drug is a TCA, bupropion, venlafaxine, or mirtazapine, or full dose as tolerated if drug is an SSRI.

If problem side effects occur, evaluate possibility of tolerance; consider temporary decrease in dose or adjunctive treatment.

If unacceptable side effects continue, taper drug over 1 week and initiate new trial; consider potential drug interactions in choice.

Evaluate response after 6 weeks at target dose; if response is inadequate, increase dose in stepwise fashion as tolerated.

If inadequate response after maximal dose, consider tapering and switching to a new drug vs adjunctive treatment; if drug is a TCA, obtain plasma level to guide further treatment.

FIGURE 463-1 A guideline for the medical management of major depressive disorder. SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressant.

A rational approach to selecting which antidepressant to use (Table 463-1) involves matching the patient's preference and medical history with the metabolic and side effect profile of the drug (Tables 463-2 and 463-3). A previous response, or a family history of a positive response, to a specific antidepressant often suggests that drug should be tried first. Before initiating antidepressant therapy, the physician should evaluate the possible contribution of comorbid illnesses and consider their specific treatment. In individuals with suicidal ideation, particular attention should be paid to choosing a drug with low toxicity if taken in overdose. Newer antidepressant drugs are distinctly safer in this regard; nevertheless, the advantages of TCAs have not been completely superseded. The existence of generic equivalents makes TCAs relatively cheap, and for secondary tricyclics, particularly nortriptyline and desipramine, well-defined relationships among dose, plasma level, and therapeutic response exist. The steady-state plasma level achieved for a given drug dose can vary more than tenfold between individuals, and plasma levels may help in interpreting apparent resistance to treatment and/or unexpected drug toxicity. The principal side effects of TCAs are antihistaminergic (sedation) and anticholinergic (constipation, dry mouth, urinary hesitancy, blurred vision). TCAs are contraindicated in patients with serious cardiovascular risk factors, and overdoses of tricyclic agents can be lethal, with desipramine carrying the greatest risk. It is judicious to prescribe only a 10-day supply when suicide is a risk. Most patients require a daily dose of 150–200 mg of imipramine or amitriptyline or its equivalent to achieve a therapeutic blood level of 150–300 ng/mL and a satisfactory remission; some patients show a partial effect at lower doses. Geriatric patients may require a low starting dose and slow escalation. Ethnic differences in drug metabolism are significant, with Hispanic, Asian, and black patients generally requiring lower doses to achieve a comparable blood level.

Second-generation antidepressants are similar to tricyclics in their effect on neurotransmitter reuptake, although some also have specific actions on catecholamine and indolamine receptors as well. Amoxapine is a dibenzoxazepine derivative that blocks norepinephrine and serotonin reuptake and has a metabolite that shows a degree of dopamine blockade. Long-term use of this drug carries a risk of tardive dyskinesia. Maprotiline is a potent noradrenergic reuptake blocker that has little anticholinergic effect but may produce seizures. Bupropion is a novel antidepressant whose mechanism of action is thought to involve enhancement of noradrenergic function. It has no anticholinergic, sedating, or orthostatic side effects and has a low incidence of sexual side effects. It may, however, be associated with stimulant-like side effects, may lower seizure threshold, and has an exceptionally short half-life, requiring frequent dosing. An extended-release preparation is available, as is a version combining it with dextromethorphan, an *N*-methyl-D-aspartate (NMDA) receptor agonist, that is proposed to result in a more rapid therapeutic response.

SSRIs such as fluoxetine, sertraline, paroxetine, citalopram, and escitalopram cause a lower frequency of anticholinergic, sedating, and cardiovascular side effects but a possibly greater incidence of gastrointestinal complaints, sleep impairment, and sexual dysfunction than do TCAs. Akathisia, involving an inner sense of restlessness and anxiety in addition to increased motor activity, may also be more common, particularly during the first week of treatment. One concern is the risk of "serotonin syndrome," which is thought to result from hyperstimulation of brainstem 5-HT_{1A} receptors and is characterized by myoclonus, agitation, abdominal cramping, hyperpyrexia, hypertension, and potentially death. Serotonergic agonists taken in combination should be monitored closely for this reason. Considerations such as half-life, compliance, toxicity, and drug-drug interactions may guide the choice of a particular SSRI. Fluoxetine and its principal active metabolite, norfluoxetine, for example, have a combined half-life of almost 7 days, resulting in a delay of 5 weeks before steady-state levels are achieved and a similar delay for complete drug excretion once their use is discontinued; paroxetine appears to incur a greater risk of withdrawal symptoms with abrupt discontinuation. All the SSRIs may impair sexual function, resulting in diminished libido, impotence, or difficulty in achieving orgasm. Sexual dysfunction frequently results in noncompliance and should be asked about specifically. Sexual dysfunction can sometimes be ameliorated by lowering the dose, by instituting weekend drug holidays (two or three times a month), or by treatment with amantadine (100 mg tid), bethanechol (25 mg tid), buspirone (10 mg tid), or bupropion (100–150 mg/d). Paroxetine appears to be more anticholinergic than either fluoxetine or sertraline, and sertraline carries a lower risk of producing an adverse drug interaction than the other two. Rare side effects of SSRIs include angina due to vasospasm and prolongation of the prothrombin time. Escitalopram is the most specific of currently available SSRIs and appears to have no significant inhibitory effects on the P450 system.

Venlafaxine, desvenlafaxine, duloxetine, and levomilnacipran block the reuptake of both norepinephrine and serotonin but produce relatively little in the way of traditional tricyclic side effects. Vortioxetine, also a 5-HT$_{1A}$ agonist, and vilazodone block reuptake of serotonin but have negligible effects on norepinephrine reuptake, although vortioxetine may increase norepinephrine levels through wide effects on serotonergic receptors, as a 5-HT$_{1A}$ agonist, 5-HT$_{1B}$ partial agonist, and a 5-HT$_{1D}$, 5-HT$_3$, and 5-HT$_7$ antagonist. Unlike the SSRIs, venlafaxine and vortioxetine have relatively linear dose-response curves. Patients on immediate-release venlafaxine should be monitored for a possible increase in diastolic blood pressure, and multiple daily dosing is required because of the drug's short half-life. An extended-release form is available and has a somewhat lower incidence of gastrointestinal side effects. Mirtazapine is a tetracyclic that has a unique spectrum of activity, as it increases noradrenergic and serotonergic neurotransmission through a blockade of central α_2-adrenergic receptors and postsynaptic 5-HT$_2$ and 5-HT$_3$ receptors. It is also strongly antihistaminic and, as such, may produce sedation. Levomilnacipran is the most noradrenergic of the SNRIs and theoretically may be appropriate for patients with more severe fatigue and anergia. Gepirone, an older drug recently approved, is a partial 5-HT$_{1A}$ agonist and 5-HT$_{2A}$ antagonist.

With the exception of citalopram and escitalopram, each of the SSRIs may inhibit one or more cytochrome P450 enzymes. Depending on the specific isoenzyme involved, the metabolism of a number of concomitantly administered medications can be dramatically affected. Fluoxetine and paroxetine, for example, by inhibiting 2D6, can cause dramatic increases in the blood level of type 1C antiarrhythmics, whereas sertraline, by acting on 3A4, may alter blood levels of carbamazepine or digoxin. Depending on drug specificity for a particular CYP enzyme for its own metabolism, concomitant medications or dietary factors, such as grapefruit juice, may in turn affect the efficacy or toxicity of the SSRI.

The MAOIs are highly effective, particularly in atypical depression, but the risk of hypertensive crisis following intake of tyramine-containing food or sympathomimetic drugs makes them inappropriate as first-line agents. Transdermal selegiline may avert this risk at low dose. Common side effects include orthostatic hypotension, weight gain, insomnia, and sexual dysfunction. MAOIs should not be used concomitantly with SSRIs, because of the risk of serotonin syndrome, or with TCAs, because of possible hyperadrenergic effects.

Electroconvulsive therapy is at least as effective as medication, but its use is reserved for treatment-resistant cases and delusional depressions. rTMS is approved for treatment-resistant depression and has been shown to have efficacy in several controlled trials. Vagus nerve stimulation (VNS) has also recently been approved for treatment-resistant depression, but its degree of efficacy is controversial. Some meta-analyses of low-intensity transcranial current stimulation (tCS) have shown a positive benefit over sham treatment, but whether this is comparable to or synergistic with antidepressant treatment is unclear. In off-label usage, intravenous ketamine, a dissociative anesthetic, and intranasal esketamine (an isomer that has FDA approval in treatment-resistant cases) have been shown to have short-term antidepressant efficacy, often after a single administration, and may decrease suicidality. Questions remain, however, about the risk/benefit ratio over the longer term. Psilocybin, a hallucinogen, has also shown some potential benefit in controlled administration. Lastly, deep brain stimulation of the ventral anterior limb of the internal capsule and of the subcallosal cingulate region has demonstrable efficacy in randomized experimental trials of treatment-resistant depression.

Postpartum depression may respond to any of the above interventions, but a neuroactive steroid, brexanolone (Zulresso), administered in a continuous intravenous infusion over 60 h, can provide symptomatic relief for at least 30 days. Sedation and loss of consciousness are possible adverse effects. An oral version, zuranolone (Zurzuvae), can be given on an outpatient basis.

Regardless of the treatment undertaken, the response should be evaluated after ~2 months. Three-quarters of patients show improvement by this time, but if remission is inadequate, the patient should be questioned about compliance, and an increase in medication dose should be considered if side effects are not troublesome. If this approach is unsuccessful, referral to a mental health specialist is advised. Strategies for treatment resistance include selection of an alternative drug, combinations of antidepressants, and/or adjunctive treatment with other classes of drugs, including lithium, thyroid hormone, L-methylfolate, *S*-adenosylmethionine, *N*-acetyl cysteine, atypical antipsychotic agents, and dopamine agonists. In switching to a different monotherapy, other drugs from the same class appear to be as likely to be efficacious as choosing a drug from a different class. A large randomized trial (STAR-D) was unable to show preferential efficacy, but the addition of certain atypical antipsychotic drugs (quetiapine extended-release; aripiprazole; brexpiprazole) has received FDA approval, as has usage of a combined medication, olanzapine and fluoxetine (Symbyax). Patients whose response to an SSRI wanes over time may benefit from the addition of buspirone (10 mg tid) or pindolol (2–5 mg tid) or small amounts of a TCA such as nortriptyline (25 mg bid or tid). Most patients will show some degree of response, but aggressive treatment should be pursued until remission is achieved, and drug treatment should be continued for at least 6–9 months to prevent relapse. In patients who have had two or more episodes of depression, indefinite maintenance treatment should be considered. Pharmacogenomic testing focusing on cytochrome p450 allelic variation may sometimes be helpful in identifying individuals who are poor or rapid metabolizers, but assessing pharmacodynamic gene variants has not been shown to be cost-effective or affect clinical outcomes.

It is essential to educate patients both about depression and the benefits and side effects of medications they are receiving. Advice about stress reduction and cautions that alcohol may exacerbate depressive symptoms and impair drug response are helpful. Patients should be given time to describe their experience, their outlook, and the impact of the depression on them and their families. Occasional empathic silence may be as helpful for the treatment alliance as verbal reassurance. Controlled trials have shown that cognitive-behavioral and interpersonal therapies are effective in improving psychological and social adjustment and that a combined treatment approach is more successful than medication alone for many patients.

BIPOLAR DISORDER

Clinical Manifestations Bipolar disorder is characterized by unpredictable swings in mood from mania (or hypomania) to depression. Some patients suffer only from recurrent attacks of *mania*, which in its pure form is associated with increased psychomotor activity; excessive social extroversion; decreased need for sleep; impulsivity and impaired judgment; and expansive, elated, grandiose, and sometimes irritable mood **(Table 463-8)**. In severe mania, patients may experience delusions and paranoid thinking indistinguishable from schizophrenia. One-half of patients with bipolar disorder present not with euphoria but with a mixture of psychomotor agitation and activation, accompanied by dysphoria, anxiety, and irritability. It may be difficult to distinguish such a mixed state from agitated depression. In some bipolar patients (*bipolar II disorder*), the full criteria for mania are lacking, and the requisite recurrent depressions are separated by periods of mild activation and increased energy (hypomania). In *cyclothymic disorder*, there are numerous hypomanic periods, usually of relatively short duration, alternating with clusters of depressive symptoms that fail, either in severity or duration, to meet the criteria of major depression. The mood fluctuations are chronic and should be present for at least 2 years before the diagnosis is made.

Manic episodes typically emerge over a period of days to weeks, but onset within hours is possible, usually in the early-morning hours. An untreated episode of either depression or mania can be as short as several weeks or last as long as 8–12 months, and rare patients have an unremitting chronic course. The term *rapid cycling* is used for patients who have four or more episodes of either depression or mania in a given year. This pattern occurs in 15% of all patients, most of whom are women. In some cases, rapid cycling is linked to an underlying thyroid dysfunction, and in others, it is iatrogenically triggered by prolonged antidepressant treatment. Approximately one-half of patients have sustained difficulties in work performance and psychosocial functioning, with depressive phases being more responsible for impairment than mania.

Bipolar disorder is common, affecting ~1.5% of the population in the United States. Onset is typically between 20–30 years of age, but many individuals report premorbid symptoms in late childhood or early adolescence. The prevalence is similar for men and women; women are likely to have more depressive and men more manic episodes over a lifetime. Recognizing a bipolar diathesis in an individual who presents with a depressive episode but no history of mania is difficult but essential in optimizing treatment planning, because antidepressants may be contraindicated and result in symptom worsening and cycle acceleration. Suggestive features of bipolarity include a childhood onset, a history of antidepressant treatment failure, atypical features of hypersomnolence and weight gain, and marked irritability or impulsivity.

Differential Diagnosis The differential diagnosis of mania includes secondary mania induced by stimulant or sympathomimetic drugs, hyperthyroidism, AIDS, neurologic disorders such as Huntington's or Wilson's disease, frontotemporal dementia, and cerebrovascular accidents. Comorbidity with alcohol and substance abuse is common, either because of shared genetic risk, poor judgment, and increased impulsivity or because of an attempt to self-treat the underlying mood symptoms and sleep disturbances.

Etiology and Pathophysiology Genetic predisposition to bipolar disorder is evident from family studies; the concordance rate for monozygotic twins approaches 80%. A number of risk genes that have been identified to date overlap with those conveying risk for other psychiatric disorders, such as schizophrenia and autism, implying some degree of shared pathophysiology. Replicated loci include the alpha subunit of the L-type calcium channel (*CACNA1C*), teneurin transmembrane protein 4 (*ODZ4*), ankyrin 3 (*ANK3*), neurocan (*NCAN*), and tetratricopeptide repeat and ankyrin repeat containing 1 (*TRANK1*). Common variants convey little individual risk but collectively account for 25% of heritability. A few rarer, more penetrant variants have also been reported, but no causative mutations have as yet been confirmed. Similarly, no clear biomarkers have been identified, but there is evidence for circadian rhythm and calcium dysregulation and oxidative stress, and mitochondrial, microRNA, and endoplasmic reticulum abnormalities. Reported MRI findings include gray matter thinning in frontal, temporal, and parietal cortex.

TREATMENT

Bipolar Disorder

(Table 463-9) Lithium carbonate is the mainstay of treatment in bipolar disorder, although paradoxically underutilized. Sodium valproate and carbamazepine, as well as a number of second-generation antipsychotic agents (e.g., aripiprazole, asenapine, cariprazine, olanzapine, quetiapine, risperidone, ziprasidone), also have FDA approval for the treatment of acute mania. Oxcarbazepine is not FDA approved but appears to enjoy carbamazepine's spectrum of efficacy. The response rate to lithium carbonate is 70–80% in acute mania, with beneficial effects appearing in 1–2 weeks. Lithium also has a prophylactic effect in prevention of recurrent mania and, to a lesser extent, in the prevention of recurrent depression, which is more difficult to treat than unipolar depression. A simple cation, lithium is rapidly absorbed from the gastrointestinal tract and remains unbound to plasma or tissue proteins. Some 95% of a given dose is excreted unchanged through the kidneys within 24 h.

Serious side effects from lithium are rare, but minor complaints such as gastrointestinal discomfort, nausea, diarrhea, polyuria, weight gain, skin eruptions, alopecia, and edema are common. Over time, urine-concentrating ability may be decreased, but significant nephrotoxicity is relatively rare. Lithium exerts an antithyroid effect by interfering with the synthesis and release of thyroid hormones. More serious side effects include tremor, poor concentration and memory, ataxia, dysarthria, and incoordination.

In the treatment of acute mania, lithium is initiated at 300 mg bid or tid, and the dose is then increased by 300 mg every 2–3 days to achieve blood levels of 0.8–1.2 meq/L. Because the therapeutic effect of lithium may not appear until after 7–10 days of treatment, adjunctive usage of lorazepam (1–2 mg every 4 h) or clonazepam (0.5–1 mg every 4 h) may be beneficial to control agitation. Antipsychotics are indicated in patients with severe agitation who respond only partially to benzodiazepines. Patients using lithium should be monitored closely, because the blood levels required to achieve a therapeutic benefit are close to those associated with toxicity.

Valproic acid may be more effective than lithium for patients who experience rapid cycling (i.e., more than four episodes a year) or who present with a mixed or dysphoric mania. Tremor and weight gain are the most common side effects; hepatotoxicity and pancreatitis are rare toxicities.

The recurrent nature of bipolar mood disorder necessitates maintenance treatment. A sustained blood lithium level of at least 0.8 meq/L is important for optimal prophylaxis and has been shown to reduce the risk of suicide, a finding not yet apparent for other mood stabilizers. Decrease in acute suicidal risk may also result from treatment with parenteral ketamine or intranasal esketamine. Combinations of mood stabilizers together or with atypical antipsychotic drugs are sometimes required to maintain mood stability. Quetiapine extended release, olanzapine, risperidone, and lamotrigine have been approved for maintenance treatment as sole agents, in combination with lithium and with aripiprazole and ziprasidone as adjunctive drugs. Lurasidone, olanzapine/fluoxetine, and quetiapine are also approved to treat acute depressive episodes in bipolar disorder. Compliance is frequently an issue and often requires enlistment and education of concerned family members. Efforts to identify and modify psychosocial factors that may trigger episodes are important, as is an emphasis on lifestyle regularity (social rhythm therapy). Mobile apps for smartphones that alert the individual and clinician to changes in activity and speech are proving useful in early detection of behavioral change and in delivering clinical interventions and education. Antidepressant medications are sometimes required for the treatment of severe breakthrough depressions, but their use should generally be avoided during maintenance treatment because of the risk of precipitating mania or accelerating the cycle frequency. Alternative off-label agents for bipolar depression include pramipexole, modafinil, omega-3 fatty acids, and *N*-acetyl cysteine; interventions such as electroconvulsive

TABLE 463-8 Criteria for Bipolar I Disorder

For a diagnosis of bipolar I disorder, it is necessary to meet the following criteria for a manic episode. The manic episode may have been preceded by and may be followed by hypomanic or major depressive episodes.

Manic Episode

A. A distinct period of abnormally and persistently elevated, expansive, or irritable mood and abnormally and persistently increased activity or energy, lasting at least 1 week and present most of the day, nearly every day (or any duration if hospitalization is necessary).

B. During the period of mood disturbance and increased energy or activity, three (or more) of the following symptoms (four if the mood is only irritable) are present to a significant degree and represent a noticeable change from usual behavior:

1. Inflated self-esteem or grandiosity.
2. Decreased need for sleep (e.g., feels rested after only 3 hours of sleep).
3. More talkative than usual or pressure to keep talking.
4. Flight of ideas or subjective experience that thoughts are racing.
5. Distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli), as reported or observed.
6. Increase in goal-directed activity (either socially, at work or school, or sexually) or psychomotor agitation (i.e., purposeless non-goal-directed activity).
7. Excessive involvement in activities that have a high potential for painful consequences (e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments).

C. The mood disturbance is sufficiently severe to cause marked impairment in social or occupational functioning or to necessitate hospitalization to prevent harm to self or others, or there are psychotic features.

D. The episode is not attributable to the physiological effects of a substance (e.g., a drug of abuse, a medication, other treatment) or another medical condition.

Note: A full manic episode that emerges during antidepressant treatment (e.g., medication, electroconvulsive therapy) but persists at a fully syndromal level beyond the physiological effect of that treatment is sufficient evidence for a manic episode and, therefore, a bipolar I diagnosis.

Note: Criteria A–D constitute a manic episode. At least one lifetime manic episode is required for the diagnosis of bipolar I disorder.

Hypomanic Episode

A. A distinct period of abnormally and persistently elevated, expansive, or irritable mood and abnormally and persistently increased activity or energy, lasting at least 4 consecutive days and present most of the day, nearly every day.

B. During the period of mood disturbance and increased energy and activity, three (or more) of the following symptoms (four if the mood is only irritable) have persisted, represent a noticeable change from usual behavior, and have been present to a significant degree:

1. Inflated self-esteem or grandiosity.
2. Decreased need for sleep (e.g., feels rested after only 3 hours of sleep).
3. More talkative than usual or pressure to keep talking.
4. Flight of ideas or subjective experience that thoughts are racing.
5. Distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli), as reported or observed.
6. Increase in goal-directed activity (either socially, at work or school, or sexually) or psychomotor agitation.
7. Excessive involvement in activities that have a high potential for painful consequences (e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments).

C. The episode is associated with an unequivocal change in functioning that is uncharacteristic of the individual when not symptomatic.

D. The disturbance in mood and the change in functioning are observable by others.

E. The episode is not severe enough to cause marked impairment in social or occupational functioning or to necessitate hospitalization. If there are psychotic features, the episode is, by definition, manic.

F. The episode is not attributable to the physiological effects of a substance (e.g., a drug of abuse, a medication, other treatment) or another medical condition.

Note: A full hypomanic episode that emerges during antidepressant treatment (e.g., medication, electroconvulsive therapy) but persists at a fully syndromal level beyond the physiological effect of that treatment is sufficient evidence for a hypomanic episode diagnosis. However, caution is indicated so that one or two symptoms (particularly increased irritability, edginess, or agitation following antidepressant use) are not taken as sufficient for diagnosis of a hypomanic episode, nor necessarily indicative of a bipolar diathesis.

Note: Criteria A–F constitute a hypomanic episode. Hypomanic episodes are common in bipolar I disorder but are not required for the diagnosis of bipolar I disorder.

Major Depressive Episode

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

Note: Do not include symptoms that are clearly attributable to another medical condition.

1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, or hopeless) or observation made by others (e.g., appears tearful). (**Note:** In children and adolescents, can be irritable mood.)
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (**Note:** In children, consider failure to make expected weight gain.)
4. Insomnia or hypersomnia nearly every day.
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
9. Recurrent thoughts of death (not just fear of dying); recurrent suicidal ideation without a specific plan; a specific suicide plan; or a suicide attempt.

B. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

C. The episode is not attributable to the physiological effects of a substance or another medical condition.

Note: Criteria A–C constitute a major depressive episode. Major depressive episodes are common in bipolar I disorder but are not required for the diagnosis of bipolar I disorder.

Note: Responses to a significant loss (e.g., bereavement, financial ruin, losses from a natural disaster, a serious medical illness or disability) may include the feelings of intense sadness, rumination about the loss, insomnia, poor appetite, and weight loss noted in Criterion A, which may resemble a depressive episode. Although such symptoms may be understandable or considered appropriate to the loss, the presence of a major depressive episode in addition to the normal response to a significant loss should also be carefully considered. This decision inevitably requires the exercise of clinical judgment based on the individual's history and the cultural norms for the expression of distress in the context of loss.

Bipolar I Disorder

A. Criteria have been met for at least one manic episode (Criteria A–D under "Manic Episode" above).

B. At least one manic episode is not better explained by schizoaffective disorder and is not superimposed on schizophrenia, schizophreniform disorder, delusional disorder, or other specified or unspecified schizophrenia spectrum and other psychotic disorder.

Source: Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*-Text Revision, 5th ed. (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

TABLE 463-9 Clinical Pharmacology of Mood Stabilizers

AGENT AND DOSING	SIDE EFFECTS AND OTHER EFFECTS
Lithium	**Common Side Effects**
Starting dose: 300 mg bid or tid Therapeutic blood level: 0.8–1.2 meq/L	Nausea/anorexia/diarrhea, fine tremor, thirst, polyuria, fatigue, weight gain, acne, folliculitis, neutrophilia, hypothyroidism Blood level is increased by thiazides, tetracyclines, and NSAIDs Blood level is decreased by bronchodilators, verapamil, and carbonic anhydrase inhibitors *Rare side effects:* Neurotoxicity, renal toxicity, hypercalcemia, ECG changes
Valproic Acid	**Common Side Effects**
Starting dose: 250 mg tid Therapeutic blood level: 50–125 μg/mL	Nausea/anorexia, weight gain, sedation, tremor, rash, alopecia Inhibits hepatic metabolism of other medications *Rare side effects:* Pancreatitis, hepatotoxicity, Stevens-Johnson syndrome
Carbamazepine/ Oxcarbazepine	**Common Side Effects**
Starting dose: 200 mg bid for carbamazepine, 150 mg bid for oxcarbazepine Therapeutic blood level: 4–12 μg/mL for carbamazepine	Nausea/anorexia, sedation, rash, dizziness/ ataxia Carbamazepine, but not oxcarbazepine, induces hepatic metabolism of other medications *Rare side effects:* Hyponatremia, agranulocytosis, Stevens-Johnson syndrome
Lamotrigine	**Common Side Effects**
Starting dose: 25 mg/d	Rash, dizziness, headache, tremor, sedation, nausea *Rare side effect:* Stevens-Johnson syndrome

Abbreviations: ECG, electrocardiogram; NSAIDs, nonsteroidal anti-inflammatory drugs.

therapy, light therapy, and rTMS may also be effective. Loss of efficacy over time may be observed with any of the mood-stabilizing agents. In such situations, an alternative agent or combination therapy is usually helpful.

SOMATIC SYMPTOM DISORDER

Many patients presenting in general medical practice, perhaps as many as 5–7%, will experience a somatic symptom(s) as particularly distressing and preoccupying, to the point that it comes to dominate their thoughts, feelings, and beliefs and interferes to a varying degree with everyday functioning. Although the absence of a medical explanation for these complaints was historically emphasized as a diagnostic element, it has been recognized that the patient's interpretation and elaboration of the experience is the critical defining factor and that patients with well-established medical causation may qualify for the diagnosis. Multiple complaints are typical, but severe single symptoms can occur as well. Comorbidity with depressive and anxiety disorders is common and may affect the severity of the experience and its functional consequences. Personality factors may be a significant risk factor, as may a low level of educational or socioeconomic status or a history of recent stressful life events. Cultural factors are relevant as well and should be incorporated into the evaluation. Individuals who have persistent preoccupations about having or acquiring a serious illness, but who do not have a specific somatic complaint, may qualify for a related diagnosis—illness anxiety disorder. The diagnosis of conversion disorder (functional neurologic symptom disorder) is used to specifically identify individuals whose somatic complaints involve one or more symptoms of altered voluntary motor or sensory function that cannot be medically explained and that cause significant distress or impairment or require medical evaluation.

In *factitious illnesses*, the patient consciously and voluntarily produces physical symptoms of illness. The term *Munchausen's syndrome* is reserved for individuals with particularly dramatic, chronic, or severe factitious illness. In true factitious illness, the sick role itself is gratifying. A variety of signs, symptoms, and diseases have been either simulated or caused by factitious behavior, the most common including chronic diarrhea, fever of unknown origin, intestinal bleeding or hematuria, seizures, and hypoglycemia. Factitious disorder is usually not diagnosed until 5–10 years after its onset, and it can produce significant social and medical costs. In *malingering*, the fabrication derives from a desire for some external reward such as a narcotic medication or disability reimbursement.

TREATMENT

Somatic Symptom Disorder and Related Disorders

Patients with somatic symptom disorder are frequently subjected to many diagnostic tests and exploratory surgeries in an attempt to find their "real" illness. Such an approach is doomed to failure and does not address the core issue. Successful treatment is best achieved through behavior modification, in which access to the physician is tightly regulated and adjusted to provide a sustained and predictable level of support that is less clearly contingent on the patient's level of presenting distress. Visits can be brief and should not be associated with a need for a diagnostic or treatment action. Although the literature is limited, some patients may benefit from antidepressant treatment.

Any attempt to confront the patient usually creates a sense of humiliation and causes the patient to abandon treatment from that caregiver. A better strategy is to introduce psychological causation as one of a number of possible explanations in the differential diagnoses that are discussed. Without directly linking psychotherapeutic intervention to the diagnosis, the patient can be offered a face-saving means by which the pathologic relationship with the health care system can be examined and alternative approaches to life stressors developed. Specific medical treatments also may be indicated and effective in treating some of the functional consequences of conversion disorder.

FEEDING AND EATING DISORDERS

CLINICAL MANIFESTATIONS

Feeding and eating disorders constitute a group of conditions in which there is a persistent disturbance of eating or associated behaviors that significantly impair an individual's physical health or psychosocial functioning. In DSM-5-TR, the described categories (with the exception of pica) are defined to be mutually exclusive in a given episode, based on the understanding that although they are phenotypically similar in some ways, they differ in course, prognosis, and effective treatment interventions.

PICA

Pica is diagnosed when the individual, aged >2 years, eats one or more nonnutritive, nonfood substances for a month or more and requires medical attention as a result. There is usually no specific aversion to food in general but a preferential choice to ingest substances such as clay, starch, soap, paper, or ash. The diagnosis requires the exclusion of specific culturally approved practices and has not been commonly found to be caused by a specific nutritional deficiency. Onset is most common in childhood, but the disorder can occur in association with other major psychiatric conditions in adults. An association with pregnancy has been observed, but the condition is only diagnosed when medical risks are increased by the behavior.

RUMINATION DISORDER

In this condition, individuals who have no demonstrable associated gastrointestinal or other medical condition repeatedly regurgitate their food after eating and then either rechew or swallow it or spit it out. The behavior typically occurs on a daily basis and must persist for at least 1 month. Weight loss and malnutrition are common sequelae, and individuals may attempt to conceal their behavior, either by covering their mouth or through social avoidance while eating. In infancy, the onset is typically between 3 and 12 months of age, and the behavior may remit spontaneously, although in some, it appears to be recurrent.

AVOIDANT/RESTRICTIVE FOOD INTAKE DISORDER

The cardinal feature of this disorder is avoidance or restriction of food intake, usually stemming from a lack of interest in or distaste of food and associated with weight loss, nutritional deficiency, dependency on nutritional supplementation, or marked impairment in psychosocial functioning, either alone or in combination. Culturally approved practices, such as fasting or a lack of available food, must be excluded as possible causes. The disorder is distinguished from anorexia nervosa by the presence of emotional factors, such as a fear of gaining weight and distortion of body image in the latter condition. Onset is usually in infancy or early childhood, but avoidant behaviors may persist into adulthood. The disorder is equally prevalent in males and females and is frequently comorbid with anxiety and cognitive and attention-deficit disorders and situations of familial stress. Developmental delay and functional deficits may be significant if the disorder is long-standing and unrecognized.

ANOREXIA NERVOSA

Individuals are diagnosed with anorexia nervosa if they restrict their caloric intake to a degree that their body weight deviates significantly from age, gender, health, and developmental norms and if they also exhibit a fear of gaining weight and an associated disturbance in body image. The condition is further characterized by differentiating those who achieve their weight loss predominantly through restricting intake or by excessive exercise (restricting type) from those who engage in recurrent binge eating and/or subsequent purging, self-induced vomiting, and usage of enemas, laxatives, or diuretics (binge-eating/purging type). Such subtyping is more state- than trait-specific, as individuals may transition from one profile to the other over time. Determination of whether an individual satisfies the primary criterion of significant low weight is complex and must be individualized, using all available historical information and comparison of body habitus to international body-mass norms and guidelines.

Individuals with anorexia nervosa frequently lack insight into their condition and are in denial about possible medical consequences; they often are not comforted by their achieved weight loss and persist in their behaviors despite having met previously self-designated weight goals. Alterations in the circuitry of reward sensitivity and executive function have been reported in anorexia, implicating disturbances in frontal cortex and anterior insula regulation of interoceptive awareness of satiety and hunger. Neurochemical findings, including the role of ghrelin, remain controversial.

Onset is most common in adolescence, although onset in later life can occur. Many more females than males are affected, with a lifetime prevalence in women of up to 4%. The disorder appears most prevalent in postindustrialized and urbanized countries and is frequently comorbid with preexisting anxiety disorders. The medical consequences of prolonged anorexia nervosa are multisystemic and can be life-threatening in severe presentations. Changes in laboratory values may be present, including leukopenia with lymphocytosis, elevations in blood urea nitrogen, and metabolic alkalosis and hypokalemia when purging is present. History and physical examination may reveal amenorrhea in females, skin abnormalities (petechiae, lanugo hair, dryness), and signs of hypometabolic function, including hypotension, hypothermia, and sinus bradycardia. Endocrine effects include hypogonadism, growth hormone resistance, and hypercortisolemia. Osteoporosis is a longer-term concern.

The course of the disorder is variable, with some individuals recovering after a single episode, while others exhibit recurrent episodes or a chronic course. Untreated anorexia has a mortality of 5.1/1000, the highest among psychiatric conditions. Maudsley Anorexia Nervosa Treatment for Adults (MANTRA) and eating disorder–focused cognitive-behavior therapy have proven to be effective therapies, with strict behavioral contingencies used when weight loss becomes critical. No pharmacologic intervention has proven to be specifically beneficial, but comorbid depression and anxiety should be treated. Weight gain should be undertaken gradually with a goal of 0.5–1 pound per week to prevent refeeding syndrome. Most individuals are able to achieve remission within 5 years of the original diagnosis.

BULIMIA NERVOSA

Bulimia nervosa describes individuals who engage in recurrent and frequent (at least once a week for 3 months) periods of binge eating and who then resort to compensatory behaviors, such as self-induced purging, enemas, use of laxatives, or excessive exercise, to avoid weight gain. Binge eating itself is defined as excessive food intake in a prescribed period of time, usually <2 h. As in anorexia nervosa, disturbances in body image occur and promote the behavior, but unlike in anorexia, individuals are of normal weight or even somewhat overweight. Subjects typically describe a loss of control and express shame about their actions, and often relate that their episodes are triggered by feelings of negative self-esteem or social stresses. The lifetime prevalence in women is ~2%, with a 10:1 female-to-male ratio. The disorder typically begins in adolescence and may be persistent over a number of years. Transition to anorexia occurs in only 10–15% of cases. Many of the medical risks associated with bulimia nervosa parallel those of anorexia nervosa and are a direct consequence of purging, including fluid and electrolyte disturbances and cardiac conduction abnormalities. Physical examination often results in no specific findings, but dental erosion and parotid gland enlargement may be present. Effective treatment approaches include SSRI antidepressants, usually in combination with cognitive-behavioral, emotion regulation, or interpersonal-based psychotherapies.

BINGE-EATING DISORDER

Binge-eating disorder is distinguished from bulimia nervosa by the absence of compensatory behaviors to prevent weight gain after an episode and by a lack of effort to restrict weight gain between episodes. Other features are similar, including distress over the behavior and the experience of loss of control, resulting in eating more rapidly or in greater amounts than intended or eating when not hungry. The 12-month prevalence in females is 1.6%, with a much lower female-to-male ratio than bulimia nervosa. Little is known about the course of the disorder, given its recent categorization, but its prognosis is markedly better than for other eating disorders, both in terms of its natural course and response to treatment. Transition to other eating disorder conditions is thought to be rare.

PERSONALITY DISORDERS

CLINICAL MANIFESTATIONS

Personality disorders are characteristic patterns of thinking, feeling, and interpersonal behavior that are relatively inflexible and cause significant functional impairment or subjective distress for the individual. The observed behaviors are not secondary to another mental disorder, nor are they precipitated by substance abuse or a general medical condition. This distinction is often difficult to make in clinical practice, because personality change may be the first sign of serious neurologic, endocrine, or other medical illness. Patients with frontal-lobe tumors, for example, can present with changes in motivation and personality while the results of the neurologic examination remain within normal limits. Individuals with personality disorders are often regarded as "difficult patients" in clinical medical practice because they are seen as excessively demanding and/or unwilling to follow recommended treatment plans. Although DSM-5-TR portrays personality disorders as qualitatively distinct categories, there is an alternative and emerging perspective that personality characteristics vary as a continuum between normal functioning and formal mental disorder, the essential features being moderate or greater impairment in self/interpersonal functioning and one or more pathological personality traits.

Personality disorders have been grouped into three overlapping clusters. *Cluster A* includes paranoid, schizoid, and schizotypal personality disorders. It includes individuals who are odd and eccentric and who maintain an emotional distance from others. Individuals have a restricted emotional range and remain socially isolated. Patients with schizotypal personality disorder frequently have unusual perceptual experiences and express magical beliefs about the external world. The essential feature of paranoid personality disorder is a pervasive mistrust and suspiciousness of others to an extent that is unjustified by available evidence. *Cluster B* disorders include antisocial, borderline, histrionic, and narcissistic types and describe individuals whose behavior is impulsive, excessively emotional, and erratic. *Cluster C* incorporates avoidant, dependent, and obsessive-compulsive personality types; enduring traits are anxiety and fear. The boundaries between cluster types are to some extent artificial, and many patients who meet criteria for one personality disorder also meet criteria for aspects of another. The risk of a comorbid major mental disorder is increased in patients who qualify for a diagnosis of personality disorder.

ETIOLOGY AND PATHOPHYSIOLOGY

Genetic studies have increasingly suggested a genetic contribution to the development of personality disorders. One study of 106,000 subjects identified nine loci significantly linked to aspects of neuroticism.

TREATMENT

Personality Disorders

Dialectical behavior therapy (DBT) is a cognitive-behavioral approach that focuses on behavioral change while providing acceptance, compassion, and validation of the patient. Several randomized trials have demonstrated the efficacy of DBT in the treatment of personality disorders. Antidepressant medications and low-dose antipsychotic drugs have some efficacy in cluster A personality disorders, whereas anticonvulsant mood-stabilizing agents and MAOIs may be considered for patients with cluster B diagnoses who show marked mood reactivity, behavioral dyscontrol, and/or rejection hypersensitivity. Anxious or fearful cluster C patients often respond to medications used for axis I anxiety disorders (see above). It is important that the physician and the patient have reasonable expectations vis-à-vis the possible benefit of any medication used and its side effects. Improvement may be subtle and observable only over time.

SCHIZOPHRENIA

CLINICAL MANIFESTATIONS

Schizophrenia is a heterogeneous syndrome characterized by perturbations of language, perception, thinking, social activity, affect, and volition. There are no pathognomonic features. The syndrome commonly begins in late adolescence, has an insidious (and less commonly, acute) onset, and, often, a poor outcome, progressing from social withdrawal and perceptual distortions to recurrent delusions and hallucinations. Patients may present with positive symptoms (such as conceptual disorganization, delusions, or hallucinations) or negative symptoms (loss of function, anhedonia, decreased emotional expression, impaired concentration, and diminished social engagement) and must have at least two of these for a 1-month period and continuous signs for at least 6 months to meet formal diagnostic criteria. Disorganized thinking or speech and grossly disorganized motor behavior, including catatonia, may also be present. As individuals age, positive psychotic symptoms tend to attenuate, and some measure of social and occupational function may be regained. "Negative" symptoms predominate in one-third of the schizophrenic population and are associated with a poor long-term outcome and a poor response to drug treatment. However, marked variability in the course and individual character of symptoms is typical.

The term *schizophreniform disorder* describes patients who meet the symptom requirements but not the duration requirements for schizophrenia, and *schizoaffective disorder* is used for those who manifest symptoms of schizophrenia and independent periods of mood disturbance. The terms *schizotypal* and *schizoid* refer to specific personality disorders and are discussed in that section. The diagnosis of delusional disorder is used for individuals who have delusions of various content for at least 1 month but who otherwise do not meet criteria for schizophrenia. Patients who experience a sudden onset of a brief (<1 month) alteration in thought processing, characterized by delusions, hallucinations, disorganized speech, or gross motor behavior, are most appropriately designated as having a brief psychotic disorder. Catatonia is recognized as a nonspecific syndrome that can occur as a consequence of other severe psychiatric/medical disorders and is diagnosed by the documentation of three or more of a cluster of motor and behavioral symptoms, including stupor, cataplexy, mutism, waxy flexibility, and stereotypy, among others. Prognosis depends not on symptom severity but on the response to antipsychotic medication. A permanent remission without recurrence does occasionally occur. About 10% of schizophrenic patients commit suicide.

Schizophrenia is present in 0.85% of individuals worldwide, with a lifetime prevalence of ~1–1.5%. An estimated 300,000 episodes of acute schizophrenia occur annually in the United States, resulting in direct and indirect costs of $155.7 billion.

DIFFERENTIAL DIAGNOSIS

The diagnosis is principally one of exclusion, requiring the absence of significant associated mood symptoms, any relevant medical condition, and substance abuse. Drug reactions that cause hallucinations, paranoia, confusion, or bizarre behavior may be dose-related or idiosyncratic; parkinsonian medications, clonidine, quinacrine, and procaine derivatives are the most common prescription medications associated with these symptoms. Drug causes should be ruled out in any case of newly emergent psychosis. The general neurologic examination in patients with schizophrenia is usually normal, but motor rigidity, tremor, and dyskinesias are noted in one-quarter of untreated patients.

EPIDEMIOLOGY AND PATHOPHYSIOLOGY

Epidemiologic surveys identify several risk factors for schizophrenia, including genetic susceptibility, early developmental insults, winter birth, and increasing parental age. Genetic factors are involved in at least a subset of individuals who develop schizophrenia. Schizophrenia is observed in ~6.6% of all first-degree relatives of an affected proband. If both parents are affected, the risk for offspring is 40%. The concordance rate for monozygotic twins is 50%, compared to 10% for dizygotic twins. Schizophrenia-prone families are also at risk for other psychiatric disorders, including schizoaffective disorder and *schizotypal* and *schizoid personality disorders*, the latter terms designating individuals who show a lifetime pattern of social and interpersonal deficits characterized by an inability to form close interpersonal relationships, eccentric behavior, and mild perceptual distortions. Large-scale GWASs have identified several hundred small effect risk loci and a few larger effect copy number variants, along with epigenetic effects, and have led to initial exploration in the clinical use of polygenic risk scores in diagnosis and prognosis. Pathways identified include ones involved in immunity, inflammation, and cell signaling. There is also recent evidence that brain gene expression in schizophrenia is similar to that seen in older aging adults without the disorder, implicating parallel mechanisms for cognitive deterioration.

TREATMENT

Schizophrenia

Antipsychotic agents **(Table 463-10)** are the cornerstone of acute and maintenance treatment of schizophrenia and are effective in the treatment of hallucinations, delusions, and thought disorders, regardless of etiology. The mechanism of action involves, at least in part, binding to dopamine D_2/D_3 receptors in the ventral striatum; the clinical potencies of traditional antipsychotic drugs parallel their affinities for the D_2 receptor, and even the newer "atypical" agents exert some degree of D_2 receptor blockade. All current neuroleptics induce expression of the immediate-early gene c-*fos* in the nucleus accumbens, a dopaminergic site connecting prefrontal and limbic cortices. The development of newer atypical neuroleptics, however, is increasingly focusing on different targets: D_3, 5-HT_{1A}, 5-HT_7, mGlu2/3, and muscarinic acetylcholine (M1, M4) receptors; α_1- and α_2-noradrenergic activity; and altering the relationship between 5-HT_2 and D_2 receptor activity, resulting in faster dissociation of D_2 binding and effects on neuroplasticity.

TABLE 463-10 Antipsychotic Agents

NAME	USUAL PO DAILY DOSE (mg)	SIDE EFFECTS	SEDATION	COMMENTS
First-Generation Antipsychotics				
Low potency				
Chlorpromazine (Thorazine)	100–1000	Anticholinergic effects; orthostasis; photosensitivity; cholestasis; QT prolongation	+++	EPSEs usually not prominent; can cause anticholinergic delirium in elderly patients
Thioridazine (Mellaril)	100–600			
Midpotency				
Trifluoperazine (Stelazine)	2–50	Fewer anticholinergic side effects	++	Well tolerated by most patients
Perphenazine (Trilafon)	4–64	Fewer EPSEs than with higher-potency agents	++	
Loxapine (Loxitane)	30–100	Frequent EPSEs	++	
Molindone (Moban)	30–100	Frequent EPSEs	0	Little weight gain
High potency				
Haloperidol (Haldol)	5–20	No anticholinergic side effects; EPSEs often prominent	0/+	Often prescribed in doses that are too high; long-acting injectable forms of haloperidol and fluphenazine available
Fluphenazine (Prolixin)	1–20	Frequent EPSEs	0/+	
Thiothixene (Navane)	2–50	Frequent EPSEs	0/+	
Second-Generation Antipsychotics				
Clozapine (Clozaril)	150–600	Agranulocytosis (1%); weight gain; seizures; drooling; hyperthermia	+ +	Requires weekly WBC count for first 6 months, then biweekly if stable
Risperidone (Risperdal)	2–8	Orthostasis	+	Requires slow titration; EPSEs observed with doses >6 mg qd
Olanzapine (Zyprexa)	10–30	Weight gain	++	Mild prolactin elevation
Quetiapine (Seroquel)	350–800	Sedation; weight gain; anxiety	+++	bid dosing
Ziprasidone (Geodon)	120–200	Orthostatic hypotension	+/++	Minimal weight gain; increases QT interval
Aripiprazole (Abilify)	10–30	Nausea, anxiety, insomnia	0/+	Mixed agonist/antagonist; extended-release available
Paliperidone (Invega)	3–12	Restlessness, EPSEs, increased prolactin, headache	+	Active metabolite of risperidone
Iloperidone (Fanapt)	12–24	Dizziness, hypotension	0/+	Requires dose titration; long-acting injectable available
Asenapine (Saphris)	10–20	Dizziness, anxiety, EPSEs, minimal weight gain	++	Sublingual tablets; bid dosing
Lurasidone (Latuda)	40–80	Nausea, EPSEs	++	Uses CYP3A4
Brexpiprazole (Rexulti)	1–4	Anxiety, dizziness, fatigue	++	CYP3A4 and 2D6 interactions
Pimavanserin (Nuplazid)	34	Edema, confusion, sedation	++	Approved for Parkinson's disease psychosis
Cariprazine (Vraylar)	1.5–6	EPSEs, vomiting	++	Preferential D_3 receptor affinity
Lumateperone (Caplyta)	42	Fatigue, dry mouth; no apparent metabolic/motor effects	++	5-HT$_A$ > D_2 receptor affinity

Abbreviations: EPSEs, extrapyramidal side effects; WBC, white blood cell.

Conventional neuroleptics differ in their potency and side effect profile. Older agents, such as chlorpromazine and thioridazine, are more sedating and anticholinergic and more likely to cause orthostatic hypotension, whereas higher-potency antipsychotics, such as haloperidol, perphenazine, and thiothixene, are more likely to induce extrapyramidal side effects. The model "atypical" antipsychotic agent is clozapine, a dibenzodiazepine that has a greater potency in blocking the 5-HT$_2$ than the D_2 receptor and a much higher affinity for the D_4 than the D_2 receptor. Its principal disadvantage is a risk of blood dyscrasias. Paliperidone is a metabolite of risperidone and shares many of its properties. Unlike other antipsychotics, clozapine does not cause a rise in prolactin levels. Approximately 30% of patients who do not benefit from conventional antipsychotic agents will have a better response to this drug, which also has a demonstrated superiority to other antipsychotic agents in preventing suicide; however, its side effect profile makes it most appropriate for treatment-resistant cases. Risperidone, a benzisoxazole derivative, is more potent at 5-HT$_2$ than D_2 receptor sites, like clozapine, but it also exerts significant α_2 antagonism, a property that may contribute to its perceived ability to improve mood and increase motor activity. Risperidone is not as effective as clozapine in treatment-resistant cases but does not carry a risk of blood dyscrasias. Olanzapine is similar neurochemically to clozapine but has a significant risk of inducing weight gain. Quetiapine is distinct in having a weak D_2 effect but potent α_1 and histamine blockade. Ziprasidone causes minimal weight gain and is unlikely to increase prolactin but may increase QT prolongation. Aripiprazole also has little risk of weight gain or prolactin increase but may increase anxiety, nausea, and insomnia as a result of its partial agonist properties. Asenapine is associated with minimal weight gain and anticholinergic effect but may have a higher than expected risk of extrapyramidal symptoms (EPSs). Cariprazine, a D_2/D_3 partial agonist, has no QT or prolactin elevation risk but can result in EPS as well.

Antipsychotic agents are effective in 70% of patients presenting with a first episode. Improvement may be observed within hours or days, but full remission usually requires 6–8 weeks. The choice of agent depends principally on the side effect profile and cost of treatment or on a past personal or family history of a favorable response to the drug in question. Atypical agents appear to be more effective in treating negative symptoms and improving cognitive function. An equivalent treatment response can usually be achieved with relatively low doses of any drug selected (i.e., 4–6 mg/d of haloperidol, 10–15 mg of olanzapine, or 4–6 mg/d of risperidone). Doses in this range result in >80% D_2 receptor blockade, and there is little evidence that higher doses increase either the rapidity or degree of response. Maintenance treatment requires careful attention to the possibility of relapse and monitoring for the development of a movement disorder. Intermittent drug treatment is less effective than regular dosing, but gradual dose reduction is likely to improve social functioning in many schizophrenic patients who have been maintained at high doses. If medications are completely discontinued, however, the relapse rate is 60% within 6 months. Long-acting injectable (LAI) preparations (risperidone, paliperidone, olanzapine, aripiprazole) are considered when noncompliance with oral therapy leads to relapses but should not be considered interchangeable because the agents differ in their indications, injection intervals and sites/volumes, and possible adverse reactions, among other factors. Extended treatment studies indicate a significant decrease in relapse with LAI usage. In treatment-resistant patients, a transition to clozapine usually results in rapid improvement, but a prolonged delay in response in some cases necessitates a 6- to 9-month trial for maximal benefit to occur.

Antipsychotic medications can cause a broad range of side effects, including lethargy, weight gain, postural hypotension, constipation, and dry mouth. Extrapyramidal symptoms such as dystonia, akathisia, and akinesia are also frequent with first-generation agents and may contribute to poor adherence if not specifically addressed. Anticholinergic and parkinsonian symptoms respond well to trihexyphenidyl, 2 mg bid, or benztropine mesylate, 1–2 mg bid. Akathisia may respond to beta blockers. In rare cases, more serious and occasionally life-threatening side effects may emerge, including hyperprolactinemia, ventricular arrhythmias, gastrointestinal obstruction, retinal pigmentation, obstructive jaundice, and neuroleptic malignant syndrome (characterized by hyperthermia, autonomic dysfunction, muscular rigidity, and elevated creatine phosphokinase levels). The most serious adverse effects of clozapine are agranulocytosis, which has an incidence of 1%, and induction of seizures, which has an incidence of 10%. Weekly white blood cell counts are required, particularly during the first 3 months of treatment.

The risk of type 2 diabetes mellitus appears to be increased in schizophrenia, and second-generation agents as a group, with the exception of lumateperone, produce greater adverse effects on glucose regulation, independent of effects on obesity, than traditional agents. Clozapine, olanzapine, and quetiapine seem more likely to cause hyperglycemia, weight gain, and hypertriglyceridemia than other atypical antipsychotic drugs. Close monitoring of plasma glucose and lipid levels is indicated with the use of these agents.

A serious side effect of long-term use of first-generation and, to a lesser extent, second-generation antipsychotic agents is tardive dyskinesia, characterized by repetitive, involuntary, and potentially irreversible movements of the tongue and lips (bucco-linguo-masticatory triad) and, in approximately half of cases, choreoathetosis. Tardive dyskinesia has an incidence of 2–4% per year of exposure and a prevalence of 20% in chronically treated patients. The prevalence increases with age, total dose, and duration of drug administration and may involve formation of free radicals and perhaps mitochondrial energy failure. Valbenazine (Ingrezza), a vesicular monoamine transporter 2 inhibitor that depletes presynaptic dopamine, is approved for treatment of tardive dyskinesia.

The CATIE study, a large-scale investigation of the effectiveness of antipsychotic agents in "real-world" patients, revealed a high rate of discontinuation of treatment after >18 months. Olanzapine showed greater effectiveness than quetiapine, risperidone, perphenazine, or ziprasidone but also a higher discontinuation rate due to weight gain and metabolic effects. Surprisingly, perphenazine, a first-generation agent, showed little evidence of inferiority to newer drugs.

Drug treatment of schizophrenia is by itself insufficient. Educational efforts directed toward families and relevant community resources have proved to be necessary to maintain stability and optimize outcome. A collaborative treatment model using social cognition interventions and involving a multidisciplinary case-management team that seeks out and closely follows the patient in the community has proved particularly effective. Attempts to prevent schizophrenia through early identification and treatment (both psychosocial and psychopharmacologic) of high-risk children and adolescents are currently being evaluated.

ASSESSMENT AND EVALUATION OF VIOLENCE

Primary care physicians may encounter situations in which family, domestic, or societal violence is discovered or suspected. Such an awareness can carry legal and moral obligations; many state laws mandate reporting of child, spousal, and elder abuse. Physicians are frequently the first point of contact for both victim and abuser. Approximately 2 million older Americans and 1.5 million U.S. children are thought to experience some form of physical maltreatment each year. Spousal abuse is believed to be even more prevalent. An interview study of 24,000 women in 10 countries found a lifetime prevalence of physical or sexual violence that ranged from 15–71%; these individuals are more likely to suffer from depression, anxiety, and substance abuse and to have attempted suicide. In addition, abused individuals frequently express low self-esteem, vague somatic symptomatology, social isolation, and a passive feeling of loss of control. Although it is essential to treat these elements in the victim, the first obligation is to ensure that the perpetrator has taken responsibility for preventing any further violence. Substance abuse and/or dependence and serious mental illness in the abuser may contribute to the risk of harm and require direct intervention. Depending on the situation, law enforcement agencies, community resources such as support groups and shelters, and individual and family counseling can be appropriate components of a treatment plan. A safety plan should be formulated with the victim, in addition to providing information about abuse, its likelihood of recurrence, and its tendency to increase in severity and frequency. Antianxiety and antidepressant medications may sometimes be useful in treating the acute symptoms, but only if independent evidence for an appropriate psychiatric diagnosis exists.

■ FURTHER READING

Alon N et al: Social determinants of mental health in major depressive disorder: Umbrella review of 26 meta-analyses and systemic reviews. Psychiatry Res 335:115854, 2024.

Anderson E et al: Depression treatment options and managing depression in primary care. N Eng J Med 390:e44, 2024.

Barry R et al: Prevalence of mental health disorders among individuals experiencing homelessness: A systematic review and meta-analysis. JAMA Psychiatry 81:691, 2024.

Bertolini F et al: Early pharmacological interventions for prevention of post-traumatic stress disorder (PTSD) in individuals experiencing acute traumatic stress symptoms. Cochrane Database Sys Rev 5:CD013613, 2024.

Birnbaum R, Weinberger DR: The genesis of schizophrenia: An origin story. Am J Psychiatry 181:482, 2024.

Borque VR et al: Genetic and phenotypic similarities across major psychiatric disorders: A systematic review and quantitative assessment. Trans Psychiatry 14:171, 2024.

Cristancho M et al: Depression-advanced treatments for treatment resistant depression. N Eng J Med 390:e44, 2024.

Gargano SP et al: A closer look to neural pathways and psychopharmacology of obsessive compulsive disorder. Front Behav Neurosci 17:1282246, 2023.

Guaiana G et al: Pharmacological treatment in panic disorder in adults: A network meta-analysis. Cochrane Database Sys Rev 11:1465, 2023.

Nestler EJ, Russo SJ: Neurobiological basis of stress resilience. Neuron 112:1911, 2024.

Park JH et al: Global perspectives on bipolar disorder treatments: In depth comparative analysis of international guidelines for medication selection. BJPsych Open 10:e75, 2024.

Reilly S et al: Collaborative care approaches for people with severe mental illness. Cochrane Database Sys Rev 5:CD009531, 2024.

Solmi M et al: An umbrella review of predictors of response, remission, recovery and relapse across mental disorders. Mol Psychiatry 28:3671, 2023.

Szuhany KL, Simon NM: Anxiety disorders: A review. JAMA 328:2431, 2022.

464 Alcohol and Alcohol Use Disorders

Marc A. Schuckit

Most patients drink alcohol, including many who take this drug at levels that can adversely affect their medical conditions or interfere with the effects of prescribed medications. Therefore, it is important to note that this chapter presents information relevant to all patients, not just those with alcohol problems. Alcohol (beverage ethanol) has diverse and widespread effects on the body and impacts directly or indirectly on almost every neurochemical system in the brain. At even relatively low doses, this drug can exacerbate most medical problems and affect medications metabolized in the liver, and at higher doses, it can temporarily mimic many medical (e.g., diabetes) and psychiatric (e.g., depression) conditions. Frequent and heavier drinking is also associated with the treatable but life-threatening condition of alcohol use disorder (the modern term for alcoholism). Physicians from all specialties play an important role in *screening*, using *brief interventions*, and treating or *referring for treatment* individuals with repetitive alcohol problems, a process abbreviated as SBIRT.

The lifetime risk for repetitive serious alcohol problems (e.g., alcohol use disorder) in patients is at least 20% for men and 10% for women, regardless of a person's education or income, and U.S. yearly costs for these disorders exceed $249 billion. Although low doses of alcohol might have healthful benefits, drinking more than three standard drinks per day enhances the risk for cancer and vascular disease, and alcohol use disorders decrease the life span by ~10 years. Unfortunately, most clinicians have had only limited training in identifying and treating alcohol-related disorders.

PHARMACOLOGY AND NUTRITIONAL IMPACT OF ETHANOL

Ethanol blood levels are expressed as milligrams or grams of ethanol per deciliter (e.g., 100 mg/dL = 0.10 g/dL), with values of ~0.02 g/dL resulting from the ingestion of one typical drink. In round figures, a standard drink is 10–12 g of ethanol, as seen in 340 mL (12 oz) of beer, 115 mL (4 oz) of nonfortified wine, and 43 mL (1.5 oz) (a shot) of 80-proof (40% ethanol by volume) beverage (e.g., whisky); 0.5 L (1 pint) of 80-proof beverage contains ~160 g of ethanol (~16 standard drinks), and 750 mL of wine contains ~60 g of ethanol. These beverages also have additional components (*congeners*) that affect the drink's taste and might contribute to adverse effects on the body. Congeners include methanol, butanol, acetaldehyde, histamine, tannins, iron, and lead. As a depressant drug, alcohol acutely decreases neuronal activity and has similar behavioral effects and cross-tolerance with other depressants, including benzodiazepines, barbiturates, and some anticonvulsants.

Alcohol is absorbed from mucous membranes of the mouth and esophagus (in small amounts), from the stomach and large bowel (in modest amounts), and from the proximal portion of the small intestine (the major site). The rate of absorption is increased by rapid gastric emptying (as seen with carbonated beverages); by the absence of proteins, fats, or carbohydrates (which interfere with absorption); and by dilution to a modest percentage of ethanol (maximum at ~20% by volume).

Between 2% (at low blood alcohol concentrations) and 10% (at high blood alcohol concentrations) of ethanol is excreted directly through the lungs, urine, or sweat, but most is metabolized to acetaldehyde, primarily in the liver. The most important pathway occurs in the cell cytosol where alcohol dehydrogenase (ADH) produces acetaldehyde, which is then rapidly destroyed by aldehyde dehydrogenase (ALDH) in the cytosol and mitochondria **(Fig. 464-1)**. A second pathway occurs in the microsomes of the smooth endoplasmic reticulum (the microsomal ethanol-oxidizing system [MEOS]) that is responsible for ≥10% of ethanol oxidation at high blood alcohol concentrations.

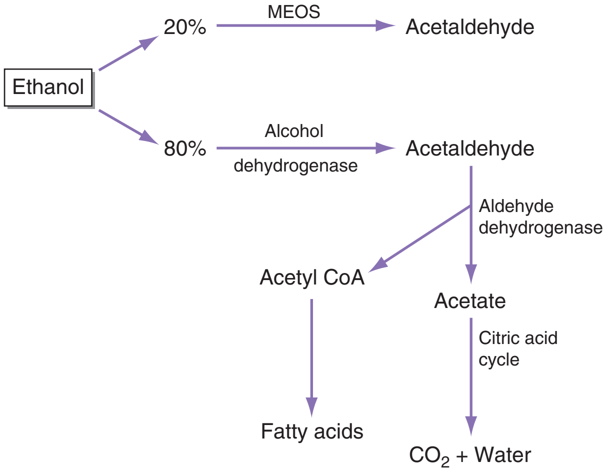

FIGURE 464-1 The metabolism of alcohol. CoA, coenzyme A; MEOS, microsomal ethanol oxidizing system.

Although a standard drink contains ~300 kJ, or 70–100 kcal, these are devoid of minerals, proteins, and vitamins. In addition, alcohol interferes with absorption of vitamins in the small intestine and decreases their storage in the liver with modest effects on folate (folacin or folic acid), pyridoxine (B$_6$), thiamine (B$_1$), nicotinic acid (niacin, B$_3$), and vitamin A.

Heavy drinking in a fasting, healthy individual can produce transient hypoglycemia within 6–36 h, secondary to the acute actions of ethanol that decrease gluconeogenesis. This can result in temporary abnormal glucose tolerance tests (with a resulting erroneous diagnosis of diabetes mellitus) until the heavy drinker has abstained for 2–4 weeks. Alcohol ketoacidosis, probably reflecting a decrease in fatty acid oxidation coupled with poor diet or persistent vomiting, can be misdiagnosed as diabetic ketosis. With alcohol-related ketoacidosis, patients show an increase in serum ketones along with a mild increase in glucose but a large anion gap, a mild to moderate increase in serum lactate, and a β-hydroxybutyrate/lactate ratio of between 2:1 and 9:1 (with normal being 1:1).

In the brain, alcohol affects almost all neurotransmitter systems, with acute effects that are often the opposite of those seen following desistance after a period of heavy drinking. The most prominent acute actions relate to boosting γ-aminobutyric acid (GABA) activity, especially at GABA$_A$ receptors. Enhancement of this complex chloride channel system contributes to anticonvulsant, sleep-inducing, antianxiety, and muscle relaxation effects of all GABA-boosting drugs. Acutely administered alcohol produces a release of GABA, and continued use increases density of GABA$_A$ receptors, whereas alcohol withdrawal states are characterized by decreases in GABA-related activity. Equally important is the ability of acute alcohol to inhibit postsynaptic *N*-methyl-D-aspartate (NMDA) excitatory glutamate receptors, whereas chronic drinking and desistance are associated with an upregulation of these excitatory receptor subunits. The relationships between greater GABA and diminished NMDA receptor activity during acute intoxication and diminished GABA with enhanced NMDA actions during alcohol withdrawal explain much of intoxication and withdrawal phenomena.

As with all pleasurable activities, alcohol acutely increases dopamine levels in the ventral tegmentum and related brain regions, and this effect plays an important role in continued alcohol use, craving, and relapse. The changes in dopamine pathways are also linked to increases in "stress hormones," including cortisol and adrenocorticotropic hormone (ACTH), during intoxication and in the context of the stresses of withdrawal. Such alterations are likely to contribute to both feelings of reward during intoxication and depression during falling blood alcohol concentrations. Also closely linked to alterations in dopamine (especially in the nucleus accumbens) are alcohol-induced changes in opioid receptors, with acute alcohol causing release of β-endorphins.

Additional neurochemical changes include increases in synaptic levels of serotonin during acute intoxication and subsequent upregulation of serotonin receptors. Acute increases in nicotinic acetylcholine

TABLE 464-1 Effects of Blood Alcohol Levels in the Absence of Tolerance

BLOOD LEVEL, g/dL	USUAL EFFECT
0.02	Decreased inhibitions, a slight feeling of intoxication
0.08	Decrease in complex cognitive functions and motor performance
0.20	Obvious slurred speech, motor incoordination, irritability, and poor judgment
0.30	Light coma and depressed vital signs
0.40	Death

systems contribute to the impact of alcohol in the ventral tegmental region, which occurs in concert with enhanced dopamine activity. In the same regions, alcohol impacts on cannabinol receptors, with resulting release of dopamine, GABA, and glutamate as well as subsequent effects on brain reward circuits.

BEHAVIORAL EFFECTS, TOLERANCE, AND WITHDRAWAL

The acute effects of a drug depend on the dose, the rate of increase in plasma, the concomitant presence of other drugs, and past experience with the agent. "Legal intoxication" with alcohol in most states is based on a blood alcohol concentration of 0.08 g/dL, some states are considering lowering acceptable levels to <0.05 g/dL, and levels of 0.04 g/dL are cited for pilots in the United States and automobile drivers in some other countries. However, behavioral, psychomotor, and cognitive changes are seen at 0.02–0.04 g/dL (i.e., after one to two drinks) **(Table 464-1)**. Deep but disturbed sleep can be seen at 0.15 g/dL in individuals who have not developed tolerance, and death can occur with levels between 0.30 and 0.40 g/dL. Beverage alcohol is probably responsible for more overdose deaths than any other drug.

Repeated use of alcohol contributes to the need for a greater number of standard drinks to produce effects originally observed with fewer drinks (acquired tolerance), a phenomenon involving at least three compensatory mechanisms. (1) After 1–2 weeks of daily drinking, *metabolic* or *pharmacokinetic tolerance* can be seen, with up to 30% increases in the rate of hepatic ethanol metabolism. This alteration disappears almost as rapidly as it develops. (2) *Cellular or pharmacodynamic tolerance* develops through neurochemical changes that maintain relatively normal physiologic functioning despite the presence of alcohol. Subsequent decreases in blood levels contribute to symptoms of withdrawal. (3) Individuals learn to adapt their behavior so that they can function better than expected under the influence of the drug (*learned* or *behavioral tolerance*).

The cellular changes caused by chronic ethanol exposure may not resolve for several weeks or longer following cessation of drinking. Rapid decreases in blood alcohol levels before that time can produce a withdrawal syndrome, which is most intense during the first 5 days, but with some symptoms (e.g., disturbed sleep and anxiety) lasting up to 4–6 months as part of a "protracted withdrawal" syndrome.

THE EFFECTS OF ETHANOL ON ORGAN SYSTEMS

Relatively low doses of alcohol (one or two drinks per day) may have mild potential beneficial effects by, for example, decreasing aggregation of platelets and potentially decreasing the risk for vascular dementia and Alzheimer's disease. However, any potential healthful effects disappear with the regular consumption of three or more drinks per day, and knowledge about the deleterious effects of alcohol can both help the physician to identify patients with alcohol use disorders and supply them with information that might help motivate changes in behavior.

NERVOUS SYSTEM

Approximately 35% of drinkers overall, including as many as 50% of drinking college students and a much higher proportion of individuals with alcohol use disorders, ever experience a *blackout*. This is an episode of temporary anterograde amnesia, in which the person was awake but forgot all (en bloc blackouts at blood alcohol levels >0.20 mg/dL) or part (fragmentary blackouts at >0.12 mg/dL) of what occurred during a drinking period.

Another common problem, one seen after as few as one or two drinks shortly before bedtime, is disturbed sleep. Although alcohol might initially help a person fall asleep, it disrupts sleep throughout the rest of the night. The stages of sleep are altered, and times spent in rapid eye movement (REM) and deep sleep early in the night are reduced. Alcohol relaxes muscles in the pharynx, which can cause snoring and exacerbate sleep apnea; symptoms of the latter occur in 75% of men with alcohol use disorders aged ≥60 years. Patients may also experience prominent and sometimes disturbing dreams later in the night. All these sleep impairments can contribute to relapses to drinking in persons with alcohol use disorders.

Other common consequences of alcohol use even at relatively low alcohol levels are impaired judgment and coordination, which increase the risk of injuries. In the United States, ~40% of drinkers have at some time driven while intoxicated. Heavy drinking can also be associated with headache, thirst, nausea, vomiting, and fatigue the following day, a hangover syndrome that is responsible for much missed work and school time and temporary cognitive deficits.

Chronic high alcohol doses cause *peripheral neuropathy* in ~10% of individuals with alcohol use disorders. Similar to diabetes, patients experience bilateral limb numbness, tingling, and paresthesias, all of which are more pronounced distally. Approximately 1% of those with alcohol use disorders develop *cerebellar degeneration* or *atrophy*, producing a syndrome of progressive unsteady stance and gait often accompanied by mild nystagmus. Perhaps 1 in 500 individuals with alcohol use disorders develop full *Wernicke's* (ophthalmoparesis, ataxia, and encephalopathy) and *Korsakoff's* (severe retrograde and anterograde amnesia) *syndromes*. These result from low levels of thiamine, especially in predisposed individuals with transketolase deficiencies. Repeated heavy drinking can contribute to *cognitive problems* and temporary memory impairment lasting for weeks to months after abstinence. Brain ventricular enlargement and widened cortical sulci on magnetic resonance imaging (MRI) and computed tomography (CT) scans occurs in ~50% of individuals with long-term alcohol use disorders; these changes are usually reversible if abstinence is maintained. Adolescents may be especially vulnerable to alcohol-related brain changes, as indicated by preclinical studies and prospective investigations in humans suggesting that alcohol exposure in the developing brain may adversely impact future cognitive processes related to cognition, reward recognition, and cue processing. There is no single "alcoholic dementia" syndrome; rather, this label describes patients who have irreversible cognitive changes (possibly from diverse causes) in the context of chronic alcohol use disorders.

Psychiatric Comorbidity Alcohol temporarily alters brain neurochemistry in a manner similar to ways observed in some psychiatric conditions, resulting in mood, anxiety, and psychotic disorders. However, those alcohol-induced psychiatric symptoms that are only observed during intense intoxication or withdrawal syndromes are likely to disappear within days to weeks of abstinence. For example, while about 40% of individuals with alcohol use disorder will at some point meet criteria for a major depressive episode, about half of those conditions are temporary substance-induced mood disorders that are likely to disappear within a month of abstinence without the use of antidepressant medications. In addition, several preexisting psychiatric disorders increase the risk for future alcohol use disorder including schizophrenia, manic-depressive disease, posttraumatic stress disorder, and anxiety syndromes such as panic disorder **(Chap. 463)**. The comorbidities of alcohol use disorders with independent psychiatric disorders might represent an overlap in genetic vulnerabilities, impaired judgment regarding the use of alcohol as a consequence of the independent psychiatric condition, or an attempt to use alcohol to alleviate symptoms of the disorder or side effects of medications.

Treatment of all forms of alcohol-induced psychopathology includes helping patients achieve abstinence and offering supportive care, as well as reassurance and "talk therapy" such as cognitive-behavioral approaches. However, with the exception of short-term antipsychotic medications for substance-induced psychoses, substance-induced psychiatric conditions only rarely require medications. Recovery is likely within several days to 4 weeks of abstinence. Conversely, because alcohol-induced conditions are temporary and do not indicate a need for long-term pharmacotherapy, a history of heavy alcohol intake is an important part of the workup for any patient who presents with any of these psychiatric syndromes.

THE GASTROINTESTINAL SYSTEM

Esophagus and Stomach Alcohol can cause inflammation of the esophagus and stomach causing epigastric distress and gastrointestinal bleeding, making alcohol one of the most common causes of hemorrhagic gastritis. Violent vomiting can produce severe bleeding through a Mallory-Weiss lesion, a longitudinal tear in the mucosa at the gastroesophageal junction.

Pancreas and Liver The incidence of acute pancreatitis (~25 per 1000 per year) is almost threefold higher in individuals with alcohol use disorders than in the general population, accounting for an estimated 10% or more of the total cases. Alcohol impairs gluconeogenesis in the liver, resulting in a fall in the amount of glucose produced from glycogen, increased lactate production, and decreased oxidation of fatty acids. These contribute to an increase in fat accumulation in liver cells. In healthy individuals, these changes are reversible, but with repeated exposure to ethanol, especially daily heavy drinking, more severe changes in the liver occur, including alcohol-induced hepatitis, perivenular sclerosis, and cirrhosis, with the latter observed in an estimated 15% of individuals with alcohol use disorders **(Chap. 353)**. Perhaps through an enhanced vulnerability to infections, individuals with alcohol use disorders have an elevated rate of hepatitis C, and drinking in the context of that disease is associated with more severe liver deterioration.

CANCER

As few as 1.5 drinks per day increases a woman's risk of breast cancer 1.4-fold. For both sexes, four drinks per day increases the risk for oral and esophageal cancers approximately threefold and rectal cancers by a factor of 1.5; seven to eight or more drinks per day produces an approximately fivefold increased risk for many other cancers. These consequences may result directly from cancer-promoting effects of alcohol and acetaldehyde or indirectly by interfering with immune homeostasis.

HEMATOPOIETIC SYSTEM

Ethanol causes an increase in red blood cell size (mean corpuscular volume [MCV]), which reflects its effects on stem cells. If heavy drinking is accompanied by folic acid deficiency, there can also be hypersegmented neutrophils, reticulocytopenia, and a hyperplastic bone marrow; if malnutrition is present, sideroblastic changes can be observed. Chronic heavy drinking can decrease production of white blood cells, decrease granulocyte mobility and adherence, and impair delayed-hypersensitivity responses to novel antigens (with a possible false-negative tuberculin skin test). Associated immune deficiencies can contribute to vulnerability toward infections, including hepatitis and HIV, and interfere with their treatment. Finally, many individuals with alcohol use disorders have mild thrombocytopenia, which usually resolves within a week of abstinence unless there is hepatic cirrhosis or congestive splenomegaly.

CARDIOVASCULAR SYSTEM

Acutely, ethanol decreases myocardial contractility and causes peripheral vasodilation, with a resulting mild decrease in blood pressure and a compensatory increase in cardiac output. Exercise-induced increases in cardiac oxygen consumption are higher after alcohol intake. These acute effects have little clinical significance for the average healthy drinker but can be problematic when persisting cardiac disease is present.

The consumption of three or more drinks per day results in a dose-dependent increase in blood pressure, which returns to normal within weeks of abstinence. Thus, heavy drinking is an important factor in mild to moderate hypertension. Chronic heavy drinkers also have a sixfold increased risk for coronary artery disease, related, in part, to increased low-density lipoprotein cholesterol, and carry an increased risk for cardiomyopathy through direct effects of alcohol on heart muscle. Symptoms of the latter include unexplained arrhythmias in the presence of left ventricular impairment, heart failure, hypocontractility of heart muscle, and dilation of all four heart chambers with associated potential mural thrombi and mitral valve regurgitation. Atrial or ventricular arrhythmias, especially paroxysmal tachycardia, can also occur temporarily after heavy drinking in individuals showing no other evidence of heart disease—a syndrome known as the "holiday heart."

GENITOURINARY SYSTEM CHANGES, SEXUAL FUNCTIONING, AND FETAL DEVELOPMENT

Heavy drinking in adolescence can affect normal sexual development and reproductive onset. At any age, modest ethanol doses (e.g., blood alcohol concentrations of 0.06 g/dL) can increase sexual drive but also decrease erectile capacity in men. Even in the absence of liver impairment, a significant minority of chronic heavy-drinking men show irreversible testicular atrophy with shrinkage of the seminiferous tubules, decreases in ejaculate volume, and a lower sperm count **(Chap. 403)**.

The repeated ingestion of high doses of ethanol by women can result in amenorrhea, a decrease in ovarian size, absence of corpora lutea with associated infertility, and an increased risk of spontaneous abortion. Drinking during pregnancy results in the rapid placental transfer of both ethanol and acetaldehyde, which may contribute to a range of consequences known as fetal alcohol spectrum disorder (FASD). One severe result is the *fetal alcohol syndrome* (FAS), seen in ~5% of children born to heavy-drinking mothers, which can include any of the following: facial changes with epicanthal eye folds; poorly formed ear concha; small teeth with faulty enamel; cardiac atrial or ventricular septal defects; an aberrant palmar crease and limitation in joint movement; and microcephaly with intellectual impairment. Less pervasive FASD conditions include combinations of low birth weight, a lower intelligence quotient (IQ), hyperactive behavior, and some modest cognitive deficits. The amount of ethanol required and the time of vulnerability during pregnancy have not been defined, making it advisable for pregnant women to abstain from alcohol completely.

OTHER EFFECTS

Between one-half and two-thirds of individuals with alcohol use disorders have skeletal muscle weakness caused by acute *alcoholic myopathy*, a condition that improves but that might not fully remit with abstinence. Effects of repeated heavy drinking on the *skeletal system* include changes in calcium metabolism, lower bone density, and decreased growth in the epiphyses, leading to an increased risk for fractures and osteonecrosis of the femoral head. *Hormonal changes* include an increase in cortisol levels, which can remain elevated during heavy drinking; inhibition of vasopressin secretion at rising blood alcohol concentrations and enhanced secretion at falling blood alcohol concentrations (with the final result that most individuals with alcohol use disorders are likely to be slightly overhydrated); a modest and reversible decrease in serum thyroxine (T_4); and a more marked decrease in serum triiodothyronine (T_3). Hormone irregularities may disappear after a month or more of abstinence.

ALCOHOL USE DISORDERS

Because many drinkers occasionally imbibe to excess, temporary alcohol-related problems are common, especially in the late teens to the late twenties. However, repeated problems in multiple life areas can indicate an alcohol use disorder as defined in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5).

DEFINITIONS AND EPIDEMIOLOGY

An *alcohol use disorder* (also called *alcoholism* or *alcohol dependence* in prior diagnostic manuals) is defined in DSM-5 of the American Psychiatric Association as repeated alcohol-related difficulties in at least 2 of 11 life areas that cluster together in the same 12-month period **(Table 464-2)**. Ten of the 11 items in DSM-5 (published in 2013) were

TABLE 464-2 ***Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Classification of Alcohol Use Disorder (AUD)**

Criteria
Two or more of the following items occurring in the same 12-month period must be endorsed for the diagnosis of an alcohol use disorder[a]:
Drinking resulting in recurrent failure to fulfill role obligations
Recurrent drinking in hazardous situations
Continued drinking despite alcohol-related social or interpersonal problems
Tolerance
Withdrawal, or substance use for relief/avoidance of withdrawal
Drinking in larger amounts or for longer than intended
Persistent desire/unsuccessful attempts to stop or reduce drinking
Great deal of time spent obtaining, using, or recovering from alcohol
Important activities given up/reduced because of drinking
Continued drinking despite knowledge of physical or psychological problems caused by alcohol
Alcohol craving

[a] *Mild AUD:* 2–3 criteria required; *moderate AUD*: 4–5 items endorsed; *severe AUD:* 6 or more items endorsed.

taken directly from the dependence and abuse criteria in DSM-IV, after deleting legal problems and adding craving. Thus, diagnoses established across the two systems agree at >.84. Severity of DSM-5 alcohol use disorder is based on the number of items endorsed: mild is two or three items; moderate is four or five; and severe is six or more of the 11 criterion items.

The lifetime risk for an alcohol use disorder in most Western countries is ~10–20% for men and 5–10% for women; higher rates are seen in individuals who seek help from health care deliverers. Between 2001 and 2013, the proportion of the U.S. population with a current (i.e., past 12 months) alcohol use disorder increased by 49% with increases of almost 100% in women, African Americans, and individuals aged ≥45. Rates are similar in the United States, Canada, Germany, Australia, and the United Kingdom; tend to be lower in most Mediterranean countries, such as Italy, Greece, and Israel; and may be higher in Ireland, France, Eastern Europe (e.g., Russia), and Scandinavia. An even higher lifetime prevalence has been reported for most native cultures, including Native Americans, Eskimos, Maori groups, and aboriginal tribes of Australia. These differences in prevalence reflect both cultural and genetic influences, as described below. In Western countries, the typical individual with alcohol use disorder has a family and a career, and the lifetime risk among physicians is similar to that of the general population.

GENETICS

Some of the most exciting recent research developments into alcohol-related disorders have clarified the contribution of genetic influences to these conditions. These investigations include how variations in genes relate to environmental and attitudinal mediators of genetic effects. Understanding how specific gene variations contribute to the risk for a condition has the potential to help with early identification of individuals at high risk, development of effective prevention efforts, and, perhaps, identifying individuals most likely to respond to specific medications.

Approximately 60% of the risk for alcohol use disorder is attributed to genes, as indicated by the fourfold higher risk in children with an alcohol use disorder parent (even if adopted early in life and raised by nonalcoholics) and a higher risk in identical twins compared to fraternal twins of affected individuals. Like most medical and psychiatric conditions that are referred to as complex genetically influenced disorders, the risk for alcohol use disorders is related to hundreds of gene variations, many of which explain <1% of the risk. As a result, vulnerabilities toward the condition are often approached by considering multiple gene variations at the same time using polygenic risk scores. These genetic variations operate primarily through intermediate characteristics that subsequently combine with environmental influences to alter the risk for heavy drinking and alcohol problems. These include genes relating to a high risk for all substance use disorders that operate through impulsivity, schizophrenia, and bipolar disorder. Another characteristic, an intense skin flushing response when drinking, decreases risk for only alcohol use disorders, and not substance use conditions related to other drugs, through gene variations for several alcohol-metabolizing enzymes, especially ALDH (a mutation only seen in Japanese, Chinese, and Korean individuals), and to a lesser extent, variations in ADH.

An additional genetically influenced characteristic that increases the risk for heavy drinking, a low level of response or low sensitivity to alcohol, can be seen very early in the drinking career and before acquired tolerance or alcohol used disorders develop. The low response per drink operates, in part, through variations in genes relating to calcium and potassium channels, GABA, nicotinic, dopamine, and serotonin systems. Prospective studies have demonstrated that this need for higher doses of alcohol to achieve effects predicts future heavy drinking, alcohol problems, and alcohol use disorders, but not problems with drugs other than alcohol. The impact of a low response to alcohol on adverse drinking outcomes is partially mediated by a range of environmental and attitudinal influences, including the selection of heavier-drinking friends, more positive expectations of the effects of high doses of alcohol, and using alcohol to cope with stress. Several studies of college freshmen demonstrated that helping students who have a low sensitivity to alcohol modify these influences was associated with lower drinking quantities and fewer alcohol-related problems over the subsequent year.

NATURAL HISTORY

Although the average age of the first drink (~15 years) is similar in individuals who do and do not go on to develop alcohol use disorders, an earlier onset of regular drinking and drunkenness, especially in the context of conduct problems, is associated with a higher risk for later alcohol-related diagnoses. By the mid-twenties, most nonalcoholic men and women begin to moderate their drinking (perhaps learning from negative consequences), whereas those with alcohol use disorders are likely to escalate their drinking despite difficulties. The first major life problem from alcohol often appears in the late teens to early twenties, and a pattern of multiple alcohol difficulties by the mid-twenties. Once established, the course is likely to include exacerbations and remissions, with little difficulty in temporarily stopping or controlling alcohol use when problems develop, but without help desistance usually gives way to escalations in alcohol intake and subsequent problems. Following treatment, for at least a year, more than half of those with alcohol use disorder maintain a marked decrease in alcohol use and related problems or achieve full abstinence, including many who stop drinking permanently. Even without formal treatment or self-help groups, there is at least a 20% chance of spontaneous remission with long-term abstinence. However, should the individual continue to drink heavily, the life span is shortened by ~10 years on average, with the leading causes of early death being enhanced rates of heart disease, cancer, accidents, and suicide.

IDENTIFICATION AND TREATMENT

The approach to treating alcohol-related conditions is relatively straightforward: (1) recognize that at least 20% of patients have an alcohol use disorder; (2) learn how to identify and treat acute alcohol-related conditions (e.g., severe intoxication); (3) know how to help patients begin to address their alcohol problems; (4) know how to treat alcohol withdrawal symptoms; and (5) learn how to appropriately treat or refer patients for additional help.

IDENTIFICATION OF PATIENTS WITH ALCOHOL USE DISORDERS

Even in affluent locales, the ~20% of patients who have an alcohol use disorder can be identified by asking questions about *alcohol problems* and noting laboratory test results that can reflect regular consumption of six to eight or more drinks per day. The two blood tests with ≥60% sensitivity and specificity for heavy alcohol consumption are γ-glutamyl transferase (GGT) (>35 U) and carbohydrate-deficient

TABLE 464-3 The Alcohol Use Disorders Identification Test (AUDIT)[a]

ITEM	5-POINT SCALE (LEAST TO MOST)
1. How often do you have a drink containing alcohol?	Never (0) to 4+ per week (4)
2. How many drinks containing alcohol do you have on a typical day?	1 or 2 (0) to 10+ (4)
3. How often do you have six or more drinks on one occasion?	Never (0) to daily or almost daily (4)
4. How often during the last year have you found that you were not able to stop drinking once you had started?	Never (0) to daily or almost daily (4)
5. How often during the last year have you failed to do what was normally expected from you because of drinking?	Never (0) to daily or almost daily (4)
6. How often during the last year have you needed a first drink in the morning to get yourself going after a heavy drinking session?	Never (0) to daily or almost daily (4)
7. How often during the last year have you had a feeling of guilt or remorse after drinking?	Never (0) to daily or almost daily (4)
8. How often during the last year have you been unable to remember what happened the night before because you had been drinking?	Never (0) to daily or almost daily (4)
9. Have you or someone else been injured as a result of your drinking?	No (0) to yes, during the last year (4)
10. Has a relative, friend, doctor, or other health worker been concerned about your drinking or suggested that you should cut down?	No (0) to yes, during the last year (4)

[a] The AUDIT is scored by simply summing the values associated with the endorsed response. A score ≥8 may indicate harmful alcohol use.

transferrin (CDT) (>20 U/L or >2.6%); the combination of the two tests is likely to be more accurate than either alone. The values for these serologic markers are likely to return toward normal within several weeks of abstinence. Other useful blood tests include high-normal MCVs (≥91 μm^3) and serum uric acid (>416 mol/L, or 7 mg/dL).

The diagnosis of alcohol use disorder ultimately rests on the documentation of a pattern of repeated difficulties associated with alcohol (Table 464-2). The criteria can be paraphrased as reaching a point where alcohol means more to the person than the significant repetitive problems that it causes. Thus, in screening, it is important to probe for marital or job problems, legal difficulties, histories of accidents, medical problems, evidence of tolerance, and so on, and then attempt to relate these issues to use of alcohol. Some standardized questionnaires can be helpful, including the 10-item Alcohol Use Disorders Identification Test (AUDIT) **(Table 464-3)**, but these are only screening tools, and a face-to-face interview is still required for a meaningful diagnosis. The diagnostic criteria in the fourth and fifth versions of the American Psychiatric Association DSM (DSM-IV and DSM-5) are very similar, both are reliable across different clinicians, and both labels are very good at predicting future problems, especially for individuals with moderate or severe disorders.

TREATMENT

Alcohol-Related Conditions

ACUTE INTOXICATION

The first priority in treating severe intoxication is to assess vital signs and manage respiratory depression, cardiac arrhythmias, and blood pressure instability, if present. The possibility of intoxication with other drugs should be considered by obtaining, if needed, toxicology screens for other central nervous system (CNS) depressants such as benzodiazepines and for opioids. Aggressive behavior should be handled by offering reassurance but also by calling for help from an intervention team. If the aggressive behavior continues, relatively low doses of a short-acting benzodiazepine such as lorazepam (e.g., 1–2 mg PO or IV) may be used and can be repeated as needed, but care must be taken not to destabilize vital signs or worsen confusion. An alternative approach is to use an antipsychotic medication (e.g., olanzapine 2.5–10 mg IM repeated at 2 and 6 h, if needed).

INTERVENTION

The steps presented here follow the acronym of SBIRT, indicating *s*creening, *b*rief *i*nterventions, and treatment or *r*eferral to *t*reatment. There are two main elements to highlighting the need for compliance with treatment in a person with an alcohol use disorder: motivational interviewing and brief interventions. During motivational interviewing, the clinician helps the patient to think through the assets (e.g., comfort in social situations) and liabilities (e.g., health- and interpersonal-related problems) of the current pattern of drinking. The clinician should listen empathetically to the responses, help the patient weigh options, and encourage the taking of responsibility for needed changes. Patients should be reminded that only they can decide to avoid the consequences that will occur if heavy drinking continues. The process of brief intervention, a similar approach, has been summarized by the acronym FRAMES: *F*eedback to the patient; *R*esponsibility to be taken by the patient; *A*dvice, rather than orders, on what needs to be done; *M*enus of options that might be considered; *E*mpathy for understanding the patient's thoughts and feelings; and *S*elf-efficacy, i.e., offering support for the capacity of the patient to make changes.

Once the patient begins to consider change, the discussions can focus more on the consequences of high alcohol consumption, suggested approaches to stopping drinking, and help in recognizing and avoiding situations likely to lead to heavy drinking such as going to bars or associating with heavy-drinking friends. Both motivational interviewing and brief interventions can be carried out in 15-min sessions, but because patients often do not change behavior immediately, multiple meetings are often required to explore the problem and possible options, discuss optimal treatments, and explain the benefits of abstinence.

ALCOHOL WITHDRAWAL

If the patient agrees to stop drinking, sudden decreases in alcohol intake can produce withdrawal symptoms, most of which are the opposite of those produced by intoxication. Features include tremor of the hands (shakes); agitation and anxiety; autonomic nervous system overactivity including an increase in pulse, respiratory rate, sweating, and body temperature; and insomnia. These symptoms usually begin within 5–10 h of decreasing ethanol intake, peak on day 2 or 3, and improve by day 4 or 5, although mild levels of these problems may persist for 4–6 months as a protracted abstinence syndrome.

About 2% of individuals with alcohol use disorder experience a withdrawal seizure, with the risk increasing in the context of older age, concomitant medical problems, misuse of additional drugs, and higher alcohol quantities. The same risk factors also contribute to the ~1% rate of withdrawal delirium, also known as *delirium tremens* (DTs), where the withdrawal includes a severe agitated delirium (mental confusion, agitation, and fluctuating levels of consciousness) associated with a tremor and autonomic overactivity (e.g., marked increases in pulse, blood pressure, and respirations). The risks for seizures and DTs can be diminished by identifying and treating underlying medical conditions early in the course of withdrawal and by instituting adequate doses of depressant medications such as benzodiazepines.

Thus, the first step in dealing with possible withdrawal phenomena is a thorough physical examination in all heavy drinkers who are considering abstinence. This includes evaluation of possible liver impairment, gastrointestinal bleeding, cardiac arrhythmias, infection, and glucose or electrolyte imbalances. It is also important to offer adequate nutrition and oral multiple B vitamins, including 50–100 mg of oral thiamine daily for a week or more. Because most patients with alcohol use disorders who enter withdrawal are either normally hydrated or mildly overhydrated, IV fluids should be avoided unless there is a relevant medical problem or significant recent bleeding, vomiting, or diarrhea.

The next step is to recognize that because withdrawal symptoms reflect the acute decrease in the usual blood levels of a CNS depressant (i.e., alcohol), the symptoms can be controlled by administering any other depressant in doses that decrease symptoms (e.g., a rapid pulse and tremor) and then tapering the dose over 3–5 days. Although most depressants are effective, benzodiazepines **(Chap. 463)** have the most supportive data for use in this situation, combining a high level of safety and low cost. Short-half-life benzodiazepines can be considered for patients with serious liver impairment or evidence of significant brain damage, but they must be given every 4 h to avoid abrupt blood-level fluctuations that may increase the risk for seizures. Therefore, most clinicians use drugs with longer half-lives (e.g., chlordiazepoxide), adjusting the dose if signs of withdrawal escalate and withholding the drug if the patient is sleeping or has orthostatic hypotension. The average patient requires 25–50 mg of chlordiazepoxide or 10 mg of diazepam given PO every 4–6 h on the first day, with doses then decreased to zero over the next 5 days. Although alcohol withdrawal can be treated in a hospital, patients in good physical condition who demonstrate mild signs of withdrawal despite low blood alcohol concentrations and who have no prior history of DTs or withdrawal seizures can be considered for outpatient detoxification. For the next 4 or 5 days, these patients should receive only 1 or 2 days of medications at a time and return daily for evaluation of vital signs. They can be hospitalized if signs and symptoms of withdrawal markedly escalate.

Treatment of patients with DTs can be challenging, and the condition is likely to run a course of 3–5 days regardless of the therapy used. However, conditions that meet the criteria for DTs outlined above represent medical emergencies that carry an estimated mortality as high as 5%, and treatment is best carried out in an intensive care unit by well-trained clinicians who closely monitor vital signs. Medications can include high-dose benzodiazepines (e.g., as much as 800 mg/d of chlordiazepoxide has been reported) or, for those who do not respond to that regimen, closely monitored doses of propofol or dexmedetomidine. The focus of care is to identify and correct medical problems and to control behavior and prevent injuries. Antipsychotic medications are not recommended for treatment of alcohol withdrawal symptoms; although antipsychotics are less likely than benzodiazepines to exacerbate confusion, they may increase the risk of seizures.

Generalized withdrawal seizures rarely require more than the administration of an adequate dose of benzodiazepines. There is little evidence that anticonvulsants such as phenytoin or gabapentin are more effective than benzodiazepines for alcohol-withdrawal seizures, and the risk of seizures has usually passed by the time effective drug levels are reached. The rare patient with status epilepticus must be treated aggressively **(Chap. 436)**.

HELPING INDIVIDUALS WITH ALCOHOL USE DISORDERS TO STOP OR SIGNIFICANTLY DECREASE DRINKING: THE REHABILITATION PHASE

An Overview After completing alcoholic rehabilitation, ≥50% of individuals with alcohol use disorders, especially highly functioning patients, maintain abstinence or significant diminution of alcohol intake for at least a year; many also achieve long-term sobriety. The ideal outcome is abstinence, but treatment trials are increasing recognizing that outcomes shy of total abstinence can still improve levels of functioning and quality of life. The core components of the rehabilitation phase of treatment include cognitive-behavioral approaches to help patients recognize the need to change, while working with them to alter their behaviors to enhance compliance. A key step is to optimize motivation toward abstinence through education of patients and their significant others about alcohol use disorders and their likely course over time. It is important to recognize that contrary to what some physicians might think, the typical person with an alcohol use disorder is likely to have a job and a family and not fit the inaccurate “down and out” stereotype. However, after years of heavy drinking, some patients require vocational or avocational counseling to help to structure their days, and all patients should try self-help groups such as Alcoholics Anonymous (AA) to assist them in developing a sober peer group and to learn how to deal with life’s stresses while remaining sober. *Relapse prevention education* helps patients identify situations in which a return to drinking is likely (e.g., stopping in a bar to meet friends but planning to only have a nonalcoholic beverage), formulate ways to avoid the risky situation, and when that is not possible, to mitigate the risks to which they are exposed. It is also important to develop coping strategies that increase the chances of a quick return to abstinence after an episode of drinking.

Although many individuals can be treated as outpatients, more intense interventions are more effective and some individuals with alcohol use disorders do not respond to just AA or outpatient groups. Whatever the setting, ongoing contact with outpatient treatment staff should be maintained for at least 6 months and preferably for a year after abstinence. Counseling focuses on areas of improved functioning in the absence of alcohol (i.e., why it is a good idea to continue abstinence), helping patients to manage free time without alcohol, encouraging them to develop a nondrinking peer group, and discussions of ways to handle stress without drinking.

The physician serves an important role in identifying the alcohol problem, diagnosing and treating associated medical and independent or substance-induced psychiatric syndromes, overseeing detoxification, referring the patient to outpatient or inpatient rehabilitation programs, providing counseling, and, if appropriate, selecting which (if any) medication might be needed. For insomnia, patients should be reassured that troubled sleep is likely to improve over subsequent weeks. They should be taught the elements of “sleep hygiene” including maintaining consistent schedules for bedtime and awakening, avoiding exercise or consumption of large meals before bedtime, and keeping the bedroom cool, dark, and quiet at night **(Chap. 33)**. Depressant sleep medications are not the optimal approach for this type of insomnia that often continues for several weeks or months. Patients are likely to develop rebound insomnia when the depressant dose is decreased or stopped. The rebound increases the chance they will increase the dose and potentially develop problems controlling the prescribed depressant drug. Sedating antidepressants (e.g., trazodone) should not be used because they interfere with cognitive functioning the next morning and disturb the normal sleep architecture, but occasional use of over-the-counter sleeping medications (sedating antihistamines) can be considered. An additional problem, anxiety symptoms, can be addressed by increasing patients’ insights into the temporary nature of the symptoms and helping them develop strategies to achieve relaxation by using forms of cognitive therapy.

Medications for the Alcohol Rehabilitation Treatment Phase The core of the rehabilitation phase for any chronic relapsing condition, including alcohol use disorder, relates to cognitive and behavioral approaches that help people comply with treatment goals and improve health and quality of life. Any medication for this disorder is likely to operate optimally in the context of such cognitive-behavioral approaches. As a result, the efficacy of a medication is best measured as the gain in functioning over and above improvements associated with the motivational interviewing, brief interventions, and related behavioral approaches. Such additional treatments (e.g., medications) are likely to have modest effect sizes, which can be difficult to document. Adding to the challenge of establishing the efficacy of a medication for this condition are the fluctuations of the intensity of alcohol-related symptoms over time and the 20% or higher spontaneous remission for alcohol use disorder. Recognizing that all treatments might cause harm through side effects and financial costs, it is important to demonstrate that a medication has a beneficial asset-to-liability ratio using double-blind controlled treatment trials.

In that light, to date, well-structured controlled trials have revealed only a few medications that have even modest benefits when used in the first 6–12 months of recovery from an alcohol use disorder. The opioid antagonist naltrexone may shorten subsequent relapses, whether used in the oral form (50–150 mg/d) or as a once-per-month 380-mg injection. By blocking opioid receptors, naltrexone decreases activity in the dopamine-rich ventral tegmental reward system and decreases the feeling of pleasure if alcohol is imbibed. A second medication, acamprosate (Campral) (~2 g/d divided into three oral doses), has similar modest effects. Acamprosate inhibits NMDA receptors, decreasing mild symptoms of protracted withdrawal. Several trials of combined naltrexone and acamprosate have reported that the combination is well tolerated, and the efficacy might be superior to either drug alone, although not all studies agree.

It is more difficult to establish the asset-to-liability ratio of a third drug, disulfiram, an ALDH inhibitor, used clinically at doses of 250 mg/d, a dose selected to avoid the side effects of the more effective 500 mg/d regimen. This drug produces vomiting and autonomic nervous system instability after drinking as a result of rapidly rising blood levels of acetaldehyde. This reaction to alcohol can be dangerous, especially for patients with heart disease, stroke, diabetes mellitus, or hypertension. The drug itself carries potential risks of temporary depressive or psychotic symptoms, peripheral neuropathy, and liver damage. Disulfiram is best given under supervision by someone (such as a spouse), especially during high-risk drinking situations (such as the Christmas holidays).

Regarding other medications, a 16-week, placebo-controlled trial in patients with histories of relatively severe acute withdrawal syndromes reported good outcomes during the rehabilitation phase with another depressant medication (gabapentin 1200 mg/d), but side effects were considerable; those results have not yet been replicated, and gabapentin can itself be misused. Additional drugs under investigation include another opioid antagonist, nalmefene; the nicotinic receptor agonist varenicline; the serotonin antagonist ondansetron; the α-adrenergic agonist prazosin, especially in combination with naltrexone; the $GABA_B$ receptor agonist baclofen; the anticonvulsant topiramate; ibudilast in individuals with low-intensity alcohol responses; and possible enhanced outcomes when talk therapies are combined with ketamine or psilocybin sessions. However, it is important to emphasize that currently there are insufficient data to determine the asset-to-liability ratio for these medications in treating alcohol use disorders, and therefore, there is insufficient support for the routine use of these medications in clinical settings.

GLOBAL CONSIDERATIONS

As described above, rates of alcohol use disorders differ across sex, age, ethnicity, and country. There are also differences across countries regarding the definition of a standard drink (e.g., 10–12 g of ethanol in the United States and 8 g in the United Kingdom) and the definition of being legally drunk. The preferred alcoholic beverage also varies across groups, even within countries. That said, regardless of sex, ethnicity, or country, the actual drug in the drink is still ethanol, and the risks for problems, course of alcohol use disorders, and approaches to treatment are similar across the world.

FURTHER READING

Bogenschutz MP et al: Percentage of heavy drinking days following psilocybin-assisted psychotherapy vs placebo in the treatment of adult patients with alcohol use disorder: A randomized clinical trial. JAMA Psychiatry 79:953, 2022.

Finn SW et al: Treatment of alcohol dependence in primary care compared with outpatient specialist treatment: Twelve-month follow-up of a randomized controlled trial, with trajectories of change. J Stud Alcohol Drugs 81:300, 2020.

Lai D et al: Evaluating risk for alcohol use disorder: Polygenic risk scores and family history. Alcohol Clin Exp Res 46:374, 2022.

Livne O et al: Agreement between DSM-5 and DSM-IV measures of substance use disorders in a sample of adult substance users. Drug Alcohol Depend 227:108958, 2021.

Mattle A et al: Gabapentin to treat acute alcohol withdrawal in hospitalized patients: A systematic review and meta-analysis. Drug Alcohol Depend 241:109671, 2022.

McPheeters M et al: Pharmacotherapy for alcohol use disorder: A systematic review and meta-analysis. JAMA 330:1653, 2023.

Meredith LR et al: The effect of neuroimmune modulation on subjective response to alcohol in the natural environment. Alcohol Clin Exp Res 46:879, 2022.

Schuckit MA et al: The low level of response to alcohol-based heavy drinking prevention program: One-year follow-up. J Stud Alcohol Drugs 77:25, 2016.

Witkiewitz K et al: Can alcohol use disorder recovery include some heavy drinking? A replication and extension up to 9 years following treatment. Alcohol Clin Exp Res 44:1862, 2020.

Witt SH et al: Acute alcohol withdrawal and recovery in men lead to profound changes in DNA methylation profiles: A longitudinal clinical study. Addiction 115:2034, 2020.

Wood E et al: Canadian guideline for the clinical management of high-risk drinking and alcohol use disorder. CMAJ 195:E1364, 2023.

465

Nicotine Addiction

David M. Burns

Ingestion of nicotine in any form (combusted or heated tobacco leaf, oral use, nicotine pouches, or inhaled by vaping nicotine or nicotine salt) can create and sustain addiction if used with sufficient intensity for a sufficient duration. Addicted users of nicotine regulate their nicotine intake by adjusting the frequency and intensity of their tobacco use, both to obtain the desired psychoactive effects and avoid withdrawal. While nicotine does cause some disease processes including complications of pregnancy, the vast majority of the diseases produced by nicotine addiction result from repetitive exposure to the carcinogens and other toxicants in various nicotine containing products. These exposures produce incremental changes that accumulate to progress toward disease, but it is the addiction to nicotine that causes the long-term, multiple times per day exposures to these agents needed to create sufficient damage to manifest as cancer, heart, and lung disease.

Over the last decade, there have been major shifts in product use, with about 40% of the nicotine delivery products being noncombusted and non-tobacco-leaf formulations. This shift has made practitioner advice about nicotine addiction more complicated, in part because many of the noncombusted products offer the potential to deliver nicotine sufficient to satisfy addiction, but with dramatic reductions in most of the toxic constituents found in smoke. Substantial reduction in disease risk can be achieved with the use of nicotine "vaping" products to aid cessation success for combusted cigarette smokers, and their use is likely to provide meaningful disease reduction with long-term complete substitution for combusted cigarette smoking even without breaking nicotine addiction.

Nicotine offers little or no benefit for the nonaddicted individual, but regular use of inhaled nicotine can produce a powerful addiction that is difficult to break and expensive to sustain. Any use of nicotine earlier in life predicts a greater use of some nicotine product later in life. For the practitioner, despite the likelihood of much lower disease risks compared to combusted cigarette use, it is hard to justify, or ignore, high-frequency use of noncombusted inhaled nicotine products in recommendations to patients who have never used combusted tobacco products.

THE PROCESS OF NICOTINE ADDICTION

When nicotine reaches the brain, it reversibly attaches to nicotinic acetylcholine receptors, which are particularly active in brain networks involved in depression, joy, excitement, and happiness. With prolonged high exposure to nicotine, nicotinic receptors are upregulated on brain cells. For most individuals, daily nicotine use is needed to produce changes in the brain that are the hallmark of addiction. The strength of addiction, and the speed with which it can develop, are influenced by the frequency of use and concentration of arterial levels of nicotine reaching the brain, which can vary widely with use of different products.

As the time since last cigarette becomes longer, nicotine levels in the blood drop, and nicotine detaches from the receptors, leaving them increasingly uncovered. Without high levels of nicotine attaching to these receptors, addicted smokers no longer feel "normal," creating a compulsive need for the next dose of nicotine. Addicted individuals perceive increasing withdrawal symptoms with increasing duration of abstinence, which can persist for 4–6 weeks.

For the practicing clinician, the diagnosis of nicotine addiction is straightforward and is manifest by the patient's loss of control over their next use of nicotine-containing products. The individual's need for the next dose of nicotine can be satisfied by the use of multiple products at different times. The loss of control is demonstrated by a history of daily nicotine use coupled with behaviors such as high numbers of uses per day, use shortly after waking in the morning, craving after periods of abstinence, and/or failure of past cessation efforts, among others.

Several genes have been associated with nicotine addiction. Some reduce the clearance of nicotine, and others have been associated with an increased likelihood of becoming dependent on tobacco and other drugs or a higher incidence of depression. It is likely that genetic susceptibility can influence the probability that adolescent experimentation with tobacco will lead to addiction as an adult. However, rates of smoking cessation have increased, and rates of nicotine addiction have decreased dramatically since the mid-1950s, suggesting that factors other than genetics are more important influences for tobacco use.

In tobacco smoke or vaping aerosol, the fraction of nicotine present in the unprotonated (freebase) form is greater at alkaline pH, and unprotonated nicotine is more easily absorbed into the bloodstream across the oral mucosa. However, high concentrations of unprotonated nicotine are irritating to the airway, reducing the tendency to inhale, limiting the amount of nicotine users will tolerate in the aerosol and slowing the rate of rise in nicotine blood levels. Products delivering nicotine that are highly addictive (cigarettes and aerosols of nicotine salts) are able to deliver high doses of nicotine and rapid rises in arterial blood concentration by producing aerosols that are mildly acidic in the mouth, reducing the irritation that inhibits inhalation. However, as the particles are inhaled into the lung, the pH of the smoke rapidly changes to the more alkaline pH of the blood (~7.4), increasing the fraction of readily absorbable unprotonated nicotine present in the alveoli. The high rate of blood flow through the alveoli rapidly removes the unprotonated nicotine, increasing conversion of the protonated nicotine or the nicotine salt to release the unprotonated form and allowing most of the nicotine in the aerosol to be absorbed into the bloodstream. The resulting rapid rise in arterial blood nicotine levels reaching the brain makes cigarettes and nicotine salt vaping devices more addictive, and more able to satisfy addiction, compared with other forms of nicotine delivery.

TRENDS IN NICOTINE PRODUCT USE

Manufactured cigarettes have been the dominant form of nicotine exposure over the last 100 years. **Figure 465-1** presents the rise and decline in U.S. per-capita consumption (total cigarettes sold divided by the U.S. population over age 18) since 1900, together with the rise and subsequent fall of the male and female lung cancer death rates resulting from that consumption. The figure demonstrates the enormous success of tobacco control efforts in changing a human risk behavior, with current adult cigarette smoking prevalence now approaching 10%.

Among high school seniors, any cigarette use in the last 30 days has declined from 31% in 2000 to 2.9% in 2023, with only 0.7% smoking daily. As these post-2000 high school students aged into their 20s, their smoking initiation rates did not increase substantially, and only 5% of 18- to 24-year-old adults currently smoke. Unfortunately, smoking prevalence rates among those over age 40, the population more likely to develop disease, have declined more slowly. In addition, the population continuing to smoke has shifted heavily toward socially disadvantaged groups, indicating that the need for health care–based smoking interventions remains a priority.

A vast array of devices to deliver nicotine without smoke have been introduced over the last decade. These include devices, many sold without U.S. Food and Drug Administration (FDA) approval, that aerosolize nicotine solution or nicotine salts, or deliver nicotine as strips or flavored nicotine pouches for oral use. With the exception of vaping, very limited data are available on patterns of use for most of these products in the United States.

For vaping among adolescents, any use in the last 30 days peaked at about 25% in high school seniors in 2019 and declined to about 17% in 2023 with daily use at about 5.8% in 2023. The dramatic fall in combusted cigarette use over the past 2 decades among adolescents resulted from changes in the regulation of cigarette advertising and

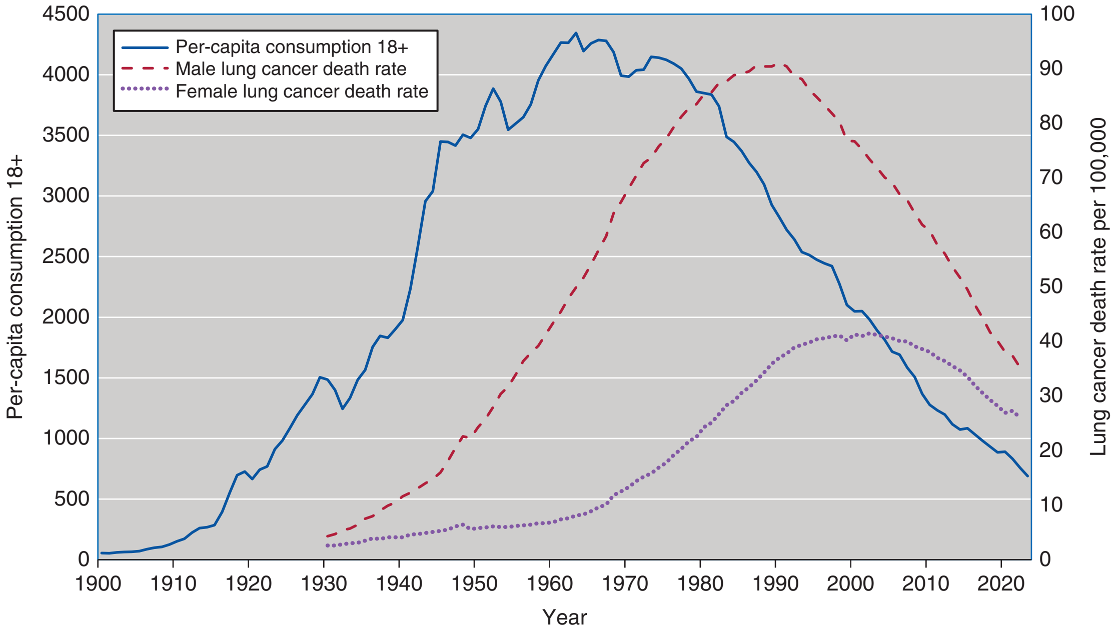

FIGURE 465-1 Changes in per-capita consumption and lung cancer death rates from 1900 to 2023.

the increasing social stigma associated with cigarette smoking. The introduction of vaping products as substitute sources of nicotine had little effect on the reduction in adolescent cigarette smoking, and the availability of vaping products has not increased adolescent smoking prevalence.

For adults, about 5% report using e-cigarettes some days or every day, with 11% of the 18–24 group reporting at least some days. Reasons for using e-cigarettes may include the following: for recreation, to sustain addiction, or to attempt to quit cigarette smoking. A substantial proportion of those using e-cigarettes for cessation report use of both e-cigarettes and combusted cigarettes interchangeably, called dual use.

In contrast to the United States, tobacco consumption in many other countries remains high, particularly among lesser developed countries. Worldwide tobacco consumption recently declined slightly for the first time after a persistent rise over the recent decades.

Changes in cost and relative rates of taxation for different tobacco products has resulted in an increase in smoking of cigars and "roll your own" cigarettes, but past recommendations about lower risk with use of cigars and pipe tobacco no longer apply. Most cigars are now manufactured in much the same way as cigarettes with small changes in weight to qualify for lower tax rates, and most pipe tobacco is now used for "roll your own" cigarettes also because of lower tax rates. As a result, smoking these products currently has risks similar to smoking manufactured cigarettes. Practitioners should ask about their use and should not recommend them as lower-risk forms of satisfying nicotine addiction.

DISEASE MANIFESTATIONS OF CIGARETTE SMOKING

Approximately 40% of cigarette smokers will die prematurely due to cigarette smoking unless they are able to quit. The major diseases caused by cigarette smoking are listed in **Table 465-1**. The ratio of smoking-related disease rates in smokers compared to never smokers (relative risk) increases with advancing age for most cancers and for chronic obstructive pulmonary disease (COPD). Relative risk declines with advancing age for cardiovascular diseases due to the increasing contribution of other risk factors to cardiovascular disease as age advances. Nevertheless, even for cardiovascular disease, the absolute difference in mortality rate between smokers and never smokers, called excess death rate, continues to increase with advancing age, as one would expect from a process of cumulative injury.

CARDIOVASCULAR DISEASES

Cigarette smokers are more likely than nonsmokers to develop both large-vessel atherosclerosis and small-vessel disease. Approximately 90% of peripheral vascular disease in the nondiabetic population can be attributed to cigarette smoking, as can ~50% of aortic aneurysms. In contrast, 24% of coronary artery disease and ~11% of ischemic and hemorrhagic strokes are attributed to cigarette smoking. There is a multiplicative interaction between cigarette smoking and other cardiac risk factors such that the increment in risk produced by smoking among individuals with hypertension or elevated serum lipids is substantially greater than the increment in risk produced by smoking for individuals without these risk factors.

In addition to its role in promoting atherosclerosis, cigarette smoking also increases the likelihood of myocardial infarction and sudden cardiac death by promoting platelet aggregation and vascular occlusion. Reversal of these effects on coagulation may explain the rapid benefit of smoking cessation for a new coronary event demonstrable among those who have survived a first myocardial infarction. This effect may also explain the substantially higher rates of graft occlusion among continuing smokers following vascular bypass surgery for cardiac or peripheral vascular disease.

Cessation of cigarette smoking reduces the risk of a second coronary event within 6–12 months. Rates of first myocardial infarction and death from coronary heart disease decline within 2–4 years following cessation among those with no prior cardiovascular history. After 15 years of abstinence, the risk of stroke, a new myocardial infarction, and death from coronary heart disease in former smokers is similar to that for those who have never smoked.

TABLE 465-1 Relative Risks for Current Smokers of Cigarettes

AGE	35–44	45–64	65–74	≥75
Males				
Lung cancer	14.33	19.03	28.29	22.51
Coronary heart disease	3.88	2.99	2.76	1.98
Cerebrovascular disease	2.17	1.48	1.23	1.12
Other vascular diseases			7.25	4.93
Chronic obstructive pulmonary disease (COPD)			29.69	23.01
All causes	2.55	2.97	3.02	2.40
Females				
Lung cancer	13.30	18.95	23.65	23.08
Other tobacco-related cancers	1.28	2.08	2.06	1.93
Coronary heart disease	4.98	3.25	3.29	2.25
Cerebrovascular disease	2.27	1.70	1.24	1.10
Other vascular diseases			6.81	5.77
COPD			38.89	20.96
All causes	1.79	2.63	2.87	2.47
Relative Risks for Selected Other Cancers				
Other cancers	Male		Female	
Larynx	14.6		13	
Lip, oral cavity, pharynx	10.9		5.1	
Esophagus	6.8		7.8	
Bladder	3.3		2.2	
Kidney	2.7		1.3	
Pancreas	2.3		2.3	
Stomach	2		1.4	
Liver	1.7		1.7	
Colorectal	1.2		1.2	
Cervix			1.6	
Acute myeloid leukemia	1.4		1.4	

CANCER

Tobacco smoking causes cancer of the lung; lip; oral cavity; naso-, oro-, and hypopharynx; nasal cavity and paranasal sinuses; larynx; esophagus; stomach; pancreas; liver (hepatocellular); colon and rectum; kidney (body and pelvis); ureter; urinary bladder; uterine cervix; and acute myeloid leukemia. There does not appear to be a causal link between cigarette smoking and cancer of the endometrium, and there is a lower risk of uterine cancer among postmenopausal women who smoke. The risks of cancer increase linearly with the increasing number of cigarettes smoked per day and logarithmically with increasing duration of smoking. Additionally, there are synergistic interactions between cigarette smoking and alcohol use for cancer of the oral cavity and esophagus. Several occupational exposures synergistically increase lung cancer risk among cigarette smokers, most notably occupational asbestos and radon exposure.

Cessation of cigarette smoking reduces the risk of developing cancer relative to continuing smoking after about 4 years of abstinence, but even 20 years after cessation, there is a persistent three- to fourfold increased risk of developing lung cancer compared to those who have never smoked.

RESPIRATORY DISEASE

Cigarette smoking is responsible for 80% of COPD. Within 1–2 years of beginning to smoke regularly, many young smokers will develop inflammatory changes in their small airways, although lung function measures of these changes do not predict subsequent development of chronic airflow obstruction. Chronic mucous hyperplasia of the larger airways results in a chronic productive cough in as many as 80% of smokers >60 years of age. Chronic inflammation and narrowing of the small airways, and/or enzymatic digestion of alveolar walls resulting in pulmonary emphysema, reduce expiratory airflow sufficiently to produce clinical symptoms of respiratory limitation in ~15–25% of smokers.

Changes in the small airways of young smokers will reverse after 1–2 years of abstinence. There is also a small increase in measures of expiratory airflow following smoking cessation among many individuals who have already developed chronic airflow obstruction, but the major change following cessation is a slowing of the rate of decline in lung function with advancing age rather than a return of lung function toward normal.

PREGNANCY

Cigarette smoking is associated with several maternal complications of pregnancy: premature rupture of membranes, abruptio placentae, and placenta previa. There is also a small increase in the risk of spontaneous abortion among smokers. Infants of smoking mothers are more likely to experience preterm delivery, have a higher perinatal mortality rate, be small for their gestational age, and have higher rates of infant respiratory distress syndrome. They are more likely to die of sudden infant death syndrome and appear to have a developmental lag for at least the first several years of life. Since it is likely that some of these pregnancy-related risks are caused or enhanced by nicotine, vaping and oral use of nicotine remain a concern in pregnancy except as a strategy to abstain from cigarette smoking.

OTHER CONDITIONS

Smoking delays healing of peptic ulcers and increases the risk of developing periodontal disease, diabetes, active tuberculosis, rheumatoid arthritis, osteoporosis, senile cataracts, and neovascular and atrophic forms of macular degeneration. It results in premature menopause, wrinkling of the skin, gallstones, cholecystitis in women, and male impotence. Patients who continue to smoke during treatment for cancer with chemotherapy or radiation have poorer outcomes and reduced survival.

ENVIRONMENTAL TOBACCO SMOKE

Long-term exposure to environmental tobacco smoke increases the risk of lung cancer and coronary artery disease among nonsmokers. It also increases the incidence of respiratory infections, chronic otitis media, and asthma in children, and it causes exacerbation of asthma in children.

LOWER TAR AND NICOTINE CIGARETTES

Filtered cigarettes with lower machine-measured yields of tar and nicotine commonly use ventilation holes in the filters and other engineering designs to artificially lower the machine measurements. Smokers compensate for the lowered nicotine delivery resulting from these design changes by changing the manner in which they puff on the cigarette or the number of cigarettes smoked per day to restore their level of nicotine intake to that needed to satisfy their addiction. As a result, actual tar and nicotine deliveries to smokers are not reduced with use of these products, negating any reduction in disease risks from switching to these products.

The amount of carcinogenic tobacco-specific nitrosamines in the tobacco used in cigarettes has increased over time, and cigarette design changes that reduce machine-measured tar and nicotine also lead to deeper inhalation of the smoke in the lung, presenting increased amounts of the more carcinogenic smoke to the alveolar portions of the lung. These changes increase the risk of adenocarcinoma of the lung above that produced by smoking older nonfiltered cigarettes. The changes in cigarette design and composition of cigarettes over the past six decades are the cause of the increase in rates of adenocarcinoma of the lung observed over the past half century, and the increased adenocarcinoma rate has increased total lung cancer rates, as there has not been a decline in the risk of other cell types with the changes in cigarette design. An increased risk for COPD may also be present. There has been no increase in risk of all lung cancer or adenocarcinoma of the lung over the same period among never smokers.

PHARMACOLOGIC INTERACTIONS

Cigarette smoking may interact with a variety of other drugs. Cigarette smoking induces the cytochrome P450 system, which may alter the metabolic clearance of drugs such as warfarin. This may result in inadequate serum levels in smokers as outpatients when the dosage is established in the hospital under nonsmoking conditions. Correspondingly, serum levels may rise when smokers are hospitalized and not allowed to smoke. Smokers may also have higher first-pass clearance for drugs such as lidocaine, and the stimulant effects of nicotine may reduce the effect of benzodiazepines or beta blockers.

OTHER FORMS OF TOBACCO USE

Other major forms of tobacco use are loose moist snuff or packets deposited between the cheek and gum and chewing tobacco. Oral tobacco use leads to gum disease and can result in oral and pancreatic cancer. There are dramatically higher risks evident for products used in Africa or Asia as, including for heart disease, compared to those used in the United States and Europe.

NICOTINE AEROSOLS

A continually expanding array of devices and oral formulations that deliver nicotine in quantities capable of creating and sustaining addiction are available. Many of these devices deliver levels of nicotine comparable to a cigarette, and different concentrations of nicotine are often offered. Nicotine salt aerosols use mild acids to facilitate inhalation of the aerosol and can deliver very high amounts of nicotine even with novice users, potentially enhancing their addictiveness compared to devices that deliver solutions of nicotine. Nicotine intake is higher for users of devices delivering nicotine salts, and they report a greater frequency of symptoms of dependence with abstinence.

Biomarker evidence on exposure to smoke toxicants demonstrates markedly lower exposures among those using e-cigarettes exclusively compared to cigarette smokers, suggesting that they have less disease risk with use. However, both biomarker and behavioral evidence demonstrate the capacity for these devices to create and sustain addiction. The long-term rates with which adults addicted to nicotine vaping quit nicotine completely or migrate to or relapse back to combusted product use as they age remain to be determined.

There is convincing randomized controlled cessation trial evidence that the use of e-cigarettes that deliver sufficient nicotine are as effective as other nicotine-replacement or varenicline medications in achieving sustained abstinence from cigarettes, but a meaningful proportion of those who achieve abstinence still use e-cigarettes at 12-month follow-up, suggesting continued nicotine addiction. Rates of longer-term relapse back to smoking among individuals with persistent nicotine addiction remain to be examined.

An additional concern is that evidence on e-cigarette use in the United States shows that approximately one-half of adult e-cigarette users continue to also smoke conventional cigarettes, negating the benefits of reduced toxicant exposure and successful abstinence.

Refillable or reloadable e-cigarette devices can be used to aerosolize a variety of liquids other than those provided by the manufacturer. Disposable “pods” and liquids for these devices can be purchased from the manufacturer but are also available from other sources that may use poor-quality manufacturing practices and control of contaminants. They may also contain marijuana oils, other drugs, and flavors not evaluated for potential lung injury with inhalation.

CESSATION

The process of stopping smoking is commonly a cyclical one, with the smoker sometimes making multiple attempts to quit, and failing, before finally being successful. Approximately 70–80% of smokers would like to quit smoking. More than one-half of current cigarette smokers attempted to quit in the last year, but only 6% quit for 6 months, and only 3% remain abstinent for 2 years. Clinician-based smoking interventions should repeatedly encourage smokers to try to quit and to use different forms of cessation assistance with each new cessation attempt.

Advice from a clinician to quit smoking, particularly at the time of an acute illness, is a powerful trigger for cessation attempts, with up to half of patients who are advised to quit making a cessation effort.

CLINICIAN INTERVENTIONS (TABLE 465-2)

The shift in the nicotine market to products containing nicotine but not tobacco, and particularly to aerosolized nicotine salt products, has complicated provider diagnosis and treatment recommendations. Two considerations should guide this process. First, assessment of nicotine addiction should include the cumulative number of episodes of use for all nicotine products used regularly when considering the intensity of nicotine ingestion. This is important particularly when there is variation in the type of product used from day to day. Inquiries about cigar and roll your own use, as well as vaping and oral nicotine products, are necessary. It is the frequency of any nicotine dosing, not the product source, that creates and sustains addiction.

Second, it is the other toxicants carried along with the nicotine that cause the majority of the disease risk, so eliminating the smoke intake can have benefits even in the presence of continued addiction to nicotine.

All patients should be asked the total daily frequency with which they use any nicotine product, how long they have used at least one nicotine product regularly, their past experience with quitting, and whether they are currently interested in quitting.

The goal is to identify whether the individual has a pattern of use that demonstrates a compulsive need for the next dose of nicotine. The number of episodes of smoking or vaping per day, how they are spaced during the day, as well as use of nicotine smoking within 30 min of waking are helpful measures of the intensity of nicotine addiction.

Even those who are not interested in quitting should be encouraged and motivated to quit by providing a clear, strong, and personalized message by the clinician that smoking cigarettes is an important health concern and that vaping can be addicting. Those uninterested individuals should be told that assistance is available if they become interested in quitting in the future. Many of those not currently expressing an interest in quitting may nevertheless make an attempt to quit in the subsequent year.

TABLE 465-2 Clinical Practice Guidelines

Physician Actions
Ask: Systematically identify all tobacco and nicotine use at every visit
Advise: Strongly urge all smokers to quit
Identify smokers willing to quit
Assist the patient in quitting
Arrange follow-up contact
Effective Pharmacologic Interventions[a]
First-line therapies
Nicotine gum (1.5)
Nicotine patch (1.9)
Nicotine nasal inhaler (2.3)
Nicotine oral inhaler (2.1)
Nicotine lozenge (2 mg: 2.0, 4 mg: 2.8)
Bupropion (2.0)
Varenicline (3.1)
Other Effective Interventions[a]
Physician or other medical personnel counseling (10 min) (1.84)
Intensive group smoking cessation programs (at least 4–7 sessions of 20- to 30-min duration lasting at least 2 and preferably 8 weeks) (1.3)
Intensive individual counseling (1.7)
Systemwide cessation tracking and assistance (5)
Telephone counseling (1.6)
Exclusive E-cigarette use (3.0)

[a] Numerical value following the intervention is the multiple for cessation success compared to no intervention.

For those interested in quitting, a quit date should be negotiated, usually not the day of the visit but within the next few weeks. A follow-up contact by office staff around the time of the quit date should be provided. There is a relationship between the amount of assistance a patient is willing to accept and the success of the cessation attempt.

Building smoking cessation as a priority into health care delivery systems by including systemwide tracking of smoking status, prompting of practitioners to ask about smoking and interest in cessation, system-based outreach to smokers to offer cessation assistance and programs between visits, and tracking of cessation outcomes can dramatically enhance sustained abstinence, with 12-month abstinence rates as high as 25%.

There are a variety of cessation products listed in Table 465-2, including over-the-counter nicotine patches, gum, and lozenges, as well as nicotine nasal and oral inhalers available by prescription. These products can be used for up to 3–6 months, and some products are formulated to allow a gradual step-down in dosage with increasing duration of smoking abstinence. Antidepressants such as bupropion (300 mg in divided doses for up to 6 months) have also been shown to be effective, as has varenicline, a partial agonist for the nicotinic acetylcholine receptor (initial dose 0.5 mg daily increasing to 1 mg twice daily at day 8; treatment duration up to 6 months). Combined use of nicotine-replacement therapy (NRT) and antidepressants, as well as the use of gum or lozenges for acute cravings in patients using patches, can increase cessation outcomes.

Pretreatment with antidepressants or varenicline is recommended for 1–2 weeks prior to the quit date. Pretreatment with nicotine patches for 2 weeks prior to a cessation date is also useful. Longer duration of nicotine replacement as a maintenance therapy for those who are unsuccessful in quitting with a shorter duration of use is a useful strategy. NRT is provided in different dosages, with higher doses being recommended for more intense smokers. Antidepressants are more effective among smokers with a history of depression symptoms.

Current recommendations are to offer pharmacologic treatment, usually with nicotine patches or varenicline, to all who will accept it, and to provide counseling and other support as a part of the cessation attempt. Cessation advice alone by clinicians or their staff is likely to increase success compared with no intervention, but a more comprehensive approach with advice, pharmacologic assistance, and counseling can increase cessation success nearly threefold.

Data from multiple studies show that switching from cigarettes to exclusive use of e-cigarettes, particularly those with nicotine salts that deliver high doses of nicotine, is as or more effective in achieving smoking abstinence compared to FDA-approved medications, and it may be more acceptable for some patients. It should be recommended only with a strong caution to avoid dual use. Dual use with combusted cigarettes is unlikely to lead to smoking cessation or long-term risk reduction.

Current recommendations suggest that FDA-approved cessation methods be tried initially, with aerosolized nicotine salt products recommended to those who fail initial attempts to quit, are quitting on their own, or are reluctant to use FDA-approved medications.

PREVENTION

Prevention of smoking initiation must begin early, preferably in the elementary school years. Practitioners who treat adolescents should be sensitive to the prevalence of this problem even in the preteen population. Practitioners should ask all adolescents whether they have experimented with nicotine or currently use nicotine products, reinforce the fact that most adolescents and adults do not smoke or use nicotine, and explain that all forms of nicotine intake can be both addictive and potentially harmful.

■ FURTHER READING

Christen SE et al: Pharmacokinetics and pharmacodynamics of inhaled nicotine salt and free-base using an e-cigarette: A randomized crossover study. Nicotine Tob Res 10:ntae074, 2024.

Eisenberg MJ et al: Effect of e-cigarettes plus counseling vs counseling alone on smoking cessation: A randomized clinical trial. JAMA 324:1844, 2020.

Hajek P et al: A randomized trial of e-cigarettes versus nicotine-replacement therapy. N Engl J Med 380:629, 2019.

Kaplan B et al: Effectiveness of ENDS, NRT and medication for smoking cessation among cigarette-only users. Tob Control 32:302, 2023.

Leone FT et al: Initiating pharmacologic treatment in tobacco-dependent adults. An official American Thoracic Society Clinical Practice Guideline. Am J Respir Crit Care Med 202:e5, 2020.

Meza R et al: Trends in US adult smoking prevalence, 2011 to 2022. JAMA Health Forum 4:e234213, 2023.

Royal College of Physicians: E-cigarettes and harm reduction: An evidence review. RCP, 2024. Available at *https://www.rcplondon.ac.uk/news/rcp-calls-regulations-protect-children-and-young-people-vaping*.

U.S. Department of Health and Human Services: The Health Consequences of Smoking: 50 Years of Progress. A Report of the Surgeon General. Atlanta, GA, 2014. Available at *https://www.ncbi.nlm.nih.gov/books/NBK179276/pdf/Bookshelf_NBK179276.pdf*.

466

Cannabis and Cannabis Use Disorder

Nora D. Volkow, Aidan Hampson, Ruben Baler

Cannabis/marijuana is used by more than >150 million people worldwide. The 2023 United States National Survey on Drug Use and Health, estimated that 42 million people used "marijuana" over the last month, while 62 million people used nicotine and 136 million drank alcohol (*https://www.samhsa.gov/data/release/2023-national-survey-drug-use-and-health-nsduh-releases*). Notably in 2023, 10 million younger people (12-25 year olds) consumed marijuana each month, while 12 million used tobacco products. For reference, nearly 19 million 12-25 year olds used alcohol on a monthly basis.

Since the Agriculture Improvement Act of 2018 (AIA; Public Law 115-334) amended the Controlled Substance Act, products made from *Cannabis sativa* species have been defined as marijuana when the plant or finished product contains >0.3% Δ-9-tetrahydrocannabinol (D9THC) by dry weight. Conversely, cannabis plant material, extracts, and derivatives that contain no more than 0.3% D9THC by dry weight are defined as "hemp." We will be using the terms marijuana and cannabis interchangeably in this report since marijuana is the term used in multiple surveys, and in state and Federal laws. Marijuana is primarily grown for its euphoric and medicinal properties, which are principally mediated by D9THC acting on neuronal type 1 cannabinoid receptors (CB1R). The contribution of minor cannabinoids (including cannabidiol [CBD]) or terpenes is limited and largely unproven to be clinically relevant at concentrations found in marijuana. Marijuana remains federally illegal as of May 2024, but 24 states, 3 U.S. territories, and Washington, DC, have passed measures to legalize nonmedicinal (adult) use, whereas 38 states, 4 U.S. territories, and Washington, DC, have legalized medicinal cannabis use. Hemp varietals are federally legal to grow and manufacture into seed/oil products, fiber, or CBD extracts. Some varietals can contain >12% (w/w) of the nonintoxicating CBD, while the D9THC level remains within the definition of "hemp" (0.3%). High CBD hemp was first envisaged as a source for the treatment of childhood seizure disorders (Federal bill H.R. 5226, 2014). However, after hemp legalization, processors began to derivatize hemp into Δ-8-tetrahydrocannabinol (D8THC) and other intoxicating hemp products (IHPs). According to one recent survey, people who use D8THC report it to be milder than D9THC and perceive it to be legal. Pharmacologic studies support lower potency of D8THC than D9THC, but other IHPs such as tetrahydrocannabiphorol can be considerably more potent (20–30×). It is worth noting that the 2023 Monitoring the Future national survey found that 11% of 12th-grade U.S. students reported past-year D8THC use.

Marijuana legalization has increased the demand for, and supply of, increasingly high D9THC-containing cannabis strains. Cannabis flower products now range from <0.3% in hemp to 20–30% in legal dispensary products. Further, extracts used as liquids in "vape pens" purportedly contain up to 75% D9THC, while advertisements of solid extracts (e.g., "shatter") containing up to 95% D9THC are not uncommon, although most of these claims appear to be exaggerated. Solid high-D9THC preparations are used for high-temperature vaporization ("dabbing"). Inhalation is a highly bioavailable (efficient) and rapid route of absorption for D9THC, so vaping products deliver high D9THC doses with rapid effect onset, conditions linked to increased risk of addiction. In contrast, edible D9THC-infused products (e.g., candies, cookies, and drinks) have slower rates of onset and are perceived to be associated with reduced harm. Edible products are less bioavailable than inhaled D9THC, but absorption can vary depending on the presence and nature of food in the stomach. Notably, "edibles" are favored by younger users and are associated with a higher probability of accidental dosing than inhaled products.

PHARMACOLOGIC EFFECTS

Cannabis is used recreationally because it enhances the subjective sense of well-being, provides rewarding sensations, and can dampen stress responses. However, consumption of high doses of D9THC can induce anxiety, paranoia, and panic. D9THC is primarily a partial agonist (activator) of G protein–coupled cannabinoid receptors (CB1R and CB2R), with the euphorigenic effects mediated through CB1Rs primarily located on excitatory glutamatergic and inhibitory γ-aminobutyric acid (GABA)-ergic interneurons and glial cells in brain regions that process stress, mood, and reward. These receptors are the effectors of the endocannabinoid system (ECS), which is physiologically activated by the endogenous ligands 2-arachidonoylglycerol (2-AG; a full agonist), and anandamide (a partial agonist). According to current understanding, 2-AG modulates synaptic signaling by inhibiting overstimulated synapses. Endocannabinoids are synthesized and eliminated *on demand* and thus provide a temporally and regionally specific modulatory signal. In contrast, the effect of intoxicating cannabinoids are defined by dose and their period of effect, not subtle "just in time" synaptic synthesis. Consequently exogenous cannabinoids disrupt important ECS neuroregulatory processes.

The rewarding effects of D9THC are thought to be mediated by modulating glutamatergic and GABAergic activity in the midbrain ventral tegmental area (VTA), from where dopaminergic neurons project into the nucleus accumbens (NAc). In the NAc, rewarding experiences are reinforced (learned) by glutamatergic and dopaminergic signal interactions. The anxiolytic effects of D9THC are mediated by its effects on the amygdala, a region critical for threat perception and emotional reactivity.

CANNABIS PHARMACOKINETICS

Smoking (e.g., joints and water pipes) remains the most common route of cannabis administration, but e-cigarette–type "vape pens" are increasingly being used. Vape pens use "e-liquid" concentrates, and the plasma concentrations achieved following their use depend on the e-liquid D9THC concentration and the "puffing profile." The ability to take a single puff from a vape pen offers easier dose control than smoking. Vaping is perceived to be less harmful than smoking, but the long-term toxicology of e-liquids in humans is not yet clear.

The subjective effects of cannabis are affected by dose, route of administration, (smoked, vaped, or ingested), and the subject's prior experience (which also modulates expectation). Smoked D9THC exhibits a bioavailability of 10–35%, with interindividual differences stemming from variations in the capacity to hold smoke in the lungs. The pharmacokinetics of heated (not burnt) cannabis flower and vaporized ethanolic extracts are similar to those of smoked flower, but no data are available to compare pharmacokinetics of smoked flower and commercial e-liquids or solid concentrates. When cannabis is smoked, plasma

D9THC concentrations become maximal within 5–10 min. After this time, plasma levels decrease with two phases governed by different elimination exponentials. There is an initial distribution phase with an (alpha) half-life ($t_{1/2}$) of ~6 min, governed by absorption into lipophilic tissues (e.g., brain). During this alpha phase, brain concentrations continue to increase even as plasma levels fall, resulting in a hysteresis (lag) between changes in plasma concentration and pharmacodynamic effects. Subjective effects tend to be maximal at 20–30 min after smoking. The plasma beta $t_{1/2}$ is approximately 1–2 h, and the total period of pharmacodynamic effect ranges from 4 to 8 h. Elimination of D9THC occurs by conversion to 11-hydroxy-THC (pharmacologically active) and subsequent conversion to a long-lasting, but pharmacologically inactive, 11-norcarboxy THC metabolite (11-COOH-THC). It is this analyte that is detected during marijuana urinalysis. Terminal elimination of 11-norcarboxy-THC exhibits a $t_{1/2}$ ranging from 20 to 35+ h, so THC use remains detectable for days in those who use occasionally and for weeks in those who use frequently and with saturated fat stores.

Regular D9THC use results in tolerance to its pharmacologic effects, so a given D9THC plasma concentration may not correlate to similar impairment levels in those using occasionally versus regularly. This challenge in correlating D9THC levels in biological matrices with behavioral effects has hampered efforts to regulate marijuana-impaired driving.

Orally consumed cannabis products (edibles and drinks) typically exhibit slower effect onset than smoked/vaped D9THC. This is partly due to stomach residence time but mainly because lipophilic cannabinoids are poorly absorbed through the water/mucus-rich layer of the intestine. The comparatively slower rate of oral cannabinoid absorption means that no hysteresis is observed, unlike after inhalation. Orally administered D9THC is extensively metabolized by intestinal and hepatic cytochromes (first-pass metabolism), so bioavailability ranges from 5 to 6% (compared to 30+% when smoked). However, when consumed with fatty foods, cannabinoids can exhibit 200–400% increased bioavailability because fatty foods (or pharmaceutical triglyceride vehicles) stimulate bile release. This emulsifies fat-associated cannabinoids, thus increasing the surface area for absorption into portal blood and stimulating absorption by enterocytes, which secrete into lymphatic lacteals. This allows a portion of fat-dissolved cannabinoids to bypass hepatic elimination, increasing bioavailability. However, lymphatic flow is slower than blood, so while lymphatically transported D9THC exhibits higher bioavailability, it also has a slower effect onset. In summary, depending on the edible product design or the consumer's feeding status, D9THC bioavailability and onset time can vary greatly. This unpredictability can complicate efforts to determine a precise dose and presents a potential overdose risk for the unwary.

HARMFUL EFFECTS

The frequency and severity of cannabis adverse effects are influenced by dose, frequency of use, route of administration, and the individual's health, age, and genetic background. Especially concerning are the potential negative effects of cannabis on the brain during early life stages. Perturbation of ECS signaling during early fetal development affects neuronal development, migration, and connectivity. A recent animal study found that deficiencies in vital fatty acids like docosahexaenoic acid could help explain the association between intrauterine exposure to D9THC and lifelong health disturbances in the offspring, including cognitive and memory impairments. The relevant human studies, which are few and confounded by the frequent use of other drugs, provide substantial evidence of lower birth weight and suggest an association between maternal marijuana use and fetal growth and preterm delivery. In addition, early results from the Adolescent Brain and Child Development (ABCD) study, a longitudinal neuroimaging, behavioral, and genetic study of close to 12,000 children in the United States, provide some evidence that exposure to cannabis during pregnancy can increase intracranial volumes and blunt development of visuospatial processing. Another analysis of the same cohort also found small but significant effects on white matter integrity during childhood, especially in the fornix, which has been implicated in the processing of emotions and memory. Not surprisingly, the American College of Obstetricians and Gynecologists recommends discouraging use of marijuana by women who are pregnant or planning a pregnancy.

Children and adolescents are also more vulnerable to the harmful effects of cannabis, the use of which increases markedly during adolescence and has been associated with lower grades, lower IQ, and higher risk of dropping out of school, although a causal relationship cannot yet be established. Brain imaging studies have revealed that use of cannabis at this stage is associated with structural and functional brain changes (not always replicated) often in the form of reduced brain connectivity and cortical thickness. Though it is not clear whether these are caused by early exposure to cannabis, research in 799 adolescents in Europe found a negative, dose-dependent correlation between self-reported cannabis use at age 14 and prefrontal cortex thickness at age 19, suggesting that cannabis use in middle to late adolescence may alter cortical development. Moreover, recent results from the ABCD study are consistent with the notion that preadolescence exposure to cannabis may contribute to lower scores on an episodic memory task and that more cannabis use may lead to poorer performances on verbal, inhibitory, working memory, and episodic memory tasks.

In 2019–2020, a syndrome known as e-cigarette or vaping product use-associated lung injury (EVALI) was observed. It became associated with vitamin E acetate, a thickening agent used to dilute black-market marijuana e-liquids. Once known, EVALI incidents diminished as this unapproved agent disappeared from use.

CANNABIS USE DISORDER

Repeated cannabis use, especially during adolescence, can result in cannabis use disorder (CUD), which the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*, defines as "a problematic pattern of cannabis use leading to clinically significant impairment or distress." The diagnostic criteria for a use disorder include drug craving, tolerance to effects, a withdrawal syndrome, and failure to fulfill role obligations due to recurrent use and drug seeking. The risk of CUD increases with earlier age of initiation, frequency of use, and exposure to cannabis with high THC content. Several studies have found a broad reduction in cannabinoid receptors in the brains of people who use cannabis when compared to control participants who do not use cannabis, but receptor density recovers rapidly, returning to values similar to those of control participants within 28 days of abstinence. In people who use cannabis regularly, abstinence results in a withdrawal syndrome that peaks within 1–3 days of drug discontinuation and manifests as anxiety, restlessness, insomnia, depression, and reduced appetite. Many of these symptoms resolve within approximately 2 weeks of discontinuation. Insomnia often persists longer and may contribute to relapse, although the degree to which this represents withdrawal rather than cannabis use to self-medicate disordered sleep is unclear.

■ PREVENTION

Preventing cannabis use during adolescence reduces the risk for CUD and the risk of other substance use disorders. There are several evidence-based prevention strategies focused on children and adolescents that have shown benefits in decreasing cannabis use during adolescence or in delaying its age of initiation. Such prevention interventions can target the individual (e.g., Keepin' It Real, Life Skills, InShape), the family (e.g., Brief Strategic Family Therapy, Coping Power Program [CPP], Familia Adelante), or the community (e.g., The Abecedarian Project, Midwestern Prevention Project, Caring School Community). School-based prevention programs are the most widely implemented. Evidence from randomized controlled trials, prospective cohort trials, and longitudinal studies all demonstrate that comprehensive interventions that include antidrug information with refusal skills, self-management skills, and social skills training provide the most effective approaches for long-term reduction of cannabis (and alcohol) use in adolescents.

■ TREATMENT

The treatment of CUD involves tapering cannabis use and providing psychosocial support to alleviate withdrawal symptoms. There are currently no U.S. Food and Drug Administration (FDA)-approved medications for CUD. There is evidence of effectiveness of several behavioral interventions as CUD treatments. Contingency management is an effective therapeutic and combined motivational enhancement, and cognitive behavioral techniques have also shown clinically significant improvements in abstinence and reduced cannabis use.

OTHER ADVERSE EFFECTS

MENTAL ILLNESS

The association between marijuana use and increased risk of mental illnesses is an area of major concern. The risk of psychosis increases with the frequent consumption of high-THC-content cannabis (>10% THC). Even upon first exposure, high-potency cannabis can trigger acute psychotic episodes, which constitute one of the main causes for emergency department (ED) visits associated with cannabis use. Most of these psychotic episodes are transient but can become chronic with regular cannabis use. In those vulnerable, cannabis may trigger or exacerbate the presentation of schizophrenia. Many earlier studies (though not all) have linked adolescent cannabis use with higher risk and earlier onset of chronic psychosis, particularly for those using cannabis at higher frequency or with higher D9THC content. A large recent study showed a stronger association between cannabis use during adolescence and risk of psychotic disorder than that documented in previous studies, consistent with the rise in cannabis potency. Concerns have also been raised regarding cannabis use during adolescence and a higher risk for depression and suicidality, though these associations have been much less studied.

ACCIDENTS

Cannabis use increases the risk of injuries when driving under its influence. D9THC impairs judgment, motor coordination, and reaction time, all necessary for safe driving. Studies have found a direct relationship between blood D9THC levels and impaired driving ability.

ACUTE AND CHRONIC TOXICITY

The increased availability of high-D9THC-content products over the past decade has been paralleled by increased marijuana-related ED visits and hospital admissions. Such emergencies can be caused by acute toxicity and chronic use syndromes. Cannabis edibles are involved in a significant portion of acute cannabis toxicity events. Patients include children accidentally consuming sweet edibles and infrequent users such as “cannabis tourists” with limited experience of the consumed products or the longer onset time of edible products. Actual D9THC dose is difficult to envisage, for both edible or inhaled products, so naïve or infrequent users are at increased risk of overdosing. Cannabis toxicity is frequently manifested by severe anxiety, tachycardia, and even acute psychoses. Chronic high-dose cannabis use can also induce a cannabis hyperemesis syndrome (presenting as severe cycles of nausea, vomiting, and abdominal pain), a growing cause for ED and hospital admissions.

THERAPEUTIC POTENTIAL

Currently, no FDA-approved medications contain cannabis-derived THC, although synthetic D9THC (or dronabinol) is approved for treatment of chemotherapy-induced nausea and appetite stimulation. Several countries have approved the cannabis-derived D9THC:CBD formulation Sativex for treating chronic pain and multiple sclerosis (MS)-induced spasticity. However, evidence of Sativex efficacy in MS is largely based on patient reports. Chronic pain is one of the most frequent indications for which medical marijuana is used. A recent analysis of data from the New York prescription drug monitoring database, from the years 2017-2019, revealed that chronic pain patients on long-term stable doses of opioid therapy (n >8000) who used medical marijuana for >8 months (compared to those who used medical marijuana for <30 days to 8 months) had a 30–50% reduction in opioid use.

FURTHER READING

Albaugh MD et al: Association of cannabis use during adolescence with neurodevelopment. JAMA Psychiat 78:1, 2021.

Evanski JM et al: The first “hit” to the endocannabinoid system? Associations between prenatal cannabis exposure and frontolimbic white matter pathways in children. Biol Psychiatry Glob Open Sci 4:11, 2024.

Hagler DJ Jr et al: Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. Neuroimage 202:116091, 2019.

Hiraoka D et al: Effects of prenatal cannabis exposure on developmental trajectory of cognitive ability and brain volumes in the Adolescent Brain Cognitive Development (ABCD) Study. Dev Cogn Neurosci 60:101209, 2023.

McDonald AJ et al: Age-dependent association of cannabis use with risk of psychotic disorder. Psychol Med 22:1, 2024.

National Survey on Drug Use and Health: Substance Use and Mental Health Services Administration, 2023. *https://www.samhsa.gov/data/release/2023-national-survey-drug-use-and-health-nsduh-releases*.

Sarikahya MH et al: Prenatal THC exposure induces long-term, sex-dependent cognitive dysfunction associated with lipidomic and neuronal pathology in the prefrontal cortex-hippocampal network. Mol Psychiat 28:4234, 2023.

Schwabe AL et al: Uncomfortably high: Testing reveals inflated THC potency on retail cannabis labels. PLoS One 18:e0282396, 2023.

Volkow ND et al: Don’t worry, be happy: Endocannabinoids and cannabis at the intersection of stress and reward. Annu Rev Pharmacol Toxicol 57:285, 2017.

Wade NE et al: Cannabis use and neurocognitive performance at 13-14 years-old: Optimizing assessment with hair toxicology in the Adolescent Brain Cognitive Development (ABCD) Study. Addict Behav 150:107930, 2024.

467

Opioid-Related Disorders

Thomas R. Kosten, Colin N. Haile

Opioid analgesics have been used since at least 300 b.c. Nepenthe (Greek for “free from sorrow”) helped the hero of the *Odyssey*, but widespread opium smoking in China and the Near East has caused harm for centuries. Since the first chemical isolation of opium and codeine 200 years ago, a wide range of synthetic opioids have been developed, and opioid receptors were cloned in the 1990s. Two of the most important adverse effects of all these agents are the development of opioid use disorder and overdose. Prescription opioids are primarily used for pain management, but due to ease of availability individuals procure and misuse these drugs with dire consequences. In 2022, for example, 8.9 million United States residents misused pain relievers and >76,000 overdose deaths involved opioids, nearly half combined with stimulants. These numbers continue to increase and have accelerated due to mixing of high-potency fentanyl derivatives with other opioids and stimulants. The accelerating death rates are partially because reversal of fentanyl overdoses can require severalfold larger doses of naloxone than the doses in the intranasal devices used for nonmedical street resuscitations. Fentanyl-associated deaths also increased during the COVID-19 pandemic. The World Drug Report attributes the greatest global burden of morbidity and mortality to opioid misuse, including disease transmission, increased health care, crime, law enforcement, family distress, and lost productivity.

The terms *dependence* and *addiction* have been replaced with opioid use disorder, opioid intoxication, and opioid withdrawal. Opioid use disorder is defined in the *Diagnostic and Statistical Manual of Mental Disorders,* Fifth Edition (DSM-5; 2022) as the repeated use of the opiate during a 12-month period while producing problems in two or more areas including tolerance, withdrawal, use of greater amounts of opioids than intended, craving, and use despite adverse consequences. This new definition reduces the diagnostic criteria from three problem areas to two, but this reduction has not changed the rates of these disorders because most opioid-using individuals meet more than three criteria.

A striking recent aspect of illicit opioid use has been its marked increase as the gateway to illicit drugs in the United States. Since 2007, prescription opiates have surpassed marijuana as the most common illicit drug that adolescents initially use, although overall rates of opioid use are far lower than marijuana. The most used opioids had been diverted prescriptions for oxycodone and hydrocodone until about 2015, when fentanyl misuse and lethal overdose rose exponentially. Two opioid maintenance treatment agents—methadone and buprenorphine—are also misused, but at substantially lower rates, and the partial opioid agonists such as butorphanol, tramadol, and pentazocine are misused even less frequently. Because the chemistry and general pharmacology of these agents are covered in major pharmacology texts, this chapter focuses on neurobiology and pharmacology relevant to opioid use disorder and its treatments. Although the neurobiology of misuse involves all four of the known opioid receptors—mu, kappa, delta, and nociceptin/orphanin—the mu receptor is the most clinically related to opioids.

TABLE 467-1 Actions of Opioid Receptors

RECEPTOR TYPE	ACTIONS
Mu (μ) (e.g., morphine, buprenorphine)	Analgesia, reinforcement euphoria, cough and appetite suppression, decreased respirations, decreased GI motility, sedation, hormone changes, dopamine and acetylcholine release
Kappa (κ) (e.g., butorphanol)	Dysphoria, decreased GI motility, decreased appetite, decreased respiration, psychotic symptoms, sedation, diuresis, analgesia
Delta (δ) (e.g., etorphine)	Analgesia, euphoria, physical dependence, hormone changes, appetite suppression, dopamine release
Nociceptin/orphanin (e.g., buprenorphine)	Analgesia, appetite, anxiety, tolerance to opioids, hypotension, decreased GI motility, 5-HT and NE release

Abbreviations: GI, gastrointestinal; 5-HT, serotonin; NE, norepinephrine.

NEUROBIOLOGY

After binding to mu opioid receptors, opioids downregulate intracellular messenger systems and activate potassium ion channels, as summarized in **Table 467-1**. All opioid receptors are G protein–linked and coupled to the cyclic adenosine monophosphate (cAMP) second messenger system and to G protein–coupled inwardly rectifying potassium channels (GIRKs). The GIRKs increase permeability to potassium ions, causing hyperpolarization and inhibiting action potential production. Thus, opioids inhibit the activity of all neurons with opioid receptors and induce analgesia, sedation, and drug reinforcement through various brain pathways.

Relevant pathways for the reinforcing euphoric effects of opioids include the mesolimbic dopaminergic pathway from the ventral tegmental area (VTA) to the nucleus accumbens (NAc), where opioids indirectly increase synaptic levels of dopamine through inhibition of GABAergic neurons that inhibit both the VTA and the NAc. The positive subjective effects of opioid drugs also include mu receptor desensitization and internalization, potentially related to stimulation of β-arrestin signaling pathways. However, the "high" only occurs when the *rate of change* in dopamine is fast. Large, rapidly administered doses of opioids block γ-aminobutyric acid (GABA) inhibition and produce a burst of VTA dopamine neuron activity that is associated with a "high" in commonly misused substances. Therefore, routes of administration that slowly increase opioid blood and brain levels, such as oral and transdermal routes, are effective for analgesia and sedation but do not produce an opioid "high" that follows smoking and intravenous routes. Other acute effects such as analgesia and respiratory depression involve opioid receptors located in other brain areas such as the periaqueductal grey for pain and medulla oblongata for respiration.

Opioid tolerance and withdrawal are related to genetic polymorphisms that impact several proteins and that after chronic opioid dosing affect the functioning of the cAMP–protein kinase A (PKA)–cAMP response-element binding protein (CREB) intracellular cascade within the locus coeruleus (LC) **(Fig. 467-1)**. Up to 50% of the risk for withdrawal is related to specific functional polymorphisms including one in the mu opioid receptor gene producing a threefold increase in this receptor's affinity for opioids and the endogenous ligand β-endorphin. Epigenetic DNA methylation changes in the mu receptor gene also appear to act as compensation by inhibiting gene transcription and reducing the number of mu receptors.

After chronic opioid dosing and its sustained inhibition of the cAMP molecular cascade as shown in Fig. 467-1, a compensation occurs in this cascade within the LC neurons that mediate opioid tolerance and withdrawal. Noradrenergic (NE) neurons in the LC activate the cerebral cortex. When large opioid doses saturate and activate all of the LC's mu receptors, action potentials cease. When this direct inhibitory effect is sustained over weeks and months of opioid use, a secondary set of adaptive changes occur to upregulate cyclic AMP enzyme activity. When the inhibiting opioid is abruptly discontinued, overactivity occurs in NE neurons of the LC that contribute to withdrawal symptoms (Fig. 467-1). This molecular model of NE neuronal activation during withdrawal has had important treatment implications, such as the use of the presynaptic α_2-agonists clonidine and lofexidine to treat opioid withdrawal by again suppressing NE neuronal activation

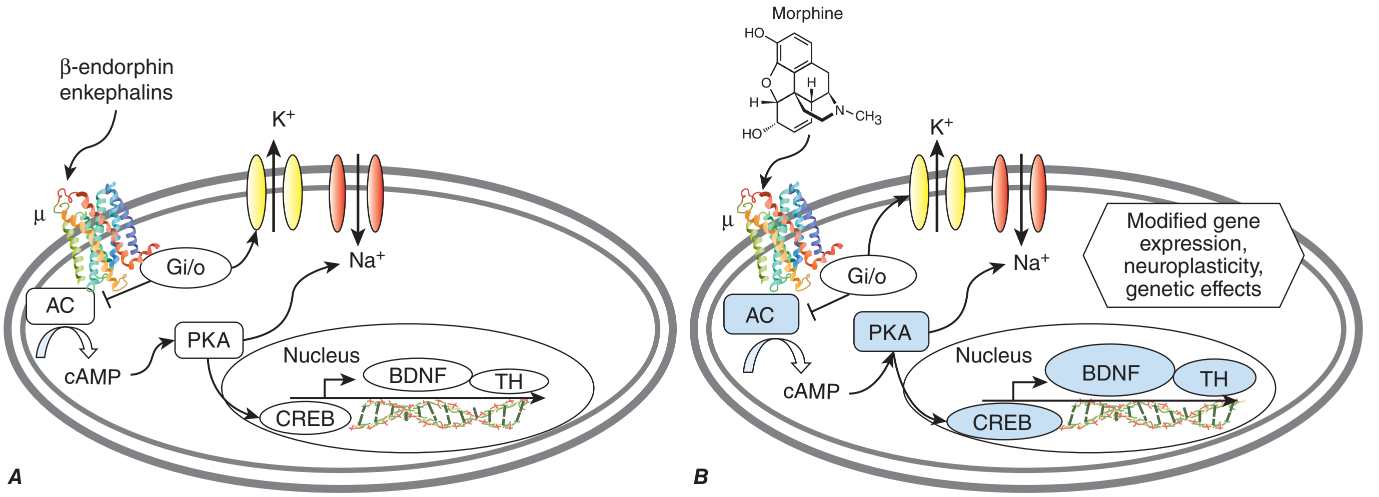

FIGURE 467-1 Normal mu-receptor activation by endogenous opioids inhibits the cyclic adenosine monophosphate (cAMP)-protein kinase A (PKA)-cAMP response-element binding protein (CREB) cascade in noradrenergic neurons within the locus coeruleus ***(A)*** through inhibitory Gi/o protein influence on adenylyl cyclase (AC). Similarly, acute exposure to opioids (e.g., morphine) inhibits this system, whereas chronic exposure to opiates ***(B)*** leads to upregulation of the cAMP pathway in an attempt to oppose opioid-induced inhibitory influence. Upregulation of this system is involved in opioid tolerance, and when the opioid is removed, unopposed noradrenergic neurotransmission is involved in opioid withdrawal. Upregulated PKA phosphorylates CREB, initiating the expression of various genes such as tyrosine hydroxylase (*TH*) and brain-derived neurotrophic factor (*BDNF*). *BDNF* is implicated in long-term neuroplastic changes in response to chronic opioids.

through feedback inhibition of this neuronal activity. Other contributors to withdrawal include deficits within the dopamine reward system and overactive neurotransmission within the glutamatergic system.

PHARMACOLOGY

Tolerance and withdrawal commonly occur with chronic daily use, developing as quickly as 6–8 weeks depending on dose concentration and dosing frequency. Tolerance is primarily pharmacodynamic with relatively limited induction of cytochrome P450 2D6 and 3A4 or other liver enzymes. Metabolism then includes conjugation to glucuronic acid and excretion of small amounts in feces. The plasma half-lives generally range from 2.5 to 3 h for morphine and >22 h for methadone. The shortest half-lives of several minutes are for fentanyl-related opioids, and the longest are for buprenorphine and its active metabolites, which can block opioid withdrawal for up to 3 days after a single dose. Tolerance to opioids leads to the need for increasing amounts of drugs to sustain the desired euphoric effects—as well as to avoid the discomfort of withdrawal. This combination has the expected consequence of strongly reinforcing misuse once it has started. Methadone taken chronically at maintenance doses is stored in the liver, which may reduce the occurrence of withdrawal between daily doses. The role of endogenous opioid peptides in tolerance and withdrawal is uncertain.

The clinical features of opioid misuse are tied to the route of administration and rapidity of the drug reaching the brain. Intravenous and smoked administration rapidly produces high drug concentrations in the brain. This produces a "rush," followed by euphoria, a feeling of tranquility, and sleepiness ("the nod"). Heroin produces effects that last 3–5 h, and several doses a day are required to forestall manifestations of withdrawal in chronic users. Symptoms of opioid withdrawal begin 8–10 h after the last dose; lacrimation, rhinorrhea, yawning, and sweating appear first. Restless sleep followed by weakness, chills, gooseflesh ("cold turkey"), nausea and vomiting, muscle aches, involuntary movements ("kicking the habit"), hyperpnea, hyperthermia, and hypertension occur in later stages of the withdrawal syndrome. The acute course of withdrawal may last 7–10 days. A secondary phase of protracted abstinence lasts for 26–30 weeks and is characterized by hypotension, bradycardia, hypothermia, mydriasis, and decreased responsiveness of the respiratory center to carbon dioxide.

Besides the brain effects of opioids on sedation and euphoria and the combined brain and peripheral nervous system effects on analgesia, a wide range of other organs can be affected. The release of several pituitary hormones is inhibited, including corticotropin-releasing factor (CRF) and luteinizing hormone, which reduces levels of cortisol and sex hormones and can lead to impaired stress responses and reduced libido. An increase in prolactin also contributes to the reduced sex drive in males. Two other hormones affected are thyrotropin, which is reduced, and growth hormone, which is increased. Respiratory depression results from opioid-induced insensitivity of brainstem neurons to increases in carbon dioxide, and in patients with pulmonary disease, this can result in clinically significant complications. In overdoses, aspiration pneumonia is common due to loss of the gag reflex. Opioids reduce gut motility, which is helpful for treating diarrhea but can lead to nausea, constipation, and anorexia with weight loss. Deaths occurred in early methadone maintenance programs due to severe constipation and toxic megacolon. Opioids such as methadone may prolong QT intervals and lead to sudden death in some patients. Orthostatic hypotension may occur due to histamine release and peripheral blood vessel dilation, which is an opioid effect usefully applied to managing acute myocardial infarction. During opioid maintenance, interactions with other medications are of concern; these include inducers of the cytochrome P450 system (usually CYP3A4) such as rifampin and carbamazepine.

Heroin users, in particular, tend to use opioids intravenously and are likely to be polydrug users, also using alcohol, sedatives, cannabinoids, and stimulants. None of these other drugs are substitutes for opioids, but they have desired additive effects. Therefore, one needs to be sure that the person undergoing a withdrawal reaction is not also withdrawing from alcohol or sedatives, which might be more dangerous and more difficult to manage.

Intravenous opioid use carries with it the risk of serious complications. The common sharing of hypodermic syringes can lead to infections with hepatitis B and HIV/AIDS, among others. Bacterial infections can lead to septic complications such as meningitis, osteomyelitis, and abscesses in various organs. Off-target effects or additions of other agents to opioids synthesized in illicit drug labs can lead to serious toxicity. For example, attempts to illicitly manufacture meperidine in the 1980s produced a parkinsonism-inducing neurotoxin, MPTP **(Chap. 446)**. More recently, adding xylazine to illicit fentanyl markedly increased fentanyl's respiratory suppression, leading to overdose deaths. Individuals who inject fentanyl and xylazine also can develop necrosis and have an increased risk of limb amputation.

TREATMENT

Opioid Overdose

The acute treatment of opioid overdose with naloxone or nalmefene is a medical emergency, and after reversal of that life-threatening complication, clinicians have two general treatment options: opioid maintenance or detoxification. Opioid agonist and partial agonist medications are commonly used for both maintenance and detoxification purposes. α_2-Adrenergic agonists are primarily used for detoxification. Antagonists are used to accelerate detoxification and then continued after detoxification to prevent relapse. The residential medication-free programs have had some success but generally less than the medication-based programs. Success of the various treatment approaches is assessed as retention in treatment and reduced opioid and other drug use; secondary outcomes, such as reduced HIV risk behaviors, crime, psychiatric symptoms, medical comorbidity, and overdoses, also indicate successful treatment.

Potentially lethal overdoses require rapid recognition and treatment with naloxone or nalmefene, two highly specific reversal agents that are relatively free of complications. The diagnosis is based on recognition of characteristic signs and symptoms, including shallow and slow respirations, pupillary miosis (mydriasis does not occur until significant brain anoxia supervenes), bradycardia, hypothermia, and stupor or coma. Blood or urine toxicology studies can confirm a suspected diagnosis, but immediate management must be based on clinical criteria. If naloxone is not administered, progression to respiratory and cardiovascular collapse leading to death occurs. With fentanyl overdoses, the naloxone dose may be twice that needed for other opioids, and recent rescue preparations contain twice the traditional dosing. Additionally, nalmefene has recently become available for treatment of overdose and has higher potency and lasts longer than naloxone. Opioids generally do not produce seizures except for unusual cases of polydrug use with the opioid meperidine, with high doses of tramadol, or in the newborn.

In addition to naloxone, management of overdose requires support of vital functions, including intubation if needed **(Table 467-2)**. If the overdose is due to buprenorphine or fentanyl, then naloxone might be required at total doses of 10 mg or greater, but primary buprenorphine overdose is nearly impossible because this agent is a partial opioid agonist, meaning that as the dose of buprenorphine is increased, it has greater opioid antagonist than agonist activity. Thus, a 0.2-mg buprenorphine dose leads to analgesia and sedation, while a hundred times greater 20-mg dose produces profound

TABLE 467-2 Management of Opioid Overdose
Establish airway. Intubation and mechanical ventilation may be necessary.
Naloxone 0.4–2.0 mg (IV, IM, or endotracheal tube). Onset of action with IV is ~1–2 min.
Repeat doses of naloxone if needed to restore adequate respiration or a continuous infusion of naloxone can be used.
One-half to two-thirds of the initial naloxone dose that reversed the respiratory depression is administered on an hourly basis (note: naloxone dosing is not necessary if the patient has been intubated).

opioid antagonism, precipitating opioid withdrawal in a person who had opioid use disorder from morphine or methadone. It is important to recognize that the goal is to reverse respiratory depression and not to administer so much naloxone that it precipitates opiate withdrawal. Because naloxone only lasts a few hours and most opioids last considerably longer, an IV naloxone drip with close monitoring is frequently employed to provide a continuous level of antagonism for 24–72 h depending on the opioid used in the overdose (e.g., morphine vs methadone). Whenever naloxone has only a limited effect, other sedative drugs that produce significant overdoses must be considered. The most common are benzodiazepines, which have produced overdoses and deaths in combination with buprenorphine. A specific antagonist for benzodiazepines, flumazenil at 0.2 mg/min can rapidly reverse overdoses, but it may precipitate seizures and increase intracranial pressure. Like naloxone, administration for a prolonged period is usually required because most benzodiazepines remain active for considerably longer than flumazenil. Support of vital functions may include oxygen and positive-pressure breathing, IV fluids, pressor agents for hypotension, and cardiac monitoring to detect QT prolongation, which might require specific treatment. Activated charcoal and gastric lavage may be helpful for oral ingestions, but intubation will be needed if the patient is stuporous.

OPIOID WITHDRAWAL

The principles of detoxification are the same for all drugs: to substitute a longer-acting, orally active, pharmacologically equivalent medication for the substance being used, stabilize the patient on that medication, and then gradually withdraw the substituted medication. Methadone and buprenorphine are the two principal medications used to treat opioid use disorder. Clonidine, a centrally acting sympatholytic agent, has also been used for detoxification in the United States. By reducing central sympathetic outflow, clonidine mitigates many of the signs of sympathetic overactivity but typically requires augmentation with other agents. Clonidine has no narcotic action and is not addictive. Lofexidine, a clonidine analogue with fewer hypotensive effects, was approved for use in the United States in 2018 for management of opioid withdrawal symptomology.

Methadone for Detoxification Dose-tapering regimens for detoxification using methadone range from 2–3 weeks to as long as 180 days, but this approach is controversial given the relative effectiveness of methadone maintenance and the low success rates of detoxification. Unfortunately, most patients tend to relapse to heroin or other opioids during or after the detoxification period, indicative of the chronic and relapsing nature of opioid use disorder.

Buprenorphine for Detoxification Buprenorphine does not appear to lead to better outcomes than methadone but is superior to clonidine in reducing symptoms of withdrawal, in retaining patients in a withdrawal protocol, and in completing treatment, although it can precipitate withdrawal in patients dependent on fentanyl or on maintenance doses of methadone (>80 mg daily).

α_2-Adrenergic Agonists for Detoxification Several α_2-adrenergic agonists have relieved opioid withdrawal by suppressing brain NE hyperactivity. Clonidine relieves some signs and symptoms of opioid withdrawal such as lacrimation, rhinorrhea, muscle pain, joint pain, restlessness, and gastrointestinal symptoms. Related agents are lofexidine, guanfacine, and guanabenz acetate. Lofexidine can be dosed up to ~2 mg/d and appears to be associated with fewer adverse effects. Clonidine or lofexidine is typically administered orally, in three or four doses per day, with dizziness, sedation, lethargy, and dry mouth as the primary adverse side effects. Outpatient-managed withdrawal regimens require close follow-up, often with naltrexone maintenance to prevent relapse.

Rapid and Ultrarapid Opioid Detoxification The opioid antagonist naltrexone, typically combined with an α_2-adrenergic agonist, has been purported to shorten the duration of withdrawal without significantly increasing patient discomfort. Completion rates using naltrexone and clonidine range from 75 to 81% compared to 40 to 65% for methadone or clonidine alone. Ultrarapid opioid detoxification is an extension of this approach using anesthetics but is highly controversial due to the medical risks and mortality associated with it.

Medications for Preventing Relapse to Illicit Opioids

Stopping opioid use is much easier than preventing relapse. Long-term relapse prevention for individuals with opioid use disorder requires combined pharmacologic and psychosocial approaches. Chronic users tend to prefer pharmacologic approaches; those with shorter histories of drug use are more amenable to detoxification and psychosocial interventions.

Methadone maintenance substitutes a once-daily oral opioid dose for three to four times daily heroin. Methadone saturates the opioid receptors and, by inducing a high level of opioid tolerance, blocks the euphoria from additional opioids. Buprenorphine, a partial opioid agonist, also can be given once daily at sublingual doses of 4–32 mg daily, and in contrast to methadone, it can be given in an office-based primary care setting.

METHADONE MAINTENANCE Methadone's slow onset of action when taken orally, long elimination half-life (24–36 h), and production of cross-tolerance at doses from 80 to 150 mg are the basis for its efficacy in treatment retention and reductions in IV drug use, criminal activity, and HIV risk behaviors and mortality. Methadone can prolong the QT interval at rates as high as 16% above rates in non-methadone-maintained, drug-injecting patients, but it has been used safely in the treatment of opioid use disorder for >40 years.

BUPRENORPHINE MAINTENANCE Sublingual buprenorphine maintenance was first approved by the U.S. Food and Drug Administration (FDA) in 2002 as a Schedule III drug for managing opioid use disorder. Unlike the full agonist methadone, buprenorphine is a partial agonist of mu-opioid receptors with a slow onset and long duration of action. Its partial agonism reduces the risk of unintentional overdose but limits its efficacy to patients who need the equivalent of only 60–70 mg of methadone, and many patients in methadone maintenance require higher doses of up to 150 mg daily. Buprenorphine is combined with naloxone at a 4:1 ratio to reduce its abuse liability. Because of pediatric exposures and diversion of buprenorphine to illicit use, mucosal films are now used rather than sublingual pills that can be crushed and snorted. A long-acting buprenorphine injection that lasts a month is also available to prevent illicit diversion and to enhance compliance.

Primary care physicians often prescribe buprenorphine for opioid use disorder, which has reduced opioid-related deaths and drug-injection-related medical morbidity with treatment retention as high as 70% over a 6-month follow-up period.

Opioid Antagonist Medications Antagonist therapy blocks the action of self-administered opioids and should eventually extinguish the habit, but this therapy is poorly accepted by patients. Naltrexone, a long-acting pure opioid antagonist, can be given orally three times a week, or by monthly injection, which markedly improves adherence, retention, and drug use. Because it is an antagonist, the patient must first be detoxified from opioids before starting naltrexone. It is safe even when taken chronically for years, is associated with few side effects (headache, nausea, abdominal pain), and can be given to patients infected with hepatitis B or C without producing hepatotoxicity. However, most providers refrain from prescribing naltrexone if liver function tests are three times above normal levels. Naltrexone maintenance combined with psychosocial therapy is effective in reducing heroin use.

Medication-Free Treatment Most opioid users enter medication-free treatments in inpatient, residential, or outpatient settings, but 1- to 5-year outcomes are very poor compared to pharmacotherapy except for residential settings lasting 6–18 months. The residential programs require full immersion in a regimented system with progressively increasing levels of independence and responsibility within a controlled community of fellow drug users. These medication-free programs, as well as the pharmacotherapy programs, also include counseling and behavioral treatments designed to teach interpersonal and cognitive skills for coping with stress and for avoiding situations leading to easy access to drugs or to craving. Relapse is prevented by having the individual very gradually reintroduced to greater responsibilities and to the working environment outside of the protected therapeutic community.

PREVENTION

Preventing the development of opioid use disorder represents a critically important challenge for physicians. Opioid prescriptions are a common source of drugs accessed by adolescents who begin a pattern of illicit drug use. The major sources of these drugs are family members, not drug dealers or the Internet. Pain management involves providing sufficient opioids to relieve the pain over as short a time as the pain warrants **(Chap. 14)**. The patient then needs to dispose of any remaining opioids, not save them in the medicine cabinet, because this behavior leads to diversion by adolescents. Finally, physicians should never prescribe opioids for themselves.

FURTHER READING

Blanco C, Volkow ND: Management of opioid use disorder in the USA: Present status and future directions. Lancet 393:1760, 2019.

Bruneau J et al: Management of opioid use disorders: A national clinical practice guideline. CMAJ 190:E247, 2018.

Food and Drug Administration: Information about medications for opioid use disorder. Accessed May 24, 2024. Available at *https://www.fda.gov/drugs/information-drug-class/information-about-medications-opioid-use-disorder-moud*.

White House Office of National Drug Control Policy (ONDCP): White House announces over $276 million for law enforcement to help address the overdose epidemic and crack down on illicit drug trafficking. Accessed May 23, 2024. Available at *https://www.whitehouse.gov/ondcp/briefing-room/2024/05/23/white-house-announces-over-276-million-for-law-enforcement-to-help-address-the-overdose-epidemic-and-crack-down-on-illicit-drug-trafficking/*.

468

Cocaine, Other Psychostimulants, and Hallucinogens

Karran A. Phillips, Wilson M. Compton

The use of cocaine, methamphetamine, other psychostimulants, and hallucinogens reflects a complex interaction between the pharmacology of the drug, the personality and expectations of the individual using the drug, and the environmental context in which the drug is used. These substances cause significant harm, although they are less commonly used than other addictive substances such as alcohol **(Chap. 464)**, nicotine **(Chap. 465)**, cannabis **(Chap. 466)**, and opioids **(Chap. 467)**. It is also important to recognize that polydrug use, involving the concurrent use of several drugs with different pharmacologic effects, is common. Sometimes one drug is used to enhance the effects of another, as with the combined use of cocaine and nicotine, or cocaine and heroin in methadone-treated patients. Some forms of polydrug use, such as the combined use of intravenous (IV) heroin and cocaine, are especially dangerous and account for many hospital emergency department visits. Cocaine and psychostimulant use (especially chronic patterns of use) may cause adverse health consequences and exacerbate preexisting disorders such as hypertension and cardiac disease. In addition, the combined use of two or more drugs may accentuate medical complications associated with use of one drug. Chronic use is often associated with immune system dysfunction and increased vulnerability to infections, including risk for HIV infection. The concurrent use of cocaine and opiates ("speedball") is frequently associated with needle sharing by people using drugs intravenously (IV). People who use IV drugs represent the largest single group of individuals with HIV infection in several major metropolitan areas in the United States as well as in many parts of Europe and Asia. Furthermore, several outbreaks of HIV in the United States since 2015 in rural and suburban areas have been attributed to clusters of injection drug use.

Psychostimulants and hallucinogens have been used for centuries to induce euphoria and alter consciousness. Hallucinogens have become popular recently, and new drugs are continually being developed. This chapter describes the subjective and adverse medical effects of cocaine, other psychostimulants including methamphetamine, 3,4-methylenedioxymethamphetamine (MDMA), and cathinones; hallucinogens such as phencyclidine (PCP), D-lysergic acid diethylamide (LSD), and *Salvia divinorum*; and emerging drugs.

PSYCHOSTIMULANTS

Psychostimulants include cocaine and methamphetamine, as well as drugs with stimulant-like properties such as MDMA and cathinones. In addition, prescribed psychostimulants such as methylphenidate, dextroamphetamine, and amphetamine are considered here.

COCAINE

Cocaine is a powerful psychostimulant drug made from the cocoa plant. It has local anesthetic, vasoconstrictor, and stimulant properties. Cocaine is a Drug Enforcement Agency (DEA) Schedule II drug, which means that it has "high potential for abuse and the potential to create psychological and/or physical dependence" but can be administered by a physician for legitimate medical uses, such as local anesthesia for some eye, ear, and throat surgeries.

Pharmacology Cocaine comes in a variety of forms, the most commonly used being the hydrochloride salt, sulfate, and a base. The salt is an acidic, water-soluble powder with a high melting point, used by snorting or sniffing intranasally or by dissolving it in water and injecting it. When used intranasally, the bioavailability of cocaine is about 60%. Cocaine sulfate ("paste") has a melting point of almost 200°C, so it has limited use, but it is sometimes smoked with tobacco. The base form can be freebase or crystallized as crack. Cocaine freebase is made by adding a strong base to an aqueous solution of cocaine and extracting the alkaline freebase precipitate. It has a melting point of 98°C and can be vaporized and inhaled. Freebase cocaine can also be crystallized and sold as crack or rock, which is also smoked or inhaled. Street dealers often dilute (or "cut") cocaine with nonpsychoactive substances such as cornstarch, talcum powder, flour, or baking soda, or adulterate it with other substances with similar effects (like procaine or amphetamine) to increase their profits. A recent concern has been the adulteration of cocaine (and other psychostimulants) with fentanyl-related opioids, resulting in overdose deaths due to opioid effects or polydrug use. Xylazine, a nonopioid sedative, analgesic, and muscle relaxant only approved for veterinary use in the United States, has also been found cut into cocaine and other psychostimulants, as described below under "Psychostimulant Clinical Manifestations".

Given the extensive pulmonary vasculature, smoked or vaporized cocaine reaches the brain very quickly, similar in speed of onset to injected cocaine. The result is a rapid, intense, transient high, which enhances its addictive potential. Cocaine binds to the dopamine (DA) transporter and blocks DA reuptake, which increases synaptic levels of the monoamine neurotransmitters DA, norepinephrine (NE), and serotonin (5HT), in both the central nervous system (CNS) and the peripheral nervous system (PNS). Use of cocaine, like other harmful drugs abuse, induces long-term changes in the brain. Animal studies have shown adaptations in neurons that release the excitatory neurotransmitter glutamate after cocaine exposure.

Epidemiology According to the National Survey on Drug Use and Health (NSDUH), in 2023 an estimated 5 million people aged 12 years or older (1.8% of the population) were past-year consumers of cocaine, including crack. Among those, 470,000 used cocaine for the first time (1287 cocaine initiates/day) including 23,000 adolescents aged 12–17 years. About 1.3 million people aged 12 years or older (0.4% of the population) in 2023 had a cocaine use disorder. According to the Centers for Disease Control and Prevention (CDC) National Center for Health Statistics, the age-adjusted rate of drug overdose deaths involving cocaine more than quintupled from 1.5 deaths per 100,000 standard population in 2011 to 8.2 in 2022. The rate of drug overdose deaths per 100,000 standard population involving both cocaine and opioids in 2021 (5.9) was 7.4 times the rate in 2011 (0.8) and was driven by the involvement of synthetic opioids including fentanyl and fentanyl analogues. The rate involving cocaine without opioid co-involvement in 2021 (1.5) was 2.1 times the rate in 2011 (0.7). The data are concerning because they describe increases over time in cocaine-related overdose both with and without the co-ingestion of opioids.

■ METHAMPHETAMINE

Methamphetamine is a psychostimulant drug usually used as a white, bitter-tasting powder or a pill. Crystal methamphetamine is a form of the drug that looks like glass fragments or shiny, bluish-white rocks. It can be inhaled/smoked, swallowed (pill), snorted, or injected (after being dissolved in water or alcohol).

Pharmacology When smoked, methamphetamine exhibits 90.3% bioavailability, compared to 67.2% for oral ingestion. Methamphetamine exists in two stereoisomers, the L- and D-forms. D-Methamphetamine, or the dextrorotatory enantiomer, is a more powerful psychostimulant, with 3–5 times the CNS activity as compared with L-methamphetamine. Methamphetamine is a cationic lipophilic molecule, which stimulates the release, and partially blocks the reuptake, of newly synthesized catecholamines in the CNS. Methamphetamine has a similar structure to the DA, NE, 5HT, and vesicular monoamine transporters and reverses their endogenous function, resulting in release of monoamines from storage vesicles into the synapse. Methamphetamine also attenuates the metabolism of monoamines by inhibiting monoamine oxidase.

Methamphetamine is more potent than amphetamine, resulting in much higher concentrations of synaptic DA and more toxic effects on nerve terminals. Outside the medical context, methamphetamine's pharmacokinetics and low cost often result in a chronic and continuous, high-dose self-administered use pattern.

Epidemiology According to the NSDUH, in 2023, 2.6 million people aged 12 years or older (0.9% of the population) used methamphetamine (not including use or misuse of prescription amphetamines or other stimulants) in the past year; of those, 78,000 used methamphetamine for the first time (213 people per day). In 2023, an estimated 1.8 million people aged 12 years or older (0.6% of the population and 69% of those with past-year use) had a methamphetamine use disorder. High rates of co-occurring substance use and mental illness exist in adults who use methamphetamine, and only about one-third of adults with past-year methamphetamine use disorder received addiction treatment. Methamphetamine availability and methamphetamine-related harms (e.g., overdose deaths, treatment admissions, infectious disease transmission) continue to increase in the United States. According to CDC data, psychostimulants (primarily methamphetamine) caused 36,251 overdose deaths in 2023. Stimulant-involved overdose deaths have risen markedly in recent years; with rates of psychostimulant overdose deaths increasing from 0.7 in 2011 to 10.4 in 2022. Further, while preliminary reports show overdose deaths involving opioids decreasing from an estimated 84,181 in 2022 to 81,083 in 2023, overdose deaths due to cocaine and psychostimulants (like methamphetamine) increased.

■ MDMA AND CATHINONES

MDMA also known as Molly, ecstasy, or X, is an illegal synthetic drug that has stimulant and psychedelic effects. Khât is a plant found in East Africa and the Middle East; it has been used for centuries for its mild stimulant-like effect. Synthetic cathinones or "bath salts" are manufactured psychostimulants that are chemically similar to the naturally occurring substance cathinone found in the khât plant and are discussed under "Emerging Drugs" below.

MDMA Molly, slang for "molecular," refers to the crystalline powder form of MDMA usually sold as powder or in capsules. The content of Molly varies and is often not MDMA at all but rather contains methylone or ethylone, which are synthetic substances commonly found in so-called bath salts and pose significant health risks. The clinician should always consider the possibility that the drug reported by the individual may be unwittingly contaminated with other substances.

With MDMA use, individuals experience increased physical and mental energy, distortions in time and perception, emotional warmth, empathy toward others, a general sense of well-being, decreased anxiety, and an enhanced enjoyment of tactile experiences. MDMA is usually taken orally in a tablet, capsule, or liquid form with first effect at 45 min on average, peak effect at 1–2 h, and duration ~3–6 h. MDMA binds to serotonin transporters and increases the release of serotonin, NE, and DA. Research in animals has shown that MDMA in moderate to high doses can cause loss of serotonin-containing nerve endings and permanent damage. MDMA is a Schedule I drug, along with other substances with no proven therapeutic value. MDMA has been given a breakthrough therapy designation by the U.S. Food and Drug Administration (FDA) as a possible treatment for posttraumatic stress disorder allowing for expedited clinical trials, but at present MDMA remains a Schedule I drug, along with other substances with no proven therapeutic value.

Adulteration of MDMA tablets with methamphetamine, ketamine, caffeine, the over-the-counter cough suppressant dextromethorphan (DXM), the diet drug ephedrine, and cocaine is common. MDMA is rarely used alone and is often mixed with other substances, such as alcohol and marijuana, making the specific impacts of its use difficult to ascertain. According to NSDUH 2023 data, among individuals aged 12 or older, 0.8% used MDMA in the past year and 507,000 people tried MDMA for the first time (>1300 per day). MDMA is predominantly used by men 18–25 years of age, with use typically beginning at age 21 years. There is evidence that gay or bisexual men and women are more likely than their heterosexual counterparts to have used MDMA in the last 30 days.

Cathinone This is an alkaloid psychostimulant structurally similar to amphetamine found in the khât (*Catha edulis*) plant, which grows at high altitudes in East Africa and the Middle East and whose leaves are chewed for their mild stimulant-like effect. The extraction of cathinone and other alkaloids from the leaves by chewing is very effective, leaving little as unabsorbed residue. The leaves and twigs can also be smoked, infused in tea, or sprinkled on food. Cathinone increases dopamine release and reduces dopamine reuptake.

Originally limited to its area of cultivation, with advances in rapid transportation and postal delivery, khât is now available in several continents including Europe and North America. Worldwide it is estimated that 10 million people chew khât, including up to 80% of all adults in some areas where the evergreen shrub is indigenous. In regions where the plant is indigenous, there have also been reports of khât use as a study aid among university students. Cathinone is a Schedule I drug in the United States, making its possession and use illegal.

■ PRESCRIBED PSYCHOSTIMULANTS

Methylphenidate, dextroamphetamine, and dextroamphetamine/amphetamine combination products are psychostimulants approved in the United States for treatment of attention-deficit hyperactivity disorder (ADHD), weight control, and narcolepsy. Prescription psychostimulants increase alertness, attention, and energy. Phenylpropanolamine, a psychostimulant used primarily for weight control, was found to be related to hemorrhagic stroke in women and removed from the market in 2005. Nonprescribed amphetamines and methylphenidate are used quite frequently by college students and as energy and productivity boosters by others. According to the 2023 NSDUH, past-year prescription stimulant misuse was reported by 3.9 million (1.4%) people aged 12 years or older. Of note, in 2023 among individuals aged

12 or older, 786,000 individuals misused prescription stimulants and cocaine; 190,000 misused prescription stimulants and methamphetamine; and 182,000 people misused or used all three. Past-year initiates of prescription stimulant misuse totaled 712,000, which averages to ~1950 people misusing prescription stimulants for the first time each day, including >1000 young adults aged 12–25 each day. Among people aged 12 years or older, 0.6% of the population in 2023 had a prescription stimulant use disorder in the past year.

PSYCHOSTIMULANT CLINICAL MANIFESTATIONS

Psychostimulants produce the same acute CNS effects: euphoria/elevated mood, increased energy/decreased fatigue, reduced need for sleep, decreased appetite, heightened sense of alertness, decreased distractibility, dose-dependent effects on focus, attention, and curiosity, increased self-confidence, increased libido, and prolonged orgasm, independent of the specific psychostimulant or route of administration. Peripheral effects may include tremor, diaphoresis, hypertonia, tachypnea, hyperreflexia, and hyperthermia. Many of the effects are biphasic; for example, low doses improve psychomotor performance, while higher doses may cause tremors or convulsions. α-Adrenergically mediated cardiovascular effects are also biphasic, with low doses resulting in increased vagal tone and decreased heart rate and high doses causing increased heart rate and blood pressure. Psychostimulant use can result in restlessness, irritability, and insomnia and, at higher doses, suspiciousness, repetitive stereotyped behaviors, and bruxism. Endocrine effects resulting from chronic use may include impotence, gynecomastia, menstrual function disruptions, and persistent hyperprolactinemia **(Table 468-1)**.

Overdose presents as sympathetic nervous system overactivity with psychomotor agitation, hypertension, tachycardia, headache, and mydriasis, and can lead to convulsions, cerebral hemorrhage or infarction, cardiac arrhythmias or ischemia, respiratory failure, or rhabdomyolysis. It is a medical emergency; treatment is largely symptomatic and should occur in an intensive care or telemetry unit. Inhalation of crack cocaine that is vaporized at high temperatures can cause airway burns, bronchospasm, and other symptoms of pulmonary disease. MDMA has also been shown to raise body temperature and can occasionally result in liver, kidney, or heart failure, or even death.

Psychostimulants are often used with other drugs, including opioids and alcohol, whose CNS-depressant effects tend to attenuate psychostimulant-induced CNS stimulation. These combinations often have additive deleterious effects, increasing the risk of morbidity and mortality. An example of this risk is the use of cocaine with alcohol, which results in the metabolite cocaethylene. Cocaethylene's effects on the cardiovascular system are additive to that of cocaine's effects, resulting in intensified pathophysiologic consequences.

Adulteration of psychostimulants, particularly cocaine, with other drugs is common and can have additional health consequences. In addition to contamination with fentanyl-related compounds, potentially resulting in fatal overdose, multiple other substances have been noted as contaminants of psychostimulants. Levamisole, an anthelminthic and immunomodulator used primarily in veterinary medicine, has been found in cocaine and can cause agranulocytosis, leukoencephalopathy, and cutaneous vasculitis, which has resulted in skin necrosis. Clenbuterol, a sympathomimetic amine used clinically as a bronchodilator, has also been found in cocaine and can result in tachycardia, hyperglycemia, palpitations, and hypokalemia. Xylazine, a nonopioid veterinary sedative, analgesic, and muscle relaxant, is most often described in the context of an opioid adulterant; however, it is also seen as an adulterant of cocaine, methamphetamine, and other stimulants. Effects associated with xylazine include dry mouth, drowsiness, hypertension, and tachycardia followed by hypotension and bradycardia, hyperglycemia, hypothermia, coma, respiratory depression, and dysrhythmia. Xylazine injection has been associated with necrotic soft tissue lesions both at the site of injection and elsewhere on the body. There is some evidence that xylazine itself can result in withdrawal symptoms such as sharp chest pains and seizures and cause physical dependence. Studies in Europe have found that, in addition to levamisole, some of the most common adulterants in cocaine include

TABLE 468-1 Complications of Psychostimulant Use

Cardiovascular	Acute • Arterial vasoconstriction • Thrombosis • Tachycardia • Hypertension • Increased myocardial oxygen demand • Increased vascular shearing forces • Coronary vasoconstriction • Cardiac ischemia • Left ventricular dysfunction/heart failure (high blood concentrations) • Supraventricular and ventricular dysrhythmias • Aortic dissection/rupture Chronic • Accelerated atherogenesis • Left ventricular hypertrophy • Dilated cardiomyopathy
Central and peripheral nervous systems	• Hyperthermia • Psychomotor agitation • Tremor • Hyperreflexia • Hypertonia • Headache • Seizures • Coma • Intracranial hemorrhage • Focal neurologic symptoms
Pulmonary	• Angioedema (inhaled) • Pharyngeal burns (inhaled) • Pneumothorax • Pneumomediastinum • Pneumopericardium • Reversible airway disease exacerbations • Bronchospasm • Shortness of breath ("crack lung") • Tachypnea • Pulmonary infarction
Gastrointestinal	• Perforated ulcers • Ischemic colitis • Bowel infarction • Impaction (body packing) • Hepatic enzyme elevation
Renal	• Metabolic acidosis • Renal infarction • Rhabdomyolysis
Endocrine	• Impotence • Gynecomastia • Menstrual function disruptions • Hyperprolactinemia
Other	• Diaphoresis • Irritability • Insomnia • Bruxism • Stereotypy • Splenic infarction • Acute angle-closure glaucoma • Vasospasm of the retinal vessels (unilateral or bilateral vision loss) • Mydriasis • Madarosis • Abruptio placentae

phenacetin, lidocaine, caffeine, diltiazem, hydroxyzine, procaine, tetracaine, paracetamol, creatine, and benzocaine.

Withdrawal from psychostimulants often includes hypersomnia, increased appetite, and depressed mood. Acute withdrawal typically lasts 7–10 days, but residual symptoms, possibly associated with neurotoxicity, may persist for several months. Debate remains whether psychostimulant withdrawal symptoms decline monotonically or occur in discrete phases, becoming worse before they improve. Psychostimulant withdrawal is not thought to be a major driver of ongoing use. Most current theories of psychostimulant addiction emphasize the primary role of conditioned craving, which can persist long after physiological withdrawal has abated. Conditioned craving includes the urge to use drugs in response to cues in the environment associated with drug use, such as associates who use drugs, drug paraphernalia, or drug-using locations.

Injection of psychostimulants places people at increased risk of contracting infectious diseases from exposure to HIV and hepatitis B or C in blood or other bodily fluids, as well as skin abscesses and endocarditis. Psychostimulant use can also increase risk for infection by causing altered judgment and decision-making, leading to risky behaviors such as unprotected sex. There is some evidence that psychostimulant use may worsen the progression of HIV/AIDS via increased injury to nerve cells exacerbating cognitive problems.

The actions and effects of khât are like those of other psychostimulants. Short-term effects include euphoria, increased alertness and arousal, loss of appetite, insomnia, headaches, and tremors. Long-term use may result in gastrointestinal disorders such as constipation, ulcers, and stomach inflammation, as well as increased risk for acute myocardial infarction and stroke due to inotropic and chronotropic effects on the heart, vasospasm of coronary arteries, and catecholamine-induced platelet aggregation. There is evidence that, rarely, heavy khât use may cause mild to moderate psychological dependence. Compulsive use has been described, with resulting grandiose delusions, paranoia, and hallucinations. A mild withdrawal syndrome from khât can include depression, nightmares, low blood pressure, and lack of energy.

DIAGNOSIS

The *Diagnostic and Statistical Manual of Psychiatric Disorders*, 5th edition (DSM-5) defines a stimulant use disorder (SUD) as a pattern of use of amphetamine-type substances, cocaine, or other stimulants leading to clinically significant impairment or distress, as manifested by at least 2 of the following 11 problems within a 12-month period: taking larger amounts, or over a longer period of time, than intended; persistent desire or unsuccessful efforts to reduce or control use; a great deal of time spent in activities necessary to obtain, use, or recover; craving; use resulting in failure to fulfill major role obligations; continued use, despite recurrent social or interpersonal problems; giving up social, occupational, or recreational activities; recurrent use in physically hazardous situations; continued use despite persistent or recurrent physical or psychological problems; tolerance; and withdrawal symptoms, or avoidance of withdrawal symptoms, by continued use.

The International Classification of Diseases (ICD) 10th Revision (ICD-10) recognizes "stimulant dependence syndrome" and "stimulant withdrawal state," and the ICD 11th Revision (ICD-11) further specifies the definition to "stimulant dependence including amphetamines, methamphetamines, or methcathinone."

TREATMENT

Acute Intoxication

As with all emergency situations, the first task is to check a patient's circulation, airway, and breathing. With cocaine use, succinylcholine is relatively contraindicated in rapid-sequence intubation; consider rocuronium (1 mg/kg IV) or another nondepolarizing agent as an alternative. If psychomotor agitation occurs, rule out hypoglycemia and hypoxemia first, and then administer benzodiazepines (e.g., diazepam 10 mg IV and then 5–10 mg IV every 3–5 hours until agitation controlled). Benzodiazepines are usually sufficient to address cardiovascular side effects. Severe or symptomatic hypertension can be treated with phentolamine, nitroglycerin, or nitroprusside. Hyperthermic patients should be cooled within ≤30 min with the goal to achieve a core body temperature of <39°C (102°F). Evaluation of chest pain in someone using cocaine should include an electrocardiogram, chest radiograph, and biomarkers to exclude myocardial infarction. The treatment approach is similar to nonstimulant-induced chest pain; however, it is recommended that whenever possible beta blockers not be used in people who use cocaine. The concern arises from the potential unopposed alpha-adrenergic stimulation that results from beta blockade possibly causing coronary arterial vasoconstriction, ischemia, and infarction and limited data supporting the benefit of beta blockers in cocaine-related cardiovascular complications. If beta blockers are to be given, it is suggested that mixed alpha/beta blockers, e.g., labetalol and carvedilol, be used rather than nonselective beta blockers, and only in situations where the benefits outweigh the risks. Because many instances of psychostimulant-related mortality have been associated with concurrent use of other illicit drugs (particularly opioids), the physician must be prepared to institute effective emergency treatment for multiple drug toxicities.

Psychostimulant Use Disorders

Treatment of psychostimulant use disorders requires the combined efforts of primary care physicians, addiction medicine physicians, psychiatrists, and psychosocial care providers. Early abstinence from psychostimulant use is often complicated by symptoms of depression and guilt, insomnia, and anorexia, which may be as severe as those observed in major affective disorders and can last for months and even years after use has stopped.

Behavioral therapies, including cognitive-behavioral therapy (CBT), the community reinforcement approach (CRA), contingency management (CM; providing structured and specific incentives to patients who remain substance free), motivational enhancement therapy (MET), combinations of these, and others, remain the mainstay of treatment for SUDs and show modest benefit. These behavioral therapies are designed to help modify the patient's thinking, expectancies, and behaviors, and to increase life-coping skills, with behavioral interventions to support long-term, drug-free recovery. There is robust evidence, including recent meta-analyses and systematic reviews, that contingency management, when implemented with fidelity to the principles of operant conditioning, is a highly effective treatment for psychostimulant use disorder, the benefit of which continues for up to 2 years beyond treatment discontinuation.

There are no FDA-approved medications for psychostimulant addiction. Current research includes several neurotransmitter-based strategies targeting DA, serotonin, γ-aminobutyric acid (GABA), and glutamate. Trials of agonist therapy with longer-acting psychostimulant medications such as dexamphetamine and methylphenidate have not been conclusive. Studies with the antidepressants mirtazapine, bupropion, sertraline, imipramine, and atomoxetine have been equivocal, as have studies with the atypical antipsychotic aripiprazole and the anticonvulsant topiramate. Other therapies being studied for the treatment of psychostimulant use disorder include acamprosate (possibly via a role in modulating the NMDA receptor), galantamine (reversible acetylcholine esterase inhibitor, which may strengthen impulse control, as well as cognitive and social abilities depleted by long-term psychostimulant use), naltrexone (opiate receptor antagonist), doxazosin (alpha-adrenergic antagonist), and varenicline (partial agonist of the $\alpha 4\beta 2$ nicotinic acetylcholine receptor and DA neurotransmission enhancer). Overall, it is promising that some of the medications studied showed significant outcome improvements over placebo, but many were also underpowered due to issues of small sample size, sample bias, low participant retention, and low treatment adherence rates. Ongoing studies are investigating lisdexamfetamine (a dexamphetamine prodrug), a combination of extended-release naltrexone with bupropion, pomaglumetad (a glutamate agonist), and several monoclonal antibodies. Special attention needs to be paid to the inclusion of underrepresented populations including women in future stimulant use disorder medication trials. Vaccines for cocaine and methamphetamine use disorders are also being developed. Finally, recent preliminary studies have brought attention to the potential use of brain stimulation techniques such as transcranial magnetic stimulation (TMS), theta-burst stimulation (TBS), and transcranial direct current stimulation (tDCS) to treat psychostimulant use disorders, although further studies will be required to determine their value, if any, in this situation.

HALLUCINOGENS

Hallucinogens are a diverse group of drugs causing alteration of thoughts, feelings, sensations, and perceptions. Some hallucinogens are found naturally in plants and mushrooms, while others are synthetic. They include ayahuasca (a tea made from Amazonian plants containing dimethyltryptamine [DMT], the primary mind-altering ingredient), DMT (aka Dimitri; can also be synthesized in a lab), LSD (clear or white odorless material made from lysergic acid found in rye and other grain fungus); peyote (mescaline, derived from a small, spineless cactus or made synthetically); and 4-phosphoryloxy-*N,N*-dimethyltryptamine (psilocybin, comes from certain South and North American mushrooms).

A subgroup of hallucinogens produces the added sensation of feeling out of control or disconnected from one's body or surroundings. These dissociative drugs include DXM (an over-the-counter cough suppressant, when used in high doses), ketamine (an FDA-approved human and veterinary anesthetic and a nasal spray [esketamine] for treatment-resistant depression); phencyclidine (PCP; a cyclohexylamine derivative and dissociative anesthetic); and *S. divinorum* (salvia, a Mexican, Central American, and South American plant). Dissociative drugs distort the way the user perceives time, motion, color, sound, and self, and their use can lead to bizarre and dangerous behavior and cause respiratory depression, heart rate abnormalities, and a withdrawal syndrome including drug craving, confusion, headache, and sweating.

Use of hallucinogens in religious and spiritual rituals goes back centuries, and they are ingested in a wide variety of ways, including orally, by smoking, intranasally, and transmucosally. Especially when taken orally, the onset of action of hallucinogens is within 20–90 min and the duration of action can be as long as 6–12 h, except for salvia, whose effects generally last about 30 min. Hallucinogens specifically disrupt the neurotransmitters serotonin and glutamate. Effects on the serotonin system can disturb mood, sensory perception, sleep, appetite, body temperature, sexual behavior, and muscle control. Glutamate system effects include perturbations in pain perception, responses to the environment, emotion, and learning and memory.

According to the NSDUH, in 2023, 2.6 million adults reported past-month use of hallucinogens and 8.8 million (3.1% of the population) reported past-year use of hallucinogens. Of these, 1.5 million used hallucinogens for the first time. Of note, these statistics include ecstasy (MDMA or "Molly") in the overall hallucinogen use category as well as LSD, PCP, peyote, mescaline, psilocybin mushrooms, ketamine, N,N-dimethyltryptamine (DMT)/Alpha-Methyltryptamine (AMT)/"Foxy," and *S. divinorum*. Past-year initiation numbers among people aged 12 years and older include 364,000 for LSD, 24,000 for PCP, and 507,000 for ecstasy. In 2023, 0.2% of people aged 12 or older had a hallucinogen use disorder.

Clinical manifestations of hallucinogen use include false sensory experiences (i.e., hallucinations), intensified feelings, heightened sensory experiences, and time perturbations. Additional physiologic responses include nausea; increases in heart rate, blood pressure, respiratory rate, or body temperature; loss of appetite; xerostomia; sleep problems; synesthesia; impaired coordination; and hyperhidrosis. Extremely negative experiences with hallucinogen use (the "bad trip") can include panic, paranoia, and psychosis, which may persist for up to 24 h. Such experiences are best treated with supportive reassurance, but benzodiazepines (e.g., diazepam 10 mg or lorazepam if liver damage is present) may be administered if agitation is severe. There is some evidence that chronic effects of hallucinogen use can occur, including persistent psychosis, memory loss, anxiety, depression, and flashbacks. Long-term effects of PCP and other dissociative drug use can include persistent speech difficulties, memory loss, depression, suicidal thoughts, anxiety, and social withdrawal that may persist for a year or more after chronic use stops.

The FDA issued breakthrough therapy designation for MDMA to expedite research into treatment of posttraumatic stress disorder and additional breakthrough therapy designations for two formulations of psilocybin for treatment of depression. In addition, some hallucinogens are being studied as potential treatment for certain SUDs, including psilocybin for alcohol and tobacco use disorders and ketamine for cocaine and methamphetamine use disorders. There is also some evidence that psilocybin and LSD may relieve pain in certain chronic pain conditions such as cluster headache, lower back pain, cancer-related pain, and phantom limb pain; studies are ongoing.

The DSM-5 defines hallucinogen use disorder as meeting 2 or more of the first 10 criteria (see above for SUD) in the past 12 months. The withdrawal criterion does not apply to hallucinogens, because hallucinogen use disorder is atypical in that use patterns are generally not chronic. There are currently no FDA-approved medications for the treatment of hallucinogen addiction. Research on behavioral treatments for hallucinogen addiction is underway.

EMERGING DRUGS

With the aid of the Internet and some basic over-the-counter (and other) ingredients, the rise of the "kitchen chemist" is upon us. The production of new psychoactive substances (NPS), such as recreationally manufactured synthetic cathinones (bath salts) and synthetic cannabinoids (K2, spice), is on the rise and has resulted in the use of unregulated psychoactive substances that are intended to copy the effects of more expensive illegal drugs such as methamphetamine and cocaine. NPS also include recreationally manufactured synthetic opioids containing brorphine and U-47700 and recreationally manufactured synthetic benzodiazepines such as bromazepams, desalkylgidazepam, and flubromazepam. In addition to NPS, nitazines (a synthetic opioid that can be more powerful than fentanyl) and tianeptine (a non-U.S.-approved antidepressant with opioid-like effects at high doses) are emerging and reemerging in the drug supply both alone and mixed into other drugs. These emerging drugs can be found online or sold in drug markets or convenience stores. Depending on the type of substance, whether a new type of opioid, depressant, synthetic cannabinoid, psychedelic, or stimulant, the effects will differ and may be unpredictable and unwanted, especially if unwittingly ingested as an adulterant in another drug. In addition, emerging substances are often not included in emergency department drug tests and are not routinely tested for when determining the cause of death after a fatal overdose.

Synthetic cathinones (bath salts) are human-made drugs chemically similar to cathinone found in khât and are often stronger and more dangerous than the natural product. They usually take the form of a white or brown crystal-like powder, packaged in small plastic or foil bundles labeled "not for human consumption," or as "plant food," "jewelry cleaner," or "phone screen cleaner," and sold online and in drug paraphernalia stores. The popular nickname Molly (slang for "molecular") often refers to the purported "pure" crystalline powder form of MDMA, usually sold in capsules. However, people who purchase powder or capsules sold as Molly often actually receive other drugs, such as synthetic cathinones. The uncertainty of what is actually in these synthetic products, whose components might change from batch to batch, makes them even more dangerous, as anyone using them is unaware of what the products actually contain and how they might respond.

The three most common synthetic cathinones are mephedrone, methylone, and MDPV (3,4-methylenedioxypyrovalerone). With oral ingestion, these drugs have an onset of action from 15–45 min, and a duration that varies from 2–7 h. Studies have found that MDPV affects the brain in a manner similar to cocaine but is at least 10 times more potent. MDPV is the most common synthetic cathinone found in the blood and urine of patients admitted to emergency departments after taking “bath salts.” High doses, or chronic use, of synthetic cathinones can lead to dangerous medical consequences, including psychosis, violent behaviors, tachycardia, hyperthermia, and even death.

The ability to synthesize addictive and dangerous drugs relatively simply and rapidly, changing just a few molecules, yet retaining the effects, has allowed many of these emerging drugs to outpace efforts to regulate them, resulting in a developing global public health concern.

SUBSTANCE USE AND MENTAL HEALTH

According to the NSDUH, in 2023, among adults aged 18 and older with no mental illness, 21% consumed illicit drugs, compared to 51.9% with severe mental illness and 42.4% with any mental illness. In 2023, among adults 18 years of age or older, 84.5 million people had either any mental illness or an SUD in the past year, 38.2 million had any mental illness in the absence of an SUD, 25.8 million had an SUD and no mental illness, and 20.4 million (7.2% of the population) had both. The percentage of adults aged 18 or older in 2023 who had both any mental illness and an SUD in the past year was highest among young adults aged 18 to 25 (14.1% or 4.8 million people). The percentage of adults aged 18 or older in 2023 who had both any mental illness and an SUD in the past year was higher among multiracial adults (13.3%) than among white (8.4%), black (7.8%), Hispanic (7.1%), or Asian adults (3.5%); the percentage could not be calculated with sufficient precision for Native Hawaiian or other Pacific Islander adults.

Among the 20.4 million adults aged 18 or older in 2023 with co-occurring any mental illness and an SUD in the past year, 62.4% (or 12.8 million people) received either substance use treatment or mental health treatment in the past year, and 37.6% (or 7.7 million people) received neither type of treatment. This equates to about two in five adults aged 18 or older with co-occurring any mental illness and an SUD in the past year who did not receive treatment for either condition. Taken together, these data point to the significant overlap of substance-related and other mental health problems and highlight the prodigious treatment gap that exists for both.

GLOBAL CONSIDERATIONS

After nicotine, alcohol, and cannabis, stimulants are the next most used drugs globally, worldwide in 2021, an estimated 22 million people used cocaine and 36 million people used amphetamines. The global cocaine supply reached a record high in 2022 with >2700 tons of cocaine produced, 20% more than the previous year, with main trafficking from the Andean region to other countries in the Americas and to Western and Central Europe. The two largest emerging methamphetamine markets in recent years have been the Near and Middle East/Southwest Asia and Southeastern Europe. The trafficking and use of synthetic stimulants, mainly cathinones, has risen notably in Central Asia and Eastern Europe. Globally, psychostimulant use has been associated with elevated mortality, increased incidence of HIV and hepatitis C infection, poor mental health (suicidality, psychosis, depression, and violence), and increased risk of cardiovascular events. The World Health Organization estimates that in 2019 global deaths attributed to cocaine and amphetamine use were 26,082 and 46,661, respectively.

Worldwide 7.4 million individuals have a stimulant use disorder, and the United Nations Office on Drugs and Crime (UNODC) estimates that only one in seven people with SUDs receives treatment. The number receiving treatment is much lower in individuals with stimulant use disorder despite the fact that cocaine treatment demand alone has risen almost 60% from 2011 to 2022 in subregions in Europe.

Globally, stigma and marginalization make treatment of drug use disorders difficult and hinder sustainable inclusive development incorporating gender and racial equity and the empowerment of women and underrepresented minorities. The existing treatment gap is further magnified when the intersectionality of gender, race, age, and ethnicity is considered, as is the treatment gap faced by populations with housing instability, low socioeconomic status, or low educational attainment, as well as LGBTQIA+, and veterans, among others.

FUTURE DIRECTIONS

Despite their prevalence and public health impact, psychostimulant and hallucinogen use disorders have no FDA-approved treatment medications. While behavioral therapies, such as contingency management and CBT, have been shown effective in psychostimulant use disorders, further research needs to be done regarding their utility for hallucinogen use disorders. Based on experience with opioid and alcohol use disorders, it is also likely that the most efficacious treatments will employ a combination of behavioral and pharmacologic therapy. Research on medications to treat psychostimulant use disorder is ongoing. Additionally, new approaches that utilize emerging technologies have considerable potential for future treatment of psychostimulant use disorders. These include neurostimulation/neuromodulation (TMS, TBS, tDCS), wearable biosensors, and mobile technology, including ecologic and geographic momentary assessment (EMA/GMA), as well as real-time interventions delivered via smartphone or other mobile devices.

FURTHER READING

Centers for Disease Control and Prevention: A stimulant guide: Answers to emerging questions about stimulants in the context of the overdose epidemic in the United States. National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, 2022.

Compton WM: Polysubstance use in the U.S. opioid crisis. Mol Psychiatry 26:41, 2021.

Farrell M et al: Responding to global stimulant use: Challenges and opportunities. Lancet 394:1652, 2019.

Substance Abuse and Mental Health Services Administration: Treatment for Stimulant Use Disorders. Treatment Improvement Protocol (TIP) Series 33. SAMHSA Publication No. PEP21-02-01 004. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2021.

Trivedi MH et al: Bupropion and naltrexone in methamphetamine use disorder. N Engl J Med 384:140, 2021.

WEBSITES

American Society of Addiction Medicine: *https://www.asam.org/public-resources*

National Institute on Drug Abuse: *https://www.drugabuse.gov/drugs-abuse*

Substance Abuse and Mental Health Services Administration: *https://www.samhsa.gov*

World Health Organization: *http://www.who.int/substance_abuse/en/*

This page intentionally left blank

469 Heavy Metal Poisoning

Howard Hu

Toxic metals (hereafter referred to simply as "metals") pose a significant threat to health through low-level as well as high level environmental and occupational exposures. One indication of their importance relative to other potential hazards is their ranking by the U.S. Agency for Toxic Substances and Disease Registry, which maintains an updated list of all hazards present in toxic waste sites according to their prevalence and the severity of their toxicity. The first, second, third, and seventh hazards on the list are heavy metals: arsenic, lead, mercury, and cadmium, respectively (*http://www.atsdr.cdc.gov/spl/*), positions they have held for many years. All four are also listed by the World Health Organization as among the 10 chemicals of highest concern in relation to health, globally. Specific information pertaining to each of these four metals, including sources and metabolism, toxic effects produced, diagnosis, and the appropriate treatment for poisoning, is summarized in **Table 469-1**.

Metals are inhaled primarily as dusts and fumes (the latter defined as tiny particles generated by combustion). Metal poisoning can also result from exposure to vapors (e.g., mercury vapor in creating dental amalgams). When metals are ingested in contaminated food or drink or by hand-to-mouth activity (implicated especially in children), their gastrointestinal absorption varies greatly with the specific chemical form of the metal and the nutritional status of the host. Once a metal is absorbed, blood is the main medium for its transport, with the precise kinetics dependent on diffusibility, protein binding, rates of biotransformation, availability of intracellular ligands, and other factors. Some organs (e.g., bone, liver, and kidney) sequester metals in relatively high concentrations for years. Most metals are excreted through renal clearance and gastrointestinal excretion; some proportion is also excreted through salivation, perspiration, exhalation, lactation, skin exfoliation, and loss of hair and nails. The intrinsic stability of metals facilitates tracing and measurement in biologic material, although the clinical significance of the levels measured is not always clear.

Some metals, such as copper and selenium, are essential to normal metabolic function as trace elements **(Chap. 344)** but are toxic at high levels of exposure. Others, such as lead and mercury, are xenobiotic and theoretically are capable of exerting toxic effects at any level of exposure. Indeed, much research is currently focused on the contribution of low-level xenobiotic metal exposure to subtle changes in health, chronic diseases, and/or increased risks of adverse events such as heart attacks, stroke, or cancer that may have significant public health consequences given the widespread nature of their exposures. Metals are well-known to be able to cause toxicity through the inhibition of enzymes, damage to subcellular organelles and DNA, covalent modification of proteins, formation of reactive oxygen species, and displacement of essential metals in metal dependent proteins. Recent research has also elucidated how metals may affect DNA methylation and other changes to the epigenome, thereby resulting in altered gene expression as well as the dysregulated expression of microRNAs. Genetic factors, such as polymorphisms that encode for variant enzymes with altered properties in terms of metal binding, transport, and effects, may modify the impact of metals on health. Nutritional status, comorbidities, and other factors can also contribute to variations in individual susceptibility to metal effects.

The most important component of treatment for metal toxicity is the recognition and termination of exposure. *Chelating agents* are used to bind metals into stable cyclic compounds with relatively low toxicity and to enhance their excretion. The principal chelating agents are dimercaprol (British anti-Lewisite [BAL]), ethylenediamine tetraacetic acid (EDTA), succimer (dimercaptosuccinic acid [DMSA]), and penicillamine; their specific use depends on the metal involved and the clinical circumstances. For the most part, they are reserved for treating acute, symptomatic metal toxicity. Activated charcoal does not bind metals and thus is of limited usefulness in cases of acute metal ingestion.

In addition to the information provided in Table 469-1, several other aspects of exposure, toxicity, or management are worthy of discussion with respect to the four most hazardous toxicants (arsenic, cadmium, lead, and mercury).

Arsenic, even at moderate levels of exposure, has been clearly linked with increased risks for cancer of the skin, bladder, renal pelvis, ureter, kidney, liver, and lung. These risks appear to be modified by smoking, folate and selenium status, genetic traits (such as ability to methylate arsenic), and other factors. Based on the accumulation of evidence, the American Heart Association concluded that low-level exposures to arsenic, lead, and cadmium are risk factors for hypertension, subclinical atherosclerosis, coronary artery stenosis, and calcification as well as ischemic heart disease and stroke, left ventricular hypertrophy and heart failure, and peripheral artery disease. Recent studies in community-based populations have also generated strong evidence that arsenic exposure is a risk factor for lung function impairment, acute respiratory tract infections, respiratory symptoms, and nonmalignant lung disease mortality. The association with cardiovascular disease may hold at levels of exposure in drinking water that are below the World Health Organization (WHO) provisional guideline value of 10 µg/L. Evidence has also continued to build indicating that low-level arsenic is a likely cause of neurodevelopmental delays in children and likely contributes to the development of type 2 diabetes and possibly nonalcoholic fatty liver disease and cirrhosis.

Serious *cadmium* poisoning from the contamination of food and water by mining effluents in Japan contributed to the 1946 outbreak of "itai-itai" ("ouch-ouch") disease, so named because of cadmium-induced bone toxicity that led to painful bone fractures. Modest exposures from environmental contamination have been associated in a growing number of studies with lower bone density, higher incidence of fractures, and faster decline in height in both men and women, effects that may be related to cadmium's calciuric and other toxic effects on the kidney. Cadmium burdens have also been associated with an increased risk of long-term kidney graft failure, and there is evidence for synergy between the adverse impacts of cadmium and lead on kidney function. Environmental exposures have also been linked to lower lung function (even after adjusting for smoking cigarettes, which contain cadmium) and, as noted earlier, the American Heart Association considers cadmium as a risk factor for cardiovascular disease and mortality, stroke, and heart failure. Cadmium triggers pulmonary inflammation, and recent population-based studies of U.S. adults found that higher cadmium burdens are associated with higher mortality from influenza, pneumonia, and COVID-19. The International Agency for Research on Cancer has classified cadmium as a known carcinogen, with evidence indicating it contributes to elevated risks of prostate, lung, breast, and endometrial cancer. Overall, this growing body of research indicates that cadmium exposure is contributing significantly to morbidity and mortality rates in the general population.

Advances in our understanding of *lead* toxicity have recently benefited by the development of x-ray fluorescence (KXRF) instruments for making safe in vivo measurements of lead levels in bone, which, in turn, reflect cumulative exposure over many years, as opposed to blood lead levels, which mostly reflect recent exposure. Higher levels of cumulative lead exposure in the general population are now well-established as a risk factor for chronic disease, even though blood lead levels have continued to decline in the general population over the past few decades following the removal of lead from gasoline, plumbing, solder in food cans, and other consumer products, with mean levels in the U.S. population now hovering in the 1–2 µg/dL range. For example, higher bone lead levels measured by KXRF have been linked to increased risk of hypertension and accelerated declines in

TABLE 469-1 Heavy Metals

MAIN SOURCES	METABOLISM	TOXICITY	DIAGNOSIS	TREATMENT
Arsenic				
Smelting and microelectronics industries; wood preservatives, pesticides, herbicides, fungicides; contaminant of deep-water wells; folk remedies; and coal; incineration of these products.	Organic arsenic (arsenobetaine, arsenocholine) is ingested in seafood and fish, but is nontoxic; inorganic arsenic is readily absorbed (lung and GI); sequesters in liver, spleen, kidneys, lungs, and GI tract; residues persist in skin, hair, and nails; biomethylation results in detoxification, but this process saturates.	Acute arsenic poisoning results in necrosis of intestinal mucosa with hemorrhagic gastroenteritis, fluid loss, hypotension, delayed cardiomyopathy, acute tubular necrosis, and hemolysis. Chronic arsenic exposure causes diabetes, vasospasm, ischemic heart disease, peripheral vascular insufficiency and gangrene, peripheral neuropathy, and cancer of skin, lung, liver (angiosarcoma), bladder, and kidney. Lethal dose: 120–200 mg (adults); 2 mg/kg (children).	Nausea, vomiting, diarrhea, abdominal pain, delirium, coma, seizures; garlicky odor on breath; hyperkeratosis, hyperpigmentation, exfoliative dermatitis, and Mees' lines (transverse white striae of the fingernails); sensory and motor polyneuritis, distal weakness. Radiopaque sign on abdominal x-ray; ECG–QRS broadening, QT prolongation, ST depression, T-wave flattening; 24-h urinary arsenic >67 µmol/d or 50 µg/d; (no seafood × 24 h); if recent exposure, serum arsenic >0.9 µmol/L (7 µg/dL). High arsenic in hair or nails.	If acute ingestion, ipecac to induce vomiting, gastric lavage, activated charcoal with a cathartic. Supportive care in ICU. Dimercaprol 3–5 mg/kg IM q4h × 2 days; q6h × 1 day, then q12h × 10 days; alternative: oral succimer.
Cadmium				
Metal plating, pigment, smelting, battery, and plastics industries; tobacco; incineration of these products; ingestion of food that concentrates cadmium (grains, cereals, organ meats).	Absorbed through ingestion or inhalation; bound by metallothionein, filtered at the glomerulus, but reabsorbed by proximal tubules (thus, poorly excreted). Biologic half-life: 10–30 y. Binds cellular sulfhydryl groups, competes with zinc, calcium for binding sites. Concentrates in liver and kidneys.	Acute cadmium inhalation causes pneumonitis after 4–24 h; acute ingestion causes gastroenteritis. Chronic exposure causes anosmia, yellowing of teeth, emphysema, minor LFT elevations, microcytic hypochromic anemia unresponsive to iron therapy, proteinuria, increased urinary β_2-microglobulin, calciuria, leading to chronic renal failure, osteomalacia, and fractures, ischemic heart disease and stroke. Cadmium exposure now known to cause lung, prostate, and kidney cancers.	With inhalation: pleuritic chest pain, dyspnea, cyanosis, fever, tachycardia, nausea, noncardiogenic pulmonary edema. With ingestion: nausea, vomiting, cramps, diarrhea. Bone pain, fractures with osteomalacia. If recent exposure, serum cadmium >500 nmol/L (5 µg/dL). Urinary cadmium >100 nmol/L (10 µg/g creatinine) and/or urinary β_2-microglobulin >750 µg/g creatinine (but urinary β_2-microglobulin also increased in other renal diseases such as pyelonephritis).	There is no effective treatment for cadmium poisoning (chelation not useful; dimercaprol can exacerbate nephrotoxicity). Avoidance of further exposure, supportive therapy, vitamin D for osteomalacia.
Lead				
Manufacturing of auto batteries, lead crystal, ceramics, fishing weights, etc.; demolition or sanding of lead-painted houses, bridges; stained glass making, plumbing, soldering; environmental exposure to paint chips, house dust (in homes built <1975), firing ranges (from bullet dust), food or water from improperly glazed ceramics or lead-contaminated cookware; lead pipes and plumbing; contaminated spices, herbal remedies, candies; exposure to environmental pollution from the combustion of leaded fuels, recycling of automobile batteries and electronic waste.	Absorbed through ingestion or inhalation; organic lead (e.g., tetraethyl lead) absorbed dermally. In blood, 95–99% sequestered in RBCs—thus, must measure lead in whole blood (not serum). Distributed widely in soft tissue, with half-life ~30 days; 15% of dose sequestered in bone with half-life of >20 years. Excreted mostly in urine, but also appears in other fluids including breast milk. Interferes with mitochondrial oxidative phosphorylation, ATPases, calcium-dependent messengers; enhances oxidation and cell apoptosis.	Acute exposure with blood lead levels (BPb) of >60–80 µg/dL can cause impaired neurotransmission and neuronal cell death (with central and peripheral nervous system effects); impaired hematopoiesis and renal tubular dysfunction. At higher levels of exposure (e.g., BPb >80–120 µg/dL), acute encephalopathy with convulsions, coma, and death may occur. Subclinical exposures in children (BPb 25–60 µg/dL) are associated with anemia; mental retardation; and deficits in language, motor function, balance, hearing, behavior, and school performance. Impairment of IQ appears to occur at even lower levels of exposure with no measurable threshold above the limit of detection in most assays of 1 µg/dL. In adults, chronic subclinical exposures (BPb >40 µg/dL) are associated with an increased risk of anemia, demyelinating peripheral neuropathy (mainly motor), impairments of reaction time and hearing, accelerated declines in cognition, hypertension, ECG conduction delays, hypertension, higher risk of cardiovascular disease and death, interstitial nephritis and chronic renal failure, diminished sperm counts, and spontaneous abortions.	Abdominal pain, irritability, lethargy, anorexia, anemia, Fanconi's syndrome, pyuria, azotemia in children with blood lead level (BPb) >80 µg/dL; may also see epiphyseal plate "lead lines" on long bone x-rays. Convulsions, coma at BPb >120 µg/dL. Clinically-apparent neurodevelopmental delays can be seen at BPb of 40–80 µg/dL with subclinical declines in IQ expected at lower levels of BPb down to 1 µg/dL. Screening of all U.S. children when they begin to crawl (~6 months) is recommended by the CDC; source identification and intervention is begun if the BPb >3.5 µg/dL. In adults, acute exposure causes similar symptoms as in children as well as headaches, arthralgias, myalgias, depression, impaired short-term memory, loss of libido. Physical examination may reveal a "lead line" at the gingiva-tooth border, pallor, wrist drop, and cognitive dysfunction (e.g., declines on the mini-mental state exam); lab tests may reveal a normocytic, normochromic anemia, basophilic stippling, an elevated blood protoporphyrin level (free erythrocyte or zinc), and motor delays on nerve conduction. U.S. OSHA requires regular testing of lead-exposed workers with removal if BPb >40 µg/dL. Newer guidelines have been proposed recommending that BPb be maintained at <10 µg/dL, removal of workers if BPb >20 µg/dL, and monitoring of cumulative exposure parameters.	Identification and correction of exposure sources are critical. In the United States, most states ask or require primary care physicians and/or laboratories to report all BPbs to the appropriate health agency. In the highly exposed individual with symptoms, chelation is recommended with oral DMSA (succimer); if acutely toxic, hospitalization and IV or IM chelation with ethylenediaminetetraacetic acid calcium disodium (CaEDTA) may be required, with the addition of dimercaprol to prevent worsening of encephalopathy. A large multicenter randomized trial of chelation showed no improvement in measures of intelligence among children with asymptomatic lead exposure (e.g., BPb 20–40 µg/dL). It is possible but remains unclear whether chelation among adults with chronic lead exposure may improve cardiovascular outcomes. Correction of dietary deficiencies in iron, calcium, magnesium, and zinc will lower lead absorption and may also improve toxicity. Vitamin C is a weak but natural chelating agent. Calcium supplements (1200 mg at bedtime) have been shown to lower blood lead levels in pregnant women.

(*Continued*)

TABLE 469-1 Heavy Metals (Continued)

MAIN SOURCES	METABOLISM	TOXICITY	DIAGNOSIS	TREATMENT
Mercury				
Metallic, mercurous, and mercuric mercury (Hg, Hg^{+}, Hg^{2+}) exposures occur in some chemical, metal-processing, electrical equipment, automotive industries; they are also in thermometers, dental amalgams, batteries. Mercury is dispersed by waste incineration. Environmental bacteria convert inorganic to organic mercury, which then bioconcentrates up the aquatic food chain to contaminate tuna, swordfish, and other pelagic fish.	Elemental mercury (Hg) is not well absorbed; however, it will volatilize into highly absorbable vapor. Inorganic mercury is absorbed through the gut and skin. Organic mercury is well absorbed through inhalation and ingestion. Elemental and organic mercury cross the blood-brain barrier and placenta. Mercury is excreted in urine and feces and has a half-life in blood of ~60 days; however, deposits will remain in the kidney and brain for years. Exposure to mercury stimulates the kidney to produce metallothionein, which provides some detoxification benefit. Mercury binds sulfhydryl groups and interferes with a wide variety of critical enzymatic processes.	Acute inhalation of Hg vapor causes pneumonitis and noncardiogenic pulmonary edema leading to death, CNS symptoms, and polyneuropathy. Chronic high exposure causes CNS toxicity (mercurial *erethism*; see Diagnosis); lower exposures impair renal function, motor speed, memory, coordination. Acute ingestion of inorganic mercury causes gastroenteritis, the nephritic syndrome, or acute renal failure, hypertension, tachycardia, and cardiovascular collapse, with death at a dose of 10–42 mg/kg. Ingestion of organic mercury causes gastroenteritis, arrhythmias, and lesions in the basal ganglia, gray matter, and cerebellum at doses >1.7 mg/kg. High exposure during pregnancy causes derangement of fetal neuronal migration resulting in severe mental retardation. Mild exposures during pregnancy (from fish consumption) are associated with declines in neurobehavioral performance in offspring. Dimethylmercury, a compound only found in research labs, is "supertoxic"—a few drops of exposure via skin absorption or inhaled vapor can cause severe cerebellar degeneration and death.	Chronic exposure to metallic mercury vapor produces a characteristic intention tremor and mercurial *erethism*: excitability, memory loss, insomnia, timidity, and delirium ("mad as a hatter"). On neurobehavioral tests: decreased motor speed, visual scanning, verbal and visual memory, visuomotor coordination. Children exposed to mercury in any form may develop *acrodynia* ("pink disease"): flushing, itching, swelling, tachycardia, hypertension, excessive salivation or perspiration, irritability, weakness, morbilliform rashes, desquamation of palms and soles. Toxicity from elemental or inorganic mercury exposure begins when blood levels >180 nmol/L (3.6 μg/dL) and urine levels >0.7 μmol/L (15 μg/dL). Exposures that ended years ago may result in a >20-μg increase in 24-h urine after a 2-g dose of succimer. Organic mercury exposure is best measured by levels in blood (if recent) or hair (if chronic); CNS toxicity in children may derive from fetal exposures associated with maternal hair Hg >30 nmol/g (6 μg/g).	Treat acute ingestion of mercuric salts with induced emesis or gastric lavage and polythiol resins (to bind mercury in the GI tract). Chelate with dimercaprol (up to 24 mg/kg per day IM in divided doses), DMSA (succimer), or penicillamine, with 5-day courses separated by several days of rest. If renal failure occurs, treat with peritoneal dialysis, hemodialysis, or extracorporeal regional complexing hemodialysis and succimer. Chronic inorganic mercury poisoning is best treated with *N*-acetyl penicillamine.

Abbreviations: ATPase, adenosine triphosphatase; BPb, blood lead; CDC, Centers for Disease Control and Prevention; CNS, central nervous system; DMSA, dimercaptosuccinic acid; ECG, electrocardiogram; GI, gastrointestinal; ICU, intensive care unit; IQ, intelligence quotient; LFT, liver function tests; OSHA, Occupational Safety and Health Administration; RBC, red blood cell.

cognition in both men and women living in urban communities. These relationships, in conjunction with other epidemiologic and toxicologic studies, persuaded multiple federal expert panels to conclude they were causal. Prospective studies have also demonstrated that higher bone lead levels, as well as blood lead levels as low as 1–7 μg/dL, are a major risk factor for resistant hypertension, coronary artery calcifications, and increased cardiovascular morbidity and mortality rates in both community-based and occupational-exposed populations. Lead exposure at community levels has also been associated with increased risks of hearing loss, Parkinson's disease, and amyotrophic lateral sclerosis. Occupational levels of lead exposures have been consistently linked with adverse indices of sperm health, such as reduction in semen volume, sperm concentration, total sperm count, sperm vitality, and total sperm motility. With respect to pregnancy-associated risks, high maternal bone lead levels were found to predict lower birth weight, head circumference, birth length, and neurodevelopmental performance in offspring by age 2 years. Offspring have also been shown to have higher blood pressures at age 7–14 years, an age range at which higher blood pressures are known to predict an elevated risk of developing hypertension. In a randomized trial, calcium supplementation (1200 mg daily) was found to significantly reduce the mobilization of lead from maternal bone into blood during pregnancy.

The toxicity of low-level organic *mercury* exposure (as manifested by neurobehavioral performance) is of increasing concern based on studies of the offspring of mothers who ingested mercury-contaminated fish. With respect to whether the consumption of fish by women during pregnancy is good or bad for offspring neurodevelopment, balancing the trade-offs of the beneficial effects of the omega-3-fatty acids (FAs) in fish versus the adverse effects of mercury contamination in fish has led to some confusion and inconsistency in public health recommendations. Overall, it would appear that it would be best for pregnant women to either limit fish consumption to those species known to be low in mercury contamination but high in omega-3-FAs (such as sardines or mackerel) or to avoid fish and obtain omega-3-FAs through supplements or other dietary sources. Well-conducted studies have convincingly debunked the contention that ethyl mercury, used as a preservative in multiuse vaccines administered in early childhood, has played a significant role in causing neurodevelopmental problems such as autism. With regard to adults, there is evidence that mercury exposure is associated with elevated markers of dyslipidemia and high-sensitivity C-reactive protein (a marker of chronic low-grade inflammation), but the direct evidence as to whether mercury exposure is associated with increased risk of hypertension and cardiovascular disease remains somewhat conflicting. There is also some evidence that mercury exposure in the general population is associated with the development of diabetes, perturbations in markers of autoimmunity, and depression. At this point, conclusions cannot be drawn and the clinical significance of these findings remains somewhat unclear.

Heavy metals pose risks to health that are especially burdensome in selected parts of the world. For example, *arsenic* exposure from natural contamination of shallow tube wells inserted for drinking water is a major environmental problem for millions of residents in parts of Bangladesh and Western India. Contamination was formerly considered only a problem with deep wells; however, the geology of this region allows most residents only a few alternatives for potable drinking water. Arsenic contamination of drinking water is also a major problem in China, Argentina, Chile, Mexico, and some regions of the United States (Maine, New Hampshire, Massachusetts). The global campaign to phase out *leaded* gasoline has successfully concluded. However, significant population exposures to lead remain, particularly in the United States with respect to older housing that contains lead paint or that receives drinking water through lead pipes, and evidence indicates that exposures are beginning to increase again in many low- and middle-income countries due to industrial pollution, the recycling of automobile batteries, electronic waste, mining, and a variety of contaminated consumer products. Populations living in the Arctic have been shown to have particularly high exposures to *mercury* due to long-range transport patterns that concentrate mercury in the polar regions, as well as the traditional dependence of Arctic peoples on the consumption of fish and other wildlife that bioconcentrate methylmercury.

A few additional metals deserve brief mention but are not covered in Table 469-1 because of the relative rarity of their being clinically encountered or the uncertainty regarding their potential toxicities. *Aluminum* contributes to the encephalopathy in patients with severe renal disease, who are undergoing dialysis **(Chap. 422)**. High levels of aluminum are found in the neurofibrillary tangles in the cerebral cortex and hippocampus of patients with Alzheimer's disease, as well as in the drinking water and soil of areas with an unusually high incidence of Alzheimer's. The experimental and epidemiologic evidence for the aluminum–Alzheimer's disease link remains relatively weak, however, and it cannot be concluded that aluminum is a causal agent or a contributing factor in neurodegenerative disease. Hexavalent *chromium* is corrosive and sensitizing. Workers in the chromate and chrome pigment production industries have consistently had a greater risk of lung cancer. The introduction of *cobalt* chloride as a fortifier in beer led to outbreaks of fatal cardiomyopathy among heavy consumers, a condition that has also recently been reported in conjunction with the cobalt exposure associated with metal implants used in hip arthroplasty. Occupational exposure (e.g., of miners, dry-battery manufacturers, and arc welders) to *manganese* (Mn) can cause a parkinsonian syndrome within 1–2 years, including gait disorders; postural instability; a masked, expressionless face; tremor; and psychiatric symptoms. In contrast to typical cases of Parkinson's disease, manganese-induced parkinsonism is nonresponsive to treatment with L-dopa. With the introduction of methylcyclopentadienyl manganese tricarbonyl (MMT) as a gasoline additive, there is concern for the toxic potential of environmental manganese exposure. Some epidemiologic studies have found an association between the prevalence of parkinsonian disorders and estimated manganese exposures emitted by local ferroalloy industries; others have found evidence suggesting that manganese may interfere with early childhood neurodevelopment in ways similar to that of lead. Manganese toxicity is clearly associated with dopaminergic dysfunction, and its toxicity is likely influenced by age, gender, ethnicity, genetics, and preexisting medical conditions. *Nickel* exposure induces an allergic response, and inhalation of nickel compounds with low aqueous solubility (e.g., nickel subsulfide and nickel oxide) in occupational settings is associated with an increased risk of lung cancer. Overexposure to selenium may cause local irritation of the respiratory system and eyes, gastrointestinal irritation, liver inflammation, loss of hair, depigmentation, and peripheral nerve damage. Workers exposed to certain organic forms of *tin* (particularly trimethyl and triethyl derivatives) have developed psychomotor disturbances, including tremor, convulsions, hallucinations, and psychotic behavior.

Thallium, which is a component of some insecticides, metal alloys, and fireworks, is absorbed through the skin as well as by ingestion and inhalation. Severe poisoning follows a single ingested dose of >1 g or >8 mg/kg. Nausea and vomiting, abdominal pain, and hematemesis precede confusion, psychosis, organic brain syndrome, and coma. Thallium is radiopaque. Induced emesis or gastric lavage is indicated within 4–6 h of acute ingestion; Prussian blue prevents absorption and is given orally at 250 mg/kg in divided doses. Unlike other types of metal poisoning, thallium poisoning may be less severe when activated charcoal is used to interrupt its enterohepatic circulation. Other measures include forced diuresis, treatment with potassium chloride (which promotes renal excretion of thallium), and peritoneal dialysis.

Chelation therapy remains the treatment of choice for most toxic metals in the setting of severe acute clinical poisoning. However, the use of chelation for treating chronic diseases remains controversial, in part because of the lack of evidence from rigorous randomized clinical trials. One area for which there is moderate evidence is the use of chelation in patients with higher than average levels of accumulated lead burdens as a means of improving kidney function. The results from a series of randomized trials conducted in Taiwan suggest that among individuals with mildly elevated lead burdens (defined as between 150 and 600 μg of lead per 72-h urine upon an EDTA mobilization test [1 g EDTA]), weekly calcium disodium EDTA chelation treatments for between 2 and 27 months can improve renal function outcomes, both in individuals with and without type 2 diabetes.

The Trial to Assess Chelation Therapy (TACT-1), a multicenter, double-blind, placebo-controlled, prospective randomized trial funded by the National Institutes of Health of 1708 patients aged ≥50 years who had experienced a myocardial infarction (MI), found that a protocol of repeated intravenous chelation with disodium EDTA, compared with placebo, modestly but significantly reduced the risk of adverse cardiovascular outcomes, many of which were revascularization procedures. The effect was particularly pronounced among those with concurrent diabetes. However, the trial did not include rigorous measures of exposure to lead or other metals or any selection criteria based on metals exposure. By contrast, the recently released results of the TACT-2 trial, which reproduced the TACT-1 protocol with the addition of measures of metals, did not show benefit. However, the metals exposure levels among subjects was very low, likely reflecting the known decline in metals exposure over time in a study conducted more than 10 years after TACT-1, and it remains unclear whether chelation has a role in treating individuals with significant chronic metals exposure.

FURTHER READING

Aalami AH et al: Carcinogenic effects of heavy metals by inducing dysregulation of microRNAs: A review. Mol Biol Rep 49:12227, 2022.

Basu N et al: Our evolved understanding of the human health risks of mercury. Ambio 52:877, 2023.

Chen H et al: Associations of blood lead, cadmium, and mercury with resistant hypertension among adults in NHANES, 1999-2018. Environ Health Prev Med 28:66, 2023.

Elkin ER et al: Metals exposures and DNA methylation: Current evidence and future directions. Curr Environ Health Rep 4:673, 2022.

Giulioni C et al: The influence of lead exposure on male semen parameters: A systematic review and meta-analysis. Reprod Toxicol 118:108387, 2023.

Lamas GA et al: Contaminant metals as cardiovascular risk factors: A scientific statement from the American Heart Association. J Am Heart Assoc 12:e029852, 2023.

Lanphear BP et al: Low-level lead exposure and mortality in US adults: A population-based cohort study. Lancet Public Health 3:e177, 2018.

Lucchini R, Tieu K: Manganese-induced parkinsonism: Evidence from epidemiological and experimental studies. Biomolecules 13:1190, 2023.

Martínez-Castillo M et al: Arsenic exposure and non-carcinogenic health effects. Hum Exp Toxicol 40:S826, 2021.

Parida L, Patel TN: Systemic impact of heavy metals and their role in cancer development: A review. Environ Monit Assess 195:766, 2023.

Park SK et al: Environmental cadmium and mortality from influenza and pneumonia in U.S. adults. Environ Health Perspect 128:127004, 2020.

Ravalli F et al: Chelation therapy in patients with cardiovascular disease: A systematic review. J Am Heart Assoc 11:e024648, 2022.

Shvachiy L et al: Uncovering the molecular link between lead toxicity and Parkinson's disease. Antioxid Redox Signal 39:321, 2023.

Smereczański NM et al: Current levels of environmental exposure to cadmium in industrialized countries as a risk factor for kidney damage in the general population: A comprehensive review of available data. Int J Mol Sci 24:8413, 2023.

Xu L et al: Positive association of cardiovascular disease (CVD) with chronic exposure to drinking water arsenic (As) at concentrations below the WHO provisional guideline value: A systematic review and meta-analysis. Int J Environ Res Public Health 17:2536, 2020.

470

Poisoning and Drug Overdose

Mark B. Mycyk

Poisoning refers to the development of dose-related adverse effects following exposure to chemicals, drugs, or other xenobiotics. To paraphrase Paracelsus, the dose makes the poison. Although most poisons have predictable dose-related effects, individual responses to a given dose may vary because of genetic polymorphism, enzymatic induction or inhibition in the presence of other xenobiotics, or acquired tolerance. Poisoning may be local (e.g., skin, eyes, or lungs) or systemic depending on the route of exposure, the chemical and physical properties of the poison, and its mechanism of action. The severity and reversibility of poisoning also depend on the functional reserve of the individual or target organ, which is influenced by age and preexisting disease.

EPIDEMIOLOGY

More than 5 million poison exposures occur in the United States each year. Most are acute, are accidental (unintentional), involve a single agent, occur in the home (>90%), result in minor or no toxicity, and involve children <6 years of age. Pharmaceuticals are involved in 47% of poisoning exposures and in 84% of serious or fatal poisonings. Household cleaning substances and cosmetics/personal care products are the most common nonpharmaceutical exposures reported to the National Poison Data System (NPDS). Fatalities most commonly occur from intentional self-harm ingestions (suicide) of pharmaceuticals or from complications involving opioids and other psychoactive drugs of abuse. According to the Centers for Disease Control and Prevention (CDC), fatalities involving drugs have steadily increased since 1999. More than 106,600 deaths were attributed to drug overdose in the United States in 2021, with an age-adjusted overdose death rate of 32.4 per 100,000. Although prescription opioids have appropriately received attention as a major reason for the ongoing substance use disorder epidemic in the United States, the ever-expanding street availability of synthetic opioids (e.g., fentanyl) and the rapid proliferation of other psychoactive drugs of abuse (e.g., xylazine) are the main drivers of drug overdose deaths. State prescription monitoring programs, enhanced training in pain management for health care professionals, expanded harm reduction services, improved outreach education to the public, and improved partnerships with regional safety officials to monitor trends in the illicit drug supply continue to be the main prevention and response priorities for addressing this worsening epidemic. Unintentional exposures can result from the improper use of chemicals at work or play; label misreading; product mislabeling; mistaken identification of unlabeled chemicals; uninformed self-medication; and dosing errors by nurses, pharmacists, physicians, parents, and the elderly. Excluding the recreational use of ethanol, attempted suicide (deliberate self-harm) is the most common reported reason for intentional poisoning. Recreational use of prescribed and over-the-counter drugs for psychotropic or euphoric effects (*abuse*) or excessive self-dosing (*misuse*) is increasingly common and may also result in unintentional self-poisoning.

About 20–25% of exposures require bedside health-professional evaluation, and 5% of all exposures require hospitalization. Poisonings account for 5–10% of all ambulance transports, emergency department visits, and intensive care unit admissions. Hospital admissions related to poisoning are also associated with longer lengths of stay and increase the utilization of resources such as radiography and other laboratory services. Up to 35% of psychiatric admissions are prompted by attempted suicide via overdosage with cases involving adolescents steadily increasing during the last decade. Overall, the mortality rate is low: <1% of all poisoning exposures. It is significantly higher (1–2%) among hospitalized patients with intentional (suicidal) overdose or complications from drugs of abuse, who account for the majority of serious poisonings. Acetaminophen is the pharmaceutical agent most often implicated in fatal poisoning. Overall, carbon monoxide is the leading cause of death from poisoning, but this prominence is not reflected in hospital or poison center statistics because patients with such poisoning are typically dead when discovered and are referred directly to medical examiners.

DIAGNOSIS

Although poisoning can mimic other illnesses, the correct diagnosis can usually be established by the history, physical examination, routine and toxicologic laboratory evaluations, and characteristic clinical course.

■ HISTORY

The *history* should include the time, route, duration, and circumstances (location, surrounding events, and intent) of exposure; the name and amount of each drug, chemical, or ingredient involved; the time of onset, nature, and severity of symptoms; the time and type of first-aid measures provided; the medical and psychiatric history; and occupation.

In many cases, the patient is confused, comatose, unaware of an exposure, or unable or unwilling to admit to one. Suspicious circumstances include unexplained sudden illness in a previously healthy person or a group of healthy people; a history of psychiatric problems (particularly depression or bipolar disorder); recent changes in health, economic status, or social relationships; and onset of illness during work with chemicals or after ingestion of food, drink (especially ethanol), or medications. When patients become ill soon after arriving from a foreign country or being arrested for criminal activity, "body packing" or "body stuffing" (ingesting or concealing illicit drugs in a body cavity) should be suspected. Relevant information may be available from family, friends, paramedics, police, pharmacists, physicians, and employers, who should be questioned regarding the patient's habits, hobbies, behavioral changes, available medications, and antecedent events. Patients need to be asked explicitly about their prescribed medications, use of over-the-counter or herbal medications, and recreational drug use. Drugs previously considered "illicit" such as cannabinoids are now legal in many states and prescribed for therapeutic purposes. A search of clothes, belongings, and place of discovery may reveal a suicide note or a container of drugs or chemicals. If available, a review of a patient's recent social media posts may be helpful to the treating clinician because self-harm plans involving medications or drugs are often described there. Without a clear history in a patient clinically suspected to be poisoned, all medications available anywhere in the patient's home or belongings should be considered as possible agents, including medications for pets. Review of the patient's record in the state prescription monitoring program (PMP) may disclose relevant history of Schedule II, III, IV, and V controlled substance use. The imprint code on pills and the label on chemical products may be used to identify the ingredients and potential toxicity of a suspected poison by consulting a reference text, a computerized database, the manufacturer, or a regional poison information center (800-222-1222). Occupational exposures require review of any available safety data sheet (SDS) from the worksite. Because of increasing globalization from travel and internet consumerism, unfamiliar poisonings may result in local emergency department evaluation. Pharmaceuticals, industrial chemicals, or novel psychoactive drugs of abuse from foreign countries may be identified with the assistance of a local medical toxicologist, a regional poison center, or via the Internet.

■ PHYSICAL EXAMINATION AND CLINICAL COURSE

The *physical examination* should focus initially on vital signs, the cardiopulmonary system, and neurologic status. The neurologic examination should include documentation of neuromuscular abnormalities such as dyskinesia, dystonia, fasciculations, myoclonus, rigidity, and tremors. The patient should also be examined for evidence of trauma and underlying illnesses. Focal neurologic findings are uncommon in poisoning, and their presence should prompt evaluation for a structural central nervous system (CNS) lesion. Examination of the eyes (for nystagmus and pupil size and reactivity), abdomen (for bowel activity and bladder size), and skin (for burns, bullae, color, warmth, moisture, pressure sores, and puncture marks) may reveal findings of diagnostic value. When the history is unclear, all orifices should be examined for the presence of chemical burns and drug packets. The odor of breath or vomitus and the color of nails, skin, or urine may provide important diagnostic clues.

The diagnosis of poisoning in cases of unknown etiology primarily relies on pattern recognition. The first step is to assess the pulse, blood pressure, respiratory rate, pulse oximetry, temperature, and neurologic status and to characterize the overall physiologic state as stimulated, depressed, discordant, or normal (Table 470-1). Obtaining a complete set of vital signs and reassessing them frequently are critical. Measuring core temperature is especially important, even in difficult or combative patients, since temperature elevation is the most reliable prognosticator of poor outcome in poisoning from stimulants (e.g., cocaine) or drug withdrawal (e.g., alcohol or γ-hydroxybutyric acid [GHB]). The next step is to consider the underlying causes of the physiologic state and to attempt to identify a pathophysiologic pattern or toxic syndrome (*toxidrome*) based on the observed findings. Assessing the severity of physiologic derangements (Table 470-2) is useful in this regard and also for monitoring the clinical course and response to treatment. In cases of polydrug overdose involving different drug classes, identifying a clear toxidrome can be challenging if the different drugs counteract the physiologic effects of one another. The final step is to attempt to identify the particular agent involved by looking for unique or relatively poison-specific physical or ancillary test abnormalities. Distinguishing among toxidromes on the basis of the physiologic state is summarized next.

The Stimulated Physiologic State Increased pulse, blood pressure, respiratory rate, temperature, and neuromuscular activity characterize the *stimulated* physiologic state, which can reflect sympathetic,

TABLE 470-1 Differential Diagnosis of Poisoning Based on Physiologic State

STIMULATED	DEPRESSED	DISCORDANT	NORMAL
Sympathetics	Sympatholytics	Asphyxiants	Nontoxic exposure
Sympathomimetics	α_1-Adrenergic antagonists	Cytochrome oxidase inhibitors	Psychogenic illness
Ergot alkaloids	α_2-Adrenergic agonists	Inert gases	"Toxic time-bombs"
Methylxanthines	ACE inhibitors	Irritant gases	Slow absorption
Monoamine oxidase inhibitors	Angiotensin receptor blockers	Methemoglobin inducers	Anticholinergics
Thyroid hormones	Antipsychotics	Oxidative phosphorylation inhibitors	Carbamazepine
Anticholinergics	β-Adrenergic blockers	AGMA inducers	Concretion formers
Antihistamines	Calcium channel blockers	Alcohol (ketoacidosis)	Extended-release phenytoin sodium capsules (Dilantin Kapseals)
Antiparkinsonian agents	Cardiac glycosides	Ethylene glycol	Drug packets
Antipsychotics	Cyclic antidepressants	Iron	Enteric-coated pills
Antispasmodics	Cholinergics	Methanol	Diphenoxylate-atropine (Lomotil)
Belladonna alkaloids	Acetylcholinesterase inhibitors	Other alcohols	Opioids
Cyclic antidepressants	Muscarinic agonists	Salicylate	Salicylates
Mushrooms and plants	Nicotinic agonists	Toluene	Sustained-release pills
Hallucinogens	Opioids	CNS syndromes	Valproate
Cannabinoids (marijuana)	Analgesics	Extrapyramidal reactions	Slow distribution
LSD and analogues	GI antispasmodics	Hydrocarbon inhalation	Cardiac glycosides
Mescaline and analogues	Heroin	Isoniazid	Lithium
Mushrooms	Sedative-hypnotics	Lithium	Metals
Phencyclidine and analogues	Alcohols	Neuroleptic malignant syndrome	Salicylate
Withdrawal syndromes	Anticonvulsants	Serotonin syndrome	Valproate
Barbiturates	Barbiturates	Strychnine	Toxic metabolite
Benzodiazepines	Benzodiazepines	Membrane-active agents	Acetaminophen
Ethanol	GABA precursors	Amantadine	Carbon tetrachloride
GHB products	Muscle relaxants	Antiarrhythmics	Cyanogenic glycosides
Opioids	Other agents	Antihistamines	Ethylene glycol
Sedative-hypnotics	GHB products	Antipsychotics	Methanol
Sympatholytics	Xylazine	Carbamazepine	Methemoglobin inducers
		Cyclic antidepressants	Mushroom toxins
		Local anesthetics	Organophosphate insecticides
		Opioids (some)	Paraquat
		Quinoline antimalarials	Metabolism disruptors
			Antineoplastic agents
			Antiviral agents
			Colchicine
			Hypoglycemic agents
			Immunosuppressive agents
			MAO inhibitors
			Metals
			Other oral anticoagulants
			Salicylate
			Warfarin

Abbreviations: ACE, angiotensin-converting enzyme; AGMA, anion-gap metabolic acidosis; CNS, central nervous system; GABA, γ-aminobutyric acid; GHB, γ-hydroxybutyrate; GI, gastrointestinal; LSD, lysergic acid diethylamide; MAO, monoamine oxidase.

TABLE 470-2 Severity of Physiologic Stimulation and Depression in Poisoning and Drug Withdrawal

Physiologic Stimulation	
Grade 1	Anxious, irritable, tremulous; vital signs normal; diaphoresis, flushing or pallor, mydriasis, and hyperreflexia sometimes present
Grade 2	Agitated; may have confusion or hallucinations but can converse and follow commands; vital signs mildly to moderately increased
Grade 3	Delirious; unintelligible speech, uncontrollable motor hyperactivity; moderately to markedly increased vital signs; tachyarrhythmias possible
Grade 4	Coma, seizures, cardiovascular collapse
Physiologic Depression	
Grade 1	Awake, lethargic, or sleeping but arousable by voice or tactile stimulation; able to converse and follow commands; may be confused
Grade 2	Responds to pain but not voice; can vocalize but not converse; spontaneous motor activity present; brainstem reflexes intact
Grade 3	Unresponsive to pain; spontaneous motor activity absent; brainstem reflexes depressed; motor tone, respirations, and temperature decreased
Grade 4	Unresponsive to pain; flaccid paralysis; brainstem reflexes and respirations absent; cardiovascular vital signs decreased

anticholinergic, or hallucinogen poisoning or drug withdrawal (Table 470-1). Other features are noted in Table 470-2. Mydriasis, a characteristic feature of all stimulants, is most marked in anticholinergic poisoning since pupillary reactivity relies on muscarinic control. In sympathetic poisoning (e.g., due to cocaine), pupils are also enlarged, but some reactivity to light remains. The anticholinergic toxidrome is also distinguished by hot, dry, flushed skin; decreased bowel sounds; and urinary retention. Other stimulant syndromes increase sympathetic activity and cause diaphoresis, pallor, and increased bowel activity with varying degrees of nausea, vomiting, abnormal distress, and occasionally diarrhea. The absolute and relative degree of vital-sign changes and neuromuscular hyperactivity can help distinguish among stimulant toxidromes. Since sympathetics stimulate the peripheral nervous system more directly than do hallucinogens or drug withdrawal, markedly increased vital signs and organ ischemia suggest sympathetic poisoning. Findings helpful in suggesting the particular drug or class causing physiologic stimulation include reflex bradycardia from selective α-adrenergic stimulants (e.g., decongestants), hypotension from selective β-adrenergic stimulants (e.g., asthma therapeutics), limb ischemia from ergot alkaloids, rotatory nystagmus from phencyclidine and ketamine (the only physiologic stimulants that cause this finding), and delayed cardiac conduction from high doses of cocaine and some anticholinergic agents (e.g., antihistamines, cyclic antidepressants, and antipsychotics). Seizures suggest a sympathetic etiology, an anticholinergic agent with membrane-active properties (e.g., cyclic antidepressants, phenothiazines), or a withdrawal syndrome. Close attention to core temperature is critical in patients with grade 4 physiologic stimulation (Table 470-2).

The Depressed Physiologic State Decreased pulse, blood pressure, respiratory rate, temperature, and neuromuscular activity are indicative of the *depressed* physiologic state caused by "functional" sympatholytics (agents that decrease cardiac function and vascular tone as well as sympathetic activity), cholinergic (muscarinic and nicotinic) agents, opioids, and sedative-hypnotic γ-aminobutyric acid (GABA)-ergic agents (Tables 470-1 and 470-2). Miosis is also common and is most pronounced in opioid and cholinergic poisoning. Miosis is distinguished from other depressant syndromes by muscarinic and nicotinic signs and symptoms (Table 470-1). Pronounced cardiovascular depression in the absence of significant CNS depression suggests a direct or peripherally acting sympatholytic. In contrast, in opioid and sedative-hypnotic poisoning, vital-sign changes are secondary to depression of CNS cardiovascular and respiratory centers (or consequent hypoxemia), and significant abnormalities in these parameters do not occur until there is a marked decrease in the level of consciousness (grade 3 or 4 physiologic depression; Table 470-2). Other clues that suggest the cause of physiologic depression include cardiac arrhythmias and conduction disturbances (due to antiarrhythmics, β-adrenergic antagonists, calcium channel blockers, digitalis glycosides, propoxyphene, and cyclic antidepressants), mydriasis (due to tricyclic antidepressants, some antiarrhythmics, meperidine, and diphenoxylate-atropine [Lomotil]), nystagmus (due to sedative-hypnotics), and seizures (due to cholinergic agents, propoxyphene, and cyclic antidepressants).

The Discordant Physiologic State The *discordant* physiologic state is characterized by mixed vital-sign and neuromuscular abnormalities, as observed in poisoning by asphyxiants, CNS syndromes, membrane-active agents, and anion-gap metabolic acidosis (AGMA) inducers (Table 470-1). In these conditions, manifestations of physiologic stimulation and physiologic depression occur together or at different times during the clinical course. For example, membrane-active agents can cause simultaneous coma, seizures, hypotension, and tachyarrhythmias. Alternatively, vital signs may be normal while the patient has an altered mental status or is obviously sick or clearly symptomatic. Early, pronounced vital-sign and mental-status changes suggest asphyxiant or membrane-active agent poisoning; the lack of such abnormalities suggests an AGMA inducer; and marked neuromuscular dysfunction without significant vital-sign abnormalities suggests a CNS syndrome. The *discordant* physiologic state may also be evident in patients poisoned with multiple agents.

The Normal Physiologic State A *normal* physiologic status and physical examination may be due to a nontoxic exposure, psychogenic illness, or poisoning by "toxic time-bombs": agents that are slowly absorbed, are slowly distributed to their sites of action, require metabolic activation, or disrupt metabolic processes (Table 470-1). Because so many medications have now been reformulated into once-a-day preparations for the patient's convenience and adherence, toxic time-bombs are increasingly common. Diagnosing a nontoxic exposure requires that the identity of the exposure agent be known or that a toxic time-bomb exposure be excluded and the time since exposure exceed the longest known or predicted interval between exposure and peak toxicity. Psychogenic illness (fear of being poisoned, mass hysteria) may also follow a nontoxic exposure and should be considered when symptoms are inconsistent with exposure history. Anxiety reactions resulting from a nontoxic exposure can cause mild physiologic stimulation (Table 470-2) and be indistinguishable from toxicologic causes without ancillary testing or a suitable period of observation.

LABORATORY ASSESSMENT

Laboratory assessment may be helpful in the differential diagnosis. Increased AGMA is most common in advanced methanol, ethylene glycol, and salicylate intoxication but can occur with any poisoning that results in hepatic, renal, or respiratory failure; seizures; or shock. The serum lactate concentration is more commonly low (less than the anion gap) in the former and high (nearly equal to the anion gap) in the latter. An abnormally low anion gap can be due to elevated blood levels of bromide, calcium, iodine, lithium, or magnesium. An increased osmolal gap—a difference of >10 mmol/L between serum osmolality (measured by freezing-point depression) and osmolality calculated from serum sodium, glucose, and blood urea nitrogen levels—suggests the presence of a low-molecular-weight solute such as acetone; an alcohol (benzyl, ethanol, isopropanol, methanol); a glycol (diethylene, ethylene, propylene); ether (ethyl, glycol); or an "unmeasured" cation (calcium, magnesium) or sugar (glycerol, mannitol, sorbitol). Ketosis suggests acetone, isopropyl alcohol, salicylate poisoning, or alcoholic ketoacidosis. Hypoglycemia may be due to poisoning with β-adrenergic blockers, ethanol, insulin, oral hypoglycemic agents, quinine, and salicylates, whereas hyperglycemia can occur in poisoning with acetone, β-adrenergic agonists, caffeine, calcium channel blockers, iron, theophylline, or *N*-3-pyridylmethyl-*N*-*p*-nitrophenylurea (PNU [Vacor]). Hypokalemia can be caused by barium, β-adrenergic agonists, caffeine, diuretics, theophylline, or toluene; hyperkalemia suggests poisoning with an α-adrenergic agonist, a β-adrenergic blocker, cardiac glycosides, or fluoride. Hypocalcemia may be seen in ethylene glycol, fluoride, and oxalate poisoning. Prothrombin time and international normalized ratio are useful for risk stratification in cases of warfarin or rodenticide poisoning but are not to be relied on when evaluating overdose or complications from direct oral anticoagulant pharmaceuticals (direct thrombin inhibitors and direct factor Xa inhibitors).

The *electrocardiogram* (ECG) can be useful for rapid diagnostic purposes. Bradycardia and atrioventricular block may occur in patients poisoned by α-adrenergic agonists, antiarrhythmic agents, beta blockers, calcium channel blockers, cholinergic agents (carbamate and organophosphate insecticides), cardiac glycosides, lithium, or tricyclic antidepressants. QRS- and QT-interval prolongation may be caused by hyperkalemia, various antidepressants, and other membrane-active drugs (Table 470-1). Ventricular tachyarrhythmias may be seen in poisoning with cardiac glycosides, fluorides, membrane-active drugs, methylxanthines, sympathomimetics, antidepressants, and agents that cause hyperkalemia or potentiate the effects of endogenous catecholamines (e.g., chloral hydrate, aliphatic and halogenated hydrocarbons).

Radiologic studies may occasionally be useful. Pulmonary edema (adult respiratory distress syndrome [ARDS]) can be caused by poisoning with carbon monoxide, cyanide, an opioid, paraquat, phencyclidine, a sedative-hypnotic, or salicylate; by inhalation of irritant gases, fumes, or vapors (acids and alkali, ammonia, aldehydes, chlorine, hydrogen sulfide, isocyanates, metal oxides, mercury, phosgene, polymers); or by prolonged anoxia, hyperthermia, or shock. Aspiration pneumonia is common in patients with coma, seizures, and petroleum distillate aspiration. Chest x-ray is useful for identifying complications from metal fume fever or elemental mercury. The presence of radiopaque densities on abdominal x-rays or abdominal computed tomography (CT) scan suggests the ingestion of chloral hydrate, chlorinated hydrocarbons, heavy metals, illicit drug packets, iodinated compounds, potassium salts, enteric-coated tablets, or salicylates.

Toxicologic analysis of urine and blood (and occasionally of gastric contents and chemical samples) can sometimes confirm or rule out suspected poisoning. Interpretation of laboratory data requires knowledge of the qualitative and quantitative tests used for screening and confirmation (enzyme-multiplied, fluorescence polarization, and radio-immunoassays; colorimetric and fluorometric assays; thin-layer, gas-liquid, or high-performance liquid chromatography; gas chromatography; mass spectrometry), their sensitivity (limit of detection) and specificity, the preferred biologic specimen for analysis, and the optimal time of specimen sampling. Personal communication with the hospital laboratory is essential to an understanding of institutional testing capabilities and limitations.

Rapid *qualitative* hospital-based urine tests for drugs of abuse are only screening tests that cannot confirm the exact identity of the detected substance and should not be considered diagnostic or used for forensic purposes. False-positive and false-negative results are common. A positive screen may result from other pharmaceuticals that interfere with laboratory analysis (e.g., fluoroquinolones commonly cause false-positive opiate screens). Confirmatory testing with gas chromatography/mass spectrometry can be requested, but it often takes weeks to obtain a reported result. A negative screening result may mean that the responsible substance is not detectable by the test used or that its concentration is too low for detection at the time of sampling. For instance, recent new drugs of abuse that often result in emergency department evaluation for unexpected complications, such as synthetic cannabinoids (spice), cathinones (bath salts), and opiate substitutes (kratom), are not detectable by hospital-based tests. In cases where a drug concentration is too low to be detected early during clinical evaluation, repeating the test at a later time may yield a positive result. Patients symptomatic from drugs of abuse often require immediate management based on the history, physical examination, and observed toxidrome without laboratory confirmation (e.g., apnea from opioid intoxication). When the patient is asymptomatic or when the clinical picture is consistent with the reported history, qualitative screening is neither clinically useful nor cost-effective. Thus, qualitative drug screens are of greatest value for the evaluation of patients with severe or unexplained toxicities, such as coma, seizures, cardiovascular instability, metabolic or respiratory acidosis, and nonsinus cardiac rhythms. In contrast to qualitative drug screens, *quantitative* serum tests are useful for evaluation of patients poisoned with acetaminophen **(Chap. 351)**, alcohols (including ethylene glycol and methanol), anticonvulsants, barbiturates, digoxin, heavy metals, iron, lithium, salicylate, and theophylline, as well as for the presence of carboxyhemoglobin and methemoglobin. The serum concentration in these cases guides clinical management, and results are often available within an hour.

The *response to antidotes* is sometimes useful for diagnostic purposes. Resolution of altered mental status and abnormal vital signs within minutes of IV administration of dextrose, naloxone, or flumazenil is virtually diagnostic of hypoglycemia, opioid poisoning, and benzodiazepine intoxication, respectively. The prompt reversal of dystonic (extrapyramidal) signs and symptoms following an IV dose of benztropine or diphenhydramine confirms a drug etiology. Although complete reversal of both central and peripheral manifestations of anticholinergic poisoning by physostigmine is diagnostic of this condition, physostigmine may cause some arousal in patients with CNS depression of any etiology.

TREATMENT

Poisoning and Drug Overdose

GENERAL PRINCIPLES

Treatment goals include support of vital signs, prevention of further poison absorption (decontamination), enhancement of poison elimination, administration of specific antidotes, and prevention of reexposure **(Table 470-3)**. Specific treatment depends on the identity of the poison, the route and amount of exposure, the time

TABLE 470-3 Fundamentals of Poisoning Management

Supportive Care	
Airway protection	Treatment of seizures
Oxygenation/ventilation	Correction of temperature abnormalities
Treatment of arrhythmias	Correction of metabolic derangements
Hemodynamic support	Prevention of secondary complications
Prevention of Further Poison Absorption	
Gastrointestinal decontamination	Decontamination of other sites
Gastric lavage	Eye decontamination
Activated charcoal	Skin decontamination
Whole-bowel irrigation	Body cavity evacuation
Dilution	
Endoscopic/surgical removal	
Enhancement of Poison Elimination	
Multiple-dose activated charcoal administration	Extracorporeal removal
Alteration of urinary pH	Hemodialysis
Chelation	Hemoperfusion
Hyperbaric oxygenation	Hemofiltration
	Plasmapheresis
	Exchange transfusion
	Continuous venovenous hemofiltration (CVVH)
Administration of Antidotes	
Neutralization by antibodies	Metabolic antagonism
Neutralization by chemical binding	Physiologic antagonism
Prevention of Reexposure	
Adult education	Notification of regulatory agencies
Child-proofing	Psychiatric referral
Naloxone distribution	Linkage to harm reduction services

of presentation relative to the time of exposure, and the severity of poisoning. Knowledge of the offending agents' pharmacokinetics and pharmacodynamics is essential.

During the *pretoxic phase*, prior to the onset of poisoning, decontamination is the highest priority, and treatment is based solely on the history. The maximal potential toxicity based on the greatest possible exposure should be assumed. Since decontamination is more effective when accomplished soon after exposure and when the patient is asymptomatic, the initial history and physical examination should be focused and brief. It is also advisable to establish IV access and initiate cardiac monitoring, particularly in patients with potentially serious ingestions or unclear histories.

When an accurate history is not obtainable and a poison causing delayed toxicity (i.e., a toxic time-bomb) or irreversible damage is suspected, blood and urine should be sent for appropriate toxicologic screening and quantitative analysis. During poison absorption and distribution, blood levels may be greater than those in tissue and may not correlate with toxicity. However, high blood levels of agents whose metabolites are more toxic than the parent compound (acetaminophen, ethylene glycol, or methanol) may indicate the need for additional interventions (antidotes, dialysis). Most patients who remain asymptomatic or who become asymptomatic 6 h after ingestion are unlikely to develop subsequent toxicity and can be discharged safely. Longer observation will be necessary for patients who have ingested toxic time-bombs.

During the *toxic phase*—the interval between the onset of poisoning and its peak effects—management is based primarily on clinical and laboratory findings. *Effects after an overdose usually begin sooner, peak later, and last longer than they do after a therapeutic dose*. A drug's published pharmacokinetic profile in standard references such as the *Physician's Desk Reference* (PDR) is usually different from its toxicokinetic profile in overdose. Resuscitation and stabilization are the first priority. Symptomatic patients should have an IV line placed and should undergo oxygen saturation determination, cardiac monitoring, and continuous observation. Baseline laboratory, ECG, and x-ray evaluation may also be appropriate. Intravenous glucose (unless the serum level is documented to be normal), naloxone, and thiamine should be considered in patients with altered mental status, particularly those with coma or seizures. Decontamination should also be considered, but it is less likely to be effective during this phase than during the pretoxic phase.

Measures that enhance poison elimination may shorten the duration and severity of the toxic phase. However, they are not without risk, which must be weighed against the potential benefit. Diagnostic certainty (usually via laboratory confirmation) is generally a prerequisite. Intestinal (gut) dialysis with repetitive doses of activated charcoal (see "Multiple-Dose Activated Charcoal," later) can enhance the elimination of selected poisons such as theophylline or carbamazepine. Urinary alkalinization may enhance the elimination of salicylates and a few other poisons. Chelation therapy can enhance the elimination of selected metals. Extracorporeal elimination methods are effective for many poisons, but their expense and risk make their use reasonable only in patients who would otherwise have an unfavorable outcome.

During the *resolution phase* of poisoning, supportive care and monitoring should continue until clinical, laboratory, and ECG abnormalities have resolved. Since chemicals are eliminated sooner from the blood than from tissues, blood levels are usually lower than tissue levels during this phase and again may not correlate with toxicity. This discrepancy applies particularly when extracorporeal elimination procedures are used. Redistribution from tissues may cause a rebound increase in the blood level after termination of these procedures (e.g., lithium). When a metabolite is responsible for toxic effects, continued treatment may be necessary in the absence of clinical toxicity or abnormal laboratory studies.

SUPPORTIVE CARE

The goal of supportive therapy is to maintain physiologic homeostasis until detoxification is accomplished and to prevent and treat secondary complications such as aspiration, bedsores, cerebral and pulmonary edema, pneumonia, rhabdomyolysis, renal failure, sepsis, thromboembolic disease, coagulopathy, and generalized organ dysfunction due to hypoxemia or shock.

Admission to an intensive care unit is indicated for the following: patients with severe poisoning (coma, respiratory depression, hypotension, cardiac conduction abnormalities, cardiac arrhythmias, hypothermia or hyperthermia, seizures); those needing close monitoring, antidotes, or enhanced elimination therapy; those showing progressive clinical deterioration; and those with significant underlying medical problems. Patients with mild to moderate toxicity can be managed on a general medical service, on an intermediate care unit, or in an emergency department observation area, depending on the anticipated duration and level of monitoring needed (intermittent clinical observation vs continuous clinical, cardiac, and respiratory monitoring). Patients who have attempted suicide require continuous observation and measures to prevent self-injury until they are no longer suicidal.

Respiratory Care

Endotracheal intubation for protection against the aspiration of gastrointestinal contents is of paramount importance in patients with CNS depression or seizures as this complication can increase morbidity and mortality rates. Mechanical ventilation may be necessary for patients with respiratory depression or hypoxemia and for facilitation of therapeutic sedation or paralysis of patients in order to prevent or treat hyperthermia, acidosis, and rhabdomyolysis associated with neuromuscular hyperactivity. Since clinical assessment of respiratory function can be inaccurate, the need for oxygenation and ventilation is best determined by continuous pulse oximetry or arterial blood-gas analysis. The gag reflex is not a reliable indicator of the need for intubation. A patient with CNS depression may maintain airway patency while being stimulated but not if left alone. Drug-induced pulmonary edema is usually noncardiac rather than cardiac in origin, although profound CNS depression and cardiac conduction abnormalities suggest the latter. Measurement of pulmonary artery pressure may be necessary to establish the cause and direct appropriate therapy. Extracorporeal measures (membrane oxygenation, extracorporeal membrane oxygenation [ECMO], venoarterial perfusion, cardiopulmonary bypass) and partial liquid (perfluorocarbon) ventilation may be appropriate for severe but reversible respiratory failure. In the last decade, ECMO has been increasingly used for critically ill poisoned patients where standard resuscitative therapy or antidotes have not been helpful, but further research is still needed to determine the right toxicologic indications for this treatment strategy.

Cardiovascular Therapy

Maintenance of normal tissue perfusion is critical for complete recovery to occur once the offending agent has been eliminated. Focused bedside echocardiography or measurement of central venous pressure may help prioritize therapeutic strategies. If hypotension is unresponsive to volume expansion and appropriate goal-directed antidotal therapy, treatment with norepinephrine, epinephrine, or high-dose dopamine may be necessary. Intraaortic balloon pump counterpulsation and venoarterial or cardiopulmonary perfusion techniques should be considered for severe but reversible cardiac failure. For patients with a return of spontaneous circulation after resuscitative treatment for cardiopulmonary arrest secondary to poisoning, therapeutic hypothermia should be used according to protocol. Bradyarrhythmias associated with hypotension generally should be treated as described in **Chaps. 251 and 252**. Glucagon, calcium, and high-dose insulin with dextrose may be effective in beta blocker and calcium channel blocker poisoning. Antibody therapy may be indicated for cardiac glycoside poisoning.

Supraventricular tachycardia associated with hypertension and CNS excitation is almost always due to agents that cause generalized physiologic excitation (Table 470–1). Most cases are mild or moderate in severity and require only observation or nonspecific sedation with a benzodiazepine. In severe cases or those associated with hemodynamic instability, chest pain, or ECG evidence of ischemia, specific therapy is indicated. When the etiology is sympathetic hyperactivity, treatment with a benzodiazepine should be prioritized. Further treatment with a combined alpha and beta blocker (labetalol), a calcium channel blocker (verapamil or diltiazem), or a combination of a beta blocker and a vasodilator (esmolol and nitroprusside) may be considered for cases refractory to high doses of benzodiazepines only when adequate sedation has been achieved but cardiac conduction or blood pressure abnormalities persist. Treatment with an α-adrenergic antagonist (phentolamine) alone may sometimes be appropriate. If the cause is anticholinergic poisoning, physostigmine alone can be effective. Supraventricular tachycardia without hypertension is generally secondary to vasodilation or hypovolemia and responds to fluid administration.

For ventricular tachyarrhythmias due to tricyclic antidepressants and other membrane-active agents (Table 470-1), sodium bicarbonate is indicated, whereas class IA, IC, and III antiarrhythmic agents are contraindicated because of similar electrophysiologic effects. Although lidocaine and phenytoin are historically safe for ventricular tachyarrhythmias of any etiology, sodium bicarbonate should be considered first for any ventricular arrhythmia suspected to have a toxicologic etiology. Intravenous lipid emulsion therapy has shown benefit for treatment of arrhythmias and hemodynamic instability from various membrane-active agents. Beta blockers can be hazardous if the arrhythmia is due to sympathetic hyperactivity. Magnesium sulfate and overdrive pacing (by isoproterenol or a pacemaker) may be useful in patients with torsades des pointes and prolonged QT intervals. Magnesium and anti-digoxin antibodies should be considered in patients with severe cardiac glycoside poisoning. Invasive (esophageal or intracardiac) ECG recording may be necessary to determine the origin (ventricular or supraventricular) of wide-complex tachycardias **(Chap. 253)**. If the patient is hemodynamically stable, however, it is reasonable to simply observe the patient rather than to administer another potentially proarrhythmic agent. Arrhythmias may be resistant to drug therapy until underlying acid-base, electrolyte, oxygenation, and temperature derangements are corrected.

Central Nervous System Therapies Neuromuscular hyperactivity and seizures can lead to hyperthermia, lactic acidosis, and rhabdomyolysis and should be treated aggressively. Seizures caused by excessive stimulation of catecholamine receptors (sympathomimetic or hallucinogen poisoning and drug withdrawal) or decreased activity of GABA (isoniazid poisoning) or glycine (strychnine poisoning) receptors are best treated with agents that enhance GABA activity, such as benzodiazepine or barbiturates. Since benzodiazepines and barbiturates act by slightly different mechanisms (the former increases the frequency via allosteric modulation at the receptor and the latter directly increases the duration of chloride channel opening in response to GABA), therapy with both may be effective when neither is effective alone. Seizures caused by isoniazid, which inhibits the synthesis of GABA at several steps by interfering with the cofactor pyridoxine (vitamin B_6), may require high doses of supplemental pyridoxine. Seizures resulting from membrane destabilization (beta blocker or cyclic antidepressant poisoning) require GABA enhancers (benzodiazepines first, barbiturates second). Phenytoin is contraindicated in toxicologic seizures: Animal and human data demonstrate worse outcomes after phenytoin loading, especially in theophylline overdose. For poisons with central dopaminergic effects (methamphetamine, phencyclidine) manifested by psychotic behavior, a dopamine receptor antagonist, such as haloperidol or ziprasidone, may be useful. In anticholinergic and cyanide poisoning, specific antidotal therapy may be necessary. The treatment of seizures secondary to cerebral ischemia or edema or to metabolic abnormalities should include correction of the underlying cause. Neuromuscular paralysis is indicated in refractory cases. Electroencephalographic monitoring and continuing treatment of seizures are necessary to prevent permanent neurologic damage.

Other Measures Temperature extremes, metabolic abnormalities, hepatic and renal dysfunction, and secondary complications should be treated by standard therapies.

PREVENTION OF POISON ABSORPTION

Gastrointestinal Decontamination Whether or not to perform gastrointestinal decontamination and which procedure to use depends on the time since ingestion; the existing and predicted toxicity of the ingestant; the availability, efficacy, and contraindications of the procedure; and the nature, severity, and risk of complications. The efficacy of all decontamination procedures decreases with time, and data are insufficient to support or exclude a beneficial effect when they are used >1 h after ingestion. The average time from ingestion to presentation for treatment is >1 h for children and >3 h for adults. Most patients will recover from poisoning uneventfully with good supportive care alone, but complications of gastrointestinal decontamination, particularly aspiration, can prolong this process. Hence, gastrointestinal decontamination should be performed selectively, not routinely, in the management of overdose patients. It is clearly unnecessary when predicted toxicity is minimal or the time of expected maximal toxicity has passed without significant effect.

Activated charcoal has comparable or greater efficacy; has fewer contraindications and complications; and is less aversive and invasive than ipecac or gastric lavage. Thus, it is the preferred method of gastrointestinal decontamination in most situations. Activated charcoal suspension (in water) is given orally via a cup, straw, or small-bore nasogastric tube. The generally recommended dose is 1 g/kg body weight because of its dosing convenience, although in vitro and in vivo studies have demonstrated that charcoal adsorbs ≥90% of most substances when given in an amount equal to 10 times the weight of the substance. Palatability may be increased by adding a sweetener (sorbitol) or a flavoring agent (cherry, chocolate, or cola syrup) to the suspension. Charcoal adsorbs ingested poisons within the gut lumen, allowing the charcoal-toxin complex to be evacuated with stool. Charged (ionized) chemicals such as mineral acids, alkalis, and highly dissociated salts of cyanide, fluoride, iron, lithium, and other inorganic compounds are not well adsorbed by charcoal. In studies with animals and human volunteers, charcoal decreases the absorption of ingestants by an average of 73% when given within 5 min of ingestant administration, 51% when given at 30 min, and 36% when given at 60 min. For this reason, charcoal given before hospital arrival by prehospital emergency medical services (EMS) increases the potential clinical benefit. Side effects of charcoal include nausea, vomiting, and diarrhea or constipation. Charcoal may also prevent the absorption of orally administered therapeutic agents, so the timing and the dose administered need to be adjusted. Complications include mechanical obstruction of the airway, aspiration, vomiting, and bowel obstruction and infarction caused by inspissated charcoal. Charcoal is not recommended for patients who have ingested corrosives because it obscures endoscopy.

Gastric lavage should be considered for life-threatening poisons that cannot be treated effectively with other decontamination, elimination, or antidotal therapies (e.g., colchicine). Gastric lavage is performed by sequentially administering and aspirating ~5 mL of fluid per kilogram of body weight through a no. 40 French orogastric tube (no. 28 French tube for children). Except in infants, for whom normal saline is recommended, tap water is acceptable. The patient should be placed in Trendelenburg and left lateral decubitus positions to prevent aspiration (even if an endotracheal tube is in place). Lavage decreases ingestant absorption by an average of 52% if performed within 5 min of ingestion administration, 26% if performed at 30 min, and 16% if performed at 60 min. Significant amounts of ingested drug are recovered from <10% of patients. Aspiration is a common complication (occurring in up to 10% of patients), especially when lavage is performed improperly. Serious complications (esophageal and gastric perforation, tube misplacement in the trachea) occur in ~1% of patients. For this reason, the physician should personally insert the lavage tube and confirm its placement, and the patient must be cooperative during the procedure. Gastric lavage is contraindicated in corrosive or petroleum distillate ingestions because of the respective risks of gastroesophageal perforation and aspiration pneumonitis. It is also contraindicated in patients with a compromised unprotected airway and those at risk for hemorrhage or perforation due to esophageal or gastric pathology or recent surgery. Finally, gastric lavage is absolutely contraindicated in combative patients or those who refuse, as most published complications involve patient resistance to the procedure.

Syrup of ipecac, an emetogenic agent that was once the substance most commonly used for decontamination, no longer has a role in poisoning management. Even the American Academy of Pediatrics—traditionally the strongest proponent of ipecac—issued a policy statement in 2003 recommending that ipecac should no longer be used in poisoning treatment. Chronic ipecac use (by patients with anorexia nervosa or bulimia) has been reported to cause electrolyte and fluid abnormalities, cardiac toxicity, and myopathy.

Whole-bowel irrigation is performed by administering a bowel-cleansing solution containing electrolytes and polyethylene glycol (Golytely, Colyte) orally or by gastric tube at a rate of 2 L/h (0.5 L/h in children) until rectal effluent is clear. The patient must be in a sitting position. Although data are limited, whole-bowel irrigation appears to be as effective as other decontamination procedures in volunteer studies. It is most appropriate for those who have ingested foreign bodies, packets of illicit drugs, and agents that are poorly adsorbed by charcoal (e.g., heavy metals). This procedure is contraindicated in patients with bowel obstruction, ileus, hemodynamic instability, and compromised unprotected airways.

Cathartics are salts (disodium phosphate, magnesium citrate and sulfate, sodium sulfate) or saccharides (mannitol, sorbitol) that historically have been given with activated charcoal to promote the rectal evacuation of gastrointestinal contents. However, no animal, volunteer, or clinical data have ever demonstrated any decontamination benefit from cathartics. Abdominal cramps, nausea, and occasional vomiting are side effects. Complications of repeated dosing include severe electrolyte disturbances and excessive diarrhea. Cathartics are contraindicated in patients who have ingested corrosives and in those with preexisting diarrhea. Magnesium-containing cathartics should not be used in patients with renal failure.

Dilution (i.e., drinking water, another clear liquid, or milk at a volume of 5 mL/kg of body weight) is recommended only after the ingestion of corrosives (acids, alkali). It may increase the dissolution rate (and hence absorption) of capsules, tablets, and other solid ingestants and should *not* be used in these circumstances.

Endoscopic or surgical removal of poisons may be useful in rare situations, such as ingestion of a potentially toxic foreign body that fails to transit the gastrointestinal tract, a potentially lethal amount of a heavy metal (arsenic, iron, mercury, thallium), or agents that have coalesced into gastric concretions or bezoars (heavy metals, lithium, salicylates, sustained-release preparations). Patients who become toxic from cocaine due to its leakage from ingested drug packets require immediate surgical intervention.

Decontamination of Other Sites Immediate, copious flushing with water, saline, or another available clear, drinkable liquid is the initial treatment for topical exposures (exceptions include alkali metals, calcium oxide, phosphorus). Saline is preferred for eye irrigation. A triple wash (water, soap, water) may be best for dermal decontamination. Inhalational exposures should be treated initially with fresh air or supplemental oxygen. The removal of liquids from body cavities such as the vagina or rectum is best accomplished by irrigation. Solids (drug packets, pills) should be removed manually, preferably under direct visualization.

ENHANCEMENT OF POISON ELIMINATION

Although the elimination of most poisons can be accelerated by therapeutic interventions, the pharmacokinetic efficacy (removal of drug at a rate greater than that accomplished by intrinsic elimination) and clinical benefit (shortened duration of toxicity or improved outcome) of such interventions are often more theoretical than proven. Accordingly, the decision to use such measures should be based on the actual or predicted toxicity and the potential efficacy, cost, and risks of therapy.

Multiple-Dose Activated Charcoal Repetitive oral dosing with charcoal can enhance the elimination of previously absorbed substances by binding them within the gut as they are excreted in the bile, are secreted by gastrointestinal cells, or passively diffuse into the gut lumen (*reverse absorption* or *enterocapillary exsorption*). Doses of 0.5–1 g/kg of body weight every 2–4 h, adjusted downward to avoid regurgitation in patients with decreased gastrointestinal motility, are generally recommended. Pharmacokinetic efficacy approaches that of hemodialysis for some agents (e.g., phenobarbital, theophylline). Multiple-dose therapy should be considered only for selected agents (theophylline, phenobarbital, carbamazepine, dapsone, quinine). Complications include intestinal obstruction, pseudo-obstruction, and nonocclusive intestinal infarction in patients with decreased gut motility. Because of electrolyte and fluid shifts, sorbitol and other cathartics are absolutely contraindicated when multiple doses of activated charcoal are administered.

Urinary Alkalinization Ion trapping via alteration of urine pH may prevent the renal reabsorption of poisons that undergo excretion by glomerular filtration and active tubular secretion. Since membranes are more permeable to nonionized molecules than to their ionized counterparts, acidic (low-pK_a) poisons are ionized and trapped in alkaline urine, whereas basic ones become ionized and trapped in acid urine. Urinary alkalinization (producing a urine pH ≥7.5 and a urine output of 3–6 mL/kg of body weight per hour by the addition of sodium bicarbonate to an IV solution) enhances the excretion of chlorophenoxyacetic acid herbicides, chlorpropamide, diflunisal, fluoride, methotrexate, phenobarbital, sulfonamides, and salicylates. Contraindications include congestive heart failure, renal failure, and cerebral edema. Acid-base, fluid, and electrolyte parameters should be monitored carefully. Although acid diuresis may make theoretical sense for some overdoses (amphetamines), it is *never* indicated and is potentially harmful.

Extracorporeal Removal Hemodialysis, charcoal or resin hemoperfusion, hemofiltration, plasmapheresis, and exchange transfusion are capable of removing any toxin from the bloodstream. Agents most amenable to enhanced elimination by dialysis have low molecular mass (<500 Da), high water solubility, low protein binding, small volumes of distribution (<1 L/kg of body weight), prolonged elimination (long half-life), and high dialysis clearance relative to total-body clearance. Molecular weight, water solubility, and protein binding do not limit the efficacy of the other forms of extracorporeal removal.

Dialysis should be considered in cases of severe poisoning due to carbamazepine, ethylene glycol, isopropyl alcohol, lithium, methanol, theophylline, salicylates, and valproate. Although hemoperfusion may be more effective in removing some of these poisons, it does not correct associated acid-base and electrolyte abnormalities, and most hospitals no longer have hemoperfusion cartridges readily available. Fortunately, recent advances in hemodialysis technology make it as effective as hemoperfusion for removing poisons such as caffeine, carbamazepine, and theophylline. Both techniques require central venous access and systemic anticoagulation and may result in transient hypotension. Hemoperfusion may also cause hemolysis, hypocalcemia, and thrombocytopenia. Peritoneal dialysis and exchange transfusion are less effective but may be used when other procedures are unavailable, contraindicated, or technically difficult (e.g., in infants). Continuous venovenous hemofiltration (CVVH) has also been used successfully when conventional intermittent hemodialysis is not well tolerated because of hemodynamic lability. Exchange transfusion may be indicated in the treatment of severe arsine- or sodium chlorate–induced hemolysis, methemoglobinemia, and sulfhemoglobinemia. Although hemofiltration can enhance elimination of aminoglycosides, vancomycin, and metal-chelate complexes, the roles of hemofiltration and plasmapheresis in the treatment of poisoning are not yet defined.

Candidates for extracorporeal removal therapies include patients with severe toxicity whose condition deteriorates despite aggressive supportive therapy; those with potentially prolonged, irreversible, or fatal toxicity; those with dangerous blood levels of toxins; those who lack the capacity for self-detoxification because of liver or renal failure; and those with a serious underlying illness or complication that will adversely affect recovery.

Other Techniques The elimination of heavy metals can be enhanced by chelation, and the removal of carbon monoxide can be accelerated by hyperbaric oxygenation.

ADMINISTRATION OF ANTIDOTES

Antidotes counteract the effects of poisons by neutralizing them (e.g., antibody-antigen reactions, chelation, chemical binding) or by antagonizing their physiologic effects (e.g., activation of opposing nervous system activity, provision of a competitive metabolic or receptor substrate). Poisons or conditions with specific antidotes include acetaminophen, anticholinergic agents, anticoagulants, benzodiazepines, beta blockers, calcium channel blockers, carbon monoxide, cardiac glycosides, cholinergic agents, cyanide, drug-induced dystonic reactions, ethylene glycol, fluoride, heavy metals, hypoglycemic agents, isoniazid, membrane-active agents, methemoglobinemia, opioids, sympathomimetics, and a variety of envenomations. Intravenous lipid emulsion has been shown to be a successful antidote for poisoning from various anesthetics and membrane-active agents (e.g., cyclic antidepressants), but the exact mechanism of benefit is still under investigation. Antidotes can significantly reduce morbidity and mortality rates but are potentially toxic if used for inappropriate reasons. Since their safe use requires correct identification of a specific poisoning or syndrome, details of antidotal therapy are discussed with the conditions for which they are indicated **(Table 470-4)**.

TABLE 470-4 Pathophysiologic Features and Treatment of Specific Toxic Syndromes and Poisonings

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Stimulated				
Sympathetics[a]				
Sympathomimetics	α_1-Adrenergic agonists (decongestants): phenylephrine, phenylpropanolamine β_2-Adrenergic agonists (bronchodilators): albuterol, terbutaline Nonspecific adrenergic agonists: amphetamines, cocaine, ephedrine	Stimulation of central and peripheral sympathetic receptors directly or indirectly (by promoting release or inhibiting reuptake of norepinephrine and sometimes dopamine)	Physiologic stimulation (Table 470-2). Reflex bradycardia can occur with selective α_1 agonists; β agonists can cause hypotension and hypokalemia.	Phentolamine, a nonselective α_1-adrenergic receptor antagonist, for severe hypertension due to α_1-adrenergic agonists; propranolol, a nonselective β blocker, for hypotension and tachycardia due to β_2 agonists; *either* labetalol, a β blocker with α-blocking activity, *or* phentolamine with esmolol, metoprolol, or another cardioselective β blocker for hypertension with tachycardia due to nonselective agents (β blockers, if used alone, can exacerbate hypertension and vasospasm due to unopposed α stimulation); benzodiazepines; propofol
Ergot alkaloids	Ergotamine, methysergide, bromocriptine, pergolide	Stimulation and inhibition of serotonergic and α-adrenergic receptors; stimulation of dopamine receptors	Physiologic stimulation (Table 470-2); formication; vasospasm with limb (isolated or generalized), myocardial, and cerebral ischemia progressing to gangrene or infarction. Hypotension, bradycardia, and involuntary movements can also occur.	Nitroprusside or nitroglycerine for severe vasospasm; prazosin (an α_1 blocker), captopril, nifedipine, and cyproheptadine (a serotonin receptor antagonist) for mild-to-moderate limb ischemia; dopamine receptor antagonists (antipsychotics) for hallucinations and movement disorders
Methylxanthines	Caffeine, theophylline	Inhibition of adenosine synthesis and adenosine receptor antagonism; stimulation of epinephrine and norepinephrine release; inhibition of phosphodiesterase resulting in increased intracellular cyclic adenosine and guanosine monophosphate	Physiologic stimulation (Table 470-2); pronounced gastrointestinal symptoms and β agonist effects (see above). Toxicity occurs at lower drug levels in chronic poisoning than in acute poisoning.	Propranolol, a nonselective β blocker, or esmolol for tachycardia with hypotension; any β blocker for supraventricular or ventricular tachycardia without hypotension; elimination enhanced by multiple-dose charcoal, hemoperfusion, and hemodialysis. Indications for hemoperfusion or hemodialysis include unstable vital signs, seizures, and a theophylline level of 80–100 μg/mL after an acute overdose and 40–60 μg/mL with chronic exposure.
Monoamine oxidase inhibitors	Phenelzine, tranylcypromine, selegiline	Inhibition of monoamine oxidase resulting in impaired metabolism of endogenous catecholamines and exogenous sympathomimetic agents	Delayed or slowly progressive physiologic stimulation (Table 470-2); terminal hypotension and bradycardia in severe cases	Short-acting agents (e.g., nitroprusside, esmolol) for severe hypertension and tachycardia; direct-acting sympathomimetics (e.g., norepinephrine, epinephrine) for hypotension and bradycardia

(*Continued*)

TABLE 470-4 Pathophysiologic Features and Treatment of Specific Toxic Syndromes and Poisonings (Continued)

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Anticholinergics				
Antihistamines	Diphenhydramine, doxylamine, pyrilamine	Inhibition of central and postganglionic parasympathetic muscarinic cholinergic receptors. At high doses, amantadine, diphenhydramine, orphenadrine, phenothiazines, and tricyclic antidepressants have additional nonanticholinergic activity (see below).	Physiologic stimulation (Table 470-2); dry skin and mucous membranes, decreased bowel sounds, flushing, and urinary retention; myoclonus and picking activity. Central effects may occur without significant autonomic dysfunction.	Physostigmine, an acetylcholinesterase inhibitor (see below), for delirium, hallucinations, and neuromuscular hyperactivity. Contraindications include asthma and non-anticholinergic cardiovascular toxicity (e.g., cardiac conduction abnormalities, hypotension, and ventricular arrhythmias).
Antipsychotics	Chlorpromazine, olanzapine, quetiapine, thioridazine	Inhibition of α-adrenergic, dopaminergic, histaminergic, muscarinic, and serotonergic receptors. Some agents also inhibit sodium, potassium, and calcium channels.	Physiologic depression (Table 470-2), miosis, anticholinergic effects (see above), extrapyramidal reactions (see below), tachycardia	Sodium bicarbonate for ventricular tachydysrhythmias associated with QRS prolongation; magnesium, isoproterenol, and overdrive pacing for torsades des pointes. Avoid class IA, IC, and III antiarrhythmics.
Belladonna alkaloids	Atropine, hyoscyamine, scopolamine	Inhibition of central and postganglionic parasympathetic muscarinic cholinergic receptors	Physiologic stimulation (Table 470-2); dry skin and mucous membranes, decreased bowel sounds, flushing, and urinary retention; myoclonus and picking activity. Central effects may occur without significant autonomic dysfunction.	Physostigmine, an acetylcholinesterase inhibitor (see below), for delirium, hallucinations, and neuromuscular hyperactivity. Contraindications include asthma and non-anticholinergic cardiovascular toxicity (e.g., cardiac conduction abnormalities, hypotension, and ventricular arrhythmias).
Cyclic antidepressants	Amitriptyline, doxepin, imipramine	Inhibition of α-adrenergic, dopaminergic, GABA-ergic, histaminergic, muscarinic, and serotonergic receptors; inhibition of sodium channels (see membrane-active agents); inhibition of norepinephrine and serotonin reuptake	Physiologic depression (Table 470-2), seizures, tachycardia, cardiac conduction delays (increased PR, QRS, JT, and QT intervals; terminal QRS right-axis deviation) with aberrancy and ventricular tachydysrhythmias; anticholinergic toxidrome (see above)	Hypertonic sodium bicarbonate (or hypertonic saline) for ventricular tachydysrhythmias associated with QRS prolongation. Use of phenytoin is controversial. Avoid class IA, IC, and III antiarrhythmics. IV emulsion therapy may be beneficial in some cases.
Mushrooms and plants	*Amanita muscaria* and *A. pantherina*, henbane, jimson weed, nightshade	Inhibition of central and postganglionic parasympathetic muscarinic cholinergic receptors	Physiologic stimulation (Table 470-2); dry skin and mucous membranes, decreased bowel sounds, flushing, and urinary retention; myoclonus and picking activity. Central effects may occur without significant autonomic dysfunction.	Physostigmine, an acetylcholinesterase inhibitor (see below), for delirium, hallucinations, and neuromuscular hyperactivity. Contraindications include asthma and nonanticholinergic cardiovascular toxicity (e.g., cardiac conduction abnormalities, hypotension, and ventricular arrhythmias).
Depressed				
Sympatholytics				
α_2-Adrenergic agonists	Clonidine, guanabenz, tetrahydrozoline and other imidazoline decongestants, tizanidine and other imidazoline muscle relaxants	Stimulation of α_2-adrenergic receptors leading to inhibition of CNS sympathetic outflow. Activity at nonadrenergic imidazoline binding sites also contributes to CNS effects.	Physiologic depression (Table 470-2), miosis. Transient initial hypertension may be seen.	Dopamine and norepinephrine for hypotension; atropine for symptomatic bradycardia; naloxone for CNS depression (inconsistently effective)
Antipsychotics	Chlorpromazine, clozapine, haloperidol, risperidone, thioridazine	Inhibition of α-adrenergic, dopaminergic, histaminergic, muscarinic, and serotonergic receptors. Some agents also inhibit sodium, potassium, and calcium channels.	Physiologic depression (Table 470-2), miosis, anticholinergic effects (see above), extrapyramidal reactions (see below), tachycardia. Cardiac conduction delays (increased PR, QRS, JT, and QT intervals) with ventricular tachydysrhythmias, including torsades des pointes, can sometimes develop.	Sodium bicarbonate for ventricular tachydysrhythmias associated with QRS prolongation; magnesium, isoproterenol, and overdrive pacing for torsades des pointes. Avoid class IA, IC, and III antiarrhythmics.

(*Continued*)

TABLE 470-4 Pathophysiologic Features and Treatment of Specific Toxic Syndromes and Poisonings (Continued)

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
β-Adrenergic blockers	Cardioselective (β_1) blockers: atenolol, esmolol, metoprolol Nonselective (β_1 and β_2) blockers: nadolol, propranolol, timolol Partial β agonists: acebutolol, pindolol α_1 Antagonists: carvedilol, labetalol Membrane-active agents: acebutolol, propranolol, sotalol	Inhibition of β-adrenergic receptors (class II antiarrhythmic effect). Some agents have activity at additional receptors or have membrane effects (see below).	Physiologic depression (Table 470-2), atrioventricular block, hypoglycemia, hyperkalemia, seizures. Partial agonists can cause hypertension and tachycardia. Sotalol can cause increased QT interval and ventricular tachydysrhythmias. Onset may be delayed after sotalol and sustained-release formulation overdose.	Glucagon for hypotension and symptomatic bradycardia. Atropine, isoproterenol, dopamine, dobutamine, epinephrine, and norepinephrine may sometimes be effective. High-dose insulin (with glucose and potassium to maintain euglycemia and normokalemia), electrical pacing, and mechanical cardiovascular support for refractory cases.
Calcium channel blockers	Diltiazem, nifedipine and other dihydropyridine derivatives, verapamil	Inhibition of slow (type L) cardiovascular calcium channels (class IV antiarrhythmic effect)	Physiologic depression (Table 470-2), atrioventricular block, organ ischemia and infarction, hyperglycemia, seizures. Hypotension is usually due to decreased vascular resistance rather than to decreased cardiac output. Onset may be delayed for ≥12 h after overdose of sustained-release formulations.	Calcium and glucagon for hypotension and symptomatic bradycardia. Dopamine, epinephrine, norepinephrine, atropine, and isoproterenol are less often effective but can be used adjunctively. High-dose insulin (with glucose and potassium to maintain euglycemia and normokalemia), IV lipid emulsion therapy, electrical pacing, and mechanical cardiovascular support for refractory cases.
Cardiac glycosides	Digoxin, endogenous cardioactive steroids, foxglove and other plants, toad skin secretions (Bufonidae spp.)	Inhibition of cardiac Na^+, K^+-ATPase membrane pump	Physiologic depression (Table 470-2); gastrointestinal, psychiatric, and visual symptoms; atrioventricular block with or without concomitant supraventricular tachyarrhythmia; ventricular tachyarrhythmias; hyperkalemia in acute poisoning. Toxicity occurs at lower drug levels in chronic poisoning than in acute poisoning.	Digoxin-specific antibody fragments for hemodynamically compromising dysrhythmias, Mobitz II or third-degree atrioventricular block, hyperkalemia (>5.5 meq/L; in acute poisoning only). Temporizing measures include atropine, dopamine, epinephrine, and external cardiac pacing for bradydysrhythmias and magnesium, lidocaine, or phenytoin, for ventricular tachydysrhythmias. Internal cardiac pacing and cardioversion can increase ventricular irritability and should be reserved for refractory cases.
Cyclic antidepressants	Amitriptyline, doxepin, imipramine	Inhibition of α-adrenergic, dopaminergic, GABA-ergic, histaminergic, muscarinic, and serotonergic receptors; inhibition of sodium channels (see membrane-active agents); inhibition of norepinephrine and serotonin reuptake	Physiologic depression (Table 470-2), seizures, tachycardia, cardiac conduction delays (increased PR, QRS, JT, and QT intervals; terminal QRS right-axis deviation) with aberrancy and ventricular tachydysrhythmias; anticholinergic toxidrome (see above)	Hypertonic sodium bicarbonate (or hypertonic saline) for ventricular tachydysrhythmias associated with QRS prolongation. Use of phenytoin is controversial. Avoid class IA, IC, and III antiarrhythmics. IV emulsion therapy may be beneficial in some cases.
Cholinergics				
Acetylcholinesterase inhibitors Muscarinic agonists Nicotinic agonists	Carbamate insecticides (aldicarb, carbaryl, propoxur) and medicinals (neostigmine, physostigmine, tacrine); nerve gases (sarin, soman, tabun, VX); organophosphate insecticides (diazinon, chlorpyrifos-ethyl, malathion) Bethanechol, mushrooms (*Boletus, Clitocybe, Inocybe* spp.), pilocarpine Lobeline, nicotine (tobacco)	Inhibition of acetylcholinesterase leading to increased synaptic acetylcholine at muscarinic and nicotinic cholinergic receptor sites Stimulation of CNS and postganglionic parasympathetic cholinergic (muscarinic) receptors Stimulation of preganglionic sympathetic and parasympathetic and striated muscle (neuromuscular junction) cholinergic (nicotine) receptors	Physiologic depression (Table 470-2). Muscarinic signs and symptoms: seizures, excessive secretions (lacrimation, salivation, bronchorrhea and wheezing, diaphoresis), and increased bowel and bladder activity with nausea, vomiting, diarrhea, abdominal cramps, and incontinence of feces and urine. Nicotinic signs and symptoms: hypertension, tachycardia, muscle cramps, fasciculations, weakness, and paralysis. Death is usually due to respiratory failure. Cholinesterase activity in plasma and red cells is <50% of normal in acetylcholinesterase inhibitor poisoning.	Atropine for muscarinic signs and symptoms; 2-PAM, a cholinesterase reactivator, for nicotinic signs and symptoms due to organophosphates, nerve gases, or an unknown anticholinesterase

(*Continued*)

TABLE 470-4 Pathophysiologic Features and Treatment of Specific Toxic Syndromes and Poisonings (Continued)

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Sedative-hypnotics[b]				
Anticonvulsants Barbiturates Benzodiazepines	Carbamazepine, ethosuximide, felbamate, gabapentin, lamotrigine, levetiracetam, oxcarbazepine, phenytoin, tiagabine, topiramate, valproate, zonisamide Short-acting: butabarbital, pentobarbital, secobarbital Long-acting: phenobarbital, primidone Ultrashort-acting: estazolam, midazolam, temazepam, triazolam Short-acting: alprazolam, flunitrazepam, lorazepam, oxazepam Long-acting: chlordiazepoxide, clonazepam, diazepam, flurazepam Pharmacologically related agents: zaleplon, zolpidem	Potentiation of the inhibitory effects of GABA by binding to the neuronal GABA–A chloride channel receptor complex and increasing the frequency or duration of chloride channel opening in response to GABA stimulation. Baclofen and, to some extent, GHB act at the GABA–B receptor complex. Meprobamate, its metabolite carisoprodol, felbamate, and orphenadrine antagonize NDMA excitatory receptors. Ethosuximide, valproate, and zonisamide decrease conduction through T-type calcium channels. Valproate decreases GABA degradation, and tiagabine blocks GABA reuptake. Carbamazepine, lamotrigine, oxcarbazepine, phenytoin, topiramate, valproate, and zonisamide slow the rate of recovery of inactivated sodium channels. Some agents also have α_2 agonist, anticholinergic, and sodium channel–blocking activity (see above and below).	Physiologic depression (Table 470-2), nystagmus. Delayed absorption can occur with carbamazepine, phenytoin, and valproate. Myoclonus, seizures, hypertension, and tachyarrhythmias can occur with baclofen, carbamazepine, and orphenadrine. Tachyarrhythmias can also occur with chloral hydrate. AGMA, hypernatremia, hyperosmolality, hyperammonemia, chemical hepatitis, and hypoglycemia can be seen in valproate poisoning. Carbamazepine and oxcarbazepine may produce hyponatremia from SIADH.	Benzodiazepines, barbiturates, or propofol for seizures. Hemodialysis and hemoperfusion may be indicated for severe poisoning by some agents (see "Extracorporeal Removal," in text). See above and below for treatment of anticholinergic and sodium channel (membrane)–blocking effects.
GABA precursors	γ-Hydroxybutyrate (sodium oxybate; GHB), γ-butyrolactone (GBL), 1,4-butanediol	Stimulation at GABA receptor complex increases chloride channel opening	Physiologic depression (Table 470-2)	Goal-directed supportive care
Muscle relaxants	Baclofen, carisoprodol, cyclobenzaprine, etomidate, metaxalone, methocarbamol, orphenadrine, propofol, tizanidine and other imidazoline muscle relaxants	Baclofen acts at GABA–B receptor complex; stimulation of α_2-adrenergic receptors inhibits CNS sympathetic outflow. Activity at nonadrenergic imidazoline binding sites also contributes to CNS effects. The others have centrally acting and various other unknown mechanisms of action.	Physiologic depression (Table 470-2)	Goal-directed supportive care; benzodiazepines and barbiturates for seizures
Discordant				
Asphyxiants				
Cytochrome oxidase inhibitors	Cyanide, hydrogen sulfide	Inhibition of mitochondrial cytochrome oxidase, with consequent blockage of electron transport and oxidative metabolism. Carbon monoxide also binds to hemoglobin and myoglobin and prevents oxygen binding, transport, and tissue uptake. (Binding to hemoglobin shifts the oxygen dissociation curve to the left.)	Signs and symptoms of hypoxemia with initial physiologic stimulation and subsequent depression (Table 470-2); lactic acidosis; normal PO_2 and calculated oxygen saturation but decreased oxygen saturation by co-oximetry. (That measured by pulse oximetry is falsely elevated but is less than normal and less than the calculated value.) Headache and nausea are common with carbon monoxide. Sudden collapse may occur with cyanide and hydrogen sulfide exposure. A bitter almond breath odor may be noted with cyanide ingestion, and hydrogen sulfide smells like rotten eggs.	High-dose oxygen; IV hydroxocobalamin or IV sodium nitrite and sodium thiosulfate (Lilly cyanide antidote kit) for coma, metabolic acidosis, and cardiovascular dysfunction in cyanide poisoning or victims from a fire; ECMO

(*Continued*)

TABLE 470-4 Pathophysiologic Features and Treatment of Specific Toxic Syndromes and Poisonings (Continued)

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
Methemoglobin inducers	Aniline derivatives, dapsone, local anesthetics, nitrates, nitrites, nitrogen oxides, nitro- and nitrosohydrocarbons, phenazopyridine, primaquine-type antimalarials, sulfonamides	Oxidation of hemoglobin iron from ferrous (Fe^{2+}) to ferric (Fe^{3+}) state prevents oxygen binding, transport, and tissue uptake. (Methemoglobinemia shifts oxygen dissociation curve to the left.) Oxidation of hemoglobin protein causes hemoglobin precipitation and hemolytic anemia (manifesting as Heinz bodies and "bite cells" on peripheral-blood smear).	Signs and symptoms of hypoxemia with initial physiologic stimulation and subsequent depression (Table 470-2), gray-brown cyanosis unresponsive to oxygen at methemoglobin fractions >15–20%, headache, lactic acidosis (at methemoglobin fractions >45%), normal PO_2 and calculated oxygen saturation but decreased oxygen saturation and increased methemoglobin fraction by co-oximetry. (Oxygen saturation by pulse oximetry may be falsely increased or decreased but is less than normal and less than the calculated value.)	High-dose oxygen; IV methylene blue for methemoglobin fraction >30%, symptomatic hypoxemia, or ischemia (contraindicated in G6PD deficiency); exchange transfusion and hyperbaric oxygen for severe or refractory cases
AGMA inducers	Ethylene glycol	Ethylene glycol causes CNS depression and increased serum osmolality. Metabolites (primarily glycolic acid) cause AGMA, CNS depression, and renal failure. Precipitation of oxalic acid metabolite as calcium salt in tissues and urine results in hypocalcemia, tissue edema, and crystalluria.	Initial ethanol-like intoxication, nausea, vomiting, increased osmolar gap, calcium oxalate crystalluria; delayed AGMA, back pain, renal failure; coma, seizures, hypotension, ARDS in severe cases	Sodium bicarbonate to correct acidemia; thiamine, folinic acid, magnesium, and high-dose pyridoxine to facilitate metabolism; ethanol or fomepizole for AGMA, crystalluria or renal dysfunction, ethylene glycol level >3 mmol/L (20 mg/dL), and ethanol-like intoxication or increased osmolal gap if level not readily obtainable; hemodialysis for persistent AGMA, lack of clinical improvement, and renal dysfunction; hemodialysis also useful for enhancing ethylene glycol elimination and shortening duration of treatment when ethylene glycol level is >8 mmol/L (50 mg/dL).
	Iron	Hydration of ferric (Fe^{3+}) ion generates H^+. Non-transferrin-bound iron catalyzes formation of free radicals that cause mitochondrial injury, lipid peroxidation, increased capillary permeability, vasodilation, and organ toxicity.	Initial nausea, vomiting, abdominal pain, diarrhea; AGMA, cardiovascular and CNS depression, hepatitis, coagulopathy, and seizures in severe cases. Radiopaque iron tablets may be seen on abdominal x-ray.	Whole-bowel irrigation for large ingestions; endoscopy and gastrostomy if clinical toxicity and large number of tablets are still visible on x-ray; IV hydration; sodium bicarbonate for acidemia; IV deferoxamine for systemic toxicity, iron level >90 μmol/L (500 μg/dL)
	Methanol	Methanol causes ethanol-like CNS depression and increased serum osmolality. Formic acid metabolite causes AGMA and retinal toxicity.	Initial ethanol-like intoxication, nausea, vomiting, increased osmolar gap; delayed AGMA, visual (clouding, spots, blindness) and retinal (edema, hyperemia) abnormalities; coma, seizures, cardiovascular depression in severe cases; possible pancreatitis	Gastric aspiration for recent ingestion; sodium bicarbonate to correct acidemia; high-dose folinic acid or folate to facilitate metabolism; ethanol or fomepizole for AGMA, visual symptoms, methanol level >6 mmol/L (20 mg/dL), and ethanol-like intoxication or increased osmolal gap if level not readily obtainable; hemodialysis for persistent AGMA, lack of clinical improvement, and renal dysfunction; hemodialysis also useful for enhancing methanol elimination and shortening duration of treatment when methanol level is >15 mmol/L (50 mg/dL)
	Salicylate	Increased sensitivity of CNS respiratory center to changes in and stimulates respiration. Uncoupling of oxidative phosphorylation, inhibition of Krebs cycle enzymes, and stimulation of carbohydrate and lipid metabolism generate unmeasured endogenous anions and cause AGMA.	Initial nausea, vomiting, hyperventilation, alkalemia, alkaluria; subsequent alkalemia with both respiratory alkalosis and AGMA and paradoxical aciduria; late acidemia with CNS and respiratory depression; cerebral and pulmonary edema in severe cases. Hypoglycemia, hypocalcemia, hypokalemia, and seizures can occur.	IV hydration and supplemental glucose; sodium bicarbonate to correct acidemia; urinary alkalinization for systemic toxicity; hemodialysis for coma, cerebral edema, seizures, pulmonary edema, renal failure, progressive acid-base disturbances or clinical toxicity, salicylate level >7 mmol/L (100 mg/dL) following acute overdose

(*Continued*)

TABLE 470-4 Pathophysiologic Features and Treatment of Specific Toxic Syndromes and Poisonings (Continued)

PHYSIOLOGIC CONDITION, CAUSES	EXAMPLES	MECHANISM OF ACTION	CLINICAL FEATURES	SPECIFIC TREATMENTS
CNS syndromes				
Extrapyramidal reactions	Antipsychotics (see above), some cyclic antidepressants and antihistamines	Decreased CNS dopaminergic activity with relative excess of cholinergic activity	Akathisia, dystonia, parkinsonism	Oral or parenteral anticholinergic agent such as benztropine or diphenhydramine
Isoniazid		Interference with activation and supply of pyridoxal-5-phosphate, a cofactor for glutamic acid decarboxylase, which converts glutamic acid to GABA, results in decreased levels of this inhibitory CNS neurotransmitter; complexation with and depletion of pyridoxine itself; inhibition of nicotine adenine dinucleotide–dependent lactate and hydroxybutyrate dehydrogenases, resulting in substrate accumulation	Nausea, vomiting, agitation, confusion; coma, respiratory depression, seizures, lactic and ketoacidosis in severe cases	High-dose IV pyridoxine (vitamin B_6) for agitation, confusion, coma, and seizures; diazepam or barbiturates for seizures
Lithium		Interference with cell membrane ion transport, adenylate cyclase and Na^+, K^+-ATPase activity, and neurotransmitter release	Nausea, vomiting, diarrhea, ataxia, choreoathetosis, encephalopathy, hyperreflexia, myoclonus, nystagmus, nephrogenic diabetes insipidus, falsely elevated serum chloride with low anion gap, tachycardia; coma, seizures, arrhythmias, hyperthermia, and prolonged or permanent encephalopathy and movement disorders in severe cases; delayed onset after acute overdose, particularly with delayed-release formulations. Toxicity occurs at lower drug levels in chronic poisoning than in acute poisoning.	Whole-bowel irrigation for large ingestions; IV hydration; hemodialysis for coma, seizures, encephalopathy or neuromuscular dysfunction (severe, progressive, or persistent), peak lithium level >4 meq/L following acute overdose
Serotonin syndrome	Amphetamines, cocaine, dextromethorphan, meperidine, MAO inhibitors, selective serotonin (5-HT) reuptake inhibitors, tricyclic antidepressants, tramadol, triptans, tryptophan	Promotion of serotonin release, inhibition of serotonin reuptake, or direct stimulation of CNS and peripheral serotonin receptors (primarily 5-HT-1a and 5-HT-2), alone or in combination	Altered mental status (agitation, confusion, mutism, coma, seizures), neuromuscular hyperactivity (hyperreflexia, myoclonus, rigidity, tremors), and autonomic dysfunction (abdominal pain, diarrhea, diaphoresis, fever, flushing, labile hypertension, mydriasis, tearing, salivation, tachycardia). Complications include hyperthermia, lactic acidosis, rhabdomyolysis, and multisystem organ failure.	Discontinue the offending agent(s); benzodiazepines for agitation or signs of stimulation; the serotonin receptor antagonist cyproheptadine may be helpful in severe cases.
Membrane-active agent	Amantadine, antiarrhythmics (class I and III agents; some β blockers), antipsychotics (see above), antihistamines (particularly diphenhydramine), carbamazepine, local anesthetics (including cocaine), opioids (meperidine, propoxyphene), orphenadrine, quinoline antimalarials (chloroquine, hydroxychloroquine, quinine), cyclic antidepressants (see above)	Blockade of fast sodium membrane channels prolongs phase 0 (depolarization) of the cardiac action potential, which prolongs QRS duration and promotes reentrant (monomorphic) ventricular tachycardia. Class Ia, Ic, and III antiarrhythmics also block potassium channels during phases 2 and 3 (repolarization) of the action potential, prolonging the JT interval and promoting early after-depolarizations and polymorphic (torsades des pointes) ventricular tachycardia. Similar effects on neuronal membrane channels cause CNS dysfunction. Some agents also block α-adrenergic and cholinergic receptors or have opioid effects (see above and Chap. 467).	QRS and JT prolongation (or both) with hypotension, ventricular tachyarrhythmias, CNS depression, seizures; anticholinergic effects with amantadine, antihistamines, carbamazepine, disopyramide, antipsychotics, and cyclic antidepressants (see above); opioid effects with meperidine and propoxyphene (see Chap. 467); cinchonism (hearing loss, tinnitus, nausea, vomiting, vertigo, ataxia, headache, flushing, diaphoresis), and blindness with quinoline antimalarials	Hypertonic sodium bicarbonate (or hypertonic saline) for cardiac conduction delays and monomorphic ventricular tachycardia; lidocaine for monomorphic ventricular tachycardia (except when due to class Ib antiarrhythmics); magnesium, isoproterenol, and overdrive pacing for polymorphic ventricular tachycardia; physostigmine for anticholinergic effects (see above); naloxone for opioid effects (see Chap. 467); extracorporeal removal for some agents (see text).

[a]See above and Chap. 468. [b]See above and Chap. 467.

Abbreviations: AGMA, anion-gap metabolic acidosis; ARDS, adult respiratory distress syndrome; CNS, central nervous system; ECMO, extracorporeal membrane oxygenation; GABA, γ-aminobutyric acid; GBL, γ-butyrolactone; GHB, γ-hydroxybutyrate; G6PD, glucose-6-phosphate dehydrogenase; MAO, monoamine oxidase; NDMA, *N*-methyl-D-aspartate; 2-PAM, pralidoxime; SIADH, syndrome of inappropriate antidiuretic hormone secretion.

PREVENTION OF REEXPOSURE

Poisoning is a preventable illness. Unfortunately, some adults and children are poison-prone, and recurrences are common. Unintentional polypharmacy poisoning has become especially common among adults with developmental delays, among the growing population of geriatric patients who are prescribed a large number of medications, and among adolescents and young adults also being prescribed a larger number of medications or experimenting with pharmaceuticals for recreational euphoria. Adults with unintentional exposures should be instructed regarding the safe use of medications and chemicals (according to labeling instructions). Confused patients may need assistance with the administration of their medications. Errors in dosing by health care providers may require educational efforts. Patients should be advised to avoid circumstances that result in chemical exposure or poisoning. Appropriate agencies and health departments (e.g., Occupational Health and Safety Administration [OSHA]) should be notified in cases of environmental or workplace exposure. The best approach to young children and patients with intentional overdose (deliberate self-harm or attempted suicide) is to limit their access to poisons. In households where children live or visit, alcoholic beverages, medications, household products (automotive, cleaning, fuel, pet-care, and toiletry products), inedible plants, and vitamins should be kept out of reach or in locked or child-proof cabinets. Depressed, bipolar, or psychotic patients should undergo psychiatric assessment, disposition, and follow-up. They should be given prescriptions for a limited supply of drugs with a limited number of refills and should be monitored for compliance and response to therapy. Expanding naloxone distribution, improving access to substance use disorder treatment, and linkage to other harm reduction services should be priorities in regions with a high incidence of opioid overdose.

SPECIFIC TOXIC SYNDROMES AND POISONINGS

Table 470-4 summarizes the pathophysiology, clinical features, and treatment of toxidromes and poisonings that are common, produce life-threatening toxicity, or require unique therapeutic interventions. In all cases, treatment should include attention to the general principles discussed above and, in particular, supportive care. Poisonings not covered in this chapter are discussed in specialized texts.

Alcohol, cocaine, hallucinogen, and opioid poisoning and alcohol and opioid withdrawal are discussed in Chaps. 464, 467, and 468; nicotine addiction is discussed in Chap. 465; acetaminophen poisoning is discussed in Chap. 351; the neuroleptic malignant syndrome is discussed in Chap. 446; and heavy metal poisoning is discussed in Chap. 469.

GLOBAL CONSIDERATIONS

Risks of poisoning in the United States and throughout the world are in transition. Patterns of travel, immigration, and internet consumerism should always be considered in patients suspected of poisoning or overdose without a clear etiology. Immigrants into various countries may have underlying poisoning from various metals from work or the environment where they previously lived; herbal remedies, food products, and cosmetics imported from overseas or ordered from the internet may be contaminated with metals, toxic plants, or other pharmaceutical contaminants; and new drugs of abuse that originate in one part of the world quickly circulate due to the ease afforded by the internet. Climate change has recently been identified as a key reason for evolving differences in the geographic pattern of poisoning injuries from snakebites and environmental toxic metal exposures. Expanding the history at the time of evaluation, recruiting the assistance of global health specialists, and ordering expanded laboratory panels may be indicated. For instance, during the COVID-19 pandemic, poisoning from household cleaning substances and novel untested therapies increased worldwide.

FURTHER READING

Dart RC et al: Expert consensus guidelines for stocking of antidotes in hospitals that provide emergency care. Ann Emerg Med 71:314, 2018.

Dart RC et al: Management of acetaminophen poisoning in the US and Canada: A consensus statement. JAMA Netw Open 2023;6(8):e2327739.

Hughes A et al: The Toxicology Investigators Consortium 2023 Annual Report. J Med Toxicol 20:350, 2024.

Gummin DD et al: 2021 Annual Report of the National Poison Data System© (NPDS) from America's Poison Centers: 39th annual report. Clin Toxicol (Phila) 60:1381, 2022.

Mattson CL et al: Trends and geographic patterns in drug and synthetic opioid overdose deaths — United States, 2013–2019. MMWR Morb Mortal Wkly Rep 70:202, 2021.

Mycyk MB: ECMO shows promise for treatment of poisoning some of the time: The challenge to do better by aiming higher. Crit Care Med 48:1235, 2020.

Mycyk MB et al: Growth in recreational cannabis markets and burden on emergency departments. JAMA Netw Open 4:e2125275, 2021.

Nelson LS et al (eds): *Goldfrank's Toxicologic Emergencies*, 11th ed. New York, McGraw-Hill, 2019.

Thompson TM et al: The general approach to the poisoned patient. Dis Mon 60:509, 2014.

471

Disorders Caused by Venomous Snakebites and Marine Animal Exposures

Nathaniel Mann, Michael D. Simpson, Erik Fisher

This chapter outlines general principles for the evaluation and management of victims of envenomation and poisoning by venomous snakes and marine animals. Because the incidence of serious bites and stings is relatively low in developed nations, true incidence and robust clinical data are often limited. As a result, therapeutic decision-making is frequently based on anecdotal information.

VENOMOUS SNAKEBITE

EPIDEMIOLOGY

The venomous snakes of the world belong to the families Viperidae (subfamily Viperinae: Old World vipers; subfamily Crotalinae: New World and Asian pit vipers), Elapidae (including cobras, coral snakes, sea snakes, kraits, and all Australian venomous snakes), Lamprophiidae (subfamily Atractaspidinae: burrowing asps), and Colubridae (a large family in which most species are nonvenomous). Death from snake envenomation represents the deadliest neglected tropical disease in the world. Most snakebites occur in developing countries with temperate and tropical climates in which populations subsist on agriculture and fishing **(Fig. 471-1)**. While estimates vary, worldwide data from 2019 suggest 63,415 (95% uncertainty interval (UI), 38,930–78,633) deaths globally from snake envenomation with a projected increase to 68,799 (95% UI, 41,372–97,613) by 2030, despite a World Health Organization (WHO) goal to half snake envenomation deaths over the same time period. There are an estimated 1.2 million to 5.5 million snakebites worldwide each year, with 421,000–1,200,000 envenomations. Such wide-ranging estimates reflect the challenges of collecting accurate data in the regions most affected by venomous snakes. Many victims in these areas either do not seek medical attention or have insufficient access to antivenom, and reporting and record-keeping are generally poor.

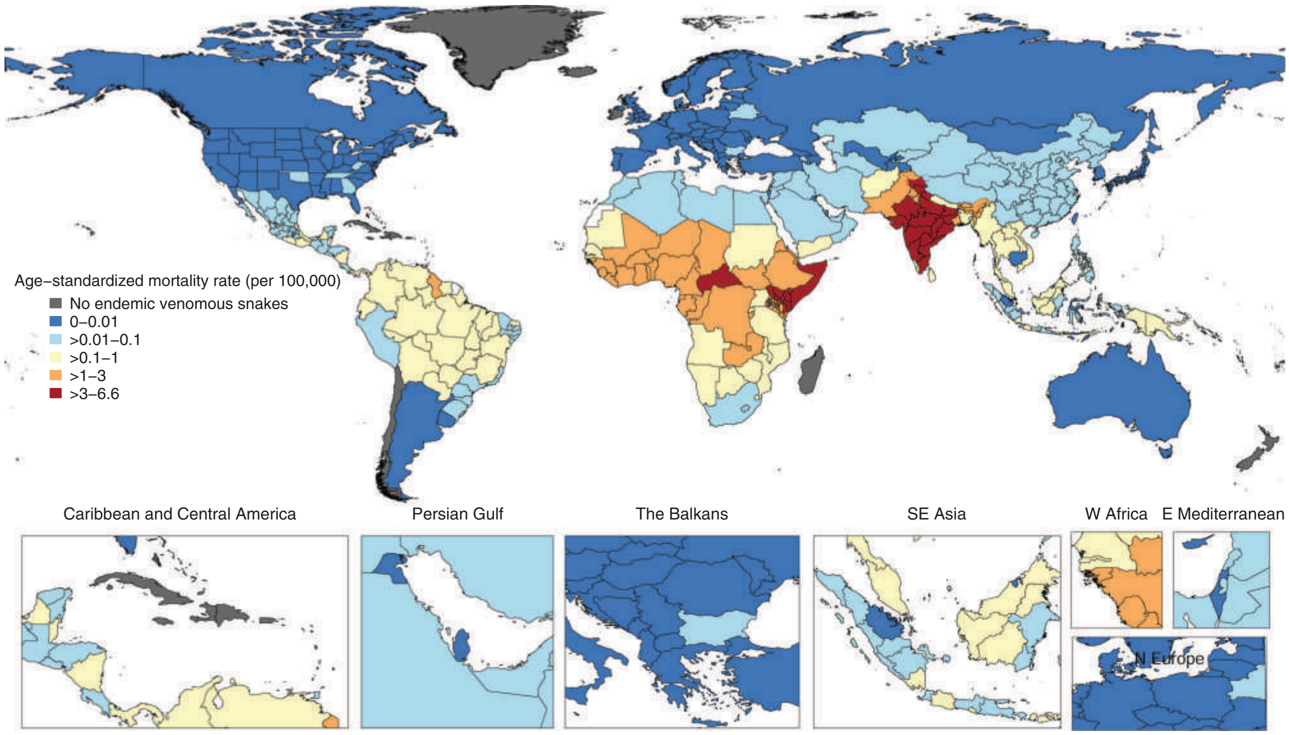
Age–standardized mortality rate (per 100,000)
No endemic venomous snakes
0–0.01
>0.01–0.1
>0.1–1
>1–3
>3–6.6
Caribbean and Central America
Persian Gulf
The Balkans
SE Asia
W Africa
E Mediterranean
N Europe

FIGURE 471-1 Age-standardized mortality rate of snakebite envenoming in 2019. GBD 2019 did not publish state-level estimates for China, so each state is colored based on China's national estimate. *(Reproduced from GBD 2019 Snakebite Envenomation Collaborators: Global Mortality of Snakebite Envenoming Between 1990 and 2019. Nat Commun 13:1, 2022.)*

SNAKE ANATOMY/IDENTIFICATION

The typical snake venom delivery apparatus consists of bilateral venom glands situated behind the eyes and hollow anterior maxillary fangs. In viperids, these fangs are long and highly mobile; they are retracted against the roof of the mouth when the snake is at rest, then brought to an upright position when about to strike. In elapids, the fangs are fixed in an erect position and smaller in size, which can lead to fewer distinct wounds after an envenomation. Venomous colubrids do not possess venom glands but instead have homologous structures known as Duvernoy's glands. Approximately 20–25% of pit viper bites and higher percentages of other snakebites (up to 75% for sea snakes) are "dry" bites, in which no venom is released.

Differentiating between venomous and nonvenomous snake species can be challenging. Identifying venomous snakes by color pattern can be misleading, as many nonvenomous snakes have color patterns that closely mimic those of venomous snakes found in the same region. Viperids are characterized by triangular heads (a feature shared with, or mimicked by, many harmless snakes), elliptical pupils (also seen in some nonvenomous snakes, such as boas and pythons), and enlarged maxillary fangs; pit vipers also possess heat-sensing organs (loreal pits) on each side of the head that assist with locating prey and aiming strikes. Rattlesnakes possess a series of interlocking hollow keratin plates (the rattle) on the tip of the tail that emit a buzzing sound when vibrated rapidly; this sound serves as a warning signal to perceived threats.

VENOMS AND CLINICAL MANIFESTATIONS

Snake venoms consist of highly complex mixtures of enzymes, polypeptides, glycoproteins, and other constituents. Venom components may vary greatly depending on the species, age, diet, and geographic location of the snake. Venom variability even exists within members of the same species. Snake venoms can cause local tissue necrosis, affect the coagulation pathway at various steps, impair organ function, or act at the neuromuscular junction to cause paralysis.

After an envenomation, the time to symptom onset and clinical presentation can be quite variable depending on the species involved, the anatomic location of the bite, and the amount of venom injected. Envenomations by most viperids and some elapids cause progressive local pain, soft tissue swelling, and ecchymosis **(Fig. 471-2)**. Hemorrhagic or serum-filled vesicles and bullae may develop at the bite site over a period of hours to days. In serious bites, tissue loss can be significant **(Fig. 471-3)**. Systemic findings are variable and can include generalized fatigue, nausea or vomiting, changes in taste, mouth numbness, tachycardia or bradycardia, hypotension, muscle fasciculations, pulmonary edema, renal dysfunction, or spontaneous hemorrhage. Envenomations by neurotoxic elapids, such as kraits (*Bungarus* species), many Australian elapids (e.g., death adders [*Acanthophis* species] and tiger snakes [*Notechis* species]), and some cobras (*Naja* species), as well as some viperids (e.g., the South American rattlesnake [*Crotalus durissus*], Mojave rattlesnake [*Crotalus scutulatus*], some Timber rattlesnakes [*Crotalus horridus*], and certain Indian Russell's vipers [*Daboia russelii*]), cause neurologic dysfunction. Early findings may consist of nausea and vomiting, headache, paresthesias or numbness, or altered mental status. Victims may develop cranial nerve abnormalities (e.g., ptosis, difficulty swallowing), followed by peripheral motor weakness. Severe envenomation may result in diaphragmatic paralysis and lead to death from respiratory failure and aspiration. Sea snake envenomation results in local pain (variable), generalized myalgias, trismus, rhabdomyolysis, and progressive flaccid paralysis; these manifestations can be delayed for several hours.

TREATMENT

Venomous Snakebite

FIELD MANAGEMENT

The most important aspect of prehospital care of a person bitten by a venomous snake is rapid transport to a medical facility equipped to provide supportive management (airway, breathing, and circulation) and antivenom therapy. Any jewelry or tight-fitting clothing near the bite should be removed to avoid constriction from anticipated soft tissue swelling. Although wound care should not delay transport, the wound should be cleaned with soap and running water and then covered with a sterile dressing. It is reasonable to apply a splint to the bitten extremity to limit movement and assist with positioning. If possible, the extremity should be maintained in

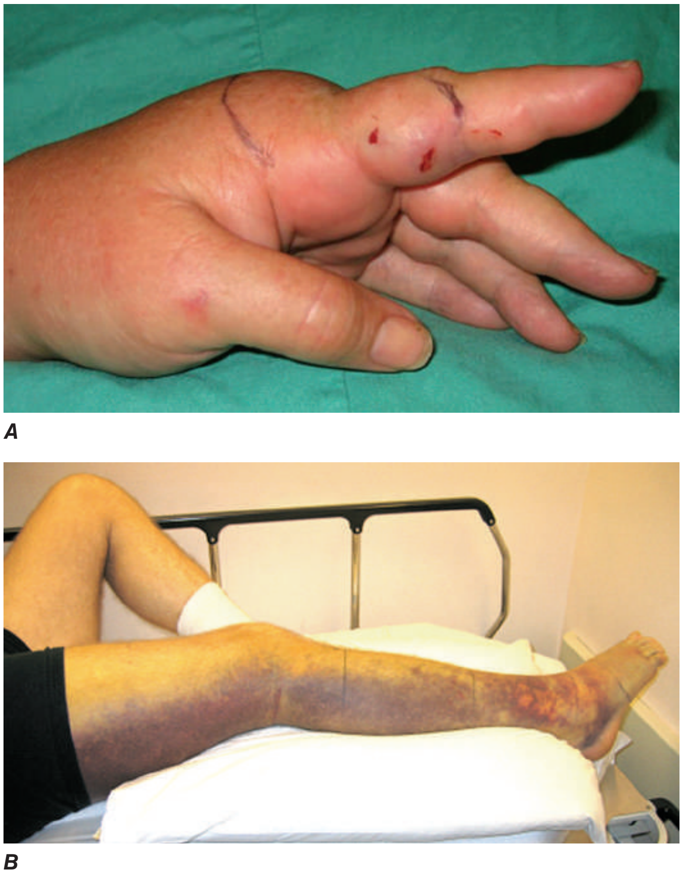

FIGURE 471-2 Northern Pacific rattlesnake (*Crotalus oreganus oreganus*) envenomations. ***A.*** Moderately severe envenomation. Note edema and early ecchymosis 2 h after a bite to the finger. ***B.*** Severe envenomation. Note extensive ecchymosis 5 days after a bite to the ankle. *(Courtesy of Robert Norris, with permission.)*

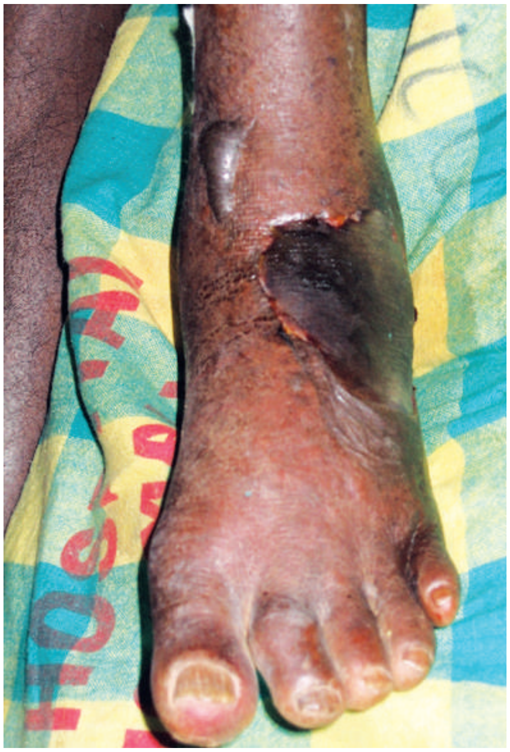

FIGURE 471-3 Early stages of severe, full-thickness necrosis 5 days after a Russell's viper (*Daboia russelii*) bite in southwestern India. *(Courtesy of Robert Norris, with permission.)*

a neutral position of comfort at approximately heart level. Attempting to capture and transport the offending snake is not advised; instead, digital photographs taken from a safe distance may assist with snake identification and treatment decisions.

Most of the first-aid measures recommended in the past are of little benefit and may worsen outcomes. Incising and/or applying suction to the bite site should be avoided, as these measures exacerbate local tissue damage, increase the risk of infection, and have not been shown to be effective. Venom sequestration devices (e.g., lympho-occlusive bandages or tourniquets) are not advised, as they may intensify local tissue damage by restricting the spread of potentially necrotizing venom and may contribute to loss of function or ischemia. In developing countries, victims should be encouraged to seek immediate treatment at a medical facility equipped with antivenom instead of consulting traditional healers and thus incurring significant delays in reaching appropriate care.

Elapid envenomations that are primarily neurotoxic and have no significant effects on local tissue may be managed with pressure-immobilization, in which the entire bitten limb is immediately wrapped with a bandage and then immobilized. This technique is meant to restrict lymphatic drainage and has been shown to delay the systemic absorption of venom from predominantly neurotoxic species but is not recommend for viperid envenomation. For pressure-immobilization to be effective, the wrap pressure should not exceed 40–70 mmHg in upper-extremity application and 55–70 mmHg in lower-extremity application. As an estimate, the bandage should be snug enough to apply pressure but loose enough for a finger to slip underneath. Additionally, the victim must be carried out of the field because walking generates muscle-pumping activity that, regardless of the anatomic site of the bite, will disperse venom into the systemic circulation. Pressure-immobilization should be used only in cases in which the offending snake is reliably identified and known to be primarily neurotoxic, the rescuer is skilled in pressure-wrap application, the necessary supplies are readily available, and the victim can be fully immobilized and carried to medical care—a rare combination of conditions, particularly in the regions of the world where such bites are most common.

HOSPITAL MANAGEMENT

Initial hospital management should focus on the victim's airway, breathing, and circulation. Patients with bites to the face or neck may require early endotracheal intubation to prevent loss of airway patency caused by rapid soft-tissue swelling. Vital signs, cardiac rhythm, oxygen saturation, and urine output should be closely monitored. Two large-bore IV lines should be established in unaffected extremities. Because of the potential for coagulopathy, venipuncture attempts should be minimized and noncompressible sites (e.g., subclavian vein) avoided. Early hypotension may be caused by bradykinin-potentiating factors, which lead to vasodilation and pooling of blood in the pulmonary and splanchnic vascular beds. Additionally, systemic bleeding, hemolysis, and loss of intravascular volume into the soft tissues may play important roles in hypotension. Fluid resuscitation with isotonic crystalloid such as saline or lactated Ringer's (20–40 mL/kg IV) should be initiated if there is any evidence of hemodynamic instability. Vasopressors (e.g., norepinephrine, epinephrine) should be considered if venom-induced shock persists after aggressive volume resuscitation and antivenom administration (see below), as the victim may be experiencing anaphylaxis to the venom components or the antivenom itself.

A thorough history (including the time of the bite and any symptoms of envenomation) should be obtained and a complete physical examination performed, with a focus on the neurovascular status of the site of envenomation. For envenomation of a limb, palpation of axillary or inguinal lymph nodes can provide information about lymphatic spread. Bandages or wraps applied in the field should be removed as soon as possible, with cognizance that the release of such ligatures may result in hypotension or dysrhythmias when stagnant acidotic blood containing venom is released into the systemic circulation. To objectively evaluate the progression of local envenomation, the leading edge of swelling, ecchymosis, and tenderness should be marked, and limb circumference should be measured at three points (e.g., at the bite site, the joint proximal, and the joint distal) every 15 min until the local tissue effects have stabilized. Once stabilized, measurements can be taken every 1–2 h. During this period of observation, the bitten extremity should be positioned above the level of the heart. Victims of neurotoxic envenomation should be monitored closely for evidence of cranial nerve dysfunction (e.g., ptosis) that may precede more overt signs of impending airway compromise (e.g., difficulty swallowing, respiratory insufficiency) necessitating endotracheal intubation and mechanical ventilation.

Blood should be drawn for laboratory evaluation as soon as possible. Important studies include a complete blood count to determine the degree of hemorrhage or hemolysis and to detect thrombocytopenia; blood type and cross-matching; assessment of renal and hepatic function; coagulation studies such as prothrombin time and fibrinogen level; measurement of fibrin degradation products such as a ᴅ-dimer; measurement of creatine kinase for suspected rhabdomyolysis; and testing of urine for blood or myoglobin. In developing regions, the 20-min whole-blood clotting test can be used to diagnose coagulopathy when access to laboratory tests is limited. To perform this test, 1–2 mL of venous blood is placed in a clean, dry glass receptacle (e.g., a test tube) and left undisturbed for 20 min. The sample is then inverted. If the blood is still liquid and a clot has not formed, coagulopathy is present. There have been some recent studies examining the utility of thromboelastography (TEG) in detecting venom-induced consumptive coagulopathy. In vitro studies have shown that TEG may be more sensitive than conventional coagulation studies in detecting coagulopathy at lower venom concentrations; however, it remains to be seen if this bears any clinical significance. Currently, there is insufficient evidence to suggest routine use of TEG. Electrocardiography and chest radiography may be helpful in severe envenomations or when there is significant comorbidity. After antivenom therapy (see below), laboratory values should be rechecked every 6 h until clinical stability is achieved.

The mainstay of treatment of a venomous snakebite resulting in significant envenomation is prompt administration of specific antivenom. Antivenoms are produced by injecting animals (generally horses or sheep) with venoms from medically important snakes. Once the stock animals develop antibodies to the venoms, their serum is harvested, and the antibodies are isolated for antivenom preparation. The goal of antivenom administration is to allow antibodies (or antibody fragments) to bind and deactivate circulating venom components before they can attach to target tissues and cause deleterious effects. Antivenoms may be monospecific (directed against a particular snake species) or polyspecific (covering several species in a geographic region). Unless the species are known to have homologous venoms, antivenoms rarely offer cross-protection against snake species other than those used in their production. Thus, antivenom selection must be specific for the offending snake; if the antivenom chosen does not contain antibodies to that snake's venom components, it will provide no benefit and may lead to unnecessary complications (see below). In the United States, assistance in finding appropriate antivenom can be obtained from a regional poison control center, which can be reached by telephone 24 h a day at (800) 222-1222.

For victims of bites by viperids or cytotoxic elapids, indications for antivenom administration include significant progressive local findings (e.g., rapidly spreading soft-tissue swelling, swelling that involves more than half the bitten limb, extensive blistering or bruising, severe pain) and any evidence of systemic envenomation (e.g., systemic symptoms or laboratory abnormalities). Caution must be used when determining the significance of isolated pain or soft-tissue swelling after the bite of an unidentified snake because the saliva of some relatively harmless species can cause mild discomfort or edema at the bite site; in such bites, antivenoms are useless and potentially harmful. Antivenoms have limited efficacy in preventing local tissue damage caused by necrotizing venoms, as venom components bind to local tissues very quickly. Nevertheless, antivenom should be administered as soon as the need for it is identified to limit further tissue damage and systemic effects. Antivenom administration after bites by neurotoxic elapids is indicated at the first sign of neurotoxicity (e.g., cranial nerve dysfunction, peripheral neuropathy). In general, antivenom is effective only in reversing active venom toxicity; it is of little benefit in reversing effects that have already been established (e.g., renal failure, established paralysis); these improve only with time and other supportive therapies.

Specific comments related to the management of venomous snakebites in the United States and Canada appear in **Table 471-1**. The package insert for the selected antivenom should be consulted regarding species covered, method of administration, starting dose, and need (if any) for redosing. Whenever possible, it is advisable for health care providers to seek advice from experts in snakebite management regarding indications for and dosing of antivenom.

Antivenom should only be administered intravenously, and the infusion should be started slowly, with the treating clinician at the bedside ready to immediately intervene at the first signs of an acute adverse reaction. In the absence of an adverse reaction, the rate of infusion can be increased gradually until the full starting dose has been administered (over a total period of ~1 h). Further antivenom may be necessary if the patient's acute clinical condition worsens or fails to stabilize, or if venom effects recur. The decision to administer further antivenom to a stabilized patient should be based on clinical evidence of the persistent circulation of unbound venom components. For viperid bites, antivenom administration should generally be continued until the victim shows definite improvement (e.g., reduced pain, stabilized vital signs, restored coagulation). Neurotoxicity from elapid bites may be more difficult to reverse with antivenom. Once neurotoxicity is established and endotracheal intubation is required, further doses of antivenom are unlikely to be beneficial. In such cases, the victim must be maintained on mechanical ventilation until recovery, which may take days to weeks.

Adverse reactions to antivenom administration include early (anaphylaxis) and late (serum sickness) hypersensitivity reactions. Clinical manifestations of early hypersensitivity may include tachycardia, rigors, vomiting, urticaria, dyspnea, laryngeal edema, bronchospasm, and hypotension. Although recommended by some antivenom manufacturers, skin testing for potential early hypersensitivity is neither sensitive nor specific and is of no benefit. The quality of antivenoms is highly variable worldwide; rates of acute anaphylactic reactions to some of these products exceed 50%. More recent data on North American snakebites suggest that the incidence of anaphylaxis to Crotalidae polyvalent immune fab (ovine) (CroFab®) (BTG International Inc., West Conshohocken, PA) antivenom may be closer to 1.1%. Incidence data on anaphylaxis to Crotalidae immune f(ab′) 2 (equine) (Anavip®) (Instituto Bioclon S.A de C.V., Tlalpan CDMX, Mexico for Rare Disease Therapeutics, Inc., Franklin, TN) are not currently available. Limited evidence from Sri Lanka supports routine pretreatment with low-dose SC epinephrine (0.25 mg of 1:1000 aqueous solution) to prevent acute anaphylactic reactions after antivenom infusion in that population. Although widely practiced, prophylactic use of antihistamines and glucocorticoids has not proved beneficial. Modest expansion of the patient's intravascular volume with crystalloids may blunt episodes of acute hypotension during antivenom infusion. Epinephrine and airway equipment should always be immediately available. If the patient develops an anaphylactic reaction to antivenom, the infusion should be stopped temporarily and the reaction treated immediately with IM epinephrine (0.01 mg/kg up to 0.5 mg), an IV antihistamine (e.g., diphenhydramine, 1 mg/kg up to 50 mg), and a glucocorticoid (e.g., hydrocortisone, 2 mg/kg up to 100 mg). Once the reaction has been controlled, if the severity of the envenomation warrants additional antivenom, the dose should be restarted as soon as possible at a slower rate (5–10 mL/h) and

TABLE 471-1 Management of Venomous Snakebites in the United States and Canada[a]

Pit Viper Bites: Rattlesnakes (*Crotalus* and *Sistrurus* spp.), Cottonmouth Water Moccasins (*Agkistrodon piscivorus*), and Copperheads (*Agkistrodon contortrix*)

- Stabilize airway, breathing, and circulation.
- Institute monitoring (vital signs, cardiac rhythm, and oxygen saturation).
- Establish two large-bore IV lines.
  - If patient is hypotensive, administer isotonic crystalloid bolus (20–40 mL/kg IV).
- Take thorough history and perform complete physical examination.
- Identify offending snake (if possible).
- Measure and record circumference of bitten extremity every 15 min until swelling has stabilized. Can take measurements every 1 h once stabilized.
- Order laboratory studies (CBC, blood type and cross-matching, metabolic panel, PT/INR/PTT, fibrinogen level, FDP, CK, urinalysis).
  - If normal, repeat CBC and coagulation studies every 4 h until it is clear that no systemic envenomation has occurred.
  - If abnormal, repeat every 6 h after antivenom administration until swelling has stabilized (see below).
- Determine severity of envenomation.
  - None: fang marks only ("dry" bite)
  - Mild: local findings only (e.g., pain, ecchymosis, nonprogressive swelling)
  - Moderate: swelling that is clearly progressing, systemic symptoms or signs, and/or laboratory abnormalities
  - Severe: neurologic dysfunction, respiratory distress, and/or cardiovascular instability/shock
- Contact regional poison control center.
- Locate and administer antivenom as indicated: Crotalidae polyvalent immune Fab (Ovine) (CroFab®) (BTG International Inc., West Conshohocken, PA) or Crotalidae immune $F(ab')_2$ (Equine) (Anavip®) (Instituto Bioclon S.A de C.V., Tlalpan CDMX, Mexico).
  - Starting dose:
    - Based on severity of envenomation
      - None or mild: none
      - Moderate: CroFab® (4–6 vials), Anavip® (10 vials)
      - Severe: CroFab® (6 vials), Anavip® (10 vials)
    - Dilute reconstituted vials in 250 mL of normal saline.
    - Infuse IV over 1 h (with medical provider in close attendance).
      - Start at rate of 25–50 mL/h for first 10 min.
      - If there is no allergic reaction, increase rate to 250 mL/h.
  - If there is an acute reaction to antivenom:
    - Stop infusion.
    - Treat with standard doses of epinephrine (IM or IV; latter route only in setting of severe hypotension), antihistamines (IV), and glucocorticoids (IV).
    - When reaction is controlled, restart antivenom as soon as possible (at rate of 5–10 mL/h; titrate up as tolerated).
  - Monitor clinical status over 1 h.
    - Stabilized or improved: Admit to hospital.
    - Progressing or unimproved: Repeat starting dose. Continue this pattern until patient's condition is stabilized or improved. Admit to ICU if possible.
- Blood products are rarely needed; if required, they should be given only after antivenom administration.
- Provide tetanus immunization as needed.
- Prophylactic antibiotics are unnecessary unless prehospital care included incision or mouth suction.
- Pain management: Administer acetaminophen and/or opioids as needed; avoid salicylates and nonsteroidal anti-inflammatory agents.
- Admit patient to hospital. (If there is no evidence of envenomation, monitor for 8–12 h before discharge.)
  - Give additional antivenom: CroFab® (2 vials every 6 h for 3 additional doses), Anavip® (4 vials for recurrent coagulopathy).
  - Monitor for evidence of rising intracompartmental pressures (see text).
  - Provide wound care (see text).
  - Start physical therapy (see text).
- At discharge, warn patient of possible recurrent coagulopathy and symptoms/signs of serum sickness. Schedule repeat laboratory testing to monitor for recurrent coagulopathy.

Coral Snake Bites: Eastern coral snake (*Micrurus fulvius*), Texas coral snake (*Micrurus tener*), and Sonoran coral snake (*Micruroides euryxanthus*)

- Stabilize airway, breathing, and circulation.
- Institute monitoring (vital signs, cardiac rhythm, and oxygen saturation).
- Establish two large-bore IV lines and initiate normal saline infusion.
- Take thorough history and perform complete physical examination.
- Identify offending snake (if possible).
- Laboratory studies are unlikely to be helpful.
- Contact regional poison control center.
- Locate and administer antivenom as indicated: Antivenin (*Micrurus fulvius*) (Equine) (commonly referred to as North American Coral Snake Antivenin; Pfizer Inc., Philadelphia, PA).[b]
  - Refer to antivenom package insert.
  - Dilute 3–5 reconstituted vials in 250–500 mL of normal saline.
  - Infuse IV over 1 h (with medical provider in close attendance).
  - If signs of envenomation progress despite initial dosing, repeat starting dose; up to 10 vials total may be required.

(*Continued*)

TABLE 471-1 Management of Venomous Snakebites in the United States and Canada[a] (Continued)

- If there is an acute adverse reaction to antivenom:
  - Stop infusion.
  - Treat with standard doses of epinephrine (IM or IV; latter route only in setting of severe hypotension), antihistamines (IV), and glucocorticoids (IV).
  - When reaction is controlled, restart antivenom as soon as possible (at rate of 5–10 mL/h; titrate up as tolerated).
- If there is evidence of neurologic dysfunction (e.g., any cranial nerve abnormalities):
  - Administer trial of acetylcholinesterase inhibitors (see Table 471-2).
  - If there is evidence of difficulty swallowing or breathing, proceed with endotracheal intubation and ventilatory support (may be required for days to weeks).
- Provide tetanus immunization as needed.
- Prophylactic antibiotics are unnecessary unless prehospital care included incision or mouth suction.
- Admit patient to hospital (ICU) even if there is no evidence of envenomation; monitor for at least 24 h.

[a] These recommendations are specific to the care of victims of venomous snakebites in the United States and Canada and should not be applied to bites in other regions of the world. [b] At the time of this writing, lot CL6814 of Antivenin has an extended expiration date of June 30, 2024.

Abbreviations: CBC, complete blood count; CK, creatine kinase; FDP, fibrin degradation products; ICU, intensive care unit; PT/INR/PTT, prothrombin time/international normalized ratio/partial thromboplastin time.

titrated upward as tolerated. In rare cases of refractory hypotension, a concomitant IV infusion of epinephrine may be initiated and titrated to clinical effect while antivenom is administered. The patient must be monitored very closely during such therapy, preferably in an emergency department or intensive care setting. Serum sickness typically develops 1–2 weeks after antivenom administration and may present as myalgias, arthralgias, fever, chills, urticaria, lymphadenopathy, or renal or neurologic dysfunction. Treatment for serum sickness consists of systemic glucocorticoids (e.g., oral prednisone, 1–2 mg/kg daily) until all symptoms have resolved, with a subsequent taper over 1–2 weeks. Oral antihistamines and analgesics may provide additional relief of symptoms.

Blood products are rarely necessary in the management of an envenomated patient. The venoms of many snake species can deplete coagulation factors and cause a decrease in platelet count or hematocrit. Nevertheless, these components usually rebound within hours after administration of adequate antivenom. If the need for blood products is thought to be great (e.g., a dangerously low platelet count in a hemorrhaging patient), these products should be given only after adequate antivenom administration to avoid fueling ongoing consumptive coagulopathy.

Rhabdomyolysis should be managed in standard fashion with IV fluids and close monitoring of urine output. Victims who develop acute renal failure should be evaluated by a nephrologist and referred for hemodialysis or peritoneal dialysis as needed. Such renal failure is usually due to acute tubular necrosis and is frequently reversible. If bilateral cortical necrosis occurs, however, the prognosis for renal recovery is less favorable, and long-term dialysis with possible renal transplantation may be necessary.

Most snake envenomations involve subcutaneous deposition of venom. On occasion, venom can be injected more deeply into muscle compartments, particularly if the offending snake was large and the bite occurred on the hand, forearm, or anterior compartment of the lower leg. Intramuscular swelling of the affected extremity may be accompanied by severe pain, decreased strength, altered sensation, cyanosis, and apparent pulselessness—signs suggesting a muscle compartment syndrome. If compartment syndrome is suspected, intracompartmental pressures may be measured, preferably with a minimally invasive technique (e.g., with a wick catheter or digital readout device). Elevation of compartment pressures above 30–40 mmHg is an indication for additional antivenom administration. The extremity should be elevated while antivenom is delivered.

Traditionally, acute compartment syndrome is treated with fasciotomy. However, in North American crotalid envenomations, there is no evidence to support fasciotomy and it should not be recommended, as modern antivenoms are effective in reducing compartment pressure. Current data suggest notable risk of increased morbidity with fasciotomy and no evidence of benefit for these patients. However, there are scant data from studies on Asian snake species that support fasciotomy in very limited circumstances. In these patients, if the intracompartmental pressure remains elevated after 1 h of such therapy, a surgical consultation may be obtained for possible fasciotomy, though published data support allowing compartment pressures up to 55 mmHg for up to 6 h before considering fasciotomy. This procedure should always be one of last resort. Fortunately, the incidence of compartment syndrome is very low after a snakebite, with fasciotomies required in <1% of cases. Nevertheless, vigilance is essential.

Acetylcholinesterase inhibitors (e.g., edrophonium and neostigmine) may promote neurologic improvement in patients bitten by snakes with postsynaptic neurotoxins. Snakebite victims with objective evidence of neurologic dysfunction may receive a test dose of acetylcholinesterase inhibitors, as outlined in **Table 471-2**. If they exhibit improvement, additional doses of long-acting neostigmine can be administered as needed. Acetylcholinesterase inhibitors are not a substitute for endotracheal intubation or the administration of appropriate antivenom when available.

Care of the bite wound includes simple cleansing with soap and water; application of a dry, sterile dressing; and splinting of the affected extremity with padding between the digits. Once antivenom therapy has been initiated, the extremity should be elevated to heart level to reduce swelling. Patients should receive tetanus immunization as appropriate. Prophylactic antibiotics are generally unnecessary after bites by North American snakes because the incidence of secondary infection is low. In some regions, secondary bacterial infection is more common, and the consequences are serious; as such, prophylactic treatment with broad-spectrum antibiotics may be appropriate. Antibiotics may also be considered if misguided first-aid efforts included incision or mouth suction of the bite site. Pain control may be achieved with acetaminophen or opioid analgesics. Salicylates and nonsteroidal anti-inflammatory agents should generally be avoided because of their potential effects on blood clotting.

TABLE 471-2 Use of Acetylcholinesterase Inhibitors in Envenomations by Neurotoxic Snakes and Cone Snails

1. Patients with clear, objective evidence of neurotoxicity (e.g., ptosis or inability to maintain upward gaze) should receive a test dose of edrophonium (if available) or neostigmine.
   a. Pretreat with atropine: 0.6 mg IV (children, 0.02 mg/kg with a minimum of 0.1 mg)
   b. Treat with:
      Edrophonium: 10 mg IV (children, 0.25 mg/kg)
      or
      Neostigmine: 0.02 mg/kg IV or IM (children, 0.04 mg/kg)
2. If objective improvement is evident after 30 min, treat with:
   a. Neostigmine: 0.5 mg IV, IM, or SC (children, 0.01 mg/kg) every 1 h as needed
   b. Atropine: 0.6 mg as IV continuous infusion over 8 h (children, 0.02 mg/kg over 8 h)
3. Closely monitor airway and perform endotracheal intubation if needed.

Wound care in the days after the bite should include careful aseptic debridement of clearly necrotic tissue once coagulopathy has been fully reversed. Intact serum-filled vesicles or hemorrhagic blebs should be left undisturbed. If ruptured, they should be debrided with sterile technique. Any debridement of damaged muscle should be conservative because there is evidence that such muscle may recover to a significant degree after antivenom therapy.

Physical therapy should be started as soon as possible to assist the patient in returning to a functional state. The incidence of long-term loss of function (e.g., reduced range of motion, impaired sensory function) is unclear.

Any patient with signs of envenomation should be observed in the hospital for at least 24 h. In North America, a patient with an apparently "dry" viperid bite should be closely monitored for at least 8–12 h before discharge, as significant toxicity occasionally develops after a delay of several hours. The onset of systemic symptoms commonly is delayed for several hours after bites by certain elapids (including coral snakes, *Micrurus* species), some non–North American viperids (e.g., the hump-nosed pit viper [*Hypnale hypnale*]), and sea snakes. Patients bitten by these snakes should be observed in the hospital for at least 24 h. Admission to an intensive care setting is advised for patients with progressive clinical findings despite initial antivenom administration; those bitten in the head, neck, or other high-risk sites; and those who develop an acute hypersensitivity reaction to antivenom.

At hospital discharge, victims of venomous snakebites should be warned about symptoms and signs of wound infection, antivenom-related serum sickness, and potential long-term sequelae, such as pituitary insufficiency from Russell's viper (*D. russelii*) bites. If coagulopathy developed in the acute stages of envenomation, it can recur during the first 2–3 weeks after the bite as venom antigens may have longer half-lives than their corresponding antivenoms. In such cases, victims should be warned to avoid elective surgery or activities posing a high risk of trauma during this period. Repeat laboratory tests (e.g., complete blood count, prothrombin time, fibrinogen level) should be scheduled to monitor for recurrent coagulopathy. Outpatient analgesic treatment, wound management, and physical therapy should be provided. For patients who present with delayed episodes of bleeding, severe isolated thrombocytopenia, or moderate thrombocytopenia with severe coagulopathy, repeat antivenom administration should be considered at typical maintenance dosing.

MORBIDITY AND MORTALITY

The overall mortality rates for victims of venomous snakebites are low in regions with rapid access to medical care and appropriate antivenoms. In the United States, for example, the mortality rate is <1% for victims who receive antivenom. Eastern and western diamondback rattlesnakes (*Crotalus adamanteus* and *Crotalus atrox*, respectively) are responsible for the majority of snakebite deaths in the United States. Snakes responsible for a large number of deaths in other countries include cobras (*Naja* species), carpet and saw-scaled vipers (*Echis* species), Russell's vipers (*D. russelii*), large African vipers (*Bitis* species), lancehead pit vipers (*Bothrops* species), and tropical rattlesnakes (*C. durissus*).

The incidence of morbidity, defined as permanent functional loss in a bitten extremity, is difficult to estimate but is substantial. Morbidity may be due to muscle, nerve, or vascular injury, or to scar contracture. Such morbidity can have devastating consequences for victims in the developing world when they lose the ability to work and provide for their families. In the United States, functional loss tends to be more common and severe after rattlesnake bites than after bites by copperheads (*Agkistrodon contortrix*) or water moccasins (*Agkistrodon piscivorus*).

GLOBAL CONSIDERATIONS

In May 2019, the WHO announced a comprehensive strategy to control and prevent snake envenomations worldwide, with the goal to reduce the number of deaths and cases of disability from snakebites by 50% by 2030. This strategy has been developed around four central tenets: empowering and engaging communities, ensuring safe and effective treatments, strengthening health systems, and improving partnerships, coordination, and resources. In many developing countries where snakebites are common, limited access to medical care and antivenoms contributes to high rates of morbidity and mortality. Recent data show that 6.85 billion people live within the regions inhabited by snakes, with >10% of people living >1 h from an urban center. Often, the available antivenoms are inappropriate and ineffective against the venoms of medically important indigenous snakes. In those regions, further research is necessary to determine the actual impact of venomous snakebites and the specific antivenoms needed in terms of both quantity and spectrum of coverage. Appropriate antivenoms must be available at the most likely first point of medical contact for patients (e.g., primary health centers) in order to minimize the common practice of referring victims to more distant, higher levels of care for the initiation of antivenom therapy. Just as important as getting the correct antivenoms into underserved regions is the need to educate populations about snakebite prevention and train medical care providers in proper management approaches. Local protocols written with significant input from experienced providers in the region of concern should be developed and distributed. Those who care for snakebite victims in these often-remote clinics must have the skills and confidence required to begin antivenom treatment (and to treat possible reactions) as soon as possible when indicated.

MARINE ENVENOMATIONS

The global incidence of marine envenomation is likely underestimated, as the majority of cases are mild. Much of the management of envenomation by marine animals is supportive, but a few specific antivenoms are available. This section provides general guidance, and patient management should be tailored to the practice patterns and known prevalence of specific venomous species in the region **(Fig. 471-4)**.

INVERTEBRATES

Cnidarians Cnidarians, such as hydroids, fire coral, jellyfish, and sea anemones, possess thousands of specialized stinging organelles called *cnidocysts* (a term that encompasses nematocysts, ptychocysts, and spirocysts) distributed along their tentacles. Nematocysts contain a barbed, coiled hollow thread bathed in venom that is discharged when provoked by mechanical stimuli, osmotic changes, or other chemical stimuli **(Fig. 471-5)**. The venom contains enzymes (phospholipases, metalloproteases), pore-forming toxins, neurotoxins, and nonprotein bioactive substances, such as tetramine, histamine, and serotonin. Venom flows through the hollow thread into the victim's skin. Because envenomation depends on intradermal delivery, only the subset of cnidarians whose barbs can pierce human skin are of toxicologic relevance.

After envenomation, immediate burning pain, pruritus, and paresthesias are common. The skin becomes reddened, darkened, edematous, and blistered, and may show signs of superficial necrosis. Rarely, systemic toxicity develops, affecting the cardiovascular, respiratory, gastrointestinal, and nervous systems, especially following stings from anemones, *Physalia* species, and scyphozoans. Anaphylaxis and secondary infection with marine bacteria can also occur.

Hundreds of deaths have been reported, many of them caused by *Chironex fleckeri, Stomolophus nomurai, Physalia physalis*, and *Chiropsalmus quadrumanus*. Irukandji syndrome is a potentially fatal condition associated with envenomation by the Australian jellyfish *Carukia barnesi* and *Malo* species. Symptoms generally manifest within 30 min and include hypertension; tachycardia; severe chest, abdominal, and back pain; nausea and vomiting; and headache. In the most serious cases, victims may develop cerebral hemorrhage, cerebral edema, cardiomyopathy, and pulmonary edema. Massive release of endogenous catecholamines induced by venom components appears to be the underlying cause of this syndrome, with catecholamine depletion occasionally resulting in delayed hypotension.

Initial management of *Cnidarian* envenomation should focus on removing the victim from the water and decontaminating the skin.

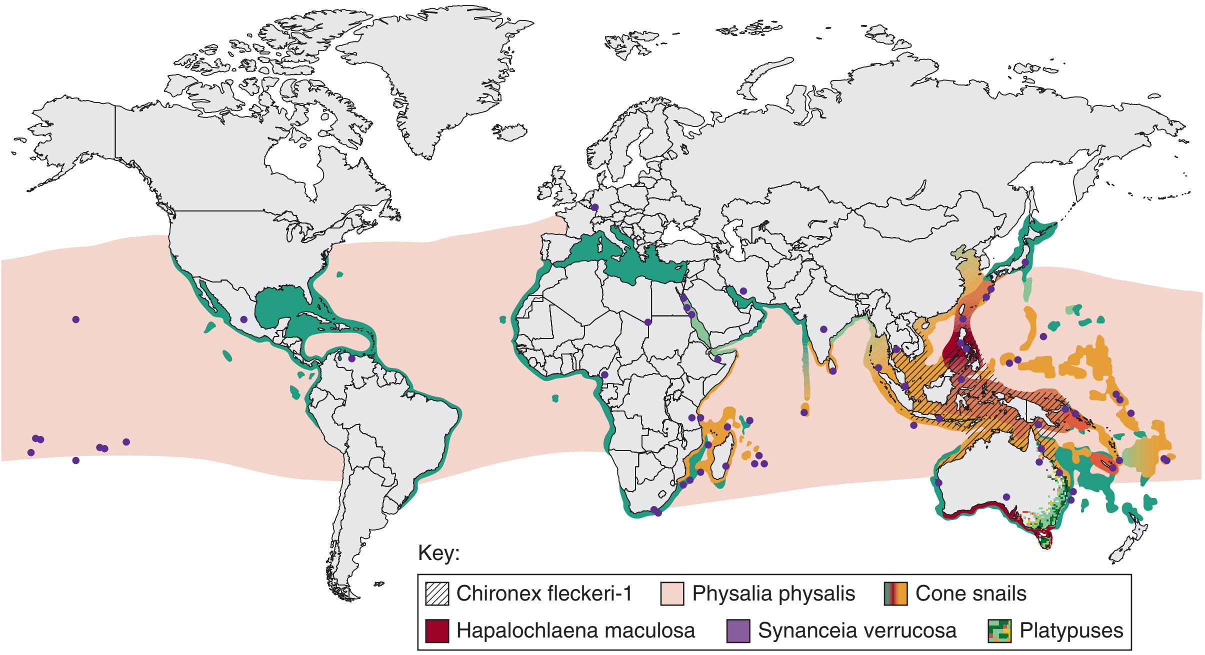

FIGURE 471-4 Geographic distribution of venomous marine animals.

Drowning is a common cause of death after significant *Cnidarian* envenomation. Once the victim is on land or a boat, the skin should be decontaminated with saline or seawater. Providers wearing protective equipment should carefully remove any adherent nematocysts. Vinegar (5% acetic acid) appears to be useful for relieving pain caused by several species, especially in the Indo-Pacific region where *C. fleckeri* and *C. barnesi* are common. In that region, decontamination with vinegar should be followed by hot water immersion up to 45°C (113°F). Vinegar may increase nematocyst discharge in *P. physalis* and *C. quinquecirrha*, species common in the United States. Victims in the United States should be decontaminated with seawater and then have affected areas immersed in hot water. Alternatively, vinegar may be tested on a small area of affected skin to assess effects before performing complete decontamination. If hot water is unavailable, commercial (chemical) cold packs or ice packs applied over a thin dry cloth or plastic membrane can be effective in alleviating pain, but do not denature venom components. Stingose, a commercial mixture of aluminum sulfate and surfactant sold in Australia, may denature venom components and provide rapid pain relief. In general, rubbing leads to further stinging by adherent cnidocysts and should be avoided.

After decontamination, topical application of a local anesthetic, antihistamine, or glucocorticoid may be helpful for symptom control. Persistent severe pain may be treated with opioid analgesics. Muscle spasms may respond to IV diazepam (2–5 mg, titrated upward as necessary). An antivenom is available from Seqirus (an affiliate of Commonwealth Serum Laboratories) for stings from *C. fleckeri* ("box jellyfish") found in Australian and Indo-Pacific waters. While initial in vitro studies demonstrated reversal of several *C. fleckeri* toxins, later animal studies were less promising, and there are no reports of human survival directly attributable to its administration. Use of the antivenom is recommended in the case of life-threatening toxicity when available, but this should not delay or replace aggressive supportive care (see "Sources of Antivenoms and Other Assistance," below). As expected, the antivenom appears less effective for envenomations from other *Cnidaria* species besides *C. fleckeri*. Adverse reactions to the antivenom include early (anaphylaxis) and late (serum sickness) hypersensitivity reactions. Acute allergic reactions should be treated with systemic antihistamines, glucocorticoids, and epinephrine when appropriate. Delayed serum sickness is relatively common, typically occurs 5–10 days after antivenom administration, and generally responds well to 5 days of oral glucocorticoids.

Treatment of Irukandji syndrome may require administration of opioid analgesics and aggressive treatment of hypertension. Appropriate antihypertensive agents include phentolamine, nicardipine, nitroprusside, nitroglycerin, and IV magnesium sulfate. Beta-adrenergic antagonists should be avoided due to the risk of worsening hypertension from unopposed alpha-adrenergic effects. All victims with systemic reactions should be observed for at least 6–8 h for rebound from any therapy and should be monitored for cardiac arrhythmias.

Safe Sea (Nidaria Technology Ltd.), a "jellyfish-safe" sunblock applied to the skin before an individual enters the water, inactivates the recognition and discharge mechanisms of nematocysts; it may prevent or diminish the effects of coelenterate stings. Whenever possible, a dive skin or wetsuit should be worn when entering ocean waters.

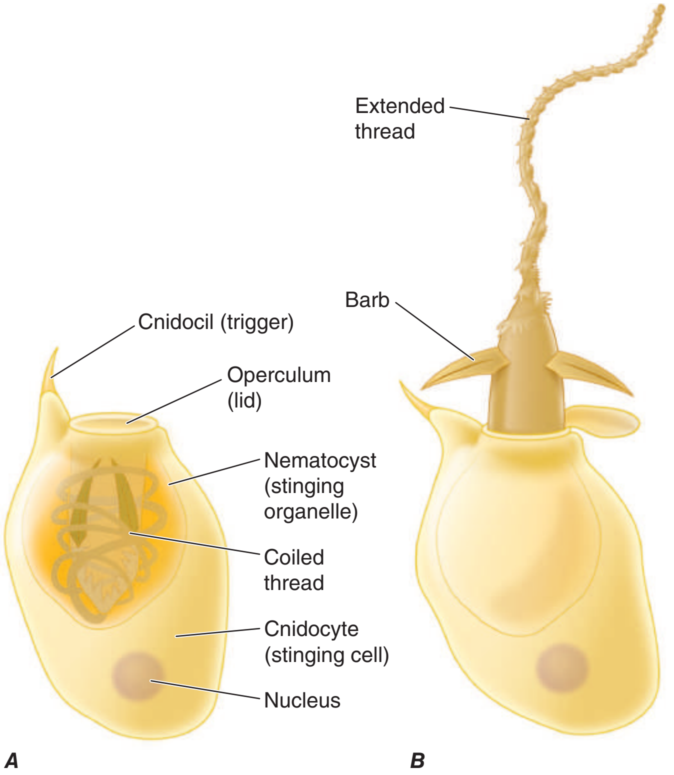

FIGURE 471-5 Schematic of a nematocyst. ***A.*** Undischarged. Note coiled hollow thread bathed in venom. ***B.*** Discharged.

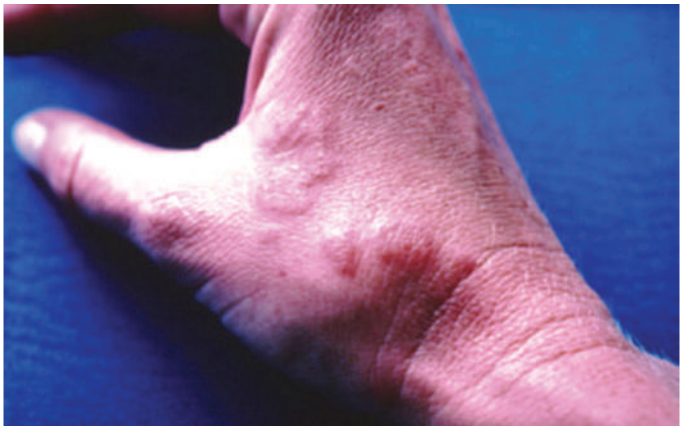

FIGURE 471-6 Rash on the hand of a diver from the spines of a bristleworm. *(Courtesy of Paul Auerbach, with permission.)*

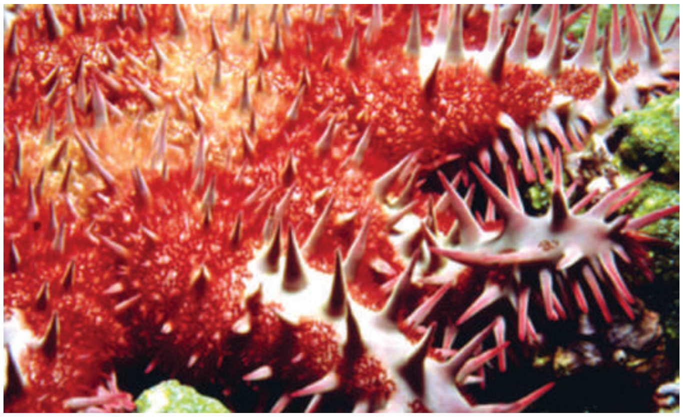

FIGURE 471-7 Spines on the crown-of-thorns sea star (*Acanthaster planci*). *(Courtesy of Paul Auerbach, with permission.)*

Sea Sponges Several species of sea sponges (Phylum *Porifera*) produce irritants known as crinitoxins. Touching a sea sponge may result in allergic contact dermatitis. Irritant dermatitis may result if the sponge's spicules of silica or calcium carbonate penetrate the skin. Affected skin should be dried and adhesive tape, a commercial facial peel, or a thin layer of rubber cement used to remove embedded spicules. Vinegar should then be applied immediately, with repeated application for 10–30 min three or four times a day thereafter. Systemic corticosteroids should be reserved for severe allergic reactions (such as severe vesiculation) or erythema multiforme. Mild reactions generally subside within 7 days. Severe envenomation can lead to fevers, chills, and muscle spasms. Desquamation of affected skin has also been described and can occur up to 2 months after exposure. Outpatient wound care is essential to monitor for secondary infection.

Annelid Worms Annelid worms (bristle worms) are covered with chitinous spines capable of penetrating human skin and producing dermal symptoms similar to those of *Cnidarian* envenomation, including pain, prickling, urticaria, and discoloration. Pain usually subsides within hours, but urticaria can persist for days and discoloration for weeks **(Fig. 471-6)**. Victims should be instructed not to scratch as it may fracture the spines, complicating their removal and increasing the risk of secondary infection. Visible bristles should be removed with forceps, adhesive tape, rubber cement, or a commercial facial peel. Soaking the affected skin in vinegar may provide additional relief. Severe inflammation can be treated with systemic antihistamines or corticosteroids.

Sea Urchins Venomous sea urchins possess hollow, venom-filled, calcified spines and pincer-like, flexible pedicellariae with venom glands. The venom contains bradykinin-like substances, steroid glycosides, hemolysins, proteases, serotonin, and cholinergic substances. Envenomation causes immediate onset of severe pain, edema, and erythema. If multiple spines penetrate the skin, the patient may develop systemic symptoms, including nausea, vomiting, numbness, muscular paralysis, and respiratory distress. Synovitis and arthritis have been reported after joint-space penetration. Retained spines can cause the formation of painful granulomas.

The affected part should be immersed in hot water (see "Cnidarians," above). Embedded spines should be removed with care as they may fracture and leave remnants lodged in the victim. Soft-tissue radiography, ultrasonography, or magnetic resonance imaging (MRI) can be used to evaluate for the presence of retained fragments. Surgical removal may be necessary, especially if the spines are in close proximity to vital structures (e.g., joints, neurovascular bundles). Granulomas from retained spines are amenable to excision or intralesional injection with triamcinolone hexacetonide. Arthritis from joint penetration has been treated with synovectomy.

Starfish The crown-of-thorns starfish (*Acanthaster planci*) produces viscous venom that coats the surface of its spines **(Fig. 471-7)**. The venom contains saponins with hemolytic, myotoxic, hepatotoxic, and anticoagulant properties. Skin puncture causes immediate pain, bleeding, and local edema. Multiple punctures may result in reactions such as local muscle paralysis. The spines fracture easily, and retained fragments may cause granulomatous lesions and synovitis. Envenomated persons benefit from immediate immersion in hot water, local anesthesia, wound cleansing, imaging, and possible exploration to remove spines and foreign material. The hemolytic and hepatotoxic components are less heat labile than other venom constituents; hot water immersion may not prevent systemic toxicity.

Sea Cucumbers Sea cucumbers excrete holothurin (a cantharidin-like liquid toxin) from their anus. This toxin is then concentrated in the tentacular organs that are projected when the animal is threatened. Underwater, holothurin induces minimal contact dermatitis but can cause significant corneal and conjunctival irritation should ocular contact occur. A severe reaction can lead to blindness. Skin should be decontaminated with vinegar. The eyes should be anesthetized with 1–2 drops of 0.5% proparacaine and irrigated copiously with normal saline, with subsequent slit-lamp examination to identify corneal defects.

Cone Snails Cone snails possess a detachable dartlike tooth bathed in venom for hunting prey and self-defense. Numerous conotoxins have been identified, with two major common mechanisms: neuromuscular acetylcholine receptor antagonism and blockade of sodium, potassium, and calcium channels. Punctures result in small, painful wounds followed by local ischemia, cyanosis, and numbness. Syncope, dysphagia, dysarthria, ptosis, blurred vision, and pruritus also have been documented. Cardiac dysrhythmias and paralysis with resultant respiratory failure have also been reported. Pressure-immobilization (see "Octopuses," below), hot-water soaks, and local anesthetics have been successful for localized symptoms. Respiratory failure may necessitate mechanical ventilation. Cardiac dysrhythmias should be treated with electrical cardioversion as ion channel-blocking antidysrhythmic medications may worsen the dysrhythmia. No antivenom is available. Edrophonium has been recommended as therapy for paralysis if an edrophonium test is positive (see Table 471-2).

Octopuses Serious envenomations and deaths have followed bites of Australian blue-ringed octopuses (*Hapalochlaena maculosa* and *Hapalochlaena lunulata*). The classic blue rings appear only when the animal is threatened. These species are not aggressive, and human envenomations have typically occurred when handling the octopus. Their salivary glands contain symbiotic bacteria that produce tetrodotoxin, a potent sodium channel blocker that inhibits transmissions in the peripheral nervous system. The bite is minimally painful, but oral and facial numbness develop within minutes of a serious envenomation. Mild weakness can rapidly progress to total flaccid paralysis. Mentation is typically unaffected. The venom can also cause peripheral vasodilation leading to profound hypotension. Deaths from these envenomations are due to respiratory failure or vasodilatory shock and hypoperfusion.

Immediately after envenomation, a wide circumferential pressure-immobilization bandage should be applied over a gauze pad placed directly over the sting. The dressing should be applied at venous-lymphatic pressure with the preservation of distal arterial pulses, and the limb should be splinted. The bandage can be released when the victim has been transported to a medical facility. There is no antidote, and treatment is supportive. Respiratory failure may necessitate mechanical ventilation. Appropriate analgesia and sedation are critical as mental status is generally unaffected even in cases of complete paralysis. Hypotension should be treated with crystalloid and vasopressors if needed. In animal studies, phenylephrine and norepinephrine were more effective than dopamine or epinephrine for treatment of vasodilatory shock. Recovery usually occurs within 24–48 h, and long-term sequelae are uncommon unless related to hypoxia or hypoperfusion. Tetrodotoxin is also found in the flesh of fish in the order Tetraodontiformes, which contains a number of "pufferfish" including blowfish, globefish, and balloonfish. The toxin can be absorbed orally when these fish are consumed. Fugu, a traditional Japanese delicacy, is a reported source of poisoning. When properly prepared by a licensed chef, fugu is meant to contain a sufficient dose of tetrodotoxin to cause mild perioral paresthesia without systemic toxicity. When larger doses of the toxin are consumed due to improper preparation, symptoms are similar to envenomation by the blue-ringed octopus.

■ VERTEBRATES

Stingrays A stingray injury is both an envenomation and a traumatic wound. Stingrays possess serrated spines with venom glands that can easily penetrate human skin. The venom contains serotonin, 5′-nucleotidase, and phosphodiesterase. Victims experience severe pain, bleeding, and edema at the site of injury that peak within 30–60 min and may persist for up to 2 days. Penetrating injuries to the thorax and heart as well as lacerations to major vessels (especially of the lower extremity) have been reported. The wound often becomes ischemic in appearance and heals poorly, with adjacent soft-tissue swelling and prolonged disability. Systemic effects of the venom include weakness, diaphoresis, nausea, vomiting, diarrhea, hypotension, dysrhythmias, syncope, seizures, muscle cramps, fasciculations, and paralysis. While the venom effects can be fatal in rare cases, most deaths are attributable to traumatic injury.

Stonefish Stonefish (*Synanceia* species) are members of the *Scorpaenidae* family and generally considered the most venomous bony fish in the world. Their venom contains pore-forming toxins, proteases, hyaluronidase, 5′-nucleotidase, acetylcholinesterase, and cardiac calcium channel blockers. The venom is delivered through 12 or 13 dorsal, 2 pelvic, and 3 anal spines when provoked by mechanical stimuli. Victims experience immediate, intense pain that peaks within 90 min, local edema, and wound cyanosis. Local symptoms most often resolve within 12 h but can persist for days. Signs of systemic toxicity include abdominal pain, vomiting, delirium, seizures, paralysis, respiratory distress, dysrhythmias, and congestive heart failure. An antivenom from Seqirus (see "Sources of Antivenoms and Other Assistance," below) can be used in cases of severe envenomation but should not replace supportive care.

Lionfish Also members of the family Scorpaenidae, lionfish (*Pterois* species) are much less toxic to humans than stonefish. The venom is delivered by curved dorsal spines and contains heat-labile, high-molecular-weight proteins. Reported symptoms include localized pain, blistering, edema, sensory changes (paresthesia, anesthesia, or hyperesthesia), and necrosis (rare).

Platypuses The platypus is a venomous mammal. The male has a keratinous spur on each hind limb that is connected to a venom gland within the upper thigh. Skin puncture causes soft-tissue edema and pain that may last for days to weeks. Care is supportive and should focus on appropriate analgesia and wound care. Hot water immersion does not appear to be beneficial.

TREATMENT

Marine Vertebrate Stings

The stings of all marine vertebrates are treated in a similar fashion. Antivenom is currently available for stonefish envenomation. The affected part should be immersed immediately in hot water up to 45°C (113°F) for 30–90 min or until there is significant pain relief. Recurrent pain may respond to repeated hot water immersion. Systemic opioids as well as wound infiltration or regional nerve block with local anesthetics can help alleviate pain and facilitate wound exploration and debridement. Advanced imaging (in particular, ultrasound or MRI) may be helpful in identification of retained foreign bodies, which may be missed on plain radiographs. After exploration and debridement, the wound should be irrigated vigorously with warm, sterile saline. Bleeding is generally controlled with local pressure. Most wounds should be left open to heal by secondary intention or treated by delayed primary closure. Early primary repair is occasionally preferred for cosmetic reasons but increases the risk of wound infection. Tetanus immunization should be provided as appropriate. Antibiotic treatment should be considered for serious wounds, immunocompromised hosts, and those undergoing hot water immersion, as the elevated temperature can increase bacterial growth. The initial antibiotics should cover *Staphylococcus* and *Streptococcus* species. If the victim is immunocompromised, if a wound is primarily repaired, or if an infection develops, antibiotic coverage should be broadened to include *Vibrio* species for wounds sustained in salt water or *Aeromonas* species for wounds sustained in freshwater.

APPROACH TO THE PATIENT

Marine Envenomations

It is useful to be familiar with the local marine fauna and to recognize patterns of injury.

Coelenterate (marine invertebrate) stings sometimes create diagnostic skin patterns. A diffuse urticarial rash on exposed skin is often indicative of exposure to fragmented hydroids or larval anemones. A linear, whiplike print pattern appears where a jellyfish tentacle has contacted the skin. In the case of the box jellyfish, a crosshatched appearance, followed by development of dark purple coloration within a few hours of the sting, heralds skin necrosis. An encounter with fire coral causes immediate pain, erythema, and swelling in the pattern of contact, similar to but more severe than the imprint left by exposure to an intact feather hydroid. Seabather's eruption, caused by thimble jellyfish and larval anemones, is a diffuse, intensely pruritic rash consisting of clusters of erythematous macules or papules that follow the pattern of bathing attire **(Fig. 471-8)**. Toxic sponges create a burning and painful red rash on exposed skin, which may blister and later desquamate. Virtually all marine stingers cause cutaneous inflammation; thus, local erythema, swelling, and adenopathy are fairly nonspecific.

A large puncture wound or jagged laceration (particularly on the lower extremity) that is more painful than one would expect from the size and configuration of the wound is likely to be a stingray envenomation. Smaller puncture wounds, sometimes associated with purple or dark discoloration, represent the activity of a sea urchin or starfish. Stony corals cause rough abrasions and, in rare instances, lacerations or puncture wounds.

■ SOURCES OF ANTIVENOMS AND OTHER ASSISTANCE

In the United States, assistance in locating a specific antivenom can be obtained from a regional poison control center (800-222-1222). Divers Alert Network, a nonprofit organization designed to assist in the care of injured divers, also may help with the treatment of marine injuries. The network can be reached at *www.diversalertnetwork.org* or by telephone 24 h a day at 919-684-9111. The antivenoms for box jellyfish

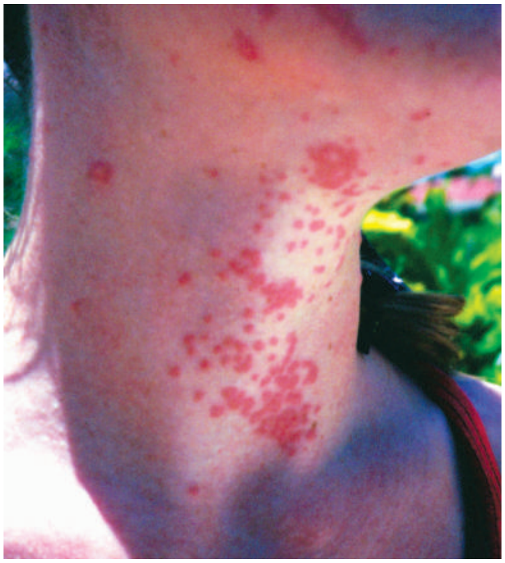

FIGURE 471-8 Erythematous, papular rash typical of seabather's eruption caused by thimble jellyfish and larval anemones. *(Courtesy of Paul Auerbach, with permission.)*

(*C. fleckeri*) and stonefish or severe scorpionfish envenomations are made in Australia by Seqirus (63 Poplar Road, Parkville, Victoria, Australia 3052; *www.seqirus.com.au*; 61-3-9389-2000). When administering the box jellyfish antivenom, time is of the essence. For cardiac or respiratory decompensation, a minimum of 1 ampule and up to 6 ampules consecutively should be given IV, preferably in a 1:10 dilution with normal saline. For stonefish or severe scorpionfish envenomation, 1 ampule of specific antivenom should be administered IM for every one or two punctures, to a maximum of 3 ampules.

MARINE POISONINGS

HISTAMINE (SCOMBROID) FISH POISONING

Histamine fish poisoning, most often referred to as scombroid or pseudoallergenic fish poisoning, may be the most common type of seafood poisoning worldwide. It was originally described after consumption of scombroid (mackerel-like) fish (including albacore, bluefin, and yellowfin tuna; mackerel; saury; needlefish; wahoo; skipjack; and bonito) but is now more commonly reported with consumption of nonscombroid fish (including dolphinfish, kahawai, sardine, black marlin, pilchard, anchovy, herring, amberjack, Australian ocean salmon, and bluefish).

Under conditions of inadequate preservation or refrigeration, the amino acid L-histidine in fish muscle undergoes decarboxylation to histamine, histamine phosphate, and histamine hydrochloride by *Morganella morganii*, *Escherichia coli*, *Proteus* species, and *Klebsiella* species. Histamine levels of 50–100 mg/100 g are noted in toxic fish, with levels >400 mg/100 g on occasion. Toxic levels can be reached within 12 h of inadequate refrigeration. The pathophysiology of this intoxication remains unclear, as large doses of oral histamine do not reproduce the condition. It has been proposed that other biogenic amines such as cadaverine and putrescine may inhibit the metabolism of histamine. Another potential mechanism is the induction of mast cell degranulation by an unidentified toxin. However, affected individuals may have normal levels of mast cell–derived prostaglandins, which argues against this mechanism.

Once formed, the toxin or toxins involved are heat stable and are not destroyed by cooking or freezing. Affected fish may have a sharply metallic or peppery taste, although more often they are normal in appearance, color, and flavor. Not all persons who eat a contaminated fish necessarily become ill, perhaps because of uneven distribution of decay within the fish.

Symptoms develop within 15–90 min of ingestion. Most cases are mild, with tingling of the lips and mouth, mild abdominal discomfort, and nausea. The more severe and commonly described presentation includes intense, sharply demarcated flushing of the face, neck, and upper trunk, pruritus, urticaria, and angioedema. This syndrome may progress to bronchospasm, nausea, vomiting, diarrhea, epigastric pain, abdominal cramps, dysphagia, headache, palpitations, tachycardia, dizziness, hypotension, and cardiogenic shock. Without treatment, the symptoms generally resolve within 8–12 h. Because of blockade of gastrointestinal tract histaminase, the reaction may be more severe in individuals taking isoniazid.

TREATMENT

Scombroid Poisoning

Therapy is directed at reversing the histamine effect with systemic antihistamines. In case literature, H2 receptor antagonists (e.g., cimetidine, ranitidine) appear to further decrease the severity and duration of illness when added to H1 receptor antagonists (e.g., diphenhydramine, hydroxyzine). If bronchospasm is severe, an inhaled bronchodilator (e.g., albuterol) can be used. In rare cases, parenteral epinephrine may be needed. The use of activated charcoal is not recommended. Protracted nausea and vomiting may be controlled with antiemetics (e.g., ondansetron, prochlorperazine). Hypotension should be treated with IV fluids. If multiple patients present from the same meal with similar symptoms, or if the fish itself can be tested for histamine content, the diagnosis can be essentially confirmed, and the patient can be informed this was not an allergic reaction.

CIGUATERA

Epidemiology and Pathogenesis Ciguatera poisoning is the most common nonbacterial food poisoning associated with fish in the United States and accounts for approximately half of all cases. Florida and Hawaii account for 90% of reported U.S. cases, although with transportation of imported fish worldwide, all clinicians need to be aware of ciguatera. The poisoning almost exclusively involves carnivorous, bottom-dwelling reef fish native to the Indian Ocean, the South Pacific, and the Caribbean Sea. More than 500 different fish species have been implicated in ciguatera poisoning, but the most common are barracuda, snapper, moray eel, grouper, sea bass, and Spanish mackerel. Global estimates of incidence vary widely from 20,000–500,000 cases per year; it is suspected that a large majority of cases go unreported. Ciguatoxins are produced primarily by the bottom-dwelling, photosynthetic marine dinoflagellate *Gambierdiscus toxicus*. These lipophilic toxins bioaccumulate in the marine food chain when large carnivorous fish consume the grazing fish that feed on these dinoflagellates. Ciguatoxins are heat stable and unaffected by freezing, drying, cooking, or gastric acid. The toxins do not affect the odor, taste, or appearance of the fish, making identification and prevention difficult. Ciguatoxins are potent activators of neuronal sodium channels but may also have other effects such as antagonism of voltage-gated potassium channels. Toxins are found in the highest concentrations in the fish's skin, head, and viscera; therefore, consumption of these portions should be avoided.

Clinical Manifestations Symptoms can develop within 15–30 min of ingestion but more commonly in 2–6 h. Most victims develop symptoms within 12 h of ingestion, and virtually all are afflicted within 24 h. Numerous ciguatoxins have been identified, and their relative abundance in different species of fish and geographic regions likely explains the wide array of reported symptoms **(Table 471-3)**. Early symptoms include nausea, vomiting, diarrhea, abdominal cramps, headache, and diaphoresis. Neurologic manifestations include vertigo, dysesthesia, paresthesia, visual disturbance, dysgeusia, and reversal of hot and cold temperature discrimination. Some victims describe a sensation of loose teeth. Bradycardia, hypotension, and orthostasis have also been reported. Gastrointestinal symptoms generally resolve after 24–48 h, but neurologic manifestations may persist for days to weeks. More severe reactions tend to occur on repeat exposure. Persons who have ingested parrotfish (scaritoxin) may develop classic ciguatera poisoning as well as a "second-phase" syndrome (after a delay of 5–10 days)

TABLE 471-3 Representative Symptoms and Signs of Ciguatera Poisoning

SYSTEM	SYMPTOMS/SIGNS
Gastrointestinal	Abdominal pain, nausea, vomiting, diarrhea
Neurologic	Paresthesias, pruritus, tongue and throat numbness or burning, sensation of "carbonation" during swallowing, odontalgia or dental dysesthesias, dysphagia, tremor, fasciculations, athetosis, meningismus, aphonia, ataxia, vertigo, pain and weakness in the lower extremities, visual blurring, transient blindness, hyporeflexia, seizures, coma
Dermatologic	Conjunctivitis, maculopapular rash, skin vesiculations, dermographism
Cardiovascular	Bradycardia, heart block, hypotension, central respiratory failure[a]
Other	Chills, dysuria, dyspnea, dyspareunia, fatigue, nasal congestion and dryness, insomnia, hypersalivation, diaphoresis, headache, arthralgias, myalgias

[a]Tachycardia and hypertension may occur after potentially severe transient bradycardia and hypotension. Death is rare.

of disequilibrium with ataxia, dysmetria, and resting or kinetic tremor. This syndrome may persist for 2–6 weeks.

Diagnosis Ciguatera poisoning is a clinical diagnosis. The toxin can be detected by liquid chromatography and tandem mass spectrometry, and fish suspected of contamination can be tested using a ciguatoxin-specific enzyme immunoassay, but these techniques are generally not available in most health care institutions. The differential diagnosis of ciguatera includes paralytic shellfish poisoning, eosinophilic meningitis, type E botulism, organophosphate insecticide poisoning, tetrodotoxin poisoning, and psychogenic hyperventilation.

TREATMENT

Ciguatera Poisoning

Therapy is supportive and based on symptoms. Volume losses from vomiting and diarrhea should be treated with crystalloids and electrolyte repletion. Hypotension may rarely be unresponsive to fluids and require vasopressors. Symptomatic bradyarrhythmias generally respond well to atropine (0.5 mg IV, up to 2 mg). Problematic orthostasis can be treated with direct alpha-adrenergic agonists (e.g., phenylephrine). IV infusion of mannitol may be beneficial in moderate or severe cases in fluid-replete patients; however, the efficacy of this therapy has not been definitively proven. An initial IV dose of mannitol at 1 g/kg may be given over 45–60 min. If symptoms are alleviated, a second dose may be given within 3–4 h and a third dose the next day. Care must be taken to avoid dehydration. The mechanism of the drug's benefit against ciguatera poisoning is unclear but may be due to decreased sodium conductance across neuronal cell membranes. Hyperosmotic water-drawing action is another proposed mechanism but no changes in neuronal cell edema have been observed in vitro. Amitriptyline (25 mg orally twice a day) reportedly alleviates pruritus and dysesthesias and may decrease rates of subsequent development of chronic nerve symptoms. Gabapentin and pregabalin have shown some efficacy in the treatment of long-term nerve pain in case reports, but evidence from controlled trials is lacking.

During recovery from ciguatera poisoning, the victim should exclude the following from the diet for 6 months: fish (fresh or preserved), fish sauces, shellfish, shellfish sauces, alcoholic beverages, nuts, and nut oils. Consumption of fish in ciguatera-endemic regions should be avoided.

PARALYTIC SHELLFISH POISONING

Paralytic shellfish poisoning is induced by ingestion of filter-feeding organisms, including clams, oysters, scallops, mussels, chitons, limpets, starfish, and sand crabs. The most common toxin is saxitoxin, produced by dinoflagellates in the genera *Alexandrium*, *Gonyaulax*, and *Pyrodinium*. These unicellular phytoplankton form the foundation of the food chain for many filter-feeding organisms and the toxin accumulates in their tissues. In the United States, paralytic shellfish poisoning is acquired primarily from seafood harvested in the Northeast, the Pacific Northwest, and Alaska. During algal blooms ("red tides") in the summer months, these planktonic species can release massive amounts of toxic metabolites into the water and cause mortality in bird and marine populations. The paralytic shellfish toxins are water soluble as well as heat and acid stable; ordinary cooking or freezing does not destroy them. Contaminated seafood looks, smells, and tastes normal. Saxitoxin appears to block sodium conductance, inhibiting neuromuscular transmission at the axonal and muscle membrane levels. A toxin concentration of >75 μg/100 g of foodstuff is considered hazardous to humans. During an algal bloom, the concentration of saxitoxin in shellfish can exceed 9000 μg/100 g. A mouse bioassay that identifies saxitoxin in suspected shellfish is currently in use. Saxitoxin can be detected in body fluids by high-performance liquid chromatography, but this method is generally not available in the clinical setting.

Intraoral and perioral paresthesia can occur within minutes to a few hours after ingestion of contaminated shellfish and can progress rapidly to involve the neck and distal extremities. Other neurologic symptoms can include headache, vertigo, ataxia, diffuse muscle weakness, hyperreflexia, and cranial neuropathies such as dysarthria, dysphagia, dysphonia, and transient vision loss. Gastrointestinal symptoms can include nausea, vomiting, diarrhea, and abdominal pain. Flaccid paralysis and respiratory insufficiency may follow 2–12 h after ingestion. In the absence of hypoxia, the victim often remains alert but paralyzed. Mortality rates in paralytic shellfish poisoning approach 12%.

TREATMENT

Paralytic Shellfish Poisoning

Treatment is supportive and based on symptoms. If the victim seeks medical attention within the first few hours after ingestion, activated charcoal (50–100 g) can be administered in the absence of vomiting. Gastric lavage and cathartics have been attempted but there is no evidence of benefit, and most authors recommend against their use.

The most serious concern is respiratory paralysis. The victim should be closely observed for respiratory distress for at least 24 h in a hospital. With prompt recognition of respiratory failure and establishment of ventilatory support, anoxic myocardial and brain injury may be prevented. If the patient survives for 18 h, the prognosis is good for a complete recovery.

AMNESIC SHELLFISH POISONING

Amnesic shellfish poisoning occurs when humans consume shellfish containing domoic acid. Marine diatoms of the genera *Nitzschia* and *Pseudonitzchia* produce the toxin, which can bioaccumulate in filter feeders during algal blooms. Clams, mussels, oysters, anchovies, and Dungeness crabs have all been found to cause amnesic shellfish poisoning. Domoic acid is an excitotoxic amino acid capable of binding to kainate and AMPA-type glutamate receptors in the central nervous system. Uncontrolled calcium influx into neurons stimulated by domoic acid binding causes neurodegeneration and apoptosis. The toxin is heat stable and is not affected by cooking or freezing. Shellfish can be tested for domoic acid by mouse bioassay and high-performance liquid chromatography (HPLC). The regulatory limit for domoic acid in shellfish is 20 parts per million. An enzyme-linked immunoassay has been developed to detect domoic acid in human body fluids but is generally not available in clinical laboratories.

Most victims will report symptoms within 5 h of ingesting contaminated shellfish, but delayed onset of up to 40 h has been reported. Symptoms include nausea, vomiting, diarrhea, abdominal cramps, and a variety of neurologic manifestations, such as severe headache, memory loss, seizures, hemiparesis, ophthalmoplegia, grimacing, purposeless chewing, agitation, emotional lability, and coma. Cardiac dysrhythmias, hypotension, and pulmonary edema have also been reported. Postmortem examination of brain tissue has shown neuronal necrosis or cell loss and astrocytosis, most prominently in the hippocampus and amygdala. Several months after the primary intoxication, victims may still display chronic residual memory deficits and motor or sensory neuropathy.

TREATMENT

Amnesic Shellfish Poisoning

Therapy is supportive and based on symptoms. IV fluids and antiemetics may be used for severe nausea, vomiting, and diarrhea. Domoic acid neurotoxicity is primarily seizure mediated; anticonvulsive therapy using GABA agonists (e.g., benzodiazepines, propofol, or barbiturates) should be instituted early. However, some patients without clinically evident seizure activity have developed neurologic sequelae.

DIARRHETIC SHELLFISH POISONING

Diarrhetic shellfish poisoning occurs with consumption of shellfish containing the lipophilic compound okadaic acid. This toxin inhibits serine and threonine protein phosphatases, with consequent protein accumulation and continued secretion of fluid by intestinal cells leading to diarrhea. Shellfish acquire these toxins by feeding on dinoflagellates, particularly of the genera *Dinophysis* and *Prorocentrum*.

Symptoms include diarrhea, nausea, vomiting, abdominal pain, and chills. Onset typically occurs between 30 min and 12 h after ingestion of contaminated shellfish. The illness is usually self-limited; most patients recover in 3–4 days and only a few require hospitalization. Treatment is supportive and focused on hydration. Toxins can be detected in food samples by a mouse bioassay, an immunoassay, and fluorometric HPLC.

Acknowledgment

Kirsten B. Hornbeak, Robert L. Norris, Alex Chen, and Charles Lei contributed to this chapter in the prior edition and material from that chapter has been retained here. We would like to dedicate this chapter to the late Dr. Paul S. Auerbach, who was a contributing author for the previous seven editions of Harrison's Principles of Internal Medicine. Dr. Auerbach had a tremendous impact on the field of emergency medicine and founded the subspecialty of wilderness medicine. Dr. Auerbach was a wonderful teacher, mentor, and friend, and will be deeply missed.

FURTHER READING

Blohm E et al: Marine envenomations, in *Goldfrank's Toxicologic Emergencies*, 11th ed. Nelson LS et al (eds). New York, McGraw-Hill Education, 2019, pp 1567-1580.

Bush SP et al: Comparison of F(ab')2 versus Fab antivenom for pit viper envenomation: A prospective, blinded, multicenter, randomized clinical trial. Clin Toxicol 53:37, 2015.

Cannon R et al: Acute hypersensitivity reactions associated with administration of crotalidae polyvalent immune Fab antivenom. Ann Emerg Med 51:407, 2008.

Fil LJ et al: Food Poisoning, in *Goldfrank's Toxicologic Emergencies*, 11th ed. Nelson LS et al (eds). New York, McGraw-Hill Education, 2019, pp 592-605.

French LK et al: Marine vertebrates, cnidarians, and mollusks, in *Critical Care Toxicology: diagnosis and management of the critically poisoned patient*, 2nd ed. Brent J et al (eds). New York, Springer, 2017, pp 2045-2074.

GBD 2019 Diseases and Injuries Collaborators: Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 396:1204, 2020. Erratum in: Lancet 396:1562, 2020.

Hornbeak KB, Auerbach PS: Marine envenomation. Emerg Med Clin North Am 35:321, 2017.

Kang AM, Fisher ES: Thromboelastography with platelet studies (TEG® with PlateletMapping®) after rattlesnake envenomation in the southwestern United States demonstrates inhibition of ADP-induced platelet activation as well as clot lysis. J Med Toxicol 16:24, 2020.

Pottier I et al: Ciguatera fish poisoning in the Caribbean Sea and Atlantic Ocean: Reconciling the multiplicity of ciguatoxins and analytical chemistry approach for public health safety. Toxins (Basel) 15:453, 2023.

Thomas EG, Thomas DJ: Mimics of allergy and angioedema: Scombroid, mast cell activation disorders, and hereditary alpha tryptasemia. Immunol Allergy Clin North Am 43:553, 2023.

Ubani CB et al: Emergency department management of North American snake envenomations. Emerg Med Pract 26:1, 2024.

Warrell DA, Williams DJ: Clinical aspects of snakebite envenoming and its treatment in low-resource settings. Lancet 401:1382, 2023.

472

Ectoparasite Infestations and Arthropod Injuries

Silas A. Davidson, Scott A. Norton

Ectoparasites include arthropods and creatures from other phyla that infest the skin or hair of animals; the host animals provide them with sustenance and shelter. The ectoparasites may remain on the superficial surfaces of the skin or hair, attached by mouthparts or with specialized claws. Other ectoparasites may penetrate the skin's surface and reside in the epidermis, dermis, or subcutaneous tissues. Ectoparasites may inflict direct mechanical injury, consume blood or nutrients, induce hypersensitivity reactions, inoculate toxins, transmit pathogens, create openings in the skin for secondary bacterial infection, and incite fear or disgust. Human beings are obligate hosts for few ectoparasites but can serve as facultative, dead-end, or paratenic (accidental) hosts for many others. Of the organisms discussed in this chapter, only scabies mites (the *hominis* variety) and human-infesting lice are obligate parasites of humans.

Arthropods that are capable of ectoparasitism or that can otherwise cause injury include insects (such as lice, fleas, bed bugs, wasps, ants, bees, and diverse kinds of flies), arachnids (spiders, scorpions, mites, and ticks), and myriapods (millipedes and centipedes). Several arthropods can cause uncomfortable reactions when they, their setae (hairlike growths), or their exudates come into contact with skin, mucous membranes, or ocular tissues.

Certain nematodes (helminths), such as the hookworms **(Chap. 239)**, are ectoparasitic in that they must enter the skin in some manner, then traverse the skin to reach deeper tissues, or migrate within the skin. Infrequently encountered ectoparasites in other animal phyla include pentastomes (armillifers or tongue worms) and leeches.

Arthropods may cause injury when they attempt to take a blood meal or as they defend themselves by biting, stinging, or exuding venoms. Papular urticaria and other lesions caused by arthropod bites and stings are so diverse and variable (depending on the host's health status and prior exposure to the arthropod's saliva, venom, or other exudates) that it is difficult to identify the precise causative organism with visual examination alone without a bona fide specimen and taxonomic expertise. Specimens of the presumably offending arthropod should, whenever possible, be sampled directly (taken from the patient, ideally by medical personnel) or indirectly by using traps or other monitoring devices in the patient's home or workplace. Samples sent to laboratories for identification should be properly fixed, preserved, and packaged. The patient's history of travel and precautionary behaviors; occupational, recreational, and environmental exposures; and proximity to animals often helps the clinician and parasitologist resolve the cause.

SCABIES

The human itch mite, *Sarcoptes scabiei* var. *hominis*, is an obligate human ectoparasite and a common cause of itchy dermatosis, affecting ~250 million persons worldwide. Both the mites and the skin condition are called scabies. Gravid female mites (~0.3 mm in length) burrow superficially within the stratum corneum, depositing several eggs per day. Gravid adult females emerge to the skin's surface about 8 days later, then (re)invade the skin of the same person or another host. Newly fertilized female mites are transferred between people mainly by direct skin-to-skin contact. Transmission is facilitated by crowding, poor hygiene, and close physical contact with other people. Generally, scabies mites die within a day or so if not on a suitable host. Transmission by sharing contaminated bedding or clothing occurs less frequently than commonly thought. Outbreaks are known to occur in preschools, hospitals, nursing homes, prisons, other institutional residences, and other congregate settings.

The rash and pruritus associated with scabies arise from a sensitization reaction to mites and their secretions/excretions. A person's initial infestation typically remains asymptomatic for up to 6 weeks before intense pruritus starts. Reinfestation, however, promptly induces a hypersensitivity reaction. Burrows become surrounded by inflammatory infiltrates composed of eosinophils, lymphocytes, and histiocytes. Infested individuals often feel generalized pruritus, not just in the most heavily involved areas.

Pruritus typically intensifies at night and after hot showers. Burrows appear as dark wavy lines in the upper epidermis and are 3–15 mm long **(Fig. 472-1)**. Classic burrows are often difficult to find because most patients have perhaps only 10–15 burrows. However, most burrows are obscured by excoriations or secondary bacterial infections. Scabietic lesions are most common on the volar wrists and along the digital web spaces. In males, the genitalia (penile glans and shaft and the scrotum) are nearly always involved. Small papules and vesicles, often accompanied by eczematous plaques, pustules, or nodules, appear symmetrically at those sites. The axillae and other intertriginous areas; around the navel and belt line; and the buttocks and upper thighs are also common sites. Except in infants, the face, scalp, neck, palms, and soles are usually spared.

Hyperinfestation with thousands of mites, a condition known as *crusted* scabies (formerly *Norwegian* scabies), may result from glucocorticoid use, immunodeficiency (including that due to HIV/AIDS), and neurologic or psychiatric disorders that dampen the itch or impair the scratch response. Crusted scabies often resembles psoriasis: both are characterized by widespread thick yellowish-white keratotic crusts, scaly plaques, and dystrophic nails. Characteristic burrows are not seen in crusted scabies, and patients are often not itchy, even though their infestations are highly contagious and have been responsible for outbreaks of common scabies in hospitals.

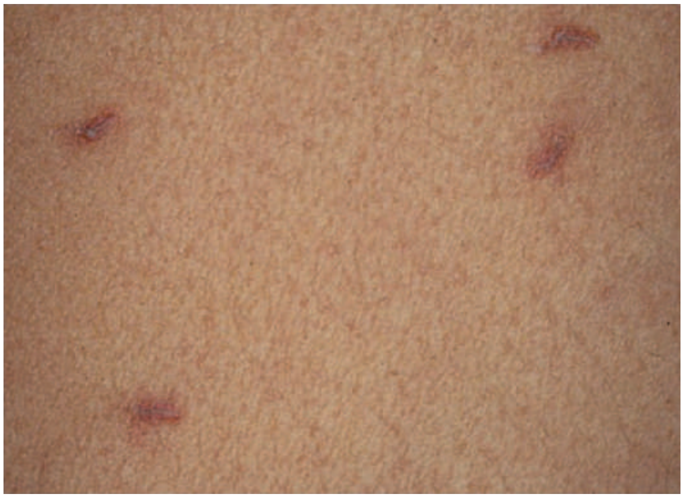

FIGURE 472-1 Scabies burrows. Scabies mites create short, delicate, linear burrows within the superficial epidermis. Although burrows are pathognomonic for scabies, most have been altered by scratching or secondary bacterial infection.

Scabies should be in the differential diagnosis for patients with pruritus and symmetric superficial, excoriated, papulovesicular skin lesions in characteristic locations, particularly if they have had direct and prolonged contact with an infested person. The diagnosis can be confirmed by microscopic examination of material scraped from burrows. Burrows should be sought and unroofed with a sterile needle or scalpel blade, and the scrapings should be examined microscopically for mites, eggs, and fecal pellets. Examination of skin scrapings; biopsy specimens; material obtained by clear cellophane tape or cyanoacrylate adhesive lifted from lesions; dermatoscopic imaging of papulovesicular lesions; and microscopic inspection of clear cellophane tape lifted from lesions also may be diagnostic. In the absence of identifiable mites or eggs, a clinical diagnosis is based on a triad of pruritus, findings on physical examination, and an epidemiologic link. Unrelated skin diseases are frequently misdiagnosed as scabies, particularly in presumed "outbreak" situations. *Sarcoptes* mites of other mammals may cause transient irritation, but they do not reside or reproduce in human hosts. In some Australian Aboriginal communities, household dogs may serve as reservoirs for human scabies mites.

TREATMENT

Scabies

The U.S. Food and Drug Administration (FDA) has approved four scabicides: permethrin, crotamiton, spinosad, and lindane. They are topical products and available by prescription only. Permethrin cream (5%) is less toxic than 1% lindane preparations and is effective against lindane-resistant infestations. In adults, scabicides are applied thinly but thoroughly from the jawline to toes after bathing, with careful application to interdigital spaces, the navel, and under the nails. and washed off 6–14 h later with soap and water. These products are relatively ineffective against scabies eggs, so a second round of treatment 1 week later is advisable.

Treating crusted scabies is difficult and requires repeated courses of both topical (permethrin) and oral (ivermectin) agents. Preapplication of a keratolytic agent, such as 6% salicylic acid, will help debulk the crusts. Permethrin should be applied to the skin's entire surface, including the scalp, face, and ears. Oral ivermectin is not FDA-approved for treating scabies but is approved to treat the nematodal diseases hookworm and strongyloidiasis. A single oral dose of ivermectin (200 μg/kg) is effective in healthy patients with common (noncrusted) scabies. Patients with crusted scabies require three to seven doses of ivermectin over 8–30 days, along with repeated applications of topical permethrin and possibly a keratolytic compound.

Within 1 day of properly administered treatment, a scabies infestation is considered noncommunicable, thereby permitting a patient to return to work, school, and other public activities. Nevertheless, remind patients that dead mites and their detritus may continue to produce the pruritic hypersensitivity dermatitis for several weeks. Unnecessary retreatment with topical agents, especially permethrin cream, may cause an irritant contact dermatitis. Topical emollients and antipruritic agents, menthol and methyl salicylate products, calamine lotion, and oral antihistamines relieve itching during treatment. Topical glucocorticoids may calm pruritus that lingers after effective treatment. To prevent reinfestations, bedding and clothing should be washed and dried on high heat or heat-pressed, and other environmental surfaces or potential sources of fomites should be cleaned. Household members and other close contacts of confirmed cases, even if asymptomatic, should be treated simultaneously to help prevent back-and-forth reinfestations.

Scabies infestations often lead to secondary bacterial infections, usually with *Staphylococcus aureus*, *Streptococcus pyogenes*, or both. Consequences of superinfections include impetigo, cellulitis, invasive bacterial infections, poststreptococcal glomerulonephritis (and subsequent kidney disease), and possibly acute rheumatic fever (and subsequent cardiac valvular disease).

CHIGGERS AND OTHER BITING MITES

Chiggers are the larvae of trombiculid (harvest) mites that feed mostly on mice and other small vertebrates in grassy or brush-covered sites in tropical, subtropical, and (less frequently) temperate areas during warm months. They reside on low vegetation and attach themselves to passing vertebrate hosts. While feeding, larvae secrete proteolytic saliva that penetrates the epidermis, creating a tube-like invagination, called a stylostome, in the host's skin. The stylostome allows the mite to imbibe tissue fluids. The saliva is highly antigenic and causes small (usually <1 cm in diameter) but exceptionally itchy papular, urticarial, or pustulovesicular lesions. In people already sensitized to salivary antigens, the papules develop within hours of attachment. While attached, chiggers appear as minute (~0.5 mm diameter) red dots on the skin. Generally, lesions have a hemorrhagic base and are slightly elevated, resembling purpuric papules seen in cutaneous small-vessel vasculitis.

Chiggers remain on their host animals for several days before falling to the ground to complete the nonparasitic stages of their life cycle. However, in humans, the intense pruritus leads to scratching that invariably destroys the chigger before it completes feeding, although the itching and burning often persist for weeks. The rash is common on the ankles and along the belt line and in areas where circumferentially tight clothing obstructs the further wanderings of the mites. Topical repellents on one's skin and insecticide-treated clothing are useful for preventing chigger bites.

Chiggers (primarily *Leptotrombidium* species) serve as vectors for intracellular rickettsial organisms in the genus *Orientia* that cause scrub typhus. Scrub typhus caused by *Orientia tsutsugamushi* was traditionally confined to the eastern half of Asia and the Indomalayan and Australasian regions. Their bites may be asymptomatic. Endemic foci of closely related *Orientia* species have recently been identified in southern Chile, East Africa, and the Arabian peninsula, where they cause scrub typhus-like illnesses. Additional areas of transmission will likely be found outside the usual endemic regions. Only larval trombiculids are predatory on mammals, and the larvae acquire their *Orientia* load from transovarial transmission from their mother mites.

Many kinds of mites feed on peridomestic birds or rodents and become particularly bothersome when they invade homes and bite people. In North America, the northern fowl mite, chicken mite, tropical rat mite, and house mouse mite normally feed on poultry, other birds, and small mammals. After their natural hosts leave the nest or die, the mites disperse and may invade homes. Although the mites are rarely seen because of their small size, their bites can be painful and pruritic. House mouse mites (*Liponyssoides sanguineus*) serve as vectors for the agent of rickettsialpox, *Rickettsia akari*, an uncommon disease characterized by mild fevers, an eschar at the bite site, and a papulovesicular eruption. Rickettsialpox **(Chap. 192)** has been recognized mainly in large northern temperate urban areas. Once these environmental mites are confirmed as causing pruritic eruptions, they are best eliminated by excluding their animal hosts, removing their nests, and cleaning and treatment of the nesting area with appropriate acaricides.

Pyemotes mites are external parasites of insect larvae and commonly occur in insect-infested products such as grain, straw, cheese, hay, and oak leaf galls. Their saliva contains a potent neurotoxin they use to immobilize their prey. The straw itch mite (*Pyemotes tritici*) occurs worldwide and is an occupational hazard for agriculture workers and others who work with dry plant material. Bites in humans are associated with pruritic rashes and may produce a unique dermatologic "comet sign" lesion—a paisley-shaped urticarial plaque **(Fig. 472-2)**. The oak leaf gall mite (*Pyemotes herfsi*) has been associated with several outbreaks of dermatosis in the central and eastern United States. These mites feed on oak gall midges and periodic cicada eggs in oak trees, and their populations fluctuate widely each year based on host availability. In the late summer and fall, mites drop or are wind-blown from oak trees and may land on a person. Bites in people are usually reported on exposed surfaces, such as neck, face, arms, and upper torso. Intense itching is reported 10–16 h after the initial bite, which most people do not recall.

Diagnosis of mite-induced dermatitis (including those caused by chiggers) relies on confirmation of the mite's identity or elicitation of a history of exposure to the mite's source. Unlike scabies mites that burrow and live in one's skin, environmental mites do not reside on humans. Therefore, treatment of the patient with acaricides (e.g., permethrin) is discouraged. Oral antihistamines or topical steroids may reduce mite-induced pruritus temporarily.

The mites that cause house dust–related allergic conditions neither bite nor infest humans.

TICK BITES AND TICK PARALYSIS

Ticks attach superficially to skin and usually feed painlessly; blood is their only food. Their salivary secretions are biologically active, preventing the host's blood from coagulating while the tick feeds. Tick saliva can transmit various pathogens and can induce several "sterile reactions" in the host, such as a local inflammatory response, fevers, and tick-bite paralysis. The two main families of ticks are the hard ticks (Ixodidae) and soft ticks (Argasidae). Because no ticks are obligate parasites on humans, all tick-borne diseases (bacterial, viral, and protozoal) are zoonotic.

Generally, soft (argasid) ticks feed quickly, attaching to a host, completing a meal in <1 h, and then dropping off the host. Because of their rapid feeding habits, soft ticks are not carried widely by their vertebrate hosts. Soft tick–associated infections usually have fairly focal distributions. On humans, red macules may develop at the bite site. Some species in Africa, the western United States, and Mexico produce painful hemorrhagic lesions.

Hard (ixodid) ticks are much more common than soft ticks, and they transmit most tick-borne infections that are familiar to physicians, patients, and the public. Hard ticks attach to the host and feed for several days to >1 week, with the exact duration depending upon the tick's species and stage of development. At the site of hard-tick bites, small areas of induration, often purpuric, develop and may be surrounded by an erythematous rim. A necrotic eschar, called a *tâche noire* ("black spot"), occasionally develops. Long-lasting dermal nodules, called persistent tick-bite granulomas, occasionally appear where the host is bitten. The nodules are roughly 1–3 cm in diameter and may linger for months after the feeding tick has fallen off or been removed. The granulomas can be treated with injected intralesional glucocorticoids, by simple local excision, or simply left to resolve on their own over perhaps a year or more. Tick-induced fever, unassociated with any pathogen, is often accompanied by headache, nausea, and malaise but usually resolves ≤ 36 h after the tick is removed.

Tick bites are also associated with a recently recognized allergic condition known as alpha-gal syndrome (AGS) or red meat allergy. This syndrome occurs in humans when exposure to tick saliva induces production of IgE antibodies that cross-react with a carbohydrate molecule, galactose-α-1,3-galactose (alpha-gal), found in muscle tissues of all mammals except humans, other apes, and Old World monkeys. In the United States, AGS is attributed to bites from the lone star tick (*Amblyomma americanum*). The geographic distribution of human cases closely aligns with that tick's geographic range. In Australia and Europe, AGS is associated with *Ixodes* spp., and in Asia, it is associated with *Haemaphysalis longicornis*. It is important for individuals with AGS to avoid additional tick bites as this may prolong or worsen their allergic reactions.

Tick paralysis, an acute ascending flaccid paralysis, is believed to be caused by one or more unidentified toxins in tick saliva that block neuromuscular transmission and decrease nerve conduction. This rare complication has been associated with >70 species of ticks. Although it has been reported worldwide, most cases occur in the Rocky Mountain region, in northwestern United States and southwestern Canada, and along Australia's eastern seaboard. In North America, dog and wood ticks (*Dermacentor* species) are most commonly implicated. Weakness begins several days after the tick attaches to the host. The Guillain-Barré-like paralysis starts in the lower legs and ascends symmetrically over several days. The paralysis may culminate in complete paralysis of the extremities and cranial nerves. Deep tendon reflexes are diminished or absent, but sensory examination and findings on lumbar puncture are typically normal. Diagnosis and treatment depend on finding the tick, which is often hidden beneath scalp hair. Removal

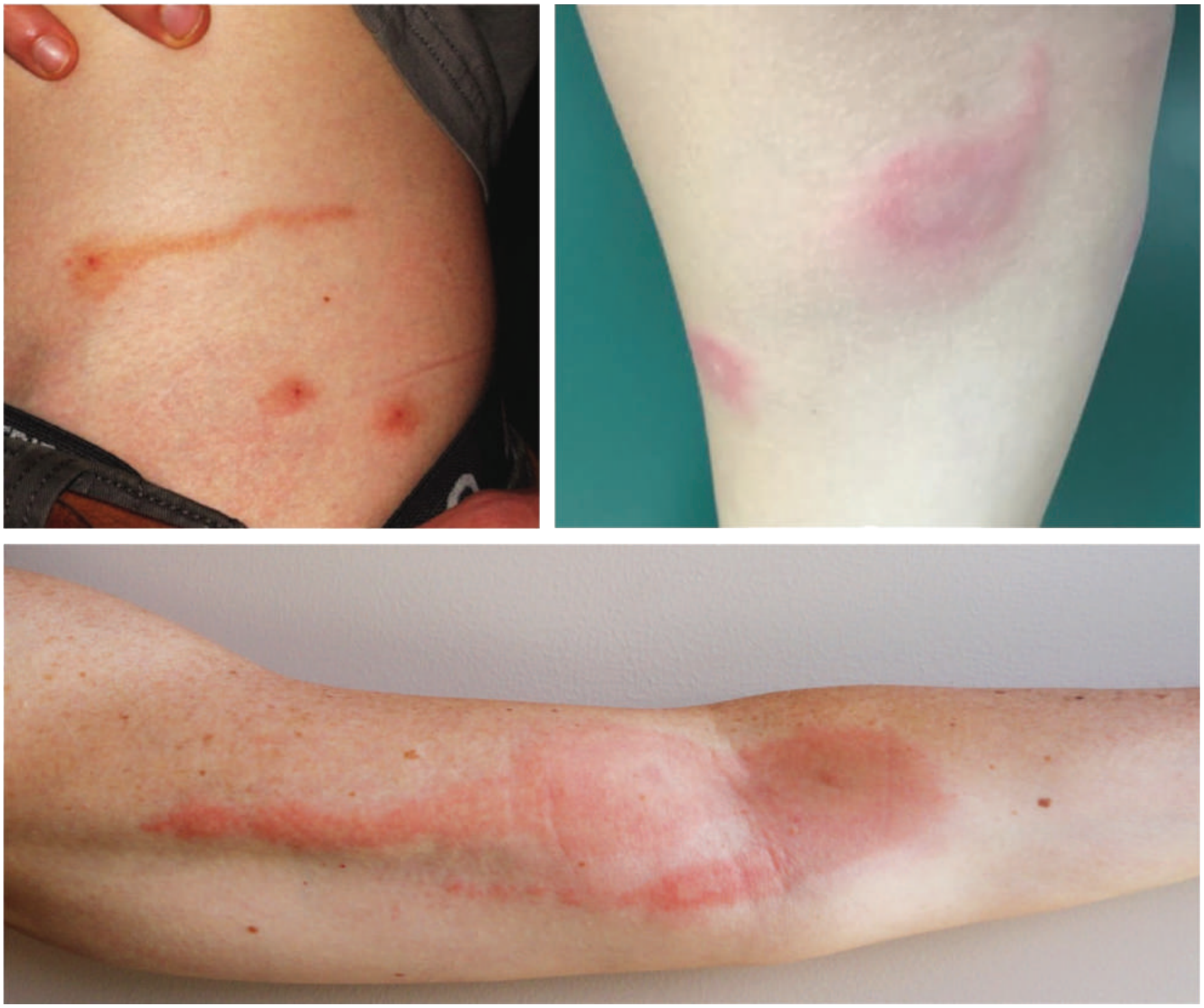

FIGURE 472-2 Comet signs in individuals with known or suspected mite-bite reactions, likely due to *Pyemotes* species. Note central punctum at bite site, surrounded by edematous erythema. Linear or serpiginous "comet tails" emanate from the central site. *Pyemotes*-induced comet tails generally do not follow typical patterns of ascending lymphatic drainage.

of the tick generally leads to improvement within a few hours and complete recovery after several days, although the patient's condition may continue to deteriorate for a full day. Failure to remove the tick may lead to dysarthria, dysphagia, and ultimately death from aspiration or respiratory paralysis.

Removal of hard ticks during the first 24 h of attachment generally prevents transmission of the agents of Lyme disease, babesiosis, anaplasmosis, and ehrlichiosis, although tick-borne viruses may be transmitted more quickly. Ticks should be removed by traction with fine-tipped forceps placed firmly around the tick's mouthparts where they enter the skin. Careful handling (to avoid rupture of ticks) and use of gloves may avert accidental contamination with pathogens contained in the tick's body fluids. Attempting to induce the tick to detach by applying heat or occlusive compounds or dressings merely delays tick removal. After removal, the site of attachment should be disinfected. Tick mouthparts sometimes remain in the skin but generally are shed spontaneously within days without the need for surgical removal. Current guidelines from the Centers for Disease Control and Prevention suggest that, rather than awaiting the onset of erythema migrans, the results of tick testing, or seroconversion to antigens diagnostic for Lyme disease, physicians may appropriately administer prophylaxis—a one-time oral dose of doxycycline (200 mg) within 72 h of tick removal—to adult patients with bites thought to be from *Ixodes scapularis* (black-legged or deer ticks) in Lyme disease–endemic areas. Whereas antibiotic prophylaxis may help prevent Lyme disease, it is not recommended as a way to prevent other tick-borne infections.

The Asian longhorned tick (*Haemaphysalis longicornis*) is a newly invasive species in the United States, first detected in the northeastern states in 2017. Although it carries several pathogens to domestic animals, wildlife, and humans in its natural range (northeastern Asia), it has not yet been implicated in disease transmission in the United States.

LOUSE INFESTATION (PEDICULOSIS AND PTHIRIASIS)

Three kinds of biting lice are obligate blood-feeding ectoparasites of human beings. These include the human head louse and the human body louse that represent distinct genetic clades of *Pediculus humanus*, and the pubic ("crab") louse (*Pthirus pubis*). Nymphs and adults of these lice feed at least once a day, ingesting human blood exclusively, and they partition ecologically on the host. Head lice infest mainly the scalp and scalp hair; body lice infest clothing; and pubic lice infest mainly pubic hair. The saliva of lice produces a pruritic morbilliform or urticarial rash in some sensitized persons. Female head lice and pubic lice cement their eggs (nits) firmly to human hair, whereas female body lice cement their eggs to clothing, particularly to threads along clothing seams. After ~10 days of development within the egg, a nymph emerges. Empty eggs may remain affixed to hair for months thereafter.

Body lice are acquired by direct contact with an infested person or their recently used clothing or bedding. These lice venture for just minutes to the skin to feed, but otherwise sequester on clothing. They generally succumb in ≤2 days if separated from their host. In developed countries, body lice are generally uncommon, found only on indigent persons who have relevant exposure and lack the wherewithal or desire to change or appropriately launder their clothing and bedding.

Body lice are vectors for the agents of louse-borne (epidemic) typhus **(Chap. 192)**, louse-borne relapsing fever **(Chap. 190)**, and trench fever **(Chap. 177)**. Body lice and their associated diseases may proliferate whenever societal upheaval and disasters limit access to clean clothes or laundry facilities. This scenario was exacerbated during the trench warfare of World War I, when troops often wore one uniform for weeks or months at a time. During damp, cold months, entrenched soldiers huddled together to stay warm, enabling body lice and their pathogens to spread easily from person to person.

Infestations result in a postinflammatory hyperpigmentation and thickening of the skin known as *vagabond's disease*. When inspecting a patient for signs of a body lice infestation, one must examine their clothing seams for nits and live lice.

Head lice are acquired mainly by direct head-to-head contact rather than via fomites such as shared hats, caps, headgear, bed linens, and grooming implements. Worldwide, the prevalence of head lice varies widely as a function of age, geography, and cultural habits. In North America, the overall prevalence of head lice infestation is generally low, but most communities, irrespective of socioeconomic circumstances, experience episodic high-prevalence focal outbreaks among younger school-aged children. Although these events generally spare adults, they can create considerable distress among families in those communities.

An infested person generally hosts 10 or fewer head lice, but the nits, even when empty or nonviable, remain cemented to one's scalp hairs as they grow. Thus, accumulated nits make it seem that a person has hundreds of head lice, even after successful treatment. Consequently, many schools retain unnecessarily onerous "no nit" policies for school attendance, even though a child has been treated successfully. Bite reactions appear as small, <1 cm diameter edematous pink wheals and are most evident along the posterior hairline. Pruritus, attributed to hypersensitivity to the louse's saliva, leads to scratching, which may lead to secondary impetigo with *S. aureus* or *S. pyogenes*.

Body lice are well known to transmit diseases, but can head lice (besides superficial secondary bacterial infections)? Several studies using polymerase chain reaction on freshly caught head lice have detected pathogenic bacteria, but clinical and epidemiologic data are generally interpreted to show that head lice are *not* natural vectors of any disease. On the other hand, experimental models suggest that head lice may have some competence as vectors for *Rickettsia prowazekii* (louse-borne typhus) and *Bartonella quintana* (trench fever). Indeed, sporadic reports indicate that cycles of trench fever are maintained in Ethiopia and elsewhere in Africa by head lice in the absence of body lice.

Crab or pubic lice are found mostly on pubic hair and are transmitted mainly by direct sexual contact. They also occur uncommonly on axillary or facial hair, including the eyelashes, as the result of either sexual or close nonsexual contact. Intensely pruritic, bluish macules <5 mm in diameter (*maculae ceruleae*) may develop at bite sites. Blepharitis commonly accompanies infestations of the eyelashes. Eyelash infestations in children should not be construed as incontrovertible evidence of inappropriate sexual contact.

Pediculosis is often suspected upon the detection of presumed nits to hairs or simply on the basis of an itchy scalp. Objects presumed to be louse eggs are often pseudo-nits composed of debris, flakes of dry skin, and hair-associated fungi. Hatched and dead eggs remain firmly affixed to scalp hair for months. Such relics are frequently misconstrued to be signs of an active louse infestation. Confirmation of a louse infestation, therefore, should rely on the discovery of a live louse.

TREATMENT

Louse Infestation

Body lice usually are eliminated by bathing and by changing to new or properly laundered clothes. Fabric is effectively deloused by heat, such as in a clothes dryer at ≥55°C (≥131°F) for 30 min or by heat-pressing. Emergency mass delousing may be warranted in congregate settings during periods of civil strife and after natural disasters to reduce the high risk of disease transmission by body lice.

Head lice and nits may be removed with a fine-toothed louse or nit comb, but this effort is time-consuming and often fails to eradicate the lice. Treatment of newly identified, active infestations traditionally relies on a 10-min topical application of a FDA-approved 1% permethrin lotion or pyrethrin-based shampoo, which are available over the counter. Also FDA-approved, but requiring a prescription, are 0.5% malathion lotion, 0.5% ivermectin lotion, and 0.9% spinosad suspension. These products do not kill viable nits, so a second application roughly 10 days later is required to clear newly hatched lice. Lice persisting after this treatment may be due to back-and-forth reinfestations within a family, or the lice may be resistant to chemical treatment. Resistance to permethrin, malathion, and lindane is well documented, with less resistance associated with ivermectin and spinosad products. Although children infested by head lice—or those who simply have remnant nits from a previous infestation—are frequently isolated or excluded from school, this practice increasingly is considered to be unjustified, ineffective, and counterproductive.

Pubic louse infestations are treated with topical pediculicides, except for eyelid infestations (*pthiriasis palpebrum*), which generally respond to a coating of petrolatum applied for 3–4 days.

MYIASIS (FLY INFESTATION)

Myiasis refers to infestations by fly larvae (maggots) that invade living or necrotic tissues or body cavities and produce different clinical syndromes, depending on the species of fly.

In forested parts of Central and South America, larvae of the human botfly (*Dermatobia hominis*) produce furuncular (boil-like) dermal and subcutaneous nodules ≤3 cm in diameter. A gravid adult female botfly captures a mosquito or another bloodsucking insect and deposits her eggs on its abdomen. When the carrier insect attacks a human or another mammalian host (often cattle) several days later, the warmth and moisture of the host's skin stimulate the eggs to hatch. The emerging larvae, ~1 mm long, promptly penetrate the skin. After 6–12 weeks of development, mature larvae emerge from the skin, drop to the ground, pupate, and finally metamorphose into adults.

The African tumbu fly (*Cordylobia anthropophaga*, also called mango fly or mputsi fly) deposits its eggs on damp sand, leaf litter, drying laundry, or damp clothing contaminated by urine or sweat. Larvae hatch from eggs upon contact with a host's body and penetrate the skin, producing boil-like lesions from which mature larvae emerge ~9–10 days later. Furuncular myiasis is suggested by uncomfortable lesions with a central breathing pore that emit bubbles when submerged in water. The sensation of movement within the patient's skin may cause severe (and understandable) emotional distress **(Fig. 472-3)**.

Larvae that cause furuncular myiasis may be induced to emerge if the breathing pore is coated with petrolatum or another occlusive substance. Removal may be facilitated by injection of a local anesthetic (or sterile injectable saline) into the subjacent tissue to uplift the larva through the breathing pore. Surgical excision is sometimes necessary because upward-pointing spines of some species hold larvae firmly in place.

Other fly larvae cause several forms of nonfuruncular myiasis. Larvae of the horse botfly (*Gasterophilus intestinalis*) emerge from eggs that typically are laid on the hairs of a horse's front legs. Direct contact with a person's skin may cause the eggs to hatch and the new larvae to invade human skin. After penetrating human skin, these larvae rarely mature but instead may migrate for weeks in the dermis. The resulting pruritic and serpiginous eruption resembles cutaneous larva migrans caused by canine or feline hookworms **(Chap. 238)**. Larvae of rabbit and rodent botflies (*Cuterebra* species) occasionally cause cutaneous or tracheopulmonary myiasis. Larvae of the sheep botfly, *Oestrus ovis*, and other flies responsible for furuncular and wound myiasis also may cause ophthalmomyiasis. Sequelae include nodules in the eyelid, retinal detachment, and destruction of the globe.

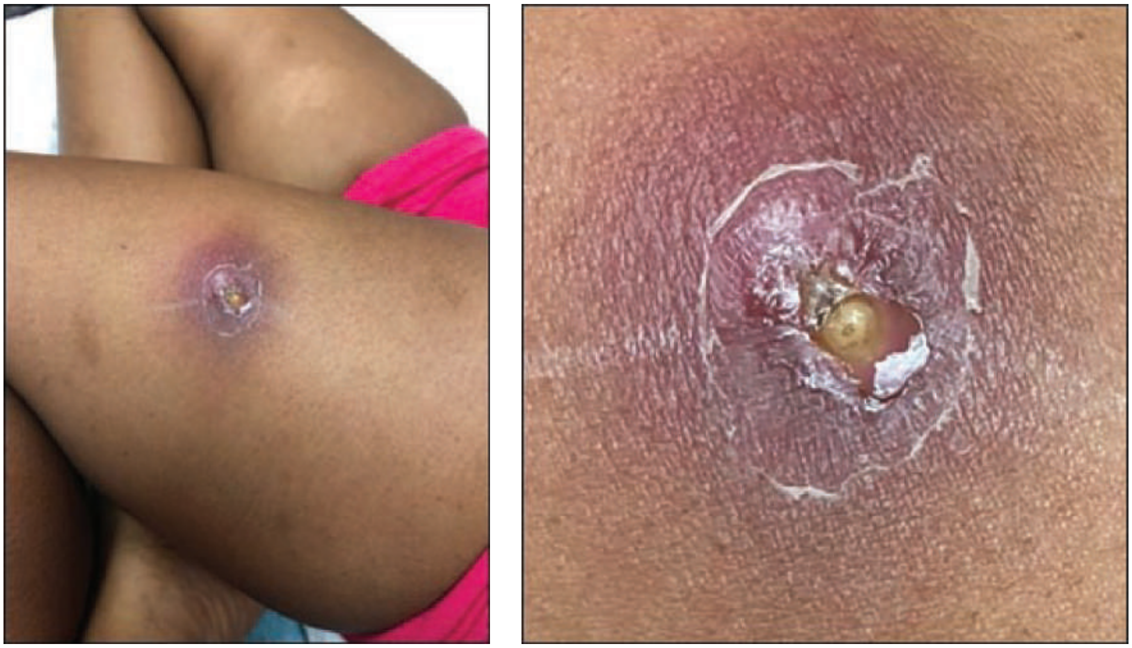

FIGURE 472-3 Furuncular myiasis with larva of *Cordylobia anthropophaga* in the thigh of a teenage girl who recently returned from visiting relatives in East Africa ***(left)***. The central pore shows the respiratory elements of the embedded fly larva ***(right)***.

Certain flies are attracted to blood and pus, laying their eggs on open or draining sores. Newly hatched larvae enter wounds or diseased skin. Larvae of several species of green bottle flies (*Lucilia* spp.) usually consume only necrotic tissues. Specially raised, sterile "surgical maggots" are available as FDA-approved medical devices for debridement of chronic wounds, such as deeply infected diabetic ulcers. On the other hand, larvae of screwworm flies (*Cochliomyia* spp.) and flesh flies (*Wohlfahrtia* spp.) can invade viable tissues more deeply and cause large suppurating lesions. These can be extraordinarily harmful to domesticated herd animals. Larvae of the sheep botfly, *Oestrus ovis*, may cause nasal or external ophthalmomyiasis, involving tear ducts, eyelids, and conjunctivae.

Larvae that infest decaying tissues may enter body cavities such as the mouth, nose, ears, sinuses, anus, vagina, and lower urinary tract, particularly in unconscious or otherwise debilitated patients. The consequences range from harmless colonization to destruction of the nose, meningitis, and deafness. Treatment involves removal of maggots, surgical debridement of tissue, and treatment of secondary infections.

Maggots are occasionally found in human feces, usually the result of a fly laying eggs on recently passed stools, and not as evidence of an intestinal maggot infestation.

PENTASTOMIASIS

Pentastomids (tongue worms), an obscure type of crustacean, inhabit respiratory passages of reptiles and carnivorous mammals. Human infestation by *Linguatula serrata* is common in the Middle East and results from eating encysted larval stages in raw liver or lymph nodes of sheep and goats, which are true intermediate hosts for the tongue worms. In areas where raw sheep and goat liver are served, pentastomid larvae migrate to the person's nasopharynx and produce an acute self-limiting syndrome—known as *halzoun* in Lebanon and *marrara* in Sudan—characterized by rapid onset (within <12 h) of pain and itching of the throat and ears, coughing, hoarseness, dysphagia, and dyspnea. Severe edema may cause obstruction that requires tracheostomy. In addition, ocular invasion has been described. Diagnostic larvae measuring ≤10 mm in length appear in copious nasal discharge or vomitus.

Another type of tongue worm, *Armillifer armillatus*, infects people who consume its eggs in contaminated food or drink or after handling the definitive host, the African python. Larvae encyst in various organs, usually the liver or peritoneum, but rarely cause symptoms. Cysts may require surgical removal as they enlarge during worm molting, but they usually are encountered as an incidental finding at autopsy. Parasite-induced lesions may be misinterpreted as a malignancy, with the correct diagnosis only confirmed histopathologically. Larva migrans–type syndromes caused by other pentastomes have been reported from Southeast Asia and Central America.

LEECH INFESTATIONS

Medically important leeches are annelid worms that attach to their hosts with chitinous cutting jaws and draw blood through muscular suckers. Medicinal leeches (Europe: *Hirudo medicinalis* and other *Hirudo* species; Asia: *Hirudinaria manillensis*; North America: *Macrobdella decora*) are still used occasionally for medical purposes to reduce venous congestion in surgical flaps or replanted body parts. This practice has been complicated by intractable bleeding, wound infections, myonecrosis, and sepsis due to *Aeromonas hydrophila*, which colonizes the gullets of commercially available leeches.

Ubiquitous aquatic leeches that parasitize fish, frogs, and turtles readily attach to human skin—most often the nasal mucosa—and avidly suck blood. Attachment is usually painless, and the leeches will detach themselves when satiated with a blood meal. Hirudin, a powerful anticoagulant secreted by the leech, causes continued bleeding after the leech has detached. Healing of a leech-bite wound is slow, and secondary bacterial infections are not uncommon. Several kinds of aquatic leeches in Africa, Asia, and southern Europe can enter the mouth, nose, and genitourinary tract and attach to mucosal surfaces at sites as deep as the esophagus and trachea. Leeches may detach on exposure to gargled saline or may be removed by forceps or medical suction.

Arboreal land leeches, which live amid rain forest vegetation, are attracted by heat and can drop from a leaf onto one's skin. Externally attached leeches generally drop off after they have engorged, but removal is hastened by gentle scraping aside of the anterior and posterior suckers the leech uses for attachment and feeding. Some authorities dispute the wisdom of removing leeches with alcohol, salt, vinegar, insect repellent, a flame or heated instrument, or applications of other noxious substances.

SPIDER BITES

All spiders are carnivores and most have venomous bites intended to immobilize and digest their prey. However, very few spiders are capable of biting humans, and the vast majority of putative spider bites have nothing to do with spiders. Of >45,000 recognized species of spiders, only about 60 are medically important to humans. In the United States, only recluse spiders (*Loxosceles* spp.) and widow spiders (*Latrodectus* spp.) are considered medically important.

Most spider bites are painful but do not require medical attention. Identification of the offending spider is important because specific treatments exist. Except when the patient actually observes a spider biting them or fleeing from a site of sudden sharp pain, most acute tender noduloulcerative lesions reported as spider-bite reactions are due to an unrelated minor injury or caused by an acute bacterial infection, particularly by methicillin-resistant *S. aureus* (MRSA).

Recluse Spider Bites and Necrotic Arachnidism Brown recluse spiders (*Loxosceles reclusa*) live mainly in the southcentral United States and have close relatives in Central and South America, Africa, the Mediterranean basin, and the Middle East. Recluse spiders are not aggressive toward humans and bite only if threatened or pressed against the skin. They generally live beneath rocks and logs or in caves and animal burrows. They invade homes and seek dark and undisturbed hiding spots in closets, garages, crawl spaces, and attics; under furniture and rubbish in storage rooms; and in folds of clothing. Despite their impressive abundance in some homes, these spiders rarely bite humans. Bites tend to occur while the victim is donning clothing in which the spider has hidden itself and are sustained primarily on the hands, arms, neck, and lower abdomen.

A brown recluse spider's bite may cause minor injury with edema and erythema, and envenomation can cause severe necrosis of skin and subcutaneous tissue and, more rarely, systemic hemolysis. Initially, the bite is painless or may produce a stinging sensation. Within a few hours, the site becomes painful and sensitive to touch, with central induration surrounded by a pale ischemic zone that itself is encircled by a zone of erythema. In most cases, the lesion resolves without treatment in just a few days. In severe cases, systemics signs, such as fever, chills, weakness, headache, nausea, vomiting, myalgia, arthralgia, and leukocytosis, may soon develop. As the bite site evolves, the erythematous zone expands, and the center becomes hemorrhagic or necrotic with an overlying bulla. A black eschar forms and sloughs several weeks later, leaving an ulcer that eventually may create a depressed scar. Healing usually takes place in ≤3 months. Local complications include injury to nerves and secondary bacterial infection. Reports of deaths attributed to bites of North American brown recluse spiders have not been verified.

The Mediterranean recluse spider (*Loxosceles rufescens*) is a widely invasive species in urban areas of both the Old and New Worlds. The dorsal surfaces of *L. rufescens* and *L. reclusa* are adorned with a fiddle-shaped pattern. *L. rufescens* is warier than *L. reclusa*, is less likely to bite, and rarely causes necrosis. Misidentification of this spider may create spurious reports of *L. reclusa* activity outside the known range of that species.

TREATMENT

Recluse Spider Bites

Initial management includes rest, ice, compression, and elevation (RICE). Analgesics, antihistamines, antibiotics, and tetanus prophylaxis should be administered if indicated. Early debridement or surgical excision of the wound without closure delays healing. Routine use of antibiotics or dapsone lacks utility. Patients should be monitored closely for signs of hemolysis, renal failure, and other systemic complications.

Widow Spider Bites The southern black widow spider (*Latrodectus mactans*) is common in the southeastern United States. Female bodies are ~1 cm in length, but the leg span may be ~5 cm across. Their bodies are shiny black with a red hourglass marking on the ventral abdomen. Other dangerous *Latrodectus* species occur elsewhere in temperate and subtropical parts of the world. The bites of the female widow spiders are notorious for their potent neurotoxins.

Widow spiders spin their webs under stones, logs, plants, or rock piles and in dark spaces in barns, garages, and outhouses. Bites are most common in the summer and early autumn and occur when a web is disturbed or a spider is trapped or provoked. The initial bite is perceived as a sharp pinprick or may go unnoticed. Fang-puncture marks are uncommon. Envenomation does not cause local tissue damage, and some persons experience no further symptoms.

α-Latrotoxin, the most active component of the venom, is a neurotoxin. It binds irreversibly to presynaptic nerve terminals and causes release and eventual depletion of acetylcholine, norepinephrine, and other neurotransmitters from those terminals. Painful cramps may spread within 60 min from the bite site to large muscles of the extremities and trunk. Extreme abdominal pain and rigidity may resemble peritonitis, but the abdomen is not tender on palpation and surgery is not warranted. The pain begins to subside during the first 12 h but may recur over the next few days or weeks before resolving spontaneously. A wide range of other sequelae, largely neurologic, may include salivation, diaphoresis, vomiting, hypertension, tachycardia, labored breathing, anxiety, headache, weakness, fasciculations, paresthesia, hyperreflexia, urinary retention, uterine contractions, and premature labor. Rhabdomyolysis and renal failure have been reported, and respiratory arrest, cerebral hemorrhage, or cardiac failure may end fatally, especially in very young, elderly, or debilitated persons.

TREATMENT

Widow Spider Bites

Treatment consists of RICE and tetanus prophylaxis. Hypertension that does not respond to analgesics and antispasmodics (e.g., benzodiazepines or methocarbamol) requires specific antihypertensive medication. The efficacy and safety of antivenin made from equine immunoglobulins are controversial for black widow bites because of potential anaphylaxis or serum sickness. Antivenins made from monoclonal antibodies are in development.

Tarantulas and Other Spiders Tarantulas are large hairy spiders of which 30 species are found in the United States, mainly in the Southwest. Several species of tarantulas have become popular household pets and are usually imported from Central or South America. Tarantulas bite people only when threatened and usually cause no more harm than a bee sting, but on occasion, the venom causes deep pain and swelling. Several species of tarantulas are covered with urticating hairs that are brushed off in the thousands when a threatened spider rubs its hind legs across its dorsal abdomen. These hairs can penetrate human skin and produce pruritic papules that may persist for weeks. Failure to wear gloves or to wash the hands after handling the Chilean Rose tarantula, a popular pet spider, has resulted in transfer of hairs to the eye with subsequent devastating ocular inflammation. Treatment of bites includes local washing and elevation of the bitten area, tetanus prophylaxis, and analgesic administration. Antihistamines and topical or systemic glucocorticoids are given for exposure to urticating hairs.

Atrax robustus, a funnel-web spider of Australia, and *Phoneutria* species, the South American banana spiders, are among the world's most dangerous spiders because of their aggressive behavior and potent neurotoxins. Envenomation by *A. robustus* causes a rapidly progressive neuromotor syndrome that can be fatal within 2 h. The bite of a banana spider causes severe local pain followed by profound systemic symptoms and respiratory paralysis that can lead to death within 2–6 h. Specific antivenins for use after bites by each of these spiders are available. Yellow sac spiders (*Cheiracanthium* species) are common in homes worldwide. Their bites, though painful, generally lead to only minor erythema, edema, and pruritus.

■ SCORPION STINGS

Scorpions are arachnids that feed on arthropods and other small animals. They paralyze their prey and defend themselves by injecting venom from a stinger on the tip of the tail. Painful but relatively harmless scorpion stings need to be distinguished from the potentially lethal envenomations from about 30 of roughly 1000 known species, which cause several thousand deaths worldwide each year. Scorpions are nocturnal and remain hidden during the day in crevices or burrows or under wood, loose bark, or rocks. They occasionally enter houses and tents and may hide in shoes, clothing, or bedding. Scorpions sting humans only when threatened.

Of approximately 40 scorpion species in the United States, only bark scorpions (*Centruroides sculpturatus*/*C. exilicauda*) in the Southwest produce venom that is potentially lethal to humans. This venom contains neurotoxins that cause sodium channels to remain open. Such envenomations usually are associated with little swelling, but prominent pain, paresthesia, and hyperesthesia can be accentuated by tapping on the affected area (the *tap test*). These symptoms soon spread to other locations; dysfunction of cranial nerves and hyperexcitability of skeletal muscles develop within hours. Patients present with restlessness, blurred vision, abnormal eye movements, profuse salivation, lacrimation, rhinorrhea, slurred speech, difficulty in handling secretions, diaphoresis, nausea, and vomiting. Muscle twitching, jerking, and shaking may be mistaken for a seizure. Complications include tachycardia, arrhythmias, hypertension, hyperthermia, rhabdomyolysis, and acidosis. Symptoms progress to maximal severity in ~5 h and subside within a day or two, although pain and paresthesia can last for weeks. Fatal respiratory arrest is most common among young children and the elderly.

Envenomations by *Leiurus quinquestriatus* in the Middle East and North Africa, by *Mesobuthus tamulus* in India, by *Androctonus* species along the Mediterranean littoral and in North Africa and the Middle East, and by *Tityus serrulatus* in Brazil cause massive release of endogenous catecholamines with hypertensive crises, arrhythmias, pulmonary edema, and myocardial damage. Acute pancreatitis occurs with stings of *Tityus trinitatis* in Trinidad, and central nervous toxicity complicates stings of *Parabuthus* and *Buthotus* scorpions of South Africa.

In Iran and adjacent countries, *Hemiscorpius lepturus* causes the most scorpion envenomations. Its stings are relatively asymptomatic at first, but its cytotoxic venom causes pain, hemolysis, and tissue necrosis after the first day. Systemic complications include hemoglobinuria and subsequent acute kidney injury.

Stings of most other species cause immediate sharp local pain followed by edema, ecchymosis, and a burning sensation. Symptoms typically resolve within a few hours, and skin does not slough. Allergic reactions to the venom may occur.

TREATMENT
Scorpion Stings

Identification of the offending scorpion helps to determine the course of treatment. Stings of nonlethal species require at most ice packs, analgesics, or antihistamines. Because most victims experience only local discomfort, they can be managed at home with instructions to return to the emergency department if signs of cranial-nerve or neuromuscular dysfunction develop. Aggressive supportive care and judicious use of antivenom can reduce or eliminate deaths from more severe envenomations. Keeping the patient calm and applying pressure dressings and cold packs to the sting site are measures that decrease the absorption of venom. A continuous IV infusion of midazolam reduces the agitation and involuntary movements produced by scorpion stings. Treating people with neuromuscular symptoms with sedatives or opiates requires close monitoring due to potential respiratory compromise. Hypertension and pulmonary edema respond to nifedipine, nitroprusside, hydralazine, or prazosin. Dangerous bradydysrhythmia can be controlled with atropine.

Commercially prepared antivenins are available in several countries for some of the most dangerous scorpion species. The FDA has approved equine-derived *C. sculpturatus* IgG F(ab')$_2$ antivenin. IV administration of antivenin rapidly reverses cranial-nerve dysfunction and muscular symptoms.

■ HYMENOPTERA STINGS

Bees, wasps, hornets, yellow jackets, and ants (all of the insect order Hymenoptera) sting in defense or to subdue their prey. Their venoms contain a wide array of amines, peptides, and enzymes that cause local and systemic reactions. Although the toxic effect of multiple stings can be fatal to a human, nearly all of the ≥100 deaths due to hymenopteran stings in the United States each year result from type 1, immediate-type allergic reactions.

Bee and Wasp Stings The stinger of the honeybee (*Apis mellifera*) is unique in being barbed. The stinging apparatus and attached venom sac tear loose from the honeybee's body, and muscular contractions of the venom sac continue to infuse venom into the skin. Other kinds of bees, ants, and wasps have smooth stinging mechanisms and can sting numerous times in succession. Generally, a person sustains just one sting from a bee or social wasp unless a nest was disturbed. Africanized honeybees (now present in South and Central America and the southern and western United States) respond to minimal intrusions more aggressively. The sting of an Africanized bee contains less venom than that of its non-Africanized relatives, but victims tend to sustain far more stings and thus receive a far greater overall volume of venom. Most patients who report having sustained a "bee sting" are more likely to have encountered stinging wasps instead.

The venoms of different kinds of hymenopterans are biochemically and immunologically distinct. Direct toxic effects are mediated by mixtures of low-molecular-weight compounds such as serotonin, histamine, acetylcholine, and several kinins. Polypeptide toxins in honeybee venom include mellitin, which damages cell membranes; mast cell–degranulating protein, which causes histamine release; the neurotoxin apamin; and the anti-inflammatory compound adolapin. Enzymes in venom include hyaluronidase and phospholipases. There appears to be little cross-sensitization between the venoms of honeybees and wasps.

Uncomplicated hymenopteran stings cause immediate pain, a wheal-and-flare reaction, and local edema, all of which usually subside in a few hours. Multiple stings can lead to vomiting, diarrhea, generalized edema, dyspnea, hypotension, and nonanaphylactic circulatory collapse. Rhabdomyolysis and intravascular hemolysis may cause renal failure. Death from the direct (nonallergic) effects of venom has followed stings of several hundred honeybees. Stings to the tongue or mouth may induce life-threatening edema of the upper airways.

Large local reactions accompanied by erythema, edema, warmth, and tenderness that spread ≥10 cm around the sting site over 1–2 days are not uncommon. These reactions may resemble bacterial cellulitis but are caused by hypersensitivity rather than by secondary infection. Such reactions tend to recur on subsequent exposure but are seldom accompanied by anaphylaxis and are not prevented by venom immunotherapy.

An estimated 0.4–4.0% of the U.S. population exhibits clinical immediate-type hypersensitivity to hymenopteran stings, and 15% may have asymptomatic sensitization manifested by positive skin tests. Persons who experience severe allergic reactions are likely to have similar or more severe reactions after subsequent stings by the same or closely related species. Mild anaphylactic reactions to insect stings, as to other causes, consist of nausea, abdominal cramping, generalized urticaria or angioedema, and flushing. Serious reactions, including upper airway edema, bronchospasm, hypotension, and shock, may be rapidly fatal. Severe reactions usually begin within 10 min of the sting and only rarely develop after 5 h.

TREATMENT
Bee and Wasp Stings

Honeybee stingers embedded in the skin should be removed as soon as possible to limit the quantity of venom delivered. The stinger and venom sac may be scraped off with a blade, a fingernail, or the edge of a credit card or may be removed with forceps. The site should be cleansed and disinfected and ice packs applied to slow the spread of venom. Elevation of the affected site and administration of oral analgesics, oral antihistamines, and topical calamine lotion help relieve symptoms.

Anaphylactic reactions to bee or wasp venom can be a life-threatening emergency that requires prompt life-saving actions.

If the individual carries a bee-sting kit, then a subcutaneous injection of epinephrine hydrochloride (0.3 mL of a 1:1000 dilution) should be considered, with treatment repeated every 20–30 min as necessary. A tourniquet may slow the spread of venom. The patient should be transferred to a hospital emergency room where treatment for profound shock, if required, can be administered safely. Such treatment may entail the use of IV epinephrine and other vasopressors, intubation or provision of supplemental oxygen, fluid resuscitation, use of bronchodilators, and parenteral administration of antihistamines. Patients should be observed for 24 h for recurrent anaphylaxis, renal failure, or coagulopathy.

Persons with a history of allergy to insect stings should carry an anaphylaxis kit with a preloaded syringe containing epinephrine for self-administration. These patients should seek medical attention immediately after using the kit.

Prophylactic immunotherapy may greatly reduce the risk of life-threatening reactions to bee and wasp stings. Repeated injections of purified venom produce a blocking IgG antibody response to venom and reduce the incidence of recurrent anaphylaxis. Honeybee, wasp, and yellow jacket venoms are commercially available for desensitization and for skin testing. Results of skin tests and venom-specific radioallergosorbent tests (RASTs) aid in the selection of patients for immunotherapy and guide the design of such treatment.

STINGING ANTS

Stinging ants are an important medical problem in the United States. Imported fire ants (*Solenopsis* species) infest southern states from Texas to North Carolina, with colonies now established in California, New Mexico, Arizona, and Virginia. Slight disturbances of their mound nests have provoked massive outpourings of ants and as many as 10,000 stings on a single person. Elderly and immobile persons are at high risk for attacks when fire ants invade dwellings.

Fire ants attach to skin with powerful mandibles and rotate their bodies while repeatedly injecting venom with posteriorly situated stingers. The alkaloid venom consists of cytotoxic and hemolytic piperidines and several proteins with enzymatic activity. The initial wheal-and-flare reaction, burning, and itching resolve in ~30 min, and a sterile pustule develops within 24 h. The pustule ulcerates over the next 48 h and then heals in ≥1 week. Large areas of erythema and edema lasting several days are not uncommon and, in extreme cases, may compress nerves and blood vessels. Anaphylaxis occurs in <2% of victims; seizures and mononeuritis have been reported. Stings are treated with ice packs, topical glucocorticoids, and oral antihistamines. Pustules should be cleansed and then covered with bandages and antibiotic ointment to prevent bacterial infection. Epinephrine administration and supportive measures are indicated for anaphylactic reactions. Fire ant whole-body extracts are available for skin testing and immunotherapy, which appear to lower the rate of anaphylactic reactions.

European fire (red) ants (*Myrmica rubra*) have recently become public health pests in the northeastern United States and southern Canada. The western United States is home to harvester ants (*Pogonomyrmex* species). The painful local reaction that follows harvester ant stings often extends to lymph nodes and may be accompanied by anaphylaxis. The bullet or conga ant (*Paraponera clavata*) of South America is known locally as *hormiga veinticuatro* ("24-hour ant"), a designation that refers to the 24 h of throbbing, excruciating pain following a sting that delivers the potent paralyzing neurotoxin poneratoxin.

DIPTERAN (FLY AND MOSQUITO) BITES

In the process of feeding on vertebrate blood and tissue fluids, adults of certain fly species inflict painful bites, inject saliva that may cause vasodilation and produce local allergic reactions, and may transmit diverse pathogenic agents. Bites of mosquitoes (culicids), tiny "no-see-um" midges (ceratopogonids), and sand flies (phlebotomines) typically produce a wheal and a pruritic papule. Small humpbacked black flies (simuliids) lacerate skin, resulting in a lesion with serosanguineous discharge that is often painful and pruritic. Regional lymphadenopathy, fever, or anaphylaxis occasionally ensues. The widely distributed deerflies and horseflies as well as the tsetse flies of Africa are stout flies that attack during the day and produce large and painful bleeding punctures. House flies (*Musca domestica*) do not consume blood but use rasping mouthparts to scarify skin and feed upon tissue fluids and salt. Beyond direct injury from bites of any kind of fly, risks include transmission of diverse pathogens and secondary bacterial infection of the lesion.

TREATMENT

Fly and Mosquito Bites

Treatment of fly bites is symptom based. Topical application of antipruritic agents, glucocorticoids, or antiseptic lotions may relieve itching and pain. Allergic reactions may require oral antihistamines. Antibiotics may be necessary for the treatment of large bite wounds that become secondarily infected.

FLEA BITES

Common human-biting fleas include the dog and cat fleas (*Ctenocephalides* species) and the rat flea (*Xenopsylla cheopis*), which infest their respective hosts and their nests and resting sites. Sensitized persons develop erythematous pruritic papules (*papular urticaria*) and occasionally vesicles and bacterial superinfection at the site of the bite. Symptom-based treatment consists of antihistamines, topical glucocorticoids, and topical antipruritic agents.

Flea infestations are eliminated by removal and treatment of animal nests, frequent cleaning of pet bedding, and application of contact and systemic insecticides to pets and the dwelling. Flea infestations in the home may be abated or prevented if pets are regularly treated with veterinary antiparasitic agents, insect growth regulators, or chitin inhibitors.

Tunga penetrans, like other fleas, is a wingless, laterally flattened insect that feeds on blood. Also known as the chigoe flea, sand flea, or jigger (not to be confused with the chigger), it occurs in tropical regions of Africa and the Americas. Adult female chigoes live in sandy soil and burrow under the skin, usually between toes, under nails, or on the soles of bare feet. Gravid chigoes engorge on the host's blood and grow from pinpoint to pea size during a 2-week interval. They produce lesions that resemble a white pustule with a central black depression and that may be pruritic or painful. Occasional complications include tetanus, bacterial infections, and autoamputation of toes (*ainhum*). Tungiasis is treated by removal of the intact flea with a sterile needle or scalpel, tetanus vaccination, and topical application of antibiotics.

HEMIPTERAN/HETEROPTERAN (TRUE BUG) BITES

Most true bugs feed on plants, but some are predaceous or feed on blood. In order to feed or to defend themselves, they may inflict bites that produce allergic reactions and are sometimes painful. Bites of the cone-nose or "kissing bugs" (family Reduviidae) tend to occur at night and are painless. Reactions to such bites depend on prior sensitization and include tender and pruritic papules, vesicular or bullous lesions, extensive urticaria, fever, lymphadenopathy, and (rarely) anaphylaxis. Bug bites are treated with topical antipruritic agents or oral antihistamines. Persons with anaphylactic reactions to reduviid bites should keep an epinephrine kit available. Some reduviids transmit *Trypanosoma cruzi*, the agent of New World trypanosomiasis (Chagas disease) **(Chap. 234)**.

The cosmopolitan and tropical bed bugs (*Cimex lectularius* and *C. hemipterus*) hide in crevices of mattresses, bed frames and other furniture, walls, and picture frames and under loose wallpaper, actively seeking blood meals at night. These bugs are now a common pest in homes, dormitories, and hotels; on cruise ships; and even in medical facilities. Their bite is painless. Bites on persons without prior exposure to bedbugs may not be noticeable. Persons sensitized to bed bug saliva develop erythema, itching, and wheals around a central hemorrhagic punctum. Reactions may manifest within minutes of the bites, or they may be delayed for days or even a week or more. Bed bugs are not known to transmit pathogens.

CENTIPEDE BITES AND MILLIPEDE DERMATITIS

Two groups of myriapods ("many-footed" arthropods) can harm humans. Centipedes, with one pair of legs per body segment, are fast-moving, aggressive, and carnivorous. They stun and kill their prey—usually other arthropods, earthworms, and rarely small vertebrates—with a venomous bite. The fangs of centipedes of the genus *Scolopendra* can penetrate human skin and deliver a venom that produces intense burning pain, swelling, erythema, and sterile lymphangitis. Dizziness, nausea, and anxiety are described occasionally, and rhabdomyolysis and renal failure have been reported. Treatment includes washing of the site, application of cold dressings, oral analgesic administration or local lidocaine infiltration, and tetanus prophylaxis.

Millipedes, with two pairs of legs per segment, are slow-moving, docile, and feed mostly on decaying plant materials. They do not bite, but some secrete defensive fluids that may burn and discolor human skin. Affected skin turns brown overnight and may blister and exfoliate. Secretions in the eye cause intense pain and inflammation that can result in corneal ulcers and even blindness. Management includes irrigation with copious amounts of water or saline, use of analgesics, and local care of denuded skin.

CATERPILLAR STINGS AND DERMATITIS

Caterpillars of several moth species are covered with hairs or spines that produce mechanical irritation and may contain or be coated with venom. Contact with these caterpillars or their hairs may lead to erucism (a pruritic urticarial or papular rash) or caterpillar envenomation. The response typically consists of an immediate burning sensation followed by local swelling and erythema and occasionally by regional lymphadenopathy, nausea, vomiting, and headache. A rare reaction to a South American caterpillar, *Lonomia obliqua*, can cause disseminated coagulopathy and fatal hemorrhagic shock.

Dermatitis is most often associated with caterpillars of io, puss, saddleback, and browntail moths in North America and with the oak processionary moth in Europe. Even contact with detached hairs of other caterpillars, such as gypsy moth larvae, can later produce erucism. Spines may be deposited on tree trunks or drying laundry or may be airborne and cause irritation of the eyes and upper airways. Treatment of caterpillar stings consists of repeated application of adhesive or cellophane tape to remove the hairs, which can then be identified microscopically. Local ice packs, topical glucocorticoids, and oral antihistamines relieve symptoms.

Few adult moths cause human health problems. Adult yellowtail moths (*Hylesia* species), found mainly in coastal mangrove zones along the eastern coast of Central and South America, have bodies that are covered by fine hairs or setae. The hairs on the ventral surface can detach and, when in contact with human skin, cause an extremely pruritic reaction called "Carapito itch." This issue is especially problematic when the moths have population booms, creating swarms around coastal communities.

BEETLE VESICATION AND DERMATITIS

Several families of beetles have independently developed the ability to produce chemically unrelated vesicating toxins. When disturbed, blister beetles (family Meloidae) exude cantharidin, a low-molecular-weight toxin that produces thin-walled blisters (≤5 cm in diameter) 2–5 h after contact. The blisters are not painful or pruritic unless broken. They resolve without treatment in ≤10 days. Nephritis may follow unusually heavy cantharidin exposure.

The hemolymph of certain rove beetles (*Paederus* species, Staphylinidae family) contains pederin, a potent vesicant. When these beetles are crushed or brushed against the skin, the released fluid causes painful, red, flaccid bullae. These beetles occur worldwide but are most numerous and problematic in parts of Africa (where they are called "Nairobi fly") and southwestern Asia. Ocular lesions may develop after impact with flying beetles at night or unintentional transfer of the vesicant on the fingers. Treatment is rarely necessary, although ruptured blisters should be kept clean and bandaged.

Larvae of common carpet beetles are adorned with dense arrays of ornate hairs called *hastisetae*. Contact with these larvae or their setae results in delayed dermal reactions in sensitized individuals. The lesions are commonly mistaken for bites of bed bugs.

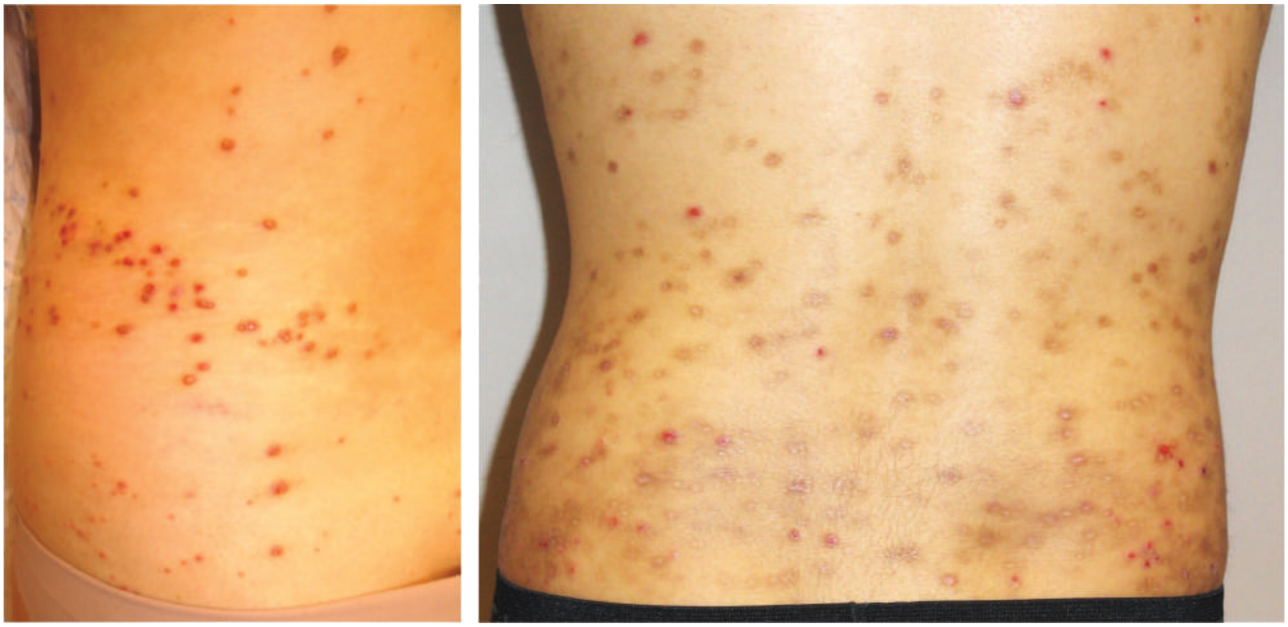

FIGURE 472-4 Real (*left*) versus delusional (*right*) infestation: comparable images of the lower backs of two young adults with multiple lesions. ***Left:*** A young woman developed innumerable widespread lesions during a camping ecotour near Manaus, Brazil. Note scattered clusters of irregularly spaced lesions, accompanied by dozens of single or isolated lesions, consistent with the semi-random feeding pattern of biting flies. Lesions appear to be in roughly the same stage of development, a feature indicating that they were acquired at roughly the same time. No lesions were present before her ecotour; none have arisen since. This patient scratches the intensely pruritic lesions and causes superficial erosions. Unexcoriated lesions are also present on her midback, where she cannot scratch. ***Right:*** A young man has innumerable widespread lesions that have accumulated for several years, with a few new lesions appearing several times a week. His lesions are in various stages of development (fresh, crusted, reepithelialized, pigmented, and scarred), a feature indicating a long-standing process. The lesions are distributed in a regular pattern consistent with periodic "excavations" to remove alleged parasites that he believes are crawling through his skin. Scarring is due to manipulations that create dermal ulcers rather than superficial excoriations and erosions. Parts of his upper midback, where he cannot scratch, are free of lesions.

DELUSIONAL INFESTATIONS

The groundless conviction that one is infested with arthropods or other parasites (Ekbom syndrome, delusory parasitosis, delusions of parasitosis, delusion of infestation, and perhaps Morgellons syndrome) is extremely difficult to treat and, unfortunately, is not uncommon **(Fig. 472-4)**. Patients describe uncomfortable sensations of something moving in or on their skin. Excoriations and self-induced ulcerations typically accompany the pruritus, dysesthesias, and imaginary insect bites. Patients often believe that some invisible or as-yet-undescribed creatures are infesting their skin, clothing, homes, or environment in general. Frequently, patients submit as evidence of infestation specimens that consist of plant-feeding and nonbiting peridomestic arthropods, pieces of skin, vegetable matter, lint, and other inanimate detritus. In the evaluation of a patient with possible delusional parasitosis, it is imperative to rule out true infestations and bites by arthropods, endocrinopathies, sensory disorders due to neuropathies, opiate and other drug use, environmental irritants (e.g., fiberglass threads), and other causes of tingling or prickling sensations. Frequently, such patients repeatedly seek medical consultations, resist alternative explanations for their symptoms, and exacerbate their discomfort by self-treatment. Recent evidence suggests that high levels of central nervous system dopamine, endogenous or pharmaceutical, may promote these delusions. Long-term pharmacotherapy with pimozide or other psychotropic agents has been more helpful than psychotherapy in treating this disorder. Patients with delusory parasitosis often develop the unshakeable conviction that they are infested by a previously unknown pathogen, while their personal lives, family support, and employment collapse around them.

ACKNOWLEDGMENT

Richard J. Pollack contributed to this chapter in previous editions and some material from that chapter has been retained here.

FURTHER READING

Amanzougaghene N et al: Where are we with human lice? A review of the current state of knowledge. Front Cell Infect Microbiol 21:9, 2020

Arlia LG, Morgan MS: A review of *Sarcoptes scabiei*: Past, present and future. Parasit Vectors 10:297, 2017.

Goddard J: *Infectious Diseases and Arthropods*, 3rd ed. Totowa, NJ, Humana Press, 2018.

Hinkle N: Ekbom syndrome: The challenge of "invisible bug" infestations. Annu Rev Entomol 55:77, 2010.

Moraru GM, Goddard J II: *The Goddard Guide to Arthropods of Medical Importance*, 7th ed. Boca Raton, FL, CRC Press, 2019.

Mullen G, Durden L: *Medical and Veterinary Entomology*, 3rd ed. London, Academic Press, 2019.

Saucier JR: Arachnid envenomation. Emerg Med Clin North Am 22:405, 2004.

Steen CJ et al: Insect sting reactions to bees, wasps, and ants. Int J Dermatol 44:91, 2005.

Thomas C et al: Ectoparasites: Scabies. J Am Acad Dermatol 82:533, 2020.

Vetter RS, Isbister GK: Medical aspects of spider bites. Annu Rev Entomol 53:409, 2008.

473 Health Effects of Climate Change

Eugene Richardson, Maxine A. Burkett

Climate change refers to the effects of accumulated greenhouse gases (GHGs) in the atmosphere on long-term weather patterns. Anthropogenic emissions—in particular from the burning of fossil fuels and land conversion—have increased mean global temperatures by approximately 1.1° Celsius above preindustrial levels. Extreme weather events, including heatwaves and 'natural disasters' (e.g., wildfires, droughts, and floods) are becoming more frequent and lead to resource scarcity (including access to safe drinking water and food), increased environmental pollution and degradation, violent conflict, and precarious migration. The climate crisis thus has direct consequences for human health **(Fig. 473-1)**, the practice of medicine, and the stability of health care systems and as such represents a health emergency. **See Chap. 474 for an overview of climate science.**

EFFECTS OF CLIMATE CHANGE ON HEALTH

The World Health Organization predicts that between 2030 and 2050, there will be an additional 250,000 deaths from undernutrition, malaria, diarrhea, and heat stress alone **(Chap. 474)**. As with much of the global burden of disease, this increase in mortality will be disproportionately borne by low- and middle-income countries (LMICs) whose health infrastructures have been weakened by neocolonialism and structural adjustment. Within wealthier countries such as the United States, climate change will amplify existing health disparities between white people and black, Indigenous, and Latinx populations.

AIR POLLUTION AND SYNERGISTIC EFFECTS

Asthma and Other Respiratory Ailments Climate change exacerbates the negative health effects of harmful air pollutants (e.g., particulate matter, ozone, sulfur dioxide, and nitrogen dioxide). While increases in fine particulate matter (<2.5 microns—$PM_{2.5}$) are a function of wildfires and the burning of fossil fuels, the latter is the predominate driver of anthropogenic climate change. Exposure to ambient air pollution and/or indoor air pollution is estimated to cause 7 million premature deaths worldwide per year, making it the largest global environmental risk factor for reversible death and disability (see **Chap. 300** for an overview). Notably, air pollution from coal-burning power plants is associated with double the mortality risk when compared with other sources of $PM_{2.5}$.

By increasing ground-level ozone and/or particulate matter concentrations in some regions, the higher temperatures associated with climate change will directly increase the global burden and severity of asthma **(Chap. 298)**, the respiratory effects of allergies **(Chap. 363)**, rhinosinusitis, chronic obstructive pulmonary disease (COPD), respiratory tract infections, interstitial lung disease, and lung cancer, resulting in increased hospital admissions and premature death. **Figure 473-2** illustrates potential pathophysiologic mechanisms by which this may occur.

Cardiovascular Disease Cardiovascular (CV) complications of climate change share similar mechanisms with climate-sensitive respiratory disease. Concentrations of $PM_{2.5}$ are the most important environmental risk factor for myocardial infarction, cerebrovascular disease, heart failure, hypertension, diabetes mellitus, arrhythmias, and venous thromboembolism.

Studies have shown that the relative risk of acute CV events is increased 1–3% in the setting of short-term elevations of $PM_{2.5}$. Longer-term exposures convey an amplified risk (~10%), which is partially attributable to the exacerbation of chronic conditions (e.g., hypertension and diabetes). This holds true for areas with low mean

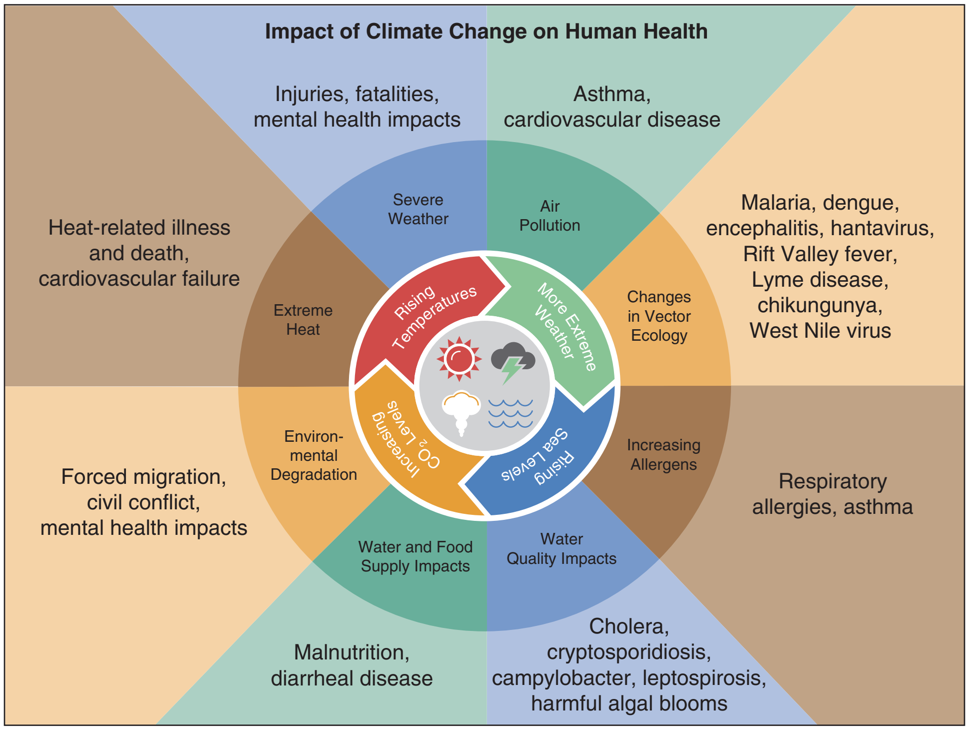
Impact of Climate Change on Human Health
Injuries, fatalities, mental health impacts
Asthma, cardiovascular disease
Heat-related illness and death, cardiovascular failure
Malaria, dengue, encephalitis, hantavirus, Rift Valley fever, Lyme disease, chikungunya, West Nile virus
Severe Weather
Air Pollution
Extreme Heat
Changes in Vector Ecology
Rising Temperatures
More Extreme Weather
Increasing CO2 Levels
Rising Sea Levels
Environmental Degradation
Increasing Allergens
Forced migration, civil conflict, mental health impacts
Respiratory allergies, asthma
Water and Food Supply Impacts
Water Quality Impacts
Malnutrition, diarrheal disease
Cholera, cryptosporidiosis, campylobacter, leptospirosis, harmful algal blooms

FIGURE 473-1 Climate change impacts a wide range of health outcomes. This figure illustrates the most significant climate change impacts (rising temperatures, more extreme weather, rising sea levels, and increasing carbon dioxide levels), their effect on exposures, and the subsequent health outcomes that can result from these changes in exposures. *(Source: https://www.cdc.gov/climateandhealth/effects/default.htm.)*

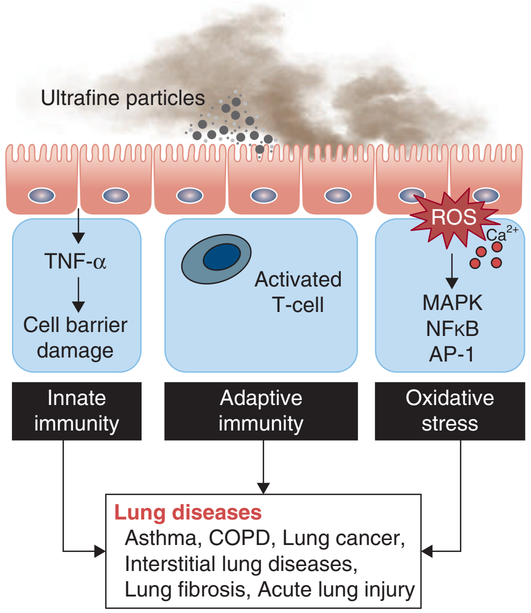

FIGURE 473-2 Mechanisms of ultrafine particle–induced respiratory health effects. *(Reproduced from GD Leikauf et al: Mechanisms of ultrafine particle-induced respiratory health effects. Exp Mol Med 52:329, 2020.)*

concentrations of $PM_{2.5}$ (i.e., where risk is determined by recurring short-term elevations). The biologic pathways whereby $PM_{2.5}$ promotes these complications are complex and multifactorial **(Fig. 473-3)**.

For clinicians, there are subtle management strategies to bear in mind for climate-sensitive CV disease. For example, in new heart failure patients, prescribers should be judicious about starting diuretics (especially diuretic-angiotensin–converting enzyme [ACE] inhibitor combinations) before the summer months or in areas with increased heatwaves in order to avoid exacerbating dehydration or heat-related illness. These patients may also benefit from increasing potassium uptake. **Figure 473-4** presents personal- and local-level interventions to reduce climate-sensitive disease associated with air pollution.

Pregnancy Women exposed to high temperatures and air pollution may be more likely to experience serious adverse pregnancy outcomes. A systematic review of studies across diverse U.S. populations found a statistically significant association of $PM_{2.5}$, ozone, and heat exposure with preterm birth, low birth weight at term, and stillbirth. As these environmental exposures become more common with climate change, an increased incidence of these complications is likely.

Potential pathophysiologic mechanisms for these outcomes are multifactorial: preterm birth may result from hematogenous transport of inhaled $PM_{2.5}$ and varied noxious chemicals and subsequent systemic inflammation or perturbation of the autonomic nervous system; low birth weight could be caused by the cumulative effect of alterations in maternal cardiac, pulmonary, and renal function, placental inflammation, and direct exposure to oxidative stress; and stillbirth may involve derangements in oxygen transport, DNA damage, and direct placental injury. Extreme heat events may lead to adverse pregnancy outcomes through dehydration and subsequent alterations in thermoregulation, blood viscosity, uterine blood flow, placental-fetal exchange, amniotic fluid volume, and hormone release (e.g., prostaglandin or oxytocin).

These risks are disproportionately borne by black mothers. Thus, failure to reduce air pollution to the World Health Organization guideline level of 10 μg/m^3 compounds structural racism.

HEAT-RELATED ILLNESSES

Renal Disease Various populations of agricultural workers who labor in hot climates around the world have been noted to suffer from chronic kidney disease (CKD), even those without common risk factors such as diabetes mellitus, hypertension, glomerular disease, or HIV. Although the cause has not been identified, potential mechanisms include genetic polymorphisms, nephrotoxicity secondary to agrochemicals or heavy metals, and heat-associated injury **(Fig. 473-5)**.

Heat Exhaustion and Heatstroke Please see **Chap. 478** for an overview of heat-related illnesses. These include heat cramps, heat exhaustion, and heatstroke and can be expected to increase in incidence as temperatures rise. Certain communities suffer from significant peaks in average temperature resulting from the urban heat island effect, which in the United States is related to policies such as redlining, racial covenants, and strategic underinvestment in neighborhoods segregated through such policies. Possible preventive measures include developing clear heatwave strategies and establishing early warning systems, mapping vulnerable populations, enhancing green space, and providing cool-down zones. It is important that clinicians provide guidance about heat-related illnesses prior to the start of summer because waiting for heat warnings will not obviate risk. Medication storage is an issue, especially for those that are carried by patients (e.g., epinephrine injections, naloxone, insulin). There are some data to suggest that exposure to extreme heat (e.g., leaving an albuterol inhaler in a vehicle on a hot day) can impair delivery mechanisms, degrade active ingredients of the medication, or cause inhalers to explode.

NATURAL DISASTERS, COASTAL FLOODING, AND DISPLACEMENT

Injury and Trauma While models predict fewer cyclones and hurricanes in a warmer late-twenty-first-century climate, they do forecast events of higher average intensity, precipitation, and the number and occurrence days of very intense category 4 and 5 storms **(Fig. 473-6)**. Such natural disasters would be expected to result in injuries and trauma seen in contemporary storms of high intensity (albeit with higher frequency), but there are other ways climate change will affect the incidence of physical trauma.

Studies have shown associations between anomalously warm temperatures and increased deaths from drownings (from people swimming longer and more frequently), transportation accidents (driving performance worsens at higher temperatures and traffic increases), and assaults (potentially from increased alcohol consumption and illicit drug use). These deaths were partly offset by decreased falls, especially among the elderly, in warmer years.

In addition, as more than half the world's population lives within 60 km of the ocean, rising sea levels will destroy homes, medical infrastructure, and other essential services, including sewage treatment systems and drinking water supplies. In concert with climate impacts such as extreme heat and lack of freshwater, subsequent displacement and migration will lead to increased mental illness, food insecurity, and communicable disease.

Finally, extreme sudden-onset events can affect the availability of medications through medical supply chain disruptions. Loss of electricity during intense storms can also compromise vaccination programs and the availability of needed medications. For example, in 2024, Hurricane Helene interrupted the production of essential parenteral drug products and intravenous (IV) fluids manufactured in North Carolina, resulting in nationwide shortages in the United States. Providers should ensure that patients with electricity-dependent medical devices have a reasonable contingency plan in case of power outages while noting that climate-related displacements can interrupt patients' access to medications for chronic diseases and limit access to clinical care in general.

Mental Illness As of 2022, an estimated 108.4 million people (i.e., 1.4% of humanity) were forcibly displaced from their homes, mainly on account of violent conflict. It is expected that millions more will be displaced in response to the effects of climate change. This has ramifications for mental health (as do natural disasters) where resultant psychosocial stress can lead to an increased incidence of anxiety, depression, and posttraumatic stress disorder (PTSD). Increased temperatures have also been found to be associated with higher rates of suicide and domestic violence. Lastly, regarding clinical practice, many psychoactive prescription drugs can interfere with thermoregulation; their use therefore confers added risk during extreme heat events.

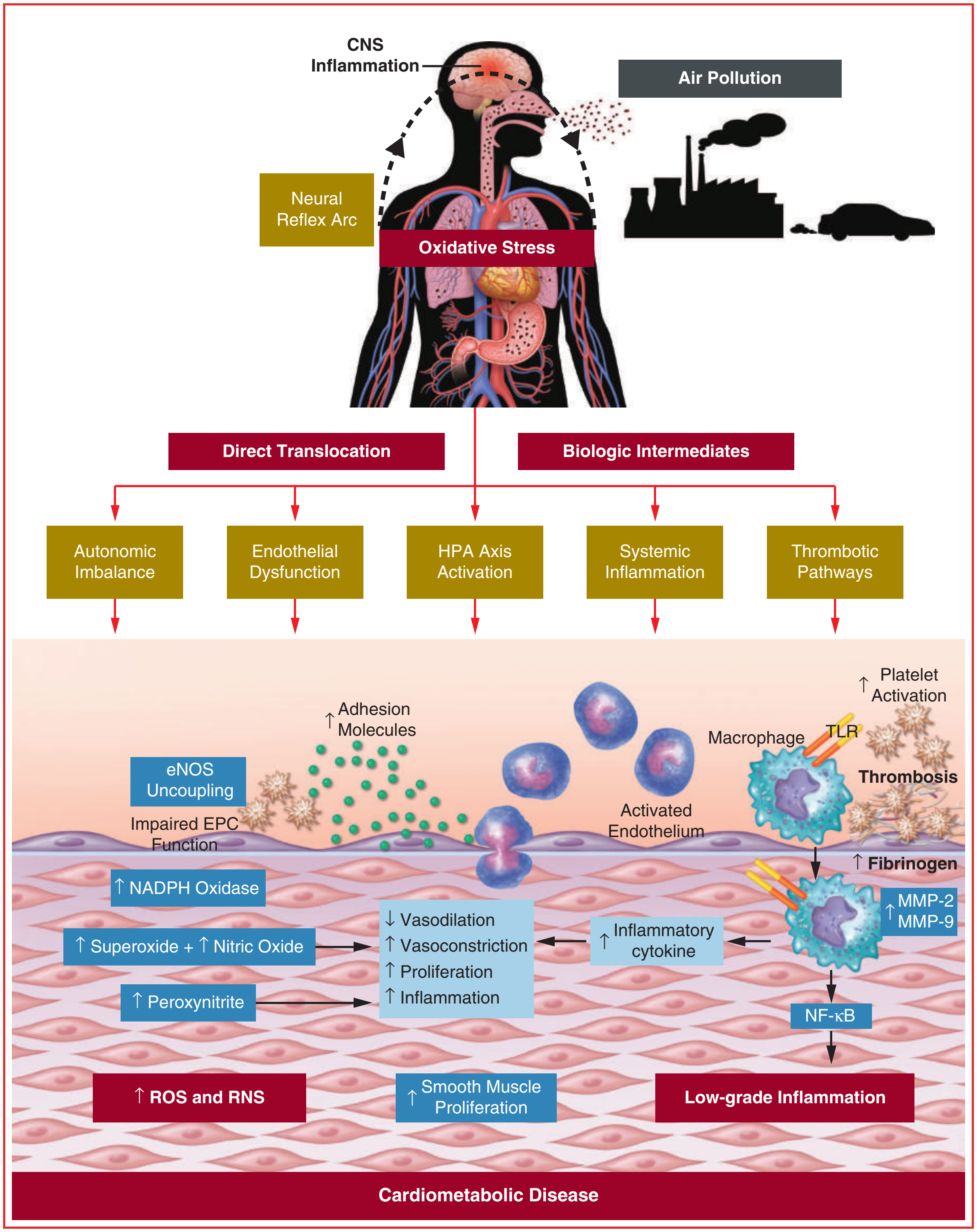
CNS
Inflammation
Air Pollution
Neural
Reflex Arc
Oxidative Stress
Direct Translocation
Biologic Intermediates
Autonomic
Imbalance
Endothelial
Dysfunction
HPA Axis
Activation
Systemic
Inflammation
Thrombotic
Pathways
↑ Platelet
Activation
↑ Adhesion
Molecules
Macrophage
TLR
Thrombosis
eNOS
Uncoupling
Impaired EPC
Function
Activated
Endothelium
↑ Fibrinogen
↑ NADPH Oxidase
↑ MMP-2
MMP-9
↓ Vasodilation
↑ Vasoconstriction
↑ Proliferation
↑ Inflammation
↑ Superoxide + ↑ Nitric Oxide
↑ Inflammatory
cytokine
↑ Peroxynitrite
NF-κB
↑ ROS and RNS
↑ Smooth Muscle
Proliferation
Low-grade Inflammation
Cardiometabolic Disease

FIGURE 473-3 Biologic pathways whereby PM$_{2.5}$ promotes cardiovascular events. *(Reproduced with permission from S Rajagopalan: Air pollution and cardiovascular disease: JACC state-of-the-art review. J Am Coll Cardiol 72:2054, 2018.)*

■ FOOD SECURITY AND OCEAN RESOURCES

Malnutrition Climate change threatens food and nutritional security through decreased crop yields in the setting of changes in precipitation, desertification, severe weather events, temperature extremes, GHG emissions, ocean warming and acidification, coastal inundation affecting dryland agriculture and aquaculture, and increasing competition from weeds and pests **(Fig. 473-7)**. Livestock and fish production are also projected to decline. For many, food will be less available, less nutritious, and more expensive.

As chronic noncommunicable diseases, both undernutrition **(Chap. 345)** and obesity **(Chap. 413)**—which affect nearly 3 billion people worldwide—share common drivers with climate change. Current dietary practices (through land conversion and overconsumption of ruminant meats) contribute to excess GHGs, decreases in biodiversity, and depletion of water supplies. Variable rainfall and increased flooding, which can contaminate freshwater supplies, will make water security a defining challenge of this century.

Plant-based diets have the potential to decrease GHG emissions; improve food security in LMICs; reduce mortality from stroke, type 2 diabetes mellitus, coronary heart disease, and cancer by 6–10%; and reduce diet-related GHGs by 29–70% by 2050 compared with a reference diet **(Fig. 473-8)**.

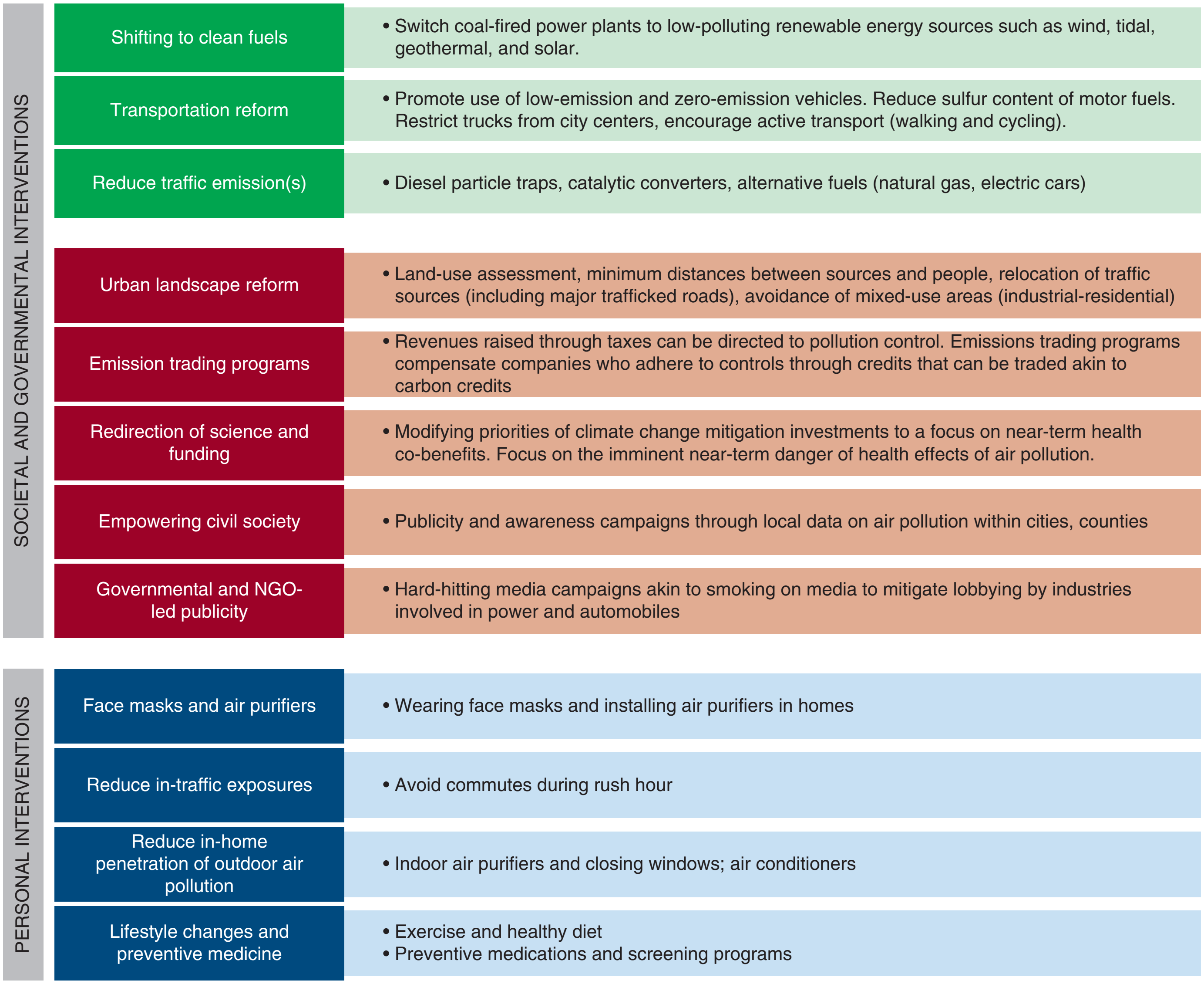

SOCIETAL AND GOVERNMENTAL INTERVENTIONS	Shifting to clean fuels	• Switch coal-fired power plants to low-polluting renewable energy sources such as wind, tidal, geothermal, and solar.
	Transportation reform	• Promote use of low-emission and zero-emission vehicles. Reduce sulfur content of motor fuels. Restrict trucks from city centers, encourage active transport (walking and cycling).
	Reduce traffic emission(s)	• Diesel particle traps, catalytic converters, alternative fuels (natural gas, electric cars)
	Urban landscape reform	• Land-use assessment, minimum distances between sources and people, relocation of traffic sources (including major trafficked roads), avoidance of mixed-use areas (industrial-residential)
	Emission trading programs	• Revenues raised through taxes can be directed to pollution control. Emissions trading programs compensate companies who adhere to controls through credits that can be traded akin to carbon credits
	Redirection of science and funding	• Modifying priorities of climate change mitigation investments to a focus on near-term health co-benefits. Focus on the imminent near-term danger of health effects of air pollution.
	Empowering civil society	• Publicity and awareness campaigns through local data on air pollution within cities, counties
	Governmental and NGO-led publicity	• Hard-hitting media campaigns akin to smoking on media to mitigate lobbying by industries involved in power and automobiles
PERSONAL INTERVENTIONS	Face masks and air purifiers	• Wearing face masks and installing air purifiers in homes
	Reduce in-traffic exposures	• Avoid commutes during rush hour
	Reduce in-home penetration of outdoor air pollution	• Indoor air purifiers and closing windows; air conditioners
	Lifestyle changes and preventive medicine	• Exercise and healthy diet • Preventive medications and screening programs

FIGURE 473-4 Social ecological interventions to reduce exposures or susceptibility to air pollution. *(Reproduced with permission from S Rajagopalan: Air pollution and cardiovascular disease: JACC state-of-the-art review. J Am Coll Cardiol 72:2054, 2018.)*

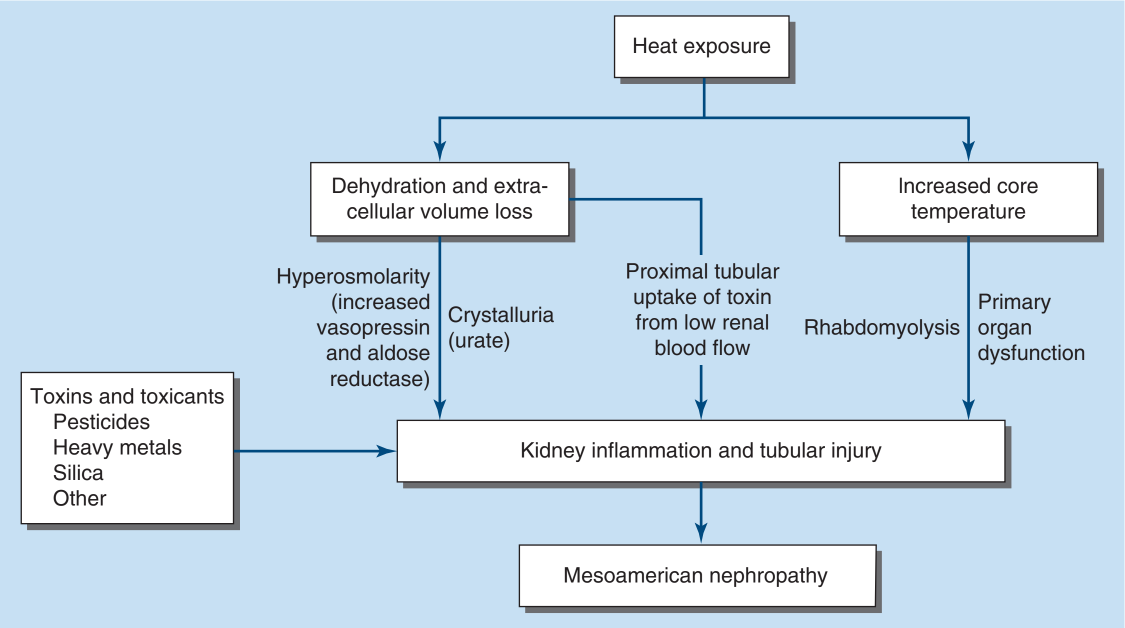
Heat exposure
Dehydration and extra-cellular volume loss
Increased core temperature
Hyperosmolarity (increased vasopressin and aldose reductase)
Crystalluria (urate)
Proximal tubular uptake of toxin from low renal blood flow
Rhabdomyolysis
Primary organ dysfunction
Toxins and toxicants
Pesticides
Heavy metals
Silica
Other
Kidney inflammation and tubular injury
Mesoamerican nephropathy

FIGURE 473-5 Possible mechanisms for the development of chronic kidney disease of unknown cause in agricultural communities. *(Adapted from RJ Johnson et al: Chronic kidney disease of unknown cause in agricultural communities. N Engl J Med 380:1843, 2019. Copyright © 2019 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

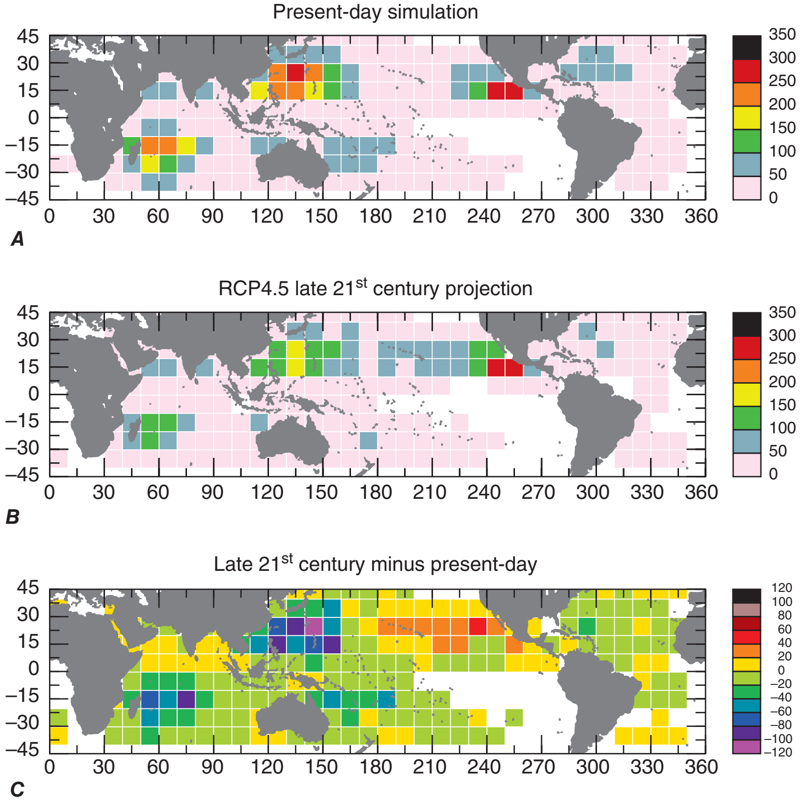
Present-day simulation
45
30
15
0
–15
–30
–45
0 30 60 90 120 150 180 210 240 270 300 330 360
350 300 250 200 150 100 50 0
A
RCP4.5 late 21st century projection
B
Late 21st century minus present-day
120 100 80 60 40 20 0 –20 –40 –60 –80 –100 –120
C

FIGURE 473-6 Simulated occurrence of all tropical storms (tropical cyclones with winds exceeding $17.5\ \mathrm{m\ s^{-1}}$) for (***A***) present-day or (***B***) late-twenty-first-century (RCP4.5; CMIP5 multimodel ensemble) conditions; unit: storms per decade. Simulated tropical cyclone tracks were obtained using the Geophysical Fluid Dynamics Laboratory (GFDL) hurricane model to resimulate (at higher resolution) the tropical cyclone cases originally obtained from the HiRAM C180 global mode. Occurrence refers to the number of days, over a 20-year period, in which a storm exceeding $17.5\ \mathrm{m\ s^{-1}}$ intensity was centered within the 10° × 10° grid region. (***C***) Difference in occurrence rate between late-twenty-first century and present day [(***B***) minus (***A***)]. *White regions* are regions where no tropical storms occurred in the simulations [in (***A***) and (***B***)] or where the difference between the experiments is zero [in (***C***)]. *(From TR Knutson et al: Global projections of intense tropical cyclone activity for the late twenty-first century from dynamical downscaling of CMIP5/RCP4.5 scenarios. J Clim 28:7203, 2015 © American Meteorological Society. Used with permission.)*

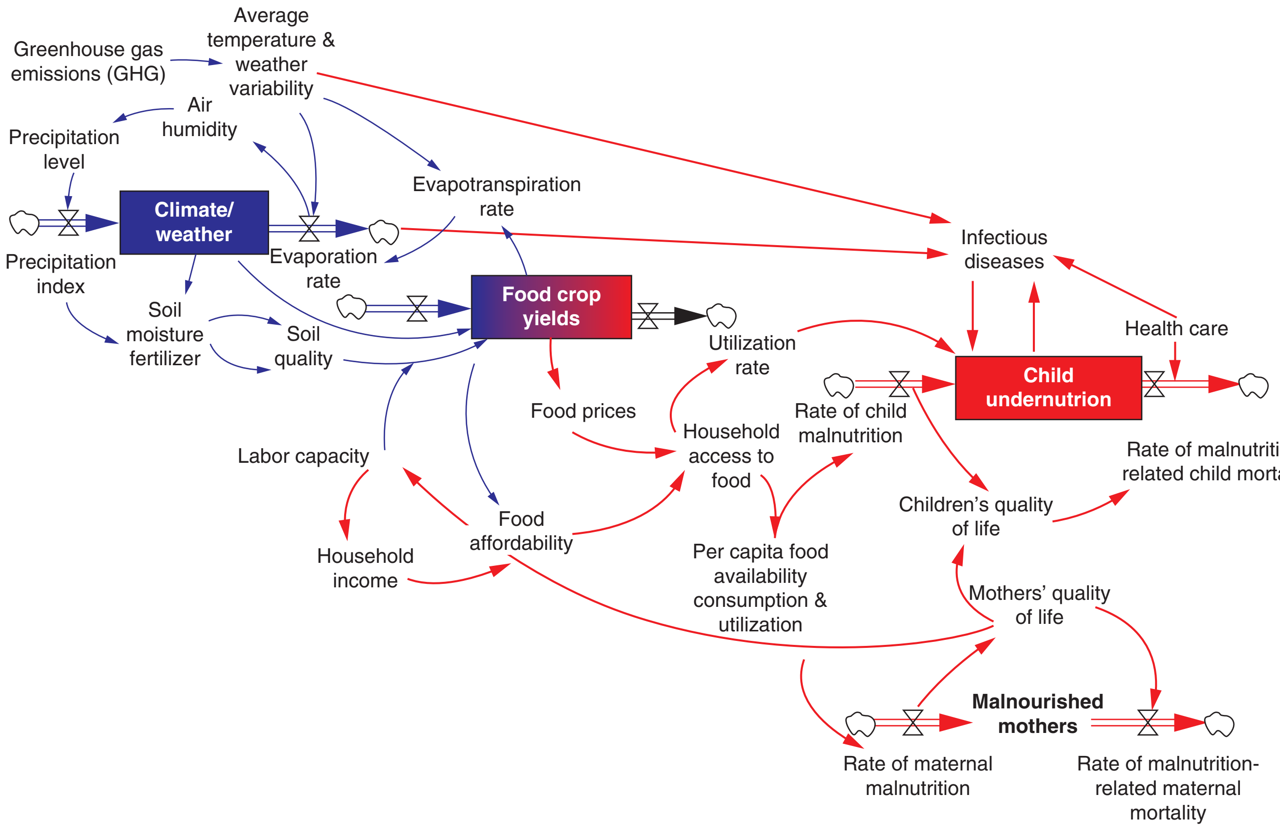
Greenhouse gas emissions (GHG)
Average temperature & weather variability
Air humidity
Precipitation level
Evapotranspiration rate
Climate/ weather
Precipitation index
Evaporation rate
Infectious diseases
Soil moisture fertilizer
Soil quality
Food crop yields
Utilization rate
Health care
Child undernutrition
Rate of child malnutrition
Food prices
Household access to food
Labor capacity
Rate of malnutrition-related child mortality
Children's quality of life
Food affordability
Household income
Per capita food availability consumption & utilization
Mothers' quality of life
Malnourished mothers
Rate of maternal malnutrition
Rate of malnutrition-related maternal mortality

FIGURE 473-7 Complex pathways from climate/weather variability to undernutrition in subsistence farming households. The factors involved in and the probable impacts of weather variables on crop yields (*blue arrows*) and of food crop yields on undernutrition (*red arrows*). *(Reproduced with permission from RK Phalkey et al: Climate change impacts on childhood undernutrition. Proc Natl Acad Sci USA 112:E4522, 2015.)*

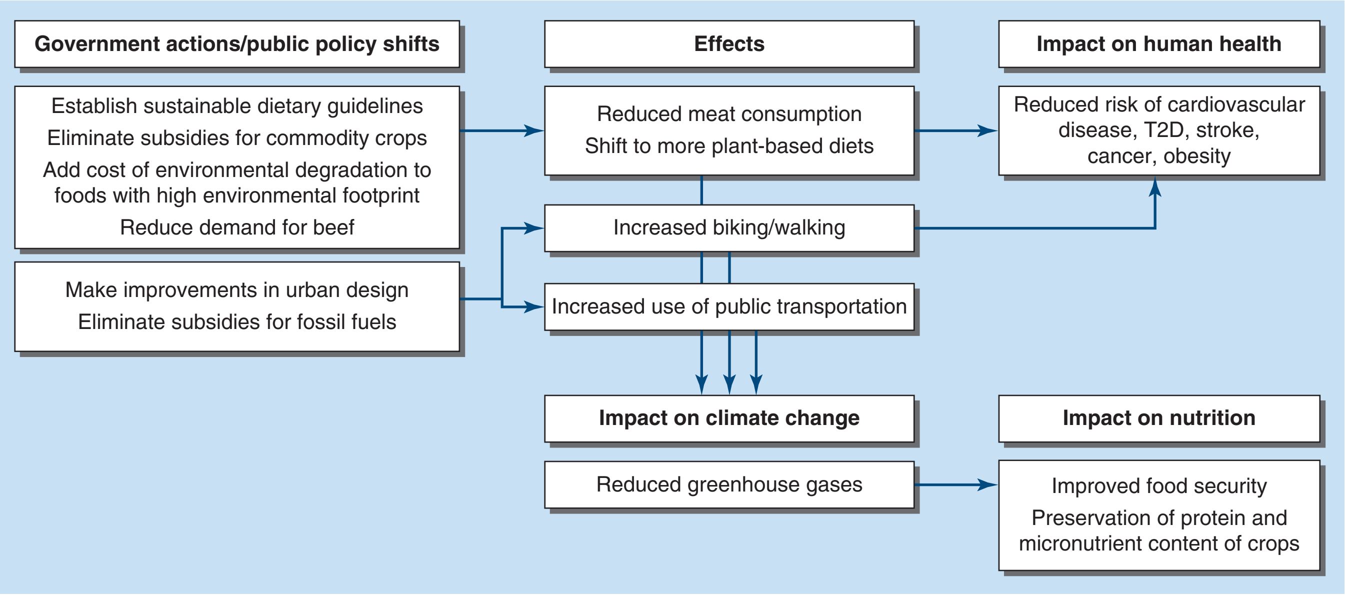
Government actions/public policy shifts
Establish sustainable dietary guidelines
Eliminate subsidies for commodity crops
Add cost of environmental degradation to foods with high environmental footprint
Reduce demand for beef
Make improvements in urban design
Eliminate subsidies for fossil fuels
Effects
Reduced meat consumption
Shift to more plant-based diets
Increased biking/walking
Increased use of public transportation
Impact on human health
Reduced risk of cardiovascular disease, T2D, stroke, cancer, obesity
Impact on climate change
Reduced greenhouse gases
Impact on nutrition
Improved food security
Preservation of protein and micronutrient content of crops

FIGURE 473-8 Strategies to mitigate the impact of climate change on human nutrition. *(Reproduced with permission of Journal of Clinical Investigation from Climate change and malnutrition: we need to act now. WH Dietz 130:556, 2020.)*

INFECTIONS AND DIARRHEAL DISEASE

(See Chap. 474.)

POTENTIAL SOLUTIONS

Reductions in GHG emissions will require health professionals to voice their evidence-based understandings of climate-sensitive pathologies to lobby for political action. Other important systemic interventions in health care include achieving universal health coverage (including financial risk protection); equitable access to quality essential health care services as well as safe, effective, and affordable medicines and vaccines; making health care and health systems net-zero carbon; adding a climate-change lens to existing lines of research; improving data quality and enhancing, standardizing, and integrating data collection in LMICs; and anticipating and correcting disaster-related health care system failures, such as impacts to supply chains or loss of electric power resulting from extreme weather events.

Interventions related to environmental policy include advocating for a tighter particulate-matter air-quality standard, supporting institutional divestment from fossil fuels, and advocating for the rapid drawdown of emissions and negative emissions strategies. Ecosocial interventions, supported by national or global institutions, include the distribution of clean cookstoves globally, switching to plant-based diets, decreased air travel, reducing air conditioner use, and increased access to more public transportation. Finally, advocating for wealth-redistribution schemes (e.g., reparations, progressive taxation, debt cancellation, improved safety nets, underemployment insurance) to empower disadvantaged populations to cope with climate hazards will have positive ancillary effects on the social determinants of health, the administration of health services, and the outcomes of clinical interventions.

CONCLUSIONS

Without sweeping reductions in GHG emissions, over the next 50 to 100 years, models predict increases in average global temperature of 2–5°C (with localized highs), rising sea levels, and more frequent and severe extreme-weather events, with resultant complications for population health globally. The hostile consequences of climate change will disproportionately affect vulnerable and marginalized groups, particularly those whose ability to cope with climate hazards is curtailed by systemic racism, colonial legacies, illicit financial flows, and human rights failings.

Health care professionals find themselves on the front line of the climate crisis and remain, in many settings, sources of information and counsel. In order to mitigate the impact of climate-sensitive diseases and resulting health disparities, they must continue to extend their clinical purview to socioecological determinants and structural interventions.

Acknowledgment

Paul Farmer contributed to this chapter in the 21st edition and some material from that chapter has been retained here.

FURTHER READING

Bekkar B et al: Association of air pollution and heat exposure with preterm birth, low birth weight, and stillbirth in the US: A systematic review. JAMA Netw Open 3:e208243, 2020.

Brief of *Amici Curiae* Public Health Experts, Public Health Organization, and Doctors In Support of Plaintiffs-Appellees' Petition for Hearing En Banc, *Juliana v. United States of America*, No. 18-36082 (9th Cir. Mar. 13, 2020).

Centers for Disease Control and Prevention: Climate Effects on Health. Available from *https://www.cdc.gov/climateandhealth/effects/default.htm*. Accessed January 19, 2024.

Hoffman JS et al: The effects of historical housing policies on resident exposure to intra-urban heat: A study of 108 US urban areas. Climate 8:12, 2020.

The Lancet: Countdown on health and climate. Available at *https://www.thelancet.com/countdown-health-climate-* Accessed January 19, 2024.

The New England Journal of Medicine: Climate Change. Available from *https://www.nejm.org/climate-crisis* Accessed January 19, 2024.

Rajagopalan S et al: Air pollution and cardiovascular disease: JACC state-of-the-art review. J Am Coll Cardiol 72:2054, 2018.

474

Climate Change and Infectious Disease

Aaron S. Bernstein, Jonathan A. Patz

Since the late nineteenth century, humans have released greenhouse gases—mainly carbon dioxide and methane—into the atmosphere, creating a new climate unseen in human times. This new climate has already altered the epidemiology of infectious diseases, and the accumulation of more greenhouse gases in the atmosphere will further alter the incidence and severity of infections. In certain instances, climate change may establish conditions promoting the emergence of infectious diseases, while in other instances, it may render areas that are presently suitable for certain diseases unsuitable. This chapter presents the current state of knowledge regarding the known and prospective infectious-disease consequences of climate change.

OVERVIEW

The term *climate change* refers to multidecadal alterations in temperature, precipitation, wind, humidity, and other components of weather outside of the natural climate variability seen in comparable time periods. Over the past 2.5 million years, the earth has warmed and cooled, cycling between glacial and interglacial periods during which average global temperatures moved up and down by 4–7°C. During the last glacial period, which ended roughly 12,000 years ago, global temperatures were, on average, 5°C cooler than in the mid-twentieth century **(Fig. 474-1)**.

The present climate period, known as the Holocene, is remarkable for its stability: temperatures have largely remained within a range of 2–3°C. This stability has enabled the successful population and cultivation of much of the earth's landmass by humanity. Current climate change differs from that in the past not only because its primary cause is human activities but also because its pace is faster. The current rate of warming on Earth is unprecedented in the last 50 million years. The 5°C of warming that occurred at the end of the last ice age about 12,000 years ago took roughly 5000 years, whereas such a temperature increment may occur within the next 150 years unless the release of greenhouse gases is substantially reduced in coming decades.

■ GREENHOUSE GASES

Greenhouse gases **(Table 474-1)** are a group of gases in Earth's atmosphere that absorb infrared radiation and thus warm the planet. Essentially, greenhouse gases act as a blanket on the earth to keep more of the sun's solar radiation in the atmosphere. Carbon dioxide, released into the atmosphere primarily from fossil fuel combustion and deforestation, has had the greatest effect on climate since the Industrial Revolution. Other greenhouse gases, such as methane, nitrous oxides, and fluorinated gases, are more potent than carbon dioxide, but make up a smaller proportion of greenhouse gases. Both carbon dioxide and methane have increased considerably from the nineteenth century until 2023, with methane concentrations more than doubling. Water vapor is the most abundant and a highly potent greenhouse gas but, given its short atmospheric life span and sensitivity to temperature, is not a major factor in recently observed warming.

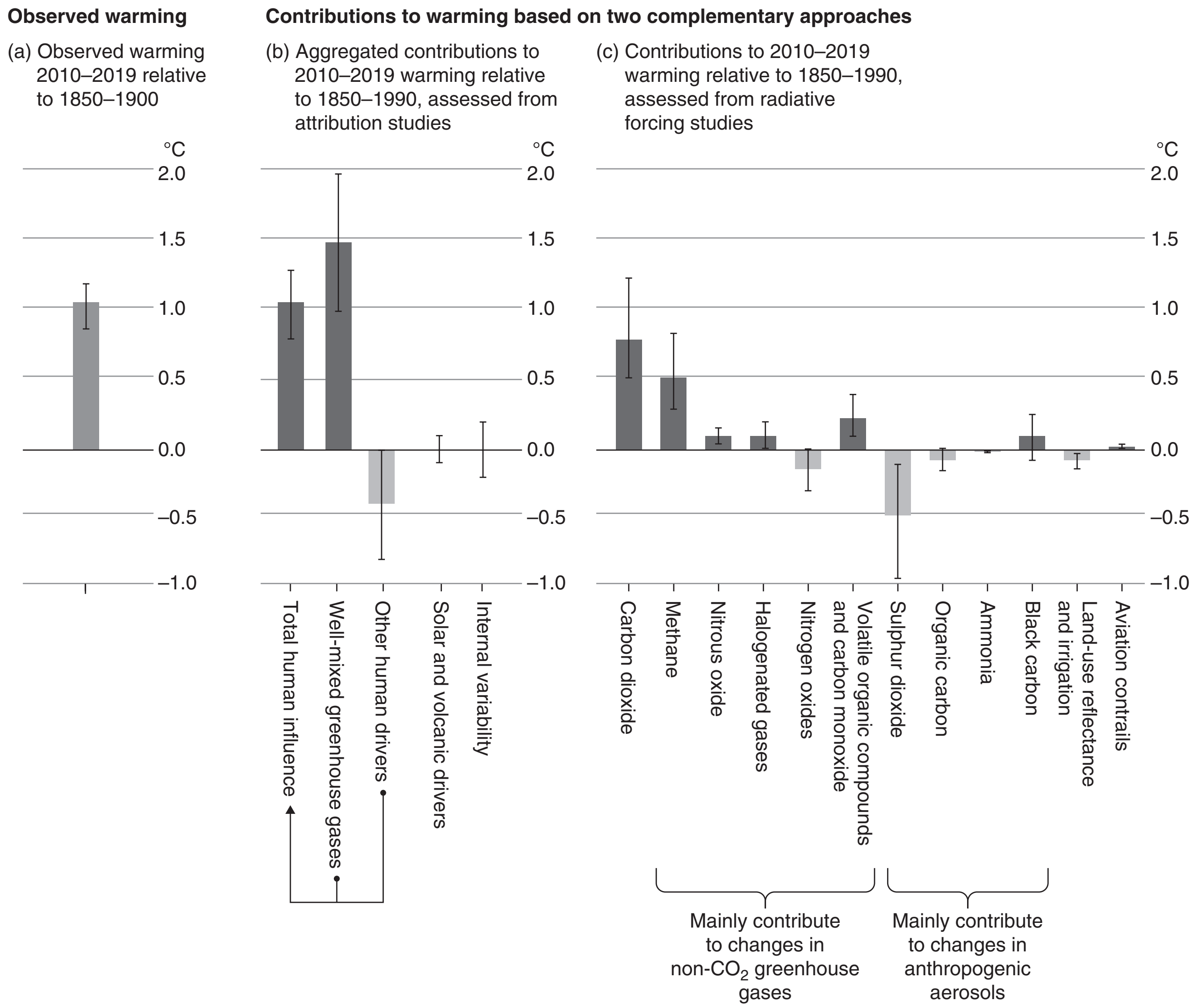

FIGURE 474-1 Assessed contributions to observed warming in 2010–2019 relative to 1850–1900. Assessed contributions to observed warming in 2010–2019 relative to 1850–1900. ***A.*** Observed global warming (increase in global surface temperature). Whiskers show the very likely range. ***B.*** Evidence from attribution studies, which synthesize information from climate models and observations. Temperature change attributed to total human influence; changes in well-mixed greenhouse gas (GHG) concentrations; other human drivers due to aerosols, ozone, and land-use change (land-use reflectance); solar and volcanic drivers; and internal climate variability. Whiskers show likely ranges. ***C.*** Evidence from the assessment of radiative forcing and climate sensitivity. Temperature changes from individual components of human influence: emissions of GHGs, aerosols, and their precursors; land-use changes (land-use reflectance and irrigation); and aviation contrails. Whiskers show very likely ranges. Estimates account for both direct emissions into the atmosphere and their effect, if any, on other climate drivers. For aerosols, both direct effects (through radiation) and indirect effects (through interactions with clouds) are considered. *(Used with permission from IPCC, 2021: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [V Masson-Delmotte et al (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 3–32, doi:10.1017/9781009157896.001. Reproduced with permission of the Licensor through PLSclear.)*

TABLE 474-1 Greenhouse Gases: Sources, Sinks, and Forcings

GAS	HUMAN SOURCES	SINK[a]	RADIATIVE FORCING[b] (95% CONFIDENCE INTERVAL)
Carbon dioxide (CO_2)	Fossil fuel combustion, deforestation	Uptake by oceans (~30%), plants	1.68 (1.33–2.03)
Methane (CH_4)	Fossil fuel production, ruminant animals, decomposition in landfills	Hydroxyl radicals in the troposphere	0.97 (0.74–1.20)
Nitrous oxide (N_2O)	Fertilizer, fossil fuel combustion, biomass burning, livestock manure	Photolysis in the stratosphere	0.17 (0.14–0.23)
Halocarbons	Refrigerants, electrical insulation, aluminum production	Hydroxyl radicals in the troposphere, sunlight in the stratosphere	0.18 (0.01–0.35)

[a] In this table, a *sink* refers to the place where greenhouse gases are naturally stored or the mechanism through which they are destroyed. [b] *Radiative forcing*, measured in watts per meter squared, refers to how much an entity can alter the balance of incoming and outgoing radiation to and from Earth's atmosphere. It is measured relative to a preindustrial (i.e., 1750) baseline. Greenhouse gases have a positive "forcing"; that is, on balance, they increase the amount of radiation (and specifically infrared radiation) that is retained in Earth's atmosphere.

Source: Intergovernmental Panel on Climate Change Fifth Assessment Report, Working Group 1, Chapter 8; American Chemical Society "Greenhouse gas sources and sinks," available at *www.acs.org/content/acs/en/climatescience/greenhousegases/sourcesandsinks.html*.

The atmosphere, some of the aerosols suspended in it, and clouds reflect a portion of incoming solar radiation back toward space. The remainder reaches Earth's surface, where it is absorbed and some is emitted back into the atmosphere. The earth emits energy absorbed from the sun at longer wavelengths, primarily infrared, that greenhouse gases can absorb. The change in wavelength that occurs as solar radiation is absorbed and re-emitted from the earth's surface is fundamental to the greenhouse effect **(Fig. 474-2)**.

TEMPERATURE

Climate change has clearly caused global warming with the average surface temperature of Earth increasing 1.09°C from 1880 to 2020. Moreover, the rate of global warming is faster now than at any time in the last 1000 years. Yet, this mean warming fails to show that warming is occurring much faster in certain regions. The Arctic has warmed twice as fast overall, and winters are warming faster than summers. Nighttime minimum temperatures are also rising faster than daytime high temperatures. Each of these nuances bears upon the incidence of infectious diseases in general and vector-borne disease specifically.

Due to climate change, extreme heat waves are expected to be more common, longer, and more severe in the future. The hottest 5-year period ever recorded since records started in the mid-nineteenth century was 2016–2020. Besides contributing directly to morbidity and mortality in human populations, heat waves wilt crops and are predicted to contribute substantially to agricultural losses. For example, the 2010 heat wave in Russia, which was unprecedented in its severity, contributed to hundreds of forest fires that generated enough air pollution to kill an estimated 56,000 people and that burned 300,000 acres of crops, including roughly 25% of the nation's wheat fields. Nutritional deficiencies underlie a substantial portion of the global burden of many infectious diseases.

PRECIPITATION

In addition to changing temperature, the emission of greenhouse gases and the consequent increase in energy in Earth's atmosphere have influenced the planet's water cycle. Since 1950, substantial increases in the heaviest precipitation events (i.e., those above the 95th percentile) have been observed in Europe and North America. Moreover, in 2022, floods harmed an estimated 58 million people globally. Other areas have seen greater drought, notably southern Australia and the southwestern United States.

A warmer atmosphere holds more water vapor. Specifically, air holds 6–7.5% more water vapor per degree (Celsius) of warming in the lower

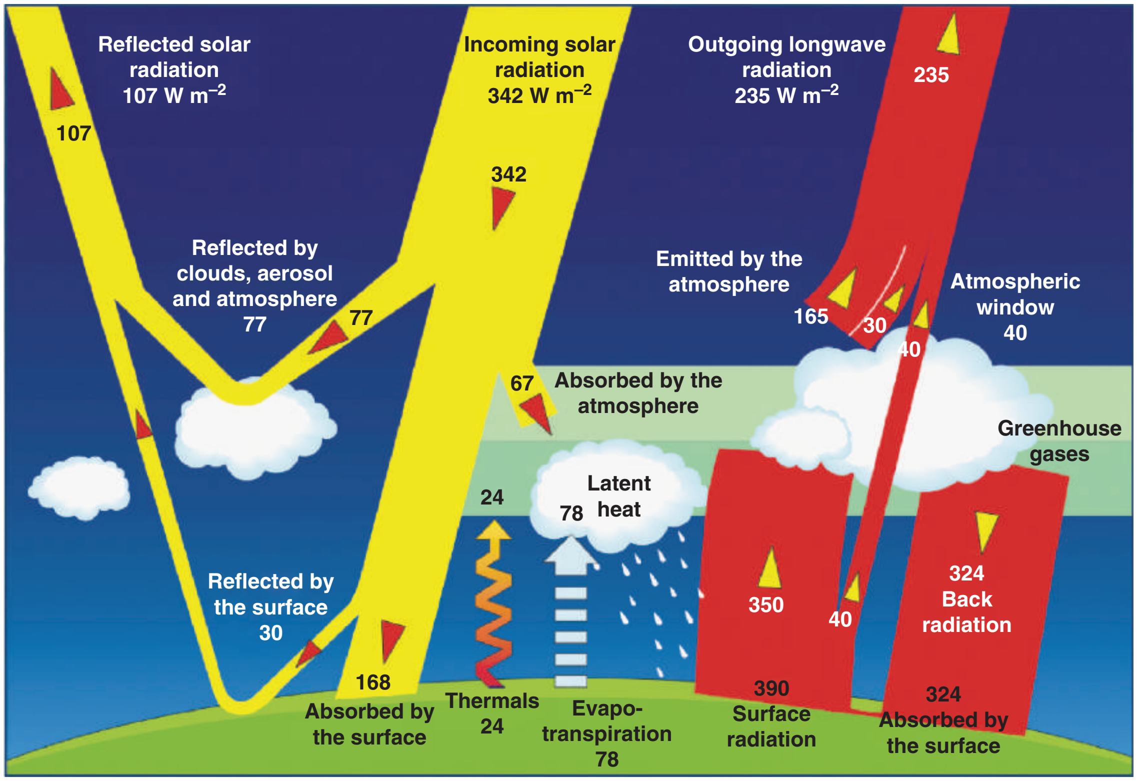
Reflected solar radiation 107 W m^{-2}
107
Incoming solar radiation 342 W m^{-2}
342
Outgoing longwave radiation 235 W m^{-2}
235
Reflected by clouds, aerosol and atmosphere 77
77
Emitted by the atmosphere
165
30
Atmospheric window 40
40
67
Absorbed by the atmosphere
Greenhouse gases
24
Latent 78 heat
Reflected by the surface 30
350
40
324 Back radiation
168
Absorbed by the surface
Thermals 24
Evapo-transpiration 78
390 Surface radiation
324 Absorbed by the surface

FIGURE 474-2 Earth's energy balance. *(Kiehl's Earth's Annual Global Mean Energy Budget, Bulletin of the American Meteorological Society, Vol. 78, No. 2, 1997 (Figure 7, page 206). © American Meteorological Society. Used with permission.)*

atmosphere. For areas that have traditionally had more precipitation on average, warming tends to promote heavier precipitation events. In contrast, in regions prone to drought, warming tends to result in greater periods between rainfalls and in the risk of drought. Floods and droughts have been associated with outbreaks of waterborne infectious diseases.

HURRICANES

The world's oceans have absorbed 90% of the excess heat that greenhouse gases have kept in Earth's atmosphere since the 1960s. Ocean heat provides energy for hurricanes, and warmer years tend to have greater hurricane activity. Stronger hurricanes (category 4 and 5) are expected with climate change, though climate change influence on hurricane frequency is uncertain. Modeling of future tropical cyclones suggests that their intensity may increase 2–11% by 2100 and that the average storm will bring 20% more rainfall.

SEA LEVEL RISE

Between 1901 and 2010, the global sea level rose ~200 mm, or ~1.7 mm per year on average. From 1993 to 2010, the rate of rise nearly doubled—i.e., to 3.2 mm annually. Most of this sea level rise has resulted from the thermal expansion of water. Glacial ice melt is the second greatest factor, and its contribution is accelerating. By 2100, global sea level may rise by 0.8–2 m, with an annual rate of rise of 8–16 mm at the century's end.

Sea level rise is not uniform. The rate of rise on the eastern seaboard of North America has been roughly double the global rate. Compounding sea level rise is the subsidence of coastal areas due to human settlement. In the absence of levee upgrades, an estimated 300 million people living near coasts worldwide will be at risk of flooding in 2050 because of the combined effects of subsidence, erosion, and sea level rise.

Along with extreme storms and overuse of coastal aquifers, rising seas also contribute to salinization of coastal groundwater. About 1 billion people rely on coastal aquifers for potable water.

EL NIÑO SOUTHERN OSCILLATION

The *El Niño Southern Oscillation* (ENSO) refers to periodic changes in water temperature in the eastern Pacific Ocean that occur roughly every 4–5 years. ENSO cycles have dramatic effects on weather around the globe. Warmer-than-average water temperatures in the eastern Pacific define *El Niño events* (see below), whereas cooler-than-average water temperatures define La Niña periods. Evidence is accruing that climate change may be increasing the frequency and severity of El Niño events.

El Niño events drive alterations in weather worldwide **(Fig. 474-3)** and are associated with extreme events and consequently higher rates of morbidity and mortality. Hurricane Mitch, one of the most powerful hurricanes ever observed, with winds reaching 290 km/h, dropped 1–1.8 m (3–6 feet) of rain over 72 h on parts of Honduras and Nicaragua. As a result of this storm, 11,000 people died and 2.7 million were displaced. Outbreaks of cholera, leptospirosis, and dengue occurred in the storm's aftermath.

POPULATION MIGRATION AND CONFLICT

The final common outcome of all climate-change effects is human displacement. Sea level rise, extreme heat and precipitation, droughts, and salinization of water supplies all conspire to make regions, including some inhabited by humans for millennia, uninhabitable. The 8 million inhabitants of low-lying South Pacific islands are vulnerable to sea level rise and could be a major source of climate migrants.

Climate change may also be contributing to humanitarian crises and conflicts. A severe 2011 drought in East Africa may have incited the Somali famine that resulted in 1 million refugees; mortality rates reached 7.4/10,000 in some refugee camps.

EFFECTS OF CLIMATE CHANGE ON INFECTIOUS DISEASE

The incidence of most, if not all, infectious diseases depends on climate. For any given infection, however, climate change is but one of many factors determining disease epidemiology. In instances in which climate change creates conditions favorable to the spread of infections,

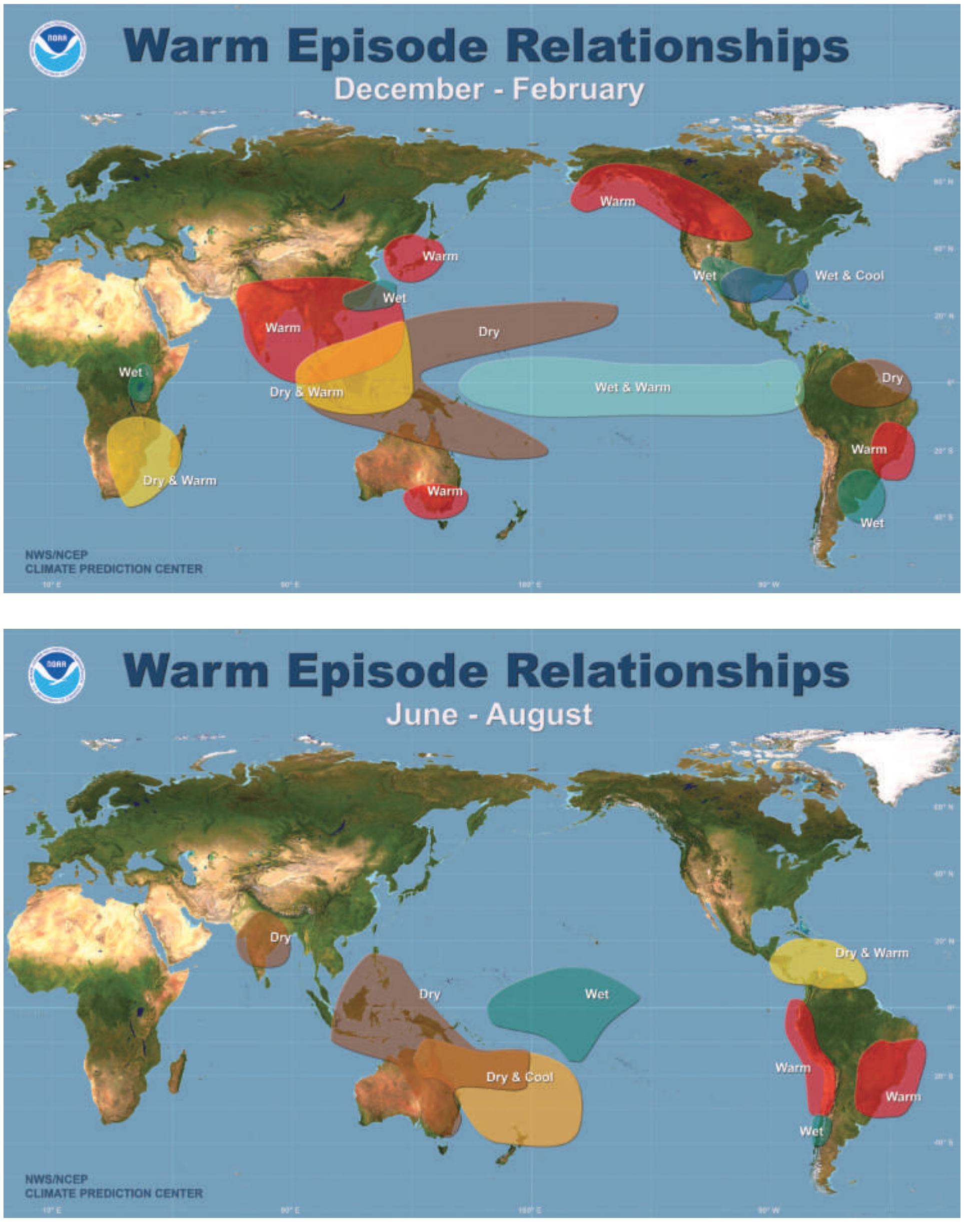

FIGURE 474-3 Characteristic weather anomalies, by season, during El Niño events. *(Source: Climate Prediction Center, https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/impacts/warm_impacts.shtml.)*

diseases may be kept in check through interventions such as vector control or antibiotic treatment.

Detecting the influence of climate change amid the many competing forces that bear on infectious disease emergence and spread can be challenging. Research on animal pathogens, which in most instances are less intervened upon than that with their human counterparts, has suggested how climate change may independently influence disease spread. Avian malaria in Hawaii, for example, has clearly moved to higher elevations where warmer temperatures have enabled disease transmission. Many putative pathways have been identified that connect greenhouse gas emissions to infectious disease risk in people. Climate change influences extend beyond contagious diseases and include diseases from microbial toxins, such as those that result from harmful algal blooms. Warmer ocean temperatures cause *Pseudonitzschia* spp., blue–green cyanobacteria, and dinoflagellates to grow faster, resulting in concurrent disease outbreaks.

VECTOR-BORNE DISEASE

Because insects are cold-blooded, ambient temperature dictates their geographic distribution. With increases in temperatures (in particular, nighttime minimum temperatures), insects are freed to move poleward and up mountainsides. At the same time, as new areas become climatically suitable, current mosquito habitats may become unsuitable because of heat extremes.

In addition, insects tend to be sensitive to water availability. Mosquitoes that transmit malaria, dengue, and other infections may breed in pools of water created by heavy downpours. As has been observed in the Amazon, breeding pools can also appear during periods of drought when rivers recede and leave behind stagnant pools of water for *Anopheles* mosquitoes. These circumstances have raised interest in the potentially favorable impact of water-cycle intensification on the spread of mosquito-borne disease.

Malaria • **TEMPERATURE** Higher temperatures promote higher mosquito-biting rates, shorter parasite reproductive cycles, and the potential for the survival of mosquito vectors of *Plasmodium* infection in locations previously too cold to sustain them. Modeling experiments have identified highland areas of East Africa and South America as perhaps most vulnerable to increased malarial incidence as a result of rising temperatures. In addition, an analysis of interannual malaria in Ecuador and Colombia has documented a greater incidence of malaria at higher altitudes in warmer years. Highland populations may be more vulnerable to malaria epidemics because they lack immunity.

Although rising temperature has the potential to expand the viable range of disease, malaria incidence is not linearly associated with temperature. While mosquitoes and parasites may adapt to a warming climate, the present optimal temperature for malaria transmission is ~25°C, with a range of transmission temperatures between 16°C and 34°C. Rising temperatures can also have differential effects on parasite development during external incubation and on the mosquitoes' gonotrophic cycle. Asynchrony between these two temperature-sensitive processes has been shown to decrease the vectorial capacity of mosquitoes.[1]

PRECIPITATION The abundance of *Anopheles* mosquitoes is strongly correlated with the availability of surface-water pools for mosquito breeding, and biting rates have been linked to soil moisture (a surrogate for breeding pools). Research in the East African highlands has documented that increased variance in rainfall over time has strengthened the association between precipitation and disease incidence. These disease-promoting effects of precipitation may be countered by the potential for extreme rainfall to flush mosquito larvae from breeding sites.

PROJECTIONS Climate models have begun to deliver output on regional scales, permitting projections of climate-suitable regions to assist national and local health authorities. Climate models speak to the temperature and precipitation ranges necessary for malaria transmission but do not account for the capacity of malaria control programs to halt the spread of disease. The global reduction in malaria distribution over the past century makes it clear that, even with climate change, malaria occurs in far fewer places today because of public health interventions. Recent vaccine trials also show promise in further reducing malaria risk.

Despite intensive efforts, malaria remains the single greatest vector-borne disease cause of morbidity and death in the world. Particularly in regions that are most affected by malaria and where the public health infrastructure is inadequate to contain it, climate modeling may provide a useful tool in determining where the disease may spread. Modeling studies in sub-Saharan Africa have suggested that, although East African nations may encompass regions that will become more climatically suitable for malaria over this century, West African nations may not. By 2100, temperatures in West Africa may largely exceed those optimal for malaria transmission, and the climate may become drier; in contrast, higher temperatures and changes in precipitation may allow malaria to move up the mountainsides of East African countries. Climate change may create conditions favorable to malaria in subtropical and temperate regions of the Americas, Europe, and Asia as well.

Dengue Like malaria epidemics, dengue fever epidemics depend on temperature **(Fig. 474-4)**. Higher temperatures increase the rate of larval development and accelerate the emergence of adult *Aedes* mosquitoes. The daily temperature range may also influence dengue virus transmission, with a smaller range corresponding to a higher transmission potential. Temperatures <15°C or >36°C substantially reduce mosquito feeding. In a *Rhesus* model of dengue, viral replication can occur in as little as 7 days with temperatures of >32–35°C; at 30°C, replication takes ≥12 days; and replication does not reliably occur at 26°C. Research on dengue in New Caledonia has shown peak transmission at ~32°C, reflecting combined effects of a shorter extrinsic incubation period, a higher feeding frequency, and more rapid development of mosquitoes. Along with temperature, peak relative humidity is a strong predictor of dengue outbreaks.

The association between dengue epidemics and precipitation is less consistent in the peer-reviewed literature, possibly because of the mosquito vector's greater reliance on domestic breeding sites than on natural pools of water. For instance, in some studies, increased access to a piped water supply has been linked to dengue epidemics, presumably because of associated increased domestic water storage. Nonetheless, several studies have established rainfall as a predictor of the seasonal timing of dengue epidemics.

The current global distribution of dengue largely overlaps the geographic spread of *Aedes* mosquitoes **(Fig. 474-5)**. The presence of *Aedes* without dengue endemicity in large regions of North and South America and Africa illustrates the relevance of variables other than climate to disease incidence. Nevertheless, coupled climatic–epidemiologic modeling suggests dramatic shifts in the relative vectorial capacity for dengue by the end of this century should little or no mitigation of greenhouse gas emissions occur **(Fig. 474-6)**.

Other Arbovirus Infections Climate change may favor increased geographic spread of other arboviral diseases, including Zika virus disease, chikungunya virus disease, West Nile virus disease, and eastern equine encephalitis. Zika virus moved to the Western Hemisphere from French Polynesia around 2013 and rapidly spread in Brazil in 2016. Although air travel was essential for the delivery of the virus to the Americas, the available evidence suggests that the 2015 El Niño event provided an optimal climate for the infection to take root and spread (see "ENSO-Related Outbreaks," below). *Aedes aegypti* is the primary vector for Zika virus. Chikungunya virus disease emerged in Italy in 2007, having previously been mostly a disease of African nations. Climate models predict that, should competent vectors be present, conditions will be suitable for the chikungunya virus to gain a foothold in Western Europe, especially France, in the first half of

---

[1] *rVc* is the vectorial capacity relative to the vector-to-human population ratio and is defined by the equation $rVc = a^2 b_h b_m e^{-\mu mn} / \mu_m$ where *a* is the vector biting rate; *b*h is the probability of vector-to-human transmission per bite; b_m is the probability of human-to-vector infection per bite; *n* is the duration of the extrinsic incubation period; and *μ*m is the vector mortality rate.

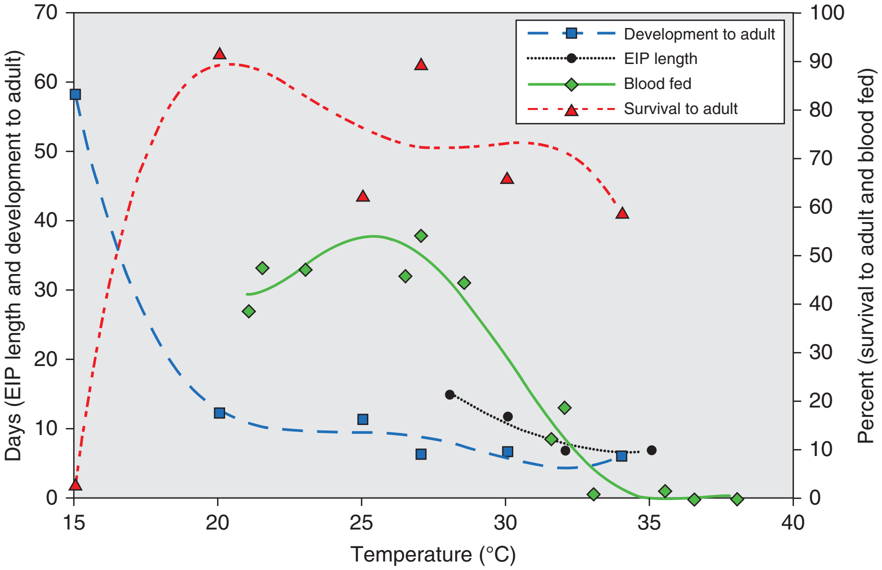

FIGURE 474-4 Effects of temperature on variables associated with dengue transmission. Shown are the number of days required for development of immature *Aedes aegypti* mosquitoes to adults, the length of the dengue virus type 2 extrinsic incubation period (EIP), the percentage of *Ae. aegypti* mosquitoes that complete a blood meal within 30 min after a blood source is made available, and the percentage of hatched *Ae. aegypti* larvae surviving to adulthood. *(Reproduced from CW Morin et al: Climate and dengue transmission: Evidence and implications. Environ Health Perspect 121:1264, 2013.)*

the twenty-first century. In North America, areas favorable to West Nile virus outbreaks are expected to shift northward in this century. Current hotspots in North America are the California Central Valley, southwestern Arizona, southern Texas, and Louisiana, which have both compatible climates and avian reservoirs for the disease. By mid-century, the upper Midwest and New England will be more climatically suited to West Nile virus; by the end of the century, the area of risk may shift even further north to southern Canada. Whether the disease will ultimately move northward will depend on reservoir availability and mosquito control programs, among other factors.

Lyme Disease In the past few decades, *Ixodes scapularis*, the primary tick vector for Lyme disease as well as for anaplasmosis and babesiosis in New England, has become established in Canada because of warming temperatures. With climate change, the range of the tick is expected to expand further **(Fig. 474-7)**.

Lyme disease, caused by the spirochete *Borrelia burgdorferi*, is the most commonly reported vector-borne disease in North America, with ~60,000 cases annually. In Europe, Lyme disease has also increased and expanded geographically. Furthermore, the timing and peaks of cases have been affected, with the annual peak case numbers in 2019 arriving 6 weeks earlier than 25 years prior.

■ WATERBORNE DISEASE

Many microorganisms, from bacteria to toxin-producing algae, cause waterborne disease **(Table 474-2)**. The outbreaks of waterborne disease are associated with heavy rainfall events. A review of 548 waterborne disease outbreaks in the United States found that 51% were preceded by precipitation levels above the 90th percentile. Since 1900, most regions of the United States except the Southwest and Hawaii have experienced an increase in heavy downpours with the greatest intensification of the water cycle in New England and Alaska.

Most disease outbreaks after heavy precipitation occur through contamination of drinking-water supplies. While outbreaks related to surface-water contamination generally occur within a month of the precipitation event, disease outbreaks from groundwater contamination tend to occur ≥2 months later. According to a review of published reports of waterborne disease outbreaks, *Vibrio* and *Leptospira* species are the pathogens most commonly involved in the wake of heavy precipitation.

Combined Sewer Systems Roughly 40 million people in the United States and millions more around the world rely on combined sewer systems in which storm water and sanitary wastewater are conveyed in the same pipe to treatment facilities. These systems were

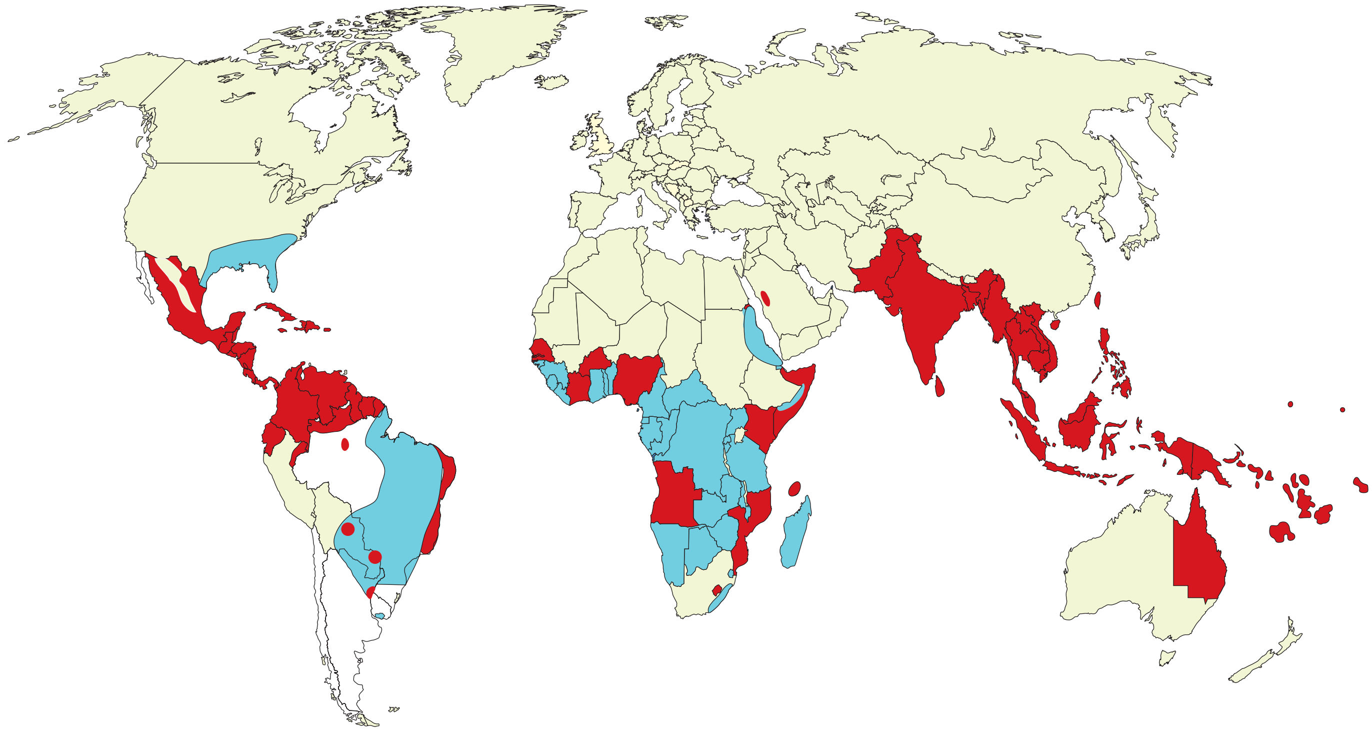

FIGURE 474-5 Distribution of *Aedes aegypti* mosquitoes (*turquoise*) and dengue fever epidemics (*red*). *(Map produced by the Agricultural Research Service of the U.S. Department of Agriculture.)*

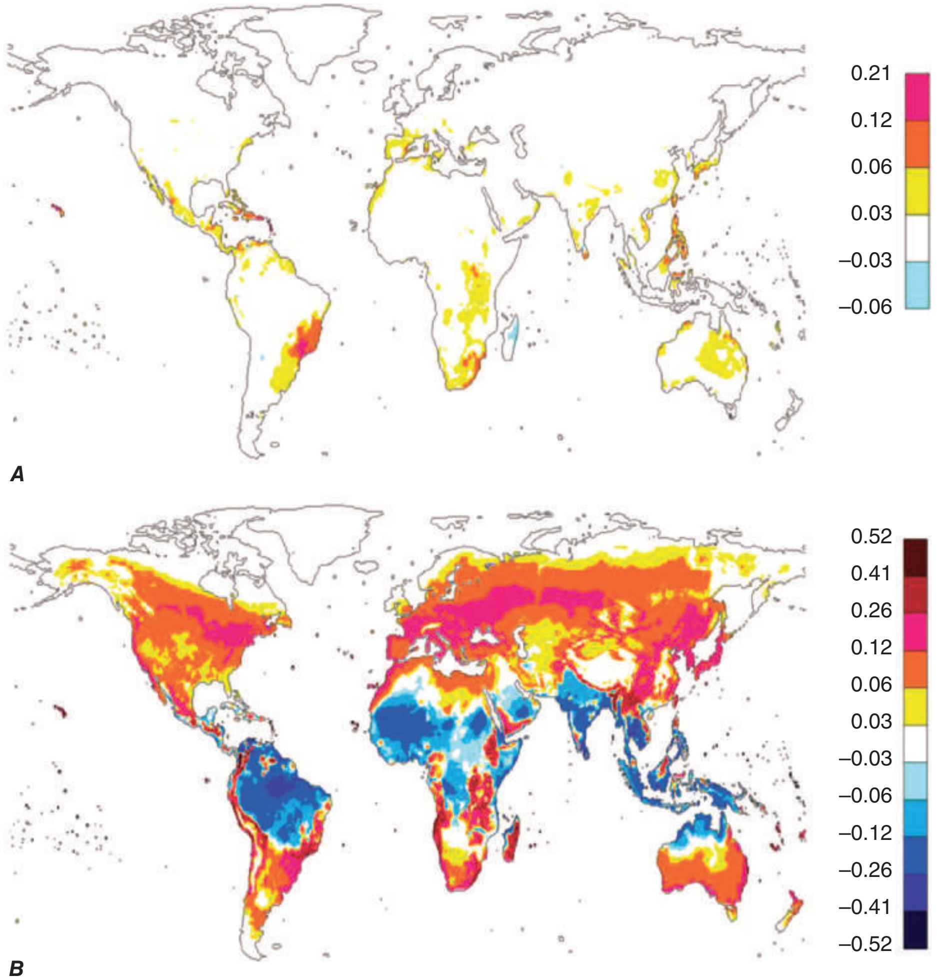

FIGURE 474-6 Trend of annually averaged global dengue epidemic potential (*rVc*). Differences in rVc are based on 30-year averages of temperature and daily temperature range. ***A.*** Differences between 1980–2009 and 1901–1930. ***B.*** Differences between 2070–2099 and 1980–2009. The mean value of rVc was averaged from five global climate models under RCP8.5, a scenario of high greenhouse-gas emission. The color bar describes the values of the rVc. *(From J Liu-Helmersson et al: Vectorial capacity of Aedes aegypti: Effects of temperature and implications for global dengue epidemic potential. PLoS One 9:e89783, 2014.)*

designed based on the nineteenth-century climate, in which heavy downpours were less frequent than they are today. The frequency of combined sewer overflows usually into freshwater bodies and resulting in untreated sewage and runoff potentially containing heavy metals and pesticides, has been increasing in cities worldwide. For instance, the channel leading into Lake Michigan from Milwaukee had its highest levels of *Escherichia coli*, up to 100 times the Environmental Protection Agency guidance for recreational waters, after combined sewer overflows **(Fig. 474-8)**.

Rising Temperatures and *Vibrio* Species Warmer temperatures favor proliferation of *Vibrio* species and disease outbreaks, as has been demonstrated in countries surrounding the Baltic Sea, Chile, Israel, northwestern Spain, and the U.S. Pacific Northwest. Around the Baltic Sea, outbreaks of *Vibrio* infection may be particularly likely because of faster warming near the poles and the sea's relatively low salt content. In the United States, levels of vibriosis roughly tripled from 1996–2010, with the highest number of cases occurring each summer.

ENSO-Related Outbreaks Weather extremes tied to El Niño events afford a window into a climate-changed future. Recent evidence indicates that climate change itself may be strengthening El Niño events. These events tend to promote epidemic infections in certain regions **(Fig. 474-9)**.

Associations of El Niño with outbreaks of Rift Valley fever in eastern and southern Africa have been known since the 1950s. El Niño favors wet conditions suitable for the insect vectors of the disease in these regions. Given the strong association between El Niño conditions and disease incidence, models have successfully predicted Rift Valley fever epidemics in humans and animals. In the 2006–2007 El Niño season, for example, outbreaks of Rift Valley fever were accurately predicted 2–6 weeks prior to epidemics in Somalia, Kenya, and Tanzania.

El Niño has had inconsistent associations with malaria incidence in African countries. Some of the strongest associations between El Niño and malaria have been identified in South Africa and Swaziland, where available data on incidence are relatively robust; however, even in these instances, the observed increased risk did not reach statistical significance. A stronger link to El Niño has been found in several studies done in South America. Research on malaria incidence in Colombia between 1960 and 2006 found that a 1°C temperature rise contributed to a 20% increase in incidence.

In the desert Southwest of the United States, hantavirus erupted following the strong El Niño of 1997. Unseasonal rainfall caused the desert to bloom, providing food resources and resulting in a boon to the mouse population. The following year saw the climate revert back to normal (arid) conditions, and as a result, increased human/mouse contact occurred as mice were forced to seek out food and nest in human dwellings.

In Bangladesh, a strong association has been observed between the El Niño Southern Oscillation Index and cholera epidemics. During El Niño, a combination of warmer sea-surface temperatures (SSTs) and rich nutrient runoff lead to phytoplankton blooms detectable by satellite. These blooms, in turn, feed zooplankton, which can harbor *V. cholerae* and, therefore, amplify their presence in the environment, causing cholera epidemics in the region **(Fig. 474-10)**.

El Niño years are often associated with an increased incidence of dengue. Research on dengue outbreaks in Thailand from 1996 to 2005 revealed that 15–22% of the variance in monthly dengue disease incidence was attributable to El Niño. In South America, data on dengue outbreaks between 1995 and 2010 showed an increased incidence during the El Niño events of 1997–1998 and 2006–2007.

El Niño may have contributed to the emergence of Zika virus in the wake of a very strong ENSO event during the winter of 2015–2016. Temperatures throughout South and Central America were exceptionally high and, just prior to the Zika epidemic, the vectorial capacity of *Ae. aegypti* was at its highest level compared with the previous 60 years. Not surprisingly (because carried by the same mosquito vector), large epidemics of dengue also occurred in Brazil and Colombia at that time. But laboratory studies show that the temperature optimum for Zika virus is 5°C warmer than for dengue. Thus, it is possible the extreme temperature played a role in the extent of the Zika epidemic.

CLIMATE CHANGE, POPULATION DISPLACEMENT, AND INFECTIOUS DISEASE EPIDEMICS

For many reasons, including freshwater shortages, flooding, food shortages, and climate change–driven conflicts, climate change has and will continue to put pressure on human populations to move. Human migrations have long been associated with epidemic disease in the

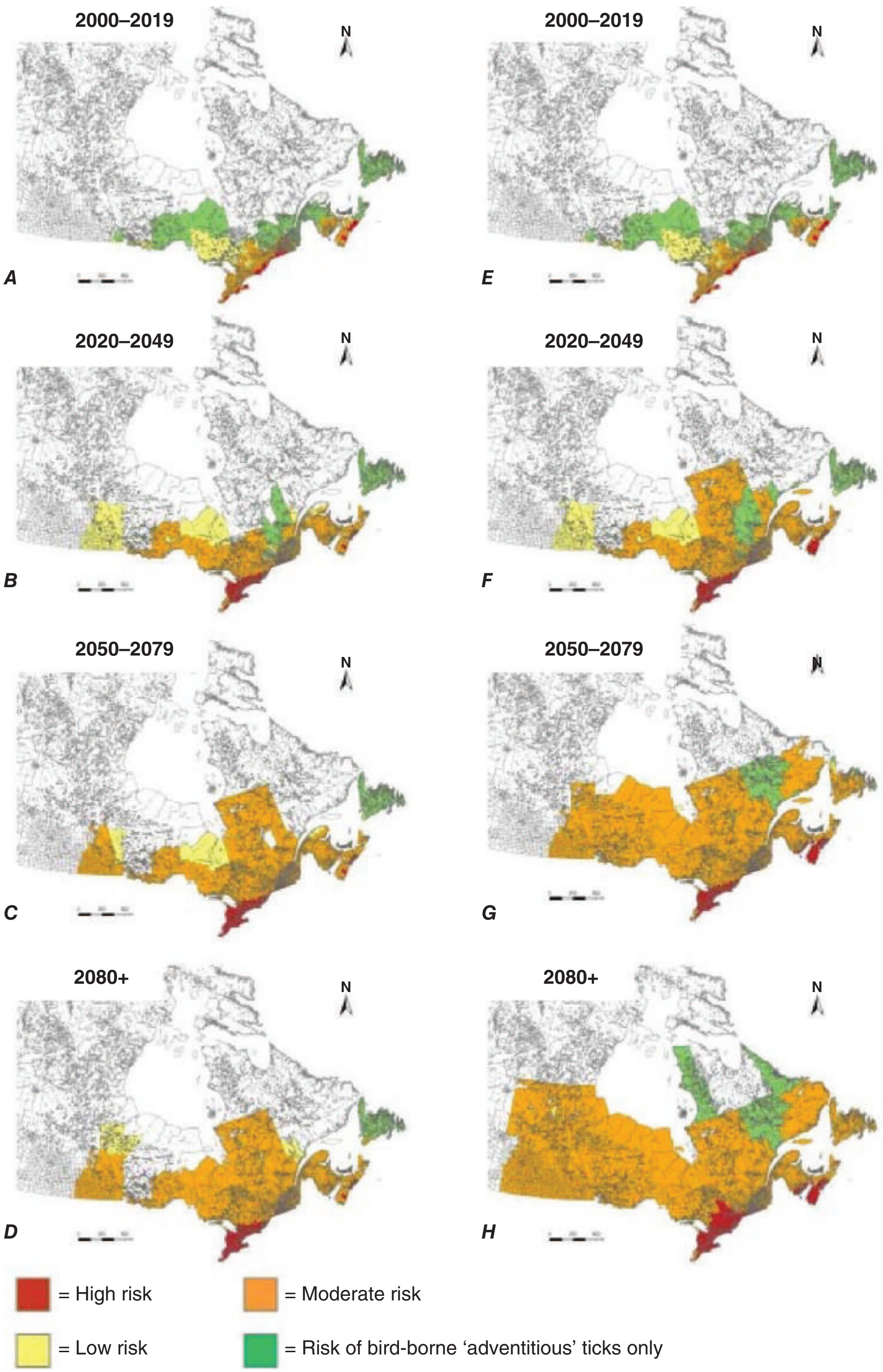
2000–2019
N
A
2000–2019
N
E
2020–2049
N
B
2020–2049
N
F
2050–2079
N
C
2050–2079
N
G
2080+
N
D
2080+
N
H
🟥 = High risk
🟧 = Moderate risk
🟨 = Low risk
🟩 = Risk of bird-borne 'adventitious' ticks only

FIGURE 474-7 Risk maps for the occurrence of the Lyme disease vector *Ixodes scapularis*. Expansion of *I. scapularis*–affected census subdivisions (CSDs) in Canada from the present (using 1971–2000 temperature normals) to the 2080s (using temperature conditions predicted by the CGCM2 climate model under emissions scenario A2). In ***A–D*** ("slow" scenario), the model assumes that by the end of each time period, only risk CSDs with an algorithm value in the top 10% will contain an *I. scapularis* population. In ***E–H*** ("fast" scenario), the model assumes that by the end of each time period, all CSDs within the "moderate"-risk zone for *I. scapularis* establishment (risk CSDs) contain an *I. scapularis* population. For both scenarios, the time steps are 2000–2019, 2020–2049, 2050–2079, and 2080– 2109. "High"-risk regions for *I. scapularis* population establishment are indicated in *red*. "Moderate"-risk regions are in *orange.* "Low"-risk regions are in *yellow*. Regions with no risk of established populations but some risk from bird-borne "adventitious" ticks are in *green*. Regions with no predicted risk of either are colorless. *(Used from NH Ogden: Risk maps for range expansion of the Lyme disease vector, Ixodes scapularis, in Canada now and with climate change. Int J Health Geogr 7:1, 2008.)*

TABLE 474-2 Climate Sensitive Agents of Water-Related Illness

PATHOGEN OR TOXIN PRODUCER	EXPOSURE PATHWAY	SELECTED HEALTH OUTCOMES AND SYMPTOMS	MAJOR CLIMATE CORRELATION OR DRIVER (STRONGEST DRIVERS LISTED FIRST)
Algae: Toxigenic marine species of *Alexandrium, Pseudo-nitzschia, Dinophysis, Gambierdiscus; Karenia brevis*	Shellfish, fish Recreational waters (aerosolized toxins)	Gastrointestinal and neurologic illness caused by shellfish poisoning (paralytic, amnesic, diarrhetic, neurotoxic) or fish poisoning (ciguatera). Asthma exacerbations, eye irritations caused by contact with aerosolized toxins (*K. brevis*).	Temperature (increased water temperature), ocean surface currents, ocean acidification, hurricanes (*Gambierdiscus* spp. and *K. brevis*)
Cyanobacteria (multiple freshwater species producing toxins including microcystin)	Drinking water Recreational waters	Liver and kidney damage, gastroenteritis (diarrhea and vomiting), neurologic disorders, and respiratory arrest.	Temperature, precipitation patterns
Enteric bacteria and protozoan parasites: *Salmonella enterica; Campylobacter* species; toxigenic *Escherichia coli; Cryptosporidium*; *Giardia*	Drinking water Recreational waters Shellfish	Enteric pathogens generally cause gastroenteritis. Some cases may be severe and may be associated with long-term and recurring effects.	Temperature (air and water; both increase and decrease), heavy precipitation, and flooding
Enteric viruses: enteroviruses; rotaviruses; noroviruses; hepatitis A and E	Drinking water Recreational waters Shellfish	Most cases result in gastrointestinal illness. Severe outcomes may include paralysis and infection of the heart or other organs.	Heavy precipitation, flooding, and temperature (air and water; both increase and decrease)
Leptospira and *Leptonema* bacteria	Recreational waters	Mild to severe flu-like illness (with or without fever) to severe cases of meningitis, kidney, and liver failure.	Flooding, temperature (increased water temperature), heavy precipitation
Vibrio bacteria species	Recreational waters Shellfish	Varies by species but include gastroenteritis (*V. parahaemolyticus, V. cholerae*), septicemia (bloodstream infection) through ingestion or wounds (*V. vulnificus*), skin, eye, and ear infections (*V. alginolyticus*).	Temperature (increased water temperature), sea level rise, precipitation patterns (as it affects coastal salinity)

Source: Climate impacts on water-related illness, in The Impacts of Climate Change on Human Health in the United States: A Scientific Assessment. U.S. Global Change Research Program, Washington, DC, 2016. *https://health2016.globalchange.gov/water-related-illness*.

migrating populations themselves and in the communities in which they settle. The specific pathogens and patterns of disease that may appear after population migration relate to endemic diseases present in the migrant populations.

Large-scale migrations are common after extreme precipitation events. In August 2022, Pakistan experienced over three times the average rainfall, resulting in extensive flooding. Roughly 33 million people were affected, over 1400 people died, and 1.7 million homes were destroyed, all resulting in large-scale displacement. Additionally, standing water provided ideal breeding sites for vectors, and contaminated flood waters in two southern provinces led to outbreaks of diarrhea, cholera, and malaria. Accelerated computing power now allows for improved attribution of single weather events to climate change. For Pakistan, climate attribution analysis suggests that intense rainfall over short periods (5 days) has become more extreme with warming temperatures.

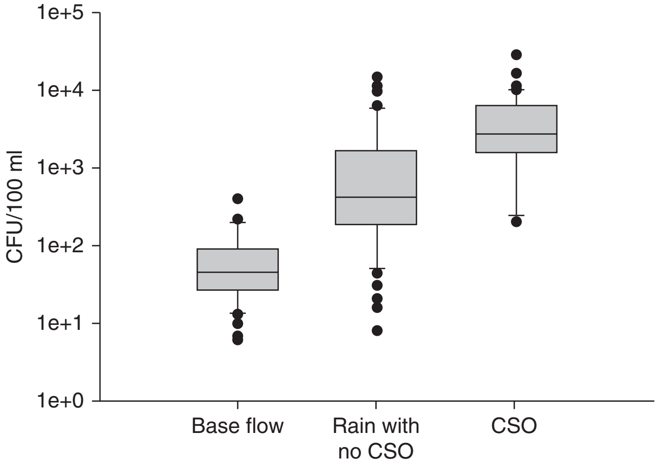

FIGURE 474-8 Levels of *Escherichia coli* in the Milwaukee estuary, which discharges to Lake Michigan, 2001–2007. Levels of *E. coli* in the Milwaukee estuary, which discharges to Lake Michigan, 2001–2007, during base flow (n = 46); following rain events with no CSO (n = 70); and following CSO events (n = 54). Boxes indicate 75% of values, with median values drawn in each. Whiskers are 95% of values, and outliers are shown as closed circles. There were significant differences in *E. coli* levels following rainfall and CSOs compared to base flow ($p \leq 0.05$). CFU, colony forming units; CSO, combined sewer overflow. *(Reproduced with permission from JA Patz et al: Climate change and waterborne disease risk in the Great Lakes region of the U.S. Am J Prev Med 35:451, 2008.)*

In low- and middle-income countries, infectious disease outbreaks associated with population displacement may be harder to detect and respond to. People forced to leave their homes en masse are at risk for contracting infections with any pathogen that may be present within the displaced population, including sexually transmitted diseases such as HIV, or airborne or droplet transmitted diseases such as tuberculosis and measles. Reducing disease risk requires overlaying of climate-related migration risk with foci of disease epidemics.

A BROADER VIEW OF CLIMATE CHANGE AND HEALTH

While climate change has far-reaching implications for the distribution and severity of infectious diseases worldwide, the greatest disease burdens related to climate change may not be due to infections, at least primarily. Climate change erodes the foundations of health, such as safe drinking water and food security, due to climatic extremes such as flooding or droughts. In addition, resource scarcity and climate instability are increasingly associated with conflicts. For example, severe droughts made more likely by climate change may have been a factor in the revolutions of the Arab Spring and the Syrian civil war. Of course, without adequate nutrition, water, or shelter, infectious disease risks rise.

The public health response to climate change entails both mitigation and adaptation measures. *Mitigation* represents primary prevention and involves the reduction of greenhouse gas emissions into the atmosphere. Although no clear safety threshold of greenhouse gas emissions has been agreed upon, in 2015 at the 21st Conference of the Parties (COP21) of the United Nations Framework Convention on Climate Change (UNFCCC) held in Paris, national governments from the major industrialized countries agreed to set a warming target of <2°C above preindustrial levels by 2050; the attainment of this goal will require reducing greenhouse gas emissions by 40–70% below 2010 levels. At COP21, a framework was established for a global carbon market and nationally determined contributions (NDCs) for countries to commit to greenhouse gas reductions to meet the Paris Climate Agreement. At the 2023 Conference of the Parties (COP28) held in Dubai,

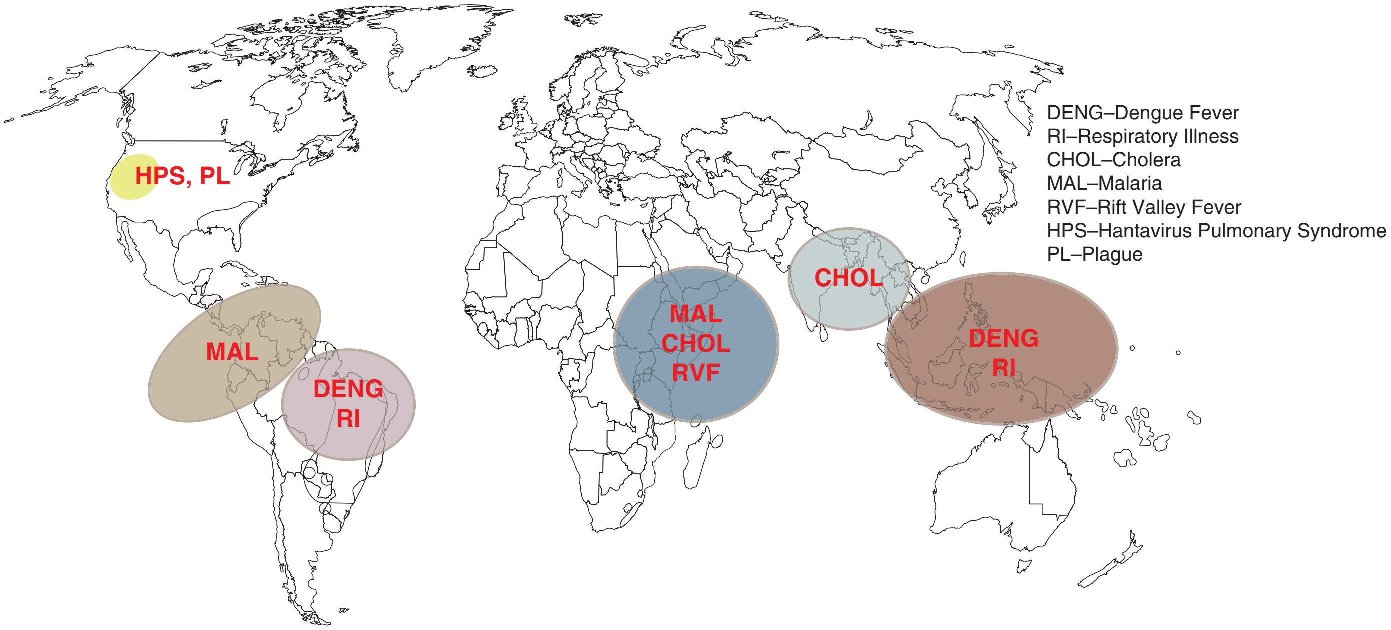

FIGURE 474-9 Characteristic patterns of disease outbreaks associated with El Niño events, determined on the basis of 2006–2007 conditions. *(From A Anyamba et al: Developing global climate anomalies suggest potential disease risks for 2006–2007. Int J Health Geogr 5:60, 2006.)*

for the first time the “need to transition from fossil fuels” was explicitly included in the final text of the agreement.

Mitigation also confers health co-benefits, including better air quality and lower incidence and severity of respiratory infections, associated with less bio- or fossil fuel combustion. One study estimated that by eliminating air quality pollutants (PM 2.5, sulfur dioxide, nitrous oxides) from energy generation across the United States, >53,000 premature deaths would be avoided. Of note, evidence has shown that long-term air pollution exposure may contribute to mortality risk from COVID-19 and influenza. Dietary and agricultural changes can also afford climate change mitigation and improve human health. Resource-intensive foods, like red meat and dairy, can lead to higher cases of diabetes and cardiovascular disease, so switching to plant-based diets can improve human health and reduce greenhouse gas emissions.

Climate *adaptation* is secondary prevention and seeks to reduce harms associated with sea level rise, heat waves, floods, droughts, wildfires, and other greenhouse gas–driven events. The efficacy of adaptation is constrained by the challenges inherent in predicting the precise location, duration, and severity of extreme weather events and flooding related to sea level rise, among other considerations **(Fig. 474-11)**.

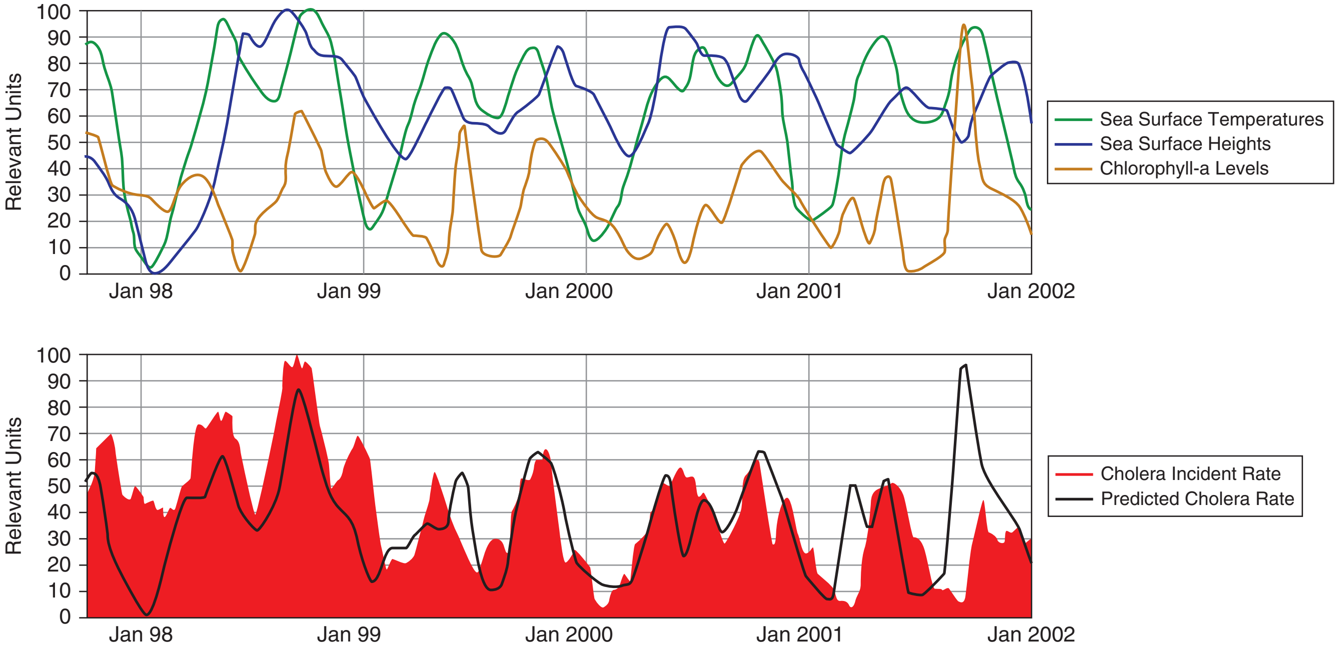

FIGURE 474-10 Predicting cholera rates. Sea surface temperature, sea surface height, and chlorophyll-a predicting cholera in Bangladesh. Top graph illustrates environmental data detected by satellite measurements from the Bay of Bengal from January 1998–January 2002 and includes sea surface temperature (*green line*), sea surface height (*blue line*), and chlorophyll-a levels (*yellow line*), an indicator for the abundance of phytoplankton. Bottom graph illustrates the incidence rate of cholera (*red*) in Bangladesh over that time period. The *black line* combines the three environmental parameters and, superimposed over cholera incidence rate, shows a strong correlation. *(Reproduced with permission from D Grimes et al: Viewing marine bacteria, their activity and response to environmental drivers from orbit: Satellite remote sensing of bacteria. Microb Ecol 67:489, 2014.)*

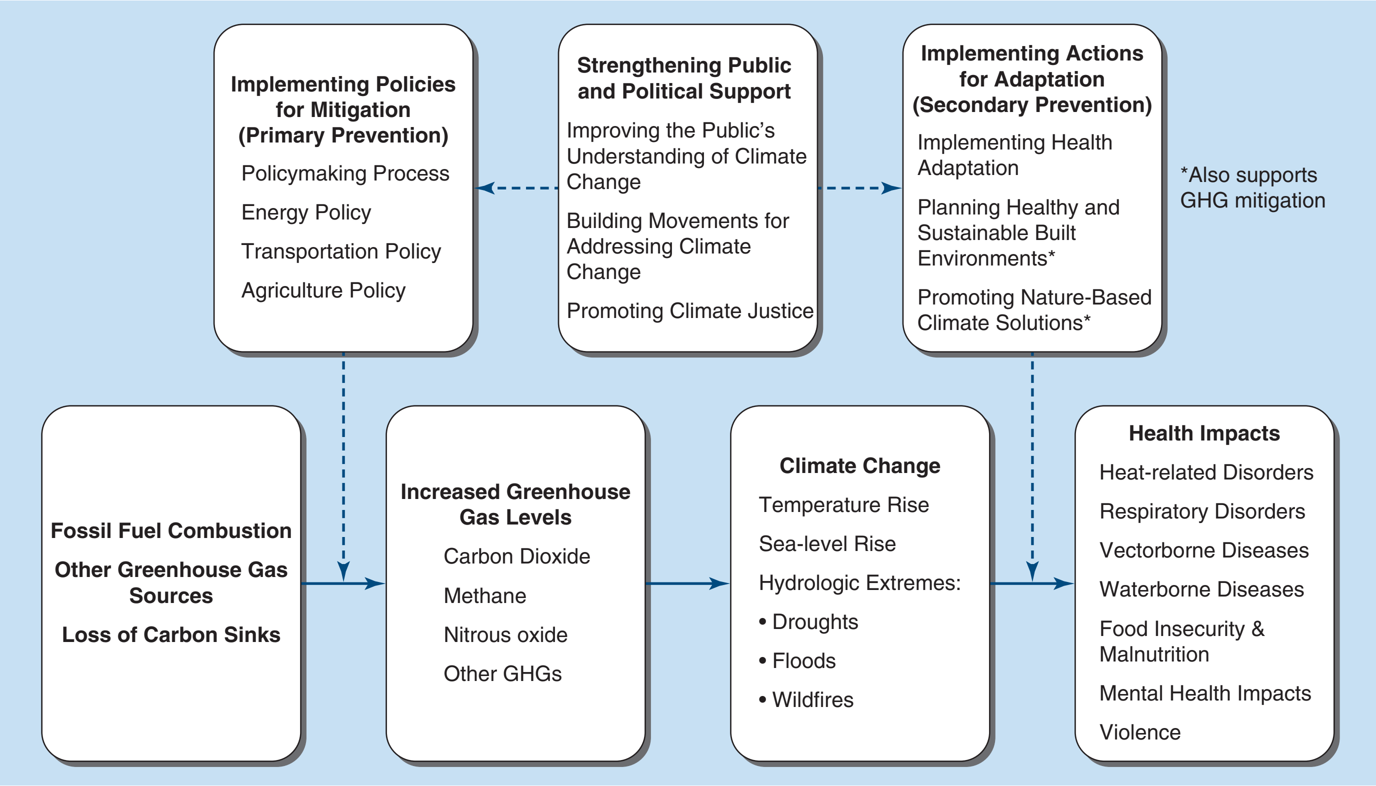
Implementing Policies for Mitigation (Primary Prevention)
Policymaking Process
Energy Policy
Transportation Policy
Agriculture Policy
Strengthening Public and Political Support
Improving the Public's Understanding of Climate Change
Building Movements for Addressing Climate Change
Promoting Climate Justice
Implementing Actions for Adaptation (Secondary Prevention)
Implementing Health Adaptation
Planning Healthy and Sustainable Built Environments*
Promoting Nature-Based Climate Solutions*
*Also supports GHG mitigation
Fossil Fuel Combustion
Other Greenhouse Gas Sources
Loss of Carbon Sinks
Increased Greenhouse Gas Levels
Carbon Dioxide
Methane
Nitrous oxide
Other GHGs
Climate Change
Temperature Rise
Sea-level Rise
Hydrologic Extremes:
• Droughts
• Floods
• Wildfires
Health Impacts
Heat-related Disorders
Respiratory Disorders
Vectorborne Diseases
Waterborne Diseases
Food Insecurity & Malnutrition
Mental Health Impacts
Violence

FIGURE 474-11 Conceptual framework of climate change and its health impacts. *(Reproduced with permission from JA Patz, BS Levy.)*

FURTHER READING

Caminade C et al: Global risk model for vector-borne transmission of Zika virus reveals the role of El Niño 2015. Proc Natl Acad Sci USA 114:119, 2017.

Colón-González FJ et al: The effects of weather and climate change on dengue. PLoS Negl Trop Dis 7:e2503, 2013.

Glass GE et al: Satellite imagery characteristics local animal reservoir populations of Sin Nombre virus in the southwestern United States. Proc Natl Acad Sci USA 99:16817, 2002.

Goren A et al: The emergence and shift in seasonality of Lyme borreliosis in Northern Europe. Proc Biol Sci 290:20222420, 2023.

Levy BS, Patz JA (eds): *Climate Change and Public Health*, 2nd ed. Oxford University Press, 2024.

Mora C et al: Over half of known human pathogenic diseases can be aggravated by climate change. Nat Climate Change 12:869, 2022.

Ogden NH: Risk maps for range expansion of the Lyme disease vector, *Ixodes scapularis*, in Canada now and with climate change. Int J Health Geogr 7:1, 2008.

Paaijmans KP et al: Temperature-dependent pre-bloodmeal period and temperature-driven asynchrony between parasite development and mosquito biting rate reduce malaria transmission intensity. PLoS One 8:e55777, 2013.

Patz JA et al: Impact of regional climate change on human health. Nature 438:310, 2005.

Patz JA et al: Climate change and waterborne disease risk in the Great Lakes region of the U.S. Am J Prev Med 35:451, 2008.

Ryan SJ et al: Warming temperatures could expose more than 1.3 billion new people to Zika virus risk by 2050. Glob Change Biol 27:84, 2021.

Trtanj J et al: Climate impacts on water-related illness, in *The Impacts of Climate Change on Human Health in the United States: A Scientific Assessment*. U.S. Global Change Research Program, Washington, DC, 2016, pp 157–188.

475

Altitude Illness

Buddha Basnyat, Geoffrey Tabin, Steven Roy

EPIDEMIOLOGY

Mountains cover one-fifth of the earth's surface; 140 million people live permanently at altitudes ≥2500 m, and 100 million people travel to high-altitude locations each year. Skiers in the Alps or Aspen; tourists to La Paz, Ladakh, or Lahsa; religious pilgrims to Kailash-Manasarovar or Gosainkunda; trekkers and climbers to Kilimanjaro, Aconcagua, or Everest; miners working in high-altitude sites in South America; and military personnel deployed to high-altitude locations are all at risk of developing acute mountain sickness (AMS), high-altitude cerebral edema (HACE), high-altitude pulmonary edema (HAPE), and other altitude-related problems. AMS is the benign form of altitude illness, whereas HACE and HAPE are life-threatening. Altitude illness is likely to occur above 2500 m but has been documented even at 1500–2500 m. In the Mount Everest region of Nepal, ~50% of trekkers who walk to altitudes >4000 m over ≥5 days develop AMS, as do 84% of people who fly directly to 3860 m. The incidences of HACE and HAPE are much lower than that of AMS, with estimates in the range of 0.1–4%. Finally, reentry HAPE, which in the past was generally limited to highlanders (long-term residents of altitudes >2500 m) in the Americas, is now being seen in Himalayan and Tibetan highlanders—and often misdiagnosed as a viral illness—as a result of recent rapid air, train, and motorable-road access to high-altitude settlements.

PHYSIOLOGY

Ascent to a high altitude subjects the body to a decrease in barometric pressure that results in a decreased partial pressure of oxygen in the inspired gas in the lungs. This change leads in turn to less pressure, driving oxygen diffusion from the alveoli and throughout the oxygen cascade. A normal initial "struggle response" to such an ascent includes increased ventilation—the cornerstone of acclimatization—mediated by the carotid bodies. Hyperventilation may cause respiratory alkalosis and dehydration. Respiratory alkalosis may be extreme, with an arterial blood pH of >7.7 (e.g., at the summit of Everest). Alkalosis may depress the ventilatory drive during sleep, with consequent periodic breathing and hypoxemia. During early acclimatization, renal suppression of carbonic anhydrase and excretion of dilute alkaline urine combat alkalosis and tend to bring the pH of the blood to normal. Other physiologic changes during normal acclimatization include increased sympathetic tone; increased erythropoietin levels, leading to increased hemoglobin levels and red blood cell mass; increased tissue capillary density and mitochondrial numbers; and higher levels of 2,3-bisphosphoglycerate, enhancing oxygen utilization. Even with normal acclimatization, however, ascent to a high altitude decreases maximal exercise capacity (by ~1% for every 100 m gained above 1500 m) and increases susceptibility to cold injury due to peripheral vasoconstriction. If the ascent is made faster than the body can adapt to the stress of hypobaric hypoxemia, altitude-related disease states can result.

GENETICS

Hypoxia-inducible factor, which acts as a master switch in high-altitude adaptation, controls transcriptional responses to hypoxia throughout the body and is involved in the release of vascular endothelial growth factor (VEGF) in the brain, erythropoiesis, and other pulmonary and cardiac functions at high altitudes. In particular, the gene *EPAS1*, which codes for transcriptional regulator hypoxia-inducible factor 2α, appears to play an important role in the adaptation of Tibetans living at high altitude, resulting in lower hemoglobin concentrations than are found in Han Chinese or South American highlanders. Other genes implicated include *EGLN1* and *PPARA*, which are also associated with hemoglobin concentration. Some evidence indicates that these genetic changes occurred within the past 3000 years, which is very fast in evolutionary terms. An intriguing question is whether the Sherpas' well-known mountain-climbing ability is partially attributable to their Tibetan ancestry, with overrepresentation of variants of *EPAS*. A striking recent finding is that some of these genetic characteristics may stem from those of Denisovan hominids who were contemporaries of the Neanderthals.

For acute altitude illness, a single gene variant is unlikely to be found, but differences in the susceptibility of individuals and populations, familial clustering of cases, and a positive association of some genetic variants all clearly support a role for genetics.

ACUTE MOUNTAIN SICKNESS AND HIGH-ALTITUDE CEREBRAL EDEMA

AMS is a neurologic syndrome characterized by nonspecific symptoms (headache, nausea, fatigue, and dizziness), with a paucity of physical findings, developing 6–12 h after ascent to a high altitude. AMS is a clinical diagnosis. For uniformity in research studies, the Lake Louise Scoring System, created at the 1991 International Hypoxia Symposium, is generally used without the sleep disturbance score. AMS must be distinguished from exhaustion, dehydration, hypothermia, alcoholic hangover, and hyponatremia. AMS and HACE are thought to represent opposite ends of a continuum of altitude-related neurologic disorders. HACE (but not AMS) is an encephalopathy whose hallmarks are ataxia and altered consciousness with diffuse cerebral involvement but generally without focal neurologic deficits. Progression to these signal manifestations can be rapid. Papilledema and, more commonly, retinal hemorrhages may develop. In fact, retinal hemorrhages occur frequently at ≥5000 m, even in individuals without clinical symptoms of AMS or HACE.

Risk Factors The most important risk factors for the development of altitude illness are the rate of ascent and a prior history of high-altitude illness. Exertion is a risk factor, but lack of physical fitness is not. An attractive but still speculative hypothesis proposes that AMS develops in people who have inadequate cerebrospinal capacity to buffer the brain swelling that occurs at high altitude. Children and adults seem to be equally affected, but people >50 years of age may be less likely to develop AMS than younger people. In general, there is no gender difference in AMS incidence. Sleep desaturation—a common phenomenon at high altitude—is associated with AMS. Debilitating fatigue consistent with severe AMS on descent from a summit is an important risk factor for death in mountaineers. A prospective study involving trekkers and climbers who ascended to altitudes between 4000 and 8848 m showed that high oxygen desaturation and low ventilatory response to hypoxia during exercise are independent predictors of severe altitude illness. However, because there may be a large overlap between groups of susceptible and nonsusceptible individuals, accurate cutoff values are hard to define. Prediction is made more difficult because the pretest probabilities of HAPE and HACE are low. Neck irradiation or surgery damaging the carotid bodies, respiratory tract infections, and dehydration appear to be other potential risk factors for altitude illness. Unless guided by clinical signs and symptoms, pulse oximeter readings alone on a trek should not be used to predict AMS.

Pathophysiology Hypobaric hypoxia is the main trigger for altitude illness. In established AMS, raised intracranial pressure, increased sympathetic activity, relative hypoventilation, fluid retention and redistribution, and impaired gas exchange have all been well noted; these factors may play an important role in the pathophysiology of AMS. Severe hypoxemia can lead to a greater than normal increase in cerebral blood flow. However, the exact mechanisms underlying AMS and HACE are unknown. Evidence points to a central nervous system process. Magnetic resonance imaging (MRI) studies have suggested that vasogenic (interstitial) cerebral edema is a component of the pathophysiology of HACE. In the setting of high-altitude illness, the MRI findings shown in **Fig. 475-1** are confirmatory of HACE, with increased signal in the white matter and particularly in the splenium of the corpus callosum. In addition, hemosiderin deposits in the corpus callosum have been characterized as long-lasting footprints of HACE. Quantitative analysis in an MRI study revealed that hypoxia is associated with mild vasogenic cerebral edema irrespective of AMS. This finding is in keeping with case reports of suddenly symptomatic brain tumors and of cranial nerve palsies without AMS at high altitudes. Vasogenic edema may become cytotoxic (intracellular) in severe HACE.

Impaired cerebral autoregulation in the presence of hypoxic cerebral vasodilation and altered permeability of the blood-brain barrier due to hypoxia-induced chemical mediators like histamine, arachidonic acid, and VEGF may all contribute to brain edema. In 1995, VEGF was first proposed as a potent promoter of capillary leakage in the brain at

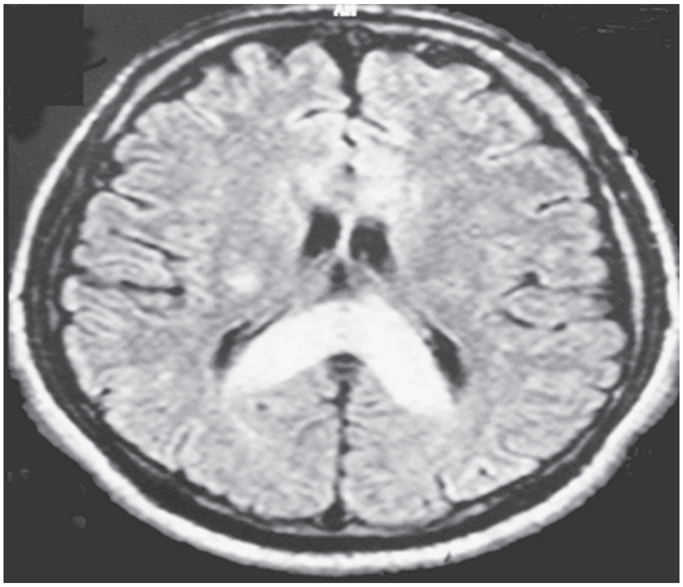

FIGURE 475-1 T2 magnetic resonance image of the brain of a patient with high-altitude cerebral edema (HACE) shows marked swelling and a hyperintense posterior body and splenium of the corpus callosum (area with dense opacity). The patient, a climber, went on to climb Mount Everest about 9 months after this episode of HACE. *(From B Basnyat et al: Clinical images. A mystery. Wilderness Environ Med 15: 53, 2004. Reused with permission from the Wilderness Medical Society. ©2004 Wilderness Medical Society.)*

high altitude, and studies in mice have borne out this role. Although studies of VEGF in climbers have yielded inconsistent results regarding its association with altitude illness, indirect evidence of a role for this growth factor in AMS and HACE comes from the observation that dexamethasone, when used in the prevention and treatment of these conditions, blocks hypoxic upregulation of VEGF. Other factors in the development of cerebral edema may be the release of calcium-mediated nitric oxide and neuronally mediated adenosine, which may promote cerebral vasodilation. Venous outflow obstruction resulting in increased brain capillary pressure is also thought to play an important role in the development of HACE. Lesions in the globus pallidum (which is sensitive to hypoxia) leading to Parkinson's disease have been reported to be complications of HACE.

The pathophysiology of the most common and prominent symptom of AMS—headache—remains unclear because the brain itself is an insensate organ; only the meninges contain trigeminal sensory nerve fibers. The cause of high-altitude headache is multifactorial. Various chemicals and mechanical factors activate a final common pathway, the trigeminovascular system. In the genesis of high-altitude headache, the response to nonsteroidal anti-inflammatory drugs and glucocorticoids provides indirect evidence for involvement of the arachidonic acid pathway and inflammation.

Prevention and Treatment (Table 475-1) Gradual ascent, with adequate time for acclimatization, is the best method for the prevention of altitude illness. Even though there may be individual variation in the rate of acclimatization, a conservative approach would be a graded ascent of ≤300 m from the previous day's sleeping altitude above 3000 m, and taking every third day of gain in sleeping altitude as an extra day for acclimatization is helpful. Spending one night at an intermediate altitude before proceeding to a higher altitude may enhance acclimatization and attenuate the risk of AMS. Another protective factor in AMS is recent high-altitude exposure; for example, the incidence and severity of AMS at 4300 m are reduced by 50% with an ascent after 1 week at an altitude ≥2000 m rather than with an ascent from sea level. However, regarding the benefits of acclimatization, clear-cut randomized studies are lacking. Repeated exposure at low altitudes to hypobaric or normobaric hypoxia is termed *preacclimatization*. Preacclimatization is gaining popularity with commercially available normobaric hypoxia "tents" used for weeks to months in preparation for the climb. However, current evidence has not shown significant effects of such technology.

TABLE 475-1 Management of Altitude Illness

CONDITION	MANAGEMENT
Acute mountain sickness (AMS), mild[a]	Discontinuation of ascent Treatment with acetazolamide (250 mg q12h) Descent[b]
AMS, moderate[a]	Immediate descent for worsening symptoms Use of low-flow oxygen if available Treatment with acetazolamide (250 mg q12h) and/or dexamethasone (4 mg q6h)[c] Hyperbaric therapy[d]
High-altitude cerebral edema (HACE)	Immediate descent or evacuation Administration of oxygen (2–4 L/min) Treatment with dexamethasone (8 mg PO/IM/IV; then 4 mg q6h) Hyperbaric therapy if descent is not possible
High-altitude pulmonary edema (HAPE)	Immediate descent or evacuation Minimization of exertion while patient is kept warm Administration of oxygen (4–6 L/min) to bring O_2 saturation to >90% Adjunctive therapy with nifedipine[e] (30 mg, extended-release, q12h) Hyperbaric therapy if descent is not possible

[a] Categorization of cases as mild or moderate is a subjective judgment based on the severity of headache and the presence and severity of other manifestations (nausea, fatigue, dizziness). [b] No fixed altitude is specified; the patient should descend to a point below that at which symptoms developed. [c] Acetazolamide treats and dexamethasone masks symptoms. For prevention (as opposed to treatment) of AMS, 125 mg of acetazolamide q12h or (when acetazolamide is contraindicated—e.g., in people with a history of sulfa anaphylaxis) 4 mg of dexamethasone q12h may be used. [d] In hyperbaric therapy (Fig. 475-2), the patient is placed in a portable altitude chamber or bag to simulate descent. [e] Nifedipine at this dose is also effective for the prevention of HAPE, as are tadalafil (10 mg twice daily), sildenafil (50 mg three times per day), and dexamethasone (8 mg twice daily). Preventative therapy should be continued for about 3 days after arriving at the target altitude. If prompt descent follows arrival at target altitude, continuation of preventative therapy is unnecessary.

Clearly, a flexible itinerary that permits additional rest days will be helpful. Sojourners to high-altitude locations must be aware of the symptoms of altitude illness and should be encouraged not to ascend further if these symptoms develop. Any hint of HAPE (see below) or HACE mandates descent. Proper hydration (but not overhydration) in high-altitude trekking and climbing, aimed at countering fluid loss due to hyperventilation and sweating, may play a role in avoiding AMS. Pharmacologic prophylaxis at the time of travel to high altitudes is warranted for people with a history of AMS or when a graded ascent and acclimatization are not possible—e.g., when rapid ascent is necessary for rescue purposes or when flight to a high-altitude location is required. Acetazolamide is the drug of choice for AMS prevention. It inhibits renal carbonic anhydrase, causing prompt bicarbonate diuresis that leads to metabolic acidosis and hyperventilation. Acetazolamide (125 mg twice daily), administered for 1 day before ascent and continued for about 3 days at the same altitude, is effective. Treatment can be restarted if symptoms return after discontinuation of the drug. Higher doses are not required. A meta-analysis limited to randomized controlled trials revealed that 125 mg of acetazolamide twice daily was effective in the prevention of AMS, with a relative-risk reduction of ~48% from values obtained with placebo. Paresthesia and a tingling sensation are common side effects of acetazolamide. Some other uncommon side effects are myopia and drowsiness. This drug is a nonantibiotic sulfonamide that has low-level cross-reactivity with sulfa antibiotics; as a result, severe reactions are rare. Dexamethasone (8 mg/d in divided doses) is also effective. A large-scale, randomized, double-blind, placebo-controlled trial in partially acclimatized trekkers clearly showed that *Ginkgo biloba* is ineffective in the prevention of AMS. In randomized studies, ibuprofen (600 mg three times daily) has been shown to be beneficial in the prevention of AMS. Recently, acetaminophen (1 g three times daily) was as effective as ibuprofen at the above dosage in a randomized, double-blind study, which did not have a placebo arm. However, more definitive studies are needed to clarify whether these medications mask the symptoms of altitude illness or whether they prevent the pathophysiology of AMS and assess the risk profile of side effects before these drugs can be routinely recommended for AMS prevention. Many drugs, including spironolactone, medroxyprogesterone, magnesium, calcium channel blockers, and antacids, confer no benefit in the prevention of AMS. Starkly conflicting results from a number of trials of inhaled budesonide for the prevention of AMS have recently been published, but, in all likelihood, the drug is ineffective. Similarly, no efficacy studies are available for coca leaves (a weak form of cocaine), which are offered to high-altitude travelers in the Andes, or for *soroche* pills, which contain aspirin, caffeine, and acetaminophen and are sold over the counter in Bolivia and Peru. Finally, a word of caution applies in the pharmacologic prevention of altitude illness. A fast-growing population of climbers in pursuit of a summit are injudiciously using prophylactic drugs such as glucocorticoids in an attempt to improve their performance; the outcome can be tragic because of potentially severe side effects of these drugs, especially if taken for a long duration.

For the treatment of mild AMS, rest alone with analgesic use may be adequate. Descent and the use of acetazolamide and (if available) oxygen are sufficient to treat most cases of moderate AMS. Even a minor descent (400–500 m) may be adequate for symptom relief. For moderate AMS or early HACE, dexamethasone (4 mg orally or parenterally) is highly effective. For HACE, immediate descent is mandatory. When descent is not possible because of poor weather conditions or darkness, a simulation of descent in a portable hyperbaric chamber (Fig. 475-2) can be very effective. Pressurization in the bag for 1–2 h often leads

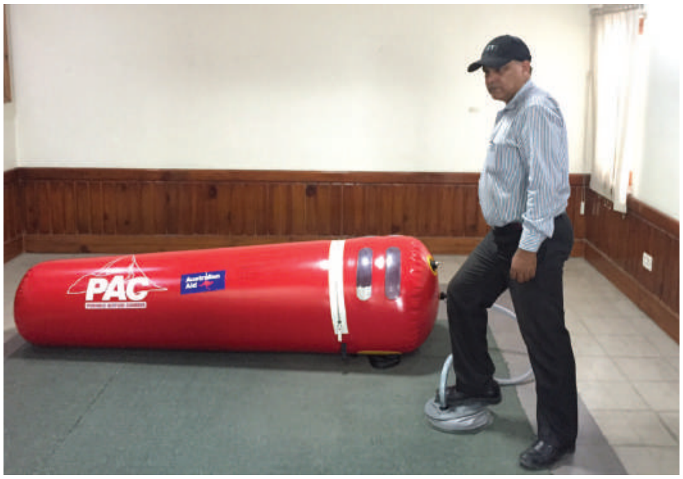

FIGURE 475-2 A hyperbaric bag. The cylindrical, portable (<7 kg) nylon bag has a one-way valve to prevent carbon dioxide buildup. A patient with severe acute mountain sickness (AMS), high-altitude cerebral edema (HACE), or high-altitude pulmonary edema (HAPE) is zipped inside the bag, which is continuously inflated with a foot pedal. The increased barometric pressure (2 psi) inside the bag simulates descent; for example, at 4250 m, the equivalent "elevation" inside the bag is ~2100 m. No supplemental oxygen is required.

to spectacular improvement and, like dexamethasone administration, "buys time." Thus, in certain high-altitude locations (e.g., remote pilgrimage sites), the decision to bring along the lightweight hyperbaric chamber may prove lifesaving. Like nifedipine, phosphodiesterase-5 inhibitors have no role in the treatment of AMS or HACE. Finally, short-term oxygen inhalation using small cannisters of oxygen or by visiting oxygen bars is unhelpful in the prevention of AMS.

■ HIGH-ALTITUDE PULMONARY EDEMA

Risk Factors and Manifestations Unlike HACE (a neurologic disorder), HAPE is primarily a pulmonary problem and therefore is not necessarily preceded by AMS. HAPE develops within 2–4 days after arrival at high altitude; it rarely occurs after >4 or 5 days at the same altitude, probably because of remodeling and adaptation that render the pulmonary vasculature less susceptible to the effects of hypoxia. A rapid rate of ascent, a history of HAPE, respiratory tract infections, and cold environmental temperatures are risk factors. Men are more susceptible than women. People with abnormalities of the cardiopulmonary circulation leading to pulmonary hypertension—e.g., mitral stenosis, primary pulmonary hypertension, and unilateral absence of the pulmonary artery—may be at increased risk of HAPE, even at moderate altitudes. Although patent foramen ovale, a common condition, is four times more common among HAPE-susceptible individuals than in the general population, there is no compelling evidence to suggest causal effect. Echocardiography is recommended when HAPE develops at relatively low altitudes (<3000 m) and whenever cardiopulmonary abnormalities predisposing to HAPE are suspected. The differential diagnosis of HAPE includes anxiety attack, pneumonia, pneumothorax, and pulmonary embolism.

The initial manifestation of HAPE may be a reduction in exercise tolerance greater than that expected at the given altitude. Although a dry, persistent cough may presage HAPE and may be followed by the production of blood-tinged sputum, cough in the mountains is almost universal and the mechanism is poorly understood. Tachypnea and tachycardia, even at rest, are important markers as illness progresses. Crackles may be heard on auscultation but are not diagnostic. HAPE may be accompanied by signs of HACE. Patchy or localized opacities (Fig. 475-3) or streaky interstitial edema may be noted on chest radiography. In the past, HAPE was mistaken for pneumonia due to the cold or for heart failure due to hypoxia and exertion. Kerley B lines or a bat-wing appearance are not seen on radiography. Electrocardiography may reveal right ventricular strain or even hypertrophy. Hypoxemia and respiratory alkalosis are consistently present in patients with HAPE. Alkalemia is often present, unless the patient is taking acetazolamide, in which case metabolic acidosis may supervene. Assessment of arterial blood gases is not necessary in the evaluation of HAPE; an oxygen saturation reading with a pulse oximeter is generally adequate. The existence of a subclinical form of HAPE has been suggested by an increased alveolar-arterial oxygen gradient in Everest climbers near the summit, but hard evidence correlating this abnormality with the development of clinically relevant HAPE is lacking. Comet-tail scoring—an ultrasound technique initially validated in cardiogenic pulmonary edema—has been used for evaluation of extravascular lung water at high altitude. However, B-lines are not just seen in patients with HAPE and are frequently detected in individuals who never go on to develop clinical HAPE. For this reason, comet-tail scoring is sensitive but not specific for HAPE, and clinical correlation is important.

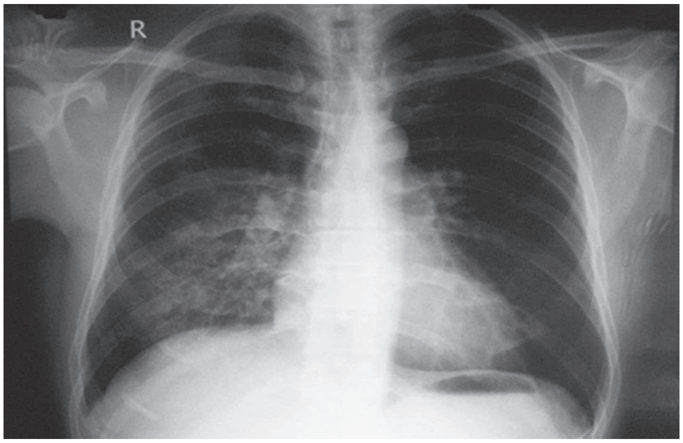

FIGURE 475-3 Chest radiograph of a patient with high-altitude pulmonary edema shows opacity in the right middle and lower zones simulating pneumonic consolidation. The opacity cleared almost completely in 2 days with descent and supplemental oxygen.

Pathophysiology HAPE is a noncardiogenic pulmonary edema with normal pulmonary artery wedge pressure. It is characterized by patchy pulmonary hypoxic vasoconstriction that leads to overperfusion in some areas. This abnormality leads in turn to increased pulmonary capillary pressure (>18 mmHg) and capillary "stress" failure. The exact mechanism for this hypoxic vasoconstriction is unknown. Endothelial dysfunction due to hypoxia may play a role by impairing the release of nitric oxide, an endothelium-derived vasodilator. At high altitude, HAPE-prone persons have reduced levels of exhaled nitric oxide. The effectiveness of phosphodiesterase-5 inhibitors in alleviating altitude-induced pulmonary hypertension, decreased exercise tolerance, and hypoxemia supports the role of nitric oxide in the pathogenesis of HAPE. One study demonstrated that prophylactic use of tadalafil, a phosphodiesterase-5 inhibitor, decreases the risk of HAPE by 65%. In contrast, the endothelium also synthesizes endothelin-1, a potent vasoconstrictor whose concentrations are higher than average in HAPE-prone mountaineers.

Exercise and cold lead to increased pulmonary intravascular pressure and may predispose to HAPE. In addition, hypoxia-triggered increases in sympathetic drive may lead to pulmonary venoconstriction and extravasation into the alveoli from the pulmonary capillaries. Consistent with this concept, phentolamine, which elicits α-adrenergic blockade, improves hemodynamics and oxygenation in HAPE more than do other vasodilators. The study of tadalafil cited above also investigated dexamethasone in the prevention of HAPE. Surprisingly, dexamethasone reduced the incidence of HAPE by 78%—a greater decrease than with tadalafil. Besides possibly increasing the availability of endothelial nitric oxide, dexamethasone may have altered the excessive sympathetic activity associated with HAPE: the heart rate of participants in the dexamethasone arm of the study was significantly lowered. Finally, people susceptible to HAPE also display enhanced sympathetic activity during short-term hypoxic breathing at low altitudes.

Because many patients with HAPE have fever, peripheral leukocytosis, and an increased erythrocyte sedimentation rate, inflammation has been considered an etiologic factor in HAPE. However, strong evidence suggests that inflammation in HAPE is an epiphenomenon rather than the primary cause. Nevertheless, inflammatory processes (e.g., those elicited by viral respiratory tract infections) do predispose persons to HAPE—even those who are constitutionally resistant to its development.

Another proposed mechanism for HAPE is impaired transepithelial clearance of sodium and water from the alveoli. β-Adrenergic agonists upregulate the clearance of alveolar fluid in animal models. In a single double-blind, randomized, placebo-controlled study of HAPE-susceptible mountaineers, prophylactic inhalation of the β-adrenergic agonist salmeterol reduced the incidence of HAPE by 50%. However, the dosage of salmeterol (125 μg twice daily) used was very high, which could result in excessive tachycardia and tremors. Other effects of β agonists may also contribute to the prevention of HAPE, and these findings are in keeping with the concept that alveolar fluid clearance may play a pathogenic role in this illness.

Prevention and Treatment (Table 475-1) Allowing sufficient time for acclimatization by ascending gradually (as discussed above for AMS and HACE) is the best way to prevent HAPE. Sustained-release nifedipine (30 mg), given twice daily, prevents HAPE in people who must ascend rapidly or who have a history of HAPE. Other drugs for the prevention of HAPE are listed in Table 475-1 (footnote e). Although dexamethasone is listed for prevention, its adverse effect profile requires close monitoring. Acetazolamide has been shown to blunt hypoxic pulmonary vasoconstriction in animal models, and this observation warrants further study in HAPE prevention. However, one large study failed to show a decrease in pulmonary vasoconstriction in partially acclimatized individuals given acetazolamide. Inhaled salmeterol is not recommended as clinical experience with this drug is limited at high altitude. Finally, potent diuretics like furosemide should be avoided in the treatment of HAPE. Early recognition is paramount in the treatment of HAPE, especially when it is not preceded by the AMS symptoms of headache and nausea. Fatigue and dyspnea at rest may be the only initial manifestations. Descent and the use of supplementary oxygen (aimed at bringing oxygen saturation to >90%) are the most effective therapeutic interventions. Exertion should be kept to a minimum, and the patient should be kept warm. Hyperbaric therapy (Fig. 475-2) in a portable altitude chamber may be lifesaving, especially if descent is not possible and oxygen is not available. Oral sustained-release nifedipine (30 mg twice daily) can be used as adjunctive therapy. No studies have investigated phosphodiesterase-5 inhibitors in the treatment of HAPE, but reports have described their use in clinical practice. The mainstays of treatment remain descent and (if available) oxygen.

In AMS, if symptoms abate (with or without acetazolamide), the patient may reascend gradually to a higher altitude. Unlike that in acute respiratory distress syndrome (another noncardiogenic pulmonary edema), the architecture of the lung in HAPE is usually well preserved, with rapid reversibility of abnormalities (Fig. 475-3). This fact has allowed some people with HAPE to reascend slowly after a few days of descent and rest. In HACE, reascent after a few days may not be advisable during the same trip.

OTHER HIGH-ALTITUDE PROBLEMS

Sleep Impairment The mechanisms underlying sleep problems, which are among the most common adverse reactions to high altitude, include increased periodic breathing; changes in sleep architecture, with increased time in lighter sleep stages; and changes in rapid eye movement sleep. Sojourners should be reassured that sleep quality improves with acclimatization. In cases where drugs do need to be used, acetazolamide (125 mg before bedtime) is especially useful because this agent decreases hypoxemic episodes and alleviates sleeping disruptions caused by excessive periodic breathing. Whether combining acetazolamide with temazepam or zolpidem is more effective than administering acetazolamide alone is unknown. In combinations, the doses of temazepam and zolpidem should not be increased by >10 mg at high altitudes. Limited evidence suggests that diazepam causes hypoventilation at high altitudes and therefore is contraindicated. For trekkers with obstructive sleep apnea who are using a continuous positive airway pressure (CPAP) machine, the addition of acetazolamide, which will decrease centrally mediated sleep apnea, may be helpful. There is evidence to show that obstructive sleep apnea at high altitude may decrease and “convert” to central sleep apnea.

Gastrointestinal Issues High-altitude exposure may be associated with increased gastric and duodenal bleeding, but further studies are required to determine whether there is a causal effect. Because of decreased atmospheric pressure and consequent intestinal gas expansion at high altitudes, many sojourners experience abdominal bloating and distension as well as excessive flatus expulsion. In the absence of diarrhea, these phenomena are normal, if sometimes uncomfortable. Accompanying diarrhea, however, may indicate the involvement of bacteria or *Giardia* parasites, which are common at many high-altitude locations in the developing world. Prompt treatment with fluids and empirical antibiotics may be required to combat dehydration in the mountains. Hemorrhoids are common on high-altitude treks; treatment includes hot soaks, application of hydrocortisone ointment, and measures to avoid constipation.

High-Altitude Cough High-altitude cough can be debilitating and is sometimes severe enough to cause rib fracture, especially at >5000 m. The etiology of this common problem is probably multifactorial. Although high-altitude cough has been attributed to inspiration of cold dry air, this explanation appears not to be sufficient by itself; in long-duration studies in hypobaric chambers, cough has occurred despite controlled temperature and humidity. The implication is that hypoxia also plays a role. Exercise can precipitate cough at high altitudes, possibly because of water loss from the respiratory tract. In general, infection does not seem to be a common etiology. Many trekkers find it useful to wear a balaclava to trap some moisture and heat. In most situations, cough resolves upon descent.

High-Altitude Neurologic Events Unrelated to “Altitude Illness” Transient ischemic attacks (TIAs) and strokes have been well described in high-altitude sojourners outside the setting of altitude sickness. However, these descriptions are not based on cause (hypoxia) and effect. In general, symptoms of AMS present gradually, whereas many of these neurologic events happen suddenly. The population that suffers strokes and TIAs at sea level is generally an older age group with other risk factors, whereas those so afflicted at high altitudes are generally younger and probably have fewer risk factors for atherosclerotic vascular disease. Other mechanisms (e.g., migraine, vasospasm, focal edema, hypocapneic vasoconstriction, hypoxia in the watershed zones of minimal cerebral blood flow, or cardiac right-to-left shunt) may be operative in TIAs and strokes at high altitude.

Subarachnoid hemorrhage, transient global amnesia, delirium, and cranial nerve palsies (e.g., lateral rectus palsy) occurring at high altitudes but outside the setting of altitude sickness have been well described. Syncope is common at moderately high altitudes, generally occurs shortly after ascent, usually resolves without descent, and appears to be a vasovagal event related to hypoxemia. Seizures occur rarely with HACE, but hypoxemia and hypocapnia, which are prevalent at high altitudes, are well-known triggers that may contribute to new or breakthrough seizures in predisposed individuals. Nevertheless, the consensus among experts is that sojourners with well-controlled seizure disorders can ascend to high altitudes.

Finally, persons with hypercoagulable conditions (e.g., antiphospholipid syndrome, protein C deficiency) who are asymptomatic at sea level may experience cerebral venous thrombosis (possibly due to enhanced blood viscosity triggered by polycythemia and dehydration) at high altitudes. Proper history taking, examination, and prompt investigations where possible will help define these conditions as entities separate from altitude sickness. Administration of oxygen (where available) and prompt descent are the cornerstones of treatment of most of these neurologic conditions.

Ocular Problems Ocular issues are common in sojourners to high altitudes. Hypoxemia induced by altitude leads to increased retinal blood flow, which can be visible as engorged retinal veins on ophthalmoscopic examination. Both high flow and hypoxemic vascular damage causing permeability have been implicated in a breakdown of the blood-retina barrier and the formation of retinal hemorrhages. Blot, dot, flame, and white-centered hemorrhages can be observed. These hemorrhages usually resolve spontaneously with descent, with only mild symptoms and no lasting visual damage in most healthy eyes. The exception is hemorrhage in the macular area. Macular hemorrhages can cause devastating initial visual loss, particularly if bilateral, and have been reported to cause permanently decreased vision in a few cases.

Stroke syndromes such as retinal vein occlusion, retinal artery occlusion, ischemic optic neuropathy, and cortical visual loss have all been reported. With unilateral vision loss, it is always important to check for a relative afferent pupillary defect. Increased hematocrit combined with dehydration may contribute to these maladies. Glaucomatous optic nerve damage may progress with hypoxemia of altitude. Acetazolamide is helpful both in combating the respiratory alkalosis that comes with increased ventilation at high altitude and in lowering the interocular pressure; its use should be considered in patients with stable controlled glaucoma. Macular degeneration and diabetic eye disease are not directly exacerbated by ascent to high altitude. Dry eye and solar damage to the cornea, known as "snow blindness," are common. Wearing of high-quality UV-blocking sunglasses, even on cloudy days, and attention to protecting and supplementing the tear film with artificial tear drops can greatly improve comfort and vision. Although modern refractive surgeries, such as photorefractive keratectomy (PRK) and laser in situ keratomileusis (LASIK), are stable at high altitude, patients who have undergone radial keratotomy should be cautioned that hypoxemia to the cornea can lead to swelling that shifts the refraction during ascent.

Psychological/Psychiatric Problems Delirium characterized by a sudden change in mental status, a short attention span, disorganized thinking, and an agitated state during the period of confusion has been well described in mountain climbers and trekkers without a prior history. In addition, anxiety attacks, often triggered at night by excessive periodic breathing, are well documented. The contribution of hypoxia to these conditions is unknown. Expedition medical kits need to include antipsychotic injectable drugs to control psychosis in patients in remote high-altitude locations.

PREEXISTING MEDICAL ISSUES

Because travel to high altitudes is increasingly popular, common conditions such as hypertension, coronary artery disease, and diabetes are more frequently encountered among high-altitude sojourners. This situation is of particular concern for the millions of elderly pilgrims with medical problems who visit high-altitude sacred areas (e.g., in the Himalayas) each year. In recent years, high-altitude travel has attracted intrepid trekkers who are taking immunosuppressive medications (e.g., kidney transplant recipients or patients undergoing chemotherapy). Recommended vaccinations and other precautions (e.g., hand washing) may be especially important for this group. Although most of these medical conditions do not appear to influence susceptibility to altitude illness, they may be exacerbated by ascent to altitude, exertion in cold conditions, and hypoxemia. Advice regarding the advisability of high-altitude travel and the impact of high-altitude hypoxia on these preexisting conditions is becoming increasingly relevant, but there are no evidence-based guidelines. In addition, recommendations made for relatively low altitudes (~3000 m) may not hold true for higher altitudes (>4000 m), where hypoxic stress is greater. Personal risks and benefits must be clearly thought through before ascent.

Hypertension At high altitudes, enhanced sympathetic activity may lead to a transient rise in blood pressure. Occasionally, nonhypertensive, healthy, asymptomatic trekkers have pathologically high blood pressure at high altitude that rapidly normalizes without medicines on descent. Sojourners should continue to take their antihypertensive medications at high altitudes. Importantly, hypertensive patients are not more likely than others to develop altitude illness. Because the probable mechanism of high-altitude hypertension is α-adrenergic activity, anti-α-adrenergic drugs such as prazosin have been suggested for symptomatic patients and those with labile hypertension. It is best to start taking the drug several weeks before the trip and to carry a sphygmomanometer if a trekker has labile hypertension. Sustained-release nifedipine may also be useful. A recent observational cohort study of 672 hypertensive and nonhypertensive trekkers in the Himalayas showed that most travelers, including those with well-controlled hypertension, can be reassured that their blood pressure will remain relatively stable at high altitude. Although blood pressure may be extremely elevated at high altitude in normotensive and hypertensive people, it is unlikely to cause symptoms.

Coronary Artery Disease Myocardial oxygen demand and maximal heart rate are reduced at high altitudes because the VO_2 max (maximal oxygen consumption) decreases with increasing altitude. This effect may explain why signs of cardiac ischemia or dysfunction usually are not seen in healthy persons at high altitudes. Asymptomatic, fit individuals with no risk factors need not undergo any tests for coronary artery disease before ascent. For persons with ischemic heart disease, previous myocardial infarction, angioplasty, and/or bypass surgery, an exercise treadmill test is indicated. A strongly positive treadmill test is a contraindication for high-altitude trips. Patients with poorly controlled arrhythmias should avoid high-altitude travel, but patients with arrhythmias that are well controlled with antiarrhythmic medications do not seem to be at increased risk. Sudden cardiac deaths are not noted with a greater frequency in the Alps than at lower altitudes; although sudden cardiac deaths are encountered every trekking season in the higher Himalayan range, accurate documentation is lacking.

Cerebrovascular Disease Patients with TIAs should avoid travel to high altitude for at least 3 months. Patients with known cerebral aneurysm should also avoid high-altitude travel because of possible rupture of the aneurysm due to increased cerebral blood flow at high altitude.

Migraine Trekkers with a history of migraine may have an increased likelihood of suffering from AMS and may also be predisposed to headaches including altered character of their migraine presenting with focal neurologic deficits. Oxygen inhalation may reduce AMS-triggered headache, whereas a migraine headache usually persists even after 10–15 min of oxygen inhalation.

Asthma Although cold air and exercise may provoke acute bronchoconstriction, asthmatic patients usually have fewer problems at high than at low altitudes, possibly because of decreased allergen levels and increased circulating catecholamine levels. Nevertheless, asthmatic individuals should carry all their medications, including oral glucocorticoids, with proper instructions for use in case of an exacerbation. Severely asthmatic persons should be cautioned against ascending to high altitudes.

Pregnancy In general, low-risk pregnant women ascending to 3000 m are not at special risk except for the relative unavailability of medical care in many high-altitude locations, especially in developing countries. Despite the lack of firm data on this point, venturing higher than 3000 m to altitudes at which oxygen saturation drops steeply seems unadvisable for pregnant women.

Obesity Although living at a high altitude has been suggested as a means of controlling obesity, obesity has also been reported to be a risk factor for AMS, probably because nocturnal hypoxemia is more pronounced in obese individuals. Hypoxemia may also lead to greater pulmonary hypertension, thus possibly predisposing the trekker to HAPE.

Sickle Cell Disease High altitude is one of the rare environmental exposures that occasionally provokes a crisis in persons with sickle cell anemia. Even when traversing mountain passes as low as 2500 m, people with sickle cell anemia have been known to have a vaso-occlusive crisis. Patients with known sickle cell anemia who need to travel to high altitudes should use supplemental oxygen and travel with caution. Thalassemia has not been known to cause problems at high altitude.

Diabetes Mellitus Well-controlled diabetes is not a contraindication for travel to high altitude. Most of the high-altitude diabetes advice is based on patients with type 1 diabetes and not type 2 diabetic patients with comorbidities. Altitude is known to both increase and decrease insulin sensitivity. An eye examination before travel may be useful. Insulin pumps are increasingly used, but bubble formation in the system may need to be closely monitored. For most diabetic patients, a continuous glucose monitor (CGM) is likely the best choice for glucose monitoring in the high-altitude environment, but a reliable glucometer should be carried by all diabetic patients. Regardless of the glucose monitoring modality selected, patients should be comfortable with the plan and transition appropriately. It is important for companions of diabetic trekkers to be fully aware of potential problems like hypoglycemia and have ready access to sweets.

Chronic Lung Disease Depending on disease severity and access to medical care, preexisting lung disease may not always preclude high-altitude travel. A proper pretravel evaluation must be conducted. Supplemental oxygen may be required if the predicted PaO_2 for the altitude is <50–55 mmHg. Preexisting pulmonary hypertension may also need to be assessed in these patients. If the result is positive, patients should be discouraged from ascending to high altitudes; if such travel is necessary, treatment with sustained-release nifedipine (30 mg twice a day) should be considered. Small-scale studies have revealed that when patients with bullous disease reach ~5000 m, bullous expansion and pneumothorax are not noted. Compared with information on chronic obstructive pulmonary disease, fewer data exist about the safety of travel to high altitude for people with pulmonary fibrosis, but acute exacerbation of pulmonary fibrosis has been seen at high altitude. A handheld pulse oximeter can be useful to check for oxygen saturation.

Chronic Kidney Disease Patients with chronic kidney disease can tolerate short-term stays at high altitudes, but theoretical concern persists about progression to end-stage renal disease. Acetazolamide, the drug most commonly used for altitude sickness, should be avoided by anyone with preexisting metabolic acidosis, which can be exacerbated by this drug. In addition, the acetazolamide dosage should be adjusted when the glomerular filtration rate falls to <50 mL/min, and the drug should not be used at all if this value falls to <10 mL/min.

Cirrhosis Of patients with cirrhosis, 16% may have portopulmonary arterial hypertension, and 32% may have hepatopulmonary syndrome; these conditions may be detrimental at high altitude as they may cause exaggerated hypoxemia. Thus, screening for these problems is important in cirrhotic patients planning a high-altitude trip. In addition, acetazolamide may be inadvisable in these patients as the drug may increase the risk of hepatic encephalopathy.

COVID-19 Evaluation is warranted before planned high-altitude travel in individuals who required care in an intensive care unit or suffered from myocarditis or arterial or venous thromboembolism. Preparticipation evaluation is also warranted in patients with persistent symptoms at least 2 weeks after a positive COVID-19 test or hospital discharge. Depending on the results of this evaluation, planned high-altitude travel may need to be modified or even deferred pending resolution of the identified abnormalities.

Dental Problems Air resulting from decay in the root system could expand on ascent and lead to increasing pain. A good dental checkup before a trekking or climbing trip may be prudent.

Malignancy Patients with current or previous malignancy may be affected in a variety of ways by altitude. Both radiation of the neck or paraganglioma/chemodectomas (which are seen at higher rates in residents of high-altitude areas) may result in carotid body dysfunction and reduced acclimatization. Pulmonary toxicities from chemotherapies (e.g., bleomycin) and radiation of the chest increase pulmonary artery pressures and impair lung function, theoretically increasing the risk of HAPE and decreasing exercise tolerance. Some chemotherapies such as bleomycin have traditionally been thought to increase susceptibility of lung tissue to oxygen toxicity and may complicate the use of supplemental oxygen, if needed. Cardiotoxic therapies for cancer decrease exercise tolerance and increase cardiovascular disease. Such patients will be at increased cardiovascular risk with strenuous activity or hypoxic stress that is expected at altitude. Patients with known metastasis to bone or those at high risk should be evaluated prior to expedition travel.

CHRONIC MOUNTAIN SICKNESS AND HIGH-ALTITUDE PULMONARY HYPERTENSION IN HIGHLANDERS

The largest populations of highlanders live in the South American Andes, the Tibetan Plateau, and parts of Ethiopia. Chronic mountain sickness (*Monge's disease*) is a disease in highlanders that is characterized by excessive erythrocytosis with moderate to severe pulmonary hypertension leading to cor pulmonale. This condition was originally described in South America and has also been documented in Colorado and in the Han Chinese population in Tibet; it is much less common in Tibetans or in Ethiopian highlanders. Migration to a low altitude results in the resolution of chronic mountain illness. Venesection and acetazolamide are helpful.

High-altitude pulmonary hypertension is also a subacute disease of long-term high-altitude residents. Unlike Monge's disease, this syndrome is characterized primarily by pulmonary hypertension (not erythrocytosis) leading to heart failure. Indian soldiers living at extreme altitudes for prolonged periods and Han Chinese infants born in Tibet have presented with the adult and infantile forms, respectively. High-altitude pulmonary hypertension bears a striking pathophysiologic resemblance to brisket disease in cattle. Descent to a lower altitude is curative.

FURTHER READING

Basnyat B: High altitude pilgrimage medicine. High Alt Med Biol 15:434, 2014.

Basnyat B, Murdoch D: High altitude illness. Lancet 361:1967, 2003.

Hillebrandt D et al: UIAA medical commission recommendations for mountaineers, hillwalkers, trekkers, and rock and ice climbers with diabetes. High Alt Med Biol 24:110, 2023.

Keyes LE et al: Blood pressure and altitude: An observational cohort study of hypertensive and nonhypertensive Himalayan trekkers in Nepal. High Alt Med Biol 18:267, 2017.

Luks AM et al: Return to high altitude after recovery from coronavirus disease 2019. High Alt Med Biol 22:119, 2021.

Luks AM et al: Wilderness Medical Society clinical practice guidelines for the prevention, diagnosis, and treatment of acute altitude illness: 2024 update. Wilderness Environ Med 35:2S, 2024.

Patterson RD, Roy S: High altitude illnesses, in *Mountain Emergency Medicine*, H Brugger et al (eds). Milan, Edra, 2021, pp 419–429.

476

Hyperbaric and Diving Medicine

Simon J. Mitchell, Richard E. Moon

Hyperbaric medicine is the treatment of certain health disorders using whole-body exposure to pressures >101.3 kPa (1 atmosphere or 760 mmHg). In practice, this invariably includes breathing oxygen during the exposure, thus the administration of hyperbaric oxygen therapy (HBO_2T). The Undersea and Hyperbaric Medical Society (UHMS) defines HBO_2T as "a medical procedure requiring a physician's prescription and oversight" in which "patients must have their entire body placed within a hard sided hyperbaric chamber that meets

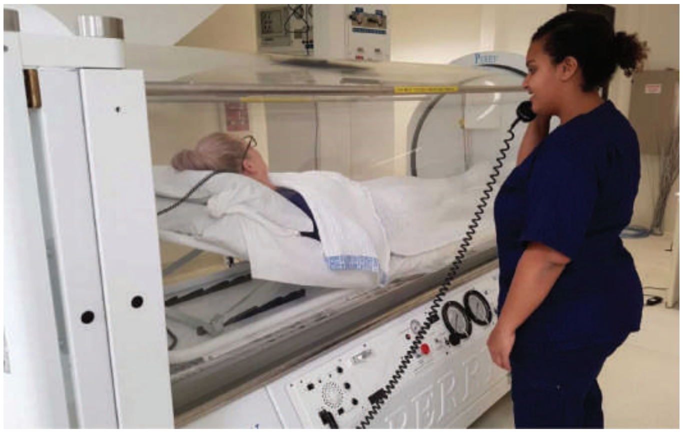

FIGURE 476-1 A monoplace chamber. The chamber is compressed with 100% oxygen. Although direct patient access is not possible, intravenous administration of fluids and drugs, ventilation, and invasive monitoring can be performed. *(Prince of Wales Hospital, Sydney.)*

the American Society of Mechanical Engineers and Pressure Vessels for Human Occupancy (ASME-PVHO-1) code, and the National Fire Protection Agency (NFPA 99) code and standards for hyperbaric chambers, at a pressure of not less than 2.0 atmospheres absolute (ATA) (202.65 kPa) while breathing physician prescribed medical grade oxygen for an amount of time that is typically between 90-120 minutes per treatment."

Historically, compression chambers were first used for the treatment of divers and compressed air workers suffering decompression sickness (DCS; "the bends"). Now they are used for treating a variety of problems, and relevant chambers are variously called a hyperbaric chamber, recompression chamber, or decompression chamber, depending on the clinical and historical context. They may be capable of compressing a single patient (a monoplace chamber) or multiple patients and attendants as required (a multiplace chamber) **(Figs. 476-1 and 476-2)**. In multiplace chambers, the chamber atmosphere is pressurized with air and the patient breathes oxygen through a special delivery system. The same may be true of monoplace chambers, although in some, the chamber environment is pressurized with oxygen and the patient simply breathes from that environment.

MECHANISMS OF HYPERBARIC OXYGEN THERAPY

In the late 1800s, the earliest systematic use of pressure as a therapeutic modality was in treating decompression sickness, a disorder caused by bubbles in blood or tissues. The use of pressure (recompression) to instantly reduce bubble size made mechanistic sense and met with obvious clinical success when instituted. In the 1960s, oxygen breathing during hyperbaric exposure was incorporated into recompression protocols to increase the diffusion gradient for inert gas from bubble to tissue to blood and alveoli. This too appeared to improve outcomes. Recompression of divers is further discussed under "Diving Medicine."

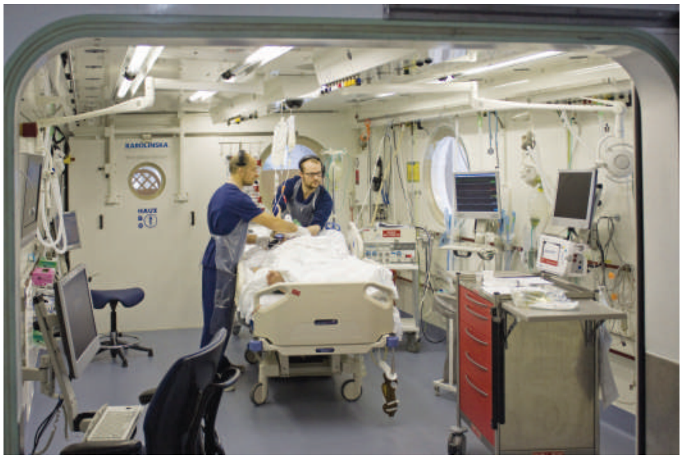

FIGURE 476-2 A multiplace chamber. Multiplace chambers are compressed with air; 100% oxygen is administered via head tent, mask, or endotracheal tube. These chambers can treat several patients simultaneously and can easily accommodate critically ill and intubated patients. *(Karolinska University Hospital.)*

Use of HBO$_2$T in nondiving disease focused initially on indications such as carbon monoxide poisoning, nonhealing wounds, and radiation tissue injury, where it was believed that hypoxia was a fundamental contributory problem and that hyperbaric oxygen could ameliorate this. Indeed, breathing oxygen at hyperbaric pressures not only ensures complete saturation of hemoglobin in blood but also results in markedly elevated quantities of dissolved oxygen in arterial plasma at a very high Po$_2$. Most HBO$_2$T regimens involve oxygen breathing at between 203 and 284 kPa (2 and 2.8 ATA). Breathing oxygen at 284 kPa (2.8 ATA) can theoretically increase the arterial Po$_2$ to more than ~270 kPa (~2025 mmHg) and the volume of oxygen dissolved in plasma to ~60 mL/L, with the latter being sufficient to sustain life without hemoglobin as proven in animal studies. The very high arterial Po$_2$ markedly improves oxygen diffusion distances through tissues. As with pressure and oxygen breathing for bubble disorders in diving, the use of markedly elevated plasma oxygen levels in treating hypoxic wounds and irradiated tissue also met with apparent success, initially reinforcing the belief that intermittent support of oxygen-dependent healing processes such as fibroblast replication, collagen deposition and cross-linking, and angiogenesis was the principal mechanism by which HBO$_2$T could be helpful.

As important as this mechanism may be, more recent research has revealed that intermittent hyperbaric hyperoxia has physical and biochemical/cell signaling effects that extend well beyond simple support of oxygen-dependent healing processes and the duration of the hyperbaric exposure. Indeed, there are pharmacologic effects that are profound and relatively long-lasting. Some of these are summarized in **Fig. 476-3**.

Thus, although removal from the hyperbaric chamber results in a rapid return of poorly vascularized tissues to their hypoxic state, even a single dose of HBO$_2$T produces changes in fibroblast, leukocyte and angiogenic functions, and antioxidant defenses that persist many hours after oxygen tensions are returned to pretreatment levels. Explanations for this effect focus on production of reactive oxygen species (ROS) such as superoxide (O_2^-), hydrogen peroxide (H_2O_2), and reactive nitrogen species (RNS) such as nitric oxide (NO) during exposure to hyperbaric hyperoxia. It appears that these species, often considered harmful in biological tissues, participate in diverse cell signaling pathways involved in production of a range of cytokines, angiogenic growth factors, and other modulators of inflammation and tissue repair whose net effect is to accelerate healing. Such mechanisms are complex and at times apparently paradoxical. For example, it is well established by in vitro and in vivo studies that notionally proinflammatory pulses of oxidative stress induced by hyperbaric hyperoxia actually result in subsequent anti-inflammatory effects such as reduced leukocyte β2 integrin adhesion molecule expression, an effect that may be protective when flaps and grafts are threatened or if endothelium is damaged by circulating bubbles.

Most of the important indications for HBO$_2$T are now recognized as benefiting in some way from activity in these cell signaling pathways. For example, when used to treat chronic hypoxic wounds, HBO$_2$T has been shown to enhance bacterial killing and phagocytosis by providing the oxygen substrate for macrophage activity; stimulate the synthesis of multiple angiogenic growth factors; inhibit leukocyte activation and adherence to damaged endothelium; and mobilize pluripotent vasculogenic progenitor cells from the bone marrow. The interactions between these mechanisms remain a very active field of investigation.

Figure 476-3 also depicts several other mechanisms of action that may be relevant in certain situations. Hyperoxia tends to induce vasoconstriction and that, along with the exaggerated standing osmotic gradient between an ultra-high Po$_2$ in arteriolar blood and the much lower Po$_2$ in tissues, may reduce tissue edema. Hyperoxia is also directly toxic or bacteriostatic to anaerobic bacteria.

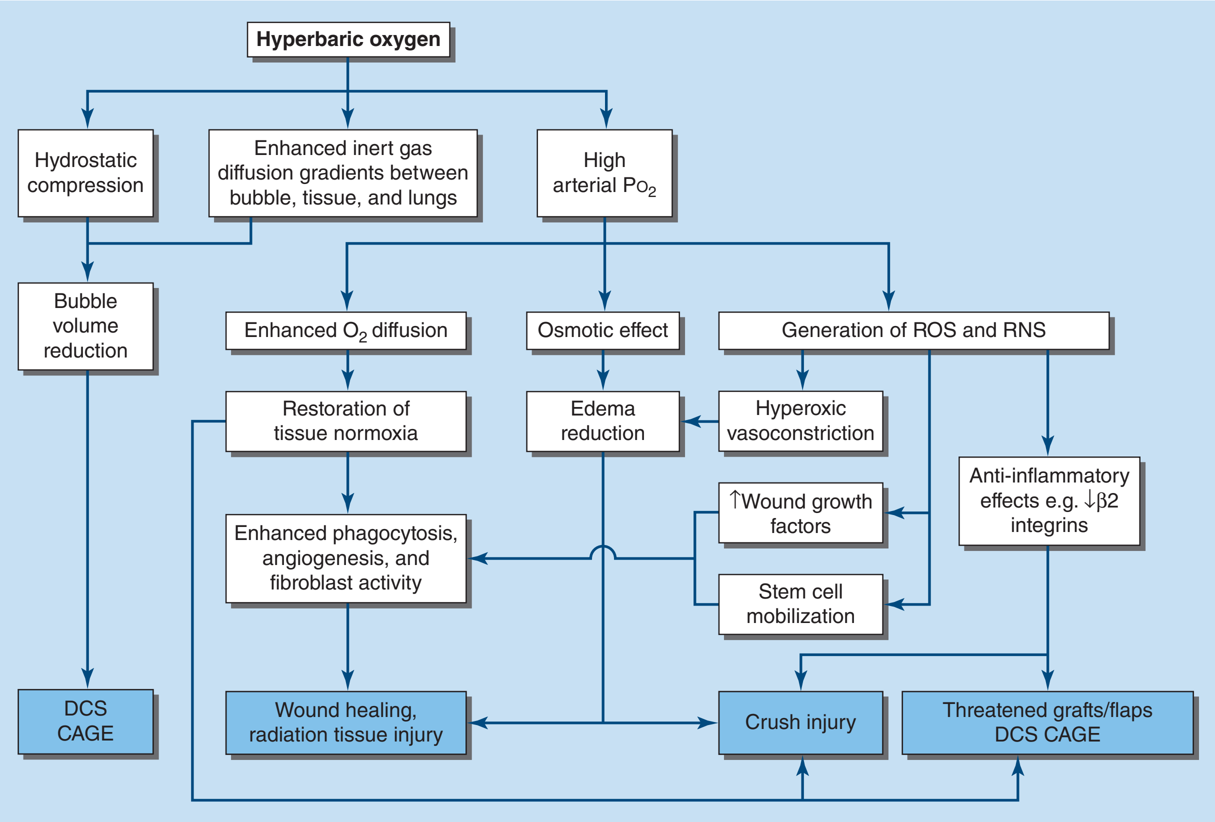
Hyperbaric oxygen
Hydrostatic compression
Enhanced inert gas diffusion gradients between bubble, tissue, and lungs
High arterial Po_2
Bubble volume reduction
Enhanced O_2 diffusion
Osmotic effect
Generation of ROS and RNS
Restoration of tissue normoxia
Edema reduction
Hyperoxic vasoconstriction
Anti-inflammatory effects e.g. ↓β2 integrins
↑Wound growth factors
Enhanced phagocytosis, angiogenesis, and fibroblast activity
Stem cell mobilization
DCS CAGE
Wound healing, radiation tissue injury
Crush injury
Threatened grafts/flaps DCS CAGE

FIGURE 476-3 Mechanisms of action of hyperbaric oxygen. There are many consequences of compression and oxygen breathing. The cell-signaling effects are potentially most important. Examples of indications are shown in the shaded boxes. CAGE, cerebral arterial gas embolism; DCS, decompression sickness; RNS, reactive nitrogen species; ROS, reactive oxygen species.

INDICATIONS FOR HYPERBARIC OXYGEN THERAPY

The "accepted" indications for HBO_2T are evolving and sometimes controversial. In 1977, the UHMS systematically examined claims for the use of HBO_2T in >100 disorders and found sufficient evidence to support routine use in only 12. The Hyperbaric Oxygen Therapy Committee of that organization has continued to update this list periodically (to the current list of 15) **(Table 476-1)** with an increasingly formalized system of appraisal for new indications and emerging evidence. Around the world, other relevant medical organizations have generally taken a similar approach. However, accepted indications vary considerably—particularly those recommended by hyperbaric medical societies in Russia and China where HBO_2T has gained much wider support than in the United States, Europe, and Australasia. Nevertheless, there are now >30 Cochrane reviews summarizing the randomized trial evidence for 27 putative indications, and numerous other systematic reviews, including attempts to examine the cost-effectiveness of HBO_2T. Following are short reviews of three important indications currently accepted by the UHMS, and a brief summary of several exploratory indications.

Late Radiation Tissue Injury Radiotherapy is a well-established treatment for suitable malignancies. In the United States alone, ~300,000 individuals annually will become long-term survivors of cancer treated by irradiation. Developments in radiotherapy techniques such as intensity-modulated radiation therapy and hypofractionated stereotactic radiation therapy have improved the precision of radiation delivery. This has allowed the use of higher doses of radiation with better cure rates, but the incidence of radiation-related complications has not changed much. Serious radiation-related complications developing months or years after treatment (late radiation tissue injury [LRTI]) will significantly affect between 5 and 10% of long-term survivors, although incidence varies widely with dose, age, and site. LRTI may involve bone or soft tissue and is most common in the head and neck, chest wall, breast, and pelvis (where it may involve the bowel or bladder).

PATHOLOGY AND CLINICAL COURSE One explanation for long-term injury to irradiated tissue is that initial radiation exposure produces fibrosis and an obliterative endarteritis from which the tissue may be incapable of recovering because of its hypovascular hypoxic nature. Indeed, its microvascularity may continue to deteriorate slowly over months to years with breakdown occasionally occurring spontaneously or when accelerated by an event like trauma or surgery. An alternative model of pathogenesis suggests that rather than a primary hypoxic state, the principal trigger is an overexpression of inflammatory cytokines that promote fibrosis, probably through oxidative stress and mitochondrial dysfunction, with tissue hypoxia as a secondary provocation. Ultimately, and once again often triggered by a further physical insult such as surgery or infection, there may be insufficient

TABLE 476-1 Current List of Indications for Hyperbaric Oxygen Therapy

1. Air or gas embolism (includes diving-related, iatrogenic, and accidental causes)
2. Carbon monoxide poisoning (including poisoning complicated by cyanide poisoning)
3. Clostridial myositis and myonecrosis (gas gangrene)
4. Crush injury, compartment syndrome, and acute traumatic ischemias
5. Decompression sickness
6. Arterial insufficiency including central retinal arterial occlusion and problem wounds
7. Severe anemia
8. Intracranial abscess
9. Necrotizing soft tissue infections (e.g., Fournier's gangrene)
10. Osteomyelitis (refractory to other therapy)
11. Delayed radiation injury (soft tissue injury and bony necrosis)
12. Skin grafts and flaps (compromised)
13. Acute thermal burn injury
14. Sudden sensorineural hearing loss
15. Avascular necrosis (aseptic osteonecrosis)

Source: The Undersea and Hyperbaric Medical Society (2024).

oxygen to sustain normal function, and the tissue becomes necrotic (radiation necrosis).

LRTI may be life-threatening and significantly reduce quality of life. Historically, the management of these injuries has been unsatisfactory. Conservative treatment is usually restricted to symptom management, whereas definitive treatment traditionally entails surgery to remove the affected part and extensive repair. Surgical intervention in an irradiated field is often disfiguring and associated with an increased incidence of delayed healing, breakdown of a surgical wound, or infection. HBO_2T may act by several mechanisms, including edema reduction, vasculogenesis, and enhancement of macrophage activity (Fig. 476-3). However, the ability of HBO_2T to promote angiogenesis in irradiated tissue represents the most important effect. The intermittent application of HBO_2T is the only intervention shown in multiple studies to increase the microvascular density in irradiated tissue in both animals and humans.

CLINICAL EVIDENCE The typical course of HBO_2T consists of 30–60 once-daily compressions to 202.6–243.1 kPa (2–2.4 ATA) for 1.5–2 h each session, often bracketed around surgical intervention if required. Although HBO_2T has been used for LRTI since at least 1975, most clinical studies have been limited to small case series or individual case reports. A 2023 Cochrane systematic review included 18 randomized controlled trials (RCTs) published since 1985 and was able to draw the following conclusions based on meta-analysis. Pooled analysis of studies across all LRTI types suggested HBO_2T improves tissue healing (risk ratio [RR] of healing with HBO_2T, 1.39; 95% confidence interval [CI], 1.02–1.89). In patients with osteoradionecrosis of the jaw or at risk of osteoradionecrosis of the jaw following planned surgery, HBO_2T reduced the risk of wound dehiscence (RR, 0.24; 95% CI, 0.06–0.94). The report also detailed individual randomized studies with notable positive results, particularly in regard to radiation proctitis and radiation cystitis. There were also negative randomized studies in both these indications.

Current UHMS guidelines (see Huang in "Further Reading") support use of HBO_2T for most forms of LRTI, a problem that is notoriously difficult to treat by other means. The American Society of Colon and Rectal Surgeons designated HBO_2T as a class 1B intervention (strongly indicated based on moderate-quality evidence) for radiation proctitis. A previous strongly positive RCT in treating bleeding radiation cystitis (RR for complete or significant improvement after HBO_2T, 3.63; 95% CI, 1.69–7.79) was recently corroborated by the Dartmouth Multicenter International Registry Initiative. In 370 bleeding patients treated with HBO_2T, the Radiation Therapy Oncology Group hematuria score fell from a median of 2 (interquartile range [IQR], 2) to 0 (IQR, 2) (see Moses et al. in "Further Reading"). Despite occasional positive case reports, results for radiation-induced neurologic injuries (brain and spinal cord) are less encouraging.

Selected Problem Wounds

A problem wound is any cutaneous ulceration that requires a prolonged time to heal, does not heal, or recurs. In general, wounds referred to hyperbaric facilities are those for which sustained attempts to heal by other means have failed. Problem wounds are common and constitute a significant health problem. It has been estimated that 1% of the population of industrialized countries will experience a leg ulcer at some time. The global cost of chronic wound care may be as high as U.S. $25 billion per year.

PATHOLOGY AND CLINICAL COURSE By definition, chronic wounds are indolent or deteriorating and resistant to the wide array of treatments applied. Although there are many contributing factors, these wounds most commonly arise in association with one or more comorbidities such as diabetes, peripheral venous or arterial disease, or prolonged pressure (decubitus ulcers). A common denominator among these various contributors is hypoxia. First-line treatments are aimed at correction of the underlying pathologies (e.g., vascular reconstruction, compression bandaging, normalization of nutritional deficiencies or normalization of blood glucose level). HBO_2T should be regarded as an adjunctive therapy to be applied simultaneously with amelioration of all modifiable risk factors for nonhealing wounds and good general wound care practice to maximize the chance of healing.

Not all problem wounds will respond to HBO_2T. A selection process based on responses to an oxygen challenge is often recommended. Wounds in hypoxic tissues often display poor or absent healing and have potential to benefit from HBO_2T depending on the cause, magnitude, and reversibility of hypoxia. Some causes of tissue hypoxia will be reversible with HBO_2T (e.g., microvascular disease and tissue edema), whereas some will respond better to surgical revascularization (e.g., critical arterial stenosis). When tissue hypoxia can be overcome by a high driving pressure of oxygen in the arterial blood, this can be demonstrated by measuring transcutaneous Po_2 using a modified transcutaneous Clarke electrode. It follows that many guidelines for patient selection for HBO_2T include the interpretation of transcutaneous oxygen tensions around the wound while breathing air and oxygen at pressure (Fig. 476-4). Other methods include simple ankle-brachial index and skin perfusion pressure.

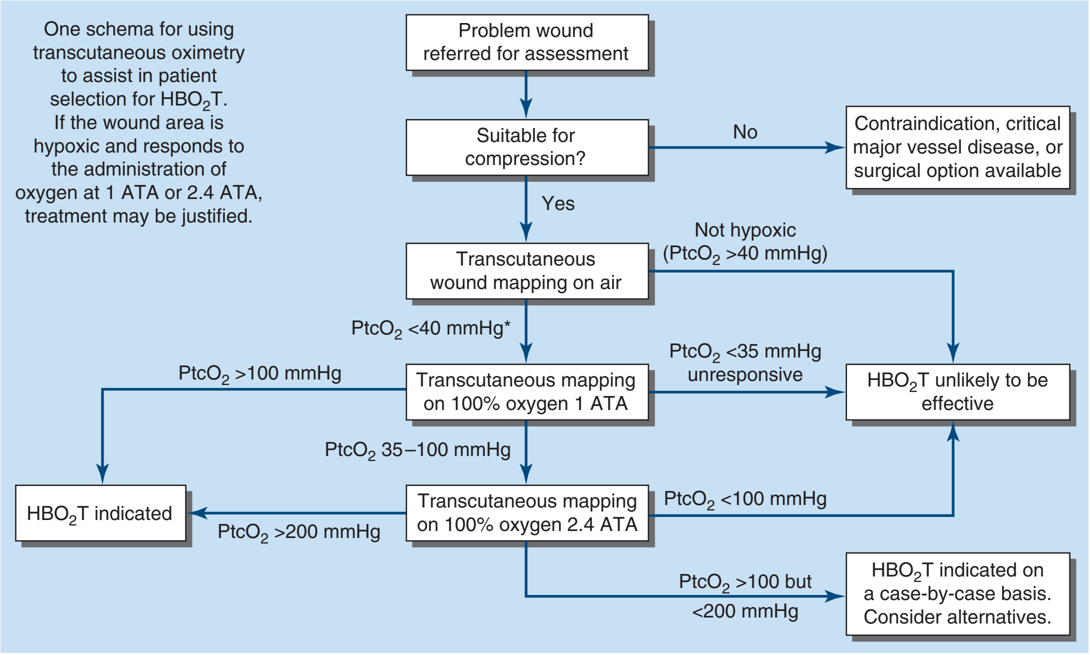
One schema for using transcutaneous oximetry to assist in patient selection for HBO_2T. If the wound area is hypoxic and responds to the administration of oxygen at 1 ATA or 2.4 ATA, treatment may be justified.
Problem wound referred for assessment
Suitable for compression?
No
Contraindication, critical major vessel disease, or surgical option available
Yes
Transcutaneous wound mapping on air
Not hypoxic ($PtcO_2$ >40 mmHg)
$PtcO_2$ <40 mmHg*
$PtcO_2$ >100 mmHg
Transcutaneous mapping on 100% oxygen 1 ATA
$PtcO_2$ <35 mmHg unresponsive
HBO_2T unlikely to be effective
$PtcO_2$ 35–100 mmHg
HBO_2T indicated
$PtcO_2$ >200 mmHg
Transcutaneous mapping on 100% oxygen 2.4 ATA
$PtcO_2$ <100 mmHg
$PtcO_2$ >100 but <200 mmHg
HBO_2T indicated on a case-by-case basis. Consider alternatives.

FIGURE 476-4 Determining suitability for hyperbaric oxygen therapy guided by transcutaneous oximetry around the wound bed. *In diabetic patients, <50 mmHg may be more appropriate. $PtcO_2$, transcutaneous oxygen pressure.

CLINICAL EVIDENCE The typical course of HBO$_2$T consists of 20–40 once-daily compressions to 2–2.4 ATA for 1.5–2 h each session but is highly dependent on the clinical response. Both retrospective and prospective cohort studies suggest that 6 months after a course of therapy, ~70% of indolent ulcers will be substantially improved or healed. Often these ulcers have been present for many months or years, suggesting the application of HBO$_2$T significantly improves the healing trajectory, either primarily or in conjunction with other strategies. A 2015 Cochrane review included 12 RCTs and concluded that the chance of a diabetic ulcer healing improved with HBO$_2$T (10 trials; RR, 2.35; 95% CI, 1.19–4.62). Although there was a trend toward reduced risk of major amputations with HBO$_2$T, it did not reach statistical significance (RR, 0.36; 95% CI, 0.11–1.18). A 2024 meta-analysis of 29 RCTs (see Chen et al. in "Further Reading") reported that in patients with diabetic foot ulcers treated with HBO$_2$T there was increased complete healing (odds ratio [OR], 2.8; 95% CI, 2.3–3.5) and reduced amputations (OR, 0.41; 95% CI, 0.18–0.95). Using data from a recent randomized trial, the cost per limb saved in treating severe (Wagner stage III/IV) diabetic foot ulcers with HBO$_2$T in the Netherlands was €19,005 (95% CI, –€18,487 to €264,334) (see Brouwer et al. in "Further Reading").

Carbon Monoxide Poisoning Carbon monoxide (CO) is a colorless, odorless gas formed during incomplete hydrocarbon combustion. Although CO is an essential endogenous neurotransmitter linked to NO metabolism and activity, it is also a leading cause of poisoning deaths and, in the United States alone, results in >50,000 emergency department visits per year and ~2000 deaths. Although there are large variations from country to country, about half of nonlethal exposures are due to self-harm. Accidental poisoning is commonly associated with defective or improperly installed heaters, house fires, and industrial exposures. The motor vehicle is by far the most common source of intentional poisoning.

PATHOLOGY AND CLINICAL COURSE The pathophysiology of CO exposure is incompletely understood. CO binds to hemoglobin with an affinity >200 times that of oxygen, directly reducing the oxygen-carrying capacity of blood and further promoting tissue hypoxia by shifting the oxyhemoglobin dissociation curve to the left. CO is also an anesthetic agent that inhibits evoked responses and narcotizes experimental animals in a dose-dependent manner. The associated loss of airway patency together with reduced oxygen carriage in blood may cause death from acute arterial hypoxia. CO also causes harm by other mechanisms including direct disruption of cellular oxidative processes, binding to heme proteins including cytochrome a/a3, and peroxidation of brain lipids.

The brain and heart are the most sensitive target organs, in part due to their high blood flow, poor tolerance of hypoxia, and high oxygen requirements. Minor exposures may be asymptomatic or present with vague constitutional symptoms such as headache, lethargy, and nausea, whereas higher doses may present with poor concentration and cognition, short-term memory loss, confusion, seizures, and loss of consciousness. Carboxyhemoglobin (COHb) levels on admission may confirm exposure but do not reliably reflect the severity or the prognosis of CO poisoning. Over the longer term, surviving patients commonly have neuropsychological sequelae that, for uncertain reasons, are sometimes delayed after an interval of days, weeks, or even months of apparent recovery. Motor disturbances, peripheral neuropathy, hearing loss, vestibular abnormalities, dementia, and psychosis have all been reported. Risk factors for poor outcome are age >35 years, exposure for >24 h, acidosis, and loss of consciousness.

CO poisoning is one of the longest-standing indications for HBO$_2$T, based traditionally on the obvious connection between exposure and tissue hypoxia and the ability of HBO$_2$T to rapidly overcome this hypoxia. Moreover, CO is eliminated rapidly via the lungs on application of HBO$_2$T, with a half-life of ~21 min breathing oxygen at 2.0 ATA versus 5.5 h or 71 min breathing air or oxygen (respectively) at sea level. In practice, it seems unlikely that HBO$_2$T could be delivered in time to prevent either acute hypoxic death or irreversible global cerebral hypoxic injury in severe cases. Thus, if HBO$_2$T is beneficial in CO poisoning, it must be by reducing the likelihood of persisting and/or delayed neurocognitive deficits through a mechanism(s) other than the simple reversal of arterial hypoxia due to high levels of COHb. There is evidence that relevant mechanisms include reversal of CO binding to cytochromes, upregulation of oxidative stress defenses, reduction in leukocyte adherence, impairment of lipid peroxidation, and others.

CLINICAL EVIDENCE The typical course of HBO$_2$T consists of two to three compressions to 2–2.8 ATA for 1.5–2 h each session. It is common for the first two compressions to be delivered within 24 h of the exposure. To date, there have been six RCTs of HBO$_2$T for CO poisoning, although only four have been reported in full. While a Cochrane review suggested there is insufficient evidence to confirm a beneficial effect of HBO$_2$T on the chance of persisting neurocognitive deficit following poisoning (34% of patients treated with oxygen at 1 atmosphere vs 29% of those treated with HBO$_2$T; OR, 0.78; 95% CI, 0.54–1.1), this may have more to do with poor reporting and inadequate follow-up than with evidence that HBO$_2$T is not effective. In the most methodologically rigorous of the analyzed studies (see Weaver et al. in "Further Reading"), at 6 weeks after poisoning, 46% of patients treated with normobaric oxygen alone had cognitive sequelae compared to 25% of those who received HBO$_2$T (p = .007; number needed to treat [NNT] = 5; 95% CI, 3–16). At 12 months, the difference remained significant (32 vs 18%; p = .04; NNT = 7; 95% CI, 4–124) despite considerable loss to follow-up.

On this basis, HBO$_2$T remains widely advocated for the routine treatment of patients with moderate to severe CO poisoning, in particular in those older than 35 years, presenting with a metabolic acidosis on arterial blood-gas analysis, exposed for lengthy periods, or with a history of unconsciousness.

EXPLORATORY INDICATIONS

Inflammatory Bowel Disease There is mounting evidence that HBO$_2$T enhances healing in flairs of fistulizing Crohn's disease or severe ulcerative colitis. Indeed, there are already two small positive RCTs along with numerous observational studies and case reports of apparent benefit. A multicenter RCT is underway. Benefit in these disorders may arise from enhancement of healing processes described earlier and/or anti-inflammatory activity.

Mild Traumatic Brain Injury After anecdotal reports of success, HBO$_2$T became a focus of interest for the treatment of veterans returning from Afghanistan with postconcussive syndromes. This led to the execution of three sham-controlled RCTs by the different branches of the U.S. military, which demonstrated improvements in patients receiving both HBO$_2$T and the sham. These improvements were considered due to placebo effects. However, a fourth study appeared to show benefit for HBO$_2$T independent of any placebo effect, and collective re-evaluation of all studies suggests that placebo effects account for some but not all of the measured improvements. It follows that there is ongoing interest in HBO$_2$T for mild traumatic brain injury, but a large definitive study is needed.

SARS-CoV-2 (COVID-19) Infection Although currently of dwindling relevance, it is notable that during the COVID-19 pandemic there were two RCTs of relatively small size and low quality that evaluated the benefit from HBO$_2$T in hypoxic patients at risk of requiring intubation and ventilation. The studies were underpowered for outcomes such as death or need for mechanical ventilation but did show a faster trajectory of recovery in patients treated with HBO$_2$T. Those studies along with several observational studies and case reports were recently summarized (see Boet et al. in "Further Reading"). To date, cases of pulmonary barotrauma precipitated by gas trapping in COVID pneumonia have not emerged, but little can be concluded about safety based on reporting of relatively low numbers of cases.

ADVERSE EFFECTS

HBO$_2$T is generally well tolerated and safe in clinical practice. Patients often experience an adverse event at some time during their treatment course, but most are mild and self-limiting. Adverse effects are associated with both alterations in pressure (barotrauma) and the administration of oxygen.

Barotrauma Barotrauma may occur when any noncompliant gas-filled space within the body does not equalize with environmental pressure during compression or decompression. Middle ear pain during compression (analogous to descent in an airplane) is the most common complication of HBO_2T, reported in ~30% of patients. Most result in either no or only trivial injury as a result, and prevention by slower compression and training in insufflation of the middle ear via the Eustachian tube is usually successful. Patients unable to insufflate the middle ear and unconscious patients typically require myringotomies or formal grommets across the tympanic membrane. Other less common sites for barotrauma of compression include the respiratory sinuses and dental caries. The lungs are potentially vulnerable to barotrauma of decompression (see "Diving Medicine" below), but the decompression following HBO_2T is so slow that pulmonary barotrauma is extremely rare and has only been seen when a patient has slowly communicating gas trapping lesions such as bullae.

Oxygen Toxicity The practical limit to the oxygen dose (whose components are inspired pressure and duration) in a single treatment session is cerebral oxygen toxicity. The most common acute manifestation is a tonic-clonic seizure, often preceded by facial twitching, anxiety, and agitation. The cause is thought to be an effect of ROSs on excitable neurons. Although clearly dose-dependent, onset is very variable both between individuals and within the same individual on different days. In large series reporting routine clinical hyperbaric practice (compressions to 2–2.4 ATA), the incidence of seizures is ~1–2/10,000 compressions.

Chronic oxygen poisoning most commonly manifests as myopic shift. This is due to alterations in the refractive index of the lens following oxidative damage that reduces the solubility of lenticular proteins, a process similar to that associated with senescent cataract formation. Up to 75% of patients show alteration in visual acuity after a course of 30 treatments at 202.6 kPa (2 ATA). Although most return to pretreatment values 6–12 weeks after cessation of treatment, a small proportion do not revert and may require a change in corrective lenses. A more rapid maturation of preexisting cataracts has occasionally been associated with HBO_2T. Although a theoretical problem, the development of pulmonary oxygen toxicity over time does not seem to be problematic in practice—probably due to the short and intermittent nature of the exposures.

CONTRAINDICATIONS

There are few absolute contraindications to HBO_2T. Untreated pneumothorax is considered a contraindication because intrapleural gas may expand on decompression and come under tension. Prior to any compression, patients with a pneumothorax should have a patent chest drain in place. The presence of other obvious risk factors for pulmonary gas trapping such as bullae should trigger a careful analysis of the risks of treatment versus benefit.

Recent systemic bleomycin or mitomycin C treatment deserves mention because of its association with a partially dose-dependent pneumonitis in ~20% of people. Patients treated with bleomycin, particularly those suffering pneumonitis, are often considered at risk of deterioration in ventilatory function following exposure to high oxygen tensions, even some years after treatment. Published experience suggests the relationship between distant bleomycin exposure and subsequent risk of pulmonary oxygen toxicity may have been overstated, and such patients have undergone uncomplicated HBO_2T. Nevertheless, any patient with a history of receiving these drugs (particularly recently and particularly if there was a consequent pneumonitis) should be carefully evaluated and counseled prior to exposure to HBO_2T.

CHALLENGES AND CONTROVERSIES IN HYPERBARIC MEDICINE

Despite an increased understanding of mechanisms and an improving evidence base, hyperbaric medicine has struggled to achieve widespread adoption in treating relevant disorders. There are several contributing factors, but high among them are a lack of relevant teaching at medical schools and a paucity of large well-conducted RCTs that clearly demonstrate efficacy in the targeted indications.

In turn, it should be appreciated that there are multiple factors that combine as an impediment to large RCTs in hyperbaric medicine. First, funding for clinical research has been difficult in an environment where the pharmacologic agent under study is abundant, cheap, and unpatentable. Second, some of the indications (such as necrotizing fasciitis and central retinal artery occlusion) are rare, sporadic, and acute. Conducting RCTs of treatment in such problems is notoriously difficult, especially where the putative treatment is available in only a small minority of hospitals. Not surprisingly, many of the accepted treatments applied in such problems are not supported by large RCTs. Third, the natural history of an evolving indication in HBO_2T is that treatment might first be tried sporadically in a small number of patients on the basis that its use makes biological sense. Unlike a new drug under development, there are no significant regulatory barriers to this. If apparently successful, then there is a risk that application of HBO_2T may become "accepted" before an RCT is proposed or organized, and clinicians then become reluctant to have patients randomized into a study. Fourth, a course of HBO_2T typically involves substantial commitment in time and logistical effort. This is a disincentive to recruitment of patients into a study with a high likelihood of being allocated to a placebo group. Finally, the potentially powerful placebo effect of HBO_2T is well recognized and highlights the importance of blinded sham controls.

Despite these impediments to RCTs, many have been successfully executed across the various indications for HBO_2T. Some have emerged long after the indication was accepted onto the UHMS list on the basis of lower-level evidence. A good example is crush injury and traumatic ischemia, for which the best supporting RCT has only recently emerged (see Miller et al. in "Further Reading"). Recognizing these challenges, however, an international clinical registry has also been created and is now beginning to produce useful outcome data that, at the very least, can be compared to the known natural history during standard management of the various monitored indications (see "Late Radiation Tissue Injury," above, and Moses et al. in "Further Reading").

Another challenge to the widespread acceptance of HBO_2T for established uses is the continuing misguided advocacy for hyperbaric therapy (sometimes using air breathing) as a panacea for treating numerous diseases, improving general well-being, and slowing aging. The prescription and delivery of HBO_2T requires no medical license in many jurisdictions. Some claims of efficacy refer to small supportive studies with a high risk of bias or placebo effect. Physicians should focus their practice on accepted or approved indications that are periodically updated (see Huang in "Further Reading").

DIVING MEDICINE

Underwater diving is both a popular recreational activity and a means of employment in a range of tasks from underwater construction to military operations. It is a complex activity with unique hazards and medical complications arising mainly as a consequence of the dramatic changes in pressure associated with both descent and ascent through the water column. For every 10.1-m increase in depth of seawater, the ambient pressure (P_{amb}) increases by 101.3 kPa (1 atmosphere) so that, for example, a diver at 20 m depth is exposed to a P_{amb} of 303.9 kPa (3 ATA), made up of 1 ATA due to atmospheric pressure and 2 ATA generated by the water column.

BREATHING EQUIPMENT

Most diving is undertaken using a self-contained underwater breathing apparatus (scuba) consisting of one or more cylinders of compressed gas connected to a pressure-reducing regulator and a demand valve activated by inspiratory effort. Some divers use rebreathers, which comprise a closed or semi-closed breathing circuit with a carbon dioxide scrubber and an oxygen addition system designed to maintain a safe inspired Po_2. Exhaled gas is recycled, and gas consumption is limited to little more than the oxygen metabolized by the diver.

Rebreathers are therefore popular for deep dives where expensive helium is included in the respired mix (see below). Occupational divers frequently use "surface supply" equipment where gas, along with other utilities such as communications and power, is supplied via an "umbilical" cable from the surface.

All these systems must supply gas to the diver at the P_{amb} of the surrounding water or inspiration would be impossible against the water pressure. For most recreational diving, the respired gas is air. Pure oxygen is rarely used because there is a dose-dependent risk (where "dose" is a function of exposure time and inspired Po_2) that oxygen may provoke seizures above an inspired Po_2 of 130 kPa (1.3 ATA). The maximum acceptable inspired Po_2 in diving is often considered to be 161 kPa (1.6 ATA), which would be achieved when breathing pure oxygen at 6 m or air at 66 m. This is a conspicuously lower Po_2 than routinely used for hyperbaric therapy (see earlier), reflecting a higher risk of oxygen toxic seizures during immersion and exercise. In order to maintain a safe Po_2 and avoid dangerous oxygen exposures, very deep diving requires the use of inspired oxygen fractions lower than in air (Fo_2 0.21), and divers tailor the oxygen content of their gases to remain within recommended exposure guidelines. Deep-diving gases include helium as a substitute for some or all of the nitrogen to reduce both the narcotic effect and high gas density that result from breathing nitrogen at high pressures.

SUITABILITY FOR DIVING

The most common reason for physician consultation in relation to diving is for the evaluation of suitability for diver training or continuation of diving after a health event. Occupational diver candidates are usually compelled to see doctors with specialist training in the field, both at entry to the industry and periodically thereafter, and their medical evaluations are usually conducted according to legally mandated standards. In contrast, in most jurisdictions, prospective recreational diver candidates simply complete a self-assessment medical questionnaire prior to diver training. If there are no positive responses, the candidate proceeds directly to training, but positive responses mandate the candidate see a clinician, often a primary care physician, for evaluation of the identified medical issue. In the modern era, such consultations have evolved from a simple proscriptive exercise of excluding those with potential contraindications to an approach in which each case is considered on its own merits and an individualized evaluation of risk is made. Such evaluations require integration of diving physiology, the impact of associated medical problems, and knowledge of the specific medical condition(s) of the candidate. A detailed discussion is beyond the scope of this chapter, but several important principles are outlined below.

There are three primary questions that should be answered in relation to any medical condition reported by a prospective diver: (1) Could the condition be exacerbated by diving? (2) Could the condition make a diving medical problem more likely? (3) Could the condition prevent the diver from meeting the functional requirements of diving? As examples of positive answers to these questions (respectively): epilepsy is usually considered to imply high risk because there are epileptogenic stimuli such as high inspired oxygen pressures encountered in diving that could make a seizure (and drowning) more likely; active asthma is considered to increase risk because it could predispose to air trapping and pulmonary barotrauma (see below); and ischemic heart disease increases risk because it could prevent a diver from exercising sufficiently to get out of a difficult situation such as being caught in a current. It can be a complex matter to recognize the relevant interactions between diving and medical conditions and to determine their impact on suitability for diving. There may follow an equally complex discussion about whether such interactions impart a disqualifying risk, and this may be influenced by the individual candidate's level of risk acceptance and the extent to which others involved (such as dive partners) might be affected. Guidelines are occasionally published on assessment of diving candidates with risk factors for important comorbidities like cardiovascular disease or who have suffered topical problems such as COVID-19 infection (see Sadler in "Further Reading"). Physicians interested in regularly conducting such evaluations should obtain relevant training, for which short courses are offered by specialist groups in most countries.

BAROTRAUMA

Barotrauma is essentially tissue injury arising as a result of ambient pressure changes. Middle-ear barotrauma (MEBT) in diving is similar to the problem that may occur during descent from altitude in an airplane, but difficulties with equalizing pressure in the middle ear are exaggerated underwater by both the rapidity and magnitude of pressure change as a diver descends or ascends. Failure to periodically insufflate the middle-ear spaces via the eustachian tubes during descent results in increasing pain. As the P_{amb} increases, the tympanic membrane may be bruised or even ruptured as it is pushed inward. Relative negative pressure in the middle ear results in engorgement of blood vessels in the surrounding mucous membranes and leads to effusion or bleeding, which can be associated with a *conductive* hearing loss **(Chap. 36)**. MEBT is much less common during ascent because expanding gas in the middle-ear space tends to open the eustachian tube automatically. Barotrauma may also affect the respiratory sinuses, although the sinus ostia are usually widely patent and allow automatic pressure equalization without the need for specific maneuvers. If equalization fails, pain usually results in termination of the dive. Difficulty with equalizing ears or sinuses may respond to oral or nasal decongestants.

Much less commonly, divers may suffer inner ear barotrauma (IEBT). Several explanations have been proposed, of which the most favored holds that forceful attempts to insufflate the middle-ear space by Valsalva maneuvers during descent result in transmission of pressure to the perilymph via the cochlear aqueduct and outward rupture of the round window, which is already under tension because of relative negative middle-ear pressure. The clinician should be alerted to possible IEBT after diving by a *sensorineural* hearing loss or true vertigo (which is often accompanied by nausea, vomiting, nystagmus, and ataxia). These manifestations can also occur in vestibulocochlear decompression sickness (DCS; see below) but should never be attributed to MEBT. Immediate review by an expert diving physician is recommended, and urgent referral to an otologist will often follow.

The lungs are also vulnerable to barotrauma but are at most risk during ascent. If expanding gas becomes trapped in the lungs as P_{amb} falls, this may rupture alveoli and associated vascular tissue. Gas trapping may occur if divers intentionally or involuntarily hold their breath during ascent or if there are bullae. The extent to which asthma predisposes to pulmonary barotrauma is debated, but the presence of active bronchoconstriction must increase risk. For this reason, asthmatics who regularly require bronchodilator medications or whose airways are sensitive to exercise or cold air are usually discouraged from diving. While possible consequences of pulmonary barotrauma include pneumothorax and mediastinal emphysema, the most feared is the introduction of gas into the pulmonary veins leading to cerebral arterial gas embolism (CAGE). Manifestations of CAGE include loss of consciousness, confusion, hemiplegia, visual disturbances, and speech difficulties appearing immediately or within minutes after surfacing. The management is the same as for DCS described below. The natural history of CAGE often includes substantial or complete resolution of symptoms early after the event. This is probably the clinical correlate of bubble involution and redistribution with consequent restoration of blood flow. Patients exhibiting such remissions should still be reviewed at specialist diving medical centers because secondary deterioration or re-embolization can occur. Unsurprisingly, these events can be misdiagnosed as typical strokes or transient ischemic attacks (TIAs) **(Chap. 438)** when patients are seen by clinicians unfamiliar with diving medicine. *All patients presenting with neurologic symptoms after diving should have their symptoms discussed with a specialist in diving medicine and be considered for recompression therapy.*

DECOMPRESSION SICKNESS

DCS is caused by the formation of bubbles from dissolved inert gas (usually nitrogen) during or after ascent (decompression) from a compressed gas dive. Bubble formation is also possible following decompression for extravehicular activity in space and with ascent to altitude in unpressurized aircraft. DCS in the latter scenarios is probably rare in comparison with diving, where the incidence is ~1 in 5000–10,000 recreational dives.

Breathing at elevated P_{amb} results in increased uptake of inert gas into blood and then into tissues. The rate at which tissue inert gas equilibrates with the inspired inert gas pressure is proportional to tissue blood flow and the blood-tissue partition coefficient for the gas. Similar factors dictate the kinetics of inert gas washout during ascent. If the rate of gas washout from tissues does not match the rate of decline in P_{amb}, then the sum of dissolved gas pressures in the tissue will exceed P_{amb}, a condition referred to as supersaturation. This is the prerequisite for bubbles to form during decompression, although other less well-understood factors are also involved. Deeper and longer dives result in greater inert gas absorption and greater likelihood of tissue supersaturation during ascent. Divers control their ascent for a given depth and time exposure using algorithms that often include periods where ascent is halted for a prescribed period at different depths to allow time for gas washout (decompression stops). Although a breach of these protocols increases the risk of DCS, adherence does not guarantee that it will be prevented. DCS should be considered in any diver manifesting postdive symptoms not readily explained by an alternative mechanism.

Bubbles may form within tissues themselves, where they cause symptoms by mechanical distraction of pain-sensitive or functionally important structures. They also appear in the venous circulation, almost certainly after forming in capillary beds as blood passes through supersaturated tissues. Some venous bubbles are tolerated without symptoms and are filtered from the circulation in the pulmonary capillaries. However, in sufficiently large numbers, these bubbles are capable of inciting inflammatory and coagulation cascades, damaging endothelium, activating formed elements of blood such as platelets, and causing symptomatic pulmonary vascular obstruction. Circulating bubbles can induce endothelial leak, intravascular hypovolemia, and high hematocrit. Moreover, if there is a right-to-left shunt through an atrial septal defect, patent foramen ovale (PFO), or an intrapulmonary shunt, then venous bubbles may enter the arterial circulation (25% of adults have a PFO). The risk of cerebral, spinal cord, inner ear, and skin manifestations appears higher in the presence of significant shunts, suggesting that these "arterialized" venous bubbles can cause harm, perhaps by disrupting flow in the microcirculation of target organs. Circulating microparticles, which are elevated in number and size after diving, are currently under investigation as indicators of decompression stress and as injurious agents in their own right. How they arise and their exact role in DCS remain unclear.

Table 476-2 lists manifestations of DCS grouped according to organ system. The majority of cases present with mild symptoms, including musculoskeletal pain, fatigue, and minor neurologic manifestations such as patchy paresthesias. Serious presentations are much less common. Pulmonary and cardiovascular manifestations can be life-threatening, and spinal cord involvement frequently results in permanent disability. Latency is variable. Serious DCS usually manifests within 60 min of surfacing, but mild symptoms may not appear for several hours. Symptoms arising >24 h after diving are very unlikely to be DCS, unless the diver is exposed to reduced ambient pressure such as during commercial air travel or surface travel at higher altitude. The presentation may be confusing and nonspecific, and there are no useful diagnostic investigations. Diagnosis is based on integration of findings from examination of the dive profile, the nature and temporal relationship of symptoms to diving, and the clinical examination. Some DCS presentations may be difficult to separate from CAGE following pulmonary barotrauma, but from a clinical perspective, the distinction is unimportant because the first aid and definitive management of both conditions are the same.

TABLE 476-2 Manifestations of Decompression Sickness

ORGAN SYSTEM	MANIFESTATIONS
Musculoskeletal	Limb pain
Neurologic	
Cerebral	Confusion
	Visual disturbances
	Speech disturbances
Spinal	Muscular weakness
	Paralysis
	Upper motor neuron signs
	Bladder and sphincter dysfunction
	Dermatomal sensory disturbances
	Abdominal pain
	Girdle pain
Vestibulocochlear	Hearing loss
	Vertigo and ataxia
	Nausea and vomiting
Peripheral	Patchy nondermatomal sensory disturbance
Pulmonary	Cough
	Dyspnea
Cardiovascular	Hemoconcentration
	Coagulopathy
	Hypotension
Cutaneous	Rash, itch
Lymphatic	Soft tissue edema, often relatively localized
Constitutional	Fatigue and malaise

TREATMENT

Decompression Sickness

First aid for either DCS or CAGE includes horizontal positioning, in the lateral decubitus position if consciousness is impaired; intravenous fluids (preferably glucose-free, isotonic intravenous, but if unavailable and the patient is sufficiently conscious, oral fluids); and sustained 100% oxygen administration. The latter accelerates inert gas washout from tissues and promotes resolution of bubbles. Definitive treatment of DCS or CAGE with recompression and hyperbaric oxygen is justified in most instances, although some mild or marginal DCS cases may be managed with first aid measures alone—an option that may be invoked by experienced diving physicians under various circumstances, but especially if evacuation for recompression is hazardous or extremely difficult. Long-distance evacuations are preferably undertaken using a helicopter flying at low altitude or a fixed-wing air ambulance pressurized to 1 ATA, although for serious cases, evacuation should not be delayed only because normobaric evacuation is not possible. Breathing oxygen en route is recommended. If oxygen toxicity is a concern during prolonged evacuation, expert opinion can provide guidance as to a schedule (see below).

Recompression reduces bubble volume in accordance with Boyle's law and increases the inert gas partial pressure difference between a bubble and surrounding tissue. At the same time, oxygen administration markedly increases the inert gas partial pressure difference between alveoli and tissue. The net effect is to significantly increase the rate of inert gas diffusion from bubble to tissue and tissue to blood, thus accelerating bubble resolution. Hyperbaric oxygen also helps oxygenate compromised tissues and may ameliorate some of the proinflammatory effects of bubbles. Various recompression protocols have been advocated, but there are no data that define the optimum approach. Recompression typically begins with oxygen administered at 2.8 ATA, the maximum pressure at which the risk of oxygen toxicity remains acceptable in a hyperbaric chamber. There follows a stepwise decompression over variable periods adjusted to symptom response. The most widely used algorithm is the U.S. Navy Table 6, whose shortest format lasts 4 h and 45 min. Typically, shorter follow-up recompressions are repeated daily while symptoms persist and appear responsive to treatment. Adjuncts to recompression include intravenous fluids and other supportive care as necessary. For cases of arterial gas embolism, evidence supports the use of intravenous lidocaine to achieve standard antiarrhythmic plasma levels. Occasionally, very sick divers require intubation, ventilation, and intensive care.

The presentation of sick divers to physicians or hospitals without diving medicine expertise creates a risk of misinterpretation of nonspecific manifestations and of consequent mistakes in diagnosis and management. Physicians finding themselves in this situation are strongly advised to expeditiously contact the 24-h worldwide diving emergency advisory service provided by the Divers Alert Network (DAN) at +1-919-684-9111.

ACKNOWLEDGMENT

The authors gratefully acknowledge the major contributions of Professor Michael Bennett (1956–2023) to previous editions of this chapter.

FURTHER READING

Boet S et al: Efficacy and safety of hyperbaric oxygen treatment to treat COVID-19 pneumonia: A living systematic review update. Diving Hyperb Med 52:126, 2022.

Brouwer R et al: Economic analysis of hyperbaric oxygen therapy for the treatment of ischaemic diabetic foot ulcers. Diving Hyperb Med 54:265, 2024.

Chen H-R et al: Application of hyperbaric oxygen therapy in diabetic foot ulcers: A meta-analysis. Int Wound J 21:e14621, 2024.

Huang E (ed): *Undersea and Hyperbaric Medicine Society Hyperbaric Medicine Indications Manual*, 15th ed. North Palm Beach, FL, Best Publishing, 2024.

Krzyżak J, Korzeniewski K: Medical assessment of fitness to dive. Parts I and II. Int Marit Health 72:36 and 115, 2021.

Lin ZC et al: Hyperbaric oxygen for late radiation tissue injury. Cochrane Database Syst Rev 8:CD005005, 2023.

Miller IL et al: Hyperbaric oxygen for lower limb trauma (HOLLT): An international multi-centre randomised clinical trial. Diving Hyperb Med 52:164, 2022.

Mitchell SJ: Decompression illness: A comprehensive overview. Diving Hyperb Med 54:1, 2024.

Mitchell SJ et al: Decompression illness. N Eng J Med 386:1254, 2022.

Moses RA et al: Patient reported outcome measures following hyperbaric oxygen therapy for radiation cystitis: Early results from the multicenter registry for hyperbaric oxygen therapy. J Urol 211:765, 2024.

Weaver LK et al: Hyperbaric oxygen for acute carbon monoxide poisoning. N Engl J Med 347:1057, 2002.

Wendling J et al (eds). *Medical Assessment for Work Under Pressure*, 2nd ed. London, International Marine Contractors Association, 2024.

Whelan HT (ed): *Hyperbaric Medicine Practice*, 5th ed. North Palm Beach, FL, Best Publishing, 2024.

477

Hypothermia and Peripheral Cold Injuries

Daniel F. Danzl

HYPOTHERMIA

Accidental hypothermia occurs when there is an unintentional drop in the body's core temperature below 35°C (95°F). At this temperature, many of the compensatory physiologic mechanisms that conserve heat begin to fail. *Primary accidental hypothermia* is a result of the direct exposure of a previously healthy individual to the cold. The mortality rate is much higher for patients who develop *secondary hypothermia* as a complication of a serious systemic disorder or injury.

TABLE 477-1 Risk Factors for Hypothermia

- Age extremes
  - Elderly
  - Neonates
- Environmental exposure
  - Occupational
  - Sports-related
  - Inadequate clothing
  - Immersion
- Toxicologic and pharmacologic
  - Ethanol
  - Anesthetics
  - Antipsychotics
  - Antidepressants
  - Anxiolytics
  - Benzodiazepines
  - Neuromuscular blockers
- Insufficient fuel
  - Malnutrition
  - Marasmus
  - Kwashiorkor
- Endocrine-related
  - Diabetes mellitus
  - Hypoglycemia
  - Hypothyroidism
  - Adrenal insufficiency
  - Hypopituitarism
- Neurologic
  - Cerebrovascular accident
  - Hypothalamic disorders
  - Parkinson's disease
  - Spinal cord injury
- Multisystemic
  - Trauma
  - Sepsis
  - Shock
  - Hepatic or renal failure
  - Carcinomatosis
  - Burns and exfoliative dermatologic disorders
  - Immobility or debilitation

CAUSES

Primary accidental hypothermia is geographically and seasonally pervasive. Although most cases occur in the winter months and in colder climates, this condition is surprisingly common in warmer regions as well. Multiple variables render individuals at the extremes of age—both the elderly and neonates—particularly vulnerable to hypothermia **(Table 477-1)**. The elderly have diminished thermal perception and are more susceptible to immobility, malnutrition, and systemic illnesses that interfere with heat generation or conservation. Dementia, psychiatric illness, and socioeconomic factors often compound these problems. Neonates have high rates of heat loss because of their increased surface-to-mass ratio and their lack of effective shivering and adaptive behavioral responses. At all ages, malnutrition can contribute to heat loss because of diminished subcutaneous fat and as a result of depleted energy stores used for thermogenesis.

Individuals whose occupations or hobbies entail extensive exposure to cold weather are at increased risk for hypothermia. Military history is replete with hypothermic tragedies. Hunters, sailors, skiers, and climbers also are at great risk of exposure, whether it involves injury, changes in weather, or lack of preparedness.

Ethanol causes vasodilation (which increases heat loss), reduces thermogenesis and gluconeogenesis, and may impair judgment or lead to obtundation. Some antipsychotics, antidepressants, anxiolytics, benzodiazepines, and other medications reduce centrally mediated vasoconstriction. Many hypothermic patients are admitted to intensive care because of drug overdose. Anesthetics can block shivering responses; these effects are compounded when patients are not insulated adequately in the operating or recovery units.

Several types of endocrine dysfunction cause hypothermia. Hypothyroidism—particularly when extreme, as in myxedema coma—reduces the metabolic rate and impairs thermogenesis and behavioral responses. Adrenal insufficiency and hypopituitarism also increase susceptibility to hypothermia. Hypoglycemia, most commonly caused by insulin or oral hypoglycemic agents, is associated with hypothermia, in part because of neuroglycopenic effects on hypothalamic function. Increased osmolality and metabolic derangements associated with uremia, diabetic ketoacidosis, and lactic acidosis can lead to altered hypothalamic thermoregulation.

Neurologic injury from trauma, cerebrovascular accident, subarachnoid hemorrhage, and a hypothalamic lesion increases susceptibility to hypothermia. Agenesis of the corpus callosum (*Shapiro's syndrome*) is one cause of episodic hypothermia. In this syndrome, profuse perspiration is followed by a rapid fall in temperature. Acute spinal cord injury disrupts the autonomic pathways that lead to shivering and will prevent cold-induced reflex vasoconstrictive responses.

Hypothermia associated with sepsis is a poor prognostic sign. Hepatic failure causes decreased glycogen storage and gluconeogenesis as well as a diminished shivering response. In acute myocardial infarction associated with low cardiac output, hypothermia may be reversed after adequate resuscitation. With extensive burns, psoriasis, erythrodermas, and other skin diseases, increased peripheral-blood flow leads to excessive heat loss.

THERMOREGULATION

Heat loss occurs through five mechanisms: radiation (55–65% of heat loss), conduction (10–15% of heat loss, increased in cold water), convection (increased in the wind), respiration, and evaporation; both of the latter two mechanisms are affected by the ambient temperature and the relative humidity.

The preoptic anterior hypothalamus normally orchestrates thermoregulation **(Chap. 20)**. The immediate defense of thermoneutrality is via the autonomic nervous system, whereas delayed control is mediated by the endocrine system. Autonomic nervous system responses include the release of norepinephrine, increased muscle tone, and shivering, leading to thermogenesis and an increase in the basal metabolic rate. Cutaneous cold thermoreception causes direct reflex vasoconstriction to conserve heat. Prolonged exposure to cold also stimulates the thyroid axis, leading to an increased metabolic rate.

CLINICAL PRESENTATION

In most cases of hypothermia, the history of exposure to environmental factors (e.g., prolonged exposure to the outdoors without adequate clothing) makes the diagnosis straightforward. In urban settings, however, the presentation is often more subtle, and other disease processes, toxin exposures, or psychiatric diagnoses should be considered. Predicting the core temperature based on the clinical presentation is very difficult.

After initial stimulation by hypothermia, there is progressive depression of all organ systems. The timing of the appearance of these clinical manifestations varies widely **(Table 477-2)**. Without knowing the core temperature, it can be difficult to interpret other vital signs. For example, tachycardia disproportionate to the core temperature suggests secondary hypothermia resulting from hypoglycemia, hypovolemia, or a toxin overdose. Because carbon dioxide production declines progressively, the respiratory rate should be low; persistent hyperventilation suggests a central nervous system (CNS) lesion or an organic acidosis. A markedly depressed level of consciousness in a patient with mild hypothermia suggests an overdose or CNS dysfunction due to infection or trauma.

Physical examination findings will also be altered by hypothermia. For instance, the assumption that areflexia is solely attributable to hypothermia can obscure the diagnosis of a spinal cord lesion. Patients with hypothermia may be confused or combative; these symptoms abate more rapidly with rewarming than with chemical or physical restraint. A classic example of maladaptive behavior in patients with hypothermia is paradoxical undressing, which involves the inappropriate removal of clothing in response to a cold stress. The cold-induced ileus and abdominal rectus spasm can mimic or mask the presentation of an acute abdomen **(Chap. 16)**.

When a patient in hypothermic cardiac arrest is first discovered, cardiopulmonary resuscitation (CPR) is indicated unless (1) a do-not-resuscitate status is verified, (2) obviously lethal injuries are identified, or (3) the depression of a frozen chest wall is not possible. Continuous CPR is normally recommended, and interruptions should be avoided if possible. In the field, when the core temperature is <28°C (82.4°F), intermittent CPR may also be effective.

As the resuscitation proceeds, the prognosis is grave if there is evidence of widespread cell lysis, as reflected by potassium levels >12 mmol/L (12 meq/L). Other findings that may preclude continuing resuscitation include a core temperature <10–12°C (<50–54°F), a pH <6.5, and evidence of intravascular thrombosis with a fibrinogen value <0.5 g/L (<50 mg/dL). The decision to terminate resuscitation before rewarming the patient past 33°C (91°F) should be predicated on the type and severity of the precipitants of hypothermia. Survival has occurred with a cardiac arrest time over 7 h. A history of asphyxia, as in an avalanche, with secondary cooling is the most important negative predictor of survival.

DIAGNOSIS AND STABILIZATION

Hypothermia is confirmed by measurement of the core temperature. When feasible, placement of a temperature probe in the lower third of the esophagus is preferable, and continuous monitoring essential. Ancillary rectal or bladder monitoring often lags behind core temperature changes.

Pulses may be undetectable in perfusing patients, and the amplitude of the QRS complex is decreased. Maximal amplification on the monitor is often helpful. Chest compressions, when not indicated, may convert a perfusing rhythm to a nonperfusing one. Bedside

TABLE 477-2 Physiologic Changes Associated with Accidental Hypothermia

SEVERITY	BODY TEMPERATURE	CENTRAL NERVOUS SYSTEM	CARDIOVASCULAR	RESPIRATORY	RENAL AND ENDOCRINE	NEUROMUSCULAR
Mild	35°C (95°F)–32.2°C (90°F)	Linear depression of cerebral metabolism; amnesia; apathy; dysarthria; impaired judgment; maladaptive behavior	Tachycardia, then progressive bradycardia; cardiac cycle prolongation; vasoconstriction; increase in cardiac output and blood pressure	Tachypnea, then progressive decrease in respiratory minute volume; declining oxygen consumption; bronchorrhea; bronchospasm	Diuresis; increase in catecholamines, adrenal steroids, triiodothyronine, and thyroxine; increase in metabolism with shivering	Increased preshivering muscle tone, then fatiguing
Moderate	<32.2°C (90°F)–28°C (82.4°F)	EEG abnormalities; progressive depression of level of consciousness; pupillary dilation; paradoxical undressing; hallucinations	Progressive decrease in pulse and cardiac output; increased atrial and ventricular arrhythmias; suggestive (J-wave) ECG changes	Hypoventilation: 50% decrease in carbon dioxide production per 8°C (17.6°F) drop in temperature; absence of protective airway reflexes	50% increase in renal blood flow; renal autoregulation intact; impaired insulin action	Hyporeflexia; diminishing shivering-induced thermogenesis; rigidity
Severe	<28°C (<82.4°F)	Loss of cerebrovascular autoregulation; decline in cerebral blood flow; coma; loss of ocular reflexes; progressive decrease in EEG abnormalities	Progressive decrease in blood pressure, heart rate, and cardiac output; reentrant dysrhythmias; maximal risk of ventricular fibrillation; asystole	Pulmonic congestion and edema; 75% decrease in oxygen consumption; apnea	Decrease in renal blood flow that parallels decrease in cardiac output; extreme oliguria; poikilothermia; 80% decrease in basal metabolism	No motion; decreased nerve-conduction velocity; peripheral areflexia; no corneal or oculocephalic reflexes

Abbreviations: ECG, electrocardiogram; EEG, electroencephalogram.

Source: From DF Danzl, RS Pozos: Accidental hypothermia. N Engl J Med 331:1756, 1994. Copyright © 1994 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

echocardiography and end-tidal carbon dioxide ($ETCO_2$) monitoring may confirm the perfusion status. If the core temperature is below 30°C (86°F) and there is no perfusion, consider up to three attempts at maximal power. If unsuccessful, rewarm to at least 30°C (86°F) before further attempts. Although cardiac pacing for hypothermic bradydysrhythmias is rarely indicated, the transthoracic technique is preferable. The J or Osborn wave at the junction of the QRS complex and ST segment suggests the diagnosis. Obvious J waves are routinely misdiagnosed by automated readings as injury current.

Supplemental oxygenation is always warranted, since tissue oxygenation is affected adversely by the leftward shift of the oxyhemoglobin dissociation curve. Pulse oximetry is often unreliable in patients with vasoconstriction. If protective airway reflexes are absent, gentle endotracheal intubation should be performed. Adequate preoxygenation will prevent ventricular arrhythmias.

Insertion of a gastric tube prevents dilation secondary to decreased bowel motility. Indwelling bladder catheters facilitate monitoring of cold-induced diuresis and can provide an ancillary approach for temperature monitoring. Dehydration is encountered commonly with chronic hypothermia, and most patients benefit from an intravenous or intraosseous crystalloid bolus. Normal saline is preferable to lactated Ringer's solution, as the liver in hypothermic patients inefficiently metabolizes lactate. The placement of a pulmonary artery catheter can cause perforation of the less compliant pulmonary artery. Insertion of a central venous catheter deeply into the cold right atrium should be avoided since this procedure, similar to transvenous pacing, can precipitate refractory arrhythmias.

Arterial blood gases should *not* be corrected for temperature **(Chap. 58)**. An uncorrected pH of 7.42 and a PCO_2 of 40 mmHg reflect appropriate alveolar ventilation and acid-base balance at any core temperature. Acid-base imbalances should be corrected gradually, since the bicarbonate buffering system is inefficient. A common error is overzealous hyperventilation in the setting of depressed CO_2 production. When the PCO_2 decreases by 10 mmHg at 28°C (82°F), it doubles the pH increase of 0.08 that occurs at 37°C (99°F). Consider $ETCO_2$ monitoring to prevent hyperventilation.

The severity of anemia may be underestimated because the hematocrit increases 2% for each 1°C drop in temperature. White blood cell sequestration and bone marrow suppression are common, potentially masking an infection. Although hypokalemia is more common in chronic hypothermia, hyperkalemia also occurs; the expected electrocardiographic changes are often obscured by hypothermia. Patients with renal insufficiency, metabolic acidosis, or rhabdomyolysis are at greatest risk for electrolyte disturbances.

Coagulopathies are common because cold inhibits the enzymatic reactions required for activation of the intrinsic cascade. In addition, thromboxane B_2 production by platelets is temperature dependent, and platelet function is impaired. The administration of platelets and fresh-frozen plasma is therefore not effective. Coagulation studies can be deceptively normal and contrast with the observed in vivo coagulopathy. This contradiction occurs because all coagulation tests are routinely performed at 37°C (99°F), and the enzymes are thus rewarmed.

■ REWARMING STRATEGIES

The key initial decision is whether to rewarm the patient passively or actively. *Passive external rewarming* simply involves covering and insulating the patient in a warm environment. With the head also covered, the rate of rewarming is usually 0.5°–2°C (1.10°–4.4°F) per hour. This technique is ideal for previously healthy patients who develop acute, mild primary accidental hypothermia. The patient must have sufficient glycogen to support endogenous thermogenesis.

The application of heat directly to the extremities of patients with chronic severe hypothermia should be avoided because it can induce peripheral vasodilation and precipitate core temperature "afterdrop," a response characterized by a continual decline in the core temperature after removal of the patient from the cold. Truncal heat application reduces the risk of afterdrop.

Active rewarming is necessary under the following circumstances: core temperature <32°C (<90°F) (*poikilothermia*), cardiovascular instability, age extremes, CNS dysfunction, hormone insufficiency, and suspicion of secondary hypothermia. *Active external rewarming* is best accomplished with forced-air heating blankets. Other options include devices that circulate water through external heat exchange pads, radiant heat sources, and hot packs. Monitoring a patient with hypothermia in a heated tub is extremely difficult. Electric blankets should be avoided because vasoconstricted skin is easily burned.

There are numerous widely available options for *active core rewarming*. Airway rewarming with heated humidified oxygen (40°–45°C [104°–113°F]) via mask or endotracheal tube is a convenient option. Although airway rewarming provides less heat than do some other forms of active core rewarming, it eliminates respiratory heat loss and adds 1°–2°C (2.2°–4.4°F) to the overall rewarming rate. Crystalloids should be heated to 40°–42°C (104°–108°F), but the quantity of heat provided is significant only during massive volume resuscitation. The most efficient method for heating and delivering fluid or blood is with a countercurrent in-line heat exchanger. Heated irrigation of the gastrointestinal tract or bladder transfers minimal heat because of the limited available surface area. These methods should be reserved for patients in cardiac arrest and then used in combination with all available active rewarming techniques.

If extracorporeal life support (ECLS) is unavailable, closed thoracic lavage is far more efficient in severely hypothermic patients with cardiac arrest. The hemithoraxes are irrigated through two inserted large-bore thoracostomy tubes. Thoracostomy tubes should not be placed in the left chest of a spontaneously perfusing patient for purposes of rewarming. Peritoneal lavage with the dialysate at 40°–45°C (104°–113°F) efficiently transfers heat when delivered through two catheters with outflow suction. Another option involves the use of endovascular temperature control catheters.

The HOPE (Hypothermia Outcome Prediction after Extracorporeal Life Support) score is a tool based on six covariates to help predict which patients may benefit from ECLS. ECLS options **(Table 477-3)** should be considered in severely hypothermic patients, especially those with *primary accidental hypothermia*. ECLS, including bypass, should be considered in nonperfusing patients without documented contraindications to resuscitation. Circulatory support may be the only effective option in patients with completely frozen extremities or those

TABLE 477-3 Options for Extracorporeal Life Support

EXTRACORPOREAL REWARMING TECHNIQUE	CONSIDERATIONS
Continuous venovenous (CVV) rewarming	Circuit: CV catheter to CV, dual-lumen CV, or peripheral catheter No oxygenator/circulatory support Flow rates 150–400 mL/min ROR 2°–3°C (4.4°–6.6°F)/h
Hemodialysis	Circuit: single- or dual-vessel cannulation Stabilizes electrolyte or toxicologic abnormalities or rhabdomyolysis Exchange cycle volumes 200–500 mL/min ROR 2°–3°C (4.4°–6.6°F)/h
Continuous arteriovenous rewarming (CAVR)	Circuit: percutaneous 8.5-Fr femoral catheters Requires systolic blood pressure of 60 mmHg No perfusionist/pump/anticoagulation Flow rates 225–375 mL/min ROR 3°–4°C (6.6°–8.8°F)/h
Cardiopulmonary bypass (CPB)	Circuit: full circulatory support with pump and oxygenator Perfusate-temperature gradient 5°–10°C (11°–22° F) Flow rates 2–7 L/min (average 3–4 L/min) ROR up to 9.5°C (20.9°F)/h
Venoarterial extracorporeal membrane oxygenation (VA-ECMO)	Decreased risk of post-rewarming cardiorespiratory failure Improved neurologic outcome

Abbreviations: CV, central venous; ROR, rate of rewarming.

with significant tissue destruction coupled with rhabdomyolysis. There is no evidence that extremely rapid rewarming improves survival in perfusing patients.

TREATMENT

Hypothermia

When a patient is hypothermic, target organs and the cardiovascular system respond minimally to most medications. Generally, medications are withheld below 30°C (86°F). In contrast to antiarrhythmics, low-dose vasopressor medications may improve the intra-arrest rates of return of spontaneous circulation. Because of increased binding of drugs to proteins as well as impaired metabolism and excretion, either a lower dose or a longer interval between doses should be used to avoid toxicity. As an example, the administration of repeated doses of digoxin or insulin would be ineffective while the patient is hypothermic, but the residual drugs would be potentially toxic during rewarming.

Achieving a mean arterial pressure of at least 60 mmHg should be an early objective. If the hypotension is disproportionate for temperature and does not respond to crystalloid/colloid infusion and rewarming, low-dose dopamine support (2–5 μg/kg per min) should be considered. Perfusion of the vasoconstricted cardiovascular system also may improve with low-dose IV nitroglycerin.

Atrial arrhythmias should be monitored initially without intervention, as the ventricular response should be slow and, unless preexistent, most will convert spontaneously during rewarming. The role of prophylaxis and treatment of ventricular arrhythmias is complex. Preexisting ventricular ectopy may be suppressed by hypothermia and reappear during rewarming. There is limited evidence to recommend any specific antiarrhythmic treatment.

Initiating empirical therapy for adrenal insufficiency usually is not warranted unless the history suggests steroid dependence or hypoadrenalism or efforts to rewarm with standard therapy fail. The administration of parenteral levothyroxine to euthyroid patients with hypothermia, however, is potentially hazardous. Because laboratory results can be delayed and confounded by the presence of the sick euthyroid syndrome **(Chap. 394)**, historic clues or physical findings suggestive of hypothyroidism should be sought. When myxedema is the cause of hypothermia, the relaxation phase of the Achilles reflex is prolonged more than is the contraction phase.

Hypothermia obscures most of the symptoms and signs of infection, notably fever and leukocytosis. Shaking rigors from infection may be mistaken for shivering. Except in mild cases, extensive cultures and repeated physical examinations are essential. Unless an infectious source is identified, empirical antibiotic prophylaxis is most warranted in the elderly, neonates, and immunocompromised patients.

FROSTBITE

Peripheral cold injuries include both freezing and nonfreezing cold injuries to tissue. Tissue freezes quickly when in contact with thermal conductors such as metal and volatile solutions. Other predisposing factors include constrictive clothing or boots, immobility, and vasoconstrictive medications. Frostbite occurs when the tissue temperature drops below 0°C (32°F). Ice-crystal formation subsequently distorts and destroys the cellular architecture. Once the vascular endothelium is damaged, stasis progresses rapidly to microvascular thrombosis. After the tissue thaws, there is progressive dermal ischemia. The microvasculature begins to collapse, arteriovenous shunting increases tissue pressures, and edema forms. Finally, thrombosis, ischemia, and superficial necrosis appear. The development of mummification and demarcation may take weeks to months.

CLINICAL PRESENTATION

The initial presentation of frostbite can be deceptively benign. The symptoms always include a sensory deficiency affecting light touch, pain, or temperature perception. The acral areas and distal extremities are the most common insensate areas. Some patients describe a clumsy or “chunk of wood” sensation in the extremity.

Deep frostbitten tissue can appear waxy, mottled, yellow, or violaceous-white. Favorable presenting signs include some warmth or sensation with normal color. The injury is often superficial if the subcutaneous tissue is pliable or if the dermis can be rolled over bony prominences.

Clinically, frostbite is superficial or deep. Superficial frostbite does not entail tissue loss but rather causes only anesthesia and erythema. The appearance of vesiculation surrounded by edema and erythema implies deeper involvement **(Fig. 477-1)**. Hemorrhagic vesicles reflect a serious injury to the microvasculature and indicate severe frostbite. Damages in subcuticular, muscular, or osseous tissues may result in amputation. An alternative classification establishes grades based on the location of presenting cyanosis; that is grade 1, absence of cyanosis; grade 2, cyanosis on the distal phalanx; grade 3, cyanosis up to the metacarpophalangeal (MP) joint; and grade 4 cyanosis proximal to the MP joint.

The two most common nonfreezing peripheral cold injuries are *chilblain* (*pernio*) and *immersion* (*trench*) *foot*. Chilblain results from neuronal and endothelial damage induced by repetitive exposure to damp cold above the freezing point. Young females, particularly those with a history of Raynaud’s phenomenon, are at greatest risk. Persistent vasospasticity and vasculitis can cause erythema, mild edema, and pruritus. Eventually plaques, blue nodules, and ulcerations develop. These lesions typically involve the dorsa of the hands and feet.

In contrast, nonfreezing cold injury includes trench foot and immersion foot that results from repetitive exposure to wet cold above the freezing point. The feet initially appear cyanotic, cold, and edematous. The subsequent development of bullae is often indistinguishable

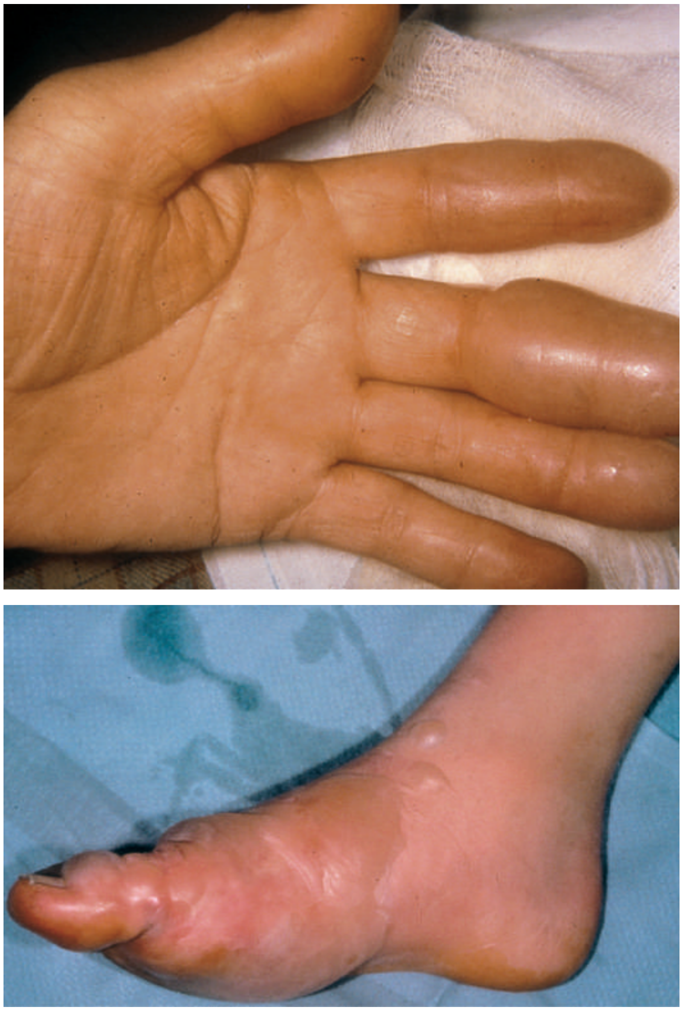

FIGURE 477-1 Frostbite with vesiculation, surrounded by edema and erythema.

from frostbite. This vesiculation rapidly progresses to ulceration and liquefaction gangrene. Consider administering amitriptyline early for pain. Patients with milder cases report hyperhidrosis, cold sensitivity, and painful ambulation for many years.

TREATMENT

Peripheral Cold Injuries

When frostbite accompanies hypothermia, hydration may improve vascular stasis. Frozen tissue should be thawed rapidly and completely by immersion in circulating water at 37°–40°C (99°–104°F) for 30–60 min and not by using hot air. Rapid rewarming often produces an initial hyperemia. The early formation of large clear distal blebs is more favorable than that of smaller proximal dark hemorrhagic blebs. A common error is the premature termination of thawing, since the reestablishment of perfusion is intensely painful. Parenteral narcotics will be necessary with deep frostbite. If cyanosis persists after rewarming, the tissue compartment pressures should be monitored carefully.

Many antithrombotic and vasodilatory treatment regimens have been evaluated. The prostacyclin analogue iloprost given within 48 h after rewarming is an option. There is no conclusive evidence that sympathectomy, steroids, calcium channel blockers, pentoxifylline, or hyperbaric oxygen salvages tissue.

Patients who have deep frostbite injuries with the potential for significant morbidity should be considered for intravenous or intraarterial thrombolytic therapy. Angiography, fluorescence microangiography, or pyrophosphate scanning may help evaluate the injury and monitor the progress of tissue plasminogen activator therapy. Heparin is recommended as adjunctive therapy. Intraarterial thrombolysis may reduce the need for digital and more proximal amputations when administered within 24 h of severe injuries. A treatment protocol for frostbite is summarized in **Table 477-4**.

Unless infection develops, any decision regarding debridement or amputation should generally be deferred. Angiography or technetium-99 bone scan may assist in the determination of surgical margins. Magnetic resonance angiography may also demonstrate the line of demarcation earlier than does clinical demarcation.

The most common symptomatic sequelae reflect neuronal injury and persistently abnormal sympathetic tone, including paresthesia, thermal misperception, and hyperhidrosis. Delayed findings include nail deformities, cutaneous carcinomas, and epiphyseal damage in children.

Management of the chilblain syndrome is usually supportive. With refractory perniosis, alternatives include nifedipine, steroids, and limaprost, a prostaglandin E_1 analogue.

TABLE 477-4 Treatment for Frostbite

BEFORE THAWING	DURING THAWING	AFTER THAWING
Remove from environment.	Consider parenteral analgesia and ketorolac.	Gently dry and protect part; elevate; place pledgets between toes, if macerated.
Prevent partial thawing and refreezing.	Administer ibuprofen (400 mg PO).	If clear vesicles are intact, aspirate sterilely; if broken, debride and dress with antibiotic or sterile aloe vera ointment.
Stabilize core temperature and treat hypothermia.	Immerse part in 37°–39°C (99°–102.2°F) (thermometer-monitored) circulating water containing an antiseptic soap until distal flush (10–45 min).	Leave hemorrhagic vesicles intact to prevent desiccation and infection.
Protect frozen part—no friction or massage.	Encourage patient to gently move part.	Continue ibuprofen (400–600 mg PO [12 mg/kg per day] q8 to 12h).
Address medical or surgical conditions.	If pain is refractory, reduce water temperature to 35°–37°C (95°–99°F) and administer parenteral narcotics.	Consider tetanus prophylaxis; elevate part. Administer hydrotherapy at 37°C (99°F). Consider dextran or phenoxybenzamine or, in severe cases, thrombolysis rt-PA (IV or intraarterial).

Abbreviation: rt-PA, recombinant tissue plasminogen activator.

FURTHER READING

Dow J et al: Wilderness medical society clinical practice guidelines for the out-of-hospital evaluation and treatment of accidental hypothermia: 2019 update. Wilderness Environ Med 30(4S):S47, 2019.

Pasquier M et al: Hypothermia outcome prediction after extracorporeal life support for hypothermic cardiac arrest patients: An external validation of the HOPE score. Resuscitation 139:321, 2019.

Pasquier M et al: On-site treatment of avalanche victims: Scoping review and 2023 recommendations of the international commission for mountain emergency medicine (ICAR MedCom). Resuscitation 184:109708, 2023.

Snijders B et al: Incidences of underlying causes of hypothermia in older patients in the emergency department: A systemic review. Eur Geriatr Med 14:411, 2023.

Takauji S et al: Outcome of extracorporeal membrane oxygenation use in severe accidental hypothermia with cardiac arrest and circulatory instability: A multicentre, prospective, observational study in Japan (ICE-CRASH study). Resuscitation 182:109663, 2023.

Teien HK et al: Training videos to prevent cold weather injuries. Int J Circumpolar Health 82:2195137, 2023.

Zafren K et al: Induced hypothermia to 4.2°C with neurologically intact survival: A forgotten case series. Wilderness Environ Med 31:367, 2020.

478

Heat-Related Illnesses

Daniel F. Danzl

Climate change is globally increasing heat-related morbidity and mortality. Extreme heat events are more frequent and are occurring in more widespread locations. Heat-related illnesses include a spectrum of disorders ranging from heat syncope, muscle cramps, and heat exhaustion to medical emergencies such as heatstroke. The core body temperature is normally maintained within a very narrow range. Although significant levels of hypothermia are tolerated **(Chap. 477)**, multiorgan dysfunction occurs rapidly at temperatures >40.5°C (104.9°C). In contrast to heatstroke, the far more common sign of fever reflects intact thermoregulation.

THERMOREGULATION

Humans are capable of significant heat generation. Strenuous exercise can increase heat generation twentyfold. The heat load from metabolic heat production and environmental heat absorption is balanced by a variety of heat dissipation mechanisms. These central integrative dissipation pathways are orchestrated by the central thermostat, which is located in the preoptic nucleus of the anterior hypothalamus. Efferent signals sent via the autonomic nervous system trigger cutaneous vasodilation and diaphoresis to facilitate heat loss.

Normally, the body dissipates heat into the environment via four mechanisms. The *evaporation* of skin moisture is the single most efficient mechanism of heat loss but becomes progressively ineffective as the relative humidity rises to >70%. The *radiation* of infrared electromagnetic energy directly into the surrounding environment occurs continuously. (Conversely, radiation is a major source of heat gain in hot climates.) *Conduction*—the direct transfer of heat to a cooler object—and *convection*—the loss of heat to air currents—become ineffective when the environmental temperature exceeds the skin temperature.

Factors that interfere with the evaporation of diaphoresis significantly increase the risk of heat illness. Examples include dripping of sweat off the skin, constrictive or occlusive clothing, dehydration, and excessive humidity. While air is an effective insulator, the thermal conductivity of water is 25 times greater than that of air at the same temperature. The *wet-bulb globe temperature* (WBGT) is a commonly used index to assess the environmental heat load. The WBGT is superior to the "heat index," which is only based on air temperature and humidity. The WBGT also incorporates radiant heat and wind speed.

The regulation of this heat load is complex and involves the central nervous system (CNS), thermosensors, and thermoregulatory effectors. The central thermostat activates the effectors that produce peripheral vasodilation and sweating. The skin surface is in effect the radiator and the principal location of heat loss, since skin blood flow can increase 25–30 times over the basal rate. This dramatic increase in skin blood flow, coupled with the maintenance of peripheral vasodilation, efficiently radiates heat. At the same time, there is a compensatory vasoconstriction of the splanchnic and renal beds.

Acclimatization to heat reflects a constellation of physiologic adaptations that permit the body to lose heat more efficiently. This process often requires 1 to several weeks of exposure and work in a hot environment. During acclimatization, the thermoregulatory set point is altered, and this alteration affects the onset, volume, and content of diaphoresis. The threshold for the initiation of sweating is lowered, and the amount of sweat increases, with a lowered salt concentration. Sweating rates can be 1–2 L/h in acclimated individuals during heat stress. Plasma volume expansion also occurs and improves cutaneous vascular flow. The heart rate lowers, with a higher stroke volume. After the individual leaves the hot environment, improved tolerance to heat stress dissipates rapidly, the plasma volume decreases, and deacclimatization occurs within weeks.

PREDISPOSING FACTORS AND DIFFERENTIAL DIAGNOSIS

When there is an excessive heat load, unacclimated individuals can develop a variety of heat-related illnesses. Heat waves exacerbate the mortality rate, particularly among the elderly and among persons lacking adequate nutrition and access to air-conditioned environments. Secondary vascular events, including cerebrovascular accidents and myocardial infarctions, occur at least 10 times more often in conditions of extreme heat.

Exertional heat illness continues to occur when laborers, military personnel, or athletes exercise strenuously in the heat. In addition to the very young and very old, preadolescents and teenagers are at risk since they may use poor judgment when vigorously exercising in high humidity and heat. Other risk factors include obesity, poor conditioning with lack of acclimatization, and mild dehydration.

Cardiovascular inefficiency is a common feature of heat illness. Any physiologic or pharmacologic impediment to cutaneous perfusion impairs heat loss. Many patients are unaware of the heat risk associated with their medications. Anticholinergic agents impair sweating and blunt the normal cardiovascular response to heat. Phenothiazines and heterocyclic antidepressants also have anticholinergic properties that interfere with the function of the preoptic nucleus of the anterior hypothalamus due to central depletion of dopamine.

Calcium channel blockers, beta blockers, and various stimulants also inhibit sweating by reducing peripheral blood flow. To maintain the mean arterial blood pressure, increased cardiac output must be capable of compensating for progressive dehydration. A variety of stimulants and substances of abuse also increase muscle activity and heat production.

Careful consideration of the differential diagnosis is important in the evaluation of a patient for a potential heat-related illness. The clinical setting may suggest other etiologies, such as malignant hyperthermia after general anesthesia. Neuroleptic malignant syndrome can be triggered by certain antipsychotic medications, including selective serotonin reuptake inhibitors. A variety of infectious and endocrine disorders as well as toxicologic or CNS etiologies may mimic heatstroke **(Table 478-1)**.

MINOR HEAT-EMERGENCY SYNDROMES

Heat edema is characterized by mild swelling of the hands, feet, and ankles during the first few days of significant heat exposure. The principal mechanism involves cutaneous vasodilation and pooling of interstitial fluid in response to heat stress. Heat also increases the secretion of antidiuretic hormone and aldosterone. Systemic causes of edema, including cirrhosis, nephrotic syndrome, and congestive heart failure, can usually be excluded by the history and physical examination. Heat

TABLE 478-1 Heat-Related Illness: Predisposing Factors and Differential Diagnosis

ILLNESS	PREDISPOSING FACTORS
Cardiovascular inefficiency	Age extremes Beta/calcium channel blockade Congestive heart failure Dehydration Diuresis Obesity Poor physical fitness
Central nervous system illness	Cerebellar injury Cerebral hemorrhage Hypothalamic cerebrovascular accident Psychiatric disorders Status epilepticus
Impaired heat loss	Antihistamines Heterocyclic antidepressants Occlusive clothing Skin abnormalities
Endocrine and immune-related illness	Diabetic ketoacidosis Multiple-organ dysfunction syndrome Pheochromocytoma Systemic inflammatory response syndrome Thyroid storm
Excessive heat load	Environmental conditions Exertion Fever Hypermetabolic state Lack of acclimatization
Infectious illness	Cerebral abscess Encephalitis Malaria Meningitis Sepsis syndrome Tetanus Typhoid
Toxicologic illness	Amphetamines Anticholinergic toxidrome Cocaine Dietary supplements Hallucinogens Malignant hyperthermia Neuroleptic malignant syndrome Salicylates Serotonin syndrome Strychnine Sympathomimetics Withdrawal syndromes (ethanol, hypnotics)

edema generally resolves without treatment in several days. Simple leg elevation or compression stockings will usually suffice. Diuretics are *not* effective and, in fact, predispose to volume depletion and the development of more serious heat-related illnesses.

Prickly heat (*miliaria rubra*, *lichen tropicus*) is a maculopapular, pruritic, erythematous rash that commonly occurs in clothed areas. Blockage of the sweat pores by debris from macerated stratum corneum causes inflammation in the sweat ducts. As the ducts dilate, they rupture and produce superficial vesicles. The predominant symptom is pruritus. In addition to antihistamines, chlorhexidine in a light cream or lotion provides some relief. In adults, localized areas may benefit from 1% salicylic acid TID, with caution taken to avoid salicylate intoxication. Clothing with breathable fabric should be clean and loose fitting, and activities or environments that induce diaphoresis should be avoided.

Heat syncope (exercise-associated collapse) can follow endurance exercise or occur in the elderly. Other common clinical scenarios include prolonged standing while stationary in the heat and sudden standing after prolonged exposure to heat. Heat stress routinely causes relative volume depletion, decreased vasomotor tone, and peripheral vasodilation. The cumulative effect of this decrease in venous return is postural hypotension, especially in nonacclimated elderly individuals. Many of those affected also have comorbidities. Therefore, other cardiovascular, neurologic, and metabolic causes of syncope should be considered. After removal from the heat source, most patients will recover promptly with cooling and rehydration.

Hyperventilation tetany occurs in some individuals when exposure to heat stimulates hyperventilation, producing respiratory alkalosis, paresthesia, and carpopedal spasm. Unlike heat cramps, heat tetany causes very little muscle-compartment pain. Treatment includes providing reassurance, moving the patient out of the heat, and addressing the hyperventilation.

HEAT CRAMPS

Heat cramps (exercise-associated muscle cramps) are intermittent, painful, and involuntary spasmodic contractions of skeletal muscles. They typically occur in an unacclimated individual who is at rest after vigorous exertion in a humid, hot environment. In contrast, cramps that occur in athletes during exercise last longer, are relieved by stretching and massage, and resolve spontaneously.

Of note, not all muscle cramps are related to exercise, and the differential diagnosis includes many other disorders. A variety of medications, myopathies, endocrine disorders, and sickle cell trait are other possible causes.

The typical patient with heat cramps is usually profusely diaphoretic and has been replacing fluid losses with copious water or other hypotonic fluids. Roofers, firefighters, military personnel, athletes, steel workers, and field workers are commonly affected. Other predisposing factors include insufficient sodium intake before intense activity in the heat and lack of heat acclimatization, resulting in sweat with a high salt concentration.

The precise pathogenesis of heat cramps appears to involve a relative deficiency of sodium, potassium, and fluid at the intracellular level. Coupled with copious hypotonic fluid ingestion, large amounts of sodium in the diaphoresis cause hyponatremia and hypochloremia, resulting in muscle cramps due to calcium-dependent muscle relaxation. Total-body depletion of potassium may be observed during the period of heat acclimatization. Rhabdomyolysis is very rare with routine exercise-associated muscle cramps.

Heat cramps that are not accompanied by significant dehydration can be treated with commercially available electrolyte solutions. Although the flavored electrolyte solutions are far more palatable, two 650-mg salt tablets dissolved in 1 quart of water produce a 0.1% saline solution. Individuals should avoid the ingestion of undissolved salt tablets, which are a gastric irritant and may induce vomiting.

HEAT EXHAUSTION

The physiologic hallmarks of heat exhaustion—in contrast to heatstroke—are the maintenance of thermoregulatory control and CNS function. The core temperature is usually elevated but is generally <40.5°C (<105°F). The two physiologic precipitants are water depletion and sodium depletion, which often occur in combination. Laborers, athletes, and elderly individuals exerting themselves in hot environments, without adequate fluid intake, tend to develop *water-depletion heat exhaustion*. Persons working in the heat frequently consume only two-thirds of their net water loss and are voluntarily dehydrated. In contrast, *salt-depletion heat exhaustion* occurs more slowly in unacclimated persons who have been consuming large quantities of hypotonic solutions.

Heat exhaustion is usually a diagnosis of exclusion because of the multitude of nonspecific symptoms. If any signs of heatstroke are present, rapid cooling and crystalloid resuscitation should be initiated immediately during stabilization and evaluation. Mild neurologic and gastrointestinal influenza-like symptoms are common. These symptoms may include headache, vertigo, ataxia, impaired judgment, malaise, dizziness, nausea, and muscle cramps. Orthostatic hypotension and sinus tachycardia develop frequently. More significant CNS impairment suggests heatstroke or other infectious, neurologic, or toxicologic diagnoses.

Hemoconcentration does not always develop, and rapid infusion of isotonic IV fluids should be guided by frequent electrolyte determinations and perfusion requirements. Most cases of heat exhaustion reflect mixed sodium and water depletion. Sodium-depletion heat exhaustion is characterized by hyponatremia and hypochloremia. Hepatic aminotransferases are mildly elevated in both types of heat exhaustion. Urinary sodium and chloride concentrations are usually low.

Some patients with heat exhaustion develop heatstroke after removal from the heat-stress environment. Aggressive cooling of nonresponders is indicated until their core temperature is 39°C (102.2°F). Except in mild cases, free water deficits should be replaced slowly over 24–48 h to avoid a decrease of serum osmolality by >2 mOsm/h.

The disposition of younger, previously healthy heat-exhaustion patients who have no major laboratory abnormalities may include hospital observation and discharge after IV rehydration. Older patients with comorbidities (including cardiovascular disease) or predisposing factors often require inpatient fluid and electrolyte replacement, monitoring, and reassessment.

HEATSTROKE

The clinical manifestations of heatstroke reflect a total loss of thermoregulatory function. Typical vital-sign abnormalities include tachypnea, various tachycardias, hypotension, and a widened pulse pressure. Although there is no single specific diagnostic test, the historical and physical triad of exposure to a heat stress, CNS dysfunction, and a core temperature >40.5°C (104.9°C) helps establish the preliminary diagnosis. Some patients with impending heatstroke will initially appear lucid. The definitive diagnosis should be reserved until the other potential causes of hyperthermia are excluded. Many of the usual laboratory abnormalities seen with heatstroke overlap with other conditions. If the patient's mental status does not improve with cooling, toxicologic screening may be indicated, and cranial computed tomography (CT) and spinal fluid analysis can be considered.

The premonitory clinical characteristics may be nonspecific and include weakness, dizziness, disorientation, ataxia, and gastrointestinal or psychiatric symptoms. These prodromal symptoms often resemble heat exhaustion. The sudden onset of heatstroke occurs when the maintenance of adequate perfusion requires peripheral vasoconstriction to stabilize the mean arterial blood pressure. As a result, the cutaneous radiation of heat ceases. At this juncture, the core temperature rises dramatically. Since many patients with heatstroke also meet the criteria for systemic inflammatory response syndrome (SIRS) and have a broad differential diagnosis, rapid cooling is essential during the extensive diagnostic evaluation. Heat-induced SIRS reflects the responses of both the innate and the adaptive immune systems (Table 478-1).

There are two forms of heatstroke with significantly different manifestations **(Table 478-2)**. Classic (epidemic) heatstroke (CHS) usually occurs during long periods of high ambient temperature and humidity, as during summer heat waves. Patients with CHS commonly have

TABLE 478-2 Typical Manifestations of Heatstroke

CLASSIC	EXERTIONAL
Older patient	Younger patient
Predisposing health factors/ medications	Healthy condition
Epidemiology (heat waves)	Sporadic cases
Sedentary	Exercising
Anhidrosis (possible)	Diaphoresis (common)
Central nervous system dysfunction	Myocardial/hepatic injury
Oliguria	Acute renal failure
Coagulopathy (mild)	Disseminated intravascular coagulation
Mild lactic acidosis	Marked lactic acidosis
Mild creatine kinase elevation	Rhabdomyolysis
Normoglycemia/calcemia	Hypoglycemia/calcemia
Normokalemia	Hyperkalemia
Normonatremia	Hyponatremia

chronic diseases that predispose to heat-related illness, and they may have limited access to oral fluids. Heat dissipation mechanisms are overwhelmed by both endogenous heat production and exogenous heat stress. Patients with CHS are often compliant with prescribed medications that can impair tolerance to a heat stress. In many of these dehydrated CHS patients, sweating has ceased and the skin is hot and dry.

The duration of core temperature elevation directly impacts morbidity and mortality. If cooling is delayed, severe hepatic dysfunction, renal failure, disseminated intravascular coagulation, and fulminant multisystem organ failure may occur. Hepatocytes are very heat sensitive. On presentation, the serum level of aspartate aminotransferase (AST) is routinely elevated. Eventually, levels of both AST and alanine aminotransferase (ALT) often increase to >100 times the normal values. Coagulation studies commonly demonstrate decreased platelets, fibrinogen, and prothrombin. Most patients with CHS require cautious crystalloid resuscitation, electrolyte monitoring, and—in certain refractory cases—consideration of central venous pressure (CVP) measurements and point of care ultrasonography **(Chap. 493)**. Hypernatremia is secondary to dehydration in CHS. Many patients exhibit significant stress leukocytosis, even in the absence of infection.

Patients with exertional heatstroke (EHS), in contrast to those with CHS, are often young and previously healthy, and their diagnosis is usually more obvious from the history. Athletes, laborers, and military recruits are common victims. Unlike those with CHS, many EHS patients present profusely diaphoretic despite significant dehydration. As a result of muscular exertion, rhabdomyolysis and acute renal failure are more common in EHS. Studies to detect rhabdomyolysis and its complications, including hypocalcemia and hyperphosphatemia, should be considered. Hyponatremia, hypoglycemia, and coagulopathies are frequent findings. Elevated creatine kinase and lactate dehydrogenase levels also suggest EHS. Oliguria is a common finding. Renal failure can result from direct thermal injury, untreated rhabdomyolysis, or volume depletion. Common urinalysis findings include microscopic hematuria, myoglobinuria, and granular or red cell casts.

With both CHS and EHS, heat-related increases in cardiac biomarker levels may be present and reversible. Heatstroke often causes thermal cardiomyopathy. As a result, the CVP may be elevated despite significant dehydration. In addition, the patient often presents with potentially deceptive noncardiogenic pulmonary edema and basilar rales despite being significantly hypovolemic. The electrocardiogram commonly displays a variety of tachyarrhythmias, nonspecific ST-T wave changes, and heat-related ischemia or infarction. Rapid cooling—not the *initial* administration of antiarrhythmic medications—is essential.

Above 42°C (107.6°F), heat can rapidly produce direct cellular injury. Thermosensitive enzymes become nonfunctional, and eventually, there is irreversible uncoupling of oxidative phosphorylation. The production of heat-shock proteins increases, and cytokines mediate a systemic inflammatory response. The vascular endothelium is also damaged, and this injury activates the coagulation cascade. Significant shunting away from the splanchnic circulation produces gastrointestinal ischemia. Endotoxins further impair normal thermoregulation. As a result, if cooling is delayed, severe hepatic dysfunction, permanent renal failure, disseminated intravascular coagulation, and fulminant multisystem organ failure may occur.

COOLING STRATEGIES

Endotracheal intubation and continuous core-temperature monitoring should be considered both at the scene and as cooling is initiated. Peripheral methods to measure temperature are *not* reliable. Hypoglycemia is a frequent finding. Since peripheral vasoconstriction delays heat dissipation, repeated administration of discrete boluses of isotonic crystalloid for hypotension is preferable to the administration of α-adrenergic agonists. Patients who require inotropic support after cooling and fluid resuscitation may require norepinephrine, followed by epinephrine and dobutamine.

Rapid cooling is essential in both CHS and EHS, and an immediate improvement in vital signs and mental status may prove valuable for diagnostic purposes. Cool water (15°C [60°F]) is sprayed on the exposed skin while fans direct continuous airflow over the moistened skin. Cold packs applied to the neck, axillae, and groin are useful cooling adjuncts. If cardiac electrodes will not adhere, they can be applied to the patient's back.

Immersion cooling in ice-cold water may be a preferable option with EHS if appropriate medical staff and materials for aggressive cooling are available on scene until emergency medical services arrival. The initial increase in temperature from peripheral vasoconstriction will rapidly be overcome by the large conductive thermal transfer into cold water. This technique presents significant monitoring and resuscitation challenges in many clinical settings. The safety of immersion cooling is best established for young, previously healthy patients with EHS (but not for those with CHS). To improve patient access and monitoring, an alternative is ice water therapy. Ice cold water is continuously poured over the supine patient lying on a porous stretcher. Muscular massage improves vasodilatation. Another option that allows patient access is ice cold immersion in a commercial "cooling body bag." To avoid hypothermic afterdrop (continued cooling), active cooling should be terminated at ~38°–39°C (100.4°F–102.2°F).

The rate of cooling with commercial cooling blankets is very slow. Other methods are less efficacious and rarely indicated, such as IV infusion of cold fluids and cold irrigation of the bladder or gastrointestinal tract. Cold thoracic and peritoneal lavage are invasive and rarely necessary. Endovascular cooling also provides effective cooling.

RESUSCITATION

Aspiration commonly occurs in heatstroke, and endotracheal intubation is usually necessary. Depolarizing agents should be avoided. The metabolic demands are high, and supplemental oxygenation is essential due to hypoxemia induced by thermal stress and pulmonary dysfunction. The oxyhemoglobin dissociation curve is shifted to the right. Pneumonitis, pulmonary infarction, hemorrhage, edema, and acute respiratory distress syndrome occur frequently in heatstroke patients. Seizures are common and can occur during therapeutic cooling. Cold-induced tonic-clonic muscular rigidity mimics seizure activity. Refractory seizures may require monitoring with an electroencephalogram.

The circulatory fluid requirements, particularly in CHS, may be deceptively modest. Aggressive cooling and modest volume repletion usually elevate the CVP to 12–14 mmHg. The reading, however, may be deceptive. Many patients present with a thermally induced hyperdynamic circulation accompanied by a high cardiac index, low peripheral vascular resistance, and an elevated CVP caused by right-sided heart failure. In contrast, most patients with EHS require far more zealous isotonic crystalloid resuscitation.

The hypotension that is initially common among patients with heatstroke results from both dehydration and high-output cardiac failure caused by peripheral vasodilation. Inotropes causing α-adrenergic stimulation (e.g., norepinephrine) can impede cooling by causing significant vasoconstriction. Vasoactive catecholamines such as dopamine or dobutamine may be necessary if the cardiac output remains depressed despite an elevated CVP, particularly in patients with a hyperdynamic circulation.

A wide variety of tachyarrhythmias are routinely observed on presentation and usually resolve spontaneously during cooling. The administration of atrial or ventricular antiarrhythmic medications is rarely indicated during cooling. Anticholinergic medications (including atropine) inhibit sweating and should be avoided. With a cardiac rhythm that sustains perfusion, electrical cardioversion of the hyperthermic myocardium should be deferred until the myocardium is cooled. Significant shivering, discomfort, or extreme agitation is preferably mitigated with short-acting benzodiazepines or propofol. On the other hand, chlorpromazine may lower the seizure threshold, has anticholinergic properties, and can exacerbate the hypotension or cause neuroleptic malignant syndrome.

Coagulopathies more commonly occur after the first day of illness. After cooling, the patient should be monitored for disseminated intravascular coagulation, and replacement therapy with fresh-frozen plasma and platelets should be considered. Consider a dose of empiric antibiotics after culturing during cooling if the etiology of the hyperthermia remains unclear.

There is no therapeutic role for antipyretics in the control of environmentally induced hyperthermia; these drugs block the actions of pyrogens at hypothalamic receptor sites. Salicylates can further uncouple oxidative phosphorylation in heatstroke and exacerbate coagulopathies. Acetaminophen may further stress hepatic function. Dantrolene is ineffective when the temperature elevation is not caused by malignant hyperthermia. Although aminocaproic acid impedes fibrinolysis, it may cause rhabdomyolysis and is not recommended in heatstroke.

DISPOSITION

Most patients with minor heat-emergency syndromes (including heat edema, heat syncope, and heat cramps) require only stabilization and treatment with outpatient follow-up. Although there are no decision rules to guide disposition choices in heat exhaustion, many of these patients have multiple predisposing factors and comorbidities that will require prolonged observation or hospital admission.

Essentially all patients with actual heatstroke require admission to a monitored setting, and most require intensive care. There are reports of very high survival rates of patients following prehospital immersion cooling without intensive care. Most or all of these patients appear to have had heat exhaustion. Many actual heatstroke patients also require prolonged tracheal intubation, invasive hemodynamic monitoring, and support for various degrees of multiorgan dysfunction syndrome. The prognosis worsens if the initial core temperature exceeds 42°C (107.6°F) or if there was a prolonged period during which the core temperature exceeded this level. Other features of a negative prognosis include acute renal failure, massively elevated liver enzymes, and significant hyperkalemia. As expected, the number of dysfunctional organ systems also correlates directly with mortality risk.

FURTHER READING

Bouchama A et al: Classic and exertional heatstroke. Nat Rev Dis Primers 8:1, 2022.

Epstein Y, Yanovich R: Heatstroke. N Engl J Med 380:2449, 2019.

Filep E et al: Exertional heat stroke, modality cooling rate, and survival outcomes: a systematic review. Medicina 56:589, 2020.

Kaewput W et al: Inpatient burden and mortality of heatstroke in the United States. Int J Clin Pract 75:e13837, 2020.

Lipman GS et al: Wilderness Medical Society practice guidelines for the prevention and treatment of heat-related illness: 2019 Update. Wilderness Environ Med 30:S33, 2019.

Platt M et al: Heat illness, in *Rosen's Emergency Medicine: Concepts and Clinical Practice*, 10th ed. Walls RM et al (eds). Philadelphia, Elsevier, 2023, pp. 1771-1780.

Rublee C et al: Evidence-based heatstroke management in the emergency department. West J Emerg Med 22:186, 2021.

Sorensen C, Hess J: Treatment and prevention of heat-related illness. N Engl J Med 387:1404, 2022.

479

Principles of Human Genetics

J. Larry Jameson, Peter Kopp

IMPACT OF GENETICS AND GENOMICS ON MEDICAL PRACTICE

Over the past four decades, novel insights into human genetics and genomics have fundamentally impacted the practice of medicine, ushering in a new area with a deeper understanding of the genetic basis of numerous health conditions, novel diagnostic technologies, disease prevention and management, personalized medicine, and targeted therapies. Human genetics refers to the study of individual genes, their role and function in disease, and their mode of inheritance. Genomics refers to an organism's entire genetic information, the *genome*, and the function and interaction of DNA within the genome, as well as with environmental or nongenetic factors, such as a person's lifestyle. With the characterization of the human genome, genomics not only complements traditional genetics in our efforts to elucidate the etiology and pathogenesis of disease, but it also plays a prominent and expanding role in diagnostics, prevention, and therapy **(Chap. 480)**. These transformative developments, originally emerging from the Human Genome Project, have been variably designated *genomic medicine*, *personalized medicine*, or *precision medicine*. Precision medicine aims at customizing medical decisions to an individual patient. For example, a patient's genetic characteristics (genotype) can be used to optimize drug therapy and predict efficacy, adverse events, and drug dosing of selected medications (*pharmacogenomics*) **(Chap. 72)**. The characterization of the mutational profile of a malignancy allows the identification of driver mutations or overexpressed signaling molecules, which then facilitates the selection of targeted therapies. Genome-wide polygenic risk scores (PRS) for common complex diseases are beginning to emerge and may impact disease prevention in the future.

Genetics has traditionally been viewed through the window of relatively rare single-gene diseases. These disorders account for ~10% of pediatric admissions and childhood mortality. Historically, genetics has focused predominantly on chromosomal and metabolic disorders, reflecting the long-standing availability of techniques to diagnose these conditions. For example, conditions such as trisomy 21 (Down's syndrome) or monosomy X (Turner's syndrome) can be diagnosed using cytogenetics. Likewise, many metabolic disorders (e.g., phenylketonuria, familial hypercholesterolemia) are diagnosed using biochemical analyses. The advances in DNA and RNA diagnostics have extended the field of genetics to include virtually all medical specialties and have led to the elucidation of the pathogenesis of most monogenic disorders. In addition, virtually every medical condition has a genetic component. As is often evident from a patient's family history, many common disorders such as hypertension, heart disease, asthma, diabetes mellitus, and mental illnesses are significantly influenced by the genetic background. These polygenic or multifactorial (complex) disorders involve the contributions of many different genes, as well as environmental factors that can modify disease risk. Genome-wide association studies (GWAS) have elucidated numerous disease-associated loci and are providing novel insights into the allelic architecture of complex traits. These studies have been facilitated by the availability of comprehensive catalogues of human single nucleotide polymorphism (SNP) haplotypes (HapMap, International Genome Sample Resource). Next-generation DNA sequencing (NGS) technologies have evolved rapidly, and the cost of sequencing whole exomes (the exons within the genome; whole exome sequencing [WES]) or genomes (whole genome sequencing [WGS]) has plummeted. Comprehensive unbiased sequence analyses are now routinely used to characterize individuals with complex undiagnosed conditions or to determine the mutational profile of advanced malignancies in order to select optimal and targeted therapies. The assembly of diploid genomes, i.e., the characterization of the complete genetic information from both sets of chromosomes in an individual's genome, will further enhance the complete resolution of genetic variation and should provide further insights into heritability and disease mechanisms.

Cancer has a genetic basis because it results from acquired somatic mutations in genes controlling growth, apoptosis, and cellular differentiation **(Chap. 76)**. In addition, the development of many cancers is associated with a hereditary predisposition. Characterization of the genome (and epigenome) in various malignancies has led to fundamental new insights into cancer biology and reveals that the genomic profile of mutations is in many cases more important in determining the appropriate therapy than the organ in which the tumor originates. The Cancer Genome Atlas (TCGA) initiative of the National Cancer Institute and the National Human Genome Research Institute has already characterized the genomic landscape across >30 malignancies. TCGA consists of comprehensive analyses of genomic and proteomic alterations and provided fundamental new insights into the molecular pathogenesis of cancer. These data, together with comprehensive catalogues of somatic mutations identified in human cancer, have direct clinical ramifications that impact cancer taxonomy, as well as the development and choice of targeted therapies.

Genetic and genomic approaches have proven invaluable for the detection of infectious pathogens and are used clinically to identify agents that are difficult to culture such as mycobacteria, viruses, and parasites, or to track infectious agents locally or globally. In many cases, molecular genetics has improved the feasibility and accuracy of diagnostic testing and has opened new avenues for therapy, including gene and cellular therapies **(Chap. 483)**. Molecular genetics has also provided the opportunity to characterize the *microbiome*, a field that characterizes the population dynamics of bacteria, viruses, and parasites that coexist with humans and other animals **(Chap. 483)**. The microbiome has significant effects on normal physiology as well as various disease states, and the field is now focusing on defining the mechanisms underlying these interactions.

Molecular biology has significantly changed the treatment of human disease. Peptide hormones, growth factors, cytokines, and vaccines can be produced in large amounts using recombinant DNA and RNA technology (e.g., mRNA vaccines against SARS-CoV-2; small interfering RNA [siRNA] to treat hypercholesterolemia). Targeted modifications of recombinant peptides provide improved therapeutic tools, as illustrated by genetically modified insulin analogues with more favorable kinetics or glucagon-like peptide 1 (GLP-1) agonists for treatment of type 2 diabetes and for weight management.

The rate at which new genetic and genomic information is being generated presents many challenges for health care providers and systems. Although many functional aspects of the genome remain unknown, there are many clinical situations where genetic and genomic information optimize patient care. Much genetic information resides in databases that provide easy access to the expanding information about the human genome, genetic disease, and genetic testing **(Table 479-1)**. For example, several thousand monogenic disorders are summarized in a large, continuously evolving compendium, the *Online Mendelian Inheritance in Man* (OMIM) catalogue (Table 479-1). The constant refinement of bioinformatics and big data analytics, together with the widespread adoption of electronic health records (EHRs), are simplifying the access, analysis, and integration of this daunting amount of new information. Importantly, genomic data can be integrated readily into EHRs and thus impact clinical practice.

■ THE HUMAN GENOME

Structure of the Human Genome The Human Genome Project was initiated in the mid-1980s as an ambitious effort to characterize

TABLE 479-1 Selected Databases Relevant for Genomics and Genetic Disorders

SITE	URL	COMMENT
National Center for Biotechnology Information (NCBI)	*http://www.ncbi.nlm.nih.gov/*	Broad access to biomedical and genomic information, literature (PubMed), sequence databases, software for analyses of nucleotides and proteins Extensive links to other databases, genome resources, and tutorials
National Human Genome Research Institute	*http://www.genome.gov/*	An institute of the National Institutes of Health focused on genomic and genetic research; links providing information about the human genome sequence, genomes of other organisms, genomic research, and legislation
Catalog of Published Genome-Wide Association Studies	*https://www.ebi.ac.uk/gwas/*	Published high-resolution genome-wide association studies (GWAS)
Ensembl Genome browser	*http://www.ensembl.org*	Maps and sequence information of eukaryotic genomes
Online Mendelian Inheritance in Man	*http://www.ncbi.nlm.nih.gov/omim*	Online compendium of Mendelian disorders and human genes causing genetic disorders
American College of Medical Genetics and Genomics	*http://www.acmg.net/*	Extensive links to other databases relevant for the diagnosis, treatment, and prevention of genetic disease
American Society of Human Genetics	*http://www.ashg.org*	Information about advances in genetic research, professional and public education, and social and scientific policies
The Cancer Genome Atlas	*https://cancergenome.nih.gov/*	Comprehensive, multidimensional characterization of the genomic and proteomic landscape of malignancies with high public health impact
COSMIC Catalogue of Somatic Mutations in Cancer	*https://cancer.sanger.ac.uk/cosmic*	Comprehensive catalogue of somatic mutations in human cancer
Genetic Testing Registry	*https://www.ncbi.nlm.nih.gov/gtr/*	International directory of genetic testing laboratories and prenatal diagnosis clinics; reviews and educational materials
Genomes Online Database (GOLD)	*http://www.genomesonline.org/*	Information on published and unpublished genomes
HUGO Gene Nomenclature	*http://www.genenames.org/*	Gene names and symbols
GENECODE	*https://www.gencodegenes.org/*	High-quality reference gene annotation and experimental validation for human and mouse genomes
MITOMAP, a human mitochondrial genome database	*http://www.mitomap.org/*	A compendium of polymorphisms and mutations of the human mitochondrial DNA
The International Genome Sample Resource (IGSR)	*http://www.internationalgenome.org*	Public catalogue of human variation and genotype data from numerous ethnic groups
Human Genome Variation Society	*https://www.hgvs.org/*	Collection and documentation of genomic variations including population distribution and phenotypic associations
ENCODE	*http://www.genome.gov/10005107*	*Enc*yclopedia *o*f *D*NA *E*lements; catalogue of all functional elements in the human genome
Dolan DNA Learning Center, Cold Spring Harbor Laboratories	*http://www.dnalc.org/*	Educational material about selected genetic disorders, DNA, eugenics, and genetic origin
The Online Metabolic and Molecular Bases of Inherited Disease (OMMBID)	*http://ommbid.mhmedical.com*	Online version of the comprehensive text on the metabolic and molecular bases of inherited disease
Online Mendelian Inheritance in Animals (OMIA)	*https://www.omia.org/home/*	Online compendium of Mendelian disorders in animals
The Jackson Laboratory	*http://www.jax.org/*	Information about murine models and the mouse genome
Mouse genome informatics	*http://www.informatics.jax.org*	Mouse genome informatics, potential mouse models of human disease, information on phenotypic similarity between mouse models and human patients

Note: Databases are evolving constantly. Pertinent information may be found by using links listed in the few selected databases.

the entire human genome and culminated in the completion of the DNA sequence for the last of the human chromosomes in 2006. The scope of a WGS analysis can be illustrated by the following analogy. Human DNA consists of ~3 billion base pairs (bp) of DNA per haploid genome, which is nearly 1000-fold greater than that of the *Escherichia coli* genome. If the human DNA sequence were printed out, it would correspond to about 120 volumes of *Harrison's Principles of Internal Medicine*.

In addition to the human genome, the genomes of thousands of organisms have been sequenced completely or partially (Genomes Online Database [GOLD]; Table 479-1). They include, among others, eukaryotes such as the mouse (*Mus musculus*), *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and *Drosophila melanogaster*; bacteria (e.g., *E. coli*); and archaea, viruses, organelles (mitochondria, chloroplasts), and plants (e.g., *Arabidopsis thaliana*). Genomic information of infectious agents has significant impact for the characterization of infectious outbreaks and epidemics. Other ramifications arising from the availability of genomic data include, among others, (1) the comparison of entire genomes (*comparative genomics*); (2) the study of large-scale expression of RNAs (*functional genomics*), proteins (*proteomics*), or protein families (e.g., the *kinome*, the complete set of protein kinases) to detect differences between various tissues in health and disease; (3) the characterization of the variation among individuals by establishing catalogues of sequence variations and SNPs; and (4) the identification of genes that play critical roles in the development of polygenic and multifactorial disorders.

CHROMOSOMES The human genome is divided into 23 different chromosomes, including 22 autosomes (numbered 1–22) and the X and Y sex chromosomes **(Fig. 479-1)**. Adult cells are *diploid*, meaning they contain two homologous sets of 22 autosomes and a pair of sex chromosomes. Females have two X chromosomes (XX), whereas males have one X and one Y chromosome (XY). As a consequence of meiosis, germ cells (sperm or oocytes) are haploid and contain one set of 22 autosomes and one of the sex chromosomes. At the time of fertilization, the diploid genome is reconstituted by pairing of the homologous chromosomes from the mother and father. With each cell division (mitosis), chromosomes are replicated, paired, segregated, and divided into two daughter cells.

STRUCTURE OF DNA DNA is a double-stranded helix composed of four different bases: adenine (A), thymidine (T), guanine (G), and cytosine (C). Adenine is paired to thymidine, and guanine is paired

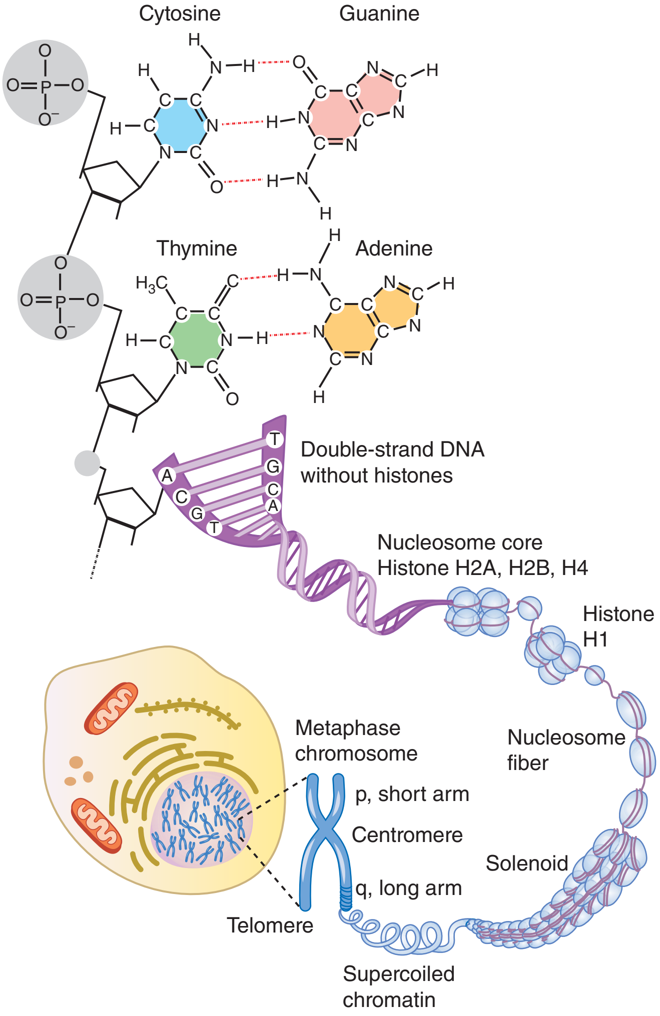

FIGURE 479-1 Structure of chromatin and chromosomes. Chromatin is composed of double-strand DNA that is wrapped around histone and nonhistone proteins forming nucleosomes. The nucleosomes are further organized into solenoid structures. Chromosomes assume their characteristic structure, with short (p) and long (q) arms at the metaphase stage of the cell cycle.

to cytosine, by hydrogen bond interactions that span the double helix (Fig. 479-1). DNA has several remarkable features that make it ideal for the transmission of genetic information. It is relatively stable, and the double-stranded nature of DNA and its feature of strict base-pair complementarity permit faithful replication during cell division. Complementarity also allows the transmission of genetic information from DNA → RNA → protein **(Fig. 479-2)**. mRNA is encoded by the so-called sense or coding strand of the DNA double helix and is translated into proteins by ribosomes.

The presence of four different bases provides surprising genetic diversity. In the protein-coding regions of genes, the DNA bases are arranged into codons, a triplet of bases that specifies a particular amino acid. It is possible to arrange the four bases into 64 different triplet codons (4^3). Each codon specifies 1 of the 20 different amino acids, or a regulatory signal such as initiation and stop of translation. Because there are more codons than amino acids, the genetic code is degenerate; that is, most amino acids can be specified by several different codons. By arranging the codons in different combinations and in various lengths, it is possible to generate the tremendous diversity of primary protein structure.

DNA length is normally measured in units of 1000 bp (kilobases, kb) or 1,000,000 bp (megabases, Mb). In the human genome, only ~1% of DNA accounts for protein-coding sequences. The noncoding DNA has multiple functional and structural roles including (1) sequences that form introns; (2) regulatory elements (promoters, enhancers, silencers, insulators); (3) sequences that generate RNAs that do not code for proteins; (4) centromeres and telomeres; (5) regions defining chromatin structure and histone modifications; (6) various forms of repetitive sequences of variable length; and (7) pseudogenes and regions without currently discernible functional or structural roles (Fig. 479-1).

GENES A *gene* is a functional unit that is regulated by transcription (see below) and encodes an RNA product, which is most commonly, but not always, translated into a protein that exerts activity within or outside the cell **(Fig. 479-3)**. Historically, genes were identified because they conferred specific traits that are transmitted from one generation to the next. Now, they are frequently characterized based on expression in various tissues (*transcriptome*). The size of genes is quite broad; some genes are only a few hundred base pairs, whereas others are extraordinarily large (2.3 Mb). The number of genes greatly underestimates the complexity of genetic expression, because single genes can generate multiple spliced messenger RNA (mRNA) products (*isoforms*), which are translated into proteins that are subject to complex posttranslational modification such as phosphorylation. *Exons* refer to the portion of genes that are eventually spliced together to form mRNA. *Introns* refer to the spacing regions between the exons that are spliced out of precursor RNAs during RNA processing. The gene locus also includes regions that are necessary to control its expression (Fig. 479-2). Current estimates predict roughly 20,000 protein-coding genes in the human genome with an average of about four different coding transcripts per gene. Remarkably, the exome only constitutes 1.14% of the genome. Of note, the number of transcripts is close to 200,000 and includes thousands of noncoding transcripts (RNAs of various length such as microRNAs [miRNA] and long noncoding RNAs [lncRNA]). These noncoding RNAs are involved in numerous cellular processes such as transcriptional and posttranscriptional regulation of gene expression, chromatin remodeling, and protein trafficking, among others. Not surprisingly, aberrant expression and/or mutations in these RNAs play a pathogenic role in numerous diseases.

SINGLE-NUCLEOTIDE POLYMORPHISMS On average, a typical genome differs from the reference human genome at 4 to 5 million sites. Some of these variants have no impact on health, whereas others may increase or lower the risk for developing a specific disease. Remarkably, however, the primary DNA sequence of humans has ~99.9% similarity compared to that of any other human. An SNP is a variation of a single base pair in the DNA. Across human populations from distinct ethnic backgrounds, there are more than 1 billion validated SNPs (Fig. 479-3). SNPs are the most common type of sequence variation and account for >90% of all sequence variation. They occur on average every 100–300 bases and are the major source of genetic heterogeneity. SNPs that are in proximity are inherited together (e.g., they are linked) and are referred to as *haplotypes* **(Fig. 479-4)**. Haplotype maps describe the nature and location of these SNP haplotypes and how they are distributed among individuals within and among populations, information that has been facilitating GWAS designed to elucidate the complex interactions among multiple genes and lifestyle factors in multifactorial disorders (see below). Moreover, haplotype analyses are useful to assess variations in responses to medications (*pharmacogenomics*) and environmental factors, as well as the prediction of disease predisposition.

COPY NUMBER VARIATIONS Copy number variations (CNVs) are relatively large genomic regions (1 kb to several Mb) that have been duplicated or deleted on certain chromosomes and hence alter the diploid status of the DNA **(Fig. 479-5)**. It has been estimated that 5–10% of the genome can display CNVs. When comparing the genomes of two individuals, ~0.4–0.8% of their genomes differ in terms of CNVs scattered throughout the genome. Some CNVs can increase or decrease gene dosage, potentially leading to detrimental effects if essential genes are impacted. Of note, de novo CNVs have been observed between monozygotic twins, who otherwise have identical genomes.

Replication of DNA and Mitosis Genetic information in DNA is transmitted to daughter cells under two different circumstances: (1) somatic cells divide by *mitosis*, allowing the diploid (2*n*) genome to replicate itself completely in conjunction with cell division; and (2) germ cells (sperm and ova) undergo *meiosis*, a process that enables the reduction of the diploid (2*n*) set of chromosomes to the haploid state (1*n*).

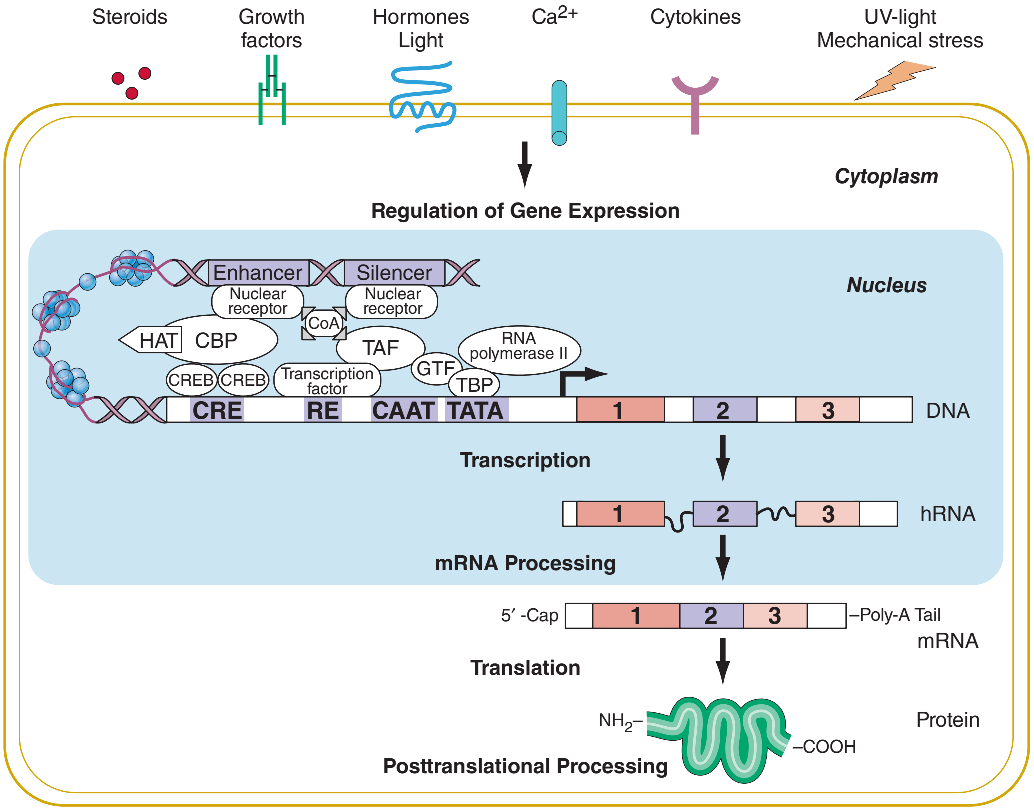

FIGURE 479-2 Flow of genetic information. Multiple extracellular signals activate intracellular signal cascades that result in altered regulation of gene expression through the interaction of transcription factors with regulatory regions of genes. RNA polymerase transcribes DNA into RNA that is processed to mRNA by excision of intronic sequences. The mRNA is translated into a polypeptide chain to form the mature protein after undergoing posttranslational processing. CBP, CREB-binding protein; CoA, co-activator; COOH, carboxyterminus; CRE, cyclic AMP responsive element; CREB, cyclic AMP response element–binding protein; GTF, general transcription factors; HAT, histone acetyl transferase; NH2, aminoterminus; RE, response element; TAF, TBP-associated factors; TATA, TATA box; TBP, TATA-binding protein.

Prior to mitosis, cells exit the resting, or G_0 state, and enter the cell cycle. After traversing a critical checkpoint in G_1, cells undergo DNA synthesis (S phase), during which the DNA in each chromosome is replicated, yielding two pairs of sister chromatids ($2n \rightarrow 4n$). The process of DNA synthesis requires stringent fidelity in order to avoid transmitting errors to subsequent generations of cells. Genetic abnormalities of DNA mismatch/repair include xeroderma pigmentosum, Bloom's syndrome, ataxia telangiectasia, and hereditary nonpolyposis colon cancer (HNPCC), among others. Many of these disorders strongly predispose to neoplasia because of the rapid acquisition of additional mutations **(Chap. 76)**. After completion of DNA synthesis, cells enter G_2 and progress through a second checkpoint before entering mitosis. At this stage, the chromosomes condense and are aligned along the equatorial plate at metaphase. The two identical sister chromatids, held together at the centromere, divide and migrate to opposite poles of the cell. After formation of a nuclear membrane around the two separated sets of chromatids, the cell divides and two daughter cells are formed, thus restoring the diploid ($2n$) state.

Assortment and Segregation of Genes During Meiosis Meiosis occurs only in germ cells of the gonads. It shares certain features with mitosis but involves two distinct steps of cell division that reduce the chromosome number to the haploid state. In addition, there is active recombination that generates genetic diversity. During the first cell division, two sister chromatids ($2n \rightarrow 4n$) are formed for each chromosome pair and there is an exchange of DNA between homologous paternal and maternal chromosomes. This process involves the formation of *chiasmata*, structures that correspond to the DNA segments that cross over between the maternal and paternal homologues **(Fig. 479-6)**. Usually there is at least one crossover on each chromosomal arm; recombination occurs more frequently in female meiosis than in male meiosis. Subsequently, the chromosomes segregate randomly. Because there are 23 chromosomes, there exist 2^{23} (>8 million) possible combinations of chromosomes. Together with the genetic exchanges that occur during recombination, chromosomal segregation generates tremendous diversity, and each gamete is genetically unique. The process of recombination and the independent segregation of chromosomes provide the foundation for performing linkage analyses, whereby one attempts to correlate the inheritance of certain chromosomal regions (or linked genes) with the presence of a disease or genetic trait (see below).

After the first meiotic division, which results in two daughter cells ($2n$), the two chromatids of each chromosome separate during a second meiotic division to yield four gametes with a haploid state ($1n$). When the egg is fertilized by sperm, the two haploid sets are combined, thereby restoring the diploid state ($2n$) in the zygote.

REGULATION OF GENE EXPRESSION

Regulation by Transcription Factors The expression of genes is regulated by DNA-binding proteins that activate or repress transcription. The number of DNA sequences and transcription factors that regulate transcription is much greater than originally anticipated. Most genes contain at least 15–20 discrete regulatory elements within 300 bp of the transcription start site. This densely packed promoter region often contains binding sites for ubiquitous transcription factors. However, factors involved in cell-specific expression may also bind to these sequences. Key regulatory elements may also reside at a large distance from the proximal promoter. The globin and the immunoglobulin genes, for example, contain *locus control regions* that are several kilobases away from the structural sequences of the gene. Specific groups of transcription factors that bind to these promoter and enhancer sequences provide a combinatorial code for regulating transcription. In this manner, relatively ubiquitous factors interact with more restricted factors to allow each gene to be expressed and regulated in a unique manner that is dependent on developmental state, cell type, and numerous extracellular stimuli. Regulatory factors also bind within the gene itself, particularly in the intronic regions. The

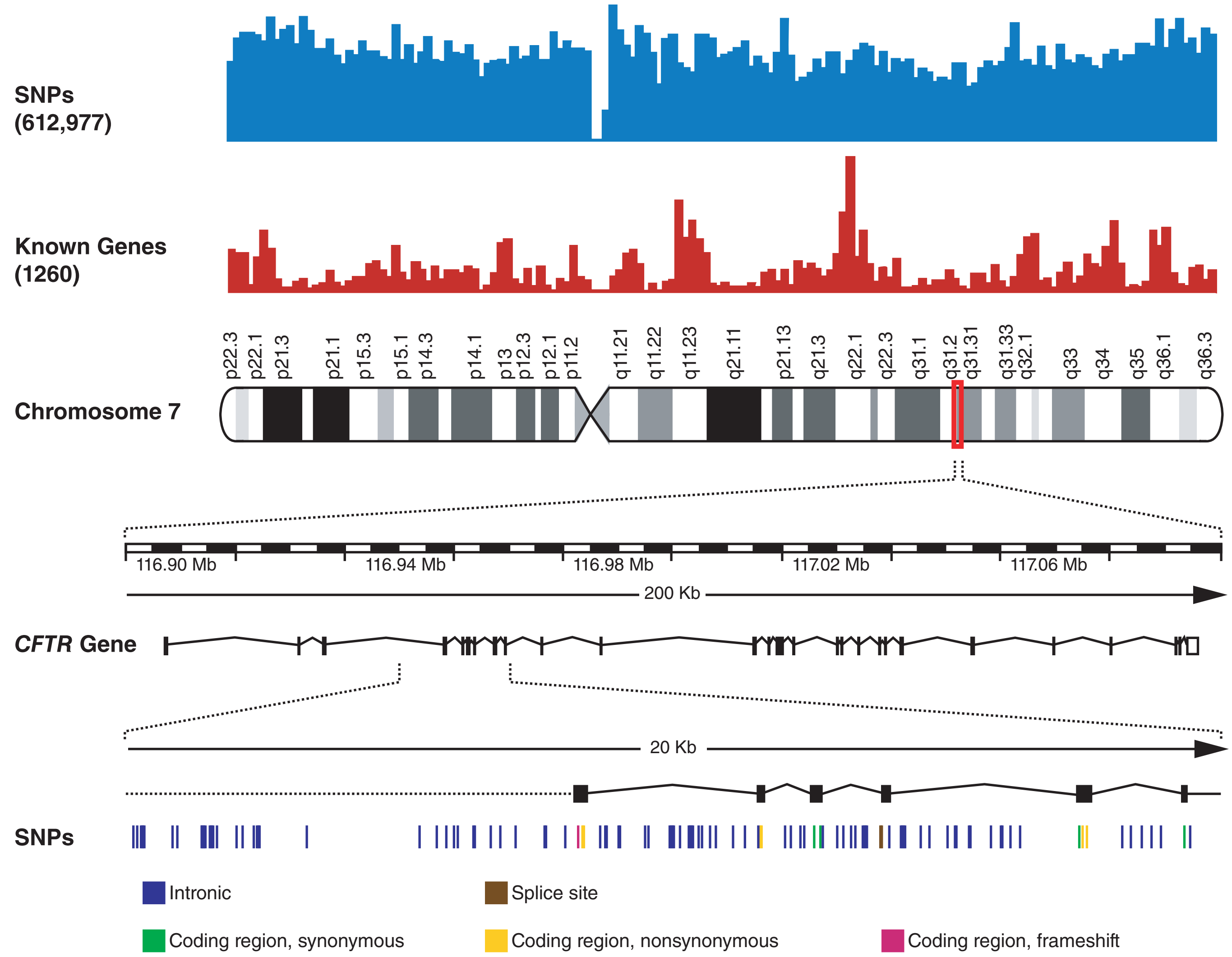
SNPs
(612,977)
Known Genes
(1260)
Chromosome 7
p22.3
p22.1
p21.3
p21.1
p15.3
p15.1
p14.3
p14.1
p13
p12.3
p12.1
p11.2
q11.21
q11.22
q11.23
q21.11
q21.13
q21.3
q22.1
q22.3
q31.1
q31.2
q31.31
q31.33
q32.1
q33
q34
q35
q36.1
q36.3
116.90 Mb
116.94 Mb
116.98 Mb
117.02 Mb
117.06 Mb
200 Kb
CFTR Gene
20 Kb
SNPs
Intronic
Splice site
Coding region, synonymous
Coding region, nonsynonymous
Coding region, frameshift

FIGURE 479-3 Chromosome 7 is shown with the density of single nucleotide polymorphisms (SNPs) and genes above. A 200-kb region in 7q31.2 containing the *CFTR* gene is shown below. The *CFTR* gene contains 27 exons. Close to 2000 mutations in this gene have been found in patients with cystic fibrosis. A 20-kb region encompassing exons 4–9 is shown further amplified to illustrate the SNPs in this region.

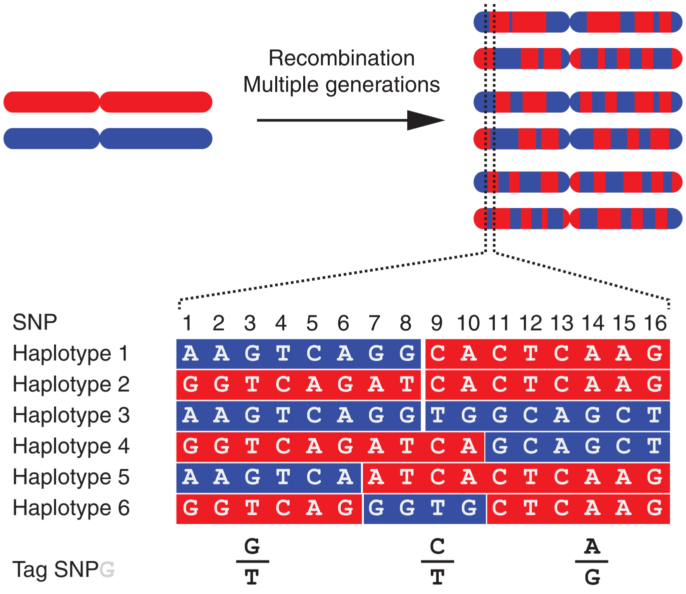
Recombination
Multiple generations

SNP	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Haplotype 1	A	A	G	T	C	A	G	G	C	A	C	T	C	A	A	G
Haplotype 2	G	G	T	C	A	G	A	T	C	A	C	T	C	A	A	G
Haplotype 3	A	A	G	T	C	A	G	G	T	G	G	C	A	G	C	T
Haplotype 4	G	G	T	C	A	G	A	T	C	A	G	C	A	G	C	T
Haplotype 5	A	A	G	T	C	A	A	T	C	A	C	T	C	A	A	G
Haplotype 6	G	G	T	C	A	G	G	G	T	G	C	T	C	A	A	G
Tag SNP			G/T						C/T					A/G		

FIGURE 479-4 The origin of haplotypes is due to repeated recombination events occurring in multiple generations. Over time, this leads to distinct haplotypes. These haplotype blocks can often be characterized by genotyping selected Tag single nucleotide polymorphisms (SNPs), an approach that facilitates performing genome-wide association studies (GWAS).

transcription factors that bind to DNA represent only the first level of regulatory control. Other proteins—*co-activators* and *co-repressors*—interact with the DNA-binding transcription factors to generate large regulatory complexes. These complexes are subject to control by numerous cell-signaling pathways and enzymes, leading to phosphorylation, acetylation, sumoylation, and ubiquitination. Ultimately, the recruited transcription factors interact with, and stabilize, components of the basal transcription complex that assembles at the site of the TATA box and initiator region. This basal transcription factor complex consists of >30 different proteins. Gene transcription occurs when RNA polymerase begins to synthesize RNA from the DNA template. A large number of identified genetic diseases involve transcription factors **(Table 479-2)**.

The field of *functional genomics* is based on the concept that understanding alterations of gene expression under various physiologic and pathologic conditions provides insight into the underlying functional role of the gene. The ENCODE (Encyclopedia of DNA Elements) project aims at identifying and annotating all functional sequences in the human genome. By revealing specific gene expression profiles, this knowledge can be of diagnostic and therapeutic relevance. The large-scale study of expression profiles is referred to as *transcriptomics* because the complement of mRNAs transcribed by the cellular genome is called the *transcriptome*.

Most studies of gene expression have focused on the regulatory DNA elements of genes that control transcription. However, it must be

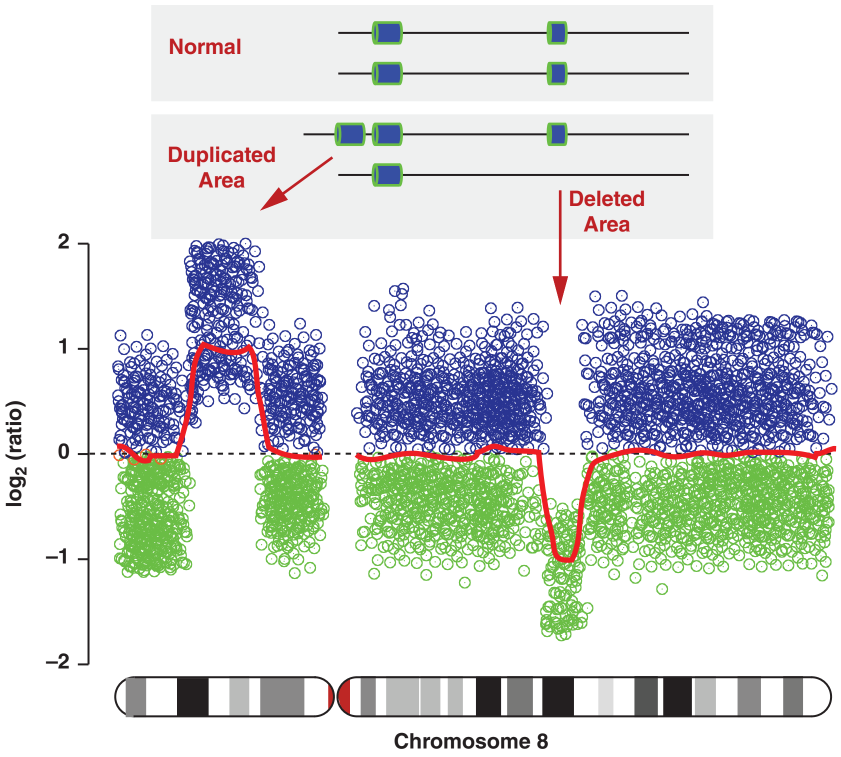

FIGURE 479-5 Copy number variations (CNV) encompass relatively large regions of the genome that have been duplicated or deleted. Chromosome 8 is shown with a CNV detected by genomic hybridization. An increase in the signal strength indicates a duplication, whereas a decrease reflects a deletion of the covered chromosomal regions.

emphasized that gene expression requires a series of steps, including mRNA processing, protein translation, and posttranslational modifications, all of which are actively regulated (Fig. 479-2).

Epigenetic Regulation of Gene Expression (see Chap. 497) *Epigenetics* describes mechanisms and phenotypic changes that are not a result of variation in the primary DNA nucleotide sequence but are caused by secondary modifications of DNA or histones. These modifications include heritable changes such as X-inactivation and imprinting, but they can also result from dynamic posttranslational protein modifications in response to environmental influences such as diet, age, or drugs. The epigenetic modifications result in altered expression of individual genes or chromosomal loci encompassing multiple genes. The term *epigenome* describes the constellation of covalent modifications of DNA and histones that impact chromatin structure, as well as noncoding transcripts that modulate the transcriptional activity of DNA. Although the primary DNA sequence is usually identical in all cells of an organism, sex- and tissue-specific changes in the epigenome contribute to determining the transcriptional signature of a cell (transcriptome) and hence the protein expression profile (proteome).

Mechanistically, DNA and histone modifications can result in the activation or silencing of gene expression (Fig. 479-7). DNA methylation involves the addition of a methyl group to cytosine residues. This is usually restricted to cytosines of CpG dinucleotides, which are abundant throughout the genome. Methylation of these dinucleotides is thought to represent a defense mechanism that minimizes the expression of sequences that have been incorporated into the genome such as retroviral sequences. CpG dinucleotides also exist in so-called *CpG islands*, stretches of DNA characterized by a high CG content, which are found in the majority of human gene promoters. CpG islands in promoter regions are typically unmethylated, and the lack of methylation facilitates transcription.

Histone methylation involves the addition of a methyl group to lysine residues in histone proteins (Fig. 479-7). Depending on the specific lysine residue being methylated, this alters chromatin configuration, making it either more open or tightly packed. Acetylation of histone proteins is another well-characterized mechanism that results in an open chromatin configuration, which favors active transcription. Acetylation is generally more dynamic than methylation, and many transcriptional activation complexes have histone acetylase activity, whereas repressor complexes often contain deacetylases and remove acetyl groups from histones. Other histone modifications include, among others, phosphorylation and sumoylation.

Furthermore, noncoding RNAs and RNA regulatory networks that bind to DNA have a significant impact on transcriptional activity.

Physiologically, epigenetic mechanisms play an important role in several instances. For example, X-inactivation refers to the relative silencing of one of the two X chromosome copies present in females. The inactivation process is a form of dosage compensation such that females (XX) do not generally express twice as many X-chromosomal gene products as males (XY). In a given cell, the choice of which chromosome is inactivated occurs randomly in humans. But once the maternal or paternal X chromosome is inactivated, it will remain inactive, and this information is transmitted with each cell division. The *X-inactive specific transcript* (*Xist*) gene encodes a long non-coding RNA (lncRNA) that mediates gene silencing on one of the X chromosomes. The inactive X chromosome is highly methylated and has low levels of histone acetylation. While the majority of X-chromosomal genes are silenced by X-inactivation, ~15% escape inactivation and are expressed.

Epigenetic gene inactivation also occurs on selected chromosomal regions of autosomes, a phenomenon referred to as *genomic imprinting*. Through this mechanism, a small subset of genes is only expressed in a monoallelic fashion. Imprinting is

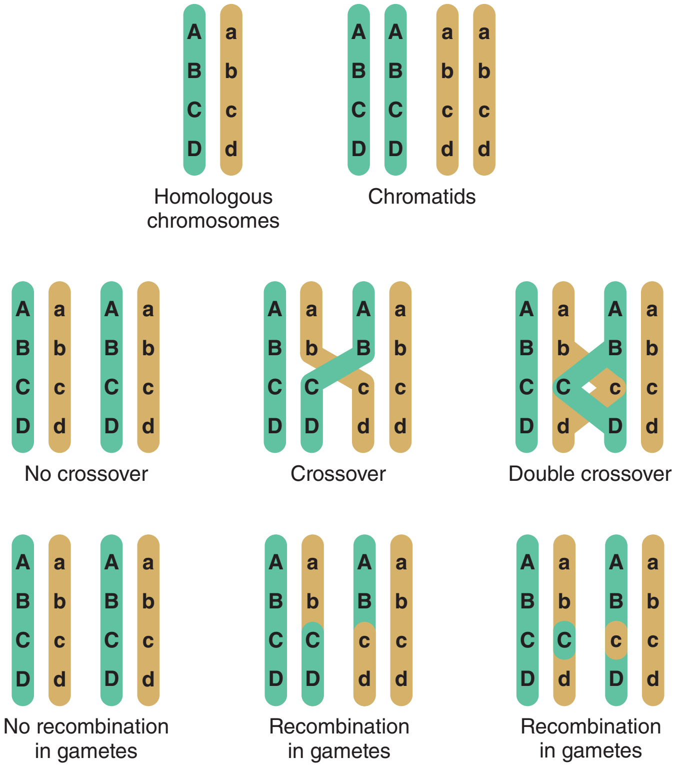

FIGURE 479-6 Crossing-over and genetic recombination. During chiasma formation, either of the two sister chromatids on one chromosome pairs with one of the chromatids of the homologous chromosome. Genetic recombination occurs through crossing-over and results in recombinant and nonrecombinant chromosome segments in the gametes. Together with the random segregation of the maternal and paternal chromosomes, recombination contributes to genetic diversity and forms the basis of the concept of linkage.

TABLE 479-2 Selected Examples of Diseases Caused by Mutations and Rearrangements in Transcription Factors

TRANSCRIPTION FACTOR CLASS	EXAMPLE	ASSOCIATED DISORDER
Nuclear receptors	Androgen receptor	Complete or partial androgen insensitivity (recessive missense mutations) Spinobulbar muscular atrophy (CAG repeat expansion)
Zinc finger proteins	WT1	WAGR syndrome: *W*ilms' tumor, *a*niridia, *g*enitourinary malformations, mental *r*etardation
Basic helix-loop-helix	MITF	Waardenburg's syndrome type 2A
Homeobox	IPF1	Maturity onset of diabetes mellitus type 4 (monoallelic mutation/ haploinsufficiency) Pancreatic agenesis (biallelic mutations)
Leucine zipper	Retina leucine zipper (NRL)	Autosomal dominant retinitis pigmentosa
High mobility group (HMG) proteins	SRY	Sex reversal
Forkhead	HNF4α, HNF1α, HNF1β	Maturity onset of diabetes mellitus types 1, 3, 5
Paired box	PAX3	Waardenburg's syndrome types 1 and 3
T-box	TBX5	Holt-Oram syndrome (thumb anomalies, atrial or ventricular septum defects, phocomelia)
Cell cycle control proteins	P53	Li-Fraumeni syndrome, other cancers
Co-activators	CREB binding protein (CREBBP)	Rubinstein-Taybi syndrome
General transcription factors	TATA-binding protein (TBP)	Spinocerebellar ataxia 17 (CAG expansion)
Transcription elongation factor	VHL	von Hippel–Lindau syndrome (renal cell carcinoma, pheochromocytoma, pancreatic tumors, hemangioblastomas) Autosomal dominant inheritance, somatic inactivation of second allele (Knudson two-hit model)
Runt	RUNX1	Familial thrombocytopenia with propensity to acute myelogenous leukemia
Chimeric proteins due to translocations	PML-RAR	Acute promyelocytic leukemia t(15;17)(q22;q11.2-q12) translocation

Abbreviations: CREB, cAMP responsive element–binding protein; HNF, hepatocyte nuclear factor; PML, promyelocytic leukemia; RAR, retinoic acid receptor; SRY, sex-determining region Y; VHL, von Hippel–Lindau.

heritable and leads to the preferential expression of one of the parental alleles, which deviates from the usual biallelic expression seen for the majority of genes. Remarkably, imprinting can be limited to a subset of tissues. Imprinting is mediated through DNA methylation of one of the alleles. The epigenetic marks on imprinted genes are maintained throughout life, but during zygote formation, they are activated or inactivated in a sex-specific manner (imprint reset) **(Fig. 479-8)**, which allows a differential expression pattern in the fertilized egg and the subsequent mitotic divisions. Appropriate expression of imprinted genes is important for normal development and cellular functions. Imprinting defects and uniparental disomy, which is the inheritance of two chromosomes or chromosomal regions from the same parent, are the cause of several developmental disorders such as Beckwith-Wiedemann syndrome, Silver-Russell syndrome, Angelman's syndrome, and Prader-Willi syndrome (see below). Monoallelic loss-of-function mutations in the *GNAS1* gene lead to Albright's hereditary osteodystrophy (AHO). Paternal transmission of *GNAS1* mutations leads to an isolated AHO phenotype (pseudopseudohypoparathyroidism), whereas maternal transmission leads to AHO in combination with hormone resistance to parathyroid hormone, thyrotropin, and gonadotropins (pseudohypoparathyroidism type IA). These phenotypic differences are explained by tissue-specific imprinting of the *GNAS1* gene, which is expressed primarily from the maternal allele in the thyroid, gonadotropes, and the proximal renal tubule. In most other tissues, the *GNAS1* gene is expressed biallelically. In patients with isolated renal resistance to parathyroid hormone (pseudohypoparathyroidism type IB), defective imprinting of the *GNAS1* gene results in decreased $G_s\alpha$ expression in the proximal renal tubules. Rett syndrome is an X-linked dominant disorder resulting in developmental regression and stereotypic hand movements in affected girls. It is caused by mutations in the *MECP2* gene, which encodes a methyl-binding protein. The ensuing aberrant methylation results in abnormal gene expression in neurons, which are otherwise normally developed.

Remarkably, epigenetic differences also occur among monozygotic twins. Although twins are epigenetically indistinguishable during the early years of life, older monozygotic twins exhibit differences in the overall content and genomic distribution of DNA methylation and histone acetylation, which would be expected to alter gene expression in various tissues.

In cancer, the epigenome is characterized by simultaneous losses and gains of DNA methylation in different genomic regions, as well as repressive histone modifications. Hyper- and hypomethylation are associated with mutations in genes that control DNA methylation. Hypomethylation is thought to remove normal control mechanisms that prevent expression of repressed DNA regions. It is also associated with genomic instability. Hypermethylation, in contrast, results in the silencing of CpG islands in promoter regions of genes, including tumor-suppressor genes. Epigenetic alterations are more easily reversible compared to genetic changes; modification of the epigenome with demethylating agents and histone deacetylases is being used in the treatment of various malignancies.

TRANSMISSION OF GENETIC DISEASE

Origins and Types of Mutations The term *mutation* or *variant* is used to designate the process of generating genetic variations as well as the effect of these alterations. A *mutation* can be defined as any change in the primary nucleotide sequence of DNA regardless of its functional consequences, although it often has a negative connotation. There has been a shift towards using the more neutral term *variant* to describe sequence changes, and it is now recommended by several professional organizations and guidelines instead of *mutation*. Some variants may be lethal, others are less deleterious, and some may confer an evolutionary advantage. Variations can occur in the germline (sperm or oocytes); these can be transmitted to progeny. Alternatively, variants can occur during embryogenesis or in somatic tissues. Variations that occur during development lead to *mosaicism*, a situation in which tissues are composed of cells with different genetic constitutions. If the germline is mosaic, a mutation can be transmitted to some progeny but not others, which sometimes leads to confusion in assessing the pattern of inheritance. Somatic mutations that do not affect cell survival can sometimes be detected because of variable phenotypic effects in tissues (e.g., pigmented lesions in McCune-Albright syndrome). Other somatic mutations are associated with neoplasia because they confer a growth advantage to cells. Epigenetic events may also influence gene expression or facilitate genetic damage. With the exception of triplet nucleotide repeats, which can expand (see below), variations are usually stable.

Sequence variants are structurally diverse—they can involve the entire genome, as in triploidy (one extra set of chromosomes), or gross numerical or structural alterations in chromosomes or individual genes. Large deletions may affect a portion of a gene or an entire gene, or, if several genes are involved, they may lead to a *contiguous gene syndrome*. Unequal crossing-over between homologous genes can result in fusion gene mutations, as illustrated by color blindness. Variations involving single nucleotides are referred to as *point mutations*.

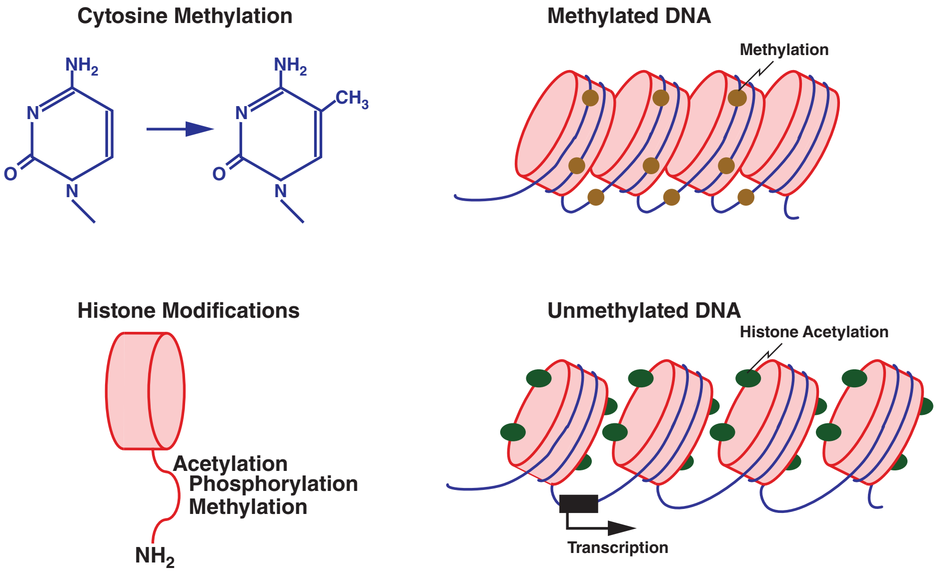

FIGURE 479-7 Epigenetic modifications of DNA and histones. Methylation of cytosine residues is associated with gene silencing. Methylation of certain genomic regions is inherited (imprinting), and it is involved in the silencing of one of the two X chromosomes in females (X-inactivation). Alterations in methylation can also be acquired, e.g., in cancer cells. Covalent posttranslational modifications of histones play an important role in altering DNA accessibility and chromatin structure and hence in regulating transcription. Histones can be reversibly modified in their amino-terminal tails, which protrude from the nucleosome core particle, by acetylation of lysine, phosphorylation of serine, methylation of lysine and arginine residues, and sumoylation. Acetylation of histones by histone acetylases (HATs), e.g., leads to unwinding of chromatin and accessibility to transcription factors. Conversely, deacetylation by histone deacetylases (HDACs) results in a compact chromatin structure and silencing of transcription.

Substitutions are called *transitions* if a purine is replaced by another purine base (A ↔ G) or if a pyrimidine is replaced by another pyrimidine (C ↔ T). Changes from a purine to a pyrimidine, or vice versa, are referred to as *transversions*. If the DNA sequence change occurs in a coding region and alters an amino acid, it is called a *missense mutation*. Depending on the functional consequences of such a missense mutation, amino acid substitutions in different regions of the protein can lead to distinct phenotypes.

Variants can occur in all domains of a gene **(Fig. 479-9)**. A point mutation occurring within the coding region leads to an amino acid substitution if the codon is altered **(Fig. 479-10)**. Point mutations that introduce a premature stop codon result in a truncated or missing protein. Large deletions may affect a portion of a gene or an entire gene, whereas small deletions and insertions alter the reading frame if they do not represent a multiple of three bases. These "frameshift" mutations, also designated as *amphigoric* amino acid changes, lead to an entirely altered carboxy terminus. Mutations in intronic sequences or in exon junctions may destroy or create splice donor or splice acceptor sites. Variants may also be found in the regulatory sequences of genes, resulting in reduced or enhanced gene transcription.

Certain DNA sequences are particularly susceptible to mutagenesis. Successive pyrimidine residues (e.g., T-T or C-C) are subject to the formation of ultraviolet light–induced photoadducts. If these pyrimidine dimers are not repaired by the nucleotide excision repair pathway, mutations will be introduced after DNA synthesis. The dinucleotide C-G, or CpG, is also a hot spot for a specific type of alteration. In this case, methylation of the cytosine is associated with an enhanced rate of deamination to uracil, which is then replaced with thymine. This C → T transition (or G → A on the opposite strand) accounts for at least one-third of point mutations associated with polymorphisms and mutations. In addition to the fact that certain types of mutations (C → T or G → A) are relatively common, the nature of the genetic code also results in overrepresentation of certain amino acid substitutions.

Polymorphisms are sequence variations that have a frequency of at least 1%. Usually, they do not result in a perceptible phenotype; the term *variant* is now preferred for the description of these sequence changes because allele frequency and functional consequences are often not known. Often, they consist of single base-pair substitutions that do not alter the protein coding sequence because of the degenerate nature of the genetic code (synonymous polymorphism), although it is possible that some might alter mRNA stability, translation, or the amino acid sequence (nonsynonymous polymorphism) (Fig. 479-10). The detection of sequence variants poses a practical problem because it is often unclear whether it creates a change with functional consequences or a benign variation. In this situation, the sequence alteration is also described as *variant of unknown significance (VUS)*.

MUTATION RATES Mutations represent an important cause of genetic diversity as well as disease. Mutation rates are difficult to determine in humans because many mutations are silent and because testing is often not adequate to detect the phenotypic consequences. Mutation rates vary in different genes but are estimated to occur at a rate of $\sim 10^{-10}$/bp per cell division. Germline mutation rates (as opposed to somatic mutations) are relevant in the transmission of genetic disease. Because the population of oocytes is established very early in development, only ~20 cell divisions are required for completed oogenesis, whereas spermatogenesis involves ~30 divisions by the time of puberty and 20 cell divisions each year thereafter. Consequently, the probability of acquiring new point mutations is much greater in the male germline than the female germline, in which rates of aneuploidy are increased. Thus, the incidence of new point mutations in spermatogonia increases with paternal age (e.g., achondrodysplasia, Marfan's syndrome, neurofibromatosis). It is estimated that about 1 in 10 sperm carries a new deleterious mutation. The rates for new mutations are calculated most readily for autosomal dominant and X-linked disorders and are $\sim 10^{-5}$–10^{-6}/locus per generation. Because most monogenic diseases are relatively rare, new mutations account for a significant fraction of cases. This is important in the context of genetic counseling because a new mutation can be transmitted to the affected individual, but this does not necessarily imply that the parents are at risk to transmit the disease to other children. An exception to this is when the new mutation occurs early in germline development, leading to *gonadal mosaicism*.

UNEQUAL CROSSING-OVER Normally, DNA recombination in germ cells occurs with remarkable fidelity to maintain the precise junction sites for the exchanged DNA sequences (Fig. 479-6). However, mispairing of homologous sequences leads to unequal crossover, with gene duplication on one of the chromosomes and gene deletion on the other chromosome. A significant fraction of growth hormone (*GH*) gene deletions, for example, involve unequal crossing-over **(Chap. 391)**. The *GH* gene is a member of a large gene cluster that includes a GH variant gene as well as several structurally related chorionic somatomammotropin genes and pseudogenes (highly homologous but functionally inactive relatives of a normal gene). Because such gene clusters contain multiple homologous DNA sequences arranged in tandem, they are particularly prone to undergo recombination and, consequently, gene duplication or deletion. Duplication of the *PMP22* gene because of unequal crossing-over results in increased gene dosage and type IA Charcot-Marie-Tooth disease. In contrast, unequal crossing-over resulting in deletion of *PMP22* causes a distinct neuropathy called *hereditary neuropathy with liability to pressure palsies* (HNPP) **(Chap. 457)**.

Glucocorticoid-remediable aldosteronism (GRA) is caused by a gene fusion or rearrangement involving the genes that encode aldosterone synthase (*CYP11B2*) and steroid 11β-hydroxylase (*CYP11B1*), normally arranged in tandem on chromosome 8q. These two genes are 95% identical, predisposing to gene duplication and deletion by

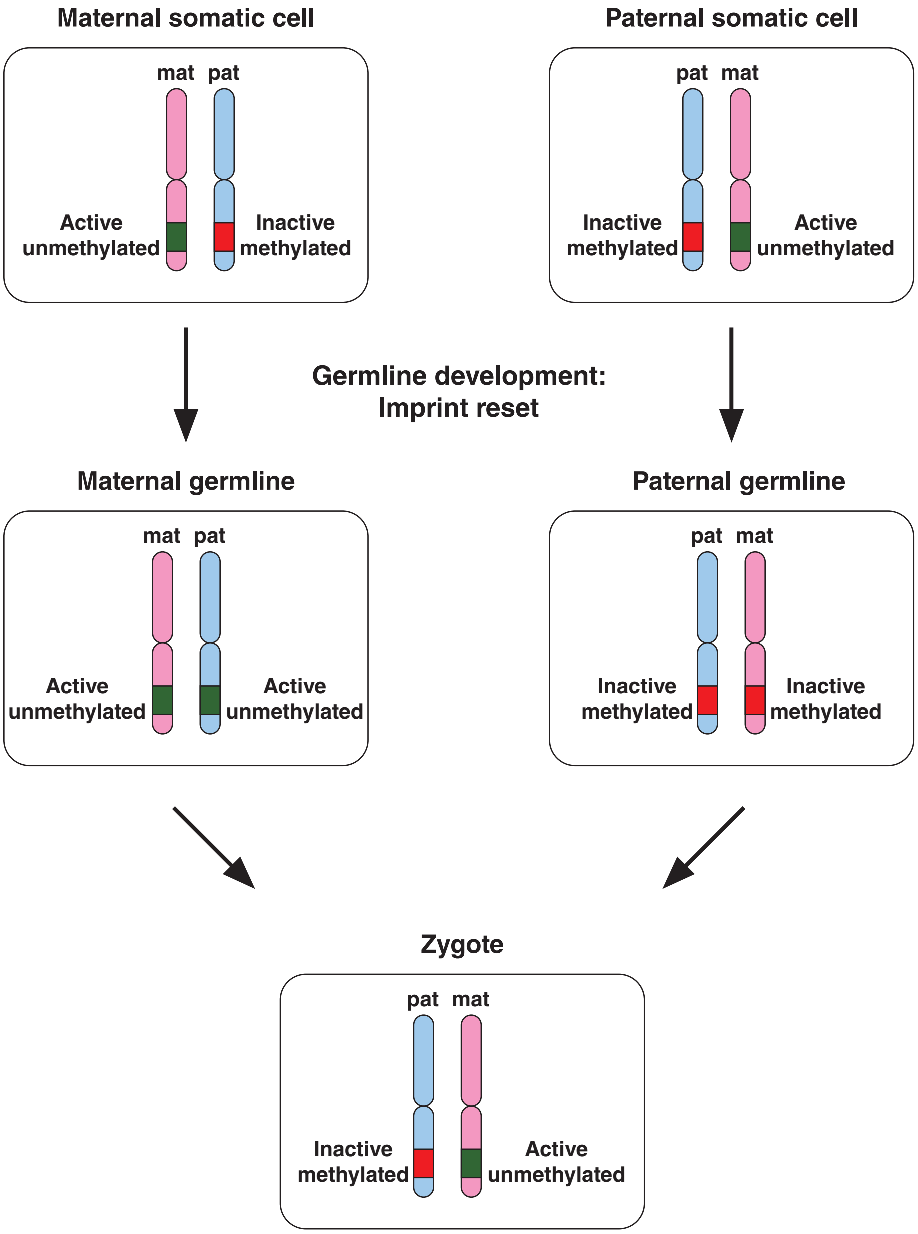

FIGURE 479-8 A few genomic regions are imprinted in a parent-specific fashion. The unmethylated chromosomal regions are actively expressed, whereas the methylated regions are silenced. In the germline, the imprint is reset in a parent-specific fashion: both chromosomes are unmethylated in the maternal (mat) germline and methylated in the paternal (pat) germline. In the zygote, the resulting imprinting pattern is identical with the pattern in the somatic cells of the parents.

unequal crossing-over. The rearranged gene product contains the regulatory regions of 11β-hydroxylase fused to the coding sequence of aldosterone synthetase. Consequently, the latter enzyme is expressed in the adrenocorticotropic hormone (ACTH)–dependent zona fasciculata of the adrenal gland, resulting in overproduction of mineralocorticoids and hypertension **(Chap. 398)**.

Gene conversion refers to a nonreciprocal exchange of homologous genetic information. It has been used to explain how an internal portion of a gene is replaced by a homologous segment copied from another allele or locus; these genetic alterations may range from a few nucleotides to a few thousand nucleotides. As a result of gene conversion, it is possible for short DNA segments of two chromosomes to be identical, even though these sequences are distinct in the parents. A practical consequence of this phenomenon is that nucleotide substitutions can occur during gene conversion between related genes, often altering the function of the gene. In disease states, gene conversion often involves intergenic exchange of DNA between a gene and a related pseudogene. For example, the 21-hydroxylase gene (*CYP21A2*) is adjacent to a nonfunctional pseudogene (*CYP21A1P*). Many of the nucleotide substitutions that are found in the *CYP21A2* gene in patients with congenital adrenal hyperplasia correspond to sequences that are present in the *CYP21A1P* pseudogene, suggesting gene conversion as one cause of mutagenesis. In addition, mitotic gene conversion has been suggested as a mechanism to explain revertant mosaicism in which an inherited mutation is "corrected" in certain cells. For example, patients with autosomal recessive generalized atrophic benign epidermolysis bullosa have acquired reverse mutations in one of the two mutated *COL17A1* alleles, leading to clinically unaffected patches of skin.

INSERTIONS AND DELETIONS Although many instances of insertions and deletions occur as a consequence of unequal crossing-over, there is also evidence for internal duplication, inversion, or deletion of DNA sequences. The fact that certain deletions or insertions appear to occur repeatedly as independent events indicates that specific regions within the DNA sequence predispose to these errors. For example, certain regions of the *DMD* gene, which encodes dystrophin, appear to be hot spots for deletions and result in muscular dystrophy **(Chap. 460)**. Some regions within the human genome are rearrangement hot spots and lead to CNVs.

ERRORS IN DNA REPAIR Because mutations caused by defects in DNA repair accumulate as somatic cells divide, these types of mutations are particularly important in the context of neoplastic disorders. Several genetic disorders involving DNA repair enzymes underscore their importance. Patients with xeroderma pigmentosum have defects in DNA damage recognition or in the nucleotide excision and repair pathway **(Chap. 81)**. Exposed skin is dry and pigmented and is extraordinarily sensitive to the mutagenic effects of ultraviolet irradiation. Variants in more than 10 different genes have been shown to cause the different forms of xeroderma pigmentosum.

Ataxia-telangiectasia is a multisystem disorder that includes progressive neurodegenerative cerebellar ataxia, immunologic defects, telangiectatic lesions, lymphomas and leukemias, and hypersensitivity to ionizing radiation **(Chap. 450)**. The discovery of the ataxia-telangiectasia mutated (*ATM*) gene revealed that it is homologous to genes involved in DNA repair and control of cell cycle checkpoints. Mutations in the *ATM* gene give rise to defects in meiosis as well as increasing susceptibility to damage from ionizing radiation. Fanconi's anemia is also associated with an increased risk of multiple acquired genetic abnormalities. It is characterized by diverse congenital anomalies and a strong predisposition to develop aplastic anemia and acute myelogenous leukemia **(Chap. 109)**. Cells from these patients are susceptible to chromosomal breaks caused by a defect in genetic recombination. It can be caused by mutations in the multiple genes forming the Fanconi's anemia pathway, which is involved in DNA repair and replication. HNPCC (Lynch syndrome) is characterized by autosomal dominant transmission of colon cancer, young age (<50 years) of presentation, predisposition to lesions in the proximal large bowel, and associated malignancies such as uterine cancer and ovarian cancer. HNPCC is predominantly caused by mutations in one of several different mismatch repair (MMR) genes including MutS

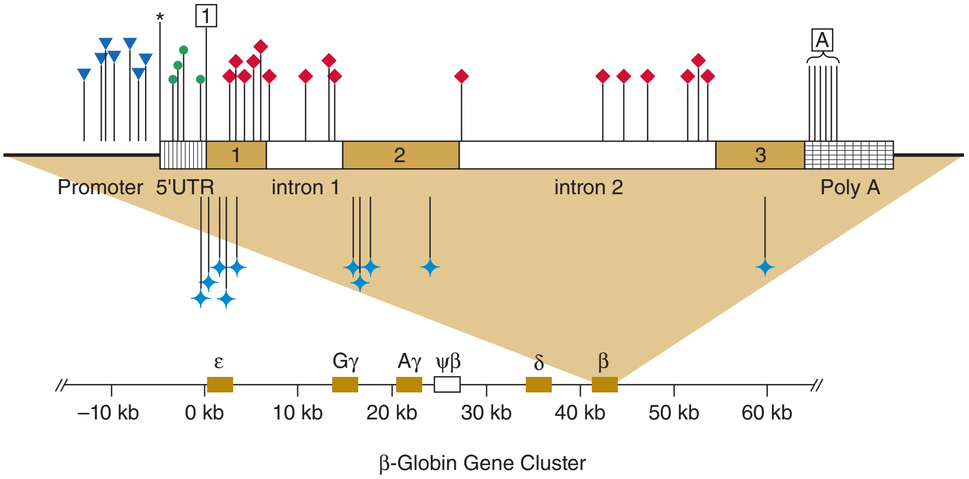
*
1
A
1
2
3
Promoter
5′UTR
intron 1
intron 2
Poly A
ε
Gγ
Aγ
ψβ
δ
β
–10 kb
0 kb
10 kb
20 kb
30 kb
40 kb
50 kb
60 kb
β-Globin Gene Cluster

FIGURE 479-9 Point mutations causing β thalassemia as example of allelic heterogeneity. The *β-globin* gene is located in the globin gene cluster. Point mutations occur in the promoter, the CAP site, the 5′-untranslated region, the initiation codon, each of the three exons, the introns, or the polyadenylation signal. Many mutations introduce missense or nonsense mutations, whereas others cause defective RNA splicing. Not shown here are deletion mutations of the β-globin gene or larger deletions of the globin locus that can also result in thalassemia. ▼, promoter mutations; *, CAP site; ●, 5′UTR; 1, initiation codon; ◆, defective RNA processing; ✦, missense and nonsense mutations; A, Poly A signal.

homologue 2 (*MSH2*), MutL homologue 1 and 6 (*MLH1, MLH6*), *MSH6*, *PMS1*, and *PMS2* **(Chap. 86)**. These proteins are involved in the detection of nucleotide mismatches and in the recognition of slipped-strand trinucleotide repeats. Germline mutations in these genes lead to microsatellite instability and a high mutation rate in colon cancer. Genetic screening tests for this disorder are now being used for families considered to be at risk. Recognition of HNPCC allows early screening with colonoscopy and the implementation of prevention strategies using nonsteroidal anti-inflammatory drugs.

UNSTABLE DNA SEQUENCES *Trinucleotide repeats* may be unstable and expand beyond a critical number. Mechanistically, the expansion is thought to be caused by unequal recombination and slipped mispairing. A premutation represents a small increase in trinucleotide copy number. In subsequent generations, the expanded repeat may increase further in length and result in an increasingly severe phenotype, a process called *dynamic mutation* (see below for discussion of anticipation). Trinucleotide expansion was first recognized as a cause of the fragile X syndrome, one of the most common causes of intellectual disability. Other disorders arising from a similar mechanism include Huntington's disease, X-linked spinobulbar muscular atrophy, and myotonic dystrophy. Malignant cells are also characterized by genetic instability, indicating a breakdown in mechanisms that regulate DNA repair and the cell cycle.

Functional Consequences of Mutations Functionally, mutations can be broadly classified as gain-of-function and loss-of-function mutations. Gain-of-function mutations are typically dominant (e.g., they result in phenotypic alterations when a single allele is affected). Inactivating mutations are usually recessive, and an affected individual is homozygous or compound heterozygous (e.g., carrying two different mutant alleles of the same gene) for the disease-causing mutations. Alternatively, mutation in a single allele can result in *haploinsufficiency*, a situation in which one normal allele is not sufficient to maintain a normal phenotype. Haploinsufficiency is a commonly observed mechanism in diseases associated with mutations in transcription factors (Table 479-2). Remarkably, the clinical features among patients with an identical mutation often vary significantly. One mechanism underlying this variability consists in the influence of modifying genes. Haploinsufficiency can also affect the expression of rate-limiting enzymes. For example, haploinsufficiency in enzymes involved in heme synthesis can cause porphyrias **(Chap. 428)**.

An increase in dosage of a gene product may also result in disease, as illustrated by the duplication of the *DAX1 (NR0B1)* gene in dosage-sensitive sex reversal **(Chap. 402)**. Mutation in a single allele can also result in loss of function due to a dominant-negative effect. In this case, the mutated allele interferes with the function of the normal (wild type) gene product by one of several different mechanisms: (1) a mutant protein may interfere with the function of a multimeric protein complex, as illustrated by mutations in type 1 collagen (*COL1A1, COL1A2*) genes in osteogenesis imperfecta **(Chap. 425)**; (2) a mutant protein may occupy binding sites on proteins or promoter response elements, as illustrated by thyroid hormone resistance β, a disorder in which

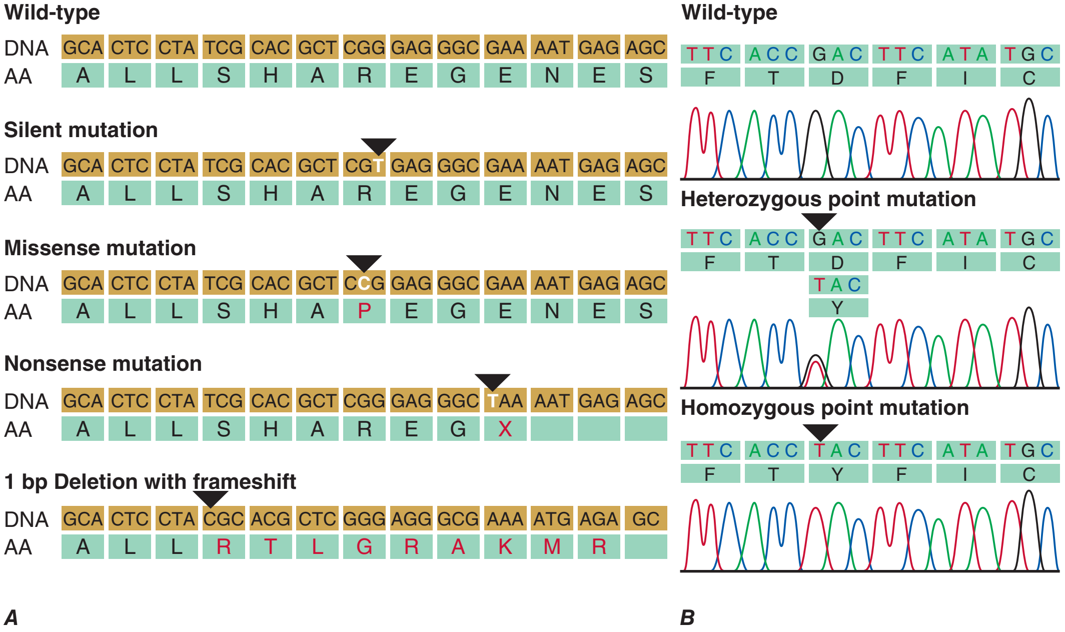
Wild-type
DNA GCA CTC CTA TCG CAC GCT CGG GAG GGC GAA AAT GAG AGC
AA A L L S H A R E G E N E S
Silent mutation
DNA GCA CTC CTA TCG CAC GCT CGT GAG GGC GAA AAT GAG AGC
AA A L L S H A R E G E N E S
Missense mutation
DNA GCA CTC CTA TCG CAC GCT CCG GAG GGC GAA AAT GAG AGC
AA A L L S H A P E G E N E S
Nonsense mutation
DNA GCA CTC CTA TCG CAC GCT CGG GAG GGC TAA AAT GAG AGC
AA A L L S H A R E G X
1 bp Deletion with frameshift
DNA GCA CTC CTA CGC ACG CTC GGG AGG GCG AAA ATG AGA GC
AA A L L R T L G R A K M R
A
Wild-type
TTC ACC GAC TTC ATA TGC
F T D F I C
Heterozygous point mutation
TTC ACC GAC TTC ATA TGC
F T D F I C
TAC
Y
Homozygous point mutation
TTC ACC TAC TTC ATA TGC
F T Y F I C
B

FIGURE 479-10 ***A.*** Examples of mutations (now commonly referred to as *variations*). The coding strand is shown with the encoded amino acid sequence. ***B.*** Chromatograms of sequence analyses after amplification of genomic DNA by polymerase chain reaction.

inactivated thyroid hormone receptor β binds to target genes and functions as an antagonist of normal receptors **(Chap. 394)**; or (3) a mutant protein can be cytotoxic as in α_1 antitrypsin deficiency **(Chap. 303)** or autosomal dominant neurohypophyseal diabetes insipidus **(Chap. 393)**, in which the abnormally folded proteins are trapped within the endoplasmic reticulum and ultimately cause cellular damage.

Genotype and Phenotype

ALLELES, GENOTYPES, AND HAPLOTYPES An observed trait is referred to as a *phenotype*; the genetic information defining the phenotype is called the *genotype*. Alternative forms of a gene or a genetic marker are referred to as *alleles*. Alleles may be polymorphic variants of nucleic acids that have no apparent effect on gene expression or function. In other instances, these variants may have subtle effects on gene expression, thereby conferring adaptive advantages associated with genetic diversity. On the other hand, allelic variants may reflect mutations that clearly alter the function of a gene product. The common Glu6Val (E6V) sickle cell mutation in the β-*globin* gene and the ΔF508 deletion of phenylalanine (F) in the *CFTR* gene are examples of allelic variants of these genes that result in disease. Because each individual has two copies of each chromosome (one inherited from the mother and one inherited from the father), an individual can have only two alleles at a given locus. However, there can be many different alleles in the population. The normal or common allele is usually referred to as *wild type*. When alleles at a given locus are identical, the individual is *homozygous*. Inheriting identical copies of a mutant allele occurs in many autosomal recessive disorders, particularly in circumstances of consanguinity or isolated populations. If the alleles are different on the maternal and the paternal copy of the gene, the individual is *heterozygous* at this locus (Fig. 479-10). If two different mutant alleles are inherited at a given locus, the individual is said to be a *compound heterozygote*. *Hemizygous* is used to describe males with a mutation in an X chromosomal gene or a female with a loss of one X chromosomal locus.

Genotypes describe the specific alleles at a particular locus. For example, there are three common alleles (E2, E3, E4) of the apolipoprotein E (*APOE*) gene. The genotype of an individual can therefore be described as *APOE3/4* or *APOE4/4* or any other variant. These designations indicate which alleles are present on the two chromosomes in the *APOE* gene at locus 19q13.2. In other cases, the genotype might be assigned arbitrary numbers (e.g., 1/2) or letters (e.g., B/b) to distinguish different alleles.

A *haplotype* refers to a group of alleles that are closely linked together at a genomic locus (Fig. 479-4). Haplotypes are useful for tracking the transmission of genomic segments within families and for detecting evidence of genetic recombination if the crossover event occurs between the alleles (Fig. 479-6). As an example, various alleles of the histocompatibility locus antigens (HLA) at the major histocompatibility complex (MHC) on chromosome 6 are used to establish haplotypes associated with certain disease states. For example, 21-hydroxylase deficiency, complement deficiency, and hemochromatosis are each associated with specific HLA haplotypes. It is now recognized that these genes lie in close proximity to the HLA locus, which explains why HLA associations were identified even before the disease genes were cloned and localized. In other cases, specific HLA associations with diseases such as ankylosing spondylitis (HLA-B27) or type 1 diabetes mellitus (HLA-DR4) reflect the role of specific HLA allelic variants in susceptibility to these autoimmune diseases. The characterization of common SNP haplotypes in numerous populations from different parts of the world has provided the necessary tools for association studies designed to detect genes involved in the pathogenesis of complex disorders (Table 479-1). The presence or absence of certain haplotypes can also be relevant for the customized choice of medical therapies (pharmacogenomics) or may have value for preventive strategies.

Genotype-phenotype correlation describes the association of a specific mutation and the resulting phenotype. The phenotype may differ depending on the location or type of the mutation in some genes. For example, in von Hippel–Lindau disease, an autosomal dominant multisystem disease that can include renal cell carcinoma, hemangioblastomas, and pheochromocytomas, among others, the phenotype varies greatly, and the identification of the specific mutation can be clinically useful in order to predict the spectrum of disease manifestations.

ALLELIC HETEROGENEITY *Allelic heterogeneity* refers to the fact that different mutations in the same genetic locus can cause an identical or similar phenotype. For example, many different mutations of the β-globin locus can cause β thalassemia **(Table 479-3)** (Fig. 479-9). In essence, allelic heterogeneity reflects the fact that many different mutations can alter protein structure and function. For this reason, maps of inactivating mutations in genes usually show a near-random distribution. Exceptions include (1) a founder effect, in which a particular mutation that does not affect reproductive capacity can be traced to a single individual; (2) "hot spots" for mutations, in which the nature of the DNA sequence predisposes to a recurring mutation; and (3) localization of mutations to certain domains that are particularly critical for protein function. Allelic heterogeneity creates a practical problem for genetic testing because one must often examine the entire genetic locus for mutations, because these can differ in each patient. For example, ~2000 variants have been identified in the *CFTR* gene to date, although some of them are very rare and some may not be disease-causing (Fig. 479-3). Mutational analysis may initially focus on a panel of mutations that are particularly frequent (often taking the ethnic background of the patient into account), but a negative result does not exclude the presence of a mutation elsewhere in the gene. Until recently, mutational analyses tended to focus on the coding region of a gene without considering regulatory and intronic regions. However, disease-causing mutations may be located outside the coding regions, so negative results need to be interpreted with caution. The increasingly widespread access to comprehensive sequencing technologies, WES and WGS, greatly facilitates unbiased mutational analyses. However, comprehensive sequencing can result in significant diagnostic challenges because the detection of a sequence alteration is not always sufficient to establish that it has a causal role (VUS).

PHENOTYPIC HETEROGENEITY *Phenotypic heterogeneity* occurs when more than one phenotype is caused by allelic mutations (e.g., by different mutations in the same gene) (Table 479-3). For example, laminopathies are monogenic multisystem disorders that result from mutations in the *LMNA* gene, which encodes the nuclear lamins A and C. Multiple autosomal dominant and recessive disorders are caused by mutations in the *LMNA* gene. They include several forms of lipodystrophies, Emery-Dreifuss muscular dystrophy, progeria syndromes, a form of neuronal Charcot-Marie-Tooth disease (type 2B1), and a group of overlapping syndromes. Remarkably, hierarchical cluster analysis has revealed that the phenotypes vary depending on the position of the mutation (*genotype-phenotype correlation*). Similarly, identical mutations in the *FGFR2* gene can result in very distinct phenotypes: Crouzon's syndrome (craniofacial synostosis) or Pfeiffer's syndrome (acrocephalopolysyndactyly).

LOCUS OR NONALLELIC HETEROGENEITY AND PHENOCOPIES *Nonallelic* or *locus heterogeneity* refers to the situation in which a similar disease phenotype results from mutations at different genetic loci (Table 479-3). This often occurs when more than one gene product produces different subunits of an interacting complex or when different genes are involved in the same genetic cascade or physiologic pathway. For example, osteogenesis imperfecta can arise from mutations in two different procollagen genes (*COL1A1* or *COL1A2*) that are located on different chromosomes and can involve multiple other genes **(Chap. 425)**. The effects of inactivating mutations in these two genes are similar because the protein products comprise different subunits of the helical collagen fiber. Similarly, muscular dystrophy syndromes can be caused by mutations in various genes, consistent with the fact that it can be transmitted in an X-linked (Duchenne or Becker), autosomal dominant (limb-girdle muscular dystrophy type 1), or autosomal recessive (limb-girdle muscular dystrophy type 2) manner **(Chap. 460)**. Mutations in the X-linked *DMD* gene, which encodes dystrophin, are the most common cause of muscular dystrophy. This feature reflects the large size of the gene (2.3 MB, 79 exons), as well as the fact that the phenotype is expressed in hemizygous males because they have only

TABLE 479-3 Selected Examples of Phenotypic Heterogeneity and Locus Heterogeneity

Phenotypic Heterogeneity

GENE, PROTEIN	PHENOTYPE	INHERITANCE	OMIM
LMNA, Lamin A/C	Emery-Dreifuss muscular dystrophy (AD)	AD	181350
	Familial partial lipodystrophy Dunnigan	AD	151660
	Hutchinson-Gilford progeria	AD	176670
	Atypical Werner's syndrome	AD	150330
	Dilated cardiomyopathy 1A	AD	115200
	Familial atrial fibrillation 3	AD	607554
	Charcot-Marie-Tooth type 2B1	AR	605588
KRAS	Noonan's syndrome	AD	163950
	Cardio-facio-cutaneous syndrome 1	AD	115150

Locus Heterogeneity

PHENOTYPE	GENE	CHROMOSOMAL LOCATION	PROTEIN
Familial hypertrophic cardiomyopathy	*MYH7*	14q11.2	Myosin heavy chain beta
Genes encoding sarcomeric proteins	*TNNT2*	1q32.1	Troponin-T2
	TPM1	15q22.2	Tropomyosin alpha
	MYBPC3	11p11q	Myosin-binding protein C
	TNNC1	19q13.4	Troponin 1
	MYL2	12q24.11	Myosin light chain 2
	MYL3	3p21.31	Myosin light chain 3
	TTN	2q31.2	Cardiac titin
	ACTC	15q14	Cardiac alpha actin
	MYH6	14q11.2	Myosin heavy chain alpha
	MYLK2	20q11.21	Myosin light-peptide kinase
	CAV3	3p25	Caveolin 3
Genes encoding nonsarcomeric proteins	*MT-T1*	Mitochondrial	tRNA isoleucine
	MT-TG	Mitochondrial	tRNA glycine
	PRKAG2	7q36.1	AMP-activated protein kinase γ2 subunit
	DMPK	19q13.32	Myotonin protein kinase (myotonic dystrophy)
	FRDA	9q21.11	Frataxin (Friedreich's ataxia)
Polycystic kidney disease	*PKD1*	16p13.3	Polycystin 1 (AD)
	PKD2	4q22.1	Polycystin 2 (AD)
	PKHD1	6p21.1-p12.2	Fibrocystin/polyductin (AR)
Noonan's syndrome	*PTPN11*	12q24.13	Protein-tyrosine phosphatase 2c
	KRAS	12p12.1	KRAS

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; OMIM, Online Mendelian Inheritance in Man.

a single copy of the X chromosome. Dystrophin is associated with a large protein complex linked to the membrane-associated cytoskeleton in muscle. Mutations in several different components of this protein complex can also cause muscular dystrophy syndromes. Although the phenotypic features of some of these disorders are distinct, the phenotypic spectrum caused by mutations in different genes overlaps, thereby leading to nonallelic heterogeneity. It should be noted that mutations in dystrophin are also associated with allelic heterogeneity. For example, mutations in the *DMD* gene can cause either Duchenne's or the less severe Becker's muscular dystrophy, depending on the severity of the protein defect.

Recognition of nonallelic heterogeneity is important for several reasons: (1) the ability to identify disease loci in linkage studies is reduced by including patients with similar phenotypes but different genetic disorders; (2) genetic testing is more complex because several different genes need to be considered along with the possibility of different mutations in each of the candidate genes; and (3) novel information is gained about how genes or proteins interact, providing unique insights into molecular physiology.

Phenocopies refer to circumstances in which nongenetic conditions mimic a genetic disorder. For example, features of toxin- or drug-induced neurologic syndromes can resemble those seen in Huntington's disease, and vascular causes of dementia share phenotypic features with familial forms of Alzheimer's dementia **(Chap. 442)**. As in nonallelic heterogeneity, the presence of phenocopies has the potential to confound linkage studies and genetic testing. Patient history and subtle differences in phenotype can often provide clues that distinguish these disorders from related genetic conditions.

VARIABLE EXPRESSIVITY AND INCOMPLETE PENETRANCE The same genetic mutation may be associated with a phenotypic spectrum in different affected individuals, thereby illustrating the phenomenon of *variable expressivity*. This may include different manifestations of a disorder variably involving different organs (e.g., multiple endocrine neoplasia [MEN]), the severity of the disorder (e.g., cystic fibrosis), or the age of disease onset (e.g., Alzheimer's dementia). MEN 1 illustrates several of these features. In this autosomal dominant tumor syndrome, affected individuals carry an inactivating germline mutation that is inherited in an autosomal dominant fashion. After somatic inactivation of the alternate allele (loss of heterozygosity; Knudson two-hit model), patients can develop tumors of the parathyroid gland, endocrine pancreas, the pituitary gland, and dermatologic lesions **(Chap. 400)**. However, the pattern of tumors in the different glands, the age at which tumors develop, and the types of hormones produced vary among affected individuals, even within a given family. In this example, the phenotypic variability arises, in part, because of the requirement for a second somatic mutation in the normal copy of the *MEN1* gene, as well as the large array of different cell types that are susceptible to the effects of *MEN1* gene mutations. In part, variable expression reflects the influence of modifier genes, or genetic background, on the effects of a particular mutation. Even in identical twins, in whom the genetic constitution is essentially the same, one can occasionally see variable expression of a genetic disease.

Interactions with the environment can also influence the course of a disease. For example, the manifestations and severity of hemochromatosis can be influenced by iron intake **(Chap. 426)**, and the course of phenylketonuria is affected by exposure to phenylalanine in the diet **(Chap. 431)**. Other metabolic disorders, such as hyperlipidemias and porphyria, also fall into this category. Many mechanisms, including genetic effects and environmental influences, can therefore lead to variable expressivity. In genetic counseling, it is particularly important to recognize this variability, because one cannot always predict the course of disease, even when the mutation is known.

Penetrance refers to the proportion of individuals with a mutant genotype that express the phenotype. If all carriers of a mutant express the phenotype, penetrance is complete, whereas it is said to be *incomplete* or *reduced* if some individuals do not exhibit features of the phenotype. Dominant conditions with incomplete penetrance are characterized by skipping of generations with unaffected carriers transmitting the mutant gene. For example, hypertrophic obstructive cardiomyopathy (HCM) caused by mutations in the *myosin-binding protein C* gene is a dominant disorder with clinical features in only a subset of patients who carry the mutation **(Chap. 267)**. Patients who have the mutation, but no evidence of the disease, can still transmit the disorder to subsequent generations. In many conditions with postnatal onset, the proportion of gene carriers who are affected varies with age. Thus, when describing penetrance, one must specify age. For example, for disorders such as Huntington's disease or familial amyotrophic lateral sclerosis, which present later in life, the rate of penetrance is influenced by the age at which the clinical assessment is performed. *Imprinting* can also modify the penetrance of a disease. For example, in patients with AHO, mutations in the $G_s\alpha$ subunit (*GNAS1* gene) are expressed clinically only in individuals who inherit the mutation from their mother **(Chap. 422)**.

SEX-INFLUENCED PHENOTYPES Certain mutations affect males and females quite differently. In some instances, this is because the gene resides on the X or Y sex chromosomes (X-linked disorders and Y-linked disorders). As a result, the phenotype of mutated X-linked genes will be expressed fully in males but variably in heterozygous females, depending on the degree of X-inactivation and the function of the gene. For example, most heterozygous female carriers of factor VIII deficiency (hemophilia A) are asymptomatic because sufficient factor VIII is produced to prevent a defect in coagulation **(Chap. 121)**. On the other hand, some females heterozygous for the X-linked lipid storage defect caused by α-galactosidase A deficiency (Fabry's disease) experience mild manifestations of painful neuropathy, as well as other features of the disease **(Chap. 429)**. Because only males have a Y chromosome, mutations in genes such as *SRY*, which causes male-to-female sex reversal, or *DAZ* (deleted in azoospermia), which causes abnormalities of spermatogenesis, are unique to males **(Chap. 402)**.

Other diseases are expressed in a sex-limited manner because of the differential function of the gene product in males and females. Activating mutations in the luteinizing hormone receptor cause dominant male-limited precocious puberty in boys **(Chap. 403)**. The phenotype is unique to males because activation of the receptor induces testosterone production in the testis, whereas it is functionally silent in the immature ovary. Biallelic inactivating mutations of the follicle-stimulating hormone (FSH) receptor cause primary ovarian failure in females because the follicles do not develop in the absence of FSH action. In contrast, affected males have a more subtle phenotype, because testosterone production is preserved (allowing sexual maturation) and spermatogenesis is only partially impaired **(Chap. 403)**. In congenital adrenal hyperplasia, most commonly caused by 21-hydroxylase deficiency, cortisol production is impaired and ACTH stimulation of the adrenal gland leads to increased production of androgenic precursors **(Chap. 398)**. In females, the increased androgen level causes ambiguous genitalia, which can be recognized at the time of birth. In males, the diagnosis may be made on the basis of adrenal insufficiency at birth, because the increased adrenal androgen level does not alter sexual differentiation, or later in childhood, because of the development of precocious puberty. Hemochromatosis is more common in males than in females, presumably because of differences in dietary iron intake and losses associated with menstruation and pregnancy in females **(Chap. 426)**.

Chromosomal Disorders *Chromosomal disorders* and the techniques used for their characterization have been discussed in detail in previous editions of this textbook. Chromosomal or cytogenetic disorders are caused by numerical (aneuploidy) or structural aberrations (deletions, duplications, translocations, inversions, dicentric and ring chromosomes, Robertsonian translocations) in chromosomes. They occur in ~1% of the general population, in 8% of stillbirths, and in close to 50% of spontaneously aborted fetuses. Indications for cytogenetic and cytogenomic chromosome analyses are summarized in **Table 479-4**. *Contiguous gene syndromes* (e.g., large deletions affecting several genes) have been useful for identifying the location of new disease-causing genes. Because of the variable size of gene deletions in different patients, a systematic comparison of phenotypes and locations of deletion breakpoints allows positions of particular genes to be mapped within the critical genomic region.

Monogenic Mendelian Disorders Monogenic human diseases are frequently referred to as *Mendelian disorders* because they obey the principles of genetic transmission originally set forth in Gregor Mendel's classic work. The continuously updated OMIM catalogue lists several thousand of these disorders and provides information about the clinical phenotype, molecular basis, allelic variants, and pertinent animal models (Table 479-1). The mode of inheritance for a given phenotypic trait or disease is determined by pedigree analysis. All affected and unaffected individuals in the family are recorded in a pedigree using standard symbols **(Fig. 479-11)**. The principles of allelic segregation, and the transmission of alleles from parents to children, are illustrated in **Fig. 479-12**. One dominant (A) allele and one recessive (a) allele can display three Mendelian modes of inheritance: autosomal dominant, autosomal recessive, and X-linked. About 65% of human monogenic disorders are autosomal dominant, 25% are autosomal recessive, and 5% are X-linked. Genetic testing is now readily available for the characterization of monogenic disorders and plays an important role in clinical medicine **(Chap. 480)**.

TABLE 479-4 Indications for Cytogenetic and Cytogenomic Analysis across the Life Span

TIMING OF TESTING	INDICATIONS FOR TESTING
Prenatal	Advanced maternal age Abnormalities on ultrasound Increased risk for genetic disorder on maternal serum screen
Neonatal and childhood	Multiple congenital anomalies Intellectual disability Autism spectrum disorders Developmental delay Failure to thrive Short stature Disorders of sexual development History of familial chromosomal alteration Cancer
Adult	Infertility Recurrent miscarriage Familial cancer

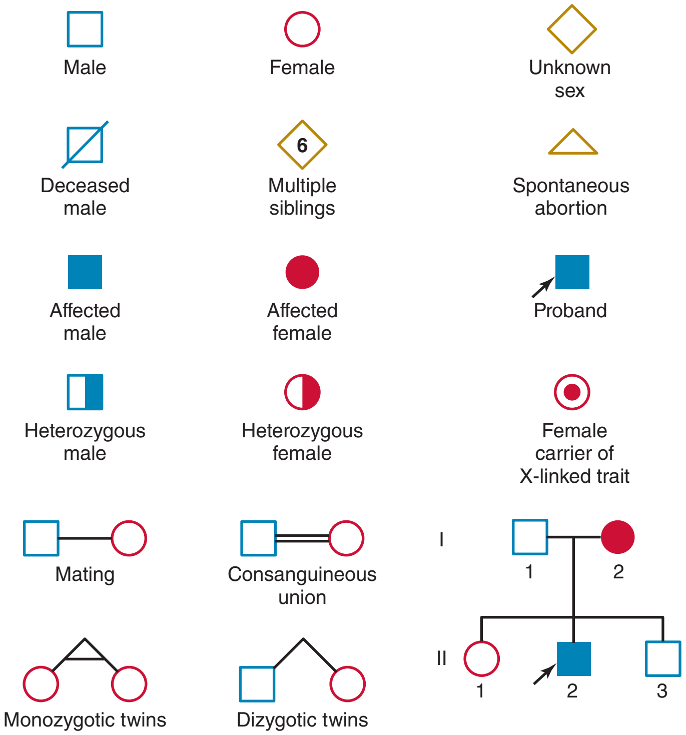

FIGURE 479-11 Standard pedigree symbols.

AUTOSOMAL DOMINANT DISORDERS In autosomal dominant disorders, mutations in a single allele are sufficient to cause the disease. In contrast to recessive disorders, in which disease pathogenesis is relatively straightforward because there is a biallelic loss of gene function, dominant disorders can be caused by various disease mechanisms, many of which are unique to the function of the genetic pathway involved. Mechanistically, the mutation may confer constitutive activation (gain of function), exert a dominant negative effect, or result in loss of function and haploinsufficiency.

In autosomal dominant disorders, individuals are affected in successive generations; the disease does not occur in the offspring of unaffected individuals. Males and females are affected with equal frequency because the defective gene resides on one of the 22 autosomes (Fig. 479-13*A*). Autosomal dominant mutations alter one of the two alleles at a given locus. Because the alleles segregate randomly at meiosis, the probability that an offspring will be affected is 50%. Unless there is a new germline mutation, an affected individual has an affected parent. Children with a normal genotype do not transmit the disorder. Due to differences in penetrance or expressivity (see above), the clinical manifestations of autosomal dominant disorders may be variable. Because of these variations, it is sometimes challenging to determine the pattern of inheritance.

It should be recognized, however, that some individuals acquire a mutated gene from an unaffected parent due to de novo germline mutations. They occur more frequently during later cell divisions in gametogenesis, which explains why siblings are rarely affected. As noted, new germline mutations occur more frequently in fathers of advanced age. For example, the average age of fathers with new germline mutations that cause Marfan's syndrome is ~37 years, whereas fathers who transmit the disease by inheritance have an average age of ~30 years.

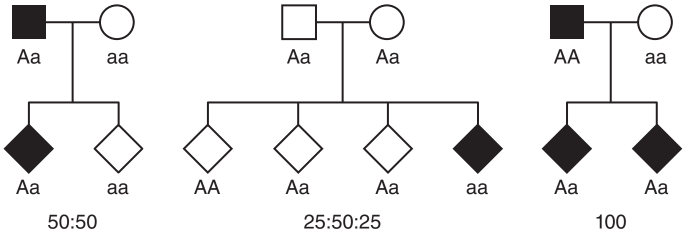

FIGURE 479-12 Segregation of alleles. Segregation of genotypes in the offspring of parents with one dominant (A) and one recessive (a) allele. The distribution of the parental alleles to their offspring depends on the combination present in the parents. Filled symbols = affected individuals.

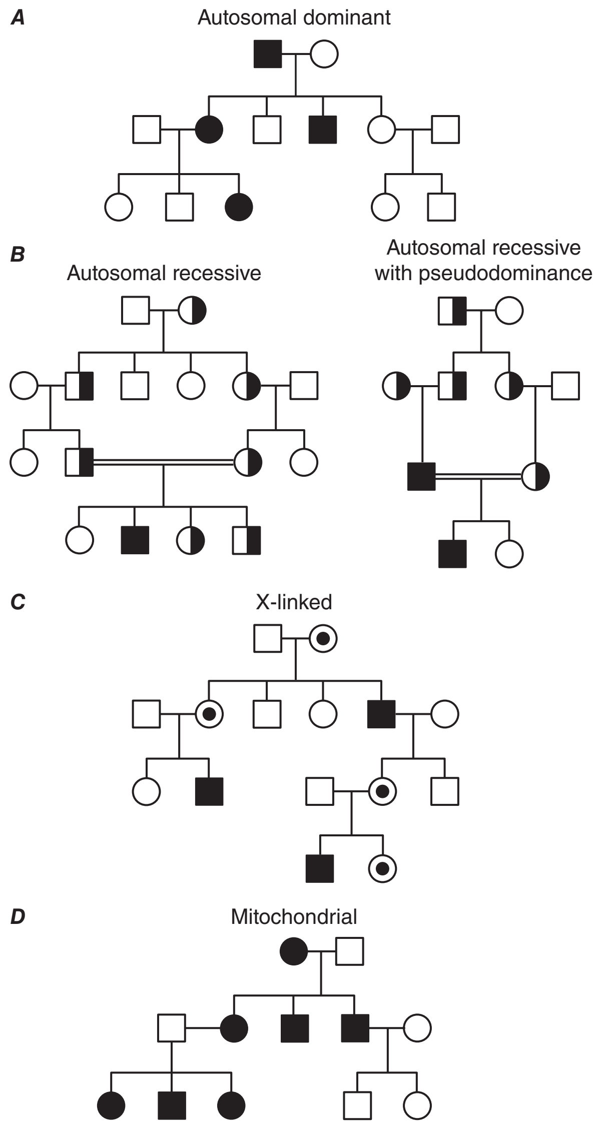

FIGURE 479-13 ***A.*** Dominant, ***B.*** recessive, ***C.*** X-linked, and ***D.*** mitochondrial (matrilinear) inheritance.

AUTOSOMAL RECESSIVE DISORDERS In recessive disorders, the mutated alleles result in a complete or partial loss of function. They frequently involve enzymes in metabolic pathways, receptors, or proteins in signaling cascades. In an autosomal recessive disease, the affected individual, who can be of either sex, is a homozygote or compound heterozygote for a single-gene defect. With a few important exceptions, autosomal recessive diseases are rare and occur more often in the context of parental consanguinity. The relatively high frequency of certain recessive disorders such as sickle cell anemia, cystic fibrosis, and thalassemia, is partially explained by a selective biologic advantage for the heterozygous state (see below). Although heterozygous carriers of a defective allele are usually clinically normal, they may display subtle differences in phenotype that only become apparent with more precise testing or in the context of certain environmental influences. In sickle cell anemia, for example, heterozygotes are normally asymptomatic. However, in situations of dehydration or diminished oxygen pressure, sickle cell crises can also occur in heterozygotes (Chap. 103).

In most instances, an affected individual is the offspring of heterozygous parents. In this situation, there is a 25% chance that the offspring will have a normal genotype, a 50% probability of a heterozygous state, and a 25% risk of homozygosity for the recessive alleles (Figs. 479-12 and 479-13*B*). In the case of one unaffected heterozygous and one affected homozygous parent, the probability of disease increases to 50% for each child. In this instance, the pedigree analysis mimics an autosomal dominant mode of inheritance (*pseudodominance*). In contrast to autosomal dominant disorders, new mutations in recessive alleles are rarely manifest because they usually result in an asymptomatic carrier state.

X-LINKED DISORDERS Males have only one X chromosome; consequently, a daughter always inherits her father's X chromosome in addition to one of her mother's two X chromosomes. A son inherits the Y chromosome from his father and one maternal X chromosome. Thus, the characteristic features of X-linked inheritance are (1) the absence of father-to-son transmission and (2) the fact that all daughters of an affected male are obligate carriers of the mutant allele **(Fig. 479-13*C*)**. The risk of developing disease due to a mutant X-chromosomal gene differs in the two sexes. Because males have only one X chromosome, they are hemizygous for the mutant allele; thus, they are more likely to develop the mutant phenotype, regardless of whether the mutation is dominant or recessive. A female may be either heterozygous or homozygous for the mutant allele, which may be dominant or recessive. The terms *X-linked dominant* and *X-linked recessive* are therefore only applicable to expression of the mutant phenotype in women. In addition, the expression of X-chromosomal genes is influenced by X chromosome inactivation.

Y-LINKED DISORDERS The Y chromosome has a relatively small number of genes. One such gene, the sex-region determining Y factor (*SRY*), which encodes the testis-determining factor (*TDF*), is crucial for normal male development. Normally, there is infrequent exchange of sequences on the Y chromosome with the X chromosome. The *SRY* region is adjacent to the pseudoautosomal region, a chromosomal segment on the X and Y chromosomes with a high degree of homology. A crossing-over event occasionally involves the *SRY* region with the distal tip of the X chromosome during meiosis in the male. Translocations can result in XY females with the Y chromosome lacking the *SRY* gene or XX males harboring the *SRY* gene on one of the X chromosomes **(Chap. 402)**. Point mutations in the *SRY* gene may also result in individuals with an XY genotype and an incomplete female phenotype. Most of these mutations occur de novo. Men with oligospermia/azoospermia frequently have microdeletions on the long arm of the Y chromosome that involve one or more of the azoospermia factor (*AZF*) genes.

Exceptions to Simple Mendelian Inheritance Patterns •

MITOCHONDRIAL DISORDERS Mendelian inheritance refers to the transmission of genes encoded by DNA contained in the nuclear chromosomes. In addition, each mitochondrion contains several copies of a small circular chromosome **(Chap. 481)**. The mitochondrial DNA (mtDNA) is ~16.5 kb and encodes transfer and ribosomal RNAs and 13 core proteins that are components of the respiratory chain involved in oxidative phosphorylation and ATP generation. The mitochondrial genome does not recombine and is inherited through the maternal line because sperm does not contribute significant cytoplasmic components to the zygote. A noncoding region of the mitochondrial chromosome, referred to as D-loop, is highly polymorphic. This property, together with the absence of mtDNA recombination, makes it a valuable tool for studies tracing human migration and evolution, and it is also used for specific forensic applications.

Inherited mitochondrial disorders are transmitted in a matrilineal fashion; all children from an affected mother will inherit the disease, but it will not be transmitted from an affected father to his children **(Fig. 479-13*D*)**. Alterations in the mtDNA that involves enzymes required for oxidative phosphorylation lead to reduction of ATP supply, generation of free radicals, and induction of apoptosis. Several syndromic disorders arising from mutations in the mitochondrial genome are known in humans, and they affect both protein-coding and tRNA genes. The broad clinical spectrum often involves (cardio) myopathies and encephalopathies because of the high dependence of these tissues on oxidative phosphorylation. The age of onset and the clinical course are highly variable because of the unusual mechanisms of mtDNA transmission, which replicates independently from nuclear DNA. During cell replication, the proportion of wild-type and mutant mitochondria can drift among different cells and tissues. The resulting heterogeneity in the proportion of mitochondria with and without a mutation is referred to as *heteroplasmia* and underlies the phenotypic variability that is characteristic of mitochondrial diseases.

Acquired somatic mutations in mitochondria are thought to be involved in several age-dependent degenerative disorders affecting predominantly muscle and the peripheral and central nervous system (e.g., Alzheimer's and Parkinson's diseases). Establishing that an mtDNA alteration is causal for a clinical phenotype is challenging because of the high degree of polymorphism in mtDNA and the phenotypic variability characteristic of these disorders. Certain pharmacologic treatments may have an impact on mitochondria and/or their function. For example, treatment with the antiretroviral compound azidothymidine (AZT) causes an acquired mitochondrial myopathy through depletion of muscular mtDNA.

MOSAICISM Mosaicism refers to the presence of two or more genetically distinct cell lines in the tissues of an individual. It results from a mutation that occurs during embryonic, fetal, or extrauterine development. The developmental stage at which the mutation arises will determine whether germ cells and/or somatic cells are involved. Chromosomal mosaicism results from nondisjunction at an early embryonic mitotic division, leading to the persistence of more than one cell line, as exemplified by some patients with Turner's syndrome **(Chap. 402)**. Somatic mosaicism is characterized by a patchy distribution of genetically altered somatic cells. The McCune-Albright syndrome, for example, is caused by activating mutations in the stimulatory G protein α ($G_s\alpha$) that occur postzygotically in early development **(Chap. 422)**. The clinical phenotype varies depending on the tissue distribution of the mutation; manifestations include ovarian cysts that secrete sex steroids and cause precocious puberty, polyostotic fibrous dysplasia, café-au-lait skin pigmentation, GH-secreting pituitary adenomas, and hypersecreting autonomous thyroid nodules.

X-INACTIVATION, IMPRINTING, AND UNIPARENTAL DISOMY According to traditional Mendelian principles, the parental origin of a mutant gene is irrelevant for the expression of the phenotype. There are, however, important exceptions to this rule. *X-inactivation* prevents the expression of most genes on one of the two X chromosomes in every cell of a female. Gene inactivation through genomic imprinting occurs on selected chromosomal regions of autosomes and leads to inheritable preferential expression of one of the parental alleles. It is of pathophysiologic importance in disorders where the transmission of disease is dependent on the sex of the transmitting parent and, thus, plays an important role in the expression of certain genetic disorders. Two classic examples are the Prader-Willi syndrome and Angelman's syndrome. Prader-Willi syndrome is characterized by diminished fetal activity, obesity, hypotonia, intellectual disability, short stature, and hypogonadotropic hypogonadism. Deletions of the paternal copy of the Prader-Willi locus located on the short arm of chromosome 15 result in a contiguous gene syndrome involving missing paternal copies of the *necdin* and *SNRPN* genes, among others. In contrast, patients with Angelman's syndrome, characterized by intellectual disability, seizures, ataxia, and hypotonia, have deletions involving the maternal copy of this region on chromosome 15. These two syndromes may also result from *uniparental disomy*. In this case, the syndromes are not caused by deletions on chromosome 15 but by the inheritance of either two maternal chromosomes (Prader-Willi syndrome) or two paternal chromosomes (Angelman's syndrome). Lastly, the two distinct phenotypes can also be caused by an imprinting defect that impairs the resetting of the imprint during zygote development (defect in the father leads to Prader-Willi syndrome; defect in the mother leads to Angelman's syndrome).

Imprinting and the related phenomenon of allelic exclusion may be more common than currently documented because it is difficult to examine levels of mRNA expression from the maternal and paternal alleles in specific tissues or in individual cells. Genomic imprinting, or uniparental disomy, is involved in the pathogenesis of several other disorders and malignancies. For example, hydatidiform moles contain a normal number of diploid chromosomes, but they are exclusively of paternal origin. The opposite situation occurs in ovarian teratomata, with 46 chromosomes of maternal origin. Expression of the imprinted gene for insulin-like growth factor 2 (IGF-2) is involved in the pathogenesis of the cancer-predisposing Beckwith-Wiedemann syndrome (BWS). These children show somatic overgrowth with organomegalies and hemihypertrophy, and they have an increased risk of embryonal malignancies such as Wilms' tumor. Normally, only the paternally derived copy of the *IGF2* gene is active, and the maternal copy is inactive. BWS can be caused by several genetic defects that result in overactivity of IGF-2, or a missing active copy of *CDKN1C*, that result in inhibition of cell proliferation. They include paternal uniparental disomy (UPD) of chromosome 11, aberrant methylation of this region, maternal chromosomal rearrangements, or deletions within the locus.

Alterations of the epigenome through gain and loss of DNA methylation and altered histone modifications play an important role in the pathogenesis of malignancies.

SOMATIC MUTATIONS Cancer can be considered a genetic disease at the cellular level **(Chap. 76)**. Cancers are monoclonal in origin, indicating that they have arisen from a single precursor cell with one or several mutations in genes controlling growth (proliferation or apoptosis) and/or differentiation. These acquired somatic mutations are restricted to the tumor and its metastases and are not found in the surrounding normal tissue. The molecular alterations include dominant gain-of-function mutations in oncogenes, recessive loss-of-function mutations in tumor-suppressor genes and DNA repair genes, gene amplification, and chromosome rearrangements. Chromothripsis refers to a mutational process including multiple clustered chromosomal rearrangements in close vicinity, for example, after injury by ionizing radiation. Rarely, a single mutation in certain genes may be sufficient to transform a normal cell into a malignant cell. In most cancers, however, the development of a malignant phenotype requires several genetic alterations for the gradual progression from a normal cell to a cancerous cell, a phenomenon termed *multistep carcinogenesis*. Genome-wide analyses of cancers using deep sequencing often reveal somatic rearrangements resulting in fusion genes and mutations in multiple genes (Table 479-1 and **Fig. 479-14)**. Comprehensive sequence analyses, now also possible through single-cell sequencing (SCS), provide insight into the evolution and genetic heterogeneity within malignancies; these include intratumoral heterogeneity among the cells of the primary tumor, intermetastatic and intrametastatic heterogeneity, and interpatient differences. These analyses further support the notion of cancer as an ongoing process of clonal evolution, in which successive rounds of clonal selection within the primary tumor and metastatic lesions result in diverse genetic and epigenetic alterations that require targeted (personalized) therapies (precision medicine). The heterogeneity of mutations within a tumor can also lead to resistance to targeted therapies because cells with mutations that are resistant to the therapy, even if they are a minor part of the tumor population, will be selected as the more sensitive cells are eliminated.

Telomeres, repeats of conserved sequences, protect the ends the chromosomes from DNA damage or fusion with neighboring chromosomes. Telomere length shortens with age. Most human tumors express telomerase, an enzyme formed of a protein and an RNA component, which adds telomere repeats at the ends of chromosomes during replication. This mechanism impedes shortening of the telomeres and is associated with enhanced replicative capacity in cancer cells. Telomerase inhibitors provide a strategy for treating advanced human cancers.

In many cancer syndromes, there is frequently an inherited *predisposition* to tumor formation. In these instances, a germline mutation is inherited in an autosomal dominant fashion inactivating one allele of an autosomal tumor-suppressor gene. If the second allele is inactivated by a somatic mutation or by epigenetic silencing in a given cell, this will lead to neoplastic growth (Knudson two-hit model). Thus, the defective allele in the germline is transmitted in a dominant mode, although tumorigenesis results from a biallelic loss of the tumor-suppressor gene in an affected tissue. The classic example to illustrate this phenomenon is retinoblastoma, which can occur as a sporadic or hereditary tumor. In sporadic retinoblastoma, both copies of the retinoblastoma (*RB*) gene are inactivated through two somatic events. In hereditary retinoblastoma, one mutated or deleted *RB* allele is inherited in an autosomal dominant manner and the second allele is inactivated by a subsequent somatic mutation. This two-hit model applies to other inherited cancer syndromes such as MEN 1 **(Chap. 400)** and neurofibromatosis types 1 and 2 **(Chap. 95)**. In contrast, in the autosomal dominant MEN 2 syndrome, the predisposition for tumor formation in various organs is caused by a gain-of-function mutation in a single allele of the *RET* gene **(Chap. 400)**.

NUCLEOTIDE REPEAT EXPANSION DISORDERS Several diseases are associated with an increase in the number of nucleotide repeats above a certain threshold **(Table 479-5)**. The repeats are sometimes located

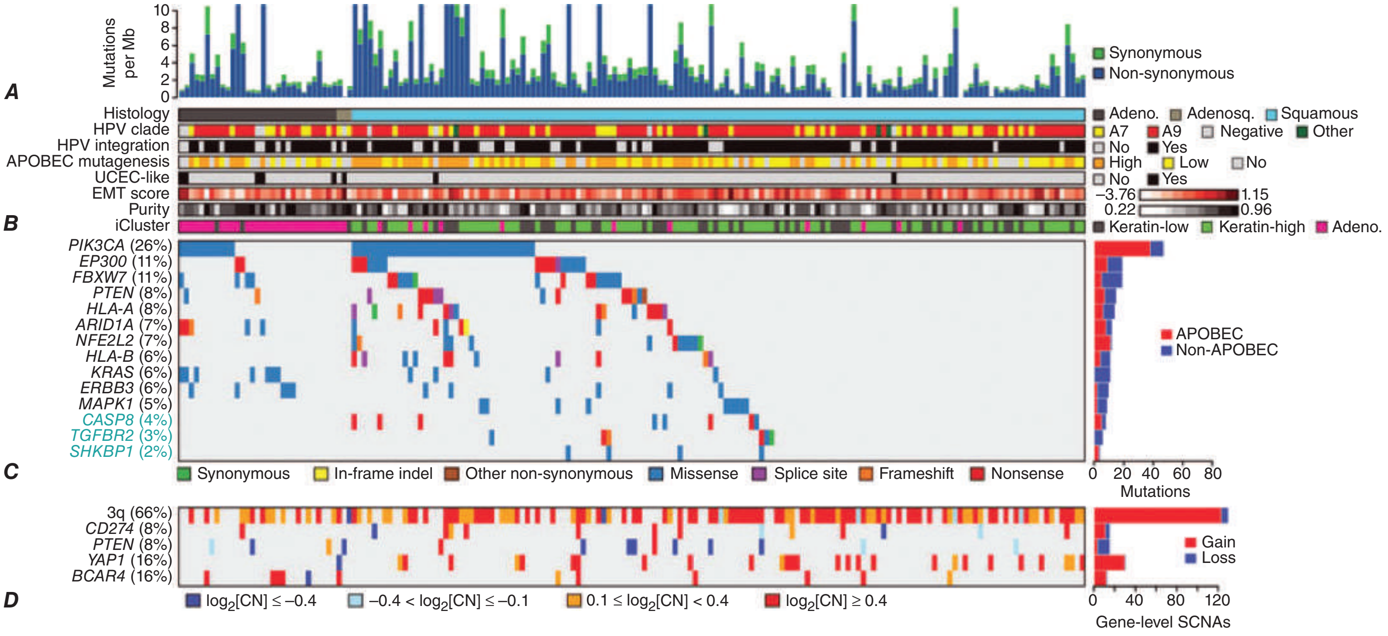

FIGURE 479-14 Somatic alterations in cervical cancer. ***A.*** Cervical carcinoma samples ordered by histology and mutation frequency; ***B.*** clinical and molecular platform features; ***C.*** significantly mutated genes (SMGs); and ***D.*** select somatic copy number alterations. SMGs are ordered by the overall mutation frequency and color-coded by mutation type. Adeno, adenocarcinomas; Adenosq, adenosquamous cancers; CN, copy number; SCNAs, somatic copy number alterations; Squamous, squamous cell carcinomas. *(Reproduced from The Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. Nature 543:378–384, 2017.)*

TABLE 479-5 Selected Trinucleotide Repeat Disorders

DISEASE	LOCUS	REPEAT	TRIPLET LENGTH (NORMAL/DISEASE)	INHERITANCE	GENE PRODUCT
X-chromosomal spinobulbar muscular atrophy (SBMA)	Xq12	CAG	11–34/40–62	XR	Androgen receptor
Fragile X syndrome (FRAXA)	Xq27.3	CGG	6–50/200–300	XR	FMR-1 protein
Fragile X syndrome (FRAXE)	Xq28	GCC	6–25/>200	XR	FMR-2 protein
Dystrophia myotonica (DM)	19q13.32	CTG	5–30/200–1000	AD, variable penetrance	Myotonin protein kinase
Huntington's disease (HD)	4p16.3	CAG	6–34/37–180	AD	Huntingtin
Spinocerebellar ataxia type 1 (SCA1)	6p22.3	CAG	6–39/40–88	AD	Ataxin 1
Spinocerebellar ataxia type 2 (SCA2)	12q24.12	CAG	15–31/34–400	AD	Ataxin 2
Spinocerebellar ataxia type 3 (SCA3); Machado-Joseph disease (MD)	14q32.12	CAG	13–36/55–86	AD	Ataxin 3
Spinocerebellar ataxia type 6 (SCA6, CACNAIA)	19p13	CAG	4–16/20–33	AD	Alpha 1A voltage-dependent L-type calcium channel
Spinocerebellar ataxia type 7 (SCA7)	3p14.1	CAG	4–19/37 to >300	AD	Ataxin 7
Spinocerebellar ataxia type 12 (SCA12)	5q32	CAG	6–26/66–78	AD	Protein phosphatase 2A
Dentatorubral pallidoluysian atrophy (DRPLA)	12p13.31	CAG	7–23/49–75	AD	Atrophin 1
Friedreich's ataxia (FRDA1)	9q21.11	GAA	7–22/200–900	AR	Frataxin

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; XR, X-linked recessive.

within the coding region of the genes, as in Huntington's disease or the X-linked form of spinal and bulbar muscular atrophy (SBMA; Kennedy's syndrome). In other instances, the repeats probably alter gene regulatory sequences. If an expansion is present, the DNA fragment is unstable and tends to expand further during cell division. The length of the nucleotide repeat often correlates with the severity of the disease. When repeat length increases from one generation to the next, disease manifestations may worsen or be observed at an earlier age; this phenomenon is referred to as *anticipation*. In Huntington's disease, for example, there is a correlation between age of onset and length of the triplet codon expansion **(Chap. 435)**. Anticipation has also been documented in other diseases caused by dynamic mutations in trinucleotide repeats (Table 479-5). The repeat number may also vary in a tissue-specific manner. In myotonic dystrophy, the CTG repeat may be 10-fold greater in muscle tissue than in lymphocytes **(Chap. 460)**.

Complex Genetic Disorders The expression of many common diseases such as cardiovascular disease, hypertension, diabetes, asthma, psychiatric disorders, and certain cancers is determined by a combination of genetic background, environmental factors, and lifestyle. A trait is called *polygenic* if multiple genes contribute to the phenotype or *multifactorial* if multiple genes are assumed to interact with environmental factors. Genetic models for these complex traits need to account for genetic heterogeneity and interactions with other genes and the environment. Complex genetic traits may be influenced by modifier genes that are not linked to the main gene involved in the pathogenesis of the trait. This type of gene-gene interaction, or *epistasis*, where the expression of a gene is altered by the expression of one or several independently inherited genes, plays an important role in polygenic traits. In aggregate, variants in multiple genes need to be present simultaneously to result in a pathologic phenotype.

Type 2 diabetes mellitus provides a paradigmatic example of a multifactorial disorder, because genetic, nutritional, and lifestyle factors are intimately interrelated in disease pathogenesis **(Table 479-6)** **(Chap. 415)**. The identification of genetic variations and environmental factors that either predispose to or protect against disease is essential for predicting disease risk, designing preventive strategies, and developing novel therapeutic approaches. The study of rare monogenic diseases may provide insight into some of the genetic and molecular mechanisms important in the pathogenesis of complex diseases. For example, the identification of the genes causing monogenic forms of permanent neonatal diabetes mellitus or maturity-onset diabetes defined them as *candidate genes* in the pathogenesis of diabetes mellitus type 2 (Tables 479-2 and 479-6) **(Fig. 479-15)**. Genome scans have identified numerous genes and loci that may be associated with susceptibility to development of diabetes mellitus in certain populations **(Fig. 479-16)**. Efforts to identify susceptibility genes require very large sample sizes, and positive results may depend on ethnicity, ascertainment criteria, and statistical analysis. Association studies analyzing the potential influence of (biologically functional) SNPs and SNP haplotypes on a particular phenotype have revealed new insights into the genes involved in the pathogenesis of these common disorders. Large variants ([micro]deletions, duplications, and inversions) present in the human population also contribute to the pathogenesis of complex disorders, but their contributions remain poorly understood.

Linkage and Association Studies There are two primary strategies for mapping genes that cause or increase susceptibility to human disease: (1) classic linkage can be performed based on a known genetic model or, when the model is unknown, by studying pairs of affected relatives; or (2) disease genes can be mapped using allelic association studies **(Table 479-7)**.

GENETIC LINKAGE *Genetic linkage* refers to the fact that genes are physically connected, or linked, to one another along the chromosomes. Two fundamental principles are essential for understanding the concept of linkage: (1) when two genes are close together on a chromosome, they are usually transmitted together, unless a recombination event separates them (Figs. 479-6); and (2) the odds of a crossover, or recombination event, between two linked genes is proportional to the distance that separates them. Thus, genes that are farther apart are more likely to undergo a recombination event than genes that are very close together. The detection of chromosomal loci that segregate with a disease by linkage can be used to identify the gene responsible for the disease and to predict the odds of disease gene transmission in genetic counseling.

Polymorphic variants are essential for linkage studies because they provide a means to distinguish the maternal and paternal chromosomes in an individual. On average, 1 out of every 1000 bp varies from one person to the next. Although this degree of variation seems low (99.9% identical), it means that >3 million sequence differences exist between any two unrelated individuals and the probability that the sequence at such loci will differ on the two homologous chromosomes is high (often >70–90%). These sequence variations include variable number of tandem repeats (VNTRs), short tandem repeats (STRs), and SNPs. Most STRs, also called *polymorphic microsatellite markers*, consist of di-, tri-, or tetranucleotide repeats that can be characterized readily using the polymerase chain reaction (PCR). Characterization of SNPs, using DNA chips or beads, permits comprehensive analyses of genetic variation, linkage, and association studies. Although these sequence variations often have no apparent functional consequences, they provide much of the basis for variation in genetic traits.

TABLE 479-6 Examples of Genes and Loci Involved in Mono- and Polygenic Forms of Diabetes

DISORDER	GENES OR SUSCEPTIBILITY LOCUS	CHROMOSOMAL LOCATION	OTHER FACTORS
Monogenic permanent neonatal diabetes mellitus	*KCNJ11* (inwardly rectifying potassium channel Kir6.2)	11p15.1	AD
	GCK (glucokinase)	7p13	AR
	INS (insulin)	11p15.5	AR, hyperproinsulinemia
	ABCC8 (ATP-binding cassette, subfamily c, member 8; sulfonylurea receptor)	11p15.1	AD or AR
	GLIS3 (GLIS family zinc finger protein 3)	9p24.2	AR, diabetes, congenital hypothyroidism
Maturity-onset diabetes of the young (MODY): monogenic forms of diabetes mellitus			
MODY 1	*HNF4α* (hepatocyte nuclear factor 4α)	20q13.12	AD inheritance
MODY 2	*GCK* (glucokinase)	7p13	
MODY 3	*HNF1α* (hepatocyte nuclear factor 1α)	12q24.31	
MODY 4	*IPF1* (insulin receptor substrate)	13q12.2	
MODY 5 (renal cysts, diabetes)	*HNF1β* (hepatocyte nuclear factor 1β)	17q12	
MODY 6	*NeuroD1* (neurogenic differentiation factor 1)	2q31.3	
MODY 7	*KLF1* (Kruppel-like factor 1)	19p13.13	
MODY 8	*CEL* (carboxyl ester lipase)	9q34.13	
MODY 9	*PAX4* (paired box transcription factor 4)	7q32.1	
MODY 10	*INS* (insulin)	11p15.5	
MODY 11	*BLK* (B-lymphocyte-specific tyrosine kinase)	8p23.1	
MODY 12	*ABCC8* (ATP-binding cassette, subfamily c, member 8; sulfonylurea receptor)	11p15.1	
MODY 13	*KCNJ11* (inwardly rectifying potassium channelKir6.2)	11p15.1	
Diabetes mellitus type 2; loci and genes linked and/or associated with susceptibility for diabetes mellitus type 2	Genes and loci identified by linkage/association studies *PPARG, KCNJ11/ABCC8, TCF7L2, HNF1B, WFS1, SLC30A8, FTO, HHEX, IGF2BP2, CDKN2A/B, CDKAL1, TSPAN8, ADAMTs9, CDC123/CAMK1D, JAZF1, NOTCH2, THADA, KCNQ1, DUSP8, MTNR1B, IRS1, SPRY2, SRR, ZFAND6, GCK, KLF14, TP53INP1, PROX1, PRC1, BCL11A, ZBED3, RBMS1, HNF1A, DGKB/TMEM195, CCND2, C2CD4A/C2CD4B, PTPRD, ARAP1/CENTD2, HMGA2, TLE4/CHCHD9, ADCY5, UBE2E2, DUSP9, GCKR, COBLL1/GRB14, HMG20A, VPS26A, ST6GAL1, AP3S2, HNF4A, BCL2, LAMA1, GIPR, MC4R, TLE1, KCNK16, ANK1, KLHDC5, ZMIZ1, PSMD6, FITM2/R3HDML/HNF4A, CILP2, ANKRD55, GLIS3, PEPD, GCC1/PAX4, ZFAND3, MAEA, BCAR1, RBM43/RND3, MACF1, RASGRP1, GRK5, TMEM163, SGCG, LPP, FAF1, TMEM154, MPHOSPH9, ARL15, POU5F1/TCF19, SSR1/RREB1, HLA-B, INS-IGF2, GPSM1, LEP, SLC16A13, PAM/PPIP5K2, SLC16A11, CCDC63, C12orf51, CCND2, HNF1A, TBC1D4, CCDC85A, INAFM2, ASB3, FAM60A, ATP8B2, MIR4686, MTMR3, DMRTA1, SLC35D3, GLP2R, GIP, MAP3K11, PLEKHA1, HSD17B12, NRXN3, CMIP, ZZEF1, MNX1, ABO, ACSL1, HLA-DQA1*		Heavily influenced by diet, energy expenditure, obesity

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; MODY, maturity-onset diabetes of the young.

In order to identify a chromosomal locus that segregates with a disease, it is necessary to characterize polymorphic DNA markers from affected and unaffected individuals of one or several pedigrees. One can then assess whether certain marker alleles cosegregate with the disease. Markers that are closest to the disease gene are less likely to undergo recombination events and therefore receive a higher linkage score. Linkage is expressed as a lod (logarithm of odds) score—the ratio of the probability that the disease and marker loci are linked rather than unlinked. Lod scores of +3 (1000:1) are generally accepted as supporting linkage, whereas a score of −2 is consistent with the absence of linkage.

ALLELIC ASSOCIATION, LINKAGE DISEQUILIBRIUM, AND HAPLOTYPES *Allelic association* refers to a situation in which the frequency of an allele is significantly increased or decreased in individuals affected by a particular disease in comparison to controls. Linkage and association differ in several aspects. Genetic linkage is demonstrable in families or sibships. Association studies, on the other hand, compare a population of affected individuals with a control population. Association studies can be performed as case-control studies that include unrelated affected individuals and matched controls or as family-based studies that compare the frequencies of alleles transmitted or not transmitted to affected children.

Allelic association studies are particularly useful for identifying susceptibility genes in complex diseases. When alleles at two loci occur more frequently *in combination* than would be predicted (based on known allele frequencies and recombination fractions), they are said to be in *linkage disequilibrium*. Evidence for linkage disequilibrium can be helpful in mapping disease genes because it suggests that the two loci are tightly linked.

Detecting the genetic factors contributing to the pathogenesis of common complex disorders is challenging. In many instances, these are low-penetrance alleles (e.g., variations that individually have a subtle effect on disease development, and they can only be identified by unbiased GWAS) (Catalog of Published Genome-Wide Association Studies; Table 479-1) (Fig. 479-16). Most variants occur in noncoding or regulatory sequences but do not alter protein structure. The analysis of complex disorders is further complicated by ethnic differences in disease prevalence, differences in allele frequencies in known susceptibility genes among different populations, locus and allelic heterogeneity, gene-gene and gene-environment interactions, and the possibility of phenocopies. Catalogues of human variation and genotype data

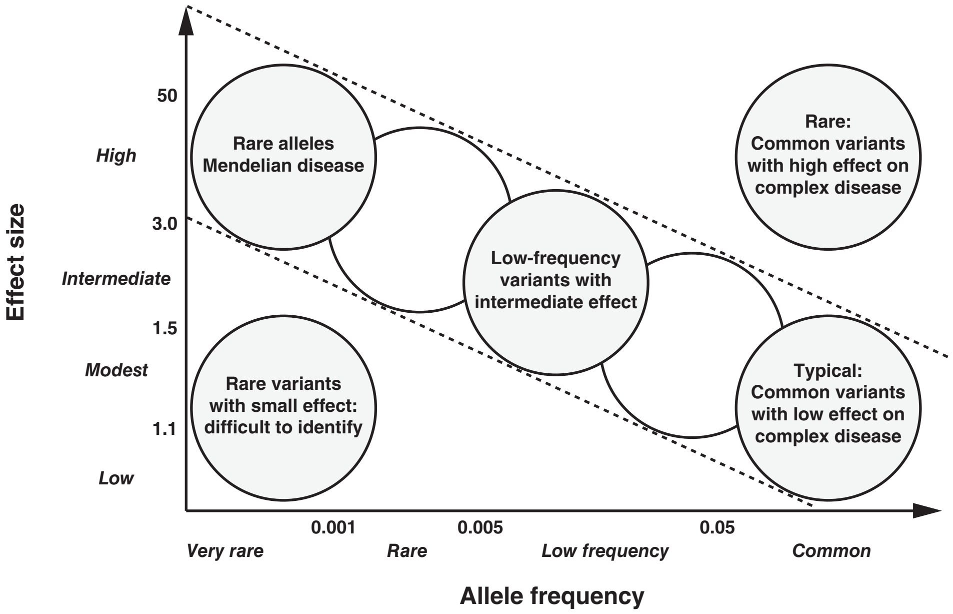
Effect size
50
High
3.0
Intermediate
1.5
Modest
1.1
Low
Rare alleles Mendelian disease
Low-frequency variants with intermediate effect
Rare: Common variants with high effect on complex disease
Rare variants with small effect: difficult to identify
Typical: Common variants with low effect on complex disease
0.001
0.005
0.05
Very rare
Rare
Low frequency
Common
Allele frequency

FIGURE 479-15 Relationship between allele frequency and effect size in monogenic and polygenic disorders. In classic Mendelian disorders, the allele frequency is typically low but has a high impact (single-gene disorder). This contrasts with polygenic disorders that require the combination of multiple low-impact alleles that are frequently quite common in the general population.

(HapMap, International Genome Sample Resource) have greatly facilitated GWAS for the characterization of complex disorders. Adjacent SNPs are inherited together as blocks, and these blocks can be identified by genotyping selected marker SNPs, so-called *Tag SNPs*, thereby reducing cost and workload (Fig. 479-4). The availability of this information permits the characterization of a limited number of SNPs to identify the set of haplotypes present in an individual (e.g., in cases and controls). This, in turn, permits performing GWAS by searching for associations of certain haplotypes with a disease phenotype of interest, an essential step for unraveling the genetic factors contributing to complex disorders.

POPULATION GENETICS In population genetics, the focus changes from alterations in an individual's genome to the distribution pattern of different genotypes in the population. In a case where there are only two alleles, A and a, the frequency of the genotypes will be $p^2 + 2pq + q^2 = 1$, with p^2 corresponding to the frequency of AA, $2pq$ to the frequency of Aa, and q^2 to aa. When the frequency of an allele is known, the frequency of the genotype can be calculated. Alternatively, one can determine an allele frequency if the genotype frequency has been determined.

Allele frequencies vary among ethnic groups and geographic regions. For example, heterozygous mutations in the *CFTR* gene are relatively common in populations of European origin but are rare in the African population. Allele frequencies may vary because certain allelic variants confer a selective advantage. For example, heterozygotes for the sickle cell mutation, which is particularly common in West Africa, are more resistant to malaria infection because the erythrocytes of heterozygotes provide a less favorable environment for *Plasmodium* parasites. Although homozygosity for the sickle cell mutation is associated with severe anemia and sickle ʻcrises, heterozygotes have a higher probability of survival because of the reduced morbidity and mortality from malaria; this phenomenon has led to an increased frequency of the mutant allele. Recessive conditions are more prevalent in geographically isolated populations because of the more restricted gene pool.

APPROACH TO THE PATIENT

Inherited Disorders

For the practicing clinician, the family history remains an essential step in recognizing the possibility of a hereditary predisposition to disease. When taking the history, it is useful to draw a detailed pedigree of the first-degree relatives (e.g., parents, siblings, and children), because they share 50% of genes with the patient. Standard symbols for pedigrees are depicted in Fig. 479-11. The family history should include information about ethnic background, age, health status, and deaths, including infants. Next, the physician should explore whether there is a family history of the same or related illnesses to the current problem. An inquiry focused on commonly occurring disorders such as cancers, heart disease, and diabetes mellitus should follow. Because of the possibility of age-dependent expressivity and penetrance, the family history will need intermittent updating. If the findings suggest a genetic disorder, the clinician should assess whether some of the patient's relatives may be at risk of carrying or transmitting the disease. In this circumstance, it is useful to confirm and extend the pedigree based on input from several family members. Emerging artificial intelligence tools analyzing facial features can aid the clinician in diagnosing patients with genetic conditions. In aggregate, this information may form the basis for genetic counseling, carrier detection, early intervention, and disease prevention in relatives of the index patient **(Chap. 480)**.

In instances where a diagnosis at the molecular level may be relevant, it is important to identify an appropriate laboratory that can perform the appropriate test. Genetic testing is available for a large number of monogenic disorders through commercial laboratories. For uncommon disorders, the test may only be performed in a specialized research laboratory. Approved laboratories offering testing for inherited disorders can be identified in continuously updated online resources (e.g., Genetic Testing Registry; Table 479-1). If genetic testing is considered, the patient and the family should be counseled about the potential implications of positive results, including psychological distress and the possibility of discrimination. The patient or caretakers should be informed about the meaning of a negative result, technical limitations, and the possibility of false-negative and inconclusive results. For these reasons, genetic testing should only be performed after obtaining *informed consent*. Published ethical guidelines address the specific aspects that should be considered when testing children and adolescents.

IDENTIFYING THE DISEASE-CAUSING GENE

Precision medicine aims to enhance the quality of medical care through the use of genotypic analysis (DNA testing) to identify genetic predisposition to disease, to select more specific

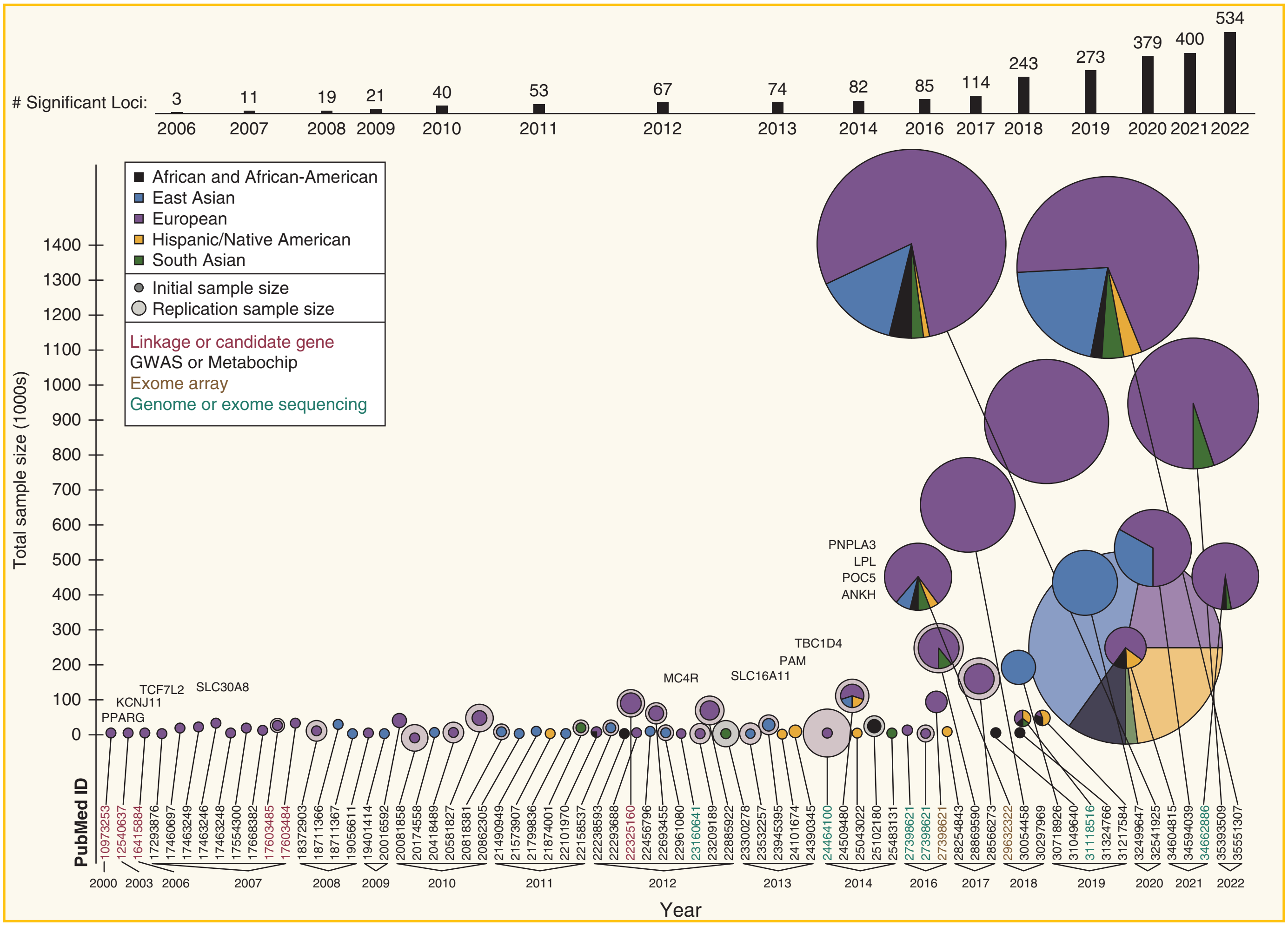

FIGURE 479-16 Genome-wide association studies (GWAS) across ancestries and discovery of loci over time. The pie charts represent type 2 diabetes GWAS, as well as candidate gene or sequencing studies. The *x* axis shows the year of publication, and the *y* axis shows discovery sample size. The inner circles are scaled in proportion to discovery sample size, and the outer circles are scaled in proportion to total (discovery + replication) sample size. Significant loci are defined as a *p* value of 5×10^{-8}. At the end of 2022, 534 type 2 diabetes distinct intervals (520 autosomal, 14 X chromosomal) were defined. *(Reproduced with permission from BF Voight.)*

pharmacotherapy, and to design individualized medical care based on genotype. Genotype can be deduced by analysis of protein (e.g., hemoglobin, apoprotein E), mRNA, or DNA. Many (pathogenic) variants can be readily identified by DNA analyses; technical advances in RNA sequencing now add increasing depth to genetic and genomic investigations (e.g., for the detection of gene fusions or aberrant gene expression patterns).

DNA testing is performed by mutational analysis or linkage studies in individuals at risk for a genetic disorder known to be present in a family. Mass screening programs require tests of high sensitivity and specificity to be cost-effective. The benefits and risks of screening newborns with genomic sequencing, and the potential impact on surveillance, preventative health care, and personalized treatment options are topics of current research (BabySeq Project). Prerequisites for the success of genetic screening programs include the following: that the disorder is potentially serious; that it can be influenced at a presymptomatic stage by changes in behavior, diet, and/or pharmaceutical manipulations; and that the screening does not result in any harm or discrimination. Screening in Jewish populations for the autosomal recessive neurodegenerative storage disease Tay-Sachs has reduced the number of affected individuals. In contrast, screening for sickle cell trait/disease in African Americans has led to unanticipated problems of discrimination by health insurers and employers. Mass screening programs harbor additional potential problems. For example, screening for the most common genetic alteration in cystic fibrosis, the ΔF508 mutation with a frequency of ~70% in northern Europe, is feasible and seems to be effective. One has to keep in mind, however, that there is pronounced allelic heterogeneity and that the *CFTR* gene can be affected by >2000 other mutations. While the search for less common mutations has been challenging in the past, next-generation genome sequencing now permits comprehensive and cost-effective mutational analyses. However, the bioinformatic analysis and the classification of the detected variants as pathogenic or benign alterations is still challenging. Occupational screening programs aim to detect individuals with increased risk for certain professional activities (e.g., α_1 antitrypsin deficiency and smoke or dust exposure). Integrating genomic data into electronic medical records is evolving and can provide significant decision support at the point of care, for example, by providing the clinician with genomic data and decision algorithms for the prescription of drugs that are subject to pharmacogenetic influences.

Mutational Analyses DNA sequence analysis is widely used as a diagnostic tool and has significantly enhanced diagnostic accuracy. It is used for determining carrier status and for prenatal testing in monogenic disorders. Numerous techniques, discussed in previous versions of this chapter, are available for the detection of mutations. Analyses of large alterations in the genome are possible using classic methods such as karyotype analysis, cytogenetics, fluorescent in situ hybridization (FISH), and array- or bead-based techniques that search for multiple single exon deletions or duplications.

TABLE 479-7 Genetic Approaches for Identifying Disease Genes

METHOD	INDICATIONS AND ADVANTAGES	LIMITATIONS
Linkage Studies		
Classical linkage analysis (parametric methods)	Analysis of monogenic traits Suitable for genome scan Control population not required Useful for multifactorial disorders in isolated populations	Difficult to collect large informative pedigrees Difficult to obtain sufficient statistical power for complex traits
Allele-sharing methods (nonparametric methods)	Suitable for identification of susceptibility genes in polygenic and multifactorial disorders	Difficult to collect sufficient number of subjects
Affected sib and relative pair analyses	Suitable for genome scan	Difficult to obtain sufficient statistical power for complex traits
Sib pair analysis	Control population not required if allele frequencies are known Statistical power can be increased by including parents and relatives	Reduced power compared to classical linkage, but not sensitive to specification of genetic mode
Association Studies		
Case-control studies	Suitable for identification of susceptibility genes in polygenic and multifactorial disorders	Requires large sample size and matched control population
Linkage disequilibrium	Suitable for testing specific allelic variants of known candidate loci	False-positive results in the absence of suitable control population
Transmission disequilibrium test (TDT)	Facilitated by comprehensive catalogs of genotypes and variants	Candidate gene approach does not permit detection of novel genes and pathways
Whole-genome association studies	Does not necessarily need relatives	Susceptibility genes can vary among different populations
Next-Generations Sequencing Technologies		
Whole exome or genome sequencing	Unbiased approach, analysis can be performed without reference sequences from parents or siblings	Requires appropriate bioinformatics, may have low sensitivity if CNV analysis is not included, detects numerous VUS, can lead to the detection of unrelated deleterious alleles
Targeted sequencing of gene panels	Captures multiple candidate genes and loci with hybridization techniques followed by deep sequencing	Permits analyses of multiple candidate genes in parallel; facilitates molecular characterization of disorders with locus heterogeneity

Abbreviations: CNV, copy number variation; VUS, variants of unknown significance.

The analysis of more discrete sequence alterations often rely on the use of PCR, which allows rapid gene amplification and analysis. Moreover, PCR makes it possible to perform genetic testing and mutational analysis with small amounts of DNA extracted from leukocytes or even from single cells, buccal cells, or hair roots. DNA sequencing can be performed directly on PCR products. The advent of comprehensive sequencing technologies analyzing the whole exome or genome, of selected chromosomes, or of numerous candidate genes in a single run with NGS platforms is now fundamentally transforming the characterization of patients with rare disorders and advanced malignancies. These techniques have the advantage of an unbiased comprehensive approach, and they are increasingly cost-effective. Analysis of cell-free DNA (cfDNA; also referred to as "liquid biopsy") present in body fluids is playing a growing role for minimally invasive diagnostics and disease monitoring. Genomic tests are also widely used for the detection of pathogens and for the identification of viral or bacterial sequence variations.

The integration of genomic tests into clinical medicine is associated with a number of ongoing challenges related to variable sensitivities of the tests, bioinformatics analyses, storage and sharing of data, and the difficulty of interpreting all genetic variants identified with comprehensive testing. The discovery of incidental (or secondary) findings that are unrelated to the indication for the sequencing analysis, but indicators of other disorders of potential relevance for patient care can pose a difficult ethical dilemma. It can lead to the detection of undiagnosed medically actionable genetic conditions but can also reveal deleterious mutations that cannot be influenced, as numerous sequence variants are of unknown significance.

A general algorithm for the approach to mutational analysis in patients with a suspected genetic disorder and (advanced) malignancies is outlined in **Fig. 479-17**. The importance of a detailed characterization of the clinical phenotype cannot be overemphasized. This is the step where one should also consider the possibility of genetic heterogeneity and phenocopies. If obvious candidate genes are suggested by the phenotype, they can be analyzed directly. After identification of a mutation, it is essential to demonstrate that it segregates with the phenotype. The functional characterization of novel mutations remains labor intensive and may require analyses in vitro or in transgenic models in order to document the relevance of the genetic alteration.

Prenatal diagnosis of numerous genetic diseases in instances with a high risk for certain disorders is possible by direct DNA analysis. *Amniocentesis* involves the removal of a small amount of amniotic fluid, usually at 16 weeks of gestation. Cells can be collected and submitted for karyotype analyses, FISH, and mutational analysis of selected genes (Table 479-4). The main indications for amniocentesis include advanced maternal age (>35 years), presence of an abnormality of the fetus on ultrasound examination, an abnormal serum "quad" test (α fetoprotein, β human chorionic gonadotropin, inhibin-A, and unconjugated estriol), a family history of chromosomal abnormalities, or a Mendelian disorder amenable to genetic testing. Prenatal diagnosis can also be performed by *chorionic villus sampling* (CVS), in which a small amount of the chorion is removed by a transcervical or transabdominal biopsy. Chromosomes and DNA obtained from these cells can be submitted for cytogenetic and mutational analyses. CVS can be performed earlier in gestation (weeks 9–12) than amniocentesis, an aspect that may be of relevance when termination of pregnancy is a consideration. Later in pregnancy, beginning at ~18 weeks of gestation, percutaneous umbilical blood sampling (PUBS; cordocentesis) permits collection of fetal blood for analysis. Prenatal cfDNA allows DNA analyses from the mother and fetus from a maternal blood sample to screen for certain chromosomal abnormalities and fetal sex. These approaches enable screening for clinically relevant and deleterious alleles inherited from the parents, as well as for de novo germline mutations, and they have the potential to identify genetic disorders in the prenatal setting.

In combination with in vitro fertilization (IVF) techniques, it is possible to perform genetic diagnoses in a single cell removed from the four- to eight-cell embryo or to analyze the first polar body from an oocyte. Preconceptual diagnosis thereby avoids therapeutic abortions but is costly and labor intensive. It should be emphasized that excluding a specific disorder by any of these approaches is never equivalent to the assurance of having a normal child.

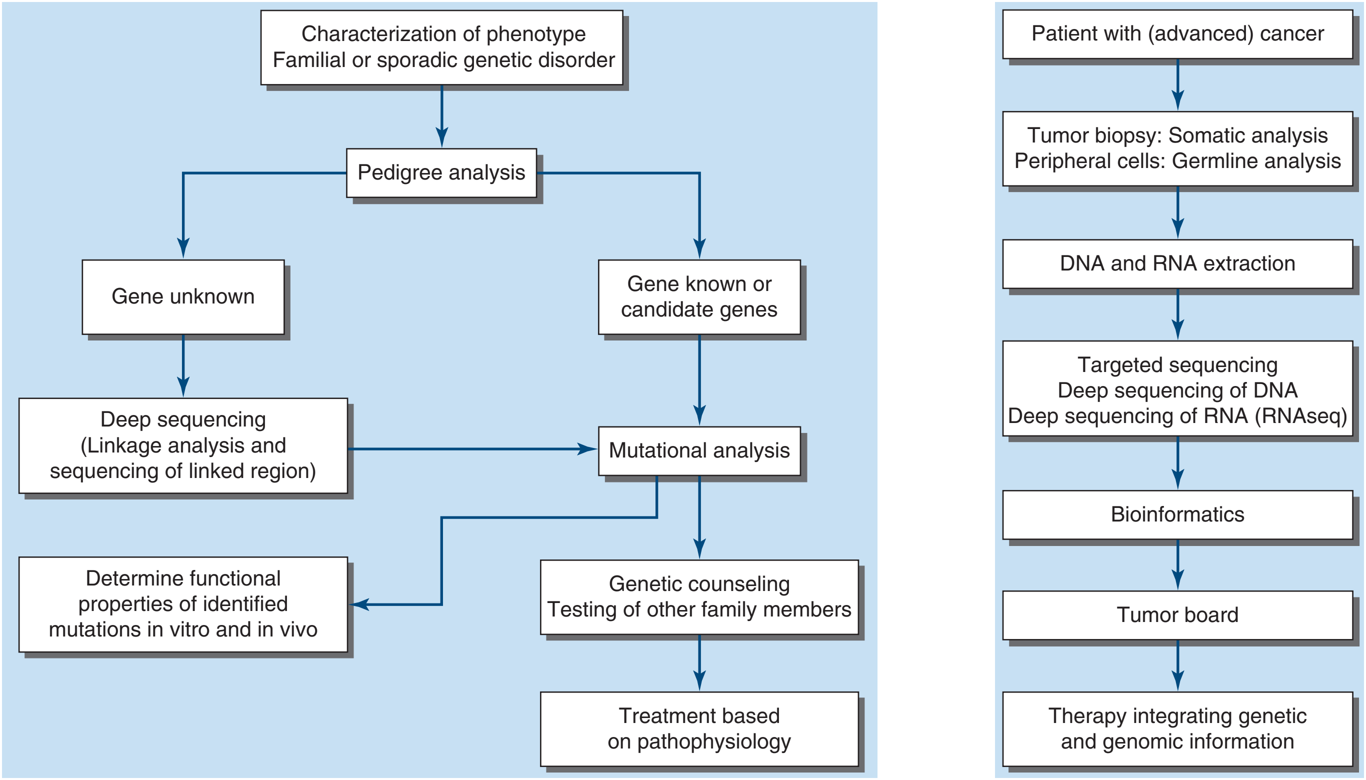
Characterization of phenotype
Familial or sporadic genetic disorder
Pedigree analysis
Gene unknown
Gene known or candidate genes
Deep sequencing (Linkage analysis and sequencing of linked region)
Mutational analysis
Determine functional properties of identified mutations in vitro and in vivo
Genetic counseling
Testing of other family members
Treatment based on pathophysiology
Patient with (advanced) cancer
Tumor biopsy: Somatic analysis
Peripheral cells: Germline analysis
DNA and RNA extraction
Targeted sequencing
Deep sequencing of DNA
Deep sequencing of RNA (RNAseq)
Bioinformatics
Tumor board
Therapy integrating genetic and genomic information

FIGURE 479-17 Approach to genetic disease.

Postnatal indications for cytogenetic analyses in infants or children include multiple congenital anomalies, suspicion of a known cytogenetic syndrome, developmental delay, dysmorphic features, autism, short stature, and disorders of sexual development, among others (Table 479-4).

Mutations in certain cancer susceptibility genes such as *BRCA1* and *BRCA2* may identify individuals with an increased risk for the development of malignancies and result in risk-reducing interventions. The detection of cytogenetic alterations and mutations is an important diagnostic and prognostic tool in leukemias, and it has also transformed the management of solid tumors. In addition to providing diagnostic information, mutational analysis can inform the choice of targeted therapies ("actionable mutations"), characterize the mutational load, identify gene signatures associated with effective immunotherapies, and be used for surveillance.

The demonstration of the presence or absence of mutations and polymorphisms is also relevant for the field of pharmacogenomics, including the identification of differences in drug treatment response or metabolism as a function of genetic background

Gene therapy through the introduction of a normal gene or the ability to make site-specific modifications to the human genome has, so far, limited clinical application. However, several gene transfer methods have now been approved for clinical use, for example, for the treatment of Leber congenital amaurosis, B-cell acute lymphoblastic leukemia, spinal muscular atrophy, and hereditary transthyretin-mediated amyloidosis. Genome editing (or gene editing) with CRISPR-Cas9 is a promising novel approach for the treatment of various diseases, for example cystic fibrosis, certain cancers, hemophilia, and sickle cell disease. The first therapies using this technology for the treatment of sickle cell disease were approved by the U.S. Food and Drug Administration in 2023 **(Chap. 483)**.

ETHICAL ISSUES

Determination of the association of genetic defects with disease, comprehensive data of an individual's genome, and studies of genetic variation raise many ethical and legal issues. Genetic information is generally regarded as sensitive information that should not be readily accessible without explicit consent (*genetic privacy*). The disclosure of genetic information may risk possible discrimination by insurers or employers. The scientific components of the Human Genome Project have been paralleled by efforts to examine ethical, social, and legal implications. An important milestone emerging from these endeavors is the Genetic Information Nondiscrimination Act (GINA), signed into law in 2008, which aims to protect asymptomatic individuals against the misuse of genetic information for health insurance and employment. It does not, however, protect the symptomatic individual. Provisions of the U.S. Patient Protection and Affordable Care Act, effective in 2014, have, in part, filled this gap and prohibit exclusion from, or termination of, health insurance based on personal health status. Potential threats to the maintenance of genetic privacy include the increasing integration of genomic data into electronic medical records, compelled disclosures of health records, and direct-to-consumer genetic testing.

It is widely accepted that identifying disease-causing genes can lead to improvements in diagnosis, treatment, and prevention. However, the information gleaned from genotypic results can have quite different impacts, depending on the availability of strategies to modify the course of disease. For example, the identification of mutations that cause MEN 2 or hemochromatosis allows specific interventions for affected family members. On the other hand, at present, the identification of an Alzheimer's or Huntington's disease gene does not currently alter therapy and outcomes. Most genetic disorders are likely to fall into an intermediate category where the opportunity for prevention or treatment is significant but limited. However, the progress in this area is unpredictable, as underscored by the finding that angiotensin II receptor blockers appear to slow disease progression in Marfan's syndrome. Genetic test results can generate anxiety in affected individuals and family members. Comprehensive sequence analyses are particularly challenging because most individuals can be expected to harbor several serious recessive gene mutations. Moreover, the sensitivity of comprehensive sequence analyses is not always greater, for example, if CNV analysis is not integrated. Genetic manipulation and patient selection for gene therapy approaches have raised ethical controversy and safety concerns that remain unresolved.

The impact of genetic testing on health care costs remains unclear. It does vary among disorders and depends on the availability of effective therapeutic modalities, and there are significant differences among health care systems. A significant problem has arisen from the marketing of genetic testing directly to consumers by commercial companies. The validity of these tests is, in part, not well defined, and there are persisting concerns about the lack of appropriate regulatory oversight, the accuracy and confidentiality of genetic information, the availability of counseling, and the handling of these results.

Many issues raised by the genome project are familiar, in principle, to medical practitioners. For example, an asymptomatic patient with increased low-density lipoprotein (LDL) cholesterol, high blood pressure, or a strong family history of early myocardial infarction is known to be at increased risk of coronary heart disease. In such cases, the identification of risk factors and an appropriate intervention are beneficial. Likewise, patients with phenylketonuria, cystic fibrosis, or sickle cell anemia are often identified as having a genetic disease early in life. These precedents can be helpful for adapting policies that relate to genetic information.

One confounding aspect of the rapid expansion of information is that our ability to make clinical decisions often lags behind initial insights into genetic mechanisms of disease. For example, when genes that predispose to breast cancer such as *BRCA1* are described, they generate tremendous public interest in the potential to predict disease, but many years of clinical research are still required to rigorously establish genotype and phenotype correlations.

Genomics may contribute to improvements in global health by providing a better understanding of pathogens and diagnostics and through contributions to drug development. There is, however, persisting concern about a widening "genomics divide" because of the costs associated with these developments and uncertainty as to whether these advances will be accessible to the populations of developing countries faced with pressing health needs associated with poverty, infectious diseases, and the relative lack of essential infrastructure. The World Health Organization's "Human Genomics in Global Health Initiative" aims at attenuating these inequities surrounding genomic medicine.

Whether related to informed consent, participation in research, or the management of a genetic disorder that affects an individual or their family, there is a great need for more information about fundamental principles of genetics. The pervasive nature of the role of genetics in medicine makes it important for physicians and other health care professionals to become more informed about genetics and to provide advice and counseling in conjunction with trained genetic counselors **(Chap. 480)**. Therefore, the application of screening and prevention strategies does require continuing patient and physician education, changes in health care financing, and legislation to protect patient's rights.

ACKNOWLEDGMENT

Selected sections and Table 479-4 have been integrated from the chapter on chromosome disorders by Dr. Nancy B. Spinner and Dr. Laura K. Conlin, published in the 19th edition of Harrison's Principles in Internal Medicine.

■ FURTHER READING

Denny JC, Collins FS: Precision medicine in 2030: Seven ways to transform healthcare. Cell 184:1415, 2021.

Doudna JA: The promise and challenge of therapeutic genome editing. Nature 578:229, 2020.

High KA, Roncarolo MG: Gene therapy. N Engl J Med 381:455, 2019.

Jarvik GP, Evans JP: Mastering genomic terminology. Genet Med 19:491, 2017.

Khera AV et al: Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 50:1219, 2018.

Khoury MJ et al: Health equity in the implementation of genomics and precision medicine: A public health imperative. Genet Med 24:1630, 2022.

Lalonde E et al: Genomic diagnosis for pediatric disorders: Revolution and evolution. Front Pediatr 8:373, 2020.

Liao WW et al: A draft human pangenome reference. Nature 617:312, 2023.

Might M, Crouse AB: Why rare disease needs precision medicine-and precision medicine needs rare disease. Cell Rep Med 3:100530, 2022.

Mighton C et al: From the patient to the population: Use of genomics for population screening. Front Genet 13:893832, 2022.

Moore AM, Richer J: Genetic testing and screening in children. Paediatr Child Health 27:243, 2022.

Shickh S et al: The clinical utility of exome and genome sequencing across clinical indications: A systematic review. Hum Genet 140:1403, 2021.

Sullivan JA et al: Exome/genome sequencing in undiagnosed syndromes. Annu Rev Med 74:489, 2023.

U.S. Food and Drug Administration: FDA approves first gene therapies to treat patients with sickle cell disease (December 8, 2023). *https://www.fda.gov/news-events/press-announcements/fda-approves-first-gene-therapies-treat-patients-sickle-cell-disease*. Accessed March 1, 2024.

Wauters A, Van Hoyweghen I: Global trends on fears and concerns of genetic discrimination: A systematic literature review. J Hum Genet 61:275, 2016.

World Health Organization: Accelerating access to genomics for global health: Promotion, implementation, collaboration, and ethical, legal, and social issues: A report of the WHO Science Council. 1-46, 2022. *https://www.who.int/publications/i/item/9789240052857*. Accessed September 28, 2024.

480

The Integration of Inherited Genetics into Clinical Practice

Susan M. Domchek, J. Larry Jameson

Genetic testing for inherited abnormalities associated with disease risk is increasingly used in the practice of clinical medicine. Germline alterations include chromosomal abnormalities, specific gene pathogenic variants (also called mutations) with autosomal dominant or recessive patterns of transmission **(Chap. 479)**, and single nucleotide polymorphisms (SNPs) with small relative risks associated with disease. Germline alterations are responsible for disorders beyond classic Mendelian conditions with genetic susceptibility to common adult-onset diseases such as asthma, hypertension, diabetes mellitus, macular degeneration, and a number of types of cancer. For many of these diseases, there is a complex interplay of genes (often multiple) and environmental factors that affect lifetime risk, age of onset, disease severity, and treatment options.

The expansion of human genetic knowledge is changing our understanding of pathophysiology and influencing our classification of diseases. Awareness of genetic etiology can have an impact on clinical management, including prevention and screening for or treatment of a range of diseases. Primary care physicians are relied upon to help patients navigate testing and treatment options. Consequently, they must understand the genetic basis for a large number of genetically influenced diseases, incorporate personal and family history to determine the risk for a specific pathogenic variant, and be positioned to provide counseling. Even if patients are seen by genetic specialists who assess genetic risk and coordinate testing, primary care providers should be able to identify patients who meet guidelines for genetic testing, and offer information regarding the indications, limitations, risks, and benefits of genetic counseling and testing. They must also be prepared to offer risk-based management following genetic risk assessment. Given the pace of genetics, this is an increasingly difficult task. The field of clinical genetics has rapidly transitioned from single-gene testing to multigene panel testing, with techniques such as whole exome and genome sequencing available as well, increasing the complexity of test selection and interpretation, as well as patient education and medical decision-making.

COMMON ADULT-ONSET GENETIC DISORDERS

INHERITANCE PATTERNS

Adult-onset hereditary diseases follow multiple patterns of inheritance. Some are autosomal dominant conditions. These include many common cancer susceptibility syndromes such as hereditary breast and ovarian cancer (due to germline *BRCA1* and *BRCA2* pathogenic variants) and Lynch syndrome (caused by germline pathogenic variants in the mismatch repair genes *MLH1*, *MSH2*, *MSH6*, *PMS2* and *EPCAM*). In both of these examples, inherited pathogenic variants are associated with a high penetrance (lifetime risk) of cancer, although penetrance is incomplete (risk is not 100%). In other conditions, although there is autosomal dominant transmission, penetrance is lower, thereby making the disorders more difficult to recognize. For example, germline pathogenic variants in *CHEK2* increase the risk of breast cancer but with a moderate lifetime risk in the range of 20–30%, as opposed to 50–70% for pathogenic variants in *BRCA1* or *BRCA2*. Other adult-onset hereditary diseases are transmitted in an autosomal recessive fashion where two mutant alleles are necessary to cause full expression of disease. Examples include hemochromatosis and *MUTYH*-associated polyposis. There are more pediatric-onset autosomal recessive disorders, such as lysosomal storage diseases and cystic fibrosis.

The genetic risk for many adult-onset disorders is multifactorial. Risk can be conferred by genetic factors at a number of loci (polygenic), which individually have very small effects (usually with relative risks of <1.5). When such risk loci (generally SNPs) are combined together, a polygenic risk score (PRS) can be provided. How PRS combine with other genes and environmental factors to impact risk is an area of active study. Despite our incomplete understanding of gene-environment interactions, data suggest that a healthy lifestyle can mitigate risk associated with elevated polygenic risk diseases, such as cardiovascular disease. SNP panels, which provide a PRS, are available to assess risk of disease, but the optimal way of using this information in the clinical setting to improve patient outcomes remains uncertain.

Many diseases have multiple patterns of inheritance, adding to the complexity of evaluating patients and families for these conditions. For example, colon cancer can be associated with a single germline highly penetrant pathogenic variant in a mismatch repair gene (Lynch syndrome; autosomal dominant), biallelic pathogenic variants in *MUTYH* (autosomal recessive), or multiple SNPs (polygenic). Many more individuals will have SNP risk alleles than germline pathogenic variants in high-penetrance genes, but the cumulative lifetime risk of colon cancer related to the former is modest, whereas the risk related to the latter is significant. Personal and family histories provide important insights into the possible mode of inheritance.

FAMILY HISTORY

When two or more first-degree relatives are affected with asthma, cardiovascular disease, type 2 diabetes, breast cancer, colon cancer, or melanoma, the relative risk for disease among close relatives ranges from two- to fivefold, underscoring the importance of family history for these prevalent disorders. In most situations, the key to assessing the inherited risk for common adult-onset diseases is the collection and interpretation of a detailed personal and family medical history in conjunction with a directed physical examination.

Family history should be recorded in the form of a pedigree, conveying health-related data on first- and second-degree relatives. When such pedigrees suggest inherited disease, they should be expanded to include additional family members. The determination of risk for an asymptomatic individual will vary depending on the size of the family, the number of unaffected relatives, the types of diagnoses, and the ages of disease onset. For example, a woman with two first-degree relatives with breast cancer is at greater risk for a specific Mendelian disorder if she has a total of 3 female first-degree relatives (with only 1 unaffected) than if she has a total of 10 female first-degree relatives (with 8 unaffected). Factors such as adoption and limited family structure (few women in a family or multiple early deaths unrelated to the target disease) should be taken into consideration in the interpretation of a pedigree. Additional considerations include young age of disease onset (e.g., a 30-year-old nonsmoking woman with a myocardial infarction), unusual diseases (e.g., male breast cancer or medullary thyroid cancer), and the finding of multiple potentially related diseases in an individual (e.g., a woman with a history of both colon and endometrial cancer). Some adult-onset diseases are more prevalent in certain ethnic groups. For instance, 2.5% of individuals of Ashkenazi Jewish ancestry carry one of three founder pathogenic variants in *BRCA1* and *BRCA2*. Factor V Leiden pathogenic variants are much more common in Caucasians than in Africans or Asians.

Additional variables that should be documented are nonhereditary risk factors among those with disease (such as cigarette smoking and myocardial infarction; asbestos exposure and lung disease; and mantle radiation and breast cancer). Significant associated environmental exposures or lifestyle factors decrease the likelihood of a specific genetic disorder. In contrast, the absence of nonhereditary risk factors typically associated with a disease raises concern about a genetic association. A personal or family history of deep-vein thrombosis in the absence of known environmental or medical risk factors suggests a hereditary thrombotic disorder. The physical examination may also provide important clues about the risk for a specific inherited disorder. A patient presenting with xanthomas at a young age should prompt consideration of familial hypercholesterolemia. The presence of trichilemmomas in a woman with breast cancer raises concern for Cowden syndrome, associated with *PTEN* pathogenic variants.

Recall of family history is often inaccurate. This is especially so when the history is remote and families lose contact or separate geographically. It can be helpful to ask patients to fill out family history forms before or after their visits, because this provides them with an opportunity to contact relatives. Ideally, this information should be embedded in electronic health records and updated intermittently. Attempts should be made to confirm the illnesses reported in the family history before making important and, in certain circumstances, irreversible management decisions. This process is often labor intensive and ideally involves interviews of additional family members or review of medical records (including pathology reports) and death certificates.

Although many inherited disorders will be suggested by the clustering of relatives with the same or related conditions, it is important to note that disease penetrance is incomplete for most genetic disorders. As a result, the pedigree obtained in such families may not exhibit a clear Mendelian inheritance pattern because not all family members carrying the disease-associated alleles will manifest clinical evidence of the condition. Furthermore, genes associated with some of these disorders often exhibit variable disease expression. For example, pathogenic variants in the breast cancer–associated gene *BRCA2* can predispose to several different malignancies in the same family, including cancers of the breast, ovary, pancreas, and prostate. For common diseases such as breast cancer, some family members without the susceptibility allele (or genotype) may develop breast cancer (or phenotype) sporadically. Such *phenocopies* represent another confounding variable in the pedigree analysis.

Some of the aforementioned features of the family history are illustrated in **Fig. 480-1**. In this example, the proband (the individual serving as the starting point for genetic assessment in a family), a

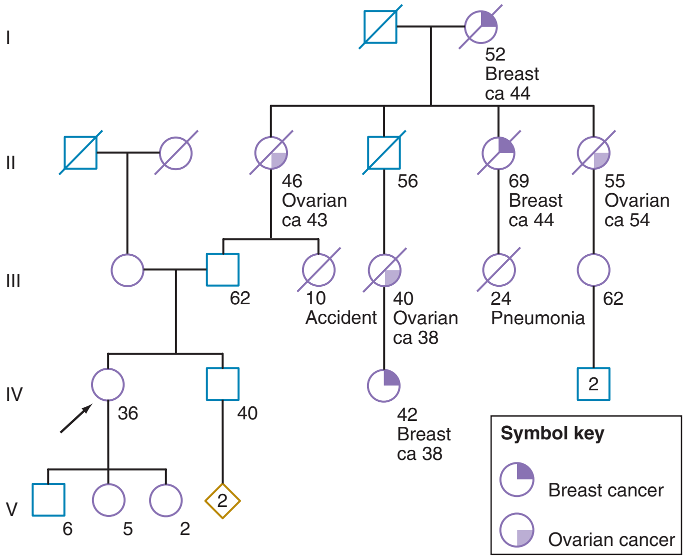

FIGURE 480-1 A 36-year-old woman (*arrow*) seeks consultation because of her family history of cancer. The patient expresses concern that the multiple cancers in her relatives imply an inherited predisposition to develop cancer. The family history is recorded, and records of the patient's relatives confirm the reported diagnoses.

36-year-old woman (IV-1), has a strong history of breast and ovarian cancer on the paternal side of her family. The early age of onset and the co-occurrence of breast and ovarian cancer in this family suggest the possibility of an inherited pathogenic variant in *BRCA1* or *BRCA2*. However, without genetic testing, it is unclear whether her father harbors such a pathogenic variant and transmitted it to her. After appropriate genetic counseling of the proband and her family, the most informative and cost-effective approach to DNA analysis in this family is to test the cancer-affected 42-year-old living cousin for the presence of a pathogenic variant in cancer susceptibility genes. If a pathogenic variant is found, then it is possible to test for this particular alteration in other family members, if they so desire. In the example shown, if the proband's cousin has a *BRCA1* pathogenic variant, testing for this alteration would be an option for the proband's father. If he tests positive, there is a 50:50 probability that the pathogenic variant was transmitted to her, and genetic testing can be used to establish the absence or presence of this alteration. In contrast, if he tests negative for the known familial *BRCA1* pathogenic variant (a true-negative result), the proband and her brother are not at risk for having inherited this variant from their father.

GENETIC TESTING FOR ADULT-ONSET DISORDERS

A critical first step before initiating genetic testing is to ensure that the correct clinical diagnosis has been made, whether it is based on family history, characteristic physical findings, pathology, or biochemical testing. Such careful clinical assessment can define the phenotype. In the traditional model of genetic testing, testing is directed initially toward the most probable genes (determined by the phenotype), which prevents unnecessary testing. Many disorders exhibit the feature of locus heterogeneity, which refers to the fact that pathogenic variants in different genes can cause phenotypically similar disorders. For example, osteogenesis imperfecta **(Chap. 425)**, long QT syndrome **(Chap. 259)**, muscular dystrophy **(Chap. 460)**, and hereditary predisposition to breast **(Chap. 84)** or colon **(Chap. 86)** cancer can each be caused by pathogenic variants in a number of distinct genes. The patterns of disease transmission, disease risk, clinical course, and treatment may differ significantly depending on the specific gene affected. Historically, the choice of which gene to test has been determined by unique clinical and family history features and the relative prevalence of candidate genetic disorders. However, rapid developments in genetic testing techniques, as discussed below, changed this paradigm. It is technically and financially feasible to sequence many genes (or even the whole exome) at one time. The incorporation of expanded testing for germline pathogenic variants has rapidly evolved both within the clinic as well as through direct-to-consumer marketing of genetic and genomic tests.

■ METHODOLOGIC APPROACHES TO GENETIC TESTING

Genetic testing is regulated and performed in much the same way as other specialized laboratory tests. In the United States, genetic testing laboratories are Clinical Laboratory Improvement Amendments (CLIA) approved to ensure that they meet quality and proficiency standards. A useful information source for various genetic tests is *www.genetests.org*. It should be noted that many tests need to be ordered through specialized laboratories.

Genetic testing is performed largely by DNA sequence analysis for pathogenic variants, although genotype can also be deduced through the study of RNA or protein (e.g., apolipoprotein E, hemoglobin S, and immunohistochemistry). For example, universal Lynch syndrome screening of colorectal and uterine cancers via immunohistochemical analysis for absence of expression of mismatch repair proteins is recommended by the National Comprehensive Cancer Network. The determination of DNA sequence alterations relies heavily on the use of polymerase chain reaction (PCR), which allows rapid amplification and analysis of the gene of interest. In addition, PCR enables genetic testing on minimal amounts of DNA extracted from a wide range of tissue sources including leukocytes (obtained from blood), leukocytes and mucosal epithelial cells (obtained via saliva or buccal swabs), and tumors (obtained from biopsies or archival tissues). Amplified DNA can be analyzed directly by DNA sequencing, or it can be hybridized to DNA chips or blots to detect the presence of normal and altered DNA sequences. Direct DNA sequencing is frequently used for determination of hereditary disease susceptibility and prenatal diagnosis. Analyses of large alterations (e.g., deletions, duplications, rearrangements, translocations) of the genome are possible using cytogenetics, fluorescent in situ hybridization (FISH), Southern blotting, multiplex ligation-dependent probe amplification (MLPA), and massively parallel sequencing.

Massively parallel sequencing has significantly altered the approach to genetic testing for adult-onset hereditary susceptibility disorder. This technology encompasses high-throughput approaches to DNA analysis that can reliably examine many genes at one time. Technically, this involves sequencing of millions of small fragments of DNA in parallel. Through bioinformatics, these fragments are pieced together by mapping the individual sequence reads to the human reference genome, a very different process than traditional Sanger sequencing, which is time-consuming and expensive.

Multiplex panels for inherited susceptibility are commercially available and include testing of a number of genes that have been associated with the condition of interest. For example, panels are available for Brugada syndrome, hypertrophic cardiomyopathy, and Charcot-Marie-Tooth neuropathy. For many syndromes, this type of panel testing makes sense as pathogenic variants in many different genes may lead to the same phenotype. A challenge is that the composition of commercially available gene panels often contains genes with no validated association with the disease of interest or a modest association for which clinical utility is uncertain. An additional problem of sequencing many genes, rather than focusing on a more limited panel of clearly associated gene, is the identification of one or more variants of uncertain significance (VUS), discussed below, or an unexpected yet clinically relevant result.

Whole exome sequencing (WES) is also now commercially available, although largely used in individuals with syndromes unexplained by traditional genetic testing. As cost declines, WES may be more widely used. Whole genome sequencing is also commercially available. Although it may be quite feasible to sequence the entire genome, there are many issues in doing so, including the daunting task of analyzing the vast amount of data generated. Other issues include (1) interpretation of frequent sequence variations of uncertain significance, (2) interpretation of alterations in genes with unclear relevance to specific human pathology, and (3) management of unexpected but clinically significant genetic findings.

Testing strategies have evolved as a result of these newer genetic testing platforms. As the costs of multiplex gene panels and WES continue to fall and as interpretation and understanding of the clinical relevance of such test results improve, there has been a shift to more extensive panel-based genetic testing in the clinic. For example, in the past, a 30-year-old woman with breast cancer but no family history of cancer and no syndromic features would undergo *BRCA1/2* testing and would be offered *TP53* testing in light of her early-onset disease. Without features consistent with other high-risk, breast cancer–related conditions, the woman would not have been routinely offered, for example, *PTEN* analysis (associated with Cowden syndrome) or testing for other breast cancer–associated genes including *PALB2*, *CHEK2*, *BARD1,* and *ATM*. It is now possible to synchronously analyze all of the aforementioned genes, along with genes such as *BRIP, RAD51C,* and *RAD51D* (which are associated with moderate ovarian cancer risk) rather than testing for *BRCA1/2* alone. Concerns about such panels include strategies related to unclear findings, including one or more VUSs, unanticipated results, and the uncertain clinical utility of some of the genes included on the panel **(Fig. 480-2)**.

DIRECT-TO-CONSUMER GENETIC TESTING

Historically, genetic testing occurred entirely, or predominantly, in the context of clinical care. However, technologic advances combined with shifts in social norms have created an increasingly complex landscape for genetic information, and patients come to providers with a wide range of information.

Clinician-directed genetic testing is currently based on the individual's personal or family medical history. Results are disclosed by the clinician with a management plan formulated based on results, within the context of the individual's personal and family medical history. Consumer-directed genetic testing differs in that the individual electively requests the tests to be performed by a direct-to-consumer (DTC) company, submits the sample directly, and results are communicated directly to the consumer.

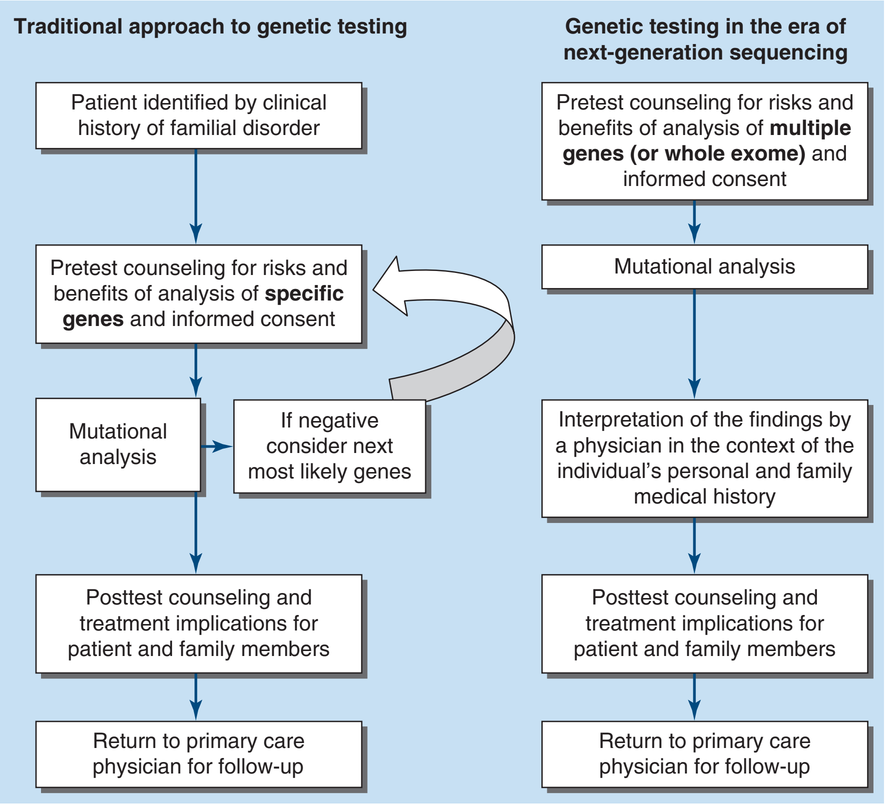
Traditional approach to genetic testing
Patient identified by clinical history of familial disorder
Pretest counseling for risks and benefits of analysis of **specific genes** and informed consent
Mutational analysis
If negative consider next most likely genes
Posttest counseling and treatment implications for patient and family members
Return to primary care physician for follow-up
Genetic testing in the era of next-generation sequencing
Pretest counseling for risks and benefits of analysis of **multiple genes (or whole exome)** and informed consent
Mutational analysis
Interpretation of the findings by a physician in the context of the individual's personal and family medical history
Posttest counseling and treatment implications for patient and family members
Return to primary care physician for follow-up

FIGURE 480-2 Approach to genetic testing.

DTC testing can include assays for disease risk or disease carrier status; others examine ancestry, pharmacogenetics, and "recreational" genetics related to traits such as aversion to cilantro and earlobe type. From a clinical utility standpoint, such testing companies most often examine sets of SNPs, common "hot spots" in genes associated with specific diseases; however, SNP analysis does not provide a definitive risk/no risk answer. In contrast, clinical genetic testing thoroughly interrogates specific genes known be associated with an individual disease or disease risk. Quality standards between DTC and clinical laboratories vary greatly. In the United States, clinical genetic testing labs meet established strict quality standards discernible as CLIA certification and College of American Pathologists (CAP) accreditation. Some DTC labs are granted U.S. Food and Drug Administration (FDA) marketing authorization, yet such results often come with a disclaimer that they should not be used for medical management. Recommendations from the American College of Medical Genetics and Genomics are summarized in **Table 480-1**.

A third, hybrid model has been developed by commercial genetic testing laboratories, particularly focused on analysis for cancer and cardiovascular disease risk. Here, after an individual initiates the testing process, either their own physician or a physician provided by the company orders the test. Many of these tests are similar to those offered to a patient in the clinical setting; however, the order and results remain separate from the medical record unless integrated by the patient's primary or specialist physician or other health care provider.

LIMITATIONS OF GENETIC TESTING

Limitations to the accuracy and interpretation of genetic testing exist. In addition to technical errors, genetics tests are sometimes designed to detect only the most common pathogenic variants. In addition, genetic testing has evolved over time. For example, it was not possible to obtain commercially available, comprehensive, large genomic rearrangement testing for *BRCA1* and *BRCA2* until 2006. Therefore, a negative result must be qualified by the possibility that the individual may have a variant that was not detectable in the test. In addition, the individual may have a pathogenic variant in another untested cancer-associated gene or in a gene not yet reported to be associated with elevated disease risk. As such, unless there is a known pathogenic variant in the family, a negative result in an individual with a suggestive personal or family history is typically classified as an uninformative negative. In this circumstance, medical management decisions should be based on personal and family history. For example, a man with a strong family history of colon cancer who receives an uninformative negative panel-based genetic test result may still be eligible earlier and more frequent colonoscopy screening.

The finding of a VUS is another limitation to genetic testing. A VUS (also termed *unclassified variant*) is a sequence variation in a gene where the effect of the alteration on the function of the protein is not known. Many of these variants are single nucleotide substitutions (also called missense alterations) that result in a single amino acid change. Although most VUSs will ultimately be reclassified as benign variants, some will prove to be functionally important. As more genes are sequenced (e.g., in a multiplex panel or through WES), the percentage of individuals found to have one or more VUSs increases significantly. The finding of a VUS is difficult for patients and providers alike and complicates decisions regarding medical management. In this setting, until there is further reclassification of the variant, ongoing screening, surveillance, and care are typically determined based on personal and family history.

TABLE 480-1 Summary of American College of Medical Genetics and Genomics Position Statement on Direct-to-Consumer Genetic Testing

The clinical laboratory must:	• Be accredited by the CLIA (Clinical Laboratory Improvement Amendments) program, the state, and/or other applicable accrediting agencies with this accreditation indicated on test result reports
Appropriately trained professionals should order, interpret, and disclose test result to avoid:	• Inadequate or lack of informed consent • Testing without appropriate indications • Selection of inadequate testing methods or incorrect test • Misinterpretation of results leading to improper clinical management
The individual undergoing testing should be adequately informed of:	• What the test can or cannot determine about their health or health risk • The potential for receiving results that neither confirm nor exclude the possibility of disease • Potential unexpected results, unrelated to the indication for testing • The potential impact of a given genetic test result on relatives
Professionals will review and provide information about:	• The validity and utility of a genetic test, including the limitations of a given test • Privacy issues to include review of who will have access to test results, the processes in place to protect results, and disposition of DNA samples once analysis is complete • Personal or family implications for life, long-term care, or disability insurance • Whether data generated will be shared with third parties

Source: Adapted from ACMG Board of Directors: Direct-to-consumer genetic testing: A revised position statement of the American College of Medical Genetics and Genomics. Genet Med 18:207, 2016.

Clinical utility is an important consideration because genetic testing for susceptibility to chronic diseases is increasingly integrated into the practice of medicine. In some situations, there is proven clinical utility to genetic testing with significant evidence-based changes in medical management options and recommendations based on results. For example, there is clear evidence that risk-reducing bilateral salpingo-oophorectomy benefits women with a documented *BRCA1/2* pathogenic variant with improvement in mortality. In some cases, the discovery of disease-associated genes has outpaced studies that assess how such information should be used in the clinical management of the patient and family. Risks associated with genes often evolve with additional data. Therefore, predictive genetic testing should be done following appropriate pretest education.

TESTING STRATEGIES

Predictive genetic testing falls into two distinct categories. Presymptomatic testing applies to diseases where a specific genetic alteration is associated with a near 100% likelihood of developing disease. In contrast, predisposition testing predicts a risk for disease that is <100%. Presymptomatic testing is available for those at risk for Huntington's disease, whereas predisposition testing is considered for those at risk for hereditary colon cancer. It is important to note that for the majority of adult-onset disorders, testing is only predictive. Test results cannot reveal with confidence whether, when, or how the disease will manifest itself. For example, not everyone with the apolipoprotein ε4 allele will develop Alzheimer's disease, and individuals without this genetic marker can still develop the disorder.

The optimal testing strategy for a family is to initiate testing in a disease-affected family member first. Identification of a pathogenic variant can direct the testing of other at-risk family members (whether symptomatic or not). In the absence of additional familial or environmental risk factors, individuals who test negative for the pathogenic variant found in the affected family member are at general population risk for that particular disease. Furthermore, they can be reassured that they are not at risk for passing the pathogenic variant on to their children. On the other hand, asymptomatic family members who test positive for the known pathogenic variant must be informed that they are at increased risk for disease development and for transmitting the alteration to their children.

Pretest education is important, as is an assessment of the patient's ability to understand and cope with test results. Genetic testing has implications for entire families, and thus, individuals interested in pursuing genetic testing should consider how test results might impact their relationships with relatives, partners, spouses, and children. In families with a known genetic pathogenic variant, those who test positive must consider the impact of their carrier status on their present and future lifestyles; those who test negative may manifest survivor guilt. Parents who are found to have a disease-associated pathogenic variant often express considerable anxiety and despair as they address the issue of risk to their children. In addition, some individuals consider options such as preimplantation genetic testing (PGT) (which allows for selection of embryos without the familial pathogenic variant of concern) in their reproductive decision-making.

When a condition does not manifest until adulthood, clinicians and parents are faced with the question of whether at-risk children should be offered genetic testing and, if so, at what age. Although the matter is debated, several professional organizations have cautioned that genetic testing for adult-onset disorders should not be offered to children. Many of these conditions have no known interventions in childhood to prevent disease; consequently, such information can pose significant psychosocial risk to the child. In addition, there is concern that testing during childhood violates a child's right to make an informed decision regarding testing upon reaching adulthood. On the other hand, testing should be offered in childhood for disorders that may manifest early in life, especially when management options are available. For example, children with multiple endocrine neoplasia 2 (MEN 2) may develop medullary thyroid cancer early in life and should be considered for prophylactic thyroidectomy **(Chap. 400)**. Similarly, children with familial adenomatous polyposis (FAP) due to a pathogenic variant in *APC* may develop polyps in their teens with progression to invasive cancer in the twenties, and therefore, colonoscopy screening is started between the ages of 10 and 15 years **(Chap. 86)**.

INFORMED CONSENT

Individuals undergoing genetic testing should understand the risks, benefits, and limitations of genetic testing, as well as the potential implications of test results. Because molecular genetic testing of an asymptomatic individual often allows prediction of future risk, the patient should understand all potential long-term medical, psychological, and social implications of testing. There have long been concerns about the potential for genetic discrimination. The Genetic Information Nondiscrimination Act (GINA) was passed in 2008 and provides some protections related to job and health insurance discrimination, but does not address disability, long-term care, and life insurance coverage. Patients should understand that alternatives remain available if they decide not to pursue genetic testing, including the option of delaying testing to a later date. The option of DNA banking should be presented so that samples are readily available for future use by family members, if needed.

FOLLOW-UP CARE AFTER TESTING

Depending on the nature of the genetic disorder, posttest interventions may include (1) cautious surveillance and awareness; (2) specific medical interventions such as enhanced screening, chemoprevention, or risk-reducing surgery; (3) risk avoidance; and (4) referral to support services. For example, patients with known deleterious pathogenic variants in *BRCA1* or *BRCA2* are strongly encouraged to undergo risk-reducing salpingo-oophorectomy at an appropriate age and are offered intensive breast cancer screening with the addition of contrast-enhanced magnetic resonance imaging (MRI) as well as the option of risk-reducing mastectomy. In addition, such women may wish to take chemoprevention with tamoxifen, raloxifene, or exemestane. Those with more limited medical management and prevention options, such as patients with Huntington's disease, should be offered continued follow-up and supportive services, including physical and occupational therapy and social services or support groups as indicated. Specific interventions will change as research continues to enhance our understanding of the medical management of these genetic conditions and more is learned about the functions of the gene products involved.

Individuals who test negative for a pathogenic variant in a disease-associated gene identified in an affected family member must be reminded that they may still be at risk for the disease. This is of particular importance for common diseases such as diabetes mellitus, cancer, and coronary artery disease. For example, a man who finds that he does not carry the disease-associated pathogenic variant in *MLH1* previously discovered in the family should be reminded that he still requires the same colon cancer screening recommended for the general population.

POPULATION SCREENING FOR GERMLINE DISEASE–ASSOCIATED VARIANTS

As costs of genetic testing have fallen, associated risks related to pathogenic variants have become more precise, and evidence for the clinical utility of interventions has increased, the conversation regarding the role of population screening has intensified. In an argument for population screening, several studies, including from large unselected biobanks, have demonstrated that many individuals with pathogenic variants in clearly established risk genes (e.g., *BRCA1*, *BRCA2*, *MLH1*, *MSH2*, *TTR*, *LDLR*) do not have a personal or family history that would have met current guidelines for genetic testing. In Israel, this had led to insurance coverage for all individuals of Ashkenazi Jewish descent for the three common *BRCA1* and *BRCA2* founder mutations (which are found in 2.5% of individuals of Ashkenazi Jewish descent). Challenges in the consideration of population screening include, but are not

TABLE 480-2 Examples of Genetic Testing and Possible Interventions

GENETIC DISORDER	INHERITANCE	GENES	INTERVENTIONS
Oncologic			
Lynch syndrome (HNPCC)	AD	*MLH1, MSH2, MSH6, PMS2*	Early endoscopic screening; risk-reducing surgery; aspirin
Familial adenomatous polyposis	AD	*APC*	Early and frequent endoscopy; prophylactic colectomy; chemoprevention
Hereditary breast and ovarian cancer	AD	*BRCA1, BRCA2*	Risk-reducing salpingo-oophorectomy; intensified breast surveillance including breast MRI; risk-reducing mastectomy; PARP inhibitors
Hereditary diffuse gastric cancer	AD	*CDH1*	Prophylactic gastrectomy; enhanced breast cancer surveillance
Hematologic			
Factor V Leiden	AD	*F5*	Avoidance of thrombogenic risk factors
Hemophilia A	XL	*F8*	Factor VIII replacement
Hemophilia B	XL	*F9*	Factor IX replacement
Glucose-6-phosphate dehydrogenase deficiency	XL	*G6PD*	Avoidance of oxidant drugs and certain foods
Sickle cell disease	AR	*HBB*	Bone marrow transplantation, hydroxyurea, gene therapy
Cardiovascular			
Hypertrophic cardiomyopathy	AD	>10 genes including *MYBPC3, MYH7, TNNT2, TPM1*	Echocardiographic screening; pharmacologic intervention; myomectomy
Long QT syndrome	AD, AR	>10 genes including *KCNQ1, SCN5A, KCNE1, KCNE2*	Electrocardiographic screening; pharmacologic intervention; implantable cardiac defibrillator devices
Marfan's syndrome	AD	*FBN1*	Echocardiographic screening; prophylactic beta blockers; aortic valve replacement as indicated
Gastrointestinal			
Familial Mediterranean fever	AR	*MEFV*	Colchicine
Hemochromatosis	AR	*HFE*	Phlebotomy
Pulmonary			
α_1 Antitrypsin deficiency	AR	*SERPINA1*	Avoidance of smoking and occupational and environmental toxins
Cystic fibrosis	AR	*CFTR*	Chest physiotherapy; agents to promote airway secretion clearance; *CFTR* modulators; lung transplantation
Endocrine			
Neurohypophyseal diabetes insipidus	AD	*AVP*	Replace vasopressin
Familial hypocalciuric hypercalcemia	AD	*CASR*	Avoidance of parathyroidectomy; calcimimetics
Multiple endocrine neoplasia type 2	AD	*RET*	Prophylactic thyroidectomy; screening for pheochromocytoma and hyperparathyroidism
Renal			
Polycystic kidney disease	AD, AR	*PKD1, PKD2, PKHD1*	Prevention of hypertension; prevention of urinary tract infections; kidney transplantation
Nephrogenic diabetes insipidus	XL, AR	*AVPR2, AQP2*	Fluid replacement; thiazides with or without amiloride
Neurologic			
Malignant hyperthermia	AD	*RYR1, CACNA1S*	Avoidance of precipitating anesthetics
Hyperkalemic periodic paralysis	AD	*SCN4A*	Diet rich in calcium and low in potassium; thiazides or acetazolamide
Duchenne's muscular dystrophy	XL	*DMD*	Corticosteroids; physical therapy; antisense oligonucleotide mediated exon skipping; abrogation of nonsense mediated decay by stop codon readthrough
Wilson's disease	AR	*ATP7B*	Zinc, trientine

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; HNPCC, hereditary nonpolyposis colorectal cancer; MRI, magnetic resonance imaging; XL, X-linked.

limited to, variable penetrance, which may be significantly impacted by family history (thus, those individuals with germline pathogenic variants detected via population screening in the absence of a classic family history may have significantly lower risks of disease), consensus of which genes should be included, management of VUS, and practical issues of implementation including insurance coverage, counseling related to the limitations of GINA, the role of primary care providers in ordering, and concerns about the possibility of worsening socioeconomic and racial disparities that already exist in genetic testing.

THERAPEUTIC INTERVENTIONS BASED ON GENETIC RISK FOR DISEASE

Specific treatments are available for a number of genetic disorders. Strategies for the development of therapeutic interventions have a long history in childhood metabolic diseases; however, these principles have been applied in the diagnosis and management of adult-onset diseases as well **(Table 480-2)**. Hereditary hemochromatosis is usually caused by pathogenic variants in *HFE* (although other genes have been less commonly associated) and manifests as a syndrome of iron overload, which can lead to liver disease, skin pigmentation, diabetes mellitus, arthropathy, impotence in males, and cardiac issues **(Chap. 426)**. When identified early, the disorder can be managed effectively with therapeutic phlebotomy. Therefore, when the diagnosis of hemochromatosis has been made in a proband, it is important to counsel other family members in order to minimize the impact of the disorder.

Preventative measures and therapeutic interventions are not restricted to metabolic disorders. Identification of familial forms of long QT syndrome, associated with ventricular arrhythmias, allows early electrocardiographic testing and the use of prophylactic antiarrhythmic therapy, overdrive pacemakers, or defibrillators. Individuals with familial hypertrophic cardiomyopathy can be screened by ultrasound, treated with beta blockers or other drugs, and counseled about the importance of avoiding strenuous exercise and dehydration. Those with Marfan's syndrome can be treated with beta blockers or angiotensin II receptor blockers and monitored for the development of aortic aneurysms. The identification of germline abnormalities that increase the risk of specific types of cancer is rapidly changing clinical management. Identifying family members with pathogenic variants that predispose to FAP or Lynch syndrome leads to recommendations of early cancer screening and prophylactic surgery, as well as consideration of chemoprevention and attention to healthy lifestyle habits. Similar principles apply to familial forms of melanoma as well as cancers of the breast, ovary, and thyroid.

There has been a significant growth in the number of molecularly directed therapies for genetic diseases including transthyretin stabilizers for *TTR*-associated cardiac amyloid; poly (ADP-ribose) polymerase (PARP) inhibitors for treatment of *BRCA1/2* and *PALB2*-associated breast, ovarian, prostate, and pancreatic cancer; and medications for Duchenne's muscular dystrophy that can either promote exon skipping or allow bypass of nonsense mutations. Gene therapy either by replacement, such as in spinal muscular atrophy, or in sickle cell disease, increasing production of fetal hemoglobin or hemoglobin A, is an exciting area with tremendous opportunities **(Chap. 483)**.

The field of pharmacogenetics identifies genes that alter drug metabolism or confer susceptibility to toxic drug reactions. Pharmacogenetics seeks to individualize drug therapy in an attempt to improve treatment outcomes and reduce toxicity. Examples include thiopurine methyltransferase (TPMT) deficiency, dihydropyrimidine dehydrogenase deficiency, malignant hyperthermia, and glucose-6-phosphate deficiency. Despite successes in this area, it is not always clear how to incorporate pharmacogenetics into clinical care. For example, although there is an association with *CYP2C6* and *VKORC1* genotypes and warfarin dosing, there is no evidence that incorporating genotyping into clinical practice improves patient outcomes compared with clinical algorithms.

Although the role of genetic testing in the clinical setting continues to evolve, such testing holds the promise of allowing early and more targeted interventions that can reduce morbidity and mortality. Rapid technologic advances are changing the ways in which genetic testing is performed. As genetic testing has become less expensive and technically easier to perform, there has been significant expansion of its use. This has created both challenges and opportunities. It is critical that physicians and other health care professionals keep current with advances in genetic medicine in order to facilitate appropriate referral for genetic counseling and judicious use of genetic testing, as well as to provide state-of-the-art, evidence-based care for affected or at-risk patients and their relatives.

■ FURTHER READING

ACMG Board of Directors: Direct-to-consumer genetic testing: A revised position statement of the American College of Medical Genetics and Genomics. Genet Med 18:207, 2016.

Anya ER et al: The Goldilocks conundrum: Disclosing discrimination risks in informed consent. J Genet Couns 31:1383, 2022.

Dewey FE et al: Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science 354:aaf6814, 2016.

Food and Drug Administration. FDA direct to consumer tests. Available at *https://www.fda.gov/medical-devices/in-vitro-diagnostics/direct-consumer-tests*. Accessed January 1, 2024.

Hampel H et al: A practice guideline from the American College of Medical Genetics and Genomics and the National Society of Genetic Counselors: Referral indications for cancer predisposition assessment. Genet Med 17:70, 2015.

Miller DT et al: ACMG SF v3.2 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). Genet Med 25:100866, 2023.

Robson ME et al: American Society of Clinical Oncology policy statement update: Genetic and genomic testing for cancer susceptibility. J Clin Oncol 33:3660, 2015.

Splinter K et al: Effect of genetic diagnosis on patients with previously undiagnosed disease. N Engl J Med 379:2131, 2018.

Turnbull C et al: Population screening requires robust evidence-genomics is no exception. Lancet 403:583, 2024.

481

Mitochondrial DNA and Heritable Traits and Diseases

Karl L. Skorecki, Bruce H. Cohen

Mitochondria are cytoplasmic organelles whose major function is to generate ATP by the process of oxidative phosphorylation under aerobic conditions. This process is mediated by the respiratory electron transport chain (ETC) multiprotein enzyme complexes I–V and the two electron carriers, coenzyme Q_{10} (CoQ_{10}) and cytochrome c, located in the inner mitochondrial membrane. Other cellular processes to which mitochondria make a major contribution include apoptosis (programmed cell death) and additional cell type–specific functions **(Table 481-1)**. The efficiency of the mitochondrial ETC in ATP production is the major determinant of overall body energy balance and thermogenesis. In addition, mitochondria are the predominant source of reactive oxygen species (ROS), whose rate of production relates to the delicately balanced coupling of ATP production to oxygen consumption in health and disease. Given the centrality of oxidative phosphorylation to the normal activities of almost all cells, it is not surprising that mitochondrial dysfunction can affect almost any organ system **(Fig. 481-1)**. Until recently, it was thought that disruption of

TABLE 481-1 Functions of Mitochondria

All Cells and Tissues
Oxidative phosphorylation
Free radical production
Calcium homeostasis
Apoptosis (programmed cell death)
Tissue- or Cell-Specific
Cholesterol metabolism
Amino and organic acid metabolism
Fatty acid beta oxidation
Sex steroid synthesis
Heme synthesis
Hepatic ammonia detoxification
Neurotransmitter metabolism

energy production was the source of the pathophysiology in those with mitochondrial dysfunction, but recent evidence suggests that free radical production and the redox state of the mitochondria may play a role as well. Thus, physicians in many disciplines might encounter patients with mitochondrial diseases and should be aware of their existence and characteristics.

The integrated activity of an estimated 1500 gene products is required for normal mitochondrial biogenesis, function, maintenance, and integrity. Aside from the 37 genes that comprise the mitochondrial DNA (mtDNA) molecule, the remaining 1400+ gene products are encoded by nuclear genes (referred to as nDNA) and thus follow the rules and patterns of nuclear genomic inheritance (Chap. 479). These nuclear-encoded proteins are synthesized in the cell cytoplasm and imported to their location of activity within the mitochondria through a complex biochemical process. This process includes unfolding of the nuclear-encoded protein, attachment to a chaperone protein that shuttles it through a specific channel to a specific mitochondrial location, and detachment from the chaperone followed by assembly with other mtDNA- and nDNA-encoded proteins. In addition, the mitochondria contain their own small genome consisting of numerous copies (polyploidy) per mitochondrion of a circular, double-strand mtDNA molecule comprising 16,569 nucleotides. This mtDNA sequence (also known as the "mitogenome") might represent the remnants of endosymbiotic prokaryotes from which mitochondria are thought to have originated. The mtDNA sequence contains a total of 37 genes, of which 13 encode mitochondrial protein components of the ETC (Fig. 481-2). The remaining 22 tRNA- and 2 rRNA-encoding genes are mitochondria-specific and dedicated to the process of translating the 13 mtDNA-encoded proteins. The mtDNA itself replicates constantly, independent of cell division, and requires its own unique polymerase, referred to as polymerase gamma (polγ), which is encoded by the nuclear gene *POLG*, disorders of which are discussed in **Chaps. 460 and 480**. However, mutations in *POLG* can disrupt the endonuclease function of polγ, resulting in somatic mutations in the mtDNA that endure with future replication. Unless this mutation occurs and is propagated in the oocyte, it is not heritable. Mutations in *POLG* can also affect the polymerase function of polγ that results in decreased replication of the mtDNA and can lead to mtDNA depletion. This dual nuclear and mitochondrial genetic control of mitochondrial function results in unique and diagnostically challenging patterns of inheritance. The current chapter focuses on heritable traits and diseases related to the mtDNA component of the dual genetic control of mitochondrial function. The reader is referred to **Chaps. 460 and 479** for consideration of mitochondrial disease originating from mutations in the nuclear genome. The former include (1) disorders due to mutations in nuclear genes directly encoding structural components or assembly factors of the oxidative phosphorylation complexes, (2) disorders due to mutations in nuclear genes encoding proteins indirectly related to oxidative phosphorylation, (3) mtDNA depletion syndromes (MDSs) characterized by a reduction of mtDNA copy number in affected tissues without mutations or rearrangements in the mtDNA, and (4) disorders due to mutations in nuclear genes that disrupt normal mitochondrial dynamics (biosynthesis, mitophagy, fission, and fusion).

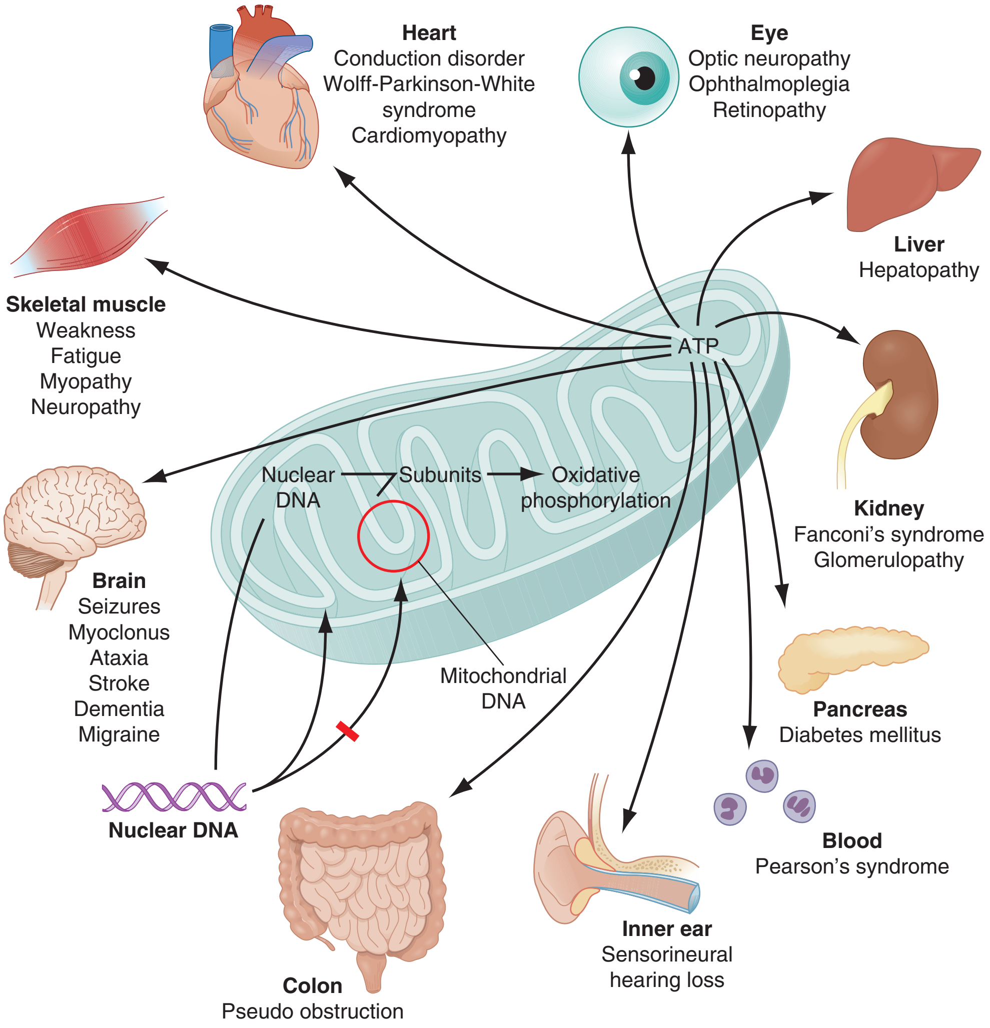
Heart
Conduction disorder
Wolff-Parkinson-White syndrome
Cardiomyopathy
Eye
Optic neuropathy
Ophthalmoplegia
Retinopathy
Liver
Hepatopathy
Skeletal muscle
Weakness
Fatigue
Myopathy
Neuropathy
ATP
Kidney
Fanconi's syndrome
Glomerulopathy
Nuclear DNA
Subunits
Oxidative phosphorylation
Brain
Seizures
Myoclonus
Ataxia
Stroke
Dementia
Migraine
Mitochondrial DNA
Pancreas
Diabetes mellitus
Nuclear DNA
Blood
Pearson's syndrome
Colon
Pseudo obstruction
Inner ear
Sensorineural hearing loss

FIGURE 481-1 Dual genetic control and multiple organ system manifestations of mitochondrial disease. *(From DR Johns: Mitochondrial DNA and disease. N Engl J Med 333:638, 1995. Copyright © 1995, Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

The classic physical structure of the mitochondria is that of a thread-like organelle, which under fixed conditions, such as observed with immunohistochemical stains or electron microscopy, has a submarine shape and measures about 1 μm in length. However, in the living state, mitochondria comprise a network, with the mitochondrial shape being highly

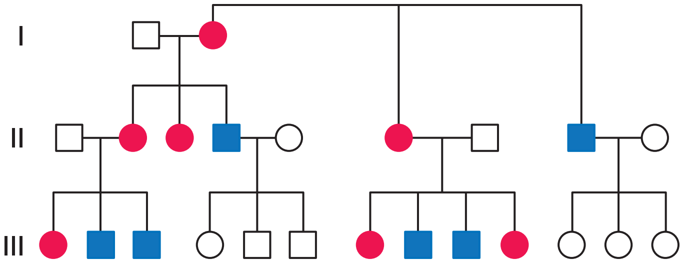

FIGURE 481-2 Maternal inheritance of mitochondrial DNA (mtDNA) disorders and heritable traits. Affected women (*filled circles*) transmit the trait to their children. Affected men (*filled squares*) do not transmit the trait to any of their offspring.

variable based on the cell type, and manifests a complex and ever-changing syncytial form, with continuous appearance and disappearance of budding structures (representing mitochondrial fission) and reorganization of separate mitochondria (representing mitochondrial fusion). Although mitochondrial number per cell type appears in the medical literature, this is no longer considered as a reliable expression of actual functional mitochondrial volume or mass.

Although the presence of mitochondria has been known for >150 years, and knowledge of their respiratory function was proposed ~100 years ago, the initial description of an illness linked to mitochondrial dysfunction was only made in 1962. The presence of mtDNA was also only reported in the 1960s, and it was not until 1988 that the first mutations in the mtDNA causing human illness were described. These included the demonstration of a large-scale mtDNA deletion causing Kearns-Sayre syndrome (KSS) and the discovery of a point mutation in *ND4*, an mtDNA-encoded complex I gene, causing Leber's hereditary optic neuropathy (LHON). Following these two discoveries, >400 pathogenic mtDNA mutations or deletions have been reported.

MITOCHONDRIAL DNA STRUCTURE AND FUNCTION

As a result of its circular structure and extranuclear location, the replication and transcription mechanisms of mtDNA differ from the corresponding mechanisms in the nuclear genome, whose nucleosomal packaging and structure are more complex. Specifically, mitochondria have their own transcription system, and the mtDNA itself replicates independently of cellular replication. Because each cell contains many copies of mtDNA and because the number of mitochondria can vary during the lifetime of each cell, mtDNA copy number is not directly coordinated with the cell cycle. Thus, vast differences in mtDNA copy number are observed between different cell types and tissues and during the lifetime of a cell. Another important feature of the mtDNA replication process is a reduced stringency of proofreading and replication error correction, leading to a greater degree of sequence variation compared to the nuclear genome. Some of these sequence variants are silent polymorphisms that do not have the potential for a phenotypic or pathogenic effect, whereas others may be considered pathogenic mutations. There are some mutations that may be considered ecogenetic, as they typically remain silent, meaning they do not cause disease, unless an external event occurs. One classic example is seen in a common (1:800) mutation in the mitochondrial 12S rRNA gene, m.A1555G, which is associated with hearing loss that is rapidly exacerbated by exposure to normal dosages of an aminoglycoside antibiotic. Because mtDNA replication is independent of cellular replication, the percentage of mutant mtDNA copies tend to increase with age, especially in cells that are terminally differentiated (nonreplicative) at birth such as neurons and myocytes, which may explain some features of mitochondrial dysfunction with aging.

With respect to transcription, initiation can occur on both strands and proceeds through the production of an intronless polycistronic precursor RNA, which is then processed to produce the 13 individual mRNA and 24 individual tRNA and rRNA products. The 37 mtDNA genes comprise fully 93% of the 16,569 nucleotides of the mtDNA in what is known as the *coding region*. The *control region*, which is contained in the D-loop, consists of ~1.1 kilobases (kb) of noncoding DNA and is thought to have an important role in replication and transcription initiation.

■ MATERNAL INHERITANCE AND LACK OF RECOMBINATION

In contrast to homologous pair recombination that takes place in the nucleus, mtDNA molecules do not undergo recombination, such that mutational events represent the only source of mtDNA genetic diversification. Moreover, it is only the maternal DNA that is transmitted to the offspring. The fertilized oocyte degrades the paternal mitochondria involving the ubiquitin proteasome system and autophagy that takes place on the inner membrane of the oocyte. However, additional studies suggest that human spermatozoa do not contain intact mtDNA and are missing TFAM, the mitochondrial transcription factor necessary for mtDNA transcription. Thus, although mothers transmit their mtDNA to both their sons and daughters, only the daughters can transmit the inherited mtDNA to future generations. Accordingly, mtDNA sequence variation and associated phenotypic traits and diseases are inherited exclusively along maternal lines, meaning both sons and daughters have equal chances of having symptomatic disease, with the only significant exception being LHON, as described below.

The phenotypic expression, including age of onset and the specific pattern and severity of organ dysfunction, of a pathogenic mtDNA mutation may vary greatly, even within families. Because of this complex relationship between mtDNA mutations and disease expression, sometimes it is difficult to recognize the maternal pattern of inheritance at the clinical or pedigree level. However, evidence of paternal transmission can almost certainly exclude an mtDNA genetic origin of phenotypic variation or disease; conversely, a disease affecting both sexes without evidence of paternal transmission strongly suggests a heritable mtDNA disorder (Fig. 481-2).

■ MULTIPLE COPY NUMBER (POLYPLOIDY), HIGH MUTATION RATE, HETEROPLASMY, AND MITOTIC SEGREGATION

Each aerobic cell in the body has multiple mitochondria, often numbering many hundreds or more in cells with extensive energy production requirements. Furthermore, the number of copies of mtDNA within each mitochondrion varies from several to hundreds; this is true of both somatic as well as germ cells, including oocytes in females. In the case of somatic cells, this means that the impact of most newly acquired somatic mtDNA mutations is likely to be very small in terms of total cellular or organ system function; however, because of the manyfold higher mutation rate during mtDNA replication, numerous different mutations may accumulate with aging of the organism. It has been proposed that the total cumulative burden of acquired somatic mtDNA mutations with age may result in an overall perturbation of mitochondrial function, contributing to age-related reduction in the efficiency of oxidative phosphorylation and increased production of damaging ROS. Because certain mtDNA (and nDNA) mutations may result in electron leak within the ETC, the ROS damage may rise to a level causing increased susceptibility to somatic mtDNA damage and disease expression. The accumulation of such acquired somatic mtDNA mutations with aging may contribute to age-related diseases, such as metabolic syndrome and diabetes, cancer, and neurodegenerative and cardiovascular disease in any given individual. However, somatic mutations are not carried forward to the next generation, and the hereditary impact of mtDNA mutagenesis requires separate consideration in the female germline.

The multiple mtDNA copy number within each cell, including the maternal germ cells, results in the phenomenon of heteroplasmy, in contrast to the much greater uniformity (homoplasy) of somatic nuclear DNA sequence. Heteroplasmy for a given mtDNA sequence variant or mutation arises as a result of the coexistence within a cell, tissue, or individual of mtDNA molecules bearing more than one version of the sequence variant **(Fig. 481-3)**. The importance of the heteroplasmy phenomena to the understanding of mtDNA-related

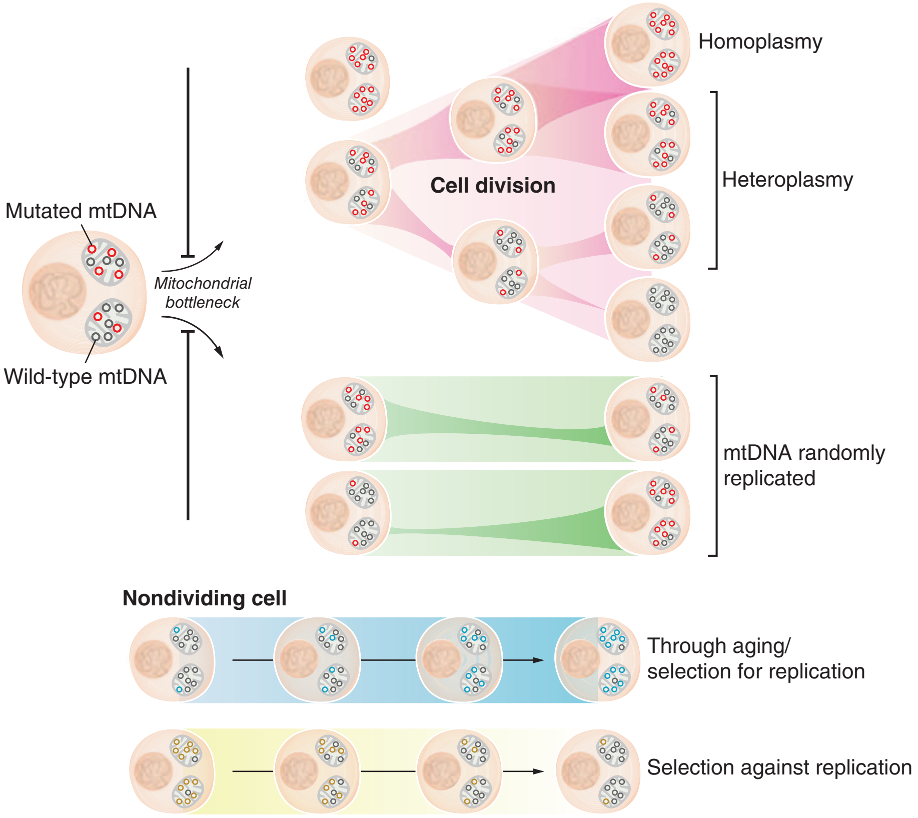

FIGURE 481-3 mtDNA genetic bottleneck and changes of heteroplasmy level throughout the lifetime. Each oocyte can inherit a different proportion of mutated mtDNA molecules from maternal mitochondria. When cells divide (shown in *pink*), heteroplasmy levels in each daughter cell can either increase, decrease, or stay approximately the same. Once inherited, mtDNA mutations can continuously "clonally expand" throughout life, even in nondividing cells (shown in *green*, *blue*, and *yellow*). If one genotype is copied more frequently than another, it will change the overall proportion of different genotypes within the cell over time. The direction of this change can be influenced by selection for or against a particular mtDNA variant (shown in *blue* and *yellow*). When a mutated mtDNA molecule has a replicative advantage, the level will increase during life and possibly exceed the biochemical threshold, and thus contribute to the age-related pathologies or the aging process (shown in the *blue box*). *(Reproduced from W Wei, PF Chinnery: Inheritance of mitochondrial DNA in humans: Implications for rare and common diseases. J Intern Med 2020; 287:634.)*

mitochondrial diseases is critical. The coexistence of mutant and nonmutant (wild-type) mtDNA and the variation of the mutant load, which can be thought of as the percentage of mutant mtDNA molecules within a specific cell, tissue, organ, or organism, contribute to the expression of a phenotype among individuals from the same maternal sibship. At the level of the oocyte, the percentage of mtDNA molecules bearing each version of the polymorphic sequence variant or mutation depends on stochastic events related to partitioning of mtDNA molecules during the process of oogenesis itself. Thus, oocytes differ from each other in the degree of heteroplasmy for that sequence variant or mutation. In turn, the heteroplasmic state is carried forward to the zygote and to the organism as a whole, to varying degrees, depending on mitotic segregation of mtDNA molecules during organ system development and maintenance. For this reason, in vitro fertilization, followed by preimplantation genetic diagnosis (PGD), is not as predictive of the genetic health of the offspring in the case of mtDNA mutations as in the case of mutations and subsequent diseases occurring in the nuclear genome. Similarly, the impact of somatic mtDNA mutations acquired during development and subsequently also shows a wide spectrum of variability. In general, a higher mutant load will result in a more severe and earlier phenotypic presentation. However, measuring heteroplasmy in one tissue (lymphocytes from blood or urine sediment containing kidney and bladder epithelial cells, for example) may not represent the percentage of mutant heteroplasmy in the tissue or organs most affected, such as the cardiac atrioventricular node or brain. Furthermore, the threshold of mutant heteroplasmy that results in clinical illness may vary depending on the specific mutation.

Mitotic segregation refers to the unequal distribution of wild-type and mutant versions of mtDNA molecules during all cell divisions that occur during prenatal development and subsequently throughout the lifetime of an individual. The phenotypic effect or disease impact will be a function not only of the inherent disruptive effect (pathogenicity) on the mtDNA-encoded gene (coding region mutations) or integrity of the mtDNA molecule (control region mutations) but also of its distribution among the multiple copies of mtDNA in the various mitochondria, cells, and tissues of the affected individual. Thus, one consequence can be the generation of a bottleneck due to the marked decline in given sets of mtDNA variants, pathogenic and nonpathogenic, consequent to such mitotic segregation. It is postulated that the main effects of this bottleneck occur between the primordial germ cell state and the primary oocyte stage of development. Heterogeneity arises from differences in the degree of heteroplasmy among oocytes of the transmitting female, together with subsequent, probably random, mitotic segregation of the pathogenic mutation during tissue and organ development and throughout the lifetime of the individual offspring. The actual expression of disease is believed to primarily depend on a threshold percentage of mitochondria whose function is disrupted by mtDNA mutations. This in turn confounds hereditary transmission patterns and hence genetic diagnosis of pathogenic heteroplasmic mutations. Generally, if the proportion of mutant mtDNA is <60%, the individual is unlikely to be affected, whereas proportions exceeding 90% likely result in clinical disease. One notable exception is LHON, in which these mutations are present either in 100% mutant homoplasmy, which causes the disease expression, or 100% wild-type homoplasmy. It is not understood why this specific phenotype and the several known mtDNA alleles that result in LHON behave in this manner.

HOMOPLASMIC VARIANTS AND HUMAN mtDNA PHYLOGENY

In contrast to classic mtDNA diseases, most of which have clinical onset during childhood and are the result of heteroplasmic mutations as noted above, during the course of human evolution, certain mtDNA sequence variants have drifted to a state of homoplasmy, wherein all of the mtDNA molecules in the organism contain the new sequence variant. This arises due to a "bottleneck" effect followed by genetic drift during the very process of oogenesis itself (Fig. 481-3). In other words, during certain stages of oogenesis, the mtDNA copy number becomes so substantially reduced that the particular mtDNA species bearing the novel or derived sequence variant may become the increasingly predominant, and eventually exclusive, version of the mtDNA for that particular nucleotide site. All of the offspring of a woman bearing an mtDNA sequence variant or mutation that has become homoplasmic will also be homoplasmic for that variant and will transmit the sequence variant forward in subsequent generations.

Considerations of reproductive fitness limit the evolutionary or population emergence of pathogenic homoplasmic mutations that are lethal or cause severe disease in infancy or childhood. Thus, with a number of notable exceptions (e.g., as noted mtDNA mutations causing LHON; and see below), most homoplasmic mutations are considered to be neutral markers of human evolution, which are useful and interesting in the population genetics analysis of shared maternal ancestry but have little significance in human phenotypic variation or disease predisposition.

More important is the understanding that this accumulation of homoplasmic mutations occurs at a genetic locus that is transmitted only through the female germline and that lacks recombination. In turn, this enables reconstruction of the sequential topology and radiating phylogeny of mutations accumulated through the course of human evolution since the time of the most recent common mtDNA ancestor of all contemporary mtDNA sequences, some 200,000 years ago. The term *haplogroup* is usually used to define major branching points in the human mtDNA phylogeny, nested one within the other, which often demonstrate striking continental geographic ancestral partitioning. At the level of the complete mtDNA sequence, the term *haplotype* is usually used to describe the sum of mutations observed for a given mtDNA sequence and as compared to a reference sequence, such that all haplotypes falling within a given haplogroup share the total sum of mutations that have accumulated since the most recent common ancestor and the bifurcation point they mark. The remaining observed variants are private to each haplotype. Consequentially, the human mtDNA sequence serves with high fidelity as a molecular prototype for a nonrecombining locus, and its variation has been extensively used in phylogenetic studies. Moreover, the mtDNA mutation rate is considerably higher than the rate observed for the nuclear genome, especially in the control region, which contains the displacement loop, or D-loop, in turn comprising two adjacent hypervariable regions (HVR-I and HVR-II). Together with the absence of recombination, this amplifies drift to high frequencies of novel haplotypes that are highly partitioned across geographically defined populations. Despite extensive research, it has not been well established that such haplotype-based partitioning has a significant influence on human health conditions. However, mtDNA-based phylogenetic analysis can be used both as a quality assurance tool and as a filter in distinguishing neutral mtDNA variants comprising human mtDNA phylogeny from potentially deleterious mutations.

MITOCHONDRIAL DNA DISEASE

The true prevalence of mtDNA disease is difficult to estimate because of the phenotypic heterogeneity that occurs as a function of heteroplasmy, the challenge of detecting and assessing heteroplasmy in different affected tissues, and the other unique features of mtDNA function and inheritance described above. It is estimated that at least 1 in 200 healthy humans harbors a pathogenic mtDNA mutation with the potential to causes disease but that heteroplasmic germline pathogenic mtDNA mutations actually result in clinical disease in ~1 in 5000 individuals.

The true disease burden relating to mtDNA sequence variation will only be known when the following capabilities become available: (1) ability to distinguish a completely neutral sequence variant from a true phenotype-modifying or pathogenic mutation, (2) accurate assessment of heteroplasmy that can be determined with high fidelity, and (3) a systems biology approach **(Chap. 499)** to determine the network of epistatic interactions of mtDNA sequence variations with mutations in the nuclear genome.

OVERVIEW OF CLINICAL AND PATHOLOGIC FEATURES OF HUMAN mtDNA DISEASE

Given the vital roles of mitochondria in all nucleated cells, it is not surprising that mtDNA mutations can affect numerous tissues with pleiotropic effects. More than 200 different disease-causing, mostly heteroplasmic mtDNA mutations have been described affecting ETC function. **Figure 481-4** provides a partial mtDNA map of some of the better characterized of these disorders. A number of clinical clues can increase the index of suspicion for a heteroplasmic mtDNA mutation as an etiology of a heritable trait or disease, including (1) familial clustering with absence of paternal transmission; (2) adherence to one of the classic syndromes (see below) or paradigmatic combinations of disease phenotypes involving several organ systems that normally do not fit together within a single nuclear genomic mutation category; (3) a complex of laboratory and pathologic abnormalities that reflect disruption in cellular energetics (e.g., lactic acidosis and neurodegenerative and myodegenerative symptoms with the finding of ragged red fibers, reflecting the accumulation of abnormal mitochondria under the muscle sarcolemmal membrane); or (4) a mosaic pattern reflecting a heteroplasmic state. There are no truly sensitive and specific biomarkers of disease, and the presence of a historically quintessential finding of ragged red fibers can be seen in numerous muscle disorders, so laboratory tests must always be interpreted in the context of their limitations and should not be used to define the disease. Because of the improved availability and decreasing cost for mtDNA sequencing, the presence or absence of a pathogenic mtDNA mutation can be diagnostic when the clinical phenotype and family history are suggestive.

Heteroplasmy can sometimes be elegantly demonstrated at the tissue level using histochemical staining for enzymes in the oxidative phosphorylation pathway, with a mosaic pattern indicating heterogeneity of the genotype for the coding region for the mtDNA-encoded enzyme.

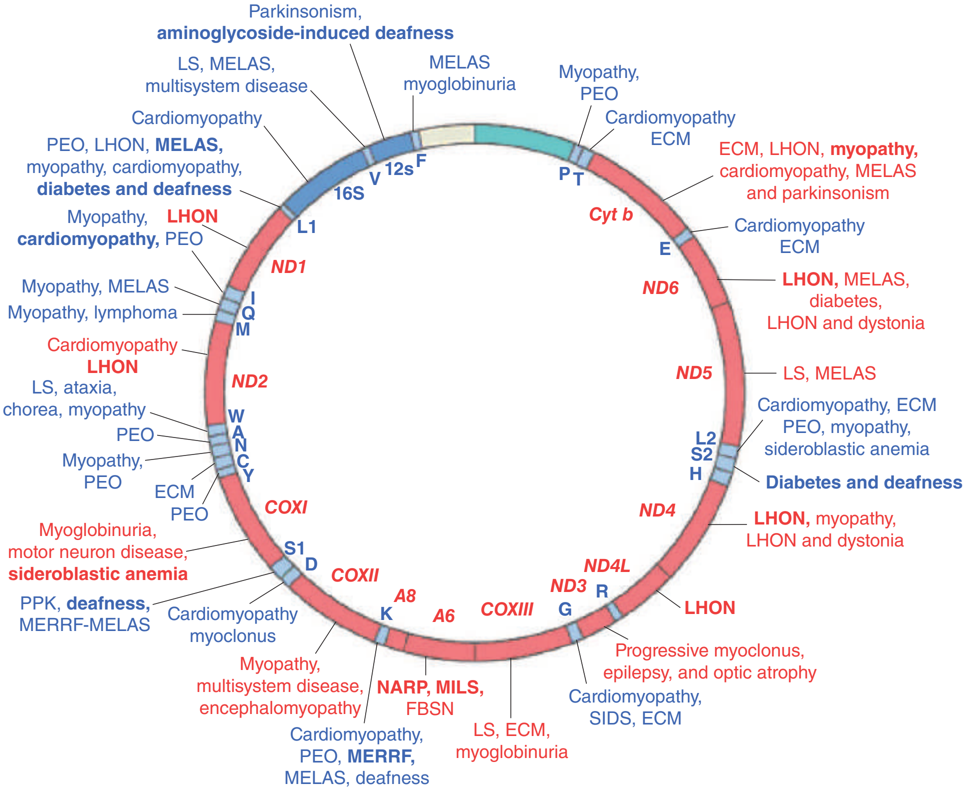
Parkinsonism,
aminoglycoside-induced deafness
LS, MELAS,
multisystem disease
MELAS
myoglobinuria
Myopathy,
PEO
Cardiomyopathy
ECM
Cardiomyopathy
PEO, LHON, MELAS,
myopathy, cardiomyopathy,
diabetes and deafness
ECM, LHON, myopathy,
cardiomyopathy, MELAS
and parkinsonism
12s
F
V
16S
P
T
Cyt b
L1
Myopathy, LHON
cardiomyopathy, PEO
Cardiomyopathy
ECM
E
ND1
ND6
LHON, MELAS,
diabetes,
LHON and dystonia
Myopathy, MELAS
Myopathy, lymphoma
I
Q
M
Cardiomyopathy
LHON
ND2
ND5
LS, MELAS
LS, ataxia,
chorea, myopathy
Cardiomyopathy, ECM
PEO, myopathy,
sideroblastic anemia
PEO
W
A
N
C
Y
L2
S2
H
Myopathy,
PEO
Diabetes and deafness
ECM
PEO
COXI
ND4
LHON, myopathy,
LHON and dystonia
Myoglobinuria,
motor neuron disease,
sideroblastic anemia
S1
D
COXII
ND4L
ND3
R
PPK, deafness,
MERRF-MELAS
Cardiomyopathy
myoclonus
A8
K
A6
COXIII
G
LHON
Progressive myoclonus,
epilepsy, and optic atrophy
Myopathy,
multisystem disease,
encephalomyopathy
NARP, MILS,
FBSN
Cardiomyopathy,
SIDS, ECM
LS, ECM,
myoglobinuria
Cardiomyopathy,
PEO, MERRF,
MELAS, deafness

FIGURE 481-4 Mutations in the human mitochondrial genome known to cause disease. Disorders that are frequently or prominently associated with mutations in a particular gene are shown in *boldface*. Diseases due to mutations that impair mitochondrial protein synthesis are shown in *blue*. Diseases due to mutations in protein-coding genes are shown in *red*. ECM, encephalomyopathy; FBSN, familial bilateral striatal necrosis; LHON, Leber's hereditary optic neuropathy; LS, Leigh's syndrome; MELAS, mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes; MERRF, myoclonic epilepsy with ragged red fibers; MILS, maternally inherited Leigh's syndrome; NARP, neuropathy, ataxia, and retinitis pigmentosa; PEO, progressive external ophthalmoplegia; PPK, palmoplantar keratoderma; SIDS, sudden infant death syndrome. *(From S DiMauro, E Schon: Mitochondrial respiratory-chain diseases. N Engl J Med 348:2656, 2003. Copyright © 2003, Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

Complex II, CoQ, and cytochrome c are exclusively encoded by nuclear DNA. In contrast, complexes I, III, IV, and V contain at least some subunits encoded by mtDNA. Just 3 of the 13 subunits of the ETC complex IV enzyme, cytochrome c oxidase (COX), are encoded by mtDNA, and therefore, this enzyme has the lowest threshold for dysfunction when a threshold level of mutated mtDNA is reached. Histochemical staining for COX activity in tissues of patients affected with heteroplasmic inherited mtDNA mutations (or with the somatic accumulation of mtDNA mutations, see below) can show a mosaic pattern of reduced histochemical staining in comparison with histochemical staining for the complex II enzyme succinate dehydrogenase (SDH) **(Fig. 481-5)**.

Next-generation sequencing (NGS) has dramatically improved the clinical genetic diagnostic evaluation of mitochondrial diseases at the level of both the nuclear genome and mtDNA. Low sequencing costs, high throughput, and short turnaround time expedite whole exome (WES) or whole genome sequencing (WGS) to identity genes and mutations with known pathogenicity or based on bioinformatics assessment of likely pathogenicity. In the context of the mtDNA, the deep coverage enabled by NGS compared to Sanger sequencing now provides rapid and reliable detection of heteroplasmy in different affected tissues. NGS yields accurate information about a patient's predominant mtDNA sequence as well as lower frequency heteroplasmic variants and can reliably reach detection of even single mutant nucleotide heteroplasmy down to levels of <10%. Lower levels are often only clinically relevant if in the setting of a striking difference in heteroplasmy in different tissues.

Clinically, the most striking overall characteristic of mitochondrial genetic disease is the phenotypic heterogeneity associated with mtDNA mutations. This extends to intrafamilial phenotypic heterogeneity for the same mtDNA pathogenic mutation and, conversely, to the overlap of phenotypic disease manifestations with distinct mutations. Thus, although fairly consistent and well-defined "classic" syndromes have been attributed to specific mutations, frequently "nonclassical" combinations of disease phenotypes ranging from isolated myopathy to extensive multisystem disease are often encountered, rendering genotype-phenotype correlation challenging. In both classical and nonclassical mtDNA disorders, there is often a clustering of some combination of abnormalities affecting the neurologic system (including optic nerve atrophy, pigment retinopathy, and sensorineural hearing loss), cardiac and skeletal muscle (including extraocular muscles), and endocrine and metabolic systems (including diabetes mellitus). Additional organ systems that may be affected include the hematopoietic, renal, hepatic, and gastrointestinal systems, although these are more frequently involved in infants and children. Disease-causing mtDNA coding region mutations can affect either one of the 13 protein-encoding genes or one of the 24 protein synthetic genes. Clinical manifestations do not readily distinguish these two categories, although lactic acidosis and specific muscle pathologic findings (e.g., ragged red and ragged blue fibers, immunohistochemical staining, paracrystalline inclusions on ultrastructure) tend to be more prominent in the latter. In all cases, either defective ATP production due to disturbances in the ETC or enhanced generation of ROS has been invoked as the mediating biochemical mechanism between mtDNA mutation and disease manifestation.

■ mtDNA DISEASE PRESENTATIONS

The clinical presentation of adult patients with mtDNA disease can be divided into three categories: (1) clinical features suggestive of mitochondrial disease **(Table 481-2)** but not a well-defined classic syndrome; (2) classic mtDNA syndromes; and (3) clinical presentation confined to one organ system (e.g., isolated sensorineural deafness, cardiomyopathy, or diabetes mellitus). It is important to note, especially when young adults come to medical attention, that symptoms of an mtDNA disorder may have begun during childhood.

Table 481-3 provides a summary of eight illustrative classic mtDNA syndromes or disorders that affect adult patients and highlights some of the most interesting features of mtDNA disease in terms of molecular pathogenesis, inheritance, and clinical presentation. The first five of these syndromes result from heritable point mutations in either protein-encoding or protein synthetic mtDNA genes; the other three result from rearrangements or deletions that usually do not involve the germline.

LHON is a common cause of maternally inherited visual failure. LHON typically presents during young adulthood with subacute painless loss of vision in one eye, with symptoms developing in the other eye 6–12 weeks later. In some instances, cerebellar ataxia, peripheral neuropathy, and cardiac conduction defects are observed. In >95% of cases, LHON is due to one of the three homoplasmic point mutations of mtDNA that affect genes encoding different subunits of complex I of the mitochondrial ETC; however, not all individuals who inherit a primary LHON mtDNA mutation develop optic neuropathy, and the male-to-female ratio is 8.2, indicating that additional environmental (e.g., tobacco exposure) or independent genetic factors are important in the etiology of the disorder. Estrogen may also play a role in the decreased clinical penetrance in women. Both the nuclear and mitochondrial genomic backgrounds modify disease penetrance. Indeed, a region of the X chromosome containing a high-risk haplotype for LHON has been identified, supporting the formulation that nuclear genes act as modifiers and affording an explanation for the male prevalence of LHON. This haplotype can be used in predictive genomic testing and prenatal screening for this disease. In contrast to the other classic mtDNA disorders, it is of interest that

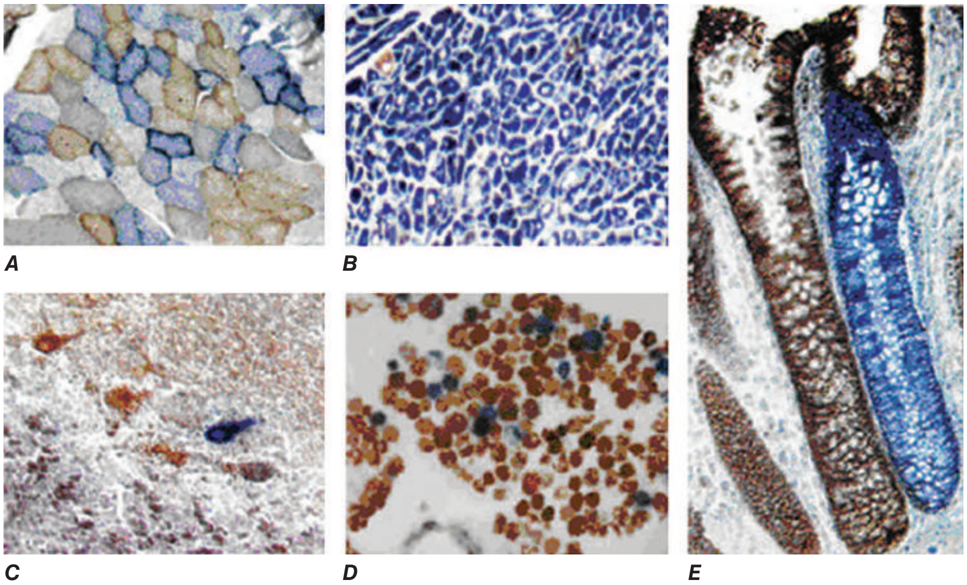

FIGURE 481-5 Cytochrome c oxidase (COX) deficiency in mitochondrial DNA (mtDNA)–associated disease. Transverse tissue sections that have been stained for COX and succinate dehydrogenase (SDH) activities sequentially, with COX-positive cells shown in *brown* and COX-deficient cells shown in *blue*. ***A.*** Skeletal muscle from a patient with a heteroplasmic mitochondrial tRNA point mutation. The section shows a typical "mosaic" pattern of COX activity, with many muscle fibers harboring levels of mutated mtDNA that are above the crucial threshold to produce a functional enzyme complex. ***B.*** Cardiac tissue (left ventricle) from a patient with a homoplasmic tRNA mutation that causes hypertrophic cardiomyopathy, which demonstrates an absence of COX in most cells. ***C.*** A section of cerebellum from a patient with mtDNA rearrangement that highlights the presence of COX-deficient neurons. ***D, E.*** Tissues that show COX deficiency due to clonal expansion of somatic mtDNA mutations within single cells—a phenomenon that is seen in both postmitotic cells (***D***; extraocular muscles) and rapidly dividing cells (***E***; colonic crypt) in aging humans. *(Reproduced with permission from R Taylor, D Turnbull: Mitochondrial DNA mutations in human disease. Nat Rev Genetics 6:389, 2005.)*

TABLE 481-2 Common Features of Mitochondrial DNA–Associated Diseases in Adults

Neurologic: stroke, epilepsy, migraine headache, peripheral neuropathy, ataxia, dystonia, myoclonus, cranial neuropathy (optic atrophy, sensorineural deafness, dysphagia, dysphasia)

Skeletal myopathy: ophthalmoplegia, exercise intolerance, myalgia, weakness

Cardiac: conduction block, cardiomyopathy

Respiratory: hypoventilation, aspiration pneumonitis

Endocrine: diabetes mellitus, premature ovarian failure, hypothyroidism, hypoparathyroidism

Ophthalmologic: cataracts, pigment retinopathy, neurologic and myopathic (optic atrophy, ophthalmoplegia)

patients with this syndrome are often homoplasmic for the disease-causing mutation. The somewhat later onset in young adulthood and modifying effect of protective background nuclear genomic haplotypes may have enabled homoplasmic pathogenic mutations to have escaped evolutionary censoring.

Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS) is a multisystem disorder with a typical onset between 2 and 10 years of age, although adult presentations also occur. Following normal early psychomotor development, the most common initial symptoms are seizures, recurrent headaches, anorexia, and recurrent vomiting. Exercise intolerance or proximal limb weakness can be the initial manifestation, followed by generalized tonic-clonic seizures. Short stature is common. Seizures are often associated with stroke-like episodes of transient hemiparesis or cortical blindness that may produce recurrent encephalopathy with impaired consciousness. It is often not possible to determine if the encephalopathy is due to refractory clinical or subclinical seizures or should be attributed to an independent effect. The cumulative residual effects of the stroke-like episodes gradually impair motor abilities, vision, and cognition, often by adolescence or young adulthood. Sensorineural hearing loss adds to the progressive decline of these individuals. A plethora of less common symptoms have been described including myoclonus, ataxia, episodic coma, optic atrophy, cardiomyopathy, pigmentary retinopathy, ophthalmoplegia, diabetes mellitus, hirsutism, gastrointestinal dysmotility, and nephropathy. The typical age of death ranges from 10 to 35 years, but some individuals live into their sixth decade. Intercurrent infections or intestinal obstructions are often the terminal events. It is proposed that the clinical diagnosis of MELAS can only be applied if the following three criteria are met: (1) stroke-like episode before age 40 years, (2) encephalopathy due to seizures and/or dementia, and (3) lactic acidosis and/or ragged red fibers. It is not atypical for some family members to have much less severe or later onset illness, presumably because of a lessor mutation load, and "MELAS" is not used as a diagnosis for these restricted phenotypes. This creates somewhat of a disconnect between the genotype for MELAS (most commonly the m.3243A>G mutation) and a diverse phenotype, which includes the syndrome MELAS, a syndrome of high-frequency hearing loss and diabetes with onset later in life, as well as many other phenotypes between these two extreme syndromes. Certain other mtDNA mutations can also cause such patterns of diverse phenotypic expression. Laboratory investigation commonly demonstrates elevated blood lactate concentrations at rest with excessive increase after moderate exercise. Magnetic resonance imaging (MRI) of the brain shows areas of involvement on T2- or fluid-attenuated inversion recovery (FLAIR) sequences, with decreased signal on perfusion-weighted sequences, which typically involve the posterior cerebrum and do not conform to the distribution of major arteries. These abnormalities may be temporary or evolve to subsequent atrophy **(Fig. 481-6)**. Electrocardiography (ECG) may show evidence of cardiomyopathy, preexcitation, or incomplete heart block. Electromyography and nerve conduction studies are consistent with a myopathic process, without or with coexisting axonal and sensory neuropathic findings. Muscle biopsy typically shows ragged red fibers with the modified Gomori trichrome stain or "ragged blue fibers" with the SDH histochemical stain, resulting from the hyperintense reaction. The diagnosis of MELAS is based on a combination of clinical findings and molecular genetic testing. Mutations in the mtDNA gene *MT-TL1* encoding tRNA^{leu} are causative. The most common mutation, present in ~80% of individuals with typical clinical findings, is an A-to-G transition at nucleotide 3243 (m.3243A>G). Mutations can usually be detected in mtDNA from leukocytes in individuals with typical MELAS; however, the occurrence of heteroplasmy can result in varying tissue distribution of mutated mtDNA. In the absence of specific treatment, various manifestations of MELAS are treated according to standard modalities for prevention, surveillance, and treatment. Recent developments in therapy are described below.

Myoclonus epilepsy with ragged red fiber (MERRF) is a multisystem disorder characterized by myoclonus, seizures, ataxia, and myopathy with ragged red fibers. Hearing loss, exercise intolerance, neuropathy, ataxia, cervical lipomas, and short stature are often present. Ataxia and lipomas can be a feature in adults or adult-onset MERRF. Cerebrospinal fluid (CSF) analysis reveals an elevated protein content. Almost all MERRF patients have a mutation in the mtDNA tRNA^{lys} gene, and the m.8344A>G mutation in the mtDNA gene encoding the lysine amino acid tRNA is responsible for 80–90% of MERRF cases.

Neuropathy, ataxia, and retinitis pigmentosa (NARP) is characterized by moderate diffuse cerebral and cerebellar atrophy and symmetric lesions of the basal ganglia on MRI **(Figs. 481-7 and 481-8)**. A heteroplasmic m.8993T>G mutation in the ATPase 6 subunit gene

TABLE 481-3 Mitochondrial Diseases Due to Mitochondrial DNA (mtDNA) Point Mutations and Large-Scale Rearrangements

DISEASE	PHENOTYPE	MOST FREQUENT mtDNA MUTATIONS	HETEROPLASMIC/ HOMOPLASMIC	MATERNAL
NARP, Leigh's syndrome	Loss of central vision leading to blindness in young adult life	m.1778G>A, m.14484T>C, m.3460G>A	Heteroplasmic	Maternal
MELAS	*M*itochondrial *e*ncephalomyopathy, *l*actic *a*cidosis, and *s*troke-like episodes; may manifest only as diabetes mellitus	Point mutation in tRNA^{leu}	Heteroplasmic	Maternal
MERRF	*M*yoclonic *e*pilepsy, *r*agged *r*ed *f*ibers in muscle, ataxia, increased CSF protein, sensorineural deafness, dementia	Point mutation in tRNA^{lys}	Heteroplasmic	Maternal
Deafness	Progressive sensorineural deafness, often induced by aminoglycoside antibiotics	m.1555A>G mutation in 12S rRNA	Homoplasmic	Maternal
	Nonsyndromic sensorineural deafness	m.7445A>G mutation in 12S rRNA	Homoplasmic	Maternal
Chronic progressive external ophthalmoplegia (PEO)	Late-onset bilateral ptosis and ophthalmoplegia, proximal muscle weakness, and exercise intolerance	Single deletions or duplications	Heteroplasmic	Mostly sporadic, somatic mutations
Pearson's syndrome	Pancreatic insufficiency, pancytopenia, lactic acidosis	Large deletion	Heteroplasmic	Sporadic, somatic mutations
Kearns-Sayre syndrome (KSS)	External ophthalmoplegia, heart block, retinal pigmentation, ataxia	The 5-kb "common deletion"	Heteroplasmic	Sporadic, somatic mutations

Abbreviations: CSF, cerebrospinal fluid; NARP, neuropathy, ataxia, and retinitis pigmentosa.

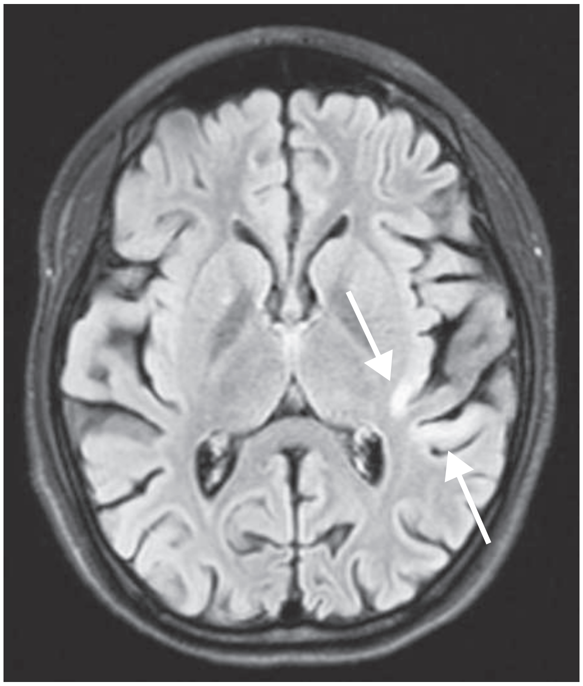

FIGURE 481-6 A 15-year-old girl with MELAS (mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes) due to m.A3243G ($\text{tRNA}^{\text{Leu(UUR)}}$), 85% mutant heteroplasmy, presenting at age 5 with focal motor seizures, ataxia, and short stature, with episodes of acute language and motor dysfunction and progress cognitive impairment. The fluid-attenuated inversion recovery (FLAIR) magnetic resonance image (MRI) shows increased signal intensity (*white arrows*) in the left temporal-parietal region in addition to global mild volume loss (increased extra-axial cerebrospinal fluid spaces).

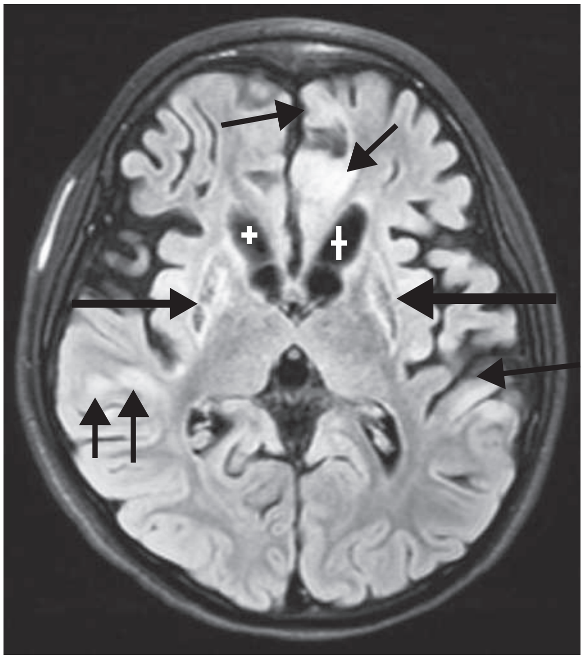

FIGURE 481-8 A 12-year-old boy with Leigh's syndrome due to m.T10191C (ND3 gene, complex I), heteroplasmy percentage not determined, presenting with infantile spasms at 8 months of life. He responded well to adrenocorticotropic hormone (ACTH), and his magnetic resonance imaging (MRI) and development were normal until 30 months when he developed dystonia and progressive medically intractable epilepsy. The fluid-attenuated inversion recover (FLAIR) MRI at 6 years of life shows global atrophy with large extra-axial cerebrospinal fluid spaces, increased signal intensity in the cortex (*thin arrows*), necrotic bilaterally symmetric lesions in the putamina, and enlarged lateral ventricles due to loss of bilateral caudate nuclei volume (*stars*).

has been identified as causative, which underscores the lack of definitive genotype-phenotype correlation in mtDNA diseases. Ragged red fibers are not observed in muscle biopsy. When >95% of mtDNA molecules are mutant, a more severe clinical, neuroradiologic, and neuropathologic picture (Leigh's syndrome) emerges. Not uncommonly, an infant is diagnosed with Leigh's syndrome due to the m.8993T>G mutation and not until several years later will the mother present with symptoms of NARP, a situation that highlights the concept of a higher threshold for lower levels of tissue heteroplasmy.

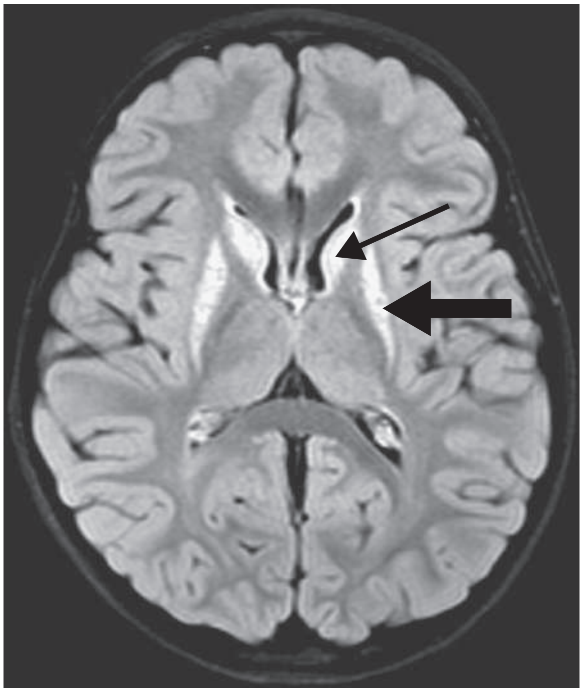

FIGURE 481-7 A 9-year-old girl with Leigh's syndrome due to m.T8993G (ATPase subunit 6), 99% heteroplasmy, presenting at age 14 months with a motor delay and who underwent magnetic resonance imaging (MRI) at 24 months, at which time she had just begun to walk. She has moderate cognitive impairment, arm chorea, and distal leg dystonia. The fluid-attenuated inversion recover (FLAIR) MRI shows symmetric bilateral increased signal in the caudate nuclei (*thin arrow*) and putamen (*thick arrow*); only left-sided lesions indicated with *arrows*.

Point mutations in the mtDNA gene encoding the 12S rRNA (m.A1555G) result in heritable nonsyndromic hearing loss. One such mutation causes heritable ototoxic susceptibility to standard dosing of aminoglycoside antibiotics, which opens a pathway for a simple pharmacogenetic test in the appropriate clinical settings. This is an example of an ecogenetic disorder in that most people with this mutation do not develop any symptoms until exposed to an external agent.

KSS, sporadic progressive external ophthalmoplegia (PEO), and Pearson's syndrome are three disease phenotypes caused by large-scale mtDNA rearrangements including partial deletions or partial duplication. The majority of single large-scale rearrangements of mtDNA are thought to result from clonal amplification of a single sporadic mutational event, occurring in the maternal oocyte during early embryonic development. The typical mtDNA deletion specifically involves 4977 nucleotides, lost at identical breakpoints, and accounting for most KSS and PEO of mtDNA deletion origin. Because germline involvement is rare, most cases are sporadic rather than inherited. KSS is characterized by the triad of onset before age 20, chronic PEO, and pigmentary retinopathy. Cerebellar syndrome, heart block, increased CSF protein content, diabetes mellitus, and short stature are also part of the syndrome. Single deletions/duplication can also result in milder phenotypes such as PEO, characterized by late-onset PEO, proximal myopathy, and exercise intolerance. In both KSS and PEO, diabetes mellitus and hearing loss are frequent accompaniments. Pearson's syndrome is characterized by infantile onset of a sideroblastic anemia accompanied by lactic acidosis and failure to thrive caused in part by exocrine pancreatic insufficiency. If the child survives, the manifestations appear phenotypically similar to those of severe KSS with myopathy, PEO, encephalopathy, and cardiomyopathy. Pearson's syndrome is generally caused by large-scale sporadic deletion of several mtDNA genes that differ from the common deletion seen in KSS. Typically, the deletion size is larger in Pearson's syndrome, and located with different breakpoints, than in KSS or PEO, but this is not always the case.

Two important dilemmas in classic mtDNA disease have benefited from recent important research insights. The first relates to the greater involvement of neuronal, muscular, renal, hepatic, and pancreatic manifestations in mtDNA disease in these syndromes. This observation has appropriately been mostly attributed to the high energy utilization of the involved tissues and organ systems and, hence, greater dependency on mitochondrial ETC integrity and health. However, because mutations are stochastic events, mitochondrial mutations should occur in any organ during embryogenesis and development. Recently, additional explanations have been suggested based on studies of the common m.3243A>G transition. The proportion of this mutation in peripheral blood cells was shown to decrease exponentially with age. A selective process acting at the stem cell level with a strong bias against the mutated form would have its greatest effect to reduce the mutant mtDNA only in highly proliferating cells, such as those derived from the hematopoietic system. Tissues and organs carrying pathogenic mtDNA mutations and having a lower cell turnover, such as brain, nerve, or retina, would not benefit from this effect and, thus, would be expected to accumulate mutational load and be the most affected. However, age-related clonal hematopoiesis might mitigate some of this disparity.

The other dilemma arises from the observation that only a subset of mtDNA mutations accounts for the majority of the familial mtDNA diseases. The random occurrence of mutations in the mtDNA sequence should yield a more uniform distribution of disease-causing mutations. However, recent studies using the introduction of one severe and one mild point mutation into the female germline of experimental animals demonstrated selective elimination during oogenesis of the severe mutation and selective retention of the milder mutation, with the emergence of mitochondrial disease in offspring after multiple generations. Thus, oogenesis itself can act as an "evolutionary" filter for the most harmful mtDNA disease.

THE INVESTIGATION OF SUSPECTED mtDNA DISEASE

The clinical presentations of classic syndromes, groupings of disease manifestations in multiple organ systems, or unexplained isolated presentations of one of the disease features of a classic mtDNA syndrome should prompt a systematic clinical investigation as outlined in **Fig. 481-9**. However, some tests are not universally available or are costly, and WGS is not only more readily available but also less costly than the cost of obtaining tissue such as muscle for biochemical and pathologic evaluation. In many medical centers, a history and exam suggestive of an mtDNA disorder will result in a molecular genetic evaluation before tissue evaluation. Indeed, mitochondrial disease should be considered in the differential diagnosis of any progressive multisystem disorder. Despite the centrality of disruptive oxidative phosphorylation, an elevated blood lactate level is neither specific nor sensitive, because there are many causes of blood lactic acidosis and many patients with mtDNA defects presenting at any age may have normal blood lactate levels. An elevated CSF lactate is a more specific test for mitochondrial disease if there is central nervous system involvement but is still not diagnostic. The serum creatine kinase may be elevated but is often normal, even in the presence of a proximal myopathy. Recently, testing for elevated levels of growth differentiating factor 15 (GDF15) has shown a high degree of sensitivity and specificity in those with a mitochondrial myopathy, but the degree of elevation for an individual patient reflects the severity of the illness and does not seem to be a sensitive marker of disease activity. Urinary organic acids (specifically TCA cycle intermediates) and amino acids (alanine, proline) may also be abnormal, reflecting metabolic as well as kidney proximal tubule dysfunction. Every patient with seizures, episodes of confusion or atypical behavioral changes, or cognitive decline should have an electroencephalogram. A brain computed tomography

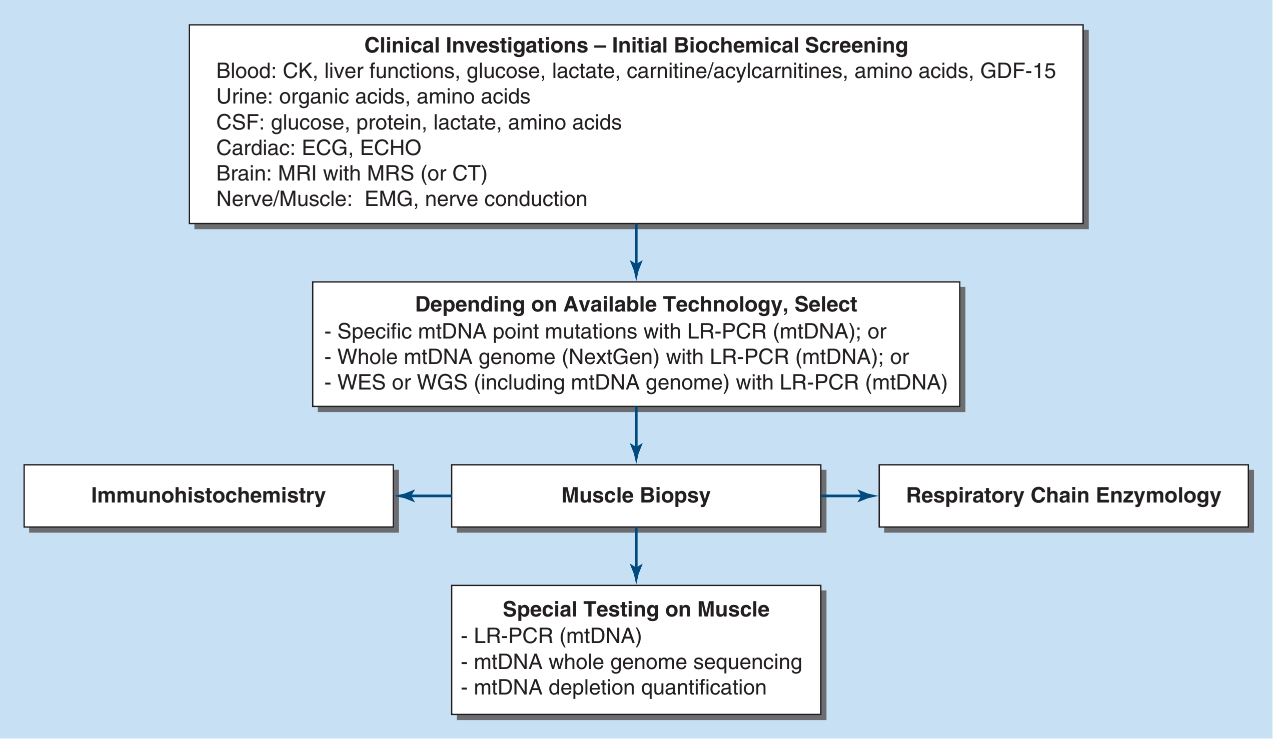
Clinical Investigations – Initial Biochemical Screening
Blood: CK, liver functions, glucose, lactate, carnitine/acylcarnitines, amino acids, GDF-15
Urine: organic acids, amino acids
CSF: glucose, protein, lactate, amino acids
Cardiac: ECG, ECHO
Brain: MRI with MRS (or CT)
Nerve/Muscle: EMG, nerve conduction
Depending on Available Technology, Select
- Specific mtDNA point mutations with LR-PCR (mtDNA); or
- Whole mtDNA genome (NextGen) with LR-PCR (mtDNA); or
- WES or WGS (including mtDNA genome) with LR-PCR (mtDNA)
Immunohistochemistry
Muscle Biopsy
Respiratory Chain Enzymology
Special Testing on Muscle
- LR-PCR (mtDNA)
- mtDNA whole genome sequencing
- mtDNA depletion quantification

FIGURE 481-9 Clinical and laboratory investigation of a suspected mitochondrial DNA (mtDNA) disorder. Following history including the family history and examination, a screening biochemical evaluation and other testing is selectively warranted, and if the evaluation suggests a mitochondrial disease, further genetic evaluation is warranted. The specific molecular genetic testing depends on available technology and costs, with clinical acumen essential for determining the extent of testing. The use of muscle investigation can support the genetic testing. Of note, mtDNA deletion disorders at times require LR-PCR on skeletal muscle tissue to find the deletion. CSF, cerebrospinal fluid; CT, computed tomography; ECG, electrocardiogram; ECHO, echocardiography; EEG, electroencephalogram; EMG, electromyogram; LHON, Leber's hereditary optic neuropathy; LR-PCR, long-range polymerase chain reaction; MELAS, mitochondrial encephalomyopathy, lactic acidosis, and stoke-like episodes; MERRF, myoclonic epilepsy with ragged red fibers; MRI, magnetic resonance imaging; MRS, magnetic resonance spectroscopy, PCR, polymerase chain reaction; RFLP, restriction fragment length polymorphism; WES, whole exome sequencing; WGS, whole genome sequencing.

(CT) scan may show calcified basal ganglia or bilateral hypodense regions with cortical atrophy. MRI is indicated in patients with brainstem signs or stroke-like episodes.

For an increasing number of mitochondrial diseases, it is possible to obtain an accurate diagnosis with a simple molecular genetic screen. For examples, 95% of patients with LHON harbor one of the three mtDNA point mutations (m.11778A>G, m.A3460A>G, or m.14484T>C). These patients have very high levels of mutated mtDNA in peripheral blood cells, and therefore, it is appropriate to send a blood sample for molecular genetic analysis by polymerase chain reaction (PCR) or restriction fragment length polymorphism (RFLP). The same is true for most MERRF patients who harbor a point mutation in the lysine tRNA gene at position 8344. In contrast, patients with the m.3243A>G MELAS mutation often have low levels of mutated mtDNA in blood. If clinical suspicion is strong enough to warrant peripheral blood testing, then patients with a negative result should have testing repeated using a saliva sample or be investigated further by performing a skeletal muscle biopsy to obtain mtDNA from a relatively nonreplicative tissue.

Muscle biopsy histochemical analysis had been the historical cornerstone for investigation of patients with suspected mitochondrial disease. Histochemical analysis may show subsarcolemmal accumulation of mitochondria with the appearance of ragged red fibers, especially in those with mtDNA mutations affecting the tRNA and rRNA genes. Electron microscopy might show abnormal mitochondria with paracrystalline inclusions. Muscle histochemistry may show COX-deficient fibers, which indicate mitochondrial dysfunction (Fig. 481-5). Respiratory chain complex assays may also show reduced enzyme function. If enzymatic or polarographic data are used to aid in the confirmation of diagnosis, a standard method of analysis should be employed. Either of these two abnormalities, within the exact context of established peer-reviewed criteria, may confirm the presence of a mitochondrial disease, to be followed by an in-depth molecular genetic analysis. In most major centers, genetic testing has become the primary means of obtaining a definitive diagnosis, using muscle pathology and biochemistry to assist with interpretation of inconclusive genetic results. It is proposed to use of the term *primary mitochondrial disease* only when a pathogenic mutation is identified that matches the clinical phenotype.

Recent evidence has provided important insights into the importance of nuclear-mtDNA genomic cross-talk and has provided a descriptive framework for classifying and understanding disorders that emanate from perturbations in this cross-talk. Although not strictly considered as mtDNA genetic disorders, manifestations do overlap those highlighted above **(Fig. 481-10)**.

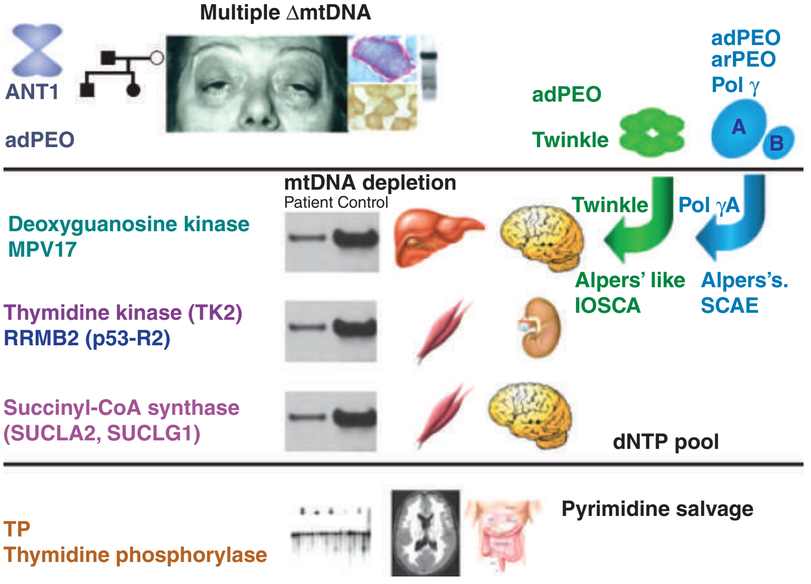

FIGURE 481-10 Disorders associated with perturbations in nuclear-mitochondrial genomic cross-talk. Clinical features and genes associated with multiple mitochondrial DNA (mtDNA) deletions, mtDNA depletion, and mitochondrial neurogastrointestinal encephalomyopathy syndromes. adPEO, autosomal dominant progressive external ophthalmoplegia; ANT, adenine nucleotide translocators; arPEO, autosomal recessive progressive external ophthalmoplegia; IOSCA, infantile-onset spinocerebellar ataxia; SCAE, spinocerebellar ataxia and epilepsy. *(Reproduced with permission from A Spinazzola, M Zeviani: Disorders from perturbations of nuclear-mitochondrial intergenomic cross-talk. J Intern Med 265:174, 2009.)*

IMPACT OF HOMOPLASMIC SEQUENCE VARIATION ON HERITABLE TRAITS AND DISEASE

The relationship among the degree of heteroplasmy, tissue distribution of the mutant mtDNA, and disease phenotype simplifies inference of a clear causative relationship between heteroplasmic mutation and disease. With the exception of certain mutations (e.g., those causing most cases of LHON), drift to homoplasmy of such mutations would be precluded normally by the severity of impaired oxidative phosphorylation and the consequent reduction in reproductive fitness. Therefore, sequence variants that have reached homoplasmy should be neutral in terms of human evolution and, hence, useful for tracing phylogeny, demography, and migration, as described above. Thus, novel homoplasmic variants are seldom pathogenic. One important exception is in the case of one or more of the homoplasmic population-level variants, which designate the mtDNA haplogroup J, and the interaction with the mtDNA mutations causing LHON. Reduced disease predilection suggests that one or more of the ancient sequence variants designating mtDNA haplogroup J appear to attenuate predisposition to degenerative disease, in the presence of other risk factors. Whether or not additional epistatic interactions between population-level mtDNA haplotypes and common health conditions will be found remains to be determined. If such influences do exist, then they are more likely to be relevant to health conditions in the postreproductive age groups, wherein evolutionary filters would not have had the opportunity to censor deleterious effects and interactions and wherein the effects of oxidative stress during aging or with poor diet or lack of exercise may play a role. Although much has been written about the possible associations between population-level common mtDNA variants and human health and disease phenotypes or adaptation to different environmental influences (e.g., climate), clinical implications have not been forthcoming.

Many studies that purport to show such associations with phenotypes such as longevity, athletic performance, and metabolic and neurodegenerative disease are limited by small sample sizes, possible genotyping inaccuracies, and the possibility of population stratification or ethnic ancestry bias. Because mtDNA haplogroups are so prominently partitioned along phylogeographic lines, it is difficult to exclude the possibility that a haplogroup for which an association has been reported is simply a marker for otherwise unappreciated population heterogeneity, wherein a nongenetic (societal or environmental) difference among the populations marked by the mtDNA haplogroup differences is actually causally related to the disease of interest. The experimental difficulty in generating cellular or animal models to test the functional influence of homoplasmic sequence variants (as a result of mtDNA polyploidy) further compounds the challenge. The most likely formulation is that the risk conferred by different mtDNA haplogroup-defining homoplasmic mutations for common diseases depends on the concomitant nuclear genomic background, together with environmental influences. Progress in minimizing potentially misleading associations in mtDNA heritable trait and disease studies should include ensuring adequate sample size taken from a large sample recruitment base, using carefully matched controls and population structure determination, and performing analysis that takes into account epistatic interactions with other genomic loci and environmental factors.

IMPACT OF ACQUIRED SOMATIC mtDNA MUTATION ON HUMAN HEALTH AND DISEASE

Studies on aging humans and animals have shown a potentially important correlation of age with the accumulation of heterogeneous mtDNA mutations, especially in organ systems that undergo the most prominent age-related degenerative tissue phenotype. Sequencing of PCR-amplified single mtDNA molecules has demonstrated an average of two-to-three-point mutations per molecule in elderly subjects when compared with younger ones. Point mutations observed include those responsible for known heritable heteroplasmic mtDNA disorders, such as the m.3344A>G and m.3243A>G mutations responsible for the MERRF and MELAS syndromes, respectively. However, the cumulative burden of these acquired somatic point mutations with age was observed to remain well below the threshold expected for phenotypic expression (<2%). Point mutations at other sites not normally involved in inherited mtDNA disorders have also been shown to accumulate to much higher levels in some tissues of elderly individuals, with the description of tissue-specific "hot spots" for acquired somatic mtDNA point mutations. Likewise, an age-associated and tissue-specific accumulation of mtDNA deletions has been observed, including deletions involved in known heritable mtDNA disorders, as well as others. The accumulation of functional mtDNA deletions in a given tissue is expected to be associated with mitochondrial dysfunction, as reflected in an age-associated patchy and reduced COX activity on histochemical staining, especially in skeletal and cardiac muscle and brain. A particularly well-studied and potentially important example is the accumulation of mtDNA deletions and COX deficiency observed in neurons of the substantia nigra in Parkinson's disease patients.

The progressive accumulation of ROS has been proposed as the key factor connecting mtDNA mutations with aging and age-related disease pathogenesis **(Fig. 481-11)**. As noted above, ROS are a by-product of normal oxidative phosphorylation and are removed by detoxifying antioxidants into less harmful moieties; however, environmental factors or mutations that result in exaggerated production of ROS or impaired removal result in ROS accumulation and subsequent cellular injury. One of the main targets for ROS-mediated injury is DNA, and mtDNA is particularly vulnerable because of its proximity to the origin of free radical production, the lack of protective histones, and less efficient injury repair systems compared with nuclear DNA. In turn, accumulation of mtDNA mutations results in inefficient oxidative phosphorylation, with the potential for excessive production of ROS, generating a "vicious cycle" of cumulative mtDNA damage. Indeed, measurement of the oxidative stress biomarker 8-hydroxy-2-deoxyguanosine has been used to measure age-dependent increases in mtDNA oxidative damage at a rate exceeding that of nuclear DNA. It should be noted that mtDNA mutations can potentially occur in postmitotic cells as well, because mtDNA replication is not synchronized with the cell cycle. Two other proposed links between mtDNA mutation and aging, besides ROS-mediated tissue injury, are the perturbations in efficiency of oxidative phosphorylation with disturbed cellular aerobic function and perturbations in apoptotic pathways, whose execution steps involve mitochondrial activity.

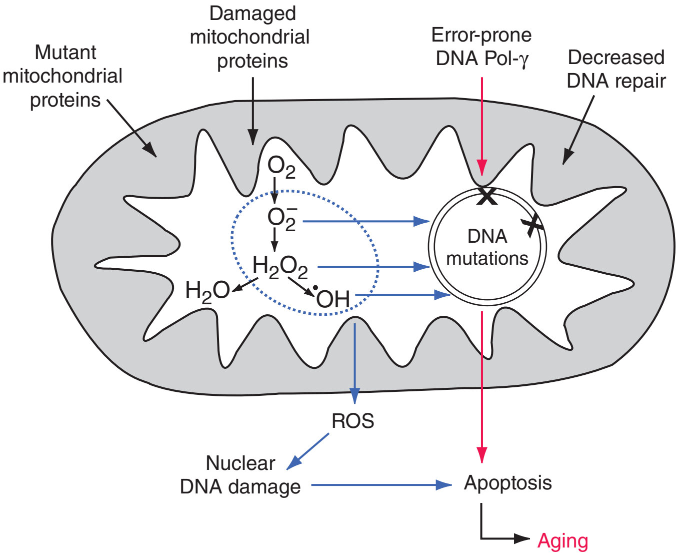

FIGURE 481-11 Multiple pathways of mitochondrial DNA (mtDNA) damage and aging. Multiple factors may impinge on the integrity of mitochondria that lead to loss of cell function, apoptosis, and aging. The classic pathway is indicated with *blue arrows*; the generation of reactive oxygen species (ROS; superoxide anion, hydrogen peroxide, and hydroxyl radicals), as a by-product of mitochondrial oxidative phosphorylation, results in damage to mitochondrial macromolecules, including the mtDNA, with the latter leading to deleterious mutations. When these factors damage the mitochondrial energy-generating apparatus beyond a functional threshold, proteins are released from the mitochondria that activate the caspase pathway, leading to apoptosis, cell death, and aging. *(Reproduced with permission from L Loeb et al: The mitochondrial theory of aging and its relationship to reactive oxygen species damage and somatic mt DNA mutations. Proc Natl Acad Sci USA 102 (52):18769-18770, 2005.)*

Genetic intervention studies in animal models have sought to clarify the potential causative relationship between acquired somatic mtDNA mutation and the aging phenotype and the role of ROS in particular. Replication of the mitochondrial genome is mediated by the activity of the nuclear-encoded *POLG*. A transgenic homozygous mouse knock-in mutation of this gene renders the polymerase enzyme deficient in proofreading and results in a threefold to fivefold increase in mtDNA mutation rate. Such mice develop a premature aging phenotype, which includes subcutaneous lipoatrophy, alopecia, kyphonia, and weight loss with premature death. Although the finding of increased mtDNA mutation and mitochondrial dysfunction with age has been solidly established, the causative role and specific contribution of mitochondrial ROS to aging and age-related disease in humans have yet to be proved. Similarly, although many tumors display higher levels of heterogeneous mtDNA mutations, a causal relationship to tumorigenesis has not been proved.

Besides the age-dependent acquired accumulation in somatic cells of heterogeneous point mutations and deletions, a quite different effect of nonheritable and acquired mtDNA mutations has been described affecting tissue stem cells. In particular, disease phenotypes attributed to acquired mtDNA mutation have been observed in sporadic and apparently nonfamilial cases involving a single individual or even tissue, usually skeletal muscle. The presentation consists of decreased exercise tolerance and myalgias, sometimes progressing to rhabdomyolysis. As in the case of the sporadic, heteroplasmic, large-scale deletion, classic syndromes of chronic PEO, Pearson's syndrome, and KSS, the absence of a maternal inheritance pattern and the finding of limited tissue distribution suggest a molecular pathogenic mechanism emanating from mutations arising de novo in muscle stem cells after germline differentiation (somatic mutations that are not sporadic and occur in tissue-specific stem cells during fetal development or in the postnatal maintenance or postinjury repair stage). Such mutations would be expected to be propagated only within the progeny of that stem cell and affect a singular tissue within a given individual, without evidence of heritability.

PROSPECTS FOR CLINICAL MANAGEMENT OF mtDNA DISEASE

■ TREATMENT OF mtDNA DISORDERS

No specific curative treatment for mtDNA disorders is currently available; therefore, the management of mitochondrial disease is largely supportive. Management issues may include early diagnosis and medical management of epilepsy, gastrointestinal dysfunction, weakness, diabetes mellitus, cardiac dysrhythmia, hearing loss, endocrinopathy, ptosis, and cataracts. Rapid identification of subclinical seizures, which may present with focal neurologic signs or even mild mental status changes, is critical in the management of MELAS and other mtDNA disorders associated with epilepsy. The value of aggressive symptom management cannot be understated. Less specific interventions in the case of other disorders involve combined treatment strategies including dietary intervention and removal of toxic metabolites. Cofactors and vitamin supplements are widely used in the treatment of diseases of mitochondrial oxidative phosphorylation, although there is little evidence, apart from anecdotal reports, to support their use. This includes administration of artificial electron acceptors, including vitamin K_3, vitamin C, and ubiquinone (CoQ_{10}); administration of cofactors (coenzymes) including riboflavin, carnitine, and creatine; and use of oxygen radical scavengers, such as vitamin E, copper, selenium, ubiquinone, and idebenone. Drugs that could interfere with mitochondrial function, such as the anesthetic agent propofol, barbiturates, and high doses of valproate, can generally be avoided if possible. The use of valproate in patients with pathogenic mutations in *POLG* and possibly other mutations affecting mtDNA stability and replication is especially contraindicated. Supplementation with the nitric oxide synthase substrate L-arginine and, more recently, L-citrulline has been advocated as a vasodilator treatment during stroke-like episodes as well as for chronic management in patients with MELAS. Open-label studies demonstrate that levoarginine and levocitrulline may be helpful in reducing the stroke-like symptoms in MELAS but may have serious side effects. As CSF folate deficiency has been reported in some cases of mitochondrial disease, this can be treated with oral folinic acid.

The physician should also be familiar with environmental interactions, such as the strong and consistent association between visual loss in LHON and smoking or ethanol consumption. A clinical penetrance of 93% was found in men who smoked. Asymptomatic carriers of an LHON mtDNA mutation should, therefore, be strongly advised not to smoke and to moderate their alcohol intake. Although not a cure, these interventions might stave off the devastating clinical manifestations of the LHON mutation. Another example is strict avoidance of aminoglycosides in the familial syndrome of ototoxic susceptibility to aminoglycosides in the presence of the mtDNA m.1555A>G mutation of the 12SrRNA encoding gene.

Clinical trials using novel agents have been initiated and launched. These agents include elamipretide (Stealth Biotherapeutics) and KL-1333 (Abliva). In an open-label study of α-tocotrienol used to treat 10 children with Leigh's syndrome, there were improvements in the primary endpoints, including the Newcastle Pediatric Mitochondrial Diseases Scale, the Gross Motor Function Measure, and the PedsQL Neuromuscular Module.

GENETIC COUNSELING, PRENATAL DIAGNOSIS, AND PGD IN mtDNA DISORDERS

The provision of accurate genetic counseling and reproductive options to families with mtDNA mutations is challenging due to the unique genetic features of mtDNA inheritance that distinguish it from Mendelian genetics. mtDNA defects are transmitted by maternal inheritance. mtDNA de novo mutations are often large deletions, affect one family member, and usually represent no significant risk to other members of the family. In contrast, mtDNA point mutations or duplications can be transmitted maternally. Accordingly, the father of an affected individual has no risk of harboring the disease-causing mutation, and a male cannot transmit the mtDNA mutation to his offspring. In contrast, the mother of an affected individual usually harbors the same mutation but may be completely asymptomatic. This wide phenotypic variability is primarily related to the phenomena of heteroplasmy and the mutation load carried by different members of the same family. Consequently, a symptomatic or asymptomatic female harboring a disease-causing mutation in a heteroplasmic state will transmit to her offspring variable amounts of the mutant mtDNA molecules. The offspring will be symptomatic or asymptomatic primarily according to the mutant load transmitted via the oocyte and, to some extent, subsequent mitotic segregation during development. Interactions with the mtDNA haplotype background or nuclear human genome (as in the case of LHON) serve as an additional important determinant of disease penetrance. Because the severity of the disease phenotype associated with the heteroplasmic mutation load is a function of the stochastic differential segregation and copy number of mutant mtDNA during the oogenesis bottleneck and, subsequently, following tissue and organ development in the offspring, it is rarely predictable with any degree of accuracy. For this reason, prenatal diagnosis (PND) and PGD techniques that have evolved into integral and well-accepted standards of practice are severely hampered in the case of mtDNA-related diseases.

The value of PND and PGD is limited, partly due to the absence of data on the rules that govern the segregation of wild-type and mutant mtDNA species (heteroplasmy) among tissue in the developing embryo. Three factors are required to ensure the reliability of PND and PGD: (1) a close correlation between the mutant load and the disease severity, (2) a uniform distribution of mutant load among tissues, and (3) no major change in mutant load with time. These criteria are suggested to be fulfilled for the NARP m.8993T>G mutation but do not seem to apply to other mtDNA disorders. In fact, the level of mutant mtDNA in a chorionic villous or amniotic fluid sample may be very different from the level in the fetus, and it would be difficult to deduce whether the mutational load in the prenatal samples provides clinically useful information regarding the postnatal and adult state.

PREVENTION OF MITOCHONDRIAL DISEASE INHERITANCE BY ASSISTED REPRODUCTIVE TECHNOLOGIES

Because the treatment options for patients with mitochondrial disease are rather limited, with no current U.S. Food and Drug Administration (FDA)–approved therapies for established mitochondrial DNA disease, preventive interventions that eliminate the likelihood of transmission of affected mtDNA into offspring are desirable. The poor reliability of prenatal and preimplantation approaches in predicting mitochondrial DNA disease has resulted in the search for alternative preventive approaches. The common purpose underlying various emerging approaches is to reduce mutant heteroplasmy levels to a level below a pathogenic threshold. This is based on the observed relationship between heteroplasmy and disease inheritance patterns, which indicates that even a small increase in copy number of nonmutant mtDNA molecules in the fertilized egg can exceed the threshold required to ameliorate serious clinical disease. Use of gene editing, with clustered regularly interspaced short palindromic repeats (CRISPR) or mitochondrial-targeted TALEN (transcription activator-like effector nucleases) technology, and others, for example, to shift the heteroplasmy load in affected tissues will require future development of corrective gene delivery techniques. Likewise, induced pluripotent cell technology has not yet met with widespread success in the preclinical research setting. This has prompted the application of mitochondrial replacement therapy (MRT) approaches **(Fig. 481-12)**. These approaches substitute in vitro the entire oocyte or zygote complement of mitochondria, together with their mtDNA from the carrier mother, with the unaffected complement of mitochondria and their unaffected mtDNA from a donor woman. This can be accomplished either by removing and transferring the carrier mother's spindle with her nuclear DNA into the unfertilized oocyte of the donor or, alternatively, by transferring the pronucleus from the fertilized oocyte of the carrier mother to the unfertilized donor oocyte from which the pronucleus has been removed. These approaches provide a "bulk" substitution and hence do not target the specific mtDNA mutation, and they are potentially applicable to a wide variety of mtDNA disorders. This is a form of germline genetic therapy, and therefore, it projects onto future generations in the case of a female offspring. Accordingly, ethical and regulatory bodies have appropriately weighed in on the societal implications of such approaches and have been tentatively supportive of human clinical investigation for situations that would prevent great suffering and when the clinical need is clear and unambiguous, subject to specified conditions and principles and subject to ethical scrutiny. Several such studies have been initiated, and careful examination and follow-up are needed to determine developmental and longer-term health and fertility of children who have undergone genetic manipulation at the earliest stages of human development and whose genomes comprise

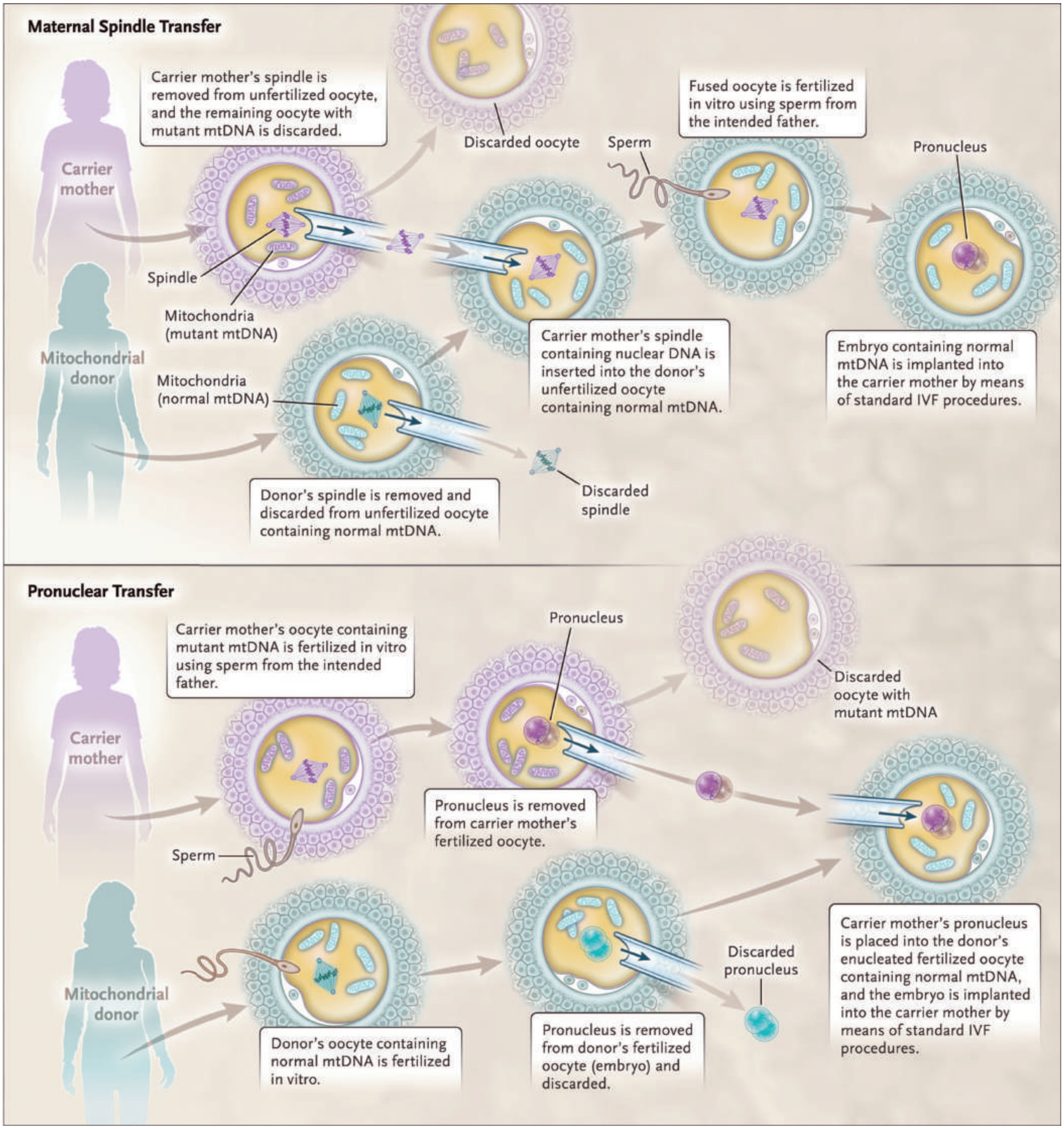
Maternal Spindle Transfer
Carrier mother's spindle is removed from unfertilized oocyte, and the remaining oocyte with mutant mtDNA is discarded.
Discarded oocyte
Sperm
Fused oocyte is fertilized in vitro using sperm from the intended father.
Pronucleus
Carrier mother
Spindle
Mitochondria (mutant mtDNA)
Mitochondrial donor
Mitochondria (normal mtDNA)
Carrier mother's spindle containing nuclear DNA is inserted into the donor's unfertilized oocyte containing normal mtDNA.
Embryo containing normal mtDNA is implanted into the carrier mother by means of standard IVF procedures.
Donor's spindle is removed and discarded from unfertilized oocyte containing normal mtDNA.
Discarded spindle
Pronuclear Transfer
Carrier mother's oocyte containing mutant mtDNA is fertilized in vitro using sperm from the intended father.
Pronucleus
Discarded oocyte with mutant mtDNA
Carrier mother
Pronucleus is removed from carrier mother's fertilized oocyte.
Sperm
Mitochondrial donor
Discarded pronucleus
Carrier mother's pronucleus is placed into the donor's enucleated fertilized oocyte containing normal mtDNA, and the embryo is implanted into the carrier mother by means of standard IVF procedures.
Donor's oocyte containing normal mtDNA is fertilized in vitro.
Pronucleus is removed from donor's fertilized oocyte (embryo) and discarded.

FIGURE 481-12 Mitochondrial replacement techniques—maternal spindle transfer and pronuclear transfer. In both procedures, some mutant mtDNA, estimated at 1–2%, might be carried over together with the spindle or pronucleus, but the levels are low enough to avoid disease risk. IVF, in vitro fertilization. *(From MJ Falk et al: Mitochondrial replacement techniques—Implications for the clinical community. N Engl J Med 374:1103, 2016. Copyright © 2016, Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

separate maternal origins of nuclear and mtDNA genomes. It has been recommended that such studies be limited to male offspring, who cannot then transmit the donor mtDNA to future generations, until such time as the health, ethical, and societal issues are well understood and live up to the exciting promise of reducing the burden of clinical mtDNA disease in the future.

■ FURTHER READING

Alston CL et al: The genetics and pathology of mitochondrial disease. J Pathol 241:236, 2017.

Camp K et al: Nutritional interventions in primary mitochondrial disorders: Developing an evidence base. Mol Genet Metab 119:187, 2016.

Di Mauro S: Mitochondrial encephalomyopathies—Fifty years on: The Robert Wartenberg Lecture. Neurology 81:281, 2013.

Di Mauro S et al: The clinical maze of mitochondrial neurology. Nat Rev Neurol 9:429, 2013.

Gorman GS et al: Mitochondrial diseases. Nat Rev Dis Primers 2:16080, 2016.

Haas RH et al: Mitochondrial disease: A practical approach for primary care physicians. Pediatrics 129:1326, 2007.

Koopman WJ et al: Monogenic mitochondrial disorders. N Engl J Med 366:1132, 2012.

Parikh S et al: Patient care standards for primary mitochondrial disease: A consensus statement from the Mitochondrial Medicine Society. Genet Med 19:1380, 2017.

Parikh S et al: Diagnosis of "possible" mitochondrial disease: An existential crisis. J Med Genet 56:123, 2019.

Russell OM et al: Mitochondrial diseases: Hope for the future. Cell 181:168, 2020.

Saneto RP et al (eds): *Mitochondrial Case Studies, Underlying Mechanisms and Diagnosis*. London, Academic Press/Elsevier, 2016.

Wei W, Chinnery PF: Inheritance of mitochondrial DNA in humans: Implications for rare and common diseases. J Intern Med 287:634, 2020.

482 Telomere Disease

Rodrigo T. Calado, Neal S. Young

DEFINITION

In telomere diseases (*telomeropathies* or *telomere biology disorders*), organ dysfunction is caused by excessive loss of the ends of chromosomes, a process termed *accelerated telomere attrition*, which results from germline mutations in genes involved in telomere maintenance. Inadequate repair or insufficient protection of telomeres and their resulting accelerated erosion induces cell death, deficient cell proliferation, and chromosome instability; affected tissues show defective regeneration, fibrosis or replacement by fat, and a proclivity for cancer. Bone marrow, lungs, liver, and skin especially suffer accelerated telomere loss and dysfunction. Telomeres shorten in humans at an average of 50 base pairs/year as measured in peripheral blood leukocytes. However, normal aging is not associated with developing manifestations of a telomere disease. In normal aging, sufficient stem cell number and function are maintained to sustain vital processes. Telomere length associates with life span in the general population. While shorter leukocyte telomeres correlate with increased mortality risk, especially from nonmalignant causes, telomere loss is not established as the cause of physiologic aging. Long telomeres due to rare inherited mutations do associate with clonal hematopoiesis and predisposition to cancer.

DISEASE MECHANISM

Telomeres, the physical termini of linear chromosomes, are repeated hexanucleotide sequences physically associated with specific proteins. Telomeres protect the chromosome ends against recognition as damaged or infectious DNA by the DNA repair machinery **(Fig. 482-1)**. During mitosis, the DNA polymerase employs an RNA oligonucleotide with a 3′ hydroxyl group to prime replication. The primer dissociates as the DNA polymerase advances along the template strand, and a gap is left at the ends of linear DNA molecules: the newly synthesized DNA strand is necessarily shorter than the original template—the "end-replication problem." Chromosome erosion is thus inevitable with mitotic cell division, but the repetitive structure buffers the loss of genetic information. In human cells, telomeres consist of hundreds to thousands of TTAGGG tandem repeats in the leading and CCCTAA in the lagging DNA strand. At birth, telomeres are long, but they inexorably shorten with aging (Fig. 482-1). In an individual cell, critically short telomere length triggers the p53 pathway, leading to proliferative arrest and apoptosis. Telomere loss is the molecular basis for the "Hayflick phenomenon," the limit to cell division. If a cell overcomes proliferation arrest, extremely short telomeres engage the DNA damage repair machinery, and chromosome end-to-end fusions, chromosome breaks, aneuploidy, and chromosome instability may occur. In addition to the telomere repeated sequences, a group of specialized proteins, collectively termed *shelterins*, directly bind to or indirectly associate with telomeres, assisting in the organization of the telomere tertiary structure and inhibiting the activity of DNA damage response proteins (Fig. 482-1).

To escape telomere attrition, cells with high proliferative demand, including embryonic and adult stem cells, lymphocytes, and most cancer cells, preserve telomere length by the synthesis of telomeric repeats. Telomerase adds GTTAGG hexanucleotides to the 5′ end of the leading DNA strand using a reverse transcriptase enzyme (TERT), and TERC, its RNA template (Fig. 482-1). The telomerase holoenzyme complex comprises two copies of TERT, TERC, and dyskerin and associated proteins. TERC binds to TERT and serves as the RNA template for its function as a reverse transcriptase. Dyskerin, encoded by *DKC1*, stabilizes the complex, and TCAB1, encoded by the *WRAP53* gene, aids telomerase trafficking to the Cajal bodies, nuclear structures for ribonucleoprotein processing where telomerase associates with telomeres for elongation. Telomerase expression is tightly regulated: MYC, sex hormones, and many additional factors stimulate *TERT* transcription. In mature cells, the *TERT* gene is highly repressed. In addition, shelterin proteins (such as POT1) regulate telomerase function and processivity (the ability to consecutively synthesize telomeric repeats in a single interaction with the telomere), modulating its catalytic activity on telomeres. Other proteins also are necessary for telomere length maintenance. RTEL1, a DNA helicase, dismantles t-loops and resolves g-quadruplexes, ensuring adequate telomere elongation.

Pathologic accelerated telomere attrition has a genetic origin. Germline loss-of-function mutations in genes involved in telomere biology impair telomere length repair, increasing the rate of telomere erosion in highly proliferative cells, reaching critically short lengths fast. The consequences are limited cell proliferation and impaired tissue regeneration. Some organs appear to be particularly susceptible to telomere erosion. Billions of blood cells are produced daily **(Chap. 101)**, and telomere attrition curtails cell proliferation, producing a hypocellular marrow and low blood counts. Lymphocytes also have high proliferative capacity and activate telomerase, but when telomeres are very short, B and T cells have an aged, more exhausted, and proinflammatory profile, and immunodeficiency may occur. The liver is also an organ with high proliferative capacity, and telomere dysfunction impairs hepatic regeneration after injury, with a variety of pathologic consequences. The lung alveolar epithelium is in contact with exogenous toxins that stimulate regeneration, and telomere loss may cooperate with other events to hinder these physiologic responses. However, it remains unclear why other regenerative tissues, like the intestine, are less affected by telomere dysfunction, or the mechanism by which telomere loss provokes a fibrotic response in the lungs (pulmonary fibrosis), an adipose response in the marrow (aplastic anemia), and both in the liver (hepatic steatosis and cirrhosis). These phenotypes appear to result from a combination of genetic, epigenetic, and environmental determinants.

When telomeres are critically short, the DNA damage response machinery may be recruited, mistaking telomeres for damaged or infectious DNA and forcing inappropriate repair. Activation of this pathway may cause chromosome instability due to end-to-end fusion of chromosomes or translocations; these alterations generate genomic instability and potentially malignant clones of cells. That telomere dysfunction increases the risk of cancer development has been demonstrated in murine models of telomerase deficiency, and patients with telomere diseases are prone to develop malignant neoplasms, particularly acute myeloid leukemia and head and neck squamous cell carcinomas.

GENETICS

The pattern of inheritance may be X-linked, autosomal recessive, and autosomal dominant, and penetrance is highly variable, even within a pedigree. The genetic architecture may be complex, and affected patients may inherit pathogenic loss-of-function variants in more than one gene involved in the same telomere biology pathway. At least 17 genes have been implicated in the etiology of telomeropathies **(Table 482-1)**.

CLINICAL MANIFESTATIONS

Presentation of telomere disease in the clinic is highly variable—in the tissues affected, in the severity of organ dysfunction, and in patterns of

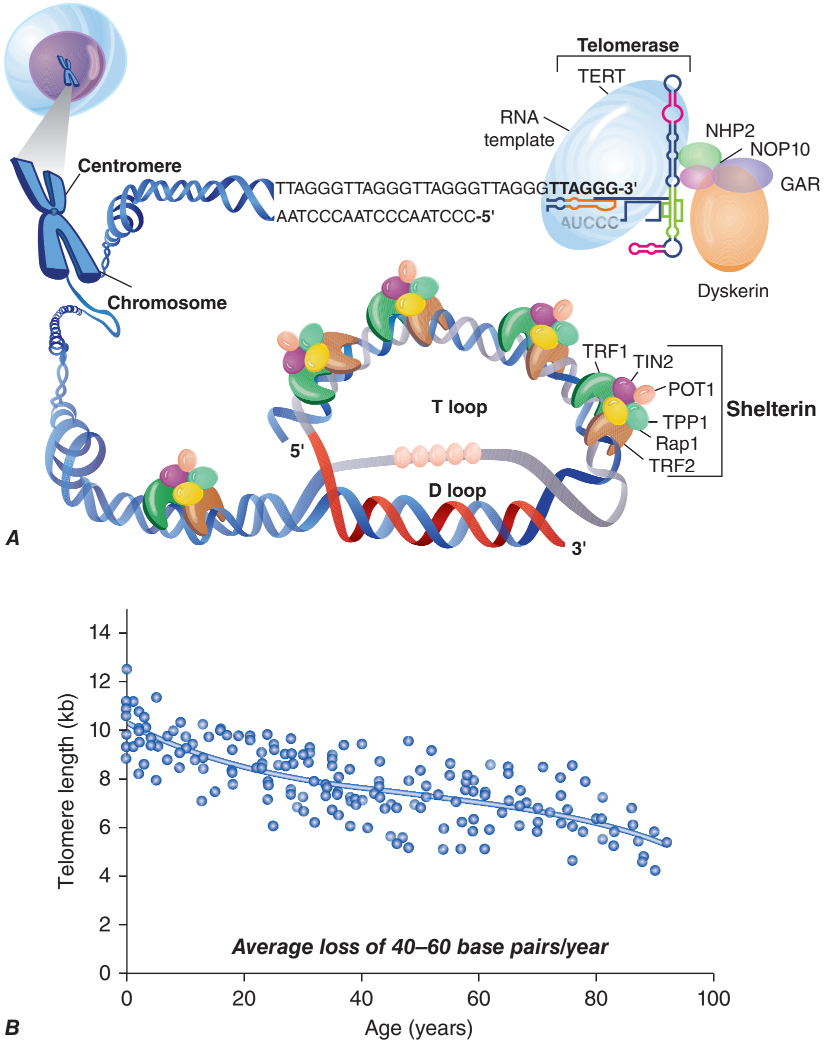

FIGURE 482-1 Telomeres and telomerase. ***A.*** Telomeres are ribonucleoprotein structures located at the termini of linear chromosomes inside the cell nucleus composed of hundreds of tandem hexameric DNA repeats. A group of proteins bind directly or indirectly to telomere sequences in order to provide protection to the structure and are collectively termed *shelterin* or *telosome* (TRF1, TRF2, TIN2, POT1, TPP1, and RAP1). As the 3′ end of the leading strand forms a single-stranded overhang, it folds back and invades the telomeric double helix, forming a lariat termed *T loop*. The telomerase complex is composed of the enzyme telomerase reverse transcriptase (TERT), its RNA component (TERC), the protein dyskerin, and associated proteins (NHP2, NOP10, and GAR). This enzymatic complex elongates telomeres by adding GTTAGG hexameric repeats to the 3′ end of the telomeric leading strand, using a sequence in TERC as the template. ***B.*** The average telomere length in human leukocytes varies: it is longer at birth (10–11 kilobases) and progressively shortens with aging (6–7 kilobases at age 90 years) at an average loss of 40–60 base pairs/year. However, there is significant variability in telomere length in each given age.

disease within a pedigree and between families with similar mutations. In a same family, one individual may be severely affected, but close relatives carrying the same mutation may be asymptomatic and with normal laboratory results. Asymptomatic carriers may have subclinical organ dysfunction, which may be detected by directed or specialized testing (reduced forced vital capacity on pulmonary function test, hypocellular bone marrow at biopsy, hepatic steatosis on ultrasound). Somatic genetic rescue is a rare spontaneous genetic event in a somatic cell, conferring a selective advantage and annulling the effect of the original pathogenic germline mutation and may occur in telomere diseases, mitigating the phenotype. Clonal hematopoiesis may be maladaptive, as acquired mutations in a specific set of genes (e.g., *TP53*) lead to myeloid neoplasms.

Environmental regenerative stresses, including factors such as smoking, alcohol consumption, and viral infection, may increase susceptibility to organ damage and contribute to disease heterogeneity.

Disease anticipation, in which clinical phenotype manifests at an earlier age in successive generations, is observed in some families with telomeropathies due to the direct inheritance of short telomeres in sperm and oocytes.

The diagnosis of a telomere disease is suggested by personal and family history, strengthened by simple measurement of leukocyte telomere length, and usually definitively established by next-generation sequencing for genes encoding telomere repair enzyme complex and shelterin components. In a machine-learning tool that differentiates acquired immune aplastic anemia from inherited bone marrow failure syndromes, telomere length is a top predictor.

Dyskeratosis Congenita Dyskeratosis congenita is the classic telomere disease of childhood and usually diagnosed in the first two decades of life. Affected children often have at least two features of the mucocutaneous triad of ungual dystrophy, reticular skin pigmentation, and oral leukoplakia **(Fig. 482-2)**. In more severe syndromes, affected newborns have cerebellar hypoplasia (Hoyeraal-Hreidarsson syndrome) or exudative retinopathy (Revesz syndrome) **(Fig. 482-3)**. Telomeres are usually extremely short, below the first percentile expected for age **(Fig. 482-4)**. Most patients with dyskeratosis congenita develop bone marrow failure, often requiring transfusions and, ultimately, bone marrow transplant. Pulmonary fibrosis appears in as many as 20% of cases and liver disease in 10%, often after bone marrow transplant for hematopoietic failure. Other tissues and organs also may be affected (Fig. 482-3). Mutations most common in dyskeratosis congenita patients are in *DKC1*, *TINF2*, *TERT*, *TERC*, and *RTEL1* genes, and triallelic inheritance (involving two genes in the same pathway) also may occur (Table 482-1).

Bone Marrow Failure Aplastic anemia **(Chap. 107)** is the most common major clinical manifestation of dyskeratosis congenita. However, young or older patients carrying a telomere defect, without typical physical stigmata, can also develop marrow failure. Genetic variants usually are monoallelic (one mutated allele and one wild-type allele), resulting in haploinsufficiency, and *TERT*, *TERC*, and *RTEL1* are the genes usually affected. Telomere loss in these cases is often less intense than in classic dyskeratosis congenita. As a result of inadequate telomerase function, the stem cell pool is limited in size and in its ability to regenerate, leading to marrow hypocellularity, and insufficient production of erythrocytes, platelets, and granulocytes **(Fig. 482-5)**. The most typical presentation is moderate aplastic anemia after a long history of macrocytic mild to moderate anemia and/or thrombocytopenia, with preservation of leukocyte numbers. A comprehensive personal and family history is important, querying especially for blood count abnormalities and cytopenia as well as lung and liver disease; early hair graying, while not specific to telomeropathies, strongly suggests telomere disease in the appropriate context.

Myeloid Neoplasms Some patients diagnosed with myelodysplastic syndrome **(Chap. 107)** or acute myeloid leukemia **(Chap. 109)** have a family history of bone marrow failure or other myeloid neoplasms. One of the genetic causes for myeloid neoplasia predisposition is a telomere defect, and these disorders are now classified together by the World Health Organization as "myeloid neoplasms associated with telomere biology disorders." Telomere length measurement may be confounded by circulating immature cells, which may have very short telomeres, precluding accurate test interpretation.

Pulmonary Fibrosis Pulmonary fibrosis appears in ~20% of children with dyskeratosis congenita. Conversely, ~10–15% of patients with idiopathic pulmonary fibrosis **(Chap. 304)** or familial pulmonary

TABLE 482-1 Genetic Variants in 13 Genes Involved in Telomere Maintenance, Inheritance Pattern, and Phenotype

GENE	DYSKERATOSIS CONGENITA	APLASTIC ANEMIA	PULMONARY FIBROSIS	CIRRHOSIS	MDS/LEUKEMIA
Telomerase					
DKC1	XL				
TERT	AD/AR	AD/AR	AD	AD	AD/AR
TERC	AD/AR	AD	AD	AD	AD
NOP10	AR				
NHP2	AR				
WRAP53	AR				
Shelterin					
TINF2	AD	AD	AD		
TERF2		AD			
ACD	AD				
Others					
RTEL1	AR	AD/AR	AD		AD
CTC1	AR	AR			
PARN			AD		
USB1	AD				
ZCCHC8		AD	AD		
NAF1			AD		

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; MDS, myelodysplastic syndrome; XL, X-linked.

fibrosis have an etiologic telomerase gene mutation. Regardless of mutation status, most pulmonary fibrosis patients have short telomeres for their age but not as short as in dyskeratosis congenita. How telomere erosion causes pulmonary fibrosis is unclear, but it might prevent adequate proliferation and regeneration of pneumocytes type II. Idiopathic pulmonary fibrosis due to a telomere disease usually appears after the fourth decade of life, with a restrictive pattern on pulmonary function testing associated with decreased diffusion capacity for carbon monoxide (DL_{CO}) and a diffuse "honeycomb" appearance on high-resolution computed tomography (CT) (Fig. 482-5). Histopathology of biopsied lung shows interstitial pneumonia. The pulmonary clinical presentation in telomere disease is indistinguishable from idiopathic pulmonary fibrosis, except that those with an underlying telomere defect may have cryptic hepatic cirrhosis, macrocytosis, cytopenias, and a family history of lung, liver, or bone marrow disease. Pulmonary arteriovenous malformation leading to right-to-left shunting is observed in patients with pulmonary fibrosis due to telomere disease. Patients with idiopathic pulmonary fibrosis or familial pulmonary fibrosis should have leukocyte telomere length assayed and, if telomeres are short, undergo screening for mutations in telomere-associated genes and telomeres; telomere length may be normal in some cases despite the presence of pathogenic mutations. *TERT*, *TERC*, *RTEL1*, and *PARN* are the most commonly mutated genes.

Liver Disease Genetic telomere defects may cause hepatic cirrhosis (Chap. 355), nodular regenerative hyperplasia of the liver, nonalcoholic fatty liver disease (Chap. 354), and hepatocellular carcinoma (Chap. 87). Hepatocytes of most patients with cirrhosis have very short telomeres. Eroded telomeres limit hepatocyte proliferation, especially upon chronic injury. Additionally, hepatocytes with short telomeres display abnormal metabolic patterns and defective mitochondrial function. Abnormal liver pathology may be uncovered incidentally during the evaluation of telomeropathy patients with aplastic anemia or pulmonary fibrosis, but cirrhosis also may be the sole or most prominent clinical presentation of a telomere defect. A minority of individuals with cirrhosis associated with virus B or C infection or alcohol-associated liver disease may carry a telomere-associated gene mutation. Liver histopathology is variable, but cirrhosis is usually associated with inflammation (Fig. 482-5), increased iron deposit, positivity for CD34 in sinusoid endothelial cells, and widening of hepatocyte plates. Defective telomere maintenance may increase the susceptibility of the liver to environmental challenges, such as alcohol and viruses, increasing the risk for developing severe hepatic disease in mutation carriers.

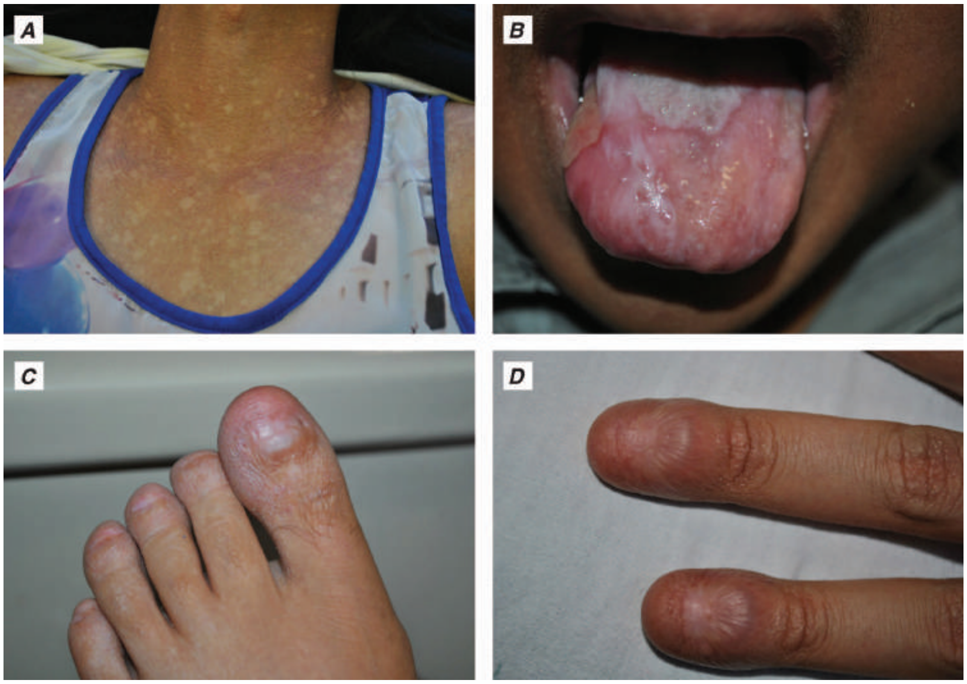

FIGURE 482-2 Skin manifestations of dyskeratosis congenita. The pediatric syndrome dyskeratosis congenita is characterized by the mucocutaneous triad of ***(A)*** reticular skin pigmentation, ***(B)*** oral leukoplakia, and ***(C, D)*** nail dystrophy.

LONG TELOMERE SYNDROME

POT1 is a shelterin component that modulates telomerase access to telomeres, thus regulating telomere elongation. Heterozygous germline loss-of-function mutations in the *POT1* gene cause excessively long telomeres that clinically translate into clonal hematopoiesis and a predisposition to benign and malignant tumors. Very long telomeres appear to augment the capacity for cell proliferation, facilitating the acquisition of harmful driver somatic mutations.

TELOMERE LENGTH MEASUREMENT

Length of telomeres can be accurately measured in peripheral blood leukocytes by commercial laboratories. Of several methods available, flow–fluorescence in situ hybridization (FISH) and quantitative real-time polymerase chain reaction (qPCR) are most widely

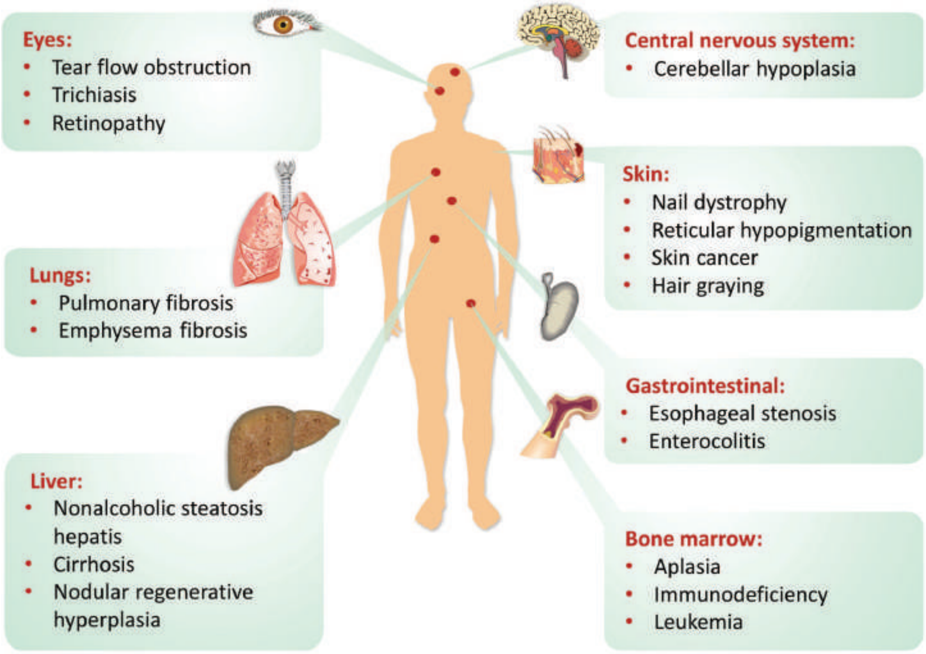

FIGURE 482-3 Clinical consequences of telomere diseases. Telomere dysfunction affects a variety of organs: cerebellum, eyes, lungs, liver, skin, gastrointestinal tract, and the bone marrow.

utilized. Both methods have advantages and limitations and require high-quality samples, usually fresh or freshly processed, as cell death and DNA degradation impact the accuracy of testing. Results are usually expressed as leukocyte telomere length in kilobases. However, the interpretation of length must account for patient age, due to physiologic telomere loss. A normal range for telomeres is available for each year of age, longest at birth and shortening at 40–60 base pairs per year (Fig. 482-1). For each age bracket, the percentile curves are calculated, and a given patient's test result is interpreted in the context of normal age variation: telomeres below the tenth percentile for age are defined as "short" and telomeres below the first percentile are considered "very short" (Fig. 482-4). Telomeres above the 99th percentile are considered "very long."

Short telomeres may also be present in some chronic conditions, such as cardiovascular disease or diabetes. In these settings, telomere erosion is not thought to be etiologic but rather a secondary consequence of chronic inflammation; telomere testing does not have clinical utility and is not recommended. Likewise, telomere length tests, despite commercial hyperbole, have not been shown to be useful in the assessment of aging and longevity or as a basis for therapeutic interventions, absent a diagnosis of genetic telomere disease.

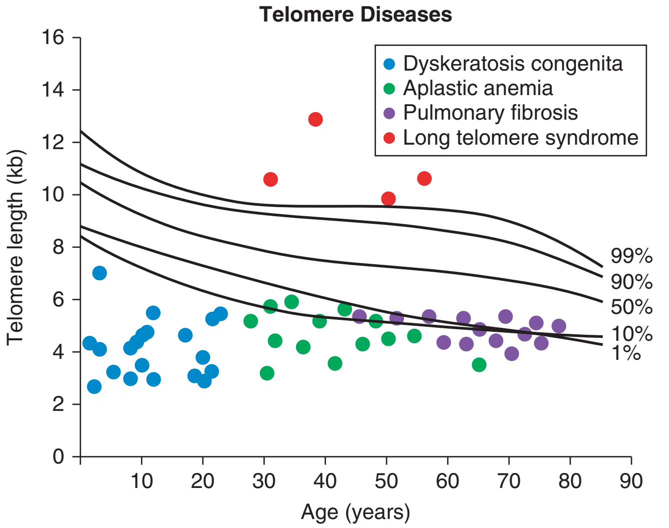

FIGURE 482-4 Telomere length measurement in telomere diseases. Telomeres shorten with aging, and solid curves represent the percentiles for age in healthy subjects. Telomeres are considered "short" when below the 10th percentile, very short when below the first percentile, and very long when above the 99th percentile. In patients with dyskeratosis congenita, telomeres are usually below the first percentile, regardless of the gene lesion, whereas in patients with aplastic anemia or pulmonary fibrosis, telomeres are usually below the 10th percentile. In patients with long telomere syndrome, telomeres are usually above the 99th percentile.

Flow-FISH uses a fluorescent-labeled nucleotide probe specific for telomere repeats to estimate telomere content in an individual cell. It has the advantage of determining telomere length in individual cells and in leukocyte subpopulations; lymphocyte telomere shortening is more specific for telomere diseases than in other cells. A limitation of flow-FISH is the requirement for intact cells for analysis, which are not always available, and neutrophils are susceptible to damage during processing, freezing, and thawing.

qPCR uses telomere-binding modified primers to measure telomere content in comparison to a housekeeping gene in the whole leukocyte population and thus does not require intact cells. qPCR provides an estimate of the average telomere length of a given sample without determining telomere length in individual cells. Good DNA quality is essential for adequate qPCR testing and automation or semi-automation is required for clinical purposes, as variability in conditions among batches may result in interassay variation.

The standard Southern blot is very accurate but requires larger amounts of DNA and is labor intensive. Other measures have been developed in research laboratories (single telomere length analysis [STELA], telomere shortest length assay [TESLA]) to assess critically short telomeres in particular.

GENETIC TESTING

When a patient with a suspected telomeropathy has short or very short telomeres, genetic screening for mutations in genes involved in telomere maintenance and biology is indicated (Table 482-1). Genetic testing has been restricted to patients with suspected telomere disorders but is increasingly incorporated into genomic screening in the bone marrow failure syndromes in general, and next-generation sequencing has been routinely used. Mutations may be biallelic or triallelic involving two genes (especially in dyskeratosis congenita), but usually only one allele is mutated. Interpretation of genetic screening results is challenging, as rare singleton polymorphisms of unknown significance have been identified in large cohorts of healthy individuals. In silico analysis, mutation location, and functional studies are helpful to interpret the mutation significance.

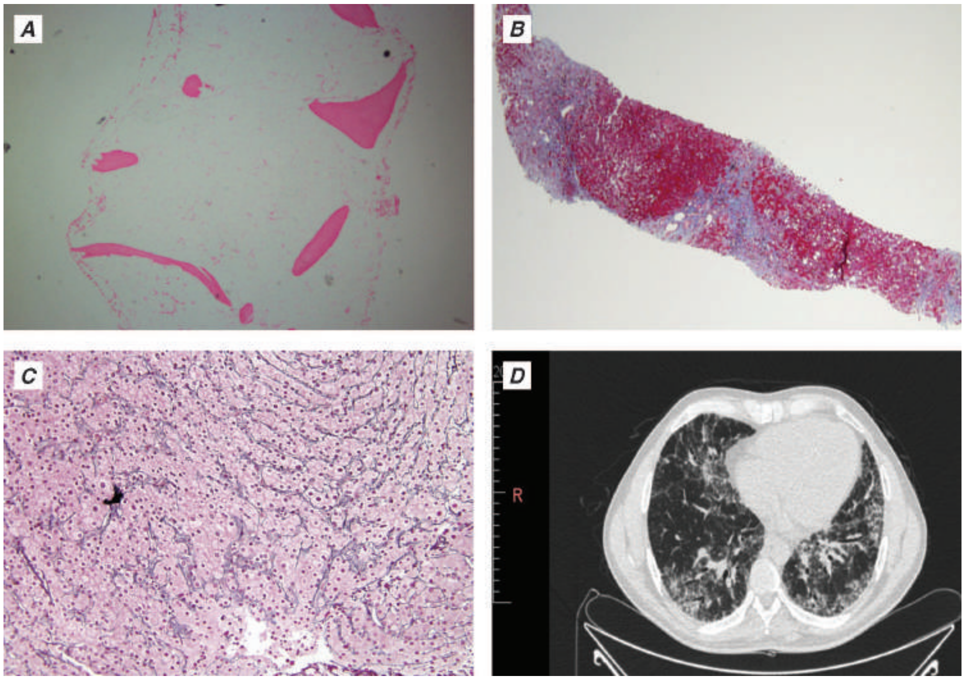

FIGURE 482-5 Pathologic manifestations of telomere diseases. ***A.*** In the bone marrow, telomere erosion predisposes to aplastic anemia, characterized by an empty hematopoietic marrow replaced by fat (hematoxylin and eosin). ***B.*** In the liver, telomere attrition predisposes to cirrhosis (hematoxylin and eosin). ***C.*** Telomere shortening may also result in nodular regenerative hyperplasia of the liver (reticulin stain). ***D.*** In the lungs, telomere dysfunction predisposes to pulmonary fibrosis mainly in the subpleural regions, which may be detected by high-resolution computed tomography scan.

Genetic counseling is necessary after screening, as the inheritance pattern may be autosomal dominant, mutation penetrance is highly variable, and phenotypes may be diverse even within a pedigree. Potential family stem cell donors must be screened before transplantation to ensure that they do not have mutations.

TREATMENT

Telomere Disease

Patients with severe aplastic anemia due to telomere disease may undergo allogeneic hematopoietic stem cell transplant when a suitable donor is available. Treatment-related mortality may be increased due to pulmonary and hepatic complications, for which reduced intensity conditioning regimens appear advantageous. Lung transplant for pulmonary fibrosis is feasible but often not performed due to coexisting cytopenias and other comorbidities. Patients with pulmonary fibrosis associated with telomere disease have a poorer outcome after lung transplant and with nontransplant therapies. Similarly, there is no specific treatment for the liver in telomere disease; liver transplant has been performed in several cases with good outcome and without excessive posttransplant mortality and improvement in the respiratory status. Telomeropathy patients should be advised to avoid toxins (metal dust, busulfan, amiodarone), ionizing radiation, cigarette smoke, and alcohol, as these can be possibly harmful.

Long-term therapy with androgens may mitigate telomere attrition and even elongate leukocyte telomere length in humans. In research trials, danazol and nandrolone improved blood counts in marrow failure patients and reduced transfusion requirements.

FURTHER READING

Blackburn EH et al: Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. Science 350:1193, 2015.

Calado RT, Young NS: Telomere diseases. N Engl J Med 361:2353, 2009.

Carvalho VS et al: Recent advances in understanding telomere diseases. Fac Rev 11:31, 2022.

Clé DV et al: Effects of nandrolone decanoate on telomere length and clinical outcome in patients with telomeropathies: A prospective trial. Haematologica 108:1300, 2023.

Collins J, Dokal I: Inherited bone marrow failure syndromes. Hematology 20:433, 2015.

DeBoy EA et al: Familial clonal hematopoiesis in a long telomere syndrome. N Engl J Med 388:2422, 2023.

Devine MS, Garcia CK: Genetic interstitial lung disease. Clin Chest Med 33:95, 2012.

Gutierrez-Rodrigues F et al: Differential diagnosis of bone marrow failure syndromes guided by machine learning. Blood 141:2100, 2023.

Townsley DM et al: Danazol treatment for telomere diseases. N Engl J Med 374:1922, 2016.

Wang M et al: Liver disease and transplantation in telomere biology disorders: An international multicenter cohort. Hepatol Commun 8:e0462, 2024.

483

Gene and Cell-Based Therapy in Clinical Medicine

Katherine A. High, Marcela V. Maus

Gene therapy is a novel area of therapeutics in which the active agent is a nucleic acid sequence rather than a protein or small molecule. One of the most powerful concepts in modern molecular medicine, gene therapy has the potential to address a host of diseases for which there are currently no available treatments. Because delivery of naked DNA or RNA to a cell is an inefficient process, most gene therapy is carried out using a vector, or gene delivery vehicle, typically engineered from viruses by deleting some or all of the viral genome and replacing it

TABLE 483-1 Characteristics of Commonly Used Gene Delivery Vehicles

	VIRAL BASE		NONVIRAL	
FEATURES	**RETROVIRAL/LENTIVIRAL**	**ADENOVIRAL**	**AAV**	**LIPID NANOPARTICLES**
Genome	RNA	DNA	DNA	RNA
Cell division requirement	G_1 phase	No	No	No
Packaging limitation	8 kb	8–30 kb	5 kb	10 kb or more
Immune responses to vector		Extensive	Few	Few
Genome integration	Yes	Poor	Poor	May be used to package either RNA or DNA
Long-term expression	Yes	No	Yes	Transient for RNA
Main advantages	Persistent gene transfer in transduced tissues	Highly effective in transducing various tissues	Elicits few inflammatory responses, nonpathogenic	RNA expressed transiently
Main disadvantages	Might induce oncogenesis in some cases; only used ex vivo	Viral capsid elicits strong immune responses	Limited packaging capacity	Immunogenic; predominantly targets the liver; challenging manufacturing process

with the therapeutic gene of interest under the control of a suitable promoter. Nonviral delivery vehicles such as lipid nanoparticles are increasingly being used (Table 483-1). Gene therapy strategies can thus be described in terms of three essential elements: (1) a gene delivery vehicle; (2) a gene to be delivered, sometimes called the transgene; and (3) a physiologically relevant target cell to which the DNA or RNA is delivered. The series of steps in which the vector and donated DNA or RNA enter the target cell and express the transgene is referred to as *transduction*. Gene delivery can take place in vivo, in which the vector is directly injected into the patient, or, in the case of hematopoietic, liver, immune, and some other target cells, ex vivo, with removal of the target cells from the patient, followed by return of the gene-modified autologous cells to the patient after manipulation in the laboratory. The latter approach effectively combines gene transfer techniques with cellular therapies.

In the past few years, gene therapy for genetic disease has moved from addressing ultra-rare inherited diseases to more common ones including sickle cell disease and hemophilia. Similarly, chimeric antigen receptor (CAR) T cells have expanded beyond hematologic malignancies to address solid tumors and autoimmune diseases (currently investigational products). Therapeutic approaches have expanded from gene transfer to gene editing, and RNA-based therapies have gained ground rapidly, partly as a key component of CRISPR-based gene editing (guide RNAs and mRNAs encoding editing enzymes) and as the active agent in multiple SARS-CoV-2 vaccines. This chapter will focus primarily on approved therapies (Table 483-2), with some discussion of investigational therapies in late-phase development and earlier trials critical to the development of the field.

Clinical trials of gene therapy have been under way since 1990; the first gene therapy product to be licensed in the United States or Europe was approved in 2012 (see below). Given that vector-mediated gene therapy is arguably one of the most complex therapeutics yet developed, typically consisting of both a nucleic acid and a protein component, this time course from first clinical trial to licensed product is noteworthy for being similar to those seen with other novel classes of therapeutics, i.e., monoclonal antibodies or bone marrow transplantation. Thousands of people have now received approved products or participated in investigational studies of gene transfer. Potential adverse events, predicted based on first principles (Table 483-3), have occurred but have been rare. Some of the initial trials were characterized by an overabundance of optimism and a failure to be appropriately critical of preclinical studies in animals; in addition, it was sometimes not fully appreciated that animal studies are only a partial guide to safety profiles of products in humans (e.g., in the setting of insertional mutagenesis or human immune responses to the vector). Clinical experience and laboratory research led to a more nuanced understanding of the actual risks (Table 483-4) and dramatic benefits of these new therapies and to more sophisticated selection of disease targets. Currently, gene therapies are being developed for a variety of disease entities. Critical aspects of the history to be assessed when evaluating a patient who has received a gene therapy product (or investigational agent) are outlined in Table 483-5.

GENE TRANSFER AND GENE EDITING FOR GENETIC DISEASES

Most approved gene therapies for genetic diseases involve gene addition therapy. Recently the first gene editing products, for sickle cell disease and β-thalassemia, have been approved. Gene therapy strategies generally involve transfer of the missing gene to a physiologically relevant target cell. However, other strategies are possible, including supplying a truncated form of the gene with comparable biological activity (e.g., a gene encoding B domain-deleted FVIII for hemophilia A or microdystrophin for Duchenne muscular dystrophy); supplying a gene that achieves a similar biologic effect through an alternative pathway (e.g., utrophin in place of dystrophin for Duchenne muscular dystrophy); or downregulating a harmful effect through a small interfering or short hairpin RNA. From a therapeutic standpoint, gene therapies for genetic diseases fall into two distinct categories: (1) they may provide treatment for diseases that have hitherto lacked any pharmacologic therapies; or (2) they may provide an alternative to complex medical regimens that are frequently characterized by significant nonadherence due to the burden of treatment (e.g., monthly red blood cell transfusions and iron chelation in transfusion-dependent β-thalassemia).

Gene therapy for genetic disease requires long-term expression of the transgene. Two distinct strategies are available to achieve this goal: one is to transduce stem cells with an integrating vector, so that all progeny cells will carry the donated gene; and the other is to transduce long-lived, postmitotic cells, such as skeletal muscle or neurons. In the case of long-lived cells, integration into the target cell genome is unnecessary. Instead, because the cells are nondividing, the donated DNA, even if stabilized predominantly in an episomal form, will give rise to expression for the life of the cell. This latter approach mitigates risks related to integration and insertional mutagenesis.

CRISPR/Cas9-based gene editing differs from gene therapy in that the therapeutic moiety utilizes a bacterial Cas9 enzyme and a guide RNA to introduce a double strand break (DSB) at a specific site in the DNA. When co-delivered with an appropriately designed gene-targeting vector, gene insertion can take place at the site of the DSB, resulting in a corrected sequence at the endogenous locus of a gene. This permanent correction in the genome is thus under the control of endogenous regulatory signals and will be passed to every daughter cell. In practice, the Cas9 enzyme is complexed with the guide RNA to form a ribonucleoprotein (RNP), which is then delivered to the cell, where the guide RNA directs the RNP complex to the genome target, introducing the DSB at a precise locus. In the *absence* of a gene-targeting vector, the DSB will be repaired using nonhomologous end joining (NHEJ), which typically results in small deletions or insertions, disrupting the expression of the target gene. In the presence of a targeting vector, repair at the DSB may occur by homology-directed repair, resulting in gene replacement at the targeted locus. With currently available systems, the

TABLE 483-2 Currently Approved Gene and Cell Therapy Products in North America and/or Europe

PRODUCT	INDICATION	AGE GROUP	YEAR FIRST APPROVED	WHERE APPROVED	VECTOR	TRANSGENE	TARGET TISSUE
Strimvelis®[a]	ADA-SCID	Pediatric	2016	Europe	Retroviral	*ADA* (adenosine deaminase)	Autologous hematopoietic stem cells (HSCs)
Kymriah® (tisagenlecleucel)	Relapsed or refractory (R/R) B-cell acute lymphoblastic leukemia (pediatric); R/R large B-cell lymphoma (adult); third-line follicular lymphoma	Pediatric and adult, different disease indications	2017	United States, Europe, China, Japan	Lentiviral	CAR directed to CD19 with 4-1BB signaling domain	Autologous T cells
Yescarta® (axicabtagene ciloleucel)	R/R and second-line large B-cell lymphomas; third-line follicular lymphoma	Adult	2017	United States, Europe, Japan	Retroviral	CAR directed to CD19 with CD28 signaling domain	Autologous T cells
Luxturna® (voretigene neparvovec)	Confirmed biallelic *RPE65* mutation–associated retinal dystrophy	Pediatric and adult	2017	United States and Europe	AAV2	*RPE65* (retinal pigment epithelial 65 kD protein)	Retinal pigment epithelial cells by single subretinal injection
Zolgensma® (onasemnogene abeparvovec)	Spinal muscular atrophy type 1 due to biallelic mutations in the *SMN1* gene	Pediatric <2 years of age	2019	United States and Europe	AAV9	*SMN1* (survival motor neuron 1)	Spinal motor neurons by single IV infusion
Zynteglo® (betibeglogene autotemcel)	Transfusion-dependent β thalassemia; sickle cell disease	Adults and pediatric ≥12 years of age	2019	Europe and United States	Lentiviral	$\beta^{A\text{-}T87Q}$ globin gene	Autologous HSCs
Libmeldy®[b] (aditarsagene autotemcel)	Metachromatic leukodystrophy due to biallelic mutations in the arylsulfatase A gene	Pediatric	2020	Europe, United States	Lentiviral	*ARSA* (arylsulfatase A)	Autologous HSCs
Tecartus® (brexucabtagene autoleucel)	R/R mantle cell lymphoma; R/R B-cell acute lymphoblastic leukemia	Adults	2020	United States and Europe	Retroviral	Same molecular construct as axicabtagene	Autologous T cells
Breyanzi® (lisocabtagene maraleucel)	R/R and second-line large B-cell lymphoma	Adult	2021	United States, Europe, and Japan	Lentiviral	CAR directed at CD19 with 4-1BB signaling domain; CD4 and CD8 T-cell products manufactured and infused separately	Autologous T cells
Abecma® (idecabtagene vicleucel)	Fifth-line treatment for multiple myeloma	Adult	2021	United States and Europe	Lentiviral	CAR directed to B-cell maturation antigen (BCMA); 4-1BB signaling domain	Autologous T cells
Carvykti® (ciltacabtagene autoleucel)	Fifth-line treatment for multiple myeloma	Adult	2022	United States and Europe	Lentiviral	CAR directed to BCMA with two single-domain antibodies; 4-1BB signaling domain	Autologous T cells
Skysona® (elivaldogene autotemcel)	Early active cerebral adrenoleukodystrophy	Boys age 4–17	2022	Europe and United States	Lentiviral	Adenosine triphosphate–binding cassette, subfamily D, member 1 (*ABCD1*)	Autologous HSCs
Upstaza® (eladocagene exuparvovec)	Confirmed AADC deficiency with severe phenotype	Children 18 months and older	2022	Europe	AAV2	Human aromatic L-amino acid decarboxylase (*AADC*)	Cells in putamen via single neurosurgical procedure
Roctavian® (valoctocogene roxaparvovec)	Severe hemophilia A and no history of inhibitors	Adults	2022	Europe and United States	AAV5	cDNA encoding human factor VIII, B domain-deleted, SQ form	Hepatocytes via single IV infusion
Hemgenix® (etranacogene dezaparvovec)	Severe or moderately severe hemophilia B	Adults	2022	Europe and United States	AAV5	cDNA encoding factor IX Padua	Hepatocytes via single IV infusion
Vyjuvek® (beremagene geperpavec)	Dystrophic epidermolysis bullosa due to mutations in *COL7A1*	Age 6 months and older	2023	United States	Herpes simplex viral vector	Collagen type VII alpha 1 chain (*COL7A1*)	Keratinocytes and fibroblasts at sites of lesions
Elevidys® (delandistrogene moxeparvovec)	Duchenne muscular dystrophy	Ages 4–5	2023	United States	AAVrh74	cDNA encoding microdystrophin	Skeletal muscle via single IV infusion
Casgevy® (exagamglogene autotemcel)	Sickle cell anemia and β thalassemia	Ages 12 and up	2023	United States and Europe	Gene editing	Inactivates *BCL11a* in RBCs	Autologous HSCs
Beqvez® (fidanacogene elaparvovec)	Severe or moderately severe hemophilia B	Adults	2024	Canada, United States and Europe	AAVrh74variant	cDNA encoding factor IX Padua	Hepatocytes via single IV infusion

[a] Autologous CD34+-enriched cell fraction that contains CD34+ cells transduced with retroviral vector that encodes for the human ADA cDNA sequence. [b] Autologous CD34+ cells encoding arylsulfatase A.

Abbreviations: AAV, adeno-associated virus; ADA-SCID, adenosine deaminase severe combined immunodeficiency; CAR, chimeric antigen receptor; RBC, red blood cell.

TABLE 483-3 Potential Complications of Gene Therapy

Gene silencing—repression of promoter
Genotoxicity—complications arising from insertional mutagenesis, or acceleration of malignant transformation in a cell on the path to oncogenesis before transduction (i.e., CAR introduced into a premalignant T cell)
Phenotoxicity—complications arising from overexpression or ectopic expression of the transgene
Immunotoxicity—harmful immune response to either the vector or transgene, or a harmful immune response of the vector (e.g., CAR T cells)
Risks of horizontal transmission—shedding of infectious vector into environment
Risks of vertical transmission—germline transmission of donated DNA

Abbreviation: CAR, chimeric antigen receptor.

first (cleavage) step is much more efficient than the second (targeting) step. The only approved gene editing product requires only the cleavage step (*vide infra*), and the same is true for most investigational products that have been published as clinical studies.

EX VIVO GENE TRANSFER

Early attempts to effect gene replacement into hematopoietic stem cells (HSCs) were stymied by the relatively low transduction efficiency of retroviral vectors, which require dividing target cells for integration. Because HSCs are normally quiescent, they are a formidable transduction target. However, identification of cytokines that induce cell division without promoting differentiation of stem cells, along with technical improvements in the isolation and transduction of HSCs, led to modest but real gains in transduction efficiency.

Immunodeficiency Disorders: Proof of Principle The first convincing therapeutic effect from gene transfer occurred in children with X-linked severe combined immunodeficiency disease (SCID), which results from mutations in the gene (*IL2RG*) encoding the γc subunit of cytokine receptors required for normal development of T and natural killer (NK) cells **(Chap. 362)**. Affected infants present in the first few months of life with overwhelming infections and/or failure to thrive. In this disorder, it was recognized that successfully transduced cells, even if few in number, would have a proliferative advantage compared to nontransduced cells, which lack receptors for the cytokines required for lymphocyte development and maturation. Isolation of autologous CD34+ cells, followed by transduction with a retroviral vector encoding the γc subunit and transplantation of the gene-modified autologous cells, led to complete reconstitution of the immune system, including documented responses to standard childhood vaccinations, clearing of infections, and remarkable gains in growth in most treated

TABLE 483-4 Adverse Events in Gene Therapy and Gene Editing

VECTOR OR TREATMENT MODALITY	SYMPTOM OR LABORATORY FINDING	MECHANISM	DOSE DEPENDENCE	MITIGATION STRATEGIES
Retroviral or lentiviral vectors	Malignancy[a]	Insertional mutagenesis[a]	Yes for retroviral vectors	Less frequent with lentiviral vectors, likely because of differences in integration patterns
AAV	Vector sequences in semen, risk of germline transmission	Based on animal studies, present in prostatic fluid but not in gametes	Yes	Barrier birth control until three sequential semen samples are negative for vector DNA
	Immune responses directed to capsid, sometimes accompanied by loss of expression	Memory T cells directed to vector capsid in humans, who are natural hosts for wild-type AAV	Yes	Reduce doses Administer immunomodulatory agents (short-term) to reduce or ablate response
	Thrombotic microangiopathy with high-dose systemic infusion	Rapid rise in antibodies to AAV, formation of antigen-antibody complexes, triggering of complement activation	Yes	Has responded to therapy with complement inhibitors including eculizumab
Ex vivo and in vivo genome editing	Off-target cleavage resulting in unintended gene silencing	Guide RNA lacks requisite specificity	Likely	Preclinical assessment for off-target effects Long-term follow-up of trial participants and patients
In vivo genome editing	Liver-directed in vivo editing has shown mild and transient transaminase elevations but excellent efficacy at doses studied clinically	Possibly immune responses to bacterial proteins in editing machinery, potentially resulting in loss of edited cells	Yes	Careful dose-ranging studies in early-phase testing Consider short-term immunomodulatory agents if needed at higher doses
CAR-T therapy	Cytokine release syndrome: fever, hypotension, tachycardia, hypoxia, multiorgan failure	Systemic inflammatory response caused by cytokines released by CAR T cells	Possibly	Tocilizumab/corticosteroids
	Neurotoxicity-cerebral edema and encephalopathy	Peripheral immune overactivation, endothelial activation-induced blood-brain barrier dysfunction, CNS inflammation	Possibly	Avoid seizure-threshold-lowering medications in early phase of treatment Treat with dexamethasone as early as possible (use specific management guidelines)
	Immunodeficiency (hypogammaglobulinemia and susceptibility to viral injections)	On-target effect against B cells and/or plasma cells Preparatory lymphodepleting chemotherapy regimen also contributes	No	Prophylaxis for opportunistic infections (PJP, VZV) for at least 1 year; vaccination schedule (specific guidelines)
	New T-cell malignancy	Insertional mutagenesis or potentially chronic activation due to new transgene	No	Report to FDA and manufacturer Consider activation of suicide gene if present in the transgene expression cassette Treat per standard guidelines

[a] In target cells, either hematopoietic stem cells or T cells.

Abbreviations: AAV, adeno-associated virus; CAR, chimeric antigen receptor; CAR-T, chimeric antigen receptor T cell; CNS, central nervous system; FDA, U.S. Food and Drug Administration; PJP, *Pneumocystis jirovecii* pneumonia; VZV, varicella-zoster virus.

TABLE 483-5 Taking History from Patients Who Have Received Gene Therapies or Gene Editing

Elements of History for Patients Who Received Gene Therapy (or Have Participated in Trials)

1. What vector was administered? Is it predominantly integrating (retroviral, lentiviral, herpesvirus, or gene editing) or nonintegrating (plasmid, adenoviral, adeno-associated viral)?
2. What were the dose and the route of administration of the vector?
3. What was the target tissue?
4. What gene was transferred in? The gene that is defective in the patient's disease? A truncated version? A gene encoding a different protein with similar properties? A knockdown approach?
5. Were there any adverse events noted after gene transfer?

Screening Questions for Long-Term Follow-Up in Gene Transfer Subjects[a]

1. Has a new malignancy been diagnosed? If so, clinicians should contact the manufacturer to report the event and obtain instructions on the collection of patient samples for testing.
2. Has a new neurologic/ophthalmologic disorder, or exacerbation of a preexisting disorder, been diagnosed?
3. Has a new autoimmune or rheumatologic disorder been diagnosed?
4. Has a new hematologic disorder been diagnosed?

[a] Factors influencing long-term risk include integration of the vector into the genome, vector persistence without integration, and transgene-specific effects.

children. However, among 20 children treated in the initial trials, five eventually developed a syndrome similar to T-cell acute lymphocytic leukemia, with splenomegaly, rising white counts, and the emergence of a single clone of T cells. Molecular studies revealed that, in most of these children, the retroviral vector had integrated within a gene, *LMO-2* (LIM only-2), which encodes a component of a transcription factor complex involved in hematopoietic development. The retroviral long terminal repeat acted as a promoter to increase the expression of *LMO-2*, resulting in T-cell leukemia.

The X-linked SCID studies were a watershed event in the evolution of gene therapy. They demonstrated conclusively that gene therapy could cure disease, with dramatic and durable clinical results. However, they also demonstrated that insertional mutagenesis leading to cancer was more than a theoretical possibility (Table 483-3). As a result of the experience in these trials, all protocols using integrating vectors in hematopoietic cells must include a plan for monitoring sites of insertion and clonal proliferation for 15 years after treatment. Initial strategies to overcome the possible complication of insertional mutagenesis included using a "suicide" gene cassette in the vector, so that errant clones can be quickly ablated, or using "insulator" elements in the cassette, which can limit the activation of genes surrounding the insertion site. However, the occurrence of malignancy in the X-linked SCID trials led to a transition to lentiviral vectors. These vectors efficiently transduce nondividing target cells and are characterized by a different pattern of integration into the genome that appears to be safer than retroviral vectors. However, recent developments in the field of CAR-T therapy, notably reports of development of T-cell lymphoma, have underscored the need for caution (*vide infra*).

Transfusion-Dependent Thalassemia: Extension of Principle Therapeutic success for inherited immunodeficiencies, though a clear unmet medical need, affects only a very small population. The success of gene therapy in β thalassemia, one of the most common genetic diseases in Asian and Mediterranean populations, and one that provided the foundation for success in sickle cell disease, the most common genetic disease in Africans, demonstrated conclusively the therapeutic impact of gene therapy. The red cell disorders β thalassemia and sickle cell disease are more challenging targets for gene therapy than the immunodeficiencies for several reasons. First, in immunodeficiency disorders, the transduced stem cells have a survival advantage over nontransduced cells, which is not the case in red cell disorders (although the fully differentiated gene-modified red blood cells [RBCs] have a survival advantage compared to thalassemic or sickle RBCs). Second, in order to achieve transfusion independence or freedom from vaso-occlusive crises, one must achieve higher transduction efficiency as well as engraftment of higher numbers of stem cells. There are now two approved products, one a lentiviral-based gene therapy and the other a gene editing approach, for both of these conditions (Table 483-2). Standard of care for transfusion-dependent β thalassemia (TDT) consists of lifelong regular RBC transfusions, typically monthly, to support hemoglobin (Hgb) levels >9 g/dL, coupled with an intensive regimen of iron chelation to minimize iron overload to the liver, heart, and endocrine system **(Chap. 103)**. Allogeneic stem cell transplantation addresses the underlying cause of the disease but carries risks of myeloablation, graft-versus-host disease (GVHD), and graft rejection and thus is reserved primarily for those with an human leukocyte antigen (HLA)-matched sibling donor (<25% of patients). The first approved gene therapy for β thalassemia consists of a lentiviral vector driving expression of an antisickling variant of β-globin (β^{T87Q}, the same product used for sickle cell disease), introduced into autologous HSCs, which are then transplanted back into the patient after myeloablation. Results of clinical trials for both β^0/β^0 genotype (the most severe) and for non-β^0/β^0 showed durable transfusion independence, defined as Hgb ≥9 g/dL and no transfusion for ≥12 months, in 20 of 22 evaluable participants in one phase 3 study and 12 of 14 patients in a second study. The remaining subjects all demonstrated reduction in the transfusion requirement, enabling iron removal therapy by either phlebotomy or iron chelation and removing risks related to iron overload. Gene therapy with lentiviral transduction of autologous cells thus dramatically simplifies the medical regimen for these patients, since it eliminates the need for ongoing transfusion and iron chelation and carries no risk of GVHD or graft rejection because it is generated from the patient's own cells. Similarly, since the transduced cells are autologous, there is no requirement for an HLA-matched donor, expanding the numbers of patients who can be treated. Safety in the initial trials has been excellent, with most adverse events related to the known risks of the myeloablative conditioning regimen.

The same lentiviral vector is also now approved for sickle cell disease; a single-arm, 24-month, open-label study assessed 36 participants who underwent apheresis followed by myeloablative conditioning and transplantation of gene-modified autologous cells. Of the 32 evaluable patients, 30 achieved complete elimination of vaso-occlusive crises between 6 and 18 months after infusion, a key efficacy endpoint in the study, and 31 of 36 achieved a globin response defined as hemoglobin A^{T87Q} (the transgene product) of at least 30% of total Hgb and an increase in total Hgb of ≥3 g/dL. Using an earlier version of this product that was prepared using a different manufacturing process and a different transplant procedure, two patients died following development of acute myeloid leukemia. Interpretation of these data is not straightforward, since patients with sickle cell disease have an increased risk of hematologic malignancy compared to the general population. The product carries a boxed warning summarizing this risk; twice yearly monitoring of a complete blood count is recommended.

Neurodegenerative Disease: Broadening of Principle The SCID trials gave support to the hypothesis that gene transfer into HSCs could be used to treat any disease for which allogeneic bone marrow transplantation was therapeutic. Moreover, the use of genetically modified autologous cells carried the advantages noted above, i.e., no risk of GVHD, guaranteed availability of a "donor" (unless the disease itself damages the stem cell population of the patient), and low likelihood of failure of engraftment. Investigators in Paris capitalized on this realization to conduct the first trial of lentiviral vector transduction of HSCs for a neurodegenerative disorder, X-linked adrenoleukodystrophy (ALD). The key to the mechanism of action is that a subpopulation of the gene-modified cells gives rise to myeloid cells that cross the blood-brain barrier and engraft as central nervous system (CNS)-resident microglia and perivascular CNS macrophages. The transduced cells carry the gene encoding the missing protein, in this case an adenosine triphosphate–binding cassette transporter (Table 483-2). Following lentiviral transduction of autologous HSCs in young boys with the disease, dramatic stabilization of disease occurred, demonstrating that stem cell transduction could work for neurodegenerative as well as immunologic disorders.

Investigators in Milan carried this observation one step further to develop a treatment for another pediatric neurodegenerative disorder that had previously responded poorly to bone marrow transplantation. Metachromatic leukodystrophy is a lysosomal storage disorder caused by mutations in the gene encoding arylsulfatase A (ARSA). The late infantile form of the disease is characterized by progressive motor and cognitive impairment and death within a few years of onset, due to accumulation of the ARSA substrate sulfatide in oligodendrocytes, microglia, and some neurons. Recognizing that endogenous levels of production of ARSA were too low to provide cross-correction by allogeneic transplant, a lentiviral vector was used to create supraphysiologic levels of ARSA expression in transduced cells. Transduction of autologous HSCs from children born with the disease, at a point when they were still presymptomatic, has led to preservation and continued acquisition of motor and cognitive milestones at time periods as long as 8 years after treatment, with observation ongoing. This product is approved in Europe and the United States for those with late infantile or early juvenile forms of the disease (Table 483-2). These results illustrate that the ability to engineer levels of expression can allow gene therapy approaches to succeed where allogeneic bone marrow transplantation cannot. A similar approach may be useful in other neurodegenerative conditions.

■ EX VIVO GENOME EDITING

The first approved genome editing product, and those furthest along in clinical development, all use strategies that require only a cleavage event, rather than both a cleavage and a targeting event. For sickle cell disease **(Chap. 103)**, the genome editing strategy is carried out ex vivo in HSCs. A Cas9/guide RNA ribonucleoprotein complex targeting the erythroid enhancer of the *BCL11A* gene, which normally represses γ-globin (the fetal β-like globin) during the fetal-to-adult β-globin switch, is introduced into autologous CD34+ cells of patients with TDT or sickle cell disease. Reduced *BCL11*A expression results in increased γ-globin expression and Hgb F production in erythroid cells, reducing the Hgb S levels and preventing sickling. The product was approved based on a study of 44 children and adults (age range 12–34 years) with sickle cell disease; of these, 31 had been followed for at least 16 months after gene editing, myeloablation, and engraftment. Of the 31 participants with adequate follow-up, 29 achieved the primary efficacy endpoint of at least 12 consecutive months without any protocol-defined vaso-occlusive crisis. The mean total Hgb at month 18 was 13.3 g/dL, compared to a baseline mean of 7.5 g/dL. This product is now approved for children (12 and older) and adults with sickle cell disease and recurrent vaso-occlusive crisis.

Other genome editing strategies, in earlier stages of clinical development, use Cas9/guide RNA to introduce a cleavage within the β-globin locus near the site of the sickle mutation and simultaneously supply a targeting vector (in this case, an AAV6 vector) encoding a short sequence of the wild-type β-globin gene. In a process dependent on the cellular homology-directed repair pathway, this sequence is used as a DNA repair template at the site of the break, resulting in replacement of the mutant β^s sequence with the wild-type sequence.

■ LONG-TERM EXPRESSION IN GENETIC DISEASE: IN VIVO GENE TRANSFER WITH RECOMBINANT ADENO-ASSOCIATED VIRAL VECTORS

Recombinant adeno-associated viral (AAV) vectors have emerged as attractive gene delivery vehicles for genetic disease. Engineered from a small replication-defective DNA virus, they are devoid of viral coding sequences and trigger very little immune response in experimental animals. They are capable of transducing nondividing target cells, and the donated DNA is stabilized primarily in an episomal form, thus minimizing risks arising from insertional mutagenesis. Because the vector has a tropism for certain long-lived cell types, such as skeletal muscle, neurons, and hepatocytes, long-term expression can be achieved even in the absence of integration. Of note, because the donated DNA is predominantly nonintegrated, long-term expression requires targeting of nondividing or slowly dividing cells; otherwise, expression is lost as cells divide. The other shortcoming of AAV as a gene delivery vehicle is that it cannot package inserts of more than ~5 kb, owing to the fact that the wild-type viral genome is only ~4.7 kb; fortunately, with some notable exceptions, most cDNAs fall below this limit.

First Approved Products for Ultra-Rare Diseases As was the case with ex vivo gene transfer, the first approved products for in vivo gene therapy were for ultra-rare disorders. In the Western world, the first approved gene therapy product for genetic disease was an AAV vector, conditionally approved in Europe in 2012, for treatment of the autosomal recessive disorder lipoprotein lipase deficiency. The sponsor allowed the approval to expire in 2017, without completing the postmarketing requirements, but the initial approval was a crucial catalyst for the current robust activity in gene therapy, and AAV vectors are now the largest category of approved products for genetic disease, including products for spinal muscular atrophy type 1, a rare form of genetic blindness, hemophilia A and B, and Duchenne muscular dystrophy (Table 483-2).

The first approved AAV therapy for genetic disease in the United States was also for an ultra-rare disease, an inherited retinal dystrophy due to mutations in the gene encoding retinal pigment epithelial-associated 65-kDa protein (*RPE65*). The retina is an attractive target for AAV-mediated gene transfer. It is a relatively immunoprivileged space, obviating problems related to immune responses, and the photoreceptors, retinal ganglion cells, and retinal pigment epithelial cells are all long-lived postmitotic cells. Routes of administration for these cell types—either intravitreal or by subretinal injection—involve standard procedures in ophthalmology. Given the small space, doses required are relatively low, lessening the manufacturing burden. Finally, canine models of a number of inherited retinal dystrophies have been well-characterized and faithfully model the human diseases. Work carried out in the 1990s had demonstrated that the canine disease could be reversed, with durable restoration of visual behavior, by subretinal injection of an AAV vector in dogs with a mutation in the gene encoding RPE65, an enzyme key to the visual cycle. Like the canine disease, the human disease is characterized by early-onset visual impairment, with most patients progressing to blindness over time. Phase 1 clinical trials by multiple groups established the safety of subretinal injections of an AAV vector expressing RPE65. A single phase 3 trial, the first randomized controlled trial in human gene therapy, demonstrated improvement in multiple measures of retinal and visual function. Of note, and likely to be a recurring theme as gene therapies address diseases for which there are no existing treatments, successful clinical development required the development and validation of a novel clinical endpoint that could measure improvements in functional vision. This product, the first licensed AAV gene therapy product in the United States, is now approved worldwide (Table 483-2). Trials for both inherited and complex acquired retinal disorders such as age-related macular degeneration, affecting millions worldwide, are now underway.

Hemophilia B: Addressing More Common Genetic Disorders and Unraveling the Human Immune Response to Systemically Administered AAV Hemophilia is the X-linked bleeding diathesis caused by mutations in the genes encoding factor VIII (hemophilia A) or factor IX (hemophilia B) **(Chap. 121)**. Current treatment relies on intravenous infusion of clotting factor concentrates or, as an alternative in hemophilia A, administration of a bispecific antibody that replaces the cofactor factor VIII by binding to the enzyme (factor IXa) and its substrate (factor X), resulting in a biologically active conformation. Gene therapy for hemophilia began with hemophilia B, a smaller patient population compared to hemophilia A, but the *F9* cDNA is smaller (2.8 kb) and is more easily accommodated in an AAV vector. The early vectors were successful at demonstrating long-term expression of factor IX at therapeutic levels in hemophilic mice and dogs when vector was delivered to the liver (hepatocytes are the normal site of synthesis of factor IX), but the initial clinical trials uncovered a plethora of problems not predicted by animal models that had to be addressed to allow successful development to proceed. Fortunately, the solutions to these problems were generalizable across multiple therapeutic indications that rely on systemic delivery of AAV vector (*vide infra*) (Table 483-4). Most complex among these was working out the human immune responses, both innate and adaptive, to the AAV vector. With systemic administration, the presence of neutralizing antibodies, harbored in a substantial portion of children and even more prevalent in adults, can prevent transduction before the target cells are reached, and the cellular immune response, not predicted by animal models, which are not natural hosts for AAV and thus lack memory T cells directed to the capsid, can result in the loss of the transduced cells in a matter of weeks after successful initial transduction. The preexisting antibodies can be screened for in advance, assuring that only those likely to benefit receive the therapy. The cellular immune response typically presents as asymptomatic transaminase elevation and concomitant loss of factor IX (or the transgene product) in the circulation. This response is dose-dependent and can be mitigated by the use of immunomodulatory agents such as glucocorticoids or by strategies that reduce the vector dose. The use of appropriately timed steroid treatment to dampen the cellular immune response resulted in the first report of durable factor IX expression in men with severe hemophilia B. The eventual widely adopted solution for hemophilia B gene therapy came from human genetics, specifically the use of a high specific activity variant of factor IX, factor IX Padua, that allowed a substantial reduction in vector dose and/or higher levels of circulating factor IX; this strategy has shown durable expression. The two currently approved hemophilia B gene therapies show an increase in mean factor IX levels well into the mild hemophilia range; these levels result in annualized bleeding rates that are noninferior to clotting factor prophylaxis in men with the disease.

Successful extension to hemophilia A required the use of a truncated version of the factor VIII cDNA. The only approved product drives therapeutic levels of expression, but durability has been less than that seen in gene therapy for hemophilia B.

Spinal Muscular Atrophy Type 1 Spinal muscular atrophy type 1 is the most common genetic cause of death in infancy and affects about 1 in 11,000 births. The disease is caused by autosomal recessive mutations in the *SMN1* gene, encoding survival motor neuron 1; affected infants undergo degeneration and loss of lower motor neurons, presenting as hypotonia, severe weakness, and failure to sit without support. In a large natural history study of untreated infants with the disease, by age 20 months, only 8% of patients with the disease were alive and free of ventilator support. In a phase 1 gene therapy study, intravenous infusion of an AAV vector (one with tropism for the nervous system [AAV9]) expressing SMN1 showed survival without ventilator support in 100% of participants (n = 15) at 20 months of age. A phase 3 trial was initiated, but the treatment was approved in the United States based on the efficacy data from the first 21 participants enrolled in that study, coupled with the safety data from the ongoing phase 3 and the completed phase 1 study (Table 483-2). The major safety concern was the risk of acute serious liver injury; because the vector is infused intravenously, there is considerable biodistribution to the liver (and to the spinal motor neurons, the therapeutic target). The approved dose of 1.1×10^{14} vector genomes/kg is quite high, and results in marked elevation of liver transaminases if untreated. Leveraging the results in the early hemophilia trials, the phase 1 study showed that the liver toxicity could be controlled using a course of corticosteroids begun 1 day before the vector infusion and continued for 30 days, with tapering begun at that point and carried out with monitoring of liver transaminases. Postmarketing studies revealed an additional rare adverse reaction, thrombotic microangiopathy (TMA). Presenting as thrombocytopenia, microangiopathic hemolytic anemia, and acute kidney injury, this constellation of findings in the setting of AAV gene therapy was first described in the AAV gene therapy trials for Duchenne muscular dystrophy, which also use very high doses of vector. When it occurs, TMA appears early after vector infusion, is complement-mediated, and has responded to complement inhibitors, i.e., eculizumab. Patients receiving onasemnogene abeparvovec are currently followed in a registry designed to assess effectiveness, long-term safety, and overall survival of patients with Spinal muscular atrophy.

IN VIVO GENOME EDITING

Clinical trials using ex vivo genome editing of HSCs by CRISPR/Cas9 systems have been ongoing for several years and have now led to the first approved genome editing product (*vide supra*). More recently, investigators have begun to explore the feasibility of in vivo genome editing, with intriguing results. As has been the case with ex vivo editing, these initial trials all center on strategies that require a cleavage step only, i.e., not a cleavage event followed by a targeting event. For in vivo editing of hepatocytes, lipid nanoparticles containing an mRNA encoding the Cas9 protein and a single guide RNA directed to the gene of interest are infused intravenously. The major safety concern involves the risk of off-target editing; an additional concern prior to clinical trials was whether a robust immune response to the bacterial Cas9 enzyme would result in unacceptable toxicity or loss of efficacy. Published experience to date does show mild, transient, dose-dependent rises in liver transaminases, but also robust efficacy in reducing circulating levels of transgene products of interest, including transthyretin in transthyretin amyloidosis and plasma kallikrein B1 in hereditary angioedema. Additional in vivo genome editing trials are underway, notably for cardiovascular conditions, to reduce circulating levels of lipoprotein(a), proprotein convertase subtilisin/kexin type 9 (PCSK9), or angiopoietin-like protein 3, all associated with atherosclerotic cardiovascular disease. An advantage of genome editing approaches is that the edit will be passed to every daughter cell; thus, changes should persist over time, and this treatment can be used even in children, without fear that the edit will be lost as the liver grows in size.

GENE THERAPY FOR CANCER

The majority of clinical gene transfer experience has been in subjects with cancer. The intent has been to increase the precision of cancer therapies and thereby make them less toxic and more effective. Most approaches have either modified the tumor directly or altered the host's response to the malignancy to produce immune effector cells that are precisely targeted to the tumor phenotype.

MODIFYING THE CANCER

Since cancer is an (acquired) genetic disorder, initial efforts were directed at correcting the genetic deficits of the tumor or introducing lethal genes. Two major and persisting obstacles, however, are the poor biodistribution and transduction efficiency of all currently available vectors, and the heterogeneity and genetic instability of the tumor targets themselves, so that correction of single driver mutations does not preclude the evolution of a resistant population.

Tumor Correction One widely used direct intratumoral approach was adenoviral-mediated expression of the tumor suppressor p53, which is mutated in many different cancers. Initial studies showed some complete and partial responses in squamous cell carcinoma of the head and neck, esophageal cancer, and non-small cell lung cancer, but as yet, there have been no successful product licensing studies for this approach except in China.

Prodrug Metabolizing Genes Efforts to overcome the above limitations have included the introduction of a prodrug or a suicide gene that would increase sensitivity of tumor cells to cytotoxic drugs. A strategy used early on was intratumoral injection of an adenoviral vector expressing the thymidine kinase (*TK*) gene. Cells that take up and express the *TK* gene can be killed after the administration of ganciclovir, which is phosphorylated to a toxic nucleoside by *TK*. The advantage of this approach is that the effects of transducing even a limited number of tumor cells are amplified by the spread of active drug to adjacent tumor cells. Although the approach continues to be examined in aggressive brain tumors and locally recurrent prostate, breast, and colon tumors, progress remains slow, and systemic benefits against metastatic disease have not been established.

MODIFYING THE HOST

Recruiting the Immune System The successful use of monoclonal antibodies that produce antitumor activity by activating the immune response has demonstrated the feasibility of manipulating the immune system to recognize the abnormal pattern of antigen expression on tumor cells. Immune cells are capable of almost unlimited expansion and persistence and can provide long-term tumor control. They can also traffic to tumor sites irrespective of location and, in principle, have the potential to evolve with the changing pattern of tumor cell phenotype and function. Antibodies targeting "checkpoint" molecules, particularly CTLA-4 and the PD-1/PD-L1 axis, which naturally limit T-cell responses and maintain tolerance, have been particularly successful.

Vaccination This strategy promotes more efficient recognition of tumor cells by the immune system, but the development of a therapeutic as opposed to the preventative vaccines required to combat infectious diseases has proved to be a considerable challenge. Approaches have included direct injection of tumor or tumor-antigen–derived RNA or DNA; transduction of tumor cells with immune-enhancing genes encoding cytokines, chemokines, or co-stimulatory molecules; and the ex vivo manipulation of dendritic cells to enhance the presentation of tumor antigens. A dendritic cell vaccine for treatment of recurrent prostate cancer has received approval in the United States, but its limited potency and high cost have constrained commercial success.

Adoptive Cell Transfer Host immune cells such as T cells, NK cells, and others can be modified to express new transgenic receptors intended to recognize tumor cells and their microenvironment (Fig. 483-1). Retargeting may use a modification of the cells' own receptor or a molecularly synthesized CAR that is usually composed of the antigen recognition portion of an antibody and the signaling components of the cell's native antigen receptor along with one or more additional signaling domains that boost T-cell activation. Both approaches have been successful, with significant responses reported with native receptors targeted to melanoma and synovial cell sarcoma and—most dramatically—with CARs targeted to CD19, an antigen expressed at high levels on normal and many malignant B cells, or B-cell maturation antigen (BCMA), an antigen expressed at high levels in normal and multiple myeloma plasma cells. Infused CAR T cells can expand many thousand-fold in vivo, persist long term, and have produced >80% complete response rates when targeting intractable B-cell acute lymphoblastic leukemia; approximately half of those patients have remained in remission for many years afterward without further cancer therapies (i.e., have been "cured"). This approach has also been successful in adult patients with relapsed or chemotherapy-refractory B-cell–derived large cell lymphoma, mantle cell lymphoma, and multiple myeloma. Many responses are sustained long term, and several of these CAR T-cell products have been approved by the U.S. Food and Drug Administration (FDA), as well as international regulatory agencies, and adopted as standard of care. In 2021, the first reports of T-cell lymphoma were reported with the use of piggyBac-modified (transposon) CAR-T therapy; 2 of 10 patients developed CAR-T lymphoma, but this may have been a result of widespread copy number variations and multiple insertions, and there was no apparent evidence of insertional mutagenesis. In November 2023, the FDA issued a warning for most approved CAR-T therapies, all of which use retroviral or lentiviral vectors for gene delivery of the CAR, suggesting that there was a higher-than-expected rate of T-cell lymphoma and that patients treated with CAR-T therapy should be monitored for secondary malignancies for life. Nevertheless, the overall benefits of these products continued to outweigh their potential risks for their approved uses. Although the FDA did not list the frequency, a flurry of subsequent reports noted that the rate of T-cell malignancies after CAR-T treatment for patients with relapsed or refractory hematologic malignances was ~20 in 34,000 patients (~0.06%); however, because reporting to the FDA on the incidence of T-cell malignancies is voluntary once products are approved, it is possibly an underestimated frequency. A boxed warning listing the possibility of secondary malignancies was added to the label of most CAR-T products in January 2024, and the FDA published instructions on reporting secondary malignancies to the community in order to gather more information. It is important to keep in mind that the vast majority of secondary malignancies in patients treated with approved CAR-T therapies are unrelated to the presence of the CAR transgene, including solid tumors and myeloid malignancies; such secondary malignancies may be related to age, the presence of clonal hematopoiesis, and exposure to prior chemotherapy or other antineoplastic treatments.

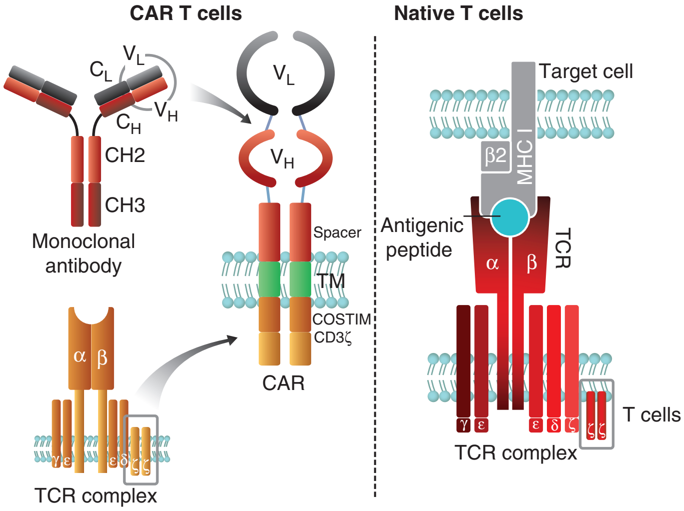

FIGURE 483-1 T-cell receptors. A native T-cell receptor (TCR) recognizes processed peptide antigens bound to major histocompatibility (MHC) molecules through its αβ chains. Signaling then occurs through a multichain intracellular CD3 complex. A chimeric antigen receptor (CAR) usually contains an extracellular receptor component derived from the antigen binding portion (V_H and V_L) of a monoclonal antibody. This produces a receptor that can recognize either protein or nonprotein antigens independent of the MHC. A transmembrane (TM) domain then connects this receptor to the ζ chain of the CD3 complex derived from the native TCR. A costimulation domain (COSTIM), such as CD28 or 4-1BB, is also present.

The general approach of CAR-T therapy is under rapid development, including trials with CAR T cells targeting different antigens for solid tumors and other hematologic malignancies and CAR T cells with different molecular structures and different gene transfer vectors. Remaining challenges in the field and application of adoptive T-cell approaches include the following: (1) the immune inhibitory microenvironment associated with most solid tumors, and recent studies further modify the T cells with countermeasures to tumor inhibitory signals; (2) acute and severe (though rarely fatal) systemic inflammatory and neurologic toxicities during the phase of T-cell expansion and tumor killing, which typically require access to intensive care for clinical management; (3) the off-target or on-target but off-tumor effects that may damage normal host tissues (e.g., normal B cells following CD19 CAR therapy); and (4) the cost, time, and complexity of manufacture, which are particular problems when antigens unique to each tumor's individual mutations are targeted (neoantigens), rather than widely shared tumor-associated antigens.

Nonimmunologic Modifications to Host Gene transfer can be used to protect normal cells from the toxicities of chemotherapy and thereby increase the therapeutic index of these drugs. The most extensively studied approach has been to transduce hematopoietic cells with genes encoding resistance to chemotherapeutic agents, including the multidrug resistance gene *MDR1* or the gene encoding O^6-methylguanine DNA methyltransferase (*MGMT*). Although such approaches reduce hematologic toxicity, cytotoxic dose escalation quickly reveals dose-limiting toxicities to other organ systems. Chemotherapy resistance can also be engineered into immune cells redirected to target cancer, to enable combination treatments with cells and chemotherapy.

Finally, gene transfer can be used to inhibit the host angiogenesis required for tumor support, for example by constitutive expression of inhibitors such as angiostatin and endostatin, or the transfer of T cells genetically modified to recognize antigens specific to newly forming vasculature. These studies are in early phases.

COMBINATION APPROACHES: MODIFICATION OF HOST AND TUMOR BY VIROTHERAPY—IMMUNO-ONCOLYTIC VIRUSES

These viruses are genetically modified to replicate in malignant but not normal cells. The replicating vectors thus proliferate and spread within the tumor, facilitating eventual tumor clearance. However, physical limitations to viral spread, including fibrosis, intermixed normal cells, basement membranes, and necrotic areas within the tumor, may reduce clinical efficacy, and their activity against metastatic disease has proved limited. Recently, the FDA granted licensing approval to talimogene laherparepvec, an oncolytic herpes virus containing the human granulocyte-macrophage colony-stimulating factor (GM-CSF) gene, for treatment of melanoma. This success has led to resurgent interest in combining the local tumor destruction and tumor antigen release mediated directly by oncolytic viruses with the recruitment of a systemic immune response mediated by immunostimulatory genes contained within the oncolytic virus. In principle, such immune-oncolytic viruses should produce responses in both local and metastatic disease. Numerous novel viral agents are now entering early-phase clinical testing.

SUMMARY AND FUTURE DIRECTIONS

Cell and gene therapies have progressed from halting beginnings to the current status as the fastest-growing sector of medicine and the health care industry. The speed of technical evolution results in a continuously changing landscape. Key advances likely to assume greater importance in the coming decade include bioengineered AAV capsids selected for tropism and increased expression, leading to lower doses, fewer adverse events, and lower manufacturing burden; extension of therapeutic trials from single-gene disorders to complex acquired disorders, including chronic heart failure, age-related macular degeneration, and Alzheimer's disease; continued expansion of in vivo genome editing; and application of CAR-T technology to solid tumors and autoimmune disorders (e.g., systemic lupus erythematosus). The power and versatility of gene therapy approaches are such that there are few serious disease entities for which gene therapies are *not* under development. Approved products and examples of clinical success are now abundant, and cell and gene therapies are likely to become increasingly important as therapeutic modalities in the twenty-first century. Realization of the therapeutic benefits of modern molecular medicine will depend on continued progress in cell and gene therapy technology.

FURTHER READING

Al-Zaidy SA et al: AVXS-101 (onasemnogene abeparvovec) for SMA1: Comparative study with a prospective natural history cohort. J Neuromuscular Dis 6:307, 2019.

Frangoul H et al: CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia. N Engl J Med 384:252, 2021.

Fumagalli F et al: Lentiviral haematopoietic stem-cell gene therapy for early-onset metachromatic leukodystrophy: Long-term results from a non-randomised, open-label, phase 1/2 trial and expanded access. Lancet 399:372, 2022.

High KA, Roncarolo MG: Gene therapy. N Engl J Med 381:455, 2019.

June CH, Sadelain M: Chimeric antigen receptor therapy. N Engl J Med 379:64, 2018.

Kanter J et al: Biologic and clinical efficacy of LentiGlobin for sickle cell disease. N Engl J Med 386:617, 2022.

Larson RC, Maus MV: Recent advances and discoveries in the mechanisms and functions of CAR T cells. Nat Rev Cancer 21:145, 2021.

Longhurst HJ et al: CRISPR-Cas9 in vivo gene editing of *KLKB1* for hereditary angioedema. N Engl J Med 390:432, 2024.

Pipe SW et al: Gene therapy with etranacogene dezaparvovec for hemophilia B. N Engl J Med 388:706, 2023.

Ruella M et al: Mechanisms of resistance to chimeric antigen receptor-T cells in haematological malignancies. Nat Rev Drug Discov 22:976, 2023.

Tabebordbar M et al: Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species. Cell 184:4919, 2021.

Verdun N, Marks P: Secondary cancers after chimeric antigen receptor T-cell therapy. N Engl J Med 390:584, 2024.

484

The Human Microbiome in Health and Disease

Neeraj K. Surana, Dennis L. Kasper

> *"All disease begins in the gut."*
>
> —Hippocrates

Nearly two and a half millennia after Hippocrates made this statement, we are just coming to truly appreciate its profundity. Since the beginning of humankind, scholars have been investigating the underpinnings of disease with an almost singular focus on the human side of the equation. Microbes were not recognized as an important cause of disease until the inception of the "germ theory" in the late nineteenth century. During the first century of medical microbiology, research largely centered on the role of microbes as pathogens. Only recently has there been a resurgence of interest in understanding how commensal organisms—the bacteria, viruses, fungi, and Archaea that make up the *microbiota*—impact human physiology. The idea that these microorganisms are vital to the well-being of humans has challenged our traditional notions of "self." Indeed, a human being can most accurately be described as a *holobiont*: a complex assemblage of human cells and microorganisms interacting in an elaborate *pas de deux* that drives normal physiologic processes.

Aimed at a better understanding of this relationship, myriad studies during the past decade have begun to catalogue the microbiota at various body sites and in a multitude of disease conditions. Diseases in virtually every organ system have been associated with changes in the microbiota. Indeed, the microbiota has been linked to intestinal disorders, disturbances in metabolic function, autoimmune diseases, and psychiatric conditions and has been shown to influence susceptibility to infection and the efficacy of pharmaceutical therapies. Knowledge of the specific mechanism(s) underlying most of these microbe–disease associations is lacking; it remains unclear whether the disease-associated alterations in the microbiota represent mere biomarkers of disease, a causal relationship, or a combination of the two. Although cause-and-effect relationships are still being elucidated for many diseases, it is clear that humans coexist in an intricate relationship with commensal organisms. This chapter explores in detail the nature of these host–commensal interactions, focusing on how this information might be translated into clinically meaningful interventions.

HISTORICAL PERSPECTIVE

Massive undertakings, such as the Human Microbiome Project (HMP) sponsored by the National Institutes of Health and MetaHIT sponsored by the European Commission, have catalogued all the bacteria present at multiple body sites in people with and without disease. Coupled with the confluence of advances in sequencing technologies **(Chap. 126)**, gnotobiotic animal availability, and microbial culture, significant progress has been made toward an understanding of the interplay between the microbiota and human health. However, current findings were foreshadowed by work done centuries ago.

The human microbiota was first explored in 1683 when Antony van Leeuwenhoek described in a letter to the Royal Society of London the "very little living animalcules, very prettily a-moving" that he had observed in the plaque between his teeth. Leeuwenhoek went on to perform the first comparative "microbiota" studies by assessing how fecal and oral bacteria differ, how oral microbes change in the setting of disease (e.g., alcoholism and tobacco use), and how microbial composition changes across the age spectrum (e.g., in young children vs old men). He attempted—unsuccessfully—to eliminate these bacteria. Although Leeuwenhoek was not taken seriously when he first reported his findings, his studies laid the groundwork for what is now the field of microbiome research, and investigators are still trying to answer many of the same overarching questions that he raised more than three centuries ago.

Although Leeuwenhoek first reported the existence of bacteria and their association with humans at the end of the seventeenth century, the significance of commensal bacteria was not realized until late in the nineteenth century. In 1885, Pasteur suggested that animals could not survive if they were "artificially and completely deprived of the common microbes." Although Pasteur's preconceived ideas were proven incorrect in 1912 by the advent of germ-free (GF) animals (animals raised without exposure to any microorganisms), the underlying concept that commensal organisms are critical to health has held up. Élie Metchnikoff made another conceptual advance in this field by suggesting at the beginning of the twentieth century that clinical outcomes could be altered by the administration of specific beneficial organisms (*probiotics*). In particular, Metchnikoff believed that aging was caused by toxic bacteria in the gut and that lactic acid–producing bacteria (e.g., *Lactobacillus* species) present in sour milk and yogurt could mitigate against this process. The data behind this specific claim are still lacking, but contemporary discoveries offer continued hope that the microbiome can be effectively harnessed to protect against and treat a variety of diseases. Thus, although the field of microbiome research is sometimes considered to have emerged over the last two decades, the basic tenets—that the microbiota varies according to body site and clinical characteristics, that microbes are critical for human health, and that specific modulation of the microbiota may lead to improved clinical outcomes—are far from new.

A PRIMER ON TAXONOMY

Given that microbiome-based studies have identified and compared microbes at different levels of taxonomic resolution **(Fig. 484-1)**, some understanding of taxonomy is essential for better comprehension of the implications of these studies. Of the ~100 bacterial phyla that exist in nature, only five (Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, and Proteobacteria) are dominant members of the human microbiome. Each of these phyla can be further categorized into multiple classes, orders, families, genera, and species. Early studies on the microbiota focused on changes in the relative abundance at the phylum level between different groups (e.g., obese vs normal-weight patients); however, these comparisons are at such a broad taxonomic level that they often provide little or no biologic insight. As illustrated in Fig. 484-1, drawing comparisons between organisms in two different bacterial phyla is analogous to comparing humans to sea stars: the evolutionary distance between the two is tremendous. Examining microbial profiles at the phylum, family, or even genus level—as is often done at present—ignores the great heterogeneity within different strains of the same bacterial species. The analytical pipelines are beginning to enable strain-level comparisons, and these improvements will likely facilitate our ongoing investigation of host–commensal interactions.

THE MICROBIOTA AND HUMAN HEALTH

■ OVERVIEW OF THE HUMAN MICROBIOTA

The overwhelming majority of microbiota studies have focused on stool, given that this sample type represents the most ecologically rich anatomic site, is easy to obtain, and can readily be followed longitudinally in the same individual. A landmark study by the HMP sought to define the "normal" microbiota throughout the entire body in healthy Western adults. To this end, the microbial populations at 15–18 body sites were characterized in 242 people. One striking finding was that all samples from a given body region (e.g., skin) were more similar to each other than they were to samples from a different body region (e.g., stool), even in the same individual **(Fig. 484-2*A*)**. In essence, the effect of the anatomic site on microbial composition is far greater than the effect of heterogeneity between individuals. That said, there was a remarkable amount of interindividual variation at any given body site **(Fig. 484-2*B*)**. In stool, for example, the abundance of the phylum Bacteroidetes ranged from ~10% in some individuals to >90% in others. Remarkably, even with person-to-person variability and differences among body sites, the functional capacity of the microbiota—assessed using metagenomic data to identify gene pathways—was quite similar across different people and different body sites **(Fig. 484-2*C*)**. This discrepancy between the substantial differences in microbial composition and the little or no resulting change in the functional properties of the microbiota reflects an important ecologic property of the microbiota: the microbial communities at different body sites and in different people assemble in such a way that all the core metabolic functions are maintained. This finding also hints at the likely possibility of significant functional redundancy within the microbiota, with different species executing the same biologic functions in different people and/or at different anatomic sites.

While the HMP provided the first large-scale catalogue of the microbiome in multiple people and at many different body sites, the amount of data generated by what, at the time, was by far the largest microbiome study has been dwarfed by subsequent studies. These more recent studies have confirmed the HMP's major tenets: the composition of the microbiota differs by body site, there is tremendous interindividual variation, and the microbial gene content is relatively conserved irrespective of the body site or individual. No microbial species are ubiquitous in all individuals and at all body sites, but some species are highly prevalent at a given body site: in the HMP study, *Staphylococcus epidermidis* was present in 93% of nares samples and *Escherichia coli* in 61% of stool samples. These findings highlight the remarkable personalization of the human microbiome. While the human genome is typically >99.5% identical in different people, the microbiotas of two individuals may not overlap at all. Although the "precision medicine" approach currently focuses on teasing out how differences in the human genome relate to different clinical end

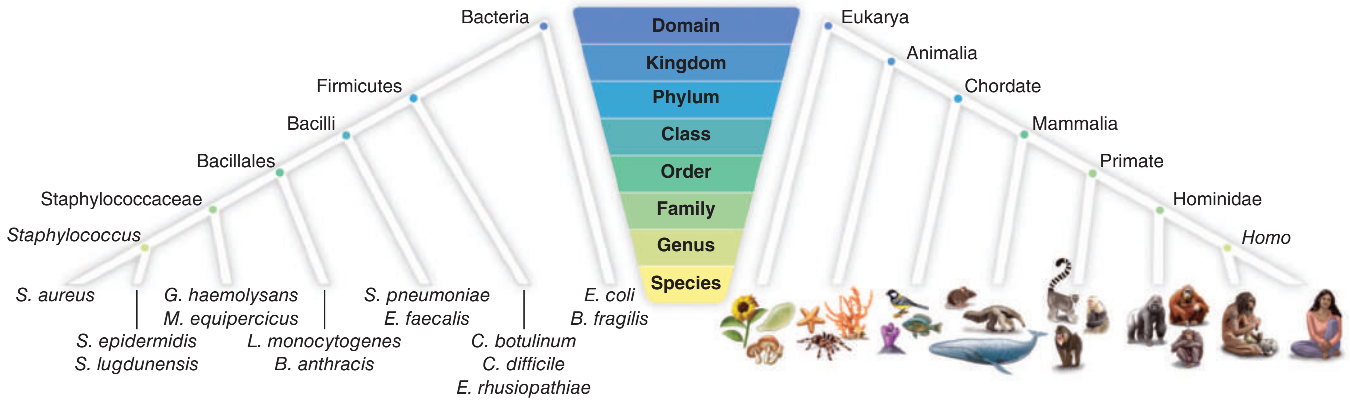

FIGURE 484-1 Juxtaposition of bacterial and human taxonomy highlights the evolutionary distance between different taxonomic levels. The listed species represent exemplars that are members of the taxon to which they are connected but that are not contained within the next-lower-level taxon listed. For example, *Clostridium botulinum*, *Clostridioides difficile*, and *Erysipelothrix rhusiopathiae* are members of the phylum Firmicutes, but are in classes other than Bacilli. Similarly, starfish and humans are both members of the kingdom Animalia, but they are in different phyla.

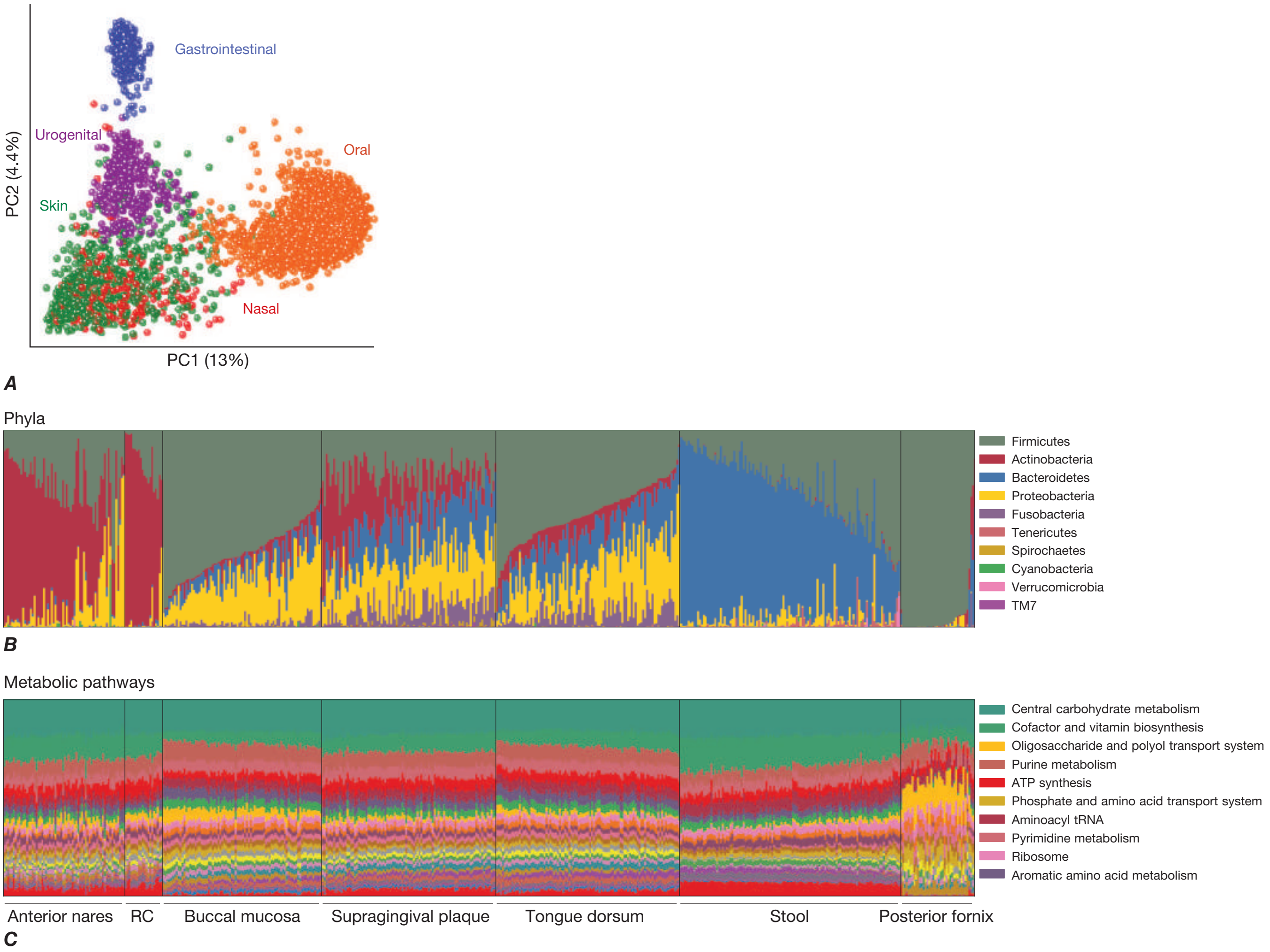

FIGURE 484-2 The human microbiome exhibits significant taxonomic variability among body sites and between individuals while maintaining core metabolic pathways. ***A.*** Principal coordinates (PC) plot showing variation among samples demonstrates that primary clustering is by body area, with the oral, gastrointestinal, skin, and urogenital habitats separate; the nares habitat bridges oral and skin habitats. Each circle represents an individual sample. ***B, C.*** Vertical bars represent microbiome samples by body habitat, with each bar within a given body site representing a different individual. Bars indicate relative abundances colored by microbial phyla (***B***) and metabolic pathways (***C***). The legend on the right indicates the most abundant phyla/pathways. RC, retroauricular crease. *(Reproduced with permission from Human Microbiome Project Consortium: Structure, function and diversity of the healthy human microbiome. Nature 486:207, 2012.)*

points, the human microbiome clearly represents a critical component for consideration.

THE MICROBIOTA BY THE NUMBERS

It has long been known that the human-associated microbiota is numerically dense. Leeuwenhoek estimated that there were more "animals living in the scrum on the teeth in man's mouth than there are men in a kingdom." Specific enumeration of the components of the microbiota has been challenging, in part because of its variability across time, space (body region), and clinical conditions. Moreover, the majority of human-associated microbes are not readily cultivable—a situation that raises questions about the best methodology for such quantitation. Initial back-of-the-envelope calculations performed in the 1970s suggested that there were roughly tenfold more bacteria in the body than there were human cells. This rather astounding estimate suggested that humans are really only ~10% "human" and that by far the greatest part of the holobiont is represented by microbes. This stark numerical discrepancy has prompted some to question "who parasitizes whom." However, it has been suggested that there are "only" ~1.3 times more bacteria in the body than there are human cells and thus that humans are ~56% "bacterial." Of note, this study does not include the numbers of viruses (known to generally be approximately tenfold more abundant than other microbes), fungi, or Archaea. Given these additional microorganisms, the notion that microbes constitute >90% of the cells present in a human body is likely correct. These ratios are even starker when one considers the genetic potential of human cells versus that of commensal organisms. In contrast to the ~20,000 genes in the human genome, the estimated total number of genes in the microbiota (which together constitute the *microbiome*)—i.e., >2,000,000—indicates that the human genome contributes <1% to the total genetic potential of the overall holobiont. Most microbiome studies have focused almost exclusively on the bacterial component; much remains to be learned about the functional interplay of bacteria, viruses, fungi, and Archaea and how these other classes of microorganisms impact human health.

In terms of overall diversity, >10,000 different bacterial species are present in the human microbiota; the intestines alone contain >1000 species. At any given time, the body of any given individual harbors 500–1000 bacterial species, with 100–200 bacterial species in the gut alone. If one considers different strains of the same bacterial species, which may be functionally different from one another, the diversity of the microbiota is probably at least an order of magnitude greater. Although marked diversity exists at the strain and species level, only limited bacterial phyla are generally found in the human microbiota at any given body site **(Fig. 484-3)**.

INFLUENCES ON THE MICROBIOTA

An individual's specific microbial configuration is dynamic and is quickly altered in response to subtle changes in the microenvironments in which the bacteria reside. On a day-to-day basis, these changes usually reflect alterations in the relative abundance of the various microbes. However, some exposures have a greater effect on

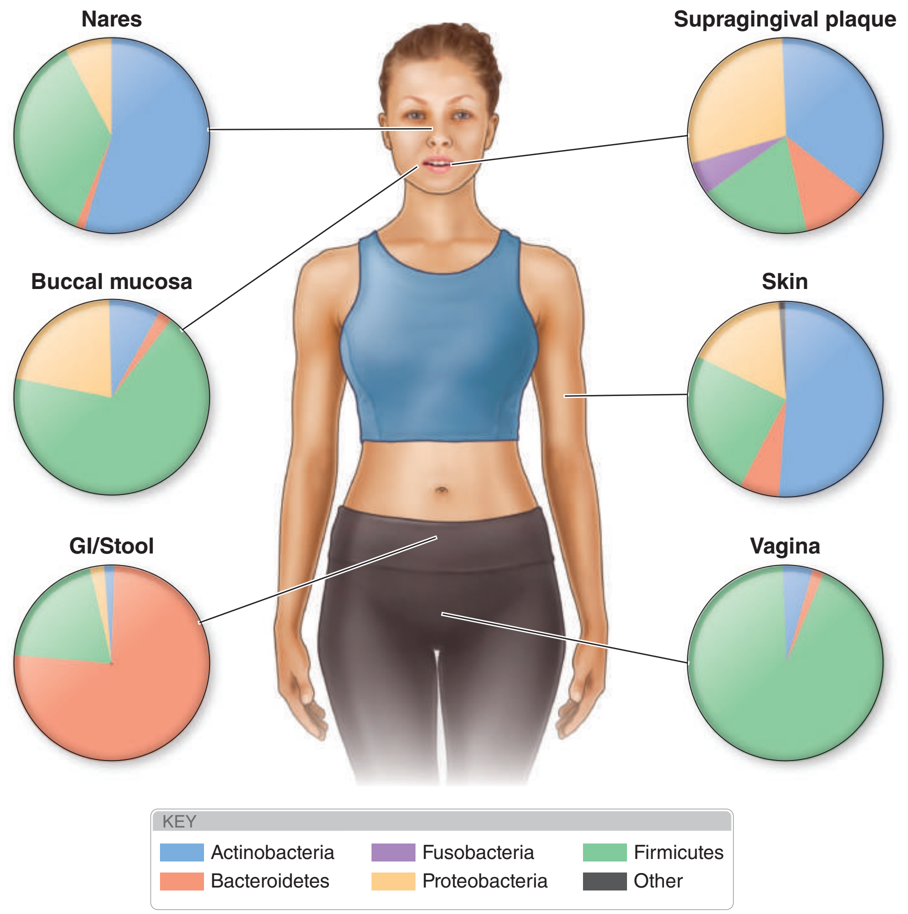

FIGURE 484-3 Different anatomic sites harbor very different microbiomes. The figure indicates the relative proportion of sequences determined at the taxonomic phylum level at six anatomic sites. *(Data for stool, vagina, nares, buccal mucosa, and supragingival plaque are from the Human Microbiome Project; data for the skin are from EA Grice et al: Topographical and temporal diversity of the human skin microbiome. Science 324:1190, 2009.)*

the microbiota and can shift the microbial population to a new equilibrium via the loss of specific species and/or the acquisition of others; this new microbial equilibrium can be associated with either health or a disease state **(Fig. 484-4)**. Identification of the factors that influence the microbiota's composition is critical to an understanding of what leads to and controls intra- and interindividual variation. Moreover, an understanding of the influences on the microbiota will facilitate the design and proper interpretation of microbiota studies. While it is clear that the microbiota can be altered through these various mechanisms, it is not yet clear whether these changes are biologically significant.

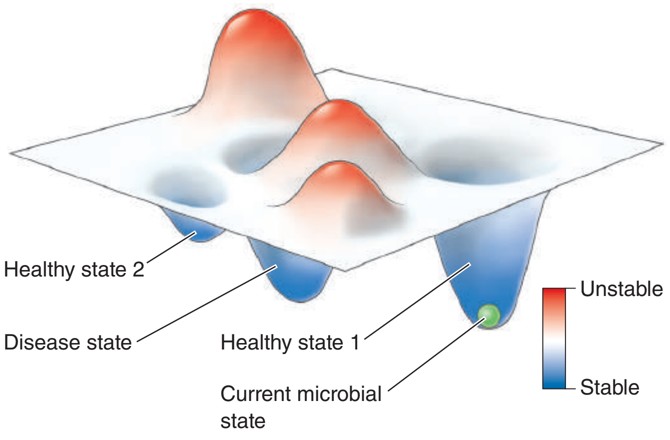

FIGURE 484-4 A stability landscape of the human microbial ecosystem. A stable state, illustrated as a depression in the landscape, can be associated with either a healthy state or a disease state. The topology of an individual's landscape reflects that person's genetics, age, diet, medications, medical history, and lifestyle. The position of the green ball represents the current microbial state. Clinical changes (e.g., administration of antibiotics, development of disease) can influence both the current state and the overall topology.

Genetics Studies of monozygotic and dizygotic twins have revealed that host genetics have a small but statistically significant effect on the microbiota's composition. Notably, some taxa, such as *Christensenella* species, are more heritable than others. A cross-sectional study of >1000 healthy individuals who have distinct ancestral origins and a relatively shared common environment confirmed a weak association between host genetics and the microbiome but highlighted that environmental factors are more prominent modulators of the microbiome. That said, the host's genetic contribution to the microbiota, albeit small, may be meaningful. Studies in mice have demonstrated that genetic variation in the major histocompatibility complex, a specific set of immune-related genes, leads to changes in the microbiota that alter susceptibility to an autoimmune disease. These studies offer a proof of concept for the notion that the genetic predisposition observed for certain diseases may actually be mediated by indirect alterations in the microbiota.

Age Burgeoning evidence now indicates that microbial exposure may begin in utero: bacterial DNA from bacteria typically associated with the oral microbiota has been identified in otherwise healthy placentas, in amniotic fluid obtained at early stages of gestation, and in meconium of term newborns. Although some controversy persists about whether these results reflect contamination and/or the presence of nonviable bacteria, they raise the possibility that human exposure to the microbial world begins before birth. The delivery mode (vaginal vs cesarean section) and the method of feeding (breast milk vs formula, timing of solid food introduction) are major determinants of an infant's early microbiota. After birth, the infant's microbiota goes through a stereotyped succession process; with increases in bacterial diversity and functional capacity, the child's microbiota resembles that of an adult by the age of 2–3 years. Cross-sectional studies that have examined the microbiota across the entire age spectrum have revealed a general stability of the fecal microbiota after 2–3 years of age; however, the microbiota of the elderly (persons >80 years of age) demonstrates notable differences from those of their younger counterparts, with increases in *Bacteroides* and *Eubacterium* species and decreases in the bacterial family Lachnospiraceae. Although there has been significant interest in defining microbial features that predispose towards longevity, there has been poor concordance of findings between studies, potentially due to very different populations being studied.

Diet Diet is a strong determinant of human health. The impact of diet is mediated, in part, by its effects on the composition of the gut microbiota. This makes intuitive sense, as the human diet provides nutrients needed not only by our own cells but also by the microbes living in the alimentary tract. In young children, this dietary influence is marked by major shifts (e.g., a decrease in *Bifidobacterium* species) in the intestinal microbiota that occur at weaning and with the introduction of solid food. In adults, long-term dietary patterns are associated with relatively stable microbial compositions. However, drastic changes in short-term macronutrient availability cause rapid (within 1 day) and reproducible fluctuations in the fecal microbiota that reflect the biologic processes needed to degrade and metabolize the nutrients in the new diet. For example, vegetarian diets are associated with a microbiota that has an increased ability to metabolize plant polysaccharides (e.g., *Roseburia* species, *Eubacterium rectale*, *Ruminococcus bromii*), while animal-based diets result in an increased abundance of bile-tolerant organisms (e.g., *Alistipes*, *Bilophila*, and *Bacteroides* species). At the completion of dietary interventions and the resumption of the individual's normal dietary pattern, the microbial communities revert back to their previous states, probably because the individual resumes their typical diet. Taken together, dietary studies confirm that the microbiota is highly adaptable and varies in relation to changes in the diet. Of note, virtually all these studies have focused on how the diet influences the fecal microbiota, with emerging evidence showing that it may similarly influence the microbiota at some nonintestinal sites.

Drugs Virtually all drugs have the capacity to change the microbiota by altering the chemical landscape in which the microorganisms live (e.g., statins, bile acid sequestrants), modulating the host's ability to recognize and react to microbes (e.g., immunosuppressants) and/or directly interfering with the microbiota's constituents (e.g., antibiotics). These potential effects have made critical interpretation of microbiota studies much more difficult. A prominent study that claimed to identify a fecal microbiota signature associated with type 2 diabetes was later found actually to have identified a signature for patients taking metformin instead; the effects of this drug on the microbiota were far greater than the effects of the disease itself. These results highlight the importance of controlling for clinical variables in microbiota studies.

Antibiotics are the most obvious and best-studied class of drugs that modulate the microbiota. Multiple groups have demonstrated that antibiotics exert a considerable effect on the gut microbiota by depleting antibiotic-sensitive strains. What is more surprising is that many strains resistant to the antibiotic tested are also eliminated. For example, treatment with ciprofloxacin, which has little to no activity against clinically relevant anaerobes, leads to a loss of roughly one-third of the bacterial taxa in the gut. This broad effect is likely mediated by the depletion of certain "keystone" species that are required for the persistence of other, unrelated species and highlights the intricate microbe–microbe interactions that are fundamental to maintenance of the overall microbial community. While many of the observed antibiotic effects (e.g., loss of specific taxa) are shared across many different individuals, some effects vary greatly among people. For example, studies found that microbiota recovery following antibiotic treatment differed significantly in terms of timing and degree. The microbiota of most healthy people who received ciprofloxacin for 5 days had completely recovered within 4 weeks, whereas microbiologic changes lasted up to 6 months in other individuals. Moreover, the degree of variation was compounded by repeated antibiotic administration, with fewer individuals reverting to their baseline microbiota after a second course of ciprofloxacin given 6 months after the first. These findings are consistent with those of microbial ecology experiments, which also showed that this type of repeated disturbance leads to less predictable results.

Lifestyle Many seemingly innocuous lifestyle decisions can impact the human microbiota. For example, a person's skin and fecal microbiotas are more similar to those of their household members, regardless of genetic relatedness, than to those of residents of different households. The degree of similarity in skin microbiotas is even greater if a dog also lives in the home; in contrast, the presence of a cat or a young child does not accentuate this microbial relatedness. The presumption is that the dog serves as a more effective "vector" for transmitting microbes during its frequent direct contact with adults in the household. The type of setting in which a person lives also impacts the microbiota. Living in a rural or farm setting leads to a different fecal microbiota than living in an urban environment. Similarly, the individual's country of residence affects the microbiota. An analysis of daily fecal samples from an individual who temporarily (i.e., for a couple of months) moved from the United States to Thailand demonstrated a large shift in the fecal microbiota that coincided with arrival in Thailand and a reversion in most respects to the "American" microbial configuration upon return to the United States. Similarly, immigration to the United States "westernizes" the microbiome of individuals coming from non-Western countries. These geography-driven changes probably reflect a combination of environmental and dietary differences between locations.

Circadian Rhythms Many human biologic processes follow a circadian clock; aspects of physiology are tuned by external cues, including the degree and timing of ambient light, temperature, and availability of nutrients. This endogenous biologic clock enables animals to efficiently adapt to changing environmental conditions. Similarly, the microbiota maintains a circadian rhythm that is linked to—and helps entrain—the host's circadian clock. If circadian oscillations are disrupted in the host, they are similarly disrupted in the microbiota, and vice versa. These bacterial vacillations occur at the level of spatial localization within the intestine, relative species abundance, and bacterial metabolite secretion. Work in the 1960s showed that mice exhibited daily periodicity of susceptibility to infection with either *Streptococcus pneumoniae* or *E. coli* lipopolysaccharide (LPS). Although the fundamental basis for this difference was not known at the time, it is likely to be related, in part, to the microbial circadian clock. Derangements of these microbial oscillations have also been linked to the development of metabolic diseases and may underlie some of the health hazards associated with shift work and jet lag.

THE MICROBIOTA AND DISEASE

THE HYGIENE HYPOTHESIS

Over the past few decades, abundant epidemiologic data have revealed an inverse correlation between exposure to microbes and the incidence of autoimmune and/or atopic diseases **(Fig. 484-5)**. This type of epidemiologic correlation led to the proposal of the "hygiene hypothesis" in 1989. Initially, this hypothesis focused on the development of atopic diseases in young children, with the idea that these epidemiologic observations could "be explained if allergic diseases were prevented by infection in early childhood, transmitted by unhygienic contact with older siblings, or acquired prenatally from a mother infected by contact with her older children."[1] In fact, this notion that differences in living conditions and environmental exposures contribute to susceptibility to hay fever (*summer catarrh*) dates back to at least the early nineteenth century. The hygiene hypothesis has continued to evolve over the past three decades and now posits that inadequacies in microbial exposure—in combination with genetic susceptibilities—lead to a collapse of the normally highly coordinated, homeostatic immune response. At its core, the hygiene hypothesis holds that specific early-life microbial exposures are required to prevent subsequent disease and that the "westernization" of society has led to a decrease in such exposures. This concept is being applied beyond atopic diseases to other inflammatory and autoimmune diseases and is thought to reflect processes that occur in later life as well.

RELATIONSHIP BETWEEN THE MICROBIOTA AND SPECIFIC DISEASE STATES

The ideas inherent in the hygiene hypothesis—in sum, that microbial exposure can affect long-term health outcomes—laid the theoretical foundation for translational microbiome studies. While most of the studies described earlier sought to describe how the microbiota responds to specific and often transient influences (e.g., a course of antibiotics, dietary interventions, travel), a multitude of studies have characterized the microbiota in patients with various diseases in the hope that a better understanding of the nature of disease-specific microbial communities will provide insight into disease pathogenesis and potentially uncover novel treatment modalities. Remarkably, virtually all these studies have demonstrated differences between the microbiotas of healthy controls and patients, irrespective of the specific disease process examined. Although it is difficult to generalize across all studies, a couple of general themes have emerged. First, disease states are typically associated with microbiotas that are less diverse than those of healthy individuals. This loss of diversity can be measured

---

[1] D. Strachan: BMJ 299:1259, 1989.

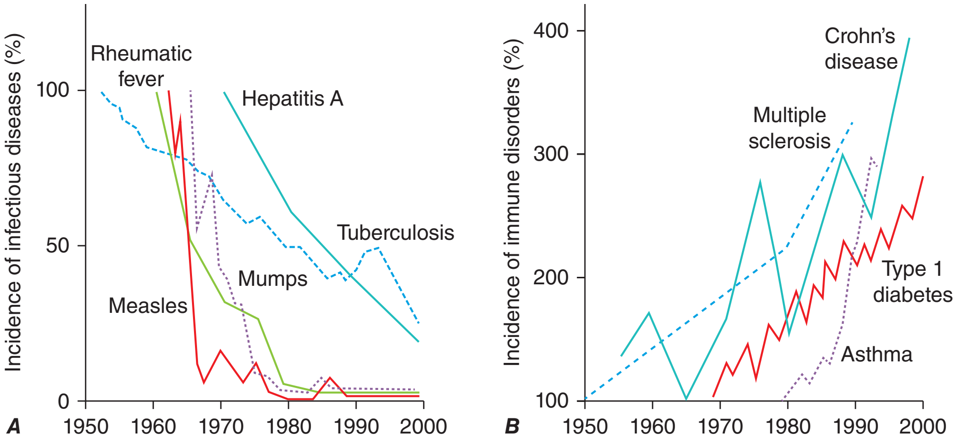

FIGURE 484-5 There was an inverse relationship between the incidence of select infectious diseases and the incidence of autoimmune disorders during the latter half of the twentieth century. *A.* Relative incidence of prototypical infectious diseases from 1950 to 2000. ***B.*** Relative incidence of select autoimmune disorders from 1950 to 2000. *(From JF Bach: The effect of infections on susceptibility to autoimmune and allergic diseases. N Engl J Med 347:911, 2002. Copyright © 2002, Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

either as a decrease in the number of species (*alpha diversity*; often measured as the number of operational taxonomic units or amplicon sequence variants, which are the bioinformatic equivalent of species) or as a reduction in the microbial relatedness of the species present (*beta diversity*). Often, both alpha and beta diversity decrease in the setting of disease. Second, states of inflammation—regardless of site or underlying disease process—are often associated with an increase in the relative abundance of the bacterial family Enterobacteriaceae and a decrease in the relative abundance of Lachnospiraceae.

Dissecting Correlation and Causality Given that most of these investigations have been designed as case-control studies, it is difficult to determine whether microbiologic findings are the cause or the effect of the disease. Even studies that examine treatment-naïve patients at the time of initial diagnosis are still confounded by this "chicken or egg" issue. Moreover, prospective, longitudinal clinical studies—still rare in the microbiome field—may simply yield correlations between the microbiome and subclinical disease rather than necessarily proving causality. Experiments in animals—specifically, studies using gnotobiotic mice (GF mice that have been colonized with specified microbial communities)—have been critical in this regard as they allow investigation of specific differences in microbial components while controlling for the host's genetics, diet, and housing conditions. Moreover, human microbes can be transplanted into gnotobiotic mice to permit in-depth mechanistic studies of how these microbial communities affect disease pathogenesis. This marriage of human samples and animal experiments has facilitated the identification of causal roles played by some microbes in disease pathogenesis; these findings provide a critical proof of concept for the interplay of the microbiota with human health. However, the vast majority of microbiome studies are still at the level of correlation. The next several sections describe the clinical and animal data for many different disease processes. Given the voluminous and rapidly changing nature of this field, it is impossible to cover all of the disease associations known to date; rather, the following discussion represents a combination of the leading exemplars of microbiome data and nascent areas of significant clinical interest. In all cases, the hope is that further study of the role of the microbiota will provide novel diagnostics, new therapeutic modalities, and/or additional insight into disease pathogenesis.

Gastrointestinal Diseases Given that the intestines harbor the largest number and greatest diversity of organisms in the body, much work has focused on how the microbiota impacts gastrointestinal diseases. Even though the luminal surface area of the gastrointestinal tract is 30–40 square meters (~90% of which is contained within the small intestine) and features marked anatomic and functional differences that result in many discrete macro- and micro-ecosystems, stool is often used as a surrogate for the intestinal microbiota given the relative ease of collecting samples. A few studies that have compared the microbial profile in stool with the mucosa-adherent organisms present in biopsy samples have demonstrated that stool is, in fact, a reasonable proxy for biopsy samples; however, the relative microbial "noise" present in stool can sometimes overwhelm the "signal," making biopsy samples more informative for some scientific questions. The key issue is to ensure that the biopsy samples evaluated represent relatively similar intestinal regions, as there are significant differences between the organisms present in the crypt and the tip of the villus and between microbes found in the ascending versus the descending colon. Newer technologies (e.g., smart capsules) are being developed that will allow for noninvasive sampling of microbial communities along the length of the gastrointestinal tract, which will provide new insight into regional differences in host–microbiota interactions.

OBESITY Obesity is a worsening epidemic throughout the world, and multiple studies have linked the composition of the intestinal microbiota to the development of obesity in animal models and in humans. Indeed, many of the initial translational microbiome studies performed in mice at the beginning of the twenty-first century focused on obesity. Gnotobiotic mouse studies have demonstrated the gut microbiota impacts host metabolism—resulting in body weight and adiposity changes—through several different mechanisms: the microbiota impacts the amount of energy extracted from the diet, promotes small-intestinal absorption of dietary fatty acids, regulates expression of lipid metabolism genes in the intestines, and induces hepatic lipogenesis and synthesis of triglycerides. Consistent with these findings, GF mice are resistant to diet-induced obesity, which establishes the requirement of the microbiota in the development of obesity. Over the past ~15 years, numerous human studies examining the relationship between the microbiome and obesity have been completed, all with mixed results. Although initial studies suggested obesity was associated with a lower ratio of the relative abundance of Bacteroidetes to Firmicutes, this has not held up in subsequent studies. Beyond this ratio of major bacterial phyla, obesity was linked to a microbiome with a lower alpha diversity. A meta-analysis of 10 studies including nearly 3000 individuals revealed an apparent lack of relationship between the Bacteroidetes/Firmicutes ratio and obesity, though there is ~2% lower diversity associated with obesity that is statistically significant but of unclear biologic significance. This finding highlights a problem common to microbiome studies: i.e., there is no sense as to what magnitude of change is biologically meaningful. Ultimately, although murine studies have indicated a causal link between the microbiota and obesity, the human data are less convincing, and their significance may be limited because the studies primarily examined only high-level taxonomic information rather than also assessing differences in bacterial products or metabolites.

The rise in obesity has elicited a plethora of ideas about the type of diet that might be most successful in leading to sustained weight loss. However, it has become clear that the same dietary ingredient can have highly diverse effects on blood glucose measurements in different people and that this effect is mediated largely by the microbiome. These observations suggest that the "optimal" diet needs to be individualized in the context of the person's microbiome, which itself may continue to change over the course of the diet. Moreover, the microbiota may also influence dietary preferences, which suggests important feedback loops between the microbiome and diet.

MALNUTRITION Representing the other end of the metabolic spectrum from obesity, malnutrition is also linked to an altered microbiome. Analysis of Malawian twin pairs (≤3 years of age) who were discordant for kwashiorkor—a severe form of malnutrition—revealed that kwashiorkor is associated with a microbiologically "immature" fecal microbiota that resembles that of a chronologically younger child. Transplantation of the fecal microbiota from these discordant twins into gnotobiotic mice that were fed a diet similar in composition to a typical Malawian diet established that the kwashiorkor-associated microbiome is causally related to poor weight gain. Subsequent studies demonstrated these same general trends in malnourished Bangladeshi children. Investigators were able to identify five bacterial species (*Faecalibacterium prausnitzii*, *Ruminococcus gnavus*, *Clostridium nexile*, *Clostridium symbiosum*, and *Dorea formicigenerans*) that—when administered together as a "cocktail" to mice colonized with a kwashiorkor-associated microbiome—were able to prevent growth impairments. Moreover, children with moderate acute malnutrition fed therapeutic food purposefully designed for its ability to alter the microbiota in defined manners have improved growth. These results demonstrate that rationally designed modulation of the microbiota may lead to improved health outcomes.

INFLAMMATORY BOWEL DISEASE Ulcerative colitis and Crohn's disease, the two predominant forms of inflammatory bowel disease (IBD), are chronic gastrointestinal inflammatory conditions that differ in their locations and patterns of inflammation **(Chap. 337)**. The following observations have led to the suggestion that IBD is the result of an immune response to a dysbiotic microbiota in a genetically susceptible individual: genes account for only ~20% of susceptibility to IBD (and many of the relevant genes are related to host–microbe interactions), antibiotic treatment reduces the clinical severity of disease, and relapses of Crohn's disease are prevented by diversion of the fecal stream. While the microbiota clearly is not the only driver of disease, it is considered to be an important element. Accordingly, numerous animal and clinical studies have been designed to tease out the nature of the relationship between the microbiota and IBD.

Most of these studies have focused on comparing the microbiome's composition in IBD patients with that in healthy controls, concentrating on microbial diversity and specific bacterial taxa that are associated with health or disease. Unfortunately, few, if any, results have been universally obtained, probably because of differences in study design, inclusion criteria, and methodology (e.g., the use of stool, rectal swabs, or biopsy samples; the choice of sequencing primers; the analysis pipeline). Even with these differences among studies, patients with IBD typically have reduced alpha and beta diversity in their fecal microbiotas. Moreover, *Clostridium* clusters IV and XIVa, which are polyphyletic and encompass several different bacterial families, are generally reduced in patients with IBD. *F. prausnitzii* is a notable example from *Clostridium* cluster IV that is often underrepresented in the stool of patients who have Crohn's disease, with more mixed results in biopsy samples. The bacterial family Lachnospiraceae, which is largely contained in *Clostridium* cluster XIVa, and other butyrate-producing organisms are also reduced in the stool of patients with IBD. Some of these species produce butyrate by using acetate generated by other members of the microbiome, and some of these acetate-producing species are similarly reduced (e.g., *Ruminococcus albus*). These complex interactions and dependencies among bacterial species pose unique challenges to definitive ascertainment of the cause–effect relationships between microbes and disease. Even before researchers were able to assess the entire microbiome at once, they often noted that patients with Crohn's disease had a higher representation of adherent invasive *E. coli* in the ileal mucosa, an observation consistent with the increased abundance of Enterobacteriaceae seen in sequencing-based microbiome studies. Beyond bacteria, burgeoning evidence supports a role for Caudovirales bacteriophages in IBD pathogenesis, though these findings may merely reflect the underlying dysbiosis related to the loss of bacterial diversity in IBD. Moreover, dysregulation of the fungal component of the microbiota (the *mycobiota*) alters the mucosal immune system and is linked to IBD disease severity. It is still unclear whether any of these microbial associations reflect the cause of IBD or merely serve as biomarkers of disease.

Studies of antibiotic-treated mice and gnotobiotic mice colonized with IBD-associated microbiotas have been useful in confirming that the microbiota affects colitis severity. Several bacterial species have been identified as either promoting colitis in mice (e.g., *Klebsiella pneumoniae*, *Prevotella copri*) or protecting against it (e.g., *Bacteroides fragilis*, *Clostridium* species); however, these organisms do not always correlate with the taxa identified as differentially abundant across multiple clinical studies. In contrast, IgA-coated commensal organisms isolated from patients with IBD promote more severe colitis in mice than either IgA-uncoated bacteria from patients with IBD or IgA-coated bacteria from healthy controls. These data suggest that functional categorization of the microbiota based on immune recognition (e.g., IgA coating) may be a useful approach for identifying pathogenic organisms.

Cardiovascular Disease Inflammation helps drive the pathogenesis of atherosclerosis, and it has long been postulated that microbes are involved in the atherosclerotic process. Early work demonstrated that patients with cardiovascular disease have higher titers of antibody to *Chlamydia pneumoniae* than control patients, that *C. pneumoniae* is present within atherosclerotic lesions, and that *C. pneumoniae* can both initiate and exacerbate atherosclerotic lesions in animal models. This type of analysis has been extended to other bacteria, such as *Porphyromonas gingivalis*, with the idea that multiple different bacteria may play some role in the pathogenesis of atherosclerosis.

Studies have demonstrated clinical correlations between serum levels of trimethylamine *N*-oxide (TMAO) and atherosclerotic heart disease. Animal studies have confirmed that transfer of the gut microbiota from atherosclerosis-susceptible strains of mice to atherosclerosis-resistant animals leads to increased serum levels of TMAO and a dietary choline-dependent increase in atherosclerotic plaques; this observation confirms the role of the gut microbiota in the generation of TMAO and atherosclerosis. Given that red meat, eggs, and dairy products are important sources of carnitine and choline (both precursors of TMAO), it is not surprising that levels of TMAO are higher in omnivores than in vegans. The gut microbiota converts carnitine into the intermediary metabolite γ-butyrobetaine, which it further metabolizes—in a diet-dependent fashion—into trimethylamine (TMA); hepatic flavin-containing monooxygenases then transform TMA into TMAO. Moreover, treatment of atherosclerosis-susceptible strains of mice with a structural analogue of choline that inhibits the first enzymatic step in TMAO formation leads to decreased circulating TMAO levels and, more importantly, restrains macrophage foam-cell formation and atherosclerotic lesion development. In a study of >4000 patients, plasma TMAO levels were also predictive of incident thrombosis risk (myocardial infarction, stroke). Gnotobiotic animals were used to demonstrate that this risk was dependent on the microbiota; although eight bacterial taxa were identified as being associated with both plasma TMAO levels and thrombotic risk, organisms with choline-utilization genes that represent the first step of TMAO production were not more abundant in animals at greater risk for thrombosis. This discrepancy highlights the complexity of the microbiota and suggests that other aspects of the overall dynamics of the microbial community may be in play.

Oncology Studies exploring the link between the microbiota and cancer have demonstrated that specific members of the microbiota can affect treatment efficacy in both a positive and a negative manner. For example, therapy with antibody to programmed cell death ligand 1 (anti-PD-L1) has proven highly effective for many different cancers **(Chap. 78)**; however, a significant proportion of patients do not respond even when their tumors have high PD-L1 expression levels. Three groups have independently performed clinical studies—sometimes coupled with gnotobiotic mouse experiments to verify causal relationships—to demonstrate that specific bacteria can potentiate checkpoint blockade inhibition in melanoma, non-small-cell lung cancer, and renal cell carcinoma. Intriguingly, these groups identified different bacteria (*Bifidobacterium*, *Faecalibacterium*, and *Akkermansia* species) as being associated with the anticancer effects, even when the same oncologic process was being studied. The biologic factors driving these differences are not yet clear but may relate to differences in adjunctive therapies, geography, and/or other as-of-yet unidentified factors. Although these seemingly disparate findings raise concern about the generalizability of microbiome studies, it may be that identifying relevant bacterial species—as opposed to their bioactive molecules—does not offer sufficient granularity for comparison across studies. The clinical relevance of the microbiota in this process was highlighted by proof-of-concept clinical trials demonstrating that fecal microbiota transplantation (FMT)—the "transplantation" of stool from one individual into another—led to clinical benefit in a few patients who previously did not respond to anti-PD-1 therapy after they received stool from patients who had previously responded to anti-PD-1 therapy, findings which still require confirmation in larger clinical trials.

In a separate set of studies, the efficacy of therapy with antibody to cytotoxic T lymphocyte–associated antigen 4 (anti-CTLA-4) was associated with T-cell responses specific for either *Bacteroides thetaiotaomicron* or *B. fragilis*. In particular, administration of *B. fragilis* to GF or antibiotic-treated mice restored the normally absent anticancer response to anti-CTLA-4 therapy. While these examples demonstrate potentiation of anticancer therapies by the microbiota, other therapies can be antagonized. Some cancers, such as pancreatic ductal adenocarcinoma, contain intratumoral bacteria, particularly Gammaproteobacteria, that can metabolize the chemotherapeutic agent gemcitabine and thereby contribute to the drug resistance of these tumors. In addition, the gut microbiota can increase the half-life of irinotecan, a chemotherapeutic agent commonly used in treating rhabdomyosarcoma and colorectal cancer, by converting an inactive metabolite back to the active form, which leads to increased drug toxicities. Overall, these examples highlight the microbiota's critical impact—both direct and indirect—on the efficacy and safety profile of drugs. Many other notable examples have been described (e.g., involving cyclophosphamide, digoxin, levodopa, and sulfasalazine), and many more likely remain to be discovered.

Using advances in computational tools for sequence decontamination and batch effect correction, reanalysis of data repositories generated by The Cancer Genome Atlas (TCGA) Research Network has identified microbial signatures within tumor genome sequences that predicted clinical outcomes in cancer, although these findings have been questioned given potential errors in the computational pipeline. This ongoing controversy highlights the complexity within the bioinformatic pipelines, the requirement for detailed reference databases, and dealing with samples that have an overall low abundance of microbes. Additional work is required to validate these signatures in prospective cohorts and to understand the biology underlying microbe–cancer interactions within the tumor milieu.

The application of microbiome science to hematopoietic stem cell transplantation (HSCT) is an area of expanding interest, particularly given the significant morbidity and mortality related to graft-versus-host disease (GVHD). In light of studies in the 1970s showing that GF mice developed less frequent and less severe gut GVHD than wild-type mice, clinicians began to use antibiotics to decontaminate the gut of patients undergoing HSCT. This decontamination approach yielded mixed results, probably because of differences in the antibiotic regimens used. The natural history of patients undergoing allogeneic HSCT includes a substantial loss of diversity in the fecal microbiota and intestinal domination (≥30% abundance in the fecal microbiota) by *Enterococcus* species and other pathogens, with a higher bacterial diversity at time of neutrophil engraftment associated with lower mortality. Moreover, a retrospective analysis of ~850 patients undergoing allogeneic HSCT revealed that receipt of imipenem-cilastatin or piperacillin-tazobactam for neutropenic fever was associated with increased GVHD-related mortality at 5 years; this observation suggested that specific bacteria may help protect against GVHD-related mortality. More detailed analyses revealed an association between the abundance of *Blautia* species and protection against GVHD and mortality, though this correlation is still being examined with regard to its causal relationship. Despite significant interest in examining these microbial relationships with HSCT, little has yet been studied in the context of solid organ transplantation, which likely represents the next frontier of transplantation-related microbiome investigation.

Autoimmune Diseases The dramatic rise in the incidence of many autoimmune diseases over the past few decades has been far more rapid than can be explained simply by genetic factors (Fig. 484-5). It is increasingly thought that environmental triggers, including the microbiome, are partially responsible for the development of these autoimmune diseases.

TYPE 1 DIABETES Type 1 diabetes (T1D) is an autoimmune disorder characterized by T cell–mediated destruction of insulin-producing pancreatic islets **(Chap. 415)**. There is a clear genetic predisposition for the disease: ~70% of patients with T1D have human leukocyte antigen (HLA) risk alleles. However, only 3–7% of children with these risk alleles actually develop disease, an observation that suggests a role for other environmental factors. Studying a prospective, densely sampled, longitudinal cohort of at-risk, HLA-matched children from Finland and Estonia, investigators detailed changes in the microbiota prior to development of disease. Although only 4 of the 33 children studied developed T1D within the time frame of the study, a marked decrease of ~25% in alpha diversity occurred after seroconversion but before disease diagnosis. The low number of cases in this study unfortunately precluded identification of any specific disease-associated taxa. A follow-up study compared the microbiomes of a larger cohort of these high-risk northern European children with those of low-risk Russian children who lived in geographic proximity. *Bacteroides* species were more abundant in the high-risk group than in the low-risk group, particularly at early ages. This difference was postulated to be associated with an altered structure of the bacterial LPS to which children were exposed at a young age. It was further suggested that *Bacteroides*-derived LPS was not able to provide the immunogenic stimulus necessary to prevent T1D. These two studies offer attractive—though logistically complicated—options for future clinical investigations aimed at exploring the role of the microbiome. The first approach—longitudinally following individuals who are at high risk for a given disease—may provide insight into host–microbe relationships by mapping temporal changes in the microbiome with disease onset. An important caveat with this type of study, though, is that the associations identified may reflect preclinical disease rather than specifically indicating causality for any observed changes. The second approach illustrates how careful selection of study participants may offer an opportunity to uncover more meaningful associations that can subsequently be experimentally verified.

RHEUMATOID ARTHRITIS Similar to many other autoimmune diseases, rheumatoid arthritis (RA) is a multifactorial disease that comes to clinical attention after an environmental factor triggers symptoms in an individual with preexisting autoantibodies. Multiple lines of evidence support the notion that RA pathogenesis is reliant on the microbiota, including the findings that GF mice do not develop symptoms in several RA models and that antibiotic treatment of mice mitigates against RA development. Several taxa (e.g., *Bacteroides* species, *Lactobacillus bifidus*, and segmented filamentous bacteria) have been implicated in promoting RA in murine models, and analysis of the fecal microbiota of patients with newly diagnosed RA has indicated that *P. copri* is a biomarker of disease. That this association with *P. copri* does not exist for chronic, treated RA or for psoriatic arthritis suggests some specificity for new-onset RA. A major limitation of this approach is that the identified association is shown to be a biomarker of disease (and, in this case, potentially of response to treatment), but no added insight is gained into a possible causal relationship between *P. copri* and RA. In fact, many of the patients with new-onset RA had no *Prevotella* detected, and several of the healthy controls had significant levels of *Prevotella*. The lack of a strict concordance between the presence (or absence) of a specific taxon and a given disease state argues against a possible causal role.

MULTIPLE SCLEROSIS Epidemiologic studies of twin pairs and at-risk individuals moving between high- and low-risk geographic areas indicate that genetics plays a minor component in multiple sclerosis

(MS) susceptibility relative to environmental factors. For example, in monozygotic twin pairs in which one sibling has MS, the other sibling also develops MS in only ~30% of cases. Although MS is a disease of the central nervous system (CNS), there is growing evidence of a link between MS and the microbiota, specifically that of the gut. In murine models of MS, GF and antibiotic-treated animals displayed reduced disease incidence and severity, and gnotobiotic mice harboring the fecal microbiota of individuals with MS—but not that of healthy controls—had increased disease activity. Clinical studies have bioinformatically associated numerous microbial changes with the presence of MS, including prior infection with Epstein-Barr virus. Importantly, a causal role has not yet been established for any of these microbes in MS pathogenesis. Although work relating the microbiome to MS is ongoing, it has opened the door to exploring this link with other neurologic diseases. Animal studies have linked the microbiota with Parkinson's disease, Alzheimer's disease, and autism, and there are clinical data assessing fecal microbiomes in relation to a variety of neurologic conditions. It is not yet clear how the gut microbiota is communicating with the CNS—i.e., whether communication takes place via bacterial metabolites that travel in the bloodstream and cross the blood-brain barrier, via migration of whole organisms into the CNS, or via feedback through the vagus nerve. Emerging data suggest that a subset of enteroendocrine cells in the intestinal epithelium is synaptically connected to the CNS, which may provide another means for the gut microbiota to impact neurologic function. Although our understanding of this brain-gut axis is still in its infancy, research in this area has elicited tremendous excitement as a tractable approach to potential treatments for these challenging diseases.

Atopic Diseases The incidence and prevalence of allergic diseases continue to steadily increase, as do more severe clinical presentations. Life-threatening food allergies are now such a public health issue that nut-free classrooms are the norm in many cities. The development of allergic diseases often follows a stereotyped progression that begins with atopic dermatitis (AD) and continues, in order, with food allergy, asthma, and allergic rhinitis. The microbiome has been linked to all of these conditions and has the potential to modulate effects anywhere along this spectrum.

ATOPIC DERMATITIS The skin is the largest organ in the body, and its different anatomic sites (e.g., antecubital fossa, volar forearm, alar crease) represent distinct ecologic niches and harbor unique microbial communities. Moreover, given that the skin serves as a critical interface between the body and the external environment (e.g., microbes), it must be able to respond to unwanted microbes with an adequate immune response. AD is an inflammatory skin disorder involving immune dysfunction and a dysbiotic skin microbiota that is typically marked by greater abundances of *Staphylococcus aureus* and reduced bacterial diversity. Effective treatment of AD does not require complete elimination of *S. aureus* but is associated with restoration of the normal level of diversity. It is likely that this increase in bacterial diversity reestablishes normal immune homeostasis in the skin; specific members of the skin microbiota have been shown to induce protective skin-restricted immune responses. Coagulase-negative staphylococci (CoNS; primarily *S. epidermidis* and *S. hominis*) obtained from lesional and nonlesional skin of patients with AD were functionally screened and compared to CoNS from healthy controls; AD-lesional CoNS were much less often able to produce antimicrobial peptides (lantibiotics) directed against *S. aureus*. To demonstrate that these lantibiotic-producing CoNS were biologically relevant, they were incorporated into a lotion and applied to the arms of patients with AD. Surprisingly, a single application of the probiotic-laced lotion led to a decrease in the abundance of *S. aureus* recovered; no such decrease was observed when lantibiotic-negative strains were used. The authors of this study did not specifically comment on the clinical improvement of the AD lesions. Nevertheless, this is one of a limited number of studies that is beginning to extend microbiome-related findings into clinical trials.

ASTHMA Asthma is characterized by the clinical triad of airflow obstruction, bronchial hyperresponsiveness, and inflammation in the lower respiratory tract. Although the long-standing dogma was that the lungs are sterile, there is now convincing evidence for a constant ebb and flow of bacteria within the lower airways. In healthy states, the mucociliary escalator continually eliminates these bacteria soon after they land in the airways; in disease states (e.g., cystic fibrosis, chronic obstructive pulmonary disease), these bacteria establish long-term colonization of the airways and influence disease pathogenesis. In asthma specifically, both fecal and airway microbes have been linked to clinical outcomes.

Early studies of the microbiome's influence on asthma used culture-based methods to assess the hypopharyngeal microbiota of asymptomatic 1-month-old infants. Intriguingly, in one study, early-life colonization with *S. pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, or a combination of these organisms—but not *S. aureus*—was significantly associated with persistent wheeze and asthma at 5 years of age. Eosinophilia and total IgE levels at 4 years of age were also increased in children who were neonatally colonized with these organisms. Although this study examined a focused set of bacteria, it laid the experimental groundwork indicating that early-life microbial exposures influence subsequent development of asthma. A later longitudinal investigation of the fecal microbiota in a general-population birth cohort of >300 children demonstrated that lower abundances of the genera *Lachnospira*, *Veillonella*, *Faecalibacterium*, and *Rothia* at 3 months of age were associated with an increased risk for development of asthma. The fact that these bacterial changes were no longer apparent when the children were 1 year of age is consistent with the notion that microbial exposures early in life are important to disease pathogenesis later in life. Transplantation of stool samples from 3-month-old children at risk for asthma into gnotobiotic mice resulted in significant airway inflammation in a murine model of asthma; pre- and postnatal exposure of mice to a four-species cocktail (*F. prausnitzii*, *Veillonella parvula*, *Rothia mucilaginosa*, and *Lachnospira multipara*) inhibited airway inflammation, with a marked reduction in neutrophil numbers in bronchoalveolar lavage fluid. These data suggest that early-life modulation of the microbiome may be an effective strategy to help prevent asthma, though the specific logistics (e.g., strains, dose, timing of exposure, patient selection) remain to be clarified.

Infectious Diseases The increased susceptibility of antibiotic-treated mice to infection with a wide range of enteric pathogens was initially observed in the 1950s and led soon thereafter to the concept of *colonization resistance*, which holds that the normal intestinal microbiota plays a critical role in preventing colonization—and therefore disease production—by invading pathogens. Seminal work in the 1970s demonstrated that this protection is largely reliant on anaerobic gram-positive organisms, and the subsequent half-century has been spent trying to identify the specific microbes involved. Although much of the work relating the microbiota to infection has focused on enteric pathogens, the intestinal microbiota has also been clearly linked to bacterial pneumonia in mouse models, and changes in the microbial composition of the gut have been causally related to changes in the severity of disease. Although this gut-lung axis clearly exists in animals, its relevance in humans is still unclear. Several groups are beginning to study the human lung microbiome in the context of pneumonia and tuberculosis. Moreover, the relationships between the microbiota and both systemic infections (e.g., HIV infection, sepsis) and the response to vaccination are starting to be explored.

ENTERIC INFECTIONS *Clostridioides difficile* infection (CDI) represents a growing worldwide epidemic and is the leading cause of antibiotic-associated diarrhea **(Chap. 139)**. Roughly 15–30% of patients who are successfully treated for CDI end up with recurrent disease. The strong association between antibiotic exposure and CDI initially raised the idea that the microbiota is inextricably linked to acquisition of disease, presumably because of the loss of colonization resistance. Consistent with the epidemiologic data, characterization of the fecal microbiota of patients with CDI revealed that it is a markedly less diverse, dysbiotic community. FMT using stool from a healthy individual was successfully used in the 1950s to treat four patients with severe CDI and has since been demonstrated in numerous studies to be an effective therapy for recurrent CDI, with clinical cure in 85–90% of patients (as detailed below). Thus, FMT for recurrent CDI has become the "poster child" for the idea that microbiome-based therapies can transform the management of many diseases previously considered to be refractory to medical therapy. Although FMT is agnostic as to the underlying mechanism of protection, work is ongoing to identify specific microbes and host pathways that can protect against CDI. Studying mice with differential susceptibilities to CDI due to antibiotic-induced changes in their microbiota, investigators identified a cocktail of four bacteria (*Clostridium scindens*, *Barnesiella intestihominis*, *Pseudoflavonifractor capillosus*, and *Blautia hansenii*) that conferred protection against CDI in a mouse model. Intriguingly, treatment of mice with just *C. scindens* offered significant, though not complete, protection in a bile acid–dependent manner. Clinical data from patients who underwent HSCT also associated *C. scindens* with protection from CDI, an observation that suggests the possibility of translating these findings from mice to humans. This study provides another example of the identification of relevant bacterial factors through examination of microbial differences in populations that differ in disease risk.

Microbiome-related changes associated with *Vibrio cholerae* infection include a striking loss of diversity (largely due to *V. cholerae* becoming the dominant member of the microbiota) and an altered composition that rapidly follows the onset of disease. These changes, which occur in a reproducible and stereotypical manner, are reversible with treatment of the disease. This recovery phase involves a microbial succession that is similar to the assembly and maturation of the microbiota of healthy infants. In addition to *V. cholerae*, streptococcal and fusobacterial species bloom during the early phases of diarrhea, and the relative abundances of *Bacteroides*, *Prevotella*, *Ruminococcus/Blautia*, and *Faecalibacterium* species increase during the resolution phase and mark the return to a healthy adult microbiota. Analysis of these microbial changes occurring in patients with cholera and in healthy children led to the selection of 14 bacteria that were transplanted into gnotobiotic mice, which were then challenged with *V. cholerae*. Bioinformatic analysis of specific taxa changing during cholera determined that *Ruminococcus obeum* restrained *V. cholerae* growth. Subsequently, this relationship was experimentally confirmed, and the *R. obeum* quorum-sensing molecule AI-2 (autoinducer 2) was found to be responsible for restricting *V. cholerae* colonization via an unclear mechanism. These studies highlight the potential for use of microbiome-based therapies to prevent and/or treat infectious diseases. Moreover, they suggest that temporal analysis of longitudinal microbiome data may be an effective strategy for identifying microbes with causal relationships to disease.

VIRAL INFECTIONS One long-standing maxim for management of infectious diseases is that antibiotics are only to be used for treatment of bacterial infections. Studies using mouse models have demonstrated, however, that a variety of viruses require the bacterial component of the microbiota for pathogenesis. Moreover, antibiotic therapy, which has bacteria-independent effects on the host, leads to reduced disease severity in some animal models of viral infection, though the clinical relevance of this is not yet clear. In addition to being required for some viral infections to proceed, commensal bacteria have also been shown to play a critical role in inducing type I interferons, which represent a potent defense mechanism against many viruses, and for modulating cellular physiology in ways that inhibit viral replication.

HIV INFECTION The augmentation of HIV pathogenesis by some viral, bacterial, and parasitic co-infections suggests that a patient's underlying microbial environment can influence the severity of HIV disease. Moreover, it has been hypothesized that the intestinal immune system plays a significant role in regulating HIV-induced immune activation; this seems particularly likely since the intestines are an early site for viral replication and exhibit immune defects before peripheral CD4+ T-cell counts decrease. Several studies of HIV-infected individuals have identified substantial differences in the HIV-associated fecal microbiota that correlate with systemic markers of inflammation. Curiously, these microbial changes do not necessarily normalize with antiretroviral therapy; this finding suggests that the microbiota may have some "memory" of the previously high HIV loads and/or that HIV infection helps reset the "normal" microbiota. This memory-like capacity of the microbiota has been demonstrated in animal models in the context of other infections and in response to dieting.

Given that the majority of new HIV transmission events follow heterosexual intercourse, there has been significant interest in examining the relationship between the vaginal microbiota and HIV acquisition. A longitudinal study of South African adolescent girls who underwent high-frequency testing for incident HIV infection facilitated the identification of bacteria that were associated with reduced risk of HIV acquisition (*Lactobacillus* species other than *L. iners*) or with enhanced risk (*Prevotella melaninogenica*, *Prevotella bivia*, *Veillonella montpellierensis*, *Mycoplasma*, and *Sneathia sanguinegens*). In mice inoculated intravaginally with *Lactobacillus crispatus* or *P. bivia*, the latter organism induced a greater number of activated CD4+ T cells in the female genital tract, a result suggesting that the increased risk of HIV acquisition associated with *P. bivia* may be secondary to the increased presence of target cells. In a separate study, the composition of the vaginal microbiota was shown to modulate the antiviral efficacy of a tenofovir gel microbicide. Although tenofovir reduced HIV acquisition by 61% in women who had a *Lactobacillus*-dominant vaginal microbiota, it reduced HIV acquisition by only 18% in women whose vaginal microbiota comprised primarily *Gardnerella vaginalis* and other anaerobes. This difference in efficacy was due to the ability of *G. vaginalis* to metabolize tenofovir faster than the target cells can take up the drug and convert it into its active form, tenofovir diphosphate. These findings illustrate how microbial ecology can be an important consideration in choosing effective treatment regimens.

RESPONSE TO VACCINATION Second only to the provision of clean water, vaccination has been the most effective public health intervention in the prevention of serious infectious diseases. Its effects are mediated by antigen-specific antibodies and, in some cases, effector T-cell responses. Although vaccines are clearly effective on a population scale, the magnitude of the immune response to vaccines can vary among individuals up to a hundredfold. Although many factors (e.g., genetics, maternal antibody levels, prior antigen exposures) can affect vaccine immunogenicity, the microbiota is now recognized as another important factor. Several cohort studies have associated differences in the fecal microbiota with altered vaccine responses, and the nasal microbiota is thought to contribute to the IgA response to live, attenuated influenza vaccines. These correlations based on clinical data have been partially confirmed in animal studies. The best example is the demonstration that the responses to nonadjuvanted viral subunit vaccines (inactivated influenza and polio vaccines) are reliant on the microbiota, whereas the responses to live or adjuvanted vaccines (live attenuated yellow fever, Tdap/alum, an HIV envelope protein/alum vaccine) are not. A causal role for the microbiome influencing vaccine-induced immunity in humans was demonstrated by comparing microneutralization titers following the inactivated influenza vaccine in individuals treated with or without antibiotics, although an antibiotic-dependent effect was only present in subjects who had low levels of preexisting immunity to influenza. These data suggest that the microbiota may serve as an adjuvant for certain vaccine types and in naïve populations. Incorporation of specific commensal bacteria and/or their products that improve vaccine responses into vaccine formulations may increase overall vaccine efficacy.

MECHANISMS OF MICROBIOME-MEDIATED EFFECTS

As highlighted in the examples above, numerous associations have been made between the microbiome and various disease states. These correlations have often been established at broad taxonomic levels, with little or no insight into causality. Given that most clinical studies of these relationships have a fairly small sample size (often <100) and are simultaneously comparing numerous variables (i.e., each of the bacterial species in the microbiota is effectively a different feature being compared), many of these studies may not be adequately powered and therefore may yield false-positive results. Testing of these correlations in animal models of disease has been critical in demonstrating a causal relationship between microbes and specific phenotypes. Because microbiome-wide association studies typically result in a long list of bacterial taxa that are correlated with a disease, it has been challenging to know which organism to test further in mechanistic studies. Moreover, even if a specific bacterial species is identified in these analyses, there is potentially enough strain-to-strain variation that the "functional" isolate may need to be recovered from the individuals studied; a publicly available representative of the species may not confer the same phenotype. Despite all these difficulties, a handful of specific microbes have now been linked to disease effects; some examples have been mentioned above. The next layer of challenges relates to identification of the specific mechanisms that underlie these causal relationships. Although the microbiota modulates most facets of host physiology, its impact on the immune system is the best-studied mechanism and helps explain its role in many diseases, particularly those that stem from misdirected immune responses.

REGULATION OF THE IMMUNE SYSTEM

The microbiota is required for the proper development, education, and maintenance of the immune system, a finding underscored by the fact that GF animals have an immature and underdeveloped immune system. Moreover, given that the microbiota has co-evolved with its host, a host-specific microbiota is critical for normal maturation of the immune system: gnotobiotic mice colonized with the microbiota from healthy humans have a small-intestinal immune system indistinguishable from GF mice. This impact on immune ontogeny begins in early life, with maternal transfer of microbially-targeted antibodies and microbe-derived metabolites augmenting neonatal immune development. Some of these early-life microbial exposures must occur during a time-sensitive window, after which subsequent exposures fail to redress the initial deficiency. Examples of these "original sins" are limited in number, but they can have long-lasting physiologic consequences that extend into adult life. In contrast, most host–microbe–immune interactions occur on an ongoing basis throughout life, with microbial perturbations (e.g., antibiotic use, changes in diet) disrupting this homeostatic immunity and potentially altering disease susceptibility.

The microbiota impacts virtually all aspects of host immunity, including its different arms (i.e., innate, adaptive), its varied anatomic niches (e.g., intestinal, skin, lung, bone marrow, CNS), and its overall immunologic tone and responsiveness. Not only does the microbiota influence the development and education of immune cells, but it also plays a critical role in modulating epithelial cell responses that contribute to immune defenses and disease pathogenesis. While the microbiota as a whole is known to drive these varied responses, not all microbes have the same immunomodulatory effects. Indeed, a broad screen of >50 taxonomically diverse commensal bacteria in GF mice demonstrated that most have the capacity to modulate the immune system, with very few bacterial taxa being immunologically quiescent; however, the immunomodulatory effects were often not detected when the same bacterium was administered to a mouse with a normal microbiota, which highlights the functional redundancy within the microbiota. Interestingly, bacterial taxonomy did not correlate with effects on the immune system, a finding that suggests the bioactive molecules may be unique rather than evolutionarily conserved. Given that all the tested bacteria express various canonical ligands (e.g., LPS, peptidoglycan, flagellin) for pattern recognition receptors, such as Toll-like receptors, commensal bacteria either modulate the immune system via a different class of products or their "canonical ligands" have unique structural motifs that trigger distinct signaling pathways—or combination of pathways—that result in education of the immune system.

Efforts are ongoing to define cognate relationships between specific commensal bacteria and their immunomodulatory effects, and approaches are being developed to define specific bacterial factors that are responsible for the phenotypic changes. Complicating factors are that many organisms, particularly those in the phylum Firmicutes, are not readily genetically tractable and that many of the phenotypes are not easy to assess with high-throughput screening. The use of mass spectrometry to detect and profile tens of thousands of different metabolites present in different bodily fluids has offered the promise of deeper insight into microbially mediated processes that underlie disease susceptibility. However, the fact that the overwhelming majority of these metabolites are not annotated, coupled with the sheer volume of data generated, has so far limited the general utility of these untargeted approaches. The few immunomodulatory bacteria and their bioactive molecules that have been identified serve as useful archetypes for how the microbiome influences the immune system and, more generally, host physiology. These commensal-derived products can generally be categorized as endobiotic microbial structures, modified dietary nutrients, and modified host-derived metabolites.

Endobiotic Microbial Immunomodulatory Molecules *B. fragilis* polysaccharide A (PSA) is perhaps the best-studied commensal-derived molecule that has been demonstrated to influence disease outcomes in mouse models. PSA—one of at least eight capsular polysaccharides expressed by *B. fragilis*—has a unique zwitterionic structure that incorporates both a positive and a negative charge within each repeating unit. Studies in which mice have been treated either with isogenic strains of *B. fragilis* that differ in PSA expression or with purified PSA have shown that PSA confers protection—prophylactically and therapeutically—against experimental colitis and MS. PSA is recognized by Toll-like receptor 2 on antigen-presenting cells, particularly plasmacytoid dendritic cells, and—in the setting of inflammation—induces interleukin 10 (IL-10)–producing regulatory T cells (Tregs) that help restrain inflammation.

B. fragilis is also the source of an immunomodulatory glycosphingolipid that, if present during neonatal life, decreases the number of colonic invariant natural killer T (iNKT) cells and improves outcomes in a model of colitis in adulthood. It is not clear whether these glycosphingolipids activate or inhibit iNKT cells; results have been discordant, probably because different glycosphingolipid species have been tested. A chemical synthesis approach confirmed that *B. fragilis* glycosphingolipids have distinct immunomodulatory functions depending on their specific structure.

There are an increasing number of commensal-derived polysaccharides and other large molecules that have been shown to modulate the immune system and/or disease outcomes. Advances in bacterial genetics have facilitated the identification of structural features that contribute to some of these host–microbiota relationships. It is likely that our general understanding of structure–function relationships of commensal-derived products will continue to grow in the coming years, mirroring what has occurred in microbial pathogenesis studies over the past several decades.

Modified Dietary Nutrients As described above, the human diet provides nutrients for the gut microbiota, which can metabolize them into new, bacteria-derived compounds. Perhaps the best example of this is fermentation of undigested dietary fibers into short-chain fatty acids (SCFAs). Several groups have demonstrated that SCFAs, the intestinal levels of which are largely determined by bacterial metabolism, are important for the induction of Tregs, though there is not agreement on which specific SCFA (propionate, acetate, or butyrate) is most relevant. Wild-type mice colonized with bacteria known to induce colonic Tregs have elevated cecal levels of SCFAs. Colonization with any of three *Bacteroides* species (*B. caccae*, *B. massiliensis*, and *B. thetaiotaomicron*) increases levels of acetate and propionate, whereas colonization with *Parabacteroides distasonis* or a mix of 17 human-derived *Clostridium* species elevates levels of all three SCFAs. In all of these cases, though, the SCFAs inhibit histone deacetylase, with a consequent increase in *Foxp3* expression. Notably, microbe-induced SCFA production has not been shown to be critical for Treg induction by any of these organisms. In contrast, there appears to be no correlation between SCFA levels and Treg numbers in mice monocolonized with various Treg-inducing bacterial species. Taken together, these data suggest important heterogeneity in the mechanisms underlying Treg development and do not rule out the possibility of other, redundant mechanisms for Treg induction. In addition to effects on Tregs, SCFAs also promote the epithelial barrier, impact cell proliferation (directionality depends on the specific cell type and SCFA), regulate host metabolism, and provide an energy source to colonocytes.

Although SCFAs represent the best-studied molecules that the microbiota generate from diet, there are many other physiologically important examples. The microbiota metabolizes tryptophan into various products (e.g., kynurenine, indole, and its derivatives) that influence immune function, metabolic diseases, viral infections, and neuronal function, among other things. Desaminotyrosine produced by *Clostridium orbiscindens* confers protection from influenza by inducing type I interferon activity. Modification of unsaturated fatty acids (e.g., linoleic acid) into different isomers regulates specific T-cell subsets embedded in the small-intestinal epithelium. These examples represent an important proof of concept that diet plays an important role in the functional output of the microbiota, not just its composition.

Modified Host-Derived Molecules Bile acids are produced in the liver but then are metabolized by intestinal bacteria to form deconjugated and secondary bile acids. These microbially produced bile acid profiles act through complex signaling pathways to balance the metabolism of lipids and carbohydrates and to affect immune responses. Therefore, bile acids are now being investigated as microbial metabolites that are critical to maintaining human health. As mentioned above, *C. scindens* helps protect mice against CDI through a bile acid–dependent process. Alterations in bile acid profiles due to underlying microbial dysbiosis have also been associated with hepatic and colonic inflammation, hepatic cellular carcinoma, colorectal cancer, and impaired gut motility. Almost all of these relationships have been documented at the level of correlation and, at best, reflect a partial change in phenotype in the setting of bile acid sequestrants (e.g., cholestyramine). Work is ongoing to determine causal relationships between bacterial metabolism of bile acids and changes in host physiology, though the most definitive evidence is that microbe-produced bile acid metabolites influence colonic Treg homeostasis.

In addition to bile acids, the gut microbiota can metabolize many other host-derived molecules, thereby regulating their levels and downstream effects. Taurine enhances NLRP6 inflammasome–induced colonic IL-18 secretion, while histamine, spermine, and putrescine suppress IL-18 secretion; the levels of all of these host-derived metabolites can be regulated by the microbiota. Inosine, the deamination product of adenosine, produced by *Bifidobacterium pseudolongum* enhances efficacy of checkpoint blockade inhibitors in mouse models. While these examples represent the tip of the iceberg, many more examples of bacterial metabolites will undoubtedly be linked to health and disease given the thousands of different bacterial metabolites present throughout the body. However, the clinical relevance of any of these bacterial metabolites remains unknown.

MOVING MICROBIOME SCIENCE FROM BENCH TO BEDSIDE

The numerous microbiome–disease associations identified thus far have generated a great deal of hope that understanding the relevant microbe–host interactions will open the door to unlimited therapeutic applications. Microbiome-based therapies offer several potential benefits. Patients often view such treatment as more "natural" than conventional drug therapy and are therefore more likely to comply with it. Biologically, microbiome-based therapies are more likely to address one of the root causes of disease (microbial dysbiosis) rather than simply affecting the downstream sequelae. Finally, a given microbiome-based therapy may serve as a "polypill" that is effective against several different diseases stemming from similar microbial changes. Despite tremendous interest in therapeutically exploiting the microbiome, there have thus far been few clinical successes along these lines.

The most successful therapeutic application of microbiome science has been the use of FMT, particularly for CDI. As mentioned earlier, FMT involves "transplanting" stool from one individual to a diseased patient, with the idea that the donor microbiota will correct whatever derangement may exist in the ill patient and therefore will alleviate symptoms. Fundamentally, this notion is agnostic as to the specific microbial dysbiosis and holds that any "healthy" microbiota will be curative, though some are now using donor stool from patients with a desired phenotype rather than any healthy individual. The idea of FMT dates to at least the fourth century, when traditional Chinese doctors used a "yellow soup" (fresh human fecal suspension) to successfully treat food poisoning and severe diarrhea. The continued use of FMT through the centuries for the treatment of diarrheal illnesses in both humans and animals, along with the growing appreciation of the importance of the microbiota, laid the groundwork for using FMT to treat CDI. Since the first major prospective trial assessing FMT for recurrent CDI in 2013, most of the numerous studies of FMT for CDI have demonstrated remarkable efficacy, with an average clinical cure rate of ~85%. The donor stool can be fresh or frozen (use of the latter allows biobanking of samples from a limited number of prescreened donors) and can be administered via nasogastric tube, nasoduodenal tube, colonoscopy, enema, or oral capsules; the cure rate is slightly higher with lower-gastrointestinal administration than with upper-gastrointestinal treatment. The optimal screening, preparation, and concentration of infused donor stool have not yet been determined, and there have been cases of antimicrobial-resistant pathogens transmitted by FMT that have led to mortality. The most common adverse effects of FMT include altered gastrointestinal motility (with constipation or diarrhea), abdominal cramps, and bloating, all of which are generally transient and resolve within 48 h. Although controlled studies of the use of FMT in immunosuppressed patients do not yet exist, meta-analyses of case reports and case series have found no serious FMT-related adverse events in >300 immunocompromised patients.

The successful use and the favorable short-term safety profile of FMT for CDI have led to its expanded application for other indications. As of July 2024, >500 trials (listed at *ClinicalTrials.gov*) were investigating the efficacy of FMT for a range of indications, including CDI, IBD (ulcerative colitis and Crohn's disease), obesity, eradication of multidrug-resistant organisms, anxiety and depression, cirrhosis, and type 2 diabetes. The few published studies regarding indications other than CDI have generally included small sample sizes and have offered mixed results. In contrast to the successes in CDI, the results have been more varied for patients with IBD, which is perhaps the second best-studied indication. It is not clear whether these discrepancies are due to heterogeneity in recipients (e.g., in terms of underlying disease mechanisms or endogenous microbiotas), the donor material, and/or the logistical details of FMT administration (e.g., route, frequency, dose). However, these results demonstrate that—under the right circumstances—modulation of the microbiota can be an effective therapy for IBD.

Although FMT offers an important proof of concept that microbiome-based therapies can be effective, treatment is difficult to standardize across large populations because of variability among stool donors and among the endogenous microbiotas of recipients. In addition, FMT is fraught with safety concerns, and its mechanisms of action are unclear. That said, there are now two microbiome-derived therapies conceptually analogous to FMT that are approved by the U.S. Food and Drug Administration (FDA) for treatment of recurrent CDI. FMT likely represents the first generation of microbiome-based therapies; subsequent generations will include the use of more refined bacterial cocktails, single strains of bacteria, or bacterial products and/or metabolites as the therapeutic intervention. The field of probiotics has a complicated history: many different strains have been tested against a multitude of diseases. Several meta-analyses have combined results across bacterial strains and/or disease indications and have generally concluded that the data are not yet convincing enough to support the use of the tested regimens. It should be noted that the tested organisms have generally been chosen based on their presumed safety profile rather than in light of a plausible biologic link to disease. The hope is that more focused, mechanistic microbiome studies will identify specific commensal organisms—and their underlying mechanisms of action—that are involved in disease pathogenesis and that will serve as the basis for the next wave of rationally chosen probiotics, a few of which are currently in clinical trials. The main hurdle in this endeavor has been identifying specific microbes that are causally related to protection from disease. Future therapeutic strategies might include administering a beneficial microbe/microbial product; targeting a deleterious microbe/microbial pathway; or modulating the microbial ecology, potentially by impacting keystone species.

PERSPECTIVE

The medical view of microbes has changed radically, moving from the early-twentieth-century notion that we are engaged in a constant struggle with microbes—an "us-versus-them" mentality that focused on the necessity of eradicating bacteria—to the current understanding that we live in a carefully negotiated state of détente with our commensal organisms. Instead of holding a simple view of microbes as enemies to be eliminated with antibiotics, scientists are increasingly recognizing the critical role these organisms play in maintaining human health; loss of these host–microbe interactions in the increasingly sterile environment typical of Western civilization may have predisposed to the increased incidence of autoimmune and inflammatory diseases. The field of microbiome research has made great strides over the past decade in cataloguing the normal microbiota and is now beginning to identify clinically actionable microbe–host relationships.

The explosion of "–omics" technologies (e.g., metagenomics, metatranscriptomics, metabolomics) has enabled the generation of vast amounts of data, but it is not yet clear how best to integrate data sets in order to gain useful insights into host–microbe relationships. The use of FMT has demonstrated that modulation of an individual's microbiota can effectively treat certain diseases; however, models with which to predict specifically how a microbiota will change after modulation—and what potentially untoward effects these changes might have—are still lacking. Implicit in this limitation is our ignorance about what microbial configuration is optimal and how a given microbiota should be rationally altered to obtain an ideal outcome.

Despite initial hyperbolic hype and a few false starts, microbiome research now stands at the forefront of an ability to treat the fundamental basis of many diseases. As the field continues to mature, it will need to move beyond correlations and address causation. The identification of causal microbes and their mechanisms of action will create a "microbial toolbox" from which relevant bioactive strains can be chosen on a per-patient basis to correct specific underlying microbial dysbioses. In the near future, our knowledge base regarding the microbiome and its relationship to health and disease will be robust enough that this information can be applied in making important treatment decisions.

FURTHER READING

Amato KR et al: The human gut microbiome and health inequities. Proc Nat Acad Sci 118:e2017947118, 2021.

Goodrich JK et al: Conducting a microbiome study. Cell 158:250, 2014.

Human Microbiome Project Consortium: Structure, function and diversity of the healthy human microbiome. Nature 486:207, 2012.

Schmidt TSB et al: The human gut microbiome: From association to modulation. Cell 172:1198, 2018.

Shalon D et al: Profiling the human intestinal environment under physiological conditions. Nature 617:581, 2023.

Stefan KL et al: Commensal microbiota modulation of natural resistance to virus infection. Cell 183:1, 2020.

Walker AW, Hoyles L: Human microbiome myths and misconceptions. Nature Microbiol 8:1392, 2023.

485 Global Issues in Medicine

Joseph J. Rhatigan, Paul E. Farmer*

WHY GLOBAL HEALTH?

Global health has emerged as an important field within medicine. Some scholars have defined global health as the field of study and practice concerned with improving the health of all people and achieving health equity worldwide, with an emphasis on addressing transnational problems. No single review can do much more than identify the leading problems in applying evidence-based medicine in settings of great poverty or across national boundaries. However, this is a moment of opportunity: only relatively recently have persistent calls for global health equity been matched by an unprecedented investment in addressing the health problems of poor people worldwide. To ensure that this opportunity is not wasted, we must strengthen health systems and improve health care delivery to address the true burden and distribution of disease. This chapter introduces the major international bodies that address global health problems; identifies the more significant barriers to improving the health of people who to date have not, by and large, had access to modern medicine; and summarizes population-based data on the most common health problems faced by people living in poverty. Examining specific problems—notably HIV/AIDS **(Chap. 208)** but also tuberculosis **(Chap. 183)**, malaria **(Chap. 231)**, Ebola **(Chap. 216)**, COVID-19 **(Chap. 205)** and key "noncommunicable" chronic diseases (NCDs)—helps sharpen the discussion of barriers to prevention, diagnosis, and care as well as the means of overcoming them. This chapter closes by discussing global health equity, drawing on concepts of social justice.

A BRIEF HISTORY OF GLOBAL HEALTH INSTITUTIONS

Concern about illness across national boundaries dates back many centuries, predating the Black Plague and other pandemics. One of the first organizations founded explicitly to tackle cross-border health issues was the Pan American Sanitary Bureau, which was formed in 1902 by 11 countries in the Americas. The primary goal of what later became the Pan American Health Organization was the control of infectious diseases across the Americas. Of special concern was yellow fever, which had been running a deadly course through much of South and Central America and halted the construction of the Panama Canal. In 1948, the United Nations formed the first truly global health institution: the World Health Organization (WHO). In 1958, under the aegis of the WHO and in line with a long-standing focus on communicable diseases that cross borders, leaders in global health initiated the effort that led to what some see as the greatest success in international health: the eradication of smallpox. Naysayers were surprised when the smallpox eradication campaign, which engaged public health officials throughout the world, proved successful in 1979 despite Cold War tensions.

At the International Conference on Primary Health Care in Alma-Ata (in what is now Kazakhstan) in 1978, public health officials from around the world agreed on a commitment to "Health for All by the Year 2000," a goal to be achieved by providing universal access to primary health care worldwide. Critics argued that the attainment of this goal by the proposed date was impossible. In the ensuing years, a strategy for the provision of selective primary health care emerged. This strategy included four inexpensive interventions collectively known as GOBI: *g*rowth monitoring, *o*ral rehydration, *b*reast-feeding, and *i*mmunizations for diphtheria, whooping cough, tetanus, polio, tuberculosis, and measles. GOBI later was expanded to GOBI-FFF, which also included *f*emale education, *f*ood, and *f*amily planning. Some public health figures saw GOBI-FFF as an interim strategy to achieve "health for all," but others criticized it as a retreat from the bolder commitments of Alma-Ata.

In 1982, UNICEF, led by James Grant, helped usher in a period of sustained and coordinated efforts to reduce childhood mortality using a similar set of interventions among other measures. Over the ensuing two decades, it is estimated that 25 million children's lives were saved in what is often called "The Child Survival Revolution."

The influence of the WHO waned during the 1980s. In the early 1990s, many observers argued that, with its vastly superior financial resources and its close—if unequal—relationships with the governments of poor countries, the World Bank had eclipsed the WHO as the most important multilateral institution working in global health. One of the stated goals of the World Bank was to help poor countries identify "cost-effective" interventions worthy of public funding and international support. At the same time, international financial institutions encouraged many of those nations to reduce public expenditures in health and education in order to stimulate economic growth as part of (later discredited) policies, generally referred to as "structural adjustment," that imposed restrictions on social sector government spending as a condition for access to credit and assistance through the World Bank, the International Monetary Fund, and regional development banks. There was a resurgence of many diseases—including malaria, trypanosomiasis, and schistosomiasis—in Africa. Tuberculosis, an eminently curable disease, remained the world's leading infectious killer of adults. Half a million women per year died in childbirth during the last decade of the twentieth century, and few of the world's largest philanthropic or funding institutions focused on global health equity.

HIV/AIDS, first described in the medical literature in 1981, precipitated a change. In the United States, the advent of this newly described infectious killer marked the culmination of a series of events that dashed previous hopes of "closing the book" on infectious diseases. In Africa, which would emerge as the global epicenter of the pandemic, HIV disease strained tuberculosis control programs, and malaria continued to claim as many lives as ever: at the dawn of the twenty-first century, these three diseases alone killed nearly 6 million people each year. New research, new policies, and new funding mechanisms were called for. The past two decades have seen the rise of important multilateral global health financing institutions such as the Global Fund to Fight AIDS, Tuberculosis, and Malaria; bilateral efforts such as the U.S. President's Emergency Plan for AIDS Relief (PEPFAR); and private philanthropic organizations such as the Bill & Melinda Gates Foundation. With its 194 member states and 150 country offices, the WHO remains important in matters relating to the cross-border spread of infectious diseases and other health threats. In the aftermath of the epidemic of severe acute respiratory syndrome in 2003, the WHO's International Health Regulations—which provide a legal foundation for that organization's direct investigation into a wide range of global health problems, including pandemic influenza, in any member state—were strengthened and brought into force in May 2007.

Even as attention to and resources for health problems in poor countries grow, the lack of coordination among global health institutions remains an issue. The global response to the COVID-19 pandemic revealed both strengths and fundamental flaws in the current state of global health governance. The WHO remains underfunded despite the ever-growing need to engage a wider and more complex range of health issues. This may be what some have called "the golden age of global health," but leaders of major global health organizations must work together to design an effective architecture that will make the most of opportunities to link new resources for and commitments to global health equity with the emerging understanding of disease burden and the unmet need to create robust and resilient national health systems. To this end, new and old players in global health must invest heavily in *discovery* (relevant basic science), *development* of new tools (preventive, diagnostic, and therapeutic), and modes of *delivery* that will ensure the equitable provision of health products and services to all who need them.

---

*Deceased.

The adoption of the Sustainable Development Goals (SDGs) in 2015 by the United Nations serves as an example of effective cooperation. The SDGs articulate 17 overarching goals across several domains to be achieved by 2030. Goal 3 specifically relates to global health and contains 13 distinct targets to be met, including reducing maternal and child mortality; ending the epidemics of HIV, tuberculosis, and malaria; and reducing the burden of NCDs.

Included in the SDGs is a commitment to achieve universal health coverage (UHC), providing universal access to high-quality essential health services at an affordable cost worldwide. Championed by the WHO, the World Bank, and many civil society organizations, Goal 3 will measure coverage of 16 essential health services and assess the financial burden of health spending by households in every country.

THE ECONOMICS OF GLOBAL HEALTH

Political and economic concerns have often guided global health interventions. As mentioned, early efforts to control yellow fever were tied to the completion of the Panama Canal. However, the precise nature of the link between economics and health remains a matter for debate. Some economists and demographers argue that improving the health status of populations must begin with economic development; others maintain that addressing ill health is the starting point for development in poor countries. In either case, there is increasing consensus that investments in health care delivery and the control of communicable diseases lead to increased productivity. The question is where to find the necessary resources to start the predicted "virtuous cycle."

During the past two decades, spending on health in poor countries has increased dramatically. According to a study from the Institute for Health Metrics and Evaluation (IHME) at the University of Washington, before the COVID-19 pandemic, total development assistance for health (DAH) worldwide (essentially a measure of "health aid") grew to $38.9 billion in 2018—up from $5.6 billion in 1990—and seemed to reach a plateau. However, the global response to COVID-19 saw the largest yearly increases in DAH ever recorded. Between 2020 and 2021, DAH grew by 86%. It is estimated that in 2021 DAH reached $64.7 billion. In 2021, the leading contributors included the United States, the United Kingdom, Germany, and private foundations.

MORTALITY AND THE GLOBAL BURDEN OF DISEASE

Refining metrics is an important task for global health: only relatively recently have there been solid assessments of the global burden of disease. The first study to look seriously at this issue, conducted in 1990, laid the foundation for the first report on *Disease Control Priorities in Developing Countries* and for the World Bank's 1993 World Development Report *Investing in Health*. Those efforts represented a major advance in the understanding of health status in developing countries. *Investing in Health* has been especially influential: it familiarized a broad audience with cost-effectiveness analysis for specific health interventions and with the notion of disability-adjusted life years (DALYs). The DALY, which has become a standard measure of the impact of a specific health condition on a population, combines absolute years of life lost and years lost due to disability for incident cases of a condition. (See **Fig. 485-1** and **Table 485-1** for an analysis of the global disease burden by DALYs.)

In 2012, the IHME and partner institutions began publishing results from the Global Burden of Diseases, Injuries, and Risk Factors (GBD) study. The GBD study is the most comprehensive effort to date to produce longitudinal, globally ambitious, and comparable estimates of the burden of diseases, injuries, and risk factors. This report reflects the expansion of the available data on health in the poorest countries and of the capacity to quantify the impact of specific conditions on a population. It measures current levels and recent trends for major diseases, injuries, and risk factors worldwide. Since its original publication, the GBD study team has revised and improved the health-state severity weight system, collated published data, and used household surveys to enhance the breadth and accuracy of disease burden data. Updated reports were released in 2013, 2015, 2017, and 2019. The report now includes fact sheets that examine 386 diseases and injuries in depth. As analytic methods and data quality improve, important trends can be identified in a comparison of global disease burden estimates from 1990 to 2019.

GLOBAL MORTALITY

Although the COVID-19 pandemic has led to significant excess deaths, it is instructive to look at the last GBD study's data as a comprehensive baseline of disease burden before the pandemic. In 2019, before the COVID-19 pandemic, of the 56.5 million deaths worldwide, 17% (9.6 million) were due to communicable diseases, maternal and neonatal conditions, and nutritional deficiencies—a marked decrease compared with figures for 1990, when these conditions accounted for 32% of global mortality. Among the fraction of all deaths related to communicable diseases, maternal and neonatal conditions, and nutritional deficiencies, 78% occurred in Sub-Saharan Africa and southern Asia. While the proportion of deaths due to these conditions has decreased significantly in the past decade, there has been a dramatic rise in the number of deaths from NCDs. The leading cause of death worldwide in 2019 was ischemic heart disease, accounting for 9.1 million deaths (16% of total deaths). In high-income countries, ischemic heart disease accounted for 16% of total deaths, and in low-income countries, it accounted for 16%. It is noteworthy that ischemic heart disease was responsible for just 5% of total deaths in Sub-Saharan Africa **(Table 485-2)**. In second place—causing 11% of global mortality—was stroke, which accounted for 8% of deaths in high-income countries, 6% in low-income countries, and 5% in Sub-Saharan Africa. Although chronic obstructive pulmonary disease (COPD) was the third leading cause of death globally and was the fifth leading cause in high-income countries (accounting for 5% of all deaths), this condition did not figure among the top 15 causes in Sub-Saharan Africa. Among the 10 leading causes of death in Sub-Saharan Africa, five were infectious diseases, with HIV/AIDS, lower respiratory infections, diarrheal diseases, and malaria ranking as dominant contributors to disease burden. In high-income countries, however, only 1 infectious disease—lower respiratory infection—ranked among the top 10 causes of death.

The number of deaths among children under 5 years dropped from 16.4 million in 1970 to 11.8 million in 1990 and to 5.0 million in 2019—a decrease that far surpassed predictions. Of childhood deaths in 2021, 2.3 million (46%) occurred in the neonatal period. Just under one-third of deaths among children under 5 years old occurred in southern Asia and slightly more than one-half in Sub-Saharan Africa, but only ~1% occurred in high-income countries.

The global burden of death due to HIV/AIDS and malaria was on an upward slope until 2004, but significant progress has been made since then. Global deaths from AIDS fell from 2.0 million in 2006 to 630,000 in 2022, while malaria deaths dropped from 1.2 million to 608,000 over the same period. Despite these improvements, malaria and HIV/AIDS continue to be major burdens in particular regions, with global implications. Although it has only a minor impact on mortality outside Sub-Saharan Africa and Southeast Asia, malaria is the fifth leading cause of death of children under 5 years of age worldwide. HIV infection ranked 30th in global DALYs in 1990 but was the 11th leading cause of disease burden in 2019, with Sub-Saharan Africa bearing the vast majority of this burden (Fig. 485-1).

The world's population is living longer: global life expectancy has increased significantly over the past 50 years from 58.8 years in 1970 to 73.5 years in 2019. This demographic change, accompanied by the fact that the prevalence of NCDs increases with age, is dramatically shifting the burden of disease toward NCDs, which have surpassed communicable, maternal, nutritional, and neonatal causes. By 2019, 74% of total deaths at all ages and 63% of all DALYs were due to NCDs. Increasingly, the global burden of disease comprises conditions and injuries that cause disability rather than death.

Worldwide, although both life expectancy and years of life lived in good health have risen, years of life lived with disability also have increased. Globally, the total burden of disability increased by 50% between 1950 and 2019. Despite the higher prevalence of diseases

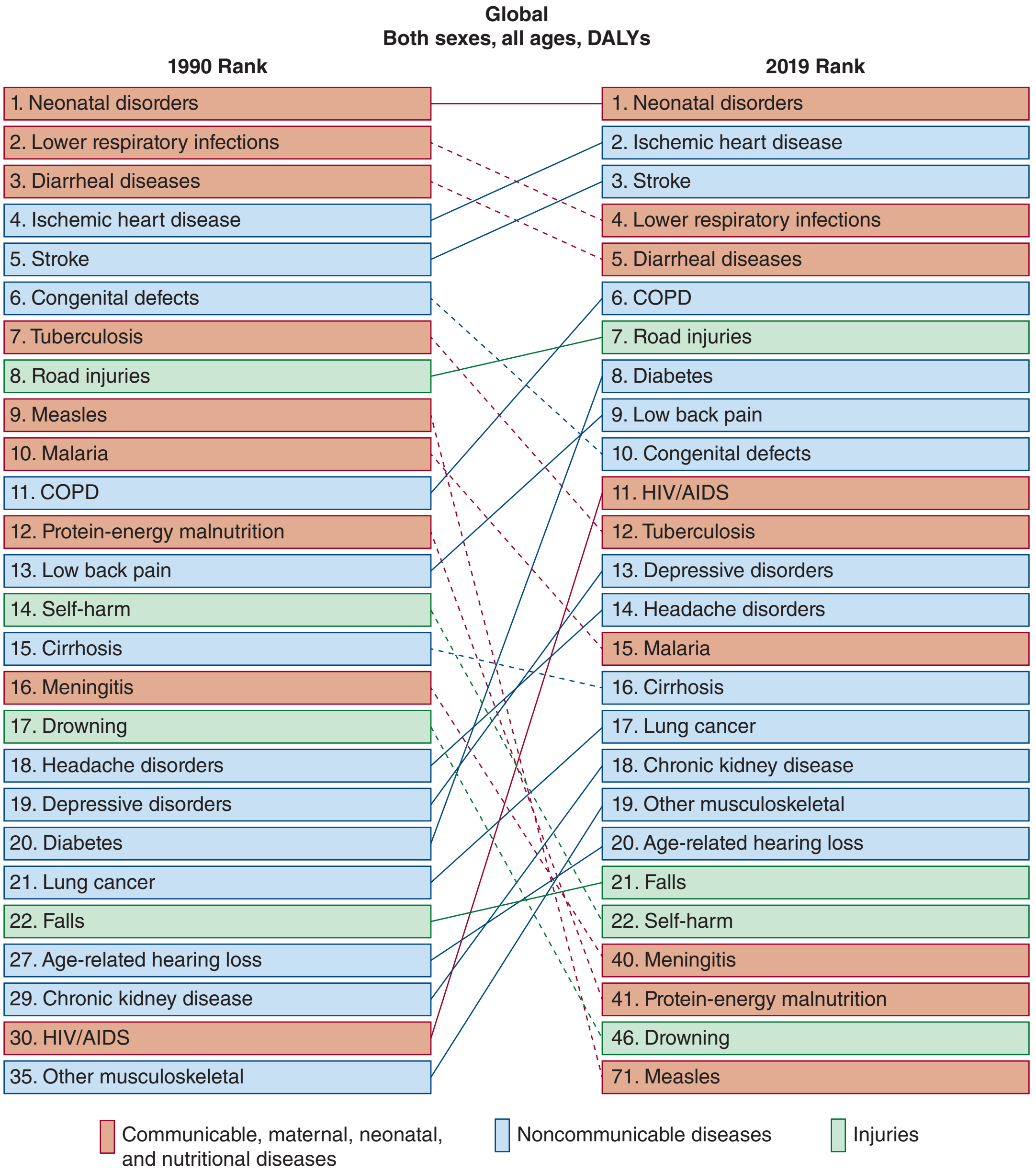

FIGURE 485-1 Global disability-adjusted life-year (DALY) ranks for the top causes of disease burden in 1990 and 2019. COPD, chronic obstructive pulmonary disease. *(From the Institute for Health Metrics and Evaluation [IHME]. GBD Compare. Seattle, WA: IHME, University of Washington, 2023. Available at http://vizhub.healthdata.org/gbd-compare. Accessed Dec 15, 2023.)*

common in older populations (e.g., dementia and musculoskeletal disease) in developed and high-income countries, best estimates from 2019 reveal that disability resulting from cardiovascular diseases, chronic respiratory diseases, and the long-term impact of communicable diseases was greater in low- and middle-income countries. In most developing countries, people lived shorter lives and experienced disability and poor health for a greater proportion of their lives.

HEALTH AND WEALTH

Clear disparities in burden of disease (both communicable and noncommunicable) across country income levels are strong indicators that poverty and health are inherently linked. Numerous studies have documented the link between poverty and health within nations as well as across them. Poverty remains one of the most important root causes of poor health worldwide, and the global burden of poverty continues to be high.

Among the 8 billion persons alive in 2023, 9% (700 million) lived on less than $2.15 per day—a standard measurement of extreme poverty—and half of these individuals lived in Sub-Saharan Africa. While children make up 31% of the overall global population, they account for 50% of persons living in poverty. The extreme poverty rate declined steadily between 1990 and 2019. Before the COVID-19 pandemic, compared with 1990, there were more than 1 billion fewer people living in poverty despite growth in the global population of more than 2 billion during that time. The COVID-19 pandemic led to significant increases in poverty, especially in low- and middle- income countries. The poorest bore the brunt of the economic shocks of COVID-19. The global poorest lost twice as much income as the richest between 2020 and 2022.

RISK FACTORS FOR DISEASE BURDEN

The GBD study found that the three leading risk factors for global disease burden in 2019 were (in order of frequency) high systolic blood pressure, smoking, and high fasting plasma glucose—a substantial change from 1990, when childhood malnutrition was ranked first. Although its prevalence has declined, maternal and childhood malnutrition remain the leading risk factors for death among children < 5 years of age. In an era that has seen obesity become a major health concern in many developed countries, the persistence of undernutrition is cause for consternation. In its rural reaches, no health care initiative, however generously funded, will be effective or comprehensive without addressing undernutrition.

In an analysis that examined how specific diseases and injuries are affected by environmental risk, the WHO estimates that 24% of all

TABLE 485-1 Leading Causes of Burden of Disease (DALYs), 2019

DISEASE OR INJURY	DALYS (MILLIONS)	PERCENTAGE OF TOTAL DALYs
World	**2540.0**	**100**
1. Neonatal disorders	185.9	7.3
2. Ischemic heart disease	182.0	7.2
3. Stroke	143.2	5.7
4. Lower respiratory infection	97.2	3.8
5. Diarrheal diseases	80.9	3.2
6. COPD	74.4	2.9
7. Road injuries	72.9	2.9
8. Diabetes	70.9	2.8
9. Low back pain	63.7	2.5
10. Congenital defects	52.8	2.1
Low-Income Countries[a]		
1. Neonatal disorders	44.4	17.4
2. Lower respiratory infection	24.5	7.6
3. Malaria	23.3	7.2
4. Diarrheal disease	23.2	7.2
5. HIV/AIDS	14.5	4.5
6. Congenital defects	12.8	4.0
7. Tuberculosis	12.2	3.8
8. Stroke	8.6	2.7
9. Road injuries	6.2	2.3
10. Ischemic heart disease	7.4	2.3
High-Income Countries[a]		
1. Ischemic heart disease	26.8	7.6
2. Low back pain	17.7	4.9
3. Stroke	15.8	4.4
4. Lung cancer	13.7	3.9
5. Diabetes	13.3	3.7
6. COPD	12.0	3.4
7. Falls	9.7	2.7
8. Alzheimer's disease	9.2	2.6
9. Other musculoskeletal	9.1	2.6
10. Depressive disorders	8.6	2.4
Sub-Saharan Africa		
1. Neonatal disorders	70.5	13.8
2. Malaria	43.2	8.5
3. Diarrheal diseases	42.0	8.2
4. Lower respiratory infection	41.1	8.0
5. HIV/AIDS	35.6	7.0
6. Congenital defects	18.1	3.6
7. Tuberculosis	17.6	3.4
8. Road injuries	11.1	2.2
9. Stroke	10.8	2.1
10. Meningitis	10.4	2.0

[a]The World Bank classifies high-income countries as those whose gross national income (GNI) per capita is ≥$14,005. Low- and middle-income countries are categorized as low income (GNI per capita, <$1,145), lower-middle income (GNI per capita, $1,146–$4,515), and upper-middle income (GNI per capita, $4,516–$14,005) (*https://datahelpdesk.worldbank.org/knowledgebase/articles/906519*).

Abbreviations: COPD, chronic obstructive pulmonary disease; DALYs, disability-adjusted life-years.

Source: Institute for Health Metrics and Evaluation, University of Washington (2023). Data available at *https://vizhub.healthdata.org/gbd-compare/*. Accessed December 15, 2023.

TABLE 485-2 Leading Causes of Death Worldwide, 2019

DISEASE OR INJURY	DEATHS (MILLIONS)	PERCENTAGE OF TOTAL DEATHS
World	**56.5**	**100**
1. Ischemic heart disease	9.1	16.1
2. Stroke	6.6	11.7
3. COPD	3.3	5.8
4. Lower respiratory infection	2.5	4.4
5. Lung cancer	2.0	3.6
6. Neonatal disorders	1.9	3.3
7. Alzheimer's disease	1.6	2.9
8. Diabetes	1.6	2.7
9. Diarrheal diseases	1.5	2.7
10. Cirrhosis	1.5	2.6
Low-Income Countries[a]		
1. Neonatal disorders	0.5	9.7
2. Lower respiratory infection	0.4	8.3
3. Diarrheal diseases	0.3	6.8
4. Stroke	0.3	6.5
5. Ischemic heart disease	0.3	6.2
6. Malaria	0.3	6.1
7. Tuberculosis	0.3	5.5
8. HIV/AIDS	0.2	5.0
9. Congenital defects	0.1	2.8
10. COPD	0.1	2.5
High-Income Countries[a]		
1. Ischemic heart disease	1.8	16.4
2. Stroke	0.9	8.4
3. Lung cancer	0.7	6.4
4. Alzheimer's disease	0.7	6.3
5. COPD	0.5	4.9
6. Lower respiratory infection	0.4	4.1
7. Colorectal cancer	0.4	3.8
8. Chronic kidney disease	0.3	3.0
9. Diabetes	0.3	2.3
10. Pancreatic cancer	0.2	2.1
Sub-Saharan Africa		
1. Neonatal disorders	0.7	9.9
2. Lower respiratory infection	0.7	8.9
3. HIV/AIDS	0.6	8.4
4. Malaria	0.6	7.8
5. Diarrheal diseases	0.6	7.8
6. Stroke	0.4	5.3
7. Tuberculosis	0.4	5.1
8. Ischemic heart disease	0.4	5.0
9. Congenital defects	0.2	2.5
10. Cirrhosis	0.2	2.5

[a]The World Bank classifies high-income countries as those whose gross national income (GNI) per capita is ≥$14,005. Low- and middle-income countries are categorized as low income (GNI per capita, <$1,145), lower-middle income (GNI per capita, $1,146–$4,515), and upper-middle income (GNI per capita, $4,516–$14,005) (*https://datahelpdesk.worldbank.org/knowledgebase/articles/906519*).

Abbreviation: COPD, chronic obstructive pulmonary disease.

Source: Institute for Health Metrics and Evaluation, University of Washington (2023). Data available at *https://vizhub.healthdata.org/gbd-compare/*. Accessed December 15, 2023.

deaths and 28% of deaths among children <5 years of age in 2019 were due to modifiable environmental factors: some 17 million children die every year from causes related to unhealthy environments, including the nearly 500,000 deaths stemming from a lack of access to clean water and sanitation. Many of these modifiable factors lead to child and adult deaths from infectious pathologies; others lead to deaths from malignancies. Risk factors such as indoor air pollution due to use of solid fuels account for 25% of DALYs due to lower respiratory infections globally. Various forms of unintentional injury and malaria top the list of health problems to which environmental factors contribute.

The third edition of *Disease Control Priorities* (DCP3), published as a set of serial volumes based on content area, provides evidence-based recommendations and cost-effectiveness analyses for numerous interventions, with attention to strategies for strengthening health systems. Cost-effectiveness analyses that compare relatively equivalent interventions in order to facilitate sound decisions under constraint are necessary; however, these analyses, as the DCP3 authors acknowledge, are unreliable when based on an incomplete knowledge of cost and evolving evidence of effectiveness. As both resources and objectives for global health initiatives grew, cost-effectiveness analyses (particularly those based on older evidence) sometimes steered policy makers and public health experts toward low-cost but ultimately ineffective interventions or away from higher-priced but effective ones. Thus, we use the term *global health equity* to emphasize the need to ensure equitable access to high-value health interventions. To illustrate these points, it is instructive to look to HIV/AIDS, which in the course of the last four decades has become one of the world's leading infectious causes of adult death.

HIV INFECTION/AIDS

Chapter 208 provides an overview of the global HIV epidemic today. Approximately 39 million people worldwide were living with HIV infection in 2022, and it was the underlying cause of death for 630,000 people that year. Approximately 67% of all people living with HIV live in Sub-Saharan Africa. Here the discussion will be limited to HIV/AIDS in the developing world. Lessons learned from tackling HIV/AIDS in low-resource settings are highly relevant to discussions of other chronic diseases, including NCDs, for which effective therapies have been developed. In the United States, after the mid-1990s, ART transformed HIV infection from an inescapably fatal disease into a manageable chronic illness. Across high-income countries, improved ART has dramatically prolonged life expectancy for people living with HIV infection, which now approaches that of the general population. This success rate exceeds that obtained with almost any treatment for adulthood cancer or for complications of coronary artery disease. In developing countries, treatment has been offered broadly only since 2003. Before 2003, many arguments were raised to justify not moving forward rapidly with ART programs for people living with HIV/AIDS in resource-limited settings. The standard litany included the price of therapy compared with the poverty of patients, the complexity of the intervention, the lack of infrastructure for laboratory monitoring, and the lack of trained health care providers. Narrow cost-effectiveness arguments that created false dichotomies—prevention *or* treatment rather than their synergistic integration—too often went unchallenged by policy makers, public health experts, and health economists. As a cumulative result of these delays in the face of health disparities both old and new, there were millions of premature deaths.

Disparities in access to HIV treatment did give rise to widespread moral indignation and a new type of health activism. In several middle-income countries, including Brazil, public programs have helped bridge the global access gap. Other innovative projects pioneered by international nongovernmental organizations (NGOs) in diverse settings such as Haiti and Rwanda have established that a simple approach to ART based on intensive community engagement and social and economic support for patients and their community-based health workers can achieve remarkable results **(Fig. 485-2)**.

During the past decade, the availability of ART has increased sharply in the low- and middle-income countries that have borne the greatest burden of the HIV/AIDS pandemic. In 2000, few people living with HIV/AIDS in these nations had access to ART, whereas by 2022, 76% of people living with HIV infection were receiving ART. In light of these dramatic gains, coverage targets have grown more ambitious; for example, in 2014, UNAIDS set the 90-90-90 targets, which aimed to have 90% of people living with HIV know their status, 90% of those with HIV treated with ART, and 90% of those on treatment achieving viral load suppression by 2020. Five countries met these ambitious goals: Botswana, Eswatini, Rwanda, the United Republic of Tanzania, and Zimbabwe. These goals were updated to "95-95-95" targets for 2030, and by 2023, these same five countries had already achieved them.

This scale-up was made possible by several developments: a staggering drop in the cost of generically manufactured ART, the development of a standardized approach to treatment, substantial investments by funders, and the political commitment of governments to afford ART as a public good. Civil-society AIDS activists spurred many of these efforts.

Starting in the early 2000s, a combination of factors, including work by the Clinton HIV/AIDS Initiative (now known as the Clinton Health Access Initiative) and Médecins Sans Frontières, led to the availability of generic ART medications. While first-line ART cost >$10,000 per patient per year in 2000, first-line regimens in low- and middle-income countries are now available for <$45 per year. At the same time, fixed-dose combinations made multidrug regimens easier to administer. Also, around this time, the WHO began advocating a public health approach to the treatment of people with AIDS in low-resource settings; this approach promised—thanks to dropping viremia—to lower transmission rates and, if universally available, to end almost all mother-to-child transmission. Derived from models of care pioneered by the NGO Partners In Health and other groups, this approach proposed the use of standard first-line treatment regimens based on a simple five-drug formulary, with a more complex (and more expensive) set of second-line options in reserve. Clinical protocols were standardized, and intensive training packages for health professionals and community health workers were developed and implemented in many countries. Early rollout efforts were supported by new funding from the Global Fund and PEPFAR. In 2003, lack of access to ART was declared a global public health emergency by the WHO and UNAIDS, and those two agencies launched the 3 by 5 Initiative, setting an ambitious target: to have 3 million people in developing countries on treatment by the end of 2005. Many countries set corresponding national targets and have worked to integrate ART into their national AIDS programs and health systems and to harness the synergies between HIV/AIDS treatment and prevention activities. External funding to

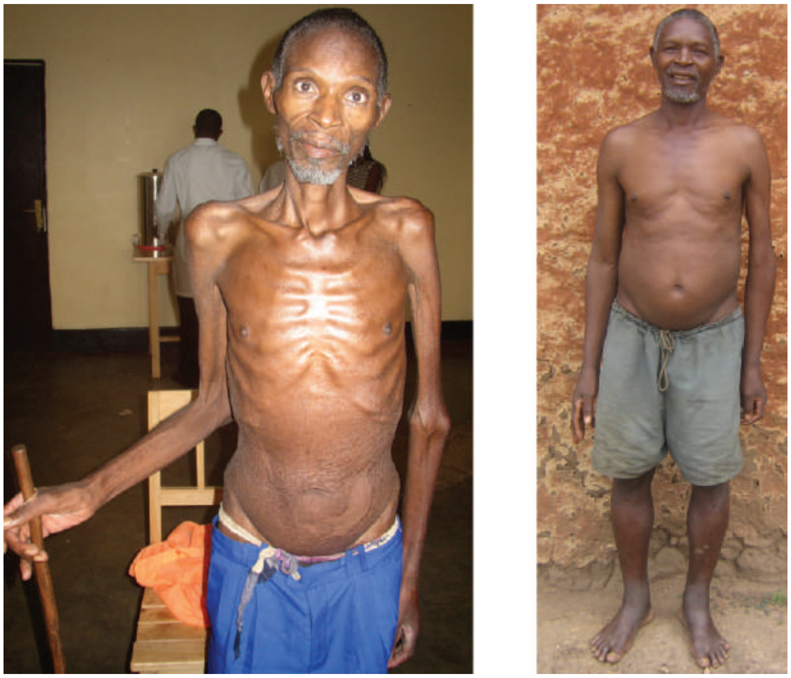

FIGURE 485-2 An HIV and tuberculosis (TB)–co-infected patient in Rwanda before (***left***) and after (***right***) 6 months of treatment.

fight HIV/AIDS in low- and middle-income countries increased dramatically during this period and beyond, rising from $332 million in 1996 to $9.9 billion in 2021. The integration of prevention and care led to a sharp drop in transmission—a 96% decline according to one review of the impact of ART rollout in heavily burdened countries in Africa and the Caribbean.

Further lessons with implications for policy and action have come from efforts now under way among lower-income countries. Rwanda provides an example: since 2000, mortality from HIV disease has fallen by 85% as the country—despite its relatively low gross national income—has provided almost universal access to ART. The reasons for this success include strong national leadership, evidence-based policy, cross-sector collaboration, community-based care, and a deliberate focus on a health-systems approach that embeds HIV/AIDS treatment and prevention in the primary health care service delivery platform. As we will discuss later in this chapter, these principles can be applied to other conditions, including NCDs.

TUBERCULOSIS

Chapter 183 provides a concise overview of the pathophysiology and treatment of tuberculosis. In 2022, an estimated 1.3 million people died from *Mycobacterium tuberculosis* infection; this figure made tuberculosis the second leading single infectious killer after COVID-19 of adults globally. The disease is closely linked to HIV infection in much of the world: of the 10.6 million estimated new cases of tuberculosis in 2022, 890,000 occurred among people living with HIV. A much more substantial proportion of the resurgence of tuberculosis registered in southern Africa is attributed to HIV co-infection. Even before the advent of HIV, however, it was estimated that fewer than one-half of all cases of tuberculosis in developing countries were ever diagnosed. Primarily because of the common failure to diagnose and treat tuberculosis, international authorities devised a single strategy to reduce the burden of disease. In the early 1990s, the World Bank, the WHO, and other international bodies promoted the DOTS strategy (*d*irectly *o*bserved *t*herapy using *s*hort-course isoniazid- and rifampin-based regimens) as highly cost-effective. Passive case-finding of smear-positive patients was central to the strategy, as was an uninterrupted drug supply.

DOTS was clearly effective for most uncomplicated cases of drug-susceptible tuberculosis, but several shortcomings were soon identified. First, the diagnosis of tuberculosis based solely on sputum smear microscopy—a method dating from the late nineteenth century—is not sensitive. Many cases of pulmonary tuberculosis and all cases of exclusively extrapulmonary tuberculosis are missed by smear microscopy, as are most cases of active disease in children. Second, passive case-finding relies on the availability of health care services, which is uneven in the settings where tuberculosis is most prevalent. Third, patients with multidrug-resistant tuberculosis (MDR-TB) are by definition infected with strains of *M. tuberculosis* resistant to isoniazid and rifampin; thus, exclusive reliance on these drugs is unwarranted in settings in which drug resistance is an established problem.

The crisis of antibiotic resistance registered in U.S. hospitals is not confined to the industrialized world or to common bacterial infections. While the great majority of patients sick with and dying from tuberculosis are afflicted with strains susceptible to all first-line drugs, a substantial minority of patients with tuberculosis in some settings are infected with strains of *M. tuberculosis* resistant to at least one first-line antituberculosis drug. Globally in 2021, an estimated 4% of all patients with new *M. tuberculosis* infections and 18% of all previously treated patients were infected with rifampin-resistant or MDR strains; most of these cases resulted from primary transmission. It was clear that poor infection control in hospitals and clinics in the face of delays in the initiation of effective therapy led to explosive and lethal epidemics due to these strains. To improve DOTS-based responses to MDR-TB, global health authorities adopted DOTS-Plus, which adds the diagnostics and drugs necessary to manage drug-resistant disease. Even as DOTS-Plus was being piloted in resource-constrained settings, however, new strains of extensively drug-resistant (XDR) *M. tuberculosis* (resistant to isoniazid and rifampin, any fluoroquinolone, and at least one injectable second-line drug) had already threatened the success of tuberculosis control programs in beleaguered South Africa, for example, where high rates of HIV infection had led to a doubling in the incidence of tuberculosis over the preceding decade. Genetic fingerprinting of cultures of infected sputum and tissues suggest that patients may be infected by more than one strain. Despite the poor capacity for detection of MDR- and XDR-TB in most resource-limited settings, an estimated 450,000 cases of MDR-TB were thought to have occurred in 2021. Approximately 8% of these cases were caused by XDR strains.

TUBERCULOSIS AND AIDS AS CHRONIC DISEASES: LESSONS LEARNED

Strategies effective against MDR-TB have implications for the management of drug-resistant HIV infection and even drug-resistant malaria, which, through repeated infections and a lack of effective therapy, has become a chronic disease in parts of Africa (see "Malaria," below). As new therapies, whether for tuberculosis or for hepatitis C infection, become available, many of the problems encountered in the past will recur. Indeed, examining AIDS and tuberculosis as chronic diseases—instead of simply communicable ones—makes it possible to draw several conclusions, many of them pertinent to global health equity in general.

First, the chronic infections discussed here are best treated with multidrug regimens to which the infecting strains are susceptible. This is true of chronic infections due to many bacteria, fungi, parasites, or viruses; even acute infections such as those caused by *Plasmodium* species are not reliably treated with a single drug.

Second, charging fees for AIDS prevention and care poses insurmountable problems for persons living in poverty, many of whom are unable to pay even modest amounts for services or medications. Like efforts to battle airborne tuberculosis, such services might best be seen as a public good promoting public health. Initially, a subsidy approach will require sustained donor contributions, but many African countries have set targets for increased national investments in health—a pledge that could render ambitious programs sustainable in the long run, as the Rwanda experience suggests. Meanwhile, as local investments increase, the price of AIDS care continues to decrease. The use of generic medications means that ART can now cost <$0.13 per day.

Third, the effective scale-up of pilot projects requires strengthening and sometimes rebuilding of health care systems, including those charged with delivering primary care. In the past, the lack of health care infrastructure has been cited as a barrier to providing ART in the world's poorest regions; however, AIDS resources, which are at last considerable, may be marshaled to rebuild public health systems in Sub-Saharan Africa and other HIV-burdened regions—precisely the settings in which tuberculosis is resurgent. Failure to pursue such a health-systems approach after civil wars ended in Sierra Leone and Liberia accounts for much of their extreme vulnerability to Ebola a decade later.

Fourth, the lack of trained health care personnel, most notably doctors and nurses, still must be addressed. The WHO recommends a minimum of 1 physician per 1000 persons, but many countries, especially in Sub-Saharan Africa, fall far short of that target. Specifically, ~45% of WHO member states report not achieving that target. In Sub-Saharan Africa, there were 0.3 physicians per 1000 people. In contrast, the United States and Cuba report 2.5 and 8.4 doctors per 1000 population, respectively. Similarly, ~50% of WHO member states report having fewer than 3 nurses and midwives per 1000 population. Sub-Saharan Africa bears >20% of the global burden of disease but has access to only 3% of the world's health workers. Further inequalities in health care staffing exist within countries. Rural–urban disparities in health care personnel mirror disparities of both wealth and health. For instance, in Sierra Leone, an estimated 75% of the national health workforce is concentrated in urban areas, where just 44% of the population lives. Even community health workers trained to provide first-line services to rural populations often transfer to urban districts.

In what is termed the "brain drain," many physicians and nurses emigrate from their home countries to pursue opportunities abroad, leaving behind health systems that are understaffed and ill-equipped to deal with either emergencies like Ebola or the usual burden of disease.

One reason doctors and nurses leave Sub-Saharan Africa and other low-income areas is that they lack the tools to practice there. Funding for "vertical" (disease-specific) programs can be used not only to strengthen health systems but also to recruit and train physicians and nurses to underserved regions where they, in turn, can help to train and then work with community health workers in supervising care for patients with AIDS and many other diseases within their communities. Such training should be undertaken even where physicians are abundant, since close community-based supervision represents the highest standard of care for chronic disease, whether in developing or developed countries. The United States, which has a dearth of health care providers in many of its poor and rural communities, has much to learn from Rwanda in this regard.

Fifth, the many barriers to adequate health care and patient adherence that are raised by extreme poverty can be removed only with the deployment of "wrap-around services": food supplements for the hungry, help with transportation to clinics, childcare, and housing. Extreme poverty makes it difficult for many patients to comply with therapy for chronic diseases, whether communicable or not. Experience shows, however, that these many barriers can be more readily surmounted than the extreme poverty itself to which chronic disease and acute infection contribute substantially. Indeed, poverty in its many dimensions is far and away the greatest obstacle to the scale-up of treatment and prevention services.

Finally, there is a need for a renewed basic-science commitment to the discovery and development of vaccines; more reliable, less expensive diagnostic tools; and new classes of therapeutic agents. This need applies not only to HIV, tuberculosis, and malaria—against none of which there is a highly effective vaccine—but also to most other neglected diseases of poverty.

■ MALARIA

Chapter 231 reviews the etiology, pathogenesis, and clinical treatment of malaria, the world's fifth-ranking infectious killer. In 2022, there were 249 million cases of malaria, and the disease killed 608,000 people; 77% of these deaths occurred among children <5 years old. The poor disproportionately experience the burden of malaria. Sub-Saharan Africa was home to about 94% of all malaria cases and 95% of all deaths. Just 4 countries—Nigeria, the Democratic Republic of the Congo, Uganda, and Mozambique—account for more than half of total malaria deaths globally.

Malaria's human cost has been enormous, with the highest toll among children—especially African children—living in poverty. In 2022, $4.3 billion was spent on malaria worldwide, but the WHO target is more than twice that sum. Macroeconomic analyses estimate that malaria may reduce the per capita gross national product of a disease-endemic country by 50% relative to that of a non–malaria-endemic country. The causes of this drag include impaired cognitive development of children, decreased schooling, decreased savings, decreased foreign investment, and restriction of worker mobility. Microeconomic analyses focusing on direct and indirect costs estimate that malaria may consume >10% of a household's annual income in malaria endemic regions.

In part because of differences in vector distribution and climate, resource-rich countries offer few blueprints for malaria control and treatment that are applicable in tropical (and resource-poor) settings. In 2001, African heads of state endorsed the WHO Roll Back Malaria (RBM) campaign, which prescribes strategies appropriate for Sub-Saharan African countries. In 2008, the RBM partnership launched the Global Malaria Action Plan (GMAP). This strategy integrates prevention and care and calls for the avoidance of single-dose regimens and an awareness of existing drug resistance; the use of insecticide-treated bed nets (ITNs); indoor residual spraying; artemisinin-based combination therapy (ACT); intermittent preventive treatment during pregnancy; prompt diagnosis; and other vector control measures such as larviciding and environmental management.

Over the past two decades, these efforts have dramatically reduced the global burden of malaria resulting in the prevention of 1.7 billion cases and saving 10.6 million lives.

Meeting the challenge of malaria control will continue to require careful study of appropriate preventive and therapeutic strategies in the context of an increasingly sophisticated molecular understanding of pathogen, vector, and host. However, an appreciation of the economic and social devastation wrought by malaria—like that inflicted by diarrhea, AIDS, and tuberculosis—on the most vulnerable populations should heighten the level of commitment to critical analysis of ways to implement proven strategies for prevention and treatment.

Funding from the Global Fund, the Gates Foundation, the World Bank's International Development Association, and the U.S. President's Malaria Initiative, along with leadership from public health authorities, is critical to sustain the benefits of prevention and treatment. Building on the growing momentum of the last decade with adequate financial support, innovative strategies, and effective tools for prevention, diagnosis, and treatment, we may yet achieve the goal of a world largely free of malaria.

■ EBOLA

Chapter 216 provides an overview of the epidemiology, pathogenesis, and clinical manifestations of Ebola virus and Marburg virus infections. The 2013–2016 outbreak of Ebola virus disease in West Africa was the largest documented Ebola epidemic to date, with >28,000 recorded cases and 11,000 recorded deaths.

Prior to the outbreak, the health systems of the three most affected countries—Liberia, Guinea, and Sierra Leone—were among the world's weakest. Histories of extractive colonial and postcolonial commerce, the conditional aid policies of international financial institutions, recent civil conflict, and underresourced health ministries left this part of West Africa bereft of the means to deliver modern medicine and promote public health. In 2013, Sierra Leone had the world's highest maternal mortality ratio, with 1180 deaths per 100,000 live births. According to one estimate, Liberia had just 51 physicians working in the entire country before the Ebola epidemic, or roughly one physician per 100,000 persons. Clinics and hospitals were scarce across the region, especially in rural areas, and routinely lacked drugs, supplies, electricity, running water, laboratories, and personal protective equipment for the prevention of nosocomial infection. Such deficits were not surprising given these countries' meager public and private expenditures on health.

The unprecedented scale of the West African Ebola epidemic was largely a symptom of these chronically weak health systems. As a result, clinicians, patients' families, and other caregivers—tasked with nursing the sick and interring the dead but lacking the means to do so safely—faced disproportionately high risks of Ebola infection. Health facilities with poor infection control and unsafe burials served as amplifiers of transmission.

The quest to contain Ebola in West Africa was one of the largest global public health efforts at that time, but it was far from ambitious clinically. As in previous Ebola outbreaks, preventing new infections was often prioritized over improving survival among those already infected, leading to substandard care for most West African patients and high case-fatality rates—by WHO estimates, ~70%. However, in settings in which quality supportive and critical care could be provided, clinical outcomes among Ebola-infected patients affirmed that Ebola virus disease is treatable, even in the absence of specific antiviral therapies and experimental drugs.

As with efforts to combat AIDS and tuberculosis, the global response to Ebola reveals the unintended consequences of pitting preventive strategies against therapeutic ones—and the pull of debates about scarcity. Misguided (and often contradictory) public health messaging, distrust of disease-control and social mobilization teams, punitive containment measures, and the unavailability of safe Ebola treatment units capable of delivering effective clinical care deterred individuals from presenting to health facilities, reporting symptomatic patients and their contacts, and cooperating with epidemic response activities. The resulting epidemic of mistrust facilitated the further spread of new infections by impeding surveillance, timely diagnosis, contact tracing, and patient isolation.

In August 2018, a new Ebola outbreak was detected in eastern Democratic Republic of the Congo (DRC) and soon became the world's second largest on record. In June 2020, when it was declared over, it had sickened ~3500 persons and killed two-thirds of them. Containment was complicated by armed conflict in the affected region, which had long experienced strife, impoverishment, and colonial and postcolonial extraction, feeding a well-founded suspicion of foreign intervention.

Despite these challenges, responders benefitted from the arrival of new tools to prevent, diagnose, and treat Ebola. A new vaccine was extensively deployed using a ring vaccination strategy. There was significant progress in the quality of supportive care provided to patients in DRC, with more routine monitoring, improved access to clinical laboratory services, better staffing of treatment centers, and more aspirational clinical protocols. Nevertheless, the overall case-fatality rate in DRC reveals that these welcome medical innovations did not reach everyone in need of them.

COVID-19

Chapter 205 provides an overview of the epidemiology, pathogenesis, and clinical manifestations of COVID-19 infection. As we near the end of the fourth year of the pandemic, there have been more than 7 million confirmed deaths due to COVID-19 infection and 780 million confirmed cases. Reasonable estimates suggest that the true total of deaths is likely ~18 million. During this period, COVID-19 was among the top five causes of death worldwide. In the early months of the COVID-19 pandemic, many health systems in high-income countries were strained to treat the surge of patients afflicted by it. The pandemic revealed deep structural deficiencies in our collective global ability to recognize and contain such "novel" pathogens. Although SARS-CoV-2 is "novel" in relation to our human immune systems, its rapid march across the globe followed the familiar pattern of many pandemic pathogens before it, such as HIV, tuberculosis, and cholera. It has particularly high attack rates among vulnerable populations, such as those experiencing homelessness and nursing home residents, and in poor communities where insufficient housing, food insecurity, and marginal employment in low-paid service jobs have resulted in inability to safely social distance and quarantine. The United States suffered during the early parts of this pandemic from underinvestment in public health systems that can provide surveillance testing at scale and perform robust contact tracing. On the global level, profound inequities occurred in access to vaccines and therapeutics. The COVID-19 Vaccines Global Access (COVAX) initiative had broad global support and ambitious goals but was unable to meet its targets. This was due, in part, to lack of access to sufficient vaccine doses as a result of excess procurement by high-income countries for their own use. COVAX successfully delivered 2 billion doses of vaccine and is thought to have averted 2.7 million deaths. We share a deep hope that this pandemic may catalyze a broader recognition about the effects of poverty on health, about the need to achieve universal access to health care for all the world's people, and about the urgency of strengthening global public health systems.

"NONCOMMUNICABLE" CHRONIC DISEASES

Although the burden of communicable diseases—especially HIV infection, tuberculosis, and malaria—still accounts for the majority of deaths in resource-poor regions within Sub-Saharan Africa and in the poorest reaches of several first-world cities, 74% of all deaths worldwide in 2022 were attributed to NCDs. Although we use this term to describe cardiovascular diseases, cancers, diabetes, and chronic lung diseases, this usage masks important distinctions. For instance, two significant NCDs in low-income countries, rheumatic heart disease (RHD) and cervical cancer, represent the chronic sequelae of infections with group A *Streptococcus* and human papillomavirus, respectively, and it is in these countries that the burden of disease due to NCDs is rising most rapidly. A little more than three-quarters of deaths attributable to NCDs occur in low- and middle-income countries, which also account for 85% of all early NCD-related deaths—a figure representing ~17 million persons and exceeding the total number of deaths due to AIDS, tuberculosis, and malaria combined. By 2030, NCDs will account for 52 million deaths annually worldwide if no additional action is taken. The recent increase in resources for and attention to communicable diseases is both welcome and long overdue, but developing countries are already carrying a "double burden" of communicable and noncommunicable diseases.

Diabetes, Cardiovascular Disease, and Cancer: A Global Perspective In contrast to tuberculosis, HIV infection, and malaria—diseases caused by single pathogens that damage multiple organs—cardiovascular diseases reflect injury to a single organ system downstream of a variety of insults, both infectious and noninfectious. Some of these insults result from rapid changes in diet and labor conditions; others are of a less recent vintage. The burden of cardiovascular disease in low-income countries represents one consequence of decades of neglect of health systems. Furthermore, cardiovascular research and investment have long focused on the ischemic conditions that are increasingly common in high- and middle-income countries.

Predictions of an imminent rise in the share of deaths and disabilities due to NCDs in developing countries have led to calls for preventive policies to improve diet, increase exercise, and restrict tobacco use, along with the prescription of multidrug regimens for persons at high-level vascular risk. Although this agenda could do much to prevent pandemic NCDs, it will do little to help persons with established heart disease stemming from nonatherogenic pathologies.

The misperception of cardiovascular diseases as a problem primarily of elderly populations in middle- and high-income countries has contributed to the neglect of these diseases by global health institutions, including regionally focused ones. Even in Eastern Europe and Central Asia, where the collapse of the Soviet Union was followed by a catastrophic surge in cardiovascular disease deaths (mortality rates from ischemic heart disease nearly doubled between 1991 and 1994 in Russia, for example), the modest flow of overseas development assistance to the health sector during these troubled years focused on the communicable causes that accounted for <1 in 20 excess deaths during that period.

DIABETES The International Diabetes Federation reports that the number of diabetic adult patients in the world is expected to increase from 537 million in 2022—~1 in 11 adults—to 783 million by 2045. Already, a significant proportion of patients with diabetes, 3 in 4 adults with the condition, live in developing countries where, because those affected are often younger, the complications of micro- and macrovascular disease take a far greater toll. Globally, these complications are a major cause of disability and reduced quality of life: a high fasting plasma glucose level ranks third among risks for disability and global mortality. The GBD 2017 study estimates that diabetes accounted for 1.6 million deaths in 2019; 84% of these deaths occurred in low- and middle-income countries.

CARDIOVASCULAR DISEASE Because systemic investigation of the causes of stroke and heart failure in Sub-Saharan Africa has begun only recently, little is known about the impact of elevated blood pressure in this portion of the continent. Modestly elevated blood pressure in the absence of tobacco use in populations with low rates of obesity may confer little risk of adverse events in the short term. In contrast, persistently elevated blood pressure goes largely undetected, untreated, and uncontrolled in this part of the world. In the cohort of men assessed in the Framingham Heart Study, the prevalence of blood pressures above 210/120 mmHg—severe hypertension—declined from 1.8% in the 1950s to 0.1% by the 1960s with the introduction of effective antihypertensive agents. Although debate continues about appropriate screening strategies and treatment thresholds, Africa's rural health centers, run largely by nurses, must quickly gain access to antihypertensive medications.

The epidemiology of heart failure also reflects inequalities in risk factor prevalence and in access to therapy. The reported burden of this condition has remained unchanged since the 1950s, but the causes of heart failure and the age of the people affected vary across the globe. Heart failure as a consequence of pericardial, myocardial, endocardial, or valvular injury is a leading cause of hospitalization in the United States and Europe, representing ~3% of all hospitalizations, and is estimated to account for a substantial proportion of medical admissions in hospitals in low-income countries as well. In high-income countries, coronary artery disease and hypertension among the elderly account for most cases of heart failure. Among the world's poorest 1 billion people, however, heart failure reflects poverty-driven exposure of children and young adults to rheumatogenic strains of streptococci and cardiotropic microorganisms (e.g., HIV, *Trypanosoma cruzi*, enteroviruses, *M. tuberculosis*), untreated high blood pressure, and nutrient deficiencies. The mechanisms underlying other causes of heart failure common in these populations—such as idiopathic dilated cardiomyopathy, peripartum cardiomyopathy, and endomyocardial fibrosis—remain unclear.

In stark contrast to the extraordinary lengths to which clinicians in wealthy countries will go to treat ischemic cardiomyopathy among elderly patients, little attention has been paid to young patients with nonischemic cardiomyopathies in resource-poor settings. Nonischemic cardiomyopathies, such as those due to hypertension, RHD, and chronic lung disease, account for >90% of cases of cardiac failure in Sub-Saharan Africa and include poorly understood entities such as peripartum cardiomyopathy (which has an incidence in rural Haiti of 1 per 300 live births) and HIV-associated cardiomyopathy. Lessons learned in the scale-up of chronic care for HIV infection and tuberculosis may be illustrative as progress is made in establishing the means to deliver heart-failure medications to these patients.

Some of the lessons learned from the chronic infections discussed above are, of course, relevant to cardiovascular disease, especially those classified as NCDs but caused by infectious pathogens. Integration of prevention and care remains as important today as in 1960 when Paul Dudley White and his colleagues found little evidence of myocardial infarction in the region near the Albert Schweitzer Hospital in Lambaréné, Gabon, but reported that "the high prevalence of mitral stenosis is astonishing." They termed it a duty to integrate prevention with penicillin prophylaxis and care, including medical management and surgery, when indicated. "The same responsibility," they agreed, "exists for those with correctable congenital cardiovascular defects."

RHD affects almost 40 million people worldwide, with ~2.8 million new cases each year. A meta-analysis of data on heart failure in Sub-Saharan Africa found that RHD was the third most common cause of heart failure in the region. This disease, which may cause endocarditis or stroke, leads to >300,000 deaths per year—almost all occurring in developing countries. Results from 14 low- and middle-income countries included in the Global Rheumatic Heart Disease Registry showed that mortality was significantly higher among patients living in low-income countries and among the less educated. Recent studies in Rwanda and Ethiopia have confirmed a high prevalence of RHD among schoolchildren, including those that are asymptomatic. In part because the prevention of RHD has not advanced since the disease's disappearance in wealthy countries, no part of Sub-Saharan Africa has eradicated RHD despite examples of success in Costa Rica, Cuba, and some Caribbean nations.

Strategies to eliminate RHD may depend on active case-finding, with confirmation by echocardiography, among high-risk groups as well as on efforts to expand access to surgical interventions among children with advanced valvular damage. Partnerships between established surgical programs and areas with limited or nonexistent facilities may help expand the capacity to provide lifesaving interventions to patients who otherwise would die early and painfully. Such partnerships can speed the further development of regional centers of excellence equipped to provide consistent, accessible, high-quality services to those now without them.

CANCER Low- and middle-income countries accounted for ~70% of the 10 million deaths due to cancer worldwide in 2020. By 2040, annual mortality from cancer is expected to increase to >16 million deaths—with developing countries experiencing a sharper increase than developed nations. "Western" lifestyle changes may be responsible for the increased incidence of cancers of the breast, colon, and prostate among populations in low- and middle-income countries, but historic realities, sociocultural and behavioral factors, genetics, and poverty itself already have a profound impact on cancer-related mortality and morbidity rates. Cancer-causing infections, such as human papillomavirus, hepatitis B virus, and *Helicobacter pylori*, are responsible for up to 50% of cancer cases in some countries in Sub-Saharan Africa. Infectious causes of cancer will continue to have a much larger impact in developing countries. Environmental and dietary factors, such as indoor air pollution and high-salt diets, also contribute to increased rates of certain cancers (e.g., lung and gastric cancers). Tobacco use (both smoking and chewing) is the most important source of increased mortality rates from lung, bladder and oral cancers. In contrast to decreasing tobacco use in many developed countries, the number of smokers is growing in developing countries, especially among women and young persons.

For many reasons, outcomes of malignancies are far worse in developing countries than in developed nations. As currently funded, overstretched health systems in poor countries are not capable of early detection; at the time of tissue diagnosis, the majority of patients already have incurable malignancies. Treatment of cancers is available for only a very small number of mostly wealthy citizens in most poor countries, and even when treatment is available, the range and quality of services are often substandard. Yet this need not be the future. Twenty years ago, MDR-TB and HIV infections were widely deemed untreatable in settings of great poverty. The feasibility of creating innovative programs that reduce technical and financial barriers to the provision of care for treatable malignancies among the world's poorest populations is now clear **(Fig. 485-3)**. Several middle-income countries, including Mexico, have expanded publicly funded cancer care to reach poorer populations. This commitment of resources has dramatically improved outcomes for cancers, from childhood leukemia to cervical cancer.

Prevention of Noncommunicable Diseases False dichotomies, including those pitting prevention against care, persist in global health and reflect, in part, outmoded paradigms or a limited understanding of shifts in disease burden and causality as well as the dramatic variations in risk within a single nation. Moreover, such dichotomies or debates are sometimes politicized as a result of vested interests. Although globalization has had many positive effects, one negative effect has been the growth in both developed and developing countries of well-financed lobbies that have aggressively promoted unhealthy dietary changes and increased consumption of alcohol and tobacco. The WHO's 2003 Framework Convention on Tobacco Control represented a major advance, committing all of its signatories to a set of policy measures shown to reduce tobacco consumption.

The WHO estimates that 80% of all cases of cardiovascular disease and type 2 diabetes as well as 40% of all cancers can be prevented through healthier diets, increased physical activity, and avoidance of tobacco. These estimates mask large local variations. Although some evidence indicates that population-based measures can have some impact on these behaviors, it is sobering to note that increasing obesity levels have not been reversed in any population. Tobacco avoidance may be the most important and most difficult behavioral modification of all. In the twentieth century, 100 million people worldwide died of tobacco-related diseases; it is projected that >1 billion people will die of these diseases in the twenty-first century, with the vast majority of those deaths in developing countries. Today, ~80% of the world's 1.1 billion smokers live in low- and middle-income countries. In 2019, tobacco use caused 8.7 million deaths, including deaths due to second-hand smoke. The majority of these deaths occurred in low- and middle-income countries. However, there is well-proven evidence that changes in policy, such as taxes on tobacco and indoor and workplace smoking bans, are effective in decreasing the number of people using tobacco, reducing the amount of tobacco consumed, and preventing young people from starting to use tobacco.

MENTAL AND NEUROLOGIC HEALTH

In 2019, 1 in every 8 people, ~970 million people worldwide, lived with a mental health disorder, including >580 million people suffering from depression and anxiety disorders. One in 4 patients visiting a health service has at least 1 mental, neurologic, or behavioral disorder, but most of these disorders are neither diagnosed nor treated. More than 700,000 people die by suicide every year, and depressive disorders are the second leading cause of years lost to disability in the world today. Most low- and middle-income countries devote <1% of their health expenditures to mental health.

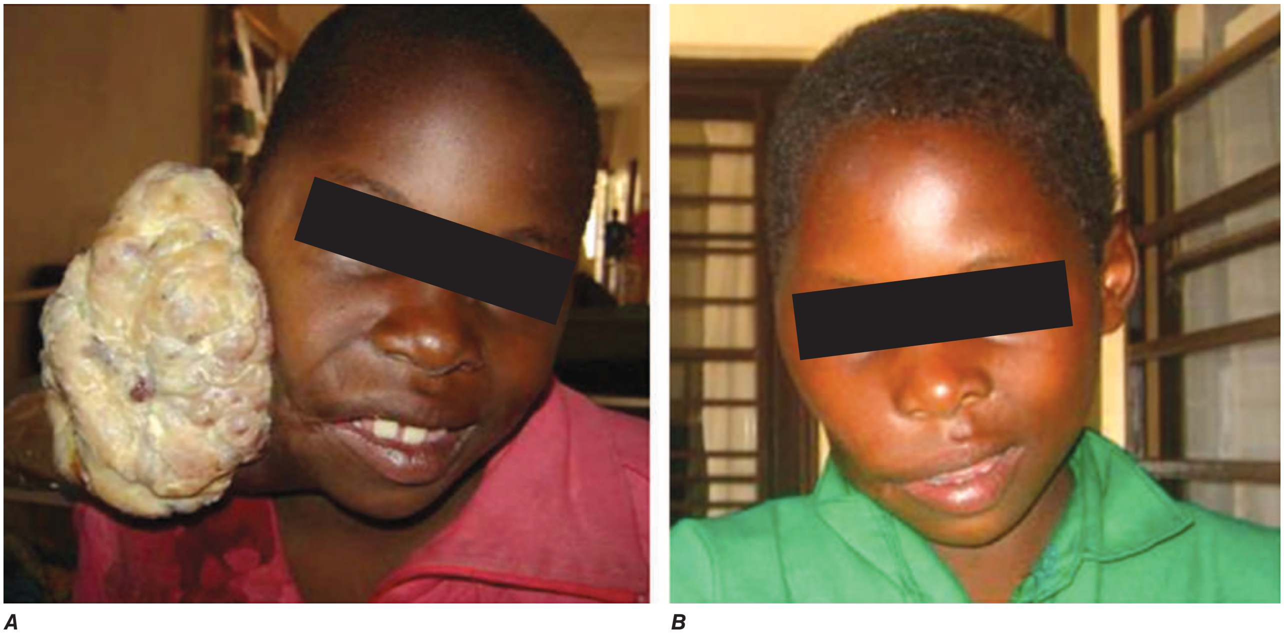

[redacted]

[redacted]

FIGURE 485-3 An 11-year-old Rwandan patient with embryonal rhabdomyosarcoma before (***left***) and after (***right***) 48 weeks of chemotherapy plus surgery. Fifteen years later, she is healthy with no evidence of disease.

Increasingly effective therapies exist for many of the major causes of mental disorders. One of the greatest barriers to delivery of such therapies is the paucity of skilled personnel. Most Sub-Saharan African countries have only a handful of psychiatrists, for example, and almost all of them practice in cities and are unavailable within the public sector or to patients living in poverty. Among the few patients who are fortunate enough to see a psychiatrist or neurologist, fewer still are able to adhere to treatment regimens: several surveys of already diagnosed patients ostensibly receiving daily therapy have revealed that, among the poor, multiple barriers prevent patients from taking their medications as prescribed. In one study from Kenya, no patients being seen in an epilepsy clinic had therapeutic blood levels of antiseizure medications, even though all had been prescribed these drugs. Moreover, many patients in this study had no detectable blood levels of these agents at all. The same barriers that prevent the poor from having reliable access to insulin or ART prevent them from benefiting from antidepressant, antipsychotic, and antiepileptic agents. To alleviate this problem, some authorities are proposing the training of health workers to provide community-based adherence support, counseling services, and referrals for patients in need of mental health services. One such program instituted in Goa, India, used lay counselors and resulted in a significant reduction in symptoms of common mental disorders among the target population.

CONCLUSION: TOWARD GLOBAL HEALTH EQUITY

There is a long way to go before evidence-based internal medicine is applied effectively and equitably among the world's poor. Public health strategies typically operate on the population level, while clinical practice, including the practice of internal medicine, often remains focused on individual patient care. However, global health equity depends on avoiding the false dichotomies of the past: neither public health nor clinical approaches alone are adequate to address the problems of global health. The integration of prevention and care, along with adequate funding, has shown that complex infectious diseases such as HIV/AIDS and tuberculosis are not impossible to manage, even though drug resistance and lack of effective health systems have complicated such work. Beyond what is usually termed communicable disease—i.e., in the arena of chronic diseases such as cardiovascular disease and mental illness—global health is still a nascent, but growing, endeavor. Efforts to address any one of these problems in settings of great scarcity need to be integrated into broader efforts to strengthen failing health systems and alleviate the growing personnel crisis within these systems. Such efforts must include the building of platforms for care delivery that are robust enough to incorporate new preventive, diagnostic, and therapeutic technologies rapidly in response to changes both in the burden of disease and in the needs not met by existing paradigms and systems of care delivery.

Academic medical centers have tried to address this gap as new technologies are introduced and assessed through clinical trials, but the reach of these institutions into settings of poverty is limited in rich and poor countries alike. When such centers link their capacities effectively to the public institutions charged with the delivery of health care to the poor, great progress can be made. For these reasons, scholarly work and practice in the field once known as "international health" and now often designated *global health equity* are changing rapidly. That scholarship and practice is still informed by the tension between clinical practice and population-based interventions, between analysis and action, and between prevention and care.

A number of university hospitals have developed training programs for physicians with an interest in global health. These programs provide training not only in the clinical problems encountered in settings of poverty, but also instruction in public health, health care service delivery, and strategies of health systems strengthening.

At least half of the world's population lacks reliable access to essential health services; the consequence is millions of preventable deaths each year. An absolute majority of these premature deaths occur in Africa, with the poorer regions of Asia not far behind. They include deaths from vaccine-preventable illness, deaths during childbirth, deaths from infectious diseases that might be cured with access to antibiotics and other essential medicines, deaths from malaria that would have been prevented by ITNs and access to therapy, and deaths from waterborne illnesses—in short, from failures of health care delivery. Other excess mortality is attributable to the inadequacy of efforts to develop new preventive, diagnostic, and therapeutic tools.

The development of these tools must be followed quickly by their equitable distribution. Those funding the discovery and development of new tools typically neglect the concurrent need for strategies to make them available to the poor. Indeed, some would argue that the biggest challenge facing those who seek to address this outcome gap is the lack of practical means of delivery in the most heavily affected regions. When new preventive and therapeutic tools are developed without concurrent attention to delivery or implementation, one encounters what are sometimes termed "perverse effects": even as new tools are developed, inequalities of outcome—lower morbidity and mortality rates among those who can afford access, with sustained high morbidity and mortality among those who cannot—grow in the absence of an equity plan to deliver the tools to those most at risk. Preventing such a future is among the most important goals of global health.

FURTHER READING

Bukhman G et al: The Lancet NCDI Poverty Commission: Bridging a gap in universal health coverage for the poorest billion. Lancet 396:991, 2020.

Cancedda C et al: Strengthening health systems while responding to a health crisis: Lessons learned by a nongovernmental organization during the Ebola virus disease epidemic in Sierra Leone. J Infect Dis 214:S153, 2016.

Farmer P: Chronic infectious disease and the future of health care delivery. N Engl J Med 369:2424, 2013.

Farmer P: *Fevers, Feuds, and Diamonds: Ebola and the Ravages of History*. New York, Farrar, Straus and Giroux, 2020.

GBD 2019 Diseases and Injuries Collaborators: Global Burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. Lancet 396:1204, 2020.

GBD 2019 Risk Factors Collaborators: Global Burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. Lancet 396:1223, 2020.

Institute for Health Metrics and Evaluation: *Financing Global Health 2021: Global Health Priorities in a Time of Change*. Seattle, Institute for Health Metrics and Evaluation, 2023.

Kim JY et al: Redefining global health-care delivery. Lancet 382:1060, 2013.

Watkins DA et al: Alma-Ata at 40 years: Reflections from the *Lancet* Commission on Investing in Health. Lancet 392:1434, 2018.

486

Emerging and Re-Emerging Infectious Diseases

David M. Morens, Anthony S. Fauci

EMERGING INFECTIOUS DISEASES: DEFINITION AND CLASSIFICATION

Pathogenic microorganisms and viruses have existed in the environment and in numerous animal species for millions of years. Humans have presumably had endemic infectious diseases since the origins of the human species, about 2 million years ago. However, until about 12,000 years ago, humans lived and moved about in small clans and tribal groups, experiencing limited contact with other humans or animals. During this pre-neolithic era, endemic infections were probably limited mostly to skin and gastrointestinal organisms. Even though contact with the environment and with animals undoubtedly led to sporadic infections with non-endemic pathogens, there was little opportunity for these pathogens to become widely spread among humans, i.e., to become epidemic.

This situation changed dramatically in the early neolithic age, about 10,000 BCE, in association with domestication of animals for food and for labor, planting and fertilization of crops, storage of water, organized disposal or diversion of sewage, growth of large settled villages and towns, labor specialization, and an enormous increase in human crowding and animal–human contacts. These new elements of human societal existence were associated with the emergence and human adaptation of existing enzootic and environmental organisms within populations large enough to sustain human-to-human spread, i.e., to cause emerging epidemics. It is believed that in this period the first significant emerging infectious diseases (EIDs) appeared.

The likelihood that numerous pandemic emergences occurred between the onset of the neolithic era and the era of microbial identification (beginning in the nineteenth century) is suggested by the large number of human pathogens that are today found globally in genetic and phenotypic forms that are identical or highly similar to each other—e.g., skin organisms such as staphylococci **(Chap. 152)**, streptococci **(Chap. 153)**, pneumococci **(Chap. 151)**, and corynebacteria **(Chap. 155)**; enteric pathogens like *Escherichia coli* **(Chap. 166)** and salmonellae **(Chap. 171)**; latently infecting neural viruses like herpesviruses **(Chap. 200)**; and sexually transmitted agents like human papillomavirus **(Chap. 203)**, gonorrhea **(Chap. 161)**, and syphilis **(Chap. 187)**. In some cases, microbial/virologic phylogenetic data allow rough estimation of the times of pathogen emergences, but even without such data, it is obvious that pandemic spread had to have occurred at some time in the past if we are to explain the global prevalence of many human pathogens today. EIDs and indeed pandemic IDs are thus a very old human phenomenon that we have only recently begun to consider.

EIDs have been defined and characterized as shown in **Table 486-1**. The importance of distinguishing between newly emerging IDs and re-emerging IDs (REIDs) has particular significance for clinicians, who usually are not only among the first to encounter the newly emerging group, but also among those with the greatest expertise in recognizing and dealing with the second group. It is also of note (Table 486-1) that subcategories of REIDs include accidental release of pathogens by human activities, e.g., vaccine-derived polioviruses or the sudden 1977 pandemic appearance of a long-extinct 1950s-era H1N1 influenza virus, presumably a result of undisclosed vaccine or other virologic research that led to viral "escape."

EMERGING INFECTIOUS DISEASES: THEIR IMPORTANCE

EIDs have been among the leading causes of death, disability, and social disruption throughout recorded human history **(Table 486-2)**. For example, it is believed by some historians that at least two of the biblical pharaonic plagues (around the thirteenth century BCE) were EIDs. The regionally pandemic "Plague of Athens" (430–425 BCE) is said to have brought about the end of the "Golden Age" of Greece.

TABLE 486-1 Emerging and Re-Emerging Infectious Diseases: Definitions, Categories, and Examples

Emerging infectious diseases (EIDs) are those recognized in humans for the first time, e.g., HIV/AIDS, Nipah virus infection, or severe acute respiratory syndrome (SARS) and COVID-19.

Re-emerging infectious diseases (REIDs) are those that have infected humans in the past and continue to reappear in new locations (e.g., West Nile virus in the United States and Russia in 1999), reappear in resistant or otherwise phenotypically different forms (e.g., influenza, methicillin-resistant *Staphylococcus aureus*, drug-resistant falciparum malaria), or reappear after apparent control or elimination (e.g., poliomyelitis in parts of Africa, cholera in Haiti in 2010 and elsewhere in association with natural disasters, wars, and mass migrations) or under unusual circumstances (e.g., deliberately released agents, including the 2001 anthrax bioterrorism attacks). Important subcategories of REIDs include the following:

- *REIDs related to accidental human release*—e.g., vaccine-derived polioviruses, epizootic vaccinia virus, and the 1979 Sverdlovsk laboratory explosion releasing anthrax spores
- *REIDs caused by human intent to harm (bioterrorism)*—e.g., the 1997 Oregon salad bar poisonings and the 2001 anthrax spore attacks in the United States

Established infectious diseases or *endemic infectious diseases* are those that have been prevalent for a sufficient period of time to allow for a relatively stable and predictable level of morbidity and mortality (e.g., many viral and bacterial respiratory and diarrheal diseases, including respiratory syncytial virus, endemic coronaviruses, noroviruses, pneumococcal disease, drug-susceptible malaria and tuberculosis, and many other tropical diseases such as helminthic and other parasitic diseases, many nosocomial infections).

TABLE 486-2 Selected Emerging Infectious Diseases of Note, 430 BCE to 2024 AD

YEAR	NAME	DEATHS	COMMENTS
430 BCE	"Plague of Athens"	~100,000	First identified transregional pandemic
541	Justinian plague (*Yersinia pestis*)	30–50 million	Pandemic; killed half of then-known world population
1340s	"Black Death" (*Yersinia pestis*)	~50 million	Pandemic; killed at least one-quarter of the known world population
1494	Syphilis (*Treponema pallidum*)	>50,000	Pandemic brought to Europe from the Americas
c. 1500	Tuberculosis	High millions	Ancient disease; became pandemic in Middle Ages
1520	*Hueyzahuatl* (*Variola major*)	3.5 million	Pandemic brought to New World by Europeans
1793–1798	"The American plague"	~25,000	Yellow fever terrorized colonial America
1832	Second cholera pandemic (Paris)	18,402	Spread from India to Europe/Western Hemisphere
1918	"Spanish" influenza	~50 million	Led to additional pandemics in 1957, 1968, 2009
1976–2020	Ebola	More than 15,000 deaths	First recognized in 1976; 29 regional epidemics to 2020
1981	Acute hemorrhagic conjunctivitis	Rare deaths	First recognized in 1969; pandemic in 1981
1981	HIV/AIDS	> 40 million	First recognized in 1981; ongoing pandemic
2002	SARS	774	Near-pandemic
2009	H1N1 "swine flu"	284,000	Fifth influenza pandemic in less than 100 years
2014	Chikungunya	Uncommon but high morbidity	Pandemic, mosquito-borne
2015	Zika	~1000?*	Pandemic, mosquito-borne

*Zika mortality has not been fully established. Most deaths are fetal or related to outcomes of severe congenital infections.

Source: Reproduced with permission from DM Morens, AS Fauci: Emerging pandemic diseases: How we got to COVID-19. Cell 182:1077, 2020.

Both the Justinian plague (544 AD) and the Black Death pandemics of bubonic/pneumonic plague of 1347–1349 AD **(Chap. 176)** depopulated large segments of Europe and surrounding regions. The cholera **(Chap. 173)** pandemic of 1831–1832 killed large numbers of Europeans and ushered in the first modern studies to characterize disease and death in modern epidemiologic terms.

The past century has featured three of the most highly fatal pandemics the world has ever experienced: the 1918 H1N1 influenza **(Chap. 206)** pandemic, thought to have been the deadliest pandemic in human history; the human immunodeficiency virus (HIV)/acquired immune deficiency syndrome (AIDS) pandemic **(Chap. 208)**, which so far has killed more than 40 million people; and the COVID-19 pandemic **(Chap. 205)**, which is now (2024) in its fifth year, having killed more than 7 million people so far, and by some estimates as many as 20 million. In addition to the highly fatal and the potentially fatal pandemics of the past century, recent decades featured a seeming Pandora's box of EIDS **(Fig. 486-1)**, including novel pathogenic agents, re-emerging agents, and agents that are known but have re-emerged in entirely new forms, e.g., dengue hemorrhagic fever **(Chap. 215)**, Zika, and numerous antibiotic-resistant bacteria **(Chap. 150)**. It appears that we have entered a new era in which emergences and re-emergences of IDs are increasing in frequency and impact.

RESPONSE OF THE MEDICAL COMMUNITY TO INFECTIOUS DISEASES

Acceptance of the concept of infection—almost complete before 1890—quickly led to treatments such as antitoxins and immune plasmas, soon thereafter to non-vaccinia immunogens, by the 1930s to powerful

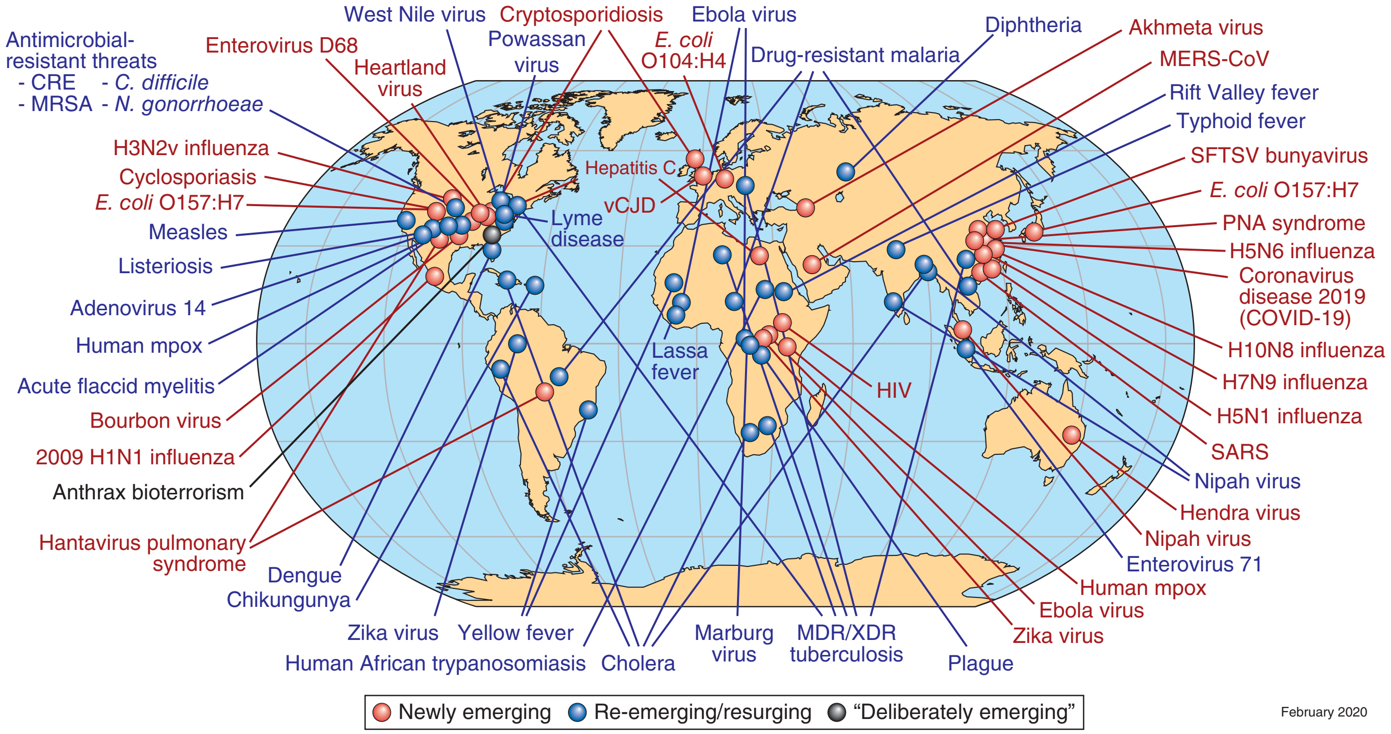
Antimicrobial-resistant threats
- CRE - *C. difficile*
- MRSA - *N. gonorrhoeae*
H3N2v influenza
Cyclosporiasis
E. coli O157:H7
Measles
Listeriosis
Adenovirus 14
Human mpox
Acute flaccid myelitis
Bourbon virus
2009 H1N1 influenza
Anthrax bioterrorism
Hantavirus pulmonary syndrome
Dengue
Chikungunya
Enterovirus D68
Heartland virus
West Nile virus
Powassan virus
Cryptosporidiosis
E. coli O104:H4
Ebola virus
Drug-resistant malaria
Diphtheria
Hepatitis C
vCJD
Lyme disease
Lassa fever
HIV
Akhmeta virus
MERS-CoV
Rift Valley fever
Typhoid fever
SFTSV bunyavirus
E. coli O157:H7
PNA syndrome
H5N6 influenza
Coronavirus disease 2019 (COVID-19)
H10N8 influenza
H7N9 influenza
H5N1 influenza
SARS
Nipah virus
Hendra virus
Nipah virus
Enterovirus 71
Human mpox
Ebola virus
Zika virus
Zika virus
Yellow fever
Human African trypanosomiasis
Cholera
Marburg virus
MDR/XDR tuberculosis
Plague
● Newly emerging
● Re-emerging/resurging
● "Deliberately emerging"
February 2020

FIGURE 486-1 Selected newly emerging, re-emerging, and human-caused emergences over recent decades. *(Reproduced with permission from DM Morens, AS Fauci: Emerging pandemic diseases: How we got to COVID-19. Cell 182:1077, 2020.)*

antibiotics, and, by the 1950s, to antivirals. Ironically, however, these successes, which many considered miraculous, also led to significant overconfidence. By the 1960s, experts were predicting that infectious diseases would be conquered. In 1981 the U.S. Centers for Disease Control and Prevention was reorganized to pivot away from IDs and toward chronic and lifestyle-associated diseases that caused the greatest U.S. mortality and years of productive life lost (YPLL). Then, before the year was out, the world was shocked by the appearance of the AIDS pandemic, caused by a previously unknown body fluid–transmitted virus: HIV.

Although the AIDS pandemic was an enormous challenge, the biomedical research community, together with the pharmaceutical industry, eventually responded, most importantly by developing effective combination antiretroviral therapies, as well as risk-reduction programs such as needle exchanges and education in safe sexual practices.

A 1992 report of the Institute of Medicine (IOM; now the National Academy of Medicine) drew attention to the enormous problem of emerging infections, coined the term and characterized the scope of EIDs, defined the variables associated with emergences, and made far-reaching recommendations for preparedness, response, research, training, and medical and public health practice. Today, clinicians, scientists, public health officials, and government leaders work together within a global infrastructure of EID awareness, in which preparedness and response capacity have been increasingly successful. Such successes include stopping SARS from becoming globally pandemic and then eradicating its virus as a human pathogen; turning AIDS from an inevitably fatal disease to one associated with a normal lifespan for most patients who take antiviral medication; establishing the United States President's Emergency Plan for AIDS Relief (PEPFAR) program to bring life-saving AIDS treatment to the poorest corners of the world; and success in controlling the 2014–2016 regional Ebola **(Chap. 216)** pandemic in West Africa, without a vaccine or proven treatment, using standard public health measures alone.

This modern global EID control infrastructure includes not only vigorous detection of and response to EIDs, including international health regulations and EID oversight by the World Health Organization (WHO), but also controlling and even eradicating existing diseases. Apart from SARS, mentioned above, in 1980 smallpox **(Chap. 201)** was declared eradicated, a feat considered by some to be the most significant accomplishment in medical history, given the millions killed by the virus over many thousands of years. In 2011, the veterinary disease rinderpest was declared eradicated as well. A number of other important IDs now appear to be either close to eradication—e.g., poliomyelitis **(Chap. 210)**, dracunculiasis **(Chap. 240)**—or significantly controlled globally with eradication at least on the horizon—e.g., measles **(Chap. 211)**, rubella **(Chap. 212)**, yaws **(Chap. 183)**. All aspects of ID control, including controlling and trying to eradicate old emerging and still-re-emerging IDs, as well as preventing and controlling the emergences of new ones, work toward the same goal of reducing the impact of IDs on global human morbidity and mortality.

MECHANISMS OF EMERGENCE

Pathogens that newly emerge into humans are acquired via several different mechanisms. Some emerging pathogens are dead-end infections, i.e., they are not usually transmitted onward to other humans, and thus are not likely to become epidemic. Common examples of such infections are those arising from environmental "point-source exposures," in which many humans are exposed to a pathogen in one place over a very narrow window of time, e.g., coccidioidomycosis **(Chap. 219)** and histoplasmosis **(Chap. 218)** outbreaks associated with excavations, norovirus **(Chap. 209)** outbreaks aboard cruise ships (due to sewage-contaminated water), or bacterial/bacterial toxin contamination of foods in restaurants or at picnic or banquet events. Such point-source emergences tend to be sporadic, unpredictable, and of very short duration, the causative organisms being from the environment and usually well known and easily diagnosed, e.g., norovirus outbreaks identified by epidemiologic and clinical characteristics, or easily tested for, e.g., bacterial culture of enteric pathogens. They are typically short outbreaks of high morbidity but low mortality.

A second category of disease emergence, animal-to-human host switching, is of relevance to both newly emerging and re-emerging diseases and accounts for virtually all novel pandemics (e.g., the influenza pandemic of 1918, HIV/AIDS) and many re-emerging IDs as well (e.g., human mpox, human Rift Valley fever). In the past two decades there has been much theoretical and microbiologic and virologic research on how such zoonotic emergences occur. Current concepts are briefly summarized. Most animal and human pathogens, and especially viruses infecting mammals, are specifically adapted to a narrow host group, such as a single host species—e.g., the many New World hantaviruses, which tend to have been finely adapted to single rodent species over thousands of years. Such pathogens may have a limited ability to infect closely related species, and they are not normally highly transmissible between members of new host species they do infect. Measles, for example, is a human-adapted virus which can infect some primate species, but despite its extraordinary contagiousness for humans, it is not naturally transmitted between primates. How then do pathogen emergences into new host species, and most importantly into humans, occur?

This is an extraordinarily difficult question to answer, since emergences tend to occur in a "black box," out of sight of scientists and epidemiologists, and indeed typically in remote locales. But it is theorized that such emergences result from uncommon constellations of otherwise low-risk molecular genomic and ecologic events **(Chaps. 125 and 126)**, which include pathogens (particularly viruses) with high mutation rates, such as many RNA viruses; intense or unusual human–animal contact; and chance. A theoretical model **(Fig. 486-2)** posits that a virus well adapted to its primary host is likely to be non-adapted or at least far less adapted to other potential hosts, even closely related ones. But since viruses have high mutation rates, there will always be some virions that have mutated to become less well adapted to their own hosts and, by chance, some of these may "accidentally" have acquired an ability to adapt to a new host. Presumably, very few of these mutated viruses come into contact with potential new hosts before they are "purged," but if one or more virions is (1) is not significantly de-adapted to its normal host, (2) is, at the same time, capable of adapting to a new host, and (3) is able to cross a "fitness valley" of de-adaptation to the old host and neo-adaptation to the new host, a host switch may occur (Fig. 486-2). This model attempts to explain why major emergences are relatively uncommon (e.g., despite countless billions of viral mutations occurring daily, looking back over the past 500 years, there has only been about one new recognized influenza pandemic every 30 or so years). It might be said, teleologically, that pathogens are constantly "trying" to emerge, but almost always failing to do so.

A third category of disease emergence relates to mutations that occur in pathogens that are already human adapted, the best-known example being development of antibiotic resistance in a bacterial species previously susceptible to particular antibiotics. Although antibiotic-resistant pathogens live in the soil and other natural environments and have done so since before human-developed antibiotics existed, bacteria are also capable of horizontally transmitting to other bacteria DNA-containing resistance genes, and these may be selected for in Darwinian fashion, e.g., by medical and hospital environments in which antibiotics are administered **(Chap. 147)**. An analogous principle even applies to viruses. Particularly as viruses gain access to larger human populations via crowding and human movement, mutation may lead to a fitness advantage, e.g., to a more transmissible phenotype that may escape natural and vaccine-induced immunity, and may even be associated with increased pathogenicity, as may have happened with the SARS-CoV-2 delta variant in 2022. A similar phenomenon was seen beginning in 2014 with the re-emergence of enterovirus D68 (EV-D68) to cause global epidemics of acute flaccid myelitis.

VARIABLES ASSOCIATED WITH DISEASE EMERGENCE

Whatever the mechanisms of pathogen emergence may be, it has become clear that the determinants of emergence—i.e., variables that provide opportunities for pathogens in disturbed ecosystems to emerge, typically beginning with the host switch of an animal pathogen

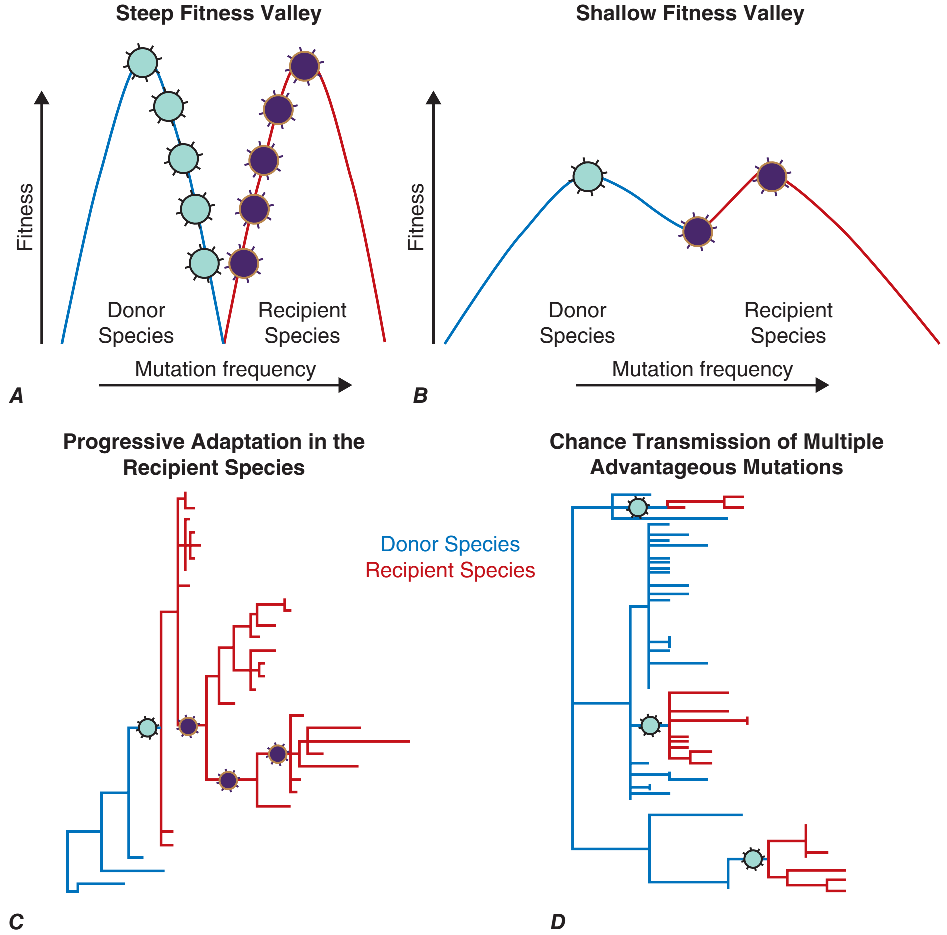

FIGURE 486-2 A theoretical model of how animal-to-human pathogen host-switching might occur, in this case the pathogen being a virus. The light-colored viruses represent those adapted to the transmitting host, and the dark-colored viruses represent those mutating in the direction of the host to which the virus is adapting. ***A*** and ***B*** compare two different situations in which there is a deep virus–host fitness valley (***A***) or a shallow fitness valley (***B***), the valleys representing the degree of challenges that mutable viruses need to overcome to be able to infect new host species. To be able to cross the steep fitness valley, a virus that is at peak adaptation to host 1 must mutate significantly in the direction of de-adaptation to be able to infect host 2, an event that is more likely if the virus has a high natural rate of mutations. In ***B***, a narrow fitness valley is more easily crossed, resulting in a host switch. The phylogenetic trees in ***C*** and ***D*** show the adaptational mutations that necessarily occur for a virus that has crossed a steep fitness valley (shown in ***A***) as it adapts to the new host (***C***). In ***D***, a virus that has crossed a narrow fitness valley (as shown in ***B***) does not "need to" adapt to the new host as significantly (***C***) to be able to initiate and sustain transmission between hosts of the new species. *(Reproduced with permission from DM Morens, AS Fauci: Emerging pandemic diseases: How we got to COVID-19. Cell 182:1077, 2020.)*

within a geographically identified global "hot spot"—are largely related to humans and human activities **(Fig. 486-3)**. Most of the important variables associated with pathogen emergence are either activities of the human host—demographics and behavior including crowding, human movement, sexual practices, and occupation (Fig. 486-3, upper right)—or variables that reflect human degradation of the environment—poverty and social ills, wars, displacements, land use practices, and inadequate public health infrastructure (Fig. 486-3, lower right). Even though HIV clades probably emerged separately and independently more than a century ago, AIDS did not become pandemic until significant viral transmission could be sustained within a modern human-dominated global environment, once composed of remote villages but more recently replaced by urban environments, transnational travel and commercial sex work, meeting places for men who have sex with men, IV drug use, and blood product transfusions. The emergence of hyperendemic dengue and dengue hemorrhagic fever after World War II was associated with urban crowding and domestic water storage. The United States epidemic of hepatitis C **(Chaps. 350 and 352)**, which began in the 1960s, was associated with blood product transfusions and injection drug use. The emergence of hantavirus pulmonary syndrome was associated with construction of human-made peridomestic outbuildings and unfinished basements that housed infected reservoir rodents during the winter. The emergence of Nipah virus **(Chap. 204)** in Malaysia in 1998 was associated with both deforestation and intensive pig farming practices. The 2003 United States mpox outbreak was associated with unregulated importation of rodent pets from enzootic areas. Emergence in China of both H5N1 and H7N9 poultry influenza A ("bird flu") was associated with crowded live animal markets, and it stopped quickly when these markets were shut down. The 2010 cholera epidemic in Haiti followed a devastating earthquake associated with human displacement, loss of access to safe water, inadequate medical and social support, and the arrival of aid from foreign locales that inadvertently imported cholera organisms.

It should be noted that emergence variables are not always identical between the newly emerging and re-emerging IDs. Re-emergences of known pathogens are more likely to result from societal failures to create and sustain safe human environments, and to disruption of balanced ecosystems in the natural world. Emergences of new pathogens such as pandemic influenza, on the other hand, are sometimes unique and highly improbable events. It is noteworthy that for diseases like influenza, in which transmission is facilitated by crowding and human movement, the intervals between pandemics are about the same today as they were 500 years ago, despite an eightfold increase in the global population and the advent of modern rapid global travel. This suggests that influenza pandemics are rare stochastic events unrelated to the size of the population that spreads them, even though population growth spreads them more widely and more quickly than they once did.

EXAMPLES OF CHALLENGES OF DISEASE EMERGENCE AND HOW THEY ARE BEING MET

The complexities of potential control of pathogen emergence and re-mergence are noteworthy: each emergent disease presents a different challenge, as briefly summarized below for three selected pathogens of global importance.

■ INFLUENZA A

The reservoir of influenza A viruses is the global pool of wild waterfowl and shorebirds. All human, other mammalian, and poultry-associated influenza viruses are derived from this pool. The 1918 pandemic influenza H1N1 "founder" virus either was derived directly from a waterfowl or had a brief period in another mammalian host before becoming pandemic. Humans immediately transmitted it to domestic pigs in 1918; descendants of those human and pig viruses still exist in whole or in part, having devolved into separate and increasingly divergent lineages over the past century. The three influenza pandemics that have occurred since 1918 were all caused by genetic descendants of the 1918 virus, having been naturally modified by genetic changes of several types. Such mutations make control and prevention difficult. The viruses have multiple mutational mechanisms to circumvent human immunity, including (1) genetic "drift" (point mutations); (2) "shift" (importation of different avian hemagglutinins with or without importation of neuraminidases); (3) intrasubtype reassortments (importation

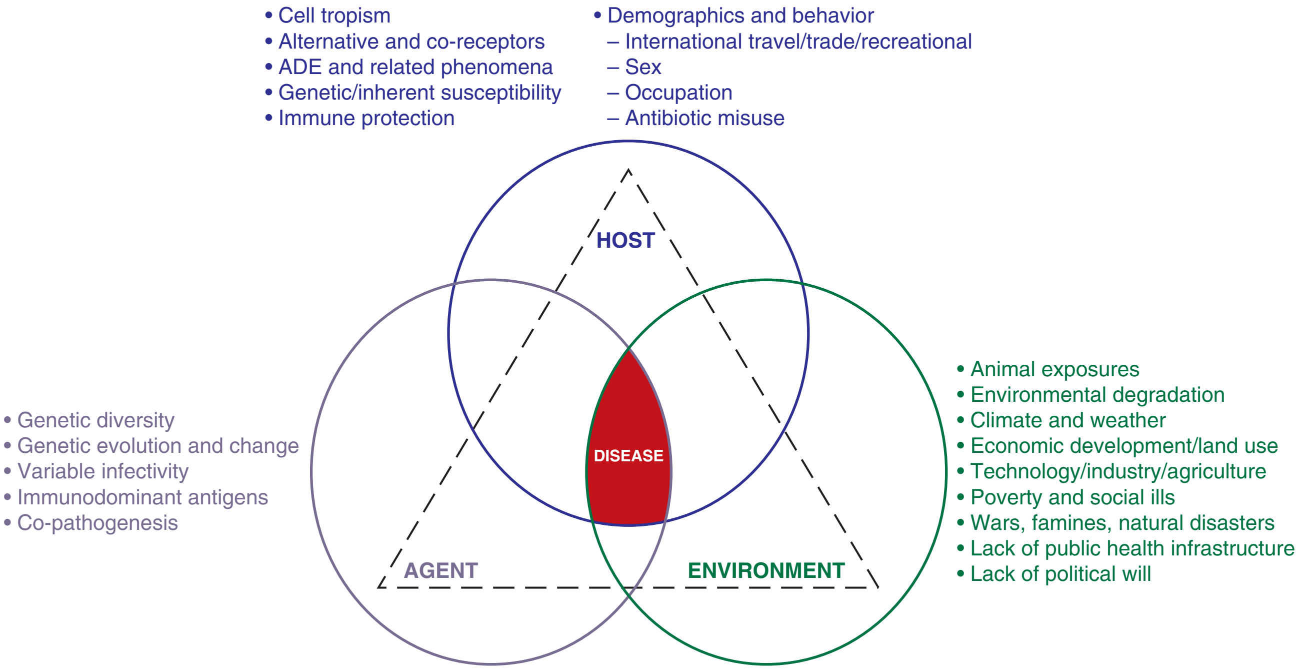
• Cell tropism
• Alternative and co-receptors
• ADE and related phenomena
• Genetic/inherent susceptibility
• Immune protection
• Demographics and behavior
– International travel/trade/recreational
– Sex
– Occupation
– Antibiotic misuse
HOST
• Genetic diversity
• Genetic evolution and change
• Variable infectivity
• Immunodominant antigens
• Co-pathogenesis
DISEASE
• Animal exposures
• Environmental degradation
• Climate and weather
• Economic development/land use
• Technology/industry/agriculture
• Poverty and social ills
• Wars, famines, natural disasters
• Lack of public health infrastructure
• Lack of political will
AGENT
ENVIRONMENT

FIGURE 486-3 Infectious agents, hosts, and the environment: determinants of disease emergence, re-emergence and persistence. Factors most closely associated with re-emergences are highlighted, including accidental and purposeful human release of infectious pathogens. *(Reproduced with permission from DM Morens, AS Fauci: Emerging pandemic diseases: How we got to COVID-19. Cell 182:1077, 2020.)*

of whole human hemagglutinin or neuraminidase subtype variants), and (4) glycosylation of external protein sites that can alter viral structure and function. With a broad repertoire of mutational possibilities, a human-adapted virus may repeatedly escape population immunity elicited by natural infection and vaccination, as has been the case with the 1968 pandemic H3N2 virus, which is still causing infections and death, and escaping vaccine immunity, 56 years after its appearance. During this interval it has also periodically, and for unknown reasons, increased in pathogenicity.

Influenza control is further confounded by the fact that influenza A is a nonsystemic virus that replicates on a one-cell-deep layer of respiratory epithelium, without fully encountering the human systemic immune system. Because influenza has a very short incubation period, there is too little time for the virus to fully elicit memory immune cells, the virus only encountering the less specific innate immune system, as well as IgA and IgM antibodies secreted across the epithelium from the systemic circulation (only fully effective in high concentration). It also has an advantage in replicating only in an immune environment of the upper respiratory tract that tolerates multiple continuous antigenic exposures, including brief viral infections, without massive immune responses that could be harmful to the host. Thus, influenza vaccines are at best incompletely effective and protect for only a matter of months. This will be an important challenge in attempts to develop universal influenza vaccines.

It is also noteworthy that pandemic and postpandemic influenza viruses are increasingly better adapted to the modern world. Before 1889, the predominant travel mechanism of influenza spread was by coach, globalization of spread taking as much as a year or more, or failing to occur at all. Until the 1700s, the Americas and Europe did not even appear to have been on the same influenza pandemic cycles. (Although the earliest genetically sequenced influenza viruses date to 1918, scholars have for several centuries presumptively identified influenza pandemics on the basis of characteristic clinical-pathologic signs and symptoms, epidemiologic patterns, and route and rapidity of global spread). In the 1889 pandemic, presumed influenza was spread from East Asia to Europe by rail. The 1918 and 1957 pandemics were spread predominantly by ship, and the 1968 and 2009 pandemics predominantly by air. Influenza spread is ideally suited to human movement and crowding. Furthermore, high rates of presymptomatic and subsymptomatic infection allow the virus to be transmitted by persons who do not know they are infectious or capable of transmitting, thwarting attempts to lower transmission by public health measures.

In short, influenza A is an avian virus emerging out of nature. It remains difficult to prevent or to identify in advance an avian influenza virus with pandemic potential. At this point we have limited ability to control constantly mutating viruses once they have become human-adapted.

■ SARS AND SARS-CoV2

The sarbecoviruses (SARS-like beta-coronaviruses) are similar to influenza in some respects and different in others. Their natural reservoir is not birds but bats, and the main human host receptor for viral infection is the angiotensin-converting enzyme 2 (ACE2) receptor, as opposed to respiratory tract sialic acid receptors for influenza. Because ACE2 receptors on bats, many other mammals, and humans are quite similar, sarbecoviruses can be thought of as potentially preadapted to humans. It is in part for this reason that many experts predict future emergences of these viruses.

Similar to influenza, SARS-CoV2 is a respiratory virus that does not produce true viremia, has a short incubation period and incomplete exposure to the systemic immune system, and often causes asymptomatic or subsymptomatic infection, bedeviling control with public health measures such as social distancing and isolation. Also similar to influenza, and despite its moderate transmissibility, it can be "superspread" in crowded environments such as restaurants and bars, churches, sports venues, or any crowded place, especially if there is imperfect airflow and humidity. In addition, like influenza, SARS-CoV2 evolves continually (not however by genetic reassortment, but by mutation and recombination), continually escapes population immunity, and is imperfectly prevented by vaccines. As has been the case with influenza, control of emerging SARS-CoV2, and prevention of population spread and large-scale mortality, remains difficult.

■ DENGUE

A third example of the complexities of emergence and response, in this case re-emergence, is dengue. An *Aedes aegypti*–borne flavivirus, "dengue virus" is actually four closely related viruses (DEN-1, DEN-2, DEN-3, and DEN-4) that interact with each other serologically and with respect to immunity and pathogenicity. Dengue is believed to have emerged more than 400 years ago, perhaps much earlier, and has been endemic and hyperendemic around the tropical belt ever since, re-emerging sporadically to cause high morbidity and sometimes high mortality. Emergences now occur repeatedly in the U.S. commonwealths of the Northern Mariana Islands and Puerto Rico, in the U.S. territories of Guam, American Samoa, and the United States Virgin Islands, and in the United States–affiliated Federated States of Micronesia. At one time, large-scale dengue epidemics were common in the continental United States, e.g., a major epidemic in Philadelphia in 1780; in recent years smaller outbreaks have occurred in southern states. Despite the fact that all dengue cases occur within a human-to-mosquito-to-human transmission cycle, without direct human-to-human transmission, dengue re-emergences can be so explosive that they mimic the most highly contagious respiratory viruses. During simultaneous 1977 outbreaks of dengue and influenza in Puerto Rico, within-household studies showed greater case-clustering of dengue than of influenza.

Dengue is highly associated with peridomestic water storage (where vector mosquitoes oviposit), lack of sanitation, crowding, and lack of screens and air conditioning. It would seem a simple matter to control dengue by controlling water storage, but even Singapore—a developed country whose residences have screens, air conditioning, excellent sanitation, and public health mosquito police who constantly patrol residences and public places to examine for breeding sites, assessing fines for even minor breeding site violations—is unable to prevent dengue outbreaks from occurring. A difficult problem seems to be that mosquitoes can oviposit in the tiniest and most inconspicuous bits of water, e.g., the slight (several millimeter) indentations in the bases of porcelain toilets where they are bolted to the floor.

Although dengue was historically considered a nuisance disease, shortly after World War II (during which only 4 of more than 90,000 dengue-infected servicemen/service women died) a new and highly fatal clinical form emerged in Southeast Asia and quickly became a significant killer of children. Studied by teams of Thai and American scientists, the new clinical form was classified as dengue hemorrhagic fever (DHF) and, in its most severe form, dengue shock syndrome (DSS), a result of sudden massive extravasation of intravascular fluids into the tissues. But the very same viruses that caused the mild form of dengue were also those that caused DHF and DSS, a puzzle that was unraveled by Halstead and others when it was learned that it was predominantly only with the second of four potential dengue infections (one with each of the four serotypes) that DHF/DSS occurred. Epidemiologic data were consistent with the belief that cross-reactive antibodies elicited by the first dengue infection, unable to neutralize the second virus, actually potentiate viral entry into cells via antibody-dependent enhancement for the acute clinician, dealing with EIDS depends upon ecologic perspectives on infectious disease occurrence.

Only in recent years have science and medicine begun to make an impact on severe dengue disease, including training of clinicians and parents, the widespread availability of oral rehydration solutions that reverse shock syndrome, and several recently developed vaccines. But fear remains that vaccines might potentiate more severe disease, and dengue remains among the greatest re-emerging disease problems for more than 3 billion people.

The three emerging disease examples noted above exemplify the complexity of disease emergence: the role of demographic and environmental variables (live poultry markets in severe human influenza; human-made peridomestic environments in dengue), pathogen interactions (partial immune crosses in influenza and dengue, related to both protection and disease severity, as well as to diagnostic difficulties), complete or partial immune escape (all three pathogens), immunopathogenesis (1918 influenza and dengue), inherent problems in vaccine prevention (all three diseases), animal reservoir hosts of progenitor viruses (all three diseases), viral evolution via mutations (all three diseases), and many other factors.

THE ROLE OF CLINICIANS AND LABORATORIANS IN EID CONTROL

Clinicians and laboratorians are on the front lines of EID control and response efforts, and they must continue to play a leading role in global pandemic preparedness and response. It is typically the astute clinician who first recognizes a new disease, usually because signs and symptoms, or complications, are not typical of similar diseases. For example, in the 1918 influenza pandemic there was never-before-seen pneumonia-associated case-fatality featuring viral infection and bacterial co-pathogenesis, the bacteria being normal oral flora and the pneumonia producing an anatomic pattern, diagnosed clinically or in some cases by x-ray, which was invariably bronchopneumonic. In the 1950s, emergence of dengue hemorrhagic fever was associated with never-before-seen complications, including shock and death, occurring in association with endemic dengue viruses and in epidemiologic patterns that were novel (shock occurring only in infants under one, and in older toddlers and school-aged children, but sparing children between these ages, as well as those in and beyond the older teenage years). In fact, it was the bizarre and unprecedented age-specific mortality patterns of DSS that suggested its pathogenetic mechanisms. That legionellosis **(Chap. 164)**, first identified in 1976, must have been an unknown emerging disease was learned when clinicians were unable to make a pathogen-specific pneumonia diagnosis and laboratories were unable to isolate a pathogen. A similar profile was seen with SARS, emerging in 2002–2003, when adults with respiratory/constitutional symptoms typical of many respiratory viruses experienced an approximate 10% case-fatality rate.

It is not only primary care providers who are in a position to recognize newly emerging diseases. Both emerging acute hemorrhagic conjunctivitis (AHC) in 1969–1970 and emerging human Rift Valley fever (RVF) in 1977 were largely identified by ophthalmologists who observed the unexpected: severe and explosively epidemic conjunctivitis with AHC, and blindness associated with macular infiltrates in RVF. In 1981, it was pathologists who identified *Pneumocystis carinii* (now *jirovecii*) pneumonia in autopsies of previously healthy young men, joining epidemiologists and clinicians in characterizing a new clinical-pathologic syndrome that was eventually recognized as AIDS. And it was pediatric neurologists who played a major role in identifying emergent enterovirus D68–associated acute flaccid myelitis beginning in about 2014. Even looking back in history to a time before the microbial era, we can identify the same phenomenon: it was a "proto-epidemiologist" who first imagined that cervical cancer was the result of a contagious disease when he showed, in the 1840s, that nuns never suffered from it.

Among the skills of medical practitioners related to a key front-line role in dealing with EIDs are the following: (1) abiding curiosity; (2) the ability to sense something out of the ordinary and pursue suspicions, using the medical history, physical examination, and laboratory tests **(Chaps. 8 and 505)**; (3) a sense of the epidemiology of different diseases that allows imagination of how an illness might have been acquired, of the mechanism of acquisition—environmental, respiratory, inoculation (including sexual inoculation), gastrointestinal, occupational, etc. (Fig. 486-3); (4) whether there are family or other case clusters, and whether such clusters represent common-source exposure, zoonotic exposure, or person-to-person transmission; (5) determination of an approximate incubation period and duration of infectivity; (6) a close relationship with the diagnostic laboratory and knowledge of when to seek additional laboratory follow-up; (7) a close relationship with allied practitioners, including specialists, and with disease researchers; (8) a close relationship with public health workers, diligence in promptly reporting reportable diseases and reportable disease syndromes, and awareness of outbreaks that may be occurring in the community; (9) regular reading of the medical literature to stay grounded within and outside of one's specialty; and (10) skill at communicating with and eliciting trust from patients, which appears to maximize the chance of finding clues that may be easily missed.

History tells us that EIDs will certainly continue to appear, including some so novel that they cannot even be imagined. An important component of the biomedical and public health enterprise is the ability to recognize, diagnose, and begin to control EIDs at the earliest possible time, and to limit their spread so that emergences do not evolve into epidemics. This requires coordinated efforts of those in multiple disciplines, communicating with each other rapidly and effectively. It is a part of the significant responsibilities of medical, biomedical, and public health practitioners, who together have the critical skill sets needed to confront the existential threats of EIDs.

FURTHER READING

Fauci AS, Morens DM: The perpetual challenge of infectious diseases. N Engl J Med 366:454, 2012.

Harper K: *Plagues Upon the Earth: Disease and the Course of Human History*. Princeton, Princeton University Press, 2023.

Keusch GT et al: Pandemic origins and a One Health approach to preparedness and prevention: Solutions based on SARS-CoV-2 and other RNA viruses. Proc Natl Acad Sci USA 119:e2202871119, 2022.

Lederberg J et al: *Emerging Infections. Microbial Threats to Health in the United States*. Washington, DC, National Academies Press, 1992.

Menachery VD et al: SARS-like WIV1-CoV poised for human emergence. Proc Natl Acad Sci USA 113: 3048, 2016.

Morens DM, Fauci AS: Emerging pandemic diseases: How we got to COVID-19. Cell 182:1077, 2020.

Morens DM, Taubenberger JK: Pandemic influenza: Certain uncertainties. Rev Med Virol 21:262, 2011.

Parrish C et al: Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Molec Biol Rev 72:457, 2008.

Wasik BR et al: Onward transmission of viruses: How do viruses emerge to cause epidemics after spillover? Philos Trans R Soc Lond B Biol Sci 374:20190017, 2019.

Wegner GI et al: Averting wildlife-borne infectious disease epidemics requires a focus on socio-ecological drivers and a redesign of the global food system. EClinicalMedicine 47:101386, 2022.

487

Primary Care and Global Health

Tim Evans, Kumanan Rasanathan

The twentieth century witnessed the rise of an unprecedented global health divide. Industrialized or high-income countries experienced rapid improvement in standards of living, nutrition, health, and health care **(Chap. 485)**. Meanwhile, in low- and middle-income countries with much less favorable conditions, health and health care progressed much more slowly. The scale of this divide is reflected in the current extremes of life expectancy at birth, with Japan at the high end (84 years) and Chad at the low end (54 years). This 30-year shortfall in Chad reflects the daunting range of health challenges faced by low- and middle-income countries. These nations must deal not only with a complex mixture of diseases (both infectious and chronic) and illness-promoting conditions but also, and more fundamentally, with the fragility of the foundations underlying good health (e.g., sufficient food, water, sanitation, and education) and of the systems necessary for universal access to good-quality health care and public health. In the last decades of the twentieth century, the need to bridge this global health divide and establish health equity was increasingly recognized. The Declaration of Alma-Ata in 1978 crystallized a vision of justice in health, regardless of income, gender, ethnicity, or education, and called for "health for all by the year 2000" through primary health care. While progress since the declaration is remarkable, >45 years later and in the aftermath of a global pandemic of COVID-19, much remains to be done to achieve global health equity.

This chapter looks first at the nature of the health challenges that underlie the health divide in low- and middle-income countries. It then outlines the values and principles of a primary health care approach, with a focus on primary care services. Next, the chapter reviews the experience of low- and middle-income countries in addressing health challenges through primary care and a primary health care approach. Finally, the chapter identifies how current challenges and global context, in particular, the global pandemic, shape an agenda for the renewal of primary health care and primary care, allied to the movement to achieve universal health coverage.

PRIMARY CARE AND PRIMARY HEALTH CARE

The term primary care has been used in many different ways: to describe a level of care or the setting of the health system, a set of treatment and prevention activities carried out by specific personnel, a set of attributes for the way care is delivered, or an approach to organizing health systems that is synonymous with the term primary health care. In 1996, the U.S. Institute of Medicine encompassed many of these different usages, defining primary care as "the provision of integrated, accessible health care services by clinicians who are accountable for addressing a large majority of personal health care needs, developing a sustained partnership with patients, and practicing in the context of family and community."[1] We use this definition of primary care in this chapter. Primary care performs an essential function for health systems, providing the first point of contact when people seek health care, dealing with most problems, and referring patients onward to other services when necessary. As is increasingly evident in countries of all income levels, without strong primary care, health systems cannot function properly or address the health challenges of the communities they serve.

Primary care is only one part of a primary health care approach. The Declaration of Alma-Ata, drafted in 1978 at the International Conference on Primary Health Care in Alma-Ata (now Almaty in Kazakhstan), identified many features of primary care as being essential to achieving the goal of "health for all by the year 2000." However, it also identified the need to work across different sectors, address the social and economic factors that determine health, mobilize the participation of communities in health systems, and ensure the use and development of technology that was appropriate in terms of setting and cost. The declaration drew from the experiences of low- and middle-income countries in trying to improve the health of their people following independence. Commonly, these countries had built hospital-based systems similar to those in high-income countries. This effort had resulted in the development of high-technology services in urban areas while leaving the bulk of the population without access to health care unless they traveled great distances to these urban facilities. Furthermore, much of the population lacked access to basic public health measures. Primary health care efforts aimed to move care closer to where people lived, to ensure their involvement in decisions about their own health care, and to address key aspects of the physical and social environment essential to health, such as water, sanitation, and education.

After the Declaration of Alma-Ata, many countries implemented reforms of their health systems based on primary health care. Most progress involved strengthening of primary care services; unexpectedly, however, much of this progress was seen in high-income countries, most of which constructed systems that made primary care available at low or no cost to their entire populations and that delivered the bulk of services in primary care settings. This endeavor also saw the reinforcement of family medicine as a specialty to provide primary care services. Even in the United States (an obvious exception to this trend), it became clear that the populations of states with more primary care physicians and services were healthier than those with fewer such resources.

Progress was also made in many low- and middle-income countries. However, the target of "health for all by the year 2000" was missed by a large margin. The reasons were complex but partly entailed a general failure to implement all aspects of the primary health care approach, particularly work across sectors to address social and economic factors that affect health and provision of sufficient human and other resources in order to make possible the access to primary care attained in high-income countries. Furthermore, despite the consensus in Alma-Ata in 1978, the global health community rapidly became fractured in its commitment to the far-reaching measures called for by the declaration. Economic recession tempered enthusiasm for primary health care, and momentum shifted to programs concentrating on a few priority measures such as immunization, oral rehydration, breastfeeding, and growth monitoring for child survival. Success with these initiatives supported the continued movement of health development efforts away from the comprehensive approach of primary health care and toward programs that targeted specific public health priorities. This approach was reinforced by the need to address the HIV/AIDS epidemic. By the 1990s, primary health care had fallen out of favor in many global-health policy circles, and low- and middle-income countries were being encouraged to reduce public sector spending on health and to focus on cost-effectiveness analysis to provide a package of health care measures thought to offer the greatest health benefits.

---

[1] Institute of Medicine. Primary Care: America's Health in a New Era (1996).

HEALTH CHALLENGES IN LOW- AND MIDDLE-INCOME COUNTRIES

Low- and middle-income countries, defined by a per-capita gross national income of <$12,535 (U.S.) per person per year, account for >85% of the world's population. Average life expectancy in these countries lags far behind that in high-income countries: whereas the average life expectancy at birth for a girl in high-income countries is 83 years, it is only 65 years for a girl in low-income countries. This discrepancy has received growing attention over the past 50 years. Initially, the situation in poor countries was characterized primarily in terms of high fertility and high infant, child, and maternal mortality rates, with most deaths and illnesses attributable to infectious or tropical diseases among remote, largely rural populations. With growing adult (and especially elderly) populations and changing lifestyles linked to global forces of urbanization, a new set of health challenges characterized by chronic diseases, environmental overcrowding, and road traffic injuries has emerged rapidly **(Fig. 487-1)**. The majority of tobacco-related deaths globally now occur in low- and middle-income countries, and the risk of a child's dying from a road traffic injury in Africa is more than twice that in Europe. Thus, low- and middle-income countries in the twenty-first century face a full spectrum of health challenges—infectious, chronic, and injury-related—at much higher incidences and prevalences than are documented in high-income countries and with many fewer resources to address these challenges.

Addressing these challenges, however, does not mean simply waiting for economic growth. Analysis of the association between wealth and health across countries reveals that, for any given level of wealth, there is a substantial variation in life expectancy at birth that has persisted despite overall global progress in life expectancy during the past 40 years **(Fig. 487-2)**. Health status in low- and middle-income countries varies enormously. Nations such as Cuba and Costa Rica have life expectancies and childhood mortality rates similar to or even better than those in high-income countries; in contrast, countries in Sub-Saharan Africa and the former Soviet bloc have at times experienced significant reversals in these health markers, particularly in the 1990s.

As Angus Deaton stated in the World Institute for Development Economics Research annual lecture on September 29, 2006, "People in poor countries are sick not primarily because they are poor but because of other social organizational failures, including health delivery, which are not automatically ameliorated by higher income." This analysis concurs with classic studies of the array of societal factors explaining good health in poor settings such as Cuba and Kerala State in India in the 1980s. Analyses conducted over the past 4 decades indeed show that rapid health improvement is possible in very different contexts. That some countries continue to lag far behind can be understood through a comparison of regional differences in progress in terms of life expectancy over this period **(Fig. 487-3)**.

As average levels of health vary across regions and countries, so too do they vary within countries **(Fig. 487-4)**. Indeed, disparities within countries are often greater than those between high-income

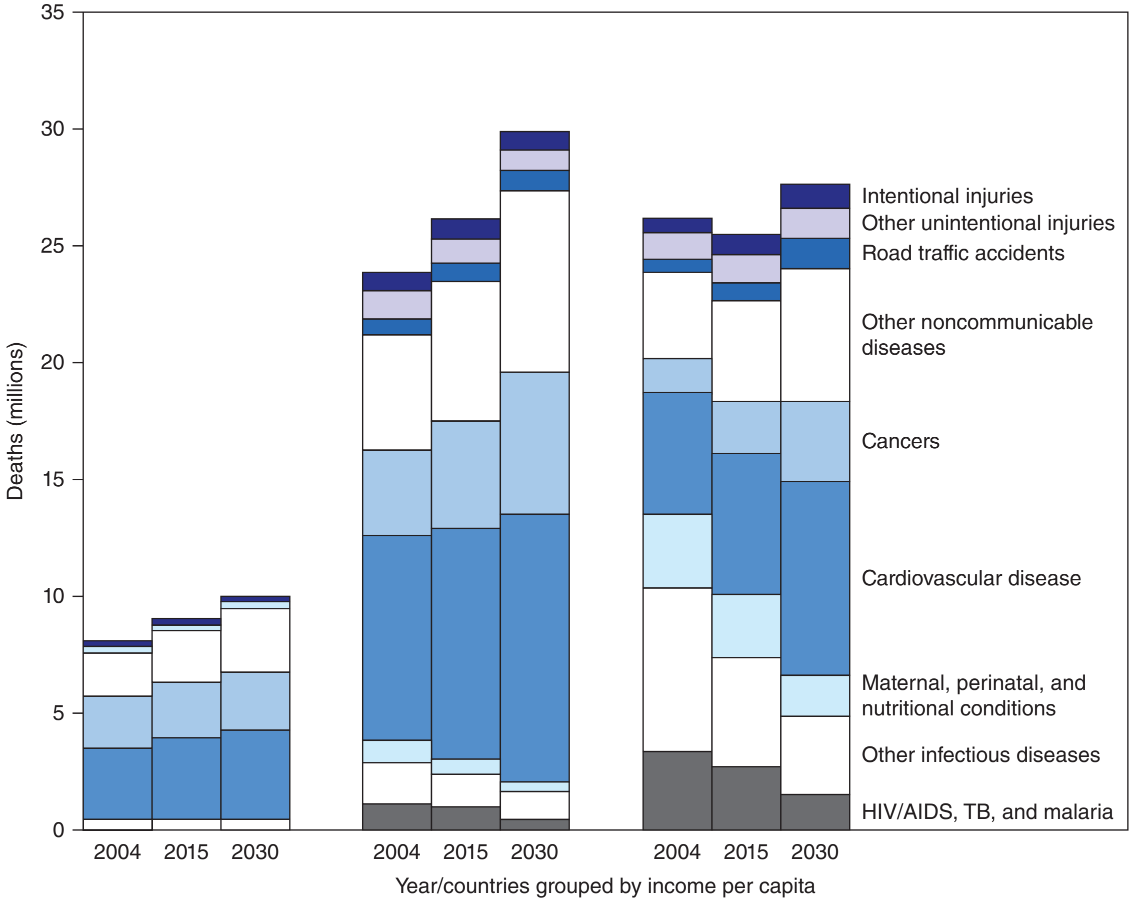

FIGURE 487-1 Projections of disease burden to 2030 for high-, middle-, and low-income countries (*left, center*, and *right*, respectively). TB, tuberculosis. *(Reproduced with permission from World Health Organization: The Global Burden of Disease 2004 Update, 2008.)*

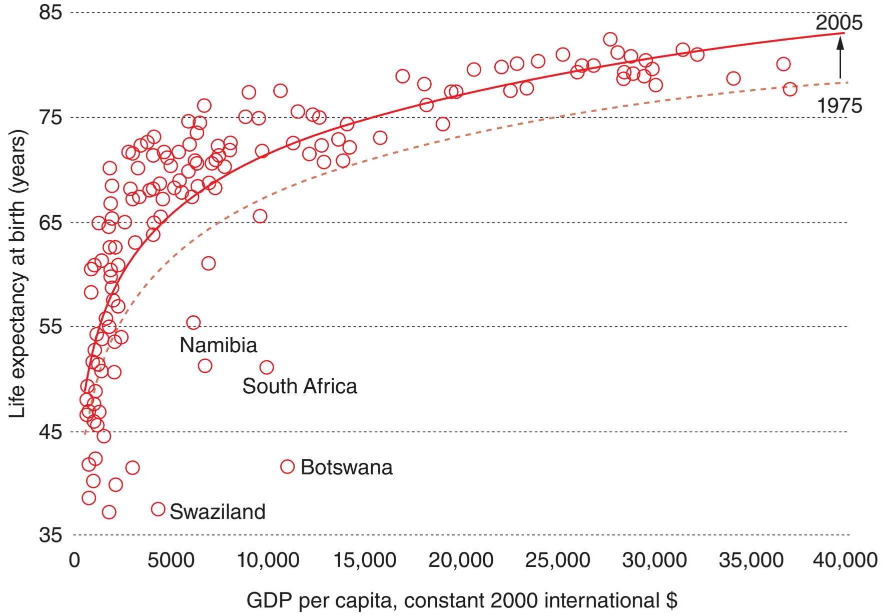

FIGURE 487-2 Gross domestic product (GDP) per capita and life expectancy at birth in 169 countries, 1975 and 2005. Only outlying countries are named. *(Reproduced with permission from World Health Organization: Primary Health Care: Now More Than Ever. World Health Report 2008.)*

and low-income countries. For example, if low- and middle-income countries could reduce their overall childhood mortality rate to that of the richest one-fifth of their populations, global childhood mortality could be decreased by 40%. Disparities in health are mostly a result of social and economic factors such as daily living conditions, access to resources, and ability to participate in life-affecting decisions. In most countries, the health care sector actually tends to exacerbate health inequalities (the "inverse-care law"); because of neglect and discrimination, poor and marginalized communities are much less likely to benefit from public health services than those that are better off. Reforming health systems toward people-centered primary care provides an opportunity to reverse these negative trends.

Health services have failed to make their contribution to reducing these pervasive social inequalities by ensuring universal access to existing, scientifically validated, low-cost interventions such as insecticide-treated bed nets for malaria, taxes on cigarettes, short-course chemotherapy for tuberculosis, antibiotic treatment for pneumonia, dietary modification and secondary prevention measures for high blood pressure and high

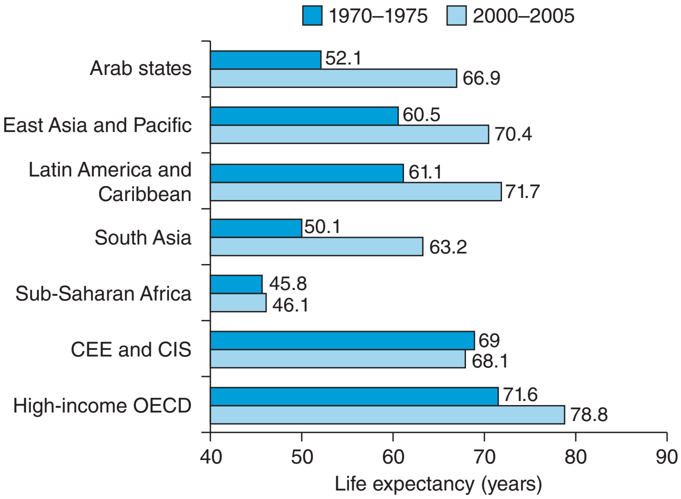

FIGURE 487-3 Regional trends in life expectancy. CEE, Central and Eastern Europe; CIS, the Commonwealth of Independent States; OECD, Organization for Economic Cooperation and Development. *(Reproduced with permission from World Health Organization: Closing the Gap in a Generation: Health Equity Through Action on the Social Determinants of Health. Commission on Social Determinants of Health Final Report, 2008.)*

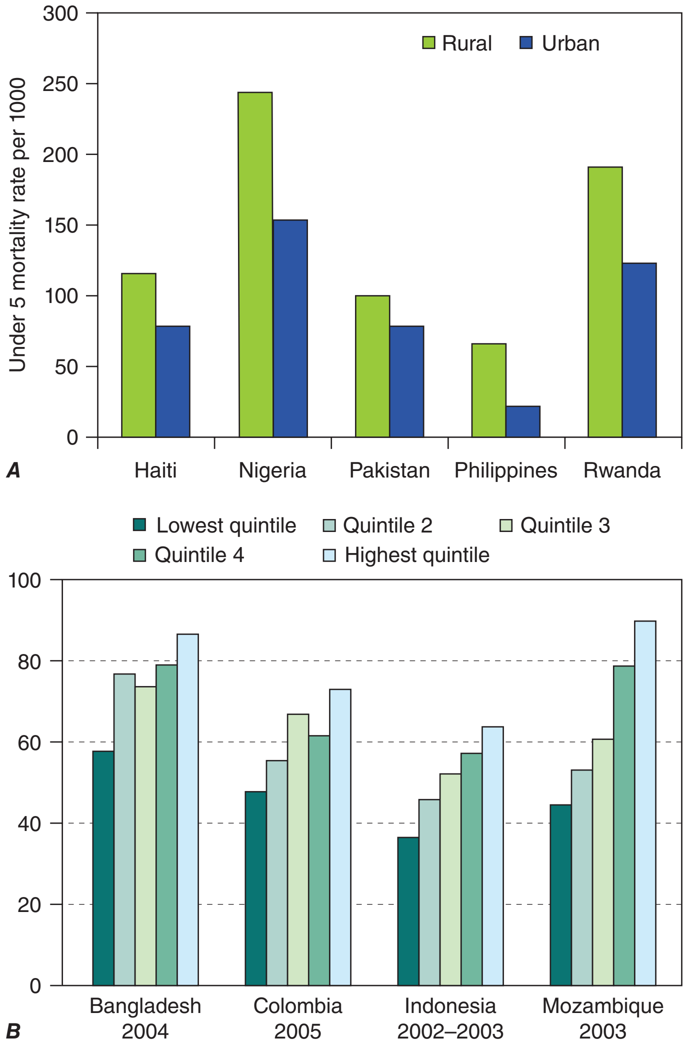

FIGURE 487-4 ***A.*** Mortality of children under 5 years old, by place of residence, in five countries. *(Reproduced with permission from World Health Organization: Data from the World Health Organization.)* ***B.*** Full basic immunization coverage (%), by income group. *(Reproduced with permission from World Health Organization: Primary Health Care: Now More Than Ever. World Health Report 2008.)*

cholesterol levels, and water treatment and oral rehydration therapy for diarrhea. Despite decades of "essential packages" and "basic" health campaigns, the effective implementation of what is already known to work with requisite scale and quality appears (deceptively) to be difficult.

Recent analyses have begun to focus on "the how" (as opposed to "the what") of health care delivery, exploring why health progress is slow and sluggish despite the abundant availability of proven interventions for health conditions in low- and middle-income countries. Three general categories of reasons are being identified: (1) shortfalls in performance of health systems; (2) stratifying social conditions; and (3) skews in science.

SHORTFALLS IN PERFORMANCE OF HEALTH SYSTEMS

Specific health problems often require the development of specific health interventions (e.g., tuberculosis requires short-course chemotherapy). However, the delivery of different interventions is often facilitated by a common set of resources or functions: money or financing, trained health workers, and facilities with reliable supplies fit for multiple purposes. Unfortunately, health systems in most low- and middle-income countries are largely dysfunctional across these core functions.

In the large majority of low- and middle-income countries, the level of public financing for health is woefully insufficient: whereas high-income countries spend, on average, >7% of the gross domestic product on health, middle-income countries spend <3%, and low-income countries <2%. External financing for health through various donor channels grew rapidly in the first decade of the twenty-first century but has grown more slowly in the second decade to its current level of $37 billion. While these funds for health are significant, they represent <2% of total health expenditures in low- and middle-income countries and therefore are neither a sufficient nor a long-term solution to chronic underfinancing. In Africa, 70% of health expenditures come from domestic sources. The predominant form of health care financing—charging patients at the point of service—is the least efficient and the most inequitable, tipping millions of households into poverty annually. Health workers, who represent another critical resource, are often inadequately trained and supported in their work, especially in locations with the greatest needs. Recent estimates indicate a shortage of >18 million health workers, constituting a crisis that is greatly exacerbated by the migration of health workers from low- and middle-income countries to high-income countries. Sub-Saharan Africa carries 24% of the global disease burden but has only 3% of the health workforce **(Fig. 487-5)**.

Critical diagnostics and drugs often do not reach patients in need because of supply-chain failures. Moreover, facilities fail to provide good-quality and safe care: new evidence suggests much higher rates of adverse events among hospitalized patients in low- and middle-income countries than in high-income countries. Weak government planning, regulatory, monitoring, and evaluation capacities are associated with rampant, unregulated commercialization of health services and chaotic fragmentation of these services as donors "push" their respective priority programs. With such fragile foundations, it is not surprising that low-cost, affordable, validated interventions are not reaching those who need them.

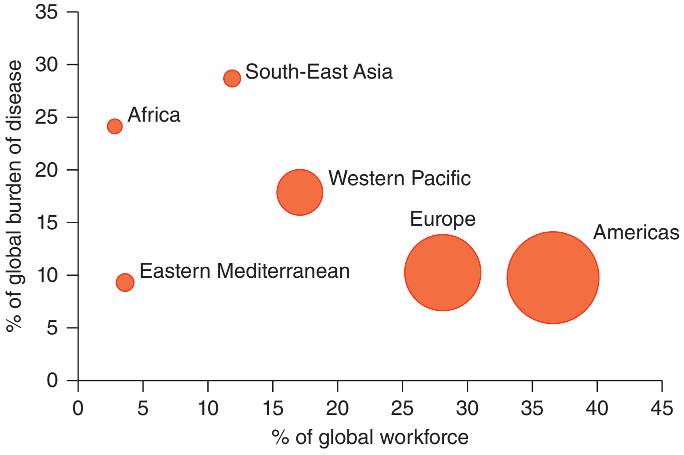

FIGURE 487-5 Global burden of disease and health workforce. *(Reproduced with permission from World Health Organization: Working Together for Health, 2006.)*

STRATIFYING SOCIAL CONDITIONS

Health care delivery systems do not exist in a vacuum but rather are embedded in a complex of social and economic forces that often stratify opportunities for health unfairly. Most worrisome are the pervasive forces of social inequality that serve to marginalize populations with disproportionately large health needs (e.g., the urban poor; illiterate mothers). Why should a poor slum dweller with no income be expected to come up with the money for a bus fare needed to travel to a clinic to learn the results of a sputum test for tuberculosis? How can a mother living in a remote rural village and caring for an infant with febrile convulsions find the means to get her child to appropriate care? Shaky or nonexistent social security systems, dangerous work environments, isolated communities with little or no infrastructure, and systematic discrimination against racialized minorities are among the myriad forces with which efforts for more equitable health care delivery must contend.

SKEWS IN SCIENCE

While science has yielded enormous breakthroughs in health in high-income countries, with some spillover to low- and middle-income countries, many important health problems continue to affect primarily low- and middle-income countries whose research and development investments are deplorably inadequate. The past two decades have seen growing efforts to right this imbalance with research and development investment for new drugs, vaccines, and diagnostics that effectively cater to the specific health needs of populations in low- and middle-income countries. For example, the TB Alliance has revitalized a previously "dry" pipeline for new tuberculosis drugs. In 2019, their new drug (pretomanid) received U.S. Food and Drug Administration approval as part of a triple oral regimen (bedaquiline, pretomanid, and linezolid [BPaL]) that treats extensively drug-resistant tuberculosis faster, better, and cheaper. Nevertheless, the global vaccine inequity witnessed during the COVID-19 pandemic is indicative of the need for further efforts to right the research and development balance globally.

As discussed above, the primary constraint on better health in low- and middle-income countries is related less to the availability of health technologies and more to their effective delivery. Underlying these health delivery challenges is a major bias regarding what constitutes legitimate "science" to improve health equity. The lion's share of health research financing is channeled toward the development of new technologies—drugs, vaccines, and diagnostics; in contrast, virtually no resources are directed toward research on how health care delivery systems can become more reliable and overcome adverse social conditions. The complexity of systems and social context is such that this issue of delivery requires an enormous investment in terms not only of money but also of scientific rigor, with the development of new research methods and measures and the attainment of greater legitimacy in the mainstream scientific establishment.

These common challenges to low- and middle-income countries partly explain the resurgence of interest in the primary health care approach and the emergence of a global movement toward universal health coverage, now enshrined as one of the Sustainable Development Goal targets adopted in Agenda 2030 by all countries at the United Nations in September 2015. In some countries (mostly middle-income), significant progress has been made in expanding coverage by health systems based on primary care and even in improving indicators of population health. More countries are embarking on the creation of primary care services despite the challenges that exist, particularly in low-income countries. Even when these challenges are acknowledged, there are many reasons for optimism that low- and middle-income countries can accelerate progress in building primary care as a key vehicle toward achieving universal health coverage.

PRIMARY HEALTH CARE IN THE TWENTY-FIRST CENTURY

The new millennium has seen a resurgence of interest in primary health care as a means of addressing global health challenges but also of familiar obstacles to its implementation. This interest has been driven by many of the same issues that led to the Declaration of Alma-Ata: rapidly increasing disparities in health between and within countries; spiraling costs of health care at a time when many people lack quality care; dissatisfaction of communities with the care they are able to access; and failure to address changes in health threats, especially noncommunicable diseases, pandemics, and challenges arising from accelerated climate change. These challenges require a comprehensive approach and strong health systems with effective primary care. Global health development agencies have partially recognized that sustaining gains in public health priorities such as HIV/AIDS and pandemic preparedness requires not only robust health systems but also the tackling of social and economic factors related to disease incidence and progression. Weak health systems have proved a major obstacle to delivering new technologies, such as COVID-19 vaccines and antiretroviral therapy, to all who need them. We discuss experiences in low- and middle-income countries in relation to primary care in greater detail below. First, we consider the features of primary health care and primary care as currently understood.

REVITALIZATION OF PRIMARY HEALTH CARE

At the 2019 World Health Assembly (an annual meeting of all countries to discuss the work of the World Health Organization [WHO]), a resolution was passed reaffirming the principles of the Declaration of Alma-Ata and the need for national health systems to be based on primary health care. This resolution reframed primary health care as three components: (1) primary care and essential public health functions as the core of integrated health services; (2) empowered people and communities; and (3) multisectoral policy and action. This reframing itself drew on the 2008 WHO World Health Report, which asserted that a primary health care approach was necessary “now more than ever” to address global health priorities, especially in terms of disparities and new health challenges.

The 2008 World Health Report highlighted four broad areas for reform (Fig. 487-6). One of these areas—the need to organize health care so that it places the needs of people first—relates to the necessity for strong primary care in health systems and what this requirement entails. The other three areas also relate to primary care. All four areas require action to move health systems in a direction that will improve health, reduce disparities, and increase satisfaction.

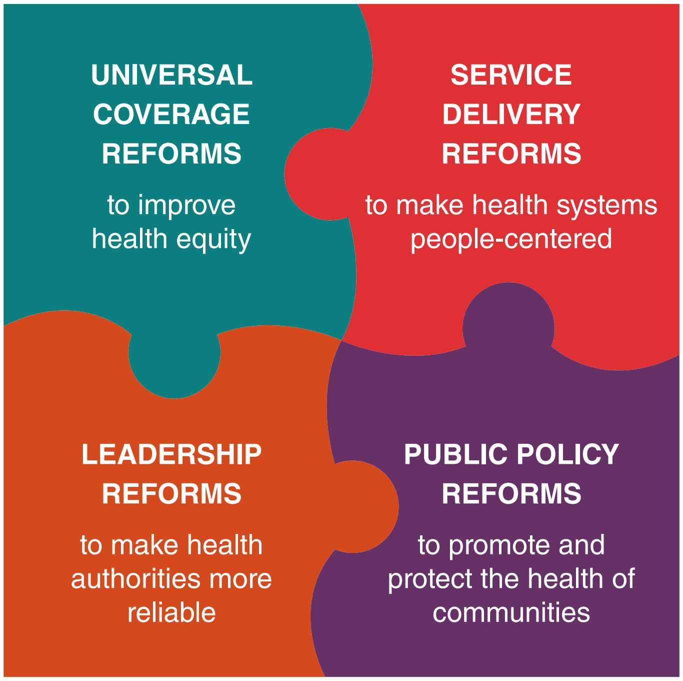
UNIVERSAL COVERAGE REFORMS
to improve health equity
SERVICE DELIVERY REFORMS
to make health systems people-centered
LEADERSHIP REFORMS
to make health authorities more reliable
PUBLIC POLICY REFORMS
to promote and protect the health of communities

FIGURE 487-6 The four reforms of primary health care renewal. *(Reproduced with permission from World Health Organization: Primary Health Care: Now More Than Ever. World Health Report 2008.)*

Universal Coverage Reforms to Improve Health Equity Despite progress in many countries, most people in the world can receive health care services only if they can pay at the point of service. Disparities in health are caused not only by a lack of access to necessary health services but also by the impact of expenditure on health. More than 100 million people are still being driven into extreme poverty each year by health care costs, with countless others deterred from accessing services at all. Moving toward prepayment financing systems for universal coverage, which ensure access to a comprehensive package of services according to need without precipitating economic ruin, has therefore emerged as a major priority in low- and middle-income countries. Increasing coverage of health services can be considered in terms of three axes: the proportion of the population covered, the range of services underwritten, and the percentage of costs paid (Fig. 487-7). Moving toward universal health coverage requires ensuring the availability of health care services to all, eliminating barriers to access, and organizing pooled financing mechanisms, such as taxation or insurance, to remove user fees at the point of service. It also requires measures beyond financing, including expansion of health services in poorly served areas, improvement in the quality of services provided to marginalized communities, and increased coverage of other social services that significantly affect health (e.g., education).

Service Delivery Reforms to Make Health Systems People-Centered Health systems have often been organized around the needs of those who provide health care services, such as clinicians and policymakers. The result is a centralization of services or the provision of vertical programs that target single diseases. The principles of primary health care, including the development of primary care, reorient care around the needs of the people to whom services cater. This “people-centered” approach aims to provide health care that is both more effective and appropriate.

The increase in noncommunicable diseases in low- and middle-income countries offers a further stimulus for urgent reform of service delivery to improve chronic disease care. As discussed above, large numbers of people currently fail to receive relatively low-cost interventions that have reduced the incidence of these diseases in high-income countries. Delivery of these interventions requires health systems that can address multiple problems and manage people over a long period within their own communities, yet many low- and middle-income countries are only now starting to adapt and build primary care services that can address noncommunicable diseases and communicable diseases requiring chronic care. Even some countries (e.g., Iran) that have had significant success in reducing communicable diseases and improving child survival have been slow to adapt their health systems to rapidly accelerating noncommunicable disease epidemics.

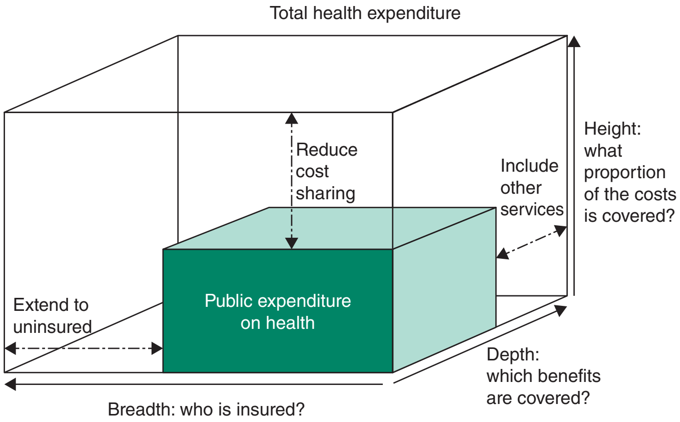
Total health expenditure
Reduce cost sharing
Include other services
Height: what proportion of the costs is covered?
Extend to uninsured
Public expenditure on health
Depth: which benefits are covered?
Breadth: who is insured?

FIGURE 487-7 Three ways of moving toward universal coverage.

People-centered care requires a safe, comprehensive, and integrated response to the needs of those presenting to health systems, with treatment at the first point of contact or referral to appropriate services. Because no discrete boundary separates people's needs for health promotion, curative interventions, and rehabilitation services across different diseases, primary care services must address all presenting problems in a unified way. Meeting people's needs also involves improved communication between patients and their clinicians, who must take the time to understand the impact of the patients' social context on the problems they develop. This enhanced understanding is made possible by improvements in the continuity of care so that responsibility transcends the limited time people spend in health care facilities. Primary care plays a vital role in navigating people through the health system; when people are referred elsewhere for services, primary care providers must monitor the resulting consultations and perform follow-up. All too often, people do not receive the benefit of complex interventions undertaken in hospitals because they lose contact with the health care system once discharged. Comprehensiveness and continuity of care are best achieved by ensuring that people have an ongoing personal relationship with a care team.

Public Policy Reforms to Promote and Protect the Health of Communities Public policies in sectors other than health care are essential to reduce disparities in health and to make progress toward global public health targets. The 2008 final report of the WHO Commission on Social Determinants of Health provided an exhaustive review of the multisectoral policies required to address health inequities at the local, national, and global levels. Advances against major challenges such as HIV/AIDS, tuberculosis, emerging infections, cardiovascular disease, cancers, and injuries require effective collaboration with sectors such as transport, housing, labor, agriculture, urban planning, trade, and energy. The COVID-19 pandemic has underscored the importance of multisectoral action to protect health; countries that have been most successful in managing the pandemic have been those best able to coordinate across their societies to implement nonpharmaceutical health measures and build social solidarity. Similarly, while tobacco control provides a striking example of what is possible if different sectors work together toward health goals, the lack of implementation of many evidence-based tobacco control measures in most countries just as clearly illustrates the difficulties encountered in such multisectoral work and the unrealized potential of public policies to improve health.

Leadership Reforms to Make Health Authorities More Reliable The Declaration of Alma-Ata emphasized the importance of participation by people in their own health care. In fact, participation is important at all levels of decision-making. Contemporary health challenges require models of leadership that acknowledge the role of government but also the many types of organizations involved in health care delivery. Governments need to engage, guide, negotiate, and regulate among these diverse actors, including but not limited to nongovernmental organizations (NGOs), the private sector, professional societies, academia, and patient groups. This difficult task requires concerted investment in leadership and governance capacity, especially if action requires the engagement of different sectors such as local government and education. The complex landscape for leadership at the national level extends to the international level. The increasing transnational character of health as seen by the interdependence of countries with respect to pandemics, climate change, health worker migration, and food security emphasizes the need for more effective global health governance mechanisms.

EXPERIENCES WITH PRIMARY CARE IN LOW- AND MIDDLE-INCOME COUNTRIES

Aspects of the primary health care approach described above, with an emphasis on primary care services, have been implemented to varying degrees in many low- and middle-income countries over the past half-century. As discussed above, some of these experiences inspired and informed the Declaration of Alma-Ata, which itself led many more countries to attempt to implement primary health care. This section describes the experiences of a selection of low- and middle-income countries in improving primary care services that have enhanced the health of their populations.

Before Alma-Ata, few countries had attempted to develop primary care on a national level. Rather, most focused on expanding primary care services to specific communities (often rural villages), making use of community volunteers to compensate for the absence of facility-based care. In contrast, in the post–World War II period, China invested in primary care on a national scale, and life expectancy doubled within roughly 20 years. The Chinese expansion of primary care services included a massive investment in infrastructure for public health (e.g., water and sanitation systems) linked to innovative use of community health workers. These "barefoot doctors" lived in and expanded care to rural villages. They received a basic level of training that enabled them to provide immunizations, maternal care, and basic medical interventions, including the use of antibiotics. Through the work of the barefoot doctors, China brought low-cost universal basic health care coverage to its entire population, most of which had previously had no access to these services.

In 1982, the Rockefeller Foundation convened a conference to review the experiences of China along with those of Costa Rica, Sri Lanka, and the state of Kerala in India. In all of these locations, good health care at low cost appeared to have been achieved. Despite lower levels of economic development and health spending, all of these jurisdictions, along with Cuba, had health indicators approaching—or in some cases exceeding—those of developed countries. Analysis of these experiences revealed a common emphasis on primary care services, with expansion of care to the entire population free of charge or at low cost, combined with community participation in decision-making about health services and coordinated work in different sectors (especially education) toward health goals. During the more than three decades since the Rockefeller meeting, some of these countries have built on this progress, while others have experienced setbacks. Recent experiences in developing primary care services show that the same combination of features is necessary for success. For example, Brazil—a large country with a dispersed population—made major strides in increasing the availability of health care from 1980 to 2010. The Brazilian Family Health Program expanded progressively across the country to reach universal coverage. This program provided communities with free access to primary care teams made up of primary care physicians, community health workers, nurses, dentists, obstetricians, and pediatricians. These teams were responsible for the provision of primary care to all people in a specified geographic area—not only those who access health clinics. Moreover, individual community health workers were responsible for a named list of people within the area covered by the primary care team. Solid evidence indicated that the Family Health Program has contributed to impressive gains in population health, particularly in terms of childhood mortality and health inequities **(Fig. 487-8)**. Nevertheless, systemic inequalities, magnified during the COVID-19 crisis, highlight how continued progress is not guaranteed, and efforts to implement and expand a primary health care approach need to adapt to new health and political challenges.

Chile has also built on its existing primary care services in the past two decades, aiming to improve the quality of care and the extent of coverage in remote areas, above all for disadvantaged populations. This effort has been made in concert with measures aimed at reducing social inequalities and fostering development, including social welfare benefits for families and disadvantaged groups and increased access to early-childhood educational facilities. As in Brazil, these steps have improved maternal and child health and have reduced health inequities. In addition to directly enhancing primary care services, Brazil and Chile have instituted measures to increase both the accountability of health providers and the participation of communities in decision-making. In Brazil, national and regional health assemblies with high levels of public participation are integral parts of the health policy-making process. Chile has instituted a patient's charter that explicitly specifies the rights of patients in terms of the range of services to which they are entitled.

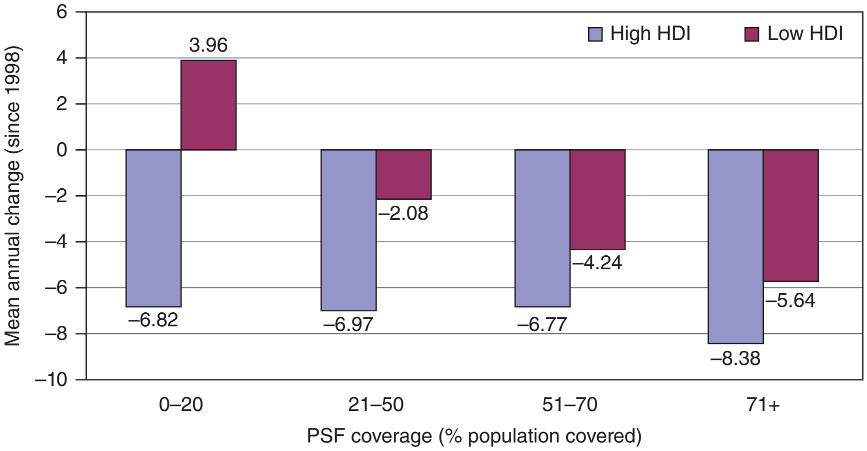

FIGURE 487-8 Improvements in childhood mortality following the Family Health Program in Brazil. HDI, Human Development Index; PSF, Program Saúde da Família (Family Health Program). *(Source: Ministry of Health, Brazil.)*

Other countries that have made recent progress with primary health care include Bangladesh, once one of the poorest countries in the world. Since achieving its independence from Pakistan in 1971, Bangladesh has seen a dramatic increase in life expectancy, and childhood mortality rates are now lower than those in neighboring nations such as India and Pakistan. The expansion of access to primary care services has played a major role in these achievements. This progress has been spearheaded by a vibrant NGO community that has focused its attention on improving the lives and livelihoods of poor women and their families through innovative and integrated microcredit, education, and primary care programs.

The above examples, along with others from the past 40 years in countries such as Thailand, Rwanda, Ethiopia, Turkey, Vietnam, and Oman, illustrate how the implementation of a primary health care approach, with a greater emphasis on primary care, has led to better access to health care services—a trend that has not been seen in many other low- and middle-income countries. This trend, in turn, has contributed to improvements in population health and reductions in health inequities. However, as these nations have progressed, other countries have shown how previous gains in primary care can easily be eroded. In Sub-Saharan Africa, undermining of primary care services contributed to catastrophic reversals in health outcomes catalyzed by the HIV/AIDS epidemic. Countries such as Botswana and Zimbabwe implemented primary health care strategies in the 1980s, increasing access to care and making impressive gains in child health. Both countries were severely affected by HIV/AIDS, with pronounced decreases in life expectancy. However, Zimbabwe has also seen political turmoil, a decline of health and other social services, and the flight of health personnel, whereas Botswana has maintained primary care services to a greater extent and has managed to organize widespread access to antiretroviral therapy for people living with HIV/AIDS.

China provides a particularly striking example of how changes in health policy relevant to the organization of health systems **(Fig. 487-9)** can have rapid, far-reaching consequences for population health. Even as the 1982 Rockefeller conference was celebrating China's achievements in primary care, its health system was unraveling. The decision to open up the economy in the early 1980s led to rapid privatization of the health sector and the breakdown of universal health coverage. As a result, by the end of the 1980s, most people, especially the poorer segments of the population, were paying directly out of pocket for health care, and almost no Chinese had insurance—a dramatic transformation. The "barefoot doctor" schemes collapsed, and the population either turned to care paid for at hospitals or simply became unable to access care. This undermining of access to primary care services in the Chinese system and the resulting increase in impoverishment due to illness contributed to the stagnation of progress in health in China at the same time that incomes in that country increased at an unprecedented rate. Reversals in primary care have meant that China now increasingly faces health care issues similar to those faced by India, although the country has more recently implemented measures to restore universal health coverage, with significant success. In both countries, rapid economic growth has been linked to lifestyle changes and noncommunicable disease epidemics. The health care systems of the two nations share two negative features that are common when primary care is weak: a disproportionate focus on specialty services provided in hospitals and unregulated commercialization of health services. China and India both saw expansion of private hospital services that cater to middle-class and urban populations who can afford care; at the same time, hundreds of millions of people in rural areas struggled to access basic services. Even in the wealthier groups, a lack of primary care services has been associated with late presentation with illness and with insufficient investment in primary prevention approaches. This neglect of prevention poses a risk of large-scale epidemics of cardiovascular disease, which could endanger continued economic growth. In addition, the health systems of both countries now depend for the majority of their funding on out-of-pocket payments by people when they use services. Thus, substantial proportions of the population must sacrifice other essential goods as a result of health expenditure and

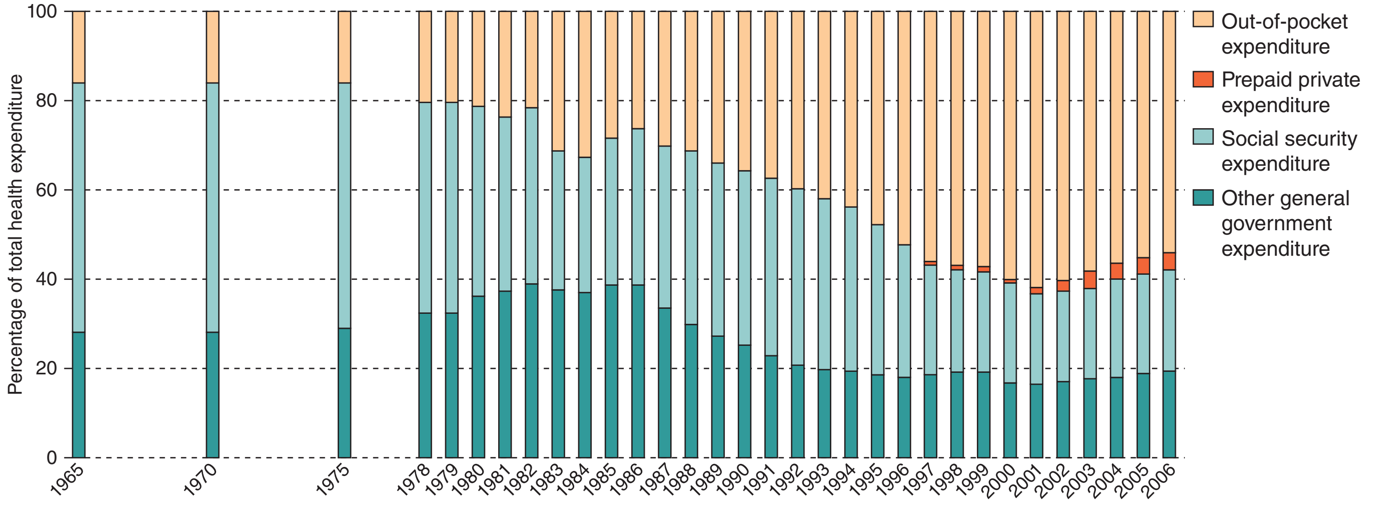

FIGURE 487-9 Changes in source of health expenditure in China over the past 40 years. *(Reproduced with permission from World Health Organization: Primary Health Care: Now More Than Ever. World Health Report 2008.)*

may even be driven into poverty by this cost. The commercial nature of health services with inadequate or no regulation has also led to the proliferation of charlatan providers, inappropriate care, and pressure for people to pay for expensive and sometimes unnecessary care.

Faced with these problems, China and India have implemented measures to strengthen primary health care. China has increased government funding of health care, has taken steps toward restoring health insurance, and has enacted a target of universal access to primary care services. India has similarly mobilized funding to greatly expand primary care services in rural areas and in urban settings. Both countries are increasingly using public resources from their growing economies to fund primary care services.

These encouraging trends are illustrative of new opportunities to implement a primary health care approach and strengthen primary care services in low- and middle-income countries. Linked to goal number 3 of the UN Sustainable Development Goals, Over the past decade, nearly all countries have adopted universal health coverage—the provision of quality health services in a timely manner at affordable cost—and the primary health care approach remains key to achieving this.

OPPORTUNITIES TO BUILD PRIMARY CARE IN LOW- AND MIDDLE-INCOME COUNTRIES

To reach global health targets, health systems must be strengthened. More money is currently being spent on health than ever before. In 2020, global health spending totaled $9 trillion (U.S.)—more than double the amount spent a decade earlier. Although most expenditure occurs in high-income countries, spending in many emerging middle-income countries has rapidly accelerated, as has the allocation of monies for this purpose by both governments in, and donors to, low-income countries. These twin trends—greater emphasis on building health systems based on primary care and allotment of more money for health care—provide opportunities to address many of the challenges discussed above in low- and middle-income countries.

Accelerating progress requires a better understanding of how global health initiatives (GHIs) can more effectively facilitate the development of primary care in low-income countries. Recent reviews including the WHO Maximizing Positive Synergies Collaborative Group and the Wellcome Trust Future of GHIs have assessed the impact on country health systems of diverse GHIs such as the Global Fund to Fight AIDS, Tuberculosis and Malaria; the Global Alliance for Vaccines and Immunization; the U.S. President's Emergency Plan for AIDS Relief; and the Global Financing Facility of the World Bank. While evidence of improved access to targeted health services is encouraging, systemic concerns related to increasing domestic resources for health and accelerating comprehensive primary care remain. If GHIs implement programs that work in tandem with other components of national health systems without undermining staffing and procurement of supplies, they have the potential to contribute substantially to the capacity of health systems to provide comprehensive primary care services.

In the context of the recent pandemic, GHIs appear even more important. The imperative of vaccinating the world's population against SARS-CoV-2 led to the creation of the multi-billion-dollar COVAX facility that failed to ensure equitable access to COVID-19 vaccines. After the pandemic, the need to recover essential health services, particularly for mothers and children, underlines the importance of GHIs like the Global Financing Facility for every mother and child as catalysts for universal access to life-saving services. The general trend is to coordinate this funding in order to reduce fragmentation of national health systems and to concentrate more on strengthening these systems. Comprehensive primary care in low-income countries must inevitably deal with the rapid emergence of chronic diseases and the growing prominence of injury-related health problems; thus, international health development assistance must become more responsive to these needs.

Beyond funding for health services, other opportunities exist. Increased social participation in health systems can help build primary care services. In many countries, political pressure from community advocates for more holistic and accountable care as well as entrepreneurial initiatives to scale up community-based services through NGOs have accelerated progress in primary care without major increases in funding. Participation of the population in the provision of health care services and in relevant decision-making often drives services to cater to people's needs as a whole rather than to narrow public health priorities.

Participation and innovation can help address critical issues related to the health workforce in low- and middle-income countries by establishing effective people-centered primary care services. Many primary care services do not need to be delivered by a physician or a nurse. Multidisciplinary teams can include paid community workers who have access to a physician if necessary but who can provide a range of health services on their own. In Ethiopia, >38,000 community health workers have been trained and deployed to improve access to primary care services, and there is increasing evidence that this measure is contributing to better health outcomes. In India, >600,000 community health advocates have been recruited as part of expanded rural primary care services. In Niger, the deployment of community health workers to deliver essential child health interventions (as a component of integrated community case management) has had impressive results in reducing childhood mortality and decreasing disparities. After the Declaration of Alma-Ata, experiences with community health workers were mixed, with particular problems regarding levels of training and lack of payment. Current endeavors are not immune from these concerns. However, with access to physician support and the deployment of teams, some of these concerns may be addressed. Growing evidence from many countries indicates that shifting appropriate tasks to primary care workers who have had shorter, less expensive training than physicians will be essential to address the human resources crisis.

Finally, recent improvements in information and communication technologies, particularly mobile phone and Internet systems, have created the potential for systematic implementation of e-health, telemedicine, and improved health data initiatives in low- and middle-income countries. These developments raise the tantalizing possibility that health systems in these countries, which have long lagged behind those in high-income countries but are less encumbered by legacy systems that have proved hard to modernize in many settings, could leapfrog their wealthier counterparts in exploiting these technologies. Although the challenges posed by poor or absent infrastructure and investment in many low- and middle-income countries cannot be underestimated and will need to be addressed to make this possibility a reality, the rapid rollout of mobile networks and their use for health and other social services in many low-income countries where access to fixed telephone lines was previously very limited offer great promise in building primary care services in low- and middle-income countries. To a partial extent, this potential has been demonstrated and even realized in many countries during the COVID-19 pandemic, with greatly increased uptake of the use of telemedicine and clinical support via digital means.

CONCLUSION

As concern continues to mount about glaring inequities in global health, there is a growing commitment to redress these egregious shortfalls, as exemplified by the central place of equity in the United Nations' Sustainable Development Goals adopted in 2015, including a specific target on the achievement of universal health coverage in all countries by 2030. This commitment begins first and foremost with a clear vision of the fundamental importance of health in all countries, regardless of income. The values of health and health equity are shared across all borders, and primary health care provides a framework for their effective translation across all contexts.

The translation of these fundamental values has its roots in four types of reform that reflect the distinct but interlinked challenges of (re)orienting a society's resources on the basis of its citizens' health needs: (1) organizing health care services around the needs of people and communities; (2) harnessing services and sectors beyond health care to promote and protect health more effectively; (3) establishing sustainable and equitable financing mechanisms for universal health coverage; and (4) investing in effective leadership to steward change locally, nationally and across borders. This common primary health care agenda highlights the striking similarity, despite enormous differences in context, in the nature and direction of the reforms that national health systems must undertake to promote greater equity in health. This shared agenda is complemented by the growing reality of global health interconnectedness due, for example, to shared microbial threats, bridging of ethnolinguistic diversity, flows in migrant health workers, and mobilization of global funds to support the neediest populations. Embracing solidarity in global health while strengthening health systems through a primary health care approach is fundamental to sustained progress in global health.

The shortfalls in health system performance, stratification of social conditions leading to unfair differences in health, and skews in science that undermine the realization of "Health for All" have never been more glaringly visible than during the COVID-19 pandemic. But they also have never commanded such global political attention at the highest level. Out of this crisis, then, is a once-in-a-lifetime opportunity to recast global and national systems to enable the genuine implementation of the primary health care approach in all countries.

FURTHER READING

Aquino R et al: Impact of the family health program on infant mortality in Brazilian municipalities. Am J Public Health 99:87, 2009.

Commission on Social Determinants of Health: Closing the Gap in a Generation: Health Equity through Action on the Social Determinants of Health: Commission on Social Determinants of Health Final Report. Geneva, World Health Organization, 2008.

Kruk ME et al: The contribution of primary care to health and health systems in low- and middle-income countries: A critical review of major primary care initiatives. Soc Sci Med 70:904, 2010.

Li X et al: The primary health-care system in China. Lancet 390:2584, 2017.

Macinko J et al: The impact of primary healthcare on population health in low- and middle-income countries. J Ambul Care Manage 32:150, 2009.

Rasanathan K, Evans T: Primary health care, the Declaration of Astana and COVID-19. Bulletin of the World Health Organization 98:801, 2020.

Rasanathan K et al: Primary health care and the social determinants of health: Essential and complementary approaches for reducing inequities in health. J Epidemiol Community Health 65:656, 2011.

Starfield B et al: Contribution of primary care to health systems and health. Milbank Q 83:457, 2005.

Tangcharoensathien V et al: Health systems development in Thailand: A solid platform for successful implementation of universal health coverage. Lancet 391:1205, 2018.

Van Lerberghe W et al: Primary Health Care: Now More Than Ever. World Health Report 2008. Geneva, World Health Organization, 2008.

488 Biology of Aging

Rafael de Cabo, David G. Le Couteur

The increase in older people over the past few decades is one of the most significant changes in human history. For the first time, people over the age of 65 years now exceed children under the age of 5 years **(Fig. 488-1)**. Aging is associated with an exponential increase in the incidence of many chronic diseases **(Fig. 488-2)**. This has significant implications for the delivery of health services and aged care in all nations.

Older people are living even longer due to advances in medical care but at the cost of longer periods of frailty and disability and of iatrogenic burdens associated with intensive medical care of multiple diseases, such as polypharmacy. Establishing the relationship between aging and disease, particularly noncommunicable disease, is one of the most important goals for biomedical research. Studies in animal models confirm that aging is malleable. The "longevity dividend" refers to the concept whereby an intervention that slows the aging process is likely to delay the onset of a wide range of age-related diseases and syndromes, as well as potentially increasing years of healthy lifespan ("health span").

DEFINITIONS OF AGING

Aging is a progressive process associated with deterioration in structure and function, leading to increased susceptibility to disease and mortality, and often associated with impaired reproductive capacity. There are statistical, biological, and phenotypic components to most definitions of aging **(Fig. 488-3)**.

THEORIES OF AGING

MUTATION ACCUMULATION AND ANTAGONISTIC PLEIOTROPY

Evolutionary theories of aging attempt to explain why aging, which impairs health and survival, has evolved, and why there is so much variability in lifespan across taxa. Aging is generally considered to be nonadaptive. This means aging has not been shaped by evolution or genetically programmed. However, many genes influence the aging process, and the initiating process of aging is most likely to involve stochastic, nonprogrammed changes in nuclear maintenance that influence gene expression and repair. Many theories of aging are based on the concept that, in the wild, mortality is secondary to extrinsic causes, such as predation, injury, and infection, and evolutionary selection pressure is generated by early life survival and reproductive success. There is minimal selection pressure to maintain health and extend life beyond early reproductive years and inevitable death from extrinsic causes. In fact, traits may evolve that are beneficial in early life and for reproduction but become harmful if the animal lives to an older age. These theories were set out by the classic "mutation accumulation" (John B.S. Haldane) and "antagonistic pleiotropy" (George C. Williams) theories of aging.

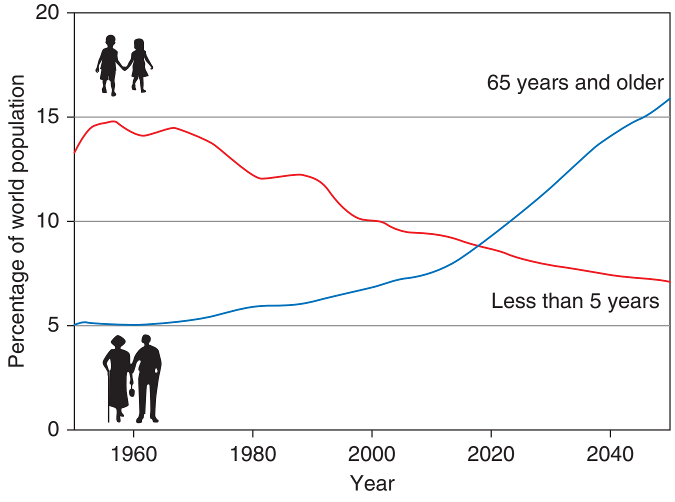

FIGURE 488-1 Globally, people over the age of 65 years now exceed children under the age of 5 years.

DISPOSABLE SOMA THEORY

There is often a trade-off between aging and reproduction. Animals with high extrinsic mortality tend to have short lives, small bodies, and greater reproductive output, while animals with low extrinsic mortality, such as humans and other primates, tend to have longer lives, larger bodies, and fewer offspring. The disposable soma theory of aging (Thomas Kirkwood) explicitly hypothesizes that evolution selects strategies that prioritize utilization of finite resources to maintain germ cells necessary for reproduction rather than for maintenance of the soma (nongerm cells), hence leading to age-related accumulation of damage to the soma.

INFORMATION THEORY OF AGING

This theory has its foundations in studies of aging biology and the role of DNA methylation in aging and as an "aging clock." With aging, there are marked changes in the epigenome (an analog information system), which have profound effects of gene expression (an information retrieval system). By contrast, the DNA code (a digital information system) is relatively stable with aging.

COMPARATIVE AGING ACROSS SPECIES

NEGLIGIBLE SENESCENCE AND PROGRAMMED AGING

There are some animals that undergo "negligible senescence," meaning that there are no obvious biological changes of aging and the rate of mortality does not increase with time. These include some strains and species of clams, sharks, hydra, and worms. The longest living vertebrate is the Greenland shark, which may live up to nearly 400 years of age. On the other hand, there are a few animals that undergo programmed aging and death. These are the semelparous animals such as Pacific salmon and marsupial mice **(Fig. 488-4)**.

GRANDMOTHER EFFECT

There are species, including humans, where evolution could influence late-life survival through what is called the grandmother effect. In these species, survival of offspring depends on the care provided by their long-lived grandmothers. This also explains the development of extended postreproductive survival in humans.

BIOLOGICAL HALLMARKS OF AGING

Aging is associated with a range of molecular processes that are remarkably similar between species. These "hallmarks of aging" are the mechanistic pathways that cause aging. The processes are highly interconnected, and impairment of one process will impact the others. The hallmarks include those that act at the various biological strata and together erode various pillars of health **(Fig. 488-5)**.

Interventions that alter the behavior of each of these pathways (via genetic manipulation, pharmacologic treatments, or nutritional interventions) influence aging and lifespan of laboratory animals such as mice, fruit flies (*Drosophila melanogaster*) or worms (*Caenorhabditis elegans*). Each of the hallmarks is a potential target for pharmacotherapies that might delay aging and the onset of age-related morbidity and increase both health span and lifespan.

GENOMIC INSTABILITY

The integrity of DNA is vulnerable to many exogenous (e.g., irradiation, chemicals, transposons) and endogenous (e.g., oxidative stress)

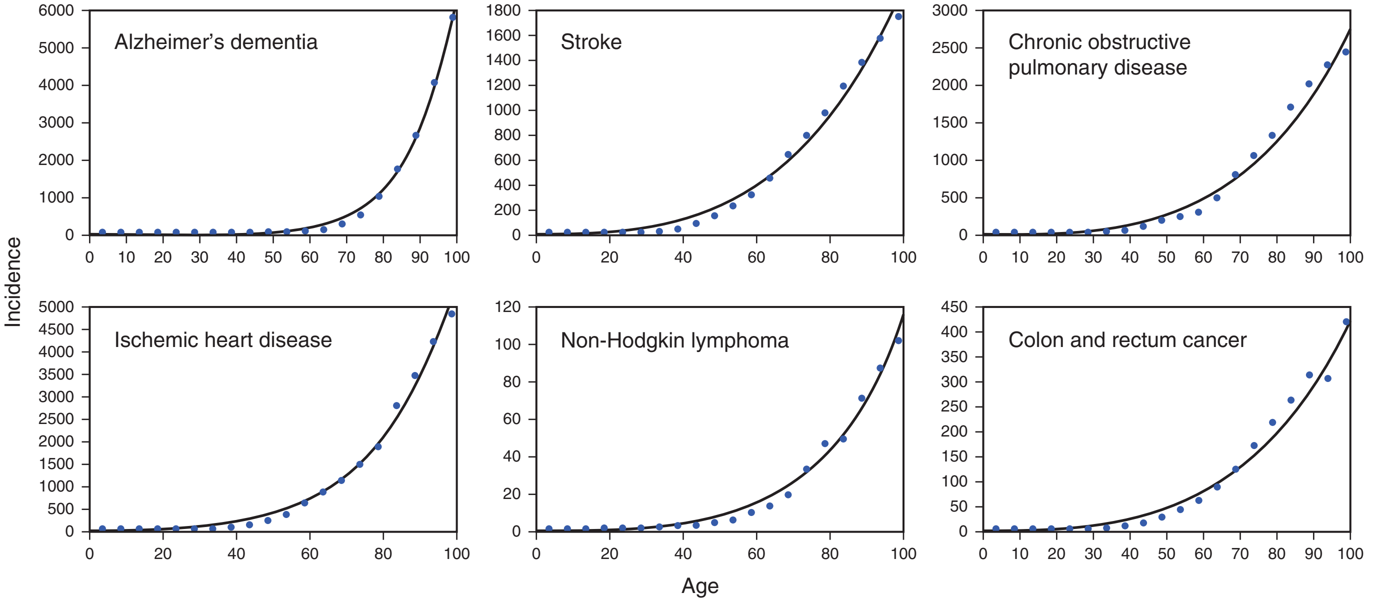

FIGURE 488-2 **Many noncommunicable diseases have an exponential increase in incidence with age**. Aging biology is likely to be an integral part of the mechanisms for these diseases.

stresses that generate largely random DNA lesions such as point mutations, translocations, and chromosomal anomalies. Mitochondrial DNA is especially susceptible to damage with aging because of its proximity to free radicals produced during oxidative phosphorylation and lack of histones and repair mechanisms. Genetic manipulation of nuclear and mitochondrial DNA repair mechanisms shortens lifespan in mice, while human premature aging syndromes are associated with deficiencies in genes necessary for nuclear maintenance. For example, Werner's syndrome is caused by mutations in a recQ helicase gene (*WRN*) required for repair of double-stranded breaks, and Hutchinson-Guildford progeria syndrome is caused by mutations in the lamin A gene (*LMNA*) required for structural support of the nucleus.

TELOMERE ATTRITION

Telomeres are repeat sequences at the ends of linear chromosomes that counter the inability of DNA polymerase to replicate the tips of chromosomes. In humans, telomeres consist of a redundant TTAGGG sequence repeated several thousand times. Some cells (e.g., germ cells, tumor cells) contain telomerase, which can reform telomeres that are shortened during replication. In most cells, after multiple divisions, the telomeres are truncated to a point where cell division cannot continue. In cell culture, this number of divisions is called the Hayflick limit, and cells that cannot undergo further division are said to have entered a phase of "replicative cellular senescence." Some studies in humans have found that telomere length in blood cells decreases with age, while mice that have been genetically manipulated to have short or long telomeres have decreased and increased lifespans, respectively. In humans, telomerase deficiency is associated with pulmonary fibrosis and aplastic anemia.

EPIGENETIC ALTERATIONS

Gene expression is regulated by DNA methylation, histone modification, chromatin remodeling, and noncoding RNAs. These all change with age, leading to altered transcription of genes, especially those involved with inflammation, mitochondrial function, and autophagy pathways. There are consistent age-related changes in the pattern of DNA methylation in human blood samples that have been called "epigenetic clocks" (e.g., "Horvath epigenetic clock") because they reflect chronologic age. Histones are proteins that package DNA into nucleosomes, thus influencing DNA available for transcription. The sirtuins are a family of NAD-dependent proteins that regulate histones through deacetylation and have significant effects on aging and lifespan. For example, overexpression of sirtuins and activation by drugs such as resveratrol increase lifespan in model organisms. De-repression of retrotransposons has been implicated in aging, and nonreverse transcriptase inhibitors decrease this and delay aging in mice.

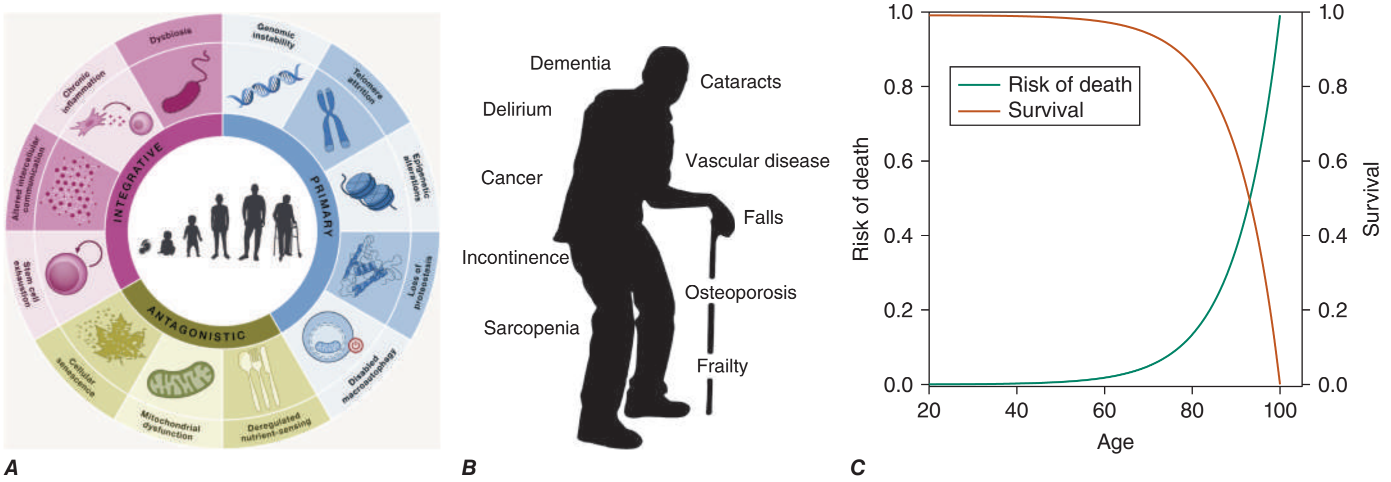
Dysbiosis
Genomic instability
Telomere attrition
Epigenetic alterations
Loss of proteostasis
Disabled macroautophagy
Deregulated nutrient-sensing
Mitochondrial dysfunction
Cellular senescence
Stem cell exhaustion
Altered intercellular communication
Chronic inflammation
INTEGRATIVE
PRIMARY
ANTAGONISTIC
A
Dementia
Cataracts
Delirium
Vascular disease
Cancer
Falls
Incontinence
Osteoporosis
Sarcopenia
Frailty
B
Risk of death
Survival
1.0
0.8
0.6
0.4
0.2
0.0
20
40
60
80
100
Age
C

FIGURE 488-3 **Definitions of aging** often include (***A***) a biological component encapsulated by the 12 hallmarks of aging, (***B***) a phenotypic component that includes many chronic diseases and syndromes of aging, and (***C***) a statistical component that in most species involves an exponential (Gompertz) increase in the risk of mortality with age.

Ocean quahog clam
Greenland shark

Negligible senescence

Antechinus
Pacific salmon

Semelparous animals

Asian elephant
Humans

Grandmother effect

FIGURE 488-4 Some animals undergo negligible senescence, while others such as the semelparous animals undergo programmed aging and death. In some long-lived species, including humans, there is a prolonged postreproductive period that can evolve because of the beneficial effects of grandmothers on survival of infants.

LOSS OF PROTEOSTASIS

Damaged proteins in cells are removed by the autophagy-lysosomal system and ubiquitin-proteosome system. These processes are impaired with aging, which can lead to intracellular and extracellular aggregates of damaged proteins and other cellular components such as lipofuscin, Lewy bodies, neurofibrillary tangles, and amylin **(Fig. 488-6)**.

DISABLED MACROAUTOPHAGY

Autophagy refers to the sequestration and digestion of proteins (i.e., proteostasis), nonprotein macromolecules (e.g., glycogen), and organelles (e.g., mitochondria, "mitophagy"). Autophagy and the expression of autophagy-related genes decline with old age. Autophagy is activated by caloric restriction (CR), spermidine, and inhibition of mechanistic target of rapamycin (mTOR) that all are associated with delayed aging. Loss of function of genes that are associated with autophagy in humans is associated with increased susceptibility to a range of age-related diseases.

DEREGULATED NUTRIENT SENSING

Nutrition has a profound effect on aging in all species. One of the most important nutritional interventions that influences aging is CR. CR involves providing animals with less food (usually ~30–50% less than eaten by ad libitum fed controls). When this is maintained over a lifetime, lifespan is increased, and many age-related pathologies and diseases are delayed. There are several interconnected pathways that mediate the effects of nutrition and CR on aging and age-related health through a wide range of downstream effects **(Fig. 488-7)**. Drugs that act on these pathways to replicate CR are called CR-mimetics and increase lifespan regardless of calorie intake. These key pathways include the following:

1. Insulin and insulin-like growth factor (IGF)-1 signaling pathway (IIS), including growth hormone (GH). Animals with genetic downregulation of this pathway have longer lifespans, including the very long-living Methuselah mice with knock out of the receptor for GH. Humans with GH receptor deficiency (Laron syndrome) have less incidence of cancer and diabetes. Many of the downstream effects on aging pathways are mediated by FOXO (forkhead box O) transcription factors. Genetic variation in some FOXO genes is associated with longevity in humans.
2. mTOR pathway. This pathway is activated by amino acids via tRNA and the IIS and therefore can be manipulated by dietary protein intake. Inactivation of mTOR by dietary restriction or rapamycin is associated with reduced protein synthesis and increased autophagy, leading to increased lifespan.
3. Sirtuins. The sirtuins are a class of seven proteins that respond to dietary energy via the cofactor NAD. Sirtuins regulate gene expression via deacetylation of histone and nonhistone proteins and have been shown to impact many processes related to aging, including apoptosis, inflammation, DNA damage/repair, and mitochondrial biogenesis. Increased sirtuin activity induced by genetic manipulation, resveratrol, or NAD supplementation has been associated with increased lifespan.
4. AMP-activated protein kinase (AMPK) responds to dietary energy restriction via cellular levels of AMP. Activation of AMPK with metformin increases lifespan in animals, and observational studies in humans suggest that metformin impacts age-related conditions. Human clinical trials of metformin are underway to determine its effects on aging.
5. Fibroblast growth factor 21 (FGF21) responds to dietary protein, with protein restriction leading to increased production and release into the blood. FGF21 promotes AMPK signaling and insulin sensitivity, and overexpression in mice increases lifespan.

MITOCHONDRIAL DYSFUNCTION

Age-related changes in mitochondria include increased electron leakage and decreased ATP production, primarily due to impaired complex IV activity. Mitochondrial DNA damage accumulates with age, and mitochondria may become swollen with disrupted cristae. Impaired mitochondrial function leads to increased generation of superoxide radicals and hydrogen peroxide, which are potent oxidants. Consequently, aging is associated with the accumulation of oxidative damage to fat, proteins, and DNA. This forms the basis of the free radical theory of aging. Antioxidants have been investigated as a method to delay age-related oxidative stress but have been ineffective in delaying aging. On the other hand, many of the nutritional and pharmacologic interventions that delay aging are associated with increased mitochondrial biogenesis, usually via their effect on the PGC1α transcription factor.

CELLULAR SENESCENCE

Senescent cells have stopped dividing because of either telomere shortening or other damage mediated by the INK4/ARF system. Senescent cells (sometimes called zombie cells) accumulate in various tissues with increasing age. They produce a range of inflammatory cytokines including interleukin 6 (IL-6) and tumor necrosis factor α (TNF-α), collectively called the senescence-associated secretory phenotype (SASP), which contributes to systemic inflammation. Elimination of senescent cells that express INK4 and several drugs that eradicate senescent cells called senolytics (e.g., dasatinib, quercetin, fisetin) delay the progression of age-related pathologies.

STEM CELL EXHAUSTION

The numbers of stem cells decline with aging, probably secondary to replicative senescence and telomere shortening. Transplantation of pluripotent stem cells from young donors to old recipients extends lifespan in mice and may improve frailty in humans. Another strategy involves induction of genes that regulate stem cells (Yamanaka OSKM factors:

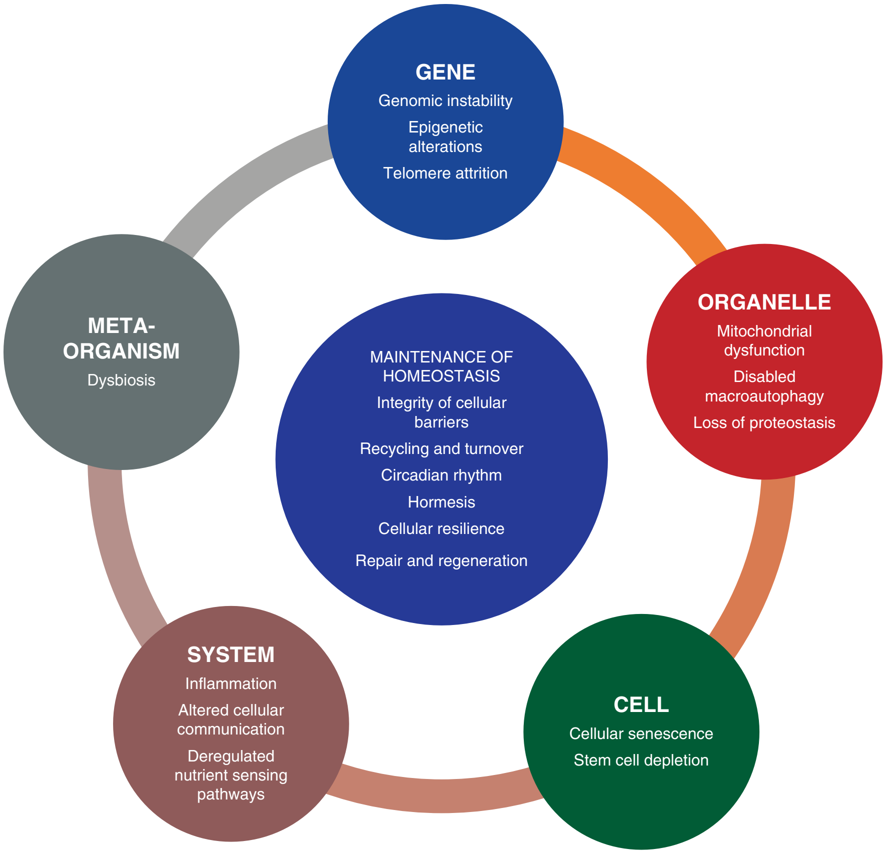
GENE
Genomic instability
Epigenetic alterations
Telomere attrition
ORGANELLE
Mitochondrial dysfunction
Disabled macroautophagy
Loss of proteostasis
CELL
Cellular senescence
Stem cell depletion
SYSTEM
Inflammation
Altered cellular communication
Deregulated nutrient sensing pathways
META-ORGANISM
Dysbiosis
MAINTENANCE OF HOMEOSTASIS
Integrity of cellular barriers
Recycling and turnover
Circadian rhythm
Hormesis
Cellular resilience
Repair and regeneration

FIGURE 488-5 The hallmarks of aging include 12 processes that form the foundations of the biological changes of aging and age-related diseases and syndromes.

OCT4, SOX2, KLF4, MYC), which delays aging in progeria mouse models. After bone marrow transplantation, no evidence suggests that the recipient runs out of donor stem cells even though the stem cell population could be compromised numerically and due to the age of the donor. Can the stem cells renew themselves in the new host?

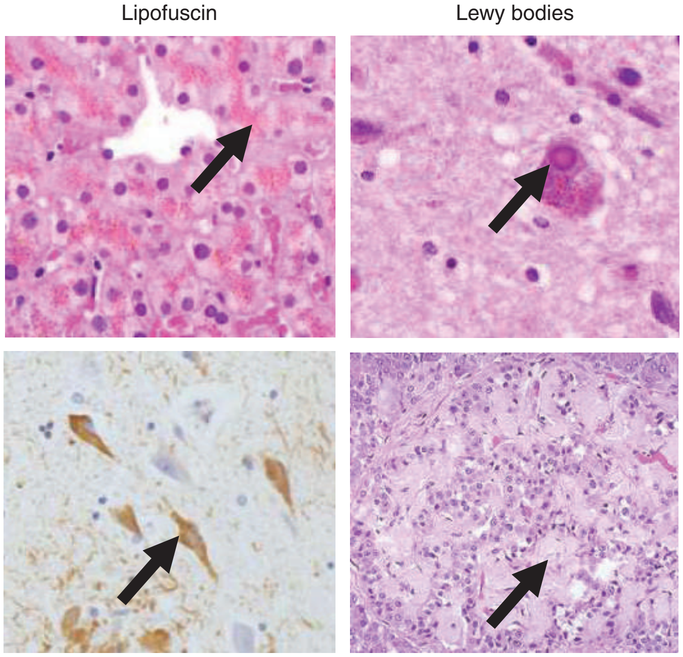
Lipofuscin
Lewy bodies
Neurofibrillary tangles
Amylin

FIGURE 488-6 Impaired proteostasis with old age contributes to the accumulation of aggregates often associated with disease (lipofuscin in the liver, Lewy bodies in the substantia nigra, neurofibrillary tangles in neurons, and amylin in pancreatic islets).

ALTERED INTRACELLULAR COMMUNICATION

Many signaling pathways undergo changes with old age including insulin/IGF-1, dopaminergic, sex hormones, growth differentiation factor 11 (GDF11), and the renin-angiotensin system. These and other bloodborne factors might explain some of the antiaging effects of blood transfusions. Age-related changes are also noted in the extracellular matrix, including increased fibrosis, that impact cell-to-cell communication.

CHRONIC INFLAMMATION

Aging is associated with low-grade activation of the innate immune system, leading to elevated levels of IL-6 and TNF-α and often elevated C-reactive protein and erythrocyte sedimentation rate (ESR) with a lower lymphocyte-to-neutrophil ratio. This has been called “inflammaging.” This may be secondary to several factors including the SASP, chronic infection with cytomegalovirus, obesity, leaky gut, and activation of the nuclear factor κB (NF-κB) pathway. Inflammation is a key factor in the pathogenesis of many chronic diseases and, in particular, frailty.

DYSBIOSIS

Complex changes occur in the gut microbiome with old age in humans. The degree of microbial heterogeneity declines, and changes are seen in some species and phyla (e.g., Bacteroidetes, *Akkermansia*). Fecal microbiome transplantation and various probiotics have some effects on aging in mice.

GEROSCIENCE

Old age is the major risk factor for many chronic diseases. The nature of the relationship has been the source of debate for millennia. On one hand, it has been argued that aging is similar to any other disease and, therefore, is amenable to therapeutic interventions. On the other hand, it has been argued that aging is an inevitable and untreatable process that increases the risk for other diseases.

Regardless, there is a marked and often exponential increase in the prevalence and incidence of most chronic diseases with age (Fig. 488-2), and there are several conditions that are generally considered to be primarily age-related disorders, including dementia, sarcopenia, frailty, and osteoporosis. These conditions are very rare below the age of 50 years.

Geroscience refers to the study of the relationship between aging biology and disease. Aging leads to impairment of many physiologic systems that will increase susceptibility to disease. For example, aging is associated with marked changes in immune function. Immunosenescence refers to age-related deterioration of the responsiveness of the immune system to infection and other antigenic challenges, caused by thymic involution with reduced naïve T cells and impaired memory T cells, a reduction in hematopoietic stem cells, and impaired antigen-presenting cell function. Aging is associated with many endocrine changes, most notably a reduction in sex steroids and GH, which contributes to sarcopenia and osteoporosis. Vascular changes including increased arterial stiffness leading to high peripheral vascular resistance and microvascular pathology have an obvious link with cardiovascular disease, but also probably contribute to other conditions such as dementia, osteoarthritis, and sarcopenia.

Although the usual dogma is that aging is a process that increases susceptibility to diseases, the relationship between chronic disease and aging may be much more fundamental. The pathogenesis of most chronic diseases includes one or more of the hallmarks of aging, and differences between disease and normal aging are defined by a quantitative difference in the expression of these hallmarks and the tissues that are affected. Likewise, the difference between aging and

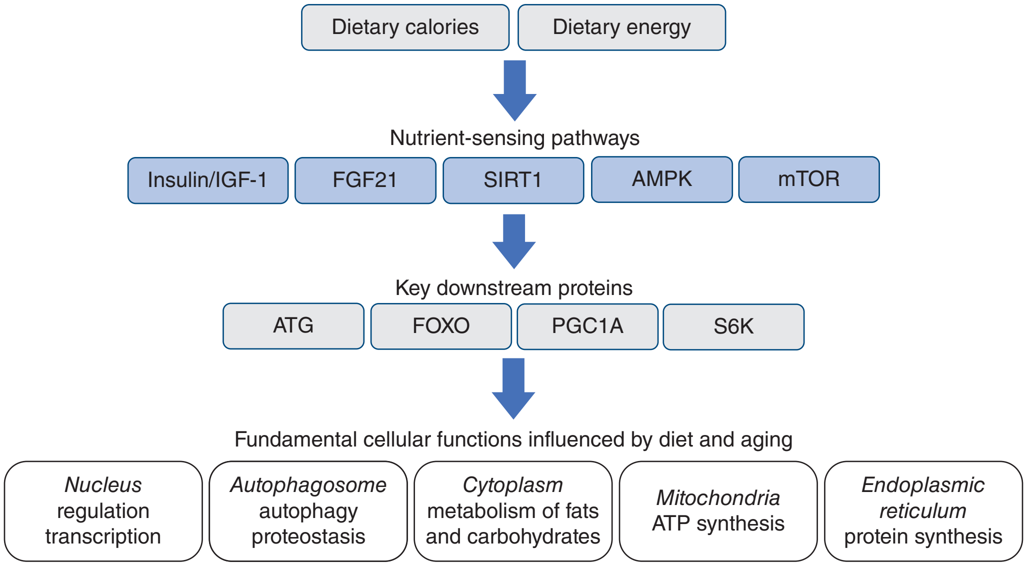
Dietary calories
Dietary energy
Nutrient-sensing pathways
Insulin/IGF-1
FGF21
SIRT1
AMPK
mTOR
Key downstream proteins
ATG
FOXO
PGC1A
S6K
Fundamental cellular functions influenced by diet and aging
Nucleus regulation transcription
Autophagosome autophagy proteostasis
Cytoplasm metabolism of fats and carbohydrates
Mitochondria ATP synthesis
Endoplasmic reticulum protein synthesis

FIGURE 488-7 Nutrient sensing pathways. The main molecular switches that respond to changes in dietary intake (*blue boxes*: insulin/IGF-1, mTOR, AMPK, SIRT1, FGF21) influence a range of downstream intermediaries (some of these are shown in the *gray boxes*). These regulate key cellular processes (*white boxes*) including metabolism, autophagy, mitochondria, and protein synthesis.

disease in terms of the clinical features is often based on quantitative differences. Therefore, there is a continuum between aging and many chronic diseases. These are often separated by so-called "prediseases" or "subclinical diseases," which are evidence for this continuum. Then, chronic disease can be considered a manifestation of aging that is predominant in a particular tissue. The presence of several chronic diseases, termed *multimorbidity*, represents aging changes that are more advanced in several tissues. Frailty can be defined as a multisystem aging syndrome where aging changes are present in most tissues, leading to multiple deficits and impaired function.

STRATEGIES THANT INCREASE HEALTH SPAN AND DELAY AGING

Aging is an intrinsic feature of human life whose manipulation has fascinated humans ever since becoming conscious of their existence. Several long-term experimental interventions (e.g., resveratrol, rapamycin, spermidine, and metformin) may open doors for pharmacologic strategies. Surprisingly, most of the effective aging interventions proposed to date converge on only a few molecular pathways: nutrient signaling, mitochondrial proteostasis, and the autophagic machinery.

Lifespan is inevitably accompanied by a gradual functional decline, steady increase of several chronic diseases, and ultimately death. For millennia, it has been a dream of mankind to prolong both lifespan and health span. Developed countries have profited from advances in medical care and technology, improvements in their public health care systems, and better living conditions derived from their socioeconomic power—to achieve remarkable increases in life expectancy during the past century. In the United States, the percentage of the population aged ≥65 is projected to increase from 13% in 2010 to 19.3% in 2030. However, old age remains the leading risk factor for major life-threatening disorders. The number of people suffering from age-related diseases is anticipated to almost double over the next two decades. The prevalence of age-related pathologies represents a major threat and an economic burden that urgently needs effective interventions.

Molecules, drugs, and other interventions that might decelerate aging processes continue to be a major focus among the general public and scientists of all biological and medical fields. Over the past three decades, this interest has taken root because many of the molecular mechanisms underlying aging are interconnected and linked with pathways that cause diseases, including cancer and cardiovascular and neurodegenerative disorders. Unfortunately, results often lack reproducibility because of the unavoidable problem of the time needed to assess the effectiveness of antiaging interventions in mammals. Experiments lasting the lifetime of animal models are prone to develop artifacts, increasing the possibilities and time windows for experimental discrepancies. Some inconsistencies in the field arise from overinterpreting the results of animal models with shortened lifespan and scenarios of accelerated aging.

Molecules, drugs, and other interventions have been proposed to have antiaging properties throughout history and into the present. In the following sections, interventions will be restricted to those that meet the following highly selective criteria: (1) promotion of lifespan and/or health span, (2) validation in at least three model organisms, and (3) confirmation by at least three different laboratories. These include CR and intermittent fasting regimens, some pharmacotherapies (resveratrol, rapamycin, spermidine, and metformin), and exercise.

CALORIC RESTRICTION

One of the most important and robust interventions that delays aging is CR. This outcome has been recorded in rodents, dogs, worms, flies, yeasts, monkeys, and prokaryotes. CR is defined as a reduction in the total caloric intake, usually of ~30%, and without malnutrition. CR reduces the nutrient-mediated release of growth factors, such as GH, insulin, and IGF-1, which have been shown to accelerate aging and enhance the probability for mortality in many organisms. Yet, the effects of CR on aging were first discovered by McCay in 1935, long before the discovery of these hormones and growth factors and signaling properties. Some of the pathways that mediate this remarkable response of CR have been elucidated in experimental models. These include the nutrient-sensing pathways (mTOR, AMPK, insulin/IGF-1, and sirtuins) and the family of FOXO transcription factors (orthologs are found in *D. melanogaster* and *C. elegans*). The transcription factor Nrf2 appears to confer most of the anticancer properties of CR in mice, even though it is dispensable for lifespan extension.

The effects of CR in monkeys have been assessed in two studies with different outcomes: one study observed prolonged life, while the other did not. In these monkey studies, there were key differences in the onset of the intervention, diet composition, feeding protocols, and genetic background that may explain this discordance. However, both studies confirmed that CR increases health span by reducing the risk for diabetes, cardiovascular disease, and cancer. In humans, CR is associated with extended lifespan and increased health span. This is most convincingly demonstrated in Okinawa, Japan, where one of the most long-lived human populations resides. In comparison to the rest of the Japanese population, Okinawan people usually combine an above-average amount of daily exercise with a below-average food intake. However, when Okinawan families moved to Brazil, they adopted a Western lifestyle that affected both exercise and nutrition, causing a rise in weight and a reduction in life expectancy by nearly two decades. In the Biosphere II project, volunteers lived together for 24 months undergoing an unforeseen severe CR that led to improvements in insulin, blood sugar, glycated hemoglobin, cholesterol levels, and blood pressure—all outcomes that would be expected to benefit lifespan. CR changes many aspects of human aging that might influence lifespan such as the transcriptome, hormonal status (especially IGF-1 and thyroid hormones), oxidative stress, inflammation, mitochondrial function, glucose homeostasis, and cardiometabolic risk factors. Epigenetic modifications are also an emerging target for CR. The first clinical trial of CR in people at average weight (a body mass index between 20 and 25) started in 2007. The Comprehensive Assessment of Long-Term Effects of Reducing Intake of Energy (CALERIE) included 143 adults between the ages of 21 and 50 years intending to reduce their caloric intake by 25% of their typical intake for 2 years; also included was a group of 75 people who remained on their normal diets and caloric intake. CALERIE has provided evidence for improvements to the quality of life, immune health, cardiometabolic integrity, liver function, and skeletal muscle quality, even though the participants only reached a moderate level of CR (11.9 ± 0.7%) over the 2-year span. CR in this clinical trial also led to a reduction in the rate of biological aging measured by a series of common clinical biomarkers of preservation of physiologic and functional integrity (e.g., liver enzymes, albumin, fasting blood glucose, insulin, and blood pressure). At the molecular level, gene expression analyses in a subset of CALERIE participants indicate that CR induces the regulation of core longevity pathways linked to the preservation of mitochondrial function and stability, lowering chronic inflammation and reducing oxidative stress.

PERIODIC FASTING

It must be noted that maintaining CR while avoiding malnutrition over a long period of time is not only arduous in humans but also linked with substantial side effects. For instance, prolonged reduction of calorie intake may decrease fertility and libido, impair wound healing, reduce the potential to combat infections, and lead to amenorrhea and osteoporosis. How can CR be translated to humans in a socially and medically acceptable way? A whole series of periodic fasting regimens are asserting themselves as suitable strategies, among them (1) the alternate-day fasting diet, (2) the "5:2" intermittent fasting diet, (3) a 48-h fast once or twice each month, and (4) daily time-restricted feeding (TRF). Periodic fasting is psychologically more viable, lacks some of the negative side effects of CR, and is only accompanied by minimal weight loss. All these dietary interventions involve a substantial reduction of caloric intake for a defined period and typically lead to an elevation of circulating ketone bodies during those low-calorie intake periods, illustrating the metabolic switch from the utilization of glucose as a fuel source to the use of fatty acids and ketone bodies. This metabolic shift results in a reduction in the respiratory exchange ratio (the ratio of carbon dioxide produced to oxygen consumed), indicating greater metabolic flexibility and energy production efficiency from use of fatty acids and ketone bodies.

It is striking that many cultures implement periodic fasting rituals, for example, some Buddhists, Christians, Hindus, Jews, Muslims, and practitioners of African animistic religions. It could be speculated that a selective advantage of fasting versus nonfasting populations is conferred by health-promoting attributes of religious routines that periodically limit caloric intake. Indeed, several lines of evidence indicate that intermittent fasting regimens exert antiaging effects. For example, improved morbidity and longevity were observed among Spanish nursing home residents who underwent alternate-day fasting. Rats subjected to alternate-day fasting live up to 83% longer than control animals fed ad libitum, and even one 24-h fasting period every 4 days is sufficient to generate lifespan extension.

Repeated fasting and eating cycles may circumvent the negative side effects of sustained CR. This strategy may even yield health benefits despite overeating behavior during the nonfasting periods. In a landmark experiment, mice fed a high-fat diet in a time-restricted manner, i.e., with regular fasting breaks, showed reduced inflammation markers, did not develop fatty liver, and were slim in comparison to mice fed ad libitum despite equivalent total calories consumed. From an evolutionary point of view, this kind of feeding pattern may reflect mammalian adaptation to food availability: overeating in times of nutrient availability (e.g., after a hunting success) and starvation in times of food scarcity. This is how some indigenous peoples who have avoided Western lifestyles live today; those who have been investigated show limited signs of age-induced diseases such as cancer, neurodegeneration, diabetes, cardiovascular disease, and hypertension.

Fasting exerts beneficial effects on health span by minimizing the risk of developing age-related diseases, including hypertension, neurodegeneration, cancer, and cardiovascular disease. The most effective and rapid repercussion of fasting is a reduction in hypertension. Two weeks of water-only fasting resulted in blood pressure <120/80 mmHg in 82% of subjects with borderline hypertension. Ten days of fasting cured all hypertensive patients who had been taking antihypertensive medication previously. Periodic fasting also dampens the consequences of many age-related neurodegenerative diseases in mouse models of Alzheimer's disease, Parkinson's disease, Huntington's disease, and frontotemporal dementia, but not amyotrophic lateral sclerosis. Fasting cycles are as effective as chemotherapy against certain tumors in mice. When combined with chemotherapy, fasting protects mice against the negative side effects of chemotherapeutic drugs, while enhancing efficacy against tumors. Combining fasting and chemotherapy rendered 20–60% mice cancer-free when inoculated with highly aggressive tumors like glioblastoma or pancreatic tumors, which have 100% mortality even with chemotherapy.

PHARMACOLOGIC INTERVENTIONS TO DELAY AGING AND INCREASE LIFESPAN

Virtually all obese people know that stable weight reduction will lower their risk of cardiometabolic disease and enhance their overall survival, and yet only 20% of overweight individuals are able to lose 10% weight for a period of at least 1 year. Even in the most motivated people (e.g., the "Cronies" who deliberately attempt long-term CR to extend their lives), long-term CR is extremely difficult to adhere to. Thus, much focus has been directed at the possibility of developing medicines that replicate the beneficial effects of CR but without the need for reducing food intake ("CR-mimetics," **Fig. 488-8**).

- *Resveratrol.* Resveratrol, an agonist of SIRT1, is a polyphenol that is found in grapes and red wine. The potential of resveratrol to promote lifespan was first identified in yeast, and it has gathered fame since, at least in part, because it has been suggested to be responsible for the so-called French paradox whereby wine reduces some of the cardiometabolic risks of a high-fat diet. Resveratrol has been reported to increase lifespan in many lower-order species such as yeast, fruit flies, worms, and fish, as well as mice on high-fat diets. In monkeys fed a diet high in sugar and fat, resveratrol had beneficial

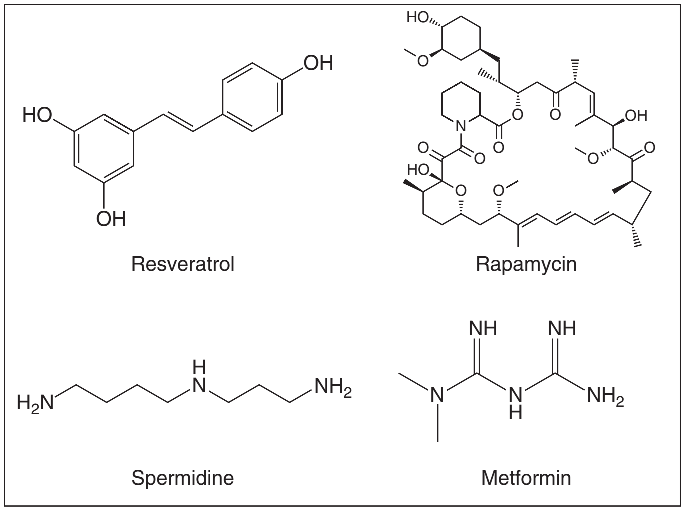

FIGURE 488-8 Chemical structure of four agents (resveratrol, rapamycin, spermidine, and metformin) that have been shown to delay aging in experimental animal models.

outcomes related to inflammation and cardiometabolic parameters. Some studies in humans have also shown improvements in cardiometabolic function, while others have not. Studies in animals and humans reveal that resveratrol mimics some of the metabolic and gene expression changes of CR. In most experimental models, resveratrol induces beneficial health effects by suppressing inflammation, oxidative damage, tumorigenesis, and immunomodulatory activities. Resveratrol also leads to improvements in mitochondrial function and protection against obesity, cancer, and cardiovascular dysfunction.

- *Rapamycin*. Rapamycin, an inhibitor of mTOR, was originally discovered on Easter Island (Rapa Nui, hence its name) as a bacterial secretion with antibiotic properties. Before its emergence in the antiaging arena, rapamycin was known as an immunosuppressant and cancer chemotherapeutic in humans. Rapamycin extends lifespan in all organisms tested so far, including yeast, flies, worms, and mice. However, the potential utility of rapamycin in lifespan extension in humans is likely to be limited by adverse effects related to immunosuppression, impaired wound healing, proteinuria, and hypercholesterolemia, among others. An alternative strategy may be the implementation of intermittent rapamycin treatment, which was found to increase mouse lifespan.
- *Spermidine*. Spermidine is a physiologic polyamine that induces autophagy-mediated lifespan extension in yeast, flies, and worms. Endogenous spermidine levels decrease during life in virtually all organisms including humans, with the remarkable exception of centenarians. Oral administration of spermidine or upregulation of bacterial polyamine production in the gut leads to lifespan extension in short-lived mouse models. The lifespan effects of spermidine are mediated through the inhibition of histone acetylases and the activation of autophagy genes, such as *atg7*, *atg11*, and *atg15* (Morselli et al., 2009). Spermidine has also been found to have beneficial effects on neurodegeneration and cardioprotection through activation of autophagy. Spermidine supplementation is safe in humans and has been associated with positive effects on cognitive function of older adults and on blood pressure maintenance.
- *Metformin*. Metformin, a biguanide first isolated from the French lilac, is widely used for the treatment of type 2 diabetes. Metformin decreases hepatic gluconeogenesis and increases insulin sensitivity. Other actions of metformin include AMPK activation, leading to mTOR inhibition and lower mitochondrial complex I activity, and activation of the transcription factor SKN-1/Nrf2. Metformin increases lifespan in different mouse strains including female mice predisposed to high incidence of mammary tumors. At a biochemical level, metformin supplementation is associated with reduced oxidative damage and inflammation and mimics some of the gene expression changes seen with CR. Based on experimental data on the positive outcomes in model organisms and the evidence emerging from epidemiologic studies, a clinical trial known as TAME (Targeting Aging with Metformin) has been initiated to assess whether metformin can delay the onset of age-related diseases beyond its effects on glucose metabolism. TAME is planning to enroll 3000 subjects, ages 65–79, in a multicenter trial in the United States.

EXERCISE AND PHYSICAL ACTIVITY

In humans and animals, regular exercise reduces the risk of morbidity and mortality. Given the marked increase in cardiovascular disease in the elderly, the effects of exercise on human health may be even stronger than those seen in laboratory mice, as mice do not develop atherosclerosis and have a far lower incidence of age-related cardiovascular disease. An increase in aerobic exercise capacity, which declines during aging, is associated with favorable effects on blood pressure, lipids, glucose tolerance, bone density, and depression in older people. Likewise, exercise training protects against aging disorders such as cardiovascular disease, diabetes mellitus, and osteoporosis. Exercise is the only intervention that can prevent or even reverse sarcopenia (age-related muscle wasting). Even moderate or low levels of exercise (30-min walking per day) have significant protective effects in obese subjects. In older people, regular physical activity has been found to increase the length of stay in independent living.

While clearly promoting health and quality of life, regular exercise does not extend lifespan. Furthermore, the combination of exercise with CR has no additive effect on maximal lifespan in rodents. However, alternate-day fasting with exercise is more beneficial for muscle mass than either treatment alone. In nonobese humans, exercise combined with CR has synergistic effects on insulin sensitivity and inflammation. From an evolutionary perspective, the responses to hunger and exercise are linked: when food is scarce, increased activity is required to hunt and gather.

HORMESIS

Paradoxically, the term *hormesis* describes the protective effects conferred by the exposure to low doses of stressors or toxins (or as Nietzsche stated, “What does not kill me makes me stronger”). Adaptive stress responses elicited by noxious agents (chemical, thermal, or radioactive) precondition an organism, rendering it resistant to subsequent higher and otherwise lethal doses of the same trigger. Hormetic stressors have been found to influence aging and lifespan, presumably by increasing cellular resilience to factors that might contribute to aging such as oxidative stress.

Yeast cells that have been exposed to low doses oxidative stress exhibit a marked anti-stress-like response that inhibits death following exposure to lethal doses of oxidants. During ischemic preconditioning in humans, short periods of ischemia protect the brain and the heart against a more severe deprivation of oxygen and subsequent reperfusion-induced oxidative stress. Similarly, the lifelong and periodic exposure to various stressors can inhibit or retard the aging process. Consistent with this concept, heat or mild doses of oxidative stress can lead to lifespan extension in *C. elegans*. CR can also be considered as a type of hormetic stress that results in the activation of antistress transcription factors (e.g., Rim15, Gis1, and Msn2/Msn4 in yeast, Nrf2 and FOXO in mammals) that enhance the expression of free radical–scavenging factors and heat shock proteins.

CONCLUSIONS

Clinicians need to understand aging biology to better manage and care for the elderly. Moreover, strategies based on aging biology are needed that delay aging, reduce the onset of age-related disorders, and increase health span for future generations. Dietary interventions and drugs that act on nutrient-sensing pathways are being developed and, in some cases, are already being tested in humans. Well-controlled human clinical trials have started to recapitulate the preclinical evidence of intermittent fasting on obesity, diabetes mellitus, cardiovascular disease, cancers, and neurologic disorders. While most animal studies show that intermittent fasting improves health throughout the lifespan, most human studies are focused on relatively short-term interventions over a few days or months. While intriguing, it remains to be seen whether people will be willing to maintain strict intermittent fasting regimens over long periods of time or if short-term clinical benefits can be obtained in combination with other therapeutic approaches.

FURTHER READING

Ferrucci L et al: Measuring biological aging in humans: A quest. Aging Cell 19:e13080, 2020.

Le Couteur DG, Thillainadesan J: What is an aging-related disease? An epidemiological perspective. J Gerontol A Biol Sci Med Sci 77:2168, 2022.

López-Otín C et al: Hallmarks of aging: An expanding universe. Cell 186:243278, 2023.

Morselli L et al: Autophagy mediates pharmacological life extension of spermidine and resveratrol. Aging 1:961, 2009.

489 Caring for the Geriatric Patient

Joseph G. Ouslander, Bernardo Reyes

DEMOGRAPHICS OF AGING AND ITS IMPLICATIONS FOR CARE OF OLDER ADULTS

The United States and other countries will continue to experience a rapid increase in the number of older adults who seek health care. The most rapidly growing segment of the population in the United States and many other developed countries is those older than 80 **(Fig. 489-1)**. According to the United Nations 2022 Aging Report, 1 in 6 people worldwide will be 65 years old or older by the year 2050. The sex and gender composition of the aging population around the world is also expected to change. Although females outlive males, an improvement in the survival of the oldest-old males could result in a more balanced gender distribution in the geriatric population in the future.

In high-income countries, consumption of health care resources will be most affected by the shift in the age distribution of the population over the next several decades. The World Health Organization continues to work actively to raise awareness of the changes necessary in current health care systems beyond increments in their budgets. As life span increases, efforts should continue to focus on promoting healthy aging to reduce the burden of disability in health care systems all over the world.

The geriatric population requires different approaches to care for several reasons. The wide variations seen in aging make it difficult to develop age-related guidelines for diagnosis and treatment. For example, acute illnesses are most often not treated in isolation but in the context of multiple comorbidities. Close to half of those older than 80 have three chronic conditions, and about one-third have four or more chronic conditions **(Fig. 489-2)**. Functional disabilities are prevalent **(Fig. 489-3)**, which require careful attention in the evaluation of the older patient, along with assessment of social support available for assistance when needed for independent and safe living.

Effectively caring for the geriatric population requires consideration of several key principles:

1. Aging is not a disease; normal aging changes generally do not cause symptoms but do increase susceptibility to many diseases and conditions due to diminished physiologic reserve.
2. Medical conditions are commonly multiple ("multimorbidity") and multifactorial in origin, requiring a comprehensive approach to evaluation and management.
3. Many potentially reversible and treatable conditions are underdiagnosed and underevaluated in this population, such as fall risk, urinary incontinence, and elder abuse and neglect; simple screening tools can help detect them.
4. Cognitive and affective disorders (e.g., mild cognitive impairment, dementia, depression, anxiety) are common and may be undiagnosed in early stages; simple screening tools can help detect them.
5. Iatrogenic illnesses are common, especially related to adverse drug reactions, immobility, and related deconditioning, as well as other complications.
6. Focus on functional ability and quality of life, as opposed to cure, are key goals of care.
7. Social history, social support, and patient preferences are critical to treating older people in a safe and person-centered manner.
8. Effective geriatric care requires multidisciplinary collaboration among many different disciplines.
9. Geriatric care is provided most commonly outside the hospital (e.g., outpatient clinics and offices, at home, in skilled nursing and assisted living settings), and attention to care transitions between settings is essential for effective care.
10. Ethical issues, palliative care, and end-of-life care are critical aspects of caring for the geriatric population.

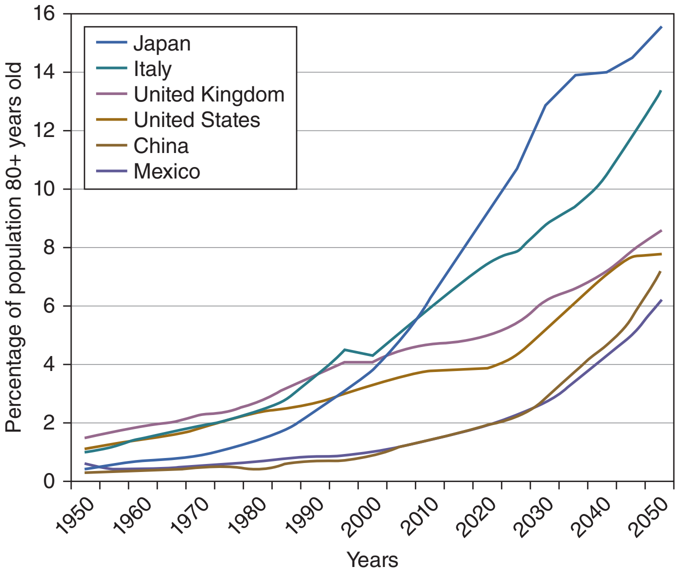

FIGURE 489-1 Percentage of the population age >80 years from 1950 to 2050 in representative nations. *(Updated data available at: https://esa.un.org/unpd/wpp/Graphs/DemographicProfiles/. Accessed December 30, 2016.)*

Another way to summarize key concepts of caring for older adults is using the "5M's of geriatrics" (mentation, medication, mobility, multicomplexity, and matters most) This framework organizes care of older adults in a person-centered manner instead of a disease-driven paradigm. The intention of the 5M's is to optimize the utilization of existing resources during the hospitalization of older adults, as well as to focus on key geriatric issues in all settings of care. At the core of the 5M's is what *matters* most to the patient in terms of goals of care and everyday living **(Fig. 489-4)**. *Mobility* is critical to individual function, quality of life, and fall risk, and ranges from the ability to move around the community to walking and transferring from a chair. Because mild cognitive impairment, dementia, delirium, and depression are all common in older adults, *mentation* is a core area for geriatric assessment. Polypharmacy and prescription of potentially inappropriate and harmful medications remain common; thus, careful *medication* reconciliation and consideration of deprescribing are core aspects of care for all older adults. Many older adults have complex clinical issues in more than one of the four M's just discussed, as well as geriatric syndromes (e.g., falls, incontinence), thus focusing attention on multiple comorbidities and *multicomplexity*. In addition to these 5M's, *social determinants of health* play a critical role in providing comprehensive care to older people.

In this chapter, these key principles serve as the background for the clinical recommendations for managing older adults. The reader is referred to textbooks of geriatric medicine for more details on each of the principles and the management of common diseases and conditions in this population.

FUNDAMENTALS OF GERIATRIC CARE

PERSON-CENTERED CARE

Person-centered care is a critical concept in caring for older people because of the complexity of their medical, functional, and psychosocial problems and, in many instances, the lack of rigorous data on the most effective strategies for caring for specific conditions in patients with multimorbidity. For almost any condition, from common disorders such as hypertension and diabetes to geriatric syndromes such as fall risk and urinary incontinence, the answer to how best to treat medical conditions in an older patient is "it depends." It depends not only on evidence-based medicine but also on careful weighing of many factors that can be important to individual older people.

The American Geriatrics Society (AGS) identifies the following elements as key to person-centered care: (1) an individualized, goal-oriented care plan based on the person's preferences; (2) ongoing review of the person's goals and care plan; (3) information sharing and

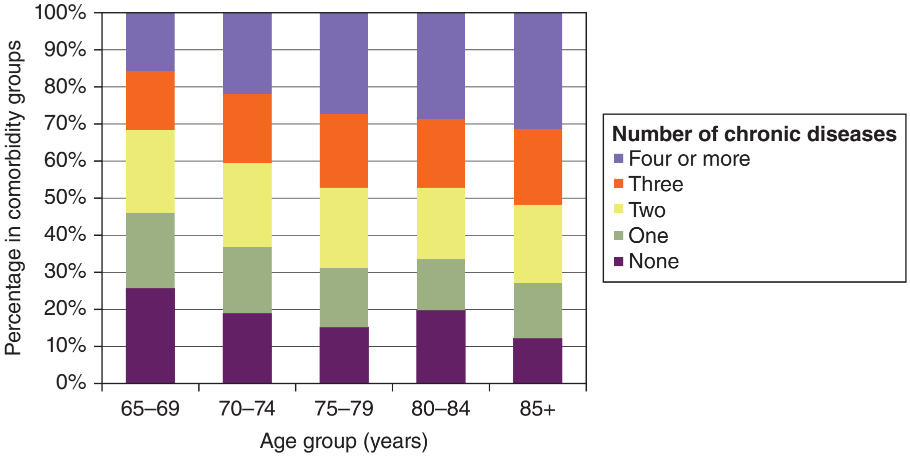

FIGURE 489-2 Prevalence of comorbidity by age group in persons ≥65 years old living in the United States and enrolled in Medicare Parts A and B in 1999. *(From JL Wolff et al: Arch Intern Med 162:2269, 2002.)*

integrated communication; (4) education and training for providers and, when appropriate, the person and those important to the person; and (5) performance measurement and quality improvement using feedback from the person and caregivers. In everyday practice with complex older patients, it is essential to have a comprehensive approach that addresses medical conditions, geriatric syndromes, improving or maintaining function and independence, quality of life, comfort, and dignity.

Several tools are available to assist in implementing evidence-based and person-centered care, including estimation of prognosis (e.g., "ePrognosis") and "Choosing Wisely" recommendations from the AGS and AMDA—The Society for Post-Acute and Long-Term Care Medicine. Examples of these recommendations that are relevant to internal medicine practice are illustrated in **Table 489-1**.

COMPREHENSIVE GERIATRIC ASSESSMENT

General Principles of Geriatric Assessment Comprehensive geriatric assessment addresses all of the components required for Medicare's annual wellness visit and more. In addition to management of common medical conditions and syndromes in older people (see below), a series of screening questions can be useful as a "geriatric review of systems" in clinical practice because of the importance and high prevalence of functional impairments and disabilities, limited social support to assist with functional limitations, cognitive and affective disorders, and geriatric conditions that may go undetected and cause patient safety issues and complications **(Table 489-2)**. Positive responses to one or more of the screening questions for each item should prompt consideration of further assessments, many of which can be accomplished using standard and validated tools available on the Internet, such as activities of daily living scales, depression scales, sleep questionnaires, and mental status examinations.

One very useful component of the geriatric assessment is the *foot exam*. Improper footwear and foot deformities, especially painful ones, are a major cause of falling (see below). If a person is not wearing socks, this may be an indication they have difficulty putting them on due to hip arthritis or another condition. Inspection between the toes is important to detect poor hygiene and ulcers, especially in diabetics. Poorly trimmed and/or fungal-infected toenails suggest that the individual cannot trim their toenails, has no one to help them do it, and does not have access to a podiatrist or regular pedicures. Other unique aspects of comprehensive geriatric assessment are described below.

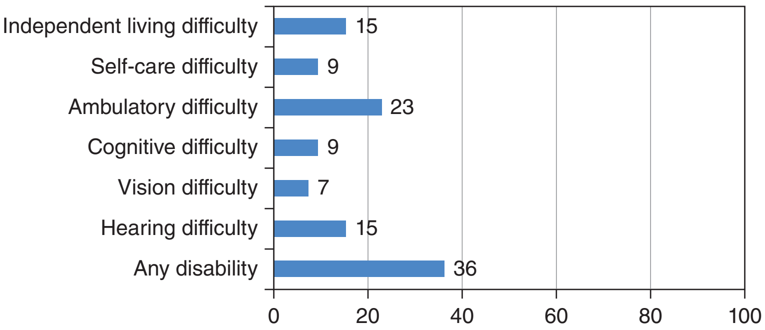

FIGURE 489-3 Percentage of people age 65+ with various disabilities. *(Source: U.S. Census Bureau, American Community Survey, 2013. Available at https://aoa.acl.gov/Aging_Statistics/Profile/2014/index.aspx. Accessed December 30, 2016.)*

Evaluation of Medical Decision-Making Capacity Key aspects of decision-making in older adults are illustrated in **Fig. 489-5**. Including the patient in the consent process for any treatment is the foundation of patient autonomy and person-centered care. Because aging is associated with an increasing potential to develop cognitive impairment, the determination of decision-making capacity is important not only to protect the patients against potential abuse but also to preserve autonomy when possible and, when it is not, to ensure an appropriate surrogate decision-making process is followed. Assessing for capacity is usually triggered by specific circumstances (e.g., the need for invasive diagnostic testing or surgery). Determination of decision-making capacity limited to medical circumstances should be differentiated from declaring a patient "incompetent" to make all decisions. Declaring someone incompetent is a legal definition and usually is reserved for court settings. Another caveat about evaluating decision-making capacity is distinguishing lack of capacity from poorly presented information, sensory impairment, language barriers, and/or low level of literacy. Diversity in values and beliefs can influence the perception of capacity for different cultures. Moreover, clinicians should be aware of the possibility of undue influence where caregivers might have secondary gain (financial or other) from taking control over decision-making on behalf of an older adult. The clinician should corroborate that the patient has received all the necessary information, comprehends the information provided, and has no major auditory or visual impairments. For geriatric patients, it is important to determine if the patient uses hearing aids or prescription glasses and ensure they are available for their use. Processing speed can slow with age and should be considered while completing a capacity assessment. Lastly, an important determinant of capacity is the presence of acute illness that could produce a transient alteration of cognition. Therefore, if the decision-making is deferred to a surrogate, this should be limited to procedures and treatments that are needed urgently.

Standard tests of cognitive function correlate poorly with the capacity to consent for specific interventions. Several standardized tools have been validated to determine decision-making capacity. The MacArthur Competence Assessment Tool–Treatment (MacCAT-T) is one of the several resources for determining decision-making capacity. It is a structured tool that has been validated, but it is lengthy and can be difficult to administer in some patients. The Capacity to Consent to Treatment Instrument (CCTI) is another tool that has been validated in patients with mild to moderate Alzheimer's disease.

Evaluation of the Older Driver For many older adults in the United States, driving is essential for maintaining independence, and driving cessation is associated with negative outcomes including social isolation and depression. On the other hand, older adults are at higher risk of being involved in fatal crashes than younger counterparts, with up to a ninefold higher risk for those ≥85 years old. Older people should be routinely assessed for their driving status and whether they have been involved in motor vehicle crashes in addition to assessment for sensory, functional, and cognitive impairments that can make driving unsafe **(Table 489-3)**. In

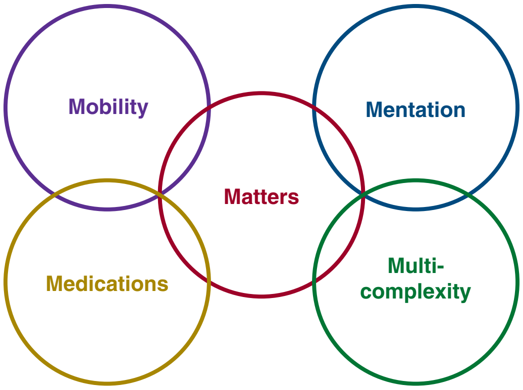

FIGURE 489-4 The 5M's of geriatrics.

addition to common geriatric conditions, several different types of drugs can impair various aspects of driving performance and should be carefully considered in older people who continue to drive, including antianxiety agents, narcotic analgesics, antipsychotics, anticonvulsants, and drugs with strong anticholinergic properties.

Suspected driving impairment can be a source of conflict between the patient (who wants to maintain independence), the family (who may want their relative to continue driving due to lack of other transportation, or may be concerned about their safety, or both), and the physician (who is concerned about all who are at risk from a car crash). These decisions involve liability since local governments might not require driving retesting for all older drivers, but in some states, physicians are required to report older people who they believe are unsafe drivers. Evaluation of driving should be aimed to correct any reversible cause of losing driving skills, such as vision and hearing impairment. Although tests of executive function such as the Trails B have been associated with poor driving performance, no single screening test predicts unsafe driving. A combination of neuropsychological testing by a psychologist and on-road testing by a trained occupational therapist can provide the physician with essential input in making the difficult decision on driving cessation. The AGS and the U.S. Department of Transportation's National Highway Traffic Safety Administration have updated the "Physician's Guide to Assessing and Counseling Older Drivers," which can be helpful to practicing clinicians and is available on the AGS website.

TABLE 489-1 Examples of Choosing Wisely Recommendations Helpful in Implementing Person-Centered Care in Complex Geriatric Patients

- Don't recommend percutaneous feeding tubes in patients with advanced dementia; instead, offer oral assisted feeding.
- Don't use antipsychotics as the first choice to treat behavioral and psychological symptoms of dementia.
- Avoid using medications other than metformin to achieve hemoglobin A_{1c} <7.5% in most older adults; moderate control is generally better.
- Don't use benzodiazepines or other sedative-hypnotics in older adults as first choice for insomnia, agitation, or delirium.
- Don't use antimicrobials to treat bacteriuria in older adults unless specific urinary tract symptoms are present.
- Don't prescribe cholinesterase inhibitors for dementia without periodic assessment for perceived cognitive benefits and adverse gastrointestinal effects.
- Don't recommend screening for breast, colorectal, prostate, or lung cancer without considering life expectancy and the risks of testing, overdiagnosis, and overtreatment.
- Don't routinely prescribe lipid-lowering medications in individuals with a limited life expectancy.
- Don't obtain a *Clostridioides difficile* toxin test to confirm "cure" if symptoms have resolved.
- Don't recommend aggressive or hospital-level care for a frail elder without a clear understanding of the individual's goals of care and the possible benefits and burdens.

Source: Adapted from *http://www.choosingwisely.org/societies/american-geriatrics-society/* and *http://www.choosingwisely.org/societies/amda-the-society-for-post-acute-and-long-term-care-medicine/amda-choosing-wisely-list/*. Accessed June 1, 2021.

Interpretation of Diagnostic Tests Physiologic changes associated with aging can affect the results of some common diagnostic tests. The large variation of many physiologic measures that are associated with normal aging makes establishing what is "normal" for many tests challenging. For this reason, the results of several diagnostic tests must be interpreted cautiously. Ambulatory cardiac monitoring may identify a variety of arrhythmias in older adults. Such arrhythmias must be linked to symptoms or adverse outcomes if left untreated before considering the use of potentially toxic medications or invasive procedures.

Advanced imaging also could demonstrate incidental abnormalities. Although a significant portion of these findings are benign, the rate of malignancy among incidental findings in the colon and extracolonic structures, as well as ovarian and thyroid gland, is ~20%. Musculoskeletal imaging, such as magnetic resonance imaging (MRI) of the spine, may reveal multiple abnormalities that may or may not be related to symptoms.

Unless further evaluation would not lead to a change in the goals of care and treatment plan, abnormal diagnostic tests require further evaluation in older patients. Examples include low hemoglobin levels, abnormal thyroid function tests, age-/sex-/weight-adjusted creatinine clearance, and elevated liver function tests. None of these abnormal findings are the result of normal aging and generally indicate the presence of a disorder that may or may not be reversible.

PREVENTION IN OLDER ADULTS

Age-Appropriate Screening Screening tests for specific diseases, as opposed to screening for geriatric conditions, require a careful person-centered approach. The focus of preventive medicine depends heavily on the ability to identify those who are at risk for specific conditions, as well as the time to benefit from interventions when considering the individual's life expectancy **(Chap. 6)**. Several professional societies have provided guidance regarding specific tests in older adults (Table 489-3). An important caveat about screening to prevent disease in older patients (e.g., colonoscopy for colon cancer, Pap smears, prostate-specific antigen testing) is that abnormal results may lead to subsequent testing and treatment among individuals who will not suffer morbidity or mortality from the disease because of limited life expectancy. Thus, geriatric patients pose a significant challenge in deciding what screening tests could offer a reasonable ratio of benefit and risk as well as being cost-effective. As an example, the U.S. Preventive Services Task Force recommends colorectal cancer screening up to the age of 75 years. Due to the limited net benefit, for those between 76 and 85 years old, the recommendation is to only consider colonoscopy if they have never been screened and are healthy enough to undergo treatment if colon cancer is detected.

Vaccinations The use of vaccines in older adults is aimed at creating immunity against common infections that could lead to serious complications and rebuilding previously obtained immunity. The U.S. Centers for Disease Control and Prevention (CDC) recommends routine vaccination against influenza, pneumococcus, and shingles as they are prevalent in this age group. CDC's recommendations on vaccination for COVID-19 and its variants, as well as respiratory syncytial virus, should be reviewed regularly because of frequent updates. Other countries in Europe and Asia have similar trends in vaccinations with small variances.

Sexually Transmitted Diseases Although most sexually transmitted diseases (STDs) occur in younger people **(Chap. 141)**, some older adults do have high-risk sexual behavior. Most Americans remain sexually active in their 60s and 70s, and up to a quarter of individuals in their 80s consider themselves sexually active. Sexually active older adults may have a lower awareness of the need for safe sexual practices, such as the risks of multiple sexual partners and condom use. The incidence of STDs in older people is still relatively low. Nonetheless, patients presenting with symptoms compatible with syphilis or

TABLE 489-2 Examples of Screening Questions and Tools and Strategies for Further Evaluation of Social Support, Functional Status, Geriatric Syndromes, and Cognition and Affect

GERIATRIC ASSESSMENT DOMAINS		RECOMMENDED SCREENS	FURTHER ASSESSMENT FOR POSITIVE SCREEN
SOCIAL	**Social Support**	Do you live alone? Do you have a caregiver? Are you a caregiver?	• Consider referral to a social worker • Refer to area agency on aging
	Elder Neglect/Abuse	Do you ever feel unsafe where you live? Has anyone ever threatened or hurt you? Has anyone been taking your money without your permission?	• Consider referral to a social worker and/or adult protective services
	Advance Directives	Would you like information or forms for a power of attorney for health care? Would you like information on a living will?	• Discussion on advance directives • Physician Orders for Life-Sustaining Treatment (POLST) (or MOLST or POST)
FUNCTIONAL	**Functional Status**	Do you need assistance with shopping or finances? Do you need assistance with bathing or taking a shower?	• Instrumental Activities of Daily Living (ADL) Scale • Basic ADL Scale
	Driving	Do you still drive? If yes: While driving, have you had an accident in the past 6 months? Driving concerns by family member?	• Vision testing • Consider occupational therapy and/or formal driving evaluation
	Vision	Do you have trouble seeing, reading, or watching TV? (with glasses, if used)	• Vision testing • Consider referral for eye exam
	Hearing	Do you have difficulty hearing conversation in a quiet room? Unable to hear whisper test 6 inches away?	• Check for cerumen in ear canals and remove if impacted • Hearing Handicap Inventory • Consider audiology referral
GERIATRIC SYNDROMES	**Medications**	Do you take five or more routine medications? Do you understand the reason for each of your medications?	• Match medications with diagnoses • Consider reducing doses, stopping drugs, adherence aides, and/or consultation with a pharmacist
	Fall Risk	Have you fallen in the past year? Are you afraid of falling? Do you have trouble climbing stairs or rising from chairs?	• “Get Up and Go” test • Consider full fall assessment • Consider physical therapy evaluation • Consider home safety assessment
	Continence	Do you have any trouble with your bladder? Do you lose urine or stool when you do not want to? Do you wear pads or adult diapers?	• Consider full continence assessment • 3IQ Questionnaire (women) • AUA 7 symptom inventory (men)
	Weight Loss	Weight <100 pounds or Unintentional weight loss ≥10 pounds over 6 months?	• Assess for common risk factors for malnutrition • Consider referral to dietician for nutritional evaluation
	Sleep	Do you often feel sleepy during the day? Do you have difficulty falling asleep at night?	• Epworth Sleepiness Scale or Pittsburgh Sleep Index • Consider referral for sleep evaluation
	Pain	Are you experiencing pain or discomfort?	• Pain assessment
	Alcohol Abuse	Do you drink >2 drinks/day?	• AUDIT-C
COGNITION AND AFFECT	**Depression**	Do you often feel sad or depressed? Have you lost pleasure in doing things over the past few months?	• PHQ-9 or Geriatric Depression Scale • Screen for suicide risk
	Cognition	Self-reported memory loss? Cognitive screen positive? (three-item recall and Clock Draw test “Mini-Cog”) Confusion Assessment (CAM) for delirium	• Montreal Cognitive Assessment or Mini Mental State Examination • If diagnosis is unclear, consider neuropsychological testing

Abbreviations: 3IQ, Three Incontinence Questions; AUA, American Urological Association; AUDIT-C, Alcohol Use Disorders Identification Test; MOLST, Medical Orders for Life-Sustaining Treatment; PHQ, Patient Health Questionnaire; POST, Physician Orders for Scope of Treatment.

Source: Adapted from RL Kane et al (eds): *Essentials of Clinical Geriatrics*, 8th ed. New York, McGraw-Hill, 2017.

gonococcal infection (cervicitis, urethritis, proctitis, epididymitis) should be screened for high-risk sexual behavior and educated if necessary. Clinical symptoms of herpes simplex infection and the possibility of becoming contagious also decrease with age. As ulcerative lesions are less frequent, herpes simplex virus 2–specific serologic testing should be considered for patients with recurrent nonspecific genital symptoms. Therapy should not be started unless the patients are symptomatic.

In the United States alone, 450 per 100,000 persons above the age of 55 are infected with HIV. Since the introduction of highly active antiretroviral therapy, the life expectancy of patients with HIV has increased, resulting in a significant increase in the number of older adults living with the disease. De novo infections have also contributed to the rising number of HIV cases in older adults. See **Chap. 208** for details on HIV infection.

TREATMENT OF COMMON DISEASES IN THE GERIATRIC POPULATION

HYPERTENSION

In the United States, 70% of older adults have hypertension. Several clinical trials have demonstrated the benefits of hypertension treatment on risk reduction of cardiovascular events in older people. Nonetheless, blood pressure targets remain controversial. The balance between the cardiovascular protective benefits versus the risk of treatment-related adverse events must be considered in individual patients based on their

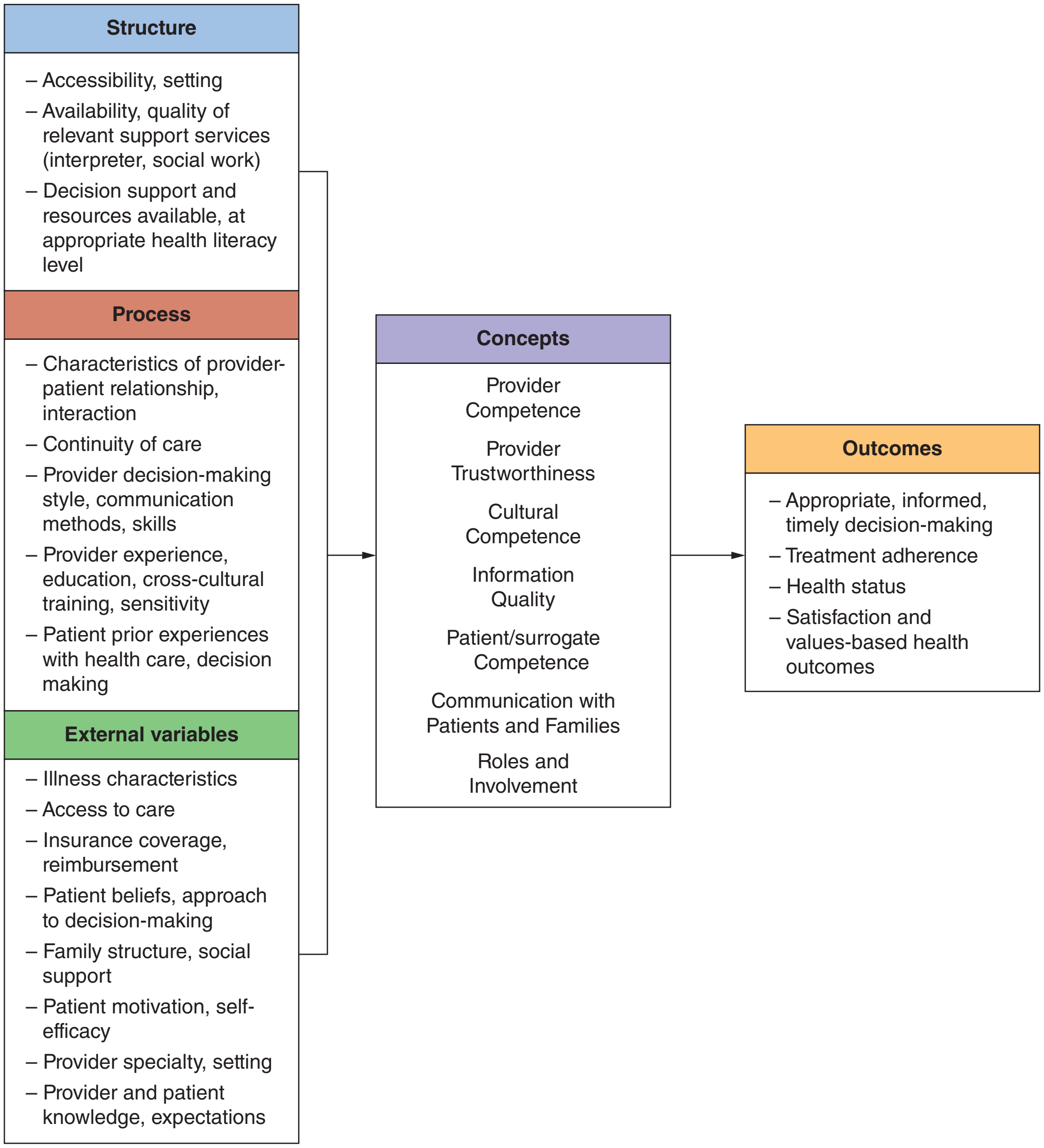
Structure
– Accessibility, setting
– Availability, quality of relevant support services (interpreter, social work)
– Decision support and resources available, at appropriate health literacy level
Process
– Characteristics of provider-patient relationship, interaction
– Continuity of care
– Provider decision-making style, communication methods, skills
– Provider experience, education, cross-cultural training, sensitivity
– Patient prior experiences with health care, decision making
External variables
– Illness characteristics
– Access to care
– Insurance coverage, reimbursement
– Patient beliefs, approach to decision-making
– Family structure, social support
– Patient motivation, self-efficacy
– Provider specialty, setting
– Provider and patient knowledge, expectations
Concepts
Provider Competence
Provider Trustworthiness
Cultural Competence
Information Quality
Patient/surrogate Competence
Communication with Patients and Families
Roles and Involvement
Outcomes
– Appropriate, informed, timely decision-making
– Treatment adherence
– Health status
– Satisfaction and values-based health outcomes

FIGURE 489-5 Key aspects of decision-making in older adults. *(Reproduced with permission from SM Dy, TS Purnell: Key concepts relevant to quality of complex and shared decision-making in health care: A literature review. Soc Sci Med 74:582, 2012.)*

comorbidities and level of function. For example, hypotension and postural hypotension related to antihypertensive therapy are common causes of syncope, falls, and related injuries in the geriatric population, especially those with multimorbidity. In addition to cardiovascular disease prevention, control of systolic blood pressure (SBP) may reduce the burden of white matter changes in the brain, which are associated with gait abnormalities and cognitive decline. A secondary analysis of the SPRINT trial (see below) suggests that intensive treatment of systolic blood pressure might reduce the risk of dementia and mild cognitive impairment (MCI) among those at higher risk for such conditions. The European Society of Cardiology/European Society of Hypertension guidelines recommend pharmacologic treatment for individuals 80 years old or older if SBP is 160 mmHg or higher. In contrast, the American College of Physicians recommends starting treatment if SBP is 150 mmHg or higher.

Two large studies (HYVET and SPRINT) have shed some light on these issues. HYVET was a multicenter study conducted in several countries involving ~3800 patients ≥80 years old. The study demonstrated that active treatment of hypertension with a target of ≤150 mmHg significantly reduced not only the risk of stroke and heart failure but also the mortality risk. As with other large hypertension studies like ALLHAT, a linear association was noted between blood pressure and stroke reduction. Nonetheless, in the HYVET study, this association was less prominent as age increased. SPRINT was another large randomized trial targeting lowering SBP to targets of <140 versus 120 mmHg (measured with an automated device) with a subgroup analysis in those aged 75 and older. Significant reductions were documented in the primary endpoint, which was a composite of cardiovascular disease events (including myocardial infarction, acute coronary syndrome, heart failure, stroke, or death from cardiovascular causes). However, it is critical to recognize that patients with diabetes, a history of stroke or heart failure, SBP <110 mmHg after 1 min of standing, as well as people with several other comorbidities, were excluded from the SPRINT trial, and aggressive treatment in the setting of these comorbidities may incur more risk of adverse effects. In addition, the reduction in adverse outcomes detected during the study was not observed long term.

Overall, these data strongly suggest a person-centered approach to hypertension in the heterogeneous older population. For older patients with minimal comorbidity, no postural hypotension, and low risk of falls and volume depletion, the benefit-risk ratio favors lower targets for SBP (<130 mmHg measured by a hand sphygmomanometer). Aggressive targets also may be more beneficial for patients with concomitant nonvalvular atrial fibrillation or coronary artery disease. However, for those with diabetes, heart failure, or postural hypotension, treatment of hypertension with higher SBP targets (e.g., <150 mmHg)

TABLE 489-3 Recommendations for Primary Prevention Screening for Specific Diseases in Older Adults from Different Professional Societies

TYPE OF SCREENING	TEST	FREQUENCY	PROFESSIONAL SOCIETY ISSUING RECOMMENDATIONS		
			USPSTF[a]	**ACS**[b]	**ACP**[c]
Colorectal	Fecal occult blood test or fecal immunochemical test (FIT) or Sigmoidoscopy or Colonoscopy	Annual Every 5 y Every 10 y	Screen all adults age 50–75; prognosis may support screening individuals of age 76–85 if never screened; not recommended for adults over age 85	Screen all adults age >50; discontinuing screening is reasonable in people with severe comorbidity that would preclude treatment	Screen all adults age 50–75 People with life expectancy <10 y should not be screened
Breast	Mammography	Every 1–2 y	Biennial screening of all women age 50–74; evidence of benefits and harms is insufficient for women age >75	Annual screening starting at age 40; continue while in good health	**ECOG**[d] Annual screening starting at age 40 y
Cervical	Pap smear HPV test	Pap only, every 3 y HPV + Pap, every 5 y	Screen women age 21–65; discontinue at age 65 if adequate prior screening	Screen women age 21–65; discontinue at age 65; discontinue at age 65 if regular screening normal	Screening should stop at age 65 if evidence of negative adequate prior screening
Lung	Low-dose CT scan	Annual	Screen age 55–80 current and former smokers with a 30+ pack-year smoking history; discontinue screening once a person has not smoked for 15 years or develops a health problem that limits their ability or willingness to have curative surgery	Screen 55–74-y-old current and former smokers in good health with a 30+ pack-year smoking history	**ACCP**[e] In settings that can deliver the comprehensive care provided to National Lung Screening Trial participants, offer screening to people age 55–74 who are current and former smokers with 30+ pack-year smoking history
Prostate	Prostate-specific antigen (PSA)	1–2 y	Do not screen men for prostate cancer with PSA if age 70 y or older	Screen men age 50 and over with a life expectancy >10 y after discussion about the risks, benefits, and uncertainties of PSA screening Follow-up screening should occur annually if PSA >2.5 ng/mL or biennially if PSA <2.5 ng/mL	**AUA**[f] Biennial PSA screening in men age 55–69 y with life expectancy >10–15 y, after shared decision-making discussions accounting for values and preferences
Osteoporosis	Dual-energy x-ray absorptiometry (DEXA) Measure height, preferably with a wall-mounted stadiometer	Perform bone mineral density testing 1–2 y after initiating medical therapy for osteoporosis and every 2 y annually thereafter	**USPSTF**[a] Screen women age 65; the current evidence is insufficient to assess the balance of benefits and harms of screening for osteoporosis to prevent osteoporotic fractures in men	**NOF**[g] Screen women age 65 and older and men age 70 and older; postmenopausal women and men age 50–69, based on risk factor profile; postmenopausal women and men age 50 and older who have had an adult-age fracture	
Carotid disease	Carotid ultrasound	Once	**SOCIETY OF VASCULAR SURGERY**		
			Age over 65, coronary artery disease, need for coronary bypass, symptomatic lower extremity arterial occlusive disease, history of tobacco use, and high cholesterol would be appropriate risk factors to prompt ultrasound in patients with a bruit		
Coronary artery disease (CAD)	Coronary calcium score (CCS)	Once	**SCCT**[h] Do not use CCS for patients with known CAD	**AHA/ACC**[i] CCS of 0 may have a strong negative predictive value for coronary events in older adults	
Abdominal aortic aneurysm	Abdominal ultrasound	Once	**USPSTF**[a] The USPSTF recommends one-time screening for abdominal aortic aneurysm with ultrasonography in men aged 65–75 y who have ever smoked; there is insufficient evidence to recommend screening for women even if they have ever smoked	**AAFP**[j] Recommended for men aged 65–75 y who have ever smoked	
Diabetes	Fasting blood glucose, glucose tolerance test, or hemoglobin A_{1c}	Annually	**USPSTF**[a] No evidence to screen for diabetes after the age of 70; recommendation being reviewed at the time of this publication	**ADA**[k] Screen people 45 y and older	

[a]U.S. Prevention Services Task Force. [b]American Cancer Society. [c]American College of Physicians. [d]Eastern Cooperative Oncology Group. [e]American College of Chest Physicians. [f]American Urology Association. [g]National Osteoporosis Foundation. [h]Society of Computed Tomography. [i]American Heart Association/American College of Cardiology. [j]American Academy of Family Physicians. [k]American Diabetes Association.

Abbreviation: CT, computed tomography.

is probably a safer approach. In addition to diet, exercise, and sodium restriction, specific pharmacologic approaches most often depend on co-occurring conditions, such as coronary artery disease, heart failure, and atrial fibrillation **(see Chap. 288)**.

DIABETES

The prevalence of type 2 diabetes in the older adult population is now >25% and is expected to increase due to adverse lifestyle changes and an increased incidence of obesity. In addition, advances in the care for type 1 diabetes have increased the number of individuals living beyond the age of 64 with this chronic illness.

Those between the ages of 65 and 74 have the highest rates of complications associated with diabetes. However, older diabetic patients are at significant risk of hypoglycemia because of potential medication errors, progressive renal insufficiency, and inconsistent oral intake. Diabetic patients age 75 or older have twice the risk of visiting the emergency department due to hypoglycemia. Hypoglycemic episodes are associated with progressive cognitive decline in older adults, especially those with existing cognitive impairment. On the other hand, chronically uncontrolled diabetes is associated with an increased risk of all-cause dementia.

Data from randomized clinical trials suggest that intensive glycemic control does not reduce major macrovascular events in older adults. Thus, the AGS guideline on diabetes in older adults (see "Further Readings") and the Choosing Wisely recommendations (Table 489-1) suggest that, in most older adults, the harms associated with a hemoglobin A$_{1c}$ (HbA$_{1c}$) target <7.5% are likely to outweigh the benefits. These recommendations are consistent with the American Diabetes Association guidelines from 2023 that recommend an HbA$_{1c}$ target of 7 or 7.5% for older adults with intact cognition and functional capacity and few comorbidities. The goals of treating diabetes in the geriatric population should be tailored to the patient's functional status, coexisting geriatric syndromes, social support, personal goals, perception of risk, and life expectancy. Insulin should be avoided when possible, and in particular, long-term use of sliding-scale insulin should be avoided because of the relatively high incidence of hypoglycemia, the challenges and discomforts of repeated glucose measurements, and the high burden it can place on caregivers. For those who do requires insulin, the use of continuous glucose monitoring may be useful.

For specifics of treatment options, see **Table 489-4 and Chap. 416**. Regardless of the therapeutic goals for HbA$_{1c}$, older diabetic patients should be regularly examined for neuropathy and retinopathy. Neuropathy can lead to unhealing foot lesions and other podiatric conditions. Retinopathy can lead to vision loss and require ophthalmologic intervention. In addition, lifestyle management is an important component of the plan of care. Diabetic older adults should exercise regularly to the extent possible and should have an adequate protein intake to try to maintain muscle mass.

TABLE 489-4 Recommendations and Considerations for Pharmacologic Therapy of Diabetes in Older Adults

MEDICATION	RECOMMENDATIONS AND CONSIDERATIONS
Metformin	• Metformin is the first-line agent for older adults with type 2 diabetes • Low risk of hypoglycemia • Recent studies suggest it may be used safely in patients with estimated glomerular filtration rate ≥30 mL/min/1.73 m^2 • Use with caution in patients with advanced renal insufficiency or significant heart failure • Can cause gastrointestinal symptoms with lower appetite
Sodium-glucose cotransporter 2 (SGLT-2) inhibitors	• Offer an oral route, which may be convenient for older adults • Low risk of hypoglycemia and hypotension • Strong evidence for renal protection and improved outcomes among patients with congestive heart failure • Caution when combined with other therapies such as diuretics
Thiazolidinediones	• If used at all, should be used very cautiously in those with, or at risk for, congestive heart failure and those at risk for falls or fractures
Sulfonylureas	• Associated with hypoglycemia and should be used with caution • Shorter-duration sulfonylureas such as glipizide are preferred • Leads to severe and persistent hypoglycemia in the settings of acute kidney injury • Glyburide is longer duration and contraindicated in older adults
Dipeptidyl peptidase 4 (DPP-4) inhibitors	• Few side effects and minimal hypoglycemia, but costs may be a barrier • No evidence of increase in major adverse cardiovascular events
Glucagon-like peptide 1 (GLP-1) agonist	• GLP-1 receptor agonists are injectable, which requires visual, motor, and cognitive skills • Associated with nausea, vomiting, diarrhea, and weight loss, which may not be desirable in some older patients, particularly those with cachexia
Insulin therapy	• Use oral medications whenever possible • Administration requires that patients or their caregivers have good visual and motor skills and cognitive ability • Insulin doses should be titrated to meet individualized glycemic targets and to avoid hypoglycemia • Once-daily basal insulin injection therapy is associated with minimal side effects and may be a reasonable option in many older patients • Multiple daily injections of insulin may be too complex for the older patient with advanced diabetes complications, life-limiting comorbid illnesses, or limited functional status • Avoid sliding scale in postacute and long-term care facilities except during periods of instability associated with acute illness; risk of hypoglycemia is high; glucose checks, insulin administration, and documentation require excessive staff time, and most patients do not like frequent fingersticks

Source: Based on recommendations from the American Diabetes Association 2020 and JS Custódio et al: Drugs Aging 37:399, 2020.

HYPERLIPIDEMIA

While good evidence exists regarding the benefits of statins for secondary prevention among those ≥75 years old, the use of these medications for primary prevention has been controversial. The use of statins in those older than 75 or 80 for primary prevention of cardiovascular events and mortality is the subject of ongoing debate in the geriatric literature. Secondary analyses of large clinical trials have demonstrated primary prevention and disability-free survival benefits with statin use among selected subgroups of older adults. Two other factors make the use of statins in older adults controversial. First, the major benefits have been demonstrated over long-term use; thus, life expectancy is a limiting factor for meaningful change in outcomes. Second, although many older adults on statins complain of muscle pain, the risk of myositis and rhabdomyolysis is increased mostly with the coexistence of other risk factors.

On the other hand, statins are safe to use in older adults, especially at moderate to low doses. Adverse effects of statins on cognitive function appear to be uncommon. Relatively healthy adults older than 75 with a life expectancy of >10 years may benefit from statin use, and the approach to hyperlipidemia should be person-centered in this population, as discussed for both hypertension and diabetes. In addition, current scientific evidence supports the use of moderate- to high-dose statins in older adults after an acute cardiovascular event (myocardial infarction, stroke) to prevent a recurrence. Even among cognitively and functionally impaired adults, another event could cause further impairment that interferes with function and quality of life. This is not true, however, for individuals at the end of life. A substantial proportion of older people (close to a third) are maintained on statins at the end of life. In a pragmatic clinical trial including older adults with a life expectancy of up to 1 year taking statin discontinuation was associated with improved quality of life and reduction of polypharmacy burden without increased risk of cardiovascular complications. Thus, discontinuation of statins in this situation is strongly recommended.

OSTEOARTHRITIS AND CHRONIC PAIN

By far, the most common cause of chronic pain in older people is osteoarthritis (OA). The approach to the management of symptomatic OA in older people differs from the approach in younger patients **(Chaps. 382 and 383)** because of the substantial toxicity of nonsteroidal anti-inflammatory drugs (NSAIDs) in this population. *Nonpharmacologic interventions, briefly discussed below, should be the first line of treatment*. While some patients older than 65 can tolerate NSAID use with concomitant protection from gastrointestinal (GI) bleeding with a proton pump inhibitor (PPI), this regimen exposes patients to two drugs with numerous potential adverse drug effects. NSAIDs are well known to be associated not only with GI bleeding but also with worsening renal function as well as fluid retention, hypertension, and exacerbation congestive heart failure. In addition, a substantial number of older patients are on anticoagulants or platelet aggregation inhibitors, which could further increase the risk of bleeding from NSAIDs. PPIs are associated with a higher incidence of pneumonia, osteoporosis, and *Clostridioides difficile*–associated diarrhea, and they may be associated with a higher risk of dementia.

Thus, in older patients with multimorbidity who have painful OA, the risks of NSAIDs most often outweigh the benefits, and older patients should be discouraged from taking nonprescription NSAIDs without consulting their primary care clinician. Topical NSAIDs are better tolerated, and 4 or 5% lidocaine patches and other nonprescription analgesic creams (e.g., 1% diclofenac) may also be effective. Acetaminophen in doses of 1 g 2–3 times daily should be the basis of pharmacologic treatment. Failure to respond could be followed up with careful trials of tramadol or a narcotic agent (started in a short-acting preparation) with appropriate attention to avoiding narcotic-induced constipation. Tramadol has less addiction potential and side effects than more potent narcotics and can be very effective for some older people. However, it is contraindicated in patients with seizure disorders and must be used cautiously with selective serotonin reuptake inhibitor and serotonin-norepinephrine reuptake inhibitor drugs to prevent serotonergic toxicity. Another pharmacologic option that can be very useful in older people with primary or secondary muscular pain due to an arthritic condition, especially those with dysphoria or depression (common in people with chronic pain), is a trial of duloxetine. Duloxetine is approved for the treatment of depression as well as pain, but it must be started in low doses (e.g., 20 mg) and titrated up to a maximum of 60 mg, if needed, over several weeks.

Although the prescription of narcotics is getting increasingly cumbersome because of high rates of abuse, this should not deter the prescription of these agents to relieve pain and disability in older patients, especially postsurgical pain. If opioids are prescribed, a bowel regimen consisting of adequate fluid, a bulk-forming agent, and use a stimulant such as senna or polyethylene glycol is highly recommended for people who have not had a movement to avoid severe constipation. Note that although docusate is often prescribed, its efficacy is highly variable. In addition, as in any other population, there should be close monitoring of issues such as diversion and abuse or addiction.

Many older patients respond well to various nonpharmacologic interventions, including stretching, strengthening, timely and appropriate use of heat and ice, massage (including the use of “massage guns”), swimming and whirlpool therapy, bracing, acupuncture, and therapeutic electrical stimulation These interventions are best carried out under the supervision of physical therapists or other professionals with appropriate expertise to avoid injury. Although now frequently offered by telehealth, they will not be optimally effective unless a trained therapist teaches the technique and assesses its proper use periodically in person. Surgical interventions, including replacement of major joints, have improved over the past several years, and even older patients with multimorbidity may benefit in terms of function and quality of life. Total knee replacement, for example, is effective in generally healthy older patients and should be considered in selected higher-risk patients. “Pre-habilitation,” with targeted strengthening and endurance exercises, and willingness to go through several weeks of postoperative physical therapy should be prerequisites for referring older patients for joint replacement.

CANCER

More than half of new cases of cancer and mortality associated with it occur after the age of 65. Data regarding older adults with multiple comorbid conditions and their response to cancer treatment are limited. While only ~10% of clinical trials have had age-stratification analyses, the available evidence suggests that age alone is not a predictor of harm. Nonetheless, making treatment decisions is challenging due to both shorter life expectancy in older adults and the cumulative effect of multiple comorbidities. Thus, a person-centered approach that integrates the principles of geriatric and palliative care is essential.

Older adults generally experience decreases in functional status after receiving chemotherapy. Most of this negative effect appears to be related to comorbidity and baseline functional status, rather than due to age alone. For this reason, specialists in geriatric oncology have proposed using comprehensive geriatric assessment, covering many of the issues addressed in Table 489-3, as a strategy to better predict which older adults will tolerate and benefit most from cancer treatment. Tools such as the Chemotherapy Risk Assessment Scale for High-Age Patients have been validated to estimate the risk of chemotoxicity. Lack of social support has been associated with poor outcomes after radiation and chemotherapy, especially in older women. Other important issues in cancer treatment planning include the availability of transportation for treatments, economic and insurance status, the patient’s ability to follow treatment plans, and family and social support available during therapy, when adverse effects and functional decline may occur.

Newer treatment options such as immunotherapy have shown promising results in the general population with lung and bladder cancer as well as lymphomas. Unfortunately, older adults are underrepresented in clinical trials using these agents, and there is a lack of robust evidence regarding tolerability and efficacy.

ANEMIA

Low hemoglobin or hematocrit is common in older adults but is not a normal age-related change. All anemic older adults should have a basic evaluation including a complete blood count, examination of a peripheral red blood cell smear, reticulocyte count, and measurement of iron, iron-binding capacity, and transferrin saturation. A serum ferritin level can help distinguish iron deficiency anemia from the anemia of chronic disease, two common forms of anemia in older adults. Iron deficiency is the most common cause of anemia in the older population, with other nutritional anemias (e.g., B_{12} deficiency) and myelodysplasia each accounting for a smaller percentage. Anemia of chronic disease is common in older people who have several chronic illnesses. The etiology of the anemia in older adults cannot be specifically explained in more than a third of the cases, and this unexplained anemia is generally normocytic, mild in degree, with a low reticulocyte count, and associated with normal or low erythropoietin levels in the face of inadequate production of new red cells. Anemia is frequently asymptomatic, but severe cases could present with symptoms such as generalized weakness and functional decline, shortness of breath, chest pain, or syncope. The unexplained anemia of aging appears to be responsive to erythropoietin, but it is unclear whether correction of the anemia improves outcomes. Thresholds for transfusion of packed red cells among older adults should be based on symptoms and associated conditions. For example, for geriatric patients suffering acute blood loss anemia after an orthopedic procedure, transfusion is indicated when hemoglobin is <8 mg/dL (instead of 7 mg/dL). Similarly, older patients with active cardiovascular disease, such as angina or heart failure, may need to be transfused at levels <8 or 9 mg/dL. For details of the general evaluation and management of anemia, please refer to **Chap. 66**.

HEART FAILURE

Although the incidence of heart failure (HF) has decreased over the years, the prevalence has seen an uptrend. This paradoxical epidemiologic trend is explained by longer survival. As many as 20% of individuals ≥80 years old carry a diagnosis of HF, with two-thirds of those having preserved ejection fraction. The management of HF is discussed in detail in Chaps. 264 and 265. The treatment of HF in frail older adults depends on cardiac ejection fraction (preserved vs reduced) and is limited by tolerability of low-sodium diets, fluid restriction, and adverse effects of medication. Medications such as diuretics, beta-blockers, angiotensin-converting enzyme inhibitors, and angiotensin receptor blockers may be poorly tolerated due to hypotension, electrolyte imbalances, volume depletion, and orthostatic hypotension. Newer therapies such as angiotensin receptor–neprilysin inhibitors, mineralocorticoid receptor antagonists, and sodium-glucose cotransporter 2 inhibitors (SGLT2i) have been shown to improve outcomes in older adults with specific types of HF. Thus, echocardiography to check ejection fraction is essential in the management of older people with HF. In contrast with other chronic disease processes, older adults have been well represented in clinical trials with these newer agents, allowing multiple subanalyses that have proven efficacy.

GERIATRIC SYNDROMES AND CONDITIONS

In this section, selected geriatric syndromes and conditions likely to be encountered by internists in hospitals, clinics, and postacute and long-term care (LTC) settings are discussed. For a more thorough discussion of these and other syndromes and conditions, the reader is referred to textbooks that focus specifically on geriatrics and gerontology (see "Further Reading").

■ FALLS

Epidemiology and Impact Among all geriatric syndromes, falls are probably the most common that internists will encounter. Falls are responsible for potentially devastating consequences for function and quality of life, as well as mortality. About one in three older community-dwelling and one in two older LTC facility residents fall annually, with many more at risk for falls. The consequences of falls include fear of falling with adverse effects on quality of life, painful injuries including hip and wrist fractures, subdural hematomas, and death. Falls are associated with loss of function and death within the year after a fall. With the increasing prescription of anticoagulants, falling presents a frequent conundrum in balancing the risks and benefits of anticoagulation. For these reasons, internists should regularly screen older people for falling using questions such as, "Have you fallen in the past year?" "Are you afraid of falling?" "Do you have trouble climbing stairs or rising from chairs?" (Table 489-2).

Evaluation The risks and causes of falls are multifactorial. Most older people at risk for a fall or who have suffered a fall have more than one potential underlying risk factor or cause. Many falls are labeled as "mechanical" and attributed to simply tripping or slipping. It is essential to recognize, however, that older people who trip or slip may have a variety of underlying reversible conditions that could have contributed to the event. Thus, a thorough evaluation of all falls is warranted. In addition to evaluating the patient who has fallen for injury, it is critical to determine, to the extent possible, whether the patient had a syncopal episode or a seizure, which dictates a very different approach to evaluation and management. As many as half of "unexplained falls" in older people with dementia (e.g., found on the floor) may be due to near-syncope or syncope related to postural hypotension.

Figure 489-6 illustrates an overview of the approach to an older person who reports a history of one or more falls in the past 6 months, and Table 489-5 provides more detail on the immediate evaluation of an older person who has fallen. Chap. 28 provides more detail on the evaluation of gait and balance disorders.

Management Table 489-6 illustrates approaches to the management of falls. Immediately after a fall, injuries and underlying acute illnesses should be identified and treated. It is common practice for older patients who come to an emergency department with a history of a fall to have a brain imaging study. While this is understandable from a potential liability standpoint, it is also reasonable to avoid such studies if no history or signs of head trauma, neurologic symptoms or signs, or anticoagulation is noted and to monitor the patient carefully over the next 48–72 h for the development of specific indications for a brain imaging study.

Because the causes of and risk factors for falls are often multifactorial, management commonly requires multiple interventions in the same patient. Among the most common and effective interventions are physical therapy for strengthening and balance; Tai Chi has also been shown to be effective in multiple trials. Although many older people who fall are vitamin D deficient, the role of vitamin D replacement in preventing falls or preventing injuries from falls when combined with interventions such as strength and balance training remains unclear. The risk/benefit ratio probably favors vitamin D replacement with at least 800 IU per day, but high-dose vitamin D (60,000 IU in one oral dose monthly) has been associated with an increase in the risk of falls. Patients who suffer a fracture after a fall should be investigated and treated for osteoporosis. Patients at high risk for recurrent falls and injuries should be encouraged to use a fall alert system; selected patients may benefit from hip protectors.

All older patients at fall risk should have an environmental evaluation for potential hazards and home improvements and a fall-alert system that can help prevent lying on the floor for many hours with the risk of rhabdomyolysis.

In older people at high fall risk and with atrial fibrillation or flutter, the risk of life-threatening bleeding associated with anticoagulants is substantially lower than the risk of stroke. Because of the devastating effects a stroke can have on function and quality of life, anticoagulation is generally preferred. Patients with this combination should also strongly consider insertion of a left atrial appendage closure device.

■ POLYPHARMACY

Epidemiology and Impact Polypharmacy is defined as the prescription of multiple medications using various thresholds (generally ranging from five up to nine simultaneous drugs) and has been identified as a major challenge in the geriatric population for decades. About 40% of the U.S. population age 65 and older take five to nine medications, and close to 20% take 10 or more. Polypharmacy is an increasingly complex challenge because of the rising prevalence of multimorbidity, a plethora of clinical practice guidelines, the proliferation of medications that can effectively treat common geriatric conditions, and rising patient and family demand for medications due in part to television advertising and information available on the Internet. For example, based on several condition-specific clinical practice guidelines (which do not account for multimorbidity), an 80-year-old person with multimorbidity including diabetes, chronic obstructive lung disease, hypertension, osteoporosis, and degenerative joint disease might be prescribed an extremely complicated nonpharmacologic regimen and over a dozen medications with the potential for multiple drug-drug and drug-disease interactions and high out-of-pocket costs.

Polypharmacy increases the risks associated with age-related changes in the pharmacology of many drugs and the risk of adverse drug events. Such events cause >100,000 hospitalizations per year; the main culprits are warfarin and other antiplatelet agents as well as insulin and other hypoglycemic agents. Other categories of drugs are also involved, including cardiovascular drugs that can cause electrolyte and volume disturbances and hypotension, falls, and syncope; central nervous system drugs associated with altered mental status and falls; and antimicrobials, which cause allergic reactions, diarrhea, and other adverse drug effects.

Evaluation Because many older patients see multiple specialists, internists should perform careful medication reconciliation at each office or clinic visit and especially at the time of care transitions, including acute hospitalization, hospital discharge, admission to a long-term care facility or home health program, and discharge from a postacute care facility to home. At each transition, all medications should be considered in terms of unclear diagnosis or indication, uncertain dose or route of administration, stop date, hold parameters, lab tests needed for monitoring, dosages different than the last care setting, medication

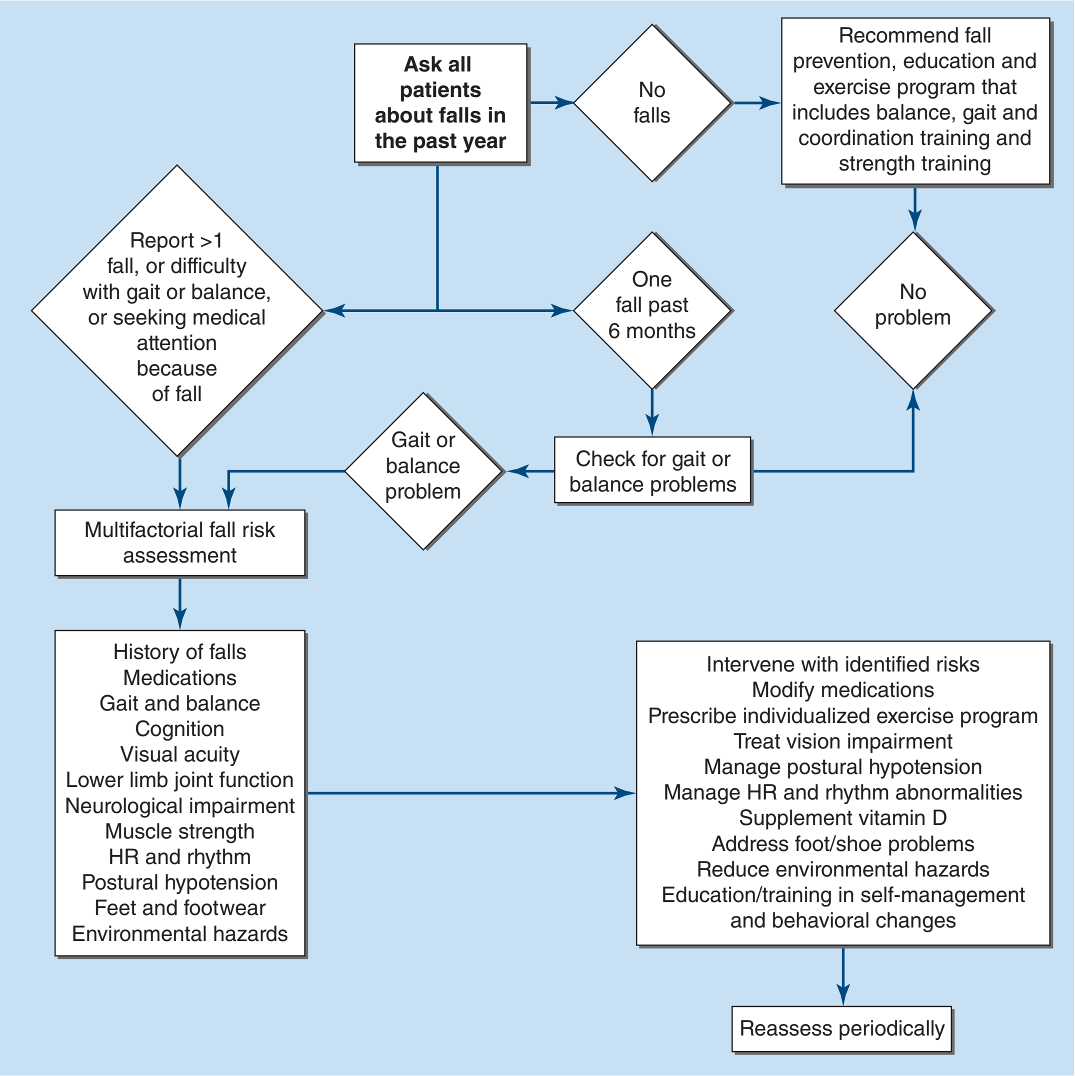
Ask all patients about falls in the past year
No falls
Recommend fall prevention, education and exercise program that includes balance, gait and coordination training and strength training
Report >1 fall, or difficulty with gait or balance, or seeking medical attention because of fall
One fall past 6 months
No problem
Gait or balance problem
Check for gait or balance problems
Multifactorial fall risk assessment
History of falls
Medications
Gait and balance
Cognition
Visual acuity
Lower limb joint function
Neurological impairment
Muscle strength
HR and rhythm
Postural hypotension
Feet and footwear
Environmental hazards
Intervene with identified risks
Modify medications
Prescribe individualized exercise program
Treat vision impairment
Manage postural hypotension
Manage HR and rhythm abnormalities
Supplement vitamin D
Address foot/shoe problems
Reduce environmental hazards
Education/training in self-management and behavioral changes
Reassess periodically

FIGURE 489-6 Algorithm depicting assessment and management of falls in older patients. HR, heart rate. *(Reproduced with permission of American Geriatrics Society. American Geriatrics Society and British Geriatrics Society: Clinical Practice Guideline for the Prevention of Falls in Older Persons. New York, American Geriatrics Society, 2010.)*

duplication, medications that should be restarted, and the potential for drug-drug and drug-disease interactions. Possible adverse drug effects, effectiveness of drug therapy, and adherence should be evaluated.

Management Internists should serve as the "quarterback" for all prescribing to help ensure adherence and minimize the potential for adverse drug effects. **Table 489-7** lists several general recommendations for geriatric prescribing that should help make drug therapy more effective and safer in older patients, especially those with multimorbidity, and **Fig. 489-7** illustrates a stepwise approach to optimizing medication management. Clinical pharmacists can be extremely helpful in achieving these recommendations and goals. **Chap. 71 also provides information on the general principles of clinical pharmacology.**

While undertreatment of certain conditions may occur in older people (e.g., osteoporosis, depression, and overactive bladder), more attention is now being paid to "deprescribing." Deprescribing must be done carefully, especially at the time of care transitions, when indications for specific drugs and patient preferences may not be clear. The AGS's updated Beers criteria include a comprehensive list of drugs that may be inappropriate in older people and the rationale for this rating. The Screening Tool of Older Persons' Prescriptions (STOPP) criteria are also useful in identifying drugs that should be reconsidered in older people.

Several commonly prescribed drugs should be considered for deprescribing efforts, including (1) diuretics and hypotensive agents when patients have systolic hypotension or postural hypotension that can precipitate near-syncope and falls; (2) overreliance on antianxiety and hypnotic medications, especially benzodiazepines; (3) psychotropic and other drugs with anticholinergic activity that can cause dry mouth and constipation and increase the long-term risk of cognitive impairment; (4) PPIs with unclear indications because of numerous reported potential adverse drug effects, including increased risk of pneumonia, osteoporosis, and dementia; (5) cholinesterase inhibitors and memantine in patients with severe cognitive impairment who have been on them for years; (6) hypoglycemic agents in patients with multimorbidity who should not have tightly controlled blood sugar with increased risk of hypoglycemia; and (7) statins and prophylactic aspirin in patients with severe chronic illness who are near the end of life.

Careful deprescribing is a critical aspect of person-centered care in the geriatric population. Several general principles, including those in Table 489-7 and Figure 489-7, may assist with deprescribing efforts. The AGS's comprehensive Deprescribing Toolkit and the *deprescribing.org* website can be very helpful these efforts.

■ COGNITIVE IMPAIRMENT: DELIRIUM AND DEMENTIA

The reader is referred to other chapters in this text **(Chaps. 442–445)** on cognitive impairment. Because these conditions are so prevalent in the older population, selected aspects highly relevant to geriatrics are briefly discussed here.

Epidemiology and Impact Delirium occurs in up to 40% of hospitalized older patients and is associated with increased morbidity, length of hospital stays, need for institutional care, health care utilization, and mortality in this population. An even higher incidence of delirium can be seen in older patients with underlying cognitive

TABLE 489-5 Evaluating the Older Person Who Falls—Immediate Postfall Evaluation

History

- Circumstances surrounding the fall
  - Relationship to changes in posture, turning of head, after a meal or medication intake, rushing to the toilet, nocturia, straining to urinate or defecate
  - Accidental trip or slip (note that many correctable factors can contribute to a reported "mechanical" fall—see text)
  - Hazards in the living environment (loose rugs, cords, unsafe steps, slippery floors, etc.)
- Premonitory or associated symptoms
  - Dizziness (lightheadedness vs vertigo); cardiovascular (postural lightheadedness, palpitations, chest pain, shortness of breath); focal neurologic symptoms suggestive of stroke or transient ischemic attack (weakness, sensory disturbance, dysarthria, ataxia, aphasia); symptoms of a seizure (witnessed clinic movements, incontinence of urine or stool, tongue biting)
  - Symptoms over the previous few days that may have led to volume depletion (poor food/fluid intake, nausea/vomiting, diarrhea, urinary frequency/polyuria)
- Exclude loss of consciousness or seizure (may be difficult without a witness)
- Medications—chronic and within the few hours before the fall
  - Diuretics and other antihypertensive drugs
  - Nitrates
  - Drugs that cause bradycardia—beta blockers; cholinesterase inhibitors
  - Psychotropics—antipsychotics, hypnotics, sedatives, antidepressants
  - Antiparkinsonian drugs
  - Hypoglycemic drugs
- Excessive alcohol intake

Physical Examination

- Exclude physical injury
  - Head trauma, hip range of motion, pubic bone tenderness, wrist pain, other signs of trauma
  - Bruising in patients on anticoagulants/platelet inhibitors
- Exclude acute illness
  - Vital signs
  - Postural vital signs (if feasible/safe)
  - Fingerstick glucose in diabetics
  - Poor skin turgor suggesting volume depletion (over chest; other areas unreliable)
  - Signs of an acute respiratory, cardiovascular, or abdominal condition
  - Focal neurologic signs suggestive of stroke
- Signs of conditions that increase risk for falls
  - Poor visual acuity; use of bifocals
  - Limited range of motion of neck (to detect possible cervical arthritis/disk disease)
  - Cardiovascular—arrhythmias, carotid bruits, aortic stenosis, mitral insufficiency, heart failure
  - Degenerative joint disease in lower extremities causing pain, limited range of motion, and/or deformity
  - Podiatric conditions (calluses; bunions; ulcerations; poorly fitted, inappropriate, or unsafe shoes)
  - Neurologic signs—lower extremity muscle weakness; peripheral neuropathy; tremor, rigidity, and/or bradykinesia suggestive of undiagnosed Parkinson's disease; cerebellar signs (abnormal heel to shin or heel tapping); abnormal reflexes that could reflect upper motor neuron disorder such as spinal cord compression or subdural hematoma; cognitive deficits that can result in poor judgement
  - Observation of gait and balance—simple Get Up and Go test (see text) with observation for short steps, poor foot elevation, wide-based gait, multiple steps to turn 180 degrees; other abnormalities that might suggest normal pressure hydrocephalus (especially in combination with symptoms of incontinence and/or cognitive impairment)

Laboratory and/or Imaging Studies

- Should be guided by history and physical examination—common examples include:
  - Complete blood count, basic metabolic panel to exclude/verify acute illness
  - Urinalysis (only when additional symptoms of urinary tract infection present)
  - Electrocardiogram (in patients suspected of acute coronary syndrome or with significant known cardiovascular disease)
  - X-rays to exclude fractures
  - Brain imaging if signs present to exclude subdural hematoma, stroke
  - Cardiac monitoring in patients with history suggestive of syncope or near-syncope
  - Electroencephalography in patients with history suggestive of seizure

Source: Adapted from RL Kane et al (eds): *Essentials of Clinical Geriatrics*, 8th ed. New York, McGraw-Hill, 2018.

impairment after surgery or while in the intensive care unit. While most episodes of delirium clear within a few days if the underlying cause(s) is identified and treated, delirium may persist for weeks or, in a few cases, for months after an acute hospitalization.

Normal aging does not cause impairment of cognitive function of sufficient severity to render an individual dysfunctional, which is the hallmark of dementia syndrome. Slowed thinking and reaction time, mild recent memory loss, and impaired executive function can occur with increasing age and may or may not progress to dementia. **Figure 489-8** illustrates the prevalence of memory impairment with increasing age. Just over 20% of people over age 70 in the United States have cognitive impairment without dementia, generally referred to as MCI. Up to 15–20% of those diagnosed with MCI will progress to dementia over a year; thus, most people with MCI will progress to dementia within 5 years. Therapeutic implications of MCI are subjects of intensive research. No nonpharmacologic or pharmacologic intervention has been shown to prevent the progression of dementia.

The definitions of Alzheimer's disease and related dementias have been updated by the American Psychiatric Association. The prevalence of dementia increases with age; by age 85, between 30 and 40% of

TABLE 489-6 Examples of Management for Underlying Causes of Falls in Older Patients

CAUSES	EXAMPLES OF TREATMENT
Cardiovascular	
Arrhythmias	Antiarrhythmic medication, ablation, pacemaker (depending on nature of arrhythmia)
Aortic stenosis with syncope or near syncope	Valve surgery (transcatheter procedure if appropriate)
Postural hypotension	Reduce or eliminate hypotensive drugs Hydration, support stockings Medication (Proamatine [midodrine], fludrocortisone, droxidopa) Adaptive behaviors (e.g., pausing and getting up slowly)
Hypertension	Manage carefully to avoid hypotension and near syncope; control may be important in patients with periventricular white matter changes in preventing further gait disturbance
Neurologic	
Autonomic dysfunction with postural hypotension	As above
Cervical spondylosis (with spinal cord compression)	Neck brace; physical therapy; consider surgery
Parkinson's disease	Antiparkinsonian drugs
Visual impairment	Ophthalmologic/optometric evaluation and specific treatment
Seizure disorder	Anticonvulsants
Normal-pressure hydrocephalus	Surgery (ventricular-peritoneal shunt)
Dementia	Supervised activities Hazard-free environment
Benign positional vertigo	Habituation exercises Anti-vertiginous medication
Others	
Foot disorders	Podiatric evaluation and treatment
Gait and balance disorders	Properly fitted shoes Physical therapy Exercise with balance training (including Tai Chi where available)
Muscle weakness, deconditioning	Lower extremity strength training
Drug adverse effects (e.g., sedatives, alcohol, other psychotropic drugs, antihypertensive)	Elimination of drug(s) when feasible
Vitamin D deficiency	Vitamin D supplementation
Recurrent falls	Fall alert system for those who live alone; hip protectors in selected patients

Source: Reproduced with permission from RL Kane et al (eds): *Essentials of Clinical Geriatrics*, 8th ed. New York, McGraw-Hill, 2018.

TABLE 489-7 General Recommendations for Geriatric Prescribing

1. Evaluate geriatric patients thoroughly to identify all conditions that could (a) benefit from drug treatment; (b) be adversely affected by drug treatment; and (c) influence the efficacy of drug treatment.
2. Manage medical conditions without drugs as often as possible.
3. Know the pharmacology of the drug(s) being prescribed.
4. Consider how the clinical status (e.g., renal function, hydration) of each patient could influence the pharmacology of the drug(s).
5. Avoid potentially serious adverse drug-drug interactions.
6. For drugs or their active metabolites eliminated predominantly by the kidney, use a formula to approximate age-related changes in renal function and adjust dosages accordingly; the Cockcroft-Gault formula (below) is probably safer as it tends to underestimate creatinine clearance.

$$\text{Creatinine clearance} = \frac{(140-\text{age}) \times \text{body weight (kg)}}{72 \times \text{serum creatinine level}} (\times\ 0.85 \text{ for women})$$

7. If there is a question about drug dosage, start with smaller doses and increase gradually until the drug is effective or intolerable side effects are observed.
8. Drug blood concentrations can be helpful in monitoring several potentially toxic drugs used in the geriatric population.
9. Help to ensure adherence by:
    a. Making drug regimens and instructions as simple as possible
    b. Using the same dosage schedule for all drugs whenever feasible (e.g., once or twice per day)
    c. Timing the doses in conjunction with a daily routine
    d. Paying attention to impaired cognitive function, diminished hearing, and poor vision when instructing patients and labeling prescriptions
    e. Instructing relatives and caregivers on the drug regimen
    f. Enlisting other health professionals (e.g., home health aides, pharmacists) to help ensure compliance
    g. Making sure the older patient can get to a pharmacist (or vice versa), can afford the prescriptions, and can open the container
    h. Using aids (e.g., special pillboxes and drug calendars) whenever appropriate
    i. Performing careful medication adjudication and patient/family education at the time of every hospital discharge
    j. Keeping updated medication records and review them at each visit
    k. Reviewing knowledge of and adherence with drug regimens regularly
10. Monitor older patients frequently for adherence, drug effectiveness, and adverse effects, and adjust drug therapy accordingly.

Source: Reproduced with permission from RL Kane et al (eds): *Essentials of Clinical Geriatrics*, 8th ed. New York, McGraw-Hill, 2017.

patients have a dementia syndrome. Alzheimer's disease and vascular dementia, which often occur together based on pathologic studies, account for most dementias in older people. Dementia with Lewy bodies accounts for up to 25% of dementia and is characterized by parkinsonian features early in the disease (as opposed to dementia in Parkinson's disease, which generally occurs years after the onset of Parkinson's), personality changes, alterations in alertness and attention, and visual hallucinations that can cause paranoia. Although most dementia syndromes are slowly progressive over several years, dementia is a terminal illness among patients who do not succumb to other comorbidities and results in a devastating loss of cognition and function in the later stages.

Evaluation Regardless of setting, the new onset of delirium should be treated as a medical emergency because it can be the manifestation of an underlying critical illness. The first step in the evaluation is to identify predisposing and precipitating factors such as hearing or visual impairment, symptoms of depression, laboratory abnormalities, uncontrolled pain, infections, exacerbation of chronic illnesses, and history of alcohol or other substances use. The most validated evaluation for delirium is the Confusion Assessment Method, which requires an acute onset and fluctuating course *and* inattention *and* disorganized thinking *or* altered level of consciousness. Because the causes and risk factors for delirium are multifactorial, evaluation requires a careful history, physical examination, and selected laboratory studies based on the findings.

The benefits of screening older community-dwelling adults for cognitive impairment are controversial, but many interventions may benefit patients and families early in the course of the disease (see below). Older patients in outpatient settings with complaints (or family reports) of early signs of cognitive impairment benefit from neuropsychological testing, which can help differentiate between MCI and dementia and identify concomitant factors such as depression and anxiety. The Mini-Cog is a sensitive screening tool for cognitive impairment and consists of a three-item recall test and clock drawing. Further evaluation of dementia includes a comprehensive history and physical examination, functional status assessment (since the diagnosis depends on impaired function), a brain imaging study, and selected laboratory tests, including a complete blood count, comprehensive metabolic panel, thyroid

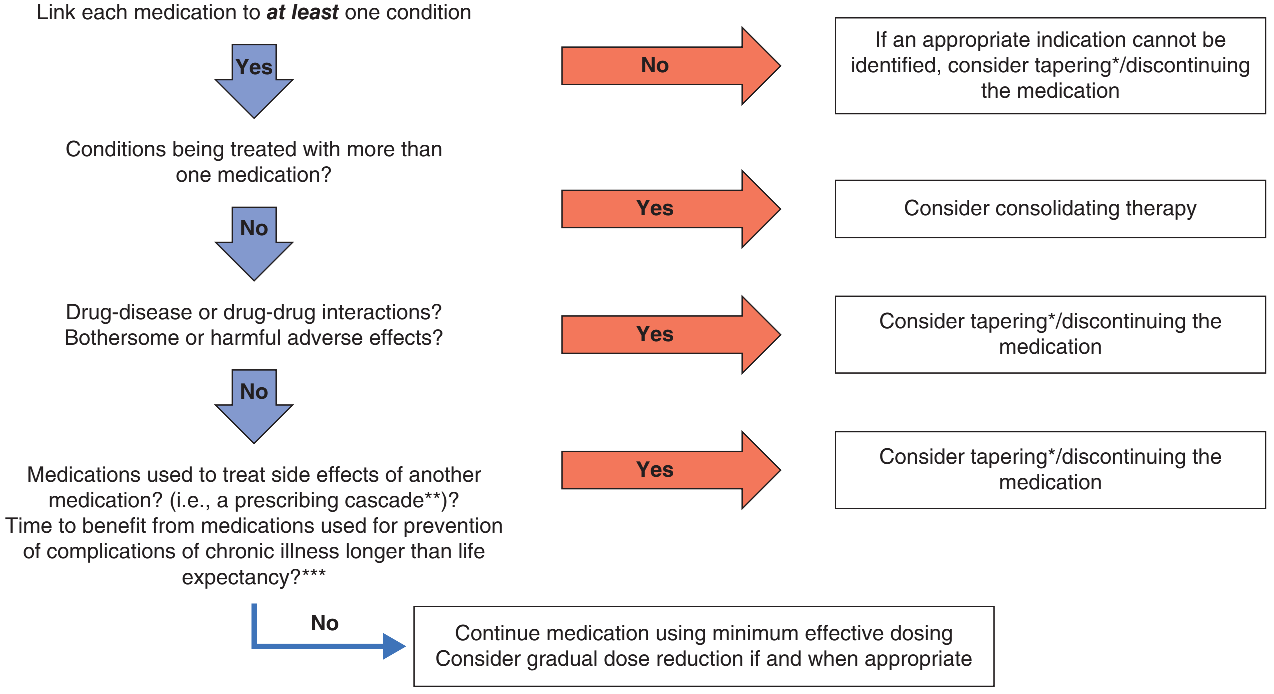
General Approach to Optimizing Medication Management
and Deprescribing in Older Adults
Link each medication to *at least* one condition
Yes
No
If an appropriate indication cannot be identified, consider tapering*/discontinuing the medication
Conditions being treated with more than one medication?
Yes
Consider consolidating therapy
No
Drug-disease or drug-drug interactions? Bothersome or harmful adverse effects?
Yes
Consider tapering*/discontinuing the medication
No
Medications used to treat side effects of another medication? (i.e., a prescribing cascade**)? Time to benefit from medications used for prevention of complications of chronic illness longer than life expectancy?***
Yes
Consider tapering*/discontinuing the medication
No
Continue medication using minimum effective dosing
Consider gradual dose reduction if and when appropriate

*Many types of drugs should be tapered to avoid withdrawal with sudden discontinuation, e.g., antidepressants, sedatives, hypnotics, analgesics.

**A common example of a prescribing cascade is treating edema associated with amlodipine with a diuretic.

***Common examples are continuing statins and inappropriate goals resulting in overtreatment of hypertension and diabetes causing hypotension and hypoglycemia in people with very limited life expectancy.

FIGURE 489-7 General approach to optimizing medication management and deprescribing in older adults. This algorithm illustrates a stepwise approach to optimizing medication management.

function tests, vitamin B_{12} level, and, if suspected, tests for syphilis and human immunodeficiency virus antibodies.

Management Table 489-8 lists pharmacologic and nonpharmacologic management strategies for various underlying risk factors and causes of delirium. Every attempt should be made to avoid or discontinue any medication that may be worsening cognitive function in a delirious geriatric patient. This may not be possible, and in some patients, psychotropic drugs may be needed to treat delirium if the patient is a danger to themselves or others and/or if delusions and hallucinations are present. Low-dose risperidone (0.25–0.5 mg) or quetiapine (12.5–50 mg at night to take advantage of sedation) may be effective; more sedating antipsychotics and benzodiazepines should be avoided unless the goal is to put the patient to sleep for a short time. If a benzodiazepine is used, it should be short-acting (e.g., lorazepam) and in a low dose; alprazolam should be avoided as many patients become dependent on it long-term. Overall, multifactorial-proactive interventions and geriatric consultation have been associated with decreased incidence and duration of delirium in the hospital setting.

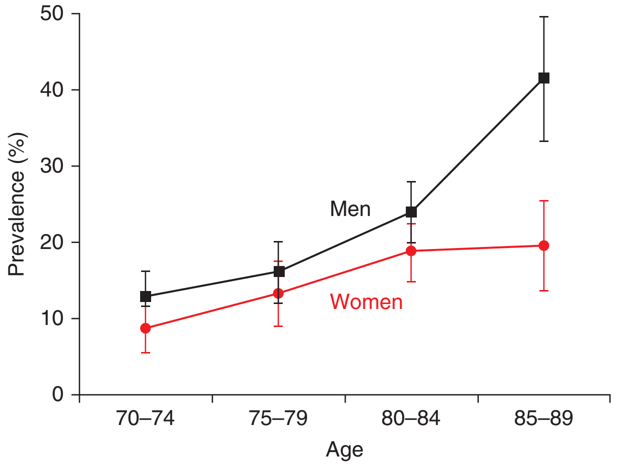

FIGURE 489-8 Prevalence of mild cognitive impairment by age and sex in Olmsted County, Minnesota. *(Reproduced with permission from RC Petersen et al: Prevalence of mild cognitive impairment is higher in men. The Mayo Clinic Study of Aging. Neurology 75:889, 2010.)*

Many nonpharmacologic interventions for older people with dementia, their families, and other caregivers may be beneficial (Table 489-9). Four basic approaches to the pharmacologic treatment of dementia are employed: (1) avoidance of drugs that can worsen cognitive function, mainly those with strong anticholinergic activity; (2) use of agents that enhance cognition and function; (3) drug treatment of coexisting depression, which is common throughout dementia; and (4) pharmacologic treatment of complications such as paranoia, delusions, psychosis, and behavioral symptoms such as agitation (verbal and physical). The use of antipsychotics to treat the neuropsychiatric symptoms of dementia is controversial. Most experts and guidelines recommend avoiding these drugs and using nonpharmacologic strategies unless patients are a danger to themselves and others or if nonpharmacologic interventions have failed. Patients with new or worsening behavioral symptoms associated with dementia should have a medical evaluation to identify potentially treatable precipitating conditions. Pain may be especially hard to detect, and if suspected, a therapeutic trial of acetaminophen should be considered.

The effectiveness of cholinesterase inhibitors and memantine in improving function and quality of life in patients with various types

TABLE 489-8 Evaluation and Management of Delirium

CONTRIBUTING FACTORS	APPROACHES TO EVALUATION AND MANAGEMENT
Drugs	Consider the etiologic role of newly initiated drugs, increased doses, interactions, over-the-counter drugs, and alcohol; consider especially the role of high-risk drugs: lower the dose, discontinue the drug, or substitute a less psychoactive medication. Consider withdraw from chronic medications. Pay special attention to psychotropic medications.
Electrolyte disturbances	Assess for and treat, especially, dehydration, hyponatremia and hypernatremia, and hypothyroidism
Infection	Urinary tract infections, pneumonia, soft tissue infection
Visual/hearing impairment	Encourage the use of glasses and hearing aids if available
Urinary and fecal disorders	Treat urinary retention and fecal impaction
Pulmonary disorders	Correct hypoxemia
Prevent or Manage Complications	
Urinary incontinence	Scheduled toileting
Immobility and falls	Encourage early mobilization to maintain baseline mobility
Pressure ulcers	Mobility and repositioning, nutrition
Sleep disturbances	Sleep hygiene, avoid sedatives, avoid unnecessary awakenings
Feeding disorders	Feeding assistance if necessary, aspiration precautions, liberalize diet if possible
Maintain Patient Comfort and Safety	
Behavioral interventions	Staff education regarding de-escalation techniques for treatment of hyperactive delirium; facilitate family visitation if possible
Pharmacologic interventions	*Only* if patients become a threat to themselves or others and other interventions have failed
Restore Function	
Hospital environment	Adequate lighting, reduce noise
Cognitive reconditioning	Reorient patient frequently
Physical reconditioning	Physical and occupational therapy
Discharge planning	Assess new needs based on predischarge functional status, evaluate social support, coordinate transitions of care, medication reconciliation

Source: Adapted from ER Marcantonio: Delirium in hospitalized older adults. N Engl J Med 378:96, 2018.

TABLE 489-9 Key Principles in the Management of Dementia

- Optimize the patient's physical and mental function through physical activity and mind plasticity principles and activities
  - Treatment underlying medical and other conditions (e.g., hypertension, Parkinson's disease, depression)
  - Avoid use of drugs with central nervous system side effects (unless required for management of psychological or behavioral disturbances—see **Chap. 31**)
  - Assess the environment and suggest alterations, if necessary
  - Encourage physical and mental activity
  - Avoid situations stressing intellectual capabilities; use memory aids whenever possible
  - Prepare the patient for changes in location
  - Emphasize good nutrition
- Identify and manage behavioral symptoms and complications
  - Driving (consider a formal driving evaluation)
  - Wandering
  - Dangerous driving
  - Behavioral disorders
  - Depression
  - Agitation or aggressiveness
  - Psychosis (delusions, hallucinations)
  - Malnutrition
  - Incontinence
- Provide ongoing care
  - Reassessment of cognitive and physical function
  - Treatment of medical conditions
- Provide information to patient and family
  - Nature of the disease
  - Extent of impairment
  - Prognosis
- Provide social service information to patient and family
  - Local Alzheimer's association
  - Community health care resources (day centers, homemakers, home health aides)
  - Legal and financial counseling
  - Use of advance directives
  - Provide family counseling for:
    - Setting realistic goals and expectations
    - Identification and resolution of family conflicts
    - Handling anger and guilt
    - Decisions on respite or institutional care
    - Legal concerns
    - Ethical concerns
    - Consideration of palliative and hospice care
- Protect the caregiver from effects of caregiver stress

Source: Reproduced with permission from RL Kane et al (eds): *Essentials of Clinical Geriatrics*, 8th ed. New York, McGraw-Hill, 2018.

of dementia is controversial, and the potential benefits of these drugs versus their risks and costs must be weighed carefully to provide optimal person-centered care. The best evidence for the effectiveness of cholinesterase inhibitors is in delaying the progression of Alzheimer's disease and increasing the time before institutional placement is needed. GI side effects can be problematic and include nausea, vomiting, and diarrhea; nightmares can be bothersome as well. In addition to these bothersome side effects, cholinesterase inhibitors can cause bradycardia and have been associated with syncope, injurious falls, and pacemaker placement. Memantine can cause dizziness, headache, confusion, and constipation. In one study, vitamin E was more effective than memantine in preventing functional decline in patients with Alzheimer's disease.

The diagnosis and management Alzheimer's disease and related disorders is changing rapidly because of the recent availability of more accurate biomarker testing and antiamyloid antibody therapy. Plasma p-tau 217 has been shown to be reasonably accurate in identifying amyloid-β and predicting cognitive decline in people with Alzheimer's disease. Antiamyloid antibodies have been approved by the U.S. Food and Drug Administration and can clear amyloid on positron emission tomography scan imaging. However, none have been shown to have major impacts on cognition as of the end of 2024. In addition, they have side effects, including intracerebral hemorrhage, and a high out-of-pocket cost **(see Chap. 442)**.

URINARY INCONTINENCE AND OVERACTIVE BLADDER

Epidemiology and Impact Urinary incontinence is curable or controllable in many older patients, especially those who have adequate mobility and mental functioning. Even when it is not curable, incontinence can be managed in a manner that keeps people comfortable, makes life easier for caregivers, and minimizes the costs of caring for the condition and its complications. Approximately one in three women and 15–20% of men older than age 65 years have some degree of urinary incontinence. Between 5 and 10% of community-dwelling older adults have incontinence more often than weekly and/or use a pad for protection from urinary accidents. The prevalence is as high as 60–80% in many nursing homes, where residents often have both urinary and stool incontinence. Many older people (~40%) suffer from "overactive bladder" (OAB), which may or may not include symptoms of incontinence. Symptoms of OAB include urinary urgency (with or without incontinence), urinary frequency (voiding every 2 h or more often), and nocturia (awakening at night to void). If nocturia alone is the predominant symptom, the patient should be asked about sleep disorders (see next section). The pathophysiology, evaluation, and management of OAB are essentially the same as for urge urinary incontinence.

Incontinence and OAB are associated with social isolation and depression and can be a precipitating factor in the decision to seek nursing home care when it cannot be managed in a manner that maintains hygiene and safety. In addition to predisposing to skin irritation and pressure ulcers, the most important potential complications of urinary incontinence and OAB are falls and resultant injuries related to rushing to get to a toilet. Older people with gait disorders, especially those with OAB and multiple episodes of nocturia or nocturnal incontinence, are at especially high risk for falls and related injuries. In addition to the bother of the condition to the older person or a caregiver, fall risk is a compelling reason for undertaking a diagnostic evaluation and specific treatment for incontinence and OAB in the geriatric population.

Evaluation Internists should ask older people about symptoms of urinary incontinence because these symptoms are often hidden out of embarrassment or fear. Simple questions can help identify incontinent patients, such as "Do you have trouble with your bladder?" "Do you ever lose urine when you don't want to?" "Do you ever wear padding to protect yourself in case you lose urine?" (Table 489-2). A substantial number of older people will respond "no" to the first two questions but "yes" to the third one.

Several points are worth noting for the practicing internist. The history and physical examination should focus on identifying potentially reversible causes and contributing factors **(Table 489-10)** and identifying the specific lower urinary tract symptoms. A simple, three-item validated questionnaire can assist in distinguishing between the most common types of incontinence **(Fig. 489-9)**. Key aspects of the history and physical exam are outlined in **Table 489-11**. Among older women, the most common symptoms are a mixture of urge and stress incontinence **(Fig. 489-10)**; the urge is usually the more bothersome. Stress incontinence can often be objectively observed during a physical examination with a comfortably full bladder by having the patient cough in the standing position; leakage of urine simultaneously with

TABLE 489-10 Reversible Conditions That Cause or Contribute to Urinary Incontinence and Overactive Bladder Symptoms in Older People

CONDITION	MANAGEMENT
Lower urinary tract conditions	
Urinary tract infection (symptomatic with frequency, urgency, dysuria, etc.)	Antimicrobial therapy
Atrophic vaginitis/urethritis	Topical estrogen (not a primary treatment for incontinence but may help prevent recurrent infections and ameliorate symptoms of overactive bladder; oral estrogens can cause or worsen incontinence)
Stool impaction with irritation of bladder/urethral innervation and/or partial bladder outlet obstruction	Disimpaction; appropriate use of stool softeners, bulk-forming agents, and laxatives if necessary; implement bowel regimen
Increased urine production	
Metabolic (hyperglycemia, hypercalcemia)	Better control of diabetes mellitus
	Therapy for hypercalcemia depends on underlying cause
Excess caffeine or fluid intake	Reduction in intake of caffeinated beverages; reduction in fluid intake (most older people with incontinence or overactive bladder self-restrict fluid intake)
Volume overload with increased urine production at night	Support stockings
Venous insufficiency with edema	Leg elevation
	Sodium restriction
	Diuretic therapy (late afternoon dose may be effective)
Congestive heart failure	Medical therapy
Impaired ability or willingness to reach a toilet	
Delirium	Diagnosis and treatment of underlying cause(s)
Chronic illness, injury, or restraint that interferes with mobility	Regular toileting
	Use of toilet substitutes
	Environmental alterations (e.g., bedside commode, urinal)
	Remove restraints if possible
Psychological (depression, anxiety)	Appropriate nonpharmacologic and/or pharmacologic treatment
Drug side effects	Remove offending drug(s) if feasible; modification of dose, frequency, or timing may also reduce symptoms for some drugs:
	Diuretics (polyuria, frequency, urgency)
	Anticholinergics (constipation, incomplete bladder emptying)
	Psychotropic drugs
	Tricyclic antidepressants (anticholinergic effects)
	Antipsychotics (immobility, sedation)
	Sedative-hypnotics (immobility, sedation)
	Narcotic analgesics (constipation, incomplete bladder emptying)
	α-Adrenergic blockers (urethral relaxation)
	α-Adrenergic agonists (urethral contraction and potential incomplete bladder emptying)
	Cholinesterase inhibitors (urinary frequency, urgency)
	Angiotensin-converting enzyme inhibitors (cough precipitating stress incontinence)
	Calcium channel blockers, gabapentin, pregabalin, glitazones (edema with nocturia)
	Alcohol (polyuria, frequency, urgency, sedation, delirium, immobility)
	Caffeine (polyuria, bladder irritation)

Source: Reproduced with permission from RL Kane et al (eds): *Essentials of Clinical Geriatrics*, 8th ed. New York, McGraw-Hill, 2018.

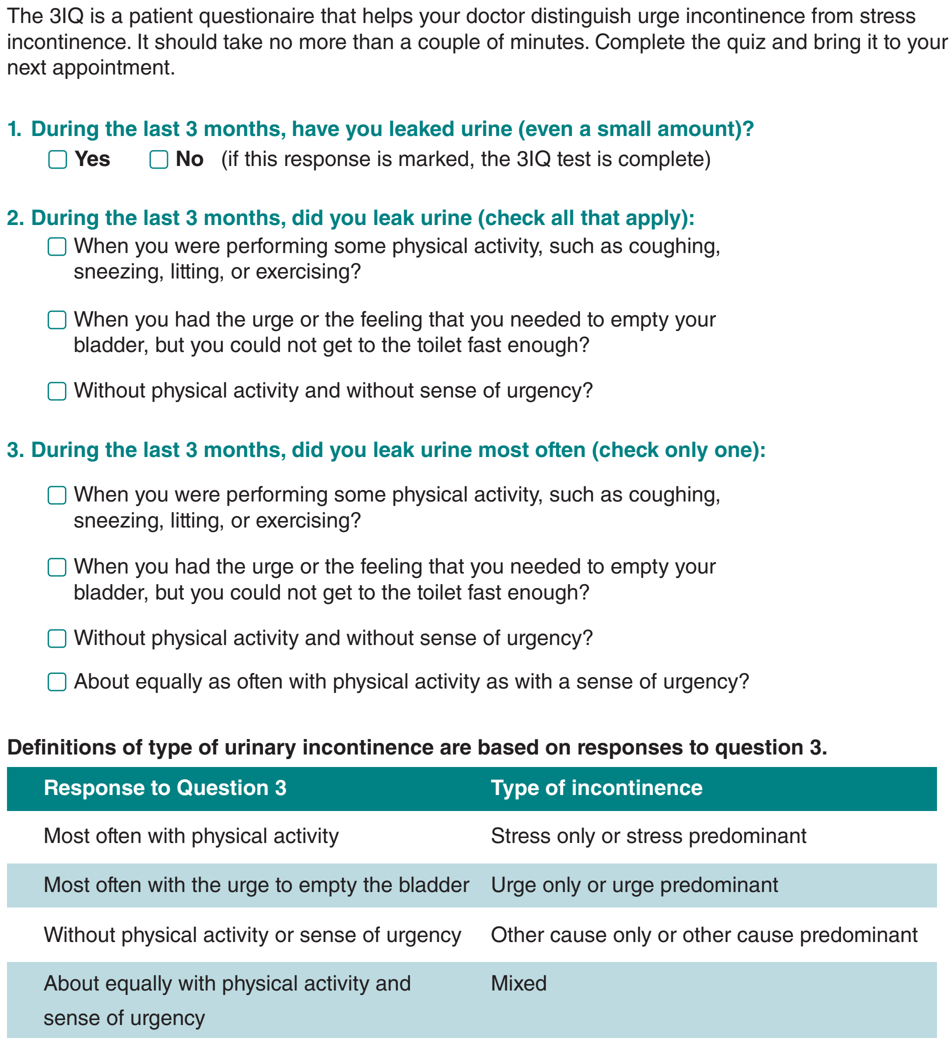

The 3IQ is a patient questionaire that helps your doctor distinguish urge incontinence from stress incontinence. It should take no more than a couple of minutes. Complete the quiz and bring it to your next appointment.

1. During the last 3 months, have you leaked urine (even a small amount)?

☐ **Yes** ☐ **No** (if this response is marked, the 3IQ test is complete)

2. During the last 3 months, did you leak urine (check all that apply):

☐ When you were performing some physical activity, such as coughing, sneezing, litting, or exercising?

☐ When you had the urge or the feeling that you needed to empty your bladder, but you could not get to the toilet fast enough?

☐ Without physical activity and without sense of urgency?

3. During the last 3 months, did you leak urine most often (check only one):

☐ When you were performing some physical activity, such as coughing, sneezing, litting, or exercising?

☐ When you had the urge or the feeling that you needed to empty your bladder, but you could not get to the toilet fast enough?

☐ Without physical activity and without sense of urgency?

☐ About equally as often with physical activity as with a sense of urgency?

Definitions of type of urinary incontinence are based on responses to question 3.

Response to Question 3	Type of incontinence
Most often with physical activity	Stress only or stress predominant
Most often with the urge to empty the bladder	Urge only or urge predominant
Without physical activity or sense of urgency	Other cause only or other cause predominant
About equally with physical activity and sense of urgency	Mixed

FIGURE 489-9 The 3 Incontinence Questions (3IQ) Assessment Tool. *(From Annals of Internal Medicine, JS Brown et al: The sensitivity and specificity of a simple test to distinguish between urge and stress urinary incontinence. 144 (10):715, 2006. Copyright © 2006 American College of Physicians. All Rights Reserved. Reprinted with the permission of American College of Physicians, Inc.)*

coughing indicates that stress incontinence is present. Older men commonly have symptoms associated with OAB and/or symptoms of voiding difficulty (hesitancy, poor or intermittent urinary stream, postvoid dribbling); the OAB symptoms are usually more bothersome. These symptoms overlap with those of both benign and malignant disorders of the prostate, and many internists may choose to consult a urologist for further management **(Chap. 92)** because a urinary flow rate and postvoid residual determination, and further evaluation if malignancy is suspected, help determine therapy.

Most older patients with symptoms of incontinence or OAB should have a postvoid residual determination, especially men, diabetics, those with neurologic disorders, and those with symptoms of voiding difficulty because incomplete bladder emptying is common in older patients and is difficult to detect by history and physical examination alone. There is no specific cutoff for an abnormal postvoid residual; the test must be done with a full bladder, and straining during the test can alter the results. In older patients, a postvoid residual between 0 and 100 mL is normal, a residual between 100 and 200 mL must be interpreted based on symptoms, and a value >200 mL is abnormal and usually influences treatment. Incomplete bladder emptying is a common contributing factor in urinary frequency and nocturia if bladder sensory function is intact or hypersensitive.

Management Some patients should be referred for further urologic, gynecologic, and/or urodynamic evaluation. Examples include a history of lower urinary tract surgery or radiation or recurrent symptomatic urinary tract infections, marked pelvic prolapse on physical examination of a woman, suspected prostate cancer, and sterile hematuria.

Potentially reversible conditions should be addressed, including the many types of medications that can affect bladder function, which should be eliminated if possible (Table 489-10). **Table 489-12** lists treatments for different types of incontinence. Many patients respond well to properly taught and adhered to behavioral interventions. Physical therapists and nurses who specialize in treating lower urinary tract symptoms can be very helpful and should be consulted if available. Pharmacologic treatment of incontinence and OAB is dictated by the innervation of the lower urinary tract. α-Adrenergic stimulation increases tone in the smooth muscle of the urethra; thus, α-agonists have been used to treat stress incontinence in women (although none are approved by the U.S. Food and Drug Administration for this indication), and α-blockers are used to decrease urethral tone in men with OAB associated with prostate enlargement. Anticholinergic/antimuscarinic agents and β_3-agonists inhibit bladder contraction and are used for OAB and urge incontinence. The latter do not have the bothersome anticholinergic effects of antimuscarinic drugs. In men with OAB and normal postvoid residual who do not respond to an α-blocker (with or without a 5α-reductase inhibitor), adding an antimuscarinic or β_3-adrenergic agent may improve symptoms with a very low risk of causing urinary retention. Patients with severe cognitive impairment and/or immobility can often be managed effectively by prompted voiding (during the daytime) and/or incontinence undergarments, as long as comfort, dignity, and safety are maintained.

Older people with incontinence or OAB who have gait disorders or a history of falling should be encouraged to use a urinal, bedside commode, or external catheter. If they insist on walking to the bathroom, they should have a clearly lit and uncluttered path to do so.

TABLE 489-11 Key Aspects of the History and Physical Examination of an Older Patient with Symptoms of Urinary Incontinence and Overactive Bladder

History

- Active medical conditions, especially neurologic disorders, diabetes mellitus, congestive heart failure, venous insufficiency
- Medication review for drugs that can contribute (see Table 489-10)
- Fluid intake pattern
  - Type and amount of fluid (especially caffeine and fluids before bedtime)
- Past genitourinary history, especially childbirth, surgery, dilatations, urinary retention, recurrent urinary tract infections
- Symptoms of incontinence
  - Onset and duration
  - Type—stress vs urge vs mixed vs other (see Fig. 489-10)
  - Frequency, timing, and amount of incontinence episodes and of continent voids (a voiding diary may be useful)
- Other lower urinary tract symptoms
  - Irritative—dysuria, frequency, urgency, nocturia
  - Voiding difficulty—hesitancy, slow or interrupted stream, straining, incomplete emptying
  - Other—hematuria, suprapubic discomfort
- Other symptoms
  - Neurologic (indicative of stroke, dementia, parkinsonism, normal-pressure hydrocephalus, spinal cord compression, multiple sclerosis)
  - Psychological (depression)
  - Bowel (constipation, stool incontinence)
  - Symptoms suggestive of volume-expanded state (e.g., lower extremity edema, shortness of breath while horizontal or with exertion)
- Environmental factors
  - Location of bathroom
  - Availability of toilet substitutes (e.g., urinal, bedside commode)
- Perceptions of incontinence
  - Patient's concerns or ideas about underlying cause(s)
  - Most bothersome symptom(s)
  - Interference with daily life
  - Severity (e.g., "Is it enough of a problem for you to consider surgery?")

Physical Examination

- Mobility and dexterity
  - Functional status compatible with ability to self-toilet
  - Gait disturbance (e.g., that may suggest parkinsonism, normal-pressure hydrocephalus)
- Mental status
  - Cognitive function compatible with ability to self-toilet
  - Motivation
  - Mood and effect
- Neurologic
  - Focal signs (especially in lower extremities) that could suggest a central nervous system condition
  - Signs of parkinsonism
  - Sacral arc reflexes (e.g., loss of perianal sensation or an anal wink in response to perianal stimulation)
- Abdominal
  - Bladder distension[a]
  - Suprapubic tenderness
  - Lower abdominal mass
- Rectal
  - Perianal sensation
  - Sphincter tone (resting and active)
  - Impaction
  - Masses
  - Size and contour of prostate (neither is diagnostic of urethral obstruction)
- Pelvic
  - Perineal skin condition
  - Perineal sensation
  - Atrophic vaginitis (friability, inflammation, bleeding)
  - Pelvic prolapse or mass
- Other
  - Lower extremity edema or signs of congestive heart failure (if nocturia is a prominent complaint)

[a] Clinically significant degrees of urinary retention may be difficult to detect on physical examination; many incontinent patients should have a postvoid residual determination done by ultrasound (see text).

Source: Reproduced with permission from RL Kane et al (eds): *Essentials of Clinical Geriatrics*, 8th ed. New York, McGraw-Hill, 2018.

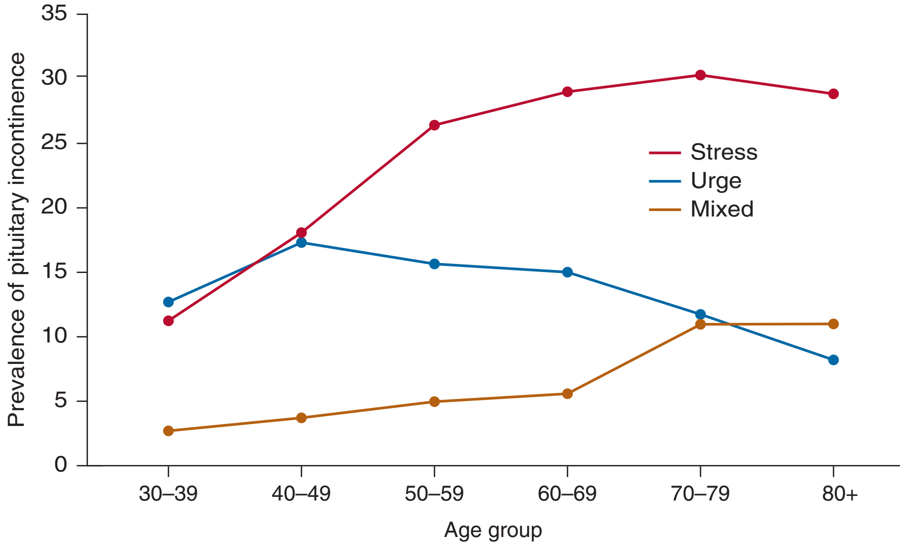

FIGURE 489-10 Rates of urge, stress, and mixed incontinence, by age group, in a sample of 3552 women. *Based on a sample of 3553 participants. *(Adapted from JL Melville, W Katon, K Delaney, K Newton: Urinary incontinence in US women: A population-based study. Arch Intern Med 165:537, 2005.)*

SLEEP DISORDERS

Sleep disorders are discussed in more detail for the general adult population in Chap. 33. Because they are so common and have some unique features in older patients, they are discussed briefly here.

Epidemiology and Impact Aging is associated with multiple changes in sleep architecture as well as multiple diseases and disorders that can disrupt sleep. Thus, complaints of sleep difficulty are common in older adults. Consequences of sleep difficulty include lower health-related quality of life, increased medication use, more cognitive decline, and greater health care utilization. Four types of primary sleep disorders are common in the geriatric population: insomnia, sleep-disordered breathing due to obstructive sleep apnea (OSA), restless leg syndrome (RLS), and periodic leg movements in sleep (PLMS). Complaints of bothersome insomnia—the inability to fall asleep or stay asleep despite a conducive environment—increase with age and occur in close to 30% of people older than 65. Insomnia is commonly associated with depression, anxiety, alcohol intake, and ingestion of caffeinated beverages later in the day. OSA occurs in ~10% of older adults but is probably underreported and underdiagnosed. It is associated with medical comorbidities, such as obesity and congestive heart failure. RLS occurs in 5–10% of adults, and its prevalence increases in those older than 70. It is almost twice as common in women than in men. Family history, iron deficiency, and intake of antihistamines and most antidepressants are risk factors. PLMS can be found in up to 45% of older people but is often of unknown clinical consequence and remains undiagnosed.

TABLE 489-12 Primary Treatments for Different Types of Geriatric Urinary Incontinence

TYPE OF INCONTINENCE	PRIMARY TREATMENTS
Stress	Pelvic muscle (Kegel) exercises Other behavioral interventions including timed voiding and double voiding to avoid residual urine α-Adrenergic agonist (none are approved by the U.S. Food and Drug Administration for this purpose) Topical estrogen to strengthen periurethral tissue (not effective alone; oral estrogens contraindicated) Periurethral injections to provide bulking and support Surgical bladder neck suspension or sling for severe incontinence, based on patient preference
Urge and overactive bladder symptoms	Pelvic muscle (Kegel) exercises Other behavioral interventions: timed voiding and double voiding to avoid residual urine Antimuscarinic and β-3-adrenergic drugs
Incontinence with incomplete bladder emptying	α-Adrenergic antagonists in men with a 5α-reductase inhibitor if the prostate is enlarged); an antimuscarinic or β-3-adrenergic drug can be added if unresponsive to the α-adrenergic agonist Bladder training, double voiding Intermittent catheterization Indwelling catheterization in selected patients in whom risks and discomforts of urinary retention outweigh risks of a chronic indwelling catheter
Incontinence with impaired physical and/or cognitive function	Behavioral interventions (prompted voiding, habit training) Environmental manipulation including use of urinal or bedside commode, safe lit path to bathroom) Incontinence undergarments and pads

Source: Reproduced with permission from RL Kane et al (eds): *Essentials of Clinical Geriatrics*, 8th ed. New York, McGraw-Hill, 2018.

Evaluation Older people should be screened for sleep difficulty with questions such as, "Do you often feel sleepy during the day?" and "Do you have difficulty falling asleep at night?" Further evaluation of the nature and impact of the complaints can be accomplished with standardized questionnaires (Table 489-3). Patients with significant sleep complaints should be asked about conditions that can interrupt sleep, such as nocturia, gastroesophageal reflux, cough due to reflux, chronic sinusitis with postnasal drip, chronic pain, and caffeine or alcohol intake. Specific questions characterizing the complaints should include inquiring about loud snoring (for OSA), the urge to move legs associated with uncomfortable sensations (RLS), and leg movements during sleep (PLMS; which may result in kicking a bed partner).

Management Patients suspected of having OSA, RLS, or PLMS should be referred for formal sleep evaluation. While hypnotics are among the most commonly prescribed drugs in the geriatric population, nonpharmacologic management of sleep should be the initial and primary approach, as many patients can benefit from properly taught and adhered to sleep hygiene interventions **(Table 489-13)**. Benzodiazepine hypnotics should be avoided whenever feasible because they are associated with next-day hangover effects, which may manifest as cognitive impairment and can precipitate falls and car crashes and rebound insomnia. Patients with sleep-onset insomnia may respond to melatonin or low-dose trazadone, both of which are safer than using a benzodiazepine chronically.

FRAILTY

Definition, Epidemiology, and Impact The term *frail* is often used to describe older adults. However, over the past several years, *frailty* has been defined as a specific syndrome, and the word *frail* is more appropriately used to describe people who meet frailty criteria. Frailty is a state of increased vulnerability characterized by a decline in physiologic reserve and function across multiple systems. Many different definitions and tools to define frailty exist. Fried criteria based on the Cardiovascular Health Study (see below) and the Frailty Index (a list of several specific diagnoses developed by Rockwood and colleagues) have been used to screen for frailty in clinical settings. The importance of screening for frailty is to mitigate disability and adverse health outcomes as well as for the assessment of benefits and risks of treatment decisions. The prevalence of frailty is higher among women and increases with age. The overall prevalence of frailty in community-dwelling adults aged 65 and older varies considerably but, on average, is 10–14% depending on the definition. The prevalence of frailty

TABLE 489-13 Nonpharmacologic Management of Insomnia in Older Adults

Sleep Hygiene Rules
Check effect of medication on sleep and wakefulness
Avoid caffeine, alcohol, and cigarettes after lunch
Limit liquids in the evening
Keep a regular bedtime-waketime schedule
Avoid naps or limit to 1 nap a day, no longer than 30 min
Spend time outdoors (without sunglasses), particularly in the late afternoon or early evening
Exercise—but limit exercise immediately before bedtime
Instructions for Stimulus-Control Therapy
Only go to bed when tired or sleepy
If unable to fall asleep within 20 min, get out of bed (and bedroom if possible); while out of bed, do something quiet and relaxing
Only return to bed when sleepy
If unable to fall asleep within 20 min, again get out of bed
Repeat these behaviors until able to fall asleep within a few minutes
Get up at the same time each morning (even if only a few hours of sleep)
Avoid naps

Source: Adapted from JB Halter et al (eds): *Hazzard's Geriatric Medicine and Gerontology*, 8th ed. New York, McGraw-Hill, 2022.

increases with age, reaching close to 16% in individuals age 80–84 and 26% in those aged 85 or older. In older hospitalized patients and institutionalized older people, the frailty prevalence varies from about 27% to up to 80%. Irrespective of the definition, the prevalence of frailty shows a U-shaped relationship with body mass index (BMI), with higher levels of frailty in individuals with both low and very high BMI.

Pathophysiology Frailty is a three-dimensional process that involves changes at the cellular, physiologic, and phenotypical levels. At the cellular level, frailty manifests as changes in mitochondrial function, the development of oxidative stress and DNA damage, telomere shortening, and stem cell exhaustion. These changes at the cellular level result in physiologic alterations including inflammation, cell mediator dysfunction such as low production of nitric oxide by the endothelium, sarcopenia, and energy unbalance. Fried and colleagues conceptualize frailty as a vicious circle of declining energetics and reserve, whose elements represent both the diagnostic criteria for the syndrome identification and the core elements of its pathophysiology. The process manifests phenotypically as an overall decline in physical function and cognitive impairment. In particular, the phenotype of frailty has been defined by Fried and colleagues by the five following characteristics: unintentional weight loss, weakness, exhaustion, slowness, and low activity (with specific operational definitions of each).

Management Although there is conflicting evidence regarding the effectiveness of specific interventions to treat or prevent frailty, person-centered physical activity programs and nutritional supplementation appear to improve components of frailty such as muscle strength, gait speed, and overall mobility. In addition, optimizing the management of chronic conditions, medication management including mitigation of polypharmacy, and identifying the individual's priorities could lead to reversing or slowing the progression of frailty. The growing field of "geroscience" is actively studying drugs including metformin and a variety of senolytic agents for their potential effects on preventing or mitigating frailty.

■ ELDER ABUSE AND NEGLECT

Epidemiology and Impact The incidences of elder abuse neglect and self-neglect are unknown because they are often unrecognized. The best data suggest that the incidence over 12 months is at least 8–10%. Abuse and neglect can result in physical injuries and related pain, worsening of chronic medical conditions, dehydration and pressure ulcers, emotional distress, and loss of income and savings.

Evaluation Because abuse and neglect are underreported, are unsuspected, and have such devastating consequences, older adults should be screened (without the presence of caregivers) with questions such as, "Do you ever feel unsafe where you live?" "Has anyone ever threatened or hurt you?" "Has anyone been taking your money without your permission?" (Table 489-2). Table 489-14 outlines the definitions, symptoms and signs, and key aspects of evaluating suspected abuse and neglect.

Management In addition to treating the physical, medical, and emotional consequences, patients suspected of elder abuse or neglect should be reported to the appropriate local or state agency to investigate and ensure the patient's safety. The reader is referred to two reviews of this topic for further information on specific aspects of management.

■ COVID-19

The COVID-19 pandemic disproportionately affected the older population, especially those residing in nursing homes and assisted living facilities. Compared with those between the ages of 18 and 29 years old, older adults are at greater risk for adverse outcomes after infection with COVID-19, especially those with multiple comorbidities. Internists should regularly refer to the CDC guidelines on vaccination as they are frequently updated. (See Chap. 205 for more details on managing COVID-19). The COVID-19 epidemic had devastating and persistent effects on nursing home staffing and heightened awareness of changes that are essential for long-term care quality in the United States (see National Academies of Sciences, Engineering, and Medicine report in "Further Reading").

END-OF-LIFE AND PALLIATIVE CARE

End-of-life and palliative care are critical aspects of caring for the geriatric population and require a comprehensive, person-centered approach; they are addressed in detail in Chap. 13, and pain management is addressed in Chap. 14. For older patients, limited life expectancy is a critical factor to consider when making end-of-life care decisions. General principles of decision-making are especially relevant when considering palliative and/or end-of-life care in older patients (Fig. 489-5). Decision-making becomes complicated, however, among older patients with multimorbidity. Without a clear terminal diagnosis, when to start palliative care/end-of-life care could be challenging. While it is sometimes clear when an older patient has a terminal condition, such as end-stage congestive heart failure or chronic obstructive pulmonary disease, many older patients with multimorbidity have combinations of conditions of varying severity. Moreover, neurogenerative disorders, including most forms of dementia, Parkinson's disease, and patients with multiple strokes, commonly have a gradually progressive course, and it can be challenging to determine when discussions about palliative and end-of-life care should be initiated. Dementia, however, should be considered a terminal illness in the advanced stages.

Internists should play a pivotal role in making the decision when to initiate these discussions and should be proactive in encouraging patients and their families to execute advance directives before a health care crisis occurs. There are good data that bear on some of the decisions. For example, the survivability of cardiopulmonary resuscitation (CPR) in hospitalized patients age 65 and older is <20%; among the old-old with multimorbidity, it is much lower. The survivability of CPR in nursing home residents is almost zero, making it a futile intervention for most in this setting. Data and recommendations from major organizations suggest that enteral feeding tubes should not be placed in patients with end-stage dementia (Table 489-1). Tools for the estimation of prognosis such as *ePrognosis.com*, for holding conversations with older people and their families about advance care planning, and for documentation of advance directives (e.g., living will, durable power of attorney for health care, Physician Orders for Life-Sustaining Treatments [POLST], and other order sets) will assist internists in paying careful attention to factors that contribute to person-centered care and in dealing with these challenging issues in end-of-life geriatric care.

TABLE 489-14 Elder Abuse and Neglect

CATEGORY	DEFINITION AND EXAMPLES	SYMPTOMS AND SIGNS	KEY ASPECTS OF EVALUATION
Physical abuse	Acts of violence that may result in pain, injury, or impairment • Pushing, slapping, hitting, force-feeding • Improper positioning or use of restraints • Improper use of medications	Abrasions Lacerations Bruises Fractures Use of restraints Burns Pain Depression Delirium or onset or worsening of dementia-related behavioral symptoms	The interview should be conducted alone with the patient; it may reveal discordant histories or findings inconsistent with the history provided by the caregiver. Ankles and wrists should be examined for abrasions suggestive of the use of restraints. Findings that are discordant with the mechanism of injury reported or multiple injuries in various stages of healing should raise the suspicion of abuse. Injuries to the head, neck, and upper arms occur in victims of physical elder abuse but must be distinguished from accidental injuries. Jaw and zygomatic fractures are more likely to be sustained from a punch than from a fall, which more typically result in fractures to orbital and nasal bones.
Psychological or verbal abuse	Conduct that causes mental or emotional distress • Verbal harassment or intimidation • Threats of punishment or deprivation • Isolation	Direct observation of verbal abuse Subtle signs of intimidation, such as deferring questions to a caregiver or potential abuser Evidence of isolation Depression, anxiety, or both	Assess the size and quality of the patient's social network (beyond the suspected abuser). Conduct standardized assessments of depression, anxiety, and cognition, directly or through referral. Ask specifically about verbal or psychological abuse with questions such as, "Does your relative/caregiver ever yell or curse at you?"; "Have you been threatened with being put into a nursing home?"; or "Are you ever prevented from seeing friends and family members whom you wish to see?"
Financial abuse	Misuse of the person's income or resources for the financial or personal gain of a caregiver or advisor • Stealing money or possessions • Denying a home • Coercing to sign contracts or spend money	Inability to pay for medicine, medical care, food, rent, or other necessities Failure to renew prescriptions, adhere to medication regimens or other treatments, or keep medical appointments Malnutrition, weight loss, or both, without an obvious medical cause Evidence of poor financial decision-making Firing of home care or other service providers by abuser Unpaid utility bills Initiation of eviction proceedings	Ask about financial exploitation with questions such as, "Has money or property been taken from you without your consent?"; "Have your credit cards or automated teller machine card been used without your consent?"; and "At the end of the month, do you have enough money left for food and other necessities?" Abrupt changes in financial circumstances of the caregiver in either direction may herald an increased risk of financial exploitation or exploitation already under way. Abuse of the power of attorney; if the person with power of attorney or health care proxy is suspected of not acting in the best interest of the patient, documents necessary to ensure that the assumption of fiduciary responsibilities is authorized.
Sexual abuse	Sexual coercion or assault	Bruising, abrasions, lacerations in the genital or anal areas or abdomen Newly acquired sexually transmitted diseases, especially in nursing home Urinary tract infection	Inquire directly about sexual assault or coercion. For patients with dementia, direct queries to caregivers about hypersexual behavior as part of a larger history regarding dementia-related behaviors and assess patient's capacity for decision-making about sexual activity. If indicated, refer to an emergency department for assessment for sexual assault and collection of specimens (forensic evidence should be collected by experienced professionals, such as nurses who have undergone Sexual Assault Nurse Examiners [SANE] training).
Neglect (by caregiver or self-neglect)	Failure to provide the materials, supplies, food and drink, or services necessary for optimal functioning or to avoid harm	Malnutrition Dehydration Poor hygiene Pressure ulcers Nonadherence to medication regimen or other treatments Worsening of dementia-related behavioral symptoms	Interview primary caregiver about their understanding of the nature of the patient's care needs and how well care is being rendered. Neglect by a caregiver may be intentional or unintentional. Assess hygiene, cleanliness, and appropriateness of dress. Examine the skin for pressure ulcers, infections, and infestations. Assess nutrition and hydration, including measuring body mass index and blood urea nitrogen and creatinine to assess hydration.

MODELS OF GERIATRIC CARE

Several innovative models of care have been developed over the past three decades designed to provide high-quality and effective care for the burgeoning geriatric population with multimorbidity, functional and cognitive impairment, and challenges with social support. These include outpatient comprehensive geriatric assessment programs, inpatient acute care for the elderly (ACE) units, consultation and co-management services, and home-based programs. These models of care are assuming greater importance in the emerging era of value-based health care services.

CARE TRANSITION INTERVENTIONS

Improving transitions of care between settings has become a major focus of governments, health systems, hospitals, postacute care (PAC) and LTC organizations and programs, physicians, and other health care professionals. Geriatric patients are especially vulnerable to complications at the time of discharge from an acute medical or psychiatric hospital, as well as at the time of discharge from a PAC facility (skilled nursing facility [SNF], acute rehabilitation or long-term hospital) or home care program. With the increasing role of hospitalists, and others who specialize in SNF care, medical care for geriatric patients has become fragmented at the time of transitions, creating opportunities for communication problems and medical errors. Changes in reimbursement and financial penalties for high rates of hospital readmissions have driven the development of many care transition interventions (**Table 489-15**).

INTERPROFESSIONAL TEAMS AND CO-MANAGED CARE

The complexity of caring for the aging population is more evident during hospitalizations due to a new acute illness or exacerbation of

TABLE 489-15 Examples Care Transitions Interventions

INTERVENTION	WEBSITE	CORE INTERVENTIONS
Re-Engineered Discharge (Project RED) (Jack et al: 2009)	*https://www.bu.edu/fammed/projectred/*	"Discharge advocate" performs the following: • Facilitates patient education and understanding • Performs medication reconciliation • Coordinates postdischarge appointments and communication with primary care provider (PCP) • Calls patient 2–3 days after discharge
Transitional Care Model (Naylor et al: 2004; Naylor et al: 1999)	*https://www.nursing.upenn.edu/ncth/transitional-care-model/*	Advanced practice nurse performs the following: • Coordinates patient care before and after discharge • Assesses each patient's needs; engages and activates the patient and family • Facilitates communication among patient, family, and health care providers • Conducts regular home visits and telephone support after discharge
Care Transitions Program® (Coleman et al: 2004)	*http://www.caretransitions.org*	"Transition coach" performs the following: • Facilitates improved self-management skills including medication management and how to respond to warning signs/symptoms • Makes postdischarge home visits and phone calls
Better Outcomes for Older Adults through Safe Transitions (BOOST) (Hansen et al: 2013)	*https://www.hospitalmedicine.org/clinical-topics/care-transitions/*	Includes toolkit facilitating the following: • Comprehensive identification and assessment of high-risk patients • Patient/caregiver education • Enhanced communication with posthospitalization care providers • Follow-up phone call with patient after discharge
Interventions to Reduce Acute Care Transfers (INTERACT) (Ouslander et al: 2013)	*http://www.interact-pathway.com*	Includes tools for skilled nursing, assisted living, and home health care, including: • Quality improvement • Communication • Decision support • Advance care planning

preexisting chronic conditions. Interprofessional teams integrate different areas of expertise to provide patient-centered care. Physicians should understand and respect the roles of nurses; physical, occupational, and speech therapists; nutritionists; pharmacists; psychologists; social workers; clergy; and other direct care staff. The evolution of interprofessional teams has resulted in a comprehensive approach to care by opening channels of communication between health professionals from different disciplines.

Co-managed medicine is an example of how enhanced communication between different providers improves outcomes, avoids common complications, and saves resources. In the era of person-centered and value-based care, effective co-managed medicine appears to deliver consistently high-quality care at a lower cost. Collaborations between internists and geriatricians are examples of this strategy. Hip fracture and trauma co-management programs also have been developed in many academic and community hospitals and are demonstrating some success in reducing complications and length of stay in older trauma patients.

AGE-FRIENDLY HEALTH SYSTEMS

A new framework for providing comprehensive, integrated, and person-centered care across settings of care has been developed called "age-friendly health systems." Health systems participating in the development of age-friendly programs focus on the 5M's discussed previously (Fig. 489-4) as a strategy to achieve high-quality care across the system. Strategies are implemented to educate and facilitate all system health care providers to focus on the 5M's of geriatrics under the leadership and mentorship of specially trained geriatrics health professionals.

FURTHER READING

2019 American Geriatrics Society Beers Criteria® Update Expert Panel: American Geriatrics Society 2019 Updated AGS Beers Criteria® for potentially inappropriate medication use in older adults. J Am Geriatr Soc 67:674, 2019.

2023 American Geriatrics Society Beers Criteria® Update Expert Panel: American Geriatrics Society 2023 updated AGS Beers Criteria® for potentially inappropriate medication use in older adults. J Am Geriatr Soc 71:2052, 2023.

AMDA—The Society for Post-Acute and Long-Term Care Medicine: Ten things clinicians and patients should question. *http://www.choosingwisely.org/societies/amda-the-society-for-post-acute-and-long-term-care-medicine/*. Accessed September 20, 2020.

American Diabetes Association: Older adults: Standards of medical care in diabetes—2020. Diabetes Care 43(Suppl 1):S152, 2020.

American Geriatrics Society: Choosing Wisely: Ten things clinicians and patients should question. *http://www.choosingwisely.org/societies/american-geriatrics-society/*. Accessed September 20, 2020.

American Geriatrics Society Panel on Pharmacologic Management of Persistent Pain in Older Persons: Pharmacologic management of persistent pain in older persons. J Am Geriatr Soc 46:1331, 2009.

Halter JB et al (eds): *Hazzard's Geriatric Medicine and Gerontology*, 8th ed. New York, McGraw-Hill, 2022.

Institute for Healthcare Improvement: Age-friendly health systems. *http://www.ihi.org/Engage/Initiatives/Age-Friendly-Health-Systems/Pages/default.aspx*. Accessed September 20, 2020.

Kane RL et al (eds): *Essentials of Clinical Geriatrics*, 8th ed. New York, McGraw-Hill, 2017.

National Academies of Sciences, Engineering, and Medicine 2022. *The National Imperative to Improve Nursing Home Quality: Honoring Our Commitment to Residents, Families, and Staff.* Washington, DC, The National Academies Press. *https://doi.org/10.17226/2652*.

National Institute on Aging: Safe driving for older adults. *https://www.nia.nih.gov/health/safety/safe-driving-older-adults*. Accessed August 20, 2024.

National Institute on Aging: Vaccinations and older adults. *https://www.nia.nih.gov/health/immunizations-and-vaccines/vaccinations-and-older-adults*. Accessed December 29, 2023.

490 Approach to Medical Consultation

Jeffrey S. Berns, Jack Ende

Effective health care requires teams of generalists and specialists with complementary expertise. Many clinical conditions require the input of more than one clinical provider, either because the diagnosis and recommended treatment is uncertain or because the required diagnostic procedure or management lie outside the principal provider's area of expertise.

To *consult* is to seek advice from someone with expertise in a particular area, whereas *consultation* refers to the meeting or comparable outcome arising from that request. Medical consultation takes several forms. Its most traditional forms include in-hospital consultation, in which physicians provide recommendations, typically recorded in the medical record, or perform procedures for a hospitalized patient, and outpatient consultations, in which patients are seen in the office setting. More contemporary forms of consultation include e-consultations, telemedicine evaluations (see "Consultation Involving Telemedicine," below), and remote medical second opinions. In these forms, the consultant may not actually see the patient but, nonetheless, assumes the responsibility of evaluating the patient's clinical condition, assessing and analyzing pertinent clinical data, and offering a synthesis and appropriate recommendations.

While forms of medical consultation evolve, basic responsibilities associated with medical consultation endure. These responsibilities can be divided into those that fall to the requesting physician or nonphysician practitioner; the consultant, who provides the consultation; and the health system, hospital, or organization that must support this important medical encounter **(Table 490-1)**.

RESPONSIBILITIES OF THE REQUESTING CLINICIAN

Before requesting a consultation, the physician or other provider should ensure that the patient endorses the purpose of the consultation, understands the role of the consultant, and anticipates the likely outcomes of the encounter. Further responsibilities of the requesting physician include being specific and communicating clearly the reason for the consultation. Vague messages such as "Please evaluate" are not as helpful as more specific inquiries such as "What is the cause of the declining kidney function?" or "How should this asymptomatic pulmonary nodule be evaluated?" To the extent possible, the requesting physician should provide the relevant clinical information, summarized as succinctly as possible. Urgency should be clearly conveyed, typically with a phone call or other direct communication.

TABLE 490-1 Stakeholder Responsibilities in the Medical Consultation Process

REFERRING PHYSICIAN OR PROVIDER	CONSULTANT PHYSICIAN	HEALTH SYSTEM, HOSPITAL, OR CARE ORGANIZATION
• Ensure patient participation and engagement • Be specific regarding clinical question and desired outcome • Communicate level of urgency • Avoid consulting for nonclinical purposes	• Maintain standards of professionalism, including those pertaining to availability, communication, respect, and collegiality • Appreciate levels of urgency and respond appropriately • Assemble and develop one's own database • Be specific in synthesis and recommendations • Understand desired outcomes, including arrangements for follow-up • Communicate with referring provider in whatever manner is mutually desirable	• Maintain adequate specialty workforce to enable appropriate access • Support systems for efficient exchange of clinical information • Develop culture of collegiality and team-based care

The requesting physician should be explicit regarding the intended outcome of the consultation, i.e., is this for a single evaluation or ongoing co-management? Communication between the requesting and the consulting physicians is paramount. Whether this communication includes direct contact is less important than that the relevant information and desired outcome be explicit and clear, regardless of communication medium. Consultations should be requested for clinical purposes and always directed to qualified consultants; they should not be driven by entrepreneurial or relationship-building purposes. Another responsibility of the referring physician is not to "overconsult." Medical care should be focused on value, not volume.

RESPONSIBILITIES OF THE CONSULTANT

Just as the referring physician should attend to clear and explicit communication, so too should the consultant follow the precepts of effective interactions between professionals, which include courtesy, availability, and clarity. Particularly on the inpatient service, where consultants may receive several requests each day, it is important that the incoming consultations are triaged and dispatched as clinically appropriate. Consultants also need to determine the requested level of involvement going forward and not assume that long-term co-management is being sought. While consultants can and should make use of available clinical data, they should also assemble independently their own database, including taking a history, performing a physical exam, and reviewing pertinent laboratory, imaging, and pathology studies. Absent that, they may be unable to provide an independent and actionable synthesis. Just as the referring physician needs to be clear and concise, so too should the consultant be specific and focused in the recommendations provided. "Possible malignant ascites" is less helpful than, "I will arrange for paracentesis to exclude the possibility of malignant ascites." For the most part, recommendations to "consider" some diagnosis or test are less helpful than more specific and concrete advice. Some referring physicians wish to be called after a patient is seen; others prefer that communication be handled as part of the medical record. How this communication is handled must also align with the complexity and urgency of the consultation and clinical circumstances.

RESPONSIBILITIES OF HEALTH SYSTEMS, HOSPITALS, AND MEDICAL ORGANIZATIONS

Health systems, hospitals, and medical organizations also have responsibilities in the consultation process. This responsibility includes ensuring that qualified consultants are accessible and available on the medical staff. Consultations within a single system are aided by common shared electronic medical records, particularly when consultations originate in the hospital, but then also involve care in the outpatient setting. Finally, health care entities should strive to foster a culture of team-based care and collegiality. Reimbursement for consultations varies among payors and may have implications for self-pay or unreimbursed expenses for providers or health systems. While it is important to understand reimbursement models, the clinical needs of the patient should be prioritized.

SPECIAL ISSUES IN MEDICAL CONSULTATION

Curbside Consults Curbside consults are requests from one physician to another for an informal and unwritten opinion about a specific patient care matter. They are typically limited in scope, mostly regarding management or questions regarding procedures, and developed from information provided by the consulting physician and perhaps the medical record (such as labs and imaging studies). Although often viewed as convenient, efficient, and a common aspect of clinical care, without a comprehensive review of the record or any direct contact with the patient, curbside consults have been found to often be incomplete or even flawed. It is not uncommon for the question being asked to be deemed too complex for a curbside consult or for it not to be the actual or only issue the consultant feels needs to be addressed. As a general rule, curbside consults should be avoided. While medicolegal liability is often cited as a reason to limit curbside consults, the risk is actually negligible as U.S. courts have ruled that curbside consults do not establish a doctor-patient relationship necessary for creating the basis for medical malpractice litigation. An important exception, however, is when a curbside consult is provided by a resident or fellow in training; in this circumstance, the trainee's supervising physician, whether aware of the curbside consult or not, is responsible for the recommendations of the trainee.

Advice Related to curbside consultation, but decidedly different, are instances when one physician reaches out to another, often one in another specialty, for advice. Examples include an internist turning to a radiologist for guidance on what is the most appropriate imaging study to diagnose a deep tissue abscess; a general internist asking a gastroenterologist for advice on management of acute diverticulitis; a hospitalist asking a neurologist for guidance on management of a patient with Parkinson's disease who is NPO; or a nephrologist asking an infectious disease specialist about immunizations in a kidney transplant recipient.

Outreach such as this is typically triggered by a specific patient encounter, but the request is more for general information that the requesting physician may use for the encounter at hand as well as for similar encounters going forward. Thus, reaching out for advice differs from formal consultation and from curbside consultation, which are specific for a particular patient. As such, requests for advice fall within the realm of collegial communication and not necessarily within that of clinical consultation per se.

Second Opinions Physicians may find themselves providing consultations requested by patients who have already been evaluated for the same problem by another physician. Not a "consult" in the usual context of one physician referring a patient to another, the service provided by the consultant is, nonetheless, very much aligned with a physician-referred consult. Second opinions, which often are encouraged by the patient's physician, may be sought by patients for reassurance that a diagnosis and treatment recommendation are correct, out of dissatisfaction with the initial physician, or with the hope of an entirely different opinion and recommendation. The physician providing the second opinion should strive to understand the patient's motivations for seeking the additional opinion. While a second opinion may have been initiated by the patient rather than referral from another physician, it is recommended that the consulting physician communicate with the patient's primary physician, as would be done following a standard consultation, unless the patient insists otherwise. In addition, professional behavior in how the consulting physician refers to the recommendations or actions of previous physicians is important, including when there is disagreement. Likewise, it is important that a transfer of care from the prior physician to the one providing a second opinion be enacted only if specifically requested by the patient or the physician who encouraged the second opinion.

Consults Involving Advanced Practice Providers Increasingly, specialist physicians may find themselves being consulted by nurse practitioners and physician assistants rather than other physicians. Whether the quality of the information provided to the consultant physician by these providers is different from physician-to-physician referrals has not been studied. Consulting physicians should know whether they should respond back to the advanced practice provider or to the supervising physician, if there is one. As with physician-to-physician consults, it is also important for the consultant to know whether the individual calling for the consult has an ongoing role in the care of the patient or is simply covering for a limited period of time. Finally, the consultant, if responding back to the advanced practice provider, should make sure that the information provided meets the needs of that provider and that questions are answered as they would be if responding back to another physician.

Consultation Involving Telemedicine and Electronic Health Records (EHRs) Consultations making use of EHRs, patient portals, and various forms of telecommunication technology, including video conferencing or cell phone communication, can improve access to care, reduce cost, and improve outcomes. This is particularly true when employed in geographic areas of health care shortage and when the clinical issues can be handled without direct contact with the patient, i.e., radiology or dermatology. However, the absence of direct contact between patient and consultant introduces special issues related to diagnostic accuracy and physician-patient relationship. Regulatory, liability, security, and confidentiality issues arise as well, as do concerns about disparities related to access to telemedicine technologies and willingness and ability to use them among some patient populations.

FURTHER READING

Ahmed S et al: Utility, appropriateness, and content of electronic consultations across medical subspecialties. A cohort study. Ann Intern Med 172:641, 2020.

Pearson SD: Principles of generalist-specialist relationships. J Gen Intern Med 14:S13, 1999.

Sulmasy DH: Policy recommendations to guide the use of telemedicine in primary care: An American College of Physicians Position Paper. Ann Intern Med 163:787, 2015.

491 Medical Disorders During Pregnancy

Sarah Rae Easter, Robert L. Barbieri

Each year, ~3.7 million births occur in the United States, and >130 million births occur worldwide. A significant proportion of births are complicated by medical disorders. Advances in medical care and fertility treatment have increased the number of people with serious medical problems seeking pregnancy. Medical problems that interfere with the physiologic adaptations of pregnancy increase the risk for poor pregnancy outcome. Conversely, in some instances, pregnancy events may have implications for an individual's long-term health.

HYPERTENSION

(See also Chap. 288) Cardiac output increases by 40% in pregnancy, with most of the increase due to an increase in stroke volume. Heart rate increases by ~10 beats/min during the third trimester. In the second trimester, systemic vascular resistance decreases, and this decline is associated with a fall in blood pressure. A blood pressure of systolic ≥140 and/or diastolic ≥90 mmHg on two measurements 4 h apart after 20 weeks' gestation is abnormal and is associated with an increase in perinatal morbidity and mortality. Hypertension during pregnancy is classified as preeclampsia, gestational hypertension, or chronic hypertension. These classifications are distinguished based on timing in pregnancy and the presence of associated features (see below). The prevalence of hypertensive disorders in pregnancy (HDP) is rising. The Centers for Disease Control and Prevention (CDC) estimated that in 2019, 16% of deliveries in the United States were complicated by a HDP.

PREECLAMPSIA

Approximately 5–7% of all pregnant women develop *preeclampsia*, the new onset of hypertension (blood pressure ≥140/90 mmHg) and proteinuria (either a 24-h urinary protein >300 mg/24 h or a protein-creatinine ratio ≥0.3) after 20 weeks of gestation. Preeclampsia can be diagnosed without proteinuria in the presence of symptoms or laboratory abnormalities raising concern for end-organ damage. Specific clinical features qualify as evidence of severe disease, including severe hypertension (blood pressure ≥160/110 mmHg), new-onset symptoms (headache not responsive to medications, visual changes, or unremitting severe epigastric pain), pulmonary edema, or laboratory abnormalities signifying thrombocytopenia (platelets <100 × 10^9/L), renal insufficiency (creatinine >1.1 mg/dL), or liver impairment (elevation of transaminases to twice the normal concentration). Eclampsia is diagnosed when a pregnant person with preeclampsia develops generalized seizures. The *HELLP syndrome* (*h*emolysis, *el*evated *l*iver enzymes, *l*ow *p*latelets) is a special subtype of preeclampsia with severe features and is a major cause of morbidity and mortality. Coagulopathy, cerebrovascular accidents (CVAs), hepatic capsule rupture, and placental abruption are additional end-organ complications of preeclampsia.

The precise pathophysiology of preeclampsia remains unknown, but chronic uteroplacental ischemia, an exaggerated maternal inflammatory response, and/or imbalance of angiogenic factors likely contribute to the clinical syndrome. In preeclampsia, there is excess production of an antiangiogenesis factor, soluble fms-like tyrosine kinase 1 (sFlt-1), and decreased production of an angiogenic factor, placental growth factor (PlGF). A ratio of circulating sFlt-1/PlGF ≥40 is associated with an increased risk of developing preeclampsia with severe features in the 2 weeks following the measurement. Abnormalities of cerebral circulatory autoregulation explain some of the neurologic manifestations of the disease and can increase the risk of stroke at even modestly elevated blood pressures. In the absence of treatment, 1 in 100 cases of preeclampsia may progress to *eclampsia*—new-onset generalized tonic-clonic seizures in a patient with preeclampsia. Low-dose aspirin initiated between 12 and 14 weeks of gestation reduces the risk in women at high risk of developing preeclampsia.

■ GESTATIONAL HYPERTENSION

The development of elevated blood pressure after 20 weeks of pregnancy in the absence of preexisting chronic hypertension or proteinuria is referred to as *gestational hypertension*. Gestational hypertension with severe features of the disease is best classified as preeclampsia. Both gestational hypertension and preeclampsia are associated with an increased risk of developing chronic hypertension, cardiovascular disease, chronic kidney disease, or diabetes mellitus later in life.

TREATMENT

Preeclampsia

The management of preeclampsia is challenging because it requires the clinician to balance the health of the mother with the health of the fetus. The definitive treatment of preeclampsia is delivery of the fetus and placenta, but preterm delivery exposes the fetus to the risks of prematurity. In preeclampsia without severe features, delivery at 37 weeks is recommended. People with preeclampsia without severe features may be managed expectantly until 37 weeks with close monitoring for development of severe features or laboratory abnormalities, frequent fetal surveillance, and limited physical activity to reduce blood pressure.

Expectant management of preeclampsia with severe features remote from term affords some benefits for the fetus but at significant risk to the mother. For women with preeclampsia with severe features, delivery is recommended unless the patient is <34 weeks and eligible for expectant management in a tertiary hospital setting. Indications for delivery prior to 34 weeks include unremitting symptoms, development of laboratory abnormalities, or severe range blood pressures refractory to medical management. The goal of prolonging pregnancy to this gestational age is to improve neonatal outcomes. Therefore, concerns about fetal well-being, such as severe fetal growth restriction or placental abruption, may also prompt delivery before 34 weeks. Timely management of blood pressures ≥160/110 mmHg reduces the risk of CVAs. Labetalol and hydralazine IV are the first-line agents to manage severe hypertension in preeclampsia with consideration of oral agents once blood pressure is controlled.

Women who have had preeclampsia are at increased risk of cardiovascular disease later in life. Interventions to reduce cardiovascular risk can be initiated in health encounters following birth.

■ CHRONIC HYPERTENSION

Pregnant people with chronic hypertension are at increased risk for superimposed preeclampsia and placental complications such as fetal growth restriction and placental abruption. These complications may necessitate preterm birth, which is associated with newborn respiratory distress syndrome. People with chronic hypertension should have a thorough prepregnancy evaluation to identify remediable causes of hypertension and to transition off antihypertensive agents associated with adverse outcomes in pregnancy. Angiotensin-converting enzyme inhibitors and angiotensin-receptor blockers are contraindicated in pregnancy because of an increased risk of stillbirth and congenital fetal anomalies. Labetalol and extended-release nifedipine are the most commonly used medications for the treatment of chronic hypertension in pregnancy. The optimal target blood pressure is <140/90 mmHg, which reduces the risk of developing preeclampsia and preterm birth. A preconception or early pregnancy assessment for end-organ impacts of hypertension, including the presence of proteinuria, may help differentiate the effects of chronic hypertension from those of superimposed preeclampsia.

■ RENAL DISEASE

Normal pregnancy is characterized by a 40% increase in glomerular filtration rate and creatinine clearance secondary to a rise in renal plasma flow. Patients with underlying renal disease may expect a worsening of existing hypertension or development of preeclampsia during pregnancy. A prepregnancy serum creatinine level ≥133 μmol/L (≥1.5 mg/dL) and low C4 levels are associated with adverse pregnancy outcomes. Certain pathologies, such as those associated with glomerular disease, increase the risk of adverse outcomes. Neither hemodialysis nor prior renal transplant is a contraindication to pregnancy, but both require close multidisciplinary management. When renal disease worsens during pregnancy, close collaboration between the internist and the maternal-fetal medicine specialist is essential so that decisions regarding delivery can be weighed to balance the sequelae of prematurity for the neonate versus long-term sequelae for the mother with respect to future renal function.

CARDIAC DISEASE

Cardiac disease is the leading cause of maternal mortality in the United States. Prepregnancy cardiac disease and cardiac disease caused by pregnancy are both major contributors. Patient education, risk stratification, optimization of hemodynamics, and multidisciplinary planning with a pregnancy heart team are the tenets of management. Patients with pulmonary hypertension **(Chap. 294)**, severe ventricular dysfunction (ejection fraction <30% or New York Heart Association class III–IV), severe mitral or aortic stenosis, severe aortic dilation, or Fontan circulation with severe complications are at the highest risk of maternal mortality. Pregnancy is contraindicated in these women, with termination of pregnancy as an option to reduce maternal risk. Risk stratification including a detailed history with attention to symptoms, echocardiography, and cardiopulmonary exercise testing guides monitoring and management for most patients. Contemporary guidelines support continuing most nonteratogenic prepregnancy medications and reserving cesarean delivery for obstetric indications with few exceptions.

■ VALVULAR HEART DISEASE

(See also Chaps. 272–279.)

Mitral Stenosis The pregnancy-induced increase in blood volume, cardiac output, and tachycardia can increase the transmitral pressure gradient and cause pulmonary edema or tachyarrhythmias in women with mitral stenosis. People with moderate to severe mitral stenosis (mitral valve area ≤1.5 cm^2) who are planning pregnancy and have either symptomatic disease or pulmonary hypertension should undergo valvuloplasty prior to conception, preferably with percutaneous balloon valvotomy. Careful control of heart rate and avoidance of hypovolemia, especially during labor and delivery, mitigate the risk of tachycardia and reduced ventricular filling times on cardiac function. The immediate postpartum period is a time of particular concern secondary to rapid volume shifts.

Aortic Stenosis People with aortic stenosis, a mean valve gradient <25 mmHg, and normal left ventricular function are likely to tolerate pregnancy. Symptomatic aortic stenosis or severe aortic stenosis with a peak gradient >50 mmHg should be evaluated for correction prior to pregnancy.

Mitral and Aortic Regurgitation The pregnancy-induced decrease in systemic vascular resistance reduces the risk of cardiac failure with these conditions, especially in those with chronic lesions. As a general rule, regurgitant lesions are well tolerated in pregnancy with acute onset of mitral or aortic regurgitation as an exception.

Mechanical Heart Valves People with mechanical heart valves are at high risk of valve thrombosis in pregnancy and warrant special consideration. Use of warfarin in pregnancy is limited to this population and avoided in the first trimester due to its association with fetal chondromalacia punctata. The risk of serious neonatal bleeding and associated neurologic injury persists throughout pregnancy, but the superiority of warfarin in preventing valve thrombosis merits its use in the second and third trimesters. Bridging from warfarin to heparin infusion at 36 weeks' gestation minimizes bleeding risk and facilitates neuraxial analgesia at delivery.

■ CONGENITAL HEART DISEASE

(See also Chap. 280) Reparative surgery has markedly increased the number of adults with congenital heart disease seeking pregnancy with a variety of disease-specific management considerations and outcomes. Repaired septal defects are comparatively low risk, whereas unrepaired septal lesions or repaired complex lesions such as tetralogy of Fallot may warrant additional surveillance by a pregnancy heart team to ensure that pregnancy is tolerated. People with complications in the setting of lower risk disease or with high-risk disease including uncomplicated Fontan circulation, systemic right ventricle, or cyanotic disease require delivery at a tertiary care center with subspecialty expertise. In *Eisenmenger's syndrome*, i.e., the combination of pulmonary hypertension with right-to-left shunting due to congenital abnormalities (Chap. 280), maternal and fetal deaths occur frequently, informing the recommendation for termination of pregnancy. The presence of a congenital cardiac lesion in the mother increases the risk of congenital cardiac disease in the newborn and is the basis for the recommendation to screen for fetal congenital heart disease with fetal echocardiography.

■ AORTOPATHY

The physiologic cardiovascular adaptations of pregnancy can predispose to aortic dissection, the majority of which occur in people with underlying aortopathies (Chap. 291). Although dissection may occur in the absence of aortic root enlargement in some genetic syndromes, sequential transthoracic echocardiographic (TTE) monitoring during pregnancy for evolution of the aortic root diameter is the mainstay of monitoring those at risk. Aortic magnetic resonance imaging (MRI) without gadolinium is useful to define thoracoabdominal aortic pathology in high-risk diseases. For most diseases, an aortic root diameter <40 mm portends a favorable pregnancy outcome, whereas a diameter >50 mm is an indication for prepregnancy repair. Beta blockers are a mainstay of therapy for most patients. A high clinical suspicion for dissection in patients presenting with chest pain is mandatory.

Marfan Syndrome (See also Chap. 425) This autosomal dominant disease is associated with an increased risk of aortic dissection and rupture. An aortic root diameter >40 mm is associated with an increased risk of aortic dissection, and an aortic root diameter >45 mm is an indication for surgical treatment. Operative vaginal delivery to limit the aortic wall stress associated with Valsalva should be considered for women with an aorta of 40–45 mm.

Ehlers-Danlos Syndrome (EDS) (See also Chap. 425) Vascular EDS (vEDS, type IV) is an autosomal dominant disease associated with an increased risk of uterine or vascular rupture that may cause death. For people with vEDS, pregnancy is relatively contraindicated because of this risk.

■ CARDIAC COMPLICATIONS IN PREGNANCY

Arrhythmias New-onset arrhythmias in healthy patients or patients with cardiac disease are common cardiac complications. Treatment is generally the same as in the nonpregnant patient, and fetal tolerance of medications such as beta blockers, calcium channel blockers, and common antiarrhythmics is acceptable. Pharmacologic or electric cardioversion may be performed to improve cardiac performance and reduce symptoms according to standard indications.

Peripartum Cardiomyopathy (PPCM) (Chap. 269) This uncommon but life-threatening condition should be considered in patients presenting in the third trimester or postpartum period with unexplained cardiogenic pulmonary edema. Treatment is directed toward symptomatic relief and improvement of cardiac function. The majority of patients recover completely. The subset of patients with left ventricular ejection fraction <30% are at high risk of progressive dilated cardiomyopathy. People with a history of PPCM may carry a genetic mutation that influences cardiac function. Approximately 10% of people with a history of PPCM have a truncating mutation in the *titin* gene encoding the sarcomere protein and experience higher rates of ongoing cardiomyopathy. Recurrence in a subsequent pregnancy is a risk, with a left ventricular ejection fraction of <50% 12 months postpartum as the threshold to advise against another pregnancy.

ENDOCRINE AND METABOLIC DISORDERS

The fetoplacental unit induces major metabolic changes to shunt glucose and amino acids to the fetus while the mother uses ketones and triglycerides to fuel her metabolic needs. The use of glucose by the fetus leads to a maternal state of accelerated ketosis during maternal fasting, characterized by lower glucose concentrations and higher hydroxybutyrate and acetoacetate levels. These metabolic changes are accompanied by maternal insulin resistance that increases during the course of pregnancy caused in part by placental production of steroids, a growth hormone variant, and placental lactogen.

■ DIABETES MELLITUS

(See also Chaps. 415–417) Pregnancy complicated by pregestational diabetes mellitus (pgDM), either type 1 or type 2 diabetes, complicates ~1% of pregnancies and is associated with increased maternal and perinatal morbidity and mortality rates. The prevalence of pgDM is greater among patients who are obese compared to normal weight (2.2 vs 0.5%). The metabolic changes of pregnancy can precipitate hyperglycemia requiring increased insulin needs, development of diabetic ketoacidosis, or hypoglycemia. Impaired glycemic control during the critical first 5–8 weeks of embryonic growth leads to an increased risk of spontaneous abortion and congenital malformations and highlights the importance of prepregnancy glycemic control. pgDM increases the risk of stillbirth, preeclampsia, and large for gestational age infants. Macrosomia increases the risk of shoulder dystocia and birth trauma, including brachial plexus injury and obstetric lacerations. Neonates are at risk of hypoglycemia, hyperbilirubinemia, polycythemia, and respiratory distress. An assessment of end-organ complications of pgDM including nephropathy, retinopathy, and neuropathy is essential to understanding the patient's risk profile.

■ GESTATIONAL DIABETES

Gestational diabetes mellitus (GDM) occurs in ~8% of pregnancies, and screening for GDM is a routine part of prenatal care, recommended by the U.S. Preventive Services Task Force with a B recommendation. The typical two-step strategy to diagnose GDM is performed at 24–28 weeks of gestation and involves the initial administration of a 50-g oral glucose challenge with a single serum glucose measurement at 60 min. Plasma glucose >7.2 mmol/L (>130 mg/dL) warrants administration of a 100-g oral glucose tolerance test (GTT) with plasma glucose measurements obtained in the fasting state and at 1, 2, and 3 h. Normal plasma glucose concentrations at these time points are <5.3 mmol/L (<95 mg/dL), <10 mmol/L (<180 mg/dL), <8.6 mmol/L (<155 mg/dL), and <7.8 mmol/L (<140 mg/dL) as the upper norms. Two elevated glucose values indicate a positive GTT diagnostic of GDM. GDM increases the risks of maternal and neonatal complications similar to those seen with pregestational diabetes. Treating GDM reduces the risk of preeclampsia, birth weight >4000 g, and shoulder dystocia.

TREATMENT
Diabetes Mellitus in Pregnancy

Preconception counseling to optimize glycemic control and assess for end-organ complications of pgDM is a cost-effective and evidence-based intervention. Guidelines encourage people with pgDM considering pregnancy to initiate insulin prior to pregnancy, targeting a preconception hemoglobin A_{1c} <6% to reduce the risk of fetal congenital malformation. Insulin is the preferred medical therapy for pgDM in pregnancy due to its safety profile and lower rates of treatment failure compared to oral hypoglycemic medications.

Once pregnancy is established, glucose control should be managed more intensively than in the nonpregnant state with assessment of blood glucose when fasting and either 1 or 2 h after a meal at a minimum. Fasting blood glucose levels should be maintained at <5.3 mmol/L (<95 mg/dL), with postprandial targets of <7.8 mmol/L (140 mg/dL) or <6.7 mmol/L (120 mg/dL) at 1 and 2 h, respectively. Continuous glucose monitoring is an evidence-based intervention to improve neonatal outcomes in type 1 diabetes. Sequential measurement of hemoglobin A_{1c} is of minimal utility for monitoring glucose control during pregnancy because of the higher rate of red blood cell turnover during pregnancy and resultant falsely low values. Average daily insulin needs increase from 0.7–0.8 units/kg in the first trimester to 0.8–1 units/kg in the second trimester and 0.9–1.2 units/kg in the third trimester. Most management strategies utilize a combination of basal insulin with short-acting insulin at mealtime or continued use of a prepregnancy insulin pump in appropriately selected patients.

Glycemic control may become more difficult to achieve as pregnancy progresses due to an increase in insulin resistance. Attention to glycemic control and frequent fetal surveillance including ultrasounds are mainstays of management in the third trimester. Ultrasound identification of a large for gestational age fetus or polyhydramnios on antenatal ultrasound can be indicators of suboptimal glycemic control. Tight glycemic control at delivery minimizes the risk of neonatal hypoglycemia due to fetal hyperinsulinemia caused by elevated maternal glucose levels. Infants of mothers with pgDM have higher rates of preterm birth, although preterm delivery is generally reserved for poor glycemic control, worsening maternal renal disease, or active proliferative retinopathy in addition to the usual obstetric indications. Induction of labor may be recommended in the early term period of 37–39 weeks of gestation. Cesarean delivery is reserved for cases of suspected macrosomia based on an estimated fetal weight of 4500 g or greater to minimize the risk of shoulder dystocia and associated birth trauma.

TREATMENT
Gestational Diabetes

Treatment of GDM begins with nutritional therapy to optimize normoglycemia and gestational weight gain, which is effective in the majority of women. Insulin is the preferred therapy for patients who exceed the aforementioned targets despite nutritional therapy. Metformin and glyburide are alternatives for patients who decline or cannot reliably take insulin. Contemporary data demonstrate lower mean birth weights, gestational weight gain, and rates of preeclampsia with metformin compared to both glyburide and insulin. The unknown long-term developmental and metabolic effects of metformin, including higher adiposity measurements in children exposed to metformin in utero, inform the preference for insulin.

GDM confers a 7- to 10-fold increase in the risk of developing type 2 diabetes later in life, with a 10% risk within 5 years of delivery. All people with GDM should have a 2-h 75-g GTT to screen for diabetes or impaired glucose tolerance 4–12 weeks following birth. The increased long-term risks of diabetes and cardiovascular disease and the need for regular follow-up with a primary care provider should be emphasized for all people with GDM. Following birth, exercise, weight loss, and treatment with metformin reduce the risk of developing diabetes in these at-risk patients.

■ OBESITY

(See also Chap. 414) Pregnant people who are obese have an increased risk for GDM, preeclampsia, cesarean delivery, congenital malformations, stillbirth, and neonatal death. A growing body of literature suggests the in utero effects of excess adipose tissue may cause changes in fetal metabolic programming that lead to adverse health outcomes in adult life. People contemplating pregnancy should attempt to attain a healthy weight prior to conception, recognizing that even a 10% reduction in weight may significantly reduce many of these risks. Those undergoing bariatric surgery should be counseled to avoid conception for 12–18 months after surgery until weight stabilizes. Bariatric surgery reduces the risks for some complications but requires increased laboratory surveillance for micronutrient deficiencies in pregnancy with appropriate supplementation. All women should be counseled to avoid weight gain in excess of the National Academy of Medicine guidelines (25–35 lb for normal weight, 15–25 lb for overweight, and 11–20 lb for obese women) with the knowledge that excess gestational weight gain increases the risk of macrosomia and cesarean delivery, independent of the presence of comorbid diabetes.

■ THYROID DISEASE

(See also Chap. 394) The estrogen-induced increase in thyroxine-binding globulin increases circulating levels of total T_3 and total T_4 in pregnancy. Placental human chorionic gonadotropin (hCG) directly stimulates the thyroid, causing an increase in free T_3 and T_4, especially in the first trimester. Interpretation of the measurement of free T_4, free T_3, and thyroid-stimulating hormone (TSH) should use trimester-specific ranges. People with a history of Graves' disease have an increased risk of fetal goiter and neonatal Graves' disease independent of treatment status due to the transplacental passage of thyroid-stimulating antibodies and stimulation of the fetal thyroid.

TREATMENT
Hyperthyroidism

Options for the treatment of symptomatic hyperthyroidism in pregnancy include beta blockers, propylthiouracil, and methimazole. Methimazole crosses the placenta to a greater degree than propylthiouracil and has been associated with fetal aplasia cutis. Propylthiouracil can be associated with maternal liver failure. Some experts recommend propylthiouracil in the first trimester and the option of continuing propylthiouracil or switching to methimazole thereafter. Radioiodine should not be used during pregnancy, either for scanning or for treatment, because of effects on the fetal thyroid.

TREATMENT
Hypothyroidism

The goal of therapy for hypothyroidism is to maintain the serum TSH in the normal range, and thyroxine is the drug of choice. The dose of thyroxine required to keep the TSH in the normal range rises during pregnancy. Since the increased thyroxine requirement occurs as early as the fifth week of pregnancy, one approach is to increase the thyroxine dose by 30% (two additional pills weekly) as soon as pregnancy is diagnosed and then adjust the dose according to TSH.

HEMATOLOGIC DISORDERS

Pregnancy has been described as a state of physiologic anemia. Red blood cell mass increases in pregnancy but to a lesser degree than plasma volume. This differential increase results in a lower hemoglobin concentration and so-called dilutional anemia. However, iron, folate, and vitamin B_{12} deficiencies are common causes of correctable anemia during pregnancy. Transfer of iron to the fetal compartment may explain why iron deficiency anemia is detected in ~11% of pregnant people in the third trimester. Screening for anemia is recommended in the first trimester and at 24 to 28 weeks' gestation. Measurement of ferritin can help detect iron deficiency before anemia is diagnosed. Iron replacement therapy is recommended in pregnancy.

Hemoglobinopathy screening is recommended for all pregnant women with testing of mean corpuscular volume, mean corpuscular hemoglobin, ferritin, and hemoglobin analysis **(Chap. 103)**. Hemoglobinopathies can be associated with increased maternal and fetal morbidity and mortality, with sickle cell disease as a particularly high-risk entity in pregnancy. Management is tailored to the specific hemoglobinopathy and is generally the same for both pregnant and nonpregnant women. Prenatal diagnosis of hemoglobinopathies in the fetus is readily available and should be discussed with prospective parents either prior to or early in pregnancy.

Thrombocytopenia occurs in 5–10% of pregnancies. The majority of cases are benign gestational thrombocytopenias, but the differential diagnosis should include immune thrombocytopenia **(Chap. 120)**, preeclampsia, and thrombotic thrombocytopenic purpura. Benign gestational thrombocytopenia is unlikely if the platelet count is <100,000/μL.

DEEP VENOUS THROMBOSIS AND PULMONARY EMBOLISM

(See also Chap. 290) Pregnancy is associated with venous stasis, endothelial injury, and a hypercoagulable state. During pregnancy, circulating coagulation factors II, VII, VIII, IX, X, and XII, fibrinogen, and von Willebrand factor increase, and protein S and antithrombin III decrease. Inherited thrombophilias and the presence of antiphospholipid antibodies increase the risk of venous thromboembolism (VTE) in pregnancy and often require prophylactic anticoagulation during pregnancy and the postpartum period to mitigate risk. Deep venous thrombosis (DVT) or pulmonary embolism (PE) occurs in about 1 in 500 pregnancies, with the highest risk in the postpartum state. In general, all diagnostic and therapeutic modalities afforded to the nonpregnant patient should be utilized in pregnancy.

TREATMENT

Venous Thromboembolism

Aggressive diagnosis and management of suspected DVT or PE optimize outcomes for both patient and fetus. Anticoagulant therapy with low-molecular-weight heparin (LMWH) or unfractionated heparin is indicated in pregnant people with VTE. Anticoagulants increase the risk of epidural hematoma in the setting of neuraxial analgesia (e.g., epidural or spinal anesthesia) in labor and must be withheld prior to placement. To ensure patient access to neuraxial analgesia in labor, prophylactic LMWH can be stopped 12 h before placement of an epidural catheter, whereas therapeutic LMWH can be discontinued for a full 24 h. Transition to unfractionated heparin as delivery approaches can shorten the time between anticoagulant administration and epidural placement. Unfractionated heparin can be used with an epidural catheter in place in those at highest risk of worsening thromboembolic disease.

NEOPLASIA

Cancer complicates ~1 in 1400 pregnancies. The four cancers that occur most commonly in pregnancy are cervical cancer, breast cancer, melanoma, and lymphoma. Also, cancer rates in pregnancy vary across the globe depending on local medical practices and exposures. In addition to cancers developing in other organs, gestational trophoblastic tumors can arise from the placenta.

Pregnancy has relatively little or no impact on the natural history of malignancies, despite the hormonal influences. People with a history of active or treated hormonally responsive breast cancer who choose to pursue pregnancy have similar rates of cancer-related events as those who do not. Spread of the cancer to the fetus (so-called vertical transmission) is exceedingly rare. However, managing cancer in a pregnant people is complex, with competing interests for mother and fetus. Generally, the management that optimizes maternal physiology is also best for the fetus. The best way to approach management of a pregnant woman with cancer is to ask, "What would one do in this clinical situation if she was not pregnant? Then, which, if any, aspect of those plans needs to be modified because she is pregnant?"

TREATMENT

Malignancy

Exposure of the developing fetus to therapeutic doses of ionizing radiation may cause adverse effects and is generally avoided during pregnancy.

Chemotherapy is associated with adverse fetal effects, but the significance of these depends on the specific agent and gestational age. Cytotoxic chemotherapy should virtually never be given in the first trimester due to risk of spontaneous abortion or malformation. If avoiding chemotherapy during this vulnerable time period could compromise maternal health, patients should be counseled about the role of therapeutic abortion to avoid serious neonatal sequelae.

A variety of single agents and combinations have been administered in the second and third trimesters, without a high frequency of toxic effects to the pregnancy or the fetus. Whether the association between chemotherapy and outcomes such as fetal growth restriction is due to the therapy or underlying malignancy is unknown. Literature supporting the short- and long-term neonatal safety of common agents is growing. For malignancies diagnosed closer to term or slowly progressive malignancies, delaying treatment until after delivery to avoid fetal exposure to chemotherapy may be desirable. If delaying therapy may compromise maternal prognosis, then treatment might be initiated after the first trimester of pregnancy with plans to deliver the fetus preterm to avoid excess cumulative exposure to chemotherapy. Neonatal prognosis is most closely linked to gestational age at delivery. Decisions regarding timing of delivery should contextualize this within the natural history of the disease, safety of the proposed treatment, and the person's goals of care.

NEUROLOGIC DISORDERS

Neurologic complaints such as headaches or neuropathies are common in pregnancy, and differentiating bothersome symptoms from life-threatening pathology is challenging **(Chap. 437)**. While most complaints are benign, *cerebrovascular accidents* (CVAs) should be high on the differential diagnosis and evaluated with MRI without gadolinium or a head computed tomography (CT) in cases of suspected stroke. Noncontrast MRI is more sensitive than CT in identifying early and small infarcts and cerebral venous thrombosis.

Exclusion of preeclampsia is important for any patient presenting with a headache after 20 weeks of gestation with a low threshold to assess for CVA due to the comparatively high prevalence in this population. Headache in preeclampsia can be associated with the *posterior reversible encephalopathy syndrome* (PRES), which is on the spectrum of *reversible cerebral vasoconstriction syndromes* (RCVS) that can present in pregnancy or the postpartum period with neurologic complaints. Peripheral nerve disorders associated with pregnancy include

Bell's palsy (idiopathic facial paralysis) **(Chap. 452)**, *carpal tunnel syndrome* (median nerve entrapment), or *meralgia paresthetica* (lateral femoral cutaneous nerve entrapment). *Restless leg syndrome* (RLS) is the most common peripheral nerve and movement disorder in pregnancy, affecting up to 20% of patients. If serum ferritin is low, oral iron supplementation is a first-line treatment option.

Pregnancy is safe for most people with neurologic disorders, with management considerations focusing on medication safety, the impact of pregnancy on the disease, and potential neonatal consequences. For those with *epilepsy* planning pregnancy, lamotrigine and levetiracetam are first-line monotherapies due to the abundance of safety data. Valproate is known to be associated with congenital malformations and should be discontinued in favor of another medication for people planning to conceive. Folic acid supplementation of 4 mg daily is recommended for those taking antiepileptic drugs (AEDs). Escalating doses of AEDs may be required due to increased clearance in pregnancy and guided by monthly monitoring of AED levels.

Patients with preexisting *multiple sclerosis* **(Chap. 455)** experience a gradual decrease in the risk of relapses as pregnancy progresses and, conversely, an increase in attack risk during the postpartum period. Prior to conception, disease-modifying therapies (DMTs) should be used to gain control of the disease. DMTs are typically withheld in pregnancy due to the decreased risk of relapse alongside limited neonatal safety data. Relapses should be treated with glucocorticoids, which may be given prophylactically after birth to reduce the risk of postpartum relapse. Finally, certain tumors, particularly pituitary adenoma **(Chap. 392)** and meningioma, may manifest during pregnancy because of accelerated growth, possibly driven by hormonal factors. Neuroimaging with noncontrast MRI may be required for women with a history of central nervous system (CNS) tumors to facilitate neuraxial analgesia.

GASTROINTESTINAL AND LIVER DISEASE

Up to 90% of pregnant people experience nausea and vomiting during the first trimester of pregnancy. *Hyperemesis gravidarum* (HG) is a severe form that prevents adequate fluid and nutritional intake and may require hospitalization to prevent dehydration and malnutrition. GDF15, which is produced by the placenta and is known to suppress appetite, may be involved in the pathogenesis of HG. Thiamine and folate supplementation and monitoring of electrolytes for evidence of refeeding syndrome should be considered in severe cases with evaluation for supplemental enteral nutrition in refractory disease.

Exacerbation of *inflammatory bowel disease* is common, and medical management, including the use of anti–tumor necrosis factor agents, parallels the nonpregnant state **(Chap. 337)**. Pregnancy is a risk factor for development or worsening of gallbladder disease such as cholelithiasis. This aggravation may be due to pregnancy-induced alteration in the metabolism of bile and fatty acids. *Intrahepatic cholestasis of pregnancy* is generally a third-trimester event presenting with profound pruritus and confirmed with an elevated level of bile acids with or without transaminitis. The association between cholestasis and stillbirth merits increased fetal surveillance and delivery before term. Symptoms can be improved with the use of ursodiol.

Acute fatty liver is a rare complication of pregnancy on the spectrum with HELLP syndrome and preeclampsia. Acute fatty liver of pregnancy is generally distinguished by markedly increased serum levels of bilirubin and ammonia and by hypoglycemia. Management of acute fatty liver of pregnancy includes delivery accompanied by supportive care. Most people recover within 7–10 days after delivery, but reports of fulminant liver failure requiring transplant underscore the importance of prompt diagnosis and management. The association between acute fatty liver in pregnancy and long-chain 3-hydroxyacyl CoA dehydrogenase (LCHAD) deficiency in the fetus inform the recommendation for close monitoring in the newborn with consideration of genetic testing.

All pregnant people should be screened for *hepatitis B virus* with HBsAg testing and *hepatitis C virus* **(Chap. 350)**. All infants receive hepatitis B vaccine, but infants born to mothers who are carriers of hepatitis B surface antigen should also receive hepatitis B immune globulin as soon after birth as possible to decrease the risk of vertical transmission. The presence of the hepatitis B E antigen in the mother and high viral load increase the risk of vertical transmission. Strategies to decrease vertical transmission of hepatitis C are limited to avoiding procedures (i.e., amniocentesis) that increase the risk. Postpartum referral to a specialist for consideration of potentially curative therapy is indicated.

INFECTIONS

BACTERIAL INFECTIONS

All pregnant patients are screened prenatally for syphilis, gonorrhea, and chlamydial infections, and the detection of any of these should result in prompt evaluation and treatment **(Chaps. 161, 187, and 194)**. Other than *bacterial vaginosis*, the most common bacterial infections during pregnancy involve the urinary tract **(Chap. 140)**. All pregnant people should be screened with a urine culture for *asymptomatic bacteriuria* at the first prenatal visit. Pregnancy is an indication for treatment of asymptomatic bacteriuria to avoid pyelonephritis. Progesterone-mediated ureteral and bladder smooth muscle relaxation, coupled with compression effects of the enlarging uterus, promote stasis and increase the risk of these conditions. Pregnant women who develop pyelonephritis should be treated with inpatient IV antibiotic administration due to the elevated risk of urosepsis and acute respiratory distress syndrome in pregnancy-associated pyelonephritis. Pregnant people with recurrent urinary tract infections or one episode of pyelonephritis should receive daily antibiotic suppressive treatment throughout the remainder of their pregnancy. Options for suppressive treatment include nitrofurantoin (50 or 100 mg at bedtime) or cephalexin (250 or 500 mg at bedtime).

VIRAL INFECTIONS

All pregnant people should be screened for hepatitis B virus, hepatitis C virus, and HIV as well as immunity to rubella and varicella. Screening for varicella immunity begins with obtaining a history of varicella immunization or infection. Testing for antibodies to varicella is reserved for cases where there is no history of prior immunization or infection, because the available varicella antibody test has low sensitivity and does not reliably detect antibodies in people who have been immunized.

Influenza **(See also Chap. 206)** Pregnant people with influenza are at increased risk of serious complications and death. All people who are pregnant or plan to become pregnant in the near future should receive inactivated influenza vaccine. The prompt initiation of antiviral treatment is recommended for pregnant people in whom influenza is suspected. Treatment can be reconsidered once the results of high-sensitivity tests are available. Prompt initiation of treatment lowers the risk of admission to an intensive care unit and death.

COVID-19 Pregnancy is a risk factor for severe COVID-19 infection. All unvaccinated people planning pregnancy or who are pregnant should receive a COVID-19 vaccine.

Cytomegalovirus Infection The most common cause of congenital viral infection in the United States is cytomegalovirus (CMV) **(Chap. 200)**. As many as 50–90% of women of childbearing age have antibodies to CMV, but only rarely does CMV reactivation result in neonatal infection. More commonly, primary CMV infection during pregnancy creates a risk of congenital CMV. No currently accepted treatment of CMV infection during pregnancy has been demonstrated to protect the fetus effectively. Severe CMV disease in the newborn is characterized most often by petechiae, hepatosplenomegaly, and jaundice. Chorioretinitis, microcephaly, intracranial calcifications, hepatitis, hemolytic anemia, and purpura may also develop. CNS involvement can result in the development of psychomotor, ocular, auditory, and dental abnormalities over time. People with a primary CMV infection should delay conception for 6 months.

Herpesvirus Infection **(See also Chap. 197)** The acquisition of genital herpes during pregnancy is associated with spontaneous abortion, prematurity, and congenital and neonatal herpes. Disseminated neonatal herpes carries with it high mortality and morbidity rates from CNS involvement. A cohort study of pregnant people without evidence of previous herpesvirus infection demonstrated that ~2% acquired a new herpesvirus infection during the pregnancy and ~60% of the newly infected women had no clinical symptoms. The risk of transmission was increased in those with infections closer to delivery. The risk of active genital herpes lesions at term can be reduced by prescribing acyclovir (400 mg three times daily) for the last 4 weeks of pregnancy to all people who had an episode of genital herpes during the pregnancy. Those with active genital herpes lesions at the time of presentation in labor should be delivered by cesarean section.

Rubella (See also Chap. 212) Rubella virus is a known teratogen; first-trimester rubella carries a high risk of fetal malformations, although the risk significantly decreases later in pregnancy. Congenital rubella infection may be diagnosed by percutaneous umbilical-blood sampling with the detection of IgM antibodies in fetal blood. All pregnant people or those considering pregnancy should be tested for their immune status to rubella.

Parvovirus Infection (See also Chap. 202) Infection with human parvovirus B19 may occur during pregnancy with the possibility of fetal transmission in the absence of immunity. It rarely causes significant maternal sequelae but can lead to erythroid aplasia in the fetus with resultant anemia, fetal hydrops, and death. Management includes screening for fetal anemia with Doppler assessment of the middle cerebral artery and consideration of intrauterine transfusion of red blood cells to the fetus to avoid the physiologic consequences of anemia while awaiting recovery of the fetal bone marrow.

HIV Infection (See also Chap. 208) The predominant cause of HIV infection in children is transmission of the virus from mother to newborn during the perinatal period. All pregnant women should be screened for HIV infection. Factors that increase the risk of mother-to-newborn transmission include high maternal viral load, low maternal CD4+ T-cell count, prolonged labor, prolonged duration of membrane rupture, and the presence of other genital tract infections, such as syphilis or herpes. Antiretroviral therapy (ART) has decreased the rate of perinatal transmission from 20% to ~1%. For those receiving antepartum ART, viral load guides the decision for vaginal versus cesarean delivery and need for adjunct intrapartum zidovudine. People with an undetectable viral load are at the lowest risk of transmission and require no additional therapy, whereas those without antepartum ART exposure or with viral loads >1000 copies/mL at delivery require IV zidovudine and a prelabor cesarean delivery, typically scheduled at 38 weeks. Cesarean delivery should be reserved for obstetric indications for women with ≥50 but ≤1000 copies/mL, and intrapartum zidovudine can be considered.

Zika Virus Zika virus (ZIKV) can be transmitted from mother to fetus throughout gestation and often results in severe microcephaly, CNS malformations, and fetal death. Symptomatic pregnant people with relevant epidemiologic exposure within 2 weeks of symptom onset should have serum and urine tested for ZIKV and dengue nucleic acids by nucleic acid amplification test and IgM serology. Sequential obstetrical ultrasound is recommended to assess for fetal growth and anomalies. Couples considering pregnancy should avoid travel to areas with known mosquito transmission of ZIKV.

VACCINATIONS

(See also Chap. 129) For rubella-nonimmune individuals contemplating pregnancy, measles-mumps-rubella vaccine should be administered, ideally at least 3 months prior to conception, but otherwise in the immediate postpartum period. All pregnant people should be vaccinated against influenza and COVID-19. Administration of one dose of the tetanus, diphtheria, and pertussis (Tdap) vaccine between 27 and 36 weeks of gestation is recommended to promote maternal IgG production and reduce the risk of neonatal pertussis due to transplacental passage of IgG.

MATERNAL MORTALITY

Maternal death is defined as death occurring during pregnancy or within 42 days of completion of pregnancy from a cause related to or aggravated by pregnancy, but not due to accident or incidental causes. The maternal mortality rate is the number of maternal deaths per 100,000 live births. From 1935 to 2007, the U.S. maternal mortality rate decreased from nearly 600/100,000 births to 12.7/100,000 births, rising thereafter. In 2021, the U.S. maternal mortality rate was 32.9/100,000 births. An increasingly complex patient population, with multiple comorbidities, and the COVID-19 pandemic likely contributed to the high rate of maternal death in 2021. There are significant racial and ethnic disparities in the maternal mortality rate, with a nearly fourfold increased risk of death for non-Hispanic black women compared to non-Hispanic white women (69.9 vs 26.6 deaths per 100,000 live births, respectively) (Chap. 11). Women over 40 years of age had a maternal mortality rate of 139 per 100,000 live births.

Pregnancy-related death is defined as the death of a person while pregnant or within 1 year of the end of pregnancy from any cause related to or aggravated by pregnancy. Cardiovascular disease, including cardiomyopathy, accounted for nearly a third of pregnancy-related deaths from 2014 to 2017 followed by infection, noncardiovascular medical conditions, hemorrhage, and thrombotic events. The relative contribution of medical disease to pregnancy-related death, coupled with knowledge that one in three pregnancy-related deaths occur 1 week to 1 year after delivery, highlights the role of the specialist in internal medicine in reducing maternal mortality.

In some countries in sub-Saharan Africa and southern Asia, the maternal mortality rate is >500/100,000 live births. The most common causes of maternal death in these countries are maternal hemorrhage, hypertensive disorders, infection, obstructed labor, and complications from unsafe pregnancy termination. The health interventions that have the greatest impact on maternal health include improving the following components of the health system: (1) access to contraceptive services in order to space births and limit total family size; (2) access to safe pregnancy termination; (3) presence of trained birth attendants at all deliveries; and (4) transportation to emergency obstetrical centers that can provide intensive medical and surgical services, including cesarean delivery. Maternal death is a global public health tragedy that could be mitigated with the application of modest resources.

SUMMARY

With improved diagnostic and therapeutic modalities as well as advances in the treatment of infertility, more patients with serious medical complications will be seeking to become pregnant and will require complex obstetric care. Improved outcomes of pregnancy in these people will be best attained by a multidisciplinary team of internists, maternal-fetal medicine (high-risk obstetrics) specialists, pediatricians, and anesthesiologists assembled to counsel these patients about the risks of pregnancy and to optimize interconception care. The importance of preconception counseling and the impact of events of pregnancy on lifelong disease cannot be overstated. It is the responsibility of all physicians caring for women in the reproductive age group to assess their patients' reproductive plans as part of their overall health evaluation and to integrate pregnancy-related diagnoses into their assessment of future risk for cardiovascular disease.

FURTHER READING

Crisafulli F et al: Variations of C3 and C4 before and during pregnancy in systemic lupus erythematosus: Association with disease flares and obstetric outcomes. J Rheumatol 50:1296, 2023.

Ford ND et al: Hypertensive disorders in pregnancy and mortality at delivery hospitalization-United States, 2017-2019. Morb Mortal Week Report 71:585, 2022.

Fu J et al: Increased risk of major congenital malformations in early pregnancy uses of angiotensin-converting enzyme inhibitors or angiotensin receptor blockers: A meta-analysis. Diabetes Metab Res Rev 37:e3453, 2021.

Hoyert DL: Maternal mortality rates in the United States, 2021. NCHS Health E-Stats. 2023. DOI: https://dx.doi.org/10.15620/cdc:124678.

Parikh NI et al: Adverse pregnancy outcomes and cardiovascular disease risk: Unique opportunities for cardiovascular disease prevention in women: A scientific statement from the American Heart Association. Circulation 143:e902, 2021.

Regitz-Zagrosek V et al: 2018 ESC guidelines for the management of cardiovascular diseases during pregnancy. Eur Heart J 39:3165, 2018.

Tita AT et al: Treatment for mild chronic hypertension during pregnancy. N Engl J Med 386:1781, 2022.

Verlohren S et al: Clinical interpretation and implementation of the sFlt-1/PlGF ratio in prediction, diagnosis and management of preeclampsia. Pregnancy Hypertens 27:42, 2022.

492

Medical Evaluation of the Patient Undergoing Noncardiac Surgery

Prashant Vaishnava, David I. Sahar, Kim A. Eagle

Cardiovascular and pulmonary complications continue to account for major morbidity and mortality in patients undergoing noncardiac surgery. Emerging evidence-based practices dictate that the internist should perform an individualized evaluation of the surgical patient to provide an accurate preoperative risk assessment and stratification that will guide optimal perioperative risk-reduction strategies. This chapter reviews cardiovascular and pulmonary preoperative risk assessment, emphasizing the goal-directed management of patients at elevated risk for adverse cardiovascular outcomes in the perioperative period. In addition, perioperative management of diabetes mellitus and prophylaxis of endocarditis and for venous thromboembolism are reviewed.

EVALUATION OF INTERMEDIATE- AND HIGH-RISK PATIENTS

Simple, standardized preoperative screening questionnaires, such as the one shown in **Table 492-1**, have been developed for the purpose of identifying patients at intermediate or high risk who may benefit from a more detailed clinical evaluation. Evaluation of such patients for surgery should always begin with a thorough history and physical examination and with a 12-lead resting electrocardiogram, in accordance with the American College of Cardiology/American Heart Association guidelines. The history should focus on symptoms of occult cardiac or pulmonary disease. The urgency of the surgery should be determined, as true emergency procedures are associated with unavoidably higher morbidity and mortality risk. Preoperative laboratory testing should be carried out only for specific clinical conditions, as noted during clinical examination. Thus, healthy patients of any age who are undergoing elective surgical procedures without coexisting medical conditions should not require any testing unless the degree of surgical stress may result in unusual changes from the baseline state.

TABLE 492-1 Standardized Preoperative Questionnaire[a]

1. Age, weight, height
2. Are you:
   Female and 55 years of age or older or male and 45 years of age or older?
   If yes, are you 70 years of age or older?
3. Do you take anticoagulant medications ("blood thinners")?
4. Do you have or have you had any of the following heart-related conditions?
   Heart disease
   Heart attack within the last 6 months
   Angina (chest pain)
   Irregular heartbeat
   Heart failure
5. Do you have or have you ever had any of the following?
   Rheumatoid arthritis
   Kidney disease
   Liver disease
   Diabetes
6. Do you get short of breath when you lie flat?
7. Are you currently on oxygen treatment?
8. Do you have a chronic cough that produces any discharge or fluid?
9. Do you have lung problems or diseases?
10. Have you or any blood member of your family ever had a problem other than nausea with any anesthesia?
    If yes, describe:
11. If female, is it possible that you are pregnant?
    Pregnancy test:
    Please list date of last menstrual period:

[a] University of Michigan Health System patient information report. Patients who answer yes to any of questions 2–9 should receive a more detailed clinical evaluation.

Source: Reproduced with permission from KK Tremper, P Benedict: Paper "Preoperative Computer." Anesthesiology 92:1212, 2000.

PREOPERATIVE CARDIAC RISK ASSESSMENT

A stepwise approach to cardiac risk assessment and stratification in patients undergoing noncardiac surgery is illustrated in **Fig. 492-1**. The evaluation begins with characterization of the combined surgical and clinical risk into categories of low (<1%) and elevated risk for major adverse cardiovascular events (MACEs). Select surgeries are associated with very low risk for MACE; these surgeries and procedures include select ophthalmologic surgeries (e.g., cataract surgery), select endoscopic procedures, and select superficial procedures. Patients undergoing these low-risk procedures should proceed to surgery without further testing. Clinical risk may be estimated with the American College of Surgeons' National Surgical Quality Improvement Program (NSQIP) risk calculator (*http://www.riskcalculator.facs.org*) or with calculation of the Revised Cardiac Risk Index (RCRI). Additional tools include the Surgical Outcome Risk Tool (SORT) or the American University of Beirut (AUB)-HAS2 Cardiovascular Risk Index. The former is based, in part, on the American Society of Anesthesiologists Physical Status (ASA-PS) grade; the latter provides for estimation of 30-day death, myocardial infarction (MI), or stroke risk.

Previous studies have compared several cardiac risk indices. The American College of Surgeons' NSQIP prospective database has identified five predictors of perioperative MI and cardiac arrest based on increasing age, American Society of Anesthesiologists class, type of surgery, dependent functional status, and abnormal serum creatinine level. However, given its accuracy and simplicity, the RCRI **(Table 492-2)** is often the favored risk index. The RCRI relies on the presence or absence of six identifiable predictive factors: high-risk surgery, ischemic heart disease, congestive heart failure, cerebrovascular disease, diabetes mellitus treated with insulin, and renal insufficiency with a creatinine >2.0 mg/dL. Each of these predictors is assigned one point. The risk of major cardiac events—defined as MI, pulmonary edema, ventricular fibrillation or primary cardiac arrest, and complete heart block—can then be predicted. Based on the presence of none, one, two, three, or more of these clinical predictors, the rate of development of one of these four major cardiac events is estimated to be 0.4%, 0.9%, 7%, and 11%, respectively **(Fig. 492-2)**. The clinical utility of the RCRI is to identify patients with three or more predictors who are at very high risk (≥11%) for cardiac complications and who may benefit from further risk stratification with noninvasive cardiac testing, initiation of preoperative preventive medical management, or avoidance of surgery.

For patients at elevated combined clinical and surgical risk for MACE, the stepwise perioperative cardiac assessment for coronary

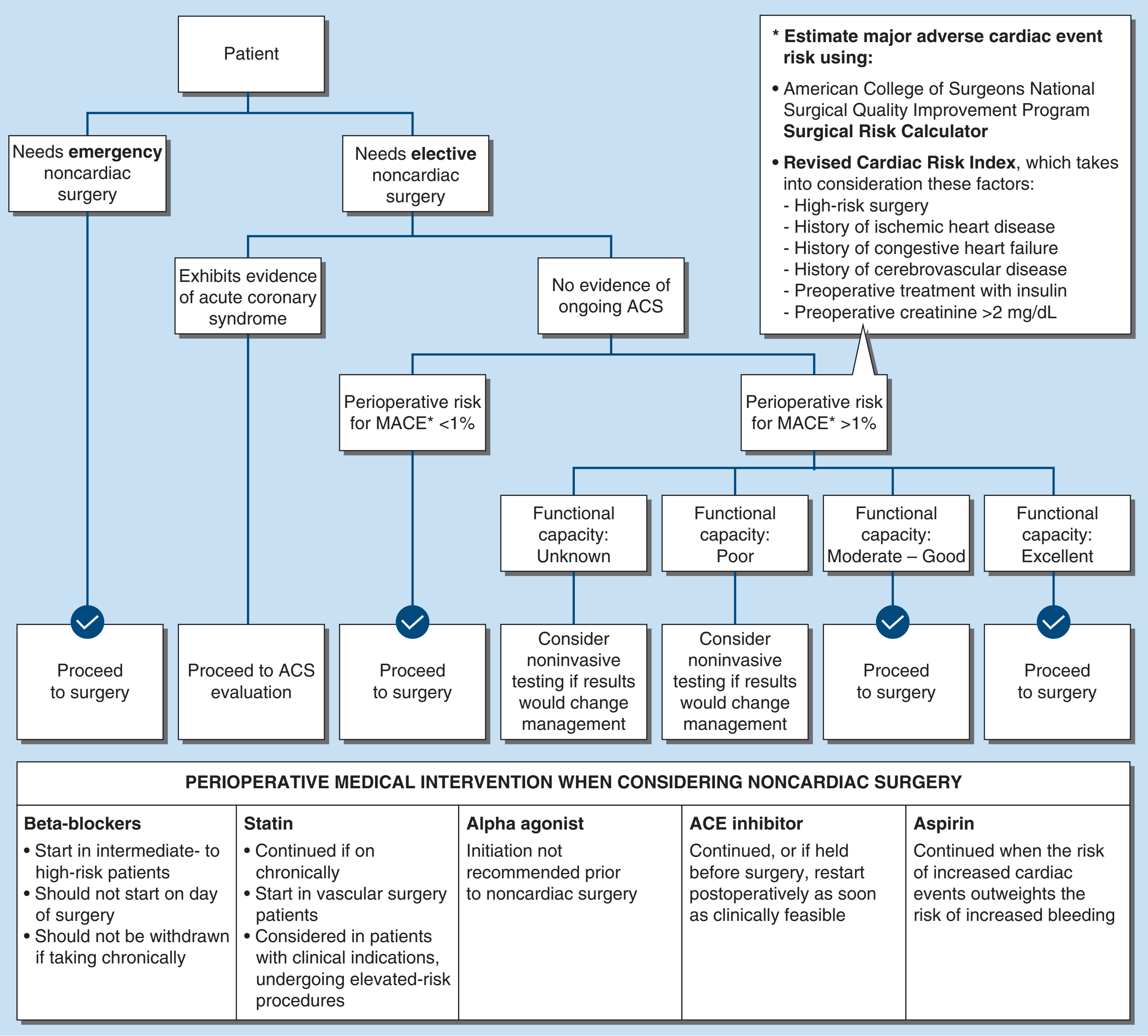
Patient
Needs **emergency** noncardiac surgery
Needs **elective** noncardiac surgery
Exhibits evidence of acute coronary syndrome
No evidence of ongoing ACS
*** Estimate major adverse cardiac event risk using:**
• American College of Surgeons National Surgical Quality Improvement Program **Surgical Risk Calculator**
• **Revised Cardiac Risk Index**, which takes into consideration these factors:
- High-risk surgery
- History of ischemic heart disease
- History of congestive heart failure
- History of cerebrovascular disease
- Preoperative treatment with insulin
- Preoperative creatinine >2 mg/dL
Perioperative risk for MACE* <1%
Perioperative risk for MACE* >1%
Functional capacity: Unknown
Functional capacity: Poor
Functional capacity: Moderate – Good
Functional capacity: Excellent
Proceed to surgery
Proceed to ACS evaluation
Proceed to surgery
Consider noninvasive testing if results would change management
Consider noninvasive testing if results would change management
Proceed to surgery
Proceed to surgery

PERIOPERATIVE MEDICAL INTERVENTION WHEN CONSIDERING NONCARDIAC SURGERY

Beta-blockers	Statin	Alpha agonist	ACE inhibitor	Aspirin
• Start in intermediate- to high-risk patients • Should not start on day of surgery • Should not be withdrawn if taking chronically	• Continued if on chronically • Start in vascular surgery patients • Considered in patients with clinical indications, undergoing elevated-risk procedures	Initiation not recommended prior to noncardiac surgery	Continued, or if held before surgery, restart postoperatively as soon as clinically feasible	Continued when the risk of increased cardiac events outweighs the risk of increased bleeding

FIGURE 492-1 Composite algorithm for cardiac risk assessment and stratification in patients undergoing noncardiac surgery. Preoperative evaluation involves a stepwise clinical evaluation. Those individuals requiring emergency surgery should proceed without further risk stratification. Acute coronary syndrome (step 2) should be evaluated and treated according to goal-directed medical therapy. For patients awaiting nonemergent surgeries and without acute coronary syndrome, perioperative risk is a combination of clinical and surgical risk. Select procedures and surgeries (e.g., select endoscopic procedures) are associated with low (<1%) perioperative risk, and no further clinical testing is generally necessary. For those procedures associated with elevated risk, an assessment of functional capacity informs the decision for further testing. Those individuals with moderate or greater functional capacity do not require further testing and should proceed to surgery. Individuals with poor or unknown functional capacity may require pharmacologic stress testing if it would change decision-making or perioperative care. ACE, angiotensin-converting enzyme; ACS, acute coronary syndrome; MACE, major adverse cardiovascular event. *(Reproduced with permission from AY Patel et al: Cardiac risk of noncardiac surgery. J Am Coll Cardiol 66:2140, 2015.)*

artery disease (CAD) proceeds with consideration of functional capacity. Participation in activities of daily living offers an expression of functional capacity, often expressed in terms of metabolic equivalents (METs). For predicting perioperative events, poor exercise tolerance has been defined as the inability to walk four blocks or climb two flights of stairs at a normal pace or to meet a MET level of 4 (e.g., carrying objects of 15–20 lb. or playing golf or doubles tennis) because of the development of dyspnea, angina, or excessive fatigue **(Table 492-3)**. Patients with moderate or greater (≥4 METs) functional capacity (e.g., climbing up a flight of stairs, walking up a hill, or walking on level ground at 4 mph) generally should not undergo further noninvasive cardiac testing prior to elective noncardiac surgery. Those patients with poor (<4 METs) or unknown functional capacity should undergo pharmacologic stress testing if the results of such testing would impact decision-making or perioperative care.

PREOPERATIVE NONINVASIVE CARDIAC TESTING FOR RISK STRATIFICATION

There is little evidence to support widespread application of preoperative noninvasive cardiac testing for all patients undergoing major surgery. The current paradigm to guide the need for noninvasive cardiac testing is to perform such testing in patients with poor or unknown capacity if it would alter clinical management or modify perioperative care. Options for pharmacologic stress testing include dobutamine stress echocardiography or myocardial perfusion imaging with coronary vasodilator stress (dipyridamole, adenosine, or regadenoson) with thallium-201 and/or technetium-99m. Similarly, there is a limited

TABLE 492-2 Clinical Markers Included in the Revised Cardiac Risk Index

High-Risk Surgical Procedures
Vascular surgery (except carotid endarterectomy)
Major intraperitoneal or intrathoracic procedures
Ischemic Heart Disease
History of myocardial infarction
Current angina considered to be ischemic
Requirement for sublingual nitroglycerin
Positive exercise test
Pathological Q waves on ECG
History of PCI and/or CABG with current angina considered to be ischemic
Congestive Heart Failure
Left ventricular failure by physical examination
History of paroxysmal nocturnal dyspnea
History of pulmonary edema
S_3 gallop on cardiac auscultation
Bilateral rales on pulmonary auscultation
Pulmonary edema on chest x-ray
Cerebrovascular Disease
History of transient ischemic attack
History of cerebrovascular accident
Diabetes Mellitus
Treatment with insulin
Chronic Renal Insufficiency
Serum creatinine >2 mg/dL

Abbreviations: CABG, coronary artery bypass grafting; ECG, electrocardiogram; PCI, percutaneous coronary intervention.

Source: Adapted from TH Lee et al: Circulation 100:1043, 1999.

TABLE 492-3 Assessment of Cardiac Risk by Functional Status

Risk	Higher	• Has difficulty with adult activities of daily living
	↑	• Cannot walk four blocks or up two flights of stairs or does not meet a MET level of 4
		• Is inactive but has no limitations
	↓	• Is active: easily does vigorous tasks
	Lower	• Performs regular vigorous exercises

Abbreviation: MET, metabolic equivalent.

Source: From LA Fleisher et al: Circulation 116:1971, 2007.

role for perioperative coronary computed tomography angiography (CCTA) in patients undergoing noncardiac surgery. While some investigators have shown in observational studies that CCTA may improve prediction of MACEs perioperatively when compared to RCRI, there are few data to demonstrate that such improved prognostication translates to improved outcomes. Furthermore, coronary revascularization before noncardiac surgery is not recommended for the express purpose of reducing perioperative cardiac events. That said, revascularization before noncardiac surgery should be considered in patients if it would be indicated regardless of the surgery planned and instead according to clinical practice guidelines. In the Coronary Artery Revascularization Prophylaxis trial, there were no differences in perioperative and long-term cardiac outcomes with or without preoperative coronary revascularization; of note, patients with left main disease were excluded.

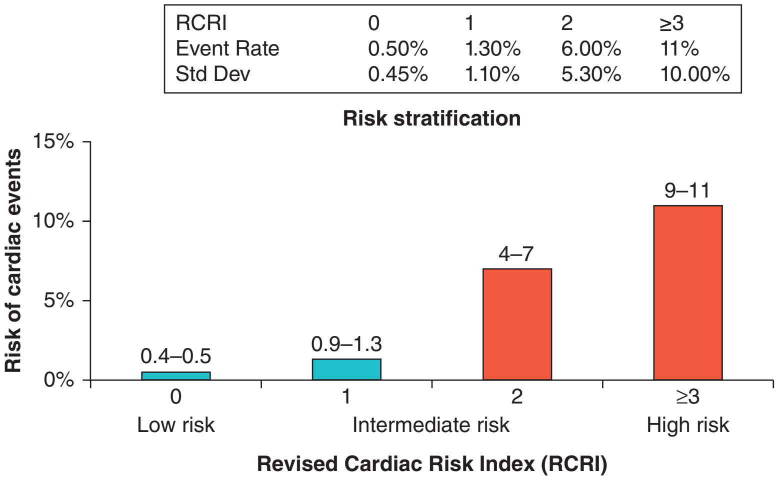

FIGURE 492-2 Risk stratification based on the Revised Cardiac Risk Index; derivation and prospective validation of a simple index for prediction of cardiac risk in patients undergoing major noncardiac surgery. Cardiac events include myocardial infarction, pulmonary edema, ventricular fibrillation, cardiac asystole, and complete heart block. *(Adapted from TH Lee et al: Circulation 100:1043, 1999.)*

RISK MODIFICATION: PREVENTIVE STRATEGIES TO REDUCE CARDIAC RISK

Perioperative Coronary Revascularization *Prophylactic* coronary revascularization with either coronary artery bypass grafting (CABG) or percutaneous coronary intervention (PCI) provides no short- or mid-term survival benefit for patients *without* left main CAD or three-vessel CAD in the presence of poor left ventricular systolic function and is *not recommended* for patients with stable CAD before noncardiac surgery. Although PCI is associated with lower procedural risk than is CABG in the perioperative setting, the placement of a coronary artery stent soon before noncardiac surgery may increase the risk of bleeding during surgery if dual antiplatelet therapy (DAPT) (aspirin and P2Y$_{12}$) is administered; moreover, stent placement shortly before noncardiac surgery increases the perioperative risk of MI and cardiac death due to stent thrombosis if such therapy is withdrawn prematurely **(Chap. 287)**. It is recommended that, if possible, elective noncardiac surgery be delayed for 6 months after elective PCI and 12 months after acute coronary syndrome. Contemporary stent platforms allow for greater flexibility in the earlier interruption of DAPT; accordingly, time sensitive noncardiac surgery could be a consideration as soon as 1 month of treatment with DAPT has passed. For patients who *must* undergo noncardiac surgery early (>14 days) after PCI, balloon angioplasty without stent placement appears to be a reasonable alternative because DAPT is not necessary in such patients.

PERIOPERATIVE PREVENTIVE MEDICAL THERAPIES The goal of perioperative preventive medical therapies with β-adrenergic antagonists, hydroxymethylglutaryl-coenzyme A (HMG-CoA) reductase inhibitors (statins), and antiplatelet agents is to reduce perioperative adrenergic stimulation, ischemia, and inflammation, all of which are heightened during the perioperative period.

β-ADRENERGIC ANTAGONISTS The use of perioperative beta blockade should be based on a thorough assessment of a patient's perioperative clinical and surgery-specific cardiac risk (e.g., as with the RCRI). The paradigm for beta blockade in the perioperative period has shifted in recent years owing, first, to the publication of the PeriOperative Ischemic Evaluation (POISE) trial demonstrating that, while perioperative beta blockade reduces the perioperative risk for MI, this is at the expense of increased death and stroke. Regarding POISE, this trial has been scrutinized for the use of an excessive dose of beta blocker in the perioperative period and one that may not be reflective of clinical practice, nor one that was titrated in the days or weeks preceding the procedure or surgery. Second, research misconduct has discredited the Dutch Echocardiographic Cardiac Risk Evaluation Applying Stress Echocardiography (DECREASE) family of studies, which previously contributed to the bedrock of data supporting the use of perioperative beta blockade but have now been retracted.

Current guidelines emphasize the following key points:

1. Continuation of beta blockade in patients undergoing surgery and who have been receiving such therapy chronically.
2. Avoidance of beta-blocker withdrawal or initiation on the day of surgery.
3. Consideration of initiation of beta-blocker therapy perioperatively (ideally far enough in advance to assess safety and tolerability) in very select high-risk patients, namely, those with intermediate- or high-risk ischemia or three or more RCRI risk factors.

HMG-COA REDUCTASE INHIBITORS (STATINS) A number of prospective and retrospective studies support the perioperative prophylactic use of statins for reduction of cardiac complications in patients with established atherosclerosis. For patients undergoing noncardiac surgery and currently taking statins, statin therapy *should be continued* to reduce perioperative cardiac risk. Initiation of statin therapy is *reasonable* for patients undergoing vascular surgery independent of clinical risk. There is equipoise about the perioperative initiation of statin therapy in patients undergoing other elevated-risk procedures. In the Lowering the Risk of Operative Complications Using Atorvastatin Loading Dose (LOAD) randomized, placebo-controlled trial, the use of 80 mg of atorvastatin within 18 h before surgery and then followed by 40 mg daily for 7 days in statin-naïve patients (24% of whom had a history of cardiovascular disease) did not reduce the risk of major adverse events.

ANGIOTENSIN-CONVERTING ENZYME (ACE) INHIBITORS It is important to maintain continuity of therapy with inhibitors of the renin-angiotensin-aldosterone system (when such therapy is used for the treatment of heart failure or hypertension).

ORAL ANTIPLATELET AGENTS The 4- to 6-week period following implantation of an intracoronary stent (bare metal or drug eluting) constitutes the period of time of greatest risk for the development of stent thrombosis. If possible, noncardiac surgery should be avoided in this vulnerable period. The duration of DAPT thereafter is dictated by the circumstances in which PCI was performed and whether the indication was stable ischemic heart disease or acute coronary syndrome. For the former among patients treated with a drug-eluting stent (DES), DAPT should be given for at least 6 months. For the latter, DAPT should be given for at least 12 months. However, DAPT may be interrupted to allow for noncardiac surgery 30 days after bare metal stent (BMS) and 6 months after DES, respectively. Elective, noncardiac surgery should be delayed for 5 days since the last dose of clopidogrel; 7 days since the last dose of prasugrel; and 3–5 days since the last dose of ticagrelor. The use of cangrelor, an intravenous reversible $P2Y_{12}$ receptor antagonist, may be an appealing bridging strategy, although studies of its use in those undergoing cardiac and noncardiac surgery are limited. If $P2Y_{12}$ inhibitor therapy (clopidogrel, prasugrel, or ticagrelor) is interrupted or discontinued in patients who have received intracoronary stents, aspirin should be continued perioperatively (save select circumstances where the risk of bleeding may be catastrophic as in neurosurgical or spinal procedures) and the $P2Y_{12}$ receptor inhibitor should be restarted as soon as possible postoperatively. Decisions surrounding antiplatelet management in the perioperative setting among patients who have received intracoronary stents are complex and should involve multidisciplinary decision-making.

α_2 AGONISTS Based on the results of POISE-2 (a large multicenter, international, blinded randomized clinical trial of aspirin and clonidine), α_2 agonists for prevention of cardiac events are not recommended in patients who are undergoing noncardiac surgery. In this trial, clonidine increased the rate of nonfatal cardiac arrest and clinically important hypotension, while reducing the rate of death or nonfatal MI.

CALCIUM CHANNEL BLOCKERS Evidence is lacking to support the use of calcium channel blockers as a prophylactic strategy to decrease perioperative risk in major noncardiac surgery.

SODIUM-GLUCOSE COTRANSPORTER 2 INHIBITORS The use of sodium-glucose cotransporter 2 (SGLT-2) inhibitors has grown in recent years owing to their beneficial impact in patients with and without type 2 diabetes mellitus and with heart failure with either reduced or preserved ejection fraction. However, euglycemic diabetic ketoacidosis is a known, but rare, complication with this therapy and one that may be precipitated by changes and fasting in the perioperative state. Accordingly, SGLT-2 inhibitor therapy should be interrupted for at least 3–4 days prior to scheduled noncardiac surgery.

TABLE 492-4 Gradation of Mortality Risk of Common Noncardiac Surgical Procedures

Higher	• Emergent major operations, especially in the elderly • Aortic and other noncarotid major vascular surgery (endovascular and nonendovascular) • Prolonged surgery associated with large fluid shift and/or blood loss
Intermediate	• Major thoracic surgery • Major abdominal surgery • Carotid endarterectomy surgery • Head/neck surgery • Orthopedic surgery • Prostate surgery
Lower	• Eye, skin, and superficial surgery • Endoscopic procedures

Source: Reproduced with permission from LA Fleisher et al: ACC/AHA 2007 Guidelines on perioperative cardiovascular evaluation and care for noncardiac surgery. Circulation 116:1971, 2007.

ANESTHETICS Mortality risk is low with safe delivery of modern anesthesia, especially among low-risk patients undergoing low-risk surgery **(Table 492-4)**. Inhaled anesthetics have predictable circulatory and respiratory effects: all decrease arterial pressure in a dose-dependent manner by reducing sympathetic tone and causing systemic vasodilation, myocardial depression, and decreased cardiac output. Inhaled anesthetics also cause respiratory depression, with diminished responses to both hypercapnia and hypoxemia, in a dose-dependent manner; in addition, these agents have a variable effect on heart rate. Prolonged residual neuromuscular blockade also increases the risk of postoperative pulmonary complications due to reduction in functional residual lung capacity, loss of diaphragmatic and intercostal muscle function, atelectasis, and arterial hypoxemia from ventilation–perfusion mismatch.

Several meta-analyses have shown that rates of pneumonia and respiratory failure are lower among patients receiving neuroaxial anesthesia (epidural or spinal) rather than general anesthesia. However, there were no significant differences in cardiac events between the two approaches. Evidence from a meta-analysis of randomized controlled trials supports postoperative epidural analgesia for >24 h for the purpose of pain relief. However, the risk of epidural hematoma in the setting of systemic anticoagulation for venous thromboembolism prophylaxis (see below) and postoperative epidural catheterization must be considered.

PREOPERATIVE PULMONARY RISK ASSESSMENT

Perioperative pulmonary complications occur frequently and lead to significant morbidity and mortality. Clinical practice guidelines recommend the following:

1. All patients undergoing noncardiac surgery should be assessed for risk of pulmonary complications **(Table 492-5)**.
2. While select studies have suggested that quitting smoking shortly before surgery increases the risk for postoperative complications through increased sputum production and/or decreased cough, meta-analysis of the available data has challenged this, and all patients should be advised of the imperative to stop smoking presurgically.
3. Patients undergoing emergency or prolonged (3–4 h) surgery; aortic aneurysm repair; vascular surgery; major abdominal, thoracic, neurologic, head, or neck surgery; and general anesthesia should

TABLE 492-5 Predisposing Risk Factors for Pulmonary Complications

1. Upper respiratory tract infection: cough, dyspnea
2. Age >60 years
3. Chronic obstructive pulmonary disease
4. Cigarette use
5. American Society of Anesthesiologists Class ≥2
6. Functional dependence
7. Congestive heart failure
8. Serum albumin <3.5 g/dL
9. Obstructive sleep apnea
10. Impaired sensorium (confusion, delirium, or mental status changes)
11. Abnormal findings on chest examination
12. Alcohol use
13. Weight loss
14. Spirometry threshold before lung resection
    a. FEV$_1$ <2 L
    b. MVV <50% of predicted
    c. PEF <100 L or 50% predicted value
    d. PCO$_2$ ≥45 mmHg
    e. PO$_2$ ≤50 mmHg

Abbreviations: FEV$_1$, forced expiratory volume in 1 s; MVV, maximal voluntary ventilation; PCO$_2$, partial pressure of carbon dioxide; PEF, peak expiratory flow rate; PO$_2$, partial pressure of oxygen.

Source: A Qaseem et al: Ann Intern Med 144:575, 2006. Modified from GW Smetana et al: Ann Intern Med 144:581, 2006, and from DN Mohr et al: Postgrad Med 100:247, 1996.

TABLE 492-6 Risk Modification to Reduce Perioperative Pulmonary Complications

Preoperatively

- Smoking cessation
- Training in proper lung expansion techniques
- Inhalation bronchodilator and/or steroid therapy, when indicated
- Control of infection and secretion, when indicated
- Weight reduction, when appropriate

Intraoperatively

- Limited duration of anesthesia
- Avoidance of long-acting neuromuscular blocking drugs, when indicated
- Prevention of aspiration and maintenance of optimal bronchodilation

Postoperatively

- Optimization of inspiratory capacity maneuvers, with attention to:
- Mobilization of secretions
- Early ambulation
- Encouragement of coughing
- Selective use of a nasogastric tube
- Adequate pain control without excessive narcotics

Source: From VA Lawrence et al: Ann Intern Med 144:596, 2006, and WF Dunn, PD Scanlon: Mayo Clin Proc 68:371, 1993.

be considered to be at elevated risk for postoperative pulmonary complications.

4. Patients at higher risk of pulmonary complications should undergo incentive spirometry, deep breathing exercises, cough encouragement, postural drainage, percussion and vibration, suctioning and ambulation, intermittent positive-pressure breathing, continuous positive airway pressure, and selective use of a nasogastric tube for postoperative nausea, vomiting, or symptomatic abdominal distention to reduce postoperative risk. Multiple pulmonary risk indices are available to estimate the postoperative risk of respiratory failure, pneumonia, and other pulmonary complications; among these is the ARISCAT risk index, which accounts for the following seven risk factors: age, low preoperative oxygen saturation, respiratory infection within the preceding month, upper abdominal or thoracic surgery, surgery lasting >2 h, hemoglobin <10 g/dL, and emergency surgery **(Table 492-6)**.
5. Preoperative spirometry and chest radiography should not be used routinely for predicting risk of postoperative pulmonary complications but may be appropriate for patients with chronic obstructive pulmonary disease or asthma.
6. Spirometry is of value before lung resection in determining candidacy for coronary artery bypass; however, it does not provide a spirometric threshold for extrathoracic surgery below which the risks of surgery are unacceptable.
7. Pulmonary artery catheterization, administration of total parenteral nutrition (as opposed to no supplementation), or total enteral nutrition has no consistent benefit in reducing postoperative pulmonary complications.

PERIOPERATIVE MANAGEMENT AND PROPHYLAXIS

DIABETES MELLITUS

(See also Chaps. 415-417) Many patients with diabetes mellitus have significant symptomatic or asymptomatic CAD and may have silent myocardial ischemia due to autonomic dysfunction. Intensive (vs lenient) glycemic control in the perioperative period is generally not associated with improved outcomes and may increase the risk of hypoglycemia. Practice guidelines advocate a target glucose range of 100–180 mg/dL in the perioperative period. Oral hypoglycemic agonists should not be given on the morning of surgery. Perioperative hyperglycemia should be treated with IV infusion of short-acting insulin or subcutaneous sliding-scale insulin. Patients whose diabetes is diet controlled may proceed to surgery with close postoperative monitoring.

INFECTIVE ENDOCARDITIS

(See also Chap. 133) Prophylactic antibiotics should be administered to the following patients before dental procedures that involve manipulation of gingival tissue, manipulation of the periapical region of teeth, or perforation of the oral mucosa: those with prosthetic cardiac valves (including transcatheter prosthetic valves); prosthetic material used in valve repair (annuloplasty ring or artificial chord); previous infective endocarditis; cardiac transplant recipients with valvular regurgitation from a structurally abnormal valve; and unrepaired cyanotic congenital heart disease or repaired congenital heart disease, with residual shunts or valvular regurgitation at the site adjacent to the site of a prosthetic patch or prosthetic device.

AORTIC STENOSIS

Previous American College of Cardiology/American Heart Association guidelines have cautioned against surgery in patients with severe aortic stenosis, citing a 10% mortality risk. More recent guidance, rooted in contemporary data, offers greater latitude for noncardiac surgery in appropriately selected patients with severe aortic stenosis. In an analysis of patients undergoing moderate- or high-risk surgery at the Mayo Clinic from 2000 to 2010, there was no significant difference in 30-day mortality between those with severe aortic stenosis and matched controls (5.9 vs 3.1%, $p = .13$); however, those with severe aortic stenosis had more MACEs (18.8 vs 10.5%, $p = .01$), mainly due to heart failure. In sum, severe aortic stenosis is associated with adverse outcomes in patients undergoing noncardiac surgery; however, in contemporary cohorts, this risk is less than has previously been stated. Patients with severe symptomatic aortic stenosis should generally undergo aortic valve intervention (surgical aortic valve replacement or transcatheter aortic valve implantation) if noncardiac surgery can be deferred. Asymptomatic patients with severe aortic stenosis and preserved ejection fraction can generally safely undergo low- or intermediate-risk noncardiac surgery. Balloon valvotomy is usually not recommended but may serve a role in the minority of patients who need "bridging" to a necessary surgery or procedure.

VENOUS THROMBOEMBOLISM

(See also Chap. 290) Perioperative prophylaxis of venous thromboembolism should follow established guidelines of the American College of

Chest Physicians. Aspirin is not supported as a single agent for thromboprophylaxis. Low-dose unfractionated heparin (≤5000 units SC bid), low-molecular-weight heparin (e.g., enoxaparin, 30 mg bid or 40 mg qd), or a pentasaccharide (fondaparinux, 2.5 mg qd) is appropriate for patients at moderate risk; unfractionated heparin (5000 units SC tid) is appropriate for patients at high risk. The use of direct oral anticoagulants may be an alternative to the use of prophylactic doses of low-dose unfractionated heparin and low-molecular-weight heparin; among patients immobilized after nonmajor orthopedic surgery, rivaroxaban 10 mg once daily when compared to enoxaparin was associated with a decrease in venous thromboembolic events, without a significant change in bleeding. Graduated compression stockings and pneumatic compression devices are useful supplements to anticoagulant therapy or as alternatives to anticoagulant therapy in patients at excessive bleeding risk.

■ FURTHER READING

Eagle KA et al: Perioperative cardiovascular care for patients undergoing noncardiac surgical intervention. JAMA Intern Med 175:835, 2015.

Fleisher LA et al: 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 130:e278, 2014.

Halvorsen S et al: 2022 ESC Guidelines on cardiovascular assessment and management of patients undergoing non-cardiac surgery: Developed by the task force for cardiovascular assessment and management of patients undergoing non-cardiac surgery of the European Society of Cardiology (ESC) Endorsed by the European Society of Anesthesiology and Intensive Care. Eur Heart J 39:3826, 2022.

Hwang JW et al: Assessment of perioperative cardiac risk of patients undergoing noncardiac surgery using coronary computed tomographic angiography. Circ Cardiovasc Imaging 8:e002582, 2015.

Levine GN et al: 2016 ACC/AHA guideline focused update on duration of dual antiplatelet therapy in patients with coronary artery disease. A report of the American College of Cardiology/American Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 68:1082, 2016.

Otto CM et al: 2020 ACC/AHA guideline for the management of patients of valvular heart disease: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 143:e72, 2021.

Smilowitz NR, Berger JS: Perioperative cardiovascular risk assessment and management for noncardiac surgery: A review. JAMA 324:279, 2020.

Tashiro T et al: Perioperative risk of major non-cardiac surgery in patients with severe aortic stenosis: A reappraisal in contemporary practice. Eur Heart J 35:2372, 2014.

493 Point-of-Care Ultrasound

Wilma Chan, Nilam J. Soni, Paul H. Mayo

DEFINITION

Point-of-care ultrasound (POCUS) is defined as the acquisition, interpretation, and clinical integration of ultrasonographic views by a treating clinician in real time at the patient's bedside. POCUS is distinct from consultative ultrasound where a clinician orders an ultrasound exam, a sonographer acquires a comprehensive set of images, an imaging specialist (most often a radiologist or cardiologist) interprets the images, and the ordering clinician receives an ultrasound report and integrates findings into clinical decision-making **(Fig. 493-1)**. The goal of POCUS is not to replace the imaging specialist or the high-resolution data provided by computed tomography (CT) or magnetic resonance imaging (MRI), but rather to improve diagnostic and therapeutic decisions made by the treating clinician at the bedside.

POCUS became part of trauma care in emergency departments in the 1980s, and subsequently, many specialties began incorporating POCUS into patient care. The 1999 House of Delegates from the American Medical Association passed a resolution (AMA HR. 802) enabling each specialty to define its own scope and appropriate use of POCUS. Specialty-based guidelines emerged supporting credentialing processes and defining standard scanning protocols to answer focused diagnostic questions. Common clinical scenarios, such as acute dyspnea, abdominal pain, and shock, can be rapidly characterized using POCUS **(Table 493-1)**.

In internal medicine, there has been expanding interest in POCUS since the 2000s. POCUS can enhance diagnostic accuracy, treatment, monitoring, and screening of patients, as well as improve patient and clinician confidence and procedural safety **(Fig. 493-2)**.

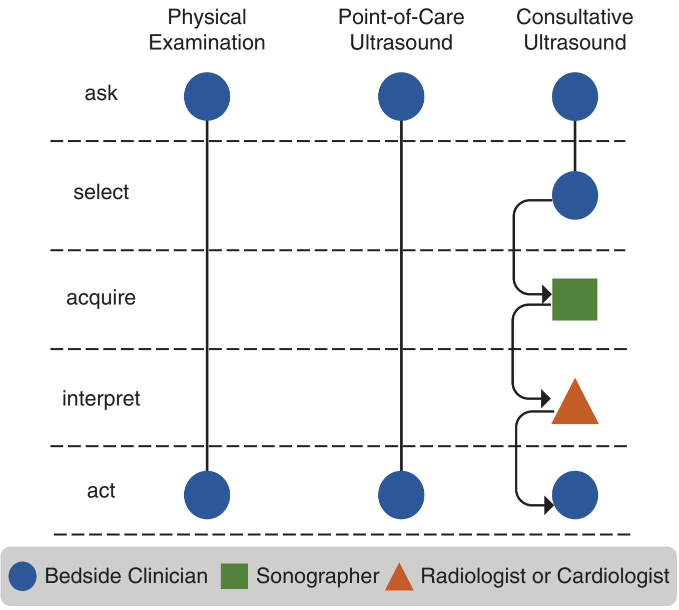

FIGURE 493-1 Workflow schematic comparing point-of-care ultrasound to physical examination and consultative ultrasound. Medical decision-making begins with asking a targeted question, selecting the diagnostic modalities, acquiring and interpreting images or other data, and ultimately, acting to incorporate the new findings into the patient's care. The three different shapes represent various personnel in this process, and *curved arrows* demonstrate the exchange of information among providers across different stages. *(Reproduced with permission from NJ Soni, BP Lucas: Diagnostic point-of-care ultrasound for hospitalists. J Hosp Med 10, 2014.)*

Portable ultrasound machines are categorized as cart-based machines versus handheld devices with wired or wireless probes connected to a tablet or mobile phone. Linear, curvilinear, and phased-array probes are commonly available, and multifunctional probes are emerging. Linear high-frequency probes have excellent image resolution but limited penetration, so they are used primarily to examine superficial structures. Deeper structures are visualized with curvilinear or phased-array probes, which have a lower frequency. Portable ultrasound devices offer two-dimensional or gray-scale imaging, and color flow and spectral Doppler imaging. Important considerations when purchasing an ultrasound machine include portability, image resolution, screen size, probe types, imaging modes, battery life, disinfection, image archiving capability, and warranty.

COMMON APPLICATIONS

CARDIAC

In the 1990s, clinicians began to perform focused cardiac POCUS exams to guide immediate management, especially for urgent and life-threatening conditions. In intensive care units and emergency departments, cardiac POCUS is routinely used to rapidly categorize shock states and acute respiratory failure. In outpatient settings, it is often used for serial monitoring of stable patients with chronic forms of heart disease.

A limited or focused cardiac POCUS exam includes five core views: parasternal long-axis, parasternal short-axis (mid-ventricular or papillary muscle level), apical four-chamber, subcostal four-chamber, and inferior vena cava views. Clinicians with comprehensive training in echocardiography, including cardiologists and intensivists certified by the National Board of Echocardiography, may perform advanced Doppler measurements of cardiac pressures and function. Cardiac POCUS and consultative echocardiography are complementary techniques where the clinical situation and operator skill determine which approach is most appropriate.

Cardiac POCUS exams can guide immediate and ongoing clinical decision-making when performed serially. To categorize shock states, left ventricular systolic function can be qualitatively categorized as normal **(Video 493-1)**, hyperdynamic **(Video 493-2)**, moderately reduced **(Video 493-3)**, or severely reduced **(Video 493-4)**. Other findings detected by cardiac POCUS that can change immediate management include acute right ventricular failure **(Video 493-5)**, cardiac tamponade **(Video 493-6)**, and gross valvular abnormalities, including severe regurgitation of the tricuspid **(Video 493-7)**, mitral **(Video 493-8)**, and aortic **(Video 493-9)** valves, as well as large valvular vegetations **(Video 493-10)**.

Competence in basic cardiac POCUS has become a mandatory component of an increasing number of specialties, including emergency medicine, pulmonary medicine, critical care medicine, and anesthesiology.

LUNG AND PLEURA

Historically, thoracic ultrasonography, comprised of lung and pleural ultrasound, was established by clinicians specialized in critical care, pulmonary, and emergency medicine. The pleural surface can be imaged through the intercostal spaces using high-frequency probes, while low-frequency probes penetrate deeper, allowing visualization of structures in the thorax. Ultrasound is superior to chest x-ray for detection of pneumothorax, early interstitial processes, and small pleural effusions and is superior to chest CT for characterization of early complex pleural effusions.

Pleural fluid is seen as a relatively hypoechoic space bounded by the diaphragm, chest wall, and atelectatic lung **(Video 493-11)**. Pleural effusions are quantified as small **(Video 493-12)**, moderate **(Video 493-13)**, or large **(Video 493-14)**, and qualitatively assessed as simple, homogeneously echogenic, complex nonseptated **(Video 493-15)**, or complex septated **(Video 493-16)**. Ultrasound guidance to identify an

TABLE 493-1 Comparison of Physical Examination Versus Point-of-Care Ultrasound Findings for Common Pathologies

PATHOLOGY	PHYSICAL EXAMINATION					POINT-OF-CARE ULTRASOUND				
	FINDING	SENSITIVITY (%)	SPECIFICITY (%)	LR+	LR–	FINDING	SENSITIVITY (%)	SPECIFICITY (%)	LR+	LR–
Pulmonary										
Pleural effusion	Dullness to percussion	89	81	4.8	0.1	Visualization of pleural fluid	93	96	23	0.07
	Decreased breath sounds	88	83	5.2	0.1					
Pulmonary edema	Crackles	19–64	82–94	3.4	NS	Bilateral B-lines	94	92	10.4	0.06
Pneumonia	Bronchial breath sounds	14	96	3.3	NS	Consolidation pattern	94–95	90–96	13.5	0.06
	Egophony	4–16	96–99	4.1	NS					
	Crackles	19–67	36–94	1.8	0.8					
Cardiac										
Elevated LV pressure	4th heart sound (S$_4$)	37–71	50–70	NS	NS	PCWP 17 if IVC >2.0	78	83	4.4	0.3
Elevated CVP (>8 cmH$_2$O)	Neck vein inspection	47–92	93–96	9.7	0.3	CVP >10 mmHg (IVC size >2 cm)	73	85	4.9	0.32
Reduced ejection fraction	3rd heart sound (S$_3$)	11–51	85–98	3.4	0.7	LV systolic dysfunction	84–91	85–88	6.5	0.14
Congestive heart failure	Rales	12–23	88–96	NS	NS	Bilateral B-lines	97	95	19.4	0.03
	Elevated JVP	10–58	96–97	3.9	NS					
	Abdominojugular test	55–84	83–98	8.0	0.3	CVP >10 mmHg (IVC size >2 cm)	73	85	4.9	0.32
	LE edema	10	93–96	NS	NS					
Abdomen										
Ascites	Bulging flanks	73–93	44–70	1.9	0.4	Visualized ascites	96	82	32	0.04
	Flank dullness	80–94	29–69	NS	0.3					
	Shifting dullness	60–87	56–90	2.3	0.4					
	Fluid wave	50–80	82–92	5.0	0.5					
Urinary retention (>400 mL)	Palpation	82	56	1.9	0.3	Bladder volume (>600 mL)	96	75	3.84	0.05
Soft Tissue and Musculoskeletal										
Lower extremity DVT	Calf swelling >2 cm	61–67	69–71	2.1	0.5	Compression venous ultrasonography	96	97	32	0.04
	Homan's sign	10–54	39–89	NS	NS					
	Wells' score (high probability)	38–87	71–99	6.3	NA					

Abbreviations: CVP, central venous pressure; DVT, deep-venous thrombosis; IVC, inferior vena cava; JVP, jugular venous pulse; LE, lower extremity; LR, likelihood ratio; LV, left ventricle or left ventricular; NA, not applicable; NS, not significant; PCWP, pulmonary capillary wedge pressure.

Source: Reproduced with permission from A Bhagra et al: Point-of-care ultrasonography for primary care physicians and general internists. Mayo Clin Proc 91:1811, 2016.

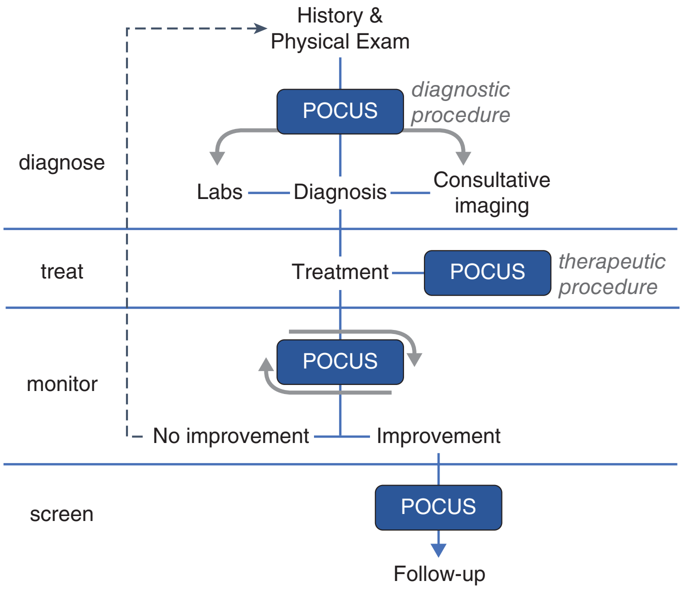

FIGURE 493-2 Clinicians can use point-of-care ultrasound (POCUS) as part of a patient's diagnosis, treatment, monitoring, and screening. A patient encounter begins with the history and physical examination, followed by a focused bedside ultrasound exam to narrow the differential diagnosis and guide workup. Treatment plans can include bedside procedures that are performed with ultrasound guidance. Serial POCUS exams can monitor disease processes and guide ongoing treatment decisions. Screening POCUS exams can detect asymptomatic, potentially treatable conditions. *(Reproduced with permission from NJ Soni, BP Lucas: Diagnostic point-of-care ultrasound for hospitalists. J Hosp Med 10:120, 2015.)*

optimal site for pleural drainage reduces the risk of pneumothorax and bleeding complications.

Normal air-filled lung tissue reflects sound waves, thereby preventing visualization of aerated lung parenchyma. Two hallmarks of normal aeration of lung on ultrasound include lung sliding **(Video 493-17)**, which results from respirophasic movement of the parietal and visceral pleural interface, and A-lines, which are horizontally orientated reverberation artifacts seen deep to the pleural line of air-filled lungs **(Video 493-18)**. Interstitial abnormalities manifest as B-lines, which are vertically orientated hyperechoic lines emanating from the pleural line to the bottom of the screen **(Video 493-19)**. Depending on their density and distribution, B-lines can support a diagnosis of cardiogenic pulmonary edema, pneumonitis, acute respiratory distress syndrome, or interstitial lung diseases. Consolidation results in lung that is tissue dense on ultrasound. Mobile air bronchograms and blood flow detected by color flow Doppler are associated with pneumonia when seen in an area of consolidation **(Video 493-20)**. Similar to chest x-ray and chest CT, identification of consolidation by lung ultrasound does not specify a diagnosis of pneumonia, and clinical correlation is required.

■ ABDOMEN

Evaluation of peritoneal free fluid is a common abdominal POCUS application. POCUS cannot specify the type of fluid (i.e., ascites, blood, urine, bile, chyme) but can detect as little as 100–500 mL of peritoneal free fluid. When ascites is present **(Video 493-21)**, POCUS can identify a safe site for paracentesis, improving procedural success and complication rates compared to landmark-based techniques. POCUS eliminates attempts at paracentesis when an insufficient volume of ascites is present **(Video 493-22)**. The best site, depth, and angle for needle insertion is determined using the ultrasound probe followed by color flow Doppler examination of the proposed trajectory of needle insertion to avoid injury to abdominal wall blood vessels **(Video 493-23)**.

POCUS is used in the initial evaluation of acute renal failure and decreased urine output. Bladder ultrasound can rapidly identify presence or absence of urine in the bladder and confirm appropriate placement and function of a urinary catheter **(Videos 493-24 and 493-25)**. Bladder ultrasound is more reliable than automated bladder scanners for urinary retention, as bladder scanners can falsely report pelvic free fluid (i.e., ascites, cysts, small bowel obstruction) as elevated bladder volume. POCUS is effective to evaluate kidney size and echogenicity; identify renal cysts, large stones, and masses; and detect and grade hydronephrosis **(Videos 493-26 to 493-28)**, thereby identifying obstructive uropathy.

POCUS can diagnose an abdominal aortic aneurysm (AAA) with high sensitivity and specificity **(Videos 493-29 and 493-30)**. A protocol that emphasizes complete visualization of the abdominal aorta from celiac trunk through the iliac bifurcation in both transverse and longitudinal planes can provide a reliable evaluation of the aorta. POCUS use for AAA screening may reduce morbidity and mortality among high-risk patients.

POCUS has utility for evaluation of small-bowel function. Normally, the small bowel is partially filled with air that obscures visualization due to scattering of sound waves. When a small-bowel obstruction (SBO) develops, the air-filled loops of bowel become fluid-filled, permitting visualization of the bowel walls. Diagnostic criteria for SBO by ultrasound include dilation of the bowel (diameter >2.5 cm), fluid-filled small-bowel loops (confirmed by appearance of plicae circularis at the perimeter), and hyperactive to-and-fro peristalsis within loops of small bowel **(Video 493-31)**. Combining patient history, physical examination, and a systematic survey of all four quadrants by ultrasound, clinicians can diagnose SBO rapidly and reliably. For a new diagnosis of SBO, POCUS can expedite early intervention and surgical consultation. For recurrent SBO, POCUS can reduce repeat radiation exposure by CT scans and expedite initiation of medical management.

■ LOWER EXTREMITY DEEP-VEIN THROMBOSIS

Two-dimensional compression ultrasound is a rapid and accurate diagnostic technique for deep-vein thrombosis (DVT) that clinicians can learn after brief training programs. A point-of-care lower extremity compression ultrasound exam yields similar diagnostic accuracy for detection of DVTs as traditional duplex or triplex ultrasound exams. DVTs commonly form at venous junctions because of high turbulence, and hence, compression ultrasound is performed at major branchpoints of the venous system. A perpendicular compression technique is required to ensure complete venous compression with wall-to-wall touching. A noncompressible vein is diagnostic of DVT **(Video 493-32)**, and visualization of intraluminal clot is not required to diagnose a DVT.

■ SKIN AND SOFT TISSUE

POCUS allows rapid differentiation between skin and soft tissue infections (SSTIs) and reactive lymph nodes, seromas, hematomas, hernias, thrombophlebitis, DVT, cysts, and bursitis. For SSTIs, POCUS can reduce unnecessary attempts at incision and drainage and avoid delays in surgical intervention. SSTIs range from cellulitis to phlegmon, abscess, and necrotizing fasciitis. POCUS can accurately distinguish abscess from cellulitis, but diagnostic accuracy is more variable for necrotizing fasciitis. To diagnose cellulitis, POCUS identifies subcutaneous edema described as "cobblestoning" **(Video 493-33)**. Abscesses appear as irregular, enclosed areas superficially with compressible material and absent central flow on color Doppler **(Video 493-34)**.

■ VASCULAR ACCESS

Current evidence supports use of ultrasound guidance for insertion of central venous catheters (CVCs) in the femoral, internal jugular, and axillary veins. Ultrasound guidance for insertion of internal jugular CVCs improves procedure success rates and reduces complications, particularly pneumothorax and arterial punctures. A preprocedure ultrasound survey identifies potential vessels to cannulate and can reveal unsuspected venous thrombosis, atypical anatomy, and venous stenosis. During insertion, real-time visualization of the needle tip reduces procedure attempts and needle redirections, which reduces the risk of complications. Sonographic confirmation of the guidewire in the target vein provides a safety check prior to venous dilation and insertion of the CVC.

For peripheral intravenous (PIV) catheter insertion, ultrasound can increase cannulation success rates while reducing puncture attempts, time to cannulation, and trauma to surrounding structures, particularly in patients with anticipated difficult PIV placement or after failed attempts using standard techniques. Ultrasound identifies peripheral veins that are large, linear, and superficial, and real-time ultrasound guidance allows visualization of the needle tip entering the vessel lumen.

TRAINING

PATHWAYS

Ultrasound training is a longitudinal process for clinicians as they progress through medical school, residency, and fellowship and enter clinical practice. Training recommendations for POCUS have been developed for different stages of medical education but with varying definitions of competence. Regardless of the clinical rank of the learner, competence in POCUS requires mastery of ultrasound knowledge (e.g., clinical indications, applications, limitations, artifacts), image acquisition, image interpretation, and clinical integration. Image acquisition and interpretation skills are learned at varying rates and require deliberate practice.

CERTIFICATION

Currently, there is no widely accepted certification for POCUS. Some residency and fellowship training programs, such as critical care and emergency medicine, require comprehensive training in POCUS, and hospitals generally grant POCUS privileges to physicians with board certification in these specialties. In contrast, internal medicine residency training does not require comprehensive POCUS training, and board certification in internal medicine does not imply competence in POCUS. Several internal medicine residency programs and professional societies have developed POCUS training courses.

CREDENTIALING AND PRIVILEGES

Clinical privileges are governed by the rules and regulations of individual hospitals. A hospital's credentialing and privileging committee is responsible for developing criteria for granting privileges for POCUS use, which may be guided by specialty-specific guidelines. Some hospitals will designate a local POCUS expert to assess competence in POCUS prior to granting privileges for POCUS use in patient care. Hospital credentialing and privileging bodies may designate POCUS as a core privilege of a specialty (e.g., emergency medicine privileges include POCUS use) or as add-on privileges separate from the primary specialty's skills. Some well-established POCUS applications, such as ultrasound-guided CVC insertion, are commonly designated as core privileges when use of ultrasound guidance is standard of care. In contrast, less common POCUS applications, such as peripheral nerve blocks, may be designated as add-on privileges.

FUTURE DIRECTIONS

The increasing portability and affordability of ultrasound devices have allowed internal medicine clinicians to incorporate POCUS into frontline patient care. Increasing POCUS use in internal medicine requires development of effective training programs during residency training and for internists in-practice. Tele-ultrasound has shown promise for training clinicians and delivering patient care remotely. In the coming years, artificial intelligence will facilitate both POCUS training and use in clinical care, and remote serial monitoring of common conditions like heart failure may be possible with patients' use of POCUS.

FURTHER READING

American College of Emergency Physicians Ultrasound Guidelines: Emergency, point-of-care, and clinical ultrasound guidelines in medicine. Available at: *https://www.acep.org/siteassets/new-pdfs/policy-statements/ultrasound-guidelines--emergency-point-of-care-and-clinical-ultrasound-guidelines-in-medicine.pdf.* Accessed December 3, 2024.

Mayo PH et al: American College of Chest Physicians/La Societe de Reanimation de Langue Francaise statement on competence in critical care ultrasonography. Chest 135:1050, 2009.

Qaseem A et al: Appropriate use of point-of-care ultrasonography in patients with acute dyspnea in emergency department or inpatient settings: A clinical guideline from the American College of Physicians. Ann Intern Med 174:985, 2021.

Soni NJ et al: Point-of-care ultrasound for hospitalists: A position statement of the Society of Hospital Medicine. J Hosp Med 14:E1, 2019.

Soni NJ, Arntfield R, Kory PD: *Point-of-Care Ultrasound*, 2nd ed. Philadelphia, Elsevier/Saunders, 2019.

Spencer KT et al: Focused cardiac ultrasound: recommendations from the American Society of Echocardiography. J Am Soc Echocardiogr 26:567, 2013.

VIDEO 493-1 Normal cardiac function.

VIDEO 493-2 Hyperdynamic cardiac function.

VIDEO 493-3 Reduced cardiac function.

VIDEO 493-4 Severely reduced cardiac function.

VIDEO 493-5 Acute right heart failure.

VIDEO 493-6 Cardiac tamponade.

VIDEO 493-7 Tricuspid valve regurgitation.

VIDEO 493-8 Mitral valve regurgitation.

VIDEO 493-9 Aortic valve regurgitation.

VIDEO 493-10 Tricuspid valve vegetation.

VIDEO 493-11 Lung atelectasis.

VIDEO 493-12 Small pleural effusion.

VIDEO 493-13 Moderate pleural effusion.

VIDEO 493-14 Large pleural effusion.

VIDEO 493-15 Homogenous pleural effusion.

VIDEO 493-16 Loculated pleural effusion.

VIDEO 493-17 Pleural sliding.

VIDEO 493-18 A-Line artifact.

VIDEO 493-19 B-Line artifact.

VIDEO 493-20 Lung consolidation.

VIDEO 493-21 Large-volume ascites.

VIDEO 493-22 Small-volume ascites.

VIDEO 493-23 Abdominal wall vessels with color Doppler.

VIDEO 493-24 Urinary catheter balloon in empty bladder.

VIDEO 493-25 Malfunctioning urinary catheter.

VIDEO 493-26 Mild hydronephrosis.

VIDEO 493-27 Moderate hydronephrosis.

VIDEO 493-28 Severe hydronephrosis.

VIDEO 493-29 Abdominal aortic aneurysm.

VIDEO 493-30 Abdominal aortic aneurysm.

VIDEO 493-31 Small-bowel obstruction.

VIDEO 493-32 Deep-vein thrombosis.

VIDEO 493-33 Cellulitis.

VIDEO 493-34 Simple abscess.

494 Improving Health and Health Care Using Behavioral Economics

Kevin G. Volpp, George Loewenstein, David A. Asch

Tobacco use, physical inactivity, unhealthy diet, excessive alcohol use, and other individual behaviors are estimated to underlie 40% of premature mortality in the United States. Approximately 75% of the $4.3 trillion currently spent on health care in the United States is attributable to cancer, heart disease, type 2 diabetes, and obesity, and each of these conditions is strongly influenced by behavior. Nearly one-half of patients prescribed medications to lower their cholesterol within 1 year following myocardial infarction stop taking these drugs—even when they are provided free of charge. Despite great advances in the science and technology of health care, a large gap separates theoretically achievable goals in health and health care from what individuals and populations actually achieve. Human behavior is both a major contributor to health problems and a barrier to the successful implementation of solutions to address them.

Recognizing that there are many reasons people do not take action to improve their health, behavior change experts have focused efforts both on strategies targeting individual behavior, such as diet and exercise, medication adherence, and smoking cessation, as well as policy-related work such as identifying problems with the existing system of health insurance, including the design of patient benefits and payments to clinicians and hospitals.

Existing policies presuppose that health care decisions are based on rational self-interest and that people have advanced computational capacity. Rational people, it is assumed, will dispassionately assess the net present value of the costs and benefits of alternative courses of behavior and pursue the best path forward. These approaches often break down when exposed to realities of human behavior such as limited attention, overconfidence, and problems of self-control. It is not just the magnitude of costs and benefits that matters, but also other critical features, such as their timing, frequency, saliency, and framing. Acknowledging these complexities has implications not only for understanding health-related behaviors, but also for intervention design. Public health programs, including those involving financial incentives, are more likely to achieve their goals if designed based not on how perfectly rational people *ought* to make health decisions but rather on how humans *actually* make them. The field of behavioral economics uses insights from psychology to identify ways that human decision-making often falls short of the ideal. The field aims to offer a more realistic account of human behavior, providing a natural framework for efforts to encourage changes in behavior.

■ CONCEPTS OF CLASSICAL ECONOMICS AND HOW THEY DIFFER FROM BEHAVIORAL ECONOMICS

Classical economics posits that individuals are rational utility maximizers, meaning that they are able to dispassionately identify alternative decisions, calculate the probabilities of and utility/disutility for each potential outcome, and then implement the decision that has the highest net present value. When it comes to health behaviors, classical economic theory would assume that if people are obese they must have decided that the costs of obesity are outweighed by the benefits of the behaviors that lead to it, and that if people smoke they must have decided that the pleasures of doing so outweigh the costs. This assumption of rational utility maximization has two major consequences for public policy, both of which are applicable in the health domain.

The first is that conventional economic thinking radically limits the range of situations in which it makes sense to intervene at a policy level. If someone has made a considered and rational decision to smoke, it becomes harder to justify government interventions to change that decision as long as their actions don't directly harm others. Under the assumptions of conventional economics, regulatory interventions such as targeted taxes and subsidies are considered appropriate only in situations characterized by externalities (costs that an individual's actions impose on others such as second-hand cigarette smoke), the presence of "market failures" (e.g., monopolies), or the presence of certain types of information asymmetries.

The second is that standard economics offers a relatively restricted array of policy tools, including taxes, subsidies, and mandates regarding patient education and information provision, and makes unrealistic assumptions about how each of these approaches will influence the behavior of individuals.

Behavioral economics builds on conventional economics by enriching its conception of individual behavior, offering enhancements on each of these dimensions. First, it broadens the range of situations in which policy interventions make sense by introducing the notion of "internalities." While externalities are costs (or benefits) individual behaviors impose on others, internalities are the costs individuals impose on themselves—typically their future selves. While interventions to address smoking and obesity can be justified on externality grounds (e.g., the health care costs are borne by individuals other than the immediately affected individual), they can also be justified on grounds of internalities—e.g., people often irrationally discount the delayed consequences of their behavior. In that sense, people might want to be protected not just from others, but from the decisions of their prior selves. Policy intervention can be further justified by the ubiquitous exploitation of individual weaknesses by businesses. Commercial enterprises may take advantage of such vulnerabilities for individual profit rather than customer health, such as with the formulation of tempting but unhealthy processed foods, the selling or prescribing of addictive substances such as cigarettes or opioids, or the pricing of meal "deals" that do not account for the health consequences of economically predatory offers. This behavior is found across industries: Credit card companies and automobile manufacturers lure new customers with "$0 down" and fleeting but appealing teaser rates of "0% interest," playing on the common propensity to focus on the present rather than on the future, as well as on people's overoptimism when it comes to their own future finances. Banks earn revenue by charging high fees for minor mistakes such as account overdrafts or breaches of minimum balance rules, often hiding the description of such fees in small print and complicated jargon. States market lottery tickets that return $0.45 on the dollar and promote these games in ways that ignore more realistic expectations using one-sided messages such as "you can't win if you don't play" rather than, for example, the equally accurate message, "you can't lose if you don't play."

Second, behavioral economics substantially broadens the range of potential policy interventions beyond those offered by traditional economics. Behavioral economics has become best known for the concepts of "libertarian paternalism" and "asymmetric paternalism." In contrast to "heavy-handed" paternalism, asymmetric paternalism attempts to protect people without limiting freedom of choice. That is, it is asymmetric because it seeks to help individuals who are prone to making irrational decisions without restricting the freedom of choice of those making informed, deliberate decisions. For example, arranging the presentation of food in a cafeteria line so that the healthy foods appear first is likely to increase the amount of healthy food chosen, without depriving those who want the unhealthy foods of the opportunity to purchase them. People who believe that individuals behave optimally should not object to asymmetric paternalism because it does not limit freedom, but those who acknowledge the limits of human rationality should endorse such measures, assuming that they are designed with individuals' interests at heart. Frequently, these measures are called "nudges."

A nudge has been defined as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid." The most prominent, and to date successful, application of nudges has been the use of defaults to increase enrollment in defined contribution retirement savings plans and, secondarily, the use of automatic escalation to encourage higher savings rates. These ideas and research findings have had a major impact on retirement savings policies worldwide. Building on this success story in savings, and bolstered by the establishment of so-called "nudge units" worldwide, the nudge agenda has positioned behavioral economics at the center of public policy.

The applicability of behavioral economics to policy, including health-related policies, however, goes well beyond nudges. Many health-relevant insights from behavioral economics do not fit the definition of a nudge. For example, there have been incentive programs aimed at changing health behaviors for maximum cost-effectiveness, improvements in the delivery of health-related information, such as nutrition labels, and new designs for physician incentives or health insurance. A common theme of this work is its harnessing of predictable decision errors that usually harm people to, instead, help them to achieve longer-term health goals or other socially valuable purposes, much the way some martial arts redirect an adversary's strength against him. Many of the same messages, incentives, and choice structures used so effectively to lure people into self-destructive health behaviors can be redirected to attract them to healthier choices that improve their long-term health and well-being. In addition, some features of human decision-making that are not recognized by traditional economics but are not really examples of decision errors, such as our propensity to experience regret when we make a bad choice and our aversion to putting ourselves into such situations (a phenomenon known as "regret aversion"), can also be exploited in the service of improving health behaviors and outcomes.

There is a common misconception that if you deploy financial incentives in order to promote behavior change, then you are engaged in behavioral economics. But that kind of activity is not behavioral economics; it is simply economics. Indeed, a large number of everyday transactions, such as being paid to go to work or getting a fine for parking in the wrong place, reflect traditional economic incentives to encourage or discourage certain behaviors. However, introducing incentives in situations, such as those characterized by internalities, in which traditional economics would suggest the incentives are unnecessary *can* be viewed as applications of behavioral economics.

Perhaps more importantly, a central lesson from the field of behavioral economics is that how incentives are delivered can matter more than their objective magnitude. There are ways of delivering large incentives that make them ineffective in changing behavior, and there are ways that can greatly magnify the effectiveness of comparatively small incentives. This observation is a source of optimism because it suggests that, with careful design, we can leverage relatively small investments to improve public health.

TABLE 494-1 Traditional Versus Behavioral Economics

TRADITIONAL ECONOMICS	BEHAVIORAL ECONOMICS
Core theory: Expected utility maximization	Core theory: Prospect theory
Assumes perfect rationality	Recognizes that people make decision errors
Starting point independent	Assessment depends on your starting point
Framing doesn't matter	Framing affects assessment even when utilities are the same
Stable preferences	Time-inconsistent preferences
People discount the future at constant rates	People discount the near future to a greater degree and have time-inconsistent discounting
Intervene only when my actions adversely affect others (negative externalities)	Consider interventions when people will harm their future selves (internalities)
Regulations and policies generally geared to protecting people from the actions of others	Regulations and policies often geared to protecting people from themselves

USING BEHAVIORAL ECONOMICS TO PROMOTE SELF-BENEFICIAL HEALTH BEHAVIORS

Over the past several decades, behavioral economics has described various ways in which people's decisions differ from standard economic models (Table 494-1). The article "Prospect Theory" by Kahneman and Tversky in 1979 is widely credited with being seminal in the development of behavioral economics. Prospect theory provides an overarching conceptual framework for describing observations about human behaviors that could not be explained by expected utility theory.

Although these deviations from expected utility theory can be seen as psychological foibles and errors, the real value of this work has come less from identifying these errors and more from recognizing that they occur in predictable ways (Table 494-2). Indeed, it is the predictability and reliability of a number of key features of human decision-making identified by behavioral economics, including but not limited to those connected to prospect theory, that allow them to be used in order to promote self-advantageous health-related behaviors.

Loss Aversion and Framing Effects Key tenets of prospect theory include the following: (1) how people feel about a set of possible outcomes depends on their starting point; this is the notion of reference-dependence in which decision-makers evaluate outcomes as gains or losses depending on their starting point; (2) people dislike losses much more than they like equivalent-sized gains, a phenomenon known as "loss aversion"; and (3) there is diminished sensitivity to both gains and losses (Fig. 494-1).

One example of how we can apply the insights of loss aversion can be seen in designing provider payment systems. Financial incentives to physicians to achieve quality goals might be presented—"framed" in the language of behavioral economics—as either rewards (in which physicians receive a bonus if, for example, they increase their colorectal cancer screening rates or increase the percentage of their patients who reach a glycosylated hemoglobin target) or penalties (in which physicians would fail to receive an expected payment if they do not meet these targets). Classical economists would consider equivalently sized rewards and penalties as identically motivating, because each reflects a structure in which one receives $X for a certain outcome (Fig. 494-1). However, loss aversion implies that whether the same incentive is framed as a gain or as a loss can make a big difference. A number of studies have shown that people have a "loss aversion ratio" in a range of 1.5 to 2.5. That means that a potential penalty of $1000 for failing to meet a quality target ought to be about as potent a motivator as a potential reward of $1500 to $2500 for meeting the same target. That multiplier is nonsensical from the standpoint of classical economics, but as an empirically verifiable descriptor of human behavior, it can be exploited in the designs of programs for clinicians or patients to improve health.

All of us—including physicians—are highly susceptible to how information is framed. In a set of experiments, patients, students, and physicians were presented with cases of lung cancer that could be treated either by surgery or radiation therapy. Among all three groups, the choice of surgery was more popular when its outcomes were framed

TABLE 494-2 Key Decision Errors and Suggestions for Addressing Them

Present-biased preferences	Provide feedback and rewards quickly
Nonlinear probability weighting	Motivate people with probabilistic rewards (lotteries)
Overoptimism and loss aversion	Get people to precommit and put money at risk as an effective motivational tool
Peanuts effect	Deliver rewards in bundles, avoiding many small rewards
Narrow bracketing	Frame rewards in terms of effort per day rather than per month or year
Regret aversion	Help people anticipate the regret of poor choices
Defaults/status quo bias	Change the architecture or environment of choice to shift the path of least resistance to favor healthy decisions
Rational-world bias	Move beyond the assumption that simply providing information will lead to desired behaviors

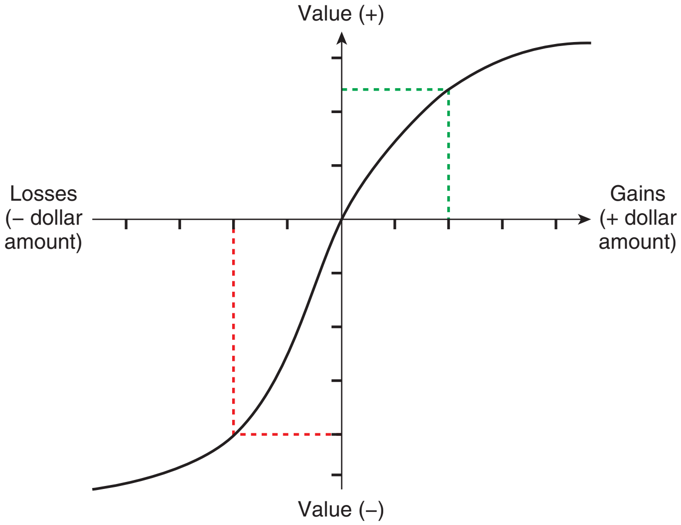

FIGURE 494-1 Prospect theory. *(Adapted from D Kahneman: Thinking, Fast and Slow. New York, NY: Farrar, Straus and Giroux, 2011.)*

in terms of the probability of survival (e.g., a 68% chance of living for >1 year) rather than in terms of the probability of death (e.g., a 32% chance of dying by the end of 1 year). We can say that such sensitivity to framing is irrational since a 68% chance of survival is logically equivalent to a 32% chance of dying—but these irrational decisions fall into predictable patterns of behavior, and that predictability can be used to influence those decisions. This means that clinicians have enormous opportunities to lead patients toward particular decisions by framing the outcomes of those decisions in specific ways, even as they remain truthful. Such an understanding could lead to the view that clinicians should be careful to balance their framing (for example, by following up the statement of 68% chance of survival with a statement like, "That means there is a 32% chance of death") in order to provide information in a way unlikely to lean toward a particular choice. Alternatively, if clinicians' roles go beyond neutral informant to trusted advisor, perhaps clinicians should deliberately frame outcomes in ways that encourage a particular choice—a much more paternalistic stance. Patients often rely on clinicians to help them make the best decisions, and that reliance may justify using the principles of behavioral economics strategically in some settings, even if the same actions in other settings might be seen as paternalistic, anti-libertarian, or coercive.

Loss Aversion and Overoptimism The power of loss aversion can be most effectively leveraged when combined with a well-documented decision error: overoptimism, or unrealistically high expectations about future outcomes. Overoptimism is especially pronounced in the context of people's predictions about their own likelihood of exerting self-control, sometimes referred to as the "false hope syndrome." The optimism that "*This* time, my efforts at weight loss will succeed!" or "*This* year, I will follow my New Year's resolutions!" often reveals itself to be false. Although in some contexts overoptimism seems to be beneficial, it can also result in suboptimal patterns of behavior. For example, people prefer paying a flat rate for gym memberships even though they would spend less if they were to pay on a per-visit basis, in part because they overestimate their future gym attendance. Indeed, many people enter into commercial weight loss programs and pay for a full year up front because they are overly optimistic about their chances of success.

Loss aversion can produce a variety of undesired behaviors, from excessive risk aversion to the tendency for people to hold on to losing investments, such as houses or stocks. However, loss aversion can advance individual or social goals. Framing reward outcomes in terms of losses by "fronting" a sum of money that gets lost if goals are not met is often more motivating than providing equivalent gains for meeting goals (the economic but not psychological equivalent). Yet, despite the greater behavioral potency of losses, program administrators are often reluctant to use loss framing, perhaps because such programs can seem more punitive than organizations may wish to appear. However, it is possible to take advantage of loss aversion by designing programs in which people voluntarily put their own money at risk in the service of achieving health-behavior goals that they themselves desire. "Deposit contracts," in which people seeking to change their own behavior put down money that they would lose if they are not successful in meeting their goals, leverage overoptimism and loss aversion to encourage behavior.

Peanuts Effects The prospect theory value function assumes diminishing marginal utility, which means that small gains and losses are disproportionately motivating relative to larger ones (e.g., two $500 rewards would be more potent than one $1000 reward). However, this may be too simple for small rewards (e.g., two $5 rewards may be less motivating than one $10 reward). This "peanuts effect" may be part of the reason why charities and retailers often describe costs in terms of "pennies a day." This observation challenges the expected efficacy of programs that emphasize efforts to repeatedly achieve small changes such as in weight loss programs. It is easy for a patient to rationalize that no single cigarette causes lung cancer, that no single snack leads to obesity, or that no single trip to the gym prevents heart disease. If self-destructive patterns of behavior, such as cigarette smoking, weight gain, or cell phone use while driving, are seen as individual instances rather than parts of a composite whole, it is easier to understand how they can be seen as acceptable. The pleasure of smoking a cigarette or eating a dessert and the convenience and engagement of conducting business or socializing via text in otherwise "dead" travel time are immediate and tangible, but the marginal costs—increased risks of developing lung cancer, being overweight, or having a car accident—seem inappreciably small. Across a lifetime or a population, however, the cumulative costs and/or probabilities are not small at all.

The tendency to underweight small outcomes can also be used to people's advantage, such as by inducing people to put away small sums for retirement savings in short cycle lengths or to make small periodic investments in their health via medication adherence. In each of these efforts, we need to think asymmetrically. In our incentive programs we should provide frequent (often daily) feedback on rewards because of present-biased preferences, but if we are *delivering* financial rewards, we want to aggregate them so that the rewards appear substantial enough to warrant attention. To use the peanuts effect to advantage, one might alert people to their rewards daily but then deliver them monthly to create larger aggregate payments.

■ PRESENT-BIASED PREFERENCES

Another central observation from behavioral economics is the concept of hyperbolic discounting, or "present bias." It is standard in conventional economics to assume that people discount the future; for example, $1000 today is worth more than $1000 a year from now, since money can be invested and earn interest. However, people tend to discount outcomes that are close in time more steeply than outcomes that are farther off in time; the degree of time discounting is disproportionately greater for short time delays than for long ones, in contrast to the assumption of standard economic models.

The medical implications of present bias are profound. For example, most people would desperately like to avoid a stroke, and many patients with hypertension have an understanding that taking their antihypertensive medications is one of the best ways to avoid a stroke in the future. While the classical economist would see daily adherence to antihypertensive treatment as a trivial cost to avoid the major cost of a future stroke, the stroke that is avoided is far in the future and uncertain; moreover, the stroke that is prevented is never noticed. In contrast, even the relatively small effort required to stay on antihypertensive medication is incurred in the present and comes without any immediate compensatory benefit. To the extent that patients overly discount the future harms of a later stroke, they will be less motivated to invest today in their own medication adherence. To the classical economist, these patients are behaving irrationally because they are failing to make small investments that, if they did the calculations, would clearly make them more likely to be substantially better off; this reasoning is parallel to how classical economists judge people who undersave for retirement (the vast majority of Americans) to be irrational. To the behavioral economist, these errors are targets for therapy, in the way that we can see genetic mutations or defects in chemical pathways as therapeutic targets in the management of illness.

Present bias is somewhat subtler than simply steep discounting of the future. In fact, it reflects *two* behavioral tendencies: (1) the tendency to overweight immediate costs and benefits relative to those occurring in the future, as just discussed, and (2) the tendency to take a more evenhanded approach to *future* costs and benefits. People are much more willing to begin dieting *tomorrow*, because the overweighting of immediate costs deters us from the immediate deprivation of dieting, and the more balanced perspective on delayed deprivation makes us willing to impose these costs on ourselves in the future.

Although present-biased preferences typically promote unhealthy behaviors, policymakers can use them for beneficial effects. The motivational impact of benefits and costs, such as rewards for good behavior and punishments for bad behavior, can be increased substantially if they are made immediate. These consequences should coincide as closely as possible with the timing of the behaviors they are meant to encourage or deter. Funds for this could be provided by employers or insurers, for whom this might be a cost-effective way to improve worker health and productivity.

Such programs have been shown to have dramatic effects in the area of addiction to drugs such as opiates and cocaine. This success is particularly striking, because many individuals with such addictions already face major adverse consequences, such as loss of livelihood and disenfranchisement from their families; yet these costs are often insufficient to motivate abstinence. Similarly, small incentives offered on proof of abstinence from tobacco have succeeded in tripling smoking-cessation rates where the far larger (but delayed) incentives in terms of improved health have failed. Small, daily incentives have significantly increased behaviors ranging from weight loss to fruit and vegetable consumption in part because they bring immediate rewards (money, excitement) to situations in which the benefits of avoiding ill health are typically distant and uncertain.

Thus, rather than requiring individuals to make decisions based on consideration of their long-term best interests, it might be useful to change short-term incentives so that beneficial actions are easier and more attractive to choose or small costs—"hassle factors"—are put in the way of behaviors adverse to health. Some school districts and employers have begun to use this approach by removing various products such as soda and candy from vending machines so that the cost of obtaining them now includes a walk off campus, while healthier food and beverage options are immediately and readily available.

Regret Aversion People dislike regretting decisions they have made, often voicing laments such as, "If only I had .... " Contemporary rhetoric recasts one form of regret aversion as "FOMO," the fear of missing out. Moreover, people are sufficiently far-sighted to anticipate possible future regret and seek to make decisions today that reduce that risk in the future. The avoidance of anticipated regret is a useful exception to present-biased preferences.

Regret aversion helps to explain the success of the Dutch postal code lottery, in which winning postal codes are selected and those living within the selected areas who purchased tickets receive prizes. Those who did not purchase tickets learn that they would have won had they done so. Individuals see their neighbors winning large prizes that they do not share, or fear that this will occur, and their desire to avoid future regret drives subsequent lottery participation.

Anticipated regret has been shown to affect a variety of preventive behaviors, such as the significant increase in vaccination use among people who experienced illness after failing to get vaccinated. Lottery-based incentive programs where eligibility is conditioned on adherence (for example, you aren't eligible to play the lottery unless you had taken your medication or checked your blood pressure the previous day) can incorporate regret aversion into their design by notifying both winners and losers. Those whose number comes up but are ineligible for a reward because of nonadherence are told they would have won had they only taken their medication, checked their blood pressure, or done whatever it was the lottery program was designed to promote. Because people hate the feeling of regret, they are more likely to engage in behaviors to avoid that feeling. Indeed, the advertising campaigns of some traditional lottery systems take advantage of regret aversion. Many people play the same favorite number when they buy lottery tickets. It easy to imagine the disappointment you would feel if you missed buying a ticket one week and that was the week your favorite number came through. Advertising slogans like "don't let your number win without you" keep people in the game in order to escape that feeling of regret. The same techniques used to promote the marketing of lottery tickets can be deployed toward health promotion. The pattern of effectiveness of a recent effort using a regret-based lottery incentive in Philadelphia to increase COVID vaccination rates highlights the importance of marketing, as larger effects were observed when the program was first announced and received extensive media coverage than in subsequent rounds of the intervention.

Defaults Although many interventions playing on behavioral economics do not qualify as "nudges," nudges remain a powerful way to influence choice. Traditional economic thinking ignores the power of the default option—the path that is "selected" when no selection is made. However, the default bias, or status quo bias, reflects our tendency to take "the path of least resistance"—to continue doing what we have been doing, or to do what comes automatically, even when superior alternatives exist. Defaults have been blamed for a wide range of suboptimal outcomes, from the failure of employees to put aside retirement funds in companies whose default contribution rate is zero, to suboptimal allocation between investment alternatives, to excessive consumption of French fries and large sodas as part of defaulted "supersized" meals at fast-food franchises. Likewise, countries and states that have "opt-in" policies for organ donation—that is, the default is nonparticipation—tend to have lower donation rates than those with opt-out defaults or (as is increasingly viewed as the best compromise) those in which citizens are required to make an active choice about whether they want to be listed as a donor. Defaults that prioritize comfort over so-called "heroic measures" have been shown to increase the rate at which patients with terminal lung diseases choose comfort-oriented plans of care. Although already employed extensively, judiciously selected defaults could be used more widely to encourage the choice of beneficial health options.

If used tactically, a choice architect (the person who makes decisions on how choices are presented to the end user) can utilize defaults such as changing scheduled automatic prescription refills from 30 to 90 days (or longer) for patients requiring lifelong chronic-disease therapy or changing the default option in fast-food restaurants in combination meals from large sodas to small sodas or water or from French fries to carrots unless you ask for French fries, to help propel people toward self-beneficial behaviors. A default toward prescription of generic medications embedded in a health system electronic order entry system moved generic prescribing to nearly 100% in a context in which previous attempts at physician education had shown no effect at all. Changing the default framing of an invitation to patients with poorly controlled diabetes to join a free remote glycemic monitoring program tripled enrollment from ~13% to 37%.

Such approaches cost nothing, since a default has to be set one way or another, preserve freedom of choice, and could change behavior substantially. We miss out on many opportunities to nudge people toward healthier lifestyles when we do not consider the default option as a tactical choice to be actively incorporated into an overall health-promoting strategy.

The Rational-World Bias Perhaps the most consequential decision error affecting health-related behaviors is the tendency of public health officials and private-sector benefit designers to make policy decisions based on the assumption that people's choices are deliberative and rational. This in turn leads to assumptions that information provision is all that is needed for optimal decision-making and that, when financial incentives are offered, the amount is all that really matters.

One significant manifestation of the rational-world bias is the complexity of health insurance plans. Health insurance is complicated for many reasons, but one reason is that health care itself is complicated. Health insurance plans are full of deductibles and copayments and other financial incentives that are designed by economists and actuaries to encourage insured patients to direct their care in specific cost-minimizing and health-maximizing ways. For these built-in incentives to exert their desired influence, however, patients need to understand the incentives they are facing, and there is considerable evidence that they do not. This complexity of insurance benefit design, rather than creating perfectly tailored incentives, is part of insurance design's undoing—because an incentive that cannot be understood cannot be effective. Most consumers lack an understanding of the most basic insurance concepts, such as deductibles, copays, and coinsurance, and, when given even a simplified version of a traditional insurance plan, are unable to compute the costs they would incur for basic services.

Giving a nod to the benefits of simplicity, the Affordable Care Act (ACA) in the United States requires insurance plans to be presented in a standardized document that describes plan features such as premiums, deductibles, and coinsurance. However, describing something that is inherently complex in simple terms raises the risk of glossing over important subtleties. A more productive approach would be to provide consumers with a truly simplified insurance product. For example, insurers might create a plan designed with copayments only, since copayments are most easily understood by consumers: copayments are analogous to paying a price for a good or service when shopping, whereas coinsurance and deductibles make it difficult for people to estimate how much their care would cost. It is difficult, if not impossible, for a consumer to estimate 10% coinsurance on a hospital stay or an emergency room visit since they have no idea of the full amount on which the 10% will be calculated. For medical markets and patient cost-sharing elements embedded in plan design to affect decision-making *ex ante*, consumers need some way of accurately gauging not only benefits and risks of treatments but also the costs. This should occur before receiving a diagnostic test or treatment as opposed to what typically happens now: a medical bill is received after the fact.

Another manifestation of the rational-world bias is the more insidious and pervasive belief that if only people knew more about how to advance their health they would do so. This error leads to making education the centerpiece of many health interventions, even though public health experts have long recognized that knowledge alone rarely translates to health-enhancing behavior. For example, nearly everyone knows the health hazards of smoking. While raising and maintaining awareness of the dangers of smoking is an important goal, the financial budgets of the Centers for Disease Control and Prevention, or the time budgets of clinicians, might be more efficiently allocated toward efforts to change behavior that do not presume that the deficit is a lack of knowledge. In cases where the deficit is not in knowledge but in behavior, reliance on education as a primary avenue for reducing health risks may divert efforts and resources from other activities that might be much more effective.

APPLICATIONS

Weight Loss Efforts to combat obesity using incentives started in the 1970s. This early work was motivated by the observation that participants who deposited money and other valuables with a therapist and signed contracts in which return of their valuables was contingent on progress toward prespecified goals lost tremendous amounts of weight. Even though participants received no training in weight loss or maintenance strategies, participants lost an average of 32 pounds. This small initial study lacked control groups and long-term follow-up but provided an important proof of concept.

In the first systematic study of deposit or precommitment contracts, participants responding to an advertisement for a weight loss program were informed that study participation required a deposit of $200 (1974 dollars), which would be fully refunded contingent on satisfactory weight loss. After an 11-session, 10-week program, those who received incentives for losing weight or for limiting calories lost significantly more weight than those who received incentives for attending sessions. Of the participants in the former two groups, 70% lost >15 pounds. A major limitation to this approach was that only 15% of the prospective participants who responded to the initial newspaper advertisement ended up enrolling, suggesting that deposit contracts that require commitment of substantial funds up front are effective for participants, but this likely deters a substantial portion of high-risk participants, and effectiveness may reflect selection of highly motivated individuals.

Subsequent studies confirmed that deposit contract participants lose significant weight.

Paying participants for weight loss using direct payments was less effective than deposit contracts. In a randomized trial, cash payments up to $25 per week for making 100% of proportional progress toward goal, $12.50 for 50% of goal, and $2.50 for not gaining weight did not result in greater weight loss in the payment group than among control subjects.

Studies that have shown no effects on either initial weight loss or maintenance typically have used incentives of small magnitude or were targeted at behaviors like attendance at weight loss programs that, by themselves, do not ensure weight loss. In recent years, weight loss incentives have become a common feature in programs used by employers and health plans, though deposit contracts have not been widely used.

Both daily lottery-type incentives and precommitment (deposit) contract incentives promote initial weight loss over 16 weeks (lottery = 13.1 pounds; $p = .014$ for lottery vs control; deposit contract = 14.0 pounds; $p = .003$ vs control; **Fig. 494-2**). However, as is the case in many weight loss studies, following the termination of incentives, participants regained most of the weight they had lost over the following 3 months. Longer trials (8 months) found no difference in effectiveness of deposit contracts for continuous 8-month weight loss versus 6-month weight loss with 2 months of maintenance, but both were successful in achieving a mean weight loss of ~10 pounds. A study that examined weight loss over 18 months using daily lotteries found modest weight loss of 5.4 pounds more than the usual care control group, but neither lottery incentives, environmental change strategies, nor the combination were significantly more effective than usual care.

Many employers use premium adjustments (increases or decreases to health insurance payments) as their standard approach to using financial incentives for health promotion. In one study, participants were randomly assigned to a control group with no other intervention or one of three financial incentive groups. Two intervention groups were offered a premium reduction of $550 if they achieved their weight loss goal by 1 year. The "delayed" group would receive the premium adjustment in the following year, spread across each pay period. The "immediate" group would have their premiums adjusted as soon as the weight loss goal was met. A third intervention group was offered a daily lottery with about a 1 in 5 chance of winning $10 and a 1 in 100 chance of winning $100. To be eligible to win each day, participants had to weigh in and be on track to lose 5% of their initial weight by 6 months, with maintenance for the subsequent 6 months. After 1 year in the program, none of the intervention approaches showed a significant degree of weight loss. The control group gained <0.1 kg.

The relative ineffectiveness of premium-based financial incentives is not surprising given their design. Premium adjustments are

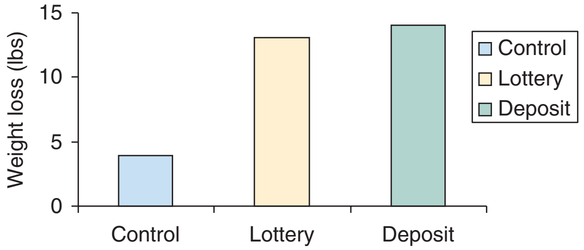

FIGURE 494-2 Weight loss in incentives versus control. *(Figure created using data from KG Volpp et al: Financial incentive-based approaches for weight loss: A randomized trial. JAMA 300:2631, 2008.)*

logistically appealing because the infrastructure to do so is already in place; however, there is little evidence for their effectiveness, and theory argues against it. Such incentives are typically deducted from, and therefore hidden, in paychecks that are directly deposited in bank accounts and may go unnoticed by the individual. While a $550 incentive seems like a large amount, it is only $20 in each biweekly paycheck. Typically, these incentives are administered on an all-or-nothing basis contingent on meeting a specific threshold such as a body mass index of ≤ 25 kg/m^2, meaning that those who are close to that target may be motivated by a goal within reach, but those who are farther away (and have the most weight to lose and the most health to gain) may be less motivated (perhaps even demotivated) by a goal that does not seem attainable. While data suggest that longer-term weight loss is more reliably attained through bariatric surgery or glucagon-like peptide 1 (GLP-1) medications, these are much higher cost alternatives (GLP-1 medications currently cost about $20,000 a year per person) that carry both known and unknown risk. The best evidence and theory suggest that the standard approach of using premium-based incentives is not very effective and that employers should consider alternative delivery channels for financial incentives.

A common problem employers and health plans struggle with is getting high levels of engagement among employees/health plan members. Weight Watchers, the largest commercial weight loss program in the United States, worked with a team of academic investigators to conduct a randomized controlled trial involving >23,000 participants testing the impact of offering employer subsidization of Weight Watchers membership fees of 50%, 80%, 100%, or 50% that could turn into 100% conditional on attending at least three Weight Watchers classes a month. Higher subsidies led to significantly higher program enrollment. The 100% subsidy produced the highest enrollment (7.7%), significantly higher than each of the lower subsidies (vs 80% subsidy: 6.2%, $p = .002$; vs 50% subsidy: 3.9%, $p < .0001$; vs hybrid: 3.7%, $p < .0001$). Enrollment in the 80% subsidy group was significantly higher than both lower subsidy groups (vs 50% subsidy: 3.9%, $p < .0001$; vs hybrid: 3.7%, $p < .0001$). Among enrollees, however, there were no differences among the four groups in attendance or weight loss. In all groups, overall weight loss was modest, with a mean weight loss ranging from 2.6 pounds in the 100% subsidy arm to 4.0 pounds in the hybrid subsidy arm. In all arms, attendance rates dropped steadily over time, suggesting that while the hypothesized trade-off between higher subsidies and lower ongoing engagement in the program did not exist, subsidies that lower the cost of participating are insufficient to achieve high levels of ongoing engagement **(Fig. 494-3)**.

In another study, 281 overweight and obese adults were randomly assigned to a control group or one of three incentive groups for a 13-week physical activity program. All participants were given a goal of achieving 7000 steps a day, tracked automatically using their smart phones. Participants in each of the three incentive arms were offered the same magnitude incentive, $1.40 per day, and were told that accumulated earnings would be sent via a check at the end of each month. However, the incentive in each group was framed differently. In the standard gain incentive group, participants were told that they could earn $1.40 each day they achieved the goal. In the regret lottery incentive group, participants were in a daily regret lottery in which they had an 18 in 100 chance of winning $5 and a 1 in 100 chance of winning $50, which averages to $1.40 per day. In the loss framing incentive group, participants were told at the beginning of each month that $42 had been placed in a virtual account and that they would lose $1.40 each day the goal was not achieved. During the 13-week intervention, participants in the control, standard gain, and regret gain arms achieved their daily step goal ~30%, 35%, and 36% of the time, respectively, but the gain-framed incentive arms were not significantly different from the control arm. However, participants in the loss framing group achieved the goal 45% of the time, a 50% relative increase, which was significantly greater than the control arm ($p = .001$). This study was one of the first to create a loss frame without requiring participants to put their own money at risk using a deposit contract, important because as previously noted fewer people are willing to engage in deposit contract-based incentives than reward-based incentives.

Experience with behavioral economics and weight loss provides several general lessons that are transferable to other health applications. While early studies mainly tested the impact of the magnitude of financial incentives, later studies have revealed that the design of the incentive strategy is at least as important. Designs can vary on a wide range of factors, including the timing and frequency of incentives, the setting of targets, the certain or probabilistic nature of incentives, the delivery channel of incentives, and the framing of incentives as gains or losses. Classical economists would see only the size of the incentive as an available lever for motivation, but behavioral economists face a much larger set of considerations in designing and testing effective approaches.

Medication Adherence Numerous studies have shown that at least a third of patients fail to adhere to medication regimens. Structural approaches to improving medication adherence such as changing some of the underlying defaults such as switching from 30-day

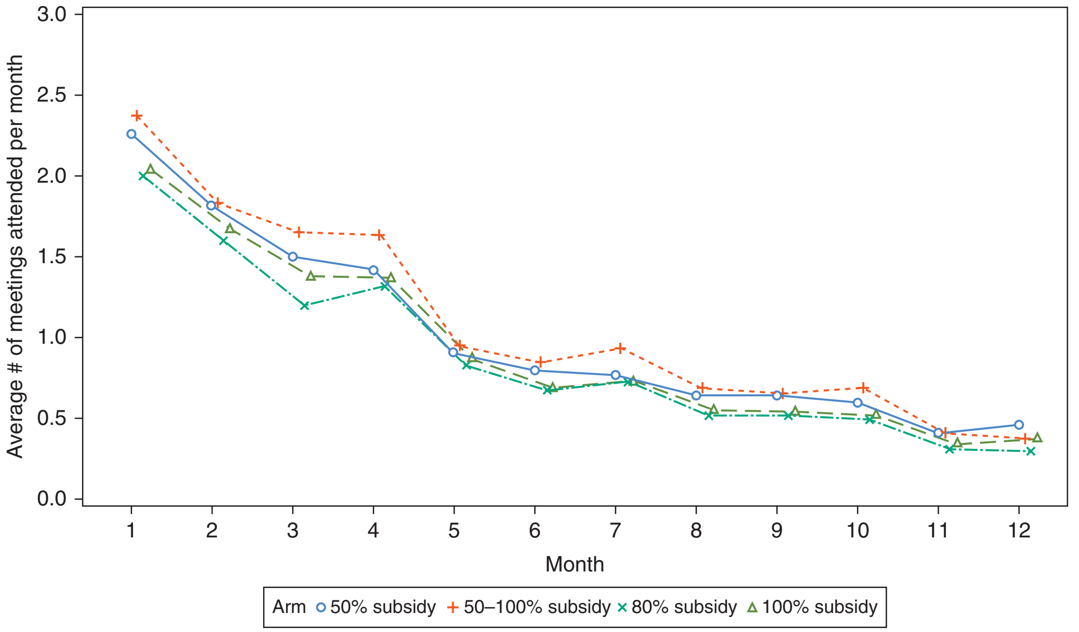

FIGURE 494-3 Attendance rates at Weight Watchers over time in different subsidy groups. *(Reproduced with permission from LK John et al: The effect of cost sharing on an employee weight loss program: A randomized trial. Am J Health Promot 32:170, 2018.)*

prescriptions to 90-day prescriptions for chronic conditions have great promise. A recent study indicated that changing to a default 90-day prescription increased adherence from 70.6% preintervention to 90.1% postintervention (p <.001). It is logical that adherence rates are higher over a year if people need 3 refills instead of 11—the latter provides more opportunities to forget, experience delays, or fall off the wagon. Automatic refills through prescription mail order might similarly help people stay adherent, giving them one less thing to think about or fail to think about.

More widely, 90-day prescriptions or automatic refills could be set up as the default, and patients or their providers could opt out if desired. However, opt-out defaults are not always possible. A large pharmacy benefits manager wanted to encourage automatic refills for patients on long-term medication but could not have members opt out of such a system because of the potential that those who missed the implications of the opt-out would be upset about finding credit card charges for automatic refills. Using active choice—i.e., requiring customers to make an explicit decision rather than only providing the service to those who proactively opt in—the company also presented options in a way that highlighted the convenience of automatic refills: “We can send your refills to you automatically or you can get your refills manually if you prefer.” This resulted in more than twice as many patients choosing to be in the automatic-refill program as compared with opt-in.

A feature of many health insurance plans is that they require patients to pay some costs out of pocket and hence discourage the use of high-value care, such as the use of statins by patients with diabetes—care that is widely seen as worth its cost. Support for the use of health insurance deductibles that require patients to have “skin in the game” for their health expenditures derives from insurance theory as well as the seminal RAND health insurance experiment, which demonstrated that these deductibles help overcome moral hazard and thereby reduce the consumption of health care services. Copayments, deductibles, and the growth of high-deductible health plans make consumers more cost conscious with the intent that this makes them better shoppers for health care services.

However, while deductibles and copayments can reduce overutilization of some lower-value health care services, deductibles and copayments make considerably less sense for medications and services that are high value and low cost—e.g., when patients receive medications to manage their hypertension, diabetes, or hyperlipidemia. Given that deductibles and copayments are designed to reduce utilization, why would we apply them to antihypertensives or statins or insulin given the high health value of these drugs? As the RAND health insurance study showed, high-deductible health plans are as likely to discourage the use of high-value services as the use of low-value services. Because patients have imperfect knowledge about which tests or services are high or low value and the relationship between price and quality, such plans discourage spending on *all* tests and services, including those of high value.

Value-based insurance design—which involves discounting, or making free, services deemed high in value—is an attempt to sharpen the blunt incentives inherent in deductibles and copayments. Value-based insurance design was inspired by research that showed the use of higher copayments significantly reduced the use of services such as prescriptions but ultimately raised costs because lower rates of medication nonadherence led to higher rates of emergency department visits and adverse outcomes. Extrapolating from these results, it was natural to conclude that lowering cost sharing for high-value activities, such as medications for chronic conditions, would increase adherence and potentially thereby reduce costs. The ACA incorporates a kind of value-based insurance design in its requirement that preventive services be offered to patients at no charge.

By and large, value-based insurance design has not delivered on the hope that it would both save money and significantly improve health. From the perspective of the purchaser (for example, the employer or insurer), the economic impact of value-based insurance design depends on whether it can make enough people adherent who were previously nonadherent—and on the health and cost consequences of that improved adherence—to offset the loss of the copayments from those who were already adherent. Improved adherence from copayment reductions for medications, even to $0, has typically been only in the range of 3–6 percentage points. Even among patients who had recent heart attacks and were given cardiovascular medications for free, average adherence was only ~45%, just a few percentage points higher than with patient engagement. Based on conventional economic thought, it might seem reasonable to assume that decreasing copayments would create effects equal and opposite to those of increasing copayments. However, behavioral economic research reveals that framing matters and that losses (in this case, higher copayments intended to reduce use) have a larger effect than equivalent size gains (lowered copayments). Furthermore, people who would be deterred by higher copayments are different from people who might become adherent with lower copayments, because the first group consists of those who take their medications while the second group consists of those who do not. In general, copayment increases lead to far smaller decreases in medication adherence than copayment decreases lead to increases in medication adherence.

Value-based insurance design is an appealing idea. Its benefits could be increased through the application of behavioral economic concepts, such as simple changes in reward delivery to increase salience (e.g., retaining the copay, but sending a rebate) and communications from insurers to patients to ensure that those who are nonadherent are aware of the benefit. Better designs might also reflect that most medication adherence happens at least daily, and reinforcements to that behavior probably need to occur more frequently than the 30- or 90-day cycles coinciding with prescription refills.

CONCEPTUAL AND PRACTICAL CONSIDERATIONS

The 5000 Hours Problem Indeed, a major challenge is determining how to reach patients and reinforce their behavior each day if we want to significantly improve their health behaviors. Even patients with chronic illnesses may spend only a few hours a year with a doctor or nurse, but they spend ~5000 waking hours a year doing just about everything else. Those 5000 h are when they live their lives and make choices about what to eat and whether to exercise, smoke, take their medications, or visit the doctor.

Although what people do in those hours almost certainly affects their health outcomes, the hours are typically ignored by the U.S. health care system, in part because current approaches to U.S. health care financing support health care during visits to the doctor, not between them, and because “hovering over” people during the hours between visits is personnel-intensive, often requiring nurses or other clinicians to call or visit patients or to staff telemedicine programs. Hovering also requires a fair amount of the very kind of engagement in their own health and health care that is so often missing in the patients these interventions aim to reach. As a result, many of the most promising efforts in telemedicine and home health care have been disappointing.

If some form of hovering is required to engage people who are otherwise hard to engage during the 5000 h, it almost certainly has to become substantially more automated—both because providers must reduce the need for expensive personnel and because patients have in many cases already revealed limits to their willingness to exert themselves to improve their health. Nevertheless, there is reason for optimism based on the increasing use of cell phones and other wireless devices that make it technologically easier to embed reminders and other forms of touch into patients’ lives. Indeed, one key lesson from behavioral economics is that rather than trying to change people’s behavior patterns to promote health, it is better to restructure their environment and circumstances so that their existing behavior patterns are more likely to lead to better outcomes. Those efforts require a substantial amount of hovering over patients. Cell phones and other wireless devices do not necessarily change behavior on their own, but because they are already part of many patients’ everyday lives, they allow previously private behaviors to be witnessed and at times acted upon.

Indeed, an error in early approaches to technology and health behavior overgeneralized from the technologies that support the “quantified self” movement. Apps and wearables that track your diet, physical activity, and biometrics were largely designed for people passionate about measuring themselves. Such individuals do not need a lot of encouragement to wear devices or enter data. The same approaches are far less likely to be useful for patients with difficult-to-manage chronic illness. Many of the internal and external challenges that make their chronic illness hard to manage also make such monitoring hard to manage. A patient who is nonadherent to medication is likely also to be nonadherent to using a new electronic device, but devices like cell phones that are already in use or other devices that require much less active involvement (like wireless pill bottles) offer more conceptual appeal.

Changing Clinician Behavior Versus Bypassing Behavior Before the introduction of electronic health record (EHR) systems, clinician decisions might be observable by reviewing charts or insurance claims, but those reviews were labor-intensive and possible only in retrospect, offering no opportunity to intercept decisions and redirect them in real time. Now, reminders and alerts are commonly programmed into EHRs, aimed at reducing errors and promoting best practices in testing and treatment—in real time, as those decisions are made. Although these alerts open up a vast array of new opportunities for changing clinical practice, they have been criticized for recreating "alert fatigue." Perhaps in part for this reason, research has yielded mixed results regarding their effectiveness. For example, while one study found that automated alerts increased hepatitis C antibody screening in a primary practice from 3.2% to 22.7%, another study found that cardiologists largely did not change practices in response to alerts that notified them of patients meeting criteria for statins.

These approaches aim to *inform* clinicians of best practices with just-in-time information, but they continue to rely on conventional pathways of clinician cognition and behavior. A potentially more efficacious approach is to change the "choice architecture" faced by clinicians. Given that those recommended for a statin or hepatitis C screening can often be identified by information already present in the EHR, and given that ordering an antibody test or prescribing a statin are also accomplished through the EHR and the benefits are generally considered much greater than the risks, consideration should be given to systems that identify patients who would benefit from the drugs or tests and automatically order them.

Given ethical reservations, or regulatory limitations, on bypassing physicians, less heavy-handed approaches are possible that make the best practice the easier choice of clinicians. For example, rather than identifying patients meeting criteria for hepatitis C screening and then reminding clinicians to order the test, EHRs could automatically create the order and execute it *unless* the clinician decides against it. A stepped-wedge randomized controlled trial demonstrated a 31.8 percentage point increase in hepatitis C virus antibody screening among hospital patients when eligible patients were defaulted to screening (unless the clinician opted *out*), as compared with a condition in which clinicians received an interruptive prompt to order the test.

These designs are potentially less likely to induce fatigue than alerts that, if attended to, interrupt the physician's train of thought. The increasing number of algorithmic guidelines combined with the increasingly digital ecosystem of clinical care offer opportunities to advance care less through reminders that aim to travel through conventional pathways of clinician cognition and behavior and more through nudges that bypass those pathways.

Enhancing External Validity Through Pragmatic Trials We would never think of introducing a new drug that touched 1000 patients without trials examining its safety and effectiveness. The same types of methodologies often are used to test behavioral interventions. Nevertheless, the evaluation of investigational drugs differs, or should differ, from the evaluation of behavioral interventions when it comes to patient consent.

Informed consent is a central element of human subject protection, and in the case of testing pharmaceuticals, devices, or surgical procedures, we generally expect consent procedures to have little or no impact on the generalizability or external validity of a study. By and large, if 300 of 1000 people approached agree to participate in a trial comparing two drugs, we have little reason to believe that the physiologic effects observed in the 300 people who consent will be much different from what would have happened had the other 700 people participated instead. The 300 people who consent can be relied upon to provide evidence generalizable to the 700 who do not, and the external validity or generalizability of the trial is not threatened by a consent process that results in the drug being tested on only a subset of the possible pool.

Trials in behavioral science, however, may be substantially biased by consent procedures. Imagine that out of 1000 potentially eligible patients with heart failure, 300 agree to participate in a trial of a behavioral intervention to improve medication adherence and 700 decline. If the intervention works among those 300, we may reasonably question whether it will also work among the other 700. It is likely that the 300 patients who consented were, in general, more motivated to improve their medication adherence and that motivation may have increased the likelihood that the intervention would succeed.

Most behavioral interventions will fall under the category of minimal risk, and contemporary human subject projections codified in the common rule permit the waiver of consent under such circumstances when the research could not practicably be carried out without a such a waiver. For these reasons, many behavioral interventions may be appropriate for waivers of informed consent, making these trials more pragmatic and therefore generalizable.

The I-Frame and the S-Frame The vast majority of research applying behavioral economics to health has focused on changing the behavior of individuals (i-frame), mostly patients but also, in some cases, physicians. A major theme of the research has been that individuals are prone to systematic mistakes when it comes to health behaviors such as medication adherence, diet and exercise, and choice between health insurance options, and that modifying these behaviors can lead to improvements in health and health care utilization. Arguably, however, many of the most severe health problems that plague the United States and other nations stem from structural or systemic (s-frame) problems rather than from individual deficiencies. The complicated nature of health care financing in the United States, for example, including fee-for-service arrangements that encourage overprovision of tests and procedures, and the atomized health insurance sector that leaves many individuals without coverage and even insured individuals struggling to pay for needed care do not stem from individual-level behavioral deficiencies but from flaws in the overall system. Likewise, while individual-level factors such as self-control clearly play a role in obesity, systemic factors, such as the subsidization of corn, dairy, and meat and the huge lobbying power of concentrated agribusinesses, help to explain why obesity rates went from 14% to >40% in the United States over the past 40 years.

Although these problems arise at a systemic rather than individual level, the solution is not to focus on s-frame to exclusion of i-frame interventions, but to recognize that both are important and that behavioral economics still has an essential role to play. First, behavioral economics research can, and has, sought to understand the underlying cause of problems such as rising levels of obesity and "deaths of despair" (e.g., suicide, drug overdoses, and liver disease). An understanding of the causes of these problems is essential for determining what types of systemic changes, presumably enacted through regulation and legislation, could successfully mitigate them. Behavioral economics can also play a major role in informing people about the systemic causes of health-related problems and in building public support for needed reforms. In addition, while structural changes might change the preconditions for success or failure, optimizing individual behavior given those structural constraints remains imperative.

Residual Challenges Behavioral economics typically takes its cue from an understanding of universal or near universal patterns of thinking. One might hope, therefore, that harnessing these patterns would have near universal effects and perhaps profound ones. For all the conceptual appeal of behavioral economics in health and medicine, however, few studies have demonstrated the kind of widespread effect we have seen with vaccination for smallpox or polio. As noted previously, switching a default toward generic prescription moved generic prescribing to nearly 100%, which is an example of a profound effect, but most other studies show modest effects, and many show no effect at all. What are we to make of the frequently disappointing outcomes?

One possibility is that we should simply recognize that human behavior is complex and that interventions traveling along a single pathway coexist with forces along other pathways in cascades that better resemble the complexity of inflammation than a single inborn error of metabolism. We should not be surprised when clever and conceptually well-grounded designs have less effect than we had hoped.

A second possibility is that we have the right pathways but not the right dose or the optimal combination of therapies. If we could think of ways to intensify the treatments of behavioral economics, perhaps a positive dose-response effect would reward us with more success.

A third possibility is that the premise of universality is false—that some individuals are more likely to respond to interventions focused on regret aversion, others on loss aversion, and so on. In this model, if we could just characterize people into their behavioral phenotypes, we might more precisely target behavioral interventions to them just as we aim to target our cancer care to the often unique molecular signatures of tumors. No coherent and consistent way to segment people into behavioral phenotypes has been developed, in part because initial studies have been large enough only to test main effects and not subgroups and in part because we have not yet identified or developed measures that effectively segment personalities for this purpose.

A fourth possibility, aligned with the first, is that some important decisions are no longer just the product of our reasoning, but also a product of our ideology and societal context. These are challenging to change if not aligned with health improvement. The COVID-19 pandemic introduced enormous interest among clinicians and public health experts in interventions that would increase health by promoting behaviors such as mask wearing or vaccination. Efforts to increase the latter, including through the use of financial incentives, were generally ineffective, perhaps because at the time they were introduced those remaining unvaccinated had strong beliefs that had driven the decision not to get vaccinated. Context and ideological or political pathways can create countercurrents stronger than the mechanistic pathways on which behavioral economics is based.

REFLECTIONS ON SARS-CoV-2 Since SARS-CoV-2 began spreading early in 2020, it has become increasingly clear that behavioral decisions among both policymakers and individuals play a major role in spread. Within countries that exercised strict control over individual behavior through measures like mandated testing, isolation, quarantine, use of contact-tracing apps, and fines for noncompliance, containment was achieved and the number of new cases dropped nearly to zero. Within countries like the United States in which there was no comparable national policy and policy decisions were left to the states, the rate of spread and degree of containment varied enormously. Leaders of many states relied on the rational-world bias, assuming individual decisions would be deliberative and rational. Individual decisions about whether to maintain social distance, wear masks, and otherwise modify routines to reduce risk reflect many of the behavioral challenges described earlier in this chapter—present bias, overoptimism, anticipated regret, narrow bracketing, status quo bias, and the peanuts effect all contributed to people either taking more or less precautions depending on their personal decision-making proclivities. The rollout of the vaccine was similarly hampered by s-level factors, most notably the absence of a centralized registry of citizens (a symptom of the atomized health insurance system in the United States), which rendered delivery of vaccines uneven and a somewhat chaotic process. Vaccination acceptance posed similar challenges; governments largely left it up to individuals to decide and did not use strong incentives or mandates. Significant variability in behavior ensued, with corresponding consequences for the populations they govern.

■ CONCLUSION

Human health derives from the interaction of basic biologic processes, environmental exposures, social structures, government policies, and behavior. The field of behavioral economics has greatly contributed to our understanding of behavior and has made significant contributions to the science of public policy. Given that understanding, we have the opportunity to replace older policies based on unrealistic normative models of rational decision-making with newer policies reflecting our most up-to-date understanding of how humans actually make decisions. Individual and population health outcomes would be very different if people were able to weigh the present and future costs of their actions carefully and dispassionately and had the necessary information and self-control to implement behavioral plans and overcome decision errors that contribute to unhealthy behaviors. Because few people can meet any of those challenges, we should not structure our behavior change interventions and public health policies around such models of behavior.

There is broad potential for improving private and public approaches to health behavior by drawing upon a more realistic understanding of human motivation rooted in behavioral economics. One major question is whether developed economies will continue to invest the majority of their health care dollars in treatment (typically ~97% of health care dollars) rather than shifting it toward innovations that seek to keep people healthy. This will be particularly important in settings where treatment options are limited and highly effective methods of prevention exist. For example, it has been estimated that a combination of low-cost cardiovascular drugs could reduce cardiovascular events by 62–88% with perfect adherence, revealing that reducing atherosclerotic cardiovascular disease risk is largely a *behavioral* challenge, given that adherence to medications remains low despite effective pharmacologic solutions. Shifts in health care financing away from fee-for-service toward various forms of payment that require health delivery systems to take on financial risk for populations of patients may drive greater interest in addressing these behavioral and social determinants of health. Research expenditures, which in many developed countries are also roughly allocated ~97% to new treatments and ~3% to prevention, similarly could be shifted to focus more on testing of innovative approaches to improve population health.

The same errors that misdirect patients and providers also misdirect policymakers. In part because of present bias, preventive services often are covered by insurance only if they show a positive return on investment, yet treatments of existing disease are not held to the same standard. An employer wondering whether to introduce a smoking cessation program for employees wonders about the return on that investment in terms of reduced illness and its cost. The same employer would likely never question the return on investment of treating lung cancer in the same employee pool, despite what would almost certainly be a negative return on investment. These asymmetries are so embedded in policy making as to be nearly invisible or at least unchallengeable. For example, the cost of treatments is not even allowed to be considered in Medicare coverage decisions. This naturally leads to overinvestment in treatments of low value and underinvestment in prevention. The same standard for assessing the impact of health programs, with the goal of achieving the most health possible with the available resources, should be used for both preventive and therapeutic services.

Despite the promise of behavioral economics in structuring policy solutions to social goals, plenty of existing policies that have nothing to do with behavioral economics are effective. For example, raising taxes on cigarettes or sugar-sweetened beverages or other unhealthy goods where it is in the public interest to consume less is a powerful policy tool derived from classical economics. Indeed, tobacco taxes represent one of the most effective ways to curb the use of tobacco and its initiation among youth. Behavioral economics can help make such policies more effective but should not be seen as a substitute for them.

For private-sector entities, the implications of choosing defaults wisely are recognized by many organizations that aim to shift the "path of least resistance" toward healthier choices. Setting up defaults in benefit program design to favor health plans that provide better coverage of preventive services, changing the environment in workplaces to make it easier to take the stairs, and serving more healthful food in cafeterias represent approaches to gently lead people toward individual and population goals.

While medical research continues to generate new tests, interventions, and drugs, all of which successfully target conditions recently seen as intractable, even the most effective drugs will not work if physicians fail to prescribe them and if patients fail to take them. Although the dominant forms of investigation in medicine seek cellular or molecular therapeutic targets to modify disease, behavioral sciences have revealed cognitive pathways that operate nearly as predictably as the genetic code. The opportunity for behavioral economics to improve health and health care delivery derives from its recognition of these behavioral pathways and the growing empirical evidence about how to best make use of them.

FURTHER READING

Asch DA et al: Automated hovering in health care—Watching over the 5000 hours. N Engl J Med 367:1, 2012.

Chater N, Loewenstein G: The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray. Behav Brain Sci 46:e147, 2023.

Loewenstein G et al: Asymmetric paternalism to improve health behaviors. JAMA 298:2415, 2007.

Thaler RH et al: Choice architecture, in *The Behavioral Foundations of Public Policy*, E. Shafir (ed). Princeton, NJ, Princeton University Press, 2013, pp 428–439.

Volpp KG et al: Financial incentive-based approaches for weight loss: A randomized trial. JAMA 300:2631, 2008.

495 Complementary and Integrative Therapies and Practices

Helene M. Langevin

The search for health and improved well-being includes many treatments, practices, and systems of care that may have originated outside conventional medicine but are gradually being folded into mainstream health care. The current health care system is fragmented, often emphasizing the pharmacologic treatment of disease alone, while often neglecting the promotion, support, and, importantly, restoration of health. Though the disease-focused model is dominant in our research and health care ecosystem, there has been a longstanding awareness that many chronic diseases, including pain conditions, can be prevented or better managed by incorporating nonpharmacologic interventions such as nutrition, exercise, and stress management into care, with an emphasis on understanding the person as a whole. Many complementary practices follow this model, and there is preliminary evidence indicating that these approaches lead to improved self-care, a better personal sense of well-being, and a greater commitment to a healthy lifestyle. Integrative health emphasizes not only the integration of complementary and conventional care but also an integrative approach to treatment of the whole person. This includes expanding our understanding of how physiologic systems interact with one another and of the connections between physical, psychological, and social aspects of health. Integrative health also includes striving for a better understanding of “salutogenesis” or pathogenesis in reverse, meaning the process by which health is restored when recovering from an injury, acute illness, or the exacerbation of a chronic disease, or when a “predisease” condition such as prediabetes or prehypertension is reversed through changes in behavior rather than pharmacologic treatment.

DEFINITIONS AND SCOPE

Complementary health therapies and practices include a broad range of practices, interventions, and natural products that are not typically part of conventional medical care **(Table 495-1)**. The term *complementary* refers to the use of these practices together with conventional therapies and is increasingly preferred to the term *alternative*, which denotes usage as a substitute for standard care.

The term *integrative health care* refers to conventional and complementary therapies and practices used together in a coordinated way. Integrative health also emphasizes care of the whole person that aims to improve health in multiple interconnected domains: social, psychological, and physical, including multiple organs and systems.

The term *whole person health* involves looking at the whole person—not just separate organs or body systems—and considering multiple factors that promote either health or disease. It means helping and empowering individuals, families, communities, and populations to improve their health in multiple interconnected biological, behavioral, social, and environmental areas. Instead of treating a specific disease, whole person health focuses on restoring health, promoting resilience, and preventing diseases across a lifespan.

The use of integrative approaches to health and well-being has grown within care settings across the United States. Researchers are currently exploring the potential benefits of integrative health in a variety of situations, including pain management for military personnel and veterans, relief of symptoms in cancer patients and survivors, and programs to promote healthy behaviors.

Although complementary therapies and practices vary widely, it is useful to classify them by their primary therapeutic input, which may be dietary (e.g., diet, herbs), psychological (e.g., meditation), physical (e.g., massage, acupuncture), or the combination of psychological and physical (e.g., yoga, tai chi). Although some complementary health practices are recommended or provided by a physician or a complementary health care provider such as a chiropractor, acupuncturist, or naturopathic practitioner, many of these practices are undertaken as “self-care.” Although some are reimbursed, most are paid for out of pocket.

PATTERNS OF USE

The first large survey of use of complementary health practices was performed by David Eisenberg and associates in 1993. It surprised the medical community by showing that >30% of Americans use complementary health products and practices. Many surveys since that time have extended those conclusions. The National Health Interview Survey (NHIS), a large, national household survey in which thousands of Americans are interviewed about their health- and illness-related experiences, is conducted annually by the National Center for Health Statistics, a component of the Centers for Disease Control and Prevention. This survey, which addressed the use of complementary health practices in 2002, 2007, 2012, 2017, and 2022 uses methods that create a nationally representative sample and has a sample size large enough to permit valid estimates about some subgroups.

An analysis of data from 27,651 adults in the most recent survey, which was conducted in 2022, evaluated changes in the U.S. adult use of seven complementary health therapies and practices over a 20-year period (from 2002 to 2022): yoga, meditation, massage therapy, chiropractic care, acupuncture, naturopathy, and guided imagery/progressive muscle relaxation. Over 20 years, U.S. adults not only increased their overall use of complementary health approaches but were also more likely to use them specifically for managing pain. In 2022, 36.7% of people used at least one of the seven approaches, compared to 19.2% in 2002. Use of yoga, meditation, and massage therapy increased the most from 2002 to 2022. Use of yoga increased from 5% in 2002 to 15.8% in 2022, rising from the fifth to the second most-used practice. Meditation increased from 7.5% in 2002 to 17.3% in 2022, and it remained the most-used complementary health practice over the 20 years. The 2012 survey, for which there are data about use of natural products, yielded the estimate that nonvitamin, nonmineral dietary supplements are used by ~18% of adults and 5% of children.

Americans often pay out-of-pocket for complementary health products and practices; the estimated out-of-pocket expenditure for complementary health practices in 2012 was $30.2 billion ($28.3 billion for adults and $1.9 billion for children), representing 1.1% of total health expenditures and 9.2% of out-of-pocket costs. On visits to complementary practitioners, Americans spent $14.7 billion out of pocket,

TABLE 495-1 Glossary of Complementary and Integrative Health Therapies and Practices

Acupuncture	A family of procedures involving stimulation of defined anatomic points, a component of the major Asian medical traditions; most common application involves penetrating the skin with thin, solid, metallic needles that are manipulated by the hands or by electrical stimulation
Ayurvedic medicine	The major East Indian traditional medicine system; treatment combines products (mainly derived from plants, but may also include animal, metal, and mineral), diet, exercise, and lifestyle
Biofeedback	The use of electronic devices to help people learn to consciously control body functions such as breathing or heart rate
Chiropractic	Chiropractic care involves the adjustment of the spine and joints to influence the body's nervous system and natural defense mechanisms to alleviate pain and improve general health; primarily used to treat back problems, headaches, nerve inflammation, muscle spasms, and other injuries and traumas
Dietary supplement	A product that is intended to supplement the diet, is taken by mouth, contains one or more dietary ingredients (including vitamins, minerals, herbs, amino acids, or certain other substances), and is labeled as being a dietary supplement
Homeopathy	A medical system with origins in Germany that is based on a core belief in the theory of "like cures like"—compounds that produce certain syndromes, if administered in very diluted solutions, will be curative
Hypnosis	The induction of an altered state of consciousness characterized by increased responsiveness to suggestion
Massage	Manual therapies that manipulate muscle and connective tissues to enhance the function of those tissues and promote muscle relaxation and well-being
Meditation	A group of practices, largely based in Eastern spiritual traditions, intended to focus or control attention and obtain greater awareness of the present moment, or mindfulness
Mind and body practices	A large and diverse group of procedures or techniques that are administered or taught by a trained practitioner or teacher; examples include acupuncture, massage therapy, meditation, relaxation techniques, spinal manipulation, tai chi, and yoga
Natural products	A variety of products such as herbs (also known as botanicals), vitamins and minerals, and probiotics, which are widely marketed, readily available to consumers, and often sold as dietary supplements
Naturopathy	A clinical discipline that emphasizes a holistic approach to the patient, herbal medications, diet, and exercise; practitioners have degrees as doctors of naturopathy
Osteopathy	A clinical discipline, now incorporated into mainstream medicine, that historically emphasized spinal manipulative techniques to relieve pain, restore function, and promote overall health
Qigong	A mind and body practice originating in China that involves using exercises to optimize energy within the body, mind, and spirit, with the goal of improving and maintaining health and well-being
Relaxation techniques	A number of practices such as progressive relaxation, guided imagery, biofeedback, self-hypnosis, and deep breathing exercises, with the goal of producing the body's natural relaxation response, characterized by slower breathing, lower blood pressure, and a feeling of increased well-being
Spinal manipulation, osteopathic manipulation	A technique where practitioners use their hands or a device to apply a controlled thrust (i.e., a force of a specific magnitude or degree in a specific direction) to a joint of the spine
Tai chi	A mind and body practice originating in China that involves slow, gentle movements and sometimes is described as "moving meditation"
Traditional Chinese medicine	A medical system that uses acupuncture, herbal mixtures, massage, exercise, and diet

which is almost 30% of what they spent out of pocket on services by conventional physicians ($49.6 billion). On natural products, such as dietary supplements, Americans spent $12.8 billion out of pocket, which was about one-quarter (24%) of what they spent out of pocket on prescription drugs ($54.1 billion).

Trends are even more striking for pain conditions. According to the NHIS surveys, painful conditions are the most common reasons why American adults use complementary health products and practices. About 40 million American adults experience severe pain in any given year, and they spend >$14 billion out of pocket on complementary therapies to manage their pain. A recent analysis of NHIS data showed a notable rise in the proportion of U.S. adults using complementary health approaches specifically for pain management. Among participants using any of the complementary health approaches, the percentage reporting use for pain management increased from 42.3% in 2002 to 49.2% in 2022.

Some patients seek out complementary health practitioners because they offer greater personal attention. For others, therapies and practices perceived as outside the mainstream reflect a "self-help" approach to health and well-being or satisfy a search for "natural" or less invasive alternatives. Since dietary supplements are labeled as "natural," they are often believed, incorrectly, to be inherently healthy.

CATEGORIES OF COMPLEMENTARY AND INTEGRATIVE HEALTH THERAPIES AND PRACTICES BASED ON PRIMARY THERAPEUTIC INPUT

■ PRIMARY DIETARY INPUT

Natural products, including plant and animal products, have a long and impressive history as sources of medicine and as important resources for biologic research. Whether as herbal supplements or as part of a diet, natural products are frequently consumed as a complex mixture of substances. This complexity can be further amplified by potential interactions with endogenous metabolic pathways, including those associated with the microbiome. The result is a collection of natural products and their metabolites that, individually and/or collectively, are associated with a network of biologic activity. Importantly, in addition to direct action on biologic targets, the activity of natural products can be influenced by an individual's diet, health, and metagenomic background. Although much remains to be understood about mechanisms of action, results of research on some natural products for a few conditions appear promising. In addition, in the 2012 NHIS, users of natural product supplements were twice as likely to report taking the natural product for a general well-being reason than for treatment of a specific health condition (88.9 vs 44.9%, respectively). Although to date, research on natural products has focused on their use for specific diseases as outlined below, a better understanding is needed about how natural products, including food, can be used most effectively to support health.

Cannabinoids An increasing amount of attention has been given recently to the nonpsychogenic effects of cannabinoids, such as cannabidiol (CBD), and terpenes found in the cannabis plant on chronic pain, particularly neuropathic pain; studies have found some limited evidence that these medicines produced better pain relief than placebos. Cannabinoids (cannabis extract, synthetic tetrahydrocannabinol [THC]) have been studied for therapeutic effects in multiple sclerosis (MS) and may relieve spasticity as well as pain in people with MS; however, no marijuana-derived medications are approved by the U.S. Food and Drug Administration (FDA) to treat MS. Sativex, an oral mucosal spray containing a mixture of THC and CBD, has received regulatory approval in >25 countries outside the United States for the treatment of spasticity (muscle stiffness/spasm) due to MS. Sativex is currently licensed in the United Kingdom for use as an add-on treatment for MS-related spasticity when people have shown inadequate response to other symptomatic treatments. Importantly, the psychoactive properties and other potential adverse effects of preparations containing cannabinoids need to be considered, including interactions with other medications and natural products; more research is needed in this area.

Melatonin Melatonin has been shown to help reduce anxiety in patients who are about to have surgery and may be as effective as standard treatment with midazolam in reducing preoperative anxiety. Findings from clinical trials support the use of melatonin supplements for sleep problems caused by shift work or jet lag and for improving sleep-onset latency and daytime sleepiness in people with insomnia. However, there are safety concerns about the use of melatonin by children and teenagers. U.S. sales of melatonin increased by about 150% during the COVID-19 pandemic, and the number of reports to U.S. poison control centers about pediatric melatonin ingestion increased from 8337 in 2012 to 52,563 in 2021. Further, according to a study published in *JAMA*, a majority of melatonin "gummy" products were inaccurately labeled, with most products exceeding the declared amount of melatonin and CBD.

Omega-3 Fatty Acids Clinical trials on rheumatoid arthritis (RA) have found that fish oil supplements can help alleviate tender joints and morning stiffness and reduce the daily nonsteroidal anti-inflammatory drug (NSAID) requirement of RA patients; however, data are not as definitive for other pain conditions. Gamma-linolenic acid (GLA) is an omega-6 fatty acid found in the oils from some plants, including evening primrose (*Oenothera biennis*), borage (*Borago officinalis*), and black currant (*Ribes nigrum*). Although oils containing GLA may have some benefit in relieving RA symptoms, only a few studies have been conducted on each of the oils. At present, it is uncertain whether omega-3 fatty acid supplementation is useful for depression. Some studies have shown small effects in adjunctive therapy in patients with a diagnosis of major depressive disorder (MDD) and in depressive patients without a diagnosis of MDD; however, most trials have been adjunctive studies. Controlled trials of omega-3 fatty acids as monotherapy are inconclusive compared to standard antidepressant medicines, and it remains unclear whether a mechanism is present to suggest that a pharmacologic or biologic antidepressant effect exists. Furthermore, there is evidence that a high dosage of fish oil supplementation is associated with a significant increased risk of atrial fibrillation (AF) compared with placebo.

Antioxidants Findings from the Age-Related Eye Disease Studies (AREDS and AREDS2) suggest that dietary supplementation with antioxidant vitamins may slow the progression of age-related macular degeneration (AMD). Compared to the original AREDS formulation, the AREDS2 formulation replaced beta-carotene with lutein and zeaxanthin due to increased risk of cancer in smokers taking high dose beta-carotene. Of note, in AREDS2, supplementation with lutein/zeaxanthin only appeared to be beneficial in participants with low dietary lutein and zeaxanthin. In a similar vein, a study using baseline data from the AREDS cohort reported that individuals eating healthier diets, characterized by higher intake of vegetables, whole grains, and seafoods, compared to those eating less healthy "Western" diets, were less likely to show signs of early AMD. It is therefore unclear at present whether the AREDS formula should be recommended for the general population regardless of diet.

Challenges of Research on Natural Products One challenge in this area is the extremely varied doses of natural products that are sold over the counter and used without much guidance or evidence of efficacy. We also know from research on vitamins that "more is not necessarily better" and that taking a "natural" substance such as a vitamin in quantities that greatly exceed what is found in food can be harmful.

Additional challenges in the assessment of plant products include their complexity and variability, including possible instability of active components or the presence of impurities, conflicting or unreliable conclusions in the literature, and low statistical power of studies. Further, there is a paucity of data on the safety of many products, including the safety of their use in a twenty-first-century context (e.g., if taken with modern prescription drugs) and their appropriate use in the context of traditional or indigenous practices.

Regulation There is an important distinction between natural products sold as dietary supplements and drugs developed from natural sources that are used to treat specific diseases. The Dietary Supplement Health and Education Act (DSHEA), passed in 1994, gives authority to the FDA to regulate dietary supplements, but with expectations that differ in many respects from the regulation of drugs or food additives. Purveyors of dietary supplements cannot claim that they prevent or treat any disease. They can, however, claim that they maintain "normal structure and function" of body systems. For example, a product cannot claim to treat arthritis, but it can claim to maintain "normal joint health."

Homeopathic products predate FDA drug regulations and are sold with no requirement that they be proved effective. Although homeopathic products are widely believed to be safe because they are highly dilute, one product, a nasal spray called Zicam, was withdrawn from the market when it was found to produce anosmia, probably because of significant zinc content. In January 2017, the FDA warned consumers about homeopathic teething tablets containing belladonna that pose a serious risk to infants and children.

Regulation of advertising and marketing claims is the purview of the Federal Trade Commission (FTC). The FTC does take legal action against promoters or websites that advertise or sell dietary supplements with false or deceptive statements. Misleading marketing of dietary supplements, homeopathic products, and indeed other complementary health products and practices contributes to the very significant risk that individuals will use them instead of effective conventional modalities. For example, in April 2020, the FTC sent warning letters to several companies allegedly selling unapproved products—some of which included high-dose dietary supplements—that may violate federal law by making deceptive or scientifically unsupported claims about their ability to treat or cure COVID-19.

Inherent Toxicity Although the public may believe that "natural" equates with "safe," it is abundantly clear that natural products can be toxic. Misidentification of medicinal mushrooms has led to liver failure. Contamination of tryptophan supplements caused the eosinophilia-myalgia syndrome. Herbal products containing particular species of *Aristolochia* were associated with genitourinary malignancies and interstitial nephritis. In 2013, dietary supplements containing 1,3-dimethylamylamine (DMAA), often touted as a "natural" stimulant, led to cardiovascular problems, including heart attacks. Among the most controversial dietary supplements is *Ephedra sinica*, or ma huang, a product used in traditional Chinese medicine for short-term treatment of asthma and bronchial congestion. The scientific basis for these indications was revealed when ephedra was shown to contain ephedrine alkaloids, especially ephedrine and pseudoephedrine. With the promulgation of the DSHEA regulations, supplements containing ephedra and herbs rich in caffeine sold widely in the U.S. marketplace because of their claims to promote weight loss and enhance athletic performance. Reports of severe and fatal adverse events associated with use of ephedra-containing products led to an evidence-based review of the data surrounding them, and in 2004, the FDA banned their sale in the United States.

A major current concern with dietary supplements is adulteration with pharmacologically active compounds. Multi-ingredient products marketed for weight loss, bodybuilding, "sexual health," and athletic performance are of particular concern. Recent FDA recalls have involved contamination with steroids, diuretics, stimulants, and phosphodiesterase type 5 inhibitors.

Herb-Drug Interactions A number of natural products have potential impacts on the metabolism of drugs. This effect was illustrated most compellingly with the demonstration in 2000 that consumption

TABLE 495-2 Resources for Dietary Supplement–Drug Interactions

National Institutes of Health National Center for Complementary and Integrative Health (NCCIH)	*https://www.nccih.nih.gov/health/know-science/how-medications-supplements-interact*	The National Institutes of Health NCCIH Know the Science initiative provides information for patients about complex scientific health topics such as drug-supplement interactions.
Medscape	*http://www.medscape.com/druginfo/druginterchecker?cid=med*	This website is maintained by WebMD and includes a free drug interaction checker tool that provides information on interactions between two or more drugs, herbals, and/or dietary supplements.
NatMed	*https://naturalmedicines.therapeuticresearch.com/tools/interaction-checker.aspx*	This website provides an interactive natural product–drug interaction checker tool that identifies interactions between drugs and natural products, including herbals and dietary supplements. This service is available by subscription.

of St. John's wort interferes with the bioavailability of the HIV protease inhibitor indinavir. Later studies showed its similar interference with metabolism of topoisomerase inhibitors such as irinotecan and with cyclosporine and many other drugs. The breadth of interference stems from the ability of hyperforin in St. John's wort to upregulate expression of the pregnane X receptor, a promiscuous nuclear regulatory factor that promotes the expression of many hepatic oxidative, conjugative, and efflux enzymes involved in drug and food metabolism.

Because of the large number of compounds that alter drug metabolism and the large number of agents some patients are taking, identification of all potential interactions can be a daunting task. Several useful Web resources are available as information sources (Table 495-2). Clearly, attention to this problem is particularly important with drugs with a narrow therapeutic index, such as anticoagulants, antiseizure medications, antibiotics, immunosuppressants, and cancer chemotherapeutic agents. Although there are many examples of substances of natural origin successfully used as pharmaceutical drugs, in general, natural products ingested as food or herbal teas, rather than concentrated extracts, are less likely to cause harm.

PRIMARY PSYCHOLOGICAL INPUT

Therapies and practices whose primary therapeutic input is predominantly mental include conventional types of psychotherapy, such as cognitive behavioral therapy (CBT), and complementary practices, such as meditation and mindfulness-based stress reduction (MBSR). Relaxation techniques, including biofeedback-assisted relaxation, also fall into this category. The boundary between conventional and complementary can be blurred, as CBT programs, for example, frequently incorporate elements of MBSR and relaxation techniques. These therapies and practices are being gradually integrated into aspects of conventional care, such as cardiac rehabilitation programs, and are playing an increasingly recognized role in the management of pain, as well as stress and sleep disturbances.

Cognitive Behavioral Therapy (CBT) The American College of Physicians practice guidelines (2016) strongly recommend the use of CBT for insomnia (also called CBT-I) as the initial treatment for chronic insomnia. Although CBT-I often includes relaxation techniques, it is not clear whether relaxation alone is beneficial. Various online applications are increasing the accessibility of these techniques at low cost.

Mindfulness-Based Stress Reduction (MBSR) Mindfulness meditation has been found to significantly reduce pain in experimental and clinical settings and to improve a wide spectrum of clinically relevant cognitive and health outcomes, including low-back pain and fibromyalgia. Recent findings from neuroimaging and randomized controlled trials confirm that mindfulness meditation reduces pain by engaging multiple, unique, nonopioidergic mechanisms that are distinct from placebo and that vary across meditative training level. There is some growing evidence that mindfulness meditation can have a beneficial effect on anxiety and help people recover from substance use disorders.

Hypnosis Findings from a few studies have demonstrated that training patients in the use of self-hypnosis significantly reduced their need for sedatives and analgesia when undergoing interventional radiologic procedures. Some studies also have suggested that hypnosis may be helpful for anxiety and health-related quality of life in people with irritable bowel syndrome (IBS). There is some evidence to suggest that hypnotherapy may improve smoking cessation, but data are not definitive.

Relaxation Techniques Relaxation techniques, including biofeedback and progressive muscle relaxation, may be helpful in managing a variety of stress-related health conditions, including anxiety associated with ongoing health problems and in those who are having medical procedures. Diaphragmatic breathing exercises may modestly lower blood pressure, reduce levels of cortisol, and reduce glycemia in people with type 2 diabetes. The efficacy of biofeedback has been evaluated in numerous studies for tension headaches, with positive results. Several studies have shown that biofeedback decreased the frequency of both pediatric and adult migraines, with some showing an effect lasting over an average follow-up phase of 17 months. Evidence suggests that relaxation techniques may also provide some benefit for symptoms of posttraumatic stress disorder and help reduce occupational stress in health care workers. Clinical practice guidelines issued by the American Cancer Society on the evidence-based use of integrative therapies during and after breast cancer treatment recommend yoga for anxiety and stress reduction. For some of these conditions, relaxation techniques are used as an adjunct to other forms of treatment.

PRIMARY PHYSICAL INPUT

A physical therapeutic input can be delivered manually (e.g., massage) or using a device (e.g., acupuncture) or can be generated by the patient (e.g., exercise).

Acupuncture The role of acupuncture in pain management has been controversial for decades, with critics pointing out its "prescientific" theoretical basis, and indeed, the rationale for the use of specific "acupuncture points" remains to be established. However, recent large-scale meta-analyses have demonstrated acupuncture to be superior to both usual care and sham acupuncture for chronic musculoskeletal pain, headache, and osteoarthritis (OA), with beneficial treatment effects persisting for up to 12 months. Clinical practice guidelines issued by the American College of Rheumatology and the Arthritis Foundation conditionally recommend acupuncture for knee, hip, and/or hand OA. The most recent (2017) American College of Physicians clinical guidelines recommend acupuncture as one of the initial treatment options for patients with acute, subacute, and chronic low-back pain. Acupuncture may provide a modest reduction in symptoms of depression, particularly when compared with no treatment or a control. Acupuncture or electroacupuncture may be an appropriate addition to drug treatment for managing chemotherapy-induced nausea and vomiting in patients with cancer. Clinical guidelines issued by the Society for Integrative Oncology and the American Society of Clinical Oncology in 2022 found intermediate level of evidence (with moderate strength) to recommend that acupuncture, reflexology, acupressure, or massage may help relieve pain in people with cancer. Acupuncture may relieve symptoms of allergic rhinitis. Clinical practice guidelines from the American Academy of Otolaryngology–Head and Neck Surgery include acupuncture among the options that health care providers may offer to interested patients with allergic rhinitis.

Spinal Manipulation The role of both osteopathic and chiropractic spinal manipulative therapies (SMTs) in management of low-back pain also has been the subject of a number of carefully performed trials and many systematic reviews. Conclusions are not consistent, but the American College of Physicians guidelines conclude that spinal manipulation has a small effect on improving function and pain compared with control—either a sham manipulation or an inert treatment.

Although evidence for spinal manipulation for chronic low-back pain is graded as low quality, the recommendation for consideration of nonpharmacologic treatment including spinal manipulation is graded as a strong recommendation, reflecting increasing concern with the impact of chronic opioid use for low-back pain. The evidence of benefit of spinal manipulation for neck pain is not as extensive, and continued concern that cervical manipulation may occasionally precipitate vascular injury clouds a contentious debate.

Massage Low- to moderate-quality evidence suggests that massage therapy is superior to nonactive therapies in reducing arthritis pain and improving functional outcomes. Massage may provide short-term relief from low-back pain, but the evidence is not of high quality. There is some evidence that massage has a positive effect on migraine, tension headaches, and neck pain.

COMBINED PSYCHOLOGICAL AND PHYSICAL INPUT

The primary therapeutic input for other mind and body practices is a combination of physical and psychological. Examples of practices in this category include yoga and tai chi, which combine movement, physical postures, and meditation.

Yoga Yoga can be beneficial for patients with fibromyalgia or chronic low-back pain, and yoga compared to nonexercise controls results in small to moderate improvements in back-related function at 3 and 6 months. There is overall evidence that yoga benefits people's general well-being by relieving stress, supporting good health habits, and improving mental/emotional health and sleep. Yoga can also help with quitting smoking, anxiety or depressive symptoms associated with difficult life situations, and quality of life for people with chronic diseases.

Tai Chi Clinical practice guidelines issued by the American College of Rheumatology and the Arthritis Foundation strongly recommend tai chi, along with other nondrug approaches such as self-management programs, for managing knee and/or hip OA. Tai chi has been shown to improve overall motor function, including balance and stability in older adults. Tai chi may help improve sleep quality in individuals with mild insomnia. Tai chi also has been shown to improve quality of life in people with heart disease, cancer, and other chronic illnesses.

MULTICOMPONENT THERAPIES AND SYSTEMS

Multicomponent approaches to health comprise two or more interventions such as lifestyle changes, physical rehabilitation, psychotherapy complementary health practices, and conventional medicine in various combinations, with an emphasis on whole person health. Complementary health therapies and practices are often multicomponent in nature, both in traditional health systems (e.g., traditional Chinese medicine, naturopathy) and in modern integrative practice. The U.S. Veterans Health Administration uses a multicomponent model of pain care that emphasizes nonpharmacologic methods, both conventional (e.g., physical therapy, CBT) and complementary (e.g., yoga, acupuncture), and may also include nutrition consultations. Several medical systems, such as chiropractic, osteopathy, naturopathy, and homeopathy, that arose in the late nineteenth century continue to be practiced today. Osteopathic medicine is mostly integrated into conventional medicine, with the addition of specific osteopathic musculoskeletal manipulation techniques. While homeopathy and naturopathy have remained largely separate from mainstream medicine, chiropractic care is increasingly available in some conventional care settings. A number of multicomponent systems, often called "whole health" systems, such as traditional Chinese medicine, Ayurveda, and homeopathy, use a diagnostic and therapeutic framework that is different from that of conventional medicine, which has posed additional challenges to their rigorous investigation.

Naturopathy Naturopathy, or naturopathic medicine, is a multicomponent therapeutic system based on philosophical principles that guide practice. Naturopaths prescribe conventional and unconventional diagnostic tests and medications, with an emphasis on relatively low doses of drugs, herbal medicines, healthy diet, and exercise.

Chiropractic The practice of chiropractic care, founded by David Palmer in 1895, is the most widespread practitioner-based complementary health practice in the United States. Although the scope of practice varies widely, chiropractic practice emphasizes manual therapies for treatment of musculoskeletal complaints.

Osteopathic Medicine Founded in 1892 by the physician Andrew Taylor Still, osteopathic medicine was originally based on the belief that manipulation of soft tissue and bone can correct a wide range of diseases of the musculoskeletal and other organ systems. Over the ensuing century, the osteopathic profession has welcomed increasing integration with conventional medicine. Today, the postgraduate training, practice, credentialing, and licensure of osteopathic physicians are virtually indistinguishable from those of allopathic physicians. Osteopathic medical schools, however, include training in manual therapies, particularly spinal manipulation, as well as diagnostic methods based on palpation of musculoskeletal tissues that are not part of conventional medical education.

Homeopathy The theoretical framework of homeopathy is based on two unconventional principles: "like cures like," the notion that a disease can be cured by a substance that produces similar symptoms in healthy people; and the "law of minimum dose," the notion that the lower the dose of the medication, the greater its effectiveness. Although the current lack of biologic underpinning for these principles has seriously limited the rationale for their use, the diagnostic framework of homeopathy could be the source of new insights that could be explored. As previously discussed, the regulatory framework for homeopathic remedies differs from that for dietary supplements, in that homeopathic products are regulated as drugs under the Federal Food, Drug, and Cosmetic Act and are subject to the same requirements related to approval, adulteration, and misbranding as other drug products. There are currently no homeopathic products approved by the FDA. Homeopathic remedies are widely available and commonly recommended by naturopathic physicians, chiropractors, and other licensed and unlicensed practitioners.

Challenges of Clinical Research on Multicomponent Therapies and Systems Classic randomized controlled trial (RCT) designs may not be well suited for research on multicomponent complementary interventions and systems such as naturopathy and Ayurvedic medicine. The dynamic relationships among an array of factors that affect health and well-being is inherent to the philosophy of these systems of care and poses methodologic challenges to the effective application of conventional RCT design. Pragmatic comparative effectiveness designs with "usual care" comparators are widely used to study these types of interventions, and trials may need to take into account the individualization of interventions and the underlying theories of these multicomponent systems. Thus, a key component of research on multicomponent therapeutic systems is the development of validated and reproducible "manualized" treatment protocols allowing for some flexibility and individual patient care. Pragmatic studies that compare multicomponent treatments with usual care cannot determine which treatment components are responsible for benefits, but other kinds of translational studies can address this issue.

THERAPEUTIC OUTPUT—SYSTEMS IMPACTED AND CHALLENGES OF MECHANISTIC RESEARCH

Complementary and integrative interventions whose therapeutic input is dietary, psychological, and/or physical may exert their effects, or therapeutic output, through a variety of mechanisms and physiologic systems. For example, peppermint oil may relieve pain associated with IBS by directly relaxing gastrointestinal smooth muscle, probiotics may have effects on the nervous system as well as the gut, and some components of traditional Chinese medicine, as well as omega-3 fatty acids and their derivatives, have immune-mediated anti-inflammatory effects. Multicomponent interventions with psychological and/or physical therapeutic input such as meditation and acupuncture can have effects on the nervous system and may also target other body systems affected by the pain condition; for example, tai chi may improve balance and stability by increasing flexibility and core strength, and the stretching involved in yoga may improve low-back pain by reducing connective tissue inflammation. For all types of therapeutic input, biopsychosocial interactions also may be important; for example, participation in an integrative group therapy pain management program may provide tools to help relieve symptoms of anxiety and depression as well as pain.

Deepening the scientific understanding of the connections that exist across domains of human health is important to better understand how conditions interrelate, identify multicomponent interventions that address these problems, and increase the support of patients through the full continuum of their health experience, including the return to health. Studies of multicomponent interventions often require multidisciplinary expertise and use state-of-the-art techniques in areas such as neuroscience, immunology, pharmacognosy, proteomics, genetics, and epigenomics. Further, there are limited preclinical models for some complementary health interventions (e.g., no relevant animal model for meditative movement practices such as yoga or tai chi). Objective, validated measurement tools are essential, as are processes and procedures to ensure quality control, whether the intervention is a mind and body practice or a natural product.

PATIENT AND PROVIDER RESOURCES

Physicians regularly face difficult challenges in providing patients with advice and education about complementary health therapies and practices. Of particular concern to all physicians are practices of uncertain safety and practices that raise inappropriate hopes. Cancer therapies, antiaging regimens, weight-loss programs, and products that claim to improve sexual function or athletic performance are frequently targeted for excessive claims and irresponsible marketing. A number of Internet resources provide critical tools for patient education **(Table 495-3)**. Because many complementary health products and practices are used as self-care and because many patients research these interventions extensively on the Internet, directing patients to responsible websites can often be very helpful.

TABLE 495-3 Internet Resources on Complementary and Integrative Health Approaches

The Cochrane Collaboration Complementary Medicine Reviews
This website offers rigorous systematic reviews of mainstream and complementary health interventions using standardized methods. It includes >800 reviews of complementary health practices. Complete reviews require institutional or individual subscription, but summaries are available to the public.
http://www.cochrane.org/evidence
MedlinePlus All Herbs and Supplements, A–Z List
MedlinePlus Complementary and Integrative Medicine
MedlinePlus Dietary Supplements
These National Library of Medicine (NLM) Web pages provide an A–Z database of science-based information on herbal and dietary supplements; basic facts about complementary and integrative health practices; and federal government sources on information about using natural products, dietary supplements, medicinal plants, and other complementary health modalities.
http://www.nlm.nih.gov/medlineplus/druginfo/herb_All.html
https://medlineplus.gov/complementaryandintegrativemedicine.html
http://www.nlm.nih.gov/medlineplus/dietarysupplements.html
National Institutes of Health National Center for Complementary and Integrative Health (NCCIH)
This National Institutes of Health NCCIH website contains information for consumers and health care providers on many aspects of complementary and integrative health products and practices. Downloadable information sheets include short summaries of complementary health approaches, uses and risks of herbal therapies, and advice on wise use of dietary supplements.
http://www.nccih.nih.gov
Resources for Health Care Providers: *http://www.nccih.nih.gov/health/providers*
NCCIH Clinical Digest e-Newsletter: *http://www.nccih.nih.gov/health/providers/digest*
Continuing medical education lectures: *http://www.nccih.nih.gov/training/videolectures*
Herbs at a Glance fact sheets: *https://www.nccih.nih.gov/health/herbsataglance*

The scientific evidence regarding complementary therapies is fragmentary and incomplete. Nonetheless, in some areas, particularly pain management, it is increasingly possible to perform the kind of rigorous systematic reviews of complementary health therapies and practices that are the cornerstone of evidence-based medicine. A particularly valuable resource in this respect is the Cochrane Collaboration, which has performed >800 systematic reviews of complementary health practices. Practitioners will find this a valuable resource to answer patient questions. Practice guidelines, particularly for pain management, are also available from several professional organizations. Links to these resources are provided in Table 495-3.

SUMMARY

The frequent use of complementary and integrative health therapies and practices reflects an active interest among the public in improving health and well-being of the whole person. The current health care system is fragmented, with diseases and comorbid conditions mostly treated separately, sometimes with drugs that interact with one another. An important step in whole person health care is considering health and disease not as separate states but as a bidirectional continuum and understanding how complementary and integrative therapies and practices, which are often multicomponent in nature, consider a patient's long-term recovery and overall health.

Acknowledgment

Dr. Josephine Briggs contributed to this chapter in prior editions and some material from prior edition chapters has been retained here.

FURTHER READING

Black LI et al: Use of complementary health approaches among children aged 4–17 years in the United States: National Health Interview Survey, 2007-2012. National health statistics reports; no 78. Hyattsville, MD, National Center for Health Statistics, 2015.

Eisenberg DM et al: Trends in alternative medicine use in the United States, 1990–1997: Results of a follow-up national survey. JAMA 280:1569, 1998.

Gaston TE et al: "Natural" is not synonymous with "safe": Toxicity of natural products alone and in combination with pharmaceutical agents. Regul Toxicol Pharmacol 113:104642, 2020.

Ijaz N et al: Whole systems research methods in health care: A scoping review. J Altern Complement Med 25:S21, 2019.

Nahin RL et al: Expenditures on complementary health approaches: United States, 2012. Natl Health Stat Rep 95:1, 2016.

Nahin RL et al: Use of complementary health approaches overall and for pain management by US adults. JAMA 331:613, 2024.

Paige NM et al: Association of spinal manipulative therapy with clinical benefit and harm for acute low back pain: Systematic review and meta-analysis. JAMA 317:1451, 2017.

Qaseem A et al: Noninvasive treatments for acute, subacute, and chronic low back pain: A clinical practice guideline from the American College of Physicians. Ann Intern Med 166:514, 2017.

Skelly AC et al: Noninvasive nonpharmacological treatment for chronic pain: A systematic review up-date. Comparative Effectiveness Review No. 227. AHRQ Publication No. 20-EHC009. Rockville, MD, Agency for Healthcare Research and Quality; April 2020.

Vickers AJ et al: Acupuncture for chronic pain: Update of an individual patient data meta-analysis. J Pain 19:455, 2018.

496 Placebo and Nocebo Effects

Kathryn T. Hall, Alia J. Crum

Placebos are sham versions of drugs, devices, or surgeries that lack the active compound, function, or procedure they are designed to simulate **(Table 496-1)**. Administration of these "inactive" treatments can have significant therapeutic benefits called placebo effects, which accounts for their use as controls in clinical trials as comparators for active drugs, devices, or surgical interventions. Key drivers of placebo effects include the patient's expectation and conscious or subconscious conditioning. Psychological studies demonstrated that expectations are shaped by factors intrinsic to the patient, including their past experiences and core beliefs or mindsets, and extrinsic factors including environmental cues (e.g., white coat), clinical practice (e.g., physical examination), and information received about a treatment (e.g., expected benefits or side effects). Neuroimaging studies have identified consistent changes in the brain in response to placebo treatment that suggest that placebo effects work by integrating incoming information about extrinsic factors with prior experience and mindsets to update expectations of treatment benefit. When expectations are negative, they can result in negative outcomes, termed, "nocebo effects." For many years, placebo effects were viewed as superfluous nuisance variables to be ignored, marginalized, and simply "controlled for" in clinical trials. With advances in psychology and neuroscience studies of placebo effects, the value of understanding their mechanism of action and harnessing these effects in clinical care and clinical trials has become increasingly apparent.

INTRODUCTION TO PLACEBO EFFECTS: A BRIEF HISTORY

The word "placebo," derived from the Latin *placere*, "to please," first appeared in medical literature in clinical lectures of William Cullen, a leading physician in the eighteenth century. In 1792 he wrote, "I prescribed therefore in pure placebo, but I make it a rule even in employing placebos to give what would have a tendency to be of use to the patient." Cullen was describing placebos being used to please, rather than treat; however, use of placebos was commonplace at that time, especially when effective therapies were exhausted or unavailable.

From the late eighteenth through the early twentieth centuries, sham treatments were also used to expose some unorthodox treatments as frauds—or, at least, as no better than placebos. In his highly publicized 1784 study of "animal magnetism," a therapy developed by Austrian physician Anton Mesmer, Benjamin Franklin and a team of leading scientists in France simulated Mesmer's elaborate rituals using fake practitioners and sham magnetism. This early placebo-controlled trial

TABLE 496-1 Glossary of Terms commonly used in Placebo Studies

TERM	DEFINITION
Additivity in clinical trials	The assumption that placebo and drug treatment responses are additive is a fundamental assumption in clinical trials. However, there are notable exceptions to this assumption in pharmacogenomic and brain imaging studies where subsets of the population have been reported to have differential effects in placebo and drug treatment arms of a trial.
Development and culture	Our caregivers and social environment influence the psychological processes that underlie the placebo effect. These processes are continuously shaped throughout life by the ideas, institutions, and interactions that constitute the culture in which we live.
Expectation	A specific belief about the future based on a prediction of what is most likely to happen. Examples: "This drug will relieve my pain"; "I will experience side effects."
Gene-(drug/placebo) interaction	Pharmacogenomic analysis has identified clinical trials in asthma, depression, pain, chronic fatigue, and cognitive function in which there are subpopulations based on genotype that have differential associations in the drug and placebo treatment arms. These differential effects often result in significant gene-(drug/placebo) interaction effects.
Implicit learning	The nonconscious acquisition of knowledge. Classical conditioning, a form of implicit learning, is implicated in certain instances of the placebo effect (e.g., implicit association of sleepiness with the administration of blue pills).
Mindset	Core belief about a domain or category that orients an individual to a particular set of beliefs, associations, and expectations, and functions to guide attentional and motivational processes (e.g., "cancer is a catastrophe"; "symptoms are signs of efficacy").
Neurobiological mechanisms	Dopamine, endogenous opioids, and endocannabinoids are three of the major neurotransmitter systems implicated in moderating the placebo effect. Placebo effects also work by activating biological properties of the body that facilitate healing, including homeostatic mechanisms and immune and inflammatory responses. These contribute to the natural history of a disease but can also be targets of placebo effects.
Nocebo effect	Sides effect or negative change in clinical outcome observed after exposure to negative information, interactions, or cues that can induce negative expectations.
Open-label placebos (OLPs)	OLPs are placebos administered to patients with their full knowledge that the treatment lacks the active pharmaceutical agent. OLP clinical trials have been conducted in irritable bowel syndrome, chronic back pain, allergic rhinitis, cancer-related fatigue, attention deficit hyperactivity disorder, major depression, and menopausal hot flushes. Meta-analysis of OLP trials found a significant overall effect.
Patient-clinician relationship	The patient-clinician relationship shapes the mindsets and expectations a patient holds about health, illness, and treatments, and affects the quality of care a patient receives. This relationship is influenced by the warmth and competence of the provider and is further shaped by characteristics like empathy and trust.
Placebo	Placebos are sham versions of drugs, devices, or surgeries that lack the active compound, function, or procedure they are designed to simulate. Placebos are often used in clinical trials as controls for placebo effects, natural history, regression to the mean, spontaneous remission, and Hawthorne effects (the tendency for people to change their behaviors when being observed).
Placebo effect	Positive change in clinical outcome observed after a placebo treatment; an exclusively attributed expectation mediated by psychological, neurological, or physiological placebo mechanisms.
Placebo response	Improvement observed among patients assigned to placebo treatment in a clinical trial.
Placebome	The genome-related products that modify placebo response. Several genes in neurotransmitter and other pathways have been implicated in modifying response to placebo treatment in clinical trials. The most well-studied of these is in the gene encoding catechol-*O*-methyltransferase (COMT).
Social and observational learning	Learning through direct observation of others undergoing treatment (i.e., other patients) and interactions with individuals who weild influence over the patient (i.e., physicians and nurses) both may powerfully drive placebo effects.
Treatment characteristics	The specific characteristics include factors like the shape, color, and labeling of the treatment; the method of administration; and the physical environment in which the treatment is administered.

demonstrated that the hugely popular and remarkably effective treatment was no better than a placebo. Early trialists attributed the positive benefits of these therapies to the power of the "imagination," reinforcing the belief that placebo treatments were the product of wayward physicians, or "quacks," who tricked gullible patients.

After World War II, advances in metabolism, physiology, and clinical pharmacology created a growing need for clinical trials to evaluate novel compounds for their ability to kill pathogens and alter disease processes. By the 1960s, the Declaration of Helsinki and Kefauver-Harris Amendments, introduced to protect patient safety by requiring informed consent, rendered use of placebos and the deception historically thought to promote placebo effects unethical, institutionalizing the transformation of placebos from salve to epistemic tool. Together, placebo controls, double-blinding, and randomization are considered the gold standard for acquiring the strongest evidence of efficacy or lack of efficacy for novel treatments. Notably, placebos are rarely used in trials of serious illnesses like cancer or when an effective treatment already exists. Then their use is limited to comparing novel treatments to standard of care plus placebo.

In randomized placebo-controlled clinical trials, the effect of the drug is calculated by simply subtracting outcomes in the placebo treatment arm or placebo response from the drug response **(Fig. 496-1*A*)**. Hence, in addition to controlling for placebo effects, placebos control for changes in the outcome of interest due to natural history (the tendency for a common cold to resolve on its own in 7–10 days), regression to the mean (a statistical phenomenon where extreme baseline measures tend to move toward the group mean), and the Hawthorne effect (the tendency for people to change behaviors when being observed). These variables, together with placebo effects, make up the placebo response. If the active treatment being tested significantly outperforms the placebo response, it is deemed efficacious and can progress through the U.S. Food and Drug Administration (FDA) approval process. However, if response to the drug is not statistically significantly greater than the placebo response, the active treatment is deemed lacking in efficacy.

Clinical trials' limited view of placebos obscures the reality that the effects associated with placebos are not, in practice, superfluous. Indeed, the effects of patient psychology (e.g., expectations, mindsets) and the social and cultural context (e.g., a clinician's demeanor and drug label and advertising information) have a meaningful impact on health outcomes that warrant a more complete understanding. In addition, the effects of these factors are difficult to isolate. In fact, the total response to drug is the product of both the drug and the social and psychological context interacting with patient biology to bring about change **(Fig. 496-1*B*)**. This perspective propels us into a new era of understanding placebo effects: not as treatment alternatives or as something to subtract, but as psychological, social, and biological mechanisms that can be considered an integral component of the overall treatment effect in medicine. By understanding placebo effects in this manner, we can optimize their benefits to improving drug discovery, maximizing existing treatments, and minimizing nocebo consequences to reduce harm.

PLACEBO RESPONSE IN CLINICAL TRIALS

Placebo effects, although often thought of in the context of a placebo pill, extend to many other treatment modalities, including sham surgeries, placebo acupuncture, and placebo diets. They have been documented in numerous conditions and diseases, including pain, depression, Parkinson's disease, anxiety disorders, cardiovascular disorders, cancer-related fatigue, asthma, and gastrointestinal disorders. Not limited to patient-reported outcomes, placebo effects can affect objective physiologic outcomes, including blood pressure measurements, immune biomarker levels, exercise endurance, and cognitive test scores.

Recently, clinical trial sponsors have invested considerable resources in reducing the impact of placebo response by adjusting patient-level variables, such as reducing patient-clinician interactions and reducing patient expectations by providing neutral information about expected benefits and side effects. In conditions like Alzheimer's disease, trialists have modeled placebo response over time and set the optimal treatment duration length to the time period beyond the 8-week period when placebo response is maximal. The placebo run-in design attempts to eliminate the influence of placebo responders by assigning all enrolled blinded patients to a placebo at the beginning of a study; patients who respond to placebo are subsequently excluded. Other more complex models, such as the sequential parallel comparison design (SPCD), randomize patients to placebo or drug at a ratio of 2:1. After a brief treatment period, placebo nonresponders are rerandomized to placebo or drug. Unlike placebo run-ins, SPCD uses all patients and, thus, should have greater power to find an effect. Still, despite these and other considerable investments, no approach reliably reduces the impact of placebo response on clinical trial failure.

MECHANISMS OF PLACEBO EFFECTS

PSYCHOLOGICAL MECHANISMS

Expectations Expectations, or beliefs about the likelihood of future events, are thought to be the key driver of placebo effects. Expectations can be conscious; for example, many patients expect nonsteroidal anti-inflammatory drugs (NSAIDs) will relieve pain, melatonin will improve sleep, and beta blockers will reduce anxiety. Expectations can also be consciously or subconsciously conditioned. Repeated use of blue sleeping pills can induce sleepiness by taking a blue placebo pill. Multiple sclerosis patients who received the immunosuppressant cyclophosphamide paired with flavored syrup later displayed drug-consistent immune responses to the flavored syrup alone. Observational learning can also play a role in eliciting placebo effects by altering expectations. Watching another person experience pain relief in response to a particular treatment can lead the observer to expect to experience similar relief, even if the stimulus is a placebo in both cases.

Mindsets Mindsets are core beliefs about a broader domain or category, such as the meaning of side effects or the nature of a disease or treatment. Mindsets orient individuals to a set of associations, expectations, and goals. A mindset such as "cancer is a catastrophe," "statins are harmful," or "my body is permanently damaged" can shape a patient's experience of pain or other side effects. While mindsets and expectations are related, they are not identical. For example, a patient in pain may have the specific expectation that a treatment will relieve their pain. But they also could have broader mindsets in which those expectations are operating, such as "injections don't work," "my condition is hopeless," or "I am in good hands."

While specific expectations can be measured and assumed to influence placebo effects in studies, mindsets may be particularly powerful in the real-world practice of medicine where individual expectations do not exist in isolation. Mindsets may also be advantageous when considering how to harness placebo and minimize nocebo effects ethically.

SOCIAL AND CULTURAL MECHANISMS

Language and Information Patients' implicit or explicit preexisting mindsets, shaped by the broader culture in which they were raised and/or reside, can be updated and informed through verbal instructions. In "open-hidden design" studies, medication administered openly by a health care professional who informs the patient that they will experience benefit (e.g., "I'm going to administer a dose of morphine, a powerful painkiller that will alleviate your pain") has a significantly greater analgesic effect compared to administering the same dose from a hidden pump without the patient's knowledge. Thus, even potent opioid analgesics lose as much as 30% of their efficacy if the patient is unaware that they received the treatment. Effects of open-hidden paradigms are also seen with objective outcomes, such as heart rate. Information is conveyed not just by a clinician's words but also through information in the health care context more broadly, such as advertising and media related to drugs and health.

Clinician Characteristics Beyond what the patient is told, trust in the source of information can also influence clinical outcomes. Social-psychological research has shown that two qualities are key: patients' perceptions of *competence*, or whether a physician "gets it" (i.e.,

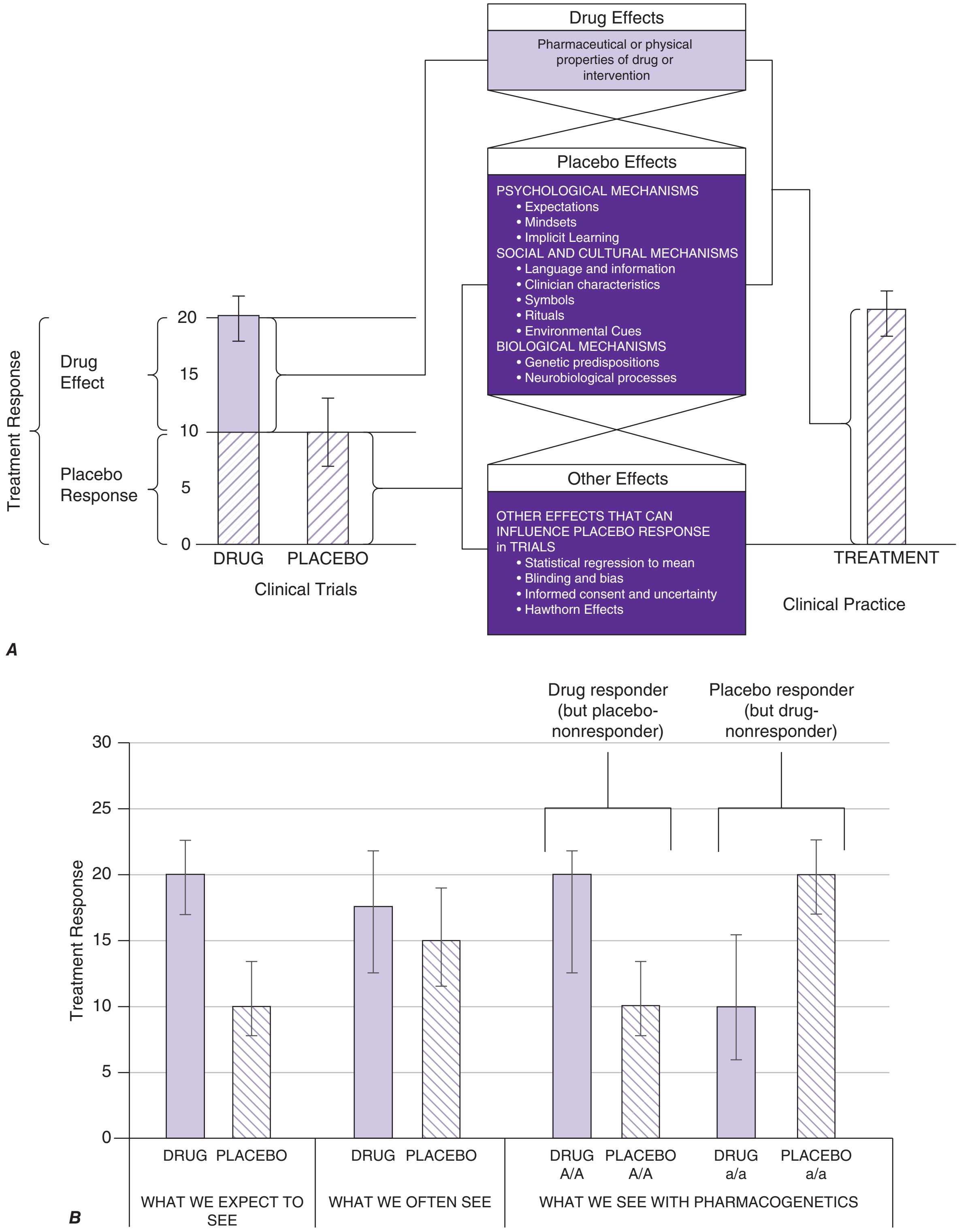

FIGURE 496-1 ***A.*** Additivity of drug and placebo response in a clinical trial. ***B.*** (1) What we expect to see: Expected outcomes from the classic view of clinical trials in which the effect of the drug exceeds placebo response. (2) What we often see: Typical results from a clinical trial in which there is no significant difference between the drug and placebo responses. (3) What we can see with pharmacogenomics: Pharmacogenomic analysis demonstrating differential effects of a genetic locus in the drug and placebo arms of a trial such that the drug effect and placebo response of one version of a variant (patients who are homozygous for the "A" allele, A/A) are opposite that of homozygotes of the alternate allele at that locus (patients who homozygous for the "a" allele, a/a). The average effects of outcomes of the two subpopulations cancel each other to give the results we often see.

displays of efficiency, knowledge, and skill), and patients' perceptions of *warmth*, or whether a physician "gets me" (i.e., displays of personal engagement, connection, and care for the patient).

Patients' assessments of clinician warmth and competence shape their treatment expectations and impact their mindsets about illness and, therefore, can influence placebo effects. In one study, an allergic reaction was induced in participants via a histamine skin prick followed by the administration of a placebo cream. The information about the cream was varied to create either positive expectations ("this cream will reduce your rash and irritation") or negative expectations ("this cream may worsen your symptoms"). When the ensuing wheal was measured 10 min later, the difference in information alone produced differences in the size of the allergic reaction. Interestingly, the effect of the information differed depending on perceived clinician warmth and competence. When the physician exhibited cues of both competence (e.g., wearing a white coat with a badge that read "Fellow at the Stanford Allergy Center") and warmth (e.g., making eye contact with the patient), the effect of the spoken message was significantly enhanced. However, when social cues were changed to induce questions about the level of competence (e.g., badge read "Student Doctor") and warmth (e.g., staring at a computer screen or making a personal connection), the information about the cream no longer mattered: both placebo and nocebo effects were minimized.

There is no one right way to signal warmth and competence, but there are many ways to fail to convey these qualities and, therefore, lose patients' trust. Indications of warmth and competence are important for building patients' trust not only with their medical providers but also with the broader clinic, hospital, or health care system. Trust in the social context can magnify placebo effects or minimize nocebo effects; it can also have a direct effect on patient care, influencing a wide array of outcomes, such as metabolic complications, immune response, symptoms, and adherence to medication.

Symbols, Rituals, and Environmental Cues Medical treatment occurs within a rich context of environmental cues, symbols, and rituals. Many patients, for example, exhibit a transient (albeit substantial) rise in blood pressure when in a medical setting, a phenomenon known as the "white coat" hypertension. Seemingly inconsequential characteristics of a drug, such as color, drug brand name, and cost, have been found to impact treatment efficacy. Patients tend to perceive capsules as stronger and more effective than tablets and tend to have a reduced response to placebos referred to as discounted or generic. In a within-subjects, repeated-measures study to examine drug labeling in migraine, patients were given either placebo or 10-mg rizatriptan labeled to create three information conditions ranging from negative ("placebo"), to neutral ("Maxalt or placebo") to positive ("Maxalt"). While patients had significantly greater relief from Maxalt labeled as Maxalt compared to placebo labeled as placebo, there was no difference in the effect of Maxalt labeled as placebo compared to placebo labeled as Maxalt. These findings demonstrate the ability of labeling and brand names to influence the effect of a drug. Interestingly, drug companies have increased their use of the letters X and Z in drug names as studies show these visually distinct letters have fewer negative associations with other medications. Furthermore, marketing research has found Z is associated with efficacy, whereas T and S are associated with greater tolerability.

■ BIOLOGICAL MECHANISMS

Pharmacologic evidence from the 1970s showing that naloxone, an opioid antagonist, could abrogate placebo effects in the experience of pain after molar tooth extraction laid the groundwork for demonstrating that psychological forces could affect patient physiology. Early neuroimaging studies revealed the release of endogenous opioids and dopamine signaling proportionate to the expectation and perception of how well a given placebo intervention worked. Using models of placebo analgesia, neuroscientists imaged the brains of healthy subjects exposed to various forms of thermal or mechanical pain-producing stimuli in the presence or absence of placebo treatments (e.g., inert creams or inactive transcutaneous electrical nerve stimulation devices) to induce placebo effects using conditioning or suggestion. These studies revealed activation of regions in the spinal cord and descending pain-control regions in the brainstem and reduced signaling in the spinothalamic tract. Later meta-analyses of 20 of these studies confirmed reduced signaling proportionate to reported placebo analgesia in pain-related activity in the thalamus, insula, and habenula, while increased signaling was observed in frontal-parietal regions.

Today, placebo treatments are hypothesized to influence brain systems involved in "meaning-making" by constructing internal models that guide our understanding of incoming signals and their source, as well as implications for anticipatory events. These internal models provide predictive signals that are incorporated with incoming sensory signals to produce the sensations and symptoms experienced. In turn, these models inform our perceptions (mindsets) and shape our reactions, amplifying or attenuating perceptual and affective circuits. Thus, contextual information around placebo treatments, including the suggestion of benefit, the visual and behavioral cues provided via the ritual of treatment, and prior experiences of benefit, can modify the neural construction of the experience and, in turn, downstream physiologic consequences in the nervous, immune, endocrine, and cardiovascular systems.

THE CHALLENGE OF IDENTIFYING "PLACEBO RESPONDERS"

In drug development, identifying and excluding placebo responders could lead to more precise, and potentially smaller and less expensive, clinical trials. Predicting which patients are likely to respond to placebos could allow clinicians to optimize patient interactions and even support gradual replacement of drugs with side effects with placebo by dosage titration.

Research into psychological predictors of placebo responders found several personality traits and constructs to be associated with placebo responders, including optimism, habitual desire for control, fun and sensation seeking, neuroticism, self-efficacy, and internal locus of control. Functional magnetic resonance imaging (fMRI) has also been used to create brain-signaling profiles that are predictive of placebo responders. In a study of patients with chronic osteoarthritis pain, right midfrontal gyrus connectivity effectively identified placebo pill responders. Interestingly, in some subjects, the active drug in this study, duloxetine, appeared to enhance the placebo response, but in others, duloxetine reduced it. This finding suggests that while drug and placebo were additive for some patients, the interaction with placebo response diminished the drug effect for others.

The observation that placebo effects are influenced by opioid and dopaminergic signaling suggested that genetic variation in the synthesis, function, or metabolism of these neurotransmitters might influence the magnitude of placebo effects. This observation gave rise to the term "placebome" to describe the group of genome-related products that potentially influence individual response to placebo treatments. Members of the placebome were identified in candidate and genome-wide association studies (GWAS) of the placebo control arms of clinical trials. For example, there is evidence of differential effects associated with the genes *COMT* and *MAO-A*, which encode enzymes that metabolize dopamine; *OPRM1*, which encodes the opioid receptor; and *TPH2* and *HTR2A*, which encode proteins involved in serotonin signaling. Of these genes, *COMT* has the strongest evidence for association with placebo response in clinical trials of irritable bowel syndrome (IBS) and pain. To date, there are 29 genes associated with response to placebo in the GWAS catalog.

Although numerous psychological, neuroimaging, and genetic profiles of placebo responders have been proposed, these biomarkers were mostly derived from small studies and have not yielded consistent results in prospective studies. This is in large part due to the broad heterogeneity in variables intrinsic to patients, including their conditions and disease severity and duration, and extrinsic study variables, including treatment duration, inclusion criteria, study location, number of study visits, outcome measures, and information about the study drug (e.g., possible side effects).

■ ADDITIVITY IN CLINICAL TRIALS

The study of placebo responders has also led to important evidence of nonadditivity in clinical trials. Additivity between drug and placebo outcomes has been a universal and fundamental, but unproven, assumption in clinical trials. Based on additivity, the drug effect is determined by subtracting the outcome in the placebo arm from the outcome in the drug treatment arm (Fig. 496-1*A*). As reported in the study investigating brain connectivity of patients with depression described above, in some patients, duloxetine enhanced brain connectivity seen in the brain region associated with placebo response. However, in other patients, this connectivity was reduced with duloxetine. Similarly, in clinical trials of chronic pain, chronic fatigue syndrome, cardiovascular disease, and asthma, significant genetic associations observed in placebo arms were null or found to be in the opposite direction in drug treatment arms. These unexpected gene-by-(drug/placebo) interaction effects suggest that, in some clinical trials, there are subsets of patients for whom the drug and placebo response is not additive (Fig. 496-1*B*). Thus, the potential for differential outcomes in the drug and placebo arms could confound outcomes in clinical trials, warranting further investigation.

NOCEBO EFFECTS

In part, because informed consent in clinical trials requires disclosure of all potential drug side effects, the side effects reported by patients randomized to placebo are often the same as those expected with randomization to drug. When this phenomenon was first documented in 1961, the word nocebo, coined from the Latin *nocere*, "to harm," was used to describe production of negative effects from negative verbal suggestions, contextual cues, or associative learning.

Although the term *nocebo* was originally defined as an adverse effect from an inert treatment, nocebo effects, like placebo effects, are the product of underlying aspects of the patients' psychology (in this case, negative expectations and mindsets) and the social context. In clinical trials, on average, 25% of participants randomized to placebo report side effects, and some studies (such as those of statins) show that the rates of side effects do not significantly differ between the active drug and placebo. Because they did not receive the active drug, we can assume the side effects arose, in part, due to expectation and not due to any active ingredients in the treatment.

Nocebo effects are not limited to clinical trials. While statins are effective cholesterol-lowering agents, the belief that they cause muscle pain is widespread, and treatment is often discontinued because of this reported side effect. This phenomenon has been extensively investigated. In one study, a "within-subjects" design was used in which each patient served as their own control. In this study, patients who reported side effects were blinded and randomized to take a placebo, statin, or no treatment on a monthly basis over a 1-year period. At the end of the study, there was no discernable difference between symptoms reported on placebo versus statin, and 50% of the patients in the trial were able to reinitiate statin therapy successfully. Negative expectations surrounding technology (e.g., wi-fi or cell phone signals), environmental agents (e.g., infrasound generated by wind turbines), or food (e.g., gluten) can also enhance the likelihood of negative symptoms related to their presence.

Ethically, it is hard to test nocebo effects deliberately, but randomized studies in laboratory contexts show that people who are told to expect side effects are more likely to experience those side effects than those who are not told to expect side effects. Expectations and mindsets can deepen negative effects through physiologic activation or by heightening awareness of symptoms that may have already been present, resulting in misattribution of their cause. Nocebo responses can also occur because of conditioned learning. For example, patients receiving chemotherapy can develop nausea when they see or smell a stimulus associated with their treatment, such as the treatment room or even a staff member.

Key drivers of placebo and nocebo effects overlap with factors that create barriers to quality clinical care for black patients and other patients of color. Poor communication, perceived discrimination, and medical mistrust are all factors demonstrated to reduce the quality of care in racially discordant dyads. Prior experiences of discrimination in the health care setting may result in expectations of discrimination and suboptimal clinician communication, enhancing nocebo and reducing placebo effects during treatment encounters. Consequently, the presence of nocebogenic and absence of placebogenic influences associated with racially discordant dyads has the potential to generate and exacerbate racial and ethnic inequities in clinical outcomes and care.

ETHICALLY AND DELIBERATELY HARNESSING PLACEBO EFFECTS

If placebo effects are understood as an integral component of the overall treatment effect, mediated by neurobiological processes and social and environmental factors, they can be personalized and maximized in the practice of medicine. Administering placebos without full knowledge of the patient is no longer acceptable for important ethical reasons. Moreover, administering "impure placebos"—pharmacologically active treatments that are prescribed at too low a dose to be effective or are known to be ineffective for the condition being treated—is reportedly common practice but still controversial and ethically problematic. The use of impure placebos varies by country. In the United States, a survey of 1200 internists and rheumatologists indicated that 62% of participants believed the practice of utilizing impure placebos—over-the-counter analgesics and vitamins—was ethically permissible. Notably, <5% reported using saline or sugar pills. The numbers are higher in Canada and the United Kingdom, where 80 and 77% respectively, of physicians surveyed reported prescribing impure placebos or treatments without proven or expected benefit. In Denmark, 86% of internists, 54% of hospitalists, and 41% of specialists in private practice report that they used placebo interventions at least one time within the previous year. Finally, the German Medical Association, after assessing placebos in medicine, published a report in 2011 acknowledging the complexity of the strong effects that placebos can have, supporting their limited use when no other therapies were available. In addition to using impure placebos, there are several other alternatives that deliberately leverage benefits of placebo effects.

OPEN-LABEL PLACEBOS

One way that researchers have addressed the ethical and legal limitations on placebos is to simply tell patients the facts about placebos. In honest or open-label placebo (OLP) studies, patients are fully informed about the absence of the active agent in the placebo, but are also told that placebo treatments can sometimes result in clinical improvement. The patients are also informed that trying a placebo might yield some benefit; even if they do not believe this to be the case, they could consider suspending their belief. To date, the findings on the effects of OLP are promising. Improvements have been observed in IBS, cancer-related pain and fatigue, depression, posttraumatic memory intrusions, allergic rhinitis, attention deficits, and hyperactivity; however, OLP has yielded no benefit in wound healing and did not enhance the cognitive abilities of healthy volunteers.

DELIBERATELY LEVERAGING PATIENT EXPECTATIONS AND MINDSETS

When understood as being driven by the psychological and social context, placebo effects can be evoked without the use of sham pills or procedures by deliberately shaping patient expectations and mindsets. At their best, doctors and patients alike already harness the forces behind placebo effects through interactions, branding, and language that inspire patients' trust, as well as useful mindsets and expectations about their condition. Once aware of this fact, health care practitioners can work ethically to leverage these forces to improve health outcomes.

In the PsyHEART trial, 124 patients undergoing coronary artery bypass surgery were randomized to standard care, supportive therapy, or an expectation of manipulation in which they were encouraged to develop clear expectations about how their life would improve after surgery (i.e., what activities they would be able to perform). Six months after surgery, patients who were randomized to the expectation manipulation showed significantly greater improvements in quality of life and reductions in disability. In the EMBRACE study, patients diagnosed with cancer were exposed to documentary-style films featuring experts in psychology and oncology and to cancer survivors who spoke about how their mindsets changed throughout and after their treatment, as well as challenges they faced along the way. Participants exposed to the films increased their health-related quality of life, such as their emotional well-being, physical health, and general functioning, by 10%, as measured by changes in industry-standard scales, compared to patients receiving treatment as usual. While these interventions were delivered directly to patients, trainings to help physicians and care teams more deliberately and effectively shape patients' expectations and mindsets in the context of their care are being developed, evaluated, and disseminated.

ETHICAL CONSIDERATIONS

In all of these applications, ethical considerations are of utmost importance; it is never acceptable to deceive or provide false information to the patient. Within these bounds, however, there is much we can do to improve patients' mindsets and expectations. Consider the nocebo effects resulting from informing patients about side effects. While it is not ethical to withhold this information from patients, providers could either provide more realistic expectations about the likelihood of side effects and set more adaptive mindsets about their meaning. In one study of children undergoing oral immunotherapy treatment (OIT) for peanut allergies, half were randomized to receive a typical warning message: side effects are negative outcomes, unrelated to treatment efficacy, that need to be managed and endured. The other half were given messages aimed to instill the mindset that some mild symptoms are often a sign that the treatment is working and signal desensitization. Compared with families informed that symptoms are negative side effects, families informed that "symptoms are positive signs of treatment efficacy" experienced significantly less anxiety, fewer symptoms during the highest doses, and improved levels of IgG4, an immune marker of allergic tolerance. Similar effects of this messaging have proven to reduce anxiety and side effects for those receiving the COVID-19 vaccine.

THE FUTURE OF PLACEBO EFFECTS

We are entering a new era of understanding about placebo effects, one in which they are not viewed as treatment alternatives or as something to subtract, but as psychological, social, and biological mechanisms that can be considered an integral component of the overall treatment effect in medicine. Work in this field is proliferating, and translation of the findings to clinical trials and clinical care is important for optimizing placebo effects to improve existing treatments while minimizing nocebo effects to reduce harm.

FURTHER READING

Colloca L et al: *Placebo Effects through the Lens of Translational Research*. New York, NY: Oxford University Press, 2023.

Evers AW et al: Implications of placebo and nocebo effects for clinical practice: Expert consensus. Psychother Psychosom 87:204, 2018.

Hall KT et al: Systems pharmacogenomics: Gene, disease, drug and placebo interactions: A case study in COMT. Pharmacogenomics 20:529, 2019.

Hall KT: *Placebos*. MIT Press Essential Knowledge Series. Cambridge, Massachusetts: The MIT Press, 2022.

Howe L et al: Changing patient mindsets about non-life-threatening symptoms during oral immunotherapy: A randomized clinical trial. J Allergy Clin Immunol Pract 7:1550, 2019.

Petrie KJ, Rief W: Psychobiological mechanisms of placebo and nocebo effects: Pathways to improve treatments and reduce side effects. Annu Rev Psychol 70:599, 2019.

Rief W et al: Preoperative optimization of patient expectations improves long-term outcome in heart surgery patients: Results of the randomized controlled PSY-HEART trial. BMC Med 15:4, 2017.

Zion SR, Crum AJ: Mindsets matter: A new framework for harnessing the placebo effect in modern medicine. Int Rev Neurobiol 138:137, 2018.

Zion SR et al: Changing cancer mindsets: A randomized controlled feasibility and efficacy trial. Psychooncology 32:1433, 2023.

Zunhammer M et al: Meta-analysis of neural systems underlying placebo analgesia from individual participant fMRIData. Nat Commun 12:1391, 2021.

497

The Role of Epigenetics in Disease and Treatment

Brian C. Capell, Shelley L. Berger

The term *epigenetics* was coined by Conrad Waddington in 1942, as he sought to explain how changes in phenotype could occur throughout development independent of any changes to genotype. Appending the prefix *epi-* (Greek, meaning "over, outside of, around") to genetics aptly describes the numerous mechanisms by which gene expression and phenotypes are influenced—and sometimes even inherited through cell division—independent of any changes to the underlying DNA sequence. Today, epigenetics occupies one of the most exciting topics in biology and medicine, offering profound opportunities for discovery, as well as promise for the development of new therapies for disease. Interdisciplinary by nature, the field crosses virtually all areas of science and medicine: chemistry and genetics, development and differentiation, immunology, cancer, aging, and neuroscience.

The continuous introduction of ever more powerful technologies for interrogating the epigenome has led epigenetics to become one of the most innovative fields within the biomedical sciences. Given the vast expanse of the topic and limitations of space, in this chapter, we provide a broad but brief overview of the field and then highlight key areas across the landscape of biomedicine where epigenetics has been revealed to play critical roles in physiology and disease, and importantly, where epigenetics-based therapies have demonstrated success in clinical medicine.

THE BIOCHEMICAL BASES OF EPIGENETICS

Fundamental to epigenetic regulation is the intricate organization into chromatin of each cell's genome **(Chap. 479)**. The fundamental unit of the packaging into chromatin is the nucleosome, consisting of 147 base pairs of DNA wrapped around an octamer of 8 histone proteins (two copies of each of the four core histone proteins: H2A, H2B, H3, and H4), and nucleosome assembly into a regular repeating spaced array along the DNA polymer. The presence of nucleosomes and level of compaction of this basic chromatin array determine the accessibility of the DNA strand to transcription factors, to DNA repair machinery, and to other DNA-binding entities. Thus, compaction has a profound influence on gene expression levels and on local DNA mutation rates. Open regions of chromatin (euchromatin) tend to be transcriptionally active, whereas compacted chromatin (heterochromatin) tends to be transcriptionally repressed. Higher order three-dimensional chromatin architecture such as folding and looping further contribute to epigenetic gene regulation and cellular phenotypes.

Histones include the four core histones, which are the most abundant and most frequently found throughout the genome, and the variant histones of H2A, H2B, and H3. The individual protein structures of both core and variant histones include amino- and carboxyl-terminal "tails," which are extended and unstructured, and highly conserved globular domains. The x-ray crystal structure of the nucleosome particle has illuminated the dynamic alterations of chromatin by an astonishing range of regulatory mechanisms, summarized below.

The three main processes that regulate chromatin compaction, and thus access to the DNA template, include direct methylation modifications (and oxidized derivatives of methylation) of the DNA strand itself, posttranslational modifications of histones, and remodeling of nucleosomes to alter their location and composition with variant histones **(Fig. 497-1)**. The major modification of DNA is cytosine methylation of CpG dinucleotides (5-mC), associated with gene repression and catalyzed by the DNMT1, DNMT3A, and DNMT3B enzymes. DNMT3A and 3B catalyze the addition of methyl groups on unmethylated DNA de novo at CpG dinucleotides that are typically located throughout transcribed genes and in intergenic regions, but lacking at promoters, while DNMT1 is critical for the maintenance of the methylation state after DNA replication and after transcription during the

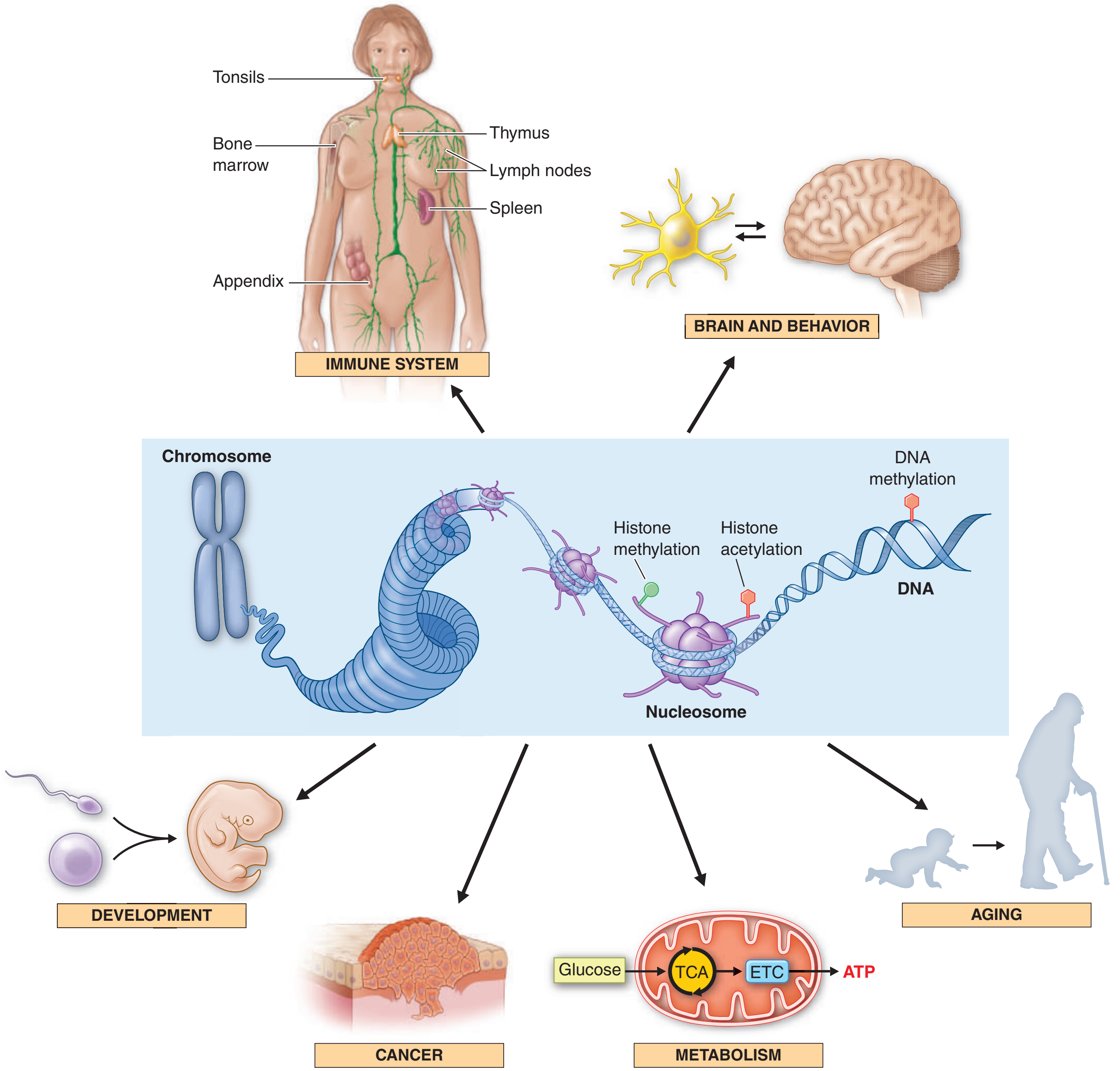

FIGURE 497-1 Epigenetic pathways influence multiple physiologic and disease pathways. As depicted in the center of the illustration, epigenetics refers to the chemical modifications of DNA and histones, which influence chromatin structure, gene expression, and susceptibility to mutations. These molecular pathways, in turn, play important roles in development, cancer, metabolism, aging, neural function, and behavior, and in the immune system. ETC, electron transport chain; TCA, tricarboxylic acid.

S phase of the cell cycle. To further alter and to remove methylation, the TET enzymes (TET1–3) progressively oxidize 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC), to 5-formylcytosine (5-fC), and to 5-carboxylcytosine (5-caC), which are unable to be recognized by DNMT1 but can be removed by additional enzymes. Hence, these are mechanisms to passively lose 5-mC following DNA replication or to actively remove 5-mC, both potentially returning to unmethylated cytosine.

Histone posttranslational modifications (hPTMs) are rich sources of diverse signaling to, and marking of, the chromatin template, including at least 60 different covalent chemical modifications on the histone N- and C-terminal tails and within the globular domains. The hPTMs are added (written) and removed (erased) by enzymes and also serve as sequence- and PTM-specific binding surfaces for effector proteins and complexes (readers) to carry out a wide range of downstream actions including transcription, replication, DNA repair, and recombination. One key point is that the staggering numbers of writers, erasers, and readers provide unlimited potential for diagnostic and therapeutic pharmacologic discovery.

Throughout this chapter, we focus on histone acetylation and methylation, the most abundant and the most well-studied hPTMs (Fig. 497-1), although a wealth of additional modifications, such as serine/threonine/tyrosine phosphorylation, lysine ubiquitination, lysine SUMOylation, and lysine ADP-ribosylation, among others, play important roles in transcriptional and chromatin regulation. For instance, histone phosphorylation targets histone H2A at Ser139 (γH2A.X), which marks DNA double-strand breaks immediately following DNA damage and is critical for the recruitment of the DNA repair machinery. Histone mono-ubiquitination functions similarly to other hPTMs, in signaling and marking the chromatin template, in particular serving to mark the initiation region or elongation of transcribed genes for future rounds of transcription, whereas histone

SUMOylation plays a role in transcriptional repression. Polyubiquitination serves to tag proteins for degradation by the proteasome, and dysfunction in this system may play a role in the pathogenesis of neurodegenerative diseases, including Alzheimer's, Parkinson's, and Huntington's. ADP-ribosylation involves a class of enzymes, the poly-ADP-ribose polymerases (PARPs), which transfer ADP-ribose units from NAD^+ to a variety of nuclear proteins. This PARylation alters the chromatin environment through the recruitment and modification of chromatin-associated proteins. In general, future studies of the profuse types and functions of hPTMs will enhance our understanding of these chromatin-based mechanisms and processes and will illuminate new opportunities and targets for therapies.

In contrast, there is extensive understanding of histone lysine acetyltransferases (KATs) and methyltransferases (KMTs). KATs, previously known as HATs, were among the first identified histone modification enzymes. They attach acetyl groups on the lysine residues of histone tails and other proteins, resulting in both a novel side chain (acetyl-lysine) and an increase in negative charge (from positive charged lysine to neutral acetyl-lysine). This alteration results in loosening of chromatin structure to become more permissive to the binding of transcription factors, and acetylation also creates a novel binding surface for the association of reader proteins. Acetylation on core histones, such as lysine 9 on histone H3 (H3K9ac) or lysine 27 (H3K27ac), is typically associated with transcriptional activation. Acetylation is very dynamic and can be rapidly removed by histone deacetylases (HDACs), of which there are multiple classes, including HDACs and sirtuins (NAD-dependent deacetylases), acting to return the lysine to unmodified ground state.

Methylation of histone tails by KMTs provides more nuanced regulation, in that particular methylated lysines are associated with transcriptional activation (e.g., H3K4me3, H3K36me3, H3K79me3), transcriptional repression (e.g., H3K27me3), or DNA repeat and centromeric silencing (e.g., H3K9me3). The output is strictly determined by effector protein binding, as methylation of lysine does not alter side chain electrostatic charge. Lysine methylation is also a more stable chemical modification than is acetylation and turns over more slowly. Lysine demethylases have been identified for several of the specific methylated sites (H3K4, H3K9, H3K36, H3K27, H3K79).

In addition to their impacts upon local chromatin structure through electrostatic alterations and through recruitment of reader effector proteins, some histone modifications can influence other epigenetic processes. For example, H3K36me3 is involved in a variety of transcriptional processes including elongation and splicing. However, through its recruitment and interaction with other methyltransferases, such as DNMT3B and METTL14, H3K36me3 impacts both DNA and RNA methylation, respectively.

Frequently coordinating with histone modification enzymes are nucleosome remodeling enzymes, which use the energy derived from the hydrolysis of ATP to reposition and remove nucleosomes along the DNA template and to exchange core histones and variant histones (including variants that are located at the transcriptional initiation sites [H2AZ] and over the transcribed genes [H3.3]). The nucleosome remodeling complexes can activate or repress transcription. The SWI/SNF family creates nucleosome-free regions for transcriptional activation, the ISWI family evenly spaces nucleosomes to repress transcription, and the INO80 family exchanges H2A with H2AZ at transcription start sites to poise transcriptional activation. Other remodeling complexes play key roles in the DNA damage response and apoptosis, among additional genomic processes.

As alluded to above, RNA can also be methylated, and "RNA epigenetics" is now an emerging area of gene regulation beyond the direct methylation of DNA and hPTMs. Methylation of RNA, such as messenger RNAs (mRNAs), has been known to exist for over half a century. However, in the last decade, the discovery of enzymes that perform reversible methylation of RNAs led to an explosion of this new field, called *epitranscriptomics*. Indeed, RNA methylation leads to mRNA degradation or facilitates translation. However, mRNA methylation itself occurs co-transcriptionally. Notably, the writer methyltransferase enzymes (METTL3, METTL14) and the demethylases (ALKBH5, FTO) have important roles in a variety of disease pathologies, and drugs targeting their clinical activities are currently in human clinical trials.

Because multiple enzymes redundantly and synergistically write, erase, and recognize these modifications on DNA, RNA, and histones, there is great complexity and the potential for fine-tuning of gene regulation. While extensive knowledge gaps remain to fully explicate these mechanisms of gene regulation, epigenetics has become a fully established discipline within biomedical research. In the coming years, it is likely that the basic understanding of these processes will be further harnessed for further betterment of human health.

TOOLS FOR THE STUDY OF EPIGENETICS

Central to the rapid pace of epigenetic discovery has been the continual development of new cutting-edge epigenetic technologies. Chromatin immunoprecipitation (ChIP), developed over three decades ago, has been a mainstay across epigenetics and molecular biology research more broadly **(Fig. 497-2)**. ChIP involves using formaldehyde to crosslink proteins to DNA and then fragmenting the DNA and reversing of the crosslinks in order to analyze the DNA. The linking of ChIP to next-generation sequencing (ChIP-seq) provided a major leap forward in allowing researchers to probe the entire genome-wide landscape of histone modifications and DNA-binding transcription factors and chromatin-modifying enzymes. This has led to fundamental discoveries regarding the role of the epigenome in the regulation of gene expression and cellular phenotypes in development and disease.

More recent refinements in these methods have expanded the applicability of these methods. Studying chromatin accessibility has become possible through the assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Fig. 497-2). Through ATAC-seq, a Tn5 transposase can be utilized to insert sequencing adapters into open regions of chromatin, allowing for the identification of DNA regulatory elements such as promoters and enhancers even at the single-cell level. Building upon both ChIP-seq and ATAC-seq, the tethering of an antibody of interest to micrococcal nuclease (MNase) allows for the cleavage of the DNA on either side of the target, sidestepping the need for any formaldehyde fixation step, significantly scaling down the signal-to-noise ratio, and reducing the number of cells and DNA required in comparison to standard ChIP-seq. This method, referred to as CUT&RUN (cleavage under targets and release using nuclease), offers the ability to obtain histone and chromatin-binding information in systems and models where cell numbers were previously rate limiting. A further modification of the CUT&RUN protocol replaces the MNase with a Tn5 transposase fused to sequencing adapters (CUT&Tag), offering the ability to profile histone modifications at the single-cell level.

While ATAC- and ChIP-seq and their derivatives have provided tremendous insights into how chromatin accessibility and histone modifications play a role in gene regulation, they did not provide information on how the physical organization and folding of the genome might contribute to gene expression. This was only able to begin to be understood by techniques that could elucidate the three-dimensional architecture and structure of chromatin. Here, techniques such as HiC (Hi-C is a high-throughput form of 3C (chromosome conformation capture) technology to study 3D genome organization) and Hi-ChIP (Hi-ChIP combines Hi-C with ChIP-sequencing to study the relationship of DNA-binding proteins to 3D genome organization) have emerged to reveal nuclear architecture and how it can either inhibit or facilitate gene expression. Collectively, these studies have revealed a model whereby enhancer state drives gene regulation. Once enhancer-promoter loops have formed, these topologically associated domains (TADs) are reinforced and can ultimately help to constrain motion of the genome and in turn increase the likelihood of further promoter-enhancer connections forming to facilitate transcription.

More recently, the latest frontier in biomedical research is spatial technologies that allow the capture of molecular data at subcellular resolution within their native tissue context. While techniques such as spatial CUT&Tag are still in development, as they continue to advance in resolution, throughput, and accessibility, they are certain to offer unprecedented insights into how tissue and disease pathology correlates with alterations in the transcriptome, epigenome, and proteome.

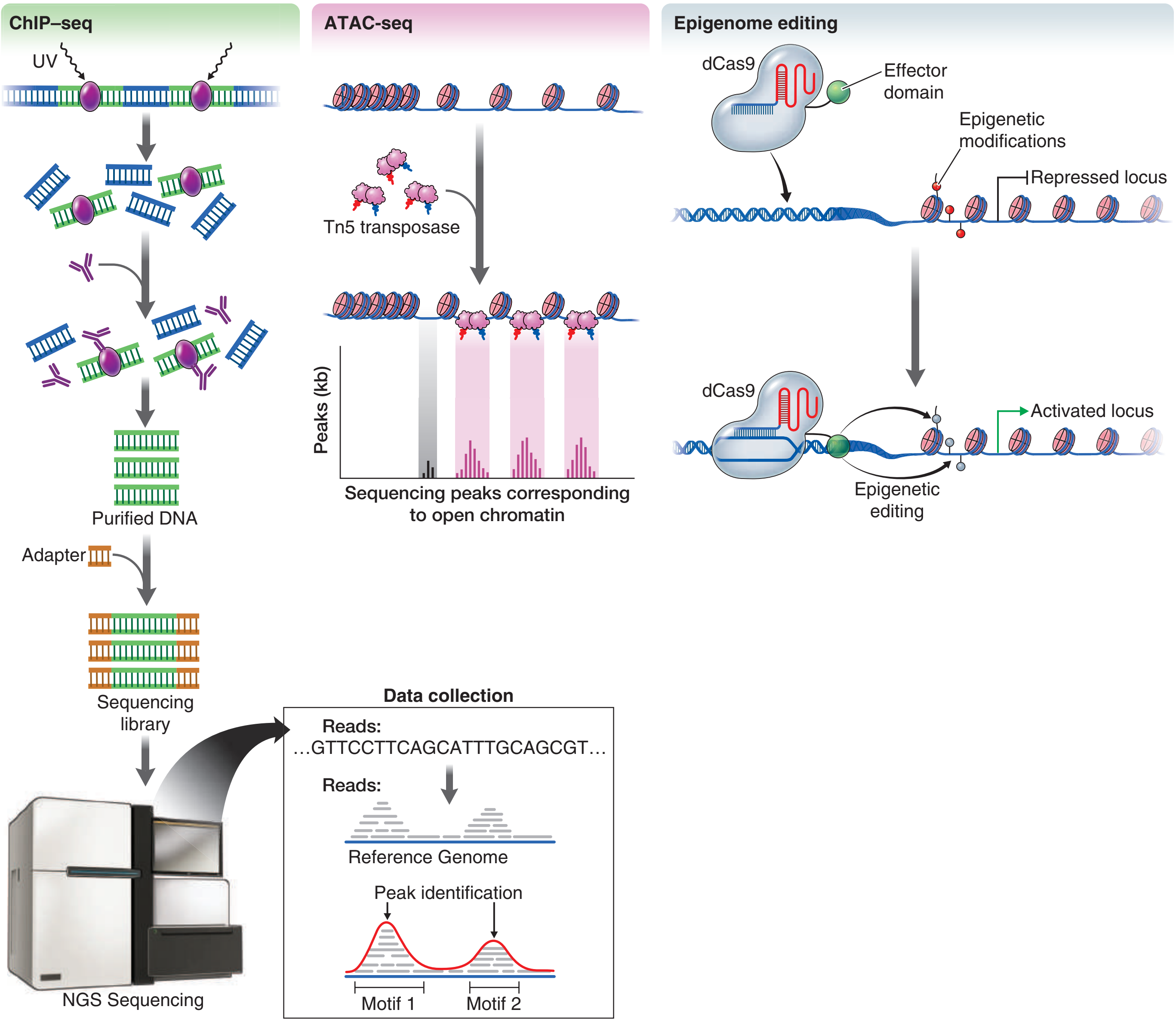
ChIP–seq
UV
Purified DNA
Adapter
Sequencing library
NGS Sequencing
ATAC-seq
Tn5 transposase
Peaks (kb)
Sequencing peaks corresponding to open chromatin
Epigenome editing
dCas9
Effector domain
Epigenetic modifications
Repressed locus
dCas9
Activated locus
Epigenetic editing
Data collection
Reads:
…GTTCCTTCAGCATTTGCAGCGT…
Reads:
Reference Genome
Peak identification
Motif 1
Motif 2

FIGURE 497-2 Core experimental techniques for the study of epigenetics. The explosion of interest and research in the past few decades has been fueled by fundamental advances in the experimental approaches and ability to profile the epigenome. Chromatin immunoprecipitation–next-generation sequencing (ChIP-seq) allows for the ability determine the genome-wide binding of a histone modification or DNA-binding protein of interest. In contrast, assay for transposase-accessible chromatin using sequencing (ATAC-seq) provides a method for determining chromatin accessibility genome-wide even down to the single-cell level. More recently, the development of CRISPR-based epigenome-editing technologies has offered a way to directly deposit histone modification in order to activate or repress specific genes. NGS, next-generation sequencing.

Finally, another major technological breakthrough, and one with tremendous therapeutic potential, is development of CRISPR-based epigenome editing approaches (Fig. 497-2). By fusing a nuclease-deactivated Cas protein to an epigenetic modifying enzyme, one can use guide RNAs to precisely target gene regulatory effectors to turn on and turn off specific genes by changing the gene's histone acetylation or methylation levels. For example, using CRISPR to guide the mRNA of an epigenetic repressor to the oncogene MYC is one strategy currently being tested for cancer treatment. These advances promise to not only elucidate new knowledge regarding principles of gene regulation but also to offer new therapeutic opportunities for disease.

EPIGENETICS IN DEVELOPMENT AND DIFFERENTIATION

Epigenetic processes are critical to organismal development and to cellular differentiation and reprogramming of cell fate (Fig. 497-1). Transcription factors establish the epigenomic landscape that enables and stabilizes cell-type-specific gene expression while simultaneously ensuring stable repression of alternative cell fates. This results in chromatin profiles that display remarkable cell-type specificity in differentiated cells, particularly at the key regulatory nodes of gene enhancers, which are gene-distal DNA elements that control transcription. In fact, epigenome profiling of the chromatin landscape in tumors of unknown cell origin can provide a better index of the origin tissue than does DNA sequencing of gene mutations within the tumor.

The cell-type-specific epigenetic program is first derived from the template of embryonic stem cells, where numerous genes required for differentiation exist in a "bivalent" state, marked by both the activating histone modification, H3K4me3, and the repressive modification, H3K27me3. Due to this unstable epigenetic state, the genes are "poised" for activation or for repression, depending on their subsequent cell fate. Critical genes directing toward a specific cell fate will be turned on, with maintained H3K4me3 and erased H3K27me3, whereas genes leading toward alternative fates will be repressed, with maintained H3K27me3 and removed H3K4me3. Once differentiated, an epigenetic barrier will prevent the cells from returning to the stem cell state. For example, constitutive heterochromatin in the form of H3K9me3 can serve as a barrier to cellular reprogramming when attempting to create induced pluripotent stem cells, and inhibiting the enzymes that catalyze H3K9me3, such as SUV39H1, can enhance reprogramming efficiency.

DNA methylation contributes to the specification of cell fate and to other developmental pathways. DNA methylation alterations are involved in critical processes ranging from sex chromosome dosage compensation to coordinating expression of imprinted genes. Disruption of this latter process can lead to imprinting disorders including Prader-Willi syndrome, Angelman syndrome, and Beckwith-Wiedemann syndrome. Recent discoveries have served to highlight the tremendous amount of interplay between epigenetic modifications and, in particular, between DNA methylation and various histone modifications.

Beyond embryonic development, epigenetics can provide the necessary coordination and balance between adult stem cell self-renewal compared to cell differentiation. This epigenetic control is critical, as impaired self-renewal can lead to stem cell exhaustion and premature aging, while excessive self-renewal may promote cancer. Key epigenetic regulators tend to play conserved roles across diverse tissue types. For instance, BMI1, a component of the polycomb repressive complex 1 (PRC1), is required for stem cell proliferation and self-renewal, and its ablation leads to stem cell depletion in hematopoietic, epidermal, muscle, intestinal, and mammary stem cells. Similarly, the DNA methyltransferase DNMT1 is required for stem cell self-renewal in hematopoietic, epidermal, and mammary stem cells. HDACs 1 and 2 possess some overlapping functions and are required for normal epidermal differentiation. Likewise, a loss of these HDAC enzymes in hematopoietic stem cells can lead to failure of differentiation and severe anemia. In a similar fashion, inhibition or loss of histone lysine demethylase 1 (LSD1), a repressor of transcription, is known to promote differentiation across multiple cellular contexts. These factors represent repressive chromatin regulation, leading to the general concept that restraining specific transcription pathways related to differentiation is crucial to maintaining undifferentiated self-renewing stem cell pools.

The epigenetic regulation of the tumor suppressor p16 (*CDKN2A*) locus during differentiation provides a prime example of this finely tuned system. For example, as mentioned above, DNMT1 is necessary for self-renewal in human epidermal stem cells. Levels of DNMT1 are high in the basal undifferentiated layer of the epidermis, decreasing progressively with epidermal stratification, leading to de-repression of the tumor suppressors p16 and p15, thereby promoting cell cycle arrest and full differentiation. BMI1 displays a similar phenotype in both hematopoietic and epidermal stem cells, repressing key genes that promote differentiation, such as *p16* and *p19ARF*. Consistently, a loss of BMI1 leads to premature differentiation and defective self-renewal. In addition to the repression provided by DNMT1 and BMI1, the p16 locus is highly decorated with the repressive H3K27me3 catalyzed by EZH2 in epidermal stem cells. Then, during epidermal differentiation, H3K27me3 is removed by the KDM6B (JMJD3) histone demethylase. Loss of this control over programmed p16 expression occurs in epithelial cancers, such as squamous cell carcinoma (SCC), where EZH2 is overexpressed and KDM6B expression is lost. Breast cancer is another example where progesterone can increase levels of EZH2 to promote mammary epithelial cell proliferation, and excessive EZH2 expression can occur in cancer. This exemplifies how epigenetics can integrate environmental signals and have a profound influence on the fine balance between stem cell maintenance and overt carcinogenesis. In general, a recurrent theme in cancer is loss of key chromatin regulation that promotes cell differentiation, combined with gain of activities that promote stemness.

Chromatin-modifying enzymes also play a major role in influencing cell-type specificity. High levels of EZH2 that modify H3K27me3 promote adipogenesis while simultaneously inhibiting osteogenesis. In contrast, the H3K27me3 demethylases, KDM6A (UTX) and KDM6B, derepress those same genes, driving stem cells toward osteogenesis. Through interactions with tissue-specific master regulators, epigenetic modifiers also shape cell-type specificity. In the epidermis, p63, the p53 family member that is a master regulator of the epidermal compartment, interacts with several chromatin regulators including HDAC1 and HDAC2, SATB1, MLL4 (KMT2D), and BRG1 to orchestrate epidermal differentiation. Similarly, the gene-activating H3K4 histone methyltransferases, MLL3 (KMT2C) and MLL4, are required for adipogenesis by forming a complex with the transcriptional activator ASC2 and the transcription factor PPARγ to induce adipogenic genes. Overall, loss of epigenetic regulation can reduce cell differentiation and increase stem cell specification to drive diseases encompassing development, cancer, and, broadly, diseases associated with aging.

EPIGENETICS OF METABOLISM

One of the fascinating aspects of epigenetics is that it represents a mechanism for direct connection between the environment and gene expression. Numerous studies in the field of metabolism have identified a complex interplay between diet, metabolism, and the epigenome (Fig. 497-1). Seminal findings in *Drosophila* and mice have shown that changes in diet, particularly the paternal diet, and other environmental factors, can influence the metabolism of offspring, ultimately promoting obesity in later generations. Epidemiologic studies in humans have supported these results, as the nutritional status of grandparents has been correlated with phenotypic effects in grandchildren. In fact, diet can directly affect the levels and activity of chromatin modifiers.

For instance, high-fat diets reduce histone acetylation through their ability to inhibit the enzymes ACLY and ACSS2, which produce acetyl-CoA. Levels of acetyl-CoA, in comparison to all measured metabolites, are indeed the best predictor of histone acetylation levels. Consistent with this, increased acetyl-CoA correlates with rising levels of total histone acetylation, including at the promoters of growth-associated genes. This increase in nuclear acetylation is associated with cell cycle progression and proliferation, and it can have clinically relevant downstream effects. For example, high levels of acetyl-CoA can delay stem cell differentiation and suppress autophagy. The oncogenes MYC and AKT can both hijack metabolic networks to enhance nutrient uptake by cancer cells, thus promoting acetyl-CoA production and resulting in both the initiation and progression of tumorigenesis. Additional evidence suggests that dietary intake of alcohol can directly contribute to acetate levels and therefore histone acetylation in the brain, with effects on the transcription of genes involved in learning and memory.

Contrary to convention that metabolic enzymes are strictly mitochondrial or cytosolic, certain metabolic enzymes can be present in the nucleus and can thereby directly regulate histone acetylation enzymes. This is the case for several enzymes that generate acetyl-CoA, including ACLY, PDH, and ACSS2, which generate acetyl-CoA from citrate, pyruvate, and acetate, respectively. Further, ACSS2 can be chromatin-bound to regulate gene expression, leading to physiologic responses such as autophagy in the liver and mammalian hippocampal function in learning. This direct metabolic-epigenetic enzyme cross-talk illuminates a crucial local role of the acetyl-CoA metabolite to effect rapid gene transcription and represents a fertile intersection for future therapeutics.

Methylation is also altered by metabolism. *S*-Adenosylmethionine (SAM) is the key metabolic cofactor for histone and DNA methylation. Dietary factors are estimated to explain 30% of the variation in human serum methionine concentration and hence can alter SAM levels and histone methylation. For example, dietary methionine availability and intracellular production of SAM affect the levels of histone H3K4me3 associated with transcriptional activation. Furthermore, these fluctuations can have critical physiologic consequences: DNA methylation levels in rectal mucosa and colonic polyps are increased by higher levels of dietary folate, and a diet low in methyl donors reduces the formation of gastrointestinal cancers in mice predisposed to these tumors. Methionine metabolism and the availability of SAM regulate stem cell differentiation and contribute to carcinogenesis. For instance, cancers with mutations in metabolic regulatory genes such as *IDH1/2*, *FH*, and *SDH* lead to the accumulation of by-products (2-hydroxyglutarate, fumarate, and succinate, respectively), which all inhibit α-ketoglutarate (α-KG)–dependent histone demethylases and thus promote hypermethylation and lead to impaired cellular differentiation. Notably, some of the α-KG–dependent demethylases, which are highly mutated in numerous cancers (i.e., KDM5A, KDM6A), also serve as cellular oxygen sensors, thus linking environmental oxygen levels to epigenetic control of methylation levels. In contrast to hypermethylated states, loss of the *MTAP* gene, which is part of the 9p21 locus containing p16 and one of the most frequent events in human cancer, disrupts normal methionine metabolism. This both lowers methylation levels, and, interestingly, also sensitizes cancer cells to inhibitors of the PRMT5 methyltransferase, therefore opening a therapeutic opportunity. These observations illustrate how connections between epigenetics and metabolism can generate unanticipated advances in medicine. Furthermore, these data highlight the tight interconnections between environmental inputs, metabolism, and epigenetics.

CANCER EPIGENETICS

Cancer is now understood to be a mixed genetic and epigenetic disease, as epigenetic dysregulation is pervasive in human cancers (Fig. 497-1). Beyond simple activation of oncogenes or reduced expression of tumor suppressors, epigenetic mechanisms can contribute to chemotherapy resistance and to failure of antitumor immunity. Accordingly, the development of drugs targeting epigenetic pathways is one of the most active areas of clinical and pharmaceutical development, with several compounds already approved for human use and shown to be effective in a variety of cancers. Epigenetic perturbations in cancer largely affect chromatin-regulating enzymes, which represent robust targets for development of novel small-molecule inhibitors, especially as compared with canonical oncogenic transcription factors (e.g., MYC) and tumor suppressors (e.g., p53).

Epigenetics can contribute to carcinogenesis in a variety of ways. First, on a global scale, chromatin organization is the single most influential factor in determining local mutation rate across the genome. Analysis of abundant tumor sequencing data has demonstrated that heterochromatic regions of the genome contain a higher frequency of mutations compared with more open euchromatic regions. This difference is due to the improved accessibility of the DNA repair machinery to less compact, more open regions of chromatin.

The first discovery of an epigenetic mutation was found in 1998 when the chromatin remodeler *SMARCB1* was shown to drive the formation of malignant rhabdoid tumors. Extensive sequencing of human tumors from the majority of cancer types has been performed by The Cancer Genome Atlas (TCGA) consortium, and remarkably, 25–30% of identified cancer driver mutations occur in chromatin regulatory proteins. Similar to *SMARCB1*, numerous other chromatin modifiers (e.g., methyltransferases *MLL3* and *MLL4*, and acetyltransferases *EP300* and *CBP*) and nucleosome remodeling enzymes and associated complex components (e.g., *SMARCA4*, *SMARCA2*, *ARID1A*) are heavily mutated and inactivated in many cancers. The majority of these mutations are loss-of-function mutations, and indeed, enzymes like MLL4 and demethylase KDM6A possess tumor-suppressive activity across a variety of tissues and cellular contexts. In contrast, the H3K27me3 histone methyltransferase *EZH2* is an oncogene, and accordingly, it is overexpressed in many advanced-stage or metastatic solid tumors such as breast cancer, prostate cancer, and melanoma. Mechanistically, EZH2 represses the *p16* tumor suppressor and other cell cycle genes required for cell cycle exit via H3K27me3 deposition. Consistent with a broad growth regulatory role, EZH2 inhibitors are therapeutically successful for a number of cancers in preclinical models and are being actively studied for B-cell lymphoma, melanoma, and other solid tumors.

In addition, provocative evidence has emerged for a direct tumorigenic role of histones based on the discovery of causative mutations, such as histone H3 mutations identified in pediatric high-grade gliomas. Specifically, the majority of these mutations are in the H3 variant H3.3, where lysine 27 is replaced by methionine (K27M). Similarly, >90% of chondroblastomas replace lysine 36 with methionine (K36M) in histone H3.3. These effects appear to be dominant negative because (1) in H3.3, these are heterozygous mutations, and (2) the mutations also occur in the canonical H3, which exists in ~30 orthologous genes in the human genome. Thus, a minority of H3/H3.3 mutant protein leads to global defects in the associated histone modifications (K27 or K36 methylation), possibly via irreversible inhibition of the cognate enzymes by the mutant histones. These "oncohistone" mutations promote resistance to apoptosis and failure of normal differentiation in a number of pediatric and adult cancers.

Beyond mutations, genetic translocations involving chromatin modifiers also implicate chromatin pathways as direct drivers in cancer. MLL1 (KMT2A), the H3K4 histone methyltransferase, is a frequent translocation partner occurring in adult and pediatric acute myeloid leukemia (AML) and in ~80% of infant acute lymphoid leukemia (ALL) cases. MLL1 can fuse with >70 translocation partners, and these mutant proteins prevent normal hematopoietic differentiation. Consistent with a causative role of MLL1 in these gene fusions, drugs inhibiting the catalytic activity of MLL1 are effective in preclinical models of AML and are currently being evaluated in human clinical trials.

Given the abundance of epigenetic abnormalities in cancer combined with the inherent reversibility of epigenetic changes, extensive efforts are underway to develop epigenetic drugs. The first epigenetic therapeutic involved the use of DNA methylation inhibitors (DNMTi) to reactivate tumor-suppressor genes. Interestingly, the mechanism of traditional chemotherapeutics, such as azacitidine and decitabine, is to inhibit DNMT1, thereby promoting global hypomethylation; these are currently in clinical use for myelodysplastic syndrome (MDS) and AML. In a second broad mechanism, loss of acetylation occurs in many cancers, and thus, HDAC inhibitors (HDACi) are under intensive development. HDACi are effective and approved for treatment in cutaneous T-cell lymphoma and multiple myeloma. Bromodomain (BRD)-containing proteins bind to lysine acetylated target proteins, including histones, and rationally designed BET inhibitors (BETi) block their binding. BETi reduce the amplified expression of oncogenes such as *MYC* in hematologic cancers. Current studies are now focused on optimizing combinatorial epigenetic therapies with conventional chemotherapies and immunotherapies, particularly given the ability of epigenetic therapeutics to promote re-expression of tumor antigens and interferon (IFN)-mediated antitumor immunity. Indeed, the development of a new generation of more specific epigenome-targeted inhibitors, combined with our increased knowledge of the underlying epigenetic mechanisms contributing to tumorigenesis, has enabled a precision medicine–based approach to harnessing the potential of these drugs. This may be particularly valuable in the context of improving patient responses to a variety of therapies beyond chemotherapies and immunotherapies, such as radiation and hormone therapies.

There are several hundred chromatin enzymes and binding proteins in the human genome, and the current focus is to identify more specific inhibitors. Indeed, targeted inhibitors of numerous mutated chromatin regulators have been developed, with >30 compounds currently in various stages of development and preclinical trials. Some notable examples showing early clinical success include EZH2 inhibitors for lymphomas, sarcomas, and melanoma; IDH inhibitors for AML and gliomas carrying mutant *IDH1* or *IDH2* genes; LSD1 inhibitors for AML and small-cell lung cancer; and DOT1L and MLL1 inhibitors for leukemias with activated MLL1. Given the broad potential effects of epigenetic regulators, it is perhaps not surprising that there have been some dose-limiting toxicities, particularly among those that are less target-specific. Collectively, the emerging picture is that the most effective and robust use of epigenetic drugs in cancer will be fine-tuning and potentiating the effects of other therapies that are either incompletely effective or marked by widespread resistance.

EPIGENETICS OF AGING

Like many diseases of aging, human aging itself results from the complex interplay between genes and the environment. Evidence that the epigenome may be the key link between these processes derives from observations that numerous environmental stimuli and stressors—ranging from diet and exercise to hormones and circadian rhythms—contribute to both aging and epigenetic alterations (Fig. 497-1). Thus, a lifetime of exposures progressively disrupts the chromatin landscape. These age-dependent changes in chromatin organization increase the susceptibility of the genome to mutations and also reduce transcriptional fidelity. Further, provocative findings in model systems demonstrate that stress-induced epigenetic changes can be transmitted over several generations and can even affect the life span of later generations. Among these global epigenetic alterations, there is dysregulation of histone modifications and a general loss of histone proteins with aging across taxa. Amazingly, experimental increases in histone levels, particularly histones H3 and H4, but not H2A or H2B, can reverse these age-related changes in mammalian cells and in the yeast *Saccharomyces cerevisiae* model.

Thus, the sum of current evidence suggests a model of aging via a general increase in activating epigenetic modifications along with a loss of repressive modifications. Together these changes create a state of transcriptional instability and "noise" that inhibits accurate transcription. Cells from patients with Hutchinson-Gilford progeria syndrome (HGPS), the most severe form of human premature aging, display reduced levels of both H3K9me3 and H3K27me3 repressive chromatin. In another premature aging disease, Werner syndrome, DNA damage induces global loss of H3K9me3 and H3K27me3 due to the inherent absence of the Werner syndrome ATP-dependent DNA helicase, which is critical for DNA repair. Such heterochromatin loss is not limited to premature aging conditions, as aged cells derived from healthy older humans display age-dependent loss of H3K9me3 leading to aberrant expression of normally repressed transposable elements. Activation of these mobile elements correlates with neurodegenerative disorders and may also promote other aging-related phenotypes such as cancer. Human fibroblasts undergoing cellular senescence (exit from cell cycle due to replicative or other stress) undergo destabilization of compact heterochromatin adjacent to the nuclear periphery, in so-called lamin-associated domains (LADs). At LADs, in addition to a reduction of repressive histone modifications as discussed above, there are broad new regions of the euchromatic histone modification H3K4me3. This general loss of heterochromatin can promote the activation of cytosolic DNA and RNA sensing pathways that promote innate immune signaling and "inflammaging."

In addition to age-associated alterations of histone modifications, direct manipulation of chromatin-modifying enzymes that control these marks affects the balance between heterochromatic and euchromatic regions, and it alters the lifespan of model organisms. Inhibiting the H3K27me3 histone demethylase KDM6A results in increased repressive H3K27me3 and extended lifespan in *Caenorhabditis elegans*. Consistent with this, genetic reduction of enzymes (*ash-2*, *set-2*, *wdr-5*) that add the activating H3K4me3 histone modification also extends lifespan in *C. elegans*. The consequences of these genetic manipulations nicely correspond to the observed changes in histone modifications as described above. Beyond histone-modifying enzymes, dysregulation of the levels or function of chromatin remodelers can also affect lifespan in model organisms. This dysregulation occurs in humans as well, as in the nucleosome remodeling deacetylase complex (NuRD), which is reduced in HGPS fibroblasts and in aged healthy donors.

In addition to age-related changes in histone methylation, histone acetylation also contributes to aging phenotypes. Dysregulation of histone acetyltransferases (HATs) and HDACs is associated with reduced longevity across model organisms. Further, sirtuin deacetylases (class III NAD$^+$-dependent HDACs) promote health span and lifespan across species as key mediators of pro-longevity effects of caloric restriction. Indeed, loss of Sirt6 results in premature aging in mice, while caloric restriction–induced increases of Sirt1 and Sirt6 expression can delay aging. As discussed previously, metabolism and acetylation are intricately linked, and the sirtuins, via NAD$^+$ levels, and other HDACs may play key roles connecting the environment, gene expression, and physiologic output. For instance, exercise in humans reduces activity of HDACs 4 and 5, leading to increased H3K36ac in skeletal muscle, which likely promotes beneficial gene expression.

Epigenetic alterations with aging are not limited to histone modifications and extend to DNA methylation. Consistent with the histone patterns, DNA methylation data support the model described above—that is, general decompaction of the epigenome with aging. Specifically, levels of 5-mC are reduced in senescent human cells, and global DNA hypomethylation occurs across the human genome with aging. Concurrent with this overall hypomethylated state, there are local regions of hypermethylation focused near CpGs at gene promoters, particularly at genes that maintain cellular differentiation and cell identity. This epigenetic disruption during aging thus leads to profound changes in transcription. For example, in hematopoietic stem cells, DNA hypermethylation blocks proper binding of transcription factors, resulting in dysregulation of normal gene expression with aging. Importantly, these patterns are not merely stochastic alterations in response to environmental stressors throughout aging. Indeed, the methylation status of a defined number of CpG sites is a highly accurate predictor of chronologic age in human tissues. This work reveals that DNA methylation status with aging outperforms previous standard biomarkers of aging, such as p16 expression levels and telomere length, and will be highly valuable in the near future to gauge effects of treatment aiming to ameliorate diseases of aging.

■ EPIGENETICS OF THE BRAIN AND BEHAVIOR

Brain disorders are among the greatest clinical challenges to understand and to treat. Most neurologic and psychiatric disorders result from complex dysregulation of numerous genes and pathways. In this interplay between underlying genetic predisposition and external environmental factors, aberrant epigenetic regulation is increasingly recognized as a potentially key modulator (Fig. 497-1).

More directly, however, several progressive neurodevelopmental disorders are caused by germline mutations in chromatin regulators. Mutations in methyl CpG binding protein 2 (*MECP2*), a protein important for binding to methylated DNA and contributing to gene repression, are the major cause of Rett syndrome. MeCP2 loss leads to overactive gene transcription in neurons and impaired presynaptic excitatory functions. Similarly, Kabuki syndrome, another progressive neurodevelopmental disorder, is caused by germline mutations in either the H3K4me1 histone methyltransferase, *MLL4* (*KMT2D*), or the H3K27me3 demethylase, *UTX* (*KDM6A*). This disorder may derive from dysregulation of transcriptional enhancers, a major class of gene regulatory elements, as both *MLL4* and *UTX* play a key role in activation of enhancers. Finally, the acetyltransferase CBP (*CREBBP*) also is important for gene enhancer function and, when mutated, can lead to Rubinstein-Taybi syndrome, a cause of intellectual disability.

Beyond germline mutations, altered methylation dynamics can drive disorders of neural development and of neurodegeneration. Fragile X syndrome, characterized by learning disabilities and cognitive impairment, is caused by mutations in the *FMR1* or *FMR2* gene or by hypermethylation of the transcriptional promoters regulating *FMR1* or *FMR2*. Similarly, Prader-Willi syndrome and Angelman syndrome, neurodevelopmental conditions caused by abnormal imprinting of the paternal or maternal chromosomal region (15q11-13), respectively, are frequently caused by aberrant DNA methylation. Further, DNA hypomethylation is implicated in some neurodegenerative conditions. For instance, in Parkinson's disease, several genes involved in pathogenesis are hypomethylated due to DNMT1 depletion, including the α-synuclein gene (*SCNA*). In Alzheimer's disease (AD), DNA hypomethylation occurs at promoters of key pathogenic genes such as amyloid precursor protein (*APP*). Indeed, APP promoter methylation is responsive to environmental factors, including aging, a major risk factor for AD. Likewise, presenilin-1 (*PSEN1*) is implicated in AD and displays altered DNA methylation in response to variations in metabolic stimuli. Recent evidence from human AD brains demonstrated significant enrichment of H3K9 and H3K27 acetylation and provided evidence that this dysregulation of the epigenome promotes gene transcription pathways involved in AD pathogenesis. Studies of Huntington's disease (HD) have demonstrated DNA hypomethylation and decreased histone acetylation, in part due to altered function of the acetyl transferase CBP, leading to transcriptional dysregulation. Together, these observations underscore altered epigenetic regulation as a crucial feature of neurodegeneration.

Additional gene regulatory proteins in the nervous system interact with and are regulated by chromatin modifiers. REST (repressor element 1–silencing transcription factor) is important in neuronal homeostasis through its ability to recruit chromatin regulatory enzymes, such as histone deacetylases and histone methyltransferases, and via its control over gene expression. REST levels increase with aging and serve a protective function in neurons against age-associated stressors and loss of cognitive function associated with AD. Similar to REST, brain-derived neurotrophic factor (BDNF), another important mediator of neural development and homeostasis, is implicated in a variety of neurologic and psychiatric disorders including HD, depression, schizophrenia, bipolar disorder, and autism. Knockdown of BDNF in the dentate gyrus leads to depression-like behavior in mouse models, and BDNF mediates effects of antidepressant therapies. Chromatin pathways, including DNA methylation/MeCP2 and H3K27me3, play a key role in BDNF regulation as observed in brains from patients with schizophrenia.

Finally, addiction medicine is another frontier where epigenetics holds great promise to reveal connections between environmental exposure and phenotypes. Although still in its early stages in terms of mechanistic understanding, emerging evidence demonstrates disruption of epigenetic homeostasis as a consequence of addictive substances ranging from alcohol to cocaine. For example, the acetylation of regulatory elements in the *FOSB* gene by the histone acetyltransferase CBP is associated with behavioral effects of cocaine. Opioid exposure appears to promote a generally more open and permissive state of chromatin marked by increases in histone acetylation and reductions in histone methylation, which may allow for a more hyperresponsive state and reinforce reward-seeking behaviors. Ethanol also induces histone acetylation and a decompacted chromatin structure with direct effects on learning and memory function.

■ EPIGENETIC INFLUENCES ON INFECTION, IMMUNITY, AND INFLAMMATION

Alterations in gene expression patterns are important determinants of immune-mediated disease, and in turn, epigenetics regulates infection, immunity, and inflammation (Fig. 497-1). Treatment with immune-stimulating agents such as lipopolysaccharide (LPS) and tumor necrosis factor α activates expression of numerous inflammatory genes within hours, with precise gene pathways and activation kinetics determined by the cellular epigenetic state. HATs and HDACs are critical components of this response, coordinating with proinflammatory transcription factors, such as AP-1 and NF-κB, to either activate (in the case of HATs) or repress (in the case of HDACs) inflammatory genes. For example, corticosteroids recruit HDAC2 to promoters of NF-κB–stimulated inflammatory genes to prevent activation during asthma treatment.

Type 1 IFN responses are exceptional examples of regulatory complexity governed by epigenetic control. In an unstimulated state, the H3K9 methyltransferases G9a (*EHMT2*) and EHMT1 suppress expression of IFN and IFN-induced genes. Upon induction of IFN-stimulated genes, STAT transcription factors recruit chromatin remodeling complexes, such as BAF (*SMARCA4*), and recruit HATs including p300, CBP, and GCN5 (*KAT2A*). In turn, chromatin remodeling and acetylation recruit chromatin binding proteins including the bromodomain protein, BRD4, which promotes transcriptional elongation and full activation. Beyond the DNA level, METTL3-mediated m^{6}A methylation on mRNAs also is a critical regulator of IFN signaling in a variety of distinct cellular contexts.

Major regulators of adaptive immunity pathways are similarly epigenetically regulated. CD4+ and CD8+ T cells undergo extensive changes in histone modification profiles during differentiation to distinct subsets of effector T cells. For example, genes associated with effector T-cell functions in CD8+ memory T cells (e.g., *PRDM1*, *KLRG1*, *IFNG*) display enrichment of H3K4me3 and low levels of H3K27me3 compared with those genes in naïve T cells. DNA methylation also plays an important regulatory role and may contribute to disease. For example, CD4+ T cells from individuals with rheumatoid arthritis (RA), systemic scleroderma, and latent autoimmune diabetes in adults display hypermethylation of the *FOXP3* gene, which activates regulatory T cells that dampen immune responses. In addition, hypermethylation of the *CTLA4* locus occurs in regulatory T cells from RA patients, impairing their immunosuppressive abilities.

During infection, epigenetic processes can play critical roles in both the immune response and defense against pathogens, as well strategies exploited by microorganisms to co-opt the host cellular machinery to advantage of the pathogen. Respiratory syncytial virus (RSV) infection promotes the expression of the histone demethylase KDM5B, which removes H3K4 methyl groups from antiviral genes such as type 1 IFNs, driving a switch from T helper 1– to T helper 2–type immune responses, thereby contributing to chronic infection. Similarly, influenza upregulates the repressive H3K9me3 methyltransferase SETDB2 to block expression of CXCL1 and a variety of NF-κB target genes involved in attracting neutrophils and host defense, both serving to lengthen the infection and contributing to bacterial superinfection. Regarding the host response to infection, studies have revealed that differences in host tissue-, age-, and sex-biased epigenetic profiles might shape susceptibility and responses to infection. For example, differential DNA methylation at the *ACE2* gene may impact expression levels of this key cellular receptor and ultimately the ability of SARS-CoV-2 to infect hosts, while alterations in antiviral IFN signaling may lead to more severe COVID-19 infection and disease. These findings are all supported by new discoveries demonstrating that epigenetics is a key component for the inflammatory memory that has been observed now across a wide variety of contexts. Numerous perturbations ranging from infections and vaccination to skin wounding and Western diets have now been shown to elicit an epigenetic memory that is maintained and propagated. This epigenetic memory extends beyond just the immune system to the involved tissues.

These findings have suggested a potential for epigenome-modifying drugs for the treatment of inflammatory and immune-related conditions. For example, the DNA methylation inhibitors azacitidine and decitabine have immunosuppressive effects possibly mediated by enhanced expression of *FOXP3*, which generally suppresses immune responses. HDACi upregulate and downregulate immune genes, and they inhibit cytokine production in macrophages from patients with RA. Further, the HDACi vorinostat and panobinostat inhibit primary B-cell responses and antibody production in vitro and in vivo. Given these broad effects, it is not surprising that the HDACi trichostatin A (TSA) has efficacy in various model systems for treatment of RA, systemic lupus erythematosus (SLE), asthma, acute kidney injury, sepsis-induced lung and cardiac damage, and acute pancreatitis. Similarly, BETi also display broad effects in blocking antigen presentation and T- and B-cell activation and thus beneficial protective effects in a variety of inflammatory settings including autoimmunity, sepsis, atherosclerosis, psoriasis, periodontitis, and arthritis. Beyond these "broad-spectrum" epigenetic inhibitors, GSK-J4, which is a specific inhibitor of the H3K27me3 demethylases KDM6A and KDM6B, has anti-inflammatory activity, presumably by preventing loss of H3K27me3 repression over inflammatory genes. Similarly, inhibition of the H3K4me3 histone methyltransferase MLL1 blocks the induction of proinflammatory cytokine gene expression in a variety of contexts.

CONCLUSION

Due to the enormity and complexity of the chromatin and epigenetics fields and their reach into all areas of biology and medicine, it is not possible to cover such a broad scope in a single chapter. Thus, here we provide a concise snapshot highlighting key areas of development in medicine. We hope to have conveyed the tremendous excitement and promise that pervades the discipline. Indeed, given the exponential growth in uncovering the interface between the epigenome and epigenetic therapies with the environment and disease, there is little doubt that the coming years will bring important additions to this field.

■ FURTHER READING

Bates SE: Epigenetic therapies for cancer. N Engl J Med 383:650, 2020.

Carter B, Zhao K: The epigenetic basis of cellular heterogeneity. Nat Rev Genet 22:235, 2021.

Dai Z et al: The evolving metabolic landscape of chromatin biology and epigenetics. Nat Rev Genet 21:737, 2020.

Dinardo AR et al: Postinfectious epigenetic immune modifications—a double-edged sword. N Engl J Med 384:261, 2021.

Hwang JY et al: The emerging field of epigenetics in neurodegeneration and neuroprotection. Nat Rev Neurosci 18:347, 2017.

Janssen SM, Lorincz MC: Interplay between chromatin marks in development and disease. Nat Rev Genet 23:137, 2021.

Millán-Zambrano G et al: Histone post-translational modifications—cause and consequence of genome function. Nat Rev Genet 23:563, 2022.

Sendinc E, Shi Y: RNA m6A methylation across the transcriptome. Mol Cell 83:428, 2023.

Vandereyken K et al: Methods and applications for single-cell and spatial multi-omics. Nat Rev Genet 24:494, 2023.

Zhang W et al: The ageing epigenome and its rejuvenation. Nat Rev Mol Cell Biol 21:137, 2020.

498

The Role of Circadian Biology in Health and Disease

Jonathan Cedernaes, Kathryn Moynihan Ramsey, Joseph Bass

Circadian rhythms are anticipatory, circa 24-h, autonomous cycles of physiology and behavior. These evolutionarily conserved rhythms have evolved at both the cell and tissue level to synchronize organismal function in anticipation of the 24-h rotation of the Earth. A common feature of modern "24/7" life is the routine disruption of these endogenous circadian cycles due to the rise in shift work, jet travel across time zones, exposure to blue light–emitting devices at night, and disrupted sleep-wake behavior. In-depth characterization of the molecular basis of circadian disorders has generated novel avenues for research on how sleep-wake disruption has been associated with aging, metabolic disease, inflammation, and cancer. This chapter provides an overview of (1) the basic biology of the circadian system; (2) primary circadian rhythm and interrelated sleep disorders; and (3) the role of the circadian system in both normal human physiology and disease states. We also include an overview of how the emerging field of chronobiology may impact drug action. A glossary of terms used in circadian biology is summarized in **Table 498-1**.

BASIC EVOLUTION AND STRUCTURE OF THE CIRCADIAN SYSTEM

Long before the emergence of multicellular life, the Earth's constant rotation around its axis gave rise to a daily cycle of light and darkness. At the emergence of the first prototypal gene involved in biological clock regulation—3.4 billion years ago in photosynthetic cyanobacteria—the period of Earth's rotation along its own axis was only 8 h. The co-occurrence in molecular evolution of the biological clock and photosynthesis hints at an interrelated and selective advantage of the clock in the regulation of energetic processes. Indeed, biological clocks coordinate oxygenic reactions with periods of sunlight each day, and perturbation of clock cycles reduces fitness, reproduction, and survival. Additionally, clocks protect photosynthetic organisms from the DNA-damaging effects of sunlight by timing the production of DNA repair processes, such as photolyase-mediated repair, to the nighttime. Across billions of years of evolution, as day length has gradually extended to today's circa 24 h, highly conserved circadian clocks (from *circa diem*, meaning "about a day") have been found in all photosensitive organisms, governing a wide range of biochemical, physiologic, and behavioral processes. A defining property of the circadian clock system is that it enables organisms to anticipate, rather than simply react to, daily changes in the external environment that are tied to the day-night cycle. In mammals, circadian systems are organized hierarchically with a light-responsive "master" circadian pacemaker located within the suprachiasmatic nucleus (SCN) of the anterior hypothalamus, which in turn presides over a network of both extra-SCN and peripheral clocks (see "Anatomic Organization of the Circadian Clock Network" below). Daily light exposure signals to the SCN and entrains the circadian system to the 24-h day (see "Entrainment and Measurement of the Circadian System," below). In turn, the SCN maintains synchrony of a diverse network of both central and peripheral clocks via a variety of signals that have as of yet to be fully identified. These signals involve direct physiologic rhythms (core body temperature), the autonomic nervous system, and neuroendocrine signals, such as cortisol, which is part of the hypothalamic-pituitary-adrenal (HPA) axis.

MOLECULAR ORGANIZATION OF THE MAMMALIAN CIRCADIAN CLOCK

At the molecular level, mammalian circadian rhythms are generated by a transcription-translation autoregulatory feedback loop. The forward limb of the clock is composed of the basic helix-loop-helix transcription factors (TFs) CLOCK (or its paralogue, NPAS2) and BMAL1. These drive expression of their own repressors (PERs and CRYs) in the negative limb in a cycle that repeats itself every 24 h **(Fig. 498-1)**. A second short feedback loop involves CLOCK/BMAL1-mediated transcription of the retinoic acid–related orphan nuclear receptor families ROR and REV-ERB, which activate and repress *Bmal1* transcription, respectively. Rhythmic posttranslational regulation of the stability and degradation of core clock TFs occurs via events such as phosphorylation by casein kinase 1 epsilon (CK1ε) and casein kinase 1 delta (CK1δ) and ubiquitination by FBXL3 and FBXL21. In addition to the circa 24-h oscillation of core clock genes, a wide array of downstream clock-controlled genes (CCGs) exhibit broad rhythmic amplitude in expression, ultimately giving rise to rhythmic physiologic processes.

The importance of localized clock gene expression has been demonstrated by genetic animal studies, such as with targeted ablation of *Bmal1*, the only clock gene that lacks a known functional paralog. Deletion of *Bmal1* either in the whole brain or in regions that span the brain region that coordinates circadian rhythms—the SCN—causes behavioral arrhythmicity, even when genetic ablation occurs in adult life. Conversely, restoring *Bmal1* expression specifically in brain in global adult *Bmal1* mutant mice rescues behavioral locomotor rhythms. Of note, whereas the protein CLOCK normally heterodimerizes with BMAL1, the paralogous protein NPAS2 can functionally substitute for CLOCK within the pacemaker neurons. Thus, while mice lacking either *Clock* or *Npas2* genes maintain rhythmicity, mutants lacking both CLOCK and NPAS2 lack circadian rhythms in locomotor activity. Further, mutations in many of the clock genes are associated with impaired circadian rhythms and physiology in both experimental animal models and humans (see "Primary Pathologies of the Circadian System" below).

A major transformation in our understanding of circadian biology came with the discovery that the molecular clock network is present not only in the SCN but also within most peripheral tissues, as well as in extra-SCN neurons in the brain. In primates, ~82% of all protein-coding transcripts exhibit daily 24-h rhythms in some tissue or other. In rodents studied under constant conditions, ~3–16% of the transcriptome in each tissue exhibits 24-h rhythms in mRNA expression levels, even though the repertoire of such genes varies substantially between tissues, in accordance with tissue-specific functions. The core clock feedback loop and the induction of transcriptional CCG rhythms also involves epigenetic mechanisms such as conformational chromatin dynamics, histone acetylation, and DNA methylation. Conversely, posttranscriptional events such as RNA polyadenylation, nucleocytoplasmic shuttling, alternative splicing, and mRNA translation also exhibit circadian variation, further increasing the repertoire of rhythmic regulation at a cellular level.

ANATOMIC ORGANIZATION OF THE CIRCADIAN CLOCK NETWORK

The molecular circadian feedback loop is synchronized with sunrise each day by photosensitive melanopsin-expressing neurons within the retina. These neurons provide input to the SCN via the retinohypothalamic tract (RHT), allowing mammals to maintain coherent

TABLE 498-1 Glossary of Terms Used in Discussion of the Circadian System

TERM	DESCRIPTION
ASPD	Advanced sleep phase disorder (see text for description).
CBT	Core body temperature. Often used as an indicator of the circadian rhythm but can be masked by sleep and exercise.
CCGs	Clock-controlled genes; output of the molecular clock.
Chronotype	Internal circadian rhythm of an individual determined by phase of entrainment, determining sleep propensity and timing of maximum alertness over a 24-h period.
Circadian period	Time required for one complete cycle or oscillation. Calculated by the time distance between two consecutive peaks or troughs of a circadian variable.
Circadian phase	Timing of the circadian rhythm. Defined by comparing, e.g., the peak (acrophase) or trough (bathyphase) to a fixed event, e.g., to a point in time. Synonymous with phase angle.
Circadian rhythm	A biological process that exhibits an endogenous, entrainable oscillation of ~24 h.
Circadian rhythm sleep disorders	Disorders of multiple etiology that have in common that they result in maladjustment of the biological clock with respect to the environment.
Constant routine	An experimental paradigm designed to study endogenous circadian rhythms in humans, by keeping behavioral and environmental factors constant. These paradigms thereby typically entail a combination of constant dim lighting, evenly distributed isocaloric energy intake, semirecumbent posture, and forced extended wakefulness.
Desynchrony	Loss of synchrony occurring either between a rhythm and its zeitgeber (external, "time giver" signal) or between two or more rhythms within an organism (internal).
Diurnal rhythm	An oscillation synchronized with the day/night cycle that repeats itself with a 24-h period. The rhythm does not have to persist when time cues (e.g., light) are absent.
DLMO	Dim-light melatonin onset; a marker of melatonin rhythm.
DSPD	Delayed sleep phase disorder (see text for description).
Entrainment	Synchronization of a circadian rhythm or other self-sustaining oscillation by a factor—zeitgeber—that enforces the oscillator. Constant entrainment between the zeitgeber and the oscillator results in a stable phase relationship between these entities.
Infradian rhythm	A recurrent cycle or period with a period length significantly greater than 24 h.
Melatonin	Hormone produced primarily by the pineal gland (chemical name *N*-acetyl-5-methoxytryptamine); derived from L-tryptophan. Various forms of melatonin can be prescribed for circadian rhythm sleep disorders or sleep disorders.
Non-24-h rhythm disorder	A syndrome in which there typically are chronic 1- to 2-h daily delays in sleep onset and wake times in an individual living in society, e.g., due to complete blindness.
Peripheral clocks	Clocks presiding outside of the suprachiasmatic nucleus, the circadian system's master pacemaker.
PRC	Phase response curve; visual representation of how a particular manipulation (e.g., light) produces phase shifts as a function of the phase (i.e., circadian time) at which the manipulation occurs. Defining the PRC to light has enabled researchers to understand and predict how entrainment to light cycles is accomplished.
SCN	The suprachiasmatic nucleus or nuclei, also known as the master pacemaker in mammalian species. A bilateral set of nuclei positioned in the anterior ventral hypothalamus. Essential for entraining extra-SCN central and peripheral oscillators to the prevailing light-dark cycle via photic input from the retina.
Shift work	Work scheduled so that it occurs outside of the traditional work schedule of 9:00 A.M. to 5:00 P.M., or 7:00 A.M. to 6:00 P.M., depending on definition. Various forms of shift work exist, such as early morning, evening, or night shifts, as well as rotating shifts.
Ultradian rhythm	A recurrent cycle or period with a period significantly shorter than 24 h—e.g., a 2-h rhythm would exhibit 12 cycles within a circadian (24-h) rhythm.

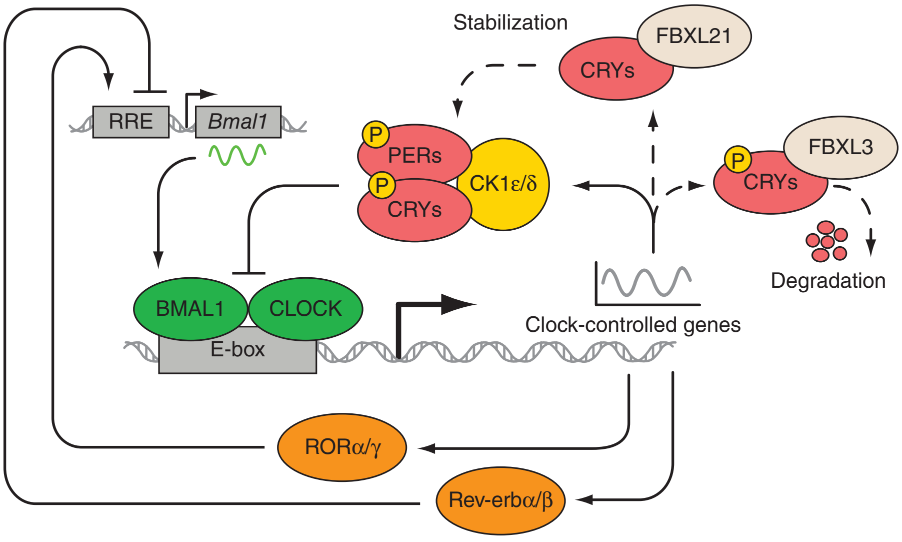

FIGURE 498-1 Central clock molecular mechanism. The core molecular clock machinery in mammals is encoded by interlocking transcription-translation feedback loops that oscillate with ~24-h periodicity. The transcription factors CLOCK and BMAL1 heterodimerize to drive transcription of downstream clock-controlled target genes containing E-box enhancer elements. Among these, the PER and CRY proteins multimerize and inhibit CLOCK/BMAL1, while RORs and REV-ERBs activate and inhibit, respectively, *Bmal1* transcription, resulting in rhythmic oscillations of clock-controlled and downstream target genes.

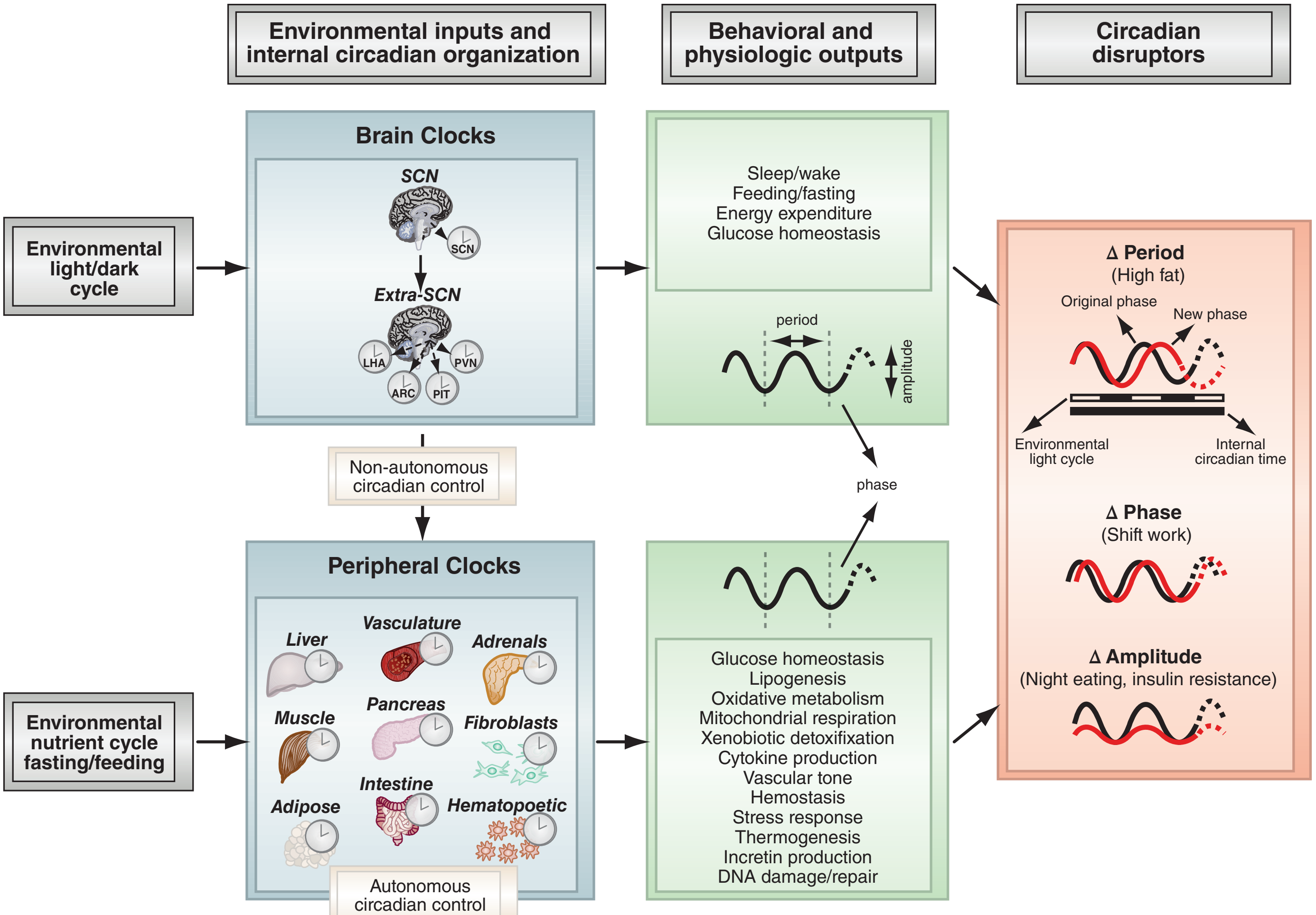
Environmental inputs and internal circadian organization
Behavioral and physiologic outputs
Circadian disruptors
Environmental light/dark cycle
Brain Clocks
SCN
SCN
Extra-SCN
LHA
PVN
ARC
PIT
Sleep/wake
Feeding/fasting
Energy expenditure
Glucose homeostasis
period
amplitude
phase
Non-autonomous circadian control
Peripheral Clocks
Liver
Vasculature
Adrenals
Muscle
Pancreas
Fibroblasts
Intestine
Adipose
Hematopoetic
Autonomous circadian control
Environmental nutrient cycle fasting/feeding
Glucose homeostasis
Lipogenesis
Oxidative metabolism
Mitochondrial respiration
Xenobiotic detoxifixation
Cytokine production
Vascular tone
Hemostasis
Stress response
Thermogenesis
Incretin production
DNA damage/repair
Δ Period
(High fat)
Original phase
New phase
Environmental light cycle
Internal circadian time
Δ Phase
(Shift work)
Δ Amplitude
(Night eating, insulin resistance)

FIGURE 498-2 Central and peripheral clocks coordinate environmental cues with behavior and physiologic outputs. Light entrains the master pacemaker neurons in the suprachiasmatic nucleus (SCN), which subsequently synchronizes extra-SCN and peripheral clocks. Brain clock output includes sleep-wake, fasting-feeding, and energy expenditure cycles, while peripheral clock output includes a wide range of physiologic processes, including glucose homeostasis, oxidative metabolism, cytokine production, and stress response. The right column indicates different ways that circadian disruptors, such as diet, shift work, or other circadian rhythm sleep disorders, may impact the clock—i.e., by changing circadian period, phase, or amplitude.

organismal rhythms in line with the external light/dark cycle. Understanding the circuit organization of the circadian clock within the brain is increasingly relevant in understanding how the master circadian pacemaker center within the SCN regulates feeding, sleep-wake activity, endocrine processes, energy expenditure, and metabolism **(Fig. 498-2)**. Identification of the SCN as the master pacemaker was first established by the observation that SCN lesioning induced complete loss of rhythms of locomotor activity, drinking behavior, and endocrine hormone secretion. The ventral "core" region of the SCN, which is composed of neurons producing vasoactive intestinal polypeptide (VIP), receives photic information directly from the retina through the RHT. At the molecular level, circadian gene transcription is induced within the SCN through the initial activation of immediate early genes, such as *Per1*, *Per2*, *c-fos*, and *jun*. Cells within the "core" region of the SCN then signal primarily via γ-aminobutyric acid (GABA)-ergic neurotransmitter release to synchronize the cells within the "shell" region of the SCN, which produce arginine vasopressin (AVP), the most important neuropeptide for maintaining intra-SCN synchronicity.

The SCN communicates to extra-SCN and peripheral clocks through both secreted factors and neuronal projections. The former was elegantly proven by the ability of SCN grafts to partially restore locomotor rhythms in SCN-lesioned animals. Efferent nerve outputs arise both from the AVP-producing shell region of the SCN and the VIP-predominated core. The SCN projects to several hypothalamic and thalamic regions, including the median preoptic nucleus (MPO), the subparaventricular zone (SPZ), the dorsomedial hypothalamus (DMH), the paraventricular nucleus of the hypothalamus (PVH), and the paraventricular nucleus of the thalamus (PVT). Some of these regions, in turn, regulate output to both sleep- and wake-promoting regions, as well as to regions involved in regulation of autonomic, body temperature, and hormonal rhythms, as well as feeding. The SCN is thereby thought to promote sleep in part through the transmission of neural signals that terminate in the sleep-promoting ventrolateral preoptic nucleus (VLPO), i.e., one of the brain regions that is active during sleep. In contrast, the SCN promotes wakefulness during the active phase by transmission of neural signals that—by passing through regions such as the SPZ and the DMH—terminate in wake-promoting regions, including the locus coeruleus, lateral hypothalamic nucleus, ventral tegmental area, and dorsal raphe nucleus.

The SCN also signals via noradrenergic fibers to the pineal gland to regulate the circadian production of the hormone melatonin. SCN control of the nighttime rise in pineal melatonin release (in both diurnal and nocturnal animals) is mediated through a pathway involving the PVH. Of note, artificial light at night delays the secretion of melatonin, ultimately affecting sleep (see "Endocrine Systems Regulated by the Circadian Clock" below). Melatonin plays a complex role in the circadian system since the MT1 and MT2 melatonin receptors are expressed in the SCN itself; thus, melatonin feeds back to modulate circadian outputs to other cells in the brain and body.

Neuronal output from the SCN also reaches peripheral tissues such as the adrenal glands, liver, and pancreas. The SCN produces rhythmic variation in multiple neuroendocrine axes, producing daily rhythms of gonadotropin, thyrotropin, and somatotropin. Prominent HPA axis rhythms ultimately give rise to daily variation in diverse pathways essential for hemodynamic stability, metabolism, and inflammation. These rhythms originate with SCN control of corticotropin-releasing hormone (CRH)–producing cells in the PVH, which may regulate sleep as well as induce daily oscillations of both pituitary adrenocorticotropic hormone (ACTH) and adrenal cortisol. Highlighting the importance of SCN output for peripheral rhythms, there is a dramatic reduction in the number of transcripts that exhibit circadian rhythms in the liver following SCN ablation in mice. Nonetheless, when the autonomous clock in the liver is ablated in mice, some key clock transcripts such as *Per2* still cycle provided the core body temperature rhythm persists. Whereas the SCN is exclusively entrained by light, meal timing can signal circadian time directly to peripheral tissues such as the liver. Thus, shifted meal timing as occurs during shift work or jetlag can uncouple peripheral clocks from the central pacemaker. Temperature can also phase shift peripheral tissue clocks, but not the SCN clock. This is an important phenomenon because, at the organismal level, the SCN generates the core body temperature rhythm as one of the major mechanisms to signal circadian time to peripheral clocks.

ENTRAINMENT AND MEASUREMENT OF THE CIRCADIAN SYSTEM

Under normal light-dark cycles, the circadian system is corrected or "entrained" on a daily basis, producing diurnal rhythms of 24 h. Such signals of entrainment are called zeitgebers (German for "time-giver" signals) and include light exposure, meal timing, and activity patterns. Light serves as the dominant zeitgeber for the circadian system, and a breakthrough in understanding photoentrainment in mammals came with the discovery of the melanopsin system, which is composed of a specialized class of photosensitive retinal ganglion cells that expresses the blue light–sensitive photopigment melanopsin in the inner retina, separate from the photoreceptive rods and cones. Blue light around this wavelength (~480 nm) suppresses melatonin, such that melatonin levels are normally low during the day, promoting subjective and objective (electroencephalography assessed) wakefulness.

The ability of light to entrain the circadian system functions according to a so-called phase response curve (PRC). When light exposure occurs prior to the critical phase of the core body temperature (CBT), defined by the CBT's minimum, light produces a phase delay in the circadian rhythm. Conversely, light exposure after this critical period causes phase advances. The circadian system can respond even to small changes in light intensity (e.g., dim light at ~100 lux can produce half of the phase delay compared with an almost 100-fold greater light exposure). This responsiveness has been found to be highly individual and varies widely. This is in part due to genetic variation, as variants in clock genes can modulate the responsiveness of the human circadian system to light.

When an organism is placed in an environment without zeitgebers, the circadian rhythm is said to free-run, as it relies on the endogenous rhythm of the circadian system. In humans, the study of endogenous circadian rhythms can be achieved by using a so-called constant routine that eliminates the risk of masking by factors such as sleep. In these paradigms, subjects are kept awake in a constant semi-recumbent posture, meals are provided on an hourly basis, and light is constantly kept below the level that can phase shift the SCN. Concurrently, circadian rhythms are assessed by frequently measuring CBT, melatonin, or peptidergic hormone rhythms over the course of more than 24 h. In animals, endogenous circadian rhythms are instead studied by examining behavior, physiologic responses, and voluntary locomotor activity following 30–36 h of complete darkness. From these measurements, key properties of the circadian system can be ascertained, such as period length (peak-to-peak or trough-to-trough time), amplitude (peak-to-trough difference), and phase (timing of peak or trough in relation to a reference point) (Fig. 498-2).

These studies have revealed that the endogenous human circadian clock runs with a period length of ~24.2 h, while that of mice runs at ~23.5 h, with some variability across strains. In humans, evidence further indicates that females may have a slightly shorter circadian clock than males (24.1 vs 24.2 h), and many circadian parameters have been found to exhibit differences that are dependent on biological sex. Notably, interindividual variability in the endogenous circadian period length is further diversified by the existence of genetic polymorphisms in clock genes (see below). These gene variants can confer extremes in the endogenous circadian period as well as phase; the latter can be advanced or delayed by ~3–4 h in each direction. This is due both to altered circadian rhythms at the cellular level and to altered SCN responsiveness to entrainment by light. For instance, *PER3* gene contains a variable-number, tandem-repeat polymorphism. Individuals homozygous for a *PER3 5/5* genotype have been reported to be more responsive than *PER3 4/4* homozygous individuals to the melatonin-suppressing effect of evening blue light exposure. By analyzing genetic variation across ~700,000 individuals, the number of genetic loci that have been identified that contribute to variability in chronotype is in the hundreds.

Using specifically developed questionnaires to establish preferred sleep-wake timing, individuals can be categorized into so-called morningness-eveningness types or chronotypes. The most commonly used questionnaires are the Horne-Östberg Morningness-Eveningness Questionnaire (MEQ) and the Munich ChronoType Questionnaire (MCTQ). A composite MEQ score allows grouping into five categories that range from definite morning-type to evening-type individuals based on preferred waking time. In contrast, the MCTQ centers on the midpoint of sleep as a circadian marker, queries age and sex across a range of geographical locations, and can be used to ascertain differences between socially imposed sleep patterns (e.g., on working days) and sleep patterns on free days (the difference constituting so-called social jetlag). According to MCTQs obtained from primarily European populations, ~1% of the general population goes to bed before 10:00 P.M. and ~8% after 3:00 A.M. Differences in chronotype are linked to altered circadian timing, including peak levels of melatonin, which can vary by up to 4 h between extreme morning and evening types. Extreme chronotypes have also been shown to be linked to various traits; i.e., low morningness scores have been associated with greater tolerance to night shift work.

Melatonin is one of the most commonly used peripheral markers of an individual's circadian rhythm, reflecting the rhythmic function of the SCN. Circadian rhythms of melatonin can be measured in saliva or plasma, whereas 6-sulphatoxymelatonin (aMT6S), a metabolite generated from the breakdown of melatonin, can also be measured in urine. Accurate estimations of melatonin rhythms are often obtained by analyzing the dim light melatonin onset (DLMO). As the name implies, this involves evening/nighttime sampling of melatonin as opposed to 24-h sampling. This makes DLMO quantification useful in both the clinical and research settings. In normally entrained individuals, the DLMO can be used to ascertain whether an individual's circadian rhythm is phase advanced or delayed, and this onset typically occurs ~2 h before the onset of sleep. The midpoint of sleep—the main marker used by the MCTQ—correlates more strongly with melatonin onset than the MEQ score. In the morning hours, the offset of melatonin ("DLMoff") can be used as a marker of circadian alignment or misalignment with the light-dark cycle. When individuals are exposed only to the natural light-dark cycle dictated by the sun, such as in an outdoor natural lighting environment, DLMO and DLMoff occur earlier. In contrast, exposure to artificial light has the overall effect of delaying the biological night and contributes to widening differences between chronotypes in modern society.

The CBT is also often utilized as an indicator of the circadian rhythm. Even though CBT is more variable than DLMO, it usually correlates well with the phase obtained using the melatonin rhythm. The CBT, however, can be masked by factors such as sleep, food intake, and activity. CBT can be recorded and registered wirelessly with relative ease. In humans, CBT can be recorded via rectal thermometers or probes that are swallowed to pass through the gastrointestinal tract. When humans are studied under normal conditions with normal lighting and sleep duration from 2300 to 0700 h, the CBT reaches around 37.2°C by 0900 h, and from there, it continues to rise slowly until it reaches 37.4°C around 11 h later. The CBT then drops to the daily low of 36.5°C in the early morning (0400 h). The minimum in body temperature also corresponds to the trough in the 24-h rhythm in resting energy expenditure.

Given the interrelationship between the circadian system and sleep-wake systems, researchers have developed paradigms that uncouple the circadian system from sleep-wake states, enabling the study of the contribution of the circadian system to investigated parameters across the entire sleep-wake cycle. These paradigms are known as “forced desynchrony” protocols and involve enforcing a significantly shortened (e.g., 20 h) or prolonged (e.g., 28 h) day length upon individuals. These protocols thus attempt to approximate what occurs during rotating shift work or “jetlag,” e.g., when travel across several time zones suddenly shifts the light-dark and behavioral cycles drastically away from the entrained 24-h rhythm. As described below, forced desynchrony protocols have contributed to uncovering how the circadian system regulates parameters such as cognitive performance, subjective alertness, and metabolic and cardiovascular health.

TABLE 498-2 Criteria for Circadian Rhythm Sleep Disorders

CRITERIA	DESCRIPTION
A	A persistent or recurrent pattern of sleep disturbance due primarily to one of the following: • Alterations of the internal circadian timekeeping system. • Misalignment between endogenous circadian rhythms and exogenous factors that affect the timing or duration of sleep.
B	A circadian-related sleep disruption that leads to insomnia, excessive daytime sleepiness, or both.
C	A sleep disturbance that is associated with impairment of social, occupational, or other areas of functioning.

PRIMARY PATHOLOGIES OF THE CIRCADIAN SYSTEM

An overarching term for disorders of the circadian system is circadian rhythm sleep disorders (CRSDs), where there is a mismatch between subjective behavioral and physiologic rhythms with the environmental light-dark or social activity-rest cycles (i.e., the body clock is out of sync with the external light-dark cycle). CRSDs can arise either due to misalignment of an exogenous environmental factor, such as light, or misalignment of the activity-rest cycle, such as occurs with shift work or jetlag, in relation to endogenous circadian timing. Mutations in the core clock genes themselves can also alter intrinsic circadian timing in relation to the external environment, which makes it difficult for these individuals to properly realign themselves. These disorders often result in adverse effects such as excessive sleepiness or depressed mood, often causing individuals to be unable to maintain a job or attend school at regular hours. The criteria for CRSDs based on the International Classification of Sleep Disorders (ICSD) are shown in **Table 498-2**.

Animal models have greatly advanced our understanding of how core molecular clock components contribute to maintaining normal sleep-wake/rest-activity cycles **(Table 498-3)**. For example, *Clock*$^{\Delta19/\Delta19}$ mice have reduced total sleep duration and less induction of rapid eye movement (REM) sleep in response to sleep deprivation. Further, mice that lack *Bmal1* have increased total sleep time, but it is more fragmented and lacks clear 24-h sleep-wake rhythms, and mice lacking the repressors *Cry1* and *Cry2* are arrhythmic and spend more time in non-REM sleep. Finally, while ablation of the circadian gene *Dbp* does not alter the specific duration of sleep stages, it does lead to an altered circadian sleep-wake distribution, with more sleep during the normal wake period and vice versa. Consistent with a key role of clock genes in regulating sleep-wake behavior, human genetic studies of twins have found that up to half of the variation in diurnal preference is heritable. Established genetic variants associated with diurnal preference and circadian sleep disorders are listed in **Table 498-4**. The following briefly mentions the most common CRSDs, but readers should refer to the chapter on sleep disorders **(Chap. 33)** for a more detailed description.

Delayed Sleep Phase Disorder Delayed sleep phase disorder (DSPD; or delayed sleep-wake phase disorder [DSWPD]) is one of the more common circadian rhythm sleep disorders, ranging from 0.2–16% of the population depending on definition used, and is most common in adolescence and early adulthood. DSPD is characterized by chronic and significant delays in both sleep onset and wake times compared to “socially acceptable” sleep-wake hours (i.e., “extreme night owls”). Rhythms of CBT and melatonin levels are also often

TABLE 498-3 Animal Models of Genetic Circadian Disruption

	AVERAGE CIRCADIAN TIME OF PEAK TRANSCRIPT LEVEL			
GENE	SCN	PERIPHERY	ALLELE	MUTANT PHENOTYPE
Bmal1 (Arntl)	15–21	22–2	*Bmal1*$^{-/-}$	Arrhythmic
CK1δ (Csnk1δ)	No rhythm	No rhythm	*Csnk1δ*$^{+/-}$	0 to 0.5-h shorter period
CK1ε (Csnk1ε)	No rhythm	No rhythm	*CK1ε*^{tau}	4-h shorter period
CK1ε	—	—	*CK1ε*$^{-/-}$	0.2- to 0.4-h longer period
Clock	No rhythm	21–3	*Clock*$^{-/-}$	0.5-h shorter period
—	—	—	*Clock*$^{\Delta19/\Delta19}$	4-h longer period/arrhythmic
Clock/Npas2	—	—	*Clock*$^{-/-}$/*NPAS2*$^{-/-}$	Arrhythmic
Cry1	8–14	14–18	*Cry1*$^{-/-}$	1-h shorter period
Cry2	8–14	8–12	*Cry2*$^{-/-}$	1-h longer period
—	—	—	*Cry2*^{A260T}	0.2-h shorter period
Dbp	—	—	*Dbp*$^{-/-}$	0.5-h shorter period
Npas2	N/A	0–4	*Npas2*$^{-/-}$	0.2-h shorter period
Per1	4–8	10–16	*Per1*$^{-/-}$	0.7-h shorter period
—	—	—	*Per1*^{brdm1}	1-h shorter period
—	—	—	*Per1*^{ldc}	0.5-h shorter period/arrhythmic
Per2	6–12	14–18	*Per2*^{brdm1}	1.5-h shorter period/arrhythmic
—	—	—	*Per2*^{ldc}	Arrhythmic
Per3	4–9	10–14	*Per3*$^{-/-}$	0 to 0.5-h shorter period
Rev-erbα (Nr1d1)	2–6	4–10	*Rev-erbα*$^{-/-}$	0.5-h shorter period/disrupted photic entrainment
Rorα	6–10	Arrhythmic/various	*staggerer*	0.5-h shorter period/disrupted photic entrainment
Rorβ	4–8	18–22	*Rorβ*$^{-/-}$	0.5-h longer period
Rorγ	N/A	16–20/various	*Rorγ*$^{-/-}$	Normal behavior

Note: Normal circadian rhythms of circadian clock and related genes, with description of circadian phenotype in mutant mice.

Abbreviation: N/A, not applicable.

Source: Adapted from Hum Mol Genet 15:R271, 2006, and Adv Genet 74:175, 2011.

TABLE 498-4 Mutations and Gene Variants Linked to Sleep-Wake Disorders and Diurnal Preference

GENE	POSITION	POPULATION	SYNDROME/SLEEP PREFERENCE
h*CKIε*	S408N	Japanese	Protection against DSPS
h*CKIγ*	T44A	Pedigree	FASPS
h*CKIΔ*	H46R	Pedigree	FASPS
h*CLOCK*	T3111C (3′-UTR)	European	Eveningness
h*CRY2*	A260T	Pedigree	FASPS
h*PER2*	S662G (missense mutations in CKIε binding region)	Pedigree	FASPS
h*PER2*	C111G (5′-UTR)	British	Extreme morningness
h*PER3*	P415A/H417R	Pedigree	FASPS and seasonal affective disorder
h*PER3*	G647	Swedish/Finish/Austrian/German	Morningness
h*PER3*	G647, P864, 4-repeat, T1037, R1158	Japanese	DSPS
h*PER3*	Increased repeats (exon 18, 54 bp)	Brazilian	DSPS
h*VIP*	rs9479402 (gene variant 54 kb upstream of VIP)	European (>97% European ancestry)	Morningness

Abbreviations: DSPS, delayed sleep phase syndrome; FASPS, familial advanced sleep phase syndrome.

delayed. DSPD is associated with polymorphisms within the circadian clock genes *CLOCK*, *PER3*, and *CRY1*, and the circadian period (tau) of these individuals may be longer. The most effective treatment includes bright-light therapy after waking in the morning (and/or dark-room therapy in the evening) in combination with melatonin administration in the evening several hours prior to the onset of sleep. These approaches attempt to realign endogenous circadian rhythms with the desired sleep-wake schedule, though often face challenges because individuals suffering from DSPD also phase delay more rapidly.

Advanced Sleep Phase Disorder Another CRSD whereby one gets the correct amount and quality of sleep but at a shifted time is advanced sleep phase disorder (ASPD; or advanced sleep-wake phase disorder [ASWPD]). The prevalence of this disorder may be <1%, but the condition may be underreported, given that it may cause fewer conflicts with societal demands (i.e., 9-to-5 schedules) compared with DSPD. Individuals with ASPD experience an advance in their major sleep episode in relation to the desired sleep-wake times. Thus, this disorder typically results both in very early evening bedtimes and morning awakenings (e.g., "extreme early birds"), resulting in reduced quality of life due to excessive sleepiness during the early evening, even in social situations. Individuals with ASPD also have phase-advanced temperature and melatonin rhythms. ASPD occurs more often in older individuals, although early-onset autosomal dominant familial variants (familial advanced sleep phase syndrome [FASPS]) have also been associated with mutations in either the *PER2* or the casein kinase 1δ (*CK1δ*) gene. *PER2* is critical for SCN resetting by light, and these *PER2* mutations have been found to shorten the endogenous circadian period to ~23.3 h compared with the normal 24.2-h period length. Treatment includes bright light or blue-enriched phototherapy in the evening hours to delay the phase of the circadian clock to a later hour.

Shift Work Sleep Disorder Given the increased prevalence of shift work in today's 24/7 society and the accumulating evidence for increased incidence of sleep and metabolic disorders, including obesity, type 2 diabetes, cardiovascular disease, and cancer, in shift workers, the need to develop effective treatments for shift work sleep disorder (SWSD) is increasingly important. SWSD is at its core defined by the primary symptom of either insomnia or excessive sleepiness arising due to work scheduled during regular sleeping hours or at irregular times. The symptoms may arise because recovery sleep consumes a large proportion of the individual's free time, potentially leading to negative social consequences such as difficulties maintaining social relationships. Older individuals are typically at an increased risk of SWSD due to the age-associated decline in the ability to maintain sleep during the time of day that would normally constitute the wake period. Therapeutic approaches include optimizing the sleep environment at home to minimize disruptions, melatonin prior to sleeping, and timed bright-light therapy. For example, for night workers, intermittent bright-light exposure during the night and avoidance of bright light during the morning, even on days off, have been shown to improve sleep and feelings of alertness. Genetic screening combined with chronotype questionnaires may become useful tools for determining whether a given individual is suited for shiftwork. For instance, a twin study indicated that a genetic variant of the circadian gene *DEC2* was associated with reduced sleep duration and shorter recovery sleep following extended sleep deprivation. More studies may reveal additional genetic variants that confer an advantage to repeated phase advances and phase delays as typically occurs in shift work.

Irregular Sleep-Wake Rhythm Damage to the SCN can produce arrhythmicity in animals and is thought to be one of the possible underlying reasons for the temporally disorganized sleep-wake pattern that characterizes the disorder known as irregular sleep-wake rhythm (ISWR). Other contributing factors may be reduced responsiveness to entraining signals such as light and physical activity, as well as decreased exposure to such signals, as often occurs with increasing age. Despite normal total sleep time, there is a relative absence of a circadian pattern to the sleep-wake cycle such that sleep occurs in several distinct randomly distributed bouts. ISWR is often associated with neurologic impairment, foremost Alzheimer's disease in older age; however, ISWR can also occur in individuals with poor sleep hygiene. Treatments involve multimodal interventions such as increased light exposure, improved sleep hygiene, and promotion of social and physical activities.

Non-24-h Sleep-Wake Rhythm Disorder Individuals with non-24-h sleep-wake rhythm disorder ("non-24"), otherwise known as free-running disorder (FRD), have endogenous circadian rhythms that are not synchronized with the external 24-h day-night cycle due to an inability to readjust the circadian clock to the 24-h day on a daily basis. This most commonly occurs in individuals who are completely blind (i.e., lacking all photoreceptors) since they are unable to respond to daily light cues that normally would reset the endogenous circadian clock (although the condition has also been reported in sighted individuals). Instead, the sleep-wake period length corresponds to the individual's endogenous circadian rhythms, which are typically slightly longer than 24 h, thereby shifting sleep and wake cycles over time in relation to the light-dark cycle. Instead of sleeping at the same time each day, their sleep time would gradually be delayed each day until their sleep period literally goes "around the clock." Depending on the individual's endogenous rhythm, the individual will take a given number of days to realign their endogenous phase (in a 360° phase plot) with the zero time point in the exogenous 24-h light-dark cycle. Because of this chronic cycling, prominent symptoms of non-24 include sleep-wake cycle disruption (insomnia and daytime sleepiness), impaired alertness and mood levels, and severe difficulties partaking in normally scheduled work, school, or social activities. Non-24 can be diagnosed following diurnal analysis of an individual's melatonin or cortisol rhythms, in combination with analyses of sleep diaries where the sleep onset and offset can be visualized over time to identify the free-running period. Treatments for sighted non-24-h patients include a combination of bright-light therapy with appropriately timed melatonin administration, whereas melatonin and dual melatonin (MT1 and MT2) receptor agonist administration in completely blind non-24-h patients has been shown to entrain free-running rhythms and improve symptoms.

Jetlag Most have experienced symptoms associated with jetlag, including insomnia, daytime sleepiness, and fatigue, when traveling from one time zone to another, as one's endogenous circadian rhythms are not yet aligned, or entrained, to the new external light-dark cycle. This is due to the slowness of the circadian system to adapt to the new time zone. Typically, the human circadian system can shift up to ~1.5 h a day in the westward direction (i.e., a phase delay), whereas it shifts more slowly (up to ~1 h daily) with eastward direction of travel (i.e., achieving a phase advance). Usually, symptoms of jetlag abate within the first couple of days after traveling and may present themselves after a first night of good sleep (which is more dependent on a high buildup of homeostatic sleep pressure). Older individuals (age >50) appear to be more at risk. While symptoms are transient, therapeutic approaches aim to hasten the synchronization of internal and external circadian cycles. Behavioral treatments include appropriately timed bright-light exposure and avoidance of bright light during the nighttime in the new destination, while pharmacologic approaches include timed melatonin administration before bedtime both prior to and following travel, resulting in improved sleep quality and decreased night waking.

Social Jetlag Individuals with a late chronotype are prone to suffer from "social jetlag," a phenomenon in which individuals are forced to awaken at a point at which their bodies are entrained to be asleep due to discrepancy between alignment of social and biological time. Social jetlag can be estimated using questionnaires, such as the MCTQ, to compare sleep timing on working or school days compared with free days. This has established that a large proportion of the European population suffers from 2 or more hours of social jetlag. Chronic social jetlag is associated with an increased risk of developing obesity and metabolic syndrome, as well as with greater alcohol consumption, smoking, and poorer academic performance in students.

The aforementioned categories of defined clinical circadian disorders have been traditionally established based on consideration of the endogenous behavioral and physiologic cycles (primarily of melatonin and temperature) with the external 24-h light-dark cycle. In the following sections, we build on the concepts of circadian behavioral disorders to consider new and emerging insight into the role of circadian disruption in organismal homeostasis **(Figs. 498-3 and 498-4)** and the availability of genetic strategies to dissect the interrelationship between clock function, health, and disease.

ROLE OF THE CLOCK SYSTEM IN PHYSIOLOGY

Endocrine Systems Regulated by the Circadian Clock In addition to regulation of behavioral rhythms such as sleep-wake and fasting-feeding cycles, the circadian clock also regulates rhythms of the endocrine system. Cortisol rhythms are regulated through a feedback loop known as the HPA axis. Hypothalamic secretion of CRH and AVP promotes secretion of pituitary ACTH, which in turn regulates rhythmic cortisol secretion from the adrenal cortex. Cortisol release increases toward the morning, and this increase is believed to prepare the brain and peripheral tissues for daytime activity and food intake. AVP secretion in mice occurs prior to sleep to promote water intake, thereby preventing dehydration during the sleep period. Several hormones, such as growth hormone (GH), cortisol, and melatonin, are influenced not only via circadian regulation, but also by sleep. For

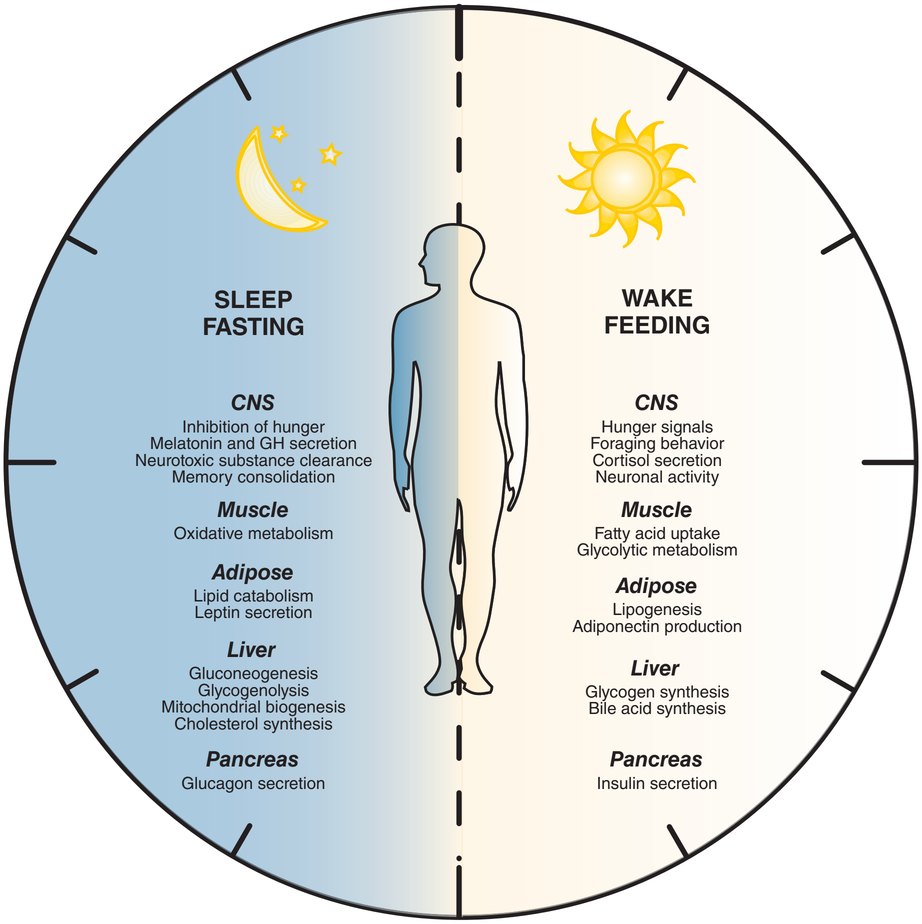
SLEEP
FASTING
CNS
Inhibition of hunger
Melatonin and GH secretion
Neurotoxic substance clearance
Memory consolidation
Muscle
Oxidative metabolism
Adipose
Lipid catabolism
Leptin secretion
Liver
Gluconeogenesis
Glycogenolysis
Mitochondrial biogenesis
Cholesterol synthesis
Pancreas
Glucagon secretion
WAKE
FEEDING
CNS
Hunger signals
Foraging behavior
Cortisol secretion
Neuronal activity
Muscle
Fatty acid uptake
Glycolytic metabolism
Adipose
Lipogenesis
Adiponectin production
Liver
Glycogen synthesis
Bile acid synthesis
Pancreas
Insulin secretion

FIGURE 498-3 The circadian clock partitions behavioral, physiologic, and metabolic processes according to time of day. The partitioning of metabolic processes to appropriate times of day is critical for the maintenance of health from cellular to mammalian organisms. This figure highlights which processes peak within the central nervous system (CNS), muscle, adipose, liver, and pancreas during either the sleep/fasting or wake/feeding cycle in humans. GH, growth hormone.

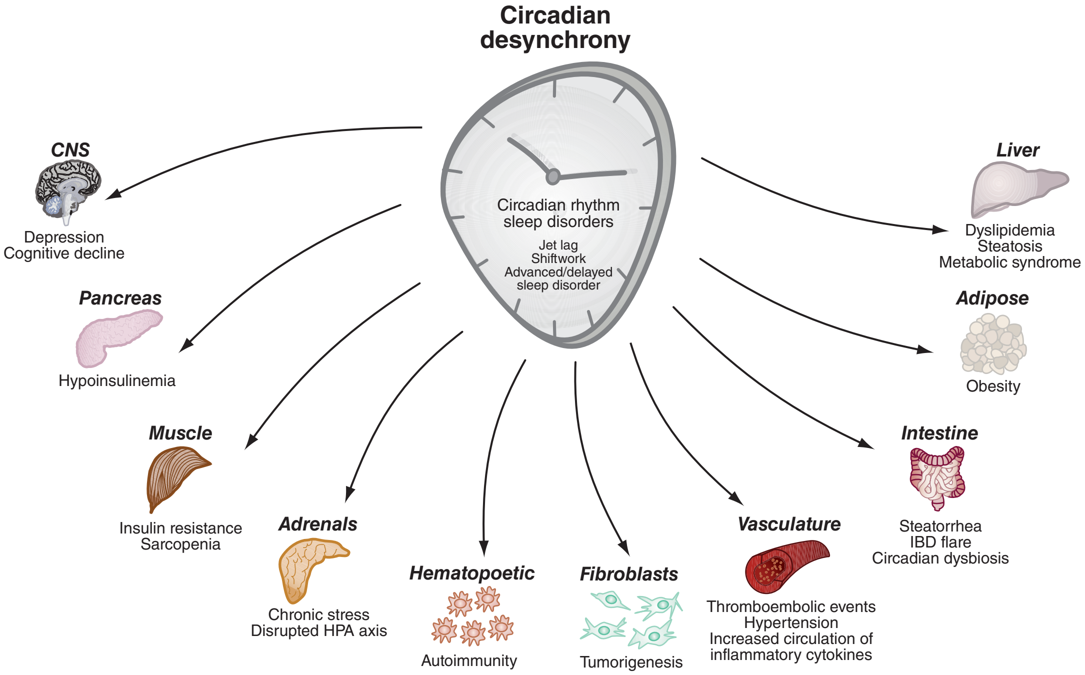

FIGURE 498-4 Pathologies resulting from circadian desynchrony. Circadian rhythm sleep disorders, including advanced/delayed sleep phase disorder, jet lag, social jet lag, and shift work, result in a desynchrony between the environmental light-dark cycle "time" and the endogenous clock "time." Pathologies can thus arise through misalignment imposed by exogenous (e.g., altered light cycle and/or feeding rhythm) and endogenous factors (e.g., mutations in core clock genes). Such desynchrony results in a host of wide-ranging pathologies across multiple tissues, including hypoinsulinemia (pancreas), disrupted hypothalamic-pituitary-adrenal (HPA) axis, autoimmunity, hypertension, obesity, and metabolic syndrome. CNS, central nervous system; IBD, inflammatory bowel disease.

instance, both GH secretion and the cortisol awakening response (CAR, i.e., the peak in cortisol soon after waking) are profoundly blunted by acute overnight wakefulness. GH secretion is primarily dependent on the occurrence of slow-wave sleep, which is a homeostatically driven sleep stage that occurs primarily in the first part of the sleep period. Both the CAR and daytime cortisol levels are also modulated by light exposure levels. Sleep also influences melatonin amplitude, such that sleep deprivation can increase melatonin levels. As sleep deprivation is often accompanied by artificial nighttime light exposure, the effect on melatonin can be combinatorial. In working environments, the effects of curtailed sleep are often confounded by mistimed exposure to light. The sensitivity to light levels that suppress melatonin can vary by an order of magnitude or more in individuals. This partly explains how even low levels of light can potently suppress melatonin secretion. Together with altered timing in light exposure, perturbed hormonal levels likely represent a mechanism through which altered timing and duration of sleep can impact central and peripheral circadian oscillators.

Centrally controlled rhythms of melatonin and cortisol are considered key regulators of extra-SCN and peripheral oscillators. Glucocorticoid receptors exist in both the central nervous system and in peripheral tissues such as skeletal muscle, liver, and adipose tissue. Upon acute shifts in light-dark or feeding cycles, rhythmic levels in cortisol appear to modulate the rate at which behavioral and physiological rhythms phase shift. Indeed, glucocorticoids regulate clock gene expression in muscle, kidney, and lung, and the powerful synthetic glucocorticoid dexamethasone is often employed in vitro for its ability to synchronize (e.g., reset) circadian rhythms of cells, including liver cells. Consistent with a role for glucocorticoid regulation of the clock, both adrenalectomy, which results in a lack of cortisol, and exogenous corticosteroid supplementation significantly disrupt the circadian clock system.

Several peripherally produced hormones and peptides are not only produced rhythmically but can also feed back to central clocks, including within the SCN. For instance, both cortisol and thyroid hormones regulate their own rhythmic synthesis by feedback to central brain regions, i.e., the hypothalamus (for cortisol) and pituitary (for both hormones). Several other peripherally produced factors have been proposed to influence the central clock, including fatty acids produced by the adipose tissue and fibroblast growth factor 21, a hormone primarily produced by the liver. Peripheral hormones that signal energy state and hunger also exhibit circadian rhythms that are regulated by local tissue clocks. The most extensively studied of these hormones are leptin, which is released from white adipose tissue cells, and ghrelin, which is released from specific endocrine cells in the upper fundus region of the stomach. Ghrelin also exhibits significant peaks related to anticipated meal timing, which persist for several days of fasting in humans. Circulating rhythms of leptin and ghrelin are disrupted in circadian mutant mice and in humans experiencing circadian misalignment, with evidence for sex-specific effects. *Per* and *Cry* mutant mice exhibit severely blunted leptin rhythms, and wild-type mice exposed to jetlag (through repeatedly altered light-dark cycles) show a reduced wake-associated decrease in leptin. Similarly, humans forced to live 28-h days exhibit increased 24-h profiles of ghrelin and decreased levels of leptin. Ghrelin and leptin signal to several regions of the brain, including integrative appetitive regions of the hypothalamus such as the arcuate and paraventricular region. The response to these hormones is rhythmically regulated by the molecular clock within several such central sites, effectively gating how these hormones influence rhythms of food intake and energy homeostasis in a time-of-day– and nutrient-dependent manner.

Role for the Clock in Metabolic Homeostasis Circadian control of glucose homeostasis has long been recognized, as early studies demonstrated variation in glucose tolerance and insulin action across the day. For example, due to a combination of circadian control of both peripheral insulin sensitivity and pancreatic β-cell insulin secretion, oral glucose tolerance is lower in the evening and afternoon compared with the morning. Another example is the "dawn phenomenon," whereby glucose levels peak prior to the onset of activity. Further, destruction of the SCN has been shown to abolish circadian regulation of glucose metabolism in rats, and daily cycles of insulin secretion and glucose tolerance are often perturbed in patients with type 2 diabetes, who also exhibit changes in gene expression rhythms in peripheral tissues such as adipose tissue. Changes in rhythmic parameters such as insulin secretion have also been observed in first-degree relatives of patients with type 2 diabetes, possibly highlighting a key hereditary role for the circadian clock in the pathogenesis of metabolic disease.

Ablating clock genes in mice has revealed a key function for both central and peripheral clocks in regulating energy homeostasis. The circadian system has been shown to regulate rhythmic insulin secretion from the pancreas via both neural signals and hormonal levels (e.g., cortisol and norepinephrine), as well as via cell-autonomous clock regulation within the pancreatic β cell itself. An early observation was that whole-body *Clock*$^{\Delta 19/\Delta 19}$ mutant mice developed obesity without displaying hyperinsulinemia, a phenomenon that indicated concurrent β-cell failure. This was later confirmed using pancreas- and β-cell–specific *Bmal1*-deficient mice, which exhibited glucose intolerance, hypoinsulinemia, and impaired glucose-stimulated insulin secretion. The molecular clock within other peripheral tissues such as liver, adipose tissue, and skeletal muscle also regulates circadian fluctuations in insulin sensitivity and glucose disposal, which are highest in the morning and decline toward the evening in humans. Liver-specific ablation of *Bmal1* in mice has revealed that the liver clock promotes gluconeogenesis, glycogenolysis, and mitochondrial oxidative metabolism in the sleep/fasting period, while promoting glycogen synthesis in the wake/feeding period. Muscle-specific *Bmal1*-deficient mice display reduced glucose tolerance, concomitant with lower levels of proteins involved in glucose uptake by muscle cells (e.g., the glucose transporter GLUT4). Ablation of the *Cry1* and *Cry2* repressors in the negative limb of the clock alters glucagon and glucocorticoid signaling in the liver, contributing to hyperglycemia and impaired glucose tolerance in these mutant mice. Together, these genetic studies in mice suggest a role for tissue-specific clocks in the partitioning of energy utilization across the sleep-wake cycle.

Importantly, peripheral clocks also interact with other environmental factors such as diet and time of feeding. For example, high-fat feeding leads not only to obesity and metabolic syndrome in mice, but also to perturbed clock gene expression across multiple peripheral tissues and a disrupted sleep-wake/fasting-feeding cycle, as revealed by increased activity and feeding during the daytime, the normal rest period in mice. Furthermore, mice that are fed a high-fat diet exclusively during their (inactive) light phase gain significantly more weight than mice that are fed the same diet during the dark period—the active period for mice. Additionally, the metabolic phenotypes arising from ad lib high-fat feeding can be significantly ameliorated by restricting the time of high-fat feeding exclusively to the dark period. Animals with disrupted clock throughout the hypothalamus and SCN exhibit mistimed eating and adverse metabolic rhythms that can be restored by dark-only feeding. Time-restricted feeding can also increase the activity of brown adipose tissue in mice and reduce hepatic glucose production to instead promote beta oxidation of fatty acids. The potential clinical utility of time-restricted eating (TRE) has been corroborated in human interventional studies. These have demonstrated that dietary interventions modulate transcriptional rhythms across tissues and that TRE can improve metabolic homeostasis as well as promote weight loss. Compared with calorie restriction, TRE has repeatedly been shown to promote weight loss by reducing calorie intake without the need to actively count calories. Time-restricted eating may also modulate the central regulation of sleep and hunger, as studies have found that humans who restrict their food intake to a shorter than ad lib period also consume fewer daily calories and report both lower hunger and improved sleep. In the setting of critical illness, nutrition is often provided at the incorrect phase of the light-dark cycle, and interventions to align feeding with environmental zeitgeibers may improve metabolic health. There is also some evidence that consuming a greater proportion of daily calories early compared with later in the day confers metabolic advantages, including weight loss. One contributing mechanism may be that diet-induced thermogenesis (energy expenditure elicited by food intake) is higher in the morning compared with evening and that daytime hunger ratings are lower when calories are preferentially consumed earlier versus later in the day.

Finally, animal studies have further shown that when the light-dark cycle is disrupted or when animals are subjected to conditions mimicking "jetlag" by artificially advancing or delaying the daily light period, there is desynchronization among circadian clocks and subsequent weight gain. Accumulating evidence in humans also finds that circadian misalignment both disrupts and desynchronizes circadian clocks across tissues. Clinical studies that have sampled tissues such as blood, skeletal muscle, and adipose tissue at regular intervals have observed disruptions in the day-night rhythms in clock and metabolic genes following sleep-wake interventions. Prolonged circadian misalignment using forced desynchrony protocols reduces insulin sensitivity in the pre- and postprandial states. Under such conditions, insulin secretion fails to suppress glucose levels, suggesting inadequate β-cell compensation. Moreover, resting metabolic rate declines significantly both in the awake and sleeping state, altogether providing potential explanations why shift work can increase the risk of obesity, type 2 diabetes, and the metabolic syndrome.

Human genetic association studies also support a role for clock genes in metabolic homeostasis and β-cell function. Carriers of a certain *BMAL1* polymorphism have a greater risk of developing type 2 diabetes, while *CLOCK* variants have been found to interact with diet, such that variants can have a protective effect on insulin sensitivity in individuals with high monounsaturated fat intake or in individuals provided a low-fat diet. In contrast, the minor allele of another *CLOCK* gene variant has been associated with increased waist circumference, but only in those with high saturated fat intake. Similarly, *NPAS2* and *BMAL1* variants have been associated with a greater risk of hypertension. Melatonin receptor *MTNR1B* gene variants that result in increased expression of MTNR1B have been associated with elevated fasting blood glucose levels and reduced insulin secretion irrespective of their level of glycemic control, although how melatonin regulates glucose homeostasis remains incompletely understood. These association studies highlight the role of the circadian system in metabolism, as well as the potential for interactions of external perturbations—such as circadian misalignment—with a protective or adverse genetic profile.

A large proportion of society recurrently shifts sleep-wake and eating times between working/nonfree days and free days. This social jetlag has been increasingly tied to metabolic disruptions, including a greater risk of obesity and type 2 diabetes. As this involves recurrent phase advances and phase delays—like shift work but often of smaller magnitude—it is possible that social jetlag, and often interlinked eating jetlag, also results in perturbed rhythms of energy expenditure in combination with disruptions to the circadian hunger drive, further increasing the risk of obesity. Repeated shifts in the food- and SCN-driven rhythm of insulin release may similarly over time increase the risk of type 2 diabetes. Shifted feeding rhythms in relation to the sleep-wake cycle and the timing of SCN activity may be causally involved in this pathogenesis. This is exemplified by the disorders known as night-eating syndrome and sleep-related eating disorder. In the former syndrome, a large part of daily calorie consumption occurs in the evening and nighttime hours, and this shifted meal pattern has been associated with a delayed timing of the internal clock. Some evidence exists that these syndromes are associated with obesity. Individuals who report sleeping fewer hours or who are subjected to restricted sleep for a few consecutive days have also been found to consume more calories, especially later in the evening, a period during which prolonged fasting favors oxidative fuel utilization. As such, this may explain why sleep restriction increases the risk of obesity. These associations have also been observed in individuals with later onset of sleep, i.e., evening chronotypes. Night-eating syndrome and later chronotypes have also been linked to type 2 diabetes and may be more common than other eating disorders such as binge-eating disorder. Both conditions have also been found to be associated with impaired glycemic control—such as a greater likelihood of hemoglobin A_{1c} values exceeding 7%—in patients already suffering from type 2 diabetes. This emphasizes how proper alignment of internal circadian rhythms with external factors are key contributing factors for long-term metabolic homeostasis.

Circadian Clocks in Relation to Brain Health and Cognition Molecular circadian clocks are present not only within the extra-SCN regions of the brain, but also in neurons, astrocytes, microglia, and cells of the blood-brain barrier. Emphasizing the functional significance of properly aligned clocks for brain health, shift workers have been found to have decreased gray matter in brain regions involved in memory and executive functions, with more notable effects in individuals who had shorter recovery periods between the onset of each shift work cycle. Adults performing rotating shift work for many years have also been shown to exhibit signs of accelerated cognitive aging. Notably, evidence suggests that these effects may be reversible, as those who stopped carrying out shift work exhibited normal cognitive performance 5 or more years later.

Studies have also uncovered an important role for perturbed circadian and sleep-wake rhythms in neurodegenerative conditions such as Alzheimer's disease (AD), Huntington's disease (HD), and Parkinson's disease (PD). Amyloid beta (Aβ), a key pathognomonic component of AD, normally exhibits circadian fluctuations in the extracellular space in the brain, as well as in the cerebrospinal fluid and plasma in humans, peaking during the active period and falling during sleep. Of note, these daily rhythms of Aβ accumulation are dampened in mice that are prone to develop AD; reduced plasma Aβ fluctuations have also been noted in older compared with younger individuals. Animal studies indicate that removal of Aβ (and other neurotoxic substances) during the nighttime sleep period is facilitated by a lymphatic-like system that relies on glial cells (the "glymphatic" system). Relevance of this system to humans is suggested by the observation that non–rapid eye movement (NREM) sleep is accompanied by hemodynamic fluctuations that alter the flow of cerebrospinal fluid, which can remove toxins such as Aβ. Consistent with a role for circadian rhythms in the pathogenesis of AD, ablation of core clock genes throughout the brain, within subregions of the brain or within glia, leads to pathology such as oxidative stress, neuronal cell death, and scarring of brain tissue (astrogliosis). Furthermore, perturbed light-dark cycles increased pathology associated with oxidative stress, and single nucleotide polymorphisms in *CLOCK* and *BMAL1* have been associated with increased risk of developing AD.

Evidence also indicates that the relationship between the circadian/sleep-wake system and AD is bidirectional. For example, patients suffering from AD exhibit several signs of perturbed circadian rhythms, the most prominent of such phenomena being "sundowning," whereby AD patients become more agitated and exhibit delirium-like symptoms in the afternoon or evening. Studies have furthermore indicated that in severe forms of AD, the circadian rhythm is phase delayed. Aged AD-prone mice also display perturbed sleep-wake patterns, which can be corrected by immunization against Aβ or by an orexin antagonist. Further research will help uncover the primary role of the circadian system in disease pathology, independent of the contribution from perturbed sleep, in conditions like AD. Notably, evidence suggests that interventions that increase daytime light exposure and that include melatonin supplementation ameliorate symptoms of AD, presumably by counteracting disrupted circadian rhythms.

Meta-analyses of cohort and longitudinal studies support an association between shift work and the risk of depression, with greater risk in women. This relationship is bidirectional, as disruption of sleep and circadian rhythms is a key feature of depression and multiple other neuropsychiatric conditions. Two important factors for why misaligned sleep increases the risk of neuropsychiatric conditions may be mistimed light exposure and disrupted 24-h rest-activity rhythms. In cross-sectional and longitudinal analyses, greater time spent outdoors during the day has been associated with fewer symptoms of insomnia, lower risk of developing depression, and less need for antidepressive medication. Similarly, decreased rest-activity rhythms are associated with lower subjective happiness and reaction time and a greater lifetime risk of major depressive or bipolar disorder. This may be partly due to genetic variation, as clock genes have been implicated in depression and mood both in human and genetic animal studies. Polymorphisms in genes that regulate sleep and circadian rhythms—for instance, a long gene variant of *PER3*—have also been linked to bipolar disorder and schizophrenia, while *CRY2* and *CLOCK* gene polymorphisms are associated with seasonal affective disorder, a type of depression arising in the fall and winter months when the levels of sunlight are lowest. Bipolar disorder is furthermore often triggered by circadian disruptions or curtailed sleep. Both bipolar disorder and schizophrenia have been linked to various forms of circadian disruption following disease onset, and a critical component of disease treatment often involves normalizing sleep and sleep-wake rhythms.

Sleep deprivation by itself is known to reduce alertness, impair decision-making, and increase risk for accidents—after 18–24 h of continuous wakefulness, several skills exhibit the same degree of decline as following mild alcohol intoxication. However, cognitive abilities may suffer even further when sleep restriction is combined with circadian misalignment as in shift work. In one study, participants were subjected to ~43-h long days in parallel with reduced sleep (equivalent to 5.6 h of sleep in a 24-h period), yielding a forced desynchrony protocol coupled with sleep loss. When subjects were tested at the nadir of their circadian period, the subjects' reaction speed dropped almost by an order of magnitude compared with controls. In another study, researchers noted almost a 36% greater incidence of serious medical errors in resident interns who regularly worked 24-h or longer shifts compared with those who were randomly assigned to work up to 16-h-long shifts. Furthermore, errors that resulted in patient death were three times more likely to occur in residents working extended hours compared with those who only worked up to 16-h-long shifts.

Circadian Regulation of Gastrointestinal Homeostasis and the Microbiota Physiologic aspects of the gastrointestinal (GI) tract exhibit day-night variations that anticipate and prepare for food intake and digestion during the active period. Gastric emptying and colonic motility are considerably greater during the active phase, as the phasic motor program supporting movement of digested material along the intestine is approximately twice as fast during the day compared with night. Bile acid secretion also exhibits circadian rhythmicity in the intestine, as do absorption and the expression of many nutrient uptake transporters in the intestinal wall, including the main glucose transporter protein SGLT1. The permeability of the intestinal wall also varies throughout the sleep-wake cycle, and mice exposed to chronic sleep fragmentation exhibit increased intestinal permeability, which may enable inflammatory molecules from bacteria to reach the systemic circulation.

The composition and function of the fecal microbiome (i.e., the gut microbiota) also display circadian rhythmicity, orchestrated by both host circadian clock gene expression and food intake rhythms. Accordingly, circadian disruption, either by environmental or genetic means, perturbs these microbial rhythms, disrupting both bacterial levels and the metabolic functions of the gut microbiota. For example, alterations in the expression and functions of the gut microbiota have been noted in humans exposed to acute jetlag, and evidence suggests that curtailing sleep, which often accompanies shift work and jetlag, can alter the gut microbiota. Corroborating the importance of the daily timing of food intake, interventions with meals scheduled earlier versus later in the day, or that involve time-restricted eating, have been found to alter the composition of the gut microbiome in humans, although the causal relevance of this remains to be ascertained.

By increasing local and systemic inflammation, circadian disruption of the gut microbiota may be causally involved in increased risk of inflammatory bowel disease (Crohn's disease and ulcerative colitis) and colon cancer in shift workers. Biological sex differences have also been reported, as female mice display more pronounced microbial rhythms. Interestingly, the gut microbiome has also been shown to influence the rhythms of host tissues, such as the intestine and liver, that also appear sex-specific. This relationship indicates that a bidirectional interaction exists between tissues that regulate metabolic processes and the gut microbiome across the sleep-wake cycle. These findings may furthermore have clinical implications, given that the gut microbiome may both directly (in the gut lumen) and indirectly (through host-microbiota interactions such as through signaling molecules) impact metabolic responses and pharmacokinetic and pharmacodynamic properties of therapeutic drugs across the 24-h day-night cycle.

Cardiovascular Health and the Circadian Clock An early epidemiologic observation was a greater incidence of myocardial infarction in the morning hours, with the lowest risk during the period preceding sleep. Other cardiovascular outcomes such as sudden cardiac death and syncope also exhibit a daily peak in the morning. Blood pressure (BP) typically peaks around 2100 h and decreases later during sleep. The postexercise recovery response of BP is faster in the late afternoon compared with the morning, and the daily timing of physical activity has been found to modulate the risk of all-cause and cardiovascular disease mortality. The lowering of BP during sleep is partially due to a circadian nighttime dip of around 3–6 mmHg in systolic BP (SBP) and 2–3 mmHg in diastolic BP (DBP). A dip in BP of either <10% or >20% during normal sleep has been associated with worse cardiovascular prognosis and risk of dementia. Nighttime BP dipping is also often disrupted in sleep-wake disorders and correlates with increased cardiovascular disease risk in conditions such as insomnia and narcolepsy. Conversely, specifically lowering nighttime BP has been found to confer a lower prospective risk of cardiovascular disease.

In addition to BP, heart rate also typically decreases during sleep, while mistimed sleep leads to higher heart rate during the sleep period. Studies also suggest that heart muscle may be more tolerant to hypoxia and thus fare better under surgery scheduled for the afternoon due to timing of cellular programs driven by the cell autonomous clock in cardiomyocytes. Thus, a combination of factors—which may also involve altered glucocorticoid levels and increased platelet aggregation—may contribute to a greater risk of cardiovascular disease in the morning. Subsequent epidemiologic studies also have demonstrated that shift work increases the risk of dyslipidemia and hypertension, as well as the risk of coronary heart disease, including myocardial infarction. These findings are in line with interventional findings in which circadian misalignment has been induced either by inverting the sleep-wake cycle or by imposing days that are far outside what the endogenous circadian clock can adapt to (i.e., either too short [e.g., 20-h] or too long [e.g., 28-h]). These studies in healthy human subjects have found that circadian misalignment elevates 24-h BP, particularly during sleep. These changes may be causally related to how the autonomic system is regulated during sleep, as evidenced by reduced vagal cardiac control when the sleep-wake cycle is inverted.

Circadian Disruption and Cancer In 2007, the International Agency for Research on Cancer (part of the World Health Organization) declared that shift work that involves circadian disruption is likely carcinogenic to humans. While evidence for an association between shift work and general cancer incidence is mixed, accruing evidence supports a link between shift work and increased risk of developing colon and breast cancer, as well as having a poorer cancer prognosis. Telomere shortening, a phenomenon in aging that destabilizes the genome, has also been observed in shift workers as well as in individuals suffering from short sleep. Such changes may reduce the ability of damaged or senescent cells to undergo apoptosis and, instead, lead to uninhibited cell growth and cancer. An indirect role for the circadian clock has also come from retrospective studies on how cancer risk is related to food timing and duration of the nighttime fast in humans. In combination with interventional studies on time-restricted feeding, these findings suggest that limiting food intake to a restricted period of the day, optimizes circadian processes thereby reducing the risk of potentially carcinogenic cell damage. Studies of recurring fasting have also shown that it lowers the risk and delays the onset of cancer.

Experimental genetic evidence has also implicated clock disruption as a factor in tumorigenesis. Genetic loss of *Per2* or *Bmal1* has been shown to promote lung tumorigenesis, while studies in *Per2* mutant mice have also revealed increased radiation-induced lymphoma associated with dysregulation of the cell cycle. However, disruption of the *Cry* gene in mice has also been implicated in tumor protection due to increased susceptibility to cell death. In contrast, pharmacologic overactivation of REV-ERB may impair growth of glioblastomas. While epidemiologic, experimental, and chronotherapeutic evidence (see section "Chronotherapy and Future Directions") suggests a link between circadian disruption and cancer, the precise role of circadian systems in tumorigenesis remains to be determined.

Circadian Regulation of the Immune System Circadian misalignment and sleep restriction both alter population levels of immune cells and decrease the ability of immune cells to produce reactive radicals, in part likely through disruption of cytokine rhythms. Chronic circadian disruption may thereby impair the immune system's ability to conduct immunosurveillance at the proper time of day. This may reduce the ability to mount an appropriate pathogen-induced effector (cytotoxic T-cell) response during the active period, as well as impair the more long-term adaptive immune response, which is favored by the cytokine milieu (e.g., surges in prolactin and GH) that accompanies the recovery/sleep phase. Instead, circadian misalignment increases a range of clinically used inflammatory markers (e.g., C-reactive protein, tumor necrosis factor α, and interleukin 6), and such changes have been noted even when the sleep-wake cycle is only prolonged to a slightly longer than normal 24.6-h day. While similar effects are also observed following acute total sleep deprivation or recurrent partial sleep restriction, circadian misalignment has been found to promote an even more pronounced elevation of such markers. Genetic clock disruption in peritoneal macrophages has also revealed clock control of Toll-like receptor 9, which is responsible for identifying molecules from foreign pathogens. *Clock* knockout mice also have reduced T-cell antigen response, and mice immunized during the day had a stronger T-cell response than mice immunized at night, supporting regulation of the immune system by the clock. Similar mechanisms likely take place in humans, as clinical studies have noted an impaired vaccine response following sleep disruption, and several studies show improved immunogenic response to various antigens when vaccinated in the morning compared with afternoon.

Aging and the Circadian Clock Instability in the clock system is an often-overlooked hallmark of aging. Aging is associated with a decline in the robustness of intrinsic rhythmic processes at the behavioral, physiologic, and molecular levels in both human and animal models. At the behavioral level, aging leads to reduced and fragmented sleep, dampened locomotor activity and feeding rhythms, and a reduced ability to entrain to light, as old rodents are 20 times less sensitive to the entraining effects of light relative to young animals. Even middle-aged individuals exposed to jetlag exhibit more symptoms of circadian misalignment, such as increased time awake and reduced alertness, compared with young individuals. On a physiologic level, some of the hallmarks of aging are a reduction in amplitude (e.g., flattening of circadian pattern) of circadian processes, which can also be seen at the cellular level in peripheral cells isolated from older compared with younger individuals. This dampening of rhythms also impacts the circadian signal during the evening period (the wake maintenance zone). Epidemiologic evidence indicates that a dampened rest-activity amplitude is associated with an increased prospective risk of a range of common health conditions, such as dementia, CVD, cancer, and all-cause mortality.

Aging also results in a phase advance (e.g., a shift in the timing of the peak or nadir) in rhythms of the endocrine and neuroendocrine systems, including sleep onset and offset. For example, cortisol, dehydroepiandrosterone (DHEA), and melatonin all have dampened rhythms and are phase advanced in aging; the combination of such changes may, for instance, contribute to more fragmented sleep and lower levels of restorative slow-wave sleep in aged individuals. Relatedly, aging results in reduced peptide expression in the SCN (VIP and AVP), cell loss in sleep-wake regions (including the SCN), and reduced amplitude of rhythms of SCN electrical activity. Further, while the SCN-dependent body temperature rhythm—a generally accepted marker for the integrity of circadian rhythms—peaks in the evening and is lowest in the early morning in young individuals, aged healthy subjects display a phase advance and a decrease in circadian amplitude in body temperature rhythms. Indeed, evidence suggests that internal desynchrony between core body temperature rhythms and the sleep-wake cycle may contribute to age-associated circadian alterations.

On a molecular level, aging is associated with decreased expression and altered diurnal profiles of several of the core clock genes, including *Clock* and *Bmal1*, within both SCN and peripheral tissues such as heart and liver. The acute induction of *Per1* in response to light was markedly reduced in the SCN of aged mice compared with young mice, potentially contributing to their delayed response to light entrainment. Mice lacking *Bmal1* die prematurely compared with control mice, consistent with premature accumulation of reactive oxygen species. These mice have an accelerated onset of numerous age-related pathologies, including cataracts, sarcopenia, reduced organ size, and decreased hair growth. Instead, deficiency of cryptochrome, a repressor of the core clock repressor, has been associated with alterations in liver regeneration, while BMAL1 and PER2 may be important for proper neurogenesis in the hippocampus, a brain region in which adult mammals normally exhibit continuous cell division. Altogether, this suggests that the highly conserved circadian clock is important for regulating a wide range of homeostatic processes, including cell-cycle pathways, which when properly phased to each other promote organismal fitness.

Shift workers have been found to exhibit molecular signs of accelerated aging, as measured by an accelerated DNA methylation clock. Measurements of altered circadian rhythms with age may serve as a useful biomarker for aging. An intriguing question is whether the decline in amplitude of rhythms correlates with a decline in function and, importantly, whether restoration of these rhythms with age, through either behavioral or pharmacologic intervention, would delay the aging process. Studies in mice indicate that behavioral and pharmacologic interventions (including exercise) can restore circadian oscillations in aging. Restoration of levels of the metabolite NAD$^+$, which are reduced with aging, in old mice by supplementation with the NAD$^+$ precursor nicotinamide riboside (NR) markedly restores rhythms of metabolic and stress response pathways, as well as late evening activity rhythms, that decline with aging through inhibition of the clock repressor PER2. Similarly, transplantation of the SCN from a young rat into an old rat "rescued" the rhythms of both locomotor activity and corticotropin hormone (CRH), suggesting that the SCN is an important target for age-related changes in clocks. Physical activity or targeted therapeutics may therefore ameliorate some of the circadian deterioration in aged humans.

CHRONOTHERAPY AND FUTURE DIRECTIONS

Chronopharmacology, also known as chronotherapy or circadian medicine, is a rapidly emerging field that studies how the timing of drug administration may impact its effectiveness. Since physiologic processes vary across the day, the timing of administration of medication may help optimize patient care. For example, since endogenous cholesterol synthesis is rhythmic in liver and peaks during the early morning hours, administration of statins (HMG-CoA reductase inhibitors) in the evening prior to bedtime has proven to be more effective than daytime administration at reducing low-density lipoprotein cholesterol (LDL-C) levels because the highest concentration of medication coincides with the peak in rhythmic endogenous cholesterol production. Given that BP exhibits a 24-h rhythm—being lowest during sleep—angiotensin-converting enzyme (ACE) inhibitors have been shown to be most effective at night to normalize the BP rhythms, restoring the nighttime dip in BP that is foremost tied to the occurrence of sleep. Numerous studies have also demonstrated that administration of cancer treatments at specific times of the day can increase chemotherapy effectiveness while also decreasing toxicity for a wide range of drugs. For example, 5-fluorouracil works best to treat colorectal cancer when administered at night, a time when the cancerous cells are more vulnerable while normal cells are quiescent and therefore less sensitive. Doxorubicin administration early in the morning to treat ovarian cancer has also been shown to be less toxic, as white blood cells recover faster than if the drug is given in the evening. Finally, the more severe morning symptoms of rheumatoid arthritis are linked to increased inflammation toward the evening; therefore, prevention of the nighttime upregulation of the immune/inflammatory reaction is more effective when glucocorticoids are administered with a nighttime release formulation.

Recognition of circadian rhythms is also critical for diagnoses and treatment of endocrine disorders. The diagnosis of Cushing's syndrome, which is characterized by hypercortisolemia, might be missed if the patient's cortisol levels are measured in the morning, when endogenous cortisol production peaks. Therefore, clinical diagnosis requires cortisol to be measured in the late evening when the levels of this hormone should typically be low. On the other hand, adrenal insufficiency is diagnosed by measuring cortisol in the morning when at its physiologic peak, and glucocorticoid therapy for these patients aims to mimic the endogenous rhythms of cortisol, as short-acting synthetic glucocorticoids are usually given several times a day in tapering doses, such that the largest amount is taken in the morning and the smallest in the evening.

Diabetes is another endocrine disorder intimately tied to circadian rhythms. Oral glucose tolerance, which is commonly used to diagnose diabetes, is worse in the afternoon and evening compared with the morning. This likely stems from greater daytime insulin sensitivity within peripheral tissues and reduced insulin secretion during the night. Similarly, due to a surge in hormone levels in the morning, diabetes patients may suffer from the dawn phenomenon (or dawn effect), an abnormally high morning increase in blood glucose due to impaired response in insulin secretion. A related phenomenon that can be tied to evening timing of insulin doses is the "rebound" or Somogyi effect. In this scenario, the initially noted clinical sign in the form of elevated glucose levels may be noted in the morning. However, the underlying cause is hypoglycemia occurring during the night, which produces a counterregulatory hormonal response that subsequently results in morning hyperglycemia. As patients with type 2 diabetes often have grossly impaired daily cycles of insulin secretion and glucose tolerance, this further highlights that time of day is an important consideration for the diagnosis and treatment of metabolic disorders such as type 2 diabetes. Another example of potential clinical relevance is how the pharmacokinetics of metformin—the most common treatment for type 2 diabetes—is significantly impacted by time of day due to rhythmicity in glomerular filtration rate and renal plasma flow. Notably, large interindividual variability in the pharmacokinetics seems to stem mostly from differences in chronotype, highlighting the need for patient-specific treatments dictated by circadian gene-environment interactions.

Continuous measurements of 24-h glucose have provided insight into sleep-wake regulation of glucose metabolism. Compared with daytime glucose levels, nighttime blood glucose levels have been found to more accurately predict a range of glucoregulatory parameters. Emerging evidence has also indicated that the daily timing of exercise may be an important determinant for more efficacious improvements in blood triglyceride and glucose levels. Furthermore, consideration of meal timing, particularly in the hospital setting, may impact patient health or responsiveness to treatments, as food in hospitals is often provided either continuously or just during the dark (rest) phase, with the latter being common in neonatal intensive care.

As our knowledge of the complexity of how circadian processes modulate physiology deepens, further advances to rationally develop new strategies for treatments of disorders affected by circadian misalignment are essential. For example, novel compounds have begun to emerge from unbiased drug discovery screens that in cell- and animal-based assays impact circadian clock components, either shortening or lengthening the period. These compounds include CRY stabilizers and various inhibitors of CKIδ, CKIε, and GSK-3. Pharmacologic control of the circadian cycle may be useful in the treatment of circadian disorders and metabolic disturbances with a circadian component. Understanding how the circadian clock controls biological functions will shed new light onto the pathogenesis of metabolic disorders with a circadian component, such as type 2 diabetes and metabolic syndrome, and will yield insight into how timing of drug delivery will impact patient care.

Acknowledgment

The authors would like to thank Billie Marcheva for her help with the figures and tables.

FURTHER READING

Allada R, Bass J: Circadian mechanisms in medicine. N Engl J Med 384:550, 2021.

Buxton OM et al: Adverse metabolic consequences in humans of prolonged sleep restriction combined with circadian disruption. Sci Transl Med 4:129ra43, 2012.

Cedernaes J et al: Transcriptional basis for rhythmic control of hunger and metabolism within the AgRP neuron. Cell Metab 29:1078, 2019.

Dibner C et al: The mammalian circadian timing system: Organization and coordination of central and peripheral clocks. Annu Rev Physiol 72:517, 2010.

Hatori M et al: Time-restricted feeding without reducing caloric intake prevents metabolic diseases in mice fed a high-fat diet. Cell Metab 15:848, 2012.

Hepler C et al: Time-restricted feeding mitigates obesity through adipocyte thermogenesis. Science 378:276, 2022.

Kervezee L et al: Metabolic and cardiovascular consequences of shift work: The role of circadian disruption and sleep disturbances. Eur J Neurosci 51:396, 2020.

Takahashi JS: Transcriptional architecture of the mammalian circadian clock. Nat Rev Genet 18:164, 2016.

499

Network Medicine and Multiomic Approaches to Health and Disease

Joseph Loscalzo

The field of human biology has progressed over the past three centuries largely as a result of the reductionist approach to the scientific problems that challenge the discipline. Biologists study the experimental response of a variable of interest in a cell or organism while holding all other variables constant. In this way, it is possible to dissect the individual components of a biologic system and assume that a thorough understanding of a specific component (e.g., an enzyme or a transcription factor) will provide sufficient insight to explain the global behavior of that system (e.g., a metabolic pathway or a gene network, respectively). Biologic systems are, however, much more complex than this approach assumes and manifest behaviors that frequently (if not invariably) cannot be predicted from knowledge of their component parts characterized in isolation. Growing recognition of this shortcoming of conventional biologic research has led to the development of the discipline of *systems biology*, which is defined as the holistic study of living organisms or their cellular or molecular network components to predict their response to perturbations. Concepts of systems biology can be applied readily to human disease and therapy and define the field of *systems pathobiology*, in which genetic or environmental perturbations produce disease and drug perturbations restore normal system behavior.

Systems biology evolved from the field of systems engineering in which a linked collection of component parts constitutes a network whose output the engineer wishes to predict. The simple example of an electronic circuit can be used to illustrate some basic systems engineering concepts. All the individual elements of the circuit—resistors, capacitors, transistors—have well-defined properties that can be characterized precisely. However, they can be linked (wired or configured) in a variety of ways, each of which yields a circuit whose response to a voltage applied across it is different from the response of every other configuration. To predict the circuit's (i.e., system's) behavior, the engineer must study its response to perturbation (e.g., voltage applied across it) holistically rather than its individual components' responses to that perturbation. Viewed another way, the resulting behavior of the system is greater than (or different from) the simple sum of its parts, and systems engineering utilizes rigorous mathematical approaches to predict these complex, often nonlinear, responses. By analogy to biological systems, one can reason that detailed knowledge of a single enzyme in a metabolic pathway or of a single transcription factor in a gene network will not provide sufficient detail in context to predict the output of that metabolic pathway or transcriptional network, respectively. Only a systems-based approach will suffice.

It has taken biologists a long time to appreciate the importance of systems approaches to biomedical problems. Reductionism has reigned supreme for many decades, largely because it is experimentally and analytically simpler than holism and because it has provided insights into biological mechanisms and disease pathogenesis that have led to successful therapies. However, reductionism cannot solve all biomedical problems. For example, the so-called off-target effects of new drugs that frequently limit their adoption likely reflect the failure of a drug to be studied in integrated context, that is, the failure to explore all possible actions aside from the principal target action for which it was developed. Other approaches to understanding biology are, therefore, clearly needed. With the growing body of genomic, proteomic, and metabolomic data sets in which dynamic changes in the expression of many genes and many metabolites are recorded after a perturbation and with the growth of rigorous mathematical approaches to analyzing those changes, the stage has been set for applying systems engineering principles to modern biology.

Physiologists historically have had more of a (bio)engineering perspective on the conduct of their studies and have been among the first systems biologists. Yet, with few exceptions, they, too, have focused on comparatively simple physiological systems that are tractable using conventional reductionist approaches. Efforts at integrative modeling of human physiological systems, as first attempted by Guyton for blood pressure regulation, represent one application of systems engineering to human biology. These dynamic physiological models often focus on the acute response of a measurable physiological parameter to a system perturbation, and do so from a classic analytical perspective in which all the conventional physiological determinants of the output parameter are known and can be modeled quantitatively.

Until recently, molecular systems analysis has been limited owing to inadequate knowledge of the molecular determinants of a biological system of interest. Although biochemists have approached metabolic pathways from a systems perspective for more than 50 years, their efforts have been limited by the inadequacy of key information for each enzyme (K_m, k_{cat}, and concentration) and substrate (concentration) in the pathway. With increasingly rich molecular data sets available for systems-based analyses, including genomic, transcriptomic, proteomic, metabolomic data, and enzymatic (kinetic), molecular biologists and biochemists are now poised to use systems biology approaches to explore biological and pathobiological phenomena.

PROPERTIES OF COMPLEX BIOLOGIC SYSTEMS

To understand how best to apply the principles of systems biology to human biomedicine, it is necessary to review briefly the building blocks of any biological system and the determinants of system complexity. All systems can be analyzed by defining their static topology (architecture) and their dynamic (i.e., time-dependent) response to perturbation. In the discussion that follows, system properties are described that derive from the consequences of topology (form) or dynamic response (function). Any system of interacting elements can be represented schematically as a *network* in which the individual elements are depicted as nodes and their connections are depicted as links. The nature of the links among nodes reflects the degree of complexity of the system. *Simple systems* are those in which the nodes are linearly linked with occasional feedback or feedforward loops modulating system throughput in highly predictable ways. By contrast, *complex systems* are nodes that are linked in more complicated, nonlinear networks; the behavior of these systems by definition is inherently more difficult to predict owing to the nature of the interacting links, the dependence of the system's behavior on its complex set of initial conditions, and the inability to measure the overall state of the system at any specific time with great precision. Complex systems can be depicted as a network of lower-complexity interacting components or modules, each of which can be reduced further to simpler analyzable canonical motifs (such as feedback and feedforward loops or negative and positive autoregulation); however, a central property of complex systems is that simplifying their structures by identifying and characterizing the individual nodes and links or even simpler substructures does not necessarily yield a predictable understanding of a system's behavior. Thus, the functioning system is greater than (or different from) the sum of its individual, tractable parts.

Defined in this way, most biological systems are complex systems that can be represented as networks whose behaviors are not readily predictable from simple reductionist principles. The nodes, for example, can be metabolites that are linked by the enzymes that cause their transformations, transcription factors that are linked by the genes whose expression they influence, or proteins in an interaction network that are linked by cofactors that facilitate interactions or by thermodynamic forces that facilitate their physical association. Biologic systems typically (but not invariably) are organized as *scale-free*, rather than stochastic, networks of nodes. Scale-free networks are those in which a few nodes have many links to other nodes (highly linked nodes, or hubs), but most nodes have only a few links (weakly linked nodes). The term *scale-free* refers to the fact that the connectivity of nodes in the network is invariant with respect to the size of the network. This is quite different from two other common network architectures: random (Poisson) and exponential distributions. Scale-free networks can be mathematically described by a power law that defines the probability of the number of links per node, $P(k) = k^{-\gamma}$, where k is the number of links per node and γ is the slope of the log P(k) versus log(k) plot; this unique property of most biologic networks is a reflection of their self-similarity or fractal nature **(Fig. 499-1)**.

There are unique properties of scale-free biological systems that reflect their evolution and promote their adaptability and survival. Biological networks likely evolved one node at a time in a process in which new nodes are more likely to link to a highly connected node than to a sparsely connected node. Furthermore, scale-free networks can become sparsely linked to one another, yielding more complex, *modular scale-free topologies*. This evolutionary growth of biologic networks has three important properties that affect system function and survival. First, this scale-free addition of new nodes promotes *system redundancy*, which minimizes the consequences of errors and accommodates adverse perturbations to the system robustly with minimal effects on critical functions (unless the highly connected nodes are the focus of the perturbation). Second, this resulting network redundancy provides a survival advantage to the system. In complex gene networks, for example, mutations or polymorphisms in weakly linked genes account for biodiversity and biologic variability without disrupting the critical functions of the system; only mutations in highly linked (*essential*) genes (hubs) can shut down the system and cause embryonic lethality. Third, scale-free biological systems facilitate the flow of information (e.g., metabolite flux) across the system compared with randomly organized biological systems; this so-called "small-world" property of the system (in which the clustered nature of the highly linked hubs defines a local neighborhood within the network that communicates through weaker, less frequent links to other clusters) minimizes the energy cost for the dynamic action of the system (e.g., minimizes the transition time between states in a metabolic network).

These basic organizing principles of complex biologic systems lead to three unique properties that require emphasis. First, biological systems are *robust*, which means that they are quite stable in response to most changes in external conditions or internal modification. Second, a corollary to the property of robustness is that complex biologic systems are *sloppy*, which means that they are insensitive to changes in external conditions or internal modification except under certain uncommon circumstances (i.e., when a hub is involved in the change). Third, complex biological systems exhibit *emergent properties*, which means that they manifest behaviors that cannot be predicted from the reductionist principles used to characterize their component parts. Examples of emergent behavior in biological systems include spontaneous, self-sustained oscillations in glycolysis; spiral and scroll waves of depolarization in cardiac tissue that cause reentrant arrhythmias; and self-organizing patterns in biochemical systems governed by diffusion and chemical reaction.

APPLICATIONS OF SYSTEMS BIOLOGY TO PATHOBIOLOGY

The principles of systems biology have been applied to complex pathologic processes with growing successes. The key to these applications is the identification of emergent properties of the system under study in

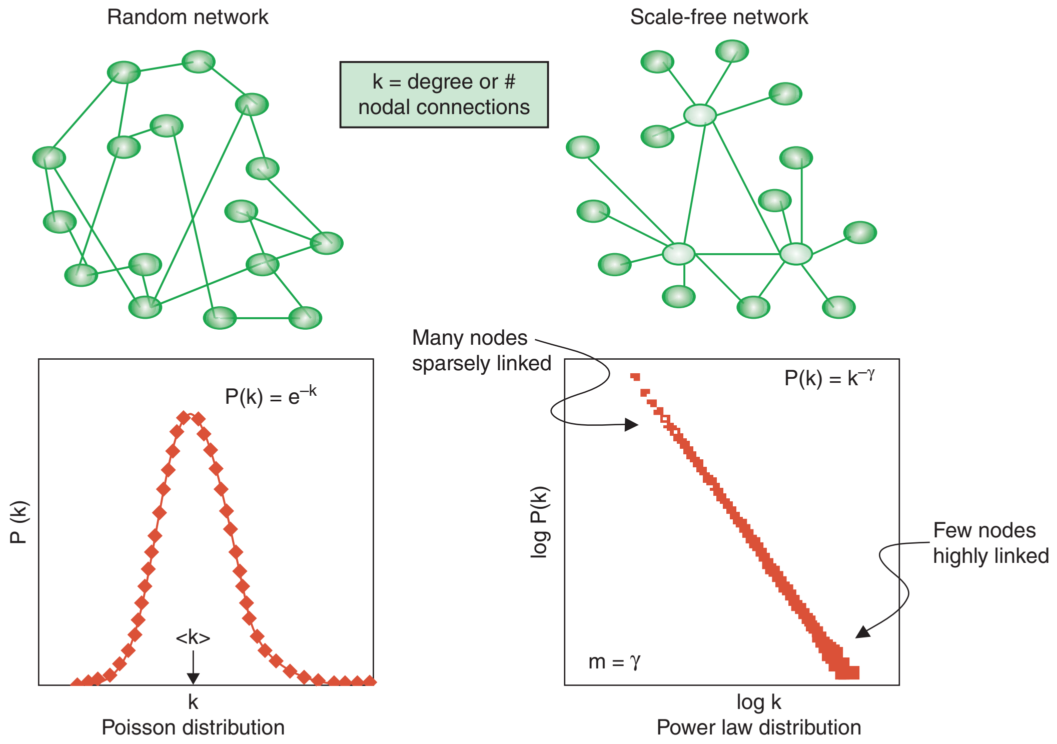

FIGURE 499-1 Network representations and their distributions. A random network is depicted on the left, and its Poisson distribution of the number of nodal connections (k) is shown in the graph below it. A scale-free network is depicted on the right, and its power law distribution of the number of nodal connections (k) is shown in the graph below it. Highly connected nodes (hubs) are lightly shaded.

order to define novel, otherwise unpredictable (i.e., from the reductionist perspective) methods for regulating the system's response. Systems biology approaches have been used to characterize epidemics and ways to control them, taking advantage of the scale-free properties of the network of infected individuals that constitute the epidemic. Through the use of a systems analysis of a neural protein-protein interaction network, unique disease-modifying proteins have been identified that are common to a wide range of cerebellar neurodegenerative disorders causing inherited ataxias. Systems analysis and disease network construction of a pulmonary arterial hypertension network led to the identification of a unique disease module involving a pathway governed by microRNA 21 and, more recently, to a novel profibrotic pathway that promotes vascular disease progression. Systems biology models have been used to dissect the dynamics of the inflammatory response using oscillatory changes in the transcription factor nuclear factor (NF)-κB as the system output. Systems biology principles also have been used to predict the development of an idiotypy–anti-idiotypy antibody network, describe the dynamics of species growth in microbial biofilms, and analyze the innate immune response. In each of these examples, a systems (patho)biology approach provided insights into the behavior of these complex systems that could not have been recognized with conventional scientific reductionism.

A unique application of systems biology to biomedicine is in the area of drug development. Conventional drug development involves identifying a potential target protein and then designing or screening compounds to identify those that inhibit (or activate) the function of that target. This reductionist analysis has identified many potential drug targets and drugs, yet only when a drug is tested in animal models or humans are the systems consequences of the drug's action revealed; not uncommonly, so-called off-target effects may become apparent and be sufficiently adverse for researchers to cease development of the agent. In addition to these types of unexpected outcomes exerted through pathways that were not considered *ab initio*, conventional approaches to drug development typically do not take into consideration the possibility of emergent behaviors of the organism or the metabolic pathway or the transcriptional network of interest that may be influenced by the drug. Thus, a systems-based analysis of potential drugs (drug-target network analysis) can benefit the development paradigm both by enhancing the likelihood that a compound of interest will not manifest unforeseen adverse effects and by promoting novel analytical methods for identifying unique control points or pathways in metabolic or genetic networks that would benefit from drug-based modulation, including the use of drug combinations. In addition and importantly, unbiased knowledge about the systems behavior of a drug can lead to rational approaches to drug repurposing.

SYSTEMS PATHOBIOLOGY AND HUMAN DISEASE CLASSIFICATION: NETWORK MEDICINE

Perhaps most important, systems pathobiology can be used to revise and refine the definition of human disease. The classification of human disease used in this and all medical textbooks derives from the correlation between pathological analysis and clinical syndromes that began in the nineteenth century. Although this approach has been very successful, serving as the basis for the development of many effective therapies, it has major shortcomings. Those shortcomings include a lack of sensitivity in defining preclinical disease, a primary focus on overtly manifest disease, failure to recognize different and potentially differentiable causes of common late-stage pathophenotypes (convergent pathophenotypes), failure to recognize common causes of different pathophenotypes (divergent pathophenotypes), and a limited ability to incorporate the growing body of molecular and genetic determinants of pathophenotype into the conventional classification scheme.

Two examples will illustrate the weakness of simple correlation analyses grounded in the reductionist principle of simplification (Occam's razor) used to define human disease. Sickle cell anemia, the "classic" Mendelian disorder, is caused by a Val6Gln substitution in the β chain of hemoglobin. If conventional genetic teaching holds, this single mutation should lead to a single phenotype in patients who harbor it (genotype-phenotype correlation). This assumption is, however, false, as patients with sickle cell disease manifest a variety of pathophenotypes (divergent pathophenotypes), including hemolytic anemia, stroke, acute chest syndrome, bony infarction, and painful crisis, as well as an overtly normal phenotype. The reasons for these different phenotypic presentations include the presence of disease-modifying genes or gene products (e.g., hemoglobin F, hemoglobin C, glucose-6-phosphate dehydrogenase), (stochastic) exposure to adverse environmental factors (e.g., hypoxia, dehydration), and the genetic and environmental determinants of common intermediate pathophenotypes or endophenotypes (i.e., variations in those generic pathological mechanisms underlying all human disease—inflammation, thrombosis/hemorrhage, fibrosis, cell proliferation, apoptosis/necrosis, immune response, etc.).

A second example of note is familial pulmonary arterial hypertension. This disorder is associated with well over 300 different mutations in three members of the transforming growth factor β (TGF-β) superfamily alone—bone morphogenetic protein receptor-2 (BMPR-2), activin receptor-like kinase-1 (Alk-1), and endoglin—as well as 16 other genes. All of these different genotypes are associated with a common pathophenotype, and each leads to that pathophenotype by molecular mechanisms that range from haploinsufficiency to dominant negative effects. As only approximately one-fourth of individuals in families that harbor these mutations manifest the pathophenotype, other disease-modifying genes (e.g., the serotonin receptor 5-HT2B, the serotonin transporter 5-HTT), genomic and environmental determinants of common intermediate pathophenotypes, and environmental exposures (e.g., hypoxia, infective agents [HIV], anorexigens) probably account for the incomplete penetrance of the disorder.

On the basis of these and many other related examples, one can approach human disease from a systems pathobiology perspective in which each "disease" can be depicted as a network that includes the following modules: the primary disease-determining elements of the genome (or proteome, if posttranslationally modified), the disease-modifying elements of the genome or proteome, environmental determinants, and genomic and environmental determinants of the generic intermediate pathophenotypes (endophenotypes). **Figure 499-2** graphically depicts these genotype-phenotype relationships as modules for the six common disease types with specific examples for each type. **Figure 499-3** shows a network-based depiction of sickle cell disease using this type of modular approach.

Goh and colleagues developed the concept of a human disease network **(Fig. 499-4)** in which they used a systems approach to characterize the disease-gene associations listed in the Online Mendelian Inheritance in Man database. Their analysis showed that genes linked to similar disorders are more likely to have products that physically associate and greater similarity between their transcription profiles than do genes not associated with similar disorders. In addition, proteins associated with the same pathophenotype are significantly more likely to interact with one another than with other proteins not associated with the pathophenotype. Finally, these authors showed that the great majority of disease-associated genes are not highly connected genes (i.e., not hubs) and are typically weakly linked nodes within the functional periphery of the network in which they operate. This set of observations led to a rigorous analysis by Menche and colleagues of 299 diseases whose associated proteins were found to cluster in discrete subnetworks or disease molecules within the global protein-protein interaction network.

This type of analysis validates the potential importance of defining disease on the basis of its systems pathobiological determinants. Clearly, doing so will require a more careful dissection of the molecular elements in the relevant pathways (i.e., more precise molecular pathophenotyping), less reliance on overt (common) manifestations of disease for their classification, and an understanding of the dynamics (not just the static architecture) of the pathobiological networks that underlie pathophenotypes defined in this way. **Figure 499-5** illustrates the elements of a molecular network within which a disease module is contained. This network is first identified by determining

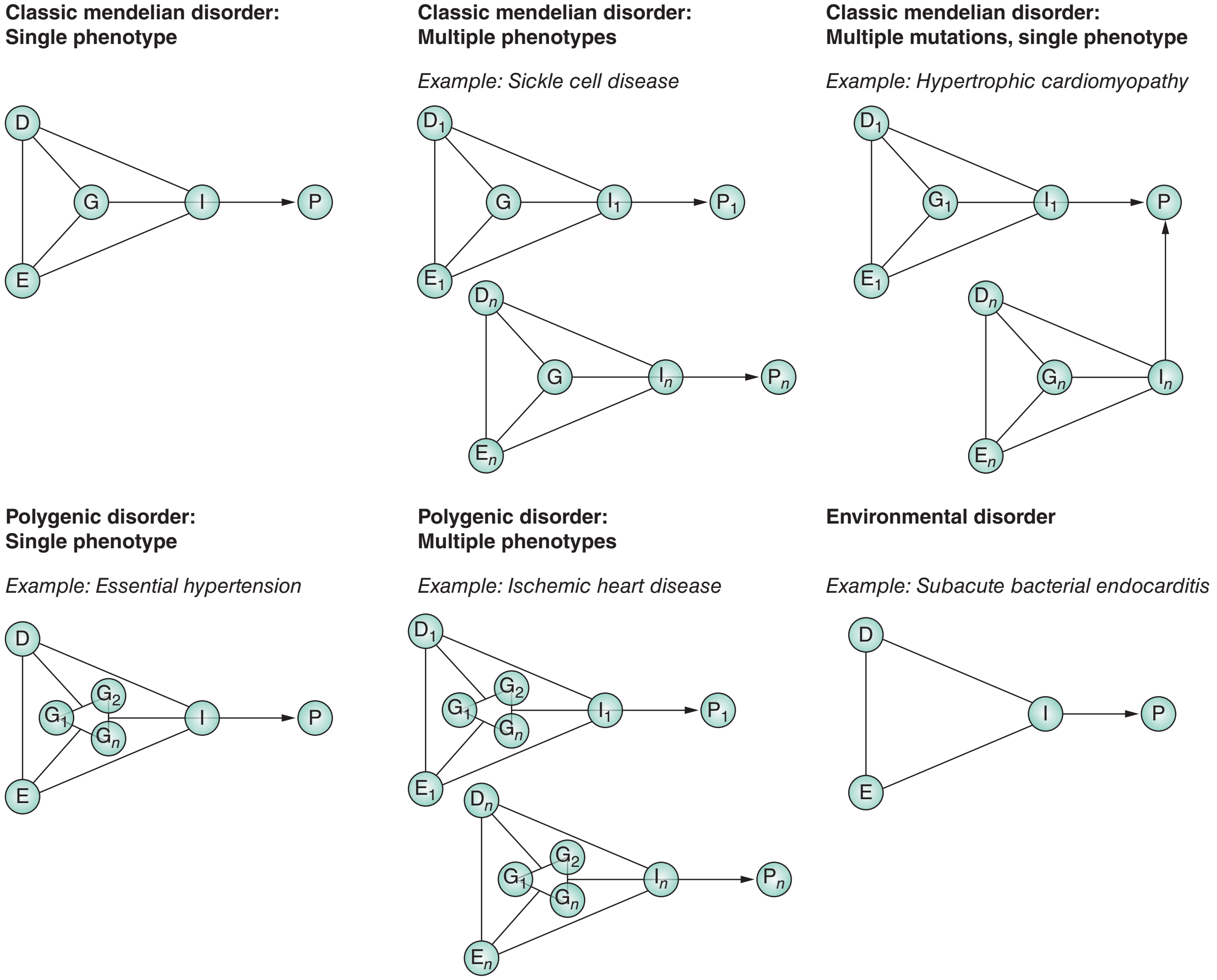
Classic mendelian disorder: Single phenotype
D
G
I
P
E
Classic mendelian disorder: Multiple phenotypes
Example: Sickle cell disease
D1
G
I1
P1
E1
Dn
G
In
Pn
En
Classic mendelian disorder: Multiple mutations, single phenotype
Example: Hypertrophic cardiomyopathy
D1
G1
I1
P
E1
Dn
Gn
In
En
Polygenic disorder: Single phenotype
Example: Essential hypertension
D
G2
G1
Gn
I
P
E
Polygenic disorder: Multiple phenotypes
Example: Ischemic heart disease
D1
G2
G1
Gn
I1
P1
E1
Dn
G2
G1
Gn
In
Pn
En
Environmental disorder
Example: Subacute bacterial endocarditis
D
I
P
E

FIGURE 499-2 Examples of modular representations of human disease. D, secondary human disease genome or proteome; E, environmental determinants; G, primary human disease genome or proteome; I, intermediate phenotype (endophenotype); P, pathophenotype. *(Reproduced with permission from J Loscalzo et al: Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. Mol Syst Biol 3:124, 2007.)*

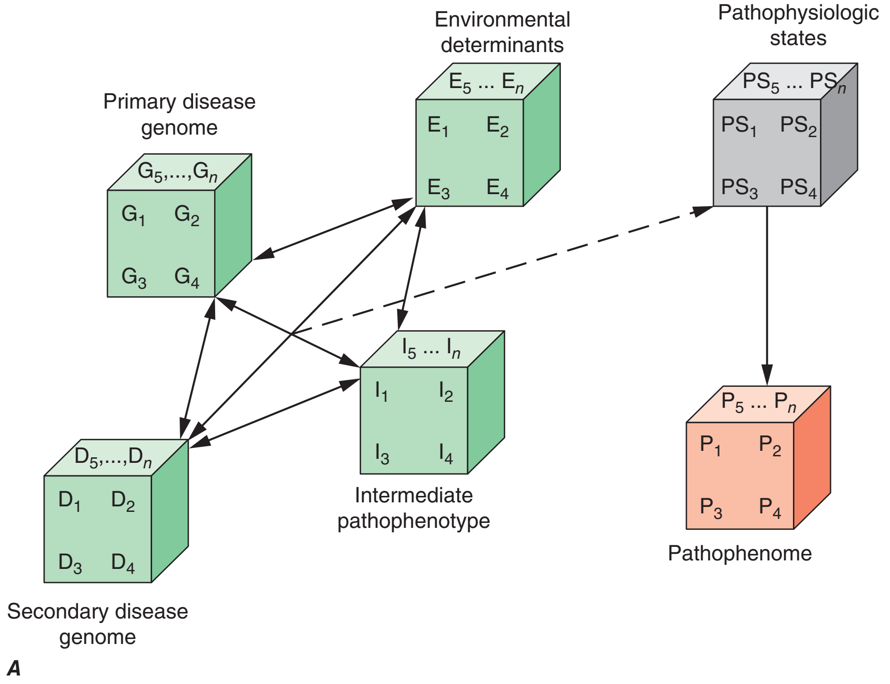
Environmental determinants
E5 ... En
E1 E2
E3 E4
Pathophysiologic states
PS5 ... PSn
PS1 PS2
PS3 PS4
Primary disease genome
G5,...,Gn
G1 G2
G3 G4
I5 ... In
I1 I2
I3 I4
Intermediate pathophenotype
P5 ... Pn
P1 P2
P3 P4
Pathophenome
D5,...,Dn
D1 D2
D3 D4
Secondary disease genome
A

FIGURE 499-3 ***A.*** Theoretical human disease network illustrating the relationships among genetic and environmental determinants of the pathophenotypes. Key: D, secondary disease genome or proteome; E, environmental determinants; G, primary disease genome or proteome; I, intermediate phenotype (endophenotype); PS, pathophysiologic states leading to P, pathophenotype. ***B.*** Example of this theoretical construct applied to sickle cell disease. Key: Red, primary molecular abnormality; gray, disease-modifying genes; yellow, intermediate phenotypes; green, environmental determinants; blue, pathophenotypes. *(Reproduced with permission from J Loscalzo et al: Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. Mol Syst Biol 3:124, 2007.)*

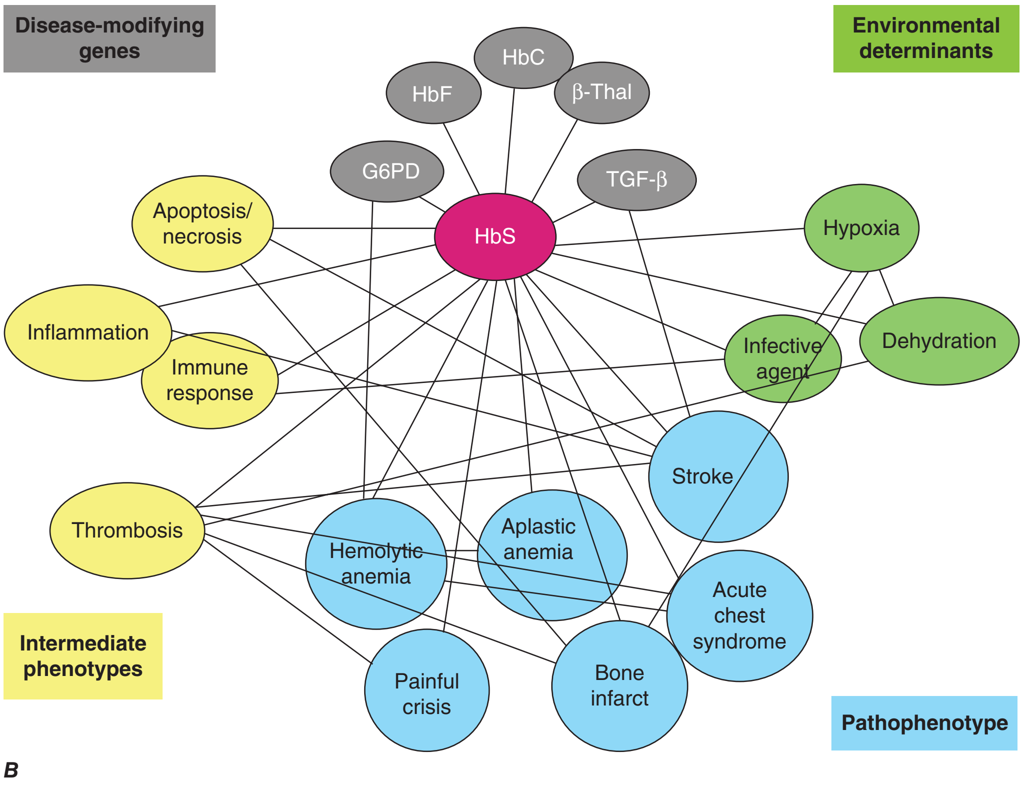
Disease-modifying genes
Environmental determinants
HbC
HbF
β-Thal
G6PD
TGF-β
HbS
Apoptosis/ necrosis
Hypoxia
Inflammation
Immune response
Infective agent
Dehydration
Stroke
Thrombosis
Aplastic anemia
Hemolytic anemia
Acute chest syndrome
Intermediate phenotypes
Painful crisis
Bone infarct
Pathophenotype
B

FIGURE 499-3 (*Continued*)

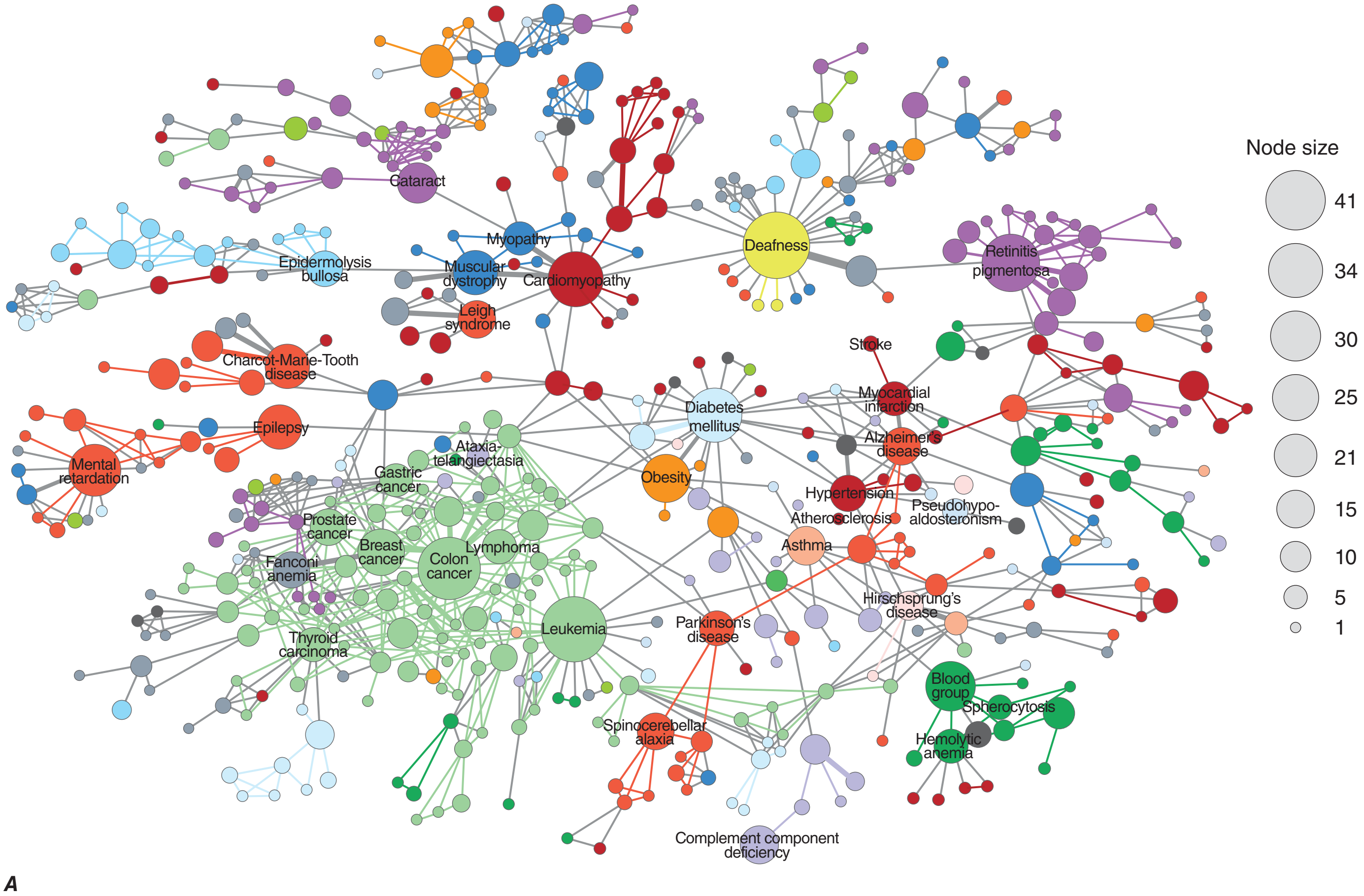
Cataract
Myopathy
Epidermolysis bullosa
Muscular dystrophy
Cardiomyopathy
Deafness
Retinitis pigmentosa
Leigh syndrome
Charcot-Marie-Tooth disease
Stroke
Myocardial infarction
Epilepsy
Mental retardation
Diabetes mellitus
Alzheimer's disease
Ataxia-telangiectasia
Gastric cancer
Obesity
Hypertension
Pseudohypo-aldosteronism
Atherosclerosis
Prostate cancer
Breast cancer
Lymphoma
Asthma
Fanconi anemia
Colon cancer
Hirschsprung's disease
Thyroid carcinoma
Leukemia
Parkinson's disease
Blood group
Spherocytosis
Spinocerebellar ataxia
Hemolytic anemia
Complement component deficiency
A
Node size
41
34
30
25
21
15
10
5
1

FIGURE 499-4 ***A.*** Human disease network. Each node corresponds to a specific disorder colored by class (22 classes, shown in the key to ***B***). The size of each node is proportional to the number of genes contributing to the disorder. Edges between disorders in the same disorder class are colored with the same (lighter) color, and edges connecting different disorder classes are colored gray, with the thickness of the edge proportional to the number of genes shared by the disorders connected by it. ***B.*** Disease gene network. Each node is a single gene, and any two genes are connected if implicated in the same disorder. In this network map, the size of each node is proportional to the number of specific disorders in which the gene is implicated. *(Reproduced with permission from KI Goh et al: The human disease network. Proc Natl Acad Sci USA 104:8685, 2007. Copyright (2007) National Academy of Sciences, U.S.A.)*

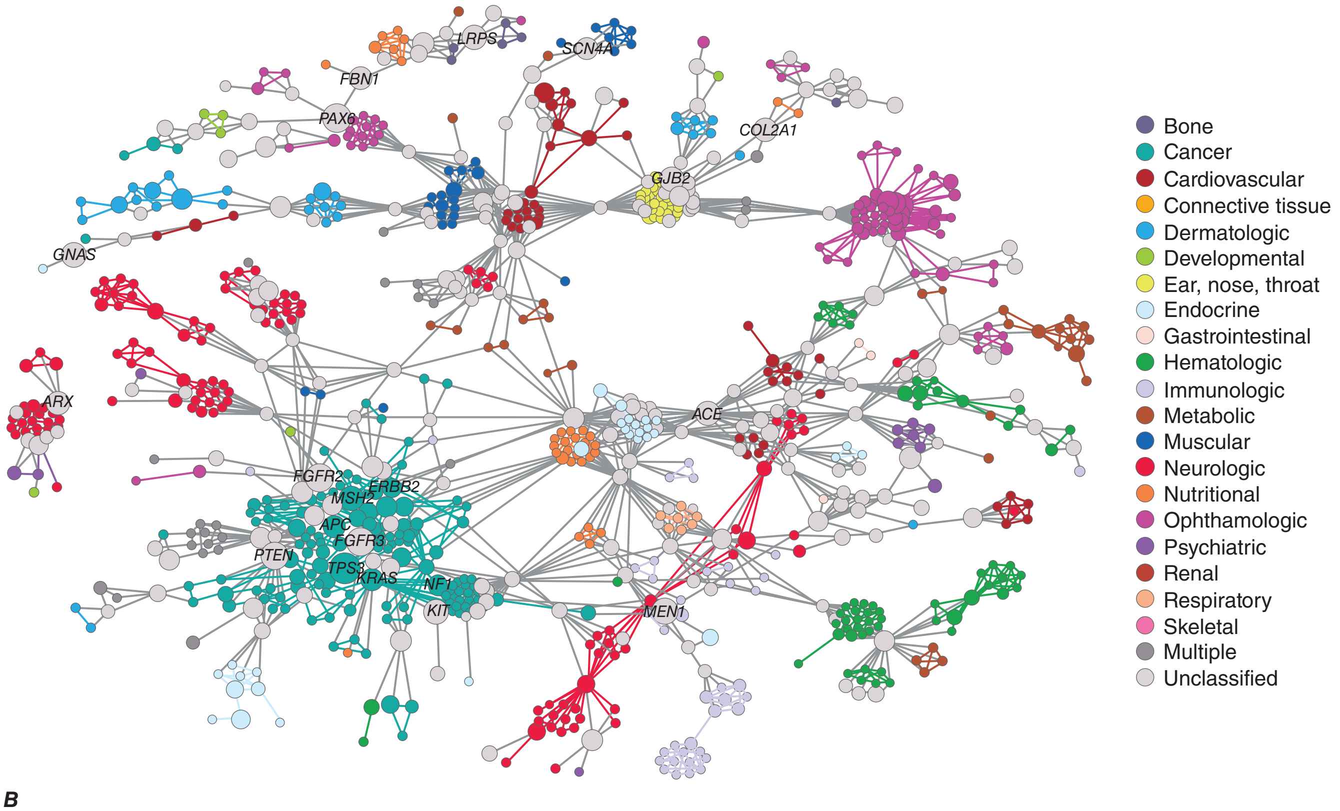
LRPS
SCN4A
FBN1
PAX6
COL2A1
GJB2
GNAS
ARX
ACE
FGFR2
ERBB2
MSH2
APC
FGFR3
PTEN
TPS3
KRAS
NF1
KIT
MEN1
B
Bone
Cancer
Cardiovascular
Connective tissue
Dermatologic
Developmental
Ear, nose, throat
Endocrine
Gastrointestinal
Hematologic
Immunologic
Metabolic
Muscular
Neurologic
Nutritional
Ophthamologic
Psychiatric
Renal
Respiratory
Skeletal
Multiple
Unclassified

FIGURE 499-4 (*Continued*)

the interactions (physical or regulatory) among the proteins or genes that comprise it (the "interactome"). These interactions then define a topologic module within which exists functional modules (pathways) and disease modules. One approach to constructing this module is illustrated in **Fig. 499-6**. Examples of the use of this approach in defining novel determinants of disease are given in **Table 499-1**, and follow from four key organizing principles that tie the interactions to human disease **(Table 499-2)**.

Interaction network complexity is itself becoming increasingly complex as multiple omic layers can now be ascertained, including the transcriptome, the proteome, the epigenome, and the metabolome. This integrative analysis leads to the formulation of a network of networks, the primary consequence of which from an evolutionary perspective is to enhance the robustness of the system to perturbations. Single-cell omic analysis adds further nuance to organ-based network analysis, which offers even greater promise for understanding individual differences in disease drivers and manifestations. **Figure 499-7** illustrates these points schematically.

Principles of network and systems biology can also be applied to clinical phenotyping. Rather than continuing to approach disease from an inclusive perspective, precision systems analysis of the molecular basis for disease must be linked to nuanced, deep phenotyping. This strategy will lead to the identification of subgroups of patients with a disease, each of which can be studied for differences in their disease modules that account for their phenotypic differences.

As yet another potential consideration, one can argue that disease reflects the later-stage consequences of the predilection of an organ system to manifest a particular intermediate pathophenotype (endophenotype) in response to injury. This paradigm reflects a reverse causality view in which a disease is defined as a tendency to heightened inflammation, thrombosis, or fibrosis, for example, after an injurious perturbation. Where the process is manifest (i.e., the organ in which it

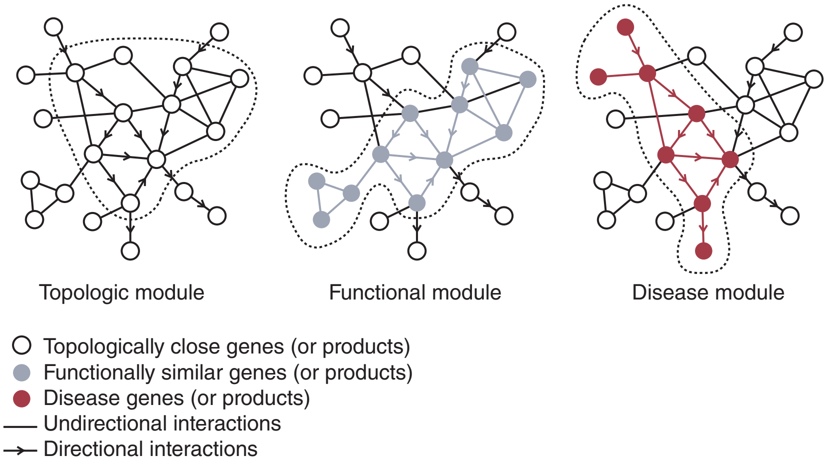
Topologic module
Functional module
Disease module
Topologically close genes (or products)
Functionally similar genes (or products)
Disease genes (or products)
Undirectional interactions
Directional interactions

FIGURE 499-5 The elements of the interactome. The interactome includes topologic modules (genes or gene products that are closely associated with one another through direct interactions), functional modules (genes or gene products that work together to define a pathway), and disease modules (genes or gene products that interact to yield a pathophenotype). *(Reproduced with permission from AL Barabási et al: Network medicine: A network-based approach to human disease. Nat Rev Genet 12:56, 2011.)*

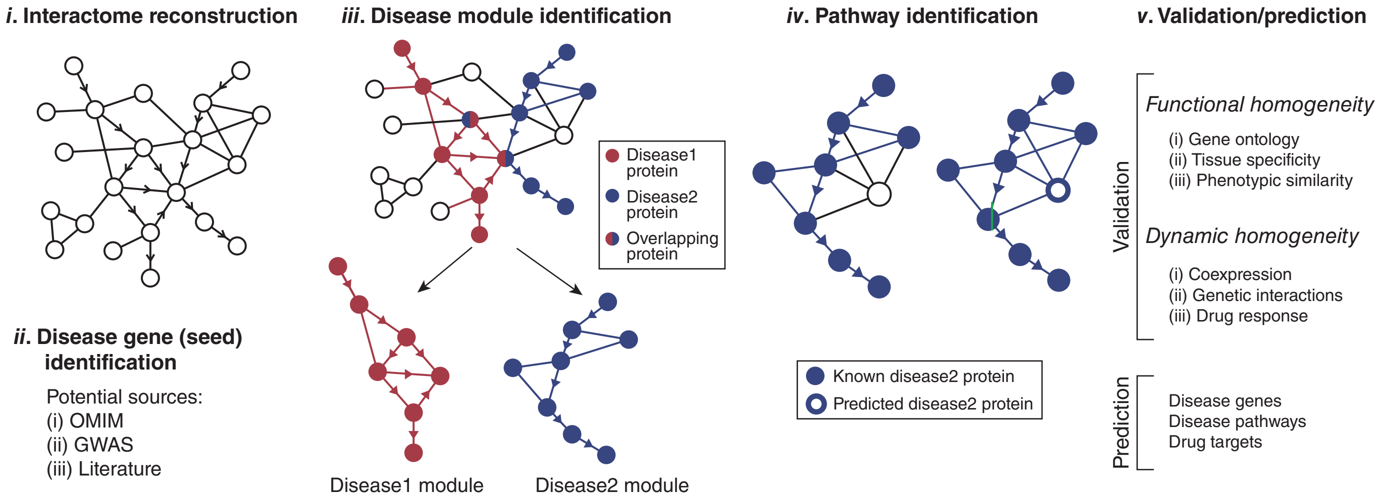
i. Interactome reconstruction
ii. Disease gene (seed) identification
Potential sources:
(i) OMIM
(ii) GWAS
(iii) Literature
iii. Disease module identification
Disease1 protein
Disease2 protein
Overlapping protein
Disease1 module
Disease2 module
iv. Pathway identification
Known disease2 protein
Predicted disease2 protein
v. Validation/prediction
Validation
Functional homogeneity
(i) Gene ontology
(ii) Tissue specificity
(iii) Phenotypic similarity
Dynamic homogeneity
(i) Coexpression
(ii) Genetic interactions
(iii) Drug response
Prediction
Disease genes
Disease pathways
Drug targets

FIGURE 499-6 Approaches to identifying disease modules in molecular networks. A strategy for defining disease modules involves (i) reconstructing the interactome; (ii) ascertaining potential seed (disease) genes from the curated literature, the Online Mendelian Inheritance in Man (OMIM) database, or genomic analyses (genome-wide association studies [GWAS], transcriptional profiling), or proteomic profiling; (iii) identifying the disease module using different modeling or statistical approaches; (iv) identifying pathways and the role of disease genes or modules in those pathways; and (v) disease module validation and prediction. *(Reproduced with permission from AL Barabási et al: Network medicine: A network-based approach to human disease. Nat Rev Genet 12:56, 2011.)*

TABLE 499-1 Examples of Systems Biology Application to Disease and Therapy

DISEASE	ANALYSIS	REFERENCE
Hereditary ataxias	Many ataxia-causing proteins share interacting partners that affect neurodegeneration	Lim et al: Cell 125:801-814, 2006
Diabetes mellitus	Metabolite-protein network analysis links three unique metabolite abnormalities in prediabetics to seven type 2 diabetes genes through four enzymes	Wang-Sattler et al: Mol Syst Biol 8:615, 2012
	The transcription factor NFATC4 regulates diabetes associated genes	Sharma et al: NPJ Syst Biol Appl 4:25, 2018
Epstein-Barr virus infection	Viral proteome exerts its effects through linking to host interactome	Gulbahce et al: PLoS One 8:e1002531, 2012
Pulmonary arterial hypertension	Network analysis indicates adaptive role for microRNA 21 in suppressing rho kinase pathway	Parikh et al: Circulation 125:1520-1532, 2012
Asthma	Disease module in the interactome explains phenotype heterogeneity and drug response	Sharma et al: Hum Mol Genet 24:3005-3020, 2015
Preeclampsia	Network analysis demonstrates the central role of vitamin D signaling in offsetting disease risk	Mirzakhani et al: J Clin Invest 126:4702-4715, 2016
Calcific aortic valve disease	Network and systems biology approaches identify the first molecular regulatory networks in vascular calcification	Schlotter et al: Circulation 138:377-393, 2018
Vascular fibrosis in pulmonary arterial hypertension	Network analysis identifies novel scaffold protein NEDD9 as a key regulator in pulmonary vascular fibrosis	Samokhin et al: Science Transl Med 10.1126/scitranslmed.aap, 2018
Chronic obstructive pulmonary disease	MicroRNA dysregulation identified from network analysis as a determinant of vascular remodeling	Musri et al: Am J Respir Cell Mol Biol 59:490-499, 2018
Drug repurposing	Network-based approach to in silico drug repurposing	Cheng et al: Nat Commun 10:3476, 2019
SARS-CoV-2	Network-based approach to identifying novel treatments for SARS-CoV-2 infection	Morselli Gysi et al: Proc Natl Acad Sci USA 118:e2025581118, 2021; Patten et al: iScience 25:104925, 2022

TABLE 499-2 Organizing Principles That Tie the Interactome to Human Disease

Local Principle: proteins involved in the same disease tend to interact.

Disease Module Principle: proteins involved in the same disease tend to cluster in connected subnetworks.

Functional Coherence Principle: proteins in a disease module are often involved in the same biological process.

Shared Components Principle: related diseases are located in the same interactome neighborhood from which unrelated diseases are separated.

Source: Reproduced from P Paci et al: NPJ Syst Biol Appl 7:3, 2021.

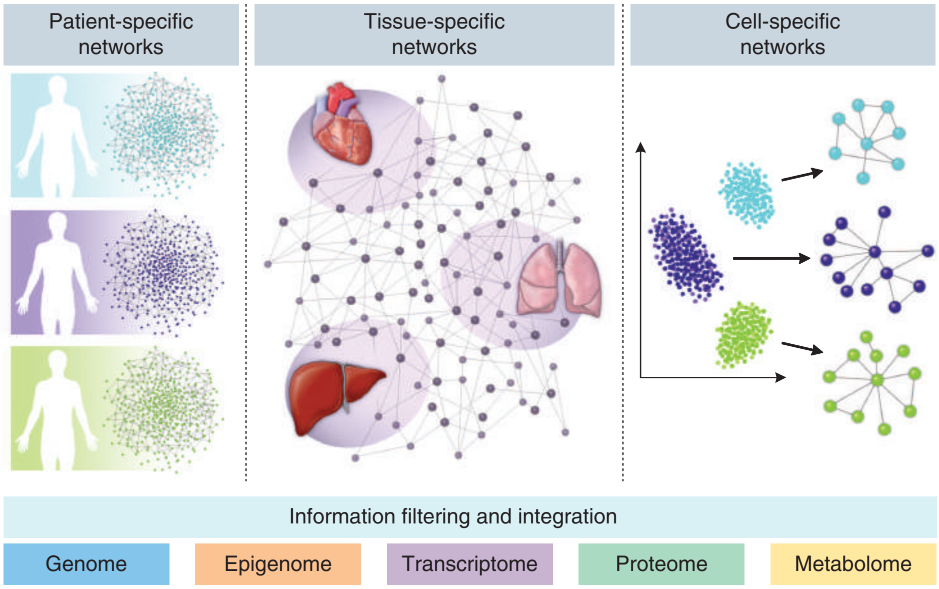
Patient-specific networks
Tissue-specific networks
Cell-specific networks
Information filtering and integration
Genome
Epigenome
Transcriptome
Proteome
Metabolome

FIGURE 499-7 Network integration of multiomics data across a wide spectrum from the patient level to the single-cell level. Patient-level omics data, tissue-specific omics data, and cell-specific omics data can be mapped to the human interactome to derive patient-specific networks, tissue-specific networks, and cell-specific networks, respectively. Such integrated networks are quite powerful for resolving the phenotypic heterogeneity in human diseases and guiding precise drug treatments.

occurs) is less important than that it occurs (with the exception of the organ-specific pathophysiological consequences that may require acute attention). For example, from this perspective, acute myocardial infarction (AMI) and its consequences are a reflection of thrombosis (in the coronary artery), inflammation (in the acutely injured myocardium), and fibrosis (at the site or sites of cardiomyocyte death). In effect, the major therapies for AMI address these intermediate pathophenotypes (e.g., antithrombotics, statins) rather than any organ-specific disease-determining process. This paradigm would argue for a systems-based analysis that would first identify the intermediate pathophenotypes to which a person is predisposed, then determine how and when to intervene to attenuate that adverse predisposition, and finally limit the likelihood that a major organ-specific event will occur. Evidence for the validity of this approach is found in the work of Rzhetsky and colleagues, who reviewed 1.5 million patient records and 161 diseases and found that these disease phenotypes form a network of strong pairwise correlations. This result is consistent with the notion that underlying genetic predispositions to intermediate pathophenotypes form the predicate basis for conventionally defined end-organ diseases.

Regardless of the specific nature of the systems pathobiological approach used, these analyses will lead to a drastic revision of the way human disease is defined and treated, continuing to validate the discipline of *network medicine, which reflects a fusion of the fields of systems biology and network science in the study of disease*. This will be a lengthy and complicated process, will be computationally intense and benefit from artificial intelligence and deep learning approaches, but ultimately will lead to better disease prevention and therapy and likely do so from an increasingly personalized, molecularly precise perspective **(Fig. 499-8)**. The analysis of pathobiology from a systems-based perspective is likely to help define specific subsets of patients more likely to respond to particular interventions based on shared disease mechanisms. Although it is unlikely that the extreme of "individualized medicine" will ever be practical (or even desirable), complex diseases can be mechanistically subclassified and interventions may be tailored to those settings in which they are more likely to work. This approach serves as a basis for the development of precision medicine.

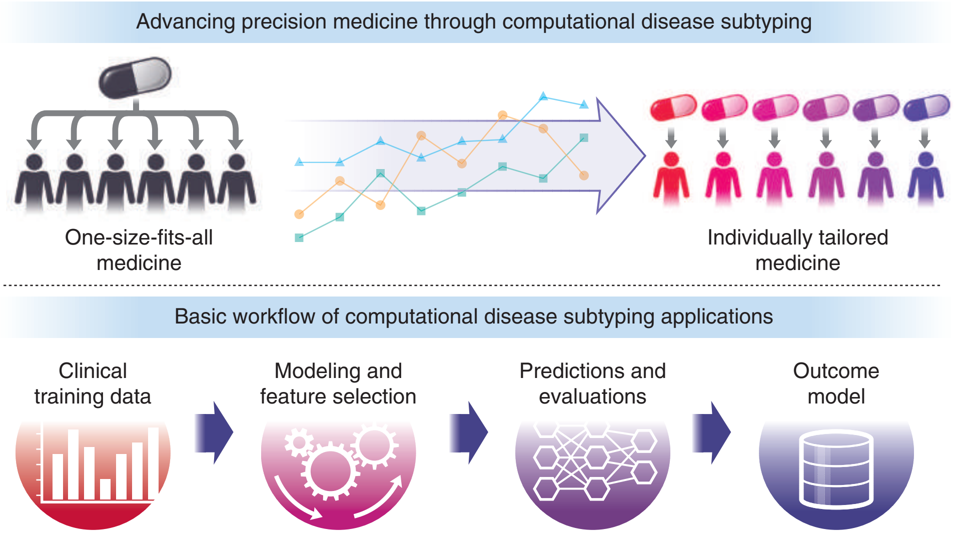
Advancing precision medicine through computational disease subtyping
One-size-fits-all medicine
Individually tailored medicine
Basic workflow of computational disease subtyping applications
Clinical training data
Modeling and feature selection
Predictions and evaluations
Outcome model

FIGURE 499-8 Advancing precision network medicine through computationally assisted molecular and phenotypic subtyping. ***Top:*** The path toward precision medicine moves from one-size-fits-all medicine to individually tailored medicine. ***Bottom:*** Basic workflow of computational disease subtyping applications.

FURTHER READING

Barabasi A-L et al: Network medicine: A network-based approach to human disease. Nat Rev Genet 12:56, 2011.

Cheng F et al: Network-based approach to prediction and population-based validation of in silico drug repurposing. Nat Commun 9:2691, 2018.

Liu X et al: Robustness and lethality in multilayer biological networks. Nat Commun 11:6043, 2020.

Loscalzo J et al (eds): *Network Medicine: Complex Systems in Human Disease and Therapeutics*. Cambridge, MA, Harvard University Press. Copyright 2017 by the President and Fellows of Harvard College. All rights reserved.

Loscalzo J et al: Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. Mol Syst Biol 3:124, 2007.

Maiorino E, Loscalzo J: Phenomics and robust multiomics data for cardiovascular disease subtyping. Arterioscl Thromb Vasc Biol 43:1111, 2023.

Menche J et al: Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science 347:1257601, 2015.

Oldham WM et al: Network analysis to risk stratify patients with exercise intolerance. Circ Res 122:864, 2018.

Paci P et al: Gene co-expression in the interactome: Moving from correlation toward causation via an integrated approach to disease module discovery. NPJ Syst Biol App 7:3, 2021.

Wang R et al: Multiomics network medicine approaches to precision medicine and therapeutics in cardiovascular diseases Arterioscl Thromb Vasc Biol 43:493, 2023.

500

Emerging Neurotherapeutic Technologies

Jyoti Mishra, Karunesh Ganguly

Neurotherapeutic technologies represent a diverse group of very promising treatment approaches with a common purpose of improving neurologic function. Decades of basic science research have paved the path for these novel technologies that have the potential to transform the lives of patients with neurologic diseases. A key goal is to minimize the consequences of lost abilities, whether they are motor, sensory, or cognitive. A common objective is to also harness the inherent plasticity of the nervous system, regardless of age, and even in the face of a degenerative process.

The technologies described below are the culmination of both an increased understanding of neural plasticity mechanisms in both the intact and the injured nervous system as well as advances in technology and computational power. While it is also clear that there may be fundamental limits on plasticity and repair mechanisms (the closing of developmental windows and/or loss of the ability of a network to compensate), the brain remains highly plastic regardless of age and even in the face of ongoing injury and/or degenerative processes. Collectively, there is now growing evidence to support neurologic restorative efforts for both "static" (e.g., stroke) and progressive neurologic disorders.

These technologies may not appear, at first glance, directly relevant to traditional medical care, but it is worth noting that clinicians have the most knowledge and experience about specific disease processes, available treatments, and the expected course of illnesses affecting the nervous system. It is thus critical that neurologic specialists and other clinicians play an important role in the future adoption of these technologies for neurologic rehabilitation. The sections below outline emerging diagnostic and therapeutic approaches that have the potential to transform the lives of patients with neurologic disorders. These include technologies to harness plasticity, neuroimaging, neurostimulation, and brain–machine interfaces.

NONINVASIVE TECHNOLOGIES TO HARNESS PLASTICITY

Neurologic rehabilitation aims to harness activity-dependent plasticity mechanisms to maximize functional restoration. This principle can be applied to a diverse range of functional domains such as movement control, sensory processing, language, pain, and cognition. For example, recent randomized controlled clinical trials for motor recovery after stroke have suggested that intensity of training may be particularly important for sustained long-term improvements. Moreover, studies of the effects of such training in rodent and nonhuman primate models further suggest that plasticity of cortical "motor maps" as well as the coordinated firing of neurons in remaining networks underlie observed functional improvements with rehabilitation. The incorporation of technology for neurologic rehabilitation has the great potential to revolutionize the delivery of care by significantly increasing access, reducing the burden for adherence to high-intensity regimens, and maximizing engagement. Below are three examples of how emerging technology can be used to harness neural plasticity and maximize functional restoration.

ROBOTICS

Rehabilitation robotics for both the upper and the lower limb have the potential to improve motor outcomes after stroke or other forms of brain injury. There is a growing recognition that focused training involving a range of tasks might be important for improved functional outcomes. While there is a growing recognition of "sensitive periods" that might represent optimal windows for rehabilitation after injury (e.g., perhaps the first several months after a stroke), such training likely has a role in the chronic period as well (e.g., maintenance therapy may also guard against known declines in function over time). Notably, the delivery of intensive training is a great challenge from both the perspective of the health care system and each patient. Outside of clinical trials, such a training program can be quite difficult to implement and maintain. It can also be costly and require significant effort.

Motor rehabilitation protocols using robotics have been developed and tested for both the upper limb and the lower limb. Such robotic therapies have often focused on the delivery of high-intensity movement practice that can surpass what is possible via existing standards of care. Moreover, robotic systems are capable of precisely measuring movement parameters (e.g., the kinematics of the movements) and providing quantitative feedback regarding the changes in performance during the training period. A particular focus has been on maximizing patient engagement and recruitment of attentional and reward pathways, both of which are increasingly recognized to drive neural plasticity. Ongoing advances in design and the user interface will continue to improve comfort and support sustained effort. For example, via close monitoring of performance and movement parameters, the system can aid at key points in order to minimize fatigue and ensure maximal engagement. Moreover, antigravity support of the upper limb can allow practice and task engagement even in the presence of severe weakness; this would be extremely challenging and labor intensive under current standards of care. Recent analysis also suggests that robotic devices may at least match outcomes realized with existing standards of care. However, rehabilitation robotics may also provide more precise feedback and permit novel quantitative rehabilitation approaches.

Figure 500-1 shows one example of an upper-limb robotic exoskeleton device that is currently being evaluated for training after stroke. A randomized, multicenter trial compared treatment with this exoskeleton system against conventional therapy provided by physical and occupational therapists. Participants were enrolled in the chronic phase and all had moderate-to-severe deficits; the groups underwent three sessions per week over an 8-week period. For robotic training, subjects trained with games to improve mobilization and to practice activities of daily living. This study provided evidence that both conventional and

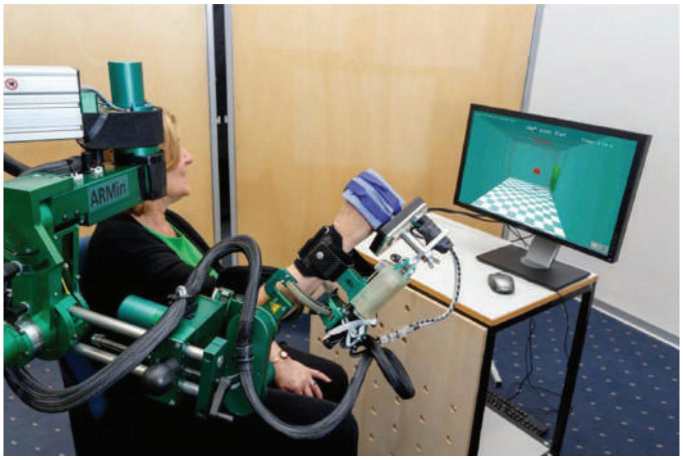

FIGURE 500-1 Photograph of a subject interacting with a complex upper-limb exoskeleton and a virtual reality system. *(From U Keller et al: Robot-assisted arm assessments in spinal cord injured patients: A consideration of concept study. PLoS One 10:e0126948, 2015.)*

robotic therapy could improve function in patients with chronic stroke. Multiple studies have also found similar gains when using either conventional or traditional approaches. Thus, a growing body of research supports the idea that such devices might complement conventional approaches to rehabilitation. Future work will need to define how rehabilitation robotics can optimally use adaptive and quantitative methods to further augment the recovery process.

VIRTUAL AND AUGMENTED REALITY

Therapeutic approaches using virtual reality (VR) and augmented reality (AR) aim to treat neurologic illnesses by specifically and quantitatively altering a patient's subjective experiences and interactions with the environment. Core components of both are advanced hardware and computational methods to generate simulated, yet realistic, perceptions. While some applications permit users to dynamically change the viewed perspective, other applications are designed to allow interactions among multiple users. Visual feedback is often a key component; this can include simple computer monitors or more immersive "head-mounted" viewers that modify the simulation based on changes in perspective. Tracking of movements (e.g., hand and head position) is often included. Multiple methods are used to allow a user to interact with the environment; interactions can be guided by straightforward means such as a keyboard, mouse, or even a joystick. More immersive methods are also frequently used. For example, gloves with embedded sensors and haptic inputs can allow the user's hand to be represented in real time in the simulated environment. Moreover, haptic interfaces can provide sensory feedback, allowing patients to interact with and "feel" virtual objects through multiple sensory modalities. A particular strength of these approaches is that therapeutic interventions can be studied in very controlled environments.

VR enables a user to interact with a simulated reality that can be precisely and quantitatively controlled. In addition to allowing patients to dynamically experience an altered reality, it can simultaneously monitor a subject's behaviors and responses. Such monitoring can allow precise measurements of clinically relevant parameters (e.g., motor actions, perception, cognitive processing) and can also be applied in specific rehabilitation training to achieve functionally meaningful goals. A growing body of literature indicates that VR environments can be tailored to individual needs and preferences, thereby maximizing engagement, motivation, and adaptation to ensure sufficient difficulty of tasks. VR environments can be designed to create powerful "gaming" platforms that are actually targeting clinically relevant parameters. For example, the upper-limb robotic systems described previously are frequently combined with VR environments that allow interaction with virtual objects.

In contrast to VR, AR overlays an artificial filter over a subject's view of the actual physical world, thus providing an "augmented" or enhanced view of the world around. AR is being tested in a diverse group of patients with neurologic impairments in the motor, sensory, or cognitive domains. AR may offer a particularly unique rehabilitation intervention for stroke patients. It is widely known that brain injuries limit patients' physical interaction with their environments. Furthermore, physical and cognitive impairments may limit social interactions. Such impoverished experiences are likely to be present during both the acute and the chronic phases. Importantly, there is clear basic scientific evidence that environmental enrichment can be a key component of rehabilitation; such enrichment may offer additive benefits to the often-limited formal rehabilitation sessions per week. Consistent with this are clinical studies suggesting that motor and cognitive outcomes may suffer when interactions with the environment are reduced; AR may be capable of increasing enrichment. For example, in the case of spatial neglect after stroke, the impaired modality may be accounted for using AR methods. Similarly, physical impairments that limit walking speeds can also limit visual feedback; both AR and VR can be used to enhance visual feedback during gait training.

Figure 500-2 shows an innovative application of AR for the treatment of "phantom limb" pain. A subset of both upper-limb and lower-limb amputees experience painful sensations that appear to originate from the missing limb. Past research has suggested that mirror therapy can be an effective treatment for phantom limb pain. During mirror therapy treatments, patients move their healthy arm in front of a mirror to produce a perception of movements of the missing limb. Previous studies have suggested that maladaptive plasticity of affected sensory cortices may be treated with mirror therapy. Importantly, in comparison to mirror therapy, AR-based therapy for phantom limb pain can be based on movements of the affected limb, i.e., using the remaining portion of the limb as opposed to the unaffected contralateral limb. This study demonstrated a novel treatment in which "phantom motor execution" is enabled using sophisticated machine-learning algorithms. More specifically, the study "decoded" phantom limb movements by measuring electromyogram (EMG) activity at the stump. Importantly, while the distal muscles responsible for movements were lost as a result of amputation, the remaining EMG activity could be used to predict presumed distal limb movements. As shown in Fig. 500-2, these inferred movements were projected onto an AR screen to create the perception of limb movements. The study showed that a subset of patients with long-term refractory phantom limb pain could experience a significant reduction in pain levels after using the AR system.

NEUROGAMING

Computerized programs that harness the power of video games have shown some evidence for ameliorating deficits in visual perception, age-related degeneration, and neuropsychiatric disorders. An essential feature of effective video game training is the progressive adjustment of the level of difficulty in line with the cognitive improvement of the patient. Important areas of active research include ways to enhance sustainability of neurogame training over long time periods and improving training *transfer*, i.e., the generalizability of task-specific training in one cognitive domain to more broad-based functional improvements. By leveraging video game technology, neurogames allow for dynamic user interaction and maintain user engagement across multiple sessions over several days of training. Important game mechanics include repetitive practice, performance-adaptive challenges, and several layers of reward feedback—from moment-to-moment point rewards to reward milestones over multiple sessions.

Notably, neurogames have therapeutic potential as they can be targeted to specific neurocognitive deficits. For instance, games have shown significant benefits in aging, by targeting speed of processing and training the abilities to multitask and suppress distractions. In each case, selective targeting is achieved by focusing the adaptive challenges to the neurocognitive domain of interest. Duration of response time windows available to the user or the level of interference are selectively targeted in the case of speed of processing training and interference training, respectively. More recent research demonstrated that it is possible to engender focused circuit neuroplasticity using such

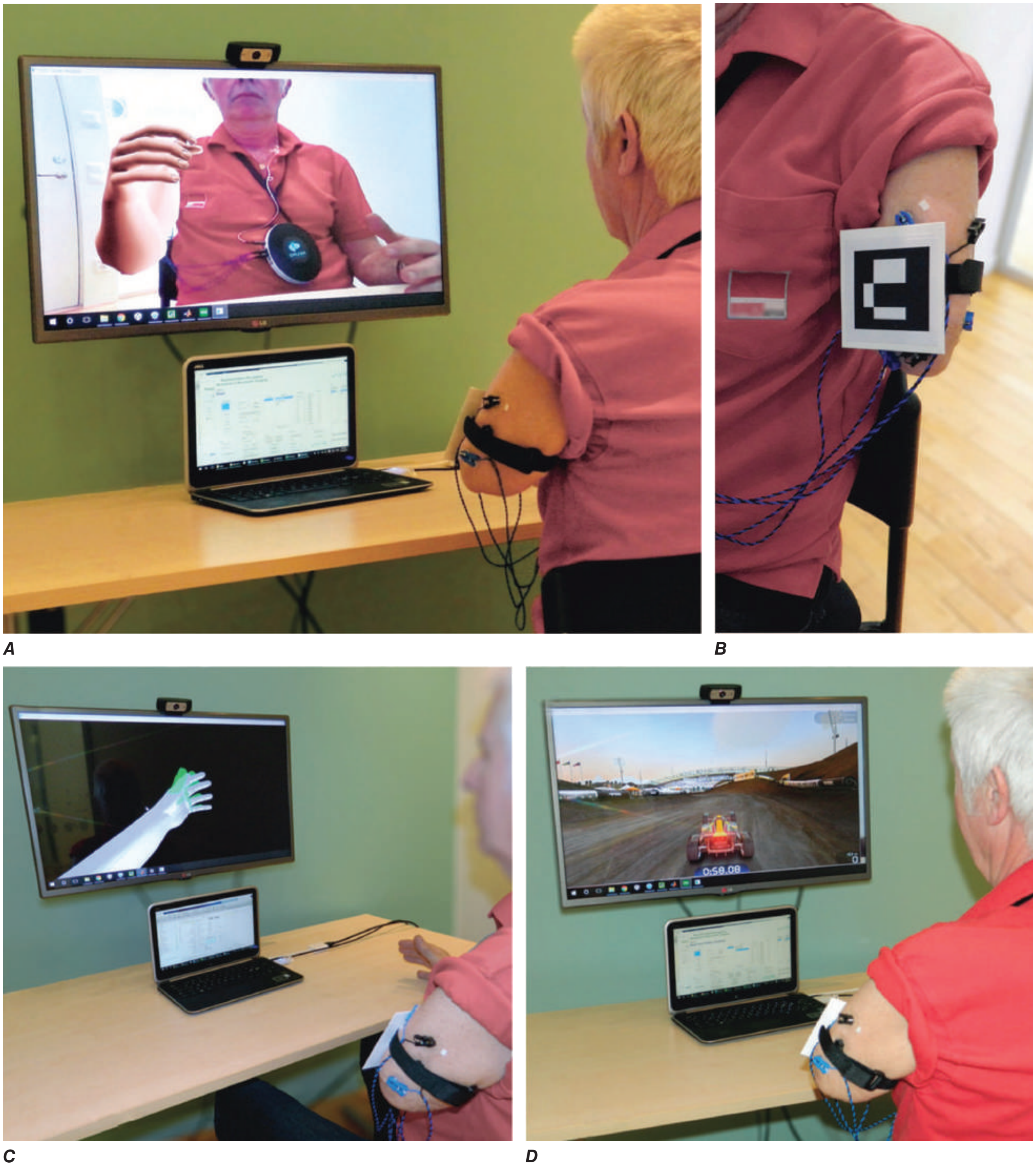

FIGURE 500-2 Augmented reality (AR) for phantom limb pain. ***A.*** A patient is shown a live AR video. ***B.*** Electromyography electrodes placed over the stump record muscle activation during training. ***C.*** The patient matches target postures during rehabilitation. ***D.*** Patient playing a game in which a car is controlled by "phantom movements." *(M Ortiz-Catalan et al: Phantom motor execution facilitated by machine learning and augmented reality as treatment for phantom limb pain: A single group, clinical trial in patients with chronic intractable phantom limb pain. Lancet 388:2885, 2016.)*

selective targeting in neurogaming. For example, older adults learned to adaptively perform within progressively more challenging distractor environments. Neuroplasticity selective to distractor processing was evidenced in this study at both the microscale, i.e., at the resolution of single neuron spiking in sensory cortex, as well as macroscale, i.e., electroencephalography (EEG)-based event-related potential recordings.

Video games have also shown promise in the treatment of visual deficits such as amblyopia, and in cognitive remediation in neuropsychiatric disorders such as schizophrenia. However, while the evidence base has been encouraging in small-sample randomized controlled trials (RCTs), larger RCTs are needed to demonstrate definitive therapeutic benefit. This is especially necessary as the commercial *brain training* industry continues to make unsubstantiated claims of the benefits of neurogaming; such claims have been formally dismissed by the scientific community. Like any other pharmacologic or device-based therapy, neurogames need to be systematically validated in multiphase RCTs establishing neural target engagement and documenting cognitive and behavioral outcomes in specific disorder populations.

Generalizability of training benefits from task-specific cognitive outcomes to more broad-based functional improvements remains the holy grail of neurogaming. Next-generation neurogames will aim to integrate physiologic measures such as heart rate variability (an index of physical exertion), galvanic skin responses, and respiration rate (indices of stress response), and even EEG-based neural measures. The objectives of such multimodal biosensor integration are to enhance the "closed-loop mechanics" that drive game adaptation and hence improve therapeutic outcomes and perhaps result in greater generalizability. These complex, yet potentially more effective, neurogames of the future will need rigorous clinical study for demonstration of validity and efficacy.

NEUROIMAGING

NEUROIMAGING OF CONNECTIVITY

Multimodal neuroimaging methods including functional magnetic resonance imaging (fMRI), EEG, and magnetoencephalography (MEG) are now being investigated as tools to study functional connectivity between brain regions, i.e., extent of correlated activity between brain regions of interest. Snapshots of functional connectivity can be analyzed while an individual is engaged in specific cognitive tasks or during rest. Resting-state functional connectivity (rsFC) is especially attractive as a robust, task-independent measure of brain function that can be evaluated in diverse neurologic and neuropsychiatric disorders. In fact, methodologic research has shown that rs-fMRI can provide more reliable brain signals of energy consumption than specific task-based fMRI approaches.

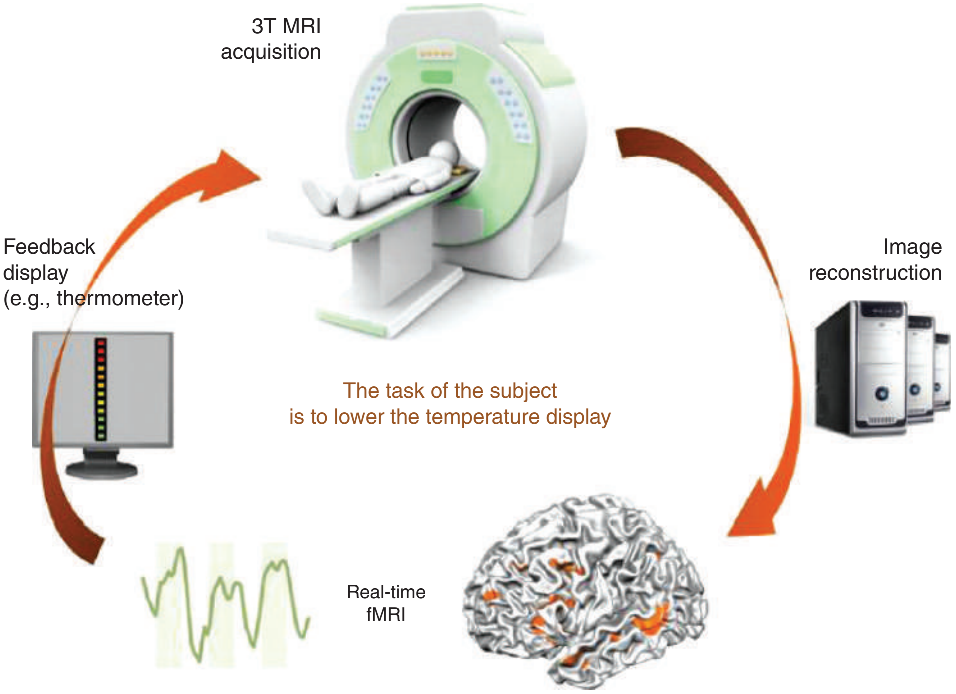

FIGURE 500-3 Neurofeedback using functional magnetic resonance imaging (fMRI). *(From T Fovet et al: Translating neurocognitive models of auditory-visual hallucinations into therapy. Front Psychiatry 7:103, 2016.)*

In recent years, there has been a surge of research to identify robust rsFC-based biomarkers for specific neurologic and neuropsychiatric disorders and thereby inform diagnoses and even predict specific treatment outcomes. For many such disorders, the network-level neurobiologic substrates that correspond to the clinical symptoms are not known. Furthermore, many are not unitary diseases, but rather heterogeneous syndromes composed of varied co-occurring symptoms. Hence, the quest for robust network biomarkers corresponding to complex neuropsychologic disorders is challenging and still in its infancy; yet some studies have made significant headway in this domain. For example, in a large multisite cohort of ~1000 depressed patients, Drysdale et al. (2017) showed that rsFC measures can subdivide patients into four neurophysiologic "biotypes" with distinct patterns of dysfunctional connectivity in limbic and frontostriatal networks. These biotypes were associated with different clinical-symptom profiles (combinations of anhedonia, anxiety, insomnia, anergia, etc.) and had high (>80%) diagnostic sensitivity and specificity. Moreover, these biotypes could also predict responsiveness to transcranial magnetic stimulation (TMS) therapy. Another recent study demonstrated utility of rsFC measures to predict diagnosis of mild traumatic brain injury (mTBI), which is clinically challenging by conventional means.

Apart from fMRI-based measures of rsFC, EEG- and MEG-based rsFC measures are also being actively investigated, as these provide a relatively lower-cost alternative to fMRI. While EEG is of lowest cost, it compromises on spatial resolution. The major strength of MEG is its ability to provide more accurate source-space estimates of functional oscillatory coupling than EEG, as well as provide measures at various physiologically relevant frequencies (up to 50 Hz shown to be clinically useful). In this regard, EEG and MEG are complementary to fMRI, which can only be used to study slow activity fluctuations (i.e., <0.1 Hz); the potential for EEG/MEG modalities to provide valid diagnostic biomarkers is currently underexploited and requires further study.

CLOSED-LOOP NEUROIMAGING

Neuroscientific studies to date are predominantly designed as "open-loop experiments," interpreting the neurobiologic substrates of human behavior via correlation with simultaneously occurring neural activity. In recent years, advances in real-time signal processing have paved the way for "closed-loop neuroimaging," wherein humans can directly manipulate experiment parameters in real-time based on specific brain signals (Fig. 500-3). Closed-loop imaging methods can not only advance our understanding of dynamic brain function but also have therapeutic potential. Humans can learn to modulate their neural dynamics in specific ways when they are able to perceive (i.e., see/hear) their brain signals in real-time using closed-loop neuroimaging-based neurofeedback. Early studies showed that such neurofeedback learning and resulting neuromodulation could be applied as therapy for patients suffering from chronic pain, motor rehabilitation in Parkinson's and stroke patients, modulation of aberrant oscillatory activity in epilepsy, and improvement of cognitive abilities such as sustained attention in healthy individuals and patients with attention-deficit hyperactivity disorder (ADHD). These approaches have also shown potential for deciphering state-of-consciousness in comatose patients, wherein a proportion of vegetative/minimally conscious patients can communicate awareness via neuroimaging-based mental imagery.

Closed-loop neuroimaging therapeutic studies have utilized real-time fMRI, EEG, and MEG methods. It is common for neural signals to be extracted from specific target brain regions for neuromodulation. However, given that distributed neural networks underlie behavioral deficits, new studies have also explored neurofeedback on combinatorial brain signals from multiple brain regions extracted using multivariate pattern analysis (MVPA). While early studies indicate therapeutic potential, clinical RCTs of closed-loop neuroimaging neurofeedback have shown mixed results. This may largely be because of the individual heterogeneity in neuropsychiatric disorders such that there is no one-size-fits-all therapy. Closed-loop neuroimaging-based therapies need to be more personalized to the preintervention cognitive and neurophysiologic states of the individual, and a better understanding developed regarding learning principles and mechanisms of self-regulation underlying neurofeedback. Clinical practitioners applying these methods also need to be well-educated on the hardware/software capabilities of these brain–computer interfaces to maximize patient outcomes.

NONINVASIVE BRAIN STIMULATION

Noninvasive brain stimulation (NIBS) is widely recognized as having great potential to modulate brain networks in a range of neurologic and psychiatric diseases; it is currently approved by the U.S. Food and Drug

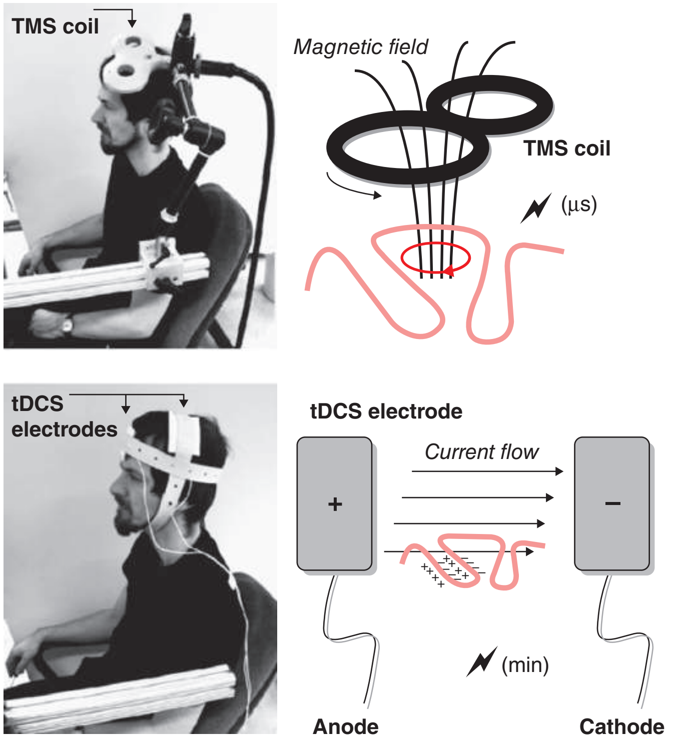

FIGURE 500-4 Illustration of transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS) setups. The upper panels show a TMS setup. Coils generate magnetic fields that can in turn generate electrical fields in the cortical tissue. The lower panels show a tDCS setup. The electrical current is believed to flow from the anode (+) to the cathode (–) through the superficial cortical areas leading to polarization. *(Reproduced with permission from R Sparing, FM Mottaghy: Noninvasive brain stimulation with transcranial magnetic or direct current stimulation [TMS/tDCS]—From insights into human memory to therapy of its dysfunction. Methods 44:329, 2008.)*

Administration (FDA) as a treatment for depression. Importantly, there is a very large body of basic research indicating that neuromodulation of the nervous system with electrical stimulation can have both short-term and long-term effects. While TMS uses magnetic fields to generate electrical currents, transcranial direct current stimulation (tDCS), in contrast, is based on direct stimulation using electrical currents applied at the scalp **(Fig. 500-4)**. TMS induces small electrical currents in the brain by magnetic fields that pass through the skull; it is known to be painless and therefore widely used for NIBS. Animal research suggests that anodal tDCS causes a generalized reduction in resting membrane potential over large cortical areas, whereas cathodal stimulation causes hyperpolarization. Prolonged stimulation with tDCS can cause an enduring change in cortical excitability under the stimulated regions. Further, changes in resting-state fMRI-based activity and functional connectivity have also been observed after tDCS. Notably, there is uncertainty regarding precisely how much electrical current is able to penetrate through the skull and modulate neural networks. Indeed, recent work has found that typical stimulation paradigms may not generate sufficient electrical fields to modulate neural activity; an alternate possibility is that peripheral nerves may be modulated and thus affect neural activity.

Neuromodulation via stimulation techniques such as tDCS and TMS have shown promise as methods to improve motor function after stroke; there are a growing number of studies demonstrating functional benefits of combining physical therapy with brain stimulation. Two commonly utilized TMS paradigms include low-frequency "inhibitory" stimulation of the healthy cortex or high-frequency "excitatory" stimulation of the injured hemisphere. Each aims to modify the balance of reciprocal inhibition between the two hemispheres after stroke. A meta-analysis of RCTs published over the past decade found a significant beneficial effect on motor outcomes. Unfortunately, a recent large multicenter trial to assess the long-term benefits of TMS on motor recovery after stroke (NICHE trial) did not find a benefit at the population level. Ongoing research aims to better understand how stimulation can directly affect neural patterns and thus allow more customization of stimulation—past trials did not record the neural responses to stimulation.

TMS and tDCS interventions are also being applied in psychiatric disorders. A substantial body of evidence supports the use of TMS as an antidepressant in major depressive disorder (MDD). TMS is also being investigated for its potential efficacy in posttraumatic stress disorder (PTSD), obsessive compulsive disorder (OCD), and treatment of auditory hallucinations in schizophrenia. Various repetitive TMS (rTMS) protocols have shown efficacy in major depression. These include both low-frequency (≤1 Hz) and high-frequency (10–20 Hz) rTMS stimulation over the dorsolateral prefrontal cortex (DLPFC). Mechanistically, low-frequency rTMS is associated with decreased regional cerebral blood flow while high-frequency rTMS elicits increased blood flow, not only over the prefrontal region where the TMS is applied but also in associated basal ganglia and amygdala circuits. Notably, the differential mechanisms of low- versus high-frequency rTMS protocols are associated with mood improvements in different sets of MDD patients, and patients showing benefits with one protocol may even show worsening with the other, again pointing to individual heterogeneity in network function. EEG-guided TMS is also being investigated in psychiatric disorders, for instance, the individual resting alpha-band (8–12 Hz) peak frequency to determine TMS stimulation rates. With respect to transcranial electrical stimulation in psychiatry, tDCS is the most commonly used protocol. In major depression, there is a documented imbalance in left versus right DLPFC activity; hence, differential anodal versus cathodal tDCS in the left versus right prefrontal cortex may be a potentially efficacious approach. Interestingly while meta-analysis shows promise for NIBS methods in psychiatric illness, large RCTs have failed to generate benefits compared to placebo treatment. Future success may require careful personalized targeting based on network dynamics and refinement of protocols to accommodate combinatorial treatments.

IMPLANTABLE NEURAL INTERFACES

Fully implantable neural interfaces that can improve clinical function already exist. Cochlear implants, for example, are sensory prostheses that can restore hearing in deaf patients. Environmental sounds are processed in real-time and then converted into patterned stimulation delivered to the cochlear nerve. Importantly, even while the patterned stimulation remains the same, there are gradual improvements in the perception of speech and other complex sounds over a period of several months after device implantation. Activity-dependent sculpting of neural circuits is hypothesized to underlie the observed perceptual improvements. Similarly, the development of deep-brain stimulation (DBS) was based on decades of work showing that surgical lesions to specific nuclei could alleviate tremor and bradykinesia in animal models. DBS involves chronic implantation of a stimulating electrode that targets specific neural structures (e.g., subthalamic nuclei or the globus pallidus in Parkinson's disease). At least for movement disorders, it is commonly thought that targeted areas are functionally inhibited by the chronic electrical stimulation.

■ IMPLANTABLE DEVICES FOR NEUROMODULATION

There has been recent progress in the development of implantable neural interfaces to treat neurologic and psychiatric illnesses. For example, for patients with refractory focal epilepsy and clearly identified seizure foci, invasive "responsive stimulation" is FDA approved. Responsive stimulation is grounded on principles of closed-loop stimulation based on real-time monitoring of brain oscillations; specifically, the device aims to detect the earliest signatures of the onset of a seizure, usually at a stage that is not symptomatic, and then deliver focused electrical stimulation to prevent further progression and generalization. A large RCT of this device was performed in patients with intractable focal epilepsy; they were assigned to either sham or active stimulation in response to seizure detection. There was a significant reduction in

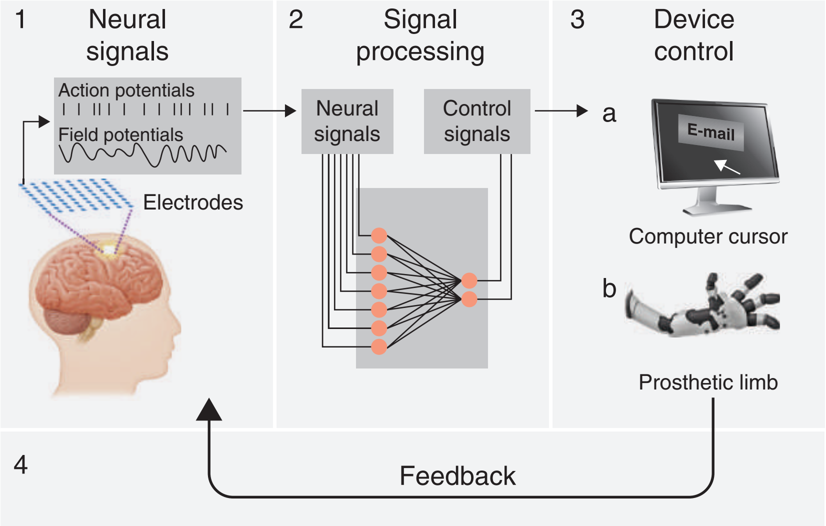

FIGURE 500-5 Components of a brain–machine interface (BMI). *(Reproduced with permission from A Tsu et al: Cortical neuroprosthetics from a clinical perspective. Neurobiol Dis 83:154, 2015.)*

seizure frequency in the stimulation group, but it was rare for patients to become seizure-free. There were also modest improvements in quality of life. Notably, there was a small increased risk of hemorrhage associated with the device. In addition to providing clinicians with another treatment option, this device has offered important avenues for research and further optimization. For example, it is now possible to monitor subclinical and clinical seizures and intracranial EEG in patients with chronic epilepsy. This has resulted in new knowledge about the association of seizures with circadian rhythms and sleep. It is also anticipated that a better understanding of the triggers of seizures and the development of better stimulation algorithms, based on real-world data, can ultimately lead to more effective treatments.

There is also great interest in the development of treatments for refractory depression. One area of focus has been on the development of DBS to treat depression. While early smaller studies were promising, a larger study failed to find benefits at the population level. Subsequent analysis has suggested the possibility that more precise tailoring of stimulation parameters to each individual is warranted, both at the level of specific pathways identified through neuroimaging as well as network activity biomarkers. Recent studies have, in fact, supported the notion that individualized patterns of network activity are predictive of a patient's symptoms and how the patient might respond to stimulation. There are now planned studies that aim to tailor stimulation to each individual with severe depression.

■ VAGUS NERVE STIMULATION TO IMPROVE RECOVERY AFTER STROKE

Vagal nerve stimulation (VNS) has recently been approved by the FDA as a therapy to enhance motor recovery after stroke. Animal studies first provided clear evidence that VNS is safe and can enhance plasticity in both intact animals as well as in models of injury. Importantly, these studies indicated that precise timing of movements is important for efficacy. For example, in animal models of stroke, stimulation of the vagus nerve was timed to the end of successful movement repetitions; these studies further indicated that the precise timing of VNS during rehabilitation is essential. VNS appears to result in rapid activation of cholinergic and noradrenergic systems; the activation of these neuromodulators may enhance attentional effects and improve "signal to noise," thus facilitating the encoding of relevant task features. This basic research culminated in smaller clinical trials and a subsequent pivotal randomized trial of VNS in stroke. In this trial, after 6 weeks of therapy paired with VNS, participants randomized to the VNS group (n = 53) had a significant increase in forelimb function compared to the control group. In addition, 90 days after the study was completed, a higher percentage of patients in the VNS group maintained clinically meaningful responses. Together, this indicates that VNS is a promising new therapy to augment rehabilitation after stroke. However, given the variability of effects for single patients, additional research is required to determine which stroke patients are the most likely to benefit. Future advances that allow VNS to be delivered in the home setting should also lead to greater use of this approach.

■ BRAIN–COMPUTER INTERFACES FOR PARALYSIS

Brain–computer interfaces (BCIs) represent a more advanced neural interface that aims to restore motor function. Multiple neurologic disorders (e.g., traumatic and nontraumatic spinal cord injury, motor neuron disease, neuromuscular disorders, stroke) can result in severe and devastating paralysis. Patients cannot perform simple activities, and they remain fully dependent for care. In patients with high cervical injuries, advanced amyotrophic lateral sclerosis (ALS), or brainstem strokes, the effects are especially devastating and often leave patients unable to communicate. While there has been extensive research into each disorder, clinically effective approaches for rehabilitation of long-term disability are lacking. BCIs offer a promising means to restore function. In the patient groups described above, while the pathways for transmission of signals to muscles are disrupted, the brain itself is largely functional. Thus, BCIs can restore function by communicating directly with the brain. For example, in a "motor" BCI, a subject's intention to move is translated in real time to control a device. As illustrated in **Fig. 500-5**, the components of a motor BCI include the following: (1) recordings of neural activity, (2) algorithms to transform the neural activity into control signals, (3) an external device driven by these control signals, and (4) feedback regarding the current state of the device.

Many sources of neural signals can be used in a BCI. While EEG signals can be obtained noninvasively, other neural signals require invasive placement of electrodes. Three invasive sources of neural signals include electrocorticography (ECoG), action potentials or spikes, and local field potentials (LFPs). Spikes and LFPs are recorded with electrodes that penetrate the cortex. "Spikes" represent high-bandwidth signals (300–25,000 Hz) that are recorded from either single neurons or multiple neurons ("multiunit"). LFPs are the low-frequency (~0.1–300 Hz) components. In contrast, ECoG is recorded from electrodes that are placed on the cortical surface. ECoG signals may be viewed as an intermediate-resolution signal in comparison with spikes/LFPs and EEG. While it is worth noting that there is still considerable ongoing research into the specific neural underpinning of each signal source, there has been great progress in the ability to decode a user's intention.

A central goal of the field of BCIs is to improve function in patients with severe disability. This can consist of a range of communication and assistive devices such as a computer cursor, keyboard control, wheelchair, or robotic limb. In the ideal scenario, the least invasive method of recording neural signals would allow the most complex level of control. Decades of research in nonhuman primates and early-phase clinical trials have demonstrated the feasibility of direct neural control of assistive technology based on recording of neural signals at multiple resolutions. There have been numerous examples of human subjects with a range of neurologic illnesses (e.g., brainstem stroke, ALS, spinal cord injury) who have demonstrated the actual use of implantable neural interfaces. This includes demonstrations of both the control of communication interfaces as well as robotic limbs. Early pilot clinical trials of BCIs based on invasive recordings of neural signals showed that relatively high rates of brain-controlled typing are possible (e.g., >30 characters per minute). A past case study additionally demonstrated that a fully implantable BCI system could allow communication in a locked-in ALS patient **(Fig. 500-6)**. At the time of the study, the patient required mechanical ventilation and could only communicate using eye movements. She was implanted with multiple subdural cortical electrodes; the neural signals were then processed and sent wirelessly to an external augmentative alternative communication (AAC) device.

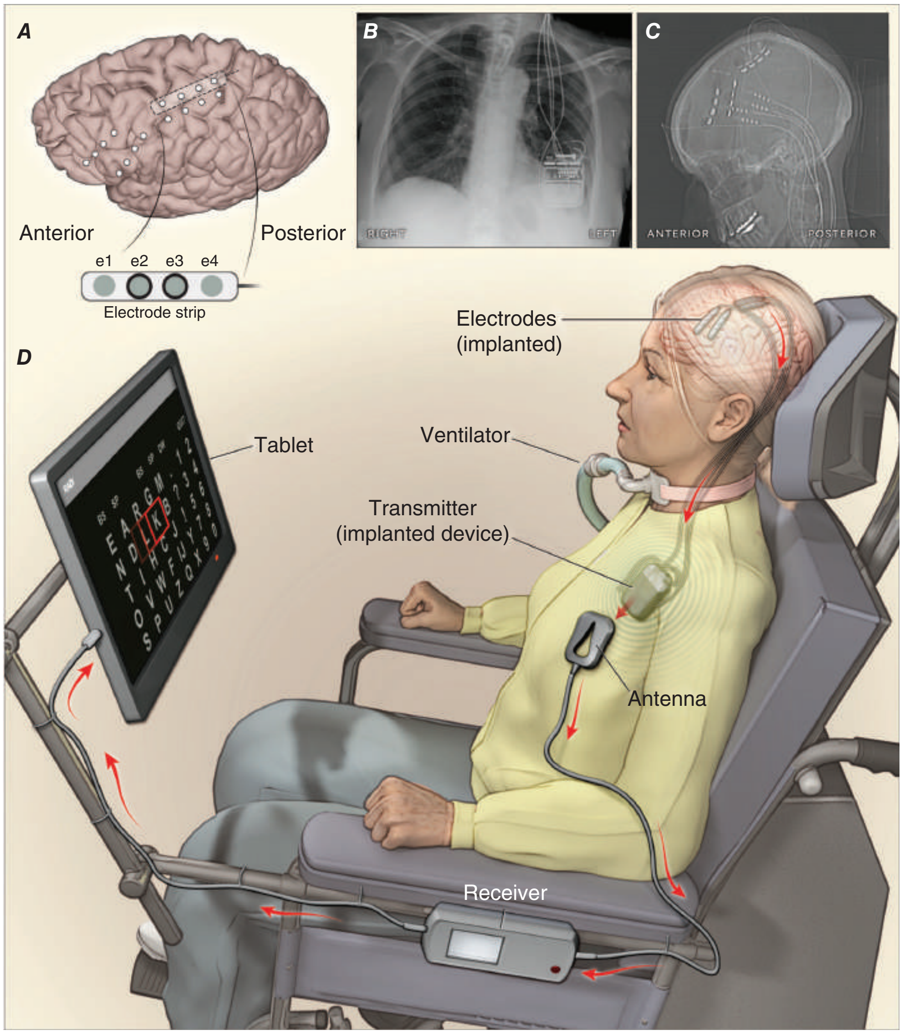

FIGURE 500-6 Illustration of an amyotrophic lateral sclerosis (ALS) patient with a fully implanted communication interface. ***A.*** Illustration of the location of electrodes on the brain. ***B.*** X-ray of chest showing the wireless module. ***C.*** X-ray of leads and wire routing. ***D.*** Schematic of the subject performing a typing task. *(From MJ Vansteensel et al: Fully implanted brain–computer interface in a locked-in patient with ALS. N Engl J Med 375:2060, 2016. Copyright © 2016 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

Importantly, she could use the interface with no supervision from research staff, albeit with a relatively low communication rate.

Over the past 5 years, there has been tremendous progress toward the goal of restoring much higher rates of communication in participants with severe impairments. These studies have used either ECoG or spike-based decoding. One of the first studies indicated that a participant with a brainstem stroke and anarthria could communicate using a set of 50 words. Two subsequent studies showed that decoding a significantly larger set of words is possible, using either spiking or ECoG. For example, one study, using spike-based recordings, indicated that decoding of a large vocabulary was possible using phoneme-based decoding; that is, an arbitrary and a remarkably large set of words could be decoded by decomposing into its set of phonemes. Together, these studies indicate the real possibility of a clinically viable speech neuroprosthetic to restore fast communication in those with anarthria or severe dysarthria.

Overall, there has been tremendous progress recently in the translation of BCIs. There are now also multiple commercial efforts to take these findings from pilot studies and to scale them to a commercially viable device. In fact, there is already a single participant with tetraplegia implanted with a first-in-class commercial device that can record spiking activity. While there are still challenges with long-term stability, this participant appears to be using this implanted device to control a computer (e.g., to control a cursor and to play video games) in the home setting. Additional work will be required to fully quantify how stable neural interfaces are and the level of performance that can be reliably achieved. As these characteristics become increasingly clear, it should allow targeted clinical translational efforts that are geared toward specific patient needs and preferences (e.g., extent of disability, medical condition, noninvasive vs invasive). For example, patients with high cervical injuries (i.e., above C4, where the arm and the hand are affected) have rehabilitation needs different from patients with lower cervical injuries (i.e., below C5–C6, where the primary deficits are the hand and fingers). Moreover, interfaces to restore communication may be different from those aimed toward movement control. We fully anticipate that over the next decade there will be larger scale clinical studies to demonstrate how BCIs allow participants with severe impairments to experience the ability to communicate and to control assistive technology.

■ FURTHER READING

Baniqued PDE et al: Brain-computer interface robotics for hand rehabilitation after stroke: A systematic review. J Neuroeng Rehabil 18:15, 2021.

Bassett DS et al: Emerging frontiers of neuroengineering: A network science of brain connectivity. Annu Rev Biomed Eng 19:327, 2017.

Dawson J et al: Vagus nerve stimulation paired with rehabilitation for upper limb motor function after ischaemic stroke (VNS-REHAB): A randomised, blinded, pivotal, device trial. Lancet 397:1545, 2021.

Drysdale AT et al: Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat Med 23:28, 2016.

Ganguly K et al: Modulation of neural co-firing to enhance network transmission and improve motor function after stroke. Neuron 110:2363, 2022.

Liu A et al: Immediate neurophysiological effects of transcranial electrical stimulation. Nat Comm 9:5092, 2018.

Metzger SL et al: A high-performance neuroprosthesis for speech decoding and avatar control. Nature 620:1037, 2023.

Mishra J et al: Video games for neuro-cognitive optimization. Neuron 90:214, 2016.

Reinkensmeyer DJ et al: Computational neurorehabilitation: Modeling plasticity and learning to predict recovery. J Neuroeng Rehabil 13:42, 2016.

Scangos K et al: State-dependent responses to intracranial brain stimulation in a patient with depression. Nat Med 27:229, 2021.

501

Machine Learning and Augmented Intelligence in Clinical Medicine

Arjun K. Manrai, Isaac S. Kohane

Machine learning has reshaped our consumer lives, with self-driving vehicles, conversant chatbots, and machine translation services so ubiquitous that they are at risk of not being considered particularly intelligent for much longer. Will the algorithms underlying these technologies similarly transform the art and practice of medicine? There is hope that modern machine learning techniques—especially the resurgence of artificial neural networks in deep learning—will foster a sea change in clinical practice that augments both the sensory and diagnostic powers of physicians while, perhaps paradoxically, freeing physicians to spend more time with their patients by performing time-consuming laborious tasks for them.

From the birth of artificial intelligence at the Dartmouth Summer Research Project on Artificial Intelligence in 1956 to powerful large language models (LLMs) ubiquitous today, machine learning methods and theory have developed in symbiosis with growing datasets and computational power. In this chapter, we discuss the foundations of modern machine learning algorithms and their emerging applications in clinical practice. Modern machine learning techniques are sufficiently capacious as to learn flexible and rich representations of clinical data and are remarkably adept at exploiting spatial and temporal structure in raw data. Recent machine learning models perform on par with expert physicians or prior state-of-the art models on tasks ranging from those involving the interpretation of images (e.g., grading retinal fundus photographs for diabetic retinopathy), to analysis of unstructured text (e.g., predicting hospital readmission from electronic health record notes or providing a differential diagnosis for a challenging clinical case), and processing of speech (e.g., detecting depression from patient speech). Over the past few years, LLMs, including the generative pretrained transformer (GPT) class of models underlying ChatGPT, have demonstrated the ability not only to excel at medical professional competency examinations but even to engage in general diagnostic reasoning abilities that many would have likely considered science fiction even a few years ago. However, many evaluations of machine learning models occur on tasks that are narrow and unrealistic and further lack the clinical context that a physician would incorporate in actual practice. The models themselves are also often divorced from considerations of actionable recommendations useful in patient care.

To help ensure these models benefit patients, this chapter aims to bring more physicians into the design and evaluation of machine learning models by providing an understanding of how modern machine learning models are developed and how they relate to more familiar methods from the epidemiologic literature.

The terms *machine learning* and *artificial intelligence* evoke distinct images today from those conjured up by the same terms in the 1950s and the 1980s, and they likely will mean something different in a decade. Computer scientist John McCarthy originally defined artificial intelligence in 1956 as "the broad science and engineering of creating intelligent machines," most often embodied today as computer software. Machine learning can be viewed as the subfield of artificial intelligence encompassing algorithms that extract generalizable patterns from data. This approach stands in contrast to approaches to create intelligence from human-engineered and explicitly programmed rules that characterized many early applications of artificial intelligence to medicine during the 1970s and 1980s (e.g., expert systems such as INTERNIST-I and MYCIN).

This chapter covers machine learning methods and applications that may augment physician expertise at the point of care. Many applications of machine learning in health care are therefore not reviewed here, such as algorithms to improve hospital planning, detect insurance fraud, and monitor new drugs for adverse events. Throughout this chapter, we discuss how new machine learning methods can learn rich representations of both clinical state and patient identity by discovering how to represent raw data. At the same time, models largely reflect the data on which they are trained and, thus, may encode and amplify biased prior practices; they may also be brittle in unfamiliar and evolving settings. If machine learning methods are positioned to tackle problems based on physician and patient needs and continuously reassessed, we envisage a future with clinics instrumented with machine learning tools that augment the ability of physicians to reason precisely and probabilistically at the point of care.

CONCEPTS OF MACHINE LEARNING

TYPES OF MACHINE LEARNING

Many physicians will be familiar with the major types of machine learning methods from methodologic counterparts discussed in the context of "traditional" statistical and epidemiologic modeling. In the current machine learning and epidemiology literature, much confusion arises over whether a method "belongs" in one camp or the other. More is gained by focusing on the computational and statistical connections, particularly in understanding how new machine learning methods compare to familiar clinical risk stratification approaches.

Broadly, there are four major types of machine learning with applications to clinical medicine: (1) supervised learning, (2) unsupervised learning, (3) semi-supervised learning, and (4) reinforcement learning. The four subfields of machine learning differ from one another in both their objectives and the degree to which algorithms have access to labeled examples from which to learn **(Fig. 501-1)**. All four subfields have roots tracing back decades with classical examples and modern counterparts **(Table 501-1)**.

Until recently, supervised machine learning approaches have dominated the medical literature. In supervised learning, paired input and labeled output examples are provided together to a machine learning algorithm that learns what combination of (potentially millions) of parameters optimizes a function that predicts output from input. The goal is to learn robust functions that work well with unseen data. If this setting is familiar, it is because clinical researchers often use well-known traditional statistical approaches like linear and logistic regression to achieve the same goal. For example, clinical risk scores, such as the American College of Cardiology (ACC)/American Heart Association (AHA) Atherosclerotic Cardiovascular Disease (ASCVD) Pooled Cohort Risk Equations or the Framingham Risk Score, are based on fitting models with paired input data (e.g., age, sex, total cholesterol, smoking status) and labeled output data (e.g., first occurrence of nonfatal myocardial infarction, coronary heart disease death, or fatal or nonfatal stroke). Contemporary deep learning methods can learn flexible representations of raw input data as opposed to relying on expert-identified features (Table 501-1). A contemporary clinical example of supervised machine learning with convolutional neural networks is the histopathologic detection of lymph node metastases in breast cancer patients (Table 501-1).

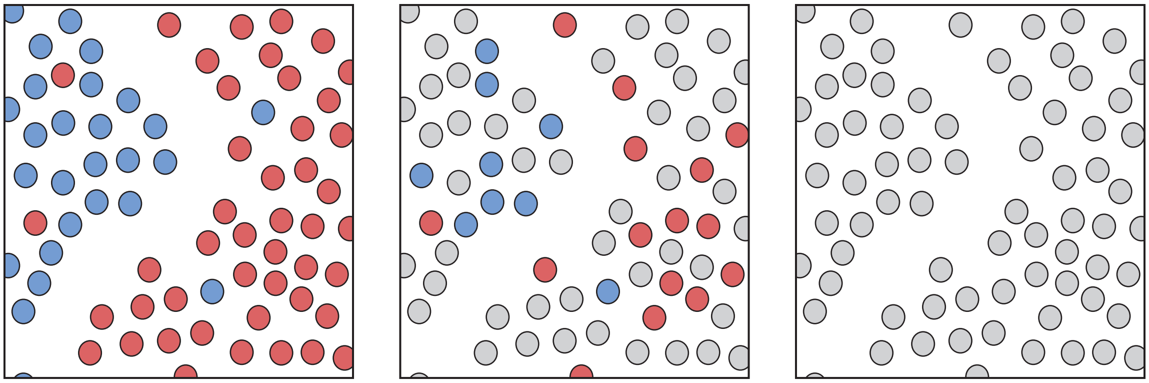

Supervised Semi-Supervised Unsupervised

FIGURE 501-1. The subfields of machine learning differ in their access to labeled examples. In supervised learning (*left*), the learning algorithm uses labeled output examples (*red, blue*) to learn a classifier for determining whether a new unlabeled data point is red or blue. Unsupervised learning methods (*right*) do not use labels but, rather, identify structure present in the data. Semi-supervised learning methods (*center*) have access to both unlabeled and labeled examples; the unlabeled data help learn a classifier with fewer labeled examples. *(Reproduced with permission from Luke Melas-Kyriazi.)*

The three remaining types of machine learning have not been as widely applied to clinical problems, but this has started to shift in recent years. These methods include unsupervised learning, semi-supervised learning, and reinforcement learning. Unsupervised learning, in contrast to supervised learning, encompasses methods that use unlabeled input data, where the goal is to discover the "structure" present in the data. A researcher may use unsupervised methods to determine whether or not their data lie on a low-dimensional "manifold" that is "embedded" in a higher-dimensional space (where dimensions here refer to the number or size of the collected elements or features of a sample or system). For example, a researcher may obtain gene expression measurements from >20,000 protein-coding genes for a large group of asthma patients and then "project" (and thereby reduce the dimensionality of) each patient into a lower-dimensional space to visualize and understand structure present in the dataset or may group asthma patients by similarity across all gene expression values. Classical linear methods include principal components analysis (PCA) and contemporary nonlinear approaches include uniform manifold approximation and projection (UMAP) (Table 501-1). Another example of unsupervised learning (or more precisely, self-supervised learning) is the pretraining process for LLMs such as ChatGPT, where massive text corpora are used to teach the model to predict the next word (Table 501-1), a task that is simple to describe but, when coupled with downstream processes to fine-tune model outputs, leads to powerful emergent behavior of these models. Semi-supervised learning is a hybrid between supervised and unsupervised learning, with methods that use both labeled data and unlabeled data. These algorithms exploit the (often low-dimensional) structure of unlabeled data to learn better models than may be possible in a purely supervised setting where labeled data may be scarce. Finally, reinforcement learning is a distinct subfield of machine learning that focuses on optimizing the iterative decision-making of an "agent" that is equipped with a cumulative "reward" function and, thus, must navigate a trade-off between exploration and exploitation of its environment. This is distinct from the other three subfields of machine learning where the entire learning signal (i.e., dataset) is presented at once. This approach to learning has been successful in teaching computers to play games at world-expert levels (e.g., Google's AlphaGo Zero).

■ MODERN MEDICAL MACHINE LEARNING

The modern machine learning toolkit includes methods that differ substantially in their complexity and ability to learn directly from raw data **(Table 501-2)**. "Traditional" statistical methods such a linear and logistic regression remain vital and often serve at minimum as useful and interpretable baselines, but often provide much more. Generalizations of these approaches as well as many other methods have been developed to learn complicated functions including highly nonlinear relationships. For example, models such as gradient boosted trees (Table 501-2) often achieve excellent performance with tabular data that lack the spatial or temporal structure that newer deep learning methods can exploit.

Modern deep learning models can learn rich and flexible representations of raw clinical data. The building blocks of these models are simple artificial "neurons," often arranged into layers **(Fig. 501-2)**. Each neuron accepts input from neurons in a preceding layer, computes a weighted sum of these inputs, and applies a nonlinear function to the weighted sum (Fig. 501-2).

TABLE 501-1 Types of Machine Learning and Clinical Examples

SUBFIELD OF MACHINE LEARNING	DEFINITION	CLASSIC EXAMPLE	CONTEMPORARY EXAMPLE	CLINICAL EXAMPLE
Supervised learning	Methods that use paired input and *labeled* output examples to learn a generalizable function that predicts output labels from input	Logistic regression	Convolutional neural network (CNN)	Histopathologic detection of lymph node metastases in breast cancer patients
Unsupervised learning	Methods that use *unlabeled* input data to discover data structure and learn efficient representations for data (e.g., clustering, dimensionality reduction)	Principle components analysis (PCA)	Pre-training of large language models (LLMs)	Visualizing structure in gene expression levels and grouping asthma patients into distinct molecular clusters; learning medical knowledge from large-scale text corpora
Semi-supervised learning	Methods that use *both* unlabeled and labeled examples to learn better functions than possible in supervised setting alone	Self-training	Consistency regularization	Use of unlabeled cardiac magnetic resonance images alongside small dataset of labeled examples to detect hypertrophic cardiomyopathy
Reinforcement learning	Methods to teach an "agent" that iteratively interacts with its environment *how* to optimize a numerical reward	Optimal control	Deep reinforcement learning (e.g., AlphaGo Zero)	Selecting fluids and vasopressor dosing iteratively to manage sepsis for patients in the intensive care unit

TABLE 501-2 Select Tools in the Modern Medical Machine Learning Toolkit

METHOD	DEFINITION	NOTES
Linear and logistic regression	Models a linear relationship between predictors and either a continuous or binary outcome variable; "traditional" statistical modeling	Necessary baseline. In small carefully curated clinical datasets, these methods often perform on par with more sophisticated methods
Gradient-boosted trees	Ensemble of "decision trees" with parameters optimized to efficiently learn accurate nonlinear functions in supervised setting	Efficient to train and often performs well on machine learning tasks with tabular data
Convolutional neural network (CNN)	Specialized deep learning architecture with groups of neurons ("convolutional filters") that exploit structure	Widely-used deep learning model for medical imaging tasks (e.g., U-Net architecture for biomedical image segmentation)
Transformer models	Deep learning architecture designed for mapping input sequences to output sequences (e.g., text, speech)	Generative Pre-trained Transformer (GPT-4) and ChatGPT; state of of the art for tasks in natural language processing and machine translation
Latent diffusion models	A generative deep learning model that can create realistic synthetic data across multiple modalities	Can generate realistic, even award-winning art; popular tools include DALL-E2 and Stable Diffusion
Transfer learning	Family of approaches to adapt models trained for one task and apply to another task (across domains)	Useful for "jump-starting" a model for a new problem, e.g., many medical computer-vision models start with a network trained for a separate (often nonmedical) task such as ImageNet and may "fine-tune" for a particular medical application

The values of the adjustable weights between neurons are learned during the training process. Neurons may be arranged into hierarchical layers that build an increasingly rich representation of input data. For example, convolutional neural networks are specialized architectures that combine groups of neurons with the mathematical operation of convolution ("convolutional filters")—to exploit spatial structure (Table 501-2). Initial layers learn low-level features (e.g., edges) and then build to higher layers that learn motifs and objects, creating a powerful representation for discriminating between inputs using output labels **(Fig. 501-3)**.

Modern neural networks may have many millions and even hundreds of billions of weight parameters. For example, the VGG-16 network by Simonyan and Zisserman, a path-breaking architecture and one of the top models in the ImageNet Large Scale Visual Recognition Challenge in 2014, had about 140 million parameters. These models would be considered small by contemporary standards; the GPT-3 model created by the company OpenAI has 175 billion parameters. The U-Net architecture by Ronneberger and colleagues, a convolutional neural network, is used frequently for biomedical image segmentation and other medical imaging tasks. Other deep learning models have architectures tailored for distinct tasks and data types including sequence to sequence and "transformer" models (the "T" in GPT-3/GPT-4), designed for sequential data, including unstructured text notes from electronic health records or patient queries (Table 501-2). In many deep learning applications, it is important to note that practitioners often do not train networks as a tabula rasa but instead benefit substantively from transfer learning (Table 501-2), where the weights may be initialized or "pretrained" from another task.

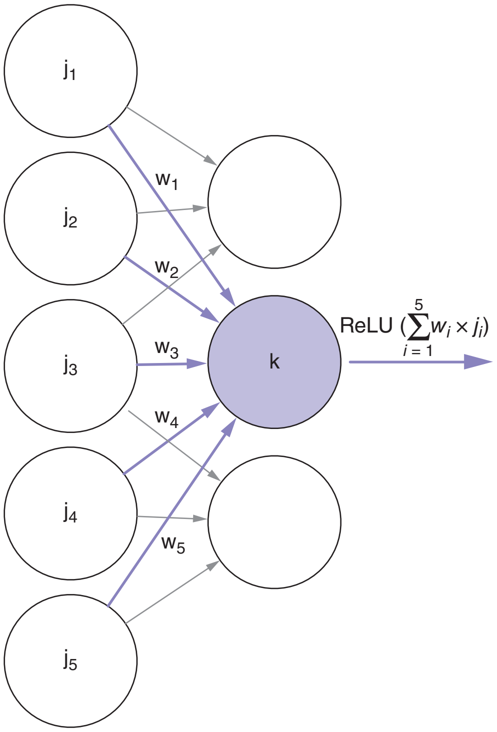

FIGURE 501-2 The artificial neuron, the building block of deep learning models. Neuron k accepts weighted input from neurons in the preceding layer and applies a function [e.g., ReLU(x) = max(0,x), the rectified linear function] to the weighted sum of inputs (often with a bias term, not shown). During training, the weights are iteratively refined to improve data fitting.

PRACTICAL CONCEPTS IN TRAINING A MODERN MACHINE LEARNING MODEL

In this section, we provide a brief overview of practical concepts that emerge when training a modern machine learning model to help readers understand the constraints and choices that confront machine learning practitioners.

Often the first task is to define what a "good" prediction is by specifying a "loss function," which quantifies the error between a model's prediction and the true label **(Table 501-3)**. There are many choices for this function. Linear regression uses quadratic loss; other examples include the cross-entropy and 0-1 loss. Given a loss function, the stochastic gradient descent method coupled with the backpropagation algorithm quantifies how to alter the adjustable weights in the network to optimize the loss function iteratively as labeled examples are provided to the network, either one by one or in batches. The weights themselves are often initialized carefully and frequently transferred over from another task. Most machine learning practitioners use specialized hardware called graphics processing units (GPUs) to perform these calculations. As with most computer hardware, GPUs range dramatically in performance and cost. Software frameworks like PyTorch and TensorFlow automate much of the otherwise cumbersome training process, abstracting away in a dozen lines what might have previously taken a team of machine learning engineers months to build. The machine learning community places great emphasis on generalization through near universal practices like widely available benchmark datasets and splitting training and testing data, with performance measures such as the area under the receiver operating curve (AUC) computed in held-out test data to improve estimates (Table 501-3).

APPLICATIONS OF MODERN MACHINE LEARNING TO CLINICAL MEDICINE

Two major classes of applications have dominated recent machine learning applications in medicine: computer vision and natural language processing. We review some of the recent applications below, highlighting the breadth of problems across clinical specialties being tackled and some emerging new directions.

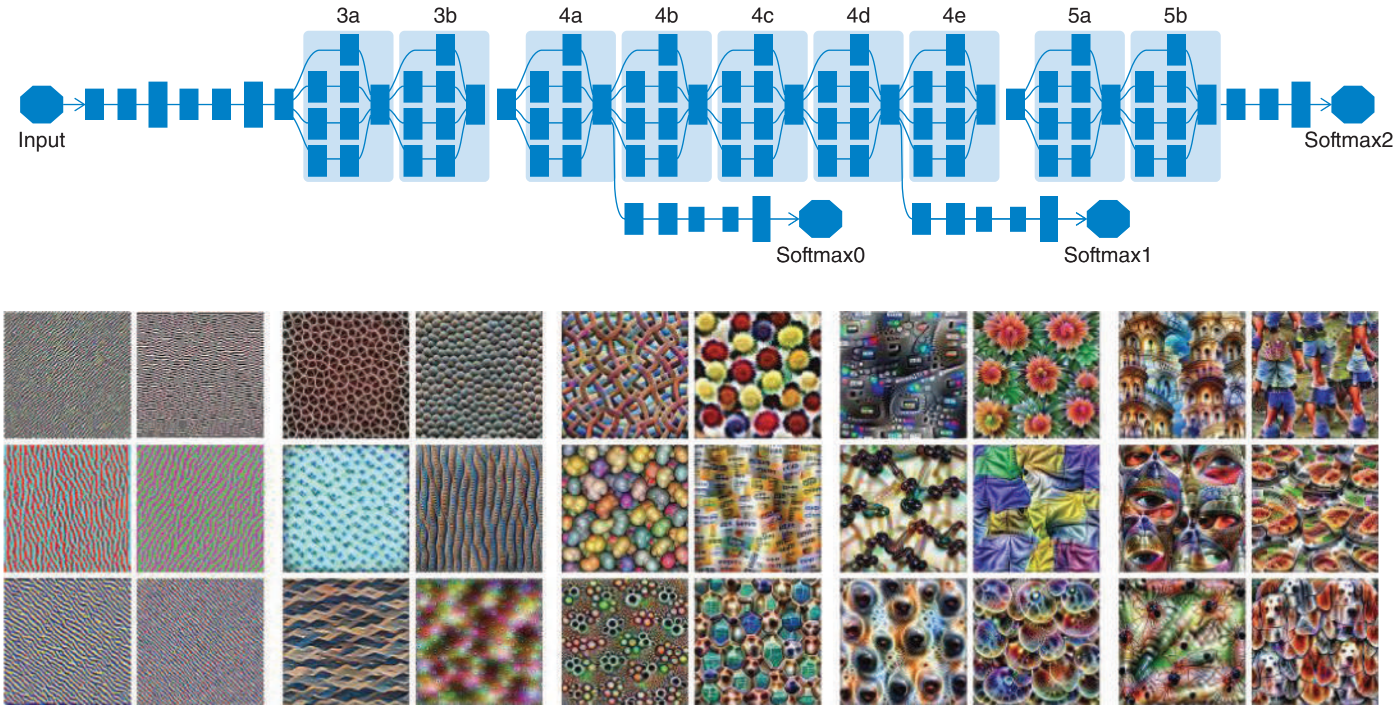

Edges (layer conv2d0) **Textures** (layer mixed3a) **Patterns** (layer mixed4a) **Parts** (layers mixed4b & mixed4c) **Objects** (layers mixed4d & mixed4e)

FIGURE 501-3 Deep learning models learn rich hierarchical representations. Visualization of what a convolutional neural network "sees" as it is processing images of common objects from the ImageNet dataset. The initial layers learn low-level features like edges, and the higher layers learn patterns and objects. *(Reproduced from distill.pub by C Olah et al: Feature visualization. Distill, 2017. https://distill.pub/2017/feature-visualization/.)*

COMPUTER VISION

Medical computer vision applications have been dominant in the medical machine learning literature over the past decade. Convolutional neural networks and newer vision transformer (ViT) models are well-suited to exploit the spatial structure in medical imaging data and are able to learn detailed representations of input images, yielding systems that often perform on par with or better than expert physicians at select tasks. A well-known application of convolutional neural networks for medical imaging was published by Gulshan and colleagues in the journal *JAMA* in 2016. The authors trained a convolutional neural network using 128,175 ophthalmologist-labeled retinal fundus photographs to develop an automated diabetic retinopathy detection system, achieving performance on par with expert ophthalmologists, with an AUC of 0.99 in two separate validation datasets. In a separate study by De Fauw and colleagues, 14,884 three-dimensional retinal optical coherence tomography scans were used to train a deep learning model that could make referral suggestions and performed with an accuracy at or superior to eight clinical experts who graded the same scans, with an AUC in test data >0.99 for urgent referral versus other referrals. Some machine learning applications in ophthalmology have already received approval from the U.S. Food and Drug Administration including the IDx-DR "device" to classify more than mild diabetic retinopathy.

TABLE 501-3 Practical Concepts for Training a Deep Learning Model

CONCEPT	DEFINITION	EXAMPLES
Loss function	Mathematical function that quantifies discrepancy between predicted label and true label	Cross-entropy, quadratic, 0–1
Back-propagation	Algorithm to compute the direction of the loss function with respect to changes in the adjustable weights ("gradient")	—
Graphics processing unit (GPU)	Specialized computer hardware to speed up the many matrix calculations involved in training a neural network	NVIDIA Tesla H100
Train-test split	How data are divided to ensure fair estimates of model performance after training	70% training, 30% test
Area under the receiver operating characteristic curve (AUC)	Common performance metric for evaluating binary classification models; 0.5 = random, 1.0 = perfect	—
Deep learning framework	Computational framework for efficiently performing matrix (tensor) calculations for training deep learning models	TensorFlow, PyTorch, Keras

Outside of ophthalmology, computer vision applications have been numerous across the many other specialties that rely on imaging data. For example, dermatologist-level classification of skin cancer was achieved in a study by Esteva and colleagues published in *Nature* in 2017. The authors trained a convolutional neural network to distinguish between keratinocyte carcinomas and benign seborrheic keratoses as well as between malignant melanomas and benign nevi. The authors concluded that the model performed at the level of the 21 board-certified dermatologists against which it was tested.

The uses of machine learning models in radiology are numerous as well, with applications including the detection of pneumonia from chest x-rays, identification of pancreatic cancer from computed tomography scans, and fast, automated segmentation of cardiac structures from cardiac magnetic resonance imaging as well as echocardiography.

Specialized deep learning architectures such as the U-Net architecture by Ronneberger and colleagues have become especially popular in the medical computer vision community. Architectures are often designed for specific imaging tasks (e.g., image segmentation) or specialized data types (e.g., three-dimensional images or videos). New frontiers of computer vision research in medicine include semi-supervised learning approaches to take advantage of extensive unlabeled data available at hospitals, particularly given the practical difficulty and cost for an individual researcher to obtain large expert-labeled datasets, and multimodal learning, where machine learning algorithms learn to reason simultaneously over different types of data (e.g., images and text).

NATURAL LANGUAGE PROCESSING

Like computer vision, natural language processing has been transformed by modern machine learning approaches, particularly the subfield of deep learning. Traditional deep learning approaches for natural language processing relied on a specialized architecture called recurrent neural networks. Contemporary deep learning methods often leverage the transformer model (Table 501-2), which is well-suited to exploit the structure of natural language and other text. Text-processing machine learning models have been successfully applied to analyze physician notes in the electronic health record, detect depression symptom severity from spoken language, and scribe patient-physician visits. For example, a study by Rajkomar and colleagues analyzed electronic health record data from 216,221 adult patients to predict in-hospital mortality, 30-day unplanned readmission, and discharge diagnoses, among other outcomes, performing at high accuracy, with an AUC of 0.93–0.94 for predicting in-hospital mortality. Importantly, much of the progress in medical natural language processing has stemmed from the widespread availability of datasets, including, for example, the Medical Information Mart for Intensive Care (MIMIC) dataset.

Many specialized deep learning architectures have been developed for natural language processing applications, including the analysis of electronic health record data, using both supervised (e.g., recurrent neural network) and unsupervised (e.g., variational autoencoder) approaches. Domain-specific language representation models have been developed for the purpose of biomedical text mining, serving as a substrate for many downstream natural language processing tasks.

Since ChatGPT was introduced in 2022, large language models including GPT-4 have rapidly been applied to diagnostic reasoning, health care documentation, and many other text-based tasks across medical specialties. The wide-ranging linguistic abilities and performance of these models across myriad tasks have surprised many physicians and machine learning practitioners alike. In a study by Kanjee and colleagues published in *JAMA* in 2023, the authors evaluated the general diagnostic reasoning abilities of GPT-4 on challenging medical cases published as part of the *New England Journal of Medicine* Clinicopathological Conferences (CPCs), also known as the Case Records of the Massachusetts General Hospital. On these challenging cases, GPT-4, which was not trained specifically for medical diagnostic reasoning tasks, included the correct diagnosis as part of its differential diagnosis in 64% of the 70 cases assessed, a surprisingly high accuracy.

OTHER APPLICATIONS

While medical computer vision and natural language processing tasks have been the focus of newer deep learning models due to the extensive structure of imaging and text data, many other application classes exist. For example, cardiologist-level performance has been achieved in deep learning approaches for detecting arrhythmias from ambulatory electrocardiograms, standing in contrast to the rule-based algorithms used traditionally to interpret electrocardiographic signals. In genomics, investigators have analyzed tumor genomes with machine learning methods to predict better survival using both deep learning and other machine learning approaches. Machine learning methods have also been used to characterize the deleteriousness of single nucleotide variants in DNA. Many other applications of machine learning to new patient data streams are emerging, for example, machine learning applied to wearables (e.g., smartwatches).

CONCLUSION

Modern machine learning offers a powerful set of techniques to learn feature representations directly from data, already performing on par with expert physicians on select tasks. If carefully trained and judiciously applied to key areas of clinician workflow, the representational power of new machine learning methods makes them likely to touch every area of clinical practice.

FURTHER READING

Gulshan V et al: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316:2402, 2016.

Haug CJ, Drazen JM: Artificial intelligence and machine learning in clinical medicine. N Engl J Med 388:1201, 2023.

Kanjee Z et al: Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. JAMA 330:78, 2023.

Krizhevsky A et al: 2012 NeurIPS paper: Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012.

LeCun YA et al: Deep learning. Nature 521:436, 2015.

Olah C et al: Feature visualization. Distill. 2017. *https://distill.pub/2017/feature-visualization/*.

Rajkomar A et al: Machine learning in medicine. N Engl J Med 380:1347, 2019.

Ronneberger O et al: U-Net: Convolutional networks for biomedical image augmentation, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, 2015, pp. 234–241.

Topol EJ: High-performance medicine: The convergence of human and artificial intelligence. Nat Med 25:44, 2019.

502 Metabolomics

Jared R. Mayers, Mathew G. Vander Heiden

Metabolism, loosely defined, represents the sum of all biochemical reactions involving small molecules with a molecular mass of ≤1000 Da within a given tissue, cell, or fluid. These small molecules are collectively referred to as *metabolites* and are involved in the biochemical processes used to create macromolecules and fulfill the energy needs of a cell or organism. *Metabolomics*, then, represents the measurement of metabolites, either qualitatively or quantitatively, often as a way to gain insight into the metabolism of a cell, tissue, or organism. No one experimental approach can characterize metabolism in its entirety; metabolomics instead strives to measure a portion of the *metabolome*, which consists of all metabolites in a given biological sample at a given time.

A link to a time-specific context is common to all "-omics" techniques, but is particularly important in metabolomics. As metabolic processes are highly connected and interdependent, with individual metabolites often being involved in multiple pathways, levels of a specific metabolite can vary in response to an alteration in either the production or the consumption of that metabolite. Because significant changes in metabolite levels can occur over a very short time frame, the levels measured can be sensitive to perturbations either upstream or downstream of the measured metabolite in a pathway. This sensitivity can make measurement challenging, but it also makes metabolomics a powerful tool with which to assess either acute or chronic changes in cells or tissues. Indeed, the metabolome can be quite dynamic and reflective of the current condition of the material being assessed, as it ultimately represents an integration of outputs from the genome, epigenome, transcriptome, and proteome **(Fig. 502-1)**.

APPROACHES AND SAMPLING CONSIDERATIONS

UNTARGETED AND TARGETED METABOLOMICS

There are two distinct approaches to measuring metabolites in biological materials: untargeted and targeted metabolomics. These strategies differ in whether a predetermined subset of metabolites is intentionally sought in a sample, with the choice of approach dictated by the question under investigation. Regardless of the method utilized, it is important to recognize that no single metabolomics technique is comprehensive. Technical considerations heavily influence metabolite measurement, even with untargeted metabolomics, and no one method is able to capture the entire metabolome. In this respect, metabolomics contrasts with some other -omics techniques, like genomics or transcriptomics—i.e., in metabolomics, if something is not measured, its absence cannot necessarily be assumed.

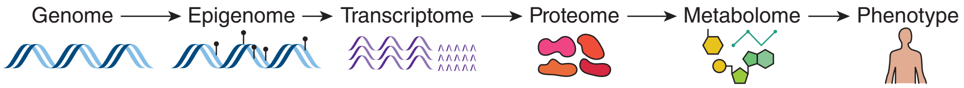

FIGURE 502-1 The metabolome is downstream of the outputs measured by other "-omics" technologies. Thus, the state of the metabolome can more closely reflect clinical and experimental phenotypes.

Untargeted Metabolomics Untargeted metabolomics is the comprehensive analysis of as many measurable analytes in a sample as possible, irrespective of their identity **(Fig. 502-2)**. Among the benefits of this approach is that it is agnostic in its measurement of the metabolome. Thus, it allows for the discovery of novel or unexpected molecules for further study. Coverage of the metabolome in an untargeted approach is influenced by the techniques used for sample preparation, metabolite separation prior to detection, and the inherent sensitivity and specificity of the analytical technique(s) employed (see "Metabolomics Technologies," below).

A major drawback of untargeted metabolomics is that molecules of interest can be measured with less confidence or missed entirely because this approach carries an inherent bias toward the detection of high-abundance molecules. Handling and interpretation of data also represent a major challenge, as each sample run generates large amounts of data whose analysis can be both complicated and time consuming. Identifying each metabolite measured requires database searching, and further experimental investigation is often needed to confirm the exact identity of a signal of interest. Finally, in most cases, this technique yields only relative metabolite quantification, thereby rendering it most useful for comparisons between biological samples.

Targeted Metabolomics Targeted metabolomics involves the measurement of a predefined group of chemically characterized metabolites—typically dictated by a hypothesis or predetermined platform—with the aim of covering a select portion of the metabolome. The metabolites measured represent only a subset of those that would be measured by an untargeted approach; thus, a targeted approach generates a much smaller data set in which individual metabolites are detected with higher confidence (Fig. 502-2). Because the identity of each signal is known in advance, standards can be added to provide absolute quantification of each metabolite measured in the sample, although the use of targeted metabolomics to compare relative metabolite levels across samples is common. In addition, sample preparation and chromatographic separation before measurement can be optimized to improve detection of specific metabolites, enabling assessment of less abundant molecules.

The key downside of targeted metabolomics is that information is gained about only those metabolites targeted by the analytical method.

SAMPLING CONSIDERATIONS

Regardless of the approach used, it is important to consider potential sources of error that can influence the conclusions drawn from a metabolomic analysis. Because of the dynamic nature of the metabolome, numerous biological confounders inherent to the samples themselves can affect levels of the metabolites measured. For this reason, the inclusion of controls or reference populations to account for these confounders can be critical for data interpretation. Established biological confounders for patient-derived material include age, sex, body mass index, time of day collected, fasting status and/or dietary differences, and comorbid conditions such as diabetes or smoking. For example, metabolites commonly altered with respect to aging are those in antioxidant and redox pathways as well as breakdown products of macromolecules. Sex differences influence a number of different metabolites, most prominently those involved in steroid and lipid metabolism. Perhaps it is not surprising that diet can also affect the metabolome, and fasting has been shown to impact almost every category of metabolite frequently measured in biological fluids.

Differences in sample handling and processing also influence metabolite measurements. Work using metabolomics to analyze material from large prospective cohort studies has shown that changes in metabolite levels introduced by sample handling can lead to falsely positive associations between specific metabolite changes and disease risk. Specific considerations include the large geographic area of distribution from which patients are drawn—e.g., a sample, such as blood, is collected locally and then exposed to variable conditions before being sent to a central lab for further processing. Moreover, because of the costs associated with obtaining and storing samples, often only one sample is available for each individual.

Time is a key variable in metabolite measurements, and efforts to assess the impact of sample handling and processing have led to improved analysis pipelines. For example, comparison of metabolites measured in samples undergoing immediate versus delayed processing can provide insight into those metabolites most affected by pre-processing storage under varying conditions. More specifically, because metabolism occurs on a very rapid time scale, some metabolite levels will continue to change after sample collection even if the sample is stored on ice. Therefore, metabolism is ideally halted or "quenched"

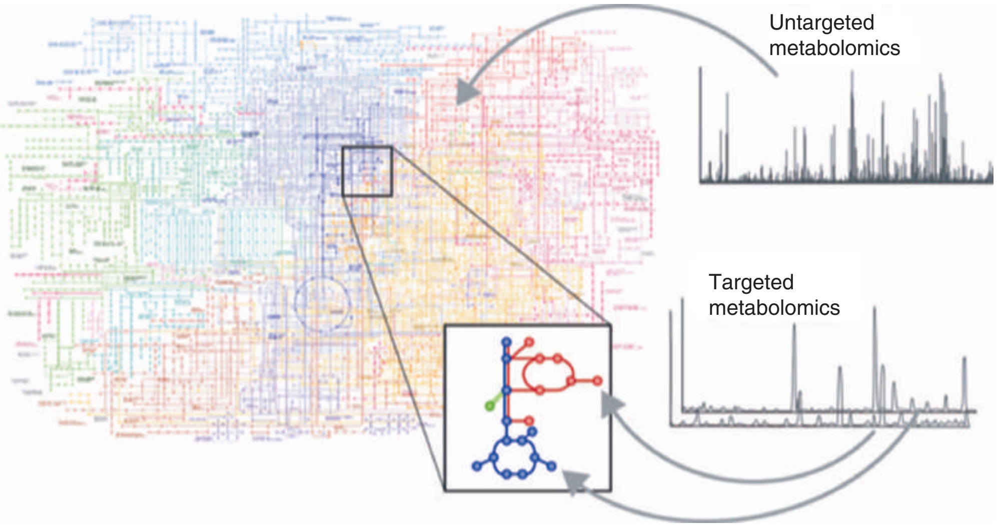

FIGURE 502-2 *Untargeted* metabolomics strives to measure as much of the metabolome as possible within a given biological sample, whereas *targeted* metabolomics focuses on measuring a predetermined subset of the metabolome. In untargeted metabolomics, a large number of signals corresponding to metabolites is generated, and further investigation is often necessary to assign a particular signal to a specific metabolite. Targeted metabolomics allows investigators to definitively measure signals that correspond to specific metabolites of interest.

immediately via rapid freezing or chemical extraction, but practical considerations involved in the collection of material from patients can sometimes make rapid quenching impossible. Therefore, focusing analysis on only those metabolites that are less sensitive to change due to delays in processing time may be important to gain biological insight.

Sequential metabolomic analyses of the same type of biological material from a patient can explore how metabolite levels vary over time. It is interesting that, when measured, many metabolites are found to be relatively stable. However, the extensive variability exhibited by some metabolites indicates that findings involving those metabolites should be interpreted with caution.

Finally, the method of sample processing can affect which metabolites are extracted from the material and thus influence what is measured.

METABOLOMICS TECHNOLOGIES

Metabolomics relies heavily on the intersection of instrumentation, software, and statistical and computational approaches for measurement of metabolite levels and downstream data analysis. While the development of new and emerging techniques to assess the metabolome is ongoing, the current, clinically applicable approaches can be separated into two broad categories: nuclear magnetic resonance (NMR)–based approaches and chromatography/mass spectrometry (MS)–based approaches. Each of these two approaches has its own set of advantages and disadvantages.

NUCLEAR MAGNETIC RESONANCE

NMR is a technique that, at its core, exploits intrinsic magnetic properties of atomic nuclei to generate data. Nuclei with an odd total number of protons and neutrons (such as ^{1}H, ^{13}C, ^{15}N, and ^{31}P) have a non-zero spin, and this spin generates a magnetic field that can interact with externally applied electromagnetic fields. NMR places compounds into a magnetic field that induces the smaller magnetic fields to align with the larger one. Samples are then exposed to a perpendicular electromagnetic field; the frequency of electromagnetic radiation needed to flip the spin of a nucleus in the exact opposite direction represents the frequency at which an atom "resonates" and can be measured. The resonance frequency of a given atom is affected by adjacent atoms and is ultimately unique for a given arrangement of atoms (i.e., each metabolite). This distribution or "spectrum" of signals is measured and recorded in an NMR experiment.

With respect to clinical applications, the primary benefits of NMR-based approaches are that they are nondestructive and can be performed on living samples, such as patients, cells, or tissues. They are also highly reproducible and require minimal sample preparation. Measurements are necessarily quantitative as the signal measured directly reflects concentration. These features ensure that multiple, comparable measurements can be made in a given sample either at a single point in time or across time. In addition, given that spins of different elements require sufficiently disparate resonance-inducing radiofrequencies in order to be entirely distinguishable, multiple elements can be assessed in a sample; this feature allows multidimensional cross-referencing of signals such as hydrogen and carbon. In an untargeted analysis, these multidimensional data can then be used for definitive metabolite identification, with comparison of results to known databases in which spectra for many metabolites in the human metabolome have been systematically recorded.

TABLE 502-1 Comparison of Nuclear Magnetic Resonance (NMR)-Based and Mass Spectrometry (MS)-Based Approaches to Metabolomic Analyses

FEATURE	NMR	MS
Reproducibility	High	Lower
Sensitivity	Low (low μM)	High (low nM)
Selectivity	Untargeted	Targeted >> untargeted
Sample preparation	Minimal	Complex
Sample measurement	Simple: single prep	Multiple preps
Metabolites per sample	50–200	>1000
Identification	Easy, using one- or two-dimensional databases	Complex; need standards and additional analyses
Quantitation	Inherently quantitative; intensity proportional to concentration	Requires standards because of varying ionization efficiency
Sample recovery	Easy, nondestructive	No
Living samples	Yes	No

Despite all these benefits, the primary challenge of NMR-based approaches is a lack of sensitivity. Because the time required to detect a signal is proportional to concentration, assessment of less abundant species is impossible or impractical. For example, while a typical NMR-based metabolomics analysis will return data on up to a couple of hundred metabolites at concentrations of >1 μM, the MS-based approaches discussed below can distinguish more than 1000 metabolites at concentrations one to two orders of magnitude lower **(Table 502-1)**.

CHROMATOGRAPHY/MASS SPECTROMETRY

A distinguishing feature of chromatography/MS–based approaches is that a multistep process that destroys the material is necessary to generate a sample for analysis. In addition, each step of the sample preparation process involves decisions that influence the metabolites measured at the time of analysis. In general, once a sample to be analyzed is prepared, that material is subjected to a combined chemical and temporal separation of compounds via chromatography, with the output delivered to a device for performance of mass-based detection (technically, measurement of a mass-to-charge [m/z] ratio)—i.e., mass spectrometry. Finally, data collected by the mass spectrometer are analyzed **(Fig. 502-3)**.

Sample Preparation Although occasionally a part of NMR-based metabolite detection protocols, MS-based approaches almost uniformly require an initial sample-preparation phase called *extraction*. This technique destroys the original sample by partitioning metabolites into distinct immiscible phases, such as polar and nonpolar. These

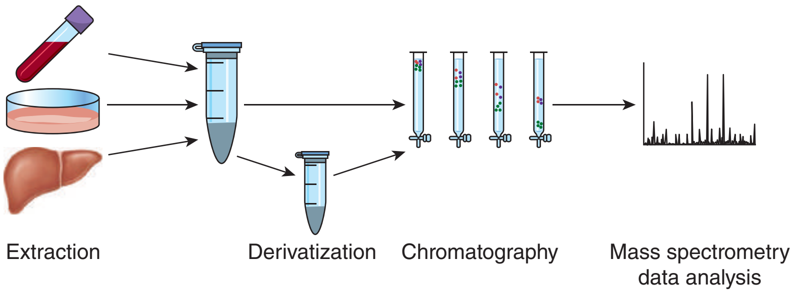

FIGURE 502-3 Metabolite measurement by chromatography/mass spectrometry–based approaches involves multiple steps, and decisions made at each step influence what is measured. First, metabolites are extracted from a biological sample in a manner that is destructive of the original sample. This process stops biochemical activity and creates metabolite samples that can be analyzed, sometimes after a chemical derivatization step that alters a subset of metabolites in a manner that facilitates their downstream analysis. Second, metabolites in the sample are separated via chromatography. Finally, the chromatographically separated compounds are analyzed by mass spectrometry. Each signal detected corresponds to a metabolite's characteristic mass per unit charge while the amplitude of that signal reflects the abundance.

phases are then mechanically separated and processed further for analysis. Given the nature of this extraction process, it is critical to determine in advance the general class of metabolites to be measured. This information will help to determine the optimal extraction protocol for specific types of metabolites of interest and to shape further downstream decisions regarding the chromatography/MS technique that also influences metabolite detection. In addition, depending on the metabolites to be analyzed and the method of separation and/or analysis used, extracted samples sometimes are processed further in a preparative step called *derivatization*: extracted metabolites are chemically modified by the addition or substitution of distinct, known chemical moieties that facilitate separation or detection of types of metabolites. By changing the chemical properties of metabolites, derivatization may improve stability, solubility, or volatility or facilitate separation from closely related compounds, enhancing measurement of specific metabolites.

Chromatography Chromatography is a ubiquitous approach used in chemistry for the separation of complex mixtures. The mixture of interest in a mobile phase is passed over a stationary phase such that compounds in the mixture interact with the stationary phase and transit through that stationary phase at different speeds, allowing their consequent separation. Two general types of chromatography are typically used in metabolomics.

LIQUID CHROMATOGRAPHY Liquid chromatography–mass spectrometry (LC-MS) is the most commonly used approach in MS-based metabolomics. In this case, chromatography is characterized by a mobile phase that is a liquid and a stationary phase that is a solid. In liquid chromatography in particular, the choice of the solid and liquid phases can dramatically influence the types of compounds separated for input into the mass spectrometer. In general, LC-MS metabolomics is highly sensitive and versatile in allowing detection of a broad range of metabolites. A downside, however, is variability in exact separation timing, especially between different instruments; which metabolites are measured is impacted by the chromatography used and how well molecules are separated.

GAS CHROMATOGRAPHY Gas chromatography–mass spectrometry (GC-MS) involves chromatography in which the mobile phase is a gas. In contrast to LC-MS, GC-based approaches have a narrower range of applications because only volatile metabolites that enter a gaseous phase are separated. When combined with appropriate derivatization, GC-MS is a robust way to detect many organic acids, including amino acids, and molecules of low polarity, such as lipids. GC-MS is more reproducible than LC-MS across platforms and requires less expensive instrumentation and less specialized training, but it also typically measures a much more restricted range of metabolites in a sample than does LC-MS.

Mass Spectrometry Once the metabolites in a sample have been separated by chromatography, they are sent into the mass spectrometer for analysis and measurement. The first step in this stage of the process is to generate charged ions, as mass spectrometers measure compounds on the basis of their m/z ratio. Charge can be imparted through various techniques, although most commonly it is attained by either applying a high voltage to a sample or striking it with a laser.

A number of different types of mass spectrometer can be employed for metabolomics. Three of the most commonly available types are discussed below.

TANDEM MASS SPECTROMETRY Tandem MS relies on three sets of quadrupole magnets arranged in series. The power of this arrangement lies in its specificity through two sequential mass analyses of the same starting compound. In the first quadrupole, the “parent” or full ion is measured before being bombarded by an inert gas in the second quadrupole; this process fragments the compound into characteristic smaller “daughter” ions. The third quadrupole then measures these daughter ions.

TIME-OF-FLIGHT MASS SPECTROMETRY While there are multiple types of time-of-flight (TOF) mass spectrometers, they all operate on similar principles. Most simply, lighter metabolites travel faster and heavier metabolites travel more slowly. TOF machines have high mass accuracy and sensitivity while also acquiring data quickly.

ION TRAP MASS SPECTROMETRY Ion trap mass spectrometers, of which the orbital trap is a subtype, offer perhaps the highest degree of flexibility when it comes to MS-based metabolomics. In general, these machines can select for a specific mass range of metabolites at multiple levels, first by filtering with a single quadrupole and then by trapping and accumulating metabolites of a particular mass or range of masses. This accumulation can be applied to low-abundance compounds, allowing increased sensitivity. It also allows repeated fragmentation of metabolites (called MS^n) to produce characteristic “daughter” ions, increasing the specificity of the analysis. Given this versatility coupled with high mass accuracy, the development of these machines is advancing rapidly; however, access to the latest versions can often be limited by cost.

CURRENT CLINICAL APPLICATIONS

Tests to assess small molecules are ubiquitous and well established throughout medicine. These include assays to measure select metabolites of known clinical relevance, such as glucose, lactate, and ammonia. Of note, many standard tests assess these metabolites one at a time; however, metabolomics can allow the assessment of many metabolites in a sample and provide more information on metabolic state at a given point in time. In some cases, metabolomics is used to detect molecules for which there is not a robust single analyte test or when multiple species measured in a sample might provide new information. Here we will focus specifically on some applications of metabolomics techniques in current clinical practice.

■ MAGNETIC RESONANCE SPECTROSCOPY

Magnetic resonance spectroscopy (MRS) is an adaptation of magnetic resonance imaging (MRI), a widely used technology in clinical practice. MRI, at its core, is essentially proton (^1H) NMR with the resulting data rendered spatially to generate an image. Recall that NMR is nondestructive and can be applied to living samples. MRS, then, is a capability built into almost every MRI machine. In practice, radiologists can focus on specific volumes of interest within a patient’s imaging and perform additional sequences to obtain an NMR spectrum in that space that can allow for the identification and quantification of specific metabolites in that space. With this approach, several different metabolites across diverse classes, including lipids, sugars, and amino acids, can be measured at a given time.

Extensive work has correlated different biological processes with altered levels and/or ratios of metabolites measured via MRS. One well-established application is in the diagnosis of brain masses. More specifically, *N*-acetylaspartate (NAA) is an amino acid derivative that is abundant in neurons, whereas choline is a metabolite whose level, as measured by MRS, correlates with cellularity and/or proliferation. Thus, an increase in the ratio of choline to NAA (and even loss of NAA signal entirely) correlates with cancer; tumors biologically are associated with the properties of increased cellularity from proliferation and the concurrent exclusion of normal neurons. A different process—for example, a brain abscess—does not result in increased choline levels (which instead may actually decrease), but does exclude neurons, resulting in an isolated NAA decrease. Metabolites such as lactate can also be helpful, depending on the clinical context, in providing insight into the metabolism of a tumor or identifying areas of early hypoxic brain injury after a stroke. Finally, among the several amino acids that can be measured, high levels of glutamine/glutamate can be helpful in a patient with altered mental status as changes in these amino acids are associated with hyperammonemia. (Glutamate serves as the central nervous system sink for ammonia, generating glutamine in the process.)

■ NEWBORN SCREENING PROGRAMS

Newborn screening programs are used to identify diseases within the first few days of life such that they can be treated or managed with early intervention. Among the classes of disease targeted by newborn screening programs are many inborn errors of metabolism, which often lead to changes in the levels of specific metabolites in blood or urine. One of the first newborn screening programs tested for phenylketonuria, which results from the inability to metabolize phenylalanine resulting in high blood and urine levels of particular metabolites. Since that time, the panel used by programs throughout the United States and around the world has expanded dramatically. The general protocol is to collect a blood sample from infants in the first few days of life (often by heel prick on a piece of paper). These samples are sent to a central lab for analysis, which typically includes metabolomics measurements with targeted LC–tandem MS. Specific inborn errors of metabolism are suggested by abnormal levels of a given metabolite or set of metabolites.

METABOLITE MEASUREMENTS IN CHILDREN AND ADULTS

Outside the window of newborn screening, direct clinical measurement of metabolite levels is also used in pediatric and adult patients. In these cases, clinical samples such as serum, cerebrospinal fluid, or urine are typically subjected to targeted LC–tandem MS to measure metabolites such as amino acids, acylcarnitines, and fatty acids. These measurements can help diagnose milder cases of inborn errors of metabolism that may have been missed by newborn screening. They can also help identify secondary metabolic defects, such as those that are related to nutritional deficiencies or are acquired in the setting of additional pathology. For example, these measurements are useful in determining the etiology of noncirrhotic hyperammonemia exposed by a catabolic stressor such as sepsis in a patient with a previously unknown subclinical or acquired urea-cycle defect.

MS-based metabolomics is used by various athletic organizations for detection of metabolites associated with banned substances and by the pharmaceutical industry for assessment of levels of pharmaceuticals and their metabolites in both blood and tissues. Such analyses can provide key pharmacokinetic information to guide drug dosing and illuminate toxicology. These approaches can also be useful in clinical practice. For example, chronic pain and its management remain a challenge, and the sequelae of opiate/opioid use and abuse are of concern to many providers, their patients, and their patients' families. Therefore, many electronic medical records systems strive to ensure appropriate and consistent patient access to pain medications, while providers may need a means to ensure that patients are adhering to their prescribed regimens. One way to monitor drug use is to perform targeted LC–tandem MS for detection of specific drug metabolites in patients' urine. This approach is more sensitive than first-generation immunoassays and can detect a range of metabolites associated with other drugs beyond the one prescribed. Given that the first-generation immunoassays also often rely on confirmatory MS testing, upfront metabolomics reduces lab turnaround time and may also reduce costs by limiting multiple tests on the same sample.

EMERGING AND EXPERIMENTAL CLINICAL APPLICATIONS

The current clinical applications of metabolomics are largely limited to the indications described above. However, ongoing efforts are aimed at expanding the use of metabolomics for detection of biomarkers that can help with disease diagnosis or prognostication.

METABOLITES AS BIOMARKERS OF DISEASE

There has been increasing work in prospective human cohort studies on the use of metabolomics, primarily MS-based approaches, to empirically identify small groups of metabolites whose altered levels are associated with the development or progression of disease. Efforts to characterize these "metabolic signatures" have been focused primarily on common, multifactorial diseases such as diabetes, cardiovascular disease, and various cancers that are well represented in large prospective cohort studies. These studies have, for example, identified altered levels of amino acids that are associated with a future diagnosis of diabetes or pancreatic cancer. Similar efforts have proliferated across conditions ranging from chronic lung diseases to neurologic/developmental disorders.

Additional efforts have been made to assess the metabolome in patient samples at the time of an acute presentation. Because altered metabolite levels can be associated with a specific clinical diagnosis and/or outcome, the idea is to identify a metabolite signature that facilitates diagnosis or provides prognostic information. This approach has been studied, for example, in the context of sepsis and septic shock, in which blood lactate levels are assessed in combination with the use of clinical tools such as the Acute Physiology and Chronic Health Evaluation (APACHE II) or the Sequential Organ Failure Assessment Score (SOFA). More recent efforts have identified a strong association between mortality and certain modified amino acids linked to mitochondrial dysfunction, highlighting a potential mechanistic link between sepsis pathogenesis and metabolic alterations.

One key limitation in all of these studies is that researchers are primarily assessing correlations between blood plasma metabolite levels and complex, multisystem diseases. It remains difficult to obtain a biological understanding of the mechanisms driving these changes or, even more simply, the primary tissue source(s) of these alterations from human data alone, without further experimentation in model systems.

REFINING DIAGNOSIS AND PREDICTION OF DRUG SUSCEPTIBILITY

In contrast to the above-described use of metabolomics-based approaches in multifactorial diseases, the application of these approaches in some specific contexts can yield an immediate diagnosis and suggest actionable therapeutic interventions. One specific example in oncology involves an understanding of the pathogenesis of oncogenic mutations in the metabolic enzyme isocitrate dehydrogenase (IDH) isoforms 1 and 2. The normal function of these enzymes is to interconvert isocitrate and α-ketoglutarate; however, cancer-specific point mutations in these enzymes alter the enzymes' function in a manner conferring neomorphic activity that converts isocitrate into 2-hydroxyglutarate (2-HG). 2-HG is a metabolite that is typically present only at very low levels in cells, but when mutant IDH protein is present, 2-HG is produced and accumulates to high levels. Elevation of 2-HG can promote changes that directly contribute to malignancy; IDH mutations and 2-HG accumulation are found in several human cancers, including specific clinical subsets of acute myeloid leukemia and glioma. Given the unique and specific accumulation of 2-HG in these mutant tumors, detection of this metabolite by LC-MS and NMR-based approaches has been studied both for diagnostic purposes and as a means of assessing drug response. For example, researchers have applied MRS-based approaches to assess the accumulation of 2-HG in gliomas, as this finding can noninvasively identify patients with an IDH-mutant subset of this cancer **(Fig. 502-4)**. This diagnosis provides prognostic information and determines if a patient could benefit from drugs targeting mutant IDH that have been shown to benefit patients with IDH-mutant gliomas. In principle, metabolomics may identify other disease biomarkers to aid with diagnosis or therapy assessments in similar ways.

PHARMACOMETABOLOMICS

The previous example positions metabolomics as a possible mechanism for achieving a more personalized approach to medicine. The emerging field of pharmacometabolomics aims to take personalization further by making this approach more widely applicable across drugs and disease states. The general idea is to link pharmacokinetics (PK) and pharmacodynamics (PD) data with baseline metabolomic profiling, with the goal of generating a predictive model for individual PK and PD responses based on a naïve patient's metabolomic profile. Ideally, this approach would allow clinicians to take a baseline set of measurements and then—a priori—choose a specific dose of a specific drug to produce the desired effect in that specific patient. If successful, this method could limit both prolonged titration of medications and medication switching, dramatically shortening and simplifying the current approach to medical therapy.

EMERGING TECHNIQUES

While efforts to improve the existing capabilities discussed above are ongoing, innovations in instrumentation and computation are allowing collection and analysis of metabolite information that previously was not possible.

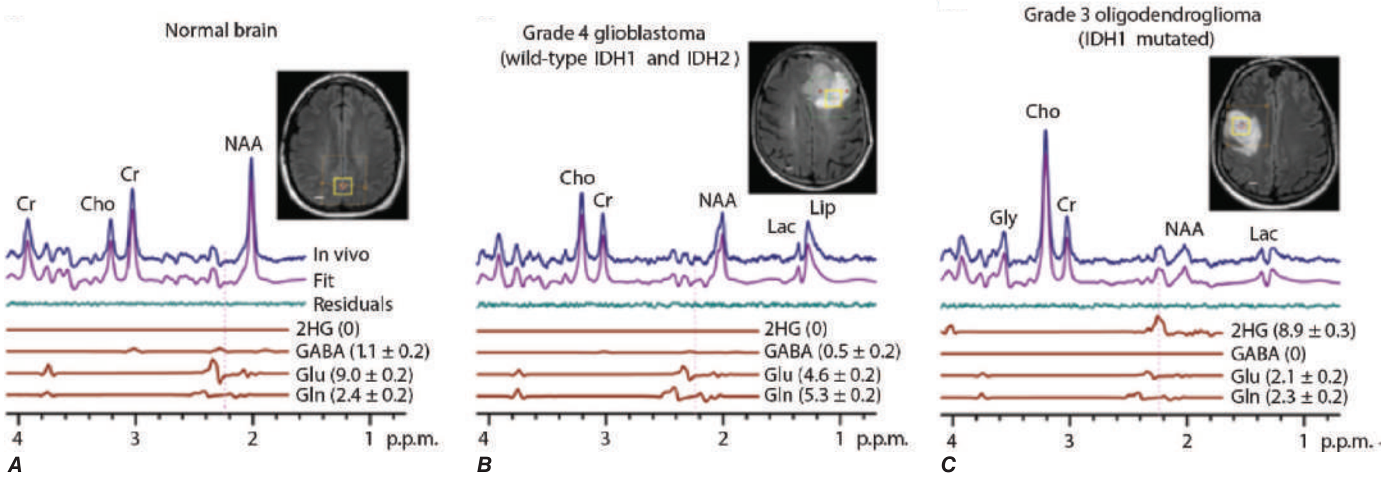

FIGURE 502-4 In vivo ^{1}H spectra and analysis demonstrating 2-hydroxyglutarate (2-HG) detection in isocitrate dehydrogenase (IDH)-mutant brain tumors. *A–C.* In vivo spectra from normal brain (***A***) and tumors (***B–C***) are shown. Components of 2-HG, γ-aminobutyric acid (GABA), glutamate, and glutamine are displayed. Measurement location is indicated by *yellow box* (voxel). 2-HG is seen only in mutant IDH brain tumors, but not normal brain or wild-type tumors. Shown in brackets is the estimated metabolite concentration (mM) ± standard deviation (s.d.). Cho, choline; Cr, creatine; Glu, glutamate; Gln, glutamine; Gly, glycine; Lac, lactate; Lip, lipids. Scale bars, 1 cm. *(Reproduced with permission from C Choi et al: 2012.)*

MASS SPECTROMETRY IMAGING

Most clinical metabolomics relies on analysis of bulk material, but in an individual patient, there are areas of normal and diseased tissue, and understanding the differences in metabolism in these areas requires both spatially sensitive resolution (imaging) and interrogation (metabolomics). While MRS can perform some of these functions, it is limited to macroscopic imaging (MRI) and relatively insensitive metabolomics approaches (NMR). In contrast, MS-based approaches, while more sensitive, by their nature rely on specimen destruction and homogenization. The premise of mass spectrometry imaging (MSI) is to overcome these limitations of MRS and mass spectrometry. MSI combines histologic evaluation of tissue with MS-based approaches to assess spatial differences in metabolites. MSI as a technique has been most highly refined in the neurosciences and can provide subcellular resolution. In general, thin slices of tissue are mounted on a slide, and metabolomics is performed at defined points across the slide, yielding spatial information on where in the tissue section metabolites are measured. One specific approach utilizes matrix-assisted laser desorption/ionization (MALDI) coupled to MS. In MALDI, tissues are coated with a special matrix and the MALDI laser scans point-by-point across a tissue slice, ionizing the metabolites at each location for analysis by a mass spectrometer. These data can then be referenced back to an image of the original tissue slice **(Fig. 502-5)**. This approach is being tested for defining tumor margins in real time during resection and thereby providing insight into boundaries between normal and abnormal tissues.

INTEGRATION WITH ADDITIONAL "-OMICS" TECHNIQUES

There is increasing interest in integrating metabolomics data with data derived from other "-omics" techniques evaluating, for example, the transcriptome or proteome (Fig. 502-1), an approach referred to as "multi-omics analysis." Integrated multi-omics may provide a more complete understanding of the biological mechanisms underpinning observed phenotypes and is being used to study heterogeneity across cell populations determined via spatial and single-cell approaches. Additionally, when applied to complex communities like the gut microbiome, these approaches can aid in the discovery of previously uncharacterized metabolic pathways that impact human health.

IMPROVING UNTARGETED METABOLOMICS

Identifying unknown signals in an untargeted metabolomics analysis remains one of the central challenges in the field. As discussed above, NMR can definitively identify unknown signals but is inferior to MS-based approaches in its sensitivity and therefore in the number of signals it can detect in a given sample. To leverage the sensitivity of MS-based detection and overcome the challenge of metabolite identification, researchers are applying computational techniques, using network-style analyses and machine learning based approaches to streamline the process. The general approach is to combine information from known biological perturbations (e.g., changes in experimental conditions or disease states), empirical mass and structural information from MS analysis, and correlations with known metabolites/pathways to place unknown metabolites within existing metabolic networks.

The growing interest in machine learning (a subset of artificial intelligence), which focuses on the training of algorithms to analyze large amounts of data, has led to it being applied to facilitate analysis and interpretation of metabolomics data. These algorithms can be used to identify unknowns in untargeted metabolomics datasets and find patterns of linked metabolites, or between metabolites and other data, that might otherwise be missed by traditional approaches.

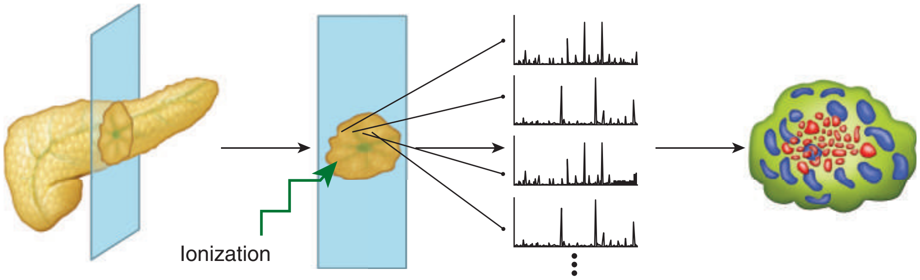

FIGURE 502-5 Mass spectrometry imaging provides spatial information around metabolites in tissues. Tissue is mounted onto a slide, and a laser or another method is used to ionize metabolites in a discreet section of the tissue for detection by mass spectrometry. The process is repeated as the laser scans across the tissue, generating an "image" based on the levels of a metabolite detected at each point in the tissue section.

SUMMARY

Metabolomics is part of a growing list of "-omics" techniques that have a growing role in both research and clinical medicine. Despite its immaturity relative to genomics, the power of metabolomics comes from being directly connected to phenotype and being very sensitive in measuring perturbations within a system or a patient. While the clinical applications of metabolomics are currently limited to specific indications, these technologies have the potential to be used more broadly in medicine. However, if these methods are to be used appropriately, clinicians need to be aware of established biological and practical confounders. Similarly, a basic knowledge of the technologies in use and their inherent limitations is critical. Nevertheless, metabolomics is becoming a part of the clinical armamentarium for diagnosis, monitoring, and treatment of disease.

FURTHER READING

Bertholdo D et al: Brain proton magnetic resonance spectroscopy: Introduction and overview. Neuroimaging Clin N Am 23:359, 2013.

Choi C et al: 2-Hydroxyglutarate detection by magnetic resonance spectroscopy in IDH-mutated patients with gliomas. Nat Med 18:624, 2012.

Coler EA et al: Metabolomics in the era of artificial intelligence. Microbiota and Host 2:e230017, 2024.

Emwas AH et al: NMR spectroscopy for metabolomics research. Metabolites 9:123, 2019.

Gencheva R et al: Clinical benefits of direct-to-definitive testing for monitoring compliance in pain management. Pain Physician 21:E583, 2018.

Kantae V: Integration of pharmacometabolomics with pharmacokinetics and pharmacodynamics: Towards personalized drug therapy. Metabolomics 13:9, 2017.

Koppel N et al: Chemical transformation of xenobiotics by the human gut microbiota. Science 356:6344, 2017.

Langley RJ et al: An integrated clinico-metabolomic model improves prediction of death in sepsis. Sci Transl Med 5:195ra95, 2013.

Mayers JR et al: Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. Nat Med 20:1193, 2014.

Rogers RS et al: Circulating N-lactoyl-amino acids and N-formylmethionine reflect mitochondrial dysfunction and predict mortality in septic shock. Metabolomics 20:36, 2024.

Townsend MK et al: Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. Clin Chem 59:1657, 2013.

Wang TJ et al: Metabolite profiles and the risk of developing diabetes. Nat Med 17:448, 2011.

503

Circulating Nucleic Acids as Liquid Biopsies and Noninvasive Disease Biomarkers

Ash A. Alizadeh, Kiran K. Khush, Yair J. Blumenfeld

While this chapter is focused on the modern use of liquid biopsies and their applications in the context of malignancies, transplantation, and noninvasive prenatal testing, it is worth considering that the history of liquid biopsies dates back many centuries. Indeed, as early as the fourth century B.C., Hippocrates is noted to have studied body fluids and humors to diagnose various maladies in his patients. By the mid-seventeenth century, the analysis of bodily fluids (especially urine) increasingly became a cornerstone of European medicine. However, arguably, the most important use of bodily fluids as liquid biopsies currently involves key clinical applications in oncology, solid organ transplantation, and prenatal diagnosis (Fig. 503-1). Below, we discuss these applications of liquid biopsies in the context of modern medical management of patients with risk of cancer and tumor progression, posttransplantation outcomes, and obstetric management.

CELL-FREE DNA IN ONCOLOGY

Major advances in the twenty-first century in cancer biology have helped transform oncology, particularly with the advent of personalized cancer therapies, which are typically selected using molecular biomarkers to identify tumor-specific vulnerabilities. However, because tumor heterogeneity remains a major obstacle to cancer monitoring and treatment, better approaches are needed to define clinically actionable targets and to facilitate the individualized therapies that embody precision medicine. While much can be learned from studying tumor biopsies directly, such as molecular profiling that informs diagnosis and monitoring of pathologic response, these are invasive clinical procedures. In addition, tumor biopsies may not yield enough material for analysis and can be risky to the patient. Therefore, an approach to analyze cancer less invasively, such as from a blood sample, serves as a potentially attractive alternative in the clinical evaluation of patients with cancer.

Tumors are well known to release analytes into body fluids, including the peripheral blood circulation, and the diagnostic use of such analytes has often been referred to as liquid biopsies (Fig. 503-2). For hematologic malignancies such as leukemias, the direct examination of circulating hematopoietic elements is a cornerstone for diagnosis, response, assessment, and disease monitoring. Indeed, molecular tools allowed the specific detection of lymphoma B cells in the blood circulation using in situ hybridization in the mid-1980s, even before the advent of polymerase chain reaction for more sensitive detection. However, even when considering hematologic neoplasms such as lymphomas, many do not circulate in large numbers in the blood. Indeed, the dominant anatomic distribution of most human cancers results in the absence of detectable circulating tumor cells as evidence for minimal residual disease (MRD) in the cellular compartment of the blood.

Aside from extravasated tumor cells circulating in the bloodstream as circulating tumor cells (CTCs), liquid biopsy methods using peripheral blood also include membrane-bound vesicles released by tumor cells called exosomes, as well as non–membrane-bound derivatives such as cell-free DNA (cfDNA) and cell-free RNA (cfRNA), which are released by apoptotic or necrotic tumor cells. Although the most common source of liquid biopsy is the peripheral blood, various bodily fluids have been used for specific liquid biopsy applications to interrogate proximally associated anatomic compartments, whether using urine, feces, pleural fluid, peritoneal fluid, bronchoalveolar lavage fluid, saliva, or cerebrospinal fluid. For example, urinary cfDNA has shown promise for noninvasive detection of genitourinary tumors including bladder carcinomas, and colorectal cancer screening can be done using multianalyte assays leveraging fecal DNA.

Among the most intensively studied tumor-derived biomarkers is circulating tumor DNA in the blood plasma, as a subset of cfDNA. cfDNA in the blood was first described in 1948 by Mandel and Metais, but it was not for another four decades that tumor-derived DNA was first observed in the plasma of cancer patients. Interestingly, cfDNA is naturally fragmented with lengths that correlate with DNA wrapped around individual nucleosomes. Since both healthy cells and cancer cells release their DNA into the circulation, a major challenge is attaining enough sensitivity to identify and quantify tumor-derived circulating DNA (circulating tumor DNA [ctDNA]) against a potentially large background of normal constitutional or germline DNA, the majority of which derives from hematopoietic sources. Methods based on next-generation sequencing offer a way to achieve this. Over the past two decades, ctDNA has been established as an important biomarker for studying tumor biology and for detection of cancers. Here, we summarize key applications of ctDNA in the context of early cancer detection,

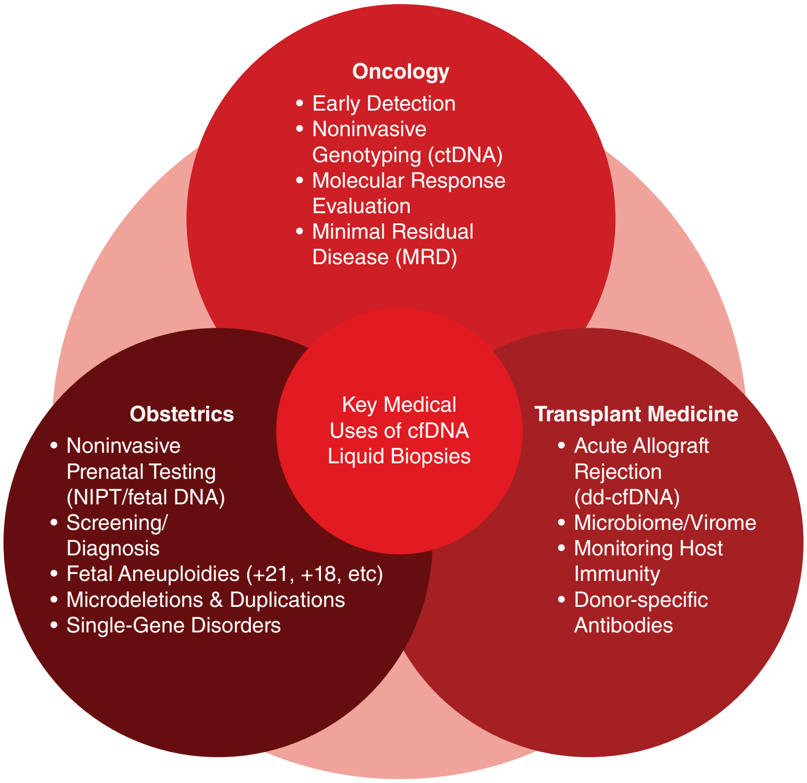
Oncology
• Early Detection
• Noninvasive Genotyping (ctDNA)
• Molecular Response Evaluation
• Minimal Residual Disease (MRD)
Key Medical Uses of cfDNA Liquid Biopsies
Obstetrics
• Noninvasive Prenatal Testing (NIPT/fetal DNA)
• Screening/ Diagnosis
• Fetal Aneuploidies (+21, +18, etc)
• Microdeletions & Duplications
• Single-Gene Disorders
Transplant Medicine
• Acute Allograft Rejection (dd-cfDNA)
• Microbiome/Virome
• Monitoring Host Immunity
• Donor-specific Antibodies

FIGURE 503-1 Key medical applications of liquid biopsies employing cell-free DNA (cfDNA), with specific focus on oncology, transplantation, and obstetrics. ctDNA, circulating tumor DNA; dd-cfDNA, donor-derived cfDNA; NIPT, noninvasive prenatal testing.

noninvasive tumor genotyping and classification, molecular response during therapy, and MRD after definitive therapy **(Fig. 503-3)**.

■ EARLY CANCER DETECTION

For several common cancer types including carcinomas of the lung, colorectal tract, and breast, early cancer detection via screening can improve outcomes for adults with established risk factors. However, such screening can pose significant risks and associated expenses that limit broad adoption. For example, while annual radiologic screening by low-dose computed tomography is recommended to screen for lung cancers in high-risk populations, implementation has been complicated by a high false discovery rate (~90%) and low compliance. Separately, distinct tumor types currently require unique corresponding screening tests. Finally, most existing cancer screening modalities have not fully integrated some of the key insights gained through molecular profiling of cancer genomes. Therefore, there is an unmet need for new methods for the early detection of cancers. Analysis of ctDNA is a promising approach that could facilitate blood-based screening. In this context, several efforts to leverage ctDNA for early detection are emerging, including approaches focused on individual cancer types, multianalyte testing combining ctDNA with other biomarkers including proteins, and broader multicancer early detection (MCED) assays leveraging mutations, tissue-specific methylation, cfDNA fragmentation profiles, and other features.

Despite much excitement about their promise, most such cancer detection efforts are relatively early in their development. As of 2021, no screening studies have prospectively applied liquid biopsies as interventions in randomized studies, with survival outcomes as clinical endpoints. Indeed, no studies performed to date have demonstrated a survival advantage of early detection using liquid biopsies in randomized trials, and this remains a key hurdle limiting widespread clinical adoption of these tests. With these limitations in mind, future prospective studies with substantially longer-term follow-up are needed to better establish the sensitivity and specificity of MCED tests for early cancer detection, when considering the common cancer types most amenable to localized interventions with curative intent. Separately, because early cancer detection may not improve cancer-specific survival, studies will also be required to determine if these tests can yield more true positives than false positives, especially because false-positive results can generate a substantial number of expensive secondary invasive procedures.

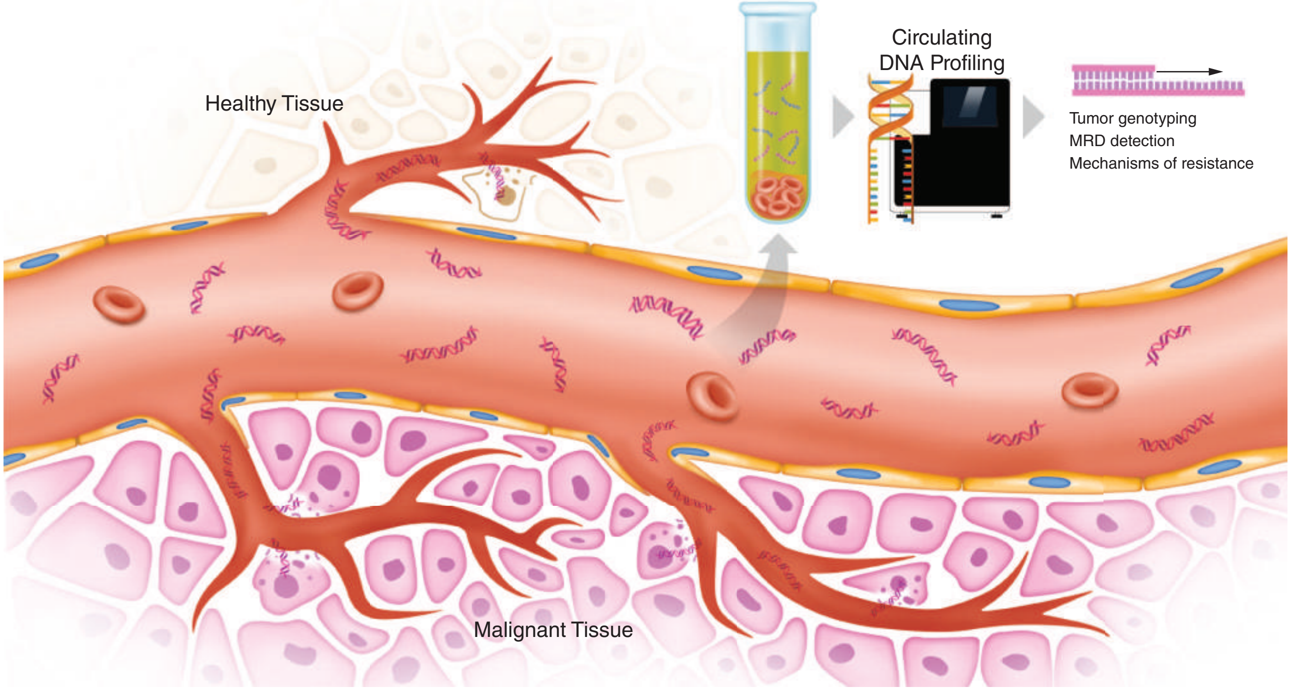
Healthy Tissue
Circulating DNA Profiling
Tumor genotyping
MRD detection
Mechanisms of resistance
Malignant Tissue

FIGURE 503-2 Cell-free DNA (cfDNA) is released both from healthy and malignant tissues. Tumors shed circulating tumor DNA (ctDNA), a minor fraction of all circulating cfDNA. Molecular profiling of cfDNA enables cancer detection and monitoring applications. MRD, minimal residual disease. *(Source: Adapted from BJ Sworder et al: Hematol Oncol, 39, 2021. https://doi.org/10.1002/hon.6_2879.)*

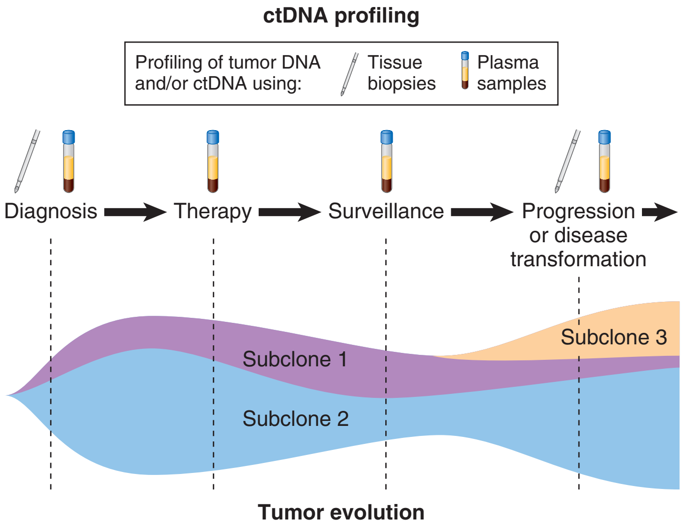

FIGURE 503-3 Framework for noninvasive identification of cancer risk groups. Schematic illustrating the application of ctDNA profiling for the identification of adverse risk at different disease milestones. A cancer patient is imagined as experiencing these disease milestones over time, depicted as an arrow progressing from left to right. During this temporal sequence, ctDNA can inform risk at diagnosis, during therapy, in surveillance of disease, and at progression or disease transformation. At diagnosis, profiling of tumor DNA obtained from either tissue biopsies (indicated by a scalpel) or plasma (depicted as blood collection tubes) allows for the identification of patients with high tumor burden and disease subtypes defined by specific genomic aberrations. Assessment of ctDNA during and after treatment facilitates the detection of both emerging resistance mutations and minimal residual disease (MRD) before progression, with potential for noninvasive prediction of relapse and treatment resistance. Tumor evolution in an anecdotal patient is illustrated, showing tumor response and clonal evolution over the course of the disease (detectable subclones at diagnosis are shown in *blue/gray*; an emergent subclone after therapy is shown in *red*). The profiling of tumor DNA and ctDNA at each milestone is shown by a double-stranded DNA molecule. *(Adapted from F Scherer et al: Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA. Sci Transl Med 8.364:364ra155-364ra155, 2016.)*

NONINVASIVE TUMOR GENOTYPING

Molecular subgroups defined by activating mutations in key oncogenic drivers are the basis for many targeted therapies in oncology across diverse cancers. For example, targetable mutations in the *EGFR* gene are the most common oncogenic driver in non-small-cell lung cancer (NSCLC), and the presence of *EGFR* mutations in NSCLC tumor biopsies is strongly correlated with a positive response to small-molecule epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs). While tissue-based determination of tumor genotype is still considered by many to represent the gold standard for diagnostic purposes, liquid biopsies have demonstrated very high positive predictive value for noninvasively determining tumor genotype in the setting of advanced malignancies. Indeed, several liquid biopsy assays have been approved by the U.S. Food and Drug Administration (FDA) for companion diagnostic use to identify mutations associated with response to targeted therapies in diverse tumor types. Currently, several comprehensive genomic profiling liquid biopsy assays targeting >50 frequently mutated genes are currently available as FDA-approved noninvasive tumor genotyping tests for various solid tumors. For example, liquid biopsy testing using the Guardant360 test (Guardant Health, Inc.) in patients with NSCLC can be performed to identify *EGFR* mutations associated with response to osimertinib or amivantamab or to identify *KRAS* mutations associated with sotorasib response. Similarly, liquid biopsy testing using the FoundationOne Liquid CDx test (Foundation Medicine, Inc.) can be performed to identify *BRCA1/2* gene mutations associated with rucaparib response in ovarian cancer, *ALK* rearrangements associated with alectinib response in NSCLC, *PIK3CA* mutations associated with alpelisib response in breast cancer, and *BRCA1*, *BRCA2*, and *ATM* mutations associated with olaparib response in metastatic castration-resistant prostate cancer.

Of note, such assays have largely been developed for noninvasive tumor genotyping in the setting of advanced malignancies and are not well suited in the context of early-stage tumors. Indeed, a recent analysis by the Sequencing Quality Control Phase 2 (SEQC2) Project Working Group led by the FDA found that ctDNA detection was less reliable below a circulating variant allele fraction of 0.5% for mutations of interest. Accordingly, the currently available assays designed for the primary purpose of noninvasive genotyping in advanced disease are generally not optimal for detection of MRD because the residual ctDNA allele fractions after definitive treatment of localized solid cancers are typically much lower than 0.5%. Furthermore, even prior to treatment of advanced cancers and especially in the context of low metastatic tumor burden, the modest sensitivity of such liquid biopsy tests currently requires that if specific lesions of interest are not initially detected in the blood, a tumor biopsy is still needed to determine if the specific mutations and alterations are present, as this could inform therapy selection.

BIOLOGIC CONSIDERATIONS

In addition to the somatic mutations found in cfDNA that originate from tumor cells, somatic mutations arising in nontumor tissues can pose as a source of biologic "background." Such mutations can confound the use of ctDNA for cancer detection and monitoring. Hematopoietic stem cells can acquire mutations through a process called age-related clonal hematopoiesis (CH), resulting in variants that can be found in both cfDNA and circulating peripheral blood leukocytes. When considering peripheral blood cells in patients who do not otherwise meet criteria for a leukemia diagnosis, identification of mutations in ~20 genes canonically associated with hematologic neoplasms (and exceeding 2% in allelic fraction) is termed *clonal hematopoiesis of indeterminate potential* (CHIP). CHIP has been identified as a risk factor for cardiovascular disease and hematologic neoplasms and is also relevant for liquid biopsies **(Fig. 503-4)**. For example, because most cfDNA derives from hematopoietic sources, CH represents a major contributor to biologic mutational background for various applications of liquid biopsies across cancers. Importantly, the prevalence of CH variants increases with patient age, larger gene panels, and more sensitive testing, with prevalence approaching 100% in adults >60 years old.

Indeed, in liquid biopsy applications for noninvasively identifying somatic alterations using cfDNA, CH represents the dominant biologic source of false-positive findings, even if this can vary as a function of the specific genes or lesions being considered. While CHIP disproportionately affects genes such as *DNMT3A*, *TET2*, *ASXL1*, and *JAK2*, which are associated with myeloid cell fitness, population sequencing studies have shown that many more genes in the genome can be affected by somatic mutations arising during CH in aging adults. For example, independent liquid biopsy studies have identified that 15–41% of cfDNA mutations in the *TP53* gene can be attributable to CH because these same lesions were found in matched blood leukocytes but not in biopsied tumor tissues.

Accordingly, direct genotyping of peripheral blood leukocytes can be very helpful to avoid such cfDNA mutations arising from CH,

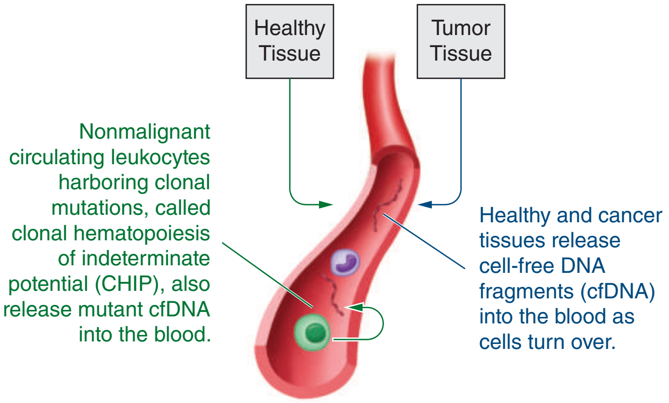
Healthy Tissue
Tumor Tissue
Nonmalignant circulating leukocytes harboring clonal mutations, called clonal hematopoiesis of indeterminate potential (CHIP), also release mutant cfDNA into the blood.
Healthy and cancer tissues release cell-free DNA fragments (cfDNA) into the blood as cells turn over.

FIGURE 503-4 Contribution of clonal hematopoiesis of indeterminate potential (CHIP) to cell-free DNA, as relevant for tumor genotyping and monitoring using circulating tumor DNA. *(Adapted from J Boegeholz et al: Hematol Oncol 39, 2021. https://doi.org/10.1002/hon.23_2879.)*

which can potentially be false-positive results masquerading as ctDNA. However, studies suggest that ~10% of mutations found in cfDNA of otherwise healthy adults may not be found in matched leukocytes. This suggests other noncirculating sources of such cfDNA mutations, including CH arising in noncirculating hematopoietic precursors in tissues such as the bone marrow or from clonal proliferations of nonmalignant, nonhematopoietic cell types. Indeed, recurrent somatic mutations in genes such as *KRAS*, *MED12*, *BRAF*, and others can be present in diverse nonmalignant cell types. These mutations (and others) can arise in a range of epithelial, endothelial, and stromal constituents of vascular malformations, endometrial leiomyomas, melanocytic nevi, and other nonmalignant proliferations. As detailed below, incorporating sequencing of tumor tissue to initially identify somatic mutations that are later monitored in plasma samples can help to guard against these nonmalignant sources of biologic background.

PRETREATMENT TUMOR BURDEN

In patients with an established cancer diagnosis, liquid biopsies can be useful as a noninvasive means of tumor genotyping using plasma, in a manner that informs therapy selection in such scenarios as described above. Aside from this use, quantitative assessments of tumor burden in the plasma prior to therapy can also provide valuable information via liquid biopsies. For example, pretreatment ctDNA levels have been shown to have significant associations with established measures of tumor burden and disease risk including stage, metabolic tumor volume, and serum protein markers such as lactate dehydrogenase in lymphomas, CA19-9 levels in pancreatic cancers, and carcinoembryonic antigen levels in colorectal cancers, among others. Importantly, in many of these scenarios and in other cancers such as NSCLC where noninvasive tumor biomarkers are not available, pretreatment ctDNA levels have been shown to have strong prognostic value for treatment failure and disease progression, providing an independently prognostic measure of risk. Indeed, such pretreatment ctDNA levels can be used to noninvasively measure tumor burden poorly captured by other indices that can lead to biases in clinical trials, including the diagnosis to treatment interval in lymphomas. Accordingly, pretreatment ctDNA levels may be used to prevent selection bias in prospective clinical trials. However, while such independent prognostic value of liquid biopsies has been validated for several tumors (e.g., using CTC levels in breast cancers), randomized trials demonstrating the clinical utility of these measurements for predicting therapeutic benefit from specific treatments have not yet been performed.

MONITORING RESPONSE TO TREATMENT

Beyond the application of liquid biopsies for early cancer detection and noninvasive tumor genotyping, their use for monitoring therapeutic responses in a quantitative and qualitative fashion deserves discussion. In many cancer types, functional imaging has utility for monitoring systemic response to treatment, when considering relative changes in volumetric and/or metabolic tumor burden to assess complete versus partial responses or stable disease. Similarly, in hematologic malignancies such as chronic myelogenous leukemia, the magnitude of response to systemic therapy with TKIs can be monitored at defined milestones while on continuous treatment, using defined thresholds for reductions in *BCR-ABL1* transcript levels in the blood. Liquid biopsies can be similarly useful for measuring quantitative changes in ctDNA while on therapy. Importantly, as with the use of functional imaging, it is critical to consider several key factors when using quantitative changes in ctDNA levels to monitor response. These include the specific tumor histology and treatment setting (frontline vs relapsed disease), treatment type, the timing of interim ctDNA response assessment, the type of ctDNA assay being used and associated sensitivity and specificity characteristics, and the thresholds for change in ctDNA levels observed as informing early molecular response. For example, liquid biopsy applications for measuring interim responses have been shown to strongly predict survival outcomes in diffuse large B-cell lymphoma and Hodgkin's lymphoma in the frontline setting. Here, 100-fold reductions (2-log) in ctDNA levels after one cycle of induction chemotherapy have been shown to reliably define an early molecular response threshold, and 2.5-log reductions in ctDNA levels after two cycles can be used to define a major molecular response, which are both strongly associated with event-free and overall survival.

MINIMAL RESIDUAL DISEASE

As in the case of several hematologic malignancies, an emerging body of evidence now demonstrates that liquid biopsies including ctDNA can detect MRD following treatment of diverse solid tumor types as a predictor for relapse risk. Indeed, detection of MRD using ctDNA-based techniques has shown strikingly high positive predictive value for predicting relapse risk in many cancer types, including carcinomas of the lung, colon, rectum, bladder, and breast, among others. For example, when considering adjuvant immunotherapy following resection of localized bladder cancers, a retrospective analysis of a randomized clinical trial strongly suggests that clinical utility for the checkpoint inhibitor atezolizumab is likely limited only to patients with ctDNA MRD detectable in the blood plasma after surgery using bespoke assays. Separately, the Centers for Medicare and Medicaid Services recently finalized the first local coverage determination to provide coverage for ctDNA MRD testing for monitoring colorectal cancers after surgery using the Signatera MRD test from Natera, and a broader draft local coverage determination is currently under consideration to enable coverage of ctDNA MRD across tumor types.

Based on these and other similar results, pivotal clinical trials are now underway that select patients for adjuvant therapy integrating liquid biopsies to detect MRD and to measure response to such adjuvant and consolidative treatments. Of note, despite the high positive predictive value of ctDNA MRD for predicting relapse, the clinical sensitivity of several current assays is generally modest, with a substantial fraction of relapses occurring in patients falsely classified as MRD negative after initial maneuvers. As in the case of early detection and noninvasive genotyping described above, for applications of liquid biopsies to MRD, it is critically important to consider the specific therapies administered, the biology of ctDNA release, the timing of MRD measurements, the liquid biopsy assay performance characteristics, and the sources of both technical and biologic background. For instance, the optimal use of liquid biopsies for their negative predictive value (e.g., by withholding unnecessary adjuvant therapy in those without evidence of MRD) is likely to require substantial additional improvements in the analytical and clinical sensitivity of MRD detection for solid tumors. Nevertheless, although the broad clinical utility of ctDNA MRD for treatment personalization has yet to be fully established, liquid biopsies hold substantial promise for guiding adjuvant and consolidative therapies.

TECHNICAL CONSIDERATIONS

Despite generally representing a small fraction of nucleic acids that circulate in the blood plasma, tumor-derived ctDNA molecules can be identified through a range of techniques related to amplification and detection. These methods broadly include assays that target tumor-specific mutations, structural variants, somatic copy number alterations, and epigenetic features and generally involve using polymerase chain reaction (PCR) and high-throughput sequencing. Recently, substantial improvements in the analytical sensitivity of liquid biopsies have been achieved through a combination of refinements in these molecular techniques and the associated computational methods for analyzing the corresponding data.

Currently, liquid biopsy applications for early cancer detection, noninvasive tumor genotyping, response monitoring, and MRD rely mainly either on sequencing-based methods or on amplicon-based techniques that do not require sequencing, such as digital droplet PCR (ddPCR) or allele-specific PCR (AS-PCR). Key factors that distinguish these methods include the cost and turnaround time for the assays (which are generally in favor of amplicon-based methods), as well as the scope of genomic aberrations simultaneously being evaluated and the breadth and depth of the associated molecular profile (which are generally in favor of sequencing-based methods). For MRD applications, the integration of multiple tumor-specific somatic alterations into a single assay can allow multiplexed sequencing-based strategies to achieve analytical sensitivities in the parts-per-million range, especially in the context of bespoke assays leveraging tumor genotypes.

Among various sequencing-based techniques, liquid biopsy applications have variably profiled the whole human genome or targeted subregions of the genome, depending on the specific applications in oncology. More specifically, selective targeting of portions of the human genome can be achieved either through enrichment by hybridization affinity capture or using locus-specific amplicons, whether focused on the whole coding exome or smaller, more focused genomic regions of interest from tens to hundreds of genes.

The biologic and technical sources of background can limit the sensitivity and specificity of liquid biopsies. Apart from the various strategies described to reduce these sources of error, many liquid biopsy studies monitoring therapeutic response and ctDNA MRD have leveraged tumor genotype–informed analyses to improve performance. Unlike noninvasive genotyping methods that rely entirely on blood plasma, this tumor genotype–informed approach includes the profiling of tumor tissue to identify mutations that are then tracked in posttreatment blood plasma. In reducing the number of mutations under consideration, this approach reduces the risk of false positives due to technical and biologic background sources of error. Separately, the tumor genotype–informed approach can be less demanding for blood sample volumes. However, due to the very low circulating levels of tumor-derived DNA at posttreatment milestones, tracking multiple tumor genotype–informed mutations and minimizing unwanted effects of biologic and technical errors are critical for optimally capturing disease risk in diverse cancers using liquid biopsies to detect ctDNA MRD.

CELL-FREE DNA IN TRANSPLANTATION

cfDNA testing offers very powerful tools for clinical monitoring of organ transplant recipients. After transplantation, cfDNA analyses have been used to assess for development of acute allograft rejection, to study microbial diversity and infection, and to quantify host immunity. This section describes current and potential future clinical applications of cfDNA testing in the transplant arena.

NONINVASIVE DETECTION OF ACUTE ALLOGRAFT REJECTION

In the setting of transplantation, cfDNA is derived from both the recipient tissues and the donated organ or cells. A transplant procedure is essentially a "genome transplant," and methods have been developed to detect and quantify levels of donor-derived cfDNA (dd-cfDNA) after transplant, with elevated levels reflecting graft injury due to acute rejection and other forms of graft damage.

In 1998, Lo and colleagues first reported detection of DNA from the organ donor in the plasma of transplant recipients. By performing PCR amplification using Y chromosome–specific primers, they were able to identify dd-cfDNA in the blood of female kidney and liver transplant recipients. This early work provided proof-of-concept of this unique approach but was limited to female recipients of male donor organs (<25% of transplant procedures).

Subsequently, a universal, sex-independent strategy was developed using whole genome shotgun sequencing to measure single nucleotide polymorphism (SNP) differences between individuals to quantify the donor signal. This approach is applicable to any organ donor and recipient combination, regardless of sex, by first genotyping the donor and recipient to identify sequence differences that can then be used to identify donor cfDNA in the recipient's blood after transplantation. Prospective studies of this approach have demonstrated that dd-cfDNA is present at very high levels during the first few days after the transplant procedure, reflecting cell death within the allograft due to ischemia and reperfusion injury. Within 1–2 weeks after transplantation, however, dd-cfDNA levels fall to a low baseline level and remain constant in the absence of acute rejection.

In the setting of acute rejection, dd-cfDNA levels increase significantly in the transplant recipient's circulation and correlate with severity of rejection. Initial studies in heart transplantation showed that, at a threshold value of 0.25%, dd-cfDNA had an area under the receiver operating characteristic curve (AUC) of 0.60 for distinguishing mild, 0.83 for distinguishing moderate-to-severe, and 0.95 for distinguishing severe rejection events, each compared to the absence of rejection. Notably, this assay can be used for surveillance of both acute cellular and antibody-mediated rejection, as both processes result in graft damage **(Fig. 503-5)**. A subsequent multicenter study confirmed the utility of dd-cfDNA monitoring for acute rejection, showing that, at a threshold of 0.25%, dd-cfDNA had a 99% negative predictive value for acute rejection and would have safely eliminated 81% of routine surveillance biopsy procedures.

Notably, increasing dd-cfDNA levels were detected several weeks to months prior to the rejection event, suggesting that dd-cfDNA is a highly sensitive marker of graft injury and can enable earlier rejection diagnosis. Early detection of graft injury may prompt augmentation of immunosuppressive therapy to halt graft damage in its early stages and to prevent a subsequent rejection event.

Finally, the investigators demonstrated higher dd-cfDNA levels associated with antibody-mediated rejection compared to cellular (T-cell–mediated) rejection, as well as different cfDNA fragment types and lengths during these two rejection processes, suggesting that dd-cfDNA patterns could potentially help differentiate these two rejection subtypes.

Parallel studies in lung, kidney, and liver transplantation have also confirmed the utility of dd-cfDNA as a noninvasive acute rejection biomarker.

A multicenter, prospective study in kidney transplantation (Diagnosing Acute Rejection in Kidney Transplant Recipients [DART]) enrolled 384 patients, on whom 107 biopsies of the transplanted kidney were performed. The dd-cfDNA level discriminated between biopsies showing any acute rejection and controls (no histologic rejection), with an AUC of 0.74, sensitivity of 59%, and specificity of 85% for acute rejection at a dd-cfDNA cutoff of ≥1.0%. A subsequent observational cohort study of 1092 adult kidney transplant recipients showed that elevated dd-cfDNA (≥0.5%) strongly correlated with clinical and subclinical allograft rejection and long-term graft function and graft survival. The study also showed that elevated dd-cfDNA levels were associated with a nearly threefold increase in risk of developing de novo donor-specific antibodies (e.g., antibodies directed against the allograft).

In kidney transplantation, dd-cfDNA could be elevated due to pathologies other than rejection, such as pyelonephritis, and is more likely a marker of severe graft injury rather than rejection alone. Thus, follow-up testing to determine the cause and nature of graft injury or use of dd-cfDNA surveillance primarily in patients at high risk of allograft rejection would be reasonable.

Studies in liver transplantation have consistently demonstrated elevated dd-cfDNA levels during episodes of biopsy-proven acute rejection. As in kidney transplantation, levels appear to rise before development of clinical manifestations of acute rejection, with data demonstrating elevated levels 4–6 days before a rise in transaminase enzyme levels and 8–15 days before biopsy confirmation of rejection. It is important to note that steady-state dd-cfDNA levels vary between transplanted organs, likely relative to the organ's mass, cellularity, or vascularization, with levels of 5–10% seen in stable liver transplant recipients and levels <0.1% seen in heart transplant patients.

Similar patterns of dd-cfDNA have also been seen after lung transplantation, with elevated levels detected weeks prior to and at the time of transbronchial biopsy demonstrating acute rejection. As in heart and kidney transplantation, dd-cfDNA levels correlate with severity of the rejection event, with greater discriminatory power seen for severe rejection and higher levels seen in antibody-mediated rejection compared to cellular (T-cell–mediated) rejection.

Donor-derived cfDNA levels also appear to have prognostic value, even in the absence of acute allograft injury. In a lung transplant cohort, patients with the highest average dd-cfDNA levels within the first 3 months after transplant, in the absence of rejection, were at high risk for subsequent chronic allograft dysfunction and allograft failure. Similarly, in kidney transplantation, donor-derived cfDNA levels at hospital discharge were shown to predict graft dysfunction at 1 year

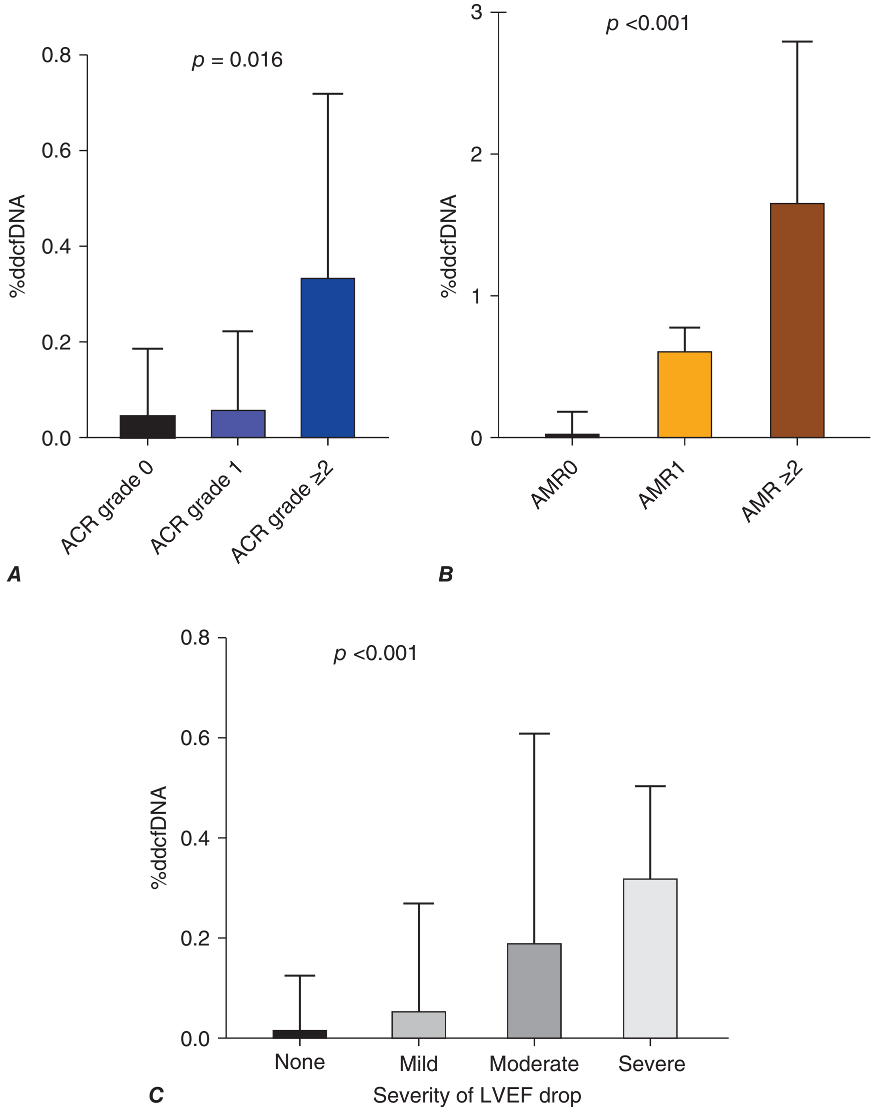

FIGURE 503-5 Correlation of percentage of donor-derived cell-free DNA (%dd-cfDNA) measures with heart transplant biopsy-graded rejection and allograft dysfunction by echocardiography. ***A.*** %dd-cfDNA in relation to severity of acute cellular rejection (ACR) by histopathologic interpretation of endomyocardial biopsy. Grade 0 rejection includes both ACR grade 0 and antibody-mediated rejection (AMR) grade 0. *P* value obtained by generalized estimating equation approach comparing all categories. ***B.*** %dd-cfDNA in relation to severity of AMR by histopathologic interpretation of the endomyocardial biopsy. Grade 0 rejection includes both ACR grades 0 or 1 and AMR grade 0. *P* value obtained by generalized estimating equation approach comparing all categories. ***C.*** %dd-cfDNA in relation to severity of allograft dysfunction measured by echocardiography. Allograft dysfunction was defined as a reduction of left ventricular ejection fraction (LVEF) by ≥5% and was further stratified by severity based on the magnitude of LVEF decline as no (<5%), mild (5% to <10%), moderate (≥10% to <15%), or severe (≥15%) allograft dysfunction. *P* value obtained by generalized estimating equation approach comparing all categories. *(Reproduced with permission from S Agbor-Enoh et al: Cell-free DNA to detect heart allograft acute rejection. Circulation 143:1184, 2021.)*

after transplant. As such, average dd-cfDNA levels over time may reflect cumulative graft injury and related long-term adverse outcomes.

Over the past few years, targeted next-generation sequencing assays (AlloSure [CareDx, Inc.] and Prospera [Natera]) have been developed to quantify dd-cfDNA levels after transplantation. These assays are composed of panels of highly polymorphic SNPs that, when sequenced and quantified, enable differentiation of donor and recipient cfDNA molecules using bioinformatics approaches. Such assays do not require donor or recipient genotyping, are rapid and relatively cost effective (particularly when compared to whole genome sequencing), and are currently available for clinical use.

In summary, dd-cfDNA can be reliably detected in the plasma of transplant recipients. Levels of dd-cfDNA are high immediately after transplantation but rapidly fall to a baseline level within several weeks, after ischemia-reperfusion injury has subsided. These dd-cfDNA levels then remain at stable low levels in the absence of graft injury, with baseline levels varying by organ type, likely related to the mass of the transplanted organ. Most studies performed to date have demonstrated rising levels with development of acute rejection and other forms of allograft injury, with highest discriminatory power for antibody-mediated rejection and severe grades of acute cellular rejection, and have shown that levels rapidly fall after successful rejection treatment.

At this time, dd-cfDNA testing cannot completely replace biopsies of the transplanted organ, which provide immunohistochemical and molecular diagnostic information that helps guide clinical treatment. Rather, this noninvasive assay reduces the need for surveillance biopsies and enables clinicians to selectively perform biopsies on patients for whom they are truly indicated, with attendant improvements in safety, cost, and patient satisfaction. The ultimate goal is to use dd-cfDNA to "personalize" immunosuppressive management by enabling weaning of therapy in stable patients with low dd-cfDNA levels and augmentation of treatment in patients with rising levels suggestive of early graft injury.

MONITORING MICROBIAL DIVERSITY AND INFECTION AFTER TRANSPLANTATION

Immunosuppressive therapies reduce the risk of allograft rejection but increase the transplant recipient's susceptibility to infectious complications. The balance between level of immune system suppression and the competing risks of rejection and infection is delicate, and the desirable "therapeutic window" for patient treatment is narrow. Furthermore, diagnosis of infectious complications after transplantation is challenging, as patients are susceptible to a wide array of pathogens and common diagnostic tests rely on an a priori hypothesis for the source of infection and performance of specific antigen-based or PCR-based molecular tests. Hypothesis-free sequencing of the microbial component of transplant recipients' cfDNA offers a window into the state of the immune response and potential diagnosis of infectious complications.

In 2013, De Vlaminck et al performed shotgun sequencing of heart and lung transplant recipient cfDNA followed by alignment to reference human and microbial databases. This work showed that 2% of cfDNA sequences were microbial in origin, including viral, bacterial, and fungal genomes. They then studied the microbiome composition in plasma at different levels of taxonomic classification and showed that viruses (73%) are more abundant than bacteria (25%) and fungi (2%). Subsequent testing showed quantitative agreement between viral counts as measured by sequencing and quantitative PCR assays for common posttransplant viral pathogens such as herpesviruses.

Available data on clinical drug treatment and dosage, including calcineurin inhibitor use and antiviral prophylaxis, were used to analyze drug-microbiome interactions. The investigators found that the structure of the virome is exquisitely sensitive to drug use and dosage—high doses or levels of immunosuppressive drugs give rise to a virome dominated by *Anelloviridae* or torque teno viruses, which are ubiquitous viruses that have no known pathogenic role in humans but have been shown to replicate in the setting of immune system suppression. Similarly, the relative amount of herpesviruses declines substantially after introduction of the prophylactic antiviral agent valganciclovir **(Fig. 503-6)**. Using this approach, the investigators

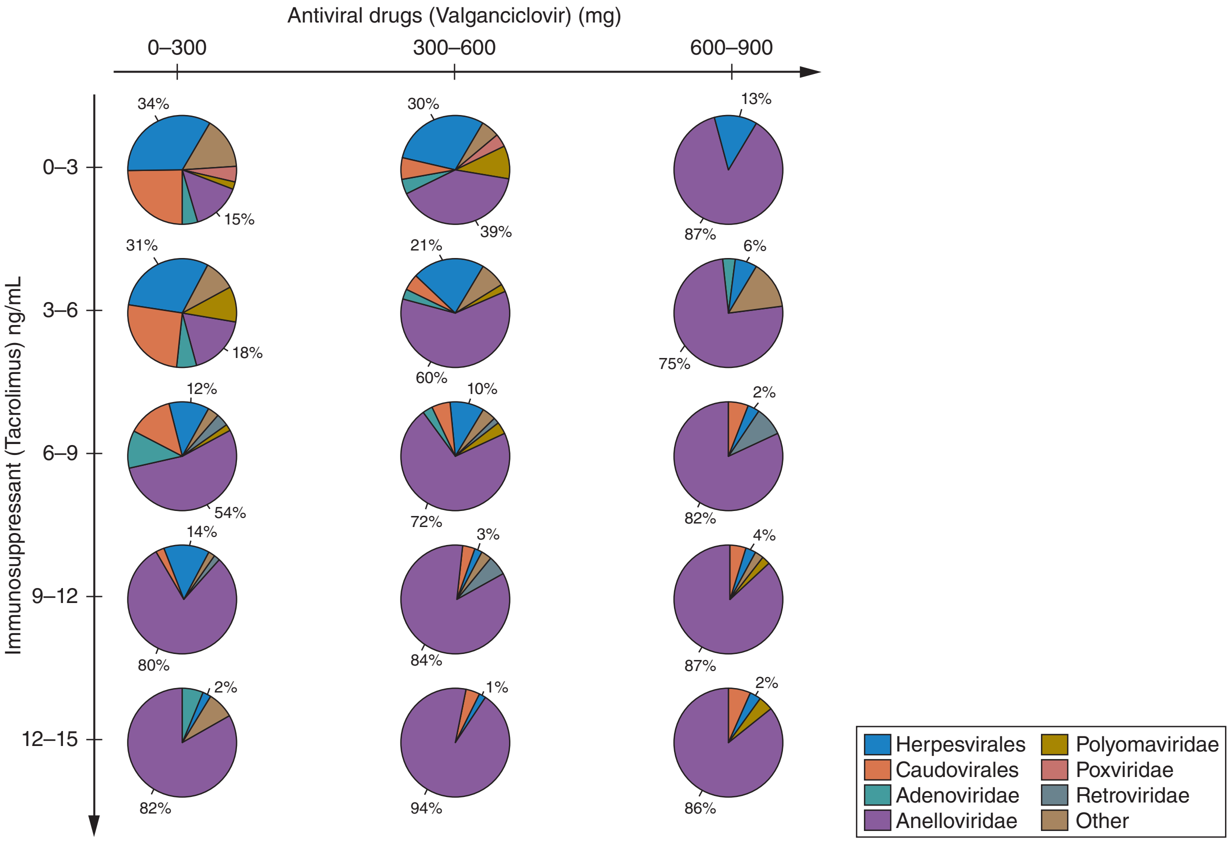

FIGURE 503-6 Relative viral genomic abundance as a function of drug dose in a cohort of heart and lung transplant recipients. Mean virome composition for patients treated with the immunosuppressant tacrolimus (47 patients, 380 samples) as a function of antiviral drug dose (valganciclovir) and concentration of tacrolimus measured in blood. To account for the delayed effect of the virome composition on drug dose, the data on drug doses were window average filtered (window size, 45 days). Herpesvirales and caudovirales dominate the virome when patients receive low doses of immunosuppressants and antiviral drugs. Conversely, anelloviridae dominate the virome when patients receive high doses of these drugs. *(Reproduced with permission from I De Vlaminck et al: Temporal response of the human virome to immunosuppression and antiviral therapy. Cell 155:1178, 2013.)*

demonstrated dramatic changes in the viral composition of the microbiome in response to introduction and weaning of immunosuppressive and antiviral therapies after transplantation.

Using a similar approach, De Vlaminck and colleagues analyzed nonhuman cfDNA as a hypothesis-free approach to test for infectious complications after lung transplantation. They showed that levels of cytomegalovirus-derived sequences in cfDNA correlated well with clinical testing results (AUC 0.91). They also identified adenovirus, polyomavirus, herpesvirus, and microsporidia sequences in patients with clinical infections that had eluded diagnosis, suggesting that cfDNA sequencing can be a powerful approach for nonbiased diagnosis of opportunistic infections after transplantation. However, several obstacles to clinical implementation must first be overcome, including the establishment of pathogen-specific thresholds to discriminate between colonization, infection, and disease.

MONITORING HOST IMMUNITY

As mentioned previously, De Vlamink and colleagues demonstrated that the *Anelloviridae* family of viruses predominated in the virome of lung transplant recipients, as analyzed by plasma cfDNA sequencing, and accounted for 68% of the total viral population. The relative abundance of anellovirus cfDNA is exquisitely sensitive to immunosuppression, ranging from <5% prior to transplant (e.g., in the nonimmunosuppressed state) to 84% during months 4.5–6 after transplantation, when immunosuppression is most profound. Starting at 6 months after transplantation, the relative amount of anellovirus cfDNA declines, corresponding to a reduction in intensity of immunosuppression **(Fig. 503-7)**.

The observation that anellovirus cfDNA levels rise and fall with degree of immune system suppression gave rise to the hypothesis that the abundance of this ubiquitous family of viruses could be used as a marker of the overall strength of the immune response. Indeed, a global marker of the net state of immunosuppression has eluded transplant physicians to date, who currently quantify doses and levels of individual immunosuppressive drugs but cannot reliably assess for their effect or potency in combination. In support of this hypothesis was the subsequent observation that anellovirus load was lower (reflecting less immunosuppression) in lung transplant recipients who developed acute rejection and was relatively high (indicating more profound immunosuppression) in patients who had a rejection-free posttransplant course, suggesting that the anellovirus cfDNA load could be a surrogate marker of immunocompetence.

Subsequent studies have confirmed these observations after different solid organ transplant procedures **(Fig. 503-8)**. A prospective lung transplant study showed a hazard ratio (HR) of 5.05 for infectious complications and an HR of 0.48 for acute rejection with every $\log_{10}$ increase in anellovirus load. Similar results have been shown after liver and kidney transplantation, with higher anellovirus levels associated with opportunistic infections and malignancy (sequelae of chronic immunosuppression) and lower levels associated with acute rejection and chronic allograft dysfunction. Prospective studies that aim to assess the potential of anellovirus levels to predict immune-related adverse events and to guide dosing of immunosuppressive medications in kidney transplant recipients are ongoing.

In summary, cfDNA sequencing has myriad clinical applications after solid organ transplantation. Donor-derived cfDNA assays are currently available for acute rejection monitoring and are being used for noninvasive surveillance after heart, lung, and kidney transplantation. These assays are highly sensitive, provide early detection of allograft injury, and can also be used to monitor response to rejection therapy. cfDNA sequencing can also be used to monitor changes in the microbiome after transplantation, in response to alterations in

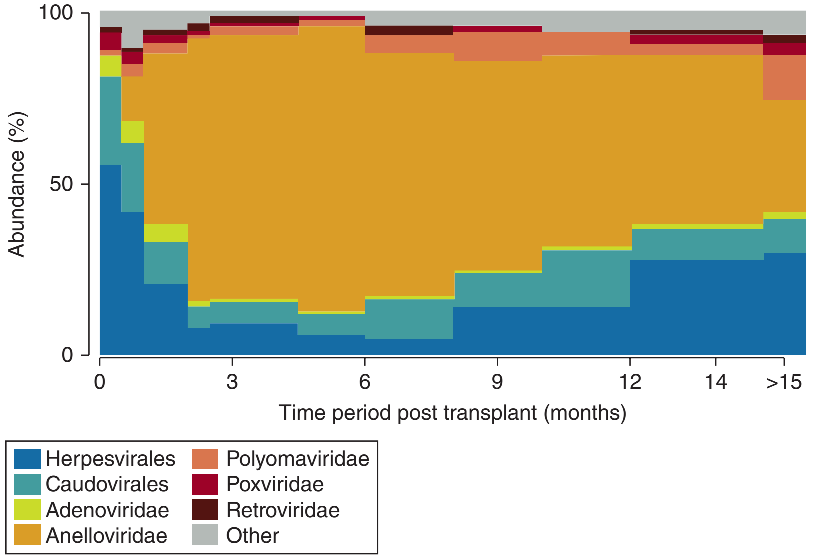

FIGURE 503-7 Viral genome abundance at the family and order level of taxonomic classification for different time periods after heart and lung transplantation. The fraction of anelloviridae expands rapidly in the first several months after transplantation. The fractions of herpesvirales, caudovirales, and adenoviridae decrease in that same time period. After 6 months, the opposite trends are observed. *(Reproduced with permission from I De Vlaminck et al: Temporal response of the human virome to immunosuppression and antiviral therapy. Cell 155:1178, 2013.)*

immunosuppression and antimicrobial therapy, and can potentially be used to diagnose opportunistic infections. Finally, quantification of anellovirus cfDNA load offers the possibility of assessing the overall status of the host immune response, which is a necessary first step for providing individualized immunosuppressive therapy to transplant recipients.

CELL-FREE DNA IN PRENATAL MEDICINE

Fetal genetic disorders are among the leading causes of stillbirth, neonatal death, and long-term developmental delay. Prenatal diagnosis of fetal genetic conditions relies on invasive sampling of fetal tissues via either amniocentesis (sampling of amniotic fluid) or chorionic villus sampling (CVS; sampling of placental tissue). Prenatal diagnostic tests are invasive, ultrasound-guided, needle-based procedures that unfortunately carry a small risk of miscarriage, preterm rupture of the membranes, and preterm birth. Historically, safe noninvasive screening for fetal genetic disorders, particularly aneuploidy (chromosomal copy number abnormalities such as trisomy 21 or Down's syndrome), was performed by combining maternal serum analyte levels (e.g., pregnancy-associated plasma protein A [PAPP-A], β-human chorionic gonadotropin [β-hCG], α-fetoprotein [AFP], inhibin) with first-trimester sonographic assessment of the fetal nuchal translucency (a measurement of the distance between the scalp and skin in the area of the fetal neck). While traditional prenatal screening is safer than invasive diagnostic tests, these modalities suffer from lower detection rates and higher false-positive rates. Advances in cell-free fetal DNA (cfDNA) technology have introduced safe noninvasive prenatal tests with much higher accuracy than traditional screening. Today, cfDNA tests have moved beyond fetal aneuploidy screening, to include fetal microdeletion and duplication syndromes and even single-gene disorders. Moreover, cell-free RNA in maternal serum has recently become a target for technologies aimed at predicting adverse pregnancy outcomes including preterm birth and preeclampsia. This section describes the use of cell-free fetal nucleic acid testing in prenatal medicine.

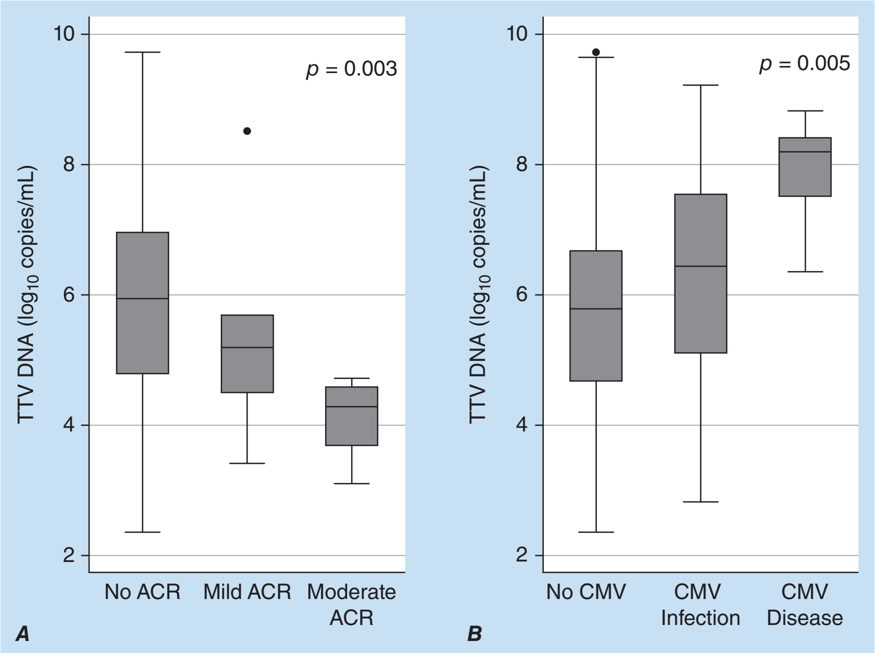

FIGURE 503-8 TTV (torque teno virus, anellovirus) load during (*A*) episodes of acute cellular rejection (ACR) and (*B*) cytomegalovirus (CMV) infection and disease after liver transplantation. *(Reproduced with permission from P Ruiz et al: Torque teno virus is associated with the state of immune suppression early after liver transplantation. Liver Transpl 25:302, 2019.)*

■ NONINVASIVE PRENATAL TESTING FOR FETAL ANEUPLOIDY

Passage of fetal cells into the maternal bloodstream was initially described almost 80 years ago. A mismatch between fetal red blood cells and maternal red blood cells was subsequently identified as the cause of isoimmunization and hydrops fetalis, a potentially lethal condition in which maternal antibodies develop against "foreign" fetal red blood cell antigens. With the development of Rho(D) immune globulin (RhoGAM) and its use to prevent RhD isoimmunization in RhD-negative mothers carrying RhD-positive pregnancies, attention then turned toward targeting fetal red blood cells for the detection of fetal aneuploidy. This is because fetal red blood cells contain a nucleus and genomic material, whereas maternal red blood cells do not. Unfortunately, initial research endeavors in the area failed to yield significant results due to several inherent limitations. First, the number of fetal red blood cells circulating among the more prevalent maternal red blood cell background is quite small. Second, the half-life of the circulating fetal red blood cells in maternal serum was found to be quite long, and circulating cells may actually originate not just from the current pregnancy but even from a prior pregnancy. Finally, the technology available at the time, namely fluorescent-activated cell sorting, was unable to accurately separate the cells to isolate the fetal cells from the maternal background.

In 1990, Lo and colleagues determined that cfDNA, and not just fetal cells, crosses the placenta and could serve as a target for fetal aneuploidy detection. This was subsequently confirmed when the presence of DNA from the Y chromosome was found in the serum of pregnant women carrying a male fetus. Over the next two decades, cfDNA circulating in maternal blood was characterized, and different genetic techniques were employed to delineate differences between fetal DNA and cell-free maternal DNA. Cell-free fetal DNA has several advantages as a diagnostic target over circulating fetal cells. First, the fraction of fetal DNA is much higher (~5–20%) than the fraction of fetal cells in maternal circulation. Second, the level of cfDNA declines within days of delivery, thereby reducing the risk of DNA from previous pregnancies leading to erroneous results.

Unfortunately, in the case of fetal aneuploidy detection, the actual sequence of fetal DNA from the chromosome of interest is identical to the maternal DNA sequence, and the challenge is to distinguish one from the other. In the early 2000s, the most promising technique for distinguishing fetal from maternal DNA was leveraging epigenetic differences between the two, particularly in the chromosomes of interest, namely chromosome 21 and chromosome 18. In 2008, two landmark studies, one by Chiu and colleagues and one by Fan and colleagues, described the utility of massive parallel sequencing technology as a counting method for noninvasive detection of fetal aneuploidy. This sequencing technology allowed for a comparison of the number of sequencing reads from chromosome 21 against an expected reference number of reads. In other words, if more sequences from chromosome 21 were found in maternal serum than what one would expect, the fetus was suspected of having Down's syndrome. While these findings were based on analysis of limited cohorts, subsequent trials in more robust and diverse cohorts confirmed these findings, and cfDNA testing became an approved clinical test in 2011.

While the initial focus of cfDNA testing was on pregnant women at high risk for fetal aneuploidy (e.g., advanced maternal age, high-risk status based on serum analyte screening, presence of fetal structural anomalies), subsequent studies confirmed the utility of cfDNA technology in both high-risk and low-risk cohorts, opening the path for widespread clinical adoption. Over time, it became clear that while the sensitivity for detecting trisomy 21 using cfDNA is high (>99%), the sensitivity for detecting trisomy 18 and 13 is slightly lower, and the positive predictive value of these tests varies by the chromosome of interest, as well as a priori risks including maternal age. Given the variable sensitivity and positive predictive values of these tests, and the fact that most societal guidelines recommend confirmatory diagnostic testing with amniocentesis of CVS in cases of positive testing, these tests are often labeled as "noninvasive prenatal screening" (NIPS) or "noninvasive prenatal testing" (NIPT), rather than "noninvasive prenatal diagnosis" (NIPD). That said, since all three terms are still used, much confusion remains among both patients and providers as to the accuracy of these tests and their inherent limitations. Today, cfDNA testing is offered by private companies, state-wide prenatal testing initiatives, and several public nationwide screening programs around the world.

NONINVASIVE PRENATAL TESTING FOR SEX CHROMOSOME ANEUPLOIDY AND RARE AUTOSOMAL TRISOMY

The accuracy (sensitivity and positive predictive value) of NIPT for the sex chromosome aneuploidies (e.g., 45,X, 47,XXY, 47,XXX, and 47,XYY) is even lower than that for the common autosomal aneuploidies. This is largely because of differences in sex chromosome mosaicism in the fetus, placenta, and even the pregnant mother. Therefore, sex chromosome testing is sometimes offered separately from autosomal aneuploidy testing, and specific pretest counseling regarding the differences in test performance is warranted.

Rare autosomal trisomy (RAT), trisomy in an autosome other than chromosome 21, 18, or 13, occurs in <1% of all pregnancies. Many RATs are usually associated with spontaneous miscarriage, but mosaic RAT has been associated with adverse pregnancy outcomes including fetal growth restriction. Since mosaic RAT can also be associated with normal outcomes, there is no universally accepted recommendation as to whether RAT should be reported by NIPT.

DEPENDENCE ON FETAL FRACTION

As clinical experience with NIPT grew, it became clear that the accuracy of the test depends on the fraction of fetal DNA in maternal serum. While fetal DNA can be found in maternal serum as early as 4–5 weeks' gestation, the fetal fraction increases with gestational age, and most NIPT programs recommend testing after 10 weeks' gestation. Although each NIPT company or national NIPT program uses a different fetal fraction cutoff to accurately report a result, a small percentage of patients will receive a "no call" or uninterpretable result due to low fetal fraction. Factors leading to "no call" also include early gestational age (usually <10 weeks), maternal obesity, underlying maternal autoimmune disease or malignancy, and vanishing twin (twins in which one fetus demises early in gestation). Several studies suggested an association between a no-call result and increased risk for aneuploidy, and the appropriate management of a no-call result remains controversial, with some recommending a repeat NIPT test, while others recommend invasive diagnostic testing given the increased aneuploidy risk.

NIPT IN MULTIPLE GESTATIONS

While initial results questioned the utility of NIPT in multiple gestations, more recent reports from large cohorts have shown that the aneuploidy detection rate in twin gestations mirrors that in singletons. Moreover, many NIPT programs report zygosity in cases of multiple gestations. This is particularly important since one would assume that NIPT results from monozygotic twins (twins originating from a single embryo) would indicate the presence of euploidy or aneuploidy in both twins, and zygosity testing is also important when the number of placentas has not been determined or remains unclear based on sonography. Monochorionic twins (all of whom are monozygotic twins) are considered higher risk than dichorionic twins (most of whom are dizygotic twins but can also be monozygotic twins) and are managed differently. Newer NIPT platforms can also distinguish the presence of aneuploidy in a single fetus in cases of dizygotic twins. The data on the accuracy of NIPT in triplet gestations or multiple gestations in which a single fetal reduction or early fetal loss occurs are more limited.

NIPT FOR FETAL MICRODELETION/DUPLICATION SYNDROMES

Advanced molecular technologies, particularly chromosomal microarray (CMA), are now the preferred diagnostic method when performing amniocentesis or CVS, particularly in the setting of a known fetal anomaly or fetal growth restriction. Unlike traditional karyotype, which relies on the visual assessment of chromosomes by a cytopathologist with a resolution of ~5–10 MB, CMA is an automated molecular technique that can identify differences as small as a single nucleotide. The main benefit of CMA over traditional karyotype is in its ability to detect copy number variants such as microdeletion/duplication syndromes (e.g., 22q11 microdeletion or DiGeorge's syndrome). Although the focus of NIPT was initially on the prenatal detection of fetal aneuploidies, the potential for NIPT to detect smaller chromosomal rearrangements, including copy number variations (CNVs), was recently evaluated and introduced into clinical care. This is important because CNVs carry a significant neonatal disease burden, and the risk of such variants may be more common than trisomy 21 among young pregnant women. Unfortunately, the detection rates for such conditions by NIPT are not as robust as they are for aneuploidy detection, and the positive predictive values are lower. Moreover, because most cfDNA originates from the placenta (syncytiotrophoblast cells), there is concern that confined placental mosaicism may result in erroneous NIPT results that target smaller genomic aberrations. High false-positive rates may lead to unnecessary parental anxiety, increased rates of invasive confirmatory testing, and pregnancy termination. Prenatal testing companies and programs offer testing for 22q11 microdeletion syndrome, Prader-Willi syndrome, Angelman syndrome, 1p36 deletion syndrome, and 5p– syndrome (cri du chat syndrome).

WHOLE GENOME NIPT AND SINGLE-GENE DISORDERS

Most NIPT platforms use next-generation sequencing to obtain DNA reads covering the entire genome but only report a fraction of the sequences analyzed. This has led to a debate as to whether whole genome NIPT should be offered and reported. Furthering the debate is the clinical incorporation of whole exome sequencing and even whole genome sequencing into testing of fetal samples obtained by amniocentesis and CVS, particularly in the setting of fetal anomalies, when CMA reports are inconclusive. As with any robust and comprehensive test, questions regarding false-positive rates and abnormalities of unclear significance are being considered. Also, whether to limit the results to known disease-causing mutations remains controversial.

While whole genome NIPT is being considered, NIPT platforms targeting single-gene disorders have been recently developed, and these are slowly becoming clinically available. This is a significant advance because expanded maternal carrier screening of single-gene disorders is becoming more prevalent, with clinically available comprehensive gene panels covering hundreds of potential disease loci. In the absence of NIPT, couples deemed at risk based on maternal and/or paternal carrier screens (e.g., an asymptomatic couple in which both parents are heterozygous for an autosomal recessive gene mutation such as cystic fibrosis) would need to undergo an invasive diagnostic procedure in order to confirm or exclude a significant fetal condition (homozygous for an abnormal cystic fibrosis gene).

NIPT USING CELL-FREE RNA

Cell-free fetal DNA is not the only nucleic acid found in maternal serum. Cell-free fetal RNA is also present, providing a unique snapshot of which genes are being transcribed at any given time in pregnancy. Cell-free fetal RNA can also be linked to tissue-specific genes, offering a unique window into developing fetal organs. Moreover, serial analysis of the pregnancy transcriptome and comprehensive analysis of cfRNA in maternal serum over time have been studied as means of early detection of common third-trimester obstetric complications including preeclampsia, fetal growth restriction, and spontaneous preterm birth. Translational studies have shown that changes in circulating cfRNA may be detected as early as the first trimester, offering an opportunity to identify high-risk pregnancies months before the onset of clinically evident disease.

NIPT DETECTING MATERNAL MALIGNANCY

Maternal malignancy complicates approximately 1 in 1000 to 1 in 1500 pregnancies. Common cancers include hematologic, breast, and gynecologic cancers, including cancers of the cervix and ovary. Because next-generation technologies sequence both maternal and fetal DNA, changes in DNA levels may not only indicate fetal disorders but also maternal disease. While most uninterpretable NIPT tests result from low fetal fraction or other fetal and/or placental conditions, a small subset may be due to undiagnosed maternal malignancy. Presence of multiple aneuploidies or autosomal monosomies has been reported by NIPT with occult maternal malignancies, and cases of maternal lymphoma, leukemia, gastrointestinal malignancy, and neuroendocrine tumors detected incidentally by NIPT have been described in the literature.

BEYOND NEXT-GENERATION SEQUENCING

While most NIPT platforms use next-generation sequencing technology, some studies have analyzed the utility of digital PCR and fetal cell analysis for NIPT. Digital PCR has been described as a useful tool for testing pregnancies at risk of X-linked and autosomal recessive single-gene disorders for both single mutations and compound heterozygous mutations. Although original attempts to isolate circulating fetal cells were abandoned >20 years ago, recent advances have led to new technologies targeting both circulating fetal nucleated red blood cells and trophoblasts. This progress, still in the research phase, suggests that it may be possible to identify and isolate single cells as early as the first trimester and analyze them by whole-gene genome amplification followed by copy number analysis using arrays or next-generation sequencing. Analysis of fetal cells would also allow for fetal gene expression profiling.

SUMMARY

Collectively, the uses described in this chapter summarize some of the myriad clinical applications of liquid biopsy methods including cfDNA profiling in oncology, transplantation medicine, and obstetrics. For each area, several existing assays have enabled transformative approaches to noninvasive diagnosis of disease, assessment of associated risk, and assessment of therapeutic responses. For example, several tumor-derived cfDNA assays are currently available for early cancer detection, noninvasive genotyping, response assessment, and MRD monitoring across diverse cancer types. Liquid biopsy assays for oncology can provide highly sensitive measures of tumor burden and associated risk and can also be used to monitor response to anticancer therapies. Prominent among applications of liquid biopsies is the detection of MRD following definitive therapy for solid cancers as a strongly prognostic biomarker with very high positive predictive value for risk of recurrence. Similarly, cell-free nucleic acids have transformed transplantation medicine. Rising levels of donor-derived cfDNA can indicate acute allograft rejection, and clinical dd-cfDNA assays are now being used for noninvasive rejection surveillance in lieu of invasive biopsy procedures. Microbial cfDNA analysis can potentially be used to screen for opportunistic infections after transplantation and for the hypothesis-free diagnosis of posttransplant infections, and such clinical assays are currently in development. Moreover, quantification of particular microbial cfDNA sequences, such as that of the *Anelloviridae* family, could one day enable transplant physicians to assess the overall status of the alloimmune response—a goal that could enable personalization of immunosuppressive therapy to maximize efficacy and to prevent the long-term toxicities associated with these medications. Finally, detection of cell-free fetal or placental DNA has improved the prenatal detection of fetal aneuploidy and other genomic abnormalities. Accordingly, the potential of these liquid biopsy techniques is rapidly expanding and places us at the cusp of a revolution in personalized management of patients across medical disciplines using innovative noninvasive techniques.

FURTHER READING

Alix-Panabières C, Pantel K: Clinical applications of circulating tumor cells and circulating tumor DNA as liquid biopsy. Cancer Discov 6:479, 2016.

Bianchi DW, Chiu RWK: Sequencing of circulating cell-free DNA during pregnancy. N Engl J Med 379:464, 2018.

Chabon JJ et al: Integrating genomic features for non-invasive early lung cancer detection. Nature 580:245, 2020.

Crowley E et al: Liquid biopsy: Monitoring cancer-genetics in the blood. Nat Rev Clin Oncol 10:472, 2013.

De Vlaminck I et al: Temporal response of the human virome to immunosuppression and antiviral therapy. Cell 155:1178, 2013.

Khush KK: Clinical utility of donor-derived cell-free DNA testing in cardiac transplantation. J Heart Lung Transplant 40:397, 2021.

Kurtz DM et al: Circulating tumor DNA measurements as early outcome predictors in diffuse large B-cell lymphoma. J Clin Oncol 36:2845, 2018.

Kurtz DM et al: Dynamic risk profiling using serial tumor biomarkers for personalized outcome prediction. Cell 178:699, 2019.

Moding EJ et al: Detecting liquid remnants of solid tumors: circulating tumor DNA minimal residual disease. Cancer Discov 11:2968, 2021.

Morain S et al: A new era in noninvasive prenatal testing. N Engl J Med 369:499, 2013.

Shaw J et al: Non-invasive prenatal testing for aneuploidy, copy-number variants and single-gene disorders. Reproduction 160:A1, 2020.

Warsof SL et al: Overview of the impact of noninvasive prenatal testing on diagnostic procedures. Prenat Diagn 35:972, 2015.

504 Protein Folding Disorders

Richard I. Morimoto, G. Scott Budinger

Many hundreds of human diseases, collectively known as protein conformational diseases or protein folding disorders, result from protein misfolding due to intrinsic and extrinsic errors amplified by aging and exposures to environmental and physiologic stress conditions. Such events challenge the functional integrity of the proteome and can lead to dysfunction, enhanced aggregation of proteins, mislocalization, and premature or inhibition of protein clearance, thus affecting cellular robustness, health, and longevity. Mismanagement of the proteome is therefore the basis of a broad class of diseases that includes orphan lysosomal storage diseases, type 2 diabetes mellitus, cystic fibrosis, some fibrotic diseases, metabolic diseases, muscle-wasting diseases, cancer, and neurodegenerative diseases exemplified by Alzheimer's disease, frontal temporal dementia, Parkinson's disease, amyotrophic lateral sclerosis (ALS), and Huntington's disease **(Fig. 504-1)**. For each of these diseases and many others described in this textbook, aging is the major contributing risk factor.

The challenge at the biochemical and molecular level is for the cell to achieve and maintain a stable and functional proteome during development that persists through young adulthood and is maintained throughout aging. For humans, this is essential for the operational health of each of the tens of trillions of cells that comprise our ~80 organs for health span and lifespan. To achieve this, our cells have evolved a remarkably efficient proteostasis network (PN) composed of ~3000 molecular chaperones and other highly conserved components essential for protein synthesis, folding, translocation, and degradation **(Fig. 504-2)** that balances input with output and ensures that every protein is functional. The PN is, therefore, essential for the robustness of all tissues and for the diverse protein-protein interactions in cell signaling, biosynthetic processes, and the structural demands and mechanical requirements for tissue shape and function. An equal, if not more important, role for the PN is to detect, prevent, and remove misfolded and aggregated proteins that accumulate in stress, aging, and disease and interfere with cellular health. Understanding how proteostasis is achieved and maintained is of fundamental biological interest and essential to prevent age-associated protein folding disorders. Consequently, there is tremendous interest in the detection and treatment of these diseases, in particular as the human population continues to live longer.

All organisms express an evolutionarily conserved set of molecular machines for the synthesis, folding, transport, and removal of unnecessary and damaged proteins. The PN is adapted for the highly specific physiologic requirements of tissues and the expression of abundant and rare proteins with wide-ranging solubilities, folding requirements, stability, and structural demands. Added to this complexity of natural clients for the PN is the additional load generated by genetic mutations carried in natural populations together with diverse environmental stressors that challenge PN capacity. Despite the central role of proteins as the workhorse of the cell, they are also highly susceptible to molecular damage, whether due to intrinsic metastability or to genetically inherited mutation or error-prone synthesis. Hence, dysfunction in the PN may clinically manifest as a gradual decline in homeostatic function in aging as occurs with genetic mutations or a loss of resilience in the face of environmental stressors. Thus, clinicians see the consequences of proteostasis failure and cellular dysfunction in both the myriad disorders that present to physicians as age-associated clinical problems and the increased morbidity and mortality associated with trauma, infection, and other acute illness requiring hospitalization in older individuals.

PROTEIN QUALITY CONTROL MECHANISMS

The PN monitors and controls the flux of protein synthesis to promote functional folding and to minimize the accumulation of off-pathway aggregation-prone intermediates by their selective disaggregation or degradation. However, unlike an automobile assembly plant for which each part is designed and engineered for a specific use and precise assembly, the PN has built-in functional redundancy with properties to tolerate tremendous chemical noise and sequence diversity generated by coding region polymorphisms and biosynthetic errors among its client components. The PN has the ability to recognize and remove kinetically unstable conformational states of proteins that accumulate in aging and that would otherwise compromise assembly and function. Proteins are highly sensitive to fluctuations in their intracellular environment caused by shifts in energetics, pH, and oxidizing and reducing conditions in addition to the myriad small molecules and metabolites that affect folding and function. Added to changes are the effects of external stress conditions caused by elevated temperatures, infection, oxidizing and reducing environments, or osmolytes that can have profound consequences on protein folding thermodynamics, kinetics, and function. These intracellular and extracellular stress conditions, if not properly responded to, are predicted to further amplify protein instability from sequence polymorphisms and biosynthetic errors that contribute to the stress of protein misfolding. The PN is organized at the cellular level into a series of highly coordinated molecular machines that direct the expression, biogenesis, and functional health of essentially all

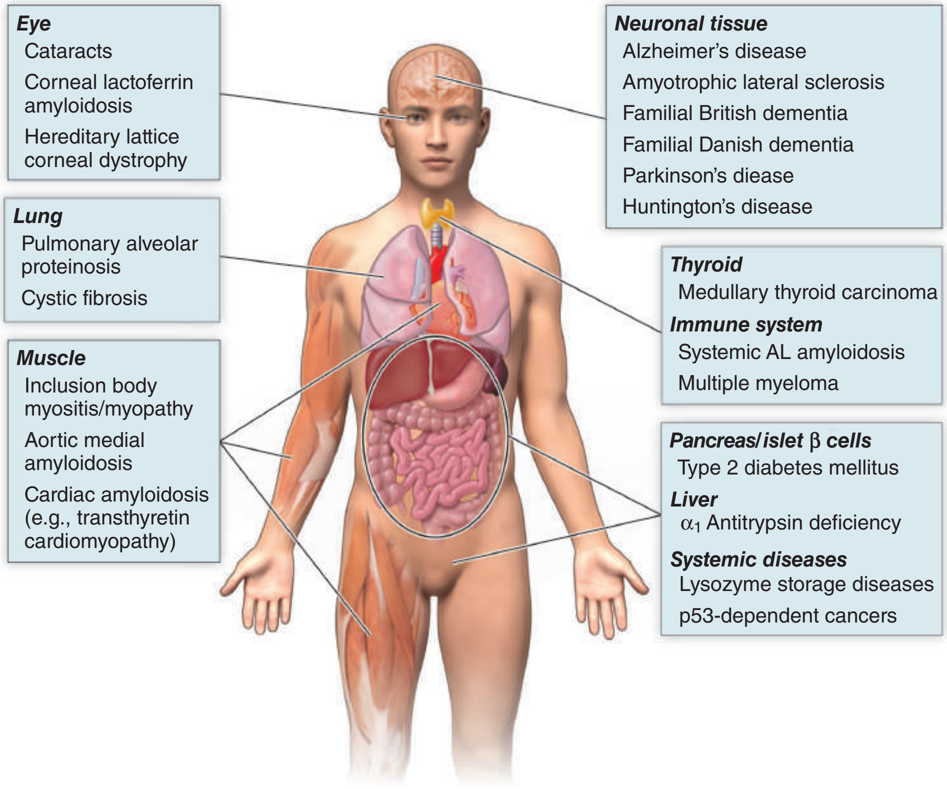
Eye
Cataracts
Corneal lactoferrin amyloidosis
Hereditary lattice corneal dystrophy
Lung
Pulmonary alveolar proteinosis
Cystic fibrosis
Muscle
Inclusion body myositis/myopathy
Aortic medial amyloidosis
Cardiac amyloidosis (e.g., transthyretin cardiomyopathy)
Neuronal tissue
Alzheimer's disease
Amyotrophic lateral sclerosis
Familial British dementia
Familial Danish dementia
Parkinson's diease
Huntington's disease
Thyroid
Medullary thyroid carcinoma
Immune system
Systemic AL amyloidosis
Multiple myeloma
Pancreas/islet β cells
Type 2 diabetes mellitus
Liver
α1 Antitrypsin deficiency
Systemic diseases
Lysozyme storage diseases
p53-dependent cancers

FIGURE 504-1 Diseases of protein folding. A representative list of tissues affected and known folding diseases.

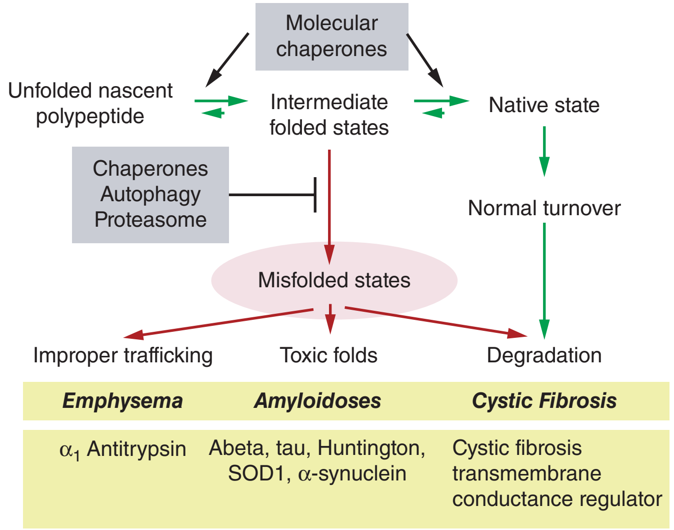
Molecular chaperones
Unfolded nascent polypeptide
Intermediate folded states
Native state
Chaperones Autophagy Proteasome
Normal turnover
Misfolded states
Improper trafficking
Toxic folds
Degradation
Emphysema
Amyloidoses
Cystic Fibrosis
α1 Antitrypsin
Abeta, tau, Huntington, SOD1, α-synuclein
Cystic fibrosis transmembrane conductance regulator

FIGURE 504-2 The proteostasis network (PN) and folding diseases. Protein biogenesis through the action of molecular chaperones ensures the transition of the nascent polypeptide to on-pathway intermediates and the folded functional native state. Such proteins then have a normal turnover mostly through the ubiquitin-proteasome system. Off-pathway species are prevented by the actions of chaperones and the recognition of nonnative misfolded states by the autophagic-lysosomal pathway and the ubiquitin-proteasome system. When misfolded species escape quality control or overwhelm the PN, they can then become improperly trafficked as occurs for α_1 antitrypsin associated with emphysema; for toxic folds as occurs for abeta, tau, Huntington, and SOD1 in amyloidogenic neurodegenerative diseases; and prematurely degraded as occurs for CFTR associated with cystic fibrosis.

proteins (Fig. 504-2). More than to regulate and orchestrate these highly synchronized events, the PN is essential for protein quality control and for the prevention of the appearance of off-pathway conformational states and condensates, with accumulation of aggregates and amyloid species. Proteome health involves the constant exchange between the intrinsic physicochemical properties of polypeptides and the biological milieu of the cellular environment in which protein sequences and function evolved.

Beginning with the synthesis of the nascent polypeptide on the ribosome, ribosome quality control (RQC) together with cytoplasmic chaperones of the HSP70, HSP90, DNAJ/HSP40, chaperonin/HSP60, and small HSPs (sHSPs) family ensure co- and posttranslational folding for the cell. Approximately 60% of the proteome resides in the cytoplasm and nucleus, for which the RQC, HSP70, HSP90, and HSP60 chaperones together with co-chaperones regulate the folded state of client proteins through cycles of ATP binding and hydrolysis. Chaperones of the HSP70 and J-domain family are particularly well studied for their ability to interact transiently with short dispersed hydrophobic regions of nascent polypeptides using the energy from nucleotide hydrolysis to regulate the release of partially folded intermediates that either reenter the chaperone cycle or are released in a folded native state. For other chaperone clients, such as transcription factors, kinases, phosphatases, and signaling molecules, their folding to the functional state is highly regulated and dependent upon interactions with the HSP90 chaperone and other regulatory co-chaperones to form stable heteromeric complexes that hold the client in a partially folded state primed for subsequent regulated release.

Consistent with the recognition that the formation of off-pathway aggregates is a kinetic component of proteostasis are the concerted activities of chaperone machines with disaggregase activity that unravel protein aggregates for refolding. These disaggregases include the AAA$^+$ protein, ClpB in bacteria, Hsp104 in yeast and plants, and the functionally analogous metazoan disaggregase composed of HSP70, HSP110, and specific J-domain proteins.

The subcellular organelles mitochondria and endoplasmic reticulum (ER) account for ~20% of the proteome. Chaperone interactions are essential for mitochondrial-targeted proteins to maintain the extended polypeptide chain in a recognition-competent state for the organellar receptors for translocation across membranes. Upon import, each translocated polypeptide is met by organellar-specific chaperones of the HSP70 and J-domain family for folding and assembly. While the mitochondrial genome encodes 13 proteins required for electron transport, the great majority of mitochondrial proteins are encoded by the nuclear genome, synthesized in the cytosol, and imported across the outer and inner mitochondrial membrane. Hence, maintenance of the mitochondrial proteome relies on the coordinated efforts of both the cytosolic and mitochondrial PN. For translocation into the lumen of the ER, the extended polypeptide interacts with a set of glycosyltransferases, calnexins, calreticulins, disulphide isomerases, and lumen localized chaperones. Proteins that misfold in the ER are recognized and retro-translocated to the cytoplasm where they are directed to the ubiquitin-proteasome system (UPS) for unfolding, ubiquitination, and degradation.

The PN is balanced by the essential catabolic processes of the UPS and the autophagy-lysosomal pathway (ALP), which recognizes proteins for degradation and recycling. The UPS is generally considered the primary pathway by which most proteins are recognized and tagged for degradation, and the ALP is highly responsive to nutrient deprivation and damage to recognize large aggregates and inclusions and engulf organelles and other subcellular compartments. In addition to their role in the regulated turnover of cellular proteins, these degradation systems are essential for protein quality control and for limiting the accumulation of misfolded and aggregated proteins during stress conditions, aging, and disease. Protein turnover by the UPS involves an enzymatic cascade of E1, E2, and E3 enzymes that utilize ubiquitin and the recognition selectivity of E3s to tag clients, followed by degradation of the polyubiquitinated substrates by the 26S proteasome in the cytoplasm. Client specificity involves the large family of ~750 ubiquitin E3 ligases. In addition to their role of marking proteins for degradation, the ubiquitination machinery has numerous additional functions in cellular processes. For example, the ubiquitin ligase listerin is associated with the ribosome to ubiquitinate nascent chains that stall during protein synthesis to prevent the accumulation of aberrant polypeptides that would subsequently aggregate. Ubiquitination of nonnative aggregated clients by the ubiquitin ligase activity of the cochaperone carboxyl terminus Hsc70 interacting protein (CHIP) is central to the triage decision of the HSP70/HSP90 complex between client folding and proteasome-mediated degradation. ER-targeted clients that are misfolded are retro-translocated to the cytoplasm where they are polyubiquitinated and degraded by cytosolic proteasomes in a process termed ER-associated degradation (ERAD). Ubiquitination also provides crosstalk between the proteasome and autophagy pathways by targeting clients for lysosomal degradation and for endosomal sorting. Chaperones co-label a protein as damaged, recruiting other proteins that place ubiquitin chains on the damaged protein for degradation by the proteasome. Alternatively, chaperones can label proteins or protein aggregates to target them to the lysosome, and in this process, damaged proteins are degraded by the lysosome, an intracellular organelle with an acidic environment enriched in proteases through autophagy.

While there is a comprehensive understanding of the process of in vivo chaperone-dependent protein folding, the details of how these decisions are made for each client, whether and for how long to be maintained in a nonnative folding state through chaperone interactions in a nucleotide-independent state, or how to assemble into a stable chaperone complex for subsequent assembly into a functional state or to interact with chaperones to directly fold to a native state remain to be fully addressed.

CELL STRESS RESPONSES: SENSORS AND REGULATORS OF PROTEIN DAMAGE

Cell stress responses are ancient genetic networks that detect, adapt, and protect all cells against toxic environmental stressors and physiologically relevant changes in their cellular environment, including changes induced during development and tissue repair after injury (Fig. 504-3). At the core of these cell stress responses are molecular switches: (1) the heat shock response (HSR) that protects proteins in the cytoplasm and nucleus regulated by HSF-1; (2) the unfolded protein response (UPR)

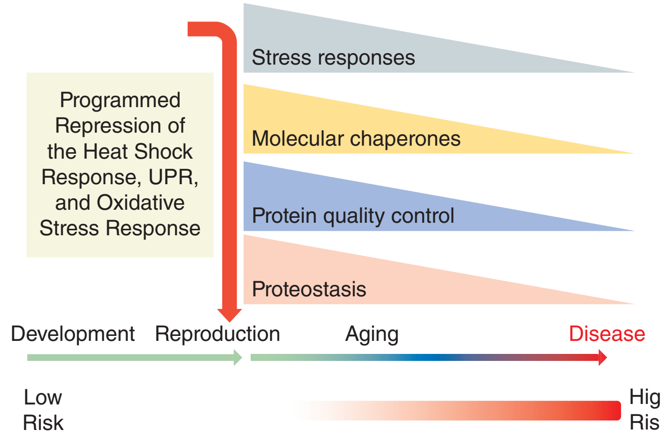

FIGURE 504-3 Aging and protein folding diseases. Aging is the major risk factor for degenerative diseases. Cell stress responses (heat shock response and the unfolded protein responses [UPR] in the endoplasmic reticulum and mitochondria) decline at reproductive maturity in studies from *Caenorhabditis elegans* and prevent adaptive and protective increased expression of molecular chaperones to prevent protein misfolding.

of the ER (UPR^{ER}) controlled by XBP1, ATF6, and ATF4; (3) the UPR of the mitochondria (UPR^{mt}) controlled by ATFS1; (4) the DAF-16/FOX-O stress response pathway associated with insulin signaling; (5) the integrated stress response (ISR) controlled by PERK, PKR, HRI, GCN2, and ATF4; and (6) the antioxidant stress response regulated by NRF-2. Collectively, these cell stress responses and their respective transcription factors (TFs) are essential for all cells and tissues regulated both autonomously and cell nonautonomously across tissues in metazoans to detect proteotoxic stress, to adapt and protect the cell against the toxic consequences of protein damage, and to regulate changes in the proteome necessary for differentiation. While each of these cell stress pathways can be activated independently, they are also induced in different combinations according to the chemical and physiologic properties of the stress signal(s) and provide cross-protective mechanisms.

The HSR is an evolutionarily conserved cellular defense mechanism that protects cells against proteotoxicity associated with misfolding, aggregation, and proteome mismanagement. HSF-1 inducibly regulates the transcription of genes encoding molecular chaperones and components of the PN. In unstressed cells, HSF-1 exists in an inert monomeric state in the cytoplasm or nucleus where it is negatively regulated by the molecular chaperones, HSP70 and HJD-1. Upon exposure to heat shock, HSF-1 undergoes a series of molecular transformations and rapidly trimerizes to acquire DNA-binding activity, undergoes extensive posttranslational modifications by phosphorylation and sumoylation, binds to heat shock elements in promoters of heat shock responsive genes, and associated with these events, forms nuclear stress bodies. Upon dissipation of the stress signal, the HSR attenuates by the active repression of HSF-1 DNA binding by acetylation and loss of HSF-1 transcriptional activity. This is accomplished by binding of HSF-1 with HSP70, HDJ-1, and HSBP1, leading to its dissociation from the trimer to the monomeric state. In addition to HSF-1 being essential for the HSR and cell and organismal stress resilience, HSF-1 is essential during early development in metazoans, functions as a maternal factor for gametogenesis, regulates oocyte maturation by activating genes that function in the meiotic cell cycle, is constitutively activated in cancer, and is necessary to maintain NAD^+ and ATP levels.

In the ER, the UPR^{ER} involves three stress response arms regulated by the transcription factors XBP1, ATF6, and ATF4 that bind to specific cis- elements for these ER-stress-responsive pathways. XBP1 is activated by IRE1, which is a transmembrane protein with kinase and endoribonuclease (RNase) activity that senses misfolding in the ER directly, leading to its autophosphorylation, oligomerization, and acquisition of RNase activity. This allows active IRE1 to cleave *XBP1* mRNA, generating a spliced transcript (*XBP1s*) that encodes XBP1 to induce the transcription of UPR target genes. ER stress also promotes the relocalization of ATF6 from the ER membrane to the Golgi apparatus, where it is cleaved by the proteases SP1 and SP2, generating a cytosolic fragment of ATF6 that translocates to the nucleus to direct transcription of a complementary set of UPR genes. Together, XBP1 and ATF6 induce the expression of genes involved in protein folding, ER-associated protein degradation, and lipid metabolism. A third ER transmembrane protein, PERK, also induced by ER stress, phosphorylates the translation initiation factor eIF2α, linking activation of the UPR^{ER} with the ISR.

In the mitochondria, the UPR^{mt} response involves ATFS1, which contains a mitochondrial targeting sequence and a nuclear localization signal. Under normal cellular conditions, ATFS1 is imported into mitochondria and degraded, but upon mitochondria stress, ATFS1 is directed only to the nucleus to regulate transcription of genes encoding mitochondrial chaperones, mitochondrial import machinery, and glycolysis. In mammals, the UPR^{mt} is regulated by ATF5, which is the orthologue of ATFS1 in *Caenorhabditis elegans*. Inhibitors of mitochondrial electron transport in mammals also activate the ISR through the release of a protease OMA1 that cleaves a cytosolic protein DELE1 to activate HRI.

The ISR is induced by one or more of four kinases: PKR, PERK, HRI, or GCN2. All four of these kinases phosphorylate eIF2a, a key protein in the ternary complex that regulates protein synthesis. The resulting global inhibition of protein synthesis paradoxically promotes translation of mRNA molecules with specific sequences in their upstream open reading frames. These include the transcription factor ATF4, which induces the expression of a gene program that maintains metabolism to preserve stress resilience. Activation of ATF4 also induces its expression and the expression of *Ddit3* (CHOP), lowering the threshold for apoptosis, and *Gadd45a*, a phosphatase that dephosphorylates eIF2a to restore translation. The ISR is linked to ER stress through PERK and to mitochondrial stress through HRI. GCN2 is activated by amino acid deprivation, and PKR is activated during viral infection. Of particular note has been the development of ISRIB, a small molecule that partially inhibits the activity of the ISR with salutary effects across diverse animal models of age-related degenerative diseases. These findings suggest that cell stress pathways that are proteome protective in youth might become pathologic in aging, making them attractive targets for therapeutic intervention. Indeed, activation of the ISR has been shown to impair stem cell differentiation, perhaps linking mitochondrial dysfunction with aging, proteostasis, and stem cell dysfunction.

In metazoans, the integration of stress survival strategies includes the antioxidant factor SKN-1/NRF2, the insulin-signaling factor DAF-16/FOXO, and the tissue identity factor PHA-4/FOXA. Oxidative and xenobiotic stresses activate OxR, which controls the expression of redox-regulatory proteins and components of protein degradative pathways mediated in mammals by NRF1/NFE2L1 and NRF2/NFE2L2, which corresponds to SKN-1 in *C. elegans*. NRF1 is an ER-resident factor that undergoes regulated proteolytic cleavage upon activation to control expression of genes encoding subunits of the proteasome and the UPS. NRF2 in the cytoplasm is negatively regulated by the redox-sensitive ubiquitin ligase KEAP-1; consequently, inactivation of KEAP-1 by oxidative and electrophilic stress leads to stabilization and nuclear translocation of NRF2, which in turn induces the expression of antioxidant proteins and detoxification enzymes.

ORGANISMAL PROTEOSTASIS IN AGING AND DISEASE

Much of our understanding of protein quality control mechanisms has come from in vitro studies with purified molecular chaperones or components of the UPS, complemented with cell extracts and cell-based assays using yeast or mammalian tissue culture cells. A common theme that emerges from these studies is that of hormesis, in which chronic low-level activation of the HSR, UPR^{ER}, and UPR^{mt} is protective against subsequent exposures to extreme and lethal cell stress conditions.

The importance of these pathways is further highlighted by studies in metazoans that indicate that cell stress responses are regulated at the organismal level by neuronal signaling. At the organismal level in

C. elegans, the HSR, UPR^{ER}, and UPR^{mt} are regulated by cell-nonautonomous control by specific sensory neurons. When neuronal signaling is impaired, the HSR reverts to cell-autonomous control. Likewise, neuronal signaling regulates the UPR^{mt} with disruption of mitochondrial function in *C. elegans* neurons activating the UPR^{mt} in nonneuronal tissues, supporting a role for a mitokine signal. Perturbation of the mitochondrial electron transfer chain (ETC) was shown to increase lifespan in both invertebrates and rodents through the activation of the UPR^{mt}. The response to mitochondrial dysfunction in *C. elegans* depends on the severity of mitochondrial impairment, with a mild reduction of ETC or reduction of mitochondrial proteostasis having hormetic effects on organismal stress resilience, proteostasis, and longevity by resetting the cytoplasmic HSR through HSF-1, independent of ATFS-1 and the UPR^{mt}. Mild perturbation of the ETC in *Drosophila* muscle also has systemic benefits on organismal health and lifespan involving the insulin signaling pathway. Communication between neurons also regulates the UPR^{ER} in peripheral tissues of *C. elegans*. During infection of *C. elegans* by pathogens, induction of the UPR^{ER} in nonneuronal tissues is mediated by sensory neurons, suggesting an organismal stress response. Cell-nonautonomous regulation of the UPR^{ER} has also been observed in mice, where overexpression of active XBP1 in pro-opiomelanocortin neurons activates the UPR^{ER} in the liver.

Several other forms of intertissue communication regulate proteostasis with beneficial effects on organismal health in model organisms. For example, muscle cell proteostasis in *C. elegans* is regulated by cholinergic signaling across the synaptic junction through modulation of HSF-1 activity. Transcellular chaperone signaling between somatic tissues, and between somatic tissues and neurons, of *C. elegans* communicate proteotoxic stress signals via the tissue code factor PHA-4/FOXA to control systemic expression of HSP90. In *Drosophila*, overexpression of small HSPs only in the flight motor muscle cells protects neurons and glial cells from elevated temperature-induced death. Enhanced expression of DAF-16/FOXO in the intestine enhances proteostasis in distant muscle cells of *C. elegans*, and likewise, overexpression of dFOXO/4E-BP in *Drosophila* muscle influences proteostasis in retina, brain, and adipose tissues to delay the age-dependent accumulation of protein aggregates.

Cell stress responses and proteostasis decline in aging with insights on the relationships between processes coming primarily from studies using *C. elegans* with support from other invertebrate and vertebrate model systems and human cells. Endogenous metastable proteins that harbor temperature-sensitive properties misfold in *C. elegans* at the permissive temperature in early aging associated with a decline in the HSR. This functional decline of proteostasis in *C. elegans* aging is regulated by cell-nonautonomous control, from the germline stem cells to the somatic tissues for the programmed repression of the organismal HSR, resulting in the loss of stress resilience and proteostasis causing age-associated protein aggregation. This HSR repressive switch is regulated by signaling from the germline stem cells to the somatic tissues, resulting in the timed placement of repressive H3K27me3 chromatin marks at the promoters of heat shock genes, causing chromatin inaccessibility for HSF-1. This age-dependent decline in the HSR can be reversed either by blocking the signal from germline stem cell signal(s) or preventing the epigenetic repressive marks. The relationship between reproduction and inducibility of the HSR observed in animals at reproductive maturity suggests that the age-associated events of cellular failure and loss of tissue robustness during aging are not random processes but rather highly regulated, perhaps to ensure that somatic tissues are programmed to decline after reproduction, consistent with the germline soma theory of aging.

Proteostasis represents one of the primary hallmarks of the biology of aging, which together with genomic instability, telomere attrition, epigenetic alterations, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication provides a mechanistic basis for the aging process. The programmed decline of proteostasis in early adulthood would support the hypothesis that failure in protein quality control would have negative consequences on the other pillars of geroscience. Whether proteostasis collapse is the first to fail or among the earliest events that fail in aging, it is consistent with very large number of human degenerative diseases in aging associated with protein misfolding.

PROPERTIES OF PROTEIN FOLDING DISEASES

The complexity that arises with protein folding diseases is that all tissues are at risk and all proteins are at risk for misfolding and loss-of-function or gain-of-function proteotoxicity. Added to this is the effect of aging and that each protein folding disease exhibits a highly variable age of onset for pathology. There is additional complexity in classification regarding whether to organize folding diseases by tissues (i.e., muscle proteinopathies or neurodegenerative diseases), according to the specific protein that misfolds such as α_1-antitrypsin deficiency, or by the biophysical nature of the misfolded or aggregated species in amyloidoses.

Disorders in which a specific mutation leads to protein misfolding or the formation of a specific insoluble protein aggregate likely represent only the tip of the iceberg of protein folding disorders. Mutations in aggregation-prone proteins coupled with changes in the cellular environment and effects on the capacity of the PN will promote misfolding and aggregation in affected tissues. Chronic stress may cause the aberrant cell stress responses and protein quality control pathways, causing further collateral damage and aggregation of other at-risk proteins. Such a mechanism may only manifest clinically after a seemingly random systemic stress like pneumonia, large bone fracture, or ischemic vascular event, possibly explaining the rapid (1–2 years) accumulation of age-related morbidities in the year following a major biologic stress, marked by a need for hospitalization. As such, the age-related decline in the function of any of the components of the PN could underlie the compounding multiple morbidity that limits health span and lifespan in many elderly individuals. Within this framework, it is useful to discuss some of the better understood mechanisms of proteostasis dysfunction that have been causally linked to diseases in humans.

■ DISORDERS THAT ENHANCE CLIENT MISFOLDING AND CAUSE PREMATURE DEGRADATION (CYSTIC FIBROSIS)

Cystic fibrosis (CF) is a recessive disorder caused by mutations in both alleles of the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene that encodes a multidomain membrane-spanning chloride ion channel protein. Thousands of mutations in *CFTR* have been identified that affect CFTR biosynthesis, folding, trafficking, and function, leading to chronic obstructive lung disease, intestinal obstruction, liver dysfunction, exocrine and endocrine pancreatic dysfunction, and male infertility. CF is a folding disease due to its recognition by the PN as misfolded protein. The most prominent mutation is deletion of phenylalanine 508 (*F508del*), present in ~90% of CF patients. Mutant ΔF508 retains partial channel function, but because it is recognized as misfolded in the ER and the cytoplasm, it is marked with ubiquitin for degradation by the UPS. Combinations of small molecules that protect the misfolded CFTR protein from degradation and enhance its function have led to substantial improved outcomes in many patients with CF **(Chap. 302)**.

■ DISORDERS THAT INDUCE TOXIC AGGREGATES AND LOSS OF FUNCTION IN MULTIPLE TISSUES (α_1-ANTITRYPSIN DEFICIENCY)

α_1-Antitrypsin deficiency (AATD) is a co-dominant inherited disease with an increased risk of chronic obstructive pulmonary disease, liver disease, and inflammation of the blood vessels. Pulmonary problems are more frequent in adults, whereas liver and skin problems may occur in adults and children. α_1-Antitrypsin is encoded by the *SERPINA1* gene and secreted into the circulation by the liver and is responsible for inactivating endogenous proteases, particularly those secreted by neutrophils and other inflammatory cells in the lung. Patients with AATD harbor mutations in *SERPINA1* that cause misfolding in the ER. The two major phenotypes resulting from this abnormality highlight the diverse consequences of misfolding on different cells and organs. In the liver, misfolding of the mutant protein results in the formation of toxic aggregates and hepatocyte death, manifest as liver injury and eventually cirrhosis—a gain-of-function toxicity. In the lung, the failure to secrete sufficient α_1 antitrypsin may lead to unchecked proteolytic damage to the delicate architecture of the alveolus, a process that is markedly worsened when neutrophils are recruited to the lung in response to cigarette smoking. This loss-of-function phenotype manifests pathologically as emphysema and clinically as chronic obstructive pulmonary disease.

INTERACTIONS WITH PN COMPONENTS THAT CHANGE CONFORMATION, STABILITY, OR FUNCTION (CANCER)

Mutations in the tumor suppressor p53 are among the most common mutations observed in patients with cancer. Deletion of p53 combined with overexpression of an oncogene is sufficient to drive metastatic cancer formation in mice, causally linking p53 mutations with cancer. Normally, p53 functions as a transcription factor that suppresses the transcription of genes involved in apoptosis resistance. While myriad mutations in p53 have been described, some result in an alternate conformation that interacts with different HSP70 chaperones within the PN. Binding of the mutant p53 protein to these chaperones affects the DNA binding property necessary for its tumor suppressor function and facilitates binding to other domains, resulting in changes in gene expression that protect malignant cells from apoptosis.

STRONGLY ENHANCED AGGREGATION PROPENSITY AND AMYLOID FORMATION (ALZHEIMER'S DISEASE, PARKINSON'S DISEASE, AMYOTROPHIC LATERAL SCLEROSIS, HUNTINGTON'S DISEASE, TYPE 2 DIABETES MELLITUS)

In some individuals, native or mutant proteins include sequence motifs that promote a highly ordered aggregation state when the cellular environment is altered. The most common of these motifs is the beta sheet, which, when exposed to the solvent environment of the cell, forms intermolecular species that bind in an iterative process that can accommodate many thousands of molecules that form cross-beta sheet amyloids. These intracellular aggregates are described as oligomers (2–24 molecules), protofibrils (rods 4–11 nm wide and 200 nm long), and amyloid fibrils with a similar width to protofibrils but microns in length. While the formation of oligomers is thermodynamically unfavorable, polymerization is favorable, causing aggregates to seed slowly but grow exponentially. In some cases, for example in Huntington's disease and familial forms of Alzheimer's disease and ALS, aggregation is accelerated by mutations or expansion of homopolymers. However, in many cases, the aggregates contain other cellular proteins that share biophysical properties of aggregation propensity or reflect dysfunction in the PN that facilitates their seeding or propagation (see below). While in most instances, damage caused by protein aggregates is localized to the cells in which they form, as occurs with islet amyloid peptide in some patients with type 2 diabetes mellitus, amyloidogenic proteins associated with neurodegenerative diseases have been shown to spread between cells and, in the case of transthyretin amyloidosis, can cause pathology in many tissues. Pathologists use staining of tissues with Congo red, which detects beta sheets, to make this diagnosis. Damage to neurons by aggregates in Alzheimer's disease can elicit a local inflammatory response by resident immune cells in the brain, both of which contribute to pathology. Much effort has been directed toward the detection of aggregates and amyloid and the development of small molecules or antibodies that block further growth or enhancing the cellular activities of the PN to suppress protein misfolding.

SECRETED AGGREGATED AND AMYLOID SPECIES CAUSING SYSTEMIC AMYLOIDOSIS

In patients with systemic amyloidosis, the secretion of large amounts of aggregation-prone proteins results in the deposition of aggregates in many tissues. These proteins can include immunoglobulins secreted from plasma cells in patients with systemic inflammation or multiple myeloma or other aggregation-prone proteins including transthyretin. Similar to other aggregate-induced diseases, mutations in transthyretin that enhance polymerization are associated with an increased risk of developing systemic amyloidosis with advancing age. These aggregates induce cellular toxicity, inflammation, and matrix reorganization, which interfere with function in an organ-specific manner.

NATIVE PROTEINS PRONE TO AGGREGATE WHEN THE CELLULAR ENVIRONMENT IS ALTERED BY STRESS AND AGING

While well-defined genetic abnormalities have been essential in elucidating the molecular mechanisms that underlie the formation of protein aggregates and causally linking them to disease, many if not most clinical diseases associated with the formation of protein aggregates develop in patients without identified mutations. In these patients, a decline in the chaperone and quality control mechanisms of the PN allows exposure of aggregation-prone domains of normal proteins to the solvent environment of the cell. Once seeded, these protein aggregates can expand rapidly to induce local or systemic injury. The decline in function of the PN that allows these aggregates to form might develop gradually with advancing age or might occur suddenly in response to an age-triggered biologic program, as occurs in *C. elegans*.

INFECTIOUS DISEASES AND IMBALANCED CELL STRESS RESPONSES IN AGING

The response to infectious diseases and the disproportionate morbidity and mortality in older individuals exposed to systemic stress are likely associated with the decline in robustness of cell stress responses and proteostasis. While these stressors include infections, surgical or accidental trauma, sepsis, and myocardial infarction, among others, pneumonia, the most common cause of death from an infectious disease in the United States, provides an illustrative example. As was evident during the COVID-19 pandemic, pneumonia morbidity and mortality disproportionately affect the elderly. Viral pneumonias, including those caused by influenza viruses and SAR-CoV-2, are primarily localized to the lung, where they activate a local and systemic inflammatory response and denude the alveolar lining. The resulting hypoxemia and systemic inflammatory response injure distant organs independent of viral injury. Impaired function of the PN during the stress might allow seeding of tissues with toxic aggregates with long-term consequences. Repair of the damaged lung and distant organs represents a major challenge to proteostasis that might be overcome in younger individuals but fails in those who are older with poor stress resilience. This loss of proteostasis resilience necessary to limit damage and allow repair could explain clinical observations in pneumonia survivors who develop persistent lung injury, skeletal muscle dysfunction impairing mobility, chronic kidney disease, cognitive dysfunction, and dementia and an increased risk of ischemic cardiovascular events in the year after hospital discharge.

FURTHER READING

Balch WE et al: Adapting proteostasis for disease intervention. Science 319:916, 2008.

Balchin D et al: In vivo aspects of protein folding and quality control. Science 353:aac4354, 2016.

Chiti F, Dobson CM: Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. Annu Rev Biochem 86:27, 2017.

Costa-Mattioli M, Walter P: The integrated stress response: From mechanism to disease. Science 368:eaat5314, 2020.

Eisele YS et al: Targeting protein aggregation for the treatment of degenerative diseases. Nat Rev Drug Discov 14:759, 2015.

Finley D, Prado MA: The proteasome and its network: Engineering for adaptability. Cold Spring Harb Perspect Biol 12:a033985, 2020.

Labbadia J, Morimoto RI: The biology of proteostasis in aging and disease. Annu Rev Biochem 84:43, 2015.

Levine B, Kroemer G: Biological functions of autophagy genes: A disease perspective. Cell 176:11, 2019.

Mallucci GR et al: Developing therapies for neurodegenerative disorders: Insights from protein aggregation and cellular stress responses. Annu Rev Cell Dev Biol 36:165, 2020.

Song J et al: Quality control of the mitochondrial proteome. Nat Rev Mol Cell Biol 22:54, 2021.

505

Novel Approaches to Diseases of Unknown Etiology

David R. Adams, Camilo Toro, Joseph Loscalzo

THE UNDIAGNOSED DISEASE STATE

The term *disease*, etymologically meaning "lack of ease" or the presence of discomfort, is defined as an abnormal state that negatively affects the structure or function of all or part of an organism and that is not due to any immediate external injury. When referring to a person experiencing a disease, the word *patient* is used in its original, meaning "the one who endures suffering." These terms are well suited to patients with undiagnosed diseases. A patient with an undiagnosed disease is one for whom a medical diagnosis is not discerned after reasonable efforts utilizing established methods and procedures. Multiple factors may contribute to a failure to reach a diagnosis **(Table 505-1)**. Patients who are affected by an undiagnosed disease for a protracted period exist in an undiagnosed state, which presents characteristic challenges for the patients, their families, and their medical providers.

THE MEANING AND CONTEXT OF A DIAGNOSIS

A diagnosis often entails hierarchical levels of information specificity with varying levels of relevance to the users (consumers) of such information (e.g.) patients and their families, health care providers, government health care agencies, insurers, epidemiologists, genetic counselors, pharmacologists, biologists). As an example, a diagnosis of Parkinson's disease in an adult is based on the progressive emergence of signs and symptoms of bradykinesia, rigidity, asymmetric rest tremor, and postural instability (clinical diagnosis), which are typically responsive to the administration of L-dopa (a therapeutic response biomarker). Together, these are cardinal features of striatonigral degeneration (a mechanistic diagnosis), a process associated with neuronal α-synuclein deposition and Lewy body pathology (histopathologic diagnosis) often based on a genetic susceptibility conferred by mutations in genes such as synuclein (*SYNCA,* a molecular diagnosis) and likely influenced by environmental exposures (e.g., manganese or other neurotoxins).

With ongoing advances in medical science and technology, the standard for what constitutes a reasonable diagnosis continues to evolve toward higher levels of specificity. For example, the utility of an antisense oligonucleotide therapy may be restricted to a specific subset of mutations associated with a given, monogenic, heritable disease. Efforts to adopt the principles of precision medicine include a growing emphasis on the context of disease within the genomic landscape, environment, social factors, medical history, nutrition, and the microbiome of any given individual. Examples include cancer susceptibility, genetically determined idiosyncratic reactions to medications, and unique pathogen susceptibilities in patients with certain immune deficiencies.

UNDIAGNOSED RARE DISEASES

Most chronically undiagnosed diseases are rare. While individual rare diseases have a low prevalence, they are numerous in aggregate. It is estimated that >6000 rare diseases affect millions of people throughout the world. Estimates of aggregate population prevalence range from 6 to 10%. Many rare diseases have a genetic basis and onset in childhood. As the cloud of uncertainty inherent in the undiagnosed disease state is removed, new disease-specific counseling, therapies, resources, community engagement, and advocacy opportunities become possible.

TABLE 505-1 Factors Contributing to the Presence of an Undiagnosed Disease

FACTOR	EXAMPLE
Misleading information	False-negative and false-positive test results
Rare disorder	Many inherited disorders have only been identified in a few individuals. For example, sialuria, a well-understood disorder of sialic acid metabolism, has been reported in 10 individuals (OMIM 269921).
Unusual causes of common diseases, including atypical course of illness	Insulin-dependent diabetes mellitus may be the presenting feature for the relatively rare autoimmune polyendocrinopathy syndrome, type I (OMIM 240300).
Presence of multiple disorders (blended phenotypes)	For an example, see PubMed ID 24863970.
Lack of characteristic symptoms of known disease	Diseases are commonly ascertained via cardinal signs or symptoms leading to incomplete ascertainment of all possible disease presentations. For instance, not all persons with Marfan's syndrome are tall relative to other family members. For progressive diseases, pathognomic signs and symptoms may be missing in early stages of disease.
New disease	No prior knowledge or record of such disease
Incorrect affected status assignments in family history	A heritable disorder may be inappropriately excluded if family history information is incorrect.
Primary disease manifestations obscured by other factors	Maladaptive behavior, medication effects, and secondary disease manifestations may obscure signs and symptoms of a primary disorder.
Disease not expected in region or population	Cystic fibrosis in persons of African ancestry, sickle cell disease in persons of northern European ancestry; infectious agents with marked geographical incidence patterns
Diseases thought to be eradicated	Poliomyelitis
Diseases occurring in unexpected time of life	Parkinson's disease in children, lysosomal storage disease in adults
Malingering	Feigned disease features intended to achieve secondary gain (Munchausen syndrome)
Rare disease mechanisms	Transmitted or sporadic prion disease, unusual zoonotic diseases

Abbreviation: OMIM, Online Mendelian Inheritance in Man.

THE EFFECT OF THE UNDIAGNOSED DISEASE STATE ON THE PATIENT

Patients with an undiagnosed disease are frequently driven to understand the basic nature of their ailment (what, when, where, how, etc.). Individuals, families, physicians, and society, however, might have a wide range of tolerance to the uncertainties associated with the undiagnosed disease state. Being undiagnosed has profound detrimental effects. Patients can go undiagnosed for decades, leading to personal and family uncertainty, high levels of stress, decreased productivity, limited accessibility to disease-specific counseling and resources, decreased quality of life, and excess utilization of medical services.

APPROACH TO CHALLENGING DISEASES OF UNKNOWN ETIOLOGY

Approaches to a patient with an undiagnosed disease can be separated into two categories. The first is a new assessment by a consultant, new provider, or diagnostic referral center. The second is periodic reassessment by an existing provider for a patient who remains undiagnosed.

TABLE 505-2 Essential Records for Undiagnosed Disease Patients

1. Any narrative summaries that detail the course of the illness
2. Copies of original test results with names, dates, testing circumstances, normal ranges, and test facility information
3. Electronic copies of imaging studies
4. Consultation notes
5. Hospitalization intake and discharge summaries
6. Accurate family history accounts and family relations

Optional but potentially useful records include:

1. Photographs and/or videos of disease manifestations
2. Longitudinal data (growth charts, symptom logs, serial lab measurements)
3. Data or specimens that could be reanalyzed, including pathology specimens and genomic sequencing of raw data

COMPREHENSIVE DATA COLLECTION

A potentially time-consuming but critical initial step is the gathering of all available medical data. Essential records are listed in Table 505-2.

The overall goal of data collection is a full understanding of the course of the disease and a verification of critical data elements used for diagnostic decision-making. Incorrect or partial second-hand accounts of prior test results contribute substantively to incorrect or missed diagnoses.

Analysis of the collected data allows for reconstruction of the process by which previous disease presentation, diagnostic thought processes, and test interpretation led to the current understanding of a patient's illness. Unintentional obfuscation of the history and findings can result from missing records, incomplete recall by the patient and fragmentation, and propagation of information (and misinformation) in the medical record. Optimally, the presence and character of key features of the illness will be reinforced by perspectives derived from multiple evaluations.

VALIDATION OF SUBJECTIVE AND OBJECTIVE FINDINGS

Teasing apart the layers of a patient's presentation often uncovers a variety of adaptive (and maladaptive) coping strategies. Some are idiosyncratic to the disease state (e.g., sun avoidance in a patient with xeroderma pigmentosum), whereas others are driven by psychosocial factors and could become primary drivers of the phenotype. It is important to consider, however, that patients believed to have "functional" or "somatoform" disorders may have unrecognized underlying illnesses, e.g., nonepileptic events (pseudo-seizures), frequently have concurrent bona fide epileptic events. Careful consideration of clinical phenomenology and associated findings on physical examination and ancillary investigations may provide clarity, affirmation, and effective redirection. Distinct clinical, radiographic, and laboratory abnormalities provide entry points to the generation of a differential diagnosis and could become effective biomarkers of disease progression and response to interventions.

Testing Strategies and New Technology The historical exclusion of a diagnostic hypothesis may be based on testing that is no longer state of the art. For example, congenital disorders of glycosylation (CDG) were historically diagnosed using transferrin isoelectric focusing. It was subsequently found that the diagnosis of many CDG types required mass spectrometric and molecular approaches. The initial assessment of a patient with an undiagnosed disease should include a reassessment of the diagnostic logic and data used in past decision-making. In the absence of concrete diagnostic leads, the use of broad scope screening tools may prove beneficial in generating meaningful diagnostic hypotheses (Table 505-3). In some cases, newer testing options may be difficult to obtain and/or be costly. Prior probability of disease and available resources will factor in determining whether new diagnostic testing is practical.

TABLE 505-3 Clinically Available Tests with Notable Utility for Undiagnosed Cases

TEST	TARGET PHENOTYPES	RATIONALE
Single nucleotide polymorphism microarray and/or karyotype	Dysmorphic features, cognitive impairment, neurodevelopmental disorders	Genomic structure abnormalities may be missed by other testing
Exome or genome sequencing	Any undiagnosed disease that is chronic and not clearly acquired	Tests a broad range of potentially unconsidered diagnostic entities
Lysosomal storage diseases (LSDs), molecular or enzymatic tests. urine organic acids, urinary glycosaminoglycans (GAGs), oxysterols	Progressive neurologic disorders, psychiatric disorders	Some LSDs have nonspecific presentations, and adult-onset cases are often missed
Congenital disorders of glycosylation, Apo CIII and N-glycan mass spectrometry	Pediatric-onset disorders, cognitive impairment, neurologic phenotypes	Large group of disorders; phenotypes for many still being characterized
Biochemical disorders, ammonia, serum polyols, urine purines and pyrimidines, plasma amino acids, very-long-chain fatty acids	Neurologic phenotypes, especially with waxing and waning course, selective speech involvement, or patients with unusual self-selected diets	Metabolic disorders may have nonspecific symptoms, and adult-onset cases are often missed
Mitochondrial sequencing and mitochondrial depletion studies; biochemical screening with serum lactate, blood pyruvate, plasma amino acids, and GDF-15	Complex multisystem disorders with neurometabolic, endocrine, and gastrointestinal symptoms, muscle dysfunction, and waxing and waning or progressive course	Large group of disorders with a wide range of presentations; yield is improved by studies in affected tissue (e.g., liver or muscle)
Cerebrospinal fluid (CSF) studies including amino acids (AAs), lactate, pterins, methyltetrahydrofolate (MTHF), or special CSF flow studies	Synthetic neurotransmitter defects in patients with unexplained fluctuating encephalopathy/movement disorders or patients with atypical neuroinflammatory syndromes	Patterns of profiles point to particular enzymatic deficits in neurotransmitter synthesis or characterization of unique immunologic profiles of inflammatory central nervous system diseases

Molecular Approaches, Including Genomics The availability and variety of clinical molecular modalities have transformed diagnostic testing in many settings. These advances have arisen from both testing scope, such as exome-wide, genome-wide, and transcriptome (RNA sequencing [RNA-Seq]) sequencing, and new medical knowledge, such as new disease-gene associations and molecular interaction networking (network medicine). Complementary screening tools, such as metabolomics, show diagnostic promise, particularly when combined with sequencing data to generate a fuller picture of disease manifestations. Simultaneous consideration of multiple data types can provide a means of appreciating overlapping and reinforcing evidence, with the potential to inform both hypothesis-driven and agnostic approaches to diagnosis.

HYPOTHESIS-DRIVEN MOLECULAR TESTING Hypothesis-driven testing implies that a defined set of heritable (or potentially heritable) disorders is the principal impetus for testing. Selection of a targeted gene sequencing panel, ideally augmented with structural variant detection, may allow for improved sensitivity, lower cost, and fewer unrelated (secondary) findings relative to full exome or genome sequencing studies. In the setting of an initial undiagnosed disease evaluation, prior sequencing panels may not include recently discovered genes. Testing with an updated panel or targeted sequencing of a newer gene is an option for consideration. In some cases, sequencing panels are generated by selective reporting of relevant genes within an exome dataset. In such cases, it may be possible to expand the analyses to new genes of interest without additional sequencing.

AGNOSTIC MOLECULAR TESTING Agnostic testing typically uses data from a broad testing platform, such as exome or genome sequencing, and considers all detectable diagnoses, even those with a low pretest probability of being present. This approach can also generate hypotheses for potentially new disease-gene associations. Analysis of the sequencing data typically includes an unrestricted search throughout the entire human genome or exome space. DNA sequence variants with potential medical relevance are identified first by bioinformatic characteristics, including known association with disease, predicted importance for protein function, interspecies conservation, population frequency, and an evolving list of other associated or functional factors. The list of candidate variants is then subject to expert review (i.e., curation). The interpretation of test results in this setting is highly influenced by both the adequacy of communication between the clinical and testing teams and the information content of the data sources used to annotate each of the thousands of variants generated in the course of sequencing.

There is a rapid proliferation of new testing platforms and analytical tools with the potential to contribute to solving undiagnosed diseases, but it remains challenging to judge their broad utility. While awaiting systematic validation and practice standards, novel techniques may be considered in special cases where a diagnostic hypothesis is closely aligned with the type of data generated by a specific testing strategy (**Table 505-4**).

PERIODIC REEVALUATION

The cornerstone for the care of a patient in an undiagnosed disease state is a plan for periodic reevaluation. The Undiagnosed Diseases Network (UDN), a 10-year National Institutes of Health–sponsored national program, was specifically designed to evaluate undiagnosed patients. The overall diagnostic rate of the UDN, including periodic reevaluation of early enrollees, was reported as ~30%. This finding illustrates the fact that many affected individuals remain in an undiagnosed state for a protracted period. For a medical provider, the care of an undiagnosed patient includes a program for symptomatic care, support related to the undiagnosed state itself, and plans for a regular reevaluation strategy seeking new insights into the diagnosis by following its time trajectory. Reevaluation is guided by emerging knowledge in the field, disease progression, and the development of new signs and symptoms. The appearance of a similar disease in a sibling or close relative may provide critical insight. Communication with the patient is an essential component. Many individuals with an undiagnosed disease report feeling abandoned by their providers once initial diagnostic ideas have been exhausted. Providers themselves may feel discouraged in being unable to provide a diagnosis. The institution and discussion of a well-defined plan for periodic reassessment and communication can help reinforce the patient-provider relationship and set reasonable expectations.

TABLE 505-4 Emerging or Special Testing Strategies and Related Diagnostic Questions

TESTING STRATEGY	RELATED DIAGNOSTIC QUESTION	AVAILABLE CLINICALLY[a]
Transcriptomics, RNA-Seq	Relevance of splice, regulatory, and other noncoding variants; correlated changes in gene expression within pathways	Yes
Metabolomics	Hypothesis generation via nontargeted approaches, correlated pathway changes, correlation with molecular findings	Yes
Epigenetics	Diseases known or suspected to be caused by methylation or parent-of-origin effects	Yes
Transcriptional profiling	Search for profile particular to certain disease states, e.g., interferon-inducible gene panels (interferon signature) in certain autoinflammatory disorders	Some
Specialized, disease-specific testing	Prion-related diseases, metabolic diseases, and many other assays	Some
Functional validation	Model organisms, cell biology, and other approaches to validating a hypothesized gene-disease association	No
Metagenomics	Search for molecular fingerprints of other organisms (e.g., infectious agents) within human samples	Yes
Long-read sequencing technology	Accurate resolution of low-complexity regions of the human genome (repeat expansion disorders) and complex genome structural rearrangements	No
Deep sequencing	Accurate resolution of low levels of mosaicism	Some
Optical genome mapping	High-resolution chromosomal structure	Yes

[a]Available clinical tests are often a small subset of approaches available via research collaboration. Clinical testing offerings are evolving rapidly and should be reassessed periodically.

Reevaluation of Differential Diagnosis The key to success for a planned reevaluation visit is preparation. The problem and differential diagnosis lists should be subject to careful, evidence-based review. New or resolved clinical features may add or remove diagnostic considerations. The passage of time may result in the emergence of distinct new phenotypic manifestations that serve as new clues in the formulation of a definitive diagnosis. Special consideration should be given to the effects of reaching maturity and aging. The establishment of a phenotype as being static versus progressive has prognostic value. Careful documentation of the rationale for including or excluding individual disorders will streamline the process for both future reevaluations and the need for consultants. Concurrent development of common diseases should be thoughtfully considered as a possible component of the primary undiagnosed condition. For example, emergence of insulin-dependent diabetes mellitus in an undiagnosed patient with a complex phenotype could be a feature of the rare autoimmune polyendocrinopathy associated with mutations in the autoimmune regulator gene *AIRE*.

New Literature Keeping abreast of current literature is an important and challenging activity for all medical providers as the body of medical knowledge continues to grow exponentially. For undiagnosed diseases, newly reported disorders and disease-gene associations are an important source of diagnostic resolution. Literature search tools such as PubMed can be augmented by online resources that connect clinical signs and symptoms (phenotypes) to disorders. For example, using the search terms "cardiomyopathy arthropathy diabetes hyperpigmentation" in the Online Mendelian Inheritance in Man website (*https://omim.org*) produces a list of disorders that includes hemochromatosis. In the context of an undiagnosed disease, this type of phenotype-driven approach can be used to search for new, relevant publications and disorders. Tools for search automation continue to be developed in both open-source and commercial settings. The success of these approaches is augmented by iterative application, ideally as part of formal, periodic reevaluation of the undiagnosed patient.

GENOMICS

The use of medical testing based on the determination of DNA sequence and structure (sometimes referred to as molecular testing) has proliferated in recent years. A wide variety of approaches are available to the clinician, from single-gene sequencing to exome or genome sequencing. RNA sequencing and optical genome mapping have recently become available as clinical tests. Many reviews of this topic are available (see Marwaha et al., 2022, in "Further Reading"). Consultation with colleagues trained in genetics can be useful when developing an optimal testing approach.

In some cases, genetic testing results may already exist in the medical record during the initial evaluation of an undiagnosed patient. This is increasingly true for younger patients; exome and genome sequencing are being used earlier, and with increasing frequency, for complex diagnostic challenges. Reanalysis of previously obtained exome and genome data should start with consideration of both the age and quality of the study and the reported patient phenotype at the time of the study report. For sequence results generated in a clinical laboratory, a discussion between the provider and the laboratory director often answers important questions about recommended next steps. The discussion should touch on how technologic advances have affected the utility of the older data and whether the laboratory offers reanalysis of the data. At a minimum, the provider, the testing laboratory, or an identified subspecialist should review previously reported DNA variants of unknown significance considering interval reports about the gene in question. More advanced reanalysis strategies are emerging and may be offered by the testing laboratory.

Some laboratories offer release of raw DNA sequencing data to their patients on request. The utility of raw data varies and depends on the identification of bioinformatics collaborators willing to reanalyze the data. Sequencing data obtained as part of a research study may not be suitable for clinical diagnostic purposes. In practice, raw and research-generated sequencing data are most useful when a collaborating researcher can be identified.

When considering a new sequencing test, the inclusion of the parents and siblings of the proband has the potential to provide enormous value in some situations. Discussion of an optimal approach with an expert colleague or the testing laboratory is encouraged.

EXPOSOME

In many cases, a detailed occupational and environmental exposure history should be obtained. Some rare disease phenotypes are pathognomonic of specific toxicant exposures (e.g., mesothelioma and asbestos exposure, clear cell adenocarcinoma of the vagina and intrauterine diethylstilbestrol [DES] exposure, chloracne and exposure to halogenated aromatic hydrocarbons). For the most part, however, chemical toxicant exposures do not produce unique phenotypes. Rather, chemical exposures operate in conjunction with lifestyle factors (e.g., smoking, alcohol intake, and nutritional status), differential host susceptibility (determined by age, sex, comorbidities, genetics, etc.), and nonchemical stressors (e.g., psychosocial stress) to produce (1) common, readily diagnosed medical diseases (e.g., asthma); (2) unusual or nonspecific phenotypes (e.g., erethism and metallic mercury exposure); or (3) atypical presentations of otherwise well-characterized disease states, initially considered an undiagnosed disease (e.g., manganese-induced parkinsonism). The nonspecificity typical of chemical-induced disease risk is further complicated by lack of exposure biomarkers for many common environmental toxicants (e.g., volatile organics), the short half-life of some contaminants (e.g., arsenic), and the possibility of decades long latency between exposure and disease onset (e.g., chemical carcinogenesis or dietary exposures to specific biochemical risk factors for atherothrombosis). In addition, we live in an era in which new chemicals are introduced into consumer products and the environment at a pace well beyond our capacity to characterize their toxicity. Within this context, one of the most powerful tools for ascertaining chemical-related disease risk is a systematic exposure history. Although there are no standardized instruments for this purpose, there are published guidelines to implement exposure assessments (Goldman and Peters, 1981; see "Further Reading" below). These include a multistep approach to exposure assessment including a job history; a review of exposures at work and at home or via hobbies and recreation; ascertainment of any temporal relationship of symptoms or disease onset to work, home, or recreational activities; and the food frequency questionnaire. If this screening identifies a potential exposure or exposures of concern with respect to patient symptoms and phenotype, a second step of evaluation involves a more detailed history to identify specific suspect agents, options for quantitative environmental exposure assessment (e.g., household tap water sampling, review of workplace Material Safety Data Sheets [MSDSs]) and biomonitoring, and etiologic plausibility for at least some aspects of the patient's phenotype.

The traditional approach to focused external exposure assessment proposed above does not, however, provide an integrated, quantitative measure of all exposures over the life course, an exposure characterization of particular interest for the risk of chronic diseases such as cancer or atherothrombosis. The exposome has been proposed as a promising means for capturing the totality of human exposure over a lifetime (analogous to the totality of genetic exposure assessed via genomic analyses), including not only external chemical or dietary/foodome (Barabasi et al., 2020; see "Further Reading" below) exposures but also internal (e.g., metabolic, hormonal, microbiome) influences and psychosocial factors. However, techniques for measuring the exposome are in relatively early stages of development, are limited by the substantial variability in human exposure experience, and have not yet been designed to capture complex combinations commonly encountered in environmental or occupational exposure settings (Peters et al., 2012; Wild, 2012; Brunekreef 2013; Barabasi et al., 2020; see "Further Reading" below). This important element of assessing patients with undiagnosed disease is, however, evolving rapidly and offers the promise of becoming a more formal part of the evaluation of many patients with undiagnosed disease.

ENGAGEMENT OF RESEARCH APPROACHES

Establishing a research collaboration for a patient with undiagnosed disease can be both challenging and rewarding. Time and effort resources are likely to limit this approach to a subset of patients with particularly compelling clinical presentations and a strong hypothesis about disease causation. The process must include early and detailed communication with the patient. Several approaches may be considered. Undiagnosed disease medicine, as a focus of specialized study, investment, and infrastructure, has proliferated around the world in the last decade, with numerous centers providing support for evaluating qualifying individuals and families.

Leveraging Phenotypic and Genotypic Similarities For a patient with a rare or undiagnosed disease and distinct presenting features, finding similarly affected individuals adds substantial benefits. It can encourage research, provide a community for affected patients, and improve the chances of finding commonalities in pathogenesis and therapeutic strategies. Phenotypic aggregation may also allow the patient to connect with consortia invested in related medical presentations. Examples include organizations dedicated to the study of related diseases such as leukodystrophies, autoinflammatory disorders, and even undiagnosed diseases; NORD, the National Organization for Rare Disorders (*rarediseases.org*), can be a useful starting point. The Office of Rare Disease Research within the National Center for Advancing Translational Science supports consortia under the Rare Disease Clinical Research Network program. Building patient cohorts may also be based on specific biological mechanisms or pathways, for example, the United Mitochondrial Disease Foundation.

Data Sharing The proliferation of DNA sequencing technology and the subsequent generation of many DNA variants of unknown clinical significance have prompted the creation of data-sharing resources specifically designed to match similar cases submitted by clinicians and researchers around the world. For example, a clinical exome report may identify variants in a gene with a potential but unproven relationship to the patient's presenting illness. The clinician could enter the gene name into a gene-matching database, and if the same gene name had been already entered by a different submitter, the database would flag a match and send contact information to both submitters. The matching procedure has the potential to identify additional cases of an ultrarare or newly described condition, while avoiding the sharing of the patient's personal health information. Embellishments of this approach involve inclusion of phenotypic features, data entry by families, and specific details of sequence variants. Example systems include GeneMatcher (the most populated database for single-gene submissions) and DECIPHER (which adds expended utility for larger structural variants). As an illustration of their utilization at the time of this chapter's publication, GeneMatcher has single-gene submissions for almost 70% of all known protein coding genes.

Collaboration Collaborations around undiagnosed disease patients may take many forms. Studies focusing on related medical conditions can sometimes be identified using the *https://clinicaltrials.gov* website, which lists many U.S. and non-U.S. clinical studies. Databases of clinical information (e.g., this textbook, GeneReviews) can be used to identify subject matter experts for related conditions. Such

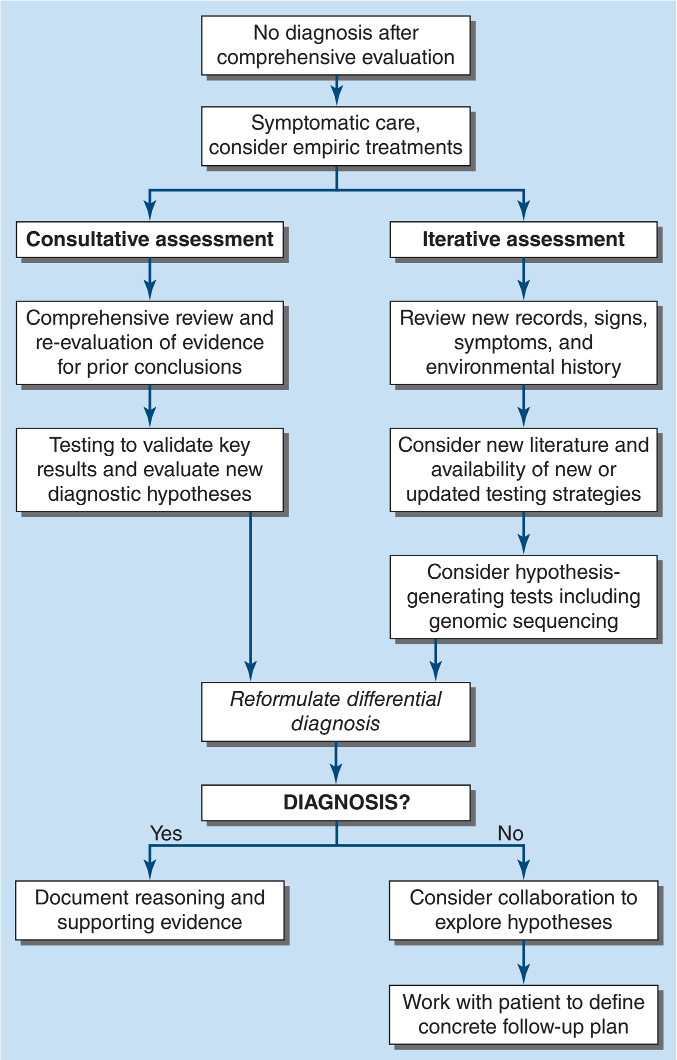
No diagnosis after comprehensive evaluation
Symptomatic care, consider empiric treatments
Consultative assessment
Iterative assessment
Comprehensive review and re-evaluation of evidence for prior conclusions
Review new records, signs, symptoms, and environmental history
Testing to validate key results and evaluate new diagnostic hypotheses
Consider new literature and availability of new or updated testing strategies
Consider hypothesis-generating tests including genomic sequencing
Reformulate differential diagnosis
DIAGNOSIS?
Yes
No
Document reasoning and supporting evidence
Consider collaboration to explore hypotheses
Work with patient to define concrete follow-up plan

FIGURE 505-1 Approach to the patient with an undiagnosed disease.

experts can be queried about ongoing studies. In some cases, a willingness to work with consenting families to provide biological specimens can open additional avenues for collaboration.

CHALLENGES

Data Portability Obtaining specimens, data, and records for a chronically undiagnosed patient can be time-consuming and challenging. Families may be charged fees for obtaining copies of old studies. Although continuing advances in record access are occurring, families should be encouraged to collect and maintain an updated collection of medical records. These should include copies of consultation notes, original laboratory results, and radiology studies (the latter preferably in electronic form). These record collections are useful for consultation, second opinions, and transitions between primary providers.

Managing Illness Behaviors, Expectations, and Secondary Manifestations Patients with undiagnosed diseases may present in any stage of the grieving process. Coping with uncertainty, loss of abilities, work, relationships, autonomy, and financial security compound the primary manifestation of the disease. Patients may have a wide range of expectations about the possible benefits of achieving a diagnosis, including successful therapy. Patients of reproductive age may find that their greatest uncertainty surrounds the potential heritability of their disorder, its effects on future reproductive decisions, and the potential risk it may represent to their children and living relatives. These factors may be equally or more disabling than the primary illness and require an individualized and multidisciplinary approach.

CONCLUSION

Chronically undiagnosed diseases present a complex challenge to patients, medical providers, and society at large. Development of a comprehensive plan for evaluation, reevaluation, and support requires a substantial investment of time and effort **(Fig. 505-1)**.

Achieving an accurate diagnosis removes at least one level of uncertainty and allows for disease-specific counseling, therapies, resources, community engagement, and advocacy opportunities otherwise not afforded to undiagnosed patients.

FURTHER READING

Barabasi AL et al: The unmapped chemical complexity of our diet. Nat Food 1:33, 2020.

Brunekreef B: Commentary: Exposure science, the exposome, and public health. Environ Mol Mutagen 54:596, 2013.

Goldman RH, Peters JM: The occupational and environmental health history. JAMA 246:2831, 1981.

Lee CE et al: Rare genetic diseases: Nature's experiments on human development. iScience 23:101123, 2020.

Marwaha S et al: A guide for the diagnosis of rare and un-diagnosed disease: Beyond the exome. Genome Med 14:23, 2022.

Peters A et al: Understanding the link between environmental exposures and health: Does the exposome promise too much? J Epidemiol Community Health 66:103, 2012.

Splinter K et al: Effect of genetic diagnosis on patients with previously undiagnosed disease. N Engl J Med 379:2131, 2018.

Wild CP: The exposome: From concept to utility. Review. Int J Epidemiol 41:24, 2012.

This page intentionally left blank

Index

Page numbers in **bold** indicate the start of the main discussion of the topic. Page numbers followed by "f " or "t" refer to the page location of figures and tables, respectively. Location entries starting with "A" (Atlases), "S" (Supplemental chapters), or "V" (Video chapters) indicate online-only chapter numbers; this content is available to all *Harrison's* readers at *accessmedicine.com/harrisons*. Index entries that end with a "v" represent book page numbers where video content is referenced and entries that begin with "E" are followed by the chapter number and "f " or "t" for the location of online-only figures and tables, respectively.

A
AA amyloidosis, **899**
  in axial spondyloarthritis, 2884
  cardiac disease in, 2027t
  clinical features of, 896t, 899, 2027t, 2415t
  diagnosis of, 2425
  etiology of, 899
  incidence of, 899
  pathogenesis of, 899, 2027t, 2425
  treatment of, 899, 2425
AAG (autoimmune autonomic ganglionopathy), 3550
AApoA1 amyloidosis, 2027t
AApoA4 amyloidosis, 2027t
AAS abuse. *See* Anabolic-androgenic steroid (AAS) abuse
AAT. *See* α_1-Antitrypsin (AAT)
AAV (adeno-associated virus), 1484, 1525, 3524, 3823t, 3827
Abacavir
  adverse effects of
    DIHS, 419, 428f, 1603. *See also* Drug-induced hypersensitivity syndrome (DIHS)
    genetic considerations in, 419, 1585, 1603
    in HIV infection, 1603, 1617t
  in combination formulations, 1621t
  genetic variations in response to, 493, 494, E72t
  for HIV infection, 1617t
  molecular structure of, 1619f
  resistance to, 1622f
Abaloparatide, 3308–3309
A band, of sarcomere, 1847
Abatacept
  adverse effects of, 2850t
  for autoimmune diseases, 2791t
  for COVID-19 disease, 1542
  for giant cell arteritis, 2903
  for granulomatosis with polyangiitis, 2899
  for psoriatic arthritis, 2890
  for rheumatoid arthritis, 2850t, 2851
  for systemic sclerosis, 2874
ABAT deficiency, 3368t
ABCA1 deficiency (Tangier disease), 1860, 3241t, 3248, 3602
ABCB1, 484
ABCB6 gene mutations, 404, 794t, 795
ABCD1 gene mutations, 3397
ABCD^2 score, 3448, 3448t
ABCDE rule, 600, A5
ABCG5/G8 gene mutations, 3243
ABCG5/G8 hemitransporter, 2727
Abciximab
  actions of, 943f, 945
  adverse effects of, 923, 923t, 946
  discontinuation before lumbar puncture, S3
  dosage of, 945, S13
  genetic variations in response to, 939t
  indications for, 945
  in PCI, 2127
  pharmacology of, 945t
Abdomen
  paradoxical movement of, 270
  physical examination of, 327, 1860
  POCUS, 3912t, 3913, 3914v
  swelling/distention of, 48t, **326,** 327f, 2591. *See also* Ascites
Abdominal angiostrongyliasis, 1819
Abdominal aortic aneurysm. *See* Aortic aneurysm, abdominal
Abdominal obesity, 3166
Abdominal pain, **109**
  in acute intermittent porphyria, 3346
  in appendicitis, 2595, 2595t
  approach to the patient, 112–113
  differential diagnosis of, 112t, 2460–2461, 2460t
  in Fabry's disease, 3355
  in familial Mediterranean fever, 2934
  history in, 109t, 112
  in IBS, 2571
  imaging in, 113
  in immunocompromised patients, 111
  in intestinal ischemia, 2587
  in intestinal obstruction, 2591
  in malabsorption disorders, 2549t
  mechanisms of, **110,** 110t
  metabolic causes of, 111
  neurogenic causes of, 111–112
  in pancreatic cancer, 673
  in pancreatitis, 112t, 2745, 2752, 2753–2754
  in pelvic inflammatory disease, 1105
  in peptic ulcer disease, 2522
  in peritonitis, 1070, 2597
  referred, 111
  "stop rules" for, 49t
Abdominal paracentesis, 328, 329
Abdominal reflexes, 3381
Abdominal wall disorders, 111, 330
Abdominojugular reflex, 1861, 1983
Abducens nerve, 3380
Abducens nerve palsy, 235, V3
Abducens nerve paresis, 234
Abduction and external rotation test, 2176
Abemaciclib, 533t, 570t, 573, 637, 753t
Aβ (amyloid beta)
  in AD, 212, 3404, 3476–3477, 3478f, 3479f
  circadian fluctuations in accumulation of, 3952
  as drug target, 3480–3481, 3481f
  in neurodegenerative diseases, 3401, 3401t
  in protein aggregation and cell death, 3399f, 3400
Aβ amyloidosis, 895
Aβ antibodies, 3480–3481, 3481f
Aβ fibers, 93, 93f
Aβ/tau protein, 195t, 197–198
Aβ vaccine, 3482
Aβ_2M amyloidosis, 895, 896t, 900
Abetalipoproteinemia (acanthocytosis, Bassen-Kornzweig syndrome)
  clinical features of, 232, 309, 3241t, 3247
  genetic considerations in, 3247, 3437
  pathophysiology of, 2545, 3247, 3437
  peripheral blood smear in, 437, 440f
  treatment of, 2613
  vitamin E deficiency in, 2612
ABI (ankle-brachial index), 1862, 1863f, 2172
Abiotrophia spp., 1041, 1214
Abiraterone, 567, 700, 702, 702f, 703
ABL1 gene mutations, 836
Ablation
  alcohol septal, A11
  catheter-based. *See* Catheter-based ablation
  for chronic venous disease, 2182
  with stereotactic radiation, 1968, 1968f
ABL gene mutations, 520t, 521
Abnormal uterine bleeding, 3132
ABO RBC blood group system, 901, 903t
Abortion, 1305, 1376, 1467, 3136, 3162
Abortive infection, 1481
ABPA. *See* Allergic bronchopulmonary aspergillosis (ABPA)
Abrocitinib, 384
Abscess
  anorectal, 2585, 2585f
  brain. *See* Brain abscess
  epidural. *See* Epidural abscess
  head and neck, 259
  intraabdominal. *See* Intraabdominal abscess/infections
  periapical, 260
  peritonsillar, 259, 1373
  perivalvular, 1038, 1047
  retropharyngeal, 259
  submandibular, 259
  subperiosteal, 255
  tuboovarian, 1365, 1365f, 1376
Abscopal effect, 481
Absence seizures, 3410, 3418f
Absent (vanishing) testis syndrome, 3098
Absolute reticulocyte count, 447
Absolute risk reduction (ARR), 28
Absorption
  of antibacterial drugs, 1167
  of carbohydrates, 2542
  disorders of. *See* Malabsorption disorders
  of drugs, 484, 1167
  of lipids, 2541–2542, 2541t, 2542f
  luminal phase, 2539, 2542
  overview of, 2541
  post-mucosal, 2539, 2548
  of proteins, 2542
  small intestinal mucosal phase, 2539, 2543–2544
Absorption spectrum, 429
Abstract thought, assessment of, 3380
Abulia, 185, 3430, 3431f
ABVD regimen, 870
Acalabrutinib
  action and targets of, 533t, 567, 568t
  adverse effects of, 568t, 753t, 754
  for CLL, 855
  for mantle cell lymphoma, 866
  for Waldenström's macroglobulinemia, 894
Acalculia, 203
ACAM2000, 1149
Acamprosate, 2710, 3681
Acanthamoeba spp., 1758, S12
Acanthamoeba spp. infections
  clinical features of, 1758–1759
    chronic meningitis, 1131t
    corneal ulcers, 1739t
    keratitis, 224, 1737, 1759
    meningoencephalitis, 1114, 1759
    seizures, 1738t
  diagnosis of, 1759, 1759f, S12
  epidemiology of, 1758, S12
  invasive, 1737
  treatment of, 1759
Acanthocytes, 437, 440f, 446f
Acanthocytosis. *See* Abetalipoproteinemia
Acantholysis, 409

Acanthosis, 2184f, 2790t
Acanthosis nigricans
  anti-insulin receptor antibodies in, 3092
  in diabetes mellitus, 3202, 3230
  disorders associated with, 397, 619, A15
  hyperpigmentation in, 397, A5
  in metabolic syndrome, 3256
  in obesity, 3187
Acarbose, 2533, 3211t
Accelerated idioventricular rhythm (AIVR), 1955f, 1961–1962, 1963f, 2124, A8
Accelerated junctional rhythm, 1940, 2124
Accessory pathways (APs)
  arrhythmias associated with, 1940
  characteristics of, 1940
  in preexcited tachycardias, 1941, 1942f
  treatment of, 1941–1942
  in Wolff-Parkinson-White syndrome. *See* Wolff-Parkinson-White syndrome
Accidents, 75t
ACD gene mutations, 3820t
ACE. *See* Angiotensin-converting enzyme (ACE)
ACE2, 3942
Acebutolol, 2100t, 3712t
ACE inhibitors. *See* Angiotensin-converting enzyme (ACE) inhibitors
Acephalgic migraine, 3428, 3462
Aceruloplasminemia, 2615, 3335
Acetaminophen
  for acute otitis media, 254
  adverse effects of
    constipation, 509
    hepatotoxicity, 489, 2667t, **2668,** A13
    metabolic acidosis, 370
    nephrotoxicity, 96
    thrombocytopenia, 923t
  drug interactions of, 1422, 2955t
  for fever, 136
  for fever in COVID-19 disease, 1541
  for fever of unknown origin, 155
  metabolism of, 2668–2669
  for migraine, 3467t
  for neck pain, 132
  for osteoarthritis, 2955, 2955t
  for osteoarthritis in older adults, 3883
  overdosage/poisoning with, 2668
  for pain, 80f, 81, 96–98, 97t
  for prevention of acute mountain sickness, 3756
  for tension-type headache, 3471
  for upper respiratory infection, 253
Acetaminophen-aspirin-caffeine, 3466, 3467t
Acetanilide, 799t
Acetazolamide
  actions of, 2368
  for acute angle-closure glaucoma, 225
  adverse effects of, 346, 375, 421, 2356, 3756
  for episodic ataxia, 3536
  for heart failure, 1993t
  for high-altitude illness, 3756, 3756t, 3758
  for hyperkalemic periodic paralysis, 3646
  for hypokalemic periodic paralysis, 3646
  for metabolic alkalosis, 375
  for paroxysmal symptoms in MS, 3588
  for pseudotumor cerebri, 229
  for raised CSF pressure headache, 117
  for salicylate-induced acidosis, 370
  for tumoral calcinosis, 3319
Acetylcholine (ACh), 293, 302, 3538
Acetylcholine receptor (AChR)
  in congenital myasthenic syndromes, 3627, 3629t
  in myasthenia gravis, 2790t, 3624–3625, 3626
Acetylcholine receptor (AChR) antibodies, 744t
Acetylcholinesterase (AChE), 3627, 3629t
Acetylcholinesterase (AChE) inhibitors
  for gastroparesis, 299, 299t
  nerve agents as, S5
  overdosage/poisoning with, 3712t
  for snakebite treatment, 3721, 3721t
Acetyl-CoA, 3939
N-Acetylglutamate synthase deficiency, 3369t, 3372
N-Acetyltransferase, 485t
Achalasia, **2507**
  chronic, esophageal cancer and, 642
  clinical features of, 2507
  diagnosis of, 2507–2508, 2507f
  differential diagnosis of, 2507
  dysphagia in, 295
  LES relaxation impairment in, 2507, 2508f
  subtypes of, 2508f
  treatment of, 2469, 2476f, 2508, V5
Achilles bursitis, 2973
Achilles tendinitis, 2968
Achondrodysplasia, 3318, 3784
Achondroplasia, 3330
AChR. *See* Acetylcholine receptor (AChR)
Achromatopsia, 207, 222
Achromobacter xylosoxidans, 1267–1268, 1268t
Acid anhydride exposure, 2239t, 2240
Acid-base disorders, **366,** 367t, 368t, S1. *See also specific disorders*
Acid-base homeostasis, 366–367
Acid-base nomogram, 367, 367f
Acidemia, 2745, 3367
Acid ethyl esters, 3244t
Acid-fast bacteria, S11
Acid fume exposure, 2239t
Acid α-1,4 glucosidase deficiency (Pompe disease, type II GSD), 3354t, **3357,** 3361t, **3363,** 3644
Acinetobacter spp., 1294
Acinetobacter spp. infections, **1294**
  antibiotic resistance in, 1152, 1186, 1186t, 1295–1297
  bloodstream, 1296
  clinical features of, 1296
  community-acquired, 1295
  complications of, 1297
  diagnosis of, 1294, S11
  in disasters, 1295
  epidemiology of, 1294–1295
  etiology of, 1294
  health care–associated, 1296
  infection control and prevention of, 997f, 1298
  meningitis, 1296
  pathogenesis of, 1295–1296
  pneumonia, 1296
  prognosis of, 1297
  in returning war veterans, 1295, S7
  skin and soft tissue, 1296
  treatment of, 1175, 1176, 1296–1297, 1297t, S7
  UTIs, 1296
  ventilator-associated pneumonia, 1031
ACIP (Advisory Committee on Immunization Practices), 1000, 1003
Acitretin, 387, 387t, 393t, 394, 421
ACKR1 gene mutations, 493
Acneiform eruptions, 395, 395t
Acne rosacea, **391**
  clinical features of, 378t, 380f, 391, 391f, 396, A5
  complications of, 391
  differential diagnosis of, 395
  treatment of, 391
Acne vulgaris, **390**
  clinical features of, 378t, 380f, 390–391, 391f, 404, A5
  drugs exacerbating, 391
  folliculitis in, 1051
  pathophysiology of, 390
  treatment of, 391
Acoramidis, 2027
Acotiamide, 302
Acoustic neuroma. *See* Vestibular schwannoma (acoustic neuroma)
Acoustic reflex, 249
Acoustic reflex decay, 249
Acral angiodermatitis, 406
Acral erythema, drug-induced, 421
Acrocephalopolysyndactyly (Pfeiffer's syndrome), 3787
Acrochordon, 378t, 380f, 398
Acrocyanosis, 1859, 2177f, **2178**
Acrodermatitis chronica atrophicans, 1451
Acrodermatitis enteropathica, 2614
Acrodysostosis, 3288t, 3289
Acrolein exposure, 2239t
Acromegaly, **3008**
  clinical features of, 3008, 3009f, 3647
    arthropathy, 2966
    cutaneous, 397
    hirsutism, 3138
    hypertension, 2140
    oral, 261
    skeletal, A15
  diagnosis of, 3008
  etiology of, 2966, 3008, 3008t
  familial, 3004
  laboratory investigation of, 3001t, 3008
  vs. pachydermoperiostosis, 3318
  paraneoplastic, 737t
  treatment of, 3008–3009, 3010f
Acro-osteolysis, 2868, 2868f
Acropachy, thyroid, 2971, 3034, 3034f
Acroparesthesia, in Fabry's disease, 3355
Acrylamide, 3610t
ActA, 1227
ACTA2 gene mutations, 3332t, A16
ACTH. *See* Adrenocorticotropic hormone (ACTH)
ACTHoma, 678t, 679
ACTH stimulation test, 3051
Actin, 1848, 1848f, 2005f, 2790t
Actinic cheilitis, 607
Actinic dermatitis, 434
Actinic keratosis
  clinical features of, 379t, 380f, 608f, A5
  conversion of, 607
  malignant conversion of, 431
  sun avoidance and, 512
  treatment of, 430
Actinic prurigo, 434
Actinobacillus spp. *See Aggregatibacter* spp.
Actinomyces spp., 1363. *See also* Actinomycosis
Actinomyces-like organisms, 1365
Actinomycetoma, nocardial, 1360, 1361f, 1362, 1362t. *See also Nocardia* spp. infections
Actinomycosis, **1363**
  clinical features of
    abdominal, 1364
    CNS, 1130t, 1365
    disseminated, 1365
    hepatic-splenic, 1364, 1364f
    musculoskeletal and soft tissue, 1365
    oral-cervicofacial, 262t, 1363–1364, 1363f
    pelvic, 1364–1365, 1364f
    thoracic, 1364, 1364f
  diagnosis of, 1365, 1366f, S11
  epidemiology of, 1363
  etiology of, 1363
  pathogenesis of, 1363
  treatment of, 1365–1366, 1366t
Action potentials, cardiac, 1849–1850, 1910–1912, 1911f, 1913, 1918f

- Action spectrum, 429
- Activase. *See* Alteplase (rtPA)
- Activated charcoal, 302, 3702, 3708, 3709, 3727
- Activated partial thromboplastin time (aPTT)
  - causes of abnormal values, 475t
  - in coagulation disorders, 928, 928t
  - coagulation factors tested in, 474, 474f, 929f
  - LA-PTT, 475
  - for monitoring heparin therapy, 947
  - in preoperative testing, 474
  - prolonged, 475t
  - reagent composition of, 474–475
- Active surveillance
  - in prostate cancer, 518, 700–701
  - in testicular cancer, 706
- Active transport, 2366
- Activin, 2991
- Acupressure, 300, 2463
- Acupuncture, 132, 302, 3925t, 3927
- Acustimulation, 2463
- Acute brain failure. *See* Delirium
- Acute cellular rejection, 2030, 2287
- Acute chest syndrome, 775
- Acute confusional state. *See* Delirium
- Acute coronary syndromes
  - diagnosis of, 2113, 2114f
  - global considerations in, 2112–2113
  - myocardial infarction
    - non-ST-segment elevation. *See* Non-ST-segment elevation acute coronary syndrome (NSTE-ACS)
    - ST-segment elevation. *See* ST-segment elevation myocardial infarction (STEMI)
  - PCI for, 2129. *See also* Percutaneous coronary interventions (PCI)
  - pulmonary edema in, 2330
  - spectrum of, 2107f
  - unstable angina, 102, 103t, 2106. *See also* Non-ST-segment elevation acute coronary syndrome (NSTE-ACS)
  - venous thromboembolism and, 2157–2158, 2158f
  - in women, 3164
- Acute disseminated encephalomyelitis (ADEM), **3593,** A16
- Acute fatty liver of pregnancy, 3903
- Acute febrile illness, **989**
  - approach to the patient, 989
  - clinical presentations of, 990–996, 991t
  - diagnostic workup in, 989
  - history in, 989
  - physical examination in, 989
  - rashes associated with, **A1**
  - treatment of, 990, 991t
- Acute febrile neutrophilic dermatosis. *See* Sweet syndrome
- Acute flaccid myelitis, 1634–1655
- Acute flaccid paralysis, 1114–1115
- Acute generalized eruptive/exanthematous pustulosis (AGEP)
  - clinical features of, 143t, 386, 425–426, 425f, 427t, A1
  - differential diagnosis of, 147, 425–426
  - drugs associated with, 425–426, 427t
  - immune pathways of, 417t, 2816f
  - timeline of, 427, 428f
- Acute inflammatory demyelinating polyneuropathy (AIDP), 3617–3618, 3617t, 3618f, 3619t
- Acute intermittent porphyria (AIP), **3346**
  - clinical features of, 3341t, 3346, 3602–3603
  - diagnosis of, 3341t, 3346
  - genetic considerations in, 3603
  - global considerations in, 3342
  - pathophysiology of, 3341f, 3602
  - treatment of, 3347
  - unsafe drugs in, 3346
- Acute interstitial nephritis (AIN), **2437**
  - allergic, 2356, 2356f, 2378, 2437–2438, 2438f
  - autoimmune, 2356, 2356f
  - clinical features of, 338t, 2356, 2379t, 2437–2438, A4
  - in crystal deposition disorders, 2440
  - diagnosis of, 338t, 2379t
  - drug-related, 2356f, 2377, 2439
  - etiology of, 2356, 2356f, 2437t
  - granulomatous, 2439
  - hypokalemia in, 356
  - hypovolemia in, 346
  - idiopathic, 2439
  - IgG4-related, 2439, 2931
  - infection-associated, 2356, 2439–2440
  - in light chain cast nephropathy, 2440, 2440f, A4
  - in lymphomatous infiltration of the kidney, 2440
  - in obstructive tubulopathies, 2440
  - renal biopsy in, A4
  - in sarcoidosis, 2928
  - in Sjögren's disease, 2438–2439
  - in SLE, 2439. *See also* Lupus nephritis
  - treatment of, 2438f, 2438t
  - in tubulointerstitial nephritis with uveitis, 2439, 2439f
- Acute interstitial pneumonia (Hamman-Rich syndrome), 2265
- Acute kidney injury (AKI), **2372**
  - approach to the patient, 339, 339f, **2353,** 2353t
  - clinical features of, 338t, 353, 355, 2378–2380, 2379t
  - complications of, **2382**
  - diagnosis of
    - biomarkers, 2382
    - blood laboratory findings, 340t, 2378, 2379t, 2381
    - history and physical examination, 2378–2380
    - imaging, 2381
    - renal biopsy, 2381–2382, A4
    - renal failure indices, 2381
    - urine findings, 340t, 2379t, 2380, 2380f
  - epidemiology of, 2372–2373
  - etiology and pathophysiology of
    - after liver transplantation, 2723
    - community-acquired, 2373
    - in critically ill patient, 2298
    - envenomation, 3719
    - hypokalemia, 355
    - intrinsic, 2375
    - ischemia-associated, 2376
    - in leptospirosis, 1442–1443
    - in malaria, 1764–1765, 1773
    - nephrotoxin-associated, 2377–2378
    - postrenal, 2378
    - prerenal, 2373–2374
    - in scleroderma renal crisis, 2384, 2866, 2870–2871, 2875, A4
    - sepsis-associated, 2375–2376
    - in systemic disease, 2360, 2361f
    - in transplant recipient, 2409
  - global considerations in, 2373
  - incidence of, 2373
  - intrinsic
    - approach to the patient, 340–341
    - diagnosis of, 2353t, 2379t
    - etiology of, 2373f, 2379t
    - pathophysiology of, 2353t, 2360, 2375, 2376f
    - treatment of, 2383–2384
  - ischemia-associated, 340, 2376, 2376f, 2379t
  - nephrotoxin-associated, 340, 2377–2378, 2379t
  - postrenal
    - approach to the patient, 340
    - diagnosis of, 2353t
    - etiology of, 2373f
    - pathophysiology of, 340, 2353t, 2360, 2378, 2380f
    - treatment of, 2384
  - prerenal
    - approach to the patient, 340
    - clinical features of, 2379t
    - diagnosis of, 2353t, 2360, 2373–2375, 2379t
    - etiology of, 2373f
    - pathophysiology of, 2353t
    - treatment of, 2383
  - prevention of, 2383–2384
  - prognosis of, 2385
  - sepsis-associated, 2375–2376, 2379t
  - in sepsis/septic shock, 2322t, 2323
  - staging of, 2310, 2372t
  - treatment of, **2383,** 2383t
- Acute lung injury (ALI), 2301f, 2323
- Acute lymphoid leukemia (ALL), **844**
  - B-cell lineage, 845, 845t, 847t, 3824t, 3829
  - in congenital disorders, 844
  - cytogenetic and molecular analysis in, 845–846, 845t
  - diagnosis of, 844, 844t
  - epidemiology of, 844
  - etiology of, 844
  - genetic considerations in, 521t
  - hyperleukocytosis in, 591
  - immunological subtypes of, 845, 845t
  - lymphadenopathy in, 476
  - minimal residual disease in, 846, 846t
  - morphological subtypes of, 844
  - peripheral blood smear in, 844, 844t, A6
  - Ph-like, 845–846
  - prevention of, 784
  - prognosis of, 846
  - T-cell lineage, 845, 845t, 847t, 849
  - treatment of
    - in adolescents and young adults, 847t, 848
    - in adults, 847, 847t
    - algorithm for, 846f
    - in central nervous system disease, 847
    - gene therapy, 3824t, 3829
    - HCT, 847–848, 919, 919t
    - immunotherapeutic approaches, 848–849, 848t
    - maintenance therapy, 847
    - in older adults, 847t, 848
    - principles for, 847
    - targeted therapies, 848–849, 848t
- Acute megakaryoblastic leukemia, 879
- Acute mesenteric ischemia. *See* Intestinal ischemia
- Acute motor axonal neuropathy (AMAN), 3617, 3617t, 3618f, 3619t
- Acute motor sensory axonal neuropathy (AMSAN), 3617, 3617t, 3618f, 3619t
- Acute mountain sickness (AMS), 3755–3756, 3756t, 3757f
- Acute myeloid leukemia (AML), **825**. *See also* Acute promyelocytic leukemia (APL)
  - cancer stem cells in, 542
  - with *CEBPA* mutation, 827
  - classification of, 826–827, 826t
  - clinical features of, 828–829
  - diagnosis of, 828, 829f, 830t
  - etiology of, 825–826
  - genetic considerations in, 520t, 825t, 826–827, 826t, 827t, 3940
  - hyperleukocytosis in, 591
  - immunophenotypic findings in, 827
  - incidence of, 825
  - infections in, 577, 832–833
  - MDS and, 815

- Acute myeloid leukemia (AML) (*Cont.*):
  - monocytopenia in, 462
  - with *NPM1* mutation, 826, 827
  - peripheral blood smear in, 828, 829f, A6
  - prognosis of, 827–828, 827t
  - as second malignancy in cancer survivors, 709, 755, 826
  - skin manifestations of, 409
  - treatment of, **829**
    - algorithm for, 831f
    - HCT, 831–832, 919, 919t
    - induction chemotherapy, 830
    - investigational agents, 833t
    - postremission therapy, 830–832
    - in relapsed or refractory disease, 833, 833t
    - supportive care, 832–833
- Acute myelomonocytic leukemia, 261, 263, 263t
- Acute myocardial infarction (AMI). *See* Myocardial infarction (MI)
- Acute necrotic collection, pancreatic, 2746t
- Acute necrotizing myopathy, 2352
- Acute necrotizing ulcerative gingivitis, 261, 262t
- Acute nephritic syndromes, 2412, 2415t, **2416**
  - ANCA small vessel vasculitis, 2415t, 2419–2420. *See also* Eosinophilic granulomatosis with polyangiitis; Granulomatosis with polyangiitis; Microscopic polyangiitis
  - antiglomerular basement membrane disease, 275, 2412, 2415t, 2418–2419, A4
  - cryoglobulinemia. *See* Cryoglobulinemia
  - endocarditis-associated glomerulonephritis, 2415t, 2417, 2428
  - Goodpasture's syndrome. *See* Goodpasture's syndrome
  - IgA nephropathy. *See* IgA nephropathy
  - IgA vasculitis. *See* IgA vasculitis
  - lupus nephritis. *See* Lupus nephritis
  - membranoproliferative glomerulonephritis. *See* Membranoproliferative glomerulonephritis
  - poststreptococcal glomerulonephritis. *See* Poststreptococcal glomerulonephritis
  - renal biopsy in, A4
- Acute pancreatitis, **2743**
  - in alcohol use disorder, 2743, 3677
  - approach to the patient, 2738, 2745
  - clinical features of, 2746
    - abdominal pain, 112t, 2745
    - AKI, 2376
    - fat necrosis, 406
    - hypocalcemia, 3286
    - hypovolemia, 346
    - nausea and vomiting, 297
  - complications of, 2749–2750, 2749t
  - diagnosis of, **2745**
    - CT, 2746t, 2747f
    - endoscopy, 2498–2499
    - ERCP, 2748f
    - imaging methods, 2738, 2738f, 2739t, 2740–2742, 2746–2747
    - laboratory tests, 113, 2739–2742, 2739t, 2745
  - differential diagnosis of, 2745–2746
  - drug-related, 2745
  - epidemiology of, 2738
  - etiology of, 2743–2744, 2744t
  - gallstone, 2489
  - genetic considerations in, 2744–2745
  - in HIV infection, 1601, 2750
  - incidence of, 2743
  - interstitial, 2744, 2746, 2746t
  - morphologic features of, 2746t
  - necrotizing. *See* Necrotizing pancreatitis
  - pancreatic enzyme activation in, 2744–2745
  - pathogenesis of, 2743–2744
  - phases of, 2746, 2747f
  - physical examination in, 2745
  - recurrent, 2750
  - severity assessment in, 2746, 2748, 2748t
  - treatment of, 2747–2749
- Acute pharyngeal cervicobrachial neuropathy, 3619t
- Acute posthemorrhagic anemia, 807–808
- Acute promyelocytic leukemia (APL)
  - chromosome abnormalities in, 539, 826, 3783t
  - clinical features of, 828
  - peripheral blood smear in, A6
  - treatment of, 539, 573, 833–834, 2984
- Acute quadriplegic neuropathy, 3648, 3648t
- Acute radiation syndrome, S6
- Acute renal failure. *See* Acute kidney injury (AKI)
- Acute respiratory disease (ARD), 1535
- Acute respiratory distress syndrome (ARDS), **2299**
  - in blastomycosis, 1701
  - clinical course of, 2299–2301, 2300f
  - in COVID-19 disease, 1541, 2294, 2299
  - diagnosis of, 2299f, 2300t, 2301f, A12
  - disorders associated with, 2299, 2299t
  - etiology of, 2299, 2299t
  - functional recovery after, 2303–2304
  - HSV infection and, 1500
  - incidence of, 2299
  - pathophysiology of, 2299–2301, 2300f, 2301f
  - pressure-volume curves in, 2304, 2304–2305f
  - prognosis of, 2303–2304
  - respiratory failure in, 2294–2295, 2295f
  - in sepsis/septic shock, 2323
  - treatment of, 2302–2303, 2303f, 2303t. *See also* Mechanical ventilation
- Acute rheumatic fever. *See* Rheumatic fever
- Acute tubular necrosis (ATN). *See* Acute kidney injury (AKI)
- AC (assist control) ventilation, 2305–2306, 2305t. *See also* Mechanical ventilation
- Acyclovir
  - actions of, 1115, 1485
  - adverse effects of
    - gastrointestinal, 1115
    - hematologic, 1115
    - musculoskeletal, 2940t
    - neurologic, 1115, 1485
    - renal, 1485, 1502, 2377
  - for Bell's palsy, 3555
  - for CMV infection prophylaxis, 1515
  - global considerations in, 1503–1504
  - for herpes zoster, 1054t, 1485, 1486t, 1507, 3566
  - for HSV infections
    - actions of, 1502
    - encephalitis, 1115, 1485, 1486t, 1500, 1502, 1503t
    - esophagitis, 1503t, 2513
    - genital, 1054t, 1109t, 1485, 1486t, 1502, 1503t
    - mucocutaneous, 1485, 1502, 1503t, 1537
    - in neonates, 1500, 1503t
    - in transplant recipient, 2410
  - for HSV prophylaxis
    - in cancer patient, 584–585
    - in HIV infection, 1502, 1595t
    - in transplant recipient, 918, 918t, 1156, 1164t
  - for oral hairy leukoplakia, 1512
  - pharmacology of, 1485
  - resistance to, 1115, 1412–1413, 1485, 1503t
  - for varicella treatment, 1485, 1486t, 1507
  - for viral encephalitis, 1115
  - for viral meningitis, 1122, 1126
  - for viral myelitis, 3566
  - for VZV prophylaxis, 584–585, 1593t
- AD. *See* Alzheimer's disease (AD)
- Adagrasib, 569t, 572
- Adalimumab
  - actions of, 2791t
  - adverse effects of, 387t, 462, 2566, 2568t, 2850t
  - for autoimmune and inflammatory diseases, 2784, 2791t
  - for axial spondyloarthritis, 2887t
  - for Behçet syndrome, 2909
  - for chronic uveitis, 225
  - for IBD, 2566, 2568t
  - monitoring during treatment with, 2568t, 2850t
  - for psoriasis/psoriatic arthritis, 387t, 2890
  - for rheumatoid arthritis, 2850t, 2851
  - for sarcoidosis, 2929t, 2930
- Adamantinoma, 729
- *ADAMTS2* gene mutations, 3328t
- ADAMTS13, 925, 2428, 2445
- Adaptive immune system, 2756, **2774**. *See also* Immune response/immune system
  - in autoimmune disease, 2824–2826
  - cells triggering, 2763t
  - definition of, 2755
  - immunoglobulins in. *See* Immunoglobulin(s)
  - inhibition of, 974
  - innate immune system and, 2756
  - intercellular interactions of, 2764f
  - primary immunodeficiencies of
    - B lymphocyte deficiencies. *See* B cell(s), deficiencies of
    - T lymphocyte deficiencies. *See* T cell(s), deficiencies of
  - response to microbes, 2779–2780, 2781
  - response to viruses, 1482, 1482f
  - T cells in. *See* T cell(s)
- Adaptive myelination, 3377
- Adaptive servoventilation, 2282
- ADCC (antibody-dependent cell-mediated cytotoxicity), 1587f, 1588, 2755, 2768
- *ADCY5* gene mutations, 3514
- Addison's disease. *See* Adrenal insufficiency, primary
- Additivity, in clinical trials, 3930t, 3932f, 3933–3934
- Adefovir
  - actions of, 1491
  - adverse effects of, 1491, 1601, 2683
  - for chronic HBV infection, 1490t, 1491, 2676t, 2677–2678, 2680f, 2682, 2682t
  - in pregnancy, 2683
  - resistance to, 2682t
- A-delta (Aδ) fibers, 93, 93f, 94
- ADEM (acute disseminated encephalomyelitis), **3593,** A16
- Adenine, 3778
- Adenine nucleotide translator (ANT), 2790t
- Adeno-associated virus (AAV), 1484, 1525, 3524, 3823t, 3825t, 3827
- Adenocarcinoma
  - bladder. *See* Bladder cancer
  - cervix. *See* Cervical cancer
  - colon. *See* Colorectal cancer
  - endometrial. *See* Endometrial cancer
  - esophagus, 641–642, 643t. *See also* Esophageal cancer
  - gastric. *See* Gastric adenocarcinoma
  - lung, 615, 616f, 617f. *See also* Lung cancer
  - prostate. *See* Prostate cancer
  - small bowel, 651
  - of unknown primary, 734–735, 734f
- Adenocarcinoma in situ, 616, 622
- *Adenocephalus pacificus,* 1838
- Adenoid cystic carcinoma, 610, 614
- Adenoidectomy, 255
- Adenoma
  - adrenal aldosterone-producing. *See* Hyperaldosteronism; Mineralocorticoid excess
  - hepatic, 670

  parathyroid, 3275. *See also* Hyperparathyroidism
  pituitary. *See* Pituitary tumors (adenomas)
  small bowel, 651
  thyroid, 3043
Adenoma sebaceum, 397, 404
Adenomyomatosis, 2734
Adenomyosis, 3136
Adenosine
  actions of, 1916
  adverse effects of, 1943
  for atrial tachycardia, 1938
  contraindications to, 1941
  for supraventricular tachycardia, 486, 1943
  as vasoconstrictor, 2364
  for ventricular tachycardia, 1964
Adenosine deaminase, 329
Adenosine deaminase deficiency, 2798–2799, 2901, 2937, 3824t, S9
Adenosine kinase deficiency, 3368t
Adenosine monophosphate–activated protein kinase (AMPK), 543, 3871
Adenosine receptor(s), 938
Adenosine receptor antagonists, 1992
S-Adenosylhomocysteine hydrolase deficiency, 3368t
Adenoviridae
  enteric, 1628f, 1628t, 1632
  for gene therapy delivery, 3823t
  laboratory identification of, S11
  replication strategies of, 1481
  serotypes of, 1535
  structure of, 1478f, 1480t
  as vaccine vector, 1484
Adenovirus infections
  clinical features of, 1535
    conjunctivitis, 224
    diarrhea, 305, 1632
    otitis media, 254
    rash, 402
    upper respiratory, 253, 1535
  in HCT recipient, 1155t, 1157
  prevention of, 1538
Adenovirus vaccine, 1538
Adenylate kinase (AK) deficiency, 796t
Adequate intake, 2599
ADH (antidiuretic hormone). *See* Arginine vasopressin (AVP)
Adhesins, bacterial, 965, 966t
Adhesive capsulitis, 2974–2975
Adhesives, dermatitis and, 423, 423f
ADHH (autosomal dominant hypocalcemic hypercalciuria), 3287
Adiaspiromycosis, 1689t, 1727
Adie's syndrome, 221
Adipocytes, 3182, 3202
Adiponectin, 3255
Adipose tissue, 2621, 3181
Adnexal mass/tumor, 404, 710, 3136
ADNFLE (autosomal dominant nocturnal frontal lobe epilepsy), 3412t
Adolescence
  cannabis use in, 3687
  iron requirements in, 765t
  seizures in, 3413–3414, 3413t
  smoking in, 511, 3685
Ado-trastuzumab emtansine, 535t, 566, 566t
ADPEAF (autosomal dominant partial epilepsy with auditory features), 3412t
ADPKD (autosomal dominant polycystic kidney disease). *See* Polycystic kidney disease, autosomal dominant
ADP receptor antagonists, 943–944. *See also* Clopidogrel; Prasugrel; Ticagrelor
ADP-ribosylation, 3936–3937
Adrenalectomy
  for Cushing's disease, 3012
  for Cushing's syndrome, 619, 3059
  for ectopic ACTH syndrome, 739
  for mineralocorticoid excess, 3063
  for pheochromocytoma, 3075
Adrenal gland
  anatomy and development of, 3050
  disorders of
    adrenal insufficiency. *See* Adrenal insufficiency
    adrenocortical carcinoma. *See* Adrenocortical carcinoma
    congenital adrenal hyperplasia. *See* Congenital adrenal hyperplasia (CAH)
    Cushing's syndrome. *See* Cushing's syndrome
    hyperaldosteronism. *See* Hyperaldosteronism
    incidentally discovered mass, 3063–3065, 3064f
    in MEN 1, 3083–3084
    mineralocorticoid excess. *See* Mineralocorticoid excess
    pheochromocytoma. *See* Pheochromocytoma
    unilateral mass, 3063t
  steroidogenesis in, 3050–3051, 3051f
    activation/inactivation of, 3052–3054, 3053f, 3054f
    regulatory control of, 3050–3052, 3052f, 3053f, 3055f
Adrenal hyperplasia
  bilateral, 3060t
  congenital. *See* Congenital adrenal hyperplasia (CAH)
  hypertension in, 2138t, 2139
  hypokalemia in, 355
  macronodular, 737t
  primary, 355
Adrenal hypoplasia congenita, 3067t, 3113
Adrenal insufficiency, **3066**
  acute, 3068
  in cancer, 593
  chronic, 3067
  diagnosis of, 3954
  epidemiology of, 3066
  etiology of, 3066–3067, 3067t
  glucocorticoid-induced, 3070–3071
  primary (Addison's disease)
    APS-1 and, 3089, 3090t
    APS-2 and, 3090t, 3091, 3091t
    autoantibodies in, 2790t, 3091
    clinical features of, 3067, 3068t, 3069f
      diarrhea, 308
      fatigue, 166
      hyperkalemia, 359
      hyponatremia, 347
      hypothermia, 3768, 3771
      oral, 264t, A15
      skin, 399, A15
    diagnosis of, 3068
    epidemiology of, 3066
    etiology of, 3066, 3067t
    genetic considerations in, 3091
    hypoglycemia in, 3234
    inflammasome mutations in, 2761t
    thymic tumors in, 3092
  secondary, 3067, 3068t
  treatment of, 3069–3070, 3070f
Adrenal vein sampling, 2139, 3061
Adrenarche, 3102, 3130
Adrenocortical carcinoma
  diagnosis of, 3065, 3066f
  differential diagnosis of, 3063–3065
  genetic considerations in, 3060t, 3065
  incidence of, 3063t, 3065
  in MEN 1, 3083–3084
  staging of, 3065, 3065t
  treatment of, 3065–3066
Adrenocortical tumors, 3083–3084
Adrenocorticotropic hormone (ACTH), **2990**
  actions of, 2982, 2991
  in adrenal steroidogenesis control, 3050–3052, 3052f, 3053f
  autoantibodies to, 2790t
  deficiency of, 2790t, 2997, 3185
  ectopic production of, 736, **738**
    clinical features of, 739, A15, S1
    diagnosis of, 739, 3011t
    etiology of, 538–739, 737t, 3054, 3056t, S1
    in lung cancer, 619
    in NETs, 679
    skin manifestations of, 399
    treatment of, 739, 3059, S1
  excess of. *See* Cushing's syndrome
  expression and regulation of, 2988t, 3949
  laboratory evaluation of, 2995t, 3011, 3068
  rhythms of, 2987
  secretion of, 2991, 3050–3051
  synthesis of, 14–15, 3050
Adrenocorticotropic hormone (ACTH) receptor, 2985t
Adrenoleukodystrophy
  cell therapy for, 3824t, 3826
  clinical features of, 3067t, 3485, 3602
  diagnosis of, 3485, 3602, A16
  gene therapy for, 762
  genetic considerations in, 3066, 3067t, 3523t, 3602
  pathogenesis of, 3397
Adrenomyeloneuropathy, 3569, 3602
Adria virus, 1662t
ADT. *See* Androgen deprivation therapy (ADT)
Aducanumab, 199–200, 3480–3481, 3481f
Adult bile ductopenia, 325
Adult growth hormone deficiency, **2996,** 2996t, 2997f
Adult polyglucosan body disease (type IV GSD), 3360t, **3362**
Adult T-cell leukemia/lymphoma (ATLL)
  acute, 1554, 1555f
  chronic, 1554, 1555f
  clinical features of, 867
  diagnosis of, 867, 868f, A6
  genetic considerations in, 521t
  HTLV-1 and, 867, 1553–1554
  lymphomatous, 1554
  pathology of, 867, 868f
  skin manifestations of, 405
  smoldering, 1554, 1555f
  Strongyloides infections and, 1554
  treatment of, 867, 1555–1556, 1555f
  tumor markers in, 504t
Advance care directives, 78–79, 88, 505
Advance care planning, 78–79, 78t, 505
Advanced glycosylation end products, 3223
Advanced practice providers, 3898
Advanced sleep-wake phase disorder, 218, 3944t, 3948
Adventitia, 1845, 1845f
Adverse drug events, 42–43
Adverse drug reactions, **489**. *See also* Cutaneous drug reactions; Drug-induced illness; *specific drugs*
  classification, 488
  diagnosis of, 490
  identification of, 483
  idiosyncratic hepatotoxicity, 2665–2668, 2667t
  immunologic mechanisms of, 417–419, 417t, 418f, 489–490
  incidence of, 489
  treatment of, 490. *See also* Drug desensitization

Adverse events
  ethical issues in, 73
  factors that increase likelihood of, 42
  frequency of, 42–43
  measurement of, 43
  national picture of, 43
  prevention of, 4, 43
  safety theory and systems theory and, 42, 42f
  vaccine-related, 19, 1008
Advisory Committee on Immunization Practices (ACIP), 1000, 1003
AE1 (red cell anion exchanger), 358
AEBP1 gene mutations, 3329t, 3330
AED (automatic external defibrillation), 2336
Aedes spp. mosquitoes
  in chikungunya virus infection, 1669
  climate change and habitat of, 3748, 3749f, 3750
  as dengue virus vector, 1668, 1678
  in Eastern equine encephalitis, 1673
  in Jamestown Canyon virus infection, 1673
  in La Crosse virus infection, 1672–1673
  in lymphatic filariasis, 1820t
  in Rift Valley fever, 1677
  in Ross River disease, 1669
  in yellow fever, 1679
  as Zika virus vector, 1668
AEDs. *See* Antiepileptic drugs (AEDs)
Aerobic glycolysis, 543, 544f
Aerobic organisms, 1371
Aeromonas spp. infections
  in animal bite-wound, 1143
  cellulitis, 1052
  clinical features of, 306t, 992, 1268
  diagnosis of, 311
  epidemiology of, 305, 1268
  in leech infestations, 3733
  treatment of, 1052, 1268, 1268t
Aerophagia, 326
Aerophobia, 1115, 1652
Aerotolerant organisms, 1371
AF. *See* Atrial fibrillation (AF)
Afamelanotide, 435, 3350
AF amyloidosis, 900. *See also* Amyloidosis
Afatinib
  actions and targets of, 568t, 572
  adverse effects of, 568t, 755
  interactions and issues, 568t
  for lung cancer, 627, 627t
Affected sib and relative pair analysis, 3797t
Afferent loop syndromes, 2533
Affordable Care Act (ACA), 68, 1009
Afib amyloidosis, 2027t
Aficamten, 2007
Aflatoxin, 530
Aflatoxin B$_1$, 660
Aflibercept, 231, 2792t
AFP. *See* A fetoprotein (AFP)
African eye worm *(Loa loa)*/loiasis. *See Loa loa* (African eye worm)/loiasis
Africanized honeybees, 3735
African python, 3733
African tick-bite fever, 148f, 1455t, 1458
African trypanosomosis, human. *See* Human African trypanosomiasis
African tumbu fly, 3732
AFRI method, hepatic fibrosis assessment, 2632t
Afterdepolarizations, 1913, 1913f
Afterload, 1851f, 1852f, 1853f
Agalsidase, 2425, 3602
Agammaglobulinemia, 2550, 2550t, 2801f, 2802
Agender, 3179t
Agent Orange, S8
AGEP. *See* Acute generalized eruptive/exanthematous pustulosis (AGEP)
Agglutination, 439f, 443
Aggrecan, 3322t, 3323
Aggregatibacter spp., 1265–1266, 1267t. *See also* Infective endocarditis, HACEK group
Aggrenox, 944
Aggressive NK cell leukemia, 874
Aging, **3869**. *See also* Older adults
  autoimmune disease and, 2823
  biological hallmarks of, 3869, 3872f
    altered intracellular communication, 3872
    cellular senescence, 3871
    chronic inflammation, 3872
    deregulated nutrient sensing, 3871, 3873f
    disabled macroautophagy, 3871
    dysbiosis, 3872
    epigenetic alterations, 3870
    genomic instability, 3869–3870
    loss of proteostasis, 3871, 3872f, 3992–3993
    mitochondrial dysfunction, 3871
    stem cell exhaustion, 3871–3872
    telomere attrition, 3870
  chronic diseases and, 3870, 3870f
  circadian disruption and, 3953–3954
  definitions of, 3869, 3870f
  epigenetics of, 3940–3941
  immune dysregulation in, 2789, 2793
  implications for health care systems, 3895–3896, 3896t
  loss of health with, 8, 8f
  microbiome and, 3833
  mtDNA mutations and, 3815, 3815f
  olfactory function and, 240, 240f, 244
  oral health and, 267
  skin effects of, 431
  sleep patterns and, 210, 210f
  species differences in, 3869, 3871f
  strategies to delay, 3873–3875, 3874f
  telomere shortening in, 3818, 3819f
  theories of, 3869
  weight regulation and, 315
Agitated saline study, 2188
Agitation, 91t, 199
Agnogenic myeloid metaplasia. *See* Primary myelofibrosis
Agnosias, 207, 3432
Agnostic molecular testing, 3997, 3997t
Agonal gasps, 191
Agoraphobia, 3657
Agouti-related peptide (AgRP), 3182, 3183f
AGR2 gene mutations, 2554t
Agrammatic primary aphasia, 204
Agrammatism, 202
Agranulocytosis, 263
Agraphesthesia, 174
Agraphia, 201
agr gene, 1197
Agrin, 3624, 3625f
AgRP (agouti-related peptide), 3182, 3183f
AH amyloidosis. *See* Amyloidosis
AH interval, 1927, 1929f
AHO. *See* Albright's hereditary osteodystrophy (AHO)
aHUS (atypical hemolytic uremic syndrome), 800–801, 925, 2412
Aicardi-Goutières syndrome, 2937
AID, 2786f
AIDP (acute inflammatory demyelinating polyneuropathy), 3617–3618, 3617t, 3618f, 3619t
AIDS. *See also* HIV infection
  deaths from, 1566f, 1569f
  definition of, 1557, 1557t
  epidemiology of, 1565–1568
AIDS dementia complex (HIV-associated dementia), 1606, 1607f, 1607t
AIFEC (autoinflammatory infantile fever with enterocolitis), 2761t
Aigai virus, 1661t
AIM2 gene, 2761–2762t
AIN. *See* Acute interstitial nephritis (AIN)
Ainhum, 3736
AION (anterior ischemic optic neuropathy), 226–227, 227f
AIP. *See* Acute intermittent porphyria (AIP)
AIP gene mutations, 3004, 3004t, 3084
Airborne precautions, 1148, 1538–1539
Air-conditioner lung, 2228t
Air-conduction threshold, 248
AIRE gene mutations
  in APS-1, 1708, 2786f, 2804, 3066, 3089, 3287
  global considerations in, 3093
Air flow
  in COPD, 2254, 2255t
  in lung, 2201–2202, 2203f, 2207
Air hunger, 267, 268f
Air plethysmography, 2618t
Air pollution
  asthma and, 2220
  cardiovascular disease and, 3739–3740, 3741f
  health effects of, 2240–2241
  hypertension and, 2136
  interventions to reduce, 3740, 3742f
  pregnancy outcomes and, 3740
  respiratory illness and, 272, 274, 2253, 3739, 3740f
Air travel, venous thrombosis risk in, 940
Airway
  in asthma, 2216–2218, 2217f
  hyperresponsive, 2216–2217, 2223, 2252
  inflammation of, 2218
  narrowing, triggers of, 2217t, 2221
  in relapsing polychondritis, 2919, 2920f
  rewarming of, 3770
Airway obstruction
  in botulism, 1235
  in cancer, 592, 592f
  dyspnea in, 269t
Airway resistance (R$_{aw}$), 270, 2206f, 2207
Airway stenting, 2288–2289
AIS. *See* Androgen insensitivity syndrome (AIS)
AIT (amiodarone-induced thyroxicosis), 3039–3040
AIVR (accelerated idioventricular rhythm), 1955f, 1961–1962, 1963f, 2124, A8
AK2 gene mutations, 458
Akathisia, 3516, 3666, 3674
AK (adenylate kinase) deficiency, 796t
Akhmeta virus, 1518t, 1519
AKI. *See* Acute kidney injury (AKI)
Akinetic mutism, 187, 2348
AKR1C2 gene mutations, 3099t, 3105
AKT, 3939
AKT gene mutations
  cancers associated with, 520t
  in Cowden's syndrome, 3080t, 3088
  in lung cancer, 617f
  in melanoma, 601–602
ALA (5-aminolevulinic acid), 430
ALA-dehydratase, 3342–3343, 3342t, 3343
ALA-dehydratase deficient porphyria, 3341f, 3341t, 3342, 3342t, 3345–3346
ALAD gene, 3342t
AL amyloidosis, **896**
  cardiac, 2022, 2027
  clinical features of, 896t, 897, 897f, 3550, 3603
  diagnosis of, 404, 896f, 897–898, 898f, 2425
  differential diagnosis of, 2415t
  etiology and incidence of, 896–897
  multiple myeloma and, 2022
  pathogenesis of, 2425
  pathology of, 897
  staging and risk stratification, 898
  treatment of, 898–899, 2425

- Alanine aminotransferase (ALT)
  - in acute viral hepatitis, 2645f, 2648f, 2656
  - in alcohol-associated liver disease, 2698
  - in autoimmune hepatitis, 2695
  - in chronic HBV infection, 2675
  - in jaundice evaluation, 324
  - in liver function evaluation, 2635, 2637t
- ALARA principle, S6
- Alarm symptoms. *See* "Red flags"
- *ALAS* gene mutations, 3342, 3342t, 3346
- Alaskapox. *See* Borealpox
- ALA-synthase, 3342, 3342t, 3343, 3349
- Albendazole
  - actions of, 1740
  - adverse effects of, 1740, 1742t
  - for ascariasis, 1815
  - for capillariasis, 1819
  - for cutaneous larva migrans, 1812t
  - for cysticercosis, 1834
  - drug interactions of, 1740, 1742t
  - for echinococcosis, 1837
  - for enterobiasis, 1818
  - for gnathostomiasis, 1812t
  - for hookworm infections, 1816
  - indications for, 1742t, 1815t
  - for loiasis, 1824
  - for lymphatic filariasis, 1821
  - for microsporidiosis, 1600
  - for neurocysticercosis, 1138
  - pharmacology of, 1740
  - in pregnancy and lactation, 1742t
  - for strongyloidiasis, 1817
  - for trichinellosis, 1812t
  - for trichuriasis, 1818
  - for visceral larva migrans, 1812t
- Albers-Schonberg disease, 3315
- Albinism, 396, 466, 606, 3368t
- Albright's hereditary osteodystrophy (AHO)
  - clinical features of, 3184, 3184t, 3288
  - genetic considerations in, 3028t, 3288–3289, 3783
  - pseudohypoparathyroidism in, 3288, 3288t
- Albumin
  - for hepatorenal syndrome, 2383
  - infusion, during paracentesis, 329
  - for peritonitis, 1070
  - serum
    - in edema, 282
    - in hypercalcemia, 364
    - in hypocalcemia, 365
    - in jaundice evaluation, 324
    - in liver function evaluation, 2635–2636
    - in nutrition assessment, 2621
  - urinary
    - in CKD, 342, 2385, 2387
    - in diabetic nephropathy, 2357, 2424, 3224f, 3226
    - in glomerular disease, 2414t
    - in HIV infection, 1601
    - measurement of, 2387
- Albuterol, 361, 2224, 3710t
- *Alcaligenes xylosoxidans. See Achromobacter xylosoxidans*
- Alcohol, **3675**. *See also* Alcohol use; Alcohol use disorder
  - absorption of, 2708, 3675
  - acetaminophen toxicity and, 489
  - actions of, 3656t
  - behavioral effects of, 3676, 3676t
  - drug interactions of, 1172t
  - metabolism of, 3675, 3675f
  - pharmacology of, 3675–3676
  - tolerance of, 3676
  - withdrawal from. *See* Alcohol withdrawal syndrome
- Alcohol-associated liver disease, **2697**
  - acetaminophen hepatotoxicity and, A13
  - cirrhosis, 2698, 2698t, 2708–2710, 2709f, 2726
  - clinical features of, 2698t
  - diagnosis of, 2637t, 2698–2699, 3336t
  - differential diagnosis of, 2699
  - epidemiology of, 2697
  - hepatitis, 324, 2660, 2698, 2698t
  - pathogenesis of, 2698, 3677
  - prognosis of, 2698, 2700
  - risk factors for, 2629, 2629t, 2698, 2698t
  - treatment of, 2633, 2699–2700, 2699f, 2719, 2726
- Alcohol dehydrogenase (ADH), 2708
- Alcoholic ketoacidosis, 348, 370
- Alcohol septal ablation, 2007, A11, A11v
- Alcohol use
  - breast cancer and, 3677
  - cardiovascular effects of, 3677
  - colorectal cancer and, 655
  - dyslipidemia and, 3246
  - esophageal cancer and, 642, 3677
  - folate deficiency and, 788
  - genitourinary effects of, 3677
  - hazardous levels of, 2702
  - HCC and, 660, 660f
  - head and neck cancer and, 610, 3677
  - hematopoietic effects of, 3677
  - hypertension and, 2137t, 2142
  - liver disease and. *See* Alcohol-associated liver disease
  - nervous system effects of, 3676
  - pancreatitis and, 2750
  - porphyria cutanea tarda and, 3348
  - in pregnancy, 3677
  - prevalence of, 2708
  - sexual dysfunction and, 3677
  - sleep disturbances and, 215, 3676
  - testicular dysfunction and, 3115
- Alcohol use disorder (alcoholism), **3675**
  - acetaminophen toxicity in, 489, 2669
  - anemia in, 450
  - autonomic dysfunction in, 3550
  - cancer risk and, 512t, 3677
  - cardiovascular disease in, 2015–2016, 3677
  - cognitive dysfunction in, 3483, 3676
  - definition of, 3677, 3678t
  - dementia in, 3483
  - dental disease and, 261
  - diagnosis of, 3678–3679, 3678t
  - epidemiology of, 3677–3678, 3681
  - erectile dysfunction in, 3157
  - gastrointestinal effects of, 318, 3677
  - gaze disorders in, 235
  - genetic considerations in, 3678
  - global considerations in, 3681
  - hepatitis in, 324
  - hormonal disturbances in, 3677
  - hypoglycemia in, 3234
  - hypomagnesemia in, 3266
  - hyponatremia in, 348–349
  - hypothermia and, 3769
  - involuntary weight loss in, 316
  - lifetime risk for, 3675
  - liver disease in. *See* Alcohol-associated liver disease
  - in military veterans, S8
  - myopathy in, 3648t, 3677
  - natural history of, 3678
  - nervous system effects of, 3676
  - neuropathy in, 3676
  - olfactory/taste dysfunction in, 242
  - palpitations in, 291
  - pancreatitis in, 2743, 3677
  - prevalence of, 2708
  - psychiatric comorbidity with, 3676–3677
  - rehabilitation for, 3680–3681
  - screening for, 13, 39t, 2629, 2629t
  - testicular defects in, 3115
  - thiamine deficiency in, 2606
  - treatment of, 3679–3680
  - in women, 3168
- Alcohol withdrawal syndrome
  - clinical features of, 3676, 3679–3680
  - delirium in, 182
  - hypokalemia in, 354, S1
  - treatment of, 3679–3680
- Aldehyde dehydrogenase (ALDH), 2708, 3678
- ALDEN (Algorithm of Drug Causality for Epidermal Necrolysis), 427
- Aldesleukin. *See* Interleukin-2 (IL-2) therapy
- *ALDH7A1* gene mutations, 3420
- Aldolase deficiency, 796t
- Aldosterone
  - actions of, 346, 355, 2308, 2367f
  - excess of. *See* Glucocorticoid-remediable aldosteronism (GRA); Hyperaldosteronism
  - in potassium regulation, 355
  - in sodium regulation, 2372
  - synthesis of, 3051f, 3054, 3055f
- Aldosterone antagonists, 1988, 1995t, 2338
- Aldosterone escape, 2310
- Aldosterone-renin-ratio (ARR), 3061, 3062t
- Aldosterone synthase, 3784
- *ALDP* gene mutations, 3523t
- ALECT2 amyloidosis, 2027t
- Alectinib
  - actions of, 533t, 568t
  - adverse effects of, 568t, 572, 753t
  - gene mutations and response to, 3982
  - indications for, 533t
  - for lung cancer, 625, 628, 628t
- Alemtuzumab
  - actions and targets of, 534t, 558, 559f, 2791t, 3577t
  - adverse effects of, 594, 597, 3033, 3577t
  - for autoimmune hemolytic anemia, 802
  - for hemophagocytic lymphohistiocytosis, 466
  - for hypereosinophilic syndrome, 880
  - for immunosuppression, 2408
  - indications for, 2791t
  - for MDS, 818
  - for MS, 3586t
  - for T-cell prolymphocytic leukemia, 874
- Alendronate
  - actions of, 3304
  - adverse effects of, 3304
  - for fracture prevention, 3305f
  - for hypercalcemia, 3285
  - for osteoporosis management/prevention, 3145, 3304, 3305f, 3311
  - for Paget's disease of bone, 3303, 3314t
- Alenquer virus, 1662t
- Alexia, 201, 201t, 203
- Alfapump system, 329
- Alfentanil, 1742t
- Algae, 3752t
- Alginic acid, 302
- Alglucosidase alfa, 3364
- Algorithm of Drug Causality for Epidermal Necrolysis (ALDEN), 427
- ALI (acute lung injury), 2301f. *See also* Acute respiratory distress syndrome (ARDS)
- Alien hand (limb) syndrome, 203, 3488, 3497
- Alirocumab, 2110, 3244t
- Aliskiren, 1994
- Alkaline phosphatase (ALP)
  - in acute viral hepatitis, 2656
  - in hypophosphatasia, 3316
  - in jaundice evaluation, 324
  - in liver abscess, 1074

Alkaline phosphatase (ALP) (*Cont.*):
  in liver function evaluation, 2634f, 2635, 2637t
  in Paget's disease of bone, 3312, 3313
  tissue nonspecific, 3316
Alkali therapy, 373. *See also* Bicarbonate therapy
Alkaptonuria, 3368t, **3371**
ALK gene mutations
  in anaplastic large-cell lymphoma, 867
  in lung cancer
    as driver mutation, 617, 617f, 617t, 628f
    as drug target, 493, 527, 533t, 572, 628, 628t, E72t
Alkhurma hemorrhagic fever virus, 1659t
ALK inhibitors, 628, 628t, 753t. *See also specific drugs*
Alkylating agents, 562t. *See also specific drugs*
  actions of, 563
  adverse effects of, 562t, 563, 2017
    anemia, 810t
    cardiotoxicity, 754
    in essential thrombocytosis, 824
    hair loss, 421, 576
    infertility, 756, 871, 3115
    late, 816, 824, 826
    long-term, 753t, 755
    myelosuppression, 574
    pulmonary, 754
    second malignancy, 755, 826
    seizures, 3414t
  carcinogenicity of, 512t, 824, 826
  for marginal zone lymphoma, 865
  for multiple myeloma, 890t
  for NETs, 685
ALL. *See* Acute lymphoid leukemia (ALL)
Allele(s), 3787
  association of, 3794–3795
  exclusion of, 2777, 3791
  frequency of, 3794–3795, 3795f
  heterogeneity of, 3786f, 3787, 3788t
  segregation of, 3789, 3790f
  types of, 3787
Allele-sharing methods, 3797t
Allen's test, 1862
Allergic angiitis and granulomatosis. *See* Eosinophilic granulomatosis with polyangiitis
Allergic bronchopulmonary aspergillosis (ABPA)
  bronchiectasis in, 2242, 2242t, 2244
  clinical features of, 1713, 1715
  in cystic fibrosis, 2248
  diagnosis of, 2221, 2233
  etiology of, 2233
  genetic considerations in, 1713
  prognosis of, 1717
  treatment of, 1717t, 2233–2234
Allergic contact dermatitis. *See* Contact dermatitis, allergic
Allergic fungal sinusitis, 257
Allergic interstitial nephritis, 2380, 2437–2438, 2438f. *See also* Acute interstitial nephritis (AIN)
Allergic reactions. *See* Hypersensitivity reactions
Allergic rhinitis, **2809,** 2811f
Allergic vasculitis, 462
Allergy, **2804**
  in asthma, 2221
  atopic, 2804
  eosinophils in, 2805
  in international traveler, 1021
  pathophysiology of, 2804–2807, 2806f
Allergy testing, 2215
Alligator bite-wound, 1143, 1145t
Alloantibodies, 901
Allodynia, 93, 172
Alloimmunization, 901, 905, 911–912
Allopurinol
  adverse effects of
    cutaneous, 393, 401, 422, 2959–2960
    DIHS, 423. *See also* Drug-induced hypersensitivity syndrome (DIHS)
    fever, 150
    genetic considerations in, 419
    racial differences in, 2960
    serious, 2959–2960
    SJS or TEN, 401, 425, 2959–2960. *See also* Stevens-Johnson syndrome (SJS); Toxic epidermal necrolysis (TEN)
    vasculitis, 2907, 2940t
  drug interactions of, 489t
    azathioprine, 2408, 3631
    6-mercaptopurine, 562t
  genetic variations in response to, E72t
  for gout, 2959–2960
  for hyperuricemia, 820, 823, 2453
  for tumor lysis syndrome, 593–594
  for uric acid nephrolithiasis, 2453
Allostasis, 211
Allotypes, 883
All-*trans*-retinoic acid (ATRA). *See also* Retinoic acid
  actions and targets of, 533t, 539, 571t, 573, 834
  for acute promyelocytic leukemia, 533t, 539, 834, 2984
  adverse effects of, 406, 420, 571t, 573, 587, 590, 834
  interactions and issues, 571t
Almotriptan, 3467t, 3468, 3468t
Alopecia, **394**
  androgenetic, 394
  androgenic, 3138
  central centrifugal cicatricial, 394
  definition of, 378t
  drug-induced, 394, 421, 576
  etiology of, 394, 394t
  in lichen planus, 388
  nonscarring, 394
  scarring, 394
  in secondary syphilis, 394
  in SLE, 2829, 2830f
  in tinea capitis, 389
  traumatic, 394
  treatment of, 395t
Alopecia areata, 394, 395t, A5
Alosetron, 312, 2577t
ALP. *See* Alkaline phosphatase (ALP)
Alpelisib, 534t, 535, 571t, 574, 640, 3982
α-adrenergic antagonists
  adverse effects of, 159, 266, 282t, 2145t
  aldosterone-renin ratio effects of, 3062t
  drug interactions of, 704
  for hypertension, 2145t
  for lower urinary tract symptoms, 704, 2456, 3175
  overdosage/poisoning with, 3705, 3712t
  for pheochromocytoma, 3075
α chains, 3322
Alpha-1,3-galactose (alpha-gal), 2813–2814, 2819
α_1-adrenergic agonists, 3710t
α_2-adrenergic agonists, 225, 2145t, 3691, 3711t, 3908
α4 integrin, 2789, 2792t
α4 β7 integrin, 2791t
"Alpha coma," 191
Alpha diversity, 3835
α fetoprotein (AFP)
  in carcinoma of unknown primary, 732, 733t
  ectopic production of, 736
  in germ cell tumors, 712
  in HCC, 660, 661
  in testicular cancer, 705, 708t
  as tumor marker, 504t
α-galactosidase, 302
Alpha-gal syndrome, 3730
α-globin genes, 770f, 771, 776, 778
Alpha-granules, 938
α-heavy chain disease (Seligman's disease), 894
α-hemolytic streptococci, 1206, 1206t, 1214
α-ketoadipic acidemia, 3369t
α-Klotho, 3318
α-linolenic acid, 2602t
α-lipoic acid, 243, 266
α motor neurons, 170f
Alpha (α) radiation, S6
α-synuclein
  in Creutzfeldt-Jakob disease, 3532
  in DLB, 195t, 3492
  in multiple system atrophy, 3548
  in neurodegenerative diseases, 3401, 3401t
  in PD, 3400, 3494, 3498–3499
  in primary autonomic failure, 3548
  in REM sleep behavior disorder, 218
α thalassemia. *See* Thalassemia syndromes
α thalassemia trait, 778t
α thalassemia with myelodysplasia (ATMDS), 778t
α-tocopherol transfer protein, 3537
Alphaviruses. *See* Togaviridae
ALPK1 gene mutations, 2937
ALPK3 gene mutations, 2004t
Alport's syndrome
  clinical features of, 248, 2360, 2414, 2415t, 2426–2427
  diagnosis of, 2426
  genetic considerations in, 2360–2361, 2412, 2427, 3320, E36t
  hearing loss in, 248, E36t
  renal biopsy in, 2426, A4
  treatment of, 2426
Alprazolam
  adverse effects of, 85
  for anxiety disorders, 3660t
  for depression, 85
  drug interactions of, 1742t
  overdosage/poisoning with, 3713t
  for panic disorder, 3658
  pharmacology of, 3660t
Alprostadil, 3160
ALPS (autoimmune lymphoproliferative syndrome), 524t, 2803
ALS. *See* Amyotrophic lateral sclerosis (ALS)
Alsin gene, 3521, 3522t
ALS/parkinsonian/dementia complex of Guam, 3485
Alström's syndrome, 3184t
Alsuviricetes. *See* Togaviridae
ALT. *See* Alanine aminotransferase (ALT)
Alteplase (rtPA)
  actions of, 955
  catheter-directed therapy with, 2164
  contraindications to, 2120
  domain structure of, 955, 955f
  indications for, 955
    acute limb ischemia, 2175
    ischemic stroke, 2298, 3441, 3441t
    pulmonary embolism, 2164
    STEMI, 2120
  intrapleural administration of, 2290
Alternative medicine, 3924. *See also* Complementary and integrative therapies and practices
Altitude illness. *See* High-altitude illness
Alum, 997t, 998
Aluminum acetate, 254
Aluminum hydroxide, 2524, S6
Aluminum phospate gel, S6
Aluminum toxicity
  hypercalcemia due to, 3282
  neurologic effects of, 3484, 3702
  pulmonary effects of, 2238

Alveolar collapse, 2302
Alveolar flooding, 2294
Alveolar gas equation, 2204
Alveolar hemorrhage, diffuse, 275
Alveolar hypoventilation, 2295
Alveolar proteinosis, primary, A12, S9
Alvimopan, 98, 2593
Alzheimer's disease (AD), **3475**
  aluminum and, 3702
  clinical features of, 194, 196t, **3475**
    aphasia, 202, 203
    Bálint's syndrome, 206
    circadian disorders, 3952
    depression, 3664
    frontal lobe syndromes, 208
    involuntary weight loss, 315
    memory loss, 207
    myoclonus, 3530
    olfactory dysfunction, 237, 241
  deaths from, 75t
  diagnosis of, 3476–3477
  differential diagnosis of, 196t, 3476
  vs. DLB, 3492, 3492t. *See also* Dementia with Lewy bodies
  in Down's syndrome, 3479
  early symptomatic, 3475
  epidemiology of, 3475, 3478
  functional anatomy of, 193, 206f
  genetic considerations in, 195t, 3404, 3479–3480, 3941
  HSV infection and, 1500
  imaging in, 198f, 3476–3477, 3477f, 3481f
  incidence of, 3475
  pathology of, 195t, 3478–3479, 3478–3479f
  pathophysiology of
    Aβ, 3400, 3404, 3478, 3478f
    Aβ amyloidosis, 895
    Aβ prions, 3401, 3525
    β-amyloid, 3398
    circadian rhythms, 3952
    complement cascade, 3395
    genetic factors, 3395, 3397, 3803
    molecular, 195t
    oligodendrocytes, 3397
    organoids, 3404
    prions, 3525
    protein aggregation, 3994
    tau prions, 3401
  treatment of, 199–200, 3480–3482, 3481f
  in women, 3163–3164
AMA (antimitochondrial antibodies), 2711
Amalgam tattoo, 264t
AMAN (acute motor axonal neuropathy), 3617, 3617t, 3618f, 3619t
Amanita mushrooms, 306, 324, 2665, 3711t
Amantadine
  actions of, 1488
  adverse effects of, 3503
  for antidepressant side effect management, 3660t
  for influenza prophylaxis, 1488t, 1594t
  for influenza treatment, 1488t
  overdosage/poisoning with, 3715t
  for PD, 3503, 3506
  resistance to, 1488, 1488t, 1537
  for SSRI-related sexual dysfunction, 3660t, 3667
Amaurosis, congenital. *See* Leber's hereditary optic neuropathy
Amaurosis fugax, 225–226, 3431
Ambiguous genitalia, 3789
Amblyomma spp. ticks, 1445, 1452, 1456, 2813, 3730. *See also* Hard tick relapsing fever (HTRF); Tick-borne diseases
Amblyopia, 234
Ambrisentan, 2194, 2195, 2197t
Amebae, free-living. *See Acanthamoeba* spp.; *Balamuthia* spp.; *Naegleria*
Amebiasis, **1753**. *See also Entamoeba histolytica*
  clinical features of, 1737, 1756
  cutaneous, 409
  diagnosis of, 1756–1757, 1757f, S12
  differential diagnosis of, 1757
  epidemiology of, 1753–1754, S12
  extraintestinal, 1756, S7
  vs. IBD, 2560–2561
  intestinal, 306t, 1739t, 1756, 2560–2561
  liver abscess in. *See* Liver abscess, amebic
  pathogenesis of, 1754–1756, 1755f
  prevention of, 1758
  in travelers, 1079t
  treatment of, 1747, 1748, 1757–1758, 1757t
  in war veterans, S7
Amebomas, 1755, 1756
Amegakaryocytic thrombocytopenia, 809t, 813
Amelogenesis imperfecta, 261
Amenamevir, 1487, 1503
Amenorrhea
  chemotherapy-related, 576
  definition of, 3132
  diagnosis of, 3132–3133, 3134f
  etiology of, 2549t, 3132–3133
  functional hypothalamic, 3134
  hypothalamic-pituitary-gonadal axis in, 3132, 3133f
  primary, 3132
  secondary, 3132
  treatment of, 3134–3135
American Society of Anesthesiologists, physical status classification, 2581t
American Thoracic Society/European Respiratory Society (ATS/ERS) criteria, 2254, 2255t
American trypanosomiasis. *See* Chagas disease
Americium exposure, S6
AMH (anti-müllerian hormone), 3094, 3126, 3128
AMI (acute myocardial infarction). *See* Myocardial infarction (MI)
Amikacin
  for *A. baumannii* infections, 1297t
  actions of, 1166, 1182t
  adverse effects of, 1177, 1399, 1426
  for *Citrobacter* infections, 1293
  for community-acquired pneumonia, 963t
  for *Edwardsiella* infections, 1294
  for *Enterobacterales* infections, 1283
  for *Enterobacter* infections, 1292
  for extraintestinal *E. coli* infections, 1287
  for hospital-acquired pneumonia, 963t, 1032t
  indications for, 1173t, 1176
  for *Klebsiella* spp. infections, 1291
  liposome inhalation suspension, 1428
  for MAC infections, 1427, 1428
  for MDR-TB, 1394t, 1399
  for *Morganella* and *Providencia* infections, 1293
  for *Nocardia* infections, 1362
  for *P. aeruginosa* infections, 1309t
  in pregnancy and lactation, 1169t
  for *Proteus* spp. infections, 1291
  resistance to, 1173t, 1182t, 1184, 1219, 1283, 1359f, 1426
  for *Serratia* infections, 1293
  for TB, 1426
  for ventilator-associated pneumonia, 1032t
Amiloride
  actions of, 2367f, 2370
  adverse effects of, 359, 2144t
  for ascites, 329
  for AVP resistance, 3019
  for hypertension, 2144t
  for Liddle's syndrome, 355, 3063
  for lithium-associated diabetes insipidus, 352, 353
  for mineralocorticoid excess, 3063
Amino acid(s)
  analysis of, 3367
  essential, 2598, 3367
  functions of, 3367
  hyperkalemia and, 358
  metabolism of, disorders of, **3367,** 3368–3370t
  in parenteral nutrition, 2622
  renal transport of, 2366f, 2369
  requirement for, 2619
  transport, disorders of, **3373,** 3374t
  types of, 3367
Aminoacyl-tRNA histidyl synthetase, 2790t
Aminoacyl-tRNA synthetase, 2790t
Aminobenzoic acid esters, 434t
ε-Aminocaproic acid
  adverse effects of, 358
  for C1INH deficiency, 2809
  for DIC, 934
  for hemophilia, 930
  for platelet dysfunction disorders, 926
  for thrombocytosis, 820, 824
  for VWD, 927
Aminoglutethimide, 3012
Aminoglycosides
  actions of, 1166, 1176, 1182t, 1183f
  adverse effects of, 1171t
    in cirrhosis with ascites, 2715
    cutaneous, 423
    hearing loss, 246, 1177, 3807
    hypokalemia, 355
    interstitial disease, 2356
    in myasthenia gravis, 3627
    nephrotoxicity, 1041, 1177, 2377
  for cystic fibrosis, 2248
  for enterococcal infections, 1219, 1219t
  indications for, 1176
  for infective endocarditis, 1041
  for osteomyelitis, 1064t
  for *P. aeruginosa* infections, 1309t
  postantibiotic effect of, 1168
  resistance to
    in *Eikenella,* 1267
    in *Enterobacterales,* 1283t, 1284–1285
    in enterococci, 1219, 1221
    mechanisms of, 1182t, 1183f, 1184
    in mycobacteria, 1397–1398
  for TB, 1426
5-Aminolevulinic acid (ALA), 430
Aminomethylcycline. *See* Omadacycline
Aminophylline, 291, 376, 423
Aminopyridine, 3588
5-Aminosalicylic acid (5-ASA) agents, 2564, 2564t
Aminosidine. *See* Paromomycin
Aminotransferase(s), 324, 2635, 2637t
Amiodarone
  actions of, 1916, 1916t
  adverse effects of, 1916t
    cutaneous, 50t, 400, 420
    hepatotoxicity, 2668, 2671
    hyperthyroxinemia, 3024t
    myopathy, 2940t, 3648t
    neuropathy, 3609t, 3610
    nonarrhythmic, 1958
    optic neuropathy, 228
    phototoxicity, 434t
    proarrhythmic, 1958
    pulmonary, 2261
    SA node dysfunction, 1920t
    secondary parkinsonism, 3497
    thrombocytopenia, 923t
    thyroid abnormalities, 2018
    thyrotoxicosis, 3039–3040

- Amiodarone (*Cont.*):
  - for atrial fibrillation, 1952
  - contraindications to, 1941
  - drug interactions of
    - calcium channel blockers, 485t
    - dabigatran, 489t
    - daclatasvir, 2689
    - digoxin, 489t
    - protease inhibitors, 485t
    - quinidine, 1745t
    - warfarin, 485t, 489t, 490
  - ECG changes caused by, 1874, 1875f
  - for focal atrial tachycardia, 1938
  - monitoring during treatment with, 1958
  - for multifocal atrial tachycardia, 1946
  - for recurrent ventricular tachycardia, 1972
  - for ventricular arrhythmias, 1958, 1961, 1964, 1965, 2336, 2337f
  - for ventricular tachycardia, 2124
- Amiodarone-induced thyroxicosis (AIT), 3039–3040
- Amithiozone, 1395
- Amitriptyline
  - adverse effects of, 3659t, 3666
    - cardiac, 1920t
    - drowsiness, 3470t
    - smell and taste distortion, 243
  - for chronic daily headache, 115
  - for ciguatera poisoning, 3727
  - for cough, 274
  - for depression, 504, 3588, 3659t, 3666
  - for fibromyalgia, 2965, 2966t
  - for functional dyspepsia, 302
  - for IC/BPS, 335
  - for insomnia, 216
  - for migraine prevention, 3469, 3470t
  - for nausea and vomiting, 299t
  - for new daily persistent headache, 117
  - for olfactory dysfunction, 243
  - overdosage/poisoning with, 3711t, 3712t
  - for pain, 97t, 506, 3588
  - for tension-type headache, 3471
- Amivantamab, 535t, 559f, 627, 627t, 3982
- AML. *See* Acute myeloid leukemia (AML)
- Amlodipine, 1996, 2101, 2101t, 2144t, 2178, 2874
- Ammonia/ammonium, 372, 2368, 2636
- Ammonia exposure, 2239t
- Ammoniagenesis, 373, 2368
- Amnesia, **207**
  - after traumatic brain injury, 3571
  - anterograde, 207
  - caring for patient with, 209
  - clinical examination in, 207, V2
  - etiology of, 207–208
  - psychogenic, 3485
  - retrograde, 207
  - transient global, 207–208, 3484–3485
- Amnesic shellfish poisoning (domoic acid intoxication), 3414, 3727–3728
- Amnestic state, 207
- *AMN* gene mutations, 787
- Amniocentesis, 3797
- Amnionless, 781
- Amniotic fluid embolism, 2159
- Amodiaquine
  - actions of, 1740
  - adverse effects of, 1741t, 1746, 1771t, 1775
  - for malaria, 1770t, 1771t
  - pharmacology of, 1746, 1771t
- Amoxapine, 3659t, 3666
- Amoxicillin
  - for actinomycosis, 1365, 1366t
  - adverse effects of, 419, 2528
  - for anthrax, S4
  - for community-acquired pneumonia, 1027t
  - for *C. trachomatis* infection, 1474
  - for endocarditis prophylaxis, 1048t
  - for enteric fever, 1318, 1318t
  - for gonococcal arthritis, 1058
  - in *H. pylori* eradication, 1302t, 2527–2528, 2527t, 2530t
  - indications for, 1173t, 1175
  - for leptospirosis, 1444, 1444t
  - for Lyme disease, 1058, 1452–1453, 1453f
  - for nontyphoidal *Salmonella* infections, 1321t
  - for osteomyelitis, 1065t
  - for otitis media, 254, 254t, 1194
  - for pneumococcal infections, 1194
  - in pregnancy and lactation, 1169t
  - prophylactic uses of, 1180t
  - resistance to, 1093, 1173t, 1265, 1291, 1324, 2528
  - for rheumatic fever, 2857
  - for sinusitis, 256
  - for streptococcal pharyngitis, 258, 258t
  - UTIs, 1219t, 1220t
  - UTIs in pregnancy, 1094
- Amoxicillin-clavulanate
  - adverse effects of, 2667, 2667t, 2671
  - for bite-wound infections, 1054t, 1144, 1145t
  - for community-acquired pneumonia, 1027t
  - for lung abscess, 1035
  - for MDR-TB, 1425–1426
  - for *Nocardia* infections, 1362
  - for osteomyelitis, 1065t
  - for otitis media, 254
  - for pelvic inflammatory disease, 1106
  - prophylactic, 1180t
  - resistance to, 1359f
  - for sinusitis, 256
- AMPA receptor antibodies, 743t, 747–748, 748f, 1114
- AmpC β-lactamase, 1181–1183, 1284, 1292, 1293, 1296
- Amphetamines
  - abuse of, 3693
  - actions of, 3656t
  - adverse effects of
    - alopecia, 394
    - cardiomyopathy, 2016
    - hypertension, 2137
    - myopathy, 3648t
    - palpitations, 291
    - vasculitis, 2940t
  - for fatigue, 167
  - overdosage/poisoning with, 3710t, 3715t
- Amphigoric amino acid changes, 3784
- Amphiphysin, 2790t
- Amphiphysin antibodies, 743t, 750, 750f
- Amphotericin B
  - adverse effects of, 1691
    - frequent, 1741t, 1784
    - hypokalemia, 355
    - interstitial disease, 2356
    - nephrotoxicity, 1601, 2377
    - occasional, 1741t
    - rare, 1784
    - thrombocytopenia, 923t
  - for aspergillosis, 1716, 1717t
  - for blastomycosis, 1702–1703, 1703t
  - for *Candida* infections
    - disseminated, 1711t, 1712
    - endocarditis, 1044
    - esophageal, 1710t
    - liver abscess, 1075
    - resistance to, 1711t
    - urinary tract, 1095
  - for *Candida* prophylaxis, 1712
  - for chronic meningitis, 1135
  - for coccidioidomycosis, 1127, 1698, 1698t
  - for cryptococcosis, 1127, 1607, 1705
  - drug interactions of, 1741t
  - for fungal infection prophylaxis, 2286, 2410
  - for fungal sinusitis, 582
  - for fusariosis, 1726t, 1729
  - for histoplasmosis, 1127, 1695, 1695t
  - indications for, 1691
  - lipid formulations of, 1691
  - for mucormycosis, 583, 1720–1721, 1721t
  - for mucosal leishmaniasis, 1787
  - for neutropenic fever, 581
  - for paracoccidioidomycosis, 1726, 1726t
  - for phaeohyphomycoses, 1726t
  - in pregnancy and lactation, 1741t
  - resistance to, 1716
  - for sporotrichosis, 1724t
  - for talaromycosis, 1726t, 1727
  - for visceral leishmaniasis, 1784
- Ampicillin
  - adverse effects of, 325, 402, 923t, A1
  - for clostridial infections, 1239t
  - drug interactions of, 1741t, 2525
  - for enterococcal infections, 1219, 1219t
  - for HACEK infections, 1267t
  - indications for, 1173t, 1175
  - for infective endocarditis, 1041, 1043t, 1045, 1048t, 1219t, 1220t
  - for leptospirosis, 1444, 1444t
  - for *L. monocytogenes* infections, 1229
  - for meningitis
    - based on pathogen, 1122, 1122t
    - *E. faecalis,* 1219t
    - empirical treatment, 1121t
    - Hib, 1263
  - for necrotizing fasciitis, 1054t
  - for nontyphoidal *Salmonella* infections, 1321t
  - for pneumococcal infections, 1194
  - in pregnancy and lactation, 1169t
  - prophylactic uses of, 1180t
  - resistance to, 1173t
    - in *Citrobacter,* 1293
    - in *E. coli,* 1287
    - in *Enterobacter,* 1292
    - in *Enterococcus* spp., 1219
    - in *H. influenzae,* 1263
    - in *Klebsiella* spp., 1291
    - in *N. gonorrhoeae,* 1255
    - in *Proteus* spp., 1291
  - for UTIs, 1219t
  - for UTIs in pregnancy, 1094
- Ampicillin-clavulanate, 1175
- Ampicillin-sulbactam
  - for *A. baumannii* infections, 1296, 1297t
  - for bite-wound infections, 1054t, 1144, 1145, 1145t
  - for *Capnocytophaga* infections, 1267, 1267t
  - for cellulitis, 1054t
  - for community-acquired pneumonia, 963t, 1028t
  - for HACEK infections, 1043t, 1267t
  - for head and neck abscesses, 259
  - indications for, 1175
  - for infective endocarditis, 1043t
  - for osteomyelitis, 1065t
  - for *P. multocida* infections, 1267t
  - resistance to, 1291, 1379, 1379t
  - for ventilator-associated pneumonia, 1032t
- AMPK (adenosine monophosphate–activated protein kinase), 543, 3871
- Amplatzer septal occluder, 2130, 2130f, A11, A11v
- Amprenavir, 1619f
- Ampullary carcinoma, 326
- Amputation, 3229, 3772
- AMS (acute mountain sickness), 3755–3756, 3756t
- AMSAN (acute motor sensory axonal neuropathy), 3617, 3617t, 3618f, 3619t
- Amsel criteria, for bacterial vaginosis, 1102
- Amur virus, 1660t, 1676

- Amygdalohippocampectomy, for epilepsy, 3425
- Amylase, 328, 2739t, 2740, 2741t, 2754
- Amyl/butyl nitrate, 3159
- Amylin, 3212, 3871, 3872f
- Amylin agonists, 3211t
- Amyl nitrite, S5
- Amyloid angiopathy, cerebral, 3453t, 3454, 3476
- Amyloid beta. *See* Aβ (amyloid beta)
- Amyloidosis, **895**
  - AA. *See* AA amyloidosis
  - AF, 900
  - AH, 896t
  - AL. *See* AL amyloidosis
  - amyloid fibril proteins in, 895
  - ATTR, 895, 896t, 899–900
  - Aβ, 895
  - Aβ$_2$M, 895, 896t, 900
  - cardiac, **2020**
    - arrhythmias in, 2027
    - clinical features of, 2021
    - diagnosis of
      - algorithm for, 2026f
      - CMR, 1894, 1896f, 2023, 2024f, A9
      - CT, 2023
      - echocardiography, 2022–2023, 2022–2023f, A9
      - PYP scan, 2023, 2024–2025f
      - ^{99m}technetium pyrophosphate scan, A9
    - differential diagnosis of, 2021–2022
    - pathology of, 2020–2021, 2021f
    - rare forms of, 2027t
    - SA node dysfunction in, 1918
    - transthyretin-mediated (TTR), 2022, 2024f, 2027, A9
    - treatment of, 2025–2028
    - typing of, 2023–2025
  - classification of, 895, 896t
  - clinical features of
    - autonomic neuropathy, 3550
    - bleeding, 473
    - cutaneous, 404, 407
    - diarrhea, 309
    - gastroparesis, 297
    - hyperkalemia, 359
    - neuropathy, 3603–3604
    - ocular, 221
    - proteinuria, 342
    - renal, 2415t, 2424–2425, A4
    - splenomegaly, 480
  - diagnosis of, 895–896, 896f, 3550
  - familial, 896t, 899
  - in familial Mediterranean fever, 899, 2935
  - global considerations in, 899
  - in IBD, 2562t
  - in leprosy, 1409
  - lichen, 404
  - localized, 896t
  - macular, 404
  - in MGUS, 897
  - in multiple myeloma, 897, A9
  - nodular, 404
  - pathogenesis of, 895–896, 3550, 3994
  - small-intestinal mucosal biopsies in, 2550, 2550t
  - systemic, 896t
  - treatment of, 2425
- Amyloid plaques, 195t, 197–198, 3400, 3529
- Amyloid polyneuropathy, familial, 3603
- Amyloid precursor protein (APP). *See also APP* gene mutations
  - in AD, 3400, 3478, 3479, 3479f, 3941
  - in amyloidosis, 896t, 898
  - in neurodegenerative disease, 3400
- Amyloid-related imaging abnormalities (ARIA), 199–200, 3480–3481, 3481f
- Amylolytic enzymes, 2743
- Amylopectinosis (Andersen disease, type IV GSD), 3360t, **3362**
- Amyotrophic lateral sclerosis (ALS), **3519**
  - clinical features of, 3521, 3596
  - differential diagnosis of, 3521, 3523
  - dysphagia in, 294, 296
  - epidemiology of, 3521
  - familial, 3521, 3522t, 3789
  - frontotemporal dementia and, 3523
  - genetic considerations in, 3522t
  - imaging in, 3520f
  - juvenile-onset, 3522t
  - parkinsonism and, 3523
  - pathogenesis of, 3523
  - pathology of, 3519–3520, 3520f
  - pathophysiology of, 3395, 3397, 3404–3405, 3994
  - treatment of, 2275–2276, 3407, 3524, 3968, 3968f
- Amyotrophic lateral sclerosis (ALS)/parkinsonian/dementia complex of Guam, 3485
- Amyotrophy, 3520, 3604
- ANA. *See* Antinuclear antibodies (ANA)
- Anabolic-androgenic steroid (AAS) abuse
  - adverse effects of, 3123, 3172–3173, 3173t
    - cardiovascular, 3123, 3172–3173
    - cholestasis, 325
    - dyslipidemia, 3249
    - edema, 282t
    - hepatic adenoma, 670
    - hepatotoxicity, 2671
    - hirsutism, 421
    - hyperbilirubinemia, 323
    - long-term, 3123, 3173
  - diagnosis of, 3123, 3173, 3174t
  - epidemiology of, 3123, 3172
  - treatment of, 3123–3124, 3173–3174
  - withdrawal symptoms, 3123
- Anaerobic bacteria, **1371**
  - vs. aerobic bacteria, 1371
  - antibiotic resistance in, 1378
  - antibiotic susceptibility testing of, 1378
  - historic perspective of, 1371
  - human health and, 1372
  - in normal human microbiota, 1371–1372, 1371t
  - specimen handling, 1378
  - virulence factors of, 1373
- Anaerobic bacterial infections, mixed
  - approach to the patient, 1373–1374
  - clinical features of
    - bacteremia, 1377–1378
    - bone and joint, 1377
    - with clostridial involvement, 1239–1240, 1239t
    - CNS, 1375
    - in dog-bite wounds, 1143
    - endocarditis, 1378
    - in human bite-wounds, 1144
    - intraabdominal, 1376
    - mouth, head, and neck, 1374
    - pelvic, 1376–1377
    - pleuropulmonary, 1375, 1376f
    - skin and soft tissue, 1377
  - definitions in, 1371
  - diagnosis of, 1378, S11
  - epidemiology of, 1374, 1374f
  - etiology of, 1372–1373, 1372f
  - pathogenesis of, 1373
  - treatment of, 1378–1380, 1379t
- Anaerobic respiration, 1371
- Anagen, 3138
- Anagen effluvium, 394, 421
- Anagrelide, 821, 824–825
- Anajatuba virus, 1660t, 1674
- Anakinra
  - adverse effects of, 2850t
  - for calcium pyrophosphate deposition disease, 2961
  - for cytokine release syndrome, 594
  - for familial Mediterranean fever, 2936
  - for fever of unknown origin, 155
  - for gout, 2959
  - for macrophage activation syndrome, 469
  - for pericarditis, 2080
  - for rheumatoid arthritis, 2850t, 2851
- Anal cancer, **659,** 1526, 1527, 1530, 1614
- Anal fissures, **2586**
- Anal fistula, 1389, **2585,** 2585f
- Analgesia (sensory abnormality), 172
- Analgesics
  - adverse effects of, 2357, 2361f, 2442
  - for lumbar puncture, S3
  - medication-overuse headache and, 115–116, 3469
  - urinary, 1094
- Anal intercourse, HIV infection risk and, 1562–1563
- Anal reflex, 3381
- Anal sphincter, 304f
- Anaphylactic reactions. *See also* Hypersensitivity reactions
  - to dialyzer, 2400
  - drug-induced, 422
  - to insect stings, 3735–3736
- Anaphylaxis, **2812**
  - clinical features of, 489, 2812
  - definition of, 2812
  - diagnosis of, 2813, 2813t
  - drug-induced, 417, 422, 489
  - exercise-induced, 2807
  - from iron therapy, 766
  - pathophysiology of, 2770, 2782, 2812
  - predisposing factors for, 2812–2813
  - prevention of, 2814
  - respiratory acidosis in, 375
  - treatment of, 2813, 3735–3736
  - vaccine-related, 1008
- *Anaplasma bovis,* 1462
- *Anaplasma capra,* 1455t, 1462
- *Anaplasma* spp. infections, 1121, 1452, 1462. *See also* Human granulocytic anaplasmosis
- Anaplastic large-cell lymphoma, 860, 860t, 867
- Anaplastic thyroid cancer, 3048
- Anasarca, 1983, 1984
- Anastrozole, 567, 634, 753t
- Anatomic dead space, 2204
- Anatomic reentry, 1913–1914
- ANCAs. *See* Antineutrophil cytoplasmic antibodies (ANCAs)
- Anchoring heuristic, 21
- *Ancylostoma* spp.
  - *A. braziliense,* 1716, 1812, S12
  - *A. caninum,* 1816
  - *A. ceylanicum,* 1816
  - *A. duodenale,* 1815t, 1816, S12
- Andersen disease (type IV GSD), 3360t, **3362,** 3645
- Andersen-Tawil syndrome, 3647
- Andersson lesions, 1062
- Andes virus, 1656, 1660t, 1674
- Andexanet, 953–954, 1951, 2163, 3456
- Androgen(s)
  - abuse of. *See* Anabolic-androgenic steroid (AAS) abuse
  - actions and metabolism of, 3105–3106, 3105f. *See also* Androgen insensitivity syndrome (AIS)
  - for adrenal insufficiency, 3070
  - adverse effects of, 420, 2703t
  - body fat distribution and, 3166

- Androgen(s) (*Cont.*):
  - deficiency of. *See* Hypogonadism; Testosterone deficiency
  - excess of, 3138
  - in hair growth and differentiation, 3138
  - synthesis of
    - disorders of, 3098–3099, 3099t
    - pathways of, 3051f, 3100f, 3104–3106, 3104f
- Androgen blockade. *See* Androgen deprivation therapy (ADT)
- Androgen-depletion syndrome, 701
- Androgen deprivation therapy (ADT)
  - adverse effects of, 701–702, 753t, 3175–3176, 3176f, 3296
  - approach to the patient, 3176–3177, 3176t
  - intermittent use of, 702
  - for prostate cancer, 567, 701–702, 3175
  - with radiation therapy, 700
- Androgen insensitivity syndrome (AIS), **3099**
  - amenorrhea in, 3133
  - clinical features of, 3099–3100, 3099t, 3133
  - complete, 3100
  - genetic considerations in, 2985, 3100, 3115, 3133, 3783t
  - gynecomastia due to, 3115
  - partial, 3099
  - testicular dysfunction due to, 3115
  - treatment of, 3133
- Androgen receptors, 536, 2985, 3099t, 3105–3106
- Androgen therapy. *See* Testosterone therapy
- Andropause, 3117
- Androstenedione, 3138
- Anejaculation, 3155
- Anemia, **443**. *See also specific types*
  - acute episodes, in sickle cell disease, 774t
  - after kidney transplantation, 2411
  - after liver transplantation, 2723
  - of aging, 769
  - in AKI, 2381, 2384
  - from blood loss, 807–808
  - chemotherapy-related, 575
  - in CKD. *See* Chronic kidney disease (CKD), hematologic abnormalities in
  - classification of, 448–450, 449t, 451f
  - clinical features of, 444
  - in colorectal cancer, 656
  - compensation for, 444
  - in critically ill patient, 2298, 2323–2324
  - definition of, 448
  - diagnosis of, 444–448
    - bone marrow examination, 447–448, 448f
    - laboratory tests, 444, 445t
    - peripheral blood smear, 445–447, 445–447f, 447t. *See also* Peripheral blood smear
  - dyspnea in, 269t, 270
  - endoscopic findings in, 2495–2496
  - etiology of, 449–450
  - in gastrointestinal bleeding, 316
  - in heart failure, 768, 1985, 1998
  - in HIV infection, 1604–1605
  - in hypothermia, 3770
  - hypoxia in, 277
  - of inflammation, 767–769, 767t, 768t
  - in kidney disease, 768
  - lactic acidosis in, 369
  - macrocytic, A6
  - in multiple myeloma, 887
  - myelophthisic, 818
  - normocytic, in rheumatoid arthritis, 2840
  - in older adults, 450, 769, 3883
  - in SLE, 2831
  - in thalassemia, 776
  - treatment of, 450
- Anergy, 2784, 2789
- Anesthesia (sensory abnormality), 172
- Anesthetics, 493, 3908, E72t
- Aneurysm
  - aortic. *See* Aortic aneurysm
  - cerebral artery, 3384f
  - coronary artery, 2907, A9
  - fusiform, 2165
  - intracranial, 3394. *See also* Subarachnoid hemorrhage
  - left ventricular, 2125
  - mycotic. *See* Mycotic aneurysm
  - popliteal artery, 2176
  - renal artery, 2153f
  - saccular, 2165, 2166, **3458**. *See also* Subarachnoid hemorrhage
  - true, 2165
- Ang1/Ang2. *See* Angiopoietins
- Angelman's syndrome, 3791, 3939, 3941
- Angiitis
  - allergic. *See* Eosinophilic granulomatosis with polyangiitis
  - cutaneous leukocytoclastic, 2905
  - isolated granulomatous, 1132t
- Angina
  - microvascular, 2091
  - stable, **2092**. *See also* Ischemic heart disease
    - in aortic regurgitation, 2044
    - in aortic stenosis, 2037
    - atypical, 2093
    - exertional, 2092
    - heart failure and, 2103
    - history in, 2092–2093
    - myocardial infarction probability with, 105f
    - in myocardial ischemia, 102–104, 103t, 105f, 2092–2093
    - physical examination in, 2093–2094
    - severity classification of, 2093, 2093t
    - threshold for development of, 2093
    - treatment of. *See* Ischemic heart disease, treatment of
  - unstable, 102, 103t, 2106. *See also* Non-ST-segment elevation acute coronary syndrome (NSTE-ACS)
  - vasospastic (Prinzmetal's variant), 1873, 2101, **2112**
- Angina decubitus, 2092
- Anginal equivalents, 2107
- Angina pectoris, 102. *See also* Angina
- Angiodermatitis, acral, 406
- Angioedema
  - in anaphylaxis, 2812
  - bradykinin-mediated, 2809
  - classification of, 2807, 2807t
  - clinical features of, 402–403, 427t, 2808
  - definition of, 2807
  - diagnosis of, 2808
  - drug-induced, 422, 427, 427t
  - etiology of, 402t, 2807, 2807t
  - hereditary, 2808
  - isolated, 2807
  - pathophysiology of, 2808
  - predisposing factors for, 2807
  - treatment of, 422, 2808–2809
  - vibratory, 2807
  - without urticaria, 2807
- Angiofibroma, 397, 404, 524t, 3080t, 3084
- Angiogenesis
  - in asthma, 2218
  - in gastric mucosa, 2516
  - growth factors in, 1847
  - inhibition of, 547–548, 547f, 548f, 3829
  - tumor, 529t, 545–546
- Angiogenic switch, 545
- Angiography
  - cerebral
    - in CNS vasculopathy, 3447f
    - complications of, 3393
    - indications for, 3393
    - in stroke, 3437–3438, 3447f
    - in subarachnoid hemorrhage, 3459, 3460f
  - in chronic thromboembolic pulmonary hypertension, 2193f
  - coronary. *See* Coronary angiography
  - CT. *See* Computed tomography (CT) angiography (CTA)
  - in lower gastrointestinal bleeding, 320
  - magnetic resonance. *See* Magnetic resonance angiography (MRA)
  - pulmonary, 2161
  - renal intraarterial, 2151f, 2153f
  - spinal, 3393–3394
- Angioimmunoblastic T cell lymphoma, 860t, 867, 875
- Angiokeratoma, 405, 3355, A15
- Angioma
  - cavernous, 3453t, 3457
  - cherry, 378t
  - senile, 380f
  - spider, 327, 2629, 2709, 2709f
- Angiomyolipoma, renal, 2434
- Angiopoietins, 546, 546f, 548f
- Angiosarcoma, 406, 729
- *Angiostrongylus* spp. infections
  - abdominal, 1819
  - clinical features of, 1131t, 1739t, 1813, 1819
  - diagnosis of, 961t, 1813, S12
  - epidemiology of, 961t, S12
  - life cycle of parasite, 1813, 1813t, 1819, S12
  - pathogenesis of, 1737, 1813
  - treatment of, 1812t, 1813
- Angiotensin-converting enzyme (ACE)
  - in glomerular filtration rate regulation, 2364, 2365f
  - in sarcoidosis, 2928
  - in SARS-CoV-2 cell entry, 3167f, 3177
  - in sodium absorption, 2372
- Angiotensin-converting enzyme (ACE) inhibitors
  - actions of, 273, 2133, 2144t
  - adverse effects of
    - acid-base disorders, 373
    - anaphylaxis, 489
    - angioedema, 2807
    - cough, 272, 273, 2143, 2221
    - cross-sensitivity and, 427
    - cutaneous, 392, 420, 422
    - erectile dysfunction, 3157
    - hyperkalemia, 359
    - hypoglycemia, 3235
    - lupus syndrome, 2940t
    - renal, 340, 2374, 2374f, 2392
    - teratogenic, 2143
  - for aortic aneurysm in Marfan's syndrome, 2167
  - for aortic dissection, 2170
  - drug interactions of, 489t
  - for heart failure, 1993f, 1994, 1995t
  - for HIV-associated nephropathy, 1601
  - for hypertension, 2143, 2144t
    - in CKD, 2147, 2392, 2396
    - in diabetic nephropathy, 3225
    - in metabolic syndrome, 3258
  - for ischemic heart disease, 2102
  - metabolism of, 486
  - for NSTE-ACS, 2110, 2112
  - for peripheral artery disease, 2173
  - preoperative continuation of, 3908
  - for prevention of sudden cardiac death, 2339
  - for pulmonary edema, 2331
  - for scleroderma renal crisis, 2446, 2875
  - for STEMI, 2121–2122
  - timing of administration of, 3954

- Angiotensin II
  - aldosterone levels and, 355
  - in glomerular filtration rate regulation, 2364, 2365f, 2373
  - in sodium reabsorption, 345, 2372
  - in urinary tract obstruction, 2455
- Angiotensin receptor blockers (ARBs)
  - actions of, 2133, 2144t
  - adverse effects of, 2143, 2144t
    - cutaneous, 422
    - hyperkalemia, 359
    - hypoglycemia, 3235
    - renal, 2374, 2374f, 2392
  - aldosterone-renin ratio effects of, 3062t
  - for heart failure, 1993f, 1994, 1995t
  - for hypertension, 2143, 2144t
    - in CKD, 2147, 2392, 2396
    - in diabetic nephropathy, 3225
    - in metabolic syndrome, 3258
  - for ischemic heart disease, 2102
  - metabolism of, 486, 493
  - for peripheral artery disease, 2173
  - for Raynaud's phenomenon, 2874
  - for STEMI, 2122
- Angle of Louis, 1860
- Anhidrotic ectodermic dysplasia, 2799
- Anidulafungin, 1692, 1711t, 1712t
- Anifrolumab, 2834, 2836
- "Animal rule," S4
- Animals
  - bite-wounds. *See* Bite-wounds
  - infectious disease and exposure to, 584, 959
  - nontyphoidal *Salmonella* infections and, 1319
- Animal-to-human host switching, 3855, 3856f
- Anion gap, 367–368
- *Anisakis*/anisakiasis, 961t, 1818
- Anismus, 313
- Anisocoria, 221
- Anisocytosis, 437, 445, 445f, 482
- Anistreplase, 955
- Ankara vaccine, 1149
- *ANK* gene mutations, 794t, 3668
- *ANKH* gene mutations, 2960
- Ankle-brachial index (ABI), 1862, 1863f, 2172
- Ankle clonus, 130
- Ankle injury, 49t
- Ankle, pulse measurement in, 1863f
- *ANKRD26* gene mutations, 825, 825t
- Ankylosing spondylitis. *See* Axial spondyloarthritis
- ANNA-3, 619
- Annelid worm stings, 3724, 3724f
- Annexin-2, 470
- Annular pancreas, 2754
- Annular skin lesion, 378t
- Annuloaortic ectasia, 287
- *ANO3* gene mutations, 3510, 3510t
- Anoctaminopathy (LGMD2L), 3641
- Anogenital warts. *See* Condyloma acuminata
- Anomia, 201, 222
- Anomic aphasia, 201t, 202
- *Anopheles* mosquitoes, 1760–1762, 1761f, 1820t, 3748
- Anorchia, bilateral, 3098
- Anorectal (rectoanal) angle, 304, 304f
- Anorectal disorders, **2582**
  - anal fissures, **2586**
  - anorectal abscess, 2585, 2585f
  - candidiasis, 1708
  - in Crohn's disease, 2558
  - *C. trachomatis* infection, 1471
  - fecal incontinence. *See* Fecal incontinence
  - fistula in ano, 1389, **2585**
  - hemorrhoidal disease, 313, 2503f, 2584t, 3758
  - in HIV infection, 1600, 1600f
  - perianal skin tags, in Crohn's disease, 2563
  - perianal ulcer, 1108–1109
  - rectal prolapse, **2582,** 2582f
  - solitary rectal ulcer syndrome, 318, 2561
  - in systemic sclerosis, 2870, 2870t
  - warts, 1527, 1527f
- Anorectal manometry, 313, 2583
- Anorectum, 304f
- Anorexia
  - in appendicitis, 2595
  - in cancer patient, 507t, 509
  - in chronic disease, 2621
  - in gastric cancer, 646
  - in terminally ill patient, 91t
- Anorexia nervosa, 298, 320, **3671**
- Anorexiants, 3191–3193, 3257
- *ANOS1 (KAL1)* gene mutations, 3111t, 3112
- Anosmia. *See* Olfactory disorders
- Anosognosia, 205, 3428, 3476
- Anovulation, 576, 3133
- Anoxic cerebral injury, in critically ill patient, 2298
- ANP. *See* Atrial natriuretic peptide (ANP)
- Anserine bursitis, 2944, 2973
- Antacids
  - adverse effects of, 306, 308, 373
  - drug interactions of, 489t, 1172t, 1741t, 1742t
  - for GERD, 302
  - for internal radionuclide contamination, S6
  - for peptic ulcer disease, 2523–2524, 2524t
- Antagonistic pleiotropy, 3869
- Antegrade urography, 2456
- Anterior cerebral artery, 170, 3430f, 3431f
- Anterior choroidal artery, 3432f
- Anterior choroidal artery infarct, A16
- Anterior dens dislocation, A16
- Anterior horn cell disorders, 169, 170
- Anterior ischemic optic neuropathy (AION), 226–227, 227f
- Anterior pituitary
  - hormone expression and regulation in, 2987, 2988t. *See also specific hormones*
  - insufficiency. *See* Hypopituitarism
  - tumors. *See* Pituitary tumors
- Anterior spinal artery syndrome, 3561, 3564–3565
- Anterocollis, 3510
- Anterograde amnesia, 207
- Anterolateral system, 172, 173f
- Anthracyclines
  - adverse effects of, 421, 574, 575t, 577, 753–754, 753t, 2016
  - for AML, 830
  - for metastatic breast cancer, 641
  - for soft tissue sarcomas, 729
- Anthralin, 395t
- Anthrax
  - in bioterrorism, 1149, S4
  - clinical features of, 995, S4
  - cutaneous, 145t, 149, 1051, A1, S4
  - diagnosis of, S4
  - gastrointestinal, S4
  - inhalational, 995, S4
  - injection, S4
  - prevention of, S4
  - treatment of, S4
- Anthrax vaccine, S4
- Anthrax vaccine adsorbed, S4
- Anthropometrics, 2617, 2618t
- Anti-AChR, 3625, 3627, 3627t
- Antiandrogens. *See* Androgen deprivation therapy (ADT)
- Antiangiogenic agents, 547–548, 547f
- Antiarrhythmics, **1915**. *See also specific drugs*
  - actions of, 1915–1916, 1916t
  - adverse effects of, 306, 1920t
  - for atrial fibrillation, 1951–1952
  - classification of, 1915–1916, 1916t
  - genetic variations in response to, 493
  - overdosage/poisoning with, 3715t
  - for pain, 97t, 100
  - for ventricular arrhythmias, 1957
- Antibacterial drugs, **1165**. *See also specific drugs and diseases*
  - absorption of, 1167
  - actions of, 1165–1167, 1182t, 1183f
  - adverse effects of, 799t, 1082, 1171t
  - allergy to, 1170
  - for animal bite-wound prophylaxis, 1145–1146, 1145t
  - approach to therapy with, **1168,** 1172t
    - duration, 1170
    - expert guidance for, 1170
    - failure to respond, 1170
    - host factors, 1168, 1170
    - inappropriate, 252
    - site of infection, 1168
  - for bronchiectasis prophylaxis, 2244
  - concentration-dependent, 1168
  - for dental procedure prophylaxis, 266, 1048, 1048t
  - distribution of, 1167
  - empirical therapy with, **1168**
    - for acutely ill febrile patient, 990, 991t
    - for community-acquired pneumonia, 1027–1028, 1027t
    - for fever of unknown origin, 155
    - for infective endocarditis, 1044–1045
    - for sepsis/septic shock, 2321t
  - excretion of, 1167
  - gut microbiome and, 3834
  - inappropriate use of, 1282–1283
  - mechanisms of action, 1165–1167, 1182t
  - metabolism of, 1167
  - pharmacodynamics of, 1168, 1168f
  - pharmacokinetics of, 1167, 1168f
  - in pregnancy and lactation, 1168, 1169t
  - prophylactic uses of, 463, **1179,** 1180t
  - resistance to, 3855. *See also specific drugs*
    - in anaerobic bacteria, 1378
    - definition of, 1181
    - in *Enterobacterales,* 1283–1284, 1283t
    - epidemiology of, 1185–1186
    - genomics in detection of, 977t, 980–982
    - in gram-negative bacteria, 1027, 1152
    - in health care–associated infections, 1151–1152
    - mechanisms of, 1026–1027, 1181, 1182t, 1183f
    - in *M. pneumoniae,* 1027
    - prevention of, 1185–1186
    - in *S. pneumoniae,* 1026–1027
    - testing for, S11
  - selection of, 983
  - stewardship of, 254, 1152–1153, 1152t, 1179–1180, 1282–1283, 1285
  - time-dependent, 1168
  - urgency of, 1185–1186, 1186t
- Anti-β_2-glycoprotein, 2837
- Antibody(ies). *See also* Immunoglobulin(s); *specific antibodies*
  - affinity maturation of, 2778
  - blood group, 901, 904–904t
  - for cancer treatment. *See* Cancer biologic therapy
  - cytotoxic reactions of, 2783
  - definition of, 2755
  - drug-dependent, 922–923
  - in neurologic paraneoplastic syndromes, 743–744, 743t
  - in parasitic infections, S12
  - virus-specific, detection of, 1484
- Antibody-dependent cell-mediated cytotoxicity (ADCC), 1587f, 1588, 2755, 2768

Antibody-drug conjugates, 535t, 536, 558, 565–566, 566t, 753t, 756
Antibody enhancement, 1588
Antibody isotype switching, 2780
Antibody-mediated rejection, 2030, 2287
Anticardiolipin (aCL), 2837
Anti-CCP, 2262, 2847, 2847t, 2945
Anticholinergics. *See also specific drugs*
  adverse effects of, 183, 266, 2575
  for asthma, 2224
  for dystonia, 3512
  for IBS, 2575
  for lower urinary tract symptoms, 704, 3175
  for nausea and vomiting, 298, 299t
  overdosage/poisoning with, 3705, 3711t
Anticholinesterase agents. *See* Cholinesterase inhibitors
Anticipation, 3793
Anticoagulant(s), **946**
  for acute limb ischemia, 2175
  adverse effects of, 407, 1951, 2163
  dental procedures and, 266
  direct oral, 952–954, 953t. *See also specific drugs*
  drug interactions of, 489t
  for DVT/PE, 3902
  guidelines for use of, S13
  in infective endocarditis, 1046
  intracerebral hemorrhage due to, 3454–3455
  in noncardiogenic stroke prevention, 3450–3451
  for NSTE-ACS, 2110t
  oral. *See* Warfarin
  parenteral. *See specific drugs*
  reversal of, 934, 953–954, 2163, 3455–3456
  for STEMI, 2111, 2121
  for stroke prevention, 1950–1951, 1951t, 3445t, 3450
  for suppurative thrombophlebitis, 1142
  for VTE/PE treatment, 2162–2163, 2163t
Anticonvulsants. *See* Antiepileptic drugs (AEDs)
Anticytokine therapy, 136
Anti-dandruff shampoo, 386
Antidepressants. *See also specific drugs*
  actions of, 3652, 3654
  adverse effects management, 3660t
  adverse effects of, 85, 306, 420, 421, 2137, 3659t
  in dementia, 199
  for depression, 3659t, 3666–3667, 3666f
  for fatigue, 167
  for functional dyspepsia, 302
  for IBS, 2575
  for insomnia, 86, 216
  for nausea and vomiting, 299
  overdose/poisoning with, 3711t, 3715t
  for pain, 97t, 99–100, 100t
  for panic disorder, 3658
  for personality disorders, 3672
  for somatic symptom disorder, 3670
  in terminally ill patient, 85
Antidiuretic hormone (ADH). *See* Arginine vasopressin (AVP)
Antidotes, 3706, 3710. *See also specific drugs/poisons*
Antidote Treatment Nerve Agent Autoinjector (ATNAA), S5
Antidromic AV reentry, 1940f, 1941
Anti-dsNDA, 2832, 2832t
Antiemetics, 298–299, 299t
Antiepileptic drugs (AEDs). *See also specific drugs*
  actions of, 3415
  adverse effects of
    cobalamin malabsorption, 786t
    cross-sensitivity of, 427–428
    cutaneous, 422
    DIHS, 424, 489. *See also* Drug-induced hypersensitivity syndrome (DIHS)
    erythroderma, 393
    hair loss, 421
    neurologic, 3421–3423t
    osteoporosis, 3296
    in pregnancy, 784, 3427
    SJS/TEN, 401, 425. *See also* Stevens-Johnson syndrome (SJS); Toxic epidermal necrolysis (TEN)
    systemic, 3421–3423t
    vitamin D deficiency, 3290
    weight gain, 3188
  for brain abscess, 1137
  breast-feeding and, 3427
  discontinuation of, 3424
  drug interactions of, 591, 1744t, 1749, 2408, 3421–3423t, 3426–3427
  for focal seizures, 3420, 3423–3424
  for generalized seizures, 3424
  genetic factors in response to, 3420
  initiation of, 3420, 3424
  monitoring of, 3424
  for nausea and vomiting, 300
  for neurocysticercosis, 1138
  overdosage/poisoning with, 3713t
  for pain, 97t, 100
  pharmacology of, 3421–3423t
  in pregnancy, 3427, 3903
  for refractory epilepsy, 3424
  for seizures in brain tumors, 716
  selection of, 3420, 3420t, 3424
Anti-factor Xa assay, 475, 947, 948
Antifibrinolytic drugs, 930
Antifungal agents, **1691**. *See also specific drugs*
  amphotericin B. *See* Amphotericin B
  azoles, 1691–1692, 1742t
  echinocandins, 1692, 1711–1712, 1711t, 1712t, 1720–1721, 1721t
  flucytosine. *See* Flucytosine
  griseofulvin, 388t, 389, 395t
  ibrexafungerp, 1692, 1698, 1710t
  for neutropenic fever, 580–581
  resistance to, 1710t, 1711–1712, S11
  terbinafine. *See* Terbinafine
  topical, 1692
Antiganglioside antibodies, 3618–3619, 3618f, 3619t
Antigen(s)
  B cell receptor for, 2755
  definition of, 2755
  macroscopic detection of, S11
  polysaccharide, selective antibody deficiency to, 2802
  T cell receptor for, 2756
  T cell recognition of, 2774–2775, 2776f
  viral, detection of, 1484
Antigenic drift, 1483, 1544
Antigenic shift, 1483, 1544
Antigen-presenting cells, 559f, 560, 2407, 2407f, 3582
Antiglomerular basement membrane (anti-GBM) disease, 275, 2412, 2415t, 2418–2419, A4
Anti-GQ1b, 3618–3619
Antihistamines. *See also specific drugs*
  for allergic rhinitis, 2810, 2811f
  for atopic dermatitis, 393t
  for contact dermatitis, 393t
  for cough, 273
  for insect stings, 3736
  for insomnia, 216
  for nausea and vomiting, 298, 299t
  overdosage/poisoning with, 3705, 3711t, 3715t
  for pityriasis rosea, 388
  for premedication, 598
  for pruritus, 383, 422
  for scombroid poisoning, 3726
  for systemic mastocytosis, 2821
  for urticaria, 2808
  for vertigo, 165t
Antihistone, 2832
Anti-HMGCR myopathy, 2912–2913
Anti-Hu paraneoplastic neuropathy, 3624
Anti-IgE. *See* Omalizumab
Anti-insulin receptor, 3091t, 3092–3093
Anti-Jo-1, 2913, 2913f, 2946t
Anti-La, 2832t, 2877, 2878t, 2946t
Antimalarial drugs. *See also specific drugs*
  adverse effects of, 1771t
    cutaneous, 386, 388, 392, 399, 420
    in GP6D deficiency, 799–800, 799t
    seizure, 3414t
  for discoid lupus erythematosus, 415
  drug interactions of, 1741–1742t
  for malaria, 1770t, 1771t
  for malaria prophylaxis, 1774t
  overdosage/poisoning with, 3714t, 3715t
  in pregnancy and lactation, 1741–1742t
  properties of, 1771t
  resistance to, 1768, 1775, 3848
  for SLE, 2833–2834
  for subacute cutaneous lupus erythematosus, 415
Antimetabolites. *See also specific drugs*
  actions of, 1166
  adverse effects of, 576, 753t
  antibacterial, 1182t
  in cancer chemotherapy, 562t, 564–565
Antimicrobial peptides, 970, 2758t
Antimicrobial stewardship, 252–253, 254, 1152–1153, 1179–1180, 1282–1283, 1285
Antimitochondrial antibodies (AMA), 2711
Antimitotic agents, 563t, 565. *See also specific drugs*
Antimonial compounds. *See* Pentavalent antimonials
Anti-müllerian hormone (AMH), 3094, 3126, 3128
Antimycobacterial agents, **1418**. *See also* Tuberculosis, treatment of; *specific drugs*
Antimyeloperoxidase (MPO) antibodies, 2419–2420, 2894, 2897
Antineutrophil antibodies, 457
Antineutrophil cytoplasmic antibodies (ANCAs)
  cytoplasmic, 2894
  in drug-induced lupus, 420
  in drug-induced vasculitis, 426, 2907
  in EGPA, 2231, 2894
  in granulomatosis with polyangiitis, 2894, 2897, 2898
  in microscopic polyangiitis, 2898
  perinuclear, 2894
  in rheumatoid arthritis, 2847
  in vasculitis, 2419–2420, 2893–2894
Antinuclear antibodies (ANA)
  in musculoskeletal disease, 2945, 2946t
  rheumatoid arthritis, 2847
  in SLE, 2831–2832, 2832t, 2946t
  in systemic sclerosis, 2871
Antioxidants, 2706, 3926
Antiparasitic agents, **1740,** 1741–1746t. *See also specific drugs*
Antiphospholipid antibodies, 475, 2832, 2832t, 2837
Antiphospholipid syndrome (APS), **2837**
  in cancer, 741
  catastrophic, 2839
  clinical features of, 31f, 150, 2415t, 2446, 2837–2838, 2838t
  diagnosis of, 2838–2839, 2838t
  differential diagnosis of, 2838–2839
  epidemiology of, 2837

  - high-altitude symptoms of, 3758
  - pathogenesis of, 2446, 2790t, 2837
  - in pregnancy, 2837, 2838t
  - renal manifestations of, 2446
  - skin manifestations of, 409
  - in SLE, 2830
  - treatment of, 2446, 2839
- α_2-Antiplasmin, 471f, 472, 954, 954f
- Antiplatelet drugs, **942**
  - actions of, 942
  - during and after PCI, 2128
  - chronic use in cardiac disease, 3445t
  - discontinuation before dental procedures, 266
  - discontinuation before lumbar puncture, S3
  - guidelines for use of, S13
  - for ischemic heart disease, 2102, S13
  - for NSTE-ACS, 2110–2111, 2110t, 2112
  - in PCI, 2104
  - perioperative, 3908
  - for peripheral artery disease, 2173
  - prophylactic, S13
  - for STEMI, 2121
  - for stroke/TIA prevention, 3445t, 3449–3450, S13
  - for stroke treatment, 3442–3443
- Antiporters (exchangers), 2366
- Antiproliferative factor, 331
- Anti-proteinase 3, 2419
- Antipsychotics
  - actions of, 3672–3673
  - for acute alcohol intoxication, 3679
  - adverse effects of, 2137, 3188, 3516–3517, 3673t, 3674, 3715t
  - in dementia, 199, 3482
  - drug interactions of, 3660t
  - overdosage/poisoning with, 3711t, 3715t
  - for schizophrenia, 3672–3674, 3673t
- Antipyretic agents, 134, 136
- Antireflux surgery, 302
- Antiretroviral agents. *See* Combination antiretroviral therapy (cART)
- Antiribosomal P, 2832t
- Anti-RNP, 2832t
- Anti-Ro, 2832, 2832t, 2836, 2877, 2878t, 2946t
- Antisense oligonucleotides, 3524
- Anti-Sm, 2832, 2832t, 2946t
- Anti-SRP myopathy, 2912–2913
- Antisynthetase syndrome (AS), 2910t, 2913, 2914–2915, 2914f, 2916t
- Antithrombin
  - actions of, 471f, 472, 475
  - activation of, 946, 946f, 948–949
  - deficiency of, 940
- Antithrombotic drugs, 942, 942f
  - anticoagulants. *See* Anticoagulant(s)
  - antiplatelet drugs. *See* Antiplatelet drugs
  - fibrinolytic drugs. *See* Fibrinolytic drugs
  - for STEMI, 2121
- Antithrombotic mechanisms, 470–472, 471f
- Antithymocyte globulin (ATG)
  - adverse effects of, 2722
  - for aplastic anemia, 812–813
  - in HCT preparation, 915
  - for immunosuppression
    - after cardiac transplantation, 2029t
    - after liver transplantation, 2722
    - before kidney transplantation, 2408
  - for MDS, 817–818
- Antithyroid drugs, 421, 3035–3036. *See also specific drugs*
- Anti-TNF-α antibodies. *See* Tumor necrosis factor-α (TNF-α) inhibitors
- Antitoxin
  - botulinum, 1236, 1237, S4
  - diphtheria, 1224
  - tetanus, 1231
- α_1-Antitrypsin (AAT), 2548
- α_1-Antitrypsin (AAT) deficiency
  - clinical features of
    - cutaneous, 406
    - granulomatosis with polyangiitis, 2420
    - hepatic, 2629, 2712
    - pulmonary, 2242t, 2243, 2250, 2253
  - diagnosis of, 2630t, A13
  - early interventions for, 3804t
  - genetic considerations in, 3787, 3993–3994
  - genetic testing for, 3796, 3804t
  - HCC and, 659
  - protein misfolding in, 3993–3994
- α_1-Antitrypsin (AAT) therapy, 2257–2258
- Antitumor antibiotics, 562t, 564
- Anti-tumor cell antibodies, 558–559, 559f
- Antitussive agents, 274
- Antivenom
  - adverse effects of, 3719
  - indications for, 3719
  - for marine envenomations, 3723, 3725
  - for scorpion stings, 3735
  - for snakebite treatment, 3719–3721, 3720–3721t
- Antiviral therapy, **1484**. *See also* Combination antiretroviral therapy (cART); *specific drugs and diseases*
  - for chronic HBV infections, 1489–1492, 1490t
  - for chronic HCV infections, 1492–1494, 1492t
  - for hepatitis D infections, 1494
  - for herpesvirus infections, 1485–1487, 1486t
  - for respiratory virus infections, 1487–1489, 1488t
- Antley-Bixler craniosynostosis, 3099
- Anton's syndrome, 3432
- Antrectomy, for duodenal ulcers, 2532–2533
- Ant(s), stinging, 3736
- Anuria, 341
- *ANXA11* gene mutations, 2928
- Anxiety
  - asthma exacerbations and, 2222
  - in cancer patient, 757
  - definition of, 3657
  - erectile dysfunction and, 3157
  - female sexual dysfunction and, 3161
  - melatonin for, 3926
  - in respiratory acidosis, 375
  - vs. syncope, 161
- Anxiety disorders, **3657**
  - alcohol-induced, 3676
  - generalized anxiety disorder, **3658,** 3660t
  - at high altitudes, 3759
  - insomnia in, 215
  - nausea and vomiting in, 298
  - obsessive-compulsive disorder, **3663**
  - palpitations in, 291
  - panic disorder, 103t, 104, **3657**
  - phobic disorders, **3661**
  - stress disorders, **3661**. *See also* Posttraumatic stress disorder (PTSD)
  - vertigo in, 165
  - in women, 3167
- Anxiolytics. *See* Benzodiazepine(s)
- Aorta
  - acute occlusion of, 2170–2171
  - aneurysm of. *See* Aortic aneurysm
  - atherosclerosis of, 2165, 2170
  - congenital anomalies of, 2165
  - dissection of. *See* Aortic dissection
  - IgG4-related disease of, 2931, 2931t
  - inflammatory disease of. *See* Aortitis
  - rupture of, 2088
  - syphilis of, 2043
  - thrombus in, S13
- Aortic aneurysm, **2165**
  - abdominal, **2167**
    - clinical features of, 327, 2167–2168
    - diagnosis of, 2168, 3914v
    - epidemiology of, 2167
    - physical examination in, 1860, 1863
    - rupture of, 2168
    - screening for, 39t, 2168, 3881t
    - treatment of, 2168–2169, 2168f
  - chest pain in, 103t, 104
  - etiology of, 2165–2166, 2166t
  - familial, 2166
  - hemoptysis in, 275
  - infectious causes of, 2166
  - in late syphilis, 1432, 2166
  - medial degeneration in, 2165–2166
  - monitoring expansion of, 2168
  - mycotic. *See* Mycotic aneurysm
  - in relapsing polychondritis, 2166
  - rupture of, 111, 112t, 2168
  - thoracic, 2166–2167, 2167f, 2903, A14
  - traumatic, 2166
  - treatment of, 2167, 2168–2169
- Aortic dissection, **2169**
  - aortic regurgitation in, 2169
  - chest pain in, 103t, 104, 105
  - classification of, 2169, 2169f
  - clinical features of, 2169–2170
  - diagnosis of, 22, 106, 2170
  - intramural hematoma and, 104, 2169
  - murmur in, 287
  - penetrating atherosclerotic ulcer and, 104, 2169
  - in pregnancy, 3900
  - risk factors for, 2169
  - treatment of, 2170
- Aortic occlusion, 2170–2171
- Aortic regurgitation, **2043**
  - in aortic dissection, 2169
  - with aortic stenosis, 2065
  - in aortitis, 2171
  - carotid pulse in, 1862f
  - clinical features of, 291, 2043–2044
  - diagnosis of
    - cardiac catheterization, 2045
    - chest radiograph, 2045
    - CMR, 1892, 1903v
    - ECG, 2044
    - echocardiography, 1891–1892, 1891f, 2044, 2045f
    - history in, 2043–2044
    - physical examination in, 2044
    - POCUS, 3914v
  - etiology of, 2043, 2043t
  - heart sounds/murmur in, 283, 287, 1860, 1865, 1865f, 2044
  - with mitral stenosis, 2065
  - pathophysiology of, 2043
  - in pregnancy, 3900
  - in rheumatic heart disease, 2855
  - treatment of, **2045,** 2046f, 2047f
- Aortic root disease, 287, 2043, 2046–2047, 2047f, 3243, 3900
- Aortic stenosis, **2033**
  - with aortic regurgitation, 2065
  - in bicuspid aortic valve, 2035
  - carotid pulse in, 1862, 1862f
  - clinical features of, 31f, 2037
  - congenital, 2036
  - diagnosis of
    - cardiac catheterization, 2038
    - chest radiograph, 2038
    - ECG, 2037–2038
    - echocardiography, 1890–1891, 2038, 2038f, A9
    - hemodynamic measurements, 1905, 1906f
    - valve area, 1907
  - epidemiology of, 2035, 2035f
  - etiology of, 2035–2036, 2035t

- Aortic stenosis (*Cont.*):
  - heart sounds/murmur in, 1866
    - characteristics of, 1865, 1865f, 2037
    - duration, 283
    - intensity and radiation, 283, 285
    - systemic vascular resistance and, 289
  - with mitral regurgitation, 2065
  - myocardial ischemia in, 2091
  - natural history of, 2038
  - palpitations in, 291
  - pathogenesis of, 2035–2037, 2036f
  - pathophysiology of, 2036–2037
  - in pregnancy, 3900
  - preoperative evaluation and management of, 3909
  - rheumatic, 2035–2036
  - stroke prevention in, 3445t
  - sudden cardiac death in, 2335t
  - systems biology applied to, 3961t
  - treatment of
    - medical, 2038–2039
    - percutaneous aortic balloon valvuloplasty for, 2040, 2131
    - strategy for, 2039f
    - valve replacement for. *See* Aortic valve replacement
- Aortic valve
  - bicuspid, 1891f, 2035, 2036
  - sclerosis, 286
  - stenosis. *See* Aortic stenosis
- Aortic valve replacement
  - indications for, 2039–2040, 2039f, 2045–2046, 2046f
  - with mitral valve replacement, 2067
  - paravalvular abscess following, A9
  - procedure and patient selection for, 2040, 2041f, 2042t
  - transcatheter. *See* Transcatheter aortic valve implantation (TAVI)
- Aortic valvuloplasty, 2040
- Aortitis, **2171**
  - aneurysm in, 2166
  - etiology of, 2166t
  - in giant cell arteritis. *See* Giant cell arteritis
  - idiopathic, 2171
  - IgG4-related, 2931, 2931t
  - imaging in, 1898, 1901f, A9
  - infective, 2171
  - rheumatic, 2171
  - syphilitic, 1432, 2171
  - in Takayasu's arteritis. *See* Takayasu's arteritis
- Aortobifemoral bypass, 2174
- Aortocolic fistula, 318
- Aortoenteric fistula, 318
- Aortography, 1907
- Aortoiliac disease, 2172. *See also* Peripheral artery disease
- APACHE II scoring system, 2292, 2292t, 2293f
- *APAF-1* gene, 539
- APAF-1 protein, 539
- Apalutamide, 701–702, 702f, 753t
- Apatite, 2961. *See also* Calcium apatite deposition disease
- *APC* gene mutations
  - in colorectal cancer, 520f, 653
  - in familial adenomatous polyposis, 522, 524t, 653
  - in Gardner's syndrome, 717t
  - germline and somatic, 524f
- Apeú virus, 1658t, 1667
- Aphagia, 293
- Aphasia, **200**
  - in AD, 3475
  - anomic, 201t, 202
  - Broca's, 201t, 202
  - caring for patient with, 209
  - clinical examination in, 201, V2
  - clinical features of, 201t
  - conduction, 201t, 202
  - fluent transcortical, 201t, 202
  - global, 201t, 202
  - isolation, 201t, 202
  - jargon, 203
  - in neurodegenerative disease, 203–204
  - nonfluent transcortical, 201t, 202
  - primary progressive, 201, 203–204, 206f
  - progressive, V2
  - pure alexia without agraphia, 201t, 203
  - pure word deafness, 201t, 203
  - subcortical, 202–203
  - Wernicke's, 201–202, 201t
- Aphemia, 203
- Apheresis, 902–903t, 1779–1780, 1824
- Aphthous ulcers
  - etiology of, 261, 456
  - in HIV infection, 1598, 1599f
  - in IBD, 2562t, 2563
  - prevention of, 463
  - recurrent, 261, 263t
- Apical ballooning syndrome, 2088
- Apixaban
  - adverse effects of, 953, 1950
  - dosage of, 952, S13
  - indications for, 952
  - management before endoscopy, 2483t
  - management before lumbar puncture, S3
  - pharmacology of, 953t
  - reversal of, 3456
  - for stroke prevention in cancer, 3451
  - for stroke prevention in cardiac disease, 952, 1950, 1951t, 3450
  - for VTE/PE treatment, 742, 2162, 2163t
  - for VTE prophylaxis, 2164t
- APL. *See* Acute promyelocytic leukemia (APL)
- APLAID (PLCγ-associated antibody deficiency and immune dysregulation), 2804
- Aplasia cutis, 3036, 3901
- Aplastic anemia, **808**
  - after liver transplantation, 2723
  - classification of, 809t
  - clinical features of, 811
  - definition of, 808
  - diagnosis of, 811–812, A6
  - epidemiology of, 809
  - etiology of, 809t
    - chemical-induced, 809, 810t, 811
    - drug-induced, 809, 810t, 811
    - immunologic, 810, 811
    - infectious, 809–810
    - inherited, 809t, 810
    - paroxysmal nocturnal hemoglobinuria, 804, 810
    - posthepatitis, 810
    - in pregnancy, 810
    - radiation-induced, 809
  - history in, 811
  - monocytopenia in, 462
  - pathophysiology of, 450, 809f, 810–811
  - physical examination in, 811
  - prognosis of, 812
  - in telomere disease, 3818, 3822, 3822f
  - treatment of, 812–813, 915, 918, 919t
- Apnea
  - sleep. *See* Sleep apnea
  - in terminally ill patient, 91t
- Apnea testing, 192
- *APOA1* gene mutations, 3241t, 3248
- *APOA5* gene mutations, 3240
- APOBEC, 1552
- *APOBEC3F* gene, 1582t, 1585
- *APOBEC3G* gene, 1582t
- APOBEC proteins, 1558, 1585
- *APOB* gene mutations, 3247
- *APOC2* gene mutations, 3240
- *APOC3* gene mutations, 3240, 3250
- *ApoE2* gene, 3245
- *ApoE* gene, 1582t, 1585, 3479, 3787
- *APOE* gene, 3238, 3245
- *ApoE ε4* allele, 3163, 3479
- Apoi virus, 1659t
- *APOL1* gene mutations
  - in CKD, 2357, 2359t, 2386–2387, 2412
  - in HIV infection, 1582t, 1585
- *APOL1/MYH9* gene, 2154
- *APOL05* gene mutations, 1793
- Apolipoprotein(s), 3237–3238, 3237t
- Apolipoprotein(a), 3237t, 3238
- Apolipoprotein A-I (ApoA-I), 3237t, 3248
- Apolipoprotein A-V (ApoA-V), 3237t, 3238, 3240
- Apolipoprotein B, 3237–3238, 3237t, 3239
- Apolipoprotein B-48, 3237t, 3238
- Apolipoprotein B-100, 3237t, 3238, 3242
- Apolipoprotein C-II (ApoC-II), 3237t, 3238
- Apolipoprotein C-II (ApoC-II) deficiency, 2744
- Apolipoprotein C-III (ApoC-III), 3237t
- Apolipoprotein E, 3237t, 3238
- *APOM* gene mutations, 2843
- Apomorphine, 3502, 3504t, S6
- Aponeurotic ptosis, 234
- Apoplexy, pituitary, 232, 2992–2993, 3067, 3068t
- Apoptosis
  - aberrant, autoimmune disease and, 2823–2824
  - of cancer cells, 529t, 539–541, 540f, 556, 557f
  - cellular and molecular control of, 2781
  - definition of, 973, 2755
  - in HIV infection, 1576
  - initiation of, 973, 1481
  - pathways of, 539–541, 540f, 557f, 973
  - of red blood cells, 443
  - in smoking, 2251
- Apotransferrin, 762, 765f
- APP. *See* Amyloid precursor protein (APP)
- Appendectomy, 2597
- Appendiceal fecalith, 2596, 2596f
- Appendicitis, **2594**
  - clinical features of, 112, 112t, 297, 2594–2596, 2595t
  - complications of, 2597
  - diagnosis of, 113, 2596, 2596f
  - differential diagnosis of, 1106, 2594–2595, 2594t
  - epidemiology of, 2594
  - in immunocompromised patients, 2596
  - liver abscess and, 1074
  - in older adults, 2596
  - pathogenesis of, 2594
  - peritonitis in, 1071
  - in pregnancy, 2596
  - treatment of, 2596–2597
- Appendix, 680–681, 2594, 2594–2595f
- Apperceptive agnosia, 207
- Appetite, 2603
- Appetite suppressants, 3191–3193, 3257
- *APP* gene mutations, 195t, 3400, 3404, 3479, 3941
- Apraxia
  - in AD, 3476
  - buccofacial, 203
  - construction, 206
  - definition of, 168, 203
  - dressing, 206
  - gait, 178
  - ideational, 203
  - ideomotor, 203
  - limb-kinetic, 203
  - oculomotor, 206, V3
  - of speech, 203
  - sympathetic, 203

Apremilast
  adverse effects of, 387t, 2896t
  for Behçet syndrome, 2909
  for psoriasis/psoriatic arthritis, 387, 387t, 2890
Aprepitant, 83, 116, 298, 299t, 300, 575
Apriso, 2564, 2564t
Aprosodia, 203
APs. *See* Accessory pathways (APs)
APS. *See* Antiphospholipid syndrome (APS)
APS-1. *See* Autoimmune polyendocrine (polyglandular) syndrome-1 (APS-1)
APS-2. *See* Autoimmune polyendocrine (polyglandular) syndrome-2 (APS-2)
aPTT. *See* Activated partial thromboplastin time (aPTT)
AQP4 antibodies, 3590, 3592
Aquagenic urticaria, 2807
Aquaporin antibodies, 744t, 750
Aquaporins
  AVP actions and, 3015, 3015f
  in CNS flow, 3398
  in collecting duct, 2370
  in loop of Henle, 2369
  in nephrogenic diabetes insipidus, 352, 2455
  in neuromyelitis optica, 35, 227
  in proximal tubule, 2367
  in water balance, 345, 345f, 2371, 2371f
Araçatuba pox, 1518t
Arachnidism, necrotic, 3734
Arachnodactyly, 3332
Arachnoid cysts, 3003
Araraquara virus, 1660t, 1674
Araucária virus, 1660t, 1674
Arboviruses, 1111–1112
Arbovirus infections, 1018, 1111–1112, 1125, 3748–3749
ARBs. *See* Angiotensin receptor blockers (ARBs)
Arcanobacterium haemolyticum infections, 140t, 402, 1226
Arcobacter spp. infections, 1303, 1304t, 1305
Arcuate fasciculus, 201
Arcus corneae, A15
Arcus senilis, A15
ARDS. *See* Acute respiratory distress syndrome (ARDS)
Area postrema (chemoreceptor trigger zone), 82, 297
Area under the receiver operating curve (AUC), 3972, 3973t
Arenaviridae
  clinical syndromes of, 993, 1666t, 1667
  geographic distribution of, 1664–1665t
  reservoirs and vectors of, 1657–1658t
  structure of, 1478f, 1479, 1480t
ARF (acute rheumatic fever). *See* Rheumatic fever
ARF gene mutations, 599, 601
Argasid (soft) tick, 3730
Argatroban
  actions of, 950
  for anticoagulation in HIT, 923, 1905
  monitoring during treatment with, 950
  pharmacology of, 950t
  for VTE/PE treatment, 2162, 2163t
Argentinian/Junin hemorrhagic fever, 1658t, 1675
Arginase deficiency, 3369t
Arginine, 3372, 3816
Arginine vasopressin (AVP), 3014
  actions of, 3015–3016, 3016–3017f
  deficiency of
    in diabetes insipidus. *See* Diabetes insipidus (DI), with primary AVP deficiency
    etiology of, 3017t
    gestational, 352, 3016, 3017t
    in hypodipsic hypernatremia, 3019, 3019t
  ectopic production of. *See* Syndrome of inappropriate antidiuresis (SIAD)
  in edema, 280
  in heart failure, 1981f
  metabolism of, 3016
  resistance to, 3017t. *See also* Diabetes insipidus (DI), with AVP resistance (nephrogenic)
  for shock, 2315
  structure of, 3014, 3015f
  synthesis and secretion of, 344–345, 3014, 3015f
  in thirst regulation, 3016, 3016f
  in urine volume regulation, 3015, 3015f
  in water balance regulation, 344–345, 345f, S1
Arginine vasopressin (AVP) receptor-2, 2985t
Argininosuccinic acidemia, 3369t
Argininosuccinic aciduria, 3372
Argon plasma coagulation, 2484, 2501f, V5
Argyll Robertson pupil, 221, 1432
Argyria, 400
ARIA (amyloid-related imaging abnormalities), 199–200, 3480–3481, 3481f
ARID1A gene mutations, 661t, 713, 734, 893, 3940
ARID2 gene mutations, 661t
Aripiprazole, 3658, 3673, 3673t, 3674
Aristolochia, 2442, 3926
Aristolochic acid, 660
Aristolochic acid nephropathy, 2357, 2377, 2442, 2444
Armodafinil, 84, 167, 219, 3588
Arnica, 2137
ARNT1 gene mutations, 3113
Aromatase deficiency, 3134–3135
Aromatase inhibitors
  actions of, 536
  adverse effects of, 364, 567, 637, 753t, 2940t, 2974, 3296
  for breast cancer prevention, 513
  for breast cancer treatment, 536, 567, 637, 639
  for endometriosis, 3137
  for male precocious puberty, 3109, 3110
Aromatic amines, 512t, 2240
Aromatic L-amino acid decarboxylase, 2790t
Arousal level, in coma, 189–190
Arousal system, 210, 211f
ARR (absolute risk reduction), 28
ARR (aldosterone-renin-ratio), 3062t
Arrhythmias. *See* Cardiac arrhythmias
Arrhythmogenic right ventricular cardiomyopathy (ARVC), 1961f, 1966, 2004t, 2009, 2010
Arsenic exposure/poisoning
  bladder cancer and, 691
  carcinogenicity of, 512t, 3699
  clinical features of, 3484, 3611, 3699, 3700t
  diagnosis of, 3700t
  diarrhea in, 306, 307
  global considerations in, 3701
  hyperpigmentation in, 400
  neuropathies in, 3610t, 3611
  optic neuropathy in, 228
  pathophysiology of, 3700t
  sources of, 3700t
  treatment of, 3700t
Arsenic trioxide, 570t, 573, 591, 834
Arteether, 1741t
Artemether
  actions of, 1746
  adverse effects of, 1741t, 1771t
  for malaria, 1768, 1770t, 1771t
  pharmacology of, 1746, 1771t
  in pregnancy and lactation, 1741t
  resistance to, 1768
Artemisinin derivatives, 1741t, 1746. *See also* Artemether; Artesunate
Arterial blood gases
  abnormal, approach to, 368–369
  in hypothermia, 3770
  in respiratory disease evaluation, 2201, 2207
  in shock, 2313
Arterial diseases, of extremities
  acrocyanosis, 1859, 2177f, **2178**
  acute limb ischemia, 2175
  arteriovenous fistula. *See* Arteriovenous fistula
  atheroembolism, 2175, 2176f
  erythromelalgia (erythermalgia), 819, 824, 2177f, **2179**
  fibromuscular dysplasia, 2150, 2174, 3447
  frostbite. *See* Frostbite
  livedo reticularis, 91, 408, 2177f, **2178,** 2837
  peripheral artery disease. *See* Peripheral artery disease
  pernio (chilblains), 2177f, **2178,** 3771
  popliteal artery aneurysm, 2176
  popliteal artery entrapment, 2176
  Raynaud's phenomenon. *See* Raynaud's phenomenon
  thoracic outlet compression syndrome, 2175–2176
  thromboangiitis obliterans, 2174–2175, 2178
Arterial hypoxemia, 277, 2307
Arterial oxygen content (CaO_2), 2310
Arterial spin labeling, in MRI, 3390, 3391f
Arterial switch procedure, 1875, 2075t, 2076
Arterial thoracic outlet syndrome, 133
Arterial thrombosis, **937**
  in acute limb ischemia, 2175
  in antiphospholipid syndrome, 2837, 2838t
  genetic considerations in, 939, 939t
  kidney, 2153
  lupus anticoagulant in, 936
  mesenteric, 2588, 2588t
  platelets in, 937–938, 937f, 941
  risk factors for, 938–939, 941
  treatment of, S13
  vascular disease and, 937
  vs. venous thrombosis, 941
Arterial tortuosity syndrome, 3331
Arterial volume, effective, reduction in, 279
Arteries. *See* Blood vessels
Arteriogenesis, 1847
Arteriolonephrosclerosis, 2154–2155
Arteriolosclerosis, 3489
Arteriovenous fistula
  acquired, 2176
  clinical features of, 406, 2176
  congenital, 2176
  diagnosis of, 2176
  for dialysis access, 2399, 2403–2404, 2403f
  dural, 3453t, 3457
  pulmonary, cyanosis in, 278
  treatment of, 2176
Arteriovenous graft, 2403
Arteriovenous malformation (AVM)
  cerebral, 3394, 3453t, 3456–3457, A16
  cervical, 131
  hemoptysis in, 275
  in inherited vessel wall disorders, 927
  in Osler-Weber-Rendu syndrome, 3456
  spinal cord, 3567, 3567f
Arteritis
  cutaneous, 2906
  giant cell (temporal). *See* Giant cell arteritis (temporal arteritis)
  necrotizing, 3447
  obliterative, 2932
  Takayasu's. *See* Takayasu's arteritis
Arterolane, 1771t
Artery of Percheron, occlusion of, 3431, A16
Artery-to-artery embolic stroke, 3445. *See also* Ischemic stroke

- Artesunate
  - actions of, 1746
  - adverse effects of, 1771t
  - drug interactions of, 1741t
  - for malaria, 991t, 994, 1768, 1770t, 1771t
  - pharmacology of, 1746, 1771t
  - in pregnancy and lactation, 1741t
  - resistance to, 1768
- Arthritis/arthropathy. *See also* Osteoarthritis (OA)
  - in acromegaly, 2966
  - bacterial. *See* Infectious (septic) arthritis
  - clinical features of
    - in familial Mediterranean fever, 2934
    - gonococcal. *See* Gonococcal infections, disseminated
    - in HIV infection, 1603
    - hypertrophic. *See* Hypertrophic osteoarthropathy
    - infectious. *See* Infectious (septic) arthritis
    - in Lyme disease, 1450
    - monoarticular, 1055t
    - polyarticular, 1055t
    - postinfectious, 1060. *See also* Reactive arthritis
    - psoriatic. *See* Psoriatic arthritis
    - in Sjögren's disease, 2877, 2877t
  - in complex regional pain syndrome, 96, 2971
  - costochondritis and, 2971
  - crystal-induced, **2956**
    - calcium apatite deposition disease, 2961–2962, 2961t, 2962f
    - calcium oxalate deposition disease, 2962, 2962f
    - calcium pyrophosphate deposition disease, 2889, 2960–2961, 2960t
    - clinical features of, 2956t
    - gout. *See* Gout
  - differential diagnosis of, 1055t
  - in hemochromatosis, 2966, 3336
  - in hemophilia, 2967
  - in hyperlipidemia, 2968
  - in hypertrophic osteoarthropathy. *See* Hypertrophic osteoarthropathy
  - in IBD, 2562t, 2563
  - in myofascial pain syndrome, 2971–2972
  - neoplasias and, 2972, 2972f
  - neuropathic joint (Charcot joint/foot) disease, 1058, 1068, 1068f, 2968–2969, 2969f, 2969t
  - ochronotic, 3371
  - paraneoplastic, 2972
  - with rash, 1666t, 1669–1670
  - in sarcoidosis, 2927
  - in sickle cell disease, 2967–2968, 2967t
  - in SLE, 2829, 2830f
  - in thalassemia, 2968
  - in Tietze syndrome, 2971
- Arthroconidia, 1696
- Arthroplasty, 2956
- Arthropod-borne viruses, **1655**. *See also specific viruses*
  - clinical syndromes of
    - arthritis and rash, 1666t, 1669–1670
    - encephalitis, 1666t, 1670–1674
    - fever and myalgia, 1665–1669, 1666t
    - hemorrhagic fevers. *See* Viral hemorrhagic fevers
    - prevention in international travelers, 1018
    - pulmonary disease, 1666t, 1674
  - diagnosis of, 1663
  - epidemiology of, 1663
  - geographic distribution of, 1664–1665t
  - reservoirs and vectors of, 1657–1663t
  - transmission of, 1656
- Arthropod injuries, 405, 3728. *See also* specific organisms
- Arthroscopy, 2956
- Artificial intelligence/deep learning, 2213, 2213f, 3970. *See also* Machine learning
  - in cardiac imaging, 1900–1901
  - decision support systems and, 26
  - ethical issues in, 74
- Artificial neuron, 3971–3972, 3972f
- Artificial saliva, 243
- Artificial tears, 224
- Art of medicine, 1
- ARVC (arrhythmogenic right ventricular cardiomyopathy), 1961f, 1966, 2004t, 2009, 2010
- Arylsulfatase A, 3827
- Arylsulfatase B deficiency (Maroteaux-Lamy disease), 3352t, 3357
- 5-ASA (5-aminosalicylic acid) agents, 2564, 2564t
- Asacol, 2564, 2564t
- Asbestos, 512t, 2235, 2236t
- Asbestos-related disease
  - asbestosis, 2235–2236, 2236t, 2237f
  - chest imaging in, 2237, 2237f
  - CT in, A12
  - lung cancer, 615, 2237. *See also* Lung cancer
  - mesothelioma, 2236. *See also* Mesothelioma
  - pleural effusion in, 2236
  - risk of, 2235–2236
  - soft tissue sarcomas, 727
- *Ascaris lumbricoides*/ascariasis, **1814**
  - clinical features of, 1739t, 1815, 1815t, 2234
  - diagnosis of, 961t, 1815, 2737, S12
  - epidemiology of, 961t, 1736, 1814, S12
  - life cycle of parasite, 1736, 1814, S12
  - treatment of, 1751, 1815–1816, 1815t
- Ascending arousal system, 210, 211f
- Asciminib
  - action and targets of, 533t, 568t
  - adverse effects of, 568t
  - for CML, 838–841, 839t
  - genetic variations in response to, 567
- Ascites, **327**
  - in cardiovascular disease, 1860
  - chylous, 328
  - diagnosis of, 328–329, 329f, 3912t, 3914v
  - differential diagnosis of, 1986t
  - etiology of, 326, 328
  - in heart failure, 1984
  - likelihood ratio of, 48t
  - in liver disease, **2715**
    - clinical features of, 2715
    - complications of, 329–330, 2716
    - definition of, 281
    - diagnosis of, 2629–2630, 2715
    - in HCC, 663
    - with jaundice, 324
    - pathogenesis of, 327–328, 2715, 2715f
    - peritonitis and, 2716
    - refractory, 2715, 2716f
    - treatment of, 329, 2715, 2716f
  - malignant, 329
  - in ovarian cancer, 710
  - in pancreatic disease, 328, 2739t, 2740
  - pathogenesis of, 327–328, 329f
  - peritonitis and, 328, 329–330, 1070
- *ASCL1* gene mutations, 825
- Ascorbic acid. *See* Vitamin C
- ASD. *See* Atrial septal defect (ASD)
- Asenapine, 3673, 3673t
- Aseptic necrosis of femoral head, 3355
- Asexual, 3179t
- Asfotase alfa, 3316
- Asherman's syndrome, 3133
- Ashkenazi Jewish population
  - *BRCA* gene testing in, 525, 633
  - dystonia in, 3510
  - factor XI deficiency in, 932
  - Gaucher disease in, 3356. *See also* Gaucher disease
  - genetic screening in, 3804
  - hearing loss in, 248
  - PD in, 3496
  - Tay-Sachs disease in, 3355. *See also* Tay-Sachs disease
  - unstable hemoglobins in, 780
- Ash leaf spot, 382, 397
- Asialoglycoprotein receptor, 2790t
- Asian longhorned tick, 3731
- Asimultanagnosia, 3432
- Askin's tumor, 730
- Asparaginase
  - actions of, 565
  - adverse effects of, 562t, 565, 597, 726t, 2703t
  - for ALL, 847, 847f, 848
  - interactions and issues, 562t
- Asparagine synthetase deficiency, 3369t
- Aspartate aminotransferase (AST)
  - in acute viral hepatitis, 2656
  - in alcohol-associated liver disease, 2698
  - in autoimmune hepatitis, 2695
  - in chronic HBV infection, 2675
  - in jaundice evaluation, 324
  - in liver function evaluation, 2635, 2637t
- Aspartylglucosaminuria, 3353t
- Aspergilloma, 1390, 1715, 1717t
- Aspergillosis, **1713**
  - after TB, 1390
  - in asthma, 2221
  - in cancer patient, 579, 582, 583
  - clinical features of, 1688t, **1713**
    - aspergilloma, 1390, 1715
    - brain abscess, 583
    - cerebral, 1714, 1714t
    - chronic meningitis, 1130t
    - cutaneous, 148, 1714, 1714t, A1, A5
    - endocarditis, 1714, 1714t
    - keratitis, 1715
    - ocular, 235
    - otitis externa, 1715
    - pneumonia, 582
    - sinusitis, 256, 257, 582
  - in cystic fibrosis, 2248
  - diagnosis of, 1713t, 1715–1716, 1715–1716f
  - disseminated, 1714, A1, A16
  - epidemiology of, 1713, 1713t
  - genetic considerations in, 1713
  - global considerations in, 1713, 1713t
  - in HCT recipient, 1155–1156, 1162, 1162f, A1
  - in HIV infection, 1597
  - immune response to, 1689, 1690f
  - pathogenesis of, 1713
  - prognosis of, 1717
  - prophylaxis for, 1155–1156, 1717, 1717t
  - respiratory tract
    - allergic bronchopulmonary. *See* Allergic bronchopulmonary aspergillosis (ABPA)
    - allergic sinusitis, 1715
    - in asthma, 1715
    - bronchitis, 1714
    - chronic pulmonary disease, 1714–1715, 1715f, 1717t
    - chronic sinusitis, 1715
    - hemoptysis in, 274
    - invasive pulmonary disease, 1713–1714, 1714t
    - invasive sinusitis, 1714
    - pulmonary hemorrhage, 592
    - tracheobronchitis, 1714
  - risk factors for, 1688t, 1713
  - in solid organ transplant recipient, 1158, 1162, 1164t
  - in transplant recipient, 2350, 2410

  - treatment of, 1716–1717, 1717t
  - in war veterans, S7
- *Aspergillus* spp.
  - antifungal susceptibility testing of, S11
  - ecology of, 1713
  - epidemiology of, 1713
  - laboratory identification of, 1688t, 1713t, 1715–1716
- Asphyxiant/blood agents, S5. *See also* Cyanide poisoning
- Aspiration
  - cough in, 272
  - in dysphagia, 296
  - in enteral nutrition, 2623
  - in opioid-related disorders, 3690
  - pneumonia and, 68, 1375
  - pneumonitis and. *See* Pneumonitis
  - of vomitus, 298
- Aspirin, **942**
  - actions of, 942–943, 943f
  - for acute pericarditis, 2079
  - adverse effects of
    - anaphylaxis, 489
    - anemia, 769
    - bleeding, 473, 943
    - bronchospasm, 2221
    - cutaneous, 419, 422
    - desensitization protocol for, 2819
    - eosinophilia, 462
    - gastric irritation/bleeding, 96, 943, 2102
    - gout, 2940t
    - hepatotoxicity, 2668
    - immunologic factors in, 419
    - nephropathy, 2442
    - Reye's syndrome, 1505, 1507
    - steatosis, 2703t
  - allergy to, 943
  - for antiphospholipid syndrome, 2839
  - for cardiac catheterization pretreatment, 1904
  - chronic use in cardiac disease, 3445t
  - with clopidogrel, 943, 3449
  - for colorectal cancer prevention, 513, 655
  - for coronary heart disease prevention, 40t, 3228
  - dosage of, 943, S13
  - drug interactions of, 3213
  - for erythromelalgia, 2179
  - for fever, 136
  - genetic variation in response to, 939, 939t
  - for giant cell arteritis, 2903
  - indications for, 943
  - for ischemic heart disease, 2102, S13
  - for Kawasaki's disease, 2907
  - for leprosy reactions, 1412
  - low-dose, 266, 485, 769
  - management of
    - before endoscopic procedures, 2483t
    - before lumbar puncture, S3
    - perioperative, 3908
  - for medication-overuse headache, 116
  - metabolism of, 485
  - for migraine, 3467t
  - for myocardial infarction
    - NSTE-ACS, 2110, 2110t, 2112
    - as secondary prevention, 2126
    - STEMI, 2118, 2121
  - for niacin-associated flushing, 2608
  - for pain, 96–98, 97t
  - in PCI, 2104, 2127, 2128
  - for peripheral artery disease, 2173
  - for preeclampsia prevention in high-risk women, 3899
  - resistance to, 943
  - for rheumatic fever, 2857
  - for stroke/TIA prevention, 3445t, 3449–3450, S13
  - for stroke treatment, 3442–3443
  - for subacute thyroiditis, 3038
  - for tension-type headache, 3471
  - for VTE prophylaxis, 2164t, S13
  - in women, 3164
- Aspirin-exacerbated respiratory disease, 2227, 2819
- Asplenia, 482, 1004t, 2795–2796, S9
- Assigned female/male at birth, 3179t
- Assist control (AC) ventilation, 2305–2306, 2305t. *See also* Mechanical ventilation
- Assisted circulation, prolonged, 2030–2033, 2031–2032f
- Assisted reproductive technologies, 3107, 3115, 3151–3152, 3151t, 3816–3817, 3817f, S13
- Association cortex, 200, 200f
- Association study, 3797t
- Associative agnosia, 207
- Associative prosopagnosia, 207
- AST. *See* Aspartate aminotransferase (AST)
- Astasia-abasia, 179
- Asteatotic eczema, 380f, 385. *See also* Eczema
- Astereognosis, 175
- Asterixis, 179, 189, 2716
- Asthenia, 3634
- Asthma, **2215**
  - approach to the patient, 2226
  - aspirin-exacerbated, 2227
  - cardiac, 1982
  - clinical features of, 2216
  - comorbidities in, 2221–2222, 2222t
  - cough in, 272
  - cough-variant, 273
  - development pathway of, 2216, 2216f, 2216t
  - diagnosis of, 2222–2223
  - differential diagnosis of, 2222t
  - dyspnea in, 269t, 270, 271
  - epidemiology of, 2216
  - etiology of, **2219**
    - airway narrowing triggers in, 2217t
    - genetic considerations in, 2219–2220
    - risk factors and exposures, 2216t, 2220–2221, 3739, 3740f
    - triggers of airway narrowing, 2221
  - exercise-induced symptoms in, 2227
  - with fungal sensitization, 1715
  - high-altitude travel and, 3759
  - high-risk patients, 2227, 2227t
  - microbiome and, 3838
  - in older adults, 2228
  - overlap with COPD, 2228
  - pathophysiology of, 2203, **2216,** 2216f, 2216t, 2219f
  - in pregnancy, 2227
  - pulmonary function values in, 2206, 2206f
  - respiratory acidosis in, 375
  - severe, 2227–2228
  - systems biology applied to, 3961t
  - treatment of, **2223**
    - alternative therapies, 2226
    - bronchial thermoplasty, 2226, 2289
    - bronchodilators, 2223–2224
    - controller therapies, 2224–2226
    - in exacerbations, 2227
    - goals of, 2223, 2223t
    - step therapy for, 2225t, 2226
    - therapies in development, 2226
    - trigger reduction, 2223
- Astigmatism, 220
- Astrocytes, 3396–3397
- Astrocytomas
  - circumscribed, 720
  - IDH-mutant, 718, 718f
  - intramedullary, 3563, 3563f
  - pilocytic, 720, A16
- Astroviruses, 1628f, 1628t, 1632
- *ASXL1* gene mutations, 761
  - in AML, 828
  - in aplastic anemia, 812
  - in chronic eosinophilic leukemia, 876
  - in chronic myelomonocytic leukemia, 877
  - MDS, 816
  - in primary myelofibrosis, 823
- Asymmetric paternalism, 3915
- Asymptomatic bacteriuria, 1086, 1089, 1092
- Asystole, 2334, 2337f. *See also* Cardiac arrest
- AT. *See* Atrial tachycardia (AT)
- ATAC-seq, 3937, 3938f
- Ataxia, **3533**
  - approach to the patient, 3533–3534
  - autosomal dominant
    - dentatorubropallidoluysian atrophy, 3536, 3793t
    - episodic ataxia, 3536
    - spinocerebellar ataxia. *See* Spinocerebellar ataxia (SCA)
  - autosomal recessive. *See* Ataxia telangiectasia (AT); Friedreich's ataxia
  - balance in, 180
  - chemotherapy-induced, 726t, 3533
  - clinical features of, 3533–3534, V1
  - etiology of, 3533–3534, 3534t
  - focal, 3533–3534, 3534t
  - gait disorders in, 177t, 178–179, 179t
  - global considerations in, 3537
  - mitochondrial, 3537
  - in MS, 3578, 3588
  - optic, 206
  - paraneoplastic, 3533
  - pathophysiology of, 3957, 3961t
  - sensory, 172, 177t, 179
  - symmetric, 3533, 3534t
  - treatment of, 3537
  - video library, V1
  - with vitamin B deficiency, 3533
  - with vitamin E deficiency, 3537
- Ataxia telangiectasia (AT)
  - cancer risk in, 755
  - clinical features of, 2799, 3537
  - diagnosis of, 2800
  - genetic considerations in
    - *ATM* gene mutations, 524t, 3537, 3785
    - DNA repair defects, 3537, 3780, 3785, S9
  - incidence of, 2799
  - lymphoid malignancies in, 858, 858t, 2800, 3537, S9
  - pathology of, 3537
  - treatment of, 3537
- *Ataxin-1,* 3534
- *Ataxin-2,* 3523, 3535
- Ataxins, 3534
- Atazanavir
  - adverse effects of, 1601, 1617t, 2640
  - drug interactions of, 2525
  - genetic variations in response to, 493, E72t
  - for HIV infection, 1617t
  - molecular structure of, 1619f
  - resistance to, 1623f
- Atelectasis, 2295, 3914v
- Atenolol
  - for adrenergic symptoms in Graves' disease, 3036
  - for aortic aneurysm, 2167
  - for ischemic heart disease, 2100t
  - overdosage/poisoning with, 3712t
- Atezolizumab
  - action and targets of, 534t, 559f, 560, 2787t
  - adverse effects of, 596, 666, 753t, 2351
  - for bladder cancer, 2787t
  - for HCC, 662, 663, 664, 664t, 665f, 666
  - indications for, 2787t
  - for lung cancer, 625, 625t, 629, 630, 630t
  - for soft tissue sarcomas, 729

- ATG. *See* Antithymocyte globulin (ATG)
- *ATG16L1* gene mutations, 2554t
- Atherectomy devices, 2128
- Atherectomy, orbital, A11
- Atheroembolism, 2152–2153, 2175, 2176f, 2361f, 2379t
- Atherosclerosis
  - aortic, 2165, 2170
  - cancer and, A10
  - carotid. *See* Carotid atherosclerosis
  - complications of, A10
  - coronary. *See* Coronary artery disease (CAD)
  - endothelium in, 1845
  - in familial hypercholesterolemia, 3243
  - immunopathogenesis of, 2790t
  - inflammatory mediators in, A10
  - intracranial, 3445, 3452
  - lipoprotein disorders and, 3237, 3248. *See also* Lipoprotein disorders
  - pathogenesis of, 2090–2091, 2107f, 2113–2114, A10, A10v. *See also* Ischemic heart disease; Myocardial infarction (MI)
  - peripheral. *See* Peripheral artery disease
  - premature, 2837
  - renal. *See* Renovascular disease
  - in rheumatologic disease, A10
  - risk factor modification for, A10
  - risk factors for, 3448–3449
  - in SLE, 2836
- Athetosis, 3509t
- Athlete(s)
  - androgen abuse in. *See* Anabolic-androgenic steroid (AAS) abuse
  - anemia in, 768
  - cardiac hypertrophy in, 2007
  - palpitations in, 291
  - traumatic brain injury in, 3574–3575. *See also* Chronic traumatic encephalopathy (CTE)
- Atidarsagene autotemcel, 3824t
- ATL. *See* Adult T-cell leukemia/lymphoma (ATL)
- *Atlastin* gene mutations, 3522t, 3525
- ATM, 530
- ATMDS (α thalassemia with myelodysplasia), 778t
- *ATM* gene mutations
  - actions of, 530, 530f, 3785
  - in ataxia-telangiectasia, 524t, 3537, 3785
  - breast cancer risk and, 524t, 633
  - in CLL, 851
  - in HCC, 661t
  - in pancreatic cancer, 671t
  - in T-cell prolymphocytic leukemia, 874
- ATN (acute tubular necrosis). *See* Acute kidney injury (AKI)
- ATNAA (Antidote Treatment Nerve Agent Autoinjector), S5
- Atogepant, 3469, 3470t
- Atomoxetine, 160, 3547
- Atonic bladder, 3588
- Atonic seizures, 3410
- Atopic conjunctivitis, 224
- Atopic dermatitis, **382**
  - allergic contact dermatitis and, 384
  - clinical features of, 383t, 393t
    - in children, 383
    - distribution of skin lesions, 378t, 380f, 383, 393t
    - in infants, 383
    - lichenification and scaling, 383f, A5
    - morphology of skin lesions, 378t, 393t
    - pruritus, 383
  - diagnosis of, 393t
  - etiology of, 382–383
  - genetic considerations in, 382–383
  - immunopathology of, 2790t
  - microbiome and, 3838
  - treatment of, 383–384, 393t, 436
- *Atopobium vaginae,* 1102
- Atopy, 2220, 2804
- Atorvastatin, 2110, 2940t, 3228, 3244t, 3257. *See also* Statins
- Atovaquone
  - actions of, 1746
  - adverse effects of, 1734t, 1742t, 1771t
  - for *Babesia* infection, 991t, 1779–1780, 1780t
  - drug interactions of, 1742t
  - for malaria, 1771t
  - for malaria prophylaxis, 1774t, 1775
  - for PCP, 1734, 1734t
  - for PCP prophylaxis, 1593t, 1734t, 1735
  - pharmacology of, 1746–1747, 1771t
  - in pregnancy and lactation, 1742t
  - resistance to, 1780
  - for toxoplasmosis, 1157
  - for toxoplasmosis prophylaxis, 1594t
- *ATP1A1* gene mutations, 355
- *ATP1A2* gene mutations, 3462
- *ATP1A3* gene mutations, 3510t, 3511
- *ATP2B3* gene mutations, 355
- *ATP7B* gene mutations, 3338, 3339, 3518
- ATP citrate lyase inhibitor. *See* Bempedoic acid
- *ATR-16* gene mutations, 778t
- ATRA. *See* All-*trans*-retinoic acid (ATRA)
- *Atrax robustus,* 3734
- Atresia
  - biliary, 2718, 2735
  - esophageal, 2507
- Atrial fibrillation (AF), **1947**
  - in AL amyloidosis, 899
  - in cardiac amyloidosis, 2027
  - central sleep apnea and, 2282
  - clinical features of, 1947–1949
  - CT pulmonary valve mapping in, A9
  - early onset, 3788t
  - ECG in, 1932t, 1948f, A8
  - echocardiography in, 1915
  - epidemiology of, 1914f, 1947
  - in heart failure, 1998
  - in hypertrophic cardiomyopathy, 2008
  - lifestyle impact on, 1954
  - mechanisms of, 1913, 1914f
  - in mitral stenosis, 2048
  - palpitations in, 291
  - paroxysmal, 1947, 1948t
  - pathophysiology of, 1848f, 1947, 1948t
  - persistent, 1947, 1948t
  - preexcited, 1941, 1942f
  - in pulmonary edema, 2331
  - risk factors for, 1947, 1954
  - in SA node dysfunction, 1919
  - stroke and, 3444, 3446t, 3450
  - subclinical, 1950
  - in thyrotoxicosis, 3033
  - treatment of, 1949
    - cardioversion and anticoagulation, 1949
    - rate control, 1949–1950
    - rhythm control, 1951–1954, 1952–1953f, 1954t
    - for stroke prevention, 1950–1951, 1950f, 1951t, 3445t, S13
- Atrial flutter, 1932t, 1943–1946, 1944–1946f, A8
- Atrial natriuretic peptide (ANP)
  - actions of, 2367f, 2370
  - in edema, 280–281
  - in heart failure, 1980
  - in hyporeninemic hypoaldosteronism, 359
  - in lung cancer, 619
- Atrial septal defect (ASD)
  - AV conduction block in, 1927
  - clinical features of, 277, A11
  - diagnosis of
    - cardiac catheterization, 1907
    - ECG, 2070, A7
    - echocardiography, 1900, 1901f, A11
    - heart sounds/murmur, 283, 285–286, 288, 1864, 1865f, 2070
  - locations of, 2070–2071, 2070f, A11
  - vs. mitral stenosis, 2050
  - mitral valve defects in, 2071
  - pathophysiology of, 2070, 2070f
  - primum, 286, 2070, 2070f, A11
  - secundum, 285, 2070, 2070f, A11
  - sinus venosus, 286, 2071
  - treatment of, 2130, 2130f, A11, A11v
- Atrial switch procedure, 2074–2075, 2075f, 2075t
- Atrial tachycardia (AT)
  - ECG in, A8
  - focal, 1937–1938, 1937f, 1938f
  - intermittent, 1936f
  - macroreentrant, 1943, 1945f
  - mechanism of, 1913
  - multifocal, 1946, 1946f, A8
  - treatment of, 1946
- Atriofascicular pathways, 1941
- Atrioventricular (AV) bundle, 1927
- Atrioventricular (AV) canal defects, 2069, 2072, 2072f
- Atrioventricular (AV) conduction block, **1925**
  - after STEMI, 2125
  - in cardiac amyloidosis, 2027
  - in cardiac mesothelioma, 2086
  - classification of, 1925, 1925t, 1926f
  - CMR in, 1897f
  - complete (third-degree), 1925, 1925t, 1926f
  - diagnosis of, 1927–1929, 1928–1929f, 1933f
  - etiology of, 1913t, 1925–1927, 1926t, 1927f
  - first-degree, 1925, 1925t, 1926f, A7
  - in Lyme disease, 1450, 1453, 1926–1927, A8
  - second-degree, 1925, 1925t, 1926f, A8
  - treatment of, 1929–1931, 1930f. *See also* Pacemaker(s)
- Atrioventricular (AV) nodal reentry tachycardia (AVNRT), 1932t, 1937f, 1939–1940, 1939f, A8
- Atrioventricular (AV) node, 1912, 1924–1925
- Atrophic vaginitis, 3161–3162
- Atrophie blanche (livedoid vasculopathy), 408, 2180
- Atrophie blanche en plaque, 2178. *See also* Livedo reticularis
- *Atrophin* gene, 3536
- Atrophy, 377t
- Atrophy, muscle, 3636–3637
- Atropine
  - adverse effects of, 291
  - in AV conduction block diagnosis, 1927
  - for AV conduction block treatment, 1927, 1929
  - for chemotherapy-related diarrhea, 564
  - for ciguatera poisoning, 3727
  - for nerve agent exposure, 3712t, S5
  - overdosage/poisoning with, 3711t
  - for SA node dysfunction, 1920
  - for snakebite treatment, 3721r
- *ATRX* gene mutations, 778t
- Attachment, bacterial, 965–967, 966f, 966t
- AT/*tat,* 1561f
- Attentional deficit, 208
- Attention, assessment of, 183
- ATTR amyloidosis, 895, 896t, 899–900. *See also* Amyloidosis
- Atypical adenomatous hyperplasia, 622
- Atypical chronic myeloid leukemia (aCML), 877t, 878
- Atypical diabetes, 3196
- Atypical hemolytic uremic syndrome (aHUS), 800–801, 925, 2412
- Atypical lipomatous tumor, 727. *See also* Soft tissue sarcomas
- AUA Symptom Index, 703t

- AUC (area under the receiver operating curve), 3972, 3973t
- AUC/MIC, 1168, 1168f
- Audiogram, 249
- Audiometry, 163, 248–249
- AUDIT (Alcohol Use Disorder Screening Test), 3679, 3679t
- Auditory brainstem responses, 248
- Auditory cortex, 200f
- Auenbrugger, Leopold, 46
- Auer rod, 826, 828, 829f
- Augmented reality, 3964, 3965f
- Aura, 3409, 3462
- Auscultation
  - abdominal, 327, 479
  - cardiac. *See* Cardiac auscultation
  - pulmonary, 270, 2200
- Austin Flint murmur, 1844, 1866
- Australian bat lyssaviruses, 1655
- Australian blue-ringed octopus bite, 3724
- Australian jellyfish sting, 3722
- Autism spectrum disorders, 3398–3399, 3403–3404, 3651–3652, 3653f, 3701
- Autoantibody(ies)
  - autoantigens and, 2790–2791t
  - in autoimmune diseases, 2780, 2824–2826, 2825t
  - drug-induced, 489
  - islet cells, 3200
  - in SLE, 2832t
  - in systemic sclerosis, 2864t, 2871
- Autocrine regulation/signaling, 531, 2987
- Autoimmune autonomic ganglionopathy (AAG), 3550
- Autoimmune diseases, **2821,** 2822t. *See also specific diseases*
  - after HCT, 918
  - aging and, 2823
  - autoantigens associated with, 2780, 2790–2791t
  - as cause of fever of unknown origin, 151t
  - cutaneous, **409**
  - definition of, 2755
  - environmental factors in, 2822–2823
  - epidemiology of, 2822
  - fever in, 136
  - genetic and epigenetic factors in, 2822, 2823t
  - in HIV infection, 1576
  - IC/BPS as, 331
  - immunopathogenesis of, 2789
  - lymphoid malignancies in, 858t, 859, 862
  - microbiome and, 3837–3838
  - myasthenia gravis and, 3628
  - prevention and treatment of, 2827
  - sex and, 2823
  - Sjögren's disease associated with, 2876
  - treatment of, 2783–2784, 2791–2793t
  - in women, 3165–3166
- Autoimmune encephalitis, 1114
- Autoimmune epithelitis, 2877
- Autoimmune hemolytic anemia
  - in CLL, 853
  - cold type, 802–803, 803f
  - diagnosis of, 449
  - in EBV infections, 1510
  - folate deficiency in, 788
  - pathophysiology of, 803f
  - warm type, 801–802, 803f
- Autoimmune hepatitis. *See* Hepatitis, autoimmune
- Autoimmune hypothyroidism. *See* Hashimoto's thyroiditis
- Autoimmune lymphoproliferative syndrome (ALPS), 524t, 2803
- Autoimmune pancreatitis, 2751–2752, 2751t, 2752f, 2930, 2931t
- Autoimmune polyendocrine (polyglandular) syndrome-1 (APS-1), **3089**
  - vs. APS-2, 3089t
  - autoantigens in, 2790t, 3089
  - clinical features of, 1708, 2804, 3066, 3067t, 3089–3090, 3090t, 3287
  - diagnosis of, 3090, 3090t
  - diseases associated with, 3089t, 3090t
  - etiology of, 3066
  - genetic considerations in, 3066, 3067t, 3089
  - global considerations in, 3093
  - treatment of, 3090
- Autoimmune polyendocrine (polyglandular) syndrome-2 (APS-2)
  - vs. APS-1, 3089t
  - autoantigens in, 3091t
  - clinical features of, 3030, 3066, 3067t, 3090t
  - diagnosis of, 3090t, 3092
  - diseases associated with, 3089t, 3090t, 3091–3092
  - genetic considerations in, 3066, 3067t, 3090–3091, 3091t
  - treatment of, 3092
- Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED). *See* Autoimmune polyendocrine (polyglandular) syndrome-1 (APS-1)
- Autoimmunity
  - cell- or organ-specific, 2790t
  - cytokine, 2790t
  - definition of, 2821–2822
  - drug-induced, 2827
  - humoral, 2864
  - mechanisms of, 2780, 2826
  - in MS, 3582
  - paraneoplastic, 2790–2791t
  - plasma protein, 2790t
  - primary immune deficiency diseases and, 2803–2804
  - systemic, 2790t
  - in systemic sclerosis, 2863–2864, 2864t
- Autoinflammatory diseases. *See also specific diseases*
  - definition of, 2755, 2934
  - fever in, 136
  - genetic considerations in, 2760, 2761–2762t
  - hereditary, 2936–2937. *See also* Hereditary recurrent fever syndromes
- Autoinflammatory infantile fever with enterocolitis (AIFEC), 2761t
- Autolysins, 1165
- Automatic external defibrillation (AED), 2336
- Automaticity, cardiac, 1913, 1913t
- Automatisms, 3409
- Autonomic dysfunction, **3538**
  - approach to the patient, 3538–3540
  - classification of, 3541t
  - clinical features of, 3538–3540
  - conditions of, **3546**
    - in diabetes mellitus, 3226, 3233, 3233f, 3549–3550
    - in HSV infection, 1500
    - in multiple system atrophy. *See* Multiple system atrophy (MSA)
    - orthostatic hypotension. *See* Orthostatic hypotension
    - paraneoplastic, 746, 751
    - in peripheral nerve and neuromuscular junction disorders, 3549–3550
    - postural orthostatic tachycardia syndrome (POTS), 1936–1937, 3548
    - pure autonomic failure, 3548
    - with spinal cord lesions, 3549
    - synucleinopathies. *See* α-synuclein
    - in tetanus, 1231
  - diagnosis of
    - heart rate variation to deep breathing, 3541, 3542f
    - laboratory evaluation, 3540–3541
    - skin biopsy, 3542, 3546
    - sudomotor function testing, 3542, 3545–3546f
    - tilt table testing, 3542, 3544–3545f
    - Valsalva maneuver, 3541–3542, 3543f
  - drug-related, 3540
  - in Guillain-Barré syndrome, 3617
  - localization of, 3540
  - peripheral, 159
- Autonomic dysreflexia, 3549
- Autonomic hyperreflexia, 3570
- Autonomic nervous system
  - anatomy of, 3538, 3539f
  - disorders of. *See* Autonomic dysfunction
  - functions of, 3538
  - parasympathetic effects of, 3540t
  - sympathetic effects of, 3540t
- Auto-PEEP, 2296, 2297f
- Autophagy
  - activation of, 3875
  - bacterial interference with, 973
  - in cancer cells, 541
  - definition of, 2755
  - impaired, aging and, 3871
  - in microbial infection, 972f, 973, 2783
  - in neurodegenerative diseases, 3400
- Autophagy-lysosomal pathway, 3991
- Autosomal dominant disorders, 3790, 3790f. *See also specific disorders*
- Autosomal dominant hypercholesterolemia, 3241t
- Autosomal dominant hypocalcemic hypercalciuria (ADHH), 3287
- Autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE), 3412t
- Autosomal dominant partial epilepsy with auditory features (ADPEAF), 3412t
- Autosomal recessive disorders, 3790–3791, 3790f. *See also specific disorders*
- Autosomal recessive hypercholesterolemia, 3241t, 3243
- Autosome, 3790
- Autotransporter proteins, 966, 966t
- Avacopan, 2896t, 2898
- Availability heuristic, 21
- Avalon virus, 1661t
- Avanafil, 704, 3156f, 3159, 3159t
- Avapritinib, 568t, 651, 729, 881, 2821
- Avascular necrosis of femoral head, 2968
- Avatrombopag, 924
- AV conduction block. *See* Atrioventricular (AV) conduction block
- Avelumab
  - action and targets of, 534t, 559f, 2787t
  - adverse effects of, 596, 753t, 2351
  - for bladder cancer, 694
  - indications for, 2787t
  - for Merkel cell carcinoma, 609
- Avian influenza viruses, 995, 1149, 1533, 1544. *See also* Influenza
- Avibactam, 1175, 1183, 1284. *See also* Ceftazidime-avibactam
- AVM. *See* Arteriovenous malformation (AVM)
- AVNRT (atrioventricular nodal reentry tachycardia), 1932t, 1937f, 1939–1940, 1939f, A8
- Avobenzone, 435, 436t
- Avoidant/restrictive food intake disorder, 313, 3671
- AVP. *See* Arginine vasopressin (AVP)
- *AVP-NPII* gene, 3016
- Axial osteomalacia, 3316–3317

- Axial spondyloarthritis (ankylosing spondylitis)
  - classification of, 2883, 2883t
  - clinical features of
    - aortic regurgitation, 2043
    - aortitis, 2171
    - extraskeletal, 2883–2884
    - murmur, 287, 1860
    - neck pain, 132
    - skeletal, 1860, 2883
  - complications of, 2886
  - definitions in, 2883
  - diagnosis of, 2884–2886, 2884f, 2885f
  - epidemiology of, 2883
  - genetic considerations in, 2780, 2880–2882, 3787
  - IBD and, 2562t, 2563
  - inflammasome mutations in, 2761t
  - KIRs with, 2772t
  - monitoring in, 2886
  - physical examination in, 121t
  - treatment of, 2784, 2886, 2887t
  - vs. vertebral osteomyelitis, 1062
- Axicabtagene ciloleucel, 535t, 559f, 863, 3824t
- Axillary dissection, 636
- Axillofemoral bypass, 2174
- *AXIN1* gene mutations, 661t
- Axitinib
  - actions and targets of, 533t, 548, 569t
  - adverse effects of, 569t, 753t, 754
  - interactions and issues, 569t
  - for NETs, 686
  - for RCC, 689, 690t
- Ayurvedic medicine, 3925t
- 5-Aza-2′-deoxycytidine, 538f, 539
- 5-Azacytidine
  - actions of, 570t, 573, 3940, 3942
  - adverse effects of, 420, 570t, 817
  - for AML, 830, 831, 833t
  - interactions and issues, 570t
  - for MDS, 817
  - for primary myelofibrosis, 823
- Azathioprine
  - actions of, 2408t
  - adverse effects of, 2408t, 2565, 2916t, 3631
    - bone marrow suppression, 2411, 2565
    - carcinogenicity, 512t
    - encephalopathy, 2350t
    - lymphoid malignancy, 2565
    - major, 2896t
    - neutropenia, 457
    - pancreatitis, 2565
    - prevention of, 2895
    - pulmonary, 2261
    - pure red cell aplasia, 814t
    - systemic, 2916
  - for autoimmune hepatitis, 2696
  - for autoimmune pancreatitis, 2752
  - for Behçet syndrome, 2909
  - for bullous pemphigoid, 412
  - for dermatomyositis, 414
  - drug interactions of, 489t, 2408, 3631
  - for EGPA, 2900
  - for epidermolysis bullosa acquisita, 413
  - genetic variations in response to, 493, E72t
  - for granulomatosis with polyangiitis, 2899
  - for IBD, 2565
  - for immunosuppression, 2029t, 2408, 2408t
  - for inflammatory myopathies, 2916, 2916t
  - metabolism of, 485t
  - for mucous membrane pemphigoid, 413
  - for myasthenia gravis, 3630–3631
  - for neuromyelitis optica, 3593
  - for pemphigus vulgaris, 410
  - for pericarditis, 2080
  - for photoallergy, 434
  - for sarcoidosis, 2929–2930, 2929t
  - for SLE, 2834
  - for vasculitis, 2895
- Azelaic acid, 391
- Azelastine nasal spray, 2810
- *AZF* gene mutations, 3791
- Azidothymidine. *See* Zidovudine (AZT)
- Azilsartan, 2144t
- Azithromycin
  - actions of, 1166, 1177, 1182t, 1427
  - adverse effects of, 1177, 1744t
  - for asymptomatic GAS carriage, 1209
  - for *Babesia* infections, 991t, 1779, 1780t
  - for *Bartonella* prophylaxis, 1594t
  - for bite-wound infections, 1144
  - for bronchiectasis prophylaxis, 2244
  - for *Campylobacter* infections, 1306
  - for Carrión's disease, 1353t
  - for cat-scratch disease, 1353t
  - for chancroid, 1109t, 1264
  - for cholera, 1329
  - for community-acquired pneumonia, 963t, 1027t, 1028t
  - for COPD exacerbation prop, 2256
  - for *C. pneumoniae* infection, 1477
  - for *C. trachomatis* infections, 1474, 1475
  - for donovanosis, 1358t
  - drug interactions of, 1744t
  - for endemic treponematoses, 1439
  - for enteric fever, 1317–1318, 1318t
  - for gonococcal infections, 1259, 1259t
  - indications for, 1173t, 1177
  - for infective endocarditis prophylaxis, 1048t
  - for *Legionella* spp. infections, 1275, 1275t
  - for leptospirosis, 1444, 1444t
  - for MAC infection, 1417, 1427
  - for MAC infection prophylaxis, 1417, 1593t
  - for meningitis prophylaxis, 1122
  - for *M. genitalium* infections, 1467
  - for *M. pneumoniae* infections, 1466, 1466t
  - for mucopurulent cervicitis, 1104
  - for nontyphoidal *Salmonella* infections, 1321t
  - for NTM infections, 1427
  - for pertussis, 1280t
  - for pneumococcal infections, 1194
  - in pregnancy and lactation, 1169t, 1744t
  - for proctitis, 1099
  - for psittacosis, 1476
  - resistance to, 1099, 1100f
    - in *Babesia,* 1780
    - mechanisms of, 1182t
    - in *M. genitalium,* 1467
    - in *N. gonorrhoeae,* 1255, 1259
    - in nontuberculous mycobacteria, 1427
    - in nontyphoidal *Salmonella,* 1320
    - prevalence of, 1173t
    - in *T. pallidum,* 1434
  - for scrub typhus, 1460
  - for *Shigella* infections, 1324t
  - for streptococcal pharyngitis, 1208
  - for travelers' diarrhea, 1019, 1081t
  - for tularemia, 1339
  - for urethritis, 1099, 1099t
- Azole antifungals, 1691–1692, 1742t. *See also specific drugs*
- Azoospermia, 576, 3100, 3107, 3114, 3789, 3791
- Azotemia, **337,** 339f, 2455. *See also* Acute kidney injury (AKI)
- AZT. *See* Zidovudine (AZT)
- Aztreonam
  - for community-acquired pneumonia, 1028t
  - indications for, 1173t, 1175, 1284
  - for *Proteus* spp. infections, 1291
  - resistance to, 1173t, 1181
- Azurophil granules, 453, 460f
- Azygous lobe, A12

B

- *B3GALT6* gene mutations, 3329t, 3330
- *B4GALT7* gene mutations, 3329t
- *Babesia* spp.
  - *B. divergens,* 1777, 1778, 1780
  - *B. duncani,* 1776, 1780
  - *B. microti,* 1774
  - *B. venatorum,* 1777, 1778, 1781
  - coinfections of, 990
  - geographic distribution of, 1776f, 1777, S12
  - laboratory identification of, 1778–1779, 1779f, A2
  - life-cycle hosts of, S12
  - transmission of, S12
- *Babesia* spp. infections, **1775**
  - clinical features of, 1740t, 1777–1778
  - co-infection with Lyme disease, 1452, 1778
  - complications of, 990
  - diagnosis of, 437, 1778–1779, 1779f, A2, S11, S12
  - epidemiology of, 1776–1777, 1776f, S12
  - etiology of, 990, 991t, 1775–1776
  - global considerations in, 1776f, 1777, 1778, 1781
  - in immunocompromised patient, 1777, 1780
  - neonatal, 1777
  - pathogenesis of, 1778
  - prevention of, 1781
  - risk factors for, 1777, 1781
  - in splenectomized patient, 482
  - transfusion-transmitted, 912, 913t, 1776–1777
  - transmission of, 1776–1777
  - treatment of, 991t, 1779–1781, 1780t
- Babes nodules, 1651
- Babinski sign, 168, 168t, 188, 192, 3381
- BabyBIG, 1236, 1237
- Bacillary angiomatosis, **1354**
  - clinical features of, 144t, 1355, 1355f
  - diagnosis of, 1355
  - epidemiology of, 1355
  - etiology of, 405, 1354–1355
  - in HIV infection, 405, 1355, 1355f, 1610
  - pathology of, 1355
  - prevention of, 1355–1356
  - skin manifestations of, 144t, 405, 1051
  - treatment of, 1054t, 1353t, 1355
- Bacillary peliosis, 1353t, 1355
- Bacille Calmette-Guérin (BCG) vaccine, 997, 1413, 1417
- *Bacillus anthracis,* 971, 973, S4. *See also* Anthrax
- Bacillus Calmette-Guérin (BCG), 551, 559f, 561, 692
- *Bacillus cereus* infections
  - in cancer patient, 583
  - clinical features of, 1080t
  - enterotoxin, 297, 306t, 965
  - food-borne, 1080, 1080t
- Bacitracin, 423, 1166, 1182t
- Back pain. *See also* Low back pain
  - in axial spondyloarthritis, 2883, 2884f
  - in cancer patient, 588–589, 589f
  - causes of
    - infections/inflammation, 994, 1062
    - inflammatory vs. mechanical, 2883t
    - metastatic disease, 3562
  - IC/BPS and, 332
  - in osteoporosis, 3300
  - in Paget's disease, 3313
- Back-propagation, 3973t
- Baclofen
  - for dystonia, 3512
  - for GERD, 302
  - overdosage/poisoning with, 3713t
  - for pain, 82
  - for spasticity, 3570, 3588
  - for tardive dyskinesia, 3517
  - for trigeminal neuralgia, 3553

- Bacteremia
  - *A. baumannii,* 1296
  - anaerobic bacterial, 1377–1378
  - *Bartonella* spp., 1351t, 1353t, 1354
  - clostridial, 1240
  - *E. coli,* 1286
  - *Enterobacter,* 1292
  - enterococcal, 1217–1218
  - *Klebsiella,* 1290
  - *L. monocytogenes,* 1228
  - in neutropenia, 581
  - nontyphoidal *Salmonella,* 1320, 1321t
  - *P. aeruginosa,* 1308–1310, 1309t
  - *Proteus,* 1292
  - streptococcal, 1211–1212
  - *V. vulnificus,* 1330–1331
- Bacteria
  - adherence of, 1076
  - antibiotic resistance in. *See* Antibacterial drugs, resistance to
  - colonic, 303
  - genome sequences of, 975, 984
  - gram-negative. *See* Gram-negative enteric bacilli
  - laboratory identification of, 975–977, 977t, **S11**
  - metabolites of, 3840–3841
  - susceptibility testing of, S11
  - taxonomy of, 3831, 3831f
  - toxin production by, 1076–1077
- Bacterial endocarditis. *See* Infective endocarditis
- Bacterial infections. *See also specific bacteria and bacterial types*
  - anaerobic, mixed. *See* Anaerobic bacterial infections, mixed
  - as cause of fever of unknown origin, 151t. *See also* Acute febrile illness; Fever of unknown origin (FUO)
  - in in HCT recipient, 1155t
  - pathogenesis of, **965**
    - avoidance of innate immune responses, **968,** 972f
      - antimicrobial peptides, 970
      - autophagy, 973
      - complement, 969, 970f
      - epigenetic mechanisms, 973
      - immune GTPases, 973
      - inflammasomes, 972–973
      - lysosomes, 969
      - neutrophil extracellular traps, 970
      - NF-κβ signaling, 971
      - nonpyroptotic cell death, 973
      - oxidative burst, 970
      - pattern recognition receptors, 971, 971t
    - cytotoxins, 974
    - host entry, 964–965
    - host response, 2779–2780, 2781
    - infection establishment
      - attachment, 965–967, 966f, 966t
      - cell entry, 967–968, 968f
      - niche, 965, 965t
      - replicative niche, 967–968
    - inhibition of adaptive immune responses, 974
    - survival in the cytosol, 967–968
    - survival in the vacuole, 967–968
    - tissue damage and pathogen dissemination, 974–975
    - transmission to new hosts, 974
  - in pregnancy, 3903
  - in scabies, 3729
  - in solid organ transplant recipient, 1158, 1159t
  - transfusion-transmitted, 912, 913t
  - in transplant recipient
    - after kidney transplantation, 2410, 2410t
    - after liver transplantation, 2723
  - withholding antipyretics in, 136
- Bacterial meningitis. *See* Meningitis, bacterial
- Bacterial overgrowth syndromes. *See* Small-bowel bacterial overgrowth
- Bacterial vaginosis
  - clinical features of, 1101t, 1102
  - diagnosis of, 1101t, 1102, 1102–1103f, S11
  - etiology of, 1377–1378
  - HIV transmission and, 1563
  - in pregnancy, 1103
  - treatment of, 1103
- Bactericidal drugs, 1165. *See also* Antibacterial drugs
- Bactericidal/permeability-increasing protein (Bpi), 2790t
- Bacteriophages, 974–975
- Bacteriostatic drugs, 1165. *See also* Antibacterial drugs
- Bacteriuria. *See* Urinary tract infections (UTIs)
- Bacteriuria, asymptomatic, 1086, 1089, 1092
- *Bacteroides* spp., 1372, 1378, 3837
- *Bacteroides fragilis*
  - antibiotic resistance in, 1379
  - influence on disease outcomes, 3840
  - in normal human microbiota, 1372, 3837
  - virulence factors of, 1373
- *Bacteroides fragilis* infections. *See also* Anaerobic bacterial infections, mixed
  - bacteremia, 1377
  - female genital tract, 1073, 1376
  - intraabdominal, 1073, 1376
  - liver abscess, 1074
  - pathogenesis of, 1373
  - peritonitis, 1071
  - skin and soft tissue, 1377
  - treatment of, 1173t, 1175, 1378–1379
- *BAD05,* 1701
- "Bad news" situation, communication in, 76–77, 77t
- BAERs (brainstem auditory evoked responses), 249
- BAFF (B-cell activating factor), 2792t, 2876
- *BAG3* gene mutations, 2004t
- Bagassosis, 2228t
- Baker's (popliteal) cyst, 2159, 2944
- *Balamuthia* spp., 1758, 1759, S12
- *Balamuthia* spp. infections
  - in chronic meningitis, 1131t, 1737
  - clinical features of, 1759
  - diagnosis of, 1759, S12
  - encephalitis, 1114, 1738t, 1759, 1759f
  - epidemiology of, 1758, S12
  - treatment of, 1759–1760
- Balance, 177, **179**. *See also* Fall(s)
- Balanitis, 1708
- *Balantidium coli*/balantidiasis, 1748, 1805t, 1808
- Bald tongue, 265t
- Bálint's syndrome, 205–206, 235, 3432
- Balkan nephropathy, 2377, 2442
- B-ALL (B cell lineage ALL), 845, 845t. *See also* Acute lymphoid leukemia (ALL)
- Ballism, 3515
- Balloon angioplasty
  - carotid, 3451–3452
  - cerebral, 3461, 3461f
  - coronary, 2127–2128, 2127f. *See also* Percutaneous coronary interventions (PCI)
- Balloon expulsion test, 313
- Balloon tamponade, 2714
- Balloon valvuloplasty, 2130–2131
- Balo's concentric sclerosis, 3589, 3589f, A16. *See also* Multiple sclerosis (MS)
- Baloxavir, 1488, 1488t, 1537, 1550
- Balsalazide, 2564, 2564t
- Baltic myoclonus (Unverricht-Lundborg disease), 3412t, 3516, 3530
- Bamboo spine, 2884
- Banana spider, 3734
- Band(s), in peripheral blood smear, 443
- Bangladesh, 3861f, 3865
- "Bangladesh regimen," for TB, 1421
- Bangui virus, 1658t
- Banna virus, 1662t
- Bannwarth's syndrome, 1130t, 1450
- Banzi virus, 1659t
- *BAP1* gene mutations, 688
- Barbiturates
  - adverse effects of
    - cross-sensitivity and, 427
    - cutaneous, 401, 423
    - folate deficiency, 787t
    - mitochondrial defects and, 3816
  - overdosage/poisoning with, 189, 3713t
  - for seizures, 3424
- Barcelona Clinic Liver Cancer (BCLC) Classification, 662–663, 663f
- Bar coding, of medications, 43
- BARDA (Biomedical Advanced Research and Development Authority), S4
- Bardet-Biedl syndrome
  - clinical features of, 2435, 2993, 3184–3185, 3184t
  - genetic considerations in, 2430t, 2435
  - pathophysiology of, 2431f
- Bariatric surgery
  - complications of, 3195, 3613
  - for diabetes mellitus, 3194–3195, 3216
  - effects of, 3194
  - hypoglycemia after, 3235
  - for MASLD, 2707
  - for metabolic syndrome, 3257
  - for obesity, 3194–3195
  - outcomes of, 3194–3195
  - procedures for, 3194, 3194f
  - vitamin deficiencies following, 366, 2613, 3195
- Baricitinib, 1542, 2851t, 2852, 2890, 2940t
- Barium radiography
  - in colorectal cancer, 656, 656f
  - for colorectal cancer screening, 656
  - in constipation, 312
  - in duodenal ulcer, 2523f
  - in dysphagia, 296
  - in malabsorption disorders, 2544f, 2550
  - in peptic ulcer disease, 2523, 2523f
- Bark scorpion, 3735
- Barlow's syndrome. *See* Mitral valve prolapse
- Barmah Forest virus, 1059, 1657t, 1669–1670
- Baroreceptors, 156, 157f, 2372
- Barotrauma, 2302, 2304, 2307, 3765, 3766
- Barracuda bite-wound, 1145t
- Barrel chest deformity, 1859, 1860
- Barrett's esophagus/metaplasia
  - dysplasia in, 2468f, 2510, 2511f
  - endoscopic findings in, 2462, 2468f, 2491, 2510, 2510f
  - endoscopic treatment of, V5
  - esophageal cancer and, 300, 642, 2510, 2510f
  - GERD and, 300, 2491, 2510
  - screening for, 301, 2491
  - in systemic sclerosis, 2870
- Barth's syndrome, S9
- *Bartonella* spp., 961t, 1350, 1351t
- *Bartonella* spp. infections
  - arthritis, 1056
  - bacillary angiomatosis. *See* Bacillary angiomatosis
  - bacillary peliosis, 1353t, 1355
  - Carrión's disease, 1351t, 1353t, 1356
  - cat-scratch disease. *See* Cat-scratch disease
  - endocarditis, 1038, 1040, 1044t, 1045, 1353t
  - global considerations in, 1356
  - Oroya fever, 1353t, 1356

- *Bartonella* spp. infections (*Cont.*):
  - prophylaxis for, 1594t
  - trench fever/chronic bacteremia, 1353t, 1354
  - verruga peruana, 1051, 1351t, 1353t, 1356
- Bartter's syndrome
  - antenatal, 355
  - classic, 355
  - clinical features of, 355, 2362, 2369, 3287
  - differential diagnosis of, 373
  - genetic considerations in, 355, 2368t, 2369, 3267
  - with sensorineural deafness, 2368t
  - subtypes of, 2368t, 3287
- Basal cell carcinoma (BCC)
  - clinical features of, 379t, 380f, 404, 606, 607–608f, A5
  - epidemiology of, 606
  - etiology of, 606
  - genetic considerations in, 524t
  - metastatic, 608
  - micronodular, 606
  - morpheaform (fibrosing), 606
  - natural history of, 607
  - nodular, 606
  - pathophysiology of, 431, 606
  - pigmented, 606
  - prevention of, 609
  - risk factors for, 606
  - sun exposure and, 431, 606
  - superficial, 606
  - treatment of, 430, 607–608
- Basal cell nevus syndrome. *See* Gorlin's syndrome (basal cell nevus syndrome)
- Basal ganglia, 3495f, 3500, 3500f
- Basal ganglia hemorrhage, 192
- Basal meningitis, 1128
- Bas-Congo virus, 1662t
- Basement membrane syndromes, 2414, 2415t, **2425**
  - Alport's syndrome. *See* Alport's syndrome
  - antiglomerular basement membrane disease, 275, 2412, 2415t, 2418–2419, A4
  - Goodpasture's syndrome. *See* Goodpasture's syndrome
  - nail-patella syndrome, 2415t, 2427
  - pathogenesis of, 2425
  - thin basement membrane disease, 343, 2415t, 2426, A4
- Basic fibroblast growth factor (bFGF), 546, 546f, 2516, 3004
- Basilar artery
  - infarction, 248
  - ischemia, 236
  - occlusion of, 3434–3435, 3436f
  - thrombosis, 190, 192
- Basilar skull fracture, 3572
- Basiliximab, 2029t, 2791t
- Basivertebral nerve ablation, 127–128
- Basophil(s)
  - in adaptive immunity, 2763t
  - hematopoietic differentiation of, 760f
  - in innate immunity, 2763t, 2770–2771
  - in peripheral blood smear, 437, 442f, 443
- Basophilic stippling, 437, 447t, 800, 803
- Bassen-Kornzweig disease. *See* Abetalipoproteinemia
- Batai virus, 1658t
- Bath salts (synthetic cathinones), 3696
- Batten disease, 3406
- Battlefield medicine, S8. *See also* Military veterans
- Battle sign, 3572
- Bax/*bax,* 574
- Bayes' rule (theorem), 4, 24, 25f
- *Baylisascaris procyonis* infections, 961t, 1114, 1131t, 1739t, 1812, S12
- Bayou virus, 1660t, 1674
- Bazedoxifene, 3145, 3148, 3303
- Bazex-Dupré-Christol syndrome, 606
- BCC. *See* Basal cell carcinoma (BCC)
- B cell(s)
  - activation of, 974
  - age-associated, 2825–2826
  - in autoimmune diseases, 2824–2825
  - deficiencies of, 2794t, 2801–2803, 2801f, S9
    - in agammaglobulinemia, 2801, 2801f
    - in cancer patient, 578t, 579
    - in common variable immunodeficiency, 2802
    - diagnosis of, 2801
    - in hyper-IgM syndrome, 2780, 2800, 2801f, 2802
    - in IgA deficiency, 2802
    - in selective antibody deficiency to polysaccharide antigens, 2802
  - definition of, 2755
  - development of, 2775f, 2777–2778
  - differentiation of, 760f, 859f
  - functions of, 2777–2778
  - in HIV infection, 1579–1580
  - Ig gene rearrangement and, 2777–2778
  - immune tolerance checkpoints in, 2789t
  - locations of, 2777
  - malignancies of. *See* Lymphoid malignancies, B-cell
  - in rheumatoid arthritis, 2846
  - in sepsis/septic shock, 2319
- B-cell activating factor (BAFF), 2792t, 2876
- B cell chronic lymphoid leukemia. *See* Chronic lymphocytic leukemia (CLL)
- B cell lineage ALL (B-ALL), 845, 845t, 847t. *See also* Acute lymphoid leukemia (ALL)
- B-cell prolymphocytic leukemia, 872, 872t
- B cell receptors
  - activation of, 2777, 2778f
  - for antigen, 2755
  - in CLL, 850
  - defective rearrangements of, 2799
- BCG. *See* Bacillus Calmette-Guérin (BCG)
- BCG (Bacille Calmette-Guérin) vaccine, 1401, 1413, 1417
- BCHE (pseudocholinesterase), 485t, E72t
- *BCL1* gene mutations, 521t, 860t
- Bcl-2, 539–540, 3312
- *BCL2* gene mutations, 521t, 731, 859, 860t
- BCL2 proteins, 540f, 557f, 859, 862
- BCL6, 862
- BCL11A, 776, 778
- *BCL11A* expression, 762, 3827
- BCLC (Barcelona Clinic Liver Cancer) Classification, 662–663, 663f
- Bcl-XL, 539
- BCMA, 886, 892
- BCNU (carmustine), 726t
- *BCOR* gene mutations, 828
- *BCR::ABL1* gene
  - in ALL, 844, 845–846, 847, 847t, 848, 848t
  - in CML
    - characteristics of, 834, 835f
    - as drug target, 521, 521t, 534t
    - identification of, 836, 837
    - pathophysiology of, 835–836
    - protein product of, 522f
    - TKI response and, 836, 841, E72t
  - imatinib response and, 493, E72t
- *BCR-ABL* inhibitors, 753t, 754, 757
- bDNA (branched-chain DNA) tests, 1590t, S11
- BDNF (brain-derived neurotrophic factor), 3185, 3185t, 3395, 3395f, 3398, 3942
- BEACOPP regimen, 870, 871
- Beatty test, 121t
- Beau's line, 421
- Beck Depression Inventory, 99
- Becker disease, 3647
- Becker's muscular dystrophy, 3641f, 3787–3788
- Beck's triad, 2081
- Beckwith-Wiedemann syndrome, 3792, 3939
- Bedaquiline
  - actions of, 1424–1425
  - adverse effects of, 1400, 1420t, 1425
  - dosage of, 1425
  - drug interactions of, 1425
  - for MDR-TB, 1394t, 1395t, 1398–1399, 1421, 1421t
  - for NTM infections, 1428
  - resistance to, 1425
  - for TB, 1420t, 1425
- Bed bugs, 3736
- Bed nets, insecticide-treated, 3849
- Bedside Index of Severity in Acute Pancreatitis (BISAP), 2748
- Beef tapeworm *(Taenia saginata)* infection, **1833,** S12
- Beer potomania, 348–349
- Bee sting, 3735–3736
- Beetle vesication/dermatitis, 3737
- Beevor's sign, 3560
- Behavioral economics, **3915**
  - applicability to policy, 3916
  - vs. classical economics, 3915–3916, 3916t
  - conceptual and practical considerations in, 3921–3923
  - decision errors in, 3916, 3916t
  - defaults/status quo bias in, 3918
  - framing effects in, 3916–3917
  - health care policy and, 3923–3924
  - hovering in, 3921
  - loss aversion in, 3916–3917
  - in medication adherence, 3920–3921
  - overoptimism in, 3917
  - peanuts effects in, 3917
  - present-biased preferences in, 3917–3918
  - rational world bias in, 3918–3919
  - in weight loss treatment, 3919–3920, 3919f, 3920f
- Behavioral therapy. *See* Cognitive behavioral therapy (CBT)
- Behçet syndrome, **2908**
  - clinical features of, 1129f, 1131, 1132t, 2908–2909
    - aortic aneurysm, 2166
    - cutaneous, 2908, 2908f
    - diarrhea, 309
    - genital ulcers, 2908, 2908f
    - neurologic, 2909, A16
    - ocular, 225, 2909
    - oral, 261, 263t, 2908f
    - vasculitis, 2906, 2909
  - definition of, 2908
  - diagnosis of, 2908, 2908t, A16
  - epidemiology of, 2908
  - HLA gene association with, 2908
  - inflammasome mutations in, 2761t
  - KIRs with, 2772t
  - pathogenesis of, 2908
  - treatment of, 155, 2909
- Bejel (endemic syphilis), 1437, 1437t, 1438, 1439f
- Belantamab mafodotin, 566, 566t, 890t, 892
- Belatacept, 2359, 2408t, 2791t
- Belimumab, 2792t, 2834, 2835, 2836
- Belinostat, 533t, 539, 570t, 573, 867
- Belladonna alkaloids, 3711t
- Bell, Joseph, 46
- Bell's palsy, **3554**
  - antiviral therapy for, 1498
  - clinical features of, 242, 265, 3554
  - diagnosis of, 3555, 3555f, A16
  - differential diagnosis of, 3554–3555
  - etiology of, 1498

  incidence of, 3554
  pathophysiology of, 3554
  in pregnancy, 3903
  in relapsing fever, 1446
  treatment of, 3555
Bell's phenomenon, 3554
Belumosudil, 917
Belzutifan, 534t, 570t, 573, 687, 2434
Bempedoic acid, 2968, 3228, 3244t, 3251, 3257
Benazepril, 2144t
Bence Jones proteins, 341, 883, 888, 2440, 2440f
Bendamustine
  action and targets of, 563
  adverse effects of, 562t, 755–756
  for CLL, 855
  for DLBCL, 862
  for follicular lymphoma, 864
  interactions and issues, 562t
  for multiple myeloma, 890t
  for peripheral T-cell lymphoma, 866
  for Waldenström's macroglobulinemia, 894
Benedikt's syndrome, 234
Beneficence, 71
Benign familial hematuria, 2426
Benign familial neonatal seizures, 3412t
Benign forgetfulness of the elderly, 194
Benign hereditary chorea of childhood, 3514
Benign paroxysmal positional vertigo, 163–164, 163t, 164f
Benign prostatic hyperplasia (BPH)
  clinical features of, 703, 703t, 2454
  diagnosis of, 703–704, 703t
  PSA levels in, 696
  treatment of, 704, 2456
Benign recurrent intrahepatic cholestasis (BRIC), 325, 2641t, 2642
Benralizumab, 880, 2233, 2901
Benserazide/levodopa, 3501, 3504t
Benzalkonium, 423
Benzathine penicillin G, 1109t, 1180t, 1208t, 1439, 2857, 2858
Benzbromarone, 2960
Benzene exposure, 512t, 809, 826
Benznidazole
  actions of, 1747
  adverse effects of, 1742t, 1747, 1791t, 2013
  for Chagas disease, 1609, 1747, 1791–1792, 1791t, 2013
  pharmacology of, 1747
Benzoate, 3372
Benzocaine, 780
Benzodiazepine(s). *See also specific drugs*
  for acute alcohol intoxication, 3679
  adverse effects of, 179, 184, 3660, 3769
  for alcohol withdrawal, 3679–3680
  for anticipatory nausea, 83
  for anxiety disorders, 3660, 3660t
  for cocaine intoxication, 3695
  for delirium, 186
  for dyspnea, 83
  for insomnia, 86, 216
  in mechanical ventilation, 2295
  for nausea and vomiting, 299, 299t, 300
  in opiate overdose management, 3691
  overdosage/poisoning with, 3691, 3713t
  for panic disorder, 3658
  pharmacology of, 3660t
  for restless legs syndrome, 217
  for stress disorders, 3663
  synthetic, 3696
  for vertigo, 165, 165t
  withdrawal from, 184, 3660
Benzodiazepine antagonists, 192
Benzonatate, 253, 274
Benzophenones, 435
Benzoyl peroxide, 391
Benzphetamine, 3191
Benztropine, 86, 3503, 3674
Benzylpenicillin, 1240
Bepiroversen, 2681
Bereavement, 76, 85, 315
Beremagene geperpavec, 3824t
Beriberi. *See also* Thiamine (vitamin B_1) deficiency
  cardiac manifestations of, 282, 2018
  clinical features of, 2605t, 2607
  dry, 2607, 3612
  wet, 2507, 3612
Berkelium exposure, S6
Berlin Questionnaire, 2275
Bermejo virus, 1660t, 1674
Bernard-Soulier syndrome, 926
Bernoulli effect, 2203
Berry aneurysm, **3458**. *See also* Subarachnoid hemorrhage
Beryllium disease/toxicity, 406, 2236t, 2238
Beryllium lymphocyte proliferation test, 2238
β-adrenergic agonists
  actions of, 1846, 1849, 1850f
  adverse effects of, 354, 2224
  for asthma, 2223–2224, 2225
  contraindications to, 361
  for COPD, 2255, 2257f, 2259
  genetic variations in response to, 493
  for high-altitude pulmonary edema, 3758
  for hyperkalemia, 361
  overdosage/poisoning with, 3705, 3710t
β-adrenergic antagonists (beta blockers)
  for accessory pathway-mediated tachycardia, 1942
  actions of, 2133, 2144t
  for acute angle-closure glaucoma, 225
  adverse effects of, 231, 487, 2101, 2102t
    alopecia, 394
    arthralgias, 2940t
    AV conduction block, 1926, 1929
    bronchospasm, 2221
    cutaneous, 386, 392, 420
    erectile dysfunction, 3157
    hair loss, 421
    heat-related illness, 3773
    hypoglycemia, 3235
    Raynaud's phenomenon, 2178
    SA node dysfunction, 1920t
    xerostomia, 266
  aldosterone-renin ratio effects of, 3062t
  for aortic aneurysm, 2167
  for aortic dissection, 2170
  for aortic regurgitation, 2044
  for atrial fibrillation, 1949, 1950
  for AV nodal reentrant tachycardia, 1940
  contraindications to, 2101, 2102t, 2109t, 2118, 2121, 2813
  drug interactions of, 2102t, 2690, 3660t
  genetic variations in response to, 492
  for glaucoma, 225, 231
  for heart failure, 1993f, 1994, 1995t
  for high-altitude pulmonary edema, 2331
  for hypertension, 2144t, 2145
  for hypertrophic cardiomyopathy, 2007
  for inappropriate sinus tachycardia, 1936
  for ischemic heart disease, 2100t, 2101, 2102
  for methylxanthine overdose/poisoning, 3710t
  for migraine prevention, 3469, 3470t
  for NSTE-ACS, 2109–2110, 2109t, 2112
  overdosage/poisoning with, 3712t, 3715t
  for palpitations, 291
  perioperative, 3907–3908
  in peripheral artery disease, 2173
  for pheochromocytoma, 3075
  for phobic disorders, 3661
  for postural orthostatic tachycardia syndrome, 3548
  for prevention of sudden cardiac death, 2339
  for respiratory alkalosis, 376
  for STEMI, 2118, 2121, 2126
  for syncope, 158
  for variceal bleeding, 317, 2714
  for ventricular arrhythmias, 1957–1958
β-amyloid. *See* Aβ amyloid
Beta blockers. *See* B-adrenergic antagonists (beta blockers)
β-carotene
  cancer and, 2611
  for erythropoietic protoporphyria, 435
  for lung cancer prevention, 512–513
  for macular degeneration, 231
β-globin gene mutations, 3786f, 3787
β-globin genes, 770, 770f, 771, 776
β-glucan test, 582, 1691
β-hemolytic streptococcal infections. *See* Group A *Streptococcus* (GAS) infections; Group B *Streptococcus* (GBS) infections
β-IV spectrin, 2791t
β-lactam antibiotics
  actions of, 1166, 1174, 1182t, 1183f
  adverse effects of, 150, 423, 425, 1171t, 1175–1176
  for anaerobic bacterial infections, 1378–1379
  for community-acquired pneumonia, 963t, 1028
  cross-sensitivity to, 1176
  for cystic fibrosis, 2248
  for cystitis, 1093, 1093t
  desensitization protocols for, 2818
  indications for, 1170, 1171t, 1174
  for infective endocarditis, 1205
  for *P. aeruginosa* infections, 1309
  for pneumococcal infections, 1192
  resistance to
    in anaerobic bacteria, 1379
    in *F. tularensis,* 1339
    mechanisms of, 1181–1182, 1182t, 1183f, 1219
β-lactamase(s), 1181, 1283, 1283t, 1295–1296
β-lactamase inhibitors, 984
β radiation, S6
β-thalassemia. *See* Thalassemia syndromes
β_2-glycoprotein, 2790t, 2837, 2862
β_2-microglobulin, 860, 895, 896t, 900
Beta diversity, 3835
Betaine, 3371
Betamethasone, 386, 3393, 3648
Betaxolol, 2100t
Bethanechol, 3588, 3660t, 3667, 3712t
Bethesda assay, 930–931, 936
Bethesda classification, thyroid cytology, 3049, 3049t
Bethlem myopathy, 3635, 3639t
Betibeglogene autotemcel, 3824t
BET inhibitors, 3940
Bevacizumab
  actions and targets of, 534t, 547f, 548, 559, 559f, 2792t
  adverse effects of, 510, 548, 559, 658, 666
    hemoptysis, 592
    hypersensitivity reaction, 597
    hypertension, 658
    long-term, 753t
    nephrotoxicity, 2377
    neurologic, 726, 726t
    thrombotic microangiopathy, 2445
  for brain metastases, 590, 724
  for colorectal cancer, 658
  for endometrial cancer, 715
  for glioblastoma, 719f
  for HCC, 662, 663, 664, 664t, 665f, 666

Bevacizumab (*Cont.*):
  indications for, 2792t
  for lung cancer, 625t
  for macular degeneration, 231
  for metastatic cervical cancer, 713
  for ovarian cancer, 712
Bexarotene, 866, 3246
Bezlotoxumab, 1085, 1085t
Bezold-Jarisch reflex, 1919, 1926
bFGF (basic fibroblast growth factor), 546, 546f, 2516, 3004
BFM-like regimen, 846f, 847
Bhanja virus, 1657t
Bias, in screening tests, 514–515
Bicalutamide, 567, 701, 702, 753t
Bicarbonate, 2366f, 2367–2368, 2742
Bicarbonate therapy
  for DKA, 3218
  for hyperkalemia, 361
  for hypovolemia, 347
  for internal radionuclide contamination, S6
  for lactic acidosis, 370
  for metabolic acidosis, 369
  for non–anion gap metabolic acidosis, 373
  for salicylate-induced acidosis, 370
  for uremic acidosis, 372
Bicarbonaturia, 348, 354
Bicipital tendinitis and rupture, 2974
Bictegravir, 1617f, 1618t, 1621, 1621t, 1623f
Bicuspid aortic valve disease, 1891f, 2035, 2036
Bifascicular block, 1930. *See also* Atrioventricular (AV) conduction block
Bifidobacterium spp., 2573, 2580, 3833, 3836, 3841
Bifid pulse, 1863
Bifid uvula, 1860
Bigeminy, 1956, 1962f
Biglycan, 3322t
Biguanides, 369, 3211t, 3212–3213, 3215, 3258. *See also* Metformin
Bilateral simultaneous stimulation test, 174
Bile
  composition of, 2727
  effects on esophageal epithelium, 2509
  functions of, 2459
  hepatic, concentration within gallbladder, 2727
  leak, 2475f
  limey, cholecystitis and, 2733
  regulation of flow of, 2727
  secretion of, 2727
Bile acid(s), 2542–2543, 2542f, 2543t, 2727, 3841
Bile acid diarrhea, 308, 2543, 2543t
Bile acid malabsorption, 311
Bile acid pool, 2727
Bile acid sequestrants
  adverse effects of, 3214, 3244t, 3258
  for bile acid diarrhea, 2543, 2543t
  for lipoprotein disorders, 3244t, 3251–3252
  for metabolic syndrome, 3257–3258
  for type 2 DM, 3214
Bile duct(s)
  common, 2727
  imaging of, 2489, 2494f, V5
  obstruction of, 588, 674, 2472f, 2473f, 2489
Bile duct disease, **2735**
  choledocholithiasis, 326, **2735**
  congenital anomalies, 2735
  diagnosis of, 2736t
  extrinsic compression, 2737
  hemobilia, 2737
  hepatobiliary parasitism, 2737
  in HIV infection, 1601
  primary sclerosing cholangitis. *See* Primary sclerosing cholangitis (PSC)
  stones, 2472f, V5
  strictures, 2736–2737
  trauma-related, 2737
Bile reflux gastropathy, surgery-related, 2534
Bile salt excretory protein, 2642
Bile salt export pump, 2727
Biliary atresia, 2718, 2735
Biliary cirrhosis, **2710**. *See also* Primary biliary cholangitis (PBC)
  coagulation disorders in, 935
  secondary, choledocholithiasis and, 2735–2736
  xanthomas in, 404
Biliary colic
  abdominal pain in, 111, 300
  differential diagnosis of, 300
  in gallbladder disease, 2730–2731
  nausea and vomiting in, 297
Biliary dyskinesia, postcholecystectomy, 2734
Biliary ectasia, congenital, 2735
Biliary hypoplasia, 2735
Biliary sludge, 2729
Biliopancreatic diversion, 3194, 3194f
Biliprotein, 321
Bilirubin. *See also* Jaundice
  biliary excretion of, 2638
  conjugated, 321, 2634, 2637–2638, 2637t
  delta fraction, 321
  extrahepatic disposition of, 2638
  in gut, 2638
  hepatocellular transport of, 2637, 2637f
  metabolism of, 320–321, 2637–2638
  production of, 320–321, 2638
  renal excretion of, 2638
  serum, 321
    elevation of. *See* Hyperbilirubinemia
    in liver function evaluation, 324, 2634, 2634f, 2637t
    measurement of, 321
    normal value, 321
  unconjugated, 321, 2634
  urine, 321, 2634, 2637t, 2638
Bilirubinuria, 320, 321
Biliverdin, 321
Biliverdin reductase, 321
Bill & Melinda Gates Foundation, 3843
Billowing mitral leaflet syndrome. *See* Mitral valve prolapse
Billroth procedures, 2532–2533, 2532f
Bimekizumab, 2887t, 2890
Binge-eating disorder, **3671**
Binimetinib, 533t, 569t, 572, 604, 605, 628
Binswanger's disease, 3490
Bioavailability, 484, 484f
Biodegradable vascular scaffold (BVS) stent, 2128
Biodosimetry Assessment Tool, S6
Bioelectrical impedance analysis, 2617, 2618t
Biofeedback training, 314, 2583, 3466, 3925t, 3927
Biofilms, 967
Bioinformatics, 3797
Biolimus, 2128
Biologic signal, S11
Biologic systems, 3955–3956, 3956f. *See also* Systems biology
Biologic therapy. *See also* Gene therapy
  for cancer. *See* Cancer biologic therapy
  for IBD, 2565–2566, 2568t
  for rheumatoid arthritis, 2849, 2850–2851t, 2851
Biomass smoke-induced illness, 2241, 2241f, 2253
Biomedical Advanced Research and Development Authority (BARDA), S4
Bioprosthetic heart valve. *See* Prosthetic heart valve
Biopsy, 499, 552–553, 552t, S11. *See also specific types and sites*
Bioterrorism, 1149, 1341, 1342, 1459, S4, S7
Biotin, 2600t, **2609**
Bipolar disorder, 3652, **3667,** 3669t, 3952
Birbeck granules, 881
Bird fancier's lung, 2228t, 2229, 2230f
Birt-Hogg-Dubé syndrome, 404, 524t, 687, 687t
Bisacodyl, 82t, 307, 314
BISAP (Bedside Index of Severity in Acute Pancreatitis), 2748
Bisexual, 3179t
Bismuth-containing preparations, 2524t, 2526
Bismuth poisoning, 264t
Bismuth subcitrate, 2527t, 2530t
Bismuth subsalicylate
  adverse effects of, 307, 420, 2526
  for diarrhea, 307
  in *H. pylori* eradication, 302, 1302t, 2527t, 2529f, 2530t
  for microscopic colitis, 309
  for travelers' diarrhea prophylaxis, 307, 1018, 1081
  for travelers' diarrhea treatment, 1081t
Bisoprolol, 1995t, 2100t
Bispecific tumor engaging (BiTe) antibodies
  actions of, 558, 559f
  adverse effects of, 578t, 579, 753t, 755, 757, 849, 892
  for ALL, 848t, 849
  for CLL, 856
  for DLBCL, 863
  for multiple myeloma, 890t, 892
2,3-Bisphosphoglycerate, 771f
Bisphosphonates
  actions of, 542
  adverse effects of, 3304
    actinomycotic osteomyelitis, 1363, 1363f
    atypical femur fracture, 3304, 3306
    hypocalcemia, 3290
    osteonecrosis of jaw, 266, 2940t, 3304
  for bone pain, 82, 731
  in early-stage breast cancer treatment, 638
  for fibrous dysplasia, 3317
  for fracture prevention, 3305f
  for hypercalcemia, 364, 3284–3285, 3284t
  for hypercalcemia of malignancy, 738
  for metastatic bone disease, 731
  in metastatic breast cancer, 639
  for osteogenesis imperfecta, 3327
  for osteoporosis treatment, 3303–3306, 3311
  for Paget's disease of bone, 3314, 3314t
  prophylactic uses of
    after radiation therapy, 756
    in multiple myeloma, 892
    in osteoporosis, 3145, 3303
    in prostate cancer, 701, 703
Bite cells, 798, 799f
Bite-wounds, **1142**
  antibiotic prophylaxis, 1145–1146, 1180t
  approach to the patient, 1144
  cat. *See* Cat(s), bite-wounds
  dog. *See* Dog(s), bite-wounds
  epidemiology of, 1142–1144
  hepatitis B prophylaxis, 1146
  human, 1051, **1143,** 1145t
  infections of, 1143, 1145t
  other animals, 1143
  rabies and tetanus prophylaxis, 1146
  treatment of, 1054t, 1144–1146, 1145t
Bithionol, 434t
Bitopertin, 3351
Bitot's spots, 2611
Bitter taste sensation, 238–239
Bivalirudin
  actions of, 950
  for anticoagulation in HIT, 923, 1905
  vs. heparin, 950
  for NSTE-ACS, 2110t, 2111
  in PCI, 2127
  pharmacology of, 950t
  for STEMI, 2121
  for VTE/PE treatment, 2162, 2163t

BK virus infection
  in HCT recipient, 597, 1155t, 1156–1157
  respiratory tract, 1536
  in solid organ transplant recipient, 1158, 1159t, 1161, 2410
Black Creek Canal virus, 1660t, 1674
Black Death, 1342, 3854, 3854t, S4
Black hairy tongue, 264t
Blackhead, 390
Blackout, 3676
Black tar heroin, 1234, 1237
Black widow spider bite, 111, 3734
Bladder
  dysfunction of, 3569–3570, 3588
  pain. *See* Interstitial cystitis/bladder pain syndrome (IC/BPS)
Bladder cancer, **690**
  carcinoma of unknown primary and, 733t
  clinical features of, 691
  deaths from, 497t, 690
  diagnosis of, 691
  epidemiology of, 690–691
  genetic considerations in, 690–691
  global considerations in, 690
  histology of, 691, 692f
  incidence of, 497t, 498f, 690
  in Lynch syndrome, 690, 691
  molecular biology of, 691
  prognosis of, 693f
  recurrence risk in, 692, 693t
  risk factors for, 690–691
  staging of, 691–692, 691f
  treatment of, 551, 561, 692–694, 693t
Bladder neck obstruction, 2378
"Blade of glass" lesion, 3312
Blastic NK cell lymphoma, 875
Blastic phase, in CML, 842
Blastocystis hominis infection, 1808
Blastomyces spp., 1699, 1700f
Blastomycosis, **1699**
  approach to the patient, 1701
  clinical features of, 1051, 1058, 1130t, 1688t, 1701–1702
  diagnosis of, 1133, 1688t, 1702
  disseminated, 1701–1702
  epidemiology of, 1130t, 1699–1700
  global considerations in, 1700
  pathogenesis of, 1700–1701
  prevention of, 1703
  prognosis of, 1703
  pulmonary, 1701
  risk factors for, 1688t, 1701
  treatment of, 1702–1703, 1703t
Blau's syndrome, 2937
Bleeding, **469**
  approach to the patient, **472**
  in CKD, 2393
  in DIC, 934
  drug-induced, 473, 947–948, 949, 952
  in filovirid infections, 1685
  gastrointestinal. *See* Gastrointestinal bleeding
  in hemophilia, 929
  history in, 472–473
  laboratory evaluation of, 474–476, 474f
  in liver disease, 935t
  systemic diseases that cause/exacerbate, 473
  variceal
    esophageal. *See* Esophageal varices
    gastric, 2482, 2488f, 2750
  in VWD, prophylaxis for, 927
Bleeding time, 474
Bleomycin
  actions of, 564
  adverse effects of, 562t, 564
    cutaneous, 2940t
    hair loss, 421
    hemolytic-uremic syndrome, 595
    in high-altitude travel, 3760
    in hyperbaric oxygen therapy, 3765
    hyperpigmentation, 399, 420
    nausea and vomiting, 575
    pulmonary, 595, 706, 754, 2261
    pulmonary fibrosis, 564, 753t
    Raynaud's phenomenon, 2178
  interactions and issues, 562t
  for ovarian germ cell tumors, 712
  for testicular cancer, 706, 707–708f, 709
Blepharitis, 223–224, 3732
Blepharoptosis, 233
Blepharospasm, 3510, 3555, V3
Blinatumomab
  action and targets of, 534t, 559, 559f, 561, 848, 848t
  adverse effects of, 594, 848
  for ALL, 848, 848t
Blind loop syndrome. *See* Small-bowel bacterial overgrowth
Blindness. *See* Visual loss
Blind spot, 222
Blister beetle, 3737
Blister cells, 798
Blistering/vesicant agents, S5
BLM gene mutations, 524t
Bloating, 326
Blomstrand's lethal chondrodysplasia, 3274, 3280f
Blood-brain barrier, 487f
Blood components
  collection and manufacturing processes, 901, 902–903t
  laboratory evaluation of, S11. *See also* Peripheral blood smear
  transfusions. *See* Transfusion(s)
Blood culture, 581, 1025, 1039–1040, S11
Blood flukes. *See Schistosoma* spp.
Blood gases. *See* Arterial blood gases
Blood glucose monitoring, 2623, 3207. *See also* Glycemic control
Blood groups, 901, 903–904, 903t, 904t
Blood pressure, **2133**. *See also* Hypertension; Hypotension
  circadian variations in, 3953
  classification of, 2134t
  control of, 1846
  home monitoring of, 2149
  measurement of, 1862, 2134–2135, 2134f
  physiology and pathophysiology of, 2133–2134, 2133f
  in pregnancy, 3898
  response to Valsalva maneuver, 3541–3542, 3543f
  screening, 39t
Blood products, contamination of
  hepatitis viruses, 930, 2653–2654
  HIV, 930, 1562t, 1564
  malaria, 1766
  screening for, 1564
  T. cruzi, 1788
  Yersinia, 1349
Blood substitutes, 808
Blood urea nitrogen (BUN)
  in AKI, 2381, 2382
  in hyponatremia, 350
  in hypovolemia, 347
  in nutrition assessment, 2618t
  in prerenal azotemia, 347
Blood vessels
  biology of, **1844**
  origin of vascular cells, 1844
  regeneration of, 1847
  smooth muscles of, 1845f, 1846
  in tumor angiogenesis, 545–546, 545f, 547f
  vascular cell biology of, 1846
Blood volume, 2371f, 2372
Bloom's syndrome, 524t, 755, 3780, S9
"Blueberry muffin syndrome." *See* Rubella, congenital
Blue halo effect, 3159
Blue light, 430, 3946
Blue macules, 400
Blue nevus, 406
Blue River virus, 1660t, 1674
Blue rubber bleb syndrome, 406
Blue sclera, 1860, 3315, 3324, 3326
"Blue toe" syndrome, 2175, 2176f
Blumer's shelf, 646
Blunt injury, cardiac, 2087–2088, 2088t
B lymphocytes. *See* B cell(s)
BMAL1, 3943
BMAL1 gene mutations, 3947t, 3951, 3953–3954
BMI. *See* Body mass index (BMI)
Bmi-1, 759f, 760
BMI1, 3939
BMP1 gene mutations, 3324, 3325t, 3326
BMPR1A gene mutations, 524t
BMPR2 gene mutations, 2185, 2186f, 2187f, 2188, 2191
BNP. *See* B-type natriuretic peptide (BNP)
Bocaparvovirus, 1525
Bocaviruses, 253, 1525, 1536
Boceprevir
  adverse effects of, 2687
  for chronic HCV infection, 2687
  drug interactions of, 2687–2688
Body composition, 2618t
Body fat, 3166. *See also* Obesity
Body fluids
  composition of, 344–346
  detection of parasites in, S12
  specimen collection and transport, S11
Body image disorder, 3123, 3172
Body lice *(Pediculus humanus),* 1446, 1458, 3731–3732
Body mass index (BMI), 2617
  calculation of, 3181, 3188
  classification of, 3181, 3188t
  in malnutrition, 2621
  visual, 2621
Body plethysmography, 2206–2207
Body temperature, 133–134
Body weight. *See* Weight
Boerhaave's syndrome, 2513
Bohr effect, 376, 771f, 807
Bolivian (Machupo) hemorrhagic fever, 1658t, 1675
Bone
  growth of, 3259
  metabolism of, 3259–3261, 3260f
  remodeling of, 3259–3260, 3261f, 3272, 3293–3294, 3293f
  resorption of, 3294f, 3299
  structure of, 3259
Bone age, 2994
Bone-anchored hearing aid, 250
Bone biopsy, 3299
Bone-conduction threshold, 248–249
Bone disease
  anaerobic bacterial infections, 1377
  in β thalassemia, 778t
  in cirrhosis, 2717
  in CKD, 2390
  with defective mineralization, 3316–3317
  dysplasias, 3318
  extraskeletal calcification and ossification, 3318–3319
  fibrous dysplasia, 3317, 3317f
  in Gaucher disease, 3355–3356
  in IBD, 2563
  metastases from other sites, 541, 619, 730–731

- Bone disease (*Cont.*):
  - in multiple myeloma. *See* Multiple myeloma
  - osteogenesis imperfecta. *See* Osteogenesis imperfecta
  - Paget's disease. *See* Paget's disease of bone
  - primary immunodeficiencies with, S9
  - sarcoma. *See* Sarcomas
  - sclerosing, 3314–3316
  - in TB, 1388
- Bone loss. *See* Osteoporosis
- Bone marrow/bone marrow examination
  - in ALL, 844, 844t
  - in anemia, 447–448, 448f
  - in aplastic anemia, 811, A6
  - cellularity in, 808t, A6
  - in CML, 837, 837t
  - collection for transplantation, 915
  - "dry tap," 873t
  - erythroid hyperplasia in, 448f, A6
  - in hairy cell leukemia, 873, 873t
  - in hemophagocytic lymphohistiocytosis, 464
  - hypocellular, 808t
  - immune tolerance checkpoints in, 2789t
  - indications for, 448
  - iron stores, 763f, 765, A6
  - in MDS, 814f, 816–817
  - in megaloblastic anemia, 784–785, 785f
  - in myelofibrosis, 822, 873t
  - in myeloid hyperplasia, 448f, A6
  - in neutrophil disorders, 463
  - normal, 448f, A6
  - procedure for, 447–448
  - stem cell niche sites in, 759
- Bone marrow failure, **808**
  - in aplastic anemia. *See* Aplastic anemia
  - bleeding tendencies in, 473
  - differential diagnosis of, 808t
  - in eosinophilic fasciitis, 416, 810
  - in paroxysmal nocturnal hemoglobinuria, 804
  - pathognomic cells in, 814f
- Bone marrow replacement, 450
- Bone marrow transplantation. *See* Hematopoietic cell transplantation (HCT)
- Bone mass/strength
  - in IBD, 2563
  - measurement of, 3297–3298, 3297f, 3297t
  - sex differences in, 3166
- Bone pain
  - at end of life, 731
  - in Gaucher disease, 3355–3356
  - in metastatic bone disease, 508, 730–731
  - treatment of, 82, 703, 731
- Bone sarcomas, **729**
  - benign, 729
  - chondrosarcoma, 524t, **730**
  - epidemiology of, 729
  - Ewing's sarcoma. *See* Ewing's sarcoma
  - osteosarcoma, 261, **729**
  - staging, 729, 729t
- Bone scan, 3313
- Bone spicules, 231–232, 232f
- Bone spurs, 3296
- *bont* genes, 1233, 1237
- BoNTs. *See* Botulinum neurotoxins (BoNTs)
- *Bordetella* spp., 1276–1277
- *Bordetella pertussis,* 966, 974, 1276–1277. *See also* Pertussis
- Borealpox, 1518t, 1519
- Bornaviridae, 1656–1657, 1658t, 1664–1665t, 1666t
- Bornholm disease (pleurodynia), 1635
- Boron, 2614t
- *Borrelia* spp., 1445, 1445t, S11
- *Borrelia burgdorferi* infections
  - complement system inhibition in, 969, 970f
  - Lyme borreliosis. *See* Lyme borreliosis
  - MALT lymphoma, 857, 865
  - myocarditis, 2013
- *Borrelia lonestari,* 1445, 1445t. *See also* Hard tick relapsing fever (HTRF)
- *Borrelia miyamotoi,* 1445, 1445t. *See also* Hard tick relapsing fever (HTRF)
- *Borrelia recurrentis,* 1445, 1445t. *See also* Louse-borne relapsing fever
- BOR (branchio-oto-renal) syndrome, 2436, E36t
- Bortezomib
  - actions and targets of, 533t, 536, 570t, 573
  - adverse effects of, 570t
    - cardiac, 2027
    - cardiovascular, 2027
    - HSV reactivation, 886, 892
    - nausea and vomiting, 575
    - neurologic, 726, 726t
    - neuropathy, 508, 573, 887, 3608t
  - for AL amyloidosis, 898
  - for cardiac amyloidosis, 2027
  - for cardiac transplant rejection, 2030
  - for cold agglutinin disease, 803
  - for kidney transplant rejection, 2410
  - for lymphoplasmacytic lymphoma, 865
  - for mantle cell lymphoma, 866
  - for multiple myeloma, 536, 573, 890t, 891–892
  - for primary effusion lymphoma, 873
  - for Waldenström's macroglobulinemia, 894
- Bosentan, 2194, 2197t, 2874
- Bosutinib
  - actions and targets of, 533t, 535, 568t
  - adverse effects of, 568t, 753t, 754, 839t, 841
  - for CML, 838–841, 839t
  - genetic variations in response to, 567
  - in imatinib resistance, 841
  - interactions and issues, 568t
  - resistance to, 567
- Botflies, 3732
- Botryomycosis, 1361, 1365
- Botswana, 3865
- Botulinum antitoxin, 1236, 1237, S4
- Botulinum neurotoxins (BoNTs). *See also* Botulism
  - as bioterrorism agent, 1234, S4
  - characteristics of, 1233
  - etiology of, 1233
  - for gastroparesis, 299
  - for IC/BPS, 336
  - therapeutic, 1234
- Botulinum toxin
  - for achalasia, 2508
  - for anal fissure, 2584
  - for detrusor hyperreflexia, 3588
  - for dystonia, 3512
  - for essential tremor, 3509
  - for hemifacial spasm, 3555
  - for ischemic ulcers, 2874
  - for migraine prevention, 3470t
  - for myoclonus, 3516
  - for Raynaud's phenomenon, 2874
  - for tics, 3515
- Botulinum toxoid, 1237, S4
- Botulism, **1233**
  - adult intestinal, 1234
  - in bioterrorism, S4
  - clinical features of, 1235–1236, 2351–2352, 3550, 3628, S4
  - diagnosis of, 1236–1237, S4
  - differential diagnosis of, 1236
  - diplopia in, 234
  - epidemiology of, 1234–1235
  - etiology of, 1233
  - foodborne, 1233–1235, 1234t
  - global considerations in, 1237–1238
  - iatrogenic, 1234, S4
  - infant, 1234, 1235, 1238, 2351–2352
  - pathogenesis of, 1233, 3628
  - prevention of, 1237, S4
  - treatment of, 1237, 3628, S4
  - weaponized inhalational, 1234
  - wound, 1234, 1235, 1237, 1238, 2352
- Bouchard's nodes, 2949f
- Bound morphemes, 202
- Bourbon virus, 1662t, 1669
- Bourneville disease. *See* Tuberous sclerosis
- Boutonneuse fever, 138t
- Boutonneuse fever (Mediterranean spotted fever), 992, 1455t, 1458
- Boutonnière deformity, 2840
- Bovine papular stomatitis, 1518t
- Bovine spongiform encephalopathy (BSE), 3526, 3526t
- Bowel obstruction. *See* Intestinal obstruction
- Bowen's disease, 390, 607
- Bowlegs, 2944, 2951, 2951f
- Bowman capsule, 2363
- Box jellyfish envenomation, 3723, 3725
- BPH. *See* Benign prostatic hyperplasia (BPH)
- Bpi (bactericidal/permeability-increasing protein), 2790t
- Brachial neuritis, 133
- Brachial plexopathy, 3614–3615, A16
- Brachial plexus, 133, 3615, 3615f
- Brachytherapy, 554, 637, 700, 728
- Bradyarrhythmias
  - in AV conduction block. *See* Atrioventricular (AV) conduction block
  - etiology of, 960t
  - junctional bradycardia, A8
  - relative bradycardia, 136
  - in SA node dysfunction. *See* Sinoatrial (SA) node disease/dysfunction
  - syncope in. *See* Cardiac syncope
- Bradykinesia, 168, 3494
- Bradykinin, 94f, 417
- Bradykinin 2 receptor antagonist, 2808
- *BRAF* gene mutations
  - in astrocytoma, 720
  - in colorectal cancer, 520f, 653, 658
  - in craniopharyngioma, 722
  - as drug target, 490, 493, 527, 532, 533t, 572
  - in glioma, 720
  - in hairy cell leukemia, 873
  - in Langerhans cell histiocytosis, 882
  - in lung cancer, 617, 617f, 617t, 628, 628f
  - in melanoma, 432, 601, 602f, 605–606
  - in multiple myeloma, 884
  - in ovarian cancer, 710
  - in pilocytic astrocytomas, 720
  - in thyroid cancer, 3045–3046
  - tumors associated with, 520t
- BRAF inhibitors
  - actions of, 569t, 572
  - adverse effects of, 405, 420, 421, 569t, 572, 753t
  - for lung cancer, 628
  - for melanoma, 604t, 605
  - resistance to, 574
- Brain
  - cerebral hemispheres, 200f, 3430–3431f
  - gray-matter heterotopia, A16
  - inferior aspect of, 3432f
  - lymphatic structures of, 3397–3398
  - metastases to, 590
  - posterior circulation of, 3432f
  - ventral view, 3551f
- Brain abscess, **1135**
  - *Actinomyces,* 1365
  - anaerobic bacterial, 1375
  - in cancer patient, 583
  - clinical features of, 994, 1136
  - definition of, 1135
  - diagnosis of, 1136–1137, 1137f, 3388f, A16

  - differential diagnosis of, 1137
  - epidemiology of, 1135–1136
  - etiology of, 991t, 994, 1136, 1375
  - histopathology of, 1136
  - nocardial, 1359, 1360f, 1361f
  - pathogenesis of, 1136
  - prognosis of, 1137
  - treatment of, 963t, 991t, 1137
- Brain biopsy, 199, 1114, 1361f, A14
- Brain branched chain amino acid deficiency, 3374t
- Brain-computer interfaces, 3968–3969, 3968–3969f
- Brain death, 192, 2298–2299
- Brain-derived neurotrophic factor (BDNF), 3185, 3185t, 3395, 3395f, 3398, 3942
- Brain drain, 3848
- Brain edema, 2342
- Brainerd diarrhea, 307
- Brain fog, 2829
- Brain injury, traumatic. *See* Traumatic brain injury (TBI)
- Brain natriuretic peptide. *See* B-type natriuretic peptide (BNP)
- Brain serine deficiency, 3374t
- Brainstem accessory optic system, 220
- Brainstem auditory evoked responses (BAERs), 249
- Brainstem disorders/lesions
  - clinical features of, 3382t
  - ischemia, 3434–3435, 3434–3436f
  - sensory abnormalities in, 176
  - weakness in, 170
- Brainstem pathways, 3464f
- Brainstem reflexes, 190–191, 190f
- Brain tumors, **716**
  - approach to the patient, 716
  - clinical features of, 716, 716t
  - coma in, 187, 188f, 192
  - epidemiology of, 716
  - germ cell tumors, 3003
  - headache in, 114, 114t, 716
  - imaging in, 716, 3392f, A16
  - intracerebral hemorrhage in, 3453t, 3455
  - in neurofibromatosis type 1, 722. *See also* Neurofibromatosis type 1 (NF1)
  - in neurofibromatosis type 2, 722–723. *See also* Neurofibromatosis type 2 (NF2)
  - primary, **717**
    - extrinsic "benign"
      - colloid cysts, 722
      - craniopharyngiomas, 222, 232, 722, 3002, 3002f
      - dermoid cysts, 722
      - epidermoid cysts, 722
      - meningiomas. *See* Meningiomas
      - pituitary tumors. *See* Pituitary tumors
      - schwannomas, 721. *See also* Vestibular schwannoma
      - solitary fibrous tumor/hemangiopericytoma, 721
    - genetic syndromes associated with, 717t
    - intrinsic malignant
      - astrocytomas, 718, 720
      - ependymoma, 720
      - gangliomas, 720
      - gliomas. *See* Gliomas
      - medulloblastomas, 721
      - PCNSL. *See* Primary central nervous system lymphoma (PCNSL)
      - pineal region tumors, 721
      - pleomorphic xanthoastrocytomas, 720
    - pathogenesis of, 717
  - treatment of, 716–717, 725–726, 726t
  - in tuberous sclerosis, 723. *See also* Tuberous sclerosis
- Branched-chain DNA (bDNA) tests, 1590t, S11
- Branching enzyme deficiency (Andersen disease, type IV GSD), 3360t, **3362,** 3645
- Branchio-oto-renal (BOR) syndrome, 2436, E36t
- Branch retinal vein occlusion, 226
- Brazil, 3864, 3865f
- *BRCA* gene mutations
  - in breast cancer, 633
  - cancers associated with, 524t, 654, 3800–3801
  - early interventions in patients with, 3800–3801
  - epigenetic silencing of, 538
  - genetic testing for, 515, 525, 3797–3798, 3800–3801, 3801f, 3804t
  - liquid biopsy for detection of, 3982
  - in ovarian cancer, 710
  - in pancreatic cancer, 671, 671t
  - population screening for, 3800–3801
  - in prostate cancer, 703
  - synthetic lethality and, 548
  - variable expression of, 3800
- Breakaway weakness, 167
- Break-bone fever. *See* Dengue virus infection
- Breakpoint method, S11
- Breast cancer, **632**
  - after Hodgkin's lymphoma treatment, 755, 871
  - carcinoma of unknown primary and, 733t, 735
  - continuum conceptual model of, 633f
  - deaths from
    - in 2014, 497t
    - by age and sex, 500t
    - by racial and ethnic group, 501t
    - trends in, 499f, 500f
  - diagnosis of, 634–635, 635f
  - gene expression profiles in, 636, 3802, 3939
  - genetic considerations in, 3800, 3801f
  - hereditary, 3800–3801, 3804t
  - incidence of, 497t, 498, 498f, 500f, 501t, 632
  - lifetime risk of, 37t
  - metastatic
    - bone marrow examination in, A6
    - to CNS, 723, 723t, 725f, 3562, 3562f
    - to eye, 232
    - goals of care, 639
    - to heart, 2087
    - lymphadenopathy in, 477
    - presentation and evaluation of, 639, 640f
    - to skin, 406
    - treatment of, 639–640
  - molecular subtypes of, 635–636
  - ovarian cancer and, 710
  - paraneoplastic syndromes in, 737t, 740t
  - pathology of, 636
  - prevention of
    - lifestyle modifications for, 511, 634
    - pharmacologic, 513, 634
    - prophylactic surgery for, 514, 554, 634, 3804t
    - tamoxifen for, 513, 634
  - radiation-induced, 755
  - risk factors for
    - alcohol use, 3677
    - genetic, 520t, 521, 524t, 633–634. *See also* *BRCA* gene mutations
    - hormone use, 3144t, 3145, 3147f
    - nongenetic, 511, 632–633
    - obesity, 2603
  - screening for
    - benefits of, 37, 37t, 515, 634
    - in older adults, 3881t
    - recommendations for high-risk women, 515
    - recommendations for normal-risk women, 39t, 515, 516t
    - six models for, 27
  - second malignancies in survivors of, 755
  - staging of, 634–635, 635t
  - survival rates in, 635t
  - synthetic lethality and, 548
  - treatment of, **636**
    - ado-trastuzumab, 566
    - anti-HER2 therapy, 536, 541, 558, 638
    - ductal carcinoma in situ, 636
    - in early-stage disease
      - bisphosphonates, 638
      - ER- and/or PR-positive, 637–638
      - HER2-positive, 638
      - in men, 639
      - neoadjuvant vs. adjuvant systemic therapy, 637
      - during pregnancy, 638
      - in premenopausal women, 638–639
      - radiation therapy, 636–637
      - surgical, 636
      - survivor follow-up, 638
      - triple-negative, 638
    - hormone receptor–directed therapy, 536
    - lobular carcinoma in situ, 636
    - in locally advanced disease, 639
    - lymphedema after, 2183
    - in metastatic disease. *See* Breast cancer, metastatic
    - molecularly targeted chemotherapy, 536, 567
  - triple-negative, 634, 636, 638, 641
  - tumor markers in, 504t
- Breast development, 3129f, 3130
- Breast-feeding
  - AED use during, 3427
  - antibacterial therapy during, 1168, 1169t
  - antiparasitic agents during, 1741–1746t
  - contraception during, 3154
  - HIV transmission by, 1565
  - IBD treatment during, 2569–2570
  - iron requirement in, 765t
  - nutrient requirements in, 2600–2602t
  - TB treatment during, 1400
  - water requirement during, 2599
- Breast milk jaundice, 2639
- Breast self-examination, 515, 516t, 634
- Breath sounds, 2200
- Breath testing, 301
- Bremelanotide, 3162
- Brentuximab vedotin
  - action and targets of, 535t, 566, 566t
  - adverse effects of, 594, 753t, 754, 755, 756
  - for anaplastic large-cell lymphoma, 867
  - for Hodgkin's lymphoma, 870
  - for mycosis fungoides, 866
  - for peripheral T-cell lymphoma, 866
  - for primary cutaneous CD30+ T-cell lymphoma, 874
- Breslow thickness, 602
- Brexanolone, 3667
- Brexpiprazole, 199, 3673t
- Brexucabtagene autoleucel, 535t, 559f, 3824t
- BRIC (benign recurrent intrahepatic cholestasis), 325, 2641t, 2642
- Brigatinib, 533t, 569t, 573, 628, 628t, 753t
- Brill-Zinsser disease, 1454, 1459, S7
- Brimonidine, 391
- Brincidofovir, 1157, 1487, 1521
- Bristleworm sting, 3724, 3724f
- Brivaracetam, 3421t, 3424
- Broadbent's sign, 2083
- Broad casts, urinary, 343
- Broca's aphasia, 201t, 202
- Broca's area, 201, 204f
- Brodalumab, 387t
- Brodie's abscess, 1061
- Brodie-Trendelenburg test, 2180
- Brolucizumab, 231
- Bromides, 405

Bromocriptine
    adverse effects of, 2178, 3007
    overdosage/poisoning with, 3710t
    for prolactinoma, 3006–3007
    for type 2 DM, 3214
Bronchial artery embolization, 276, 592
Bronchial hygiene, 2244
Bronchial thermoplasty, 2226, 2289
Bronchiectasis, **2242**
    after TB, 1390
    approach to the patient, 2243
    breath sounds in, 2200
    clinical features of, 2243
    clubbing/hypertrophic osteoarthropathy in, 279
    complications of, 2244
    with COPD, A12
    in cystic fibrosis, A12
    definition of, 2242
    diagnosis of, 2242t, 2243, 2243f
    epidemiology of, 2242
    etiology of, 2242, 2242t
    hemoptysis in, 274
    imaging in, 2243, 2243f, A12
    nontuberculous mycobacterial, 1416, 2242
    pathogenesis of, 2242–2243
    pathology of, 2242–2243
    prevention of, 2244
    prognosis of, 2244
    traction, 2243
    treatment of, 2243–2244
Bronchiolitis obliterans, 411, 2287
Bronchitis
    Aspergillus, 1714
    chronic, 272, 274, 2239t, 2249. *See also* Chronic obstructive pulmonary disease (COPD)
    eosinophilic, 273
    hemoptysis in, 274
    pathophysiology of, 2203
Bronchoalveolar lavage, 2208, 2928
Bronchodilators
    anticholinergics. *See* Anticholinergics
    β-adrenergic agonists. *See* Β-adrenergic agonists
    theophylline. *See* Theophylline
Broncholithiasis, 1694
Bronchoplasty, 2288
Bronchopulmonary aspergillosis. *See* Allergic bronchopulmonary aspergillosis (ABPA)
Bronchopulmonary dysplasia, A12
Bronchoscopic ablation procedures, 2288
Bronchoscopy
    bronchoalveolar lavage, 582, 2208
    brushing and endobronchial biopsy, 2208
    endobronchial ultrasound-guided transbronchial needle aspiration, 2209–2210, 2209f, 2288, 2288f
    fiberoptic, 620
    guided peripheral, 2210, 2210f
    in hemoptysis, 276
    in hypersensitivity pneumonitis, 2230
    in interstitial lung disease, 2262
    in lung cancer diagnosis, 620
    peripheral, 2288
    in respiratory disease evaluation, 2201, 2208
    robotic, 2210, 2210f
    therapeutic. *See* Interventional pulmonary medicine
    transbronchial biopsy, 2209
    transbronchial needle aspiration, 2209
Bronzing, in hemochromatosis, 3336
Brown recluse spider bite, 3734
Brown-Séquard syndrome, 170, 176, 3561
Brown tumor, 2390
Brucella spp., 1331–1332
Brucella spp. infections, **1331**
    chronic meningitis, 1130t, 1131
    clinical features of, 1332–1333, 1333t, S7
    diagnosis of, 1133, 1333–1334
    differential diagnosis of, 1333
    epidemiology of, 1331–1332
    global considerations in, 1332
    immune response to, 1332
    incubation period for, 1333, S7
    monocytosis in, 462
    osteomyelitis, 1062
    pathogenesis of, 1332
    postexposure prophylaxis for, 1335
    prevention of, 1335, S7
    prognosis of, 1335
    skin manifestations of, 406
    vs. TB, 1333, 1333t
    treatment of, 1334–1335, S7
    in war veterans, S7
Brudzinski's sign, 1119
Brugada's sign, 359
Brugada's syndrome
    clinical features of, 160, 1970
    ECG in, 1915, 1961f, 1970
    epidemiology of, 1914
    genetic considerations in, 160, 1970
    recurrent VT/VF in, 1975
    sudden cardiac death in, 2334, 2335t
    treatment of, 1970, 2340t
Brugia malayi/timori, 961t, 1708, 1820t, S12. *See also* Lymphatic filariasis
Bruns-Garland syndrome, 3604
Brush border, 2366
Brushing, bronchial, 2209
Bruxism, 217, 261, 265
BSE (bovine spongiform encephalopathy), 3526, 3526t
BTK gene mutations, 2802
BTK inhibitors, 554, 567, 753t, 855, 894, 1689
BTNL2 gene mutations, 2266, 2928
B-type natriuretic peptide (BNP)
    in amyloidosis, 896, 898
    in dyspnea evaluation, 271
    in edema, 280–281
    in heart failure, 328, 1980, 1985
    in pulmonary hypertension, 2190, 2869
Bubo, 1051, 1343, 1344f, S4. *See also* Plague
Buccal mucosa, microbiota of, 3833f
Buccofacial apraxia, 203
Budd-Chiari syndrome (hepatic vein thrombosis)
    abdominal pain in, 112t
    ascites in, 328
    vs. autoimmune hepatitis, 2695
    vs. cardiac cirrhosis, 2712
    in HCC, 2668
    liver transplantation for, 2719
    in paroxysmal nocturnal hemoglobinuria, 804
    in polycythemia, 451, 819
    recurrence of, after liver transplantation, 2724
    splenomegaly in, 480
Budesonide, 309, 2564
Buerger's disease (thromboangiitis obliterans), 2174–2175, 2178
Bufadienolide, 358
Bufavirus, 1525
Buffalo hump, 1602, 1602f
Buffalopox, 1518, 1518t
Buffy coat, 443
Bufo marinus (cane toad), 358
Bufonidae spp., 3712t
"Building-associated illness," 2241
Bulbospinal pathway, 169f
Bulevirtide, 1494, 2681, 2684
"Bulge sign," 2941
Bulimia nervosa, 261, 298, 356, **3671**
Bullae, 137, 378f, 1049f, 1050t, 1051. *See also* Vesicles/bullae
Bullet ant, 3736
Bullous dermatosis of hemodialysis, 401
Bullous impetigo, 389, 401, 1051, 1210. *See also* Impetigo
Bullous pemphigoid. *See* Pemphigoid, bullous
Bumetanide, 374, 1993t, 2331, 2356
BUN. *See* Blood urea nitrogen (BUN)
Bundibugyo virus, 1680, 1680–1683f
Bundle branch reentry ventricular tachycardia, 1967
Bundle of His, 1912, 1925
Bunyamwera virus, 1658t
Bunyaviridae, 993, 1539, 1675–1677
    clinical syndromes of, 1666t
    geographic distribution, 1664t
    reservoirs and vectors of, 1656, 1657t
    structure of, 1656
Buprenorphine, 97t, 100, 3691
Bupropion
    adverse effects of, 85, 615, 3659t, 3666
    for depression, 85, 3659t, 3666
    with naltrexone, for obesity, 3191, 3192t, 3193f
    for smoking cessation therapy, 615, 2255, 3685, 3685t
    for SSRI-related sexual dysfunction, 3660t, 3667
Buried bumper syndrome, 2474, 2484f
Burkholderia cepacia infections, 1309t, 1313, 2248, 2284
Burkholderia mallei, 1313
Burkholderia pseudomallei, 968, 971, 985, 1309t, 1313, S7
Burkitt's leukemia, 845, 847, 847t, 848, 848t. *See also* Acute lymphoid leukemia (ALL)
Burkitt's lymphoma, **862**
    clinical features of, 862
    diagnosis of, 862, 862f, A6
    EBV infection and, 512t, 525, 1510
    epidemiology of, 862
    genetic considerations in, 860t, 862
    in HIV infection, 1613
    malaria and, 1766
    oral manifestations of, 261
    pathogenesis of, 859f
    treatment of, 862
    tumor lysis syndrome in, 593
Burnett's syndrome, 3282
Burning mouth syndrome (glossodynia, glossalgia), 65, 242
Burnout, in health workers, 52, 52f
Burn patient
    AKI in, 2376
    hypocalcemia in, 365
    hypothermia in, 3769
    hypovolemia in, 347
    infections in, 1296, 1307
    respiratory acidosis in, 375
Burn pit emissions, 2240, S8
Burosumab, 3265
Burr cells, 445, 446f
Burrowing asps, 3716. *See also* Snakebites
Bursitis, 2943–2944, 2944f, **2973**
Buschke-Ollendorff syndrome (dermatofibrosis lenticularis disseminata), 3316
Buspirone
    for antidepressant side effect management, 3660t, 3667
    for anxiety, 3660–3661
    for anxiety disorders, 3660–3661, 3660t
    for depression, 3667
    drug interactions of, 1742t
    for functional dyspepsia, 299t, 302
    pharmacology of, 3660t
Bussuquara virus, 1659t

- Busulfan
  - adverse effects of, 842
    - hepatic, 2667
    - hyperpigmentation, 400, 420
    - nausea and vomiting, 575
    - neurologic, 591, 726t
    - pulmonary, 595
  - for CML, 842
  - drug interactions of, 1742t
  - in HCT preparation, 915
- Butabarbital, 3713t
- Butorphanol, 97t, 98, 100, 3467t
- BVS (biodegradable vascular scaffold) stent, 2128
- Bwamba virus, 1658t
- Byler's disease, 2642
- Byssinosis, 2236t, 2238–2239
- Bystander killing, 1588
- Bystander mechanisms, 2826

C

- C1 inhibitor (C1INH) deficiency, 2790t, 2807, 2808
- C1q, 2786f, 2790t, 3395
- C1R/S, 2786f, 3329t, 3330
- C2, 2786f
- C3, 2359, 2415t, 3395
- C3 glomerulopathy, A4
- C4, 2786f
- *C4A* gene mutations, 2823t
- C5a, 2893
- *C9orf72* gene mutations, 3397, 3486, 3514, 3522t, 3523
- *C13orf1* gene mutations, 2554t
- C282Y mutation, 3337
- Ca^{2+}/calmodulin pathway, 1383
- CA15-3, 639
- CA19-9, 504t, 667, 674
- CA-125
  - in ovarian cancer, 517t, 518, 710, 712
  - in pancreatic cancer, 674
  - as tumor marker, 504t
- CA1527/29, 639
- Cabazitaxel, 565, 703
- Cabergoline, 3006, 3010, 3059
- CABG. *See* Coronary artery bypass grafting (CABG)
- Cabotegravir, 1617f, 1618t, 1621, 1621t, 1623f
- Cabozantinib
  - actions and targets of, 533t, 535, 548, 569t
  - adverse effects of, 569t, 666
  - for HCC, 664t, 665f, 666
  - interactions and issues, 569t
  - for medullary thyroid carcinoma, 3085
  - for NETs, 684t, 686
  - for RCC, 689, 690t
- Cachectic myopathy, 2352
- Cache valley virus, 1658t
- Cachexia, 509, 2621
- Cacipacoré viruses, 1659t
- *CACNA1A* gene mutations, 3462
- *CACNA1C* gene mutations, 3404, 3652, 3668
- *CACNA1* gene mutations, 3536
- *CACNAB2* gene mutations, 3652
- CAD. *See* Coronary artery disease (CAD)
- CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy), 3448, 3485, A16
- Cadherins, 542f
- Cadmium exposure/poisoning, 671, 2239t, 2443, 3699, 3700t
- Ca-DTPA, S6
- *Caenorhabditis elegans,* 1809
- Café-au-lait macules/spots (CALMs), 399, 403, 3109, 3317
- Caffeine
  - actions of, 211
  - adverse effects of, 86, 215, 291, 354, 2137, 2442
  - for hypnic headache, 3475
  - infertility and, 3151
  - for low CSF volume headache, 117
  - for migraine, 3467t
  - overdosage/poisoning with, 3710t
  - for post–lumbar puncture headache, S3
  - for shift-work disorder, 219
- CagA protein, 1299
- CAGE questionnaire, 2629, 2629t
- CAH. *See* Congenital adrenal hyperplasia (CAH)
- Calabar swellings, 1737, 1824
- Calcification
  - of coronary arteries, 1880, 1886f, 2095–2097
  - dystrophic, 3318t, 3319
  - extraskeletal, 3318–3319, 3318t
  - metastatic, 3318, 3318t
  - of solitary pulmonary nodule, 622
- Calcific tendinitis, 2974
- Calcinosis
  - in calcium apatite deposition disease, 2961
  - tumoral, 3318–3319
- Calcinosis circumscripta, 3319
- Calcinosis cutis
  - clinical features of, 403, 2869f
  - metastatic, 403
  - in systemic sclerosis, 415, 2861t, 2868, 2869f
- Calcinosis universalis, 3319
- Calciphylaxis, 404, 2391, 2391f
- Calcipotriene, 387
- Calcitonin, **3274**
  - actions of, 3274, 3306
  - adverse effects of, 3303
  - for bone pain, 82
  - circulating level of, 3274
  - ectopic production of, 737t
  - for hypercalcemia, 3284t, 3285
  - hypocalcemic activity of, 3274
  - in medullary thyroid carcinoma, 3085
  - in MEN 2 and MEN3, 3085t
  - for osteoporosis management/prevention, 3306
  - for Paget's disease of bone, 3303, 3314t
  - sources of, 3274
  - as tumor marker, 504t
- Calcitonin gene-related peptide (CGRP), 94, 3274
- Calcitonin gene-related peptide (CGRP) receptor antagonists (gepants), 3467t, 3468–3469, 3470t
- Calcitriol, 366, 2369, 2390, 2391, 3271, 3287
- Calcium, **362**
  - absorption of, 3262
  - bone remodeling and, 3294
  - in cardiac contraction, 1846, 1848f, 1849, 2018
  - in CKD, 2390–2391, 2391f
  - deficiency of. *See* Hypocalcemia
  - excess of. *See* Hypercalcemia
  - extracellular, 362, 362f, 3262
  - food sources of, 3301
  - homeostasis of, 2369, 3261–3262, 3261f
  - intake of, 3262
  - malabsorption of, postgastrectomy, 2534
  - metabolism of, 3261–3262, 3261f
  - nephrolithiasis risk and intake of, 2449, 2452
  - recommended intake of, 2601t, 3145, 3294–3295, 3301, 3301t
  - renal transport of, 2367f, 2369
  - supplements. *See* Calcium supplements
  - tolerable upper intake level of, 2614t
  - toxicity of, 2614t
  - urinary excretion of, 2449
  - in vascular smooth-muscle cell function, 1846, 1847f
- Calcium apatite deposition disease, 2961–2962, 2961t, 2962f
- Calcium carbonate, 2524, 3301, 3301t
- Calcium channel(s)
  - autoantigen, 2791t
  - cardiac, 1911–1912, 1912f
  - disorders of. *See* Hypokalemic periodic paralysis (HypoKPP)
  - muscle, 1846, 1847f
- Calcium channel blockers (CCBs)
  - actions of, 2133, 2144t
  - adverse effects of, 2102t, 2144f, 3157
    - cutaneous, 426
    - edema, 282t, 1860, 2180
    - erythromelalgia, 2179
    - gingival hyperplasia, A3
    - gingival overgrowth, 261
    - heat-related illness, 3773
    - lupus syndrome, 2940t
    - peripheral edema, 3188
    - SA node dysfunction, 1920t
    - secondary parkinsonism, 3497
    - xerostomia, 266
  - aldosterone-renin ratio effects of, 3062t
  - for aortic dissection, 2170
  - for arrhythmias in pregnancy, 3900
  - for atrial fibrillation, 1949, 1950
  - contraindications to, 2102t, 2109t
  - drug interactions of, 2102t, 3660t
  - for focal atrial tachycardia, 1938
  - for heart failure, 1996
  - for high-altitude pulmonary edema, 2331
  - for hypertension, 2143, 2144t
  - for hypertension in CKD, 2396
  - for hypertrophic cardiomyopathy, 2007
  - for ischemic heart disease, 2101–2102, 2101t
  - metabolism of, 485t
  - for NSTE-ACS, 2109t, 2110
  - overdosage/poisoning with, 3712t
  - for Prinzmetal's variant angina, 2101, 2112
  - for Raynaud's phenomenon, 2178, 2874
  - for ventricular arrhythmias, 1958
- Calcium citrate, 366, 3301, 3301t
- Calcium docusate, 82t
- Calcium gluconate, 361, 366, 3301t
- Calcium infusion study, 2536
- Calcium lactate, 3301t
- Calcium oxalate crystals/stones
  - arthropathy from, 2962, 2962f
  - in ethylene glycol poisoning, 2440
  - in IBD, 2562t, 2563
  - renal, 2448, 2451f, 2452, A4. *See also* Nephrolithiasis
  - in short-bowel syndrome, 2548
- Calcium phosphate product, elevated, 403
- Calcium phosphate stones, 2448, 2453. *See also* Nephrolithiasis
- Calcium polycarbophil, 2577t
- Calcium pyrophosphate deposition (CPPD) disease, 2889, 2960–2961, 2960t
- Calcium-sensing receptors, 363, 364, 2790t, 2985t
- Calcium signaling defects, 2799
- Calcium sparks, 1846
- Calcium supplements
  - absorption of, 3262
  - adverse effects of, 3301
  - calcium content of, 3301t
  - for cancer prevention, 511
  - for hyperkalemia, 361
  - for hypocalcemia, 366
  - for hypoparathyroidism, 3287
  - for osteoporosis management/prevention, 2612, 3145, 3300–3301
  - for vitamin D deficiency, 3271
- Calciviridae. *See also* Norovirus infection
  - characteristics of, 1628
  - replication strategies of, 1479
  - structure of, 1478f, 1480t, 1628f
- Calcofluor white method, S11

Calculation ability, assessment of, 3380
Calculus, 260, A3
Calicheamicin, 566
California encephalitis virus, 1658t, 1672–1673
Californium exposure, S6
CALMs (cafe-au-lait macules/spots), 399, 403
Calor, 454f
Caloric irrigation, 236
Caloric needs, 2598
Caloric restriction, aging effects and, 3871, 3873–3874
Caloric testing, 163, 190–191
Calpainopathy (LGMD2A), 3641
Calretinin, 616, 733t, 735
CALR gene mutations, 819, 822, 823, 824
Calsequestrin, 1849, 2004t
Calymmatobacterium granulomatis, 1356. *See also* Donovanosis
CAMP factor, 1213
cAMP response-element binding protein (CREB), 3654, 3654f, 3689, 3689f, 3873t
Camptothecins, 564
Campylobacter spp., 1303, 1304t, 3398
Campylobacter spp. infections, **1303**
  antibiotic resistance in, 1186t
  climate change and, 3752t
  clinical features of, 306t, 1060, 1079t, 1304t, 1305
  colitis, vs. IBD, 2559
  complications of, 1305
  diagnosis of, 1080, 1305–1306, S11
  differential diagnosis of, 1306
  epidemiology of, 1079t, 1304
  etiology of, 1304, 1304t
  food-borne, 305, 1080t, 1081t, 1304
  Guillain-Barré syndrome and, 3617
  in HIV infection, 1598
  immunoproliferative small intestinal disease and, 894, 1305
  MALT lymphoma of small intestine and, 857, 865
  pathogenesis of, 1304–1305
  prognosis of, 1306
  treatment of, 1081, 1306
Camurati-Engelmann disease (progressive diaphyseal dysplasia), 3315
Canadian Cardiovascular Society, classification of cardiovascular disease, 2093t
Canagliflozin, 3211t, 3213
Canakinumab, 2792t, 2936, 2959
Canasa suppositories, 2564
c-ANCA (cytoplasmic antineutrophil cytoplasmic antibody), 2894, 3624
Cancer. *See also specific types*
  anemia in, 768, 769
  approach to the patient, **497**
  atherosclerosis and, A10
  autoimmunity and, 2826–2827
  cell growth/death pathways, 556, 557f
  cellular senescence in, 531
  chromosome abnormalities in, 502. *See also* Cancer genetics
  chronic meningitis in, 1132t, 1134f
  circadian disruption and, 3953
  clonal origin of, 519–520, 520f
  death and dying issues, 505
  deaths from
    by age, 500t
    averted, since early 1990s, 502f
    by race and ethnicity, 501t
    by sex, 497t, 499f, 500t, 3163f
    by site, 497t, 499f
    trends in, 498, 499f, 502f
    in United States and Britain, 75t
    worldwide, 503f
  definition of, 551–552
  depression in, 3664
  diagnosis of, 499, 552–553, 552t
  early detection of, 3981
  emergencies in, **585**
    adrenal insufficiency, 593
    airway obstruction, 592, 592f
    biliary obstruction, 588
    cytokine release syndrome, 561, 579, 594–595, 754, 2351, 2788
    hemolytic-uremic syndrome, 595. *See also* Hemolytic-uremic syndrome (HUS)
    hemoptysis, 591–592
    hemorrhagic cystitis, 597
    hypercalcemia. *See* Hypercalcemia
    hypersensitivity reactions to antineoplastic drugs, 597–598
    hypoglycemia, 592–593
    increased intracranial pressure, 590
    intestinal obstruction. *See* Intestinal obstruction
    lactic acidosis, 592
    neoplastic meningitis, 590–591
    neutropenic enterocolitis (typhilitis), 582, 597, 597f, 1376
    pericardial effusion/tamponade, 587
    pulmonary and intracerebral leukostasis, 591
    pulmonary infiltrates, 595–596
    seizures, 591
    SIAD. *See* Syndrome of inappropriate antidiuresis (SIAD)
    spinal cord compression. *See* Spinal cord compression, neoplastic
    superior vena cava syndrome. *See* Superior vena cava syndrome
    tumor lysis syndrome. *See* Tumor lysis syndrome
    urinary obstruction, 588
  extent of disease, 499, 501
  fatigue in, 166
  fever of unknown origin in, 150, 151t
  genetic considerations in. *See* Cancer genetics
  genetic testing in, 525, 525f, 3804t
  genomes of, 525–526
  global considerations in, 498, 3851, 3852f
  high-altitude travel and, 3760
  immune evasion in, 544
  immunopathology of, 548–551, 550f
  incidence of, 497–498, 497t, 498f, 500f, 501t
  infections in, **576**
    approach to, 576–577
    bacteremia, 581
    C. difficile–associated disease, 582–583
    cholecystitis, 583
    CNS, 583
    gastrointestinal tract, 597
    intrinsic risk factors for, 577
    neutropenia and. *See* Neutropenia, in cancer patient
    neutropenic enterocolitis, 582
    pneumonia, 580f, 582
    prevention of, 584–585
    prophylaxis for, 577
    sinusitis, 581–582
    skin and soft tissue, 583, 584f, A5
    treatment-related risk factors for, 577–579, 578t
    urinary tract, 583–584
  infectious causes of, 957
  involuntary weight loss in, 315
  in low- and middle-income countries, 3851
  lymphatic vessels of, 546
  metastasis of, 529t. *See also* Metastatic disease
  microbiome and, 3836–3837
  new symptoms in, 504
  nutritional issues and support in, 2626
  obesity and, 512, 544–545, 3187
  in older adults, 3883
  as organ that ignores its niche, 529–530
  pathogenesis of. *See* Cancer cells; Carcinogenesis
  patient management in, **498**
  pediatric vs. adult, 529
  physiologic reserve of patient, 502, 503t
  in pregnancy, 3902
  prevention of. *See* Cancer prevention
  prognosis of, 499, 501
  progressive, 504
  recurrent, 505
  risk factors for, 498, 3677, 3683, 3683t
  screening for. *See* Cancer screening
  second malignancy in, 554, 755, 871
  self-renewing capacity of, 761
  signaling pathway alterations in, 526, 526t
  stable disease, 504
  staging of, 499, 553
  stem cells and, 761
  stroke in, 3451
  survivorship issues, 752–753, 757–758
  symptom control in, **506**
    anorexia and cachexia, 509
    assessment of, 506, 507f
    constipation, 509
    dyspnea, 509–510
    fatigue, 506
    gastrointestinal and oral mucositis, 507t, 508
    hot flashes, 510
    nasal vestibulitis, 510
    nausea and vomiting, 508–509
    pain, 506–508, 507–508t
  treatment of
    assessing response to, 504, 504t, 3983
    cancer stem cells as target of, 542
    complete response to, 504
    deaths averted by, 502f
    gene therapy, 3824t, 3828
    genetic considerations in, 527–528
    goal of, 553
    localized, 554–555
      radiation therapy. *See* Radiation therapy
      surgical. *See* Cancer surgery
    long-term follow-up and late effects, 504–505, 532t, **752,** 753t, 757–758
    palliative, 504, 553, 576
    partial response to, 504
    planning, 502
    principles of, **551**
    psychosocial support, 505
    resistance to, 574
    spectrum of, 552t, 553
    supportive care, **505, 574**
    systemic
      biologic therapy. *See* Cancer biologic therapy
      checkpoint inhibitors. *See* Immune checkpoint blockade/inhibitors
      cytotoxic chemotherapy. *See* Cancer cytotoxic chemotherapy
      gene therapy to enhance, 3830
      molecular targeted. *See* Cancer molecular targeted therapy
      overview, 555–558
      strategies, 556–558, 558t
  tumor growth rates in, 553, 553f
  tumor heterogeneity in, 525–526
  tumor microenvironment in, 544
  vaccines for, 551, 3829
  viruses and, 525

- Cancer biologic therapy, **558**
  - actions and targets of, 555, 559f
  - adverse effects of, 421–422, 737t, 740, 753t, 2672, 2819
  - antibody-mediated, 558
    - antiangiogenic agents, 547–548, 548f
    - antibody-drug conjugates, 535t, 536, 558, 559f, 565–566, 566t
    - anti-tumor cell antibodies, 558–559, 559f
    - bispecific tumor engaging (BiTe) antibodies, 558, 559f
    - immunoregulatory, 548–549, 550f, 558, 559f, 560
    - stroma-directed antibodies, 559–560, 559f
  - CAR-T cells. *See* Chimeric antigen receptor (CAR) T cells
  - cytokines, 561
  - immune checkpoint inhibitors, 2672, 2786f, 2787–2788, 2787t. *See also* CTLA-4 antibodies; PD-1 antibodies
  - nontargeted immunomodulators, 559f, 561
  - oncolytic viruses, 559f, 561
  - principles of, 548–549, 558
  - T cell–mediated, 560t, 561
- Cancer cells
  - apoptosis in, 529t, 539–541, 540f, 556, 557f
  - autophagy in, 541
  - biology of, **528**
  - cell cycle in, 530–531
  - chromatin structure alterations in, 538, 538f
  - evasion of immune system by, 529t, 548–549, 550f
  - failure to differentiate, 529t
  - gene expression in, 529t, 538–539, 538f
  - Gompertzian growth curve/kinetics in, 529, 553, 553f
  - induction of differentiation in, 556
  - loss of replicative senescence in, 529t, 531
  - metabolism in, 543–544, 544f
  - mutations in, 528
  - oncogene addiction in, 536, 537f
  - phenotypic characteristics of, 529t
  - plasticity of, 542–543
  - resistance of, 542–543
  - signal transduction pathways in, 526f, 531–537
  - synthetic lethality in, 537f
  - telomerase activation in, 531
  - tissue invasion by, 529t, 541
  - in tumor angiogenesis, 528, 545–546
- Cancer cytotoxic chemotherapy, **555, 561**. *See also specific drugs and cancers*
  - action and targets of, 556
  - adjuvant, 556
  - adverse effects of
    - cardiovascular, 1895
    - cutaneous, 420
    - depression, 504
    - desensitization protocols for, 2818–2819
    - long-term, 504–505, 753t
      - cardiovascular, 753–754, 1895, 2016–2017
      - hematologic, 826
      - hepatic and gastrointestinal, 757
      - immune system dysfunction, 755
      - neurologic, 756–757
      - pulmonary, 754–755
      - renal/urinary, 757
      - reproductive/endocrine, 755–756, 3113, 3115
      - second malignancy, 755
    - neurologic, 726, 726t
    - neuropathies, 3608, 3608t
    - neutropenia. *See* Neutropenia, in cancer patient
    - toxic erythema, 421–422, 421f
  - antimetabolites, 562t, 564–565
  - antitumor antibiotics, 562t, 564
  - complete response to, 555
  - development of, 59–60, 556f
  - DNA-interactive agents, 561, 562t, 563, 564
  - dose-limiting toxicity, 555
  - high-dose regimens, 556
  - intraperitoneal, 712
  - maximum-tolerated dose, 555
  - metabolism modulators for, 543–544
  - mitotic spindle inhibitors, 563t, 565
  - neoadjuvant, 556
  - in older adults, 3883
  - palliative, 556
  - partial response to, 555
  - during pregnancy, 576
  - principles for use, 555–558
  - resistance to, 542–543, 574
  - response to, 555
  - strategies in, 556–558, 558t
  - supportive care during, **574**
  - timing of, 219
  - topoisomerase poisons, 562t, 564
- Cancer field effect, 660
- Cancer genetics, **519**. *See also specific cancers*
  - cancer gene classes, 520. *See also* Oncogene(s); Tumor-suppressor genes
  - chromosomal instability in solid tumors, 521
  - clonal origin and multistep nature of cancer, 519–520, 520f
  - CpG islands in, 3782, 3783
  - DNA methylation in, 3783, 3784
  - epigenetic control mechanisms, 522, 3940
  - epigenetic influences on gene transcription, 538–539, 538f
  - familial cancer syndromes, 522–525, 524f, 524t
  - future of, 528
  - genomic profiling, 32–33, 33f, 35, 525–526, 3982
  - historical perspective, 519
  - somatic mutations in, 3792, 3792f
  - treatment strategy and, 527–528, 3777
  - tumor heterogeneity and, 525–526, 527f
  - viruses in, 525
- Cancer Genome Atlas, 3777, 3940
- Cancer molecular targeted therapy, 533–535t, 555, **566**
  - action and targets of, 556, 557f
  - adverse effects of, 753t
  - chromatin-modifying agents, 570t, 573
  - combination, 536
  - cyclin-dependent kinase inhibitors, 570t, 573
  - development of, 555–556, 556f
  - diagnostically guided, 527–528, 566
  - DNA methyltransferase inhibitors, 538f
  - histone deacetylase inhibitors. *See* Histone deacetylase (HDAC) inhibitors
  - hormone receptor–directed, 536–537, 566–567
  - KRAS/RAF/MEK inhibitors, 569t, 572
  - metabolism modifiers, 571t, 573–574
  - multikinase inhibitors, 526f, 569–570t, 572–573
  - non-receptor-linked tyrosine kinase antagonists, 567, 568t
  - PARP inhibitors. *See* PARP (poly-ADP-ribose polymerase) inhibitors
  - principles of, 555
  - protein homeostasis modulators. *See* Protein homeostasis modulators (proteasome inhibitors)
  - receptor-linked tyrosine kinase antagonists, 567, 568–569t, 572
  - resistance to, 574
- CancerNet, 502
- Cancer predisposition syndromes, 522–525, 524t
- Cancer prevention, **511,** 3851. *See also specific cancers*
  - BMI and, 512
  - cancer biology and, 551
  - chemoprevention, 512–514
  - deaths averted by, 502f
  - diet modification, 511
  - education and healthful habits, 511–512
  - physical activity, 511
  - smoking cessation, 511
  - sun avoidance, 512
  - surgical methods, 514, 554
  - vaccines and, 513–514
- Cancer screening, 37, **514**. *See also specific cancers*
  - accuracy of, 513, 514t
  - assessment of tests for, 514
  - biases of tests for, 514–515
  - in cancer survivors, 758
  - cell-free DNA methods for, 3981, 3982f
  - recommendations for asymptomatic normal-risk subjects, 516–517t
  - risks from, 515
- Cancer stem cells, 542, 543f, 553, 617
- Cancer surgery, 514, **553**
- Cancer syndromes, familial, 3804t
- Cancrum oris (noma), 261, 1374
- Candesartan, 1987, 1994, 1995t, 2144t, 3469, 3470t
- *Candida* spp., 1688t, 1706–1707, 1723, S11
- *Candida auris*
  - antibiotic resistance in, 1186t
  - antifungal resistance in, 1711–1712
  - clinical features of, 1150. *See also Candida* spp. infections
  - diagnosis of, 1709–1710, 1710t
  - epidemiology of, 1150, 1707, 1723
- *Candida* spp. infections, **1706**
  - in cancer patient, 577, 581
  - in diabetes mellitus, 3229
  - diagnosis of, 381, 1709–1710, 1709f, 1710t
  - disseminated/deeply invasive, 1688t, 1708–1709, 1708f
    - in brain, 583, A16
    - chronic meningitis, 1130t
    - endocarditis, 1037t, 1038, 1044
    - gastrointestinal, 1071
    - liver abscess, 1075
    - ocular, 1708–1709, 1709f
    - osteomyelitis, 1062
    - renal abscess, 1075
    - risk factors for, 1688t, 1707–1708, 1707t
    - skin manifestations of, 148, 583, 1708, 1708f, A1
    - treatment of, 1710–1712, 1711–1712t
  - epidemiology of, 1707
  - esophageal
    - complications of, 2513
    - diagnosis of, 1599f, 2495f, 2512f
    - in HIV infection, 1598, 1599f, 2513
    - treatment of, 1710t, 2513
  - etiology of, 388t, 389–390
  - health care–associated, 1150
  - immune response to, 1689
  - in immunocompromised patient, 390
  - inflammasome mutations in, 2761t
  - joint, 1059
  - mucocutaneous, **389**
    - chronic, 1708, 2801, 3089, 3090t
    - clinical features of, 267, 378t, 388t, 389–390, 1688t
    - diagnosis of, 1723
    - etiology and pathogenesis of, 1723
    - intertriginous, 390
    - perianal, 1708
    - prevention of, 463
    - risk factors for, 1688t, 1689
    - thrush. *See* Thrush
    - treatment of, 388t, 390, 1710t, 1723
  - oropharyngeal, 65, 264t, 267, 295, A3. *See also* Thrush
  - pathogenesis of, 1707–1708, 1707t
  - prophylaxis for, 1595t, 1712

Candida spp. infections (*Cont.*):
  in solid organ transplant recipient, 1161–1162, 1164t
  in transplant recipient, 2410
  urinary tract, 1087, 1095
  vulvovaginal, 1101–1102, 1101t, 1602, 1708, 1710t
Candidate genes, 491, 3793
Candidatus Neoehrlichia mikurensis, 1461
Candiru virus, 1662t
CANDLE (chronic atypical neutrophilic dermatosis with lipodystrophy and elevated temperature) syndrome, 2937
Cane toad *(Bufo marinus),* 358
Cangrelor, 944, 2110, 2110t, 2127, S13
Cannabidiol, 299, 299t
Cannabidiol (CBD), 3421t, 3686
Cannabinoid(s). *See also* Cannabis/marijuana use
  adverse effects of, 184, 2016
  for multiple sclerosis, 3688
  for nausea and vomiting, 83, 298, 299t
  for pain, 100, 3688, 3925–3926
  for spasticity, 3925–3926
  synthetic, 3696
Cannabinoid hyperemesis syndrome, 298, 3688
Cannabis edibles, 3687
Cannabis/marijuana use, **3686**
  accidents and, 3688
  adverse effects of, 1920t, 2137, 3687, 3688
  head and neck cancer risk and, 610
  hypogonadism and, 3114
  inhaled, 3686–3687
  initial actions of, 3656t
  mental illness and, 3688
  oral consumption of, 3687
  pharmacokinetics of, 3686–3687
  pharmacologic effects of, 3686
  prevention of, 3687
  therapeutic potential, 83, 299t, 3688
  toxicity of, 3688
  usage of, 3686
Cannabis use disorder, **3687**
Cano Delgadito virus, 1660t, 1674
Cantalago pox, 1518t
Cantharidin, 3737
CANVAS (cerebellar ataxia, neuropathy, and vestibular areflexia syndrome), 3537, 3603
CaO$_2$ (arterial oxygen content), 2310
Capacity to Consent to Treatment Instrument, 3877
CAPD (continuous ambulatory peritoneal dialysis), 1072, 2384–2385, 2400. *See also* Peritoneal dialysis
Capecitabine
  actions of, 562t
  adverse effects of, 420, 562t, 641
  for cholangiocarcinoma, 668, 669, 669f
  for colorectal cancer, 658
  for gallbladder cancer, 670
  for gastric cancer, 649
  genetic variations in response to, 493, E72t
  interactions and issues, 562t
  for metastatic breast cancer, 641
  for NETs, 684t, 685
Capgras syndrome, 195
Capillaria philippinensis/capillariasis, 961t, 1736, 1739t, 1818–1819, S12
Capillaries. *See* Blood vessels
Capillaritis, 407
Capillary leak syndrome, 574, 1247, 1845, 2361f
Capillary telangiectasias, 3453t, 3457
Capitation, 23
Capivasertib, 569t, 640, 753t, 754
Caplacizumab, 925
Caplan's syndrome, 2840
Caplan syndrome, 2238
Capmatinib, 533t, 568t, 628
Capnocytophaga spp. infections
  antibiotic resistance in, 1267
  clinical features of, 993, 1051, 1267
  complications of, 993
  etiology of, 993
  treatment of, 1145, 1145t, 1177, 1267, 1267t
Capreomycin, 1399, 1426
CAPS (cryopyrin-associated periodic syndromes), 155, 1132t, 2761t, 2936–2937
Capsaicin, 299t, 300, 2955t, 3227, 3604t
Capsid, viral, 1477, 1478f
Capsular polysaccharide complex, 1073
Capsule endoscopy
  in celiac sprue, 2500f
  in Crohn's disease, 2558, 2559f
  in gastrointestinal bleeding, 320
  indications for, 2462, 2462t
  of jejunal vascular ectasia, 2471f
  small-bowel, 2465, 2468, 2471f
Capsulitis, adhesive, 2974
Captopril, 411, 923t, 1994, 1995t
Captopril renography, 2150, 2151t
Caput medusae, 323, 327, 2630
Caraparú virus, 1658t, 1667
Carapito itch, 3737
Carbamazepine
  for acroparesthesia, 3355
  adverse effects of, 3553
    cardiotoxicity, 2017
    chemokine release syndrome, 595
    cholestasis, 2667
    common, 3670t
    cross-sensitivity and, 427
    cutaneous, 393, 401
    delayed hypersensitivity, 408. *See also* Drug-induced hypersensitivity syndrome (DIHS)
    DIHS, 424
    erythroderma, 393
    fever, 150
    genetic considerations in, 419
    hair loss, 421
    lupus syndrome, 2835, 2940t
    nail disorders, 421
    neurologic, 3421t
    in pregnancy, 3427
    rare, 3670t
    SJS/TEN, 425. *See also* Stevens-Johnson syndrome (SJS); Toxic epidermal necrolysis (TEN)
    systemic, 3421t
    thrombocytopenia, 923t
    weight gain, 3188
  for bipolar disorder, 3668, 3670t
  for chorea, 2858
  contraindications to, 3420
  drug interactions of, 489t, 1422, 1742t, 2721, 3421t, 3427
  genetic variations in response to, 491, 494, E72t
  for neuropathic pain, 82, 506
  for neuropathy, 3604t
  overdosage/poisoning with, 3713t, 3715t
  for pain, 97t, 100, 3588
  for paroxysmal symptoms in MS, 3588
  pharmacology of, 3421t, 3670t
  for seizures, 3420, 3421t
  for SUNCT/SUNA prevention, 3473
  for trigeminal neuralgia, 3553
Carbamoylphosphate synthase-1 deficiency, 3369t, 3372
Carbapenem(s)
  actions of, 1166, 1182t, 1183f
  for anaerobic bacterial infections, 1379, 1379t
  for *Campylobacter* infections, 1306
  cross-sensitivity to, 1176
  drug interactions of, 1172t
  for gram-negative bacterial infections, 1268t, 1283–1284
  indications for, 1175
  for *Klebsiella* spp. infections, 1291
  for MDR-TB, 1425–1426
  for neutropenic enterocolitis, 582
  for *Proteus* spp. infections, 1291
  resistance to, 1175
    in *Acinetobacter* spp., 1296, 1297
    in anaerobic bacteria, 1379
    detection of, 977t, 981
    in *Enterobacterales,* 977t, 981, 1152, 1283–1284, 1283t
    mechanisms of, 1181, 1182t, 1183f
    urgency of, 1186, 1186t
  for staphylococcal infections, 1204
Carbapenemases, 1181, 1283t, 1284
Carbenicillin, 355, 374
Carbenoxolone, 355, S1
Carbidopa/levodopa, 3501, 3504t
Carbidopa/levodopa/entacapone, 3502, 3504t
Carbimazole, 394, 3035
Carbinolamine dehydratase deficiency, 3368t
Carbohydrate(s)
  digestion and absorption of, 2542
  disorders of metabolism of
    fructose metabolism disorders, 3359f, 3362t, 3366
    galactose metabolism disorders, 3359f, 3361t, 3365–3366
    glycogen storage diseases. *See* Glycogen storage diseases (GSD)
  high intake of, dyslipidemia and, 3246
  malabsorption of, 309
  metabolism of, 3358, 3359f
  in parenteral nutrition, 2622
  recommended intake of, 2602t
  requirement for, 2598
Carbohydrate loading, 3347
Carbon dioxide (CO$_2$), 3744, 3745f, 3746t
Carbon disulfide, 3610t
Carbonic anhydrase, 2366f, 2368
Carbonic anhydrase deficiency, 3315
Carbonic anhydrase II, 2790t
Carbonic anhydrase inhibitors, 228, 231, 1745t, 2133
Carbon monoxide poisoning
  clinical features of, 3753
  complications of, 2348
  epidemiology of, 3753
  hypoxia in, 277
  imaging in, A16
  lactic acidosis in, 369
  optic neuropathy in, 228
  pathophysiology of, 780, 3753
  polycythemia in, 402
  smoke inhalation and, 2239–2240
  treatment of, 780, 2348, 3753
Carbon tetrachloride, 2665, 2667t
Carboplatin
  adverse effects of, 562t, 563, 753t
    anaphylaxis, 2812
    desensitization protocol for, 2817f, 2819
    hypersensitivity reaction, 597
    nausea and vomiting, 575
    nephrotoxicity, 2377
  for anal cancer, 659
  for bladder cancer, 693, 694
  for breast cancer, 637
  for carcinoma of unknown primary, 735
  for endometrial cancer, 715
  interactions and issues, 562t
  for lung cancer, 624, 625t, 627t
  for metastatic cervical cancer, 713

  for ovarian cancer, 712
  as substitute for cisplatin, 756
  for testicular cancer, 707f, 708
  for thymic tumors, 630
Carboxyhemoglobin, 452, 780
Carboxypeptidase E, 3185, 3185t
Carbuncles, 1054, 1198
Carcinoembryonic antigen (CEA), 504t, 639, 658, 674
Carcinogen(s), 431, 512
Carcinogenesis
  epigenetics in, 3940
  initiation step in, 512
  Knudson two-hit hypothesis/model of, 3783t, 3788, 3792
  multistep nature of, 519–520, 520f, 3792
  promotion, 512
  in skin, 432
Carcinoid syndrome, **681**
  carcinoid crisis, 682–683
  clinical features of, 309, 681–682, 2608
  diagnosis of, 682
  diarrhea in, 309, 681–682
  heart disease in, 683
  pulmonic stenosis/regurgitation in, 2063
  serotonin synthesis and secretion in, 682f
  skin manifestations of, 395, 400, 681
  treatment of, 312, 682, 683f
  tricuspid stenosis/regurgitation in, 683, 683f
Carcinoma, 528
Carcinoma in situ, bladder, 692
Carcinoma of unknown primary, **731**
  approach to the patient, 731
  biology of, 731
  cervical adenopathy in, 611
  genetic considerations in, 731
  imaging in, 732
  immunohistochemical analysis in, 732–733, 733t
  incidence of, 731
  molecular profiling in, 733
  next-generation sequencing in, 733–734
  occult germ cell tumor, 709
  pathology of, 732–734, 732t
  prognosis of, 735
  treatment of, 734–735, 734f
  tumor markers in, 731–732, 732f
Carcinosarcoma, uterine, 715
CARD8 gene, 2761–2762t
CARD9 deficiency, 1689, 1712, 2801
CARD15 gene mutations. *See NOD2* gene mutations
carD gene, 1384
Cardiac allograft vasculopathy, 1931f, 2030
Cardiac arrest
  coronary artery disease and, 2334–2335
  definition of, 2332, 2333t
  epidemiology of, 2332–2333
  etiology of, 2332, 2335t
  hypothermic, 3769
  precipitating factors for, 2333
  sudden, 2332
  survival rates after, 2336, 2338
  treatment of, **2335**
    initial, 2336
    long-term, 2338
    postresuscitation, 2336–2338
    rhythm-based, 2336, 2337f
  ventricular tachycardia and, 1956
Cardiac arrhythmias. *See also specific arrhythmias*
  after STEMI, 2123–2125
  bradyarrhythmias. *See* Bradyarrhythmias
  in cardiac amyloidosis, 2027
  diagnosis of, 1914–1915
  evaluation of, 1914–1915
  in heatstroke, 3775
  in hemochromatosis, 3336
  in hyperkalemia, 359
  in hypokalemia, 356
  in hypomagnesemia, 3267
  in hypothermia, 3770
  in ischemic heart disease, 2091–2092
  mechanisms of, **1912,** 1913f, 1913t
  in pregnancy, 3900
  in respiratory alkalosis, 376
  respiratory sinus, A8
  sudden cardiac death and, 2334, 2334f, 2335t
  syncope in. *See* Cardiac syncope
  tachyarrhythmias. *See* Supraventricular tachyarrhythmias; Tachycardia
  treatment of
    antiarrhythmic drugs, 1915–1916, 1916t
    catheter-based ablation. *See* Catheter-based ablation
    implantable devices. *See* Cardiovascular implantable electronic devices (CIEDs)
  warning arrhythmias, 2123
Cardiac asthma, 1982
Cardiac auscultation
  dynamic, 1867, 1867t
  in heart murmurs, 289, 289t, 1865–1866, 1865–1866f. *See also* Heart murmurs
  heart sounds in, 1864, 1864f. *See also* Heart sounds
  in pericardial disease, 1867
  of prosthetic valves, 1867
Cardiac biomarkers. *See also* Troponins
  in chest pain evaluation, 107
  in myocardial injury, 2109t
  in NSTE-ACS, 2108
  in STEMI, 2115–2116, 2117f
  in VTE, 2160
Cardiac cachexia, 1983, 1983t
Cardiac catheterization, **1903**
  in aortic regurgitation, 2045
  in aortic stenosis, 2038
  aortography during, 1907
  in cardiogenic shock, 2327
  coronary angiography during. *See* Coronary angiography
  hemodynamic assessment in, **1905**
    cardiac output, 1906
    in constrictive/restrictive conditions, 1905–1906, 1906t
    intracardiac shunts, 1907
    normal waveforms, 1905f
    reference values for, 1905t
    valve area, 1906–1907
    in valvular disease, 1905, 1906f
    vascular resistance, 1906
  indications for, 1903–1904, 1903t
  postprocedure care, 1909
  preprocedure management, 1903–1904
  in pulmonary hypertension, 2869
  racial/ethnic disparities in referral for, 62
  risks of, 1903–1904
  technique of, 1904–1905
  vascular access for, 1904
  ventriculography during, 1907, 1907f
Cardiac cirrhosis, 2712
Cardiac conduction system, 1844, 1868, 1868f, 1912
Cardiac contractility modulation, 1998
Cardiac contraction, **1847**
  cardiac activation in, 1849–1850, 1851f
  cardiac ultrastructure and, 1847
  contractile process in, 1847–1849, 1848f, 1849f, 1850f
  excitation-contraction coupling in, 1850f
  sliding filament model for, 1847–1849
Cardiac electrophysiology, **1910**. *See also* Cardiac arrhythmias
  catheter-based ablation. *See* Catheter-based ablation
  cellular basis of, 1910–1912, 1911–1912f
  electrocardiography. *See* Electrocardiography (ECG)
  historical perspective on, 1910
  implantable rhythm management devices. *See* Cardiovascular implantable electronic devices (CIEDs)
  impulse propagation in, 1912
  invasive testing in, 1915
Cardiac glycosides, 3712t
Cardiac imaging, **1876**
  after STEMI, 1893–1894, 1894f, 1902f
  artificial intelligence in, 1900–1901
  in CAD evaluation. *See* Coronary artery disease (CAD), diagnosis of
  in cardiac mass evaluation, 1898, 1900f
  in cardiac structure and function assessment, 1878–1879
  in congenital heart disease, 1899–1900, 1901f, 1902f. *See also specific defects*
  contrast agents for, 1883–1884
  CT. *See* Computed tomography (CT), in cardiac disease
  CT angiography. *See* Computed tomography (CT) angiography (CTA)
  diagnostic accuracy of, 1884t
  echocardiography. *See* Echocardiography
  in infectious and inflammatory disease, 1898–1899, 1900f, 1901f
  MRI. *See* Cardiac magnetic resonance imaging (CMR)
  in new-onset heart failure, 1894–1895, 1896f, 1897f, 1903v
  nuclear, 1878–1879. *See also* Myocardial perfusion imaging
  in pericardial disease, 1895–1898, 1897f, 1898f, 1899f, A9
  POCUS, 3911, 3912t, 3914v
  of prosthetic heart valves, 1907–1908, 1908f
  radiation exposure in, 1883
  in valvular heart disease, 1890–1893. *See also specific conditions*
Cardiac intensive care units, 2120–2121
Cardiac jelly, 1844
Cardiac magnetic resonance imaging (CMR)
  after STEMI, 1893, 1894f
  in aortic regurgitation, 1892, 1903v, 2044
  in atrial myxoma, 2086f
  in CAD evaluation, 1887, 1888f, 1902v
  in cardiac amyloidosis, 1894, 1896f, 2023, 2024f, A9
  in cardiac masses, 1898, 1900f
  in chest pain evaluation, 109, 1889–1890, 1889v, 1890f
  in congenital heart disease, 1900, 1902f
  in dilated cardiomyopathy, 1965, 1966f
  in heart failure, 1985
  in hypertrophic cardiomyopathy, 1894, 1896f, A9
  in iron overload, 1894
  in ischemic heart disease, 2095–2096f
  in left ventricular dysfunction, 1881, 1902v
  in metastatic lung cancer, A9
  in mitral regurgitation, 1893
  in myocarditis, 2012f
  in patient with pacemaker/defibrillator, 1884
  in pericarditis, 1897–1898, 1899f, 1903v, 2081f, A9
  principles of, 1880–1881
  pulse sequences for, 1881, 1881t
  in right ventricular dysfunction, 1882–1883, 1902v
  in STEMI evaluation, A9
  stress testing with, 290, 2095, 2095–2096f, A9
  in valvular heart disease, 1881
  in ventricular tachycardia, 1957, 1958f

- Cardiac mass, 1898, 1900f. *See also* Cardiac tumors
- Cardiac murmurs. *See* Heart murmurs
- Cardiac output
  - control of, **1850,** 2133
  - measurement of, 1906
  - oxygen delivery and, 2310
  - in pregnancy, 3898
  - in shock, 2292–2293
- Cardiac rehabilitation, 1988
- Cardiac resynchronization therapy, 1931, 1950, 1998–1999
- Cardiac risk assessment, 3905–3906, 3906f, 3907f, 3907t
- Cardiac sarcoidosis, 2925–2926. *See also* Sarcoidosis
- Cardiac-specific troponins. *See* Troponins
- Cardiac syncope
  - in arrhythmias, 160, 1914–1915
  - causes of, 155t
  - ECG in, 160
  - pattern of, 1841
  - in structural heart disease, 160
  - treatment of, 160, 2339t
- Cardiac tamponade, **2081**
  - cardiogenic shock in. *See* Cardiogenic shock
  - clinical features of, 2081–2082, 2081t
  - diagnosis of, 2082, 3914v
  - differential diagnosis of, 1905–1906, 1906t, 2081t, 2329t
  - dyspnea in, 269, 269t
  - ECG in, 1874, 1875f, 2081, 2081t, A7
  - malignant, 587
  - paradoxical pulse in, 2081–2082
  - pathophysiology of, 2081, 2082f
  - treatment of, 2082
- Cardiac thrombus, 1898, 1900f
- Cardiac transplantation, **2028**
  - allograft rejection in, 2029–2030
  - donor criteria and allocation for, 2028–2029
  - immunosuppression for, 2029–2030, 2029t
  - indications for, 2028, 2028t
  - infections after, 1159t, 1162f. *See also* Solid organ transplant recipient, infections in
  - late complications after, 1920, 2030
  - pig-to-human, 2033
  - sinus bradycardia following, 1920
  - surgical technique for, 2029
- Cardiac trauma, 2082, **2087**
- Cardiac tumors, **2084**
  - benign, 2086
  - clinical features of, 2086
  - echocardiography in, 1898, 1900f, 2085–2086, 2085–2086f, 2085t
  - imaging of, 2085, 2085t
  - metastatic, 247f, 1898, 2087, A9
  - primary, 1898, 2085–2086. *See also* Myxoma
  - sarcoma, 2086
- *Cardiobacterium hominis,* 1266, 1267t
- Cardioembolic stroke, 3444–3445. *See also* Ischemic stroke
- Cardio-fascio-cutaneous syndrome, 3788t
- Cardiogenic shock, **2324**
  - classification of, 2324–2325, 2326f
  - clinical features of, 2326
  - diagnosis of, 2326–2327, 2329t
  - etiology of, 2325–2326, 2327t
  - in free wall rupture, 2330
  - hemodynamic patterns in, 2311t, 2329t
  - history in, 2312
  - iatrogenic, 2331
  - incidence of, 2325
  - in mitral regurgitation, 2330
  - in myocarditis, 2330
  - pathophysiology of, 2325, 2325f
  - patient profile in, 2325
  - physical examination in, 2312
  - prognosis of, 2329
  - in right ventricular failure, 2329–2330
  - in STEMI, 2123
  - timing of, 2325–2326
  - treatment of, 2315, 2327–2329, 2328f
  - in ventricular septal rupture, 2330
- Cardiolipin, 475, 2790t
- Cardiomyocytes, 1847, 1848f
- Cardiomyopathy, **2001**
  - arrhythmogenic right ventricular, 1961f, 1966, 2009, 2010
  - cardiogenic shock in. *See* Cardiogenic shock
  - in Chagas disease, 1790, 1791, 1791f
  - classification of, 2001, 2002t
  - clinical features of, 2001, 2002t
  - definition of, 2001
  - dilated. *See* Dilated cardiomyopathy
  - dyspnea in, 270
  - genetic considerations in, 2003–2004, 2004t, 2005f
  - HIV-associated, 1598, 2012
  - hypertrophic. *See* Hypertrophic cardiomyopathy
  - ICD indications in, 2339t
  - in inherited disorders of metabolism, 2010, 2010f
  - initial evaluation of, 2003t
  - iron-overload, 1894, 1903v
  - ischemic, 2092
  - in low- and middle-income countries, 3851
  - in metabolic disorders, 2010
  - peripartum, 2017–2018, 3900
  - restrictive. *See* Restrictive cardiomyopathy
  - in rheumatoid arthritis, 2840
  - sudden cardiac death and, 2335, 2335t
  - syncope in. *See* Cardiac syncope
  - Takotsubo (stress-related), 104, 2019, 2019f
  - toxic, 2015–2017
  - ventricular tachycardia in, 1965, 1966f, 1971
- Cardiopulmonary bypass, 923, 2376, 3770t
- Cardiopulmonary exercise testing (CPET), 271, 292, 2189
- Cardiopulmonary resuscitation, 88, 2336, 3769
- Cardiorenal syndrome, 1981–1982, 1991, 2361f
- Cardiovascular collapse, 2332, 2333t. *See also* Cardiac arrest
- Cardiovascular disease
  - air pollution and, 3739–3740, 3741f
  - in AL amyloidosis, 899
  - in alcohol use disorder, 3677
  - in amyloidosis, 897
  - approach to the patient, **1841**
  - arrhythmias in. *See* Cardiac arrhythmias
  - in β thalassemia, 778t
  - biology of, **1843**
  - cardiac symptoms in, 1841
  - circadian clock and, 3952–3953
  - in CKD. *See* Chronic kidney disease (CKD), cardiovascular disorders in
  - congenital. *See* Congenital heart disease (CHD)
  - coronary artery. *See* Coronary artery disease (CAD)
  - cyanosis in, 278
  - deaths from, 75t, 498
    - vs. other causes, 1856f
    - sex differences in, 3163f
    - worldwide by income group, 1856–1857, 1856f, 1857f, 3846t
  - depression in, 3664
  - in diabetes mellitus
    - cardiomyopathy, 2018
    - heart failure, 1986, 1986t, 1998
    - hypertension, 2147
    - ischemic heart disease, 2100
    - morbidity and mortality data, 3227
    - risk factors for, 3228
    - treatment and prevention of, 3228
  - diagnosis of, 1841–1842
    - imaging tests for. *See* Cardiac imaging
    - physical examination. *See* Cardiovascular system, physical examination of
    - pitfalls in, 1843
    - in women vs men, 3164–3165
  - dyspnea in, 269, 269t, 270, 271
  - edema in, 281, 281t
  - epidemiologic transition of, 1854–1856, 1855t
  - epidemiology of, **1854,** 1856f
  - erectile dysfunction in, 3155
  - functional classification of, 1841–1842, 1842t, 2093t
  - global burden of, 1856–1857, 1856f, 3846t, 3850
  - heart failure. *See* Heart failure (HF)
  - in hemochromatosis, 1894, 1927, 2018–2019, 3336
  - hypertension treatment in, 2146
  - in hypothermia, 3769t
  - in IBD, 2562t, 2564
  - immunizations recommendations for, 1004t
  - impact of COVID-19 pandemic on, 1858
  - ischemic. *See* Ischemic heart disease
  - ischemic stroke in, 3444–3445, 3444t, 3445t
  - as late consequence of cancer/cancer therapy, 2016–2017
  - in Loeys-Dietz syndrome, 3332
  - in low- and middle-income countries, 3846t, 3850–3851
  - in Marfan's syndrome, 3332
  - MASLD and, 2702
  - in megaloblastic anemia, 784
  - metabolic syndrome and, 3254
  - microbiome and, 3836
  - myocardial infarction. *See* Myocardial infarction (MI)
  - natural history of, 1841
  - obesity and, 3187
  - in osteogenesis imperfecta, 3324
  - in Paget's disease of bone, 3313
  - parasitic infections, 1739t
  - in pregnancy, 3899–3900
  - prevention of, 1843, 3851
  - rheumatic. *See* Rheumatic fever, cardiovascular complications of
  - in rheumatoid arthritis, 2840–2841
  - risk factors for, 1857–1858, 3677. *See also* Atherosclerosis
    - diabetes mellitus, 1858. *See also* Diabetes mellitus (DM)
    - diet, 1857
    - family history, 1842
    - genetic, 1858
    - hypercholesterolemia, 1858. *See also* Lipoprotein disorders
    - hypertension, 1858. *See also* Hypertension
    - metabolic syndrome, 1841
    - obesity, 1858. *See also* Obesity
    - physical inactivity, 1857
    - smoking, 1857. *See also* Smoking
    - in women, 3164–3165, 3165f
  - in sepsis/septic shock, 2322–2323, 2322t
  - smoking and, 3683, 3683t
  - sudden cardiac death and, 2324–2334, 2334f
  - syncope in. *See* Cardiac syncope
  - in syphilis. *See* Syphilis, secondary
  - in systemic sclerosis, 2871
  - testosterone therapy and, 3122–3123
  - thyroid disease and. *See* Hyperthyroidism; Hypothyroidism
  - treatment of, 1843
  - valvular. *See* Valvular heart disease
  - in women, 3164–3165, 3165f

- Cardiovascular health, 9
- Cardiovascular implantable electronic devices (CIEDs)
  - CMR in patient with, 1884
  - complications of, 1958
  - contraindications to, 2341, 2341t
  - for coronary artery disease, 2339t
  - for heart failure, 1999, 1999t, 2339t
  - for hypertrophic cardiomyopathy, 2008, 2339–2340t
  - indications for, 1958, 2339–2340, 2339–2340t
  - infective endocarditis in
    - antimicrobial therapy for, 1045
    - diagnosis of, 1039–1040
    - etiology of, 1037, 1037t
    - incidence of, 1037
    - surgical treatment of, 1047
  - shocks from, 1972, 1973f
  - for STEMI, 2124, 2124f
  - for ventricular arrhythmias, 1917, 1958–1959, 1959f, 2339–2340t
    - after STEMI, 1961, 1965, 2338
    - in dilated cardiomyopathy, 1965
    - electrical storm, 1972
- Cardiovascular system
  - biology of, **1843**
    - assessment of cardiac function, 1852–1854, 1853f
    - blood vessels. *See* Blood vessels
    - cardiac contraction. *See* Cardiac contraction
    - control of cardiac performance and output, 1850–1852, 1852f, 1852t
    - developmental, 1843–1844, 1845f
    - in hypoxia, 277
  - imaging of. *See* Cardiac imaging
  - physical examination of, **1859**
    - abdomen, 1860
    - approach to, 1859
    - arterial pulses, 1862–1863, 1862–1863f
    - auscultation. *See* Cardiac auscultation
    - blood pressure, 1862
    - chest, 1860
    - extremities, 1860
    - head and neck, 1860
    - heart inspection and palpation, 1864
    - jugular venous pressure and waveform, 1860–1861, 1861f
    - skin, 1859–1860
  - in sleep, 211
- Cardioversion
  - anticoagulation before, 1949
  - for atrial fibrillation, 1949
  - for atrial flutter, 1944
  - for atrial tachycardia, 1938
  - for preexcited tachycardias, 1941
- Caregiver(s), 87
- Care transitions, for older adults, 3876, 3895, 3896t
- Carey-Coombs murmur, 288
- Carfilzomib
  - action and targets of, 533t, 536, 570t, 573
  - adverse effects of, 570t, 573, 754, 2017
  - for multiple myeloma, 536, 890t, 891–892
  - for Waldenström's macroglobulinemia, 894
- Carglumic acid, 3372
- Caries, dental, 260, 1374, A3
- Cariprazine, 3673, 3673t
- Carisoprodol, 3713t
- Carmustine (BCNU), 726t
- Carney syndrome/complex (LAMB/NAME syndrome)
  - adrenal nodules in, 3055, 3063, 3088
  - clinical features of, 3004, 3004t, 3088
  - genetic considerations in, 3004, 3080t, 3088
  - myxoma in, 2085
  - skin manifestations of, 399, 1860
- Carnitine metabolism disorders, 2010, 2018
- Carnitine palmitoyltransferase 2 (CPT2) deficiency, 3645
- Caroli's disease/syndrome, 667, 2434, 2719
- Carotenemia, 2612
- Carotenoderma, 320
- Carotenoids, 2610
- Carotid artery, 1585, 1844
- Carotid artery dissection, 131, 3445
- Carotid artery pulse, 1862, 1862f
- Carotid atherosclerosis. *See also* Atherosclerosis
  - asymptomatic, 3451
  - risk factors for, 3451
  - screening for, 3881t
  - stroke due to, 3445, 3446t
  - treatment of, 2131, 2131f, **3451**
- Carotid body, 267, 268f
- Carotid cavernous fistula, 233
- Carotid sinus hypersensitivity, 1920, 1927
- Carotid sinus massage, 161, 1942
- Carpal spasm, 365
- Carpal tunnel syndrome
  - in acromegaly, 2966
  - in amyloidosis, 900
  - clinical features of, 3613
  - conditions associated with, 2943
  - diagnosis of, 3613
  - in familial amyloid polyneuropathy, 3603
  - in hypothyroidism, 3030
  - neck pain in, 133
  - in pregnancy, 3903
  - in systemic sclerosis, 2871
  - treatment of, 3613
  - wrist pain in, 2943
- Carpenter's syndrome, 3184t
- Carpet beetle, 3737
- "Carpet tacking," 415
- Carriers, 2366
- Carrión's disease, 1351t, 1353t, 1356
- CAR T cells. *See* Chimeric antigen receptor (CAR) T cells
- Cartilage
  - joint failure related to, 2950, 2950f
  - regeneration of, for osteoarthritis, 2956
- Cartilage hair hypoplasia, 457, 2800, S9
- Cartilage oligomeric matrix protein (COMP), 3321t
- Carvajal syndrome, 2004t
- Carvallo's sign, 285, 286, 1866, 2049
- Carvedilol
  - adverse effects of, 2144t
  - genetic variations in response to, 492, E72t
  - for heart failure, 1995t
  - for hypertension, 2144t
  - metabolism of, 485t
  - overdosage/poisoning with, 3712t
- Casal's necklace, 2608
- Cascara, 307, 312
- Case-control study, 3797t
- Caseous necrosis, 1384
- Case reports, 27
- *CASP1* gene, 2761t
- *CASP5* gene, 2762t
- *CASP8* gene, 713
- CASP10, 2786f
- Caspase, 539, 540f, 2781
- Caspase-1, 972
- Caspase-8, 539, 557f
- Caspase-9, 539, 557f
- Caspofungin
  - for *Candida* infections, 1710t, 1711t, 1712, 1712t
  - for fungal infections in cancer patient, 833
  - indications for, 1692
  - for neutropenic fever, 581
  - resistance to, 1711t
- Caspr2 antibodies, 743t, 746f, 748, 751, 3627
- *CaSR* gene mutations, 363, 2369, 2744, 3084, 3279, 3287
- Castell's method, splenic percussion, 480
- Castelo dos Sonhos virus, 1661t, 1674
- Castleman's disease
  - AA amyloidosis and, 899
  - follicular dendritic cell carcinoma and, 882
  - HHV-8 infection and, 857, 1516
  - in HIV infection, 1614
  - paraneoplastic pemphigus and, 411
  - vs. TB, 1387
  - treatment of, 1517
- Cast(s), urinary, 343, A4
- Cat(s)
  - bite-wounds, **1143**
    - approach to the patient, 1144
    - cellulitis in, 1051
    - infections from, 1143, 1267
    - microbiology of, 1143
    - prophylaxis for, 1145–1146
    - treatment of, 1054t, 1144, 1145t
  - cat-scratch disease. *See* Cat-scratch disease
  - fleas, 1351, 3736
  - tapeworms, 1839
  - in *T. gondii* life cycle, 1797, 1797f
- Catacamas virus, 1661t, 1674
- Catagen, 3138
- Catalepsy, 187
- Catamenial epilepsy, 3427
- Cataplexy, 161, 214–215, 220v
- Cataract
  - etiology of, 230
  - glucocorticoid-related, 2836
  - in HCT recipient, 916
  - radiation-induced, 554
  - sunflower, 3339
  - surgical treatment of, 230
  - visual loss in, 230
- Catatonia, 187
- Catecholamine(s), 504t, 3074, 3074t
- Catecholaminergic polymorphic ventricular tachycardia, 160, 1971, 2340t
- Catechol-*O*-methyltransferase (COMT) inhibitors, 3503, 3504t
- Caterpillar stings/dermatitis, 3737
- Cathartics, 3709
- Cathepsin K, 3315
- Catheter-associated urinary tract infection, 1087, 1090, 1094, 1147, 1148t
- Catheter-based ablation
  - for accessory pathways, 1941–1942
  - for atrial fibrillation, 1950, 1952–1953f, 1952–1954, 1954t
  - for atrial flutter, 1944
  - for atrial tachycardia, 1938
  - for AV nodal reentrant tachycardia, 1940
  - for cardiac arrhythmias, 1914, 1916–1917, 1916f
  - complications of, 1920, 1953–1954
  - for PVCs, 1960, 1962f
  - for sinus tachycardia, 1936
  - for ventricular arrhythmias, 1959, 1965, 1966t, 1971, 1972
- Catheter-related bloodstream infections
  - *A. baumannii,* 1296
  - in cancer patient, 581
  - enterococcal, 1217–1218
  - epidemiology of, 1148
  - gram-negative bacteria in, 1269
  - health care–associated, 1148
  - in parental nutrition, 2624
  - pathophysiology of, 1148
  - prevention of, 43, 1148, 1148t
- Cathinones, 3693, 3696–3697. *See also* Psychostimulants

Cation-exchange resin, for hyperkalemia, 361
Cat-scratch disease, **1351**
  approach to the patient, 1353
  clinical features of, 477, 1051, 1351–1353, 1352f
  diagnosis of, 1353
  epidemiology of, 1351
  etiology of, 1351
  in HIV infection, 1610
  pathogenesis of, 1351
  prevention of, 1354
  treatment of, 1353, 1353t
Catu virus, 1658t
Cauda equina syndrome, 120, 126t, 170, 171, 589, 2454
Causalgia, 96
Cautious gait, 177t, 178
Cavernous angiomas, 3453t, 3457
Cavernous malformations, 3389f
Cavernous sinus syndrome, 3558, 3558f
Cavernous sinus thrombosis, 994, 1142, 3558
Cavotricuspid isthmus-dependent atrial flutter, 1943
CBFA2 gene mutations, 3783t
CBG (corticosteroid-binding globulin), 3128
CBG (cortisol-binding globulin), 3053–3054
CBL gene mutations, 877
CBT. *See* Cognitive behavioral therapy (CBT)
CBT (core body temperature), 3944t, 3946
CCBE1 gene mutations, 2183
CCBs. *See* Calcium channel blockers (CCBs)
C-C chemokine family, 456, 2773
CCDC134 gene mutations, 3325t, 3327
CCGs (clock-controlled genes), 3943, 3944t
CCK (cholecystokinin), 2742, 3182, 3182f
CCK (cholecystokinin)–releasing factor, 2743
CCL2 gene, 1581t, 1585
CCL3 gene, 1581t, 1585
CCL3L1 gene, 1582t
CCL3L gene, 1581t
CCL4L gene, 1581t
CCL5 gene, 1581t
CCM2 gene mutations, 3457
CCND1 gene mutations, 520t, 660, 661t
CCNE1 gene mutations, 660
CCNU (lomustine), 562t, 718
CCPD (continuous cyclic peritoneal dialysis), 2398, 2400. *See also* Peritoneal dialysis
CCR1, 2767t, 2769t
CCR2, 1581t, 1583f, 2767t, 2769t
CCR3, 1583f, 2767t, 2769t
CCR4, 2767t, 2769t
CCR5 receptor
  in HIV infection
    pathogenesis, 461, 1577, 1578f, 1579, 2784
    susceptibility and response, 1581t, 1582t, 1583, 1583f
    viral replication cycle, 1557–1558, 1560f
  ligands, cell types, and disease connection, 2769t
  maraviroc response and, 1583, E72t
  source, cell target, and biologic activity, 2767t
CCR6, 2767t, 2769t
CCR6 gene mutations, 2554t
CCR7, 2767t, 2769t
CCR8, 2767t, 2769t
CCR9, 2768t, 2769t
CCR10, 2769t
CCRL2, 1581t, 1585
CD1a (T6, HTA-1), 2757t
CD1b, 2757t
CD1c, 2757t
CD1d, 2757t
CD2 (T12, LFA-2), 2757t
CD3 (T3, Leu-4), 2757t
CD4 (T4, Leu-3), 2757t
CD4 molecule, 1557–1558, 1560f
CD4+ T cells
  in adaptive immunity, 1482, 2764f, 2779–2780
  in CLL, 852
  definition of, 2755
  in drug reactions, 419
  epigenetics of, 3942
  functions of, 2774
  in giant cell arteritis, 2902
  in HBV infection, 2649
  in HCV infection, 2650
  helper, 2756
  in HIV infection
    dysfunction and depletion, 1569–1570, 1572, 1578–1579
    early effects, 1569–1570
    early events, 1570f
    HIV-specific, 1588
    latently infected, resting, 1572, 1572f
    for monitoring immunologic competence, 1590, 1591f
    viral invasion, 1558, 1559f, 1571
  in IgG4-related disease, 2933
  in immunity to abscesses, 1073
  in *M. tuberculosis* infection, 1385
  in NTM infections, 1415
  in rheumatoid arthritis, 2844, 2845f
  in sepsis/septic shock, 2319
  subsets of, 2756
  in *T. gondii* infection, 1799
  in transplant rejection, 2407, 2407f
  in viral infections, 1482, 1482f
CD4+ T lymphocytopenia, idiopathic, 1614–1615
CD7 (3A1, Leu-9), 2757t
CD8 (T8, Leu-2), 2757t
CD8+ T cells
  in adaptive immunity, 1482, 2764f
  in COPD, 2200
  definition of, 2755
  in drug reactions, 419
  epigenetics of, 3942
  in HBV infection, 2649
  in HCV infection, 2650
  in hemophagocytic lymphohistiocytosis, 465
  in HIV infection, 1571, 1578, 1588
  in HSV infection, 1497
  in *L. monocytogenes* infection, 1227
  in *M. tuberculosis* infection, 1385
  in sepsis/septic shock, 2319
  in spondyloarthritis, 2880–2882
  in *T. gondii* infection, 1798, 1799
  in transplant rejection, 2407, 2407f
  in viral infections, 1482, 1482f
CD14 (LPS-receptor), 2757t, 2758
CD16 (FcγRIIIa), 2768
CD16a (FcγRIIIa), 2757t
CD18 gene mutations, 458
CD19 (B4), 535t, 862, 2757t, 2788, 2788f, 3829
CD20 (B1), 534t, 558, 2757t, 2791t
CD21 (B2, C2, EBV-R, C3dR), 2757t
CD22 (BL-CAM), 536, 2757t
CD23 (FcεRII, B6, Leu-20, BLAST-2), 2757t
CD25, 504t, 2786f, 2789, 2791t
CD25 deficiency, 2803
CD28, 2757t, 2784
CD30, 504t, 536
CD32 (FcγRII), 2757t
CD33, 536
CD34 antigen, 915
CD34+ stem cells, 776, 778, 822
CD40, 2757t
CD40 deficiency, 2800
CD40 ligand, 938, 2800
CD40 ligand deficiency, 458
CD40 surface antigen, 938
CD45 (LCA, T200, B220), 2757t
CD45RA, 2757t
CD45RB, 2757t
CD45RC, 2757t
CD45RO, 2757t
CD52, 534t, 558, 2791t
CD56, 616, 2768
CD64 (FcγRI), 2757t
CD79b, 536
CD80 (B7-1, BB1), 2757t, 2791t
CD86 (B7-2, B70), 2757t, 2791t
CD89 (FCaR), 2757t
CD95 (APO-1, Fas), 2758t
CD112 (nekton-2, PVRL2), 2758t
CD122 deficiency, 2803
CD134 (OX40), 2758t
CD152 (CTLA-4), 2758t
CD154 (CD40L), 2758t
CD155 (PVR), 2758t
CD223 (LAG-3), 2758t
CD226 (DNAM-1), 2758t
CD252 (OX40L), 2758t
CD272 (BTLA), 2758t
CD274 (PD-L1), 2758t
CD278 (ICOS), 2758t
CD279 (PD-1), 2758t, 2777
CD319, 2792t
CD357 (GTTR), 2758t
CDC. *See* Centers for Disease Control and Prevention (CDC)
CDC73 gene mutations, 3084
CDCA (chenodeoxycholic acid), 2727
CD (cluster of differentiation) classification, 2755, 2757–2758t
CDH1 gene mutations, 524t, 646, 691
CDI. *See Clostridioides difficile*–associated infection (CDI)
CDK4 gene mutations, 599
CDK5RAP2, 3402
CDK12 gene mutations, 710
CDK inhibitors, 570t, 573, 753t, 757
CDKN1B gene mutations, 3004t
CDKN1C, 761
CDKN1C gene mutations, 3792
CDKN2A gene mutations
  in bladder cancer, 691
  in brain tumors, 717
  in carcinoma of unknown primary, 734
  in epigenetic differentiation, 3939
  in familial melanoma, 524t, 599
  in gastric adenocarcinoma, 646f
  in HCC, 660, 661t
  in lung cancer, 617
  in pancreatic cancer, 671t, 672
  in skin cancer, 431
CDNKIB gene mutations, 3080t
CDNKL-5 gene mutations, 3412t
CDX-2, in carcinoma of unknown primary, 733, 733t
CEA (carcinoembryonic antigen), 504t, 514, 639, 658
CEACAMs, 967
CEAP classification, chronic venous insufficiency, 2181, 2181t
CEBPA gene mutations, 825, 825t, 834
CEBPE gene mutations, 460
Cefaclor, 403, 419, 1175
Cefadroxil, 258t, 1175, 1203t
Cefamandole, 1175
Cefazolin
  for CAPD peritonitis, 1072
  for cellulitis, 1054t
  for endocarditis prophylaxis, 1048t
  indications for, 1173t, 1175, 1180t
  for infective endocarditis, 1043t
  for intraabdominal infections, 963t
  for nongonococcal bacterial arthritis, 1057
  prophylactic uses of, 1180t

    resistance to, 1173t
    for staphylococcal infections, 1203t
Cefdinir, 1175
Cefepime
    for *A. baumannii* infections, 1296
    adverse effects of, 1175
    for CAPD peritonitis, 1072
    for *Citrobacter* infections, 1293
    for community-acquired pneumonia, 1028t
    empirical uses of, 2321t
    for *Enterobacterales* infections, 1283–1284
    for *Enterobacter* infections, 1292
    for hospital-acquired pneumonia, 963t
    indications for, 1173t, 1175
    for intraabdominal infections, 963t
    for meningitis, 1121t, 1122, 1122t
    for neutropenic fever, 580
    for nongonococcal bacterial arthritis, 1057
    for osteomyelitis, 1064t
    for *P. aeruginosa* infections, 1309t
    for peritonitis, 1071
    for *Proteus* infections, 1291
    for pyelonephritis, 1094
    resistance to, 1173t, 1183, 1292
    for sepsis/septic shock, 963t, 991t
    for *Serratia* infections, 1293
    for ventilator-associated pneumonia, 1032t
Cefiderocol
    for *A. baumannii* infections, 1296–1297, 1297t
    for *E. coli* infections, 1287
    for *Enterobacterales* infections, 1283, 1284
    indications for, 1174t
    for *Klebsiella* spp. infections, 1291
    for *P. aeruginosa* infections, 1309t, 1313
    in pregnancy and lactation, 1169t
    resistance to, 1182t, 1183, 1183f, 1314
    for ventilator-associated pneumonia, 1032
Cefixime
    for enteritic fever, 1317–1318
    for gonococcal infections, 1259t
    indications for, 1175
    in pregnancy and lactation, 1169t
    resistance to, 1100f, 1255, 1259
Cefoperazone, 1175
Cefotaxime
    for community-acquired pneumonia, 963t
    for enteric fever, 1318
    for epiglottitis, 1263
    for gonococcal infections, 1259t
    indications for, 1175
    for infective endocarditis, 1042
    for intraabdominal infections, 963t
    for leptospirosis, 1444, 1444t
    for Lyme disease, 1453, 1453f
    for meningitis, 1122, 1122t, 1194, 1263
    for meningococcal infections, 1249
    for nongonococcal bacterial arthritis, 1057
    for peritonitis, 329, 1070
    for pneumococcal infections, 1026, 1192
    for yersiniosis, 1350
Cefotetan
    adverse effects of, 802t
    indications for, 1173t, 1175
    for pelvic inflammatory disease, 1106, 1106t
    prophylactic uses of, 1180t
    resistance to, 1173t, 1284, 1379
Cefoxitin
    for anaerobic bacterial infections, 1379
    for bite-wound infections, 1145t
    for gas gangrene, 1054t, 1239t
    indications for, 1173t, 1175
    for intraabdominal infections, 963t
    for NTM infections, 1428
    for pelvic inflammatory disease, 1106, 1106t
    prophylactic uses of, 1180t
    resistance to, 1173t, 1284, 1379
Cefpirome, 1175
Cefpodoxime, 1027t, 1144, 1169t, 1175
Cefprozil, 1175
Ceftaroline
    actions of, 1204
    adverse effects of, 1175
    for community-acquired pneumonia, 1028t
    indications for, 1173t, 1175
    for infective endocarditis, 1044
    for MRSA infections, 1183, 1203t, 1204
    resistance to, 1173t, 1183
Ceftazidime
    for CAPD peritonitis, 1072
    for community-acquired pneumonia, 1028t
    for febrile neutropenia, 1309t
    for glanders, 1309t
    for hospital-acquired pneumonia, 963t
    indications for, 1173t, 1175
    for intraabdominal infections, 963t
    for melioidosis, 1309t
    for meningitis, 1121t, 1122, 1122t
    for neutropenic fever, 580
    for nongonococcal bacterial arthritis, 1057
    for osteomyelitis, 1064t
    for *P. aeruginosa* infections, 1309t
    for peritonitis, 1070, 1071
    resistance to, 1173t
    for ventilator-associated pneumonia, 1032t
Ceftazidime-avibactam
    indications for, 1173t, 1175, 1283, 1284
    for *Klebsiella* spp. infections, 1291
    for *P. aeruginosa* infections, 1309t, 1313
    for peritonitis, 1070, 1072
    in pregnancy and lactation, 1169t
    for *Proteus* infections, 1291
    resistance to, 1173t
    for ventilator-associated pneumonia, 1032
Ceftizoxime, 1259t
Ceftobiprole, 1202, 1203t, 1204–1205
Ceftolozane-tazobactam
    indications for, 1173t, 1175
    for *Klebsiella* spp. infections, 1291
    for *P. aeruginosa* infections, 1309t, 1313
    for peritonitis, 1070, 1072
    in pregnancy and lactation, 1169t
    for *Proteus* infections, 1291
    resistance to, 1173t
    for ventilator-associated pneumonia, 1032
Ceftriaxone
    for actinomycosis, 1365, 1366t
    adverse effects of, 802t
    for *Bartonella* endocarditis, 1353t
    for bite-wound infections, 1144, 1145
    for *Brucella* infections, 1334
    for Carrión's disease, 1353t
    for chancroid, 1109t, 1264
    for community-acquired pneumonia, 963t, 1028t
    drug interactions of, 1172t
    empirical uses of, 963t, 991t, 2321t
    for enteric fever, 1317–1318, 1318t
    for enterococcal infections, 1219, 1219t
    for epididymitis, 1099
    for epidural abscess, 963t, 991t
    for epiglottitis, 1263
    for gonococcal arthritis, 1058
    for gonococcal infections, 1099, 1099t, 1258–1259, 1259t
    for head and neck abscesses, 259
    indications for, 1173t, 1175
    for infective endocarditis, 963t, 991t
        Bartonella, 1044t
        enterococcal, 1042
        HACEK, 1043t, 1266–1267, 1267t
        streptococcal, 1043t
    for intraabdominal infections, 963t
    for intracranial infections, 991t
    for leptospirosis, 1444, 1444t
    for Lyme disease, 1058, 1453
    for meningitis
        H. influenzae, 1263
        as initial empirical therapy, 963t, 991t, 1121t, 1122
        N. meningitides, 1122, 1122t
        S. pneumoniae, 1122–1123, 1122t, 1194
    for meningococcal infections, 1249
    for meningococcal prophylaxis, 1253
    for *Nocardia* infections, 1362
    for nongonococcal bacterial arthritis, 1057
    for nontyphoidal *Salmonella* infections, 1321t
    for osteomyelitis, 1064t, 1065t
    for otitis media, 254
    for pelvic inflammatory disease, 1106, 1106t, 3137
    for peritonitis, 1070, 1071
    for pneumococcal infections, 1026, 1192
    for post-splenectomy sepsis, 991t
    in pregnancy and lactation, 1169t
    for proctitis, 1109
    prophylactic uses of, 1048t, 1122
    for purpura fulminans, 991t
    for pyelonephritis, 1094
    for relapsing fever, 1447, 1447f
    resistance to, 1100f, 1173t, 1255, 1319, 1359f
    for *Shigella* infections, 1324t
    for syphilis, 1435
    for Whipple's disease, 1370
    for yersiniosis, 1350
Cefuroxime
    for community-acquired pneumonia, 1027t
    indications for, 1175
    for intraabdominal infections, 963t
    for Lyme disease, 1453, 1453f
    for pneumococcal infections, 1192
    in pregnancy and lactation, 1169t
    for sinusitis, 256
Celecoxib
    adverse effects of, 97–98
    genetic variations in response to, E72t
    for gout, 2959
    for osteoarthritis, 2955
    for pain, 96–98, 97t
Celiac disease (gluten-induced enteropathy), **2545**
    clinical features of, 2545
        ataxia, 3533
        diarrhea, 309, 2545
        dyspepsia, 300
        hypokalemia, 355
        neuropathy, 3605
        oral, 261
    cobalamin malabsorption in, 787
    complications of, 2545, 3092, 3115
    diagnosis of, 2545
        barium radiography in, 2544f
        endoscopic, 2495, 2500f
        small-intestinal mucosal biopsies in, 2547f, 2550, 2550t
        tTg antibody test in, 3092
    diseases associated with, 2545
        APS-2, 3090t, 3091, 3091t
        dermatitis herpetiformis, 412, 2545
        diabetes, 2545
    enteropathy-type T cell lymphoma and, 867, 874
    epidemiology of, 309, 2545
    genetic considerations in, 2545, 3091t
    vs. IBS, 2574
    immunology of, 2781, 2790t
    inflammasome mutations in, 2761t
    pathology of, 309
    refractory, 2545–2546
    treatment of, 301, 311, 2546

- Celiac disease (gluten-induced enteropathy) (*Cont.*):
  - vitamin A deficiency in, 2611
  - vitamin K deficiency in, 2613
- Celiprolol, 2167
- Cell cycle
  - in cancer cells, 529t, 530–531
  - checkpoints, 521, 530–531
  - DNA replication in, 3779–3780, 3780f
  - segregation of mtDNA in, 3808
- Cell-free DNA (cfDNA)
  - clinical applications of, 3980, 3981f
  - contribution of clonal hematopoiesis of indeterminate potential to, 3982f
  - in oncology, 3980–3981, 3981f. *See also* Circulating tumor DNA (ctDNA)
  - in prenatal medicine, 3987–3989
  - in transplantation
    - for detection of acute allograft rejection, 3983f, 3984–3985, 3985f
    - for monitoring host immunity, 3986–3987
    - for monitoring microbial diversity and infection, 3985–3986, 3986f, 3987f
- Cell-free RNA, 3988
- Cell growth/death pathways, 556, 557f
- Cell lysis, in hypothermia, 3769
- Cell-mediated immunity, 1384
- Cell necrosis, 2755
- Cell stress responses, 3991–3992, 3992f
- Cellular metabolism, genetic defects of, 2004t
- Cellular senescence, 531, 539, 3871
- Cellular transport, 2365–2366
- Cellulitis
  - anaerobic bacterial, 1377
  - in cancer patient, 583
  - clostridial, 1243
  - crepitant, 1243
  - diagnosis of, 50, 1052, 3914v
  - differential diagnosis of, 405
  - *E. coli,* 1286
  - etiology of, 1050t, 1052
  - *H. influenzae,* 1262
  - *Klebsiella,* 1290
  - *Nocardia,* 1360, 1362t
  - orbital, 233
  - pathophysiology of, 1049f, 1052
  - periorbital, 1052
  - sterile, 421
  - streptococcal, 1052, 1210–1211
  - treatment of, 1052, 1054t, 1208t, 2321t
- Cell wall synthesis inhibitors, 1165–1166, 1182t, 1183f
- Cementum, 260
- Cemiplimab, 534t, 559f, 596, 608, 630t, 753t, 2351
- Cenobamate, 3421t, 3424
- Centers for Disease Control and Prevention (CDC)
  - on antibiotic resistance threats, 1186, 1186t
  - antimicrobial stewardship guidelines of, 1179
  - infection-control guidance and oversight, 1147
  - opioid prescribing guidelines for chronic pain, 101t
  - opioid prescribing guidelines of, 100
  - Vaccines Information Statements site, 1007
- Centipede bite, 3737
- Central airway obstruction, 2288, 2289f, 2289t
- Central auditory processing disorders, 247
- Central cord syndrome, 3561
- Central cyanosis, 278–279, 278t, 1859. *See also* Cyanosis
- Central fever, 189
- Central hypoventilation syndrome, 2276
- Central nervous system (CNS)
  - cytokine production in, 135
  - in hypothermia, 3769t
  - in hypoxia, 277
  - lymphatic structures of, 3397–3398
  - paraneoplastic disorders of. *See* Paraneoplastic syndromes, neurologic
  - vasculitis of. *See* Vasculitis, primary CNS
- Central nervous system (CNS) cancer. *See also specific types and sites*
  - brain tumors. *See* Brain tumors
  - deaths from, 497t, 500f
  - dementia in, 3484
  - incidence of, 497t
  - lymphoma. *See* Primary central nervous system lymphoma (PCNSL)
  - metastatic, 590
- Central nervous system (CNS) infections
  - *Actinomyces,* 1365
  - anaerobic bacterial, 1375, 1379
  - *Aspergillus,* 1714, A16
  - brain abscess. *See* Brain abscess
  - in cancer patient, 583
  - cranial epidural abscess, 1141, 1141f
  - dementia due to, 3484
  - empyema. *See* Subdural empyema
  - encephalitis. *See* Encephalitis
  - intracranial, 991t, 994
  - *L. monocytogenes,* 1228
  - meningitis. *See* Meningitis
  - *M. tuberculosis,* 1388–1389
  - mucormycosis, 1718–1719, 1719f
  - mumps, 1648
  - neurocysticercosis. *See* Neurocysticercosis
  - *Nocardia,* 1359–1360
  - nontyphoidal *Salmonella,* 1320
  - *P. aeruginosa,* 1309t, 1311
  - rabies. *See* Rabies
  - sarcoidosis, 1132t
  - sporotrichosis, 1724, 1724t
  - *T. gondii,* 1800, 1800f, 1801t
  - in transplant recipient, 2350–2351
  - treatment of, 1379
  - *T. whipplei,* 1368
  - VZV, 1504. *See also* Herpes zoster
- Central pontine myelinolysis (osmotic demyelination syndrome), 350, 352, 2348–2349, 2349f, 3461, S1
- Central precocious puberty, 3109. *See also* Precocious puberty
- Central retinal artery occlusion, 225, 226f
- Central retinal vein occlusion, 226, 226f
- Central sensitization, 93
- Central sleep apnea, 213, 1982, 1985, 2280f, 2282
- Central tolerance, 2789, 2789t
- Central transtentorial herniation, 187f, 188
- Central venous catheter
  - infections related to. *See* Catheter-related bloodstream infections
  - insertion of, 3913–3914
  - for parenteral nutrition, 2624
  - in shock, 2314
  - superior vena cava syndrome and, 585, 586
- Central venous pressure (CVP), 1860–1861, 2295
- Centromere-associated proteins, 2790t, 2864t
- Cepacia syndrome, 1309t, 1313
- Cephalexin
  - for asymptomatic pharyngeal GAS carriage, 1209
  - for atopic dermatitis, 383
  - for endocarditis prophylaxis, 1048t
  - for impetigo, 1210
  - indications for, 1175
  - in pregnancy and lactation, 1169t
  - for staphylococcal skin/soft tissue infections, 964t, 1203t
  - for streptococcal pharyngitis, 258t
- Cephalosporins
  - actions of, 1166, 1182t
  - adverse effects of, 395, 411, 422, 457, 1175, 2377
  - clinical use of, 1175
  - for community-acquired pneumonia, 1027t
  - cross-sensitivity to, 427
  - for diverticular disease, 2580
  - for gonococcal infections, 1258–1259, 1259t
  - for gram-negative bacterial infections, 1268t
  - for mastoiditis, 255
  - for nongonococcal bacterial arthritis, 1057
  - for otitis media, 254
  - for plague, S4
  - in pregnancy and lactation, 1169t
  - resistance to, 1041
    - in *Acinetobacter* spp., 1296
    - in *Enterobacter* spp., 1292
    - in *Enterobacterales,* 1283t, 1284
    - mechanisms of, 1182–1183, 1182t
    - in *N. gonorrhoeae,* 1255
    - in *Salmonella,* 1319
    - in *S. pneumoniae,* 1122
  - for staphylococcal infections, 1204
- Cerebellar ataxia. *See* Ataxia
- Cerebellar ataxia, neuropathy, and vestibular areflexia syndrome (CANVAS), 3537, 3603
- Cerebellar degeneration/atrophy
  - in alcohol use disorder, 3676
  - disease-modifying protein identification in, 3957
  - falls in, 161
  - gait disorders in, 177, 178–179
  - paraneoplastic, 619, 749, 2791t
- Cerebellar hemorrhage, 192, 3454. *See also* Intracerebral hemorrhage
- Cerebral amyloid angiopathy, 3453t, 3454, 3478, 3480, 3489–3490
- Cerebral angiography. *See* Angiography, cerebral
- Cerebral arterial gas embolism, 3766
- Cerebral arteriopathy, 3391f, A16
- Cerebral artery(ies), 3430f. *See also* Middle cerebral artery
- Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), 3448, 3485, A16
- Cerebral blood flow, 2343, 2344f
- Cerebral contusion, 3573, 3573f
- Cerebral edema
  - after subarachnoid hemorrhage, 3458
  - high-altitude, 277, 3755–3756, 3755f, 3756t
  - in hyponatremia, 349
- Cerebral hemispheres, 200f, 3430–3431f
- Cerebral herniation, 187–188, 187–188f
- Cerebral ischemia, 3428
- Cerebral malaria, 994
- Cerebral microbleeds, 3490, 3490f
- Cerebral palsy, 50t, 178
- Cerebral perfusion pressure, 2343
- Cerebral salt wasting, 348
- Cerebral small-vessel disease, 3489–3490, 3490f
- Cerebral venous sinuses, 1141–1142, 1141f
- Cerebritis, 1135
- Cerebrospinal fluid (CSF), 3398
- Cerebrospinal fluid (CSF) analysis
  - in ALL, 844
  - in bacterial meningitis, 960, 962t, 1120, 1121t, S2
  - in chronic meningitis, 1131t, 1133
  - in coma, 191
  - in delirium, 185
  - in dementia, 198
  - in encephalitis, 960, 962t
  - in fever of unknown origin, 153
  - in fungal meningitis, S2
  - in Guillain-Barré syndrome, 3619
  - in HIV infection, 1606
  - inflammatory profile in, S2

  in meningococcal disease, 1249
  in MS, 3583
  in neurocysticercosis, S2
  in neuromyelitis optica, S2
  in neurosarcoidosis, S2
  in neurosyphilis, S2
  normal values, 962t, S3
  in parasite infections, S12
  in PML, 1117
  in RMSF, 992
  specimen collection and transport, S11
  in subacute meningitis, 1127
  in undiagnosed disease, 3996t
  in viral encephalitis, 1112
  in viral meningitis, 1124
Cerebrospinal fluid (CSF) leakage, 116, 116f, 3484, 3572, A16
Cerebrovascular diseases, **3428**. *See also* Stroke
  approach to the patient, 3428–3429, 3429f
  deaths from, 3846t
  definition of, 3428
  dementia associated with. *See* Vascular dementia
  global considerations in, 3428, 3846t
  incidence of, 3428
  seizures due to, 3414
  types of, 3428
  vascular anomalies, 3453t, 3456–3457
Ceritinib, 533t, 568t, 570, 628, 753t
Cerivastatin, 1742t
Cerliponase alfa, 3357
Certolizumab
  actions of, 2791t
  adverse effects of, 387t, 462, 2566, 2568t, 2850t
  for axial spondyloarthritis, 2887t
  for IBD, 2566, 2568t
  indications for, 2791t
  in pregnancy, 2570
  for psoriasis/psoriatic arthritis, 387t, 2890
  for rheumatoid arthritis, 2850t, 2851
  testing before and during treatment, 2568t
Ceruloplasmin, 324, 3339
Cerumen impaction, 248
Cervical adenopathy, 611, 611f
Cervical cancer, **713**
  clinical features of, 217
  deaths from, 500f, 501t
  diagnosis of, 217
  epidemiology of, 217, 1526
  etiology of, 713
  genetic considerations in, 713, 3792f
  global considerations in, 713, 1526
  HIV infection and, 1614
  HPV in. *See* Human papillomavirus (HPV) infection, cervical cancer and
  incidence of, 217, 498, 500f, 501t
  KIRs with, 2772t
  lifetime risk of, 37t
  paraneoplastic syndromes in, 737t
  prevention of, 514, 713
  prognosis of, 711t
  risk factors for, 713
  screening for, 37, 39t, 516t, 517, 713, 1529–1530, 3881t
  staging of, 711t, 713, 714f
  treatment of, 713
Cervical dystonia, 3510
Cervical radiculopathy, 129–130, 129t
Cervical spine
  atlantoaxial involvement in rheumatoid arthritis, 2840
  degenerative disk disease, 130, 131f
  degenerative myelopathy, 130, 131f
  lesions of, 3560
  trauma to, 132
  tumors, 130–131, 131f
Cervical venous hum, 1866
Cervicitis, mucopurulent, 1097t, 1103–1104, 1104f, 1256, 1471, 1473t
Cervix, ectopy of, 1104
CESD (cholesteryl ester storage disease, lysosomal acid lipase deficiency), 3354t, **3357**
Cesium-137 exposure, S6
Cestodes/cestode infections, 1736, 1810, **1832**
  coenurosis, 1839
  Diphyllobothrium latum (fish tapeworm), 787, S12
  Dipylidium caninum/dipylidiasis, 1839
  Echinococcus. See Echinococcosis (hydatid disease)
  fish tapeworms, 1736, 1739t, 1838–1839
  Hymenolepiasis diminuta, 1838
  by organ system and signs/symptoms, 1738–1740t
  Rodentolepis nana (dwarf tapeworm), 961t, 1736, 1838, S12
  sparganosis, 1839
  Taenia spp. *See under Taenia*
Cetirizine, 2808
CETP (cholesteryl ester transfer protein), 3239, 3239f
Cetuximab
  actions and targets of, 534t, 558, 559f
  adverse effects of
    anaphylaxis, 2812
    cutaneous, 558, 614
    desensitization protocol for, 2819
    hypersensitivity reactions, 597, 658
    pulmonary, 755
  for colorectal cancer, 558, 658
  genetic variations in response to, E72t
  for head and neck cancer, 613
  for Ménétrier's disease, 2540
Cevimeline, 266, 2878
CF. *See* Cystic fibrosis (CF)
cfDNA. *See* Cell-free DNA (cfDNA)
C fibers, 93, 94
CFI gene mutations, S2
CFTR gene mutations
  in cystic fibrosis, 967, 2245–2246, 2246f, 3781f, 3787, 3993, E72t
  genetic testing for, S10
  malaria resistance and, 3795
  in pancreatitis, 2744, 2751
  types of, 3787
CFTR protein
  in cystic fibrosis, 2245, 2246f, 2247, 2249v
  P. aeruginosa and, 1308
CGDs. *See* Chronic granulomatous diseases (CGDs)
C gene, of HBV, 2644, 2644f
cGMP (cyclic guanosine monophosphate), 1981, 2194, 3155, 3156f
CGRP (calcitonin gene-related peptide), 94, 3274
CGRP receptor antagonists (gepants), 3467t, 3468–3469, 3470t
CHA$_2$DS$_2$-VASc score, 1950, 1950t, 3444, 3445t
Chagas disease (American trypanosomiasis), **1788**
  approach to the patient, 1790
  cardiac manifestations of, 1739t, 1791, 1791f, 1927, 1965, 2013
  clinical features of, 1789–1790, 1790t, 2013
  diagnosis of, 1790–1791, 1790t, S12
  differential diagnosis of, 2507
  epidemiology of, 1789, 2013, S12
  global considerations in, 1789, 1792
  in HIV infection, 1609
  inflammasome mutations in, 2761t
  pathology of, 1789, 1789f
  prevention of, 1792
  prognosis of, 1790t
  in solid organ transplant recipient, 1159t
  staging of, 1790t, 1791
  transmission of, 913t, 1788
  treatment of, 1747, 1750, 1791–1792, 1791t, 2013
Chagoma, 1790t
Chagres virus, 1662t
Chalazion, 224
Chancre
  syphilitic, 261–262, 262t, 1430, 1430f, A1, A5
  trypanosomal, 1794, A1
Chancroid
  clinical features of, 1052, 1107, 1107f, 1108t, 1264, 1264f, A5
  diagnosis of, 1108, 1264
  epidemiology of, 1107, 1264
  etiology of. *See Haemophilus ducreyi*
  global considerations in, 1107
  in HIV infection, 1602
  HIV transmission and, 1263
  lymphadenopathy in, 477
  treatment of, 1108–1109, 1109t, 1264
Chandipura virus, 1663t, 1673
Channelopathy(ies)
  in cardiomyopathy, 2004t, 2009
  genetic considerations in, 795
  muscular, 3646–3647
  renal disease in, 2359t
Chapare virus, 1657t
Charcot foot (neuropathic joint disease), 1058, 1068, 1068f
Charcot joint (neuropathic joint disease), 2968–2969, 2969f, 2969t
Charcot-Leyden crystals, 309, 462, 1715
Charcot-Marie-Tooth (CMT) disease
  classification of, 3598, 3599–3600t, 3601
  genetic considerations in, 3599–3600t, 3784
  mitochondrial dysfunction in, 3400
  phenotypic heterogeneity in, 3787, 3788t
Charcot's triad, 2489, 2735
CHARGE (colomba of the eye, heart anomaly, choanal atresia, retardation, genital and ear anomalies) syndrome, 2799, S9
Charter on Medical Professionalism, 6
Chart review, 28
CHCHD2 gene mutations, 3498t
CHD. *See* Congenital heart disease (CHD)
CHD7 deficiency, 2799
CHD7 gene mutations, 3111t, 3112, S9
Checkpoint inhibitors. *See* Immune checkpoint blockade/inhibitors
Chédiak-Higashi syndrome
  cellular/molecular defects in, 459t, 460, 2768
  clinical features of, 261, 459t, 460, 466, 2803, S9
  diagnosis of, 442f, 443, 459t, 460, 460f, 466
  lymphoid malignancies in, 858, 858t
  treatment of, 466, 2609
Cheek, traumatic lesion inside of, A3
Cheese washer's lung, 2228t
Cheilitis, 264t, 607, A3
Cheilosis, 265, 765
CHEK2 gene mutations, 3800
Chelation therapy, 777, 3699, 3700t, 3702, S6
Chemical burn, oral, 263t
Chemical meningitis, 1131
Chemical restraints, 186
Chemical terrorism/chemical weapons, 3712t, **S5**
Chemical Warfare Convention, S5
Chemical worker's lung, 2228t, 2229
"Chemo brain," 756
Chemo cap, 576
Chemokine(s)
  in asthma, 2219
  definition of, 2755
  functions of, 455–456
  in HIV infection. *See* CCR5 receptor, in HIV infection
  in pneumonia, 1023

- Chemokine(s) (*Cont.*):
  - receptors, 2769t
  - in sepsis/septic shock, 2317, 2318f
- Chemoreceptors, 267
- Chemoreceptor trigger zone (CTZ), 82, 297, 575
- Chenodeoxycholic acid (CDCA), 2727
- Cherry angioma, 378t
- Cherry hemangioma, 405, A5
- Cherry red spot, A15
- Chest compressions, 2336
- Chest discomfort/pain, **102**
  - approach to the patient, **104,** 105t
    - cardiac biomarkers, 107, 107f
    - chest radiography, 107
    - clinical algorithms, 2108f
    - CMR, 109, 1889, 1890f, 1902v
    - CT angiography, 108, 1889, 1889f, 1890f
    - decision aids for, 108, 108f
    - ECG, 107
    - echocardiography, 109, A9
    - in the emergency department, 1889–1890
    - history, 105–106, 105f
    - in patients with known CAD, 1889, 1889f
    - physical examination, 106–107
    - stress testing, 108–109, 1887–1888, 1889f
  - of cardiopulmonary origin, 103t
    - aortic disease, 104
    - exertional, 1842
    - lung cancer, 618–619, 618t
    - myocardial ischemia. *See* Angina; Myocardial infarction (MI)
    - nonischemic myocardial disease, 104
    - pericardial disease, 104, 2078–2079, 2092, 2115, 2125
    - pneumothorax, 104
    - pulmonary disease, 104
    - pulmonary embolism, 104
    - in respiratory disease, 2199–2200
  - with dysphagia, 295
  - epidemiology of, 102, 102f
  - of gastrointestinal origin, 103t
    - esophageal disorders, 104, 106, 2504, 2507
    - GERD, 104, 106, 301, 2509
  - of musculoskeletal origin, 104
  - myocardial infarction probability with, 105f
  - natural history of, 102
  - of neuromuscular origin, 103t
  - outpatient evaluation of, 109
  - in pericarditis, 2078–2079
  - of psychological origin, 103t, 104, 3658
  - in sickle cell anemia, 775
  - in TB, 1387
- Chest pain units, 2109
- Chest physiotherapy, 2244, 2248
- Chest radiography, **A12**
  - in anthrax, S4
  - in ARDS, 2300f
  - in cardiogenic shock, 2326
  - in chest pain evaluation, 107
  - in cough evaluation, 273–274
  - in COVID-19 disease, 1541, A12
  - in drowning, A12
  - in dyspnea, 270–271
  - in heart failure, 1985
  - in hemoptysis, 276
  - in hypersensitivity pneumonitis, 2229
  - in interstitial lung disease, 2262
  - in *Legionella* pneumonia, 1273, 1273–1274f
  - in lung cancer screening, 518, 617–618, 618t
  - normal, A12
  - in PCP, 1731, 1733f
  - in pleural effusion, A12
  - in pneumococcal pneumonia, 1192, 1192f
  - in pneumonic plague, 1344f
  - in pneumoperitoneum, 1070f
  - in pulmonary hypertension, 2188, 2190f
  - in respiratory disease evaluation, 2201, 2210–2211, 2211f
  - in sarcoidosis, 2925, 2925f
  - in thoracic lymphadenopathy, 478
  - in tularemia, 1339, 1339f
- Chest tightness, 267, 268f
- Chest trauma, 2088–2089, 2166
- Chest wall
  - abnormality of, ventilatory restriction due to, 2205, 2273–2274
  - compliance of, 2202, 2296
  - disorders of, 269, 269t
  - pressure-volume curves of, 2202f
  - tenderness in, 107
  - trauma to, 284
- Chewing tobacco, 511, 3684
- Cheyne-Stokes respiration
  - in central sleep apnea, 2280f, 2282
  - in coma, 191
  - in heart failure, 1983
  - in terminally ill patient, 91, 91t
- CHF (congestive heart failure). *See* Heart failure (HF)
- Chiari malformation, 132, 237, 3567, 3568f, V3
- Chiasmal tumors, 232
- Chiasmata, 3780
- Chickenpox. *See* Varicella
- Chief cells, 2518
- Chiggers, 3730
- Chigoe flea, 3736
- Chikungunya virus infection
  - climate change and, 3748
  - clinical features of, 140t, 146, 402, 1059, 1669, A1
  - epidemiology of, 140t, 146, 1112, 1669, 3748–3749
  - reservoirs and vectors of, 1657t, 1669
  - treatment of, 1669
  - vaccine, 1014t, 1017
- Chilblains (pernio), 146, 2177f, **2178,** 3771–3772
- Child abuse, 71, 1144, 1257, 3674
- Child-Pugh score, in liver disease, 663, 2632–2633, 2633t
- Children
  - atopic dermatitis in, 383
  - drug use in, 488
  - ECG in, A7
  - fever treatment in, 137
  - GH deficiency in, 2994
  - Graves' disease in, 3037
  - HIV infection in, 1566
  - hypothyroidism in, 3030
  - infectious diarrhea in, 1079
  - iron requirement in, 765t
  - liver transplantation in, 2718–2719, 2718t
  - malaria in, 1765–1766, 1765t
  - mortality rates in, 3844, 3861f, 3865f
  - nutrient requirements in, 2600–2602t
  - rheumatic fever in, 2854
  - seizures in, 3411, 3413, 3413t
  - smoking in, 511
  - vaccinations in, 1002
- Child-Turcotte-Pugh (CTP) score, 2720
- Chile, 3864
- Chilean Rose tarantula, 3734
- Chimeric antigen receptor (CAR) T cells
  - action and targets of, 535t, 549, 559f, 561, 2783, 2788, 2788f
  - adverse effects of, 3825t
    - cytokine release syndrome, 561, 2351, 2788
    - hemophagocytic lymphohistiocytosis, 469
    - immune system dysfunction, 755
    - infections, 753t
    - neurologic, 757, 2351
    - seizures, 591
    - T-cell malignancies, 549, 3829
  - for ALL, 848t, 849
  - for CLL, 856
  - definition of, 2783
  - for DLBCL, 863, 3829
  - for follicular lymphoma, 865
  - for Hodgkin's lymphoma, 870
  - infection prophylaxis with, 579
  - for mantle cell lymphoma, 866
  - for multiple myeloma, 892, 3824t, 3829
  - for neuromyelitis optica, 3593
  - resistance to, 549
- Chimpanzee, as HIV reservoir, 1559
- China, 3865–3866, 3865f
- Chinese herbal nephropathy, 2377, 2442
- Chios virus, 1662t
- ChIP (chromatin immunoprecipitation), 3937, 3938f
- CHIP (clonal hematopoiesis of indeterminate potential), 769, 826, 1858–1859, 3982, 3982f, A10
- CHIPS, 1198
- Chiropractic, 3925t, 3928
- *Chlamydia* spp.
  - apoptosis inhibition by, 973
  - autophagy inhibition by, 973
  - genomes of, 1469
  - growth cycle of, 1468, 1468f
  - microbiology of, 1467–1468, 1468f
  - survival in cytosol, 969
  - survival in vacuole, 967
- *Chlamydia pneumoniae* infections, **1476**
  - clinical features of, 406, 1476
  - diagnosis of, 1476–1477
  - epidemiology of, 1476
  - pathogenesis of, 143t, 259, 1476, 3836
  - treatment of, 1477
- *Chlamydia psittaci* infections (psittacosis), 1475–1476
- *Chlamydia trachomatis,* 971, 976, 1468
- *Chlamydia trachomatis* infections, 961t, **1469**
  - anorectal, 1109
  - cutaneous, 401, 406
  - genitourinary, **1469**
    - diagnosis of, 976, 1472–1473, 1473t, S11
    - epidemiology of, 1097, 1469–1470
    - epididymitis, 1099, 1470
    - global considerations in, 1469
    - with gonococcal infection, 1259, 1259t
    - incidence of, 1097
    - lymphogranuloma venereum. *See* Lymphogranuloma venereum (LGV)
    - mucopurulent cervicitis, 1471
    - partner-delivered therapy for, 1474
    - pathogenesis of, 1468–1469
    - proctitis, 1471
    - screening for, 39t, 1110, 1474
    - spectrum of, 1469
    - spondyloarthritis and, 2882
    - treatment of, 1099, 1099t, 1474
    - urethral syndrome in women, 1471–1472
    - urethritis, 1098–1099, 1470
  - inflammasome mutations in, 2761t
  - lung, 961t
  - in neonates, 1472, 1473t, 1474
  - pelvic inflammatory disease. *See* Pelvic inflammatory disease (PID)
  - perihepatitis, 1471
  - in pregnancy, 1472
  - prevention of, 1474
  - reactive arthritis, 1470–1471
  - trachoma, 224, 1474–1475
- *Chlamydophila psittaci,* 857, 865
- Chloral hydrate, 3713t
- Chlorambucil, 562t, 591, 855
- Chloramphenicol
  - actions of, 1166, 1182t

- adverse effects of, 1171t, 1179
  - aplastic anemia, 810t
  - bone marrow failure, 826
  - neutropenia, 457
  - optic neuropathy, 228
  - pure red cell aplasia, 814t
- for anaerobic bacterial infections, 1379, 1379t
- for Carrión's disease, 1353t
- for donovanosis, 1358
- drug interactions of, 1179
- for enteric fever, 1318, 1318t
- for Hib meningitis, 1263
- indications for, 1174t, 1179
- for meningococcal infections, 1249
- for plague, 1346, 1346t, S4
- in pregnancy and lactation, 1169t
- resistance to, 1182t, 1183f, 1184
- RMSF, 1457
- for scrub typhus, 1460
- for tick-borne spotted fevers, 1458
- for tularemia, 1340, S4

Chlordiazepoxide
- for alcohol withdrawal, 3680
- for anxiety disorders, 3660t
- overdosage/poisoning with, 3713t
- pharmacology of, 3660t

Chlorhexidine bath, 1148
Chlorhexidine mouthwash, 243
Chloride
- recommended intake of, 2601t
- renal transport of, 2366f, 2369

Chloride channel activators. *See* Lubiprostone
Chloride channel disorders, 3647
Chloridorrhea, congenital, 308
Chlorine, as chemical terrorism agent, S5
Chloroma, cutaneous (granulocytic sarcoma), 405
Chloroquine
- actions of, 1747
- adverse effects of
  - bone marrow failure, 810t
  - cardiotoxicity, 2017
  - common, 1771t
  - cutaneous, 399
  - hair discoloration, 421
  - myopathy, 3648t
  - neuropathy, 3608, 3609t
  - occasional, 1741t, 1775
  - ocular, 232, 1775, 2833
  - rare, 1741t, 1747, 1771t
- drug interactions of, 1741t
- genetic variations in response to, E72t
- for hypercalcemia, 365
- for malaria, 1741t, 1768, 1770t, 1771t
- for malaria prophylaxis, 1774t, 1775
- overdosage/poisoning with, 3715t
- pharmacology of, 1747, 1771t
- for porphyria cutanea tarda, 3348
- resistance to, 1768, 1775
- for retinitis pigmentosa, 232
- for SLE, 2833
- for Whipple's disease, 1370

Chlorproguanil, 1743t
Chlorpromazine
- adverse effects of, 3673, 3673t
  - cutaneous, 420
  - hepatic, 324
  - lupus syndrome, 2940t
  - photoallergy, 434t
- for delirium, 86, 86t
- for dyspnea, 83
- for medication-overuse headache, 116
- for migraine, 3467t, 3469
- for nausea and vomiting, 83
- overdosage/poisoning with, 3711t
- for schizophrenia, 3673, 3673t

Chlorpropamide, 923t
Chlorthalidone, 2143, 2144t
Choclo virus, 1661t, 1674
Choice architecture, 3922
Choking/pulmonary-damaging agents, S5
Cholangiocarcinoma, **667**
- classification of, 667, 667f
- diagnosis of, 667
- epidemiology of, 667
- extrahepatic, 326, 669
- intrahepatic, 667
- molecular drivers of, 667
- pathogenesis of, 667, 668f
- risk factors for, 667
- in sclerosing cholangitis, 2737
- staging of, 667, 669f, 2474f
- treatment of, 667–669, 669f, 2474f, 2724

Cholangiopancreatography
- endoscopic. *See* Endoscopic retrograde cholangiopancreatography (ERCP)
- magnetic resonance. *See* Magnetic resonance cholangiopancreatography (MRCP)

Cholangitis
- abdominal pain in, 112t
- ascending, 2489
- in choledocholithiasis, 2472f, 2735
- IgG4-associated, 326
- immunoglobulin G4–associated, 2737
- in parasitic infections, 1739t
- primary biliary, 324, 400, 2695, **2711,** 2790t, A13
- primary sclerosing. *See* Primary sclerosing cholangitis (PSC)
- in solid organ transplant recipient, 1159t
- treatment of, 2321t

Cholecystectomy
- complications of, 2731, 2734
- for gallbladder cancer, 670
- for hyperplastic cholecystoses, 2734
- indications for, 2731
- with splenectomy, in hereditary spherocytosis, 795

Cholecystitis, **2732**
- acalculous, 2732
- in cancer patient, 583
- chronic, 2733
- clinical features of, 111, 112t, 297, 2732–2733, 2735
- complications of, 2733
- diagnosis of, 112–113, 2732
- differential diagnosis of, 2660
- emphysematous, 2732–2733
- pancreatitis and, 2735
- treatment of, 2733–2734

Cholecystokinin (CCK), 2516f, 2742, 3182, 3182f
Cholecystokinin (CCK)–releasing factor, 2743
Cholecystopathy, acalculous, 2732
Cholecystosis(es), hyperplastic, 2734
Choledochal cysts, 2735
Choledocholithiasis, 326, 2731f, **2735**
Cholelithiasis. *See* Gallstone(s)
Cholera, **1325**
- clinical features of, 306t, 1079t, 1327–1328, 1327f, 1328t
- diagnosis of, 1328
- differential diagnosis of, 306t
- epidemiology of
  - climate change and, 3750, 3753f
  - environmental factors, 1326–1327
  - genomics in, 985, 986f
  - historical studies, 984
- global considerations in, 1326–1327, 1326f
- microbiome and, 3839
- pancreatic, 308
- pandemics, 3750, 3750t, 3856
- pathogenesis of, 1327. *See also* *Vibrio cholerae*
- prevention of, 1329. *See also* Cholera vaccine
- treatment of, 1328–1329, 1328t, 1329t

Cholera gravis, 1326, 1327
Cholera vaccine
- for international travelers, 1014t, 1015
- killed whole-cell, 1329
- live attenuated viral, 1329

Cholestasis
- benign postoperative, 325
- benign recurrent intrahepatic, 325, 2641t, 2642
- drug-induced, 325, 2667, 2667t, 2669t, 2671
- dyslipidemia and, 3246–3247
- extrahepatic, 324–325, 325t, 326
- imaging in, 324–325
- intrahepatic, 325, 325t
- jaundice in, 324
- liver function tests in, 2635, 2637t
- during parenteral nutrition, 2625
- of pregnancy, 325, 2729, 3903
- progressive familial intrahepatic, 325, 2641t, 2642

Cholestatic diseases, 2627, 2628t. *See also* Liver disease/failure
Cholestatic hepatitis, 2659
Cholesteatoma, 245, 255
Cholesterol absorption inhibitor. *See* Ezetimibe
Cholesterol emboli, 407, 2415t, 2427, A4
Cholesterolosis, 2734
Cholesterol side-chain cleavage enzyme (CYPIIa), 2790t
Cholesterol stones, gallbladder, 2728–2729, 2728f, 2730t. *See also* Gallstone(s)
Cholesteryl ester storage disease (lysosomal acid lipase deficiency), 3241t, 3244–3245, 3354t, **3357**
Cholesteryl ester transfer protein (CETP), 3239, 3239f
Cholestyramine
- adverse effects of, 309, 3244t
- drug interactions of, 489t
- for IBS, 2575, 2577t
- for lipoprotein disorders, 3244t, 3251–3252
- for metabolic syndrome, 3257–3258
- for pruritus, 2737
- vitamin A absorption and, 2611

Cholic acid, 2727
Choline, 2600t, **2610**
Cholinergic urticaria, 402, 2807
Cholinesterase inhibitors
- for AD, 3480
- adverse effects of, 3480, 3629, 3889
- for dementia, 199, 3888–3889
- for DLB, 3493
- for myasthenia gravis, 3628–3629

Chondrocalcinosis, 356, 2961
Chondrocyte(s), 2950, 2950f, 2956
Chondrodysplasia(s), **3330**
Chondroitin, 2956
Chondromatosis, synovial, 2972
Chondrosarcoma, 524t, **730**
Choose My Plate, 2604, 2604t
CHOP regimen, 863, 864, 866, 867, 872, 873
Chordoma, 729, 3003
Chorea(s), **3512**
- childhood-onset, 3514
- clinical features of, 3509t
- disorders associated with, 3514
- drug-related, 3514
- Huntington's disease. *See* Huntington's disease
- Sydenham's, 2856, 3514
- treatment of, 2858

Chorea-acanthocytosis, 3514
Choriocarcinoma. *See also* Gestational trophoblastic disease
- diagnosis of, 705
- ovarian, 712
- paraneoplastic syndromes in, 737t
- parasellar, 3003

- Choriomeningitis, lymphocytic, 1059, 1131t
- Chorionic villus sampling (CVS), 3797
- Chorioretinopathy, central serous, 231
- Choroid
  - gyrate atrophy of, 3369t
  - melanoma of, 232, 232f, 599f, 601t, 605
- Choroidal artery infarct, A16
- Chrm1, 3397
- *CHRNA4* gene mutations, 3412t
- *CHRNA5* gene mutations, 3652
- Chromatin
  - in cancer cells, 538, 538f
  - in DNA, 3782, 3784f
  - immunoprecipitation, 3937, 3938f
  - modification, 526t, 3935–3937
  - in tumors of unknown cell origin, 3938
- Chromatin-modifying agents, 570t, 573
- Chromatography, 3976–3977, 3976f
- Chromium, 512t, 2601t, 2614t, **2615,** 3702
- *Chromobacterium violaceum,* 1268t, 1269
- Chromoblastomycosis, 1689t, 1725
- Chromogranin
  - in carcinoma of unknown primary, 733t, 735
  - in lung cancer, 616
  - in NETs, 677
- Chromophobe tumors, renal, 687, 688t
- Chromophores, 429, 430
- Chromosomal microarray, 3988
- Chromosomal sex, 3094, 3094f
- Chromosome(s)
  - analysis of, 3789t
  - crossing-over and recombination of, 3780, 3782f
  - imprinting in, 3782–3783, 3785f
  - instability of, 521, 653. *See also* Cancer genetics
  - replication of, 3778
  - structure of, 3778, 3779f
- Chromosome abnormalities/disorders. *See also* Genetic disorders; *specific diseases*
  - in cancer, 520t
  - contiguous gene syndromes, 3789
  - copy number variations, 3779, 3782f
  - diagnosis of, 3789t
  - incidence of, 3789
  - mosaicism, 3791
  - noninvasive prenatal testing for, 3987–3989
  - in radiation-exposed patient, S6
  - in sex chromosomes, 3095t. *See also* Disorders of sex development (DSDs)
  - single-nucleotide polymorphisms. *See* Single-nucleotide polymorphisms (SNPs)
  - translocation, 521, 521t, 522f
  - unequal crossing-over, 3784–3785
- Chromothripsis, 521
- Chronic atypical neutrophilic dermatosis with lipodystrophy and elevated temperature (CANDLE) syndrome, 2937
- Chronic beryllium disease/toxicity, 406
- Chronic bronchitis, 272, 274, 2239t, 2249
- Chronic Care Model, 44, 44f
- Chronic cutaneous lupus erythematosus. *See* Discoid (chronic cutaneous) lupus erythematosus
- Chronic disease, aging and, 3869, 3870f
- Chronic disease risk reduction intake, 2599
- Chronic eosinophilic leukemia, 819, 876, 878t
- Chronic fatigue syndrome. *See* Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS)
- Chronic granulomatous diseases (CGDs)
  - cellular/molecular defects in, 459t, 460, 1689, 2796
  - clinical features of, 459t, 460, 2796, 3115
  - complications of, 460
  - diagnosis of, 459t
  - differential diagnosis of, 2924t
  - treatment of, 463, 2784, 2796
  - X-linked, 2796, S9
- Chronic inflammatory demyelinating polyneuropathy (CIDP), 3619t, **3622**
- Chronic kidney disease (CKD), **2385**
  - in agricultural workers, 2397, 3740, 3742f
  - in amyloidosis, 896. *See also* Amyloidosis
  - approach to the patient
    - establishment of diagnosis and etiology, 2395
    - genetic testing, 2395
    - history and physical examination, 2394
    - imaging, 2395
    - laboratory studies, 2387, 2388t, 2394–2395
    - renal biopsy, 2395
    - staging, 2387–2388
  - breath odor in, 267
  - calcium and phosphate metabolism disorders in
    - bone manifestations of, 2390
    - calciphylaxis, 2391, 2391f
    - cardiovascular manifestations of, 2390–2391
    - hyperphosphatemia, 3265–3266
    - secondary hyperparathyroidism, 2390, 3282
    - treatment of, 2391, 3282, 3290
    - tumoral calcinosis, 2390, 2390f
    - vitamin D deficiency and, 3290
  - cardiovascular disorders in
    - epidemiology of, 2391
    - heart failure, 2391–2392
    - hyperphosphatemia and, 2390–2391
    - hypertension, 226f, 2138t, 2139, 2146–2147, 2392
    - left ventricular hypertrophy, 2392
    - pericardial disease, 2083, 2392
    - risk factors for, 2400
    - treatment of, 2392
    - vascular disease, 2391
  - classification of, 2385, 2386f
  - clinical features of, 338t, 2362
  - contrast agent use in, 1904, 2394
  - deaths from, 75t, 2397, 3846t
  - diagnosis of, 338t, 2360
  - drug concentration and response and, 488
  - dyslipidemia in, 3246
  - endocrine-metabolic disturbances in, 2394, 3115
  - epidemiology of, 2388, 2397
  - etiology of, 2359t, 2360–2361, 2388
    - diabetes mellitus. *See* Diabetic nephropathy
    - glomerulonephritis. *See* Glomerulonephritis
    - urinary tract obstruction, 2455
  - in Fabry's disease, 3355
  - fluid, electrolyte, and acid-base disorders in
    - edema, 281, 281t
    - fluid, 2389
    - hyperkalemia, 359, 2389
    - hypokalemia, 356, 2389
    - hyponatremia, 347–348, 2389
    - metabolic acidosis, 369, 371–372, 2389–2390
    - treatment of, 372, 2390
  - gastrointestinal and nutritional abnormalities in, 2394
  - genetic considerations in, 2357, 2359t, 2368t, 2386–2387, 2388t, 2395
  - global considerations in, 2397, 2401–2402
  - glomerular filtration rate in, 337–339, 2388t
  - HCV treatment in, 2693
  - hematologic abnormalities in
    - abnormal hemostasis, 2393
    - anemia, 768, 2362
    - bleeding, 473
    - burr cells, 446f
    - etiology of, 2392–2393
  - high-altitude travel and, 3760
  - immunization in, 1004t
  - in multiple myeloma, 886–887
  - neurologic disorders in, 2393–2394
    - encephalopathy, 188
    - fatigue, 166
    - gait disorders, 179
    - neuropathy, 2393–2394, 3605
  - pathophysiology of, 2385, 2387f, 2388–2389
  - in pregnancy, 2394, 3899
  - renal response to, 2387f
  - risk factors for, 2385–2387, 2388t
  - in Sjögren's disease, S1
  - skin manifestations of, 2394
  - in SLE. *See* Lupus nephritis
  - specialized nutritional support in, 2626
  - stage 5 (end-stage renal disease)
    - cardiovascular disease and, 2401
    - global considerations in, 2401–2402
    - in HIV infection, 1601
    - hypertension in, 2427
    - incidence of, 2398
    - outcomes of, 2401
    - treatment of, 2398. *See also* Hemodialysis; Kidney transplantation; Peritoneal dialysis
    - uremic syndrome in, 2388–2389
  - stages of, 2361–2362, 2362t
  - staging of, 2386f, 2387–2388
  - TB treatment in, 1400
  - treatment of
    - diabetic nephropathy. *See* Diabetic nephropathy
    - erythrocyte-stimulating agents, 769
    - intraglomerular hypertension reduction, 2396, 2396f
    - medication dose adjustment, 2396–2397
    - patient education, 2397
    - racial/ethnic disparities in, 63, 63f
    - renal replacement therapy, 2397. *See also* Hemodialysis; Kidney transplantation; Peritoneal dialysis
  - tuberculous, 1388, 1388f
- Chronic lung allograft dysfunction (CLAD), 1163, 2286, 2287
- Chronic lymphocytic leukemia (CLL), **849**
  - B-cell receptor signaling in, 850
  - cell of origin in, 850, 850f
  - clinical features of, 852
  - complications of, 852–854
    - autoimmune, 802, 853
    - infections, 852–853
    - pure red cell aplasia, 813
    - Richter's transformation, 853–854
    - second malignancies, 853
  - cytogenetic abnormalities in, 850, 851f
  - diagnosis of, 852, 852t
  - epidemiology of, 849–850
  - genetic considerations in, 851–852, 851t, 858
  - immunology of, 851–852, 872t
  - lymphadenopathy in, 477
  - mu heavy chain disease in, 894
  - peripheral blood smear in, 443, 850, 850f, A6
  - prognosis of, 854, 854t
  - renal involvement in, 2440
  - skin manifestations of, 411
  - splenomegaly in, 480
  - staging of, 854, 854t
  - treatment of, 854–856, 854t, 856t, 920
  - tumor lysis syndrome in, 593
- Chronic mesenteric ischemia. *See* Intestinal ischemia
- Chronic mountain sickness, 278, 3760
- Chronic mucocutaneous candidiasis, 1708, 2801, 3089, 3090t
- Chronic myeloid leukemia (CML), **834**
  - accelerated blastic phase of, 836, 838

  - atypical (*BCR-ABL1* negative), 877t, 878
  - blastic phase of, 838
  - cancer stem cells in, 542
  - chromosome abnormalities in, 521, 521t, 522f, 819, 834, 835–836, 835f
  - clinical features of, 836–837, 837t
  - diagnosis of, 837–838, A6
  - epidemiology of, 834
  - etiology of, 834–835
  - global considerations in, 843, 843f
  - incidence of, 834
  - pathophysiology of, 835–836
  - peripheral blood smear in, 443, A6
  - prognosis of, 834, 838, 839f
  - progression of, 836, 838
  - skin manifestations of, 409
  - splenomegaly in, 837
  - treatment of, **838,** 839t
    - in accelerated or blastic phases, 842
    - chemotherapeutic agents, 842
    - in cytogenetic abnormalities of Ph-negative cells, 843
    - goals for, 840
    - HCT, 841–842, 919–920, 919t
    - interferon, 842
    - leukapheresis, 842
    - monitoring, 842
    - in pregnancy, 842–843
    - splenectomy, 480
    - TKI toxicity management, 840–841
    - tyrosine kinase inhibitors, 536, 567, 839t, 840–841
- Chronic myelomonocytic leukemia (CMML), 815, 877–878, 877t
- Chronic neutrophilic leukemia, 819, 875–876, 877t
- Chronic nonspherocytic hemolytic anemia (CNSHA), 797, 800
- Chronic obstructive pulmonary disease (COPD), **2249**
  - with bronchiectasis, A12
  - chronic hypercapnia in, 2276
  - cor pulmonale in, 2254. *See also* Cor pulmonale
  - deaths from, 2249
  - definition of, 2249
  - dyspnea in, 269t, 270, 2254
  - ECG in, A7, A8
  - exacerbations of, 1262f, 1265, 2258–2259, 2306
  - exercise intolerance in, 291
  - genetic considerations in, 2253
  - high-altitude travel and, 3760
  - history in, 2254
  - imaging in, 2250f, 2255, A12
  - incidence of, 2249
  - laboratory findings in, 2254–2255, 2255t
  - natural history of, 2253–2254, 2254f
  - overlap with asthma, 2228
  - pathogenesis of, 2249–2251, 2250f
  - pathophysiology of, 2251–2252
  - physical findings in, 2254
  - pulmonary hypertension in, 2192
  - risk factors for, 2252–2253
  - severity of, 2249, 2254, 2256f
  - systems biology applied to, 3961t
  - treatment of
    - lung transplantation, 2258
    - lung volume reduction surgery, 2258, 2289–2290
    - pharmacotherapy, 2255–2258, 2257f
    - pulmonary rehabilitation, 2258
- Chronic pancreatitis, **2750**
  - approach to the patient, 2738, 2740f
  - autoimmune, 2751–2752, 2751t, 2752t
  - clinical features of, 309, 326, 406, 2752
  - cobalamin malabsorption in, 787
  - complications of, 2752–2753, 2753t
  - diagnosis of, 2499, 2739t, 2740–2742, 2741t, 2752, 2753f
  - epidemiology of, 2738
  - etiology of, 2750t, 2751
  - genetic considerations in, 2751
  - pathophysiology of, 2750–2751
  - treatment of, 2753–2754
- Chronic pelvic pain syndrome, 1089
- Chronic progressive external ophthalmoplegia. *See* Progressive external ophthalmoplegia (PEO)
- Chronic prostatitis/chronic pelvic pain syndrome, 330
- Chronic pulmonary aspergillosis (CPA), 1390
- Chronic spontaneous urticaria, 422
- Chronic thromboembolic pulmonary hypertension (CTEPH)
  - diagnosis of, 2190, 2193f
  - etiology of, 2192
  - treatment of, 2164, 2193f, 2196, 2196f
- Chronic traumatic encephalopathy (CTE), 194, 3401, 3484, 3523, 3576–3577
- Chronic venous disease, **2179**
  - anatomic considerations, 2179
  - classification of, 2181, 2181t
  - clinical features of, 2180, 2180f
  - diagnosis of, 2181
  - differential diagnosis of, 2180–2181
  - epidemiology of, 2179
  - pathophysiology of, 2180
  - treatment of, 2181–2182
- Chronic wasting disease (CWD), 3526, 3526t
- Chronotherapy, 3954
- Chronotropic incompetence, 1919
- Chronotype, 3944t
- *Chryseobacterium indologenes* infections, 1268t
- *Chryseobacterium meningosepticum* infections. *See Elizabethkingia meningoseptica* infections
- Chrysiasis, 400
- *CHS1* gene, 460
- *CHST14* gene mutations, 3329t, 3330
- Churg-Strauss. *See* Eosinophilic granulomatosis with polyangiitis
- Chvostek's sign, 365
- Chylomicron(s), 2541, 2542f, 2545, 3238–3239, 3239f
- Chylomicronemia syndrome, familial, 3240, 3241t
- Chylothorax, 2270
- *CIAS1* gene mutations, 462
- Ciclopirox, 386
- Cidofovir
  - actions of, 1116, 1487
  - for adenovirus infection, 1157, 1537
  - adverse effects of, 1116, 1486t, 1487, 1537, 1601, 2377
  - for CMV infection, 1116, 1416, 1486t, 1487, 3566
  - for CMV prophylaxis, 1594t
  - for HSV infection, 1502
  - indications for, 1487
  - pharmacology of, 1487
  - for poxvirus infections, 1521
  - for smallpox, S4
- CIDP (chronic inflammatory demyelinating polyneuropathy), 3619t, **3622**
- CIEDs. *See* Cardiovascular implantable electronic devices (CIEDs)
- Cigarette smoking. *See* Smoking
- Cigar smoking, 511
- Ciguatera poisoning, 306, 3726–3727, 3727t
- Cilium, 2430, 2431f
- Cilostazol, 2112, 2174
- Ciltacabtagene autoleucel, 535t, 559f, 890t, 892, 3824t
- Cimetidine
  - actions of, 2369
  - adverse effects of, 324, 923t, 2524, 2940t
  - drug interactions of, 489t
    - chloroquine, 1741t
    - macrolides, 1172t
    - mebendazole, 1742t
    - metronidazole, 1744t
    - PDE-5 inhibitors, 3159
    - phenytoin, 2524
    - praziquantel, 1751
    - quinidine, 1745t
    - quinine, 1752
    - theophylline, 2524
    - warfarin, 2524
  - for gastrinoma, 3081
  - for GERD, 301
  - for IC/BPS, 335
  - for peptic ulcer disease, 2524, 2524t, 3585
  - as premedication for taxanes, 565
- Cimino fistula, 2402
- Cinacalcet, 738, 740, 2391
- Cinefluoroscopy, 1907–1908, 1908f
- Cingulate cortex, 204, 204f
- Cinnamates, 435
- Cinnarizine, 3497
- Ciprofloxacin
  - actions of, 1166, 1177–1178, 1182t
  - adverse effects of, 228, 799t, 1178, 1745t, 2667l, 2667t
  - for *Aeromonas* infections, 1268
  - for anthrax, S4
  - for bite-wound infections, 1054t, 1145, 1145t
  - for bronchiectasis prophylaxis, 2244
  - for *Brucella* infections, 1334l
  - for *Campylobacter* infections, 1306
  - for Carrión's disease, 1353t
  - for chancroid, 1109t, 1264
  - for cholera, 1329
  - for clostridial infections, 1239t
  - for community-acquired pneumonia, 963t
  - for cystitis, 1093
  - for diverticulitis, 2580
  - drug interactions of, 1745t
  - for enteric fever, 1318
  - for gonococcal arthritis, 1058
  - gut microbiome and, 3834
  - for hospital-acquired pneumonia, 963t, 1032t
  - indications for, 1174t, 1178, 1745t
  - for infectious diarrhea, 307, 1081t
  - for intraabdominal infections, 963t
  - for *Legionella* spp. infections, 1275t
  - for necrotizing fasciitis, 1054t
  - for neutropenic fever, 581
  - for nongonococcal bacterial arthritis, 577
  - for nontyphoidal *Salmonella* infections, 1321t
  - for osteomyelitis, 1064t, 1065t
  - for otitis externa, 254
  - for *P. aeruginosa* infections, 1309t
  - for peritonitis, 330
  - for peritonitis prophylaxis, 1071
  - for plague, 1346
  - for plague prophylaxis, 1346, 1347, 1347t
  - for plague treatment, 1346t
  - for pouchitis in IBD, 2565
  - prophylactic uses of, 1180t
  - for pyelonephritis, 1094
  - resistance to
    - in *Campylobacter* spp., 1306
    - mechanisms of, 1182t, 1185
    - in *M. pneumoniae,* 1466
    - in *N. gonorrhoeae,* 1100f
    - in *Nocardia* spp., 1359f
    - rates of, 1174t
    - in *Salmonella,* 1317, 1319
    - in *Shigella,* 1324

Ciprofloxacin (*Cont.*):
  for *Shigella* infections, 1324, 1324t
  for tularemia, 1339, 1340t
  for tularemia prophylaxis, 1340t
  for tularemia treatment, S4
  for ventilator-associated pneumonia, 1032t
  for yersiniosis, 1350
Circadian clock
  aging and, 3953–3954
  anatomic organization of, 3943–3944, 3944–3945f
  in brain health and cognition, 3951–3952
  cancer and, 3953
  cardiovascular health and, 3952–3953
  drug administration and, 3954
  endocrine systems regulated by, 3949–3950
  gastrointestinal homeostasis and, 3952
  gut microbiota and, 3952
  human microbiome and, 3834
  immune system and, 3953
  in metabolic homeostasis, 3950–3951
  molecular organization of, 3943, 3944f
  partitions of, 3949f
  period length of, 3946
Circadian cycle, 2987
Circadian pacemaker, 218
Circadian period, 3944t
Circadian phase, 3944t
Circadian rhythm, 211–212, 218–219, 3050, 3052f, 3944t
Circadian rhythm sleep disorders, 3944t, 3947t
Circadian system
  desynchrony of, 218–219, 3947, 3947f, 3947t, 3950f
  entrainment and measurement of, 3946–3947
  evolution and structure of, 3943
Circle of Willis aneurysm, 234
Circulating endothelial precursors, 545f
Circulating tumor DNA (ctDNA)
  in breast cancer, 639
  in carcinoma of unknown primary, 734
  clonal hematopoiesis of indeterminant potential contribution to, 3982, 3982f
  in colorectal cancer, 658
  for early cancer detection, 3981, 3982f
  in lung cancer, 620
  for minimal residual disease detection, 3983
  for monitoring response to treatment, 3983
  for noninvasive tumor genotyping, 3982
  for quantitative assessment of tumor burden, 3983
  technical considerations, 3983–3984
  from tumors, 3980, 3981f
Circulatory hypoxia, 277
Circulatory integrity, maintenance of, 345–346, 346f
Circumcision, 1563, 1626
Cirrhosis, **2708**
  in acute-on-chronic liver failure, 2717
  alcohol-associated, 2698, 2698t, 2708–2710, 2709f, 2726. *See also* Alcohol-associated liver disease
  in α_1-antitrypsin deficiency, 2712
  autoimmune hepatitis and, 2710
  biliary, 2710–2711. *See also* Primary biliary cholangitis (PBC)
  cardiac, 2712
  chronic hepatitis and, 2660. *See also* Hepatitis
  clinical features of
    ascites. *See* Ascites, in liver disease
    clubbing, 279
    cyanosis, 278
    diarrhea, 309
    edema, 281, 281t
    esophageal varices, 317
    hyponatremia, 348
    jaundice, 323
    pleural effusion, 2270
    splenomegaly, 480
    testicular atrophy, 3115
  coagulation disorders in, 935
  complications of, **2713,** 2713t
    AKI, 2383
    bone disease, 2717
    coagulopathy, 2717
    encephalopathy. *See* Encephalopathy, hepatic
    hematologic abnormalities, 2717
    hepatopulmonary syndrome, 2717
    hepatorenal syndrome, 2383, 2716
    malnutrition, 2717
    portal hypertension. *See* Portal hypertension
    spontaneous bacterial peritonitis, 2716
  cryptogenic, vs. autoimmune hepatitis, 2695
  in cystic fibrosis, 2713
  definition of, 2673
  etiology of, 2708t
  HBV infection and, 2673–2674, 2710
  HCC and, 659–660, 2660
  HCV infection and, 2685, 2710
  in hematochromatosis, 2712, 3335
  high-altitude travel and, 3760
  liver biopsy in, A13
  liver transplantation for, 2719
  in MASH, 2710. *See also* Metabolic dysfunction–associated steatohepatitis (MASH)
  in MASLD, 2700. *See also* Metabolic dysfunction–associated steatotic liver disease (MASLD)
  pathogenesis of, 327–328
  physical examination in, 323–324, 2629–2630
  postnecrosis, vs. autoimmune hepatitis, 2695
  recurrent, after liver transplantation, 2724
  staging of, 2632–2633, 2632–2633t
  surveillance of, 661
  telomere dysfunction and, 3818, 3820, 3822f
  treatment of, 2633
  tumor markers in, 504t
  in Wilson's disease, 2712
Cisatracurium, 2295, 2352
Cisgender, 3179t
Cisplatin
  actions of, 563
  adverse effects of, 562t, 563, 614, 709
    cardiovascular, 709, 2017
    desensitization protocol for, 2819
    electrolyte disturbances, 355, 563, S1
    encephalopathy, 2350t
    gonadal dysfunction, 706
    hearing loss, 246, 706, 726, 756
    hemolytic-uremic syndrome, 595
    late, 709
    long-term, 753t
    nausea and vomiting, 297, 563, 575
    nephropathy, 706, 757
    nephrotoxicity, 756, 2377
    neurologic, 726, 726t, 756, 3608t
    neutropenic enterocolitis, 597
    Raynaud's phenomenon, 563, 2178
  for bladder cancer, 693, 694
  for breast cancer, 637
  for carcinoma of unknown primary, 735
  for cervical cancer, 713
  for cholangiocarcinoma, 668, 669
  for gallbladder cancer, 670
  for gastric cancer, 649
  for gestational trophoblastic disease, 715
  for head and neck cancer, 613
  interactions and issues, 562t, 563
  for lung cancer, 624, 625t
  for metastatic cervical cancer, 713
  for osteosarcoma, 730
  for ovarian germ cell tumors, 712
  for salivary gland tumors, 614
  for testicular cancer, 706, 707–708f, 708–709
  for thymic tumors, 630
Citalopram
  adverse effects of, 3659t, 3666
  for depression, 85, 3659t, 3666
  genetic variations in response to, E72t
  for menopausal symptoms, 510, 3143
Citrate, urinary, 2449, 2452
Citrobacter spp. infections, 1087, 1293
Citrulline, 3373, 3816
Citrullinemia type 1, 3369t
Citrullinemia type 2 (citrin deficiency), 3369t, **3373,** 3374t
CJD. *See* Creutzfeldt-Jakob disease (CJD)
CK. *See* Creatine kinase (CK)
CKD. *See* Chronic kidney disease (CKD)
CKD-EPI: eGFR, 337–338
c-kit, 2805
c-kit mutations, 493, 602f, 605, 728, E72t
CKs. *See* Cytokeratins (CKs)
CLAD (chronic lung allograft dysfunction), 1163, 2286, 2287
Cladophialophora, 1725, 1728, 1728f
Cladribine, 2821
  adverse effects of, 873
  for B-cell prolymphocytic leukemia, 872
  for hairy cell leukemia, 873
  for mastocytosis, 881
  for MS, 3586t, 3587
CLAH (congenital lipoid adrenal hyperplasia), 3067t
Clarithromycin
  actions of, 1166, 1177, 1182t, 1427
  adverse effects of, 1177, 1427
  for *Bartonella* infection prophylaxis, 1594t
  for community-acquired pneumonia, 1027t, 1028t
  for *C. pneumoniae* infection, 1477
  drug interactions of, 1177, 1427, 1742t, 2959, 3213
  for endocarditis prophylaxis, 1048t
  in *H. pylori* eradication, 1302t, 2527t, 2528, 2529f, 2530t
  indications for, 1173t, 1177
  for *Legionella* spp. infections, 1275, 1275t
  for leprosy, 1411
  for MAC infection, 1417, 1426, 1427
  for MAC infection prophylaxis, 1417, 1593t
  for NTM infections, 1427
  for pertussis, 1280t
  for pneumococcal infection, 1194
  in pregnancy and lactation, 1169t
  resistance to, 1173t
    in *H. pylori,* 1303, 2527, 2528, 2529f
    mechanisms of, 1182t
    in *Nocardia* spp., 1359f
  for streptococcal pharyngitis, 258, 258t
Clasp-knife phenomenon, 168
Class switch, 883
Claude's syndrome, 234, 3431, 3432f
Claudication
  in aortic occlusion, 2170
  intermittent, 2172. *See also* Peripheral artery disease
  neurogenic, 120
  physical examination in, 1863, 1863f
  in thromboangiitis obliterans, 2174
  treatment of, 2173–2174
Claudin 18.2, 647, 649
Claustrophobia, 3661
Clavulanate, 1183. *See also* Ampicillin-clavulanate
Clavulanic acid, 325
CLCN1 gene mutations, 3647
CLCN5 gene mutations, 2369, 3264

- *CLCN7* gene mutations, 3315
- *CLDN16* gene mutations, 2369
- Clearance, of drugs, 486
- Clear cell carcinoma, renal, 688, 688t. *See also* Renal cell carcinoma (RCC)
- Clear liquid diet, 2603
- Cleidocranial dysplasia, 3259
- Clemastine, 3397
- Clenbuterol, 3694
- Clenched-fist injury, 1144, 1145t
- CLIA (Clinical Laboratory Improvement Amendments Act), S10
- Climate change, **3744**
  - adaptation to, 3753, 3754f
  - El Niño events and, 3747, 3747f, 3750, 3753f
  - greenhouse gases and, 3745–3746, 3745f, 3746f, 3746t
  - health effects of, **3739,** 3739f, 3754f
    - air pollution and. *See* Air pollution
    - heat-related illness, 3740, 3742f
    - infectious diseases, **3747**
      - arborvirus infections, 3748–3749
      - dengue, 3748, 3749–3750f
      - El Niño–related disease outbreaks, 3750, 3753f
      - Lyme disease, 3749, 3751f
      - malaria, 3748
      - schistosomiasis, 1827
      - vector-borne disease, 3748–3749, 3749–3750f
      - waterborne disease, 3749–3750, 3752f, 3752t
    - malnutrition, 3741, 3743–3744f
    - in natural disasters, flooding, and displacement, 3740, 3743f
    - potential solutions, 3744
  - hurricane severity and, 3747
  - mitigation of, 3752–3753
  - overview of, 3745, 3745f
  - population displacement and, 3752
  - population migration and conflict and, 3747
  - precipitation patterns and, 3746–3747
  - sea levels and, 3747
  - temperature and, 3745–3746, 3745f
  - waterborne disease and, 1269
- Clindamycin
  - for acne vulgaris, 391
  - for actinomycosis, 1366t
  - actions of, 1166, 1177, 1182t
  - adverse effects of, 1171t, 1177, 1734t, 1742t
  - for anaerobic bacterial infections, 1379
  - for anthrax, S4
  - for asymptomatic pharyngeal GAS carriage, 1209
  - for *Babesia* infections, 991t, 1779–1780, 1780t
  - for bacterial vaginosis, 1101t, 1103
  - for bite-wound infections, 1054t, 1144, 1145t
  - for *Capnocytophaga* infections, 1267
  - for cellulitis, 1054, 1054t
  - for clostridial infections, 1239t
  - for clostridial myonecrosis, 991t, 1054t, 1239t, 1242
  - for CNS toxoplasmosis, 1138
  - for head and neck abscesses, 259
  - indications for, 1173t, 1177, 1742t
  - for invasive group A streptococcal infections, 1054
  - for lung abscess, 1036
  - for malaria treatment, 1770t
  - for MRSA infections, 383, 1203t, 1204
  - for necrotizing fasciitis, 991t, 1054t, 1211
  - for osteomyelitis, 1064t, 1065t
  - for PCP, 1734t
  - for pelvic inflammatory disease, 1106, 1106t
  - for pneumococcal infections, 1194
  - in pregnancy and lactation, 1169t, 1742t
  - prophylactic uses of, 1180t
  - resistance to
    - in anaerobic bacteria, 1379, 1379t
    - in *Babesia,* 1780
    - in *C. perfringens,* 1242
    - in *Eikenella,* 1267
    - in group A streptococci, 1212
    - mechanisms of, 1182t, 1184
    - prevalence of, 1173t
  - for skin and soft tissue infections, 964t, 1054, 1203t
  - for streptococcal pharyngitis, 258, 258t
  - for toxic shock syndrome, 991t, 993, 1205, 1208t, 1212
  - for toxoplasmosis prophylaxis, 1593t
- Clinical breast examination, 515, 516t, 634
- Clinical decision-making, 3–4, 65–66. *See also* Clinical reasoning
- Clinical Frailty Scale, 2619t
- Clinical Laboratory Improvement Amendments Act (CLIA), S10
- Clinical laboratory tests, S10
  - accuracy of, 23–24, 24t, 48–49
  - batteries of, 2
  - critical values, S10
  - diagnostic principles of, S10
  - error sources in, S10
  - genetic. *See* Genetic testing
  - global considerations in, S10
  - home, S10
  - inappropriate use of, 1152–1153, 1152t
  - indications for, S10
  - in older adults, 3878
  - physician's evaluation and use of, 2
  - point-of-care, S10
  - predictive value of, 514, 514t
  - reference values for, S10
  - regulation of laboratories performing, S10
  - sensitivity and specificity of, 23–24, 24t, 514, 514t, S10
- Clinical microbiology laboratory, **S11**. *See also* Infectious diseases, laboratory diagnosis of; Virus(es), laboratory identification of
- Clinical practice guidelines, 4, 28–29, 74
- Clinical reasoning, **20**
  - clinical expertise in, 20
  - cognitive shortcuts in, 21–22
  - decision support tools for, 26–27
  - diagnostic test accuracy and, 23–24, 24t. *See also* Clinical laboratory tests
  - economic incentives and, 23
  - evidence-based medicine and, 27–29
  - hypothetico-deductive model of, 22
  - intuitive vs. analytic, 20–22
  - pattern recognition in, 21
  - personalization in, 22–23
  - practice setting and, 23
  - practice style and, 23
- Clinical research. *See* Research
- Clinical skills, 1–2
- Clinical trials
  - additivity in, 3930t, 3933–3934
  - adverse events in, 2667
  - for behavioral interventions, 3922
  - in evidence-based medicine, 27–28
  - placebo response in, 3931, 3932f
  - terms used in, 3930t
  - in vaccine development, 998–999
- Clinical uncertainty, 66
- Clinton Foundation HIV/AIDS Initiative, 3847
- CLIPPERS, 3566
- Clitoral engorgement, 3161
- Clitoral vacuum device, 3162
- CLL. *See* Chronic lymphocytic leukemia (CLL)
- CLN2 (neuronal ceroid lipofuscinosis type 2), 3354t, 3357
- Clobazam, 3421t
- Clobetasol, 386
- CLOCK, 3943
- Clock-controlled genes (CCGs), 3943, 3944t
- *CLOCK* genes, 3943, 3947t, 3952
- Clodronate, 638
- Clofazimine
  - actions of, 1425
  - adverse effects of, 400, 420, 1411, 1412, 1425, 1606
  - for leprosy, 1410, 1411t
  - for leprosy reactions, 1412
  - for MDR-TB, 1394t, 1395t, 1399, 1421t, 1425
  - for NTM infections, 1428
- Clofibrate, 2940t
- Clomiphene, 3124, 3135, 3151, 3174
- Clomipramine, 215, 420, 3659t, 3663
- Clonal dynamics, of stem cells, 761
- Clonal hematopoiesis of indeterminate potential (CHIP), 769, 825, 1858–1859, 3982, 3982f, A10
- Clonazepam
  - adverse effects of, 3421t
  - for anxiety disorders, 3660t
  - for ataxia/tremor in MS, 3588
  - for dyspnea, 83t
  - for glossodynia, 266
  - for insomnia, 216
  - for myoclonus, 3515
  - overdosage/poisoning with, 3713t
  - for pain, 97t
  - for panic disorder, 3658
  - pharmacology of, 3660t
  - for posthypoxic myoclonus, 2348
  - for REM sleep behavior disorder, 218
  - for seizures, 3421t
  - for sleep disorders in PD, 3507
  - for status epilepticus, 3426f
  - for tardive dyskinesia, 3517
  - for vertigo, 165t
- Clonidine
  - adverse effects of, 282t, 1920t, 2145t
  - for ascites, 329
  - for diabetic diarrhea, 312
  - for hypertension, 2145t
  - for inappropriate sinus tachycardia, 1936
  - for menopausal symptoms, 3143
  - for opioid withdrawal syndrome, 3691
  - overdosage/poisoning with, 3711t
  - for pain, 82
  - perioperative, 3908
  - for smoking cessation therapy, 615
  - for Tourette's syndrome, 3515
- *Clonorchis sinensis*/clonorchiasis, 667, 961t, 1739t, 1826t, 1830–1831, 1830t, 1831t, S12
- Clopidogrel
  - actions of, 943, 943f
  - adverse effects of, 473, 925, 944, 2445
  - with aspirin, 943, 3449
  - for cardiac catheterization pretreatment, 1904
  - dosage of, 944, S13
  - drug interactions of, 301, 489t, 1422, 2525–2526
  - genetic variations in response to, 486, 492, 494, 939, 939t, 2110, E72t
  - indications for, 943–944
  - for ischemic heart disease, 2102
  - management before endoscopic procedures, 2483t
  - management before lumbar puncture, S3
  - metabolism of, 485t, 486
  - for myocardial infarction
    - NSTE-ACS, 2110, 2110t, 2112
    - as secondary prevention, 2126
    - STEMI, 2121

- Clopidogrel (*Cont.*):
  - in PCI, 2127, 2128
  - in peripheral artery disease, 2173
  - resistance to, 944
  - for stroke/TIA prevention, 3449–3450
- Clorazepate, 3660t
- Closed-loop obstruction, 2590. *See also* Intestinal obstruction
- Clostridial myonecrosis (gas gangrene), **1241**
  - of bowel, 1240
  - clinical features of, 993, 1051, 1241
  - diagnosis of, 1241, 1241f, 3229
  - etiology and pathogenesis of, 991t, 993, 1053, 1241–1242, 1242f
  - histologic findings in, 1242, 1242f
  - nontraumatic (spontaneous), 1241, 1241f
  - prevention of, 1243
  - prognosis of, 1243
  - traumatic, 1241
  - treatment of, 991t, 1054, 1054t, 1239t, 1242–1243
- *Clostridioides difficile*–associated infection (CDI), **1082**
  - antibiotic resistance in, 1186t
  - antibiotics associated with, 1082
  - antibiotic stewardship in, 1152t
  - in cancer patient, 582–583, 597
  - clinical features of, 306t, 1060, 1084
  - definition of, 1082
  - diagnosis of, 976, 1080, 1084, 1084t
  - differential diagnosis of, 597
  - enterococci's effect on toxin production in, 1215
  - epidemiology of, 1079–1080, 1082
  - etiology of, 1080, 1082
  - fulminant, 1085–1086
  - global considerations in, 1083
  - in HCT recipient, 1156
  - health care–associated, 305, 1149–1150
  - vs. IBD, 2560
  - inappropriate testing in, 1152–1153, 1152t
  - microbiome and, 3838–3839
  - pathogenesis of, 1082–1083, 1083f
  - PPIs and, 1082, 2525
  - prevention and control of, 1086, 1150, 1179, 1186
  - prognosis of, 1086
  - recurrent, 1085, 3838–3839
  - risk factors for, 1082, 1149–1150
  - in solid organ transplant recipient, 1155, 1155t
  - treatment of, 1084–1085, 1085t, 1176, 3838–3839, 3841
- *Clostridium* spp., 1238, 1372, 3836
  - *C. baratii,* 1233
  - *C. botulinum,* 965, 1233, S4. *See also* Botulism
  - *C. butyricum,* 1233
  - *C. difficile,* 974, 1082, 1186, 1186t. *See also Clostridioides difficile*–associated infection (CDI)
  - *C. perfringens,* 1238–1239, 1238f, 1242. *See also Clostridium perfringens* infections
  - *C. septicum,* 1238. *See also Clostridium septicum* infections
  - *C. tetani,* 1230. *See also* Tetanus
- *Clostridium* spp. infections
  - bacteremia, 1240, 1377
  - *C. bifermentans,* 1243
  - *C. botulinum. See* Botulism
  - *C. difficile. See Clostridioides difficile*–associated infection (CDI)
  - *C. histolyticum,* 1240, 1243
  - *C. perfringens. See Clostridium perfringens* infections
  - *C. ramosum,* 1243
  - *C. septicum,* 1240, 1241–1242
  - *C. sordellii,* 993, 1053, 1240, 1243
  - *C. tertium,* 993, 1239t, 1240, 1241, 1243
  - *C. tetani. See* Tetanus
  - enteric, 1240
  - epidemiology of, 1239
  - global considerations in, 1239
  - polymicrobial infections involving, 1239–1240, 1239t
  - skin and soft tissue, 1240–1241. *See also* Clostridial myonecrosis
  - treatment of, 1239t
  - wound contamination, 1239
- *Clostridium perfringens* infections
  - bacteremia, 1180
  - clinical features of, 306t
  - clostridial myonecrosis. *See* Clostridial myonecrosis
  - food-borne, 1080, 1080t, 1240
  - genetic considerations in, 1240
  - hemolytic anemia in, 801
  - necrotizing enterocolitis, 1240
  - necrotizing fasciitis, 993. *See also* Necrotizing fasciitis
  - treatment of, 1242–1243
- *Clostridium septicum* infections, 583, 1376
- Clotrimazole, 390, 1101t, 1103, 1692, 2721
- Clotting factors. *See also specific factors*
  - defective, 407
  - engineered, 931
  - laboratory evaluation of, 929f
  - in normal hemostasis, 470f
  - specific factor assays, 475
- Cloxacillin, 1169t
- Clozapine
  - adverse effects of, 2017, 3188, 3506, 3673, 3673t, 3674
  - for DLB, 3493
  - overdosage/poisoning with, 3711t
  - for psychosis in Huntington's disease, 3514
  - for psychosis in PD, 3506
  - for schizophrenia, 3673, 3673t
- CLRs (C-type lectin receptors), 2758, 2759t
- Clubbing, of digits, **279**
  - in cardiovascular disease, 1860, 2971
  - disorders associated with, 279, 2970–2971, 2971t
  - in Graves' disease, 3034f
  - in hypertrophic osteoarthropathy, 279, 2970–2971, 2970f, 3318
  - in lung cancer, 279, 619
  - in respiratory disease, 270, 2200
  - in thyroid disease, 2971
- Clue cells, 1102, 1103f
- *CLU* gene mutations, 3479
- Cluster headache, **3472**
  - clinical features of, 3471t, 3472
  - differential diagnosis of, 115, 115t
  - imaging in, 3465f
  - prevalence of, 113t
  - prevention of, 3472t
  - treatment of, 3471t
- Cluster of differentiation (CD) classification, 2755, 2757–2758t
- Cluster-tic, 3552
- Clutton's joint, 1058
- CML. *See* Chronic myeloid leukemia (CML)
- CMML (chronic myelomonocytic leukemia), 815, 877–878, 877t
- CMR. *See* Cardiac magnetic resonance imaging (CMR)
- CMT (Charcot-Marie-Tooth) disease. *See* Charcot-Marie-Tooth (CMT) disease
- CMV infections. *See* Cytomegalovirus (CMV) infections
- Cnidarian envenomation, 3722–3723, 3723f
- CNS. *See* Central nervous system (CNS)
- CNTN1 (contactin-1), 3623
- CO$_2$ (carbon dioxide), 3745, 3745f, 3746t
- Co-activators, 3781
- Coagulase-negative staphylococci (CoNS), 1062, 1064, 1195, 3838. *See also* Staphylococcal infections
- Coagulation, 469–470, 470f, 921
- Coagulation cascade
  - extrinsic pathway, 470, 929f
  - intrinsic pathway, 470, 470f, 929f
  - overview of, 940, 940f
  - red blood cells fate in, 805f
- Coagulation disorders, **928**
  - acquired factor inhibitors and, 935–936
  - in cirrhosis, 2717
  - in CKD, 2393
  - disseminated intravascular coagulation. *See* Disseminated intravascular coagulation (DIC)
  - familial multiple coagulation deficiencies, 932–933
  - genetic characteristics of, 928t
  - in heatstroke, 3775
  - hemophilia. *See* Hemophilia
  - in hypothermia, 3770
  - laboratory evaluation of, 928t, 929f
  - in liver failure, 935, 935t
  - in sepsis/septic shock, 2319
  - in vitamin K deficiency. *See* Vitamin K deficiency
- Coagulation factor(s). *See also specific factors*
  - in hepatobiliary diseases, 2636, 2637t
  - multiple deficiencies of, 932
  - vitamin K–dependent, 932–933, 933f
- Coagulation factor inhibitors, 928, 930–931, 935–936
- Coagulation tests, 474–475, 475t
- Coal dust, 2236t
- Coal tar, 393t
- Coal worker's pneumoconiosis, 2238
- Coarctation of the aorta, **2076,** 2076f, 2140
- Coats' disease, 3642
- Cobalamin (vitamin B$_{12}$), **780**
  - absorption of, 781
  - deficiency of. *See* Cobalamin (vitamin B$_{12}$) deficiency
  - dietary sources of, 781, 785
  - folate and, 782
  - functions of, 2606f
  - for hypogammaglobulinemia, S9
  - malabsorption of, 243, 786–787, 786t
  - metabolism of, disorders of, 787
  - postgastrectomy levels of, 2534
  - requirement/recommended intake of, 781, 2600t
  - serum, 788
  - structure of, 780–781, 2606f
  - transport in blood, 781
- Cobalamin (vitamin B$_{12}$) deficiency. *See also* Megaloblastic anemia
  - after cystectomy, 694
  - clinical features of, 782–784, 2605t, 3484
    - anemia, 449
    - cardiovascular, 784
    - dementia, 3483–3484
    - frontal lobe syndrome, 208
    - gait disorders, 179, 3533
    - hyperpigmentation, 400
    - jaundice, 323
    - myelopathy, 3483, 3568, A16
    - neuropathies, 3612
    - olfactory dysfunction, 243
    - optic neuropathy, 228
    - oral, 65
    - peripheral neuropathy, 783
  - diagnosis of, 788–789, 3612, A16
  - etiology of, 785–786t, 785–787, 3483, 3611–3612

  - in fish tapeworm infestations, 1839
  - hematologic findings in, 439f, 443, 453, 784–795, 785f
  - in HIV infection, 1605
  - in infants, 785, 787
  - postgastrectomy, 786
  - in pregnancy, 784
  - risk factors for, 2605t
  - in small-bowel bacterial overgrowth, 2543
  - small-intestinal mucosal biopsies in, 2550t
  - treatment of, 789–790, 3484, 3612
  - in tropical sprue, 2546
- Cobalt exposure, 2017, 2238, 3702, S6
- Cobblestone papillae, 224
- Cobicistat, 1601, 1621, 1621t
- Cobimetinib
  - actions of, 569t
  - adverse effects of, 569t, 753t
  - in combination therapy, 572
  - for metastatic melanoma, 604, 604t, 605
- Cobras, 3717, 3722. *See also* Snakebites
- Cocaine. *See also* Psychostimulants
  - actions and effects of, 3656t, 3692
  - adverse effects of
    - cardiomyopathy, 2016
    - hypertension, 2137
    - intracerebral hemorrhage, 3447, 3454
    - myopathy, 2940t, 3648t
    - nasal and auricular, 2898
    - palpitations, 291
    - skin necrosis, 421
    - steatosis, 2703t
    - vasculitis, 2940t
  - forms of, 3692
  - overdose and abuse of
    - clinical features of, 3705, 3710t, 3715t
    - epidemiology of, 3692–3693
    - treatment of, 3710t, 3715t
  - pharmacology of, 3692
- *Coccidioides* spp., 1696, 1697f
- Coccidioidomycosis, **1696**
  - arthritis, 1058
  - clinical features of, 225, 406, 1688t, 1697, 1698t
  - complications of, 1697
  - diagnosis of, 961t, 1688t, 1697–1698
  - disseminated, 1697, 1699
  - epidemiology of, 961t, 1696
  - in HIV infection, 1594t, 1597
  - meningitis, 1130t, 1697, 1698, 1699, A16
  - pathogenesis of, 1696–1697
  - prevention of, 1699
  - risk factors for, 1688t
  - skin manifestations of, 1051, 1697
  - in solid organ transplant recipient, 1162, 1699
  - subacute meningitis, 1126
  - treatment of, 1698–1699, 1698t
- Cochlea, 245f
- Cochlear hydrops, 246
- Cochlear implant, 246, 250–251, 251f
- Cochrane Database of Systematic Reviews, 27
- Cockcroft-Gault equation, 337
- Coclé virus, 1662t, 1668
- Codeine
  - adverse effects of, 492
  - for cough, 274
  - for diarrhea, 312
  - drug interactions of, 489t, 3660t
  - for dyspnea, 83t
  - genetic variations in response to, 486, 492, E72t
  - metabolism of, 485t, 486
  - for pain, 80f, 81, 97t
- Coding region, of mtDNA, 3807
- Codman's triangle, 730
- Codons, 3779
- Coenurosis, 1813
- Coffee worker's lung, 2228t
- Cogan's syndrome, 2166, 2906–2907
- Cognitive behavioral therapy (CBT)
  - for alcohol use disorder, 3680
  - for fatigue, 167
  - for fibromyalgia, 2965
  - for glossodynia, 266
  - for IC/BPS, 335
  - for insomnia, 216, 3927
  - for low back pain, 126
  - for menopausal symptoms, 3145
  - for metabolic syndrome, 3257
  - for obesity, 3191
  - for phobic disorders, 3661
  - for stimulant use disorders, 3695
- Cognitive dysfunction
  - alcohol use and, 3676
  - assessment of, 183, V2
  - caring for patient with, 209
  - chemotherapy-related, 756
  - delirium and, 182
  - in folate/cobalamin deficiency, 784
  - in MS, 3578, 3588
  - in older adults, 3885–3886, 3888f
  - postmenopausal hormone therapy and, 3144t, 3146, 3164
- Cognitive rehabilitation procedures, 209
- Cogwheel rigidity, 3381
- Cohen's syndrome, 3184t
- *COL1A1/2* gene mutations
  - in chondrodysplasias, 3331
  - in Ehlers-Danlos syndrome, 3328t, 3330
  - in osteogenesis imperfecta, 3259, 3320, 3325t, 3787
- *COL3A1* gene mutations, 3301t, 3302t, 3330
- *COL4A1* gene mutations, 3448
- *COL4A3/4/5* gene mutations, 2359, 2360
- *COL5A1/2* gene mutations, 3328t, 3330
- *COL9A* gene mutations, 3331
- *COL12A1* gene mutations, 3328t, 3330
- *COL17A1* gene mutations, 3785
- Colchicine
  - for acute idiopathic pericarditis, 2080
  - adverse effects of, 2936
    - alopecia, 394
    - cobalamin malabsorption, 786t
    - diarrhea, 309, 2080
    - myopathy, 2940t, 3648t
    - neuropathy, 3609t, 3610
  - for Behçet syndrome, 2909
  - for calcium apatite deposition disease, 2962
  - for calcium pyrophosphate deposition disease, 2961
  - contraindications to, 2080
  - drug interactions of, 489t, 2959
  - for familial Mediterranean fever, 899, 2936
  - for fever of unknown origin, 155
  - for gout, 2959
  - for ischemic heart disease, 2103
  - for SLE, 2835
  - for urticaria, 2809
- Cold abscess, 463
- Cold agglutinin disease, 802–803, 803f
- Cold exposure, cyanosis in, 279
- Cold-stimulus headache, 3474
- Cold urticaria, 402, 2807
- Colectomy, 314, 514, 654, 1085
- Colesevelam, 3214, 3244t, 3251–3252, 3258
- Colestipol, 3244t, 3251–3252, 3257–3258
- Colic, biliary. *See* Biliary colic
- Colistin. *See* Polymyxin E (colistin)
- Colitis
  - amebic, 1756
  - bleeding in, 318
  - CMV, in HIV infection, 1600
  - collagenous, 309, 2561
  - colonoscopy in, 2465, 2470f
  - in Crohn's disease, 2558
  - diarrhea in, 309
  - diversion, 2561
  - etiology of, 2470f
  - vs. IBD, 2561
  - ischemic, 306, 2470f, 2561, 2589
  - lymphocytic, 2561
  - microscopic, 309
  - primary immune deficiency diseases and, 2803–2804
  - pseudomembranous, 1079, 1082–1083, 1083f, 2470f. *See also Clostridioides difficile*–associated infection (CDI)
  - sexually acquired, 1109
- Collagenoma, 3080t, 3084
- Collagenous colitis, 2561
- Collagens
  - autoantibodies to, 2790t
  - in connective tissues, 3320, 3321t
  - fibrillar, 3320, 3322–3323, 3323f
  - genes for, 3320
  - subepithelial deposition of, 2218
- Collagen type IV-3α-chain, 2790t
- Collagen vascular disease, 270, 2514
- Collapsing falls, 181. *See also* Fall(s)
- Collecting duct, 2367f, 2368t, 2370
- Collectins, 2758
- Colles' fracture, 3292f
- Collier's sign, 236
- Colloid cysts, 722, A16
- Colon
  - bleeding from, 318
  - blood supply to, 2587, 2587f
  - cathartic, 313
  - diverticula of. *See* Diverticular disease
  - functions of, 2459
  - obstruction of. *See* Intestinal obstruction
  - physiology of, 303, 303t
  - polyps of, 517, 2465f, 2470f, 2480–2481f. *See also* Colorectal cancer
  - storage and salvage in, 303–304
  - in systemic sclerosis, 2870, 2870t
  - transit time, measurement of, 313
- Colonic obstruction. *See* Intestinal obstruction
- Colonic pseudoobstruction. *See* Intestinal pseudoobstruction
- Colonization, 964
- Colonization resistance, 3838
- Colonoscopy, 2465
  - in colitis, 2465, 2470f
  - colonic mucosa on, 2468f
  - for colorectal cancer screening
    - in average-risk patients, 39t, 516t, 655, 2502t
    - costs of, 40
    - effectiveness of, 517
    - vs. flexible sigmoidoscopy, 655–656
    - in inflammatory bowel disease, 2502t
    - with personal or family history, 2502t
    - polyp appearance in, 2465f, 2470f, 2471f, 2479f, 2503f
    - sensitivity of, 2465
    - strategies for, 2496–2497, 2502t
  - in constipation, 312
  - in diarrhea, 307, 311
  - in diverticular disease, 2580
  - in gastrointestinal bleeding, 320
  - indications for, 2462, 2462t
  - in intestinal ischemia, 2589
  - normal, 2469f
  - open-access, 2499–2500
  - patient preparation for, 2500
  - polyp removal during, V5
  - therapeutic, 2464
  - in ulcerative colitis, 2557, 2557f
  - virtual, 656, 2497, 2503f

Colony-stimulating factors, 453
Colorado tick fever, 1657t, 1669
Color anomia, 203, 222
Color blindness, 221–222, 3535
Colorectal cancer, **653**
  adenomatous polyps and, 653
  carcinoma of unknown primary and, 735
  clinical features of, 656, 656f
  deaths from, 497t, 499f, 500f, 500t, 501t, 653
  endoscopic appearance of, 2471f
  epidemiology of, 653
  etiology of, 653–654, 653t
  genetic considerations in
    chromosome abnormalities, 520t, 521–522
    familial syndromes, 521–522, 524t, 653–654, 654t
  IBD and, 2570, 2570f
  incidence of, 497t, 498, 498f, 500f, 501t, 653
  inflammatory bowel disease and, 655
  lifetime cumulative risk of, 37t
  metastatic, 406, 657, 658, 723t
  pathogenesis of, 520f, 653
  postmenopausal hormone therapy and, 3144t, 3146, 3147f
  prevention of, 511, 513, 514, 655
  prognosis of, 657, 657f, 657t
  recurrence of, 656–657
  risk factors for, 653–654, 653t
  screening for. *See also* Colonoscopy, for colorectal cancer screening
    benefits of, 38, 40, 655
    in inflammatory bowel disease, 2502t
    methods for, 517–518
    in older adults, 3881t
    with personal or family history, 2502t
    recommendations for normal-risk adults, 39t, 516t, 656, 2502t
    techniques of, 655–656, 655t
  spread of, 657
  staging of, 656–657, 657f
  treatment of, 657–658
    endoscopic dissection, 2479–2480f, V5
    endoscopic stent placement, 2494f, V5
  tumor markers in, 504t
Color flow Doppler, 1876–1877, 1877f. *See also* Echocardiography
Color vision, assessment of, 221–222
Coltiviruses, 1657, 1657t, 1669
Coma, **186**
  anatomy of, 187, V4
  approach to the patient, **189,** 190f, V4
  brain death determination in, 192
  definition of, 186
  diagnosis of, 188f, 191, V4
  differential diagnosis of, 191–192, 191t
  pathophysiology of, 187–189, 187f
  prognosis of, 192–193
  treatment of, 192
Coma bullae, 401
Co-managed medicine, 3895–3896
Combat veterans. *See* Military veterans
Combination antiretroviral therapy (cART). *See also specific drugs*
  adverse effects of, 1617–1619t, 2672, 3033, 3609t, 3611
  antineoplastic therapy and, 579
  combination formulations of, 1621t
  in developing countries, 3847, 3847f
  entry inhibitors, **1618t,** 1620f, 1621, 1623f
  formulations of, 1621t
  in HIV infection, 1617t
  indications for changing, 1624–1625, 1624t
  initiation of, 976, 1621, 1624t
  integrase inhibitors, **1618–1619t,** 1620f, 1621, 1623f
  in pregnancy, 3904
  principles of, 1616, 1621–1625, 1624t
  protease inhibitors, **1617–1618t**
    actions of, 1621
    adverse effects of, 2940t
    drug interactions of, 485t, 1400, 1402
    for HCV infection, 2687–2689, 2688t
    for HIV infection, 1617–1618t
    molecular structure of, 1619f
    resistance to, 1623f
  resistance testing for, 1590
  reverse transcriptase inhibitors, **1617t**
    actions of, 1616
    adverse effects of, 369, 1616
    drug interactions of, 1402
    for HIV infection, 1616, 1617t
    molecular structure of, 1619f
    resistance to, 1622–1623f
  selection of initial drugs, 1624–1625, 1624t
Combined pituitary hormone deficiency (CPHD), 3068t
Comedones, 390, 391f
"Comet sign," 3730, 3731f
Commensal bacteria, 3831
Common carotid artery, 3431
Common cold, 253, 1537
Common variable immunodeficiency (CVID), 858, 858t, 2802
Commonwealth Serum Laboratories, 3723
Commotio cordis, 2088
Communication
  of bad news, 75–76, 77t
  racial/ethnic factors in, 66, 66f
Community-acquired pneumonia. *See* Pneumonia, community-acquired
COMP (cartilage oligomeric matrix protein), 3321t
Comparative genomics, 3778
COMP gene mutations, 3331
Complacency, in vaccine acceptance, 15, 15f
Complementary and integrative therapies and practices, 3924
  with combined psychological and physical input, 3928
  definitions and scope of, 3924
  drug interactions of, 3926–3927, 3927t
  for gastrointestinal disease, 2463–2464
  multicomponent therapies and systems, 3928
  for pain, 3925
  patient and provider resources, 3929, 3929t
  patterns of use of, 3924–3925
  with primary dietary input, 2672, 3925–3926. *See also* Herbal medicines/supplements
  with primary physical input, 3927–3928
  with primary psychological input, 3927
  regulation of, 3926
  research challenges in, 3928–3929
  toxicity of, 3926
Complement C3, 2353
Complement C5, 2792t
Complement factor H, 231
Complement receptor for C3bi, 2786f
Complement system
  in bacterial infection, 969, 970f
  deficiencies of, 403, 1132t, 1247, 2782, 2783t, 2786f, 2797
  definition of, 2755
  in HIV infection, 1588
  innate immune system and, 2758t, 2771
  pathways of, 2353, 2771, 2773f, 2783t
Complete blood count (CBC), 478, 2618t
Complete gonadal dysgenesis (Swyer's syndrome), 3098
Complete (third-degree) heart block, 1925, 1925t, 1926f, A8
Complete response, to chemotherapy, 555
Complex partial status epilepticus, 3484
Complex regional pain syndrome (CRPS), 96, 2971
Complex sleep apnea, 2282
Complex systems, 3955
Compound heterozygote, 3787
Comprehension, assessment of, 201
Compression, in physical examination, 121t
Compression stockings
  for chronic venous disease, 2181
  for DVT, 2162
  for lymphedema, 2184
  for stasis dermatitis/ulceration, 385
Compression ultrasonography, of leg veins, 742
Computed tomography (CT)
  in adrenocortical carcinoma, 3066f
  in appendicitis, 2596, 2596f
  in ARDS, 2301f, A12
  in asbestosis, A12
  in aspergillosis, 1715f, 1716, 1716f
  in asthma, 2223
  in axial spondyloarthritis, 2885f
  in bile duct disease, 2736t
  in bladder cancer, 691
  for body composition study, 2618t
  in bone mass measurement, 3297
  in bronchiectasis, 2243, A12
  in bronchopulmonary dysplasia, A12
  in CAH, 3072f
  in carcinoma of unknown primary, 732
  in cardiac disease
    angiography. *See* Computed tomography (CT) angiography (CTA)
    CAD, 1883–1887, 1886f, 1887f, 1888–1889
    for calcium scoring, 1880, 1881f, 1886f, A12
    contrast agents in, 1883–1884
    infective endocarditis, 1898–1899, 1901f
    for pulmonary vein mapping, A10
  chest, **A12**
  in cholestatic conditions, 324
  in chronic cough, 274
  in chronic thromboembolic pulmonary hypertension, 2193f
  in cirrhosis, 327
  in colorectal cancer, 656
  complications of, 3384, 3386
  in COPD, 2250f, 2255
  in COVID-19 disease, 1541
  in Cushing's syndrome, 3058, 3059f
  in echinococcosis, 1837f
  in EGPA, 2233
  in gastrointestinal disease evaluation, 2462
  in granulomatosis with polyangiitis, 2896f
  halo sign on, 1716
  in HCC diagnosis, 661
  in hearing loss evaluation, 250
  in hemoptysis, 276
  in HSV encephalitis, 1499f
  in hypersensitivity pneumonitis, 2229, 2230f
  in interstitial lung disease, 2262, 2263f, A12
  in intestinal obstruction, 2592, 2592f
  in *Legionella* infections, 1273, 1274f
  in liver disease, 2631, 2631t
  in lung cancer, 620, 625, A12
  for lung cancer screening, 516t, 518, 617–618, 618t
  in lymphadenopathy, 478
  in lymphangioleiomyomatosis, A12
  in mastoiditis, 255
  in McCune-Albright syndrome, A16
  in mediastinal masses, 2271
  in mineralocorticoid excess, 3061
  in musculoskeletal disease evaluation, 2947, 2947t, 2948f
  in neck pain, 130

  in nephrolithiasis, 2450, 2452
  in neurologic disease
    adrenoleukodystrophy, A16
    anterior choroidal artery infarct, A16
    artery of Percheron infarct, A16
    carbon monoxide poisoning, A16
    cavernous malformation, 3389f
    cerebral contusion, 3573f
    CNS vasculitis, A16
    coccidiomycosis meningitis, A16
    coma, 191
    encephalitis, 1113–1114
    ependymoma, A16
    epidural hematoma, 3572f, A16
    glioblastoma, A16
    guidelines for, 3383t
    Huntington's disease, A16
    hyperglycemic hemichorea, A16
    indications for, 3384
    intracerebral hemorrhage, 3453f, 3455
    Jefferson fracture, A16
    middle cerebral artery occlusion, 3385f
    mucormycosis, A16
    neurologic critical illness, 2344
    posterior choroidal artery infarct, A16
    seizures, 3418
    spinal ligament injury, A16
    stroke, 3435, 3437, 3437f, 3447f
    subarachnoid hemorrhage, 3459, 3460f
    subdural hematoma, 3572, 3572f
    superior sagittal sinus thrombosis, A16
    techniques for, 3383–3384
    traumatic axonal injury, 3573f
    vertebral fracture, A16
    Wallenberg syndrome, A16
  in neutropenic enterocolitis, 597f
  in osteomyelitis, 1062
  in otosclerosis, A16
  in pancreatic cancer, 673, 673f
  in pancreatic disease evaluation, 2738, 2738f, 2739t, 2741
  in pancreatitis, 2746–2747, 2746t, 2747f, 2752, 2753f
  in PCP, 1731, 1733f
  in peripheral bronchoscopy, 2210
  in pheochromocytoma, 3074t, 3075f
  photon-counting, 3384, 3386f
  in pleural effusion, A12
  in pneumatocele, A12
  in pneumothorax, A12
  in pulmonary embolism, 2161, 2161f, A12
  in pulmonary hypertension, 2189
  in pulmonary MAC infection, 1417, 1417f
  in RCC, 688
  in respiratory disease evaluation, 2201, 2211–2212, 2212f
  in sarcoidosis, 2925, 2926f
  in sinusitis, 257
  in superior vena cava syndrome, 586, 586f
  of temporal bone, 248, 250
  in tularemia, 1339, 1340f
  in urinary tract obstruction, 2456
Computed tomography angiography (CTA)
  coronary
    in anomalous right coronary artery, A9
    in aortic aneurysm, 2168f
    in aortic stenosis, 1891
    in aortitis, A9
    in CAD evaluation, 1880, 1881f, A9
    in chest pain evaluation, 108, 1889–1890, 1889f, 1890f
    contrast agents in, 1883–1884
    in coronary aneurysm, A9
    in infective endocarditis, 1040
    in ischemic heart disease, 2097, 2097f
    in murmur evaluation, 290
    in pericardial disease, 1899f
    principles of, 1880
  neurologic, 3384
    in cerebral artery aneurysm, 3384f
    in cerebral artery occlusion, 3385f
    in small-vessel stroke, 3447f
    in subarachnoid hemorrhage, 3460f
  in pulmonary hypertension, 2190
  of renal arteries, 2150, 2151t, 2153f
Computed tomography (CT) colonography, 516t, 518, 2497, 2503f
Computed tomography (CT) myelography, 130, 3384, 3392–3393
Computerized physician order entry (CPOE), 43
Computer vision, 3973
COMT gene, 3933
COMT (catechol-*O*-methyltransferase) inhibitors, 3503, 3504t
Concentration gradients, 2366
Concussion, 3571, 3574–3575, 3575t, S8
Condoms, 1260, 1626, 3154
Conduction aphasia, 201t, 202, 204
Condyloma acuminata (anogenital warts)
  clinical features of, 1051, 1527, 1527f, A5
  in HIV infection, 1606
  treatment of, 390, 1230–1231, 1489, 1531t
Condylomata, 1051
Condylomata lata, 392, 1602, A1, A5
Cone(s), 220, 221
Cone-nose bug bite, 3736
Cone snail envenomation, 3724
Confabulation, 207
Confidence, in vaccine acceptance, 15f, 16
Confidentiality, in patient-physician relationship, 71
Conflicts of interest, 73
Confocal microscopy, 607f
Confusion, 182, 186. *See also* Delirium
Confusion Assessment Method, 183
Conga ant, 3736
Congeners, 3675
Congenital absence of vagina, 3101
Congenital adrenal hyperplasia (CAH), **3071**
  clinical features of
    ambiguous genitalia, 3101, 3789
    hirsutism, 3101, 3138
    hypokalemia, 355
    precocious puberty, 3109
    salt-wasting crisis, 3101
    sex differences in, 3071
  diagnosis of, 3071, 3072f
  etiology of, 3100–3101
  genetic considerations in
    CYP gene mutations, 3060t, 3071, 3071t, 3101, 3101t
    gene conversion, 3785
    sex-limited factors, 3071, 3789
  treatment of, 3071–3072, 3101
  variants of, 3071, 3071t
Congenital amaurosis. *See* Leber's hereditary optic neuropathy (congenital amaurosis)
Congenital central hypoventilation syndrome, 2276
Congenital erythropoietic porphyria (CEP), 3341t, 3342, **3349**
Congenital generalized lipodystrophy, A15
Congenital heart disease (CHD), adult, 2067
  atrial septal defect. *See* Atrial septal defect (ASD)
  AV conduction block in, 1927
  cardiac development and, 2069
  chest wall deformity in, 1860
  clubbing/hypertrophic osteoarthropathy in, 279
  coarctation of the aorta, **2076,** 2076f
  cyanosis in, 278
  Ebstein's anomaly, 1941, 2071, 2071f
  genetic considerations in, 2069
  global considerations in, 2078
  head and neck features of, 1860
  imaging in, 1899–1900, 1901f, 1902f
  intracardiac shunt in, 1907, 2071
  multi-organ considerations in, 2067–2068, 2068t
  palliative care in, 2078
  partial anomalous pulmonary venous return, 2071, 2071f
  patent ductus arteriosus (PDA), 1866, 2066, 2072, 2072f
  percutaneous interventions for, 2130
  polycythemia in, 451
  pregnancy and, 2068, 2069t, 3900
  prevalence of, 2067
  protein-losing enteropathy in, 2549
  pulmonary arterial hypertension in, 2191
  right heart dilation, 2070, 2070t
  single ventricle physiology, 2076, 2077f
  special considerations in, 2067–2068
  as subspecialty in cardiology, 2067
  sudden cardiac death in, 2335, 2335t
  terminology in, 2068–2069
  tetralogy of Fallot. *See* Tetralogy of Fallot
  transposition of the great arteries. *See* Transposition of the great arteries
  in Turner's syndrome, 3097
  unrepaired. *See* Eisenmenger's syndrome
  ventricular septal defect. *See* Ventricular septal defect (VSD)
Congenital hypothyroidism, 2983, 3021, **3027,** 3028t
Congenital infections. *See* Neonates, infections in
Congenital lipoid adrenal hyperplasia (CLAH), 3067t
Congenital myasthenic syndromes (CMS), 3627–3628, 3629t
Congenital renal disease, 2359t, 2360–2361
Congenital rubella syndrome. *See* Rubella, congenital
Congenital syphilis. *See* Syphilis, congenital
Congenital vascular malformations, 3456–3457
Congestive heart failure. *See* Heart failure (HF)
Conivaptan, 351, 738
Conjugated equine estrogens, 3148, 3302
Conjugate eye movement, 190, 190f
Conjunctivitis, **224**
  acute hemorrhagic, 1636, 1636f
  adenoviral, 1535
  allergic, 224
  atopic, 224
  C. trachomatis, 1472, 1473t, 1474–1475
  gonococcal, 1257, 1259t
  in measles, 1534
  in mucous membrane pemphigoid, 413
  Onchocerca, 1823
  physical examination in, 47t
  in tularemia, 1338
Connective tissue(s), 3319–3323, 3321–3322t, 3323f
Connective tissue disorders, **3319**
  Alport's syndrome. *See* Alport's syndrome
  chondrodysplasia, **3330**
  classification of, 3319–3320
  Ehlers-Danlos syndrome. *See* Ehlers-Danlos syndrome
  epidermolysis bullosa, 296, 3320, 3785
  interstitial lung disease in. *See* Interstitial lung diseases (ILDs)
  Marfan's syndrome. *See* Marfan's syndrome
  osteogenesis imperfecta. *See* Osteogenesis imperfecta
Connexin(s), 247
Connshing syndrome, 3061

- Conn's syndrome. *See* Hyperaldosteronism, primary
- CoNS (coagulase-negative staphylococci), 1064, 1195, 3838. *See also* Staphylococcal infections
- Consciousness, assessment of, 3379
- Constant routine, 3944t
- Constipation, **312**
  - abdominal swelling in, 327
  - approach to the patient, 312, 313f
  - assessment of, 82
  - in cancer patient, 509
  - definition of, 312
  - differential diagnosis of, 2460–2461
  - epidemiology of, 302–303
  - etiology of, 82, 312, 312t
  - fecal incontinence and, 2584
  - in IBS, 2571
  - idiopathic, 312
  - in MS, 3588
  - opioid-induced, 81, 98
  - in PD, 3495, 3506
  - rectal prolapse and, 2582
  - severe, 314
  - in terminally ill patient, 82
  - treatment of, 82, 82t, 98, 314
- Construction apraxia, 206
- Consultative medicine. *See* Medical consultation
- "Consumption." *See* Tuberculosis (TB)
- Contact dermatitis, **384**
  - allergic, 379, 384, 393t, 423, 423f, A5
  - clinical features of
    - acute phase, 381f
    - chronic phase, 381f
    - distribution of skin lesions, 378t, 380f, 393t
    - erythroderma, 393t
    - morphology of skin lesions, 378t, 393t
    - vesicles/bullae, 400
  - diagnosis of, 384, 393t
  - drug-induced, 423, 423f
  - irritant, 384, 393t, 423
  - systemic, 423
  - treatment of, 384, 393t
- Contactin-1 (CNTN1), 3623
- Contact precautions, 1148, 1538
- Contact urticaria, 2807
- Contig, 976t, 980f
- Contiguous gene syndromes, 3783, 3789
- Continuing education, 7–8
- Continuous ambulatory peritoneal dialysis (CAPD), 1072, 2384–2385, 2400. *See also* Peritoneal dialysis
- Continuous arteriovenous rewarming, 3770t
- Continuous cyclic peritoneal dialysis (CCPD), 2398, 2400. *See also* Peritoneal dialysis
- Continuous glucose monitoring (CGM), 3207, 3208f
- Continuous positive airway pressure (CPAP)
  - adverse effects of, 2281t
  - for obesity hypoventilation syndrome, 2276
  - for sleep apnea, 2281–2282
- Continuous quality improvement, 44
- Continuous renal replacement therapy, 2323, 2330, 2384–2385
- Continuous wave Doppler, 1876–1877, 1877f. *See also* Echocardiography
- Contraception, **3152**
  - AEDs and, 3427
  - barrier methods, 3154
  - during breastfeeding, 3154
  - counseling for, 3154
  - effectiveness of, 3153t
  - emergency, 3154
  - hormonal methods, 3152t, 3153–3154. *See also* Oral contraceptives
  - intrauterine devices, 3153–3154
  - permanent, 3152–3153
- Contraction alkalosis, 374
- Contractures
  - in Emery-Dreifuss muscular dystrophy, 3641
  - etiology of, 2942, 2942t
  - in myopathies, 3634
- Contrast agents
  - adverse effects of
    - allergic reactions, 1904, 3386
    - immediate cutaneous, 417
    - mechanisms of, 422
    - nephrogenic fibrosing dermopathy, 2394
    - nephrogenic systemic fibrosis, 420, 3388, 3389
    - nephropathy, 1884, 1904, 2377, 2379t, 2394, 3384, 3386
    - pustular eruptions, 426. *See also* Acute generalized eruptive/exanthematous pustulosis (AGEP)
    - seizure, 3414t
    - urticaria/anaphylaxis, 422
  - for cardiac imaging, 1883–1884
  - for MRI, 3387–3488
  - precautionary measures for, 2394, 3386, 3387t
- Contrast enema. *See* Barium radiography
- Control chart, 44
- Controlled attenuation parameter (CAP), 2705
- Controlled human infection models (CHIMs), 999
- Control, of disease, 1002
- Control region, of mtDNA, 3807
- Contusion, cerebral, 3573, 3573f
- Conus medullaris syndrome, 3560
- Convenience, in vaccine acceptance, 15f, 16
- Convergence-projection hypothesis, of referred pain, 95f
- Convergent phenotype, 31, 31f
- Conversion disorders, 3485, 3670
- Convolutional neural network, 3972, 3972t, 3973, 3973f
- Cooley's anemia. *See* Thalassemia syndromes
- Cooling strategies, in heat-related illness, 137, 3775
- Coombs test, 801, 908, 910f
- Coordination, assessment of, 3382
- COP (cryptogenic organizing pneumonia), 2263f, 2264–2265, 2264f
- COPA, 2786f
- Copanlisib, 571t, 574, 864
- COPA syndrome, 2937
- COPD. *See* Chronic obstructive pulmonary disease (COPD)
- Copeptin, 344, 353, 3018
- Copper, **2615**
  - chelation, 3338, 3340
  - deficiency of, 817, 2534, 2614t, 2615, 3612–3613
  - functions of, 2615
  - recommended intake of, 2601t
  - tolerable upper intake level of, 2614t
  - toxicity of, 2614t, 2615. *See also* Wilson's disease
  - urine, 3339
- Copperhead bite, 3720t, 3722. *See also* Snakebites
- Coprolalia, 3515
- COPRO-oxidase, 3341t, 3342t, 3343, 3348
- Coproporphyria, hereditary. *See* Hereditary coproporphyria (HCP)
- Coproporphyrin, 2642
- Copy number variations (CNVs), 3779, 3782f, 3988–3989
- Coral snake bite, 3720–3721t, 3722. *See also* Snakebites
- Cord blood screening, 776
- Cord blood stem cells, 915
- Cord blood transplantation, 915
- Core body temperature (CBT), 3944t, 3946
- Core-needle biopsy, 553, 2213–2214
- Co-repressors, 3781
- CoreValve, A11
- Cori syndrome (type IIIa GSD), **3359,** 3360t
- "Corkscrew" esophagus, 2508, 2509f
- Corneal abrasions, 222
- Corneal blink reflex, 191, 191f
- Corneal light reflex, 221
- Corneal ulcers, 1739t
- Corollary discharge, 268
- Coronary angiography
  - in cardiogenic shock, 2327
  - coronary artery visualization on, 1908, 1908f
  - CT. *See* Computed tomography (CT) angiography (CTA)
  - fractional flow reserve measurement in, 1909, 1910f
  - intravascular ultrasound in, 1908–1909, 1909f
  - optical coherence tomography during, 1908–1909, 1909f
  - in STEMI, 2120
  - for vasospastic angina, 2112
- Coronary arteriography, 2097–2098
- Coronary artery(ies)
  - aneurysm of, A9
  - calcification of, 1880, 1881f, 1886f, 2095–2097, A12
  - chronic total occlusion of, A11, A11v
  - CT angiography of. *See* Computed tomography (CT) angiography (CTA)
  - CT of. *See* Computed tomography (CT), in cardiac disease
  - embryonic development of, 1844
  - MRI of. *See* Cardiac magnetic resonance imaging (CMR)
  - normal, 1908, 1908f
  - perforation of, during PCI, A11, A11v
  - perfusion imaging of. *See* Myocardial perfusion imaging
  - stenosis of, 1908, 1908f, 1909f
  - thrombosis in, A10
- Coronary artery aneurysm, 2907
- Coronary artery bypass grafting (CABG)
  - brain injury following, 2348
  - for cardiogenic shock, 2327
  - complications of, 2118t
  - for heart failure, 1999
  - for ischemic heart disease, 2104–2105, 2105f, 2129
  - for NSTE-ACS, 2111
  - vs. PCI, 2105, 2105f
  - pleural effusion after, 2270
  - prophylactic, before noncardiac surgery, 3907
  - racial/ethnic disparities in referral for, 62
  - survival benefit of, 2104
  - in women, 3164
- Coronary artery disease (CAD). *See also* Atherosclerosis; Ischemic heart disease
  - in alcohol use disorder, 3677
  - in allografts, 2030, 2031f
  - anticoagulant/antiplatelet therapy for, S13
  - AV conduction block in, 1926
  - clinical features of, 1841
    - angina. *See* Angina
    - cardiac symptoms, 1841
    - dyspnea, 269, 269t
    - oral pain, 65
    - palpitations, 291
  - continuum of, A10
  - deaths from, 3163f
  - diagnosis of
    - CMR, 1887, 1888f, 1895f
    - coronary angiography, 1908, 1908f
    - CT, 1883–1887, 1886f, 1887f, 1888–1889
    - CT angiography, 2097
    - ECG, A7
    - imaging tests, 1884–1890, 1884t

    - myocardial perfusion imaging, 1884t, 1885f, 1886f, 1894f, 1895f, A9
    - pre- and posttest probability analysis, 25–26, 25f, 1884
    - sex differences in, 3164–3165
    - stress echocardiography, 1884–1885, 1902v, A9
  - diet and, 2603
  - dyslipidemia and, 3243
  - heart failure in. *See* Heart failure (HF)
  - high-altitude travel and, 3759
  - in HIV infection, 1598
  - metabolic syndrome and, 3254
  - pathophysiology of, 2091–2092, 2790t, A10
  - PCI for. *See* Percutaneous coronary interventions (PCI), for revascularization
  - periodontal disease and, 261
  - peripheral artery disease and, 2172–2173
  - postmenopausal hormone therapy and, 3144t, 3145–3146, 3147f
  - preoperative testing for, 3906–3907, 3906f
  - radiation-associated, 871
  - risk factor modification for, A10
  - screening for, 3881t
  - sex differences in, 3163f
  - sudden cardiac death and, 2334–2335, 2334f, 2335t
  - vasoconstrictor use in, 266
  - ventricular tachycardia in, 1965
  - in women, 3164–3165
- Coronary artery spasm. *See* Vasospastic angina (Prinzmetal's variant angina)
- Coronary atherosclerosis. *See* Coronary artery disease (CAD)
- Coronary blood flow, 2091
- Coronary revascularization. *See* Coronary artery bypass grafting (CABG); Percutaneous coronary interventions (PCI)
- Coronaviridae, 1478f, 1479, 1480t, 1535
- Coronavirus infections
  - clinical features of, 253
  - COVID-19 disease. *See* COVID-19 disease
  - diarrhea in, 305
  - Middle East respiratory syndrome, 995, 1535
  - SARS. *See* Severe acute respiratory syndrome (SARS)
- Coronin A deficiency, S9
- Cor pulmonale
  - vs. constrictive pericarditis, 2084
  - in COPD, 2254
  - diagnosis of, 1871
  - in polycythemia, 451
  - right ventricular dysfunction in, 1883
- Corpus callosum, agenesis of, 3768
- Corpus luteum, 3129
- Corrigan's (water-hammer) pulse, 287, 1863, 2044
- Corrosive esophagitis, 2513
- Corticobasal degeneration, 3482, 3488f
- Corticobasal syndrome
  - clinical features of, 178, 196, 203, 3488, 3497
  - neuropathology of, 3487f, 3488, 3488f, 3497
- Corticospinal pathway, 169f
- Corticosteroid-binding globulin (CBG), 3128
- Corticosteroids. *See* Glucocorticoid(s)
- Corticotropin-releasing hormone, ectopic production of, 737t, 739
- Cortisol
  - actions of, 355, 2986, 3231f
  - circadian rhythm of, 3050, 3052f
  - deficiency of, 3234
  - inactivation of, 3054, 3054f
  - laboratory evaluation of, 3050–3052
  - synthesis of, 3051f
- Cortisol awakening response, 3950
- Cortisol-binding globulin (CBG), 3053–3054
- Cortisone acetate, 2996t
- *Corynebacterium* spp., 1225–1226
  - *C. afermentans,* 1226
  - *C. amycolatum,* 1226
  - *C. auris,* 1226
  - *C. diphtheriae,* 1221, 1224, S11. *See also* Diphtheria
  - *C. glucuronolyticum,* 1226
  - *C. imitans,* 1226
  - *C. jeikeium,* 1225–1226
  - *C. minutissimum,* 382, 1226
  - *C. pseudodiphtheriticum,* 1226
  - *C. pseudotuberculosis,* 1225
  - *C. striatum,* 1226
  - *C. ulcerans,* 1225
  - *C. urealyticum,* 1226
  - *C. xerosis,* 1226
- Cost-effectiveness, of screening tests, 40
- Co-stimulatory molecules, 2755, 2777
- Costochondritis (Tietze's syndrome), 103t, 104, 2919–2920, 2971
- Costoclavicular maneuver, 2176
- Cosyntropin test, 3051, 3068, 3091
- Cotransporters (symporters), 2366
- Cotton dust exposure, 2236t, 2238–2239
- Cottonmouth water moccasin bite, 3720t. *See also* Snakebites
- Cotton-wool spots, 225f, 3313
- Cough, **272**
  - assessment of, 272–273, 2199
  - chronic, 272–273
  - duration of, 272
  - etiology of, 272–273
  - in GERD, 301
  - global considerations in, 274
  - in helminthic infections, 1739t
  - high-altitude, 3758
  - impaired, 272, 272t
  - infectious droplets in, 964
  - in interstitial lung disease, 2261
  - in lung cancer, 618, 618t
  - mechanism of, 272, 272f
  - in *Nocardia* infections, 1359
  - nocturnal, 1982
  - with normal chest radiograph, 272–273
  - in pertussis, 1278–1279, 1278t. *See also* Pertussis
  - in respiratory disease, 2199
  - with swallowing, 295
  - symptomatic, 272
  - in TB, 1387
  - treatment of, 274
- Cough headache, primary, 3473
- Cough hypersensitivity syndrome, 272, 274
- Cough suppressants, 274
- Countercurrent in-line heat exchanger, 3770
- Countercurrent multiplication, 2369
- Courvoisier's sign, 673
- COVAX, 3866
- Cover test, 221, 234
- COVID-19 disease, **1539**. *See also* Long COVID syndrome
  - amylase levels in, 2740
  - in cancer patient, 579
  - in children, 141t, 402, 1540, 2907
  - clinical course of, 1541
  - clinical features of, 141t, 995, 1540
    - ADEM, 3594
    - AKI, 2360, 2376
    - ARDS, 1541, 2294, 2299
    - coagulation abnormalities, 940, 2156f
    - cutaneous, 141t, 143t, 146, 402
    - diarrhea, 1632
    - Guillain-Barré syndrome, 3606
    - hemophagocytic lymphohistiocytosis, 468
    - microvascular thrombosis, 2448
    - mild/asymptomatic, 141t
    - moderate/severe, 141t
    - mucormycosis, 1718
    - myelopoiesis, 768
    - myocarditis, 2012
    - myositis, 2917
    - olfactory dysfunction, 240–241
    - pneumonia, 3994
    - pneumonitis, 596, 596f
    - renal, 2429, 2437
    - respiratory, 252, 995, 1540
    - severe, 141t, 996
    - Tapia syndrome, 3559
    - thrombosis, 2154, 2156f, 2448
    - vascular damage, 927
  - complications of, 1541, 1543
  - diabetes mellitus and, 3203, 3230
  - diagnosis of, 988, 995, **1541**
    - chest radiograph, A12
    - peripheral blood smear, 453, 456f
  - differential diagnosis of, 1548, 2907
  - dysregulated immune response in, 973
  - epidemiology of, 141t, 1539–1540, 3850
  - etiology of, 141t, 1539. *See also* SARS-CoV-2
  - in HCT recipient, 1157
  - high-altitude travel following, 3760
  - hospitalization indications for, 1542
  - in IBD patients, 2570
  - mortality rates of, 995, 996, 1546, 1547f
  - in myasthenia gravis patients, 3633
  - in obese patients, 3187
  - in older adults, 1540, 3894
  - pathogenesis of, 2154, 2156f, 3942, 3994
  - PD and, 3501
  - in pregnancy, 3903
  - prevention of, 1489, 1540, 1543. *See also* COVID-19 vaccines
  - risk factors for severe illness in, 1540
  - sex differences in, 3162, 3167, 3167f, 3177
  - systems biology applied to, 3961t
  - transmission of, 995, 1148, 1490–1804
  - treatment of, 996, **1541**
    - antibody-based, 1542–1543
    - for complications, 1543
    - early observational reports of, 27
    - hyperbaric oxygen therapy for, 3764
    - inpatient management, 1542
    - molnupiravir, 1488t, 1489
    - nirmatrelvir/ritonavir, 1488t, 1489
    - outpatient management, 1542
    - remdesivir, 1116, 1488t, 1489, 1542
  - in vulnerable populations, 3850
- COVID-19 pandemic
  - behavioral decisions affecting, 3923
  - early case reports in, 27
  - fentanyl-associated deaths during, 3688
  - GAS infections rates and, 1207
  - genomics in response to, 988
  - global considerations in, 988, 995, 3850, 3854
  - health worker burnout during, 53
  - impact on cardiovascular disease, 1858
  - influenza circulation during, 1544
  - measles vaccine disruption during, 1639
  - poisoning cases during, 3716
  - resource allocation during, 71
  - vaccine misinformation and, 16
- COVID-19 vaccines
  - bivalent, 1543
  - development process for, 988, 1543
  - global access to, 3866
  - misinformation about, 7, 16
  - mRNA-based, 1543
  - myocarditis and, 2015
  - precautions for, 1007t

- COVID-19 vaccines (*Cont.*):
  - recommended administration of
    - in health care workers, 1153t
    - in international travelers, 1013–1014
    - in pregnancy, 1005, 3903
  - subunit protein approach to, 1543
  - thrombocytopenia following, 923
  - updates of, 1543
  - viral vectored, 1543
- Cowden's syndrome (multiple hamartoma syndrome)
  - breast cancer risk in, 633
  - clinical features of, 404, 524t, 3088
  - genetic considerations in, 404, 524t, 717t, 3080t, 3088, 3800
  - tumors associated with, 404, 717t
- Cowpox, 1518, 1518t, 1519, 1521
- COWS mnemonic, 191
- COX. *See* Cyclooxygenase (COX)
- *Coxiella burnetii,* 968, 973, 1038, 1040, 1462. *See also* Q fever
- COX inhibitors. *See* Cyclooxygenase (COX) inhibitors
- Coxsackievirus infections
  - acute hemorrhagic conjunctivitis, 1636, 1636f
  - clinical syndromes associated with, 1634t
  - diagnosis of, 1636
  - hand-foot-and mouth disease. *See* Hand-foot-and-mouth disease
  - herpangina, 262t, 1535, 1636
  - vs. herpes zoster, 1506
  - meningitis, 1124, 1634
  - myocarditis and pericarditis, 1635
  - oral manifestations of, 262t
  - pleurodynia (Bornholm disease), 1635
  - skin manifestations of, 137, 138t, 144t, 402, 1049, 1635, 1635f
- CPA (chronic pulmonary aspergillosis), 1390
- CPAP. *See* Continuous positive airway pressure (CPAP)
- *cpe* gene, 1240
- C peptide, 3198, 3235
- CPET (cardiopulmonary exercise testing), 271, 292
- CpG (cytosine-phosphate-guanine), 997t, 998
- CpG island methylation pathway (CIMP), 653
- CpG islands, 539, 3782, 3783
- CPHD (combined pituitary hormone deficiency), 3068t
- *CPOX* gene mutations, 3342t, 3348
- CPPD (calcium pyrophosphate deposition) disease, 2889, 2960–2961, 2960t
- CPT2 (carnitine palmitoyltransferase 2) deficiency, 3645
- *CR1* gene mutations, 3480
- Crab lice *(Pthirus pubis),* 3731–3732
- Crab yaws, 1438
- Crackles, 2200
- Cradle cap, 386
- Cramps, muscle, 3634
- Cranberry, 1095
- Cranial epidural abscess, 1141, 1141f
- Cranial nerve disorders
  - anatomic considerations, 3551, 3551f
  - in botulism, 1235, 1236
  - dysphagia and dysphonia in, 3556–3557
  - etiology of, 3558t
  - facial nerve, 3553–3555. *See also* Bell's palsy
  - glossopharyngeal neuralgia, 265, 3555–3556
  - in Guillain-Barré syndrome, 3559, 3617
  - injuries, 3573–3574
  - multiple, 3558–3559
  - neck weakness in, 3558
  - in subarachnoid hemorrhage, 3458
  - tongue paralysis in, 3558
  - trigeminal nerve, 242, 3551–3553, 3553t. *See also* Trigeminal neuralgia
- Cranial nerve examination, 3380
- Cranial radiation, 756
- Craniofacial synostosis (Crouzon's syndrome), 3787
- Craniopharyngiomas, 222, 232, 722, 3002, 3002f
- "Crazy paving," 2237, 2237f
- C-reactive protein (CRP)
  - in CKD, 2389
  - in community-acquired pneumonia, 1025–1026
  - in fever evaluation, 135
  - high-sensitivity, 1591t
  - in immune response to pneumococci, 1191
  - in infectious diseases, 960
  - in nutrition assessment, 2617, 2618t, 2621
  - in SLE, 2786f
  - in ulcerative colitis, 2557
- Creatine kinase (CK)
  - myocardial band (CK-MB)
    - in cardiac trauma, 2088
    - in STEMI, 2116, 2117f
  - serum
    - in distal myopathies, 3643
    - in glycogen storage disorders, 3644
    - in glycolytic defects, 3645
    - in inflammatory myopathies, 2910
    - in muscle disease, 3639
    - in muscular dystrophy, 3639, 3642
- Creatine phosphate, 1854
- Creatinine
  - serum, 340t
    - in AKI, 2378
    - in estimation of glomerular filtration rate, 2388t, S10
    - in hypovolemia, 347
    - in nutrition assessment, 2618t
    - as surrogate for glomerular filtration rate, 337
  - urine, 340t
- Creatinine clearance, 337, 488, 3899
- Creatinine height index, 2618t
- CREB (cAMP response-element binding protein), 3654, 3654f, 3689, 3689f, 3783t
- *CREB3L1* gene mutations, 3324, 3325t
- *CREBBP* gene mutations, 3941
- Cremasteric reflex, 3381
- Crepitant cellulitis, 1243
- Crepitus, 2942, 2942t
- CREST syndrome, 396, 416, 2859, 2945. *See also* Systemic sclerosis
- Cretinism, 261, 3022
- Creutzfeldt-Jakob disease (CJD), **3529**
  - after dura matter grafts, 3529
  - clinical features of, 194, 196t, 3482, 3525–3526, 3530
  - diagnosis of, 197, 199, 3530–3532, 3531f
  - differential diagnosis of, 3482–3483, 3532
  - familial, 3526t, 3528t
  - genetic considerations in, 195t, 3528
  - iatrogenic, 3526, 3526t, 3529
  - infection control in, 3532
  - molecular basis of, 195t
  - neuropathology of, 195t, 3400, 3401t, 3529–3530
  - prevention of, 3532–3533
  - as prion disease, 3526
  - treatment of, 3532
  - variant, 3526, 3526t, 3529
- Cricopharyngeal bar, 294
- Crigler-Najjar syndrome, **2639**
  - clinical features of, 323, 2634, 2639, 2639t
  - diagnosis of, 2639t
  - genetic considerations in, 2639
  - vs. Gilbert syndrome, 2639t
  - subtypes of, 323, 2639t
  - treatment of, 2640, 2718
- Crimean-Congo hemorrhagic fever, 993, 1661t, 1677
- Crisaborole, 383
- CRISPR/Cas, 74, 762, 776, 778, 979, 3798, 3938, 3938f
- CRISPRs, 984
- Cristoli virus, 1658t
- Critical care medicine, **2291**
- Critically ill patient
  - AKI in, 2298
  - anemia in, 768, 2298
  - approach to, **2291**
  - DVT in, 2297. *See also* Deep venous thrombosis (DVT)
  - enteral nutrition for, 2624
  - glycemic control in, 2297–2298
  - hypoglycemia in, 3234
  - illness severity assessment in, 2291–2292, 2292t, 2293f
  - monitoring of, 2296–2297, 2297f
  - multiorgan system failure in, 2296
  - neurologic disorders in, **2342,** 2343t. *See also* Coma
    - after solid organ transplant, 2350–2351
    - approach to the patient, **2343**
      - diagnostic studies, 2344–2345
      - increased ICP monitoring and treatment, 2345, 2345f, 2345t. *See also* Intracranial pressure (ICP), increased
      - secondary brain insults, 2345–2346
    - brain death, 2298–2299
    - coma. *See* Coma
    - delirium. *See* Delirium
    - hyperperfusion disorders, 2349–2350, 2350f, 2350t
    - hypoxic-ischemic encephalopathy, 2346–2348
    - metabolic encephalopathies, 2348
    - myopathy, 2352, 3647
    - neuromuscular transmission disorders, 2351–2352
    - neuropathy, 2351
    - osmotic demyelination syndrome (central pontine myelinolysis), 350, 2348–2349, 2349f
    - pathophysiology of, 2342–2343, 2344f
    - peripheral nervous system, 2343t, 2351–2352
    - polyneuropathy, 3606
    - postcardiac bypass, 2348
    - respiratory evaluation in, 2351
    - sepsis-associated encephalopathy, 2348
    - status epilepticus. *See* Status epilepticus
    - stroke. *See* Stroke
    - subarachnoid hemorrhage. *See* Subarachnoid hemorrhage
    - Wernicke's disease/encephalopathy. *See* Wernicke's disease/encephalopathy
  - nutritional status of, 2297–2298
  - parenteral nutrition for, 2625
  - prevention of complications in, 2297–2298
  - respiratory failure in, 2294–2295, 2295f. *See also* Acute respiratory distress syndrome (ARDS)
  - sepsis in, 2296
  - shock in, 2292–2294, 2294f. *See also* Shock
  - stress-related ulcers in, 2297, 2537
  - weakness in, 2298
  - withholding/withdrawing care from, 2299
- Critical values, in laboratory tests, S10
- Crizanlizumab, 775
- Crizotinib
  - actions and targets of, 527, 533t, 568t, 572, 2187f
  - adverse effects of, 568t, 753t, 755
  - for anaplastic large-cell lymphoma, 867

- genetic variations in response to, 493, E72t
  - interactions and issues, 568t
  - for lung cancer, 572, 628, 628t
- CRMP5 antibodies, 619, 743t, 744, 751
- CroFab (Crotalidae polyvalent immune fab), 3720t
- Crofelemer, 312
- Crohn's disease. *See also* Inflammatory bowel disease (IBD)
  - cancer risk in, 2570
  - classification of, 2558t
  - clinical features of
    - colitis, 2558
    - diarrhea, 309
    - extraintestinal, 309, 2539
    - gastroduodenal disease, 300, 2558
    - gastrointestinal bleeding, 318
    - ileocolitis, 2557–2558
    - jejunoileitis, 2558
    - nausea and vomiting, 297
    - oral, 261
    - perianal disease, 2558, 2584
  - colorectal cancer in, 2570
  - complications of, 2559
  - diagnosis of
    - endoscopic, 2470f, 2496, 2503f, 2558, 2559f
    - MRI in, 2559, 2559–2560f
    - radiology in, 2544f, 2558–2559
    - small-intestinal mucosal biopsies in, 2550t
  - differential diagnosis of, 1306, 1471, 2559–2561, 2560t
  - epidemiology of, 2551–2552, 2551t
  - etiology of, 2552f, 2553
  - familial patterns of, 2552
  - genetic considerations in, 2553–2554, 2553–2554t
  - immune regulation in, 2554–2555
  - inflammasome mutations in, 2761–2762t
  - inflammatory cascade in, 2555
  - metastatic, 2562t
  - pathogenesis of, 2552f, 2781, 2790t
  - pathology of, 2556, 2556f
  - in pregnancy, 2569–2570
  - racial and ethnic differences in, 2551
  - serologic markers in, 2559
  - smoking effects on, 2552
  - "string sign" in, 2558
  - treatment of, 2567, 2569, 2784. *See also* Inflammatory bowel disease (IBD), treatment of
  - vitamin K deficiency in, 2613
- Cromolyn sodium, 2225, 2810
- Cronkhite-Canada syndrome, 309
- *Cronobacter sakazakii,* 1292
- Cross-bridge cycling, in cardiac contraction, 1848f, 1849
- Cross-cultural education, 68
- Crossed straight leg raising, 120, 121t
- Crossing-over, 3782f, 3784–3785
- Cross-sensitivity, drug reactions, 427–428
- Cross-talk, nuclear-mitochondrial, 3814, 3814f
- Crotamiton, 3729
- Croup (laryngotracheobronchitis), 259, 1548
- Crouzon's syndrome (craniofacial synostosis), 3787
- Crowned dens syndrome, 132
- Crown, tooth, 260
- CRP. *See* C-reactive protein (CRP)
- CRPS (complex regional pain syndrome), 96, 2971
- CRS (cytokine release syndrome), 561, 579, 594–595, 754, 2351, 2788
- *CRTAP* gene mutations, 3324, 3325t, 3326
- Cruise ship passenger, 305
- Crust (skin lesion), 377t, 961, 1049f, 1050t
- Crusted (Norwegian) scabies, 1605, 3729
- *CRY1/2* gene mutations, 3947t, 3952
- Cryoablation, for ventricular tachycardia, 1959
- Cryoglobulin(s), 2905
- Cryoglobulinemia
  - in chronic HCV infection, 2685, 2693
  - clinical features of, 2356, 2415t
  - monoclonal, 407
  - vascular purpura in, 927
- Cryoglobulinemic vasculitis, **2905,** A14
- Cryohydrocytosis, 795
- Cryoprecipitate, 903t, 907t
  - for coagulation disorders in CKD, 2393
  - for coagulation disorders in liver disease, 935
  - for DIC, 934, 2322t
  - for hemophilia, 930
- Cryopyrin-associated periodic syndromes (CAPS), 155, 1131, 1132t, 2761t, 2936–2937
- Cryosurgery, 555, 607
- Cryotherapy, 390, 508, 1530, 1531t
- Cryothermy, 1917
- Cryptic miliary tuberculosis, 1390
- Cryptococcomas, 1705
- Cryptococcosis, **1703**
  - approach to the patient, 1704
  - chronic meningitis in, 1130t, 1131, 1134f
  - clinical features of, 235, 1688t, 1705, 1705f
  - complications of, 1706
  - diagnosis of, 961t, 1688t, 1705, S11
  - disseminated, 148, 148f, 1705, 1705f, A1
  - epidemiology of, 961t, 1131, 1704
  - in HIV infection, 1594t, 1606–1607, 1705–1706
  - immune reconstitution inflammatory syndrome and, 1706
  - pathogenesis of, 1704, 1704f
  - prevention of, 1706
  - prognosis of, 1706
  - risk factors for, 1688t
  - skin manifestations of, 1705, 1705f, A1
  - in solid organ transplant recipient, 1162
  - subacute meningitis, 1126
  - treatment of, 1127, 1705–1706
- *Cryptococcus* spp., 961t, 1703–1704
- Cryptogenic organizing pneumonia (COP), 2263f, 2264–2265, 2264f
- Cryptogenic sensory and sensorimotor polyneuropathy (CSPN), 3613
- Cryptorchidism, 704, 3100, 3114
- *Cryptosporidium*/cryptosporidiosis, **1805**
  - climate change and, 3752t
  - clinical features of, 305, 306t, 1737, 1739t, 1806, S7
  - diagnosis of, 1805t, 1806, S7, S12
  - differential diagnosis of, 306t
  - epidemiology of, 1737, 1739t, 1806, S12
  - in HIV infection, 1599
  - incubation period for, S7
  - life cycle of, 1806, S12
  - pathophysiology of, 1806
  - in travelers, 1079t
  - treatment of, 1750, 1806
  - in war veterans, S7
- Crystal(s)
  - arthritis induced by. *See* Arthritis, crystal-induced
  - renal tubular obstruction by, 2440
  - urinary, 2450, 2451f. *See also* Nephrolithiasis
- Crystallopathies, 2755, 2760
- Crystalluria, in ethylene glycol poisoning, S1
- *CSBT* gene mutations, 3515
- Csendes' procedure, 2533
- Csf1r, 3394
- *CSF3R* gene mutations, 875
- CSF (cerebrospinal fluid) analysis. *See* Cerebrospinal fluid (CSF) analysis
- CSF (cerebrospinal fluid) leakage, 116, 116f, 3483, 3572
- CSPN (cryptogenic sensory and sensorimotor polyneuropathy), 3613
- *CSTB* gene mutations, 3412t
- CT. *See* Computed tomography (CT)
- CTA. *See* Computed tomography (CT) angiography (CTA)
- *CTC1* gene mutations, 3820t
- CTCL (cutaneous T cell lymphoma). *See* Mycosis fungoides
- ctDNA. *See* Circulating tumor DNA (ctDNA)
- CTE (chronic traumatic encephalopathy), 194
- *Ctenocephalides felis,* 1458
- CTEPH. *See* Chronic thromboembolic pulmonary hypertension (CTEPH)
- CTLA-4, 550f, 560, 2785–2786, 2786f, 2804
- CTLA-4 antibodies. *See also specific drugs*
  - action and targets of, 534t, 549–550, 550f, 560, 2783, 2786f, 2787–2788, 2787t
  - adverse effects of, 3092
    - endocrinopathy, 740, 756
    - hepatic, 2672
    - immune-related, 605, 2561
    - myocarditis, 2015
    - myositis, 2917
    - toxic erythema, 421–422
  - indications for, 2787t
  - microbiome and effectiveness of, 3837
- *CTLA4* gene mutations, 2553, 3942
- *CTLRN* gene mutations, 3375
- *CTNNB1* gene mutations
  - cancers associated with, 520t
  - in HCC, 660, 661t
  - in hepatoblastoma, 670
  - in hepatocellular adenoma, 660, 670
  - in thyroid cancer, 3046
- *CTNS* gene mutations, 3375
- *CTRC* gene mutations, 2744
- C-type lectin receptors (CLRs), 2317, 2758, 2759t
- CTZ (chemoreceptor trigger zone), 82
- Cubilin, 781
- Cubital tunnel syndrome, 3613
- CUL3, 359
- *Culex* mosquitoes, 1657–1663t, 1669–1670, 1671–1672, 1820t
- *Culicoides paraensis,* 1667
- *Culiseta* mosquitoes, 1670
- Cullen's sign, 807, 2745
- Cultural competence, 69
- Cultural safety, 19
- "Cupping," in glaucoma, 230, 230f
- CURB-65 criteria, for pneumonia, 1026
- Curbside consults, 3897–3898
- Curcumin, 2956
- Curium exposure, S6
- Curling's ulcer, 2537
- Cushing's disease, **3010**
  - clinical features of, 3011, 3011t
  - vs. Cushing's syndrome, 3054
  - definition of, 3054
  - diagnosis of, 3011–3012
  - etiology of, 3010–3011
  - pathogenesis of, 2977
  - prevalence of, 3010–3011
  - screening tests for, 3001t
  - treatment of, 3012, 3012f
- Cushing's syndrome, **3054**
  - in Carney complex, 3088
  - caused by ectopic ACTH production. *See* Adrenocorticotropic hormone (ACTH), ectopic production of
  - clinical features of, 3011t, 3056–3057, 3056t
    - bleeding, 472
    - cutaneous, 397, 407, 3057f, A15
    - dyslipidemia in, 3247
    - hirsutism, 3138
    - hypertension, 2140

Cushing's syndrome, clinical features of (*Cont.*):
    hypokalemia, 355, S1
    metabolic alkalosis, 375
    obesity/abdominal fat, 326, 3057f, 3185, A15
    vessel wall disorders, 927
  vs. Cushing's disease, 3054
  diagnosis of, 3057, 3058f, 3059f, 3185, 3954
  differential diagnosis of, 3011t, 3057–3058, 3058f, 3059f
  epidemiology of, 3054
  etiology of, 3054–3056, 3055t, 3060t
  genetic considerations in, 3060t
  glucocorticoid withdrawal and, 567
  hormonal rhythms in, 2987
  in lung cancer, 619
  periodic, 3088
  treatment of, 3058f, 3059–3060, S1
Cushing's ulcer, 2537
Cutaneous anthrax, 145t, 149, 1051, S4
Cutaneous arteritis, 2906
Cutaneous CD30+ T cell lymphoma, 875
Cutaneous diphtheria, 1222, 1223, 1223f, 1224
Cutaneous drug reactions, **416**
  exanthematous eruption, A1
  genetic considerations in, 419
  global considerations in, 419–450
  immune
    DIHS. *See* Drug-induced hypersensitivity syndrome (DIHS)
    fixed drug eruption, 423, 423f
    irritant/allergic contact dermatitis, 423, 423f
    morbilliform/maculopapular drug eruption, 138t, 145, 401–402, 422, 422f
    overlap hypersensitivity syndromes, 426
    pathogenesis of, 417–419, 417t, 418f, 489–490
    pruritus, 422
    pustular eruptions, 425–426. *See also* Acute generalized eruptive/exanthematous pustulosis (AGEP)
    Stevens-Johnson syndrome. *See* Stevens-Johnson syndrome
    toxic epidermal necrolysis. *See* Toxic epidermal necrolysis
    urticaria/angioedema, 422
    vasculitis, 426, 426f
    vesicles/bullae, 400–401
  incidence of, 417f
  management of, 426
    confirmation of drug reaction, 426
    cross-sensitivity, 427–428
    drug desensitization for. *See* Drug desensitization
    drugs to suspect and withdraw, 427
    early diagnosis in severe reactions, 426, 426t
    prevention of recurrence, 427
    reporting, 429
    testing and drug rechallenge, 428–429
    "treating through" an eruption, 427
  nonimmune, 417
    exacerbation or induction of dermatologic disease, 420
    hair disorders, 421
    nail disorders, 421
    photosensitivity eruptions, 420, 434, 434t
    pigmentation changes, 399, 420
    toxic erythema, 421–422, 421f
    warfarin necrosis, 420–421, 421f
  timeline of, 427, 428f
  vasculitis, 2907
Cutaneous emboli, 401
Cutaneous larva migrans. *See* Larva migrans, cutaneous
Cutaneous leishmaniasis
  clinical features of, 1738t, 1786, 1786f, S7
  diagnosis of, 1738t, 1786–1787, S7, S12
  differential diagnosis of, 1786
  epidemiology of, 1782t, 1783f, 1786, S7
  etiology of, 1786
  immunopathogenesis of, 1786
  incubation period for, S7
  prevention of, S7
  treatment of, 1741t, 1787, S7
  in war veterans, S7
Cutaneous leukocytoclastic angiitis, 2905
Cutaneous lymphoid hyperplasia (pseudolymphoma), 393, 405
Cutaneous mastocytosis, 396, 402, 406, 880, 2820–2821
Cutaneous reflexes, 168, 3381
Cutaneous sarcoidosis, 2927, 2927f. *See also* Sarcoidosis
Cutaneous small-vessel vasculitis. *See* Leukocytoclastic (cutaneous small-vessel) vasculitis
Cutaneous T cell lymphoma. *See* Mycosis fungoides
Cutaneous ulcers, **408**
  definition of, 377t
  description of, 137
  etiology of, 408t, 410t, 1050t, 1051
  with fever, 145t, 148–149
  hyperbaric oxygen therapy for, 3763–3764, 3763f
  in parasitic infections, 1738t
  venous, 2180, 2180f
Cutaneous vasculitis. *See* Vasculitis
Cutavirus, 1525
Cutibacterium acnes, 390, 1061, 1065t
Cutis verticis gyrata, 2970
CUT&RUN, 3937
CVID (common variable immunodeficiency), 858, 858t, 2802
CVP (central venous pressure), 1860–1861, 2295
CVS (chorionic villus sampling), 3797
c-v wave, 286
CWD (chronic wasting disease), 3526, 3526t
CX$_3$CR1, 2768t, 2769t
CX3CR1, 1581t
CXC chemokine family, 456, 464, 2768t, 2769t, 2773
CXCL4, 2864
CXCL12, 759, 760, 914, 1582t
CXCR1, 2769t
CXCR2, 2768t, 2769t
CXCR3-A, 2768t, 2769t
CXCR3-B, 2768t, 2769t
CXCR4 gene mutations, 458, 893, 894, 2795, 2802
CXCR4 receptor
  in hematopoietic stem cells, 759, 914
  in HIV infection, 1557–1558, 1577, 1578f
  ligands, cell types, and disease connection, 2769t
  source, target, and biologic activity, 2768t
CXCR5, 2769t
CXCR6, 2769t
CXCR6 gene, 1581t
Cyanide poisoning
  in chemical terrorism, S5
  clinical features of, 50t, 3713t, S5
  complications of, 2346, 2348
  differential diagnosis of, S5
  mechanisms of, 277, 3713t, S5
  prognosis of, S5
  treatment of, 3713t, S5
Cyanobacteria, 3752t
Cyanocobalamin, for cobalamin deficiency, 789, 3612
Cyanosis, **278**
  approach to the patient, 279
  central, 278–279, 278t, 1859
  in coma, 189
  in congenital heart disease, 451
  degree of, 278
  differential diagnosis of, 278–279
  etiology of, 278t
  in hemoglobinopathy, 780
  pathophysiology of, 278
  peripheral (acrocyanosis), 278t, 279, 1859
CYB5A gene mutations, 3099t
CYB5R3 deficiency, 780
CYBC1, 460
Cyclic adenosine 5'-monophosphate (cyclic AMP), 135, 135f, 974, 3449, 3689, 3689f
Cyclic guanosine monophosphate (cGMP), 1981, 2194, 3155, 3156f
Cyclic neutropenia, 136, 261, 457
Cyclic vomiting syndrome, 298, 300
Cyclin B1, 2790t
Cyclin-dependent kinase(s), 573
Cyclin-dependent kinase inhibitors. *See* CDK inhibitors
Cyclobenzaprine, 3713t
  for fibromyalgia, 2965, 2966t
  for neck pain, 132
  for spasticity, 3588
Cyclobutane dimers, 429–430
Cyclomodulins, 973
Cyclooxygenase (COX), 136, 2517, 2518f
Cyclooxygenase (COX) inhibitors, 96–97, 136, 943f. *See also* Aspirin
Cyclooxygenase 2 (COX-2) inhibitors. *See also* Nonsteroidal anti-inflammatory drugs (NSAIDs)
  adverse effects of, 97–98, 359, 2517
  for colorectal cancer prevention, 513
  for fever, 136
  in NSAID-induced ulceration prevention, 2530
  for osteoarthritis, 2955, 2955t
  for pain, 96–98
  in peptic ulcer disease, 317
Cyclooxygenase 3 (COX-3), 136
Cyclopentolate hydrochloride, 222
Cyclophosphamide
  actions of, 563
  adverse effects of, 562t, 563, 2873, 2916t
    bladder cancer, 2895
    cardiovascular, 754, 2017
    cutaneous, 400, 420
    cystitis, 563, 597, 757, 916
    hepatotoxicity, 2667
    immune system dysfunction, 755
    infertility, 756, 2418, 3115
    leukopenia, 2899
    long-term, 753t
    major, 2896t
    male infertility, 3115
    nausea and vomiting, 575
    pulmonary, 754
  for AL amyloidosis, 898
  for autoimmune hemolytic anemia, 802
  for Behçet syndrome, 2909
  for breast cancer, 637, 638
  for Burkitt's lymphoma, 862
  for cardiac amyloidosis, 2027
  for CLL, 855
  for DLCBL, 863
  for EGPA, 2233, 2900
  for Ewing's sarcoma, 730
  for follicular lymphoma, 864
  for gastric lymphoma, 650
  for gestational trophoblastic disease, 715
  for granulomatosis with polyangiitis, 2899
  for inflammatory myopathies, 2916t
  interactions and issues, 562t
  for lupus nephritis, 2418, 2834–2835
  for lymphoplasmacytic lymphoma, 865
  for metastatic breast cancer, 641

  - for metastatic pheochromocytoma, 3077
  - for mucous membrane pemphigoid, 413
  - for multiple myeloma, 890t
  - for myasthenia gravis, 3631
  - for pemphigus vulgaris, 410
  - for photoallergy, 434
  - for systemic sclerosis, 2873, 2874
  - for thymic tumors, 632
  - for vasculitis, 2895
  - for Waldenström's macroglobulinemia, 894
- *Cyclops* spp., 1813, 1825, 1839
- Cyclosarin (GF), S5
- Cycloserine, 1394t, 1399, 1425, 2609
- *Cyclospora cayetanensis*/cyclosporiasis, 305, 1805t, **1806,** S12
- Cyclosporin A, 466, 469
- Cyclosporine
  - actions of, 2408, 2408t, 2565, 2721
  - adverse effects of, 387t, 813, 2408t, 2565, 2721, 2916t
    - carcinogenicity, 512t
    - dyslipidemia, 3247
    - gingival overgrowth, 261
    - gout, 2940t
    - hirsutism, 3138
    - hyperkalemia, 359
    - hypertrichosis, 421
    - hypoperfusion syndrome, 2350, 2350f, 2350t
    - microangiopathic hemolytic anemia, 925
    - myopathy, 2940t
    - nephrotoxicity, 2408, 2443, 2721
    - osteoporosis, 3296
    - seizures, 2565, 3414t
    - skin cancer, 433
    - thrombotic microangiopathy, 2445
  - for aplastic anemia, 813
  - for atopic dermatitis, 383, 393t
  - for Behçet syndrome, 2909
  - contraindications to, 387
  - for DIHS, 424
  - drug interactions of, 485t, 489t, 2408, 3927
    - azithromycin, 1744t
    - azoles, 1742t
    - chloroquine, 1741t
  - for epidermolysis bullosa acquisita, 413
  - for GVHD prophylaxis, 908
  - for IBD, 2565
  - for IC/BPS, 335–336
  - for immunosuppression
    - after cardiac transplantation, 2029t
    - after kidney transplantation, 2408, 2408t
    - after liver transplantation, 2721
  - for inflammatory myopathies, 2916t
  - for MDS, 817–818
  - metabolism of, 485t
  - for myasthenia gravis, 3631
  - P-glycoprotein and, 485t
  - for photoallergy, 434
  - for psoriasis, 387, 387t, 393t
  - for SJS/TEN, 425
  - for urticaria, 2809
- Cyclothymic disorder, 3668
- CYP1A1, 615
- *CYP2B6,* 1582t, 1585
- CYP2C9
  - in drug metabolism pathways, 485t
  - drug responses and, 491, 493, E72t
  - frequency of variants in, E72t
  - warfarin dose requirements and, 951, 951t
- CYP2C19
  - in drug metabolism pathways, 485t
  - drug responses and, 492–493, 939, 944, E72t
  - frequency of variants in, E72t
- CYP2D6
  - in autoimmune hepatitis, 2790t
  - in drug metabolism pathways, 492f
  - drug responses and, 485t, 492, E72t
  - frequency of variants in, E72t
- CYP3A, 485t
- CYP3A4, 492, 939, 1167, 2688, 3271
- CYP3A5, 492, E72t
- *CYP7A1* gene mutations, 2728
- CYP11A1, 3052
- *CYP11A1* gene mutations, 3098, 3099t
- *CYP11B1* gene mutations, 355, 3101, 3101t, 3784
- *CYP11B2* gene mutations, 355, 3784
- CYP17 (17α-hydroxylase), 2790t
- *CYP17* (17α-hydroxylase) gene mutations, 3052, 3098, 3099t
- *CYP19* gene mutations, 3101, 3101t, 3115
- CYP21 (21-Hydroxylase), 2790t
- *CYP21A1P* pseudogene, 3785
- *CYP21A2* (21-hydroxylase gene) mutations, 3071, 3071t, 3101, 3101t, 3785
- CYPIIa (cholesterol side-chain cleavage enzyme), 2790t
- Cyproheptadine
  - for antidepressant side effect management of, 3660t
  - for Cushing's disease, 3012
  - for ergot alkaloid overdosage/poisoning, 3710t
  - for migraine prevention, 3469
  - for serotonin syndrome, 3715t
  - for urticaria, 2809
- Cyproterone acetate, 3141
- CysLT (cysteinyl leukotrienes), 2218, 2805
- CysLT$_1$ receptor antagonists. *See* Montelukast; Zafirlukast
- Cyst, 378t. *See also specific types and sites*
- Cystathionine metabolism disorders, 3369t
- Cystathionine synthase, 783f, 784
- Cystathioninuria, 3369t
- Cystatin B, 3412t
- Cystatin C, 339
- Cystectomy, 336, 693–694, 693t
- Cysteine metabolism disorders, 3369t
- Cystic duct stump syndrome, 2734
- Cysticercosis
  - chronic meningitis in, 1131t
  - clinical features of, 225, 1051, 1833
  - diagnosis of, 1833–1834, S12
  - etiology of, 1138
  - neurologic. *See* Neurocysticercosis
  - treatment of, 1834
- Cystic fibrosis (CF), **2245**
  - clinical features of
    - bronchiectasis, 2242, 2242t, A12
    - clubbing, 2970
    - exercise intolerance, 291
    - gastrointestinal, 309, 2245
    - hemoptysis, 275
    - hepatic, 2245, 2713
    - other organ systems, 2245
    - pancreatic, 2245, 2751
    - respiratory, 2245
  - CT in, A12
  - diagnosis of, 2215, 2246–2247
  - disease spectrum of, 2247
  - genetic considerations in, 2245–2246, 2246f, 3782f, 3787, S10
  - genetic testing for, 3796, 3804t, S10
  - global considerations in, 2248
  - infections in
    - *Aspergillus,* 2233, 2248
    - *B. cepacia,* 1313, 2248, 2284
    - nontuberculous mycobacterial, 1415
    - *P. aeruginosa,* 1307, 1310, 2245, 2248
  - newborn screening for, 2246–2247
  - pathogenesis of
    - airflow resistance, 2249
    - autoantigens, 2790t
    - CFTR deficiency, 2245–2246, 2246f, 2249v
    - immunodeficiency, S9
    - protein misfolding, 3993
    - pulmonary inflammation and remodeling, 2245
  - treatment of
    - *CFTR* modulation, 2247
    - chronic management, 2247–2248
    - early interventions for, 3804t
    - lung transplantation, 2248, 2283
    - personalized molecular therapies, 2247
    - pulmonary exacerbation, 2248
    - quality improvement in, 2248
  - vitamin E deficiency in, 2613
- Cystic fibrosis transmembrane conductance regulator. *See* CFTR protein
- Cystine stones, 2448, 2451f, 2453. *See also* Cystinuria; Nephrolithiasis
- Cystinosis, 3374t, **3375**
- Cystinuria, 2368t, 2369, 2450, **3373,** 3374t
- Cystitis. *See also* Urinary tract infections (UTIs)
  - alkaline-encrusted, 1226
  - chemotherapy-induced, 563
  - clinical features of, 1089
  - definition of, 1086
  - diagnosis of, 1089, 1091–1092, 1091f
  - *E. coli,* 1286
  - epidemiology of, 1087
  - hematuria in, 343
  - hemorrhagic, 597, 916
  - in men, 1089, 1092, 1093
  - prognosis of, 1095
  - recurrent, 1095, 1180t
  - sexually transmitted, 1097t
  - in solid organ transplant recipient, 1159t
  - treatment of, 1092–1094, 1093t
- *Cystoisospora belli*/cystoisosporiasis
  - clinical features of, 305, 1739t, 1806, 2560
  - diagnosis of, 961t, 1805t, 1806, S12
  - epidemiology of, 961t
  - in HIV infection, 1600, 1806, 2560
  - life-cycle hosts of, S12
  - treatment of, 1806
- Cystopathy, diabetic, 3227
- Cystoscopy, 334, 704
- Cytarabine (cytosine arabinoside)
  - actions of, 565
  - adverse effects of, 562t, 565
    - hematologic, 785, 790
    - nausea and vomiting, 575
    - neurologic, 726t, 2350t, 3608t
    - pulmonary, 595
    - toxic erythema, 421
  - for AML, 830–831
  - for CNS leukemia, 847
  - for DLBCL, 863
  - interactions and issues, 562t
  - for mantle cell lymphoma, 866
  - for neoplastic meningitis, 591
  - for PCNSL, 721
- Cytidine deaminase, 574, 2786f
- Cytoadherence, 1762
- Cytochrome b5 (CYB5A), 3099
- Cytochrome b5 reductase deficiency, 796t
- Cytochrome *c* oxidase, 539
- Cytochrome *c* oxidase deficiency, 3810, 3810f
- Cytochrome oxidase inhibitors, 3713t
- Cytochrome P450 monooxygenase, 484
- Cytogenetic analysis, 3789t, 3797
- Cytogenomic analysis, 3789t. *See also* Single-nucleotide polymorphisms (SNPs)
- Cytokeratins (CKs)
  - in carcinoma of unknown primary, 733, 733t
  - in lung cancer, 616

- Cytokine(s), **2771**
  - in asthma, 2218, 2219f
  - in autoimmune and inflammatory diseases, 2784, 2790t
  - in autoimmune hypothyroidism, 3029
  - biologic activity of, 2765–2768t
  - in cancer treatment, 561
  - definition of, 2755, 2784
  - in diabetes mellitus, 3200
  - expression of, 2755
  - of innate immune system, 2758t, 2762
  - in nutrition assessment, 2618t
  - production in CNS, 135
  - proinflammatory
    - in anemia, 762, 763f, 768f
    - in atherosclerosis, A10
    - in autoimmune disease, 2826
    - in HIV infection, 1576–1577
    - in host response to pathogens, 2780
    - in IBS, 2572–2573
    - in metabolic syndrome, 3255
    - sepsis/septic shock, 2318f
    - in sepsis/septic shock, 2317
  - pyrogenic, 134–135, 135f
  - signaling deficiency, in SCID, 2798
  - sources and targets of, 2765–2768t
  - in spondyloarthritis, 2881f
  - in thyroid-associated ophthalmopathy, 3033
  - in thyroid hormone synthesis, 3023
- Cytokine inhibitors, 168, 2784. *See also* Tumor necrosis factor-α (TNF-α) inhibitors
- Cytokine receptor–linked kinase, 2981t
- Cytokine receptors, 2765–2768t, 2773, 2983, 2983f. *See also specific receptors*
- Cytokine release syndrome (CRS), 561, 579, 594–595, 754, 2351, 2788
- Cytokine storm, 974, 1198, 1545
- Cytolysins, 974
- Cytolytic T cells. *See* Natural killer T (NK-T) cells
- Cytomegalovirus (CMV), 1512
- Cytomegalovirus (CMV) immune globulin, 1515
- Cytomegalovirus (CMV) infections, **1512**
  - congenital, 1510, 1513t, 3903
  - diagnosis of, 1515, S11
  - diarrhea in, 305
  - differential diagnosis of, 1513t
  - epidemiology of, 1512–1513
  - esophagitis, 2495f, 2513
  - in HCT recipient, 918, 918t, 1155t, 1156, 1160, 1513t
  - in hemophagocytic lymphohistiocytosis, 468
  - in HIV infection, 1513t, 1514, 1595f, 1600, 1610
  - in liver allograft, A13
  - lymphadenopathy in, 477
  - mononucleosis, 1513–1514
  - nausea and vomiting in, 297
  - neuropathy in, 3607
  - pathogenesis of, 1513
  - pathology of, 1513
  - perinatal, 1513
  - pneumonia, 1533, 1537
  - in pregnancy, 1513, 3903
  - prevention of, 1515
  - proctitis, vs. IBD, 2560
  - prophylaxis for, 918, 1160, 1486t, 1487, 1594t
  - reactivation in DIHS, 424
  - retinitis, 1485, 1513f, 1514, 1515, 1610
  - skin manifestations of, 402
  - in transplant recipient
    - after HCT, 1513t, 1514
    - after solid organ transplant, 1158, 1160–1161, 1513t, 1514
    - brain, 2350
    - prophylaxis for, 1513t, 2410
    - treatment of, 1513t, 2410
  - treatment of, 1485–1487, 1486t, 1515–1516, 1537, 3566
  - vascular composite allografts and, 1158
- Cytopathology, S11
- Cytoplasmic antineutrophil cytoplasmic antibody (c-ANCA), 2894, 3624
- Cytosine, 3778, 3784f
- Cytosine arabinoside. *See* Cytarabine (cytosine arabinoside)
- Cytosine-phosphate-guanine (CpG), 997t, 998
- Cytoskeletal proteins, 2004t, 2005f, 2008
- Cytotoxic T cells. *See* CD8+ T cells
- Cytotoxins, bacterial, 974

D

- d4T (stavudine), 1616, 1619f, 2672
- D9THC, 3686–3687. *See also* Cannabis/marijuana use
- Dabigatran
  - adverse effects of, 953, 1951
  - dosage of, 952–953, S13
  - drug interactions of, 489t
  - indications for, 952
  - management before endoscopic procedures, 2483t
  - management before lumbar puncture, S3
  - monitoring treatment with, 475, 953
  - for NSTE-ACS, 2110t
  - pharmacology of, 953t
  - in pregnancy, 954
  - reversal of, 953–954, 2163, 3456
  - for stroke prevention, 1951, 1951t, 3450
  - for VTE/PE treatment, 2162, 2163t
  - for VTE prophylaxis, 2164t
- Dabrafenib
  - actions and targets of, 533t, 569t, 572
  - adverse effects of, 569t, 605
  - interactions and issues, 569t
  - for lung cancer, 628
  - for melanoma, 604, 604t, 605
- Dacarbazine (DTIC)
  - actions of, 563
  - adverse effects of, 434t, 562t, 563, 575, 726t
  - interactions and issues, 562t
  - for melanoma, 605
  - for metastatic pheochromocytoma, 3077
  - for soft tissue sarcomas, 729
- Daclatasvir, 1492t, 1493, 2689
- Daclizumab, 2791t
- Dacomitinib, 533t, 568t, 572, 627, 627t
- Dacryocystitis, 224
- Dacryocytes, 437
- Dactinomycin, 562t, 564, 715
- Dactylitis
  - characteristics of, 1049, 2943
  - etiology of, 1049, 2943
  - in psoriatic arthritis, 2887
  - in sickle cell disease, 2968
- DADs (delayed afterdepolarizations), 1913, 1913f, 1913t
- DAEC (diffusely adherent *E. coli*), 1288
- Dakar bat virus, 1659t
- Dakin's solution, S5
- Dalbavancin, 1045
  - actions of, 1166, 1182t, 1204
  - adverse effects of, 1171t, 1176
  - for enterococcal infections, 1219–1220
  - indications for, 1173t, 1176
  - in pregnancy and lactation, 1169t
  - resistance to, 1173t, 1182t, 1184
  - for staphylococcal infections, 1203t, 1204
- Dalfampridine, 3588
- Dalfopristin. *See* Quinupristin-dalfopristin
- Dalteparin, 2163t, 2164t
- DALY (disability-adjusted life years), 3844, 3845f, 3846t
- Damage-associated molecular patterns (DAMPs), 2317, 2318f, 2950f
- Danazol, 923t, 1556, 2721
- Dandruff, 1723
- Danicopan, 805f, 806
- Danon disease (LAMP2 deficiency), 2004t, 2010, 3354t, 3361t, 3365
- Dantrolene, 3570, 3588
- Dapagliflozin, 1988, 1995t, 1996, 3211t, 3213
- *Daphnia,* 1052
- Dapsone
  - for acne vulgaris, 391
  - actions of, 1411
  - adverse effects of
    - DIHS, 424, 1411. *See also* Drug-induced hypersensitivity syndrome (DIHS)
    - frequent, 1743t
    - in G6PD deficiency, 799t, 800, 1605
    - hemolysis, 413
    - methemoglobinemia, 413, 780
    - neuropathy, 3609t
    - occasional, 1743t
  - for dermatitis herpetiformis, 412–413
  - drug interactions of, 1743t
  - for epidermolysis bullosa acquisita, 413
  - for leprosy, 1410, 1411t
  - for linear IgA disease, 413
  - for mucous membrane pemphigoid, 413
  - overdosage/poisoning with, 3714t
  - for PCP prophylaxis, 1593t, 1734t, 1735
  - in pregnancy and lactation, 1743t
  - for toxoplasmosis prophylaxis, 1593t
  - for urticaria, 2809
- Daptomycin
  - actions of, 1167, 1182t, 1183f, 1204
  - adverse effects of, 1171t, 1176, 1204, 1220
  - for enterococcal infections, 1042, 1219–1220, 1220t
  - indications for, 1173t, 1176
  - for infective endocarditis, 1042
  - for MRSA infections, 1203t, 1204
  - for osteomyelitis, 1064t
  - for peritonitis, 1070, 1072
  - resistance to, 1173t, 1182t, 1185, 1204, 1221
  - for staphylococcal infections, 1203t, 1204
- Daratumumab, 534t, 559f, 890t, 891, 892, 898, 2027
- *DARC* gene, 1581t
- Dar es Salaam virus, 1662t
- Daridorexant, 216
- Darier's sign, 406, 2820
- Darifenacin, 3549
- Darmbrand (enteritis necroticans), 1241
- Darolutamide, 701–702, 702f, 753t
- Dartmouth Atlas of Health Care, 44
- Darunavir, 1606, 1618t, 1619f, 1621, 1621t, 1623f
- DAS181, 1489
- Dasabuvir, 1494, 2689
- Dasatinib
  - actions and targets of, 533t, 535, 567, 568t, 838
  - adverse effects of, 568t, 753t, 754, 839t, 840–841
  - for ALL, 848, 848t
  - for CML, 533t, 535, 838–841, 839t
  - genetic variations in response to, 567, E72t
  - in imatinib resistance, 841
  - interactions and issues, 568t
- DASH diet, 2141
- Data analysis, in precision medicine, 36, 36f
- Data Safety and Monitoring Committee (DSMC), 999
- Data sharing, 3998
- Daunorubicin
  - actions of, 564
  - adverse effects of, 394, 420, 562t, 564
  - for AML, 830

  for APL, 833
  interactions and issues, 562t
  liposomal, for Kaposi's sarcoma, 1613
David procedure, 2047, 2047f
"Dawn phenomenon," 3950, 3954
Dawson fingers, 3593
DAX-1, 2988, 2988t, 3050
DAX1 gene mutations, 3113, 3786
DAX1/NR0B1 gene mutations, 3094
Day-care centers, 305
Daytime sleepiness, 212–213, 213t, 2281
DAZ gene mutations, 3789
DBS. *See* Deep brain stimulation (DBS)
DBT (dialectical behavior therapy), 3672
DC-CK1/PARC, 2767t
DCP (des-γ-carboxy prothrombin), 661
DC-SIGN, 1558
DDAVP. *See* Desmopressin (DDAVP)
ddC (zalcitabine), 1619f
ddI (didanosine), 1619f, 2672
ᴅ-dimer test
  in DIC, 934, 934t
  in HIV infection, 1591t
  in VTE/PE, 2159
DDR2 gene mutations, 617
DDX41 gene mutations, 825, 825t
Deafness. *See* Hearing loss
Death
  from cancer, 497t, 498, 499f, 500f, 500t
  in cancer patient, 505
  cardinal signs of, 92
  childhood, 3844
  infectious disease–related, 957, 958f
  leading causes of, 3162–3163
    age-specific, 41t
    cardiovascular disease, 1854–1856, 1856f, 1857f
    epidemiologic transition in, 1854–1856
    in men, 3163f, 3168
    in United States and Britain, 75t
    in women, 3163f
    worldwide, 1856f, 3844, 3846t
  site of, 75–76, 75f
  of spouse, 92
  sudden cardiac. *See* Sudden cardiac death
Death adders, 3717
Death rattle, 91, 91t
Death receptors, 539, 540f, 2781
deaths from, 75t
Debranching enzyme deficiency (type IIIa/b GSD), **3359,** 3360t, 3645
DEB-TACE (drug-eluting bead chemoembolization), 666
DEC. *See* Diethylcarbamazine (DEC)
Deception, 72
Decerebrate rigidity, 189
Decision analysis, 26–27
Decision limits, S10
Decision-making. *See* Clinical reasoning
Decision-making, in older adults, 3877, 3880f
Decision support tools, 26
Decitabine
  actions of, 570t, 573, 3940, 3942
  adverse effects of, 570t, 817
  for AML, 830
  interactions and issues, 570t
  for MDS, 817
  for primary myelofibrosis, 823
Decitabine-cedazuridine, 817
Decompression sickness, 3766–3768, 3767t
Deconditioning, 269t, 270
Decongestants, 253, 2810, 3705, 3711t
Decorin, 3322t
Decorporation, S6
Decorticate rigidity, 189
Decremental conduction, 1925
Dectin-1, 1708
Decubitus (pressure) ulcers, 1051, 3570
Deep brain stimulation (DBS)
  for depression, 3667, 3967
  for dystonia, 3512
  for headache, 3472
  for obsessive-compulsive disorder, 3664
  for PD, 3505, 3967
  for seizures, 3425
Deep gluteal syndrome, 121t
Deep learning framework, 3973t, 3974
Deep venous thrombosis (DVT), **2154**
  asymptomatic, 2162
  classification of, 2158
  in critically ill patient, 2297
  diagnosis of
    bedside rules for, 49t
    clinical decision rules, 2159, 2159f, 2159t
    integrated approach, 2162f
    invasive imaging modalities, 2161
    nonimaging modalities, 472, 2159–2160, 2160f
    noninvasive imaging modalities, 2160–2161, 2160f, 2161f
    physical examination in, 3912t
    POCUS, 3912t, 3913, 3914v
  differential diagnosis of, 2159, 2159t
  embolization of, 2158, 2158f
  epidemiology of, 939–940
  history in, 473
  in HIV infection, 1605
  lower extremity, 2161, S13
  oral contraceptives and, 3153
  paraneoplastic, 741–742, 742t
  perioperative prophylaxis for, 3909–3910
  postmenopausal hormone therapy and, 3144t, 3145, 3147f
  in pregnancy, 3902
  prevention of, 742, 2162, 2164, 2164t, 2297, 3461
  risk factors for, 473, 474f
  treatment of, 2161–2162, 3902, S13
  upper extremity, 2161, S13
Deerpox, 1518t
Deer tick. *See Ixodes* spp. ticks
Defaults/status quo bias, 3916t, 3918
Defecation, 304–305, 304f
Defecography, 314
Defensins, 453
Defensive medicine, 23
Deferasirox, 777, 1721–1722, 3337
Deferiprone, 777, 3337
Deferoxamine, 777, 2019, 3337, 3714t
Defibrillator, implantable. *See* Cardiovascular implantable electronic devices (CIEDs)
Deformity, joint, 2942, 2942t
Degarelix, 701, 702f
Deglutitive inhibition, 293
Dehalogenase 1 (IYD), 3028t
Dehydrated stomatocytosis, 795
Dehydration
  in cholera, 1327–1328, 1328t, 1329t
  delirium in, 183
  in diarrhea, 307
  in hypothermia, 3770
  mechanisms of, 2371, 2371f
  in terminally ill patient, 91t
Dehydroemetine. *See* Emetine
Dehydroepiandrosterone (DHEA), 3051f, 3070, 3138
Deiodinase(s), 3021f, 3024
Déjérine-Roussy syndrome, 176, 3432, 3454
Delafloxacin, 1166, 1174t, 1182t, 1203t, 1204
Delamanid, 1394t, 1395t, 1398–1400, 1420t, 1425
Delandistrogene moxeparvovec, 3824t
Delavirdine, 1619f
Delayed afterdepolarizations (DADs), 1913, 1913f, 1913t
Delayed puberty
  in β thalassemia, 778t
  female, 3131–3132, 3131t
  male, 3109t, 3110
Delayed sleep-wake phase disorder, 212, 218, 3947–3948, 3948t
Delayed-type hypersensitivity reaction, 419, 1384–1385, 1408, 2783
Deletion. *See* Insertions/deletions
Delirium, **181**
  approach to the patient, 183
  assessment of, 85, 183, 185t
  clinical features of, 181–182
  in critically ill patient, 182, 186, 2298, 2348
  dementia and, 182
  diagnosis of, 184–185t, 185
  in DLB, 3492
  drug-induced, 182–183, 184t
  epidemiology of, 182
  etiology of, 85, 184–185, 184t
  at high altitudes, 3759
  history in, 183
  morbidity of, 182
  in older adults, 182, 1089, 2348, 3885–3889, 3889t
  pathogenesis of, 182–183
  physical examination in, 183–184
  postoperative/ICU, 2298
  prevention of, 186
  reversibility of, 182
  risk factors for, 182, 3889t
  in terminally ill patient, 85–86, 86t
  treatment of, 85–86, 86t, 185–186, 186f, 2348
Delirium tremens, 18, 3679
Delta fraction, conjugate bilirubin, 321
Delta hepatitis. *See* Hepatitis D virus (HDV) infection
Delta receptors, 3689t
Deltaretroviruses. *See* Human T-cell lymphotropic virus (HTLV-1)
Delusional infestations, 3737f, 3738
Delzicol, 2564, 2564t
Demeclocycline, 351, 619, 738, 3020, S1
Dementia, **193**
  in AD. *See* Alzheimer's disease
  aging and, 3872
  aphasia in, 203–204
  approach to the patient, **195**
  clinical features of, 195, 196t, 241
  in Creutzfeldt-Jakob disease. *See* Creutzfeldt-Jakob disease (CJD)
  definition of, 193
  delirium and, 182
  diagnosis of, 195t, 197–199
  differential diagnosis of, 196t
  epidemiology of, 3885–3887
  etiology of, 193–194, 194–195t, 3482–3485
  frontotemporal. *See* Frontotemporal dementia (FTD)
  functional anatomy of, 193
  genotypes of, 35
  global considerations in, 199
  history in, 195–196
  HIV-associated, 1606, 1607f, 1607t
  imaging in, 198f
  insomnia in, 216
  with Lewy bodies. *See* Dementia with Lewy bodies
  molecular basis of, 195t
  neurologic examination in, 197
  nutritional support in, 2626
  obesity and, 3187
  in PD, 3476, 3491–3492, 3506
  physical examination in, 197

Dementia (*Cont.*):
  postmenopausal hormone therapy and, 3144t, 3146
  prevalence of, 193
  risk factors for, 194–195
  semantic, 207
  treatment of, 199–200, 3888–3889, 3889t
  types of, 194
  vascular, 194
Dementia pugilistica. *See* Chronic traumatic encephalopathy
Dementia with Lewy bodies
  vs. AD, 3492, 3492t
  clinical features of, 195, 196t, 3492
    autonomic dysfunction, 3549
    delirium, 182
    hallucinations, 196
    myoclonus, 3530
    olfactory dysfunction, 241
    REM sleep behavior disorder, 218
  vs. Creutzfeldt-Jakob disease, 3532
  diagnosis of, 3492t, 3493t
  glucocerebrosidase in, 3400
  molecular basis of, 195t
  vs. Parkinson's disease dementia, 3491–3492, 3506, 3887
  pathogenesis of, 3492–3493
  pathology of, 3492
  prions in, 3399f
  treatment of, 3493
Demethylation, of DNA, 539
Demodex spp., 395
Demyelination, in MS, 3581, 3582, 3582f
Dendritic cells
  in autoimmune disease, 2824
  definition of, 2755
  functions of, 2762–2763, 2763t, 2764f, 2768
  hematopoietic differentiation of, 760f, 2763, 2764f
  in HIV infection, 1580
  myeloid, 2762, 2763t, 2824
  neoplasms of, 872t, 881–882
  plasmacytoid 1, 2762, 2763t
  for prostate cancer, 2784
  in sepsis/septic shock, 2319
Dengue hemorrhagic fever, 3858
Dengue shock syndrome, 3858
Dengue viruses, 1659t
Dengue virus infection
  climate change and, 3748, 3749–3750f
  clinical features of, 139t, 402, 468, 993, 1665, 1668, 3858, A1
  diagnosis of, 1668
  epidemiology of, 139t, 145, 993, 1668, 3857–3858
    Aedes distribution and, 3748, 3749f
    El Niño and, 3750, 3753f
  severe, 993, 1668, 1678–1679
  vaccine, 1017, 1678–1679
Dennie-Morgan folds, 383
Denosumab
  actions of, 534t, 542, 3307
  adverse effects of, 3290, 3307
  discontinuation of, 3307
  for fibrous dysplasia, 3317
  for hypercalcemia, 365, 738, 3284t, 3285
  indications for, 3307
  for metastatic bone disease, 731
  in metastatic breast cancer, 639
  for osteoporosis management/prevention, 3145, 3306–3307, 3306f
  for Paget's disease of bone, 3303
  prophylactic, 701, 703, 892–893
Dense deposit disease, 2420, A4. *See also* Membranoproliferative glomerulonephritis
Dense-granules, 938
Dental care/procedures
  bleeding with, 472
  in medically complex patient, 266
  preventive, 261
  prophylactic antibiotics for, 266, 1048, 1048t, 1180t
  prophylaxis for bleeding in, 927
Dental disease
  brain abscess and, 1135
  calculus, A3
  caries, 260, 1374, A3
  ear pain in, 253
  gingivitis. *See* Gingivitis
  high-altitude travel and, 3760
  involuntary weight loss in, 315
  lymphadenopathy in, 477
  in osteogenesis imperfecta, 3324, 3326
  periodontitis. *See* Periodontal disease (periodontitis)
  plaque, 260, 1374, A3
  root cavity, A3
Dentatorubropallidoluysian atrophy, 3536, 3793t
Dentin, 260
Dent's disease, 2368t, 2369, 3264
Dentures, 267, 1708, A3
Deoxycoformycin, 873
Deoxycorticosterone (DOC), 3051f, 3054, 3061
Deoxyhemoglobin, 771f
DEPDC5 gene mutations, 3412t
Dependence
  alcohol. *See* Alcohol use disorder (alcoholism)
  substance. *See* Substance use disorders
Dependoparvovirus, 1525. *See also* Adeno-associated virus (AAV)
Depilatory, 3141
"Deprescribing," 3885, 3888f
Depression/depressive disorders, **3664**
  agitated, 3668
  assessment of, 84
  asthma exacerbations and, 2222
  in cancer patient, 504, 757, 3664
  in cardiovascular disease, 3664
  chemosensory alterations in, 242
  circadian disruption and, 3952
  clinical features of, 3664–3665, 3665t
  in diabetes mellitus, 3664
  diagnosis of, 3665t
  dizziness in, 165
  drug-related, 3664
  etiology of, 84, 3665–3666
  headache and, 114
  heart failure outcome and, 1986, 1998
  in HIV infection, 3664
  in hypothyroidism, 3664
  insomnia in, 215
  involuntary weight loss in, 315–316
  major depression, 3664–3665, 3665t, 3666f
  in medical illness, 3664
  in military veterans, S8
  nausea and vomiting in, 298
  in neurologic disorders, 3664
    AD, 3482
    epilepsy, 3426–3427
    Huntington's disease, 3514
    MS, 3578, 3588
    PD, 3506
  pain and, 99–100
  pathophysiology of, 3654, 3655f, 3666
  persistent, 3665
  postpartum, 3167, 3667
  screening for, 39t
  sexual dysfunction in, 3157, 3161
  shift work and, 3952
  in terminally ill patient, 84–85
  treatment of, **3666**
    algorithm for, 3666f
    antidepressants, 3659t, 3660t, 3666–3667
    in cerebral function disorders, 209
    complementary and integrative therapies, 3926, 3927
    supportive, 3666
    in terminally ill patient, 84–85
    transcranial magnetic stimulation, 3667, 3967
  unipolar, 3665
  in women, 3167, 3665
de Quervain's tenosynovitis, 2942f, 2943, 2974
de Quervain's (subacute) thyroiditis, 265, 3038, 3038f, 3038t
Dermacentor spp. ticks, 1456, 1669, 3730. *See also* Tick-borne diseases
Dermal papillae, 1049f
Dermatitis, **382**
  atopic. *See* Atopic dermatitis
  cercarial (swimmer's itch), 1051, 1736, 1828
  contact. *See* Contact dermatitis
  seborrheic. *See* Seborrheic dermatitis
  stasis. *See* Stasis dermatitis
Dermatitis herpetiformis, **412**
  clinical features of, 379, 380f, 410t, 412, A5
  diagnosis of, 412
  gluten sensitivity in, 412, 2545
  histology of, 410t, 412
  HLA gene association with, 412
  immunopathology of, 410t, 412
  treatment of, 412–413
Dermatofibroma, 378t, 380f
Dermatofibrosarcoma protuberans, 609f, 610, 729
Dermatofibrosis lenticularis disseminata (Buschke-Ollendorff syndrome), 3316
Dermatographism (dermographism), 402, 2807, 2807f, A5
Dermatoheliosis, 431
Dermatomes, 175f
Dermatomyositis (DM), **2911**
  clinical features of, 172, 413–414, 2910t, 2911
    cutaneous, 413–414
      on face, 414f
      on hands, 414f, 2911, 2911f, A5
      rash, 2911, 2911f, A5
      telangiectasias, 396, 414
    interstitial lung disease, 2266
    myocarditis, 2015
    Raynaud's phenomenon, 2178
  diagnosis of, 2910t, 2911
  drug-induced, 420, 2940t
  histopathology of, 2911–2912, 2912f
  immunopathogenesis of, 2790t
  in lung cancer, 619
  pathogenesis of, 2911–2912
  prognosis of, 2912
  treatment of, 414, 2914–2915, 2916t
Dermatomyositis sine myositis, 413
Dermatophagoides spp., 2809
Dermatophyte infections, **389**
  clinical features of, 378t, 386t, 388t, 389, 1688t, 1723
  diagnosis of, 382, 389, 1688t, 1723
  epidemiology of, 1723
  etiology of, 388t, 389, 1723
  histologic features of, 386t
  pathogenesis of, 1723
  risk factors for, 1688t
  treatment of, 388t, 389, 1723–1724, 1724t
Dermographism (dermatographism), 402, 2807, 2807f, A5
Dermoid cysts, 712, 722
Dermopathy, thyroid, 3034f, 3037
Dermoscopy, 382, 600, 607f
DES (diethylstilbestrol), 512t, 701
Descending perineum syndrome, 313

- Desensitization. *See* Drug desensitization
- *DES* gene mutations, 2004t
- Desipramine
  - adverse effects of, 3659t, 3666
  - for depression, 504, 3588, 3659t, 3666
  - for neuropathic pain, 101
  - for pain, 97t, 99, 3588
- Desirudin, 950, 950t
- Desloratadine, 2810
- Desmin, 2004t, 2008, 2790t
- Desmogleins, 409, 410, 2790t
- Desmoplakin, 2791t
- Desmopressin (DDAVP)
  - for adult hypopituitarism, 2996t
  - adverse effects of, 926, 3018
  - for coagulation disorders in CKD, 2393
  - for diabetes insipidus, 353, 3018–3019, S1
  - in diabetes insipidus, 344
  - for hemophilia, 930
  - for hypernatremia, 353
  - for hypodipsic hypernatremia, 3019t
  - for hyponatremia, 351, S1
  - for orthostatic hypotension, 160, 3548
  - for platelet function disorders, 926
  - for sleep enuresis, 217
  - for VWD, 927
- Desmosome complexes, 2004t, 2005f
- Desvenlafaxine, 3143, 3227, 3659t, 3667
- Desynchrony, 3944t, 3947
- Des-γ-carboxy prothrombin (DCP), 661
- Detergent worker's lung, 2228t
- Detrusor hyperreflexia/dyssynergia, 3588
- Detumescence, 3155, 3156f
- Deucravacitinib, 387, 387t, 2835, 2890
- Developing world. *See* Low- and middle-income countries
- Device-related infections. *See* Health care–associated infections
- Devic's disease. *See* Neuromyelitis optica (NMO)
- DEXA (dual-energy x-ray absorptiometry), 2618t, 2717, 3277, 3296–3297
- Dexamethasone
  - actions of, 1168
  - adverse effects of, 2027, 3648
  - for AL amyloidosis, 898
  - for brain abscess, 1137
  - for brain metastases, 590
  - for CAH, 3072
  - for cardiac amyloidosis, 2027
  - for cerebral edema, 716
  - for COVID-19 disease, 996
  - drug interactions of, 1742t
  - for enteric fever, 1318
  - epidural injection of, 3393
  - for fatigue, 84
  - for hemophagocytic lymphohistiocytosis, 882
  - for high-altitude illness, 3756, 3756t, 3758
  - for high-altitude illness prevention, 3757
  - for high-altitude pulmonary edema, 2331
  - for hirsutism, 3141
  - for immune thrombocytopenic purpura, 924
  - for increased intracranial pressure, 2345t
  - for lymphoplasmacytic lymphoma, 865
  - for meningitis, 963t, 991t, 994, 1121, 1122, 1263
  - for multiple myeloma, 890t, 891–892
  - for nausea and vomiting, 83, 299t, 508, 575
  - for neurocysticercosis, 1138, 1834
  - for pain, 82
  - as premedication for taxanes, 565
  - for prenatal treatment of CAH, 3101
  - for spinal cord compression, 3562
  - for tuberculous meningitis, 1127
- Dexamethasone suppression test, 3050, 3141
- Dexlansoprazole, 301, 2524t, 2525
- Dexmedetomidine, 186, 2298, 2348
- Dexrazoxane, 564
- Dextroamphetamine, 81, 84, 85, 215
- Dextrocardia, A7
- Dextromethorphan, 253, 274, 3693, 3715t
- Dextromethorphan-bupropion, 3659t
- Dextrose, 347, 353, 361
- DF (discriminant function) value, 2709
- DGM (diphosphoglycerate mutase) deficiency, 796t
- *DHCR7* gene mutations, 3099t
- DHEA. *See* Dehydroepiandrosterone (DHEA)
- DHF (dihydrofolate), 781, 783f
- DHF (dihydrofolate) reductase, 574, 1182t, 1185
- DHF (dihydrofolate) reductase inhibitors, 1743t, 1751
- Dhori virus, 1662t
- DHPR (dihydropteridine reductase) deficiency, 3368t
- DHT (dihydrotestosterone), 3102, 3105f, 3138–3139. *See also* Testosterone
- *DHX377* gene mutations, 3099t
- Diabetes insipidus (DI), **3016**
  - adipsic, 352
  - with AVP resistance (nephrogenic)
    - clinical features of, 344, 352, 2362, S1
    - diagnosis of, 343f, 344, 352
    - drug-induced, 3016
    - early interventions for, 3804t
    - etiology of, 343, 343f, 352, 3016, 3017t, S1
    - genetic considerations in, 2368t, 3017t
    - genetic testing for, 3804t
    - lithium-associated, S1
    - pathophysiology of, 352, 3017
    - treatment of, 353, 3019, S1
  - clinical features of, 3016
  - differential diagnosis of, 3017–3018, 3018f
  - with primary AVP deficiency (pituitary, central)
    - clinical features of, 343, 352, 3016
    - diagnosis of, 343f, 344, 352
    - differential diagnosis of, 3017–3018
    - early interventions for, 3804t
    - etiology of, 343f, 344, 3015, 3016, 3017t
    - genetic considerations in, 3015, 3017t, 3787
    - genetic testing for, 3804t
    - pathophysiology of, 3017
    - treatment of, 353, 3018–3019
- Diabetes mellitus (DM), **3195**
  - abdominal pain in, 111, 112t
  - acromegaly and, 3008
  - approach to the patient, 3203–3204
  - atypical, 3196, 3204
  - autoantibodies in, 2825, 2825t
  - autoimmune in the adult, 3204
  - celiac disease and, 2545
  - classification of, 3195–3196, 3196f, 3196t, 3204
  - colorectal cancer and, 655
  - complications of, **3222,** 3223t
    - cardiovascular. *See* Cardiovascular disease, in diabetes mellitus
    - carotenoderma, 320
    - cutaneous, 3230
    - detection and prevention of, 3205, 3205t
    - diabetic ketoacidosis. *See* Diabetic ketoacidosis (DKA)
    - dyslipidemia, 3228
    - erectile dysfunction, 3157, 3227
    - female sexual dysfunction, 3161, 3227
    - gastrointestinal, 297, 3227
    - genitourinary, 1087, 3227
    - glycemic control and, 3222–3223, 3222f
    - hyperglycemic hypermolar state, 3216, 3217t, 3218–3219
    - hyperkalemia, 358
    - hypertension, 3225
    - hypoglycemia, 3230, 3232–3234, 3233f
    - infections, 3229–3230
    - lactic acidosis, 369
    - lower extremity, 49t, 1061, 1068, 1068f, 3204, 3228–3229
    - MASLD, 2702
    - mechanisms of, 3223
    - mucormycosis, 1718
    - muscle involvement, 3647
    - neuropathy. *See* Diabetic neuropathy
    - ocular
      - abducens nerve palsy, 235
      - cataracts, 230
      - epiretinal membrane, 232
      - oculomotor nerve palsy, 234–235
      - retinopathy. *See* Diabetic retinopathy
      - tonic pupil, 221
      - vitreous degeneration, 229
    - renal. *See* Diabetic nephropathy
    - retinopathy. *See* Diabetic retinopathy
  - contrast agent risk in, 1904, 3224
  - COVID-19 disease and, 3203
  - cystic fibrosis–related, 2245
  - deaths from, 75t
  - definition of, 3195
  - depression in, 3664
  - in developing countries, 1858, 3850
  - diagnosis of, 3196t, 3197–3198, 3198t, 3204
  - diarrhea in, 309, 312
  - epidemiology of, 3196–3197
  - etiology of, 3195, 3196t
  - genetic considerations in, 3195, 3196t, 3202–3203, 3783t, 3787, 3793, 3794t
  - gestational, 3165, 3196, 3221, 3900–3901
  - global considerations in, 1858, 3196–3197, 3197f, 3850
  - HCV infection and, 2685
  - in hemochromatosis, 3336
  - high-altitude travel and, 3760
  - history in, 3203–3204
  - immune checkpoint–related, 3203
  - immunization recommendations for, 1004t
  - incidence of, 3196–3197
  - involuntary weight loss in, 315
  - in Kearns-Sayre syndrome, 3645
  - maturity-onset of the young (MODY), 3195, 3202–3203, 3783t, 3794t
  - nausea and vomiting in, 297
  - obstructive sleep apnea and, 2278
  - oral manifestations of, 65, 261
  - pancreatic cancer risk and, 671
  - pancreatic exocrine disease and, 3203
  - parenteral nutrition in patient with, 3220
  - physical examination in, 3204
  - in pregnancy, 3221, 3900–3901
  - preoperative evaluation and management of, 3909
  - prevalence of, 3196–3197, 3197f
  - prevention of, 3851
  - screening for, 39t, 3197–3198, 3198t, 3881t
  - skin manifestations of, 397, 401, 404, 3230, A15
  - stroke risk in, 3446t
  - systems biology applied to, 3961t
  - taste disorders in, 242
  - treatment and management of, **3205**
    - blood glucose monitoring in, 3207
    - emerging therapies, 3216
    - during glucocorticoid therapy, 3220
    - glycemic control monitoring, 3207–3208, 3207t
    - glycemic control targets, 3208–3209, 3209t
    - goals and principles of, 3205, 3205t
    - in hospitalized patients, 3219–3220
    - lifestyle management, 3205
    - multidisciplinary team in, 3205
    - nutrition therapy in, 3206, 3206t
    - in older adults, 3209t, 3220–3221, 3882, 3882t

- Diabetes mellitus (DM), treatment and management of (*Cont.*):
    - ongoing aspects of, 3205–3207, 3205t
    - outpatient, improvement of, 44
    - during parenteral nutrition therapy, 3220
    - physical activity in, 3206
    - in pregnancy, 3221. *See also* Gestational diabetes mellitus
    - psychosocial care in, 3206–3207
    - self-management education and support, 3205–3206
  - type 1, **3199**
    - APS-1 and, 3090t, 3227
    - APS-2 and, 3091, 3091t
    - environmental factors in, 3200–3201
    - genetic considerations in, 3200, 3202–3203
    - HLA gene association with, 3837
    - hypertension management, 3225
    - immunologic markers in, 3199–3200
    - immunopathogenesis of, 2782, 2790t
    - inflammasome mutations in, 2762t
    - KIRs with, 2772t
    - microbiome and, 3837
    - pathogenesis of, 3199–3200, 3199f
    - pathophysiology of, 3200
    - staging of, 3199f, 3200
    - treatment and management of, **3209,** 3211t
      - adverse effects of, 3216
      - emerging therapies, 3216
      - goals of, 3208–3209, 3209t
      - insulin preparations, 3209t, 3210
      - insulin regimens, 3210, 3212
      - intensive management, 3208f, 3209–3210
      - pramlintide, 3212
  - type 2, **3201**
    - clinical features of, 3204
    - dyslipidemia and, 3246
    - genetic considerations in, 3201, 3793, 3794t, 3796f
    - hypertension management, 3225
    - insulin resistance syndromes in, 3202
    - ketosis-prone, 3204
    - metabolic abnormalities in, 3201–3202, 3201f
    - metabolic syndrome and, 3254, 3256
    - obesity and, 3187
    - pathophysiology of, 3201, 3201f
    - postmenopausal hormone therapy and, 3144t
    - prevalence of, 2979t
    - prevention of, 3202
    - protein aggregation in, 3994
    - risk factors for, 3196–3197
    - in schizophrenia, 3674
    - screening/testing for, 2979t, 3198–3199, 3198t
    - treatment and management of, 3211t, **3212,** 3212f
      - adverse effects of, 3216
      - bariatric surgery, 3194–3195, 3216
      - biguanides, 3212–3213
      - bile acid–binding resins, 3214
      - bromocriptine, 3214
      - combination therapy, 3215–3216
      - emerging therapies, 3216
      - α-glucosidase inhibitors, 3214
      - insulin secretagogues, 3213
      - insulin therapy, 3214
      - selection of initial agent, 3214–3215, 3215f
      - SGLT-2 inhibitors, 3213
      - thiazolidinediones, 3214
  - in women, 3165
- Diabetic diarrhea, 309, 312
- Diabetic ketoacidosis (DKA), **3216**
  - clinical features of, 3216–3217, 3217t
  - complications of, 3218
  - diagnosis of, 3217, 3217t, S1
  - euglycemic, 3217, 3217t
  - hypokalemia in, 354
  - hyponatremia in, 347
  - pathophysiology of, 370, 3217
  - prevention of, 3218
  - treatment of, 370, 3217–3218, 3218t, S1
- Diabetic nephropathy, **3224**
  - albuminuria in, 3224, 3224f, 3226
  - autonomic dysfunction in, 3549–3550
  - clinical features of, 2415t
  - epidemiology of, 2423
  - glycemic control and, 3223, 3224f, 3225
  - membranous nephropathy in, 2359
  - microalbuminuria in, 2424
  - natural history of, 2423–2424, 3224, 3224f
  - pathogenesis of, 2388, 2412, 2424, 3224
  - renal biopsy in, 2423, A4
  - treatment of, 2359, 2424, 3225–3226
  - in type 1 vs. type 2 DM, 2424, 3224
- Diabetic neuropathy, **3226**
  - autonomic, 3226, 3227, 3228, 3549–3550, 3604
  - constipation in, 314
  - distal symmetric sensory and sensorimotor polyneuropathy, 3604
  - mononeuropathy, 3226, 3604
  - pathophysiology of, 96
  - polyneuropathy, 3226, 3604
  - polyradiculopathy, 3226
  - prevalence of, 3226
  - radiculoplexus, 3604
  - risk factors for, 3226
  - screening for, 3204
  - treatment of, 99, 3226–3227, 3604t
- Diabetic retinopathy
  - advanced, 231, 231f
  - clinical features of, 3223, 3223f
  - detection system for, 3973
  - glycemic control and, 3223
  - treatment of, 231f, 3224
- "Diabetic skin spots," 3230
- Diacetyl exposure, 2240
- Diagnosis
  - "don't miss," 57–58, 59t
  - errors in, 56, 56f, 57t
  - failure to reach. *See* Undiagnosed disease
  - health information technology and, 59t, 60
  - improving, 56–57, 56–57t, 60
  - meaning and context of, 3995
  - pitfalls in, 58–59, 59t
  - quality of care and, 27
  - safety culture and, 59–60
  - uncertainty in, 57
- Diagnostic imperative, 22
- Diagnostic refinement, 22
- Diagnostic tests. *See* Clinical laboratory tests
- Diagnostic verification, 22
- Dialectical behavior therapy (DBT), 3672
- Dialysate, 2398–2399
- Dialysis. *See* Hemodialysis; Peritoneal dialysis
- Dialysis access, 2399–2400
- Dialyzer, 2398, 2400
- Diaminopyridine, 3588, 3628
- Diamond-Blackfan anemia, 810, 813
- Diapedesis, 454, 457f
- Diaper rash, 1708
- Diaphyseal aclasis, 3318
- Diarrhea
  - acute, **305**. *See also* Diarrhea, infectious
    - approach to the patient, 306–307, 307f
    - etiology of, 305–306
    - treatment of, 307
  - after cholecystectomy, 2734
  - antibiotic-associated, 307. *See also* *Clostridioides difficile*–associated infection (CDI)
  - bile acid, 2543, 2543t
  - bloody, 305, 306t, 1739t
  - Brainerd, 307
  - in carcinoid syndrome, 681–682
  - in celiac disease, 2545
  - chemotherapy-induced, 508, 576
  - chronic, **307**
    - approach to the patient, 310–311, 310f, 311t
    - definition of, 305
    - dysmotility causes of, 308t, 309–310
    - endoscopic findings in, 2470f, 2497–2498, 2503f
    - factitial causes of, 310
    - inflammatory causes of, 308t, 309
    - osmotic causes of, 308–309, 308t
    - secretory causes of, 307–308, 308t
    - treatment of, 311–312
    - in war veterans, S7
  - in Crohn's disease, 2557
  - definition of, 305, 2540
  - diabetic, 312
  - differential diagnosis of, 2460t, 2461, 2540
  - with enteral nutrition, 2623
  - epidemiology of, 302–303
  - fatty acid, 2543t
  - in HIV infection, 1598–1600, 1600f
  - hypernatremia in, 352
  - hypokalemia in, 354
  - hyponatremia in, 347
  - hypovolemia in, 346
  - in IBS, 2571
  - infectious, 305–306, **1076**. *See also* Gastroenteritis; Travelers' diarrhea
    - approach to the patient, 1078–1079, 1078f
    - bacterial vs. viral, 1629t
    - *Campylobacter. See Campylobacter* spp. infections
    - *C. difficile. See Clostridioides difficile*–associated infection (CDI)
    - clinical features of, 305–306, 306t, 1629t
    - complications of, 1079, 1079t
    - *C. perfringens,* 1240. *See also Clostridium perfringens* infections
    - *E. coli. See Escherichia coli* infections, intestinal
    - *Edwardsiella,* 1294
    - epidemiology of, 305, 1079–1080
    - food-related, 305, 1080, 1080t. *See also* Food-related illness
    - *Giardia* spp. *See Giardia* spp. infections
    - host defenses against, 1077
    - in immunodeficient persons, 305
    - inflammatory, 1077t
    - in institutionalized persons, 305
    - laboratory evaluation of, 1080
    - noncholera *Vibrio,* 1330, 1330t
    - noninflammatory, 1077t
    - parasitic, 1739t
    - pathogenesis of, 305–306, 1076–1077, 1077t
    - penetrating, 1077t
    - prevention of, 1081–1082
    - protozoal, 1805–1808
    - *Salmonella. See Salmonella* spp. infections
    - *Shigella. See Shigella* spp. infections
    - with systemic manifestations, 306
    - treatment of, 1081, 1081t, 1288
    - *T. whipplei,* 1368
    - *V. cholera. See* Cholera
    - *Yersinia,* 1348
  - in malabsorption disorders, 2540, 2549t
  - metabolic acidosis in, 346, 372
  - osmotic, 352, 2540, 2548

  overflow, 305
  paraneoplastic, 737t
  persistent, 305
  postvagotomy, 2533–2534
  in protein-losing enteropathy, 2548
  secretory, 352, 2541
  in short bowel syndrome, 2547
  in small-bowel bacterial overgrowth, 2543
  travelers'. *See* Travelers' diarrhea
  in tropical sprue, 2546
  in ulcerative colitis, 2556–2557, 2557t
  water loss in, 2598
  in Whipple's disease, 2546
  in ZES, 2535
  zinc losses in, 2626
Diarrhetic shellfish poisoning, 3728
Diascopy, 382
Diastolic dysfunction, 1853, 1853f, 1882, 1883f
Diastolic failure. *See* Heart failure (HF)
Diastolic murmur. *See* Heart murmurs
Diatrizoate meglumine, 923t
Diazepam
  adverse effects of, 1231–1232
  for agitation, 3696
  for alcohol withdrawal, 3680
  for anxiety disorders, 3660t
  for cocaine intoxication, 3695
  drug interactions of, 1742t, 2525
  for insomnia, 86
  for marine envenomations, 3723
  for nerve agent exposure, S5
  overdosage/poisoning with, 3713t
  pharmacology of, 3660t
  for spasms in tetanus, 1231
  for spasticity, 3570, 3588
  for vertigo, 165t
Diazoxide, 421, 3235
Dibasic aminoaciduria (lysinuric protein intolerance), 2368t, 3373, 3374t
Dibothriocephalus latus (fish tapeworm), 787, 1736, 1739t, 1838–1839, S12
DIC. *See* Disseminated intravascular coagulation (DIC)
Dicarboxylic aminoaciduria, 3374t
Dichloralphenazone, 3467t, 3469
Dichromat, 222
Dick test, 1209
Diclofenac
  adverse effects of, 434t, 923t, 2955
  cream, 3883
  with gastric protective agent, 2531
  for migraine, 3467t
  for osteoarthritis, 2955t
  for pain, 96
Dicloxacillin
  adverse effects of, 325, 355
  for impetigo, 1210
  indications for, 1175
  for staphylococcal skin/soft tissue infections, 383, 964t, 1203t
Dicyclomine, 2575
Didanosine (ddI), 1616, 1619f, 1622f, 2672, 3611
Dideoxynucleosides, 1610
DIDMOAD (Wolfram's) syndrome, 3093
Dientamoeba fragilis, 961t, 1748, 1805t, 1808
Diet
  after STEMI, 2121
  asthma risk and, 2220
  autoimmune disease and, 2823
  for cancer prevention, 511
  cardiovascular disease and, 1857
  circadian clock and, 3951
  colorectal cancer and, 654
  composition of, 2599
  dyslipidemia and, 3246
  energy density of, 3190
  global considerations in, 1857
  head and neck cancer and, 610
  for hypertension prevention, 2137t, 2140–2141
  infectious disease exposure and, 959
  microbiome and, 3833–3834
  nephrolithiasis and, 2449, 2452–2453
  pancreatic cancer and, 671
  peptic ulcer disease and, 2522
  prostate cancer and, 695
Dietary assessment. *See* Nutrition assessment
Dietary fiber
  cancer and, 511
  colorectal cancer and, 654
  for constipation, 312, 314
  for diverticular disease, 2580, 2580t
  fermentation of, 3840
  for IBS, 2574–2575
  recommended intake of, 2602t
Dietary Guidelines for Americans, 2020–2025, 10t, 11, 2142
Dietary reference intakes, 2599
Dietary supplements, 2599, 2672, 3925–3927, 3925t, 3927t
Dietary Supplements Health and Education Act (DSHEA), 3926t
Diet history, 2603
Diethylcarbamazine (DEC)
  actions of, 1747
  adverse effects of, 1745t, 1822
  for loiasis, 1824
  for lymphatic filariasis, 1821
  pharmacology of, 1747
  for tropical pulmonary eosinophilia, 1822
Diethylene glycol poisoning, 2377, S1
Diethylpropion, 3191
Diethylstilbestrol (DES), 512t, 701
Diet therapy. *See* Nutritional support/therapy
Dieulafoy's lesion, 318, 2482, 2488f, V5
Differentiation, in cancer cells, 556
Diffuse alveolar hemorrhage, 275
Diffuse cutaneous leishmaniasis, 1787
Diffuse fasciitis with eosinophilia, 416
Diffuse idiopathic skeletal hyperostosis, 2886
Diffuse infiltrative lymphocytosis syndrome (DILS), 1603
Diffuse large B-cell lymphoma (DLBCL), **862**
  clinical features of, 862
  CLL transformation to, 853–854
  diagnosis of, 862–863, 863f
  frequency of, 858f
  gastric, 650
  genetic considerations in, 860t
  in HIV infection, 1613
  lymph node examination in, A6
  treatment of, 536, 650, 863, 3824t
Diffuse leprosy of Lucio and Latapí, 1404, 1408
Diffusely adherent *E. coli* (DAEC), 1288
Diffuse nontoxic goiter, 3040–3041. *See also* Thyroid nodular disease
Diffuse (periventricular) white matter disease (leukoaraiosis), 198f, 3448
Diffusion capacity of the lung for carbon monoxide (DL_{CO}), 2206f, 2207
Diffusion tensor imaging (DTI), 3390, 3392f, 3490
Diffusion-weighted imaging, 3437, 3438f, 3530, 3531f
Diflunisal, 900
DiGeorge's (velocardiofacial) syndrome, 2799, 3286, 3652, S9
Digestion. *See* Absorption
Digitalis
  actions of, 1849
  adverse effects of, 228, 1937, 3116
  ECG effects of, 1874
  for pulmonary edema, 2331
Digital necrosis, in systemic sclerosis, 2867, 2867f
Digital rectal examination, 517, 517t, 656, 698
Digit span tests, 183, 208
Digoxin
  adverse effects of, 923t, 1920t, 1926, 1929
  for atrial fibrillation, 1949
  drug interactions of, 485t, 489t
    azithromycin, 1744t
    azoles, 1742t
    PPIs, 2525
    pyronaridine, 1752
  for heart failure, 1996
  metabolism of, 485t
  overdosage/poisoning with, 358, 3712t
  for supraventricular arrhythmias, 2124
  toxicity of, 356, 358, 361
DIHS. *See* Drug-induced hypersensitivity syndrome (DIHS)
Dihydroartemisinin, 1741t, 1746
Dihydroergotamine
  for exercise headache, 3474
  for medication-overuse headache, 116
  for migraine, 3467t, 3468, 3468t
Dihydrofolate (DHF), 781, 783f
Dihydrofolate (DHF) reductase, 574, 1166, 1182t, 1185
Dihydrofolate (DHF) reductase inhibitors, 1743t, 1751
Dihydropteridine reductase (DHPR) deficiency, 3368t
Dihydropteroate synthetase, 1166, 1182t
Dihydropteroate synthetase inhibitors, 1743t
Dihydropyridines. *See* Calcium channel blockers (CCBs)
Dihydropyrimidine dehydrogenase (DPYD), 493, E72t
Dihydrorhodamine oxidation test, 460
Dihydrotestosterone (DHT), 3102, 3105f, 3138–3139. *See also* Testosterone
1,25-Dihydroxyvitamin D. *See also* Vitamin D
  actions of, 362, 362f, 3268
  overproduction of, 362, 737t, 738
  synthesis of, 431
  underproduction of, 740
Diisocyanate exposure, 2240
Dilated cardiomyopathy, **2015**
  alcohol-related, 2015–2016
  chemotherapy-related, 2016
  in CKD, 2392
  clinical features of, 2002t
  diagnosis of, 2002t
  echocardiography in, 2009f
  etiology of, 2008–2009
  familial, 2016t
  genetic considerations in, 1971, 2004t, 2008–2009, 3788t
  heart failure in, 1984. *See also* Heart failure (HF)
  in hemochromatosis, 2018, 2019f
  in HIV infection, 2012
  metabolic causes of, 2016t, 2018–2019
  myocarditis and, 2012, 2016t. *See also* Myocarditis
  natural history of, 1841
  pathophysiology of, 2008–2009, 2009f, 2790t
  prognosis of, 2009–2010
  Takotsubo (stress-related), 104, 2019, 2019f
  toxic, 2015–2017, 2016t
  ventricular tachycardia in, 1965, 1966f, 1971
DILIN (Drug-Induced Liver Injury Network), 2668
Diloxanide furoate, 1742t, 1747
DILS (diffuse infiltrative lymphocytosis syndrome), 1603

- Diltiazem
  - for accessory pathway-mediated tachycardia, 1942
  - adverse effects of, 1926, 1929, 2101t, 2144t, 2940t
  - for aortic dissection, 2170
  - for atrial fibrillation, 1949
  - for AV nodal reentrant tachycardia, 1940
  - contraindications to, 1941
  - drug interactions of, 1745t, 2101, 2408, 2721
  - for hypertension, 2144t
  - for hypertension in CKD, 2396
  - for ischemic heart disease, 2101, 2101t
  - for NSTEMI-ACS, 2110
  - overdosage/poisoning with, 3712t
  - for Raynaud's phenomenon, 2178, 2874
  - for sex headache prevention, 3474
  - for supraventricular arrhythmia, 2124
- Dilute Russell's viper venom test, 475, 936, 2832
- Dimenhydrinate, 165t, 298, 299t
- Dimercaprol, 3700–3701t
- D-Dimer test, 472
- 1,3-Dimethlamylamine (DMAA), 3926
- Dimethyl fumarate (DMF), 3586t
- Dimethyl sulfoxide (DMSO), 336
- Dimethyltryptamine (DMT), 3696
- Dim-light melatonin onset (DLMO), 218, 3944t, 3946
- Dimorphic fungi, 1725
- Dinutuximab, 534t, 559f
- Dioxin exposure, S8
- Dipeptidyl peptidase-4 (DPP-IV) inhibitors, 412, 3211t, 3213, 3882t
- Diphenhydramine
  - adverse effects of, 2808
  - for anaphylaxis, 3719
  - for insomnia, 216
  - for nausea and vomiting, 509
  - overdosage/poisoning with, 3711t, 3715t
  - as premedication for taxanes, 565
  - for pruritus, 383, 422
- Diphenoxylate, 312, 2583
- Diphenoxylate hydrochloride/atropine sulfate, 2577t
- Diphenylhydantoin, 476
- Diphosphoglycerate mutase (DGM) deficiency, 796t
- Diphtheria, **1221**
  - approach to the patient, 1113
  - clinical features of, 257t, 259, 1223, 1223f, 2013, 3606
  - complications of, 1224
  - cutaneous, 1222, 1223, 1223f, 1224
  - diagnosis of, 1224, 2013
  - epidemiology of, 1222
  - etiology of, 1222
  - global considerations in, 1222
  - incidence pre- and postvaccine, U.S., 1002t, 1222
  - pathogenesis and immunology of, 1222
  - prevention of, 1004t, 1225
  - prognosis of, 1225
  - prophylaxis for close contacts, 1225
  - treatment of, 1224–1225
- Diphtheria antitoxin, 1222, 1225
- Diphtheria-tetanus-pertussis vaccine (DTaP), 1225, 1232, 1277. *See also* Tetanus, diphtheria, acellular pertussis (Tdap) vaccine
- Diphtheria-tetanus vaccine. *See* Tetanus-diphtheria (Td) vaccine
- *Diphyllobothrium latum* (fish tapeworm). *See Dibothriocephalus latus* (fish tapeworm)
- Diploid, 3778
- Diplopia, **234**
  - in abducens nerve palsy, 235
  - in botulism, 1235
  - evaluation of, 234
  - in Graves' disease, 3035
  - in horizontal gaze disorders, 235–236
  - monocular vs. binocular, 234
  - in multiple ocular nerve palsies, 235
  - in myasthenia gravis, 234. *See also* Myasthenia gravis
  - in nystagmus. *See* Nystagmus
  - in oculomotor nerve palsy, 234–235
  - in opsoclonus, 237
  - in supranuclear disorders of gaze, 235
  - in trochlear nerve palsy, 235
  - in vertical gaze disorders, 236
- Diplotype, 491
- Dipteran bite, 3736
- *Dipylidium caninum*/dipylidiasis, 1839
- Dipyridamole, 944–945, 945f, 2483t, 3449
- Direct antiglobulin (Coombs) test, 908, 910f
- Direct thrombin inhibitors, parenteral, 950, 950t, S3. *See also* Argatroban; Bivalirudin
- *Dirofilaria* infections, 1825
- Diroximel fumarate, 3586t
- Disability, 2240
- Disability-adjusted life years (DALY), 3844, 3845f, 3846t
- Disaster medicine, 1295
- *DISC1* gene mutations. *See* DiGeorge's (velocardiofacial) syndrome
- Discogenic pain, 119, 119f, 120, 123t, 127
- Discoid (chronic cutaneous) lupus erythematosus
  - alopecia in, 394
  - clinical features of, 415, 2829
  - diagnosis of, 415
  - oral manifestations of, 261
  - skin manifestations of, 415, 415f, A1, A5
  - telangiectasias in, 396
  - treatment of, 415
- Discriminant function (DF) value, 2709
- Discrimination, 67
- Disease
  - classification of, 3957, 3958f
  - control of, 1002
  - definition of, 964
  - elimination, 1002
  - eradication, 1002
  - genotype-phenotype relationships in, 3957, 3958f, 3960
  - germ theory of, 957
  - networks, 3957–3962, 3958–3961f
  - nosology, 29, 30f
  - probability, measures of, 24–25
  - of unknown etiology. *See* Undiagnosed disease
- *Disease Control Priorities,* 3847
- *Disease Control Priorities in Developing Countries,* 3844
- Disease-modifying antirheumatic drugs (DMARDs), 2849, 2850–2851t, 2851–2852
- Disequilibrium, 179, 181. *See also* Balance
- Disk/agar diffusion method, S11
- Diskography, 3393
- Disk-space infection. *See* Osteomyelitis, vertebral
- Dislocation, 2942, 2942t
- Disopyramide, 1938, 1952, 1958, 2007
- Disorders of sex development (DSDs), **3094**
  - chromosomal sex disorders. *See* Sex chromosome disorders
  - classification of, 3095t
  - clinical presentation by life stage, 3095t
  - global considerations in, 3102
  - gonadal and phenotypic sex disorders
    - 46,XX disorders. *See* 46,XX disorders of sex development
    - 46,XY disorders. *See* 46,XY disorders of sex development
  - holistic approach to, 3094
  - long-term follow-up, 3101–3102
- Disposable soma theory of aging, 3869
- Disseminated intravascular coagulation (DIC), **933**
  - in acute promyelocytic leukemia, 826
  - chronic, 934
  - clinical features of, 934, 990, 992
  - diagnosis of, 934, 934t
  - differential diagnosis of, 934
  - etiology of, 933, 933t
  - in filovirid infections, 1684
  - in hemophagocytic lymphohistiocytosis, 464
  - in leptospirosis, 1442
  - pathophysiology of, 933–934, 933f
  - in sepsis/septic shock, 2322t, 2323–2324
  - skin manifestations of, 407
  - treatment of, 934
- Distal cholangiocarcinoma, 667f, 669
- Distal convoluted tubule, 2367f, 2368t, 2370
- Distal esophageal spasm, 295, 2508–2509, 2509f
- Distal myopathies, 3640t, 3641f, 3643–3644
- Distal weakness, 171
- Distraction (gapping test), 121t
- Distribution, of drugs, 486, 1167
- Distributive shock
  - hemodynamic characteristics of, 2311t
  - history in, 2312
  - pathophysiology of, 2310–2311
  - physical examination in, 2312
  - treatment of, 2315
- Disulfiram, 228, 1744t, 3609t, 3681
- Ditans (5-HT_{1F} receptor agonists), 3467t, 3469
- Diuresis
  - for hypercalcemia, 3284, 3284t
  - postobstructive, 2457
  - water, 3015, 3015f
- Diuretics. *See also* Loop diuretics; Thiazide diuretics
  - abuse of, 356
  - actions of, 347, 2144t
  - adverse effects of, 246
    - acid-base disorders, 374
    - gout, 2940t
    - hypokalemia, 355, 356
    - hyponatremia, 347
    - hypovolemia, 346
    - renal, 340
    - xerostomia, 266
  - aldosterone-renin ratio effects of, 3062t
  - for ascites, 329
  - for cardiac amyloidosis, 2026
  - for chronic venous disease, 2182
  - for heart failure, 1990–1991, 1993t, 1996, 2123
  - for hyperkalemia, 361
  - for hypertension, 2143, 2144t
  - for hypertension in CKD, 2147
  - for pulmonary edema, 2331
  - for vertigo, 165t
- Diurnal rhythm, 3944t
- Divalent metal transporter 1 (DMT1), 762, 763, 3334f
- Divalproex sodium. *See* Valproate/valproic acid
- Divergent phenotype, 31, 31f
- Divers Alert Network, 3725, 3768
- Diversion colitis, 2561
- Diverticular disease, **2578**
  - anatomy of, 2578–2579
  - bleeding in, 318, 2579
  - clinical features of, 2579, 2579t
  - diagnosis of, 2484, 2491f, 2579
  - epidemiology of, 2578

    esophageal, 2506, 2506f
    genetic considerations in, 2579
    jejunal, 2544f
    pathophysiology of, 2578–2579, 2578f
    recurrent, 2581
    staging of, 2579, 2579f
    treatment of
        endoscopic, 2491f
        medical, 2579–2580, 2580t
        surgical, 2580–2581, 2581t, 2582f, V5
    Zenker's. *See* Zenker's diverticulum
Diverticulitis
    clinical features of, 112t, 150, 306, 2579, 2579t
    definition of, 2578
    diagnosis of, 2579
    Hinchey classification of, 2579, 2580f
    vs. IBD, 2561
Diving medicine, **3762**
Dix-Hallpike maneuver, 163
Dizziness, 49, **162**. *See also* Vertigo
DKA. *See* Diabetic ketoacidosis (DKA)
DKC1 gene mutations, 810, 3818, 3820t
DKK-1, 2846
DLBCL. *See* Diffuse large B-cell lymphoma (DLBCL)
DL$_{CO}$ (diffusion capacity of the lung for carbon monoxide), 2206f, 2207
DLMO (dim-light melatonin onset), 218, 3944t, 3946
DLT (dose-limiting toxicity), 555
DM. *See* Dermatomyositis (DM); Diabetes mellitus (DM)
DMAA (1,3-dimethlamylamine), 3926
DMARDs (disease-modifying antirheumatic drugs), 2849, 2850–2851t, 2851–2852
DMD gene mutations, 2004t, 2008, 3785, 3787–3788
DMF (dimethyl fumarate), 3586t
DMPK gene mutations, 2004t
DMSA (succimer), 3700–3701t
DMSO (dimethyl sulfoxide), 336
DMT (dimethyltryptamine), 3696
DMT1 (divalent metal transporter 1), 762, 763, 3334f
DMXL2 gene mutations, 3112t
DNA
    cell-free. *See* Cell-free DNA (cfDNA)
    circulating tumor. *See* Circulating tumor DNA (ctDNA)
    demethylation of, 539
    epigenetic modifications of, 3782–3783, 3784f
    hypermethylation of, 538
    instability of, aging and, 3869–3870
    methylation of, 3782, 3784f, 3785f, 3939, 3941
    mitochondrial. *See* Mitochondrial DNA (mtDNA)
    repair, 430
    repair defects, 2799–2800, 3785–3786
    repetitive, 3779
    replication of, 3779–3780, 3780f
    structure of, 3779, 3779f
    unstable sequences, 3786
DNA-dependent nucleoside-stimulated ATPase, 2790t
DNA gyrase, 1167, 1185
DNA-interactive agents, 561, 562t, 563. *See also specific drugs*
DNAJB11 gene mutations, 2431
DNAJC12 deficiency, 3368t
DNA methylation inhibitors, 3940
DNA methyltransferase inhibitors, 538f
DNA repair, 3785–3786
DNA repair defects, S9
DNA repair pathway modulators. *See* PARP (poly-ADP-ribose polymerase) inhibitors
DNA sequencing, 3796, 3998
DNA topoisomerase II, 2790t
DNA vaccines, 997t
DNA viruses, 1478f, 1479–1481, 1480t. *See also specific viruses*
DNMT1, 3935–3936, 3939
DNMT3A/3B, 3935
DNMT3a gene mutations, 761
DNMT3A gene mutations, 812, 825, 877
DNMT3B gene mutations, 2186
Dobrava virus, 1661t, 1676
Dobutamine
    for cardiogenic shock, 2315, 2328
    for heart failure, 1992, 1993t
    for pulmonary edema, 2331
    in stress testing, 1878
DOC (deoxycorticosterone), 3051f, 3054, 3061
Docetaxel
    actions of, 565
    adverse effects of, 563t, 565, 753t, 2940t, 3608t
        desensitization for, 2818
        premedication for, 563t, 565
        pulmonary, 595
    for bladder cancer, 693
    for breast cancer, 637
    drug interactions of, 1742t
    for head and neck cancer, 613
    interactions and issues, 563t
    for lung cancer, 624, 627t, 629, 630
    for prostate cancer, 702, 703
    for soft tissue sarcomas, 729
DOCK8 deficiency, 459t, 463
Docosanol, 1485
Doctor, The, 1f
Dofetilide
    actions of, 1916t
    adverse effects of, 1916t, 1952, 1958
    for atrial fibrillation, 1952
    for atrial flutter, 1944
    dosage in kidney disease, 488
    dosage monitoring, 487
    drug interactions of, 1742t
    for ventricular arrhythmias, 1958
Dog(s)
    bite-wounds, 993, **1142**
        antibiotic prophylaxis for, 1145–1146
        approach to the patient, 1144
        cellulitis in, 1052
        epidemiology of, 1142–1143
        infections from, 1143, 1267
        microbiology of, 1143
        rabies prophylaxis for, 1146
        treatment of, 1054t, 1144–1145, 1145t
    filarial infections, 1821–1822
    fleas, 3736
    tapeworms, 1839
    ticks, 1456
Döhle bodies, 442f, 443, 453, 455f
Dolasetron, 83, 575
"Doll's eyes," 190
Dolor, 454f
Dolutegravir, 1402, 1617f, 1618t, 1621t, 1623f
Domestic abuse/violence
    human bite-wounds in, 1143
    incidence of, 3168, 3674
    lifetime cumulative risk of, 37t
    mandated reporting of, 71
    screening for, 39t, 3674
Domoic acid intoxication (amnesic shellfish poisoning), 3414, 3727–3728
Domperidone, 116, 299, 299t, 2874, 3469, 3549
Donanemab, 199–200, 3480
Donath-Landsteiner antibodies, 802
Donepezil, 178, 199, 243, 1920t, 3480, 3491, 3493
Donor after cardiac death, 2285
Do-not-resuscitate (DNR) orders, 89
Donovan bodies, 1357, 1357f
Donovanosis, **1356**
    clinical features of, 1108t, 1357, 1357f
    control and prevention of, 1358
    diagnosis of, 1357, 1357f
    epidemiology of, 1107, 1357
    etiology of, 1289, 1356
    treatment of, 1357–1358, 1358t
"Don't miss" diagnoses, 57–58, 59t
L-Dopa. *See* Levodopa
Dopamine
    deficiency of, in PD, 3500
    for heart failure, 1992
    for hypothermia, 3771
    in migraine, 3462
    in PD, 3500
    in placebo effect, 3933
    for pulmonary edema, 2331
Dopamine-2 receptor antibodies, 748
Dopamine agonists
    for acromegaly, 3010
    adverse effects of, 217, 3007, 3502–3503
    for PD, 3501–3502, 3504t
    for prolactinoma, 3006–3007
    for restless legs syndrome, 217
Dopamine receptor antagonists, 299, 299t, 3467t, 3469
Dopa-responsive dystonia (Segawa syndrome), 3368t, 3497, 3511
Doppler ultrasound, 1876–1877, 1877f. *See also* Echocardiography
Doravirine, 1621t, 1622f
Doripenem
    for anaerobic bacterial infections, 1379
    indications for, 1175
    for intraabdominal infections, 963t
    for *P. aeruginosa* infections, 1309t
    resistance to, 1181
Dormancy, in cancer cells, 556
Dornase, 2244
Dorsal midbrain syndrome (Parinaud's syndrome/complex), 221, 236, 721, 1353, V3
Dorsal respiratory group, 2273
Dorsal root ganglia, 93, 93f, 94, 172
Dorsal root ganglionopathy, 750
Dose-limiting toxicity (DLT), 555
Dostarlimab, 534t, 559f, 703, 753t, 2787t
Dosulepin, 3470t
DOT strategy, for TB, 1396, 3848
Double-balloon endoscopy/enteroscopy, 2462, 2462t, 2468, 2471f, 2472f, 2501f, V5
Doublecortin, 3412t
Double equipoise, 665
Double minutes, 521
Double-prion diseases, 3401, 3401t
Double-stranded DNA viruses, 1479
Double-stranded RNA viruses, 1479
Double vision. *See* Diplopia
Downbeat nystagmus, 237
Down's syndrome (trisomy 21)
    AD in, 3401, 3479
    ALL in, 844
    AML in, 826
    autoimmune polyendocrine deficiencies in, 3093
    clinical features of, 2183
    hypothyroidism in, 3029
    MDS in, 816
    noninvasive prenatal testing for, 3988
    oral manifestations of, 261
    transient myeloproliferative disorder in, 879
Doxazosin, 2145t, 2178
Doxepin
    adverse effects of, 3659t
    for depression, 3659t
    for insomnia, 216
    overdosage/poisoning with, 3711t, 3712t

- Doxepin (*Cont.*):
  - for pain, 97t
  - for pruritus, 383
  - for urticaria, 2809
- Doxercalciferol, 3271
- Doxorubicin
  - actions of, 564
  - adverse effects of, 562t, 564
    - acral erythema, 421
    - cardiotoxicity, 564, 754, 2016
    - desensitization protocol for, 2819
    - myelosuppression, 564
    - pigmentation changes, 420
  - for breast cancer, 637, 638
  - for DLBCL, 863
  - for endometrial cancer, 715
  - for Ewing's sarcoma, 730
  - for gastric lymphoma, 650
  - for HCC, 665
  - for hepatoblastoma, 670
  - for insulinoma, 3082
  - for osteosarcoma, 730
  - for ovarian cancer, 712
  - for salivary gland tumors, 614
  - for soft tissue sarcomas, 729
  - for thymic tumors, 632
  - timing of administration of, 3954
- Doxycycline
  - for acne vulgaris, 391
  - for actinomycosis, 1366t
  - actions of, 1166, 1177, 1182t
  - adverse effects of, 400, 1177, 1771t, 2703t
  - for anthrax, S4
  - for bacillary angiomatosis/peliosis, 1054t, 1353t
  - for *Bartonella* endocarditis, 1353t
  - for *Bartonella* prophylaxis, 1594t
  - for bite-wound infections, 1054t, 1145, 1145t
  - for *Brucella* infections, 1334
  - for bullous pemphigoid, 412
  - for cat-scratch disease, 1353t
  - for chlamydial infections, 1259, 1259t
  - for cholera, 1329
  - for community-acquired pneumonia, 1027t
  - for *C. pneumoniae* infection, 1477
  - for *C. trachomatis* infections, 1474, 1475
  - for *C. trachomatis* prophylaxis, 1474
  - for donovanosis, 1358t
  - for ehrlichioses, 1461
  - for endemic treponematoses, 1439
  - for epidemic typhus, 1458
  - for epididymitis, 1099
  - for gonococcal arthritis, 1058
  - for hard tick relapsing fever, 1448
  - for human granulocytotropic anaplasmosis, 1462
  - indications for, 1173t, 1177
  - for infective endocarditis, 1044t
  - for leptospirosis, 1444, 1444t
  - for Lyme disease, 1058, 1453, 1453f
  - for Lyme disease prophylaxis, 1453, 3731
  - for lymphatic filariasis, 1821
  - for malaria, 1770t, 1771t
  - for malaria prophylaxis, 1774t, 1775
  - for *M. perstans* infection, 1824
  - for MRSA infections, 383, 1203t, 1204
  - for mucopurulent cervicitis, 1104
  - for *Mycoplasma* infections, 1466, 1466t, 1467
  - for onchocerciasis, 1823
  - for pelvic inflammatory disease, 1106, 1106t, 3137
  - for plague prophylaxis, 1347t
  - for plague treatment, 1346, 1346t, S4
  - for proctitis, 1109
  - prophylactic uses of, 1180t
  - for psittacosis, 1476
  - for Q fever, 1464
  - for relapsing fever, 1447, 1447f
  - resistance to, 1173t, 1182t, 1184, 1359f
  - for rickettsial disease, 992
  - for rickettsialpox, 1458
  - for RMSF, 991t, 992, 1457
  - for scrub typhus, 1460
  - for sinusitis, 256
  - for syphilis, 1434t, 1435
  - for syphilis prophylaxis, 1429
  - for tick-borne spotted fevers, 1458
  - for trench fever, 1353t
  - for tularemia, 1339, 1340t
  - for tularemia prophylaxis, 1340t
  - for tularemia treatment, S4
  - for *Ureaplasma* infections, 1467
  - for Whipple's disease, 1370
- Doxylamine, 300, 3711t
- *DPD* gene mutations, 658
- DPP-IV (dipeptidyl peptidase-4) inhibitors, 412, 420, 3211t, 3213, 3882t
- DPPX antibodies, 744t, 748–749
- *DPYD* gene mutations, 641, E72t
- *Dracunculus medinensis*/dracunculiasis (guinea worm infection)
  - clinical features of, 1051, 1059, 1738t, 1825
  - diagnosis of, 1825, S12
  - etiology and epidemiology of, 1825
  - treatment of, 1752, 1825
- Dravet's syndrome, 3412t
- Drawer sign, 2944
- DRESS (drug reaction with eosinophilia and systemic symptoms). *See* Drug-induced hypersensitivity syndrome (DIHS)
- Dressing apraxia, 206
- Dressings, for chronic venous disease, 2181
- Driver mutations, 526, 617, 617f
- Driving ability/privileges, 200, 209, 3427, 3877–3878, 3879t
- *DRK,* 1701
- Dronabinol, 298, 509, 3688
- Dronedarone
  - actions of, 1916t
  - adverse effects of, 1920t, 1952
  - for arrhythmias, 1952, 1958, 1998
  - drug interactions of, 1422
- Drop arm test, 2974
- Drop attacks, 181. *See also* Fall(s)
- Droplet precautions, 1148, 1538
- Dropped head syndrome, 3634
- Drospirenone, 359, 3141
- Drowning, A12
- Drowsiness, 81, 186
- Droxidopa, 160, 3227, 3547
- Drp1, 3400
- Drug(s)
  - adherence to regimens for, 3920–3921
  - adverse reactions to. *See* Adverse drug reactions; Cutaneous drug reactions; Drug-induced illness
  - benefits of, 483
  - bioavailability of, 484
  - clearance of, 486
  - development of
    - for cancer, 499f, 555–556
    - genetics and, 494–495
    - genomics in, 983–984
    - systems biology and, 3957, 3961t
    - target molecules in, 490
  - distribution of, 486
  - dose-response relations, 483, 483f
  - dose selection, **487**
  - elimination of, 486
  - failure of efficacy of, 487
  - genetic determinants of response to, **491**
  - half-life of, 485, 485f
  - indications for, 483
  - interactions between. *See* Drug interactions
  - intravenous administration of, 486
  - medication errors, 4–5
  - metabolism of
    - active metabolites in, 486
    - enzymes in, 1167
    - first-pass effect, 484, 484f, 1167
    - genetic variations in, 491, 492f, E72t
    - in men vs. women, 3166–3167
  - microbiome and, 3834
  - names of, 3933
  - network-based precision repurposing of, 32, 34f
  - new, skews in research for, 3862
  - overdosage of. *See* Poisoning/drug overdose
  - plasma concentration of, 484, 484f, 487f, 489
  - prescribing for older adults, 3884–3885, 3887t, 3888f
  - prescription, use in U.S., 416
  - risks of, 483
  - slow-release formulation of, 484
  - sustained-release formulation of, 484
  - time-of-day effects on, 219, 3954
- Drug allergy/hypersensitivity, 2815, 2816f. *See also* Cutaneous drug reactions
- Drug-dependent antibodies, 922–923
- Drug desensitization, **2814**
  - for antibiotics, 2818
  - breakthrough reactions and outcomes, 2815
  - for chemotherapeutic agents, 2817f, 2818–2819
  - contraindications to, 2818f
  - definitions and mechanisms of, 2815
  - impact of, 2819
  - indications for, 428–429, 2814, 2816–2818, 2816f, 2818f
  - for monoclonals/biologicals, 2817f, 2819
  - for NSAIDs, 2819
  - principles of, 2818f
  - protocols for, 2815–2816, 2817f
  - for small molecules, 2819
- Drug disposition. *See* Drug(s), metabolism of
- Drug dosage
  - dosing interval, 485
  - in heart failure and shock, 488
  - in kidney disease, 488
  - in liver disease, 488
  - loading dose, 485f, 486
  - maintenance dose, 487
  - in older adults, 488
  - plasma concentration as guide to therapy, 487
  - selection of, 487
- Drug-eluting bead chemoembolization (DEB-TACE), 666
- Drug fever, 409
- Drug history, 488
- Drug-induced hypersensitivity syndrome (DIHS)
  - clinical features of
    - erythroderma, 393
    - fever and rash, 142t, 147, 150, 423, 423f, A1
    - frequent signs and symptoms, 427t
    - systemic, 423
  - diagnosis of, 424, 426t
  - drugs associated with, 423–424, 427t
  - etiology of, 142t, 393, 423–424
  - genetic considerations in, 419, 493
  - in HIV infection, 1603
  - pathogenesis of, 417t, 418f, 419, 423
  - time of onset, 423–424, 427, 428f
  - treatment of, 424. *See also* Drug desensitization
- Drug-induced illness. *See also* Adverse drug reactions; Poisoning/drug overdose
  - anaphylaxis, 422
  - chronic meningitis, 1132t
  - cutaneous. *See* Cutaneous drug reactions
  - fatigue, 166
  - fever of unknown origin, 150

  gingival overgrowth, 261
  liver injury. *See* Liver disease/failure, drug-induced
  lupus. *See* Lupus erythematosus, drug-induced
  movement disorders, **3516**
  musculoskeletal conditions, 2940t
  neutropenia, 456–457, 457t
  primary immunodeficiencies, S9
  tendinopathies, 2974
  timeline of, 427, 428f
  vasculitis, 2940t
  in women vs. men, 3166–3167
Drug-Induced Liver Injury Network (DILIN), 2668
Drug interactions, 488, **488**. *See also specific drugs*
  with herbal/dietary supplements, 3926–3927, 3927t
  pharmacodynamic, 489t
  pharmacokinetic, 489t
Drug reaction with eosinophilia and systemic symptoms (DRESS). *See* Drug-induced hypersensitivity syndrome (DIHS)
Drug-receptor interaction, 486
Drug-target network analysis, 3957
Drusen, 228, 228f, 231, 231f
Dry beriberi, 2607, 3612
Dry eye. *See* Keratoconjunctivitis sicca (dry eye)
Dry mouth. *See* Xerostomia
"Dry tap," 873, 873t
DSDs. *See* Disorders of sex development (DSDs)
DSE gene mutations, 3329t
DSG1/2 gene mutations, 2004t
DSHEA (Dietary Supplements Health and Education Act), 3926
DSP gene mutations, 2004t
DTaP. *See* Diphtheria-tetanus-pertussis vaccine (DTaP)
DTIC. *See* Dacarbazine (DTIC)
DTPA, S6
Dual-energy x-ray absorptiometry (DEXA), 2618t, 2717, 3277, 3296–3297
Dual-process theory, 20
Dual-task walking, 177
Dubin-Johnson syndrome, 323, 2638, 2641–2642, 2641t, 2727
Duchenne muscular dystrophy, 3638–3639, 3641f
Duchenne's muscular dystrophy, 3787–3788, 3804t, 3823, 3824t
Duck fever, 2228t
Ductal carcinoma in situ (DCIS), 636
Duffy RBC group system, 904, 904t
Dugbe virus, 1661t
Duke criteria, for infective endocarditis, 1039, 1040t
Dulaglutide, 3211t, 3213
Duloxetine
  adverse effects of, 3604t, 3659t
  for depression, 85, 3659t, 3667
  for fibromyalgia, 2965, 2966t
  for low back pain, 126
  for neuropathic pain, 100, 506, 508, 3227, 3604t
  in older adults, 3883
  for osteoarthritis, 2955
  for pain, 97t, 100, 3588
Dumping syndrome, 2533
Duodenal ulcers
  clinical features of, 2522
  complications of, 2522–2523
  diagnosis of, 2465, 2467f, 2484f, 2523, 2523–2524f
  differential diagnosis of, 2523
  epidemiology of, 2518
  gastroduodenal mucosal defense and, 2517–2518
  H. pylori and. *See Helicobacter pylori* infection
  NSAID-induced, 2530–2531, 2530t
  pathology of, 2518
  pathophysiology of, 2518–2519
  recurrence of, 2533
  refractory, 2531
  treatment of, 2481, 2484–2485f, 2530–2531, V5. *See also* Peptic ulcer disease, treatment of
Duodenitis, 318
Duodenoscope, infections related to, 1152
DUOX, 3023, 3028t
Dupilumab, 383, 2226, 2514, 2792t
Duplication cyst, 318
Dupuytren contracture, 2630
Durable power of attorney for health care, 79
Dural sinus thrombosis, 3452, 3453t
Dura matter grafts, 3529
Dura, vascular malformations of, 3453t, 3457, 3567, 3567f
Durlobactam, 1175, 1176, 1183, 1296, 1297t
Duroziez's sign, 2044
Durvalumab
  action and targets of, 559f, 2787t
  adverse effects of, 596, 753t, 2351
  with bevacizumab, 548
  for cholangiocarcinoma, 668, 669
  for gallbladder cancer, 670
  for HCC, 664t, 665f, 666
  indications for, 2787t
  for non-small-cell lung cancer, 626, 630t
  for small-cell lung cancer, 630
DUSP6 gene mutations, 3112, 3112t
Dutasteride, 513, 702f, 3141, 3175
Duvelisib, 571t, 574, 864
Duvenhage virus, 1655
DUX4 gene mutations, 3641–3642
DVT. *See* Deep venous thrombosis (DVT)
Dwarf tapeworm *(Rodentolepis nana),* 961t, 1750, 1838, S12
DXM, 3696
Dynamic mutation, 3786
Dynamic visual acuity, 163, 164
Dysbetalipoproteinemia, familial, 3241t, 3245
Dysbiosis, 3871
Dyschondroplasia (Ollier's disease), 3318
Dyschromic macules, 1439
Dysentery
  in amebiasis. *See* Amebiasis
  etiology of, 305, 1076–1077
  in *Shigella* infections. *See Shigella* spp. infections
  treatment of, 307, 1081
  V. parahaemolyticus, 1330
Dysesthesia, 172, 175. *See also* Sensation, abnormalities of
Dysferlinopathies (LGMD2B/DR2), 3641
Dysfibrinogenemia, 932, 935
Dysgerminoma, 712
Dysgeusia. *See* Taste disorders
Dysgraphia, 201, 203
Dyshidrotic eczema, 378t, 380f, 384, 384f, A5
Dyskeratosis congenita (Hoyeraal-Hreidarsson syndrome)
  clinical features of, 399, 531, 606, 810, 2800, 3819
  genetic considerations in, 810, 2800, 3819, 3820t, S9
  pathogenesis of, 531
  telomere length measurement in, 3819, 3821f
Dyskerin, 810, 3818, 3819f
Dyskinesia(s)
  diphasic, levodopa-related, 3502
  exercise-induced, 3517
  paroxysmal, 3517
  pathophysiology of, 3500f
  tardive, 178, 299, 3516–3517, 3674
Dyskinetic gain, 177t
Dyslipidemias. *See* Lipoprotein disorders
Dysmenorrhea, 3137
Dyspareunia, 3161
Dyspepsia. *See also* Indigestion
  clinical features of, 2490
  endoscopy in, 2490
  functional, 300, 301, 1300, 2523
  new-onset, approach to, 2531, 2531f
  nonulcer, 2491–2492
Dysphagia, **293**
  approach to the patient, 295–296, 295f
  in botulism, 1235
  in cranial nerve disorders, 3556–3557
  diagnosis of, 296, 2492–2493, 2496f
  esophageal, 294–295, 2504
  in esophageal cancer, 642–643
  in gastric cancer, 646
  in GERD, 2509
  history in, 295
  involuntary weight loss and, 315
  in iron-deficiency anemia, 765
  in myasthenia gravis, 3626
  oropharyngeal, 294
  pathophysiology of, 293–294
  physical examination in, 295–296
  propulsive (motor), 293
  radiation therapy-related, 614
  with solids vs. liquids, 295
  structural, 293
  in terminally ill patient, 91t
  in tetanus, 1231
  transfer, 2492
  treatment of, 296, 2492–2493, 2496–2499f, V5
Dysphonia, 3556–3557
Dysplasia(s)
  of bone. *See* Bone disease
  fibromuscular, 2150, 2174, 3447
  pituitary, 2992
  skeletal (chondrodysplasias), **3330**
Dysplastic nevus, 600, 602f, A5
Dyspnea, **267**
  after transfusion, 908t
  in aortic stenosis, 2037
  approach to the patient, **269,** 270f
  assessment of, 83t, 268, 268t, 2199
  in cancer patient, 507t, 509–510
  in cardiovascular disease, 269, 269t, 271, 1841
  in COPD, 2254
  definition of, 267
  differential diagnosis of, 268–269, 269t, 1986t
  efferent-afferent mismatch in, 267, 268f
  epidemiology of, 267
  in heart failure, 1982
  history in, 270
  imaging in, 270–271
  in interstitial lung disease, 2261
  laboratory studies in, 271
  nocturnal, 270
  in orthostatic hypotension, 159
  pathophysiology of, 267–268, 268f
  physical examination in, 270
  in pulmonary hypertension, 2188
  in respiratory acidosis, 375
  in respiratory disease, 268–269, 269t, 271, 2199
  in terminally ill patient, 83, 83t
  treatment of, 83, 83t, 271
Dyspraxia, sympathetic, 203
Dysreflexia, autonomic, 3549
Dysthymic disorder, 3665
Dystonia, **3509**
  clinical features of, 178, 3509–3510, 3509t
  combined, 3510t, 3511
  complex, 3511
  dopa-responsive (Segawa syndrome), 3368t, 3497, 3511
  drug-induced, 86, 3511, 3516

Dystonia (*Cont.*):
  focal, 3510
  generalized, 3510–3511
  genetic considerations in, 3510–3511, 3510t
  isolated, 3510t
  pathophysiology of, 3512
  tardive, 3516–3517
  treatment of, 3512
  in Wilson's disease, 3338
Dystonic storm, 3512
Dystroglycans, 3640, 3641f
Dystrophia myotonica. *See* Myotonic dystrophy
Dystrophic calcification, 403, 3318t, 3319
Dystrophin, 2005f, 2008–2009, 3639, 3641f, 3787–3788
Dystrophin gene mutations, 3638–3639

E
EACA. *See* E-Aminocaproic acid
EADs (early afterdepolarizations), 1913, 1913f, 1913t
EAEC (enteroaggregative *E. coli*), 1285t, 1288. *See also Escherichia coli* infections, intestinal
EAP (extracellular adherence protein), 1198
Ear. *See also* Hearing loss
  anatomy and physiology of, 244, 245f
  infections of
    otitis externa, 1309t, 1311, 1715
    otitis media. *See* Otitis media
    P. aeruginosa, 1311
  in relapsing polychondritis, 253, 2918, 2918–2919f
Early afterdepolarizations (EADs), 1913, 1913f, 1913t
Early repolarization syndrome, 1971
Ear pain, 253. *See also* Otitis externa; Otitis media
Eastern Cooperative Oncology Group (ECOG), 84, 502, 503t, 557
Eastern equine encephalitis, 1657t, 1673
Eating disorders
  anorexia nervosa, 298, **3671**
  avoidant/restrictive food intake disorder, 3671
  binge-eating disorder, 3671
  bulimia nervosa, 261, 298, 356, 3671
  constipation in, 313
  diarrhea in, 310
  hypokalemia in, 356
  pica, 3670
  rumination disorder, 298, 3670–3671
  sleep-related, 3951
Eaton-Lambert syndrome. *See* Lambert-Eaton myasthenic syndrome (LEMS)
Ebola virus disease. *See also* Filovirid infections
  clinical features of, 993, 1684f, 1685, S4
  complications of, 1686
  deaths from, 3849
  epidemiology of, 985, 986f, 1680–1681, 1682–1683f
  global considerations in, 3849–3850
  prevention of, S4
  prognosis of, 1686
  treatment of, 1685–1686, S4
Ebola virus vaccine, 1687
Ebstein's anomaly, 1941, 2071, 2071f
Eburnation, 1860
EBV infections. *See* Epstein-Barr virus (EBV) infections
E-cadherin, in gastric cancer, 646
Ecallantide, 2809
Ecamsule, 435, 436t
Ecchymoses, 137, 897, 897f
ECCO$_2$ removal (extracorporeal CO$_2$ removal), 2309, 2309t
ECFa (eosinophil chemotactic factor of anaphylaxis), 462
ECG. *See* Electrocardiography (ECG)
Echarate virus, 1662t
Echinocandins, 1692, 1711–1712, 1711t, 1712t, 1720–1721, 1721t, 2513
Echinococcosis (hydatid disease)
  clinical features of, 1739t, 1836, 2737, S7
  diagnosis of, 1836–1837, 1836–1837f, S12
  epidemiology of, 1736, 1835–1836, S7
  etiology of, 1836, S12
  incubation period for, S7
  joint involvement in, 1059
  prevention of, 1838, S7
  staging of, 1836, 1837f
  treatment of, 1837–1838, S7
Echinococcus granulosus, 1736, 1835–1836, S12
Echinococcus multilocularis, 1736, 1835–1836, S12
Echinocytes, 437, 440f, 446f
Echinostoma spp., 1826t, 1830t, 1831t
Echocardiography, **1876**
  after STEMI, 1893, 1894f
  in aortic dissection, 2170
  in arrhythmia evaluation, 1915
  in atrial septal defect, 1898–1899, 1900, 1901f, A11
  in cardiac amyloidosis, A9, A9v
  in cardiac trauma, 2088, 2089f
  in cardiac tumors, 1898, 1900f, 2085–2087, 2085–2087f, 2085t
  in cardiogenic shock, 2326, 2329t
  in chest pain evaluation, 109, A9, A9v
  contrast agents in, 1883–1884
  in dilated cardiomyopathy, 2009f
  Doppler, 1876–1877, 1877f
  in dyspnea evaluation, 271
  in heart failure, 1985
  in heart murmurs, 290, 1843
  in hypertrophic cardiomyopathy, A9, A9v
  in infective endocarditis, 1040–1041, 1042f, 1898–1899, 1901f
  in multiple/mixed valvular heart disease, 2066
  in myocardial function assessment, 1852–1853
  in penetrating cardiac injury, 2088, 2089f
  in pericardial effusion, 1895–1896, 1898f, 2079, 2080f, A9
  POCUS, 1878, 1878f, 3911, 3912t, 3914v
  principles of, 1876–1878, 1877f
  in pulmonary edema, 2329t
  in pulmonary embolism evaluation, 2161
  in pulmonary hypertension, 2189, 2190f, A9, A9v
  in pulsus paradoxus, A9
  in rheumatic heart disease, 2856f, 2856t, 2857t, 2859v
  in shock, 2314
  in STEMI, 2116–2117, A9, A9v
  stress
    advantages of, 1884–1885
    in CAD evaluation, 1884–1885, 1902v, A9, A9v
    in chest pain evaluation, 108–109
    indications for, 1878
    in ischemic heart disease evaluation, 2095, 2095f
    for risk stratification, 1888, 1889f, 3906–3907
  in syncope evaluation, 161
  three-dimensional, 1876, 1877f, 2044
  transesophageal. *See* Transesophageal echocardiography (TEE)
  two-dimensional, 1876, 1877f
  in valvular heart disease, 1890
    aortic regurgitation, 1891–1892, 1891f, 2044, 2045f
    aortic stenosis, 1890–1891, 2038, 2038t, A9, A9v
    bicuspid aortic valve, 1891f
    mitral regurgitation, 1892, 1892f, 2044, 2045f, 2053, 2054f, 2859v
    mitral stenosis, 1892, 1893f, 2049, 2049f, 2859v
    mitral valve prolapse, 1892, 1893f, 2058, 2058f
    pulmonic regurgitation, 2064, 2064f
    pulmonic stenosis, 2063
    tricuspid regurgitation, 2060f, 2062
    tricuspid stenosis, 2060
  in ventricular septal defect, 1900
Echolalia, 201t, 202, 3515
Echo-planar MRI, 3390
Echopraxia, 3515
Echoviruses, 1535, 1632
Echovirus infections, 137, 138t, 148, 402, 1131t, 1535, 1634
E-cigarettes, 615, 3683–3685, 3685t
Eclampsia, 2140, 3899
ECM (extracellular matrix), 966
ECMO. *See* Extracorporeal membrane oxygenation (ECMO)
ECOG (Eastern Cooperative Oncology Group), 84, 502, 503t, 557
Econazole, 390
Economics
  behavioral. *See* Behavioral economics
  classical, 3915–3916, 3916t
  health and, 3845
  in medical care, 23
eCOT-NS (external concurrent occipital and trigeminal neurostimulation), 3467t
ECQMs (electronic clinical quality metrics), 45
Ecstasy (MDMA), 184, 349, 3693, 3696
Ectasias, vascular. *See* Vascular ectasias
Ecthyma, 389
Ecthyma gangrenosum
  in cancer patient, 583
  clinical features of, 143t, 147, 1308, A5
  epidemiology of, 143t, 147
  etiology of, 143t, 992, 1052, 1268, 1308
  skin manifestations of, 143t, 408, 992, 1308f, A1
Ectopia lentis, A15
Ectopic lymphoid follicles, 3581
Ectopic ossification, 3318t, 3319
Ectopic (tubal) pregnancy, 112t, 113, 3136
Ectopy, cervical, 1104
Ectropion, 224, 2585
Eculizumab
  actions of, 2792t
  for antiphospholipid antibody syndrome, 2446, 2839
  for C$_3$ glomerulopathy, 2420
  for cardiac transplant rejection, 2030
  for familial hemolytic-uremic syndrome, 2445
  for familial (atypical) hemolytic-uremic syndrome, 800–801
  for hemolytic-uremic syndrome, 595, 2428
  for myasthenia gravis, 3631, 3632t
  for neuromyelitis optica, 227, 3565, 3592, 3592t
  for paroxysmal nocturnal hemoglobinuria, 805–806, 805f, 806f, 2792t
  for SLE, 2835
ECV (epidemiologic cutoff value), S11
Eczema, **382**. *See also* Atopic dermatitis
  asteatotic, 380f, 384
  clinical features of, 382
  dyshidrotic, 378t, 380f, 384, 384f, A5
  hand, 380f, 384, 384f
  nummular, 385
Eczema herpeticum, 1498, A1
ED. *See* Erectile dysfunction (ED)
EDACS score, 108f
Edaravone, 3524
EDD (extended daily dialysis), 2385

Edema, **275**
  approach to the patient, 282
  in cardiovascular disease, 1841
  in cirrhosis, 281, 281t
  clinical features of, 281
  differential diagnosis of, 1986t
  distribution of, 282
  drug-induced, 282, 282t
  etiology of, 281–282, 281t
  in heart failure, 281, 281t, 1982
  heat-related, 3773–3774
  hypoalbuminemia and, 342
  in kidney disease, 281, 281t
  localized, 282
  localized hypoxia and, 277
  lower extremity, 1860
  in nephrotic syndrome, 281, 281t
  of nutritional origin, 282
  pathogenesis of, 279–281, 280f, 2371f
  pitting, 281, 1860
  in pneumonia, 1023
  pulmonary. *See* Pulmonary edema
  refeeding, 282
  stasis dermatitis and, 385
  vasogenic, 2349
Edentulousness, 267
Edge Hill virus, 1659t
Edinger-Westphal nuclei, 220
Edmonton Symptom Assessment System, 506, 507f
Edoxaban
  dosing of, 953, S13
  indications for, 952
  management before lumbar puncture, S3
  management of bleeding during treatment with, 953
  monitoring of, 953
  for NSTEMI after PCI, 2110t
  pharmacology of, 953t
  for stroke prevention in atrial fibrillation, 1951, 1951t, 3450
  for VTE/PE treatment, 2162
Edrophonium, 3627, 3627t, 3721, 3721t
EDSS (Expanded Disability Status Score), 3580t, 3585
EDTA (ethylenediaminetetraacetic acid), 3702
^{153}Sm-EDTMP, 703
Edwardsiella tarda infections, 1281, 1294
EEG. *See* Electroencephalography (EEG)
Efavirenz, 485t, 981, 1585, 1617t, 1619f, 1621t, 1622f
Effector proteins, 967, 969
Efferocytosis, 973
Efgartigimod, 3622, 3631, 3632t
Eflornithine
  actions of, 1747
  adverse effects of, 1743t
  for human African trypanosomiasis, 1795, 1796t
  pharmacology of, 1747–1748
  in pregnancy and lactation, 1743t
EGF (epidermal growth factor), 2516, 2516f, 2981t, 3023
eGFR (estimated glomerular filtration rate), 2362, 2362t. *See also* Glomerular filtration rate (GFR)
EGFR (epidermal growth factor receptor), 615, 2539, 3004
EGFR gene mutations
  in bladder cancer, 691
  as drug target, 533t, 534t
  in head and neck cancer, 611
  liquid biopsy for detection of, 3982
  in lung cancer
    as driver mutation, 617, 617f
    as drug target, 567, 572, 627, 628f
    frequency of, 617t
    in never smokers, 615
    point mutation, 520t
  in primary brain tumors, 717
EGFR inhibitors
  actions of, 558, 567, 572
  adverse effects of, 395, 420, 421, 572, 755
  for colorectal cancer, 558
  for head and neck cancer, 613
  for lung cancer, 567, 572, 627, 627t
Egophony, 2200
EGPA. *See* Eosinophilic granulomatosis with polyangiitis
Egyptian rousettes, 1681, 1686
EHEC. *See* Enterohemorrhagic *E. coli* (EHEC)
Ehlers-Danlos syndromes, **3327**
  classification of, 3327, 3328–3329t
  clinical features of
    bleeding, 472
    cardiac, 2058, 2166, 2167
    cutaneous, 407, 3327
    ligament and joint changes, 50t, 3327
    vascular, 3330
    vascular disorders, 927
  diagnosis of, 3330
  genetic considerations in, 3328–3329t, 3330
  incidence of, 3327
  pregnancy in, 3900
  treatment of, 3330
Ehrlichia spp., 1460, S11
  E. chaffeensis, 139t, 1121, 1452, 1455t, 1460
  E. ewingii, 1460
  E. muris, 1460
Ehrlichioses, 1455t, 1460–1462, 1460f, 2013
EIA (enzyme immunoassay), 1589, S11
Eicosanoids, 432
EIEC (enteroinvasive *E. coli*), 1285t, 1288. *See also* *Escherichia coli* infections, intestinal
Eikenella corrodens, 1143, 1266, 1267t. *See also* Infective endocarditis, HACEK group
Eiken syndrome, 3274
Eisenmenger's syndrome
  clinical features of, 451
  hypoxia in, 277
  pathophysiology of, 2078
  pregnancy in, 3900
  prognosis of, 2078
  pulmonary arterial hypertension in, 2191
  treatment of, 2078
Ejaculation, 3155
Ejection fraction, 1852
Ekbom's syndrome, 3738
Ekiri syndrome, 1323
Ekpoma virus 1/2, 1662t
Ektacytometry, 796
Elacestrant, 640
Elagolix, 3137
Elamipretide, 3816
ELANE gene mutations, 457, 2795
Elapids, 3717, 3718. *See also* Snakebites
Elastase:antielastase hypothesis, emphysema, 2251
Elastase, stool, 2739t, 2742
Elastin, 2251, 3321t, 3323
Elastography, hepatic, 2631t, 2632, 2636, 2699, 2705
Elbasvir/grazoprevir, 1492t, 1493–1494, 2690
Elbow bursitis, 2973
Elbow neuropathy, 3613
ELC/MIP-3β, 2767t
Elder abuse and neglect, 3674, 3894, 3895t
Elderly. *See* Older adults
Eldocagene exuparvovec, 3824t
Electrical alternans, 1874, 2080f, 2081, 2087
Electrical injury, 2178, 3566
Electrical remodeling, 1913
Electrical storm, **1972,** 1972f, 1974f, 1975f
Electrocardiography (ECG), **1868, A7**. *See also specific disorders*
  ambulatory, 1915
  in chest pain evaluation, 107
  clinical interpretation of, 1842, 1876
  computerized, 1876
  in dyspnea evaluation, 271
  electrophysiology in, 1868, 1868f
  in heart failure, 1985
  intervals of, 1868–1869, 1868f
  leads of, 1869, 1869f
  normal, 1869–1870, 1870f, A7, A8
  in overdosage/poisoning, 3706
  P wave, 1869
  QRS complex, 1869–1870, 1870f
  in respiratory sinus arrhythmia, A8
  sensitivity and specificity of, 1868
  stress
    in arrhythmia evaluation, 1915
    in chest pain evaluation, 109, 1888
    contraindications to, 2094
    in heart failure evaluation, 1985
    in ischemic heart disease evaluation, 2094, 2095f
    pharmacologic, 2094
    protocols for, 2097t
    in syncope evaluation, 161
  ST segment elevations on, 1872, 1873–1874, 1874t
  T wave, 1870
  U wave, 1870
  waveforms of, 1868–1869, 1868f
Electrochemical potential, 2366
Electrocochleography, 249
Electroconvulsive therapy, 3667
Electrocorticography, 3968–3919
Electrodiagnostic studies
  electroencephalography. *See* Electroencephalography (EEG)
  evoked potentials, 3583–3584
  magnetoencephalography, 3418
  in myasthenia gravis, 3627
  in myopathies, 3637–3638
  nerve conduction studies. *See* Nerve conduction studies
  sudomotor axon reflex, 3597
  sudomotor function testing, 3542, 3545–3546f
Electroencephalography (EEG)
  in brain death, 192
  in coma, 191
  in Creutzfeldt-Jakob disease, 3531–3532
  in delirium, 185
  in dementia, 199
  in hypoxic-ischemic encephalopathy, 2346, 2347f
  in neurologic critical illness, 2344–2345
  in seizures, 3417–3418, 3417f, 3418f
  stereotactic, 3425
Electrogenic transport, 2366
Electrolysis, for hair removal, 3141
Electrolyte(s)
  for diarrhea, 307
  pancreatic secretion of, 2743
  in stool, measurement of, 2549–2550
Electrolyte disturbances. *See also specific conditions*
  delirium in, 184
  in enteral nutrition, 2623
  in heart failure, 1984
  in hypothermia, 3770
Electromyography (EMG)
  fibrillation potentials on, 169
  in inclusion body myositis, 2914
  in inflammatory myopathies, 2910
  in low back pain, 126
  in muscle disease, 3637–3638
  in myasthenia gravis, 3627, 3627t
  in neck pain, 130
  in peripheral neuropathy, 3597, 3597t

Electroneutral transport, 2366
Electronic cigarettes, 13, 511
Electronic clinical quality metrics (ECQMs), 45
Electronic medical records, 5, 40, 3898
Electrophoresis, of serum proteins, 883, 885f
Electrophysiologic studies
  cardiac. *See* Cardiac electrophysiology
  electromyography. *See* Electromyography (EMG)
  nerve conduction studies. *See* Nerve conduction studies
Elementary body, 1468, 1468f
Elephantiasis, 1820, 1821f. *See also* Lymphatic filariasis
Eletriptan, 3467t, 3468, 3468t
Elexacaftor, 2247
ELF (Enhanced Liver Fibrosis) test, 2632t, 2636, 2673, 2704
Eliglustat tartrate, 3356
Elimination
  of disease, 1002
  of drugs, 485, 1167
ELISA (enzyme-linked immunosorbent assay), 1484
Elivadogene autotemcel, 3824t
Elizabethkingia meningoseptica infections, 1268, 1268t
Elliptocytes, 437, 440f, 445
Elliptocytosis, 440f, 794f, 795
El Moro Canyon virus, 1661t
El Niño events, 1677, 3747, 3747f, 3750, 3753f
Elongin (pVHL), 3087
Elosulfase alfa, 3357
Elotuzumab, 534t, 559f, 890t, 892, 2792t
Elranatamab, 534t, 890t, 892
Elsberg's syndrome, 3566
Eltrombopag, 812–813, 924
Eluxadoline, 312, 2575, 2577t
Elvitegravir, 1617f, 1618t, 1621, 1621t, 1623f
Emapalumab, 466, 882, 1690
*emb*B gene, 1397, 1423
Embolectomy, 2164
Embolic stroke. *See* Ischemic stroke
Embolism
  in acute limb ischemia, 2175
  cerebral arterial gas, 3766
  cutaneous, 401, 408
  infectious, 408
  in infective endocarditis, 1039f, 1045–1046, 1047, A5
  pulmonary. *See* Pulmonary embolism (PE)
Embryologic development
  of blood vessels, 1844
  of cardiovascular system, 1843–1844, 1844f
  of kidney, 2363–2364, 2364f
Embryonal carcinoma, 705, 712, 737t, 3003
Emergency contraception, 3154
Emergent behavior, in biologic systems, 3957
Emerging infectious diseases, **3853**. *See also specific infections*
  clinicians and laboratorians in control of, 3858
  definition and classification of, 3853, 3853t, 3854f, 3854t
  health care facilities preparedness for, 1149
  mechanisms of, 3855, 3856f
  medical community response to, 3854–3855
  significance of, 3853–3854
  variables associated with, 3855–3856, 3857f
Emergomycosis, 1689t, 1726t, 1727
Emerin proteins, 2004t, 2009
Emery-Dreifuss muscular dystrophy, 3641–3642, 3641f, 3787, 3788t
Emesis, 296. *See also* Nausea and vomiting
Emetine, 228, 1743t, 1747
EMG. *See* Electromyography (EMG)
Emicizumab, 931, 931f
Emmetropia, 220
Emmonsia crescens, 1727
Emollients, 383, 385
Empagliflozin, 1988, 1992, 1995t, 1996, 3211t, 3213
Emphysema. *See also* Chronic obstructive pulmonary disease (COPD)
  bullous, A12
  centrilobular, 2250, 2250f, A12
  chest radiograph in, A12
  CT in, A12
  cyanosis in, 278
  with idiopathic pulmonary fibrosis, A12
  panlobular, 2250, 2250f, A12
  paraseptal, 2250, 2250f, A12
  pathogenesis of, 2250–2251, 2251f
  pathophysiology of, 2249
  pulmonary function values in, 2206, 2206f
  subcutaneous, A12
Emphysematous cholecystitis, 2732–2733
Empty sella, 2994
Empyema
  abdominal pain in, 112t
  actinomycotic, 1364, 1364f
  anaerobic bacterial, 1375–1376
  in cholecystitis, 2733
  definition of, 2268
  in pneumococcal pneumonia, 1192
  spontaneous bacterial, 330
  in streptococcal pneumonia, 1211
  subdural. *See* Subdural empyema
  tuberculous, 1387–1388
EMS (eosinophilia-myalgia syndrome), 416, 462, 3926
Emtansine, 566
Emtricitabine (FTC)
  adverse effects of, 1490t
  for chronic HBV infection, 1490t, 2680
  for HIV infection, 1617t, 1621, 1621t
  resistance to, 1622f
  structure of, 1619f
Enalapril, 411, 1995, 1995t, 2144t
Enalaprilat, 2170
Enamel hypoplasia, 261
Enasidenib, 571t, 574, 833, 833t
Encephalitis, **1111**
  amebic, 1114
  autoimmune, 191, 746f, 1114
  brainstem, 746
  chronic
    PML. *See* Progressive multifocal leukoencephalopathy (PML)
    progressive rubella panencephalitis, 1118
    subacute sclerosing panencephalitis, 1117–1118, 1641
  cortical, 746, 746f
  definition of, 1111
  delirium in, 83
  granulomatous amebic, 1758–1759
  hyponatremia in, 348
  in influenza, 1548
  limbic, 746, 746f, 3484, A16
  L. monocytogenes, 1228
  measles-associated, 1641
  M. pneumoniae, 1465
  paraneoplastic, 746, 746f
    with AMPA receptor antibodies, 747–748, 748f
    with antibodies to intracellular antigens, 745f
    with Caspr2 antibodies, 748
    with dopamine-2 receptor antibodies, 748
    with DPPX antibodies, 748–749
    with GABA receptor antibodies, 746f, 748
    with GlyR antibodies, 748
    with IgLON5 antibody, 738
    with LGI1 antibodies, 748
    with mGluR5 antibodies, 748
    with MOG antibodies, 738, 746f
    with NMDAR antibodies, 745f, 746–747
    treatment of, 749
  pinworm, 1114
  in rabies, 1652, 1653t
  in RMSF, 1456–1457
  toxoplasmic, 1800, 1800f
  viral, **1111, 1670**
    arthropod-borne, 1670–1671
    in cancer patient, 583
    clinical features of, 1111, 1670–1671
    CMV, 1514
    diagnosis of, 962t, 1112–1114, 1113f, 1114t, 1671
    differential diagnosis of, 1114–1115
    enterovirus, 1634, 1634t
    epidemiology of, 1115
    etiology of, 1111–1112, 1112t
    flavivirus, 1671–1672
    HHV-6, 1516
    HSV. *See* Herpes simplex virus (HSV) infections, encephalitis
    mumps-associated, 1126, 1648
    peribunyavirus, 1672–1673
    rhabdovirus, 1673
    sequelae of, 1116
    togavirus, 1673–1674
    treatment of, 1115–1116
Encephalitozoon spp., 1748, 1807, S12
Encephalomyelitis
  acute demyelinating, 1548
  acute disseminated (ADEM), **3593**
  experimental allergic, 3396, 3396f
  paraneoplastic, 744, 746, 2791t
  postinfectious, 1641, 3593–3594
Encephalopathy. *See also* Delirium
  chemotherapy-induced, 726, 726t
  chronic traumatic, 194, 3401, 3484, 3523, 3576–3577
  epileptic, 3374t
  Hashimoto's, 3483, A16
  hepatic, **2716**
    clinical features of, 2348, 2716, 3484
    diagnosis of, 2630, 2716
    imaging in, A16
    pathophysiology of, 188
    treatment of, 2716–2717
  high-altitude, 3755–3756, 3755f, 3756t
  HIV. *See* HIV infection, neurologic disease in
  hypertensive, 3455
  hyponatremic, 349
  hypoxic-ischemic, 2346–2348, 2347f, 3428
  in Lyme disease, 1451
  in malaria, 991t, 994, 1763
  metabolic, 188, 202, 2348
  posterior reversible. *See* Posterior reversible encephalopathy syndrome (PRES)
  in renal failure, 188
  sepsis-associated, 184, 2321, 2348
Enchondromatosis (Ollier's disease), 3318
Encorafenib, 533t, 569t, 572, 604, 605, 658
Endarterectomy, 2174, 2193f, 2196, 3451–3452
Endarteritis, infective, 1036
Endemic goiter, 3040
Endemic syphilis (bejel), 1437, 1437t, 1438, 1439f
Endemic treponematoses. *See* Treponematoses
Endemic (murine) typhus, **1459**
  clinical features of, 138t, 145, 1455t, 1459–1460
  diagnosis of, 1460
  epidemiology of, 1455t, 1459
  etiology of, 1455t
  treatment of, 1460
Endobronchial biopsy, 2208
Endobronchial intratumoral chemotherapy, 2289

- Endobronchial ultrasound-guided transbronchial needle aspiration, 2209–2210, 2209f, 2288, 2288f
- Endocarditis
  - infective. *See* Infective endocarditis
  - Libman-Sacks, 2830
  - marantic, 150, 408, 619, 1038
  - nonbacterial thrombotic, 1038
- Endochondromatosis, 3318
- Endocrine disorders, **2977**. *See also specific disorders*
  - cardiomyopathy in, 2018
  - clinical evaluation of, 2978–2980
  - fatigue in, 166
  - in hypothermia, 3769t
  - in lung cancer, 619
  - MASLD and, 2702
  - myopathy in, 3647
  - obesity in, 3185, 3187
  - paraneoplastic. *See* Paraneoplastic syndromes, endocrinologic
  - pathologic mechanisms of, 2977–2978
  - in pregnancy, 3900–3902
  - prevalence of, 2979t
  - radiation-induced, 755–756
  - screening/testing for, 2979t, 2980
  - types of, 2977, 2978t
- Endocrine glands, 2977. *See also specific glands*
- Endocrine tumors. *See* Neuroendocrine tumors (NETs)
- Endocrinology, 2977
- End-of-life care, **74**. *See also* Terminally ill patient
  - advance care planning in, 78–79, 78t
  - assessment and planning in, 76
  - for cancer patient, 505
  - communicating bad news to patient in, 76–77, 77t
  - continuous goal assessment in, 77–78
  - epidemiology of, 74–75, 75f, 75t
  - hospice care, 76, 88
  - in older adults, 3894
  - principles of, 5–6
  - withdrawing/withholding life-sustaining treatment during, 5–6, 88–89, 2299
- Endolymphatic hydrops. *See* Ménière's disease
- Endolymphatic sac tumors, 247
- Endometrial cancer, **714**
  - clinical features of, 714
  - deaths from, 497t, 499f, 501t, 714
  - epidemiology of, 714
  - genetic considerations in, 714
  - incidence of, 497t, 498f, 500f, 501t
  - in Lynch syndrome, 714
  - oral contraceptives and, 3153
  - pathology of, 714
  - postmenopausal hormone therapy and, 3144t, 3145, 3147
  - prognosis of, 711t, 714
  - risk factors for, 714
  - staging of, 711t, 714
  - tamoxifen use and, 714
- Endometriosis, 275, 3137, 3152
- Endometritis
  - actinomycotic, 1365
  - alone, vs. with salpingitis, 1105
  - clinical features of, 3136
  - clostridial, 1239t, 1243
  - *C. trachomatis,* 1471
- Endomyocardial biopsy, 2011, 2012f
- Endomyocardial fibrosis, 2015
- Endopep-MS assay, 1237
- Endophthalmitis
  - *Candida,* 1708, 1712, 1712f
  - clinical features of, 225, 225f
  - clostridial, 1243
  - etiology of, 225
  - *Nocardia,* 1360
  - *P. aeruginosa,* 1309t, 1311
- β-Endorphin, 430
- Endoscopic retrograde cholangiopancreatography (ERCP)
  - in acute biliary obstruction, 2489
  - in bile duct disease, 2472f, 2475f, 2489, 2736t
  - of biliary tract, 2731f
  - in cholangiocarcinoma, 2474f
  - in cholecystitis, 2732
  - in cholestasis, 324
  - in gallstone pancreatitis, 2489
  - indications for, 2462, 2462t, 2468–2469
  - in liver disease, 2631
  - in pancreatic disease, 2739t, 2741
  - in pancreatitis, 2748f
  - pancreatitis after, 2741, 2744
  - in primary sclerosing cholangitis, 2563, 2712, 2731f
  - procedure, 2468–2469
  - risks of, 2473–2474
- Endoscopic secretin test, 2739t
- Endoscopic sphincterotomy, 2468–2469, 2474f
- Endoscopic therapy. *See also* Gastrointestinal endoscopy
  - for achalasia, 2508, V5
  - for Barrett's esophagus/metaplasia, V5
  - for chronic pancreatitis, 2753–2754
  - for colonic obstruction, V5
  - for colonic pseudoobstruction, 2493f
  - for diverticular disease, 2491f, V5
  - for esophageal varices, 317, 2482, 2487f, 2714, V5
  - full-thickness resection, 2469, 2478f, V5
  - for gastric outlet obstruction, 2492f, V5
  - for gastrointestinal bleeding, 317, 2476, 2478, 2483–2487f, V5
  - for gastrointestinal malignancies, 2474f, 2478f, 2493, 2494f, 2495, 2499f, 2500f, V5
  - for gastrostomy tube complications, 2482t, 2483f
  - for ileal stricture, 2472f
  - indications for, 2464
  - for intestinal obstruction, 587, 2493–2494f
  - for Mallory-Weiss tears, 317
  - mucosal resection, 2471, 2479–2480f, V5
  - natural orifice transluminal endoscopic (NOTES) surgery, 2469
  - for obesity, 2493, 3195, V5
  - for peptic ulcer disease, 317, 2484–2485f
  - peroral endoscopic myotomy (POEM), 2469, 2476f
  - peroral endoscopic tumorectomy (POET), 2469, 2477f
  - for poisoning, 3709
  - sphincterotomy, 2474f
  - submucosal dissection, 2471, 2478f, 2479–2480f, 2495f, V5
  - suturing, 2460, 2478–2481f, V5
  - for vascular ectasia, 318, 2496, 2501f
  - for Zenker's diverticulum, V5
- Endoscopic ultrasound (EUS)
  - in bile duct disorders, 2489, 2494f, 2736t
  - in cholestatic conditions, 324
  - in chronic pancreatitis, 2752
  - in esophageal disease, 2504
  - with fine-needle aspiration, 2469, 2475f, 2499
  - for gastrointestinal cancer staging, 2462, 2469, 2475, 2499
  - indications for, 2462, 2462t, 2469
  - in pancreatic disease evaluation, 2739t, 2741–2742, 2741t
  - for pancreatic necrosis, 2469, V5
  - in ZES, 2536, 2536t
- Endoscopy. *See* Gastrointestinal endoscopy
- Endospores, 1696, 1697f
- Endostatin, 548f
- Endothelial lipase, 3239
- Endothelial nitric oxide synthase (eNOS), 328, 3161
- Endothelin(s)
  - in cardiovascular function, 1845, 1846
  - in edema, 280
  - in heart failure, 1980
  - in male sexual function, 3155, 3156f
  - in thyroid function, 3023
- Endothelin receptor antagonists, 2187f, 2194, 2197t
- Endothelium
  - activation in sepsis/septic shock, 2319
  - antithrombotic effects of, 470
  - biology of, 1845
  - in blood pressure control, 2133
  - clinical assessment of, 1846f
  - functions of, 1845, 1845t
  - lymphocyte interactions with, 2781–2782
  - in tumor angiogenesis, 544, 545f, 546, 546f, 547f
- Endothelium-derived hyperpolarizing factor, 1845
- Endotoxin. *See* Lipopolysaccharide (LPS)
- Endotoxin challenge test, 463
- Endotracheal intubation, 276. *See also* Mechanical ventilation
- Endotracheal suctioning, 1030
- Endovascular procedures. *See* Percutaneous interventions
- End-stage renal disease (ESRD). *See* Chronic kidney disease (CKD), stage 5
- End-systolic left ventricular pressure-volume relationship, 1853, 1853f
- Eneboparatide, 3287
- Enema
  - barium. *See* Barium radiography
  - for constipation, 82, 82t
- Energy balance
  - definition of, 2598
  - earth, 3746, 3746f
  - physiologic regulation of, 3181–3182, 3182f
- Energy deficiency, 2619
- Energy density, of food, 3190
- Energy expenditure, 2598, 3183, 3183f
- Energy requirements, 2598, 2619
- Enfortumab-vedotin, 535t, 566, 566t, 694
- Enfuvirtide
  - actions of, 1621
  - adverse effects of, 1618t
  - for HIV infection, 1618t, 1621
  - molecular structure of, 1617f
  - resistance to, 1623f
- Enhanced external counterpulsation, 1997, 2105
- Enhanced Liver Fibrosis (ELF) test, 2632t, 2636, 2673, 2704
- Enolase, 1188f
- β-Enolase deficiency (type XIII GSD), 3361t
- Enophthalmos, 232
- eNOS (endothelial nitric oxide synthase), 328, 3161
- Enoxaparin
  - discontinuation before lumbar puncture, S3
  - dosage of, S13
  - for NSTE-ACS, 2110t, 2111
  - in PCI, 2127
  - for VTE/PE treatment, 2163t
  - for VTE prophylaxis, 2164t, 3910
- Ensartinib, 628t
- Entacapone, 3503, 3504t
- *Entamoeba bangladeshi,* 1754
- *Entamoeba dispar,* 1754
- *Entamoeba histolytica. See also* Amebiasis
  - epidemiology of, S12
  - immune response to, 1755
  - laboratory identification of, 1756–1757, S12

Entamoeba histolytica. See also Amebiasis (*Cont.*):
  life-cycle hosts of, S12
  life cycle of, 1753, 1754f
  tissue invasion by, 1737, 1754
  virulence factors of, 1754
Entamoeba moshkovskii, 1754
Entecavir
  actions of, 1491
  adverse effects of, 1490t
  for chronic HBV infection, 1490t, 2679
    vs. other agents, 2676t, 2680f, 2682t
    recurrent after liver transplantation, 2724
    treatment recommendations, 2682–2683
  for HBV prophylaxis, 1164t
  monitoring during treatment with, 1490t
  pharmacology of, 1491, 2679
  resistance to, 1490t, 1491, 2682t
Enteral nutrition
  complications of, 2623–2624
  contraindications to, 2623
  in critically ill patient, 2624, 2625
  definition of, 2622
  in esophageal cancer, 645
  formulas for, 2622
  for IBD, 2567
  indications for, 2622
  initiation, progress, and monitoring of, 2623
  in sepsis/septic shock, 2322t, 2323
Enteric (typhoid) fever, **1315**
  vs. *Brucella* infection, 1333
  clinical features of, 139t, 1077, 1316–1317, 1316f, 1317f
  diagnosis of, 1317
  epidemiology of, 145, 1315–1316, 1316f
  global considerations in, 1315–1316, 1316f, 1318
  multidrug-resistant, 1219f, 1315–1316
  pathogenesis of, 1077, 1315
  prevention and control of, 1318–1319. *See also* Typhoid vaccines
  skin manifestations of, 402, 1316f, A1
  treatment of, 1317–1318, 1318t
Enteric nervous system, 303
Enteritis
  drug-induced, 2550t
  radiation, 554, 2550t
  sexually acquired, 1109
Enteritis necroticans, 1240
Enteroaggregative *E. coli* (EAEC), 1285t, 1288. *See also Escherichia coli* infections, intestinal
Enterobacterales. See also Gram-negative enteric bacilli
  ampC β-lactamase–producing, 1284
  antibiotic resistance in
    to aminoglycosides, 1283t, 1284
    to carbapenems, 957, 977t, 981, 985, 1152, 1186t, 1281, 1283t
    to cephalosporins, 1283t
    to fluoroquinolones, 1283t, 1284
    to fosfomycin, 1283t
    mechanisms of, 1182, 1283–1284, 1283t
    multidrug, 1029t, 1087–1088, 1152
  carbapenemase-producing, 1284
  extended-spectrum β-lactamase–producing, 1283, 1283t
Enterobacterales infections
  brain abscess, 1135
  diagnosis of, 1091–1092
  endocarditis, 1038
  epidemiology of, 1281
  epididymitis, 1099
  in human bite-wounds, 1144
  meningitis, 1119
  osteomyelitis, 1064t, 1065t
  peritonitis, 1070
  pneumonia, 1024
  prevention of, 1285
  risk factors for, 1024
  UTIs, 1087, 1091
  ventilator-associated pneumonia, 1029t, 1031
Enterobacteriaceae. *See Enterobacterales;* Gram-negative enteric bacilli
Enterobacter spp. infections, **1291**
  antibiotic resistance in, 1027, 1292
  clinical features of, 1292
  diagnosis of, 1292
  epidemiology of, 1281, 1291–1292
  osteomyelitis, 1065t
  treatment of, 1292
Enterobius vermicularis/enterobiasis (pinworm), 1815t, 1818, S12
Enterocele, 314
Enterococcal infections, **1215**
  bacteremia, 1217–1218, 1219t
  bone and joint, 1218
  endocarditis. *See* Infective endocarditis, enterococcal
  epidemiology of, 1216–1217
  etiology of, 1215, 1215f
  global considerations in, 1217
  health care–associated, 1216–1217
  intraabdominal, 1218
  meningitis. *See* Meningitis, bacterial
  neonatal, 1218
  osteomyelitis, 1065t
  pathogenesis of, 1215–1216
  peritonitis, 2716
  prostatitis, 1217
  in sepsis/septic shock, 2316
  soft tissue, 1218
  treatment of, 1042, 1218–1220, 1219t, 1220t
  urinary tract, 1087, 1217, 1219t, 1220t
  vancomycin-resistant. *See* Vancomycin-resistant enterococci (VRE)
Enterococci. *See also* Enterococcal infections
  antimicrobial resistance in, 1173t, 1184, 1185, 1217, 1220–1221, S11
  characteristics of, 1215, 1215f
  genome of, 984, 1217
  laboratory identification of, 1215, 1215f
  virulence factors of, 1216
Enterocolic fistula, 308
Enterocolitis
  necrotizing, 912, 1240
  neutropenic, 597, 597f, 1376
  Pseudomonas, 1312
  radiation, 309
  sexually acquired, 1109
Enterocytozoon bieneusi, 1600
Enterohemorrhagic *E. coli* (EHEC). *See also Escherichia coli* infections, intestinal
  apoptosis inhibition by, 973
  characteristics of, 1076, 1285t, 1287–1288
  epidemiology of, 1285t, 1287–1288
  as extracellular pathogen, 973
  food-borne, 1080t, 1081
Enteroinvasive *E. coli* (EIEC), 1285t, 1288. *See also Escherichia coli* infections, intestinal
Enterokinase, 2743
Enteropathogenic *E. coli* (EPEC), 967, 973, 1076, 1285t, 1288. *See also Escherichia coli* infections, intestinal
Enteropathy
  gluten-induced. *See* Celiac disease (gluten-induced enteropathy)
  HIV, 1600
  protein-losing, **2548**
Enteropathy-type T cell lymphoma, 867, 874
Enteroscopy, 2472f, V5
Enterotoxigenic *E. coli* (ETEC). *See also Escherichia coli* infections, intestinal
  adherence of, 1076
  characteristics of, 1285t, 1288
  epidemiology of, 1285t, 1288
  food-borne, 1080t
  toxin production in, 1076
Enterotoxins, 965
Enterovirus(es), 1534t, 1535, 1632–1633
Enterovirus infections, **1633**
  acute hemorrhagic conjunctivitis, 1636, 1636f
  acute otitis media, 254
  climate change and, 3752t
  clinical syndromes associated with, 1634t, 1636
  coxsackievirus. *See* Coxsackievirus infections
  diagnosis of, 1636
  encephalitis, 1634
  epidemiology of, 1633
  exanthems, 1635, 1635f
  generalized disease of the newborn, 1634
  global considerations in, 1633
  hand-foot-and-mouth disease. *See* Hand-foot-and-mouth disease
  herpangina, 1636
  immunity to, 1633
  meningitis, 1114, 1124, 1634
  myelitis, 3566
  myocarditis, 1635
  nonspecific febrile illness (summer grippe), 1634
  pathogenesis of, 1633
  pericarditis, 1635
  pleurodynia, 1635
  poliovirus. *See* Poliovirus infection
  rash in, 137, 144t, 1635, 1635f
  treatment of, 1636–1637
  upper respiratory, 1636
Enthesitis, 2882, 2883, 2887, 2940t, 2942t, 2944f
Entomologic inoculation rate, 1762
Entrainment, 3944t
Entrapment neuropathy, 3903
Entrectinib, 533t, 569t, 572, 573, 628
Entropion, 224, 1475
Entry inhibitors. *See* Combination antiretroviral therapy (cART)
Enuresis, sleep, 217
Enveloped viruses, 1478f, 1479
Envelope protein, HIV, 1558, 1559
env gene, 1551, 1552f, 1559, 1561f
Environmental enteric dysfunction, 2546
Environmental tobacco smoke. *See* Smoking, passive exposure to
Enzalutamide, 567, 701–702, 702f, 703, 753t
Enzyme(s). *See specific enzymes*
Enzyme immunoassay (EIA), 1589, S11
Enzyme-linked immunosorbent assay (ELISA), 1484
Eosinopenia, 462
Eosinophil(s), **462**
  in adaptive immunity, 2763t
  in asthma, 2223
  disorders of, 462
  functions of, 462
  hematopoietic differentiation of, 760f
  in inflammation, 2769–2770
  in innate immunity, 2763t, 2769–2770
  morphology of, 456f
  in parasitic infections, 462, S12
  in peripheral blood smear, 442f, 443
Eosinophil chemotactic factor of anaphylaxis (ECFa), 462
Eosinophil granules, 462
Eosinophilia. *See also* Hypereosinophilia
  diseases associated with, 462
  drug-induced, 462
  infectious causes of, 960, 961t

  - paraneoplastic, 740t, 741
  - parasites associated with, 1809, 1813, S12
  - tropical pulmonary, 462
- Eosinophilia-myalgia syndrome (EMS), 416, 462, 3926
- Eosinophilic bronchitis, 273
- Eosinophilic esophagitis, 295, 296, 300, 2491, **2511,** 2512f
- Eosinophilic fasciitis, 416, 462, 810, 2876
- Eosinophilic folliculitis, 395
- Eosinophilic gastritis, 2539
- Eosinophilic gastroenteritis, 300, 309, 2550, 2550t
- Eosinophilic granulomatosis with polyangiitis (EGPA) (Churg-Strauss), **2900**
  - ANCAs in, 2233, 2894
  - clinical features of, 406, 2014, 2232, 2266, 2420, 2899, 3623
  - definition of, 2900
  - diagnosis of, 2233, 2420, 2900, A14
  - incidence and prevalence of, 2900
  - pathology and pathogenesis of, 2894, 2900
  - pericardial effusion in, A14
  - prognosis of, 2233
  - renal involvement in, 2415t, 2420
  - treatment of, 2233, 2900–2901
- Eosinophilic leukemia, chronic, 819, 876, 878t
- Eosinophilic myocarditis, 2014–2015
- Eosinophilic pneumonia. *See* Pulmonary infiltrates with eosinophilia
- Eosinophilic pustular folliculitis, 1605
- Eotaxin, 2767t
- *EPB41* gene mutations, 794t
- *EPB42* gene mutations, 794t
- Epcoritamab-bysp, 534t, 559f, 753t, 863
- EPEC (enteropathogenic *E. coli*), 967, 973, 1076, 1285t, 1288. *See also Escherichia coli* infections, intestinal
- Ependymomas, 720, A16
- *Ephedra sinica* (ma huang), 2137, 3414t, 3926
- Ephedrine, 354, 3710t
- Ephelides, 399
- EPH receptors, 546, 548f
- Ephrins, 545, 546f
- Epicondylitis, 2942t
- Epidemic, 1544
- Epidemic polyarthritis (Ross River virus disease), 1059, 1657t, 1669–1670
- Epidemic potential, 987
- Epidemic (louse-borne) typhus, **1458**
  - clinical features of, 138t, 992, 1455t, 1459
  - diagnosis of, 1459
  - epidemiology of, 138t, 145, 992, 1455t, 1458–1459, 3732
  - global considerations in, 1459
  - prevention of, 1459
  - skin manifestations of, 138t
  - treatment of, 1459
- Epidemiologic cutoff value (ECV), S11
- Epidemiologic transition, **1854,** 1855T
- Epidermal growth factor (EGF), 2516, 2516f, 2981t, 3023
- Epidermal growth factor receptor (EGFR), 615, 2539, 3004
- Epidermal growth factor receptor (EGFR) inhibitors. *See* EGFR inhibitors
- Epidermoid cysts, 403, 722, A5
- Epidermolysis bullosa, 296, 3320, 3785
- Epidermolysis bullosa acquisita, 401, 410t, **413**
- Epididymitis, 1097t, 1099, 1099t, 1470, 1473t
- Epidural abscess
  - anaerobic bacterial, 1375
  - cervical, 131, 131f
  - clinical features of, 994
  - cranial, 1140, 1141f
  - etiology of, 991t, 994
  - intracranial, 994
  - spinal, 991t, 994, 3563–3564, 3563f
  - treatment of, 963t, 991t, 3564
- Epidural anesthesia, 3908
- Epidural hematoma, 131, 3564, 3572, 3572f, A16
- Epidural metastases, 588, 725, 725f
- Epidural spinal cord compression, 3562, 3562f
- Epidural spinal hemorrhage, 3455
- Epidural steroid injections, 128, 132
- Epigenetic clock, 3870
- Epigenetics, **3935**
  - of aging, 3870, 3940–3941
  - biochemical bases of, 3935–3937, 3936f
  - of brain and behavior, 3941–3942
  - in cancer, 538–539, 538f, 3940
  - definition of, 6, 538, 3935
  - in development and differentiation, 3938–3939
  - gene expression and, 522, 3782–3783, 3784f
  - in infection, immunity, and inflammation, 3942
  - of metabolism, 3939–3940
  - tools for study of, 3937–3938, 3938f
  - in undiagnosed disease, 3997t
- Epigenome, 3782
- Epiglottitis, **259,** 1262, 1263
- Epilatory, 3141
- Epilepsia partialis continua, 746, 3409
- Epilepsy, **3408**. *See also* Seizure(s)
  - autonomic, 160
  - catamenial, 3427
  - coma in, 189
  - continuous seizures in. *See* Status epilepticus
  - definition of, 3408–3409
  - employment and, 3427
  - etiology of, 3411, 3413–3414, 3413t, 3414t
  - genetic testing in, 3418
  - incidence of, 3408
  - loss of driving privileges in, 3427
  - mortality data, 3427
  - in older adults, 3428
  - pregnancy and, 3427, 3903
  - psychiatric disorders associated with, 3426–3427
  - psychosocial issues in, 3427
  - refractory, 3424–3425
  - sudden unexpected death in, 3427
  - syndromes, **3411**
    - genetic considerations in, 3412t, 3415
    - juvenile myoclonic epilepsy, 3411
    - Lennox-Gastaut syndrome, 3411
    - mesial temporal lobe epilepsy, 3411, 3413f, 3413t
  - treatment of, **3419**
    - antiepileptic drug therapy. *See* Antiepileptic drugs (AEDs); *specific drugs*
    - in refractory disease, 3424–3425
    - stem cell therapy, 3407
    - surgical, 3424–3425
  - in women, 3427
- Epileptic encephalopathy, 3374t
- Epileptic spasms, 3410
- Epileptiform spikes, 3409, 3417
- Epileptogenesis, 3411, 3415
- Epinephrine
  - in advanced cardiac life support, 2336, 2337f
  - adverse effects of, 2315
  - for anaphylaxis, 2813, 3719, 3736
  - for drug-induced urticaria/angioedema, 422
  - in glucose regulation, 3232t
  - in hypoglycemia, 3232t
  - injection, for ulcer hemostasis, 2481, 2485f
  - in local anesthetics, 266
  - neutrophil count and, 458
  - for scombroid poisoning, 3726
  - for sepsis/septic shock, 2322t, 2323
  - for shock, 2315
  - in vascular smooth-muscle cells, 1846
- Epinephrine challenge test, 463
- Epiretinal membrane, 232
- Epirubicin, 562t, 692
- Episcleritis, 224, 2562t, 2563, 2920
- Episodic ataxia, 3374t, 3536
- Epistasis, 3793
- Epistaxis
  - approach to the patient, 472
  - vs. hemoptysis, 274
  - in nasopharyngeal cancer, 611
  - in Osler-Weber-Rendu syndrome, 927
  - in VWD, 926
- Epithelial cells/epithelium
  - in adaptive immunity, 2763t
  - airway, in asthma, 2218
  - bacterial avoidance of, 969
  - gastric, 2514–2515
  - high-resistance, 2370
  - in innate immunity, 969
  - intestinal, 2543–2544
  - permeability of, 2365, 2781
- Epithelial-mesenchymal transition, 541, 542f
- Epitope spreading, 2826
- Epitranscriptomics, 3937
- Eplerenone
  - drug interactions of, 1422
  - for heart failure, 1994, 1995t
  - for hypertension, 2144t, 2146
  - for mineralocorticoid excess, 3063
  - for STEMI, 2011
- Eplets, 2407
- Epley maneuver, 163–164, 164f
- Eplontersen, 900, 2027
- *EPM2A* gene mutations, 3412t, 3515
- EPOCH regimen, 650, 873
- EPOCH-R regimen, 862
- Epoprostenol, 2190, 2194, 2197t, 2303, 2303t, 2307, 2874
- Epothilones, 508, 565
- EPP. *See* Erythropoietic protoporphyria (EPP)
- Epratuzumab, 848t, 849
- Eprodisate, 899
- *Epsilon-sarcoglycan* gene, 3515
- Epsilon wave, 1966
- Epstein-Barr virus (EBV), 1508
- Epstein-Barr virus (EBV) infections, **1508**
  - cancer and, 512t, 1510
  - diagnosis of, 1510–1511, 1511f, S11
  - differential diagnosis of, 1511, 1511t
  - encephalitis, 1111, 1114t
  - epidemiology of, 1508
  - gastric adenocarcinoma and, 642, 645, 1510
  - global considerations in, 1510
  - in HCT recipient, 1155t, 1161
  - hemophagocytic lymphohistiocytosis and, 466, 468, 882
  - infectious mononucleosis. *See* Infectious mononucleosis
  - latent, 434
  - lymphoid malignancies and, 858t, 1510–1511
    - B cell lymphoproliferative disease, 755
    - Burkitt's lymphoma, 512t, 525, 844, 857, 1510
    - extranodal NK/T-cell lymphoma, nasal type, 857, 867–868, 874, 2898
    - Hodgkin's lymphoma, 1510
    - PCNSL, 720
    - post-solid organ transplant, 2351
    - primary effusion lymphoma, 873
  - malaria and, 1766
  - ME/CFS and, 1510
  - meningitis, 1125
  - MS and, 3581
  - myelitis, 3566
  - nasopharyngeal cancer and, 610, 1510
  - neuropathy in, 3607

Epstein-Barr virus (EBV) infections (*Cont.*):
  oral hairy leukoplakia, 1010f, 1510
  pathogenesis of, 1508–1509
  pharyngitis, 257, 1536
  posttransplant lymphoproliferative disorder and, 1161
  reactivation in DIHS, 417
  rheumatoid arthritis and, 2843
  skin manifestation of, 402
  in solid organ transplant recipient, 1158, 1161
  systems biology applied to, 3961t
  treatment of, 1512
  vascular composite allografts and, 1158
Epstein's syndrome, 924
Eptifibatide
  actions of, 943f, 945
  adverse effects of, 923t, 946
  discontinuation before lumbar puncture, S3
  dosage of, 945, S13
  genetic variations in response to, 939t
  indications for, 945
  for NSTE-ACS, 2110t
  for PCI, 2127
  pharmacology of, 945t
Eptinezumab, 3469, 3470t
Epulis (gingival hypertrophy), A3
Epulis fissuratum, A3
Epworth sleepiness scale, 2275
Equine antitoxin
  botulinum, 1237
  tetanus, 1231
Eradication, of disease, 1002
Eravacycline
  for *A. baumannii* infections, 1297, 1297t
  actions of, 1166, 1177, 1182t
  adverse effects of, 1171t
  for *Citrobacter* infections, 1293
  for *Enterobacterales* infections, 1283–1284
  for enterococcal infections, 1220
  indications for, 1174t, 1177
  for MRSA infections, 1205
  for NTM infections, 1418
  for peritonitis, 1071
  in pregnancy and lactation, 1169t
  resistance to, 991, 1174t, 1182t, 1184
ERBB2 gene mutations
  amplification of, 521
  in liver cancer, 667
  in lung cancer, 617, 627, 628f
  in upper gastrointestinal tract cancer, 642, 647, 651
  in urothelial carcinoma, 691
ERBB3 gene mutations, 713
ERCP. *See* Endoscopic retrograde cholangiopancreatography (ERCP)
Erdafitinib
  action and targets of, 533t, 568t, 572
  adverse effects of, 508, 694
  for bladder cancer, 694
Erdheim-Chester disease, 3566
Erectile dysfunction (ED), **3155**
  after prostatectomy, 699
  after radiation therapy, 700, 756
  alcohol use and, 3677
  approach to the patient, 3157–3158, 3158f
  in diabetes mellitus, 3227
  drug-related, 3157, 3157t
  epidemiology of, 3155
  in MS, 3588
  in MSA, 3548
  pathophysiology of, 3156–3157
  prevalence of, 2979t
  screening/testing for, 2979t
  in systemic sclerosis, 2871
  testosterone deficiency and, 3118
  treatment of, 3158–3160, 3159t
Erection, 3155, 3156f
Erenumab, 3469, 3470t
Ergot alkaloids, 2178, 3705, 3710t
Ergotamine
  adverse effects of, 3468
  for exercise headache prevention, 3474
  for migraine, 3467t, 3468, 3468t
  overdosage/poisoning with, 3710t
Eribulin, 508, 563t, 565
  for metastatic breast cancer, 641
  for soft tissue sarcomas, 729
Erlotinib
  actions and targets of, 533t, 568t, 572
  adverse effects of, 421, 568t, 572, 595
  drug interactions of, 591
  for head and neck cancer, 614
  interactions and issues, 568t
  for lung cancer, 572, 627, 627t
  for pancreatic cancer, 676t
  resistance to, 572
ermB gene, 1026
EROS, 460
Erosio interdigitalis blastomycetica, 1708
Erosion (skin lesion), 377t
Ertapenem
  for anaerobic bacterial infections, 1379
  for community-acquired pneumonia, 1028t
  indications for, 1173t, 1175
  in pregnancy and lactation, 1169t
  resistance to, 1173t, 1181
Erucism, 3737
Ervebo, S4
Erve virus, 1661t
Erysipelas
  etiology of, 1050t, 1052, 1210
  in familial Mediterranean fever, 2934
  pathophysiology of, 1049f, 1052
  skin manifestations of, 1052, 1210, 1210f, A1, A5
  treatment of, 1052, 1208t
Erysipelothrix rhusiopathiae, cellulitis, 1052
Erythema annulare centrifugum, 394
Erythema elevatum diutinum, 406
Erythema gyratum repens, 394
Erythema induratum, 405
Erythema infectiosum (fifth disease)
  clinical features of, 138t, 1522–1523, 1523f
  diagnosis of, 1524, 1524t
  epidemiology of, 138t
  etiology of, 138t
  skin manifestations of, 138t, 145, 402, 1523f, A1, A5
Erythema marginatum, 139t, 146, 394, 2856. *See also* Rheumatic fever
Erythema migrans, 139t, 145–146, 394, 1449f, 1450, A1, A5. *See also* Lyme borreliosis (Lyme disease)
Erythema multiforme
  classification of, 401
  clinical features of, 379, 381f, 394, 401, 424, 1465
  in coccidioidomycosis, 1697
  differential diagnosis of, 140t
  epidemiology of, 140t
  etiology of, 401, 424
  HSV infection and, 1498, 1503t, A1
  oral manifestations of, 263t
  skin manifestations of, 140t, 146, A1, A5
Erythema nodosum
  clinical features of, 148, 405
  in coccidioidomycosis, 1697
  epidemiology of, 144t
  etiology of, 144t
  in IBD, 2561, 2562t
  skin manifestations of, 144t, 148, A1, A5
Erythema nodosum leprosum, 1408–1409, 1412
Erythema, sunburn, 430
Erythrasma, 382, 1226
Erythrocyte blood groups, 901, 904t
Erythrocyte sedimentation rate (ESR), 115, 135, 960, 962t
Erythrocyte-stimulating agents, 764, 769
Erythrocytosis, 450. *See also* Polycythemia
  etiology of, 819, 820t
  in hemoglobinopathy, 780
  in myelofibrosis, 822
  in obstructive uropathy, 2455
  paraneoplastic, 740–741, 741t
  stress (spurious), 819
  testosterone therapy and, 3121
Erythroderma, **392**
  clinical features of, 392, 393t, 993
  drug-induced, 393
  etiology of, 392–394, 393t
  in malignant diseases, 393–394
  in primary cutaneous diseases, 393t. *See also specific diseases*
  in toxic shock syndrome, 991t, 993. *See also* Toxic shock syndrome (TSS)
  treatment of, 393t
Erythroferrone, 762, 763
Erythroid hyperplasia, 448f, 791, A6
Erythroid precursors, 762, 763f
Erythroleukemia, A6
Erythromelalgia (erythermalgia), 819, 2177f, **2179**
Erythromycin
  for acne vulgaris, 391
  for actinomycosis, 1366t
  actions of, 1166, 1182t
  adverse effects of, 299, 324, 1177, 2667
  for bacillary angiomatosis/peliosis, 1353t
  for bite-wound infections, 1145t
  for bronchiectasis prophylaxis, 2244
  for cellulitis, 1054t
  for chancroid, 1264
  for cholera, 1329
  for *C. trachomatis* infection, 1474
  for diphtheria, 1224
  for donovanosis, 1358t
  drug interactions of, 485t, 1177
    azoles, 1742t
    cyclosporine, 2408, 2721
    PDE-5 inhibitors, 3159
    statins, 485t
    tacrolimus, 2408, 2721
  for endoscopy pretreatment, 319
  for gastroparesis, 299, 299t
  indications for, 1173t
  for *Legionella* spp. infections, 1275t
  for Lyme disease, 1453, 1453f
  for pertussis, 1280t
  for pertussis prophylaxis, 1280
  in pregnancy and lactation, 1169t
  prophylactic, 2858
  prophylactic uses of, 1180t
  for psittacosis, 1476
  for relapsing fever, 1447–1448, 1447f
  resistance to, 1173t, 1182t, 1208–1209
  for streptococcal pharyngitis, 258, 258t
Erythron, 443–444
Erythroparvovirus, 1522. *See also* Parvovirus B19 infections
Erythrophagocytosis, in aggressive lymphoma, A6
Erythroplakia, 261, 264t, 610, 611
Erythropoiesis
  ineffective, 322, 2638
  iron-deficient, 762
  iron utilization in, 763
  megaloblastic, A6
  physiology of, 443–444, 444f

- Erythropoiesis-stimulating agents (ESA), 2393
- Erythropoietic protoporphyria (EPP)
  - clinical features of, 402, 433, 435, 3341t, 3343t, 3350, 3350f
  - diagnosis of, 435, 3343t, 3350
  - global considerations in, 3342
  - pathophysiology of, 3341f, 3349–3350
  - treatment of, 435, 3350
- Erythropoietin
  - actions of, 443
  - in anemia, 444, 444f, 762
  - deficiency of, 450
  - ectopic production of, 451, 740–741
  - elevated, 451
  - in hematopoietic differentiation, 760f
  - plasma, 444
  - regulation of production of, 443, 443f
- Erythropoietin therapy
  - adverse effects of, 575, 814t
  - for anemia, 450, 769, 2393
  - for MDS, 818
  - for orthostatic hypotension, 160
- ESBL (extended-spectrum β-lactamase), 1151, 1181, 1283, 1283t
- Escape mutants, in HBV, 2647
- Eschar
  - after tick bite, 148f
  - in anthrax, 145t, 1051, A1, S4
  - characteristics of, 137, 1049f, 1051
  - in ecthyma gangrenosum, 143t
  - in rickettsialpox, 143t, 149
  - in scrub typhus, 138t, 149
- *Escherichia coli*
  - adherence of, 967, 1076, 1088
  - antibiotic resistance in, 1027, 1093, 1174t, 1181, 1185, 1283, 1286–1287
  - in combined sewer systems, 3750, 3752f
  - commensal strains, 1285
  - epidemiology of, 1281
  - extraintestinal pathogenic strains, 965, 966t, 1281, 1282t, 1285
  - genome of, 984, 1285
  - in inflammatory bowel disease, 3836
  - intestinal pathogenic strains, 1278t, 1287
    - diffusely adherent, 1288
    - enteroaggregative, 1288
    - enterohemorrhagic. *See* Enterohemorrhagic *E. coli* (EHEC)
    - enteroinvasive, 1288
    - enteropathogenic, 1076, 1288
    - enterotoxigenic. *See* Enterotoxigenic *E. coli* (ETEC)
    - Shiga toxin–producing. *See* Shiga toxin–producing *E. coli* (STEC)
  - laboratory identification of, S11
  - pili of, 1088
  - structure and function of, 1281
  - toxin production by, 1076–1077
  - virulence factors of, 1088, 1281
- *Escherichia coli* infections, **1285**
  - epidemiology of, 1281
  - extraintestinal, 1282
    - abdominal/pelvic, 1071, 1286
    - bacteremia, 1286
    - cellulitis, 1286
    - diagnosis of, 1286
    - ecthyma gangrenosum, 408
    - endovascular, 1286
    - meningitis, 1286
    - osteomyelitis, 1062
    - perinephric abscess, 1075
    - peritonitis, 2716
    - pneumonia, 1286
    - treatment of, 1286–1287
    - urinary tract, 1087–1088, 1100, 1285–1286
  - intestinal. *See also* Travelers' diarrhea
    - clinical features of, 305, 1080, 1080t, 1285t
    - diagnosis of, 306, 1289
    - enterohemorrhagic: strain O157:H7, 305, 985, 1080, 1289, 2444. *See also* Hemolytic-uremic syndrome (HUS)
    - etiology of, 1286–1287
    - vs. IBD, 2560
    - Shiga toxin–producing, 2444
    - transmission of, 1287
    - treatment of, 1289
  - sepsis/septic shock, 2316
  - treatment of, 1174t
- Escitalopram
  - adverse effects of, 3659t, 3666
  - for depression, 85, 199, 3659t, 3667
  - for depression in AD, 3482
  - for depression in MS, 3588
  - for generalized anxiety disorder, 3661
  - genetic variations in response to, E72t
  - for menopausal symptoms, 3143
- Esketamine, 3659t, 3667
- Eslicarbazepine, 3421t
- Esmolol
  - for aortic dissection, 2170
  - for ischemic heart disease, 2100t
  - for MAOI overdosage/poisoning, 3710t
  - overdosage/poisoning with, 3712t
- Esomeprazole, 301, 2524–2525, 2524t, 2536
- Esophageal atresia, 2507
- Esophageal cancer, **641**
  - adenocarcinoma, 641–642, 643t, 2468f
  - Barrett's esophagus and. *See* Barrett's esophagus/metaplasia
  - clinical features of, 296, 642–643, 2507
  - deaths from, 497t, 641, 2507
  - dysphagia in, 2493, 2499f
  - early-onset, 642
  - endoscopic appearance of, 2510, 2510f
  - etiology of, 642
  - genomic alterations in, 642
  - global considerations in, 641–642
  - *H. pylori* infection and, 1300
  - incidence of, 497t, 498, 641–642, 2506–2507
  - prevention of, 511
  - prognosis of, 643t
  - risk factors for, 300, 642, 3677, 3683t
  - screening for, 642
  - smokeless tobacco and, 511
  - squamous cell, 641–642, 643t, 2491, 2495f
  - staging of, 643, 643t, 644f, 645t
  - treatment of, 643–644, 2477f, 2481f, V5
- Esophageal disease, **2503**
  - clinical features of
    - chest pain, 103t, 104, 2504, 2508
    - dysphagia, 293–296, 2504, 2506
    - globus sensation, 2504
    - heartburn, 300, 2460, 2503
    - indigestion, 300. *See also* Indigestion
    - odynophagia, 2504
    - regurgitation, 2503
    - water brash, 2504
  - congenital anomalies, 2507
  - dermatologic disorders and, 2514
  - diagnosis of, 2504–2505, 2505f
  - diverticula, 296, 2506, 2506f. *See also* Diverticular disease
  - esophagitis. *See* Esophagitis
  - GERD. *See* Gastroesophageal reflux disease (GERD)
  - hiatal hernia, 2505, 2505f
  - iatrogenic injury, 2513
  - mechanical trauma, 2513
  - motility disorders, 2507–2509
    - achalasia. *See* Achalasia
    - diffuse esophageal spasm, 2508–2509
    - distal esophageal spasm, 295, 2509f
  - obstruction, 2484–2485, 2492f
  - rings, 2505–2506, 2505f
  - rupture, 2272
  - Schatzki's ring, 294, 296, 2492, 2497f, 2506, 2512f
  - stricture, 294, 2493, 2496f, 2498f
  - structural disorders, 2505–2507, 2505f, 2506f
  - in systemic sclerosis, 2870, 2870t
  - tumors, 652, 2477f, 2506–2507. *See also* Esophageal cancer
  - varices. *See* Esophageal varices
  - webs, 2505–2506
- Esophageal manometry
  - in distal esophageal spasm, 2509f
  - in dysphagia, 296
  - in GERD, 301
  - indications for, 2463, 2504
  - normal, 2505f
- Esophageal pseudodiverticulosis, 2506, 2506f
- Esophageal spasm
  - chest pain in, 103t
  - diffuse, 2508–2509, 2509f
  - distal, 295, 2509f
  - jackhammer esophagus, 2509f
- Esophageal varices
  - bleeding in, 317–318
  - diagnosis of, 2482, 2487f, 2713
  - in portal hypertension, 2713
  - in superior vena cava syndrome, 585
  - treatment of, 2482, 2487f, 2714, 2714f, V5
- Esophageal web, 642
- Esophagitis, 2495f
  - clinical features of, 112t, 295–296, 2513–2514
  - corrosive, 2513
  - diagnosis of, 2490–2491
  - eosinophilic. *See* Eosinophilic esophagitis
  - in GERD. *See* Gastroesophageal reflux disease (GERD)
  - in HIV infection, 1598, 1599f, 2512f, 2513–2514
  - infectious, 2512–2513
    - *Candida*. *See Candida* spp. infections, esophageal
    - CMV, 2495f, 2513
    - HSV, 1500, 1503t, 2495f, 2513
  - peptic, 2510f
  - pill-induced, 295, 2513
  - radiation, 2513
  - treatment of, 301
- Esophagogastroduodenoscopy, 2465, 2466–2467f, 2468f, 2504
- Esophagus
  - anatomy of, 293, 2503
  - Barrett's. *See* Barrett's esophagus/metaplasia
  - "corkscrew," 2508, 2509f
  - foreign bodies in, 2484–2485, 2492f, 2514
  - functions of, 2503
  - gastric juice effects on, 2510
  - jackhammer, 2509f
  - in scleroderma, 2514
  - systemic disease effects on, 2514
- ESR (erythrocyte sedimentation rate), 115, 135, 960, 962t
- *ESR1* gene mutations, 639, 640
- ESRD (end-stage renal disease). *See* Chronic kidney disease (CKD), stage 5
- Essential mixed cryoglobulinemia, 2685, 2693. *See also* Cryoglobulinemic vasculitis
- Essential thrombocytosis (ET), 819, **823**
- Essential tremor, 3497t, 3509
- Estazolam, 3713t
- Esthesioneuroblastoma, 730
- Esthiomene of genitalia, 1472
- Estimated average requirement, 2599

- Estimated glomerular filtration rate (eGFR), 2362, 2362t. *See also* Glomerular filtration rate (GFR)
- Estrogen(s)
  - actions of, 3127f
  - asthma symptoms and, 2221
  - bone remodeling and, 3294f, 3295, 3303
  - breast cancer and, 633
  - carcinogenicity of, 512t
  - cardioprotective actions of, 3164
  - excess of, 3116
  - neuroprotective actions of, 3163–3164
  - in perimenopause, 3142, 3142f
  - production of, 3104f, 3127–3128, 3127f
- Estrogen deprivation therapy, 637
- Estrogen receptors
  - actions of, 3128, 3302
  - in breast cancer, 536
  - in carcinoma of unknown primary, 733t
- Estrogen therapy. *See also* Postmenopausal hormone therapy
  - for adult hypopituitarism, 2996t
  - adverse effects of, 3302, 3303f
    - cholestasis, 2667, 2667t
    - cutaneous, 399, 435
    - dyslipidemia, 3246
    - edema, 282t
    - erectile dysfunction, 3157
  - colorectal cancer and, 513, 655
  - for contraception. *See* Contraception; Oral contraceptives
  - endometrial cancer and, 714
  - for female sexual dysfunction, 3161–3162
  - forms of, 3148
  - for gonadotropin deficiency, 2998
  - for hirsutism, 3141
  - for hypogonadism, 2998
  - for migraine treatment, 3468t
  - for osteoporosis, 3302–3303
  - for PCOS, 3135
  - transdermal, 3145, 3148
  - for Turner's syndrome, 3097
  - vaginal, 1088, 3145, 3149
- Eszopiclone, 216, 242
- ET (essential thrombocytosis), 819, **823**
- Etanercept
  - actions of, 2784, 2791t
  - adverse effects of, 387t, 462, 1056, 2835, 2850t
  - for autoimmune and inflammatory diseases, 2791t
  - for axial spondyloarthritis, 2887t
  - monitoring during treatment with, 2850t
  - for psoriasis/psoriatic arthritis, 387t, 2890
  - for rheumatoid arthritis, 2850t, 2851
- ETEC. *See* Enterotoxigenic *E. coli* (ETEC)
- Etelcalcetide, 2391
- Ethambutol
  - actions of, 1423
  - adverse effects of, 1397, 1423
    - gout, 2940t
    - neuropathy, 3609t
    - optic neuritis, 1420t, 1423
    - optic neuropathy, 228
    - thrombocytopenia, 923t
    - zinc deficiency, 2614
  - dosage of, 1423
  - for MAC infection prophylaxis, 1593t
  - for MDR-TB, 1394t, 1395t, 1397–1399, 1421t
  - for NTM infections, 1418, 1423, 1427, 2244
  - pharmacology of, 1423
  - resistance to, 1397, 1423
  - for TB, 1394–1396, 1394t, 1395t, 1419t, 1420t, 1423
  - for tuberculous meningitis, 1127
- Ethanol
  - beverage. *See* Alcohol
  - for ethylene glycol poisoning, 371
- Ethchlorvynol, 228
- Ethical issues, **70**
  - avoiding deception, 72
  - beneficence, 71
  - claims of conscience, 72
  - clinical practice guidelines and, 74
  - in clinical research, 73–74
  - confidentiality, 71
  - conflicts between patient's choices and their best interests, 71
  - conflicts of interest, 73
  - futile care, 89
  - in genetic testing, 3798–3799
  - in global health field experiences, 74
  - informed consent, 5, 70–71
  - justice, 72
  - medical errors. *See* Adverse events
  - moral distress, 73
  - new technologies and, 7, 74
  - occupational risks, 72
  - patient autonomy, 70–71
  - patient who lacks decision-making capacity, 71
  - personal values and, 72
  - physician impairment, 73
  - placebo effects and, 3934
  - principles in, 70–72
  - professional oaths and codes and, 72
  - resource allocation, 72
  - virtue ethics, 72
- Ethinyl estradiol, 3141, 3142, 3148, 3153. *See also* Estrogen therapy
- Ethionamide, 1394t, 1395, 1395t, 1399, 1421t, 1426
- Ethnic groups. *See also* Racial/ethnic disparities in health care
  - drug response in, 491
  - waist circumference values in, 3188t
- Ethosuximide, 3421t, 3713t
- Ethylenediaminetetraacetic acid (EDTA), 3702
- Ethylene glycol poisoning
  - acidosis in, 371, 3705, S1
  - AKI in, 2377
  - clinical features of, 3714t
  - diagnosis of, 2381, S1
  - mechanism of, 3714t
  - optic neuropathy in, 228
  - treatment of, 192, 371, 3714t
- Ethylene oxide, 3610t
- Ethynodiol diacetate, 3141
- Etidronate, 738, 3270, 3314t
- Etomidate, 739, 3012, 3059, 3713t
- Etonogestrel, 3154
- Etoposide
  - actions of, 564
  - adverse effects of, 562t, 564
    - DIHS, 597
    - late, 709
    - nausea and vomiting, 575
    - neurologic, 591, 3608t
  - for Burkitt's lymphoma, 862
  - drug interactions of, 591
  - for Ewing's sarcoma, 730
  - for gestational trophoblastic disease, 715
  - for hemophagocytic lymphohistiocytosis, 466, 469, 882
  - interactions and issues, 562t
  - for lung cancer, 630
  - for macrophage activation syndrome, 469
  - for ovarian germ cell tumors, 712
  - for peripheral T-cell lymphoma, 866
  - for soft tissue sarcomas, 729
  - for testicular cancer, 706, 707–708f, 709
- Etranacogene dezaparvovec, 3824t
- Etravirine, 1617t, 1619f, 1622f
- *ETV6* gene mutations, 825, 825t
- Euchromatin, 3935
- Eumelanin, 430
- Eumycetoma, 1688t, 1725
- Eunuchoid proportions, 3107
- European bat lyssaviruses, 1655
- European fire ants, 3736
- EUS. *See* Endoscopic ultrasound (EUS)
- Eustachian tube dysfunction, 245
- Euthanasia, 89–90, 90t
- Euthyroid hyperthyroxinemia, 3024, 3024t
- Evacuation disorders, 312
- Everglades virus, 1657t, 1673
- Everolimus
  - actions and targets of, 533t, 535, 571t, 574, 2408t
  - adverse effects of, 508, 571t, 574
    - pulmonary, 595–596
    - in transplant recipients, 2408, 2408t, 2722
  - for breast cancer, 574, 640
  - in drug-eluting stents, 2128
  - drug interactions of, 591
  - for immunosuppression, 2029t, 2408, 2408t, 2722
  - for insulinoma, 678, 678t, 3235
  - interactions and issues, 571t
  - for lymphoplasmacytic lymphoma, 865
  - for NETs, 684t, 686, 735
  - for RCC, 689, 690t
  - for reducing skin cancer risk in transplant recipients, 433
  - for small-bowel carcinoid tumors, 652
  - for subependymal giant cell astrocytomas, 723
  - for tuberous sclerosis, 2434
- Evidence-based medicine, 4, 27–29
- Evinacumab, 3244t, 3252
- EVLP (ex vivo lung perfusion), 2285
- Evoked potentials, 3583–3584
- Evolocumab, 2110, 3244t
- Ewart's sign, 2079
- Ewingii ehrlichiosis, 1455t, 1461
- Ewing's sarcoma, **730**
  - clinical features of, 261
  - diagnosis of, 730
  - genetic considerations in, 521t, 730
  - lactate dehydrogenase in, 504t
  - treatment of, 730
- Exagamglogene, 2447, 3824t
- Exanthems. *See* Rash
- Exanthem subitum (roseola), 138t, 145, 1516, A1
- Exchangers (antiporters), 2366
- Exchange transfusion, 776, 1779–1780, 1780t
- Excisional biopsy, 552
- Excitable gap, 1914
- Excitation-contraction coupling, genetic defects of, 2004t
- Excoriation, 377t
- Executive functions, 208
- Exemestane, 513, 567, 634, 637
- Exenatide, 3211t, 3213
- Exercise/exertional headache, primary, 113t, 3473–3474
- Exercise intolerance/tolerance, **291**
  - in glycolytic defect disorders, 3645
  - in ischemic heart disease, 2099
  - preoperative, 3906f
- Exercise/physical activity
  - anaphylaxis due to, 2807
  - asthmatic bronchoconstriction and, 2221, 2227
  - in AV conduction block diagnosis, 1927
  - benefits of, 38t
  - for cancer prevention, 511, 634
  - cardiac response to, 1852, 1852f
  - cardiovascular disease and, 1857
  - for diabetes mellitus, 3206
  - energy required for, 2598
  - energy requirements for common activities, 2099, 2099t

  - for fatigue, 167
  - gonadotropin deficiency due to, 3113
  - for heart failure, 1988, 1997
  - for hypertension management, 2137t, 2142
  - hypoglycemia and, 3236
  - for intermittent claudication, 2173–2174
  - for lifespan enhancement, 3875
  - in metabolic syndrome, 3253, 3257
  - for older adults, 3875
  - for osteoarthritis, 2954
  - for osteoporosis management/prevention, 3301–3302
  - for PD, 3507
  - recommendations for, 11, 12t
  - tissue oxygen requirements in, 277
  - for type 2 DM prevention/delay, 3202
  - for weight loss, 3190–3191
- Exercise testing, 109, 271, 292, 1888, 1889f. *See also* Echocardiography, stress; Electrocardiography (ECG), stress
- Exertional heatstroke, 3775, 3775t
- Exertion-induced dyskinesia, 3517
- Exfoliatin, 1051
- Exfoliative erythroderma syndrome, 141t, 147, A1
- Exhaled breath condensate, 2214–2215
- Exocrine, 2977
- Exons, 3779
- Exophthalmos, 232
- Exostoses, multiple, 524t, 3318
- Exotoxin A, 1307t, 1308
- Expanded Disability Status Score (EDSS), 3580t, 3585
- Expectation, in clinical trials, 3930t
- Expected utility theory, 3916t
- Expedited partner therapy, 1111
- Experimental allergic (autoimmune) encephalomyelitis, 3396, 3396f
- Experimental allergic neuritis, 3618
- Exportin 1, 573
- Exposomics/exposome, 6, 3998
- Expressivity, variable, 3789
- *Exserohilum,* 985, 1728
- *Exserohilum rostratum* infection, 1059, 1130t
- *EXT1/2* gene mutations, 524t, 3318
- Extended daily dialysis (EDD), 2385
- Extended-spectrum β-lactamase (ESBL), 1151, 1181, 1283, 1283t
- Extensive metabolizers, 491, 492f
- Extensor plantar response, 3381
- External beam radiation therapy, 636–637, 644, 659, 700, 728, 3047
- External concurrent occipital and trigeminal neurostimulation (eCOT-NS), 3467t, 3469
- Externalities, in behavior, 3915
- External pressure headache, 3474
- Extinction (sensory phenomenon), 174, 205
- Extracellular adherence protein (EAP), 1198
- Extracellular blood volume, 2371f
- Extracellular fluid, 2621
- Extracellular matrix (ECM), 966
- Extracellular pathogens, 967
- Extracorporeal CO$_2$ removal (ECCO$_2$ removal), 2309, 2309t
- Extracorporeal membrane oxygenation (ECMO)
  - for ARDS, 2302, 2303t
  - for hantavirus pulmonary syndrome, 996
  - in lung transplantation, 2285, 2286
  - for pulmonary edema, 2331
  - types and key features, 2309, 2309t
- Extracorporeal shockwave lithotripsy, 2450
- Extramammary Paget's disease, 609–610, 609f
- Extramedullary syndromes, 3561
- Extranodal NK/T cell lymphoma, nasal type, 857, 867–868, **874,** 2898
- Extrapyramidal reactions, 3674, 3715t
- "Extreme delta brush," 191
- Extreme endurance activities, 11
- Extremities
  - arterial diseases of. *See* Arterial diseases, of extremities
  - in cardiovascular examination, 1860, 1862–1864, 1863f
  - lymphatic disorders of. *See* Lymphedema
  - periarticular disorders of, **2973**
  - in shock, 2312
  - venous disorders of, 2179. *See also* Chronic venous disease; Venous thromboembolism (VTE)
- Extrinsic allergic alveolitis. *See* Hypersensitivity pneumonitis
- Extrinsic incubation, 1656
- Extrinsic pathway, coagulation, 470, 929f
- Extubation, of terminally ill patient, 89
- Ex vivo gene transfer, 3825–3827
- Ex vivo genome editing, 3825t, 3827
- Ex vivo lung perfusion (EVLP), 2285
- Eyach virus, 1657t
- Eye(s). *See also* Ocular disorders
  - alignment of, 221
  - "cherry red" spot in, A15
  - dry. *See* Keratoconjunctivitis sicca
  - as immune-privileged site, 2826
  - melanoma of, 232, 232f, 599f, 601t
  - movements of. *See* Ocular movements
  - nerve agent injury to, S5
  - painful or red, 222–225
  - tumors of, 232
  - vesicant injury to, S5
- Eye fields, frontal, 200f, 204, 204f
- Eye-Hand-Foot score, in leprosy, 1404–1405
- Eyelids, disorders of, 233–234, V3
- Ezetimibe, 2110, 3228, 3244t, 3251, 3257, 3647
- EZH2, 3939, 3940
- *EZH2* gene mutations, 816, 828, 864
- EZH2 inhibitors, 3940

F

- FABER (flexion, abduction, and external rotation) (Patrick's test), 121t
- Fab exchange, in IgG4-related disease, 2933
- Fabry's disease, **3355**
  - carriers of, 2425, 3789
  - clinical features of
    - abdominal pain, 3355
    - cardiomyopathy, 2010, 2010f
    - cerebrovascular disease in, 3355, 3448
    - cutaneous, 405, 3353t, 3355, A15
    - hematologic, 3353t
    - neuropathy, 3601
    - renal, 2412, 2415t, 2425
  - enzyme deficiency in, 3353t
  - genetic considerations in, 2004t, 2010, 2425, 3355, 3601–3602
  - pathogenesis of, 2425
  - prevalence of, 3355
  - renal biopsy in, 2425, A4
  - treatment of, 2010, 2425, 3355, 3602
- Facet joint pain, 119, 123t
- Facial flushing, 391
- Facial myokymia, 3555, 3578
- Facial nerve
  - anatomy of, 239f, 3553–3554, 3554f
  - examination of, 3380
  - injuries of, 247, 3573–3574
  - in taste sensation, 239
- Facial pain/numbness, 3551–3553. *See also* Trigeminal neuralgia
- Facial palsy, 172, 1235, 1450, 1452, 1453
- Facial weakness/paralysis
  - anatomic considerations in, 3554f
  - in Bell's palsy. *See* Bell's palsy
  - in facial myokymia, 3555
  - in hemifacial spasm, 3555
  - in MS, 3578
- Facilitated diffusion, 2366
- Facioscapulohumeral muscular dystrophy (FSHD), 3634, 3637f, 3643–3644
- F-actin, 2790t
- Factitious illness, 150, 230, 310, 3235–3236, 3670
- Factor V
  - deficiency of, 472, 928t, 932
  - inhibitors, 928, 935–936
  - screening assays, 474f, 475
- Factor V Leiden (factor Va)
  - early interventions for, 3804t
  - in fibrin clot formation, 470
  - genetic testing for, 3804t
  - protein C resistance and, 471
  - thrombosis risk in, 473, 474f, 941
- Factor VII
  - in coagulation cascade, 940
  - deficiency of, 472, 932
  - screening assays, 474f, 475
- Factor VIII
  - deficiency of, 472, 475, 932. *See also* Hemophilia
  - inhibitors, 930, 935–936
  - in pregnancy, 932
  - replacement, for hemophilia, 930
  - screening assays, 474f, 475
  - in VWD, 926
- Factor IX
  - deficiency of, 472. *See also* Hemophilia
  - inhibitors, 930, 935
  - in pregnancy, 932
  - replacement, for hemophilia, 930
- Factor IX Padua, 931
- Factor X
  - deficiency of, 472, 928t, 932
  - inhibitors, 935
  - screening assays, 474f, 475
- Factor Xa, 470, 940, 948
- Factor XI
  - deficiency of (hemophilia C), 928t, 932
  - in fibrin clot formation, 470, 471f
  - inhibitors, 932, 935
- Factor XII deficiency, 928t
- Factor XIII deficiency, 928t, 932
- Facultative anaerobes, 1371
- FADD (Fas-associated death domain), 539, 540f
- FADIR (flexion, adduction, and internal rotation), 121t
- Faget's sign (relative bradycardia), 959, 960t
- FAIR (flexion, adduction, and internal rotation), 121t
- Fall(s), **180**
  - assessment and management of patient with, 180–181, 3884, 3885f, 3886t
  - epidemiology of, 180, 3884
  - hyponatremia and, 349
  - interventions to reduce, 181
  - mechanical, 180
  - in older adults. *See* Older adults, falls in
  - patterns of, 181
  - recurrent, 180
  - risk factors for, 180, 180t
  - sensory deficits and, 181
  - vs. syncope, 159
- Fallopian tube cancer, 712
- False-negative tests, 24–25
- False-positive tests, 24–25
- *FAM13A,* 2253
- *FAM46A* gene mutations, 3325t, 3326
- *FAM111A,* 3286
- Famciclovir
  - adverse effects of, 1485
  - for herpes zoster, 1054t, 1485, 1486t, 1507

Famciclovir (*Cont.*):
  for HSV esophagitis, 2513
  for HSV infections, 1054t, 1109t, 1485, 1486t, 1502, 1503t, 1537
  for HSV prophylaxis, 1502, 1595t
  pharmacology of, 1485
  resistance to, 1485
  for varicella, 1507
  for viral meningitis, 1126
Familial adenomatous polyposis (polyposis coli)
  colonic polyps in, 2501f
  colorectal cancer screening in, 2496, 2502t
  early interventions for, 654, 3804t, 3805
  genetic considerations in, 157, 522–523, 524f, 524t
  genetic testing for, 3803, 3804t
  treatment of, 654
Familial amyloid polyneuropathy, 3603
Familial amyotrophic lateral sclerosis, 3521, 3522t
Familial chylomicronemia syndrome, 3240, 3241t
Familial cold autoinflammatory syndrome (FCAS), 2935t, 2936–2937
Familial combined hypolipidemia, 3241t, 3248–3249
Familial defective ApoB, 3242
Familial dysalbuminemic hyperthyroxinemia (FDH), 3024, 3024t
Familial dysbetalipoproteinemia, 3241t, 3245
Familial focal epilepsy with variable foci, 3412t
Familial gingival fibromatosis, 261
Familial glucocorticoid deficiency, 3067t
Familial hemiplegic migraine, 3462
Familial Hibernian fever. *See* Tumor necrosis factor-α receptor-associated periodic syndrome (TRAPS)
Familial hyperaldosteronism, 355. *See also* Glucocorticoid-remediable aldosteronism (GRA)
Familial hypercholesterolemia, 3241t, 3242–3243, 3252, A15
Familial hyperkalemia, 795
Familial hypobetalipoproteinemia, 3247
Familial hypocalciuric hypercalcemia (FHH), 363, 364, 3084, 3279–3280, 3804t
Familial isolated hyperparathyroidism (FIHP), 3084
Familial Mediterranean fever, **2934**
  ascites in, 328
  clinical features of, 2935t
    abdominal pain, 111, 112t, 2934
    in acute attacks, 2934–2935
    amyloidosis, 899, 2935
    arthritis, 2934
    cutaneous, 2934
    fever pattern/duration, 136, 2934
  diagnosis of, 2935
  differential diagnosis of, 2935
  early interventions for, 3804t
  genetic considerations in, 2934, 2935t
  genetic testing for, 2935, 3804t
  inflammasome molecule mutations in, 2761t
  interleukin 1 in, 462
  pathophysiology of, 2934
  treatment of, 155, 899, 2935t, 2936
Familial partial lipodystrophy, 3240–3241, 3241t
Familial polyposis of the colon. *See* Familial adenomatous polyposis (polyposis coli)
Family history, 2
Famotidine, 301, 923t, 2524, 2524t, 2808, 3660t
Fam-trastuzumab, 535t
Fam-trastuzumab deruxtecan-nxki, 566, 566t
FAN1 gene mutations, 2436, 2442, 3513
Fanconi-Bickel syndrome, 2368t, 3236, 3360t, 3363
Fanconi's anemia, 755, 810, 3785
Fanconi's syndrome
  clinical features of, 372, 2362
  hereditary tubal disease in, 2359t
  in multiple myeloma, 886–887, 2359
  tubulointerstitial nephritis in, 2356
  in Wilson's disease, 3338
Farber disease, 3354t
Far Eastern tick-borne encephalitis, 1671
Farmer's lung, 2228t, 2229, 2236t, 2239
Farnesoid X receptor, 2642, 2727
Farnesyl transferase inhibitors, 526f
Fas, 2786f, 2803
Fas-associated death domain (FADD), 539, 540f
Fasciculations, 168t, 169, 2910, S5
Fasciitis
  eosinophilic, 416, 810, 2876
  necrotizing. *See* Necrotizing fasciitis
  palmar, 2972
Fasciola/fascioliasis, **1831**
  clinical features of, 1739t, 1831, 1831t
  diagnosis of, 961t, 1831, S12
  epidemiology of, 961t, 1736, 1826t, 1831, S12
  life-cycle hosts of, S12
  pathogenesis of, 1831
  transmission of, 1826t
  treatment of, 1753, 1830t
Fasciolopsis buski, 1826t, 1830t, 1831, 1831t, S12
FAS gene mutations, 524t
FasL, 2786f
FASL gene mutations, 524t
Fas ligand, 539
Fasting hypoglycemia, 3236
Fasting, periodic, 3874
Fasting plasma glucose (FPG), 3196t, 3197
Fat(s)
  dietary, 511, 654, 3246
  malabsorption of, 309. *See also* Steatorrhea
  metabolism of, in type 2 DM, 3201–3202
  recommended intake of, 2602t
  requirement for, 2598
Fatal familial insomnia (FFI), 216, 3526t, 3528, 3528t, 3530
Fat embolism, 407–408, 2159
Fatigability, 166, 168
Fatigue, **165**. *See also* Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS)
  in anemia, 444
  approach to the patient, 167
  assessment of, 83–84, 506
  in cancer patient, 506, 507t
  in chronic hepatitis, 2675, 2685
  definition of, 165, 3634
  differential diagnosis of, 166–167
  epidemiology of, 166
  etiology of, 83, 165
  in fibromyalgia, 2963
  global considerations in, 166
  health care errors and, 42
  in heart failure, 1983
  in liver disease, 2628
  in medical residents, 52
  in MS, 3578, 3588
  in myasthenia gravis, 3626
  in myopathies, 3634
  in primary biliary cholangitis, 2711
  prognosis of, 168
  radiation-induced, 554
  vs. sleepiness, 213
  in terminally ill patient, 83–84, 91t
  treatment of, 84, 167–168, 506
  vs. weakness, 165
Fat necrosis, 406
Fatty acid(s), 2541–2542, 2549t. *See also* Free fatty acids (FFAs); Lipid(s)
Fatty acid diarrhea, 2543t
Fatty acid mediators, 2218–2219
Fatty liver disease
  alcoholic, 2697. *See also* Alcohol-associated liver disease
  drug-induced, 2669t
  in IBD, 2562t, 2563
  nonalcoholic. *See* Metabolic dysfunction–associated steatohepatitis (MASH); Metabolic dysfunction–associated steatotic liver disease (MASLD)
  in pregnancy, 3903
Favipiravir, 1489
Favism, 800, 801t
Fazio-Londe syndrome, 3524
FBN1 gene mutations, 3331, 3332t
FCAS (familial cold autoinflammatory syndrome), 2935t, 2936
Fc (Ig Fc) receptors, 2755, 2762, 2769
FcRn, 2792t
FcRn inhibitors, 3631, 3632t
FcγRIII/III, 2786f
FDH (familial dysalbuminemic hyperthyroxinemia), 3024, 3024t
Feather hydroid envenomation, 3725
Febrile neutropenia, 574–575, 575t, 580–581, 1312
Febrile neutrophilic dermatosis. *See* Sweet syndrome
Febrile nonhemolytic transfusion reaction, 911
Febrile seizures, 137, 3411, 3413
Febuxostat, 593–594, 2453, 2960
Fecal DNA testing, 516t, 656
Fecal elastase, 2739t, 2742
Fecal immunochemical testing, 516t, 517
Fecal impaction, 305, 308
Fecal incontinence, **2583**
  clinical features of, 2583
  constipation and, 2584
  epidemiology of, 2583
  evaluation of, 2583
  in MS, 3588
  pathophysiology of, 305, 2583, 2583t
  rectal prolapse and, 2582
  in systemic sclerosis, 2874
  treatment of, 2583–2584
Fecalith, appendiceal, 2596, 2596f
Fecal microbiota transplantation, 1085, 1085t, 2464, 3838–3839, 3841
Fecal occult blood
  in colorectal cancer screening, 38, 39t, 320, 516t, 517, 655
  endoscopic evaluation of, 2496
Feces. *See* Stool
FECH, 3343, 3349
FECH gene mutations, 3342, 3342t, 3350
Fechtner syndrome, 924
Fecundability, 3149
Federal Trade Commission, 3926
Fedratinib, 569t, 573, 821, 823
Feedback control, 2986–2987, 2986f
Fee-for-service reimbursement, 23
Feinstein, Alvan, 47
Felbamate, 3421t, 3713t
Felodipine
  adverse effects of, 2101t, 2144t
  for heart failure, 1996
  for hypertension, 2144t
  for ischemic heart disease, 2101, 2101t
Felty's syndrome, 458, 480, 2841
Female orgasmic disorder, 3161
Female pseudohermaphroditism. *See* 46,XX disorders of sex development
Female reproductive system disorders
  cancer. *See specific organs*
  delayed puberty, 3131–3132, 3131t
  in diabetes mellitus, 3221
  infertility. *See* Infertility
  menstrual disorders. *See* Menstrual disorders

  obesity and, 3187
  ovarian disorders. *See* Ovary, disorders of
  precocious puberty, 712, 3130–3131, 3130–3131t
  sex chromosome disorders. *See* Sex chromosome disorders
  sexual dysfunction, 3160–3162, 3227
  tubal disease, 3151
  in veterans, S8
  46,XX disorders. *See* 46,XX disorders of sex development
Femoral artery
  for PCI access, 2127
  percutaneous intervention for occlusion of, 2132f
  pulse, in aortic regurgitation, 2044
Femoral-femoral bypass, 2174
Femoral fracture, 3304, 3306
Femoral head, aseptic necrosis of, 3355
Femoral head, avascular necrosis of, 2968
Femoral neuropathy, 3614
Femoral-popliteal artery disease, 2172. *See also* Peripheral artery disease
Femoral stretch test, 120, 121t
FeNa (fractional excretion of sodium), 2381
Fenfluramine, 683, 3415, 3421t, 3660t
FeNO (fraction of exhaled nitric oxide), 2214–2215, 2223
Fenofibrate, 3244t, 3258
Fenoprofen, 97t
Fentanyl, 80f, 97t, 98, 100, 3688, 3690
Fermentation, 1371
FERMT3 gene mutations, 460
Ferric citrate, 767t, 769
Ferric maltol, 767t
Ferriman and Gallwey scale for hirsutism, 3138, 3139f
Ferritin
  functions of, 762
  serum
    in alcohol-associated liver disease, 3336t
    in anemia of inflammation, 767t
    factors affecting, 449
    in hemochromatosis, 3335f, 3336, 3336t
    in hemophagocytic lymphohistiocytosis, 464, 466
    in iron-deficiency states, 765, 767t
    normal, 449
    in thalassemia, 767t
Ferrochelatase, 3341t, 3342t
Ferroportin, 762–763, 3334, 3334f, 3335
Ferumoxytol, 767t
Fetal alcohol spectrum disorder, 3677
Fetal alcohol syndrome, 3677
Fetal hemoglobin (HbF), 762, 771. *See also* Hereditary persistence of fetal hemoglobin (HPFH)
Fetal nigral dopamine cells, 3505
Fetor hepaticus, 2630
FEV$_1$ (forced expiratory volume in 1 s), 2207
  age-related changes in, 2253–2254, 2254f
  in COPD, 2251–2252
  in interstitial lung disease, 2262
  in respiratory disorders, 2206f
  smoking and, 2252, 2252f
Fever, **133**. *See also* Acute febrile illness; Fever of unknown origin (FUO)
  after transfusion, 908t
  approach to the patient, 135–136
  in autoimmune/autoinflammatory disease, 136
  as cardinal sign, 135, 135t
  central, 189
  in coma, 189
  definition of, 134
  factitious, 150
  fraudulent, 150
  heart rate and, 959
  vs. hyperthermia, 134
  hypothalamic, 134
  hypoxia and, 277
  with myalgia, 1665–1668, 1666t
  in parasitic infections, 1738, 1740t
  pathogenesis of, 134–135, 135f
  Pel-Ebstein, 136, 869
  quadriplegic, 3569
  with rash, **137,** 138–145t, 409, 1248, **A1**. *See also* Rash
  in recipient of anticytokine therapy, 136
  recurrent, 136. *See also* Hereditary recurrent fever syndromes
  relapsing. *See* Relapsing fever
  returned traveler, 1316
  in returned traveler, 1317
  in rheumatic fever, 2857
  in schistosomiasis, 1828–1829
  in splenectomized patient, 482
  treatment of, 136–137
  water loss in, 2598
Fever of unknown origin (FUO), **149**. *See also* Acute febrile illness
  approach to the patient, 152–155, 152f, 155f
  definition of, 149
  differential diagnosis of, 150, 151t
  etiology and epidemiology of, 149–150, 149t, 151t
  global considerations in, 149, 149t
  prognosis of, 155
  treatment of, 154–155
Fexinidazole, 1795, 1796, 1796t
Fexofenadine, 2808
Fexolinetant, 510, 3143
FEZ1 gene mutations, 3111t, 3112
FFAs. *See* Free fatty acids (FFAs)
FFI (fatal familial insomnia), 216, 3526t, 3528, 3528t
FFR (fractional flow reserve), 1909, 1910f, A11
FGF2, 3404
FGF8 gene mutations, 3111t, 3112
FGF14 gene mutations, 3497
FGF17 gene mutations, 3111t, 3112
FGF-19 (fibroblast growth factor-19), 308
FGF19 gene mutations, 660, 661t
FGF-21 (fibroblast growth factor-21), 3871
FGF23 (fibroblast growth factor 23), 740, 2369, 2390, 3319
FGF23 gene mutations, 3264, 3319
FGFR1 gene mutations, 617, 3111t, 3112
FGFR1-mutated eosinophilia, 879–880
FGFR2 gene mutations, 533t, 667, 668, 3787
FGFR3 gene mutations, 691, 3331
FGFR gene mutations, 720
FH gene mutations, 687t
FHH (familial hypocalciuric hypercalcemia), 363, 3084, 3279–3280, 3804t
FHIT gene mutations, 617
FIB-4 method, hepatic fibrosis assessment, 2632t, 2636, 2704–2705
Fiber, dietary. *See* Dietary fiber
Fibric acid derivatives (fibrates)
  adverse effects of, 3244t, 3250, 3648t
  for hypertriglyceridemia, 3244t, 3250
  for metabolic syndrome, 3258
Fibrillarin, 2790t, 2864t
Fibrillary glomerulopathy, 2425, A4
Fibrillary-immunotactoid glomerulopathy, 2415t
Fibrillation potentials, EMG, 169
Fibrillin-1, 2166, 3321t
Fibrillin-2, 3321t
Fibrillin aggregates, 3323
Fibrin, 471, 921, 942, 942f
Fibrin clot, 470, 471f
Fibrin degradation products, 934
Fibrin(ogen) degradation products, 471, 471f, 475
Fibrinogen, 471, 472, 928t, 943f
Fibrinolysis
  in DIC, 933, 933f
  regulation of, 471–472, 471f
  thrombosis and, 941
Fibrinolytic drugs, **954**
  actions of, 954, 954f
  for acute limb ischemia, 2175
  adverse effects of, 2120
  alteplase. *See* Alteplase (rtPA)
  anistreplase, 955
  contraindications to, 2120, 2164
  for DVT, 2161–2162
  indications for, 954
  for pulmonary embolism, 2164
  reteplase. *See* Reteplase
  for STEMI, 2119f, 2120
  streptokinase, 954, 955f, 2120
  tenecteplase. *See* Tenecteplase
  urokinase, 955
Fibroblast growth factor-19 (FGF-19), 308, 2541
Fibroblast growth factor-21 (FGF-21), 3871, 3873f
Fibroblast growth factor-23 (FGF-23), 2369, 2390, 3259, 3264, 3319
Fibroblast growth factor signaling, 572
Fibrocystin/polyductin, 2431f, 2433
Fibrodysplasia ossificans progressiva, 3319
Fibroelastoma, papillary, 2086
Fibrofolliculomas, 404
Fibrogenesis imperfecta ossium, 3317
Fibroids, uterine (leiomyoma), 3136, 3137, 3152
Fibrolamellar hepatocellular carcinoma, 670
Fibromas
  cardiac, 1898, 1900f, 2086
  ovarian, 712
  paraneoplastic syndromes in, 737t
Fibromuscular dysplasia, 2139, 2150, 2174, 2360, 3447
Fibromyalgia, **2963**
  approach to the patient, 2965
  clinical features of, 2941, 2963–2964, 2964f, 3636
  comorbidities of, 2941, 2963
  definition of, 2963
  diagnosis of, 2964–2965, 2965t
  differential diagnosis of, 1452, 2964, 2965t
  epidemiology of, 2941, 2963
  genetic considerations in, 2964–2965
  in HIV infection, 1604
  IC/BPS and, 332
  physiology of, 2964–2965
  treatment of, 2965, 2966t
Fibronectin, 2790t, 3321t
Fibronectin-binding protein, 1038
Fibrosis
  pulmonary. *See* Pulmonary fibrosis
  in systemic sclerosis, 2862, 2863f, 2864–2865, 2865f
Fibro Test method, hepatic fibrosis assessment, 2632t, 2636
Fibrous dysplasia, 3317, 3317f. *See also* McCune-Albright syndrome
Fibrous histiocytoma, malignant, 727. *See also* Soft tissue sarcomas
FIC1 gene mutations, 2642
Fick method, cardiac output measurement, 1906
Fidanacogene elaparvovec, 3824t
Fidaxomicin, 583, 1084–1085, 1085t
Fiduciary obligations, 71
Fifth disease. *See* Erythema infectiosum
Figurate skin lesions, 394, 395t
FIHP (familial isolated hyperparathyroidism), 3084
Filaggrin, 383, 432
Filamentous hemagglutinin, 1277, 1278

Filariae/filarial infections. *See* Nematodes/nematode infections, filarial
Filarial dance sign, 1821, 1825v
Fildes, Luke, 1f
Filiform warts, 390
Filoviridae
  genome of, 1680
  global considerations in, 993
  laboratory safety precautions for, 1680
  phylogeny/evolution of, 1681f
  replication strategies of, 1479
  structure of, 1478f, 1480t, 1680, 1681f
  taxonomy of, 1680f
Filovirid infections
  clinical features of, 993, 1685
  complications of, 1686
  control and prevention of, 1686–1687
  diagnosis of, 1685
  differential diagnosis of, 1685
  epidemiology of, 1680–1681, 1682–1683f
  etiology of. *See* Filoviridae
  pathogenesis of, 1683–1685
  prognosis of, 1686
  treatment of, 1685–1686
Filovirid vaccine, 1687
Fimbriae. *See* Pili (fimbriae), bacterial
Financial abuse, 3895t
Financial incentives, 73
Finasteride
  for androgenetic alopecia, 395t
  for hirsutism, 3141
  for lower urinary tract symptoms, 3175
  for prostate cancer prevention, 513
Fine-needle aspiration biopsy, 478, 553, 2213–2214, 3048–3049
Finerenone, 2139, 2144t, 2146, 2424, 3258
Finger(s)
  clubbing of. *See* Clubbing, of digits
  diagnostic clues in, 50t
  in gouty arthritis, A15
Finger anomia, 203
Finger-to-nose testing, 3382
Fingolimod, 3586t
Finkelstein sign/test, 2974
Finkelstein's sign/test, 2943
Fire ants, 3736
Fire coral envenomation, 3722
First-pass elimination/effect, 484, 484f, 1167
FISH (fluorescence in situ hybridization), 3796
Fisher's syndrome, 235, V3
Fish eye disease, 3241t, 3248
Fish meal worker's lung, 2228t
Fish odor syndrome (trimethylaminuria), 2610
Fish oils. *See* Omega-3 fatty acids (fish oils)
Fish tapeworm, 1838–1839
Fish tapeworm *(Dibothriocephalus latus),* 787, 1736, 1739t, 1838–1839, S12
Fissured (scrotal) tongue, 265t, A3
Fissures, anal, **2586**
Fistula(s). *See also* Arteriovenous fistula
  carotid cavernous, 233
  CSF–venous, A16
  gallbladder, 2733
  hepatobronchial, 1756
  pancreaticopleural, 2748f
Fistula in ano, 1389, **2585,** 2585f
Fitz-Hugh–Curtis syndrome (perihepatitis), 1105, 1471
Fitzpatrick skin types, 430, 430t
5-day fever (trench fever), 1353t, 1354
Fixed drug eruption, 423, 423f
FKBP10 gene mutations, 3324, 3325t
FKBP14 gene mutations, 3328t, 3330
Flaccidity, 168
Flagella
  of bacteria, 966, 966t
  of *P. aeruginosa,* 1307, 1307t
Flagellin, 971
Flail leaflet, 286
FLAIR. *See* Fluid-attenuated inversion recovery (FLAIR) MRI
Flatus/flatulence, 326, 2571
Flat warts, 390
Flatworms/flatworm infections. *See* Cestodes/cestode infections; Trematodes/trematode infections
Flaviviridae
  clinical syndromes of, 1666t, 1671
  geographic distribution, 1664–1665t
  global considerations in, 993
  replication strategies of, 1479
  reservoirs and vectors of, 1656, 1659–1660t
  structure of, 1478f, 1480t, 1656
Flavoenzymes, 2608
Flavonoids, **2610**
Flavopiridol, 526f
Flavor enhancers, 243
FLCN gene mutations, 687, 687t
Flea(s)
  bites, 3736
  cat, 1351, 3736
  dog, 3736
  in endemic typhus. *See* Endemic (murine) typhus
  in plague, S4
Flea-borne spotted fever, 1455t, 1458
Flecainide
  actions of, 1916t
  adverse effects of, 1916t, 1920t, 1952
  for atrial fibrillation, 1952
  for AV nodal reentry tachycardia, 1940
  for focal atrial tachycardia, 1938
  metabolism of, 485t
  for ventricular arrhythmias, 1958
Flesh fly, 3733
Flexal virus, 1657t
Flexible sigmoidoscopy. *See* Sigmoidoscopy
Flexion, abduction, and external rotation (FABER) (Patrick's test), 121t
Flexion, adduction, and internal rotation (FADIR; FAIR), 121t
Flexner Report, 52
Flibanserin, 3162
FLI-EWSR1 gene mutations, 521t
Flinders Island spotted fever, 1458
FLIP, 541
FLIP (functional lumen imaging probe), 296, 2504–2505
FLNC gene mutations, 872, 2004t
Floaters, 229
Floppy iris syndrome, 2456
Floppy-valve syndrome. *See* Mitral valve prolapse
Flow-limited breath, 2279t, 2280f
Flow limit, of lung, 2203
Flow-volume loop, 2203f, 2205, 2207
Floxuridine, 2667
FLRT3 gene mutations, 3112, 3112t
FLT3 gene mutations, 520t, 826, 832
FLT3 inhibitors, 420, 753t, 757
FLT3-ITD gene mutations, 834
Flubendazole, 1748
Flucloxacillin, 491, 493, 1043t, 2667
Fluconazole
  advantages of, 1691
  adverse effects of, 423, 923t, 1691, 1742t
  for blastomycosis, 1702, 1703t
  for *Candida* infections
    disseminated, 1711t
    endocarditis, 1712
    esophageal, 1710t, 2513
    liver abscess, 1075
    oral, 1710t
    resistance to, 1711t
    tissue-invasive, 2410
    urinary tract, 1095
    vulvovaginal, 1101t, 1103, 1691, 1710t
  for *Candida* prophylaxis, 918, 918t, 1155, 1164t, 1595t, 1691, 1712
  for coccidioidomycosis, 1127, 1698
  for cryptococcosis, 1127, 1607, 1705
  for cryptococcosis prophylaxis, 1706
  for cutaneous leishmaniasis, 1787
  drug interactions of, 485t, 1742t, 2408, 2721, 3213
  for fungal infection prophylaxis, 1691
  for hypercalcemia, 3285
  indications for, 1691
  for *Malassezia* infections, 1723
  in pregnancy and lactation, 1742t
  prophylactic uses in HIV infection, 1594t
  susceptibility testing for, S11
Flucytosine
  actions of, 1692
  adverse effects of, 457, 1692
  for *Candida* endocarditis, 1044
  for *Candida* endophthalmitis, 1712
  for *Candida* UTIs, 1095
  for cryptococcal meningitis, 1127, 1607, 1705
  indications for, 1692
Fludarabine
  action and targets of, 565
  adverse effects of, 562t, 565, 575
    hemolytic anemia, 802t
    long-term, 753t, 755
    neurologic, 726, 726t
    pulmonary, 595, 754
  for B-cell prolymphocytic leukemia, 872
  for CLL, 855
  in HCT preparation, 915
  interactions and issues, 562t
  for Waldenström's macroglobulinemia, 894
Fludrocortisone
  for adrenal insufficiency, 3070
  for hypoaldosteronism, 373
  for orthostatic hypotension, 160, 3506, 3547, 3660t
  for postural orthostatic tachycardia syndrome, 1936, 3548
  for syncope, 158
Fludrocortisone suppression test, 3061
Fluid-attenuated inversion recovery (FLAIR) MRI
  in ARIA, 3480, 3481f
  in Balo's concentric sclerosis, A16
  in Behçet syndrome, A16
  in brain abscess, A16
  in brain tumors, 716, 718f, A16
  in cerebral arteriopathy, A16
  in cerebral small-vessel disease, 3490, 3490f
  characteristics of, 3387
  in CNS aspergillosis, A16
  in Creutzfeldt-Jakob disease, 3530, 3531f
  in Hashimoto's encephalopathy, A16
  in Huntington's disease, 3513f
  in limbic encephalitis, A16
  in MELAS, A16
  in MS, 3589f
  in MSA, 3549f
  in neurofibromatosis type 1, A16
  in viral encephalitis, 1113–1114, 1113f, 1121
Fluid overload, 912
Fluid requirements, 2598–2599
Fluid restriction, 738, 2302, 3020
Fluid retention, 2626
Fluid therapy/resuscitation
  for alcoholic ketoacidosis, 370
  for constipation, 82

  for diabetic ketoacidosis, 370
  for diarrhea, 307
  for DKA, 3218
  for HHS, 3219
  for hypercalcemia, 364, 3284t
  for hypovolemia, 347
  for meningococcal disease, 1249
  for prevention of contrast-induced nephropathy, 1904
  for sepsis/septic shock, 2322t, 2323
  for shock, 2314–2315
Flukes. *See* Trematodes/trematode infections
Flumazenil, 3691
Flunarizine, 3469, 3470t, 3475, 3497
Flunitrazepam, 3713t
Fluorescence in situ hybridization (FISH), 3796, 3821
Fluorescent auramine microscopy, S12
Fluorescent treponemal antibody–absorbed (FTA-ABS) test, 1433
Fluoridation, 261
Fluoride, 358, 2601t, 2614t, 2615
Fluorochrome stains, S11
Fluoropyrimidine, 657
Fluoroquinolones
  absorption of, 1424
  actions of, 1166–1167, 1177, 1182t
  adverse effects of, 1093, 1093t, 1171t, 1177, 1424
    in children, 1263
    mechanisms of, 422
    nails, 420
    onycholysis, 421
    photosensitivity, 420, 434t
    in pregnancy, 1095, 1263
    vesicles/bullae, 400
  for *Aeromonas* spp. infections, 1268, 1268t
  for anaerobic bacterial infections, 1379
  for bite-wound infections, 1144
  for community-acquired pneumonia, 1027t, 1028
  contraindications to, 1474
  for *C. trachomatis* infections, 1474
  for cystitis, 1093, 1093t
  drug interactions of, 1172t, 3213
  efficacy of, 1168
  empirical uses of, 963t
  for enteric fever, 1317–1318
  for gram-negative bacterial infections, 1268t
  for *H. influenzae* infections, 1263
  indications for, 1177–1178
  for isoniazid-resistant TB, 1398
  for *Legionella* spp. infections, 1275, 1275t
  for MDR-TB, 1421t
  for *M. pneumoniae* infections, 1466
  for plague, 1346
  in pregnancy and lactation, 1169t
  prophylactic uses of, 1180t
  for prostatitis, 1094
  for pyelonephritis, 1094
  resistance to, 1093
    in anaerobic bacteria, 1379
    in *Campylobacter* spp., 1306
    detection of, 981
    in *E. coli,* 1027, 1087, 1287
    in *Enterobacter,* 1292
    in *Enterobacterales,* 1283t, 1284
    in *Klebsiella* spp., 1291
    mechanisms of, 1182t, 1185
    in mycobacteria, 1397, 1424
    in pneumococci, 1026
    in *Proteus* spp., 1291
    in *Salmonella,* 1318
    in *Shigella,* 1324
    in *V. cholera,* 1329
  for *Serratia* infections, 1292
  for *Shigella* infections, 1324
  for TB, 1420t, 1424
  for yersiniosis, 1350
Fluoroscopy, in dysphagia, 294, 296
Fluorosis, 2615
5-Fluorouracil (5-FU)
  for actinic keratoses, 609
  actions of, 565
  adverse effects of, 434t, 562t, 565
    cardiotoxicity, 2017
    cutaneous, 400, 420
    encephalopathy, 2350t
    gastrointestinal, 597
    long-term, 753t
    mucositis, 508
    nail disorders, 421
    nausea and vomiting, 575
    neurologic, 726t
    in pregnancy, 638
    steatosis, 2703t
  for anal cancer, 659
  for basal cell carcinoma, 607
  for bladder cancer, 693
  for breast cancer, 637
  for colorectal cancer, 658
  for gastric cancer, 644, 649
  for gastrinoma, 3082
  genetic variations in response to, 493, 641, E72t
  for head and neck cancer, 613
  for insulinoma, 3082
  interactions and issues, 562t
  for metastatic breast cancer, 641
  for pancreatic cancer, 674, 676t
  timing of administration of, 3954
Fluoxetine
  adverse effects of, 2703t, 3659t, 3666
  for cataplexy, 215
  for depression, 85, 504, 3588, 3659t, 3667
  drug interactions of, 485t, 489t, 492
  for menopausal symptoms, 3143
  metabolism of, 485t, 492
  for obsessive-compulsive disorder, 3664
  for pain, 100
  for panic disorder, 3658
  pharmacology of, 3667
Fluoxymesterone, 567, 3119
Fluphenazine, 3673t
Flurazepam, 3660t, 3713t
Flushing
  in acne rosacea, 391
  in carcinoid syndrome, 681
  nicotinic acid therapy and, 2609
Flutamide, 567, 702, 753t, 3135, 3141
Fluvastatin, 3244t. *See also* Statins
Fluvoxamine, 3659t, 3664, E72t
Fly bites, 3736, 3737f
Fly infestation, 3732–3733
FMR1/2 gene mutations, 3134, 3403, 3941
fMRI (functional MRI), 95f, 3390, 3933, 3966, 3966f
Foam cells, 404
Focal ataxia, 3533–3534
Focal atrial tachycardia, 1932t, 1937–1938, 1937f, 1938f
Focal nodular hyperplasia, of liver, 670
Focal segmental glomerulosclerosis
  clinical features of, 2415t, 2422
  COVID-19 related, 2429
  genetic considerations in, 2411–2412
  HIV-related, 2357, 2428
  pathogenesis of, 2421–2422, 2422f, 2422t
  renal biopsy in, 2422, A4
  treatment of, 2422–2423
Focal seizures, 3409, 3420, 3423–3424, 3456–3457
Focal sensory seizures, 176
FODMAP intolerance, 309
FODMAPs, 2575, 2576–2577, 2576t, 2577f, 2577t
Fogo selvagem, 410
Folate, **781**
  absorption of, 781
  for alcohol-induced acidosis, S1
  cobalamin and, 782
  deficiency of. *See* Folate deficiency
  dietary sources of, 781
  functions of, 782, 782t, 783f, 2606f
  red cell, 789
  requirement/recommended intake of, 781, 2600t
  serum, 789
  structure of, 2606f
  supplements. *See* Folate supplements
  transport in blood, 781–782
Folate deficiency. *See also* Megaloblastic anemia
  anemia in, 449
  causes of, 787–788, 787t
  clinical features of, 784, 2605t
  diagnosis of, 789
  drug-induced, 788
  hematologic findings in, 784–785
  jaundice in, 323
  myelopathy in, 3568
  neural tube defects and, 784
  neutrophil segmentation in, 453
  optic neuropathy in, 228
  peripheral blood smear in, 439f
  in pregnancy, 784, 790, 1844
  in premature infants, 790
  prophylaxis for, 790
  skin manifestations of, 400
  small-intestinal mucosal biopsies in, 2550t
  treatment of, 790
  in tropical sprue, 2546
Folate supplements
  in infants and children, 790
  for methotrexate toxicity prophylaxis, 2895
  for paroxysmal nocturnal hemoglobinuria, 806
  in pregnancy, 40t, 784, 790, 3903
  prophylactic, 790
  for pyruvate kinase deficiency, 796
  for tropical sprue, 2546
Folate synthesis antagonists, 788, 1166, 1182t, 1183f
FOLFIRINOX regimen, 674, 675, 675t, 676t
FOLFIRI regimen, 658
FOLFOXIRI-bevacizumab regimen, 658
FOLFOX regimen, 649, 658, 669f
Folic acid, 781. *See also* Folate
Folinic acid
  for folate deficiency, 788, 790
  for methotrexate toxicity prophylaxis, 2895
  for pancreatic cancer, 674, 676t
FOLIRINOX-bevacizumab regimen, 658
Follicle(s)
  Graafian, 3126f
  ovarian, 3125–3126, 3125–3126f
Follicle-stimulating hormone (FSH)
  actions of, 2991
  deficiency of, 2996t
  expression and regulation of, 2988t
  in females, age-related changes in levels of, 3126–3127, 3127f
  GnRH and, 2987
  laboratory evaluation of, 2995t, 3107
  in male pubertal development, 3102
  in menstrual cycle regulation, 3128–3129, 3129f
  in ovarian follicle development, 3126
  in perimenopause, 3142, 3142f
  secretion of, 2991
  in spermatogenesis, 3106
  synthesis of, 2991

- Follicle-stimulating hormone (FSH) receptor, 2985t, 3106, 3789
- Follicle-stimulating hormone (FSH) therapy, 3119, 3134
- Follicular dendritic cell(s), 1574
- Follicular dendritic cell sarcoma, 882
- Follicular lymphoma, **863**
  - bone marrow examination in, A6
  - clinical features of, 864
  - diagnosis of, 864–865, 864f
  - differential diagnosis of, 864
  - frequency of, 858f
  - genetic considerations in, 521t, 860t, 863–864
  - immunophenotype of, 852t, 872t
  - lymph node examination in, A6
  - pathogenesis of, 859f
  - prognosis of, 861t, 864
  - transformation to DLBCL, 865
  - treatment of, 864–865
- Follicular plugging, 415
- Follicular thyroid cancer, 3046–3047
- Folliculitis, **395**
  - *Candida,* 1708
  - clinical features of, 378t, 380f, 1198
  - eosinophilic, 395
  - etiology of, 1050t, 1051
  - in HIV infection, 1605
  - hot-tub, 142t, 147, 395, 1051, 1312
  - *Malassezia,* 1723
- Folliculitis decalvans, 394, 395t
- Follistatin, 2984
- Fomepizole, 192, 371, S1
- Fomites, 1538
- Fondaparinux, **949**
  - actions of, 946f, 949
  - adverse effects of, 950
  - for anticoagulation in HIT, 923
  - dosage of, S13
  - vs. low-molecular-weight heparin, 949t
  - management before lumbar puncture, S3
  - for NSTE-ACS, 2111
  - pharmacology of, 949
  - for STEMI, 2121
  - for VTE/PE treatment, 2162, 2163t
  - for VTE prophylaxis, 2164t, 3910
- *Fonsecaea,* 1725
- Fontan procedure, 2076, 2077f
- Food-exchange list, 2603
- Food-frequency questionnaire, 2603
- Food guide, 2603
- Food-related illness
  - bacterial, 1080, 1080t
    - *B. cereus,* 1080, 1080t
    - *Campylobacter,* 980t, 1080t, 1304
    - *C. botulinum,* 1233–1235, 1234t, 1237
    - *C. jejuni. See Campylobacter* spp. infections
    - *C. perfringens,* 1080, 1080t, 1240
    - *E. coli. See Escherichia coli* infections, intestinal
    - genome sequencing for tracking, 977
    - *L. monocytogenes,* 305, 1227, 1228t
    - non-01/039 *V. cholera,* 1330, 1330t
    - *Salmonella. See Salmonella* spp. infections, food-borne
    - *S. aureus. See Staphylococcus aureus* infections, food-borne
    - *Shigella. See Shigella* spp. infections
    - *V. cholerae. See* Cholera
    - *V. parahaemolyticus,* 1080t, 1330, 1330t
    - *V. vulnificus,* 992–993, 1051, 1330–1331, 1330t
    - *Yersinia* spp., 1348
  - in cancer patient, 584
  - ciguatera poisoning, 306, 3726–3727, 3727t
  - diarrhea in, 305, 1080, 1080t. *See also* Diarrhea, infectious
  - diarrhetic shellfish poisoning, 3728
  - differential diagnosis of, 305, 306t
  - domoic acid intoxication, 3414, 3727–3728
  - nematode-borne. *See* Nematodes/nematode infections
  - nonbacterial, 1080
  - paralytic shellfish poisoning, 3727
  - prevention of, 584
  - scombroid poisoning, 306, 3726
  - trematode-borne, 1830–1832, 1830t, 1831t
  - viral, 306t
- Food restriction, 3951
- Foot drop, 3614
- Foot examination, 3204, 3877
- Foot, frostbite of, A5
- Foot osteomyelitis, 1068–1069, 1068f
- Foramen magnum syndrome, 3561
- Foraminal herniation, 187f, 188
- Forbes syndrome (type IIIa GSD), 3360t
- Forced desynchrony protocols, 3947
- Forced expiratory volume in 1 s (FEV_1). *See* FEV_1
- Forced vital capacity (FVC), 2206f, 2207, 2262
- Forchheimer spots, 137
- Fordyce spots, 264t
- Fordyce's spots, 137
- Forearm exercise test, 3638
- Foreign body
  - in esophagus, 2484–2485, 2492f, 2514
  - in eye, 222
  - granuloma, 406
- for gastric lymphoma, 650
- Formaldehyde exposure, 384, 2239t
- Formalin-ether sedimentation, S12
- Formiminoglutamic aciduria, 3368t
- Formoterol, 2224, 2225t, 2226, 2255
- for multiple myeloma, 890t, 891–892
- Fortification spectra, 230, 3378
- Fort Sherman virus, 1658t
- Fosaprepitant, 83
- *FOSB* gene, 3942
- Foscarnet
  - actions of, 1116, 1486
  - adverse effects of
    - electrolyte abnormalities, 1160
    - electrolyte disturbances, 1486t, 1516
    - hypokalemia, 355
    - nephrotoxicity, 1116, 1486, 1486t, 1516, 1601, 2377
  - for CMV infection, 1016, 1160, 1486–1487, 1486t, 1516, 3566
  - for CMV prophylaxis, 918
  - for HHV-6 infection, 1156, 1487
  - for HSV infection, 1486, 1502, 1503t
  - resistance to, 1486–1487
- Fosfomycin
  - for *A. baumannii* infections, 1297t
  - actions of, 1166, 1179, 1182t
  - adverse effects of, 1171t, 1179
  - for *Citrobacter* infections, 1293
  - for *Enterobacterales* infections, 1283, 1283t
  - indications for, 1174t, 1179
  - resistance to, 1093, 1174t, 1179, 1182t, 1283, 1283t, 1291
  - for UTIs, 1093, 1093t, 1179, 1219t, 1220t, 1287
- Fosmanogepix, 1698, 1729, 1730
- Fosmidomycin, 1748
- Fosphenytoin, 2298
- Fostamatinib, 802
- Fostemsavir, 1618t, 1621
- Foster Kennedy syndrome, 242
- Founder effect, 3787
- 4Kscore Test, 697
- Fournier's gangrene, 1052, 1377, 2321t
- Fovea, 220
- Foveal hypoplasia, 3374t
- Foveation, 220
- Foville's syndrome, 235
- *FOXC2* gene mutations, 2183
- Foxglove, 358, 3712t
- *FOXO* transcription factors, 3871, 3873f
- FOXP3, 2846, 3092
- *FOXP3* gene mutations, 2786f, 2789, 2803, 2823t, 3092, 3942
- FPG (fasting plasma glucose), 3196t, 3197
- Fractalkine, 2768t
- Fractional excretion of sodium (FeNa), 2313, 2381
- Fractional flow reserve (FFR), 1909, 1910f, 2130, A11, A11v
- Fraction of exhaled nitric oxide (FeNO), 2214–2215, 2223
- Fracture(s)
  - atypical femur, 3304, 3306, 3307
  - Colles', 3292f
  - hip. *See* Hip(s), fracture of
  - hyponatremia and, 349
  - IBD–related, 2563
  - Lisfranc, 2969
  - in osteogenesis imperfecta, 3324
  - osteoporotic. *See* Osteoporosis, fractures in
  - in Paget's disease of bone, 3313
  - pelvic, 3291
  - risk assessment for, 3145, 3297, 3298f, 3300
  - skull, 3571–3572
  - temporal bone, 247
  - vertebral. *See* Vertebral fractures
  - wrist, 3291
- Fracture Risk Assessment (FRAX) score, 3145, 3297, 3298f
- Fragile X pre-mutation, 179, 3134
- Fragile X syndrome, 3403, 3652, 3786, 3793t, 3941
- Frailty, 2717, 3873, 3893–3894
- Frameshift mutation, 3784, 3786f
- Framework Convention on Tobacco Control (WHO), 3851
- Framing effects, 3916–3917
- *Francisella* spp., 1335–1336
- *Francisella tularensis. See also* Tularemia
  - autophagy evasion by, 973
  - as biologic weapon, 1341, S4
  - characteristics of, 1335
  - epidemiology of, 1336
  - immune invasion by, 970f, 971, 1337
  - reservoirs and modes of transmission, 1336–1337
  - secretion systems of, 968
  - subtypes of, 1335, 1336
- Franklin's disease (gamma heavy chain disease), 894
- *Frataxin* gene, 3537
- Fraudulent fever, 150
- FRAX (Fracture Risk Assessment) score, 3145, 3297, 3298f
- FRC (functional residual capacity), 2202, 2205, 2206f, 2296
- Freckles (ephelides), 399
- Free erythrocyte protoporphyrin, 3345
- Free fatty acids (FFAs)
  - in cardiac metabolism, 1853–1854
  - in glucose utilization, 3202
  - in insulin resistance, 3254, 3254f
- Free radical theory of aging, 3871
- Free T_3 index, 3026
- Free T_4 index, 3026
- Free water deficit, S1
- Freezing gait, 178, 180, 3507. *See also* Parkinson's disease
- Freiberg's test/sign, 120, 121t
- Fremanezumab, 3469, 3470t
- Frenzel eyeglasses, 163
- Fresh-frozen plasma, 934, 935, 2322t, 3456

Friedreich's ataxia, **3536**
  clinical features of, 3536
  genetic considerations in, 2004t, 3536–3537, 3793t
  imaging in, 3536, 3536f
  mitochondrial dysfunction in, 3400
  treatment of, 3537
Froment sign, 3613
Frontal abulic syndrome, 208
Frontal disinhibition syndrome, 208
Frontal eye fields, 204, 204f, 235
Frontal gait disorder, 178, 179t
Frontal lobe syndromes, 208–209
Frontal network syndrome, 209
Frontotemporal dementia (FTD), **3485**
  ALS and, 3522t, 3523
  behavioral variant, 207, 208, 3486, 3486f
  clinical features of, 194, 196, 196t, 3486
  functional anatomy of, 193, 206f
  genetic considerations in, 35, 195t, 3395, 3486
  imaging in, 3486f
  molecular basis of, 195t
  neuropathology of, 195t
  nonfluent/agrammatic variant, 3486, 3486f
  pathogenesis of, 206f, 3395, 3400, 3486–3487, 3487f
  pathology of, 3486, 3487f, 3488f
  prions in, 3525
  semantic variant, 3486, 3486f
  syndromes
    corticobasal syndrome. *See* Corticobasal syndrome
    neuropathology of, 3487f
    in PD, 3488
    Pick's disease, 3487, 3488f
    progressive supranuclear palsy. *See* Progressive supranuclear palsy syndrome
  treatment of, 3487
Frontotemporal lobar degeneration (FTLD), 203, 204, 3486, 3487f, 3488f
Frostbite, **3771**
  clinical features of, 2177f, 2179, 3771–3772, 3771f, A5
  complications of, 3772
  pathophysiology of, 3771
  treatment of, 2179, 3772, 3772t
Frovatriptan, 3467t, 3468, 3468t
"Frozen shoulder," 2974–2975
Fructose, 309, 3358
Fructose metabolism disorders, 3362t, 3366, 3369f
Fruquintinib, 568t, 658
FSH. *See* Follicle-stimulating hormone (FSH)
FSHD (facioscapulohumeral muscular dystrophy), 3634, 3637f, 3643–3644
FSHβ gene mutations, 3106, 3112t, 3113
FS (Functional Status) Score, 3580t
FTA-ABS (fluorescent treponemal antibody–absorbed) test, 1433
FTC (emtricitabine), 1617t, 1619f
FTD. *See* Frontotemporal dementia (FTD)
FTLD (frontotemporal lobar degeneration), 203, 204, 3485, 3487f
5-FU. *See* 5-Fluorouracil (5-FU)
Fucosidosis, 3353t
Fugue states, 3485
Fukutin-related protein deficiency (LGMD2I), 3641–3642
Full liquid diet, 2603
Fulminant diabetes, 3196
Fulvestrant, 567, 640, 753t
Fumagillin, 1743t, 1748
Fumarate reductase, 1752
Functional gait disorders, 179
Functional genomics, 3778, 3781
Functional lumen imaging probe (FLIP), 296, 2504–2505
Functional MRI (fMRI), 95f, 3390, 3933, 3966, 3966f
Functional (psychogenic) neurologic symptom disorder, 3519
Functional reentry, 1914
Functional residual capacity (FRC), 2202, 2205, 2206f, 2296
Functional status, 3907t
Functional Status (FS) Score, 3580t
Fundoplication, 302, 2511
Fungal balls, 1715, 1715f, 1717t
Fungal infections, **1687**. *See also specific diseases*
  arthritis, 1058–1059
  in asthma, 2221
  in cancer patient, 578, 582, 584
  as cause of fever of unknown origin, 151t
  classification of, 1687
  definition of, 1687
  diagnosis of, 1690–1691
  drug-resistant, 1687
  endemic, 1687
  etiology of, 1687
  in HCT recipient, 918, 918t, 1155t, 1156
  health-care associated, 1150
  invasive, 582
  in lung transplant recipient, 1162, 2286
  multiple nerve palsies in, 235
  opportunistic, 1687, 1728–1729
  pathogenesis of, 1687, 1689–1690, 1690f
  septic shock in, 990
  sinusitis, 256, 257, 3559
  in solid organ transplant recipient, 1159t, 1161–1163, 1705f, 2410, 2723
  treatment of. *See* Antifungal agents
Fungi, 1687, S11
Funnel chest (pectus excavatum), 1860, 3332
Funnel-web spider, 3734
FUO. *See* Fever of unknown origin (FUO)
Furazolidone, 1743t, 1748, 2528
Furosemide
  adverse effects of
    acute interstitial nephritis, 2356
    bullous pemphigoid, 412, 420
    fever, 150
    hearing loss, 246
    metabolic alkalosis, 374
    phototoxicity, 434t
    pseudoporphyria, 401
    thrombocytopenia, 923t
  for AKI, 2384
  for ascites, 329, 2715
  for heart failure, 1991, 1993t, 2123
  for hypercalcemia of malignancy, 738
  for hypertension, 2145t
  for pulmonary edema, 83, 2331
  for SIAD, 351, 738, S1
Furrier's lung, 2228t
Furuncles, 1054, 1198
Furuncular myiasis, 3732
Furunculosis, 389
Fusariosis
  in cancer patient, 582
  clinical features of, 1688t, 1729, 1729f
  diagnosis of, 1688t, 1729
  epidemiology of, 1728–1729
  risk factors for, 1688t
  treatment of, 1726t, 1729
  in war veterans, S7
Fusarium spp., 1728
FUS gene mutations, 3400, 3486, 3487f
Fusidic acid, 1064t, 1065t
Fusiform aneurysm, 2165
Fusobacterium spp., 1372, 1373
Fusobacterium spp. infections. *See also* Anaerobic bacterial infections, mixed
  in dog-bite wound, 1143
  in human bite-wounds, 1144
  infectious arthritis, 1377
  Lemierre's disease/syndrome. *See* Lemierre's disease/syndrome
  in lung abscess, 1035
FUS protein, 195t
FUS/TLS gene, 3522t, 3523
FUT2, 966
Futibatinib, 533t, 563t, 568t
Futile care, 89, 2299
FVC (forced vital capacity), 2206f, 2207
FXN gene mutations, 2004t

G
G6PC3 (glucose-6-phosphatase) deficiency, S9
G6PD deficiency. *See* Glucose-6-phosphate dehydrogenase (G6PD) deficiency
G20210A, 941
G20210A mutation, 941
GABA (γ-aminobutyric acid), 210, 211f, 3368t, 3675
GABA (γ-aminobutyric acid) agonists, 302
Gabapentin
  for acroparesthesia, 3355
  adverse effects of, 82, 3188, 3422t, 3604t
  for alcohol withdrawal treatment, 3681
  for anxiety disorders, 3661
  for cough, 274
  for fibromyalgia, 2965
  for hot flashes in cancer treatment, 510, 638
  for IC/BPS, 335
  for menopausal symptoms, 3143
  for metastatic bone pain, 508
  for nausea and vomiting, 299, 299t, 300
  for neuropathy, 80f, 81–82, 100, 506, 3227, 3604t
  for new daily persistent headache, 117
  overdosage/poisoning with, 3713t
  for pain, 97t, 3588
  for paroxysmal symptoms in MS, 3588
  pharmacology of, 3422t
  for restless legs syndrome, 217
  for seizures, 3422t, 3424
  for SUNCT/SUNA prevention, 3473
GABA (γ-aminobutyric acid) receptor antibodies, 743t, 744t, 746f, 747f, 748, 750, 750f, 1114
GAD1 gene mutations, 3655
GAD65 (glutamate decarboxylase), 1114, 2790t
GAD antibodies, 3533
GAD (glutamic acid decarboxylase) antibodies, 744t, 746, 749, 750, 750f, 1114
Gadolinium contrast, 420, 1884, 2377, 2394, 3387–3388. *See also* Contrast agents
Gadolinium exposure, S6
Gaenslen's test (pelvic torsion), 121t
Gag/*gag,* 1551, 1552f, 1558, 1561f
Gag reflex, 3380
Gaisböck's syndrome, 451, 819
Gait
  anatomy and physiology of, 177
  cerebellar, 746
  diagnostic clues in, 50t
  examination of, 3382
Gait apraxia, 178
Gait cycle, 177
Gait disorders, **177**
  approach to the patient, 179
  cautious gait, 178
  cerebellar ataxia, 178–179, 179t
  etiology and prevalence of, 177t
  freezing gait, 178, 180, 3507. *See also* Parkinson's disease
  frontal, 178, 179t
  functional, 179
  in neuromuscular disease, 179

- Gait disorders (*Cont.*):
  - sensory ataxia, 179, 179t
  - steppage, 179
  - stiff-legged gait, 178
  - in toxic and metabolic disorders, 179
  - video library, V1
- Galactocerebrosidase deficiency, 3357
- Galactomannan test, 1691, 1716
- Galactorrhea, 3005
- Galactose, 3358
- Galactose metabolism disorders, 3359f, 3361t, 3365–3366
- α-Galactosidase A gene mutations. *See* Fabry's disease
- Galantamine, 199, 3480, 3491, 3695
- Galcanezumab, 3469, 3470t, 3472, 3472t
- Gallavardin effect, 283, 285, 1867, 2037
- Gallbladder
  - bile concentration within, 2727
  - fistula formation in, 2733
  - gangrene of, 2733
  - hypomotility of, 2729
  - palpable, 673
  - perforations of, 2733
  - porcelain, 2733
  - strawberry, 2734
- Gallbladder disease, **2728**
  - abdominal pain in, 110
  - cancer, 326, 669–670
  - chest discomfort/pain in, 103t
  - cholecystitis. *See* Cholecystitis
  - clinical features of, 2730–2731
  - congenital anomalies, 2728
  - diagnosis of, 2730, 2730t, 2731f
  - gallstones. *See* Gallstone(s)
  - hyperplastic cholecystoses, 2734
  - in IBD, 2562t, 2563
  - natural history of, 2731
  - postmenopausal hormone therapy and, 3144t, 3145
  - in pregnancy, 3903
- Gallium-68, 681, 681f
- Gallium nitrate, 726t
- Gallstone(s), **2728**
  - asymptomatic, 2731
  - biliary sludge and, 2729
  - cholesterol stones, 2728–2729, 2728f, 2730t
  - clinical features of, 2730–2731
  - complications of, 2731
  - diagnosis of, 2730, 2730t
  - epidemiology of, 2728
  - genetic considerations in, 2728
  - in IBD, 2562t, 2563
  - jaundice in, 323
  - natural history of, 2731
  - obesity and, 3187
  - pathogenesis of, 2728–2729
  - pigment stones, 2729–2730, 2730t
  - as risk factor for gallbladder cancer, 174
  - in sickle cell disease, 774t
  - treatment of, 2731
- Gallstone ileus, 2593, 2733
- Gallstone pancreatitis, 2489, 2748
- *GALNT3* gene mutations, 3265, 3319
- GALT (gut-associated lymphoid tissue), 1570, 1574
- γ-aminobutyric acid (GABA), 210, 211f, 3182, 3368t, 3675
- γ-aminobutyric acid (GABA) agonists, 302
- γ-aminobutyric acid (GABA) receptor antibodies, 744t, 748, 1114
- γ-carboxylase (GGCX), 934
- γ-glutamylcarboxylase, 934
- γ-glutamyl transferase, 464
- γ-glutamyl transpeptidase (GGT), 2635, 2698
- Gamma heavy chain disease (Franklin's disease), 894
- γ-hydroxybutyric acid (GHB), 184, 215, 3713t
- Gamma knife surgery, 3001, 3553
- Gamma-linolenic acid (GLA), 3926
- γ motor neurons, 170f
- γ radiation, 554, S6
- *GANAB* gene mutations, 2431
- Ganaxolone, 3412t, 3415, 3422t
- Ganciclovir
  - actions of, 1115, 1485
  - adverse effects of, 1115–1116, 1160, 1485, 1486t, 3414t
  - for CMV infection
    - prophylaxis, 918, 918t, 1160, 1485
    - in transplant recipient, 1537
    - treatment, 1485, 1486t, 1515, 2513, 3566
  - for HHV-6 infection, 1156
  - indications for, 1485–1486
  - pharmacology of, 1485
  - resistance to, 1485, 1515
- Gan Gan virus, 1658t, 1669
- Gangliocytomas, 3003
- Gangliogliomas, 720
- Ganglionopathy, dorsal root, 750, 3607
- Gangosa, 1438
- Gangrene
  - anaerobic bacterial, 1377
  - Fournier's, 1052, 1377, 2321t
  - gallbladder, 2733
  - gas. *See* Clostridial myonecrosis
  - hemolytic streptococcal. *See* Necrotizing fasciitis
- GAPD (glyceraldehyde-3-phosphate dehydrogenase) deficiency, 796t
- Gapping test (distraction), 121t
- Gardener's pupil, 221
- Gardner-Diamond syndrome, 408
- *Gardnerella vaginalis,* 1102, 3839, S11. *See also* Bacterial vaginosis
- Gardner's syndrome
  - clinical features of, 403, 654, 654t, 717t
  - genetic considerations in, 717t
  - polyp distribution and characteristics in, 654t
  - skin manifestations of, 403
- Garetosmab, 3319
- Gas chromatography, 3977
- Gas cysts, intestinal, 1240
- Gasdermin D (GSDMD), 972
- Gas exchange, 2204–2205, 2205f, 2207
- Gas gangrene. *See* Clostridial myonecrosis
- GAS infections. *See* Group A *Streptococcus* (GAS) infections
- Gas, intestinal, 326, 2571
- Gas-liquid chromatography, S11
- Gastrectomy
  - cobalamin deficiency after, 786
  - complications of, 2533–2534, 3613
  - for gastric adenocarcinoma, 647
  - for gastric lymphoma, 650
  - for peptic ulcer disease, 2532
- Gastric acid, 1077, 2515–2517, 2518f, 2743
- Gastric adenocarcinoma, **645**
  - clinical features of, 646
  - deaths from, 499f, 501t, 645
  - diffuse type, 647
  - epidemiology of, 645
  - etiology of, 645–646
  - genetic considerations in, 524t, 646–647, 646f, 1299
  - global considerations in, 645
  - histopathologic classification of, 646–647, 646f
  - *H. pylori* infection and, 1300, 1300f, 2526–2527
  - incidence of, 498, 501t, 645
  - indigestion in, 300
  - postgastrectomy, 2534
  - screening for, 646
  - staging of, 645t, 647, 648f, 2475f
  - treatment of, 647–649
- Gastric antral vascular ectasia, 2870, 2874
- Gastric balloons, intraluminal, 3194
- Gastric biopsy, 786
- Gastric bypass surgery. *See* Bariatric surgery
- Gastric emptying, delayed. *See* Gastroparesis
- Gastric emptying test, 301
- Gastric fundal varices, 2482, 2488f, V5
- Gastric glands, 2514, 2515f
- Gastric inhibitory peptide, ectopic production of, 737t, 739
- Gastric lavage, 370, 3708–3709
- Gastric lymphoma, 649–650. *See also* Mucosa-associated lymphoid tissue (MALT) lymphoma
- Gastric neuroendocrine tumors, 652, 679–680, 680f. *See also* Neuroendocrine tumors (NETs)
- Gastric outlet obstruction, 649, 2485, 2492f, 2523, V5
- Gastric peroral endoscopic myotomy (G-POEM), 299
- Gastric ulcers. *See* Peptic ulcer disease
- Gastric varices, 318
- Gastrin, in ZES, 2535–2536, 2536t
- Gastrinoma. *See also* Zollinger-Ellison syndrome (ZES)
  - clinical features of, 308, 678, 678t
  - in MEN 1, 2535, 3080t, 3081–3082
  - screening tests for, 2535–2536, 2536t, 3081t
  - treatment of, 312, 678, 678t, 2536–2537, 3081–3082
- Gastritis, **2537**
  - abdominal pain in, 112t
  - acute, 2537–2538, 2538t
  - after cholecystectomy, 2734
  - autoimmune, 2790t
  - bleeding in, 318, 437
  - chronic, 2538–2539, 2538t, 2539f
  - classification of, 2537, 2538t
  - eosinophilic, 2539
  - granulomatous, 2539
  - halitosis in, 267
  - immune checkpoint inhibitor–induced, 2539
  - lymphocytic, 2539
  - multifocal atrophic, 2538
  - Russell body, 2539
  - stress-related, 318, 2537
  - treatment of, 2539
  - varioliform, 2539
- Gastroenteritis. *See also* Diarrhea, infectious
  - vs. appendicitis, 2595
  - eosinophilic, 300, 309, 2550, 2550t
  - food-related. *See* Food-related illness
  - *L. monocytogenes,* 1228
  - *Salmonella. See Salmonella* spp. infections, food-borne
  - *T. whipplei,* 1368
  - viral, **1627**
    - adenovirus, 1535, 1628f, 1628t, 1632
    - astrovirus, 1628f, 1628t, 1632
    - vs. bacterial, 1629t
    - calicivirus, 1628f, 1628t. *See also* Norovirus infection
    - clinical features of, 1629t
    - diagnosis of, 1629t
    - picobirnavirus, 1628f, 1632
    - rotavirus. *See* Rotavirus infections
    - sapovirus, 1628, 1628t
    - torovirus, 1628f, 1632
    - treatment of, 1629t
- Gastroesophageal junction, 2505, 2505f
- Gastroesophageal junction cancer, 642
- Gastroesophageal reflux disease (GERD), **2509**
  - approach to the patient, 301
  - asthma exacerbations and, 2221

  clinical features of
    alarm symptoms, 301, 301t, 2490
    chest pain, 103t, 104, 106
    cough, 273
    dysphagia, 296
    esophagitis, 2490, 2510f
    extraesophageal, 2510
    indigestion/heartburn, 300, 2503, 2510
    nausea and vomiting, 297
    oral, 261
  complications of, 2510–2511, 2510f. *See also* Barrett's esophagus/metaplasia
  diagnosis of, 2490, 2505
  differential diagnosis of, 300, 2510
  eosinophilic esophagitis and, 2511–2512
  epidemiology of, 300, 2509
  H. pylori and, 1300
  mechanisms of, 300
  pathophysiology of, 2509–2510
  in systemic sclerosis, 2870, 2874
  treatment of, 301–302, 2511
Gastrointestinal bleeding, **316**
  approach to the patient, 317f, **318,** 319f, 2474, 2476, 2478
  aspirin-induced, 943
  in bleeding disorders, 473
  in Crohn's disease, 2558
  diagnosis of, 318–319
  in Dieulafoy's lesion, 2482, 2488f, V5
  differential diagnosis of, 2460t
  diverticular, 2484, 2491f, 2579
  endoscopic evaluation of, 2476, 2478
  epidemiology of, 316
  from esophageal varices. *See* Esophageal varices
  from Mallory-Weiss tears, 2482, 2489f, V5
  minor, endoscopic findings in, 2498, 2503f
  in Osler-Weber-Rendu syndrome, 927
  in peptic ulcer disease. *See* Peptic ulcer disease, complications of
  sources of
    colonic, 318, 320, 2461. *See also* Hemorrhoidal disease; Rectal bleeding
    obscure, 320
    occult, 316, 320, 2496
    small intestinal, 318, 320, 2461
    upper gastrointestinal, 317–318, 317f, 319–320, 319t, 2461
  vascular ectasias and, 2482, 2490f
Gastrointestinal cancer
  anal, 659, 1527
  colorectal. *See* Colorectal cancer
  early-onset, 642
  esophageal. *See* Esophageal cancer
  gastric. *See* Gastric adenocarcinoma
  gastric lymphoma, 649–650. *See also* Mucosa-associated lymphoid tissue (MALT) lymphoma
  small bowel, 651
Gastrointestinal disease, **2459**. *See also specific diseases*
  approach to the patient, 2461–2463, 2462t
  classification of, 2459–2460
  clinical features of, 2460, 2460t
    abdominal pain, 2460. *See also* Abdominal pain
    altered bowel habits, 2460–2461. *See also* Constipation; Diarrhea
    ascites, 326
    chest pain, 103t, 104
    gastrointestinal bleeding, 2461. *See also* Gastrointestinal bleeding
    heartburn, 2460
    involuntary weight loss, 315. *See also* Involuntary weight loss
    jaundice, 2461. *See also* Jaundice
    nausea and vomiting, 297, 2460. *See also* Nausea and vomiting
  in diabetes mellitus, 3227
  endoscopy in, 2461, 2461t
  functional testing in, 2463
  genetic considerations in, 2460
  gut microbiome and, 3835
  at high altitudes, 3758
  histopathology of, 2463
  history in, 2461
  imaging in, 2462
  laboratory tests in, 2461–2462
  mucormycosis, 1719
  in nerve agent exposure, S5
  in obesity, 3187
  parasitic infections, 1738–1740t
  physical examination in, 2461
  in pregnancy, 3903
  in sulfur mustard exposure, S5
  in travelers. *See* Travelers' diarrhea
  treatment of, 2463–2464
  in Whipple's disease, 1368, 1368f
Gastrointestinal endoscopy, **2464**. *See also specific types, e.g.,* Capsule endoscopy
  in anemia, 2495–2496
  antibiotic prophylaxis for, 2473, 2482t
  antithrombotic drug management before, 2473, 2483t
  in Barrett's esophagus. *See* Barrett's esophagus/metaplasia
  in biliary obstruction, 2473f, 2489
  in cancer staging, 2475f, 2499
  in celiac disease. *See* Celiac disease (gluten-induced enteropathy)
  in cobalamin deficiency, 788
  in colonic pseudo-obstruction, 2487, 2493f
  for colorectal cancer screening. *See* Colonoscopy
  in diarrhea, 307, 2497–2498
  in duodenal ulcer, 2524f
  in dyspepsia, 2490
  in dysphagia, 296, 2492–2493, 2496f
  equipment for, 2464–2465, 2465f
  in esophageal disease, 2504
  in gastric outlet obstruction, 2485, 2492f
  in gastrointestinal bleeding. *See* Gastrointestinal bleeding
  in gastrointestinal obstruction, 2485, 2494f
  in GERD, 2490, 2497f
  in hematochezia, 2498, 2503f
  in *H. pylori* infections, 1301, 1301t
  indications for, 2462, 2462t
  in malabsorption disorders, 2498
  narrow-band imaging, 2465, 2465f
  in nausea and vomiting, 298
  open-access, 2499–2500
  in pancreatitis, 2498–2499
  patient preparation for, 2500
  in peptic ulcer disease, 2523, 2524f. *See also* Peptic ulcer disease
  risks of, 2473–2474
  therapeutic. *See* Endoscopic therapy
  upper tract, 301, 2465, 2466–2467f
  urgent, **2474**
  in Whipple's disease, 1368, 1368f
Gastrointestinal hemorrhage. *See* Gastrointestinal bleeding
Gastrointestinal motility test, 298
Gastrointestinal neuroendocrine tumors (GI-NETs). *See* Neuroendocrine tumors (NETs)
Gastrointestinal obstruction. *See* Intestinal obstruction
Gastrointestinal pseudoobstruction. *See* Intestinal pseudoobstruction
Gastrointestinal stromal tumor (GIST)
  clinical features of, 650
  diagnosis of, 650
  genetic considerations in, 728
  metastatic, 728
  pathogenesis of, 650–651
  pathology of, 728
  small-bowel, 651
  treatment of, 567, 651, 729, 2478f
Gastrointestinal tract, 1372, 2459. *See also* Absorption
Gastroparesis
  abdominal pain in, 300
  in diabetes mellitus, 3227
  diagnosis of, 298
  etiology of, 297
  indigestion in, 300
  nausea and vomiting in, 298
  in systemic sclerosis, 2870
  treatment of, 299, 299t
Gastropathy, 318, 2714
Gastrostomy tube, 296, 2483f
GATA1 gene mutations, 443, 826, 879, 3349
GATA2 deficiency
  clinical features of, 458, 816, 1416, 2796
  diagnosis of, 459t
  pathogenesis of, 459t, 2795f, 2796
GATA2 gene mutations, 577, 816, 825, 825t, 2183
GATA3, 733t, 3286
GATA6 gene mutations, 3203
Gatekeeper genes, 523
Gatekeeper, physician as, 72
Gates Foundation, 3843
Gaucher cells, **3355**
Gaucher disease, 480, 818, 3353t, **3355,** 3499
Gay, 3179t
Gaze disorders, 235–236, 236f
Gaze-evoked nystagmus, 236–237
Gaze stabilization reflex, 220
GBA gene mutations, 3355–3356, 3400, 3497, 3498t, 3499
GBD (Global Burden of Diseases, Injuries, and Risk Factors Study), 3845
GBP (guanylate binding protein), 973
GBS. *See* Guillain-Barré syndrome (GBS)
GBS infections. *See* Group B *Streptococcus* (GBS) infections
GB virus C (hepatitis G virus) infection, 1601
GCH1 gene mutations, 3510t, 3511
G-CSF (granulocyte colony-stimulating factor), 454f, 1023, 2766t
G-CSF therapy. *See* Granulocyte colony-stimulating factor (G-CSF) therapy
GDF15, 3813
GDNF (glial-derived neurotrophic factor), 3407
Gefitinib
  actions and targets of, 568t, 572
  adverse effects of, 568t, 572, 595
  drug interactions of, 591
  for head and neck cancer, 614
  interactions and issues, 568t
  for lung cancer, 627, 627t
  resistance to, 572
GEFs (guanine nucleotide exchange factors), 969
Gegenhalten (paratonia), 168
Gelesis100, 3194
Gelsolin amyloidosis, 2027t
Gemcitabine
  actions of, 565
  adverse effects of, 562t, 565, 595, 754, 2377, 2445
  for bladder cancer, 692, 693
  for carcinoma of unknown primary, 735
  for cholangiocarcinoma, 669
  for DLBCL, 863
  for gallbladder cancer, 670
  interactions and issues, 562t
  for lung cancer, 627t, 630
  for metastatic breast cancer, 641

- Gemcitabine (*Cont.*):
  - for ovarian cancer, 712
  - for pancreatic cancer, 674, 676t
  - for peripheral T-cell lymphoma, 867
  - resistance to, 3837
  - for soft tissue sarcomas, 729
  - for testicular cancer, 709
- *Gemella morbillorum,* 1041, 1043t, 1214
- Gemfibrozil, 3244t, 3251, 3258
- Gemifloxacin
  - actions of, 1166, 1182t
  - for community-acquired pneumonia, 963t, 1027t, 1028t
  - for gonococcal infections, 1259, 1259t
  - indications for, 1178
  - resistance to, 1182t
- Gemtuzumab, 593, 594, 597, 753t
- Gemtuzumab-ozogamicin, 535t, 566, 566t, 830, 833t
- Gender-affirming chest surgery, 3179t
- Gender-affirming genital surgeries, 3179t
- Gender-affirming hormone therapy, 3179t
- Gender dysphoria, 3094, 3179t
- Gender expression, 3179t
- Gender identity, 3098, 3101, 3178, 3179t
- Gender nonconforming, 3179t
- Gender queer, 3179t
- Gene(s). *See also specific genes*
  - assortment and segregation during meiosis, 3780, 3782f
  - definition of, 3779
  - deletions/duplications, E72t
  - mutations. *See* Mutation(s)
  - regulatory regions of, 3779, 3780f
  - structure of, 3779
- Gene conversion, 3785
- Gene expression
  - alterations of, 538–539, 538f, 3870
  - in cancer cells, 528
  - epigenetic regulation of, 3782–3783, 3784f
  - phenotype and, 33f
  - regulation by transcription factors, 3780–3782, 3780f, 3783t
- Gene expression profile (GEP), 602, 733
- Gene network, 3955, 3959–3961f
- Generalized anxiety disorder, **3658,** 3661t
- Generalized epilepsy with febrile seizures, 3412t
- Generalized seizures, 3410–3411
- Generative AI, 26
- Generative pretrained transformer (GPT), 3970, 3972f
- Gene silencing, 522
- Gene therapy, **3823**
  - for acute intermittent porphyria, 3347
  - for adenosine deaminase deficiency, 3824t
  - for adrenoleukodystrophy, 3826
  - for angiogenesis inhibition, 3829
  - for cancer, 3824t, 3828
  - clinical trials, 3826t
  - complications of, 3823, 3825t
  - currently approved products, 3824t
  - delivery vehicles for, 3823t
  - ethical issues in, 74
  - for familial hypercholesterolemia, 3243
  - for genetic disorders, 3823–3827, 3824t
  - for heart failure, 2000
  - for hemophilia, 931, 3827–3828
  - for immunodeficiency disorders, 463, 2799, 3825–3826
  - for infectious diseases, 3824t
  - for leukemia, 3824t, 3829
  - long-term follow-up in, 3826t
  - for LPL deficiency, 3827
  - for metachromatic leukodystrophy, 3824t, 3827
  - for PD, 3506
  - for *RPE65* mutation-associated retinal dystrophy, 3824t, 3827
  - for sickle cell disease, 776, 3826
  - for spinal muscular atrophy type 1, 3824t, 3828
  - stem cells as target of, 761–762
  - for thalassemia, 778, 3824t, 3826
  - viral vectors for, 1553, 3823t
  - for Wilson's disease, 3340
- Genetic counseling
  - ethical issues in, 3798–3799
  - in familial cancer syndromes, 525, 525f505f
  - indications for, 3801
  - in mtDNA diseases, 3816
  - pretest, 3803
- Genetic disorders. *See also specific disorders*
  - adult-onset
    - family history in, 3800–3801, 3801f
    - genetic testing for, 3801–3804, 3802f, 3804t. *See also* Genetic testing
    - inheritance patterns in, 3800
    - therapeutic interventions for, 3804t, 3805
  - approach to the patient, **3795,** 3798f
  - chromosome disorders. *See* Chromosome abnormalities/disorders
  - complex, 3793
  - databases/websites, 3778t
  - diagnosis of, 3777
  - family history in, 3795
  - identifying disease-causing gene in, 3795–3798, 3797t
  - Mendelian. *See* Mendelian disorders
  - mitochondrial. *See* Mitochondrial DNA diseases
  - multifactorial, 3777, 3793, 3800, 3958f
  - mutational analysis in, 3796–3797
  - penetrance of, 3800
  - polygenic, 3793, 3794t
  - population screening for, 3804–3805
  - prenatal diagnosis, 3797
  - transmission of. *See* Mutation(s)
  - variable expression of, 3800
  - X-linked. *See* X-linked disorders
  - Y-linked, 3789, 3791
- Genetic Information Nondiscrimination Act (GINA), 525, 3798, 3803, S10
- Genetic linkage, 3793–3794, 3797t
- Genetic privacy, 3798
- Genetics
  - databases/websites, 3778t
  - impact on medical practice, 3777
  - integration into clinical practice, 3799–3800
  - population, 3795
- Genetic testing
  - approach to, 3801–3804, 3802f
  - for breast cancer risk, 633–634
  - direct-to-consumer, 3802, 3803t
  - ethical issues in, 3798–3799
  - for familial cancer syndromes, 525, 525f
  - family relationships and, 3803
  - follow-up care after, 3803–3804
  - indications for, 3789t
  - informed consent in, 3795, 3803, S10
  - limitations of, 3802–3803, S10
  - methods for, 3796–3797, 3797t, 3801–3804
  - in myopathies, 3638
  - predisposition, 3803
  - presymptomatic, 3803
  - screening programs, 3796
  - in telomere disease, 3821–3822
  - in undiagnosed disease, 3997–3998
- Gene transfer, horizontal, 976t
- Genital tract, anaerobic flora of, 1372
- Genital tract infections
  - abscesses, 1073
  - anaerobic bacterial, 1376–1377
  - HPV. *See* Human papillomavirus (HPV) infections
  - HSV. *See* Herpes simplex virus (HSV) infections, genital
  - parasitic, 1739t
  - tuberculous, 1388
- Genital ulcer(s)
  - in Behçet syndrome, 2908, 2908f
  - chancroid. *See* Chancroid
  - clinical features of, 1107f, 1108t
  - diagnosis of, 1108t
  - donovanosis. *See* Donovanosis
  - etiology of, 1097t, 1107
  - HIV transmission and, 1563, 1563f
  - HSV. *See* Herpes simplex virus (HSV) infections, genital
  - lymphogranuloma venereum. *See* Lymphogranuloma venereum
  - treatment of, 1108–1109, 1109t
- Genome, 976t, 3777. *See also* Human genome
- Genome-wide association studies (GWAS)
  - advantages and limitations of, 3797t
  - of complex traits, 3777, 3794–3795, 3796f
  - microbial, 976t
  - multiple data types in, 494–495
  - in pharmacogenomics, 491
- Genomic imprinting. *See* Imprinting
- Genomic medicine, 3777, 3799
- Genomics, 6. *See also* Pharmacogenomics
  - in cancer. *See* Cancer genetics
  - comparative, 3778
  - databases/websites, 3778t
  - functional, 3778, 3781
  - global health and, 3799
  - impact on medical practice, 3777
  - in infectious disease. *See* Infectious diseases, genomics of
  - in precision medicine, 31–34, 33f
  - terminology of, 976t
  - in undiagnosed disease, 3997–3998
- Genotoxicity, 3825t
- Genotype, 3787
- Genotype-phenotype correlation
  - in common disease types, 3958f
  - definition of, 3787
  - in sickle cell anemia, 3958–3959f
  - in sickle cell disease, 3957
  - in von Hippel–Lindau disease, 3787
- Genotyping, 491, 494
- Gentamicin
  - actions of, 1166, 1182t
  - adverse effects of, 1041, 1177
  - for *Bartonella* endocarditis, 1353t
  - for *Brucella* infections, 1334l
  - for *Campylobacter* infections, 1306
  - for group C or G *Streptococcus* infections, 1213
  - for hospital-acquired pneumonia, 963t
  - indications for, 1173t, 1177
  - for infective endocarditis, 1041, 1043t, 1044, 1044t, 1205
  - intratympanic, for Ménière's disease, 247
  - for *L. monocytogenes* infections, 1229
  - for meningitis, 1121t, 1122t, 1123
  - for necrotizing fasciitis, 991t
  - for pelvic inflammatory disease, 1106, 1106t
  - for plague, 1346, 1346t, S4
  - in pregnancy and lactation, 1169t
  - prophylactic, 1180t
  - resistance to, 1041, 1173t, 1182t, 1184, 1219, 1221
  - for septic shock, 991t
  - for trench fever, 1353t
  - for tularemia, 1339–1340, 1340t, S4
  - for ventilator-associated pneumonia, 1032t
- Genu valgum, 2944, 2951, 2951f
- Genu varum, 2944, 2951, 2951f
- Geographic tongue, 265t, 380f, A3
- Gepants (CGRP receptor antagonists), 3467t, 3468–3469, 3470t
- Gephyrin, 2791t

Gepirone, 3659t
Gepotidacin, 1341
Gerbil bite-wound, 1143
GERD. *See* Gastroesophageal reflux disease (GERD)
Geriatric care. *See* Aging; Older adults
German measles. *See* Rubella virus infection (German measles)
Germ cell(s)
  male, 3118
  ovarian, 3094, 3124, 3125–3126f
  primordial, 3094, 3124
Germ cell tumor
  as carcinoma of unknown primary, 709
  carcinoma of unknown primary and, 733t, 735
  extragonadal, 709, 735
  ovarian, 712
  retroperitoneal/mediastinal, 709, 2271
  testicular. *See* Testicular cancer
  tumor markers in, 504t
Germinoma, 737t, 3003
Germiston virus, 1658t
Germline mutations, 3784
Germ theory of disease, 957
Geroscience, 3872–3873
Gerstmann's syndrome, 203, 204
Gerstmann-Sträussler-Scheinker (GSS) disease, 3401, 3401t, 3526t, 3528, 3530
Gestational AVP deficiency, 352, 3016, 3017t
Gestational diabetes mellitus, 3165, 3196, 3221, 3900–3901
Gestational hypertension, 2140, 3899
Gestational trophoblastic disease, 504t, **715,** 737t
GFAP antibodies, 744t, 746
GFAP (glial fibrillary acidic protein) autoimmunity, 3593
GFR. *See* Glomerular filtration rate (GFR)
GGCX (γ-carboxylase), 934
GGT (γ-glutamyl transpeptidase), 2635, 2698
GH. *See* Growth hormone (GH)
GHB (γ-hydroxybutyric acid), 184, 215, 3713t
GH gene mutations, 3784
GH-N gene mutations, 2994
Ghrelin, 2517, 2990, 3182f, 3950
GHRH (growth hormone-releasing hormone), 737t, 2990, 2994, 3004
Gianotti-Crosti syndrome, 2660
Giant cell arteritis (temporal arteritis), **2902**
  clinical features of, 2902
    aortic aneurysm, 2166, A14
    aortitis, 2171
    ear pain, 253
    fever of unknown origin, 153
    headache, 114–115, 2902
    ocular, 227, 227f, 235, 2902
    oral pain, 65
  definition of, 2902
  diagnosis of, 115, 2902–2903, A14
  incidence and prevalence of, 2902
  ischemic stroke in, 3447
  pathogenesis of, 2166, 2902
  treatment of, 155, 2903
Giant cell myocarditis, 2014
Giant cell tumor, 729, 737t, 740
Giant orf, 1518
Giant platelets, 441f
Giardia spp., 1804, 1804f, S12
Giardia spp. infections, **1804**
  climate change and, 3752t
  clinical features of, 305, 306t, 1739t, 1805, S7
  diagnosis of
    serologic and molecular tests, 306, 311, 1805t, S12
    small-intestinal mucosal biopsies, 2547f
    stool analysis, 1805, 1805f, 1805t, S7
  differential diagnosis of, 306t
  epidemiology of, 305, 1079, 1804
  incubation period for, S7
  pathophysiology of, 1804–1805
  prevention of, 1805
  in travelers, 1079t
  treatment of, 1805, S7
    furazolidone, 1748
    metronidazole, 1805
    nitazoxanide, 1750, 1805
    quinacrine, 1752
    tinidazole, 1753, 1805
  in war veterans, S7
Gibbus, 1388
Giemsa stain, 381, S12
Gigantism. *See* Acromegaly
Gilbert's syndrome, **2640**
  clinical features of, 323, 2639t, 2640, 2729
  vs. Crigler-Najjar syndrome, 2639t
  diagnosis of, 2634, 2634f, 2637t, 2639t
  drug sensitivity in, 564
  epidemiology of, 323
  genetic considerations in, 323, 493, 2640–2641
  treatment of, 2640
  unconjugated bilirubin in, 321
Gilteritinib, 533t, 568t, 572, 833, 833t
GINA (Genetic Information Nondiscrimination Act), 525, 3798, 3803, S10
GI-NETs. *See* Neuroendocrine tumors (NETs)
Ginger, 300, 2463
Gingival erythema, linear, 261
Gingival fibromatosis, idiopathic familial, 261
Gingival hyperplasia, 261, A3
Gingival hypertrophy (epulis), A3
Gingival recession, A3
Gingivitis
  acute necrotizing ulcerative (Vincent angina), 261, 262t, 1374
  clinical features of, 261, A3
  etiology of, 1374
  prevention of, 463
  as risk factor lung abscess, 1034
Gingivostomatitis, herpetic. *See* Herpes simplex virus (HSV) infections, oral-facial
Gingko biloba, 3414t, 3482
Ginseng, 2137
GIRK channels, 355, 3689
GIST. *See* Gastrointestinal stromal tumor (GIST)
Gitelman's syndrome
  clinical features of, 355–356, 374, 2362, 2370, S1
  diagnosis of, 355–356, S1
  differential diagnosis of, 374
  genetic considerations in, 356, 2368t, 2370, 3267
Givosiran, 3347, 3603
GJB2 gene mutations, 247
GLA (gamma-linolenic acid), 3926
GLA gene mutations, 2004t, 2010
Glanders, 1309t, 1313
Glanzmann's thrombasthenia, 926
Glasdegib, 570t, 573, 833t
Glasgow-Blatchford score, 319, 319t
Glasgow Coma Scale, 192, 2292t, 3571, 3571t
Glatiramer acetate, 3585, 3587t
Glaucoma
  acute angle-closure, 225, 230
  clinical features of, 230
  headache in, 115
  normal-tension, 226
  pathophysiology of, 230–231, 230f
  treatment of, 231
  visual field assessment in, 222
  visual loss in, 230–231
GLCE gene mutations, 3112t
GLE1 gene mutations, 3523
Gleason grading system, for prostate cancer, 697–698
Glecaprevir/pibrentasvir, 1492t, 1493, 2686f, 2691
Glial cells, 3952
Glial-derived neurotrophic factor (GDNF), 3407
Glial fibrillary acidic protein (GFAP) autoimmunity, 3593
Glial substances, 1404
Gliflozins. *See* Sodium-glucose co-transporter 2 (SGLT2) inhibitors
Glimepiride, 3211t, 3213
Gliomas
  adult-type diffuse, 718, 718t
    glioblastoma, 717, 718–719, 719f, 3392f, A16
    IDH-mutant astrocytoma, 718, 718f
    oligodendroglioma, 718
  brainstem, A16
  cerebral, 3392f
  clinical features of, 716t
  genetic considerations in, 520t, 3940, 3978, 3979f
  hypothalamic, 3003
  imaging in, 2999
  ocular signs in, 222
  optic, 3003
  optic nerve, 232
  pediatric-type diffuse high-grade, 718t, 720
  pediatric-type diffuse low-grade, 718t, 720
  treatment of, 560
Glioneuronal tumors, 720
Gliosarcomas, 720
Glipizide, 485t, 493, 3211t, 3213
Global aphasia, 201t, 202
Global burden of disease. *See* Global medicine
Global Burden of Diseases, Injuries, and Risk Factors Study (GBD), 3845
Global Financing Facility, 3866
Global Fund to Fight AIDS, Tuberculosis, and Malaria, 3843
Global health equity, 3847, 3852, 3863–3864
Global Initiative for Lung Disease (GOLD) criteria, for COPD, 2254, 2255t
Global Leadership Initiative on Malnutrition, 2616, 2616f
Global Malaria Action Plan (GMAP), 3849
Global medicine, **3843**. *See also* Low- and middle-income countries
  AA amyloidosis, 899
  abdominal angiostrongyliasis, 1819
  acute coronary syndrome, 2112–2113
  acyclovir effectiveness, 1503–1504
  addressing problems in, 3852
  AKI, 2373
  alcohol use disorder, 3681
  amebiasis, 1753–1754
  aplastic anemia, 809
  APS-1, 3093
  aspergillosis, 1713, 1713t
  ataxias, 3537
  Babesia infections, 1776f, 1777, 1778, 1781
  biomass smoke-induced illness, 2241, 2241f
  bladder cancer, 690
  blastomycosis, 1700
  botulism, 1237–1238
  brain drain, 3848–3849
  Brucella infections, 1331–1332
  cancer, 498, 503f, 3851
  Candidatus Neoehrlichia mikurensis infection, 1461
  cardiovascular disease, 1856–1857, 1856f, 3850–3851
  Carrión's disease, 1356
  C. difficile infections, 1083
  cervical cancer, 713, 1526
  Chagas disease, 1789, 1792
  chemical terrorism, S5
  cholera, 1326–1327, 1326f
  chronic cough, 274

- Global medicine (*Cont.*):
  - CKD, 2397, 2401–2402
  - climate change and. *See* Climate change
  - clinical laboratory resources, S10
  - clostridial infections, 1239
  - CML, 843, 843f
  - congenital heart disease, 2078
  - congenital syphilis, 1432
  - COVID-19 disease, 3850
  - cryptococcosis, 1704
  - *C. trachomatis* infections, 1469
  - cystic fibrosis, 2248
  - dementia, 199
  - diabetes mellitus, 1858, 3196–3197, 3197f, 3850
  - dietary assessment, 2604
  - diphtheria, 1222
  - disease burden
    - causes of, 3845f, 3846t
    - health workforce and, 3862f
    - by income, 3860f
    - mortality and, 1856f, 3844–3845, 3846t
    - by region, 3846t
    - risk factors, 3845–3847
    - wealth and, 3845
  - disease eradication, 1002
  - disorders of sex development, 3102
  - donovanosis, 1357
  - drug hypersensitivity, 419–420
  - Ebola virus disease, 3849–3850
  - EBV infections, 1510
  - economics of, 3844
  - endemic typhus, 1459
  - enteric (typhoid) fever, 1315–1316, 1316f, 1319
  - enterovirus infections, 1633
  - epidemic typhus, 1458
  - epidemiologic transition, 1854–1856
  - esophageal cancer, 641–642
  - familial Mediterranean fever, 2937
  - fatigue, 166
  - fever of unknown origin, 149, 149t
  - field experiences in, ethical issues in, 74
  - filovirid infections, 1682–1683f
  - foot osteomyelitis, 1069
  - G6PD deficiency, 797, 799f
  - gastric adenocarcinoma, 645
  - genital ulcers, 1107
  - genomics and, 988, 3799
  - genomic technologies in, 988
  - gestational trophoblastic disease, 715
  - glycogen storage diseases, 3366–3367
  - gram-negative bacterial infections, 1281
  - group A *Streptococcus* infections, 1206–1207, 1207f
  - head and cancer, 610
  - heart failure, 1976, 1999–2000
  - heavy metal poisoning, 3699
  - hepatitis E infections, 2629
  - hepatocellular carcinoma, 659–660, 660f
  - *HFE* gene mutations, 3338
  - HHV-6, 7, and 8 infections, 1516
  - *H. influenzae* infections, 1261
  - HIV infection
    - incidence, deaths, and prevalence, 1565–1566f, 1565–1568, 1566–1567f
    - in sub-Saharan Africa, 3847–3849, 3847f
    - subtype distribution, 1561, 1562f
  - HSV infection, 1497
  - HTLV-1 infections, 1553
  - HTLV-2 infections, 1556
  - human African trypanosomiasis, 1796
  - hyperlipidemia, 1858
  - hyperparathyroidism, 366
  - hypersensitivity pneumonitis, 2231
  - hypertension, 1858
  - IC/BPS, 336–337
  - infectious disease-associated glomerulonephritis, 2428–2429
  - inflammatory bowel disease, 2551, 2552–2553
  - inflammatory myopathies, 2917
  - influenza, 1544
  - institutions for, 3843–3844
  - interstitial lung diseases, 2266
  - iodine deficiency, 3022–3023, 3022f
  - iron-deficiency anemia, 767
  - ischemic heart disease, 2090, 2112–2113
  - jaundice, 326
  - kidney disease, 2443–2444
  - *Klebsiella* spp. infections, 1289
  - left ventricular assist systems, 2033
  - leprosy, 1404–1405, 1405–1406f, 3606
  - leptospirosis, 1441
  - liver function tests, 2637
  - Lyme borreliosis, 1449
  - lymphogranuloma venereum, 1472
  - malaria, 1761–1762, 1761f, 1768, 3748, 3849
  - MASLD, 2707
  - maternal mortality, 3904
  - measles, 1003, 1638–1639, 1642
  - melanoma, 1942–1943
  - meningococcal disease, 992, 1245, 1246f
  - meningococcal vaccines, 1251–1252
  - mesoamerican nephropathy, 2436
  - metabolic syndrome, 3252
  - micronutrient deficiencies, 2604
  - mpox, 1517
  - multiple myeloma, 885
  - mumps, 1646
  - myasthenia gravis, 3633
  - myopathies, 3648
  - neglected tropical diseases, 1413
  - neurologic disease, 3377, 3377f
  - neuromyelitis optica, 3590–3591
  - nocardiosis, 1358
  - nontyphoidal *Salmonella* infections, 1319
  - NTM infections, 1414–1415, 1418
  - obesity, 1858, 3181, 3851
  - osteoarthritis, 2949
  - osteomyelitis, 1063, 1066
  - Paget's disease of bone, 3311–3312
  - paroxysmal nocturnal hemoglobinuria, 804
  - pertussis, 1277, 1277f
  - plague, 1342–1343, 1342f
  - pneumococcal infections, 1195
  - pneumonia, 1033
  - poisoning, 3716
  - poliovirus infections, 1002, 1637–1638, 1637t
  - porphyrias, 3342
  - prion diseases, 3526–3527, 3526t
  - prosthetic joint infections, 1067
  - psychiatric disorders, 3657, 3851–3852
  - pulmonary infiltrates with eosinophilia, 2234
  - pulmonary paragonimiasis, 274
  - Q fever, 1462
  - rabies, 1650, 1650f
  - RCC, 690
  - relapsing fever, 1445
  - rheumatic fever, 2854
  - rheumatic heart disease, 1207f, 1979, 2033, 2034, 2854
  - rheumatoid arthritis, 2842, 2842f, 2853
  - rickettsial diseases, 992, 1458
  - rotavirus infections, 1630–1631, 1630f, 1632f
  - rubella, 1643, 1645–1646, 1645f
  - sarcoidosis, 2921
  - schistosomiasis, 1825, 1828f
  - sepsis/septic shock, 2316
  - sexually transmitted infections, 1097, 1107
  - *Shigella* infections, 1322
  - sickle cell disease, 772
  - skin cancer, 432
  - snakebite care, 3722
  - substance use disorders, 3697
  - syphilis, 1429
  - TB, 1381–1382f, 1381–1383, 3848–3849
  - testicular cancer, 704
  - tetanus, 1230, 1232
  - tobacco use, 3851
  - toxoplasmosis, 1800
  - trachoma, 1474
  - treponematoses, 1436–1438, 1437f
  - tularemia, 1336, 1341
  - vaccine-preventable disease, 1003
  - valvular heart disease, 2033–2035, 2034f
  - vascular dementia, 3489
  - *Vibrio* spp. infections, 1330t
  - viral hemorrhagic fevers, 993
  - viral respiratory infections, 1539
  - vitamin D deficiency, 366
  - *Yersinia* spp. infections, 1348, 1350
- Global Pneumococcal Sequence Clusters, 1188
- Global travel. *See* International travelers
- Globi, 1404
- Globin gene clusters, 770, 770f
- Globin gene switching, 770f, 771
- Globoside RBC group system, 904t
- Globus pharyngeus, 293
- Globus sensation, 2504
- Glofitamab, 534t, 559f, 753t, 863
- Glomerular diseases, **2411**. *See also specific diseases*
  - acute nephritic syndromes. *See* Acute nephritic syndromes
  - approach to the patient, 2414–2416, 2414t
  - basement membrane syndromes. *See* Basement membrane syndromes
  - clinical syndromes, 2414–2416, 2415–2416t
  - genetic considerations in, 2411–2412
  - glomerular-vascular syndromes, 2415t, 2416, 2427
  - glomerulonephritis. *See* Glomerulonephritis
  - hematuria in, 341, 343, 2414, 2415–2416t
  - infectious disease-associated syndromes, 2415–2416t, 2416, 2428–2429
  - microalbuminuria in, 2414t
  - nephrotic syndromes. *See* Nephrotic syndromes
  - pathogenesis of, 2411–2412, 2413f
  - progression of, 2413–2414
  - proteinuria in, 2414, 2414t, 2415–2416t
  - pulmonary-renal syndromes, 2414, 2415t, 2425
  - renal biopsy in, 2416, A4
- Glomerular filtration rate (GFR)
  - age-related decline in, 2387
  - assessment of, 337–339, 2387, 2388t
  - in CKD, 2362
  - in kidney disease, 337–339, 2362t
  - in pregnancy, 3899
  - race modifier for, 338
  - reduced, 337
  - regulation of, 2364, 2373–2374, 2374f
  - serum creatinine and, 337
- Glomerulonephritis, **2354**
  - acute, 2379t
  - C3, A4
  - classification of, 2416
  - endocarditis-associated, 2415t, 2417
  - fibrillary, A4
  - hematuria in, 341, 2414
  - immune-mediated, 2412, 2413f
  - membranoproliferative. *See* Membranoproliferative glomerulonephritis
  - membranous. *See* Membranous glomerulonephritis
  - mesangioproliferative, 2415t, 2421
  - pathogenesis of, 2412–2413, 2790t
  - pauci-immune, 2420

  poststreptococcal/postinfectious. *See* Poststreptococcal glomerulonephritis
  primary, 2416
  progression of, 2413–2414, A4
  rapidly progressive, 2354f, 2355–2356
  rapidly progressive (crescentic), 2414, 2426f
  renal biopsy in, 2416, A4
  secondary, 2416
  in Sjögren's disease, 2877
  in SLE. *See* Lupus nephritis
  tubulointerstitial abnormalities with, 2442
Glomerulosclerosis
  age-related, 2416
  focal segmental. *See* Focal segmental glomerulosclerosis
Glomerulus (glomeruli), 2363–2364, 2387f, 2411, 2412f
Glossitis, 265t, 765, 2549t
Glossodynia (burning mouth syndrome, glossalgia), 65, 242
Glossopharyngeal nerve, 239, 239f, 3380, 3555, 3556f
Glossopharyngeal neuralgia, 265, 3555–3556, 3578
Glossopharyngeal neuropathy, 3559
Glucagon
  for bradycardia due to poisoning/overdose, 3712t
  for food impaction dislodgement, 2514
  for hypoglycemia, 3232t, 3236
Glucagon-like peptide-1 (GLP-1), 3182, 3182f, 3199
Glucagon-like peptide-1 (GLP-1) receptor agonists. *See also* Liraglutide; Tirzepatide
  for diabetes mellitus, 3211t, 3213, 3215
  for insulin resistance, 3258
  for ischemic heart disease, 2100
  for MASH, 2706
  in older adults, 3882t
  for OSA, 2282
  for primary polydipsia, 3019
  in women, 3166
Glucagonoma, 678–679, 678t, 679f, 3082
Glucocerebrosidase, 3400, 3499
Glucocorticoid(s)
  for ABPA, 2233
  for ACTH deficiency, 2997
  actions of, 2408t
  for ADEM, 3594
  for adrenal insufficiency, 3069
  adverse effects of, 716, 2225, 2836, 2916t, 3585, 3622
    adrenal insufficiency, 3070–3071
    dyslipidemia, 3247
    lithium for, 3585
    major, 2896t
    myopathy, 2940t, 3648, 3648t
    osteoporosis, 2612, 2849, 3296, 3310–3311
    protection from, 3585
    steatosis, 2703t
    tendinopathies, 2974
  for alcohol-associated liver disease, 2699–2700, 2709
  for allergic conjunctivitis, 224
  for anterior ischemic optic neuropathy, 227
  for asthma, 2224–2225, 2225t, 2226
  for autoimmune hemolytic anemia, 802
  for autoimmune hepatitis, 2696
  for autoimmune pancreatitis, 2751–2752
  for Behçet syndrome, 2909
  for bronchiectasis, 2244
  for CAH, 3072
  for cancer, 567
  carcinogenicity, 512t
  cataracts, 230
  for chronic inflammatory demyelinating neuropathy, 3622
  for cluster headache prevention, 3472, 3472t
  for contact dermatitis, 393t
  cutaneous, 395, 420, 423
  depression, 84
  diabetes mellitus and, 3220
  for dyspnea, 83
  for ectopic ACTH syndrome, 739
  edema, 282t
  for EGPA, 2230, 2900
  eosinopenia, 462
  for eosinophilia, 741
  epidural injections, 128, 132, 3393
  equipotency, 3069
  for fatigue, 84
  for fever, 136
  for fever of unknown origin, 155
  genomics of response to, 983
  for giant cell arteritis, 2903
  for gout, 2959
  for granulomatosis with polyangiitis, 2895
  for Graves' ophthalmopathy, 3037
  half-lives of, 2982
  for hypercalcemia, 365, 738, 3284t, 3285
  for hypereosinophilic syndromes, 462, 2233
  hypertension, 2137
  hypertrichosis, 421
  for IBD, 2564–2565
  for idiopathic cutaneous vasculitis, 2906
  for IgA vasculitis, 2905
  for immune thrombocytopenic purpura, 924
  for immunosuppression, 2408, 2408t
  infection, 578, 578t, 2225
  for inflammatory myopathies, 2915–2916, 2916t
  inhaled
    adverse effects of, 2256
    for asthma, 2224
    for COPD, 2256
    for cough, 274
  intraarticular, for osteoarthritis, 2955, 2955t
  intradiscal injections, 127
  intralesional
    for alopecia areata, 395t
    for mucous membrane pemphigoid, 413
    for pemphigus foliaceus, 411
  intranasal
    adverse effects of, 2810
    for allergic rhinitis, 2810
    for sinusitis, 242, 256, 257
  for macrophage activation syndrome, 469
  for MS, 3585
  for multiple myeloma, 890t
  myopathy, 2352
  for nausea and vomiting, 83, 299t
  for neurocysticercosis, 1834
  for neuromyelitis optica, 227, 3593
  for optic neuritis, 227
  for orbital pseudotumor, 233
  osteoporosis, 2225
  for pain, 82
  for PCP, 1734
  for polyarteritis nodosa, 2902
  in pregnancy, 412
  for relapsing polychondritis, 2921
  resistance to, 3060t
  for rheumatic fever, 2857–2858
  for rheumatoid arthritis, 2849
  for sarcoidosis, 2929, 2929t
  in STEMI, 2119
  synthesis of, 3050–3052, 3051f, 3100f
  systemic
    adverse effects of, 2256, 2896t
    for atopic dermatitis, 383
    for Behçet syndrome, 2909
    for bullous pemphigoid, 412
    for cerebral edema, 716
    for COPD exacerbation, 2259
    for cutaneous lupus erythematosus, 415
    for DIHS, 424
    for drug-induced urticaria/angioedema, 422
    for Duchenne muscular dystrophy, 3640
    for epidermolysis bullosa acquisita, 413
    for hypercalcemia, 365
    for IgG4-related disease, 2933
    for leprosy reactions, 1412
    for lichen simplex chronicus, 384
    for lupus nephritis, 2418
    for Ménière's disease, 247
    for meningitis, 1194, 1263
    for meningococcal infection, 1250
    for minimal change disease, 2421
    for *M. pneumoniae* pneumonia, 1466
    for MS, 3585
    for myasthenia gravis, 3630
    for neuromyelitis optica, 3593
    for pemphigoid gestationis, 412
    for pemphigus foliaceus, 411
    for pemphigus vulgaris, 410
    for phototoxicity, 434
    for serum sickness, 3721
    for sinusitis, 241
    for SJS/TEN, 425
    for SLE, 2834
    for systemic mastocytosis, 2821
  for systemic sclerosis, 2873, 2874
  for Takayasu arteritis, 2904
  for thrombotic thrombocytopenic purpura, 925
  topical
    for atopic dermatitis, 383
    for bullous pemphigoid, 412
    for candidiasis, 390
    for contact dermatitis, 384
    for dermatitis, 393t
    for dermatomyositis, 414
    for hand eczema, 384
    for herpes simplex keratitis, 224
    for lichen planus, 388
    for lichen simplex chronicus, 384
    for morbilliform drug eruption, 422
    for mucous membrane pemphigoid, 413
    for pemphigoid gestationis, 412
    for pemphigus foliaceus, 411
    for phototoxicity, 434
    for pityriasis rosea, 388
    for psoriasis, 388, 393t
    for seborrheic dermatitis, 386
    for stasis dermatitis, 385
    for sulfur mustard exposure, S5
    for uveitis, 225
    for vitiligo, 398t
  in transplant recipient, 2408, 2410, 2411
  for tuberculous meningitis, 983
  tumor lysis syndrome, 593
  for urticaria, 2809
  for uveitis, 225
  for vasculitis, 2895
  vessel wall disorders, 927
  withdrawal from, 420, 567, 3071
Glucocorticoid-remediable aldosteronism (GRA, familial hyperaldosteronism type I)
  complications of, 3061
  diagnosis of, 356
  genetic considerations in, 355, 3060t, 3061, 3785
  treatment of, 3063
Gluconeogenesis, 2598, 3231
Glucosamine, 2956

Glucose
  in ascitic fluid, 328
  balance and counterregulation of, 3231–3232, 3231f, 3232f
  for hyperkalemia, 361
  metabolism in cancer cells, 543–544
  plasma levels of, 3196t, 3197–3198, 3198t
  for porphyria, 3603
  regulation of, 3198, 3198–3199f
  renal transport of, 2366f, 2369
Glucose-6-phosphatase deficiency (type Ia GSD, Von Gierke disease), **3358,** 3360t, A15, S9
Glucose-6-phosphate dehydrogenase (G6PD) deficiency, **797**
  clinical features of, 323, 493, 798–800, E72t
  diagnosis of, 799f, 800
  drug contraindications in, 799–800, 799t
    dapsone, 1605
    furazolidone, 1748
    primaquine, 1751, 1770t, 1771t, 1772, 1775
    rasburicase, 594
    tafenoquine, 1780
  drug responses in, 493, E72t
  early interventions for, 3804t
  epidemiology of, 797–798, 799f
  genetic considerations in, 796t, 797, 797f
  genetic testing for, 3804t
  global considerations in in, 797, 798, 799f
  malaria resistance and, 797, 1762–1763
  pathophysiology of, 792f, 797, 798f
  red cell enzyme abnormalities in, 796t
  RMSF in, 1457
  treatment of, 800
Glucose-6-phosphate isomerase, 2790t
Glucose-6-phosphate isomerase (GPI) deficiency, 792f, 796t
Glucose-galactose malabsorption, 2545
Glucose tolerance, 3197–3198, 3255, 3258, 3950–3951, 3954
Glucose transporter 1 (GLUT-1), 276, 1553
Glucose transporter 1 (GLUT-1) deficiency, 3412t, 3420
Glucose transporter 2 (GLUT-2), 276
Glucose transporter 2 (GLUT-2) deficiency (Fanconi-Bickel syndrome, type XI GSD), 2368t, 3236, 3360t, 3363
α-Glucosidase inhibitors, 3211t, 3214, 3215
Glue sniffer's neuropathy, 3611
GluK2 antibodies, 748
GLUR (glutamate receptor), 2790t
GLUT-1 (glucose transporter 1), 276, 1553
GLUT-1 (glucose transporter 1) deficiency, 3412t, 3420
GLUT-2 (glucose transporter 2), 276
Glutamate, 94
Glutamate decarboxylase (GAD65), 2790t
Glutamate receptor (GLUR), 2790t
Glutamic acid decarboxylase (GAD) antibodies, 744t, 746, 749, 750, 750f
Glutaminase deficiency, 3369t
ʟ-Glutamine, 775
Glutamine metabolism disorders, 3369t
Glutamine synthase deficiency, 3369t
Glutaric acidemia, 3369t
Glutathione peroxidase deficiency, 800
Glutathione-S-transferase proteins, 321
Glutathione synthase deficiency, 796t
Gluten-free diet
  for ataxic disorders, 3537
  for celiac disease, 2546
  for dermatitis herpetiformis, 412–413
Gluten-induced enteropathy. *See* Celiac disease
Gluten intolerance, nonceliac, 309, 2546
Glutethimide, 489t
Glyburide, 923t, 3211t, 3901
Glycan shield, 1587
Glycemic control
  circadian rhythm disorders and, 3950–3951
  in critically ill patient, 2297–2298, 3219–3220
  in diabetes mellitus, 3207–3209, 3207t, 3209t, 3222–3223, 3223f
  diabetic nephropathy and, 3224
  diabetic retinopathy and, 3223
  in metabolic syndrome, 3258
  in older adults, 3209t, 3882, 3882t
  perioperative, 3219–3220, 3909
  in pregnancy, 3900–3901
Glyceraldehyde-3-phosphate dehydrogenase (GAPD) deficiency, 796t
Glycine, 3322
Glycine encephalopathy, 3368t, 3374t
Glycine metabolism disorders, 3368t
Glycine *N*-methyltransferase deficiency, 3368t
Glycine receptor (GlyR) antibodies, 744t, 748, 750, 750f
Glycogen, 3358
Glycogen branching enzyme deficiency (type IV GSD), 3360t, **3362**
Glycogenolysis, myocardial, 1853–1854
Glycogen storage diseases (GSD), **3358**
  clinical features of, 3358
  epidemiology of, 3358
  global considerations in, 3366–3367
  IBD and, 2553t
  late-onset, 3354t
  liver
    Fanconi-Bickel syndrome, 2368t, 3236, 3360t, 3363
    type 0, 3360t
    type Ia, **3358,** 3360t, A15, S9
    type Ib, 3360t
    type III, **3359,** 3360t, 3645
    type IV, 3360t, **3362,** 3645
    type VI, 3360t, **3362**
    type IX, 2368t, 3360t, **3363**
  muscle
    Danon disease (LAMP2 deficiency), 2004t, 2010, 3354t, 3361t, 3365
    PRKAG2 syndrome, 3361t, 3365
    type II (Pompe disease, acid α-1,4 glucosidase deficiency), 3354t, **3357,** 3361t, **3363,** 3644
    type V (McCardle disease), 3361t, **3364,** 3644
    type VII (Tarui disease), 796t, 3361t, **3364**
    type IX (muscle phosphorylase kinase deficiency), 3361t, **3365**
    type X (phosphoglycerate mutase deficiency), 3361t, 3365
    type XI (lactate dehydrogenase deficiency), 3361t, 3365
    type XII (fructose 1,6-bisphospate aldolase A deficiency), 3361t
    type XIII (β-enolase deficiency), 3361t
    type XV (polyglucosan body myopathy), 3361t, **3365**
  pathogenesis of, 3358, 3359f
Glycolysis, 792f, 796, 1854, 3644
Glycopeptides, 753t, 1166, 1176, 1182t, 1183–1184
Glycoprotein hormones, 2981
Glycoprotein Ib/IX, 2790t
Glycoprotein Ib-IX-V complex, 938
Glycoprotein IIb/IIIa, 470, 942, 943f
Glycoprotein IIb/IIIa antagonists. *See also specific drugs*
  actions of, 945
  adverse effects of, 946
  discontinuation before lumbar puncture, S3
  dosage of, 945
  indications for, 945
  for NSTE-ACS, 2110–2111
  in PCI, 2127
  pharmacology of, 945t
  for STEMI, 2121
Glycoprotein IIb/IIIa receptors, 921, 938
Glycoprotein IIb/IIIg, 2790t
Glycoproteinosis(es), 3353t
Glycosaminoglycans, 331
Glycosphingolipodoses, neutral, 3353t
Glycosuria, 347, 2368t
Glycylcyclines. *See* Tigecycline
Glycyrrhizinic acid, 355, S1
Glymphatic system, 3398
GMAP (Global Malaria Action Plan), 3849
GM-CSF (granulocyte-macrophage colony-stimulating factor), 454f, 455, 2766t
GM-CSF (granulocyte-macrophage colony-stimulating factor) therapy, 561, 575t
GM gangliosides, 966–967, 2790t, 3522t
GNA11 gene mutations, 3279
GNAL gene mutations, 3510t, 3511
GNAS1 gene mutations, 3056, 3063, 3088, 3184, 3317, 3783, 3789
GNAS gene mutations, 3288–3289
Gnathostoma spinigerum/gnathostomiasis
  clinical features of, 1131t, 1738t, 1739t, 1814
  diagnosis of, 961t, 1814, S12
  epidemiology of, 961t, 1131, 1813–1814
  pathogenesis of, 1737, 1814
  treatment of, 1812t, 1814
GnRH. *See* Gonadotropin-releasing hormone (GnRH)
GnRH1 gene mutations, 3112t, 3113
GnRHR gene mutations, 3112t, 3113
GOBI, 3843
GOBI-FFF, 3843
Goiter, 3029, 3040. *See also* Thyroid nodular disease
Goitrogens, 3041
Goitrous thyroiditis. *See* Hashimoto's thyroiditis
Gold, adverse effects of, 388, 400, 420, 923t, 2907, 3610t
GOLD (Global Initiative for Lung Disease) criteria, for COPD, 2254
Golden Rice, 2612
Golgin, 2790t
Golimumab
  actions of, 2791t
  adverse effects of, 387t, 462, 2566, 2568t, 2850t
  for autoimmune and inflammatory diseases, 2784, 2791t
  for axial spondyloarthritis, 2887t
  for IBD, 2566, 2568t
  for psoriasis/psoriatic arthritis, 387t, 2890
  for rheumatoid arthritis, 2850t, 2851
  testing before and during treatment, 2568t
Gomori methenamine silver stain, S11
Gompertzian growth curve/kinetics, 553, 553f
Gonadal development, 3096f
Gonadal dysgenesis, 3098
Gonadal mosaicism, 3784
Gonadal sex, 3094, 3096f
Gonadectomy, 3133
Gonadotropin(s), **2991**. *See also* Follicle-stimulating hormone (FSH); Luteinizing hormone (LH)
  actions of, 2991
  deficiency of. *See* Hypogonadism
  in male reproductive function, 3107
  for ovulatory induction, 3151
  pituitary tumor production of, 3012–3013
  secretion of, 2991
  synthesis of, 2991
  for testosterone deficiency, 3119
Gonadotropinoma, 3001t
Gonadotropin-releasing hormone (GnRH)
  actions of, 2989f, 2991
  deficiency of. *See* Hypogonadotropic hypogonadism

  - hypothalamic, 2980
  - in male pubertal development, 3102–3103
  - in menstrual cycle regulation, 3128
  - in ovarian function regulation, 3126–3127, 3127f
  - pulse frequency of, 2987, 2991
- Gonadotropin-releasing hormone (GnRH) analogues, 3110
- Gonadotropin-releasing hormone (GnRH) receptor, 2985t
- Gonadotropin-releasing hormone (GnRH) stimulation test, 3107
- Gonioscopy, 225
- Gonococcal arthritis. *See* Gonococcal infections, disseminated
- Gonococcal infections, **1253**
  - anorectal, 1256
  - in children, 1257
  - definition of, 1253
  - diagnosis of, 1258, S11
  - disseminated
    - clinical features of, 144t, 1057–1058, 1257, 1257f
    - complications of, 1257
    - epidemiology of, 144t, 1257
    - joint involvement in, 1257, 1258f
    - risk factors for, 1257
    - skin manifestations of, 144t, 148, 408, 1257f, A1, A5
    - treatment of, 959t, 1058, 1258–1260
  - epidemiology of, 1097, 1253–1254
  - epididymitis, 1099
  - in HIV infection, 1258
  - in men, 1098–1099, 1255–1256, 1255f
  - in neonates, 1257, 1259t
  - ocular, 1256, 1259t
  - oropharyngeal, 259, 262t, 1254, 1256, 1259t
  - pathogenesis of, 1254–1255. *See also Neisseria gonorrhoeae*
  - pelvic inflammatory disease. *See* Pelvic inflammatory disease (PID)
  - in pregnancy, 1256–1257
  - prevention and control of, 1260
  - screening for, 39t, 1110
  - treatment of, 1099, 1099t, 1258–1260, 1259t
  - in women, 1099–1100, 1256
- Go–no go test, 208
- Gonorrhea. *See* Gonococcal infections
- Goodpasture's syndrome
  - autoantibodies in, 2825t
  - clinical features of, 2414, 2415t, 2416, 2418
  - diagnosis of, 2418
  - pathophysiology of, 2412, 2790t, 2825t
  - prognosis of, 2418
  - treatment of, 2418
- Gordon's syndrome. *See* Pseudohypoaldosteronism type II
- Gore atrial septal occluder, 2130, 2130f
- Gorlin formula, valve area measurement, 1906–1907
- Gorlin's syndrome (basal cell nevus syndrome)
  - brain tumors associated with, 717t, 721
  - clinical features of, 404, 1900f
  - genetic considerations in, 524t, 717t
  - nonmelanoma skin cancer in, 606
- Goserelin, 567, 637, 701
- Gottron's papules/sign, 414, 415f, 2911, 2911f
- Gout, **2956**
  - clinical features of, 2958
  - diagnosis of, 2958–2959, 2959f
  - drug-induced, 2940t
  - nephropathy in, 2443
  - obesity and, 3187
  - pathogenesis of, 2956–2957, 2957–2958f
  - vs. rheumatoid arthritis, 2847
  - tophaceous, 404, A15
  - treatment of, 2957f, 2959–2960
  - in von Gierke disease, 3359
- Gou virus, 1661t
- Gower sign, 3634, 3637f
- gp41, 1571, 1587
- gp120, 1558, 1571, 1579, 1627
- gp160, 1579
- GPI (glucose-6-phosphate isomerase) deficiency, 792f, 796t
- *GPIHPB1* gene mutations, 3240
- G-POEM (gastric peroral endoscopic myotomy), 299
- G protein(s), 2983, 2984f, 3689
- G protein α, 2983, 2984f
- G protein–coupled receptors (GPCRs), 237
  - effectors and pathways of, 2981, 2981t, 2984f
  - mutations in, 2978, 2983–2984, 2985t
- GPT (generative pretrained transformer), 3970
- GRA. *See* Glucocorticoid-remediable aldosteronism (GRA)
- Graafian follicle, 3126f
- Graded exercise therapy, 167
- Gradenigo's syndrome, 235, 1142
- Gradient-boosted trees, 3972t
- Graft-versus-host disease (GVHD)
  - chronic, 917–918
  - clinical features of, 911, 917
  - cobalamin malabsorption in, 787
  - diarrhea in, 306, 309
  - grading and staging of, 917, 917t
  - in HCT recipient, 914, 917, 1154, 1156, 1216
  - incidence of, 917
  - microbiome and, 3837
  - neuropathy in, 3607
  - prevention of, 914, 917
  - skin manifestations of, 388, 392, 401, 402
  - transfusion-associated, 810, 911
  - treatment of, 917
- Graft-versus-leukemia effect, 831, 1154
- Graft-versus-tumor effect, 914, 1154
- Graham Steell murmur, 287, 2049, 2064
- Grain dust exposure, 2236t, 2239
- Gramicidin, 1180t
- Grammar, assessment of, 201
- Gram-negative enteric bacilli. *See also Enterobacterales; specific bacteria*
  - antibiotic resistance in, 1027, 1152, 1182t, 1183f, 1184, 1282–1285, 1283t
  - antibiotic targets in, 1183f
  - epidemiology of, 1281
  - global considerations in, 1281
  - infectious syndromes of, 1282
  - laboratory identification of, 1282, S11
  - pathogenesis of, 1281–1282, 1282t
  - structure and function of, 1281
- Gram-positive bacteria, S11. *See also specific bacteria*
- Gram stain, 1025, S11
- Granada virus, 1662t
- Grandmother effect, 3869, 3871f
- Granisetron, 82, 116, 299t, 575
- Granular casts, A4
- Granular lymphocytes, 443
- *Granulicatella* spp., 1041, 1214
- Granulocyte(s), 441f, 443, 458t, 760f, 903t, 907t. *See also* Basophil(s); Eosinophil(s); Neutrophil(s)
- Granulocyte colony-stimulating factor (G-CSF), 454f, 1023, 2766t
- Granulocyte colony-stimulating factor (G-CSF) therapy
  - adverse effects of, 405, 420
  - in HCT, 759, 915
  - indications for, 561, 575t
  - for MDS, 818
  - for neutropenia, 463
  - for sulfur mustard exposure, S5
- Granulocyte-macrophage colony-stimulating factor (GM-CSF), 454f, 455, 2766t
- Granulocyte-macrophage colony-stimulating factor (GM-CSF) therapy, 561, 575t
- Granulocytic sarcoma (cutaneous chloroma), 405
- Granulocytosis, 740t, 741
- Granuloma(s)
  - *Candida,* 1708
  - caseating, 2356
  - in coccidioidomycosis, 1696
  - diagnosis of, 382
  - etiology of, 2924t
  - hepatic, 2669t, A13
  - palisading, 404
  - periapical, 260
  - pyogenic, 261
  - in sarcoidosis, 406, 2356, 2925, 2927. *See also* Sarcoidosis
  - in schistosomiasis, 1828
  - tick-bite, 3730
  - in vasculitis, 2894
- Granuloma annulare, 394, 403, 3230
- Granuloma inguinale. *See* Donovanosis
- Granulomatosis
  - eosinophilic, with allergic angiitis. *See* Eosinophilic granulomatosis with polyangiitis
  - lymphomatoid, 873–874, 2898
- Granulomatosis infantiseptica, 1229
- Granulomatosis with polyangiitis (Wegener's), **2896**
  - clinical features of, 1132t, 2420, 2897
    - bacterial pneumonia, A14
    - bilateral nodular infiltrates, A14
    - chronic meningitis, 1132t
    - cutaneous, 406, 2897
    - hemoptysis, 275
    - interstitial lung disease, 2266
    - intracranial infarcts, 3447
    - ocular, 224, 2897, A14
    - oral, 261, 266
    - renal, 2412, 2415t, 2420, 2897
    - saddle nose deformity, 1860, 2897
    - sinusitis, A14
  - definition of, 2896
  - diagnosis of, 2897–2898
    - ANCAs, 2898
    - chest imaging, A14
    - CT, 2896f, A14
    - lung biopsy, A14
    - lung histology, 2896f, 2897
    - renal biopsy, 2420, A4, A14
  - differential diagnosis of, 2898
  - incidence and prevalence of, 2896
  - pathogenesis of, 2894, 2896–2897
  - pathology of, 2896–2897, 2896f
  - treatment of, 2894, 2898–2899
- Granulomatous (necrotizing) arteritis, 3447
- Granulomatous diseases, chronic. *See* Chronic granulomatous diseases (CGDs)
- Granulomatous gastritis, 2539
- Granulomatous interstitial nephritis, 2439
- Granulomatous (subacute) thyroiditis, 3038, 3038f
- Granulosa cell(s), 3124
- Granulosa cell tumor, 712
- Grapefruit juice, 1748, 2102
- Graphesthesia, 174, 3382
- Graphics processing unit, 3972, 3973t
- Grasp reflex, 3382
- Graves' disease, **3032**. *See also* Thyrotoxicosis
  - anti-*Yersinia* antibodies in, 1349
  - APS-2 and, 3091, 3091t
  - autoantibodies in, 2825, 2825t
  - in children, 3037

- Graves' disease (*Cont.*):
  - clinical features of, 3033–3035, A15
    - goiter, 3040
    - hypopigmentation, 396
    - ophthalmopathy. *See* Graves' (thyroid-associated) ophthalmopathy
    - pretibial myxedema, A5, A15
    - signs and symptoms, 3033t, 3034f
    - thyroid acropachy, 2971, 3034f
    - thyroid dermopathy, 3034f
  - diagnosis of, 2979t, 3026, 3034, 3034f
  - differential diagnosis of, 3034
  - edema in, 282
  - epidemiology of, 3032
  - genetic considerations in, 3032–3033
  - insulin autoimmune syndrome in, 3093
  - pathogenesis of, 2790t, 2978, 3032–3033
  - in pregnancy, 2983, 3036–3037
  - pregnancy and, 3901
  - prevalence of, 2979t
  - radiation therapy-related, 756
  - thymic tumors in, 3092
  - treatment of, 3035–3037
- Graves' (thyroid-associated) ophthalmopathy
  - autoantibodies in, 2790t
  - clinical features of, 233, 3034f, A15, V3
  - diagnosis of, 233
  - pathogenesis of, 3033
  - treatment of, 233, 3037
- Gray (Gy), 554, S6
- Gray hepatization, 1023
- Gray platelet syndrome, 437
- Grazoprevir, 2690
- Grazoprevir/elbasvir. *See* Elbasvir/grazoprevir
- Great saphenous vein, ligation of, 2182
- Green bottle fly, 3733
- Greenhouse gases, 3739, 3741, 3745–3746, 3745f, 3746f, 3746t
- Green nail syndrome, 1312
- Green tea extracts, 390
- Grey Turner sign, 807
- Griffiths' point, 2587, 2587f
- *GRIN2A* gene mutations, 3412t
- Griscelli syndrome type 2, 465, 466, 2803, S9. *See also* Hemophagocytic lymphohistiocytosis (HLH)
- Griseofulvin, 388t, 389, 395t
- *GRN* gene mutations, 195t, 3395, 3486
- "Groove sign," 416
- Gross cystic disease fibrous protein, 733, 733t
- Ground-glass opacity
  - bleomycin-related, 754
  - in COVID-19 disease, 996
  - in granulomatosis with polyangiitis, A14
  - in interstitial pneumonitis, A12
  - in lung cancer, A12
  - pulmonary, 582, 2236
  - in pulmonary alveolar proteinosis, A12
  - radiation therapy-related, 754
  - treatment of, 622–623
- Group A *Streptococcus* (GAS) infections, **1206**. *See also Streptococcus pyogenes*
  - asymptomatic carrier of, 1209
  - cellulitis, 1052, 1210–1211
  - complications of. *See* Poststreptococcal glomerulonephritis; Rheumatic fever
  - COVID-19 pandemic and, 1207
  - dactylitis, 1049
  - diagnosis of, S11
  - ecthyma, 389
  - endocarditis. *See* Infective endocarditis
  - epidemiology of, 1206–1207
  - erysipelas. *See* Erysipelas
  - global considerations in, 1206–1207, 1207f
  - impetigo. *See* Impetigo
  - joints, 1056. *See also* Infectious (septic) arthritis
  - myositis, 1053, 1211
  - necrotizing fasciitis. *See* Necrotizing fasciitis
  - pathogenesis of, 1207
  - pharyngitis. *See* Pharyngitis, streptococcal
  - pneumonia, 1211
  - prevention of, 1212
  - puerperal, 1211
  - pyomyositis, 1053
  - rheumatic fever, 1206–1207, 2854, 2856, 2858. *See also* Rheumatic fever
  - in scabies, 3729
  - scarlet fever. *See* Scarlet fever
  - toxic shock syndrome. *See* Toxic shock syndrome (TSS), streptococcal
  - treatment of, 1208–1209, 1208t
- Group B *Streptococcus* (GBS) infections
  - cellulitis, 1052
  - clinical syndromes, 1213–1214
  - diagnosis of, 1213, S11
  - endocarditis. *See* Infective endocarditis
  - epidemiology of, 1213
  - meningitis, 1119
  - pharyngitis, 259
  - in pregnancy, 1213
  - prevention of, 1213
  - treatment of, 1213, 1214
- Group C *Streptococcus* infections, 259, 1212–1213
- Group D *Streptococcus* infections, 1214
- Group G *Streptococcus* infections, 1212–1213
- Grover's-like rash, 421
- Growth
  - disorders of, **2994**. *See also* Short stature
  - hormones in, 2986. *See also* Growth hormone (GH)
  - somatic, skeletal maturation and, 2994
- Growth hormone (GH), **2989**
  - actions of, 2990, 3232t
  - Creutzfeldt-Jakob disease after, 3529
  - ectopic production of, 737t
  - effectors and signaling pathways of, 2981t
  - excess. *See* Acromegaly
  - expression and regulation of, 2988t, 3949–3950
  - insensitivity, 2994
  - laboratory evaluation of, 2996t
  - in puberty, 3102, 3130
  - recombinant
    - for adult growth hormone deficiency, 2997, 2997f
    - adverse effects of, 2997
    - for growth hormone deficiency in children, 2996
    - for Prader-Willi syndrome, 3184
  - secretion of, 2990
  - synthesis of, 2989–2990
- Growth hormone (GH) deficiency
  - in adults, **2996,** 2996t, 2997f
  - in children, 2994, 3235. *See also* Short stature
  - idiopathic, 2994
  - laboratory investigation of, 2995t
  - radiation therapy-related, 756
- Growth hormone (GH) receptor(s), 2984
- Growth hormone (GH) receptor antagonists, for acromegaly, 3010
- Growth hormone-releasing hormone (GHRH), 737t, 2990, 2994, 3003
- Growth hormone-releasing hormone (GHRH) agonists, 3157
- Growth hormone-releasing hormone (GHRH) neuroendocrine tumors, 3083
- GRO-α/MGSA, 2768t
- GRO-β/MIP-2α, 2768t
- $G_s\alpha$, 3280f, 3288–3289, 3288t
- GSD. *See* Glycogen storage diseases (GSD)
- GSDMD (gasdermin D), 972
- GSK-J4, 3942
- GSS (Gerstmann-Sträussler-Scheinker) disease, 3401, 3401t, 3526t, 3528, 3530
- GTPases, 969, 972f, 973
- GTP-cyclohydrolase 1 deficiency, 3497
- GTP cyclohydrolase I deficiency, 3368t
- Guamá virus, 1658t
- Guanabenz, 3691, 3711t
- Guanarito virus, 1657t
- Guanethidine, 282t
- Guanfacine, 2145t, 3515, 3691
- Guanine, 3778
- Guanine nucleotide exchange factors (GEFs), 969
- Guanylate binding protein (GBP), 973
- Guarana, 2137
- Guaroa virus, 1657t
- Guillain-Barré syndrome (GBS), **3617**
  - after *Campylobacter* infection, 1305
  - after infectious diarrhea, 1079t
  - antecedent events, 3617–3618
  - vs. botulism, 1236
  - vs. CIDP, 3622
  - clinical features of, 3617, 3621t
    - autonomic dysfunction, 3550, 3617
    - cranial nerve palsies, 3559, 3617
    - facial nerve disorders, 3555
    - weakness, 170, 172, 187, 3617
  - in CMV mononucleosis, 1514
  - in COVID-19 disease, 3606
  - diagnosis of, 3617t, 3619–3620, A16
  - differential diagnosis of, 3619–3620
  - epidemiology of, 3617–3618
  - immunopathogenesis of, 2790t, 3398, 3618–3619, 3619t, 3620f
  - in influenza, 1548
  - influenza vaccine and, 1549, 3617
  - Miller Fisher variant. *See* Miller Fisher syndrome
  - pathophysiology of, 3619
  - prognosis of, 3621–3622
  - subtypes of, 3617, 3617t, 3618f
  - treatment of, 3620–3621
- Guinea worm infection (dracunculiasis). *See Dracunculus medinensis*/dracunculiasis
- Gulf War syndrome, S7, S8
- Gumboil, 260
- Gummas, 1051, 1430, 1432
- Gums, strawberry, 261
- Günther's disease (congenital erythropoietic porphyria), 3341t, 3342, **3349**
- *GUSB* gene mutations, 3357
- Guselkumab, 387t, 2793t, 2890, 2892
- Gut-associated lymphoid tissue (GALT), 1570, 1574
- Gut microbiota. *See also* Human microbiome
  - aging and, 3871
  - anaerobic bacteria in, 1371, 1372
  - benefits of, 1372
  - circadian regulation of, 3952
  - composition of, 3832, 3833f
  - in gastrointestinal disease, 1077, 3835
  - in heart failure, 1982
  - in IBD, 2552f, 2554, 2573, 2573f, 2574f, 2576–2577, 3836
  - immune checkpoint blockade and, 2788
  - in metabolic syndrome, 3255
  - metabolites of, 3841
  - in neurologic disease, 3398–3399
  - nutrients and, 3840–3841
  - in obesity, 3835
  - in PD, 3499
  - in rheumatoid arthritis, 2843
  - in spondyloarthritis, 2882
- Guttate psoriasis, 386. *See also* Psoriasis
- GVHD. *See* Graft-versus-host disease (GVHD)

GWAS. *See* Genome-wide association studies (GWAS)
Gy (Gray), 554, S6
GYG1 gene mutations, 3365
Gynecomastia, **3115**
  with androgen deprivation therapy, 702, 3176–3177
  definition of, 3115
  diagnosis of, 2979t
  etiology of, 3116–3117
  evaluation of, 3116–3117, 3116f
  hCG levels in, 740
  in Klinefelter syndrome, 3096
  in liver disease, 323, 327
  prevalence of, 2979t, 3115
  spironolactone-related, 329
  treatment of, 3117
Gypsy moth larvae, 3737
Gyrate atrophy, choroid and retina, 3369t

H

H1N1 influenza
  epidemic potential of, 987
  epidemiology of, 995, 3856
  pandemic, 1149, 1533, 1545, 3750, 3750t
  pathogen evolution of, 987
  pathophysiology of, 3398
  risk factors for, 995
H_1 receptor antagonists. *See* Antihistamines
H_2 receptor antagonists. *See* Histamine H_2 receptor antagonists
H3FA gene mutations, 720
H3K4me3, 3939, 3941
H3K9ac, 3939
H3K27ac, 3939
H3K27me3, 3939, 3940–3941
H3N2 influenza, 987, 1533
H5N1 influenza, 468, 987, 995, 1533, 1545
H6PDH (hexose-6-phosphate dehydrogenase), 3054
H7N9 influenza, 995, 1533
HAART (highly active antiretroviral therapy). *See* Combination antiretroviral therapy (cART)
HACE (high-altitude cerebral edema), 277, 3755–3756, 3755f, 3756t
HACEK group, 1265–1266
HACEK group infections, 1265–1266, 1267t. *See also* Infective endocarditis, HACEK group
HAD (HIV-associated dementia), 1606, 1607f, 1607t
HAE (hereditary angioedema), 2808
Haemaphysalis longicornis tick, 1677
Haemophilus ducreyi, 1107, 1108, 1263–1264. *See also* Chancroid
Haemophilus influenzae
  antibiotic resistance in, 1263
  host response to, 1261
  laboratory identification of, 1262–1263, S11
  microbiology of, 1260
  nontypable strains, 1260, 1260t
  type b, 1260, 1260t
Haemophilus influenzae infections, **1260**
  bronchiectasis, 2243
  clinical features of, 1260t, 1261–1262
    cellulitis, 1052, 1262
    epiglottitis, 259, 1262
    hemoptysis, 274
    laryngitis, 259
    mastoiditis, 255
    meningitis, 1119, 1261–1262, 1263
    otitis media, 254, 1262
    pharyngitis, 257t, 259
    pneumonia, 1262, 1262f
  in COPD, 2259
  in cystic fibrosis, 2245
  diagnosis of, 1262–1263
  epidemiology of, 1260–1261, 1261f
  global considerations in, 1261
  in human bite-wound, 1144
  incidence pre- and postvaccine, U.S., 1002t
  pathogenesis of, 1261
  in post-splenectomy sepsis, 482, 990
  prevention of, 1263
  in solid organ transplant recipient, 1155t
  treatment of, 1173t, 1185, 1263
Haemophilus influenzae vaccine, 482, 1263
Haemophilus parainfluenzae, 1266, 1267t. *See also* Infective endocarditis, HACEK group
Haemophilus vaginitis. *See* Bacterial vaginosis
HAI (hemagglutinin inhibition) antibodies, 1546
Hair
  drug-induced disorders of, 394, 421, 576
  excessive. *See* Hirsutism
  growth cycle of, 3138
  loss of. *See* Alopecia
  removal of, 3141
  terminal, 3137
  vellus, 3137
Hairy cell leukemia, **872**
  clinical features of, 872–873, 2901, 2972
  immunophenotype of, 872t, 873
  infections in, 577
  monocytopenia in, 462
  splenomegaly in, 480
  treatment of, 481, 873
  tumor markers in, 504t
Hakki formula, aortic valve area measurement, 1907
Half-life
  of circulating hormone, 2982
  drugs, 485f
  of drugs, 485, 1167
Half-normal saline, 347. *See also* Saline
Halide exposure, 2239t
Halitosis, 266–267
Hallervorden-Spatz disease. *See* Pantothenate kinase–associated neurodegeneration (PANK)
Hallucinations
  chemosensory, 242
  in delirium, 182
  in dementia, 196, 196t
  in DLB, 3491–3492
  hypnagogic, 214
  hypnopompic, 214
  in PD, 3506
  in respiratory acidosis, 375
Hallucinogens, 3696
Hallucinosis, peduncular, 3432, 3432f
Halocarbons, 3746t
Halofantrine
  actions of, 1748
  adverse effects of, 1741t, 1771t
  drug interactions of, 1741t, 1744t, 1748
  for malaria, 1771t
  pharmacology of, 1748, 1771t
  in pregnancy and lactation, 1741t
Haloperidol
  adverse effects of, 3673t
  for delirium, 86, 86t, 2348
  for nausea and vomiting, 83, 299t, 300, 509, 575
  overdosage/poisoning with, 3711
  for schizophrenia, 3673, 3673t
Halo sign, 1716, A14
Halting rules, in clinical trials, 999
Halzoun, 3733
HAM (HTLV-1-associated myelopathy), 1553–1555, 1556, 3567
Hamartin, 687t, 723, 2434
Hamartoma
  hypothalamic, 3003, A16
  iris, 3087
  pulmonary, 622
  of small bowel, 651
Hamman-Rich syndrome (acute interstitial pneumonia), 2265
Hamman's sign, 2273
Hampton's hump, 107, 2160
HAND (HIV-associated neurocognitive disorders), 1585, 1606, 1607f, 1607t
Hand(s)
  cryoglobulinemia in, A14
  diagnostic clues in, 50t
  eczema of, 380f, 384, 384f
  frostbite of, A5
  injury, bite-wounds, 1143–1144. *See also* Bite-wounds
  osteoarthritis of, 2949, 2949f
  pain in, 2942–2943, 2942f
Hand-foot-and-mouth disease
  epidemiology of, 1635–1636
  global considerations in, 1635–1636
  oral manifestations of, 262t, 1535, 1635, 1635f, A1, A5
  skin manifestations of, 140t, 146, 1049, 1635, 1635f
Hand-foot syndrome, 573, 772, 2967
Handgrip strength, 2619t
Hand hygiene, 1148, 1538
Handoffs, 3, 7
Hand-Schüller-Christian disease, 3003
Hansen's disease. *See* Leprosy
Hantaan virus, 1661t, 1676
Hantaviridae
  clinical syndromes of, 1666t
  geographic distribution of, 1664–1665t
  reservoirs and vectors of, 1660–1661t
  structure of, 1478f, 1480t
  virology of, 1656
Hantavirus infections
  clinical features of, 1539
  epidemiology of, 993, 996, 1539, 1674, 3750, 3753f, 3856
  hemorrhagic fever with renal syndrome, 993, 1676–1677
  pulmonary syndrome, 996, 1539, 1674
HAPE. *See* High-altitude pulmonary edema (HAPE)
Haplogroup, 3807
Haploid, 3777
Haploinsufficiency, 3786
Haplotype(s), 3779, 3781f, 3787
Haplotype block, 3781f
HapMap project, 3777, 3794
Haptocorrins, 781
Hard (ixodid) tick, 1445, 3730. *See also* Tick(s)
Hard tick relapsing fever (HTRF)
  clinical features of, 1446
  diagnosis of, 1446
  epidemiology of, 1445, 1445t
  etiology of, 1445
  pathogenesis of, 1446
  treatment of, 1447f, 1448
Hartmann's procedure, 2582, 2582f
Hartnup disease/syndrome, 2609, 3374t, **3375**
Harvester ants, 3736
Hashimoto's encephalopathy, 3030, 3483, A16
Hashimoto's thyroiditis (autoimmune hypothyroidism)
  after DIHS, 424
  autoantibodies in, 2825, 2825t
  classification of, 3028
  clinical features of, 396, 3028t, 3029–3030, 3029f, 3039, 3040
  differential diagnosis of, 3030–3031

- Hashimoto's thyroiditis (autoimmune hypothyroidism) (*Cont.*):
  - diseases associated with, 3030
  - inflammasome mutations in, 2761t
  - laboratory evaluation of, 3030, 3030f
  - pathogenesis of, 2790t
  - prevalence of, 3028
  - thyroid lymphoma in, 3048
  - treatment of, 3031–3032
- Hastisetae, 3737
- HAV (hepatitis A virus), 2643, 2643–2644f, 2645t. *See also* Hepatitis A virus (HAV) infection
- Haverhill fever (rat-bite fever), 141t, 1143
- Haversian systems, 3259, 3261
- HAV infections. *See* Hepatitis A virus (HAV) infection
- Hawkinsinuria, 3368t
- Haw River syndrome, 3536
- *HAX-1* gene mutations, 457, S9
- Hayflick limit/phenomenon, 3818, 3870
- Hb. *See* Hemoglobin (Hb)
- HBAT (heptavalent botulinum antitoxin), S4
- HBcAg (hepatitis B core antigen), 2644, 2645t
- HBeAg (hepatitis B e antigen), 2644, 2645t
- HBIG (hepatitis B immunoglobulin), 1146, 1153, 2662–2664
- HBsAg (hepatitis B surface antigen), 2644, 2645t
- HBV. *See* Hepatitis B virus (HBV)
- HBV infection. *See* Hepatitis B virus (HBV) infection
- HBxAG (hepatitis B x antigen), 2644f, 2645
- *HBZ,* 1552
- HCC. *See* Hepatocellular carcinoma (HCC)
- hCG. *See* Human chorionic gonadotropin (hCG)
- HCOM. *See* Hypertrophic cardiomyopathy
- HCoV-NL63, 1535
- HCP. *See* Hereditary coproporphyria (HCP)
- *HCP5* gene, 1581t, 1583f, 1585
- HCT. *See* Hematopoietic cell transplantation (HCT)
- HCV (hepatitis C virus), 2645t, 2647–2648, 2648f. *See also* Hepatitis C virus (HCV) infection
- HDAC. *See* Histone deacetylase (HDAC)
- HDL-C. *See* High-density lipoprotein cholesterol (HDL-C)
- HDR (hypoparathyroidism, deafness, and renal dysplasia) syndrome, 3286
- HDV (hepatitis D virus), 2645t, 2647. *See also* Hepatitis D virus (HDV) infection
- Headache, **113, 3462**
  - in acute mountain sickness, 3756
  - acute, new-onset, 114, 114t, 3458
  - after traumatic brain injury, 3574
  - anatomy and physiology of, 113–114
  - with autonomic symptoms, 114
  - in brain abscess, 1136
  - in brain tumor, 114, 716
  - chronic daily, 114t, **115**
  - classification of, 3463t
  - cluster. *See* Cluster headache
  - cold-stimulus, 3474
  - depression and, 114
  - external pressure, 3474
  - in giant cell arteritis, 114–115, 2902
  - in glaucoma, 115
  - hemicrania continua, **3473**
  - in hemicrania continua, 115, 115t
  - in hyperperfusion disorders, 2350
  - hypnic, 3475
  - medication-overuse, **115**
  - in meningitis, 114
  - migraine. *See* Migraine
  - new daily persistent, **116**
    - clinical features of, 116, 3475
    - differential diagnosis of, 116t
    - low CSF volume headache, 116, 116f
    - posttraumatic, 117
    - pseudotumor cerebri, 117
    - raised CSF pressure headache, 117
    - treatment of, 117, 3475
  - nummular, 3474–3475
  - in papilledema, 228
  - paroxysmal hemicrania, 115, 115t, 3471t, **3472**
  - post–lumbar puncture, S3
  - in pregnancy, 3899
  - primary, 113, 113t, 3463t
  - primary care management of, 117
  - primary cough, 3473
  - primary exercise, 3473–3474
  - primary sex, 3474
  - primary stabbing, 3474
  - primary thunderclap, 3474
  - in respiratory acidosis, 375
  - secondary, 113, 113t, 114–115
  - in subarachnoid hemorrhage, 114, 3458
  - SUNCT/SUNA, 115, 115t, 3471t, **3473**
  - as symptom of serious underlying disorder, 114t
  - tension-type, 115t, 3463t, 3471
- Head and neck abscesses, 259
- Head and neck cancer, **610**
  - carcinoma of unknown primary and, 735
  - clinical features of, 611–612
  - diagnosis of, 611–612, 611f
  - epidemiology of, 610
  - etiology of, 610
  - genetic considerations in, 610
  - global considerations in, 610
  - histopathology of, 610–611
  - incidence of, 610
  - lymphadenopathy in, 476
  - paraneoplastic syndromes in, 737t
  - premalignant lesions, 610
  - second malignancies in survivors of, 611
  - staging of, 611–612, 612f
  - treatment of, 613–614
- Head impulse test, 163
- Head injury. *See* Concussion; Traumatic brain injury (TBI)
- Head lice *(Pediculus capitis),* 3731–3732
- Head tilt test, 235
- Health care–associated infections, **1147**
  - *A. baumannii,* 1295, 1297f, 1298
  - antibiotic-resistant, 1151–1153, 1151f, 1152t
  - *Candida,* 1150
  - in critically ill patient, 2297
  - diarrheal
    - *C. difficile*–associated infection, 1082, 1149–1150. *See also Clostridioides difficile*–associated infection (CDI)
    - etiology of, 1079
    - norovirus, 1150. *See also* Norovirus infection
  - emerging infectious diseases, 1149
  - endocarditis, 1039
  - enterococcal, 1216–1217
  - epidemiology of, 1147
  - genomic analysis in, 985
  - health care environment and, 1150–1151
  - HIV, 1564–1565
  - *Klebsiella* spp., 1289
  - *Legionella* pneumonia, 1272–1273. *See also Legionella* spp. infections
  - measles, 1639
  - mold, 1150
  - nosocomial and device-related
    - catheter-associated urinary tract infection, 1087, 1090, 1094, 1147
    - central venous catheter-related. *See* Catheter-related bloodstream infections
    - pneumonia, 1147–1148. *See also* Ventilator-associated pneumonia (VAP)
    - surgical-site infections. *See* Surgical-site infections
  - pneumonia, 963t, 1023, 1032t, 1033, 1200, 1286
  - prevention of, 43, 1146–1147, 1205–1206, 2297
  - *S. aureus,* 1196, 1205–1206
  - sinusitis, 256
  - surveillance, 1146–1147
- Health care costs, 86, 3844
- Health care disparities. *See* Racial/ethnic disparities in health care
- Health care proxy, 79
- Health care quality, **43**
  - data about current state of, 44
  - diagnosis in, 27
  - factors relating to, 44
  - national state of measurement of, 45
  - pay-for-performance and, 45
  - public reporting of, 45
  - quality theory and, 43–44, 44f
  - racial/ethnic disparities and, 62. *See also* Racial/ethnic disparities in health care
  - strategies for improving, 44–45, 44f
- Health care safety. *See* Adverse events
- Health care workers
  - brain drain, 3848
  - burnout in, 52–53, 52f, 54f
  - errors of, 42
  - global burden of disease and, 3862, 3862f
  - health risks of, 72, 1153, 1153t, 1564–1565, 1625
  - immunizations in, 1005, 1005t, 1010, 1153
  - on medical missions, 1022t
  - minority, 68
  - shift-work disorder in, 219
  - shortages of, 55
- Health insurance, 1021, 3919, 3921
- Health promotion, 9
- Health systems reform, 3859–3860, 3863–3864, 3863f
- Healthy volunteer effect, 515
- Hearing, 244, 245f
- Hearing aids, 246, 250
- Hearing loss, **244**
  - age-related, 246, 246f
  - approach to the patient, 248, 248t
  - audiologic assessment of, 248–249
  - from barotrauma, 3766
  - conductive, 245, 246f
  - in congenital rubella, 1644
  - drug-induced, 246
  - evoked responses in, 249–250
  - genetic causes of, 247–248, 3374t, 3807, 3811t, 3812, E36t
  - imaging studies in, 250
  - mixed, 246f, 247
  - in mumps, 1648
  - noise-induced, 246–247, 249, 251, 251–252t
  - in osteogenesis imperfecta, 3326
  - in otitis media, 255
  - physical examination in, 248
  - prevention of, 251–252, 251t
  - sensorineural, 246, 246f, 247, 3811t
  - sudden, 248
  - in syphilis, 1431, 1433–1434
  - treatment of, 250–251, 251f
  - vertigo with, 163
- Heart
  - control of performance and output of, 1850–1852, 1851f, 1852t
  - diastolic function of, 1853, 1853f
  - embryogenesis of, 1843–1844, 1844f

  - functional assessment of, 1852–1854, 1853f
  - metabolism in, 1853–1854
  - normal blood flow in, 2068f
  - physical examination of, 1864. *See also* Cardiac auscultation; Heart murmurs; Heart sounds
  - regenerative potential of, 1854
- Heart block. *See* Atrioventricular (AV) conduction block
- Heartburn, 300, 2460, 2503. *See also* Gastroesophageal reflux disease (GERD); Indigestion
- Heart disease. *See* Cardiovascular disease
- Heart failure (HF), **1975**
  - in AA amyloidosis, 899
  - acute, 1976
  - approach to the patient, 142–144, 1982t, 1983t
  - in carcinoid syndrome, 683
  - cardiac cirrhosis in, 2712
  - cardiogenic shock in. *See* Cardiogenic shock
  - chronic, 1976
  - in CKD, 2391–2392
  - clinical features of
    - abdomen and extremities, 1984
    - anemia, 768, 769, 1998
    - ascites, 328
    - atrial arrhythmias, 1998
    - cardiac, 1984
    - central sleep apnea, 2282
    - depression, 1998
    - dyspnea, 270, 1882, 2199
    - edema, 281, 282
    - exercise intolerance, 292, 292f
    - fatigue, 166, 1983
    - general appearance and vital signs, 1983–1984, 1983t
    - hypokalemia, 356
    - hyponatremia, 348
    - hypoxia, 277
    - jaundice, 325
    - jugular venous pulse, 1983–1984
    - likelihood ratios of, 48t, 49f
    - orthopnea, 1982
    - paroxysmal nocturnal dyspnea, 1982
    - pleural effusion, 2268
    - pulmonary, 1984
    - sleep-disordered breathing, 1997–1998
    - splenomegaly, 480
    - vital signs, 1983
  - comorbidities, 1986, 1986t, 1989
  - costs of, 1976
  - definitions in, 1975–1976, 1976f
  - diagnosis of, **1984,** 1984f
    - biomarkers, 1985
    - chest radiograph, 1985
    - ECG, 1985
    - exercise testing, 1985
    - invasive studies, 1985
    - laboratory tests, 1984–1985
    - in new-onset disease, 1894–1895, 1896f, 1897f, 1902v
    - noninvasive imaging, 1985
    - physical examination in, 3912t
    - POCUS, 3912t, 3914v
  - differential diagnosis of, 1986, 1986t
  - effect on drug concentration and response, 488
  - epidemiology of, 1976, 1976f, 3850–3851
  - folate deficiency in, 788
  - functional classification of, 2093t
  - global considerations in, 1976, 1999–2000, 3850–3851
  - in hemochromatosis, 3336
  - high-output, 1978t, 1982
  - in infective endocarditis, 1038, 1046–1047
  - in ischemic heart disease, 2092, 2103
  - low-output, 1983
  - with mildly reduced ejection fraction, 1977
  - mitral regurgitation in, 1999
  - morbidity and mortality from, 1976, 1977t
  - in older adults, 3883–3884
  - pathophysiology of
    - cardiorenal and abdominal interactions in, 141–142
    - gut congestion, the microbiome, and inflammation in, 1982
    - high-output states in, 1982
    - mechanisms of disease progression, 1980–1981, 1981f
    - ventricular remodeling, 1979–1980, 1980f, 1980t
  - phenotypes and causes of, 1976–1979, 1978f, 1978t, 1979t
  - population attributable risk for, 1977–1978, 1978–1979f
  - precipitating factors, 1983, 1983t
  - with preserved ejection fraction, 137, 1978t, 1987–1989, 1989f, 2193
  - pulmonary hypertension in, 2193
  - with reduced ejection fraction, 1976–1977, 1978t
  - in refeeding syndrome, 2626
  - in rheumatic fever, 2858
  - STEMI and, 2123
  - sudden cardiac death prevention in, 1999, 1999t
  - treatment of, **1987**
    - in acute decompensated disease
      - clinical guiding principles, 1992
      - general principles, 1989–1990, 1990f
      - inotropic therapy, 1992, 1993t
      - vasoactive therapy, 1991, 1993t
      - volume management, 1990–1991, 1993t
    - cardiac resynchronization therapy, 1998–1999
    - cardiac transplantation. *See* Cardiac transplantation
    - cellular and gene-based, 2000
    - comorbidity management, 1997–1998
    - global considerations in, 1999–2000
    - precision medicine in, 34–35
    - with preserved ejection fraction, 1987–1989, 1989f
    - racial/ethnic disparities in, 62
    - with reduced ejection fraction, 1992, 1993f
      - ACE inhibitors and ARBs, 1994
      - anticoagulants and antiplatelets, 1997
      - anticytokine agents, 1997
      - calcium channel antagonists, 1996
      - cardiac contractility modulation, 1998
      - CIED, 1999, 1999t
      - digoxin, 1996
      - diuretics, 1996
      - enhanced external counterpulsation, 1997
      - exercise, 1997
      - fish oil, 1997
      - heart rate modification, 1995–1996
      - micronutrients, 1997
      - mineralocorticoid receptor antagonists, 1994
      - myosin activation, 1996
      - novel neurohormonal antagonists, 1994–1995
      - SGLT-2 inhibitors, 1996
      - soluble guanylyl cyclase stimulation, 1996
      - statins, 1997
      - vagal nerve stimulation, 1998
      - vasodilators, 1994
    - supportive care, 2000
    - surgical, 1999–2000
  - tricuspid regurgitation and, 2060
- Heartland virus, 1657t, 1668
- HeartMate II, 2030
- HeartMate 3, 2030–2031, 2031f
- Heart murmurs
  - in aortic regurgitation, 283, 283f, 289, 1865f, 1866, 2044
  - in aortic stenosis, 283, 284f, 285, 289, 1865, 1865f, 2037
  - approach to the patient, **283,** 1842–1843, 1842f
  - in atrial septal defect, 283, 285–286, 288, 1864
  - auscultation of, 288f, 289–290, 289f, 1865–1866, 1865f
  - benign, 1866
  - clinical context of, 290
  - continuous, 283, 283f, 284t, 288–289, 288f, 1866
  - diastolic, 283, 283f, 284t, 287–288, 287f, 1866
  - duration and character of, 283, 283f
  - dynamic evaluation of, 287f, 289–290, 289t, 1867, 1867t
  - echocardiography in, 290, 1843
  - etiology of, 284t
  - in hypertrophic cardiomyopathy, 1865, 2006
  - in infective endocarditis, 1038, 1038t
  - intensity of, 283, 284f, 1867, 1867t
  - location and radiation of, 283–284, 284f
  - in mitral regurgitation, 283f, 284f, 1865, 1865f, 2053
  - in mitral stenosis, 283, 283f, 287f, 288, 1865f, 1866, 2049
  - in mitral valve prolapse, 284, 285f, 286, 289, 1866, 1866f, 2058
  - in patent ductus arteriosus, 283f, 1866
  - in prosthetic valves, 1867
  - in pulmonic regurgitation, 288, 1866, 2064
  - in pulmonic stenosis, 283f, 284f, 285, 286f, 1866
  - splitting in, 288–289, 289f
  - systolic, 283–287, 284f, 284t, 285f, 287f
  - in tricuspid regurgitation, 2061
  - in tricuspid stenosis, 283f, 1866, 2060
  - in ventricular septal defect, 284f, 285, 1866
- Heart rate
  - deep breathing and, 3541, 3542f
  - intrinsic, 1917, 1918
  - in pregnancy, 3898
  - resting, 1917
  - in shock, 2292t, 2312
  - during sleep, 212
  - Valsalva maneuver and, 3541–3542, 3543f
- HEART score, 108f
- Heart sounds. *See also* Heart murmurs
  - in aortic regurgitation, 2044
  - in aortic stenosis, 2037
  - diastolic, 1865
  - in heart failure, 1984
  - in ischemic heart disease, 2093–2094
  - in mitral regurgitation, 2053
  - in mitral stenosis, 2049
  - in mitral valve prolapse, 2058
  - in myxoma, 288, 2085
  - normal, 1864, 1865f
  - in pericardial disease, 1867
  - in prosthetic valves, 1867
  - in pulmonary hypertension, 2188
  - second, splitting of, 288, 288f, 289f, 1864, 1865f
  - in STEMI, 2115
  - systolic, 1864
  - in tricuspid regurgitation, 2061
  - in tricuspid stenosis, 2059–2060
- Heart transplantation. *See* Cardiac transplantation
- Heartworm, 1825
- Heat conservation, 134
- Heat cramps, 3774
- Heat edema, 3773–3774
- Heat exhaustion, 3774
- Heat-labile enterotoxin, *E. coli,* 1076
- Heat loss, 3768
- Heat production, 134

- Heat-related illness, **3772**
  - climate change and, 3740
  - cooling strategies for, 3775
  - differential diagnosis of, 134, 3773, 3773t
  - heat cramps, 3774
  - heat exhaustion, 3774
  - heatstroke, 134, 3774–3775, 3775t
  - minor syndromes of, 3773–3774
  - prognosis of, 3776
  - renal disease and, 3740, 3742f
  - resuscitation in, 3775–3776
  - risk factors for, 3773, 3773t
- Heat sensitivity, in MS, 3578, 3588
- Heat shock factor-1 (HSF-1), 3992–3993
- Heat shock protein, 2790t
- Heat shock response, 3992–3993, 3992f
- Heat-stable enterotoxin, *E. coli,* 1076
- Heatstroke, 134, 3774–3775, 3775t
- Heat syncope, 3774
- Heat waves, 3746
- Heavy chain diseases, **894**
- Heavy metal pigmentation, 264t, 400
- Heavy metal poisoning, 2356, 2443, **3699**. *See also specific metals*
- Heberden's nodes, 2949f
- Hedgehog pathway, 431, 526t, 573, 606, 670
- Heel
  - bursitis of, 2973
  - plantar fasciitis of, 2975
- Heel-knee-shin maneuver, 3382
- Heel-to-toe gait, 130
- Heerfordt syndrome, 2928
- Heinz bodies
  - after splenectomy, 482
  - in G6PD deficiency, 798
  - laboratory identification of, 437, 441f
  - in splenomegaly, 479
  - unstable hemoglobin variants and, 780
- HELEX septal occluder, A11
- *Helicobacter cinaedi* infection, 1304, 1304t, 1305
- *Helicobacter fennelliae* infection, 1304, 1304t
- *Helicobacter pylori*
  - antibiotic resistance in, 1184, 2528
  - characteristics of, 1298, 2519
  - establishment of infection, 965, 965t, 966
  - evasion of immune system by, 971
  - prevalence of, 2519
- *Helicobacter pylori* infections, **1298**
  - breath odor in, 267
  - clinical features of, 1300, 1300f
  - diagnosis of, 301, 1300–1301, 1301t, 2523, 2524t
  - duodenal ulcers and, 2519
  - epidemiology of, 1299, 2519
  - functional dyspepsia and, 1300
  - gastric cancer and, 645, 1300, 2538–2539
  - gastritis and, 2537–2538, 2539f
  - host factors in, 2519f, 2520
  - immune thrombocytopenic purpura and, 924
  - indigestion and, 300
  - iron deficiency and, 1300
  - MALT lymphoma and, 650, 857, 858t, 865, 1300
  - natural history of, 2520, 2521f
  - pathology and pathogenesis of, 1299–1300
  - pathophysiology of, 2518–2519, 2519–2520f
  - in peptic ulcer disease, **2519**
    - bacterial factors in, 2519–2520f
    - diagnosis of, 2523, 2524
    - epidemiology of, 1300, 2519
    - host factors in, 2519f
    - mechanisms of, 2518–2519, 2519–2520f
    - natural history of, 2520, 2521f
    - pathophysiology of, 1300, 2518–2519
    - recurrent bleeding in, 317
  - in polycythemia vera, 820
  - prevention of, 1303
  - protective role of, 1300
  - treatment of
    - adverse effects of, 1303, 2527–2528
    - antibiotic resistance in, 1303, 2528, 2529f
    - approach to, 1301–1302, 1302f, 2526–2529, 2529f
    - failure in, 1303, 2528
    - first-line regimens, 2527t
    - in gastric lymphoma, 650
    - indications for, 2526–2527
    - investigational approaches to, 2528–2529
    - quadruple therapy, 2528
    - regimens for, 302, 1302–1303, 1302t, 2528
    - reinfection after, 2529
    - salvage regimens, 2528, 2530t
    - test-and-treat approach, 2526
    - triple therapy, 2527–2528
  - upper gastrointestinal tract disease and, 1300f
- Heliotrope rash, 413, 414f, 2911
- Heller myotomy, for achalasia, 2508
- HELLP syndrome, 2139–2140, 2446–2447, 2660, 3899, 3903
- Helminths/helminthic infections, 1736, **1809**
  - cestodes. *See* Cestodes/cestode infections
  - nematodes. *See* Nematodes/nematode infections
  - by organ system and signs/symptoms, 1738–1740t
  - trematodes. *See* Trematodes/trematode infections
- Helper T cells. *See* CD4+ T cells
- Hemagglutinin, filamentous, 1277, 1278
- Hemagglutinin inhibition (HAI) antibodies, 1546
- Hemangioblastoma, 687, 687t, 740t
- Hemangioendothelioma, 729
- Hemangioma
  - bone, 729
  - cardiac, 2086
  - cerebellar, 452
  - cherry, 405, A5
  - of liver, 670, 740
  - of small bowel, 651
- Hemangiopericytoma, 721, 737t, 740
- Hemarthrosis, 473, 929, 930, 2967
- Hematemesis, 274, 298, 316. *See also* Gastrointestinal bleeding
- Hematin, 3347, 3603
- Hematochezia, 316, 320, 2498, 2503f. *See also* Gastrointestinal bleeding
- Hematocrit, 444, 445t, 451, 474
- Hematologic disease
  - in cirrhosis, 2717
  - fatigue in, 166
  - paraneoplastic. *See* Paraneoplastic syndromes, hematologic
  - in pregnancy, 3902
  - in rheumatoid arthritis, 2841
  - in sepsis/septic shock, 2322t, 2323–2324
- Hematoma
  - epidural, 3564, 3572, 3572f, A16
  - in hemophilia, 929, 930
  - muscle, spontaneous, 473
  - spinal epidural, 3564
  - subdural. *See* Subdural hematoma
- Hematomyelia, 3564
- Hematopoiesis, 440f, 443–444, 443f, 479, 785
- Hematopoietic cell transplantation (HCT), **914**
  - allogenic, 561, 914, 915t, 1154
  - autologous, 561, 914, 1154
  - bone marrow collection for, 915
  - categories of, 914–915
  - complications of
    - diffuse alveolar hemorrhage, 275
    - early direct chemoradiotoxicities, 916, 916f
    - graft failure, 917
    - graft-versus-host disease. *See* Graft-versus-host disease (GVHD)
    - hemorrhagic cystitis, 597
    - immune-mediated hemolysis, 908
    - infections, **1154**
      - bacterial, 918, 1155–1156, 1216
      - CMV, 1513t, 1514
      - early, 1156, 1156t
      - fungal, 918, 1155–1156, 1162, 1718
      - HHV-6, 1516
      - late, 1157
      - prophylaxis for, 832–833, 918, 918t, 1164–1165, 1164t
      - *T. gondii,* 1157, 1800, 1803
      - timing of, 916f, 1155t
      - very early, 1155–1156, 1155t
      - viral, 918, 1156–1157, 1160–1161
      - VZV, 1506
    - late direct chemoradiotoxicities, 916–917
    - microangiopathic kidney injury, 2445–2446, 2445t
    - neuropathy, 3607
    - neutropenic fever, 1155
    - thrombotic microangiopathy, 2445–2446
  - cord blood, 915
  - engraftment and immune reconstitution, 916
  - indications for
    - AL amyloidosis, 898
    - ALL, 847–848, 919, 919t
    - AML, 830, 831–832, 919, 919t
    - aplastic anemia, 812, 915, 918, 919t, 3822
    - ATL, 1555f
    - autoimmune and inflammatory diseases, 2784
    - β thalassemia, 778
    - chronic myelomonocytic leukemia, 878
    - chronic neutrophilic leukemia, 876
    - CLL, 856, 919t, 920
    - CML, 841–842, 919–920, 919t
    - DLBCL, 863
    - follicular lymphoma, 865
    - hemoglobinopathies, 918–919
    - hemophagocytic lymphohistiocytosis, 465–466
    - Hodgkin's lymphoma, 870, 919t, 920
    - hyper-IgM syndrome, 2800
    - immunodeficiency disorders, 918
    - juvenile myelomonocytic leukemia, 879
    - mantle cell lymphoma, 866
    - MDS, 817, 919t, 920
    - MS, 3587
    - mucopolysaccharidoses, 3356
    - multiple myeloma, 891–892, 919t, 920
    - non-Hodgkin's lymphoma, 919t, 920
    - nonmalignant diseases, 919
    - osteopetrosis, 3315
    - paroxysmal nocturnal hemoglobinuria, 806
    - peripheral T-cell lymphoma, 867
    - phagocyte disorders, 463
    - primary amyloidosis, 2425
    - primary myelofibrosis, 823
    - pyruvate kinase deficiency, 797
    - SCID, 915, 918, 919t, 2799
    - sickle cell disease, 776, 915, 919
    - systemic sclerosis, 2874
    - testicular cancer, 920
    - thalassemia, 915, 919, 919t
  - jaundice following, 325
  - microbiome and, 3837
  - peripheral blood stem cells for, 915–916
  - preparative regimens for, 915, 1155
  - procedure for, 915–916
  - relapse following, 920
  - survival rates following, 919t

- Hematopoietic growth factor (type 1) receptor, 2773
- Hematopoietic stem cells, 443, **758**
  - defects in, 450
  - developmental biology of, 759
  - differentiation of, 760–761, 760–761f
  - excess capacity of, 759–760
  - features of, in HCT, 914. *See also* Hematopoietic cell transplantation (HCT)
  - functions of, 758–759, 759f
  - in gene therapy and gene editing, 761–762
  - microenvironment of, 759
  - mobility of, 759
  - relationship to cancer, 761
  - self-renewal and clonal dynamics in, 761
  - in tumor angiogenesis, 545f, 546f
- Hematopoietins, 2773
- Hematuria, **342**
  - approach to the patient, 341f
  - benign familial, 2426
  - in bladder cancer, 691
  - in bleeding disorders, 473
  - differential diagnosis of, 2359
  - in glomerular disease, 341, 2414, 2415–2416t, A4
  - gross, 343
  - in hemophilia, 929
  - isolated, 343
  - in lower urinary tract syndromes, 2359
  - microscopic, 343
  - in nephrolithiasis, 2450
  - in parasitic infections, 1739t
  - in polycystic kidney disease, 2432
  - in RCC, 688
  - synpharyngitic, 2355
- Heme, 435, 3341f, 3342t, 3343–3344, 3344f
- Heme albumin, 3347
- Heme arginate, 3347
- Heme oxygenase, 321
- Hemianopia, 222, 223f
- Hemiballismus, 3515
- Hemichorea, A16
- Hemicrania continua, 115, 115t, **3473**
- Hemidesmosomal protein, 2790t
- Hemifacial spasm, 3554, 3555, 3578, V3
- Hemin, 3347
- Hemiparesis, 170, 171f
- Hemiplegia, 189
- Hemipteran bite, 3736
- Hemispatial neglect, 204–205, 205f
- Hemispherectomy, 3425
- Hemizygous, 3787
- Hemobilia, 2737
- Hemoccult test. *See* Fecal occult blood
- Hemochromatosis, **3333**
  - classification of, 3333, 3333t
  - clinical features of, 3335–3336, 3805
    - arthropathy, 2943, 2966, 3336
    - cardiac, 1894, 1927, 2018–2019, 3336
    - cutaneous, 400, 1859, 3336
    - diabetes mellitus, 3336
    - hepatic, 2629, 2712, 3335
    - hypogonadism, 3114, 3336
  - definition of, 3333
  - diagnosis of
    - CMR, 1894, 1903v
    - endomyocardial biopsy, 2019, 2019f
    - laboratory tests, 2630t, 3336, 3336t
    - liver biopsy, 3336–3337, A13
  - early interventions for, 3804t, 3805
  - environmental factors in, 3789
  - in females vs. males, 3789
  - HCC and, 659, A13
  - hereditary
    - clinical features of, 2712
    - diagnosis of, 2712
    - gene mutations in, 3333, 3333t, 3334–3335
    - genetic testing for, 2629, 3804t
    - screening for, 3337f
  - pathophysiology of, 3335, 3335f
  - prevalence of, 3333–3334
  - prognosis of, 3337
  - screening for, 3337, 3337f
  - secondary. *See* Iron overload
  - treatment of, 2019, 3337
  - *V. vulnificus* infection in, 1330
  - *Yersinia* septicemia and, 1349
- Hemodialysis
  - access for
    - arteriovenous grafts and fistulas, 2399–2400, 2403–2404, 2403f
    - catheters, 2402–2403
    - complications of, 2403
    - history of, 2402
    - physical examination in, 2404, 2404v
    - physiology and pathophysiology of, 2402–2404
    - preservation of, 2404
  - for AKI, 2384–2385
  - bullous dermatosis of, 401
  - for CKD, **2398,** 2399f
  - complications of, 2400
  - in diabetes mellitus, 3226
  - for drug- and toxin-induced acidosis, 370–371, S1
  - for Fabry's disease, 3355
  - folate deficiency in, 788
  - goals of, 2400
  - for hypercalcemia, 365, 738, 3284t, 3285
  - for hyperkalemia, 361
  - for hypermagnesemia, 3268
  - oxalosis in, 2962
  - pericarditis in, 2083
  - for poisoning/drug overdose, 3709, 3710t, 3713t
  - for pulmonary edema, 2330
  - for rewarming, 3770, 3770t
- Hemodynamic assessment
  - calculation and normal values, 1905t, 2329t
  - in cardiac catheterization, **1905**
  - in cardiac tamponade, 1906t
  - normal waveforms, 1905f
  - in pericarditis, 1906t
  - in restrictive cardiomyopathy, 1906t
  - in shock, 2329t
  - in STEMI, 2122–2123, 2329t
  - in valvular heart disease, 1905, 1906f
- Hemoglobin (Hb), **769**
  - age-related changes in, 445t
  - developmental biology of, 770
  - disorders of
    - diagnosis of, 771–772
    - hemoglobinopathies. *See* Hemoglobinopathies
    - hereditary persistence of fetal hemoglobin, 770t, 771, 777t
    - thalassemias. *See* Thalassemia syndromes
  - elevated, 451, 452f. *See also* Polycythemia
  - functions of, 770–771
  - globin gene clusters, 770, 770f
  - globin gene switching, 771
  - high O_2 affinity, 779t, 780
  - low O_2 affinity, 779t, 780
  - normal, 444, 445t
  - in pregnancy, 445t
  - reduced, 277, 279
  - structure of, 770
  - testosterone therapy and, 3122
- Hemoglobin A (HbA), 770, 772t
- Hemoglobin A_{1c} (HbA_{1c}), 770, 3196t, 3197, 3207–3208, 3207t, 3209t
- Hemoglobin A_2 (HbA_2), 770, 772t, 777t
- Hemoglobin Bart's hydrops fetalis, 778t, 779
- Hemoglobin-based oxygen carriers (HBOC), 808
- Hemoglobin C, 759, 772t, 774f, 779t
- Hemoglobin E, 759, 779t
- Hemoglobinemia, A6
- Hemoglobin F (HbF), 762, 771. *See also* Hereditary persistence of fetal hemoglobin (HPFH)
- Hemoglobin Hasharon, 780
- Hemoglobin H disease, 759, 778t
- Hemoglobin Köln, 780
- Hemoglobin Lepore, 777t
- Hemoglobinopathies
  - acquired, 780
  - arthropathies associated with, 2967–2968, 2967t
  - HCT for, 918–919
  - with hemoglobins with altered oxygen affinity, 779t, 780
  - with M hemoglobins, 779t, 780
  - in pregnancy, 3902
  - with unstable hemoglobins, 779t, 780
- Hemoglobin-oxygen dissociation curve, 277, 770–771, 771f
- Hemoglobin S, 772, 772t, 773f. *See also* Sickle cell disease
- Hemoglobin SC disease, 772t
- Hemoglobin SE, 772t
- Hemoglobinuria
  - after transfusion, 908t
  - march, 801, 801t
  - paroxysmal cold, 801t, 802
  - paroxysmal nocturnal. *See* Paroxysmal nocturnal hemoglobinuria
- Hemoglobin Zurich, 780
- Hemojuvelin, 3334, 3334f, 3335
- Hemolysis
  - in *Babesia* spp. infections, 1778
  - compensated, 792
  - extravascular, 803, 803f
  - hyperbilirubinemia and, 2638
  - intravascular, 801t, 803, A6
  - transfusion-related, 905, 908, 909f, 910f
- Hemolytic anemia, **791**
  - acquired, 791t, **801,** 801t
    - immune, 801–803, 802t
      - autoimmune hemolytic anemia. *See* Autoimmune hemolytic anemia
      - cold agglutinin disease, 802–803
      - microangiopathic, 323
      - paroxysmal cold hemoglobinuria, 802, 802t
      - paroxysmal nocturnal hemoglobinuria. *See* Paroxysmal nocturnal hemoglobinuria
    - in infections, 801, 801t
    - from mechanical destruction of red cells, 801
    - microangiopathic. *See* Microangiopathic hemolytic anemia
    - from toxic agents and drugs, 802t, 803–804
  - in alcohol-associated liver disease, 2709
  - bone marrow examination in, A6
  - chronic nonspherocytic, 797
  - classification of, 791, 791t
  - clinical features of, 791–792, 791t
  - vs. compensated hemolysis, 792
  - congenital Heinz body, 780
  - definitions in, 791
  - diagnosis of, 437, 440f, 449, 791, 791t
  - differential diagnosis of, 791
  - drug-induced, 489

- Hemolytic anemia (*Cont.*):
  - folate deficiency in, 788
  - inherited, 791t
    - with cation transport disorders, 794t
    - channelopathies, 795–796
    - with enzyme abnormalities, 796, 796t
    - familial hemolytic-uremic syndrome, 800–801
    - G6PD deficiency. *See* Glucose-6-phosphate dehydrogenase (G6PD) deficiency
    - hereditary elliptocytosis, 794f, 794t, 795
    - hereditary spherocytosis. *See* Hereditary spherocytosis
    - pyrimidine-5'-nucleotidase (P5N) deficiency, 796t
    - pyruvate kinase deficiency, 323, 792f, 796–797, 796t, 797f
    - with red cell membrane-cytoskeleton abnormalities, 792–793, 794t
  - pathophysiology of, 791–792, 792f
  - in thalassemia, 777
- Hemolytic streptococcal gangrene. *See* Necrotizing fasciitis
- Hemolytic-uremic syndrome (HUS)
  - after infectious diarrhea, 1079t, 2444
  - AKI in, 341, 2379t, A4
  - atypical (familial), 800–801, 925, 2381, 2412, 2444
  - chemotherapy-related, 595
  - clinical features of, 145t, 2444
    - acute nephritic syndrome, 341
    - encephalopathy, 2350t
    - jaundice, 323
    - rash, 148
  - diagnosis of, 2381
  - *E. coli* infections and, 305–306, 1285t, 1288, 2428, 2444
  - etiology of, 145t, 305–306
  - genetic considerations in, 2428
  - not associated with diarrhea, 925
  - pathogenesis of, 595, 925, 1845, 2427–2428, 2444–2445
  - renal biopsy in, 2428, A4
  - *Shigella* infections and, 1323, 1325, 2428, 2444
  - treatment of, 595, 925, 2428, 2445
- Hemophagocytic lymphohistiocytosis (HLH), 146, 464, 594, 2803, S9
  - primary, **464,** 464t
    - anemia in, 768
    - approach to the patient, 465–466
    - Chédiak-Higashi syndrome, 466. *See also* Chédiak-Higashi syndrome
    - diagnosis of, 465, 882
    - familial, 464–465
    - genetic considerations in, 882
    - Griscelli syndrome type 2, 466
    - other causes of, 466
    - pathophysiology of, 465, 465t
    - treatment of, 465–466, 882
    - X-linked lymphoproliferative disease, 466, 1510, 1512
  - secondary, 464t, **466**
    - clinical features of, 466, 882
    - CNS involvement in, 466–467
    - diagnosis of, 467–468, 882
    - epidemiology of, 466
    - infection-associated, 468, 882
    - laboratory findings in, 466
    - macrophage activation syndrome, 469
    - malignancy-associated, 468–469, 882
    - pathogenesis of, 467, 467f
    - transplant- and CAR-related, 469
    - treatment of, 468–469, 882
    - triggering factors for, 466
- Hemophagocytic syndrome, 866, 867
- Hemophilia, **929**
  - arthropathy in, 2967
  - autosomal, 927. *See also* Von Willebrand disease
  - bleeding in, 472
  - carriers of, 932
  - clinical features of, 929
  - complications of, 930–931
  - diagnosis of, 928t
  - genetic considerations in, 928t, 929
  - late clinical problems in, 931–932
  - pathogenesis of, 929
  - treatment of, 928t, 929–932, 3827–3828
- Hemophilia A, 929, 3789, 3804t, 3823. *See also* Hemophilia
- Hemophilia B, 929, 3804t, 3824t. *See also* Hemophilia
- Hemophilia C (factor XI deficiency), 928t, 932
- Hemoptysis, **274**
  - anatomy and physiology of, 274
  - in cancer, 591–592
  - etiology of, 274–275
  - evaluation of, 275–276, 275f
  - global considerations in, 274–275
  - in lung cancer, 618, 618t
  - massive, 274
  - in mitral stenosis, 2048
  - in pulmonary embolism, 106
  - in respiratory disease, 2199
  - in TB, 1387
  - treatment of, 276
- Hemorrhage
  - anemia of, 807–808
  - epidural spinal, 3455
  - gastrointestinal. *See* Gastrointestinal bleeding
  - hypovolemia in, 347
  - intracerebral. *See* Intracerebral hemorrhage
  - postpartum, 472
  - splinter, 1860
  - subconjunctival, 223
- Hemorrhagic cystitis, 597, 757
- Hemorrhagic fever, viral. *See* Viral hemorrhagic fevers
- Hemorrhagic fever with renal syndrome, 1676–1677
- Hemorrhagic telangiectasia, hereditary. *See* Osler-Weber-Rendu syndrome
- Hemorrhagic thrombocythemia. *See* Essential thrombocytosis (ET)
- Hemorrhoidal cushions, 2584
- Hemorrhoidal disease, 313, 2503f, **2584,** 2584t, 3758
- Hemorrhoidectomy, 2584–2585, 2584t
- Hemosiderin, 385
- Hemostasis, **469**
  - antithrombotic mechanisms in, 470–471, 471f
  - disorders of, 472–473, 472t, 475t. *See also specific disorders*
  - endothelium in, 1845
  - fibrin clot formation in, 470f, 471
  - fibrinolytic system in, 471–472, 471f
  - platelet plug formation in, 470–471
  - tests of, 928, 929f
- Hemosuccus pancreaticus, 318
- Hemothorax, 2270
- Henbane, 3711t
- Henderson-Hasselbalch equation, 365
- Henderson-Patterson bodies, 1519
- Hendra virus, 1539
- Hennekam's syndrome, 2183
- Henoch-Schönlein. *See* IgA vasculitis
- Hensin, 2370
- Hepadnaviridae, 1478f, 1480t. *See also* Hepatitis B virus (HBV)
- Heparan proteoglycans, 470
- Heparan sulfate glycosaminoglycans, 1845
- Heparin, **946**
  - actions of, 946–947, 946f
  - adverse effects of
    - alopecia, 394
    - bleeding, 947–948
    - osteoporosis/osteopenia, 948, 2940t
    - skin necrosis, 421
    - thrombocytopenia. *See* Heparin-induced thrombocytopenia
    - transaminase elevation, 948
  - in cardiac catheterization, 1904
  - coagulation test abnormalities and, 475
  - for DIC, 934
  - dosage of, 947, S13
  - drug interactions of, 562t
  - for DVT/PE, 3902
  - for frostbite, 3772
  - for IC/BPS, 336
  - limitations of, 947, 947t
  - vs. low-molecular-weight heparin, 948t. *See also* Low-molecular-weight heparin (LMWH)
  - management of
    - before endoscopic procedures, 2483t
    - before lumbar puncture, S3
    - treatment monitoring, 947
  - measurement of activity of, 475
  - for NSTE-ACS, 2110t, 2111
  - in PCI, 2127
  - pharmacology of, 946–947, 947t
  - for STEMI, 2121
  - for VTE/PE treatment, 742, 2162, 2163t
  - for VTE prophylaxis, 2164t, 3461
- Heparin-binding proteins, 947
- Heparin-induced thrombocytopenia (HIT)
  - clinical features of, 407, 948
  - delayed onset, 923
  - diagnosis of, 923, 948
  - pathogenesis of, 489
  - time course of, 923, 923f
  - treatment of, 923–924, 948, 948t
- Hepatic adenoma, 670
- Hepatic artery embolization, 682, 684, 3082
- Hepatic encephalopathy. *See* Encephalopathy, hepatic
- Hepatic flukes. *See* Liver flukes
- Hepatic hydrothorax, 281, 330, 2270
- Hepatic lipase deficiency, 3241t, 3245
- Hepatic necrosis, 489, 2665, 2669t. *See also* Hepatitis, fulminant
- Hepatic steatosis
  - drug-associated, 2703t
  - etiology of, 2703t
  - metabolic dysfunction-associated. *See* Metabolic dysfunction–associated steatotic liver disease (MASLD)
- Hepatic vein thrombosis. *See* Budd-Chiari syndrome
- Hepatitis
  - acute viral, **2643**
    - vs. autoimmune hepatitis, 2694
    - clinical features of, 324, 2652t, 2655–2656
    - complications and sequelae of, 2659–2660
    - diagnosis of
      - algorithm for, 2658–2659, 2658t
      - liver biopsy, A13
      - liver function tests, 324, 2637t, 2656
      - serologic tests, 2656–2657
    - differential diagnosis of, 1511t, 2660
    - etiology of. *See specific hepatitis viruses*
    - extrahepatic manifestations of, 2650
    - HIV infection and, 1601
    - in military forces, S7
    - in older adults, 2660
    - pathogenesis of, 2649–2650

    pathology of, 2650–2651
    prognosis of, 2659
    prophylaxis for, 2661–2662
    risk factors for, 2628–2629
    treatment of, 2661
  alcohol-associated, 324, 2660, 2698, 2698t, 2699f, 2709–2710. *See also* Alcohol-associated liver disease
  aplastic anemia and, 810
  autoimmune, **2694**
    cirrhosis due to, 2710
    clinical features of, 324, 2695–2696
    definition of, 2694
    diagnosis of, 2630t, 2673t, 2695, A13
    differential diagnosis of, 2695–2696
    immunopathogenesis of, 2694, 2790t
    prognosis of, 2694
    treatment of, 2696
  cholestatic, 2659
  chronic, 2660, **2673**. *See also* Hepatitis B virus (HBV) infection, chronic; Hepatitis C virus (HCV) infection, chronic
    classification of, 2673, 2673t
    grading of, 2673, 2674t
    laboratory evaluation of, 2673t
    staging of, 2673, 2674t
  coagulation disorders in, 935
  differential diagnosis of, 2660
  drug-induced. *See* Liver disease/failure, drug-induced
  fibrosing cholestatic, 325, 2649, 2724
  fulminant, 2659, 2661
  HSV, 1500
  ischemic, 325
  jaundice in, 325
  neuropathy in, 3607
  physical examination in, 324
  toxic, 2665, 2667t. *See also* Liver disease/failure, drug-induced
  transfusion-associated, 2653–2654
  tumor markers in, 504t
Hepatitis A virus (HAV), 2643, 2643–2644f, 2645t
Hepatitis A virus (HAV) infection
  climate change and, 3752t
  clinical features of, 2643–2644f, 2652t, 2655–2656, S7
  complications and sequelae of, 2659–2660
  diagnosis of, 2630t, 2643, 2643–2644f, 2656, 2658t, S7
  epidemiology of, 2651–2652, 2652t
  etiology of. *See* Hepatitis A virus (HAV)
  global considerations in, 2651–2652
  incidence pre- and postvaccine, U.S., 1002t
  incubation period for, S7
  in military forces, S7
  pathogenesis of, 2650
  postexposure prophylaxis for, 2662
  preexposure prophylaxis for. *See* Hepatitis A virus (HAV) vaccine
  recurrent, after liver transplantation, 2724
  risk factors for, 2628–2629
  serology in, 2656
  transfusion-transmitted, 912
  transmission of, 2651–2652
  treatment of, 2661
Hepatitis A virus (HAV) vaccine
  contraindications to, 1007t
  recommended administration of
    in adults, 1004t, 2662, 2662t
    in children, 2662, 2662t
    in HIV infection, 1594t
    in international travelers, 1013, 1014t, 2662
    in liver disease, 2633, 2662
Hepatitis B core antigen (HBcAg), 2644, 2645t
Hepatitis B e antigen (HBeAg), 2644, 2645t
Hepatitis B immunoglobulin (HBIG), 1146, 1153, 2662–2664
Hepatitis B surface antigen (HBsAg), 2644, 2645t
Hepatitis B virus (HBV)
  extrahepatic sites of, 2647
  genomic structure of, 2643, 2644f
  molecular variants of, 2646–2647
  proteins and particles of, 2644–2645, 2644f, 2645t
  serologic and viral markers of, 2645–2646, 2645f
  structure of, 1480t
Hepatitis B virus (HBV) infection
  acute
    clinical features of, 2652t, 2655–2656, S7
      cholestasis, 325
      erythema nodosum, 406
      polyarteritis nodosa, 2901
      rash, 143t, 148
      serum sickness–like, 1059
      time course of, 2645f
      urticarial vasculitis, 403
    diagnosis of, **2657**
      algorithm for, 2658–2659, 2658t
      laboratory tests in, 2630t
      quantitative nucleic acid test in, S11
      serology in, 2656–2657, 2657t, 2658t, S7
    epidemiology of, 2652–2653, 2652t
    escape mutants in, 2647
    etiology of. *See* Hepatitis B virus (HBV)
    extrahepatic manifestations of, 2650, 2659
    fulminant, 2659
    genotypes in, 2644
    global considerations in, 2652–2653
    HCC and, 660, 660f, 661
    in health care workers, 1625
    immunotolerant vs. immunoreactive, 2649–2650
    incidence of, 1002t
    incubation period for, S7
    in military forces, S7
    pathogenesis of, 2649–2650
    pathology of, 2650–2651
    prevention of, 1625, S7. *See also* Hepatitis B virus (HBV) vaccine
    prognosis of, 2659
    risk factors for, 2628–2629, 2653t
    screening for, 1110, 2653, 2653t, 3903
    transmission of, 2652–2653, 2652t
      by needlestick, 1153, 1153t
      perinatal, 2649–2650, 2653, 3903
      by transfusion, 913t
    treatment of, 2652t, 2661
  in cancer patient, 579
  chronic, **2673**
    after kidney transplantation, 2411
    after liver transplantation, 2683
    clinical features of, 2660, 2675, 2710
    complications and sequelae of, 2659–2660
    diagnosis of, 2673t, 2675, A13
    HBeAg-negative, 2674–2675, 2677
    HBeAg-reactive, 2674
    HDV infection with, 2653, 2660, 2684
    histology of, 2673
    HIV infection and, 1601, 2674, 2683
    infectivity of, 2646
    nonreplicative phase, 2674
    prognosis of, 2673
    renal involvement in, 2416t, 2428–2429
    replicative phase of, 2674
    serology and virologic markers of, 2645–2646, 2645f
    treatment of, 1489–1492, **2675**
      adefovir, 1490t, 1491, 2676t, 2677–2678, 2680f, 2682t
      during cancer treatment, 2683–2684
      combination therapy, 2680
      entecavir, 1490t, 1491, 2676t, 2679, 2680f, 2682t
      HBeAg-negative, 2681t
      HBeAg-reactive, 2681t
      interferons, 1490t, 1491–1492, 2675–2676
      lamivudine, 1490t, 1491, 2676–2677, 2676t, 2680f, 2682t
      liver transplantation, 2719
      novel antivirals and strategies, 2680–2681
      in patients with cirrhosis, 2683
      PEG IFN, 2676t, 2678–2679, 2680f, 2682, 2682t
      recommendations for, 2681–2682t, 2681–2683, 2681t
      telbivudine, 1490t, 1491, 2676t, 2678, 2680f, 2682t
      tenofovir, 1490t, 1491, 2676t, 2679–2680, 2680f
  recurrent, after liver transplantation, 2719, 2724–2725
  in solid organ transplant recipient, 1164t
Hepatitis B virus (HBV) vaccines
  adverse effects of, 2940t
  comparison of, 2662–2663, 2663t
  contraindications to, 1007t
  effectiveness of, 2653, 2663
  liver cancer risk and, 514, 661
  for postexposure prophylaxis, 1146, 2663–2664
  recommended administration of
    in adults, 1004t, 1110, 2662–2663, 2663t
    in health care workers, 1153t
    in hemodialysis patients, 2663t, 2664
    in HIV infection, 1004t, 1594t
    in infants and children, 2662–2663, 2663t
    in international travelers, 1013, 1014t
    in liver disease, 2633
Hepatitis B x antigen (HBxAg), 2644f, 2645
Hepatitis C virus (HCV), 2645t, 2647–2648, 2648f
Hepatitis C virus (HCV) infection
  acute
    clinical features of, 2652t, 2655–2656, S7
    diagnosis of, 2630t, 2648f, 2657–2658, 2658t, S7, S10, S11
    epidemiology of, 2652t, 3856
    extrahepatic manifestations of, 2650
    genetic considerations in, 2650
    global features of, 2654
    incubation period for, S7
    in military forces, S7
    pathogenesis of, 2650
    pathology of, 2650–2651
    prevention of, 2664
    prognosis of, 2659
    risk factors for, 2628–2629, 2655t
    screening for, 39t, 2654–2655, 2655t, 3903, S10
    transfusion-associated, 913t, 2653–2654, 2664
    transmission of, 2652t, 2653–2655
    treatment of, 2652t, 2661, 2692, S7
  chronic, **2684**
    alcohol-associated liver disease and, 3677
    aminotransferase activity in, 2685
    clinical features of, 2685
      arthritis, 1059
      cholestasis, 325
      cirrhosis, 2685, 2710
      cryoglobulinemic vasculitis, 2693, 2905
      depression, 3664
      lichen planus, 388, 392
      osteosclerosis, 3316
      porphyria cutanea tarda, 435
      urticarial vasculitis, 403
    complications and sequelae of, 2660
    deaths from, 2654

- Hepatitis C virus (HCV) infection, chronic (*Cont.*):
    - diagnosis of, 2673t, 2685, A13
    - epidemiology of, 2654–2655, 2684–2685
    - genetic considerations in, 2684–2685
    - HCC and, 660, 660f, 661
    - in hemophiliacs, 931
    - HIV infection and, 1601, 2672
    - inflammasome mutations in, 2761t
    - KIRs with, 2772t
    - liver disease progression in, 2684–2685
    - lymphoid malignancies and, 858, 858t, 865
    - outcomes of, 2685
    - prognosis of, 2685
    - renal involvement in, 2416t, 2428–2429
    - treatment of, **2686**
        - after liver transplantation, 2692–2693, 2719, 2725–2726
        - after relapse or nonresponse, 2692–2693
        - combination therapy, 1492–1493, 1492t, 2688–2691, 2688t
        - drug targets in, 2686f
        - elbasvir/grazoprevir, 1492t, 1493, 2690
        - genotypes and, 2686, 2688t, 2691
        - glecaprevir/pibrentasvir, 1492t, 1493, 2691
        - in HIV infection, 2690, 2693
        - IFN-based, 2686–2687
        - paritaprevir/ritonavir/ombitasvir + dasabuvir, 2689
        - in patients with cirrhosis, 2692–2693
        - in patients with essential mixed cryoglobulinemia, 2693
        - in persons who inject drugs, 2693
        - in pregnancy, 2693
        - protease inhibitors, 2688–2689, 2688t
        - recommendations for, 2688t, 2691–2692
        - in renal failure, 2693
        - ribavirin, 1492t, 1494
        - selection of regimen for, 2693
        - simeprevir, 2689
        - sofosbuvir, 1492–1493, 1492t
        - sofosbuvir/daclatasvir, 1492t, 1493, 2689
        - sofosbuvir/ledipasvir, 1492t, 1493, 2689–2690
        - sofosbuvir/velpatasvir, 1492t, 1493, 2690–2691
        - sofosbuvir/velpatasvir/voxilaprevir, 1492t, 1493, 2691
        - streamlining, 2693
  - kidney transplantation in, 2405
  - recurrent, after liver transplantation, 2719, 2725–2726
  - in solid organ transplant recipient, 1158, 1164t
  - transmission by needlestick, 1153, 1153t
- Hepatitis D virus (HDV), 2645t, 2647
- Hepatitis D virus (HDV) infection
  - chronic, 2660, 2673t, 2684
  - clinical features of, 2652t, 2655–2656
  - complications and sequelae of, 2660
  - diagnosis of, 2630t, 2658–2659
  - epidemiology of, 2652t, 2653
  - etiology of. *See* Hepatitis D virus (HDV)
  - global considerations in, 2653
  - HBV infection with, 2647, 2653, 2660, 2684
  - pathology of, 2651
  - prevention of, 2664
  - prognosis of, 2659
  - transmission of, 2652t, 2653
  - treatment of, 1494, 2661–2663, 2684
- Hepatitis E virus (HEV), 1478f, 1479, 1480t, 2645t, 2649
- Hepatitis E virus (HEV) infection
  - climate change and, 3752t
  - clinical features of, 2652t, 2655–2656, S7
  - diagnosis of, 2630t, S7
  - epidemiology of, 2629, 2652t, 2655, S7
  - etiology of. *See* Hepatitis E virus (HEV)
  - fulminant, 2661
  - global considerations in, 2655
  - incubation period for, S7
  - pathogenesis of, 2650
  - pathology of, 2651
  - prevention of, 2664
  - prognosis of, 2659
  - risk factors for, 2629
  - transfusion-transmitted, 912, 913t
  - transmission of, 2652t, 2655
  - treatment of, 2661
  - in war veterans, S7
- Hepatitis E virus (HEV) vaccine, 2664
- Hepatitis G virus (GB virus C) infection, 1601
- Hepatoblastoma, 670
- Hepatocellular adenoma, 670
- Hepatocellular carcinoma (HCC), **659**
  - clinical features of, 300
  - deaths from, 497t, 499f, 500t, 501t
  - diagnosis of, 661–662, 662f, 2630t, A13
  - epidemiology of, 659, 660f
  - etiology of, 530
  - fibrolamellar, 670
  - global considerations in, 659–660, 660f
  - in hemophilia, 932
  - hepatitis and, 512t, 525, 660, 661, 2660, 2692
  - immunopathology of, 2790t, 2791t
  - incidence of, 497t, 498, 498f, 501t, 659, 660f
  - molecular pathogenesis of, 658f, 660–661, 661t
  - paraneoplastic syndromes in, 737t
  - prevention of, 661
  - risk factors for, 659–660, 3348
  - screening for, 661
  - staging of, 662–664, 663f
  - treatment of, **662**
    - algorithm for, 662–664, 663f
    - liver transplantation, 664–665, 2719
    - locoregional, 665–666
    - summary of key results, 664t
    - surgical, 664
    - systemic, 665f, 666–667
  - tumor markers in, 504t
- Hepatocellular diseases, 2627, 2628t. *See also* Liver disease/failure
- Hepatocyte growth factor/c-Met pathway, 542f
- Hepatocytes, 2627
- Hepatojugular reflux, 1983
- Hepatolenticular degeneration. *See* Wilson's disease
- Hepatomegaly
  - abdominal swelling with, 327
  - in Gaucher disease, 3355
  - in heart failure, 1984
  - in hemochromatosis, 3335
  - in liver disease, 2629
  - in parasitic infections, 1739t
  - in schistosomiasis, 1829
- Hepatopulmonary syndrome, 48t, 270, 2630, 2717
- Hepatorenal syndrome, 330, 2361f, 2374, 2383, 2716
- Hepatosplenic T-cell lymphoma, 860t, 867, 874–875, 2566
- Hepcidin, 3334, 3334f, 3335
  - in anemia of inflammation, 450, 768
  - in iron deficiency, 764, 764f, 765f
  - in iron hemostasis, 762–763, 763f
  - in polycythemia vera, 821
- Hepeviridae, 1478f, 1479, 1480t. *See also* Hepatitis E virus (HEV)
- Hephaestin, 763
- Heptavalent botulinum antitoxin (HBAT), S4
- *HER2* gene mutations
  - in breast cancer, 521
  - in carcinoma of unknown primary, 731
  - as drug target, 534t, 535t, 536
  - in esophageal cancer, 642, 644
  - in gastric adenocarcinoma, 647, 649
  - in lung cancer, 617f, 617t, 627
- HER2/neu, 541, 558, 566, 636, E72t
- Herbal medicines/supplements
  - causing increased bleeding, 473
  - drug interactions of, 489, 3926–3927, 3927t
  - for functional dyspepsia, 302
  - nephrotoxicity of, 2377, 2442
  - regulation of, 3926
  - toxicity of, 2672, 3926
- Herd immunity, 14, 1002
- Hereditary angioedema (HAE), 2808
- Hereditary coproporphyria (HCP)
  - clinical features of, 3341t, 3348, 3602
  - diagnosis of, 3341t, 3343t, 3348
  - genetic considerations in, 3603
  - pathophysiology of, 3341f, 3603
  - treatment of, 3348, 3603
- Hereditary elliptocytosis, 440f, 794f, 795, 795f
- Hereditary endotheliopathy, retinopathy, nephropathy, and stroke (HERNS), 3448
- Hereditary hemochromatosis. *See* Hemochromatosis, hereditary
- Hereditary hemorrhagic telangiectasia. *See* Osler-Weber-Rendu syndrome
- Hereditary hypertension with hyperkalemia. *See* Pseudohypoaldosteronism type II
- Hereditary nephritis, 343
- Hereditary neuralgic amyotrophy (HNA), 3600t, 3601
- Hereditary neuropathy(ies). *See* Peripheral neuropathy, hereditary
- Hereditary neuropathy with liability to pressure palsies (HNPP), 3599t, 3601, 3784
- Hereditary nonpolyposis colon cancer. *See* Lynch syndrome
- Hereditary persistence of fetal hemoglobin (HPFH), 770t, 771, 777t
- Hereditary recurrent fever syndromes
  - clinical features of, 2935t
  - cryopyrin-associated. *See* Cryopyrin-associated periodic syndromes (CAPS)
  - familial Mediterranean fever. *See* Familial Mediterranean fever
  - genetic considerations in, 2935t
  - global considerations in, 2937
  - hyperimmunoglobulinemia D with periodic fever syndrome, 2936
  - treatment of, 2935t
  - tumor necrosis factor-α receptor-associated periodic syndrome. *See* Tumor necrosis factor-α receptor-associated periodic syndrome (TRAPS)
- Hereditary sensory and autonomic neuropathy (HSAN), 3600t, 3601
- Hereditary spastic paraplegia (HSP), 3522–3523t, **3525,** 3569
- Hereditary spherocytosis
  - clinical features of, 793–794
  - diagnosis of, 794
  - epidemiology of, 793
  - folate deficiency in, 788
  - genetic considerations in, 794, 794f, 794t
  - jaundice in, 323
  - peripheral blood smear in, 437, 439f, 447, 795f
  - splenomegaly in, 480
  - treatment of, 795
- Hereditary stomatocytosis, 440f, 794f, 794t
- Hermansky-Pudlak syndrome, 926, 2266, 2553t, 2803, S9
- Hermaphroditism. *See* Ovotesticular disorders of sexual development
- Hernia(s), 330, 2505, 2505f, A12

- Herniation, cerebral, 187–188, 187f
- HERNS (hereditary endotheliopathy, retinopathy, nephropathy, and stroke), 3448
- Heroin, 3648t, 3690. *See also* Opioid-related disorders
- Herpangina, 262t, 1636
- Herpes gestationis (pemphigoid gestationis), 410t, 412, 2790t
- Herpes gladiatorum, 1049, 1499
- Herpes labialis, 262t, 380f, 1049, 1485, 1486t
- Herpes simplex virus (HSV)
  - acyclovir resistance in, 1503–1504
  - genome of, 995
  - host response to, 1496–1497
  - laboratory identification of, 1501–1502
  - reactivation mechanisms of, 1495–1496
  - structure of, 1495
- Herpes simplex virus (HSV) infections, **1495**
  - actinomycosis and, 1363
  - anorectal, 1108, 1498–1499, 2560
  - "asymptomatic," 1497
  - autonomic nervous dysfunction in, 1500
  - in cancer patient, 577
  - diagnosis of, 381, 1501–1502
  - disseminated
    - clinical features of, 143t, 147, 297, 305
    - in immunosuppressed patients, 1496, 1500
    - rash in, 143t
    - treatment of, 1502, 1503t
  - encephalitis
    - in cancer patient, 583, 747
    - clinical features of, 1499, 3514
    - complications of, 1500
    - diagnosis of, 1112, 1113f, 1114t, 1499f
    - differential diagnosis of, 1114
    - genetic considerations in, 1499
    - vs. meningitis, 1121
    - MRI in, 3388f
    - pathogenesis of, 1499
    - treatment of, 1486t, 1499–1500, 1502, 1503t
  - epidemiology of, 1497
  - genital, 378t
    - clinical features of, 1107f, 1108, 1108t, 1498–1499, 1498f
    - diagnosis of, 1108
    - epidemiology of, 1096, 1107, 1497
    - lymphadenopathy in, 477
    - in pregnancy, 3905–3906
    - prophylaxis for, 1485–1487, 1486t, 1502
    - recurrence rates of, 1498
    - skin manifestations of, 1049
    - treatment of, 1054t, 1485–1487, 1486t, 1502, 1503t
  - global considerations in, 1497
  - in HCT recipient, 918, 1155t, 1156
  - herpes gladiatorum, 1049, 1499
  - herpetic whitlow, 1499, 1503t, 1606
  - in HIV infection
    - epidemiology of, 1497
    - perirectal, 1600, 1600f, 1606
    - prophylaxis for, 1595t
    - treatment of, 1502, 1503t, 1606
  - in immunocompromised patients, 1496, 1503t
  - inflammasome mutations in, 2761t
  - keratitis, 224, 1486t, 1487, 1499, 1503t
  - latent, 1495
  - lymphadenitis, 1500, 1503t
  - meningitis, 1125, 1131, 1131t, 1500
  - neonatal, 1500, 1503t, 1504, 3904
  - oral-facial
    - clinical features of, 262t, 1498, 1535, A1
    - course of, 261, 262t
    - pathogenesis of, 1049, 1498
    - treatment of, 1485, 1486t, 1503t, 1537
  - pathogenesis of, 1495
  - peripheral nervous system, 1500
  - pharyngitis, 1498, 1535
  - in pregnancy, 1500, 1501, 1504, 3903–3904
  - prevention of, 1504
  - reactivation of, 1495, 1503t
  - skin manifestations of
    - distribution, 378t, 380f
    - erythema multiforme, 424
    - morphology, 378t
    - vesicles/bullae, 143t, 147, 401, 1049
  - in solid organ transplant recipient, 918t, 1164t
  - in transplant recipient, 2350
  - treatment of, 1502–1504, 1503t
  - visceral, 1500
    - esophagitis, 1500, 1503t, 2495f, 2513
    - liver, 1500
    - pneumonitis, 1500
- Herpesviruses (Orthoherpesviridae), 1478f, 1480t, 1481, 1535–1536, S11. *See also specific viruses*
- Herpes zoster (shingles), **1505**
  - clinical features of, 378t, 1505–1506
    - abdominal pain, 111
    - chest discomfort, 50, 103t, 104
    - glossopharyngeal neuropathy, 3559
    - neuropathy, 3607
    - ocular, 224
    - oral manifestations, 262t, 265
    - postherpetic neuralgia, 265, 1506, 1507, 3607
    - skin manifestations, 1049, 1505–1506, 1507f, A1, A5
    - trigeminal neuropathy, 3553
  - diagnosis of, 1506–1507
  - differential diagnosis of, 1506
  - epidemiology of, 1505
  - in HIV infection, 1506, 1506f, 1605, A5
  - pathogenesis of, 96, 1504
  - prevention of, 892, 1507, 1508t, 1594t
  - in transplant recipient, 1506
  - treatment of, 1054t, 1485–1486, 1486t, 1507
- Herpes zoster encephalitis, 583
- Herpes zoster ophthalmicus, 224, 1506
- Herpes zoster oticus (Ramsay Hunt syndrome), 163, 253, 265, 1506, 3554
- Herpes zoster vaccine
  - contraindications to, 1007t
  - recommended administration of, 40t, 1004t, 1126, 1507, 1594t
  - storage and handling of, 1007
- Herpetic whitlow, 1049, 1499, 1503t, 1606
- Herpetiform lesions, 378t
- Hers disease (liver phosphorylase deficiency, type VI GSD), 3360t, **3362**
- Hertel exophthalmometer, 232
- HES. *See* Hypereosinophilic syndromes (HES)
- *HESX1* gene, 2992
- *HESX1* gene mutations, 3068t, 3112t, 3113
- Heterochromatin, 3935
- *Heterophyes heterophyes,* 961t, 1826t, 1830t, 1831t, S12. *See also* Intestinal flukes
- Heteroplasmy/heteroplasmia, 3791, 3807–3808, 3809, 3810f
- Heterosexual, 3179t
- Heterosexual sexual precocity, 3108
- Heterotopia, A16
- Heterozygous, 3787
- Heuristics, 21–22
- HEV (hepatitis E virus), 1478f, 1479, 1480t, 2645t, 2649. *See also* Hepatitis E virus (HEV) infection
- HEVs (high endothelial venules), 2781
- Hexacarbons, 3610t, 3611
- Hexagonal-phase phospholipids test, 936
- Hexokinase (HK) deficiency, 796t
- Hexosaminidase A deficiency. *See* Tay-Sachs disease
- Hexose-6-phosphate dehydrogenase (H6PDH), 3054
- HF. *See* Heart failure (HF)
- *HFE* gene mutations
  - global considerations in, 3338
  - in hereditary hemochromatosis, 3333, 3335. *See also* Hemochromatosis
  - in liver disease, 659, 2629, 3335, 3338
  - in porphyria cutanea tarda, 435, 3335, 3348
- 2HG (2-hydroxyglutarate), 544, 3978, 3979f
- *hGH-N* gene, 2990
- *hGH-V* gene, 2990
- *HHIP* gene mutations, 2251, 2253
- HHS (hyperglycemic hyperosmolar state), 3216, 3217t, 3218–3219
- HHV-6 infection. *See* Human herpesvirus-6 (HHV-6) infection
- HHV-7 infection. *See* Human herpesvirus-7 (HHV-7) infection
- HHV-8 infection. *See* Human herpesvirus-8 (HHV-8) infection
- 5-HIAA, 682, 682f
- Hiatal hernia, 2505, 2505f
- Hic (pneumococcal H inhibitor), 1187
- Hidradenitis, neutrophilic eccrine, 406, 421
- Hidradenitis suppurativa, 1051, 1377, 2892, 3187
- HIDS (hyperimmunoglobulinemia D with periodic fever syndrome), 155, 2935t, 2936
- HIFs (hypoxia-inducible factors), 276, 444, 541, 545, 2393
- High-altitude cerebral edema (HACE), 277, 3755–3756, 3755f, 3756t
- High-altitude illness, **3754**
  - acute mountain sickness, 3755–3757, 3756t
  - cerebral edema, 277, **3755,** 3755f, 3756t
  - chronic mountain sickness, 278, 3760
  - cough, 3758
  - epidemiology of, 3754
  - gastrointestinal, 3758
  - genetic considerations in, 3755
  - hypoxia in, 277
  - neurologic events, 3758
  - physiology of, 3754–3755
  - polycythemia, 451
  - preexisting medical issues and, 3759–3760
  - psychological/psychiatric problems, 3759
  - pulmonary edema. *See* High-altitude pulmonary edema (HAPE)
  - pulmonary hypertension and, 3760
  - sleep impairment, 3758
  - in travelers, 1019–1020
- High-altitude pulmonary edema (HAPE)
  - clinical features of, 3757, 3757f
  - pathophysiology of, 277, 3757–3758
  - prevention of, 2331, 3758
  - risk factors for, 3757
  - treatment of, 2331, 3756t, 3758
- High-amplitude propagated contractions, 304
- High-density lipoprotein cholesterol (HDL-C)
  - composition of, 3237, 3237f
  - high levels of, 3242t
  - low levels of, 3248–3249
    - in metabolic syndrome, 3255, 3258
    - pathophysiology of, 3248
    - primary (genetic) causes of, 3241t
      - *APOA1* gene mutations, 3248
      - familial LCAT deficiency, 3248
      - primary hypoalphalipoproteinemia, 3248–3249
      - Tangier disease (*ABCA1* deficiency), 1860, 3248, 3602
    - secondary causes of, 3242t, 3249
  - metabolism of, 3239, 3239f
- High endothelial venules (HEVs), 2781
- High-flow nasal oxygen, 510

High-frequency ventilation, 2303t. *See also* Mechanical ventilation
Highly active antiretroviral therapy (HAART). *See* Combination antiretroviral therapy (cART)
High-molecular-weight kininogen deficiency, 928t
High-sensitivity C-reactive protein (hs-CRP), in HIV infection, 1591t
Hilar adenopathy, 477
Hinchey classification, of diverticulitis, 2579, 2580f
HINTS battery, for dizziness, 49, 49t
Hip(s)
  bursitis of, 2944f, 2945, 2973
  fracture of
    bisphosphonates for prevention of, 3303–3306, 3305f
    denosumab for prevention of, 3306–3307, 3306f
    epidemiology of, 3291–3292, 3292f
    lifetime cumulative risk of, 37t
    management of, 3300
    PPIs and, 2525
    risk screening for, 38
  osteoarthritis of, 2951, 2953
  pain in, 2944–2945, 2944f, 2953
Hip replacement, 2956
Hirata's syndrome (insulin autoimmune syndrome), 2790t, 3093
Hirschsprung disease, 397
Hirsutism, **3137**
  after menopause, 3140
  approach to the patient, 3139–3141, 3139f
  drug-induced, 421, 3138
  etiology of, 3138, 3138t
  hormonal evaluation in, 3138–3141
  prevalence of, 2979t
  scoring scale for, 3138, 3139f
  screening/testing for, 2979t
  treatment of, 3141
Hirudin, 3733
His bundle, 1912, 1925
His bundle pacing, 1931, 1931f
Histamine, 94f, 417
Histamine H_1 receptor antagonists. *See* Antihistamines
Histamine H_2 receptor antagonists
  adverse effects of, 2524
  drug interactions of, 489t, 1742t
  for gastritis, 318
  for GERD, 301, 2511
  for peptic ulcer disease, 2524, 2524t
  for premedication in chemotherapy, 598
  for scombroid poisoning, 3726
  for systemic mastocytosis, 2821
  for urticaria, 2808
Histidine metabolism disorders, 3368t
Histidinemia, 3368t
Histidinuria, 3374t
Histiocytic sarcoma/malignant histiocytosis, 881
Histiocytosis X (Langerhans cell histiocytosis), 261, 881–882, 2993, 3003, A12, A16
Histochemical stains, S11
Histoid leprosy, 1408
Histone(s), 538, 3782, 3784f, 3935
Histone acetylase, 538
Histone deacetylase (HDAC)
  in COPD, 2251
  in epigenetics of aging, 3941
  in epigenetics of cancer, 538–539, 538f, 3940
  in high-fat diets, 3939
Histone deacetylase (HDAC) inhibitors
  actions of, 538f, 539, 573, 3942
  adverse effects of, 570t
  epigenetics of, 3940, 3941, 3942
  indications for, 533t, 539, 573, 890t
Histone H2A-H2B-DNA, 2790t
Histone post-translational modifications (hPTMs), 3936–3937
Histopathology, S11
Histoplasma capsulatum, 1692–1693, 1693f
Histoplasmosis, **1693**
  clinical features of, 1688t, 1693–1694
    arthritis, 1058, 1059f
    cavitary, 1694
    chronic meningitis, 1130t
    cutaneous, 406
    hemophagocytic lymphohistiocytosis, 468
    lymphadenopathy, 477, 1694, A12
    mediastinitis, 1694, 1696
    oral, 263t
    subacute meningitis, 1126
  complications of, 1694, 1696
  diagnosis of, 1133, 1688t, 1691, 1694–1695, 1694f, 1695t, A16, S7
  disseminated
    clinical course of, 1694
    in HIV infection, 1594t, 1597, 1610–1611
    treatment of, 1695–1696, 1695t
  epidemiology of, 1130t, 1693
  incubation period for, S7
  pathogenesis of, 1693, 1694f
  of pons, A16
  prevention of, S7
  risk factors for, 1688t, 1694
  in solid organ transplant recipient, 1162, 1164t
  treatment of, 1127, 1695–1696, 1695t, S7
  in war veterans, S7
Histotoxic hypoxia, 277
HIT. *See* Heparin-induced thrombocytopenia (HIT)
HIV, **1556**
  antibodies to, 1587, 1587f
  binding and fusion with target cell, 1557–1558, 1560f
  cellular targets of, 1578
  classification of, 1550t
  genome of, 1552, 1552f, 1558–1559, 1561f
  genotypes of, 981
  geographic distribution of, 1562f
  molecular heterogeneity of, 1559–1561
  phylogeny of, 1557, 1558f
  receptors and co-receptors of, 1577–1578, 1578f
  replication cycle of, 1557–1558, 1560f
  schematic structure of, 1551, 1552f
  structure of, 1557, 1559f
  subtypes of, 1557, 1558f
HIV-1, 985, 1557, 1558f, 1559–1560f, 1559–1561, 1562f. *See also* HIV infection
HIV-2, 1557, 1558f
HIV-associated dementia (HAD), 1606, 1607f, 1607t
HIV-associated neurocognitive disorders (HAND), 1585, 1606, 1607f, 1607t
Hives. *See* Urticaria
HIV infection, **1562**
  acute syndrome, 257, 257t, 457, 1570, 1592, 1592f, 1592t
  arthritis syndromes in, 1059
  asymptomatic stage (clinical latency), 1573, 1592
  cardiovascular disease in, 1598, 2012–2013, 2083
  as chronic disease, 3848–3849
  cobalamin deficiency in, 787
  diagnosis of, 976, **1588,** 1589f, 1590t, S11
  differential diagnosis of, 1511t
  drug desensitization in, 428–429
  drug reactions in, 417, 1603
  drug resistance in, 981
  drug toxicity in, 2672
  elite nonprogressor, 1592
  endocrine and metabolic disorders in, **1602**
    adrenal insufficiency, 359
    hypogonadism, 1603, 3113
    lipodystrophy, 1602, 1602f, 3254
    metabolic syndrome, 1602
    SIAD, 1603
    thyroid disorders, 1603
  epidemiology of
    global, 1096–1097, 1565–1568, 1565–1568f, 3843, 3847–3848, 3856
    in infants and children, 1566
    mortality, 1569, 1569f, 3844
    racial/ethnic disparities in, 1568, 1569f
    by subtypes, 1559–1561, 1562f
    by transmission category, 1568, 1568f
    in U.S., 1568–1569, 1568–1569f, 3847
  etiology of, 1601–1602
  fatigue in, 166
  gastrointestinal disease in, **1598**
    esophagitis, 1598, 1599f, 2512–2513, 2512f
    HIV enteropathy, 1600
    infectious diarrhea, 305, 1598–1600, 1600f
    proctocolitis, 305
  generalized wasting in, 1611
  genetic considerations in
    antiretroviral-associated hypersensitivity, 1585
    CCR5 genotype, 1583–1584, 1583f
    chemokine receptors, 1581t
    chemokines, 1581t
    CYP2B6, 1582t
    CYP2B6 gene, 1582t
    cytokines, 1582t
    gene-gene interactions, 1582t
    HLA-B alleles, 1581t, 1583f, 1584
    innate immunity genes, 1582t
    MHC locus genes, 1581t, 1584–1585, 1584f
    for specific conditions, 1582t, 1585
  genitourinary tract infections in, 1601–1602
    chancroid, 1263
    gonococcal, 1258
    syphilis, 1434, 1601–1602
    vulvovaginal candidiasis, 1602
    warts, 1526
  global considerations in, 2428
  health care workers and, 1564–1565, 1625
  hearing loss in, 247
  hematopoietic system disease in
    anemia, 1604–1605
    bone marrow suppression, 1604, 1604t, 1607
    hemophagocytic lymphohistiocytosis, 468
    lymphadenopathy, 477, 1574, 1604
    MGUS, 1604
    monocyte dysfunction, 461
    neutropenia, 1605
    thrombocytopenia, 922, 924, 1605
    VTE, 1605
  in hemophiliacs, 931
  hepatobiliary disease in, 1600–1601
    cholangiopathy, 326, 2737
    drug-induced, 2672
    hepatitis, 1601, 2660, 2683
    liver transplantation for, 2719
    pancreatitis, 1601, 2750
  herpes zoster in, 1505, 1506f
  hyperkalemia in, 359
  idiopathic CD4+ T lymphocytopenia in, 1614–1615
  immune reconstitution inflammatory syndrome in, 1117, 1391, 1597, 1604, 1604t
  immune response to, 1586–1588, 1586t, 1587f
  immunizations in, 1004t, 1005, 1529, 1594t
  immunologic and rheumatologic disease in, **1603**
  inflammasome mutations in, 2761t

  in injection drug users, 3692
  KIRs with, 2772t
  laboratory monitoring of, **1590**, 1591f, 1591t
  long-term nonprogressor, 1592
  long-term survivor, 1573
  medical community response to, 3855
  microbiome and, 3839
  natural history of, 978f
  neoplastic disease in, 579
    anal cancer, 659, 1526, 1614
    cervical cancer, 1526–1527, 1614
    HPV-related, 1526–1527
    Kaposi's sarcoma. *See* Kaposi's sarcoma
    lymphoid malignancies, 858, 858t, 868
    lymphomas, 1585, 1613–1614, 1613f, 1614f
    multicentric Castleman's disease, 1614
    PCNSL, 720
    primary effusion lymphoma, 873
  nephropathy in, 2357
  neurologic disease in, **1606,** 1606t
    Chagas disease, 1609
    depression, 3664
    HIV-associated dementia, 1606, 1607f, 1607t, 3484
    HIV-associated neurocognitive disorders, 1585, 1606, 1607f, 1607t
    inflammatory demyelinating polyradiculoneuropathy, 3606–3607
    meningitis
      aseptic, 1606
      chronic, 1131t, 1135
      clinical features of, 1125
      cryptococcal, 1606–1607, 1705–1706
      fungal, 1606–1607
      treatment of, 1126
    mononeuropathies, 3607
    myelopathy, 1609
    myopathy, 1610
    neurologic deficits, 1608
    pathogenesis of, 1585–1586, 1606
    peripheral neuropathy, 1609–1610
    PML, 1117, 1609
    polyneuropathy, 1609, 3606
    polyradiculopathy, 3607
    seizures, 1608, 1608t
    sensory neuronopathy/ganglionopathy, 3607
    stroke, 1609
    toxoplasmic encephalitis, 1608, 1608f, 1796f, 1800
  in older adults, 3879
  ophthalmologic disease in, **1610**
    choroid lesions, 1610
    CMV retinitis, 1387, 1514, 1610
    progressive outer retinal necrosis, 1610
  opportunistic/disseminated infections in
    Actinomyces, 1363
    Bartonella, 405, 1355–1356, 1355f, 1610
    Blastomyces, 1702
    CMV, 1514, 1515
    Cryptococcus, 1606–1607, 1705–1706
    Cryptosporidium, 1806
    herpes zoster, 1505, 1506
    Histoplasma, 1597, 1610–1611
    HSV, 1501, 1502, 1503t
    incidence of, 1595, 1595f
    JC virus, 1609
    Leishmania, 1611, 1783, 1785
    malaria, 1611, 1765
    molluscum contagiosum, 1519
    mpox, 1521
    Nocardia, 1361
    nontuberculous mycobacterial, 1415
    P. aeruginosa, 1312
    prophylaxis for, 1417, 1593–1594t
    talaromycosis, 1726t, 1727
    T. cruzi, 1609, 1789
    T. gondii, 1608, 1608f, 1800, 1800f, 1803
    T. marneffei, 1611
    T. whipplei, 1369
  oropharyngeal disease in, 265t, **1598**
    aphthous ulcers, 261, 1598, 1599f
    dysphagia, 295
    hairy leukoplakia, 1010f, 1510, 1598, 1599f
    linear gingival erythema, 261
    parotid enlargement, 266
    periodontitis, 261
    during primary infection, 262t, 1536
    sicca syndrome, 2878t
    thrush, 390, 1598, 1599f
  pathogenesis of, **1569**
    advanced disease, **1573**
    apoptosis in, 1576
    autoimmune phenomenon in, 1576, 2789
    B cell defects in, 1579–1580
    CD4+ T cell defects in, 1578–1579
    CD8+ T cell defects in, 1579
    cellular targets, 1578
    chronic infection
      clinical latency vs. microbiologic latency, 1573
      immune system evasion, 1571–1572
      persistent virus replication, 1571
      reservoirs of HIV-infected cells, 1572–1573, 1572f
      viral dynamics, 1572–1573, 1573f
    clinical course in untreated patient, 1570f
    in CNS. *See* HIV infection, neurologic disease in
    cytokine network in, 1576–1577
    dendritic cell defects in, 1580
    early events in, 1569–1571, 1570–1571f
    immune activation, inflammation and, 1575–1576, 1576t
    in long-term nonprogressors, 1573–1574
    in long-term survivors, 1573–1574
    lymphocyte turnover in, 1577, 1577f
    lymphoid organs and, 1574, 1574f, 1575f
    macrophage defects in, 1580
    monocyte defects in, 1580
    NK cell defects in, 1580
    viral receptors and co-receptors in, 1577–1578, 1578f
  postexposure prophylaxis, 1153, 1625–1626
  in pregnancy, 1565, 3904
  prevention of, 1625–1626, 1626f
  renal disease in
    etiology of, 1601–1602
    glomerulonephritis, 2415t
    kidney transplantation for, 2405
    nephropathy, 1585
    renal biopsy in, A4
    thrombotic microangiopathy, 2446
  respiratory disease in, **1595**
    bacterial pneumonia, 1595–1596
    fungal infections, 1597
    idiopathic interstitial pneumonia, 1597
    MAC infections, 1593t, 1597
    PCP, 1585, 1593t, 1595f, 1596, 1730–1731, 1734, 1734t. *See also Pneumocystis* pneumonia (PCP)
    pulmonary arterial hypertension, 2192
    pulmonary hypertension, 1598
    R. equi infection, 1597
    TB. *See* Tuberculosis (TB), in HIV infection
  screening for, 39t, 1589, 1625
  skin manifestations of, **1605**
    bacillary angiomatosis, 1610
    condyloma acuminatum, 1606
    cutaneous drug reactions, 417, 1606
    folliculitis, 395, 1605
    herpes zoster, 1605, A5
    HSV infections, 1600, 1600f, 1606
    ichthyosis, 1605
    molluscum contagiosum, 1606
    onychomycosis, 389
    porphyria cutanea tarda, 435
    in primary infection, 137, 138t, 145, 148f, 402, A1
    psoriasis, 1605
    seborrheic dermatitis, 385, 1605
    skin cancer, 606
  stage 3. *See* AIDS
  staging, 1557, 1557t
  symptomatic disease, **1592**
  thyroid hormone levels in, 3039
  transmission of
    antiretroviral therapy and, 1562
    in blood and blood products, 913t, 931, 1562t, 1564
    by breast-feeding, 1565
    by human bite, 1562t, 1565
    mother-to-child, 1564, 1617t, 3904
    by needlestick, 1153, 1153t, 1564
    occupational, 1564–1565
    prevention of, 1617t, 1625–1626
    in saliva, 1562t, 1565
    sexual, 1562–1563, 1562t, 1563f
    in solid organ transplants, 1158
    vaginal microbiota and, 3839
  travel considerations for, 1020–1021
  treatment of, **1615**
    combination antiretroviral therapy. *See* Combination antiretroviral therapy (cART)
    experimental, 1625
    HCT, 2784
    initial evaluation, 1615–1616, 1615t
    Internet resources for, 1616t
    milestones in, 978f
    in pregnancy, 3904
  vaccine development for, 1626–1627
  in war veterans, S7
  in women, 3166
H^+,K^+-ATPase, 2518, 2790t
HK (hexokinase) deficiency, 796t
HLA-B8 gene mutations, 3091
HLA-B27
  in axial spondyloarthritis, 2880–2882, 3787
  in psoriatic arthritis, 2887
  in reactive arthritis, 1060, 2890
HLA-B alleles. *See also* Human leukocyte antigen (HLA) complex
  cutaneous drug reactions and, 419, 1585
  drug response and, 493, 494, 2960, 3420, E72t
  in HIV infection, 1581t, 1584, 1585
HLA-B gene mutations, 2823t
HLA-C gene, 1581t, 1583f, 1584, 1585, 1585f
HLA complex. *See* Human leukocyte antigen (HLA) complex
HLADQ2/8, 2545
HLA-DR3, 3066, 3091
HLA-DR4, 2902, 3787
HLA-DR5, 2878t
HLA-DRB1 gene mutations
  autoimmune disease risk and, 2823t
  in autoimmune hepatitis, 2694
  in giant cell arteritis, 2902
  in PD, 3499
  in rheumatoid arthritis, 2842–2843, 2842f
  in systemic sclerosis, 2861
HLA-DR polymorphisms, 3029
HLH (hemophagocytic lymphohistiocytosis), 146, 2803, S9
HMBS gene, 3342t
HMBS gene mutations, 3346
HMB-synthase, 3341t, 3342, 3342t
hMG (human menopausal gonadotropin), 3119
HMG-CoA reductase inhibitors. *See* Statins

- hMPV (human metapneumovirus), 253, 254, 1533–1534, 1533t
- HNA (hereditary neuralgic amyotrophy), 3600t, 3601
- *HNF* gene mutations, 670, 3202–3203, 3783t
- HNPCC (hereditary nonpolyposis colon cancer). *See* Lynch syndrome
- HNPP (hereditary neuropathy with liability to pressure palsies), 3599t, 3601, 3784
- Hoarseness
  - with dysphagia, 295
  - in GERD, 301
  - in laryngeal cancer, 611
  - in lung cancer, 619
- Hockey puck sign, 1265
- Hodgkin's lymphoma, **868**
  - approach to the patient, 869
  - classical, 869
  - classification of, 868, 868t
  - clinical features of, 869
  - CLL and, 853
  - diagnosis of, 868f, 869
  - differential diagnosis of, 869
  - EBV infection and, 868, 1510
  - epidemiology of, 868
  - etiology of, 868–869
  - fever in, 136
  - vs. mediastinal B cell lymphoma, 873
  - mixed cellularity, 868, 869
  - nodular lymphocyte-predominant, 871
  - nodular sclerosing, 868, 873, A6
  - paraneoplastic syndromes in, 740t, 741, 743–744t, 748
  - peripheral blood smear in, A6
  - second malignancies in survivors of, 755, 871
  - skin manifestations of, 405
  - splenomegaly in, 480
  - staging of, 480, 869, 869t
  - treatment of, 869–870
    - in advanced-stage disease, 870
    - brentuximab vedotin, 566
    - dacarbazine, 563
    - in early-stage disease, 870
    - HCT, 919t, 920
    - late consequences of, 871, 3115
    - long-term effects, 752, 755, 871
    - in relapsed disease, 871
  - tumor markers in, 504t
- Hofmann's test, 129
- "Holiday heart," 3677
- Hollenhorst plaques, 225, 226f, 1860
- Holobiont, 3830
- Holter monitoring, 1915
- Holt-Oram syndrome, 50t, 1859, 1860, 3783t
- Homan's sign, 1860
- Home-hospital care, 3
- Homeopathy, 3925t, 3926, 3928
- Homeostasis, 2986
- Homer-Wright rosettes, 721
- Homocitrullinuria, 3369t, 3374t
- Homocysteine, 788–789, 2608
- Homocystinuria(s), **3370**
  - cardiovascular disease in, 784
  - classic, 3370
  - clinical features of, 3369t, 3370–3371
  - enzyme defects in, 3369t
  - folate deficiency in, 788
  - with methylmalonic acidemia, 3369t
  - pathways, enzymes, and coenzymes involved in, 3370–3371, 3371f
  - treatment of, 3371
- Homogeneous staining region, 521
- Homogentisic acid oxidase, 3371
- Homoplasmy, 3808–3809, 3814
- Homozygous, 3786f, 3787
- Honeybee sting/venom, 3735–3736
- Honey, in infant botulism, 1234, 1237
- Hookworm infections
  - clinical features of, 1738t, 1815t, 1816
  - diagnosis of, 961t, 1816, S12
  - eosinophilia in, 2234, S12
  - epidemiology of, 961t, 1736, 1738t, 1816, S12
  - life-cycle hosts of, 1816, S12
  - skin manifestations of, 1051
  - treatment of, 1815t, 1816
- Hoover's sign, 121t
- Hordeolum (sty), 224
- Horizontal gaze disorders, 235–236
- Horizontal gene transfer, 976t
- Hormesis, 3875
- Hormiga veinticuatro, 3736
- Hormone(s), **2980**. *See also specific hormones and types*
  - autocrine control, 2987
  - classes of, 4–5
  - deficiency of, 2978, 2978t
  - degradation of, 2982–2983
  - ectopic production of, 736–737, 737t
  - eutopic production of, 736
  - excess, 2977–2978, 2978t
  - feedback regulatory systems, 2986–2987, 2986f
  - functions of, 2980, 2986
  - half-life of, 2982
  - measurements of, 2980
  - paracrine control, 2987
  - resistance to, 2978, 2978t
  - role of, 2977
  - secretion of, 2982–2983
  - serum-binding proteins and, 2982–2983
  - stimulation tests, 2980
  - suppression tests, 2980
  - synthesis and processing of, 2982
  - transport of, 2982–2983
  - unbound, 2982–2983
- Hormone receptor(s)
  - characteristics of, 2981
  - families of, 2981, 2981t
  - membrane receptors, 7–8, 2983f
  - nuclear receptors, 2983, 2985f
- Hormone receptor–directed therapy, 536–537, 566–567. *See also specific drugs*
- Hormone replacement therapy
  - for hypopituitarism, 2994, 2996t
  - postmenopausal. *See* Postmenopausal hormone therapy
- Hormone resistance syndromes, 2978
- Horner's syndrome
  - in cervical cord lesions, 3560
  - in lung cancer, 619, 626
  - ocular manifestations of, 221, 232, 234
- Horse botfly, 3732
- Hospice care, 76, 88, 576. *See also* End-of-life care
- Hospital-acquired infections. *See* Health care–associated infections
- *Hospital Compare* (website), 45
- Hospitalists, 3
- Hospital(s), patient care in, 3
- Hospital Quality Initiative, 45
- "Hot cross buns sign," 3548, 3549f
- Hot flashes, 507t, 510, 638
- "Hot potato" voice, 259
- Hotspots, mutational, 522
- Hot-tub folliculitis, 142t, 147, 395, 1051, 1312, A1
- Hot-tub lung, 1415, 2228t, 2229
- Household air pollution, 2241
- House mouse mites, 3730
- Howell-Jolly bodies
  - in asplenia, 2796
  - laboratory identification of, 437, 440f, 445, 446f
  - post-splenectomy, 482
  - in splenomegaly, 479
- Howship's lacunae, 3260, 3277
- Hoyeraal-Hreidarsson syndrome. *See* Dyskeratosis congenita
- HPA (hypothalamic-pituitary-adrenal) axis, 2987, 2991, 3050–3051, 3052f
- HPeVs (human parechoviruses), 1534, 1534t
- HPFH (hereditary persistence of fetal hemoglobin), 770t, 771, 777t
- HPT-JT (hyperparathyroidism-jaw tumor) syndrome, 3080t, 3087, 3087t, 3275, 3276
- HPV. *See* Human papillomavirus (HPV)
- HPV infections. *See* Human papillomavirus (HPV) infections
- *HRAS* gene mutations, 614
- H RBC group system, 904t
- *HRPT2* gene mutations, 3276
- *HS6ST1* gene mutations, 3111t, 3112
- HSAN (hereditary sensory and autonomic neuropathy), 3600t, 3601
- hs-CRP (high-sensitivity C-reactive protein), 1591t
- *HSD3β-2* gene mutations, 3098, 3099t, 3101, 3101t
- *HSD17β-3* gene mutations, 3099, 3099t
- HSF-1 (heat shock factor-1), 3992–3993
- HSP (hereditary spastic paraplegia), 3522–3523t, **3525,** 3569
- HSV. *See* Herpes simplex virus (HSV)
- HSV infections. *See* Herpes simplex virus (HSV) infections
- 5-HT. *See* Serotonin
- hTERT, 531
- HTLV-1. *See* Human T-cell lymphotropic virus (HTLV-1)
- HTLV-2 (human T-cell lymphotropic virus-2) infections, 1556
- Human African trypanosomiasis (sleeping sickness), **1792**
  - approach to the patient, 1794
  - epidemiology of, S12
  - global considerations in, 1796
  - pathology and pathogenesis of, 1793–1794
  - prognosis of, 1796
  - skin manifestations of, A1
  - *T.b. gambiense*
    - clinical features of, 140t, 1737, 1740t, 1794, S7
    - diagnosis of, 1795, S7, S12
    - epidemiology of, 1792–1793, 1793f, S12
    - treatment of, 1747, 1795, 1796f, 1796t, S7
  - *T.b. rhodesiense*
    - clinical features of, 140t, 1737, 1738t, 1740t, 1794, S7
    - diagnosis of, 1795, 1795f, S7, S12
    - epidemiology of, 1792–1793, 1793f, S12
    - treatment of, 1747, 1796, 1796t, S7
- Human bite-wounds, 1052, **1143,** 1145t
- Human bocaviruses, 253, 1525, 1536
- Human chorionic gonadotropin (hCG)
  - for androgenic-anabolic steroid use disorder, 3124
  - in carcinoma of unknown primary, 732, 733t
  - ectopic production of, 736, 737t, 740
  - in germ cell tumors, 712
  - in gestational trophoblastic disease, 715
  - in testicular cancer, 705, 708t
  - as tumor marker, 504t, 705
- Human chorionic gonadotropin (hCG) stimulation test, 3108
- Human chorionic gonadotropin (hCG) therapy, 3119, 3174
- Human evolution, 3809
- Human genome, **3777**. *See also* Chromosome(s); Gene(s)
  - databases/websites, 3778t
  - replication of, 3779–3780
  - structure of, 3777–3780, 3779f, 3780f

Human Genome Project, 3777–3778, 3798
Human granulocytic anaplasmosis, 1452, 1455t, 1460f, 1461–1462
Human growth hormone. *See* Growth hormone (GH)
Human herpesvirus-6 (HHV-6), 1516
Human herpesvirus-6 (HHV-6) infection, **1516**
  clinical features of, 1516
  diagnosis of, 1156
  differential diagnosis of, 1511t
  diseases associated with, 1516
  encephalitis, 583
  epidemiology of, 1516
  exanthem subitum, 138t, 145, 1516, A1
  in HCT recipient, 1155t, 1156, 1516
  myocarditis in, 2012
  reactivation in DIHS, 424
  in transplant recipient, 2350
  treatment of, 1156, 1516
Human herpesvirus-7 (HHV-7), 1516
Human herpesvirus-7 (HHV-7) infection, **1516**
  clinical features of, 1516
  epidemiology of, 1516
  in solid organ transplant recipient, 1155t, 1156
  treatment of, 1516
Human herpesvirus-8 (HHV-8) infection, **1516**
  clinical features of, 1517
  epidemiology of, 1516, 1586
  HIV infection and, 1517
  Kaposi's sarcoma and, 727, 1516, 1586
  lymphoid malignancies and, 857, 858t, 1516
  primary effusion lymphoma and, 873, 1516
  treatment of, 1517
Human immunodeficiency virus. *See* HIV
Human leukocyte antigen (HLA) complex
  disease associations with, 3091t, 3787
  HLA class I alleles, 1581t
  HLA class I genes, 1584, 2799, 2842–2843
  HLA class II alleles, 1581t
  HLA class II genes, 2799
  in transplantation, 914, 2406, 2407, 2407f
Human leukocyte surface antigens, 2757–2758t
Human menopausal gonadotropin (hMG), 3119
Human metapneumovirus (hMPV), 143t, 253, 254, 1533–1534, 1533t
Human microbiome, **3830**. *See also* Gut microbiota
  atopic diseases and, 3838
  autoimmune diseases and, 3837–3838
  bacteria in, 1371–1372, 3832, 3832f
  cancer and, 3836–3837
  cardiovascular disease and, 3836
  circadian rhythms and, 3834, 3952
  definition of, 3777
  diet and, 3840–3841
  disease states and, 3834–3835, 3835f
  drugs and, 3834
  genetics and, 3833
  in health vs. disease, 3833f
  historical perspective on, 3829–3830
  in immune system regulation, 3840
  infectious diseases and, 957–958, 3834, 3835f, 3838–3839
  influences on, 3832–3834, 3833f
  lifestyle and, 3834
  in malnutrition, 3835–3836
  mechanisms of effects of, 3839–3840
  neurologic disease and, 3398–3399
  overview of, 3830–3831, 3832f
  therapeutic applications of, 3841
  vaccine response and, 3839
Human Microbiome Project (HMP), 3830–3831
Human monocytic ehrlichiosis, 1121
Human monocytotropic ehrlichiosis, 139t, 1455t, 1460–1461
Human papillomavirus (HPV), 713, 1480t, 1525
Human papillomavirus (HPV) infections, **1525**
  anal cancer and, 659, 1527
  cervical cancer and
    mechanisms of, 525
    risk factors for, 217, 390
    screening for, 1529–1530
    types associated with, 713, 1526
  clinical features of, 1527, 1527f
  counseling for, 1531
  diagnosis of, S11
  epidemiology of, 1526
  genetic considerations in, 713
  HIV infection and, 1526–1527, 1614
  immune response to, 1526
  incidence of, 1096
  inflammasome mutations in, 2761t
  oral cancer risk and, 261
  oropharyngeal cancer and, 512, 610, 613, 1526
  pathogenesis of, 1525–1526
  prevention of, 1527–1528. *See also* Human papillomavirus (HPV) vaccines
  screening for, 1529–1530
  treatment of, 1489, 1530–1531, 1531t
  warts, 264t, 1527f. *See also* Condyloma acuminata
Human papillomavirus (HPV) vaccines
  for anogenital wart prevention, 390
  bivalent, 1528
  for cervical cancer prevention, 390, 514, 713
  contraindications to, 1006t
  cross-protection from, 1529
  effectiveness of, 514
  nine-valent, 1528–1529
  precautions for, 1006t
  quadrivalent, 1528
  recommended administration of
    in adults, 40t, 1004t, 1110, 1529
    in children and adolescents, 514, 713, 1110, 1529
    in people with HIV, 1529, 1594t
    single-dose, 1529
  shared decision-making for, 1529
  targeting of, 984
  virus-like particles in, 1528
Human parechoviruses (HPeVs), 1534, 1534t
Human stem cell leukemic-lymphoma syndrome (8p11 myeloproliferative syndrome), 880
Human T-cell lymphotropic virus (HTLV-1)
  carcinogenicity of, 512t, 844
  classification of, 1550t
  discovery of, 1553
  genomic structure of, 1552, 1552f
  molecular biology of, 1553
  schematic structure of, 1552f
Human T-cell lymphotropic virus-1 (HTLV-1) infections
  adult T cell leukemia/lymphoma. *See* Adult T cell leukemia/lymphoma
  arthritis syndromes in, 1059
  clinical features of, S7
  epidemiology of, 1553–1554, S7
  global considerations in, 1553, 1554f
  inflammasome mutations in, 2761t
  lymphoid malignancies and, 857, 858t
  myelopathy, 1553–1554, 3525, 3567
  prevention of, 1556
  transfusion-transmitted, 913t
  treatment of, 1555–1556, 1555f
  in war veterans, S7
Human T-cell lymphotropic virus-2 (HTLV-2) infections, 1556
Human tetanus immunoglobulin (HTIG), 1146, 1231
Humidifier fever, 2228t
"Hummingbird sign," 3497
Humoral hypercalcemia of malignancy. *See* Hypercalcemia, malignancy-associated
Hunner lesion, 330, 336
Hunter disease, 3352t, 3356
Hunt-Hess scale, for subarachnoid hemorrhage, 3459t
Huntingtin, 3405, 3513
Huntingtin gene mutations, 3405, 3513
Huntington's disease, **3512**
  clinical features of, 178, 208, 235, 241, 3483, 3512–3513
  etiology of, 3513
  functional anatomy of, 193
  genetic considerations in, 3405, 3513, 3786, 3789, 3793, 3793t, 3941
  genetic testing for, 3803
  imaging in, 3513f, A16
  pathophysiology of, 3395, 3400, 3405, 3994
  prevalence of, 3512
  treatment of, 3513–3514
Huntington's disease–like disorders, 3514–3515
Hu protein(s), 2791t
Hu protein antibodies
  in limbic encephalitis, 746f
  neurologic syndromes associated with, 743t, 745f, 746
  in paraneoplastic autonomic neuropathy, 751
  in paraneoplastic sensory neuronopathy, 750
  in small-cell lung cancer, 619
Hurler-Scheie syndrome, 3352t, 3356
Hurler's disease, 3352t, 3356, A15
Hurricanes, 3747
HUS. *See* Hemolytic-uremic syndrome (HUS)
Hutchinson-Gilford progeria syndrome, 3787, 3788t, 3870, 3941
Hutchinson's sign, 224, 599f, 1506
Hutchinson's teeth (incisors), 261, 1433
HV interval, 1927
Hyaline casts, A4
Hyalomma ticks, 1677
Hyaluronan, 2844
Hyaluronate lyase, 1188f
Hyaluronic acid, 2949, 2955, 2955t
Hybrid capture, S11
Hybridization, for pathogen identification, 979
Hydatid disease. *See* Echinococcosis (hydatid disease)
Hydatidiform mole, 715, 3791–3792. *See also* Gestational trophoblastic disease
Hydralazine
  adverse effects of, 2145t, 2940t
    edema, 282t
    hepatotoxicity, 2667
    lupus-like syndrome, 489, 493, 2835
    vasculitis, 426, 2907
    vitamin B_6 deficiency, 3612
  genetic variations in response to, 493, E72t
  for heart failure, 1994, 1995t
  for hypertension, 2145t
  for hypertension in pregnancy, 2140
  metabolism of, 485t
  for preeclampsia, 3899
Hydroa vacciniforme, 434
Hydrocele, 1820, 1820f
Hydrocephalus
  adipsic diabetes insipidus in, 352
  after subarachnoid hemorrhage, 3459, 3461
  coma in, 192
  communicating, 197, 198f
  in ependymoma, A16
  gait disorders in, 178
  in intracranial meningitis, 1073, 1128
  in neurocysticercosis, 1833–1834
  normal-pressure, 197, 198f, 3483
Hydrochloric acid therapy, for metabolic alkalosis, 375

Hydrochlorothiazide, 923t, 2144t, 2835
Hydrocodone, 80f, 83t, 274
Hydrocortisone
  for ACTH deficiency, 2997
  for adrenal insufficiency, 3069
  for adult hypopituitarism, 2996t
  for anaphylaxis, 3719
  for CAH, 3072, 3101
  for candidiasis, 390
  for hypercalcemia, 365
  for IBD, 2565–2566
  for myxedema coma, 3032
  for otitis externa, 254
  for prevention of transfusion reaction, 908
  for thyroid storm, 3037
Hydrodistention, 334, 336
Hydrogen peroxide (H_2O_2), in endothelium, 1845
Hydrogen sulfide exposure/poisoning, 2239t, 3713t
Hydroids, 3725
Hydromorphone, 80f, 83t, 97t, 98
Hydronephrosis, 693, 2454, 2455, 3914v
Hydrophobia, 1115, 1652, 1653f
Hydrops
  endolymphatic. *See* Ménière's disease
  fetalis, 1524, 1524t
  gallbladder, 2733
  Hemoglobin Bart's, 778t, 779
Hydrothorax, 281, 330, 2270
Hydroureter, 2454
Hydroxocobalamin (vitamin B_{12A}), 789, 3713t, S5
β-Hydroxybutyrate, 369, 370, 3217, 3235
4-Hydroxybutyric aciduria, 3368t
Hydroxychloroquine
  adverse effects of
    cardiotoxicity, 2017
    common, 2850t
    cutaneous, 399
    myopathy, 2940t, 3648t
    neuropathy, 3608, 3609t
    ocular, 232, 1464
    serious, 2850t
  for antiphospholipid antibody syndrome, 2839
  for arthritis in Sjögren's disease, 2879
  for calcium pyrophosphate deposition disease, 2961
  drug interactions of, 1489
  for hypercalcemia, 365
  for infective endocarditis, 1044t
  for malaria prophylaxis, 1774t
  monitoring during treatment with, 2850t
  overdosage/poisoning with, 3715t
  for porphyria cutanea tarda, 435, 3348
  for Q fever, 1464
  for retinitis pigmentosa, 232
  for rheumatoid arthritis, 2849, 2850t, 2852
  for sarcoidosis, 2929t, 2930
  for SLE, 2833–2834
  for urticaria, 2809
  for Whipple's disease, 1370
2-Hydroxyglutarate (2HG), 544, 828
Hydroxykynureninuria, 3368t
11β-Hydroxylase deficiency, 355, 3054, 3071t, 3101
17α-Hydroxylase (CYP17), 2790t
17α-Hydroxylase deficiency
  clinical features of, 355, 3098–3099
  diagnostic markers of, 3071t
  genetic considerations in, 3099
  hypergonadotropic hypogonadism in, 3134–3135
17β-Hydroxylase deficiency, 3099
21-Hydroxylase (CYP21), 2790t
21-Hydroxylase antibodies, 3090
21-Hydroxylase deficiency
  classic. *See* Congenital adrenal hyperplasia (CAH)
  nonclassic, 3071, 3140–3141
21-Hydroxylase gene *(CYP21A2)* mutations, 3071, 3071t, 3101, 3101t, 3785
3-Hydroxy-3-methylglutaric aciduria, 3369t
17-Hydroxyprogesterone (17OHP), 3051f, 3071t
3β-Hydroxysteroid dehydrogenase deficiency, 3071t, 3098
11β-Hydroxysteroid dehydrogenase, 355, 739, 3054, S1
11β-Hydroxysteroid dehydrogenase deficiency. *See* Syndrome of apparent mineralocorticoid excess (SAME)
17β-Hydroxysteroid dehydrogenase deficiency, 3099
5-Hydroxytryptophan, 3515
Hydroxyurea
  actions of, 565, 775
  adverse effects of, 420, 562t, 790
  for chronic myelomonocytic leukemia, 878
  for chronic neutrophilic leukemia, 876
  for CML, 842
  for essential thrombocytosis, 824
  for hypereosinophilic syndromes, 462, 880
  for hyperleukocytosis, 591
  interactions and issues, 562t
  for polycythemia vera, 821
  for sickle cell anemia, 775
Hydroxyzine, 335, 383, 422, 2809
Hygiene hypothesis, 2826, 3834
Hymen, imperforate, 3133
Hymenolepiasis diminuta, 1838, S12
Hymenolepiasis nana. See Rodentolepis nana (dwarf tapeworm)
Hymenoptera venom/sting, 2812, 2814, 2821, 3735–3736
Hyoscyamine, 1094, 3711t
Hypalgesia, 172
Hyperabduction maneuver, 2176
Hyperacute rejection, 2406, 2407
Hyperaldosteronism
  diagnosis of, 356
  familial type I. *See* Glucocorticoid-remediable aldosteronism (GRA)
  familial types II and III, 355
  genetic considerations in, 355, 356
  hyperreninemic, secondary, 373–374
  hypertension in, 2138t, 2139
  hypokalemia in, 355
  idiopathic, 355
  metabolic acidosis in, 372
  primary, 355, 356, 3060t, 3647
  secondary, 355
Hyperalgesia, 93, 100, 172
Hyperammonemia, 2636, 3367, 3369t, 3371, 3374t
Hyperamylasemia, 2740, 2741t
Hyperamylasuria, 2741t
Hyperargininemia, 3374t
Hyperbaric medicine, 3760–3761
Hyperbaric oxygen therapy, **3761**
  adverse effects of, 3764–3765
  for altitude illness, 3757f
  for carbon monoxide poisoning, 3764
  challenges and controversies in, 3765
  contraindications to, 3765
  equipment for, 3761, 3761f
  exploratory indications for, 3764
  for gas gangrene, 1054, 1243
  for hemorrhagic cystitis, 597
  for high-altitude illness, 3756–3757, 3756tl, 3758
  indications for, 3762–3764, 3762t, 3763f
  for late radiation tissue injury, 3762–3763
  mechanisms of, 3761, 3762f
  for problem wounds, 3763–3764, 3763f
Hyperbilirubinemia, **2638**
  in alcohol-associated liver disease, 2698
  approach to the patient, **322,** 322f
  conjugated
    in benign recurrent intrahepatic cholestasis, 325, 2641t, 2642
    differential diagnosis of, 323
    in Dubin-Johnson syndrome, 323, 2641–2642, 2641t
    measurement of, 321
    with other liver abnormalities
      approach to the patient, 322f, 323–324
      in cholestatic disorders, 324–326, 325t
      in hepatocellular disease, 324, 324t
    in progressive familial intrahepatic cholestasis, 325, 2641t, 2642
    in Rotor syndrome, 323, 2641t, 2642
  isolated, 322–323, 323t
  in pancreatitis, 2745
  pathophysiology of, 322–323
  unconjugated
    from decreased hepatic bilirubin clearance, 2639
    differential diagnosis of, 322–323, 323t
    from hereditary defects. *See* Crigler-Najjar syndrome; Gilbert's syndrome
    from increased bilirubin production, 2638
    measurement of, 321
Hypercalcemia, **362, 3274**
  after kidney transplantation, 2411
  aluminum intoxication and, 3282
  asymptomatic, 3283
  in ATL, 1554
  bone turnover–associated, 3281–3282
  chronic, 364, 2443, 3283, 3283f
  in CKD, 2390, 3282
  clinical features of, 166, 363–364, 2140, 2614t, 3275
  diagnosis of, 364, 3274
  differential diagnosis of, 3274–3275, 3283–3284, 3283f
  ECG in, 364, 1874, 1875f, A7
  etiology of, 362–363, 362t, 3275, 3275t, 3283
  familial hypocalciuric, 363, 3084, 3279–3280, 3804t
  genetic considerations in, 3279–3280, 3280f
  in hyperparathyroidism. *See* Hyperparathyroidism, primary
  hyperthyroidism and, 3281–3282
  immobilization and, 3282
  of infancy, 3281
  lithium therapy and, 3279
  malignancy-associated, **737,** 3280–3281
    clinical features of, 737–738
    clinical syndromes related to, 3280–3281
    diagnosis of, 364, 738, 3278f
    etiology of, 737, 737t
    in lung cancer, 619
    mechanisms of, 362–363, 3280–3281
    in metastatic bone disease, 730
    in multiple myeloma, 364, 887, 892
    pathogenesis of, 737t
    treatment of, 738, 3281
  in milk-alkali syndrome, 363, 3262, 3282–3283
  in nephrogenic diabetes insipidus, 352
  renal effects of, 2443
  in sarcoidosis, 2927–2928, 3281
  thiazides and, 3282
  treatment of, 364–365, 3284–3285, 3284t
  vitamin A intoxication and, 3282
  vitamin D–related, 3281
Hypercalciuria
  in Bartter's syndrome, 350, 3287
  hematuria and, 343
  hypocalcemic, 3287
  loop diuretics and, 355
  nephrolithiasis risk and, 2449

- Hypercapnia
  - adverse effects of, 2307, 2307t
  - coma in, 188
  - in COPD, 2276
  - etiology of, 367
  - in mechanical ventilation, 2307
  - metabolic alkalosis following, 374
  - permissive, 375
  - in respiratory acidosis, 375
- Hypercholesterolemia (elevated LDL-C)
  - approach to the patient, 3249
  - in metabolic syndrome, 3255, 3257–3258
  - multifactorial, 3245
  - pathophysiology of, 3242
  - primary (genetic) causes of
    - autosomal dominant, 3241t
    - autosomal recessive, 3241t, 3243
    - familial, 494, 2968, 3241t, 3242–3243, 3249, A15
    - lysosomal acid lipase deficiency, 3241t, 3244–3245
    - sitosterolemia, 3241t, 3243–3244
  - secondary causes of, 3242t, 3246–3247
  - treatment of, 3250–3252, 3257–3258
- Hypercortisolism, 3011. *See also* Cushing's syndrome
- Hyper-CVAD regimen, 846f, 847
- Hyperekplexia-3, 3374t
- Hyperemesis gravidarum, 298, 3023, 3903
- Hypereosinophilia, 878t, **879**
  - chronic eosinophilic leukemia, 876, 878t
  - clinical features of, 880, 2014, 2233
  - diagnosis of, 878t
  - diseases associated with, 462
  - idiopathic (hypereosinophilic syndrome), 462, 878t, 879, 880
  - lymphocytic variant, 878t, 879
  - pathophysiology of, 462, 2014
  - subtypes of, 879, 2233
  - treatment of, 880, 2233
  - with tyrosine kinase gene fusions, 878t, 879–880
- Hypereosinophilic syndromes (HES), 2770, 3605
- Hyperesthesia, 172
- Hypergammaglobulinemia, 2695
- Hypergastrinemia, 2534
- Hyperglobulinemia, S1
- Hyperglycemia
  - acute disorders related to
    - diabetic ketoacidosis. *See* Diabetic ketoacidosis (DKA)
    - hyperglycemic hyperosmolar state, 3216, 3217t, 3218–3219
  - β-agonists and, 361
  - chorea in, 3514, A16
  - clinical features of, 3203
  - hypernatremia in, 352
  - hyponatremia in, 350
  - in overdosage/poisoning, 3705
  - in pancreatic cancer, 673
  - in pancreatitis, 2745
  - in parental nutrition, 2624–2625
  - perioperative, 3909
  - spectrum of, 3196t
- Hypergonadotropic hypogonadism, 3134–3135
- Hyperhomocysteinemia, 941, 2609, 3371
- Hyperhydroxyprolinemia, 3368t
- Hyper-IgD syndrome, 136
- Hyper-IgE (Job) syndrome, **462**
  - autosomal dominant, 2800, S9
  - autosomal recessive, 2800
  - cellular/molecular defects in, 459t, 462–463
  - clinical features of, 459t, 463, 1052, 2800, S9
  - diagnosis of, 459t, 2800
  - monocytosis in, 463
- Hyper-IgM syndrome, 2780, 2800, 2801f, 2802
- Hyperimmunoglobulinemia D with periodic fever syndrome (HIDS) (mevalonate kinase deficiency), 155, 2935t, 2936
- Hyperinflation, 270, 2251
- Hyperinsulinemia, 3140, 3202
- Hyperinsulinism, endogenous, 3235
- Hyperkalemia, **358**
  - in AKI, 2382, 2383t
  - in CKD, 2389
  - clinical features of, 359
  - definition of, 358
  - diagnosis of, 359–360, 360f
  - drug-induced, 359
  - ECG in, 359, 1873, 1874f
  - etiology of, 353, 358–359, 358t
  - familial, 795
  - in heart failure, 1984
  - in HIV infection, 359
  - in overdosage/poisoning, 3706
  - rebound, 358
  - renal effects of, 2443
  - treatment of, 361
  - in tumor lysis syndrome, 593
  - urinary tract obstruction and, 2455
  - ventricular tachycardia in, 1956
- Hyperkalemic periodic paralysis (HyperKPP), 359, 3635, 3646, 3804t
- Hyperkinetic movement disorders, **3508,** 3509t
  - chorea. *See* Chorea(s)
  - drug-induced, 3516–3517
  - dystonia. *See* Dystonia
  - gait disorders in, 178
  - hemiballismus, 3515
  - myoclonus. *See* Myoclonus
  - paroxysmal dyskinesias, 3517
  - restless legs syndrome. *See* Restless legs syndrome
  - tics, 3515
  - Tourette's syndrome, 3515
  - tremor, **3508,** 3509t
- Hyperleukocytosis, 591
- Hyperlipidemia. *See* Lipoprotein disorders
- Hyperlipoproteinemia, type III (familial dysbetalipoproteinemia), 3241t, 3245
- Hyperlysinemia, 3369t
- Hypermagnesemia, **3267,** 3267t
- Hypermethioninemia, 3368t
- Hypernatremia, **352**
  - clinical features of, 352
  - diagnosis of, 352–353, 352f
  - etiology of, 352, S1
  - free water deficit in, S1
  - hypodipsic/adipsic, **3019,** 3019t
  - treatment of, 352f, 353
- Hyperopia, 220
- Hyperornithinemia, 3369t, 3374t
- Hyperostosis corticalis generalisata (van Buchem's disease), 3316
- Hyperoxaluria, 2449, 2548, 3368t
- Hyperparathyroidism
  - familial isolated, 3084
  - global considerations in, 366
  - muscle weakness in, 3647
  - prevalence of, 2979t
  - primary, **3275**
    - asymptomatic, 3275, 3277, 3278t
    - clinical features of, 364, 3277–3278
    - diagnosis of, 3278, 3278f, 3283–3284
    - etiology of, 3275–3276
    - genetic considerations in, 364, 3275–3277, 3276f
    - incidence of, 3275
    - in MEN syndromes. *See entries at* Multiple endocrine neoplasia
    - natural history of, 3275
    - pathology of, 3277
    - treatment of, 3278–3279, 3278t
  - screening/testing for, 2979t
  - secondary, 363t, 2390, 3282
  - tertiary, 3282
- Hyperparathyroidism-jaw tumor (HPT-JT) syndrome, 3080t, 3087, 3087t, 3275, 3276
- Hyperparathyroid-like syndromes, 3279–3280, 3280f
- Hyperpathia, 96, 172
- Hyperperfusion disorders, 2349–2350, 2350f, 2350t
- Hyperphagia, 3002, 3185
- Hyperphenylalaninemias, 3367, 3368t, 3370
- Hyperphosphatemia, **3265**
  - in AKI, 2381, 2382, 2383t, 2384
  - in CKD, 2390
  - clinical features of, 3266
  - diseases associated with, 3318t
  - etiology of, 3265–3266, 3265t, 3290
  - hypocalcemia and, 3289, 3290
  - severe acute, 3290
  - treatment of, 3266, 3290
  - in tumoral calcinosis, 3319
  - in tumor lysis syndrome, 593
- Hyperpigmentation, **397**
  - in adrenal insufficiency, 3067, 3069f
  - diffuse, 399–400
  - drug-induced, 400, 420
  - etiology of, 399t
  - in liver disease, 2630
  - localized, 399
  - in Whipple's disease, 1369
- Hyperplastic polyp, 653
- Hyperprolactinemia, **3004**
  - clinical features of, 3004
  - diagnosis of, 3004–3005, 3158
  - erectile dysfunction in, 3157
  - etiology of, 3004, 3005t, 3083
  - galactorrhea and, 3005
  - hirsutism in, 3138
  - hypogonadotropic hypogonadism and, 3114
  - laboratory investigation of, 3005
  - prevalence of, 2979t
  - screening/testing for, 2979t
  - treatment of, 3005
- Hyperprolinemia, 3368t
- Hyperpyrexia, 134, 137
- Hyperreactive malarial splenomegaly, 1766
- Hypersensitivity myocarditis, 2014
- Hypersensitivity pneumonitis, **2228**
  - clinical features of, 2229
  - clinical prediction rule, 2230
  - definition of, 2228
  - diagnosis of, 2229–2230, 2230f, A12
  - differential diagnosis of, 2230
  - etiology of, 2228–2229, 2228t
  - global considerations in, 2231
  - pathophysiology of, 2229
  - treatment of, 2230
- Hypersensitivity reactions. *See also* Anaphylactic reactions
  - to antineoplastic drugs, 597–598
  - to contrast agents. *See* Contrast agents
  - delayed-type, 419, 1384–1385, 2783
  - drug-related, 2815. *See also* Cutaneous drug reactions; Drug-induced hypersensitivity syndrome (DIHS); Drug-induced illness
  - immediate-type, 2782–2783
  - to insect stings, 3735–3736
  - in leprosy, 1408–1409
  - pathophysiology of, 2807–2808
  - to transfusions, 911
  - to vaccine or vaccine component, 1005

- Hypersensitivity vasculitis, 2905
- Hypersplenism, 480, 2714–2715
- Hypertension
  - after kidney transplantation, 2411
  - cardiovascular disease risk in, 2135–2136
  - classification of, 2134, 2134t
  - in coma, 189
  - in diabetes mellitus, 3225–3226
  - diet and, 2603
  - ECG in, A7
  - global considerations in, 1858
  - high-altitude travel and, 3759
  - hypokalemic, 356, 3061, S1
  - laboratory tests, 2135
  - malignant. *See* Malignant hypertension
  - MASH progression and, 2702
  - in metabolic syndrome, 3255, 3258
  - mineralocorticoid-mediated, S1
  - in older adults, 2147–2148, 3879–3880
  - paraneoplastic, 737t
  - pathologic consequences of
    - cardiovascular, 1858, 2100
    - erectile dysfunction, 3155
    - heart failure, 1984. *See also* Heart failure (HF)
    - renal, 338t, 2415t, 2427, A4
    - retinopathy, 226, 226f
    - stroke, 3448–3449, 3453–3454, 3453t. *See also* Intracerebral hemorrhage
  - physical examination in, 1862
  - portal. *See* Portal hypertension
  - portopulmonary, 2717
  - prevalence of, 2134, 2148–2149
  - prevention of, 2140, 2141f
  - primary, 2136, 2137t
  - progesterone-induced, 3060t
  - pulmonary arterial. *See* Pulmonary arterial hypertension (PAH)
  - screening for, 39t
  - secondary, 2138t
    - drug-related, 2137
    - less common causes of, 2140
    - in OSAHS, 2136–2137
    - in pheochromocytoma, 3073
    - in pregnancy, 2139–2140, 3898–3899
    - in primary aldosteronism, 2139
    - in renal parenchymal disease, 2139
    - in renovascular disease, 2139
  - stroke and, 3446t
  - supine, 159
  - treatment of, **2143**
    - behavioral economics in, 3917–3918
    - best practices for, 2149, 2149f
    - in CKD, 2392
    - in metabolic syndrome, 3258
    - nonpharmacologic, 2137t, **2140**
    - pharmacologic, **2143**
      - approach to, 2140, 2141f, 2146f
      - blood pressure target in, 2145–2146
      - in cardiovascular disease, 2146
      - in CKD, 2146–2147
      - combination, 2145
      - in diabetes mellitus, 2147
      - drug classes for, 2143–2145, 2144–2145t. *See also specific drugs*
      - efficacy of, 2143
      - in older adults, 2147–2148
      - racial/ethnic considerations in, 2148
      - for resistant hypertension, 2146, 2147f
      - for stroke prevention, 3448–3449
      - in women, 2148
    - in polycystic kidney disease, 2433
    - renal nerve denervation for, 2132, 2146
  - in urinary tract obstruction, 2455
  - vasoconstrictor use in, 266
  - white coat, 1862
  - in women, 3165
- Hypertensive emergencies, 2148. *See also* Malignant hypertension
- Hypertensive encephalopathy, 3455
- Hypertensive urgency, 2148
- Hyperthecosis, 3138
- Hyperthermia, 134, 150. *See also* Heat-related illness
- Hyperthyroidism
  - clinical features of, 166
    - alopecia, 394
    - cardiac, 2018
    - cutaneous, 400, 402
    - diarrhea, 309
    - edema, 282
    - hypercalcemia, 3281–3282
    - hypertension, 2140
    - hypokalemia, 354, S1
    - muscle involvement, 3647
  - definition of, 3032
  - in gestational trophoblastic disease, 715
  - in Graves' disease. *See* Graves' disease
  - in HIV infection, 1603
  - vs. myasthenia gravis, 3628
  - paraneoplastic, 737t
  - in pregnancy, 3901
- Hyperthyroxinemia, euthyroid, 3024, 3024t
- Hypertrichosis, 421, 435
- Hypertriglyceridemia
  - approach to the patient, 3249
  - definition of, 3240
  - in diabetes mellitus, 3228
  - diagnosis of, 3249
  - with elevated LDL-C. *See* Mixed hyperlipidemia
  - with low HDL-C, 3249
  - in metabolic syndrome, 3253f, 3255, 3258
  - multifactorial, 3241–3242
  - pancreatitis and, 2744, 2748–2749
  - in parenteral nutrition, 2625
  - primary (genetic) causes of, 3240–3242, 3241t
  - secondary causes of, 3242t, 3246
  - treatment of, 3250, 3258
- Hypertrophic cardiomyopathy, **2005**
  - apical, 2006
  - clinical features of, 2002t, 2006
    - heart palpation in, 1864
    - heart sounds/murmurs, 284f, 285, 289, 1865, 1867
    - syncope. *See* Cardiac syncope
  - convergent phenotypes in, 31f
  - in Danon disease, 2004t, 2010, 3361t, 3365
  - diagnosis of, 2006–2007
    - CMR, 1894, 1897f, A9
    - ECG, A7
    - echocardiography, 2006f, A9, A9v
    - initial approach to, 2002t
  - early interventions for, 3804t, 3805
  - epidemiology of, 2005
  - in Fabry's disease, 3355
  - genetic considerations in, 2004t, 2005, 3788t, 3789
  - genetic testing for, 3804t
  - incidence of, 1971
  - mitral regurgitation in, 2052
  - natural history of, 1841
  - pathology of, 2005–2006, 2006f
  - prognosis of, 2008
  - pulmonary hypertension in, 2192, 2193f
  - sudden death and, 1971, 2008, 2008t
  - treatment of, 1971, 2339t, A11, A11v
  - ventricular tachycardia in, 1971
- Hypertrophic osteoarthropathy, **2969**
  - clinical features of, 2970–2971, 3318
  - clubbing in, 279, 1860, 2970–2971, 3318
  - diagnosis of, 279
  - differential diagnosis of, 3318l
  - disorders associated with, 2970–2971, 2971t
  - familial, 2970
  - in lung cancer, 619, 2970
  - pathogenesis of, 279
  - pathophysiology of, 2970
  - treatment of, 2971
- Hyperuricemia
  - in AML, 829
  - in gout. *See* Gout
  - metabolic syndrome and, 3256
  - renal effects of, 2443
  - treatment of, 2959–2960
  - in tumor lysis syndrome, 593
- Hyperuricosuria, 343
- Hypervalinemia, 3369t
- Hyperventilation, 376, 2276–2277
- Hyperventilation tetany, 3774
- Hyperviscosity, 451, 2178
- Hypervolemia, 345f, 2382
- Hypesthesia, 172
- Hypnagogic jerks, 3515
- Hypnic headache, 115t, 3475
- Hypnosis, 3925t, 3927
- Hypnozoites, *Plasmodium,* 1760
- Hypoactive sexual desire, female, 3161
- Hypoalbuminemia, 282, 342, 346, 1860, 2621, 2636
- Hypoaldosteronism, 347, 358t, 359, 372
- Hypoalphalipoproteinemia, primary, 3248–3249
- Hypobetalipoproteinemia, familial, 3241t, 3247
- Hypocalcemia, **365, 3285**
  - after parathyroidectomy, 3279
  - in AKI, 2382, 2383t
  - chronic, 3285–3286, 3291
  - classification of, 3286, 3286t
  - clinical features of, 365, 2018, 2614t, 3279, 3286t
  - diagnosis of, 365–366
  - differential diagnosis of, 3291
  - ECG in, 1874, 1875f, A7
  - etiology of, 363t, 365, 3286–3287
  - genetic syndromes associated with, 3286–3287
  - global considerations in, 366
  - hyperphosphatemia and, 3265, 3290
  - hypomagnesemia and, 3286–3287
  - inherited disorders, 2368t, 2370
  - in pancreatitis, 2745, 3286
  - pathophysiology of, 3285–3286, 3294
  - PTH-related
    - hypoparathyroidism. *See* Hypoparathyroidism
    - pseudohypoparathyroidism. *See* Pseudohypoparathyroidism (PHP)
  - treatment of, 366, 3279
  - vitamin D–related, 3286t. *See also* Vitamin D deficiency
- Hypocalciuria, 355
- Hypocaloric nutrition, 2620
- Hypocapnia, 367, 376
- Hypochondria, 478
- Hypochromia, 437, 445, 447t
- Hypocretin (orexin), 205, 211f, 215, 242, 3398
- Hypocupric myelopathy, 3568
- Hypodipsic/adipsic hypernatremia, **3019,** 3019t
- Hypoesthesia, 172
- Hypogammaglobulinemia
  - bronchiectasis in, 2242, 2242t
  - clinical features of, 309
  - in CLL, 852
  - disorders associated with, S9
  - infectious diarrhea in, 1080
  - in multiple myeloma, 886
  - in X-linked lymphoproliferative syndrome, 466
- Hypogeusia. *See* Taste disorders
- Hypoglossal nerve, 3380, 3558
- Hypoglycemia, **3230**
  - accidental, 3235–3236
  - approach to the patient, 3236

  - in cancer, 502–593
  - clinical features of, 3232, 3232t
  - in critical illness, 3234
  - in diabetes mellitus, 3230, 3232–3234, 3233f
  - diagnosis of, 3236
  - drug-induced, 3234
  - endogenous hyperinsulinism and, 3235
  - etiology of, 3230, 3231t, 3232
  - exercise-related, 3206, 3236
  - factitious, 3235–3236
  - in heatstroke, 3775
  - hormonal defense against, 2986
  - hormone deficiencies and, 3234–3235
  - hypothermia in, 3768
  - in inborn errors of metabolism, 3236
  - in insulin autoimmune syndrome, 3093
  - insulin-induced, 2997
  - with insulin plus glucose, 361
  - in malaria, 1764, 1773
  - non–beta-cell tumors and, 3235
  - in overdosage/poisoning, 3705
  - in parental nutrition, 2625
  - pathophysiology of, 3231f, 3232, 3232t
  - perioperative, 3909
  - recurrent, prevention of, 3236
  - risk factor reduction, 3233–3234
  - syncope and, 160–161
  - treatment of, 3236
  - tumor-induced, 737t, 739–740
  - unawareness, 3233
- Hypogonadism. *See also* Testosterone deficiency
  - acquired testicular causes of, 3114–3115
  - androgen abuse and, 3123
  - in cryptorchidism, 3114
  - erectile dysfunction in, 3157
  - in hemochromatosis, 3336
  - in HIV infection, 1603
  - hypergonadotropic, 3134–3135
  - hypogonadotropic. *See* Hypogonadotropic hypogonadism
  - in Klinefelter's syndrome, 3114. *See also* Klinefelter's syndrome
  - prevalence of, 2979t
  - screening/testing for, 2979t
- Hypogonadotropic hypogonadism, **3111**. *See also* Testosterone deficiency
  - acquired, 3113–3114
  - amenorrhea in, 3133–3134
  - clinical features of, 2997–2998
  - congenital disorders associated with, 3111–3112t, 3111–3113
  - delayed puberty in, 3110, 3131t, 3132
  - diagnosis of, 2997–2998
  - etiology of, 2997
  - familial, 3111
  - genetic considerations in, 3110
  - hemochromatosis and, 3114
  - hyperprolactinemia and, 3114
  - infertility in, 3151
  - isolated, 2997, 3133–3134
  - obesity and, 3114, 3185
  - pathophysiology of, 3111
  - sellar mass lesions and, 3114
  - treatment of, 2998
- Hypokalemia, **354**
  - arrhythmia in, 1913
  - in CKD, 2389
  - clinical features of, 356
  - diagnosis of, 356, 357f
  - in DKA, 3218
  - ECG in, 356, 1874, 1875f, A7
  - etiology of, 354–355
  - in Gitelman's syndrome, S1
  - in heart failure, 1984
  - in hyperthyroidism, S1
  - hyponatremia and, 352
  - magnesium deficiency in, 356, S1
  - in mineralocorticoid excess, 3061
  - in overdosage/poisoning, 3705
  - renal effects of, 2443
  - treatment of, 356–357, S1
- Hypokalemic periodic paralysis (HypoKPP)
  - clinical features of, 354, 3646
  - etiology of, 354
  - familial, 354
  - genetic considerations in, 354, 3646
  - thyrotoxicosis and, 3033, 3647, S1
  - treatment of, 357
- Hypolipidemia, familial combined, 3241t, 3248–3249
- Hypomagnesemia, **3266**
  - after parathyroidectomy, 3279
  - arrhythmia in, 1913
  - clinical features of, 3267
  - etiology of, 3266–3267, 3266t
  - familial, 2368t, 2370
  - genetic considerations in, 3267
  - hypocalcemia and, 3287–3288
  - hypokalemia and, 354, 356, S1
  - hypoparathyroidism and, 3287–3288
  - primary, 2368t
  - treatment of, 356, 3267, 3279, 3288, S1
  - vitamin D deficiency and, 3267
- Hypomelanosis, 396, 398t
- Hypomelanosis of Ito (linear nevoid hypopigmentation), 397
- Hyponatremia, **347**
  - in adrenal insufficiency, 3067
  - after subarachnoid hemorrhage, 348, 3459, 3461
  - in AKI, 2382, 2383t
  - chronic, 349–350, S1
  - in CKD, 2389
  - clinical features of, 188, 349–350
  - diagnosis of, 348f, 350
  - etiology of, 349t, 2371f
  - euvolemic, 348. *See also* Syndrome of inappropriate antidiuresis (SIAD)
  - exercise-associated, 349
  - in heart failure, 1984
  - hypervolemic
    - disorders associated with, 348
    - pathophysiology of, 348
    - vs. SIAD, 3020
    - treatment of, 351
  - hypokalemia and, 351
  - hypovolemic
    - differential diagnosis of, 350, 3020
    - disorders associated with, 347–348
    - pathophysiology of, 347–348
    - treatment of, S1
  - with low solute intake, 348–349
  - in lung cancer, 619
  - paraneoplastic, 738
  - treatment of, 350–351, 351t, 3020, S1
- Hypoparathyroidism
  - acquired, 3287
  - APS-1 and, 3089, 3090t
  - APS-2 and, 3091t
  - chronic, 3287
  - clinical features of, 3286–3287
  - diagnosis of, 3278f
  - etiology of, 363t
  - genetic syndromes associated with, 3286–3287
  - hypocalcemia and, 363t, 365, 3286–3287, 3286f
  - isolated, 3287
  - muscle involvement in, 3647
  - transient, 3287
  - treatment of, 3287
- Hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome, 3286
- Hypophosphatasia, 261, 3316
- Hypophosphatemia, **3263**
  - clinical features of, 2018, 2614t, 3264
  - etiology of, 3263–3264, 3263t
  - hypercalcemia and, 738
  - hypokalemia and, S1
  - osteomalacia and, 740, 3270. *See also* Osteomalacia
  - treatment of, 3264–3265, 3265t
- Hypophysitis, 2993, 3090t, 3091t
- Hypopigmentation, **396**
  - diffuse, 396
  - etiology of, 397t, 398t
  - linear nevoid, 397
  - localized, 396
  - postinflammatory, 397, 398t
  - systemic diseases associated with, 396–397
- Hypopituitarism, 261, **2992,** 2996t, 3768
  - acquired causes of, 2993–2994
  - clinical features of, 2994
  - diagnosis of, 2994, 2995t
  - etiology of, 2992–2994, 2992t
  - treatment of, 2994
- Hypoplastic left heart syndrome, 2076
- Hypoproliferative anemias. *See* Aplastic anemia; Iron-deficiency anemia
- Hypoproteinemia, 282
- Hypopyon, 2909
- Hyposensitization. *See* Immunotherapy
- Hyposmia. *See* Olfactory disorders
- Hypospadias, isolated, 3100
- Hyposplenism, 482. *See also* Splenectomy
- Hypotension
  - in coma, 192
  - in gastrointestinal bleeding, 318
  - in heatstroke, 3775
  - in hemodialysis, 2400
  - in hypovolemia, 347
  - intracranial, 3483
  - in malabsorption disorders, 2549t
  - orthostatic. *See* Orthostatic hypotension
  - in shock, 2292–2293, 2294f, 2312
  - transfusion-related, 908t, 912
- Hypothalamic disease/masses
  - body temperature in, 134, 3002
  - clinical features of, 2999t, 3002
  - gliomas, 3003
  - hamartomas, 3003, A16
  - hypopituitarism due to, 2992–2993. *See also* Hypopituitarism
  - IgG4-related, 2931t
  - obesity in, 3185–3186
- Hypothalamic-pituitary-adrenal (HPA) axis, 2987, 2991, 2997, 3050–3051, 3052f
- Hypothalamic-pituitary-gonadal axis, 3103f, 3109, 3124, 3133, 3133f
- Hypothalamic-pituitary-testicular axis, 3117, 3123
- Hypothalamic-pituitary vasculature, 2989f
- Hypothalamic set point, 134
- Hypothalamus
  - developmental dysfunction of, 2992–2993
  - in energy balance regulation, 3182, 3182f, 3183f
  - in ovarian function regulation, 3126–3127, 3127f
  - in sleep onset, 211
- Hypothermia, **3768**
  - clinical features of, 3769, 3769t
  - coma and, 189
  - diagnosis of, 3769–3770
  - drug-induced, 3768, 3768t
  - ECG in, 1874, 1875f
  - etiology of, 3768–3769, 3768t
  - in myxedema coma, 3032
  - in older adults, 3768, 3771
  - physiologic changes in, 3769, 3769t
  - primary accidental, 3768
  - rewarming strategies for, 3770–3771, 3770t

- Hypothermia (*Cont.*):
  - risk factors for, 3768t
  - secondary, 3768
  - therapeutic, 2338
  - treatment of, 3771
- Hypothyroidism, **3027**
  - amiodarone-induced, 3039–3040
  - autoimmune. *See* Hashimoto's thyroiditis
  - clinical (overt), 3028, 3031
  - clinical features of, 166, 3029–3030, 3029f
    - alopecia, 394
    - ascites, 328
    - ataxia, 3533
    - bleeding, 473
    - cardiac, 1918, 2018
    - carotenemia, 2612
    - carotenoderma, 320
    - cutaneous, 402, A15
    - depression, 3664
    - dyslipidemia, 3246
    - edema, 282
    - hirsutism, 3138
    - hypertension, 2140
    - hyponatremia, 348
    - hypothermia, 3768, 3771
    - muscle involvement, 3647
    - neuropathy, 3604
    - obesity, 3185
  - congenital, 2983, 3021, **3027,** 3028t
  - consumptive, 740
  - diagnosis of, 3030–3031, 3030f
  - differential diagnosis of, 3030–3031
  - diseases associated with, 3030
  - etiology of, 3027, 3027t, 3031
  - genetic considerations in, 3028t
  - in HIV infection, 1603
  - iatrogenic, 3031
  - in pregnancy, 3023, 3031–3032, 3901–3902
  - prevalence of, 2979t
  - radiation therapy-related, 614, 756, 871
  - screening/testing for, 2979t
  - secondary, 3027t, 3031
  - subclinical, 3028, 3031
  - in systemic sclerosis, 2871
  - transient, 3027t
  - treatment of, 3031–3032
- Hypotonic saline, 353. *See also* Saline
- Hypouricemia, 2368t
- Hypoventilation. *See also* Obstructive sleep apnea/hypopnea syndrome (OSAHS)
  - central hypoventilation syndrome, 2276
  - clinical features of, 2273–2274, 2274t
  - diagnosis of, 2274–2275, 2275f
  - hypoxia in, 277
  - in obesity, 2276, 3187
  - pathophysiology of, 2273, 2274f
  - respiratory acidosis in, 375
  - treatment of, 2275–2276
- Hypovolemia, **346**
  - in AKI, 2382
  - AVP levels and, 345f
  - clinical features of, 346–347
  - diagnosis of, 346–347
  - etiology of, 346–347
  - in STEMI, 2123
  - treatment of, 347
- Hypovolemic shock
  - clinical features of, 347
  - hemodynamic patterns in, 2311t, 2329t
  - initial assessment of, 2292–2294, 2311
  - pathophysiology of, 2311
  - physical examination in, 2312
  - predictors of fluid responsiveness in, 2293–2294, 2294f
- Hypoxemia, 444, 2745
- Hypoxia, **276**
  - adaptation to, 277–278
  - in angiogenesis, 545, 547f
  - effects of, 276–277
  - etiology of, 277
- Hypoxia-inducible factors (HIFs), 276, 444, 541, 545, 2187, 2187f, 2393
- Hypoxic-ischemic encephalopathy, 2346–2348, 2347f, 3428
- Hy's Law, 2667
- Hysterectomy, 711, 713, 714
- Hysterosalpingogram, 3150

I

- I-309/TCA-3, 2767t
- IA-2 (ICA512), 2790t
- IABP (intraaortic balloon pump), 2132, 2328, 2331
- IAP (inhibitor of apoptosis proteins), 539, 540f
- IAPP (islet amyloid polypeptide), 3198
- Iatrogenic hypothyroidism, 3031
- Ibalizumab, 1617f, 1618t, 1621
- I band, of sarcomere, 1847
- Ibandronate, 638, 3145, 3304, 3305f
- IBD. *See* Inflammatory bowel disease (IBD)
- IBM. *See* Inclusion body myositis (IBM)
- IBMPFD (inclusion body myopathy with Paget's disease and frontotemporal dementia), 3312
- IBPN (immune-mediated brachial plexus neuropathy), 3614–3615
- Ibrexafungerp, 1692, 1698, 1710t
- Ibritumomab, 2791t
- Ibrutinib
  - action and targets of, 533t
  - actions of, 567, 568t
  - adverse effects of, 568t, 579, 753t, 754, 757, 2017
  - for CLL, 855
  - for cold agglutinin disease, 803
  - for DLBCL, 862
  - for GVHD, 917
  - interactions and issues, 568t
  - for lymphoplasmacytic lymphoma, 865
  - for mantle cell lymphoma, 866
  - for nodal marginal zone lymphoma, 865
  - for Waldenström's macroglobulinemia, 894
- IBS. *See* Irritable bowel syndrome (IBS)
- Ibuprofen
  - for acute idiopathic pericarditis, 2079
  - advantages of, 2530
  - adverse effects of, 923t, 2955t
  - drug interactions of, 2955, 2955t
  - for dysmenorrhea, 3137
  - for fever, 136
  - for frostbite, 3772t
  - for gout, 2959
  - for migraine, 3467t
  - for osteoarthritis, 2955t
  - for pain, 80, 80f, 97t
  - for prevention of acute mountain sickness, 3756
- Ibutilide, 1916t, 1941, 1949
- ICAM-1 (intercellular adhesion molecule 1), 1585
- Icatibant, 2809
- IC/BPS. *See* Interstitial cystitis/bladder pain syndrome (IC/BPS)
- ICD (implantable cardioverter-defibrillator). *See* Cardiovascular implantable electronic devices (CIEDs)
- Ice-pack test, 3626–3627
- ICF (immunodeficiency with centromere instability and facial anomalies), 2800, S9
- Ichthyosis, in HIV infection, 1605
- Icodextrin, 2401
- Icosapent, 3228, 3244t
- *ICOS* gene mutations, 2802
- ICP. *See* Intracranial pressure (ICP)
- Icterus, 2629
- Ictotest, 321
- ICU patient. *See* Critically ill patient
- ICVP (International Certificate of Vaccine Prophylaxis), 1012
- Idarubicin, 562t, 566, 830
- Idarucizumab, 953, 1951, 2163, 3456
- IDDM (insulin-dependent diabetes mellitus). *See* Diabetes mellitus, type 1
- Ideational apraxia, 203
- Idecabtagene vicleucel, 535t, 890t, 892, 3824t
- Idelalisib
  - action and targets of, 534t, 535, 571t, 574
  - adverse effects of, 571t
  - for follicular lymphoma, 864
  - interactions and issues, 571t
- Ideomotor limb apraxia, 203
- *IDH1/IDH2* gene mutations
  - 2-HG levels and, 3978, 3979f
  - in AML, 828, 3940
  - in cholangiocarcinoma, 667, 668
  - in chronic eosinophilic leukemia, 876
  - in glioma, 520t, 718
  - in myelofibrosis, 823
  - in oligodendroglioma, 718
- IDH inhibitors, 544, 571t, 574, 3940
- Idiopathic CD4+ T lymphocytopenia, 1614–1615
- Idiopathic cutaneous vasculitis, **2905**
- Idiopathic pulmonary fibrosis (IPF). *See also* Interstitial lung diseases (ILDs)
  - acute exacerbations of, 2265
  - clinical features of, 2260t, 2262
  - diagnosis of, 2260t, 2262, 2263f
  - with emphysema, A12
  - etiology of, 2260t
  - imaging of, A12
  - lung transplantation for, 2283
  - pathology of, 2262, 2264f
  - treatment of, 2262–2263
- Idiotypes, 883
- Idioventricular rhythms, 1955, 1955f, 1963. *See also* Premature ventricular contractions (PVCs)
- IDLs (intermediate-density lipoproteins), 3237, 3237f
- Idoxuridine, 1502
- IF (intrinsic factor), 781, 786, 2790t
- *IFI16* gene, 2761t
- *IFITM5* gene mutations, 3325t, 3326, 3327
- *IFNG* gene mutations, 2823t
- *IFNGR1/2* gene mutations, 1415
- *IFNL3,* E72t
- *IFNL3* gene mutations, 2684, 2687
- *IFNL4,* 1582t
- IFN-α. *See* Interferon-α (IFN-α)
- IFN-α2a. *See* Pegylated interferon-α2a (PEG IFN-α2a)
- IFN-α2b. *See* Pegylated interferon-α2b (PEG IFN-α2b)
- IFN-γ-inducible protein-10 (IP-10), 2768t
- Ifosfamide
  - actions of, 563
  - adverse effects of, 562t
    - cardiotoxicity, 754, 2017
    - cystitis, 597, 757
    - hypokalemia, 355
    - interstitial disease, 2356
    - long-term, 753t
    - nausea and vomiting, 575
    - nephrogenic diabetes insipidus, 352
    - nephrotoxicity, 2377
    - neurologic, 563, 591, 726, 726t
    - pulmonary, 595, 754
  - for endometrial cancer, 715
  - for Ewing's sarcoma, 730

  - interactions and issues, 562t
  - for osteosarcoma, 730
  - for soft tissue sarcomas, 729
  - for testicular cancer, 708f, 709
- Ig. *See* Immunoglobulin(s)
- IgA nephropathy
  - in axial spondyloarthritis, 2884
  - clinical features of, 2415t, 2419
  - epidemiology of, 2419
  - hematuria in, 343
  - vs. IgA vasculitis, 2419
  - pathogenesis of, 2419, 2419f, 2771
  - pathophysiology of, 2355
  - renal biopsy in, 2419, A4
  - treatment of, 2355, 2419
- IgA vasculitis (Henoch-Schönlein), **2904**
  - clinical features of, 148f, 408, 927, 2356, 2415t, 2904
  - definition of, 2904
  - diagnosis of, 2904
  - vs. IgA nephropathy, 2419
  - incidence and prevalence of, 2904
  - inflammasome mutations in, 2762t
  - pathogenesis of, 927, 2904
  - treatment of, 2905
- IGF. *See* Insulin-like growth factor (IGF)
- Ig Fc receptors, 2755, 2769
- IgG4-related disease (IgG4-RD), **2930**
  - clinical features of, 2166, 2171, 2439, 2930–2931, 2931t, 2932f
  - epidemiology of, 2932
  - hypertrophic pachymeningitis, 1132t
  - pathology of, 2930, 2932–2933, 2932–2933f
  - pathophysiology of, 2933
  - sclerosing cholangitis, 2737, 2751
  - serologic findings in, 2931–2932
  - treatment of, 2933
- IGHV (immunoglobulin heavy chain) mutation, 850, 854t
- IgLON5 antibody, 744t, 749
- IGRAs (interferon-γ release assays), 149, 1384, 1394, 1417
- IKK (IκB kinase), 536, 971
- IKKγ deficiency, 2799
- iKNRs (inhibitory NK cell receptors), in HIV infection, 1580
- *IL2/IL2R* gene mutations, 2823t
- *IL2RG* gene mutations, 3825
- *IL6* gene mutations, 2823t
- *IL10* gene mutations, 2553, 2554t, 2823t
- *IL12B* gene mutations, 2553, 2554t
- *IL17A* gene mutations, 2823t
- *IL17RD* gene mutations, 3112, 3112t
- *IL21R-BCL6* gene mutations, 521t
- *IL21R-RB1* gene mutations, 1416
- *IL23R* gene mutations, 2553, 2554t, 2823t
- *IL28B* gene mutations, 2650
- *IL33* gene mutations, 2823t
- ILCs. *See* Innate lymphoid cells (ILCs)
- ILDs. *See* Interstitial lung diseases (ILDs)
- Ileal pouch-anal anastomosis (IPAA), 2565, 2567
- Ileal resection, 786
- Ileal stricture, 2472f
- Ileorectostomy, 314
- Ileostomy, 694
- Ilesha virus, 1658t
- Ileus
  - etiology of, 2590, 2590t, 2593
  - gallstone, 2593, 2733
  - hypokalemia and, 356
  - postoperative, 98
  - treatment of, 2593
- Ilhéus virus, 1659t
- Iliopsoas bursitis, 2944f, 2945, 2973
- Iliopsoas sign, 2595t
- Iliotibial band syndrome, 2974
- Iloperidone, 3673t
- Iloprost, 2195, 2197t, 3772
- IMAGe syndrome, 3067t
- Imatinib
  - actions and targets of, 532, 533t, 535, 567, 568t, 2187f
  - adverse effects of, 567, 568t
    - cardiac, 567
    - common, 840
    - growth deceleration, 756
    - hepatic, 567
    - management of, 841–842
    - myopathy, 2940t
    - pulmonary, 595
    - serious, 840–841
  - for ALL, 847
  - for CML, 521, 533t, 536, 567, 839t, 840–841
  - cost of, 843
  - drug interactions of, 591
  - genetic variations in response to, 493, 567, E72t
  - for GIST, 535, 567, 651, 729
  - for hypereosinophilia, 879
  - for hypereosinophilic syndromes, 462, 2233
  - interactions and issues, 568t
  - for melanoma, 605
  - resistance to, 536, 567, 836
  - for soft tissue sarcomas, 729
- Imerslund-Gräsbeck syndrome, 786–787
- Imerslund's syndrome, 786–787
- Iminoglycinuria, 3374t
- Imipenem
  - for *A. baumannii* infections, 1296
  - actions of, 1183f
  - adverse effects of, 1175
  - for anaerobic bacterial infections, 1379, 1379t
  - for bite-wound infections, 1145t
  - for *Campylobacter* infections, 1306
  - for community-acquired pneumonia, 1028t
  - for enterococcal infections, 1219t
  - for glanders, 1309t
  - for hospital-acquired pneumonia, 963t, 1032t
  - indications for, 1173t, 1175
  - for intraabdominal infections, 963t
  - for *M. abscessus* infection, 1428
  - for melioidosis, 1309t
  - for neutropenic fever, 580
  - for *Nocardia* infections, 1362
  - for NTM infections, 1428
  - for osteomyelitis, 1064t, 1065t
  - for *P. aeruginosa* infections, 1309t
  - for peritonitis, 1071
  - resistance to, 1173t, 1181, 1183f, 1359f
  - for sepsis/septic shock, 963t
  - for ventilator-associated pneumonia, 1032t
- Imipenem-cilastin
  - for *A. baumannii* infections, 1297t
  - for actinomycosis, 1366t
  - empirical uses of, 2321t
  - for MDR-TB, 1394t, 1426
  - in pregnancy and lactation, 1169t
- Imipenem-cilastin-relebactam, 1169t, 1174t
- Imipenem-relebactam, 1032, 1309t
- Imipramine
  - adverse effects of, 324, 420, 3659t
  - for depression, 3659t, 3666
  - for detrusor spasticity, 3570
  - for functional dyspepsia, 302
  - for IBS, 2575
  - overdosage/poisoning with, 3711t, 3712t
  - for pain, 97t
  - for sleep enuresis, 217
- Imiquimod
  - for actinic keratoses, 609
  - for basal cell carcinoma, 608
  - for cutaneous leishmaniasis, 1787
  - for extramammary Paget's disease, 610
  - for genital warts, 390, 1489, 1530, 1531t
  - for HSV infection, 1502
  - for lentigo maligna, 603
  - for melanoma, 603
- Immersion cooling, 3775
- Immersion (trench) foot, 3771
- Immune checkpoint blockade/inhibitors. *See also* CTLA-4 antibodies; PD-1 antibodies
  - action and targets of, 2783, 2786f, 2787–2788
  - adverse effects of, 605
    - autoimmune conditions, 755, 2788, 2827, 2972, 3092
    - autoimmune endocrine disease, 740, 756, 2788
    - bullous pemphigoid, 412
    - cardiotoxicity, 753t, 754, 2015
    - colitis, 2561, 2788
    - cutaneous, 412, 421–422
    - diabetes mellitus, 3203
    - gastritis, 2539
    - hepatotoxicity, 2672, A13
    - myasthenia gravis, 3633
    - myopathies, 3648t
    - nephrotoxicity, 2439
    - neurologic, 2351, 2917
    - optic neuritis, 755
  - enterococcal potentiation by, 1216
  - indications for, 2787t
  - for lung cancer, 628–629
  - for melanoma, 604–605
  - for NETs, 686
  - for nonmelanoma skin cancer, 432
  - resistance to, 574
  - for thyroid cancer, 3048
- Immune checkpoints, 526, 2789, 2789t
- Immune-complexes, 2782, 2782f, 2826
- Immune deficiency diseases, primary. *See* Primary immune deficiency diseases
- Immune-enhancing formulas, for enteral nutrition, 2622
- Immune homeostasis, 2784
- Immune-mediated brachial plexus neuropathy (IBPN), 3614–3615
- Immune-mediated necrotizing myopathy, 2910t, 2912–2913, 2913f, 2914–2915, 2916t
- Immune memory, 2756
- Immune-privileged sites, 2826, 3397–3398
- Immune reconstitution inflammatory syndrome (IRIS)
  - in cryptococcosis, 1706
  - diagnosis of, 1164
  - in histoplasmosis, 1695
  - in HIV infection
    - clinical features of, 1604, 1604t
    - with cryptococcosis, 1706
    - with molluscum contagiosum infections, 1521
    - pathogenesis of, 1576
    - with PML, 1117
    - with TB, 1391, 1420
    - with tuberculosis, 1597
  - pathogenesis of, 1576, 1604
  - with PCP treatment, 1163
  - in Whipple's disease, 1369
- Immune response/immune system, **2755**
  - adaptive. *See* Adaptive immune system
  - aging and, 3872
  - antitumor effects, 555
  - cell-mediated immunity, 555, 1384
  - cellular interactions in regulation of, 2779–2780
  - cellular migration in, 2782f
  - circadian regulation of, 3953
  - clinical evaluation of, 2783
  - components of, 2793–2794

- Immune response/immune system (*Cont.*):
  - dysregulation of
    - in aging, 2789, 2793
    - in autoimmune disease, 2789. *See also* Autoimmune diseases
    - genetic considerations in, 2786f
  - endogenous derangements of. *See* Autoimmunity
  - endothelium in, 1845
  - epigenetic influences on, 3942
  - evasion of, by cancer, 544, 548–549, 550f
  - in functional dyspepsia, 300
  - of gut, dysfunction of, 2460
  - in HIV infection, 1586–1588, 1586t, 1587f
  - in IBD, 2554–2555
  - innate. *See* Innate immune system
  - microbiota in regulation of, 3840
  - at mucosal surfaces, 2780–2781
  - pattern recognition in. *See* Pattern recognition receptors (PRRs)
  - in sepsis/septic shock, 2317–2319, 2318f
  - suppression by cancer cells, 548–549, 550f
  - terminology related to, 2755–2756
  - to vaccines, 996–997
  - in women, 3166
- Immune thrombocytopenia, 853
- Immune thrombocytopenic purpura, 460, 924
- Immune tolerance, 931, 2780
- Immunization, **1001**. *See also* Vaccine(s)
  - for adults, 40t, 1003
    - administration of, 1008, 1008f
    - contraindications to, 1003, 1006–1007t
    - determining need for, 1003
    - for international travel, 1011–1015, 1012t, 1014t
    - practice standards, 1003
    - provider recommendation for, 1009–1010
    - recommendations for, 1003
    - requirements for, 1010
    - schedules for, 1003, 1004t
    - system support for, 1010
  - consumer access to and demand for, 1009
  - disruption during COVID-19 pandemic, 1639
  - documentation of, 1008–1009
  - for HCT recipients, 918
  - for health care workers, 1005, 1005t, 1153, 1153t
  - in HIV infection, 1004t, 1594t
  - for immunocompromised persons, 1005
  - for international travelers, 1011–1013, 1014t
  - microbiome and response to, 3839
  - in nonmedical settings, 1010
  - in older adults, 1005, 3878
  - performance monitoring, 1010
  - in pregnancy, 1004t, 1005, 3904
- Immunization Action Coalition, 1007
- Immunoblastic lymphomas, 1613. *See also* Diffuse large B cell lymphoma (DLBCL)
- Immunocompromised patient. *See also* Transplant recipient
  - abdominal pain in, 111
  - actinomycosis in, 1363
  - antibacterial therapy in, 1168
  - bite-wound infection in, 1144
  - *Candida* infections in, 390
  - CMV infections in, 1513t, 1514
  - diarrhea in, 305, 1080
  - dysphagia in, 295
  - immunization in, 1004t, 1005
  - international travel by, 1020
  - *Legionella* spp. infections in, 1271–1272
  - primary CNS lymphoma in, 721
  - skin cancer in, 432–433
- Immunodeficiency viruses, 1557, 1558f, 1559–1561. *See also* HIV
- Immunodeficiency with centromere instability and facial anomalies (ICF), 2800, S9
- Immunodysregulation polyendocrinopathy enteropathy X-linked syndrome (IPEX), 3092
- Immunodysregulation polyendocrinopathy enteropathy X-linked (IPEX) syndrome, 2789, 2803
- Immunoediting, 2784
- Immunofluorescent stains, S11
- Immunoglobulin(s). *See also* Antibody(ies)
  - in adaptive immunity, 2778–2779
  - classification of, 883
  - electrophoresis of, 883, 885f
  - gene rearrangements in, 883, 884f, 2778–2779
  - polyvalent, 908, 911
  - properties of, 2779t
  - replacement of, 2802–2803
  - structure of, 883, 884f, 2778–2779, 2779t
  - types of, 2773
- Immunoglobulin A (IgA), 2779, 2779t, 2780, 2790t
- Immunoglobulin A (IgA) deficiency, 2545, 2802
- Immunoglobulin A (IgA) nephropathy. *See* IgA nephropathy
- Immunoglobulin A (IgA) vasculitis. *See* IgA vasculitis
- Immunoglobulin D (IgD), 2779, 2779t
- Immunoglobulin E (IgE)
  - assays for, 2810
  - in asthma, 2223
  - desensitization, 2814
  - functions of, 2779t, 2792t
  - in immediate drug reactions, 417t, 418f
- Immunoglobulin E (IgE) antibody. *See* Omalizumab
- Immunoglobulin E (IgE) receptor, 2790t
- Immunoglobulin G (IgG), 2802
  - in paraneoplastic pemphigus, 411
  - in pemphigus foliaceus, 410
  - in pemphigus vulgaris, 409
  - properties of, 2779, 2779t
  - receptors for, 2769
  - transplacental transfer of, 458
- Immunoglobulin G (IgG) antibodies
  - mumps, 1649
  - in red blood cells, 901
  - rubella, 1644
- Immunoglobulin gene superfamily, 2775
- Immunoglobulin heavy chain (IGHV) mutation, 850, 854t, 855
- Immunoglobulin M (IgM), 865, 2779, 2779t
- Immunoglobulin M (IgM) antibodies, 901, 1334, 1644
- Immunoglobulin replacement, 2802–2803
- Immunohistochemical stains, S11
- Immunologic memory, 2774
- Immunologic priming, 2774
- Immunologic synapse, 2777
- Immunomodulation, transfusion-related, 912
- Immunoproliferative small intestinal disease, 885, 894, 1305
- Immunoreceptor tyrosine-based activation motifs (ITAMs), 2776f, 2777
- Immunoregulatory antibodies, 549, 550f, 558, 559f, 560
- Immunosenescence, 3872
- Immunosuppression
  - by cancer cells, 549, 550f
  - in lung transplantation, 2286
  - in sepsis/septic shock, 2319
  - UV-induced, 432
- Immunosuppressive therapy
  - after liver transplantation, 2721–2722
  - in cardiac transplantation, 2030
  - chronic meningitis associated with, 1135
  - complications of, 755, 2030, 2351
  - COVID-19 disease risk and, 3633
  - dyslipidemia during, 3247
  - immunization during, 1005
  - in kidney transplantation, 2407–2409, 2408t
  - lymphoid malignancies and, 858
  - osteoporosis associated with, 3296
  - skin cancer associated with, 432–433, 606
- Immunotactoid glomerulopathy, 2425
- Immunotherapy
  - for allergic rhinitis, 2810–2811
  - for anaphylaxis prevention, 2814
  - for asthma, 2223
  - for autoimmune and inflammatory diseases, 2783–2784
  - for cancer, 549, 737t, 740. *See also* Cancer biologic therapy
  - contraindications to, 2811
  - oral, for peanut allergy, 2814
- Immunotoxicity, 3825t
- Impaired physician, 73
- Impella device, 2132
- Impetigo, **388**
  - bullous, 389, 401, 1051, 1210
  - clinical features of, 378t, 388–389, 388t, 1210, 1210f
  - differential diagnosis of, 1210
  - etiology of, 388–389, 388t, 1210
  - *S. aureus,* **388,** 388t, 1009–1210, A1, A5
  - streptococcal, 1009–1210, A5
  - treatment of, 388t, 389, 1208t, 1210
- Impetigo contagiosa, 1051
- Impingement syndrome, 2973–2974
- Implantable cardioverter-defibrillator (ICD). *See* Cardiovascular implantable electronic devices (CIEDs)
- Implicit learning, 3930t
- Implicit memory, 207
- Imprinting
  - disorders caused by, 3782–3783, 3791–3792
  - genomic, 3782–3783, 3785f
  - penetrance and, 3789
- Inactivated vaccines, 997t
- Inappropriate antidiuresis. *See* Syndrome of inappropriate antidiuresis (SIAD)
- Incapacitating/behavior-altering agents, S5
- Incidentalomas, adrenal, 3063–3065, 3063t, 3064f
- Incident (movement) pain, 508, 508t
- Incisional biopsy, 552
- Inclisiran, 3228, 3244t
- Inclusion body myopathy with Paget's disease and frontotemporal dementia (IBMPFD), 3312
- Inclusion body myositis (IBM), **2913**
  - clinical features of, 2910t, 2913, 2914f, 3634
  - diagnosis of, 2910t, 2913–2914
  - differential diagnosis of, 2910
  - global considerations in, 2917
  - histopathology of, 2914–2915, 2915f
  - pathogenesis of, 2915
  - prognosis of, 2915
  - treatment of, 2915
- Inclusion cyst, epidermoid, 380f
- Incompetent patient
  - advance directives for, 79
  - assessing capacity to make medical decisions, 71
  - choice and role of surrogates, 71
  - end-of-life care for, 88
- Incomplete penetrance, 3789
- Incontinence
  - fecal. *See* Fecal incontinence
  - urinary. *See* Urinary incontinence
- Incontinentia pigmenti, 399, 3783t
- Incretin mimetics, 2706, 3213
- Indacaterol, 2224, 2255

- Indapamide, 2143, 2144t
- India, 3865–3866
- India ink, S11
- "Indiana pouch," 694
- Indian hedgehog, 3259
- indications for, 1154–1155
- Indigestion, **300**
  - approach to the patient, 301
  - definition of, 297
  - differential diagnosis of, 300, 2490
  - in esophageal disease, 2503
  - in GERD. *See* Gastroesophageal reflux disease (GERD)
  - mechanisms of, 300
  - treatment of, 301–302
- Indinavir
  - adverse effects of, 1601, 2640, 2672
  - drug interactions of, 1742t, 3927
  - metabolism of, 485t
  - molecular structure of, 1619f
  - resistance to, 1623f
- Indirect antiglobulin (Coombs) test, 908, 910f
- Indirect effect, of vaccines, 1195
- Indirect fraction, bilirubin, 2634
- Indomethacin
  - for acute idiopathic pericarditis, 2079
  - adverse effects of, 3235
  - for cough headache, 3473
  - for exercise headache, 3474
  - for gout, 2959
  - for hemicrania continua, 3473
  - for hypnic headache, 3475
  - for medullary thyroid cancer, 312
  - for pain, 97t
  - for paroxysmal hemicrania, 3473
  - for prevention of post-ERCP pancreatitis, 2742
  - for sex headache, 3463
  - for stabbing headache, 3474
- Indoor air pollution, 2241
- Indoor residual spraying, 3849
- Indoor tanning, 432
- Induced pluripotent stem cells, 3402
- Inducible laryngeal obstruction (vocal cord dysfunction), 2222
- Inebilizumab, 227, 3565, 3592, 3592t
- Inert gas dilution, 2206–2207
- Infant botulism, 1235, 1236, 1237, 1238
- Infantile poikilocytosis, 800
- Infants. *See also* Neonates
  - atopic dermatitis in, 383
  - cobalamin deficiency in, 785
  - cradle cap in, 385
  - *C. trachomatis* infections in, 1474
  - hypercalcemia in, 3281
  - microbiome of, 3833
  - nutrient requirements in, 2600–2602t
  - seizures in, 3411, 3413, 3413t
  - water requirement in, 2599
- Infection, 964
- Infection-control programs
  - for *A. baumannii,* 1297f, 1298
  - in critical care setting, 2297
  - for filovirid infections, 1686–1687
  - genomics in, 985
  - organization, responsibilities, and oversight of, 1146
  - for respiratory tract infections, 1538–1539
  - for *S. aureus,* 1205
  - surveillance by, 1146–1147
- Infection susceptibility syndromes, 460, 461
- Infectious (septic) arthritis, **1055**. *See also specific types*
  - approach to the patient, 1055
  - bacterial
    - anaerobic bacterial, 1377
    - *Brucella,* 1333
    - gonococcal, 1057–1058, 1257, 1257f, 1258f, 1259t
    - microbiology of, 1056
    - mycobacterial, 1058
    - nongonococcal, 1056–1057, 1056f
    - nontyphoidal *Salmonella,* 1320
    - pathogenesis of, 1055–1056
    - *S. aureus,* 1199–1200, 1205
  - from bite-wound infections, 1143
  - etiology of, 1055, 1055t
  - fungal, 1058–1059, 1059f
  - in HIV infection, 1604
  - Lyme arthritis, 1058, 2790t
  - parasitic, 1059
  - postinfectious, 1060. *See also* Reactive arthritis
  - prevention of, 1060
  - in prosthetic joints. *See* Prosthetic joint infections
  - in sickle cell disease, 2968
  - syphilitic, 1058
  - viral, 1059
- Infectious diseases, **957**. *See also specific diseases and agents*
  - approach to the patient, **958**. *See also* Acute febrile illness
  - climate change and. *See* Climate change
  - emerging and re-emerging. *See* Emerging infectious diseases
  - empiric antibiotic treatment of, 962, 963–964t
  - epigenetic influences on, 3942
  - eradication of, 3855
  - fatigue in, 166
  - genomics of, **975**
    - in antibiotic resistance detection, 977t, 980–982
    - COVID-19 pandemic and, 988
    - in drug and vaccine development, 983–984
    - epidemic potential and, 987
    - epidemiology and, 977t, 984–987, 986f
    - global considerations in, 988
    - host-based diagnostics and, 982–983
    - limitations and progress of, 975–977
    - in microbial diagnostics, 975, 977t
    - in organism identification, 977t, 978–979
    - in pathogen discovery, 977t, 979–980, 980f
    - in pathogen evolution, 987–988
    - public health implications of, 977
    - in therapeutics, 983–984
  - global considerations in, 957, 958f
  - historical perspective, 957
  - history in, 958–959
  - imaging in, 961–962
  - immune status in, 959
  - infection control for, 962
  - laboratory diagnosis of, 960–961, **S11**
  - microbiome and, 3834, 3835f, 3838–3839
  - perspective on, 964
  - physical examination in, 959–960
  - sexually transmitted. *See* Sexually transmitted infections (STIs)
  - skin signs of, 960, A5
  - specialist consultation for, 962
  - vaccine-preventable, 1002, 1002t
  - in war veterans, S7
- Infectious mononucleosis
  - clinical features of, 138t, 166, 257t, 258–259, 1509, 1509t, 1536
  - CMV, 1513–1514
  - complications of, 810, 1509–1510
  - diagnosis of, 1509, 1509f, 1511, 1511f
  - differential diagnosis of, 1511, 1511t, 1801t
  - EBV, 1509–1510, 1509f, 1509t
  - lymphadenopathy in, 477
  - oral manifestations of, 262t
  - rash in, 137, 138t, 402, 1509, 1509f
  - splenomegaly in, 480
  - treatment of, 1522
- Infective endarteritis, 1036
- Infective endocarditis, **1036**
  - *Abiotrophia,* 1041
  - abscesses in, 1038
  - *Actinomyces,* 1363
  - acute, 1037
  - anaerobic bacterial, 1378
  - aortic regurgitation and, 2043t
  - *Aspergillus,* 1714
  - AV conduction block in, 1927
  - *Bartonella,* 1038, 1040, 1044t, 1045, 1351t, 1353t, 1354
  - *Brucella,* 1334–1335
  - *Candida,* 1037t, 1044, 1712
  - as cause of fever of unknown origin, 150
  - *C. burnetii,* 1038, 1040, 1044t, 1048, 1463, 1464
  - clinical features of, 146, 995, 1038–1039, 1038t, 1039f, 1860
  - clubbing/hypertrophic osteoarthropathy in, 279
  - complications of, 1047
  - *Corynebacterium,* 1039
  - culture-negative, 150, 1037–1038, 1037t
  - diagnosis of
    - blood cultures, 1039–1040
    - cardiac imaging, 1040–1041
    - CT-PET, 1899, 1901f
    - Duke criteria, 1039, 1040t
    - echocardiography, 1040–1041, 1041t, 1042f, 1898–1899, 1901f
    - laboratory studies, 1038t, 1040
  - embolic events in, 1039, 1039f, 1045–1046
  - *Enterobacter,* 1041t
  - enterococcal
    - clinical features of, 1217–1218
    - complications of, 1218
    - echocardiography in, 1041, 1041t
    - incidence of, 1037t
    - recurrent, 1218
    - treatment of, 1041–1042, 1043t, 1219t, 1220t
  - epidemiology of, 1037, 2035
  - etiology of, 991t, 995, 1037–1038, 1037t
  - *Gemella morbillorum,* 1041, 1043t
  - glomerulonephritis associated with, 2415t, 2417
  - *Granulicatella,* 1041
  - HACEK group
    - clinical features of, 1038, 1266–1267
    - diagnosis of, 1039
    - epidemiology of, 1266–1267
    - etiology of, 1037, 1266–1267
    - incidence of, 1037t
    - pathogenesis of, 1266–1267
    - treatment of, 1044t, 1045, 1266–1267, 1266t
  - health care–associated, 1039, 1200
  - heart failure in, 1038, 1046–1047
  - IgM antibodies in, 2779
  - murmur in, 284
  - mycotic aneurysm in, 1039, 1047
  - during neutropenia, 581
  - *Nocardia,* 1360
  - nontyphoidal *Salmonella,* 1321, 1321t
  - osteomyelitis and, 1062
  - outcome of, 1047–1048
  - *P. aeruginosa,* 1044, 1309t, 1310
  - paravalvular infection in, 1038, 1047
  - pathogenesis of, 1038
  - in persons who inject drugs, 1037, 1039, 1047, 1200
  - pneumococcal, 1037t, 1044
  - polymicrobial, 1037t
  - prevention of, 1048, 1048t
    - antibiotic prophylaxis for dental care, 266, 1048, 1048t
    - prophylaxis for dental care, 1180t
    - in surgical patient, 3909
  - in prosthetic valve. *See* Prosthetic valve endocarditis (PVE)

Infective endocarditis (*Cont.*):
- retinal hemorrhage in, 225
- septic emboli in, 1039, 1039f, 3414, A1, A5
- skin manifestations of, 141t, 146, 995
- splenomegaly in, 479
- staphylococcal
  - clinical features of, 141t, 995, 1038, 1039, 1039f, 1200, A1
  - diagnosis of, 1200
  - echocardiography in, 1041, 1041t
  - emboli in, 1039, 1039f, 1200, 1200f
  - epidemiology of, 1200
  - etiology of, 1037
  - incidence of, 1037t, 1200
  - pathogenesis of, 1038
  - treatment of, 1042, 1043t, 1044, 1047, 1205
- streptococcal
  - clinical features of, 137t, 1038, A5
  - diagnosis of, 1041t
  - group C or D, 1212–1213
  - group D, 1214
  - incidence of, 1037t
  - treatment of, 1041, 1043t
- subacute, 1037
- in transcatheter implanted aortic valve. *See* Transcatheter aortic valve implantation (TAVI)
- treatment of
  - antimicrobial therapy, 991t, 1041–1045, 1043–1044t, 1219t
  - antithrombotic therapy, 1045–1046
  - empirical, 963t, 1044–1045
  - monitoring, 1045
  - outpatient, 1045
  - partial oral antibiotic, 1045
  - surgical, 1046–1047, 1046t
- *T. whipplei,* 1038, 1040, 1368–1369, 1370

Inferior petrosal venous sampling, 3011–3012

Inferior vena cava diameter, in shock, 2314

Inferior vena cava filter, 742, 2132, 2163

Infertility, **3149**
- after Hodgkin's lymphoma, 871
- in CAH, 3101
- chemotherapy-related, 576, 706, 756, 871
- in cystic fibrosis, 2245
- definition of, 3149
- diagnostic evaluation of, 2979t, 3149–3150
- etiology of, 3149, 3150f
- in Klinefelter syndrome, 3096, 3097
- male, 3107, 3150, 3150f, 3151, 3171, 3677
- ovulatory dysfunction/failure in, 3134–3135, 3151, 3789
- pelvic inflammatory disease and, 1107
- preconception counseling, 3150–3151
- prevalence of, 2979t, 3149
- psychological aspects of, 3152
- radiation therapy-related, 756
- STIs causing, 1097t
- treatment of, 3151–3152, 3151t
- tubal disease and, 3151
- in Turner's syndrome, 3097

Infestations
- delusional, 3737f, 3738
- fly, 3732–3733
- leech, 3733
- louse, 3731–3732
- maggot, 3732–3733
- scabies. *See* Scabies
- tongue worms, 3733

Infigratinib, 533t, 563t, 568t, 3318

Inflammaging, 3872

Inflammasome(s)
- characteristics of, 2755, 2760
- diseases associated with, 2936–2937
- functions of, 972
- mutations associated with clinical disease, 2761–2762t
- pathogen avoidance of, 972–973, 972f

Inflammation
- aging and, 3872
- anemia and, 767–769, 767t
- in asthma, 2218, 2219f
- atherosclerosis and, A10
- cardinal signs of, 454f
- endothelium in, 1845
- heart failure and, 1982
- in IC/BPS, 331
- immune cells at sites of, 2782f
- laboratory identification of, S11
- in malnutrition, 2621
- nociceptor-induced, 94
- in obesity, 545, 3186
- platelets and thrombosis in, 938
- in psychiatric disorders, 3656
- in sepsis/septic shock, 2317, 2318f
- thrombosis and, A10
- in vivo assessment of, 463

Inflammation of unknown origin (IUO), 149

Inflammatory bowel disease (IBD), **2551**. *See also* Crohn's disease; Ulcerative colitis
- after infectious diarrhea, 1079t
- atypical colitides mimicking, 2561
- cancer risk in, 2570
- clinical features of, 309, 354
- colorectal cancer in, 655, 2502t, 2570, 2570f
- COVID-19 disease and, 2570
- differential diagnosis of, 1306, 2559–2561, 2560t
- disorders associated with, 2553t
- dysplasia in, 2570, 2570f
- early-onset, 2551, 2553t
- epidemiology of, 2551–2552, 2551t
- etiology of, 2552f, 2553
- extraintestinal manifestations of, 2562t
  - cardiopulmonary, 2564
  - dermatologic, 409, 2561, 2563
  - hepatobiliary, 2563
  - metabolic bone disorders, 2563
  - neuropathy, 3605
  - ocular, 225, 2563
  - oral, 261
  - rheumatologic, 2563
  - spondyloarthritis, 2881f, 2891–2892
  - thromboembolic disorders, 2460
  - urologic, 2460
- familial patterns of, 2552
- genetic considerations in, 2552f, 2553–2554, 2553–2554t
- global considerations in, 2551, 2552–2553
- gut microbiome in, 2552f, 2554, 2576–2577, 2577f, 3836
- immune regulation in, 2553–2554
- inflammasome mutations in, 2761–2762t
- inflammatory cascade in, 2555
- noninfectious diseases mimicking, 2561
- pathogenesis of, 2552f, 2553, 3836
- in pregnancy, 2569–2570, 3903
- racial and ethnic differences in, 2551
- serologic markers in, 2559
- smoking and, 2552
- treatment of, **2564,** 2569f
  - antibiotics, 2566
  - anti-integrins, 2566
  - anti-TNF therapies, 2565–2566
  - 5-ASA agents, 2564, 2564t
  - azathioprine, 2565
  - biologic therapies, 2565–2566, 2568t
  - cyclosporine, 2565
  - glucocorticoids, 2565–2566
  - hyperbaric oxygen therapy for, 3764
  - 6-mercaptopurine, 2565
  - methotrexate, 2565
  - nutritional therapies, 2567
  - during pregnancy, 2569–2570
  - small molecules, 2566–2567
  - surgery, 2567–2569, 2569t
  - tacrolimus, 2565
- tumor markers in, 504t

Inflammatory diarrhea, 309. *See also* Diarrhea

Inflammatory mediators, in sensitization, 93

Inflammatory myopathies, **2909**. *See also* Dermatomyositis; Inclusion body myositis; Polymyositis
- autoantibodies in, 2825, 2825t
- diagnosis of, 2909–2911
- differential diagnosis of, 2909–2911, 2910t
- drug-induced, 2917, 3647–3648
- epidemiology of, 2909
- global considerations in, 2917
- overlap syndromes, 2912
- treatment of, 2914–2917, 2916t

Inflammatory response, 2555

Infliximab
- adverse effects of
  - common, 2568t, 2850t
  - cutaneous, 419
  - desensitization protocol for, 2819
  - infections, 462, 2566
  - lupus syndrome, 2835
  - in pregnancy, 2570
  - septic arthritis, 1056
  - serious, 387t, 2566, 2568t, 2850t
- antibodies to, 2566
- for autoimmune and inflammatory diseases, 2784
- for axial spondyloarthritis, 2887t
- for Behçet syndrome, 2909
- for COVID-19 disease, 1542
- for IBD, 2565–2566, 2568t
- monitoring during treatment with, 2568t, 2850t
- for psoriasis/psoriatic arthritis, 387t, 2890
- for rheumatoid arthritis, 2850t, 2851
- for sarcoidosis, 2929t, 2930

Influenza, **1544**
- clinical features of, 995, 1047, 1533
- complications of, 1533, 1547–1548
- COVID-19 and, 1544, 1547f
- definition of, 1544
- diagnosis of, 984, 1548
- differential diagnosis of, 253, 1548
- epidemiology of, 987, 995, 1544–1546
- etiology of, 1544
- global considerations in, 1544
- in HCT recipient, 1157
- health care facility preparedness for, 1149
- immune response in, 1546
- monocyte dysfunction in, 461
- morbidity and mortality from, 75t, 1545f, 1546, 1546f, 1547f
- myocarditis in, 2012
- pandemics, 1533, 1544–1545, 1545f, 1545t, 3856–3857
- pathogenesis of, 1436
- pathogen evolution of, 1544
- pneumonia and, 1547–1548
- in pregnancy, 1547, 1549, 3903
- prevention of, 1487–1488, 1488t. *See also* Influenza vaccine
- risk factors for, 1546t
- treatment of, 1157, 1487–1488, 1488t, 1537, 1549–1550

Influenza A viruses, 1544–1545, 1544f, 1545t, 3856

Influenza B viruses, 143t, 1545–1546

Influenza C viruses, 1545–1546

Influenza H1N1. *See* H1N1 influenza

Influenza H2N2, 1533

Influenza H3N2, 987

Influenza H5N1, 987, 995, 1533
Influenza H7N9, 995
Influenza vaccine
  adverse effects of, 1549
  categories of, 1538, 1548–1549, 1548t
  contraindications to, 1006t, 1549
  in COPD, 2258
  Guillain-Barré syndrome and, 3617
  inactivated, 1004t, 1006t, 1549
  live attenuated, 1004t, 1006t, 1007, 1549
  microbiome and response to, 3839
  performance of, 1549
  precautions for, 1006t, 1549
  recommended administration of
    in adults, 40t, 1004t, 1548t, 1549
    in children, 1548t, 1549
    in health care workers, 1153t
    in HIV infection, 1594t
    in international travelers, 1014–1015, 1014t
    in liver disease, 2633
    in older adults, 1005
    in pregnancy, 1005, 3903
Influenza viruses
  adaptation to modern world, 3857
  antigenic drift in, 1544
  antigenic shift in, 1544
  epigenetics of, 3942
  genetic changes in, 3856–3857
  pathogen evolution of, 987, 1533
  replication strategies of, 1479
  subtypes of, 1545t
  virology of, 1544
Information technology, 1, 5
Information theory of aging, 3869
Informed consent, 5, 70–71, 3795, 3803, 3922
Infradian rhythm, 3944t
Ingenol mebutate, 609
Inguinal hernia, 330
Inguinal lymph node, 477
Inguinal syndrome, 1472
INH. *See* Isoniazid (INH)
*inh*A gene, 1397
Inhalational anthrax, 995, S4
inherited tubule growth and development disorders. *See* Tuberous sclerosis
Inhibin, 2991, 3107, 3128
Inhibitor of apoptosis proteins (IAP), 539, 540f
Inhibitory NK cell receptors (iNKRs), in HIV infection, 1580
Injection drug users. *See* Persons who inject drugs (PWID)
INK4/ARF system, 3871
Inkoo virus, 1658t
Innate immune system, **2756**. *See also* Immune response/immune system
  adaptive immune system and, 2756
  autoimmunity and, 2823–2824
  components of, 2756, 2758t
  definition of, 2755
  effector cells of, **2760,** 2763t
    basophils. *See* Basophil(s)
    dendritic cells. *See* Dendritic cells
    eosinophils. *See* Eosinophil(s)
    macrophages. *See* Macrophage(s)
    monocytes. *See* Monocyte(s)
    neutrophils. *See* Neutrophil(s)
    NK cells. *See* Natural killer (NK) cells
  epigenetic control of, 973–974
  pathogen avoidance of. *See* Bacterial infections, pathogenesis of
  pattern recognition receptors in, 2756, 2758, 2759t, 2760
  primary deficiency of. *See* Primary immune deficiency diseases, of innate immune system
  response to viruses, 1482
Innate lymphoid cells (ILCs)
  in adaptive immunity, 2763t
  in asthma, 2218
  definition of, 2756
  development and function of, 2763, 2770–2771f
  in innate immunity, 2763t
Inorganic dusts, 2236t, 2238
Inotersen, 900
Inotuzumab ozogamicin, 566, 566t, 848t, 849
Inoue balloon technique, mitral commissurotomy, 2051f
Inpatient care, 3
Inpatient deaths, 75–76, 75f
INR (International Normalized Ratio), 474, 951
Insect bites/stings, 378t, 405, 410, **3735**. *See also specific insects*
Insecticides, 489t, 3712t
Insecticide-treated bed nets, 3849
Insertional metagenesis, 3826
Insertions/deletions, 491, 522, 3785, E72t
Insomnia
  assessment of, 86
  drug-induced, 215
  etiology of, 86
  evaluation of, 215–216
  fatal familial, 216
  in medical disorders, 215
  in neurologic disorders, 215–216
  psychiatric disorders and, 211, 215
  psychophysiologic factors in, 215
  sleep hygiene and, 215
  in terminally ill patient, 86
  treatment of, 86, 211f, 216–217, 216t
Institute for Healthcare Improvement, 53
Institute of Medicine. *See* National Academy of Medicine
Insulin. *See also* Insulin therapy
  actions of, 2986, 3199, 3231, 3231f, 3232t
  biosynthesis of, 3198
  ectopic production of, 737t
  effectors of, 2981t
  secretion of, 3198–3199, 3199f, 3202
  signaling pathways of, 2981t
Insulin aspart, 3209t, 3210
Insulin autoimmune syndrome (Hirata's syndrome), 2790t, 3093
Insulin-dependent diabetes mellitus (IDDM). *See* Diabetes mellitus, type 1
Insulin detemir, 3209t, 3210
Insulin glargine, 3209t, 3210
Insulin glulisine, 3209t, 3210
Insulin infusion devices, 3208f, 3212
Insulin-like growth factor 1 (IGF-1) receptor, 2791t
Insulin-like growth factor 2 gene, 3792
Insulin-like growth factor I (IGF-I)
  in acromegaly, 3008
  actions of, 2990
  adverse effects of, 2990
  factors affecting levels of, 2990
  in gastroduodenal mucosal defense and repair, 2516f
  hypoglycemia and, 3235
  injected, 2990
  in obesity, 544
  in puberty, 3102
  signaling pathway, 660, 2981t, 3871, 3873f
  in thyroid hormone synthesis, 3023
Insulin-like growth factor II (IGF-II), 727, 737t, 739–740, 2990
Insulin-like growth factor II gene, 739–740
Insulin lispro, 3209t, 3210
Insulinoma
  clinical features of, 678, 678t, 3082
  diagnosis of, 3082, 3235
  endogenous hyperinsulinism due to, 3235
  incidence of, 3235
  in MEN 1, 3082
  screening tests for, 3081t
  treatment of, 678, 678t, 3082, 3235
  weight gain in, 3185
Insulin receptor, 2790t, 2984
Insulin receptor substrates, 3199
Insulin resistance
  autoimmune, 2790t, 3092–3093
  colorectal cancer and, 654, 655
  dyslipidemia and, 3246, 3248
  free fatty acids in, 3254
  in metabolic syndrome. *See* Metabolic syndrome
  obesity and, 3186
  in pregnancy, 3901
  syndromes of, 3202
  treatment of, 3258
Insulin secretagogues
  incretins, 3211t, 3213, 3215
  sulfonylureas. *See* Sulfonylureas
Insulin therapy
  adverse effects of, 3188
  for DKA, 3218, S1
  for hyperglycemic hyperosmolar state, 3219
  for hyperkalemia, 361
  hypokalemia and, 354
  intensive, 3208f
  in older adults, 3882t
  in parenteral nutrition, 2625
  in pregnancy, 3901
  for type 1 DM, 3209t, 3210, 3211t, 3212
  for type 2 DM, 3211t, 3214, 3215–3216
Insulin tolerance test, 2995t, 3051
Insulitis, 3200
Integrase, 1558
Integrase inhibitors. *See* Combination antiretroviral therapy (cART)
Integrative health care, 3924. *See also* Complementary and integrative therapies and practices
Integrative medicine, 127
Integrin(s)
  alpha-4, 2789, 2792t
  of innate immune system, 2758
  in neutrophil adhesion, 454
  stem cell, 759, 914
  in tumor angiogenesis, 545f, 546, 547f
Integrin α4β7, 1577–1578
Intellectual disability, 3374t
Intensivists, 3, 2291
Interactome, 32, 33f, 3960–3961f, 3961t
Intercalated cells, 2367f, 2370
Intercellular adhesion molecule 1 (ICAM-1), 1585
Interdigitating dendritic cell sarcoma (reticulum cell sarcoma), 882
Interface hepatitis, 2651, 2695
Interferon(s)
  adverse effects of, 394, 421, 1491, 2676
  genetic variations in response to, E72t
  pharmacology of, 1491–1492
  structure of, 2773
  in viral infections, 1491–1492
Interferon-α (IFN-α)
  actions and targets of, 561
  adverse effects of, 561, 2676
    cardiotoxicity, 2017
    cutaneous, 420
    depression, 84
    thyroid dysfunction, 3039
  for anogenital warts, 1489
  for ATL, 1555–1556, 1555f
  for chronic HBV infection, 2675–2676
  for chronic HCV infection, 2686
  for chronic HDV infection, 1494, 2684
  for CML, 842

Interferon-α (IFN-α) (*Cont.*):
  for genital warts, 1530, 1531t
  for hairy cell leukemia, 873
  for hypereosinophilic syndromes, 462, 880
  for Kaposi's sarcoma, 1612t, 1613
  for mastocytosis, 881
  for NETs, 686
  for polycythemia vera, 820
  production of, 2762
  source, target, and biologic activity of, 2766t
  for systemic mastocytosis, 2821
Interferon-α2a (IFN-α2a). *See* Pegylated interferon-α2a (PEG IFN-α2a)
Interferon-α2b (IFN-α2b). *See* Pegylated interferon-α2b (PEG IFN-α2b)
Interferon-β (IFN-β), 2766t, 3585, 3587t
Interferon-β1a (IFN-β1a), 3587t
Interferon-β1b (IFN-β1b), 3587t
Interferon-γ (IFN-γ)
  in anemia of inflammation, 767–768, 768f
  in *Brucella* infections, 1332
  for chronic granulomatous disease, 463, 2784
  functions of, 461
  in fungal infection susceptibility, 1689–1690, 1690f
  in hemophagocytic lymphohistiocytosis, 465, 468
  in HIV infection, 461
  indications for, 463
  in NTM infections, 1415, 1415f
  source, target, and biologic activity of, 2766t
  in T cell lymphoma, 461
Interferon-γ (IFN-γ) release assays (IGRAs), 149, 1384, 1394, 1417
Interleukin(s), 2765–2766t
Interleukin-1 (IL-1), 2792t
  in familial Mediterranean fever, 462
  in fever induction, 135, 135f, 155
  source, target, and biologic activity of, 2765t
  in vasculitis syndromes, 2894
Interleukin-1 receptor–associated kinase (IRAK-4), 1191
Interleukin-1β (IL-1β)
  in anemia of inflammation, 767–768
  in gout, 2957
  in heart failure, 1980
  in HIV infection, 1580
  in osteoarthritis, 2950, 2950f
Interleukin-2 (IL-2), 2765t
Interleukin-2 (IL-2) receptor, 2407
Interleukin-2 (IL-2) receptor antagonist, 2286
Interleukin-2 (IL-2) therapy
  actions and targets of, 561
  adverse effects of, 84, 282t, 420, 561, 754
  for melanoma, 561, 604t
  for RCC, 561
Interleukin-3 (IL-3), 2218, 2765t
Interleukin-4 (IL-4), 432, 2765t
Interleukin-4 (IL-4) receptor alpha subunit, 2792t
Interleukin-5 (IL-5), 462, 741, 2218, 2765t, 2793t
Interleukin-6 (IL-6)
  adverse effects of, 754
  in blastomycosis risk, 1700
  in depression, 3656
  excess of, 2792t, 3872
  in fever induction, 135, 135f
  in HIV infection, 1582t, 1591t
  in multiple myeloma, 884
  in sepsis/septic shock, 2317
  source, target, and biologic activity of, 2765t
Interleukin-6 (IL-6) inhibitors, 754
Interleukin-6 (IL-6) receptor, 2792t
Interleukin-7 (IL-7), 1579, 2765t
Interleukin-7A (IL-7RA), 1582t
Interleukin-8 (IL-8), 455, 2251, 2765t
Interleukin-9 (IL-9), 2765t
Interleukin-10 (IL-10), 432, 1582t, 2765t
Interleukin-11 (IL-11), 2765t
Interleukin-12 (IL-12), 461, 1415, 1415f, 1689–1690, 2765t
Interleukin-12B (IL-12B), 2266
Interleukin-13 (IL-13), 2765t
Interleukin-14 (IL-14), 2765t
Interleukin-15 (IL-15), 2765t
Interleukin-16 (IL-16), 2765t
Interleukin-17 (IL-17)
  in fungal infection restriction, 1689
  in psoriasis, 2789
  in rheumatoid arthritis, 2845f, 2846
  source, target, and biologic activity of, 2765t
  in spondyloarthritis, 2881f, 2882
Interleukin-17A (IL-17A), 2793t
Interleukin-18 (IL-18), 2765t
Interleukin-21 (IL-21), 2765t
Interleukin-22 (IL-22), 2765t
Interleukin-23 (IL-23), 2765t, 2793t
Interleukin-24 (IL-24), 2765t
Interleukin-25 (IL-25, IL-17E), 2766t
Interleukin-26 (IL-26), 2766t
Intermediate-density lipoproteins (IDLs), 3237, 3237f
Intermediate metabolizers, 491
Internal carotid artery, 3430–3431, 3431f, 3432f, A14
Internalities, in behavior, 3915
International Certificate of Vaccine Prophylaxis (ICVP), 1012
International Health Regulations, 3843
International Normalized Ratio (INR), 474, 951
International Prognostic Index, for non-, 861
International Prognostic Index, for non-Hodgkin's lymphoma, 861t
International Sensitivity Index (ISI), 474, 951, 2636
International travelers, **1010**
  activity-specific infection risks in, 1019
  epidemiology of medical problems in, 1011
  gastrointestinal illness in. *See* Travelers' diarrhea
  to high altitudes, 1019–1020, 3759–3760. *See also* High-altitude illness
  immunizations for, 1011–1015, 1014t, 1318
  infectious disease exposure in, 959
  infectious disease spread and, 957, 1152
  injuries in, 1020
  jet lag in, 219, 1013t, 1020
  leptospirosis in, 1441
  malaria prevention for, 1773–1775, 1774t
  online resources for, 1013t
  outbreaks and emerging infectious diseases, 1022
  posttravel care, 1022
  with preexisting medical conditions, 1013t, 1020–1021
  pre-travel consultation for, 1011, 1012t
  pre-travel preparations, 1021
  prevention of arthropod-borne infections in, 1018
  special populations, 1021, 1022t
  venous thromboembolism in, 1019
Internet resources
  on complementary and integrative therapies, 3929t
  on dietary supplement–drug interactions, 3927t
  on genomics and genetic disorders, 3778t
  on HIV disease, 1616t
  for medical information, 7
  vaccine hesitancy and, 16
Internuclear ophthalmoplegia, 236, 236f, V3
Interobserver agreement, 47–48
Intersex, 3179t. *See also* Disorders of sex development (DSDs)
Interstitial cystitis/bladder pain syndrome (IC/BPS), **330**
  approach to the patient, 332, 332t
  clinical features of, 332
  complications of, 336
  definition of, 330
  diagnosis of, 333–334, 333t
  epidemiology of, 331–332
  etiology and pathogenesis of, 330–331
  global considerations in, 336–337
  prognosis of, 336
  treatment of, 334–336, 335f
Interstitial fluid, 279
Interstitial lung diseases (ILDs), **2259**
  acute exacerbations of, 2265
  classification of, 2259–2260, 2261f, 2262
  in dermatomyositis/polymyositis, 2266
  diagnosis of, 2260–2262
  dyspnea in, 269t, 270, 2261
  eosinophilic. *See* Pulmonary infiltrates with eosinophilia
  exercise intolerance in, 291
  genetic considerations in, 2266
  global considerations in, 2266
  granulomatous, 2266. *See also* Eosinophilic granulomatosis with polyangiitis; Granulomatosis with polyangiitis
  in hypersensitivity pneumonitis. *See* Hypersensitivity pneumonitis
  idiopathic pulmonary fibrosis. *See* Idiopathic pulmonary fibrosis
  imaging in, A12
  molecular targeted therapy and, 595
  pneumonia. *See* Interstitial pneumonia
  pneumonitis, A12
  pulmonary hypertension in, 2191, 2192
  pulmonary lymphangioleiomyomatosis, A12
  in rheumatoid arthritis, 2265–2266, 2840
  in sarcoidosis, 2260t, 2263f, 2266. *See also* Sarcoidosis
  smoking-related, 2260t, 2263–2264
  in systemic sclerosis, 2260t, 2265, 2868–2869, 2869f, 2873f, 2874. *See also* Systemic sclerosis
Interstitial nephritis. *See* Acute interstitial nephritis (AIN)
Interstitial pancreatitis, 2744, 2746, 2746t
Interstitial pneumonia
  in HCT recipient, 916
  in HIV infection, 1597–1598
  lymphoid, 1597–1598
  non-specific (NSIP), 1597–1598, 2260t, 2263, 2263f, 2264f
Interstitial pneumonitis, A12
Intertrigo, *Candida,* 1708
Interventional cardiology, 2126, 2132. *See also* Percutaneous coronary interventions (PCI)
Interventional nephrology, **2402**
Interventional neuroradiology, 3393
Interventional pulmonary medicine, **2287**
  for asthma, 2289
  for central airway obstruction, 2288, 2289f, 2289t
  for COPD, 2289–2290
  for early-stage lung cancer, 2289
  endobronchial ultrasound-guided transbronchial needle aspiration, 2209–2210, 2209f, 2288, 2288f
  lung volume reduction, 2290
  peripheral bronchoscopy, 2288
  for pleural disease, 2290
  for pneumothorax and persistent air leak, 2290
Intervertebral disks, 119, 119f, 120, 123t, 124t
Intestinal flukes, 1826t, 1830t, 1831, 1831t, S12

Intestinal ischemia, **2586**
  acute, 2586–2588, 2588t
  anatomy of, 2587, 2587f
  arterio-occlusive, 2587, 2588t
  chronic, 2587
  clinical features of, 112t, 2587–2588
  diagnosis of, 2587
  epidemiology of, 2586–2587
  incidence of, 2586
  mesenteric venous thrombosis, 2587, 2588t, 2589
  nonocclusive, 2587, 2588t, 2589
  pathophysiology of, 2587
  treatment of, 2587–2589, 2588t
Intestinal lymphangiectasia
  clinical features of, 309, 2183
  congenital, 309
  in protein-losing enteropathy, 2548
  small-intestinal mucosal biopsies in, 2547f
Intestinal necrosis, 361
Intestinal obstruction, **2589**
  in cancer, 587
  clinical features of, 587, 2591
    abdominal pain, 110, 112t, 2591
    abdominal swelling, 327, 2591
    lactic acidosis, 369
    nausea and vomiting, 297, 2591
  closed-loop, 2590
  in cystic fibrosis, 2245
  diagnosis of, 113, 2487, 2494f, 2591–2592, 2592f
  differential diagnosis of, 110, 112
  epidemiology of, 2589–2590, 2590t
  etiology of, 2589–2590, 2590t
  pathophysiology of, 2590, 2591f
  physical examination in, 2591
  postoperative, 2593
  in systemic sclerosis, 2870, 2874
  treatment of, 587, 2487, 2494f, 2593, V5
Intestinal pseudoobstruction (Ogilvie's disease)
  clinical features of, 297, 354, 2487
  diagnosis of, 2487, 2493f
  etiology of, 2590
  treatment of, 299, 299t, 2487, 2489, 2493f, V5
Intestinal resection, 308, 2547
Intestinal stagnant loop syndrome, 786
Intima, 1845
Intimal dissection, 2127, 2127f
Intimate partner violence. *See* Domestic abuse/violence
Intraabdominal abscess/infections
  actinomycosis, 1364, 1364f
  anaerobic bacterial, 1376, 1379
  C. septicum, 1376
  enterococcal, 1218
  etiology of, 1286
  intraperitoneal, 1072–1074, 1073f
  liver. *See* Liver abscess
  neutropenic enterocolitis, 582, 597, 597f
  nontyphoidal *Salmonella,* 1320
  perinephric, 1075
  peritonitis. *See* Peritonitis
  pleural effusion in, 2270
  psoas, 1075
  renal, 1075–1076
  in solid organ transplant recipient, 1159t
  splenic. *See* Splenic abscess
  treatment of, 963t, 1379, 2321t
Intraaortic balloon pump (IABP), 2132, 2328, 2331, A11
Intraarticular injections, 2955, 2955t
Intracerebral hemorrhage, **3452**
  aphasia in, 201
  brain tumor and, 3455
  clinical features of, 3454
  coma in, 192
  diagnosis of, 3453, 3453f, 3455
  emergency management of, 3453
  etiology of, 3428, 3453t, 3454–3455
  headache in, 114
  hypertension and, 3453–3454
  lobar, 3455
  ocular manifestations of, 230
  prevention of, 3456
  prognosis of, 3455t
  in traumatic brain injury, 3455
  treatment of, 3455–3456
  vascular malformations and, 3456–3457
Intracranial atherosclerosis, 3445, 3452
Intracranial hypotension, 3483, A16
Intracranial infections, 991t, 994
Intracranial meningitis, 1128
Intracranial pressure (ICP), increased
  in bacterial meningitis, 1120, 1123
  from brain metastases, 590
  in chronic meningitis, 1073, 1128
  in coma, 189
  idiopathic (pseudotumor cerebri), 117, 229, 3031
  in intracerebral hemorrhage, 3456
  monitoring of, 2345, 2345f
  nausea and vomiting in, 298
  ocular manifestations of, 228–229, 228f, 235
  pathophysiology of, 2343
  signs of, 2345
  in subarachnoid hemorrhage, 3460
  treatment of, 590, 2345, 2345t, 3460–3461
Intracytoplasmic sperm injection (ICSI), 3107, 3151, 3151t
Intradiscal injections, 127
Intraductal papillary neoplasms, 672
Intrahepatic cholangiocarcinoma, 667–669, 668f, 669f
Intrahepatic cholestasis of pregnancy, 3903
Intraluminal gastric balloons, 3194
Intramedullary syndromes, 3561
Intramuscular injection, of vaccines, 1008, 1008f
Intraocular pressure, increased, 225, 230. *See also* Glaucoma
Intraperitoneal abscess, 1072–1074, 1073f
Intraperitoneal cavity, 1069, 1069f
Intrathecal administration, opioids, 98
Intrauterine devices (IUDs), 1365, 1365f, 3135, 3153–3154, 3153t
Intrauterine growth retardation, 3100
Intrauterine insemination, 3151–3152, 3151t
Intravascular device-related infections. *See* Catheter-related bloodstream infections
Intravascular large B cell lymphoma, 873
Intravascular lithotripsy, 2128
Intravascular ultrasound, 2128
Intravenous immunoglobulin (IVIg)
  adverse effects of, 2916t, 3632
  for antiphospholipid syndrome, 2839
  for autoimmune and inflammatory diseases, 2784
  for CIDP, 3622
  for dermatomyositis, 414
  for DIHS, 424
  for epidermolysis bullosa acquisita, 413
  for Guillain-Barré syndrome, 3620
  for hemophagocytic lymphohistiocytosis, 468
  for hypogammaglobulinemia, 892
  for immune thrombocytopenic purpura, 924
  for inflammatory myopathies, 2916, 2916t
  for Kawasaki's disease, 2907
  for measles prophylaxis, 1642
  for MS, 3587
  for mucous membrane pemphigoid, 413
  for multifocal motor neuropathy with conduction block, 3524
  for myasthenia gravis, 3630, 3631–3632
  for parvovirus B19 infection, 815, 1524
  for pemphigus vulgaris, 410
  for rheumatic fever, 2858
  for streptococcal infections, 1054
  for streptococcal toxic shock syndrome, 1208t, 1212
  for toxic epidermal necrolysis, 1051
  for vaccine-induced immune thrombocytopenia and thrombosis, 924
  for viral meningitis, 1126
Intraventricular hemorrhage, primary, 3455
Intravesical anesthetic bladder challenge, 334
Intravesical therapy, 336, 692–693
Intrinsic factor (IF), 781, 786, 2790t
Intrinsic pathway, coagulation, 470, 470f, 929f
Introns, 3779
Intussusception, 297, 314, 318, 2593
Invariant natural killer (NK) cells, 3840
Invasin, 1348
Inverse psoriasis, 386. *See also* Psoriasis
Investing in Heath, 3844
In vitro fertilization (IVF), 3107, 3151–3152, 3151t, 3797
In vivo genome editing, 3825t
Involuntary active euthanasia, 89t
Involuntary weight loss, **314**
  assessment of, 316, 316t
  in chronic pancreatitis, 2752
  in esophageal cancer, 642
  etiology of, 315–316, 315t
  in HIV infection, 1611
  in liver disease, 2629
  in lung cancer, 618t, 619
  in malabsorption disorders, 2549t
  in older adults, 315
  in pancreatic cancer, 673
  significance of, 315
  in thyrotoxicosis, 3033
  treatment of, 316
Iodides, adverse effects of, 405, 462
Iodine, 2615
  deficiency of. *See* Iodine deficiency
  excess intake of, 3031
  for goiter, 3041
  intake during pregnancy, 3023
  recommended intake of, 2601t, 3023
  tolerable upper intake level of, 2614t
  toxicity of, 2614t
Iodine deficiency
  clinical features of, 2614t, 3022–3023
  global considerations in, 3022–3023, 3022f
  goiter and, 3040–3041
  hypothyroidism and, 3031
  prevention of, 3031
Iodine dermatitis, 1748
^{131}Iodine-ethiodol, S6
^{131}Iodine-MIGB, 3077
Iodoquinol
  actions of, 1748
  adverse effects of, 1743t, 1748
  for amebiasis, 1748, 1757t
  contraindications to, 1748
  indications for, 1743t
  pharmacology of, 1748
  in pregnancy and lactation, 1743t
Ion absorption, congenital defects in, 308
Ion channels, cardiac, 1911–1912, 1912f
Ion exchangers, S6
Ionizing radiation
  AML and, 826
  aplastic anemia and, 809
  from cardiac imaging, 1883
  CML and, 834
  lung cancer and, 615

- Ionizing radiation (*Cont.*):
  - pulmonary fibrosis and, A12
  - skin damage from, 396
- Ion trap mass spectrometry, 3977
- Iopanoic acid, 923t, 3024t
- IP-10 (IFN-γ-inducible protein-10), 2768t
- IPAA (ileal pouch-anal anastomosis), 2567
- Ipecac syrup, 3709, S6
- IPEX (immunodysregulation polyendocrinopathy enteropathy X-linked) syndrome, 2553t, 2789, 2803, 3092
- IPF. *See* Idiopathic pulmonary fibrosis (IPF)
- *IPF1* gene mutations, 3783t
- Ipilimumab
  - action and targets of, 534t, 559f, 560, 2783, 2787t
  - adverse effects of, 560
    - autoimmunity, 593
    - cutaneous, 421
    - hepatotoxicity, 2672
    - long-term, 753t
    - nephrotoxicity, 2377
    - neurologic, 726, 2351
  - for esophageal cancer, 645
  - for HCC, 665f
  - for high-grade NETs, 686
  - indications for, 560, 2783, 2787t
  - for lung cancer, 629, 630t
  - for melanoma, 560, 604, 604t, 605
  - for RCC, 689, 690t
  - resistance to, 574
- Ipodate, 3024t
- *ipr1* gene, 1384
- Ipratropium bromide, 253, 2255, 2810
- Iptacopan, 805f, 806
- Iquitos virus, 1659t, 1667
- IRAK-4 (interleukin-1 receptor–associated kinase), 1191
- IRAK-4 deficiency, 459t
- Iraqi Jewish population, factor XI deficiency in, 932
- *IREB2,* 2253
- *IRGM* gene mutations, 2554t
- Iridectomy, laser, 225
- Iridium exposure, S6
- Iridocyclitis. *See* Uveitis
- Irinotecan
  - actions of, 566
  - adverse effects of, 493, 562t, 564, 597, 658, 754, 2703t
  - for carcinoma of unknown primary, 735
  - for colorectal cancer, 658
  - drug interactions of, 591, 3927
  - for Ewing's sarcoma, 730
  - for gastric cancer, 649
  - genetic variations in response to, 493, E72t
  - in Gilbert syndrome, 2640
  - interactions and issues, 562t
  - for lung cancer, 630
  - metabolism of, 485, 485t
  - for pancreatic cancer, 674, 676t
  - for soft tissue sarcomas, 729
- IRIS. *See* Immune reconstitution inflammatory syndrome (IRIS)
- Iritis. *See* Uveitis
- Irkut virus, 1655
- Iron
  - absorption of, 762–763, 763f
  - accumulation in brain, 3518–3519
  - in bone marrow, A6
  - daily requirements of, 762–763, 765t
  - deficiency of, 65, 217, 763–764, 766t, 2614t, 2625–2626, 3902. *See also* Iron-deficiency anemia
  - dietary sources of, 762–763
  - distribution in body, 762, 763f
  - excessive ingestion of, 3335
  - excess of. *See* Iron overload
  - in hemoglobin molecule, 449
  - homeostasis, pathways of, 3334f
  - macrophage recycling of, 763
  - metabolism of, 762–763, 763f
  - parenchymal, in liver, 3335
  - recommended intake of, 2601t
  - serum, 449. *See also* Ferritin, serum
  - storage of, 763
  - stores of, 3335, 3336t
  - supplements. *See* Iron replacement therapy
  - toxicity of. *See* Iron overload
  - utilization of, 763
- Iron carboxymaltose, 766, 767t
- Iron cycle, 762–763, 763f
- Iron-deficiency anemia, **763**
  - clinical features of, 765
  - diagnosis of, 765–766, 767t
  - etiology of, 764, 766t
  - from gastrointestinal bleeding, 320, 2495–2496
  - gastrointestinal cancer and, 656
  - global considerations in, 764, 767
  - in hookworm infections, 1816
  - in-hospital, 2625–2626
  - iron-refractory, 764
  - jaundice in, 323
  - menorrhagia and, 472
  - pathophysiology of, 449, 764, 764–765f
  - peripheral blood smear in, 437, 438f, 445f
  - prevention of, 767
  - surgery-related, 2534
  - treatment of, 766–767, 767t
- Iron-deficient erythropoiesis, 762
- Iron derisomaltose, 767t
- Iron fumarate, 767t
- Iron gluconate, 767t
- Iron-hematoxylin stain, S12
- Iron overload
  - chronic, 792
  - clinical features of, 2614t
  - CMR in, 1894, 1903v
  - etiology of, 3333t, 3335
  - in hemochromatosis. *See* Hemochromatosis
  - in porphyria cutanea tarda, 435
  - secondary, 3333, 3335
  - transfusion-related, 776, 813, 908t, 912
  - in Wilson's disease. *See* Wilson's disease
- Iron poisoning, 3714t
- Iron replacement therapy, 766–767, 767t, 769
- Iron sucrose, 767t
- Iron sulfate, 766, 767t
- Irregular sleep-wake rhythm, 3948
- Irreversibly sickled cells (ISCs), 772, 773f
- Irritable bowel syndrome (IBS), **2571**
  - approach to the patient, 2573–2574
  - clinical features of, 2571–2572
    - abdominal pain/pressure, 112, 112t, 326, 2571
    - altered bowel habits, 310, 312t, 2571
    - diarrhea, 310
    - gas and flatulence, 2571
    - nausea and vomiting, 297
    - upper gastrointestinal, 2571–2572
  - diagnostic criteria for, 2571, 2571t
  - differential diagnosis of, 311, 2573–2574
  - IC/BPS and, 332
  - incidence of, 302
  - pathophysiology of, 2572–2573, 2572f, 2573f, 2574f
  - postinfectious, 1019, 1079t, 2572
  - psychological features in, 2572
  - severity spectrum of, 2577t
  - treatment of, 312, 2574–2578, 2576t, 2577f, 2577t, 3927
- Irritant contact dermatitis. *See* Contact dermatitis
- Irritant inhalation, 2220, 2221
- Irukandji syndrome, 3722, 3723
- Isaacs' syndrome, 751, 3635
- Isatuximab, 534t, 559f, 890t, 891–892
- Isavuconazole
  - for aspergillosis, 1716, 1717t
  - for coccidioidomycosis, 1698
  - for cryptococcosis, 1706
  - for mucormycosis, 1719, 1721t, 1722
  - pharmacology of, 1692
  - resistance to, 1711t
- Isavuconazonium sulfate, 1702
- Ischemia
  - acute limb, 2175
  - cardiac. *See* Ischemic heart disease
  - cerebral, 3428, 3439, 3439f. *See also* Ischemic stroke
  - intestinal. *See* Intestinal ischemia
- Ischemic cascade, 2342–2343
- Ischemic colitis, 306, 307, 2470f, 2561, 2589
- Ischemic heart disease, **2090**
  - angina in. *See* Angina
  - arrhythmia in, 2092
  - asymptomatic (silent), 2092, 2105–2106
  - cardiomegaly in, 2092
  - cardiomyopathy and, 2092
  - in CKD, 2391
  - diagnosis of
    - approach to, 2095f
    - cardiac imaging, 2094–2097, 2095f
    - CMR, 2095–2096f
    - coronary arteriography, 2097–2098
    - CTA, 2097f
    - ECG, 1872–1873, 1872f, 1873f, 1874t, 2094, 2095f
    - history, 2092–2093
    - laboratory tests, 2094
    - physical examination, 2093–2094
    - stress testing, 2094, 2095–2096f, 2097t
  - epidemiology of, 2090
  - exercise intolerance in, 291
  - functional classification of, 1841–1842, 1842t, 2093t
  - global considerations in, 2090, 2112–2113, 3844
  - heart failure in, 2092, 2103
  - pathogenesis of, 1890–2091, 2090–2091f. *See also* Atherosclerosis; Coronary artery disease (CAD)
  - prognosis of, 2098
  - risk factors for, 2099–2100
  - symptomatic, 2092
  - treatment of, **2098**
    - ACE inhibitors, 2102
    - activity adaptation, 2099, 2099t
    - aggravating conditions, 2099
    - algorithm for, 2103f
    - anticoagulants, S13
    - antiplatelet drugs, 2102, S13
    - ARBs, 2102
    - beta blockers, 2100t, 2101, 2102t
    - CABG. *See* Coronary artery bypass grafting (CABG)
    - calcium channel blockers, 261–262, 2101t, 2102t
    - colchicine, 2103
    - explanation and reassurance, 2098–2099
    - ICD, 2339t
    - ivabradine, 2103
    - lipid-lowering therapy, 2100
    - nicorandil, 2103
    - nitrates, 2100–2101, 2100t, 2102t
    - PCI. *See* Percutaneous coronary interventions (PCI), for revascularization
    - principles, 1843

    ranolazine, 2102, 2102t
    risk factor management, 2099–2100
    SGLT-2 inhibitors, 2102–2103
    trimetazidine, 2103
    in women, 2100
Ischemic penumbra, 3439
Ischemic stroke, **3439**
  after TIA, 3448, 3448t
  clinical features of, 3428
  etiology of, **3443,** 3443f, 3444t
    artery-to-artery embolic, 3445
    CADASIL, 3448
    cardioembolic, 3444–3445, 3445t
    drug-related, 3447
    Fabry's disease, 3355, 3448
    fibromuscular dysplasia, 3447
    giant cell arteritis, 3447
    hypercoagulable disorders, 3446–3447
    leukoaraiosis, 3448
    moyamoya disease, 3447
    necrotizing arteritis, 3447
    posterior reversible encephalopathy syndrome, 3447–3448
    sickle cell anemia, 3447
    small-vessel, 3446, 3447f
    vasculopathy, 3447, 3447f
    venous sinus thrombosis, 3447
  imaging of. *See* Stroke, imaging studies in
  localization of. *See* Stroke, localization of
  pathophysiology of, 3439, 3439f
  prevention of, **3448**
    anticoagulation therapy for, 3450–3451
    antiplatelet agents for, 3445t, 3449–3450
    in atrial fibrillation, 1950–1951, 1950f, 1951t
    risk factor management in, 3448–3449
  treatment of, **3440**
    algorithm for, 3429f, 3440f
    anticoagulants, 3443, S13
    endovascular techniques, 3441–3442
    intravenous thrombolysis, 3441, 3441t
    medical support, 3440–3441
    neuroprotection, 3443
    platelet inhibitors, 3442–3443
    stroke centers and rehabilitation in, 3443
Ischial bursitis, 2973
ISCs (irreversibly sickled cells), 772, 773f
Isfahan virus, 1663t
Ishak fibrosis scale, liver disease, 2632
Ishihara color plates, 222
ISI (International Sensitivity Index), 474, 951, 2636
Islet-1, 1843
Islet amyloid polypeptide (IAPP), 3198
Islet cell(s)
  autoantibodies, 3200
  pancreatic, 3198
  transplantation of, 2754
  tumors of. *See* Neuroendocrine tumors (NETs), pancreatic
Isobutyl-CoA dehydrogenase deficiency, 3369t
Isocarboxazid, 3659t
Isocyanate exposure, 2239t
Isoflavones, 2610
Isoforms, 3779
Isolation aphasia, 201t, 202
Isolation techniques, 1538–1539
Isoleucine metabolism disorders, 3370t
Isometheptene, 3467t
Isomorphic (Koebner) phenomenon, 386, 1612
Isoniazid (INH)
  actions of, 1421–1422
  adverse effects of, 1422
    cutaneous, 400
    fever, 150
    hepatotoxicity
      features of, 2665, 2667t, 2670
      genetic considerations in, 493
      monitoring for, 1403, 1420t, 1422
    lupus syndrome, 2835, 2940t
    neuropathy, 1422, 3609t, 3611
    optic neuropathy, 228
    pellagra, 2608
    pure red cell aplasia, 814t
    pyridoxine for prevention of, 1396, 1422, 2608, 3611, 3612
    in renal failure, 1400
    seizure, 3414t
  dosage of, 1422
  drug interactions of, 1422, 1742t
  genetic variations in response to, 493, E72t
  metabolism of, 485t
  for *M. kansasii* infections, 1418
  overdosage/poisoning with, 3710, 3715t
  pharmacology of, 1402
  resistance to, 1397–1399, 1402, 1422
  for TB
    active infection, 1394–1396, 1394t, 1395t, 1419t, 1420t, 1421–1422
    high-dose, 1399, 1421t
    monitoring and management of, 1420t
    prophylaxis, 1593t
  for TB preventive treatment, 1402–1403, 1402t, 1419, 1419t, 1420t
  for tuberculous meningitis, 1127
Isoprinosine, 1118
Isopropyl alcohol poisoning, 371
Isoproterenol, 1927, 1929
Isosexual precocity, 3108
Isosorbide, 1994, 1995t, 2100t, 2123, 2331
Isospora belli/isosporiasis. *See Cystoisospora belli*/cystoisosporiasis
Isotretinoin
  for acne vulgaris, 391
  adverse effects of, 391, 394, 421, 2940t, 3246
  for oral leukoplakia, 512
  for pityriasis rubra pilaris, 393t
Isotypes, 883
Isovaleric acidemia, 3369t
Isradipine, 2101, 2101t
Issyk-Kul virus, 1661t
Istradefylline, 3503, 3504t
ITAMs (immunoreceptor tyrosine-based activation motifs), 2776f, 2777
Itaquí virus, 1659t, 1667
ITITM5 gene mutations, 3324, 3325t
ITLN1 gene mutations, 2554t
ITPA gene, E72t
Itraconazole
  for ABPA, 2233
  adverse effects of, 355, 1691, 1742t
  for aspergillosis, 1716, 1717t
  for aspergillosis prophylaxis, 1717
  for blastomycosis, 1702, 1703t
  for *Candida* infection prophylaxis, 1717
  for *Candida* infections, 1710t
  for coccidioidomycosis, 1698
  contraindications to, 389
  drug interactions of, 389, 485t, 1742t, 2721, 3159
  for eumycetoma, 1725
  for histoplasmosis, 1127, 1695, 1695t
  indications for, 1691
  for *Malassezia* infections, 1723
  for onychomycosis, 389, 1724, 1724t
  for paracoccidioidomycosis, 1726, 1726t
  for phaeohyphomycoses, 1726t, 1728
  in pregnancy and lactation, 1742t
  prophylactic uses of, 463, 1594t, 1712
  resistance to, 1711t, 1716
  for sporotrichosis, 1724, 1724t
  susceptibility testing for, S11
  for talaromycosis, 1726t
  for tinea infections, 1724, 1724t
IUDs (intrauterine devices), 1365, 1365f, 3153–3154, 3153t
IUO (inflammation of unknown origin), 149
Ivabradine, 1920t, 1936, 1993f, 1995, 2103
Ivacaftor, 2247, E72t
IV drug users. *See* Persons who inject drugs (PWID)
Ivermectin
  for acne rosacea, 391
  actions of, 1748
  adverse effects of, 1743t, 1748–1749, 1822, 1823, 1824
  for ascariasis, 1815
  contraindications to, 1823
  for cutaneous larva migrans, 1812t
  for gnathostomiasis, 1812t
  indications for, 1743t, 1748, 1815t
  for loiasis, 1824
  for lymphatic filariasis, 1821
  for onchocerciasis, 1823
  pharmacology of, 1748
  in pregnancy and lactation, 1743t
  resistance to, 1748
  for scabies, 3729
  for strongyloidiasis, 1817
  for trichuriasis, 1818
IVF (in vitro fertilization), 3107, 3151–3152, 3797
IVIg. *See* Intravenous immunoglobulin (IVIg)
"Ivory vertebra," 3313
Ivosidenib, 534t, 571t, 574, 668, 833, 833t
Ixabepilone, 563t, 565, 641
Ixazomib, 533t, 536, 570t, 573, 890t, 892
Ixekizumab, 387t, 2793t, 2887t, 2890
Ixodes spp. ticks. *See also* Tick(s)
  in *Babesia* infections, 1776. *See also Babesia* spp. infections
  characteristics and life cycle of, 1449
  climate change and habitat of, 3749, 3751f
  in human granulocytotropic anaplasmosis, 1461
  in Lyme borreliosis, 1449, 3731. *See also* Lyme borreliosis (Lyme disease)
  in relapsing fever. *See* Hard tick relapsing fever (HTRF)
  removal of, 3731
Ixodid (hard) ticks, 3730
IYD (dehalogenase 1), 3028t
Izokibep, 2890
Izumi fever, 1349

J

Jaccoud arthropathy, 2829, 2830f
Jackhammer esophagus, 2509f
"Jacksonian march," 3409
Jacobsen's syndrome, S9
Jaffe reaction, S10
JAK (Janus family of protein tyrosine kinases), 2846
JAK (Janus kinases) 1 and 2, 567, 2773–2774
JAK2 gene mutations
  in essential thrombocytosis, 823, 824
  in IBD, 2554t
  in polycythemia vera, 452, 819, 820
  in primary myelofibrosis, 821, 822
  in pulmonary arterial hypertension, 2187
JAK2 inhibitors, 821
JAK3 deficiency, 2774
JAK3 gene mutations, 874, 879
JAK inhibitors
  action and targets of, 532
  adverse effects of, 2886
  for fungal infections, 1690
  for IBD, 2567
  for psoriatic arthritis, 2890, 2891t
  for rheumatoid arthritis, 2851–2852, 2851t
  for spondyloarthritis, 2881f, 2882, 2886

JAK/STAT pathway, 526t
Jamaica bush tea, 324
Jamestown Canyon virus, 1659t, 1673
Janeway lesions, 995, 1038, 1039f, 1860, A1
Jansen’s chondrodystrophy, 3264
Jansen’s disease, 3280, 3280f
Japanese encephalitis virus, 1660t, 1671
Japanese encephalitis virus vaccine, 1014t, 1017, 1671
Japanese spotted fever, 1458
Jargon aphasia, 202
Jarisch-Herxheimer reaction
  in endemic treponematoses, 1439
  in leptospirosis, 1444
  in Lyme disease, 1453
  in relapsing fever, 1446, 1448
  in syphilis, 1435, 1602
  in Whipple’s disease, 1370
Jaundice, **320**
  in acute viral hepatitis, 2656
  approach to the patient, **322,** 322f
  with ascites, 324
  breast milk, 2639
  in cardiovascular disease, 1859
  in cholangiocarcinoma, 667
  in choledocholithiasis, 2735
  in cholestatic disorders, 324–325, 325t, 2732
  in chronic hepatitis, 2675
  in clinical drug trials, 2667
  differential diagnosis of, 320, 2460t, 2461
  drug-induced, 324, 325t
  global considerations in, 326
  in hemolytic anemia, 791
  in hepatocellular conditions, 324, 325t
  history in, 323
  laboratory evaluation of, 321, 324
  in leptospirosis, 1442
  in liver disease, 2628
  neonatal, 796, 798, 2639
  in pancreatic cancer, 673
  pathophysiology of, 320–321
  physical examination in, 323–324, 2629
jaundice following, 325
Jaw, osteonecrosis of, 266, 3285, 3304
J chain, 2779
JC virus infection
  in HIV infection, 1609
  natalizumab treatment and, 3586t
  in PML, 1117, 1161
  respiratory tract, 1536
  in solid organ transplant recipient, 1161
Jefferson fracture, A16
Jejunal vascular ectasia, 2471f, 2501f
Jejunostomy tube feeding, 299
Jellyfish sting, 3722–3723
Jerk nystagmus, 236
Jervell and Lange-Nielsen syndrome, 248, E36t
Jet lag disorder, 219, 1013t, 1020
Jetlag disorder, 3949
Jigger, 3736
Jimson weed, 3711t
Jin Bu Huan, 2672
JMML (juvenile myelomonocytic leukemia), 879
Job syndrome. *See* Hyper-IgE (Job) syndrome
Jod-Basedow effect, 3039, 3041
Joint(s)
  disorders of. *See* Arthritis; Musculoskeletal disorders; Osteoarthritis
  failure of, 2949–2950, 2950f
  malalignment of, 2951, 2951f
  position testing, 173, 174t
  prosthetic, infections in. *See* Prosthetic joint infections
  protective mechanisms of, 2949–2950
Joint Commission, antimicrobial stewardship guidance of, 1179
Joint replacement, 2956
Joint United Nations Program on HIV/AIDS (UNAIDS), 3847
Jones criteria, 2856, 2857t
Joubert’s syndrome, 2431f, 2435
JPH3 gene mutations, 3514
Jugular septic thrombophlebitis. *See* Lemierre’s disease/syndrome
Jugular venous pressure, 1860–1861, 1861f
Jugular venous pulse/distension
  in ascites, 326
  differential diagnosis of, 1986t
  in heart failure, 1983–1984
  in hypovolemia, 347
  in shock, 2293, 2312
Junctional bradycardia, A8
Junctional tachycardia, 1940
Junín/Argentinian hemorrhagic fever, 1658t, 1675
Juquitiba virus, 1661t, 1674
Justice, 71–72
Justinian plague, 3854, 3854t
Juvenile arthritis, 2762t
Juvenile goiter, 3040
Juvenile granulosa cell tumors, 712
Juvenile idiopathic arthritis, 469
Juvenile myelomonocytic leukemia (JMML), 879
Juvenile myoclonic epilepsy, 3411
Juvenile Paget’s disease, 3312
Juvenile pernicious anemia, 786
Juvenile polyposis, 524t, 646, 654t
Juvenile spring eruption, 433
Juxtaglomerular apparatus, 2365f

K
K^+. *See* Potassium
Kabuki syndrome, 3941
KAL1 gene mutations, 3112–3113
Kala-azar, S12. *See also* Visceral leishmaniasis
KAL gene mutations, 2992
Kallikrein inhibitor, 2809
Kallikreins, 697
Kallmann syndrome, 242, 2992–2993, 3111–3112
Kanamycin
  actions of, 1166
  adverse effects of, 1399, 1426
  for MDR-TB, 1421
  in pregnancy, 1169t
Kaolin-pectin, 1741t
Kaposi’s sarcoma
  clinical features of, 406, 609f, 1599f, 1611–1612, 1611f, A5
  diagnosis of, 1612
  epidemiology of, 1586, 1595f
  etiology of, 610, 1586
  HHV-8 infection and, 727, 1516, 1586
  in HIV infection, 1586, 1595f, 1599f, 1611–1612, 1611f, 1612t
  oropharyngeal, 264t
  pathogenesis of, 1586
  pericardial, 1598
  staging of, 1612, 1612t
  treatment of, 1612–1613, 1612t
Kaposi’s sarcoma-associated herpesvirus (KSHV) infection. *See* Human herpesvirus-8 (HHV-8) infection
Kappa (κ), 47–48
Kappa (κ) receptors, 3689t
Karnofsky Performance Scale, 502, 503t, 557
Karshi virus, 1660t
Karyomegalic tubulointerstitial nephritis, 2436, 2442
Karyorrhexis, 539
Katayama fever, 1828–1829
katG gene, 1383, 1397
KATs, 3937
Kava, 324, 489
Kawasaki’s disease
  clinical features of, 141t, 146, 402
  epidemiology of, 2907
  inflammasome mutations in, 2762t
  prognosis of, 2907
  rash in, 141t, 402, A1
  treatment of, 2907
  vasculitis in, 2907
Kayser-Fleischer ring, 2630, 3339, 3339f, 3518, A15
KCl. *See* Potassium chloride (KCl)
KCNH2 gene mutations, 1970t
KCNJ1 gene mutations, 2369
KCNJ2 gene mutations, 354
KCNJ5 gene mutations, 355
KCNJ18 gene mutations, 3647
KCNN4 gene mutations, 794t, 795
KCNQ1 gene mutations, 1970t
KCNQ2 gene mutations, 3412t
KDELR2 gene mutations, 3325t, 3327
KDM6A/B, 3939, 3940
KDPI (Kidney Donor Profile Index), 2405
Keap1 gene mutations, 661t
Kearns-Sayre syndrome (KSS)
  clinical features of, 3287, 3645, 3811t, 3812–3813
    AV conduction block, 1930
    endocrine, 3067t, 3093
    ocular, 232, 234
  diagnosis of, 3645
  genetic considerations in, 3067t, 3645
  mtDNA mutation in, 3093, 3645, 3811t, 3812
  treatment of, 3645
Kédougou virus, 1660t
Kelch-like protein 11 antibodies, 743t
Kelling-Madlener procedure, 2533
Kell RBC group system, 904, 904t
Keloid, 378t, A5
Kemerovo virus, 1658t, 1669
Kennedy’s disease. *See* Spinobulbar muscular atrophy
Kenney-Caffey syndrome, 3286
Keratic precipitates, 224
Keratin, 2790t
Keratinocyte(s), 430
Keratitis, **224**
  A. baumannii, 1296
  Acanthamoeba, 1759, 1759f
  Aspergillus, 1715
  fungal, 1728
  Fusarium, 1728
  HSV, 224, 1486t, 1499, 1503t
  Nocardia, 1360, 1362, 1362t
  Onchocerca, 1823
  P. aeruginosa, 1309t, 1311
  Scedosporium, 1728
Keratoacanthoma, 606, 608f, A5
Keratoconjunctivitis, 220
Keratoconjunctivitis sicca (dry eye), **224**
  diagnosis of, 2877
  differential diagnosis of, 2878t
  in Sjögren’s disease, 2877, 2878t
  in SLE, 2831
  in systemic sclerosis, 2871
  treatment of, 224, 2878
Keratoconus, 3330
Keratolytic agents, 390
Keratomalacia, 2611
Keratosis pilaris, 378t, 380f
Kerion, 389, 395t
Kerley B lines, 2066
Kernicterus, 2639t, 2640
Kernig’s sign, 1119
Kernohan-Woltman sign, 188
Keshan’s disease, 2018, 2615

- Ketamine
  - abuse of, 184, 2667, 3696
  - antidepressant effects of, 3654, 3667
  - overdosage/poisoning with, 3705
- Ketoacidosis
  - alcoholic, 348, 370
  - breath odor in, 267
  - diabetic. *See* Diabetic ketoacidosis (DKA)
  - nausea and vomiting in, 298
  - treatment of, 374
- α-Ketoadipic acidemia, 3369t
- Ketoconazole
  - adverse effects of, 359, 389, 1742t, 3012, 3115
  - for Cushing's disease, 3012
  - for Cushing's syndrome, 3059
  - CYP3A and, 485t
  - for dermatophyte infections, 389
  - drug interactions of
    - calcineurin inhibitors, 2408
    - CYP3A in, 485t, 2102
    - major, 1742t
    - PDE-5 inhibitors, 3159
    - proton pump inhibitors, 489t, 2525
    - sulfonylureas, 3213
    - tacrolimus, 2721
  - for ectopic ACTH syndrome, 739
  - for hypercalcemia, 365, 3285
  - for male precocious puberty, 3110
  - in pregnancy and lactation, 1742t
  - for seborrheic dermatitis, 385
  - shampoo, for tinea, 395t
  - for topical candidiasis, 1692
- Ketogenic diet, 3412t, 3420
- Ketolides, 1166, 1177, 1182t, 1184. *See also* Telithromycin
- Ketonuria, 3219
- Ketoprofen, 3137
- Ketorolac, 96, 97t, 224, 3772t
- Ketosis-prone type 2 diabetes mellitus, 3204
- Keystone virus, 1659t
- KFR (Kidney Failure Risk) equation, 2387
- Khât, 3693, 3694. *See also* Psychostimulants
- KI. *See* Potassium iodide (KI)
- Ki67, 3065
- Ki-67 proliferative index, 677, 677t, 686
- Kidd RBC group system, 904, 904t
- Kidney. *See also* Nephron(s)
  - calcium absorption/excretion by, 3262
  - congenital abnormalities of, 2436
  - disease/failure
    - acute. *See* Acute kidney injury (AKI)
    - chronic. *See* Chronic kidney disease (CKD)
    - glomerular. *See* Glomerular diseases
    - heart failure and, 1981
    - nephrolithiasis. *See* Nephrolithiasis
    - tubulointerstitial diseases. *See* Tubulointerstitial diseases
    - vascular. *See* Renovascular disease
  - drug clearance in, 1167
  - embryological development of, 2363–2364, 2364f
  - erythropoietin production in, 444, 444f
  - in hypothermia, 3769t
  - phosphate absorption/excretion in, 3264
  - potassium absorption/excretion by, 355, 2367f, 2370, S1
  - sodium absorption/excretion in, 345–346, 345f, 346f, 2366–2367f, 2371f, 2372, S1
  - vasculature of, 2150, 2364, 2365f
  - in vitamin D metabolism, 3268, 3268f, 3269f
  - water absorption/excretion by, 345, 345f, 2366–2367f, 2371, 2371f, S1
- Kidney biopsy. *See* Renal biopsy
- Kidney cancer. *See also* Renal cell carcinoma (RCC)
  - deaths from, 497t, 687
  - hereditary papillary, 524t
  - incidence of, 497t, 687
  - paraneoplastic syndromes in, 737t, 740t
- Kidney Donor Profile Index (KDPI), 2405
- Kidney Failure Risk (KFR) equation, 2387
- Kidney injury molecule-1 (KIM-1), 2382
- Kidney mass, 2362–2363
- Kidney transplantation, **2405**
  - for acute intermittent porphyria, 3347
  - in diabetic nephropathy, 3226
  - donor selection for, 2405t, 2406–2407
  - for glucose-6-phosphatase deficiency, 3359
  - graft survival rates after, 2405, 2406t
  - HLA typing in, 2406
  - mortality rates after, 2405, 2406t
  - organ allocation system for, 2405
  - presensitization in, 2406–2407
  - racial/ethnic disparities in referral for, 62, 63, 63f
  - recipient evaluation for, 2397, 2405–2406
  - recipient management
    - algorithm for, 2409, 2409f
    - chronic lesions of graft, 2410
    - complications, 2410–2411
    - induction therapy, 2407–2408
    - infections, 1159t, 1162f, 2410, 2410t. *See also* Solid organ transplant recipient, infections in
    - maintenance therapy, 2408–2409, 2408t
    - malignancy, 2410
    - in rejection, 2409–2410
  - rejection in, 2407, 2407f
  - results of, 2405, 2406t
  - in systemic sclerosis, 2875
  - tissue typing and immunogenetics in, 2406
- Kikuchi's disease, 1387
- Killer cell immunoglobulin-like activating or inhibitory receptors (KIRs), 2768, 2772t
- Killian's dehiscence, 294
- Killian's triangle, 2506, 2506f
- KIM-1 (kidney injury molecule-1), 2382
- Kimmelstiel-Wilson nodules, 2423, A4
- Kimura's disease, 1387
- Kinase inhibitors, 526f, 536. *See also specific drugs*
- Kinases, 532
- Kindler syndrome, 396
- *Kingella kingae,* 1266, 1267t. *See also* Infective endocarditis, HACEK group
- *Kir 2.1* gene mutations, 3647
- Kirby-Bauer method, S11
- *KIR* gene, 1582t
- *KISS1* gene mutations, 3112t, 3113, 3131
- *KISS1R* gene mutations, 3127, 3131
- Kissing bug bite, 3736
- Kisspeptin, 2985t, 2991, 3113, 3127
- *KIT* gene mutations, 880–881, 2819, 2821
- *KITLG* gene mutations, 705
- Kit ligand, 759
- Kjer's dominant optic atrophy, 222
- *Klebsiella* spp.
  - antibiotic resistance in, 1283, 1291
  - *K. granulomatis,* 1356. *See also* Donovanosis
  - *K. oxytoca,* 1289
  - *K. pneumoniae,* 1181, 1186, 1281, 1289, 1290f
  - microbiology of, 1281
- *Klebsiella* spp. infections, **1289**
  - abdominal, 1290, 1290f
  - atrophic rhinitis, 1289
  - bacteremia, 1290
  - cellulitis, 1290
  - clinical features of, 306t, 993, 1289–1290
  - diagnosis of, 1290, 1290f
  - diarrhea, 306t, 1079
  - donovanosis. *See* Donovanosis
  - ecthyma gangrenosum, 408
  - epidemiology of, 1281
  - health care–associated, 985
  - liver abscess, 1074
  - pneumonia, 1289–1290
  - rhinoscleroma, 1289, 1720
  - sepsis/septic shock, 2316
  - treatment of, 1174t, 1291
  - urinary tract, 2453
  - UTI, 1087, 1290
- *KLF1,* 771
- KLHL3, 359
- Klinefelter's syndrome, **3095**
  - clinical features of, 2183, 3096, 3097t, 3107, 3114, 3116
  - diseases associated with, 3114
  - genetic considerations in, 3114
  - lymphoid malignancies in, 858t
  - mediastinal nonseminoma germ cell tumors in, 709
  - pathophysiology of, 3095
  - prevalence of, 2979t
  - screening/testing for, 2979t
  - treatment of, 3096–3097, 3114
- Klippel-Trénaunay syndrome, 2180, 2183
- *KLLN* gene mutations, 3080t, 3088
- *Kluyvera,* 1294
- *KMT2A. See MLL*
- *KMT2B* gene mutations, 3510t, 3511
- *KMT2C* gene mutations, 661t
- *KMT2D* gene mutations, 803
- KMTs, 3937
- Knee(s)
  - bursitis of, 2973
  - CPPD arthropathy of, 2960
  - gout tophi of, A15
  - iliotibial band syndrome in, 2974
  - osteoarthritis of, 2949, 2951, 2951f, 2953, 2953f
  - pain in, 2944, 2952, 2953f
- Knee replacement, 2956
- Knock-knees, 2944, 2951, 2951f
- Knudson two-hit hypothesis/model, 3783t, 3788, 3792
- Koebner (isomorphic) phenomenon, 386, 1612
- KOH (potassium hydroxide) wet-mount preparation, 381, S11
- Koilonychia, 765
- Kokobera virus, 1660t, 1669
- Kommerell's diverticulum, 2165
- Koplik's spots, 137, 401, 1641, A1. *See also* Measles (rubeola)
- Korsakoff 's psychosis/syndrome, 2349, 3483, 3676
- Kostmann's syndrome, 457, 826
- Koutango virus, 1660t
- KPCs (*Klebsiella pneumoniae,* carbapenemase-producing), 1181
- Krabbe disease, 3353t, **3357,** A16
- Kraits, 3717. *See also* Snakebites
- *KRAS* gene mutations
  - cancers associated with, 520–521, 520t
  - in carcinoma of unknown primary, 734
  - in cervical cancer, 713
  - in cholangiocarcinoma, 667
  - in colorectal cancer, 520t, 653
  - drug response and, E72t
  - as drug target, 536, 572
  - liquid biopsy for detection of, 3982
  - in lung cancer, 617, 617f, 617t, 628, 628f
  - in multiple myeloma, 784, 884
  - in ovarian cancer, 710
  - in pancreatic cancer, 672, 675–676
  - in skin cancer, 431
  - in small-bowel neoplasms, 651
  - syndromes associated with, 3788t
  - treatment-resistant, 574
- KRAS/RAF/MEK inhibitors, 569t, 572
- *KRIT1* gene mutations, 3457

Krukenberg tumors, 646, 710, 711
4Kscore Test, 697
KSHV (Kaposi's sarcoma-associated herpesvirus) infection. *See* Human herpesvirus-8 (HHV-8) infection
KSS. *See* Kearns-Sayre syndrome (KSS)
Ku-DNA-protein kinase, 2790t
Kugelberg-Welander disease, 3524
Ku-nucleoprotein La phosphoprotein, 2790t
Kupffer cells, 2627
Kurkino virus, 1661t, 1676
Kuru, 3526, 3526t
Kussmaul respiration, 191, 3216
Kussmaul's sign, 327, 1861, 1983, 2084
Kwashiorkor. *See* Malnutrition
Kyasanur Forest disease virus, 1660t, 1671, 1678
Kyphoplasty, 3300, 3310, 3393
Kyphoscoliosis, 269, 269t

L

Labetalol
  for aortic dissection, 2170
  for hyperperfusion disorders, 2350
  for hypertension in pregnancy, 2140, 3899
  for hypertensive emergency, 2148
  for ischemic heart disease, 2100t
  overdosage/poisoning with, 3712t
  for preeclampsia, 3899
  for sympathomimetic overdosage/poisoning, 3710t
Laboratory diagnosis/tests. *See also* Clinical laboratory tests
  infectious diseases, **S11**
  parasitic infections, **S12**
Laboratory worker's lung, 2228t
Labyrinthectomy, 247
Labyrinthine infarction, 247
Labyrinthitis, 83, 299t
Lacosamide, 716, 3422t, 3424, 3426f
Lacrimal glands, 224, 2932f
Lacrimation, with headache, 114
La Crosse virus, 1659t, 1672–1673
Lactase deficiency, **2544**
  diagnosis of, 301, 311, 2544–2545
  diarrhea in, 309
  indigestion in, 300
  pathophysiology of, 2544
  postinfectious, 1079t
  treatment of, 301, 309, 311, 2545
Lactate dehydrogenase (LDH)
  in anemia, 449
  in ascitic fluid, 328
  in hemophagocytic lymphohistiocytosis, 464
  with left ventricular assist systems, 2032
  in melanoma, 602
  in non-Hodgkin's lymphoma, 860
  in PCP, 1731
  in testicular cancer, 705, 708t
  in thromboembolic renal disease, 2153
  as tumor marker, 504t
Lactate dehydrogenase (LDH) deficiency, 3361t
Lactation. *See* Breast-feeding
Lactic acidosis
  approach to the patient, 369–370
  in cancer, 592
  etiology of, 369
  in malaria, 1764
  in shock, 2313
  treatment of, 369–370, 374
Lactobacillus spp., 1095, 1102, 2528, 2573, 2580, 3398, 3839
Lactobacillus gasseri, 302
Lactoferrin, 453, 455
Lactophenol cotton blue stain, S11
Lactose, 2542
Lactose breath hydrogen test, 311, 2544–2545
Lactose intolerance/malabsorption. *See* Lactase deficiency
Lactose tolerance test, 2544
Lactulose, 82t, 309, 314, 2575, 2577t, 2717
Lacunar cells, A6
Lacunar infarction, 3430, 3446, 3490, 3490f, A16
LAD (leukocyte adhesion deficiency), 458, 459t, 460, 2796, S9
LADs (lamin-associated domains), 3941
Lady Windermere syndrome, 1416
Lafora's disease (progressive myoclonus epilepsy), 3412t, 3516
LAG-3, 605
Laguna Negra virus, 1661t, 1674
Laherparepvec, 3830
Laing distal myopathy, 3640t, 3643–3644
LAM (lymphangioleiomyomatosis), 2260, 2261
LAMB4 gene mutations, 2579
Lambert-Eaton myasthenic syndrome (LEMS)
  vs. botulism, 1236
  clinical features of, 3628
  diagnosis of, 3628
  immunopathogenesis of, 2791t
  in lung cancer, 619, 744, 3628
  vs. myasthenia gravis, 3628
  ocular manifestations of, 235
  prognosis of, 3633
  reflex responses in, 169
  treatment of, 3628
LAMB syndrome. *See* Carney syndrome/complex (LAMB/NAME syndrome)
Lamellar bone, 3261
Lamin-associated domains (LADs), 3941
Laminectomy, 128, 132
Lamin proteins, 2009
Lamivudine (3TC)
  adverse effects of, 1617t, 2677, 3611
  for chronic HBV infection, 1490t, 1491, 2676–2677
    with cirrhosis, 2677
    with HIV co-infection, 2677
    vs. other agents, 2676t, 2680f
    vs. PEG IFN, 2682t
  in combination formulations, 1621t
  for HIV infection, 1617t
  molecular structure of, 1619f
  resistance to, 1491, 1622f, 2677, 2682t
Lamotrigine
  adverse effects of
    common, 3670t
    cutaneous, 422, 3420
    DIHS, 424. *See also* Drug-induced hypersensitivity syndrome (DIHS)
    fever, 150
    neurologic, 3422t
    SJS or TEN, 401, 425, 3420, 3670t. *See also* Stevens-Johnson syndrome (SJS); Toxic epidermal necrolysis (TEN)
    systemic, 3422t
  contraindications to, 3420
  drug interactions of, 3420, 3422t
  for mood disorders, 3668, 3670t
  for neuropathic pain, 82, 506
  overdosage/poisoning with, 3713t
  pharmacology of, 3422t
  for seizures, 717, 3422t, 3424
  for SUNCT/SUNA prevention, 3473
  for trigeminal neuralgia, 3553
LAMP2 deficiency (Danon disease), 2004t, 2010, 3354t, 3361t, 3365
LAMP2 gene mutations, 2002t, 2004t, 2010, 3365
Lancinations, 172
Langerhans cell(s), 432
Langerhans cell histiocytosis (histiocytosis X), 261, 881–882, 2993, 3003, A12, A16
Langerhans cell sarcoma, 882
Language, 201, 3380, V2
Language barriers, 66
Laninamivir, 1489, 1537
Lanreotide
  for acromegaly, 3009
  adverse effects of, 3009
  for gastrinoma, 3082
  for gastrointestinal NETs, 652, 683f, 684, 684t
Lansoprazole, 301, 2524t, 2525, 2531, 3081
Lanthanum, 2391
Laparoscopic adjustable gastric banding, 3194, 3194f
Laparoscopic adrenalectomy, 3063
Laparoscopic appendectomy, 2597
Laparoscopic sleeve gastrectomy, 3194, 3194f
Laparoscopy, 113, 553, 1105
Laparotomy, 2589
Lapatinib
  actions of, 533t, 568t, 570
  adverse effects of, 568t, 754
  for breast cancer, 637, 641
  genetic variations in response to, E72t
  interactions and issues, 568t
Laplace's law, 1851
LA-PTT, 475
LARC/MIP-3α/Exodus-1, 2767t
Large granular lymphocytes, 443, 457. *See also* Natural killer (NK) cells
Large language models (LLMs), 74
Laron syndrome, 3871
Laron-type dwarfism, 2984, S9
Larotrectinib, 533t, 568t, 572, 628
Larva currens, 1816
Larva migrans
  cutaneous
    clinical features of, 1051, 1812
    diagnosis of, S12
    pathogenesis of, 1812
    from pentastomes, 3733
    treatment of, 1812, 1812t
  ocular, 1812
  visceral
    clinical features of, 1738t, 1739t, 1812, S7
    diagnosis of, 1812, S7, S12
    epidemiology of, 1812
    incubation period for, S7
    pathogenesis of, 1812
    treatment of, 1747, 1812, 1812t, S7
    in war veterans, S7
Laryngeal cancer, 611, 613. *See also* Head and neck cancer
Laryngeal nerve palsy, 3556–3557
Laryngitis, **259**
Laryngotracheobronchitis (croup), 259, 1548
LAS (lung allocation score), 2283
Laser iridectomy, 225
Laser photocoagulation, 3223, 3224
Laser therapy
  for basal cell carcinoma, 607
  for esophageal cancer, 645
  for hair removal, 3141
  for warts, 1530, 1531t
LASIK, 220
Lasmiditan, 3467t, 3468t, 3469
Lassa virus, 993, 1658t, 1675–1676
LATE. *See* Limbic-predominant aging-related TDP-43 encephalopathy (LATE)
Latency-associated transcripts (LATs), of HSV, 1495
Latent diffusion models, 3972t
Lateral epicondylitis, 2975
Lateral femoral cutaneous neuropathy (meralgia paresthetica), 3614, 3903
Lateral geniculate body, 220
Lateral inferior pontine stroke syndrome, 3434f

Lateral medullary stroke syndrome, 3433, 3433f, A16
Lateral midpontine stroke syndrome, 3435f
Lateral sclerosis, 3520, 3569
Laterocollis, 3510
Latex agglutination assay, S11
α-Latrotoxin, 3734
LATs (latency-associated transcripts), of HSV, 1495
Laugier-Hunziker syndrome, 399
Laurence-Moon syndrome, 3113
Laxatives
  abuse of, 308–309, 355
  adverse effects of, 306, 313, 355
  for constipation, 82, 82t, 313, 314
  for IBS, 2575
  for internal radionuclide contamination, S6
Lazertinib, 627t
LCAT deficiency, familial, 3241t, 3248
LCK-TCRB gene mutations, 521t
LCM (lymphocytic choriomeningitis), 1124, 1126, 1131t, 1658t, 1665
LCZ696. *See* Sacubitril-valsartan
LDH. *See* Lactate dehydrogenase (LDH)
LDL-C. *See* Low-density lipoprotein cholesterol (LDL-C)
LDLRAP1 gene mutations, 3243
LDLR gene mutations, 3242
Lead poisoning
  vs. ALS, 3523
  clinical features of, 3484, 3699, 3700t
  diagnosis of, 3611, 3700t
  hemolytic anemia in, 803
  hypertension and, 3701
  metabolism of, 3700t
  neuropathy in, 3610t, 3611
  optic neuropathy in, 228
  oral manifestations of, 264t
  peripheral blood smear in, 437, 441f
  pregnancy-associated risks, 3701
  sources of, 3700t
  treatment of, 3484, 3611, 3700t, 3702
Lead-time bias, 38, 515, 617
Leaky epithelia, 2365
Leber’s hereditary optic neuropathy (congenital amaurosis)
  clinical features of, 227, 2430t, 3810, 3811t
  diagnosis of, 3810, 3814
  genetic considerations in, 227, 2430t
  genetic counseling in, 3816
  mtDNA mutation in, 3810, 3811t, 3814
  treatment of, 227, 232, 3816
  visual fields in, 222
Lebombo virus, 1658t, 1669
Lebrikizumab, 2793t
Lecanemab, 199–200, 3480–3481, 3481f
Lechiguanas virus, 1661t, 1674
Lectins, 759
Ledipasvir, 2686f
Ledipasvir/sofosbuvir, 1492t, 2688t, 2689–2690
Leeches, 1268
Leech infestations, 3733
Lefamulin
  actions of, 1166, 1178, 1182t, 1183f
  adverse effects of, 1171t, 1178
  drug interactions of, 1172t, 1178
  indications for, 1174t, 1178, 1185
  in pregnancy and lactation, 1169t
  resistance to, 1174t, 1182t, 1183f, 1184–1185
Leflunomide, 2834
  adverse effects of, 2850t, 2853, 3609t
  monitoring during treatment with, 2850t
  for relapsing polychondritis, 2921
  for rheumatoid arthritis, 2849, 2850t, 2852
  for sarcoidosis, 2929t, 2930
Left atrial appendage closure, A11, A11v
Left bundle branch, 1925
Left bundle branch block, 1865f, 1871–1872, 1872f, 1981, A7, A8
Left ventricular aneurysm, 2125, S13
Left ventricular assist systems, 2030–2032, 2032–2033f, 2328–2329, A11, S13. *See also* Mechanical circulatory support
Left ventricular dysfunction
  in aortic regurgitation, 2043
  cardiogenic shock in, 2324, 2328f
  diagnosis of, 1882, 1883f, 1902v, 1907, 1907f
  in heart failure. *See* Heart failure (HF)
  in mitral regurgitation, 2052–2053
  pulmonary hypertension and, 2192
  in STEMI, 2121–2122
Left ventricular fascicular VT, 1967
Left ventricular hypertrophy
  in CKD, 2392
  ECG in, 1871, 1871f, A7
  hypertension and, A7
  with myocardial ischemia, 2091
Left ventricular noncompaction, 2004t
Left ventricular pressure-volume relationship, 1853, 1853f
Left ventricular remodeling, 1979–1980, 1980t, 2121–2122. *See also* Heart failure (HF)
Legal blindness, 221
Legal issues, 71, 79, 89–90
Leg elevation, for stasis dermatitis/ulceration, 385, 393t
Legionella spp.
  apoptosis inhibition by, 973
  autophagy inhibition by, 973
  cellular trafficking of, 969
  epigenetic modulation by, 973
  free-living ameba and, 1758
  laboratory identification of, S11
  microbiology of, 1269
  NF-κβ inhibition by, 971
  non-*pneumophilia,* 1270
  secretion system of, 968
  in water systems, 1269–1271, 1271f, 1271t, 1275–1276
Legionella spp. infections, **1269**
  in cancer patient, 582, 1271
  clinical presentation of, 1272–1273, 1272t
  diagnosis of, 1025, 1273–1275, 1273f, 1274f, S11
  as emerging disease, 3858
  epidemiology of, 1269–1270, 1270f
    common-source outbreaks, 1270–1271, 1271f, 1271t, 1276
    seasonality and climate in, 1270
    sporadic outbreaks, 1271
    transmission, 1272
  health care–associated, 1150–1151
  non-*pneumophilia,* 1272, 1273, 1274f
  prevention of, 1275–1276
  prognosis of, 1275
  risk factors for, 1024, 1271–1272
  treatment of, 1185, 1275, 1275t
Legionnaire’s disease. *See* *Legionella* spp. infections
Leg length discrepancy, 120
Leigh’s syndrome, 3646, 3811t, 3812, 3812f, 3816, A16
Leiomyoma (uterine fibroids), 218, 3136, 3137, 3152
Leiomyomatosis, hereditary, 687t
Leiomyosarcoma, 714, 715. *See also* Soft tissue sarcomas
Leishmania spp., 1737
  characteristics of, 1781, 1782t
  entry into host, 2780
  geographic distribution of, 1782t
  laboratory identification of, S12
  L. aethiopica, 1782t, 1786. *See also* Cutaneous leishmaniasis
  L. braziliensis, 1782t, 1786, 1787, S12. *See also* Cutaneous leishmaniasis; Mucosal leishmaniasis
  L. donovani, 1781, 1782t, S12. *See also* Visceral leishmaniasis (kala-azar)
  L. guyanensis, 1782t, 1787. *See also* Cutaneous leishmaniasis
  life cycle of, 1781, 1782t, S12
  L. infantum, 1781, 1782t. *See also* Visceral leishmaniasis (kala-azar)
  L. major, 1782t, 1786, 1787. *See also* Cutaneous leishmaniasis
  L. mexicana, 1782t, 1786, 1787. *See also* Cutaneous leishmaniasis
  L. tropica, 1782t, 1786, S12. *See also* Cutaneous leishmaniasis
Leishmaniasis, **1781**
  cutaneous. *See* Cutaneous leishmaniasis
  diffuse cutaneous, 1782t, 1787
  epidemiology of, 1781, 1782t, 1783f, S7, S12
  mucosal, 1782t, 1787, 1788f
  prevention of, 1788
  visceral. *See* Visceral leishmaniasis
Lemborexant, 216
Lemierre’s disease/syndrome, 259, 994–995, 1035, 1375
LEMS. *See* Lambert-Eaton myasthenic syndrome (LEMS)
Lenacapavir, 1619t, 1620f, 1621, 1623f
Lenalidomide
  actions and targets of, 559f, 561
  adverse effects of, 753t
    constipation, 509
    diarrhea, 757
    DVT, 887
    heart failure, 2017
    hemolytic anemia, 802t
    immune-related, 755
    myelosuppression, 817
    neuropathy, 508
    second malignancy, 892
  for AL amyloidosis, 898
  for ATL, 1555f, 1556
  for DLBCL, 863
  for follicular lymphoma, 864
  for MDS, 817
  for POEMS syndrome, 3093
  for primary effusion lymphoma, 873
  for SLE, 2834
Lenègre’s disease, 1925
Length bias, 38, 617
Length-biased sampling, 515
Lennox-Gastaut syndrome, 3411
Lenticulostriate arteries, 3429
Lentigo, 398
Lentigo maligna, 601t, 602f, 603, A5
Lentiviruses. *See also* HIV
  in gene therapy, 1553, 3402, 3823t, 3825t, 3826
  nononcogenic, 1557
  phylogenetic tree of, 1558f
  structure and life cycle, 1552
Lenvatinib
  action and targets of, 533t, 548, 569t
  adverse effects of, 569t, 666, 754
  for endometrial cancer, 715
  for HCC, 664, 664t, 665f, 666
  interactions and issues, 569t
  with pembrolizumab, 548
  for RCC, 535, 689, 690t
LEOPARD syndrome, 398
LEP gene mutations, 3112t
Lepirudin, 950
Lepra cells, 1404
LEPR gene mutations, 3112t

Lepromatous leprosy, 1051, 1404, 1408, 1408f, A5
Lepromin test, 1410
Leprosy, **1404**
  classification of, 1406
  clinical features of, 1051, 1406–1407, S7
    borderline lepromatous, 1407–1408, 1407f
    borderline tuberculoid, 1407, 1407f
    lepromatous, 1051, 1404, 1408, 1408f, A5
    mid-borderline, 1407
    neuropathy, 3555, 3606
    renal, 2416t, 2429
    tuberculoid, 397, 1407, 1407f
  complications of, 1408–1409, 1409f
  diagnosis of, 1409–1410
  differential diagnosis of, 1410
  epidemiology of, 1404–1405, 1405–1406f
  etiology of, 1404
  global considerations in, 1404–1405, 1405–1406f, 3606
  incubation period for, 1405–1406, S7
  indeterminate, 1406–1407
  inflammasome mutations in, 2761t
  lepromatous, 3606
  multibacillary, 1406, 1411t
  pathogenesis of, 1406
  paucibacillary, 1406, 1411t
  postexposure prophylaxis for, 1413–1414
  prevention and control of, 1413–1414
  primary neuritic, 1408
  rehabilitation and social aspects, 1413
  transmission of, 1405
  treatment of, 3606, S7
    adverse effects of, 1411
    disease management during and after, 1412–1413
    leprosy reactions, 1411–1412
    multidrug, 1410–1411, 1411t
    for nerve function impairment and neuropathic pain, 1412
    relapse, 1411
    second-line drugs, 1411
  tuberculoid, 3606
  vaccines, 1413
Leprosy reactions, 1408–1409, 1409f, 1411–1412
Leptin
  actions of, 3182, 3950
  deficiency of, 2993
  in hypogonadotropic hypogonadism, 3113, 3134
  in male pubertal development, 3102–3103
  in obesity, 3184, 3185t
  resistance to, 3254
Leptin receptors, 3185, 3185t
Leptomeningeal metastases, 620, 724, 725f, S2
Leptospira/leptospirosis (Weil's syndrome), **1439**
  climate change and, 3752t
  clinical features of, 139t, 145, 326, 1130t, 1442–1443
  diagnosis of, 1443, S7, S11
  differential diagnosis of, 1443
  epidemiology of, 1019, 1440–1441
  etiologic agent of, 1440, 1440–1441f
  global considerations in, 1041
  incubation period for, 1441, 1441f, S7
  in international travelers, 1441
  pathogenesis of, 1441–1442, 1441–1442f
  prevention of, 1444, 1444t, S7
  prognosis of, 1444
  treatment of, 1443–1444, 1444t, S7
Leriche's syndrome, 2170
LERPE1 gene mutations, 3325t
LES (lower esophageal sphincter), 293, 2507, 2508f
Lesbian, 3179t
Lesbian, gay, bisexual, transgender, queer or questioning, intersex, asexual and more (LGBTQIA+) patients, **3178**
  future directions in, 3180
  health disparities in, 3178
  outdated and insensitive terms, 3180t
  positive health experiences for, 3178–3180
  terminology and definitions, 3179–3180t
Leser-Trelat sign, 398
Letermovir
  for CMV prophylaxis, 1156, 1160, 1487, 1515
  for CMV treatment, 1516
  drug interactions of, 1487
  for HSV prophylaxis, 1486t
  for infection prophylaxis after HCT, 918
  pharmacology of, 1487
Lethal midline granuloma. *See* Extranodal NK/T cell lymphoma, nasal type
Letrozole, 567, 593, 637, 753t, 3135, 3151
Leu7, 616
Leucine metabolism disorders, 3369t
Leucovorin
  for colorectal cancer, 658
  for DLBCL, 863
  5-FU and, 565
  for gastric cancer, 649
  methotrexate and, 562t, 564
  for osteosarcoma, 730
  for pancreatic cancer, 674
  for PCP prophylaxis, 1593t, 1734t
  for toxoplasmosis prophylaxis, 1593t
Leukapheresis, for CML, 842
Leukemia
  acute lymphoid. *See* Acute lymphoid leukemia (ALL)
  acute myeloid. *See* Acute myeloid leukemia (AML)
  acute promyelocytic. *See* Acute promyelocytic leukemia (APL)
  adult T cell. *See* Adult T-cell leukemia/lymphoma
  aggressive NK cell, 874
  chemotherapy-associated, 871
  chromosome abnormalities in, 539
  chronic eosinophilic, 819, 876, 878t
  chronic lymphocytic. *See* Chronic lymphocytic leukemia (CLL)
  chronic myeloid. *See* Chronic myeloid leukemia (CML)
  chronic myelomonocytic, 815, 877–878, 877t
  chronic neutrophilic, 819, 875–876, 877t
  deaths from, 497t, 499f, 500f
  definition of, 528
  hairy cell. *See* Hairy cell leukemia
  incidence of, 497t
  joint involvement in, 2972
  juvenile myelomonocytic, 879
  lymphadenopathy in, 476
  mast cell, 2820
  neuropathy in, 3607
  ocular invasion by, 232
  oral manifestations of, 261
  paraneoplastic syndromes in, 740t
  polycythemia vera and, 820
  splenomegaly in, 480
  T-cell large granular lymphocytic, 874, 2841
  T-cell prolymphocytic, 481, 874
  viruses and, 525
Leukemia cutis, 405, 583
Leukemia inhibitory factor (LIF), 2767t
Leukemoid reaction, 458
Leukoaraiosis (diffuse/periventricular white matter disease), 198f, 3448
Leukocyte(s), 443, 453, A4. *See also* Basophil(s); Eosinophil(s); Lymphocyte(s); Monocyte(s); Neutrophil(s)
Leukocyte adhesion deficiency (LAD), 458, 459t, 460, 2796, S9
Leukocyte common antigen, 733t
Leukocyte esterase test, 1092
Leukocyte function-associated antigen (LFA-1), 2790t
Leukocytoclastic (cutaneous small-vessel) vasculitis
  clinical features of, 145t, 148, 148f, 408, A1, A5
  drug-induced, 426, 426f
  epidemiology of, 145t
  etiology of, 145t, 408
Leukoderma, 397, 398t
Leukodystrophy(ies), 3353–3354t, 3485
Leukoencephalopathy
  progressive multifocal. *See* Progressive multifocal leukoencephalopathy (PML)
  radiation therapy–related, 726
  reversible posterior, 591, 3447–3448, 3455
Leukoerythroblastosis, 818. *See also* Primary myelofibrosis
Leukopenia. *See* Neutropenia
Leukoplakia
  characteristics of, 261, 380f, A3
  in dyskeratosis congenita, 3819, 3820f
  etiology of, 261
  oral hairy. *See* Oral hairy leukoplakia
  as premalignant lesion, 512, 610, 611
  smoker's, 264t
  sublingual, A3
Leukostasis, 591
Leukotriene modifiers, 2225, 2225t, 2226. *See also* Montelukast; Zafirlukast
Leukotrienes, 93, 417
Leuprolide, 567, 637, 701
Levamisole, 420, 421, 923t, 2898, 2940t, 3694
Levetiracetam
  adverse effects of, 3422t, 3423
  for cyclic vomiting syndrome, 299t, 300
  for myoclonus, 3516
  overdosage/poisoning with, 3713t
  for seizures, 591, 716, 3422t, 3424
  for status epilepticus, 3426f
Levine's sign, 106, 2092
Levodopa
  actions of, 3501, 3501f
  adverse effects of, 3501–3502, 3507
  for dystonia, 3512
  for PD, 3501–3502, 3501f, 3504t, 3507
Levofloxacin
  actions of, 1166, 1178, 1182t
  adverse effects of, 2667
  for *Campylobacter* infections, 1306
  for community-acquired pneumonia, 963t, 1027t, 1028t
  for *C. trachomatis* infection, 1474
  for cystitis, 1093
  for epididymitis, 1099
  for HACEK endocarditis, 1266–1267, 1267t
  for hospital-acquired pneumonia, 963t, 1032t
  in *H. pylori* eradication, 1302t, 2527t, 2528, 2529f
  indications for, 1174t, 1178
  for infectious diarrhea, 1081t
  for intraabdominal infections, 963t
  for *Legionella* spp. infections, 1275t
  for leprosy, 1411
  for MDR-TB, 1394t, 1395t, 1398–1399, 1421t, 1424
  for neutropenic fever, 581
  for nongonococcal bacterial arthritis, 1057
  for osteomyelitis, 1064t, 1065t
  for *P. aeruginosa* infections, 1309t
  for peritonitis, 1071
  for plague prophylaxis, 1346, 1347, 1347t
  for plague treatment, 1346, 1346t, S4
  for pneumococcal infections, 1194

    prophylactic uses of
        in HCT recipients, 918t
        in high-risk neutropenia, 577, 1180t
        postsplenectomy, 579
    for pyelonephritis, 1094
    resistance to, 1026, 1174t, 1182t, 1303
    for sinusitis, 256
    for TB preventive treatment, 1402t, 1403
    for tularemia, 1339, 1340t
    for ventilator-associated pneumonia, 1032t
Levoketoconazole, 3012, 3059
Levomilnacipran, 3659t, 3667
Levonorgestrel, 3153–3154
Levorphanol, 97t, 100
Levosimendan, 1992, 1993t, 2000–2001
Levothyroxine (LT4)
    adverse effects of, 3031
    for diffuse nontoxic goiter, 3041
    drug interactions of, 3031
    for hypothyroidism, 3031–3032
    for myxedema coma, 3032
    in older adults, 3032
    in pregnancy, 3031–3032
    for subacute thyroiditis, 3038
Lev's disease, 1925
Lewisite, S5
Lewis RBC group system, 904t
Lewis-Sumner syndrome, 3622
Lewy body(ies)
    characteristics of, 3400, 3492
    dementia with. *See* Dementia with Lewy bodies
    in PD, 3494, 3494f
    prions and, 3399f
    proteostasis loss and, 3871, 3873f
Leydig cell, 3104–3106
LFA-1 (leukocyte function-associated antigen), 2790t
LGBTQIA+ patients. *See* Lesbian, gay, bisexual, transgender, queer or questioning, intersex, asexual and more (LGBTQIA+) patients
LGI1 antibodies, 744t, 746f, 748
LGI1 gene mutations, 3412t
LGV. *See* Lymphogranuloma venereum (LGV)
LH. *See* Luteinizing hormone (LH)
LHCRG gene mutations, 3099t
Lhermitte's sign/symptom, 129
    in cervical spine lesions, 176, 3378
    in MS, 3378, 3578
    radiation therapy–related, 726, 871
    in spinal cord compression, 588
LHON. *See* Leber's hereditary optic neuropathy (congenital amaurosis)
LHX3 gene mutations, 3112t, 3113
LHβ gene mutations, 3112t, 3113
Lialda, 2564, 2564t
Libertarian paternalism, 3915
Libido
    female, 3161
    in hypothyroidism, 3029
    in malabsorption disorders, 2549t
    male, 3155
Lice
    body *(Pediculus humanus),* 1446, 1458, 3731–3732
    diseases carried by
        epidemic typhus. *See* Epidemic (louse-borne) typhus
        relapsing fever. *See* Louse-borne relapsing fever
    head *(Pediculus capitis),* 3731–3732
Licensure, vaccine, 999–1000
Lichen amyloidosis, 404
Lichenification, 377t, 383
Lichen myxedematosus, generalized (papular mucinosis), 404
Lichenoid lesions, 378t
Lichen planopilaris, 388
Lichen planus, **388**
    clinical features of, 386t
        alopecia, 394
        distribution, 378t, 380f, 388
        esophageal, 296
        morphology, 378t
        oral, 263, 263t, 264t
        oral manifestations, A3
        papules and plaques, 388, 388f, A5
        papulosquamous lesions, 392
    in diabetes mellitus, 3230
    drug-related, 388
    erosive, A3
    HCV infection and, 388, 392, 2685
    histologic features of, 386t
    treatment of, 388
Lichen simplex chronicus, 380f, 384
Lichen tropicus, 3774
Licorice, 355, S1
Liddle's syndrome (pseudoaldosteronism)
    clinical features of, 355, 374, 2370
    diagnosis of, 356, 3061
    genetic considerations in, 355, 2368t, 3060t, 3061
    pathogenesis of, 374
    pathophysiology of, 2370
    treatment of, 356, 375, 3063
Lidocaine
    for arrhythmias, 1916t, 1958
    for cardiac arrest due to VT or VF, 2336, 2337f
    for IC/BPS, 336
    inhaled, for cough, 274
    metabolism of, 485t
    for neuropathy, 100, 3604t
    for SUNCT/SUNA, 3473
    for topical anesthesia in lumbar puncture, S3
    viscous, for mucositis, 508
Lidoderm patch, 3604t
LIF (leukemia inhibitory factor), 2767t
Life expectancy
    at birth by census tract, 62f
    genetic considerations in, 3815
    global burden of disease and, 3844
    global differences in, 3860, 3861f
    by race and sex, U.S., 61f
    screening procedures and increase in, 38t
    of women, 3162–3163
Lifestyle modifications
    for GERD, 2511
    for metabolic syndrome, 3257
    for obesity, 3190–3191
Lifileucel, 535t
Li-Fraumeni syndrome
    breast cancer risk in, 633
    clinical features of, 717t, 727
    genetic considerations in, 524t, 530, 615, 717t, 727, 3783t
    lung cancer risk in, 615
Light chain deposition disease/nephropathy
    clinical features of, 2415t, 2424, 2440
    diagnosis of, 2440
    etiology of, 2440
    pathophysiology of, 2440
    in POEMS syndrome, 2447
    renal biopsy in, 2424, 2440f, A4
    treatment of, 2424, 2440
Light-headedness, 162
Likelihood ratio, 25, 25f, 48–49
Lille score, 2700
LILRB2 gene, 1582t
Limaprost, 3772
Limb dystonia, 3510
Limb-girdle muscular dystrophies (LGMDs)
    autosomal dominant, 3639t
    autosomal recessive, 3639t
    classification of, 3640
    clinical features of, 3640
    diagnosis of, 3640
    genetic considerations in, 3787–3788, 3788t
    global considerations in, 3648
    prevalence of, 3641–3642
    proteins involved in, 3641f
    treatment of, 3629t
Limbic encephalitis, 744, 746, 3484, A16
Limbic network/system, 207
Limbic-predominant aging-related TDP-43 encephalopathy (LATE)
    clinical features of, 194, 195, 196t, 197
    diagnosis of, 197
    differential diagnosis of, 196t
    molecular basis of, 194, 195t
    neuropathology of, 195t, 196
Limb ischemia, 2175. *See also* Peripheral artery disease
Limb-kinetic apraxia, 203
Limb-sparing surgery, 728, 730
Limey bile, 2733
Limit dextrinosis (type IIIa/b GSD), **3359,** 3360t
Linaclotide, 313, 314, 2576, 2577t, 3549
Lincosamides. *See* Clindamycin
Lindane, 3729, 3732
Linear and logistic regression, 3972t
Linear gingival erythema, 261
Linear IgA disease, 410t, **413,** 420
Linear nevoid hypopigmentation (hypomelanosis of Ito), 397
Linear scleroderma, 394
Linezolid
    actions of, 1166, 1182t, 1204, 1425
    adverse effects of, 923t, 1171t, 1178, 1204, 1219, 1420t, 1425
    for community-acquired pneumonia, 963t, 1028t
    drug interactions of, 1172t, 1178, 1425
    empirical uses of, 2321t
    for enterococcal infections, 1219, 1219t, 1220t
    indications for, 1174t, 1178
    for MDR-TB, 1394t, 1395, 1395t, 1398–1399, 1421, 1425
    for MRSA infections, 963t, 1028, 1032, 1203t, 1204
    for necrotizing fasciitis, 1211
    for *Nocardia* infections, 1362
    for nocardiosis, 1164
    for NTM infections, 1427, 1428
    for osteomyelitis, 1065t
    for peritonitis, 1072
    in pregnancy and lactation, 1169t
    resistance to, 1174t, 1182t, 1184, 1204, 1221
    for staphylococcal infections, 1054t, 1204
    for streptococcal infections, 1054
    for TB, 1420t, 1425
    for toxic shock syndrome, 1205
    for ventilator-associated pneumonia, 1032, 1032t
LINGO-1, 3397
Linkage analysis, 3793–3794, 3797t
Linkage disequilibrium, 3794, 3797t
Linoleic acid, 2602t
Lionfish sting, 3725
Liothyronine, 3032, 3047
LIPA gene mutations, 3357
LipA lipase, 1295
Lipase, 2738, 2739t, 2753
LIPC gene mutations, 3245
Lipedema, 2184
Lipemia retinalis, 3240

Lipid(s)
  defects in metabolism of, 3645
  deposits of, in familial hypercholesterolemia, A15
  digestion and absorption of, 2541–2542, 2541t, 2542f
  as energy source, 3645
  neutral, disorders of, 3354t
  in parenteral nutrition, 2622–2623, 2625
  production of, 3202
  transport of, 3238, 3239f
Lipid mediator generation, 2805f, 2977
Lipid membrane microdomain (lipid raft), 2776
Lipid nanoparticles, 997t, 998, 3823, 3823t
Lipoarabinomannan, 1383, 1404
Lipoatrophy, 3230
Lipodermatosclerosis, 405, 2180
Lipodystrophy
  clinical features of, A15
  familial, 3240–3241, 3254, 3787, 3788t
  in HIV infection, 1602, 1602f, 1616, 3254
  metabolic syndrome and, 3254
  treatment of, 3240–3241
Lipofuscin, 3871, 3873f
Lipoglycopeptides, 1166, 1176, 1182t, 1183–1184
Lipohypertrophy, 3230
Lipoic acid, 243, 266
Lipoidal tests, for syphilis, 1433
Lipolysis, 2541, 2542f
Lipolytic enzymes, 2743
Lipomas, 403, 2086, 3080t, 3084
Lipooligosaccharide (LOS), *N. gonorrhoeae,* 1254
Lipopeptides. *See* Daptomycin
Lipopolysaccharide (LPS)
  B. cepacia, 1313
  Brucella, 1332, 1334
  epigenetics of, 3942
  in heart failure, 1982
  P. aeruginosa, 1307t, 1308
  as pyrogen, 134
  in sepsis/septic shock, 2317
Lipoprotein(a), 3237t, 3239, 3242t
Lipoprotein(s)
  classification of, 3237–3238, 3237f
  composition of, 3237–3238. *See also* High-density lipoprotein cholesterol (HDL-C); Low-density lipoprotein cholesterol (LDL-C); Very low–density lipoproteins (VLDLs)
  metabolism of, A10, A10v
  in nutrition assessment, 2618t
  proteins associated with, 3237–3238, 3237t
  structure and metabolism of, **3237**
  in transport of dietary lipids, 3238, 3238f
Lipoprotein disorders
  approach to the patient, **3249**
  arthritis in, 2968
  cardiovascular disease risk and, 1858, 2099, 2100
  in diabetes mellitus, 3228
  diagnosis of, 3249
  with elevated apoB-containing lipoproteins, **3239**
    hypercholesterolemia. *See* Hypercholesterolemia
    hypertriglyceridemia. *See* Hypertriglyceridemia
    mixed hyperlipidemia. *See* Mixed hyperlipidemia
  erectile dysfunction in, 3155
  global considerations in, 1858
  with low HDL-C. *See* High-density lipoprotein cholesterol (HDL-C), low levels of
  with low LDL-C. *See* Low-density lipoprotein cholesterol (LDL-C), low levels of
  metabolic syndrome and, 3253f
  musculoskeletal disorders associated with, 2968
  in obesity, 3186–3187
  in older adults, 3882–3883
  screening for, 39t, 2979t, 3239, S10
  secondary causes of, 2744, 2748–2749
  skin manifestations of, 404
Lipoprotein lipase (LPL)
  in hypertriglyceridemia, 3240
  in hypolipidemia, 3248
  in insulin resistance, 3254
  in lipoprotein metabolism, 3238, 3239f
  in metabolic syndrome, 3246
Lipoprotein lipase (LPL) deficiency, 3240, 3827
Liposarcoma, 728. *See also* Soft tissue sarcomas
Lip reading, 251
Liquid biopsy, 3980, 3981f. *See also* Cell-free DNA (cfDNA); Circulating tumor DNA (ctDNA)
Liquid chromatography, 3977
Liquid nitrogen therapy, 390
Liraglutide
  actions of, 3192, 3192t
  adverse effects of, 3192, 3192t, 3257
  contraindications to, 3192t
  for diabetes mellitus, 3211t, 3213, 3215
  for metabolic syndrome, 3257
  for NASH, 2706
  for obesity, 3192, 3192t, 3193f
  for OSA, 2282
Lisch nodules, 3077, 3087
Lisdexamfetamine, 503t
Lisfranc fracture–dislocation, 2969
Lisinopril, 1995t, 2144t
Lisocabtagene maraleucel, 535t, 856, 863, 3824t
Lispro, 3209t, 3210
Lissencephaly, 3402–3403, 3412t
Listeria monocytogenes
  autophagy evasion by, 973
  characteristics of, 1227
  epigenetic modulation by, 973
  genome sequencing of, 987–988
  laboratory identification of, S11
  survival in cytosol, 968
Listeria monocytogenes infections, **1227**
  in cancer patient, 583
  clinical features of
    bacteremia, 1228
    encephalitis, 1114
    febrile gastroenteritis, 1228
    focal infections, 1228–1229
    meningitis, 994, 1119, 1122, 1122t
    neurolisteriosis, 1228
  diagnosis of, 1229
  epidemiology of, 1227–1228
  food-related, 305, 1227–1228, 1228f
  immune response to, 1227
  neonatal, 1229
  pathogenesis of, 1227
  in pregnancy, 1229
  prevention of, 1229
  treatment of, 583, 1229
Listeriolysin O (LLO), 968, 1227
Lithium
  actions of, 3652
  adverse effects of
    acid-base disorders, 368
    alopecia, 394
    cardiotoxicity, 2017
    common, 3668, 3670t
    cutaneous, 391, 392, 395, 420
    diabetes insipidus, 352, 3016, 3019, S1
    hypercalcemia, 3279
    lupus syndrome, 2940t
    nail disorders, 421
    nephropathy, 2442–2443
    neuropathy, 3609t
    nystagmus, 237
    psoriasis, 386
    rare, 3670t
    SA node dysfunction, 1920t
    secondary parkinsonism, 3497
    seizures, 3414t
    weight gain, 3188
  for bipolar disorder, 3668, 3670t
  for cluster headache prevention, 3472, 3472t
  drug interactions of, 1744t
  for glucocorticoid adverse effects, 3585
  for hypnic headache, 3475
  overdosage/poisoning with, 3715t
  pharmacology of, 3670t
Litifilimab, 2835
Live attenuated vaccines, 996, 997t
Livedo reticularis, 91, 407, 408, 2177f, **2178,** 2837
Liver
  alcohol effects on. *See* Alcohol-associated liver disease
  anatomy of, 2627, A13
  drug metabolism in, 484
  drug transport in, 484, 484f
  functions of, 2627
  iron recycling in, 763, 763f
  palpation of, 327
Liver abscess, **1074**
  Actinomyces, 1364, 1364f
  amebic
    clinical features of, 1756, S7
    complications of, 1756
    diagnosis of, 1074, 1757, 1757f, S12
    epidemiology of, 1753
    pathogenesis of, 1755–1756
    treatment of, 1757t, 1758, S7
  Candida, 1074
  clinical features of, 1074
  diagnosis of, 1074, 1074f
  etiology of, 1074
  K. pneumoniae, 1074, 1289, 1290f, 1290t
  treatment of, 1074–1075
Liver biopsy, 2632, **A13**
  in acetaminophen hepatotoxicity and alcoholic liver injury, A13
  in acute hepatitis, A13
  in alcohol-associated liver disease, 2699
  in α_1-antitrypsin deficiency, A13
  in autoimmune hepatitis, 2695, A13
  in chronic HBV infection, A13
  in chronic HCV infection, 2685, A13
  in cirrhosis, A13
  in cirrhosis secondary to hemochromatosis with HCC, A13
  in CMV infection with liver allograft, A13
  contraindications to, 2636
  in HCC, 662, A13
  in hemochromatosis, 3337, A13
  in hepatic granuloma, A13
  indications for, 2634f, 2636
  in metabolic dysfunction–associated steatohepatitis, A13
  in primary biliary cholangitis, A13
  in primary sclerosing cholangitis, A13
  in rejection of liver allograft, A13
  in Wilson's disease, 3339, A13
Liver disease/failure, **2627**. *See also* Cirrhosis
  acute-on-chronic, 2717
  in alcohol abuse. *See* Alcohol-associated liver disease
  in α_1-antitrypsin deficiency, 2629
  in β thalassemia, 778t
  in cancer patient, 757
  cholestatic, 324–325, 2627, 2628t. *See also* Biliary cirrhosis
  clinical features of, 326, 2628–2629

    - AKI, 2374
    - ascites, 328, 328f, 2629
    - bleeding, 473
    - coagulation disorders, 935, 935t, 936f
    - cutaneous, 2630
    - dyslipidemia, 3246–3247
    - encephalopathy. *See* Encephalopathy, hepatic
    - fatigue, 166, 2628
    - halitosis, 267
    - hepatomegaly, 2629
    - hepatorenal syndrome, 2374
    - icterus, 2629
    - itching, 2628
    - jaundice, 324, 2628
    - lactic acidosis, 369
    - nausea and vomiting, 2628
    - neuropathy, 3605–3606
    - odor, 50t
    - pain, 2628
    - respiratory, 376
    - splenomegaly, 2629
    - thromboembolism, 935, 936f
  - clinical history, 2628–2629, 2629t
  - in cystic fibrosis, 2245
  - diagnosis of
    - algorithm for, 2631f
    - biopsy in, 2632. *See also* Liver biopsy
    - vs. DIC, 934
    - imaging in, 2631–2632, 2631t, 2636
    - in jaundiced patient, 324
    - laboratory tests in, 2630–2631, 2630t, 2636
    - liver function tests in. *See* Liver function tests
  - drug-induced, **2665,** 2669t
    - acetaminophen in, 324, 489, 2667t, 2668–2670, A13
    - amiodarone in, 2671
    - amoxicillin-clavulanate in, 2671
    - anabolic steroids in, 2671
    - antiretrovirals in, 2672
    - clinical features of, 2667t
    - in exacerbation of preexisting liver disease, 2633
    - genetic considerations in, 419
    - herbal medicines in, 2672
    - idiosyncratic, 2665–2668, 2667t
    - isoniazid in, 2667t, 2670
    - mechanisms of, 2665–2668, 2666f
    - niacin in, 2608
    - nitrofurantoin in, 2670–2671
    - oral contraceptives in, 2667t
    - statins in, 2671–2672
    - in TB, 1397
    - treatment of, 2668
    - trimethoprim-sulfamethoxazole, 2671
    - valproate in, 2670
  - in dyskeratosis congenita, 3819
  - effect on drug concentration response, 488
  - etiology of, 2627–2628, 2628t
  - family history in, 2629
  - genetic considerations in, 2629, 3338
  - grading of, 2632–2633
  - in GVHD, 917, 917t
  - in hemochromatosis. *See* Hemochromatosis
  - in hemophagocytic lymphohistiocytosis, 464
  - in hepatitis. *See* Hepatitis
  - hepatocellular, 2627, 2628t
  - in IBD, 2563
  - immunization recommendations for, 1004t
  - in leptospirosis, 1441–1442
  - liver transplantation for. *See* Liver transplantation
  - obesity and, 3187
  - during parenteral nutrition, 2625
  - patterns of, 2627–2628, 2628t
  - peripheral blood smear in, 440f, 446f
  - physical examination in, 323–324, 2629–2630
  - in pregnancy, 3903
  - risk factors for, 2628–2629
  - in schistosomiasis, 1829
  - in sepsis/septic shock, 2322t, 2323
  - in sickle cell disease, 774t
  - specialized nutritional support in, 2626
  - staging of, 2632–2633
  - steatotic, 2700–2702. *See also* Metabolic dysfunction–associated steatohepatitis (MASH); Metabolic dysfunction–associated steatotic liver disease (MASLD)
  - telomere defects in, 3820, 3822f
  - thyroid hormone levels in, 3039
  - toxic, 2665, 2667t
  - tumors. *See* Liver tumors
  - in type 2 DM, 3202
  - in Wilson's disease, 2630, 3338
- "Liver flap," 2716
- Liver flukes
  - cholangiocarcinoma and, 667
  - clinical features of, 1736, 1830–1831, 1831t
  - diagnosis of, 1830–1831, 2737, S12
  - epidemiology of, 1826t, S12
  - transmission of, 1826t
  - treatment of, 1830t, 1832
- Liver function tests, **2633**
  - abnormal, evaluation of, 2630–2631, 2631f, 2633, 2634f
  - blood ammonia, 2636
  - coagulation factors, 2636
  - global considerations in, 2637
  - serum albumin. *See* Albumin, serum
  - serum bilirubin. *See* Bilirubin, serum
  - serum enzymes, **2634**. *See also* Alanine aminotransferase (ALT); Alkaline phosphatase (ALP); Aspartate aminotransferase (AST); γ-glutamyl transpeptidase (GGT)
  - serum globulins, 2636
  - shortcomings of, 2633
  - urine bilirubin, 321, 2634, 2637t, 2638
  - use of, 2636–2637, 2637t
- Liver Imaging Reporting and Data System (LI-RADS), 661
- Liver pain, 2628
- Liver phosphorylase deficiency (type VI GSD, Hers disease), 3360t, **3362**
- Liver phosphorylase kinase deficiency (type IX GSD), 3360t, **3363**
- LiverTox website, 2668
- Liver transplantation, **2718**
  - for acute intermittent porphyria, 3347
  - for alcohol-associated liver disease, 2700, 2726
  - for ATTR amyloidosis, 900
  - for autoimmune hepatitis, 2696
  - for branching enzyme deficiency, 3362
  - candidate selection for, 2719, 2720t
  - in children, 2718–2719, 2718t
  - for cholangiocarcinoma, 668
  - complications of, **2722**
    - graft rejection, 2723–2724, A13
    - hepatic, 2722t, 2723
    - infections, 1158, 1159t, 2723. *See also* Solid organ transplant recipient, infections in
    - nonhepatic, 2722t, 2723–2724
    - recurrence of primary disease, 2724–2726
  - contraindications to, 2719–2720, 2719t
  - for Crigler-Najjar syndrome, 2640
  - deceased-donor selection for, 2720
  - for erythropoietic protoporphyria, 3351
  - frequency of, 2718
  - for glucose-6-phosphatase deficiency, 3359
  - HAV recurrence following, 2724
  - HBV reinfection/recurrence following, 2683, 2724–2725
  - for HCC, 664–665, 664t
  - HCV treatment before or after, 2692–2693, 2725–2726
  - HDV reinfection following, 2684
  - for hemochromatosis, 3337
  - history of, 2718
  - immunosuppressive therapy for, 2721–2722
  - indications for, 2718, 2718t, 2724–2725f
  - living-donor, 2720–2721
  - for MASLD, 2707
  - MELD score and, 2633
  - for metastatic NETs, 684
  - Milan criteria for, 664
  - orthotopic, 2718
  - outcome of, 2718, 2724–2726
  - PELD system in, 2633
  - quality of life after, 2726
  - surgical technique for, 2721, 2721f
  - timing of, 2718
  - for Wilson's disease, 3340
- Liver tumors, **659**
  - benign, 670
  - cholangiocarcinoma. *See* Cholangiocarcinoma
  - fibrolamellar HCC, 670
  - hepatoblastoma, 670
  - hepatocellular carcinoma. *See* Hepatocellular carcinoma (HCC)
  - metastatic. *See* Metastatic disease, to liver
- Liver X receptor, 2728
- Living will, 79
- Lixisenatide, 3211t, 3213
- *LKB1* gene mutations, 617
- LKM antibodies, 2656, 2686, 2695
- LLMs (large language models), 74
- LLO (listeriolysin O), 968, 1227
- *LMF1* gene mutations, 3240
- *LMNA* gene mutations, 2004t, 3787, 3788t, 3870
- *LMO-2* gene, 3826
- LMWH. *See* Low-molecular-weight heparin (LMWH)
- *LMX1B* gene mutations, 2427
- *LMYC* gene mutations, 521
- *LNK* gene mutations, 819
- Loading dose, 485, 485f
- *Loa loa* (African eye worm)/loiasis
  - characteristics of parasite, 1820t, 1824
  - clinical features of, 1738t, 1739t, 1824, 1824f
  - diagnosis of, 961t, 1824, S12
  - epidemiology of, 961t, 1736, 1738t, 1820t, 1824
  - pathogenesis of, 1736, 1824
  - prophylaxis for, 1824
  - treatment of, 1824
- Lobar hemorrhage, 230, 3454. *See also* Intracerebral hemorrhage
- Lobectomy
  - lung, 621
  - temporal, 3425
- Lobeline, 3712t
- Lobular carcinoma in situ (LCIS), 633, 636
- Local anesthetics, 266, 3715t
- Local heat urticaria, 2807
- Locked-in state, 189, 191, 3434
- Lockjaw (trismus), 1231
- Locus coeruleus, 3654, 3654f
- Locus control region, 771, 3780
- Locus (nonallelic) heterogeneity, 3787–3788, 3788t, 3801
- Lod score, 3794
- Loeys-Dietz syndrome, 1860, 2166, 2167, 3331–3333, 3332t
- Lofexidine, 3691
- Löffler's endocarditis, 462, 2015
- Löffler's (Loeffler's) syndrome, 462, 1736, 1739t, 2234, 2234t

- Löfgren syndrome, 2927, 2928
- Logopenic primary aphasia, 204
- LOH (loss of heterozygosity), 522, 523f
- *Lomentospora* infections, 1726t, 1729–1730
- Lomitapide, 3244t, 3252
- Lomustine (CCNU), 562t, 563, 718
- Lonafarnib, 2684
- Loncastuximab tesirine, 535t, 566, 566t, 863
- Lone Star tick, 1460, 3730
- Long-chain fatty acids/triglycerides, 2541, 2549t
- Long COVID syndrome
  - autonomic dysfunction in, 1936
  - cognitive impairment in, 119, 1541
  - exercise intolerance in, 181
  - insomnia in, 215
  - ME/CFS in, 3649
  - pain syndromes in, 1541
- Longevity. *See* Life expectancy
- Longevity dividend, 3869
- Long QT syndrome
  - arrhythmia in, 1913
  - clinical features of, 160, 1969
  - congenital, 493, 1969–1970, A8
  - ECG in, 1969f, A8
  - etiology of, 1970t
  - genetic considerations in, 160, 1969–1970, 1970t
  - genetic testing for, 3804t
  - sudden cardiac death in, 2335, 2335t
  - treatment of, 1974–1975, 2340t, 3804t, 3805
- Long telomere syndrome, 3820
- Long terminal repeats (LTRs), 1551, 1559, 1561f
- *Lonomia obliqua,* 3737
- Loop diuretics
  - actions of, 347, 2133, 2145t, 2367f, 2369
  - adverse effects of, 355, 356, 374
  - for CKD, 2389
  - for heart failure, 1990
  - for hypercalcemia, 365, 3284t
  - for hypercalcemia of malignancy, 738
  - for hyperkalemia, 361
  - for hypertension, 2145t
  - for hypokalemia, 357
  - for hyponatremia, 351
  - for pulmonary edema, 2331
- Loop of Henle
  - disorders involving, 2368t
  - embryologic development of, 2364
  - functions of, 2367f, 2369
  - in water absorption/excretion, 345, 345f
- Looser's zones, 3271
- Loperamide
  - for diarrhea, 307, 312, 508, 564, 566, 576
  - for fecal incontinence, 2583
  - for IBS, 2575, 2577t
  - for microscopic colitis, 309
  - for travelers' diarrhea, 1081t
- Lopinavir, 1619f, 1623f
- LOPS (loss of protective sensation), 3204
- Loratadine, 2808
- Lorazepam
  - for acute alcohol intoxication, 3679
  - for anxiety disorders, 3660t
  - for delirium, 86, 86t
  - for dyspnea, 83t
  - for insomnia, 86, 216
  - for lumbar puncture analgesia, S3
  - for nausea and vomiting, 83, 299, 299t
  - overdosage/poisoning with, 3713t
  - pharmacology of, 3660t
  - for status epilepticus, 2298, 3426f
  - during terminal weaning, 89
- Lordotic posture, 3634, 3638f
- Lorlatinib, 533t, 568t, 570, 628, 628t, 753t
- Losartan
  - adverse effects of, 2144t
  - for aortic aneurysm, 2167
  - for aortic regurgitation, 2044
  - drug interactions of, 489t
  - genetic variations in response to, 493, E72t
  - for heart failure, 1995t
  - for hypertension, 2144t
  - metabolism of, 485, 485t, 493
  - for Raynaud's phenomenon, 2874
- Loss aversion, 3916–3917, 3916t
- Loss function, 3972, 3973t
- Loss of heterozygosity (LOH), 522, 523f
- Loss of protective sensation (LOPS), 3204
- Lottery-based reward systems, for health behavior, 3918, 3919, 3919f, 3921
- Louse-borne relapsing fever
  - clinical features of, 1446
  - diagnosis of, 1446–1447
  - epidemiology of, 1445, 1445t
  - prevention of, 1448
  - prognosis of, 1448
  - treatment of, 1447–1448, 1447f
- Louse-borne typhus. *See* Epidemic (louse-borne) typhus
- Lovastatin, 1742t, 2940t, 3244t. *See also* Statins
- Lovotibeglogene autotemcel, 2447
- Low- and middle-income countries
  - health challenges in
    - brain drain, 3848
    - burden of disease, 3844, 3846t, 3860–3861, 3860f
    - cancer, 3851
    - cardiovascular disease, 1856–1857, 1856f, 3850–3851
    - child mortality, 3860–3861, 3861f
    - diabetes, 3850
    - Ebola virus disease, 3849–3850
    - health system performance, 3862, 3862f
    - HIV/AIDS, 1565–1568, 1565f, 3847–3848, 3847f
    - life expectancy, 3860, 3861f, 3862
    - malaria, 3849
    - population displacement, 3752
    - rheumatic heart disease, 2854, 2855f
    - skews in research, 3862
    - social stratification, 3862
    - TB, 3848–3849
  - primary care in, **3859**
    - experiences with, 3865f, 3866
    - opportunities to build, 3866
    - revitalization of, 3863–3864
  - smoking patterns in, 1857
- Low back pain. *See also* Back pain
  - categorization of, 118–119, 118t
  - electrodiagnostic testing in, 126
  - history and physical examination in, 120, 121t, V8
  - imaging in, 122–123, 3393
  - mixed type, 125t
  - natural course of, 119
  - nociceptive/mechanical, 119–120, 119f, 124t
  - radicular, 119, 120, 124t
  - "red flags" in, 120–122, 126t
  - treatment of, **126**
    - interventional pain procedures, 127–128
    - multimodal and interdisciplinary strategies, 126–127
    - pharmacotherapy, 126
    - physical therapy, 127
    - psychological, 126
    - with radicular symptoms, 128
    - selective nerve root blocks, 126, 3393
- Low-density lipoprotein (LDL) apheresis, 3243, 3252
- Low-density lipoprotein cholesterol (LDL-C)
  - atherosclerosis and, A10
  - composition of, 3237f, 3238
  - elevated levels of. *See* Hypercholesterolemia (elevated LDL-C)
  - estimate of, 3239
  - low levels of, 3241t, 3242t, 3247–3248. *See also* Abetalipoproteinemia
  - oxidized, 2790t
  - in transport of hepatic lipids, 3238–3239, 3239f
- Lower esophageal sphincter (LES), 293, 2507, 2508f
- Lower motor neuron diseases, 731, 3522t, 3524. *See also* Amyotrophic lateral sclerosis (ALS)
- Lower motor neuron pathways, 170f
- Lower motor neuron (neuropathic) weakness, 168, 168t, 169
- Lower urinary tract symptoms (LUTS), 3174–3175
  - in benign prostatic hyperplasia, 703, 703t, 2454. *See also* Benign prostatic hyperplasia (BPH)
  - hematuria in, 2359
  - in IC/BPS, 330, 333
- Low-fat diet, 511, 634
- Low FODMAP diet, 301
- Low-molecular-weight dextran, 767t
- Low-molecular-weight heparin (LMWH), **948**
  - actions of, 946f, 948–949
  - adverse effects of, 949
  - for antiphospholipid antibody syndrome, 2839
  - assessment of, 475
  - discontinuation before lumbar puncture, S3
  - dosage of, 949
  - for DVT/PE treatment, 3902
  - vs. fondaparinux, 949t
  - vs. heparin, 948t
  - monitoring treatment with, 949
  - for NSTE-ACS, 2111
  - pharmacology of, 949
  - for STEMI, 2110t, 2121
  - thrombocytopenia induced by. *See* Heparin-induced thrombocytopenia
  - for VTE/PE treatment, 742, 2162, 2163t
  - for VTE prophylaxis, 2164t, 3910
- Loxapine, 3673t
- *LOX* gene mutations, 3332t
- Lp(a). *See* Lipoprotein(a)
- LPL. *See* Lipoprotein lipase (LPL)
- LPL (lipoprotein lipase) deficiency, 3240, 3827
- LPS. *See* Lipopolysaccharide (LPS)
- LPS-binding protein, 2758
- LPS receptor. *See* CD14 (LPS-receptor)
- Lp-X, 3246
- *LRBA* gene mutations, 2786f
- LRP4 antibodies, 3625
- *LRP5* gene mutations, 3293
- *LRRK2* gene mutations, 2554t, 3400, 3497, 3498t, 3499
- *LSC1* gene mutations, 2183
- LSD, 3696
- LT4. *See* Levothyroxine (LT4)
- *LTA4H* gene, 983
- LTRs (long terminal repeats), 1551, 1559, 1561f
- LT-β, 2766t
- Lubiprostone, 81, 313, 314, 2576, 2577t, 3549
- Lubricin, 2844, 2949
- Lucey-Driscoll syndrome, 2639
- Lucio's phenomenon, 1051, 1409
- ^{177}Lu-DOTA-octreotate, 571t, 684t, 685
- ^{177}Lu-dotatate, 735
- Ludwig's angina, 260, 1375
- Lues maligna, 1430, 1602
- Lujo virus, 1658t
- Lumacaftor, 2247, E72t
- Lumateperone, 3673t
- Lumbar disk herniation, 120, 124t
- Lumbar disk surgery, 128
- Lumbar fractures. *See* Vertebral fractures
- Lumbar lordosis, 326
- Lumbar medial branch nerve blocks, 127

Lumbar nerve root decompression, 128
Lumbar puncture, **S3**
  analgesia for, S3
  in coma, 191
  headache following, 116, S3
  imaging and laboratory studies prior to, S3
  in meningitis, 1120
  in meningococcal infections, 1249
  in neurologic critical illness, 2332
  in patient receiving anticoagulants or antiplatelet agents, S3
  positioning for, S3
  in subarachnoid hemorrhage, 3459
  technique, S3
Lumbar spinal lesions, 3560
Lumbar spine, 122f
Lumbosacral plexopathies, 3615, 3616t
Lumbosacral plexus, 3615, 3616f
Lumbo virus, 1659t
Lumefantrine, 1741t, 1749, 1771t
Luminal bile salt deficiency, 2542f, 2543, 2543t
Luminal digestion, 2540, 2542. *See also* Malabsorption
Lumpectomy, 636
Lung
  airways resistance in, 2207
  blood flow through, 2201
  capacity of, 2204
  compliance of, 2202, 2296
  elastic recoil pressure of, 2202
  flow-related properties of, 2202–2203, 2203f, 2204
  shunt, 2204, 2205f
  volume-related properties of, 2201–2202, 2202f, 2203f, 2206–2207
Lung abscess, **1034**
  anaerobic bacterial, 1375, 1376f
  approach to the patient, 1036
  clinical features of, 1034
  complications of, 1036, 1136
  definition of, 1034
  diagnosis of, 1034
  differential diagnosis of, 1034
  epidemiology of, 1034
  etiology of, 1034, 1034t, 1375
  hemoptysis in, 274
  pathogenesis of, 1034–1035, 1034t
  pathology and microbiology of, 1034t, 1035
  in pneumonia, 1029
  prevention of, 1036
  prognosis of, 1036
  treatment of, 1035–1036
Lung allocation score (LAS), 2283
Lung biopsy
  in granulomatosis with polyangiitis, A14
  in hypersensitivity pneumonitis, 2230, 2230f
  in interstitial lung disease, 2262, 2264f
  in microscopic polyangiitis, A14
  in vasculitic syndromes, A14
Lung cancer, **614**
  adenocarcinoma, 615–616, 616f, 617f
  airway obstruction in, 592
  asbestos-related, 2236
  carcinoma of unknown primary and, 733, 733t
  classification of, 615–616, 616f
  clinical features of, 618–619, 618t
  clubbing in, 279
  deaths from, 497t, 499f, 500t, 501t
  diagnosis of, 620, A12
  epidemiology of, 614–615
  genetic considerations in, 520t, 615–617, 617f, 617t
  hemoptysis in, 275, 591
  immunohistochemistry of, 616
  incidence of, 497t, 498, 498f, 500f, 501t
  inherited predisposition to, 615
  large cell, 615–616, 616f
  lymphadenopathy in, 477
  metastatic
    to bone, 730
    clinical features of, 619, 619t, 2970
    to CNS, 723f, 723t
    to eye, 232
    to heart, 2087, A9
    incidence of, 619
    to skin, 406
    treatment of
      cytotoxic chemotherapy, 629, 630t
      immunotherapy, 628–629, 629f, 630t
      molecular targeted therapy, 532, 627–628, 627–628t, 628f
      second-line therapy, 629
      supportive care, 629–630
  non-small-cell
    biomarker testing in, 620
    genotyping for, 3982
    metastatic, 723f
    pathology of, 615–616, 616f
    prognosis of, 623f
    staging of, 620–621, 621f, 622t
    treatment of
      ablative therapies for early-stage, 2289
      chemotherapy in stages I and II, 624–625, 625t
      evolution of, 623
      for occult and stage 0 carcinomas, 622
      for Pancoast tumors, 626
      professional guidelines references for, 624t
      radiation therapy in stages I and II, 622–623
      for solitary pulmonary nodule and "ground-glass" opacities, 622–623, 623t, 624f
      in stage III, 625–626
      in stage IV. *See* Lung cancer, metastatic
      surgical, 623, 626–627
      survival, 623f
      treatment of, **622,** 623f
    tumor markers in, 504t
  occupational exposures and, 2240
  paraneoplastic syndromes in, 619, 737t, 740t
  pathogenesis of, 616–617
  pathology of, 615–616, 616f
  prevention of, 512–513
  racial/ethnic disparities in treatment of, 62
  radon exposure and, 2240
  risk factors for, 615
  screening for, 516t, 518, 617–618, 618t, 3881t
  second malignancies in survivors of, 755
  small-cell
    hypokalemia in, S1
    immunohistochemistry of, 616, 2790t
    intestinal pseudoobstruction in, 298
    Lambert-Eaton myasthenic syndrome in, 619, 744, 3628
    paraneoplastic syndromes in, 737, 737t, 743–744t, 746, 3624
    pathology of, 615–616, 616f
    SIAD in, 348, S1
    staging of, 620–621
    treatment of, 631f
      chemotherapy, 630
      prophylactic cranial irradiation, 631
      surgery, 630
      thoracic radiation therapy, 630–631
  smoking and, 511, 615, 3683t
  squamous cell, 615–616, 616f
  staging of, 620–622, 621f, 622t
  stem cells in, 617
  superior vena cava syndrome in, 586f, 618t, 619, 2272
  in women, 3168
Lung disease. *See also* Respiratory disorders; *specific disorders*
  air pollution and, 2240
  barotrauma-related, 3766
  biomass smoke-induced, 2241, 2241f
  cancer. *See* Lung cancer
  cyanosis in, 278
  in cystic fibrosis, 2245. *See also* Cystic fibrosis (CF)
  deaths from, 75t
  global considerations in, 2241
  high-altitude travel and, 3760
  IgG4-related, 2931t
  indoor exposures and, 2241
  in leptospirosis, 1442f
  occupational. *See* Occupational illness, lung disease
  in osteogenesis imperfecta, 3324
  parasitic infections, 1739t
  pulmonary hypertension in, 2192
  respiratory acidosis in, 375
  in rheumatoid arthritis, 2840
  in sarcoidosis, 2924–2925, 2925–2926f
  in systemic sclerosis, 2865–2866, 2865f, 2868–2869, 2869f
Lung flukes (*Paragonimus* spp.)
  clinical features of, 1739t, 1831t, 1832
  diagnosis of, 961t, S12
  epidemiology of, 274–275, 961t, 1736, 1739t, 1826t, 1831, S12
  hemoptysis caused by, 274–275
  life-cycle hosts of, S12
  treatment of, 1753, 1830t, 1832
Lung injury, acute, 2301f, 2323. *See also* Acute respiratory distress syndrome (ARDS)
Lung lavage, S6
Lung transplantation, **2283**
  contraindications to, 2283–2284, 2283t
  for COPD, 2258
  for cystic fibrosis, 2248, 2283
  donor considerations, 2285, 2285t
  donor management, 2285
  donor operation, 2285–2286
  for idiopathic pulmonary fibrosis, 2263, 2283, 3822
  immunosuppression for, 2286
  indications for, 2283
  infections in recipient, 1158, 1159t, 1162, 1164. *See also* Solid organ transplant recipient, infections in
  outcomes of, 2287t
  perioperative considerations and complications, 2286–2287
  recipient management, 2284–2285, 2287
  recipient operation, 2286
  in systemic sclerosis, 2874
Lung volume(s), 2201–2202, 2203f, 2206–2207, 2251
Lung volume reduction surgery, 2258
Lungworm, rat. *See Angiostrongylus* spp. infections
Lupus anticoagulants
  in antiphospholipid syndrome, 2837
  laboratory evaluation of, 475, 2832
  in thrombotic disease, 936
  warfarin therapy and, 952
Lupus erythematosus
  acute cutaneous, **414,** 415f
  clinical features of, 406
  discoid. *See* Discoid (chronic cutaneous) lupus erythematosus
  drug-induced, 387–388, 420, 489, 493, 2827, 2835–2836, 2940t
  neonatal, 2836
  subacute cutaneous, 405, 414, 420, 420f, A1
  systemic. *See* Systemic lupus erythematosus (SLE)

- Lupus hairs, 394
- Lupus nephritis
  - classification of, 2418, 2418t, 2829, 2831f
  - clinical features of, 2415t, 2417–2418, 2439
  - pathogenesis of, 2412, 2417
  - prognosis of, 2418
  - renal biopsy in, 2418, A4
  - treatment of, 2418, 2834–2835, 2834f
- Lupus panniculitis, 405
- Lupus pernio, 406, 2927, 2927f
- Lupus vulgaris, 406
- Lurasidone, 3668, 3673t
- Lurbinectedin, 562t, 564, 630
- Lusitropic effect, 1850f
- Luspatercept, 777, 818
- Lutein, 2611
- Luteinizing hormone (LH), **2991**
  - actions of, 2982, 2991, 3127–3128
  - deficiency of, 2996t
  - ectopic production of, 737t
  - expression and regulation of, 2988t, 2989f
  - feedback control of, 2987
  - in females, age-related changes in levels of, 3126–3127, 3127f
  - GnRH and, 2987
  - laboratory evaluation of, 2995t, 3107
  - in male pubertal development, 3103–3104
  - in menstrual cycle regulation, 3129, 3129f
  - in ovarian follicle development, 3126
  - ovulation and, 3129
  - in perimenopause, 3142, 3142f
  - secretion of, 212, 2991, 3126
  - synthesis of, 2991
  - testosterone deficiency and, 3118
- Luteinizing hormone (LH) receptor, 2978, 2983, 2985t, 3789
- Luteinizing hormone-releasing hormone (LHRH) agonists, 537, 567
- Luteinizing hormone (LH) therapy, 3119
- 177-Lutetium, 574, 685, 735
- 177-Lutetium PSM-617, 703
- LUTS. *See* Lower urinary tract symptoms (LUTS)
- ^{177}Lu-vipivotide tetraxetan, 571t
- Lycopene, 2612
- Lyme borreliosis (Lyme disease), **1448**
  - climate change and, 3749, 3751f
  - clinical features of, 139t, 394, **1450**
    - arthritis, 1058, 2790t
    - AV conduction block, 1450, 1926–1927, A8
    - chronic meningitis, 1130t
    - disseminated infection, 1450
    - erythema migrans, 139t, 394, 1101–1102, 1449f, 1450, A1
    - evolution of, 139t, 145, 394, 1450
    - facial palsy, 3554
    - localized infection, 1450
    - myocarditis, 2013
    - neuropathy, 3523, 3606
    - persistent infection, 1450–1451
  - co-infection with *Babesia,* 1452, 1778
  - diagnosis of, 976, 1451–1452, 1451t, A16
  - differential diagnosis of, 146, 1452
  - epidemiology of, 1449
  - etiology of, 1448–1449
  - global considerations in, 1449
  - pathogenesis of, 1449–1450
  - posttreatment syndrome, 1451, 1453
  - prevention of, 1453, 3731
  - prognosis for, 1453
  - regional variations in, 1451
  - reinfection in, 1453
  - ticks in, 1449, 3731
  - treatment of, 1058, 1451t, 1452–1453, 1453f
- Lymphadenitis, 1387, 1387f, 1500
- Lymphadenopathy, **476**
  - approach to the patient, 476–477
  - benign, 476
  - diseases associated with, 477t
  - in fever of unknown origin, 150
  - in Hodgkin's lymphoma, 869
  - IgG4-related, 2931t, 2933
  - infections associated with, 476–477, 477t, 959–960
    - *Bartonella* spp., 1351t
    - *Brucella,* 1333
    - cat-scratch disease, 1351, 1352f
    - EBV, 1509
    - HIV infection, 1574, 1604
    - human African trypanosomiasis, 1794
    - nontuberculous mycobacterial, 1416
    - plague, 1343, 1344f
    - tularemia, 1338, 1338f
    - *T. whipplei,* 1369
  - laboratory evaluation of, 478
  - localized/regional, 476–477
  - in sarcoidosis, 2928
  - in Sezary syndrome, 394
  - with splenomegaly, 477
  - thoracic/abdominal, 477
- Lymphangiectasia, intestinal. *See* Intestinal lymphangiectasia
- Lymphangiogenesis, therapeutic, 2185
- Lymphangiography, 2184
- Lymphangioleiomyomatosis (LAM), 2260, 2261, A12
- Lymphangitis, 1211, 2183, 2184
- Lymphatic filariasis, **1819**
  - clinical features of, 1059, 1820–1821, 1820f, 2183, S7
  - diagnosis of, 1821, 1825v, S7, S12
  - differential diagnosis of, 1821
  - epidemiology of, 1736, 1819, 1820t
  - incubation period for, S7
  - pathogenesis of, 1736, 1819–1820
  - prevention and control of, 1822
  - treatment of, 1059, 1747, 1821–1822, S7
- Lymphatic obstruction, 282, 309
- Lymphedema, **2182**
  - after melanoma treatment, 606
  - anatomic considerations in, 2182–2183
  - clinical features of, 2183–2184, 2184f
  - congenital, 2183
  - diagnosis of, 2184
  - differential diagnosis of, 2184
  - etiology of, 2183, 2183t
  - in lymphatic filariasis, 1820, 1820f, 2183
  - pathophysiology of, 282, 2183
  - stages of, 2183–2184, 2184t
  - treatment of, 606, 1822, 2184–2185
- Lymphedema praecox (Meige's disease/syndrome), 2183, 3510
- Lymphedema tarda, 2183
- Lymph node biopsy, 477, 478, 620
- Lymphocutaneous syndrome, 1360
- Lymphocyte(s)
  - autoreactivity of, 2824–2826
  - B. *See* B cell(s)
  - endothelial cell interactions of, 2781–2782, 2782f
  - in hemophagocytic lymphohistiocytosis, 464, 465f
  - large granular, 443, 457. *See also* Natural killer (NK) cells
  - in nutrition assessment, 2618t
  - in peripheral blood smear, 442f, 443
  - reactive, 443
  - T. *See* T cell(s)
  - tumor-infiltrating, 2785
- Lymphocytic choriomeningitis (LCM)
  - clinical features of, 1059, 1126, 1665, 1667
  - diagnosis of, 1124, 1126, 1131t, 1667
  - reservoirs and vectors of, 1658t
  - risk factors for, 1131t
- Lymphocytic colitis, 2561
- Lymphocytic gastritis, 2539
- Lymphocytic hypophysitis, 2993
- Lymphocytic meningitis, 1132t
- Lymphocytic variant hypereosinophilia, 878t, 879
- Lymphocytoma cutis, 405, 406
- Lymphogranuloma venereum (LGV)
  - clinical features of, 1108f, 1108t, 1472, 2183
  - complications of, 1472
  - diagnosis of, 1473t
  - epidemiology of, 1097, 1107, 1472
  - etiology of, 1472
  - lymphadenopathy in, 477
- Lymphohistiocytosis, hemophagocytic, 146, 2803, S9
- Lymphoid hyperplasia, cutaneous (pseudolymphoma), 394, 405
- Lymphoid malignancies, **871**. *See also* Hodgkin's lymphoma; Leukemia; Non-Hodgkin's lymphoma; Plasma cell disorders
  - approach to the patient, 860–862
  - B-cell
    - B-cell differentiation and, 858–859, 859f
    - B-cell prolymphocytic lymphoma, 872
    - hairy cell leukemia. *See* Hairy cell leukemia
    - immunophenotypes of, 872t
    - intravascular large B cell lymphoma, 873
    - lymphomatoid granulomatosis, 873–874, 2989
    - mediastinal (thymic) large B cell lymphoma, 873
    - nodal marginal zone B cell lymphoma, 865, 873
    - primary effusion lymphoma, 873, 1516, 1613
    - splenic marginal zone B cell lymphoma. *See* Splenic marginal zone lymphoma
    - treatment of, 3824t
  - carcinoma of unknown primary and, 733t
  - classification of, 857t, 872t
  - CNS, primary. *See* Primary central nervous system lymphoma (PCNSL)
  - deaths from, 497t
  - diagnosis of, 860–861
  - differential diagnosis of, 1511t
  - EBV infection and, 1510
  - erythrophagocytosis in, A6
  - etiology of, 858t, 1510
  - extranodal NK/T cell lymphoma, nasal type, 857, 867–868, **874,** 2898
  - fever in, 136, 150
  - genetic considerations in, 858–860, 860t
  - hemophagocytic lymphohistiocytosis and, 468
  - in HIV infection, 1613–1614, 1613f, 1614f
  - incidence of, 497t
  - infections in, 578t
  - lymphadenopathy in, 476
  - mediastinal, 2272
  - neuropathy in, 3607
  - nodular follicular, A6
  - non-Hodgkin's subtypes. *See* Non-Hodgkin's lymphoma, subtypes
  - ocular invasion by, 232
  - oral manifestations of, 263t, 1613f
  - paraneoplastic syndromes in, 736, 737t
  - proteinuria in, 342
  - in rheumatoid arthritis, 2841
  - in Sjögren's disease, 2877, 2877t
  - skin manifestations of, A5
  - small intestinal, 651
  - splenomegaly in, 480
  - staging of, 861, 861t
  - T cell
    - extranodal T/NK-cell lymphoma, nasal type, 2898

    hepatosplenic T-cell lymphoma, 860t, 867, 874–875, 2566
    large granular lymphocytic leukemia, 2841
  T-cell and NK cell
    aggressive NK cell leukemia, 874
    angioimmunoblastic T-cell lymphoma, 860t, 867, 875
    blastic NK cell lymphoma, 875
    enteropathy-type T-cell lymphoma, 867, 874
    extranodal NK/T-cell, nasal type. *See* Extranodal NK/T cell lymphoma, nasal type
    hepatosplenic T-cell lymphoma, 860t, 867, 874–875
    primary cutaneous CD30+ T cell lymphoma, 875
    subcutaneous panniculitis-like T cell lymphoma, 867, 875
    T-cell differentiation and, 859–860, 860f
    T-cell large granular lymphocytic leukemia, 874
    T-cell prolymphocytic leukemia, 874
  in thymus. *See* Thymoma
  in thyroid, 3048
  in transplant recipient, 2351
  tumor markers in, 504t
  urticaria in, 148
Lymphoid tissue, mucosa-associated. *See* Mucosa-associated lymphoid tissue (MALT)
Lymphoma. *See* Lymphoid malignancies
Lymphoma cutis, 405, 406
Lymphomatoid granulomatosis, 873–874, 2898
Lympho-occlusive bandage, for snakebite, 3718
Lymphoplasmacytic lymphoma, 857, 860t, 865
Lymphoplasmacytic sclerosing pancreatitis. *See* IgG4-related disease (IgG4-RD)
Lymphoscintigraphy, 2184
Lymphotactin/SCM-1, 2767t
Lynch syndrome (hereditary nonpolyposis colon cancer)
  associated lesions in, 654, 654t
  bladder cancer and, 690
  bladder cancer in, 690, 691
  characteristics of, 654
  clinical features of, 654
  early interventions for, 3804t, 3805
  endometrial cancer in, 714
  gastric adenocarcinoma in, 646
  genetic considerations in, 524t, 654
    APC tumor-suppressor gene, 520f, 523
    DNA mismatch/repair, 3780, 3785
    MLH1 gene mutations, 523, 524t, 654, 671t, 3785–3786
    MSH2 gene mutations, 654
    somatic mutational steps, 520f
  genetic testing for, 3804t
  ovarian cancer in, 710
  polyp distribution and characteristics in, 654t
  screening for, 654, 3786
  small-bowel adenocarcinoma in, 651
  type II, 710
Lyonization, 932
Lysine metabolism disorders, 3369t
Lysine methylation, 3937
Lysinuric protein intolerance, 2368t, **3373,** 3374t
Lysosomal acid lipase deficiency (cholesteryl ester storage disease), 466, 3241t, 3244–3245, 3354t, **3357**
Lysosomal storage diseases, **3351**
  genetic considerations in, 3355
  glycogen disorders, 3354t. *See also* Pompe disease (acid α-1,4 glucosidase deficiency)
  glycoproteinoses, 3353t
  GM_2 gangliosidoses, 3352–3353t, 3522t. *See also* Tay-Sachs disease
  leukodystrophies, 3353–3354t
  mucolipidoses, 3353t
  mucopolysaccharidoses, 3352t, 3356–3357, A15
  neutral glycosphingolipidoses
    Fabry's disease. *See* Fabry's disease
    Gaucher disease. *See* Gaucher disease
    Niemann-Pick diseases, 3353t, 3356
  neutral lipid disorders, 3354t
  pathogenesis of, 3351, 3355
  treatment of, 3355
Lysosome(s), 969
Lysozyme, 969
Lyssaviruses, 1650, 1655. *See also* Rabies
LYST gene mutations, 466
LYST protein, 460

M
MacArthur Competence Assessment Tool-Treatment, 3877
Machado-Joseph disease (SCA3), 3535, 3793t
Machine learning, **3970**
  applications to clinical medicine, 3972–3973
  ethical issues in, 74
  medical, 3971–3972, 3972t, 3973f
  natural language processing and, 3973–3974
  practical concepts for, 3972, 3973f
  types of, 3970–3971, 3971f, 3971t
Machine operator's lung, 2228t
Machupo/Bolivian hemorrhagic fever, 1658t, 1675
Maciel virus, 1661t, 1674
MAC infections. *See Mycobacterium avium* complex (MAC) infections
Macitentan, 2194, 2197t
MACOP-B regimen, 873
Macroamylasemia, 2754
Macrocytosis, 438f, 445f, 449, 450
Macroglossia, 265t, 897, 897f
Macrolides. *See also specific drugs*
  actions of, 1166, 1177, 1182t
  adverse effects of, 426, 1171t, 1177, 1744t
  for community-acquired pneumonia, 1027t, 1028
  drug interactions of, 1172t, 1744t
  indications for, 1177
  for *Legionella* spp. infections, 1275t
  for pertussis, 1279, 1280t
  resistance to
    in *F. tularensis,* 1339
    in group A streptococci, 1054
    manifestations of, 1026
    mechanisms of, 1182t, 1184
    in *M. pneumoniae,* 1027, 1466
    in pneumococci, 1026, 1192
    prevalence of, 1208–1209
    in *T. pallidum,* 1439
Macronutrient malnutrition, 2620–2621. *See also* Malnutrition
Macro-ovalocytes, 438f, 447, 447t, 450
Macrophage(s)
  in adaptive immunity, 2763t, 2764f
  in autoimmune disease, 2824
  brain, 3394–3396
  development of, 2764f
  functions of, 461
  in HIV infection, 1580
  in innate immunity, 2760, 2763t
  in iron recycling, 763
  in pneumonia, 1023
  in sepsis/septic shock, 2317, 2318f, 2319
  in smokers, 2251
  in TB, 1384–1385
Macrophage activation syndrome, 469, 594
Macrophage colony-stimulating (M-CSF), 2767t
Macrophage-defined chemokine (MDC), 2767t
Macrophage inflammatory proteins, 455, 1585, 2767t
Macroreentrant atrial tachycardia, 1943, 1945f
Macrothrombocytopenia, 922f
Macula, 220, 222
Macula densa, 2364, 2365f
Maculae ceruleae, 3732
Macular amyloidosis, 404
Macular degeneration, 231, 231f
  complementary and integrative therapies for, 3926
  gene therapy for, 3827
  prevention of, 2611
  stem cell therapy for, 3407–3408
Macular hole, 232
Maculatum disease, 1455t
Macules, 137, 264t, 377t, 378f
Maculopapular cutaneous mastocytosis, 2820
Madariaga virus, 1657t
Madrid virus, 1659t, 1667
Madura foot, 1725
Madurella, 1725
Mafenide acetate, S5
Maffucci's syndrome, 406, 3318
MAG (myelin-associated glycoprotein), 893, 2790t
Maggots, 3732
MAGIC syndrome, 2920
Magnesium, 2601t, 3266
  deficiency of. *See* Hypomagnesemia
  excess of. *See* Hypermagnesemia
  renal transport of, 2367f, 2369
Magnesium citrate, 82t
Magnesium hydroxide, 82t, 2524, 2575, 2577t
Magnesium oxide, 509
Magnesium sulfate, 1231, 3899
Magnesium supplements, 243, 3301
Magnesium-wasting syndromes, 3266, 3266t
Magnetic resonance angiography (MRA)
  in frostbite, 3772
  in neurologic disease, 3389–3390
  in peripheral artery disease, 2172, 2173f
  phase-contrast, 3390
  of renal arteries, 2150, 2151t
  of thoracic aneurysm, 2167f
  time-of-flight, 3389–3390, A16
  vessel wall, 3390, 3391f
Magnetic resonance cholangiopancreatography (MRCP)
  in bile duct disease, 2489, 2494f, 2736t
  in cholestatic disease, 324
  in liver disease, 2631
  in pancreatic disease evaluation, 2738, 2738f, 2739t, 2741
  in primary sclerosing cholangitis, 2712
Magnetic resonance elastography, 2631t, 2632, 2636
Magnetic resonance imaging (MRI)
  in adrenocortical carcinoma, 3066f
  in AVP deficiency, 3018, 3019
  in axial spondyloarthritis, 2885, 2885f
  for body composition study, 2618t
  for breast cancer screening, 516t, 517, 634
  in CAH, 3072f
  in cardiac disease. *See* Cardiac magnetic resonance imaging (CMR)
  in coma, 191
  contraindications to, 3389, 3390t
  in Crohn's disease, 2559, 2559–2560f
  in Cushing's syndrome, 3057–3058, 3059f
  in echinococcosis, 1837f
  FLAIR. *See* Fluid-attenuated inversion recovery (FLAIR) MRI
  in HCC diagnosis of, 661
  in hearing loss evaluation, 250
  in liver disease, 2631–2632, 2631t
  in low back pain, 122
  in lymphadenopathy, 478
  in lymphedema, 2184

Magnetic resonance imaging (MRI) (*Cont.*):
  in mastoiditis, 255
  in mediastinal masses, 2271–2272
  in Ménière's disease, 247
  in musculoskeletal disease evaluation, 2947, 2947t, 2948f
  in neck pain, 130
  in neurologic disease. *See also* Neuroimaging, MRI
    acute disseminated encephalomyelitis, A16
    acute transverse myelitis, A16
    adrenoleukodystrophy, A16
    amyotrophic lateral sclerosis, 3520f
    arteriovenous malformation, A16
    Bell's palsy, 3555, 3555f, A16
    brachial plexopathy, A16
    brain abscess, 1136, 1137f
    brain metastases, 723, 723f
    brain tumors, 716, 718f, 719f, 3392f, A16
    CADASIL, A16
    Candida infection, in brain, A16
    cavernous malformation, 3389f
    cerebral abscess, 3388f
    cerebral arteriopathy, 3391f
    cerebral small-vessel disease, 3490f
    cobalamin deficiency myelopathy, A16
    coccidiomycosis meningitis, A16
    cranial epidural abscess, 1141
    Creutzfeldt-Jakob disease, 3530, 3531f
    dementia, 197, 198f
    DLB, 3492t
    encephalitis, 1113–1114, 1113f, 1114t, 3388f
    epidural metastases, 725f
    FTD, 3486f
    Guillain-Barré syndrome, A16
    Hashimoto's encephalopathy, A16
    hemiplegic migraine, A16
    hepatic encephalopathy, A16
    histiocytosis, A16
    Huntington's disease, 3513f, A16
    hyperperfusion disorder, 2350, 2350f
    hypoxic-ischemic encephalopathy, 2347f
    intracerebral hemorrhage, 3455
    intramedullary astrocytoma, 3563f
    Krabbe disease, A16
    lacunar infarction, A16
    Leigh's syndrome, A16
    leptomeningeal metastases, 725f
    limbic encephalitis, A16
    Lyme disease, A16
    melanosis, A16
    meningioma, 188f, 722f
    meningitis, 1120, 1133, 1134f
    mesial temporal lobe epilepsy, 3413f
    metachromatic leukodystrophy, A16
    migraine after radiation therapy, A16
    moyamoya disease, A16
    MS, 3583, 3584f, A16
    MSA, 3548, 3549f
    MS variants, 3589f, A16
    multiple system atrophy, A16
    myopathy, 3638
    neurocysticercosis, 1139f
    neurofibromatosis type 1, A16
    neurofibromatosis type 2, A16
    neuromyelitis optica, 3590, 3591f, A16
    neurosarcoidosis, A16
    neurosyphilis, A16
    osmotic demyelination syndrome, 2349, 2349f
    PCNSL, 720, 720f
    pilocytic astrocytoma, A16
    PML, 1117
    seizures, 3418
    sickle cell disease, A16
    spinal arteriovenous malformation, 3567
    spinal cord infarction, 3565, A16
    spinal epidural abscess, 3563–3564, 3563f
    stroke, 3437, 3438f
    Sturge-Weber syndrome, A16
    subdural empyema, 1140f
    subdural hematoma, A16
    superior sagittal sinus thrombosis, A16
    syringomyelia, 3568f
    thoracic meningioma, 3563f
    traumatic brain injury, 3574
    tuberculosis, A16
    tuberous sclerosis, A16
    vascular dementia, 3491
    vasculitis, A16
    vertebral fracture/dislocation, A16
    vestibular schwannoma, 722, 722f
    von Hippel–Lindau disease/syndrome, A16
  in osteomyelitis, 1062, 1063f
  in pancreatic disease, 2738, 2738f, 2739t, 2741
  in pheochromocytoma, 3074, 3074t, 3075f
  of pituitary tumors, 2998–2999, 3000f, 3002f, 3018, A16
  in polymyositis, 2912f
  in prolactinoma, 3006, 3007f
  in prostate cancer staging, 698
  in psoriatic arthritis, 2888
  in RCC, 688
  in respiratory disease evaluation, 2212, 2213f
  in rheumatoid arthritis, 2847–2848
  in sarcoidosis, 2926
  of sellar masses, 2998–2999
  in sinonasal aspergillosis, A16
  in spinal cord compression, 589, A16
  in viral encephalitis, 1113f, 3388f
  in Wernicke's disease, 2349f
Magnetic resonance neurography, 3390
Magnetic resonance spectroscopy, 3977
Magnetic sphincter augmentation, 302
Magnetoencephalography (MEG), 3418
Maguari virus, 1659t
Mahaim fibers, 1941
Ma huang *(Ephedra sinica),* 3414t, 3926
Maintenance dose, 487
Majocchi's granuloma, 395
Major depression, 3664–3665, 3665t, 3666f, 3967. *See also* Depression/depressive disorders
Major histocompatibility complex (MHC). *See* Human leukocyte antigen (HLA) complex
MAL (methyl aminolevulinate), 430
Malabsorption disorders, **2540**
  approach to the patient, **2549**
  classification of, 2541t
  clinical features of, 2549, 2549t, 3537
  diagnosis of, 2547f, 2549–2550, 2549t, 2550t
  in luminal phase of digestion
    bile acid diarrhea, 2543, 2543t
    with disordered intestinal motility, 2542
    fatty acid diarrhea, 2543t
    in hepatobiliary disease, 2543
    with ileal resection or disease, 2543, 2543t
    luminal bile salt deficiency, 2542f, 2543, 2543t
    in pancreatic disease, 2542
    small-bowel bacterial overgrowth, 300, 309, 311, **2543,** 2870, 2874
    surgery-related, 2534, 2542
  in mucosal phase of digestion
    abetalipoproteinemia. *See* Abetalipoproteinemia
    celiac disease. *See* Celiac disease
    environmental enteric dysfunction, 2546
    glucose-galactose malabsorption, 2545
    lactase deficiency. *See* Lactase deficiency
    short-bowel syndrome, 369, **2547**
    sucrase-isomaltase deficiency, 2545
    tropical sprue, 309, 787, **2546**
    Whipple's disease. *See* Whipple's disease
  pathophysiology of, 2549t
  protein-losing enteropathy, **2548**
Maladie de Roger, 287
Malaria, **1760**
  in children, 1765–1766, 1765t
  climate change and, 3748, 3749f
  clinical features of, **1763,** 1764t, S7
    acidosis, 369, 1764
    encephalopathy, 991t, 994, 1763
    fever, 136, 1740t, 1763
    hematologic abnormalities, 1765
    hypoglycemia, 1764, 1773
    jaundice, 326
    liver dysfunction, 1765
    pulmonary edema, 1764, 1773
    renal failure, 1764–1765, 1773, 2416t, 2429
    retinal hemorrhage, 1763, 1765f
    seizures, 1738t, 1763
    splenomegaly, 1739t
  complications of, 1738t, 1765, 1765t, 1766, 1773
  diagnosis of, S7, S12
    alternative procedures, 1769t
    blood films of parasites, 437, 1766–1767, 1766–1769f, 1769t, A2, S12
    vs. *Brucella* infection, 1333
    serologic and molecular tests, 1767–1768, 1769t, S12
  epidemiology of, 1761–1762, 1761f, 3748, 3750, S7, S12
  erythrocyte changes in, 1762
  etiology of. *See Plasmodium* spp.
  global considerations in, 1761f, 1773, 3748, 3849
  in HIV infection, 1575, 1611, 1765
  host defense mechanisms in, 1762–1763
  incubation period for, S7
  pathogenesis of, 1760–1761, 1761f
  in pregnancy, 1765, 1765t
  prevention of
    chemoprophylaxis, 1773–1775, 1774t. *See also specific drugs*
    in children, 1774, 1774t
    indoor residual spraying, 3849
    insecticide-treated bed nets, 3849
    in international travelers, 1012t, 1017–1018
    organizational efforts, 3849
    personal protection, 1773
    in pregnancy, 1773–1774
    in travelers, 1774–1775, 1774t
  prognosis of, 1764t
  resistance to, 776, 797, 1763, 3795
  in returned traveler, 1737
  transfusion-transmitted, 913t, 1766
  transmission of, 1760–1761, 1761f
  treatment of
    antimalarial drugs, 1741–1742t, 1771t. *See also specific drugs*
    drug resistance and, 1768, 3849
    regimens for, 1770t, 3849
    in severe disease, 1768–1772, 1770t
    in uncomplicated disease, 1770t, 1772–1773
  vaccine, 1773
  in war veterans, S7
Malassezia spp. infections, 388t, 389, 395, 1722–1723, S11. *See also* Tinea versicolor
Malathion, 3732
Maldonado virus, 1662t
Male feminization, paraneoplastic, 737t
Male pseudohermaphroditism. *See* 46,XY disorders of sex development

- Male reproductive system disorders, **3107**
  - in adulthood, **3111**
    - aging-related. *See* Testosterone deficiency
    - androgen insensitivity syndromes. *See* Androgen insensitivity syndrome (AIS)
    - diagnosis of, 3107–3108
    - erectile dysfunction. *See* Erectile dysfunction (ED)
    - gynecomastia. *See* Gynecomastia
    - hypogonadism. *See* Hypogonadism
    - hypogonadotropic hypogonadism. *See* Hypogonadotropic hypogonadism
    - infertility, 3107, 3150f, 3151, 3171
    - obesity and, 3187
  - cancer. *See specific organs*
  - delayed puberty, 3109t, 3110
  - precocious puberty, 3789
  - sex chromosome disorders. *See* Sex chromosome disorders
  - testicular. *See* Testes/testicles, disorders of
  - 46,XY disorders. *See* 46,XY disorders of sex development
- Maleylacetoacetate isomerase deficiency, 3368t
- MALFDI-TOF MS. *See* Matrix-assisted laser desorption/ionization/time-of-flight mass spectometry (MALFDI-TOF MS)
- Malic dehydrogenase, 1749
- Malignant conversion, 431
- Malignant fibrous histiocytoma, 727. *See also* Soft tissue sarcomas
- Malignant hypertension, 2153–2154, 2412
- Malignant hyperthermia, 493, 3804t
- Malignant otitis externa, 254
- Malingering, 3670
- Mallenkot catheter, 2585
- Mallory-Weiss tears
  - bleeding in, 317, 2513
  - diagnosis of, 2482, 2489f
  - hematemesis in, 298
  - treatment of, 317, 2482, 2513, V5
- Malnutrition, **2615**
  - acute disease- or injury-associated, 2550t
  - in AKI, 2383, 2384
  - assessment of. *See* Nutrition assessment
  - in cirrhosis, 2717
  - in CKD, 2394
  - climate change and, 3741, 3743–3744f
  - diagnosis in clinical settings, 2616, 2616f, 2616t
  - diagnosis of, 2621–2622
  - fatigue in, 166
  - gonadotropin deficiency due to, 3113
  - grading of, 2616t
  - gut microbiome and, 3835–3836
  - hypothermia in, 3768
  - in malabsorption disorders, 2549t
  - micronutrient, 2620–2621
  - protein-calorie, 2616
  - risk of, 2603
- Malocclusion, 261
- MALT (mucosa-associated lymphoid tissue), 2780
- Maltese cross formation, 2425, A4
- MALT lymphoma. *See* Mucosa-associated lymphoid tissue (MALT) lymphoma
- Malt worker's lung, 2228t
- Mammalian target of rapamycin. *See* mTOR
- *Mammarenavirus,* 1656, 1657–1658t
- Mammary souffle, 289, 1866
- Mammography
  - in carcinoma of unknown primary, 732
  - screening, 38, 38t, 39t, 515, 516t, 634. *See also* Breast cancer, screening
- Managed care programs, 6
- MANAs (mutation-associated neoantigens), 526
- Manganese, 2601t, 2614t, 2615, 3702
- Manganese-56 exposure, S6
- Mania, 215, 3668, 3669t. *See also* Bipolar disorder
- Mannitol, 346, 352, 358, 2345t, 3727
- Mannosidosis, 3353t, 3357–3358
- Manometry
  - anorectal, 314
  - esophageal. *See* Esophageal manometry
  - intestinal, 298
- *Mansonella ozzardi,* 1820t, 1824
- *Mansonella perstans,* 1820t, 1824
- *Mansonella streptocerca*/streptocerciasis, 1820t, 1824, S12
- Mantle cell lymphoma, **865**
  - genetic considerations in, 521t, 859, 860t, 866
  - immunophenotype of, 852t, 872t
  - pathogenesis of, 859f
  - prognosis of, 862, 866
  - subsets of, 865–866
  - treatment of, 481, 561, 866, 3824t
- *MAO-A,* 3933
- MAO-B (monoamine oxidase type B) inhibitors, 3503, 3504t
- MAOIs. *See* Monoamine oxidase inhibitors (MAOIs)
- MAPK pathway, 526t, 628, 734, 971, 3045
- Maple syrup urine disease, 3369t
- Ma protein antibodies, 743t, 746
- Maprotiline, 3659t, 3666
- *MAPT* gene mutations, 195t, 3400, 3486, 3487f, 3497, 3522t
- Marantic endocarditis, 150, 408, 619, 1038
- Maraviroc
  - actions of, 1583, 1621
  - adverse effects of, 1618t
  - drug interactions of, 1422
  - genetic variations in response to, 1583, 1591, E72t
  - for HIV infection, 1618t, 1621
  - molecular structure of, 1617f
  - resistance to, 1623f
- Marble bone disease (osteopetrosis), 3314–3315
- Marburg's variant, MS, 3589, 3589f
- March hemoglobinuria, 801, 801t
- Marchiafava-Bignami disease, 3483
- Marcus Gunn pupil, 221
- Marfan's syndrome
  - aortic regurgitation in, 2043
  - cardiac manifestations of, 106, 3332
    - aortic aneurysm, 2166, 2167, 3332
    - aortic dissection, 2169
    - mitral valve prolapse, 2058
    - murmur, 284, 287
  - classification of, 3331
  - clinical features of
    - hand deformities, 1860
    - ocular, 3332
    - oral deformities, 1860
    - skeletal, 50t, 3331–3332, A15
  - diagnosis of, 3332
  - genetic considerations in, 3332
  - genetic testing for, 3804t
  - incidence of, 3331
  - paternal age and, 3784, 3790
  - pregnancy in, 3900
  - treatment of, 3333, 3798, 3804t, 3805
  - vessel wall disorders in, 927
- Margetuximab, 534t, 641
- Marginal zone lymphoma, **865**
  - clinical features of, 865
  - diagnosis of, 865
  - extranodal. *See* Mucosa-associated lymphoid tissue (MALT) lymphoma
  - immunophenotype of, 852t
  - nodal, 865
  - pathophysiology of, 865
  - splenic, 481, 857, 860t, 865, 872, 872t
  - treatment of, 481, 865
  - vs. Waldenström's macroglobulinemia, 893
- Maribavir, 1160, 1486t, 1487, 1516
- Marijuana. *See* Cannabis/marijuana
- Marine envenomations, **3722**
  - antivenoms for, 3723, 3725
  - approach to the patient, 3725
  - epidemiology of, 3722, 3723f
  - invertebrate, 3722–3725, 3723f, 3724f
  - treatment of, 3725
  - vertebrate, 3725, 3726f
- Marine poisoning, **3726**
  - ciguatera, 306, 3726–3727, 3727t
  - domoic acid intoxication, 3414, 3727–3728
  - scombroid, 306, 3726
  - shellfish, 3414, 3727–3728
- Maripa virus, 1661t, 1674
- Marituba virus, 1659t, 1667
- Markesbery-Griggs distal myopathy, 3640t, 3643–3644
- Maroteaux-Lamy disease (arylsulfatase B deficiency), 3352t, 3357
- Marrara, 3733
- Marshall's syndrome, 3331
- MASH. *See* Metabolic dysfunction–associated steatohepatitis (MASH)
- MASLD. *See* Metabolic dysfunction–associated steatotic liver disease (MASLD)
- Massage therapy, 3925t, 3928
- Massively parallel sequencing. *See* Next-generation sequencing
- Massive transfusion, 912
- Mass movements, intestinal, 304
- Mass spectrometry, 3976–3977, 3976f
- Mass spectrometry imaging, 3979, 3979f
- Mast cell(s)
  - activation of, 2804–2805, 2805f
  - in adaptive immunity, 2763t
  - in allergy, 2805–2806, 2806f
  - in anaphylaxis, 2812
  - distribution of, 2805
  - hematopoietic differentiation of, 760f
  - in innate immunity, 2763t, 2770
  - malignancies of, 872t. *See also* Mastocytosis
  - membrane-derived lipid mediators from, 2805, 2805f
  - non–IgE mediated activation of, 422, 428f
  - secretory granules of, 2805
  - in urticaria pigmentosa, 406
- Mast cell leukemia, 2820
- Mast cell sarcoma, 2820
- Mastectomy, 514, 553–554, 634, 636
- Master's disease (Southern tick-associated rash illness), 139t, 145, 1452
- Mastitis, 1199, 1390, 1648, 1708
- Mastocytoma, 2820
- Mastocytosis
  - cutaneous, 396, 402, 406, 880, 2819–2820, 2820t
  - epidemiology of, 2819
  - small-intestinal mucosal biopsies in, 2550t
  - systemic
    - aggressive, 880, 2820, 2820t
    - classification of, 2819–2820, 2820t
    - clinical features of, 291, 300, 308, 880–881, 2820
    - with clonal hematologic non–mast cell lineage disease, 2820
    - definition of, 2819
    - diagnosis of, 880–881, 2820–2821t, 2821
    - differential diagnosis of, 2821
    - genetic considerations in, 2820
    - indolent, 2820
    - prognosis of, 881
    - smoldering, 2820, 2820t
    - treatment of, 881, 2821
    - well-differentiated, 2821

Mastoiditis, 255, 1119
MATE1/2, 493
Maternal inheritance, 3807, 3807f. *See also* Mitochondrial DNA diseases; X-linked disorders
Maternal mortality, 3162, 3904
Maternal spindle transfer, 3817f
Matonaviridae, 1478f, 1480t. *See also* Rubella (German measles)
Matrix-assisted laser desorption/ionization/time-of-flight mass spectometry (MALFDI-TOF MS)
  for cerebral arteriopathy, A16
  characteristics of, 3389–3390, 3977
  for microbial identification, S11
    Brucella spp., 1333
    C. auris, 1709, 1710t
    F. tularensis, 1339
    gram-negative bacteria, 1282
    HACEK group, 1266
    Yersinia spp., 1349
Matrix-M, 997t, 998
Matrix metalloproteinase 8, 2187
Matrix metalloproteinase 9, 1384
Matrix metalloproteinase 12, 2251
Matrix metalloproteinases (MMPs)
  in cancer cell signaling, 542f
  in photoaging of skin, 431
  in rheumatoid arthritis, 2846
  in tumor angiogenesis, 546, 548f
Mat telangiectasias, 396
Maturity-onset diabetes of the young (MODY), 3195, 3202–3203, 3783t, 3793, 3794t
Mavacamten, 2007
MAX gene mutations, 3078–3079, 3079t
Maxillary sinusitis, 65
Maximum-tolerated dose (MTD), 555
Mayaro virus, 1059, 1657t
Mayer-Rokitansky-Kuster-Hauser syndrome, 3101, 3133
May-Hegglin anomaly, 924
Mazzotti reaction, 1747
MBC (minimal bactericidal concentration), S11
MBI (mild behavioral impairment), 194
MBL gene, 1582t
MBP (myelin basic protein), 2790t
MBTPS2 gene mutations, 3324, 3325t, 3326
MC1R gene mutations, 430
MC4R gene mutations, 3182, 3185
McAdam's criteria, relapsing polychondritis, 2920
McArdle disease (muscle phosphorylase deficiency, type V GSD), 3361t, **3364,** 3644
McBurney's point, 2595
McConnell's sign, 1883, 2161
McCune-Albright syndrome
  clinical features of
    adrenal tumors, 3063
    Cushing's syndrome, 3056
    cutaneous, 399, 3109, 3317
    endocrinologic, 3004, 3088
    fibrous dysplasia, 3004t, 3317
    phenotype and, 3791
    precocious puberty, 3109
  diagnosis of, 3317, 3317f, A16
  genetic considerations in, 3280f
    GNAS1 gene mutations, 3056, 3088, 3317, 3783
    $G_s\alpha$ mutations, 2978, 2984, 3004, 3043, 3080t, 3088, 3109, 3317
    mosaicism, 3791
  treatment of, 3317
M cells, 969
MCH (mean corpuscular hemoglobin), 444, 445t
MCI (mild cognitive impairment), 193, 3886, 3888f
MCIR gene mutations, 599–600
McLeod blood group, 440f, S9
McLeod syndrome, 3514
McMurray test, 2944
M component
  in heavy chain diseases, 894
  of immunoglobulin, 883
  in multiple myeloma, 887–888, 892
  in plasma cell disorders, 883
  in Waldenström's macroglobulinemia, 893
MCP (monocyte chemotactic protein), 1585, 2767t, 3656
mcr-1, 1185
M-CSF (macrophage colony-stimulating factor), 2766t
MCT8 gene mutations, 3024
MCTD. *See* Mixed connective tissue disease (MCTD)
MCV (mean corpuscular volume), 437, 444, 445t, 450, 765
MDC, 2767t
MDCT (multidetector computed tomography), 3384. *See also* Computed tomography (CT)
MDM4 gene mutations, 520t
MDMA (ecstasy), 184, 349, 3693, 3696
MDPV, 3696
MDR3 gene mutations, 2729
MDRD (Modification of Diet in Renal Disease) equation, 337, 2388t, S10
MDR (multidrug resistance) gene/phenotype, 541, 574
MDS. *See* Myelodysplastic syndromes (MDS)
Mean arterial pressure, in shock, 2292–2293
Mean corpuscular hemoglobin (MCH), 444, 445t
Mean corpuscular hemoglobin concentration, 444, 445t
Mean corpuscular volume (MCV), 437, 444, 445t, 450, 765
Measles (rubeola), **1638**. *See also* Rubella
  approach to the patient, 1640
  atypical, 402
  clinical features of, 138t, 1534, 1640–1641, 1640f
    in impaired immunity, 1641
    Koplik's spots, 137, 402, 1534, 1641, A1
    rash, 137, 138t, 146f, 402, 1641, A1
  complications of, 1117–1118, 1641–1642, 3594
  diagnosis of, 1641
  differential diagnosis of, 1641
  epidemiology of, 1639
  eradication of, 1642
  etiology of, 1639
  global considerations in, 1003, 1638–1639, 1642
  immune response to, 1640, 1640f
  incidence pre- and postvaccine, U.S., 1002t
  incubation period for, 1482
  mortality rates of, 993
  outbreaks in North America since 2010, 14, 15t
  pathogenesis of, 1639–1640, 1640f
  prevention of, 1642. *See also* Measles-mumps-rubella vaccine (MMR)
  prognosis of, 1642
  transmission of, 1639
  treatment of, 1641
  vitamin A deficiency in, 2611
Measles and Rubella Partnership, 1639
Measles-mumps-rubella vaccine (MMR)
  adverse effects of, 1642, 1649
  contraindications for, 1006t
  coverage with, 14
  effectiveness of, 1649
  precautions for, 1006t
  recommended administration of
    in adults, 40t, 1004t, 1645
    in children, 1642, 1645, 1649
    in health care workers, 1153t
    in international travelers, 1014, 1014t
    in pregnancy, 3904
  storage and handling of, 1007
Measles vaccine, 1642
Measles virus, 1481–1482, 1534, 1639. *See also* Measles (rubeola)
Mebendazole
  actions of, 1749
  adverse effects of, 1742t, 1749
  for ascariasis, 1815
  for capillariasis, 1819
  drug interactions of, 1742t
  for enterobiasis, 1818
  for hookworm infections, 1816
  indications for, 1742t, 1815t
  pharmacology of, 1749
  in pregnancy and lactation, 1742t
  for trichinellosis, 1812t
  for trichuriasis, 1818
mecA gene, 1204
Mecamylamine, 3570
ME/CFS. *See* Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS)
mec gene, 981, 1027, 1183
Mechanical circulatory support, 2328–2329, 2328f, A11. *See also* Left ventricular assist systems
Mechanical heart valve. *See* Prosthetic heart valve
Mechanical ptosis, 233
Mechanical ventilation, **2304**
  for ARDS, 2302, 2303t
  for asthma attacks, 2227
  in coma, 192
  complications of
    airway, 2307
    alveolar collapse, 2302
    cardiopulmonary, 2307
    lung injury, 2294–2295, 2302
    pneumonia. *See* Ventilator-associated pneumonia (VAP)
    respiratory acidosis, 375
    tension pneumothorax, 2271
    volutrauma, 2294–2295
  continuous positive airway pressure. *See* Continuous positive airway pressure (CPAP)
  for COPD, 2259
  indications for, 2304
  liberation from, 2308–2309, 2308f
  modes of, 2305–2306, 2305t
  patient management during, 2295–2296, 2302–2303
  principles of, 14–15, 2304–2305f
  for pulmonary edema, 2330
  for respiratory acidosis, 376
  for sepsis/septic shock, 2323
  for shock, 2294, 2315
  strategies to optimize gas exchange in, 2306–2307
  withdrawal of, 88–89
Mechanistic target of rapamycin (mTOR). *See* mTOR
Mechanoreceptors, in lung, 267
Mechanotransduction, 244
Mechlorethamine (nitrogen mustard), 512t, 726t, 3115, S5
Meckel-Gruber syndrome, 2430t, 2431f
Meckel's diverticulum, 318, 320
Meclizine, 83, 165t, 298, 299t, 300
Meclofenamate, 923t
Meconium ileus, 2245
MECP2 gene mutations, 3403, 3783, 3941
Médecins Sans Frontières, 3847
Medial degeneration, 2165–2166
Medial epicondylitis, 2975

Median neuropathy. *See* Carpal tunnel syndrome
Median sternotomy, 3615
Mediastinal germ cell tumor, 709
Mediastinal (thymic) large B cell lymphoma, 873
Mediastinal lymphadenopathy, 477, 478
Mediastinal mass, 291, 2271–2272, 2272f. *See also* Thymoma
Mediastinitis
  acute, 2272
  anaerobic bacterial, 1375
  chronic, 2272–2273
  fibrosing, 1694, 1696, 2272–2273
  sternal osteomyelitis and, 1067
  in thoracic organ transplants, 1159t
Mediastinoscopy, 2208
Mediastinotomy, 2208
Mediastinum, 2271, 2271f, 2272t
Media, vaccine hesitancy and, 16
Medical consultation
  curbside, 3897–3898
  involving advanced practice providers, 3898
  involving telemedicine and EHRs, 3898
  in pregnancy. *See* Pregnancy
  second opinions, 3898
  stakeholder responsibilities in, 3897, 3897t
Medical decision-making, 3–4, 66
Medical education, 7, 73
Medical errors, 4–5, 49–51, 73. *See also* Adverse events
Medical evacuation insurance, 1021
Medical practice, 1, 1f. *See also* Physician
  accountability in, 7
  clinical skills and, 1–2
  ethical issues in. *See* Ethical issues
  evaluation of outcomes, 4
  expanding frontiers of, 6–8
  globalization of, 6–7. *See also* Global medicine
  humane aspects of, 6
  Internet and, 7
  legal issues in. *See* Legal issues
  patient care management in, 2–5
  patient-physician relationship in, 5–6
  practice guidelines in, 4
  public expectations of, 7
  research and, 8
  science and art of, 1
  team-based care and, 2–3
Medical professionalism, 6
Medical records, electronic, 5
Medical tourists, 1022t
Medication-overuse headache, **115,** 3469
Medicinal leech, 3733
Medicine, **1**
  art of, 1
  evidence-based, 3–4, 27–29
  practice of. *See* Medical practice
  science of, 1, 7
Meditation, 3924, 3925t, 3927
Mediterranean fever, familial. *See* Familial Mediterranean fever
Mediterranean lymphoma, 894
Mediterranean spotted (boutonneuse) fever, 991t, 1455t, 1458
Medium-chain fatty acids, 2541, 2549t
Medium-chain triglycerides, 2542
Medroxyprogesterone
  adverse effects of, 3176
  for breast cancer, 567
  for contraception, 3154
  for hot flashes in cancer treatment, 510
  for hot flushes with androgen deprivation therapy, 3176
  for menorrhagia in perimenopause, 3142
  for PCOS, 3135
  for postmenopausal hormone therapy, 3148
Medulla, 3433f
Medullary cystic kidney disease, 2435
Medullary sponge kidney, 2436
Medullary stroke syndromes, 3433, 3433f
Medullary thyroid carcinoma (MTC), **3085**
  clinical features of, 308
  diagnosis of, 3085, 3086f
  familial, 3048, 3085
  genetic considerations in, 3085
  in MEN 2, 3048, 3077, 3085
  paraneoplastic syndromes in, 737t
  treatment of, 312, 3048, 3085
  tumor markers in, 504t
Medulloblastomas, 721
Mees' lines, 3484
Meester-Loeys syndrome, 3331
Mefenamic acid, 3137, 3142
mef gene, 1026
Mefloquine
  actions of, 1749
  adverse effects of, 1744t, 1749, 1771t, 1772, 1775, S8
  contraindications to, 1775
  drug interactions of, 1741t, 1744t, 1749
  for malaria, 1770t, 1771t
  for malaria prophylaxis, 1774t, 1775
  pharmacology of, 1749, 1771t
  in pregnancy, 1744t, 1749, 1775
MEFV gene, 2761t
MEG (magnetoencephalography), 3418
MEG3 gene mutations, 3004
Megacolon, 313, 314, 1791. *See also* Toxic megacolon
Megalin, 2983
Megaloblastic anemia, **782**
  autosomal recessive, 786–787
  biochemical basis of, 782
  clinical features of, 782–784
  diagnosis of, 443, 450
  drug-induced, 790
  etiology of, 781t. *See also* Cobalamin (vitamin B_{12}) deficiency; Folate deficiency
  hematologic findings in, 784–785, 785f
  nutrient deficiencies in, 449
  in orotic aciduria, 790
  peripheral blood smear in, 785f
  thiamine-responsive, 790
  treatment of, 789–790
Megarectum, 314
Megestrol, 84, 359, 509, 510, 593
Meglumine antimonate, 1741t, 1746, 1784
Meibomian gland carcinoma, 224
Meibomitis, 224
Meige's disease (lymphedema praecox), 2183, 3510
Meigs' syndrome, 712, 2270
Meiosis, 3779, 3782f
MEK1 gene mutations, 617f
MEK inhibitors
  action and targets of, 526f, 532, 557f
  adverse effects of, 395, 420, 757, 2017
  for melanoma, 604, 604t
Melamine, 2377
Melanocortin-2 receptor, 2991
Melanocortin 2 receptor, 3052
Melanocortin-2 receptor, 3053f
Melanocortin-4 receptor, 2985t, 3184, 3185t
Melanocyte(s), 430
α-Melanocyte-stimulating hormone (MSH)
  in appetite control, 3182, 3183f
  ectopic ATCH production and, 739
  overproduction of, 400
  sun response in skin and, 430
Melanocytic nevus, 398
Melanoma, **598,** A5
  acral lentiginous, 601t, A5
  amelanotic/hypomelanotic, 599f
  carcinoma of unknown primary and, 733t
  choroidal, 232, 232f, 599f, 601t, 605
  clinical features of, 599f, 601t
  deaths from, 497t
  desmoplastic, 601t
  diagnosis of, 600–601, A5
  early detection of, 600
  epidemiology of, 432, 598
  familial, 524t
  genetic considerations in
    BRAF, 494
    genomic, 601–602, 602f
    germline mutations, 599–600
    oncogenes, 520t
  global considerations in, 598–599
  histologic subtypes, 601t
  incidence of, 497t, 498f
  KIRs with, 2772t
  lentigo maligna, 601t, 602f, A5
  leukoderma in, 397
  lymphadenopathy in, 477
  metastatic, 398, 406, 477, 723t, 2087, A16, S2
  mortality from, 598
  mucosal, 599f, 601t
  nevi and, 598, 599f
  nodular, 599f, 601t, 602f, A5
  oral, 264t
  pathogenesis of, 430, 432, 601–602, 2790t
  prevention of, 600
  prognosis of, 602, 603t
  retinopathy in, 751
  risk factors for, 598, 600t, 2566
  screening for, 518
  sentinel node biopsy in, 603
  staging of, 602–603, 603t
  sun/UV exposure and, 430, 432, 512, 512t, 598, 600
  superficial spreading, 377f, 599f, 601t, A5
  treatment of
    biologic therapy, 494, 550, 560, 561
    clinically localized disease, 603
    follow-up, 606
    immunotherapy, 604–605, 2783
    metastatic disease, 604–606, 604t
    regionally metastatic disease, 603–604
    targeted therapy, 532, 572, 605
    virotherapy, 3830
Melanopsin, 220, 3946
Melanosis coli, 312
Melanosomes, 430
Melarsoprol, 1744t, 1749, 1796, 1796t
Melasma, 378t, 400, 420
MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis, and strokelike episodes) syndrome, 3287, 3646, 3811, 3811t, 3812f, A16
Melatonin
  for anxiety, 3926
  in circadian system, 3944t, 3945
  for cluster headache prevention, 3472t
  for migraine prevention, 3469, 3470t
  for REM sleep disorder, 3493
  in sleep, 212
  for sleep disorders, 218, 3926
MELD (Model for End-Stage Liver Disease) score, 2633, 2720, 2720t
Melena, 316
Meleney gangrene, 1377
Meleney's ulcer, 409
Melflufen, 890t, 892
Melioidosis, 1309t, 1313, S7
Melkersson-Rosenthal syndrome, 3555
Melodic intonation therapy, 202
Melorheostosis, 3316

- Melphalan
  - actions of, 563
  - adverse effects of, 562t, 563, 754, 2667
  - for AL amyloidosis, 898
  - for cardiac amyloidosis, 2027
  - in HCT preparation, 915
  - interactions and issues, 562t
  - for multiple myeloma, 890t, 891
- Memantine, 199, 3470t, 3480, 3491, 3889
- Membrane receptor(s), 7–8, 2983f
- Membrane receptor families, 2981t
- Membrane transport, defects of, **3373,** 3374t
- Membranoproliferative glomerulonephritis
  - autoantibodies in, 2790t
  - clinical features of, 2359, 2415t
  - pathogenesis of, 2412, 2420–2421, 2420f
  - renal biopsy in, 2420, A4
- Membranous glomerulonephritis (nephropathy), 2359, 2412, 2415t, 2423, 2423f, 2423t, A4
- Memory
  - assessment of, 207, 3380
  - immune, 2756
  - immunologic, 2774
  - implicit, 207
  - limbic network for, 207
  - loss of. *See* Amnesia
  - working, 208
- *MEN1/MENIN* gene mutations
  - genetic testing for, 3084
  - in MEN syndromes, 717t, 2535, 2977, 3004, 3080t, 3084, 3276, 3276f
  - NETs and, 677
  - tumors associated with, 524t
  - variable expressivity of, 3789
- Menaquinone, 2613. *See also* Vitamin K
- Menarche, 3130, 3132
- Mendelian disorders, 3789–3791
  - allele frequency and effect size in, 3795f
  - autosomal dominant, 3790, 3790f
  - autosomal recessive, 3790–3791, 3790f
  - genotype-phenotype relationships in, 3958f
  - *Online Mendelian Inheritance in Man,* 3777, 3778t
  - X-linked. *See* X-linked disorders
  - Y-linked, 3789, 3791
- Mendelian susceptibility to mycobacterial disease (MSMD), 2796
- Mendelson syndrome, 1375
- Ménétrier's disease, 300, **2539,** 2549
- Ménière's disease
  - clinical features of, 163, 164, 246, 248
  - diagnosis of, 164, 249
  - epidemiology of, 246–247
  - hearing loss in, 246, 248, 249
  - nystagmus in, 237
  - pathophysiology of, 164, 247
  - treatment of, 164, 247
- Meningeal biopsy, 1135
- Meningeal syphilis, 1432
- Meningiomas
  - clinical features of, 716t, 721
  - epidemiology of, 721
  - hearing loss in, 247
  - imaging in, 188f, 721, 722f, 2999
  - in MEN 1, 3084
  - ocular signs in, 222, 232
  - olfactory groove, 242
  - optic nerve sheath, 232
  - pathophysiology of, 721
  - sellar, 3003
  - thoracic, 3563, 3563f
  - treatment of, 721
- Meningitis
  - bacterial, **1118**
    - *A. baumannii,* 1296
    - clinical features of, 994, 1119–1120
    - complications of, 1218
    - deaths from, 1123
    - definition of, 1118
    - diagnosis of, 960, 962t, 1120–1121, 1121t, S2
    - differential diagnosis of, 1121–1122
    - *E. coli,* 1286
    - enterococcal, 1218, 1219t, 1220t
    - epidemiology of, 992, 1118
    - etiology of, 991t, 994, 1118–1119
    - *F. tularensis,* 1340
    - global considerations in, 992
    - gram-negative, 1023, 1122t
    - *H. influenzae,* 1119, 1261–1262, 1263
    - incidence of, 1118
    - increased ICP in, 1123
    - in infants, 1213
    - leptospiral, 1443
    - *L. monocytogenes,* 1119, 1122t, 1123, 1228
    - meningococcal, 991t, 1119, 1122, 1122t, 1248
    - nontyphoidal *Salmonella,* 1320, 1321t
    - in organ donor, 1160
    - *P. aeruginosa,* 1311
    - pathophysiology of, 1119, 1120f
    - pneumococcal
      - clinical features of, 1193
      - diagnosis of, 1193
      - empirical treatment of, 991t, 994
      - pathophysiology of, 1119
      - treatment of, 1121–1122, 1122t, 1193–1194
    - prognosis of, 1123
    - prophylaxis for exposure to, 1180t
    - risk factors for, 994
    - staphylococcal, 1119, 1122t, 1123
    - streptococcal, 1213
    - *T. pallidum,* 1432
    - treatment of
      - adjunctive, 1123
      - antimicrobial therapy based on pathogen, 1122–1123, 1122t, 1193–1194, 1219t, 1220t, 1263
      - empirical antimicrobial therapy, 991t, 1121t, 1122, 1249, 2321t
      - ICP-related, 1123
    - *Y. pestis,* 1344
  - in cancer patient, 583, 590–591
  - chronic, **1128**
    - amebic, 1758
    - approach to the patient, 1128–1135
    - bacterial, 1130t
    - basal, 1128
    - clinical features of, 1128–1131, 1128t, 1129f
    - CSF analysis in, 1073
    - diagnosis of, 1128–1129, 1132t, 1134f
    - epidemiology of, 1131
    - etiology of, 1129f
    - fungal, 1130t
    - helminthic, 1131t
    - in HIV infection, 1131t, 1135
    - imaging in, 1133
    - in immunosuppressed patients, 1135
    - intracranial, 1128
    - laboratory investigation of, 1134–1135
    - meningeal biopsy in, 1135
    - noninfectious causes of, 1132t
    - protozoal, 1131t
    - spinal, 1128
    - systemic manifestations of, 1128, 1129f
    - treatment of, 1135, 1180t
    - viral, 1131t
  - delirium in, 184
  - eosinophilic, 1737, 1739t, 1813
  - fungal, A16, S2
    - *Candida,* 1712
    - *Coccidioides,* 1130t, 1697, 1698, 1699
    - *Cryptococcus,* 1705, 1706
    - etiology of, 1126
    - in HIV infection, 1606–1607
    - treatment of, 1127
  - genetic considerations in, S2
  - headache in, 114
  - hearing loss in, 246
  - hyponatremia in, 348
  - Mollaret's, 1500
  - neck pain in, 131
  - neoplastic, 590–591
  - in neurosyphilis, 1432
  - pyogenic, 1739t
  - in secondary syphilis, 1431
  - subacute, 1126–1127
  - syphilitic, 1127, 1130t
  - tuberculous, 962t, 983, 1126–1127, 1130t, 1388–1389, 1400
  - viral, **1123**
    - arborvirus, 1125
    - chronic, 1131t
    - clinical features of, 1123
    - diagnosis of, 960, 962t, 1124
    - differential diagnosis of, 1124–1125
    - EBV, 1125
    - enteroviral, 1125, 1634, 1634t
    - epidemiology of, 1124
    - etiology of, 1123–1124, 1124t, 1125–1126
    - in HIV infection, 1125, 1606
    - HSV, 1125, 1131t, 1500, 1503t
    - lymphocytic choriomeningitis. *See* Lymphocytic choriomeningitis (LCM)
    - in mumps infection, 1126, 1648
    - prognosis of, 1126
    - treatment of, 1126
    - VZV, 1125
  - weakness in, 170
- Meningococcal infections, **1244**
  - clinical features of, 1248–1249
    - cutaneous, 402, 991t, 1248, 1248f
    - meningitis. *See* Meningitis
    - meningococcemia/septicemia. *See* Meningococcemia
    - postmeningococcal reactive disease, 1249
    - post-splenectomy sepsis, 990, 991t
  - complications of, 1250
  - definition of, 1244
  - diagnosis of, 1249
  - epidemiology of, 1245–1247, 1246f, 1247f
  - etiology of. *See Neisseria* spp. (meningococci); *Neisseria meningitidis*
  - genetic considerations in, 1245
  - global considerations in, 992, 1245, 1246f
  - pathogenesis of, 1247–1248
  - prevention of. *See* Meningococcal vaccine
  - prognosis of, 1250
  - prophylaxis for contacts, 1252–1253
  - risk factors for, 1246–1247, 1247f
  - treatment of, 1174t, 1249–1250
- Meningococcal vaccine
  - based on subcapsular antigens, 1252, 1252f
  - conjugate, 992, 1250–1252
  - contraindications to, 1007t
  - effects of, 1118
  - global considerations in, 1251–1252
  - polysaccharide, 1250, 1251f
  - recommended administration of
    - in adolescents and adults, 1004t, 1251–1252
    - in children, 1250
    - in complement inhibitor treatment, 3631, 3632t
    - in health care workers, 1153t
    - in international travelers, 1014t, 1015
    - in people with HIV, 1594t
    - in splenectomized patient, 482
  - storage and handling of, 1007

Meningococcemia
  acute, 144t, 148
  chronic, 144t, 148, 1248–1249, A1
  clinical features of, 992, 1248
  complications of, 992, 1248
  epidemiology of, 992
  etiology of, 991t, 992
  fulminant, A1, A5
  pathogenesis of, 1247–1248
  prognosis of, 1250
  shock in, 1248
  skin manifestations of
    papular lesions, A1
    petechiae, 992, 1120
    purpura, 144t, 379f, 408, 1248, A1, A5
    rash, 402
  treatment of, 991t, 1249–1250
Meningoencephalitis
  Balamuthia, 1759–1760
  Cryptococcus, 1705, 1706
  L. monocytogenes, 1228
  primary amebic *(Naegleria),* 1758
  in solid organ transplant recipient, 1160
Meningovascular syphilis, 1432
Menkes' kinky hair syndrome, 2615
Menopause, **3143**
  bone remodeling and, 3295
  cardiovascular disease risk and, 3164
  clinical features of, 3143, 3143f
  definition of, 3142
  health risks following, 3162–3163
  hirsutism after, 3140
  hormone therapy in. *See* Postmenopausal hormone therapy
  in military veterans, S8
  physiology of, 3142, 3142f
  premature, after cancer treatment, 576, 639
  prevalence of, 2979t
  screening/testing for, 2979t, 3142, 3143f
  sexual dysfunction and, 3161
  transition to, 3142–3143, 3143f
Menorrhagia, 472, 926, 3142
Men's health, **3168**
  ALS, 3164
  anabolic-androgen steroid abuse. *See* Anabolic-androgenic steroid (AAS) abuse
  androgen deprivation therapy. *See* Androgen deprivation therapy (ADT)
  breast cancer, 639
  COVID-19 disease, 3177
  leading causes of death, 3163f, 3168
  lower urinary tract symptoms, 3174–3175. *See also* Benign prostatic hyperplasia (BPH)
  muscle dysphoria syndrome, 3172
  nutrient requirements, 2600–2602t
  PD, 3164
  prostate cancer. *See* Prostate cancer
  reproductive and sexual function, 3171–3172. *See also* Erectile dysfunction (ED); Male reproductive system disorders
  sexually transmitted infections, 3177. *See also* Sexually transmitted infections (STIs)
  testosterone and, 3169–3171, 3169f, 3169t. *See also* Testosterone; Testosterone therapy
Menstrual cycle
  duration of, 2987
  follicular phase of, 3129, 3129f
  hormonal integration of, 2986, **3128,** 3129f
  luteal phase of, 3129, 3129f
Menstrual disorders, **3132**
  amenorrhea. *See* Amenorrhea
  definition of, 3132
  diagnosis of, 3132–3133, 3134f
  dysmenorrhea, 3137
  etiology of, 3132–3133
  hypothalamic-pituitary-gonadal axis in, 3132, 3133f
  menorrhagia, 472, 926
  obesity and, 3187
  prevalence of, 3132
  treatment of, 3133–3135
MEN syndromes. *See under* Multiple endocrine neoplasia
Mental disorders. *See* Psychiatric/psychological disorders
Mental status assessment, 3379–3380. *See also* Mini-Mental Status Examination (MMSE)
Men who have sex with men (MSM)
  definition of, 3179t
  epididymitis in, 1099
  gonococcal infections in, 1256
  HBV vaccine for, 1110
  HPV infection in, 1527
  intestinal infections in, 1097t, 1109
  proctitis in, 1109, 1472
  sexually transmitted infections in, 1097, 3177
  syphilis in, 1429, 1431
  urethritis in, 1099
MEONs. *See* Multiple endocrine and other organ neoplasias (MEONs)
MEOS (microsomal ethanol oxidizing system), 2709
Meperidine
  adverse effects of, 3414t, 3648t
  dosage in kidney disease, 488, 2397
  dosage in liver disease, 488
  for migraine, 3469
  overdosage/poisoning with, 3715t
  for pain, 97t
Mephedrone, 3696
Mephenytoin, 485t
Mepolizumab
  adverse effects of, 2896t
  for asthma, 2225–2226, 2793t, 2900–2901
  for eosinophilic granulomatosis with polyangiitis, 2233, 3624
  for hypereosinophilic syndrome, 880, 2233
Meprobamate, 3713t
Meralgia paresthetica (lateral femoral cutaneous neuropathy), 3614, 3903
Mercaptoethane sulfonate (mesna), 563, 597, 757, 916
6-Mercaptopurine (6-MP)
  actions of, 565
  adverse effects of, 457, 562t, 565, 790, 2565
  for ALL, 846f, 847
  for autoimmune hepatitis, 2696
  drug interactions of, 489t, 562t
  genetic variations in response to, 493, 2565, E72t
  for IBD, 2565
  metabolism of, 485t, 493
Mercury exposure/poisoning
  autism and, 3701
  clinical features of, 3484, 3701t
  diagnosis of, 3701t
  global considerations in, 3702
  neuropathy due to, 3610t, 3611
  in pregnancy, 3701
  skin manifestations of, 264t
  treatment of, 3701t
Merkel cell carcinoma
  clinical features of, 607, 609f
  epidemiology of, 431
  natural history of, 607
  pathophysiology and etiology of, 431–432, 606
  treatment of, 609
Merkel cell polyoma virus, 432, 1481
Merogony, 1761
Meropenem
  for *A. baumannii* infections, 1296, 1297t
  actions of, 1183f
  for anaerobic bacterial infections, 1379, 1379t
  for *B. cepacia* infection, 1309t
  for *Campylobacter* infections, 1306
  for community-acquired pneumonia, 963t, 1028t
  for enteric fever, 1318t
  for glanders, 1309t
  for head and neck abscesses, 259
  indications for, 1173t, 1175, 1284
  for intraabdominal infections, 963t
  for MDR-TB, 1394t, 1426
  for melioidosis, 1309t
  for meningitis, 583, 1121t, 1122, 1122t
  for *Nocardia* infections, 1362
  for osteomyelitis, 1065t
  for *P. aeruginosa* infections, 1309t, 1313
  for peritonitis, 1071
  in pregnancy and lactation, 1169t
  resistance to, 1173t, 1181, 1183f
  for sepsis/septic shock, 963t
  for ventilator-associated pneumonia, 1032t
  for Whipple's disease, 1370
Meropenem-vaborbactam
  actions of, 1284
  indications for, 1173t, 1174t, 1175, 1284
  for *P. aeruginosa* infections, 1309t, 1313
  in pregnancy and lactation, 1169t, 1170t
  for pyelonephritis, 1094
  resistance to, 1173t, 1174t, 1175
  for ventilator-associated pneumonia, 1032
Merozoites, *Plasmodium,* 1760
MERRF (myoclonic epilepsy with ragged red fibers) syndrome, 3646, 3811, 3811t, 3814
MERS (Middle East respiratory syndrome), 995, 1539
Mesalamine, 2564, 2564t
Mesangial cells, 2363
Mesangioproliferative glomerulonephritis, 2415t, 2421
Mescaline, 3696
MESD gene mutations, 3325t, 3326–3327
Mesenchymal chondrosarcoma, 730
Mesenchymal stem cells, 3408
Mesenchymal tumor, 737t
Mesenteric fibrosis, 680
Mesenteric ischemia/vascular insufficiency. *See* Intestinal ischemia
Mesenteric lymphadenitis, 112t
Mesenteric venous thrombosis, 2587, 2588t, 2589, 2750
Mesial temporal lobe epilepsy, 3411, 3413f, 3413t
Mesna (mercaptoethane sulfonate), 563, 597, 757, 916
Mesoamerican nephropathy, 2436
Mesothelin, 733t
Mesothelioma
  asbestos-related, 2236
  AV conduction block in, 2086
  carcinoma of unknown primary and, 733t, 735
  immunohistochemistry of, 616
  intramyocardial, 2086
  pleural effusion in, 2270, A12
Messenger RNAs (mRNAs), 1484, 1543, 2788, 3777
Mestranol, 3141
Meta-analysis, 28
Metabolic acidosis, **369**
  acid-base nomogram, 367f
  in AKI, 2382, 2383t, 2384
  alcoholic ketoacidosis, 348, 370
  in CKD, 2389–2390
  clinical features of, 369
  compensatory responses in, 367f

Metabolic acidosis (*Cont.*):
  diabetic ketoacidosis. *See* Diabetic ketoacidosis (DKA)
  drug- or toxin-induced, 370–371. *See also* Ethylene glycol poisoning; Methanol poisoning
  high–anion gap, 369, 369t, S1
  hyperkalemia in, 358, 359
  lactic acid. *See* Lactic acidosis
  in mixed acid-base disorders, 367, 368t
  non–anion gap, 372–373, 372f, S1
  in overdosage/poisoning, 3705
  treatment of, 369
Metabolic alkalosis, **373**
  acid-base nomogram, 367f
  clinical features of, 374–375
  compensatory responses in, 367t
  differential diagnosis of, 373
  etiology of, 373, 374t
  hypokalemia in, 356, S1
  in mixed acid-base disorders, 367, 368t
  pathogenesis of, 373
  treatment of, 375
Metabolic disease, 188, A15. *See also specific diseases*
Metabolic dysfunction–associated steatohepatitis (MASH)
  diagnosis of, 2630t, A13
  diagnostic criteria for, 2700
  differential diagnosis of, 2660, 2695
  HCC and, 659, 660, 660f
  incidence of, 2700
  natural history of, 2700–2701
  outcomes of, 2702
  pathogenesis of, 2701–2702
  treatment of, 2705–2707
Metabolic dysfunction–associated steatotic liver disease (MASLD), **2700**
  cirrhosis due to, 2710
  clinical features of, 2703–2704
  diagnosis of, 2702–2703
  diagnostic criteria for, 2700
  differential diagnosis of, 2660
  epidemiology of, 2700
  fibrosis severity in, 2701
  global considerations in, 2707
  histologic spectrum of, 2700, 2701t
  metabolic syndrome and, 2702, 3256
  natural history of, 2700–2701
  obesity and, 3187
  risk factors for, 2702
  staging and risk stratification in, 2704–2705, 2704f
  treatment of, 2705–2707
Metabolic encephalopathy, 202, 2348
Metabolic equivalent tasks (METs), 2097t, 2099, 2099t
Metabolic memory, 3223
Metabolic syndrome, **3252**
  adiponectin in, 3255
  age and, 3253
  cardiovascular disease and, 1841, 3254, 3256
  clinical features of, 3256
  diabetes mellitus and, 3202, 3254, 3256
  diagnosis of, 3253t, 3256
  dyslipidemia in, 3255, 3258
  epidemiology of, 3252, 3253f
  etiology of, 3254–3255
  global considerations in, 3252
  glucose intolerance in, 3255, 3258
  HCC and, 660
  HCV infection and, 2685
  in HIV infection, 1602
  hypertension in, 3255, 3258
  hyperuricemia/uric acid stones in, 3256
  insulin resistance in, 3254–3255, 3258
  lipodystrophy and, 3254
  MASLD and, 2702, 3256
  obesity/overweight and, 3252, 3256–3257
  obstructive sleep apnea and, 3256
  pathophysiology of, 3254f, A10, A10v
  polycystic ovarian syndrome and, 3256
  prevalence of, 2979t
  proinflammatory cytokines in, 3255
  in psoriasis, 387
  risk factors for, 3252–3254
  screening/testing for, 2979t
  thrombosis risk in, 941
  treatment of, 3256–3258
  waist circumference in, 3253t, 3255
Metabolism
  of drugs. *See* Drug(s), metabolism of
  epigenetics of, 3939–3940
Metabolism modifiers, 571t, 573–574
Metabolites, 3978
Metabolome, 3974, 3975f
Metabolomics, **3974**
  clinical applications in
    magnetic response spectroscopy, 3977
    metabolic measurements in children and adults, 3978
    newborn screening programs, 3977–3978
  definition of, 6, 3974
  emerging clinical applications in
    diagnosis of endocrine disease, A15
    disease biomarker identification, 3978
    drug susceptibility prediction, 3978, 3979f
    pharmacometabolomics, 3978
  emerging techniques for, 3978–3979, 3979f
  sampling considerations in, 3975–3976
  targeted, 3975, 3975f
  technologies for
    chromatography/mass spectrometry, 3976–3977
    nuclear magnetic resonance, 3976, 3976t
  in undiagnosed disease, 3996, 3997t
  untargeted, 3975, 3975f, 3979
Metaboreceptors, 268
Metacarpophalangeal joint, in rheumatoid arthritis, 2839, 2840f, 2848f, 2854v
Metachromatic leukodystrophy (MLD), 3353t, 3485, 3824t, 3827, A16
Meta-effectors, 969
Metagenomics. *See also* Human microbiome
  definition of, 6, 976t
  in infectious disease diagnosis, 977t
  in pathogen discovery, 979–980, 980f
  in undiagnosed disease, 3997t
Metagonimus yokogawai, S12. *See also* Intestinal flukes
Metal fume fever, 2240
Metamorphopsia, 231, 232
Metamyelocytes, 453, 454f
Metanephrines, 3074t
Metapneumovirus, 253, 254, 1533, 1533t
Metastasis-suppressor genes, 541
Metastatic calcification, 3318, 3318t
Metastatic disease, **541**
  to adrenal glands, 593
  in bladder cancer, 694
  to bone, 541, 619, 730–731
  bone marrow examination in, A6
  to brain, **723**
    clinical features of, 716t
    diagnosis of, 723–724, 723f
    epidural, 724–725, 725f. *See also* Spinal cord compression, neoplastic
    increased intracranial pressure and, 590
    leptomeningeal, 724, 724f, S2
    sources of, 619, 723, 723t
    treatment of, 724
  characteristics of, 529t, 541, 542f
  in colorectal cancer, 657
  to eye, 232
  gene therapy for, 3830
  to heart, 1898, 2087, 2087f, A9
  to hypothalamus, 593
  to liver
    ascites with, 328
    colorectal cancer, 657
    NETs, 684
  to lung, 275, 1029
  in lung cancer, 618t, 619
  lymphadenopathy in, 477
  to pericardium, 2083
  to pituitary, 593, 3003
  in RCC, 689–690
  to skin, 394, 397, 406, A5
  to spine, 724–725, 725f, A16
  splenomegaly in, 480
  surgery for, 554
Metastatic pheochromocytoma, 3077
Metastatic potential, 541, 542f
METAVIR scale, liver disease, 2632
Metaxalone, 3713t
Metformin
  actions of, 544, 3211t, 3212
  adverse effects of, 786t, 1904, 2394, 2397, 3211t, 3212–3213
  contraindications to, 3211t, 3225
  genetic variations in response to, 493
  for lifespan enhancement, 3875
  for MASH, 2706
  for metabolic syndrome, 3258
  in older adults, 3882t
  for PCOS, 3135, 3151
  for prediabetes, 3202
  structure of, 3874f
  tumor-related effects of, 544
  for type 2 DM, 3211t, 3212–3213, 3215
MET gene mutations, 617f
  in hereditary papillary RCC, 524t, 525, 687t, 688
  in lung cancer, 617, 617t, 628, 628f
Methacholine, 2223
Methadone
  adverse effects of, 3690
  drug interactions of, 489t
  in opioid-related disorders, 3691
  for pain, 82, 97t, 100
Methamphetamine. *See also* Psychostimulants
  actions of, 3656t
  epidemiology of use of, 3693
  intracerebral hemorrhage due to, 3447, 3454
  overdosage/poisoning with, 3708
  pharmacology of, 3693
  pulmonary hypertension and, 2190
Methane, 3745f, 3746t
Methanol poisoning, 228, 371, 3714t, S1
Methemoglobin, 279, 780
Methemoglobinemia, 780
Methemoglobin inducers, 3714t
Methenamine, 1094
Methenamine hippurate, 1095
Methicillin, 923t, 1027, 1204
Methicillin-resistant *Staphylococcus aureus* (MRSA) infections
  in atopic dermatitis, 383
  cellulitis, 233, 1052
  community-acquired
    epidemiology of, 1196, 1197f
    furunculosis, 380
    life-threatening, 1201
    pathogenesis of, 1198
    pneumonia, 1024, 1027, 1027f, 1028t
  control of, 1152
  in cystic fibrosis, 2248

- daptomycin resistance in, 1185
- diagnosis of, 981
- endocarditis, 1042, 1043t, 1200f
- genomics of, 977t, 981, 984–985, 1027
- health care–associated, 1152
- impetigo, 1051
- in neutropenic fever, 580
- osteomyelitis, 1064t
- prevalence of, 1204
- risk factors for, 1024, 1027
- sepsis/septic shock, 2316
- treatment of, 383, 1028t, 1054, 1054t, 1203–1205, 1203t
- in ventilator-associated pneumonia, 2308

Methimazole
- adverse effects of, 3035, 3093
- for Graves' disease, 3035
- for hyperthyroidism in pregnancy, 3901
- for radioactive iodine exposure, S6
- for thyroid storm, 3037

Methionine, 3939

Methionine aminopeptidase, 1748

Methionine metabolism disorders, 3368t

Methionine synthase, 781, 783f, 784

Methocarbamol, 3713t

Methotrexate
- actions of, 564, 782, 2565
- adverse effects of, 562t, 564, 2565, 2916, 2916t
    - common, 2850t
    - cutaneous, 420
    - folate deficiency, 788
    - hepatic, 2668, 2703t
    - in HIV infection, 1603
    - leukoencephalopathy, 2350t, 2351
    - myelopathy, 726t
    - in pregnancy, 638, 2570, 2853
    - prevention of, 2895
    - pulmonary, 595, 754, 2261
    - renal, 562t, 564
    - seizures, 726t
    - serious, 387t, 2850t, 2896t
    - teratogenic, 2836
    - vasculitis, 426
- for ALL, 846f, 847
- for breast cancer, 637
- for calcium pyrophosphate deposition disease, 2961
- for CNS leukemia, 847
- for dermatomyositis, 414
- for DLBCL, 863
- drug interactions of, 562t
    - NSAIDs, 562t, 564
    - probenecid, 564
    - salicylates, 564
    - sulfonamides, 1743t
    - trimethoprim, 1172t, 1743t
- for ectopic pregnancy, 3136
- for EGPA, 2900
- for gestational trophoblastic disease, 715
- for giant cell arteritis, 2171, 2903
- for granulomatosis with polyangiitis, 2899
- for hand osteoarthritis, 2955
- for IBD, 2565
- for inflammatory myopathies, 2916, 2916t
- for metastatic breast cancer, 641
- for morphea, 416
- for neoplastic meningitis, 591
- for neurocysticercosis, 1835
- in older adults, 2853
- for osteosarcoma, 730
- for PCNSL, 721
- for polymyalgia rheumatica, 2903
- for psoriasis, 387, 387t
- for relapsing polychondritis, 2921
- resistance to, 574
- for rheumatoid arthritis, 2849, 2850t, 2853
- for sarcoidosis, 2929, 2929t
- for SLE, 2834
- for systemic sclerosis, 2873, 2875
- for Takayasu arteritis, 2904
- therapeutic monitoring of, 2850t
- for vasculitis, 2895

2-Methoxy estradiol, 548f

Methyl alcohol, 228

Methyl aminolevulinate (MAL), 430

Methylation cycle, 783f

2-Methylbutyryl-glycinuria, 3370t

Methylcellulose, 2577t

6-Methylcoumarin, 434t

3-Methylcrotonyl glycinuria, 3369t

Methyldopa
- adverse effects of, 282t, 802t, 804, 1920t, 2145t, 2667, 2835, 2940t
- for hypertension, 2145t

Methylene blue, 780, 799t, 1094

5,10-Methylenetetrahydrofolate (5-MTHF), 782, 782t

Methylenetetrahydrofolate reductase (MTHFR), 783f, 784

Methylfolate trap, 782

3-Methylglutaconic aciduria, 3369t

2-Methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency, 3370t

Methylmalonate, serum, 788–789

Methylmalonic acidemia/aciduria, 783, 3369t, 3370t

Methylnaltrexone, 81, 98, 2593

7α-Methyl-19-nortestosterone, 3119

Methylone, 3696

Methylphenidate
- abuse of, 3693–3694
- for depression, 85
- for fatigue, 84, 3588
- for narcolepsy, 213t, 215
- for opioid-induced drowsiness, 81

Methylprednisolone
- adverse effects of, 2916t
- for alcohol-associated liver disease, 2699f, 2700
- drug interactions of, 1742t
- for giant cell arteritis, 2903
- for granulomatosis with polyangiitis, 2898
- for Graves' ophthalmopathy, 3037
- for IBD, 2564
- for inflammatory myopathies, 2916t
- for macrophage activation syndrome, 469
- for MS, 3585
- for MS-associated myelitis, 3565
- for nausea and vomiting, 299t
- for neuromyelitis optica, 3592, 3593
- for optic neuritis, 227
- for PCP, 1734t
- for serum sickness, 813
- for transplant rejection, 2410

Methyl violet, 798

Methylxanthines, 376, 3710t

Methysergide, 2178, 3469, 3710t

Metoclopramide
- adverse effects of, 299, 3497
- for constipation, 82
- drug interactions of, 1742t, 2721
- for gastroparesis, 299, 299t, 3227
- for migraine, 3466, 3467t, 3469
- for MSA, 3549
- for nausea and vomiting, 82, 509, 575

Metoprolol
- adverse effects of, 2144t, 3470t
- for aortic dissection, 2170
- drug interactions of, 1752
- genetic variations in response to, 492, E72t
- for heart failure, 1995t
- for hypertension, 2144t
- for ischemic heart disease, 2100t
- metabolism of, 485t
- for migraine prevention, 3470t
- overdosage/poisoning with, 3712t
- for STEMI, 2118

Metrifonate, 1744t, 1749–1750

Metronidazole
- for acne rosacea, 391
- actions of, 1167, 1182t, 1183f
- adverse effects of, 392, 1171t, 1178, 1744t, 2717, 3609t
- for amebic colitis, 1757, 1757t
- for amebic liver abscess, 1757t, 1758
- for anaerobic bacterial infections, 1379, 1379t
- for bacterial vaginosis, 1101t, 1103
- for bite-wound infections, 1144
- for blastocystosis, 1808
- for *C. difficile* infection, 1085, 1085t
- for clostridial infections, 1239t, 1243
- for *D. fragilis* infections, 1808
- for diverticulitis, 2580
- drug interactions of, 1172t, 1178, 1744t, 3213
- empirical uses of, 2321t
- for epidural abscess, 963t
- for *Giardia* infections, 307, 1805
- in *H. pylori* eradication, 1302t, 2527–2528, 2527t, 2530t
- indications for, 1174t, 1178, 1180t, 1744t
- for intraabdominal infections, 963t
- for intracranial infections, 991t
- for meningitis, 1121t, 1122, 1122t
- for necrotizing fasciitis, 1054t
- for osteomyelitis, 1065t
- for pelvic inflammatory disease, 1106, 1106t
- for peritonitis, 1071
- for pouchitis in IBD, 2565
- in pregnancy and lactation, 1169t, 1744t
- resistance to, 2527
    - in anaerobic bacteria, 1378–1379, 1379t
    - in *Eikenella,* 1268
    - in *H. pylori,* 1303, 2528
    - mechanisms of, 1182t, 1185
    - prevalence of, 1174t, 1178
- for tetanus, 1231
- for *T. vaginalis* infection, 1101t, 1102, 1808

METs (metabolic equivalent tasks), 2097t, 2099, 2099t

Metyrapone, 739, 3012, 3059

Metyrapone test, 2995t

Mevalonate kinase deficiency (hyperimmunoglobulinemia D with periodic fever syndrome), 155, 2935t, 2936

Mexiletine
- adverse effects of, 101, 3604t
- for arrhythmias, 1958
- for hyperkalemic periodic paralysis, 3646
- metabolism of, 485t
- for myotonia, 3642
- for neuropathy, 3604t
- for pain, 3588
- for paramyotonia congenita, 3647

Meyer's loop, 222

MFAP5 gene mutations, 3332t

mGluR1 antibodies, 744t

mGluR5 antibodies, 743t, 748

MGMT, 685, 719

MGUS. *See* Monoclonal gammopathy of undetermined significance (MGUS)

MHC (major histocompatibility complex). *See* Human leukocyte antigen (HLA) complex

MHC class-restricted T cells, 2774

M hemoglobins, 779t, 780

MI. *See* Myocardial infarction (MI)

Miasma, 957

Mibefradil, 3159

MIC (minimal inhibitory concentration), 1026, 1168, 1181, S11
mic-2 gene, 730
MIC-A5.1 allele, 3091
Micafungin, 1692, 1710t, 1711t, 1712t
MICA gene, 1581t, 1583f, 1585
Mice. *See* Rodent(s); Rodent-borne viruses
Mice bite-wounds, 1143, 1145t
Micelles, 2540, 2727
Michet's criteria, relapsing polychondritis, 2920
Miconazole, 390, 1101t, 1103, 1692
Microaerophilic organisms, 1371
Microalbuminuria. *See* Albumin, urinary
Microangiopathic hemolytic anemia. *See also* Thrombotic microangiopathy
  jaundice in, 323
  pathogenesis of, 800, 801t
  peripheral blood smear, 922f
  peripheral blood smear in, 440f
  renal vascular injury and, 2444
  in TTP, 925
Microarray technology, 733, 976t, 977t
Microbial bioterrorism, 1149, 1341, 1342, 1459, S4, S7
Microbial pathogenesis. *See* Bacterial infections, pathogenesis of
Microbiome, 3777, 3830. *See also* Human microbiome
Microbiomics, 6
Microbiota, 964, 3830. *See also* Gut microbiota
Microcephaly, 1668, 3402–3403, 3904
Microcytosis, 445, 448
Microfilaria, S12
Microglia, 3394–3396, 3395f, 3582
β_2-Microglobulin
  in amyloidosis, 895, 896t, 900
  laboratory evaluation of, 475
  in multiple myeloma, 889, 889t
Microimmunofluorescence test, 1473
Microinfarcts, 3490, 3490f
Micronucleus assay, S6
Micronutrient malnutrition, 2620. *See also specific micronutrients*
MicroRNAs (miRNAs)
  in carcinoma of unknown primary, 733
  gene expression silencing by, 539
  in rheumatoid arthritis, 2843
  in testicular cancer, 705
Microsatellite instability, 520f, 523, 653, 714
Microscopic polyangiitis, 2415t, 2420, **2899,** A14
Microsomal ethanol oxidizing system (MEOS), 2709
Microsporida/microsporidiosis, 1600, 1739t, 1805t, **1807,** 1807f, S12
Microtubules, 561, 565
Microvascular angina, 2091
MIDAS (Migraine Disability Assessment Score), 3464, 3466f
Midazolam
  for conscious sedation, 485
  for delirium, 86, 86t
  dosage in liver disease, 488
  drug interactions of, 1742t
  for dyspnea, 83t
  ICU delirium and, 2298
  for nerve agent exposure, S5
  overdosage/poisoning with, 3713t
  for palliative sedation, 84
  for scorpion stings, 3735
  for spasms in tetanus, 1231
  for status epilepticus, 3426f
  during terminal weaning, 89
Midbrain, 3436f
Midbrain stroke syndromes, 3436f
Middle cerebral artery
  anatomy of, 3430f
  lesion in territory of, 202
  occlusion of, 3385f, 3429–3430, 3430f, 3437f, A16
Middle East respiratory syndrome (MERS), 995, 1535, 1539
Middleton maneuver, 479
Midline destructive disease, 2898
Midline granuloma, 261
Midodrine
  adverse effects of, 3547
  for cirrhotic ascites, 329
  for dialysis-related hypotension, 2400
  for hepatorenal syndrome, 2383
  for orthostatic hypotension, 160, 3227, 3547
  for postural orthostatic tachycardia syndrome, 1936
Midostaurin, 526f, 569t, 832, 833t, 881, 2821
Midpontine stroke syndrome, 3435f
Mifepristone, 739, 3012, 3059
MIF gene mutations, 2823t
MIG (monokine induced by IFN-γ), 2768t
Migalastat, 2010, 2425, 3355, 3602
Miglustat, 3356
Migraine, **3462**
  acephalgic, 3428, 3464
  after radiation therapy, A16
  chronic, 3462
  classification of, 3463t
  clinical features of, 3462, 3464, 3464t
    autonomic symptoms, 114
    cyclic vomiting, 298
    fatigue, 166
    ocular, 230
    Raynaud's phenomenon, 2177
    taste disorders, 242
    vertigo, 162
  diagnosis of, 3462, 3466t
  familial hemiplegic, 3462
  genetic considerations in, 3462
  hemiplegic, A16
  high-altitude travel and, 3759
  imaging in, 3462, 3465f
  pathogenesis of, 96, 3462, 3464f
  prevalence of, 113t, 3462
  prevention of, 3468t, 3469, 3470t
  retinal, 225
  vs. stroke, 3428
  treatment of
    5-HT_1 agonists, 3467t, 3468
    for acute attack, 3466, 3467t, 3468t
    combination therapy, 3466
    disability assessment in, 3464, 3466f
    dopamine antagonists, 3467t, 3469
    neuromodulation, 3469
    nonpharmacologic, 3466
    NSAIDs, 3466, 3467t
    opioids, 3467t, 3469
  triggers, 3462
  vestibular, 164
Migraine aura, 3462
Migraine Disability Assessment Score (MIDAS), 3464, 3466f
Migrainous neuralgia, 265
Migrainous stroke, V3
Migrating motor complex, 303
Mikulicz's disease, 2930. *See also* IgG4-related disease (IgG4-RD)
Milan criteria, for liver transplantation, 664
Mild behavioral impairment (MBI), 194
Mild cognitive impairment (MCI), 193, 3475, 3492, 3492t, 3506, 3886, 3888f
Miliaria rubra, 3774
Miliary tuberculosis, 153, 1386f, 1389–1390, 1400
Military culture, S8
Military veterans
  barriers to health care for, S8
  demographics of, S8
  health care needs of, S8
  health risks in, S8
  infections in, 1295, **S7**
  medical evaluation of, S8
  military and deployment history, S8
  neuropsychiatric illnesses in, S8
  postcombat care, S8
  postwar syndromes in, S8
  toxic environmental exposures in, 2240, S8
  whole health approach to, S8
  women, S8
Milk-alkali syndrome, 363, 364, 3262, 3282–3283
Milk of calcium bile, 2733
Milk of magnesia, 82t, 314
Millard-Gubler syndrome, 235
Miller-Dieker syndrome, 3403
Miller Fisher syndrome. *See also* Guillain-Barré syndrome (GBS)
  clinical features of, 1236, 3559, 3617, 3617t
  diagnosis of, 3617t, 3621t
  pathogenesis of, 3617, 3617t, 3618f, 3619t
Miller's lung, 2228t
Millipede dermatitis, 3737
Milnacipran, 2965, 2966t
Milrinone, 1992, 1993t, 2328, 2331
Milroy's disease (congenital lymphedema), 2183, 2548
Miltefosine
  actions of, 1750
  adverse effects of, 1744t, 1785
  for cutaneous leishmaniasis, 1787
  for mucosal leishmaniasis, 1787
  for *Naegleria* infection, 1737, 1758
  pharmacology of, 1750
  for visceral leishmaniasis, 1784, 1785
Milwaukee shoulder, 2961
Mind and body practices, 3925t, 3927
Mindfulness-based stress reduction, 3927
Mindsets, 3930t, 3931
Mineral(s), **2613**. *See also specific minerals*
  body stores of, 2605
  deficiencies of, 2614t
  in parenteral nutrition, 2623
  recommended intake of, 2601t
  tolerable upper intake levels of, 2614t
  toxicities of, 2614t
Mineralocorticoid(s)
  actions of, 2986
  for adrenal insufficiency, 3070
  resistance to, 346
  synthesis of, 3051f, 3052, 3053f
Mineralocorticoid excess, **3060**. *See also* Syndrome of apparent mineralocorticoid excess (SAME)
  clinical features of, 3061
  diagnosis of, 3061, 3062f
  differential diagnosis of, 3061, 3062f, 3063
  epidemiology of, 3060
  etiology of, 3060–3061, 3060t
  pathophysiology of, S1
  treatment of, 3062f, 3063
Mineralocorticoid receptors, 346
Mineral oil, vitamin A absorption and, 2611
Minimal bactericidal concentration (MBC), S11
Minimal change disease, **2421**
  clinical features of, 2415t, 2421
  etiology of, 2357
  pathogenesis of, 2421, 2421f
  proteinuria in, 342, 2357, 2380, 2414, 2421
  renal biopsy in, 2421, A4
  treatment of, 2421
Minimal effective concentration (MEC), S11
Minimal erythema dose, 434

- Minimal inhibitory concentration (MIC), 1026, 1168, 1181, S11
- Minimally conscious state, 187, 193
- Minimally invasive adenocarcinoma, 616, 622
- Minimal residual disease
  - in ALL, 846, 846t
  - in CLL, 856, 856t
  - ctDNA for detection of, 3983
  - in multiple myeloma, 892
- Mini-Mental Status Examination (MMSE), 85, 183, 3379
- Mini Nutritional Assessment, 316
- Minitransplant, 561
- Minocycline
  - for *A. baumannii* infections, 1296–1297, 1297t
  - for acne vulgaris, 391
  - for actinomycosis, 1366t
  - actions of, 1166, 1177, 1182t
  - adverse effects of, 1177
    - cutaneous, 420
    - DIHS, 424. *See also* Drug-induced hypersensitivity syndrome (DIHS)
    - fever, 150
    - hepatotoxicity, 2667
    - hyperpigmentation, 399, 420
    - lupus syndrome, 2835
    - oral lesions, 264t
    - steatosis, 2703t
    - vasculitis, 403, 426, 2940t
  - indications for, 1173t, 1177
  - for leprosy, 1411
  - for MRSA infections, 383, 1203t, 1204
  - for *Nocardia* infections, 1362
  - for osteomyelitis, 1065t
  - resistance to, 1173t, 1182t, 1184, 1359f
  - for Whipple's disease, 1370
- Minoxidil
  - adverse effects of, 282t, 421, 1920t, 3138
  - for androgenetic alopecia, 395t
  - development of, 488
  - for hypertension, 2145t
- Miosis, 190, 221, 3705
- MIP-1α (macrophage inflammatory protein-1α), 1585, 2767t
- MIP-1β (macrophage inflammatory protein-1β), 2767t
- Mipomersen, 3244t, 3252
- *miR-155* overexpression, 851
- Mirabegron, 3549, 3588
- Mirikizumab, 2793t
- Mirizzi's syndrome, 2732
- miRNAs. *See* MicroRNAs (miRNAs)
- Mirtazapine
  - adverse effects of, 3188, 3659t, 3667
  - for anorexia, 509
  - for depression, 85, 3659t, 3667
  - for functional dyspepsia, 302
  - for insomnia, 86
  - for involuntary weight loss, 316
  - for nausea and vomiting, 299, 299t
  - for PML, 1117
- Mirvetuximab soravtansine-gynx, 535t, 566, 566t
- MIS-C (multisystem inflammatory syndrome in children), 141t, 402, 1132t, 1540, 2907
- Mismatch repair (MMR) pathway, 653
- Misonidazole, 3609t
- Misoprostol, 2524t, 2526, 2530t
- Missense mutation, 3784, 3786f
- Mistake (error), 42
- Mitapivat, 797
- Mites, biting, 3730, 3731f
- *MITF*, 397, 430, 599, 687t, 3783t
- Mitiglinide, 3211t
- Mitochondria
  - dysfunction, 3400, 3871
  - in energy production, 3645
  - functions of, 3805–3807, 3806t
  - genetic defects of, 2004t
- Mitochondrial ataxias, 3537
- Mitochondrial cytochrome c, 2781
- Mitochondrial DNA (mtDNA), **3807**. *See also* Mitochondrial DNA diseases
  - aging and, 3815, 3815f, 3870
  - functions of, 3791, 3806t, 3807–3809
  - heteroplasmy in, 3807–3808, 3808f
  - homoplasmic variants of, 3808–3809
  - inheritance of, 3645, 3807, 3807f
  - mitotic segregation of, 3807–3808
  - nuclear cross-talk perturbations with, 3814, 3814f
  - phylogeny of, 3809
  - polyploidy in, 3807–3808
  - replication of, 3807
  - structure of, 3791, 3807–3809
- Mitochondrial DNA diseases, **3805**
  - age-associated, 3815, 3815f
  - cardiomyopathy in, 2003, 2004t
  - clinical features of, 248, 3806f, 3810–3813, 3811t
  - depletion syndromes, 3646
  - diagnosis of, 3813–3814, 3813f
  - examples of, 3809f, 3810–3813, 3811t
  - genetic control of, 3806f
  - genetic counseling in, 3816
  - genetic mutations causing, 3806, 3809f, 3810, 3813–3814
  - heteroplasmy in, 3807–3808, 3808f
  - homoplasmic sequence variation in, 3808–3809, 3814
  - inheritance of, 3790f, 3791, 3806, 3807f
  - with large-scale DNA rearrangements, 3811t
  - organ manifestations of, 3806f
  - pathologic features of, 3809–3810, 3810f
  - phenotypic heterogeneity of, 3808f
  - with point mutations, 3811t
  - prevention of, 3816–3817, 3817f
  - renal, 2436
  - sporadic, 3816
  - treatment of, 3815–3816
- Mitochondrial myopathy(ies), **3645**
  - Kearns-Sayre syndrome. *See* Kearns-Sayre syndrome
  - MELAS syndrome, 3287, 3646, 3811, 3811t, 3812f
  - MERRF syndrome, 3645–3646, 3811, 3811t, 3814
  - pathogenesis of, 3645
  - progressive external ophthalmoplegia. *See* Progressive external ophthalmoplegia
- Mitochondrial myopathy, encephalopathy, lactic acidosis, and strokelike episodes (MELAS) syndrome, 3287, 3646, 3811, 3811t, 3812f, A16
- Mitomycin C
  - adverse effects of, 595, 925, 2377, 2445
  - for anal cancer, 659
  - for bladder cancer, 692
- Mitophagy, 3871
- Mitosis, 3779
- Mitotane, 739, 3012, 3059, 3066
- Mitotic segregation, 3808
- Mitotic spindle inhibitors, 563t, 565
- Mitoxantrone, 562t, 564, 703
- MitraClip, 2130f, 2131, A11
- Mitral regurgitation, **2052**
  - with aortic stenosis, 2065
  - cardiogenic shock in, 2328f, 2330
  - clinical features of, 2053
  - diagnosis of
    - chest radiograph, 2053
    - CMR, 1893
    - ECG, 2053, A7
    - echocardiography, 1892–1893, 1892f, 2053, 2054f, 2859v
    - hemodynamic patterns, 2329t
    - physical examination, 2053–2054
    - POCUS, 3914v
  - in dilated cardiomyopathy, 2015
  - etiology of, 2052, 2052t
  - heart failure and, 1981, 1999
  - heart sounds/murmur in, 283f, 284f, 287f, 289f, 1865, 1865f, 2053
  - in hypertrophic cardiomyopathy, 2006
  - left atrial compliance in, 2053
  - with mitral stenosis, 2067
  - pathophysiology of, 2052–2053
  - in pregnancy, 3900
  - radiation-associated, 753
  - in rheumatic heart disease, 2855, 2859v
  - sudden cardiac death in, 2335t
  - treatment of
    - algorithms for, 2054–2055f
    - medical, 2054–2055
    - surgical, 2055–2056, 2055t
    - transcatheter repair and replacement, 2056–2057, 2056–2057f, 2130f, 2131
- Mitral stenosis, **2048**
  - with aortic regurgitation, 2065
  - associated lesions, 2049, 2065
  - clinical features of, 2048–2049
  - diagnosis of
    - cardiac catheterization, 2050
    - chest radiograph, 2049
    - ECG, 2049, A7
    - echocardiography, 1892–1893, 1893f, 2049, 2859v
    - hemodynamic measurements, 1905, 1906f, 2048
    - physical findings, 2049
    - valve area, 1907
  - differential diagnosis of, 2049–2050
  - etiology of, 2048, 2048t
  - heart sounds/murmur in, 283, 283f, 287f, 288, 1865f, 1866, 2049
  - with mitral regurgitation, 2067
  - pathology of, 2048
  - pathophysiology of, 2048
  - in pregnancy, 3900
  - in rheumatic heart disease, 288, 2048, 2855, 2859v
  - skin manifestations of, 1859
  - treatment of
    - commissurotomy, 2051–2052, 2051f
    - medical therapy, 2050–2051
    - strategy for, 2050f
    - valve replacement, 2051–2052, 2055t
    - valvuloplasty, 2130–2131
  - with tricuspid stenosis, 2059
- Mitral valve, 1892, 1892f
- Mitral valve prolapse, **2057**
  - clinical features of, 291, 1860, 2058
  - congenital or developmental, 2058, 2058f
  - echocardiography in, 1892, 1893f, 2057f, 2058, 2058f
  - etiology of, 2057–2058, 2057f
  - heart sounds/murmur in, 285, 285f, 289, 1866, 1866f, 2058
  - in polycystic kidney disease, 2432
  - stroke prevention in, 3445t
  - sudden cardiac death in, 2335t
  - treatment of, 2059
- Mitral valve repair/replacement
  - with aortic valve replacement, 2067
  - commissurotomy, 2051, 2051f
  - indications for, 2051–2052, 2055–2056, 2059
  - mortality after, 2055t
  - transcatheter, 2056–2057, 2056–2057f, 2059, 2067, A11

Mittelschmerz, 3130, 3137
Mixed connective tissue disease (MCTD)
  AV conduction block in, 1927
  clinical features of, 2875–2876
  diagnosis of, 2875–2876
  lymphadenopathy in, 476
  neuropathy in, 3605
Mixed gonadal dysgenesis (45,X/46,XY mosaicism), 3097t, 3098
Mixed hyperlipidemia (elevated TG and LDL-C)
  approach to the patient, 3249
  definition of, 3245
  primary (genetic) causes, 3241t
    familial combined hyperlipidemia, 3245–3246
    familial dysbetalipoproteinemia, 3245
    hepatic lipase deficiency, 3245
  secondary factors in, 3246–3247
  treatment of. *See* Hypercholesterolemia (elevated LDL-C), treatment of
Mixed shock, 2311
Mixing studies, in coagulation evaluation, 475
Miyoshi distal myopathy, 3637, 3641, 3643–3644
MKRN3 gene mutations, 3109, 3131
MLD (metachromatic leukodystrophy), 3353t, 3485, 3824t, 3827, A16
MLH1 gene mutations
  in bladder cancer, 691
  in endometrial cancer, 714
  in Lynch syndrome, 523, 524t, 654, 671t, 3785–3786
  in ovarian cancer, 710
  in pancreatic cancer, 671t
MLH6 gene mutations, 710, 3786
MLL, 3940, 3941
MLL gene mutations, 660, 661t, 784, 3941
M-mode echocardiography, 1876, 1877f
MMPs. *See* Matrix metalloproteinases (MMPs)
MMR. *See* Measles-mumps-rubella vaccine (MMR)
MMSE. *See* Mini-Mental Status Examination (MMSE)
MNase, 3937
MNS RBC group system, 904, 904t
MoAB114, S4
Mobile genetic element, 976t
Modafinil
  for depression, 85
  for fatigue, 84, 167, 3588
  for narcolepsy, 213t, 215
  for opioid-induced drowsiness, 81
  for shift-work disorder, 219
Model for End-Stage Liver Disease (MELD) score, 2633, 2720, 2720t
Modification of Diet in Renal Disease (MDRD) equation, 337, 2388t, S10
Modified Medical Research Council Dyspnea Scale, 268, 268t
Modoc virus, 1660t
Modular scale-free topology, 3956
Modulins, 974
MODY (maturity-onset diabetes of the young), 3195, 3202–3203, 3783t, 3794t
Moffett's position, 243
MOG (myelin oligodendrocyte glycoprotein), 35, 227, 2790t
Mogamulizumab, 559f, 866, 867, 1555f, 1556
MOG (myelin oligodendrocyte glycoprotein) antibodies, 744t, 746, 746f, 749
MOG (myelin oligodendrocyte glycoprotein) antibody-associated disease, 3593, 3593t
Mohs micrographic surgery, 603, 608, 610
Mokola virus, 1655
Molar pregnancy. *See* Gestational trophoblastic disease
Mole, atypical, 600
Molecular mimicry, 2826, 3398
Molindone, 3673t
Mollaret's meningitis, 1131, 1132t, 1500
Molluscum contagiosum
  clinical features of, 1519, 1521f
  diagnosis of, 1419
  epidemiology of, 1518, 1518t
  in HIV infection, 1606
  skin manifestations of, 1049, A5
  treatment of, 1521
Molly (MDMA), 349, 3693, 3696. *See also* Psychostimulants
Molnupiravir, 1156, 1488t, 1542
Molybdenum, 2601t, 2614t, 2615, S6
Momelotinib, 821
"Monday chest tightness," 2238–2239
Monge's disease (chronic mountain sickness), 278, 3760
Monkey, bite-wounds from, 1143, 1145t
Monkeypox. *See* Mpox
Monkeypox viruses, 1517. *See also* Mpox
Monoamine oxidase inhibitors (MAOIs)
  adverse effects of, 3188, 3658, 3659t, 3667
  for depression, 3659t, 3667
  drug interactions of, 1743t, 1748, 3660t, 3667
  overdosage/poisoning with, 3710t, 3715t
  for panic disorder, 3658
Monoamine oxidase type B (MAO-B) inhibitors, 3503, 3504t
Monobactams. *See* Aztreonam
Monoballism, 3515
Monoblasts, 461
Monoclonal B cell lymphocytosis, 849, 852
Monoclonal cryoglobulinemia, 407
Monoclonal gammopathy, 744, 2424
Monoclonal gammopathy, malignant. *See* Chronic lymphocytic leukemia (CLL); Multiple myeloma; Waldenström's macroglobulinemia
Monoclonal gammopathy of undetermined significance (MGUS)
  amyloidosis in, 897
  diagnosis of, 887–888, 888t, 896t
  in HIV infection, 1604
  polyneuropathy in, 3619t, 3623
  treatment of, 889–890
  tumor markers in, 504t
  VWD in, 927
Monoclonal immunoglobulin, 504t
Monoclonal mast cell activation syndrome, 2821
Monocyte(s), **461**
  in autoimmune disease, 2824
  development of, 2764f
  disorders of, 458t, 461–462
  hematopoietic differentiation of, 760f
  in HIV infection, 1580
  in innate immunity, 2760
  in peripheral blood smear, 442f, 443
  in sepsis/septic shock, 2317, 2318f
  in TB, 1385
Monocyte chemotactic protein (MCP), 1585, 2767t, 3656
Monocytopenia, 462
Monocytosis, 461
Monogenic disorders. *See* Mendelian disorders
Monokine induced by IFN-γ (MIG), 2768t
Mononeuropathies, diabetic, 3604. *See also* Diabetic neuropathy
Monongahela virus, 1661t, 1674
Monoparesis, 171, 171f
Monophosphoryl lipid A (MPL), 997t, 998
Monosodium glutamate, 243
Monosodium urate crystals, 2956–2957, 2959f
Mono-ubiquination, histone, 3936–3937
Montelukast, 257, 2225, 2808, 2810, 2940t
Mood disorders, **3664**
  alcohol-induced, 3676
  bipolar disorder, 3652, **3667,** 3669t, 3952
  depression. *See* Depression/depressive disorders
  in fibromyalgia, 2963
  neuropathology of, 3654, 3655f
Mood stabilizers, 3670t. *See also specific drugs*
MOPP regimens, 871
Moral distress, 73
Moraxella catarrhalis, 1264
Moraxella catarrhalis infections
  clinical features of, 1265
    hemoptysis, 274
    laryngitis, 259
    otitis media, 254, 1265
    pneumonia, 1262f, 1265
    sinusitis, 1265
  in COPD, 1262f, 1265, 2259
  diagnosis of, 1265, S11
  epidemiology of, 1264
  pathogenesis of, 1264–1265
  treatment of, 1185, 1265
Moray eel bite-wound, 1145t
Morbilliform eruptions. *See also* Rash
  definition of, 378t
  drug-induced, 417, 417t, 422, 422f
  etiology of, 401–402, 401t
Morganella infections, 1093, 1293
Morgellons syndrome, 3738
Morningness-eveningness types, 3946
Morphea, 394, **416,** 2790t
Morphine
  adverse effects of, 2118
  contraindications to, 2109t
  for cough, 274
  dosage in liver disease, 488
  for dyspnea, 83t
  for NSTE-ACS, 2109t, 2110
  for pain, 80f, 81, 97t, 98
  for pulmonary edema, 2331
  routes of administration of, 98
  for STEMI, 2118
  sustained-release, 97t, 100
  during terminal weaning, 88
Morquio syndrome, 3352t, 3357
Mortality. *See* Death
Morumbi virus, 1662t
Morvan's syndrome, 743t, 748, 751
Mosaicism, 3783, 3791
Mosapride, 2593
Mosquitoes
  Aedes. See Aedes mosquitoes
  avoidance of, 1017–1018
  bite, 3736
  Culex. See Culex mosquitoes
  in malaria transmission, 1760–1762, 1761f
Mosunetuzumab-axgb, 534t, 559f, 753t
Motilin agonists, 299t
Motion sickness, 83, 297, 298, 299t, 1020
Motor cortex, 200f
Motor examination, 3379–3380
Motor neuron disorders. *See also* Amyotrophic lateral sclerosis (ALS)
  differential diagnosis of, 3521, 3523
  etiology of, 3520t
  genetic, 3522–3523t
  lower, 3520t, 3522t, 3524
  sporadic, 3520t
  upper, 3520t, 3522–3523t, 3525
Motor neuropathy, 3623
Motor tic, 3515
Mounier-Kuhn syndrome, 2242
Mountain sickness, chronic, 278, 3760
Mouth. *See* Oral cavity

Movement disorders
  choreas. *See* Chorea(s)
  drug-induced, **3516**
  dystonia. *See* Dystonia
  functional (psychogenic), 3519
  Huntington's disease. *See* Huntington's disease
  hyperkinetic. *See* Hyperkinetic movement disorders
  myoclonus. *See* Myoclonus
  in neurodegeneration with brain iron accumulation, 3518–3519
  in Parkinson's disease. *See* Parkinson's disease
  restless legs syndrome. *See* Restless legs syndrome
  Tourette's syndrome, **3515**
  in Wilson's disease. *See* Wilson's disease
Moxetumomab, 848t
Moxidectin, 1743t, 1750, 1815, 1815t
Moxifloxacin
  actions of, 1166, 1182t
  adverse effects of, 1178
  for anaerobic bacterial infections, 1379
  for bite-wound infections, 1144
  for community-acquired pneumonia, 963t, 1027t, 1028t
  indications for, 1174t, 1178
  for *Legionella* spp. infections, 1275t
  for leprosy, 1411l
  for lung abscess, 1036
  for MDR-TB, 1394t, 1395t, 1399, 1421, 1421t, 1424, 1425
  for mucopurulent cervicitis, 1104
  for *Mycoplasma* infections, 1466t, 1467
  for *Nocardia* infections, 1362
  for pelvic inflammatory disease, 1106
  for plague, 1346, 1346t
  for pneumococcal infections, 1194
  resistance to, 1026, 1174t, 1182t, 1359f, 1379, 1467
Moxonidine, 1993f, 1994
Moyamoya disease, 3447, 3455, A16
6-MP. *See* 6-Mercaptopurine (6-MP)
MPL gene mutations, 821, 822, 823, 824
MPN-U (myeloproliferative neoplasm, unclassifiable), 876
Mpox
  clinical features of, 1519, 1520
    perianal ulcers, A1
    pustular lesion, A1
    rash, 142t, 147, 1051
    vesicles, A5
  diagnosis of, 1499, 1519
  differential diagnosis of, 1499
  epidemiology of, 1051, 1149, 1518–1519, 1518t, 3856
  health care facility preparedness for, 1149
  in HIV infection, 1521
  in travelers, 1506
  treatment of, 1489, 1499, 1521
  vaccine, 1019
  zoonotic characteristics of, 1518t
Mpox vaccine, 1522, 1594t
M protein, 1207
MPTP, 3497
MPZ (myelin protein zero), 3598
MRA. *See* Magnetic resonance angiography (MRA)
MRCP. *See* Magnetic resonance cholangiopancreatography (MRCP)
MRE11 gene mutations, 2800, S9
MRGPRX2, 422
MRI. *See* Magnetic resonance imaging (MRI)
mRNAs (messenger RNAs), 1484, 1543, 2788, 3777
MRP (multidrug resistance–associated protein), 2637f, 2638
MRSA infections. *See* Methicillin-resistant *Staphylococcus aureus* (MRSA) infections
MS. *See* Multiple sclerosis (MS)
MSA. *See* Multiple system atrophy (MSA)
MSCRAMMs, 966, 966t, 1038, 1197, 1216
MSH. *See* A-Melanocyte-stimulating hormone (MSH)
MSH2 gene mutations
  in endometrial cancer, 714
  in Lynch syndrome, 523, 524t, 654, 671t, 3786
  in ovarian cancer, 710
  in upper urinary tract cancer, 691
MSH3 gene mutations, 3513
MSH6 gene mutations, 523, 524t, 671t, 691, 3786
MSM. *See* Men who have sex with men (MSM)
MSMD (Mendelian susceptibility to mycobacterial disease), 2796
MST1 gene mutations, 2554t
MSY region, 3107, 3107f
MTAP gene mutations, 3939
MTC. *See* Medullary thyroid carcinoma (MTC)
MTD (maximum-tolerated dose), 555
mtDNA. *See* Mitochondrial DNA (mtDNA)
5-MTHF (5,10-methylenetetrahydrofolate), 782, 782t
MTHFR (methylenetetrahydrofolate reductase), 783f, 784
MTHFR gene mutations, 784
MTNR1B gene mutations, 3951
mTOR
  in COPD, 2251
  in HCC, 660
  inactivation of, 973
  pathway, in aging, 3871
mTOR inhibitors, 533t, 535, 571t, 573–574, 686. *See also* Everolimus; Temsirolimus
mtrR gene, 1255
MTTP gene mutations, 3247
MUC1 gene mutations, 2435
MUC5B gene mutations, 2261, 2266
Mucambo virus, 1657t, 1673
Mucin, 2729
Mucinous cystic neoplasm, 672
Mucinous cyst, pancreatic, 2738, 2738f
Muckle-Wells syndrome (MWS), 462, 2935t, 2936
Mucociliary escalator, 1191
Mucolipidoses, 3353t
Mucopolysaccharidoses, 3352t, 3356–3357, A15
Mucorales, 1718, 1718t
Mucormycosis, **1717**
  in cancer patient, 582
  clinical features of, 148, 235, 256, 1688t, 1718–1719
  diagnosis of, 1688t, 1719–1720, 1719f
  differential diagnosis of, 1720
  epidemiology of, 1718
  etiology of, 1718, 1718t
  follow-up radiology in, 1722
  in HIV infection, 1595
  in immunocompromised patient, A5
  invasive, A16
  pathogenesis of, 1718
  prognosis of, 1722
  rhinocerebral, 995
  risk factors for, 1688t
  treatment of, 1720–1722, 1721t
Mucosa-associated lymphoid tissue (MALT), 2780
Mucosa-associated lymphoid tissue (MALT) lymphoma
  clinical features of, 649–650
  etiology of, 857, 865
  genetic considerations in, 860, 860t, 865
  H. pylori and, 649, 857, 1299–1300
  pathology of, 865
  treatment of, 650, 865
Mucosal bleeding, 473
Mucosal leishmaniasis, 1782t, 1787, 1788f
Mucosal malabsorption, 309
Mucositis
  acute necrotizing ulcerative (noma), 261
  cancer treatment-related, 507t, 508, 576, 578t
  in HCT recipient, 916
  radiation therapy-related, 614
Mucous membrane pemphigoid, 263, 263t, 410t, **413**
Mucous retention cyst, oral, 264t
Mucus hypersecretion, 2218
Mucus, intestinal, 1077
Mudri virus, 1662t
Muehrcke's lines, 2357
Mu heavy chain disease, 894–895
Muir-Torre syndrome, 404, 606
Muju virus, 1661t
Mulberry molars, 261, 1433
Müllerian agenesis, 3101, 3133
Müllerian inhibiting substance (antimüllerian hormone), 3094, 3126, 3128
Müllerian tumors, 715
Multicentric reticulohistiocytosis, 404
Multidetector computed tomography (MDCT), 3384. *See also* Computed tomography (CT)
Multidrug resistance–associated protein (MRP), 2637f, 2638
Multidrug resistance (MDR) phenotype, 541, 574
Multifactorial disorders, 3777, 3793
Multifactorial inheritance, 3800
Multifocal atrial tachycardia, 1932t, 1946, 1946f, A8
Multifocal motor neuropathy, 3619t, 3623
Multifocal motor neuropathy with conduction block, 3521, 3524
Multikinase inhibitors, 557f, 569–570t, 572–573
Multilocus sequence typing, 976t, 1188
Multimorbidity, 3873, 3876, 3877f
Multinodular goiter, 3041–3042. *See also* Thyroid nodular disease
Multiomics data, 3960, 3962f
Multiorgan system failure, 774t, 2296
Multiple carboxylase/biotinidase deficiency, 3370t
Multiple endocrine and other organ neoplasias (MEONs), 3079–3080, 3080t
  Carney syndrome. *See* Carney syndrome/complex
  Cowden's syndrome. *See* Cowden's syndrome
  McCune-Albright syndrome. *See* McCune-Albright syndrome
  neurofibromatosis type 1. *See* Neurofibromatosis type 1 (NF1)
  neurofibromatosis type 2. *See* Neurofibromatosis type 2 (NF2)
  von Hippel–Lindau disease. *See* Von Hippel–Lindau disease/syndrome
Multiple endocrine neoplasia type 1 (MEN 1, Wermer's syndrome), **3080**
  clinical features of, 3080–3081, 3080t
    adrenocortical tumors, 3083–3084
    angiofibromas, 404, 3084
    carcinoid tumors, 3083
    collagenomas, 3084
    gastrinoma, 2535, 3081–3082
    glucagonoma, 3082
    hyperparathyroidism, 3275–3276
    insulinoma, 3082
    lipomas, 3084
    meningiomas, 3084
    NETs, 676, 679, 3082–3083. *See also* Neuroendocrine tumors (NETs)
    parathyroid tumors, 3081, 3275
    pituitary tumors, 3004, 3004t, 3083
    pNETs, 3081, 3082–3083, 3082f
    thyroid tumors, 3084
    VIPoma, 3082
    ZES, 3081

- Multiple endocrine neoplasia type 1 (MEN 1, Wermer's syndrome) (*Cont.*):
  - epidemiology of, 3079–3080
  - genetic considerations in, 3084
    - chromosomes involved, 524t, 3084, 3276, 3276f
    - *MEN1/Menin* gene mutations, 717t, 3004, 3004t
    - variable expressivity, 3788
  - genetic testing in, 3084
  - pathogenesis of, 2977
  - screening tests in, 3081t, 3084
- Multiple endocrine neoplasia type 2 (MEN 2)
  - clinical features of, 3084–3085
  - early interventions for, 3804t
  - genetic considerations in, 523, 3077, 3085t, 3792
  - genetic testing for, 3803, 3804t
  - medullary thyroid carcinoma in, 3048, 3077, 3080t, 3085, 3086f. *See also* Medullary thyroid carcinoma (MTC)
  - medullary thyroid carcinoma only, 3080t, 3085
  - parathyroid tumors in, 3086
  - pathogenesis of, 2977
  - pheochromocytoma in, 3077, 3086
  - screening tests in, 3085t
  - type 2A
    - clinical features of, 404, 3080t, 3085, 3086
    - genetic considerations in, 524t, 3080t
    - hyperparathyroidism, 3275
  - type 2B (MEN 3), 3080t, 3085–3086, 3275, A15
- Multiple endocrine neoplasia type 4 (MEN 4), 3004t, 3080t, 3086
- Multiple endocrine neoplasia type 5 (MEN 5), 3086–3087
- Multiple exostoses, 3318
- Multiple hamartoma syndrome. *See* Cowden's syndrome
- Multiple/mixed valvular heart disease, **2065**
- Multiple myeloma, **883**
  - AL amyloidosis and, 2022
  - clinical features of, 885–887
    - amyloidosis, 897, A9
    - anemia, 450, 887
    - bone lesions/pain, 885–886, 887f, 892–893, 3299, A16
    - cryoglobulinemia, A14
    - hypercalcemia, 364, 737, 738, 887
    - hypogammaglobulinemia, 886
    - infections, 577, 886, A5
    - neuropathy, 751, 887, 3607–3608, 3623
    - vs. osteoporosis, 3299
    - proteinuria, 342, 887
    - renal, 2359
      - contrast nephropathy, 2377, 2379t
      - light chain cast nephropathy, 2380, 2424, 2440, 2440f, A4
      - renal biopsy in, A4
      - tubular damage, 886–887
    - vascular purpura, 927
  - definition of, 883
  - diagnosis of, 885f, 887–889, 888t, A6
  - epidemiology of, 885
  - epigenetic influences in, 538
  - etiology of, 883–884
  - genetic considerations in, 860t, 883–884, 889
  - global considerations in, 885
  - M component in, 887–888
  - nonsecretory, 888t
  - pathogenesis of, 541–542, 859f, 885, 886f
  - prognosis of, 889, 889t
  - smoldering, 888t, 890
  - staging of, 889, 889t
  - treatment of, **889,** 890t, 891f
    - CAR T-cells, 3824t
    - chemotherapy, 890–891
    - denosumab, 542
    - HCT, 891–892, 919t, 920
    - HDAC inhibitors, 539
    - lenalidomide, 891–892
    - proteasome inhibitors, 536, 573, 891
    - relapsed disease, 892
    - supportive care, 892–893
    - thalidomide, 891
  - tumor markers in, 504t
  - VWD in, 927
- Multiple sclerosis (MS), **3577**
  - active progressive, 3579
  - clinical features of, 3577–3578, 3577t, 3587–3588
    - autonomic dysfunction, 3549, 3588
    - constipation, 314, 3588
    - cranial nerve palsies, 3578
    - erectile dysfunction, 3157
    - fatigue, 166, 3588
    - frontal lobe syndrome, 208
    - gait disorders, 178
    - hearing loss, 247
    - myelitis, 3565
    - myelopathy, 3568
    - ocular, 225, 235, 236, 3578
    - olfactory dysfunction, 241
    - weakness, 170, 3578, 3588
  - clinical variants of, 3589, 3589f
  - course of, 3578–3580, 3579f
  - diagnosis of, 3582–3584, 3583t, 3584f, A16
  - differential diagnosis of, 3584, 3585t, 3590, 3594
  - epidemiology of, 3580–3581
  - genetic considerations in, 3581, 3581t
  - microbiome and, 3398, 3837–3838
  - opticospinal, 3590
  - pathogenesis of, **3581**
    - autoimmunity, 2790t, 2825
    - autoreactive T cells, 3582
    - demyelination, 3397, 3581
    - humoral autoimmunity, 3582
    - immunology, 3582
    - inflammasome mutations, 2762t
    - neurodegeneration, 3581–3582
    - physiology, 3582, 3582f
    - reactive astrocytes, 3397
  - pregnancy and, 3588, 3903
  - primary progressive, 3579
  - prognosis of, 3588
  - progression independent of relapsing activity, 3579–3580
  - relapse associated worsening, 3579
  - relapsing/remitting, 3578, 3579f
  - scoring systems for, 3578–3579, 3580t
  - secondary progressive, 3578–3579, 3579f
  - treatment of, **3584**
    - for acute attacks, 3585
    - dietary, 2612
    - disease-modifying therapies, 3585–3587, 3586–3587t
    - experimental therapies, 3587
    - glucocorticoids, 3585
    - for initial demyelinating episodes, 3585
    - off-label options in, 3587
    - symptomatic, 3587–3588
    - unproven methods for, 3587
- Multiple sleep latency test, 212, 213
- Multiple sulfatase deficiency, 3354t
- Multiple system atrophy (MSA)
  - α-synuclein in, 3400
  - clinical features of, 3496, 3497t, 3506, 3548
  - diagnosis of, 3496, 3497t
  - vs. DLB, 3492
  - epidemiology of, 3548
  - imaging in, 3548, 3549f, A16
  - ocular involvement in, 221
  - parkinsonism in, 178, 3496
  - prions in, 3401, 3532
  - sleep disorders in, 218
  - treatment of, 3549
- Multiplex panels, 3801
- Multisystem inflammatory syndrome in children (MIS-C), 141t, 402, 1132t, 1540, 2907
- Multisystem proteinopathies, 3644, 3644t
- Multivariate pattern analysis, 3966
- Multivariate statistical models, 4
- Mumps, **1646**
  - clinical features of, 1059, 1647–1648, 1648f, 3114
  - complications of, 1126, 1131t, 1648
  - diagnosis of, 1639
  - differential diagnosis of, 1648
  - epidemiology of, 1646–1647, 1647f
  - etiology of, 1646
  - immunity to, 1647, 1649
  - incidence pre- and postvaccine, U.S., 1002t
  - pathogenesis of, 1647
  - prevention of, 1649. *See also* Measles-mumps-rubella vaccine (MMR)
  - transmission of, 1647
  - treatment of, 1649
- Munchausen's syndrome, 310, 3670
- Mupirocin
  - actions of, 1166, 1182t, 1183f
  - for impetigo, 1210
  - indications for, 1174t
  - nasal, in staphylococcal skin infections, 383
  - prophylactic uses of, 1180t
  - resistance to, 1174t, 1182t, 1185
- Mu receptors, 3689, 3689f, 3689t
- Murine leukemia virus, 1550t, 1552f
- Murine (endemic) typhus. *See* Endemic (murine) typhus
- Murmurs. *See* Heart murmurs
- Murphy's sign, 324, 2732
- Murray valley encephalitis virus, 1660t
- Murutucú virus, 1659t, 1667
- Muscarinic agonists, 3712t
- Muscle(s)
  - atrophy of, 2621, 3636–3637
  - electrophysiologic studies of. *See* Electromyography (EMG)
  - enlargement of, 3636–3637
  - fatigability of, 165, 168
- Muscle biopsy
  - in antisynthetase syndrome, 2913, 2914f
  - in dermatomyositis, 2911–2912, 2912f
  - in immune-mediated necrotizing myopathy, 2910, 2913f
  - in inclusion body myositis, 2914, 2915f
  - in inflammatory myopathies, 2910
  - in mtDNA diseases, 3810f, 3814
  - in muscular dystrophy, 3638
  - in myopathies, 3638
  - in polymyositis, 2912, 2913f
  - in trichinellosis, S12
- Muscle bulk, 168
- Muscle compartment syndrome, 3721
- Muscle contracture, 3634–3635
- Muscle cramps, 2400, 3634, 3774
- Muscle disorders
  - calcium channel. *See* Hypokalemic periodic paralysis (HypoKPP)
  - chloride channelopathies, 3647
  - energy metabolism, 3645. *See also* Glycogen storage diseases
  - muscular dystrophy. *See* Muscular dystrophy
  - myopathy. *See* Myopathy(ies)
  - potassium channelopathies, 3647
  - sodium channel, 359, 3635
  - sodium channelopathies, 3646–3647

Muscle dysphoria syndrome, 3172
Muscle mass, 2621
Muscle pain, 3636
Muscle palsy, V3
Muscle phosphorylase deficiency (McArdle disease, type V GSD), 3361t, **3364,** 3644
Muscle phosphorylase kinase deficiency (type IX GSD), 3361t, **3365**
Muscle relaxants
  adverse effects of, 1920t, 2352, 3412t
  for low back pain, 126
  in mechanical ventilation, 2295
  for neck pain, 132
  overdose/poisoning with, 3711t, 3713t
Muscle-specific tyrosine kinase (MuSK), 3624
Muscle spindle, 170f
Muscle stiffness, 3635–3636
Muscle strength, 2942, 3381
Muscle stretch (tendon) reflexes, 168, 168t, 170f, 3381
Muscle tone, 168, 168t, 3381
Muscle wasting, 1611
Muscle weakness. *See* Weakness
Muscular dystrophy, **3638**
  Becker's, 3638–3640
  diagnosis of, 3636f, 3637–3638
  differential diagnosis of, 2910
  distal, 3640t, 3643–3644
  Duchenne, 3638–3640, 3787–3788, 3804t, 3823
  Emery-Dreifuss, 3641–3642, 3787
  facioscapulohumeral, 3634, 3637f, 3642–3643
  genetic considerations in, 3785, 3787–3788, 3788t
  global considerations in, 3648
  limb-girdle. *See* Limb-girdle muscular dystrophies (LGMDs)
  with myasthenia, 3629t
  myotonic. *See* Myotonic dystrophy
  oculopharyngeal, 234, 3643
  proteins associated with, 3641f
Musculoskeletal disorders
  approach to the patient, **2938,** 2938t, 2939–2940f
    hand pain, 2942–2943, 2942f
    hip pain, 2944–2945, 2944f, 2953
    knee pain, 2944, 2953
    shoulder pain, 2943–2944, 2943f
    telehealth evaluation, 2945
  arthritis. *See* Arthritis/arthropathy
  articular vs. nonarticular, 2938
  chest discomfort/pain in, 103t, 104
  drug-induced, 2940t
  in hemochromatosis, 3336
  history in, 2940
  in hospitalized patient, 2941
  hyperlipidemia and, 2968
  imaging in, 2946–2947, 2947t, 2948f
  infections
    Actinomyces, 1365
    in joints. *See* Infectious (septic) arthritis
    parasitic, 1740t
    in prosthetic joints. *See* Prosthetic joint infections
  inflammatory vs. noninflammatory, 2938–2939
  laboratory studies in, 2945–2946, 2946t
  obesity and, 3187
  in older adults, 2941
  osteoarthritis. *See* Osteoarthritis
  physical examination in, 2941–2942
  prevalence of, 2938
  in rheumatic fever, 2855–2856
  in sarcoidosis, 2927
  in systemic sclerosis, 2871
Mushrooms, poisonous, 306, 324, 3711t, 3712t
Mushroom worker's lung, 2228t
MuSK (muscle-specific tyrosine kinase), 3624
Musk ambrette, 434t
MuSK antibodies, 234, 3627
Mussels, contaminated, 3727
Mustard gas, S5
Mustard (atrial switch) procedure, 2074–2075, 2075f, 2075t
Mutation(s), 3783, **3783**. *See also specific genes*
  alleles, 3787
  allelic heterogeneity in, 3787
  in cancer cells, 528. *See also* Cancer genetics
  chromosomal disorders. *See* Chromosome abnormalities/disorders
  definition of, 3783
  dynamic, 3786
  functional consequences of, 3786–3787
  germline, 3784
  identifying, 3795–3798
  imprinting of, 3791–3792
  locus heterogeneity in, 3787–3788, 3788t
  in Mendelian disorders. *See* Mendelian disorders
  mitochondrial. *See* Mitochondrial DNA diseases
  mosaicism in, 3791
  nucleotide repeat disorders, 3792–3793, 3793t
  origins and types of, 3783–3784, 3786f
  penetrance of, 3788
  phenocopies, 3787–3788
  phenotypic heterogeneity in, 3787, 3788t
  rate of, 3784
  sex-influenced, 3788
  somatic, 3783, 3792
  uniparental disomy in, 3791–3792
  variable expressivity of, 3788–3789
  X-inactivation of, 3791–3792
Mutation accumulation theory of aging, 3869
Mutation-associated neoantigens (MANAs), 526
Mutator phenotype, 520
Mutism, 201
MUTYH gene mutations, 524t, 654
MWS (Muckle-Wells syndrome), 462, 2935t, 2936
Myalgia, 3634–3636
Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), **3649**
  approach to the patient, 3649–3651
  clinical features of, 3650, 3650t
  comorbid conditions, 3650, 3650t
  course and prognosis of, 3651
  diagnosis of, 3649–3650, 3650t
  differential diagnosis of, 3650
  EBV infection and, 1510
  epidemiology of, 3649
  IC/BPS and, 332
  vs. Lyme disease, 1452
  management of, 3650–3651
  pathophysiology of, 3649, 3649f
  risk factors for, 3649
Myasthenia gravis, **3624**
  APS-2 and, 3090t, 3091t
  autoantibodies in, 35, 2790t, 2825t, 3625–3626, 3627
  vs. botulism, 1236
  clinical features of, 166, 3626
    diplopia, 234, 235
    dysphagia, 296
    nystagmus, 237
    ocular, 3626, V3
    pemphigus, 411
    ptosis, 234
    thymoma, 3092
    weakness, 166, 171, 172, 3626
  COVID-19 risk and, 3633
  diagnosis of, 3626–3627, 3627t, 3630t
  differential diagnosis of, 2910, 3627–3628
  disorders associated with, 3628, 3630t
  drug-induced, 3648t
  drugs to avoid in, 3633, 3633t
  epidemiology of, 3626
  global considerations in, 3633
  immune checkpoint inhibitor–associated, 3633
  pathophysiology of, 3624–3626, 3625–3626f
  prognosis of, 3633
  pulmonary function values in, 2205–2206, 2206f
  reflex responses in, 169
  treatment of, **3628**
    algorithm for, 3630f
    assessment of response to, 3633
    cholinesterase inhibitors, 3628–3629
    complement inhibitors, 3631, 3632t
    FcRn inhibitors, 3631, 3632t
    immunotherapy, 3629–3631
    investigational, 3632
    IVIg, 3631–3632
    plasmapheresis, 3631–3632
    thymectomy, 3629
Myasthenic crisis, 3632–3633
Myasthenic fatigue, 3625
MYB gene, 771
MYB gene mutations, 720
MYBPC3 gene mutations, 2004t, 2005
MYC, 431, 862, 3940
Mycetoma, 274, 1360, 1361f, 1365
MYC gene mutations, 520t, 660, 859, 860t
MYCL1 gene mutations, 520t
MYCN gene mutations, 520t
Mycobacterium spp.
  M. africanum, 1380
  M. bovis, 1380
  M. canetti, 1380
  M. leprae, 1051, 1404. *See also* Leprosy
  M. lepromatosis, 1404
  M. orygis, 1380
  M. pinnipedii, 1380
  M. tuberculosis. See Mycobacterium tuberculosis
  nontuberculous. *See* Nontuberculous mycobacterial (NTM) infections
Mycobacterium abscessus infections, 1051, 1163, 1414, 1418, 1427–1428, 2284
Mycobacterium avium complex (MAC) infections
  bronchiectasis, 2242, 2243. *See also* Bronchiectasis
  diagnosis of, 1417, S11
  disseminated, 1415–1416
  epidemiology of, 1416
  in HIV infection, 309, 1417, 1427, 1595f, 1597
  lymphadenopathy, 1416
  prevention of, 1417
  prophylaxis for, 1593t
  pulmonary, 1416, 1426
  in solid organ transplant recipient, 1163
  treatment of, 1417–1418, 1426–1427
Mycobacterium bovis bacille Calmette-Guérin. *See* Bacille Calmette-Guérin (BCG) vaccine
Mycobacterium spp. infections
  arthritis, 1058
  chronic meningitis, 1130t
  diagnosis of, S11
  inherited susceptibility to, 459t, 461, 461f
  M. abscessus. See Mycobacterium abscessus infections
  M. intracellulare. See Mycobacterium avium complex (MAC) infections
Mycobacterium kansasii infections
  arthritis, 1058
  diagnosis of, 1417
  in HIV infection, 1415
  pulmonary, 1416
  treatment of, 1418, 1423, 1427

- *Mycobacterium marinum* infections
  - cellulitis, 1052
  - diagnosis of, 1416–1417
  - skin, 405, 1051, 1415, 1416
  - treatment of, 1052, 1418, 1423, 1427
- *Mycobacterium tuberculosis. See also* Tuberculosis (TB)
  - delayed hypersensitivity and, 1385–1386
  - host response to, 1384, 2781, 2783
  - immune response inhibition by, 974, 1384
  - innate resistance to, 1384
  - lipids and proteins of, 1385
  - microbiology of, 1380–1381, 1380f
  - mTOR reactivation by, 973
  - phagosome of, 1383
  - rifampin-resistant, 981, 988
  - survival inside vacuole by, 968
  - transmission of, 1382
  - virulence of, 1383–1384
- Mycophenolate mofetil/sodium
  - actions of, 2408t
  - adverse effects of
    - gastrointestinal, 3630
    - in kidney transplant recipient, 2408t, 2411
    - in liver transplant recipient, 2722
    - mucosal malabsorption, 309
    - in myopathies, 2916t
    - teratogenic, 2836
    - in vasculitis, 2896t
  - for autoimmune hepatitis, 2696
  - for bullous pemphigoid, 412
  - for dermatomyositis, 414
  - for DIHS, 424
  - for epidermolysis bullosa acquisita, 413
  - for granulomatosis with polyangiitis, 2899
  - for immunosuppression
    - after cardiac transplantation, 2029t
    - after kidney transplantation, 2408, 2408t
    - after liver transplantation, 2722
  - for inflammatory myopathies, 2916–2917, 2916t
  - for lupus nephritis, 2418, 2834–2835
  - for mucous membrane pemphigoid, 413
  - for myasthenia gravis, 3630
  - for neuromyelitis optica, 3593
  - for pemphigus vulgaris, 410
  - for photoallergy, 434
  - for relapsing polychondritis, 2921
  - for sarcoidosis, 2929t, 2930
  - for SLE, 2834
  - for systemic sclerosis, 2873, 2874
  - for vasculitis, 2895
- *Mycoplasma* spp., 1143, 1464–1465
- *Mycoplasma genitalium* infections, 1098–1099, 1099t, 1104, 1466t, 1467
- *Mycoplasma hominis* infections, 1466–1467, 1466t
- *Mycoplasma* spp. infections, 424
- *Mycoplasma pneumoniae,* 1465
- *Mycoplasma pneumoniae* infections, **1465**
  - clinical features of
    - erythema multiforme, 401, 1465
    - erythema nodosum, 406
    - extrapulmonary, 1465–1466
    - rash and mucositis, 143t, 147, A1
    - upper respiratory tract, 257t, 259, 1465
  - diagnosis of, 1466, 1466t
  - epidemiology of, 1465
  - macrolide-resistant, 1027
  - pathogenesis of, 1465
  - treatment of, 1466, 1466t
- Mycosis fungoides (cutaneous T cell lymphoma)
  - clinical features of, 609f, 866
  - peripheral blood smear in, A6
  - skin manifestations of, 392, 393–394, 405, 866, A5
  - treatment of, 436, 539, 866
- Mycotic aneurysm
  - circle of Willis, 3458. *See also* Subarachnoid hemorrhage
  - clinical features of, 2171
  - diagnosis of, 2171
  - epidemiology of, 2171
  - etiology of, 2166
  - in infective endocarditis, 1039, 1047
  - treatment of, 2171
- MyD88, 2759f
- MyD88 deficiency, 459t
- *MYD88* gene mutations, 851, 851t, 865, 873, 893, 894
- Mydriasis, 221, 3705
- Myelin, 3397
- Myelin-associated glycoprotein (MAG), 893, 2790t
- Myelination, adaptive, 3397
- Myelin basic protein (MBP), 2790t, 3582
- Myelin oligodendrocyte glycoprotein (MOG), 35, 227, 2790t
- Myelin oligodendrocyte glycoprotein (MOG) antibodies, 746f, 751
- Myelin oligodendrocyte glycoprotein (MOG) antibody-associated disease, 3593, 3593t
- Myelin protein zero (MPZ), 3598
- Myelitis. *See* Myelopathy, inflammatory and immune
- Myeloblasts, 453, 454f
- Myelocytes, 453, 454f
- Myelodysplastic syndromes (MDS), **815**. *See also* Myeloid malignancies
  - classification of, 815, 815t
  - clinical features of, 816
  - definition of, 815
  - diagnosis of, 814f, 816–817
  - differential diagnosis of, 817
  - epidemiology of, 816
  - etiology of, 816
  - genetic considerations in, 816
  - MDS/MPN with ring sideroblasts and thrombocytosis, 878–879
  - pathophysiology of, 450, 816
  - prognosis of, 817, 817t
  - as second malignancy in cancer survivors, 755
  - telomere disorders and, 3819
  - treatment of, 817–818, 919t, 920, 3940
- Myelofibrosis
  - primary. *See* Primary myelofibrosis
  - secondary, 818
- Myelography, 3392–3393
- Myeloid dendritic cells, 2762, 2763t. *See also* Dendritic cells
- Myeloid/erythroid ratio, 448f
- Myeloid hyperplasia, 448f, A6
- Myeloid malignancies, **875**
  - acute myeloid leukemia. *See* Acute myeloid leukemia (AML)
  - atypical chronic myeloid leukemia, 877t, 878
  - chronic myelomonocytic leukemia, 815, 877–878, 877t
  - chronic neutrophilic leukemia, 819, 875–876, 877t
  - classification of, 875, 876t
  - with eosinophilia. *See* Eosinophilia
  - juvenile myelomonocytic leukemia, 879
  - myelodysplastic syndromes. *See* Myelodysplastic syndromes (MDS)
  - myeloproliferative neoplasm, unclassifiable, 876
  - telomere disorders and, 3819
  - transient myeloproliferative disorder, 879
- Myeloid metaplasia. *See* Primary myelofibrosis
- Myeloma. *See* Multiple myeloma
- Myeloma kidney, 2440, 2440f, A4
- Myeloma light chains, 2377, 2380, 2440
- Myelopathy
  - acute spondylitic, 3564
  - acute transverse, A16
  - cancer-related, 3569
  - chronic, **3566**
    - in adrenomyeloneuropathy, 3569
    - in cobalamin deficiency, 3568
    - hereditary spastic paraplegia, 3522–3523t, **3525,** 3569
    - HTLV-1 associated, 1553–1555, 1556, 3567
    - hypocupric, 3568
    - in MS, 3568
    - primary lateral sclerosis, 3525, 3569
    - spondylitic, 3566–3567
    - in syringomyelia, 3567–3568, 3568f
    - in tabes dorsalis, 1432, 3568–3569
    - in tethered cord syndrome, 3569
    - toxic, 3569
    - in vascular malformations of cord and dura, 3567, 3567f
  - compressive. *See* Spinal cord compression
  - compressive vs. noncompressive, 3562
  - degenerative cervical, 130, 131f
  - evaluation of, 3564t
  - gait disorders in, 178
  - in high-voltage electrical injury, 3566
  - in HIV infection, 1609
  - inflammatory and immune (myelitis), **3565**
    - acute infectious, 3566
    - in multiple sclerosis, 3565. *See also* Multiple sclerosis (MS)
    - in neuromyelitis optica, 3565
    - paraneoplastic, 731
    - postinfectious, 3566
    - in sarcoidosis, 3565–3566
    - in systemic immune-mediated disorders, 3566
  - in spinal cord infarction, 3565
- Myeloperoxidase, 455, 2790t
- Myeloperoxidase deficiency, 459t, 460
- Myelophthisic anemias, 818
- Myeloproliferative neoplasms, **818**. *See also* Myeloid malignancies
  - classification of, 815, 818t
  - essential thrombocytosis, 819, 823
  - evolution into AML, 825
  - JAK inhibitors for, 532
  - polycythemia vera. *See* Polycythemia vera
  - primary myelofibrosis. *See* Primary myelofibrosis
  - splenomegaly in, 480
  - unclassifiable, 876
- Myelosuppression, chemotherapy-induced, 574–575, 575t. *See also* Neutropenia
- Myenteric plexus, 293, 303
- *MYH7* gene mutations, 2004t, 2005
- *MYH9* gene, 1585
- *MYH9* gene mutations, 924
- *MYH11* gene mutations, 3332t
- MYH-associated polyposis, 654, 654t
- Myiasis, 3732
- *MYL* gene mutations, 2004t
- *MYLK* gene mutations, 3332t
- Myoadenylate deaminase deficiency, 3638
- Myocardial contusion, 2087–2088
- Myocardial infarction (MI)
  - abdominal pain in, 111, 112t
  - vs. acute pericarditis, 2079, 2083
  - arrhythmias in
    - AV conduction block, 1926
    - polymorphic ventricular tachycardia, 1969
    - PVCs and NSVT, 1961
    - ventricular fibrillation, 1972, 1975f
  - cardiac biomarkers in, 2108, 2109t
  - cardiac imaging following, 1893–1894, 1894f, 1902f, 1907f

  - cardiogenic shock in. *See* Cardiogenic shock
  - chest pain in, 105f, 2106–2107, 2114–2115
  - circadian variations in, 219, 3952–3953
  - classification of, 2118t
  - dental care after, 266
  - diagnosis of, 105f, 106, A7, A9
  - erectile dysfunction and, 3158
  - hemodynamic patterns in, 2329t
  - in HIV infection, 1598
  - hypokalemia in, 354, S1
  - hypothermia in, 3769
  - incidence of, 2106, 2113
  - in kidney transplant recipients, 2410–2411
  - mitral regurgitation in, 2052
  - murmur after, 284
  - natural history of, 1841
  - nausea and vomiting in, 298
  - non-ST-segment elevation. *See* Non-ST-segment elevation acute coronary syndrome (NSTE-ACS)
  - phases of, 2116f
  - postmenopausal hormone therapy and, 3144t
  - prevention of, 943
  - Q wave abnormalities in, 1873, 1873f, A7
  - radiation-associated, 753
  - right ventricular, 2081t
  - in SLE, 2829–2830
  - ST-segment elevation. *See* ST-segment elevation myocardial infarction (STEMI)
  - sudden cardiac death and, 2334, 2338
  - systems-based analysis of, 3962
  - treatment of, 954, 1930
  - in women, 3164
- Myocardial ischemia. *See also* Ischemic heart disease
  - anterior wall, 1873, 1873f, A7
  - anterolateral wall, A7
  - in aortic regurgitation, 2043
  - in aortic stenosis, 2091
  - cardiac biomarkers in, 2108, 2109t
  - clinical features of, 1841
  - diagnosis of, 1841
  - dyspnea in, 270
  - ECG in, 1872–1873, 1873f, 2091–2092, A7
  - effects of, 2091–2092, 2092f
  - lateral, A7
  - with left ventricular hypertrophy, 2091
  - nonspecific reaction to, 2116
  - pathophysiology of, 2090–2091, 2090f, 2092f
  - silent, 2092, 2105–2106
  - transmural, 1872f, 1873
  - ventricular tachycardia/fibrillation in, 1969
- Myocardial perfusion imaging
  - in CAD evaluation, 25, 25f, 1885, 1885f, 1886f, 1895f, A9
  - in chest pain evaluation, 108, A9
  - ECG-gated, A9
  - indications for, 1879
  - in ischemic heart disease, 2095, 2095–2096f, A9
  - multidimensional, 1880f
  - PET vs. SPECT, 1879
  - protocols for, 1878–1879
  - radiation exposure from, 1883
  - radiopharmaceuticals for, 1878t, 1879
  - in sarcoidosis, A9
  - in STEMI, 1894, 1894f
  - in wall motion abnormalities, 1879f
- Myocardial rupture, 2088
- Myocarditis, **2011**
  - acute, 2011–2012, 2012f
  - autoantibodies in, 2825
  - cardiogenic shock in, 2330. *See also* Cardiogenic shock
  - chronic, 2012
  - diagnosis of, 1889, 1890f, 2011, 2012f
  - dilated cardiomyopathy and, 2012, 2016t. *See also* Dilated cardiomyopathy
  - in heart transplant, 1159t
  - infective
    - bacterial, 2013
    - in COVID-19 disease, 2012
    - in diphtheria, 1223
    - eosinophilic, 2014–2015
    - giant cell, 2014
    - in HIV infection, 2012–2013
    - in influenza, 1548
    - in mumps, 1648
    - parasitic, 2013
    - in relapsing fever, 1446
    - tick-borne, 2013
    - viral, 1634t, 1635, 2012–2013, 2330
  - noninfective
    - checkpoint inhibitor–induced, 2015
    - eosinophilic, 2014–2015
    - giant cell, 2014
    - in sarcoidosis, 2014, 2014f
    - in systemic inflammatory diseases, 2015
    - vaccine-related, 2015
  - palpitations in, 291
  - pathophysiology of, 2011, 2790t
  - treatment of, 2011–2012
  - VT/VF storm in, 1975
- Myocardium, 1854, 2090–2091, 2330
- Myoclonic epilepsy with ragged red fibers (MERRF) syndrome, 3645–3646, 3811, 3811t, 3814
- Myoclonic seizures, 3410, 3516
- Myoclonus, **3515**
  - Baltic (Unverricht-Lundborg disease), 3412t, 3516
  - clinical features of, 3509t, 3515–3516
  - coma and, 189
  - in Creutzfeldt-Jakob disease, 3530
  - with dystonia, 3510t, 3511
  - posthypoxic, 2347–2348
  - treatment of, 3516
- Myofascial pain, 99, 119, 123t, 128, 132
- Myofascial pain syndrome, 2971–2972
- Myofibrillar myopathies, 3640t, 3643–3644
- Myogenic ptosis, 234
- Myogenic reflex, 2364, 2373
- Myoglobin, 2377
- Myoglobinuria, 3645
- Myokymia, 3555, 3578, V3
- Myonecrosis
  - anaerobic bacterial, 1377
  - clostridial. *See* Clostridial myonecrosis
  - etiology of, 1050t
  - with necrotizing fasciitis, 1053, 1053f
  - synergistic nonclostridial anaerobic, 1053
- Myopathy(ies)
  - acute necrotizing, 749, 2352
  - alcoholic, 3677
  - clinical features of, **3634**
    - gait disorders in, 179
    - muscle enlargement and atrophy, 3636–3637
    - pain, cramps, and stiffness, 3634–3636, 3635t
    - weakness, 168, 169, 3634, 3635t, 3636f, 3637t, 3638f
  - in critical/systemic illness, 2352, 3647
  - diagnosis of, 3636f, 3637–3638
  - distal, 3640t, 3643–3644
  - drug-induced, 2940t, 3251, 3648–3649, 3648t
  - endocrine and metabolic, 3647
  - global considerations in, 3648
  - hereditary. *See* Muscular dystrophy
  - in HIV infection, 1610
  - inflammatory. *See* Inflammatory myopathies
  - in lung cancer, 619
  - mitochondrial. *See* Mitochondrial myopathy(ies)
  - myofibrillary, 3640t
  - oculopharyngeal, 3643
  - thick-filament, 2352
- Myopia, 220, 3332
- Myosin, 1848, 1848f, 2005f, 2790t
- Myosin-binding protein C gene mutations, 3789
- Myosin light chain, 1846
- Myositis
  - autoimmunity in, 2790t
  - clinical features of, 1053
  - etiology of, 1050t, 1053
  - immune checkpoint inhibitor–associated, 3633
  - inclusion body. *See* Inclusion body myositis
  - in influenza, 1548
  - streptococcal, 1053, 1211
  - treatment of, 1054
  - in trichinellosis, 1811
- Myositis ossificans progressiva, 3319
- Myositis-specific antibodies (MSA), 2825, 2910, 2911
- Myotonia, 3635
- Myotonia congenita, 3635
- Myotonic dystrophy, **3642**
  - AV conduction block in, 1927
  - clinical features of, 50t, 3642
  - diagnosis of, 3642
  - genetic considerations in, 3642, 3786, 3793, 3793t
  - ocular involvement in, 234
  - proteins involved in, 3641f
  - testicular dysfunction in, 3115
  - treatment of, 3642
- MyPlate Plan, 2599
- Myringotomy, 255
- Myxedema
  - clinical features of, 3029
  - edema in, 282
  - etiology of, 2180
  - hypothermia in, 3768, 3771
  - pericardial effusion in, 2083
  - pretibial, 2180, A5, A15
- Myxedema coma, 3032
- Myxofibrosarcoma, 727
- Myxoma, **2085**
  - clinical features of, 2085
  - CMR in, 1898, 2086f
  - dyspnea in, 270
  - echocardiography in, 2085–2086, 2085–2086f
  - epidemiology of, 2085
  - heart sounds/murmur in, 288, 2085
  - vs. mitral stenosis, 2050
  - palpitations in, 291
  - pathology of, 2085
  - skin manifestations of, 399
  - syncope in. *See* Cardiac syncope
  - treatment of, 2086
- Myxovirus resistance protein A (MxA), 2911, 2912f

N

- Na^+. *See* Sodium
- NAAT (nucleic acid amplification test), 976, 976t
- Nab-paclitaxel, 563t, 565, 676t
- Nab-sirolimus, 729
- Nabumetone, 2955
- *N*-acetyl-benzoquinone-imine (NAPQI), 2668
- *N*-acetylcysteine (NAC)
  - for acetaminophen overdose, 489, 2670
  - for alcohol-associated hepatitis, 2710
  - in *H. pylori* eradication, 2528
  - for prevention of contrast-induced nephropathy, 1904
- Na^+-Cl^- co-transporter, thiazide-sensitive, 345, 359
- Nadofaragene firadenovec, 559f, 561, 693

Nadolol, 1970, 2100t, 3712t
NADPH (nicotinamide adenine dinucleotide phosphate) oxidase, 453, 460, 970, 1689, 1690f
Naegleria, 1758
Naegleria infections
  clinical features of, 1114, 1739t, 1758
  diagnosis of, 1739t, 1758
  epidemiology of, 1758, S12
  laboratory evaluation of, 1758, S12
  pathogenesis of, 1758
  treatment of, 1750, 1758
Nafamostat, 359
Nafcillin
  adverse effects of, 355
  for cellulitis, 1054t
  drug interactions of, 1172t
  indications for, 1173t, 1175
  for infective endocarditis, 1044
  for meningitis, 1121t, 1122t
  for nongonococcal bacterial arthritis, 1057
  for osteomyelitis, 1064t, 1065t
  resistance to, 1041, 1173t
  for staphylococcal infections, 1203t, 1204
Nafcillin/oxacillin, 964t
Nail(s)
  in amyloidosis, 897, 897f
  in cardiovascular disease, 1860
  in dermatomyositis, 2911, 2911f
  drug-induced disorders of, 421
  in dyskeratosis congenita, 3820f
  in HIV infection, 1606
  in lichen planus, 388, 388f, A5
  onychomycosis, 389
  in psoriasis, 386
  in systemic sclerosis, 415, 2867, 2867f, 2868f
  telangiectasia, 396
Nail-patella syndrome, 2415t, 2427
NAIP-NLRC4 inflammasome, 972
Nairobi fly, 3737
Nairobi sheep disease virus, 1662t
Nairoviridae, 1478f, 1480t, 1656, 1661–1662t, 1664–1665t, 1666t
Na^+-K^+-$2Cl^-$ cotransporter, 345, 2369
Na^+/K^+-ATPase, 345, 2369
Nalidixic acid, 434t, 799t, 923t, 1166
NALIFIRINOX regimen, 674
NALIRIFOX regimen, 676T
Nalmefene, 3681
Naloxone, 98, 3690–3691, 3690t
Naltrexone
  in alcoholism management, 3680–3681
  with bupropion, for obesity, 3191, 3192t, 3193f, 3257
  for opioid-related disorders, 3691
NAME syndrome. *See* Carney syndrome/complex
Naming deficit, 201
^{13}N-ammonia, 1878t, 1879
NAP-2 (neutrophil-activating protein-2), 2768t
Naphthalene, 798, 799t
NAPQI (*N*-acetyl-benzoquinone-imine), 2668
Naproxen
  adverse effects of, 401, 923t, 2955, 2955t
  for dysmenorrhea, 3137
  for gout, 2959
  for medication-overuse headache, 115
  for migraine, 3467t
  for osteoarthritis, 2955t
  for pain, 97t
  for rheumatic fever, 2857
Napsin-A (Nap-A), 616
Naratriptan, 3467t, 3468, 3468t
Narcolepsy, **214**
  cataplexy in, 161, 214, 220v
  diagnosis of, 213t, 214, 214f
  immunopathogenesis of, 3398
  olfactory loss in, 241–242
  treatment of, 215
Narcotics. *See* Opioid(s)
NARES (perennial nonallergic rhinitis with eosinophilia syndrome), 2810
Nares, microbiota of, 3833f
NARP (neuropathy, ataxia, retinitis pigmentosa) syndrome, 3811–3812, 3811t
Nasal congestion, 253, 1537, 2758, 2811f
Nasal polyposis, 2222
Nasal vestibulitis, 507t, 510
NASBA test, for HIV RNA, 1590t
NASH (nonalcoholic steatohepatitis). *See* Metabolic dysfunction–associated steatohepatitis (MASH)
Nasogastric tube, 295, 296, 2623
Nasopharyngeal cancer, 611, 613, 1510, 1512, 2239t. *See also* Head and neck cancer
Nasu-Hakola disease, 3480
NAT (*N*-acetyl transferase), 493
NAT2, 493, E72t
Natalizumab
  actions of, 2792t
  adverse effects of, 3586t
  for IBD, 2566
  indications for, 2792t
  for MS, 3585, 3586t
Nateglinide, 3211t, 3213
National Academy of Medicine, 52, 53, 64
National Childhood Vaccine Injury Act (NCVIA), 1007
National Cholesterol Education Project Adult Treatment Panel III (ATP III), S10
National Family Caregivers Association, 87
National Healthcare Safety Network, 1146
National Institutes of Health National Center for Complementary and Alternative Medicine (NCCAM), 3927t, 3929t
National Lung Cancer Screening Trial (NLST), 617–618, 618t
National Quality Forum, 43
National Vaccine Injury Compensation Program (NVICP), 1001
Natpara, 366
Natural cytotoxicity receptors (NCRs), 1580, 2768
Natural killer (NK) cells
  in autoimmune disease, 2824
  defects in, 465, 1497
  definition of, 2785
  development of, 2770–2771f
  functions of, 2756, 2763t, 2768, 2771f
  invariant, 3840
  in sepsis/septic shock, 2319
  signaling of, 2768, 2773f
  surface receptors of, 2763, 2768
Natural killer T (NK-T) cells
  in adaptive immunity, 2763t
  definition of, 2756
  functions of, 2756, 2768
  in HBV infection, 2649
  hematopoietic differentiation of, 760f
  in HIV infection, 1580, 1588
  in innate immunity, 2763t, 2768–2769
  neoplasms of, 874
  receptors of, 2768
  signaling of, 2768
Natural language processing, 3973–3974
Natural orifice transluminal endoscopic (NOTES) surgery, 2469
Natural products, 3924–3925, 3925t
Naturopathy, 3925t, 3928
Nausea, 296, 297
Nausea and vomiting, **297**
  in appendicitis, 2595
  approach to the patient, 298
  in cancer patients, 507t, 509
  chemotherapy-induced, 507t, 508–509, 575
    acute emesis, 298
    anticipatory, 83, 298, 575
    assessment of, 508
    delayed emesis, 298
    treatment of, 82–83, 299–300, 508–509, 575
  differential diagnosis of, 297–298, 297t, 2460t
  etiology of, 82
  gastrointestinal causes of, 2460
  in liver disease, 2628
  mechanisms of, 297
  in myocardial infarction, 106
  in pregnancy, 298, 300, 3903
  radiation-induced, 554
  in terminally ill patient, 82–83
  treatment of, 298–299, 299t, 508–509
Naxitamab, 534t
Naxos syndrome, 2004t
NBIA (neurodegeneration with brain iron accumulation), 3518–3519
NBTE (nonbacterial thrombotic endocarditis), 1038
NCAM, 616
NCRs (natural cytotoxicity receptors), 1580, 2768
NCVIA (National Childhood Vaccine Injury Act), 1007
ND4 gene mutations, 3807
NDM carbapenemases (New Delhi metallo-β-lactamases), 1181
Nebivolol, 1988, 2100t, 2144t
Necator americanus, 1815t, 1816, S12. *See also* Hookworm infections
NECDIN gene, 2993, 3791
Necitumumab, 534t
Neck pain, **129**
  alarm symptoms, 130t
  approach to the patient, 129
  autoimmune/inflammatory, 132
  congenital, 132
  degenerative, 130
  endocrine/metabolic, 132
  infectious, 131
  neoplastic, 130–131
  in orthostatic hypotension, 159
  radicular, 129–130, 129t
  traumatic, 132
  treatment of, 132
  vascular, 131
Neck weakness, 3558
Necrobiosis lipoidica, 404
Necrobiosis lipoidica diabeticorum, 3230, A15
Necrolytic migratory erythema, 394, 678, 679f
Necroptosis, 973
Necrosis
  digital, in systemic sclerosis, 2867, 2867f
  fat, 406
  intestinal, 361
  pancreatic, 2469, 2649, 2747f, V5
  pulp, 260
Necrotic arachnidism, 3734
Necrotizing arteritis, 3447
Necrotizing enterocolitis, 912, 1240
Necrotizing fasciitis
  clinical features of, 993, 1051, 1052–1053, 1211
  etiology of, 991t, 993, 1050t, 1052, 1211, 1377
  imaging of, 1053f
  pathophysiology of, 1052–1053, 1211
  treatment of, 991t, 1054, 1054t, 1208t, 1211, 2321t
Necrotizing myopathy, 751, 2909, 2910t, 2913f, 2915–2916, 2916t
Necrotizing pancreatitis. *See also* Acute pancreatitis
  diagnosis of, 2746t, 2747f
  intraabdominal abscess in, 1073
  pathogenesis of, 2744
  treatment of, 2469, V5

- Necrotizing vasculitis, 379, 379f, 382, 406, 2769
- Nedd4-2, 2372
- *NEDD9* gene mutations, 2187f
- Neer test, 2943
- Nef/*nef,* 1552f, 1561f, 1576, 1585
- Negative predictive value, 514, 514t
- Negative-strand RNA viruses, 1477, 1478f, 1479, 1480t
- Neglect (sensory phenomenon), 174, 203–204, 205f, 209
- Neglected tropical diseases, 1413, 1810
- Neglect, elder, 3895t
- Negligible senescence, 3869, 3871f
- Negri body, 1651, 1653f
- *Neisseria* spp. (meningococci). *See also* Meningococcal infections
  - capsular groups of, 1244, 1244t
  - laboratory identification of, S11
  - pili, 966, 966t
- *Neisseria gonorrhoeae. See also* Gonococcal infections
  - antibiotic resistance in, 1100f, 1181, 1183, 1186, 1186t, 1255
  - host response to, 1254–1255
  - laboratory identification of, 1258, S11
  - lipooligosaccharide of, 1254
  - microbiology of, 1253
  - outer-membrane proteins of, 1254
- *Neisseria meningitidis. See also* Meningococcal infections
  - antibiotic resistance in, 1122, 1183
  - asymptomatic carriage of, 1247
  - colonization by, 1247
  - genotyping of, 1245
  - host response to, 1247–1248
  - laboratory identification of, S11
  - outer-membrane vesicles, 1252, 1252f
  - structure of, 1244–1245, 1244f, 1244t, 1245f
  - virulence factors of, 1247
- *Neisseria meningitidis* vaccine. *See* Meningococcal vaccine
- Nelarabine, 849
- *NELF* gene mutations, 3111t, 3112
- Nelfinavir, 1619f, 1623f, 1744t
- Nelson's syndrome, 400, 3012
- Nemaline rod myopathy, 3634, 3641f
- Nematocyst, 3722, 3723f
- Nematodes/nematode infections, 1809–1810
  - filarial
    - *Brugia* spp., 1819, 1820t, S12. *See also* Lymphatic filariasis
    - *Dracunculus medinensis*/dracunculiasis. *See Dracunculus medinensis*/dracunculiasis
    - *Loa loa. See Loa loa* (African eye worm)/loiasis
    - *Mansonella* spp., 1820t, 1824, S12
    - *Onchocerca volvulus. See Onchocerca volvulus*/onchocerciasis
    - tropical pulmonary eosinophilia, 462, 1822
    - *Wuchereria bancrofti,* 961t, 1819, 1820t, 2234, S12. *See also* Lymphatic filariasis
    - zoonotic, 1824–1825
  - intestinal, 1736, **1814,** 1815t
    - anisakiasis, 1818
    - *Ascaris lumbricoides. See Ascaris lumbricoides*/ascariasis
    - *Capillaria philippinensis*/capillariasis, 961t, 1818–1819, S12
    - *Enterobius vermicularis*/enterobiasis, 1815t, 1818, S12
    - hookworm. *See* Hookworm infections
    - *Strongyloides stercoralis. See Strongyloides stercoralis*/strongyloidiasis
    - trichostrongyliasis, 1818
    - *Trichuris trichiura*/trichuriasis (whipworm), 961t, 1736, 1815t, 1818
  - by organ system and signs/symptoms, 1738–1740t
  - tissue, 1736–1737, **1810**. *See also* Larva migrans
    - *Ancylostoma* spp. *See Ancylostoma* spp.
    - *Angiostrongylus. See Angiostrongylus* spp. infections
    - *Baylisascaris procyonis,* 961t, 1114, 1131t, 1812, S12
    - *Gnathostoma. See Gnathostoma spinigerum*/gnathostomiasis
    - *Toxocara canis/cati,* 961t, 1739t, 1812, S12
    - *Trichinella. See Trichinella*/trichinellosis
- NEMO (NF-κB essential modulator)
  - deficiency of, 459t, 460, 2802, 2804, S9
  - in NTM infections, 1415, 1415f, 1416
- Neoadjuvant therapy, 554, 625
- Neoantigens, tumor, 2785, 2788
- Neobladder, orthotopic, 694
- Neomycin
  - adverse effects of, 309, 423, 786t, 2717
  - for IBS, 2576
  - for otitis externa, 254
  - prophylactic, 1180t
  - vitamin A absorption and, 2611
- Neonatal-onset multisystem inflammatory disease (NOMID), 2935t, 2936
- Neonates
  - consumptive hypothyroidism in, 740
  - diabetes in, 3203
  - folate requirements in, 788, 790
  - infections in
    - babesiosis, 1777
    - *Candida,* in brain, A16
    - CMV, 1513, 1513t, 3903
    - *C. trachomatis,* 1472, 1473t, 1474
    - enterococcal, 1218
    - enterovirus (generalized disease of the newborn), 1634, 1634t
    - gonococcal, 1257, 1259t
    - hepatitis B, 2650, 2653
    - herpesvirus, 3903–3904
    - HIV, 1568–1569
    - HSV, 1500–1501, 1503t, 1504
    - *L. monocytogenes,* 1229
    - rubella. *See* Rubella, congenital
    - syphilis. *See* Syphilis, congenital
    - *T. cruzi,* 1788, 1790, 1790t
    - tetanus, 1230, 1232
    - *T. gondii,* 1797, 1798, 1800–1801, 1801t, 1802
    - varicella, 1505
  - jaundice in, 796, 798, 2639
  - lupus in, 2836
  - neutropenia in, 458
  - screening programs, 3977–3978
  - seizures in, 3411, 3413t
  - vitamin K deficiency in, 2613
- *Neorickettsia,* 1460
- Neostigmine, 2593, 3712t, 3721, 3721r
- Nephrectomy, 689
- Nephrocalcinosis, 2357, 2443
- Nephrogenic diabetes insipidus. *See* Diabetes insipidus (DI), with AVP resistance (nephrogenic)
- Nephrogenic fibrosing dermopathy, 2394
- Nephrogenic systemic fibrosis, 420, 1884, 3389
- Nephrolithiasis, **2448**
  - abdominal pain in, 112t
  - approach to the patient, **2450**
  - clinical features of, 338t, 2450
  - conditions associated with, 1075, 2448
  - diagnosis of, 338t, 2450–2451, 2451f
  - differential diagnosis of, 2450
  - epidemiology of, 2448
  - follow-up for, 2453
  - genetic considerations in, 2450
  - history in, 2451
  - in hyperparathyroidism, 3277
  - in IBD, 2562t, 2563
  - inherited, 2368t
  - pathogenesis of, 2448–2449
  - prevention of, 2452
  - risk factors for, 2449–2450
  - treatment of, 2450–2451, 2452–2453
- Nephron(s)
  - embryologic development of, 2364, 2364f
  - functions of, **2365**
    - cellular transport, 2365–2366
    - in collecting duct, 2367f, 2370
    - in distal convoluted tubule, 2367f, 2370
    - epithelial solute transport, 2365
    - in loop of Henle, 2367f, 2369
    - membrane transport, 2365–2366
    - in proximal tubule, 2366–2367, 2366f
  - inherited disorders involving, 2368t
- Nephronophthisis, 2430t, 2431f, 2435
- Nephropathia epidemica, 1676
- Nephropathy
  - analgesic, 2357, 2361f, 2442
  - aristolochic acid, 2357, 2377, 2442, 2444
  - atherosclerotic, 2415t, 2427
  - calcineurin-inhibitor, 2410, 2443
  - contrast-induced, 1904, 2377, 2379t, 2394, 3384, 3386
  - diabetic. *See* Diabetic nephropathy
  - endemic, 2397
  - heavy metal, 2443
  - HIV-associated, 1585, 1601, 2428, A4
  - hypercalcemic, 2443
  - hypokalemic, 2444
  - IgA. *See* IgA nephropathy
  - in lead poisoning, 2443
  - light chain cast. *See* Light chain deposition disease/nephropathy
  - lithium-associated, 2442–2443
  - membranous. *See* Membranous glomerulonephritis
  - phosphate, 2440, A4
  - polyomavirus-associated, 1161
  - quartan malarial, 1766
  - reflux, 2441–2442, 2441f
  - sickle cell, 774t, 2415t, 2427, 2442, 2447
  - urate, 2440
  - uric acid. *See* Uric acid nephropathy
- Nephrosclerosis, hypertensive, 2154, 2427, A4
- Nephrotic syndromes, **2421**
  - clinical features of, 338t, 2357–2359, 2358f, 2414, 2415t, 2421
    - acid-base disorders, 368
    - ascites, 328
    - edema, 281, 281t
    - hypoalbuminemia, 342
    - hyponatremia, 348
  - diabetic nephropathy. *See* Diabetic nephropathy
  - diagnosis of, 338t
  - dyslipidemia in, 3247
  - etiology of, 2357–2359, 2358f
  - in Fabry's disease. *See* Fabry's disease
  - focal segmental glomerulosclerosis. *See* Focal segmental glomerulosclerosis
  - genetic considerations in, 2357, 2359t
  - glomerular deposition diseases, 2424–2425
    - fibrillary glomerulopathy, 2425
    - fibrillary-immunotactoid glomerulopathy, 2415t
    - immunotactoid glomerulopathy, 2425
    - light chain deposition disease. *See* Light chain deposition disease
    - renal amyloidosis, 2415t, 2424–2425, A4
  - membranous glomerulonephritis. *See* Membranous glomerulonephritis

Nephrotic syndromes (*Cont.*):
  minimal change disease. *See* Minimal change disease
  paraneoplastic, 2359
  treatment of, 2421
Nephroureterectomy, 694
Neprilysin, 1980
Neprilysin inhibitors. *See* Sacubitril-valsartan
Nepuyo virus, 1659t, 1667
Neratinib, 568t, 570, 637, 641, 753t, 754
Nerve agents, 3712t, S5
Nerve biopsy, 3598
Nerve conduction studies
  in inflammatory myopathies, 2910
  in leprosy, 1410
  in low back pain, 126
  in muscle disease, 3637
  in neck pain, 130
  in peripheral neuropathy, 3597, 3597t
Nerve function testing, 1410, 1412
Nesidioblastosis, 3235
Nesiritide, 1991, 1993t, 2331
Netherlands, physician-assisted suicide in, 89, 89t
Netilmicin, 1166
NETs. *See* Neuroendocrine tumors (NETs)
NETs (neutrophil extracellular traps), 317, 942, 970, 2317, A10
Network medicine, 6, 3962, 3962f
Networks
  disease, 3957–3962, 3958–3959f, 3959–3960f
  drug repurposing using, 32, 34f
  gene, 33f, 3955, 3959–3961f
  integration of multiomics data in, 3960, 3962f
  neural, 201
  organization of, 3955–3956, 3956f
Neuralgia(s)
  glossopharyngeal, 265, 3555–3556
  migrainous, 265
  postherpetic, 265, 1506, 1507, 3607
  trigeminal. *See* Trigeminal neuralgia
Neural interfaces, implantable, 3967–3969, 3968–3969f
Neural network, 201
Neural stem cells, 3401–3402, 3405
Neural tube defects, 784
Neuraminidase, 1187, 1188f
Neuraminidase inhibitors, 1487–1488, 1549–1550
Neurexin 3 alpha antibodies, 749
Neuritic plaque, in AD, 3478, 3478f
Neuritis
  in leprosy, 1409, 1412
  optic. *See* Optic neuritis
  oral, 265
Neuroacanthocytosis, 3514
Neuroblastoma, 504t, 520t, 521
Neurocysticercosis
  brain abscess in, 1136
  clinical features of, 1138, 1833
  diagnosis of, 1138, 1139f, 1833–1834, 1834t, 1835f, S2
  etiology of, 1138
  treatment of, 1138, 1834–1835
Neurodegeneration with brain iron accumulation (NBIA), 3518–3519
Neuroendocrine tumors (NETs)
  appendix, 680–681
  carcinoid syndrome in. *See* Carcinoid syndrome
  carcinoma of unknown primary and, 733t, 735
  clinical features of, 309
  ectopic ACTH production in, 679, 738
  esophageal, 652
  gastric, 652, 679–680, 680f, 3081
  high-grade, 686
  histologic classification and molecular features, 677, 677t
  incidence and prevalence, 676–677, 677f
  in MEN 1, 676–677, 679, 3079, 3080t. *See also* Multiple endocrine neoplasia type 1 (MEN 1, Wermer's syndrome)
  metastatic, 681, 681f, 684
  pancreatic
    glucagonoma, 678–679, 678t, 679f, 3082
    growth hormone-releasing, 3083
    insulinoma. *See* Insulinoma
    in MEN 1, 679, 3080t, 3082
    nonfunctioning, 679, 3082–3083, 3082f
    other secretory, 679
    paraneoplastic syndromes in, 737t
    screening tests for, 3081t
    somatostatinoma, 309, 679
    VIPoma. *See* VIPoma
  rectal, 681
  small bowel, 652, 680, 680f, 681f
Neuroepithelioma, peripheral, 730
Neurofascin 155, 3623
Neurofibrillary tangles (NFTs)
  in AD, 3475, 3477–3478f, 3478f
  in chronic traumatic encephalopathy, 3484
  proteostasis loss and, 3871, 3873f
Neurofibroma, 403, 727
Neurofibromatosis type 1 (NF1)
  clinical features of, 717t, 722, 3077, **3087**
  genetic considerations in, 722, 3088, 3784
    chromosomal location, 524t, 3080t
    NF1 gene mutations, 524t, 722, 3077
  MRI in, A16
  pheochromocytoma in, 3077
  skin manifestations of, 399, 403, 722, 3077, A5
  soft tissue sarcomas in, 727
Neurofibromatosis type 2 (NF2)
  clinical features of, 248, 717t, 723, 3087
  CNS tumors in, 717t, 723
  genetic considerations in, 524t, 717t, 722
  MRI in, A16
Neurofibromin, 717t, 722, 3088
Neurogenic bladder, 3549
Neurogenic claudication, 120
Neurogenic ptosis, 234
Neurohormonal system, in heart failure, 1980, 1981f
Neurohypophysis. *See* Posterior pituitary
Neuroimaging, **3383, A16**. *See also specific disorders*
  angiography, 3393–3394
  closed-loop, 3966, 3966f
  of connectivity, 3966
  CT, **3383,** 3383t. *See also* Computed tomography (CT), in neurologic disease
  in interventional procedures, 3393
  MRI, **3386**. *See also* Magnetic resonance imaging (MRI), in neurologic disease
    arterial spin labeling, 3390
    complications of, 3389
    contraindications to, 3389, 3390t
    contrast agents, 3387–3389
    diffusion tensor, 3390, 3392f
    diffusion-weighted, 3385f, 3388f, 3390
    echo-planar, 3390
    FLAIR. *See* Fluid-attenuated inversion recovery (FLAIR) MRI
    functional, 95f, 3390, 3966, 3966f
    guidelines for, 3383t
    magnetic resonance neurography, 3390
    perfusion, 3390
    technique, 3386–3387, 3387t
  myelography, 3392–3393
  positron emission tomography, 3392. *See also* Positron emission tomography (PET), in neurologic disease
  in spine interventions, 3393
Neurokinin (NK_1), 297
Neurokinin (NK_1) receptor antagonists, 274, 298, 299t, 575
Neuroleptic malignant syndrome (NMS), 3517
Neuroleptics. *See also* Antipsychotics
  adverse effects of, 86, 3157, 3516–3517
  for delirium, 86, 86t
  for Tourette's syndrome, 3515
Neurolisteriosis, 1228
Neurologic disease. *See also specific diseases*
  approach to the patient. *See* Neurologic examination/consultation
  in cobalamin deficiency, 783
  in critically ill patient. *See* Critically ill patient, neurologic disorders in
  depression in, 3664
  electrodiagnostic studies of. *See specific techniques and* Electrodiagnostic studies
  fatigue in, 166
  global burden of, 3377, 3377f
  in HIV infection. *See* HIV infection, neurologic disease in
  in hypothermia, 3769t
  hypoventilation in, 2274, 2274f
  imaging studies in. *See* Neuroimaging
  insomnia in, 215
  involuntary weight loss in, 315
  as late consequence of cancer/cancer therapy, 756–757
  localization of, 3377–3378, 3382–3383, 3382t
  paraneoplastic. *See* Paraneoplastic syndromes, neurologic
  pathobiology of
    astrocytes, 3396–3397
    lymphatics, 3397–3398
    microbiota, 3398–3399
    microglia and macrophages, 3394–3396, 3395f
    myelin, 3397
    neuroimmunology and neuroinflammation, 3394
    oligodendrocytes, 3397
    pathologic proteins, 3399–3400
    prions. *See* Prion(s)
  in pregnancy, 3902–3903
  prevalence of, 3377
  rehabilitation for. *See* Neurotherapeutic technologies
  in sarcoidosis, A16
  in sepsis/septic shock, 2321–2322, 2322t
  stem cell therapy in, 3401–3402
  in syphilis. *See* Neurosyphilis
  testicular dysfunction in, 3115
  in Wilson's disease, 3338
Neurologic examination/consultation, 3377
  in botulism, 1236
  in coma, 189–190, 190f
  comprehensive, V7
  coordination examination, 3382
  cranial nerve examination, 3380
  in delirium, 183–184
  in dizziness and vertigo, 162–163
  gait examination, 3382
  in headache, 114, 117
  imaging studies. *See* Neuroimaging
  localization of findings in nervous system, 3382–3383, 3382t
  mental status examination, 3379–3380
  motor examination, 3380–3381
  neurologic history in, 3378–3379
  neurologic method in, 3377–3378
  in olfactory/taste dysfunction, 243
  overview of, 3379, V7
  reflex examination, 3381
  screening, V6
  sensory examination, 3381–3382

- Neurologic-myopathic syndromes, in lung cancer, 619
- Neuroma, 404, 3085. *See also* Vestibular schwannoma (acoustic neuroma)
- Neuromodulation therapy, 336, 3467t, 3469, 3470t, 3472, 3967–3968. *See also* Vagus nerve stimulation
- Neuromuscular blocking agents. *See* Muscle relaxants
- Neuromuscular junction, 3624, 3625–3626f
- Neuromuscular junction disorders, **3624**
  - autonomic dysfunction in, 3549–3550
  - botulism. *See* Botulism
  - clinical features of, 3382t
  - congenital myasthenic syndrome, 3627–3628, 3629t
  - in critically ill patient, 2351–2352
  - Lambert-Eaton myasthenic syndrome. *See* Lambert-Eaton myasthenic syndrome (LEMS)
  - myasthenia gravis. *See* Myasthenia gravis
  - progressive external ophthalmoplegia. *See* Progressive external ophthalmoplegia
  - weakness in, 169
- Neuromyelitis optica (NMO), **3590**
  - antibodies in, 35, 227, 750, 3565, 3590
  - associated conditions, 3591–3592
  - clinical features of, 3590
  - course of, 3590
  - diagnosis of, 227, 3565, 3590, 3590t, 3591f, A16, S2
  - global considerations in, 3590–3591
  - immunology of, 3565
  - molecular mimicry in, 3398
  - vs. MS, 3590
  - treatment of, 227, 3565, 3592–3594
- Neuromyotonia, 748, 751, 3635
- Neuronal ceroid lipofuscinosis type 2 (CLN2), 3354t, 3357
- Neuronal nicotinic acetylcholine receptor, 2791t
- Neuron-specific enolase (NSE)
  - as brain damage indicator, 2346
  - in carcinoma of unknown primary, 733t
  - as tumor marker, 504t, 616
- Neuro-ophthalmology, V3
- Neuropathic joint (Charcot joint/foot) disease, 1058, 1068, 1068f, 2968–2969, 2969f, 2969t
- Neuropathic (lower motor neuron) weakness, 168, 168t, 169
- Neuropathy. *See also* Peripheral neuropathy; Polyneuropathy
  - acute quadriplegic, 3648
  - brachial plexus, 3614–3615
  - diabetic. *See* Diabetic neuropathy
  - entrapment, 3903
  - femoral, 3614
  - gait disorders in, 179
  - in HIV infection. *See* HIV infection, neurologic disease in
  - immune-mediated, 3619t
    - anti-Hu paraneoplastic neuropathy, 3624
    - CIDP, 3622
    - Guillain-Barré syndrome. *See* Guillain-Barré syndrome (GBS)
    - in MGUS, 3623
    - Miller Fisher syndrome. *See* Miller Fisher syndrome
    - multifocal motor neuropathy, 3623
    - in multiple myeloma, 3623
    - vasculitic neuropathy, 3623–3624
  - lateral femoral cutaneous, 3614, 3903
  - in leprosy, 1412
  - low back pain and, 118, 118t, 124t
  - median, 3613
  - optic. *See* Optic neuropathy
  - pain in, 80, 80f, 96, 100–101, 506, 508, 508t
  - paraneoplastic autonomic, 751
  - peroneal, 3614
  - radial, 3613–3614
  - sacral, 314
  - sciatic, 3614
  - sensory deficits in, 175–176
  - treatment of, 3604t
  - trigeminal, 3553
  - ulnar, 3613
- Neuropathy, ataxia, retinitis pigmentosa (NARP) syndrome, 3811–3812, 3811t
- Neuropeptide Y, 3182
- Neuroprotective agents, 3443, 3504–3505
- Neuroregulin 1, 572
- Neurosarcoidosis, 1132t, A16, S2. *See also* Sarcoidosis
- Neurosyphilis
  - asymptomatic, 1431–1432
  - clinical features of, 111, 179, 1126, 1130t, 1432, 3484
  - in congenital syphilis, 1433
  - diagnosis of, 1127, 1433, A16, S2
  - in HIV infection, 1602
  - treatment of, 1434t, 1435
  - vs. viral meningitis, 1124
- Neurotherapeutic technologies, **3963**
  - implantable neural interfaces, 3967–3969, 3968–3969f
  - neurogaming, 3964–3966
  - neuroimaging, 3966, 3966f
  - noninvasive brain stimulation, 3425, 3966–3967. *See also* Transcranial magnetic stimulation (TMS)
  - robotics, 3963–3964, 3964f
  - virtual and augmented reality, 3964, 3965f
- Neurotransmitter(s), 1846
- Neurotrophins, 2218
- Neurotropic tyrosine kinase receptor (NRTK), 572
- Neutral protamine Hagedorn (NPH) insulin, 3209t, 3210
- Neutron (η) particles, S6
- Neutropenia, **456**
  - acute, 456
  - antibacterial prophylaxis for, 1180t
  - in aplastic anemia, 813
  - autoimmune, 457, 457t
  - in cancer patient. *See also* Cancer, infections in
    - antibiotic prophylaxis in, 577
    - chemotherapy-induced, 574, 575t, 577, 578t
    - fever with, 579–581
    - risk factors for, 578t
    - treatment of, 574–575, 575t
  - chronic, 456, 463
  - cyclic, 136, 261, 457
  - definition of, 453
  - diagnosis of, 462
  - drug-induced, 456–457, 457t
  - enterocolitis associated with, 1376
  - etiology of, 456, 457t
  - febrile, 1312
  - genetic considerations in, 493
  - in HCT recipient. *See* Hematopoietic cell transplantation (HCT), complications of
  - hereditary, 457–458
  - in HIV infection, 1605
  - refractory. *See* Myelodysplastic syndromes (MDS)
  - severe congenital, 2794–2795, 2795f
  - skin manifestations of, A5
  - treatment of, 463
  - in viral hepatitis, 2656
- Neutropenic enterocolitis (typhilitis), 582, 597, 597f, 1376
- Neutrophil(s), **453**
  - abnormalities of, **456**
    - autoimmune disease and, 2824
    - chronic granulomatous diseases. *See* Chronic granulomatous diseases (CGDs)
    - granule disorders, 459t, 460
    - inherited disorders, 459t
    - leukocyte adhesion disorders, 458, 459t, 460
    - neutropenia. *See* Neutropenia
    - phagocyte activation disorders, 459t, 460–461
  - in adaptive immunity, 2763t
  - apoptosis of, 455
  - circulating pools of, 453, 456f
  - in control of fungal infections, 1689
  - hypersegmented, 439f
  - in innate immunity, 2763t, 2769
  - marginated, 453, 456f
  - marrow release of, 453–455
  - maturation of, 453, 454f
  - morphology of, 455f
  - in peripheral blood smear, 442f, 443, 455f
  - senescent, 445
  - shift to the left, 453
  - travel through pulmonary capillaries, 453–454, 457f
- Neutrophil-activating protein-2 (NAP-2), 2768t
- Neutrophil bands, 453, 454f
- Neutrophil count, 457
- Neutrophil extracellular traps (NETs), 317, 942, 970, 2317, A10
- Neutrophil gelatinase associated lipocalin (NGAL), 2382
- Neutrophil granules, 453, 459t, 460, 2317
- Neutrophilia, 458, 458t
- Nevirapine
  - adverse effects of
    - cutaneous, 401, 422, 1617t
    - genetic considerations in, 1585
    - hepatotoxicity, 1601, 2672
    - in HIV infection, 1617t
    - SJS or TEN, 401, 425, 1585, A3. *See also* Stevens-Johnson syndrome (SJS); Toxic epidermal necrolysis (TEN)
  - drug interactions of, 1402
  - for HIV infection, 1617t
  - resistance to, 1622f
  - structure of, 1619f
- Nevoid basal cell carcinoma syndrome. *See* Gorlin's syndrome
- Nevomelanocytic nevus, 377f
- Nevus
  - atypical, melanoma and, 600
  - dysplastic, 600, A5
  - nevomelanocytic, 377f, A5
  - oral, 264t
- Nevus depigmentosus, 397
- Newborn(s). *See* Neonates
- New daily persistent headache. *See* Headache, new daily persistent
- New Delhi metallo-β-lactamases (NDM carbapenemases), 1181
- New York Heart Association (NYHA), classification of cardiovascular disease, 1841–1842, 1842t, 1982, 1982t, 2093t
- New York virus, 1661t, 1674
- *NEXN* gene mutations, 2004t
- Next-generation sequencing, 801, 976t, 3797t, 3801, 3810
- NF1. *See* Neurofibromatosis type 1 (NF1)
- *NF1* gene mutations
  - actions of, 722, 3088
  - in CNS tumors, 717t, 722
  - in juvenile myelomonocytic leukemia, 879
  - in melanoma, 601, 602f

NF1 gene mutations (*Cont.*):
  NETs and, 677
  in neurofibromatosis 1, 3077
  in ovarian cancer, 710
  in pheochromocytoma/paraganglioma, 3077–3078, 3079t
  in soft tissue sarcomas, 727
  tumors associated with, 524t
NF2. *See* Neurofibromatosis type 2 (NF2)
NF2 gene mutations, 524t, 717t, 723
NFE2L2 gene mutations, 661t
NFTs. *See* Neurofibrillary tangles (NFTs)
NF-κB. *See* Nuclear factor-κB (NF-κB)
NF-κB essential modulator. *See* NEMO (NF-κB essential modulator)
NGAL (neutrophil gelatinase associated lipocalin), 2382
Ngari virus, 1659t
NHLRC1 gene mutations, 3516
NHP2 gene mutations, 3820t
Niacin (vitamin B_3), **2608**
  adverse effects of, 3648t
  deficiency of. *See* Pellagra
  functions of, 2606f, 2608
  for metabolic syndrome, 3258
  metabolism of, 2608
  recommended intake of, 2600t
  structure of, 2606f
  toxicity of, 2608
Nicardipine
  adverse effects of, 2101t, 2940t
  for hyperperfusion disorders, 2350
  for hypertensive emergency, 2148
  for ischemic heart disease, 2101, 2101t
Niche site, stem cells, 759
Nickel allergy, 381f, 384, A5
Nickel exposure, 512t, 610, 3702
Niclosamide, 1744t, 1750
Nicoladoni-Branham sign, 2176
Nicorandil, 2103, 2112
Nicotine, 3656t, 3681–3683, 3712t. *See also* Smoking
Nicotine aerosols, 3681, 3683, 3684
Nicotine replacement therapy, 615, 2255, 3685, 3685t
Nicotinic acid. *See* Niacin (vitamin B_3)
Nicotinic agonists, 3712t
NIDDM (non–insulin-dependent diabetes mellitus). *See* Diabetes mellitus (DM), type 2
Niemann-Pick diseases, 3353t, 3356
Nifedipine
  adverse effects of, 2101t, 2144t
  for chilblain, 3772
  dosage in liver disease, 488
  drug interactions of, 1745t
  for ergot alkaloid overdosage/poisoning, 3710t
  for high-altitude illness, 3756t, 3758, 3760
  for high-altitude pulmonary edema, 2331
  for hypertension, 2144t
  for hypertension in pregnancy, 2140, 3899
  for ischemic heart disease, 2101, 2101t
  ointment, for anal fissure, 2586
  overdosage/poisoning with, 3712t
  for Raynaud's phenomenon, 2178, 2874
Nifurtimox
  actions of, 1750
  adverse effects of, 1744t, 1750, 1791t, 2013
  for Chagas disease, 1609, 1744t, 1791–1792, 1791t, 1796t, 2013
  for human African trypanosomiasis, 1795
  pharmacology of, 1750
Night blindness, 2549t, 2605t, 2611
Night-eating syndrome, 3951
Nightshade, 3711t
Nijmegen breakage syndrome, 2800, S9
Nikolsky's sign, 141t, 146, 409
Nil lesion. *See* Minimal change disease
Nilotinib
  actions and targets of, 533t, 535, 567, 568t, 838
  adverse effects of, 568t, 753t, 754, 839t, 840–841
  for ALL, 848, 848t
  for CML, 838–841, 839t
  genetic variations in response to, 567, E72t
  interactions and issues, 568t
Nilutamide, 701, 753t
Nimesulide, 3137
nim genes, 1185
Nimodipine, 3474
Nintedanib, 754, 2230, 2262, 2265, 2874
Nipah virus, 1112, 1539
Nipa virus, 3856
Nippostrongylus brasiliensis infection, 2769
Niraparib, 533t, 571t, 712
Niridazole, 799t
Nirmatrelvir, 996, 1156
Nirmatrelvir/ritonavir, 1488t, 1489, 1542
Nirogacestat, 571t, 573, 729
Nirsevimab-alip, 1489
NIS (sodium iodide symporter), 2790t, 3022, 3028t
NIS gene mutations, 3022
Nisoldipine, 2101t
Nissen fundoplication, 2511
Nit(s), 3732
Nitazenes, 3696
Nitazoxanide
  actions of, 1750
  adverse effects of, 1744t
  for *Cryptosporidium* infections, 1600, 1806
  for *Cyclospora* infections, 1807
  for *Giardia* infections, 1805
  in *H. pylori* eradication, 2527t
  pharmacology of, 1750
  in pregnancy and lactation, 1744t
  for *R. nana* infections, 1838
Nitisinone, 3371
Nitrates
  actions of, 2100–2101
  adverse effects of, 2102t, 2118
  contraindications to, 2102t, 2109t
  dietary, 645
  drug interactions of, 2102t
  for heart failure, 1988, 1994, 2123
  interaction with PDE-5 inhibitors, 3159
  interaction with phosphodiesterase type 5 inhibitors, 2102t, 2109, 2118
  for ischemic heart disease, 2100–2101, 2100t, 2102t
  long-acting, 2101
  for NSTE-ACS, 2109, 2109t
  overdosage/poisoning with, 3714t
  for pulmonary edema, 2331
  for STEMI, 2118
  tolerance to, 2101
  for vasospastic angina, 2112
Nitric oxide
  for arterial hypoxemia, 2307
  in asthma, 2219
  in blood vessels, 1845, 1846
  in cirrhosis, 328
  in clitoral engorgement, 3161
  in erection, 3155, 3156f
  in gastric mucosal integrity, 2517
  for high-altitude pulmonary edema, 2331
  in pulmonary hypertension, 2194
  in swallowing, 293
Nitric oxide synthase. *See* Endothelial nitric oxide synthase (eNOS)
Nitric oxide therapy, 2303, 2303t
Nitrites, 3714t
Nitroblue tetrazolium dye test, 463
Nitrofurantoin
  actions of, 1167, 1178, 1182t, 1183f
  adverse effects of, 1171t, 1178
    common, 1093t
    eosinophilia, 462
    folate deficiency, 787t
    in G6PD deficiency, 799t
    hepatotoxicity, 2667, 2670–2671
    lupus syndrome, 2835
    neuropathy, 3609t
    pulmonary, 1093, 2261
  for cystitis, 1093, 1093t
  for cystitis prophylaxis, 1095
  for *E. coli* infections, 1283
  indications for, 1174t, 1178
  in pregnancy and lactation, 1169t
  prophylactic uses of, 1180t
  resistance to, 1087, 1093, 1174t, 1178, 1182t, 1185, 1291
  for UTIs, 1219t, 1220t, 1287
Nitrogen dioxide exposure, 2239t
Nitrogen mustard (mechlorethamine), 512t, 726t, 3115, S5
Nitroglycerin
  adverse effects of, 2101, 2118
  for cocaine intoxication, 3695
  for ergot alkaloid overdosage/poisoning, 3710t
  for heart failure, 1991, 1993t, 2122
  for hypothermia, 3771
  intravenous, 2122
  for ischemic heart disease, 2100–2101, 2100t
  metabolism of, 485
  for pulmonary edema, 2331
  for Raynaud's phenomenon, 2874
  for STEMI, 2118, 2122
Nitroimidazole, 2527t, 2530t
Nitroprusside
  for aortic dissection, 2170
  for cocaine intoxication, 3695
  for ergot alkaloid overdosage/poisoning, 3710t
  for heart failure, 1991, 1993t
  for hypertensive emergency, 2148
  for MAOI overdosage/poisoning, 3710t
  for pulmonary edema, 2331
Nitroprusside ketone reaction, 370
Nitrosoureas, 563, 574, 595
Nitrous oxide
  as greenhouse gas, 3745f, 3746t
  inhalation of, cobalamin deficiency in, 787, 3568
NIV. *See* Noninvasive ventilation (NIV)
Nivolumab
  action and targets of, 534t, 559f, 560, 2787t
  adverse effects of, 560
    acute interstitial nephritis, 2377
    autoimmune hemolytic anemia, 802
    hepatotoxicity, 2672
    long-term, 753t
    neurotoxicity, 726, 2351
    pneumonitis, 596, 596f
  for bladder cancer, 694
  for esophageal cancer, 645
  for gastric cancer, 649
  for glioblastoma, 719
  for HCC, 665f, 666
  for head and neck cancer, 614
  for high-grade NETs, 686
  for Hodgkin's lymphoma, 870
  indications for, 2787t
  for lung cancer, 625, 625t, 629, 630t
  for melanoma, 603, 604, 604t, 605, 2783
  for pancreatic cancer, 675
  for RCC, 689, 690t
  resistance to, 574

- Nixon's method, splenic percussion, 480
- Nizatidine, 2524t, 2525
- NK$_1$ (neurokinin), 297
- NK$_1$ (neurokinin) receptor antagonists, 274, 298, 299t, 508–509
- NK cells. *See* Natural killer (NK) cells
- NK-T cells. *See* Natural killer T (NK-T) cells
- *NKX2-1,* 3028t, 3514
- NKX2-5, 3028t
- *NLRC4* gene mutations, 466, 2761t, 2937
- NLRP1 inflammasome, 972
- *NLRP3* gene mutations, 2934, 2936. *See also* Cryopyrin-associated periodic syndromes (CAPS)
- NLRP3 inflammasome, 972, 2936
- *NLRP* gene mutations, 2760, 2761–2762t
- NLRs (NOD-like receptors), 2758, 2759t, 2760
- NLST (National Lung Cancer Screening Trial), 617–618, 618t
- NMDA receptor antibodies, 743t, 745f, 746–747, 1114, 2827, 3484
- NMO. *See* Neuromyelitis optica (NMO)
- NMR (nuclear magnetic resonance), 3976, 3976t
- NMS (neuroleptic malignant syndrome), 3517
- *NMYC* gene mutations, 521
- NNT (number needed to treat), 28
- *Nocardia* spp., 1358, 1359f
- *Nocardia* spp. infections, **1358**
  - actinomycetoma, 1360, 1361f, 1362, 1362t
  - brain abscess, 583
  - in cancer patient, 578, 582, 583
  - chronic meningitis, 1130t
  - clinical features of, 405, 1359–1360, 1359–1361f
  - diagnosis of, 961f, 1360–1361, S11
  - disseminated, 1361
  - epidemiology of, 1358
  - global considerations in, 1358
  - pathogenesis of, 1358–1359
  - pathology of, 1359
  - prevention of, 1362
  - in solid organ transplant recipient, 1162f, 1164
  - in transplant recipient, 2350
  - treatment of, 1362
- Nocebo effects, 95, 3930t, **3934**
- Nociceptin/orphanin receptors, 3689t
- Nociceptive pain, 80, 80f, 118, 118t, 119, 506
- Nociceptor(s), 93f, 94–95, 94f
- Nociceptor-induced inflammation, 94, 94f
- Nociplastic pain, 118, 118t, 124t
- Nocturia, 1983, 2455
- *NOD1/2* gene mutations, 971
- *NOD2* gene mutations, 2554, 2554t, 2937
- Nodal marginal zone B cell lymphoma, 865, 873
- NOD-like receptors (NLRs), 2758, 2759t, 2760
- Nodopathies, 3623
- Nodular amyloidosis, 404
- Nodular vasculitis, 405
- Nodule
  - *Aspergillus,* 1715
  - cutaneous, 137, 377t, 378f. *See also* Papulonodular lesions
  - thyroid. *See* Thyroid nodular disease
- Noise exposure, 251–252, 251–252t
- Noma (cancrum oris), 261, 1374
- NOMID (neonatal-onset multisystem inflammatory disease), 2935t, 2936
- Non-24-hour sleep-wake rhythm disorder, 218, 3944t, 3948–3949
- Nonaka distal myopathy, 3640t, 3643–3644
- Nonalcoholic steatohepatitis (NASH). *See* Metabolic dysfunction–associated steatohepatitis (MASH)
- Nonallelic (locus) heterogeneity, 3787–3788, 3788t, 3801
- Nonbacterial thrombotic endocarditis (NBTE), 1038
- Nonchromogens, 1426
- Nonenveloped viruses, 1478f
- Nonhistone proteins, 3779f
- Non-Hodgkin's lymphoma, **857**. *See also* Lymphoid malignancies
  - anti-TNF associated, 2566
  - approach to the patient, 860–862
  - diagnosis of, 860–861
  - epidemiology of, 857–858, 858f
  - etiology of, 857–858, 858t
  - extranodal, 649–650
  - genetic considerations in, 521t, 858–860, 860t
  - in HIV infection, 1585
  - immunology of, 858–860, 859f, 860f
  - primary central nervous system, 1613, 1614f
  - prognosis of, 861–862, 861t
  - renal involvement in, 2440
  - skin manifestations of, 405, 411, A5
  - splenomegaly in, 480
  - staging of, 861, 861t
  - subtypes
    - adult T-cell leukemia/lymphoma. *See* Adult T-cell leukemia/lymphoma
    - anaplastic large-cell lymphoma, 867
    - angioimmunoblastic T cell lymphoma, 860t, 867, 875
    - Burkitt's lymphoma. *See* Burkitt's lymphoma
    - chronic lymphocytic leukemia. *See* Chronic lymphocytic leukemia (CLL)
    - DLBCL. *See* Diffuse large B-cell lymphoma (DLBCL)
    - enteropathy-type T-cell lymphoma, 867, 874
    - extranodal T/NK cell lymphoma, nasal type, 857, 867–868, **874,** 2898
    - follicular lymphoma. *See* Follicular lymphoma
    - frequency of, 858f
    - hairy cell leukemia. *See* Hairy cell leukemia
    - hepatosplenic T-cell lymphoma, 860t, 867, 874–875
    - lymphomatoid granulomatosis, 2898
    - lymphoplasmacytic lymphoma, 865
    - MALT lymphoma. *See* Mucosa-associated lymphoid tissue (MALT) lymphoma
    - mantle cell lymphoma. *See* Mantle cell lymphoma
    - marginal zone. *See* Marginal zone lymphoma
    - mycosis fungoides. *See* Mycosis fungoides
    - PCNSL. *See* Primary central nervous system lymphoma (PCNSL)
    - peripheral T-cell lymphoma, not otherwise specified, 860t, 862, 866–867
    - primary effusion lymphoma, 1516, 1613
    - subcutaneous panniculitis-like T-cell lymphoma, 867, 875
  - testicular non-GCT in, 709
  - treatment of, 919t, 920
- Nonhuman primates, bite-wounds from, 1143, 1145t
- Non–insulin-dependent diabetes mellitus (NIDDM). *See* Diabetes mellitus, type 2
- Noninvasive brain stimulation, 3966–3967
- Noninvasive ventilation (NIV)
  - contraindications for, 2259, 2306t
  - for COPD, 2259, 2306
  - for hypoventilation, 2275–2276
  - indications for, 2306
  - for pulmonary edema, 2330
- Nonlinear probability weighting, 3916t
- Nonocclusive mesenteric ischemia, 2588t, 2589
- Nonoliguria, 341
- Nonoxynol-9, 1088, 1260
- Nonreactive miliary tuberculosis, 1390
- Nonreplicating viral vector vaccines, 997t
- Nonseminomatous germ cell tumor, 706, 707–708f, 708t. *See also* Testicular cancer
- Nonsense-mediated decay, 522
- Nonsense mutation, 3786f
- Non-small-cell lung cancer. *See* Lung cancer, non-small-cell
- Non-specific interstitial pneumonia (NSIP), 2260t, 2263, 2263f, 2264f
- Nonsteroidal anti-inflammatory drugs (NSAIDs)
  - actions of, 2517
  - for acute otitis media, 254
  - adverse effects of
    - bleeding, 473
    - in cirrhosis with ascites, 2715
    - colitis, vs. IBD, 2561
    - cross-sensitivity and, 427–428
    - cutaneous, 399, 417, 420, 422
    - desensitization protocol for, 2819
    - diarrhea, 306
    - edema, 282t
    - fixed drug eruption, 423
    - gastrointestinal bleeding, 311, 513, 2955, 2955t
    - gastrointestinal ulcers, 300. *See also* Peptic ulcer disease, NSAID-induced
    - hyperkalemia, 359
    - hypertension, 2137
    - immunologic factors in, 419
    - in ischemic heart disease, 2103
    - in older adults, 3883
    - photosensitivity, 420
    - proximal ileal stricture, 2472f
    - pulmonary eosinophilia, 2234
    - renal, 96, 340, 2373, 2374f, 2397, 2438, 2955
    - SJS/TEN, 425. *See also* Stevens-Johnson syndrome (SJS); Toxic epidermal necrolysis (TEN)
    - steatosis, 2703t
    - in STEMI, 2119
  - for axial spondyloarthritis, 2887t
  - for calcium apatite deposition disease, 2962
  - for calcium pyrophosphate deposition disease, 2961
  - for colorectal cancer prevention, 655
  - discontinuation before lumbar puncture, S3
  - drug interactions of, 489t, 562t
  - for dysmenorrhea, 3137
  - for episcleritis and scleritis, 224
  - for fever, 136
  - for fever of unknown origin, 155
  - gastric protection with, 2530, 2530t
  - for gout, 2959
  - for low back pain, 126
  - for menorrhagia in perimenopause, 3142
  - for migraine, 3466, 3467t
  - for neck pain, 132
  - for nephrogenic diabetes insipidus, 353
  - for osteoarthritis, 2955, 2955t
  - for pain, 80, 96–98, 97t
  - for phototoxicity, 434
  - for polyposis syndromes, 654
  - for rheumatic fever, 2857
  - for rheumatoid arthritis, 2849
  - for subacute thyroiditis, 3038
  - for sunburn, 430
  - for tension-type headache, 3471
  - topical, 2955, 2955t, 3883
- Non-ST-segment elevation acute coronary syndrome (NSTE-ACS), **2106**
  - cardiac biomarkers in, 2108, 2109t
  - cardiogenic shock in, 2328f. *See also* Cardiogenic shock
  - chest pain in, 103t, 2106
  - clinical features of, 2106–2108
  - diagnosis of, 2106–2108, 2107f, 2107t, 2108f, 2114f

- Non-ST-segment elevation acute coronary syndrome (NSTE-ACS) (*Cont.*):
  - ECG in, 2108
  - global considerations in, 2112–2113
  - incidence of, 2106
  - pathophysiology of, 2106, 2107f
  - pulmonary edema in, 2331
  - risk stratification in, 2107t, 2108
  - treatment of, **2109**
    - ACE inhibitors, 2110
    - antithrombotic therapy, 943, 2110–2111, 2110t
    - beta blockers, 2109–2110, 2109t
    - calcium channel blockers, 2109t, 2110
    - invasive vs. conservative strategy, 2111, 2111f
    - long-term management, 2112
    - monitoring, 2109
    - morphine, 2109t, 2110
    - nitrates, 1909, 2109t
    - PCI, 2111, 2111f. *See also* Percutaneous coronary interventions (PCI)
    - statins, 2110
- Nonsustained ventricular tachycardia (NSVT). *See* Ventricular tachycardia (VT), nonsustained
- Nonthyroidal illness. *See* Sick euthyroid syndrome
- Nontoxic multinodular goiter, 3041–3042. *See also* Thyroid nodular disease
- Nontropical sprue. *See* Celiac disease (gluten-induced enteropathy)
- Nontuberculous mycobacterial (NTM) infections, **1414**
  - in children, 1416
  - clinical features of
    - bronchiectasis, 1416, 2242, 2243. *See also* Bronchiectasis
    - cervical lymph nodes, 1416
    - disseminated disease, 1415–1416
    - osteomyelitis, 1058, 1062
    - pulmonary, 1416
    - skin and soft tissue, 1416
  - diagnosis of, 1416–1417, 1417f, 1426
  - epidemiology of, 1414–1415, 1426
  - global considerations in, 1414, 1418
  - *M. abscessus. See Mycobacterium abscessus* infections
  - *M. avium* complex. *See Mycobacterium avium* complex (MAC) infections
  - *M. chelonae,* 1428
  - *M. chimaera,* 978, 985
  - *M. fortuitum,* 1414, 1415, 1416, 1427
  - microbiology of, 1426
  - *M. kansasii. See Mycobacterium kansasii* infections
  - *M. marinum. See Mycobacterium marinum* infections
  - *M. ulcerans,* 1051, 1414, 1416
  - pathobiology of, 1415, 1415f
  - prevention of, 1418
  - prognosis of, 1418
  - in solid organ transplant recipient, 1163–1164
  - treatment of, 1417–1418, 1426–1428
- Nontyphoidal salmonellosis. *See Salmonella* spp. infections
- Noonan's syndrome, 2063, 2183, 3788t
- *NOP10* gene mutations, 3820t
- Norepinephrine
  - actions of, 3538
  - in cardiac contraction, 1849
  - for cardiogenic shock, 2327–2328
  - for hepatorenal syndrome, 2383
  - for sepsis/septic shock, 2322–2323, 2322t
  - for shock, 2315
  - in vascular smooth-muscle cells, 1846
- Norethindrone, 3142, 3148
- Norfloxacin
  - actions of, 1166, 1182t
  - adverse effects of, 799t
  - for donovanosis, 1358
  - for infectious diarrhea, 1081t
  - for peritonitis, 330
  - for peritonitis prophylaxis, 1071
  - resistance to, 1182t, 1185
  - for *Shigella* infections, 1324
- Norgestimate, 3141
- Normal-pressure hydrocephalus, 178, 197, 198f, 3483
- Normal saline. *See* Saline
- Normeperidine, 98
- Normochromia, 437
- Noroviruses, 1628, 1628f
- Norovirus infection
  - climate change and, 3752t
  - clinical features of, 305, 306t, 1629, 1629t
  - diagnosis of, 1629
  - epidemiology of, 305, 1079, 1079t, 1628, 3855
  - etiology of, 1628, 1628t
  - health care–associated, 1150
  - immunity to, 1629
  - pathogenesis of, 1023–1629
  - prevention of, 1629–1630
  - treatment of, 1629
- Norrie's disease, E36t
- North Queensland tick typhus, 139t
- Nortriptyline
  - adverse effects of, 3659t
  - for chronic daily headache, 115
  - for depression, 3588, 3659t, 3666, 3667
  - genetic variations in response to, 492
  - for migraine prevention, 3469, 3470t
  - for nausea and vomiting, 299t
  - for neuropathic pain, 82, 99, 101, 506
  - for pain, 97t, 99, 3588
  - for smoking cessation therapy, 615
- Norwalk virus infection. *See* Norovirus infection
- Norwegian (crusted) scabies, 1605, 3729
- Nose
  - in granulomatosis with polyangiitis, 1860, 2897
  - in lepromatous leprosy, 1408
  - in relapsing polychondritis, 1860, 2918, 2919f
  - swab, specimen collection and transport, S11
- No-see-um midges, 3736
- *Nosema* spp., 1808
- Nosocomial infections. *See* Health care–associated infections
- NO SPECS mnemonic, in Graves' disease, 3034
- *NOTCH1* gene mutations, 431, 850, 2036
- *NOTCH2* gene mutations, 872
- *NOTCH4* gene mutations, 2861
- Notch pathway, 526t, 573
- NOTES (natural orifice transluminal endoscopic) surgery, 2469
- Nothnagel's syndrome, 234
- Novichok nerve agents, S5
- NOX2, 453, 460
- *NPAS2* genes, 3943, 3947t, 3951
- *NPC1/2* gene mutations, 3356
- NPH insulin, 3209t, 3210
- *NPHP2* gene mutations, 2435
- *NPHS1/2* gene mutations, 2357, 2411
- *NPM1* gene mutations, 826, 827, 827t
- *NPRL3,* 771
- *NR0B1* gene mutations, 3112t
- *NR1H4* gene mutations, 2642
- *NR5A1* gene mutations, 2992, 3099t, 3101t
- NRAMP1, 1384
- *NRAS* gene mutations
  - cancers associated with, 520t
  - as drug target, 572
  - in lung cancer, 617f, 617t
  - in melanoma, 432, 601, 602f, 605
  - in multiple myeloma, 784, 884
  - in nonmelanoma skin cancer, 606
- NREM sleep, 210–211, 210f
- NRF2, 2251, 3537
- *NRG1* gene, 572
- *NRL* gene mutations, 3783t
- *NRTK* gene mutations, 572
- *NRXN1* gene mutations, 3652
- NSAIDs. *See* Nonsteroidal anti-inflammatory drugs (NSAIDs)
- NSE. *See* Neuron-specific enolase (NSE)
- NSIP (non-specific interstitial pneumonia), 2260t, 2263, 2263f, 2264f
- *NSMF* gene mutations, 3112t
- NSTE-ACS. *See* Non-ST-segment elevation acute coronary syndrome (NSTE-ACS)
- NSVT (nonsustained ventricular tachycardia). *See* Ventricular tachycardia (VT), nonsustained
- Ntaya virus, 1660t
- N-terminal pro-BNP (NT-proBNP), 898, 1985, 2190, 2268
- NTM infections. *See* Nontuberculous mycobacterial (NTM) infections
- *NTRK* gene mutations, 617, 617f, 617t, 628, 628f
- NTS (nucleus of the solitary tract), 239
- Nuchal rigidity, 1119–1120
- Nuclear factor-κB (NF-κB)
  - activation defects of, 459t, 2799
  - essential modulator. *See* NEMO (NF-κB essential modulator)
  - inhibition of, 536, 540f, 971
  - receptor activator of. *See* RANK (receptor activator of NFκB)
  - systems biology analysis of, 3957
- Nuclear magnetic resonance (NMR), 3976, 3976t
- Nuclear membrane, inherited defects of, 2004t, 2009
- Nuclear-mitochondrial genomic cross-talk, 3814, 3814f
- Nuclear receptor(s), 8–9, 2981, 2983, 2985f, 3783t
- Nuclear weapons, S6, S8
- Nucleic acid tests
  - amplification, 976, 976t, 1483, 1590, 1590t, S11
  - applications, S11
  - next-generation and metagenomic, S11
  - probes for direct detection of pathogens, S11
  - quantitative strategies, S11
- Nucleocapsid, viral, 1477, 1478f
- Nucleoside reverse transcriptase inhibitors. *See* Combination antiretroviral therapy (cART)
- Nucleosome, 3779f, 3784f, 3935
- Nucleosome remodeling enzymes, 3937
- 5′-Nucleotidase, 2635
- Nucleotide repeat expansion disorders, 3792–3793, 3793t
- Nucleotide reverse transcriptase inhibitors. *See* Combination antiretroviral therapy (cART)
- Nucleus of the solitary tract (NTS), 239, 239f
- Nucleus pulposus, herniated, 111, 3565
- Nudge, in behavioral economics, 3915–3916, 3918
- *NUDT15,* 493
- Number needed to treat (NNT), 28
- Numbness, 172. *See also* Sensation
- Nummular eczema, 385
- Nummular headache, 3474–3475
- Nummular skin lesions, 378t
- Nurse-like cell, in CLL, 852
- Nusinersen, 3524
- Nutrient(s), 2459, **2541,** 2598. *See also specific nutrients*
- Nutrient requirements, **2598**
  - in critically ill patient, 2297–2298
  - factors altering, 2599, 2602–2603
  - recommended intakes, **2599,** 2600–2602t

- Nutrient-sensing pathways, 3871, 3873f
- Nutritional monitoring, 2603
- Nutritional physiology, 2619–2620
- Nutritional Screening Initiative, 316
- Nutritional short stature, 2995
- Nutritional supplements, 2599
- Nutritional support/therapy. *See also* Enteral nutrition; Parenteral nutrition
  - for acute pancreatitis, 2749
  - after bariatric surgery, 3195
  - for dementia, 2626
  - for diabetes mellitus, 3206, 3206t
  - for gastrointestinal disease, 2463
  - for IBS, 2574, 2576t
  - for metabolic syndrome, 3257
  - for obesity, 3190
  - optimized voluntary, 2622
  - for STEMI, 2121
- Nutrition assessment, **2616**
  - in acute care setting, 2603
  - in ambulatory setting, 2603–2604, 2616
  - components of
    - anthropometric data, 2617, 2618t
    - clinical signs and physical examination, 2617
    - diet history, 2617
    - functional outcomes, 2617, 2619t
    - laboratory evaluation, 2617, 2618t
    - medical/surgical history and clinical diagnosis, 2616–2617
  - global considerations in, 2604
  - in high-risk patient, 2604
  - instruments for, 2622
  - in protein-calorie malnutrition, 2616
- NVICP (National Vaccine Injury Compensation Program), 1001
- NY-014 virus, 1518t
- Nyando virus, 1659t
- NYHA (New York Heart Association), classification of cardiovascular disease, 1842t, 1982t, 2093t
- Nylon flock exposure, 2240
- Nystagmus, **236**
  - congenital, 236, V3
  - downbeat, 237
  - etiology of, 236
  - gaze-evoked, 236–237
  - jerk, 236
  - patterns of, 162–163, 163t
  - rotary, in stimulant overdose, 3705
  - seesaw, V3
  - upbeat, 237
  - vestibular, 237
  - in vestibular dysfunction, 180
- Nystatin, 388t, 390, 1692, 1710t, 2410

O

- OA. *See* Osteoarthritis (OA)
- Oak leaf gall mite, 3730
- OATP1B1 (organic anion transport protein 1B1), 2637f, 2638, 2642
- Obesity, **3187**
  - central adiposity in, 3252
  - comorbid conditions, 3188–3189, 3189t
    - asthma, 2221
    - bone and joint disorders, 2952
    - diabetes mellitus, 3196–3197, 3201–3202
    - dyslipidemia, 3186, 3246, 3249
    - erectile dysfunction, 3155
    - PCOS, 3135
  - complications of, **3186,** 3186f
    - cancer, 512, 544–545, 655, 687, 3187
    - cardiovascular disease, 1858, 2018, 2099, 3187
    - cutaneous disorders, 3187
    - diabetes mellitus, 3187
    - endocrine disorders, 3187
    - gastrointestinal disorders, 3187
    - hypogonadotropic hypogonadism, 3114
    - impaired response to infection, 3187
    - insulin resistance, 3186, 3186f
    - liver disease, 3187
    - MASLD, 2702
    - mechanistic considerations in, 3186, 3186f
    - neurologic disease, 3187
    - reproductive disorders, 3187
    - rheumatologic disorders, 3187
    - sleep deprivation, 3188
  - definition of, 3181, 3181f
  - dyspnea in, 270
  - emergency contraception and, 3154
  - epidemiology of, 3181
  - evaluation of, 3188–3189, 3188t
  - fatigue and, 166
  - global burden of, 1858, 3851
  - global considerations in, 3181
  - heart failure and, 1986
  - high-altitude travel and, 3759
  - history focused on, 3188
  - hyperpigmentation in, 397
  - in hypothalamic disorders, 3002
  - infertility and, 3150–3151
  - in metabolic syndrome, 3252, 3253f, 3253t
  - pancreatic cancer risk and, 671–672
  - pathobiology of, **3181**
    - energy intake and expenditure, 3183
    - environmental factors, 3183–3184
    - genetic considerations in, 3184–3185, 3184t, 3185t
    - genetic vs. environmental factors in, 3183
    - gut microbiome in, 3835
    - homeostatic regulation, 3181–3182, 3182f, 3183f
    - leptin in, 3184, 3185, 3185t
    - nutrient storage in adipose tissue, 3182
  - pregnancy and, 3901
  - prevalence of, 12, 2979t, 3181
  - pulmonary function values in, 2205, 2206f
  - screening/testing for, 39t, 2979t
  - secondary to other disorders, 3185
  - severity staging in, 3188, 3189f
  - treatment of, **3190**
    - algorithm for, 3189f
    - bariatric surgery, 3194–3195, 3194f
    - behavioral therapy, 3191
    - diet therapy, 3190
    - endoscopic, 2493, 3195, V5
    - goal of, 3190
    - incentive programs for, 3919–3920, 3919f, 3920f
    - in metabolic syndrome, 3257
    - pharmacotherapy, 3191–3194, 3192t
    - physical activity, 3190–3191
    - readiness to change and, 3189
    - selection of, 3189f, 3190
  - in women vs. men, 3166
- Obesity hypoventilation syndrome, 2276
- Obiltoxaximab, S4
- Obinutuzumab
  - actions of, 558, 559f, 2791t
  - adverse effects of, 593
  - for CLL, 855
  - for follicular lymphoma, 864
  - indications for, 2791t
  - for SLE, 2835
- Object recognition, 206–207
- Obligate anaerobes, 1371
- Observational data, 27
- Obsessive-compulsive disorder (OCD), **3663**
- Obstructive shock, 2311, 2311t
- Obstructive sleep apnea/hypopnea syndrome (OSAHS), **2277**
  - approach to the patient, 2278
  - comorbidities of, 2280–2281
  - daytime sleepiness in, 213, 2279, 2281
  - definition of, 2277
  - diagnosis of, 213t, 2275, 2279, 2279t, 2280f
  - epidemiology of, 12, 2138t, 2278
  - fatigue and, 166
  - in heart failure, 1997–1998
  - heart failure and, 1986
  - hypertension and, 2136–2137, 2138t, 2280–2281
  - MASLD and, 2702
  - in metabolic syndrome, 3256
  - obesity and, 2278, 3187
  - in older adults, 3893
  - pathophysiology of, 2138t, 2206, 2277–2278, 2278f, 2280f
  - polysomnography in, 2279, 2280f
  - pulmonary hypertension in, 2190
  - quality of life in, 2281
  - risk factors for, 2278
  - screening for, 12
  - severity scale for, 2280t
  - sudden cardiac death and, 2333
  - treatment of, 213t, 2138t, 2281–2282, 2281t
- Obturator sign, 2595t
- Occam's razor, 3957
- Occipital adenopathy, 476
- Occipital lobe stroke, 230
- Occipital nerve injection, 3472, 3472t
- Occipitotemporal network, 206–207
- Occupational illness. *See also* Health care workers
  - animal bite-wounds, 1143
  - asthma, 2221
  - contact dermatitis, 384
  - hand eczema, 384
  - hearing loss, 251–252, 251–252t
  - lung disease, 2200, **2234**
    - asbestos-related. *See* Asbestos-related disease
    - categories of, 2236t
    - characteristics of agents causing, 2235
    - chest imaging in, 2235
    - chronic beryllium disease, 2238
    - coal worker's pneumoconiosis, 2238
    - COPD, 2253
    - diagnosis of, 2235
    - disability assessment in, 2240
    - exposure assessment in, 2235
    - history in, 2235
    - hypersensitivity pneumonitis. *See* Hypersensitivity pneumonitis
    - incidence of, 2234
    - inorganic dusts, 2236t, 2238
    - in military veterans, 2240. *See also* Military veterans
    - organic dusts, 2236t, 2238–2239
    - pulmonary function tests in, 2235
    - silicosis, 2235–2238, 2237f
    - toxic chemicals, 2239–2240, 2239t
    - in World Trade Center rescuers, 42
  - osteoarthritis, 2952
  - shift-work disorder, 219
  - tularemia, 1337
- OCD (obsessive-compulsive disorder), **3663**
- OCG (oral cholecystography), 2730
- *Ochlerotatus* mosquitoes, 1672–1673
- *Ochrobactrum* spp., 1033, 1269, 1331
- Ochronosis, 3371
- Ochronotic arthritis, 3371
- Ocrelizumab, 3585, 3586t, 3587
- OCT (optical coherence tomography), 1908–1909, 1909f
- Octopus envenomation, 3724–3725
- Octreotide
  - for acromegaly, 3009
  - adverse effects of, 3009
  - for carcinoid syndrome, 312, 652, 682

- Octreotide (*Cont.*):
  - for chemotherapy-related diarrhea, 564, 576
  - for dumping syndrome, 2533
  - in end-of-life care, 82
  - for gastrinoma, 3082
  - for hepatorenal syndrome, 2383
  - for hypoglycemia, 3236
  - for hypomotility of small bowel, 2874
  - for insulinoma, 3235
  - for intestinal pseudoobstruction, 299, 299t
  - for Ménétrier's disease, 2540
  - for NETs, 684, 684t
  - for oncogenic osteomalacia, 740
  - for orthostatic hypotension, 160
  - for refractory diarrhea, 508
  - structure, 683f
  - for variceal bleeding, 317, 2714
  - for vascular ectasias, 318
  - for ZES, 2536
- Ocular alignment, 221
- "Ocular bobbing," 190
- Ocular disorders, **220**. *See also* Optic neuritis; Visual loss
  - acne rosacea and, 391
  - in Behçet syndrome, 225, 2909
  - in cat-scratch disease, 1352f, 1353, 1353t
  - in diabetes mellitus. *See* Diabetes mellitus (DM), complications of; Diabetic retinopathy
  - in granulomatosis with polyangiitis, 2897
  - at high altitudes, 3758–3759
  - in IBD, 2563
  - infections. *See also* Conjunctivitis; Keratitis; Retinitis
    - *Candida,* 1709, 1709f, 1712
    - *L. loa. See Loa loa* (African eye worm)/loiasis
    - *Nocardia,* 1360, 1362t
    - *Onchocerca,* 1823
    - *P. aeruginosa,* 1309t, 1311
    - parasitic, 1739t
    - syphilis, 1431, 1433, 1434t
    - *T. gondii,* 1799, 1801, 1802
    - *Toxocara,* 1812
  - in Marfan syndrome, 3332
  - in MS, 3578
  - in mucormycosis, 1718–1719
  - in myasthenia gravis, 3626
  - in osteogenesis imperfecta, 3326
  - in sarcoidosis, 2927, 3565–3566
  - in Sjögren's disease, 2877
  - trachoma, 224, 1474–1475
  - in tularemia, 1338, 1338t
  - tumors, 232, 232f
- Ocular flutter, 237, V3
- Ocular lubricants, 224
- Ocular movements, 162–163, 190–191, 190f, 221, V3
- Oculocephalic reflexes, 190, 190f, 3487
- Oculocutaneous albinism, 396
- Oculomotor apraxia, 206
- Oculomotor nerve, 3380
- Oculomotor nerve palsy, 221, 234–235, V3
- Oculopharyngeal distal myopathy, 3643
- Oculopharyngeal dystrophy, 234, 3643
- Oculovestibular response, 190, 190f
- Odontoid fracture, A16
- Odors, 50t
- Odynophagia, 293, 295, 300, 642–643, 2504, 2513
- *ODZ4* gene mutations, 3668
- Ofatumumab
  - actions of, 558, 559f, 2791t
  - adverse effects of, 593
  - for ALL, 848t
  - for MS, 3585, 3586t
- Ofloxacin
  - actions of, 1182t
  - for *Brucella* infections, 1334
  - for leprosy, 1411
  - resistance to, 1182t, 1424, 1466
  - for *Shigella* infections, 1324
- Ogilvie's disease. *See* Intestinal pseudoobstruction
- 17OHP (17-hydroxyprogesterone), 3051f, 3072
- Oil-in-water emulsion, 997t, 998
- Oka. *See* Varicella vaccine
- OKT3, 282t
- Olanzapine
  - for acute alcohol intoxication, 3679
  - adverse effects of, 3188, 3673, 3673t, 3674
  - for anorexia, 509
  - for bipolar disorder, 3668
  - for delirium, 86, 86t
  - for nausea and vomiting, 81, 299, 299t, 300, 509, 575
  - overdosage/poisoning with, 3711t
  - for schizophrenia, 3673, 3673t, 3674
- Olaparib
  - action and targets of, 533t, 571t, 574
  - adverse effects of, 571t
  - for breast cancer, 637, 638, 641
  - gene mutations and response to, 3982
  - interactions and issues, 571t
  - for ovarian cancer, 712
  - for pancreatic cancer, 676
  - for prostate cancer, 703
- Olaratumab, 534t
- Older adults. *See also* Aging
  - abuse and neglect, 71, 3894, 3895t
  - ALL in, 848
  - anemia in, 450, 769, 3883
  - antidepressant use in, 3666
  - appendicitis in, 2596
  - assessment of, 3877–3878, 3879t
  - asthma in, 2228
  - benign forgetfulness of, 194
  - cancer in, 3883
  - care transitions for, 3876, 3895, 3896t
  - cerumen impaction in, 248
  - co-managed care of, 3895–3896
  - COVID-19 disease in, 3894
  - decision-making skills in, 3877, 3880f
  - delirium in, 182, 3885–3888, 3889t. *See also* Delirium
  - dementia in, 3885–3888, 3889t. *See also* Dementia
  - demographics of, 3869, 3869f, 3876, 3886f
  - diabetes mellitus in, 3209t, 3220–3221, 3882
  - diagnostic test interpretation in, 3878
  - driving impairment in, 3877–3878, 3879t
  - drug-induced illness in, 488
  - drug management in, 3884–3885, 3887t, 3888f
  - drug use in, 488
  - end-of-life and palliative care in, 3894. *See also* End-of-life care; Palliative care
  - epilepsy in, 3428
  - falls in
    - causes of, 3886t
    - epidemiology and impact of, 180, 3884
    - evaluation of, 3879t, 3885f, 3886t
    - interventions to reduce, 181
    - management of, 3884, 3886t
    - risk factors for, 180, 180t
  - frailty in, 3893–3894
  - hearing loss in, 246, 246f
  - heart failure in, 3883–3884
  - heat-related illness in, 3773
  - hemoglobin/hematocrit in, 445t
  - hepatitis in, 2660
  - hyperlipidemia in, 3882
  - hypertension in, 2148, 3879–3880
  - hypothermia in, 3768, 3771
  - immunizations in, 1005
  - involuntary weight loss in, 315
  - levothyroxine dosage in, 3032
  - liver abscess in, 1074
  - memory impairment in, 3886, 3888f
  - metabolic syndrome in, 3253
  - multimorbidity in, 3873, 3876, 3877f
  - musculoskeletal disease in, 2941, 2949
  - olfactory dysfunction in, 240, 240f, 244
  - oral health, 267
  - osteoarthritis treatment in, 3883
  - PD in. *See* Parkinson's disease
  - person-centered care of, 3876–3877, 3878f, 3878t
  - polypharmacy in, 3884–3885, 3887t, 3888f
  - preventive care for, 3878–3879, 3881t
  - rheumatoid arthritis in, 2853
  - screening tests in, 41t
  - seizures in, 3413t, 3414
  - sleep disorders in, 3893, 3894t
  - specialized nutritional support in, 2626
  - urinary incontinence in, 3889–3891, 3890t, 3891–3893f, 3893t
  - vitamin D deficiency in, 3269
  - water requirement of, 2599
  - weight regulation in, 315
- Olecranon bursitis, 2973
- Oleoresin, 384
- Olfactory bulb, 237, 238f
- Olfactory disorders, **240**
  - after traumatic brain injury, 3573–3574
  - in COVID-19 disease, 1541
  - diseases and conditions associated with, 240–242, 241t
  - in early PD, 3495
  - epidemiology of, 240–241, 240f
  - etiology of, 240–241, 241t
  - evaluation of, 242–243
  - in gonadotropin deficiency, 3111–3112t, 3111–3113
  - in LBD, 3492
  - taste disorders and, 242
  - treatment of, 243–244
- Olfactory nerve, 3380
- Olfactory system, 237, 238f
- Oligodendrocyte precursor cells, 3397
- Oligodendrocytes, 3397
- Oligodendrogliomas, 718
- Oligomenorrhea, 3132
- Oligomer(s), 3400, 3478
- Oliguria, 341, 2362
- Olivopontocerebellar atrophy. *See* Spinocerebellar ataxia (SCA)
- Ollier's disease (enchondromatosis), 3318
- Ollier's disease (juvenile granulosa cell tumors), 712
- Olmesartan, 307, 309, 2144t
- Olodaterol, 2224, 2255
- Olorofim, 1698, 1729, 1730
- Olutasidenib, 534t, 571t, 833, 833t
- Omacetaxine, 839t, 842
- Omadacycline
  - actions of, 1166, 1177, 1182t
  - adverse effects of, 1171t, 1177
  - for *Citrobacter* infections, 1293
  - for *Enterobacter* infections, 1292
  - for enterococcal infections, 1220
  - for gram-negative bacterial infections, 1283–1284
  - indications for, 1174t, 1177
  - for *Klebsiella* spp. infections, 1291
  - for MRSA infections, 1203t, 1205
  - for NTM infections, 1418
  - in pregnancy and lactation, 1170t
  - resistance to, 1174t, 1182t, 1184, 1283, 1291
  - for *Serratia* infections, 1293

Omalizumab
  for ABPA, 2233
  actions of, 2792t
  adverse effects of, 419
  for allergic rhinitis, 2811
  for asthma, 2225
  for EGPA, 2233
  indications for, 2792t
  for solar urticaria, 434
  for urticaria, 2809
Omapatrilat, 1994–1995
Omaveloxolone, 3537
Ombitasvir, 2689
Omecamtiv mecarbil, 1995t, 1996
Omega-3 fatty acids (fish oils), 473
  adverse effects of, 3244t, 3250, 3926
  antiplatelet effects of, 473
  for heart failure, 1997
  for hypertriglyceridemia, 3244t, 3250
  for metabolic syndrome, 3258
  for rheumatoid arthritis, 3926
Omenn syndrome, 2799, 2803
Omentectomy, 711
Omeprazole
  actions of, 2524–2525
  drug interactions of, 489t
  formulations of, 2525
  for gastrinoma, 3081
  genetic variations in response to, 492, E72t
  for GERD, 301
  in *H. pylori* eradication, 1302t
  metabolism of, 485t, 492
  for peptic ulcer disease, 2524t, 2525, 2531
  for ZES, 2536
OmpA protein, 1295
Omsk hemorrhagic fever virus, 1660t, 1678
OMT (orofacial myofunctional therapy), 2282
OMVs (outer-membrane vesicles), 1252, 1252f, 1295
Onabotulinum toxin A. *See* Botulinum toxin
Onasemnogene abeparvovec, 3824t
Onchocerca volvulus/onchocerciasis
  characteristics of parasite, 1820t, 1822
  clinical features of, 225
    lymph nodes, 1823
    ocular, 1739t, 1823
    onchocercomata, 1823
    skin, 398, 1051, 1738t, 1823, 1823f
    systemic, 1823
  diagnosis of, 961t, 1823, S12
  epidemiology of, 961t, 1736, 1738t, 1820t, 1822
  etiology of, 1822
  Loa Loa infection and, 1823
  pathology of, 1822–1823
  prevention of, 1823–1824
  treatment of, 1750, 1823
Oncogene(s), 519, 520, 520t, 521t, 528, 542f. *See also* Cancer genetics
Oncogene addiction, 536, 537f, 572
Oncogene checkpoint, 530f
Oncologic emergencies. *See* Cancer, emergencies in
Oncology, 553
Oncolytic viruses, 559f, 561, 3830
Oncostatin M (OSM), 2767t
Ondansetron
  for ataxia/tremor in MS, 3588
  for nausea and vomiting, 82, 116, 298, 299t, 300, 509, 575
  for scombroid poisoning, 3726
Ondine's curse, 2276
One-and-a-half syndrome, 236, V3
"Onionskin" lesion, 2153
"Onion skinning," 2446
Online Mendelian Inheritance in Man, 3777, 3778t
Onycholysis, 386, 390, 421
Onychomadesis, 421
Onychomycosis, 389, 1692, 1713, 1724, 1724t
O'nyong-nyong virus, 1059, 1657t, 1669
Oocyte donation, 3135, 3816, 3817f
Oogonia, 3124, 3125f
Oophorectomy, 554, 634, 710, 712
Oophoritis, 1648
OPA1, 966t
OPA1 gene mutations, 3400
Opacity-associated protein (Opa), 1253
OPC-31260, 2433
OPC-67683. *See* Delamanid
Open-access endoscopy, 2499–2500
Open-label placebos, 3930t, 3934
OPG. *See* Osteoprotegerin (OPG)
Ophthalmia neonatorum, 1257, 1259t
Ophthalmomyiasis, 3732
Ophthalmopathy
  diabetic. *See* Diabetic retinopathy
  Graves'. *See* Graves' (thyroid-associated) ophthalmopathy
Ophthalmoplegia
  internuclear, 236, 236f, V3
  progressive external. *See* Progressive external ophthalmoplegia
Opicapone, 3503, 3504t
Opioid(s)
  actions of, 98, 3654, 3654f, 3656t
  adverse effects of, 81, 253, 2955t
    constipation, 509
    delirium, 184
    dysphagia, 295
    hypogonadism, 3113
    management of, 98
    pruritus, 422
    respiratory depression, 98
    SA node dysfunction, 1920t
  chronic/long-term use of, 75t, 98, 101t
  for cough, 253
  COX inhibitor combinations, 98–99
  cross-tolerance among, 81
  dosage of, 81
  for dyspnea, 83, 271, 510
  endogenous, 96
  extended-release formulations of, 100
  for frostbite, 3772
  guidelines for prescribing, 98, 101t
  for migraine, 3467t, 3469
  most misused, 3689. *See also* Opioid-related disorders
  neurobiology of, 3689–3690, 3689t, 3692f
  in older adults, 3883
  for osteoarthritis, 2955, 2955t
  overdosage/poisoning with, 190, 3690–3691, 3690t, 3705, 3715t
  for pain, 80f, 81, 97t, 98
  pharmacology of, 3690
  for restless legs syndrome, 217
  routes of administration of, 98
  for sickle cell pain, 773
  withdrawal from, 3690, 3691
Opioid antagonists, 3691
Opioid receptors, 96, 3689–3690, 3692f, 3692t
Opioid-related disorders, **3688**
  clinical features of, 3690
  complications of, 3690
  diagnosis of, 3688
  epidemiology of, 3689–3690
  pathogenesis of, 3654, 3654f
  prescription opioids and, 100, 3689
  prevention of, 3692
  screening for, 13
  in terminally ill patient, 81
  treatment of, 3691–3692
Opioid use disorder, 3688
Opisthorchis/opisthorchiasis, 667, 961t, 1826t, 1830–1834, 1830t, 1831t
Opsoclonus, 237
Opsoclonus-myoclonus syndrome, paraneoplastic, 749, 2791t
Opsonization, 2771
Optical coherence tomography (OCT), 1908–1909, 1909f, 2128
Optic ataxia, 206
Optic atrophy, 228, 228f, 1432, 3400
Optic chiasm, 222, 2999t
Optic cup, 230
Optic disc, 222, 228–229, 228f
Optic disc drusen, 229, 229f
Optic gliomas, 3003
Optic nerve, 220
  disorders of, 222. *See also* Optic neuritis; Optic neuropathy
  examination of, 3380
  injuries of, 3573–3574
  tumors of, 232
Optic neuritis, 221
  color desaturation in, 222
  epidemiology of, 227
  ethambutol-related, 1423
  in MS, 227, 3577
  in neuromyelitis optica. *See* Neuromyelitis optica (NMO)
  recurrent, 227
  retrobulbar, 221, 227, 227f
  treatment of, 227, 3397
  visual field defects in, 222
Optic neuropathy
  anterior ischemic, 226–227, 227f
  drug-induced, 228–229
  in giant cell arteritis, 2902
  Leber's. *See* Leber's hereditary optic neuropathy
  posterior ischemic, 227
  in sarcoidosis, 2927
  toxic, 228–229, 228f
Opt-in vs. opt-out policies, 3918
Optokinetic reflex, 220
Oral appliances, for OSA, 2281
Oral cancer
  clinical features of, 261, 263t, 611, A3
  diagnosis of, 261
  incidence of and deaths from, 497t
  prevention of, 512
  risk factors for, 261, 388, 3677
  smoking and, 511
Oral cavity. *See also* Oral mucosa
  anaerobic bacterial infections of, 1373, 1374
  dry. *See* Xerostomia
  healthy, A3
  systemic disease signs in, **260, A3**
Oral cholecystography (OCG), 2730
Oral contraceptives, **3153**
  for acne vulgaris, 391
  adverse effects of
    cholestasis, 324, 2667t
    cutaneous, 144t, 399, 405
    dyslipidemia, 3246
    hair loss, 421
    hepatic adenoma, 670
    hirsutism, 421
    hyperkalemia, 359
    hypertension, 2137, 3165
    liver function test impairment, 2667
    oral pigmentation, 264t
    pigmentation changes, 420
    venous thrombosis, 940
  for amenorrhea, 3135
  benefits of, 3153
  contraindications to, 639, 3141, 3152t, 3153
  drug interactions of, 489t, 3427
  for dysmenorrhea, 3137

Oral contraceptives (*Cont.*):
  effectiveness of, 3153t
  for endometriosis, 3137
  for hirsutism, 3141
  HIV infection risk and, 1563
  for PCOS, 3135
  in perimenopause, 3142
  progestin-only, 3153–3154
Oral-facial-digital syndrome type I, 2430t
Oral glucose tolerance test, 3196–3197
Oral hairy leukoplakia, 264t
  in EBV disease, 1510, 1510f, A5
  in HIV infection, 1598, 1599f
  treatment of, 1512
Oral mucosa
  age-related changes in, 267
  pigmented lesions, 264t
  vesicular, bullous, or ulcerative lesions, 258–263t
  white lesions, 264t
Oral pain, nondental, **265**
Oral rehydration solution (ORS)
  for cholera, 1328–1329, 1328t
  composition of, 1081, 1325, 1328, 1328t
  for infectious diarrhea, 307, 1081, 1081t
  for *Shigella* infections, 1325
Oral sex, HIV transmission and, 1562t, 1563
Oral ulcers, 261–262, 263t, 2908, 2908f, A3
Orán virus, 1661t, 1674
Orbital atherectomy, A11, A11v
Orbital cellulitis, 233, 3558
Orbital pseudotumor, 233, 2931t, 2932f, V3
Orbital tumors, 233
Orbit, disorders of, A16
Orbitofrontal cortex, 237
Orbivirus, 1658t, 1663
Orchidopexy, 705, 3100
Orchiectomy, 554, 567, 705, 706
Orchitis, 1648, 3114
Orexin (hypocretin), 210, 211f, 215, 242
Orf, 1049, 1518t
Organic acid(s), renal transport of, 2369
Organic anion transport protein 1B1 (OATP1B1), 2637f, 2638, 2642
Organic dusts, 2236t, 2238–2239
Organ of Corti, 244, 245f
Organoids, 3402
Organophosphates, 306, 3610t, 3712t
Organ transplantation. *See* Solid organ transplantation
Orgasmic disorder, female, 3161
Oriboca virus, 1659t, 1667
Orientia tsutsugamushi, 992, 1455t, 1460. *See also* Scrub typhus
Oritavancin
  actions of, 1166, 1182t
  adverse effects of, 1171t, 1176
  drug interactions of, 1172t
  for enterococcal infections, 1220
  indications for, 1173t, 1176
  in pregnancy and lactation, 1169t
  resistance to, 1173t, 1182t, 1184
  for staphylococcal infections, 1203t, 1204
Orlistat, 2613, 3192–3193, 3192t, 3257
ORMDL2 gene mutations, 2553, 2554t
Ornithine decarboxylase, 1751
Ornithine metabolism disorders, 3369t
Ornithine transcarbamylase deficiency, 3369t, 3372
Ornithodoros spp. tick, 1445, 1445f
Orofacial myofunctional therapy (OMT), 2282
Oromandibular dystonia, 3510
Oropharyngeal disease
  cancer, 512–513, 611. *See also* Head and neck cancer
  dysphagia in, 294
  in systemic sclerosis, 2870, 2870t
  in tularemia, 1338, 1338t
Oropouche virus, 1659t, 1667
Orotic aciduria, 790
Oroya fever, 1353t, 1356
Orphenadrine, 3713t, 3715t
ORS. *See* Oral rehydration solution (ORS)
Orthobunyavirus, 1656, 1658–1659t, 1669
Orthodeoxia, 2630
Orthodromic AV reciprocating tachycardia, 1932t, 1940f, 1941
Orthoebolavirus, 1680, 1680–1682f. *See also* Ebola virus disease; Filoviridae
Orthoflavirus, 1656, 1659t, 1668, 1678
Orthohantavirus, 1656, 1660–1661t, 1674, 1676
Orthoherpesviridae (herpesviruses), 1478f, 1480t, 1481, 1535–1536, S11. *See also specific viruses*
Orthomarburgvirus, 1680, 1680–1682f. *See also* Filoviridae
Orthomyxoviridae
  clinical syndromes of, 1533, 1666t, 1669. *See also* Influenza
  geographic distribution of, 1664–1665t
  reservoirs and vectors of, 1656, 1662t
  structure of, 1478f, 1480t
Orthonairovirus, 1656, 1661t
Orthopnea, 270, 1982
Orthopoxvirus infections. *See also specific viruses*
  clinical features of, 1519
  differential diagnosis of, 1519
  epidemiology of, 1517–1518
  in humans, 1518t
  pathogenesis of, 1518–1519
  treatment of, 1521
  zoonotic characteristics of, 1518t
Orthostatic hypotension
  causes of, 155t, 159
  classification of, 3547
  clinical features of, 159, 3547
  definition of, 1862, 3546
  diabetic neuropathy and, 3227
  diagnosis of, 3547
  dizziness in, 162, 3547
  in hypovolemia, 347
  in opioid-related disorders, 3690
  pathophysiology of, 159, 159f, 1862, 3546
  in PD, 3506
  postprandial, 3547
  prevalence of, 3546
  treatment of, 160, 3227, 3493, 3547–3548
  variants of, 3547
Orthostatic proteinuria, 2414
Orthostatic tachycardia, 347
Orthostatic tremor, 178
Ortner's syndrome, 2049
Orungo virus, 1658t, 1669
OSAHS. *See* Obstructive sleep apnea/hypopnea syndrome (OSAHS)
Osborn wave, 1874
Oscillopsia, 162
Oseltamivir
  actions of, 1487
  adverse effects of, 1487
  for influenza, 1028, 1157, 1488, 1488t, 1537, 1550
  for influenza prophylaxis, 1488t, 1594t
  pharmacology of, 1487
  resistance to, 1487, 1488
Osilodrostat, 3012, 3059
Osimertinib
  action and targets of, 532, 533t, 569t, 572
  adverse effects of, 569t, 595
  gene mutations and response to, 3982
  interactions and issues, 569t
  for lung cancer, 625, 625t, 627, 627t
Osler's nodes, 1038, 1860
Osler-Weber-Rendu syndrome (hereditary hemorrhagic telangiectasia)
  bleeding in, 472
  clinical features of, 396
  gastrointestinal bleeding in, 318
  skin manifestations of, 396, 1859
  vascular malformations in, 927, 3456
Osler, William, 52
OSM (oncostatin M), 2767t
Osmolar gap, 371
Osmolytes, 350
Osmoreceptive, 2371
Osmoreceptors, 2371f, 3014
Osmoregulation, 2371, 2371f
Osmotic demyelination syndrome (central pontine myelinolysis), 350, 352, 2348–2349, 2349f, 3461, S1
Osmotic diarrhea, 308, 352, 2540, 2549. *See also* Diarrhea
Osmotic diuresis, 346
Osmotic fragility, 793
Osmotic laxatives. *See* Laxatives
OspA, 1449
OspC, 1449
Ospemifene, 3145, 3149
Ossa virus, 1659t, 1667
Ossification, extraskeletal, 3318–3319, 3318t
Osteitis condensans ilii, 2886
Osteitis fibrosa cystica, 366, 2390, 3277, 3290–3291
Osteoarthritis (OA), **2949**
  in acromegaly, 2966
  clinical features of, 2952–2953
  definition of, 2949
  diagnosis of, 2949, 2953, 2953f
  genetic considerations in, 2951
  global considerations in, 2949
  joints affected by, 2949, 2949f
    hand, 2949, 2949f
    hip, 2951, 2953
    knee, 2949, 2951, 2951f, 2953, 2953f
  oral pain in, 265
  pain sources in, 2952
  pathogenesis of, 2949–2950, 2950f
  pathology of, 2952
  prevalence of, 2949
  vs. psoriatic arthritis, 2889
  risk factors for, 2950–2951, 2951f
  treatment of, **2953,** 2955t
    acetaminophen, 2955
    cartilage regeneration, 2956
    COX-2 inhibitors, 2955
    exercise, 2954
    intraarticular injections, 2955
    malalignment correction, 2954
    NSAIDs, 2955
    in older adults, 3883
    opioid analgesics, 2955
    surgical, 2956
Osteoarthropathy, hypertrophic. *See* Hypertrophic osteoarthropathy
Osteoblast(s), 3259, 3261f, 3272, 3312f
Osteoblastic lesion, 730, 731
Osteoblastoma, 737t
Osteochondrodysplasias, 3318
Osteochondroma, 729, 730, 3318
Osteoclast(s), 3260–3261, 3260f, 3261f, 3272, 3312f
Osteoclast activating factor, 885
Osteocyte(s), 3259
Osteocytic osteolysis, 3260
Osteodystrophy, renal, 339, 3290
Osteogenesis, 730

Osteogenesis imperfecta, **3324**
  classification of, 3324, 3325t
  clinical features of, 245, 1860, 2043, 2058, 3324–3325
  diagnosis of, 3327
  genetic considerations in, 3325t, 3326–3327, 3787
  incidence of, 3324
  treatment of, 2609, 3327
Osteolytic lesion, 730
Osteoma cutis, 404
Osteomalacia
  axial, 3316–3317
  diagnosis of, 3270–3271
  in hypophosphatasia, 3316
  hypophosphatemia and, 3270
  neck pain in, 132
  oncogenic (tumor-induced), 737t, 740, 3264
  pathogenesis of, 3270
  postgastrectomy, 2534
  vitamin D deficiency and, 3289
Osteomyelitis, **1061**
  Actinomyces, 1363f, 1365
  anaerobic bacterial, 1377
  from bite-wound infections, 1143
  Blastomyces, 1701, 1702, 1703t
  Brucella, 1333
  classification of, 1061
  in diabetes mellitus, 3229
  E. coli, 1287
  enterococcal, 1218
  foot, 49t, 1068–1069, 1068f, 1311
  in long bones, 1063–1065, 1064t, 1066f
  nontyphoidal *Salmonella,* 1320
  P. aeruginosa, 1309t, 1311
  in prosthetic joints. *See* Prosthetic joint infections
  psoas abscess in, 1075
  S. aureus, 1199, 1199f
  skull, 1377
  sternal, 1067–1068, 1067f
  treatment of, 1205, 3229
  tuberculous, 1058, 1062, 1066
  vertebral, 1311
    clinical features of, 131, 150, 1062, 1199
    complications of, 1063
    diagnosis of, 1062, 1063f, 1199f
    epidemiology of, 1062
    global considerations in, 1063
    microbiology of, 1062
    pathogenesis of, 1061–1062
    treatment of, 1062–1063, 1064t
  in war veterans, S7
Osteonecrosis
  ART-associated, 1604
  drug-induced, 2940t
  femoral head, 2948f
  in IBD, 2562t, 2563
  of jaw, 266, 579, 3285, 3304, 3307, A3
  in sickle cell disease, 775
Osteopathic medicine, 3925t, 3928
Osteopenia, 2752, 2940t
Osteopetrosis, 3314–3315
Osteopoikilosis, 3316
Osteopontin, 759
Osteoporosis, **3291**
  androgen deprivation therapy and, 701, 3176
  approach to the patient, 3298–3299
  in axial spondyloarthritis, 2886
  chemotherapy-related, 756
  definition of, 3291
  diagnosis of
    biochemical markers, 3299, 3299t
    bone biopsy, 3299
    bone mass measurement, 3296–3297, 3297f, 3297t, 3298f
    laboratory evaluation, 3299
    vertebral imaging, 3297, 3298t
  diet and, 2603
  differential diagnosis of, 3299
  epidemiology of, 3291–3293
  fractures in
    epidemiology of, 3291–3292, 3292f
    management of, 3300
    prevention of, 3300, 3305f. *See also* Osteoporosis, treatment of
    risk factors for, 3145, 3292–3293, 3292f, 3292t, 3296–3297, 3298f
    vertebral. *See* Vertebral fractures
  glucocorticoid-induced, 2612, 2849, 3296, 3310–3311
  heparin-induced, 948, 2940t
  in HIV infection, 1604
  in IBD, 2562t, 2563
  in liver disease, 2717
  neck pain in, 132
  pathophysiology of, 3293–3294f, 3293–3296
  prednisone and, 2849
  prevalence of, 2979t
  prevention of, 3144t, 3145
  in rheumatoid arthritis, 2842
  risk factors for, 3291, 3292t
    chronic disease, 3295–3296, 3296t
    cigarette smoking, 3296
    estrogen status, 3295
    low calcium intake, 3294–3295
    medications, 3296, 3296t
    pancreatitis, 2752
    physical inactivity, 3295
    reduction in, 3300
    vitamin D deficiency, 3295
  screening for, 38, 39t, 2979t, 3881t
  in short-bowel syndrome, 2548
  treatment of, **3300**
    abaloparatide, 3308–3309
    bisphosphonates, 3303–3306, 3305f
    calcitonin, 3306
    calcium, 3300–3301, 3301t
    denosumab, 3306–3307, 3306f
    estrogens, 3302–3303
    exercise, 3301–3302
    indications for, 3302
    monitoring of, 3310
    nonpharmacologic, 3309–3310
    nutritional recommendations, 3300–3301
    progestins, 3303, 3303f
    risk factor reduction, 3300
    romosozumab, 3309, 3310f
    sequence of therapy, 3309
    SERMs, 3303
    teriparatide, 3307–3308, 3307f, 3308f
    unapproved agents, 3309
    vitamin D, 3301
  in women, 3166
Osteoprotegerin (OPG)
  in bone metastasis, 541
  in bone remodeling, 3294, 3294f
  in osteopetrosis, 3312f, 3315
  in primary myelofibrosis, 821
  in rheumatoid arthritis, 2845f, 2846
Osteosarcoma, 261, **729**
Osteosclerosis, 821, 3316. *See also* Paget's disease of bone
Ostium primum, 286, A11
Ostium secundum, 285, A11
Oteseconazole, 1692, 1698
Otitis externa, **254,** 1309t, 1311, 1715, 3229
Otitis media
  acute, 254–255, 254t, 1193, 1194
  anaerobic bacterial, 1374
  antibiotics for, 254–255, 254t, 1194
  C. auris, 1226
  chronic, 1374
  clinical features of, 1193
  with effusion, 255
  hearing loss in, 245
  H. influenzae, 1262
  in influenza, 1548
  M. catarrhalis, 1265
  in nasopharyngeal cancer, 611
  treatment of, 1194
Otoacoustic emissions, 247
Otoconia, 163
Otorrhea, 254
Otosclerosis, 245, 249, 3386f, A16
Otoscopy, 245
OTUD4 gene mutations, 3112t
Outcome (quality theory), 43
Outcome evaluation, 4
Outer-membrane proteins, 1253
Outer-membrane vesicles (OMVs), 1252, 1252f, 1295
Outflow tract VT, 1967
Outpatient care, 3
Ova and parasite examination, 306
Ovalocytosis, malaria resistance and, 1762–1763
Ovarian cancer, **709**
  breast cancer and, 710
  carcinoma of unknown primary and, 735
  clinical features of, 327, 710
  deaths from, 497t, 499f, 500t
  epidemiology of, 709–710
  epithelial, 710–711
  genetic considerations in, 520t, 524t, 710
  germ cell tumors, 712
  incidence of, 497t, 710
  in Lynch syndrome, 710
  metastatic, 406, 477
  oral contraceptives and, 3153
  paraneoplastic syndromes in, 737t, 738, 2827
  pathogenesis of, 710
  pathology of, 709–710
  postmenopausal hormone therapy and, 3144t, 3147
  prognosis of, 711t, 712
  protective factors for, 710
  recurrent, 712
  risk factors for, 710
  screening for, 517t, 518, 710
  staging of, 710, 711t
  stromal tumors, 712
  surgical prevention of, 514
  treatment of, 711–712, 711f
  tumor markers in, 504t, 710
Ovarian follicles, 3125–3126, 3125–3126f
Ovarian hyperstimulation syndrome, 2270
Ovarian reserve, 3150
Ovary
  development of, 3094, 3124–3126, 3125–3126f
  disorders of
    borderline tumors, 712
    cancer. *See* Ovarian cancer
    cyst, 112, 113, 472, 3136
    dysfunction/failure, 3133–3134, 3151, 3789
    teratoma. *See* Teratoma, ovarian
  function of, **3126**
    assessment of, 3130
    hypothalamic secretion in, 3126–3127, 3127f
    ovarian peptides in, 3128
    ovarian steroids in, 3127–3128, 3127f. *See also* Estrogen(s); Progesterone
    pituitary secretion in, 3127f, 3128
  metastatic disease to, 711
Overactive bladder, 332, 3174, 3889–3890, 3890t, 3893t
Overdiagnosis, 514, 617
Overdosage. *See* Poisoning/drug overdose
Overflow diarrhea, 305

Overhydrated stomatocytosis, 795
Overlapping sphincteroplasty, 2583
Overoptimism, 3916t, 3917
Overweight, 12, 3181, 3181f, 3188t. *See also* Obesity
Ovotesticular disorders of sexual development, 3097t, 3098
Ovulation assessment, 3150, 3151
OXA-48 (OXA group carbapenemase), 1181
Oxacillin
  adverse effects of, 355, 2667
  for cellulitis, 1054t
  for epidural abscess, 991t
  indications for, 1173t, 1175
  for infective endocarditis, 1044
  for nongonococcal bacterial arthritis, 1057
  for osteomyelitis, 1064t, 1065t
  resistance to, 1041, 1173t
  for staphylococcal infections, 1203t, 1204
Oxalate, 369, 371, 2449
Oxaliplatin
  actions of, 563
  adverse effects of, 562t, 563, 753t
    desensitization protocol for, 2819
    hepatic, 2667
    peripheral neuropathy, 658
  for colorectal cancer, 657, 658
  for gastric cancer, 644, 649
  interactions and issues, 562t
  for pancreatic cancer, 674, 676t
  for testicular cancer, 709
Oxalosis, 2962, 2962f, A4
Oxamniquine, 1745t, 1750
Oxandrolone, 3119
Oxazepam, 3660t, 3713t
Oxazolidinones. *See* Linezolid; Tedizolid
Oxcarbazepine
  adverse effects of, 3420, 3422t, 3670t
  for anxiety disorders, 3661
  for bipolar disorder, 3668, 3670t
  drug interactions of, 3422t
  for neuropathic pain, 82
  overdosage/poisoning with, 3713t
  for pain, 97t
  pharmacology of, 3422t
  for seizures, 3420, 3422t
  for trigeminal neuralgia, 3553
Oxicam, 425
Oxidative burst, 970
Oxidative phosphorylation, 3645
Oxidative stress, 3255, 3871, 3875
P450 Oxidoreductase deficiency, 3071t
Oximes, S5
5-Oxoproline, 371
3-Oxothiolase deficiency, 3370t
Oxybutynin, 217, 510, 3143, 3506, 3549, 3570, 3588
Oxycodone
  abuse of, 100. *See also* Opioid-related disorders
  for dyspnea, 83t
  for pain, 97t
Oxycodone extended-release, 97t, 100
Oxygen delivery, 292, 292f, 2296–2297, 2310
Oxygen-hemoglobin dissociation curve, 277, 770–771, 771f
Oxygen therapy
  for anaphylaxis, 2813
  for cluster headache, 3472
  for COPD, 2256
  for COPD exacerbation, 2259
  for dyspnea, 83, 271, 510
  for high-altitude illness, 3756t, 3758
  hyperbaric. *See* Hyperbaric oxygen therapy
  for hyponatremia, 351
  for hypothermia, 3770
  for pulmonary edema, 2330
  for respiratory acidosis, 376
  for sepsis/septic shock, 2322t, 2323
  for shock, 2315
  for STEMI, 2118
Oxygen toxicity, 3765, 3767
Oxygen uptake, 2204
Oxymetazoline, 253, 391, 2810
Oxyntic cell, 2514, 2515f
Oxyntic gland, 2514, 2515f
Oxyphenbutazone, 923t
Oxytocin, 3016
Ozanimod, 2567, 2568t, 3586t
Ozone exposure, 2239t

P

P1PK RBC group system, 904t
P2X3 channel antagonists, 274
P3H1 gene mutations, 3324, 3326
p14ARF, 530
p16, 599, 601, 671t, 672, 836, 3940
p16Ink4a, 530, 538
p21Cip2/Waf1, 530
p24, 976, 978f
p24 antigen capture assay, 1589–1590, 1590t
p27Kip1, 530
p40, 616
P_{50}, 770–771, 771f
p53
  activation of, 530–531, 530f
  adenoviral-mediated expression of, 3828
  in brain tumors, 717
  in cancer cells, 530–531, 546, 2781, 2791t
  in melanoma, 599
  in SLE, 2791t
  VEGF inhibition and, 547f
p53 gene mutations
  in B-cell prolymphocytic leukemia, 872
  in cancer, 530–531, 530f, 3783t, 3994
  in carcinoma of unknown primary, 731
  in colorectal cancer, 653
  in head and neck cancer, 611
  in Li-Fraumeni syndrome, 717t, 727
  in skin cancer, 431, 606
  in skin response to sun, 430
  in thyroid cancer, 3046
p57, 761
p62, 2791t, 2914, 2915f
p63, 616, 3939
p70S6K, 526f
p-80-Collin, 2790t
$\text{p190}^{\text{BCR:ACBL1}}$, 835, 835t
$\text{p210}^{\text{BCR:ACBL1}}$, 835, 835t
$\text{p230}^{\text{BCR:ACBL1}}$, 835, 835t
P450 oxidoreductase deficiency, 3071t, 3099
PA 824. *See* Pretomanid
PABA (*p*-aminobenzoic acid), 435, 436t, 1166, 1185
PABP2 gene mutations, 3643
Pacemaker(s), **1920**
  after STEMI, 2125
  for atrial fibrillation, 1950
  for AV conduction block, 1930–1931f, 1930–1932
  CMR in patient with, 1884
  complications of, 1924
  five-letter codes for, 1923–1924
  indications for, 1917, 1922–1923, 1923f, 1923t, 1930–1931, 1930f
  MRI in patient with, 3389
  nomenclature of, 1923–1924
  pacing mode, 1923–1924
  for pulmonary edema, 2331
  for SA node dysfunction, 1922–1923, 1923t
  for STEMI, 1930
  temporary, 1920, 1930
  types of, 1930–1931, 1931f
Pacemaker cells, 1917
Pace test, 121t
Pachydermoperiostosis, 2970. *See also* Hypertrophic osteoarthropathy
Packed red blood cells. *See* Transfusion(s), packed red blood cell
Paclitaxel
  actions of, 565
  adverse effects of, 563t, 565
    desensitization to, 2818
    DIHS, 598
    hypersensitivity reaction, 2813
    long-term, 753t
    lupus syndrome, 2940t
    myopathy, 2940t
    nasal vestibulitis, 510
    neurologic, 3608t
    neuropathy, 508
    premedication for, 563t, 565
    pulmonary, 595
  for anal cancer, 659
  for bladder cancer, 693
  for carcinoma of unknown primary, 735
  in drug-eluting stents, 2128
  for endometrial cancer, 715
  for gastric cancer, 649
  interactions and issues, 563t
  for Kaposi's sarcoma, 1613
  for lung cancer, 625t, 627t
  for metastatic cervical cancer, 713
  for ovarian cancer, 712
  for pancreatic cancer, 674
  for testicular cancer, 709
  for thymic tumors, 630
  for triple-negative breast cancer, 641
Pacritinib, 821
PACT Act, S8
PAD14 gene mutations, 2843
PAF (pure autonomic failure), 3548
Paget-Schroetter syndrome, 2176
Paget's disease, extramammary, 609, 609f
Paget's disease of bone, **3311**
  clinical features of, 3312–3313
    multiple-site, A15
    neck pain, 132
    oral, 261
    tibia, A15
  complications of, 3313
  diagnosis of, 3313–3314, 3313f
  epidemiology of, 3311–3312
  etiology of, 3312, 3312f
  genetic considerations in, 3312
  juvenile, 3312
  pathophysiology of, 3312
  treatment of, 3314, 3314t
PAH. *See* Pulmonary arterial hypertension (PAH)
PAI-1 (plasminogen activator inhibitor I), 941, 954, 954f, 1845
Pain
  acute, treatment of, **96**
    acetaminophen for. *See* Acetaminophen
    aspirin for. *See* Aspirin
    nonsteroidal anti-inflammatory drugs for. *See* Nonsteroidal anti-inflammatory drugs
    opioids for. *See* Opioid(s)
  assessment of, 80, 507f, 3382
  in cancer patient, 506–508, 507–508t, 703
  chronic, **99,** 100t, 101t, 3883, S8. *See also* Opioid(s)
  complementary and integrative therapies for, 3924
  depression and, 99
  etiology of, 80
  frequency of, 80
  incident (movement), 508, 508t

  neuropathic, 80, 80f, 96, 100–101, 506, 508, 508t. *See also* Neuropathy
  nociceptive, 80, 80f, 508t
  pathophysiology of, **93**
    central mechanisms, 94–95, 95f
    neuropathic pain, 96
    pain modulation, 95–96, 95f
    peripheral mechanisms, 93–94, 93f, 94f
  perception of, 113
  psychological variables, 99
  racial/ethnic disparities in treatment of, 62
  referred, 94, 95f, 110t, 111
  somatic, 80
  sympathetically maintained, 96
  in terminally ill patient, 80–82, 80f
  testing sense of, 172, 174t
  visceral, 80, 508t
Painful articular syndrome, 1603
Painless (silent) thyroiditis, 3038–3039
PALB2 gene mutations, 633, 671, 671t, 710
Palbociclib
  action and targets of, 533t, 570t, 573
  adverse effects of, 570t, 753t
  for breast cancer, 573
  interactions and issues, 570t
Palifermin, 916
Palilalia, 3515
Paliperidone, 3673, 3673t, 3674
Palivizumab, 1489, 1538
Palla’s sign, 2160
Palliative care, **76,** 87–88, 502, 553, 576, 3894. *See also* End-of-life care
Palliative sedation, 84
Pallister-Hall syndrome, 3003
Palmar crease, 1860
Palmar erythema, 2629, 2709, 2709f
Palmar fasciitis, 2972
Palmomental response, 3381
Palonosetron, 83, 575
Palopegteriparatide, 364, 3287
Palovarotene, 3319
Palpation
  of abdomen, 111
  of chest, 2200
  of heart, 1864
  of spleen, 479
Palpitations, **291**
Palsy
  Bell’s. *See* Bell’s palsy
  cranial nerve. *See* Cranial nerve disorders
2-PAM Cl (2-pralidoxime chloride), 3712t, S5
Pamidronate
  for bone pain, 82
  for hypercalcemia, 364, 738, 3284t, 3285
  for metastatic bone disease, 731
  for multiple myeloma, 892
  for Paget’s disease of bone, 3314t
p-aminobenzoic acid (PABA), 435, 436t, 1166, 1185
PAMPs. *See* Pathogen-associated molecular patterns (PAMPs)
Pan American Sanitary Bureau, 3843
pANCA (perinuclear ANCA), 2894
Pancoast tumor (superior sulcus tumor), 133, 619, 626, 2943, 3615
Pancreas
  anatomy of, 672, 672f
  annular, 2754
  beta cells of, 3198, 3200
  biopsy of, 2741, 2742
  disorders of. *See* Pancreatic disease
  exocrine secretion in, **2742**
    in autoprotection of pancreas, 2743
    enteropancreatic axis and feedback inhibition in, 2743
    enzymes in, 2743
    insufficient, 2752. *See also* Chronic pancreatitis
    laboratory evaluation of, 2739t, 2742
    regulation of, 2742–2743
    in steatorrhea, 2542
    water and electrolyte content of, 2743
  transplantation, 3216, 3225f, 3234
Pancreas divisum, 2754
Pancreas-kidney transplantation, 3226
Pancreatectomy, 2754
Pancreatic agenesis, 3783t
Pancreatic cancer, **671**
  clinical features of, 406, 673
  deaths from, 497t, 499f, 500t, 671
  diagnosis of, 673–674, 673f
  epidemiology of, 671
  extrahepatic cholestasis in, 326
  genetic considerations in, 520t, 538, 671, 671t
  incidence of, 497t
  obstruction due to, 2500f
  paraneoplastic syndromes in, 737t
  pathology and molecular considerations in, 672, 672f
  risk factors for, 671–672
  staging of, 674, 675t
  treatment of, 674–676, 675t, 676t
  tumor markers in, 504t, 674
Pancreatic cholera. *See* VIPoma (vasoactive intestinal peptidoma)
Pancreatic disease. *See also* Pancreatitis
  abdominal pain in, 110
  annular pancreas, 2754
  approach to the patient, **2738**
  ascites in, 328
  cancer. *See* Pancreatic cancer
  in cystic fibrosis, 2245. *See also* Cystic fibrosis (CF)
  diabetes mellitus and, 3203
  diagnosis of
    biopsy, 2741, 2742
    imaging, 2738f, 2739t, 2740–2742, 2741t
    pancreatic enzymes, 2739t, 2740
    pancreatic function tests, 2739t, 2742
  endocrine tumors. *See* Neuroendocrine tumors (NETs), pancreatic
  epidemiology of, 2738
  etiology of, 2738
  indigestion in, 300
  macroamylasemia, 2754
  mucinous cysts, 672, 2738, 2738f
  pancreas divisum, 2754
  skin manifestations of, 406
Pancreatic enzyme(s)
  activation of, in pancreatitis, 2744
  in pancreatic disease diagnosis, 2739t, 2740, 2741t
  preparations, 312, 674, 2753
Pancreatic intraepithelial neoplasias, 672
Pancreatic islet transplantation, 3216
Pancreatic necrosis, 2469, 2499, 2747f, 2749, V5
Pancreatic neuroendocrine tumors (pNETs). *See* Neuroendocrine tumors (NETs), pancreatic
Pancreaticopleural fistula, 2748f
Pancreatic polypeptide-secreting tumors (PPomas), 3082–3083
Pancreatitis. *See also* Acute pancreatitis; Chronic pancreatitis
  ascites in, 328
  autoimmune, 2751t, 2930, 2931t
  biliary tract disease and, 2735
  cobalamin malabsorption in, 787
  diagnosis of, 80, 112
  differential diagnosis of, 2745–2746
  endoscopic findings in, 2498–2499
  gallstone, 2489, 2748
  hereditary, 2754
  hypocalcemia in, 365, 3290
  hypovolemia in, 346
  in IBD, 2562t
  in mumps, 1648
  nausea and vomiting in, 297
  pancreatic cancer risk and, 671
  pleural effusions in, 2745
  post-ERCP, 2473
  tumor markers in, 504t
Pancuronium, 2352
Pancytopenia, 804, 808, 808t
Pandemic, 1544
Pandoraea commovens, 1269
Panencephalitis, 1117–1118
Pangender, 3179t
Panic attack, 291
Panic disorder, 103t, 107, **3657**
Panitumumab
  action and targets of, 534t, 558, 559f
  adverse effects of, 558, 594, 658
  for colorectal cancer, 558, 658
  genetic variations in response to, E72t
Panniculitis, 405, 2937
Panobinostat, 533t, 539, 570t, 573, 890t, 892, 3942
Pansexual, 3179t
Panton-Valentine leukocidin (PVL) toxin, 1027, 1052, 1198
Pantoprazole, 301, 2524–2525, 2524t, 2537, 3585
Pantothenate kinase–associated neurodegeneration (PANK), 3497, 3518–3519
Pantothenic acid (vitamin B_5), 2600t, **2610**
Panuveitis, 225
PAPA (pyogenic arthritis, pyoderma gangrenosum, and acne) syndrome, 462, 2937
Papez circuit, 207
Papillary fibroelastoma, 2086
Papillary muscle rupture, 2122t
Papillary necrosis, 2357, 2442, 2442t
Papillary renal carcinoma, 524t, 525, 688, 688t. *See also* Renal cell carcinoma (RCC)
Papillary thyroid cancer, 3046, 3046f
Papilledema, 189, 228–229, 228f
Papillomaviridae, 1478f, 1480t. *See also* Human papillomavirus (HPV)
Papillon-Lefevre syndrome, 261
Pap smear, 38, 38t, 39t, 516t, 517, 713
Papular acrodermatitis of childhood, 2660
Papular eruptive xanthomas, A15
Papular mucinosis, 404
Papules, 137, 377t, 378f
Papulonodular lesions, **403**
  blue, 403t, 406
  etiology of, 1050t, 1051
  pink, 403t, 404
  purple, 403t, 406
  red, 403t, 405–406
  red-brown, 403t, 406
  skin-colored, 403t, 404
  violaceous, 403t, 406
  white, 403, 403t
  yellow, 403t, 404–405
Papulosquamous lesions, **386,** 386t, 392, 392t
PAR1 (protease activation receptor 1), 938
Para-aminosalicylic acid, 786t, 923t, 1391t, 1426
Paraballism, 3515
Paracellular transport, 2365
Paracentesis, abdominal, 328, 329
Paracoccidioidomycosis, 1688t, 1725–1726, 1726t
Paracrine regulation, 2987
Paradoxical pulse, 241–242
Paradoxical undressing, 3769
Paraesophageal hernia, 2505
Parafacial respiratory group, 2273

- Parafibromin, 3080t, 3084, 3087
- Paraganglioma
  - cardiac, 2086
  - definition of, 3073
  - diagnosis of, 3074, 3074t, 3075f
  - genetic considerations in, 3073, 3078–3079, 3079t
  - metastatic, 3077
  - pathogenesis of, 3073, 3073f
  - pathology of, 3074, 3075f, 3076f
  - syndromes, 3075f, 3077–3078
  - treatment of, 3075–3077
- *Paragonimus* spp. *See* Lung flukes (*Paragonimus* spp.)
- Parainfluenza virus infections
  - clinical features of, 143t, 253, 259, 1534
  - virology of, 1534, 1534t
- Paralysis
  - in botulism, 1236
  - brain-computer interfaces for treatment of, 3968–3969, 3968–3969f
  - definition of, 168
  - periodic
    - hyperkalemic (HyperKPP), 359, 3635, 3646, 3804t
    - hypokalemic. *See* Hypokalemic periodic paralysis (HypoKPP)
    - thyrotoxic, 3647
  - in rabies, 1652, 1653t
  - secondary hyperkalemic, 359
  - sleep, 213t, 214
  - tick-bite, 3730
  - Todd's, 3409
  - tongue, 3558
- Paralytic poliomyelitis, 1633–1634. *See also* Poliovirus infection
- Paralytic shellfish poisoning, 3727
- Paramyotonia congenita, 3635, 3646–3647
- Paramyxoviridae, 1478f, 1479, 1480t, 1534, 1534t
- Paranasal sinus cancer. *See* Head and neck cancer
- Paraneoplastic syndromes
  - ataxia in, 3533
  - autoimmunity and, 2826–2827
  - definition of, 736
  - endocrinologic, **736**
    - cancer immunotherapy-associated, 740
    - consumptive hypothyroidism, 740
    - ectopic ATCH production. *See* Adrenocorticotropic hormone (ACTH), ectopic production of
    - ectopic hCG production, 740
    - etiology of, 736–737, 737t
    - hypercalcemia. *See* Hypercalcemia, malignancy-associated
    - oncogenic osteomalacia, 740
    - tumor-associated SIAD, 738. *See also* Syndrome of inappropriate antidiuresis (SIAD)
    - tumor-induced hypoglycemia, 739–740
  - hematologic, **740,** 740t
    - eosinophilia, 741
    - erythrocytosis, 740–741
    - granulocytosis, 741
    - thrombocytosis, 741
    - venous thromboembolism, 741–742, 742t
  - immunopathology of, 2790–2791t
  - intrahepatic cholestasis, 325
  - joint involvement in, 2972
  - in lung cancer, 619
  - membranous nephropathy, 2359
  - neurologic, **743,** 743t
    - anti-Hu neuropathy, 3624
    - approach to the patient, 744
    - autoimmune autonomic ganglionopathy, 3550
    - cerebellar degeneration/atrophy, 749, 2791t
    - vs. Creutzfeldt-Jakob disease, 3532
    - diagnosis of, 744, 745f
    - dorsal root ganglionopathy, 750
    - encephalitides with intermediate- or low-risk antibodies. *See* Encephalitis, paraneoplastic
    - encephalitides with NMDA receptor antibodies, 3484
    - encephalomyelitis and focal encephalitis, 2791t
    - encephalomyelitis and focal encephalitis with high-risk antibodies, 744, 746, 746f
    - immune-related necrotizing myopathy, 751
    - Lambert-Eaton myasthenic syndrome. *See* Lambert-Eaton myasthenic syndrome (LEMS)
    - opsoclonus-myoclonus syndrome, 749, 2791t
    - pathogenesis of, 743–744, 743–744t
    - peripheral neuropathies, 750–751
    - sensory neuroneuropathy, 750
    - sensory neuroneuropathy/ganglionopathy, 3607
    - spinal cord, 749–750
    - stiff-person syndrome, 750, 750f, 2791t
    - visual syndromes, 751
  - pemphigus, 401, 410t, **411,** 2791t
  - in RCC, 688
- Paranoá virus, 1661t, 1674
- Paranodopathies, 3623
- Paraparesis, 170, 171f
- Paraphasia, 201
- *Paraplegin* gene, 3522t, 3525
- Parapneumonic effusion, 1192, 2268–2269
- *Parapoxvirus* infection, 1518, 1519, 1521. *See also specific viruses*
- Paraproteinemia, 473. *See also* Plasma cell disorders
- Parapsoriasis, 392
- Parasitic infections. *See also* Cestodes/cestode infections; Helminths/helminthic infections; Nematodes/nematode infections; Protozoa/protozoal infections; Trematodes/trematode infections
  - approach to the patient, 1738
  - arthritis in, 1059–1060
  - as cause of fever of unknown origin, 151t
  - hepatobiliary, 2737
  - laboratory diagnosis of, **S12**
  - myocarditis in, 2013
  - by organ system, 1738–1740t
  - by signs/symptoms, 1738–1740t
  - treatment of, **1740**
- Parasomnias, 217–218
- Paraspinal fullness, 120
- Parathyroid disorders
  - adenoma, 3275
  - carcinoma, 3275, 3277
  - genetic considerations in, 3276–3277, 3276f
  - hyperparathyroidism. *See* Hyperparathyroidism
  - hyperparathyroidism-jaw tumor (HPT-JT) syndrome, 3080t, 3087, 3087t, 3275, 3276
  - hyperplasia, 3275
  - hypoparathyroidism. *See* Hypoparathyroidism
  - in MEN syndromes. *See entries at* Multiple endocrine neoplasia
- Parathyroidectomy, 3278–3279, 3290–3291
- Parathyroid hormone (PTH), **3271**
  - actions of, 362f, 363, 2986, 3271–3272
  - analogues. *See* Teriparatide
  - assessment of, 2987
  - bone effects of, 3307
  - in calcium homeostasis, 2369, 3262, 3280f
  - in CKD, 2390
  - deficiency of, 365
  - ectopic production of, 365, 619, 736, 737t
  - excess of, 362f, 3272
  - half-life of, 2982
  - in hypercalcemia, 3283–3284
  - for hypoparathyroidism, 366, 3287
  - metabolism of, 3273
  - in primary hyperparathyroidism, 3278, 3278f
  - vs. PTHrP, 3273, 3273f
  - receptor cross-talk of, 2981
  - regulation of, 3272–3273
  - resistance to, 3288, 3289f
  - structure and physiology of, 3272, 3273f
  - synthesis of, 3272
- Parathyroid hormone (PTH)/PTHrP receptor, 2981, 2985t, 3273–3274, 3274f, 3280f
- Parathyroid hormone–related protein (PTHrP), **3273**
  - actions of, 3273–3274, 3274f
  - analogue. *See* Abaloparatide
  - ectopic production of, 363, 736, 737, 737t
  - vs. PTH, 2981, 3273, 3273f
  - structure and physiology of, 3273–3274, 3273f
- Paratonia (gegenhalten), 168
- Paravalvular leak closure, A11, A11v
- Parcopa, 3504t
- *PARD3B* gene, 1582t
- Parechoviruses, 1112, 1534t, 1535, 1638
- Parenteral nutrition
  - cholestasis in, 325
  - complications of, 2624–2625
  - in critically ill patient, 2625
  - diabetes mellitus and, 3220
  - discontinuation of, 2624
  - hypercalcemia in, 364
  - initiation and monitoring of, 2624
  - iron concentration during, 2625–2626
  - nutrient composition of, 2602, 2622–2623
- Parenting trends, vaccine hesitancy and, 16
- Paresis, **168,** 1432. *See also* Periodic paralysis
- Paresthesia, 172, 365, 2549t
- Paretic/hypotonic gait, 177t
- Parietal cell, 2514, 2515f, 2517–2518
- Parietal cell antibodies, 786
- Parietal cortex, posterior, 200f, 204
- Parietal lesions, 235
- Parietofrontal network, 200f, 204
- Parinaud's syndrome/complex (dorsal midbrain syndrome), 221, 236, 721, 1353, V3
- Paritaprevir/ritonavir/ombitasvir/dasabuvir, 1494, 2689
- *PARK7* gene mutations, 3498t, 3499
- Parkes-Weber syndrome, 2180
- *Parkin* gene mutations, 3400, 3499
- Parkinsonism
  - ALS and, 3523
  - atypical, 3496, 3497t
  - definition of, 3495
  - differential diagnosis of, 3494f
  - drug-induced, 3497
  - with dystonia, 3510t
  - gait disorders in, 177t, 178
  - lower body, 178
  - manganese-induced, 3702
  - in MSA, 3548
  - secondary, 3497, 3497t
  - in Wilson's disease. *See* Wilson's disease
- Parkinson's disease (PD), **3493**
  - AD and, 3476
  - chronic traumatic encephalopathy and, 3576
  - clinical features of, 3494–3495, 3494t
    - autonomic disturbances, 3506
    - cardinal, 3494, 3494t
    - constipation, 314, 3506
    - delirium, 182

    - dementia, 194, 3476, 3488, 3491–3492, 3506
    - dysphagia, 294, 296
    - frontal lobe syndrome, 208
    - motor, 50t, 177t, 178, 3494–3495, 3494t, 3507, V1
    - nonmotor, 3494t, 3506–3507
    - ocular, 235
    - olfactory dysfunction, 237, 241
    - psychoses, 3506
    - seborrheic dermatitis, 385
    - sleep disorders, 212, 218, 3506–3507
  - COVID-19 disease and, 3501
  - differential diagnosis of, 3495–3496, 3495t, 3497t
  - early-onset, 3499
  - ECG in, A8
  - etiology of, 3497–3500
  - falls in, 181
  - familial, 3498–3499
  - genetic considerations in, 3400, 3497–3500, 3498t
  - imaging in, 3495–3496, 3496f
  - microbiome and, 3400
  - pathogenesis of, 3497–3500, 3500f
  - pathology of, 3494, 3494f
  - pathophysiology of, 3396, 3500, 3500f
  - prevalence of, 3494
  - prions in, 3498–3499, 3525
  - vs. progressive supranuclear palsy, 3488
  - protein aggregation in, 3941, 3994
  - sex differences in, 3164
  - treatment of, **3501**
    - amantadine, 3503
    - anticholinergics, 3503
    - approach to, 3507–3508, 3508f
    - COMT inhibitors, 3503, 3504t
    - continuous dopaminergic delivery, 3504
    - deep brain stimulation, 3505
    - dopamine agonists, 3502–3503, 3504t
    - experimental, 3505–3506
    - gene therapy, 3506
    - levodopa, 3501–3502, 3501f, 3504t
    - MAO-B inhibitors, 3503, 3504t
    - neuroprotective agents, 3504
    - for nonmotor and nondopaminergic features, 3506–3507
    - nonpharmacologic, 3507
    - on-demand therapies for "off" periods, 3504
    - stem cell therapy, 3405–3406, 3505–3506
    - surgical, 3505
- *PARN* gene mutations, 3820t
- Paromomycin
  - actions of, 1750–1751
  - adverse effects of, 1741t, 1751, 1785
  - for amebiasis, 1751, 1757, 1757t
  - for cutaneous leishmaniasis, 1787
  - for *D. fragilis* infections, 1808
  - for *Giardia* infections, 1805
  - indications for, 1741t
  - pharmacology of, 1750–1751
  - in pregnancy and lactation, 1741t
  - for visceral leishmaniasis, 1784–1785
- Paronychia, 390, 421, 1312, 1708
- Parotid gland, 610, 2928
- Parotitis, 266, 1647–1648, 1648f, 2878t, 2928
- Paroxetine
  - adverse effects of, 3188, 3659t, 3667
  - for anxiety disorders, 3661
  - for depression, 85, 504, 3659t, 3666
  - drug interactions of, 485t, 489t, 492
  - for menopausal symptoms, 3143
  - metabolism of, 485t, 492
  - for panic disorder, 3658
  - pharmacology of, 3667
  - for PTSD, 3663
- Paroxysmal cold hemoglobinuria, 801t, 802
- Paroxysmal dyskinesias, 3517
- Paroxysmal exertion-induced dyskinesia, 3517
- Paroxysmal hemicrania, 115, 115t, 3471t, **3472**
- Paroxysmal kinesigenic dyskinesia (PKD), 3517
- Paroxysmal nocturnal dyspnea, 1982
- Paroxysmal nocturnal hemoglobinuria, **804**
  - clinical features of, 323, 450, 804, 810
  - diagnosis of, 804
  - global considerations in, 804
  - natural history of, 804
  - pathophysiology of, 804, 805f
  - treatment of, 805–806, 806f
- Paroxysmal nonkinesigenic dyskinesia (PNKD), 3517
- Paroxysmal supraventricular tachycardia (PSVT). *See also* Supraventricular tachyarrhythmias
  - accessory pathways and, 1940–1942. *See also* Wolff-Parkinson-White syndrome
  - AV nodal reentry tachycardia, 1932t, 1937f, 1939–1940, 1939f, A8
  - definition of, 93
  - junctional tachycardia, 1940
  - mechanisms of, 1937f
  - orthodromic AV reciprocating tachycardia, 1941
  - preexcited tachycardias, 1941
  - treatment of, 1941–1943, 1943f
- PARP (poly-ADP-ribose polymerase), 537, 537f, 3937
- PARP (poly-ADP-ribose polymerase) inhibitors
  - action and targets of, 528, 533t, 537f, 571t, 574
  - adverse effects of, 571t
  - interactions and issues, 571t
  - for ovarian cancer, 537, 712
  - for pancreatic cancer, 671, 675
  - for prostate cancer, 703
- Parrotfish poisoning, 3726
- Parrot's pseudoparalysis, 1058
- Parsonage-Turner syndrome, 3614
- Pars planitis, 225
- Partial anomalous pulmonary venous return, 2071, 2071f
- Partial gonadal dysgenesis, 3098
- Partial response (PR), to chemotherapy, 555
- Partner notification, 1111, 1260
- Parulis, 260
- Parvoviridae, 1478f, 1480t, 1481, 1522, 1536
- Parvovirus 4, 1525
- Parvovirus B19 infections, **1522**
  - aplastic crisis in, 792, 814
  - clinical features of, 1522–1524, 1523f. *See also* Erythema infectiosum (fifth disease)
  - diagnosis of, 1524, 1524t
  - epidemiology of, 1522
  - hydrops fetalis, 1524
  - myocarditis in, 2012
  - pathogenesis of, 1522, 1523
  - polyarthritis in, 1059
  - in pregnancy, 1524, 3904
  - prevention of, 1524
  - pure red cell aplasia, 814f, 1524. *See also* Pure red cell aplasia
  - pure red cell aplasia in, 814
  - transfusion-transmitted, 912, 913t
  - treatment of, 815, 1524
- Parvus et tardus pulse, 285, 1865
- PASCAL device, 2130f, 2131
- Pasireotide, 3012, 3059
- Passenger mutation, 526
- Passive euthanasia, 89t
- Passive immunity, 996–997
- Passive leg raise test, 2314
- Passive transport, 2366
- Pasteur effect, 276
- *Pasteurella* spp. infections
  - in cat-bite wound, 1052, 1143, 1267
  - cellulitis, 1052
  - clinical features of, 1056, 1143, 1267
  - in dog-bite wound, 1143, 1267
  - epidemiology of, 1267
  - treatment of, 1052, 1054t, 1145, 1145t, 1267, 1267t
- Pastia's lines, 146, 1209, 1209f
- Patch (skin lesion), 377t
- Patched gene, 431
- Patch testing, 382, 2215
- Patellar tendinitis, 2974
- Patent ductus arteriosus (PDA), 277, 288, 288f, 1859, 2066, 2072, 2072f
- Patent foramen ovale, 2130, 3445t, 3452, 3757
- Paternal age, 3118, 3151, 3171, 3784, 3790
- Paternalism, in behavioral economics, 3915
- Pathogen-associated molecular patterns (PAMPs)
  - characteristics of, 2756
  - masking of, 971
  - in response to viruses, 1482
  - in sepsis/septic shock, 2317, 2318f
- Patient
  - autonomy, 71
  - expectations of, 7
  - feedback from, 4
  - hospital experience of, 3
  - Internet use by, 7
  - satisfaction, 66, 66f
- Patient care, 2–5
- Patient-centeredness, 69
- Patient-controlled analgesia, 98
- Patient Health Questionnaire-2 (PHQ-2) Depression Screen, S8
- Patient-physician relationship. *See* Physician-patient relationship
- Patiromer, 361, 373
- Patisiran, 900
- Patrick's test (flexion, abduction, and external rotation), 121t
- Pattern recognition, 21
- Pattern recognition receptors (PRRs). *See also* Toll-like receptors (TLRs)
  - definition of, 2756
  - of inflammasomes, 972–973
  - in innate immune system, 971, 971t, 2758, 2758t, 2763
  - ligands of, 971, 971t, 2759t
  - pathogen avoidance of, 971
  - in sensing of cytosolic infection, 971–972
  - in sepsis/septic shock, 2317, 2318f
  - signaling pathways of, 2759f
- Patterson-Kelly syndrome, 642
- *PAX3-FOXO1* gene mutations, 521t
- *PAX3* gene mutations, 397, 3783t
- PAX5, 858
- *PAX8* gene mutations, 3021, 3028t, 3045
- *PAX8-PPARG* gene mutations, 521t
- Pay-for-performance system, 45
- Pazopanib
  - actions and targets of, 548, 570t, 573
  - adverse effects of, 570t, 573, 753t, 754
  - interactions and issues, 570t
  - for NETs, 684t, 686
  - for RCC, 690t
  - for soft tissue sarcomas, 729
- PBC. *See* Primary biliary cholangitis (PBC)
- PBPs (transpeptidase enzymes), 1183
- *PC1* gene mutations, 3112t
- PCA2 antibodies, 247t
- PCC (prothrombin complex concentrate), 930, 933, 935, 954
- PCI. *See* Percutaneous coronary interventions (PCI)

PCNSL. *See* Primary central nervous system lymphoma (PCNSL)
PCOS. *See* Polycystic ovarian syndrome (PCOS)
PCP. *See* Phencyclidine (PCP); *Pneumocystis* pneumonia (PCP)
PC-PTSD (Primary Care PTSD) Screen, S8
PCR. *See* Polymerase chain reaction (PCR)
PCSK1 (prohormone convertase 1), 3185, 3185t
PCSK1 gene mutations, 3112t
PCSK9 deficiency, 3241t, 3247
PCSK9 gene mutations, 494, 3242, 3247
PCSK9 inhibitors, 494, 3243, 3244t, 3245, 3251, 3257
PCT. *See* Porphyria cutanea tarda (PCT)
PCT (procalcitonin), 1025–1026
PCV (pressure-control ventilation), 2305–2306, 2305t, 2306f. *See also* Mechanical ventilation
PD. *See* Parkinson's disease (PD)
PD-1, 549, 550f, 560, 660, 2785–2787, 2786f
PD-1 antibodies. *See also specific drugs*
  action and targets of, 534t, 549, 550, 550f, 560, 2783, 2786f, 2787–2788
  adverse effects of, 605
    adrenal insufficiency, 593
    autoimmune endocrine disease, 740
    colitis, 2561
    cutaneous, 392, 420, 422
    endocrine dysfunction, 756, 3092
    hepatotoxicity, 2672
    myocarditis, 2015
    myositis, 2917
    Sjögren's disease, 2878
  for anal cancer, 659
  for bladder cancer, 694
  for colorectal cancer, 657
  for esophageal cancer, 645
  for gastric adenocarcinoma, 649
  for head and neck cancer, 614
  indications for, 2787t
  for melanoma, 432, 604, 604t, 605
  microbiome and effectiveness of, 3836
  for pancreatic cancer, 675
PDA (patent ductus arteriosus), 277, 288, 288f, 1859, 1866, 2066, 2072, 2072f
PDC (pyruvate dehydrogenase complex), 2711, 2790t
PDCD10 gene mutations, 3457
PDGF (platelet-derived growth factor), 546f, 2773
PDGFRA gene mutations, 650, 651, 728
PDGF (platelet-derived growth factor) receptors, 546f, 567
PDGFR gene mutations, 717
PDGFR-mutated eosinophilia, 879–880
PD-L1
  in breast cancer, 639
  in esophageal cancer, 645
  in HCC, 660
  in head and neck cancer, 614
  in Hodgkin's lymphoma, 869, 870
  in immune system evasion by cancer, 549, 550f, 560
  in lung cancer, 619, 620, 629, 629f
PD-L1 antibodies. *See also specific drugs*
  action and targets of, 534t, 549, 550, 550f, 560, 2787t, 2788
  adverse effects of, 593, 740, 2015, 2672, 2878, 2917, 3092
  for bladder cancer, 694
  indications for, 2787t
  for Merkel cell carcinoma, 609
PE. *See* Pulmonary embolism (PE)
Peanut allergy, 2814
Pearson's syndrome, 3811t, 3812
Peau d'orange skin, 1210, 2184
Pectoriloquy, 2200
Pectus carinatum (pigeon chest), 1860, 3332
Pectus excavatum (funnel chest), 1860, 3332
Pedal edema, 2200
PEDF gene mutations, 3326
Pediculiasis, 3731–3732
Pediculicides, 3732
Pediculus capitis (head lice), 3731–3732
Pediculus humanus (body lice), 1445, 1458, 3731–3732
Pedigree analysis
  in family history, 3795
  in genetic counseling, 3800–3801, 3801f
  in Mendelian disorders, 3790, 3790f
  in mtDNA diseases, 3807, 3807f
  standard symbols for, 3790f
Peduncular hallucinosis, 3432, 3432f
PEEP (positive end-expiratory pressure), 2302, 2304, 2305f. *See also* Mechanical ventilation
Pegcetacoplan, 805f, 806, 806f
Pegfilgrastim, 575t
Pegloticase, 800, 2960
Pegunigalsidase, 3602
Pegvaliase, 3370
Pegvisomant, 3010
Pegylated interferon-α2a (PEG IFN-α2a)
  adverse effects of, 1491
  for chronic HBV infection, 1488t, 1491–1492, 2676t, 2678–2679, 2680f, 2682t
  for chronic HCV infection, 2686–2687
  for chronic HDV infection, 1494, 2684
  contraindications to, 1491
  for myelofibrosis, 823
  for polycythemia vera, 820–821
Pegylated interferon-α2b (PEG IFN-α2b), 1488t, 2678–2679
Pegylated interferon-β-1a (PEG IFN- β-1a), 3587t
PELD system, in liver transplantation, 2633
Pel-Ebstein fever, 136, 869
Pelger-Hüet anomaly, 442f, 453, 455f
Peliosis hepatitis, 2668
Pelizaeus-Merzbacher disease, 3397, 3406, 3525
Pellagra
  in carcinoid syndrome, 682
  clinical features of, 400, 2605t, 2608, 3484, 3612, A15
  risk factors for, 2608
  treatment of, 2608, 3612
Pelvic abscess, 1376
Pelvic congestion, 3137
Pelvic floor dysfunction, 314
Pelvic floor retraining, 314, 335
Pelvic fractures, 3292
Pelvic inflammatory disease (PID), **1104**
  abdominal pain in, 110
  anaerobic, 1376
  clinical features of, 1105
  C. trachomatis, 1471
  diagnosis of, 1105–1106
  epidemiology of, 1105
  etiology of, 1097t, 1104–1105, 1467, 3136
  follow-up for, 1107
  pelvic pain in, 3136
  prevention of, 1107
  prognosis of, 1107
  treatment of, 1106–1107, 1106t
Pelvic organ crosstalk, 331
Pelvic pain, **3135**
  acute, 3136
  approach to the patient, 3135–3136
  in cervical cancer, 713
  chronic, 3136–3137
  etiology of, 3135, 3136t
  in IC/BPS, 332, 334
  in pelvic inflammatory disease. *See* Pelvic inflammatory disease (PID)
Pelvic torsion (Gaenslen's test), 121t
PEM (postexertional malaise), 3649, 3650t
Pemberton's sign, 3041
Pembrolizumab
  action and targets of, 534t, 559f, 560, 2787t
  adverse effects of, 596, 753t, 2351, 2377, 2672
  for bladder cancer, 693, 694
  for breast cancer, 637, 638
  for carcinoma of unknown primary, 734
  for cholangiocarcinoma, 668, 669
  for endometrial cancer, 715
  for esophageal cancer, 644, 645
  for gastric cancer, 649
  for HCC, 665f, 666
  for head and neck cancer, 614
  for Hodgkin's lymphoma, 870
  indications for, 2787t
  for lung cancer, 625, 625t, 629, 630t
  for melanoma, 603, 604, 604t, 605
  for Merkel cell carcinoma, 609
  for NETs, 686
  for pancreatic cancer, 675
  for PML, 1117
  for prostate cancer, 703
  for RCC, 689, 690t
  for triple-negative breast cancer, 641
Pemetrexed
  actions of, 565
  adverse effects of, 562t, 565
  interactions and issues, 562t
  for lung cancer, 624, 629
  for PCNSL, 721
Pemigatinib, 533t, 563t, 569t, 572, 833t
Pemivibart, 1489, 1543
Pemoline, 85
Pemphigoid
  bullous, **411**
    clinical features of, 410t, 411, 411f, A5
    drug-induced, 420
    histology of, 410t, 412
    immunopathology of, 410t, 412, 2790t
    treatment of, 412
  esophageal involvement/dysphagia in, 296
  gestational, 410t, 412
  mucous membrane, 263, 263t, 410t, **413**
Pemphigus
  drug-induced, 420
  paraneoplastic, 401, 410t, **411,** 2791t
Pemphigus foliaceus, **410**
  clinical features of, 410, 410t
  drug-induced, 411
  histology of, 410t
  immunopathology of, 410, 410t, 2790t
  treatment of, 411
Pemphigus vulgaris, **409**
  clinical features of, 401, 409–410, 410t, 411f, A5
  diagnosis of, 409
  histology of, 410t
  immunopathology of, 409–410, 410t, 2790t
  oral manifestations of, 263, 263t, 411f
  treatment of, 410
penA, 1255
penB, 1255
Penciclovir, 1485, 1503t
Pendred's syndrome, 248, 3022, 3028t, E36t
Pendrin, 3022, 3028t
Penetrance, 3789, 3800
Penetrating injury, cardiac, 2088–2089
Penetrating radiation, S6
D-Penicillamine
  adverse effects of
    autoimmune myasthenia gravis, 3627
    cutaneous, 388, 411, 420, 3340
    lupus syndrome, 2940t
    myopathy, 2940t, 3648t
    optic neuropathy, 228

    - pyridoxine and, 2609
    - vasculitis, 2940t
    - zinc deficiency, 2613
  - for cystine stone prevention, 2453
  - for cystinuria, 3373
  - for mercury poisoning, 3701t
  - for systemic sclerosis, 2874
  - for Wilson's disease, 3340, 3518
- Penicillin(s)
  - for actinomycosis, 1365, 1366t
  - actions of, 1166, 1182t
  - adverse effects of
    - acid-base disorders, 374
    - cross-sensitivity in, 427–428
    - cutaneous immunologic, 419
    - desensitization to, 428–429, 2818
    - erythema nodosum, 405
    - erythroderma, 393
    - hemolytic anemia, 489, 804
    - hypokalemia, 355
    - neutropenia, 457
    - pemphigus, 411
    - platelet dysfunction, 926
    - rash, 401, 422
    - renal, 2377
    - seizure, 3414, 3414t
    - serum sickness–like reactions, 419
    - toxic epidermal necrolysis, 401
    - urticaria, 422
    - vasculitis, 2907
  - for anaerobic bacterial infections, 1379
  - for anthrax, S4
  - for asymptomatic pharyngeal GAS carriage, 1209
  - for bite-wound infections, 1145t
  - for *Capnocytophaga* infections, 1267t
  - for clostridial myonecrosis, 991t, 1054
  - for enterococcal infections, 1219t
  - for epidural abscess, 991t
  - for erysipelas, 1208t
  - for gas gangrene, 1054t, 1242
  - for GAS infections, 1208t
  - for group B *Streptococcus* infections, 1213
  - for group C or G streptococcal infections, 1212–1213
  - for HACEK infections, 1267t
  - indications for, 1173t, 1174–1175
  - for infective endocarditis, 1041, 1043t, 1044
  - for leptospirosis, 1444, 1444t
  - for meningococcemia, 991t
  - for necrotizing fasciitis, 991t, 1208t
  - for pharyngitis, 1208t
  - for pneumonia, 1208t
  - in pregnancy and lactation, 1169t
  - resistance to
    - in anaerobic bacteria, 1379
    - in *C. perfringens,* 1242–1243
    - mechanisms of, 1181–1182, 1182t, 1283, 1283t
    - in meningococci, 1249
    - in *N. gonorrhoeae,* 1100f, 1255
    - in pneumococci, 1026, 1123, 1192
    - in staphylococci, 1204
  - for rheumatic fever, 2857
  - for streptococcal pharyngitis, 258, 258t
  - for syphilis, 1435
  - for tetanus, 1231
  - for toxic shock syndrome, 1208t
- Penicillin-binding proteins, 1166, 1181, 1188f, 1193, 1204, 1220
- Penicillin G
  - for diphtheria, 1224
  - for endemic treponematoses, 1439
  - for group B *Streptococcus* infections, 1213, 1214
  - indications for, 1173t, 1175
  - for *L. monocytogenes* infections, 1229
  - for Lyme disease, 1453, 1453f
  - for meningitis, 1121t, 1122, 1122t
  - for neurosyphilis, 1434t, 1435
  - for nongonococcal bacterial arthritis, 1057
  - for osteomyelitis, 1064t, 1065t
  - for pharyngitis, 1208t
  - for pneumococcal infections, 1192
  - for relapsing fever, 1447, 1447f
  - resistance to, 1173t
  - for staphylococcal infections, 1203t
  - for syphilis, 1434–1435, 1434t
  - for syphilis prophylaxis, 1434
  - for syphilitic meningitis, 1127
- Penicillin V, 1224, 2858
- *Penicillium marneffei* infection (penicilliosis). *See Talaromyces marneffei* infection
- Penile cancer, 1527
- Penile prosthesis, 3160
- Penile tumescence, 3155
- Penile warts. *See* Condyloma acuminata
- Pentamidine
  - actions of, 1751, 2370
  - adverse effects of
    - acid-base disorders, 373
    - bronchospasm, 1734t
    - frequent, 1745t, 1795
    - hyperkalemia, 359
    - hypotension, 1734
    - hypovolemia, 346
    - nephrotoxicity, 2377
    - occasional, 1745t, 1795
    - pancreatic, 1601
    - rare, 1745t
    - renal, 1601
  - for cutaneous leishmaniasis, 1787
  - for human African trypanosomiasis, 1795, 1796f, 1796t
  - for PCP, 1734, 1734t
  - for PCP prophylaxis, 1593t, 1596, 1734t
  - pharmacology of, 1751
  - in pregnancy and lactation, 1745t
- Pentasa, 2564, 2564t
- Pentastomiasis, 3733
- Pentavalent antimonials, 1741t, 1746, 1784, 1787
- Pentazocine, 2940t
- Pentobarbital, 3426f, 3713t
- Pentosan polysulfate sodium (PPS), 335
- Pentostatin, 595
- Pentoxifylline
  - for alcohol-associated liver disease, 2710
  - for intermittent claudication, 2174
  - for leprosy reactions, 1412
- Penumbra, 2342
- PEO. *See* Progressive external ophthalmoplegia (PEO)
- PEP. *See* Postexposure prophylaxis (PEP)
- PEPFAR (United States President's Emergency Plan for AIDS Relief), 3843, 3855
- Peppermint oil, 2575, 3928
- Pepsin, 2518
- Peptic ulcer disease, **2514**
  - clinical features of, 103t, 106, 112t, 300, 2522
  - complications of
    - bleeding
      - endoscopic treatment of, 2478, 2481, 2484–2487f
      - incidence of, 2522
      - recurrent, prevention of, 317, 2481–2482, 2486f, V5
      - treatment of, 317
    - gastric outlet obstruction, 2523
    - perforation, 2522–2523
  - conditions related to
    - gastritis. *See* Gastritis
    - Ménétrier's disease, **2539**
    - stress-related mucosal injury, 318, 2297, 2537
    - Zollinger-Ellison syndrome. *See* Zollinger-Ellison syndrome (ZES)
  - deaths from, 2514
  - diagnosis of, 2523
    - barium studies, 2523, 2523f
    - endoscopic, 2467f, 2481, 2491, 2523, 2524f
    - with hemorrhage, 2481, 2484–2485f, 2484f
    - *H. pylori* detection, 2523, 2524t
  - differential diagnosis of, 2523
  - duodenal ulcers in. *See* Duodenal ulcers
  - epidemiology of, 2514, 2518
  - etiology of, 300, 2521, 2522t
  - gastroduodenal mucosal defense and, 2517–2518
  - genetic considerations in, 2521–2522
  - *H. pylori*–related. *See Helicobacter pylori* infections, in peptic ulcer disease
  - NSAID-induced, 317, **2520,** 2521f, 2529–2531, 2530t
  - pathology of, 2518
  - pathophysiology of, 2519
  - in polycythemia vera, 820
  - recurrence of, 2533
  - refractory, 2531
  - risk factors for, 2518
  - treatment of, **2523,** 2524t
    - antacids, 2523–2524
    - approach to, 2531, 2531f
    - bismuth-containing preparations, 2526
    - H_2 receptor antagonists, 2524
    - *H. pylori* eradication. *See Helicobacter pylori* infections, treatment of
    - NSAID-induced, 2529–2531, 2530t
    - PPIs, 2524–2526, 2526f
    - prostaglandin analogues, 2526
    - in refractory disease, 2531
    - sucralfate, 2526
    - surgical, 2532–2534, 2532f
  - ZES and, 2534. *See also* Zollinger-Ellison syndrome (ZES)
- Peptide(s)
  - antimicrobial, 2758t
  - ovarian, 3128
- Peptide receptor radioligand therapy, 685
- Peptide YY, 3182, 3182f
- Peptidoglycan, 1166
- *Peptostreptococcus* spp. infections, 1144, 1206t, 1372, 1374, 1377. *See also* Anaerobic bacterial infections, mixed
- *PER1* gene mutations, 3947t
- *PER2* gene mutations, 218, 3947t, 3954
- *PER3* gene mutations, 3947t, 3952
- Peramivir, 1157, 1488, 1488t, 1537, 1550
- Perampanel, 3422t, 3424
- Perchlorate, 3040
- Percussion
  - of chest, 46, 270, 2200
  - of spleen, 480
- Percutaneous coronary interventions (PCI), **2126, A11**
  - for adult congenital heart disease, 2130
  - antiplatelet agents in, 2110, 2127, 2128
  - aortic valve replacement. *See* Transcatheter aortic valve implantation (TAVI)
  - aortic valvuloplasty, 2040
  - atherectomy, 2128
  - atrial septic defect repair, 2130, A11
  - balloon angioplasty and stenting, 2127–2128, 2127f
  - for cardiogenic shock, 2327, 2328f
  - circulatory support during, 2132
  - complications of, 2118t, 2128–2129, A11
  - distal protection device for, 2128, A11
  - embolic protection device for, 2128

Percutaneous coronary interventions (PCI) (*Cont.*):
  imaging for, 2128
  indications and patient selection for, 2129
  intimal dissection in, 2127, 2127f
  for left atrial appendage closure, A11
  vs. medical therapy in acute coronary syndromes, 2130
  mitral balloon commissurotomy, 2051
  mitral balloon valvotomy, 2051f
  mitral valve repair and replacement, 2056–2057, 2056–2057f
  multivessel, A11, A11v
  for myocardial infarction
    NSTE-ACS, 2111, 2111f
    rescue, 2120
    STEMI, 2119–2120, 2119f, A11
    time from symptom onset to, 2119
  for paravalvular leak, A11
  prophylactic, before noncardiac surgery, 3906f, 3907
  for revascularization in stable ischemic heart disease, **2103**
    of bifurcation lesions, A11, A11v
    vs. CABG, 2105, 2105f, 2129
    of chronic total occlusion, A11
    efficacy of, 2104
    fractional flow reserve-guided, A11, A11v
    indications and patient selection for, 2103–2104
    orbital atherectomy and shockwave lithotripsy, A11
    restenosis after, 2104
    risks of, 2104
    saphenous vein graft with distal protection, 2128, A11
    stent thrombosis in proximal LAD lesion following, A11
    in unprotected left main coronary artery, A11, A11v
  steerable guidewire in, 2127
  STEMI related to, 2128
  success of, 2128, 2129f
  techniques for, 2127–2128
  transluminal coronary angioplasty, 2126
  tricuspid valve repair and replacement, 2062, 2062f
  valvular heart disease, 2130–2131
Percutaneous interventions
  for carotid atherosclerosis, 3451–3452
  for cerebral vasospasm, 3461, 3461f
  coronary. *See* Percutaneous coronary interventions (PCI)
  for gastrostomy tube placement, 2469
  for ischemic stroke, 3441–3442
  mechanical thrombectomy, 2175, 3441–3442, A11
  peripheral artery, 2131, 2132f, 2174
  for pulmonary embolism, 2132
  for refractory hypertension, 2132
Percutaneous tibial nerve stimulation, 2584
Percutaneous transhepatic cholangiogram, 2732, 2736t
Percutaneous umbilical blood sampling (PUBS), 3797
Perennial nonallergic rhinitis, 2810
Perennial nonallergic rhinitis with eosinophilia syndrome (NARES), 2810
Perfringolysin (PFO), 1242
Perfusion MRI, 3390
Pergamino virus, 1661t, 1674
Pergolide, 2179, 3710t
Perianal disease. *See* Anorectal disorders
Periapical disease, 260, 1374
Periappendicitis, 1105
Peribunyaviridae
  clinical features of, 1667
  clinical syndromes of, 1666t
  reservoirs and vectors of, 1656, 1658–1659t
  structure of, 1478f, 1479, 1480t
Pericardial effusion
  after cardiac injury, 2088
  cardiac imaging in, 1895–1898, 1898f
  chronic, 2083
  ECG in, 1874, 1875f, A7
  echocardiography in, 2080f, A9
  in eosinophilic granulomatosis with polyangiitis, A14
  etiology of, 1895
  malignant, 587
  in pericarditis, 2079
  physical examination in, 47t, 1864
  tuberculous, 2084
Pericardial friction rub, 1867, 2079, 2115, 2125
Pericardial knock, 327
Pericardial tamponade, 587, 1867
Pericardiocentesis, 2082, 2087
Pericarditis, **2078**
  abdominal pain in, 111, 112t
  acute
    clinical features of, 2078–2079, 2092, 2115, 2125
    CMR in, 2081f, A9
    differential diagnosis of, 2082–2083
    ECG in, 107, 2079, 2080f, A7
    echocardiography in, 2079, 2080f
    postcardiac injury, 2082
    treatment of, 2079–2080
    viral/idiopathic, 2082
  cardiac tamponade in. *See* Cardiac tamponade
  chest pain in, 103t, 104, 106, 2078–2079
  in CKD, 2083, 2392
  classification of, 2079t
  in collagen vascular disease, 2083
  constrictive
    clinical features of, 2081t, 2083–2084
    CMR in, 1899f, 1903v, A9
    differential diagnosis of, 1905–1906, 1906t, 2081, 2081t, 2084
    dyspnea in, 269, 269t
    ECG in, 2081t, 2083
    echocardiography in, 2081t, 2083
    hemodynamic profile in, 2083
    pathophysiology of, 2082f, 2083
    pulse pressure in, 2083
    treatment of, 2084
  differential diagnosis of, 2082–2083
  etiology of, 2079t
  in heart transplant, 1159t
  infectious causes of, 1548, 1634t, 1635, 2083
  in neoplastic disease, 2083
  pericardial effusion in, 2079, 2080f
  pericardial friction rub in, 1867, 2079
  pyogenic, 2083
  radiation-induced, 554, 587, 753, 2083
  recurrent, 2083
  in rheumatoid arthritis, 2840
  in SLE, 2083, 2830
  in STEMI, 2125
  subacute effusive-constrictive, 2084
  tuberculous, 1389, 2082, 2084
Pericardium
  congenital defects of, 2078
  metastases to, 587
  structure and function of, 2078
Pericytes, 1844, 1845f
Perihepatitis (Fitz-Hugh-Curtis syndrome), 1105, 1471
Perihilar cholangiocarcinoma, 667f, 669
Perilymphatic fistula, 247
Perimenopause, 3142–3143, 3142f
Perineal descent, 304f, 313
Perinephric abscess, 1075, 1094
Perinuclear ANCA (pANCA), 2894
Periodic acid–Schiff stain, 389, S11
Periodic fever with aphthous stomatitis, pharyngitis, and cervical adenitis (PFAPA), 2762t, 2937
Periodic health evaluations, 9–10
Periodic hypothermia syndrome, 3002
Periodic leg movements, 3517–3518
Periodic limb movement disorder, 217
Periodic paralysis
  hyperkalemic, 359, 3635, 3646, 3804t
  hypokalemic. *See* Hypokalemic periodic paralysis (HypoKPP)
  thyrotoxic, 3647
Periodontal disease (periodontitis)
  abscess in, 1374
  AIDS-related, 261
  clinical features of, 261, A3
  in Ehlers-Danlos syndrome, 3330
  etiology of, 260–261, 1374
  localized juvenile, 261
  in older adults, 267
  prevention of, 261, 463
  rheumatoid arthritis and, 2843
  risk factors for, 261
Periodontium, 260
Periostitis, in lung cancer, 619
Peripartum cardiomyopathy, 2017–2018, 3900
Peripharyngeal space infections, 1374
Peripheral artery disease, **2162**
  cardiovascular disease and, 943, 2172–2173
  clinical features of, 2172
  diagnosis of, 1863, 1863f, 2172, 2173f
  pathophysiology of, 2172
  preoperative risk assessment in, 2174
  prognosis of, 2172–2173
  treatment of
    exercise program, 2173–2174
    pharmacologic, 2173, S13
    revascularization, 2131, 2132f, 2174
Peripheral artery pulses, 1862
Peripheral blood smear, **437**. *See also specific disorders*
  Howell-Jolly bodies in. *See* Howell-Jolly bodies
  normal, 438f, 442f, 445f, 922f
  red cell defects in, 445–447
  red cell fragmentation in, 446f
  reticulocytes in, 438f, 447f
  spur cells in, 446f
  target cells in, 440f, 446f
Peripheral clocks, 3944t
Peripheral cyanosis, 278t, 279, 1859. *See also* Cyanosis
Peripherally inserted central catheter (PICC), 2161
Peripheral nerve(s)
  components of, 93f
  cutaneous fields of, 174f
  hyperexcitability, 751
  pain mechanisms, 95–96
Peripheral neuropathy, **3594**. *See also* Plexopathies; Polyneuropathy
  in acute intermittent porphyria, 3346
  in alcohol use disorders, 3676
  in amyloidosis, 3603–3604. *See also* Amyloidosis
  approach to, **3595,** 3595f
    electrodiagnostic studies in, 3597, 3597t
    key questions in, 3595–3597, 3596t
    laboratory studies in, 3597–3598
    nerve biopsy in, 3598
    pattern recognition in, 3595–3597, 3596t
    sensory involvement in, 3596–3597
    skin biopsy in, 3598
  autonomic dysfunction and, 3549–3550
  in celiac disease, 3605

  - in chronic liver disease, 3605–3606
  - in CKD, 2393–2394, 3605
  - in CMV infections, 3607
  - in cobalamin deficiency, 783
  - in COVID-19 disease, 3606
  - critical illness polyneuropathy, 2351, 3606
  - cryptogenic sensory and sensorimotor polyneuropathy, 3613
  - in diabetes mellitus. *See* Diabetic neuropathy
  - in diphtheria, 1224, 3606
  - drug-induced/toxic, **3608,** 3609–3610t
    - amiodarone, 3610
    - antiretrovirals, 3611
    - arsenic, 3611
    - cancer chemotherapy, 726, 726t, 756, 3608, 3608t
    - chloroquine, 3608
    - colchicine, 3610
    - hexacarbons, 3611
    - hydroxychloroquine, 3608
    - isoniazid, 3611
    - lead, 3611
    - mercury, 3611
    - pyridoxine, 3611
    - thalidomide, 3610
    - thallium, 3611
  - in EBV infections, 3607
  - in hepatitis, 3607
  - hereditary, 3598, 3599–3600t, 3602t
    - adrenoleukodystrophy. *See* Adrenoleukodystrophy
    - adrenomyeloneuropathy, 3602
    - cerebellar ataxia, neuropathy, and vestibular areflexia syndrome (CANVAS), 3537, 3603
    - Charcot-Marie-Tooth disease. *See* Charcot-Marie-Tooth (CMT) disease
    - in Fabry's disease, 3601–3602. *See also* Fabry's disease
    - familial amyloid polyneuropathy, 3603
    - hereditary neuralgic amyotrophy, 3601
    - hereditary neuropathy with liability to pressure palsies, 3601
    - hereditary sensory and autonomic neuropathy, 3601
    - in porphyria, 3602–3603. *See also* Porphyria(s)
    - in Refsum's disease, 232, 3602
    - sorbitol dehydrogenase deficiency with peripheral neuropathy, 3601
    - in Tangier disease, 1860, 3241t, 3248, 3602
  - in herpes zoster, 3607
  - in HIV infection, 1609–1610, 3606–3607
  - in hypereosinophilic syndrome, 3605
  - in hypothyroidism, 3604
  - in IBD, 3605
  - in leprosy, 1410, 3606
  - in Lyme disease, 3523, 3606
  - in malabsorption disorders, 2549t
  - malignancy-related, 3607–3608
  - in mixed connective tissue disease, 3605
  - mononeuropathies, 3613–3614
  - in nutritional deficiencies, 3611–3612
  - orthostatic hypotension and, 159
  - paraneoplastic, 750–751
  - plexopathies, 3614–3616
  - in pregnancy, 3903
  - radiculopathies, 3614, 3614t
  - in rheumatoid arthritis, 3605
  - in sarcoidosis, 2927, 3605
  - in Sjögren's disease, 3605
  - in SLE, 3605
  - in systemic sclerosis, 3605
  - weakness in, 170, 171, 3595–3596, 3596t
- Peripheral node addressin (PNad), 2781
- Peripheral sensitization, 93
- Peripheral T-cell lymphoma, 860t, 862, 866–867
- Peripheral tolerance, 2789, 2789t
- Peripherin, 232
- Periplasmic space, 1165
- Peristalsis, 293
- Peristaltic sounds, 112
- Perisylvian network, 201
- Peritoneal carcinomatosis, 328, 328f, 735
- Peritoneal dialysis, **2400**
  - access for, 2401, 2402
  - complications of, 2401
  - continuous ambulatory, 1072, 2384–2385, 2400
  - continuous cyclic, 2398, 2400
  - dose for, 2401
  - for hypercalcemia, 3285
  - principles of, 2400
  - solutions for, 2400–2401
- Peritoneal equilibrium test, 2401
- Peritoneal inflammation, 110
- Peritoneal lavage, as rewarming strategy, 3770
- Peritoneal papillary serous carcinoma, 735
- Peritonitis, **1069**
  - abdominal pain in, 110, 112, 1070
  - acute, **2597**
  - ascites and, 329–330, 2716
  - aseptic, 2597
  - clinical features of, 329, 1070, 2597
  - diagnosis of, 1070, 1070f
  - etiology of, 1070, 1286, 2597, 2597t
  - hypovolemia in, 346
  - pathophysiology of, 1069–1070, 1069f
  - in peritoneal dialysis, 1072, 2401
  - postoperative, 1072
  - prevention of, 1071
  - primary (spontaneous bacterial), 329–330, 1070–1071, 2597, 2716
  - prognosis of, 1071, 2597
  - recurrence of, 1071, 1180t
  - secondary, 329, 1071–1072, 2597
  - treatment of, 1070–1072, 2597
  - tuberculous, 1389
  - tumor markers in, 504t
- Peritonsillar abscess, 259, 1374
- Periventricular (diffuse) white matter disease (leukoaraiosis), 3448
- Perleche, 380f, 390
- Permanent junctional reciprocating tachycardia, 1941
- Permethrin, 3729, 3732
- Permissive underfeeding, 2620
- Pernicious anemia
  - APS-1 and, 3090t
  - APS-2 and, 3090t, 3091t
  - autoantibodies in, 2790t, 2825, 2825t, 3091t
  - cobalamin absorption in, 786
  - diagnosis of, 786
  - epidemiology of, 785–786
  - gastritis in, 2538
  - juvenile, 786
  - oral manifestations of, 65
- Pernio (chilblains), 146, 2177f, **2178,** 3771–3772, A1
- Peroneal neuropathy, 3614
- Peroral endoscopic myotomy (POEM), 2469, 2476f, 2493, V5
- Peroral endoscopic tumorectomy (POET), 2469, 2477f, 2493, V5
- Peroxisomal catalase, 2709
- Peroxisome proliferator-activated receptor γ (PPARγ), 2564, 3214
- Perphenazine, 3673, 3673t, 3674
- PERRLA, 221
- Persistent light reaction, 434
- Persistent müllerian duct syndrome, 3100
- Persistent vegetative state, 187, 193
- Personality disorders, **3671**
- Personalized medicine. *See* Precision medicine
- Person-centered care, of older adults, 3876–3877, 3878t
- Persons who inject drugs (PWID)
  - bacterial arthritis in, 1056
  - *Candida* infections in, 1059
  - endocarditis in, 1037, 1039, 1047, 1200
  - HIV infection in, 3692
  - HIV transmission in, 1562t, 1564, 1626
  - HTLV infections in, 1556
  - tetanus in, 1230
  - wound botulism in, 1234, 1237
- Pertactin, 1278, 1280
- Perthes test, 2180
- Pertussis, **1276**
  - clinical features of, 1278–1279, 1278f
  - complications of, 1279
  - diagnosis of, 1279
  - differential diagnosis of, 1279
  - epidemiology of, 1277–1278, 1277f, 1278f
  - global considerations in, 1277, 1277f
  - immune response in, 1278
  - infection control measures, 1280
  - pathogenesis of, 1278. *See also Bordetella pertussis*
  - prevention of, 1280
  - treatment of, 1279–1280, 1280t
- Pertussis toxin, 1277, 1278
- Pertussis vaccine, 1278, 1280. *See also* Tetanus, diphtheria, acellular pertussis (Tdap) vaccine
- Pertuzumab
  - actions and targets of, 534t, 536, 558, 559f
  - adverse effects of, 558, 638, 756
  - for breast cancer, 558, 637, 638, 641
- PET. *See* Positron emission tomography (PET)
- Petechiae. *See also* Purpura
  - in coma, 189
  - etiology of, 1248t
  - in fat embolism, 407
  - in meningococcemia, 992, 1248
  - in RMSF, 1456, 1456f
- Petrous bone fractures, 3574
- Peutz-Jeghers syndrome
  - breast cancer risk in, 633
  - gastrointestinal carcinoma in, 646, 651
  - lesions associated with, 654t
  - oral manifestations of, 264t
  - ovarian sex cord tumors in, 712
  - pancreatic cancer in, 671t
  - polyp distribution in, 654t
  - skin manifestations of, 398–399
- Peyer's patches
  - in immune-mediated damage, 2781
  - in *Salmonella* infections, 1315, 1317
  - in yersiniosis, 1348
- Peyote, 3696
- PF-4 (platelet factor 4), 923, 947, 2768t
- PFA (pulsed field ablation), 1917
- PFA-100, 474, 476
- PFAPA (periodic fever with aphthous stomatitis, pharyngitis, and cervical adenitis), 2762t, 2937
- Pfeiffer's syndrome (acrocephalopolysyndactyly), 3787
- PGD (primary graft dysfunction), 2286
- PGE_2. *See* Prostaglandin E_2 (PGE_2)
- *P* gene mutations, 396
- 3-PGHD (phosphoglycerate dehydrogenase) deficiency, 3368t
- PGL-1 (phenolic glycolipid), 1404, 1409
- P-glycoprotein (PGP), 484, 485t, 487f, 541, 574
- pH
  - arterial, 367
  - esophageal, 301
  - urine, 2450

Phaeohyphomycoses, 1688t, 1726t, 1728, 1728f
Phagocyte(s). *See also* Neutrophil(s)
  mononuclear. *See* Monocyte(s)
  in primary immune deficiency diseases, 2794, 2795f
  in vitro tests of, 463
Phagocyte activation defects, 459t, 460–461
Phagocytosis, 1383
Phagophobia, 293
Phalen's sign, 3613
Phantom pain
  limb, 3964, 3965f
  oral, 265
Pharmaceutical companies, physicians' relationships with, 73
Pharmacodynamics
  of antibacterial drugs, 1168, 1168f
  definition of, 483
  genetic considerations in, 493
  interactions in, 489t
  principles of, 486
Pharmacogenes, 491, E72t
Pharmacogenetics, **491**
  candidate gene studies in, 491
  in clinical practice, 493–494
  drug development and, 494–495
  of platelet inhibitors, 939, 939t
Pharmacogenomics, **491,** 3779, 3798, 3805
Pharmacokinetics, **484**
  absorption, 484, 1167
  of antibacterial drugs, 1167, 1168f
  distribution, 486, 1167
  elimination/excretion, 486, 1167
  genetic variants affecting, 491
  high-risk, 486
  interactions in, 489t
  metabolism, 486, 1167
Pharmacology, **483**
  clinical, 483
  diseases affecting drug concentration and response, 488
  genetic determinants of response to drugs, 491
  global considerations in, 483
Pharmacometabolomics, 3978
Pharyngitis
  arcanobacterial, 140t, 1226
  C. imitans, 1226
  clinical features of, 257, 257t
  C. ulcerans, 1225
  differential diagnosis of, 257, 257t, 1223
  in diphtheria, 1223, 1223f
  ear pain in, 253
  etiology of, 1208t
  in GERD, 301
  gonococcal, 259, 1256, 1259, 1259t
  HSV, 1498
  in infectious mononucleosis, 1509, 1509t
  nonstreptococcal, 257t, 258–259
  streptococcal
    clinical features of, 257, 1208
    complications of, 258, 1209
    diagnosis of, 257–258, 258t, 1208
    differential diagnosis of, 1511t
    epidemiology of, 1208
    treatment of, 258, 258t, 1208–1209, 1208t
  viral, 1208t, 1636, 2355
  Y. pestis, 1344
Phase response curve (PRC), 3944t, 3946
Phenacetin, 512t, 2442, 2444
Phenazopyridine, 1094, 3714t
Phencyclidine (PCP)
  abuse of, 3696
  actions of, 3656t
  delirium and, 184
  overdosage/poisoning with, 3705, 3708
Phendimetrazine, 3191
Phenelzine
  adverse effects of, 3469, 3659t
  for depression, 3659t
  for migraine prevention, 3469
  for new daily persistent headache, 117
  overdosage/poisoning with, 3710t
Phenformin, 485t, 786t
Phenobarbital
  adverse effects of, 323, 391, 424, 3422t
  drug interactions of, 489, 1742t, 1744t, 1745t, 2721, 3422t
  overdosage/poisoning with, 3713t
  for seizures, 3422t, 3424
Phenocopy, 491, 3787–3788
Phenol exposure, 320
Phenolic glycolipid (PGL-1), 1404, 1409
Phenolphthalein, 423
Phenothiazines
  for acute intermittent porphyria, 3347
  adverse effects of
    cardiac, 1920t
    cutaneous, 388, 434t
    heat-related illness, 3773
    ocular, 232
    orthostatic hypotension, 159
  for dyspnea, 83
  for nausea and vomiting, 575
  overdose/poisoning with, 3705
  for retinitis pigmentosa, 232
Phenotoxicity, 3825t
Phenotype, 3787
  convergent, 31, 31f
  divergent, 31, 31f
  oversimplification of, 30
  in precision medicine, 31–32, 35–36
  sex-influenced, 3789
Phenotypic heterogeneity, 3787, 3788t
Phenotypic sex, 3094–3095, 3094f, 3096f
Phenoxybenzamine, 3075, 3588, 3772t
Phentermine, 3191, 3192t, 3257
Phentermine/topiramate (PHEN/TPM), 3191, 3192t, 3193f, 3257
Phentolamine
  for cocaine intoxication, 3695
  for high-altitude pulmonary edema, 3757
  for Irukandji syndrome, 3723
  for nerve agent exposure, S5
  for pheochromocytoma, 3075
  for sympathomimetic overdosage/poisoning, 3710t
Phenuiviridae, 1656, 1657t, 1662t, 1663–1665t, 1666t, 1667
Phenylacetate, 3372
Phenylalanine metabolism disorders, 3368t, 3370
Phenylbutazones, 489t, 826
Phenylephrine, 2810, 3710t
Phenylketonuria, 3368t, **3370,** 3789
Phenylpropanolamine, 489, 3710t
Phenytoin
  adverse effects of, 3604t
    cross-sensitivity in, 427
    cutaneous, 391, 393, 399, 402
    DIHS, 424. *See also* Drug-induced hypersensitivity syndrome (DIHS)
    fever, 150
    folate deficiency, 787t, 788
    hepatotoxicity, 2665
    hirsutism, 3138
    hypertrichosis, 421
    in kidney disease, 488
    long-term, 3423
    lupus syndrome, 2781, 2835, 2940t
    neurologic, 3422t, 3609t
    osteopenia, 2940t
    SJS/TEN, 401, 425. *See also* Stevens-Johnson syndrome (SJS); Toxic epidermal necrolysis (TEN)
    systemic, 3422t
    thrombocytopenia, 923t
    vasculitis, 2907
  drug interactions of, 489t, 591
    acetaminophen, 489
    azoles, 1742t
    cyclosporine, 489t, 2721
    fluoxetine, 3422t
    glipizide, 485t
    isoniazid, 1422
    losartan, 489t
    methadone, 489t
    oral contraceptives, 489t
    PPIs, 2525
    quinidine, 489t, 1745t
    rifampin, 1423
    rufinamide, 3422t
    sulfonamides, 1743t, 3422t
    tacrolimus, 2721
    TMP-SMX, 1172t
    warfarin, 489t
  genetic variations in response to, 493, E72t
  metabolism of, 485t
  for neuropathy, 3604t
  overdosage/poisoning with, 3713t
  for pain, 100, 3588
  for paroxysmal symptoms in MS, 3588
  pharmacology of, 3422t, 3423
  for seizures, 3422t, 3423
  for seizures from CNS metastases, 591
  for status epilepticus, 2298, 3426f
  for trigeminal neuralgia, 3553
Pheochromocytoma, **3073**
  clinical features of, 3074
    cardiac, 291, 2018
    by frequency of occurrence, 3074t
    hypertension, 2140
    paraneoplastic syndromes, 738
    sympathetic overactivity, 3086, 3548
  definition of, 3073
  diagnosis of, 3086, 3086f
    biochemical testing, 3074, 3074t
    imaging, 3074, 3074t, 3075–3076f
    pathology, 3063, 3074, 3076f
  differential diagnosis of, 3063–3065, 3075
  epidemiology of, 3073
  etiology of, 3073
  familial, 3078
  genetic considerations in, 3073, 3078–3079, 3079t
  in MEN syndromes, 3077, 3086
  metastatic, 3077
  in neurofibromatosis, 3078, 3087
  in paraganglioma syndromes, 3077–3078
  pathogenesis of, 3073, 3073f
  in pregnancy, 3077
  treatment of, 3075–3077, 3086
  tumor markers in, 504t
  in von Hippel-Lindau syndrome, 3077, 3078f
Pheomelanin, 430, 432
PHEX gene mutations, 3264
PHI (Prostate Health Index), 697
Philadelphia chromosome. *See BCR::ABL1* gene
PHKA1 gene mutations, 3365
PHKG2 gene mutations, 3363
Phlebectasia corona, 2180
Phlebotomus sandfly, 1667
Phlebotomy
  for erythrocytosis, 741
  for hemochromatosis, 3337
  for polycythemia, 452, 820, 821
  for porphyria cutanea tarda, 435, 3348
Phlebovirus, 1662t, 1667
Ph-like ALL, 845–846, 848t

Phobic disorders, **3661**
Phonemic paraphasia, 201, 202
Phoneutria spp., 3734
Phonic tics, 3515
Phonological loop, 201
Phosgene, 2239t, S5
Phosphate nephropathy, 2440, A4
Phosphate/phosphorus
  adverse effects of, 364
  deficiency of. *See* Hypophosphatemia
  for DKA, S1
  excess of. *See* Hyperphosphatemia
  homeostasis of, 2369
  for hypercalcemia, 738, 3284t, 3285
  metabolism of, 3262–3263
  recommended intake of, 2601t
  renal tubular transport of, 2366f, 2369
  tolerable upper intake level of, 2614t
  toxicity of, 2614t
Phosphatidylethanol (Peth), 2709
Phosphatidylinositol-3-kinase (PI3K), 536, 3199
Phosphatonin, ectopic production of, 737t
Phosphodiesterase type 3 inhibitors, 2331
Phosphodiesterase type 4 inhibitors, 387t
Phosphodiesterase type 5 inhibitors
  actions of, 488, 3155, 3156f
  adverse effects of, 3159t
  drug interactions of
    nitrates, 2109, 2118, 3159
    tamsulosin, 704
  for erectile dysfunction, 3155, 3159, 3159t
  for high-altitude pulmonary edema, 3758
  for lower urinary tract symptoms, 704, 3175
  pharmacology of, 3159t
  for pulmonary arterial hypertension, 2187f, 2194, 2197t, 2874
  for Raynaud's phenomenon, 2178, 2874
Phosphofructokinase deficiency (type VII GSD, Tarui disease), 796t, 3361t, **3364**
Phosphoglycerate dehydrogenase (3-PGHD) deficiency, 3368t
Phosphoglycerate kinase deficiency, 796t, 3644
Phosphoglycerate mutase deficiency (type X GSD), 3361t, 3365
Phospholamban, 1849, 1850f, 2004t
Phospholipase A_2, 136, 430t
Phospholipase Cγ2, 2786f
Phospholipid transfer protein (PLTP), 3239
Phosphorus-32 exposure, S6
Phosphorylase kinase deficiency (type IX GSD), 2368t, 3361t, **3363**
Phosphorylation, histone, 3936–3937
Photoaging, 431
Photoallergy, 434, 434t
Photochemotherapy, 433, 436. *See also* PUVA photochemotherapy
Photochromogens, 1426
Photodynamic therapy, 430, 607, 645
Photoimmunology, 432–433
Photon, 554
Photon-counting CT, 3384, 3386f
Photopatch testing, 434
Photoproducts, 430
Photoprotection, 435–436, 436t. *See also* Sunscreen(s)
Photopsia, 229, 232
Photoreactive keratectomy (PRK), 220
Photoreceptors, 220
Photosensitivity diseases, **433**
  classification of, 433t
  diagnosis of, 433, 436f
  drug-induced, 420, 434, 434t
  management of, 434, 436f
  polymorphous light eruption, 405, 433–434
  porphyrias. *See* Porphyria(s)
Photosensitizer, 434
Phototest procedures, 434
Phototherapy. *See also* PUVA photochemotherapy
  actions of, 436
  adverse effects of, 436
  for atopic dermatitis, 393t
  for cancer treatment, 555
  for delayed sleep-wake phase disorder, 218
  for morphea, 416
  for neonatal jaundice, 2639
  for pityriasis rosea, 388
  for psoriasis, 387, 393t
  for shift-work disorder, 219
  for vitiligo, 398t
Phototoxic eruptions, 400
Phototoxicity, 434, 434t. *See also* Photosensitivity diseases
PHP. *See* Pseudohypoparathyroidism (PHP)
PHQ-2 (Patient Health Questionnaire-2) Depression Screen, S8
Phrenic nerve stimulator, 2282
Phrygian cap, 2728
Phthalocyanines, 555
Phthisis. *See* Tuberculosis (TB)
Phylloquinone, 2613. *See also* Vitamin K
Physical abuse, 3895t
Physical activity. *See* Exercise/physical activity
Physical Activity Guidelines for Americans, 11, 12t
Physical examination
  clues to asymptomatic disorders in, 49, 50t
  historical evolution of, 46
  improved diagnostic accuracy from, 47–49, 48t–49t, 49f
  as meaningful ritual for patient and physician, 51
  medical error from oversights in, 49–51
  physical signs of "stop rules" for, 49, 49t
  purpose of, 2
  as sole means of diagnosis, 47, 47t
Physical restraints, in delirium, 86, 182, 186
Physical signs, 48–49, 48t
Physical therapy
  for axial spondyloarthritis, 2886
  for low back pain, 127
  for lymphedema, 2184
  for neck pain, 132
  for vertigo, 165t
Physician. *See also* Medical practice
  brain drain, 3848
  as citizen, 8
  claims of conscience by, 72
  conflicts of interest for, 73
  continuing education for, 7–8
  errors of, 4–5, 42, 73
  as gatekeeper, 72
  humanistic qualities of, 6
  impaired, 73
  moral distress of, 73
  oaths and codes of, 72
  occupational risks of, 72
  personal values of, 72
  research and teaching by, 8
  as student, 7–8
  trainee, ethical issues for, 73
  virtue ethics and, 72
  well-being of
    challenges in current health care landscape, 52–53, 52f
    historical perspective on, 52
    moral distress and, 73
    strategies for improving, 53–55, 54f
  working hours of, 7, 42, 219
Physician-assisted suicide, 89–90, 89t
Physician-induced demand, 23
Physician Orders for Life-Sustaining Treatment (POLST), 79
Physician-patient relationship, 2, **5–6,** 66–67, 66f
Physostigmine
  for anticholinergic overdose, 192, 3706, 3711t
  for belladonna alkaloid overdose, 3711t
  for mushroom/plant poisoning, 3711t
  overdosage/poisoning with, 3712t
PI3K, 532, 573, 714
PI3K/AKT pathway, 532, 541, 601, 611
PI3K inhibitors, 571t, 574, 757, 856
PI3Kδ, 2786f
Pibrentasvir, 2686f
Pibrentasvir/glecaprevir. *See* Glecaprevir/pibrentasvir
Pica, 765, 3670
PICALM gene mutations, 3479
PICC (peripherally inserted central catheter), 2161
Pick's disease, 3487, 3487f, 3488f
Picobirnaviruses, 1628f, 1632
Picornaviridae
  name changes in, 1534, 1534t
  replication strategies of, 1479
  structure of, 1478f, 1480t
"Picture frame" vertebra, 3313
PID. *See* Pelvic inflammatory disease (PID)
Piebaldism, 397, 398t
PIEZO1 gene mutations, 794t, 795
PIG-A gene mutations, 810
Pigbel (enteritis necroticans), 1240
Pigeon chest (pectus carinatum), 1860, 3332
Pigmented pretibial papules, 3230
Pigmented villonodular synovitis (PVNS), 2972, 2972f
Pigment stones, 2729–2730, 2730t
PIK3CA gene mutations, 520t
  in bladder cancer, 691
  in breast cancer, 639, 640, 3982
  in Cowden's syndrome, 3080t, 3088
  in lung cancer, 617, 617f
  in ovarian cancer, 710
PIK3CA pathway, 526t
Pili (fimbriae), bacterial, 965–966, 966t
  of *B. Pertussis,* 1277
  E. coli, 965
  of enterococci, 1216
  of *Neisseria* spp., 966
  of *N. gonorrhoeae,* 1253
  of *P. aeruginosa,* 1307, 1307t
  of *S. pneumoniae,* 1188f
  type IV, 965
Pill esophagitis, 295, 300, 2513–2514
Pilocarpine
  for acute angle-closure glaucoma, 225
  miosis produced by, 221
  for oral dryness, 243
  overdosage/poisoning with, 3712t
  for sicca symptoms, 2878
  for xerostomia, 266
Pilocytic astrocytoma, 720, A16
Pilosebaceous unit, 3138
Pimavanserin, 3493, 3506, 3673t
Pimecrolimus, 383, 393t
Pimozide, 1742t, 3738
Pinch purpura, 404
Pindolol, 2100t, 3658, 3712t
Pineal gland, 3945
Pineal region tumors, 721
Pinguecula, 223
PINK1 gene mutations, 3400, 3498t, 3499
Pinna. *See* Ear
Pinta, 1437–1438, 1437t, 1439, 1439f
Pintides, 1438
Pinworm. *See Baylisascaris procyonis* infections; *Enterobius vermicularis*/enterobiasis (pinworm)
Pioglitazone, 691, 2611, 2706, 3211t, 3214, 3258
Piperacillin, 802t, 923t, 1175, 2580

- Piperacillin-tazobactam
  - for actinomycosis, 1366t
  - for anaerobic bacterial infections, 1379t
  - for community-acquired pneumonia, 1028t
  - for *Enterobacterales* infections, 1283
  - for hospital-acquired pneumonia, 963t
  - indications for, 1173t, 1175
  - for intraabdominal infections, 963t
  - for neutropenic enterocolitis, 582
  - for neutropenic fever, 580
  - for osteomyelitis, 1064t, 1065t
  - for *P. aeruginosa* infections, 1309t
  - for peritonitis, 1070, 1071
  - for *Proteus* spp. infections, 1291
  - for pyelonephritis, 1094
  - resistance to, 1173t, 1291, 1379, 1379t
  - for septic shock, 963t, 991t
  - for ventilator-associated pneumonia, 1032t
- Piperaquine, 1741t, 1751, 1771t
- Piperazine, 1745t, 1751
- Piracetam, 3516
- Piranha bite-wound, 1143, 1145t
- Pirfenidone, 2230, 2262
- Piriformis syndrome, 121t
- Piroxicam, 411, 434t
- Pirtobrutinib, 533t, 567, 568t, 866
- Piry virus, 1663t
- Pit-1 (PIT-1)
  - in anterior pituitary development, 2988, 2988t, 2992
  - in congenital hypothyroidism, 3028t
  - in growth hormone deficiency, 2994
- Pitavastatin, 3244t. *See also* Statins
- Pitolisant, 215
- Pitting, 479
- Pitting edema, 281, 1860. *See also* Edema
- Pituitary apoplexy, 232, 2993–2994, 3067, 3068t
- Pituitary disorders
  - anterior
    - insufficiency. *See* Hypopituitarism
    - tumors. *See* Pituitary tumors (adenomas)
  - IgG4-related, 2931t
  - posterior
    - diabetes insipidus. *See* Diabetes insipidus (DI)
    - hypodipsic hypernatremia. *See* Hypodipsic/adipsic hypernatremia
- Pituitary gland
  - anatomy of, 2988, 2989f
  - anterior. *See* Anterior pituitary
  - development of, 2988
  - insufficiency of. *See* Hypopituitarism
  - posterior. *See* Posterior pituitary
  - tests of function of, 2995t
- Pituitary snuff taker's lung, 2228t
- Pituitary tumors (adenomas), **3003**
  - ACTH-producing. *See* Cushing's disease
  - aggressive, 3014
  - classification of, 3003, 3003t
  - clinical features of, 2998, 2999t, 3083
  - etiology of, 3003–3004
  - gonadotropin-producing, 3001t, 3012–3013
  - histologic evaluation in, 2999
  - hormone-excess syndromes due to, 3003–3004
  - hypogonadotropic hypogonadism and, 3114
  - laboratory investigation of, 2999, 3001t
  - local mass effects of, 2998, 3000f
  - in MEN 1, 3004, 3004t, 3080t, 3083
  - metastatic, 3003
  - MRI in, 2998–2999, 3000f, 3002f, A16
  - nonfunctioning, 3012–3013
  - ocular signs in, 222, 232, 233f
  - pathogenesis of, 3003–3004
  - secondary adrenal insufficiency and, 3068t
  - treatment of, 2999–3000, 3001f, 3083
  - TSH–secreting, 3001t, 3013–3014, 3037
- Pit viper bite, 3717, 3720t, 3722
- Pityriasis rosea, **388**
  - clinical features of, 386t
    - distribution, 378t, 380f, 386t
    - lesion characteristics, 388
    - morphology, 378t
    - papulosquamous lesions, 392
    - patches on trunk, 388f, A5
  - differential diagnosis of, 146
  - drug-induced, 392
  - histologic features of, 386t
  - treatment of, 388
- Pityriasis rubra pilaris, 393t
- Pityriasis versicolor. *See* Tinea versicolor
- Pivmecillinam, 1093, 1093t, 1094, 1283, 1324t
- Pizotifen, 3470t
- PKCδ (protein kinase Cδ) deficiency, 2786f
- PKD (paroxysmal kinesigenic dyskinesia), 3517
- *PKD* gene mutations, 2430
- *PKHD1* gene mutations, 2433, 2434
- Placebo, 3930t
- Placebo effects, **3930**
  - biological mechanisms of, 3933
  - in clinical trials, 3931, 3932f
  - definition of, 3930t
  - future of, 3934
  - genetic considerations in, 3933
  - harnessing of, 3934–3935
  - history of, 3930–3932
  - imaging of, 95f
  - in pain relief, 95
  - psychological mechanisms of, 3931
  - social and cultural mechanisms of, 3931–3933
  - terminology in studies of, 3930t
- Placebome, 3930t, 3933
- Placebo responders, 3933
- Placebo response, 3930t
- Placental abruption, 3899
- Placental growth factor, 3899
- Plague, **1341**
  - in bioterrorism, 1342, S4
  - bubonic, 1343–1344, 1344f, S4
  - case definitions of, 1345t
  - clinical features of, 1051, 1343–1344, 1344f, S4
  - diagnosis of, 1344–1346, 1345f, 1345t, S4
  - epidemiology of, 1342–1343, 1342f, 3753f
  - etiology of, 1342. *See also Yersinia pestis*
  - global considerations in, 1342–1343, 1342f
  - mortality from, 1343
  - multidrug-resistant, S4
  - pathogenesis of, 1343
  - pneumonic, 990, 1344, 1344f, S4
  - prevention of, 1347, 1347t, S4
  - septicemic, 1344, S4
  - treatment of, 1346, 1346t, S4
- Plague of Athens, 3853–3854, 3854t
- Plague vaccine, 1347, S4
- PLAID (PLCγ-associated antibody deficiency and immune dysregulation), 2804
- Plan-Do-Check-Act cycle, 44, 44f
- Plantar fasciitis, 2887, 2975
- Plantar reflex, 3381
- Plantar warts, 390
- Plant-based diets, 3741, 3744f
- Plaque
  - atherosclerotic. *See* Atherosclerosis, pathogenesis of
  - dental, 260, 1374, A3
  - microbiota of, 3833f
  - neuritic, in AD, 3478, 3478f
- Plaque (skin lesion), 137, 377t, 378f
- Plasma
  - clinical uses, 906t
  - collection and manufacturing processes, 902–903t
  - fresh-frozen, 934, 935, 3456
- Plasma cell(s), A6
- Plasma cell disorders, **883**
  - acid-base disorders in, 368
  - amyloidosis. *See* Amyloidosis
  - genetic considerations in, 883, 884f
  - heavy chain diseases, **894**
  - multiple myeloma. *See* Multiple myeloma
  - paraproteinemia in, 473
  - pathophysiology of, 883, 885f
  - POEMS syndrome. *See* POEMS syndrome
  - proteinuria in, 342
  - Waldenström's macroglobulinemia. *See* Waldenström's macroglobulinemia
- Plasma cell leukemia, 889
- Plasmacytic lymphoma of the oral cavity, 1613
- Plasmacytoid 1 dendritic cells, 2762, 2763t. *See also* Dendritic cells
- Plasmacytoma, 888, 888t, 889, 890
- Plasma exchange/plasmapheresis
  - for ADEM, 3594
  - for CIDP, 3622
  - efficacy evaluation of, 906t
  - for Guillain-Barré syndrome, 3620
  - for hemolytic-uremic syndrome, 2445
  - for hyperviscosity syndromes, 893
  - indications for, 906t
  - for MGUS, 889
  - for MS, 3585
  - for myasthenia gravis, 3630, 3631
  - for neuromyelitis optica, 3592
  - for pemphigus vulgaris, 410
  - for SLE, 2835
  - for thrombotic thrombocytopenic purpura, 925, 2428, 2445
  - for Waldenström's macroglobulinemia, 893
- Plasma protein binding, 486
- Plasma renin activity, 355
- Plasmin, 471–472, 471f, 954, 954f
- Plasminogen, 471–472, 471f, 954, 954f
- Plasminogen activator, 954, 954f
- Plasminogen activator inhibitor I (PAI-1), 941, 954, 954f, 1845
- Plasminogen activator inhibitors, 471–472, 471f
- *Plasmodium* spp. *See also* Malaria
  - characteristics of, 1760–1761, 1760t, A2
  - drug resistance in, 1768, 1775
  - epidemiology of, S12
  - host response to, 1762–1763
  - laboratory identification of
    - blood findings, 1769t
    - demonstration of parasite, 1766–1767, 1766f, 1767f, 1768f, A2, S12
    - PCR, S12
  - life-cycle hosts of, S12
  - *P. falciparum*
    - adherence of, 1737
    - characteristics of, 1760–1761, 1760t, A2
    - erythrocyte changes caused by, 1762
    - G6PD deficiency and, 799
    - geographic distribution of, 1761–1762
    - hemoglobin S and, 772
    - laboratory identification of, 1766f, 1767f, 1769t, A2
    - ligand-receptor interactions of, 1763
    - temperature sensitivity of, 1762
  - *P. knowlesi*
    - characteristics of, 1760–1761, A2
    - geographic distribution of, 1762
    - laboratory identification of, 1769f, 1769t, A2, S12
  - *P. malariae*
    - characteristics of, 1760t, A2
    - geographic distribution of, 1762
    - laboratory identification of, 1769f, 1769t, A2

  P. ovale
    characteristics of, 1760t, A2
    geographic distribution of, 1762
    laboratory identification of, 1768f, 1769t, A2
  P. vivax
    characteristics of, 1760t, A2
    geographic distribution of, 1762
    inflammasome mutations in, 2761t
    laboratory identification of, 1767f, 1769t, A2
    temperature sensitivity of, 1762
  tissue invasion by, 1737
  transmission cycle, 1761f
Plasticity, of cancer cells, 542–543
Plateau pressure, 2296
Platelet(s), 921
  abnormal function of, 407
  activation of, 316–317, 470, 921, 937f, 938, 943f
  adhesion of, 470, 921, 938, 942, 943f
  aggregation of, 921, 938, 943f
  in AML, 828
  in arterial thrombosis, 937–938
  characteristics of, 921, 937–938
  collection and manufacturing processes, 902t
  disorders of, 472t, **921**
    acquired, 926
    hemolytic-uremic syndrome. *See* Hemolytic-uremic syndrome (HUS)
    inherited, 926
    thrombocytopenia. *See* Thrombocytopenia
    thrombocytosis. *See* Thrombocytosis
    von Willebrand disease. *See* Von Willebrand disease
  elevated levels of, 824
  giant, 441f
  hematopoietic differentiation of, 760f
  in hemostasis, 469–470, 921
  in inflammation, 938
  in peripheral blood smear, 437
  transfusion of. *See* Platelet transfusion
Platelet-activating factor, 417, 2812
Platelet aggregometry, 476
Platelet count
  in DIC, 934, 934t
  normal, 473, 921
  in preoperative evaluation, 473
Platelet-derived growth factor (PDGF), 546f, 2773
Platelet-derived growth factor (PDGF) receptors, 546f, 567
Platelet factor 4 (PF-4), 923, 947, 2768t
Platelet function tests, 476
Platelet inhibitors. *See* Antiplatelet drugs
Platelet plug, 469–470, 921
Platelet refractoriness, 911–912
Platelet storage pool disorder, 926
Platelet transfusion
  for AML, 832
  for aplastic anemia, 813
  for coagulation disorders in liver disease, 935
  for DIC, 934
  in intracerebral hemorrhage, 3456
  of platelet concentrates, 906t
  for thrombocytopenia, 575
Platypnea, 270, 2630
Platypus, 3725
Plazomicin
  actions of, 1166, 1182t
  for *Citrobacter* infections, 1293
  indications for, 1174t, 1176
  in pregnancy and lactation, 1170t
  resistance to, 1182t, 1184
  for ventilator-associated pneumonia, 1032
PLCγ-associated antibody deficiency and immune dysregulation (PLAID, APLAID), 2804
Plecanatide, 314, 2576, 2577t, 3549
Pleconaril, 1537
Pleistophora spp., 1808
Pleomorphic adenoma, 266
Pleomorphic xanthoastrocytomas, 720
Plerixafor, 759, 915
Pleural biopsy, 2208
Pleural catheter, indwelling, 2290
Pleural effusion, **2267**
  in chylothorax, 2270
  in cirrhosis, 323, 2270
  diagnosis of, 2267–2268, 2267f, 2268t, 3912t, 3914v, A12
  differential diagnosis of, 1986t
  dyspnea in, 270
  etiology of, 2267, 2270
  exudative, 2267–2268, 2268t
  in heart failure, 1984, 2268
  in hemothorax, 2270
  malignant, 619, 620, 2269–2270
  in mesothelioma, 2270
  in pancreatitis, 2745, 2748f
  parapneumonic, 2268–2269, 2269f
  in pneumonia, 1029
  in pulmonary embolism, 2270
  transudative, 2267–2268, 2268t
  tuberculous, 1387, 2270
  in viral infection, 2270
Pleural (primary) effusion lymphoma, 1516
Pleuritis
  abdominal pain in, 111
  chest pain in, 103t
  in rheumatoid arthritis, 2840
  in SLE, 2830
  tuberculous, 2270
Pleurocentesis (thoracentesis), 2208
Pleurodynia (Bornholm disease), 1635
Pleuromutilin. *See* Lefamulin
Pleuroscopy, 2208, 2290
Plexogenic remodeling, pulmonary, 2185, 2186f, 2187
Plexopathies
  brachial, 3614–3615
  lumbosacral, 3615, 3616t
  perioperative, 3615
  radiation-induced, 3615–3616
PLOD1 gene mutations, 3328t, 3330
PLOD2 gene mutations, 3325t
PLP (pyridoxal 5'-phosphate), 3316
PLP gene mutations, 3523t
PLTP (phospholipid transfer protein), 3239
Plummer-Vinson syndrome, 265, 642, 765
Plutonium exposure, S6
PM. *See* Polymyositis (PM)
PML. *See* Progressive multifocal leukoencephalopathy (PML)
PML gene, 826
PML-RARA, 827, 832, 834
*PML-RAR*α, 534t, 539, 573
PML-RAR gene mutations, 3783t
PMNs (polymorphonuclear leukocytes), 1104, 2116, A10, S11
PMP22 gene mutations, 3784
PMS1 gene mutations, 710, 714, 3786
PMS2 gene mutations, 523, 524t, 671t, 710, 714, 3786
PNad (peripheral node addressin), 2781
*pnc*A gene, 1397
pNETs. *See* Neuroendocrine tumors (NETs), pancreatic
PNETs (primitive neuroectodermal tumors), 728, 730
Pneumatic compression device, for VTE prophylaxis, 3910
Pneumatocele, A12
Pneumatosis intestinalis, 597, 597f, 2870
Pneumococcal H inhibitor (hic), 1187
Pneumococcal infections, **1187**
  approach to the patient, 1191
  case-fatality ratios for, 1190
  clinical features of
    endocarditis, 1037t, 1044. *See also* Infective endocarditis
    meningitis, bacterial. *See* Meningitis
    otitis media. *See* Otitis media
    pneumonia. *See* Pneumonia, bacterial
    sepsis/septic shock, 1194
    sinusitis, 1193
  in COPD, 2259
  epidemiology of, 988–1190, 1189f
  etiology of, 1187. *See also Streptococcus pneumoniae*
  global considerations in, 1195
  in HCT recipient, 1157
  in HIV infection, 1595–1596
  host defense mechanisms, 1191
  incidence pre- and postvaccine, U.S., 1002t, 1189f, 1195
  inflammasome mutations in, 2761t
  pathogenesis of, 1190–1191
  post-splenectomy, 482
  prevention of, 1195. *See also* Pneumococcal vaccine
  risk factors for, 1190, 1190t
  treatment of, 1174t, 1193–1194
Pneumococcal surface adhesion A, 1187
Pneumococcal surface antigen A, 1188f
Pneumococcal surface protein A, 1188f
Pneumococcal surface protein C, 1187, 1188f
Pneumococcal vaccines
  capsular polysaccharide, 1194
  effectiveness of, 1189f, 1195
  genomics and, 984, 987
  global considerations in, 1194–1195
  polysaccharide–protein conjugate, 1189f, 1194–1195
  recommended administration of
    in adults, 40t, 1004t
    after splenectomy, 482
    contraindications to, 1007t
    in HIV infection, 1594t
    in liver disease, 2633
    in multiple myeloma, 893
Pneumocystis spp., 1730, 1731
Pneumocystis jiroveci, 961t, 1730, 2781, 3858. *See also Pneumocystis* pneumonia (PCP)
Pneumocystis pneumonia (PCP)
  clinical features of, 1596, 1731
  diagnosis of, 582, 1163, 1596, 1689t, 1731–1732, 1732f, 1733f
  epidemiology of, 1730–1731
  in HCT recipient, 1156, 1162
  in HIV infection, 1585, 1595f, 1596, 1730–1731, 1734
  pathogenesis of, 1731, 1732–1733
  prognosis of, 1732–1733
  prophylaxis for, 578–579, 582, 585, 1164t, 1734–1735, 1734t, 2410
    in HIV infection, 1593t, 1596
    pentamidine, 1735
    trimethoprim-sulfamethoxazole, 463, 918, 918t, 1156, 1735, 2895
  risk factors for, 1689t
  in solid organ transplant recipient
    after high-dose glucocorticoids, 567
    after lung transplant, 1162–1163, 2286–2287
    prophylaxis for, 1734–1735, 2410
  treatment of, 1596, 1733–1734, 1734t
Pneumocystosis, 1689t
Pneumolysin, 1187, 1188f
Pneumomediastinum, 2273
Pneumonectomy, for lung cancer, 623, 626

Pneumonia, **1023**
- abdominal pain in, 111, 112t
- acute interstitial (Hamman-Rich syndrome), 2265
- aspiration, foreign body, 1375
- atypical, 1024
- bacterial. *See also* Pneumonia, community-acquired
  - *A. baumannii,* 1296
  - *Actinomyces,* 1364
  - after influenza, 1200, 1547–1548
  - aspiration, 1375
  - case-fatality ratios for, 1190
  - clinical features of, 1191–1192, 1211
  - complications of, 1192, 1211
  - *C. pneumoniae,* 1476–1477
  - *C. psittaci,* 1475–1476
  - diagnosis of, 1192, 1192f
  - differential diagnosis of, 1192
  - *E. coli,* 1286
  - *Enterobacter,* 1292
  - in granulomatosis with polyangiitis, A14
  - *H. influenzae,* 1262
  - in HIV infection, 1595–1596. *See also Pneumocystis* pneumonia (PCP)
  - incidence pre- and postvaccine, U.S., 1002t, 1189f, 1195
  - *K. pneumoniae,* 1289–1290
  - *Legionella,* 1272–1273, 1272t, 1275. *See also Legionella* spp. infections
  - *M. catarrhalis,* 1265
  - *M. pneumoniae,* 1465
  - *Nocardia,* 1359, 1359–1360f, 1362, 1362t
  - nontyphoidal *Salmonella,* 1320
  - *P. aeruginosa,* 1309t, 1310
  - parapneumonic effusion in, 2268
  - *R. equi,* 1226
  - treatment of, 1193–1194, 1208t
  - *T. whipplei,* 1369
- *Blastomyces,* 1701
- in cancer patient, 582
- *Coccidioides,* 1697
- community-acquired
  - clinical features of, 48t, 1024–1025
  - complications of, 1029
  - diagnosis of, 1025–1026
  - epidemiology of, 1024, 1024t
  - etiology of, 1023–1024, 1024t
  - follow-up, 1029
  - PPIs and, 2525
  - prevention of, 1029
  - prognosis of, 1029
  - risk factors for, 1024, 1024t
  - severity indicators, 1026, 1026t
  - treatment of
    - antibiotic resistance in, 1026–1027
    - empirical antibiotics, 963t, 1027–1029, 1027t, 2321t
    - failure to respond to, 1029
    - racial/ethnic disparities in, 62
    - site of care, 1026, 1026t
- in COVID-19 disease, 995
- cryptogenic organizing, 2263f, 2264–2265, 2264f
- deaths from, 1024
- definition of, 1023
- diagnosis of, 582, 3912t, 3914v
- eosinophilic. *See* Pulmonary infiltrates with eosinophilia
- global impact of, 1033
- health care–associated
  - *A. baumannii,* 1296
  - diagnosis of, 1033
  - *E. coli,* 1286
  - empirical antibiotics for, 963t
  - etiology of, 1023, 1147–1148
  - pathogenesis of, 1033
  - *S. aureus,* 1200
  - treatment of, 2321t
- incidence of, 1024
- infant *(C. trachomatis),* 1473t, 1474
- interstitial. *See* Interstitial pneumonia
- mixed viral and bacterial, 1547–1548
- mortality rates, 1545f, 1547f
- necrotizing, 274, 1024, 1032
- pathology of, 1023
- pathophysiology of, 1023, 2203, 3994
- Q fever, 1462
- typical, 1023
- ventilator-associated. *See* Ventilator-associated pneumonia (VAP)
- viral
  - CMV, 1514
  - enterovirus, 1636
  - influenza, 1547–1548
  - varicella, 1505

Pneumonia Severity Index (PSI), 1026

Pneumonic plague, 1344, 1344f, S4. *See also* Plague

Pneumonic tularemia, 1338, 1339f, 1340

Pneumonitis
- in cancer patient, 596, 754
- checkpoint inhibitor–induced, 596, 596f
- chemical, 1375
- in COVID-19 disease, 596, 596f
- giant-cell, 1641
- HSV, 1500
- hypersensitivity. *See* Hypersensitivity pneumonitis
- radiation-induced, 596, 754
- VZV, 1504

Pneumoperitoneum, 1070, 1070f

Pneumothorax
- chest pain in, 103t, 104
- as contraindication to hyperbaric oxygen therapy, 3765
- imaging of, A12
- secondary, 2270–2271
- spontaneous, 104, 2270
- tension, 104, 2271
- traumatic, 2271
- treatment of, 2290

Pneumoviridae, 1480t, 1533t. *See also specific viruses*

PNKD (paroxysmal nonkinesigenic dyskinesia), 3517

PNPLA3 gene mutations, 660, 2701, 2702

PNPLA6 gene mutations, 3112t

Pocapavir, 1489

Podagra, with gouty inflammation, A15

Podoconiosis, 2183

Podophyllotoxin, 390, 1531, 3609t

POEM (peroral endoscopic myotomy), 2469, 2476f, 2493, V5

POEMS syndrome, **894**
- clinical features of, 400, 888t, 894, 2447, 3093, 3623
- diagnosis of, 888t
- differential diagnosis of, 400
- pathogenesis of, 894, 3093
- renal involvement in, 2447
- treatment of, 894, 3093

POET (peroral endoscopic tumorectomy), 2469, 2477f, 2493, V5

Poikilocytes, 437

Poikilocytosis, 437, 445, 445f, 482, 777, 800

Poikiloderma, 378t, 396, 414

Poikilothermia, 3770

Point mutation, 520–521, 522, 3784, 3786f

Point-of-care ultrasound (POCUS), **3911**
- common applications of, 3911, 3913f
  - abdomen, 3912t, 3913, 3914v
  - cardiac, 3911, 3912t, 3914v
  - deep venous thrombosis, 3913, 3914v
  - lung and pleura, 3911–3913, 3912t, 3914v
  - skin and soft tissue, 3912t, 3913, 3914v
  - vascular access, 3913–3914
- vs. consultative ultrasound, 3911f
- future directions in, 3914
- vs. physical examination, 3911f, 3912t
- in shock, 2313–2314
- training for, 3914

Point-source exposures, 3855

Poisoning/drug overdose, **3703**. *See also specific substances*
- epidemiology of, 3703
- global considerations in, 3716
- history in, 3703
- hypothermia in, 3768
- laboratory evaluation in, 3705–3706
- pattern recognition in, 3704–3705, 3704t
- physical examination in, 3703–3704
- prevention of, 3716
- severity assessment in, 3704–3705, 3705t
- treatment of
  - antidote administration, 3710
  - cardiovascular therapy, 3707–3708
  - CNS therapy, 3708
  - decontamination at other sites, 3709
  - enhancement of poison elimination, 3709–3710
  - gastrointestinal decontamination, 3708–3709
  - general principles, 3706–3707, 3706t
  - prevention of reexposure, 3716
  - respiratory care, 3707
  - supportive care, 3707–3708

Poison ivy, 384

Poison oak, 384

Poison sumac, 384

Poisson distribution, 3956, 3956f

Polar bear bite-wound, 1143

Polarized cells, 2365

Polatuzumab vedotin, 535t, 566, 566t, 753t, 863

pol gene, 1551, 1552f, 1558, 1561f

POLG gene mutations, 3645, 3806, 3815

Poliomyelitis. *See* Poliovirus infection

Poliovirus infection
- clinical features of, 1633
- diagnosis of, 1633
- elimination and eradication of, 1002
- global considerations in, 1002, 1637–1638, 1637t
- incidence pre- and postvaccine, U.S., 1002t
- paralytic poliomyelitis, 1633–1634
- pathogenesis of, 1633
- postpolio syndrome, 1634, 3523
- prevention and eradication of, 1637–1638, 1638t
- vaccine-associated, 1634, 3853
- weakness in, 170

Poliovirus vaccine
- effectiveness of, 1637
- poliomyelitis associated with, 1634, 1637
- recommended administration of
  - in adults, 1638, 1638t
  - in international travelers, 1014t, 1015
- types of, 1637–1638

Polonium-210 exposure, S6

POLR3B gene mutations, 3112t

POLST (Physician Orders for Life-Sustaining Treatment), 79

Poly-ADP-ribose polymerase. *See* PARP

Polyangiitis
- with granulomatosis. *See* Granulomatosis with polyangiitis
- microscopic. *See* Microscopic polyangiitis

- Polyangiitis overlap syndromes, 2907
- Polyarteritis nodosa, **2901**
  - clinical features of, 2083, 2901, 2901t
  - definition of, 2901
  - diagnosis of, 2901, A14
  - drug-induced, 426
  - incidence and prevalence of, 2901
  - neuropathy in, 3623
  - pathology and pathogenesis of, 2893, 2901
  - skin manifestations of, 406
  - treatment of, 2901–2902
- Polyarthropathy syndrome, 1523, 1524t
- Polychondritis, relapsing. *See* Relapsing polychondritis
- Polychromasia, 445
- Polychromatophilia, 437, 438f, 447t
- Polycomb family, 761
- Polycyclic hydrocarbons, 512t, 727
- Polycyclic lesions, 378t
- Polycystic kidney disease
  - autosomal dominant, **2430**
    - clinical features of, 2166, 2360, 2431–2432
    - diagnosis of, 2432–2433
    - early interventions for, 3804t
    - genetic considerations in, 2430–2431, 2430t, 3788t
    - genetic testing for, 3804t
    - pathogenesis of, 2430–2431, 2431f, 2454
    - treatment of, 2433
  - autosomal recessive, 2430t, **2433**
- Polycystic ovarian syndrome (PCOS)
  - in CAH, 3101
  - comorbid conditions, 3135
  - hirsutism in, 3138, 3140
  - infertility in, 3151
  - insulin resistance in, 3202
  - MASLD and, 2702
  - menstrual disorders in, 3135
  - metabolic syndrome and, 3256
  - obesity and, 3135, 3187
  - prevalence of, 2979t
  - screening/testing for, 2979t, 3140
  - skin manifestations of, 391, 397
  - treatment of, 3135
- Polycystin-1/2, 2430, 2431f
- Polycythemia, **450**
  - as adaptation to hypoxia, 278
  - approach to the patient, 452, 452f
  - clinical features of, 451–452
  - cyanosis and, 278
  - definition of, 450–451
  - erythromelalgia in, 2179
  - secondary, 278
  - smoker's, 452
  - spurious, 451
- Polycythemia vera, **819**
  - classification of, 818–819
  - clinical features of, 451, 819
  - complications of, 820
  - diagnosis of, 452, 819–820
  - epidemiology of, 819
  - etiology of, 819, 820t
  - genetic considerations in, 819
  - splenomegaly in, 480, 820
  - treatment of, 820–821
- Polydipsia
  - etiology of, 343–344
  - hyponatremia in, 350
  - primary. *See* Primary polydipsia
- Polydrug abuse, 3692. *See also* Substance use disorders
- Polyethylene glycol, 81, 314, 509, 2575, 2577t
- Polygenic disorders, 3777, 3793, 3794t, 3795f, 3958f
- Polygenic risk score, VTE, 2156
- Polyglandular autoimmune insufficiency. *See* Autoimmune polyendocrine (polyglandular) syndrome-1 (APS-1)
- Polymerase chain reaction (PCR)
  - definition of, 976t
  - for genetic testing, 3796, 3801
  - in infectious disease diagnosis, 977t, 979
  - in infective endocarditis diagnosis, 1040
  - in leprosy diagnosis, 1410
  - in pneumonia diagnosis, 1025
  - quantitative, 3820–3821
  - in rabies diagnosis, 1653
  - reverse transcriptase, 1590, 1590t
  - in sputum analysis, 2214
  - in viral encephalitis diagnosis, 1112–1113
  - in viral meningitis diagnosis, 1124
- Polymerase gamma, 3806
- Polymer fume fever, 2240
- Polymeric formulas, for enteral nutrition, 2622
- Polymorphic microsatellite markers, 3793
- Polymorphic reticulosis. *See* Extranodal NK/T cell lymphoma, nasal type
- Polymorphic ventricular tachycardia. *See* Ventricular tachycardia (VT), polymorphic
- Polymorphism, 3784. *See also* Single-nucleotide polymorphisms (SNPs)
- Polymorphonuclear leukocytes (PMNs), 1104, 1373, 2116, A10, S11
- Polymorphous light eruption, 405, 433–434
- Polymyalgia rheumatica, 132, 155, 227, 2847, **2902,** 3636
- Polymyositis (PM), **2912**
  - clinical features of, 172, 2912
  - diagnosis of, 2912, 2912f
  - dysphagia in, 296
  - histopathology of, 2912, 2913f
  - immunopathology of, 2790t
  - interstitial lung disease in, 2266
  - myocarditis in, 2015
  - prognosis of, 2912
  - Raynaud's phenomenon in, 2178
  - treatment of, 2915–2916, 2916t
- Polymyxin B
  - for *A. baumannii* infections, 1297, 1297t
  - actions of, 1167, 1178–1179, 1182t, 1183f
  - adverse effects of, 423, 1171t, 1179
  - for *Enterobacterales* infections, 1284
  - for gram-negative bacterial infections, 1268t
  - indications for, 1179
  - for otitis externa, 254
  - for *P. aeruginosa* infections, 1312
  - resistance to, 1182t, 1183f, 1185, 1283
  - for ventilator-associated pneumonia, 1032t
- Polymyxin-bacitracin, 224
- Polymyxin E (colistin)
  - for *A. baumannii* infections, 1297, 1297t
  - actions of, 1167, 1178–1179, 1182t, 1183f
  - adverse effects of, 1171t, 1179, 1284
  - for *Enterobacterales* infections, 1284
  - indications for, 1174t, 1179
  - for *P. aeruginosa* infection, 1309t, 1312
  - resistance to, 1174t, 1182t, 1183f, 1185, 1268, 1283
  - for ventilator-associated pneumonia, 1032t
- Polyneuropathy. *See also* Peripheral neuropathy
  - acute inflammatory demyelinating, 3617–3618, 3617t, 3619t
  - acute toxic, 171
  - autoantibodies in, 2790t
  - chemotherapy-related, 756
  - chronic inflammatory demyelinating, 3619t, **3622**
  - critical illness, 2351, 3606
  - cryptogenic sensory and sensorimotor, 3613
  - diabetic. *See* Diabetic neuropathy
  - in diphtheria, 1224
  - familial amyloid, 3603
  - in Guillain-Barré syndrome. *See* Guillain-Barré syndrome (GBS)
  - hereditary sensory and autonomic, 3601
  - in HIV infection. *See* HIV infection, neurologic disease in
  - large-fiber, 176
  - in MGUS, 3623
  - in multiple myeloma, 3623
  - in POEMS syndrome, 894
  - sensory deficits in, 175–176
  - small-fiber, 176
  - weakness in, 171
- Polyomaviridae, 432, 1478f, 1480t, 1536
- Polyp(s)
  - colorectal
    - bleeding from, 318
    - endoscopic findings in, 2465, 2470f, 2471f
    - endoscopic treatment of, 2479–2480f, V5
    - in familial adenomatous polyposis, 2501f. *See also* Familial adenomatous polyposis
    - flat serrated, 2465, 2471f
    - pathogenesis of cancer in, 653. *See also* Colorectal cancer
  - gallbladder, 2734
- Polypharmacy, in older adults, 3884–3885, 3887t, 3888f
- Polypill, 1859
- Polyploidy, of mtDNA, 3777–3808
- Polyposis coli. *See* Familial adenomatous polyposis (polyposis coli)
- Polyposis, juvenile, 524t, 654t
- Polyradiculopathy
  - diabetic. *See* Diabetic neuropathy
  - in HIV infection. *See* HIV infection, neurologic disease in
- Polysaccharide A, 1073, 3840
- Polysaccharide antigens, 2802
- Polysaccharide capsule
  - meningococcal, 1244, 1244t
  - pneumococcal, 1187, 1188f
- Polysaccharide iron complexes, 767t
- Polysaccharide-protein conjugate vaccines, 997t
- Polysomnography, 210, 212, 214f, 2279, 2279f
- Poly-ubiquination, histone, 3937
- Polyuria
  - in AKI recovery, 2384
  - approach to the patient, 343–344, 343f
  - hypokalemic, 356
  - pathophysiology of, 343–344
  - treatment of, 353
  - in urinary tract obstruction, 2455
  - in water diuresis, S1
- Pomalidomide, 559f, 561, 753t, 887, 890t, 892, 1613
- POMC (proopiomelanocortin), 3068t, 3182, 3183f, 3185t
- *POMC* gene, 739, 2988, 2990, 3067
- *POMC* gene mutations, 3068t, 3185
- Pompe disease (acid α-1,4 glucosidase deficiency, type II GSD), 3354t, **3357,** 3361t, **3363,** 3644
- Ponatinib
  - actions and targets of, 532, 533t, 535, 567, 568t, 838
  - adverse effects of, 567, 568t, 754, 839t, 841
  - for ALL, 848, 848t
  - for CML, 838–841, 839t
  - genetic variations in response to, 567
  - in imatinib resistance, 841
  - interactions and issues, 568t
- Poncet's disease, 1058
- Ponesimod, 3586t, 3587
- Pongola virus, 1659t

Pontiac fever, 1272t, 1273. *See also Legionella* spp. infections
Pontine hemorrhage, 192, 3454. *See also* Intracerebral hemorrhage
Pontine stroke syndromes, 3434–3436f, A16
Poor metabolizers, 491, 492f
Popliteal artery
  aneurysm, 2176
  entrapment, 2176
  pulse, 1863f
Popliteal (Baker's) cyst, 2159, 2944
Population genetics, 3795, 3804–3805
Population migration, climate change and, 3747
Porcelain gallbladder, 2733
Pore-forming toxins, 974
POR gene mutations, 3071t, 3099, 3099t, 3101t
Porins, 1183, 1253, 1295
Pork tapeworm infection. *See Taenia solium* (pork tapeworm) infection
Porphyria(s), **3340**
  acute, 3343t, 3345
  classification of, 3340, 3341t, 3344–3345
  clinical features of, 3341–3342, 3341t, 3343t, 3344–3345
    abdominal pain, 111, 3602
    autonomic dysfunction, 3550
    cutaneous, 401
    neurologic, 3603
    photosensitivity, 430, 435
  cutaneous (chronic), 3343t, 3345
  diagnosis of, 3341–3342, 3343t, 3345
  differential diagnosis of, 3345
  erythropoietic
    clinical features of, 3341, 3341t, 3344
    congenital, 3342, **3349**
    diagnosis of, 3341t, 3343t
    pathophysiology of, 3341f
    protoporphyria. *See* Erythropoietic protoporphyria (EPP)
    X-linked protoporphyria. *See* X-linked protoporphyria (XLP)
    X-linked sideroblastic anemia, **3349**
  global considerations in, 3342
  hepatic
    acute intermittent porphyria. *See* Acute intermittent porphyria (AIP)
    ALA-dehydratase deficient, 3341f, 3341t, 3342, 3342t, 3345–3346
    clinical features of, 3340–3341, 3341t
    diagnosis of, 3341t, 3345
    hereditary coproporphyria. *See* Hereditary coproporphyria (HCP)
    pathophysiology of, 3341f
    porphyria cutanea tarda. *See* Porphyria cutanea tarda (PCT)
    unsafe drugs in, 3346
    variegate porphyria. *See* Variegate porphyria (VP)
Porphyria cutanea tarda (PCT), **3347**
  clinical features of, 435, 3343t, 3347–3348, A15
  diagnosis of, 3343t, 3348
  differential diagnosis of, 401
  genetic considerations in, 435, 3347
  HCC and, 3348
  HCV infection and, 2685
  iron overload in, 3335
  pathogenesis of, 3341f
  photosensitivity in, 435
  risk factors for, 3347–3348
  skin manifestations of
    blistering lesions, 3347, 3347f
    hyperpigmentation, 400, 435, A15
    vesicles/bullae, 401, 435, A15
  treatment of, 435, 3348
Porphyria variegata, 401
Porphyrinogens, 3343
Porphyrins, 430, 435, 555
Porphyromonas spp., 971, 1372, 1373
Porphyromonas spp. infections. *See also* Anaerobic bacterial infections, mixed
  in dog-bite wound, 1143
  in human bite-wound, 1144
  of lung, 1375
  orofacial, 1374
  rheumatoid arthritis and, 2843
Portal hypertension, 2708, **2713**
  classification of, 2713t
  clinical features of, 2713
    ascites. *See* Ascites
    esophageal varices. *See* Esophageal varices
    gastropathy, 2714
    splenomegaly and hypersplenism, 480, 2714–2715
  definition of, 2713
  diagnosis of, 2713–2714
  etiology of, 2713, 2713t
  in HCC, 660
  in hemochromatosis, 3337
  pulmonary arterial hypertension and, 2191–2192
Portopulmonary hypertension, 2717
Posaconazole
  adverse effects of, 355
  for aspergillosis, 1716, 1717t
  for blastomycosis, 1703
  for *Candida* infections, 1711t
  for coccidioidomycosis, 1698
  for cryptococcosis, 1706
  drug interactions of, 1692
  for eumycetoma, 1725
  for fungal sinusitis, 582
  for fusariosis, 1726t, 1729
  indications for, 1691–1692
  for mucormycosis, 1721t, 1722
  for phaeohyphomycoses, 1726t, 1728
  prophylactic uses
    aspergillosis, 1717t
    Candida, 1595t, 1712
    in neutropenia, 577, 581, 585, 1691
    in transplant recipients, 1155, 1156, 1162, 1691
  resistance to, 1711t
  for *Scedosporium* infections, 1726t, 1730
Positive end-expiratory pressure (PEEP), 2302, 2304, 2305f. *See also* Mechanical ventilation
Positive predictive value (PPV), 514, 514t
Positive-strand RNA viruses, 1477, 1478f, 1479, 1480t
Positron emission tomography (PET)
  in adrenocortical carcinoma, 3065, 3066f
  in carcinoma of unknown primary, 732
  in cardiac disease
    aortitis, A9
    heart failure, 1985
    infective endocarditis, 1040–1041, 1899, 1901f
    ischemic heart disease, 2094, 2095–2096f
    for myocardial perfusion imaging, 1879. *See also* Myocardial perfusion imaging
    Takayasu's arteritis, A14
  in fever of unknown origin, 153–154, 154f
  in lung cancer, 620
  in medullary thyroid carcinoma, 3086f
  in MEN 2A, 3086f
  in musculoskeletal disease evaluation, 2947
  in NETs, 681, 681f
  in neurologic disease, 3392
    AD, 3476–3477, 3477f
    cluster headache, 3465f
    dementia, 197–198, 198f
    DLB, 3492t
    intracerebral hemorrhage, 3455
    migraine, 3462, 3465f
    PD, 3495–3496, 3496f
    stroke, 3438–3439
  in osteomyelitis, 1062, 1063f
  in pancreatic cancer, 673, 673f
  in pheochromocytoma, 3074, 3075–3076f
  principles of, 3392
  in prostate cancer, 697, 698, 701
  in respiratory disease, 2213
  in tularemia, 1339, 1340f
  in tumor detection, 543
  in ZES, 2536
POST (posterior shear test), 121t
Post-acute infection syndromes, 1482
Postal workers, anthrax exposure in, S4
Postanginal septicemia. *See* Lemierre's disease/syndrome
Postantibiotic effect, 1168
Postcardiac injury syndrome, 2082, 2088
Postcoital contraception, 3154
Postconcussion syndrome/disorders, 3575–3576, S8
Posterior capsulotomy, 230
Posterior cerebral artery
  anatomy of, 3430f, 3432f
  lesion in territory of, 203, 207–208
  occlusion of, 222, 3431, 3431f, 3432f, 3436f
Posterior choroidal artery infarct, A16
Posterior column-medial lemniscal pathway, 173, 173f
Posterior cortical atrophy syndrome, 206, 206f
Posterior fossa brain tumor, 114
Posterior ischemic optic neuropathy, 227
Posterior parietal cortex, 204, 204f
Posterior pituitary
  disorders of
    diabetes insipidus. *See* Diabetes insipidus (DI)
    hypodipsic hypernatremia, **3019**
  hormone production in, 3014, 3015f. *See also* Arginine vasopressin (AVP)
Posterior reversible encephalopathy syndrome (PRES), 912, 2349–2350, 2350f, 2350t, 3447–3448, 3902
Posterior shear test (POST), 121t
Posterior spinal artery infarction, 3565
Postexertional malaise (PEM), 3649, 3650t
Postexposure prophylaxis (PEP)
  HAV, 2662
  HBV, 2663–2664
  HIV, 1564, 1625–1626
  measles, 1642
  rabies, 1654–1655, 1654f
  syphilis, 1429
  VZV, 1507, 1508t, 1593t
Postextubation stridor, 2307
Postherpetic neuralgia, 265, 1506, 1507, 3607
Postictal state, 189
Post–kala-azar dermal leishmaniasis, 1785–1786, 1786f
Postlaminectomy syndrome, 125t
Postmenopausal hormone therapy, **3143**
  adverse effects of, 940, 3302, 3303f
  approach to the patient, 3148–3149, 3148t
  benefits of, 3143–3145, 3144t, 3147f, 3165
  breast cancer risk and, 3145, 3303f
  cognitive function and, 3146, 3164
  colorectal cancer risk and, 513, 655, 3146
  contraindications to, 3148
  coronary heart disease and, 3145–3146, 3165, 3303f
  dyslipidemia and, 3246
  endometrial cancer and, 714, 3145, 3147
  fracture reduction due to, 3302, 3303f
  gallbladder disease and, 3145

  health status changes after discontinuation of, 3147–3148
  options for, 3148–3149
  ovarian cancer and, 3147
  risk-benefit profile, 3144t, 3147, 3147f
  short-term vs. long-term use of, 3148–3149
  stroke risk and, 3302, 3303f
  VTE and, 3145, 3302, 3303f
Postnasal drip, 273
Postoperative period
  AKI in, 2376
  nausea and vomiting in, 298
Postparalytic syndrome, 2295
Postpartum period
  depression in, 3167, 3667
  DVT in, 3902
  hemorrhage in, 472
  infections in, 1467
  thyroiditis in, 3038–3039
Postpolio syndrome, 1634
Postprandial orthostatic hypotension, 3547
Post-premature ventricular contraction, 289
Poststreptococcal glomerulonephritis
  clinical features of, 2415t, 2417
  diagnosis of, 2355
  epidemiology of, 2416
  etiology of, 2416–2417
  pathophysiology of, 2354–2355, 2412, 2416–2417, 2417f
  renal biopsy in, 2417, A4
  treatment of, 2417
Posttest probability of disease, 24, 25f, 48
Postthrombotic syndrome, 2154, 2157, 2157f, 2157t
Posttransfusion purpura, 911
Posttranslational processing, 3780f
Posttransplant lymphoproliferative disorder (PTLD), 1161, 1512
Posttraumatic stress disorder (PTSD), 3661–3663, 3662–3663t, S8
Postural control, 177
Postural hypotension. *See* Orthostatic hypotension
Postural orthostatic tachycardia syndrome (POTS), 1936–1937, 3548
Postvoid residual, 3891
Postwar syndromes, S8
POT1 gene mutations, 599, 3820
Potassium, **353**
  in cardiac contraction, 1849
  depletion of, 354–355, 374
  excess intake of, 358
  intake, hypertension and, 2137t, 2142
  recommended intake of, 2601t
  renal transport of, 2367f, 2369
  transport and reabsorption of, 345–346, 346f, 353–354
Potassium balance, 353–354, 2389
Potassium bicarbonate, 356
Potassium-binding resins, 2390
Potassium channels
  ATP-sensitive, 3199, 3199f
  cardiac, 1911–1912, 1912f
  muscle, disorders of, 3647
  in seizure initiation, 3414
Potassium chloride (KCl)
  administration guidelines, 356–357
  adverse effects of, 356–357, 786t
  for DKA, S1
  for hypokalemia, 356, S1
  for hypokalemic periodic paralysis, 3646
Potassium citrate, 356
Potassium-competitive acid pump antagonists, 2526
Potassium hydroxide (KOH) wet-mount preparation, 381, S11
Potassium iodide (KI)
  adverse effects of, 1724–1725
  prior to thyroidectomy, 3036
  for radioactive iodine exposure, S6
  for sporotrichosis, 1724–1725, 1724t
  for thyroid storm, 3037
Potassium phosphate, S1
Potassium-sparing diuretics, 2144t, 2370, 2390
Potato riddler's lung, 2228t
POTS (postural orthostatic tachycardia syndrome), 1936–1937, 3548
Pott's disease, 1058, 1075, 1388, 1388f
Pouchitis, 2565, 2567
Poultry worker's lung, 2228t, 2229
Poverty, health and, 3845. *See also* Low- and middle-income countries
Powassan virus, 1452, 1660t, 1671
Power law distribution, 3956, 3956f
Power of attorney for health care, 79
Poxviridae, 1478f, 1480t, 1481, 1517
Poxvirus infections
  clinical features of, 1519, 1520f, 1521f
  complications of, 1521
  diagnosis of, 1519
  differential diagnosis of, 1519
  epidemiology of, 1517–1518
  in humans, 1518t
  pathogenesis of, 1518–1519
  prevention of, 1521–1522
  prognosis of, 1521
  treatment of, 1521
Pozelimab, 2792t
PPARγ (peroxisome proliferator-activated receptor γ), 2564, 3214
PPD. *See* Purified protein derivative (PPD)/tuberculin skin test (TST)
PPHP (pseudopseudohypoparathyroidism), 3288, 3288t, 3783
PPIB gene mutations, 3324, 3325t, 3326
PPIs. *See* Proton pump inhibitors (PPIs)
PPKAR1A gene mutations, 3080t, 3088
PPNAD (primary pigmented nodular adrenal disease), 3056, 3088
PPomas (pancreatic polypeptide-secreting tumors), 3082–3083
PPOX gene, 3342t
PPOX gene mutations, 3349
PPV (positive predictive value), 514, 514t
Practice guidelines, 4
PRAD 1, 3276f
Prader-Willi syndrome
  clinical features of, 2993, 3113, 3184, 3184t, 3791
  genetic considerations in
    deletions, 2993
    imprinting, 3113, 3791, 3939, 3941
    uniparental disomy, 3791
    X-inactivation, 3791
Pragmatics, 203
Pralatrexate, 867
2-Pralidoxime chloride (2-PAM Cl), 3712t, S5
Pralsetinib, 533t, 628
Pramipexole
  adverse effects of, 3502
  for PD, 3502, 3504t
  for restless legs syndrome, 213t, 217, 3518
Pramlintide, 3211t, 3212
Prasterone, 3145, 3149
Prasugrel
  actions of, 943
  adverse effects of, 473, 944
  for cardiac catheterization pretreatment, 1904
  vs. clopidogrel, 943–944, 2110
  contraindications to, 2110
  dosage of, 944, S13
  genetic variation in response to, 939t
  for ischemic heart disease, 2102
  management before endoscopic procedures, 2483t
  for NSTE-ACS, 2110, 2110t, 2112
  in PCI, 2104, 2110, 2127, 2128
  resistance to, 944
  for STEMI, 2121
Pravastatin, 2940t, 3244t. *See also* Statins
Prazepam, 3660t
Praziquantel
  actions of, 1751
  adverse effects of, 1745t, 1751
  for cysticercosis, 1834
  for diphyllobothriasis, 1839
  for echinococcosis prevention, 1838
  for food-borne trematode infections, 1830t
  indications for, 1745t
  for neurocysticercosis, 1138
  pharmacology of, 1751
  in pregnancy and lactation, 1745t
  for *R. nana* infections, 1838
  for schistosomiasis, 1751, 1830, 1830t
  for *Taenia* spp. infections, 1834
Prazosin
  for ergot alkaloid overdosage/poisoning, 3710t
  for pheochromocytoma, 3075
  for Raynaud's phenomenon, 2178, 2874
PRC (phase response curve), 3944t, 3946
PRDM5 gene mutations, 3329t, 3330
Prealbumin. *See* Transthyretin (TTR)
Prebiotics, 2576
Pre-Bötzinger complex, 2273
Precipitation, patterns of, 3746–3747
Precision medicine
  applications of, 34–35
  in cancer treatment of, 527–528
  disease nosology and, 29–31, 31f
  drug repurposing using, 32, 34f
  future development of, 35–36, 36f
  in genetic disease, 3777, 3795–3796
  molecular and phenotypic subtyping and, 3962, 3962f
  overview of, 4
  requirements of, 31–34, 31f
Precocious puberty
  female, 712, 3130–3131, 3130–3131t
  male, 740, **3108,** 3109t, 3789. *See also* McCune-Albright syndrome
Preconceptual diagnosis, 3797
Prediabetes, 3196f, 3197
Prediction models, 26, 26t
Predictive value of test, 24
Predisposition testing, 3803
Prednisolone
  for alcohol-associated liver disease, 2699f, 2700
  for chorea, 2858
  for COPD exacerbation, 2259
  for gastric lymphoma, 650
  for hand osteoarthritis, 2955
  for neurocysticercosis, 1834
  for rheumatic fever, 2858
Prednisone
  for ACTH deficiency, 2997
  for acute idiopathic pericarditis, 2080
  for adult hypopituitarism, 2996t
  adverse effects of, 1734t, 2849, 2916t, 3622, 3648. *See also* Glucocorticoid(s), adverse effects of
  for asthma, 2224–2225
  for autoimmune hemolytic anemia, 802
  for autoimmune hepatitis, 2696
  for autoimmune pancreatitis, 2751–2752
  for Behçet syndrome, 2909
  for Bell's palsy, 3555
  for cancer treatment-induced pneumonitis, 754
  for chorea, 2858

- Prednisone (*Cont.*):
  - for CIDP, 3622
  - for cluster headache prevention, 3472, 3472t
  - for contact dermatitis, 384
  - for DIHS, 424
  - for DLBCL, 863
  - for EBV-associated disease, 1512
  - for eosinophilic fasciitis, 416
  - for epidermolysis bullosa acquisita, 413
  - for follicular lymphoma, 864
  - for giant cell arteritis, 2903
  - for gout, 2959
  - for granulomatosis with polyangiitis, 2898
  - for Graves' ophthalmopathy, 233, 3037
  - for GVHD, 917
  - for hirsutism, 3141
  - for hypercalcemia, 365
  - for hypercalcemia of malignancy, 738
  - for hypereosinophilic syndrome, 880
  - for hypersensitivity pneumonitis, 2230
  - for IBD, 2564
  - for idiopathic cutaneous vasculitis, 2906
  - for IgA vasculitis, 2905
  - for IgG4-related disease, 2933
  - for immune thrombocytopenic purpura, 924
  - for immunosuppression, 2408
  - for inflammatory myopathies, 2915–2916, 2916t
  - for leprosy reactions, 1411–1412
  - for loiasis, 1824
  - for minimal change disease, 2421
  - for MS, 3585
  - for mucous membrane pemphigoid, 413
  - for multiple myeloma, 890t, 891
  - for myasthenia gravis, 3630
  - for neurocysticercosis, 1138
  - for neuromyelitis optica, 3592, 3593
  - for optic neuritis, 227
  - for PCP, 1734t
  - for pemphigoid gestationis, 412
  - for pemphigus vulgaris, 410
  - for polymyalgia rheumatica, 2903
  - for postherpetic neuralgia, 1507
  - for relapsing polychondritis, 2921
  - for rheumatic fever, 2858
  - for rheumatoid arthritis, 2849
  - for serum sickness, 3721
  - for sinusitis with hyposmia, 243
  - for SJS/TEN, 425
  - for subacute thyroiditis, 3038
  - for systemic sclerosis, 2875
  - for trichinellosis, 1812t
- Preeclampsia, **3898**
  - clinical features of, 2140, 3899
  - definition of, 3898–3899
  - diagnosis of, 3899
  - in gestational trophoblastic disease, 715
  - KIRs with, 2772t
  - pathophysiology of, 3899
  - risk factors for, 3899
  - systems biology applied to, 3961t
  - treatment of, 3899
- Preexcited tachycardias, 1932t, 1939
- Prefrontal cortex, 208
- Prefrontal network, 208
- Pregabalin
  - adverse effects of, 3604t
  - for anxiety disorders, 3661
  - for cough, 274
  - for fibromyalgia, 2965, 2966t
  - for menopausal symptoms, 3143
  - for neuropathy, 3227, 3604t
  - for pain, 82, 97t, 100, 506, 3588
  - for pancreatitis pain, 2753
  - for restless legs syndrome, 217
  - for status epilepticus, 3426f
- Pregnancy, **3898**
  - abdominal swelling in, 326–327
  - acute fatty liver of, 3903
  - alcohol use in, 3677
  - antibacterial therapy in, 1168, 1169t
  - antiparasitic agents in, 1741–1746t
  - antiphospholipid syndrome in, 2837, 2838t, 2839
  - antithrombotic prophylaxis in, S13
  - aortic regurgitation in, 3900
  - aortic stenosis in, 3900
  - aplastic anemia in, 810
  - appendicitis in, 2596
  - asthma in, 2227
  - asymptomatic bacteriuria in, 1086, 3903
  - AVP deficiency in, 352, 3016, 3017t
  - bacterial infections in, 3903
  - bacterial vaginosis in, 1103
  - breast cancer treatment during, 638
  - cancer chemotherapy during, 576, 3902
  - cancer diagnosis in, 3989
  - cannabis/marijuana use during, 3687
  - cardiomyopathy in, 2017, 3900
  - cholestasis in, 325, 2729, 3903
  - chorea in, 3514
  - in CKD, 2394, 3899
  - climate change and adverse outcomes in, 3740
  - CML treatment during, 842–843
  - CMV infection in, 1513, 3903
  - congenital heart disease and, 2068–2069, 2069t, 3900
  - COVID-19 disease in, 3903
  - *C. trachomatis* infection in, 1472, 1474
  - depression in, 3167
  - diabetes insipidus in, 352
  - in diabetes mellitus, 3900–3901
  - diabetes mellitus in, 3165, 3196, 3221
  - DIC in, 934
  - direct oral anticoagulants in, 954
  - DVT in, 3902
  - ectopic, 112t, 113, 3136
  - edema in, 282
  - epilepsy and, 3427, 3903
  - erythema infectiosum in, 146
  - fatigue in, 167
  - filovirid infections in, 1686
  - folate deficiency in, 784, 788, 790
  - folate requirement in, 788, 790, 1844
  - gallbladder disease in, 2729, 3903
  - gastrointestinal disease in, 3903
  - GBS infection in, 1213
  - genital herpes in, 3903–3904
  - gonococcal infections in, 1256–1257
  - Graves' disease in, 3036–3037
  - HBV infection in, 3903–3904
  - HBV treatment in, 2683
  - HCV treatment in, 2693
  - heart disease in, 3899–3900
  - heart murmurs in, 288
  - HELLP syndrome in, 2446–2447, 2660, 3899, 3903–3904
  - hematologic disease in, 3902
  - hemoglobin disorders screening in, 776
  - hemoglobin/hematocrit in, 445t
  - hemoglobinopathy in, 3902
  - in hemophilia carriers, 932
  - herpesvirus infection in, 3903–3904
  - high-altitude travel during, 3759
  - HIV infection in, 1565, 3904
  - hormonal integration of, 2986
  - HSV infection in, 1501, 1503t, 1504, 3903–3904
  - hydronephrosis in, 2454
  - hypertension in, 2138t, 2139–2140, 3898–3899
  - hyperthyroidism in, 3901
  - hypothyroidism in, 3023, 3031–3032, 3901–3902
  - IBD and, 2569–2570
  - immunization during, 1004t, 1005, 3904
  - influenza in, 1547, 1550
  - international travel during, 1021
  - iron requirement in, 765t
  - lead poisoning in, 3701
  - liver disease in, 3903
  - *L. monocytogenes* infections in, 1229
  - Lyme disease in, 1453
  - malaria in, 1765, 1765t
  - malaria prevention in, 1775
  - in Marfan's syndrome, 3900
  - maternal mortality, 3904
  - mercury consumption in, 3701
  - metabolic changes in, 3900
  - in military veterans, S8
  - mitral regurgitation in, 3900
  - mitral stenosis in, 2051, 3900
  - molar. *See* Gestational trophoblastic disease
  - in MS, 3588–3589, 3903
  - mumps in, 1648
  - nausea and vomiting in, 298, 300, 2609, 3903
  - neurologic disorders in, 3902–3903
  - nutrient requirements in, 2600–2602t
  - obesity and, 3901
  - oral disease in, 261
  - parvovirus B19 infection in, 1524, 3904
  - pemphigoid gestationis in, 410t, 412
  - peripheral nerve disorders in, 3902–3903
  - phenylketonuria in, 3370
  - pheochromocytoma in, 3077
  - prolactinoma in, 3007–3008
  - psoriasis in, 386
  - pulmonary embolism in, 3902
  - pulmonary hypertension in, 3899
  - Q fever in, 1462, 1464
  - respiratory alkalosis in, 376
  - retinoic acid in, 2612
  - rheumatoid arthritis in, 2853
  - rubella in, 137, 1644–1645, 3904. *See also* Rubella, congenital
  - in sickle cell disease, 774t
  - in SLE, 2836
  - smoking effects in, 3684
  - Sweet syndrome in, 144t
  - syphilis in, 1432, 1434t, 1435
  - in systemic sclerosis, 2871
  - taste function in, 242
  - TB treatment in, 1395t, 1400
  - *T. cruzi* infection in, 1788, 1790, 1790t
  - tetanus in, 1230
  - tetanus vaccine in, 1232
  - *T. gondii* infection in, 1798, 1800–1801, 1802
  - thiazide diuretics in, 458
  - thrombocytopenia in, 3902
  - thyroid function during, 3023, 3901
  - tularemia in, 1341
  - tumor markers in, 504t
  - UTIs in, 1094–1095
  - varicella in, 1505
  - viral infections in, 3903–3904
  - vitamin A deficiency in, 1844
  - vitamin A in, 2612
  - vitamin A toxicity in, 2612
  - VTE prophylaxis in, S13
  - VTE risk in, 940
  - warfarin in, 952
  - water requirement in, 2599
  - in Wilson's disease, 3340
  - Zika virus infection in, 140t, 1021, 1668
- Pregnenolone, 3051f, 3052–3053
- Prehypertension, 3258
- Preimplantation genetic diagnosis, 3816
- Prejudice, 67
- Prekallikrein deficiency, 928t
- Preload, 1850, 1852t, 1853f, 2310

Premature atrial contractions, palpitations in, 291
Premature closure, in clinical reasoning, 21
Premature ejaculation, 3155
Premature menopause. *See* Primary ovarian insufficiency
Premature ovarian failure. *See* Primary ovarian insufficiency
Premature ventricular contractions (PVCs)
  in acute coronary syndromes, 1961
  in acute illness, 1960
  characteristics of, 1960
  in congenital heart disease, 1961
  definition of, 1954
  diagnosis of, 1960
  etiology of, 1960
  in heart failure, 1961, 1985
  in hypertrophic cardiomyopathy, 1961
  idiopathic, 1960
  idioventricular rhythms, 1955f, 1961, 1962f
  mechanisms of, 1913, 1913t
  multifocal, 1954, 1955f
  palpitations in, 291
  treatment of, 1960–1961, 1962f
  unifocal, 1954, 1955f
  ventricular dysfunction and, 1961, 1962f
  VT/VF from, 1974, 1975f
Premunition, 1763
Prenatal diagnosis/testing
  amniocentesis, 3797
  chorionic villus sampling, 3797
  hemoglobin disorders, 776
  in mtDNA disorders, 3816
  noninvasive methods of, 3987–3989
Preoperative questionnaire, 3905, 3905t
Preosteoclasts, 2846
Prepatellar bursitis, 2973
PRES (posterior reversible encephalopathy syndrome), 2349–2350, 2350f, 2350t
Presatovir, 1489
Presbycusis, 246, 246f
Presbyopia, 220
Presbyosmia, 240
Presenilins, 3479
Presensitization, in transplantation, 2406–2407
Present-biased preferences, 3916t, 3917–3919
Preserved ratio–impaired spirometry (PRISm), 2254
Pre-S genes, of HBV, 2644, 2644f
Pressure-control ventilation (PCV), 2305–2306, 2305t, 2306f. *See also* Mechanical ventilation
Pressure-immobilization, for snakebite, 3718
Pressure-regulated volume control ventilation (PRVC), 2305t, 2306
Pressure-support ventilation (PSV), 2305t, 2306. *See also* Mechanical ventilation
Pressure (decubitus) ulcers, 1051, 3570
Pressure urticaria, 2807
Pressure-volume curve
  in ARDS, 2304, 2304–2305f
  of chest wall, 2202f
  of lung, 2201–2202, 2202f
  normal, 2304f
Presymptomatic testing, 3802
Presynaptic rundown, 3625
Presyncope, 155
Presystemic elimination, 484, 484f
Presystolic accentuation, 288
Pretest probability of disease, 24, 25f, 48, 48t
Pretomanid, 1395t, 1398–1399, 1421, 1421t, 1425
Pretransfusion testing, 905
Preventive care. *See also specific diseases*
  accident and suicide risk assessment in, 13
  in adult-onset genetic disorders, 3804t, 3805
  approaches to, 8–9
  approach to the patient, 13
  counseling in, 41t
  health behavior and lifestyle assessment in, 10
  health promotion in, 9
  healthy eating guidelines for, 10–11, 10t
  insurance coverage of, 3921
  mental health and addiction screening in, 13
  in older adults, 3878–3879, 3881t
  periodic health evaluations and, 9–10
  physical activity recommendations in, 11, 12t
  prioritizing strategies for, 9
  vs. screening, 9
  screening tests in, 37, 39t
  sleep hygiene, 11–12
  smoking cessation in, 13. *See also* Smoking cessation
  vaccination in, 13. *See also* Immunization
  vaccine counseling in, 18–19, 18t
  weight management in, 12–13
Prevotella spp., 1372, 2843, 3837
Prevotella spp. infections, 1102. *See also* Anaerobic bacterial infections, mixed
  in dog-bite wound, 1143
  of head, neck, and mouth, 1374
  in human bite-wound, 1144
  orofacial, 1374
PRF1 gene mutations, 465
Priapism, 774t, 3155
Prickly heat, 3774
Pridopidine, 3514
Primaquine
  actions of, 1751
  adverse effects of, 1734t, 1751, 1775
    in G6PD deficiency, 799, 799t, 1751
    genetic considerations in, 493
    major, 1771t
    minor, 1771t
    in parasitic infections, 1741t
  contraindications to, 1741t, 1770t, 1772, 1775
  drug interactions of, 1741t, 1745t
  genetic variations in response to, 493, E72t
  for malaria, 1770t, 1771t, 1772
  for malaria prophylaxis, 1774t, 1775
  overdosage/poisoning with, 3714t
  for PCP, 1734t
  pharmacology of, 1751, 1771t
Primary afferent(s), 93, 93f
Primary afferent nociceptors, 93, 93f, 94f
Primary alveolar proteinosis, A12, S9
Primary autonomic failure, 3548
Primary biliary cholangitis (PBC), 324, 400, 2695, **2711,** 2790t, 2825, A13
Primary care
  definition of, 3859
  duties of, 3
  revitalization of, 3863–3864, 3863f
  vaccine counseling and vaccination in, 18–19, 18t
  worldwide access to, 3859. *See also* Low- and middle-income countries
Primary Care PTSD (PC-PTSD) Screen, S8
Primary central nervous system lymphoma (PCNSL), **720**
  diagnosis of, 720–721
  EBV infection and, 857
  epidemiology of, 720, 857
  in HIV infection, 1613, 1614f
  imaging of, 720f, 1614f
  in immunocompromised patients, 720, 721
  treatment of, 721
Primary cough headache, 3473
Primary cutaneous CD30+ T cell lymphoma, 875
Primary effusion lymphoma, 873, 1516, 1613
Primary eosinophilia. *See* Eosinophilia
Primary exercise headache, 3473–3474
Primary graft dysfunction (PGD), 2286
Primary immune deficiency diseases (PIDs), **2793**
  of adaptive immune system, 2794t, **2797**
    B lymphocyte deficiencies. *See* B cell(s), deficiencies of
    T lymphocyte deficiencies. *See* T cell(s), deficiencies of
  B cell deficiencies in, 2801
  B cell differentiation in, 2801f
  classification of, 2794t
  colitis and, 2803–2804
  consequences of, 2794
  definition of, 2794
  diagnosis of, 2794, 2795t, 2801
  diseases associated with, S9
  effector pathways in, 2797, 2797f
  gene therapy for, 3825–3826
  genetic considerations in, 2794
  HCT for, 918, 1164
  of innate immune system, **2794**
    asplenia, 482, 1004t, 2795–2796, S9
    chronic granulomatous diseases. *See* Chronic granulomatous diseases (CGDs)
    complement deficiency, 2782, 2783t, 2797
    diseases associated with, S9
    GATA2 deficiency. *See* GATA2 deficiency
    leukocyte adhesion deficiency, 458, 459t, 460, 2796
    Mendelian susceptibility to mycobacterial disease, 2796
    severe congenital neutropenia, 457, 2794–2795, 2795f, S9
    toll-like receptor pathway deficiencies, 2796–2797
  lymphoid malignancies and, 858t
  phagocytic cells in, 2794, 2795f
  regulatory defects and, 2794t, 2803
    in autoimmune lymphoproliferative syndrome, 524t, 2803
    in autoimmune polyendocrinopathy-candidiasis-ectodermal dysplasia syndrome. *See* Autoimmune polyendocrine (polyglandular) syndrome-1 (APS-1)
    in hemophagocytic lymphohistiocytosis, 2803
    in immunodysregulation polyendocrinopathy X-linked syndrome, 2789, 2803
    in PLCγ-associated antibody deficiency and immune dysregulation, 2804
  T cell differentiation in, 2797f, 2799
  variable expression of, 2794
Primary intraventricular hemorrhage, 3455
Primary lateral sclerosis, 3525, 3569
Primary myelofibrosis, **821**
  clinical features of, 821–822
  complications of, 822–823
  diagnosis of, 440f, 441f, 446f, 821f, 822, 822f
  diseases associated with, 821t
  etiology of, 821, 821t
  genetic considerations in, 819, 821
  prognosis of, 822–823, 822t
  splenomegaly in, 480, 821
  treatment of, 823
Primary neuritic leprosy, 1408
Primary olfactory cortex, 237, 238f
Primary ovarian failure, 3789
Primary ovarian insufficiency, 3134
Primary peritoneal serous carcinoma, 735
Primary pigmented nodular adrenal disease (PPNAD), 3056, 3088
Primary polydipsia
  AVP stimulation in, 3015
  differential diagnosis of, 3017–3018, 3018f
  etiology of, 3016, 3017t
  pathophysiology of, 3017
  treatment of, 3019

- Primary prevention. *See* Preventive care
- Primary progressive aphasia, 201, 203–204, 206f, 3486. *See also* Frontotemporal dementia (FTD)
- Primary sclerosing cholangitis (PSC), **2711**
  - cholangiocarcinoma and, 667
  - clinical features of, 324, 326, 2563, 2712, 2737
  - diagnosis of
    - cholangiography in, 324, 2563, 2712, 2731f, 2737
    - laboratory tests in, 2630t, 2712
    - liver biopsy in, A13
    - ultrasound in, 324
  - diseases associated with, 2737
  - drug-induced, 2667
  - in HIV infection, 2737
  - in IBD, 2562t, 2563
  - IgG4-related, 2737, 2751, 2931, 2931t. *See also* IgG4-related disease (IgG4-RD)
  - natural history of, 2737
  - recurrence after liver transplantation, 2724
  - small duct, 2563, 2737
  - treatment of, 2712, 2737
- Primary sex headache, 3474
- Primary stabbing headache, 3474
- Primary taste cortex (PTC), 239
- Primary thunderclap headache, 3474
- Primate immunodeficiency viruses, 1557, 1558f, 1559. *See also* HIV
- Primidone
  - adverse effects of, 787t, 788, 3422t
  - for ataxia/tremor in MS, 3588
  - drug interactions of, 3422t
  - for myoclonus, 3516
  - overdosage/poisoning with, 3713t
  - for seizures, 3422t
  - for tremor, 3509
- Primitive neuroectodermal tumors (PNETs), 728, 730
- Primitive reflexes, 3381
- Primordial prevention, 8–9
- Principal cells, 353, 2367f, 2370, S1
- PR interval, 1912
- Prinzmetal's variant angina (vasospastic angina), 1873, 2101, **2112**
- Prion(s)
  - decontamination of, 3532
  - definition of, 3525, 3527t
  - diseases due to. *See* Prion disease(s)
  - as drug target, 3401
  - strains of, 3527, 3528t
  - terminology related to, 3527t
  - transmissibility of, 3526–3527
- Prion disease(s), **3525**
  - Creutzfeldt-Jakob disease. *See* Creutzfeldt-Jakob disease
  - epidemiology of, 3526–3527
  - global considerations in, 3526–3527, 3527t
  - Huntington disease–like disorders, 3514–3515
  - inherited, 3528
  - neurodegenerative, 3399f, 3400–3401, 3401t, 3482–3483
  - pathogenesis of, 3399–3400, 3526t, 3527–3528, 3527f
  - research in, 3532–3533
  - species barrier in, 3527–3528
  - sporadic, 3528
  - types of, 3526, 3526t
- Prion proteins, 3525, 3527f
- Prion rod, 3526t
- PRISm (preserved ratio–impaired spirometry), 2254
- Pritelivir, 1487, 1502, 1503
- PRK (photoreactive keratectomy), 220
- *PRKACA-DNAJB1* gene mutations, 670
- *PRKAG2* gene mutations, 2002t, 2004t, 2010, 3365
- PRKAG2 syndrome, 3361t, 3365
- *PRKAR1A* gene mutations, 2085, 3004, 3004t, 3056
- *PRKG1* gene mutations, 3332t
- *PRKN* gene mutations, 3498t, 3499
- *PRKRA* gene mutations, 3510t, 3511
- *PRNP* gene, 195t, 3514, 3526t, 3527
- *PRNP* gene mutations, 3528, 3533
- Probenecid
  - adverse effects of, 323
  - with cidofovir, 1116
  - drug interactions of, 489t, 1745t
  - for gout, 2960
  - for neurosyphilis, 1434t
  - for syphilitic meningitis, 1127
- Probiotics
  - actions of, 3928
  - for diverticular disease, 2580
  - historical perspective, 3831
  - for *H. pylori* eradication, 2528
  - for IBS, 2576, 2577t
  - for indigestion, 302
  - for infectious diarrhea and IBS, 2464
  - for traveler's diarrhea prophylaxis, 1081
- Procainamide
  - actions of, 1916
  - adverse effects of
    - lupus syndrome, 489, 493, 2781, 2835, 2940t
    - in myasthenia gravis, 3627
    - pure red cell aplasia, 814t
    - SA node dysfunction, 1920t
    - thrombocytopenia, 923t
  - genetic variations in response to, 493, E72t
  - metabolism of, 485t
  - for preexcited tachycardias, 1941
  - for ventricular arrhythmias, 1958
  - for ventricular tachycardia, 2123
- Procalcitonin (PCT), 1025–1026
- Procarbazine
  - actions of, 563
  - adverse effects of, 562t, 563
    - DIHS, 597
    - male infertility, 3115
    - myelosuppression, 563, 574
    - neurologic, 726t
    - pulmonary, 595
  - for astrocytoma, 718
  - interactions and issues, 562t
- Procaspase-9, 540f
- Process (quality theory), 44
- Prochlorperazine
  - adverse effects of, 324
  - for migraine, 3467t, 3469
  - for nausea and vomiting, 83, 299t, 300, 509, 575
  - for scombroid poisoning, 3726
- Procollagen, 3322
- Proctitis
  - *C. trachomatis,* 1471, 1473t
  - HSV, 1498–1499, 1503t, 2560
  - sexually acquired, 1109, 1257, 1471
  - ulcerative colitis and, 2556–2557
- Proctocolitis, sexually acquired, 1109
- Proctopathy, radiation, V5
- Proctosigmoidoscopy. *See* Sigmoidoscopy
- Prodrugs, 486
- Progenitor cells, 760f
- Progeria, 3787, 3788t
- Progesterone, 376, 510, 2996t, 3128, 3135
- Progesterone receptors, 733t
- Progestins
  - adverse effects of, 282t
  - for contraception, 3153–3154, 3153t
  - hypersensitivity to, 2819
  - for osteoporosis management/prevention, 3303, 3303f
  - for PCOS, 3135
  - in perimenopause, 3142
  - in postmenopausal hormone therapy. *See* Postmenopausal hormone therapy
- Proglottids, 1810
- Prognathism, 261
- Programmed aging, 3869
- Programmed cell death. *See* Apoptosis
- Progranulin, 3395, 3486
- Progressive bulbar palsy, 3559
- Progressive diaphyseal dysplasia (Camurati-Engelmann disease), 3315–3316
- Progressive external ophthalmoplegia (PEO)
  - clinical features of, 234, 3645, 3811t, 3812, V3
  - diagnosis of, 3645
  - differential diagnosis of, 3628
  - genetic considerations in, 3645, 3811t, 3812, 3814f
- Progressive familial intrahepatic cholestasis, 325, 2641t, 2642
- Progressive massive fibrosis, 2238
- Progressive multifocal leukoencephalopathy (PML), **1116**
  - chemotherapy-related, 757
  - clinical features of, 1116–1117
  - diagnosis of, 1117
  - in HIV infection, 1116, 1609
  - JC virus and, 1117, 1161
  - natalizumab and, 2566
  - pathology of, 1116–1117
  - rituximab and, 2851
  - in solid organ transplant recipient, 1161
  - TNF-α inhibitors and, 387
  - in transplant recipient, 2351
  - treatment of, 1117
- Progressive muscular atrophy. *See* Spinal muscular atrophy (SMA)
- Progressive myoclonus epilepsy, 3412t
- Progressive rubella panencephalitis, 1118
- Progressive supranuclear palsy syndrome
  - clinical features of, 196, 3487, 3497, 3497t
  - eye movements in, 3487, 3497
  - gait disorders in, 178, 3497
  - neuropathology of, 3487–3488, 3487f, 3488f, 3497
- Progressive systemic sclerosis. *See* Systemic sclerosis (scleroderma)
- Proguanil
  - actions of, 1751
  - adverse effects of, 1743t, 1771t, 1775
  - drug interactions of, 1743t
  - for malaria, 1771t
  - for malaria prophylaxis, 1774t, 1775
  - pharmacology of, 1751, 1771t
  - in pregnancy and lactation, 1743t, 1775
  - resistance to, 1775
- Prohormone convertase 1 (PCSK1), 3185, 3185t
- Proinsulin, 3235
- Project BioShield, S4
- *PROK2* gene mutations, 3111t, 3112
- *PROK2R* gene mutations, 3111t, 3112
- Prokinetic agents, 299t, 302
- Prolactin, **2988**
  - actions of, 2989
  - expression and regulation of, 2988t
  - laboratory evaluation of, 2995t
  - secretion of, 212, 2988–2989
  - synthesis of, 2988
- Prolactinoma, 3001t, **3005,** 3006f, 3007f, 3080t, 3083
- Prolidase deficiency, 3368t
- Proline metabolism disorders, 3368t
- Prolyl hydroxylase inhibitors, 769
- Promethazine, 165t, 434t
- PROMM (proximal myotonic myopathy), 3642
- Promonocytes, 461
- Promoter (carcinogenesis), 512

Promyelocytes, 453, 454f
Promyelocytic leukemia. *See* Acute promyelocytic leukemia
Pronator drift, 3381
Prone positioning, in ARDS, 2302, 2307
Pronormoblasts, 443, 814, 814f
Pronouns, 3179t
Pronuclear transfer, 3816, 3817f
Pronucleating factors, 2729
Proopiomelanocortin (POMC), 3068t, 3182, 3183f, 3185t
Prop-1 (PROP-1)
  in adrenal insufficiency, 2992
  in anterior pituitary development, 2988t, 2992
  in congenital hypothyroidism, 3028t
  in GH deficiency, 2994
PROP1 gene mutations, 3068t, 3112t, 3113
Propafenone
  actions of, 1916t
  adverse effects of, 1916t, 1920t, 1952
  for atrial fibrillation, 1952
  for focal atrial tachycardia, 1938
  genetic variations in response to, 492, E72t
  metabolism of, 485t
Propantheline bromide, 3588
Prophylactic cranial irradiation, 631
Propionibacterium acne, 2892
Propionic acidemia, 3370t
Proplatelets, 937–938
Propofol, 84, 86, 86t, 1231, 1920t, 3710t, 3713t, 3816
Propoxyphene, 3715t
Propranolol
  for adrenergic symptoms in Graves' disease, 3036
  adverse effects of, 3024t, 3470t
  for aortic dissection, 2170
  for ataxia/tremor in MS, 3588
  for consumptive hypothyroidism, 740
  for drug-induced movement disorders, 3517
  for hypokalemia, 356
  for ischemic heart disease, 2100t
  for long QT syndrome, 1970
  for methylxanthine overdosage/poisoning, 3710t
  for migraine prevention, 3469, 3470t
  overdosage/poisoning with, 3712t
  for pheochromocytoma, 3075
  for phobic disorders, 3661
  for postpartum thyroiditis, 3039
  for sex headache prevention, 3474
  for sympathomimetic overdosage/poisoning, 3710t
  for tremor, 3509, 3660t
Proprioception, 172, 173
Proptosis, 232–233, 3034, A14. *See also* Graves' (thyroid-associated) ophthalmopathy
Propylene glycol poisoning, 371, S1
Propylthiouracil
  adverse effects of, 394, 426, 2907, 2940t, 3035
  for Graves' disease, 3035
  for hyperthyroidism in pregnancy, 3901
  for radioactive iodine exposure, S6
  for thyroid storm, 3037
Prosody, 203
Prosopagnosia, 206
Prospect theory, 3916, 3916t, 3917
Prostacyclin, 1845, 2187f, 2193–2194, 2285, 2517f
Prostaglandin(s)
  in asthma, 2219
  in drug reactions, 417
  in dysmenorrhea, 3137
  in gastroduodenal mucosal defense and repair, 2516f, 2517
  in heart failure, 1980
  in male sexual response, 3155
  pain and, 93, 94f
  for Raynaud's phenomenon, 2874
  synthesis of, 2517
Prostaglandin analogues
  adverse effects of, 2526
  for glaucoma, 225, 231
  for peptic ulcer disease, 2524t, 2526
Prostaglandin E_2 (PGE_2)
  ectopic production of, 737t
  in fever, 135, 136
  in gastric epithelia defense and repair, 2517
  in photoimmunosuppression, 430
  in sunburn erythema, 430
  synthesis of, 2517f
Prostate, 695
Prostate biopsy, 697
Prostate cancer, **695**
  anemia in, 450
  autopsy-detected, 695
  bone lesions in, 450, 703
  carcinoma of unknown primary and, 735
  clinical states model of, 695–696, 696f
  deaths from, 497t, 499f, 500t, 501t, 695, 3175
  diagnosis of, 695–699
  epidemiology of, 695
  erectile dysfunction after treatment of, 3155
  genetic considerations in, 520t, 524t, 695, 3982
  incidence of, 497t, 498, 498f, 501t, 695, 3175
  lymphedema in, 2183
  metastatic, 723, 723t, 730
  paraneoplastic syndromes in, 737t, 739
  pathology of, 697–698
  prevention of, 513, 696, 2612
  prostate-specific antigen in, 696–697, 699
  screening for, 38, 40, 517t, 518, 696–697, 3881t
  staging of, 498–699, 698f, 698t
  testosterone deficiency in, 3175
  testosterone therapy and, 3122, 3175
  treatment of
    vs. active surveillance, 518, 700–701
    adverse effects of, 3175
    androgen deprivation therapy. *See* Androgen deprivation therapy (ADT)
    cell-based, 2784
    hormone receptor–directed agents, 567, 701–702
    hot flashes and, 510
    late consequences of, 756
    in localized disease, 699–701
    in metastatic disease, 701–703
    pain management, 703
    prostatectomy, 699
    radiation therapy, 699–700
    rising PSA after, 701
    tumor vaccines, 561
  tumor markers in, 504t
Prostatectomy, 699
Prostate disorders
  benign hyperplasia. *See* Benign prostatic hyperplasia (BPH)
  cancer. *See* Prostate cancer
  prostatitis. *See* Prostatitis
Prostate Health Index (PHI), 697
Prostate-specific antigen (PSA)
  in carcinoma of unknown primary, 732, 733t, 735
  free, 696
  increasing, after prostate cancer treatment, 701
  in prostate cancer, 696–697, 699
  for prostate cancer screening, 517t, 518, 696–697
  testosterone therapy and, 3122
  as tumor marker, 504t
Prostate-specific membrane antigen (PMSA) PET scans, 698, 701
Prostatitis
  C. glucuronolyticum, 1226
  classification of, 1087
  clinical features of, 1089
  definition of, 1086
  diagnosis of, 1089
  enterococcal, 1217
  treatment of, 1094
  tumor markers in, 504t. *See also* Urinary tract infections (UTIs)
Prosthetic heart valve
  abscess in, A9
  anticoagulant therapy with, 3445t, 3450–3451, S13
  cinefluoroscopy of, 1907–1908, 1908f
  dental procedures and, 266
  endocarditis in. *See* Prosthetic valve endocarditis (PVE)
  heart sounds/murmurs in, 1867
  in pregnancy, 3900
Prosthetic joint infections
  classification of, 1066
  clinical features of, 1060, 1066, 1066f
  dental procedures and, 266
  diagnosis of, 1060, 1066–1067
  enterococcal, 1218
  epidemiology of, 1066
  global considerations in, 1067
  gram-negative bacteria in, 1268
  microbiology of, 1066
  pathogenesis of, 1066
  prevention of, 1060, 1067
  treatment of, 1060, 1065t, 1067
Prosthetic valve endocarditis (PVE)
  diagnosis of, 1039, 1040–1041
  etiology of, 1037, 1037t
  incidence of, 1037, 1037t
  late-onset, 1039
  staphylococcal, 1200–1201
  in transcatheter implanted aortic valve. *See* Transcatheter aortic valve implantation (TAVI)
  treatment of, 1043t, 1044, 1045, 1046, 1205
Protamine sulfate, 948, 2163
Protease activation receptor 1 (PAR1), 938
Protease inhibitors. *See* Combination antiretroviral therapy (cART)
Proteasome, 573, 2776
Proteasome inhibitors. *See* Protein homeostasis modulators (proteasome inhibitors)
Protein(s)
  aggregation of, in neurologic disease, 3399–3400
  digestion and absorption of, 2542
  energy deficiency and, 2619–2620
  lipoprotein-associated, 3238–3239
  prion, 3525
  quality control mechanisms. *See* Proteostasis network
  recommended intake of, 2602t
  requirement for, 2598, 2619–2620
  secretory, 2779
  serum binding, 3023–3024
  TAP, 2776
  thyroid hormone binding, 3023–3024
Proteinase 3, 2790t, 2894
Protein C, 470–471, 471f, 475
Protein-calorie malnutrition, 2616
Protein-catabolic intensity, 2621–2622
Protein C deficiency
  DIC in, 933
  high-altitude symptoms of, 3758
  thrombosis risk in, 940
  warfarin-induced skin necrosis in, 407, 421, 952
Protein-enhanced formulas, for enteral nutrition, 2622

Protein folding disorders, **3990**
  with aggregation propensity, 3994
  aging and, 3992–3993, 3992f, 3994
  with enhanced misfolding, 3993. *See also* Cystic fibrosis (CF)
  infections and, 3994
  proteostasis network and, 3990, 3991f
  with secretion of aggregated protein, 3994. *See also* Amyloidosis
  stress and, 3994
  with toxic aggregates and loss of function, 3993–3994. *See also* A_1-Antitrypsin (AAT) deficiency
  types of, 3990, 3993
Protein homeostasis modulators (proteasome inhibitors)
  action and targets of, 536, 540f, 557f, 570t, 573
  adverse effects of, 570t, 573, 595, 753t, 757, 2017
  interactions and issues, 570t
  for multiple myeloma, 890t, 891–892
Protein kinase A, 1847f, 1849, 3052, 3689, 3689f
Protein kinase Cδ (PKCδ) deficiency, 2786f
Protein kinase G, 1847f
Protein-losing enteropathy, **2548**
Protein S, 471, 471f, 475, 940, 952
Protein S deficiency, 421
Protein synthesis inhibitors, 1166, 1182t, 1183f
Protein tyrosine phosphatase, 532
Proteinuria, **341**
  in AKI, 2378
  approach to the patient, 341–342, 342f
  in diabetic nephropathy, 2424
  in glomerular disease, 2414, 2414t, 2415–2416t
  nephrotic-range, 2414
  pathophysiology of, 342, 2357, 2414
  in preeclampsia, 3899
  in SLE. *See* Lupus nephritis
Proteoglycans, 417, 3323
Proteolytic enzymes, 2743
Proteomics, 6, 3778
Proteostasis network
  in aging and disease, 3871, 3873f, 3992–3993
  in cancer, 3994
  control of protein synthesis by, 3990–3991, 3991f
  in pneumonia, 3994
  in protein folding diseases. *See* Protein folding disorders
Proteus spp., 1179, 1184, 1284, 1291
Proteus spp. infections, **1291**
  clinical features of, 1291
  diagnosis of, 1291
  epidemiology of, 1281
  renal abscess, 1075
  treatment of, 1291
  urinary tract, 2453
  UTI, 1087, 1291
Prothionamide, 1394t, 1395t, 1399
Prothrombin, 471f
Prothrombin complex concentrate (PCC), 930, 933, 935, 954, 3455–3456
Prothrombin deficiency, 473, 928t
Prothrombin gene mutations, 473, 941
Prothrombin time (PT), 928
  in acute viral hepatitis, 2656
  causes of abnormal values, 475t
  in coagulation disorders, 928t
  coagulation factors measured in, 474, 474f, 929f
  in DIC, 934, 934t
  elevated, 324
  in hepatobiliary diseases, 2636, 2637t
  for monitoring warfarin therapy, 951
  prolonged, 475t
  sensitivity of, 474
Proton pump inhibitors (PPIs)
  adverse effects of, 301, 420, 2511, 2525, 2526f
    acute interstitial nephritis, 2356, 2438
    cobalamin malabsorption, 786t
    musculoskeletal, 2940t
  for diarrhea, 311–312
  drug interactions of, 489t, 1742t, 2525–2526
  for gastrinoma, 678, 678t
  for GERD, 301, 2511, 2874
  in *H. pylori* eradication, 1302t, 2527t, 2530t
  for metabolic alkalosis, 375
  for NSAID-induced ulcer prophylaxis, 2530, 2530t
  for peptic ulcer disease, 317, 318, 2524–2526, 2524t, 2530
  pharmacology of, 2524–2525
  for stress-related mucosal injury, 2537
  for systemic mastocytosis, 2821
  for ZES, 2536
PROTO-oxidase, 3341t, 3342t, 3343, 3349
Protoparvovirus, 1525
Protoporphyria, erythropoietic. *See* Erythropoietic protoporphyria
Protoporphyrins, 3345
Protozoa/protozoal infections, **1737**
  blood and tissue, 1737–1738
    Babesia. See Babesia spp.
    Leishmania. See Leishmania spp.
    Plasmodium. See Plasmodium spp.
    Toxoplasma gondii. See Toxoplasma gondii
    Trypanosoma. See Trypanosoma spp.
  free-living amebae, 1737, 1758
    Acanthamoeba. See Acanthamoeba spp.
    Balamuthia. See Balamuthia spp.
    Naegleria. See Naegleria infection
  vs. helminthic parasites, 1809
  intestinal, 1737
    Cryptosporidium. See Cryptosporidium/cryptosporidiosis
    Cyclospora cayetanensis. See Cyclospora cayetanensis/cyclosporiasis
    Cystoisospora belli. See Cystoisospora belli/cystoisosporiasis
    Entamoeba histolytica. See Entamoeba histolytica
    Giardia. See Giardia spp.
    Microsporida. *See* Microsporida/microsporidiosis
  laboratory diagnosis of, **S12**
  by organ system and signs/systems, 1738–1740t
Protriptyline, 215, 3659t
Providencia infections, 1293, 2453
Provirus, 1551
PROX1 gene, 1582t
Proximal interphalangeal (PIP) joint, rheumatoid arthritis in, 2839–2840
Proximal myotonic myopathy (PROMM), 3642
Proximal tubule, 2366–2369, 2366f, 2368t
Proximal weakness, 171
Prozone effect/phenomenon, 1433, 2932
PrP amyloid, 3526t
PrP^{c}, 3526, 3526t, 3527, 3527f
PrP^{Sc}, 3526, 3526t, 3527, 3527f, 3529, 3530
PrP^{Sc} prions, 195t, 3400, 3401t. *See also* Prion disease(s)
PRRs. *See* Pattern recognition receptors (PRRs)
PRSS1 gene mutations, 671t, 2744–2745, 2751, 2754
PR (partial response), to chemotherapy, 555
Prucalopride, 299, 299t, 302, 314, 2593, 3549
Prune juice, 82t
Pruno, 1234t, 1235
Pruritus
  in atopic dermatitis, 383
  in CKD, 2362, 2394
  definition of, 378t
  drug-induced, 422
  in liver disease, 2628
  in pancreatic cancer, 673
  in polycythemia vera, 820
  in primary biliary cholangitis, 2711
  in scabies, 3729
  treatment of, 383
Prussian blue, 3702, A6, S6
PRVC (pressure-regulated volume control ventilation), 2305t, 2306
PSA. *See* Prostate-specific antigen (PSA)
PSAT1 deficiency, 3368t
PSC. *See* Primary sclerosing cholangitis (PSC)
P-selectin, 938
PSEN1/2 gene mutations, 2004t, 3479, 3941
Pseudoachalasia, 2507
Pseudoaldosteronism. *See* Liddle's syndrome
Pseudoallergenic fish poisoning, 3726
Pseudoaneurysm(s)
  aortic, 2165
  pancreatic, 2750
  pericardial, 2125
Pseudoappendicitis, 1348
Pseudoathetosis, 172
Pseudobradycardia, 1956
Pseudobulbar palsy, 169, 295
Pseudocholinesterase (BCHE), 485t, E72t
Pseudocowpox, 1518t
Pseudocyst(s), pancreatic, 2746t, 2749, 2753f
Pseudodementia, 3485, 3665
Pseudodiarrhea, 305
Pseudodiverticulosis, esophageal, 2506, 2506f
Pseudodominance, 3791
Pseudoephedrine
  adverse effects of, 354, 423
  for allergic rhinitis, 2810
  for bladder dysfunction, 3588
  contraindications to, 2810
  for upper respiratory infection, 253
Pseudogene, 3784
Pseudogout, 2960, 3336
Pseudohalitosis, 267
Pseudohermaphroditism
  female. *See* 46,XX disorders of sex development
  male. *See* 46,XY disorders of sex development
Pseudo-Hurler polydystrophy, 3353t
Pseudohyperkalemia, 358, 358t, 593
Pseudohypoaldosteronism type I, 359, 2368t
Pseudohypoaldosteronism type II (Gordon's syndrome)
  clinical features of, 359
  genetic considerations in, 359, 2368t
  treatment of, 359
Pseudohypokalemia, 354
Pseudohyponatremia, 350
Pseudohypoparathyroidism (PHP). *See also* Albright's hereditary osteodystrophy (AHO)
  classification of, 3288, 3288t
  clinical features of, 3288
  genetic considerations in, 3280f, 3288–3289, 3289f
  hypocalcemia and, 3288
  treatment of, 3289
Pseudo-Kaposi's sarcoma, 406
Pseudolymphoma (cutaneous lymphoid hyperplasia), 393, 405
Pseudomembrane(s)
  in diphtheria, 1223, 1223f
  in sulfur mustard exposure, S5
Pseudomembranous colitis, 1079, 1082–1083, 1083f, 2470f. *See also Clostridioides difficile*–associated infection (CDI)
Pseudomonas aeruginosa, **1307**
  antibiotic resistance in, 1032, 1173t, 1182, 1184, 1186t, 1312–1313
  avoidance of host defenses by, 1308

  epidemiology of, 1307
  epigenetic modulation by, 973
  inflammatory response to, 1308
  laboratory identification of, 1307, S11
  motility of, 1307–1308
  tissue injury by, 1308
  treatment of, 1175
  virulence factors of, 1307–1308, 1307t
Pseudomonas aeruginosa infections, **1307**
  arthritis, 1056, 1057
  bacteremia, 1308–1310, 1309t
  bronchiectasis, 2242, 2243
  in cancer patient, 582
  cellulitis, 1052
  chronic respiratory, 1308, 1310
  CNS, 1309t, 1311
  in cystic fibrosis, 1307, 1310, 2245, 2248
  diagnosis of, 382, 1307, S11
  ecthyma gangrenosum, 147, 1052, 1308, 1308f, A5. *See also* Ecthyma gangrenosum
  endovascular, 1309t, 1310
  epidemiology of, 1307
  in febrile neutropenia, 1309t, 1312
  folliculitis, 142t, 147, 395, 1051, 1312, A1
  gastrointestinal, 1312
  in HIV infection, 1312
  in infective endocarditis, 1038
  in lung abscess, 1035
  in lung transplantation, 2284
  mastoiditis, 255
  meningitis, 583, 1122, 1122t
  multidrug-resistant, 1309t, 1312–1313
  nail, 1312
  ocular, 985, 1309t, 1311
  osteomyelitis, 1062, 1064t, 1065t, 1067, 1309t, 1310–1311
  otitis externa, 254, 1309t, 1311
  pathogenesis of, 1307–1308, 1307t
  pneumonia, 1309t, 1310
  risk factors for, 1024
  sepsis/septic shock, 2316
  skin and soft tissue, 1312
  treatment of, 1027f, 1028, 1028t, 1052, 1173t, 1175, 1309t
  UTIs, 1087, 1293, 1309t, 1312
  ventilator-associated pneumonia, 1031
  in war veterans, S7
Pseudomonas oryzihabitans, 1269
Pseudoobstruction, intestinal. *See* Intestinal pseudoobstruction
Pseudopapilledema, 229, 229f
Pseudoporphyria, 401, 420, 435
Pseudopseudohypoparathyroidism (PPHP), 3288, 3288t, 3783
Pseudoterranova decipiens, 1818
Pseudothrombocytopenia, 921, 922f
Pseudotumor(s)
  in hemophilic arthropathy, 2967
  in IgG4-related disease, 2931, 2932f
  orbital, 233
Pseudotumor cerebri, 117, 229, 3031
Pseudotumor syndrome, 929
Pseudoxanthoma elasticum, 404, 927, 1860
PSI (Pneumonia Severity Index), 1026
Psilocybin, 3667, 3696
Psittacosis (*C. psittaci* infection), 1475–1476
Psoas abscess, 1075–1076
Psoralens, 400, 421, 434t, 436
Psoriasis, **386**
  clinical features of, 393t
    distribution of skin lesions, 378t, 379, 380f, 386t, 393t
    erythroderma, 393t
    morphology of skin lesions, 378t, 393t
    papules and plaques, 380f, 386t, A5
  diagnosis of, 393t
  drugs exacerbating, 420, 2940t
  etiology of, 386–387
  guttate, 386
  histologic features of, 386t
  in HIV infection, 1605
  hypothermia in, 3769
  in IBD, 2562t, 2563
  inflammasome mutations in, 2762t
  inverse, 386
  KIRs with, 2772t
  plaque-type, 386
  pustular
    vs. AGEP, 425
    clinical features of, 386, 395, A5
    drugs exacerbating, 420
  treatment of, **387,** 387t, 393t, 431, 436, 2612, 2784
Psoriatic arthritis, **2886**
  aortic aneurysm in, 2166
  classification of, 2888, 2889t
  clinical features of, 386, 392, 2887–2888
  diagnosis of, 2888
  differential diagnosis of, 2847, 2889
  disease activity assessment in, 2889–2890
  epidemiology of, 2887
  imaging in, 2888, 2889f
  KIRs with, 2772t
  pathogenesis of, 2881f
  phenotypes of, 2887
  treatment of, 387t, 2890, 2891t
PSORS1C3 gene, 1581t, 1583f, 1584
PSORSC1 gene mutations, 2861
PSP deficiency, 3368t
P-SPIKES, for communicating bad news, 77, 77t
PSTPIT1 gene mutations, 462, 2937
PSV (pressure-support ventilation), 2305t, 2306. *See also* Mechanical ventilation
PSVT. *See* Paroxysmal supraventricular tachycardia (PSVT)
Psychiatric/psychological disorders, **3657**
  abdominal pain in, 112t
  in alcohol use disorder, 3676–3677
  anxiety disorders. *See* Anxiety disorders
  biology of, **3651**
    genetic considerations in, 3651–3652, 3653f
    inflammatory mechanisms, 3656
    signal transduction, 3652–3654, 3654f
    systems neuroscience, 3654–3656, 3655f
  bipolar disorder. *See* Bipolar disorder
  in cancer survivors, 757
  classification of, 3657
  vs. dementia, 3485
  depression. *See* Depression/depressive disorders
  in displaced persons, 3740
  fatigue in, 166
  feeding and eating disorders, **3670**
  in fibromyalgia, 2963
  gait disorders in, 179
  global considerations in, 3657, 3851–3852
  at high altitudes, 3759
  IBS and, 2572
  insomnia in, 215
  in LGBTQIA+ people, 3178
  low back pain in, 126
  in military veterans, S8
  mood disorders. *See* Mood disorders
  movement disorders in, 3519
  palpitations in, 291
  personality disorders, **3671**
  prevalence of, 3657
  primary care physician's responsibility in, 13, 3657
  schizophrenia. *See* Schizophrenia
  seizures in, 3419
  sensory symptoms of, 176
  somatic symptom disorder, **3670**
  substance use and, 3697
  syncope in, 160
  systems neuroscience of, 3656t
  taste disorders in, 242
  in Wilson's disease, 3338–3339
  in women, 3167
Psychogenic amnesia, 3485
Psychosis(es). *See also* Schizophrenia
  alcohol-induced, 3676–3677
  D9THC-induced, 3688
  in Huntington's disease, 3514
  in PD, 3506
  postictal, 3426–3427
Psychosocial short stature, 2995
Psychostimulants. *See also specific substances*
  acute intoxication, 3695
  adulteration of, 3694–3695
  clinical effects of, 3656t, 3694
  combined with other drugs, 3694
  overdose of, 3694
  prescribed use of, 85, 3693
  withdrawal from, 3695
Psychostimulant use disorder, 3695
Psychotherapy
  for depression, 3666
  for panic disorder, 3658
  for phobic disorders, 3661
  for PTSD, 3663
  referral for, 3657
Psyllium, 81, 2575, 2577t
Psylocibin, 85
PT. *See* Prothrombin time (PT)
PTCH gene mutations, 524t, 606
PTEN gene mutations
  in brain tumors, 717, 717t
  in breast cancer, 633, 3800
  cancers associated with, 524t
  in cervical cancer, 713
  in Cowden's syndrome, 404, 3080t, 3088
  in endometrial cancer, 714
  genetic testing for, 3802
  in melanoma, 601–602
  in ovarian cancer, 710
Pterygium, 223
PTGER4 gene mutations, 2554t
PTH. *See* Parathyroid hormone (PTH)
PTH(1-84), 3287
PTH gene mutations, 736, 3287
Pthiriasis palpebrum, 3732
Pthirus pubis (pubic lice), 3731–3732
PTH-related peptide, 363
PTHrP. *See* Parathyroid hormone–related protein (PTHrP)
PTLD (posttransplant lymphoproliferative disorder), 1161
Ptosis, 233–234, 1235
PTPN2 gene mutations, 2553, 2554t
PTPN11 gene mutations, 879
PTPN22 gene mutations, 2822, 2823t, 2843
PTPS (6-pyruvoyl-tetrahydropterin synthase) deficiency, 3368t
PTSD (posttraumatic stress disorder), 3661–3663, 3662–3663t, S8
Puberty
  definition of, 3102
  delayed. *See* Delayed puberty
  female, 3130
  male, 3102, 3102f
  precocious. *See* Precocious puberty
Pubic lice *(Pthirus pubis),* 3731–3732
Puborectalis muscle, 304f
PUBS (percutaneous umbilical blood sampling), 3797
Pudendal nerve terminal motor latency, 2583
"Puff of smoke" lesion, 3447

Pulmonary abscess. *See* Lung abscess
Pulmonary angiography, 2161
Pulmonary arterial hypertension (PAH). *See also* Pulmonary hypertension
  in β thalassemia, 778t
  in congenital heart disease, 2191
  in connective tissue disease, 2191
  diagnosis of, A9
  familial, 3957, 3961t
  future research needs in, 2197
  genetic considerations in, 2185–2187, 2186f
  genotypes of, 35
  in HIV infection, 1597, 2192
  idiopathic, 2191, A9
  pathobiology of, 2185, 2186–2187f
  pathophysiology of, 2187–2188, 2188–2189f
  portal hypertension and, 2191–2192
  in sarcoidosis, 2926
  in sickle cell disease, 774t
  in systemic sclerosis, 2191, 2865f, 2866, 2869, 2874–2875
  systems biology applied to, 3957, 3961t
  treatment of, 35, 2187f, 2193–2194, 2196f, 2197t
Pulmonary artery catheter, 1985, 2297, 2315, 2326, 2330
Pulmonary artery wedge pressure, 1861, 2190
Pulmonary capillary wedge pressure, 2315, 2330
Pulmonary edema, **2330**
  acute, 1976
  cardiogenic, 2330
  chest radiography in, A12
  cyanosis in, 278
  diagnosis of, 2329t, 2330, 3912t
  hemodynamic patterns in, 2329t
  hemoptysis and, 275
  high-altitude. *See* High-altitude pulmonary edema (HAPE)
  in hyponatremia, 349
  in malaria, 1764, 1773
  noncardiogenic, 2330
  pathophysiology of, 2203
  physical examination in, 2200
  reexpansion, 2331
  treatment of, 2330–2331
  in upper airway obstruction, 2332
Pulmonary embolism (PE), **2154**
  adequacy of ventilation in, 2204
  after subarachnoid hemorrhage, 3461
  chest pain in, 103t, 104
  classification of, 2158
  diagnosis of
    blood tests, 2159–2160
    cardiac biomarkers, 2160
    chest CT, 2161, 2161f
    chest radiograph, 107, 2160
    clinical decision rules, 2159, 2159t
    clinical evaluation, 2159f
    clinical pearls, 2159
    CT, A12
    ECG, 107, 2160, 2160f, A7
    echocardiography, 2161
    imaging algorithms, 2159f, 2162f
    lung scan, 2161
    MRI, 2161
    pulmonary angiography, 2161
    Wells clinical prediction rule, 26t
  differential diagnosis of, 2159t
  dyspnea in, 270, 2159
  epidemiology of, 939–940, 1854–2156, 2155f
  hemoptysis in, 106, 275
  in HIV infection, 1605
  incidence of, 937
  long-term sequelae of, 2154, 2157f
  in mitral stenosis, 2048
  mortality from, 2154, 2155f
  nonthrombotic, 2159
  palpitations in, 291
  paraneoplastic, 741–742
  pathophysiology of, 2158–2159, 2158f
  pleural effusion in, 2270
  postmenopausal hormone therapy and, 3144t, 3145
  in pregnancy, 3902
  prevention of, 2164, 2164t
  readmission rate for, 2154, 2156f
  right ventricular dysfunction in, 1883, 2158
  risk stratification in, 2162
  septic, 1200, 1200f
  socioeconomic status and, 2154, 2155f
  treatment of
    algorithm for, 2162f
    anticoagulants, 2162–2163, 2163t, S13
    catheter-directed therapy, 2132, 2164
    duration of anticoagulation, 2163
    emotional support, 2164
    fibrinolysis, 2164
    inferior vena cava filter, 2163
    in massive PE, 2163–2164
    pulmonary embolectomy, 2164, A11, A11v
    pulmonary thromboendarterectomy, 2164
  ventilation-perfusion heterogeneity in, 2204
Pulmonary fibrosis
  bleomycin-related, 754
  bronchiectasis in, 2242, 2242t, 2243
  chemotherapy-associated, 563, 564
  in dyskeratosis congenita, 3819–3820
  idiopathic. *See* Idiopathic pulmonary fibrosis
  as late consequence of cancer/cancer therapy, 754
  pathophysiology of, 2203
  pulmonary function values in, 2204, 2206f
  radiation-induced, 554
  in rheumatoid arthritis, A12
  in sarcoidosis, 2925
  in systemic sclerosis, 2865–2866, 2865f
  telomere dysfunction and, 3819–3820, 3822f
  treatment of, 754–755
Pulmonary function tests
  in asthma, 2222–2223
  common abnormalities in, 2206f
  in COPD, 2251–2252, 2254
  in hypersensitivity pneumonitis, 2229
  in interstitial lung disease, 2262
  in occupational lung disease, 2235
  in pulmonary hypertension, 2190
  in respiratory disease evaluation, 2200–2201, 2240
Pulmonary heart disease. *See* Cor pulmonale
Pulmonary hypertension, **2185**
  arterial. *See* Pulmonary arterial hypertension
  chronic thromboembolic. *See* Chronic thromboembolic pulmonary hypertension (CTEPH)
  classification of, 2191, 2192t
  clinical features of, 2188–2189
    chest pain, 103t, 104
    dyspnea, 269t, 270, 2188
    exercise intolerance, 291
    tricuspid regurgitation, 2188
  diagnosis of, 2188–2190, 2191f, 2195f
    chest radiography, 2190f
    CT, 2189
    ECG, 2188, 2190f, A7
    echocardiography, 2188–2189, 2190f, A9
    heart sounds, 1865f, 1866, 2188
    pulmonary function tests, 2189
    right heart catheterization, 2190
    serology, 2190
    sleep studies, 2190
    vasoreactivity testing, 2190
  epigenetics in, 2186–2187
  future research needs in, 2197
  high-altitude, 3760
  in left heart disease, 2192, 2193f, 2195
  in lung disease, 2192, 2195–2196
  in mitral stenosis, 2048
  pathobiology of, 2185, 2186f
  pathophysiology of, 2158, 2187–2188
  in polycythemia vera, 821
  in pregnancy, 3899
  Raynaud's phenomenon in, 2178
  in sarcoidosis, 2193
  in schistosomiasis, 2193
  in sickle cell disease, 2193
  treatment of, 2195–2196
Pulmonary infarction, 111, 2159
Pulmonary infiltrates with eosinophilia
  ABPA. *See* Allergic bronchopulmonary aspergillosis (ABPA)
  acute eosinophilic pneumonia, 2231–2232, 2231t
  chronic eosinophilic pneumonia, 2232
  classification of, 2231, 2231t
  C. pseudotuberculosis, 1225
  drug- and toxin-induced, 2234
  in EGPA. *See* Eosinophilic granulomatosis with polyangiitis
  global considerations in, 2234
  hypereosinophilic syndromes. *See* Hypereosinophilic syndromes (HES)
  infectious causes of, 2234, 2234t
  pathophysiology of, 2231
Pulmonary lymphangioleiomyomatosis, A12
Pulmonary mass/nodule, 622, 623f, 623t, 1697, A12. *See also* Lung cancer
Pulmonary rehabilitation, 2258
Pulmonary renal syndrome, 2361f
Pulmonary risk assessment, preoperative, 3908–3909, 3909t
Pulmonary valve replacement, 2064, 2074, 2131
Pulmonary vein mapping, A9
Pulmonic regurgitation, 288, 289, 1866, **2063,** 2064f
Pulmonic stenosis, **2063**
  clinical features of, 2063
  diagnosis of, 2063
  etiology of, 2063, 2063t
  heart sounds/murmur in, 283f, 284f, 285, 286f, 1864, 1866, 2063
  pathophysiology of, 2063
  treatment of, 2063, 2130, 2131
Pulp, dental, 260
Pulpitis, 260, 1374
Pulp necrosis, 260
"Pulsatile tinnitus," 3457
Pulse, 1862, 1862f, 1863f, 2172
Pulsed-field ablation (PFA), 1917, 1952, 1953f
Pulsed-field gel electrophoresis, S11
Pulsed wave Doppler, 1876–1877, 1877f. *See also* Echocardiography
Pulseless disease. *See* Takayasu's arteritis
Pulseless electrical activity, 2334, 2335t, 2337f. *See also* Cardiac arrest
Pulse oximetry, 2200, 2296
Pulse pressure, 2293, 2294f, A10, A10v
Pulsus alternans, 1863, 1983
Pulsus paradoxus
  in dyspnea evaluation, 269t, 270
  echocardiography in, A9
  etiology of, 1863
  measurement of, 1863
  in obstructive lung disease, 2200
  in pericardial tamponade, 1867
Pulsus parvus et tardus, 1862, 2037
Punch biopsy, 381

“Punch drunk” syndrome. *See* Chronic traumatic encephalopathy (CTE)
Punding, 3502
Punta Toro virus, 1662t, 1668
Pupils
  afferent defect of, V3
  assessment of, 221
  in coma, 190, 190f
Pure alexia without agraphia, 201t, 203
Pure autonomic failure (PAF), 3548
Pure red cell aplasia, **813**
  classification of, 814t
  clinical associations of, 813–814
  congenital (Diamond-Blackfan anemia), 813
  definition of, 813
  differential diagnosis of, 813, 814f
  drug-induced, 489, 814t
  etiology of, 813–814
  parvovirus B19 infection and, 450, 814, 814f, 1524, 1524t
  pathophysiology of, 450
  thymoma and, 813
  treatment of, 814–815
Pure tone audiometry, 248–249
Pure white cell aplasia, 813
Pure word deafness, 201t, 203
Purified protein-based vaccines, 997t
Purified protein derivative (PPD)/tuberculin skin test (TST)
  in fever of unknown origin, 153
  limitations of, 975–976, 1393
  in NTM infections, 1417
  in TB diagnosis, 1385
  as TB preventive treatment indication, 1401–1402, 1401t
  in TB screening, 1393–1394
Purine metabolism disorders, 2798–2799, S9
Purine nucleoside phosphorylase deficiency, 2799
Purkinje fibers, 1844, 1912
Purpura, 137, **406**
  nonpalpable, 137, 407
  palpable, 137, 406, 408, 2905
  posttransfusion, 911
  senile, 472
  solar, 407
  vascular, 927
  wet, 924
Purpura fulminans
  clinical features of, 144t, 148, 407
  epidemiology of, 144t
  etiology of, 991t, 992
  genetic considerations in, 933, 941
  in meningococcemia, 1248, A1, A5
  treatment of, 934, 991t, 992
Pus, 455
Push enteroscopy, 320, 2468
Push packages, S4
Pustular lesions, 137, 377t, **395,** 425–426
Pustular psoriasis. *See* Psoriasis, pustular
Puumala virus, 1661t, 1676
PUVA photochemotherapy
  actions of, 436
  adverse effects of, 387
  for atopic dermatitis, 393t
  for morphea, 416
  for mycosis fungoides, 866
  for polymorphous light eruption prophylaxis, 433–434
  for psoriasis, 387, 393t, 436
  for vitiligo, 398t, 436
PVCs. *See* Premature ventricular contractions (PVCs)
PVE. *See* Prosthetic valve endocarditis (PVE)
pVHL (elongin), 3087
PVL (Panton-Valentine leukocidin) toxin, 1027, 1052, 1198
PVNS (pigmented villonodular synovitis), 2972, 2972f
P wave, 1869, 1910–1911
PWID. *See* Persons who inject drugs (PWID)
PYCR1 (pyrroline-5-carboxylate reductase 1) deficiency, 3368t
PYCR2 (pyrroline-5-carboxylate reductase 2) deficiency, 3368t
Pyelonephritis. *See also* Urinary tract infections (UTIs)
  actinomycotic, 1364
  approach to the patient, 1089–1090
  clinical features of, 1089
  complications of, 1094
  definition of, 1086
  emphysematous, 1089, 1090f, 1094
  M. hominis, 1467
  in pregnancy, 3903
  prognosis of, 1095
  renal biopsy in, A4
  risk factors for, 1087
  treatment of, 1094, 2321t
  xanthogranulomatous, 1089–1090, 1090f, 1094
Pyemotes spp., 3730, 3731f
Pyknodysostosis, 3315
Pyoderma. *See* Impetigo
Pyoderma gangrenosum
  clinical features of, 409, 1051, A5
  differential diagnosis of, 406
  diseases associated with, 409
  in IBD, 2561, 2562t
Pyoderma vegetans, 2562t, 2563
Pyogenic arthritis, pyoderma gangrenosum, and acne (PAPA) syndrome, 462, 2937
Pyogenic granuloma, 261, 421
Pyomyositis, 1053, 1200
Pyostomatitis vegetans, 2562t, 2563
PYP (technetium pyrophosphate) scan, 2023, 2024–2025f
Pyramidal (corticospinal) system, 169f
Pyrantel pamoate, 1240, 1745t, 1751, 1815, 1815t, 1818
Pyrazinamide
  actions of, 1423
  adverse effects of, 1397, 1400, 1423
  dosage of, 1423
  genetic variations in response to, E72t
  for MDR-TB, 1394t, 1395t, 1397–1399
  pharmacology of, 1423
  in pregnancy, 1423
  resistance to, 1397, 1423
  for TB, 1394–1396, 1394t, 1395t, 1419t, 1420t, 1423
  for tuberculous meningitis, 1127
Pyrethrins, 3732
Pyridostigmine
  adverse effects of, 3629, S8
  for gastroparesis, 299, 299t
  for Lambert-Eaton myasthenic syndrome, 3628
  for myasthenia gravis, 3629–3630
  for orthostatic hypotension, 160, 3547
Pyridoxal 5’-phosphate (PLP), 3316
Pyridoxine (vitamin B_6), **2608**
  adverse effects of, 3609t
  for alcohol-induced acidosis, 371, S1
  deficiency of, 400, 817, 2605t, 2608–2609, 3612
  dietary sources of, 2608
  for ethionamide toxicity, 1426
  functions of, 2606f, 2608
  for homocystinuria, 3371
  for isoniazid toxicity, 1396, 1403, 1419t, 1422, 3611, 3708, 3715t
  for nausea and vomiting in pregnancy, 300, 2609
  recommended intake of, 2600t
  for sideroblastic anemia, 3349
  structure of, 2606f
  for TB prophylaxis, 1593t
  toxicity of, 2609, 3609t, 3611
  for tuberculous meningitis, 1127
Pyridoxine-dependent seizures, 3369t
Pyrilamine, 3711t
Pyrimethamine
  actions of, 1751–1752
  adverse effects of, 788, 1743t, 1752, 1771t, 1775
  for *C. belli* infections, 1806
  for CNS toxoplasmosis, 1138
  drug interactions of, 1743t
  for malaria, 1770t, 1771t
  for PCP prophylaxis, 1593t, 1734t
  pharmacology of, 1752, 1771t
  in pregnancy and lactation, 1743t
  resistance to, 1775
  for toxoplasmosis, 1157, 1802–1803
  for toxoplasmosis prophylaxis, 1593t, 1803
Pyrimidine-5’-nucleotidase (P5N) deficiency, 796t, 800
PYRIN domains, 462, 2934
Pyrin inflammasome, 972, 2934
Pyrin protein, 2934
Pyrogens, 134
Pyroglutamic acid, 371
Pyronaridine, 1745t, 1752, 1771t
Pyroptosis, 972, 1576
Pyrosis. *See* Heartburn
Pyrroline-5-carboxylate reductase 1 (PYCR1) deficiency, 3368t
Pyrroline-5-carboxylate reductase 2 (PYCR2) deficiency, 3368t
Δ^1-Pyrroline-5-carboxylate synthase deficiency, 3368t
Pyruvate dehydrogenase complex (PDC), 2711, 2790t
Pyruvate kinase deficiency, 323, 792f, 796–797, 796t, 797f
6-Pyruvoyl-tetrahydropterin synthase (PTPS) deficiency, 3368t
Pyuria, 343, 989, 1086

Q

Q fever, **1462**
  chronic, 1463
  clinical features of, 150, 1455t, 1462, 2013, S7
  diagnosis of, 150, 1463, 1463t
  epidemiology of, 1455t, 1462
  etiology of, 1455t, 1462
  fatigue syndrome following, 1463
  follow-up, 1464
  global considerations in, 1462
  incubation period for, S7
  prevention of, 1464, S7
  treatment of, 1048, 1463–1464, S7
  in war veterans, S7
Q fever vaccine, 1464
QRS complex, 1868, 1869–1870, 1870f, 1912
QRS interval, prolonged, 1981
QRS voltage, low, 1874, 1875f
QS-21, 997t, 998
qSOFA (quick Sequential Organ Failure Assessment Score), 2313
QT interval
  prolonged, 356, 365, 487, 1874, 1875f
  shortened, 1874, 1875f
Quadrantanopia, 223f
Quadriparesis, 170–171, 170f, 171t
Quadriplegic fever, 3569
Quality in health care. *See* Health care quality
“Quality of life” issues, 6
Quality theory, 43–44, 44f
QuantiFERON-TB Gold test, 1394
Quantitative sensory testing (QST), 331
Quaranfil virus, 1662t

- *Quaranjavirus,* 1656, 1662t
- Quartan malarial nephropathy, 1766
- Queensland tick typhus, 1458
- Queer, 3179t
- Questioning, in LGBTQIA+ terminology, 3179t
- Quetiapine
  - adverse effects of, 3673t, 3674
  - for bipolar disorder, 3668
  - for delirium, 86t
  - for dementia, 199
  - overdosage/poisoning with, 3711t
  - for psychosis in Huntington's disease, 3514
  - for psychosis in PD, 3506
  - for schizophrenia, 3673, 3673t
- Quick Sequential Organ Failure Assessment Score (qSOFA), 2313
- Quinacrine
  - actions of, 1752
  - adverse effects of, 320, 420, 1745t
  - drug interactions of, 1741t, 1745t
  - pharmacology of, 1752
  - in pregnancy and lactation, 1745t
  - for SLE, 2833–2834
- Quincke's pulse, 288, 2044
- Quinidine
  - actions of, 1916t
  - adverse effects of
    - arthralgias, 2940t
    - cardiac, 1752
    - common/frequent, 1745t, 1771t
    - cutaneous, 392
    - lupus syndrome, 2940t
    - occasional, 1745t
    - rare, 1771t
    - renal, 1752
    - SA node dysfunction, 1920t
    - thrombocytopenia, 489, 923t
  - for arrhythmias, 1958
  - for Brugada's syndrome, 1970
  - drug interactions of, 485t, 489t, 1745t
    - azoles, 1742t
    - SSRIs, 3660t
  - for malaria, 991t, 1770t, 1771t
  - metabolism of, 485t
  - pharmacology of, 1752, 1771t
  - in pregnancy and lactation, 1745t
- Quinine
  - actions of, 1752
  - adverse effects of
    - common/frequent, 1745t, 1771t
    - hearing loss, 246
    - hypoglycemia, 3234
    - occasional, 1745t
    - oral pigmentation, 264t
    - rare, 1771t
    - renal, 1752
    - thrombocytopenia, 489, 922, 923t
    - TTP, 2445
  - for *Babesia* infection, 991t, 1779–1780, 1780t
  - drug interactions of, 1745t, 1752
  - for malaria, 991t, 1771t
  - overdosage/poisoning with, 3715t
  - pharmacology of, 1752
  - in pregnancy and lactation, 1745t
- Quinolones. *See also* Fluoroquinolones
  - actions of, 1166–1167, 1182t, 1183f
  - adverse effects of, 400, 2377, 2940t, 2974, 3234
  - resistance to, 1182t, 1183f, 1185, 1204, 1255
- Quinsy. *See* Peritonsillar abscess
- Quintan (trench) fever, 1353t, 1354
- Quinupristin-dalfopristin
  - actions of, 1166, 1179, 1182t
  - adverse effects of, 1171t, 1179
  - drug interactions of, 1172t, 1179
  - indications for, 1174t, 1179
  - in pregnancy and lactation, 1169t
  - resistance to, 1174t, 1182t, 1184
- Quizartinib, 534t, 569t, 753t, 833, 833t
- Quorum sensing, 1327
- Q wave, 1868, 1873, 1873f, 2115, A7

R

- R5 viruses, 1577
- R5X4 viruses, 1577
- RA. *See* Rheumatoid arthritis (RA)
- RAA system. *See* Renin-angiotensin-aldosterone (RAA) system
- *RAB27A* gene mutations, 464, 465
- *RAB32* gene mutations, 3497t, 3499
- Rabeprazole
  - adverse effects of, 2525
  - for GERD, 301
  - for peptic ulcer disease, 2524t, 2525
  - for ZES, 2536
- Rabies, **1650**
  - clinical features of, 1115, 1652, 1653f, 1653t
  - diagnosis of, 1115, 1653, 1653f, S7
  - differential diagnosis of, 1115, 1653–1654
  - epidemiology of, 1650–1651, 1650–1651f
  - global considerations in, 1650, 1650f
  - incubation period for, S7
  - in military forces, S7
  - pathogenesis of, 1651–1652, 1652f
  - postexposure prophylaxis for, 1145t, 1146, 1654–1655, 1654f
  - prognosis of, 1654
  - treatment of, 1654
- Rabies immune globulin, 1146, 1654–1655
- Rabies vaccine
  - after bite-wounds, 1146
  - Guillain-Barré syndrome and, 3617
  - for international travelers, 1014t, 1017
  - postexposure, 1655
  - preexposure, 1655
  - for travelers, 1655
- Rabies virus, 1482, 1650
- Raccoon pinworm. *See Baylisascaris procyonis* infections
- Raccoon sign, 897f, 3572
- "Rachitic rosary," 3271
- Racial/ethnic disparities in health care, **61**
  - health system interventions for, 67–68
  - hospice care use and, 76
  - in hypertension treatment, 2148
  - implications for clinical practice, 69
  - life expectancy and, 61–62, 61–62f
  - mistrust and, 67
  - nature and extent of, 61–62, 63–64f
  - patient interventions for, 69
  - placebo/nocebo effects and, 3934
  - provider interventions for, 68–69
  - recommendations to address, 67–68
  - root causes of, 62–64
  - social determinants of health and. *See* Social determinants of health
  - sudden cardiac death rate, 2332–2333
- Racism, 62
- Rad, 554, S6
- Radial artery, for PCI access, 2127
- Radial neuropathy, 3613–3614
- Radiation, S6. *See also* Ionizing radiation
- Radiation dose, 554
- Radiation enteritis, 2550t
- Radiation esophagitis, 2513
- Radiation sensitizers, 554
- Radiation syndrome, acute, S6
- Radiation terrorism, **S6**
  - decontamination methods, S6
  - dispersal events, S6
  - hospital management of casualties, S6
  - medical assay of exposure level, S6
  - medical management of casualties, S6
- Radiation therapy, **554**
  - for acromegaly, 3010
  - adverse effects of, 554
    - bowel complications, 309, 700
    - breast cancer, 633
    - cardiovascular, 753, 2083
    - cataract, 230
    - cobalamin malabsorption, 787
    - dry eye, 224
    - dysphagia, 293
    - endocrine dysfunction, 755–756
    - head and neck, 611
    - hepatic, 757
    - in high-altitude travel, 3760
    - in HIV infection, 1612
    - hyperbaric oxygen therapy for, 3762–3763
    - hypopituitarism, 2993, 3001
    - infections, 578t, 579
    - long-term, 726, 753t, 871
    - lymphedema, 2183
    - neurologic, 725–726, 756, A16
    - pericarditis, 587, 753, 2083
    - pituitary, 3067t
    - plexopathy, 3615–3616
    - pneumonitis, 596
    - pulmonary, 596, 754, 2234
    - in rectal cancer, 658
    - reproductive dysfunction, 756
    - second malignancy, 554, 755
    - sexual function, 700
    - testicular dysfunction, 3115
    - thyroid cancer, 3045
    - vascular, 554
    - xerostomia, 266
  - for anal cancer, 659
  - for astrocytoma, 718
  - biologic effects of, 554
  - for bone pain, 82
  - brachytherapy. *See* Brachytherapy
  - for brain metastases, 724
  - for breast cancer, 636–637
  - cranial, prophylactic in lung cancer, 631
  - curative, 554
  - dosage of, 554
  - for epidural metastases, 725
  - for esophageal cancer, 644
  - external beam. *See* External beam radiation therapy
  - for follicular lymphoma, 864
  - for gastric cancer, 649
  - for glioblastoma, 719
  - for head and neck cancer, 613
  - for Hodgkin's lymphoma, 870–871
  - for leptomeningeal metastases, 724
  - for lung cancer, 622–623, 626, 630–631
  - for melanoma, 604
  - modalities of, 554
  - palliative, 639
  - for PCNSL, 721
  - physical properties of, 554
  - for pituitary tumors, 3001–3002
  - for plasmocytoma, 890
  - for prostate cancer, 699–700
  - quantitation of, 554
  - for RCC, 689
  - for rectal cancer, 658
  - second malignancy and, 871, 3002
  - for skin cancer, 607
  - for soft tissue sarcomas, 728
  - systemic, 554
  - for testicular cancer, 706
- Radiative forcing, 3745f
- Radicular low back pain, 120, 124t, 128
- Radicular neck pain, 129–130, 129t

Radiculopathy, 3393, 3614–3616, 3614t
Radioallergosorbent tests, 2223
Radiocontrast media. *See* Contrast agents
Radioembolization, 666
Radiofrequency ablation
  for cardiac arrhythmias, 1916–1917, 1916f
  characteristics of, 555
  for chronic venous disease, 2182
  dysplastic Barrett's esophagus, V5
  for HCC, 664t, 665
  for metastatic lesions, 555, 684
Radiofrequency thermal rhizotomy, 3553
Radioiodine therapy
  for Graves' disease, 3036
  for hyperfunctioning solitary thyroid nodule, 3043–3044
  for hyperthyroidism, 3901
  for nontoxic multinodular goiter, 3041–3042
  safety precautions for, 3036
  for thyroid cancer, 3047
  for toxic multinodular goiter, 3043
Radionuclide contamination, S6
Radionuclide scans/scintigraphy
  for colonic transit measurement, 313
  in gallbladder disease, 2730, 2730t
  in gastrointestinal bleeding, 320
  in musculoskeletal disease evaluation, 2947, 2947t
  in prostate cancer staging, 698–699
  in STEMI, 2117
  in thyroid evaluation, 3026–3027
Radium-226 exposure, S6
Radium-233, 703, 731
Radon exposure, 615, 2240
Raf kinase inhibitors, 526f, 557f
RAF/MEK pathway antagonists, 569t, 572
RAG1/2 gene mutations, 2786f
Ragged red fibers, 3645
Rales, 1984, 2200
Raloxifene
  for breast cancer prevention, 513, 634
  for gynecomastia, 3117
  for osteoporosis prevention/management, 3145, 3303
Ralstonia spp., 1269
Raltegravir, 1617f, 1618t, 1623f
Ramelteon, 186
Ramipril, 2144t
Ramosetron, 83
Ramsay Hunt syndrome (herpes zoster oticus), 163, 253, 265, 1506, 3554
Ramucirumab
  action and targets of, 534t, 548, 559f, 2792t
  adverse effects of, 592, 666
  for gastric cancer, 649
  for HCC, 662, 664t, 665f, 666
  indications for, 2792t
  for lung cancer, 629
Randall's plaque, 2449
Range of motion, 2942, 2942t
Ranibizumab, 231, 2792t
Ranitidine
  for gastrinoma, 3081
  for peptic ulcer disease, 2524t, 2525
  for urticaria, 2808
RANK (receptor activator of NFκB)
  in bone metastasis, 541
  in bone remodeling, 3294, 3294f
  in osteoclast and osteoblast differentiation, 3312f
  in Paget's disease of bone, 3312
  in rheumatoid arthritis, 2845f
RANKL (RANK ligand)
  antibody. *See* Denosumab
  in bone remodeling, 3260, 3294, 3294f
  receptors for, 541
  in rheumatoid arthritis, 2845f, 2846
Ranolazine, 1916t, 2102, 2102t
RANTES, 2767t
Raoultella, 1294
Rapamycin. *See* Sirolimus
Rapeseed oil, 2940t
Rapid acetylators, 493
Rapid cycling, in bipolar disorder, 3668
Rapid detection tests
  for Group A streptococcal infections, 258
  for parasites, S12
Rapidly progressive dementia, 194
Rapid plasma reagin (RPR) test, 1433
RAS (reticular activating system), 187
Rasagiline, 178, 3503, 3504, 3504t
Rasburicase
  adverse effects of, 594, 799t, 800
  genetic variations in response to, 493, E72t
  for tumor lysis syndrome treatment/prophylaxis, 594
  for uric acid nephropathy, 829
RAS gene mutations
  cancers associated with, 536, 3004
  as drug target, 536, 572
  in juvenile myelomonocytic leukemia, 879
  in melanoma, 601, 605
  in thyroid cancer, 3046
Rash. *See also* Skin disorders/lesions; *specific diseases*
  after transfusion, 908t
  approach to the patient, 137
  with arthritis, 1666t, 1669–1670
  centrally distributed maculopapular eruptions, 137, 138–140t, 146
  classification of, 137
  confluent desquamative erythemas, 141–142t, 146–147
  drug-induced. *See* Cutaneous drug reactions
  with fever, **137,** 138–145t, 409, 1248, **A1**
  in glucagonoma, 678, 679f
  in infections, 990–993. *See also specific infections*
  morbilliform, 377t, 401, 402t
  nodular eruptions, 144t, 148
  in parasitic infections, 1738t
  peripheral eruptions, 140–141t, 146
  petechial, 1248t, 1456, 1456f, A1
  purpuric eruptions, 144–145t, 148, 1248t
  pustular eruptions, 142–143t, 147
  scarlatiniform, 402, 402t
  ulcers and eschars, 145t, 148–149
  urticaria-like eruptions, 143–144t, 147–148, 147f
  vesiculobullous eruptions, 142–143t, 147
Rasmussen's aneurysm, 274, 1387
Rasmussen's encephalitis, 2790t
RAS pathway, 526t
Raspberry tongue, 265t
Ras protooncogene, 526f
RASSF1A gene mutations, 617
Rastelli procedure, 2075, 2075t
Rat(s). *See* Rodent(s); Rodent-borne viruses
Rat-bite fever (Haverhill fever), 141t, 1143
Rat-bite fever (sodoku), 139t, 1143
Rat bite-wound, 1143
Rat flea, 3736
Rathke's cysts, 3002
Rational world bias, 3916t, 3918–3919
Rationing, 44
Rat lungworm. *See Angiostrongylus* spp. infections
Rattlesnake bite, 3717, 3718f, 3720t, 3722. *See also* Snakebites
Ravn virus, 1680, 1680–1682f
Ravulizumab, 595, 805f, 806
  for familial hemolytic-uremic syndrome, 2445
  for myasthenia gravis, 3631, 3632t
  for neuromyelitis optica, 3592, 3592t
R_{aw} (airway resistance), 270, 2204, 2206f
Raxibacumab, S4
Raynaud's phenomenon, **2176**
  chilblain in, 3771
  clinical features of, 2176–2177, 2177f
  cyanosis in, 279
  in dermatomyositis, 414
  drug-induced, 595, 2178, 2940t
  localized hypoxia in, 277
  pathophysiology of, 2177, 2790t, 2859
  primary (Raynaud's disease), 270, 2177–2178
  secondary causes of, 2177t, 2178
  in Sjögren's disease, 2877, 2877t
  in SLE, 2829
  in systemic sclerosis, 416, 2178, 2859, 2861t, 2867, 2867f
  treatment of, 2178, 2874
RB1 gene mutations
  in bladder cancer, 691
  cancers associated with, 524t, 727
  in CML, 836
  in HCC, 661t
  in lung cancer, 617
  in ovarian cancer, 710
RBD. *See* REM sleep behavior disorder (RBD)
Rb gene mutations, 3276–3277, 3792
RBM20 gene mutations, 2004t
Rb (retinoblastoma) protein, 530, 573, 599, 615
RCC. *See* Renal cell carcinoma (RCC)
R-CHOP regimen, 650, 863, 864
RCRI (Revised Cardiac Risk Index), 3905, 3907f, 3907t
RDW (red blood cell distribution width), 437, 445
Reactive arthritis, **2890**
  after *Campylobacter* infection, 1305, 2890
  after *C. trachomatis* infection, 1470–1471, 1473t
  after infectious diarrhea, 1079t
  after nontyphoidal *Salmonella* infections, 1320, 2890
  after *Shigella* infections, 1323–1324, 2890
  after *Yersinia* infections, 1350, 2890
  clinical features of, 1060, 2890
    aortic aneurysm, 2166
    diarrhea, 305
    oral, 261
  diagnosis of, 2890
  vs. gonococcal arthritis, 1257, 1258f
  in HIV infection, 1603
  treatment of, 1350, 2890–2891
Reactive hyperemia, 2177
Reactive infectious mucocutaneous eruption (RIME), 143t, 147, 424, A1
Reactive lymphocytes, 443
Reactive oxygen species (ROS)
  in aging, 3815
  in asthma, 2219
  damage to mtDNA by, 3815, 3815f
  in endothelium, 1845
  generation of, 430
  in hyperbaric oxygen therapy, 3761
  in sepsis/septic shock, 2317, 2318f, 2319
  skin damage caused by, 430, 431
Reactive skin decontamination lotion (RSDL), S5
"Readback," 43
Reading, assessment of, 201
Reassortment, genetic, 3856
Rebound tenderness, 112
Rebuck skin window test, 463
Receiver operative characteristic curve (ROC), 24, 24f
Receptors, in bacterial adherence, 966–967
Receptor tyrosine kinase, 526f, 532, 2981t
Receptor tyrosine kinase receptors, 540f
RECIST (Response Evaluation Criteria in Solid Tumors), 504, 555
Recluse spider bite, 3734

Recombinant tissue plasminogen activator (rtPA). *See* Alteplase (rtPA)
Recombination, genetic, 3780, 3782f
Recommended dietary allowance, 2599
Recoverin, 2791t
Rectal bleeding
  in colorectal cancer. *See* Colorectal cancer
  in hemorrhoidal disease. *See* Hemorrhoidal disease
  in rectal prolapse, 2582
  in ulcerative colitis, 2556–2557, 2557t
Rectal cancer, 658. *See also* Colorectal cancer
Rectal neuroendocrine tumors, 681. *See also* Neuroendocrine tumors (NETs)
Rectal prolapse, 1325, **2582,** 2582f
Rectal snip, S12
Rectal temperatures, 134
Rectoanal (anorectal) angle, 304, 304f
Rectoanal inhibitory reflex, 313
Rectocele, 314
Rectopexy, 2583
Rectus sheath, hematoma of, 111
Red blood cell(s)
  agglutination of, 439f, 443
  in anemia, 444. *See also* Anemia
  antigens and antibodies, 901, 903–904, 903–904t
  complement activation and, 805f
  excess of. *See* Polycythemia
  folate in, 789
  fragmented, 439f, 446f
  glycolysis in, 791, 792f
  hematopoietic differentiation of, 760f
  immune destruction of, 801, 803f
  life cycle of, 763, 792
  mechanical destruction of, 801, 801t
  membrane-cytoskeleton complex of, 792–793, 793f
  nucleated, 437, 440f, 447t
  osmotic fragility of, 793
  parasites in, 437
  in peripheral blood smear, 437, 441–442f
  production of, 443–444, 443f
  rare phenotypes, 904–903
  stippled, 437, 441f
  transfusions. *See* Transfusion(s)
  urinary casts, 341, A4
Red blood cell concentrate (RBCC), 902t
Red blood cell distribution width (RDW), 437, 445
Red blood cell indices, 445t
Red blood cell scintigraphy, 320
Red blood cell survival study, 792
Red cell anion exchanger (AE1), 358
"Red flags"
  in diagnosis, 57–58, 59t
  in GERD, 301, 301t, 2490
  in low back pain, 120–122, 126t
  in neck pain, 130t
  in upper respiratory tract infections, 130t
Red hair color (RHC) phenotype, 599–600
Red hepatization, 1023
Red light, for basal cell carcinoma, 430
Red man syndrome, 422, 1171t, 1176
Red pulp, of spleen, 478, 479f
"Red tide," 3727
5α-Reductase, 3138
5α-Reductase inhibitors, 513, 696, 704
5α-Reductase type 2 deficiency, 3099
Reductionism, 29, 31f, 3955
Reduviid bugs, 1737, 1789, 1792, 2013, 3737
Reed-Sternberg cells, 868, 869f, A6
Re-emerging infectious diseases, 3853, 3853t. *See also* Emerging infectious diseases
Reentry, 1913–1914, 1913t
Refeeding edema, 282
Refeeding syndrome, 2626
Reference values, clinical laboratory tests, S10
Referred pain, 94, 95f, 110t, 112
Reflex myoclonus, 3515
Reflex sympathetic dystrophy. *See* Complex regional pain syndrome (CRPS)
Reflex syncope. *See* Syncope, neurally mediated
Reflex testing, 3381. *See also specific reflexes*
Reflux nephropathy, 2441–2442, 2441f
Refractive state, 220–221
Refsum's disease, 232, 3602
Refusal of care, 72
Registries, 27
REGN-EB3, S4
Regorafenib
  actions and targets of, 533t, 548, 570t
  adverse effects of, 508, 570t, 666
  for colorectal cancer, 535, 573, 658
  for GIST, 729
  for HCC, 535, 564, 664t, 665f, 666
  interactions and issues, 570t
Regret aversion, 3916, 3916t, 3918
Regulatory intramembrane proteolysis (RIP), 3326
Regurgitation
  in achalasia, 2507
  in esophageal disease, 2504
  in GERD, 300, 2510
  nasal, 293, 295
Rehabilitation
  after leprosy treatment, 1413
  in alcohol use disorder, 3680–3681
  in ischemic stroke management, 3443
  neurotherapeutic. *See* Neurotherapeutic technologies
  in opioid-related disorders, 3691–3692
  pulmonary, 2258
  in spinal cord disorders, 3569–3570, 3569t
Reifenstein's syndrome (partial androgen insensitivity syndrome), 3099
Rejection
  in cardiac transplantation, 2030
  cfDNA in detection of, 3983f, 3984–3985, 3985f
  in kidney transplantation, 2407, 2409–2410
  in liver transplantation, 2723–2724, A13
  in lung transplantation, 2286
Relapsing fever, **1444**
  clinical features of, 136, 139t, 1446, S7
  diagnosis of, 1446–1447, 1447f, S7
  differential diagnosis of, 1447
  epidemiology of, 1445, 1445t
  etiology of, 1445, 1445f, 1445t
  global considerations in, 1445, 1445t
  hard tick. *See* Hard tick relapsing fever (HTRF)
  incubation period for, S7
  louse-borne. *See* Louse-borne relapsing fever
  pathogenesis of, 1446
  prevention of, 1448
  prognosis of, 1448
  treatment of, 1447–1448, 1447f, S7
Relapsing polychondritis, **2917**
  clinical features of, 253, 1860, 2166, 2918–2919f, 2918–2920
  diagnosis of, 2920–2921
  differential diagnosis of, 2920
  epidemiology of, 2917–2918
  pathology of, 2918
  pathophysiology of, 2918
  treatment of, 2921
Relative afferent pupil defect, 221
Relative bradycardia (Faget's sign), 136, 959, 960t
Relatlimab, 534t, 559f, 560, 604, 605, 2787t
Relaxation techniques, 3925t, 3927
Relaxin, 3128
Relebactam, 1175, 1183
Relugolix, 567, 701, 702f, 3137
Remdesivir, 996, 1116, 1156, 1488t, 1489, 1542
Remibrutinib, 2809
Reminder systems, 40
REM sleep, 210–211, 210f
REM sleep behavior disorder (RBD)
  clinical features of, 218, 220v, 241
  in DLB, 3491–3493, 3492t
  in PD, 3495, 3506–3507
  prevalence of, 218
  treatment of, 218
Renal abscess, 1075
Renal amyloidosis, 2424–2425, A4
Renal artery stenosis, **2150**
  azotemia in, 340
  diagnosis of, 2150, 2151t
  hyperaldosteronism and, 356
  impaired oxygen delivery in, 444
  incidence of, 2150
  pathophysiology of, 2150
  treatment of, 2150–2151, 2152t
Renal artery thrombosis, 341
Renal biopsy, **A4**
  in acute interstitial nephritis, A4
  in acute nephritic syndromes, A4
  in AKI, 2381–2382, A4
  in Alport's syndrome, 2426, A4
  in antiglomerular basement membrane disease, 2418, A4
  assessment of, 2416
  in CKD, 2395
  in diabetic nephropathy, 2423, A4
  in Fabry's disease, 2425, A4
  in focal segmental glomerulosclerosis, A4
  in glomerular diseases, 2416, A4
  in granulomatosis with polyangiitis, 2420, A4, A14
  in hemolytic-uremic syndrome, 2428, A4
  in hypertensive nephrosclerosis, 2427, A4
  in IgA nephropathy, 2419, A4
  indications for, 2363
  in lupus nephritis, 2418, A4
  in membranoproliferative glomerulonephritis, 2420, A4
  in membranous glomerulonephritis, 2423, A4
  in minimal change disease, 2421, A4
  in multiple myeloma, 2440f, A4
  in oxalosis, A4
  in poststreptococcal glomerulonephritis, 2417, A4
  in pyelonephritis, A4
  stains for, 2416
  in vasculitic syndromes, A14
Renal cell carcinoma (RCC), **687**
  clinical features of, 688
  deaths from, 501t
  diagnosis of, 688, 2362–2363
  epidemiology of, 687
  genetic considerations in, 687–688, 687t
  global considerations in, 690
  incidence of, 501t
  intrahepatic cholestasis in, 325
  metastatic, 723t
  paraneoplastic syndromes in, 688, 741
  pathology of, 687–688, 688t
  polycystic kidney disease and, 2432
  prognosis of, 688, 689f
  staging of, 688, 689f
  treatment of, 573, 689–690, 690t
  tuberous sclerosis and, 687t, 2434
  von Hippel-Lindau syndrome and, 687, 687t, 3087
Renal colic, 2448, 2450, 2451f, 2455, 2456. *See also* Nephrolithiasis
Renal coloboma syndrome, 2436
Renal cysts. *See* Polycystic kidney disease
Renal cysts and diabetes syndrome, 2430t

- Renal disease/failure
  - acute. *See* Acute kidney injury (AKI)
  - chronic. *See* Chronic kidney disease (CKD)
  - glomerular. *See* Glomerular diseases
  - inherited tubule growth and development disorders, **2430**
    - autosomal dominant, 2435
    - congenital abnormalities of kidney and urinary tract, 2436
    - with cystic phenotype, 2430t
    - diagnostic considerations in, 2436
    - global considerations in, 2436
    - karyomegalic tubulointerstitial nephritis, 2436, 2442
    - medullary sponge kidney, 2436
    - in mitochondrial disease, 2436
    - nephronophthisis, 2430t, 2435
    - polycystic kidney disease. *See* Polycystic kidney disease
    - tuberous sclerosis. *See* Tuberous sclerosis
    - von Hippel-Lindau disease. *See* Von Hippel-Lindau disease/syndrome
  - nephritis/nephritic syndromes, **2354**
    - AKI. *See* Acute kidney injury (AKI)
    - glomerulonephritis. *See* Glomerulonephritis
    - lupus nephritis. *See* Lupus nephritis
    - pyelonephritis. *See* Pyelonephritis
    - tubulointerstitial. *See* Tubulointerstitial diseases
  - nephrolithiasis. *See* Nephrolithiasis
  - tubulointerstitial diseases. *See* Tubulointerstitial diseases
  - vascular. *See* Renovascular disease
- Renal medullary carcinoma (*SMARCB1*-deficient), 687–688, 688t
- Renal natriuretic peptide, 2370
- Renal nerve denervation, 2132, 2146
- Renal osteodystrophy, 339, 3290
- Renal pelvis, carcinoma of. *See* Kidney cancer; Renal cell carcinoma (RCC)
- Renal transplantation. *See* Kidney transplantation
- Renal tubular acidosis (RTA)
  - distal (type I), 356, 372, 373, 2368t, 2455, S1
  - generalized distal (type IV), 372, 3224
  - proximal (type II), 372, 373, 2368t
  - treatment of, 2879
- Renal tubular defects, 338t
- Renal vein thrombosis, **2447**
- Renin, 737t
- Renin-angiotensin-aldosterone (RAA) system
  - assessment of, 3052
  - in blood pressure control, 2133
  - in edema, 279–280
  - in glomerular filtration rate regulation, 2364, 2365f
  - in heart failure pathogenesis, 1980, 1981f
  - in mineralocorticoid synthesis, 3052, 3053f
  - in potassium regulation, 353–354, 359
  - in sodium regulation, 2372
- Renovascular disease, **2150**
  - atheroembolic, 2152–2153, 2379t
  - atherosclerotic nephropathy, 2427
  - diagnosis of, 2138t, 2150, 2151t
  - hypertension and, 2153–2154
  - hypertension in, 2138t, 2139
  - hypertensive nephrosclerosis, 2154, 2427, A4
  - renal artery stenosis. *See* Renal artery stenosis
  - thromboembolic, 2153, 2153f
  - thrombotic microangiopathic. *See* Thrombotic microangiopathy
  - treatment of, 2138t
- Reoviruses, 1478f, 1479, 1480t, 1638, 1664–1665t
- Repaglinide, 3211t, 3213
- Repetition, assessment of, 201
- Replicating viral vector vaccines, 997t
- R-EPOCH regimen, 863
- Repotrectinib, 533t, 569t, 570, 628
- Representativeness heuristic, 21
- Reproduction number (Ro), 987
- Reproductive system disorders. *See* Female reproductive system disorders; Male reproductive system disorders
- Reptilase time, 475
- Reptile bites. *See* Snakebites
- Research
  - ethical issues in, 73–74
  - physician engaged in, 8
  - skews in, global effects of, 3862
- Resident physicians, working hours of, 7, 52, 219
- Residual volume (RV), 2202, 2202f, 2203f, 2206f
- Resistance
  - antibiotic, 980–982, 1151–1153, 1152t, 3855
  - to cancer treatments, 574
  - intrinsic, 1481
- Resistance strength training, 181
- Resistance to thyroid hormone (RTH), 3024t, 3025, 3786
- Reslizumab, 2225–2226, 2793t
- Resmetirom, 2706
- Respiration
  - in hypothermia, 3769t
  - in shock, 2291t, 2292t
  - signaling pathways in, 267–268, 268f
- Respiratory acidosis, **375**
  - acid-base nomogram, 367f
  - clinical features of, 375
  - compensatory responses in, 367t, 375
  - diagnosis of, 375
  - drug-induced, 375
  - etiology of, 375, 375t
  - hyperkalemia in, 359
  - in mixed acid-base disorders, 367, 368t
  - treatment of, 376
- Respiratory alkalosis, **376**
  - acid-base nomogram, 367f
  - compensatory responses in, 367t
  - diagnosis of, 376
  - etiology of, 375t, 376
  - in hypoxia, 277
  - in mixed acid-base disorders, 367, 368t
  - treatment of, 376
- Respiratory bronchiolitis–associated ILD, 2260t, 2263–2264
- Respiratory disorders. *See also* Lung disease; Respiratory tract infections
  - air pollution and, 3739, 3740f
  - approach to the patient, **2199**
  - aspirin-exacerbated, 2227
  - categories of, 2199, 2199t
  - chest pain in, 2199–2200
  - clinical features of, 2199–2200
  - clubbing in, 2200
  - cough in, 2199
  - diagnosis of, **2200, 2207**
    - allergy testing, 2215
    - arterial blood gases, 2204
    - artificial intelligence/deep learning in, 2213, 2213f
    - bronchoscopy. *See* Bronchoscopy
    - chest radiography, 2201, 2210–2211, 2211f
    - CT, 2201, 2211–2212, 2212f
    - exhaled breath condensate, 2214–2215
    - gas exchange measurement, 2204
    - mediastinoscopy, 2208
    - mediastinotomy, 2208
    - MRI, 2212, 2213f
    - PET, 2213
    - pleural biopsy, 2208
    - pulmonary function testing, 2200–2201
    - sputum analysis. *See* Sputum analysis
    - sweat testing, 2215
    - thoracentesis, 2208
    - thoracoscopy, 2208, 2208f
    - thoracotomy, 2208
    - transthoracic needle aspiration, 2213–2214
    - ventilatory function measurement, 2206–2207
  - dyspnea in, 269t, 270, 271, 2199
  - hemoptysis in, 2199
  - history in, 2199–2200
  - inhaled agents associated with, 2200
  - obesity and, 3187
  - obstructive patterns, 2206
  - pathophysiology of, **2205**
  - physical examination in, 2200
  - pulmonary function values in, 2205–2206, 2206f
  - restrictive patterns, 2205–2206, 2206f
  - risk factors for, 2200
  - in sepsis/septic shock, 2322t, 2323
  - smoking and, 2200, 3683–3684, 3683t
- Respiratory effort–related arousal (RESA), 2279, 2279t, 2280f
- Respiratory failure
  - in ARDS. *See* Acute respiratory distress syndrome (ARDS)
  - in botulism, 1235–1236
  - in critically ill patient, 2294–2295
  - hypercapneic, 2295, 2304
  - in hyponatremia, 349
  - hypoxemic, 277, 2294–2295, 2304
  - pathophysiology of, 2294–2295
  - perioperative, 2295
  - in shock, 2294–2295
- Respiratory function
  - diffusion, 2204
  - disturbances of. *See* Respiratory disorders
  - flow-related properties of, 2202–2203, 2203f, 2296, 2297f
  - during sleep, 212
  - ventilation-perfusion heterogeneity, 2204–2205, 2205f
  - ventilatory adequacy, 2204
  - volume-related properties of, 2201–2202, 2202f, 2296
  - work of breathing, 2203–2204
- Respiratory muscles
  - assessment of strength of, 2207
  - fatigue/weakness of, 2206, 2294
  - hypoperfusion of, 2295
- Respiratory quotient, 2204
- Respiratory sinus arrhythmia, A8
- Respiratory syncytial virus (RSV) infections
  - antigenic variability in, 1533
  - clinical features of, 1533
  - epidemiology of, 1533
  - epigenetics of, 3942
  - in HCT recipient, 1157
  - otitis media, 254
  - prevention of, 996
  - prophylaxis for, 1489, 1538
  - treatment of, 1488t, 1489, 1537
- Respiratory syncytial virus (RSV) vaccine, 1005, 1538
- Respiratory tract infections, **1531**
  - ambulatory antibiotic prescribing for, 252
  - *Aspergillus*. *See* Aspergillosis, respiratory tract
  - in asthma, 2220, 2221
  - chronic *P. aeruginosa,* 1310
  - COPD exacerbations and. *See* Chronic obstructive pulmonary disease (COPD), exacerbations of
  - *C. pneumoniae,* 1476–1477
  - in cystic fibrosis, 2245
  - hemoptysis in, 274
  - lymphadenopathy in, 477
  - mastoiditis, 1119
  - in mitral stenosis, 2048–2049

- Respiratory tract infections (*Cont.*):
  - *M. pneumoniae,* 1465, 1466t
  - mucormycosis, 1719
  - *Nocardia,* 1359, 1359–1360f, 1362t
  - nontuberculous mycobacterial, 1416
  - olfactory dysfunction in, 240–241
  - pathogen entry mechanisms, 964
  - pneumonia. *See* Pneumonia
  - prevention of, 584
  - *S. aureus,* 1200
  - sporotrichosis, 1724, 1724t
  - upper
    - antibiotic prescribing for, 252–253
    - epiglottitis, **259,** 1262, 1263
    - nonspecific, 253
    - pharyngitis. *See* Pharyngitis
    - red flags for, 253
    - sinusitis. *See* Sinusitis
  - viral
    - adenovirus. *See* Adenovirus infections
    - anatomic sites of, 1532
    - approach to the patient, 1536–1537
    - in asthma, 2221
    - clinical features of, 1537
    - complications of, 1538
    - coronavirus. *See* Coronavirus infections
    - diagnosis of, 1537
    - epidemiology of, 1536
    - global considerations in, 1539
    - hantavirus. *See* Hantavirus infections
    - in HCT recipient, 1157
    - human metapneumovirus, 143t, 253, 254, 1533–1534, 1533t
    - in immunocompetent hosts, 1532
    - in immunocompromised hosts, 1532–1533
    - influenza virus. *See* Influenza
    - measles. *See* Measles (rubeola)
    - parainfluenza virus. *See* Parainfluenza virus infections
    - paramyxovirus. *See* Paramyxoviridae
    - picornaviruses. *See* Enterovirus infections; Parechoviruses; Rhinoviruses
    - polyomavirus. *See* Polyomavirus(es)
    - prevention of, 1538–1539
    - rhinovirus, 253, 1534–1535, 1537
    - risk factors for, 1536
    - RSV. *See* Respiratory syncytial virus (RSV) infections
    - syndromes, 1532
    - transmission of, 1536
    - treatment of, 1536–1537
- Respite care, 86
- Response Evaluation Criteria in Solid Tumors (RECIST), 504, 555
- Responsive neurostimulation, 3425
- REST, 3941
- Restan virus, 1659t, 1667
- Restenosis, 2104, 2128
- Resting energy expenditure (REE), 2598, 2619
- Restitution, 2516
- Restless legs syndrome, **3517**
  - in CKD, 2393
  - clinical features of, 213t, 217, 3517–3518
  - diagnosis of, 213t, 217
  - in older adults, 3893
  - in pregnancy, 3903
  - treatment of, 213t, 217, 3518
- Reston virus, 1681–1682f, 1683
- Restrictive cardiomyopathy, **2020**
  - in amyloidosis. *See* Amyloidosis, cardiac
  - clinical features of, 2002t, 2020, 2081t
  - diagnosis of, 2002t, 2081t
  - differential diagnosis of, 1905–1906, 1906t, 2081t, 2084
  - etiology of, 2020, 2020t
  - genetic considerations in, 2004t, 2010, 2020
  - treatment of, 2020
- Resveratrol, 3870, 3874f, 3875
- Retapamulin, 1184
- Reteplase
  - actions of, 956
  - for acute limb ischemia, 2175
  - contraindications to, 2120
  - domain structure of, 955f, 956
  - indications for, 956
  - for STEMI, 2120
- *RET* gene mutations
  - actions of, 3073
  - as drug target, 527
  - genetic testing for, 3086
  - in lung cancer, 617, 617f, 628, 628f
  - in MEN 2, 3080t, 3792
    - hyperparathyroidism and, 3277
    - medullary thyroid carcinoma and, 524t, 525, 3046, 3048
    - pheochromocytoma and, 3077, 3086
    - screening for, 2977–2978
  - in pheochromocytoma/paraganglioma, 3073, 3078–3079, 3079t, 3086
  - prophylactic thyroidectomy in, 3085, 3085t
  - screening tests in, 3085t, 3086
  - in thyroid cancer, 3045–3046, 3085
- Reticular activating system (RAS), 187
- Reticular dysgenesis, 458, 2799, S9
- Reticulate body, 1468, 1468f
- Reticulin stain, 441f
- Reticulocyte(s), 447f
- Reticulocyte count
  - in anemia, 450
  - correction of, 447
  - definition of, 438f
  - determination of, 447
  - in hemolytic anemia, 791, 791t
  - interpretation of, 447f
- Reticulocyte production index, 480
- Reticulohistiocytosis, multicentric, 404
- Reticulospinal pathway, 169f
- Reticulotype, 32, 34f
- Reticulum cell sarcoma (interdigitating dendritic cell sarcoma), 882
- Retifanlimab-dlwr, 534t, 559f, 2787t
- Retina
  - detached, 222, 229–230, 229f
  - early degeneration of, 3374t
  - functions of, 220
  - gyrate atrophy of, 3369t
  - hemorrhages in
    - etiology of, 225, 225f
    - in high-altitude illness, 3755
    - in malaria, 1763, 1765f
  - lesions of, 222
  - radiation-induced damage to, 554
- Retinal artery occlusion, 225–226, 226f, 1860
- Retinaldehyde, 2611
- Retinal dystrophy, optic nerve edema, splenomegaly, anhidrosis, and headaches (ROSAH) syndrome, 2937
- Retinal pigment epithelium, 3407
- Retinal vein occlusion, 226, 226f
- Retinitis
  - CMV, 1485, 1487, 1514, 1515, 1610
  - differential diagnosis of, 1801t
  - *T. gondii,* 1799, 1801, 1801t
- Retinitis pigmentosa, 231–232, 232f, 2435, 3374t, 3783t
- Retinoblastoma
  - hereditary (familial), 522, 523f, 524t, 3792
  - sarcomas and, 727
  - sporadic, 3792
- Retinoblastoma (Rb) protein, 530, 573, 599, 661t
- Retinoic acid, 391, 2610, 2612. *See also* All-*trans*-retinoic acid (ATRA)
- Retinoic acid-inducible gene-1-like receptors (RLRs), 2758, 2758t, 2760
- Retinoic acid receptors, 2610, 2984
- Retinoid acid X receptors (RXRs), 3025, 3025f
- Retinoids, 243, 387, 421, 2610–2611
- Retinol, 512–513
- Retinol activity equivalent, 2611
- Retinol-binding protein, 2618t
- Retinopathy
  - cancer-associated, 2791t
  - CMV, 1514, 1514f
  - diabetic. *See* Diabetic retinopathy
  - hypertensive, 226, 226f
  - paraneoplastic, 751
  - in sickle cell disease, 774t
- Retrobulbar tumor, of optic nerve, 232
- Retrocalcaneal bursitis, 2973
- Retrocollis, 3510
- Retrograde amnesia, 207
- Retrograde ejaculation, 3155, 3227
- Retrograde urography, 2456
- Retromer complex, 3400
- Retroperitoneal adenopathy, 477
- Retroperitoneal fibrosis, 2931, 2931t
- Retroperitoneal hemorrhage, 929
- Retroperitoneal lymph node dissection, 705, 707f
- Retropharyngeal abscess, 259, 1374
- Retropharyngeal mass, A16
- Retropharyngeal space infections, 1374
- Retroviridae. *See also specific viruses*
  - cancer development and, 520, 525
  - gene expression and replication in, 1481, 1551, 1551f
  - for gene therapy delivery, 1484, 1553, 3823t, 3825t
  - life cycle of, 1551–1553, 1551f
  - structure of, 1478f, 1480t, 1551
  - subfamilies of, 1550, 1550t
  - virology of, 1550–1551
- Rett syndrome, 3403, 3783, 3941
- Revaprazan, 2526
- Reverse cholesterol transport, 3239
- Reverse T$_3$ (rT$_3$), 3021f
- Reverse transcriptase enzyme, HIV, 1558
- Reverse transcriptase inhibitors. *See* Combination antiretroviral therapy (cART)
- Reverse transcriptase polymerase chain reaction (RT-PCR), 976, 1590, 1590t
- Reverse vaccinology, 984
- Reversible cerebral vasoconstriction syndrome, 3448
- Reversible posterior leukoencephalopathy, 591, 3447–3448, 3455
- Revesz syndrome, 3819
- Revised Cardiac Risk Index (RCRI), 3905, 3907f, 3907t
- Rev/*rev,* 1552, 1552f, 1561f
- Revumenib, 833t
- Rewarming
  - for frostbite, 3772, 3772t
  - for hypothermia, 3770–3771, 3770t
- Reye's syndrome, 136, 1505, 1507, 1548
- Rezafungin, 1692
- RFC1-related CANVAS syndrome, 3537
- Rh$_0$(D) immune globulin, 924
- Rhabdomyolysis
  - AKI in, 2377, 2379t, 2380, 2384
  - in heatstroke, 3775
  - hyperkalemia in, 359
  - hypernatremic, 352
  - hypocalcemia in, 365
  - hypokalemia and, 356
  - hypophosphatemia-induced, 3264
  - skin infections and, 1053
  - treatment of, 2384

- Rhabdomyoma, cardiac, 2086, 2087f
- Rhabdomyosarcoma, 521t, 727, 728, 3852f. *See also* Soft tissue sarcomas
- Rhabdoviridae
  - clinical syndromes of, 1666t, 1673
  - geographic distribution, 1664–1665t
  - reservoirs and vectors of, 1662t, 1663t
  - structure of, 1478f, 1479, 1480t
  - virology of, 1656–1657
- *RHAG* gene mutations, 794t, 795
- *RHBDF2* gene mutations, 642
- RHC (red hair color) phenotype, 599–600
- Rheumatic fever
  - cardiovascular complications of
    - antithrombotic prophylaxis for, 3445t
    - aortic valve disease, 2035–2036, 2043
    - diagnosis of, 2855, 2856f, 2857t
    - echocardiography in, 2855, 2856f, 2857t, 2859v
    - global considerations in, 1206–1207, 1207f, 1979, 2033, 2034f, 2854, 3851
    - heart failure, 2858
    - mitral stenosis, 288, 2048, 2855, 2859v. *See also* Mitral stenosis
    - mitral valve prolapse, 2058
    - myocarditis, 2013
    - pathogenesis of, 2854, 2855f
    - pericarditis, 2083
    - prophylaxis for, 2858, 2858t
    - staging of, 2855, 2856t
  - clinical features of, 139t, 2855–2856
    - chorea, 2856, 2858
    - erythema marginatum, 139t, 146, 394, 2856
    - fever, 2856
    - joint involvement, 1060, 2855–2856
    - nodules, 2856
  - diagnosis of, 2856–2857, 2857t
  - epidemiology of, 2854
  - etiology of, 139t
  - genetic considerations in, 2854
  - global considerations in, 2854
  - group A streptococcal infection and, 2856
  - immunopathology of, 2790t, 2854
  - inflammasome mutations in, 2762t
  - pathogenesis of, 2854, 2855f
  - prevention of, 2858, 2858t
  - prognosis of, 2858
  - treatment of, 2857–2858
- Rheumatic heart disease. *See* Rheumatic fever, cardiovascular complications of
- Rheumatoid arthritis (RA), **2839**
  - approach to the patient, 2852
  - autoantibodies in, 2825, 2825t
  - clinical features of, 2839–2842
    - anemia, 450
    - cardiovascular, 2840–2841
      - aortic aneurysm, 2166
      - AV conduction block, 1927
      - pericarditis, 2083
    - constitutional, 2840
    - cutaneous, 403, 2840, A5
    - extraarticular, 2840, 2841f
    - hematologic, 2841
    - lymphoma, 2841
    - neck pain, 132
    - neuropathy, 3605
    - oral pain, 265
    - pulmonary, 2265–2266, 2840
    - Raynaud's phenomenon, 2178
    - Sjögren's disease, 2840
    - spine, 2840
    - T cell large granular lymphocytic leukemia, 2841
    - vasculitis, 2841
    - wrists and fingers, 2839–2840, 2840f
  - conditions associated with, 2841–2842, 2841f
  - diagnosis of, 47t, 2846–2848, 2847t, 2848f, 2854v
  - differential diagnosis of, 2847, 2889
  - disability in, 2848
  - EBV and, 2843
  - environmental factors in, 2843
  - epidemiology of, 2842, 2842f
  - genetic considerations in, 494–495, 2842–2843, 3942
  - global considerations in, 2842, 2842f, 2853
  - KIRs with, 2772t
  - microbiome and, 3837
  - natural history of, 2848
  - pathogenesis of, 2779, 2790t, 2825t, 2844–2845, 2845f
  - pathology of, 2843–2844
  - prognosis of, 2848
  - remission in, 2852, 2852t
  - in SLE, 2829
  - treatment of, **2848**
    - approach to, 2848–2849, 2852
    - assistive devices, 2853
    - biologic agents, 2849, 2850–2851t, 2851
    - complementary and integrative therapies, 3926
    - DMARDs, 2849, 2850–2851t, 2851
    - global considerations in, 2853
    - glucocorticoids, 2849
    - NSAIDs, 2849
    - in older adults, 2853
    - physical therapy, 2853
    - in pregnancy, 2853
    - surgery, 2853
    - timing of, 3954
    - TNF-α inhibitors, 2784
- Rheumatoid nodules, 403
- Rheumatoid vasculitis, 2772t, 2841
- Rheumatologic diseases. *See* Musculoskeletal disorders
- Rhinitis, **2809,** 2811f
- Rhino-orbital-cerebral mucormycosis, 995, 1718–1719
- Rhinophyma, 391
- Rhinoscleroma, 1289
- Rhinosinusitis, 255, 1534, 2222. *See also* Sinusitis
- Rhinoviruses, 1534–1535, 1534t
- Rhinovirus infections, 143t, 253, 1534–1535, 1537
- *Rhizobium radiobacter,* 1268, 1268t
- *Rhizopus* spp., 1718, 1718t, 1719f. *See also* Mucormycosis
- *Rhodococcus* spp. infections, 1226, 1597
- Rho kinase, 3156f
- Rhonchi, 1984, 2200
- Rh RBC blood cell group system, 901, 903, 904t
- Ri antibodies, 743t, 749
- Ribavirin
  - actions of, 1488–1489, 1494, 1537
  - for adenovirus infection, 1157
  - adverse effects of, 1116, 1489, 1494, 2661, 2687
  - for chronic HCV infection, 1492t, 1494, 2686–2687, 2688–2691, 2688t
  - drug interactions of, 2672
  - for fulminant hepatitis, 2661
  - genetic variations in response to, E72t
  - pharmacology of, 1494
  - in pregnancy, 2687
  - for RSV infection, 918, 1157, 1488t, 1489, 1537
  - for viral encephalitis, 1116
  - for viral hemorrhagic fevers, 1675–1676, S4
- Ribociclib
  - action and targets of, 533t, 570t, 573
  - adverse effects of, 570t, 753t
  - for breast cancer, 573, 637, 638
  - interactions and issues, 570t
- Riboflavin (vitamin B_2), 243, 2600t, 2606f, **2608**
- Riboflavin (vitamin B_2) deficiency, 2605t, 2608
- Ribosomal RNA (rRNA), 978
- Rice water stool, 1327, 1327f
- Richardson syndrome. *See* Progressive supranuclear palsy syndrome
- Richter's transformation, in CLL, 853–854
- Rickets
  - clinical features of, 2605t
  - diagnosis of, 3270–3271
  - hereditary vitamin D–resistant, 3270
  - in hypophosphatasia, 3316
  - hypophosphatemic, 2368t, 3264
  - pathogenesis of, 2605t, 3270
  - vitamin D–dependent type I (pseudo-vitamin D–resistant), 3270, 3290
  - vitamin D–dependent type II, 3290
- Rickettsia spp., 139t, 968, 1454, 1455t, 1456, 1459
- *Rickettsia akari. See* Rickettsialpox
- *Rickettsia conorii,* 992, 1458. *See also* Mediterranean spotted (boutonneuse) fever
- Rickettsial diseases, **1454**
  - boutonneuse fever, 139t
  - clinical features of, 139t, 145, 402, 1454, 1455t
  - diagnosis of, S11
  - ehrlichioses, 1460–1461
  - endemic typhus. *See* Endemic (murine) typhus
  - epidemic typhus. *See* Epidemic (louse-borne) typhus
  - epidemiology of, 1454, 1455t
  - etiology of, 1454, 1455t
  - global considerations in, 992
  - human granulocytic anaplasmosis, 1452, 1455t, 1460f, 1461–1462
  - Mediterranean spotted (boutonneuse) fever, 992, 1458
  - Q fever. *See* Q fever
  - rickettsialpox. *See* Rickettsialpox
  - Rocky Mountain spotted fever. *See* Rocky Mountain spotted fever (RMSF)
  - scrub typhus. *See* Scrub typhus
  - tick-borne spotted fevers, 1458
- Rickettsialpox, **1458**
  - clinical features of, 143t, 147, 1455t, 1458
  - diagnosis of, 1458
  - differential diagnosis of, 1506
  - epidemiology of, 143t, 1049, 1455t, 1458
  - etiology of, 143t
  - rash in
    - eschar, 143t, 147, 1458f, A1
    - evolution of, 143t, 1049, 1458, 1459f
    - papules and plaques, 143t, 147, 402, 1049, A1
    - vesicles/pustules, 147, 1459f
  - transmission of, 1455t, 3730
  - treatment of, 1459
  - vs. varicella, 147, 1506
- *Rickettsia prowazekii,* 992, 1454, 1458. *See also* Epidemic (louse-borne) typhus
- *Rickettsia rickettsii,* 992, 1456. *See also* Rocky Mountain spotted fever (RMSF)
- *Rickettsia typhi,* 1454. *See also* Endemic (murine) typhus
- Ridley-Jopling classification, leprosy, 1406
- Riedel's thyroiditis, 3039
- Rifabutin
  - actions of, 1167, 1182t, 1424
  - adverse effects of, 1418, 1424
  - drug interactions of, 1742t
  - for HIV-associated TB, 1400
  - in *H. pylori* eradication, 2528, 2529f, 2530t
  - for MAC infection, 1417, 1424
  - for MAC infection prophylaxis, 1417, 1593t
  - pharmacology of, 1424
  - resistance to, 1182t, 1185, 1424
  - for TB, 1424
  - for TB prophylaxis, 1593t

- Rifampin
  - for *A. baumannii* infections, 1297t
  - actions of, 1167, 1182t, 1183f, 1422
  - adverse effects of, 1423
    - allergic interstitial nephritis, 2438
    - gastrointestinal, 1178
    - hepatotoxicity, 1171t, 1178, 1400
    - hyperbilirubinemia, 323
    - in leprosy, 1411
    - lupus syndrome, 2835
    - nephritis, 2356, 2377
    - thrombocytopenia, 923t, 1397
  - for anthrax, S4
  - for asymptomatic GAS carriage, 1209
  - for *Brucella* infections, 1334
  - for Carrión's disease, 1353t
  - for cat-scratch disease, 1353t
  - dosage of, 1423
  - drug interactions of, 489t, 1172t, 1400, 1402, 1423, 2721
    - atovaquone, 1742t
    - dapsone, 1743t
    - quinidine, 1745t
  - genetic variations in response to, E72t
  - indications for, 1174t, 1178
  - for infective endocarditis, 1043t, 1044, 1205
  - for leprosy, 1410, 1411t
  - for leprosy prophylaxis, 1414
  - for MAC infections, 1426
  - for meningitis, 1122, 1194
  - for meningitis prophylaxis, 1122, 1180t
  - metabolism of, 1178
  - for NTM infections, 1418, 1427, 2243
  - for osteomyelitis, 1064t, 1065, 1065t
  - pharmacology of, 1422–1423
  - for prosthetic joint infections, 1060, 1067, 1205
  - for Q fever, 1464
  - resistance to
    - detection of, 988
    - mechanisms of, 1182t, 1185
    - in *M. leprae,* 1411
    - in *M. tuberculosis,* 1397–1399, 1398f, 1423
    - in mycobacteria, 981, 988, 1423
    - in *N. meningitidis,* 1253
    - prevalence of, 1174t
  - for TB
    - active infection, 1394–1396, 1394t, 1395t, 1419t, 1420t, 1422–1423
    - isoniazid-resistant, 1398
    - monitoring and management of, 1420t
  - for TB preventive treatment, 1402, 1402t, 1419, 1419t, 1420t, 1593t
  - for tuberculous meningitis, 1127
- Rifamycins, 1167, 1178, 1182t
- Rifapentine
  - actions of, 1167, 1182t
  - adverse effects of, 1424
  - drug interactions of, 1402
  - pharmacology of, 1424
  - resistance to, 1182t, 1185, 1424
  - for TB, 1395, 1424
  - for TB preventive treatment, 1402–1403, 1402t, 1419t, 1424
- Rifaximin
  - for diverticular disease, 2580
  - for dyspepsia, 302
  - for hepatic encephalopathy, 2717
  - for IBS, 312, 2576, 2577t
  - for peritonitis prophylaxis, 1071
  - for travelers' diarrhea, 307, 1019, 1081t
  - for travelers' diarrhea prophylaxis, 1081
- Rift Valley fever, 993, 1662t, 1665, 1677, 3750, 3753f, 3858
- Right atrial overload, ECG in, 1871f
- Right atrial pressure, in shock, 2293, 2294f
- Right atrium, 1935f
- Right bundle branch, 1925
- Right bundle branch block
  - in Chagas disease, 1791
  - conditions associated with, 1871–1872
  - ECG in, 1791f, 1871, 1872f, A7, A8
  - heart sounds in, 1865f
- Right heart dilation, congenital, 2070, 2070t
- Right-left confusion, 203
- Right-to-left shunt
  - intracardiac, 1907
  - intrapulmonary, 277, 278
- Right ventricular dysfunction
  - after atrial switch procedure, 2075, 2075t
  - cardiogenic shock in, 2328f, 2329–2330
  - diagnosis of, 1882–1883, 1902v, A7
  - etiology of, 1882–1883
  - in pulmonary embolism pathophysiology, 2158, 2158f
- Right ventricular failure/infarction, 2081t, 2123, 2158, 2187–2188, 2189f
- Rigidity, 168
- Rilonacept, 2792t, 2936
- Rilpivirine, 1616, 1617t, 1619f, 1621, 1621t, 1622f
- Riluzole, 3524
- Rimantadine, 1488, 1488t, 1537, 1594t
- RIME (reactive infectious mucocutaneous eruption), 143t
- Rimegepant, 3467t, 3468–3469, 3468t, 3470t
- Ring(s), esophageal, 2505–2506, 2505f
- Ringworm, 1723. *See also* Dermatophyte infections
- Rinne test, 248, 3380
- Rio Bravo virus, 1660t
- Riociguat, 2194, 2196, 2197t, 2874
- Rio Mamoré virus, 1661t, 1674
- Ripretinib, 533t, 729
- Ri protein, 2791t
- Risankizumab, 387t, 2566, 2568t, 2793t, 2890, 2892
- Risedronate
  - adverse effects of, 3304
  - for fracture prevention, 3305f
  - for osteoporosis management/prevention, 3145, 3304–3306, 3305f, 3311
  - for Paget's disease of bone, 3303, 3314t
- Risperidone
  - adverse effects of, 3188, 3673t
  - for bipolar disorder, 3668
  - for delirium, 86t
  - overdosage/poisoning with, 3711t
  - for psychosis in Huntington's disease, 3514
  - for schizophrenia, 3673, 3673t, 3674
- Ritodrine, 354
- Ritonavir
  - actions of, 1621
  - adverse effects of, 1617t, 2672
  - for COVID-19 disease, 996
  - drug interactions of, 485t, 1742t, 2721, 3071
  - for HIV infection, 1617t, 1621
  - metabolism of, 485t
  - molecular structure of, 1619f
  - resistance to, 1623f
- Rituximab
  - actions and targets of, 534t, 558, 559f, 2791t, 2851
  - adverse effects of
    - B cell depletion, 579
    - common, 2850t, 2851
    - cutaneous, 419
    - cytokine release syndrome, 594
    - desensitization protocol for, 2817f, 2819
    - DIHS, 597
    - IgM flare, 894
    - infections, 2851, 2895, 2916t
    - infusion reaction, 594
    - long-term, 755
    - paraneoplastic pemphigus, 410
    - PML, 757, 2851, 2916t
    - pulmonary, 2261
    - serious, 2850t, 2896t
    - tumor lysis syndrome, 593
  - for ALL, 848, 848t
  - for antiphospholipid antibody syndrome, 2446, 2839
  - for autoimmune hemolytic anemia, 802, 853
  - for autoimmune hepatitis, 2696
  - for autoimmune pancreatitis, 2752
  - for B-cell prolymphocytic leukemia, 872
  - for bullous pemphigoid, 412
  - for Burkitt's lymphoma, 862
  - for cardiac transplant rejection, 2030
  - for Castleman's s disease, 1517
  - for CLL, 855
  - for coagulation factor inhibitor eradication, 931
  - for cold agglutinin disease, 803
  - for dermatomyositis, 414
  - for DLBCL, 536, 864
  - for EGPA, 2900
  - for epidermolysis bullosa acquisita, 413
  - for follicular lymphoma, 864
  - for gamma heavy chain disease, 894
  - for gastric lymphoma, 650
  - for granulomatosis with polyangiitis, 2899
  - for hairy cell leukemia, 873
  - for hemolytic-uremic syndrome, 595
  - for hepatitis C–associated cryoglobulinemic vasculitis, 2905
  - for IgG4-related aortitis, 2171
  - for IgG4-related disease, 2933
  - for immune thrombocytopenic purpura, 924
  - for inflammatory myopathies, 2916t, 2917
  - for insulin autoimmune syndrome, 3093
  - for kidney transplant rejection, 2410
  - for lymphoplasmacytic lymphoma, 865
  - for mantle cell lymphoma, 866
  - for marginal zone lymphoma, 865
  - for membranous glomerulonephritis, 2423
  - for MGUS, 889–890
  - monitoring during treatment with, 2850t
  - for MS, 3585, 3586t
  - for mucous membrane pemphigoid, 413
  - for myasthenia gravis, 3631
  - for neuromyelitis optica, 3593
  - for PCNSL, 721
  - for pemphigus vulgaris, 410
  - for rheumatoid arthritis, 2850t, 2851
  - for Sjögren's disease, 2879, S1
  - for SLE, 2835
  - for splenic marginal zone lymphoma, 872
  - for systemic sclerosis, 2874
  - for thrombotic thrombocytopenic purpura, 925
  - for vasculitic neuropathy, 3624
  - for vasculitis, 2895
  - for Waldenström's macroglobulinemia, 894
- Rivaroxaban
  - adverse effects of, 953, 1951
  - for anticoagulation in HIT, 923
  - dosage of, 952–953, S13
  - indications for, 952
  - management before endoscopic procedures, 2483t
  - management before lumbar puncture, S3
  - monitoring treatment with, 953
  - for NSTE-ACS, 2110t
  - for peripheral artery disease, 2173
  - pharmacology of, 953t
  - in pregnancy, 954
  - reversal of, 3456
  - for stroke prevention, 1951, 1951t, 3450
  - for VTE/PE treatment, 2162, 2163t
  - for VTE prophylaxis, 2164t, 3910

- Rivastigmine, 178, 199, 3480, 3491, 3493, 3506
- River blindness. *See Onchocerca volvulus*/*onchocerciasis*
- Rizatriptan, 3467t, 3468, 3468t, 3933
- RLRs (retinoic acid-inducible gene-1-like receptors), 2758, 2758t, 2760
- Rmp protein, 1255
- *RMRP* gene mutations, 457, 2800, S9
- RMSF. *See* Rocky Mountain spotted fever (RMSF)
- RNA methylation, 3937
- RNA polymerase (RNP), 2790t, 2864t
- RNA vaccines, 997t, 998
- RNA viruses, 1478f, 1479–1481, 1480t. *See also specific viruses*
- *RNF216* gene mutations, 3112t
- Ro (reproduction number), 987
- Robotic bronchoscopy, 2210, 2210f
- Robotics, for neurologic rehabilitation, 3963–3964, 3964f
- ROC (receiver operative characteristic curve), 24, 24f
- Rocio virus, 1660t
- "Rocker foot," 2969
- Rocky Mountain spotted fever (RMSF), **1456**
  - clinical features of, 1455t, 1456–1457
    - cutaneous
      - characteristics of, 137, 140t, 146f, 1121
      - evolution of, 142f, 146, 408, 992, 1121, 1456
      - petechial rash, 992, 1456, 1456f, A1, A5
      - purpura, 408
    - myocarditis, 2013
    - vascular damage, 927
  - diagnosis of, 1457
  - differential diagnosis of, 146, 1121, 1457
  - epidemiology of, 1455t, 1456
  - etiology of, 991t, 992
  - fulminant, 1457
  - pathogenesis of, 1456
  - prevention of, 1457–1458
  - treatment of, 991t, 992, 1457
- Rocuronium, 2352, 3695
- Rod(s), 220
- Rodent(s)
  - bite-wounds, 1143, 1145t
  - in plague transmission, 1343, S4
- Rodent-borne viruses, **1655**. *See also specific viruses*
  - clinical syndromes of, 1663, 1665, 1666t
    - arthritis and rash, 1666t, 1669–1670
    - encephalitis, 1666t, 1670–1674
    - fever and myalgia, 1665–1669, 1666t
    - hemorrhagic fevers. *See* Viral hemorrhagic fevers
    - pulmonary disease, 1666t, 1674
  - diagnosis of, 1663
  - epidemiology of, 1663
  - geographic distribution of, 1664–1665t
  - reservoirs and vectors of, 1657–1663t
  - transmission of, 1656
- *Rodentolepis nana* (dwarf tapeworm), 961t, 1750, 1838
- Roflumilast, 2256
- Roll Back Malaria campaign, 3849
- Romaña sign, 1790t
- Romberg maneuver, 3382
- Romberg's sign, 164, 172, 180
- Rombo's syndrome, 606
- Romidepsin, 533t, 539, 570t, 573, 867, 874
- Romiplostim, 924
- Romosozumab, 3309, 3310f
- Root canal therapy, 260, 267
- Root cavity, A3
- Ropinirole, 213t, 217, 3504, 3504t, 3518
- ROS. *See* Reactive oxygen species (ROS)
- *ROS1* gene mutations, 617, 617f, 617t, 628, 628f
- Rosacea. *See* Acne rosacea
- ROSAH (retinal dystrophy, optic nerve edema, splenomegaly, anhidrosis, and headaches) syndrome, 2937
- Rosenbaum card, 221
- Roseola (exanthem subitum), 138t, 145, 1516, A1
- Rose spots, 1316f, A1
- Rosiglitazone
  - actions of, 2611, 3214
  - adverse effects of, 3211t, 3214
  - contraindications to, 3211t
  - for diabetes mellitus, 3211t, 3214
- Ross procedure, 2040
- Ross River virus disease (epidemic polyarthritis), 1059, 1657t, 1669–1670
- Rosuvastatin, 2110, 3228, 3244t, 3257. *See also* Statins
- Rotational atherectomy, 2128
- Rotator cuff tears, 2974
- Rotator cuff tendinitis, 2943, 2973–2974
- Rotaviruses, 1479, 1628f, 1628t, 1630
- Rotavirus infections
  - climate change and, 3752t
  - clinical features of, 297, 306t, 1631
  - diagnosis of, 306, 1628t, 1631
  - epidemiology of, 1079, 1079t, 1630–1631, 1630f
  - global considerations in, 1630–1631, 1630f, 1632f
  - immunity to, 1631
  - incidence pre- and postvaccine, U.S., 1002t
  - pathogenesis of, 1631
  - prevention of, 1631, 1632f
  - treatment of, 1631
- Rotavirus vaccine, 1631, 1632f
- *Rothia mucilaginosa,* 583
- Roth's spots, 225, 225f, 995
- Rotigotine, 3502, 3504t, 3518
- Rotor syndrome, 323, 2641t, 2642
- Rouleaux formation, 439f, 443
- Round pneumonia, 1192
- Roundworms, 1736–1737. *See also* Nematodes/nematode infections
- Rous sarcoma virus, 520, 1550t
- Roux-en-Y gastric bypass, 3194, 3194f, 3195, 3235
- Rove beetle, 3737
- Rovsing's sign, 2595, 2595t
- Rowasa enemas, 2564
- Roxifiban, 923t
- Rozanolixizumab, 2792t, 3593, 3631, 3632t
- *RPE65* gene mutations, 3824t, 3827
- *rpo*B gene, 1397
- RPR (rapid plasma reagin) test, 1433
- *RPS6KA3* gene mutations, 661t
- rRNA (ribosomal RNA), 978
- RS3PE syndrome, 2847
- RSDL (reactive skin decontamination lotion), S5
- RSV604, 1489
- RSV infections. *See* Respiratory syncytial virus (RSV) infections
- rT$_3$ (reverse T$_3$), 3021f
- RTA. *See* Renal tubular acidosis (RTA)
- RTEL1, 3818
- *RTEL1* gene mutations, 3819, 3820t
- RTH (resistance to thyroid hormone), 3024t, 3025, 3786
- rtPA (recombinant tissue plasminogen activator). *See* Alteplase (rtPA)
- RT-PCR (reverse transcriptase polymerase chain reaction), 1590, 1590t
- RT-QuIC assay, 3531
- Rubber band ligation, 2584
- Rubella virus, 1643
- Rubella virus infection (German measles), **1643**
  - clinical features of, 138t, 402, 1643
    - arthralgia, 1059
    - oral, 261
    - skin, 137, 138t, 402, 1644f, A1
  - congenital, 1002t, 1118, 1643–1644, 1644t, 3904
  - diagnosis of, 1644–1645
  - differential diagnosis of, 1511t
  - epidemiology of, 1643
  - global considerations in, 1645–1646, 1645f
  - incidence pre- and postvaccine, U.S., 1002t
  - pathogenesis of, 1643
  - in pregnancy, 137, 1644–1645, 3904
  - prevention of, 1645–1646, 1645f
  - treatment of, 1645
- Rubella virus vaccine, 1645–1646, 1645f, 3904. *See also* Measles-mumps-rubella vaccine (MMR)
- Rubeola. *See* Measles
- Rubidium-82, 1878t, 1879
- Rubinstein-Taybi syndrome, 3783t, 3941
- *Rubivirus,* 1643. *See also* Rubella virus infection
- Rubor, 454f, 2172
- Rubrospinal pathway, 169f
- Rucaparib, 533t, 571t, 703, 712, 3982
- Rufinamide, 3423t
- Rumination, 296, 2504
- Rumination disorder/syndrome, 298, 3670–3671
- *RUNX1* gene mutations, 815t, 816, 825, 825t, 828, 836, 878
- *Runx2,* 3259
- Russell body gastritis, 2539
- Russell's viper bite, 3717, 3718f, 3722. *See also* Snakebites
- Ruxolitinib
  - actions of, 533t, 567, 568t, 2187f
  - adverse effects of, 568t, 821, 823
  - for atopic dermatitis, 383
  - for atypical chronic myeloid leukemia, 878
  - for GVHD, 917
  - for hemophagocytic lymphohistiocytosis, 466, 882
  - interactions and issues, 568t
  - for polycythemia vera, 820–821
  - for primary myelofibrosis, 823
- RV (residual volume), 2202, 2202f, 2203f, 2206f
- RV144, 1627
- R value, 2667
- R wave, 1868, A7
- *RYR1* gene, 493, E72t
- *RYR2* gene mutations, 2004t
- *RYR3* gene mutations, 1582t, 1585

S

- S-100 protein, in carcinoma of unknown primary, 733t
- SAAG (serum ascites-albumin gradient), 328
- Saaremaa virus, 1661t, 1676
- Sabiá virus, 1658t
- Saccades, 221, 237
- Saccadomania, 237
- Saccharopinuria, 3369t
- Saccular aneurysm, 2165, 2166, **3458**. *See also* Aortic aneurysm; Subarachnoid hemorrhage
- Sacituzumab govitecan
  - actions of, 535t, 566t
  - adverse effects of, 597, 694
  - for bladder cancer, 694
  - for breast cancer, 641
  - indications for, 566, 566t
- Sacral insufficiency fracture, A16
- Sacral nerve modulation, 336, 2583. *See also* Neuromodulation therapy
- Sacral nerve stimulation, 2584
- Sacral spinal lesions, 3560
- Sacroiliac joint pain, 119–120, 121t, 123t, 127
- Sacroiliitis
  - in axial spondyloarthritis, 2883
  - in familial Mediterranean fever, 2934
  - grading of, 2883, 2883t

- Sacroiliitis (*Cont.*):
  - in IBD, 2562t, 2563
  - in psoriatic arthritis, 2888
- SACT (sinoatrial conduction time), 1918
- Sacubitril-valsartan, 1988, 1992, 1995, 1995t
- Saddle nose deformity, 1860, 2897, 2918, 2919f
- Safety in health care, **42**. *See also* Adverse events
- Safety theory, 42
- Safinamide, 3503, 3504t
- Sagging brain syndrome, 3483
- St. John's wort, 434t, 489t, 2137, 3927
- St. Louis encephalitis, 1660t, 1671–1672
- St. Vitus' dance. *See* Sydenham's chorea
- *salA* gene, 1216
- Salicylates
  - adverse effects of, 246, 370, 376
  - drug interactions of, 489t, 564
  - overdosage/poisoning with, 370, 3705, 3714t
  - for rheumatic fever, 2857
- Salicylic acid
  - for acne vulgaris, 391
  - for prickly heat, 3774
  - for scabies, 3729
  - for tinea versicolor, 389
  - for warts, 390
- Salience network, 207
- Saline
  - half-normal, for hypovolemia, 347
  - hypertonic, 351, 358, 2345t
  - hypotonic, 347, 353
  - isotonic, 347, 353, 375, 3718
  - normal
    - for hypercalcemia, 365
    - for hypercalcemia of malignancy, 738
    - for hyperkalemia, 361
    - for hyponatremia, 351
    - for hypovolemia, 347
    - for SIAD, 738, 3020
    - for sinusitis, 256, 257
    - for sore throat, 258
- Saline infusion test, 3061
- Saliva, 243, 266, 1564
- Salivary duct obstruction, 266
- Salivary gland(s)
  - benign tumors of, 610, 614
  - diseases of, 266–267
  - enlargement of, 2878, 2878t
  - IgG-4 related lesions of, 2931, 2931t, 2932f
  - malignant tumors of, 266, 610, 614
- Salivary stone (sialolithiasis), 266, A3
- Saliva substitutes, 266
- Salmeterol, 1422, 2224, 2255
- *Salmonella* spp.
  - antibiotic resistance in, 1174t, 1319–1320, 1321
  - chronic carriage of, 1317, 1318
  - inherited susceptibility to, 461f
  - laboratory identification of, S11
  - *S.* enteritidis, 1319
  - *S.* typhimurium. *See Salmonella* Typhimurium
  - subspecies/serotypes, 1314–1315
- *Salmonella* spp. infections, **1314**. *See also* Enteric (typhoid) fever
  - aortitis, 2171
  - bacteremia, 1320
  - bone, joint, and soft tissue, 1320
  - climate change and, 3752t
  - clinical features of, 1060
  - CNS, 1320
  - diagnosis of, 1080, 1320, S11
  - endovascular, 1320
  - epidemiology of, 1319–1320
  - etiology of, 1314–1315
  - food-borne, 2559
    - clinical features of, 305, 306t, 1080t, 1320
    - epidemiology of, 1319
    - pathogenesis of, 1315
    - treatment of, 1081, 1081t, 1320–1321, 1321t
  - genital, 1320
  - global considerations in, 1319
  - in HIV infection, 1594t, 1598
  - intraabdominal, 1320
  - prevention and control of, 1321
  - pulmonary, 1320
  - splenic abscess, 1075
  - treatment of, 1174t, 1320–1321, 1321t
  - UTIs, 1320
- *Salmonella* Paratyphi, 1314–1315. *See also* Enteric (typhoid) fever
- *Salmonella* Typhi, 1314–1315. *See also* Enteric (typhoid) fever
- *Salmonella* Typhimurium
  - adherence of, 967
  - antibiotic resistance in, 1186t, 1319–1320
  - apoptosis inhibition by, 973
  - autophagy inhibition by, 973
  - as bioweapon, S4
  - cytotoxins, 974
  - entry into host, 964
  - immune response inhibition by, 974
  - inflammasomes of, 972
  - MAPK signaling inhibition by, 971
  - secretion systems, 968
  - serotypes, 1314–1315
- Salmonellosis, nontyphoidal. *See Salmonella* spp. infections
- Salmon River virus, 1657t
- Salp15, 1449
- Salpingitis
  - abdominal pain in, 112t, 113, 1105
  - clinical features of, 1105
  - *C. trachomatis,* 1471, 1473t
  - diagnosis of, 1105–1106
  - treatment of, 1107
- Salpingo-oophorectomy, 710, 712
- Salsalate, 2955t
- Salt-depletion heat exhaustion, 3774
- Salt therapy, 351, 375
- Salt-wasting disorders, 348
- Salty taste sensation, 238
- *Salvia divinorum,* 3696
- SAM, 3939
- ^{153}Samarium therapy, 82, 554, 731
- SAME. *See* Syndrome of apparent mineralocorticoid excess (SAME)
- Same-sex attraction, 3180t
- Sand flea, 3736
- Sandfly
  - in Carrión's disease, 1356
  - in Coclé virus disease, 1668
  - in *Leishmania* infections, 1781, 1782t
  - in pemphigus foliaceus, 410
  - in Punta Toro virus disease, 1668
  - in Toscana virus disease, 1667–1668
- Sandfly fever, S7
- Sandfly fever viruses, 1662t, 1667–1668
- Sandhoff disease, 3352t, 3355, 3522t
- Sanfilippo syndromes, 3352t
- Sanjad-Sakati syndrome, 3286
- SA (sinoatrial) node, 1868, 1868f, 1912, 1917–1918, 1918f
- SA (sinoatrial) node disease/dysfunction. *See* Sinoatrial (SA) node disease/dysfunction
- Saphenous vein bypass graft, 2128, 2174, A11, A11v
- SAPHO syndrome, 1067, 2892
- SAPIEN valve, 2130f, 2131, A11
- Saponins, 997t, 998
- Sapoviruses, 1628, 1628f, 1628t
- Saquinavir, 485t, 1619f, 1623f, 1742t
- SARC-F, 2619t
- Sarcocystosis, 1808
- Sarcoglycans, 3640, 3641f
- Sarcoidosis, **2921**
  - approach to the patient, 2930
  - clinical course and prognosis of, 2928
  - clinical features of, 2923, 2924f
    - alopecia, 394
    - azotemia, 341
    - cardiac, 2925–2926, A9
      - myocarditis, 2014, 2014f
      - treatment of, 2340t
      - ventricular tachycardia, 1965, 1975
    - cough, 273
    - cutaneous, 394, 405, 406, 1860, 2927, 2927f, A5
    - extrahepatic cholestasis, 325
    - fever of unknown origin, 150
    - gastrointestinal, 300, 2928
    - hypercalcemia, 363, 2927–2928, 3281
    - hypocalcemia, 365
    - hypopituitarism, 2993
    - liver, 2927
    - lung disease, 2260t, 2263f, 2266, 2924–2925, 2925–2926f
    - lymphadenopathy, 477, 2928
    - musculoskeletal, 2927
    - neurologic, 1132t, 2926–2927
      - facial palsy, 3555
      - myelitis, 3565–3566
      - neuropathy, 3605
    - ocular, 224, 225, 2927
    - oral, 266
    - parotid glands, 2928
    - pulmonary hypertension, 2193, 2926
    - renal, 2356, 2928, A4
    - reproductive system, 2928
    - weakness, 170
  - diagnosis of, 2263f, 2923–2924, 2924t, 2925f, 2926f, A4, A16
  - differential diagnosis of, 406, 2878t, 2924, 2924t
  - drug-induced, 420
  - epidemiology of, 2014, 2921
  - etiology of, 2922
  - follow-up and monitoring, 2928–2929
  - mortality from, 2921, 2928
  - vs. occupational lung disease, 2238
  - pathogenesis of, 2922–2923, 2922f
  - pathology of, 2923–2924, 2923f
  - staging of, 2924f, 2925
  - treatment of, 2929–2930, 2929t
- Sarcomas, **727**
  - bone. *See* Bone sarcomas
  - cardiac, 2086
  - deaths from, 500t
  - definition of, 528
  - follicular dendritic cell, 882
  - granulocytic, 405
  - histiocytic, 881
  - interdigitating dendritic cell, 882
  - Langerhans cell, 882
  - mast cell, 2820
  - metastatic, 723t
  - paraneoplastic syndromes in, 737t
  - soft tissue. *See* Soft tissue sarcomas
  - synovial, 2972
- Sarcomere, 1847
- Sarcomeric proteins, 2004t, 2005f, 2009
- Sarcopenia, 2188, 2717, 3875
- Sarcoplasmic reticulum, 1847f, 1848f, 1980
- *Sarcoptes scabei,* 3729. *See also* Scabies
- Sarcosinemia, 3368t
- *sar* gene, 1197
- Sargramostim, 3499
- Sarilumab, 2850t, 2851, 2896t
- Sarin (GB), 3712t, S5, S8

SARMs (selective androgen receptor modulators), 3120–3121
SARS. *See* Severe acute respiratory syndrome (SARS)
SARS-CoV-1 (severe acute respiratory syndrome coronavirus 1), 1483, 1535. *See also* Severe acute respiratory syndrome (SARS)
SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2). *See also* COVID-19 disease; Multisystem inflammatory syndrome in children (MIS-C)
  ACE2 and, 3167, 3167f, 3942
  cell entry mechanisms, 3167, 3167f
  control of, 3857
  early global spread of, 985
  epidemiology of, 3857
  genomic sequencing of, 979, 980, 984, 985–986
  laboratory detection of, 976, 979
  Spike surface protein of, 988, 998
  structure of, 1539
  TMPRSS2 and, 3167, 3167f
  vaccine development for, 984
  variants of, 1483
  wastewater-based surveillance, 986
SASP (senescence-associated secretory phenotype), 3871
Satiety, 3191
Satralizumab, 227, 3565, 3592, 3592t, 3593
Saturated solution of potassium iodide (SSKI), 1724–1725, 1724t
Sauna taker's lung, 2228t
Saxitoxin, 3727
SBDS gene mutations, 457
SBP (spontaneous bacterial peritonitis), 329–330, 1070–1071, 2716
SBRT (stereotactic body radiotherapy), 590, 596, 622–623
SCA. *See* Spinocerebellar ataxia (SCA)
SCA (sudden cardiac arrest), 2332
Scabicides, 3729
Scabies, **3729**
  clinical features of, 378t, 3729, A5
  diagnosis of, 3729
  epidemiology of, 3729
  treatment of, 1748, 3729
Scale (skin lesion), 377t
Scale-free network, 3956, 3956f
Scalene lymph node, 477
Scalene maneuver, 2176
Scapular winging, 3634, 3637f, 3642
Scar (skin lesion), 377t
Scaritoxin, 3726
Scarlatiniform eruptions, 402, 402t. *See also* Rash
Scarlet fever (second disease)
  clinical features of, 141t, 402, 1209
  differential diagnosis of, 1209
  epidemiology of, 141t
  etiology of, 141t, 1209
  rash in
    differential diagnosis of, 137, 146
    erythematous, A1
    evolution of, 141t
    scarlatiniform, 402, 1209f
SCC*mec,* 1204
Scedosporium spp. infections, 1688t, 1726t, 1729–1730
SCF (stem cell factor), 2767t
Schatzki's ring, 294, 296, 2492, 2497f, 2506, 2512f
Scheie disease, 3352t, 3356
Schimke immunoosseous dysplasia, 2800, S9
Schinzel-Giedion syndrome, 878
Schirmer strip, 224
Schistocytes, 437, 447t, 934
Schistosoma spp. *See also* Schistosomiasis
  geographic distribution of, 961t, 1826t, 1828f, S12
  laboratory identification of, S12
  life cycle of, 1736, 1826, 1827f, S12
  S. haematobium, 1826, 1828, 1829
  S. japonicum, 1826, 1829
  S. mansoni, 1826, 1829
  transmission of, 1826, 1826t
Schistosomes, 1051
Schistosomiasis, **1826**
  bladder cancer and, 512t, 690, 691, 1829
  clinical features of, 1736, **1828,** S7
    acute, 1828–1829
    CNS, 1829
    eosinophilia, S12
    genitourinary, 1739t
    hepatosplenic, 1829
    intestinal, 1829
    myelitis, 3566
    neurologic, 1738t
    portal hypertension, 1828, 1828f
    pulmonary, 1829
    pulmonary arterial hypertension, 2193
    renal, 2416t, 2429
    skin, 1051
    swimmer's itch, 1828
    urogenital, 1829
  diagnosis of, 961t, 1829–1830, S7, S12
  epidemiology of, 961t, 1019, 1826–1827, 1826t, 1828f, S12
  etiology of. *See Schistosoma* spp.
  incubation period for, S7
  pathogenesis of, 1736, 1827–1828
  prevention and control of, 1830
  in travelers, S12
  treatment of, 1751, 1830, 1830t, S7
  in war veterans, S7
Schizoaffective disorder, 3672
Schizogony, 1761
Schizoid personality disorders, 3672
Schizont, *Plasmodium,* 1761
Schizophrenia, **3672**
  CATIE study of, 3674
  chemosensory alterations in, 242
  circadian disruption and, 3952
  clinical features of, 3672
  vs. dementia, 3485
  diabetes mellitus type 2 in, 3674
  differential diagnosis of, 3672
  epidemiology of, 3672
  genetic considerations in, 3652, 3655, 3672
  insomnia in, 215
  neuropathology of, 3655
  pathophysiology of, 3404, 3672
  treatment of, 3672–3674, 3673t
Schizophreniform disorder, 3672
Schizotypal personality disorders, 3672
Schmid metaphyseal chondrodysplasia, 3331
Schnitzler's syndrome, 150, 155, 402, 2937
Schober test, 121t
Schwann cell tumor, benign (neurofibroma), 403
Schwannomas, 721. *See also* Vestibular schwannoma (acoustic neuroma)
Schwannomatosis, familial, 717t
Sciatica, 2944f, 2945
Sciatic nerve, 3615, 3616f
Sciatic neuropathy, 3614
SCID. *See* Severe combined immunodeficiency (SCID)
Science of medicine, 1
Scintigraphy
  in fever of unknown origin, 154
  in gastrointestinal disease, 2462
  in pheochromocytoma, 3074t
  thyroid, 3026–3027
"Scissoring" gait, 178
Scl-70 (topoisomerase I), 2790t, 2946t
Sclera, blue, 1860, 3315, 3323, 3326
Scleral icterus, 320
Scleritis, 224
Sclerodactyly, 414, 2177, 2188, 2867f
Scleroderma esophagus, 2514
Scleroderma renal crisis, **2870**. *See also* Systemic sclerosis (scleroderma)
  clinical features of, 2870–2871
  diagnosis of, 2870–2871
  epidemiology of, 2870
  pathology of, 2446, 2866, 2871f, A4
  treatment of, 2384, 2446, 2875
Scleromyxedema (papular mucinosis), 404
Sclerosing cholangitis. *See* Primary sclerosing cholangitis (PSC)
"Sclerosis of the left cardioskeleton," 1927
Sclerosteosis, 3316
Sclerostin, 3294, 3309
Sclerotherapy, 2182, 2584
SCN (severe congenital neutropenia), 457, 2794–2795, 2795f, S9
SCN (suprachiasmatic nucleus), 220, 3943, 3944t
SCN1A gene mutations, 3462
SCN2A1 gene mutations, 3412t, 3420
SCN4A gene mutations, 359, 3646
SCN5A gene mutations, 1927, 1970, 1970t, 1971, 2004t, 2009
SCN8A gene mutations, 3420
SCN9A gene mutations, 2179
SCNA gene mutations, 3941
Scolex, 1810
Scoliosis, 120, 3324, 3330
Scombroid poisoning, 306, 3726
Scopolamine
  for excess secretions, 83, 91, 91t
  for nausea and vomiting, 83, 298, 299t
  overdosage/poisoning with, 3711t
  in terminal weaning, 89
  for vertigo, 165t
Scorpionfish envenomation, 3726
Scorpion sting, 3734–3735
Scotochromogens, 1426
Scotoma, 222, 223f, 230
Scrapie, 3526, 3526t
Screening tests, 2, 37. *See also specific diseases and tests*
  adverse outcomes of, 38–39
  age-specific recommendations, 41t
  evidence-based recommendations for, 39–40, 39t
  health benefits of, 38, 38t
  in newborns, 3977–3978
  vs. preventive care, 9
  principles of, 37t
Screening trials, 515
Screwworm fly, 3733
Scrub typhus, **1460**
  clinical features of, 138t, 149, 1455t, 1460, S7
  diagnosis of, 1460, S7
  epidemiology of, 138t, 1455t, 1460, 3730
  etiology of, 992, 1455t, 3730
  incubation period for, S7
  prevention of, S7
  rash in, 138t, 149, 992, 1460
  treatment of, 1460, S7
Scuba equipment, 3765–3766
Scurvy
  clinical features of, 407, 2605t, 2609, A15
  etiology of, 2609
  oral manifestations of, 65
  pathogenesis of, 3322
  treatment of, 2609
  vessel wall disorders in, 927
SDF (stromal cell–derived factor), 823, 1577, 2768t

SDGs (Sustainable Development Goals), 3844
SDH (succinate dehydrogenase), 3073
SDHA gene mutations, 3079t
SDHB gene mutations
  in Cowden's syndrome, 3080t, 3088
  in pheochromocytoma and paraganglioma, 3076f, 3077, 3079, 3079t
SDHC gene mutations, 3079t
SDHD gene mutations, 3078–3079, 3079t
Seabather's eruption, 3725, 3726f
Sea cucumber envenomation, 3724
Seadornavirus, 1662t, 1663
Seal bite-wound, 1143
Sea levels, rise in, 3747
Seal finger, 1143, 1145
Sealpox, 1518t
Sea snake envenomation, 3717
Seasonal affective disorder, 3665
Sea sponge, 3724
Sea urchins, 3724
Sebaceous adenoma, 404
Sebaceous carcinoma, 404
Sebaceous hyperplasia, 404
Sebastian syndrome, 924
Sebopsoriasis, 385
Seborrheic dermatitis, **385**
  clinical features of
    distribution and morphology of skin lesions, 378t, 380f, 385, 393t
    erythroderma, 393t
    facial, 386f, A5
    scalp, A5
  in HIV infection, 1605
  Malassezia infection in, 1723
  vs. pemphigus foliaceus, 410
  treatment of, 385–386, 393t
Seborrheic keratosis, 378t, 380f, 397–398, A5
SEC23B gene mutations, 3080t
Secnidazole, 1103
Secobarbital, 3713t
Secondary (specific) granule, 453
Secondary hyperkalemic paralysis, 359
Secondary prevention, 8
Secondary syphilis. *See* Syphilis, secondary
Secondary taste cortex, 239
Second disease. *See* Scarlet fever
Second-hand smoke. *See* Smoking, passive exposure to
Second-impact syndrome, 3575
Second opinions, 3898
"Second wind" phenomenon, 3364
Secretagogues, 2576
Secretin test, 2536, 2739t, 2742, 2752
Secretion defects, platelet, 926
Secretion systems, bacterial, 966, 966f, 968–969, 1295
Secretory diarrhea, 307–308, 308t, 352, 2543, 2548. *See also* Diarrhea
Secretory leukocyte protease inhibitor (SLPI), 1565
Secretory protein, 2779
Secukinumab, 387t, 2793t, 2887t, 2890
Sedation, 84, 2295
Sedative-hypnotics, 3705, 3713t. *See also specific drugs*
Sedoreoviridae
  clinical syndromes of, 1666t
  geographic distribution of, 1664–1665t
  reservoirs and vectors of, 1658t, 1662t
  structure of, 1478f, 1480t, 1663
  transmission of, 1666t
Segawa syndrome (dopa-responsive dystonia), 3368t, 3497, 3511
Seizure(s), **3408**. *See also* Epilepsy
  absence, 3410, 3418f
  after traumatic brain injury, 3411, 3574
  in alcohol withdrawal, 3679
  approach to the patient, 3415, 3415f
  atonic, 3410
  in brain abscess, 1136
  in brain tumor, 716–717
  in cancer patient, 591, 726, 726t
  chemotherapy-induced, 726t
  classification of, 3409–3410, 3409t, 3418f
  coma in, 189
  continuous. *See* Status epilepticus
  definition of, 3408
  delirium and, 185, 3484
  differential diagnosis of, 3418–3419, 3419t
  drug-induced, 3414t
  eclamptic, 3899
  EEG in, 3417–3418, 3417f, 3418f
  in encephalitis, 1111
  etiology of, 3411, 3413–3414, 3413t, 3414t
  evolution of, 3409–3410
  febrile, 137, 3411, 3413
  focal, 3409, 3456–3457
  focal sensory, 176
  generalized, 3410
  genetic testing in, 3418
  in heatstroke, 3775
  in hemophagocytic lymphohistiocytosis, 464
  history and examination in, 3415–3417
  in HIV infection, 1608, 1608t
  in hyperperfusion disorders, 2350
  in hypocalcemia, 365
  in hyponatremia, 349
  imaging in, 3418
  laboratory studies in, 3417
  lactic acidosis in, 369
  magnetoencephalography in, 3418
  in malaria, 1763
  mechanisms of, 3414–3415
  in meningitis, 1120
  myoclonic, 3410
  in neurocysticercosis, 1833, 1834
  in overdosage/poisoning, 3705, 3708
  in parasitic infections, 1738t
  in pregnancy, 3899
  psychogenic, 3419
  pyridoxine-dependent, 3369t
  in schistosomiasis, 1829
  sudden cardiac death risk and, 2333
  vs. syncope, 160, 3418–3419, 3419t
  temporal lobe, 208
  tonic-clonic, 3410
  treatment of, 716–717, 3419. *See also* Epilepsy, treatment of
Selectins, 454, 457f, 914
Selection bias, 515
Selective androgen receptor modulators (SARMs), 3120–3121
Selective estrogen response modulators (SERMs), 526f, 536, 2981, 3303
Selective IgG2 deficiency, 2802
Selective serotonin reuptake inhibitors (SSRIs)
  adverse effects of, 3659t, 3666–3667
    osteoporosis, 3296
    peptic ulcer disease, 2518
    SA node dysfunction, 1920t
    serotonin syndrome. *See* Serotonin syndrome
    sexual dysfunction, 3157, 3667
    SIAD, 348
  in dementia, 199
  for depression, 3659t, 3666–3667, 3666f
  for depression in AD, 3482
  for depression in terminally ill patient, 85
  drug interactions of, 492, 3660t, 3667
  for functional dyspepsia, 302
  genetic variations in response to, 492
  for IBS, 2575, 2577t
  for inappropriate sinus tachycardia, 1936
  overdosage/poisoning with, 3715t
  for pain, 100
  for panic disorder, 3658
  pharmacology of, 3667
  for premature ejaculation, 3155
  for PTSD, 3663
  for vertigo, 165, 165t
Selegiline
  adverse effects of, 3659t
  for depression, 3659t
  for hypertensive crisis prophylaxis, 3667
  overdosage/poisoning with, 3710t
  for PD, 178, 3503, 3504, 3504t
Selenium, 513, 2601t, 2614t, **2615**
Selenium deficiency, 2018, 2614t, 2615
Selenium exposure, 3702
Selenium sulfide, 388t, 389, 395t, 398t
Selenocysteine, 2615
Selenosis, 2615
Selexipag, 2194, 2197t
Self-neglect, 3895t
Self-renewal, in stem cells, 761
Selgantolimod, 2681
Seligman's disease (alpha heavy chain disease), 894
Selinexor, 534t, 570t, 573, 890t, 892
Sellar chordomas, 3003
Sellar mass lesions, **2998**
  clinical features of, 222, 2998, 2999t
  histologic investigation of, 2999
  hypogonadotropic hypogonadism and, 3114
  hypothalamic. *See* Hypothalamic disease/masses
  laboratory investigation of, 2999, 3001t
  mass effects of, 2998, 3000f
  MRI in, 2998–2999, 3000f
  ophthalmologic evaluation in, 2999
  pituitary. *See* Pituitary tumors (adenomas)
Selpercatinib, 533t, 569t, 570, 628, 3085
Selumetinib, 722, 753t
SEMA3A gene mutations, 3112, 3112t
Semaglutide
  actions of, 3192t, 3193
  adverse effects of, 3192t, 3193, 3211t
  contraindications to, 3192t, 3211t
  for diabetes mellitus, 3193, 3211t, 3213
  for heart failure, 1988
  for metabolic syndrome, 3257
  for NASH, 2706
  for obesity, 3192f, 3192t, 3193
Semantic dementia, 207
Semantic paraphasia, 201
Semantic primary aphasia, 204
Semen analysis, 3108, 3150
Seminiferous tubules, 3106–3107
Seminoma. *See also* Testicular cancer
  pathology of, 505
  risk classification of, 706, 708t
  staging of, 706, 707–708f
  treatment of, 706, 707–708f, 2271
Semliki Forest virus, 1657t
Senataxin, 3521, 3522t
Sendai virus, 3402
Senescence. *See also* Aging
  carcinogenesis and, 529t, 531
  cellular, 3871
  definition of, 539
  negligible, 3869, 3871f
Senescence-associated secretory phenotype (SASP), 3871
Senile purpura, 472, 927
Senior-Loken syndrome, 2430t, 2435
Senna, 81, 82t, 307, 312, 509
Sennetsu fever, 1460
Senning (atrial switch) procedure, 2074–2075, 2075f, 2075t

- Sensation
  - abnormalities of, **172**
    - in brainstem lesions, 176
    - in cortical lesions, 176
    - falls and, 180–181
    - in focal seizures, 176
    - localization of, 175–176
    - in nerve and root lesions, 175–176
    - psychogenic causes of, 176
    - in spinal cord lesions, 176
    - in thalamic lesions, 176
    - treatment of, 176
  - anatomy of, 172–173, 173f
  - examination of, **173**
    - cortical, 173–175
    - primary, 173, 174–175f, 174t
    - quantitative sensory, 175
  - positive and negative symptoms, 172
  - terminology in, 172
- Sensitivity, 23–24, 514, 514t, S10
- Sensitization, 93, 2804
- Sensory ataxia, 172, 177t, 179, 179t
- Sensory balance training, 181
- Sensory dissociation, 176
- Sensory examination, 3381–3382
- Sensory neuronopathy, 750
- Sensory neuropathy. *See* Peripheral neuropathy
- Sentinel bleeds, subarachnoid, 3459
- Sentinel lymph node biopsy, 603, 636
- Seoul virus, 1661t, 1676
- Sepik virus, 1660t
- Sepsis/septic shock
  - anaerobic bacteremia and, 1377
  - without clear focus, 990, 991t
  - clinical features of, 990, 991t, 993, 2320
  - clostridial, 1239t, 1243
  - complications of
    - acute lung injury, 2323
    - AKI, 2323
    - cardiovascular, 2322–2323
    - encephalopathy, 2348
    - gastrointestinal, 2323
    - hematologic, 2323–2324
    - hypoglycemia, 3234
    - hypothermia, 3769
    - hypovolemia, 346
    - intracerebral hemorrhage, 3455
    - long-term, 2324
    - neurologic, 184, 2321–2322
    - renal, 2375–2376, 2379t
  - in critically ill patient, 2296
  - epidemiology of, 2316, 2316f
  - etiology of, 990, 991t, 2316–2317
  - hemodynamic patterns in, 2329t
  - in hemophagocytic lymphohistiocytosis, 468
  - hospital-onset, 2316–2317
  - outcomes of, 2324
  - pathogenesis of, 2317–2320, 2318f
  - in pneumococcal infections, 1194
  - puerperal, 1211, 1262
  - recognition of, 2320
  - risk and prognostic factors, 2317
  - with skin findings, 990–993, 991t
  - with soft tissue findings, 991t, 993
  - treatment of, 991t
    - antibiotics, 963t, 2315
    - empirical, 991t, 2321t
    - host-targeted, 2324
    - initial, 2320–2321
    - vasopressin, 2315
- *SEPT9* gene, 517
- Septal ablation, A11, A11v
- Septal panniculitis. *See* Erythema nodosum
- Septal reduction therapy, 2007
- Septic arthritis. *See* Infectious (septic) arthritis
- Septic diskitis. *See* Osteomyelitis, vertebral
- Septic emboli, 1038, 1039f, A1, A5
- Septicemia
  - hemolysis in, 801t
  - meningococcal. *See* Meningococcemia
  - in plague, 1344
  - respiratory alkalosis in, 376
  - in yersiniosis, 1349
- Septic thrombosis, 1142
- Septic thrombosis, intracranial, 994
- Seqirus, 3723, 3726
- Seralutinib, 2187f
- SERCA2a, 2000
- Serelaxin, 1991, 1993t
- Serine kinase receptors, 2981t, 2983f, 2984
- Serine metabolism disorders, 3368t
- Serine protease autotransporters (SPATEs), 974
- SERMs (selective estrogen response modulators), 526f, 536, 2981, 3303
- Seronegative autoimmune autonomic neuropathy, 3550
- Serotonin (5-HT)
  - in carcinoid syndrome, 682, 682f
  - in IBS, 2573
  - in migraine, 3462
  - pain and, 94f
- Serotonin norepinephrine reuptake inhibitors (SNRIs), 101, 199, 3588, 3659t
- Serotonin ($5\text{-HT}_{1B/1D}$) receptor agonists (triptans), 300, 3462, 3467t, 3468
- Serotonin (5-HT_{1F}) receptor agonists (ditans), 3467t, 3469
- Serotonin (5-HT_{4}) receptor agonists, 299t, 300, 2575, 3462
- Serotonin (5-HT_{3}) receptor antagonists
  - adverse effects of, 299
  - for IBS, 2575
  - for nausea and vomiting, 83, 116, 298, 299t, 508–509, 575
- Serotonin release assay, 948
- Serotonin syndrome
  - clinical features of, 3666–3667, 3715t
  - etiology of, 3517, 3715t
  - mechanism of, 3715t
  - with SSRIs, 3666–3667
  - treatment of, 3715t
- *SERPINA1,* 2253, 3993–3994
- *SERPINF1*/PEDF gene mutations, 3324, 3325t, 3326
- *SERPINH1* gene mutations, 3324, 3325t
- Serra Norte virus, 1662t
- *Serratia* spp. infections, **1292**
  - antibiotic resistance in, 1293
  - clinical features of, 408, 1292
  - diagnosis of, 1292
  - epidemiology of, 1281, 1292
  - treatment of, 1293
- Sertoli-Leydig tumor, 712
- Sertraline
  - adverse effects of, 3659t, 3667
  - for depression, 85, 504, 3588, 3659t, 3666–3667
  - for menopausal symptoms, 3143
  - for obsessive-compulsive disorder, 3664
  - for panic disorder, 3658
  - for phobic disorders, 3661
  - for PTSD, 3663
- Serum albumin. *See* Albumin, serum
- Serum ascites-albumin gradient (SAAG), 328
- Serum bilirubin. *See* Bilirubin, serum
- Serum-binding proteins, 2982–2983
- Serum creatinine. *See* Creatinine, serum
- Serum ferritin. *See* Ferritin, serum
- Serum proteins, electrophoresis of, 883, 885f
- Serum sickness
  - with antivenom, 3721, 3723
  - with ATG treatment, 813
  - clinical features of, 419, 427t, 2907, 3719
  - immunologic mechanisms of, 2893
  - pathogenesis of, 417t, 418f, 419, 422, 489
  - treatment of, 813, 3719
  - urticaria/angioedema in, 403, 422, 2808
  - urticaria-like eruptions in, 143t, 147–148
- Serum sickness–like illness, 403, 419, 427t, 2659, 2907
- Setmelanotide, 3184, 3192
- “Setting sun” sign, 236
- Sevelamer, 2391, 3266
- Seven transmembrane helix family, 2773. *See also* G protein–coupled receptors (GPCRs)
- Severe acute respiratory syndrome (SARS)
  - clinical features of, 995, 1535
  - as emerging disease, 3858
  - epidemiology of, 995
  - etiology of, 995, 1535
  - imaging in, A12
- Severe acute respiratory syndrome coronavirus 2. *See* SARS-CoV-2
- Severe combined immunodeficiency (SCID)
  - clinical features of, 2798, S9
  - diagnosis of, 2798
  - etiology of, 2798–2799
  - genetic considerations in, 2773
  - HCT in, 915, 918, 919t
  - T cell differentiation and, 2798f
  - treatment of, 3824t, 3825–3826
- Severe congenital neutropenia (SCN), 457, 2794–2795, 2795f, S9
- Severe fever with thrombocytopenia syndrome, 1657t, 1677–1678
- Severe sepsis, 2316
- Severity-of-illness (SOI) scoring systems, 2291, 2291t, 2292t
- Sex chromosome(s), 3778, 3789
- Sex chromosome disorders, 3095t
  - 46,XX DSD. *See* 46,XX disorders of sex development
  - 46,XY DSD. *See* 46,XY disorders of sex development
  - Klinefelter’s syndrome. *See* Klinefelter’s syndrome
  - noninvasive prenatal testing for, 3988
  - ovotesticular DSD, 3097t, 3098
  - Turner’s syndrome. *See* Turner’s syndrome
  - 45,X/46,XY mosaicism (mixed gonadal dysgenesis), 3097t, 3098
- Sex cord tumors, ovarian, 712
- Sex development, 3094–3095, 3096f
- Sex development disorders. *See* Disorders of sex development (DSDs)
- Sex differences, 3162, 3168–3169. *See also* Men’s health; Women’s health
- Sex headache, primary, 3474
- Sex hormone-binding globulin (SHBG), 3105, 3105f, 3128, 3139, 3160
- Sex-influenced trait, 3789
- Sex-limited trait, 3789
- Sex reversal, 3783t, 3786, 3789
- Sex therapy, 3160
- Sexual abuse, 332, 1257, 1258, 3157, 3895t
- Sexual assault, 3168, S8
- Sexual dysfunction, 3155
  - alcohol use and, 3677
  - female, 3160–3162, 3160t, 3227
  - male, 3172–3173. *See also* Erectile dysfunction (ED)
  - in MS, 3578, 3588
  - SSRI-related, 3667
- Sexual interest arousal disorder, female, 3161
- Sexually transmitted infections (STIs), **1096**. *See also specific infections*
  - cervical cancer and, 713
  - in children, 1257
  - circumcision and reduced risk of, 1563
  - classification of, 1096–1097, 1096t

Sexually transmitted infections (STIs) (*Cont.*):
  epidemiology of, 1096–1097, 1096t
  etiology of, 1097t
  global considerations in, 1097
  HIV transmission and, 1563
  in LGBTQIA+ people, 3178
  in older adults, 3878–3879
  partner notification in, 1110–1111
  prevention and control of, 1109–1110, 1110f, 3177
  prevention of, 584
  risk assessment for, 1097–1098, 1098t, 3177
  screening for, 1110–1111, 3177
  in travelers, 1019
  UTIs and, 1090–1091
Sexual orientation, 3178, 3180t
Sexual pain disorder, 3161
Sexual response, physiology of, 3155, 3156f, 3161
Sézary's syndrome, 394, 436, 866, A6
SF1 (steroidogenic factor 1), 3050
SF1 gene mutations, 3094, 3099t, 3101t, 3112t
SF3B1 gene mutations, 815t, 816, 828, 850, 851t
SGCE gene mutations, 3510t, 3511
S genes, of HBV, 2644, 2644f
SGLT2 (sodium-glucose co-transporter 2), 1981
SGLT2 inhibitors. *See* Sodium-glucose co-transporter 2 (SGLT2) inhibitors
SH2B1 gene mutations, 3185, 3185t
Shadow plaques, 3581
Shah-Waardenburg syndrome, 397, 3783t
Shaking rigors, 3771
Shapiro's syndrome, 3768
Shared epitope, 2842
Shark bite-wound, 1145t
Shawl sign, 2911, 2911f
SHBG (sex hormone-binding globulin), 3105, 3105f, 3128, 3139, 3160
Sheehan's syndrome, 2993, 3067
Shellfish poisoning, 993, 3414, 3727–3728
Shelterin, 3819f
Shewanella spp. infections, 1268–1269, 1268t
Shift work, 3944t
Shift-work sleep disorder, 218–219, 3948
Shiga toxin–producing *E. coli* (STEC), 801, 1077, 1081, 1285t, 1287–1288. *See also* *Escherichia coli* infections, intestinal
Shigella spp.
  antibiotic resistance in, 1324
  characteristics of, 965, 966, 1322
  genomic signatures of, 1322
  infectivity of, 1322–1323, 1323f
  laboratory identification of, 1324, S11
  mTOR reactivation by, 973
  NF-κβ inhibition by, 971
  niche, 965, 965t
  survival in cytosol, 968
  survival in vacuole, 968
Shigella spp. infections, **1322**
  clinical features of, 306t, 1060, 1080t, 1323–1324
  complications of, 305, 1325
  diagnosis of, 1080, 1324, S11
  epidemiology of, 1322
  global considerations in, 1322
  in HIV infection, 1598
  vs. IBD, 1324
  pathogenesis of, 1322–1323, 1323f
  prevention of, 1325
  treatment of, 1324–1325, 1324t
Shingles. *See* Herpes zoster (shingles)
Shingrix, 1507
Shivering, 134, 3768
Shock
  in acute pancreatitis, 2745
  approach to the patient, 2292–2294, 2293f, **2309**
  cardiogenic. *See* Cardiogenic shock
  classification of, 2310–2311, 2310t
  cyanosis in, 278
  definition of, 2309
  distributive, 2310–2311, 2311t
  ECG in, 2313
  effect on drug concentration and response, 488
  history in, 2312
  hypotension in, 2292–2293
  hypovolemic. *See* Hypovolemic shock
  hypoxia in, 277
  laboratory evaluation of, 2313, 2313t
  lactic acidosis in, 369
  mixed, 2311
  obstructive, 2311, 2311t
  pathophysiology of, 1248, 2310
  physical examination in, 2312–2313
  point-of-care ultrasound in, 2313–2314
  respiratory failure in, 2294–2295
  septic. *See* Sepsis/septic shock
  stages of, 2311
  treatment of
    antibiotics, 2315
    initial, 2314
    mechanical ventilation, 2294, 2294f
    oxygenation and ventilation support, 2315
    principles of, 2312, 2312t
    vasopressor and inotropic support, 2315
    volume resuscitation, 2314–2315
  undifferentiated, 2311
Shock index (SI), 2313
Shockwave intravascular lithotripsy, A11, A11v
Shohl's (sodium citrate) solution, 369, 372, S1
Shokwe virus, 1659t
Shone's complex, 2036
Shopping cart sign, 120
Short-bowel syndrome, 369, **2547,** 2611
Short-chain fatty acids, 2541, 2543, 2544, 2549t, 3840
Short QT syndrome, 1970
Short stature, **2994**
Short tandem repeats, 3793
Shotgun sequencing, 979
Shoulder
  anatomy of, 2973f
  bursitis of, 2973
  "frozen," 2974–2975
  pain in, 2943–2944, 2943f
  tendinitis of, 2973–2974
Shoulder pain, 133
Shprintzen-Golberg syndrome, 3331
Shuni virus, 1659t
Shunt
  intracardiac, 1907
  pulmonary, 277, 278, 2204, 2205f
Shwachman-Diamond syndrome, 457, 810
Shy-Drager syndrome. *See* Multiple system atrophy (MSA)
SI (shock index), 2313
SIAD. *See* Syndrome of inappropriate antidiuresis (SIAD)
Sialadenitis, 242, 266
Sialidosis, 3353t
Sialolithiasis, 266, A3
Siberian tick typhus, 138t
Sib pair analysis, 3797t
Sick euthyroid syndrome, 3024, **3039,** 3771
Sickle cell anemia, 772t
Sickle cell crisis
  clinical features of, 773, 2967
  hyperbilirubinemia in, 323
  splenic infarction in, 479
  treatment of, 773
Sickle cell disease, **772**
  clinical features of
    hand-foot syndrome, 2967
    musculoskeletal, 2967–2968, 2967t
    nephropathy, 2415t, 2427, 2442, 2447
    priapism, 3155
    pulmonary hypertension, 2193
    splenic abscess, 1075
    stroke, 3447
    testicular atrophy, 3115
  complications of, **773,** 774t
    acute chest syndrome, 775
    acute painful episodes, 773. *See also* Sickle cell crisis
    autosplenectomy, 482
    cholestasis, 325
    folate deficiency, 788
    jaundice, 323
    leg ulcers, 775
    osteonecrosis, 775
    stroke, 774t
  diagnosis of, 772–773, 772t, 774f, A16
  genetic considerations in, 3787, 3790, 3957, 3958f
  genetic testing for, 3796
  genotypes of, 772, 772t
  high-altitude travel and, 3759
  malaria resistance and, 1763, 3795
  origin, spread, and epidemiology of, 772
  pathophysiology of, 772, 772t, 773f
  peripheral blood smear in, 439f
  screening, 776
  systems biology applied to, 3957, 3958–3959f, 3958f
  transfusion reactions in, 911
  treatment of
    crizanlizumab, 775
    gene therapy, 762, 776, 2447, 3824t, 3826
    L-glutamine, 775
    HCT, 776, 915, 919
    hydroxyurea, 775
    transfusion, 776
    voxelotor, 775
Sickle cells, 439f, 447, 447t
Sickle cell trait (HbAS), 772t, 775
Sickle thalassemia syndrome, 772t
Sick sinus syndrome, 1918, A8
Sideroblastic anemia, X-linked, 3341f, 3341t, 3349
Sideroblasts, ringed, 814f, A6
Sievert (Sv), S6
Sigmoidoscopy
  for colorectal cancer screening, 39t, 40, 516t, 517, 655–656
  in constipation, 312
  in diarrhea, 307, 311
  in gastrointestinal bleeding, 320
  indications for, 2465
  in ulcerative colitis, 2557
Sigmoid resection, 2581, 2582f
Sigmoid volvulus, 2487, 2493f, 2590, 2593
Signal recognition protein (SRP54), 2790t
Signal transduction pathways
  in cancer cells, 526, 526f, 526t, 531–537, 540f
  membrane receptor families and, 2981t
  in psychiatric disorders, 3652–3653
"Signet-ring" sign, 2243
Sign of Leser-Trelat, 398
Sildenafil
  actions of, 3155, 3156f
  adverse effects of, 3159, 3159t
  contraindications to, 3159, 3159t
  development of, 488
  drug interactions of, 704
  for erectile dysfunction, 503t, 3159, 3159t
  for heart failure, 1988
  interaction with nitrates, 2109, 2118
  for lower urinary tract symptoms, 704
  pharmacology of, 3159t
  for pulmonary arterial hypertension, 2194, 2197t, 2874
  for Raynaud's phenomenon, 2178, 2874
Silent mutation, 3786f

Silent nociceptors, 94
Silent (painless) thyroiditis, 3039–3040
Silibinin, 2668
Silica, 2236, 2236t
Silica exposure, 2420
Silicosis, 2236–2238, 2236t, 2237f
Siltuximab, 594, 2792t
Silver, skin deposits, 400
Silver sulfadiazine, S5
SIM1 gene mutations, 3183f, 3185, 3185t
Simeprevir, 2689
Simethicone, 302, 2575
Simian immunodeficiency virus (SIV), 1558f
Simple goiter, 3041
Simple systems, 3955
Simplicity heuristic, 21–22
Simultanagnosia, 205f, 206, 235
Simultaneous bilateral stimulation test, 205
Simvastatin. *See also* Statins
  adverse effects of, 923t, 2940t, 3244t
  drug interactions of, 1742t
  genetic variations in response to, 491, 493, E72t
  for lipoprotein disorders, 3244t
Sindbis virus, 1657t, 1670
Sinecatechins, 390, 1531
Single-balloon enteroscopy, 2468
Single-minded 1. *See SIM1* gene mutations
Single-nucleotide polymorphisms (SNPs)
  in *CFTR* gene, 3781f
  definition of, 976t
  drug response and, 491, E72t
  effects of, 3793
  frequency of, 3779
  haplotype, 3779, 3781f, 3787
  in undiagnosed disease, 3996t
Single photon emission computed tomography (SPECT). *See* SPECT
Single pulse transcranial magnetic stimulation, 3467t, 3468t, 3469, 3470t
Single-stranded DNA viruses, 1481
Single ventricle physiology, 2076, 2077f
Sin Nombre virus, 1661t, 1674
Sinoatrial (SA) conduction time (SACT), 1918
Sinoatrial (SA) nodal artery ischemia and infarction, 1919–1920
Sinoatrial (SA) node, 1868, 1868f, 1912, 1917–1918, 1918f
Sinoatrial (SA) node disease/dysfunction, **1918**
  diagnosis of, 1918, 1919f, 1921f
  reversible causes of, 1918, 1920, 1920t
  subtypes of
    carotid sinus hypersensitivity and neurally mediated bradycardia, 1920
    chronotropic incompetence, 1919
    ischemia and infarction, 1919–1920
    sinus node exit block, 1919, 1919f, 1922f
    sinus node fibrosis, 1919
    tachycardia-bradycardia syndrome, 160, 1919, 1922f
  treatment of, 1920–1924, 1923f, 1923t. *See also* Pacemaker
Sinus arrhythmia, 1935
Sinus bradycardia, 2124–2125, A7
Sinusitis
  Actinomyces, 1364
  acute, **255**
  allergic fungal, 1715
  anaerobic bacterial, 1374
  antibiotic prescribing criteria for, 256, 256t
  Aspergillus, 1712, 1715, A16
  in cancer patient, 581–582
  chronic, 256–257, 1374
  in cystic fibrosis, 2245
  differential diagnosis of, 256
  ear pain in, 253
  fungal, 582
  H. influenzae, 1262
  in HIV infection, 1595
  hospital-associated, 256
  maxillary, 265
  M. catarrhalis, 1265
  mucormycosis, A16
  olfactory dysfunction in, 240–241, 243
  pneumococcal, 1192
  recurrent, 256
  subdural empyema and, 1138–1139
Sinus node, 1935, 1935f
Sinus node fibrosis, 1919
Sinus node recovery time (SNRT), 1918
Sinusoidal obstructive syndrome, 916
Sinus pause, 1918, 1918f
Sinus rhythm, 1935, 1935f
Sinus tachycardia, **1936**
  ECG in, A7, A8
  epidemiology of, 1914
  evaluation and treatment of, 1936f
  inappropriate, 1932t, 1936–1937
  physiologic, 1936, 1936f, 1936t
  in pulmonary edema, 2331
  in pulmonary embolism, 107
  in thyrotoxicosis, 3033
Siponimod, 3586t, 3587
Sipuleucel-T, 551, 559f, 561, 703
Sirolimus (rapamycin)
  actions of, 526f, 574, 2408t
  adverse effects of, 2408, 2408t, 2410, 2445, 2722, 3247
  for Castleman's disease, 34
  in drug-eluting stents, 2128, 2129f
  drug interactions of, 1742t
  for immunosuppression, 2029t, 2408, 2408t, 2722
  for lifespan enhancement, 3875
  for polycystic kidney disease, 2433
  for reducing skin cancer risk in transplant recipients, 433
  structure of, 3874f
  for subependymal giant cell astrocytomas, 723
Sirolimus protein bound particles, 571t
SIRS (systemic inflammatory response syndrome), 2316, 2325f
SIRT1 agonist. *See* Resveratrol
Sirtuins, 3870, 3871, 3873f
Sister Mary Joseph's nodule, 323, 646, 673
Sister of p-glycoprotein (bile salt excretory protein), 2642
Sitosterolemia, 3241t, 3243–3244
Sit-to-stand test, 2619t
SIV (simian immunodeficiency virus), 1558f
6-meter walk test, 2619t
Sixth disease. *See* Exanthem subitum (roseola)
Sixth nerve palsy, 3458
Sjögren's disease, **2876**
  autoantibodies in, 2825t
  autoimmune diseases associated with, 2876
  clinical features of, 2876–2877, 2877f, 2877t
    cutaneous, 403, 408
    extraglandular, 2877t
    lymphoma, 2877, 2878f
    myelopathy, 3566
    nephritis, 2356
    neuropathy, 3598, 3605
    ocular, 224, 2877
    renal, 2438–2439, S1
    trigeminal neuropathy, 3553
    vasculitis, 2877
    xerostomia, 266, 2876–2877
  definition of, 2876
  diagnosis of, 2877t, 2878, 2878t, 2879f
  differential diagnosis of, 2878, 2878t
  epidemiology of, 2876
  in HCV infection, 2685
  in HIV infection, 1603
  pathogenesis of, 2790t, 2876
  in rheumatoid arthritis, 2840
  vs. rheumatoid arthritis, 2847
  treatment of, 2878–2879
SJS. *See* Stevens-Johnson syndrome (SJS)
Skeletal dysplasias, **3330**
Skeletal maturation, somatic growth and, 2994
Skew deviation, 236
Skin
  actinically damaged, 396
  age-related changes in, 431
  in cardiovascular system examination, 1859–1860
  color of, 430
  complete examination of, 517t, 518
  in cyanosis, 278
  disorders/lesions. *See* Skin disorders/lesions
  Fitzpatrick classification of, 430, 430t
  microbiota of, 3833, 3833f
  structural components of, 1049, 1049f
  ultraviolet radiation effects on, 429–430. *See also* Sun exposure
  yellowing, 320. *See also* Jaundice
Skin biopsy
  in autonomic dysfunction, 3542, 3546
  in *Leishmania* spp. infections, S12
  in leprosy, 1409
  in peripheral neuropathy, 3598
  technique for, 381
  in vasculitic syndromes, A14
Skin blister assay, in vivo, 463
Skin cancer, **598**
  in CLL, 853
  genetic considerations in, 521t
  global considerations in, 432
  melanoma. *See* Melanoma
  nonmelanoma, 431–432, **606,** 609–610, A5
    basal cell carcinoma. *See* Basal cell carcinoma
    extramammary Paget's disease, 609–610
    Kaposi's sarcoma. *See* Kaposi's sarcoma
    Merkel cell carcinoma. *See* Merkel cell carcinoma
    pathophysiology and etiology of, 606
    risk factors for, 432
    squamous cell carcinoma. *See* Squamous cell carcinoma, skin
    in transplant recipient, 433, 2410
  paraneoplastic syndromes in, 737t
  prevention of, 609
  screening for, 517t, 518
  self-examination for, 512
  sun/UV exposure and, 431–432, 512, 512t, 606
  UV light therapy and, 387
Skin disorders/lesions. *See also* Rash; *specific diseases*
  approach to the patient, **377**
  burrowing, 1051
  cancer. *See* Skin cancer
  crusts, 1049f, 1050t, 1051
  descriptions of, 377t
  diagnosis of, 381–382
  distribution of, 378t, 379, 380f
  drug-induced. *See* Cutaneous drug reactions; *specific drugs*
  eschars, 1050t, 1051
  esophageal manifestations of, 2514
  folliculitis. *See* Folliculitis
  history in, 381
  immunologically mediated, **409, A5**
  induration, 2859, 2860t
  infections, **1049**. *See also specific infections*
    anaerobic bacterial, 1377
    in cancer patient, 583
    diagnosis of, 1053, 3913, 3914v
    parasitic, 1738t
    pathophysiology of, 1049, 1049f
    treatment of, 964t, 1054, 1054t

- Skin disorders/lesions (*Cont.*):
  - light-induced. *See* Photosensitivity diseases
  - as manifestations of internal/systemic disease, **392, A5**. *See also specific diseases*
    - acneiform eruptions, 395
    - alopecia, 394, 395t. *See also* Alopecia
    - erythroderma, 392–394, 393t
    - exanthems, 401, 402t. *See also* Rash
    - figurate lesions, 394, 395t
    - hyperpigmentation, 397–400, 399t
    - hypopigmentation, 396–397, 397t, 398t
    - papulosquamous lesions, 392, 392t
    - purpura, 137, 406–408
    - pustular lesions, 395–396
    - telangiectasias, 396, 396t
    - urticaria, 402–403
  - morphology of, 378t
  - nodules, 1050t, 1051
  - papules, 1050t, 1051
  - papulonodular lesions. *See* Papulonodular lesions
  - physical examination in, 379
  - purpura. *See* Purpura
  - shape of, 379
  - telangiectasias. *See* Telangiectasia(s)
  - ulcers. *See* Cutaneous ulcers
  - vesicles/bullae. *See* Vesicles/bullae
- Skin snip, S12
- Skin tag (acrochordon), 379t, 380f, 398, 2563
- Skull fractures, 3571–3572
- SLAMF7, 2792t
- Slapped-check disease. *See* Erythema infectiosum
- *SLC01B1* gene, 491, 493, E72t
- *SLC2A1* gene mutations, 794t, 795, 3412t, 3420, 3517
- *SLC3A1* gene mutations, 2369
- *SLC4A1* gene mutations, 358, 794t, 795
- *SLC5A1* gene mutations, 2545
- *SLC7A7* gene, 3373
- *SLC7A9* gene mutations, 2369
- *SLC12A3* gene mutations, 2369
- *SLC19A2* gene mutations, 790
- *SLC22A5* gene mutations, 2554t
- *SLC25A13* gene, 3373
- *SLC35C1* gene mutations, 460
- *SLC39A13* gene mutations, 3329t, 3330
- *SLC39A19* gene mutations, 3375
- SLC/TCA-4/Exodus 2, 2767t
- SLE. *See* Systemic lupus erythematosus (SLE)
- SLED (slow low-efficiency dialysis), 2385
- Sleep, **204**
  - age-related changes in, 209–210, 211f
  - AV conduction block during, 1927, 1927f
  - behavioral changes in, 212
  - insufficient, 213
  - neuroanatomy of, 210–211
  - neurochemistry of, 210–211
  - NREM, 210, 210f
  - optimal amounts of, 11
  - organization of, 210, 211f
  - physiology of, 209–212
  - REM, 210, 211f
  - states and stages of, 210, 211
- Sleep apnea, **213**. *See also* Obstructive sleep apnea/hypopnea syndrome (OSAHS)
  - in acromegaly, 3008
  - central, 213, 1982, 1986, 2280f, 2282
  - complex, 2282
  - daytime sleepiness and, 213
  - heart failure and, 1986, 1997–1998
  - mixed, 213–214
  - SA node dysfunction in, 1918
- Sleep bruxism, 217–218
- Sleep deprivation, 210, 2987, 3188, 3952
- Sleep disorder(s), **212**
  - approach to the patient, 212
  - circadian rhythm, 218–219
  - costs of, 209
  - daytime sleepiness in, 212–213, 213t
  - driving rights and, 212
  - epidemiology of, 12, 209
  - fatigue and, 166
  - in fibromyalgia, 2963
  - at high altitudes, 3758
  - insomnia. *See* Insomnia
  - narcolepsy. *See* Narcolepsy
  - in older adults, 3893, 3894t
  - parasomnias, 217–218
  - in PD, 3506–3507
  - periodic limb movement disorder, 217
  - restless legs syndrome, 217
  - screening for, 12
- Sleep enuresis, 217
- Sleep hygiene, 11–12, 84, 216, 216t, 3894t
- Sleepiness, vs. fatigue, 213
- Sleeping sickness. *See* Human African trypanosomiasis
- Sleep paralysis, 213t, 214
- Sleep terrors, 217
- Sleep-wake cycles, 210–211f, 211, 3946
- Sleepwalking (somnambulism), 217
- Sliding filament model, for muscle contraction, 1847–1849
- “Slip” (error), 42
- “Slipping clutch” syndrome, 178
- Slipping rib syndrome, 2971
- Slit-skin smear, 1409
- Slow acetylators, 493
- Slow low-efficiency dialysis (SLED), 2385
- SLPI (secretory leukocyte protease inhibitor), 1565
- SLR (straight leg raising), 121t
- Sly syndrome, 3357
- SMA (spinal muscular atrophy), 3522t, 3524, 3824t, 3828
- SMAC, 539, 540f
- *SMAD2* gene mutations, 3331, 3332t
- *SMAD3* gene mutations, 2166, 3331, 3332t
- *SMAD4* gene mutations, 520f, 524t, 667
- Smads, 2984
- SM-AHNMD (systemic mastocytosis associated with clonal hematologic non–mast cell lineage disease), 2820
- Small airways disease. *See* Chronic obstructive pulmonary disease (COPD)
- Small-bowel bacterial overgrowth, 300, 309, 311, **2543,** 2870, 2874
- Small-bowel ultrasound, 2559
- Small-cell carcinoma, bladder, 691, 692f. *See also* Bladder cancer
- Small-cell lung cancer. *See* Lung cancer, small-cell
- Small intestine
  - adenocarcinoma of, 651
  - benign neoplasms of, 651
  - bleeding from, 318
  - blood supply to, 2587, 2587f
  - double-balloon enteroscopy of, 2468, 2472f
  - endoscopy of, 2465, 2468. *See also* Capsule endoscopy
  - fluid absorption and secretion in, 303, 304f
  - functions of, 2459
  - GISTs of, 651
  - iron absorption in, 763, 763f
  - ischemia of. *See* Intestinal ischemia
  - motility of, 303, 303t
  - mucosal biopsy of, 2547f, 2550–2551, 2550t
  - neuroendocrine tumors of, 652, 680, 680f, 681f. *See also* Neuroendocrine tumors (NETs)
  - nutrient digestion and absorption in, 2540, 2543–2544
  - obstruction of, 110, 297, 2589, 3914v. *See also* Intestinal obstruction
  - resection of, 2613
  - in systemic sclerosis, 2870, 2870t
- Small-molecule inhibitors, 52, 540f
- Small noncleaved cell lymphoma. *See* Burkitt’s lymphoma
- Smallpox
  - in bioterrorism, 1149, S4
  - clinical features of, 142t, 1519, S4
  - diagnosis of, S4
  - differential diagnosis of, 147, 1506
  - eradication of, 1002
  - hemorrhagic, S4
  - incidence pre- and postvaccine, U.S., 1002t
  - malignant, S4
  - mortality rates of, 1521
  - prevention of, S4
  - skin manifestations of, 142t, 147, 1049, A1, S4
  - treatment of, 1489, 1521, S4
- Smallpox vaccine, 1149, 1521, 1522, 2015, S4
- Small-vessel stroke, 3446, 3446f. *See also* Ischemic stroke
- *SMARCB1*-deficient renal medullary carcinoma, 687–688, 688t
- *SMARCD2* gene mutations, 460
- *SMARC* gene mutations, 3940
- “Smart pumps,” 43
- *SMCHD1* gene mutations, 3643
- ^{153}Sm-EDTMP, 703
- Smell. *See* Olfactory disorders; Olfactory system
- Smith-Lemli-Opitz syndrome, 3067t, 3099t
- *SMN1* gene mutations, 3828
- *SMO* gene mutations, 606
- Smoke inhalation, 2239
- Smokeless tobacco, 261, 264t, 511, 610
- Smoking. *See also* Smoking cessation
  - asthma and, 2220
  - autoimmune disease and, 2823
  - cancer and, 512t, 3683, 3683t
    - bladder, 691
    - cervical, 713
    - colorectal, 655
    - esophageal, 642
    - HCC, 660
    - head and neck, 610
    - lung, 615
    - oral, 261
    - pancreatic, 671
    - RCC, 687
    - skin, 606
  - cardiovascular disease and, 3683, 3683t
    - global considerations in, 1857
    - ischemic heart disease, 2099
    - peripheral artery disease, 2173
    - stroke risk, 3446t
  - deaths related to, 3683
  - drug interactions with, 3684
  - epidemiology of, 3682, 3682f
  - erectile dysfunction and, 3155
  - global considerations in, 3851
  - IBD and, 2552
  - infertility and, 3151
  - interstitial lung disease and, 2260t, 2261, 2263–2264
  - in Leber’s hereditary optic neuropathy, 3816
  - leukoplakia and, 264t
  - of lower tar and nicotine cigarettes, 3684
  - meningococcal infections and, 1247
  - MS risk and, 3581
  - neutrophil counts in, 458
  - osteoporosis and, 3296
  - pancreatitis and, 2750
  - passive exposure to
    - in children, 2220, 3684
    - COPD and, 2253

    lung cancer and, 511, 615, 3684
    respiratory disease and, 2200, 2241
  peptic ulcer disease and, 2518, 2521
  polycythemia and, 452
  in pregnancy, 3684
  prevalence of, 3683
  prevention of, 3685
  respiratory disease and, 1536, 2200, 3683–3684, 3683t
    COPD, 2250–2251, 2252, 2252f
    Legionella spp. infections, 1271
  rheumatoid arthritis and, 2843
  second malignancy and, 512
  toxic dust exposure and, 2238
  in women, 3168
Smoking cessation
  after lung cancer diagnosis, 615
  complementary and integrative therapies for, 3927
  in COPD, 2251, 2255
  global considerations in, 1857, 3851
  in ischemic heart disease treatment, 2099
  life expectancy increase following, 38t
  in peripheral artery disease, 2173
  pharmacologic therapy for, 615, 2255, 3685, 3685t
  physician interventions in, 3685, 3685t
  preoperative, 3909t
  in preventive care, 13, 511
  stages of, 511
  success rates of, 511, 3685
  in thromboangiitis obliterans, 2174
Smoothened gene, 431
Smooth muscle, vascular, 1845f, 1846, 1847f
Smudge cells, 443
Snakebites, **3716**
  clinical features of, 804, 3717, 3718f
  epidemiology of, 3716, 3717f
  field management of, 3717–3718
  global considerations in, 3722
  hospital management of, 3718–3722, 3720–3721t
  morbidity and mortality from, 3722
  snake identification, 3717
  wound infections following, 1143, 1145, 1145t
SNAREs, 969, 1233
SNCA gene mutations, 3498–3499, 3498t
Snellen chart, 221, E34f
Snowshoe hare virus, 1659t
SNPs. *See* Single-nucleotide polymorphisms (SNPs)
SNRIs (serotonin norepinephrine reuptake inhibitors), 101, 199, 3588, 3659t
SNRPN gene, 2993, 3791
SNRT (sinus node recovery time), 1918
SNS (Strategic National Stockpile), S4
Sochi virus, 1661t, 1676
Social determinants of health, 64, 65f
  health system factors, 64–66
  patient-level factors, 67, 67f
  provider-level factors, 66–67, 66f
Social history, 2
Social jetlag, 219, 3949, 3951
Social justice, 6
Social media, 16–17
Social phobia, 3661
Social Vulnerability Index (SVI), 64
SOD1, 3521, 3522t, 3523, 3524
SOD1 gene mutations, 3397, 3400, 3404
SOD3, 2251
Sodium
  in CKD, 2389
  fractional excretion of, 340t
  hormonal regulation of, 2371f, 2372
  intake, hypertension and, 2137t, 2141–2142
  recommended intake of, 2601t
  renal absorption/excretion of, 345–346, 345f, 346f, 2366–2367f, 2370–2372, 2371f, S1
  urine levels, in AKI, 340, 340t
Sodium bicarbonate
  adverse effects of, 2524
  for alcohol-induced acidosis, S1
  for hyperkalemia, 361
  for prevention of contrast-induced nephropathy, 1904
  for ventricular arrhythmias in overdose/poisoning, 3711t
Sodium channels
  cardiac, 1911–1912, 1912f
  muscle, 359
  muscle, disorders of, 3635, 3646–3647
  in seizure initiation, 3414
Sodium citrate (Shohl's) solution, 369, 372, 373, S1
Sodium current, in cardiac action potential, 1912f
Sodium diarrhea, congenital, 308
Sodium docusate, 82t
Sodium-glucose co-transporter 2 (SGLT2), 1981
Sodium-glucose co-transporter 2 (SGLT2) inhibitors
  adverse effects of, 3213
  for diabetes mellitus, 2359, 3211t, 3213, 3215, 3882t
  for heart failure, 1981, 1988, 1992, 1995t, 1996
  for ischemic heart disease, 2100, 2102–2103
  for MASH, 2706
  for OSA, 2282
  for peripheral artery disease, 2173
  renal protective effects of, 2373, 2392, 2396f
  for SIAD, 3020
Sodium hippurate, 1213
Sodium hypochlorite, 383, S5
Sodium iodide symporter (NIS), 2790t, 3022, 3028t
Sodium ipodate, 3024t
Sodium loading test, 3061
Sodium modeling, in dialysis, 2399
Sodium nitrite, S5
Sodium phosphate enema, 82t
Sodium picosulfate, 314
Sodium polystyrene sulfonate (SPS), 361, 373
Sodium restriction, 329
Sodium retention, 280, 280f
Sodium stibogluconate, 1746, 1784, 1785t, 1787
Sodium sulfacetamide, 391
Sodium thiosulfate, S5
Sodium tyropanoate, 3024t
Sodium valproate. *See* Valproate/valproic acid
Sodoku (rat-bite fever), 139t, 1143
SOFA score, for sepsis, 2291–2292, 2291t
Sofosbuvir
  actions of, 1492–1493, 2686f, 2689
  for chronic HCV infection, 1492–1493, 1492t, 2689
    with daclatasvir, 1493, 2689
    with ledipasvir, 1492t, 1493, 2688t, 2689–2690
    with velpatasvir, 1492t, 1493, 2688t, 2690–2691
    with velpatasvir and voxilaprevir, 1492t, 1493, 2688t, 2691
  pharmacology of, 1492–1493, 2689
Soft (argasid) tick, 3730. *See also* Tick(s)
Soft tick relapsing fever. *See* Relapsing fever
Soft tissue, 1049f
Soft tissue infections. *See also* Cellulitis
  A. baumannii, 1296
  Actinomyces, 1365
  anaerobic bacterial, 1377
  in cancer patient, 583
  deep. *See* Myositis; Necrotizing fasciitis
  enterococcal, 1218
  treatment of, 1054, 1054t
  V. vulnificus, 1330–1331
  in war veterans, S7
Soft tissue sarcomas, **727**
  classification of, 717–718
  diagnosis of, 728
  epidemiology of, 727
  genetic considerations in, 727
  incidence of, 727
  staging and prognosis of, 728, 728f
  treatment of, 728–729
SOI (severity-of-illness) scoring systems, 2291, 2291t, 2292t
Solar elastosis, 430, 431
Solar purpura, 407
Solar radiation, 429. *See also* Sun exposure
Solar urticaria, 402, 434, 2807
Solid organ transplantation
  heart. *See* Cardiac transplantation
  kidney. *See* Kidney transplantation
  liver. *See* Liver transplantation
  lung. *See* Lung transplantation
  organisms transmitted by, 1777
Solid organ transplant recipient. *See also* Immunocompromised patient
  abdominal pain in, 111
  allograft rejection in. *See* Rejection
  cfDNA for monitoring immunity in, 3986–3987, 3987f
  hemophagocytic lymphohistiocytosis in, 469
  infections in, **1154**
    cfDNA for diagnosis of, 3985–3986, 3986f, 3987f
    CMV, 1514
    CNS, 2350–2351
    cytomegalovirus, 1160–1161
    donor-derived, 1158–1160
    Epstein-Barr virus, 1161
    fungal, 1161–1163
    Legionella, 1271
    management of, 1159t
    mycobacterial, 1162f, 1163–1164
    Nocardia, 1162f, 1164
    polyomaviruses, 1161
    prophylaxis for, 1164–1165, 1164t
    risk factors for, 1157–1158, 1159t
    by site and transplant type, 1159t
    time course of, 1158
    vascular composite allografts and, 1158
  lipoprotein disorders in, 3247
  neurologic complications in, 2350–2351
  skin cancer in, 387, 432–433, 606, 2410
Solifenacin, 3549, 3588
Solitary fibrous tumor/hemangiopericytoma, 721
Solitary metastases, 735
Solitary pulmonary nodule, 622, 623t, 624f. *See also* Lung cancer
Solitary rectal ulcer syndrome, 318, 2561
Solriamfetol, 215
Soluble urokinase plasminogen activator receptor (suPAR), 2382
Somali famine, 3747
Soman (GD), 3712t, S5
Somatic hypermutation, 2777
Somatic mosaicism, 3791
Somatic mutations
  cancer development and, 519, 520f, 3792, 3792f
  effects of, 3783
  in mitochondrial DNA, 3791
Somatic symptom disorder, **3670**
Somatization, 291
Somatosensory cortex, 94, 94f, 200f
Somatosensory deficits, 180
Somatosensory evoked potentials (SSEPs), 2346
Somatosensory pathways, 173f

Somatostatin
  actions of, 2517, 2990
  half-life of, 2982
  synthesis of, 2990
Somatostatin analogues
  for acromegaly, 3009
  adverse effects of, 682, 2433, 3009
  for carcinoid syndrome, 312, 682, 683f
  for esophageal varices, 317
  for gastrinoma, 678, 678t
  for glucagonoma, 678, 678t
  for nausea and vomiting, 299t
  for NETs, 684–685
  for polycystic kidney disease, 2433
  structure, 683f
  for TSH-secreting adenomas, 3014
  for VIPoma, 678t, 679
  for ZES, 2537
Somatostatinoma, 309, 679
Somatotropin, 2995t
Somnambulism (sleepwalking), 217
Somnolence, 165
Somogyi effect, 3954
Sonelokimab, 2890
Söngling virus, 1662t
Sonic hedgehog factor pathway, 573
Sonidegib, 571t, 573, 608
Sorafenib
  actions and targets of, 527, 533t, 535, 548, 570t, 573
  adverse effects of, 320, 421, 508, 570t, 573, 753t, 755
  drug interactions of, 591
  for HCC, 535, 664, 664t, 665f, 666
  interactions and issues, 570t
  for NETs, 686
  for RCC, 535
  for thyroid cancer, 535
Sorbitol, 309, 2575, 2577t
Sorbitol dehydrogenase deficiency with peripheral neuropathy, 3601
Sore throat, **257**. *See also* Pharyngitis
Sotagliflozin, 1995t
Sotalol
  actions of, 1916t
  adverse effects of, 1920t, 1958
  for atrial fibrillation, 1952
  for atrial flutter, 1944
  dosage in kidney disease, 488
  dosage monitoring, 487
  for focal atrial tachycardia, 1938
  overdosage/poisoning with, 3712t
  for ventricular arrhythmias, 1958, 1972
Sotatercept, 2187f, 2194
Sotorasib, 569t, 572, 3982
Sour taste sensation, 238
Southeast Asia ovalocytosis, 794t, 795
Southern tick-associated rash illness (Master's disease), 139t, 146, 1452
Sowdah, 1823
SOX1 antibodies, 247t
SOX2 gene mutations, 3113
SOX3 gene mutations, 3068t, 3287
SOX9 gene, 3094, 3096f
SOX9 gene mutations, 3098, 3099t, 3100
SOX10, 2790t
SOX10 gene mutations, 397, 3111t, 3112
SOX17 gene mutations, 2185
SOX18 gene mutations, 2183
SP7 gene mutations, 3324, 3325t
Sp110 deficiency, 2801, S9
SPARC gene mutations, 2187f, 3324, 3325t, 3326
Sparganosis, 1839
Spasmodic dysphonia, 3510
Spasms
  hemifacial, 3554, 3555
  in MS, 3578, 3588
  muscle, 3635
  in tetanus, 1230
Spasticity, 168, 177t, 178, 3570, 3588
Spastin gene mutations, 3522t, 3525
SPATEs (serine protease autotransporters), 974
Spatial orientation, 204–206, 204f
Specialized nutritional support. *See* Enteral nutrition; Parenteral nutrition
Specific (secondary) granule, 453
Specific granule deficiency, 459t, 460f
Specificity
  in immune response, 2756
  of tests, 24, 514, 514t, S10
SPECT (single photon emission computed tomography)
  for myocardial perfusion imaging. *See* Myocardial perfusion imaging
  in parathyroid surgery, 3279
  radiopharmaceuticals for, 1878t, 1879
Spectinomycin, 1166, 1255, 1259, 1259t
Speech, 201, 3380, V2
Speech audiometry, 249
Speech reception threshold, 249
Spermatogenesis, 3106–3107, 3107f, 3784, 3789
Sperm banking, 576, 706, 3115
Spermidine, 3871, 3874f, 3875
Sperm, mutations in, 3784
SPG7 gene mutations, 3522t
SPG11 gene mutations, 3523t
Spherocytes, 437, 447, 447t
Spherocytosis. *See* Hereditary spherocytosis
Spherules, 1696, 1697f
Sphincter of Oddi, 2728
Sphincteroplasty, 2583
Sphincterotomy, endoscopic, 2473f
Sphingobacterium, 1269
Sphingomonas spp., 1269
Spider angioma, 327, 2629, 2709, 2709f
Spider bites, 405, 804, **3733**
Spike surface protein, 988, 998
SPIN11 gene mutations, 671t
Spina bifida, 120
Spinal accessory nerve, 3380
Spinal angiography, 3393–3394
Spinal cord
  anatomy of, 3559–3560, 3561f
  compression of. *See* Spinal cord compression
  disorders of. *See* Spinal cord disease/injury
  pain mechanisms of, 94, 95f
Spinal cord compression
  in hematomyelia, 3564
  IgG4-related, 2931t
  neoplastic
    clinical features of, 588–589, 3562
    diagnosis of, 589, 589f, 620, 3562, 3563f
    in intradural mass lesions, 3563, 3563f
    in lung cancer, 619, 620
    metastatic, 588, 725, 725f, 3562, A16
    pathogenesis of, 3562
    in primary intramedullary tumors, 3563, 3563f
    treatment of, 589–590, 589f, 3562
  in spinal epidural abscess, 991t, 3563–3564, 3563f
  in spinal epidural hematoma, 3564
Spinal cord disease/injury, **3559**
  anatomic considerations in, 3559–3560, 3560t, 3561f
  approach to the patient, 3559–3561, 3560t
  autonomic dysfunction in, 3549, 3570
  cervical. *See* Cervical spine
  compressive. *See* Spinal cord compression
  constipation in, 314
  in HIV infection, 1609
  imaging in, A16
  infarction, 3564–3565, A16
  infectious. *See* Meningitis
  level of lesion in, 3382t, 3560, 3560t
  in multiple sclerosis, A16
  myelopathies. *See* Myelopathy
  osteoporosis and, 3296
  paraneoplastic syndromes, 731–732
  patterns of, 3560–3561
  priapism in, 3155
  prognosis of, 3569t
  radiation-induced, 554
  referred abdominal pain from, 111
  sensory abnormalities in, 176
  sinus bradycardia in, 1920
  stem cell therapy for, 3406–3407
  testicular function in, 3115
  treatment of, 3569–3570
  tumors, 131
  vascular malformations, 3567, 3567f
Spinal epidural abscess
  clinical features of, 994, 3563–3564
  diagnosis of, 3563–3564, 3563f
  etiology of, 994, 3563
  risk factors for, 3563
  treatment of, 991t, 3564
Spinal epidural hematoma, 3564, A16
Spinal longitudinally extensive collection, A16
Spinal manipulative therapies, 127, 3925t, 3927–3928
Spinal meningitis, 1128
Spinal muscular atrophy (SMA), 3522t, 3524, 3824t, 3828
Spinal osteomyelitis. *See* Osteomyelitis, vertebral
Spinal stenosis, 120, 122f, 124t, 128, 179
Spinareoviridae, 1478f, 1480t, 1657t, 1663, 1664–1665t, 1666t, 1669
Spindle cell tumor, 411
Spine
  cervical. *See* Cervical spine
  lumbar. *See* Lumbar spine
SPINK1 gene mutations, 2744, 2751
Spinobulbar muscular atrophy (Kennedy's disease)
  vs. ALS, 3524
  clinical features of, 3115, 3524
  genetic considerations in, 3522t, 3524, 3783t, 3786, 3793, 3793t
Spinocerebellar ataxia (SCA)
  classification of, 3534
  clinical features of, 3534
  genetic considerations in, 3534, 3783t
  SCA1, 3534–3535, 3535f
  SCA2, 3535, 3793t
  SCA3 (Machado-Joseph disease), 3535, 3793t
  SCA6, 3535, 3793t
  SCA7, 3535–3536, 3793t
  SCA8, 3536
  SCA12, 3793t
  SCA27B, 3536
  treatment of, 3537
Spinosad, 3729
Spinothalamic tract, 94, 95f, 172, 173f, 176
Spiramycin, 1744t, 1752, 1802
Spiration Valve System, 2290
Spirillum infection, 1143
Spirillum minus, 139t
Spirometra, 1839
Spirometry
  in COPD, 2251, 2254, 2255t
  preoperative, 3909, 3909t
  in respiratory disease evaluation, 2200–2201, 2202f, 2206–2207
Spironolactone
  for acne vulgaris, 391
  actions of, 2370
  adverse effects of, 2146
    androgen blockade, 3115
    erectile dysfunction, 2144t, 3157
    gynecomastia, 329, 2144t
    hyperkalemia, 329, 2144t, 3141

    - hyponatremia, 329
    - hypotension, 3141
  - for ascites, 329, 2715
  - for heart failure, 1988, 1993t, 1994, 1995t
  - for hirsutism, 3141
  - for hypertension, 2139, 2144t, 2146
  - for male precocious puberty, 3110
  - for mineralocorticoid excess, 3063
  - for PCOS, 3135
  - for STEMI, 2011
  - for syndrome of apparent mineralocorticoid excess, 356
- Spirulina, 420
- Spleen
  - accessory, 478
  - congenital absence of, 482, 1004t, 2795–2796, S9
  - infarction of, 112t, 479
  - palpation of, 479
  - percussion of, 480
  - rupture of, 112t, 479, 481, 1510, 1780
  - structure and function of, 478–479, 479f
- Splenectomy, **480**
  - for autoimmune hemolytic anemia, 802
  - for β thalassemia, 778t
  - complications of, 482
  - for congenital erythropoietic porphyria, 3349
  - for hairy cell leukemia, 873
  - for hereditary spherocytosis, 795
  - for Hodgkin's lymphoma staging, 480
  - for immune thrombocytopenic purpura, 924
  - immunizations after, 1004t
  - immunizations before, 579, 795
  - infections after, 578t, 579, 795, 990, 991t
  - in leukemia, 480–481
  - long-term effects of, 482
  - for lymphoma, 481
  - for marginal zone lymphoma, 865
  - for myelofibrosis, 823
  - for polycythemia vera, 821
  - for pyruvate kinase deficiency, 796
  - for sickle cell disease, 774t
  - for splenic abscess, 1075
  - for splenic marginal zone lymphoma, 872
- Splenic abscess
  - abdominal pain in, 111, 112t, 1075
  - *Actinomyces,* 1364, 1364f
  - clinical features of, 1075
  - etiology of, 1075
  - in infective endocarditis, 1047, 1075
  - treatment of, 1075
- Splenic marginal zone lymphoma, 481, 857, 860t, 865, 871, 872, 872t
- Splenic vein thrombosis, 2750
- Splenomegaly, **478**
  - abdominal swelling in, 327
  - approach to the patient, 479–480
  - clinical features of, 479–480
  - in CML, 837
  - congestive, 480
  - differential diagnosis of, 480, 1784
  - diseases associated with, 481t
  - in Gaucher disease, 3355
  - in HCC, 664
  - hyperreactive malarial, 1766
  - laboratory evaluation of, 480
  - in liver disease, 2629
  - with lymphadenopathy, 477
  - massive, 479, 481t
  - in MDS, 816
  - in myelofibrosis, 821, 823
  - neutrophil destruction and, 458
  - in parasitic infections, 1739t
  - in polycythemia vera, 451, 820
  - in portal hypertension, 2714–2715
  - in schistosomiasis, 1829
  - in β thalassemia, 778t
  - in visceral leishmaniasis, 1784, 1784f
- Splenosis, 481
- Splinter hemorrhages, 1860
- Spondweni virus, 1660t
- Spondylitis
  - ankylosing. *See* Axial spondyloarthritis (ankylosing spondylitis)
  - in psoriasis, 386
  - tuberculous (Pott's disease), 1075, 1388, 1388f
- Spondyloarthritis, **2879**
  - axial. *See* Axial spondyloarthritis (ankylosing spondylitis)
  - genetic considerations in, 2880–2882
  - IBD-associated, 2891–2892
  - KIRS in, 2772t
  - pathogenesis of, 2880–2882, 2881f
  - pathology of, 2880
  - peripheral, 2892, 2892t
  - psoriatic. *See* Psoriatic arthritis
  - reactive arthritis. *See* Reactive arthritis
  - spectrum of, 2879–2880, 2880t
- Spondylodiskitis. *See* Osteomyelitis, vertebral
- Spondylosis, 178
- Spondylotic myelopathy, 3566–3567
- Spongiosis, 382
- Spontaneous bacterial empyema, 330
- Spontaneous bacterial peritonitis (SBP), 329–330, 1070–1071, 2716. *See also* Peritonitis
- Spontaneous breathing trial, 2308, 2308f
- Spontaneous speech, 201
- Sporadic goiter, 3040
- Sporadic late-onset nemaline myopathy, 3644
- Spore, anthrax, S4
- Sporogony, 1762
- *Sporothrix schenckii,* 1724
- Sporotrichosis, **1724**
  - clinical features of
    - arthritis, 1058
    - crusted lesions, 1051
    - lymphocutaneous, 1689t
    - meningitis, 1130t
    - nodules, 405, 1416, 1724, 1724f
  - diagnosis of, 1689t, 1724
  - etiology of, 1724
  - pathogenesis of, 1724
  - risk factors for, 1689t
  - treatment of, 1724–1725, 1724t
- Sporozoites, *Plasmodium,* 1760, 1761f
- Sports. *See* Athlete(s)
- Spousal abuse. *See* Domestic abuse/violence
- Sprouting, 545, 545f. *See also* Angiogenesis
- *SPRY4* gene mutations, 3112, 3112t
- SPS (sodium polystyrene sulfonate), 361, 373
- *SPTA1* gene mutations, 794t
- *SPTB* gene mutations, 794t
- Spur cell(s), 446f, 447
- Spur cell anemia, 323
- Spurling's maneuver, 129
- Sputum analysis
  - in asthma, 2223
  - in cough, 273
  - Gram stain, 1025
  - in *Legionella* pneumonia, 1273
  - in lung cancer, 619
  - with PCR assays, 2214
  - in pneumonia, 1025
  - in respiratory disease evaluation, 2201, 2214
  - specimen collection and transport, S11
  - in TB, 1392
- *SQSTM1* gene, 3312
- Squamous cell carcinoma
  - anal, 659
  - bladder, 691, 692f. *See also* Bladder cancer
  - cervix. *See* Cervical cancer
  - endometrium. *See* Endometrial cancer
  - esophagus, 641–642, 643t. *See also* Esophageal cancer
  - genetic considerations in, 3939
  - head and neck region, 610. *See also* Head and neck cancer
  - lung, 615–616, 616f. *See also* Lung cancer
  - oral, 261, 263t, 388
  - skin
    - BRAF inhibitors and, 605
    - clinical features of, 379t, 606–607, 608f, A5
    - distribution of, 380f
    - epidemiology of, 606
    - genetic considerations in, 606
    - natural history of, 607
    - pathophysiology of, 606
    - prevention of, 609
    - risk factors for, 606
    - screening for, 518
    - in solid organ transplant recipient, 606
    - sun exposure and, 431, 606
    - treatment of, 608
  - of unknown primary, 734f, 735
- Square root sign, 2083
- Square-wave jerks, V3
- *SRD5A2* gene mutations, 3099, 3099t, 3115
- *SREB,* 1701
- SRP54 (signal recognition protein), 2790t
- SRS (stereotactic radiosurgery), 590, 724, 3001
- *SRSF2* gene mutations, 823, 828
- *SRY* gene, 3102
- *SRY* gene mutations
  - in disorders of testis development, 3094, 3098, 3099t, 3789, 3791
  - in testicular/ovotesticular disorders of sex development, 3100, 3101t
  - transcription factor in, 3783t
- SSc sine scleroderma, 2859. *See also* Systemic sclerosis (scleroderma)
- SSEPs (somatosensory evoked potentials), 2346
- SSKI (saturated solution of potassium iodide), 1724–1725, 1724t
- SSPE (subacute sclerosing panencephalitis), 1117–1118, 1641
- SSRIs. *See* Selective serotonin reuptake inhibitors (SSRIs)
- *SSX1/2* gene mutations, 728
- Stabbing headache, primary, 3474
- Stab phlebectomy, 2182
- *STAG2* gene mutations, 828
- Staghorn calculi, 2453
- Stagnant bowel syndrome. *See* Small-bowel bacterial overgrowth
- Staining, S11
- Standard precautions, 1148, 1538
- Stanford V regimen, 870
- Stapedectomy, 245, 249f, 250
- Stapes, 244, 245f
- Staphylococcal infections, **1195**
  - aortitis, 2171
  - in cancer patient, 581
  - cellulitis, 1052
  - clinical syndromes, 1202
  - diagnosis of, 1198, 1202
  - in dog-bite wound, 1143
  - endocarditis. *See* Infective endocarditis, staphylococcal
  - food-borne, 1080, 1080t
  - meningitis. *See* Meningitis
  - osteomyelitis, 1062, 1064t, 1065t, 1067
  - pathogenesis of, 1202
  - of prosthetic devices, 1202
  - *S. aureus. See Staphylococcus aureus* infections
  - in scabies, 3729
  - treatment of, –1174t, 1202–1204, 1203t
- Staphylococcal scalded skin syndrome
  - clinical features of, 141t, 146, 1201
  - pathogenesis of, 1198
  - vs. scarlet fever, 146–147

Staphylococcal scalded skin syndrome (*Cont.*):
  skin manifestations of
    generalized desquamation, 1201, 1202f, A1
    Nikolsky's sign, 146
    progression, 401
    rash, 141t, 146, 401
    vesicles/bullae, 401, 1051
  vs. toxic epidermal necrolysis, 147, 401, 1051
Staphylococcus argenteus, 1196
Staphylococcus aureus
  antibiotic resistance in, 1173t, 1181, 1183, 1184, 1185, 1203–1204
  antimicrobial susceptibility testing, S11
  enterotoxins, 965
  evasion of host defenses by, 1198
  genome of, 1196–1197
  laboratory identification of, 1195–1196, 1196f, 1198
  methicillin-resistant. *See* Methicillin-resistant *Staphylococcus aureus* (MRSA) infections
  pathologic response to, 1196
  T cell activation by, 974
  vancomycin-intermediate, 1044, 1176, 1184, 1204
  virulence regulation in, 1197
Staphylococcus aureus infections
  in atopic dermatitis, 383, 3838
  CAPD peritonitis, 1073
  clinical syndromes, 1199t
  in diabetes mellitus, 3229, 3230
  diagnosis of, 1198, S11
  endocarditis. *See* Infective endocarditis, staphylococcal
  epidemiology of, 1196
  food-borne
    clinical features of, 306t, 1080, 1080t
    diagnosis of, 1201
    differential diagnosis of, 306t, 1201
    pathogenesis of, 1198
    sources of, 305, 1080t, 1201
  host response to, 1198
  in human bite-wound, 1144
  in leukocyte adhesion deficiency, 460
  meningitis. *See* Meningitis, bacterial
  methicillin-resistant. *See* Methicillin-resistant *Staphylococcus aureus* (MRSA) infections
  musculoskeletal, 1199–1200
    osteomyelitis, 1062, 1067, 1068, 1199, 1199f
    pyomyositis, 1053, 1200
    septic arthritis, 1056, 1199–1200. *See also* Infectious (septic) arthritis
  nausea and vomiting in, 297
  otitis externa, 254
  parotitis, 266
  pathogenesis of, 1196–1198
  peritonitis, 2716
  prevention of, 1205–1206
  prosthetic device–related, 1200–1201
  psoas abscess, 1075–1076
  respiratory tract, 274, 1035, 1200
  risk factors for, 1196
  in sepsis/septic shock, 2316
  skin and soft tissue, 1198–1199
    carbuncles, 1198
    cellulitis, 1052
    folliculitis, 395, 1051, 1198
    furuncles, 1198
    furunculosis, 389
    impetigo, 388, 388t, 1209–1210. *See also* Impetigo
    mastitis, 1199
    staphylococcal scalded skin syndrome. *See* Staphylococcal scalded skin syndrome
    vesicles/bullae, 401
  splenic abscess, 1075
  toxic shock syndrome. *See* Toxic shock syndrome, staphylococcal
  toxin-mediated disease, 1198
  treatment of, 1174t, 1202–1205, 1203t
  urinary tract, 1087, 1201
  vaginal, 1103
Staphylococcus epidermidis infections, 1202
Staphylococcus lugdunensis infections, 1038, 1202
Staphylococcus pseudintermedius, 1196
Staphylococcus saprophyticus infections, 1087, 1100, 1202
Staphylococcus schleiferi infections, 1202
StAR (steroidogenic acute regulatory protein), 2982, 3098, 3099t
Starfish envenomation, 3724, 3724f
STARI (Southern tick-associated rash illness), 139t, 146, 1452
Starvation-related malnutrition, 2620
Stasis dermatitis, **385**
  clinical features of, 378t, 380f, 385, 393t, A5
  evolution of, 385f
  treatment of, 385f, 393t
Stasis ulcer, 380f, 385, 385f
STAT1 gene mutations, 1415f, 1416, 1708, 2786f, 2801
STAT3 deficiency. *See* Hyper-IgE (Job) syndrome
STAT3 gene mutations, 463, 1689, 1690f, 2554t, 2786f, 2804
STAT4 gene mutations, 2822, 2823t
STAT5b gene mutations, S9
STAT factors, 526f, 2774
Statins
  adverse effects of, 3244t
    hepatotoxicity, 2665, 2671, 3251, 3258
    myopathy, 2912, 2968, 3251, 3647–3648, 3648t
    tendinitis, 2940t, 2974
  for colorectal cancer prevention, 513
  for dementia prevention, 3482
  in diabetes mellitus, 3228
  drug interactions of, 489t, 1742t, 3251
  for heart failure, 157
  for intermittent claudication, 2174
  for lipoprotein disorders
    approach to, 3251
    dosage of, 3244t
    in ischemic heart disease, 2099
    in metabolic syndrome, 3257
    in older adults, 3882
  for MASLD/MASH, 2707
  metabolism of, 485t
  for NSTE-ACS, 2110, 2112
  perioperative, 3908
  for stroke prevention, 3449
  timing of administration of, 3954
Status asthmaticus, 2227
Status epilepticus, **3425**
  complex partial, 3484
  in critically ill patient, 2298, 2346, 2347, 2347f
  definition of, 3425
  in nerve agent exposure, S5
  nonconvulsive, 3425
  treatment of, 3426f
Stauffer's syndrome, 688
Stavudine (d4T), 1616, 1619f, 1622f, 2672, 3611
Steady state, 485, 485f, 486
Steatohepatitis
  drug-associated, 2703t
  metabolic dysfunction–associated. *See* Metabolic dysfunction–associated steatohepatitis (MASH)
  pathogenesis of, 2701–2702
Steatorrhea
  definition of, 2540
  diagnosis of, 2540, 2549–2550
  etiology of, 309, 2543t, 2549
  lipid digestion and absorption defects in, 2541t
  in pancreatic cancer, 673
  in pancreatitis, 2541t, 2542, 2753
  in protein-losing enteropathy, 2548
  in small-bowel bacterial overgrowth, 2543
  in tropical sprue, 2546
  in Whipple's disease, 309
STEC (Shiga toxin–producing *E. coli*), 1077, 1081, 1285t, 1287–1288. *See also Escherichia coli* infections, intestinal
Steele-Richardson-Olszewski syndrome. *See* Progressive supranuclear palsy syndrome
Steinert's disease, S9
Stem cell(s)
  cancer. *See* Cancer stem cells
  cord blood, 915
  exhaustion, aging and, 3871–3872
  hematopoietic. *See* Hematopoietic stem cells
  induced pluripotent, 3402
  mesenchymal, 3408
  mtDNA mutations, 3816
  neural, 3405
Stem cell factor (SCF), 2767t, 2793
Stem cell therapy
  for age-related macular degeneration, 3407–3408
  for ALS, 3407
  for PD, 3405–3406, 3505–3506
  sources of cells for, 3402
  for spinal cord injury, 3406–3407
Stem cell transplantation. *See* Hematopoietic cell transplantation (HCT)
STEMI. *See* ST-segment elevation myocardial infarction (STEMI)
Stemmer's sign, 2184
Stenotrophomonas maltophilia infections, 582, 1309t, 1314, S11
Stent(s)
  for aortic aneurysm, 2168, 2168f
  biliary, in pancreatic cancer, 2495, 2500f
  for chronic venous disease, 2182
  coronary artery. *See also* Percutaneous coronary interventions (PCI)
    characteristics of, 2127–2128, 2127f
    drug-eluting, 943, 1928, 2104, 2128, 2129f, A11, A11v
    restenosis in, 2104, 2128–2129, A11, A11v
  esophageal, for malignant dysphagia, 2493, 2499f
  for peripheral artery disease, 2174
  for renal artery stenosis, 2151
Steppage gait, 179
Stereognosis, 174, 3382
Stereopsis, 221
Stereotactic body radiotherapy (SBRT), 590, 596, 622–623, 728
Stereotactic EEG, 3425
Stereotactic radiation cardiac ablation, 1968, 1968f
Stereotactic radiosurgery (SRS), 590, 724, 3001
Stereotyping, 66, 69
Sterilization, 3152–3153, 3153t
Sternal osteomyelitis, 1067–1068, 1067f
Sternoclavicular joint, infectious arthritis of, 1056, 1056f
Steroid(s). *See specific hormones*
Steroidogenesis. *See specific hormones and glands*
Steroidogenic acute regulatory protein (StAR), 2982, 3098, 3099t
Steroidogenic enzyme pathways, disorders of, 3098–3099, 3099t, 3100f
Steroidogenic factor 1 (SF1), 3050
Steroid purpura, 407
Stevens-Johnson syndrome (SJS)
  clinical features of, 142t, 147, 401, 424–425, 424–425f, 426t, 427t
  diagnosis of, 426t
  differential diagnosis of, 424

  drugs associated with, 425, 427t
  epidemiology of, 142t
  etiology of, 142t, 401
  genetic considerations in, 419, 493
  in HIV infection, 1606
  in *M. pneumoniae* infection, 1465
  oral manifestations of, 263t, A3
  pathogenesis of, 417t, 418f, 489, 2816f
  prevention of future episodes, 427
  timeline of, 427, 428f
  toxic epidermal necrolysis overlap with, 424, 424f
  treatment of, 425
Stickler syndrome, 3331
Stiff-legged gait, 178
Stiff-person syndrome
  clinical features of, 3635
  gait disorders in, 178
  paraneoplastic, 750, 750f, 2791t
  pathogenesis of, 2790t, 3635
Still's disease
  clinical features of, 139t
    anemia, 768
    fever of unknown origin, 150
    macrophage activation syndrome, 469
    nodules, 403
    rash, 146, A1
    urticaria, 402
  epidemiology of, 139t
  treatment of, 155
Still's murmur, 286
Stimulant use disorder, 3695. *See also* Psychostimulants
STING, 2937
Stinging ants, 3736
Stingray injury, 3725
STIs. *See* Sexually transmitted infections (STIs)
STK11 gene mutations, 538, 633, 671t
Stokes-Adams attack, 160
Stomach
  acid and hormone secretion in, 2517–2518, 2518f
  biopsy in cobalamin deficiency, 786
  epithelial anatomy of, 2514, 2515f
  functions of, 2459
  mucosal defense and repair in, 2515–2517, 2516–2517f
Stomach tumors
  adenocarcinoma. *See* Gastric adenocarcinoma
  lymphoma, 649–650
  neuroendocrine tumors, 652. *See also* Neuroendocrine tumors (NETs)
Stomatitis, 380f, 388, 2549t
Stomatocytes, 437, 440f
Stomatocytosis, 440f, 794f, 794t, 795
Stonefish sting, 3725
Stony coral injury, 3725
Stool
  microbiota of, 3831, 3833f
  osmolality of, 2549–2550
  retained, abdominal swelling and, 327
Stool analysis
  antigen test, for *H. pylori,* 1301, 1301t, 2524t
  in chronic diarrhea, 311
  for intestinal parasite detection, S12
  in malabsorption disorders, 2549–2550
  for occult blood. *See* Fecal occult blood
  in pancreatic disease evaluation, 2739t, 2742
  specimen collection and transport for, S11
Stool-bulking agents, 2574–2575
Stool culture, 306, 1084t
Stool DNA testing, 517
Stool softeners, 82, 82t
STOP-Bang survey, 2275
"Stop rules," 49t
Storage pool disorder, platelet, 926
Straight back syndrome, 1860, 2058
Straight leg raising (SLR), 120, 121t
Strategic National Stockpile (SNS), S4
Strawberry gallbladder, 2734
Strawberry gums, 261
Strawberry tongue, 265t, 402, 1209
Straw itch mite, 3730
Strength testing, 3381
Streptobacillus moniliformis, 141t, 1143
Streptocerciasis/*Mansonella streptocerca,* 1820t, 1824, S12
Streptococcal infections. *See also* Group A *Streptococcus* (GAS) infections; Group B *Streptococcus* (GBS) infections
  aortitis, 2171
  brain abscess. *See* Brain abscess
  cellulitis, 1052, 1210–1211
  in dog-bite wound, 1143
  endocarditis. *See* Infective endocarditis, streptococcal
  epiglottitis, 259
  global considerations in, 2033
  glomerulonephritis following. *See* Poststreptococcal glomerulonephritis
  in human bite-wound, 1144
  mastoiditis, 255
  meningitis. *See* Meningitis
  osteomyelitis, 1062, 1064t, 1065t
  otitis media, 254
  pharyngitis, 257–258
  in sepsis/septic shock, 2316
  skin infections, 1051, 1052
  in solid organ transplant recipient, 1155t
  toxic shock syndrome. *See* Toxic shock syndrome (TSS), streptococcal
Streptococci
  anaerobic, 1214. *See also* Anaerobic bacteria
  characteristics of, 1206, 1206t
  classification of, 1206, 1206t
  laboratory identification of, S11
  nutritionally variant, 1214
  viridans, 1174t, 1206, 1206t, 1214
Streptococcus agalactiae, 1206, 1206t, 1213. *See also* Group B *Streptococcus* (GBS) infections
Streptococcus bovis infections. *See* Group D *Streptococcus* infections
Streptococcus dysgalactiae, 1212
Streptococcus iniae, 1214
Streptococcus milleri, 1206t, 1212, 1214
Streptococcus mutans, 260
Streptococcus pneumoniae. See also Pneumococcal infections
  antibiotic resistance in, 1026–1027, 1122, 1173t, 1181, 1183, 1190, 1193
  genome of, 1188
  immune response to, 1191
  laboratory identification of, 1187, 1187f
  molecular typing of, 1188
  nasopharyngeal carriage of, 1189, 1189f
  replicative niche, 967
  serotype distribution of, 1189
  virulence factors in, 1187–1188, 1188f
Streptococcus pyogenes, 1206
Streptococcus pyogenes infections. *See* Group A *Streptococcus* (GAS) infections; Necrotizing fasciitis
Streptococcus suis, 1214
Streptogramins. *See* Quinupristin-dalfopristin
Streptokinase, 954–955, 954f, 2121
Streptomycin
  actions of, 1166, 1426
  adverse effects of, 228, 1340, 1400, 1426
  for *Brucella* infections, 1334
  dosage of, 1426
  for MDR-TB, 1394t
  pharmacology of, 1426
  for plague, 1346, 1346t, S4
  in pregnancy and lactation, 1169t
  resistance to, 1184, 1219, 1221, 1395, 1426
  for TB, 1395, 1400, 1426
  for tularemia, 1339–1340, S4
  for verruga peruana, 1353t
Streptozotocin, 685, 3082
Stress, 3, 2986
Stress disorders, **3661,** 3927. *See also* Posttraumatic stress disorder (PTSD)
Stress incontinence, 3890–3891, 3893f, 3893t
Stress testing
  ECG. *See* Electrocardiography (ECG), stress
  echocardiography. *See* Echocardiography, stress
  myocardial perfusion imaging. *See* Myocardial perfusion imaging
  in peripheral arterial disease, 2172
  risk stratification using, 1888, 1889f, 3906–3907
Stress ulcers, 2297
Stridor, 2200
Strimvelis, 3824t
"String sign," 2558
Stroke, **3428**. *See also* Intracerebral hemorrhage; Ischemic stroke
  approach to the patient, 3428–3429, 3429f
  vs. botulism, 1236
  circadian variations in, 219
  clinical features of, 3428
    aphasia, 201–203
    coma, 192
    delirium, 185
    depression, 3664
    dysphagia, 294
    exercise intolerance, 291
    gait disorders, 178
    ocular, 230, 237
    seborrheic dermatitis, 385
    weakness, 166
  in critically ill patient, 2298
  deaths from, 75t
  definition of, 3428
  dementia and, 3489
  dental care after, 266
  etiology of, 3428
  fatigue following, 167
  global considerations in, 1855–1856
  global mortality from, 3844
  at high altitudes, 3759
  in HIV infection, 1609
  hypokalemia and progression of, 356
  hypothermia in, 3768
  imaging studies in
    cerebral angiography, 3437–3438
    CT, 3435, 3437, 3437f, 3453f
    MRI, 3437, 3438f, A16
    perfusion studies, 3437f, 3438–3439
    ultrasound, 3438
  with left ventricular assist systems, 2032
  localization of
    anterior cerebral artery, 3430, 3430f, 3431f
    anterior choroidal artery, 3430, 3432f
    basilar artery, 3434–3435
    common carotid artery, 3431
    internal carotid artery, 3430–3431, 3431–3432f
    middle cerebral artery, 3429–3430, 3430–3431f
    posterior cerebral artery, 3431, 3432f, 3436f
    vertebral and posterior inferior cerebellar arteries, 3432–3433, 3432f, 3433f
  osteoporosis following, 3296
  postmenopausal hormone therapy and, 3144t, 3146, 3147f
  in pregnancy, 3902
  prevention of. *See* Ischemic stroke, prevention of
  risk factors for, 3446t, 3448–3449
  in sickle cell disease, 774t

- Stroke (*Cont.*):
  - in SLE, 2830
  - in transplant recipient, 2350
  - treatment of, 954, 3429, 3429f, 3963–3964, 3964f, 3968
- Stroke centers, 3443
- Stroke volume, 1850, 1852f, 1852t, 2310
- Stroma-directed antibodies, 559–560, 559f
- Stromal cell–derived factor (SDF), 1577, 2768t
- Stromal tumors, ovarian, 712
- *Strongyloides stercoralis*/strongyloidiasis
  - ATL and, 1554
  - clinical features of, 1738t, 1815t, 1816–1817, 2234, S7
  - diagnosis of, 1817, S7, S12
  - eosinophilia in, 961t, S12
  - epidemiology of, 961t, 1736, 1738t, 1816, 2234, S7, S12
  - incubation period for, S7
  - life cycle of parasite, 1815t, 1816, 1817f, S12
  - prevention of, 1019
  - in solid organ transplant recipient, 1164t
  - treatment of, 1748, 1815t, 1817, S7
  - in war veterans, S7
- ^{90}Strontium exposure, S6
- ^{89}Strontium therapy, 82, 731
- Structure (quality theory), 43
- Struma ovarii. *See* Teratoma, ovarian
- Struvite stones, 2453. *See also* Nephrolithiasis
- ST segment elevation, 1872–1873, 1874t, 1970
- ST-segment elevation myocardial infarction (STEMI), **2113**. *See also* Myocardial infarction
  - circadian variations in, 2114
  - clinical features of, 103t, 2114–2115
  - complications of, **2122**
    - accelerated idioventricular rhythm, 2124
    - cardiogenic shock, 2123, 2325, 2328f. *See also* Cardiogenic shock
    - conduction disturbances, 2125
    - heart failure, 2123
    - hypovolemia, 2123
    - left ventricular aneurysm, 2125
    - papillary muscle rupture, 2122t
    - pericarditis, 2125
    - pulmonary edema, 2330
    - pump failure, 2122–2123
    - recurrent chest pain, 2125
    - right ventricular infarction, 2012
    - sinus bradycardia, 2124–2125
    - supraventricular arrhythmias, 2124
    - thromboembolism, 2125
    - ventricular dysfunction, 2122
    - ventricular free wall rupture, 2122t
    - ventricular premature beats, 2123
    - ventricular septal rupture, 2122t
    - ventricular tachycardia and fibrillation, 2123–2124, 2124f
  - diagnosis of
    - biomarkers, 2115–2116, 2117f
    - ECG, 1873, 1874t, 2114f, 2115, 2119t, A7, A11
    - echocardiography, 2116–2117, A9
    - physical examination, 2115
    - radionuclide imaging, 2117
  - differential diagnosis of, 2115
  - epidemiology of, 2113
  - incidence of, 2106, 2113
  - nonspecific reaction to, 2116
  - painless, 2115
  - pathophysiology of, 2113–2114, 2114–2115f. *See also* Atherosclerosis; Coronary artery disease (CAD)
  - during PCI, 2118t, 2128
  - recurrent, 2125–2126
  - risk factors for, 2113–2114
  - risk stratification and management after, 2125–2126
  - secondary prevention after, 2126
  - silent, 2118t
  - stages of, 2115, 2116f
  - TIMI grading system, 2119
  - treatment of
    - ACE inhibitors, 2121–2122
    - activity restriction, 281
    - antiplatelet agents, 2121
    - antithrombotic therapy, 2121
    - ARBs, 2122
    - aspirin, 2118
    - beta blockers, 2118, 2121
    - cardioverter-defibrillator, 2124, 2124f
    - diet therapy, 281
    - in emergency department, 2117–2118, 2119f
    - fibrinolysis, 2120
    - in hospital/coronary care unit, 2120–2121
    - integrated approach to, 2119, 2119f
    - limitation of infarct size, 2119
    - morphine, 2118
    - nitrates, 2118, 2122
    - oxygen therapy, 2118
    - pacemaker, 2125
    - pain control, 2118
    - PCI, 2119–2120, 2119f, A11
    - prehospital care, 2117
    - reperfusion therapy, 2119f, 2120
- *STUB1* gene mutations, 3112t
- Stupor, 186
- Sturge-Weber syndrome, A16
- STW 5, 302, 2464
- *STX11* gene mutations, 465
- *STXBP2* gene mutations, 464, 465
- Sty (hordeolum), 224
- Stylostome, 3730
- Subacromial bursitis, 2973
- Subacute combined degeneration, 3568
- Subacute cutaneous lupus erythematosus, 405, 414–415, 420, 420f
- Subacute sclerosing panencephalitis (SSPE), 1117–1118, 1641
- Subarachnoid hemorrhage, **3457**
  - vs. bacterial meningitis, 1121
  - clinical features of, 3458–3459
  - in critically ill patient, 2298
  - diagnosis of, 3459–3460, 3460f
  - DVT prophylaxis after, 3461
  - ECG in, 1874, 1875f
  - epidemiology of, 3458
  - fundus examination in, 189
  - grading of, 3459t
  - headache in, 114, 114t, 3458
  - at high altitudes, 3758
  - hyponatremia in, 348, 3459
  - hypothermia in, 3768
  - neurologic deficits after, 3459
  - pathophysiology of, 3458, 3458f
  - in polycystic kidney disease, 2432
  - traumatic, 3573
  - treatment of, 3460–3461, 3461f
- Subconjunctival hemorrhage, 223
- Subcortical aphasia, 202–203
- Subcutaneous panniculitis-like T cell lymphoma, 867, 875
- Subdural empyema, **1138**
  - clinical features of, 994, 1139–1140
  - definition of, 1138, 1140f
  - diagnosis of, 1140, 1140f
  - differential diagnosis of, 1140
  - epidemiology of, 1138
  - etiology of, 994, 1138, 1375
  - pathophysiology of, 1139–1140
  - prognosis of, 1141
  - treatment of, 1140
- Subdural hematoma
  - acute, 3572, 3572f
  - chronic, 3572–3573, 3572f
  - dementia in, 3484
  - imaging in, 3572f, A16
  - spinal, A16
  - weakness in, 170
- Subependymal giant cell astrocytomas, 720
- Suberosis, 2228t
- Subfascial endoscopic perforator surgery, 2182
- Subjective cognitive decline, 194
- Subjective Global Assessment, of nutrition, 2616
- Subluxation, 2942, 2942t
- Submandibular abscess, 259
- Submandibular space infections, 260, 1374
- Submucosal plexus, 303
- Subperiosteal abscess, 255
- Sub-Saharan Africa, 3861f, 3862, 3865
- Substance P, 94, 94f, 293, 303
- Substance use disorders. *See also* Persons who inject drugs (PWID); *specific substances*
  - delirium in, 184
  - dementia in, 3485
  - emerging drugs in, 3696–3697
  - future directions in treatment of, 3697
  - genetic considerations in, 3942
  - global considerations in, 3697
  - in LGBTQIA+ people, 3178
  - in men, 3168
  - mental health issues and, 3697
  - neuropathology of, 3656, 3656t
  - pathogenesis of, 3654, 3654f, 3655f, 3656
  - polydrug, 3692
  - prevalence of, 3657
  - in women, 3168
- Substernal goiter, 3041
- Substrate reduction therapy, 3356
- Succimer (DMSA), 3700–3701t
- Succinate dehydrogenase (SDH) enzyme, 3073
- Succinylcholine, 358, 485t, E72t
- Suck response, 3381
- Sucralfate, 302, 2524t, 2526
- Sucrase-isomaltase deficiency, 2545
- Sucrosomial iron, 767t
- Sudan virus, 1680, 1680–1682f
- Sudden cardiac arrest (SCA), 2332. *See also* Cardiac arrest
- Sudden cardiac death (SCD)
  - circadian variations in, 219
  - definition of, 2332, 2333t
  - epidemiology of, 2332–2333, 2334f
  - etiology of, 2333–2335, 2335t
  - during exercise, 11
  - in heart failure, 1999
  - in hypertrophic cardiomyopathy, 2008, 2008t
  - precipitating factors in, 2333
  - prevention of
    - in general population, 2342
    - in high-risk populations, 2339–2340t, 2339–2341
    - in low-risk patients, 2341–2342, 2341f
    - risk stratification for, 2338–2339
  - risk factors for, 2333, 2334f, 2341f
- Sudden unexpected death in epilepsy, 3427
- Sudeck's point, 2587, 2587f
- Sudomotor function testing, 3542, 3544–3545f
- Sugars, malabsorption of, 309
- Suicide
  - assessment for risk of, 13
  - deaths from, 75t
  - in LGBTQIA+ people, 3178
  - physician-assisted, 89–90, 89t
  - in physicians, 42
  - in veterans, S8
  - worldwide, 3851

Sulbactam, 1175, 1176, 1183, 1296, 1297t. *See also* Ampicillin-sulbactam
Sulfacetamide, 224
Sulfadiazine
  actions of, 1166, 1182t
  adverse effects of, 1601, 1743t
  drug interactions of, 1743t
  in pregnancy and lactation, 1743t
  resistance to, 1182t, 1370
  for toxoplasmosis, 1138, 1157, 1803
  for toxoplasmosis prophylaxis, 1593t
  for Whipple's disease, 1370
Sulfadoxine, 1743t, 1770t, 1773
Sulfamethoxazole
  actions of, 1166, 1182t
  adverse effects of, 324, 417, 427, 799t, 1743t, 2356
  drug interactions of, 1743t, 3213
  in pregnancy and lactation, 1743t
  resistance to, 1182t
Sulfasalazine
  adverse effects of, 150, 787t, 2438, 2564, 2835, 2850t
  for IBD, 2564, 2564t
  monitoring during treatment with, 2850t
  for rheumatoid arthritis, 2849, 2850t
Sulfhemoglobin, 279
Sulfisoxazole, 1166, 1182t
Sulfocystinuria, 3369t
Sulfonamides
  actions of, 1166, 1182t, 1183f
  adverse effects of
    cross-sensitivity in, 427–428
    cutaneous
      erythroderma, 393
      exanthem, 402
      fixed drug eruption, 423
      hyperpigmentation, 399
      maculopapular eruptions, 422
      red lesions, 405
      vesicles/bullae, 401
    desensitization to, 428, 2818
    DIHS, 424. *See also* Drug-induced hypersensitivity syndrome (DIHS)
    eosinophilia, 462
    fever, 150
    frequent, 1743t
    in G6PD deficiency, 799t
    hepatotoxicity, 2667
    hypoglycemia, 3234
    nephritis, 2356, 2377
    nephrotoxicity, 2438
    neutropenia, 457
    optic neuropathy, 228
    phototoxicity, 434t
    in pregnancy, 1095
    rare, 1743t
    SJS or TEN, 401, A1. *See also* Stevens-Johnson syndrome (SJS); Toxic epidermal necrolysis (TEN)
    thrombocytopenia, 922, 923t
    vasculitis, 2907
  drug interactions of, 1743t
  indications for, 1174t
  metabolism of, 485t
  overdosage/poisoning with, 3714t
  in pregnancy and lactation, 1169t
  resistance to, 1182t, 1183f, 1185
Sulfonylureas
  actions of, 3211t, 3213
  adverse effects of, 434t, 3188, 3213
  contraindications to, 3211t
  drug interactions of, 3213
  in older adults, 3882t
  for type 2 DM, 3211t, 3213, 3215
Sulfur-containing shampoo, 388t
Sulfur dioxide exposure, 2239t
Sulfur granules, 1363, 1365, 1366f
Sulfur mustard, S5
Sulindac, 324
Sumatriptan, 299t, 300, 3467t, 3468, 3468t, 3472
Summer catarrh, 3834
Summer grippe, 1634
Summer-type pneumonitis, 2228t
Sunburn, 430, 430t, 512
SUNCT/SUNA, 115, 115t, 3471t, **3473**
Sundowning, 183, 185, 3952
Sun exposure, **430**
  melanoma and, 430, 432, 512, 512t, 598, 600
  nonmelanoma skin cancer and, 431–432, 512, 512t, 606
  skin effects of
    acute, 430–431, 430t
    immunologic, 432
    immunosuppressive, 432–433
    malignant, 431–432
    molecular targets of, 429–430
    nonmalignant, 431
  vitamin D synthesis and, 431
Sunflower cataracts, 3339
Sunitinib
  actions and targets of, 527, 533t, 535, 548, 570t, 573
  adverse effects of, 320, 570t, 573, 685, 753t, 754, 755
  drug interactions of, 591
  for GIST, 535, 729
  for HCC, 535
  interactions and issues, 570t
  for metastatic pheochromocytoma, 3077
  for NETs, 684t, 685, 735
  for RCC, 535, 573, 689, 690t
Sunscreen(s)
  active ingredients of, 435, 436t
  in chronic actinic dermatitis, 434
  ingredients in, 436t
  jellyfish-safe, 3723
  for melanoma prevention, 600
  for polymorphous light eruptions prophylaxis, 433
  for skin cancer prevention, 512
  SPF of, 435
  water resistance of, 435
suPAR (soluble urokinase plasminogen activator receptor), 2382
Superantigens, 134, 2777
Superficial venous thrombosis, 2158
Superior mesenteric artery occlusion, 111, 297
Superior orbital fissure, 3559
Superior sagittal sinus thrombosis, 994, 1142, A16
Superior sulcus tumor (Pancoast tumor), 133, 619, 626, 2943, 3615
Superior vena cava syndrome, **585,** 586f, 618t, 619, 1860, 2272
Supersaturation, 2448
Supine hypertension, 159
Suppurative thrombophlebitis, intracranial, 1141–1142
Suprachiasmatic nucleus (SCN), 220, 3943, 3944f, 3944t, 3945–3946
Supraclavicular lymph node, 477
Supraspinatus tendon, 2973–2974, 2973f
Supraventricular tachyarrhythmias, 1932
  atrial fibrillation. *See* Atrial fibrillation
  atrial flutter. *See* Atrial flutter
  atrial tachycardia. *See* Atrial tachycardia
  clinical features of, 1932–1933
  definition of, 1938–1939
  ECG in, 1932t, 1934f
  initial evaluation of, 1933–1935, 1933f, 1934f
  mechanisms of, 1932t, 1933
  in overdosage/poisoning, 3707
  paroxysmal. *See* Paroxysmal supraventricular tachycardia (PSVT)
  sinus tachycardia. *See* Sinus tachycardia
  in STEMI, 2124
  treatment of, 1933, 1934t
  vs. ventricular tachycardia, 1964, 1964f
Sural nerve biopsy, A14
Suramin
  actions of, 1752
  adverse effects of, 726t, 1746t, 1796, 3608t
  for human African trypanosomiasis, 1796, 1796t
  pharmacology of, 1752
Surfactant replacement therapy, 2303, 2303t
Surfer's myelopathy, 3565
Surgical ablation, for ventricular arrhythmias, 1954, 1959
Surgical myectomy, 2007
Surgical patient
  aortic stenosis in, 3909
  blood loss in, 807–808
  diabetes mellitus in, 3909
  infective endocarditis prophylaxis in, 3909
  mortality risk in, 3908t
  perioperative risk modification strategies
    ACE inhibitors, 3908
    α_2 agonists, 3908
    β-adrenergic antagonists, 3907–3908
    coronary revascularization, 3907
    oral antiplatelet agents, 3908
    pulmonary, 3908, 3909t
    SGLT-2 inhibitors, 3908
    statins, 3908
  preoperative evaluation of, **3905**
    for anesthesia, 3908
    cardiac risk assessment, 3905–3907, 3906–3907f, 3907t
    functional status, 3907t
    intermediate- to high-risk patient, 3905
    preoperative questionnaire, 3905, 3905t
    pulmonary risk assessment, 3908–3909, 3909t
  specialized nutritional support for, 2626
  VTE prophylaxis in, 2164, 2164t, 3909–3910, S13
Surgical-site infections
  enterococcal, 1218
  prevention of, 1148t, 1179, 1180t
  in solid organ transplant recipient, 1159t
  sternal, 1067–1068
Surgical ventricular restoration, 1999
Surrogates, for health care decisions, 71
Surufatinib, 684t, 685–686
Susceptibility testing, S11
Sustainable Development Goals (SDGs), 3844
Sutezolid, 1425
Sutimlimab, 803
Suvorexant, 216
Sv (sievert), S6
Swallowing, 293, 293f. *See also* Dysphagia
Swan-Ganz catheter. *See* Pulmonary artery catheter
Swan-neck deformity, 2840
S wave, 1868
Sweat/sweating
  excess, 3540
  initiation of, 3773
  in pheochromocytoma, 3074
  potassium loss in, 354
  sodium loss in, 346
  water loss in, 346, 2598–2599
Sweat testing, 2215, 2246–2247
Sweet syndrome
  in cancer patient, 583, 584f
  clinical features of, 144t, 148, 405, 583
  diagnosis of, 580, 583

- Sweet syndrome (*Cont.*):
  - differential diagnosis of, 406
  - disorders associated with, 405
  - epidemiology of, 144t
  - etiology of, 144t, 2937
  - in IBD, 2562t, 2563
  - MDS and, 816
  - skin manifestations of, 144t, 148, 405, 409, 584f, A1, A5
- Sweet taste sensation, 238
- Swimmer's ear (otitis externa), 254, 1309t, 1311
- Swimmer's itch, 1051, 1736, 1828
- Swine influenza viruses, 995, 1149, 1545. *See also* Influenza
- Swinging flashlight test, 220
- Swyer's syndrome (complete gonadal dysgenesis), 3098
- Sydenham's chorea, 2856, 3514
- Syk, 2804
- Symmetric drug-related intertriginous and flexural exanthema, 423
- Sympathectomy, 2178
- Sympathetic dyspraxia, 203
- Sympathetic postganglionic neuron, 93f
- Sympatholytic agents, 3705, 3711t
- Sympathomimetics, 3060t, 3710t
- Symplicity device, 2132
- Symporters (cotransporters), 2366
- Synaptobrevin (VAMP2), 1230
- Synaptophysin, 616, 733t
- Synaptotagmin, 2791t
- Syncope, **155**
  - after vaccination, 1008
  - in aortic stenosis, 2037
  - approach to the patient, 160–161
  - cardiac. *See* Cardiac syncope
  - definition of, 155
  - diagnosis of, 161
  - differential diagnosis of, 160–161, 3658
  - ECG in, 161
  - epidemiology of, 155–156
  - heat-related, 3774
  - high-risk features of, 156t
  - natural history of, 155–156
  - neurally mediated (reflex, vasovagal)
    - causes of, 155t
    - classification of, 156–158
    - clinical features of, 158
    - pacemaker for, 1920
    - pathophysiology of, 156, 157f
    - treatment of, 158
  - in orthostatic hypotension. *See* Orthostatic hypotension
  - pathophysiology of, 156
  - in psychiatric disorders, 161
  - vs. seizure, 160, 3418–3419, 3419t
  - tilt table testing in, 3542
  - in ventricular tachycardia, 1956, 1957
- Syndrome of apparent mineralocorticoid excess (SAME, 11β-hydroxysteroid dehydrogenase deficiency)
  - clinical features of, 355, 3061
  - diagnosis of, 356
  - genetic considerations in, 355, 3060t, 3061
  - treatment of, 356
- Syndrome of inappropriate antidiuresis (SIAD)
  - vs. cerebral salt wasting, 348
  - clinical features of, 3019
  - differential diagnosis of, 350, 3019–3020
  - etiology of, 348
  - in HIV infection, 1603
  - hyponatremia in, 348, 3019, S1
  - in lung cancer, 619
  - pathophysiology of, 348, 3019
  - subtypes of, 348
  - treatment of, 351, 619, 3020, S1
  - tumor-associated, 737t, 738, S1
- Syndrome X. *See* Metabolic syndrome
- Syndromic panels, 979
- Synkinesis, 3554
- Synovectomy, 2967, 2972
- Synovial effusion, 2944
- Synovial fluid analysis
  - in calcium oxalate deposition disease, 2962, 2962f
  - in calcium pyrophosphate deposition disease, 2961
  - interpretation of, 2945–2946, 2946f
  - in rheumatoid arthritis, 2847
- Synovial sarcoma, 728. *See also* Soft tissue sarcomas
- Synovitis, 2950, 2950f, 2972, 2972f
- Synthetic lethality, 537, 537f
- Synucleinopathies. *See* α-synuclein
- Syphilis, **1428**
  - CNS involvement in. *See* Neurosyphilis
  - congenital
    - clinical features of, 1058, 1432–1433
    - global considerations in, 1432
    - incidence of, 1097, 1429, 1432
    - oral manifestations of, 261, 262t, 1433
    - treatment of, 1435
  - diagnosis of, 976, 1433–1434
  - endemic. *See* Endemic syphilis (bejel)
  - epidemiology of, 1097, 1428, 1428f
  - etiology of, 1428–1429
  - follow-up of response to therapy, 1435–1436
  - global considerations in, 1429
  - in HIV infection, 1434, 1601–1602
  - immunity to, 1436
  - incidence of, 1097, 1429f
  - latent, 1431
  - pathogenesis of, 1429–1430
  - in pregnancy, 1429, 1432, 1434t, 1435
  - prevention of, 1429, 1434–1435
  - primary
    - clinical features of, 1108, 1108t, 1430
    - differential diagnosis of, 1430
    - natural course of, 1429–1430
    - skin manifestations of, 1052, 1430, 1430f, A1, A5
    - treatment of, 1434–1435, 1434t
  - reinfection, 1431
  - risk factors for, 1429
  - secondary, **1430**
    - articular manifestations of, 1058
    - cardiovascular manifestations of, 1432, 2043, 2166, 2171
    - epidemiology of, 1429, 1429f
    - hepatitis in, 1431
    - lymphadenopathy in, 477
    - natural course of, 1430
    - neurologic manifestations of. *See* Neurosyphilis
    - ocular involvement in, 1431
    - oral manifestations of, 262t
    - renal involvement in, 1431, 2416t, 2429
    - skin manifestations of, 140t, 146, 392, 402, 406, 1430
      - alopecia, 394
      - condylomata lata, 1431, A1
      - figurate lesions, 394
      - histopathology of, 1430
      - mucous patch, 1431, A1
      - on palms and soles, 1430, 1431f, A1, A5
      - papulosquamous eruption, 1051, 1431f, A1, A5
      - vs. pityriasis rosea, 146, 386t, 388
      - vs. psoriasis, 386
    - treatment of, 1434, 1434t
  - tertiary
    - clinical features of, 262t, 1051, 1432
    - treatment of, 1434t, 1435
  - transfusion-transmitted, 913t
  - transmission of, 1429
  - untreated, natural course of, 1429–1430
- Syringomyelia, 3567–3568, 3568f
- Syrinx cavities, 3568
- Systematic reviews, 27
- Systemic inflammatory response syndrome (SIRS), 2316, 2325f
- Systemic lupus erythematosus (SLE), **2827**
  - autoantibodies in, 2825t, 2831–2832, 2832t
  - classification of, 2827
  - clinical features of, 139t
    - antiphospholipid syndrome, 2830. *See also* Antiphospholipid syndrome (APS)
    - autoimmune hemolytic anemia, 802
    - cardiovascular, 2832t
      - AV conduction block, 1927
      - coronary heart disease, 2829–2830
      - pericarditis, 2083, 2830
      - sterile endocarditis, 150, 2830
    - cutaneous, 405, 2829, 2832t
      - alopecia, 394, 395t, 2830f
      - characteristics of, 139t, 146
      - discoid. *See* Discoid (chronic cutaneous) lupus erythematosus
      - distribution of, 405
      - malar erythema, A1, A5
      - malar rash, 2829, 2830f
      - panniculitis, 405
      - papulonodular red lesions, 405
      - papulosquamous lesions, A5
      - subacute, 2829, 2830f, A1
      - telangiectasias, 396
      - vasculitis, 403
    - fever of unknown origin, 150
    - gastrointestinal, 2831
    - hematologic, 810, 2830–2831, 2832t
    - hemoptysis, 275
    - lymphadenopathy, 476
    - musculoskeletal, 2829, 2830f, 2832t
    - neurologic, 1131, 1132t, 2829, 2832t, 3514, 3566, 3605
    - ocular, 2831, 2832t
    - oral, 261
    - pulmonary, 2830, 2832t
    - Raynaud's phenomenon, 2178, 2829
    - renal, 2359, 2439, 2829, 2832t. *See also* Lupus nephritis
    - splenomegaly, 480
    - systemic, 2829
  - diagnosis of, 2827, 2831–2833
  - drug-induced. *See* Lupus erythematosus, drug-induced
  - epidemiology of, 2827
  - genetic considerations in, 2828
  - hemophagocytic lymphohistiocytosis and, 469
  - inflammasome mutations in, 2762t
  - outcomes in, 2836
  - pathogenesis of, 2790t, 2791t, 2825t, 2828–2829, 2829f, 2832t
  - pregnancy and reproductive health issues in, 2836
  - treatment of, **2833,** 2833f
- Systemic lytic state, 954
- Systemic mastocytosis. *See* Mastocytosis, systemic
- Systemic mastocytosis associated with clonal hematologic non–mast cell lineage disease (SM-AHNMD), 2820
- Systemic necrotizing vasculitis, 2769
- Systemic sclerosis (scleroderma), **2859**
  - cancer associated with, 2871
  - classification of, 2860–2861t
  - clinical features of, 2860f, **2866,** 2867f
    - acro-osteolysis, 2868, 2868f
    - cardiac, 1927, 2083, 2866, 2871
    - cutaneous, **415,** 2860t–2861t, 2867–2868
      - alopecia, 394
      - calcinosis, 415, 2868, 2869f

      vs. dermatomyositis, 414
      digital necrosis, 2867, 2867f
      dystrophic calcification, 403
      facial, 415–416, 416f, A5
      fingers and hand, 415f, A5
      hyperpigmentation, 400
      hypopigmentation, 397
      limited vs. diffuse, 2861t
      nails, 2867, 2867f, 2868f
      pathology of, 2865, 2865f
      sclerodactyly, 2867f
      telangiectasias, 396, 396t, 415, 2868f
      vascular, 2867–2868, 2868f
    esophageal involvement/dysphagia, 293, 295–296, 2514
    gastrointestinal, 297, 2866, 2869–2870, 2870t, 2874
    initial presentation, 2866
    musculoskeletal, 2871
    neuropathy, 3605
    ocular, 2871
    pulmonary
      interstitial lung disease, 2260t, 2265, 2868–2869, 2869f, 2874
      pathology of, 2865–2866, 2865f
      pulmonary arterial hypertension, 2191, 2865f, 2866, 2869, 2874–2875
    Raynaud's phenomenon, 416, 2178, 2859, 2867, 2867f
    renal, 2446, 2866, 2875, A4. *See also* Scleroderma renal crisis
  definition of, 2859
  diagnosis of, 2860–2861t, 2872, 2872f
  diffuse cutaneous, 2861t
  drug-induced, 2940t
  environmental factors in, 2861–2862, 2862f
  epidemiology of, 2859, 2861
  genetic considerations in, 2861, 2862f
  inflammasome mutations in, 2762t
  limited cutaneous, 2861t
  localized, 2875
  natural history of, 2875
  pathogenesis of, 2772t, 2790t, 2862–2863f, 2862–2865, 2864t
  pathology of, 2865–2866, 2865f
  pregnancy in, 2871
  prognosis of, 2875
  sex differences in, 2861
  staging of, 2872
  treatment of, 416, **2872**
    antifibrotic therapy, 2874
    for gastrointestinal complications, 2874
    immunosuppressive agents, 2873–2874
    for interstitial lung disease, 2873f, 2874
    for musculoskeletal complications, 2875
    principles of, 2872–2873, 2873t
    for pulmonary arterial hypertension, 2874–2875
    for renal complications. *See* Scleroderma renal crisis
    skin care, 2875
    vascular therapy, 2874
Systemic vascular resistance, 289, 1906, 2292–2293
Systems biology, **3955**
  human disease classification and, 3957–3962, 3958–3960f, 3961t
  in medical practice, 6
  pathobiology applications of, 3956–3957
  in pharmacology, 483
  properties of biologic systems, 3955–3956
Systems neuroscience, 3654–3656
Systems theory, 42, 42f
Systolic click-murmur syndrome. *See* Mitral valve prolapse
Systolic failure. *See* Heart failure (HF)
Systolic murmurs. *See* Heart murmurs
SYT gene mutations, 728

T
T_3. *See* Triiodothyronine (T_3)
T_3-resin uptake test, 3026
T_4. *See* Thyroxine (T_4)
T790M germline mutation, 615
Tabes dorsalis, 1432, 3568–3569
Tabun (GA), 3712t, S5
TAC3 gene mutations, 3112t, 3127
TAC3R gene mutations, 3112t
Tacaiuma virus, 1659t
TACE (transarterial chemoembolization), 662, 664t, 665–666
Tăchéng tick virus, 1662t, 1663t
Tâche noire, 1458, 3730
Tachycardia
  atrioventricular nodal reentry, 1939–1940, 1939f
  in hypovolemia, 347
  preexcited, 1941, 1942f
  in pulmonary edema, 2331
  in shock, 2312
  supraventricular. *See* Supraventricular tachyarrhythmias
  syncope in. *See* Cardiac syncope
  ventricular. *See* Ventricular tachycardia
Tachycardia-bradycardia syndrome, 160, 1919, 1920, 1922f
Tachyphylaxis, 3285
Tachypnea, 189, 2312
TACO (transfusion-associated circulatory overload), 905, 908t, 912
Tacrine, 3712t
Tacrolimus
  actions of, 2408t, 2721
  adverse effects of, 2408, 2408t, 2721, 2916t
    hyperkalemia, 359
    hyperperfusion syndrome, 2350, 2350t
    microangiopathic hemolytic anemia, 925
    myopathy, 2940t
    nephrotoxicity, 2443, 2721
    osteoporosis, 3296
    seizures, 3414t
    skin cancer, 433
    thrombotic microangiopathy, 2445
  for atopic dermatitis, 383, 393t
  for autoimmune hepatitis, 2696
  drug interactions of, 1742t, 2408, 2721–2722, 2959
  genetic variations in response to, 492, E72t
  for IBD, 2565
  for immunosuppression
    after cardiac transplantation, 2029t
    after kidney transplantation, 2408, 2408t
    after liver transplantation, 2721–2722
  for inflammatory myopathies, 2916t
  metabolism of, 485t, 2721
  for myasthenia gravis, 3631
TACT (Trial to Assess Chelation Therapy), 3702
Tactile fremitus, 2200
Tadalafil
  actions of, 3155, 3156f
  adverse effects of, 3159, 3159t
  contraindications to, 3159, 3159t
  for erectile dysfunction, 503t, 3159, 3159t
  for high-altitude pulmonary edema, 3757
  for lower urinary tract symptoms, 704
  pharmacology of, 3159t
  for pulmonary arterial hypertension, 2194, 2195, 2197t
  for Raynaud's phenomenon, 2178
Taenia asiatica, 1832–1833
Taenia multiceps, 1839
Taenia saginata (beef tapeworm) infection, **1832,** S12
Taenia serialis, 1839
Taenia solium (pork tapeworm) infection
  clinical features of, 1736, 1738t, 1739t, 1833. *See also* Cysticercosis
  diagnosis of, 1833–1834, S12
  epidemiology of, 1133, 1833
  geographic distribution of, S12
  life cycle of parasite, 1833, S12
  pathogenesis of, 1833
  prevention of, 1835
  treatment of, 1834–1835
Tafamidis, 900, 2027
Tafasitamab, 559f, 863
Tafenoquine, 1741t, 1752, 1780
TAFI (thrombin-activatable fibrinolysis inhibitor), 472, 941
TAFI gene mutations, 3510t, 3511
Tagraxofusp, 572t, 574
Tag SNPs, 3794
Tahyña virus, 1659t
Tai chi, 3302, 3925t, 3928
Taï Forest virus disease, 1680, 1680–1683f
Takayasu's arteritis, 2166, 2171, **2903,** 2903t, 3431, 3447, A14
Takotsubo (stress-related) cardiomyopathy, 104, 2019, 2019f, 2331
Takotsubo syndrome, 2088, 3164
TAL1-TCTA gene mutations, 521t
Talaromyces marneffei infection (talaromycosis, penicilliosis), 1594t, 1611, 1726t, 1727
Talazoparib, 533t, 571t, 641, 703
Talc dust, 2238
Talimogene laherparepvec, 551, 559f, 561, 604, 604t
T-ALL (T-cell lineage ALL), 845, 846t, 849. *See also* Acute lymphoid leukemia (ALL)
Talquetamab, 534t, 890t, 892
Tamdy virus, 1662t
Tamm-Horsfall protein (uromodulin), 342, 343, 2377, 2435, 2440
Tamoxifen
  actions of, 526f, 536, 567
  adverse effects of
    depression, 84
    endometrial cancer, 567, 634, 714
    hepatotoxicity, 2668
    steatosis, 2703t
    thrombocytopenia, 923t
    thromboembolism, 567, 634
    tumor lysis syndrome, 593
    visual loss, 726t
  for breast cancer, 536, 567, 637
  for breast cancer in men, 639
  for breast cancer prevention, 40t, 513, 634, 3303
  drug interactions of, 492, 1422
  genetic variations in response to, 492, E72t
  for gynecomastia, 702, 3117
  metabolism of, 486
  for osteoporosis management/prevention, 3303
  for uterine cancer, 714
Tamsulosin, 704
Tanapox, 1518, 1518t, 1519
TandemHeart, 2132
Tandem mass spectrometry, 3977
Tandospirone, 299t, 302
Tangier disease (ABCA1 deficiency), 1860, 3241t, 3248, 3602
Tanning, 430, 432, 600, 606
Tanzavirus, 1662t
Tapentadol, 101, 3227
Tapeworms/tapeworm infections, 1736, 1739t. *See also* Cestodes/cestode infections
Tapia syndrome, 3559
TAP (transporters associated with antigen processing) proteins, 2776
Tap test, 3735
Tarantulas, 3734
TARC, 2767t
Tardive dyskinesia, 178, 299, 3516–3517, 3674

Tardive dystonia, 3517
Tardive syndromes, 3516–3517
TARDP gene mutations, 3486, 3522t
Target cells, 437, 440f, 446f, 447, 447t
Target of rapamycin complex 1 (TORC1), 3654
Target volume, in radiation therapy, 554
Tarlatamab, 630
Tartrate-resistant acid phosphatase (TRAP), 3315
Tarui disease (phosphofructokinase deficiency, type VII GSD), 796t, 3361t, **3364**
Taste buds, 237–238, 239f, 242
Taste disorders
  after traumatic brain injury, 3573–3574
  in COVID-19 disease, 1541
  diseases and conditions associated with, 242
  drug-related, 242
  etiology of, 242
  evaluation of, 242–243
  olfactory dysfunction and, 242
  radiation therapy-related, 614
  treatment of, 243–244
  in zinc deficiency, 2613
Taste pore, 238, 239f
Taste system, 237–238, 239f
TATA-binding protein, 3783t
Tataguine virus, 1659t
TAT/*tat,* 1552, 1552f, 1585
Tau gene mutations, 3521
Tau protein
  in AD, 3482
  in chronic traumatic encephalopathy, 3576
  in neurodegenerative diseases, 195t, 3531
  PET imaging of, 197–198
Tau protein antibodies, 3482
TAVI (transcatheter aortic valve implantation). *See* Transcatheter aortic valve implantation (TAVI)
Taxanes. *See also specific drugs*
  adverse effects of, 421, 508, 565, 575, 597, 638, 753t, 3608t
  characteristics of, 565
  drug interactions of, 591
  resistance to, 565
  for soft tissue sarcomas, 729
Taxonomy, 3831, 3831f
Tax/*tax,* 1551, 1552f, 1553
Tay-Sachs disease, **3355**
  adult, 3355, 3522t, 3524
  "cherry red" spot in eye in, A15
  deficiency in, 3352t
  genetic screening for, 3355, 3796
  infantile, 3355
  juvenile, 3352t, 3355
Tazarotene, 387
Tazemetostat, 534t, 570t, 573, 641, 864
TAZ gene mutations, 2004t
Tazobactam, 1183. *See also* Ceftolozane-tazobactam; Piperacillin-tazobactam
TB. *See* Tuberculosis (TB)
TBCE protein, 3286
TBG (thyroxine-binding globulin), 3023–3024, 3024t, 3026
TBI. *See* Traumatic brain injury (TBI)
TBP gene mutations, 3514
TBX gene mutations, 2185, 3286, 3783t
TCAB1, 3818
TCAs. *See* Tricyclic antidepressants (TCAs)
T cell(s), 974
  activation of, 974, 2785–2787, 2785f
  age-related reduction in, 3872
  antigen recognition by, 2774–2775, 2776f
  in autoimmune diseases, 2780, 2824
  autoreactive, in MS, 3582
  central memory, 2774
  co-stimulation of, 2784
  cytotoxic. *See* CD8+ T cells
  deficiencies of, 2794t, 2797–2798, 2797f, S9
    in cancer patient, 578, 578t
    cartilage hair hypoplasia, 457, 2800
    with DNA repair defects, 2799–2800, S9. *See also* Ataxia telangiectasia (AT)
    functional defects, 2799
    with hyper IgE. *See* Hyper-IgE (Job) syndrome
    in hyper-IgM syndrome, 2800
    in Omenn syndrome, 2799
    in severe combined immunodeficiency. *See* Severe combined immunodeficiency (SCID)
    in thymic defects, 2799
    in Wiskott-Aldrich syndrome, 458, 2553–2554t, 2800
  definition of, 2756
  development of, 2774, 2775f
  differentiation of, 760f, 860f, 2797–2798, 2797f, 2798f
  effector memory, 2774
  exhaustion of, 2785
  helper. *See* CD4+ T cells
  immune tolerance checkpoints in, 2789t
  inactivation by tumor cells, 549, 550f
  locations of, 2774
  malignancies of. *See* Lymphoid malignancies, T-cell
  MHC class-restricted, 2774
  negative selection, 2774
  NK. *See* Natural killer T (NK-T) cells
  positive selection, 2774
  in psoriasis, 386–387
  in response to *S. aureus,* 1198
  in sepsis/septic shock, 2319
  in systemic sclerosis, 2864
  in TB, 1385
  in vasculitis, 2894
  in viral infections, 1482, 1482f
T cell exhaustion, 2756, 2777
T-cell large granular lymphocytic leukemia, 874, 2841
T-cell lineage ALL (T-ALL), 845, 845t, 847t. *See also* Acute lymphoid leukemia (ALL)
T cell–mediated therapies, 560t, 561
T-cell prolymphocytic leukemia, 481, 874
T cell receptors (TCRs)
  in adverse drug reactions, 417t, 419
  for antigens, 2756, 2774–2775
  defective rearrangements of, 2799
  diversity of, 2774
  retargeting of, 3829, 3829f
  signaling through, 2776f
T cell superantigens, 2777
TCIRG1 gene mutations, 3315
TCL1 gene mutations, 874
Tdap vaccine. *See* Tetanus, diphtheria, acellular pertussis (Tdap) vaccine
TDF (testis-determining factor), 3791
TDP43 gene mutations, 3400, 3521
TDP43 protein, 195t, 3486, 3521, 3523. *See also* Limbic-predominant aging-related TDP-43 encephalopathy (LATE)
Td vaccine. *See* Tetanus-diphtheria (Td) vaccine
Teaching, physician engaged in, 8
Team-based care, 2–3
Teardrop cells, 440f, 447
Teardrop fracture, A16
Tebentafusp-tebn, 535t, 559f, 605
^{99m}Technetium, 1878t, 1879, S6
Technetium pyrophosphate (PYP) scan, 2023, 2024–2025f, A9
TECK (thymus expressed chemokine), 2768t
Teclistamab, 534t, 559f, 890t, 892
Tecovirimat, 1489, 1521, S4
Tedizolid
  actions of, 1166, 1182t, 1204
  adverse effects of, 1171t
  drug interactions of, 1172t
  for *E. faecalis* infections, 1220
  indications for, 1174t, 1178
  for MRSA infections, 1203t
  in pregnancy and lactation, 1169t
  resistance to, 1174t, 1182t, 1184, 1221
  for staphylococcal infections, 1204
Teduglutide, 2548
TEE. *See* Transesophageal echocardiography (TEE)
TEE (total energy expenditure), 2619
Teeth. *See also* Dental *entries*
  age-related changes in, 267
  development of, 260
  discolored, 261
  diseases of, **260, A3**
  extraction of, 265, 266
  Hutchinson's, 261
  loss of, 261, 267
  in osteogenesis imperfecta, 3324, 3326
  structure of, 260
TEF (thyrotrope embryonic factor), 2988, 2988t
Tegafur, 493, E72t
Tegaserod, 2575
Tegoprazan, 2526
Teicoplanin, 1166, 1182t
Telangiectasia(s), **396**. *See also* Ataxia telangiectasia (AT)
  in acne rosacea, 391
  capillary, 3457
  in cardiovascular disease, 1859
  definition of, 377t, 2180
  in dermatomyositis, 414
  etiology of, 396t
  hereditary hemorrhagic. *See* Osler-Weber-Rendu syndrome
  linear, 396
  mat, 396
  nailfold, 396
  in systemic sclerosis, 415, 2867–2868, 2868f
  treatment of, 2875
Telaprevir, 2687
Telavancin
  actions of, 1166, 1182t, 1204
  adverse effects of, 1171t, 1176, 1204
  for enterococcal infections, 1220
  indications for, 1173t, 1176
  for MRSA infections, 1203t
  in pregnancy and lactation, 1169t
  resistance to, 1182t, 1184
  for staphylococcal infections, 1204
Telbivudine, 1490t, 1491
  adverse effects of, 2678
  for chronic HBV infection, 2676t, 2678, 2680f, 2682t
  in pregnancy, 2683
  resistance to, 2682t
Telemedicine, 2945, 3898
Teletherapy, 554
Telitacicept, 2835
Telithromycin
  actions of, 1166, 1177, 1182t
  adverse effects of, 1177
  drug interactions of, 1177
  indications for, 1177
  for NTM infections, 1428
  resistance to, 1182t, 1184
Telmisartan, 3449
Telogen, 3138
Telogen effluvium, 394, 395t, 421
Telomerase, in cancer cells, 531, 3792
Telomere(s), 531, 3792, 3818, 3819f, 3870
Telomere disease, **3818**
  circadian disruption and, 3953
  clinical features of, 3818–3820, 3820–3821f
  dyskeratosis congenita. *See* Dyskeratosis congenita (Hoyeraal-Hreidarsson syndrome)

  genetic considerations in, 3818, 3820t, 3821–3822
  mechanisms of, 3818
  pathophysiology of, 3822f
  telomere length measurement in, 3820–3821, 3821f
  treatment of, 3822
Telotristat, 678t, 682
Temazepam
  for anxiety disorders, 3660t
  for insomnia, 216
  overdosage/poisoning with, 3713t
  pharmacology of, 3660t
  for sleep impairment at high altitudes, 3758
TE method, hepatic fibrosis assessment, 2632t
Temozolomide
  actions of, 563
  adverse effects of, 562t, 726
  for astrocytoma, 718
  for glioblastoma, 719
  interactions and issues, 562t
  for metastatic pheochromocytoma, 3077
  for NETs, 684t, 685
  for oligodendroglioma, 718
  for PCNSL, 721
  resistance to, 574
  for ZES, 2537
Temperature sense, 173, 174t
Temporal arteritis. *See* Giant cell arteritis (temporal arteritis)
Temporal bone, fracture of, 247
Temporal lobectomy, 3425
Temporal lobe epilepsy, 207, 3415
Temporomandibular joint disorder, 265
Temsirolimus
  actions and targets of, 533t, 535, 571t, 574
  adverse effects of, 571t, 574, 595
  drug interactions of, 591
  for endometrial cancer, 715
  for mantle cell lymphoma, 866
TEN. *See* Toxic epidermal necrolysis (TEN)
Tenapanor, 314, 2391, 2576, 2577t
Tenatoprazole, 2526
Tendinitis, 2940t, 2944, 2968, 2973–2974
Tendinopathies, drug-induced, 2974
Tendinous xanthomas, 2630, A15
Tenecteplase (TNK)
  actions of, 955
  for acute limb ischemia, 2175
  contraindications to, 2120
  domain structure of, 955, 955f
  indications for, 956
  for ischemic stroke, 3441, 3441t
  for STEMI, 2120
Tenofovir alafenamide (TAF)
  adverse effects of, 1616
  for chronic HBV infection, 1488t, 2676t, 2679–2680
  in combination formulations, 1621t
  for HIV infection, 1491, 1617t, 1621
  structure of, 1619f
Tenofovir disoproxil fumarate (TDF)
  actions of, 1491
  adverse effects of, 1491, 1601, 1616, 1617t, 2377
  for chronic HBV infection, 2679–2680
    dosage for, 1490t, 1491
    in HIV infection, 2683
    vs. other agents, 2676t, 2680f, 2683
    with PEG IFN-α2a, 1491
    therapeutic monitoring of, 1490t, 2683
  in combination formulations, 1621t
  drug interactions of, 1402
  for HIV infection, 1491, 1617t, 1621, 3839
  for HIV pre-exposure prophylaxis, 1621
  resistance to, 1491, 1622f, 2682t
  structure of, 1619f
Tenosynovitis, 2919, 2974
Tension pneumothorax, 104, 2271, 2311
Tension-type headache, 113t, 115, 3463t, 3471
Teplizumab, 3200, 3216
Tepotinib, 533t, 569t, 628, 628f
Teprenone, 2526
Teprotumumab, 2791t, 3037
Teratoma
  midline, A16
  ovarian, 712, 743t, 744, 747, 748f, 3037, 3792
  parasellar, 3003
  pineal, 721
  testicular, 705
Terazosin, 2178, 3570, 3588
Terbinafine
  adverse effects of, 242, 420, 2940t
  for onychomycosis, 389, 1692, 1724, 1724t
  for ringworm, 1692
  for sporotrichosis, 1724t
  for tinea, 395t, 1724, 1724t
Terbutaline, 3710t
TERC, 3818
TERC gene mutations, 810, 3819, 3820t
TERF2 gene mutations, 3820t
Teriflunomide, 3585, 3587t
Teriparatide, 364, 3145, 3304, 3307–3308, 3307–3308f, 3311
Terizidone, 1394t
Terlipressin, 317, 2383
Terminal extubation, 88
Terminal hair, 3137
Terminally ill patient, **76**
  advance care planning for, 78–79, 78t
  assessment of, 76
  with cancer, 505
  clinical courses of, 91t
  communicating bad news to, 76–77, 77t
  continuous goal assessment, 77–78
  delirium in, 185
  existential needs of, 87
  managing last stages for
    care during last hours, 91–92, 91t
    euthanasia and physician-assisted suicide in, 78t, 89–90
    futile care in, 89
    palliative care services in, 87–88
    withholding/withdrawal of care in, 88–89
  physical symptoms of, 79–80, 79t
    anorexia, 91t
    apnea, 91t
    constipation, 82, 82t
    dehydration, 91t
    delirium, 91t
    dysphagia, 91t
    dyspnea, 83, 83t
    fatigue, 83–84, 91t
    incontinence, 91t
    nausea and vomiting, 82–83
    pain, 80–82, 80f
  psychological symptoms of, 79t, 84–86, 86t
  social needs of, 86–87
Terminal weaning, 88
Terrorism
  chemical, 3712t, **S5,** S8
  microbial/biological, 1149, **S4,** S7
  radiation, **S6**
TERT, 3818, 3819f
TERT gene, 3818
TERT gene mutations
  in bone marrow failure, 810
  in glioblastoma, 719
  in HCC, 660, 661t
  in interstitial lung disease, 2261, 2267
  in melanoma, 599, 601
  in telomerase disease, 3819, 3820t
  in thyroid cancer, 3046
Tertiary prevention, 8
Testes/testicles, **3102**
  biopsy of, 3108
  development of, 3094, 3102, 3102f
  disorders of. *See also* Male reproductive system disorders
    acquired defects, 3114–3115
    adrenal rest tumors, 3072, 3072f
    alcohol use disorder and, 3677
    cancer. *See* Testicular cancer
    in Carney complex, 3088
    cryptorchidism, 3100, 3114
    in disorders of sex development. *See* 46,XY disorders of sex development
    in Klinefelter's syndrome, 3114. *See also* Klinefelter's syndrome
    non-germ cell tumors, 709
    swelling/mass, 705
    torsion, 1099, 3115
  dysgenetic, 3098, 3099t
  as immune-privileged site, 2826
  physical examination of, 3107
  referred pain from, 111
  regulation of function of
    androgen in, 3103–3106, 3103f, 3105f
    hypothalamic-pituitary-testis axis in, 3103–3104, 3103f
    seminiferous tubules in, 3106–3107
  structure of, 3102
Testicular cancer, **704**
  clinical features of, 705
  diagnosis of, 705
  epidemiology of, 704–705
  genetic considerations in, 705
  global considerations in, 704
  incidence of, 704
  non-germ cell tumors, 709
  nonseminoma, 706
  paraneoplastic syndromes in, 737t
  pathology of, 705
  risk factors for, 704–705
  seminoma, 706
  staging of, 705–706, 707–708f
  treatment of
    chemotherapy, 706
    HCT, 920
    initial, 705
    late effects of, 709
    postchemotherapy surgery, 706, 708
    in relapsed disease, 708–709
    stage-based, 706–708, 707–708f
  tumor markers in, 705
Testicular dysgenesis, 3098, 3099t
Testicular feminization syndrome. *See* Androgen insensitivity syndrome (AIS)
Testis-determining factor *(TDF),* 3791
Testosterone
  age-related changes in levels of, 3169–3171, 3169f
  assays of, 3107–3108, 3118
  bioavailable, 3108
  in hirsutism, 3138–3139
  in men with prostate cancer, 3175
  outcomes in older men and levels of, 3169, 3169t
  synthesis of, 3051f, 3104f
  total, tests of, 3107–3108
  transport and metabolism of, 3103–3104, 3104f
  unbound, measurement of, 3108
Testosterone cypionate, 3119, 3120t
Testosterone deficiency
  age-related, 3169
  approach to the patient, 3118–3119
  diagnosis of, 3118, 3118f
  gynecomastia and, 3115–3116
  prevalence of, 3117
  in prostate cancer, 3175
  treatment of
    gonadotropins, 3119
    testosterone therapy. *See* Testosterone therapy

Testosterone enanthate, 3119, 3120t
Testosterone-lowering drugs. *See* Androgen deprivation therapy (ADT)
Testosterone therapy
  abuse of. *See* Anabolic-androgenic steroid (AAS) abuse
  for adult hypopituitarism, 2996t
  adverse effects of, 421, 3117, 3121–3122t, 3160, 3249
    carcinogenicity, 512t
    cholestasis, 2667t
    cutaneous, 420
    hirsutism, 421
    in older men, 3171, 3171t
  for aplastic anemia, 813
  contraindications to, 3121, 3121t
  for delayed puberty, 3111
  efficacy and safety of, 3117
  for erectile dysfunction, 3159–3160
  for female sexual dysfunction, 3161–3162
  formulations of, 3119–3121, 3120t
  for gonadotropin deficiency, 2998
  for hypogonadism, 2998
  indications for, 3117
  in men with history of prostate cancer, 3175
  monitoring of, 3121–3122, 3121t, 3160
  novel formulations of, 3120–3121
  in older men, 3170–3171, 3170f, 3170t, 3172f
  pharmacology of, 3120t
  regimens for, 3120
  for testosterone deficiency, 3117, 3119–3120
Testosterone undecanoate, 3119, 3120t
Testotoxicosis, 3109
TET2 gene mutations, 761, 816, 825, 867, 2186
Tetanus, **1230**
  approach to the patient, 1230–1231
  clinical features of, 1231, 1231t
  complications of, 1231, 1232
  definition of, 1230
  diagnosis of, 1231
  differential diagnosis of, 1231
  epidemiology of, 1230
  etiology of, 1230
  global considerations in, 1230, 1232
  incidence pre- and postvaccine, U.S., 1002t, 1230
  neonatal, 1230, 1232
  pathogenesis of, 1230
  in pregnancy, 1230
  prevention of, 40t, 1004t, 1232
  prognosis of, 1232, 1232f
  prophylaxis, for bite-wounds, 1146
  severity classification of, 1231t, 1232t
  treatment of, 1231–1232
Tetanus, diphtheria, acellular pertussis (Tdap) vaccine
  contraindications to, 1006t
  precautions for, 1006t
  recommended administration of
    in adults, 1004t, 1225, 1232, 1280
    in children, 1225, 1232, 1280
    in health care workers, 1153t
    in international travelers, 1014–1015
    in pregnancy, 1004t, 1005, 1232, 1280, 3904
Tetanus-diphtheria (Td) vaccine, 40t, 1004t, 1006t, 1225
Tetanus Severity Score, 1232t
Tetany, in malabsorption disorders, 2549t
TET enzymes, 3936
Tethered cord syndrome, 3569
Tetherin, 1558
Tetrabenazine, 3497, 3512, 3513–3514, 3516–3517
Tetracycline(s)
  for acne rosacea, 391
  for acne vulgaris, 391
  for actinomycosis, 1365, 1366t
  actions of, 1166, 1177, 1182t, 1183f
  adverse effects of, 1171t, 1177, 2528
    cutaneous, 399, 401, 420, 1746t
    fixed drug eruption, 423
    folate deficiency, 787t
    frequent, 1746t, 1771t
    lupus syndrome, 2940t
    nail disorders, 421
    phototoxicity, 434t
    rare, 1746t
    renal failure, 1771t
    steatosis, 2703t
    teeth discoloration, 261, 420, 1346
  for balantidiasis, 1808
  for cholera, 1329
  contraindications to, 1447
  for *C. trachomatis* infection, 1474
  for donovanosis, 1358t
  drug interactions of, 489t, 1172t, 1742t, 1746t
  for ehrlichioses, 1461
  for endemic treponematoses, 1439
  in *H. pylori* eradication, 1302t, 2527t, 2528, 2530t
  indications for, 1177
  for malaria, 1770t, 1771t
  for meibomitis, 224
  for *Mycoplasma* infections, 1466, 1466t
  for parasitic infections, 1746t
  pharmacology of, 1771t
  for plague, 1346, 1346t
  for plague prophylaxis, 1347t
  in pregnancy and lactation, 1169t, 1746t
  for psittacosis, 1476
  for relapsing fever, 1447, 1447f
  resistance to
    in *Enterobacterales,* 1283, 1283t
    in *H. pylori,* 2528
    mechanisms of, 1182t, 1183f, 1184
    in *N. gonorrhoeae,* 1100f, 1255
    in *V. cholerae,* 1329
  for syphilis, 1434t, 1435
  for tropical sprue, 2546
  for tularemia, 1340
Tetrahydrobiopterin, 3370
Tetrahydrocannabinol (THC), 299t, 575, 3686. *See also* Cannabis/marijuana use
Tetrahydrofolate, 782, 783f
Tetrahydrozoline, 3711t
Tetralogy of Fallot
  clinical features of, 451
  epidemiology of, 2073
  hypoxia in, 277
  murmur in, 286f, 288
  pathophysiology of, 2073, 2073f
  primary repair of, 2073, 2073f
  repaired, sequelae of, 2073–2074, 2074t
    arrhythmias, 1967, 2074
    interventions for, 2074
    in pregnancy, 3900
    pulmonic regurgitation, 2064, 2073–2074
    sudden cardiac death, 2074, 2335, 2335t
Tetraparvovirus, 1525
Tetrathiomolybdate, 3340, 3518
TF (tissue factor), 470, 470f
TFE3, 687, 688t
TFEB, 688t
T follicular helper cells (Tfh), 2407, 2407f, 2756. *See also* CD4+ T cells
Tg. *See* Thyroglobulin (Tg)
TGFB2/3 gene mutations, 3331, 3332t
TGFRB1 gene mutations, 3331, 3332t
TGFRB2 gene mutations, 713, 3331, 3332t
TGF-α (transforming growth factor-α), 546, 2516, 2516f, 2539
TGF-β. *See* Transforming growth factor-β (TGF-β)
T_h1 T cells
  characteristics of, 2756
  in MS, 3582
  in rheumatoid arthritis, 2846
T_h2 T cells, 2756
T_h9 T cells, 2756
T_h13 T cells, 2756
T_h17 T cells
  in autoimmune diseases, 2784
  characteristics of, 2756
  in rheumatoid arthritis, 2846
Thalamic astasia, 181
Thalamic hemorrhage, 192, 3454. *See also* Intracerebral hemorrhage
Thalamus, 176
Thalassemia syndromes, 770t, **776**
  α thalassemia, **778,** 778t
  α thalassemia trait, 778t
  α thalassemia with myelodysplasia, 778t
  arthropathy in, 2968
  β thalassemia, **776,** 2968, 3786f, 3787
    classification of, 776, 777t
    clinical features of, 777t
    complications of, 777, 778t
    diagnosis of, 777, 777f, 777t
    epidemiology of, 776
    global considerations in, 776
    pathophysiology of, 776
    screening for, 777
    treatment of, 762, 777–778, 3824t, 3826
  β thalassemia intermedia, 777, 777f, 777t, 2968
  β thalassemia minor, 2968
  β thalassemia trait, 777, 777t
  differential diagnosis of, 767t
  genetic considerations in, 776
  HCT for, 915, 919, 919t
  jaundice in, 323
  malaria resistance and, 776, 1762
  peripheral blood smear in, 437, 440f, 446f
  splenomegaly in, 480
Thalidomide
  actions and targets of, 559f, 561
  adverse effects of, 509, 561, 726, 753t, 887
    neuropathy, 3609t, 3610–3611
    in pregnancy, 1412
  for leprosy reactions, 1412
  for multiple myeloma, 561, 890t, 891–892
  for POEMS syndrome, 3093
  for primary myelofibrosis, 823
  for SLE, 2834
  for vascular ectasias, 318
Thallium-201, 1878t, 1879
Thallium poisoning, 228, 3610t, 3611, 3702
THAP1 gene mutations, 3510, 3510t
Thawing, for frostbite, 3772, 3772t
THBR (thyroid hormone binding ratio), 3026
THC (tetrahydrocannabinol), 299t, 575, 3686. *See also* Cannabis/marijuana use
Thelarche, 3130
Theophylline
  adverse effects of, 354, 376, 2224, 3414t
  drug interactions of, 489t
    ciprofloxacin, 1745t
    fluoroquinolones, 1172t
    PPIs, 2525
    thiabendazole, 1742t, 1752
  overdosage/poisoning with, 356, 3710t
  for SA node dysfunction, 1920
Therapeutic diet, 2603
Therapeutic hypothermia, 2338
Therapeutic index/ratio/window, 483, 483f, 555
Thermal ablation, in chronic venous disease, 2182
Thermoregulation, 134, 151t, 212, 3769, 3772–3773
THF starvation, 782
Thiabendazole, 1742t, 1752–1753

- Thiamine (vitamin B_1), **2605**
  - for alcohol-induced acidosis, S1
  - deficiency of. *See* Thiamine (vitamin B_1) deficiency
  - food sources of, 2605, 3612
  - functions of, 2605, 2606f
  - recommended intake of, 2600t
  - structure of, 2606f
  - supplements, 371, 2607, 3679
  - toxicity of, 2608
- Thiamine (vitamin B_1) deficiency. *See also* Beriberi
  - in alcoholism, 2606
  - clinical features of, 2605t, 2606–2607. *See also* Wernicke's disease/encephalopathy
    - acid-base disorders, 369
    - ataxia, 3533
    - cardiomyopathy, 2018
    - delirium, 184
    - dementia, 3483
    - neuropathy, 3612
    - optic neuropathy, 228
  - diagnosis of, 3612
  - in refeeding syndrome, 2626
  - risk factors for, 2605t, 2606–2607
  - treatment of, 2607, 3612
- Thiazide diuretics
  - actions of, 2133, 2367f
  - adverse effects of, 2143, 3157
    - acid-base disorders, 374
    - cutaneous, 388, 392, 400, 420
    - hypercalcemia, 364, 3282
    - hypokalemia, 355, 356
    - hyponatremia, 347, 350
    - nephritis, 2356
    - phototoxicity, 434t
    - in pregnancy, 458
    - vasculitis, 2907
  - for AVP resistance, 3019
  - for CKD, 2389
  - drug interactions of, 1743t, 1745t
  - for heart failure, 1990–1991
  - for hyperkalemia, 361
  - for hypertension, 2144t
  - for hypoparathyroidism, 3287
  - for nephrogenic diabetes insipidus, 353
  - for nephrolithiasis, 2453
  - for pseudohypoaldosteronism, 361
- Thiazolidinediones
  - actions of, 3211t, 3214
  - adverse effects of, 282t, 3188, 3211t, 3296
  - contraindications to, 3211t, 3214
  - for MASH, 2706
  - for metabolic syndrome, 3258
  - in older adults, 3882t
  - for type 2 DM, 3211t, 3214, 3215–3216
- Thick-filament myopathy, 2352
- Thienopyridines, 943–944. *See also* Clopidogrel; Prasugrel; Ticlopidine
- Thiethylperazine, 299t, 575
- Thigh thrust (posterior shear test), 121t
- Thimerosal, 384
- Thin basement membrane disease, 343, 2415t, 2426, A4
- 6-Thioguanine, 562t, 565, E72t
- Thiopental, 3426f
- Thiopurine *S*-methyltransferase (TPMT), 485t, 493, 2408, 2835, 2895, E72t
- Thiopurine *S*-methyltransferase (TPMT) deficiency, 3805
- Thioridazine, 86t, 420, 3673, 3673t, 3711t
- Thiotepa, 591, 726t
- Thiothixene, 3673, 3673t
- Third disease. *See* Rubella
- Third nerve palsy, 50, 188f, 3458
- Third spacing, 346
- Third window, inner ear, 245, 249
- Thirst, 352, 2371, 2371f, 3016, 3016f
- Thogotoviruses, 1662t
- Thomsen disease, 3647
- Thoracentesis (pleurocentesis), 2208
- Thoracic adenopathy, 477
- Thoracic aortic aneurysm. *See also* Aortic aneurysm
  - clinical features of, 2166–2167
  - diagnosis of, 2167f
  - in giant cell arteritis, 2903, A14
  - heritable, 3331–3333, 3332t. *See also* Marfan's syndrome
  - treatment of, 2167
- Thoracic fracture, 3292. *See also* Vertebral fractures
- Thoracic lavage, as rewarming strategy, 3770
- Thoracic meningioma, 3563f
- Thoracic outlet syndromes, 133, 2175–2176
- Thoracic spinal lesions, 3560
- Thoracoscopy, 2208, 2208f, 2290
- Thoracotomy, 2208
- Thorium-232, S6
- 3TC. *See* Lamivudine (3TC)
- Threonine kinase gene mutations, 399
- Threonine metabolism disorders, 3370t
- Throat swab, 258, S11
- Thrombectomy, percutaneous mechanical, 2175, 3441–3442, A11, A11v
- Thrombin, 470, 470f, 928, 936, 942, 942f
- Thrombin-activatable fibrinolysis inhibitor (TAFI), 472, 941
- Thrombin time, 475, 475t, 928t, 929f
- Thromboangiitis obliterans, 2174–2175, 2178
- Thrombocytopenia, **921**
  - approach to the patient, 921, 922f
  - bleeding in, 473
  - chemotherapy-related, 575
  - drug-induced, 489, 922–923, 923t
  - familial, with propensity to acute AML, 3783t
  - in filovirid infections, 1685, 1686
  - heparin-induced. *See* Heparin-induced thrombocytopenia
  - in HIV infection, 922, 925, 1605
  - immune. *See* Immune thrombocytopenic purpura
  - infection-induced, 921–922
  - inherited, 924
  - in leptospirosis, 1442, 1443
  - peripheral blood smear in, 441f, 921, 922f
  - in pregnancy, 3902
  - refractory. *See* Myelodysplastic syndromes (MDS)
  - skin manifestations of, 407
  - in SLE, 2831
  - treatment of, 575
- Thrombocytopenia with absent radii, 809t
- Thrombocytopenic purpura, 1301, 2790t. *See also* Thrombotic thrombocytopenic purpura (TTP)
- Thrombocytosis
  - erythromelalgia in, 2179
  - essential, 819, **823**
  - etiology of, 823t, 925–926
  - paraneoplastic, 740t, 741
- Thromboembolectomy, 2175
- Thromboembolic disease
  - acute limb ischemia in, 2175
  - in COVID-19 disease, 1541
  - in IBD, 2562t, 2563
  - in mitral stenosis, 2049
  - pulmonary hypertension in. *See* Chronic thromboembolic pulmonary hypertension (CTEPH)
  - renal, 2153, 2153f
  - in STEMI, 2125
- Thromboendarterectomy, 2164, 2174
- Thrombogenesis, 942, 943f
- Thrombolysis in Myocardial Infarction (TIMI) trial, 2107f, 2108, 2119
- Thrombolytic therapy, 2175, 3441, 3441t
- Thrombomodulin, 471f
- Thrombophilia, 475
- Thrombophlebitis, 1141–1142, 1231–1232. *See also* Venous thromboembolism (VTE)
- Thromboplastin, 474
- Thrombopoietin, 760f, 823–824, 921, 937
- Thrombopoietin receptor agonists, 913
- Thrombosis, **472**
  - approach to the patient, 472
  - arterial. *See* Arterial thrombosis
  - coronary artery, A10
  - dural sinus, 3452
  - endothelium in, 1845
  - epidemiology of, 937
  - history in, 473
  - in HIT, 923
  - in infection, A10
  - in inflammation, 938
  - laboratory evaluation of, 474–475
  - in liver disease, 935, 935t, 936f
  - in meningococcal disease, 1247–1248
  - pathophysiology of, 937, 937f
  - in polycythemia, 451
  - renal vein, 2447
  - risk factors for, 473, 473t, 474f
  - venous. *See* Venous thromboembolism (VTE)
- Thrombotic microangiopathy, **2444**
  - in antiphospholipid antibody syndrome, 2446
  - cancer-related, 2446
  - convergent phenotypes in, 31f
  - drug-related, 2445
  - etiology of, 2361f
  - HCT-related, 916, 2445–2446, 2445t
  - in HELLP syndrome, 2446–2447
  - in hemolytic-uremic syndrome. *See* Hemolytic-uremic syndrome (HUS)
  - HIV-related, 2446
  - pathophysiology of, 2444
  - in POEMS syndrome, 2447
  - primary, 2428t
  - secondary, 2428t
  - in sickle cell disease. *See* Sickle cell disease
  - in systemic sclerosis. *See* Systemic sclerosis (scleroderma)
  - in thrombotic thrombocytopenic purpura. *See* Thrombotic thrombocytopenic purpura
- Thrombotic thrombocytopenic purpura (TTP), **924**
  - AKI in, 2379t
  - congenital, 2445
  - diagnosis of, 925, 934
  - drug-induced, 2445
  - as drug reaction, 595
  - epidemiology of, 145t
  - genetic considerations in, 2428
  - vs. hemolytic-uremic syndrome, 2445
  - idiopathic, 925, 2445
  - inherited, 925
  - pathology and pathogenesis of, 925, 925f, 1845, 2428, 2445
  - peripheral blood smear in, 924
  - renal involvement in, 2428
  - risk factors for, 2427
  - skin manifestations of, 145t, 148, 407
  - treatment of, 925, 2428, 2445
- Thromboxane A_2, 942, 943f, 2113
- Thrush
  - clinical features of, 257, 264t, 390, 1708, A3
  - in HIV infection, 1598, 1599f
  - treatment of, 264t, 1710t
- "Thumb sign," 259

Thunderclap headache, 3458, 3474
Thymectomy, 3629
Thymidine, 3778
Thymidylate synthase, 629, 658
Thymoma, **631**
  in autoimmune diseases, 3092
  clinical features of, 631, 810
  diagnosis of, 2271, 3628
  hypogammaglobulinemia in, 2802
  in myasthenia gravis, 3626, 3629
  paraneoplastic syndromes in, 411, 737t, 743t
  pemphigus and, 411
  pure red cell aplasia and, 813
  staging of, 631–632, 632t
  treatment of, 632
Thymus
  atrophy in aging, 2793
  in HIV infection, 1578
  immune tolerance checkpoints in, 2789t
  in myasthenia gravis, 3626, 3628, 3630t
  primary defects of, 2799, S9
  T-cell production in, 2774
  tumors of. *See* Thymoma
Thymus- and activation-regulated chemokine (TARC), 2767t
Thymus expressed chemokine (TECK), 2768t
Thyroglobulin (Tg)
  in autoimmune thyroiditis, 2790t
  in carcinoma of unknown primary, 732
  in congenital hypothyroidism, 3021, 3028t
  in thyroid cancer follow-up, 3047
  in thyrotoxicosis, 3026
Thyroglobulin (Tg) antibodies, 3029
Thyroid acropachy, 2971, 3034f
Thyroid and eye muscle shared protein, 2790t
Thyroid-associated ophthalmopathy. *See* Graves' (thyroid-associated) ophthalmopathy
Thyroid cancer, **3045**
  anaplastic, 3048
  carcinoma of unknown primary and, 733t
  classification of, 3045
  deaths from, 497t
  diagnosis of, 2979t, 3026–3027, 3043f
  follicular, 3046–3047
  genetic considerations in, 520t, 3045–3046
  incidence of, 497t, 498f, 2979t, 3045
  medullary. *See* Medullary thyroid carcinoma (MTC)
  papillary, 3046, 3046f
  pathogenesis of, 3045–3046
  in patients with thyroid nodule, 3045
  radiation therapy-related, 755
  risk factors for, 3045, 3045t
  staging of, 3045
  survival rates, 3046f
  treatment of, 3046–3047
Thyroid cysts, 3044
Thyroid dermopathy, 3034f, 3037
Thyroidectomy
  for diffuse nontoxic goiter, 3041
  for Graves' disease, 3036
  prophylactic, in *RET* mutations, 3085, 3085t
  for thyroid cancer, 3047, 3085
Thyroid eye disease. *See* Graves' (thyroid-associated) ophthalmopathy
Thyroid gland, **3020**
  amiodarone effects on, 3039–3040
  anatomy and development of, 3020–3021
  disorders of
    cancer. *See* Thyroid cancer
    in HIV infection, 1603
    hyperthyroidism. *See* Hyperthyroidism
    hypopigmentation in, 396
    hypothyroidism. *See* Hypothyroidism
    lymphoma, 3048
    nodular. *See* Thyroid nodular disease
    radiation-induced, 755
    thyroiditis. *See* Thyroiditis
    thyrotoxicosis. *See* Thyrotoxicosis
  hormones of. *See* Thyroid hormone(s)
  laboratory studies of, **3025**
  pain in, differential diagnosis of, 3037
  physical examination of, 3025
  regulation of, 3021
Thyroid hormone(s). *See also* Thyroxine (T_4); Triiodothyronine (T_3)
  actions of, 2986, 3024–3025, 3025f
  half-lives of, 2982
  laboratory evaluation of, 3025–3026
  receptor binding of, 3024–3025, 3025f
  resistance to, 3024t, 3025, 3786
  structure of, 3021f
  synthesis of, 3021–3022, 3021f
  transport and metabolism of, 3024–3025, 3024t
  unbound, 3026
Thyroid hormone binding proteins, 3023–3024
Thyroid hormone binding ratio (THBR), 3026
Thyroid hormone receptor (TR), 2985, 3024–3025, 3025f
Thyroiditis, **3037**
  acute, 3037–3038, 3038t
  atrophic, 3028
  autoimmune. *See* Hashimoto's thyroiditis (autoimmune hypothyroidism)
  chronic, 3038t, 3039
  destructive, 3037
  drug-induced, 3039
  etiology of, 3038t
  postpartum, 3038–3039
  radiation therapy-related, 755
  Riedel's, 2931t, 3039
  silent (painless), 3038–3039
  subacute, 65, 3038, 3038f, 3038t
Thyroid nodular disease, **3040**
  approach to the patient, 3048–3050, 3049f
  benign, 3044
  classification of, 3044t, 3048, 3049t
  evaluation of, 2979t, 3026–3027, 3042f, 3043f
  goiter, 3040–3042, A15
  hyperfunctioning solitary nodule, 3043–3044, 3044f
  pathophysiology of, 3040
  prevalence of, 2979t
  thyroid cancer and, 3045. *See also* Thyroid cancer
Thyroid peroxidase (TPO), 2790t, 3021, 3023, 3028t
Thyroid peroxidase (TPO) antibodies, 3026, 3029, 3033
Thyroid response elements (TREs), 3025, 3025f
Thyroid scintigraphy, 3026–3027
Thyroid-stimulating hormone (TSH), **2991**
  actions of, 2991
  assessment of, 2987
  deficiency of, 2996t
  ectopic production of, 737t
  expression and regulation of, 2988t
  laboratory evaluation of, 2995t, 3026
  pituitary adenoma secretion of, 3013–3014, 3037
  in pregnancy, 3023
  secretion of, 2991
  suppression therapy, 3047
  synthesis of, 2991
Thyroid-stimulating hormone receptor (TSH-R)
  in congenital hypothyroidism, 3028t
  in hyperfunctioning thyroid solitary nodule, 3043, 3044f
  mutations of, 2983, 2985t, 3028t, 3043, 3044f
  in thyroid-associated ophthalmopathy, 3033
  in thyroid development, 3021
Thyroid-stimulating hormone receptor (TSH-R) antibody (TRAb), 3026, 3029
Thyroid-stimulating immunoglobulin (TSI), 3026, 3033
Thyroid storm (thyrotoxic crisis), 3037
Thyroid transcription factor-1 (TTF-1)
  in carcinoma of unknown primary, 733
  in congenital hypothyroidism, 3021, 3028t
  in lung cancer, 616, 733
Thyrotoxicosis, **3032**. *See also* Graves' disease
  amiodarone-induced, 3039–3040
  apathetic, 3033
  clinical features of, 3033t, 3034f
  definition of, 3032
  etiology of, 3033t, 3037
  fetal/neonatal, 3036–3037
  hypoxia in, 277
  palpitations in, 291
  subclinical, 3041
Thyrotoxicosis factitia, 3026, 3037
Thyrotoxic periodic paralysis (TPP). *See* Hypokalemic periodic paralysis (HypoKPP)
Thyrotrope embryonic factor (TEF), 2988, 2988t
Thyrotropin receptor, 2790t
Thyrotropin-releasing hormone (TRH), 2985t, 2988, 2991
Thyroxine (T_4)
  characteristics of, 3024t
  free, 3023–3024, 3026
  half-life of, 2982
  laboratory evaluation of, 3024t, 3026
  receptor binding of, 3025, 3025f
  in sick euthyroid syndrome, 3039
  structure of, 3021f
  in subacute thyroiditis, 3038, 3038f
  total, 3026
  transport and metabolism of, 3024–3025
Thyroxine-binding globulin (TBG), 3023–3024, 3024t, 3026
Thyroxine-binding prealbumin. *See* Transthyretin
Thyroxine (T_4) therapy, 291, 2996t
TIA. *See* Transient ischemic attack (TIA)
Tiagabine, 3423t, 3424, 3661, 3713t
Tianeptine, 3696
Tibrovirus, 1656, 1662t
Tic(s), 3509t, 3515
Ticagrelor
  actions of, 944
  adverse effects of, 473, 944
  cardiac catheterization pretreatment, 1904
  dosage of, 944, S13
  indications for, 944
  for ischemic heart disease, 2102
  management before endoscopic procedures, 2483t
  for NSTE-ACS, 2110, 2110t, 2112
  in PCI, 2104, 2127, 2128
  for peripheral artery disease, 2173
  for STEMI, 2121
  for stroke risk reduction, 3449
Ticarcillin, 355, 374
Ticarcillin-clavulanate
  for anaerobic bacterial infections, 1379t
  indications for, 1175
  for peritonitis, 1071
  for *S. maltophilia* infection, 1309t
Tic douloureux. *See* Trigeminal neuralgia
Tick(s)
  Amblyomma spp., 1445, 1452, 1456, 2812–2813, 3730
  bite characteristics of, 3730
  Dermacentor, 1456, 3730
  Ixodes. See Ixodes spp. ticks
  Lyme disease prophylaxis following bite, 1453, 3731
  Ornithodoros spp., 1445, 1445f
  removal of, 3731
Tick-borne diseases
  Babesia infections. *See Babesia* spp. infections
  Colorado tick fever, 1657t, 1669

  - ehrlichioses, 1455t, 1460–1461
  - encephalitis, 1014t, 1017, 1660t, 1671
  - human granulocytic anaplasmosis, 1452, 1455t, 1460f, 1461–1462
  - Lyme borreliosis. *See* Lyme borreliosis
  - relapsing fever. *See* Hard tick relapsing fever (HTRF); Relapsing fever
  - rickettsial. *See* Rickettsial diseases
  - RMSF. *See* Rocky Mountain spotted fever (RMSF)
  - Southern tick-associated rash illness, 139t, 146, 1452
  - tick-borne lymphadenopathy, 1455t
  - tick granuloma, 3730
  - tick paralysis, 3730
  - tularemia. *See* Tularemia
- Ticlopidine
  - actions of, 943f
  - adverse effects of, 925, 2940t
  - management before endoscopic procedures, 2483t
  - management before lumbar puncture, S3
- Tie2, 545f, 546, 546f
- Tietze's syndrome (costochondritis), 103t, 104, 2919–2920, 2971
- Tigecycline
  - for *Acinetobacter* infections, 1296–1297, 1297t
  - actions of, 1166, 1177, 1182t, 1183f
  - adverse effects of, 1177
  - for anaerobic bacterial infections, 1379
  - for *Enterobacterales* infections, 1283, 1284
  - for *Enterobacter* infections, 1292
  - for enterococcal infections, 1220
  - for gram-negative bacterial infections, 1268t
  - indications for, 1174t, 1177
  - for *Klebsiella* spp. infections, 1291
  - for *Legionella* spp. infections, 1275
  - for *Nocardia* infections, 1362
  - for NTM infections, 1428
  - in pregnancy and lactation, 1169t
  - resistance to
    - in *Acinetobacter,* 1295
    - in *Enterobacterales,* 1283
    - mechanisms of, 1182t, 1183f, 1184, 1221
    - rates of, 1174t
  - for staphylococcal infections, 1204
- Tiger snakes, 3717. *See also* Snakebites
- "Tight-building" syndrome, 2241
- Tight epithelia, 2365
- Tight junction, 2365
- Tildrakizumab, 387t
- Tilt table test, 157f, 159f, 161, 1915, 3542, 3544–3545f
- Tiludronate, 3314t
- Timed Up and Go (TUG) test, 180
- Time-of-flight imaging. *See* Matrix-assisted laser desorption/ionization/time-of-flight mass spectometry (MALFDI-TOF MS)
- Time-restricted feeding, 3951
- TIMI (Thrombolysis in Myocardial Infarction) trial, 2107f, 2108, 2119
- Timolol
  - genetic variations in response to, 492, E72t
  - for ischemic heart disease, 2100t
  - metabolism of, 485t
  - for migraine prevention, 3469
  - overdosage/poisoning with, 3712t
- Timothy syndrome, 3404
- Tincture of opium, 312
- Tinea, 394, 1723, 1723t, 1724t. *See also* Dermatophyte infections
- Tinea capitis, 382, 394, 395t, 1723
- Tinea corporis, 389, 1723, A5
- Tinea cruris, 380f, 389, 1723
- Tinea manum, A5
- Tinea pedis, 380f, 389, 1723, A5
- Tinea unguium, 389, A5
- Tinea versicolor, **389**
  - clinical features of, 378t, 388t, 389, 398t, 1722
  - diagnosis of, 382, 389, 398t
  - etiology of, 388t, 389
  - treatment of, 388t, 389, 398t
- Tinel's sign, 2943, 3613
- Tin exposure, 3702
- *TINF3* gene mutations, 3820t
- Tingling, 172. *See also* Sensation
- Tinidazole
  - actions of, 1753
  - adverse effects of, 1744t
  - for amebiasis, 1757, 1757t
  - for bacterial vaginosis, 1103
  - drug interactions of, 1744t
  - for *Giardia* infections, 1805
  - in *H. pylori* eradication, 1302t, 2528
  - pharmacology of, 1753
  - in pregnancy and lactation, 1744t
  - for *T. vaginalis* infections, 1101t, 1102, 1808
- Tinnitus, 247, 251, 3457
- Tinnitus masker, 251
- TINU (tubulointerstitial nephritis with uveitis), 2356, 2439, 2439f
- Tinzaparin, 2163t
- Tiopronin, 2453, 3373
- Tiotropium, 2255
- TIP39, 3274
- Tipifarnib, 591, 614
- Tipranavir, 1619f
- TIPS (transjugular intrahepatic portosystemic shunt), 317, 330, 2714, 2715
- Tirabrutinib, 894
- Tirofiban
  - actions of, 943f, 945
  - adverse effects of, 923t, 946
  - dosage of, 945, S13
  - genetic variation in response to, 939t
  - indications for, 945
  - for NSTE-ACS, 2110t
  - in PCI, 2127
  - pharmacology of, 945t
- Tirzepatide
  - actions of, 3192t, 3193
  - adverse effects of, 3192t, 3193–3194, 3211t
  - contraindications to, 3192t, 3194, 3211t
  - for diabetes mellitus, 3193, 3211t, 3213
  - for metabolic syndrome, 3257
  - for NASH, 2706
  - for obesity, 3192t, 3193, 3193f
  - for OSA, 2282
- Tisagenlecleucel, 535t, 559f, 3824t
- Tisotumab vedotin, 535t, 566t, 713
- Tissue factor (TF), 470, 470f, 942, 2318f
- Tissue factor pathway inhibitor, 471, 471f, 947
- Tissue nonspecific alkaline phosphatase (TNSALP), 3316
- Tissue plasminogen activator (tPA)
  - for acute limb ischemia, 2175
  - arterial thrombotic disease and, 941
  - contraindications to, 2120
  - in fibrinolysis, 471f
  - recombinant. *See* Alteplase (rtPA)
  - for STEMI, 2120
- Tissue transglutaminase, 2790t
- Tissue transglutaminase IgA, 2545
- Titanium dioxide, 435, 436t, 2238
- Titin, 1848
- Tivozanib, 690t
- Tixagevimab-cilgavimab, 1543
- Tizanidine, 3570, 3588, 3711t, 3713t
- *TK* gene, 3828
- TKIs. *See* Tyrosine kinase inhibitors (TKIs)
- TLC (total lung capacity), 2202, 2202f, 2203f, 2206f
- *TLR7* gene, 1582t
- *TLR7* gene mutations, 2823t
- TLR (toll-like receptor) pathway deficiencies, 2796–2797
- TLRs. *See* Toll-like receptors (TLRs)
- T lymphocytes. *See* T cell(s)
- TMAO (trimethylamine N-oxide), 3836
- TMA test, for HIV RNA, 1590t
- *TMEM38B* gene mutations, 3324, 3325t
- *TMEM*43 gene mutations, 2004t
- *TMEM127* gene mutations, 3078–3079, 3079t
- *TMEM173* gene mutations, 2937
- TMPRSS2, 3167, 3167f, 3177
- *TMPRSS2-ERG* gene mutations, 521, 521t
- *TMPRSSG* gene mutations, 764
- TMP-SMZ. *See* Trimethoprim-sulfamethoxazole (TMP-SMZ)
- TNF. *See* Tumor necrosis factor (TNF)
- *TNFAIP3* gene mutations, 2553, 2554t
- *TNF* gene mutations, 2823t
- *TNFRSF11A/B* gene mutations, 2936
- *TNFRSF11A/B* genes, 3312
- *TNFSF15* gene mutations, 2579
- TNK. *See* Tenecteplase (TNK)
- TNM staging, 499, 501
  - of bladder cancer, 692, 693f
  - of bone sarcoma, 729, 729t
  - of breast cancer, 635
  - of colorectal cancer, 656–657, 657f
  - of head and neck cancer, 611–612, 612f
  - of lung cancer, 621–622, 622t
  - of prostate cancer, 698–699, 698f, 698t
  - of RCC, 688, 689f
  - of soft tissue sarcomas, 728, 728t
- *TNNC1* gene mutations, 2004t
- TNSALP (tissue nonspecific alkaline phosphatase), 3316
- *TNXB* gene mutations, 3328t, 3330
- Tobacco use, 3168. *See also* Smoking
- Tobin index, 2308
- Tobramycin
  - actions of, 1166, 1182t
  - for hospital-acquired pneumonia, 963t, 1032t
  - indications for, 1173t
  - for infective endocarditis, 1044
  - inhaled, for bronchiectasis prophylaxis, 2244
  - resistance to, 1173t, 1182t, 1219, 1359f
  - for ventilator-associated pneumonia, 1032t
- Tocilizumab
  - actions of, 2784, 2792t, 2851
  - adverse effects of, 2850t, 2851, 2895, 2896t, 2903
  - for CAR-T neurotoxicity, 2351
  - for COVID-19 disease, 1542
  - for cytokine release syndrome, 594, 754, 2351
  - for giant cell arteritis, 227, 2171, 2903
  - indications for, 2792t
  - for rheumatoid arthritis, 2850t, 2851
  - for systemic sclerosis, 2873–2874
  - for Takayasu arteritis, 2904
- α-Tocopherol, 512, 513, 3612. *See also* Vitamin E
- α-Tocotrienol, 3816
- Todd's paralysis, 3409
- Toe-brachial index, 2172
- Toe(s), gout in, A15
- Tofacitinib
  - actions of, 2852
  - adverse effects of, 2568t, 2570, 2851t, 2852
  - for axial spondyloarthritis, 2887t
  - for IBD, 2567, 2568t
  - for IBD-associated arthritis, 2892
  - for psoriatic arthritis, 2890
  - for rheumatoid arthritis, 2851t, 2852
  - testing before and during treatment, 2568t
- TOF (time-of-flight) imaging. *See* Matrix-assisted laser desorption/ionization/time-of-flight mass spectometry (MALFDI-TOF MS)

- To-fro murmur, 288, 288f
- Togaviridae
  - clinical syndromes of, 1666t, 1673
  - geographic distribution of, 1664–1665t
  - reservoirs and vectors of, 1657t
  - structure of, 1478f, 1480t, 1656
- Tolbutamide, 324
- Tolcapone, 3503, 3504t
- Tolerable upper levels, 2599
- Tolerance, 81, 1218, 2756
- Tolerance induction, 2814
- Tolfenamic acid, 3467t
- Toll-like receptor (TLR) agonists, 997t, 998
- Toll-like receptor (TLR) pathway deficiencies, 2796–2797
- Toll-like receptors (TLRs)
  - in adaptive immune response, 971, 971t, 2758, 2760
  - fever and, 135
  - in glomerular disease, 2412
  - in HSV infection, 1497
  - in IC/BPS, 331
  - in innate immune response, 1191, 1482
  - ligands and functions of, 2759t
  - in platelets, 938
  - in rheumatoid arthritis, 2844, 2845f
  - in sepsis/septic shock, 2317
  - signaling pathways of, 2759f
  - in Sjögren's disease, 2876
- Tolosa-Hunt syndrome, 233, 234, 1720, 3559
- Tolterodine, 3549, 3588
- Tolvaptan
  - adverse effects of, 351, 619
  - for heart failure, 1992
  - for polycystic kidney disease, 2433, 2434
  - for SIAD, 351, 619, 738, 3020, S1
- TOMAC (tonic motor activation), 3518
- Tonate virus, 1657t, 1673
- Tone, 168
- Tongue
  - alterations of, 265t
  - bald, 265t
  - black hairy, 264t
  - fissured (scrotal), 265t, A3
  - geographic, 265t, 380f, A3
  - paralysis of, 3558
  - raspberry, 265t
  - in Sjögren's disease, 2876, 2877f
  - strawberry, 265t
  - ulcer on, A3
  - white coated, A3. *See also* Thrush
- Tongue worms, 3733
- Tonic-clonic seizures, 3410, 3419t. *See also* Seizure(s)
- Tonicity, 2371, 2371f
- Tonic motor activation (TOMAC), 3518
- Tonic pupil, 221, V3
- Tonsillectomy, 472–473
- Tonsillitis, 253
- Tonsils, orange, 1860
- Tooth. *See* Teeth
- Tophi, 404, 2948f
- Topiramate
  - adverse effects of, 242, 3191, 3423, 3423t, 3470t
  - for cluster headache prevention, 3472t
  - for hemicrania continua, 3473
  - for migraine prevention, 3469, 3470t
  - for nausea and vomiting, 299t, 300
  - for neuropathic pain, 82
  - overdosage/poisoning with, 3713t
  - pharmacology of, 3423t
  - with phentermine, for obesity, 3191, 3192t, 3193f
  - for posttraumatic headache, 117
  - for raised CSF pressure headache, 117
  - for seizures, 716, 3423, 3423t
  - for SUNCT/SUNA prevention, 3473
- "Top of the basilar" syndromes, 3434
- Topoisomerase I, 564
- Topoisomerase I (Scl-70), 2790t, 2864t
- Topoisomerase II, 564, 574
- Topoisomerase IV, 1167
- Topoisomerase inhibitors, 755, 816, 826
- Topoisomerase poisons, 562t, 564. *See also specific drugs*
- Topology, of biologic system, 3955, 3956f
- Topotecan
  - actions of, 564
  - adverse effects of, 562t, 564
  - for Ewing's sarcoma, 730
  - interactions and issues, 562t
  - for lung cancer, 630
  - for ovarian cancer, 712
- Toppling falls, 181. *See also* Fall(s)
- *TOR1A* gene mutations, 3510, 3510t
- TORC1 (target of rapamycin complex 1), 3654
- Toremifene, 637
- Toripalimab, 534t, 2787t
- Toroviruses, 1628f, 1632
- Torsades de pointes. *See* Ventricular tachycardia, polymorphic
- Torsemide, 374, 1991, 1993t, 2145t, 2331
- Torticollis, 3510
- TOS (true neurogenic thoracic outlet syndrome), 133
- Toscana virus, 1112, 1662t, 1667–1668
- Tositumomab, 2791t
- Total artificial heart, 2033
- Total body water, 2371, 2371f, S1
- Total energy expenditure (TEE), 2619
- Total iron-binding capacity. *See* Transferrin saturation
- Total lung capacity (TLC), 2202, 2202f, 2203f, 2206f
- Total parenteral nutrition. *See* Parenteral nutrition
- Touch localization test, 174
- Touch sensation, 173, 174t, 1410, 3381
- Touraine-Solente-Golé syndrome, 2970. *See also* Hypertrophic osteoarthropathy
- Tourette's syndrome, **3515**
- Tourniquet, for snakebite, 3718
- Toxic adenoma, 3043, 3044f
- Toxic epidermal necrolysis (TEN)
  - vs. AGEP, 425
  - clinical features of, 142t, 147, 147f, 400, 424–425f, 426t, 427t, A1
  - diagnosis of, 426t
  - drugs associated with, 401, 425, 427t
  - epidemiology of, 142t
  - etiology of, 142t
  - genetic considerations in, 419
  - pathogenesis of, 417t, 418f, 419, 489, 2816f
  - prevention of future episodes, 427
  - vs. staphylococcal scalded skin syndrome, 147, 401, 1051
  - timeline of, 427, 428f
  - treatment of, 425, 1051
- Toxic erythema of chemotherapy, 421–422, 421f
- Toxic granulations, neutrophils, 453
- Toxic megacolon, 1323, 1325, 2557
- Toxic multinodular goiter, 3042–3043. *See also* Thyroid nodular disease
- Toxic optic neuropathy, 227–228
- Toxic shock syndrome (TSS)
  - clostridial, 1239t, 1243
  - diarrhea in, 306
  - etiology of, 993
  - focal infections and, 993
  - skin manifestations of, 402, 993
  - staphylococcal
    - clinical features of, 141t, 146, 402, 993, 1201, 1201t
    - diagnosis of, 1201
    - pathogenesis of, 134, 1198, 1201, 2777
    - rash in, 141t, 146
    - tampon-associated, 1205
    - treatment of, 1205
    - vaginitis in, 1103
  - streptococcal
    - case definition of, 1211, 1212t
    - clinical features of, 141t, 146, 402, 993, 1211, 1212t
    - diagnosis of, 1211–1212
    - epidemiology of, 141t
    - etiology of, 141t, 1211
    - rash in, 141t, 146
    - treatment of, 1208t, 1212
- Toxic sponges, 3725
- Toxidrome, 3704
- Toxins. *See also specific toxins*
  - bacterial, 965
  - hepatic effects of, 2665, 2667t
  - neurologic effects of, 3608, 3609–3610t, 3610–3611
  - pulmonary effects of, 2236t, 2239–2240, 2239t
  - testicular effects of, 3115
  - in undiagnosed disease, 3998
- *Toxocara canis/cati,* 961t, 1812, S12. *See also* Larva migrans, visceral
- *Toxoplasma gondii,* 1737–1738, 1797, 1797f, S12
- Toxoplasmosis, **1797**
  - asymptomatic, 1799
  - in cancer patient, 583
  - clinical features of, 1738, **1799**
    - cardiac, 1799, 1927
    - chronic meningitis, 1131t
    - CNS, 583, 1138, 1608, 1608f, 1799, 1800, 1800f
    - gastrointestinal, 1799
    - in immunocompetent patient, 1799–1800, 1802
    - in immunocompromised patient, 1738t, 1800, 1800f, 1802
    - lymphadenopathy, 477, 1799
    - myelopathy, 3566
    - ocular, 225, 1739t, 1799, 1801, 1802
    - pulmonary, 1799
    - renal, 2429
    - seizures, 1738t
    - weakness, 170
  - congenital, 1738t, 1797, 1798, 1800–1801, 1802
  - diagnosis of, 1801–1802, S12
  - differential diagnosis of, 1511t, 1801, 1801t
  - epidemiology of, 1797–1798, S12
  - etiology of, 1797, 1797f
  - global considerations in, 1800
  - in HCT recipient, 1155t, 1157
  - in HIV infection, 1593t, 1595f, 1608, 1608f, 1800, 1800f, 1802
  - immune response to, 1799
  - pathogenesis of, 1798
  - pathology of, 1798–1799
  - in pregnancy, 1738, 1738t, 1798, 1800–1801
  - prevention of, 1803
  - prophylaxis for, 1803
  - in solid organ transplant recipient, 1159t, 1164t, 1798
  - transmission of, 1797f, 1798
  - in transplant recipient, 2350
  - treatment of, 1138, 1157, 1752, 1802–1803
- *TP53* gene mutations
  - aflatoxin B_1 and, 660
  - in AML, 826, 828
  - in bladder cancer, 691
  - in breast cancer, 633
  - in carcinoma of unknown primary, 734
  - in CLL, 851, 851t, 854t
  - in CML, 836
  - in esophageal cancer, 642
  - in gastric adenocarcinoma, 646f, 647
  - genetic testing for, 3802

  in HCC, 660, 661t
  in lung cancer, 615, 617
  in MDS, 815t, 816
  in multiple myeloma, 884
  in ovarian cancer, 710
  in pancreatic cancer, 672
  tumors associated with, 524t
tPA. *See* Alteplase (rtPA); Tissue plasminogen activator (tPA)
TPIT gene mutations, 2992
TPM1 gene mutations, 2004t
TPMT (thiopurine *S*-methyltransferase), 485t, 493, 2408, 2835, E72t
TPMT (thiopurine *S*-methyltransferase) deficiency, 3805
TPN (total parenteral nutrition). *See* Parenteral nutrition
TPO (thyroid peroxidase), 2790t, 3021, 3023, 3028t
TPO (thyroid peroxidase) antibodies, 3026, 3029, 3033
TPP (thyrotoxic periodic paralysis). *See* Hypokalemic periodic paralysis (HypoKPP)
TPP1 gene mutations, 3357
TPPA (*T. pallidum* particle agglutination) test, 1433
TR (thyroid hormone receptor), 2985, 3024–3025, 3025f
TRα, 3024–3025
TRβ, 3024–3025
TRAb (thyroid-stimulating hormone receptor antibody), 3026, 3029
Trabectedin, 564, 715, 729
Trabeculectomy, 231
Trace minerals. *See* Mineral(s)
Tracheal intubation, 192
Tracheitis, HSV, 1500
Trachelectomy, 713
Tracheobronchitis, *Aspergillus,* 1714
Tracheomalacia, 2203, 2919, 2920f
Trachoma, 224, 1474–1475
Traditional Chinese medicine, 3925t, 3928
TRAIL, 539, 540f, 541
Trail-making test, 2630
Trained immunity, 2756, 2824
Train-test split, 3973t
TRALI (transfusion-related acute lung injury), 905, 908t, 911
Tralokinumab, 384
Tramadol
  adverse effects of, 3414t, 3604t
  for fibromyalgia, 2965, 2966t
  for neuropathy, 3604t
  in older adults, 3883
  overdosage/poisoning with, 3715t
  for pain, 80f, 97t, 101
Trametinib
  actions and targets of, 533t, 569t, 572
  adverse effects of, 569t, 753t, 754
  interactions and issues, 569t
  for lung cancer, 628
  for melanoma, 572, 604, 604t, 605
"Tram tracks" sign, 2243, 2420, A4
Trandolapril, 1995t, 2144t
Tranexamic acid, 820, 824, 913, 926, 927, 930
Tranilast, 923t
TRANK1 gene mutations, 3668
Transaminases, 948, 2635
Transarterial chemoembolization (TACE), 662, 664t, 665–666
Transarterial radioembolization (TARE), 666
Transbronchial biopsy, 2209, 2262
Transbronchial needle aspiration, 2209–2210
Transcatheter aortic valve implantation (TAVI)
  AV block following, 1926
  endocarditis following
    clinical features of, 1039
    diagnosis of, 1039–1040
    etiology of, 1037, 1037t
    incidence of, 1037, 1037t
    outcome, 1047–1048
    treatment of, 1045, 1046
  with mitral valve repair, 2067
  outcomes of, 2041, 2041f, 2131, A11
  patient selection for, 2040, 2041f, 2042t, 2046, A11
  procedure for, A11, A11v
  valve models for, 2040, 2041f, 2130f, 2131, A11
Transcatheter mitral valve repair and replacement, 2056–2057, 2056–2057f
Transcatheter mitral valve replacement, 2067, A11, A11v
Transcatheter tricuspid valve repair and replacement, 2062, 2062f
Transcobalamin, 781
Transcobalamin deficiency, 787, S9
Transcortical aphasia, 201t, 202
Transcranial direct current stimulation, 3967, 3967f
Transcranial magnetic stimulation (TMS)
  for major depression, 3667, 3967
  for migraine, 3467t, 3468t, 3469
  principles of, 3966–3967, 3967f
Transcription coactivator p75, 2790t
Transcription factors
  defects in, diseases caused by, 3783t
  in gene expression, 3940
  gene expression and, 3780–3782, 3780f
  modulation of, 570t, 573
  in signal transduction in cancer cells, 531–537
  in stem cells, 760, 760f
Transcriptome, 976t, 3779, 3781
Transcriptomics, 3781
Transcutaneous supraorbital nerve stimulation, 3467t, 3469, 3470t
Transcutaneous tibial nerve stimulation, 2584
Transcytosis, 969
Transduction, 3823
Transendothelial migration, 2781
Transesophageal echocardiography (TEE)
  in aortic dissection, 2170
  in aortic regurgitation, 2044, 2045f
  in aortic stenosis, 1891, 2038, A11
  in atrial fibrillation, 1949
  in atrial septal defect, A11
  indications for, 1878
  in infective endocarditis, 1040–1041, 1042f
  in mitral regurgitation, 2053
  in mitral stenosis, 2049
  in mitral valve prolapse, 2057f, 2058
  in murmur evaluation, 290
  principles of, 1877–1878
Transfalcial herniation, 187f, 188
Trans fats, 3246
Transferase variants, 493
Transfer dysphagia, 293, 2492. *See also* Dysphagia
Transfer learning, 3972t
Transferrin
  diferric, 762
  iron content of, 762
  in iron metabolism, 762, 763f
  laboratory measurement of. *See* Transferrin saturation
  monoferric, 762
  in nutrition assessment, 2618t
Transferrin-binding proteins, 1253
Transferrin receptor(s), 762, 3334, 3334f, 3335
Transferrin saturation, 449, 764, 765, 767t
Transformation, bacterial, 1181
Transformer models, 3972t
Transforming growth factor-α (TGF-α), 546, 2516, 2516f, 2539
Transforming growth factor-β (TGF-β)
  in aortic aneurysm, 2166
  driver genes of, 526t
  in familial pulmonary arterial hypertension, 3957
  in heritable thoracic aortic disease, 3332
  in hypercalcemia of malignancy, 737
  in Loeys-Dietz syndrome, 3332t
  mutations in, 3957
  in oncogene signaling pathway, 542f
  in pulmonary arterial hypertension, 2185
  in radiation-induced fibrosis, 754
  source, target, and biologic activity of, 2767t
  in thyroid hormone synthesis, 3023
Transfusion(s)
  ABO compatibility in, 903t, 905
  adverse reactions to, **905**
    immune-mediated
      allergic, 911
      alloimmunization/platelet refractoriness, 911–912
      febrile nonhemolytic, 911
      graft-versus-host disease, 911
      hemolytic reactions, 905, 908, 909f, 911
      immunomodulation, 912
      posttransfusion purpura, 911
      transfusion-related acute lung injury, 911
      treatment of, 908
    incidence of, 905
    nonimmunologic
      fluid overload, 912
      hypotension, 912
      iron overload, 776, 813, 912
      massive transfusion-associated reactions, 323, 358–359, 912
      of uncertain cause, 912
    prevention of, 905, 908, 912
    warnings signs for, 908t
  alternatives and perspectives, 913
  for AML, 832
  for anemia, 450, 769
  for aplastic anemia, 813
  for β thalassemia, 777
  for chemotherapy-associated anemia/thrombocytopenia, 575
  for cold agglutinin disease, 803
  component collection and manufacturing process for, 901, 902–903t
  cryoprecipitate, 907t
  for DIC, 934
  donor mononuclear cells, 907t
  efficacy evaluation of, 906–907t
  granulocyte concentrates, 907t
  for hemophilia, 930
  indications for, 906–907t
  infections transmitted by
    Babesia, 913t, 1776–1777
    bacterial, 912, 913t
    CMV, 1513, 1515
    Creutzfeldt-Jakob disease, 3529
    hepatitis, 2653–2654
    malaria, 1766
    parasites, 913t
    prevalence of, 912–913, 913t
    viruses, 912–913, 913t
    West Nile virus, 1672
    Yersinia, 1349
  metabolic alkalosis in, 373
  multicomponent, 907t
  packed red blood cell, 813, 832
  for paroxysmal nocturnal hemoglobinuria, 806
  plasma, 906t
  platelet. *See* Platelet transfusion
  reducing need for, 913
  for sickle cell disease, 776
  whole blood, 907t

Transfusion-associated circulatory overload (TACO), 905, 908t, 912
Transfusion-related acute lung injury (TRALI), 905, 908t, 911
Transgender, 3178, 3180t
Transhemorrhoidal dearterialization, 2584
Transient acantholytic dermatosis, 379t
Transient aplastic crisis, 1523–1524, 1524t
Transient elastography, 2632, 2636, 2699
Transient global amnesia, 207–208, 3484–3485
Transient ischemic attack (TIA), **3448**
  in basilar artery distribution, 3432
  definition of, 3428
  at high altitudes, 3758
  prevention of, 3448–3449, S13
  small-vessel, 3446
  stroke risk following, 3448, 3448t
  treatment of, 3429f, 3448
  visual symptoms of, 230
Transient monocular blindness, 225
Transient myeloproliferative disorder (TMD), 879
Transition (mutation), 3784
Transitional cell (urothelial) carcinoma, 691, 692f, 694. *See also* Bladder cancer
Transit volume, in radiation therapy, 554
Transjugular intrahepatic portosystemic shunt (TIPS), 317, 330, 2714, 2715
Translocation, chromosome, 521, 521t, 522f
Trans man/transgender man, 3180t
Transmission, 964
Transmission disequilibrium test, 3797t
Transmission networks, 984–985
Transpeptidase enzymes (PBPs), 1183
Transplantation. *See specific types*
Transplant recipient. *See* Hematopoietic cell transplantation (HCT); Solid organ transplant recipient
Transporters associated with antigen processing (TAP) proteins, 2776
Transporter variants, 493
Transposition of the great arteries, **2074**
  arterial switch procedure for, 2075, 2075f, 2075t
  atrial switch procedure for, 2074–2075, 2075f, 2075t
  AV conduction block in, 1927
  congenitally corrected, 2075–2076, 2076f
  hypoxia in, 277
  pathophysiology of, 2073–2074, 2074f
  Rastelli procedure for, 2075, 2075t
Transrectal ultrasonography, 697
Transsphenoidal surgery
  for acromegaly, 3009
  complications of, 3001
  for Cushing's disease, 3012
  for Cushing's syndrome, 3059
  for pituitary tumors, 2999–3000, 3001f, 3013
Transtentorial herniation, 187, 187f
Transthoracic (transtracheal) echocardiography. *See* Echocardiography
Transthoracic needle aspiration, 2213–2214
Transthyretin (TTR), 895, 896t, 899, 3024t, A9
Transtubular K^+ concentration gradient (TTKG), 356, 360, 360f, S1
Transurethral resection of bladder tumor (TURBT), 692, 693, 693t
Transvaginal ultrasound, 517t, 518
Transvenous phrenic nerve stimulator, 2282
Transverse sinus thrombosis, 1142
Transversion (genetic), 3784
Tr (DNER) antibodies, 743t, 749
Tranylcypromine, 3659t, 3710t
TRAP (tartrate-resistant acid phosphatase), 3315
TRAPS (tumor necrosis factor-α receptor-associated periodic syndrome), 136, 155, 462, 1131, 1132t, 2935t, 2936
Trastuzumab
  actions and targets of, 534t, 536, 541, 558, 559f
  adverse effects of
    cardiotoxicity, 558, 638, 753t, 754, 2017
    desensitization protocol for, 2819
    DIHS, 597
    infusion reaction, 594
    long-term, 753t
  for breast cancer, 536, 541, 558, 637, 638, 641
  for colorectal cancer, 658
  for esophageal cancer, 644
  for gastric cancer, 649
  genetic variations in response to, 493, E72t
Trastuzumab deruxtecan
  adverse effects of, 754
  for gastric cancer, 649
  for lung cancer, 627
  for metastatic breast cancer, 641
Trastuzumab emtansine, 638, 641
Traube's sign, 480, 2044
Trauma
  aortic, 2088, 2166
  bile duct, 2737
  cardiac, 2083, **2087**
  chest wall, 284
  head-related. *See* Traumatic brain injury (TBI)
  hormonal response to, 2986
  hypothermia in, 3768
  pneumothorax in, 2271
  testicular, 3115
Traumatic axonal injury, 3573, 3573f
Traumatic brain injury (TBI), **3570**
  amnesia in, 207
  axonal injury, 3573, 3573f
  brain abscess and, 1135
  cerebral contusion, 3573, 3573f
  classification of, 3571, 3571t
  concussion/mild, 3571, 3574, 3575t, S8
  cranial nerve injuries, 3573–3574
  definition of, 3571
  dementia and, 194, 3484
  epidural hematoma, 3572, 3572f
  fatigue in, 166
  frontal lobe syndromes and, 208
  headache in, 115t, 117
  hyperbaric oxygen therapy for, 3764
  hypokalemia in, 354
  hyponatremia in, 348
  imaging in, 3574
  incidence of, 3570
  of intermediate severity, 3576
  long-term outcomes in, 3576–3577
  olfactory dysfunction in, 240
  outcome prediction in, 3574
  postconcussion syndrome/disorders, 3575–3576
  recurrent. *See* Chronic traumatic encephalopathy (CTE)
  second-impact syndrome, 3575
  seizures and, 3411, 3414, 3574
  severe, 3576
  skull fracture, 3571–3572
  sports-related, 3574–3575, 3575t. *See also* Chronic traumatic encephalopathy (CTE)
  stem cell therapy for, 3408
  subarachnoid hemorrhage in, 3454, 3573
  subdural hematoma. *See* Subdural hematoma
  types and pathologies of, 3571–3574
  in war veterans, S7, S8
Travelers. *See* International travelers
Travelers' diarrhea, **1018**
  epidemiology of, 305, 1079–1080
  etiology of, 305, 1079, 1079t
  prevention of, 307, 1012t, 1018, 1019t, 1081
  prophylaxis for, 1018, 1081
  treatment of, 1018–1019, 1019t, 1081, 1081t
Travel health insurance, 1021
Travel health kit, 1021
Trazodone
  adverse effects of, 85, 3659t
  for depression, 85, 3659t
  for insomnia, 86, 216, 3660t
  for PTSD, 3663
Treacher Collins syndrome, E36t
"Tree-in-bud" pattern, 2243
T regulatory cells (T-regs), 1579, 2756, 2780, 2785, 2789, 2846. *See also* CD4+ T cells
  induction of, 3840
TREM2 gene mutations, 3395, 3480
Trematodes/trematode infections, 1736, 1810, **1825**
  approach to the patient, 1825–1826
  blood flukes. *See* Schistosomiasis
  control and prevention of, 1832
  epidemiology of, 1826t
  intestinal flukes. *See* Intestinal flukes
  liver flukes. *See* Liver flukes
  lung flukes. *See* Lung flukes (*Paragonimus* spp.)
  by organ system and signs/symptoms, 1738–1740t
Tremelimumab
  actions of, 559f, 2783
  adverse effects of, 2377
  for HCC, 664t, 665f, 666, 2787t
  for NSCLC, 630t
Tremor, **3508,** 3509t
Trench fever, 1351t, 1353t, 1354, 3732
Trench foot (immersion foot), 3771
Trench mouth (acute necrotizing ulcerative gingivitis), 261, 262t, 1374
Trendelenburg gait, 49
Treponema spp., 1184, 1428, 1437t, 1438. *See also* Treponematoses
Treponema pallidum, 1428–1429, 1433, 1437t, 1438. *See also* Syphilis
Treponema pallidum particle agglutination (TPPA) test, 1433
Treponematoses, **1436**
  clinical features of, 1438–1439, 1438–1439f
  comparison of, 1437t
  control of, 1439
  diagnosis of, 1439
  epidemiology of, 1436–1437, 1437f
  treatment of, 1439
Treprostinil, 2194, 2197t, 2874
TREs (thyroid response elements), 3025, 3025f
Tretinoin. *See* All-*trans*-retinoic acid (ATRA)
TRH (thyrotropin-releasing hormone), 2985t, 2988, 2991
Trial to Assess Chelation Therapy (TACT), 3702
Triamcinolone, 3648
Triamterene, 2144t, 2370
Triazolam
  for anxiety disorders, 3660t
  drug interactions of, 1742t
  for insomnia, 216
  overdosage/poisoning with, 3713t
  pharmacology of, 3660t
Tribeč virus, 1658t, 1669
Tribendimidine, 1753
Trichiasis, 1475
Trichinella/trichinellosis, **1810**
  clinical features of, 1131t, 1740t, 1810–1811, 2013
  diagnosis of, 961t, 1700, 1811f, S12
  eosinophilia in, S12
  epidemiology of, 961t, 1810
  life cycle of parasite, 1810, 1811f
  pathogenesis of, 1736, 1810–1811
  prevention of, 1812
  treatment of, 1752, 1811, 1812t

Trichloroethylene, 2665
Tricholemmoma, 404
Trichomonas vaginalis, 1808
Trichomonas vaginalis infections, 1101t, 1102, 1110, 1739t, **1808,** S11
Trichophyton, 388t, 389, 395t, 1723. *See also* Dermatophyte infections
Trichosporon spp. infections, 1688t, 1726t, 1730
Trichostatin A, 3942
Trichostrongylus/trichostrongyliasis, 1818
Trichotillomania, 395t
Trichromat, 222
Trichrome stain, S12
Trichuris trichiura/trichuriasis (whipworm), 961t, 1736, 1739t, 1815t, 1818, S12
Triclabendazole, 1742t, 1753, 1830t
Tricuspid atresia, 2076
Tricuspid regurgitation, **2060**
  in carcinoid heart disease, 683, 683f
  clinical features of, 327, 2061
  diagnosis of, 2060f, 2062, 3914v
  etiology of, 2059t, 2060
  in mitral valve disease, 2065
  murmur in, 283f, 285, 287, 287f
  pathophysiology of, 2061
  in pulmonary hypertension, 2188
  treatment of, 2061f, 2062, 2062f, 2067
  with tricuspid stenosis, 2067
Tricuspid stenosis, **2059**
  clinical features of, 1860, 2059–2060
  vs. constrictive pericarditis, 2084
  diagnosis of, 2060
  etiology of, 2059, 2059t
  heart sounds/murmur in, 283f, 288, 1866, 2060
  pathophysiology of, 2059
  treatment of, 2060
  with tricuspid regurgitation, 2067
Tricuspid valve repair/replacement, 2062, 2062f, 2067
Tricyclic antidepressants (TCAs)
  actions of, 3652–3653
  adverse effects of, 99–100, 3604t, 3659t, 3666
   erectile dysfunction, 3157
   hypothermia, 3768
   orthostatic hypotension, 159
   SA node dysfunction, 1920t
   weight gain, 3188
  contraindications to, 3666
  for depression, 3659t, 3666, 3666f
  drug interactions of, 485t, 489t, 1742t, 3660t
  for functional dyspepsia, 302
  genetic variations in response to, 492, E72t
  for headache, 117
  for IBS, 2575, 2577t
  for low back pain, 126
  metabolism of, 485t
  for migraine prevention, 3470t
  for narcolepsy, 215
  for nausea and vomiting, 298, 299t, 300
  for neuropathy, 3227, 3604t
  overdosage/poisoning with, 1874, 1875f, 3705, 3711t, 3712t, 3715t, A7
  for pain, 97t, 99–100, 100t, 3588
  pharmacology of, 3666
Trientine, 3340, 3518
Trifascicular block, 1930. *See also* Atrioventricular (AV) conduction block
Trifluoperazine, 3673t
Trifluorothymidine, 1502
Trifluridine, 1486t, 1487
Trifluridine/tipiracil, 562t, 565, 649, 658
Trigeminal autonomic cephalalgias, 114, 3463t, 3471–3472, 3471t
Trigeminal nerve
  anatomy of, 3551, 3552f
  disorders of, 242, 3551–3553, 3553t
  examination of, 3380
Trigeminal neuralgia, **3553**
  clinical features of, 3551–3552
  diagnosis of, 3553
  differential diagnosis of, 3552–3553
  ear pain in, 253
  epidemiology of, 3551–3552
  in MS, 3578
  oral pain in, 265
  pathophysiology of, 3552f
  treatment of, 100, 3553
  vs. trigeminal autonomic cephalgias, 3472
Trigeminal neuropathy, 3553
Trigeminovascular system, 114
Triggered arrhythmias/activity, 1913, 1913t
Trigger mechanism, bacterial, 967, 968f
Trigger, migraine, 3462
Trigger point injections, 128, 336
Triglycerides
  elevated levels of. *See* Hypertriglyceridemia
  long-chain, 2539, 2542f
  medium-chain, 2542
Trihexyphenidyl, 3503, 3512, 3674
Triiodothyronine (T_3)
  characteristics of, 3024t
  free, 3023–3024, 3026
  half-life of, 2982
  laboratory evaluation of, 3024t, 3026
  receptor binding of, 3025, 3025f
  in sick euthyroid syndrome, 3039
  structure of, 3021f
  in subacute thyroiditis, 3038, 3038f
  total, 3026
  transport and metabolism of, 3024–3025
Trilostane, 3012
TRIM5-α protein, 1558
Trimetazidine, 2103
Trimethoprim
  actions of, 782, 788, 1166, 1182t, 1183f, 2369, 2370
  adverse effects of, 346, 359, 373, 788, 1743t
  drug interactions of, 1743t
  indications for, 1743t
  for PCP, 1734t
  resistance to, 1182t, 1183f, 1185
Trimethoprim-polymyxin, 224
Trimethoprim-sulfamethoxazole (TMP-SMZ)
  actions of, 1177
  adverse effects of, 1093t, 1171t, 1177, 1733, 1734t
   cholestatic hepatitis, 324
   desensitization to, 2818
   on fecal flora, 1092
   in G6PD deficiency, 799t
   hepatotoxicity, 2665, 2671–2672
   in HIV infection, 1596, 1601, 1603
   neutropenia, 457
   photosensitivity, 420
   in pregnancy, 1229
   serum sickness–like reaction, 419
   thrombocytopenia, 923t
   vasculitis, 2940t
  for *B. cepacia* infections, 1309t
  for bite-wound infections, 1144, 1145t
  for blastocystosis, 1808
  for *Brucella* infections, 1334
  for *C. belli* infections, 1806
  for chronic granulomatous disease, 463
  for *Cyclospora* infections, 1807
  for cystitis, 1092, 1093t
  for donovanosis, 1358t
  drug interactions of, 1172t
  for enteric fever, 1318t
  for glanders, 1309t
  for granulomatosis with polyangiitis, 2899
  indications for, 1174t, 1177
  for *L. monocytogenes* infections, 1229
  for melioidosis, 1309t
  for meningitis, 1123
  for MRSA infections, 383, 964t, 1054, 1203t, 1204
  for *Nocardia* infections, 1362
  for nontyphoidal *Salmonella* infections, 1321t
  for osteomyelitis, 1064t, 1065t
  for paracoccidioidomycosis, 1726t
  for PCP, 1596, 1733, 1734t
  for PCP prophylaxis, 578–579, 1593t, 1734t, 1735, 2410, 2895
  for peritonitis prophylaxis, 1071
  for pertussis, 1279, 1280t
  for plague, S4
  prophylactic uses of
   in chronic neutropenia, 463
   in HCT recipient, 918t
   in HIV infection, 1593t
  for pyelonephritis, 1094
  resistance to
   in *B. cepacia* and *S. maltophilia,* 1309t
   in *E. coli,* 1087, 1174t
   in *Enterobacterales,* 1283, 1283t
   mechanisms of, 1185
   in *Nocardia* spp., 1359f
   in *Proteus* spp., 1291
   in *S. aureus,* 1174t
  for *S. maltophilia* infections, 1309t, 1314
  for staphylococcal soft tissue infections, 1203t
  for toxoplasmosis, 1157
  for toxoplasmosis prophylaxis, 1803
  for UTI prophylaxis, 1095
  for Whipple's disease, 1370
Trimethylamine N-oxide (TMAO), 3836
Trimethylaminuria (fish odor syndrome), 50t, 2610
Trimetrexate, 1742t
Trimipramine, 3659t
Trinucleotide repeat disorders, 3786, 3793, 3793t
Triose phosphate isomerase deficiency, 796t
Tripeptidyl peptidase II, 2786f
Triple A syndrome, 3067t
Triple-negative breast cancer, 634, 636, 638, 641
Tripod sign, 121t
Triptans, 3462, 3467t, 3468, 3715t
Trismus (lockjaw), 1231, 3553
Trisomy 12, 850, 851f
Trisomy 13, 2183
Trisomy 18, 2183
Trisomy 21. *See* Down's syndrome (trisomy 21)
Tritium exposure, S6
TrkB gene mutations, 3185, 3185t
TRK gene mutations, 3045
Trochanteric bursitis, 2944f, 2945, 2973
Trochlear nerve, 3380, 3574
Trochlear nerve palsy, 235
Troglitazone, 2667
Tropheryma whippelii, 1038, 1040, 1062, 1367, 1369, 2546. *See also* Whipple's disease
Trophic factors, 3505
Trophozoites, *Plasmodium,* 1760
Tropical pulmonary eosinophilia, 462, **1822**
Tropical spastic paraparesis. *See* HAM (HTLV-1-associated myelopathy)
Tropical sprue, 309, 787, **2546**
Tropical storms, 3740, 3742f
Tropisetron, 83
Tropomyosin, 1848–1849, 2005f
Troponins
  in cardiac contraction, 1848–1849, 1848f
  in cardiac trauma, 2088
  in chest pain evaluation, 107
  in CKD, 2391
  in NSTE-ACS, 2108
  in STEMI, 2115–2116, 2117f

Trousseau's sign, 365, 674
Trousseau's syndrome, 619, 673, 741
TRPC6 gene mutations, 2411
TRPM6 gene mutations, 2370
TRP proteins, 238
TRPV1 channel antagonists, 274
TRPV1 channels, 2573
Trubanaman virus, 1659t, 1669
True bug bites, 3737
True-negative tests, 24
True neurogenic thoracic outlet syndrome (TOS), 133
True-positive tests, 24
Trust, between patient and health care provider, 7, 69
Trypanids, 1794
Trypanosoma spp.
  T. brucei gambiense/rhodesiense, 1131t, 1737, 1740t, S12. *See also* Human African trypanosomiasis
  T. cruzi, 1737, 1740t, 1788, 2013, S12. *See also* Chagas disease
Trypanothione reductase, 1746
Trypsinogen, 2542
L-Tryptophan
  adverse effects of, 489
  contamination of, 462, 3926
  drug interactions of, 3660t
  overdosage/poisoning with, 3715t
Tryptophan hydroxylase, 682, 682f, 2573, 2790t
Tryptophan metabolism disorders, 3368t, 3369t
TSC1/2 gene mutations
  in autism spectrum disorder, 3653f
  in HCC, 661t
  in lymphangioleiomyomatosis, 2267
  in NETS, 686
  in tuberous sclerosis, 524t, 686, 687t, 723, 2434
  in urothelial carcinoma, 691
T-scores, 3291, 3297, 3297f
Tsetse flies, 1792, 1793
TSH. *See* Thyroid-stimulating hormone (TSH)
TSI (thyroid-stimulating immunoglobulin), 3026
T-SPOT.TB assay, 1394
TSS. *See* Toxic shock syndrome (TSS)
TTF-1. *See* Thyroid transcription factor-1 (TTF-1)
TTF-2, 3028t
tTg antibody test, 3092
TTKG (transtubular K^+ concentration gradient), 356, 360, 360f, S1
TTN gene mutations, 2004t, 2009, 2017
TTP. *See* Thrombotic thrombocytopenic purpura (TTP)
TTR (transthyretin), 895, 896t, 899, 3024t, A9
TTR cardiac amyloidosis, 2020–2021, 2027
Tubal (ectopic) pregnancy, 112t, 113, 3136
Tubal sterilization, 3152–3153, 3153t
TUBB3 gene mutations, 3112, 3112t
Tubby syndrome, 3184t
Tuberculin skin test (TST). *See* Purified protein derivative (PPD)/tuberculin skin test (TST)
Tuberculoid leprosy, 397, 1407, 1407f
Tuberculoma, 1389
Tuberculosis (TB), **1380**
  bronchiectasis in, 2242
  as cause of fever of unknown origin, 150, 153
  in children, 1385–1386, 1386f, 1400
  as chronic disease, 3848–3849
  complications of, 1390
  diagnosis of, **1391**
    biomarkers, 1393
    culture, 1392
    drug susceptibility testing, 1392–1393
    genomics in, 977t
    IFN-γ release assays, 1394
    imaging, 1386f, 1393, A12, A16
    invasive procedures, 1393
    limitations of, 975–976
    microscopy, 1392
    nucleic acid amplification, 1391–1392, S11
    tuberculin skin testing, 1385, 1393–1394
  epidemiology of, 1381–1382f, 1381–1383, 1419, 3848–3849
  etiology of. *See Mycobacterium tuberculosis*
  extensively drug-resistant, 1382, 1395t, 3848
  extrapulmonary
    adrenal, 1390, 3066
    aortic aneurysm, 2166
    congenital, 1390
    gastrointestinal, 328–329, 1389
    genital, 1105
    genitourinary, 1388, 1388f, 3133
    in HIV infection, 1390–1391, 1390f
    lymphadenitis, 477, 1387, 1387f
    lymphedema, 2183
    mastitis, 1390
    meningitis, 962t, 983, 1130t, 1388–1389, A16
    miliary, 153, 1386f, 1389–1390, 1400
    myocardial, 2013
    ocular manifestations, 1390
    oral manifestations, 261, 262t
    otitis, 1390
    pericardial, 1389, 2083, 2084
    pleural, 1387–1388
    skin manifestations, 406, 1051, 1390
    spinal/skeletal
      vs. *Brucella* infection, 1333, 1333t
      clinical features of, 1058, 1388
      diagnosis of, 150, A16
      osteomyelitis, 967, 1058, 1062
      Poncet's disease, 1058
      Pott's disease, 1075, 1388, 1388f
      treatment of, 1400
    subacute meningitis, 1126, 1388–1389
    treatment of, 1400
    tuberculoma, 1389
    upper airway, 1388
  global burden of, 1381–1382f, 1381–1383, 1419, 3848–3849
  in HIV infection
    antiretroviral therapy and, 1390–1391, 1419–1420
    clinical features of, 1390–1391, 1596–1597
    control of, 1403
    in developing countries, 1390, 1403, 3847f, 3848
    diagnosis of, 1393
    epidemiology of, 1382, 1390, 1390f, 1596
    immune reconstitution inflammatory syndrome in, 1391, 1420, 1597
    interaction of, 1575
    prophylaxis for, 1401–1402, 1401t, 1593t
    risk of, 1382–1383, 1383t
    treatment of, 1400, 1424, 1597
  incubation period for, S7
  infection control measures, 1150, 1625
  inflammasome mutations in, 2761t
  innate resistance to, 1384
  latent infection
    diagnosis of, 983
    screening for, 1001–1402, 1401t, 2851
    in solid organ transplant recipient, 1163
    treatment of, 1401–1403, 1402t, 1419, 1419t
  multidrug-resistant
    detection of, 977t, 981
    epidemiology of, 1381–1382, 1397–1398, 1398f
    treatment of, 1395t, 1397–1400, 1420–1421, 1421t, 3848
  natural history of, 1383
  pathogenesis of, 1383–1385
  prevention and control of, 1163, 1164t, 1403, 3848
  pulmonary
    cough in, 274, 1387
    hemoptysis in, 274, 1387
    pleuritis in, 2270
    postprimary disease, 1383, 1386–1387, 1386f
    primary disease, 1382, 1385–1386, 1386f
  risk factors for, 1382–1383, 1383t
  silica exposure and, 2238
  in solid organ transplant recipient, 1158, 1160, 1162f, 1163
  transmission of, 1382–1383
  treatment of, **1394, 1418**
    adherence to, 1396
    adverse effects of, 1397, 1420–1424. *See also specific drugs*
    in children, 1394t, 1400
    DOTS plus strategy, 1396
    DOT strategy, 1396, 1420, 3848
    for drug-resistant strains, 1394t, 1397–1400, 1421, 1421t
    drug susceptibility testing, 1392–1393
    failure and relapse, 1397
    first-line antituberculosis drugs, 1394t, 1421–1424
      ethambutol. *See* Ethambutol
      isoniazid. *See* Isoniazid (INH)
      pyrazinamide. *See* Pyrazinamide
      rifabutin. *See* Rifabutin
      rifampin. *See* Rifampin
      rifapentine. *See* Rifapentine
    fixed-dose combination products, 1396
    monitoring response to, 1397, 1420t
    patient care and support, 1396
    in pregnancy, 1395t, 1400
    preventive, 1401–1403, 1402t, 1419, 1419t
    principles of, 1419–1421, 1419t
    regimens, 1394t, 1395–1396, 1419t
    second-line antituberculosis drugs, 1424–1426
    simplified approach to, 1419, 1419t
    in solid organ transplant recipient, 1163
    streptomycin, 1395
  in war veterans, S7
Tuberous sclerosis (Bourneville disease)
  brain tumors in, 720, 723
  clinical features of, 397, 717t, 723, 2266, 2430t, 2434, 3652
  genetic considerations in, 524t, 687t, 717t, 723, 2430t, 2434
  MRI in, A16
  NETs and, 686
  pathogenesis of, 723, 2434
  RCC and, 687t, 2434
  skin manifestations of, 397, 404
  treatment of, 723, 2434
Tuberous xanthoma, 404, A15
Tublin, 687t, 2790t
Tuboovarian abscess, 1365, 1365f, 1376
Tubuloglomerular feedback, 2364, 2371f, 2373, 2396, 2396f
Tubulointerstitial diseases, **2437**
  acute interstitial nephritis. *See* Acute interstitial nephritis (AIN)
  autosomal dominant, 2435
  chronic, 2437t, **2440**
    analgesic nephropathy, 2442
    aristolochic acid nephropathy, 2377, 2442, 2444
    calcineurin-inhibitor nephropathy, 2410, 2443
    clinical features of, 2437
    in glomerulonephritis, 2442
    heavy metal nephropathy, 2443
    karyomegalic tubulointerstitial nephritis, 2436, 2442
    lithium-associated nephropathy, 2442–2443
    sickle cell nephropathy, 2415t, 2427, 2442
    vesicoureteral reflux and reflux nephropathy, 2441–2442, 2441f

  chronic/noninflammatory, 2356–2357
  classification of, 2437t
  global considerations in, 2443–2444
  hypokalemic nephropathy, 2444
  metabolic, 2437t, 2443
  uric acid nephropathy. *See* Uric acid nephropathy
Tubulointerstitial nephritis with uveitis (TINU), 2356, 2439, 2439f
Tucatinib, 533t, 569t, 570, 641, 658
TUG (Timed Up and Go) test, 180
Tularemia, **1335**
  animal bite-related, 1143
  approach to the patient, 1337
  in bioterrorism, S4
  clinical features of, 145t, 1051, 1337–1338, 1338f, 1338t, A1, S4
  complications of, 1340
  diagnosis of, 1338–1339, 1339–1340f, S4
  epidemiology of, 1336–1337
  etiology of. *See Francisella tularensis*
  global considerations in, 1336, 1341
  pathogenesis of, 1337
  pathology of, 1339
  prevention of, 1341, S4
  prognosis of, 1340–1341
  septic shock in, 990
  treatment of, 1339–1340, 1340t, S4
  type A, 1336, 1336f, 1340–1341
  type B, 1336, 1336f, 1341
Tula virus, 1661t, 1676
Tullio phenomenon, 245
Tumbu fly, 3732
Tumor, 528. *See also specific types and sites*
Tumor (sign of inflammation), 454f
Tumor (skin lesion), 377t
Tumoral calcinosis, 2390f, 2391, 3265, 3318–3319
Tumor burden, 499, 501, 3983
Tumor heterogeneity, 525–526, 527f, 574
Tumor-induced osteomalacia, 737t, 740, 3264
Tumor-infiltrating lymphocytes, 2785
Tumor lysis syndrome
  AKI in, 2377, 2379t, 2381
  clinical features of, 593, 2377
  hyperkalemia in, 358
  hypocalcemia in, 365
  treatment of, 593–594, 594f
  uremic syndrome in, 2361f
Tumor markers, 504, 504t, 733
Tumor necrosis factor (TNF), 135, 135f
Tumor necrosis factor receptor(s), 539
Tumor necrosis factor (type III) receptor family, 2773
Tumor necrosis factor-α (TNF-α)
  adverse effects of, 754
  in anemia of inflammation, 767–768, 768f
  in *Brucella* infections, 1332
  in depression, 3656
  elevated levels, in aging, 3872
  functions of, 462, 2791t
  in heart failure, 1980
  in hemophagocytic lymphohistiocytosis, 465
  in HIV infection, 1577
  monoclonal antibodies targeting, 2791t
  in MS, 3582
  in NTM infections, 1415
  in rheumatoid arthritis, 2845f, 2846
  source, target, and biologic activity of, 2766t
  in spondyloarthritis, 2881f, 2882
  in vasculitis syndromes, 2894
Tumor necrosis factor-α (TNF-α) inhibitors. *See also specific drugs*
  adverse effects of, 2566, 2850t
    antibody development, 2566
    cardiotoxicity, 2017
    cutaneous, 392, 420, 606, 2566
    encephalopathy, 2350t
    hepatosplenic T cell lymphoma, 2566
    immunosuppression, 432
    infections, 136, 2566, 2851
    latent tuberculosis reactivation, 2851
    non-Hodgkin's lymphoma, 2566
    periodic syndrome (TRAPS), 155, 462, 2935t, 2936
    pulmonary, 2261
    serious, 387–388, 387t, 2850t
  for autoimmune and inflammatory diseases, 2784, 2826
  contraindications to, 387, 463
  for IBD, 2565–2566
  monitoring during treatment with, 2850t
  for psoriasis or psoriatic arthritis, 387, 387t
  for rheumatoid arthritis, 2850t, 2851
Tumor necrosis factor-α receptor-associated periodic syndrome (TRAPS), 136, 155, 462, 1131, 1132t, 2935t, 2936
Tumor necrosis factor-β (TNF-β), 2766t
Tumor neoantigens, 2785
Tumor-suppressor genes
  definition of, 520
  epigenetic silencing of, in cancer, 538–539, 538f
  inactivation of, 522
  in lung cancer, 617
  role of, 522
  in thyroid cancer, 3045–3046
Tumor vaccines, 561
Tumstatin, 548f
Tunari virus, 1661t, 1674
Tunga penetrans, 3736
Tungiasis, 3736
Tungsten carbide exposure, 2238
Tunica media, 1845, 1845f
Tunneled catheters, 2399–2400, 2401, 2402
Turcot's syndrome, 654, 654t, 717t, 721
Turkey handling disease, 2228t
Turner sign, 2745
Turner's syndrome, **3097**
  aortic aneurysm in, 2165
  aortic valve disease in, 2036
  autoimmune polyendocrine deficiencies in, 3091t, 3093
  clinical features of, 3097, 3097t
  genetic considerations in, 3791
  IBD and, 2553t
  lymphedema in, 2183
  pathophysiology of, 3097
  prevalence of, 2979t
  primary ovarian insufficiency in, 3134
  screening/testing for, 2979t
  treatment of, 3097–3098
Tuskegee study, 67
T wave, 1868, 1870, 1912
Twiddler's syndrome, 1924
Twinkle gene mutations, 3645
Two-point discrimination test, 174, 3382
Two-spirit, 3180t
Tylosis palmaris et plantaris, 642
Tympanic membrane, 244, 245, 245f, 248, 254
Tympanic membrane perforation, 245
Tympanic membrane thermometer, 134
Tympanogram, 249
Tympanometry, 249
Tympanoplasty, 245
Tympanostomy tubes, 245, 254–255
Typhilitis (neutropenic enterocolitis), 582, 596, 597f, 1376
Typhoidal tularemia, 1338
Typhoid fever. *See* Enteric (typhoid) fever
Typhoid vaccines, 1014t, 1015, 1318–1319
Typhus
  endemic. *See* Endemic (murine) typhus
  epidemic. *See* Epidemic (louse-borne) typhus
  scrub. *See* Scrub typhus
Tyramine, 3403
Tyrosinase, 430, 2790t
Tyrosinase gene mutations, 396
Tyrosine hydroxylase, 2790t
Tyrosine kinase inhibitors (TKIs). *See also specific drugs*
  actions and targets of, 535
  adverse effects of, 420, 421, 614, 754, 756, 840–841, 2017
  for ALL, 737t, 848
  for CML, 839t, 840–841
  cost of, 843
  discontinuation of, 841
  global considerations in, 843, 843f
  for head and neck cancer, 614
  for lung cancer, 627–628, 627t
  for NETs, 685–686
  non-receptor-linked, 567, 568t
  in pregnancy, 842–843
  receptor-linked, 567, 568–569t, 572
  resistance to, 841
Tyrosine kinase receptors, 2983f, 2984
Tyrosine metabolism disorders, 3368t
Tyrosinemia, 3368t
Tzanck smear, 381

U

Ubiquitin, 3513, 3519
Ubiquitination, 971
Ublituximab, 3586t
Ubrogepant, 3467t, 3468–3469, 3468t
UDCA. *See* Ursodeoxycholic acid (UDCA)
Udd distal myopathy, 3640t, 3643–3644
UDP (uridine diphosphate) glucuronosyltransferase, 321, 2637f, 2638
UDPGT gene mutations, 323
UGT1A1
  in bilirubin metabolism, 2637f, 2638
  in Crigler-Najjar syndrome, 2640
  drug disposition and, 485t
  drug response and, E72t
  frequency of variants in, E72t
  in Gilbert's syndrome, 493, 2640
  in neonatal jaundice, 2639
UGT1A1 gene, 485t, 2638, 2638f
UGT1A1 gene mutations, 798, 2729
UGT1 gene complex, 2638, 2638f, 2640
UGT2B17 gene, 3174
UGT gene mutations, 658
Ularitide, 1991, 1993t
Ulcer(s). *See also specific types*
  cutaneous. *See* Cutaneous ulcers
  definition of, 2514, 2518
  foot, 3228–3229. *See also* Diabetes mellitus, complications of
  genital. *See* Genital ulcer(s)
  leg, 775, 778t
  oral, 261–262, A3
  pressure, 1051, 3570
Ulcerative colitis. *See also* Inflammatory bowel disease (IBD)
  classification of, 2557t
  clinical features of, 309, 2556–2557, 2557t
  colorectal cancer in, 655, 2570, 2570f
  complications of, 2557
  diagnosis of, 2557, 2557f, 2557t
  differential diagnosis of, 1306, 2559–2561, 2560t
  epidemiology of, 2551–2552, 2551t
  etiology of, 2553
  familial patterns of, 2552
  genetic considerations in, 2553–2554, 2553–2554t
  global considerations in, 2551
  immune regulation in, 2554–2555
  inflammasome mutations in, 2761–2762t

- Ulcerative colitis (*Cont.*):
  - pathogenesis of, 2781
  - pathology of, 2555–2556, 2555–2556f
  - in pregnancy, 2569–2570
  - racial and ethnic differences in, 2551
  - smoking effects on, 2552
  - treatment of. *See* Inflammatory bowel disease (IBD), treatment of
  - tumor markers in, 504t
- Ulipristal acetate, 3154
- Ulnar neuropathy, 3613
- Ultradian rhythm, 3944t
- Ultrafiltration, 1991, 2399
- Ultra-rapid metabolizers, 491, 492f
- Ultrasonography
  - in appendicitis, 2596
  - in bile duct disease, 2736t
  - in bone mass measurement, 3297
  - in cholestatic conditions, 324
  - in chronic venous disease, 2181
  - of deep leg veins, 2160, 2160f
  - in echinococcosis, 1836f, 1837f
  - endobronchial, 620
  - endoscopic. *See* Endoscopic ultrasound (EUS)
  - in fever of unknown origin, 153
  - in gallbladder disease, 2730, 2730t, 2731f
  - for HCC screening, 661
  - in infertility evaluation, 3150
  - intravascular coronary artery, 1908–1909, 1909f
  - in liver disease evaluation, 2631–2632, 2631t, 2636
  - in lymphadenopathy, 478
  - in musculoskeletal disease evaluation, 2947
  - in pancreatic disease evaluation, 2739t, 2741
  - in peripheral arterial disease, 2172
  - pleural, 2208
  - point-of-care. *See* Point-of-care ultrasound (POCUS)
  - in psoriatic arthritis, 2888
  - of renal arteries, 2151t
  - in renal disease evaluation, 2363
  - in rheumatoid arthritis, 2848, 2848f, 2854v
  - scrotal, 705
  - small-bowel, 2559
  - in splenomegaly, 480
  - in stroke, 3438
  - in thyroid dysfunction evaluation, 3027
  - in thyroid nodule evaluation, 3027, 3042f, 3043f, 3048
  - transvaginal, 517t, 518
  - in urinary tract obstruction, 2455
- Ultraviolet light therapy. *See* Phototherapy
- Ultraviolet radiation, 429–430. *See also* Sun exposure
- Umami taste sensation, 238
- Umbilical cord stem cells, 915
- Umbilical venous hum, 327
- Umbralisib, 534t
- Umbrella cells, 331
- *UMOD* gene mutations, 2435
- UNAIDS (Joint United Nations Program on HIV/AIDS), 3847
- Una virus, 1657t
- *UNC13D* gene mutations, 465
- Uncal transtentorial herniation, 187, 187f
- Underweight, 3188t
- Undiagnosed disease, **3995**
  - approach to the patient, 3999f
  - challenges in, 3999
  - data collection for, 3996, 3996t
  - effect on patient, 3995
  - exposure assessment in, 3998
  - factors contributing to, 3995t
  - genomics in, 3997–3998
  - periodic reevaluation in, 3997
  - rare, 3995
  - research approaches in, 3998–3999
  - testing strategies and new technologies, 3996, 3996t, 3997t
  - validation of subjective and objective findings in, 3996–3997
- Undifferentiated pleomorphic sarcoma, 727
- Undulant fever. *See Brucella* spp. infections
- Unenveloped viruses, 1479
- *Unequal Treatment,* 64
- Unfractionated heparin. *See* Heparin
- Unintentional weight loss. *See* Involuntary weight loss
- Uniparental disomy, 715, 3791
- Unipolar depressive disorders, 3665. *See also* Depression/depressive disorders
- Uniporters, 2366
- United Nations, 3844
- United Network of Organ Sharing (UNOS), 2720, 2720t
- United States President's Emergency Plan for AIDS Relief (PEPFAR), 3843, 3855
- United States Preventive Services Task Force, 39t, 40, 40t
- Universal coverage reforms, 3863, 3863f
- University of Pennsylvania Smell Identification Test, 240f, 243
- Unstable angina, 103t, 104, 2106. *See also* Non-ST-segment elevation acute coronary syndrome (NSTE-ACS)
- Unstable hemoglobins, 779t, 780
- Unverricht-Lundborg disease (Baltic myoclonus), 3412t, 3516, 3530
- Upadacitinib
  - adverse effects of, 2568t
  - for atopic dermatitis, 384
  - for axial spondyloarthritis, 2887t
  - for IBD, 2567, 2568t
  - for IBD-associated arthritis, 2892
  - for psoriatic arthritis, 2890
  - for rheumatoid arthritis, 2851t
- Upbeat nystagmus, 237
- UPC-LFA, 1410
- UPOINT, 334
- Upper airway cough syndrome, 273
- Upper airway neurostimulation, 2282
- Upper esophageal sphincter, 293
- Upper motor neuron diseases, 3522–3523t, 3525, 3597
- Upper motor neuron pathways, 169f
- Upper motor neuron weakness, 168, 168t, 169
- Upper respiratory infections (URIs). *See* Respiratory tract infections
- Upshaw-Schulman syndrome, 925, 2445
- Uranium-235 exposure, S6
- Urate nephropathy (urate nephrosis), 2440
- Urate oxidase. *See* Rasburicase
- Urea, 3020
- Urea breath test, 1301, 1301t, 2524t
- Urea cycle, 3348–3349, 3372f
- Urea cycle defects, 3369t, **3371**
- *Ureaplasma* infections, 1098, 1159t, 1466–1467, 1466t
- Urea, urinary, 2621–2622
- Uremia
  - abdominal pain in, 111
  - in AKI, 2382
  - peripheral blood smear in, 440f, 446f
  - polyneuropathy in, 3605
- Uremic acidosis, 372
- Uremic fetor, 2394
- Uremic syndrome, 2360, 2361f, 2362, 2388–2389. *See also* Chronic kidney disease (CKD)
- Ureteral obstruction, 110
- Ureteral reflux, 2361
- Ureterectomy, 694
- Ureteric bud, 2363, 2364f
- Urethritis
  - in men
    - approach to the patient, 1098–1099
    - *C. glucuronolyticum,* 1226
    - chlamydial, 1470, 1473t
    - clinical features of, 1098, 1255–1256
    - etiology of, 1087, 1097t, 1098, 1255–1256. *See also* Sexually transmitted infections (STIs)
    - gonococcal, 1255–1256, 1255f
    - nongonococcal, 1470, 1473t
    - postgonococcal, 1473t
    - treatment of, 143, 1099t
  - *Mycoplasma,* 1467
  - *Ureaplasma,* 1467
  - in women, 343, 1099–1100, 1257, 1471–1472, 1473t
- Urge incontinence, 3891, 3893f, 3893t
- Uric acid
  - decreased excretion of, 2368t, 2453
  - increased production of. *See* Hyperuricemia
  - laboratory evaluation of, 2945
  - tissue deposition of (tophi), 2948f
  - urine, 2450, A4
- Uric acid nephropathy
  - in AML, 829
  - hyperuricemia and, 2443
  - pathophysiology of, 2443
  - treatment of, 2443
- Uric acid stones, renal, 2448, 2453. *See also* Nephrolithiasis
- Uridine diphosphate galactose 4-epimerase deficiency (galactosemia type III), 3361t, 3366
- Uridine diphosphate-glucuronosyl transferase (UDPGT), 321, 323
- Urinalysis
  - in abdominal pain evaluation, 113
  - in AKI, 2380–2381, 2380f
  - atlas of, A4
  - in hematuria, 342–343. *See also* Hematuria
  - in IC/BPS, 333
  - in nephrolithiasis, 2450, 2451f
  - parasites in, S12
  - in proteinuria, 341–342. *See also* Proteinuria
  - specimen collection and transport, S11
  - in urinary tract obstruction, 2455
  - in UTIs, 1091–1092
- Urinary alkalinization, for overdosage/poisoning, 3709
- Urinary analgesics, 1094
- Urinary diversion, 694
- Urinary electrolyte free-water clearance, 351t, 353
- Urinary incontinence
  - after prostatectomy, 699
  - in older adults, 3889–3891, 3890t, 3891–3893f, 3893t
  - postmenopausal hormone therapy and, 3144t
  - in terminally ill patient, 91t
- Urinary tract
  - bleeding from. *See* Hematuria
  - congenital abnormalities of, 2436, 2453–2454, 2454t
- Urinary tract infections (UTIs), **1086**
  - approach to the patient, 1088–1089
  - in cancer patient, 583–584
  - catheter-associated, 1087, 1090, 1094, 1147, 1148t
  - clinical features of, 1089–1090
  - complicated, 1086–1087
  - cystitis. *See* Cystitis
  - definitions in, 1086–1087
  - delirium in, 184
  - in diabetes mellitus, 3227, 3230
  - diagnosis of, 1090–1091, 1091f
  - epidemiology of, 1087

- etiology of, 1087–1088
  - *A. baumannii,* 1296
  - *Citrobacter,* 1293
  - *C. urealyticum,* 1226
  - *E. coli,* 1087, 1286, 1293
  - enterococcal, 1217, 1219t, 1220t
  - *K. pneumoniae,* 1291, 2453
  - *Morganella,* 1293
  - nontyphoidal *Salmonella,* 1320
  - *P. aeruginosa,* 1293, 1309t, 1312
  - *Proteus,* 1291, 1293, 2453
  - *Providencia,* 1293, 2453
  - *S. aureus,* 1201
  - *S. saprophyticus,* 1202
  - tuberculous, 1388
- genetic considerations in, 1088
- in IC/BPS, 330–331, 334
- vs. IC/BPS, 333
- in men, 1087, 1093
- in MS, 3588
- nephrolithiasis and, 2453
- pathogenesis of, 1088, 1088f
- in pregnancy, 1094–1095, 3903
- prevention of, 1095
- prognosis of, 1095
- pyelonephritis. *See* Pyelonephritis
- recurrent, 1087, 1095
- renal abscess and, 1075
- risk factors for, 1087
- in solid organ transplant recipient, 1158, 1159t
- treatment of
  - antibiotic stewardship in, 1152t
  - asymptomatic bacteriuria, 1092
  - *Candida* infection, 1095
  - catheter-associated, 1094
  - considerations for antimicrobial therapy, 1092
  - cystitis, 1092–1094, 1093t
  - drug resistance and, 1087–1088, 1092
  - *E. coli,* 1287
  - in pregnancy, 1094–1095
  - prostatitis, 1094
  - pyelonephritis, 1094
- uncomplicated, 1086
- urinary tract obstruction in, 2455

Urinary tract obstruction, **2453**
- AKI and, 340, 2378, 2380f
- in cancer, 588
- clinical features of, 338t, 588, 2454–2455
- diagnosis of, 338t, 2455–2456, 2456f
- diuresis following, 2457
- etiology of, 2454–2455, 2454t
- fetal, 2454
- pathophysiology of, 2454–2455, 2455t
- prognosis of, 2456–2457
- treatment of, 588, 2384, 2456

Urinary D-xylose test, 2550

Urination, painful
- in cystitis. *See* Cystitis
- in urinary tract infections. *See* Urinary tract infections (UTIs)

Urine
- abnormalities of, 338t, **341**. *See also* Urinalysis
  - casts, 343, A4
  - crystals, 3373, A4
  - hematuria. *See* Hematuria
  - polyuria. *See* Polyuria
  - proteinuria. *See* Proteinuria
  - pyuria, 343
- nephrolithiasis risk and, 2449–2450
- osmolality of, 340t, 343, 350, 353
- sodium loss in, 347, 350
- tea/cola colored, 320, 2354

Urine bilirubin, 321, 2634, 2637t, 2638

Urine culture, 1092, 1152–1153, 1152t

Urine dipstick test, 321, 1091

Urine microscopy, 1092

Urine output, 2312, 2598

Urobilinogens, 321, 2638

Urobilins, 321

Urocanic acid, 432

Urocanic aciduria, 3368t

Urochromes, 2394

URO-decarboxylase, 3342t, 3343, 3343t, 3348

UROD gene mutations, 435, 3347, 3348

Urogenital secretions, specimen collection and transport, S11

Urography, 2456

UROIII, 733

Urokinase, 955

Urokinase-type plasminogen activator (u-PA), 471, 471f

Urologic chronic pelvic pain syndrome (UCPPS), 330

Uromodulin (Tamm-Horsfall protein), 342, 343, 2377, 2435, 2440

Uropathogenic *E. coli* (UPEC), 965, 966t. *See also Escherichia coli* infections, extraintestinal

UROS gene, 3342t

URO-synthase, 3342t, 3343, 3349

Urothelial (transitional cell) carcinoma, 691, 692f, 694. *See also* Bladder cancer

Urothelial dysfunction, 331

Urschel's sign, 2158

Ursodeoxycholic acid (UDCA)
- for gallstone dissolution, 2732
- for gallstone prevention during rapid weight loss, 2729
- for primary biliary cholangitis, 2711
- for primary sclerosing cholangitis, 2712

Urticaria, **402, 2807**
- in anaphylaxis, 2812
- cholinergic, 402, 2807, 2808
- chronic idiopathic, 2790t
- chronic spontaneous, 422
- classification of, 2807, 2807t
- clinical features of, 147, 147f, 379t, 394, 402–403, 2807–2808
  - distribution, 379t
  - papules and plaques, 382f, A1, A5
- cold, 402, 462, 2807
- definition of, **2807**
- dermographic, 2806, 2807f, A5
- diagnosis of, 382, 2808
- drug-induced, 417, 417t, 422
- etiology of, 147–148, 402, 402t, 2807, 2807t
- in mastocytosis, 2820. *See also* Mastocytosis
- pathophysiology of, 2807–2808
- physical, 402
- predisposing factors for, 2807
- treatment of, 422, 2808–2809

Urticarial vasculitis, 143t, 147–148, 403

Urticaria pigmentosa, 402, 406

Urushiol, 384

U.S. *See* United States

USB1 gene mutations, 3820t

Usher's syndrome, 248, E36t

USP8 gene mutations, 3004

Ustekinumab, 387t, 2566, 2568t, 2890, 2892

Usutu virus, 1660t, 1671

UTAF1 gene mutations, 828

Uterus, disorders of
- amenorrhea in, 3132–3133, 3134f
- clostridial myonecrosis (gas gangrene), 1053
- endometrial cancer. *See* Endometrial cancer
- endometriosis. *See* Endometriosis
- fibroids (leiomyoma), 714, 3136, 3137, 3152
- pelvic pain in. *See* Pelvic pain
- sarcoma, 715
- treatment of, 3133

UTIs. *See* Urinary tract infections (UTIs)

UTX gene mutations, 3941

Uukuniemi virus, 1663t

Uukuviruses, 1663t

Uveitis
- anterior, 225
- in Behçet syndrome, 2909
- in IBD, 2562t
- posterior, 225
- in psoriatic arthritis, 2888
- in sarcoidosis, 2927
- in secondary syphilis, 1431
- with tubulointerstitial nephritis, 2439
- in Whipple's disease, 1369, 1370

Uvula, bifid, 1860

Uvulopalatopharyngoplasty, 2281

U wave, 1868, 1870

V

Vaborbactam, 1175, 1183. *See also* Meropenem-vaborbactam

VacA, 1299

Vaccine(s). *See also* Immunization; *specific vaccines*
- adjuvants to, 996, 997t, 998
- administration of, 1008, 1008f
- adverse event reporting, 19, 1001, 1009
- approach to the patient, **17,** 18t
- assessment of safety of, 1001
- for cancer, 551
- in cancer patient, 585
- cancer prevention and, 513–514, 3829
- categories and characteristics of, 996, 997t
- contraindications to, 1003, 1005, 1006–1007t
- culturally safe care and, 19
- development of, **998,** 999f
  - Advisory Committee on Immunization Practices process in, 1000, 1000f
  - conceptual and technological advances in, 997–998, 998f
  - controlled human infection models in, 999
  - genomics in, 983–984, 987
  - phases of, 998–999
- documentation of, 1008–1009
- future trends in, 1010
- global considerations in, 1003
- in health promotion, 13
- hypersensitivity to, 1005
- immune responses to, 996–997
- impact of, 1001–1003, 1002t
- inequities in access to, 19
- information statements, 1005, 1007
- injury compensation program for, 1001
- licensure of, 999–1000
- myocarditis and, 2015
- policy and postlicensure surveillance, 1000
- precautions for, 1006–1007t
- reduced response to, 579
- routinely administered in U.S., 1002t
- safety monitoring, 1009
- storage and handling of, 1007–1008
- tumor, 561

Vaccine Adverse Event Reporting System (VAERS), 1001, 1009

Vaccine and Related Biological Products Advisory Committee, 999–1000

Vaccine-induced immune thrombocytopenia and thrombosis (VITT), 923, 924

Vaccine Information Sheet (VIS), 1005, 1007

Vaccine opposition and hesitancy
- approach to the patient, **17,** 18t, 1001
- causes of, 14
- COVID-19 pandemic and, 16
- definition of, 14
- ethical issues for physicians, 70
- factors in, 14–15, 15f, 1001
- public health threats and, 1149
- social and cultural trends affecting, 16–17

- Vaccine-preventable disease, 1002, 1002t
- Vaccine Safety Datalink Project, 1001
- Vaccines for Children (VFC) program, 1000, 1002
- Vaccinia immune globulin, 1521
- Vaccinia virus, 1518, 1518t, 1519, 1521
- Vaccinia virus vaccine, 1521
- Vacuoles, E1 enzyme, X-linked, autoinflammatory, somatic (VEXAS) syndrome, 150, 2937
- Vacuum constriction device, 3160
- VA-ECMO (venoarterial extracorporeal membrane oxygenation), 2309, 2309t, 2328
- VAERS (Vaccine Adverse Event Reporting System), 1001, 1009
- Vagabonds' disease, 3732
- Vagal reflex maneuvers, for tachycardias, 1933f, 1934t, 1938, 1939, 1942
- Vagina
  - atrophy of, 3161
  - blind, 3132
  - cancer, 1526
  - congenital absence of, 3101
  - dilatation of, 3133
  - discharge from, 1101–1102. *See also* Vulvovaginal infections
  - microbiota of, 1088, 3833f, 3839
- Vaginal ring, 3152
- Vaginal septum, 3133
- Vaginoplasty, 3101
- Vagotomy, 2532–2533
- Vagus nerve, 239, 239f, 3380, 3398, 3556–3557, 3557f
- Vagus nerve stimulation
  - for cluster headache, 3472t
  - for depression, 3667
  - for heart failure, 1998
  - for migraine prevention, 3470t
  - for migraine treatment, 3467t, 3468t, 3469
  - for motor recovery after stroke, 3968
  - for refractory epilepsy, 3425
- Valacyclovir
  - actions of, 1485
  - adverse effects of, 1485, 1486t
  - for Bell's palsy, 3555
  - for CMV prophylaxis, 1515
  - for herpes zoster, 1054t, 1485, 1486t, 1507, 3566
  - for HSV esophagitis, 2513
  - for HSV infections, 1502–1504, 1503t
    - genital, 1054t, 1109t, 1485
    - mucocutaneous, 1486t, 1537
  - for HSV prophylaxis, 1502, 1504, 1595t
  - pharmacology of, 1485
  - for varicella, 1507
  - for viral encephalitis, 1115
  - for viral meningitis, 1126
- Valbenazine, 3513, 3517, 3674
- Valdecoxib, 97t
- Valganciclovir
  - for CMV infection, 2410, 2513
    - prophylaxis, 1485, 1594t
    - treatment, 1485, 1486t, 1515
  - indications for, 1485–1486
  - pharmacology of, 1485
  - resistance to, 1485
- Valgus, 2944, 2951, 2951f
- Valine metabolism disorders, 3369t
- Valley fever. *See* Coccidioidomycosis
- Valoctocogene roxaparvovec, 3824t
- Valproate/valproic acid
  - adverse effects of, 3470t
    - bone marrow suppression, 3423
    - common, 3670t
    - hair loss, 421
    - hepatotoxicity, 2665, 2670, 3423
    - mitochondrial defects, 3816
    - neurologic, 3423t
    - rare, 3670t
    - steatosis, 2703t
    - systemic, 3423t
    - teratogenicity, 3423, 3903
    - thrombocytopenia, 923t
    - weight gain, 3188
    - zinc deficiency, 2613
  - for bipolar disorder, 3668, 3670t
  - drug interactions of, 1172t, 3422t, 3423t
  - for migraine prevention, 3469, 3470t
  - for myoclonus, 3516
  - for new daily persistent headache, 117
  - overdosage/poisoning with, 3713t
  - for panic disorder, 3658
  - pharmacology of, 3423t, 3670t
  - for posthypoxic myoclonus, 2348
  - for seizures, 591, 717, 3423, 3423t
  - for status epilepticus, 3426f
  - for tardive dyskinesia, 3517
- Valrubicin, 693
- Valsalva maneuver
  - in autonomic dysfunction evaluation, 3541–3542, 3543f
  - in focal atrial tachycardia diagnosis, 1938
  - heart murmurs and, 289, 1867, 1867t
  - for PSVT, 1942
- Valsalva ratio, 3542, 3543f
- Valsartan, 1994, 1995t, 2144t
- Valve area, measurement of, 1906–1907
- Valvular heart disease. *See also specific conditions*
  - cardiac imaging in, 1890–1893
  - dyspnea in, 270, 271
  - exercise intolerance in, 291
  - global burden of, 2033–2035, 2034f
  - multiple and mixed, **2065**
  - murmurs in, **283**
  - natural history of, 1841
  - in polycystic kidney disease, 2432
  - in pregnancy, 3900
  - radiation-induced, 753
  - sudden cardiac death and, 2334, 2334f
  - traumatic, 2088
- VAMP2 (synaptobrevin), 1230
- *vanA* gene, 1204
- van Buchem's disease (hyperostosis corticalis generalisata), 3316
- Vancomycin
  - actions of, 1166, 1176, 1182t
  - adverse effects of, 923t, 1171t, 1176
    - cutaneous, 420, 422
    - DIHS, 424
    - fever, 150
    - infusion syndrome, 422
    - mechanisms of, 422
    - nephrotoxicity, 2377
  - for asymptomatic GAS carriage, 1209
  - for CAPD peritonitis, 1072
  - for *C. difficile* infections, 583, 1084–1085, 1085t
  - for clostridial infections, 1239t, 1243
  - for community-acquired pneumonia, 1028t
  - for corynebacterial infections, 1226
  - dosage monitoring, 487
  - efficacy of, 1168
  - empirical uses of, 963t, 2321t
  - for enterococcal infections, 1219t, 1220
  - for epidural abscess, 963t, 991t
  - indications for, 1173t, 1176
  - for infective endocarditis
    - as empirical therapy, 963t, 991t, 1045
    - enterococcal, 1041, 1043t, 1219t
    - staphylococcal, 1042, 1043t, 1205
    - streptococcal, 1043t
  - for intraabdominal infection, 963t
  - for intracranial infections, 991t
  - for mastoiditis, 255
  - for meningitis
    - as empirical therapy, 963t, 991t, 1121t, 1122, 1249
    - enterococcal, 1219t
    - MRSA, 1121, 1122t
    - pneumococcal, 1121, 1193
  - for MRSA infections, 1203t, 1204, 1205
    - as empirical therapy, 963t
    - endocarditis, 1042
    - meningitis, 1121, 1122t
    - skin, 1054t
  - for necrotizing fasciitis, 1054t
  - for nongonococcal bacterial arthritis, 1057
  - for osteomyelitis, 1064t, 1065t
  - for pneumococcal infections, 1193
  - for post-splenectomy sepsis, 991t
  - in pregnancy and lactation, 1169t
  - prophylactic uses of, 1180t
  - for purpura fulminans, 991t
  - resistance to. *See also* Vancomycin-resistant enterococci (VRE)
    - in *E. faecalis,* 1219, 1220t
    - mechanisms of, 1182t, 1183–1184, 1204, 1220
    - rates of, 1173t, 1176
  - for sepsis/septic shock, 963t, 991t
  - for staphylococcal infections, 1203t
  - for toxic shock syndrome, 991t, 1205
  - for ventilator-associated pneumonia, 1032t
- Vancomycin-intermediate *S. aureus* (VISA), 1044, 1176, 1184, 1204
- Vancomycin-resistant enterococci (VRE)
  - *C. difficile* fitness and, 1215
  - diagnosis of, 981
  - epidemiology of, 1216–1217
  - genomics of, 977t, 981, 1217
  - global considerations in, 1217
  - health care–associated, 1152, 1216–1217
  - laboratory identification of, S11
  - pathogenesis of, 1215–1216
  - resistance mechanism of, 1220–1221
  - in sepsis/septic shock, 2316
  - treatment of, 1042, 1072, 1219t, 1220
- van den Bergh reaction, 321
- Vandetanib
  - action and targets of, 533t, 570t
  - adverse effects of, 509, 570t
  - interactions and issues, 570t
  - for thyroid cancer, 3085
- *van* gene, 981, 1183–1184
- Vanishing bile duct syndrome, 325, 2667, 2671
- Vanishing (absent) testis syndrome, 3098
- *van* operon, 1221
- VAP. *See* Ventilator-associated pneumonia (VAP)
- Vaping
  - D9THC concentrates, 3686–3687
  - nicotine concentrates, 13, 511, 615, 3682–3683, 3685
- Vapreotide, 317
- Vaptans (vasopressin antagonists), 351, 738, S1
- Vardenafil, 503t, 704, 3155, 3156f, 3159, 3159t
- Varenicline, 615, 2255, 3685, 3685t, 3695
- Variable expressivity, 3788–3789
- Variable number of tandem repeats (VNTRs), 3793
- Variants of unknown significance (VUS), 525, 3784, 3787, 3802
- Variation, gene, 3783. *See also* Mutation(s)
- Variceal bleeding
  - esophageal. *See* Esophageal varices
  - gastric, 2482, 2488f, 2750, V5
- Varicella (chickenpox), **1505**
  - clinical features of, 142t, 147, 147f, 378t, 1505
    - extracutaneous, 1505
    - oral, 262t
    - skin

    characteristics of, 142t, 147, 147f
    evolution of, 1505, 1505f, A1, A5
    rash, 402
    vesicles/bullae, 401, 1049
  complications of, 1505, 3592, 3594
  congenital, 1505
  differential diagnosis of, 147, 1506
  epidemiology of, 142t, 1505
  incidence pre- and postvaccine, U.S., 1002t
  pathogenesis of, 1504
  perinatal, 1505
  prevention of, 1507, 1508t. *See also* Varicella vaccine
  in transplant recipient, 2350
  treatment of, 1485, 1486t, 1507
Varicella vaccine
  contraindications for, 1006t
  precautions for, 1006t
  recommended administration of
    in adults, 40t, 1004t, 1507
    in children, 1507
    in health care workers, 1153t
    in international travelers, 1015
    in pregnancy, 1004t
  storage and handling of, 1007
Varicella-zoster immune globulin (VZIG), 1507, 1508t, 1594t
Varicella-zoster virus (VZV), 1504, 1506–1507
Varicella-zoster virus (VZV) infections. *See also* Herpes zoster (shingles); Varicella (chickenpox)
  in cancer patient, 578
  diagnosis of, 381
  encephalitis, 1111, 1114t
  etiology of, 1504
  in HCT recipient, 918, 918t
  meningitis, 1125
  neuropathy in, 3607
  pathogenesis of, 1504
  prevention of, 1507, 1508t, 1593t
  treatment of, 1507
Varicella-zoster virus (VZV) vaccine, 1007, 1126, 1507
Varicocele, 3107
Varicose veins, 2180. *See also* Chronic venous disease
Variegated squirrel bornavirus 1, 1658t
Variegate porphyria (VP)
  clinical features of, 435, 3341t, 3348–3349, 3602
  diagnosis of, 3343t, 3349
  global considerations in, 3342
  pathophysiology of, 3341f, 3603
  vs. porphyria cutanea tarda, 435
  treatment of, 3349, 3603
Variola virus, 1517, 1518t, S4. *See also* Smallpox
Varioliform gastritis, 2539
Varus, 2944, 2951, 2951f
Vasa recta, 2364
Vasa vasorum, 1845
Vascular access-related infections. *See* Catheter-related bloodstream infections
Vascular addressins, 548f
Vascular cell adhesion molecule-1 (VCAM-1), 759, 1585, 2854
Vascular composite allografts, 1158
Vascular dementia, **3489**
  approach to the patient, 195, 3491
  cerebral small-vessel disease and, 3489–3490, 3490f
  cerebral strokes and, 3489
  clinical features of, 196
  global considerations in, 199, 3489
  risk factors for, 195
  treatment of, 3491
Vascular ectasias
  bleeding in, 318
  clinical features of, 406
  diagnosis of, 2482, 2490f
  treatment of, 2482, 2501f
Vascular endothelial growth factor (VEGF)
  in diabetic retinopathy, 3223
  functions of, 2792t
  in high-altitude illness, 3755–3756
  in kidney development, 2363
  monoclonal antibodies targeting, 2792t
  in tumor angiogenesis, 545–546, 545f, 546–548f
Vascular endothelial growth factor (VEGF) receptors, 545
Vascular mimicry, 546
Vascular resistance, 1906
Vascular smooth muscle, 1845f, 1846–1847, 1847f
Vascular tone, 1846–1847
Vasculitis. *See also* Vasculitis syndromes
  allergic, 462
  as cause of fever of unknown origin, 150, 151t
  definition of, 2893
  diagnosis of, 2894–2895
  differential diagnosis of, 2894, 2894t
  diseases associated with, 2907
  drug-induced, 426, 426f, 2907, 2940t
  in familial Mediterranean fever, 2934
  leukocytoclastic (cutaneous small-vessel). *See* Leukocytoclastic (cutaneous small-vessel) vasculitis
  necrotizing. *See* Necrotizing vasculitis
  neuropathy in, 3623–3624
  paraneoplastic, 751
  primary CNS
    brain biopsy in, A14
    clinical features of, 2906
    vs. Creutzfeldt-Jakob disease, 3532
    dementia in, 3484
    differential diagnosis of, 2906, 3447
    imaging in, 2906f, A14, A16
  rheumatoid, 2841
  secondary, 2893t, 2907
  single-organ, 2905
  in Sjögren's disease, 2877, 2877t
  skin manifestations of, 406, A5, A14
  treatment of, 2895
  urticarial, 143t, 147, 403
Vasculitis syndromes, **2893**
  approach to the patient, 2894–2895, 2895f
  Behçet syndrome. *See* Behçet syndrome
  classification of, 2893, 2893t
  Cogan's syndrome, 2166, 2906–2907
  cryoglobulinemic, **2905,** A14
  cutaneous, A14
  definition of, 2893
  diagnosis of, 2894–2895, 2895f, A14
  differential diagnosis of, 2894, 2894t
  EGPA. *See* Eosinophilic granulomatosis with polyangiitis
  giant cell arteritis. *See* Giant cell arteritis
  granulomatosis with polyangiitis. *See* Granulomatosis with polyangiitis
  idiopathic cutaneous, **2905**
  IgA vasculitis. *See* IgA vasculitis
  Kawasaki's disease. *See* Kawasaki's disease
  microscopic polyangiitis, 2415t, 2420, **2899,** A14
  pathogenesis of, 2893, 2893t
  polyangiitis overlap syndromes, 2907
  polyarteritis nodosa. *See* Polyarteritis nodosa
  polymyalgia rheumatica, 132, 155, 227, 2847, **2902,** 3636
  primary CNS. *See* Vasculitis, primary CNS
  Takayasu's arteritis, 2166, 2171, **2903,** 2903t, 3431, 3447, A14
  treatment of, 2895, 2895f, 2896t
  vessel damage in, 2893, 2893t
Vasculogenesis, 545
Vasculopathy. *See specific types*
Vasectomy, 3152–3153, 3153t
Vasoactive intestinal peptide (VIP), 293, 737t, 2743
Vasoactive intestinal peptide tumor (VIPoma). *See* VIPoma
Vasoconstriction
  cold-induced, 3769
  cyanosis in, 279
  heat conservation by, 134
  localized hypoxia and, 277
  in thrombosis, 921
Vasodilation
  as adaptation to hypoxia, 278
  effects of, 921
  in exercising muscle, 1852
  heat loss by, 134
Vasodilators
  actions of, 2133, 2145t
  adverse effects of, 2145t
  for heart failure, 1991, 1993t, 1995t
  for hypertension, 2145t
Vasogenic edema, 2342, 2349
Vasomotor rhinitis, 2810
Vasopressin. *See also* Arginine vasopressin (AVP)
  actions of, 2367f, 2370, 2986
  deficiency, 2996t
  ectopic production of. *See* Syndrome of inappropriate antidiuresis (SIAD)
Vasopressin agonist. *See* Desmopressin (DDAVP)
Vasopressin antagonists (vaptans), 351, 738, S1
Vasoreactivity testing, 2190
Vasospasm, after subarachnoid hemorrhage, 3459, 3461, 3461f
Vasospastic angina (Prinzmetal's variant angina), 1873, 2101, **2112**
Vasovagal syncope. *See* Syncope, neurally mediated
Vaughan-Williams classification, of antiarrhythmic drugs, 1915–1916, 1916t
VCAM-1 (vascular cell adhesion molecule-1), 759, 1585, 2854
VCL gene mutations, 2004t
VCTE (vibration-controlled elastography), 2705
Vd (volume of distribution), 1167
VDR (vitamin D receptor), 3269
VDRL (Venereal Disease Research Laboratory) test, 1433, 1602
Vecuronium, 2352
Vedolizumab, 2566, 2568t, 2791t, 2792t, 2892
Vegetarian diet, 349, 785, 3833–3834
Vegetations, valvular, 1036, 1036f, 1040, 3914v
Vegetative state, 186, 193
VEGF. *See* Vascular endothelial growth factor (VEGF)
VEGFA gene mutations, 660, 661t
VEGFR2 gene mutations, 691
VEGFR3 gene mutations, 2183
Veins. *See* Blood vessels
"Velcro" crackles/rales, 754, 2868
Vellus hair, 3137
Velocardiofacial (DiGeorge's) syndrome, 2803, 3286, 3652, S9
Velpatasvir/sofosbuvir, 1492t, 1493, 2688t, 2690–2691
VEMP (vestibular-evoked myogenic potential test), 250
Vemurafenib
  actions and targets of, 527, 533t, 569t, 572
  adverse effects of, 400, 569t, 605
  drug interactions of, 591
  genetic variations in response to, 493
  for hairy cell leukemia, 873
  interactions and issues, 569t
  for melanoma, 572, 604, 604t, 605
  relapse following treatment with, 490, 494
  resistance to, 574

Venereal Disease Research Laboratory (VDRL) test, 1433, 1602
Venetoclax
  action and targets of, 533t, 540, 571t
  adverse effects of, 571t, 593, 855
  for ALL, 855
  for AML, 831, 833t
  for CML, 830
  for cold agglutinin disease, 803
  interactions and issues, 571t
  for mantle cell lymphoma, 866
  for MDS, 818
Venezuelan equine encephalitis viruses, 1657t, 1673
Venlafaxine
  adverse effects of, 3176, 3604t, 3659t, 3667
  for anxiety disorders, 3661
  for cataplexy, 215
  for depression, 85, 3588, 3659t, 3667
  for hot flashes, 638
  for hot flushes with androgen deprivation therapy, 3176, 3176t
  for menopausal symptoms, 3143
  for migraine prevention, 3470t
  for neuropathic pain, 101, 3227, 3604t
  for pain, 97t, 100, 3588
  for panic disorder, 3658
  pharmacology of, 3667
  for PTSD, 3663
Venoarterial extracorporeal membrane oxygenation (VA-ECMO), 2309, 2309t, 2328–2329
Venomous snakebite. *See* Snakebites
Venoocclusive disease with immunodeficiency (VODI) syndrome (Sp110 deficiency), S9
Venous anomalies, 3457
Venous disorders. *See* Chronic venous disease; Lymphedema; Venous thromboembolism (VTE)
Venous hum, 288–289
Venous hypertension, 396, 406
Venous lake, 406
Venous malformation, 406
Venous pressure, 282
Venous return, 289
Venous sinus thrombosis, 3447
Venous thoracic outlet syndrome, 133
Venous thromboembolism (VTE), **939**. *See also* Deep venous thrombosis (DVT); Pulmonary embolism (PE)
  abdominal/pelvic, S13
  in antiphospholipid syndrome, 2837, 2838t
  vs. arterial thrombosis, 941
  atherothrombosis and, 2157–2158, 2158f
  in β thalassemia, 778t
  with brain tumor, 717
  coagulation cascade in, 940
  epidemiology of, 939–940, 2154–2156
  etiology of, 939–940, 940t
  genetic considerations in, 939t, 941
  history in, 473
  in HIV infection, 1605
  incidence of, 537
  in liver disease, 935, 935t, 936f
  lupus anticoagulant in, 936
  mesenteric, 2587, 2588t
  oral contraceptives and, 3153
  paraneoplastic, 741–742, 742t
  pathophysiology of, 2156–2158, 2158f
  polygenic risk score for, 2156
  postmenopausal hormone therapy and, 3144t, 3145
  postthrombotic syndrome, 2156, 2157f, 2157t, 2162
  prevention of
    anticoagulant dosage for, S13
    antiplatelet therapy for, S13
    fondaparinux for, 949
    in hospitalized patients, 2164, 2164t
    in surgical patient, 3910
  prothrombotic states and, 2156
  renal, 2447
  risk factors for, 473t, 2156
    acquired, 940–941, 940t
    age, 474f
    in cancer patient, 742, 742t
    history of thrombosis, 473
    inflammation-linked, 2157t
    inherited, 939t, 941
    travel, 1019
Venovenous extracorporeal membrane oxygenation (VV-ECMO), 2309, 2309t
Ventilation-perfusion mismatch/heterogeneity, 277, 2190, 2204–2205, 2205f
Ventilator-associated pneumonia (VAP), **1029**
  A. baumannii, 1296
  clinical features of, 1030
  complications of, 1032–1033
  diagnosis of, 1030–1031
  epidemiology of, 1030
  etiology of, 1029, 1029t
  failure to improve, 1032
  follow-up, 1033
  MRSA, 1032
  P. aeruginosa, 1031–1032, 1310
  pathophysiology of, 1030, 1030t, 1147, 2307–2308
  prevention of, 43, 44, 1030t, 1033, 1147–1148, 2308
  prognosis of, 1033
  treatment of, 1031–1032, 1032t, 2308
Ventilatory function
  adequacy of, 2204
  disturbances of. *See* Hyperventilation; Hypoventilation; Respiratory disorders
  flow-related properties of, 2202–2203, 2203f
  measurement of, 2206–2207
  physiology of, 2273, 2274f
  volume-related properties of, 2201–2202, 2202f
  work of breathing, 2203–2204
Ventral respiratory column, 2273
Ventricular afterload, 1851, 1852f
Ventricular aneurysm, 2125
Ventricular arrhythmias, **1954**
  after STEMI, 2123–2124
  approach to the patient, 1956–1957
  characteristics of, 1954–1956, 1955–1956f
  clinical features of, 1956
  diagnosis of, 1956–1957, 1958f
  origins of, 1957f
  premature ventricular beats. *See* Premature ventricular contractions (PVCs)
  treatment of, 1957–1959, 1959f
  ventricular fibrillation. *See* Ventricular fibrillation (VF)
  ventricular tachycardia. *See* Ventricular tachycardia (VT)
Ventricular couplets, 1954
Ventricular depolarization, 1869, 1870f
Ventricular dysplasia, arrhythmogenic, 1961f, 2001
Ventricular end-diastolic pressure, 1850
Ventricular end-diastolic volume, 1850, 1852f
Ventricular fibrillation (VF)
  in chest trauma, 2088
  ECG in, 1956f
  etiology of, 1971
  in hypothermia, 3769t
  mechanism of, 1914, 1915
  PVC-initiated, 1974
  in STEMI, 2124
  sudden cardiac death and, 2334
  treatment of, 1971, 2336, 2337f, 2338, 2339t
  ventricular tachycardia and, 1971
Ventricular filling, assessment of, 1853
Ventricular flutter, 1956, 1956f. *See also* Ventricular tachycardia
Ventricular free wall rupture, 2122t, 2328f, 2330
Ventricular function curve, 1850
Ventricular premature beats, 2092, 2123
Ventricular remodeling, 1979–1980, 1980f, 1980t, 2121–2122. *See also* Heart failure (HF)
Ventricular septal defect (VSD)
  atrioventricular canal type, 2072, 2072f
  AV conduction block in, 1927
  cardiac catheterization in, 1907
  cardiogenic shock in, 2328f
  echocardiography in, 1900
  epidemiology of, 2072
  heart sounds/murmur in, 283f, 284f, 285, 1866
  hypoxia in, 277
  pathophysiology of, 2071–2072, 2072f
  prognosis of, 2072
  subpulmonary, 2072, 2072f
  traumatic, 2088, 2089f
Ventricular septal rupture, 2122t, 2329t, 2330
Ventricular tachycardia (VT)
  bundle branch reentry, 1967
  definition of, 1954–1955
  idiopathic, 1967–1968, 1967f
  incessant, 1972–1975, 1972f, 1974f
  in ischemic heart disease, 2092
  mechanism of, 1913
  monomorphic, 1955–1956, 1955f, 1956f, 1957f, 1959f, A8
  nonsustained
    characteristics of, 1955, 1955f
    diagnosis of, 1960
    diseases associated with, 1960–1961
    prognosis of, 1960
    treatment of, 1960–1961
  outflow tract, 1967
  polymorphic, **1968**
    in Brugada's syndrome, 1971. *See also* Brugada's syndrome
    catecholaminergic, 160, 1971
    characteristics of, 1968–1969
    diseases associated with, 1970t
    drug-induced, 1970t
    in early repolarization syndrome, 1971
    ECG in, 1956, 1956f, 1969f
    in genetic dilated cardiomyopathy, 1971
    in hypertrophic cardiomyopathy, 1971
    in long QT syndrome, 160, 1969–1970, 1969f. *See also* Long QT syndrome
    in myocardial infarction/ischemia, 1969
    treatment of, 2336, 2337f
  sinusoidal, 1956, 1956f
  in STEMI, 2123–2124
  sustained
    after myocardial infarction, 1964, 1965f
    characteristics of, 1963–1964
    diagnosis of, 1963–1964
    etiology of, 1964t, 1965
    idiopathic, 1967
    prognosis of, 1964
    vs. supraventricular tachycardia, 1964, 1964f
    treatment of, 1967–1968, 1968f
  sustained monomorphic, 1970t
  treatment of
    antiarrhythmic drugs, 1957–1958
    approach to, 2337f
    automated external defibrillator, 2336
    cardiovascular implantable device, 1958–1959, 1959f, 2339t
    catheter ablation, 1965, 1966t

  ventricular fibrillation. *See* Ventricular fibrillation (VF)
  ventricular flutter, 1956, 1956f
Ventricular tachycardia (VT) storm, 1972, 1974f
Ventriculography, 1847, 1907f
Ventriculostomy, 2345, 2345f
Verapamil
  for accessory pathway-mediated tachycardia, 1942
  adverse effects of, 1926, 1929, 2101t, 2144t, 3472
  for aortic dissection, 2170
  for atrial fibrillation, 1949
  for AV nodal reentrant tachycardia, 1940
  for cluster headache prevention, 3472, 3472t
  contraindications to, 1941
  drug interactions of, 485t, 2101, 2408, 2721
    azoles, 1742t
    quinidine, 1745t
  for hypertension, 2144t
  for hypertension in CKD, 2396
  for hypertrophic cardiomyopathy, 2007
  for hypnic headache, 3475
  for ischemic heart disease, 2101t
  for LV fascicular VT, 1967
  metabolism of, 485, 485t
  for NSTE-ACS, 2110
  overdosage/poisoning with, 3712t
  for supraventricular arrhythmia, 2124
Verbal abuse, 3895t
Vericiguat, 1993f, 1995t, 1996
Vermiculite dust, 2238
Verner-Morrison syndrome. *See* VIPoma
Verruca plana, 380f, 390
Verruca vulgaris, 380f, 390, 1051
Verruga peruana, 1051, 1353t, 1356
Vertebrae
  "ivory," 3313
  metastases in, A16
  "picture frame," 3313
Vertebral artery dissection, 131, 3445
Vertebral artery, occlusion of, 3432–3433, 3432f, 3433f
Vertebral augmentation, 127–128
Vertebral fractures
  compression, A16
  diagnosis of, 3292, 3292f, 3297, A16
  epidemiology of, 3292, 3292f
  estrogen deficiency and, 3295
  estrogen therapy and, 3145
  pain in, 3300
  pathologic, A16
  prevention of
    bisphosphonates for, 3303–3306, 3305f
    denosumab for, 3306–3307, 3306f
    raloxifene for, 3303
    teriparatide for, 3308, 3308f
  screening for, 3297, 3298t
  treatment of, 3300
Vertebral osteomyelitis. *See* Osteomyelitis, vertebral
Vertebrogenic pain, 123t, 127–128
Vertebroplasty, 3300, 3310, 3393
Vertical diplopia, 221, 235
Vertical gaze disorders, 236
Vertigo, **162**
  acute prolonged (vestibular neuritis), 163
  approach to the patient, 162–163
  benign paroxysmal positional, 163–164, 163t, 164f
  central, 163t
  differential diagnosis of, 163
  evoked by loud sounds, 246
  functional, 165
  in Ménière's disease. *See* Ménière's disease
  in migraine, 164
  in MS, 3578
  peripheral, 163t
  prevention of, 165
  treatment of, 165, 165t
  in vestibular dysfunction, 164–165, 180
Very-low-calorie diet, 3190, 3216
Very low–density lipoproteins (VLDLs)
  in ataxia, 3537
  characteristics of, 3237f, 3238–3239
  estimate of, 3239
  in transport of hepatic lipids, 3238–3239, 3239f
Very-low-fat diet, 3249
Vesicants, S5
Vesicles/bullae, 377t, **400**
  clinical features of, 378f, 378t
  description of, 137, 377t
  drug-induced, 400–401
  etiology of, 401t, 1049, 1050t
  with fever, 142–143t, 147
  in herpes simplex virus infections. *See* Herpes simplex virus (HSV) infections
  infections associated with, 1049, 1050t
  in smallpox. *See* Smallpox
  in sulfur mustard exposure, S5
  in varicella. *See* Varicella (chickenpox)
Vesicoureteral reflux, 2441–2442, 2441f, 2454
Vesiculovirus, 1650, 1663, 1663t
Vessel wall, 921
Vessel wall disorders, **927**
Vessel wall MR imaging, 3390, 3391f
Vestibular ataxic gait, 177t
Vestibular dysfunction, 163t, 164–165, **180**
Vestibular-evoked myogenic potential test (VEMP), 250
Vestibular migraine, 164
Vestibular neuritis, 163, 163t. *See also* Vertigo
Vestibular nystagmus, 237
Vestibular rehabilitation therapy, 165
Vestibular schwannoma (acoustic neuroma)
  diagnosis of, 164, 247, 722
  hearing loss in, 164, 247, 249
  imaging in, 250, 722f
  in neurofibromatosis, 721, 3087
  tinnitus in, 247
  treatment of, 722
Vestibulitis, nasal, 510
Vestibulocochlear nerve, 3380
Vestibulo-ocular reflex, 163
Vestibulospinal pathway, 169f
Veterans. *See* Military veterans
Veterans Administration, S8
VEXAS syndrome, 150, 2937
VF. *See* Ventricular fibrillation (VF)
VGCC antibodies, 743t, 749
VHL gene mutations
  actions of, 3073, 3087
  epigenetic silencing of, 538
  NETs and, 677
  in pheochromocytoma/paraganglioma, 3073, 3079, 3079t
  in primary brain tumors, 717t
  in RCC, 687–688, 687t
  in von Hippel–Lindau disease/syndrome. *See* Von Hippel–Lindau disease/syndrome
VHL protein, 3073
Vibegron, 3549, 3588
Vibration-controlled elastography (VCTE), 2705
Vibration injury, 2178
Vibration sense, 173, 174t, 3382
Vibratory angioedema, 2807
Vibrio alginolyticus infections, 1330t, 1331
Vibrio cholerae. See also Cholera
  adherence of, 1076
  characteristics of, 964, 1325–1326, 3839
  cytotoxin, 974
  epidemiology of, 1326–1327, 1326f
  as extracellular pathogen, 967
  genome of, 985, 986f, 1327
  laboratory identification of, 1328, S11
  microbiology of, 1326
  niche, 965, 965t
  non-O1/O139, 1330, 1330t
  serotypes of, 1326
  tissue damage by, 974
  toxin production by, 1327
Vibrio spp. infections, 3750, 3752t
Vibrio parahaemolyticus infections, 1080t, 1330, 1330t
Vibrio vulnificus infections
  clinical features of, 143t, 147, 583, 992–993, 1051, 1330–1331
  epidemiology of, 1330, 1330t
  treatment of, 1331
Vibrotactile positional therapy, 2282
Vici's syndrome, S9
Vidarabine, 457, 1486t, 1487, 1502
Video-assisted thoracoscopic surgery, 2208
Video capsule endoscopy, 320
Video game training, 3964–3965
Vif/*vif,* 1552, 1552f, 1561f
Vilanterol, 2224, 2255
Vilazodone, 3659t, 3667
Villous adenoma, 308, 653
Vimentin, 542f, 2790t
Vinblastine
  actions of, 565
  adverse effects of, 563t, 753t
    alopecia, 421
    myelotoxicity, 565
    neurologic, 3608t
    phototoxicity, 434t
    Raynaud's phenomenon, 2178
  interactions and issues, 563t
  for Kaposi's sarcoma, 1613
Vinca alkaloids. *See also specific drugs*
  adverse effects of, 421, 508, 575, 753t, 756, 3608t
  drug interactions of, 1742t
Vincristine
  actions of, 565
  adverse effects of, 563t, 565, 753t
    depression, 84
    neurologic, 726, 726t
    neuropathy, 565, 3608t
  for ALL, 847
  for astrocytoma, 718
  for DLBCL, 863
  for Ewing's sarcoma, 730
  for follicular lymphoma, 864
  for gastric lymphoma, 650
  genetic variations in response to, 492, E72t
  for gestational trophoblastic disease, 715
  interactions and issues, 563t
  for metastatic pheochromocytoma, 3077
  for soft tissue sarcomas, 729
Vinorelbine
  actions of, 565
  adverse effects of, 563t, 595, 597, 3608t
  interactions and issues, 563t
  for lung cancer, 625t, 630
  for metastatic breast cancer, 641
  resistance to, 565
Vinyl chloride, 324, 512t, 2703t, 2940t
Violence, 3674. *See also* Domestic abuse/violence
VIP (vasoactive intestinal peptide), 293, 737t, 2743
Vipers, 3717, 3722. *See also* Snakebites
VIPoma (vasoactive intestinal peptidoma)
  clinical features of, 308, 354, 678t, 679, 3082
  in MEN 1, 3082
  treatment of, 678t, 679, 3082
Viral encephalitis. *See* Encephalitis, viral

- Viral hemorrhagic fevers, **1674**. *See also specific diseases*
  - in bioterrorism, S4
  - clinical features of, 144t, 148, **993,** 1674–1675, S4
  - diagnosis of, 1675, S4
  - epidemiology of, 144t, 993
  - etiology of, 993, 1666t, 1674–1679
  - health care facility preparedness for, 1149
  - pathogenesis of, 1675
  - prevention of, S4
  - treatment of, 1675, S4
- Viral hepatitis. *See* Hepatitis
- Viral infections. *See also specific diseases*
  - AKI in, 2376
  - arthropod-borne. *See* Arthropod-borne viruses
  - asthma exacerbations and, 2221
  - cancer development and, 525
  - as cause of fever of unknown origin, 151t
  - diagnosis of, 977t, 979–980, 1483–1484
  - in HCT recipient, 1155t
  - host resistance of, 1481
  - long-term effects of, 1482
  - orchitis and, 3115–3116
  - with pandemic potential, preparedness for, 1149
  - pleural effusion in, 2270
  - in pregnancy, 3903–3904
  - renal involvement in, 2428–2429
  - respiratory tract. *See* Respiratory tract infections, viral
  - rodent-borne. *See* Rodent-borne viruses
  - in solid organ transplant recipient, 1159t, 2287
  - transfusion-transmitted, 912–913, 913t
  - in transplant recipient
    - after kidney transplantation, 2410, 2410t
    - after liver transplantation, 2723
- Viral load, S11
- Viral meningitis. *See* Meningitis, viral
- Viral (subacute) thyroiditis, 3038, 3038f
- Virchow's node, 323, 327, 477, 673
- Virchow's triad, 2156
- Viremia, 1481
- Viridans streptococci, 1062, 1174t, 1206, 1206t, 1214
- Virilization, 3138
- Virtual colonoscopy, 655, 2497, 2503f
- Virtual reality, 3964, 3964f
- Virtue ethics, 72
- Virus(es). *See also specific viruses*
  - classification of, 147, 1478f, 1480t
  - drugs targeting. *See* Antiviral therapy
  - effects on host cell, 1481
  - evolution of, 1482–1483
  - host responses to, 1482, 1482f
  - laboratory identification of, 977t, 979–980, 1483–1484
  - molecular epidemiology of, 1483
  - oncolytic, 559f, 561, 3830
  - replication in cells, 1479–1481
  - structure of, 1477–1479, 1478f
  - as therapeutic tool or agents, 3824t
  - therapeutic uses of, 1484
  - vaccines for, 1484
- Virus-like particles, in vaccines, 997t
- VISA (vancomycin-intermediate *S. aureus*), 1044, 1204
- Visceral afferent hypersensitivity, **300,** 2572
- Visceral larva migrans. *See* Larva migrans, visceral
- Visceral leishmaniasis (kala-azar), **1781**
  - clinical features of, 468, 1737, 1739t, 1783, 1784f, S7
  - diagnosis of, 1783–1784, 1783f, S7, S12
  - differential diagnosis of, 1784
  - epidemiology of, 1737, 1781, 1782t, 1783f
  - etiology of, 1781
  - in HIV infection, 1611, 1737, 1783, 1785
  - immunopathogenesis of, 1781–1783, 2790t
  - incubation period for, S7
  - prognosis of, 1785
  - skin lesions after, 1785–1786, 1786f
  - treatment of, 1741t, 1750, 1784–1785, S7
  - in war veterans, S7
- Visceral pain, 80, 506
- Vismodegib, 533t, 571t, 573, 608
- Visual acuity, 21, 163, 164, 3380
- Visual agnosia, 3432
- Visual aura, 230
- Visual cortex, 200f
- Visual function, assessment of
  - color vision, 221–222
  - eye movements and alignment, 221
  - pupils, 221
  - refractive state, 220–221
  - stereopsis, 221
  - visual acuity, 221
  - visual fields, 222, 223f, 3380
- Visual loss
  - chemotherapy-induced, 726t
  - chronic/permanent
    - in cataract. *See* Cataract
    - in central serous chorioretinopathy, 231
    - in diabetic retinopathy. *See* Diabetic retinopathy
    - in epiretinal membrane, 232
    - in glaucoma. *See* Glaucoma
    - in macular degeneration. *See* Macular degeneration
    - in retinitis pigmentosa, 231, 232f, 3783t
    - trachoma and, 1474
    - tumor-related, 232, 232f
  - CMV retinitis and, 1514
  - factitious, 230
  - in giant cell arteritis, 114
  - HSV keratitis and, 1499
  - in legal blindness, 221
  - in leprosy, 1405
  - in parasitic infections, 1739t
  - with sellar masses, 2998
  - transient or sudden, 225–230, 226f
- Visual object agnosia, 206
- Visual target cancellation test, 205
- Vital capacity, 2202, 2203f
- Vitamin(s), **2605**. *See also specific vitamins*
  - body stores of, 2605
  - deficiencies of, 2605t
  - functions of, 2606–2607f
  - recommended intake of, 2600t
  - structures of, 2606–2607f
- Vitamin A, **2610**
  - deficiency of. *See* Vitamin A deficiency
  - dietary sources of, 2611
  - functions of, 2607f, 2610
  - metabolism of, 2610–2611
  - recommended intake of, 2600t
  - for retinitis pigmentosa, 231
  - structure of, 2607f
  - supplements, 243, 1641, 2611
  - toxicity of, 2612, 2672, 3282
- Vitamin A deficiency
  - clinical features of, 2605t, 2611
  - global considerations in, 2611
  - keratitis in, 224
  - in pregnancy, 1844, 2611
  - risk factors for, 2605t
- Vitamin B_1. *See* Thiamine (vitamin B_1)
- Vitamin B_2. *See* Riboflavin (vitamin B_2)
- Vitamin B_3. *See* Niacin (vitamin B_3)
- Vitamin B_5 (pantothenic acid), 2600t, **2610**
- Vitamin B_6. *See* Pyridoxine (vitamin B_6)
- Vitamin B_{12}. *See* Cobalamin (vitamin B_{12})
- Vitamin B_{12A} (hydroxocobalamin), 789, 3713t, S5
- Vitamin C, **2609**
  - absorption of, 2609
  - deficiency of, 450, 2605t, 2609. *See also* Scurvy
  - dietary sources of, 2609
  - functions of, 407, 2607f, 2609, 3322
  - for macular degeneration, 231
  - for methemoglobinemia, 780
  - recommended intake of, 2600t
  - structure of, 2607f
  - supplements, 2449, 2609, 2962
  - toxicity of, 2609
- Vitamin D, **2612, 3268**
  - beneficial effects of, 431
  - colorectal cancer risk and, 655
  - deficiency of. *See* Vitamin D deficiency
  - functions of, 2607f, 2612
  - for hypoparathyroidism, 3287
  - malabsorption of, surgery-related, 2534
  - metabolism of, 3268, 3269f
  - photochemistry of, 431
  - recommended intake of, 2600t, 3295, 3301
  - resistance to, 365
  - sources of, 2612
  - structure of, 2607f
  - supplements. *See* Vitamin D supplements
  - synthesis of, 431, 3268, 3268f
  - toxicity of, 2612, 3281
- Vitamin D deficiency
  - anticonvulsants and, 3290
  - asthma risk and, 2220
  - in CKD, 3290
  - clinical features of, 2605t, 3270, 3295
  - colorectal cancer and, 655
  - diagnosis of, 365–366, 3270–3271
  - etiology of, 3269–3270, 3270t, 3289–3290
  - genetic considerations in, 3269–3270
  - global considerations in, 366
  - hypocalcemia and, 365, 3290
  - hypomagnesemia and, 3267
  - intestinal disease and, 3268
  - MS and, 3581
  - nonskeletal disease and, 2612
  - in older adults, 3269
  - osteomalacia due to. *See* Osteomalacia
  - osteoporosis and, 3295
  - prevalence of, 2979t
  - rickets due to. *See* Rickets
  - risk factors for, 2605t, 2612
  - screening/testing for, 2979t
  - treatment of, 366, 2612, 3267, 3271
  - in women, 3166
- Vitamin D receptor (VDR), 3269
- Vitamin D supplements
  - for chemotherapy-induced taste dysfunction, 243
  - for hypocalcemia, 366
  - for MS, 3587
  - in older adults, 40t, 3884
  - for osteoporosis prevention/management, 2612, 3145, 3301
  - for psoriasis, 393t
  - for vitamin D deficiency, 2612, 3271
- Vitamin E, **2613**
  - deficiency of. *See* Vitamin E deficiency
  - functions of, 2607f, 2612–2613
  - for macular degeneration, 231
  - metabolism of, 2613
  - recommended intake of, 2600t, 2613
  - structure of, 2607f
  - supplements, 473, 2706
  - toxicity of, 2613
- Vitamin E deficiency
  - ataxic disorders in, 3537
  - clinical features of, 2605t, 2613, 3612
  - diagnosis of, 3612
  - familial, 2613, 3612

  - risk factors for, 2605t, 2613
  - treatment of, 2613, 3612
- Vitamin K, **2613**
  - bone mass and, 3301
  - deficiency of. *See* Vitamin K deficiency
  - dietary sources of, 2613
  - functions of, 934, 2607f, 2613
  - metabolism of, 932, 933f
  - recommended intake of, 2600t
  - structure of, 2607f
  - supplements. *See* Vitamin K supplements
  - toxicity of, 2614
  - for warfarin reversal, 2163
- Vitamin K antagonists, 950. *See also* Warfarin
- Vitamin K cycle, 933f, 950–951, 950f
- Vitamin K deficiency, **934**
  - bleeding in, 473
  - clinical features of, 2605t, 2613
  - etiology of, 933–934, 2613
  - risk factors for, 2605t, 2613
  - treatment of, 935
- Vitamin K epoxide reductase complex (VKORC1), 932–933, 950–951
- Vitamin K supplements
  - for coagulation disorders in liver disease, 935, 2717
  - for neonates, 935
  - for vitamin K deficiency, 2613
  - for warfarin-induced bleeding, 952
- VITAMINS TRAP mnemonic, 364
- Vitiligo
  - APS-2 and, 3091t
  - clinical features of, 378t, 382f, 398t, A5
  - diagnosis of, 382, 398t
  - disorders associated with, 396–397, 3230
  - inflammasome activity in, 2762t
  - pathophysiology of, 398t, 2790t
  - treatment of, 398t, 436
- Vitreous disorders, 229
- VITT (vaccine-induced immune thrombocytopenia and thrombosis), 923
- *Vittaforma* spp., 1808
- *VKORC1* polymorphisms, 493, 951, 951t, E72t
- VKORC1 (vitamin K epoxide reductase complex), 932–933
- VLA-4, 759
- VLDLs. *See* Very low–density lipoproteins (VLDLs)
- VNTRs (Variable number of tandem repeats), 3793
- Vocal cord dysfunction (inducible laryngeal obstruction), 2222
- Vocal cord injury, 2307
- Vocal tic, 3515
- Voclosporin, 2835, 2836
- VODI (venoocclusive disease with immunodeficiency) syndrome (Sp110 deficiency), S9
- Vogt-Koyanagi-Harada syndrome, 225, 397, 1122, 1131, 1132t
- Voiding cystourethrography, 2456
- Volanesorsen, 3250
- Volume of distribution (Vd), 1167
- Volume overload, 1982, 2626
- Voluntary active euthanasia, 89t
- Volutrauma, ventilator-induced, 2294–2295
- Volvulus, 2487, 2493f, 2589–2590, 2593
- Vomiting. *See also* Nausea and vomiting
  - in acute cholecystitis, 2732
  - after traumatic brain injury, 3574
  - definition of, 296
  - in gastroenteritis. *See* Gastroenteritis
  - hypokalemia in, 356
  - hyponatremia in, 347
  - hypovolemia in, 346
  - in intestinal obstruction, 2591
  - Mallory-Weiss tear following. *See* Mallory-Weiss tears
  - mechanisms of, 297
  - metabolic alkalosis in, 373
  - water loss in, 2598
- von Gierke disease (glucose-6-phosphatase deficiency, type I GSD), **3358,** 3360t, A15, S9
- von Hippel–Lindau disease/syndrome
  - clinical features of, 3087
    - CNS tumors, 717t, 2430t, 3078f, A16
    - hemangioblastoma, 687t, 3087
    - renal, 687, 687t, 2430t, 2434
    - tumors and cysts, 247, 3078, 3078f
  - diagnosis of, 2434, A16
  - genetic considerations in
    - genotype-phenotype correlation, 3787
    - locus and phenotype, 524t
    - transcription factor mutations, 3783t
    - *VHL* gene mutations, 3078, 3080t, 3087
  - treatment of, 573, 2434
- Vonoprazan, 301, 1303, 2526
- von Recklinghausen's disease. *See* Neurofibromatosis type 1 (NF1)
- von Willebrand disease (VWD), **926**
  - acquired, 819, 824, 927, 2032
  - bleeding in, 472
  - clinical features of, 926–927
  - diagnosis of, 926t
  - inheritance patterns in, 926–927
  - pathophysiology of, 472
  - subtypes of, 926, 926t
  - treatment of, 927
- von Willebrand factor
  - in platelet activation, 469, 937f
  - in platelet aggregation, 937f, 938
  - in thrombotic thrombocytopenia purpura, 925, 925f
  - in VWD, 926–927, 926t. *See also* Von Willebrand disease
- von Willebrand factor concentrate, 927
- Vorapaxar, 946, 2173
- Vorasidenib, 574, 718
- Voretigene neparvovec, 3824t
- Voriconazole
  - for ABPA, 2233
  - adverse effects of, 400, 401, 420, 1691
  - for aspergillosis, 1716, 1717t
  - for blastomycosis, 1702–1703, 1703t
  - for *Candida* infections, 1711t
  - for CNS aspergillosis, 583
  - for coccidioidomycosis, 1698
  - contraindications to, 1691
  - for cryptococcosis, 1706
  - for eumycetoma, 1725
  - for fungal infections in cancer patient, 833
  - for fusariosis, 1726t, 1729
  - genetic variations in response to, 1691, E72t
  - indications for, 1691
  - for phaeohyphomycoses, 1726t, 1728
  - resistance to, 1711t, 1716
  - for *Scedosporium* infections, 1726t, 1730
  - for *Trichosporon* infections, 1726t, 1730
- Vorinostat, 533t, 539, 570t, 573, 874, 3942
- Vortioxetine, 3659t, 3667
- Vosoritide, 3318
- Voxelotor, 775
- Voxilaprevir, 1492t, 1493, 2686f, 2691
- V pre-B, 2777
- Vpr/*vpr,* 1552, 1552f, 1561f
- *VPS13A* gene mutations, 3514
- VPS35, 3400
- *VPS35* gene mutations, 3498t
- Vpu/*vpu,* 1552, 1552f, 1559, 1561f
- Vpx/*vpx,* 1552, 1552f, 1559
- VR, S5
- VRE. *See* Vancomycin-resistant enterococci (VRE)
- VSD. *See* Ventricular septal defect (VSD)
- V-sign, 2911, 2911f
- VT. *See* Ventricular tachycardia (VT)
- VTE. *See* Venous thromboembolism (VTE)
- Vulvar cancer, 599f, 1526
- Vulvar warts. *See* Condyloma acuminata
- Vulvovaginal infections, **1101**
  - bacterial vaginosis, 1101t, 1102–1103, 1103f
  - candidiasis, 1101t, 1103, 1602, 1708, 1710t
  - clinical features of, 1101, 1101t
  - etiology of, 1097t, 1101t
  - gonococcal, 1257
  - in HIV infection, 1602
  - mucopurulent cervicitis, 1097t, 1103–1104, 1104f, 1257, 1471, 1473t
  - treatment of, 1102
  - *Trichomonas vaginalis,* 1101t, 1102, 1110, S11
  - ulcerative. *See* Genital ulcer(s)
- VUS (variants of unknown significance), 525, 3784, 3787, 3802
- Vutrisiran, 900, 2027
- VV-ECMO (venovenous extracorporeal membrane oxygenation), 2309, 2309t
- VWE. *See* Von Willebrand disease (VWD)
- VX, S5
- VZIG (varicella-zoster immune globulin), 1507, 1508t
- VZV (varicella-zoster virus), 1504, 1506–1507. *See also* Varicella-zoster virus (VZV) infections

W

- Waardenburg's syndrome, 248, 397, 3783t, E36t
- Wada test, 3425
- Waddell signs, 121t
- WAGR syndrome, 3783t
- Waist circumference, 3188, 3188t, 3253t, 3255
- Wakefulness, 210, 210f
- Waldenström's hypergammaglobulinemia purpura, 408
- Waldenström's macroglobulinemia, **893**
  - amyloidosis in, 897
  - clinical features of, 865, 893, 2424
  - diagnosis of, 865
  - genetic considerations in, 893
  - paraneoplastic syndromes in, 411
  - pathogenesis of, 859f
  - Raynaud's phenomenon in, 2178
  - treatment of, 893–894
  - vascular purpura in, 927
  - VWD in, 927
- Walled-off necrosis, pancreatic, 2746t, 2747f
- Wallenberg's syndrome, 3433, 3433f, A16
- Walrus bite-wound, 1143
- Warburg effect, 543, 544f, 573, 2187
- Warfarin, **950**
  - actions of, 950–951, 950f
  - adverse effects of
    - alopecia, 394
    - bleeding, 952, 1951, 2163
    - calciphylaxis, 2391
    - fetal, 3900
    - in pregnancy, 952, 3900
    - skin necrosis, 407, 420–421, 421f, 427t, 952, A5
  - for antiphospholipid syndrome, 2446
  - as contraindication to lumbar puncture, S3
  - dosage of, 493, 951–952, 951t, 2163
  - drug interactions of
    - amiodarone, 485t, 489t, 490
    - azoles, 1742t
    - cimetidine, 489t
    - ciprofloxacin, 1745t
    - isoniazid, 1422

- Warfarin, drug interactions of (*Cont.*):
    - metronidazole, 1172t, 1744t
    - nafcillin, 1172t
    - orlistat, 2613
    - PPIs, 2525
    - SSRIs, 3660t
    - sulfonamides, 1743t
    - sulfonylureas, 3213
    - tetracyclines, 1746t
    - trimethoprim, 1172t, 1743t
  - genetic variations in response to, 491, 493, 950–951, 951t, 3805, E72t
  - management of
    - before cardiac catheterization, 1904
    - before gastrointestinal endoscopy, 2483t
    - with left ventricular assist systems, 2032
    - preoperative, 952
    - treatment monitoring, 951
  - metabolism of, 485t
  - for NSTE-ACS, 2110t
  - in patient with HIT, 922–923
  - pharmacology of, 950–951
  - in pregnancy, 952
  - for stroke prevention, 1951, 2050, 3450–3451
  - for VTE/PE treatment, 742, 2163, 2163t
  - for VTE prophylaxis, 2164t
- Wart(s), **390**
  - etiology of, 390, 1051
  - flat, 390
  - genital. *See* Condyloma acuminata
  - oral, 264t
  - plantar, 390
  - treatment of, 390
- Warthin-Starry stain, S11
- Warts, hypogammaglobulinemia, infections, and myelokathexis (WHIM) syndrome, 453, 457, 2795, 2802
- War veterans. *See* Military veterans
- WASp, 2800
- *WASP* gene mutations, 2795, 2800
- Wasp sting, 3735–3736
- Wastewater-based surveillance, 986
- Wastewater drains, as infection source, 1151
- Wasting syndrome, 1611
- Watchman LAA closure device, A11, A11v
- Water
  - recommended intake of, 2602t
  - renal transport of, 2371, 2371f
  - requirement for, 2598–2599
- Water balance, 344–345, 2371, 2371f, 2389
- Waterborne disease
  - climate change and, 1269, 3749–3750, 3752f, 3752t
  - health care–associated, 1150–1151
  - *Legionella* spp. infections. *See Legionella* spp. infections
  - *V. cholerae* infections. *See* Cholera
- Water brash, 301, 2504
- Water-depletion heat exhaustion, 3774
- Water deprivation, 351, S1
- Water deprivation test, 344, 353, S1
- Water diuresis, 3015, 3015f, S1
- Water-hammer (Corrigan's) pulse, 287, 1863, 2044
- Water intoxication, 2371, 2371f
- Water loss, 346, 347–348, 2598–2599
- Watermelon stomach, 2874
- Water pipes, 3686
- Waxy casts, urinary, 343
- WBRT (whole-brain radiation therapy), 721, 724
- WDHA syndrome. *See* VIPoma
- *WDR11* gene mutations, 3111t, 3112
- Weakness, **168**
  - abnormalities associated with, 168
  - in botulism, 1235, 3628
  - in critically ill patient, 2298, 2351–2352
  - definition of, 168
  - distal, 171
  - in distal myopathies, 3643
  - distribution of, 168, 170–172
  - in endocrine disorders, 3647
  - evaluation of, 171f, 2909–2910, 3634, 3636f, 3637f, 3637t, 3638f
  - facial. *See* Facial weakness/paralysis
  - in facioscapulohumeral muscular dystrophy, 3642
  - falls and, 181
  - vs. fatigue, 165, 167
  - generalized (quadriparesis), 170–171, 171f, 171t
  - in Guillain-Barré syndrome, 3617
  - in hemiparesis, 170
  - in hypokalemia, 356, S1
  - in hypokalemic periodic paralysis, 3646
  - in hypovolemia, 347
  - in limb-girdle muscular dystrophy, 3640
  - lower motor neuron (neuropathic), 168, 168t, 169
  - in malabsorption disorders, 2549t
  - in monoparesis, 171
  - in MS, 3578, 3588
  - in muscular dystrophy, 3640
  - in myasthenia gravis, 3626
  - myopathic, 168, 168t, 169
  - in myotonic dystrophy, 3642
  - neuromuscular junction, 169
  - in paraparesis, 170
  - in parathyroid disorders, 3647
  - patterns of, 3596t
  - in peripheral neuropathies, 3595–3596, 3596t
  - in potassium channel disorders, 3647
  - in progressive diaphyseal dysplasia, 3315
  - proximal, 171
  - psychogenic, 168t, 169
  - in restricted distribution, 171–172
  - in sodium channel disorders, 3646–3647
  - upper motor neuron, 168, 168t, 169
- Weaning, from mechanical ventilation, 88–89, 2308–2309, 2308f
- Weaponization, of microbes, S4
- Weathering effect, 63
- Weber's syndrome, 234, 3431, 3432f
- Weber test, 248, 3380
- Web(s), esophageal, 2505–2506
- Wegener's syndrome. *See* Granulomatosis with polyangiitis
- Weight
  - hypertension and, 2137t
  - regulation of, 315, 3181–3182, 3182f, 3183f
- Weight gain. *See also* Obesity
  - abdominal fat from, 326
  - circadian clock desynchrony and, 3951
  - in edema, 281
  - in hypothyroidism, 3029
  - physiology of, 3181–3182, 3182f, 3183f
  - in thyrotoxicosis, 3033
- Weight loss
  - biliary sludge formation and, 2729
  - cholesterol stone formation and, 2729
  - for hypertension management, 2137t, 2141
  - involuntary. *See* Involuntary weight loss
  - for MASH, 2706
  - for metabolic syndrome, 3257
  - for obesity treatment. *See* Obesity, treatment of
  - in preventive care, 12–13
- Weil's syndrome. *See Leptospira*/leptospirosis
- Weiss score, adrenocortical carcinoma, 3065
- Welander distal myopathy, 3640t, 3643–3644
- Wells criteria, for pulmonary embolism, 26t
- Werdnig-Hoffmann disease, 3524
- Wermer's syndrome. *See* Multiple endocrine neoplasia type 1
- Werner's syndrome, 3788t, 3870, 3941
- Wernicke-Korsakoff syndrome, 207, 2607
- Wernicke's aphasia, 201–202, 201t, 204
- Wernicke's area, 201, 204f
- Wernicke's disease/encephalopathy
  - in alcohol use disorder, 3676
  - clinical features of, 235, 2349, 2607, 3483, 3559, V3
  - imaging in, 2349, 2349f
  - pathogenesis of, 2349
  - treatment of, 2349, 3483
- Wesselsbron virus, 1660t
- Westermark's sign, 107, 2160
- Western blot, 1484
- Western equine encephalitis, 1657t, 1673–1674
- Western/European tick-borne encephalitis, 1671
- West Nile virus, 1660t
- West Nile virus infection
  - climate change and, 3748–3749
  - clinical features of, 140t, 170, 402, 1114–1115, 1125, 1672
  - diagnosis of, 1113, 1114, 1114t
  - epidemiology of, 1112, 1672, 3748–3749
  - transfusion-transmitted, 912, 913t, 1672
  - in transplant recipient, 1672
  - treatment of, 1116
  - vaccines, 1116
- Westphal variant, Huntington's disease, 3512
- Wet beriberi, 2607, 3612
- Wet-bulb globe temperature, 3773
- Wetland virus, 1662t
- Wet mount, S11, S12
- *WFS1* gene mutations, 3093
- WHDA syndrome. *See* VIPoma
- Wheal, 137, 377t
- Wheat intolerance, nonceliac, 309
- Wheezing, 1984, 2199
- WHIM (warts, hypogammaglobulinemia, infections, and myelokathexis) syndrome, 453, 457, 2795, 2802
- Whiplash injury, 132
- Whipple procedure, 669
- Whipple's disease, **1367**
  - asymptomatic colonization/carriage, 1368
  - clinical features of, 235, 1368–1369, 2546
    - acute infection, 1368
    - cardiovascular, 1368–1369, 2013
    - chronic infection, 400, 1130t, 1368, 1368f
    - dermatologic, 1369
    - gastrointestinal, 1368, 1368f
    - isolated infection, 1369
    - lymphatic, 1369
    - neurologic, 1368
    - ocular, 1369
    - pulmonary, 1369
    - reinfection/relapsing infection, 1369
    - steatorrhea, 309
  - diagnosis of, 150, 1369–1370, 2546, 2547f
  - epidemiology of, 1367
  - etiology of, 1367, 2546. *See also Tropheryma whippelii*
  - pathogenesis of, 1367–1368
  - treatment of, 311, 1370, 2546
- Whipple's triad, 3230
- Whipworm (*Trichuris trichiura*/trichuriasis), 961t, 1736, 1739t, 1815t, 1818
- Whispered pectoriloquy, 2200
- White blood cells, 443, A4. *See also* Basophil(s); Eosinophil(s); Lymphocyte(s); Monocyte(s); Neutrophil(s)
- White coat hypertension, 1862, 2135
- Whitehead, 390
- White matter disease, diffuse (leukoaraiosis), 198f, 3448
- White matter hyperintensities, 3490, 3490f
- White pulp, of spleen, 478, 479f
- White sponge nevus, 264t
- Whitewater Arroyo virus, 1658t

- Whitlow, herpetic, 1049
- WHO (World Health Organization). *See* World Health Organization (WHO)
- Whole blood transfusion, 907t
- Whole-body scanning, 3027, 3047
- Whole-bowel irrigation, 3709
- Whole-brain radiation therapy (WBRT), 721, 724
- Whole-exome sequencing, 3801
- Whole-genome sequencing, 976t, 3989
- Whooping cough. *See* Pertussis
- Wickham's striae, 388
- Widow spider bite, 3734
- Wildfires, 2241
- Wild type allele, 3787
- Williams-Campbell syndrome, 2242
- Williams myopathy, 3640t, 3643
- Williams' syndrome, 3281
- Wilms' tumor, 524t, 3792
- Wilms' tumor gene-1 mutations. *See WT1* gene mutations
- Wilson's disease, **3338**
  - vs. autoimmune hepatitis, 2695
  - clinical features of, 3338–3339
    - anemia, 2615
    - hepatic, 324, 2630, 2712, 3338
    - Kayser-Fleischer ring, 2630, 3339, 3339f, 3518, A15
    - neurologic, 3338, 3518
    - psychiatric, 3338–3339
  - diagnosis of, 324, 2630t, 2712, 3339, 3339t, A13
  - epidemiology of, 2712, 3518
  - future outlook for, 3340
  - genetic testing for, 3804t
  - HCC and, 659
  - history of, 3338
  - pathogenesis of, 2615, 2712
  - prevalence of, 3518
  - treatment of, 3340, 3518, 3804t
  - zinc deficiency in, 2614
- Wine maker's lung, 2228t
- "Winking owl" sign, 589
- Winterbottom sign, 1794
- Wiskott-Aldrich syndrome, 458, 2553t, **2800**
- Withdrawal syndrome
  - alcohol. *See* Alcohol withdrawal syndrome
  - benzodiazepine, 184, 3660
  - opioid, 3691
  - psychostimulants, 3695
- Withdrawal/withholding of care, 5–6, 88–89, 2299
- WNK kinases, 359
- *WNT1* gene mutations, 3324, 3325t
- *WNT4* gene mutations, 3096f, 3133
- WNT/APC pathway, 526t
- WNT/ β-catenin pathway, 431, 660, 661t
- Wnt pathway, 3294, 3397
- Wnt proteins, 2846
- *Wolbachia,* 1460
- Wolff-Chaikoff effect, 3023, 3037, 3040
- Wolff-Parkinson-White syndrome
  - accessory pathways in, 1940–1941, 1940–1941f
  - developmental defects in, 1844
  - ECG in, 1940, A8
  - sudden cardiac death in, 2335, 2335t
  - treatment of, 1941–1942
- Wolfram's (DIDMOAD) syndrome, 3093
- Wolman's disease. *See* Lysosomal acid lipase deficiency
- Women's health, **3162**
  - AD, 3163–3164
  - ALS, 3164
  - autoimmune disease, 2823
  - autoimmune disorders, 3165–3166
  - body temperature variations, 134
  - breast cancer. *See* Breast cancer
  - cardiovascular disease, 3164–3165
  - cardiovascular disease risk reduction, 2100
  - COVID-19 disease, 3167
  - depression, 3665
  - diabetes mellitus, 3165
  - disease risk, 3162–3163
  - domestic violence. *See* Domestic abuse/violence
  - epilepsy, 3427
  - hirsutism. *See* Hirsutism
  - HIV infection, 3166
  - hypertension, 3165
  - hypertension treatment, 2148
  - leading causes of death, 3162–3163, 3163f
  - life expectancy, 3162–3163
  - NSTE-ACS treatment in, 2112
  - nutrient requirements, 2600–2602t
  - obesity, 3166
  - osteoporosis, 3166
  - PD, 3164
  - pharmacology, 3166–3167
  - psychological disorders, 3167
  - smoking, 3168
  - substance abuse, 3168
  - sudden cardiac death, 2333
  - veterans, S8
- Women who have sex with women (WSW), 3179t
- Wood's lamp examination, 382, 397
- Woodworker's lung, 2228t
- Work of breathing, 268, 2203–2204
- World Bank, 3843
- World Health Organization (WHO)
  - building primary care in low- and middle-income countries, 3866
  - clinical classification of leprosy, 1406
  - *Disease Control Priorities,* 3847
  - emerging infection disease control, 3855
  - formation of, 3843
  - Global Leprosy Strategy 2021–2030, 1414
  - monitoring of COVID-19 information on social media, 16
- World Trade Center disaster, 2240
- Wound(s)
  - chronic. *See* Cutaneous ulcers
  - infections. *See also* Surgical-site infections
    - anaerobic bacterial, 1377
    - botulism, 1234, 1235, 1237, 1238, 2351
    - clostridial, 1239. *See also* Clostridial myonecrosis
    - specimen collection and transport, S11
    - in war veterans, S7
- Woven bone, 3261
- *WRAP53* gene mutations, 3818, 3820t
- Wrap-around services, 3849
- Wright's stain, 381, 462, S12
- Wrist fractures, 3292
- *WRN* gene mutations, 3870
- "Wrong-way" eyes, 190
- WSW (women who have sex with women), 3179t
- *WT1* gene mutations
  - cancers associated with, 524t
  - in gonadal development, 3094, 3096f
  - in mesothelioma, 616
  - in testis development disorders, 3098, 3099t
  - in WAGR syndrome, 3783t
- *Wuchereria bancrofti,* 961t, 1819, 1820t, 2234, S7, S12. *See also* Lymphatic filariasis
- Wyeomyia virus, 1659t

X

- X4 viruses, 1577
- Xanthelasma, 380f, 404, 2630, A15
- Xanthine oxidase inhibitors, 489t
- Xanthoastrocytomas, pleomorphic, 720
- Xanthoma, 404
  - in cardiovascular disease, 1860
  - in cholestasis, 3247
  - eruptive, 404, 3240, A15
  - in familial chylomicronemia syndrome, 3240
  - in familial dysbetalipoproteinemia, 3245
  - in familial hypercholesterolemia, 3243, A15
  - in lipoprotein disorders, 2968, 3241t, A15
  - normolipemic plane, 404
  - palmar, 3245
  - palmar crease, 1860
  - plane, 404
  - tendon, 404, 3243, A15
  - tuberous, 404, 3249
- *XBP1* gene mutations, 2554t
- XCR1, 2769t
- Xenotransplantation, 2033
- Xerocytosis, 795
- Xeroderma pigmentosum, 430, 606, 3780, 3785
- Xerophthalmia, 2549t, 2611
- Xerosis, 379t, 1475
- Xerostomia
  - dental disease and, 266
  - differential diagnosis of, 2878t
  - drug-induced, 266
  - etiology of, 266
  - radiation therapy-related, 614
  - in Sjögren's disease, 2876–2877
  - in systemic sclerosis, 2871
  - treatment of, 2878–2879
- Xerotic eczema (asteatotic eczema), 380f, 385. *See also* Eczema
- X-inactivation, 3782, 3791. *See also* X-linked disorders
- Xingu virus, 1659t
- *Xist* gene, 3782
- X-linked disorders
  - adrenoleukodystrophy. *See* Adrenoleukodystrophy
  - agammaglobulinemia, 458
  - chronic granulomatous disease, 2794, S9. *See also* Chronic granulomatous diseases (CGDs)
  - color blindness, 221–222
  - incidence of, 3789
  - inheritance of, 3789, 3790f, 3791
  - NEMO deficiency, 459t, 460, 2802, 2804, S9
  - spinobulbar muscular atrophy. *See* Spinobulbar muscular atrophy (Kennedy's disease)
- X-linked dystonia-parkinsonism, 3511
- X-linked lymphoproliferative disease, 466, 1510, 1512. *See also* Hemophagocytic lymphohistiocytosis (HLH)
- X-linked proliferative syndrome (XLP), 2798
- X-linked protoporphyria (XLP)
  - clinical features of, 3345, 3350
  - diagnosis of, 3341t, 3343t, 3345, 3350
  - genetic considerations in, 3350
  - global considerations in, 3342
  - pathophysiology of, 3341t
  - treatment of, 3350–3351
- X-linked sideroblastic anemia, 3341f, 3343t, **3349**
- X-rays, 554, S6
- 46,XX disorders of sex development, 3095t, 3101t
  - congenital absence of the vagina, 3101
  - increased androgen synthesis, 3100, 3101t. *See also* Congenital adrenal hyperplasia (CAH); 21-Hydroxylase deficiency
  - testicular/ovotesticular DSD, 3100, 3101t
- 45,X/46,XY mosaicism (mixed gonadal dysgenesis), 3097t, 3098
- 46,XY disorders of sex development, 3095t, 3097t, 3099t
  - androgen actions disorders, 3099–3100
  - androgen synthesis disorders, 3098–3099
  - cryptorchidism, 3100
  - isolated hypospadias, 3100
  - persistent müllerian duct syndrome, 3100
  - testis development disorders, 3098
- *Xylohypha* infections, 1130t
- D-Xylose test, in malabsorption disorders, 2550

Y

y (Gray), 554
Yamaguchi's syndrome, ECG in, A7
Yamanaka factors, 3871
Yao's syndrome, 2937
Yatapoxviruses. See Tanapox
Yaws
  clinical features of, 1438, 1438f
  epidemiology of, 1436, 1437f
  vs. other treponematoses, 1437t
  treatment of, 1439
Y chromosome, 3107f, 3778, 3791
^{90}Y-DOTA-TOC, 685
Yeasts, 1688t. *See also Candida* spp. infections; Cryptococcosis; *Trichosporon* spp. infections
Yellow fever, 1660t, 1679, 3843, 3844
Yellow fever vaccine, 1014t, 1015, 1016f, 1679
Yellow nail syndrome, 2183
Yellow oleander, 358
Yellow sac spider, 3734
Yellowtail moth, 3737
Yergason's test, 2974
Yersinia spp.
  adhesin proteins of, 966, 966t
  apoptosis inhibition by, 973
  as extracellular pathogen, 967
  immune response inhibition by, 974
  MAPK signaling inhibition by, 971
  subspecies of, 1347–1348
Yersiniabactin, 1348
Yersinia enterocolitica, 973, 1348, 1349
Yersinia spp. infections (yersiniosis)
  antithyroid antibodies in, 1349
  clinical features of, 305, 306t, 1060, 1348–1349
  complications of, 305
  diagnosis of, 1349
  enteric fever, 1077
  epidemiology of, 1348
  etiology of, 1347–1348
  food-borne, 1348
  global considerations in, 1348, 1350
  vs. IBD, 2559–2560
  pathogenesis of, 1348
  postinfective phenomena of, 1349
  prevention and control of, 1350
  transfusion-associated, 912
  treatment of, 1349–1350
Yersinia pestis. See also Plague
  as bioterrorism agent, 1342, S4
  global distribution of, 1342, 1342f
  immune response avoidance by, 971
  laboratory identification of, 1344–1346, 1345f, S11
  life cycle of, 1343
  pandemics caused by, 3854, 3854t
  pathogenesis of, 1343
Yersinia pseudotuberculosis, 967, 1342, 1348, 1349. *See also Yersinia* spp. infections (yersiniosis)
Yezo virus, 1662t
Y-linked disorders, 3789, 3791
Yoga, 2965, 3924, 3928
Yohimbine, 160
Yolk sac tumor, of ovary, 712
Yo protein, 2791t
Yo protein antibodies, 619, 743t, 749
^{90}Yttrium, 685, 1556, S6

Z

Zafirlukast, 2225
Zalcitabine (ddC), 1616, 1619f, 3611
Zaleplon, 216, 3713t
Zanamivir, 1157
  actions of, 1487
  contraindications to, 1550
  for influenza prophylaxis, 1488t
  for influenza treatment, 1488, 1488t, 1537, 1550
  pharmacology of, 1487
  resistance to, 1488
Zanubrutinib
  action and targets of, 533t, 567, 568t
  adverse effects of, 567, 568t, 753t, 754
  for CLL, 855
  for lymphoplasmacytic lymphoma, 865
  for mantle cell lymphoma, 866
  for nodal marginal zone lymphoma, 865
  for Waldenström's macroglobulinemia, 894
ZAP70 deficiency, 2799
Zavegepant, 3467t, 3468t
ZBTB7A, 771
Z-disk, genetic defects of, 2004t
Zeaxanthin, 2611
Zeitgebers, 3946
Zenker's diverticulum
  clinical features of, 298, 2506
  diagnosis of, 2506f
  dysphagia in, 294
  endoscopy of, 2492, 2497f, V5
  pathophysiology of, 2506
ZES. *See* Zollinger-Ellison syndrome (ZES)
Zicam, 3926
Zidovudine (AZT)
  adverse effects of, 1616, 1617t
    anemia, 1604
    hepatotoxicity, 2672
    hypertrichosis, 421
    myopathy, 1610, 2940t, 3648t
    nail disorders, 420, 421, 1606
    neutropenia, 457
    oral pigmentation, 264t
    steatosis, 2703t
  for ATL, 1555–1556, 1555f
  drug interactions of, 1741t, 1743t
  for HIV infection, 1617t, 3904
  molecular structure of, 1619f
  resistance to, 1622f
Zieve's syndrome, 2709
Zika virus, 980, 1660t
Zika virus disease
  clinical features of, 140t, 1059, 1668
  epidemiology of, 1668, 3748, 3750
  fetal microcephaly and, 1668, 3403, 3904
  in pregnancy, 1021, 3904
  skin manifestations of, 137, 140t, 146, A1
  transmission of, 913, 1668
Zileuton, 2225
Zilucoplan, 3631, 3632t
Zimbabwe, 3865
Zinc, **2614**
  absorption of, 2614
  deficiency of, 2611, 2614, 2614t
  functions of, 2614
  for macular degeneration, 231
  recommended intake of, 2601t
  in specialized nutritional support, 2626
  supplements, 242, 243, 1329, 2614
  tolerable upper intake level of, 2614t
  toxicity of, 2614, 2614t, 3926
  for Wilson's disease, 3340, 3518
Zinc gluconate lozenges, 2614
Zinc oxide, 435, 436t
Zinc oxide exposure, 2240
Zinc sulfate flotation, S12
Zinc supplements, 2717
Zipper mechanism, bacterial, 967, 968f
Ziprasidone, 3668, 3673, 3673t
Zirconium, 406
Ziv-aflibercept, 534t, 548, 572t, 574
Z-line deformity, 2840
Z lines, of sarcomere, 1847
Zn-DTPA, S6
ZNF469 gene mutations, 3329t, 3330
ZNRD1 gene, 1581t
Zolbetuximab, 649
Zoledronate (zoledronic acid)
  adverse effects of, 3303, 3304
  for fracture prevention, 3305f
  for hypercalcemia, 364, 738, 3284t, 3285
  for multiple myeloma, 892
  for osteoporosis management/prevention, 3145, 3304, 3305f, 3311
  for Paget's disease of bone, 3303, 3314t
  prophylactic, 638, 703
Zoliflodacin, 1259
Zollinger-Ellison syndrome (ZES), **2534**
  clinical features of, 2534–2535
  cobalamin malabsorption in, 787
  diagnosis of, 2535–2536, 2535–2536t, 2550t
  differential diagnosis of, 2531, 2535
  epidemiology of, 2534
  gastrinoma in, 678
  gastroduodenal ulcer in, 300
  in MEN 1, 2535, 3081
  pathophysiology of, 2534
  treatment of, 2536–2537
  tumor distribution in, 2534
  tumor localization in, 2536, 2536t
Zolmitriptan, 3467t, 3468, 3468t, 3472
Zolpidem, 86, 216, 3713t, 3758
Zombie cells, 3871
Zonisamide, 299t, 300, 3423t, 3424, 3504, 3713t
Zoonotic infection, 1483
Zoster. *See* Herpes zoster (shingles)
Zoster ophthalmicus. *See* Herpes zoster ophthalmicus
Zoster sine herpetica, 1506
Zotarolimus, 2128
ZRSR2 gene mutations, 828
ZRT1, 1701
ZS-9, 361
Z-scores, 3297, 3297f
Zungarococha virus, 1659t, 1667
Zuranolone, 3659t
Zwitterionic characteristics, 1073